# INTRODUCTION: DRUG DISCOVERY IN THE 21st CENTURY

SHAYNE COX GAD

## 1 INTRODUCTION

The discovery, development, and registration of a pharmaceutical is an immensely expensive operation and represents a rather unique challenge. For every 9000 to 10,000 compounds specifically synthesized or isolated as potential therapeutics, one (on average) will actually reach the market. This process is illustrated diagrammatically in Figure 1. Each successive stage in the process is more expensive, making it of great interest to identify as early as possible those agents that are likely not to go the entire distance, allowing a concentration of effort on the compounds that have the highest probability of reaching the market. Compounds "drop out" of the process primarily for three reasons:

1. Toxicity or (lack of) tolerance
2. (Lack of) efficacy
3. (Lack of) bioavailability of the active moiety in humans

Early identification of poor or noncompetitive candidates in each of these three categories is thus extremely important [1], forming the basis for the use of screening in pharmaceutical discovery and development. How much and which resources to invest in screening and each successive step in support of the development of a potential drug are matters of strategy and phasing that are detailed later in this Introduction. In vitro methods are increasingly

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

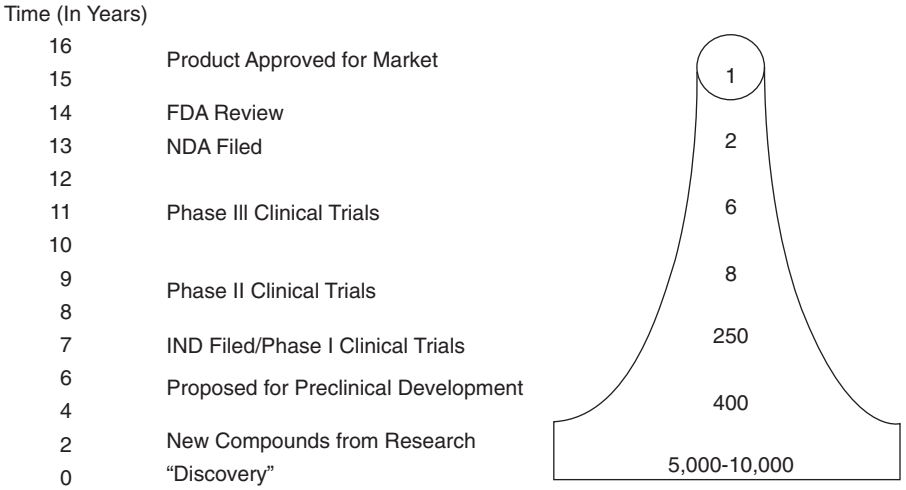


**Figure 1** Attrition during the development of new molecules with a promise of therapeutic potential. Over the course of taking a new molecular entity through scale-up, safety, and efficacy testing, and, finally, to market, typically only 1 out of every 9000 to 10,000 will go to the marketplace.

providing new tools for use in both early screening and the understanding of mechanisms of observed toxicity in preclinical and clinical studies [2, 3], particularly with the growing capabilities and influence of genomic and proteomic technologies. This is increasingly important as the societal concern over drug prices has grown [4]. Additionally, the marketplace for new drugs is exceedingly competitive. The rewards for being either early (first or second) into the marketplace or achieving a significant therapeutic advantage are enormous in terms of eventual market share. Additionally, the first drug approved sets agency expectations for those drugs that follow. In mid-2004, there were 263 pharmaceutical products awaiting approval (93 of these biotech products)—the "oldest" having been in review 10 years) and some 2300 additional agents in the IND stage. Not all of these (particularly the oldest) will be economically successful).

The usual way in which transition (or "flow") of new molecules between the different phases is handled in drug discovery/development is to use a tiered screening or testing approach. Each tier generates more specific data (and costs more to do so) and draws on the information generated in earlier tiers to refine the design of new studies. Different tiers are keyed to the support of successive decision points (go/no-go points) in the development process, with the intent of reducing risks (as to efficacy bioavailability and safety) as early as possible.

The first real critical decisions concerning the potential advancement of a compound to evaluation in clinical trials are the most difficult. They require an understanding of how well particular in vitro or in vivo work in predicting

adverse effects in humans (usually very well, but there are notable lapses; for example, giving false positives and false negatives) and an understanding of what initial clinical trials are intended to do. Though an "approved" IND grants one entry into limited evaluations of drug effects in humans, flexibility in the execution and analysis of these studies offers a significant opportunity to also investigate efficacy [5].

Once past the discovery and initial lead or candidate selection stages, each aspect of development becomes more tightly connected with the other aspects of the development of a compound, particularly the potential clinical aspects. These interconnections are coordinated by project management systems. Many times during the early years of the development process, biological evaluation of efficacy and safety constitutes the rate-limiting step—it is, in the language of project management, on the critical path.

Another way in which pharmaceutical development varies from toxicology as practiced in other industries is that it is a much more multidisciplinary and integrated process. This particularly stands out in the incorporation of the evaluation of ADME (absorption, distribution, metabolism, and excretion) aspects in the safety evaluation process. These pharmacokinetic/metabolism (PKM) aspects are evaluated for each of the animal model species utilized to predict the safety of a potential drug prior to evaluation in humans. Frequently, in vitro characterizations of metabolism for model (or potential model) species and humans are performed to allow optimal model selection and understanding of findings. This allows for an early appreciation of both the potential bioavailability of active drug moieties and the relative predictive values of the various biological models. Such data early on are also very useful (in fact, sometimes essential) in setting does levels for later animal studies and in projecting safe dose levels for clinical use. Unlike the case in most other realms of development of biologically active molecules, one is not limited to extrapolating the relationships between administered dose and systemic effects. Rather, one has significant information on systemic levels of the therapeutic moiety—typically, total area under the curve (AUC), peak plasma levels ($C_{max}$), and plasma half-lives, at a minimum.

## 2 SCREENS: THEIR USE AND INTERPRETATION IN DRUG DISCOVERY

Much (perhaps even most) of what is performed in safety assessment can be considered screening—trying to determine if some effect is or is not (to an acceptable level of confidence) present [6]. The general concepts of such screens are familiar to toxicologists in the pharmaceutical industry because the approach is a major part of the activities of the pharmacologists involved in the discovery of new compounds. But the principles underlying screening are not generally well recognized or understood. And such understanding is essential to the proper use, design, and analysis of screens [7, 8]. Screens are

the biological equivalent of exploratory data analysis, or EDA [9]. Each test or assay has an associated activity criterion, that is, a level above which the activity of interest is judged to be present. If the result for a particular test compound meets this criterion, the compound may pass to the next stage. This criterion could be based on statistical significance (e.g., all compounds with observed activities significantly greater than the control at the 5 percent level could be tagged). However, for early screens, such a formal criterion may be too strict, resulting in few compounds begin identified as "active."

A useful indicator of the efficacy of an assay series is the frequency of discovery of truly active compounds. The frequency is related to the probability of discovery and to the degree of risk (hazard to health) associated with an active compound passing a screen undetected. These two factors in turn depend on the distribution of activities in the series of compounds being tested and the chances of rejecting or accepting compounds with given activities at each stage.

Statistical modeling of the assay system may lead to the improvement of the design of the system by reducing the interval between discoveries of active compounds. The objectives behind a screen and considerations of (1) costs for producing compounds and testing and (2) the degree of uncertainly about test performance will determine desired performance characteristics of specific cases. In the most common case of early toxicity screens performed to remove possible problem compounds, preliminary results suggest that it may be beneficial to increase the number of compounds tested, decrease the numbers of animals per group, and increase the range and number of doses. The result will be less information on more structure, but there will be an overall increase in the frequency of discovery of active compounds (assuming that truly active compounds are entering the system at a steady rate).

The methods described here are well suited to analyzing screening data when the interest is truly in detecting the absence of an effect with little chance of false negatives. There are many forms of graphical analysis methods available, including some newer forms that are particularly well suited to multivariate data (the types that are common in more complicated screening test designs). It is intended that these aspects of analysis will be focused on in a later publication.

The design of each assay and the choice of the activity criterion should, therefore, be adjusted, bearing in mind the relative costs of retaining false positives and rejecting false negatives. Decreasing the group sizes in the early assays reduces the chance of obtaining significance at any particular level (such as 5 percent), so that the activity criterion must be relaxed, in a statistical sense, to allow more compounds through. At some stage, however, it becomes too expensive to continue screening many false positives, and the criteria must be tightened accordingly. Where the criteria are set depends on what acceptable noise levels are in a screening system.

## Characteristics of Screens

An excellent introduction to the characteristics of screens in Redman's [10] interesting approach, which identifies four characteristics of an assay. Redman assumes that a compound is either active or inactive and that the proportion of activities in a compound can be estimated from past experience. After testing, a compound will be classified as positive or negative (i.e., possessing or lacking activity). It is then possible to design the assay so as to optimize the following characteristics:

1. Sensitivity: the ratio of true positives to total activities
2. Specificity: the ratio of true negatives to total inactives
3. Positive accuracy: the ratio of true to observed positives
4. Negative accuracy: the ratio of true to observed negatives
5. Capacity: the number of compounds that can be evaluated
6. Reproducibility: the probability that a screen will produce the same result at another time (and, perhaps, in some other lab)

An advantage of testing many compounds is that it gives the opportunity to average activity evidence over structural classes or to study quantitative structure–activity relationships (QSARs). Quantitative structure–activity relationships can be used to predict the activity of new compounds and thus reduce the chance of in vivo testing on negative compounds. The use of QSARs can increase the proportion of truly active compounds passing through the system.

To simplify this presentation, data sets drawn only from neuromuscular screening activity were used. However, the evaluation and approaches should be valid for all similar screening data sets, regardless of source. The methods are not sensitive to the biases introduced by the degree of interdependence found in many screening batteries that use multiple measures (such as the neurobehavioral screen).

1. Screens almost always focus on detecting a single end point of effect (such as mutagenicity, lethality, neurotoxicity, or development toxicity) and have a particular set of operating characteristics in common.
2. A large number of compounds are evaluated, so ease and speed of performance (which may also be considered efficiency) are very desirable characteristics.
3. The screen must be very sensitive in its detection of potential effective agents. An absolute minimum of active agents should escape detection; that is, there should be very few false negatives (in other words, the type II error rate or $\beta$ level should be low). Stated yet another way, the signal gain should be way up.

4. It is desirable that the number of false positives be small (i.e., there should be a low type I error rate of $\alpha$ level).
5. Items 2 to 4, which are all to some degree contradictory, require the involved researchers to agree on a set of compromises, starting with the acceptance of a relatively high $\alpha$ level (0.10 or more), that is, an increased noise level.
6. In an effort to better serve item 2, safety assessment screens are frequently performed in batteries so that multiple end points are measured in the same operation. Additionally, such measurements may be repeated over a period of time in each model as a means of supporting item 3.
7. This screen should use small amounts of compound to make item 1 possible and should allow evaluation of materials that have limited availability (such as novel compounds) early on in development.
8. Any screening system should be validate initially using a set of blind (positive and negative) controls. These blind controls should also be evaluated in the screening system on a regular basis to ensure continuing proper operation of the screen. As such, the analysis techniques used here can then be used to ensure the quality or modify the performance of a screening system.
9. The more that is known about the activity of interest, the more specific the form of screen that can be employed. As specificity increases, so should sensitivity.
10. Sample (group) sizes are generally small.
11. The data tend to be imprecisely gathered (often because researchers are unsure of what they are looking for) and therefore possess extreme within-group variability. Control and historical data are not used to adjust for variability or modify test performance.
12. Proper dose selection is essential for effective and efficient screen design and conduct. If insufficient data are available, a suitably broad range of doses must be evaluated (however, this technique is undesirable on multiple grounds, as has already been pointed out).

It should be kept in mind that there are a number of common mistakes (in both the design and conduct of studies and in how information from studies is used) that have led to unfortunate results, ranging from losses in time and money to the discarding of perfectly good potential drugs. Such outcomes are indeed the great disasters in drug discovery—especially since many of them are avoidable if attention is paid to a few basic principles.

It is quite possible to design a study for failure. Common shortfalls include:

1. Using the wrong animal model.
2. Using the wrong route or dosing regimen.

3. Using the wrong vehicle or formulation of test material.
4. Using the wrong dose level. In studies where several dose levels are studied, the worst outcome is to have an effect at the lowest dose level tested (i.e., the safe dosage in animals remains unknown). The next worst outcome is to have no effect at the highest dose tested (generally meaning that the signs of toxicity remain unknown, invalidating the study in the eyes of many regulatory agencies).
5. Making leaps of faith. An example is to set dosage levels based on others' data and to then dose all test animals. At the end of the day, all animals in all dose levels are dead. The study is over; the problem remains.
6. Using the wrong concentration of test materials in a screen. Many effects are very concentration dependent.

The design and conduct of discovery screens and programs also require an understanding of some basic concepts:

1. The studies are performed to establish or deny a specific activity of a compound, rather than to characterize the toxicity of a compound.
2. Because pharmaceuticals are intended to affect the functioning of biological systems and safety assessment characterizes the effects of higher-than-therapeutic doses of compounds, it is essential that one be able to differentiate between hyperpharmacology and true (undesirable) adverse effects.
3. Focus of the development process for a new pharmaceutical is an essential aspect of success but is also difficult to maintain. Clinical research units generally desire to pursue as many or as broad claims as possible for a new agent and frequently also apply pressure for the development of multiple forms for administration by different routes.

This volume will present a wide variety of approaches to the discovery and identification of potential new drugs. In assembling this volume, these approaches were derived thinking of four large categories. This approach is a traditional one, focusing on using some accepted (at least to the researcher) screens for a specific therapeutic activity to identify active or promising structures.

## Therapeutic Area Approach: Diseases Seeking Drugs

This approach is perhaps the most traditional one, focusing on using some accepted (at least to the researcher) screen for a specific therapeutic activity to identify active or promising structure. Most drug discovery in the past started with such mass screening of selected molecules.

### Mechanism Approach: Drugs Seeking Diseases

This approach is still currently popular. One or more compounds possessing a mechanism of activity are evaluated for activity in a specific disease. Or several: The challenge is the plausibility of suitable therapeutic activity and the degree of validility of one by hypothesis for a disease process.

### Medicinal Chemistry Approach

This is also a traditional approach starting with molecular structures of known properties and seeking to modify structure to improve and optimize desirable features. The medicinal chemist utilizes knowledge of what past structural modifications have meant in terms of functional differences.

### Technique-Based Approaches

Since the early 1990s, the most popular approaches have utilized new technologies based on expanded knowledge of how to characterize genetic and molecular level processes in humans and other species. Regrettably, as of late 2004, these methods have not yet borne the expected fruit. But perhaps the necessary period of learning how to use such tools is almost over.

***Genomics*** The field of genomics, particularly high-throughput sequencing and characterization of expressed human genes, has created new opportunities for drug discovery. Knowledge of all the human genes and their functions may allow effective preventive measures and change drug research strategy and drug discovery development processes. Pharmacogenomics (dating back to the 1970s) is the application of genomic technologies such as gene sequencing, statistical genetics, and gene expression analysis to drugs in clinical development and on the market. It applies the large-scale systematic approaches of genomics to speed the discovery of drug response markers, whether they act at the level of the drug target, drug metabolism, or disease pathways. The potential implication of genomics and pharmacogenomics in drug discovery and development is that treatments for diseases could be identified according to genetic and specific individual markers, selecting medications and dosages that are optimized for individual patients.

***Combinatorial Chemistry*** By reacting a set of starting chemicals in every possible combination, combinatorial chemistry gives life scientists the ability to create molecules in huge numbers and to test them for sought-after properties. Those abilities have attracted the attention of pharmaceutical companies. In recent years, virtually every large drug company has set up a combinatorial chemistry group.

The groups have revolutionized drug search programs. Before the mid-1990s, a single chemist could make perhaps 5 compounds per week. Now, he

or she can create up to 100 new compounds in the same period of time. The work helps to find leads for potential drugs and significantly speeds up optimization of those leads.

***Proteomics*** Information can be used to identify proteins associated with a disease, which computer software can then use as targets for new drugs. For example, if a certain protein is implicated in a disease, the three-dimensional structure of that protein provides the information a computer program needs to design drugs to interfere with the action of the protein. A molecule that fits the active site of an enzyme but cannot be released by the enzyme will inactivate the enzyme. This is the basis of new drug discovery tools, which aim to find new drugs to inactivate proteins involved in disease. As genetic differences among individuals are found, researchers will use these same techniques to develop personalized drugs that are more effective for the individual.

## REFERENCES

1. Fishlock, D. (1990, April 24). Survival of the fittest drugs. *Financial Times*, pp. 16–17.
2. Gad, S. C. (1998b). A tier testing strategy incorporating *in vitro* testing methods for pharmaceutical safety assessment. *Humane Innovations Alternatives Animal Experimentation*, *3*, 75–79.
3. Gad, S. C. (2001). *In Vitro Toxicology*, 2nd ed. Taylor and Francis, Philadelphia, PA.
4. Littlehales, C. (1999). The Price of a Pill, *Modern Drug Discovery*, Jan/Feb 1999, pp. 21–30.
5. O'Grady, J., Linet, O. I. (1990). *Early Phase Drug Evaluation in Man.* CRC Press, Boca Raton, FL.
6. Zbinden, G., Elsner, J., Boelsterli, U. A. (1984). Toxicological screening. *Reg. Toxicol. Pharamcol. 4*, 275–286.
7. Gad, S. C. (1988a). An approach to the design and analysis of screening studies in toxicology, *J. Am. Coll. Of Toxicol. 7*(2), 127–138.
8. Gad, S. C. (1989a). Principles of screening in toxicology: With special emphasis on applications to neurotoxicology, *J. Am. Coll. Toxicol. 8*(1), 21–27.
9. Tukey, J. W. (1977). *Exploratory Data Analysis.* Addison-Wesley, Reading, MA.
10. Redman, C. (1981). Screening compounds for clinically active drugs. In C. R. Buncher, J. Tsay (Eds), *Statistics in the Pharmaceutical Industry.* Marcel Dekker, New York, pp. 19–42.

## BIBLIOGRAPHY

Alder, S., Zbinden, G. (1988). *National and International Drug Safety Guidelines.* MTC Verlag, Zollikon, Switzerland.

Angell, M. (2004). *The Truth About the Drug Companies.* Random House, New York.

Beyer, K. (1978). Discovery, development and delivery of new drugs. *Monographs in Pharmacology and Physiology*, No. 12. Spectrum, New York.

FDA (Food and Drug Administration) (1987a). Good laboratory practice regulations: Final rule. 21 CFR Part 58, *Fed. Reg.*, September 4, 1987.

FDA (Food and Drug Administration) (1987b). New drug, antibiotic, and biologic drug produce regulations. 21 CFR Parts 312, 314, 511, and 514, *Federal Register*, *52*(53), 8798–8857.

French, S. (1982). *Sequencing and Scheduling.* Halsted, New York.

Gad, S. C. (2000). *Product Safety Evaluation Handbook*, 2nd ed. Marcel Dekker, New York.

Goozner, M. (2004). *The $800 Million Pill.* University of California Press, Los Angeles.

Guarino, R. A. (1987). *New Drug Approval Process.* Marcel Dekker, New York.

Hamner, C. E. (1982). *Drug Development.* CRC Press, Boca Raton, FL, pp. 53–80.

Kliem, R. L. (1986). *The Secrets of Successful Project Management.* Wiley, New York.

Knutson, J. R. (1980). *How to Be a Successful Project Manager.* American Management Associations, New York.

Leber, P. (1987). FDA: The federal regulations of drug development. In H. Y. Meltzer (Ed.), *Psychopharmacology: The Third Generation of Progress.* Raven, New York, pp. 1675–1683.

Levy, A. E., Simon, R. C., Beerman, T. H., Fold, R. M. (1977). Scheduling of toxicology protocol studies. *Comput. Biomed. Res. 10*, 139–151.

Matoren, G. M. (1984). *The Clinical Process in the Pharmaceutical Industry.* Marcel Dekker, New York, pp. 273–284.

Sneader, W. (1986). *Drug Development: From Laboratory to Clinic.* Wiley, New York.

Spector, R., Park, G. D., Johnson, G. F., Vessell, E. S. (1988). Therapeutic drug monitoring. *Clin. Pharmacol. Therapeu. 43*, 345–353.

Spilker, B. (1994). *Multinational Drug Companies*, 2nd ed. Raven, New York.

Zbinden, G. (1992). *The Source of the River Po.* Haag and Herschen, Frankfurt, Germany.

# NATURAL PRODUCTS

CHARLES B. SPAINHOUR

1 INTRODUCTION
2 HISTORY AND BACKGROUND OF THE USE OF NATURAL PRODUCTS AS THERAPEUTIC AGENTS
3 NATURAL PRODUCTS RESEARCH AND DEVELOPMENT—AN UPDATE
4 DISCOVERY OF NATURAL PRODUCTS
Literature Sources
Environmental Sources
5 ESSENTIAL PHARMACODYNAMICS
Protein Targets of Drug Action
General Principles of Drug Action
Molecular Aspects of Drug Action—Receptors
Molecular Aspects of Drug Action—Ion Channels
Molecular Aspects of Drug Action—G-Protein-Coupled Systems
Molecular Aspects of Drug Action—Receptors as Enzymes
Molecular Aspects of Drug Action—Transcription Factors
Molecular Aspects of Drug Action—Other Targets
Drug Tolerance
6 SCREENING FOR NATURAL PRODUCT ACTIVITY
7 ISOLATION AND PURIFICATION OF NATURAL PRODUCTS
8 STRUCTURE IDENTIFICATION OF NATURAL PRODUCTS
9 SYNTHESIS OF NATURAL PRODUCTS
10 DEVELOPMENT OF NATURAL PRODUCTS
Regulatory Guidelines and Nonclinical Development
Learning from the Mistakes of the Past in the Development of Natural Products
11 FUTURE OF NATURAL PRODUCTS
References

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

## 1 INTRODUCTION

By definition, the word *natural* is an adjective referring to something that is present in or produced by nature and not artificial or man-made. When the word *natural* is used in verbiage or written, many times it is assumed that the definition is something good or pure. However, many effective poisons are natural products [145]. The term *natural products* today is quite commonly understood to refer to herbs, herbal concoctions, dietary supplements, traditional Chinese medicine, or alternative medicine [72]. That will not be the case in this chapter. The information presented here will be restricted to the discovery and development of modern drugs that have been isolated or derived from natural sources. While in some cases, such discovery and development may have been based on herbs, folklore, or traditional or alternative medicine, the research and discovery of, along with the development of, herbal remedies or dietary supplements typically present different challenges with different goals [93, 152]. So while the stories of herbs and drugs are very much intertwined, it needs to be fully appreciated that the use of herbs as natural product therapy is different than the use of herbs as a platform for drug discovery and further development.

## 2 HISTORY AND BACKGROUND OF THE USE OF NATURAL PRODUCTS AS THERAPEUTIC AGENTS

Natural products are generally either of prebiotic origin or originate from microbes, plants, or animal sources [115, 116]. As chemicals, natural products include such classes of compounds as terpenoids, polyketides, amino acids, peptides, proteins, carbohydrates, lipids, nucleic acid bases, ribonucleic acid (RNA), deoxyribonucleic acid (DNA), and so forth. Natural products are not just accidents or products of convenience of nature. More than likely they are a natural expression of the increase in complexity of organisms [76]. Interest in natural sources to provide treatments for pain, palliatives, or curatives for a variety of maladies or recreational use reaches back to the earliest points of history.

Nature has provided many things for humankind over the years, including the tools for the first attempts at therapeutic intervention [115, 116]. Neanderthal remains have been found to contain the remnants of medicinal herbs [72]. The *Nei Ching* is one of the earliest health science anthologies ever produced and dates back to the thirtieth century BC [115, 116]. Some of the first records on the use of natural products in medicine were written in cuneiform in Mesopotamia on clay tablets and date to approximately 2600 BC [29, 30, 115, 116]. Indeed, many of these agents continue to exist in one form or another to this day as treatments for inflammation, influenza, coughing, and parasitic infestation. Chinese herb guides document the use of herbaceous plants as far back in time as 2000 BC [72]. In fact, *The Chinese Materia Medica* has been

repeatedly documented over centuries starting at about 1100 BC [29, 30]. Egyptians have been found to have documented uses of various herbs in 1500 BC [29, 30, 72]. The best known of these documents is the Ebers Papyrus, which documents nearly 1000 different substances and formulations, most of which are plant-based medicines [115, 116]. Asclepius (in 1500 BC) was a physician in ancient Greece who achieved fame in part because of his use of plants in medicine [72]. A collection of Ayurvedic hymns in India from 1000 BC and earlier describes the uses of over 1000 different herbs. This work served as the basis for *Tibetan Medicine* translated from Sanskrit during the eighth century [29, 30]. Theophrastus, a philosopher and natural scientist in approximately 300 BC, wrote a *History of Plants* in which he addressed the medicinal qualities of herbs and the ability to cultivate them. The Greek botanist Pedanious Dioscorides in approximately AD 100 produced a work entitled *De Materia Medica*, which today is still a very well-known European document on the use of herbs in medicine. Galen (AD 130–200), practiced and taught pharmacy and medicine in Rome and published over two dozen books on his areas of interest. Galen was well-known for his complex formulations containing numerous and multiple ingredients. Monks in monasteries in the Middle Ages (fifth to the twelfth centuries) copied manuscripts about herbs and their uses [29, 30, 72]. However, it should not go unrecognized that it was the Arabs who were responsible for maintaining the documentation of much of the Greek and Roman knowledge of herbs and natural products and expanding that information with their own knowledge of Chinese and Indian herbal medicine [29, 30]. The Persian philosopher and physician Avicenna produced a work entitled *Canon Medicinae*, which is considered to be the definitive summarization of Greek and Roman medicine. Li Shih-Chen produced a Chinese drug encyclopedia during the Ming Dynasty entitled *Pen-ts'as kang mu* in AD 1596, which records 1898 herbal drugs and 8160 prescriptions [115, 116]. John Wesley, the founder of Methodism, had a profoundly negative view on the status of physicians within society and in 1747 wrote a book entitled *Primitive Physic*, which was a popular reference book of the time detailing numerous natural cures [72]. When the colonists originally came to America, they lacked trained physicians and so turned to the Native Americans for advice in healing practices. Such a lack of conventional medicine and physicians in early America spawned the production of various types of almanacs and other publications that contained various natural product-based recipes and assorted tidbits of medical information. Indeed, in an effort to curry favor with commoners, physicians themselves turned to the production of self-treatment guides for the general public. Various types of societies and botanical clubs held meetings and published different types of communiqués to educate the public with regard to the availability of natural products and how they could be helpful to an individual's health. Samuel Thompson's *Thompson's New Guide to Health* was one very popular publication. For a variety of different reasons, the interest in natural products continues to this very day [6, 8, 17, 39, 72, 81, 88, 90, 104]. The first commercial pure natural product introduced for therapeutic use is gen-

erally considered to be the narcotic morphine, marketed by Merck in 1826 [118]. The first semisynthetic pure drug based on a natural product, aspirin, was introduced by Bayer in 1899.

## 3 NATURAL PRODUCT RESEARCH AND DEVELOPMENT—AN UPDATE

The World Health Organization estimates that approximately 80 percent of the world's population relies primarily on traditional medicines as sources for their primary health care [44]. Over 100 chemical substances that are considered to be important drugs that are either currently in use or have been widely used in one or more countries in the world have been derived from a little under 100 different plants. Approximately 75 percent of these substances were discovered as a direct result of chemical studies focused on the isolation of active substances from plants used in traditional medicine [29, 30]. The number of medicinal herbs used in China in 1979 has been estimated to be numbered at 5267 [115, 116]. More current statistics based on prescription data from 1993 in the United States show that over 50 percent of the most prescribed drugs had a natural product either as the drug or as the starting point in the synthesis or design of the actual end chemical substance [118]. Thirty-nine percent of the 520 new drugs approved during the period 1983 through 1994 were either natural products or derivatives of natural products [65]. Indeed, if one looks at new drugs from an indication perspective over the same period of time, over 60 percent of antibacterials and antineoplastics were again either natural products themselves or based on structures of natural products. Of the 20 top-selling drugs on the market in the year 2000 that are not proteins, 7 of these were either derived from natural products or developed from leads generated from natural products. This select group of drugs generates over 20 billion U.S. dollars of revenue on an annual basis [60, 65].

Drug development over the years has relied only on a small number of molecular prototypes to produce new medicines [65]. Indeed, only approximately 250 discrete chemical structure prototypes have been used up to 1995, but most of these chemical platforms have been derived from natural sources.

While recombinant proteins and peptides are gaining market share, low-molecular-weight compounds still remain the predominant pharmacologic choice for therapeutic intervention [60]. Just a small sampling of the many available examples of the commercialization of modern drugs from natural products along with their year of introduction, indication, and company are: Orlistat, 1999, obesity, Roche; Miglitol, 1996, antidiabetic (Type II), Bayer; Topotecan, 1996, antineoplastic, SmithKline Beecham; Docetaxel, 1995, antineoplastic, Rhône-Poulenc Rorer; Tacrolimus, 1993, immunosuppressant, Fujisawa; Paclitaxel, 1993, antineoplastic, Bristol-Myers Squibb.

The overwhelming concern today in the pharmaceutical industry is to improve the ability to find new drugs and to accelerate the speed with which

new drugs are discovered and developed. This will only be successfully accomplished if the procedures for drug target elucidation and lead compound identification and optimization are themselves optimized. Analysis of the human genome will provide access to a myriad number of potential targets that will need to be evaluated [60, 65]. The process of high-throughput screening enables the testing of increased numbers of targets and samples to the extent that approximately 100,000 assay points per day are able to be generated. However, the ability to accelerate the identification of pertinent lead compounds will only be achieved with the implementation of new ideas to generate varieties of structurally diverse test samples [60, 65, 66]. Experience has persistently and repeatedly demonstrated that nature has evolved over thousands of years a diverse chemical library of compounds that are not accessible by commonly recognized and frequently used synthetic approaches. Natural products have revealed the ways to new therapeutic approaches, contributed to the understanding of numerous biochemical pathways and have established their worth as valuable tools in biological chemistry and molecular and cellular biology. Just a few examples of some natural products that are currently being evaluated as potential drugs are (natural product, source, target, indication, status): manoalide, marine sponge, phospholipage-$A_2$ $Ca^{2+}$-release, anti-inflammatory, clincial trials; dolastatin 10, sea hare, microtubules, antineoplastic, nonclinical; staurosporine, streptomyces, protein kinase C, antineoplastic, clinical trials; epothilone, myxobacterium, microtubules, antineoplastic, research; calanolide A, B, tree, DNA polymerase action on reverse transcriptase, acquired immunodeficiency syndrome (AIDS), clinical trials; huperzine A, moss, cholinesterase, alzheimer's disease, clinical trials [60].

The costs of drug discovery and drug development continue to increase at astronomical rates, yet despite these expenditures, there is a decrease in the number of new medicines introduced into the world market. Despite the successes that have been achieved over the years with natural products, the interest in natural products as a platform for drug discovery has waxed and waned in popularity with various pharmaceutical companies. Natural products today are most likely going to continue to exist and grow to become even more valuable as sources of new drug leads. This is because the degree of chemical diversity found in natural products is broader than that from any other source, and the degree of novelty of molecular structure found in natural products is greater than that determined from any other source [31, 65, 142].

Where are these opportunities? Well, research into the use of plant-derived natural products alone in just the field of medicine covers a broad spectrum of activities [35, 67, 166, 168, 169]. Examples of such biological activity profiles would include, but are not limited to, nootropics, psychoactive agents, dependence attenuators, anticonvulsants, sedatives, analgesics, anti-inflammatory agents, antipyretics, neurotransmission modulators, autonomic activity modulators, autacoid activity modulators, anticoagulants, hyoplipidemics, antihypertensive agents, cardioprotectants, positive ionotropes, antitussives,

antiasthmatics, pulmonary function enhancers, antiallergens, hypoglycemic agents, antifertility agents, fertility-enhancing agents, wound healing agents, dermal healing agents, bone healing agents, compounds useful in the prevention of urinary calculi as well as their dissolution, gastrointestinal motility modulators, gastric ulcer protectants, immunomodulators, hepato-protective agents, myelo-protective agents, pancreato-protective agents, oculo-protective agents, membrane stabilizers, hemato-protective agents, antioxidants, agents protective against oxidative stress, antineoplastics, antimicrobials, antifungal agents, antiprotozoal agents, antihelminthics, and nutraceuticals [35]. Many frontiers remain within the field of natural products that can provide opportunities to improve our quality of life.

Fungal disease has historically been a difficult clinical entity with which to effectively deal. Fungal diseases can include more than just a mycosis and can also include allergic reactions to fungal proteins and toxic reactions to fungal toxins. Mycoses as a group include diseases that are significantly more serious and life-threatening than nail infestations, athletes foot, or "jock-itch." Indeed, increasing numbers of overtly healthy individuals are becoming victims of the complications of fungal infestation. The reasons for this are that increasing numbers of people are receiving immunomodulatory treatment for an organ transplant or some underlying chronic systemic pathology, antineoplastic chemotherapy for cancer, or have been the recipients of proper or improper use of powerful antibiotics. Additionally there are a number of individuals within society that are infected with the human immunodeficiency virus (HIV). The available drugs to treat mycoses have been limited [5]. Furthermore, in this armamentarium, there are problems with dose-limiting nephrotoxicity, the rapid development of resistance, drug–drug interactions of concern, and a fungistatic mechanism of action. Thus there is an urgent need for the development of more efficacious antifungal agents with fewer limitations and less side effects. Ideally such compounds should possess good distribution characteristics, a novel mechanism of action, and a broad-spectrum cidal antifungal activity. The discovery and isolation of an echinocandin-type lipopeptide (FR901379) and lipopeptidolactone (FR901469) from microbes has been a significant achievement. These compounds are water soluble and inhibit the synthesis of 1,3-β-glycan, a key component of the fungal cell wall. Furthermore, since the cell wall is a feature particular to fungi and is not present in eukaryotic cells, such inhibitors certainly have the potential to demonstrate selective toxicity against the fungi and not against the animal or human host. The ultimate modifications of the lipopeptide and lipopeptidolactone referenced above led to the discovery of micafungin (FK463), which is currently in phase III clinical trials. This work along with the relatively recent approval of caspofungin (Merck) as a therapeutic agent for the treatment of disseminated aspergillosis are significant achievements in that they demonstrate that a melding of the proper research to identify and develop appropriate targets with the chemical and biological diversity found in natural products can be very rewarding.

Much ado has been made over recent years about endocrine disruptors and their effects on humans [33]. It needs to be recognized that endocrine disruptors are not just synthetic chemicals but can also be natural products. The use of natural product endocrine disruptors may provide significant insight into our understanding of the mechanisms by which the evolution of the genome can protect transactivation of the sex hormone receptors and aid in the development of drugs, which can protect the embryo during its development from hormone disruptive effects.

Diabetes is a multisystemic affliction, having impact on nearly every body organ. As a disease, it kills more individuals on a per annum basis than AIDS and breast cancer combined [148]. The impact on the quality of life of an individual suffering with diabetes is profound. A number of natural products currently exist that demonstrate hypoglycemic activity. Indeed, depending upon the source that one might use, there are approximately 800 to 1200 plants that exhibit hypoglycemic activity. While research and development efforts in this particular area thus far are largely restricted to traditional medicine uses, future research may well identify a potent antidiabetic agent.

The incidences of neuropsychiatric disorders are steadily increasing as our population increases in size and age. Such disorders include, but are not limited to, seizure disorders, schizophrenia, dementia, mania, aggression, memory loss, psychoses, age-related cognitive decline, depression, anxiety states, mood disorders, substance abuse, and substance dependence. There is a large body of data available that suggests the use of many natural products as potential treatments for these conditions and other neuropsychiatric disorders [18, 91, 92]. Indeed, a number of plant extracts have been associated with the treatment of various categories of mental symptoms and various types of receptor selectivity [18]. A very controversial potential psychotherapeutic agent is *Gingko biloba* [52]. A lack of understanding of mechanism of action, misidentification of materials, contamination of materials, intrinsic toxicity, and absence of standardization all contribute to this controversy. Further fractionation, isolation, and characterization of active components of these and other plants will undoubtedly lead to the discovery of novel neuropsychiatric agents as well as the debunking of other alleged therapies.

There are numerous blood-based diseases that afflict humans. These would include, but are not limited to, anemia, blood group incompatibility, blood protein disorders, bone marrow diseases, hemoglobinopathies, hemorrhagic diatheses, leukemia, disorders of leukocyte dysfunction, platelet disorders, and erythrocyte aggregation disorders. A number of natural products have been reported in the literature to be of value in the treatment of Epstein-Barr virus infection, leukemia, thrombosis and coagulopathy, malaria, anemia, and bone marrow diseases [113]. Extracts from the fungus *Trichothecium roseum*, the sea cucumber *Cucumaria japonica*, the legume *Amorpha fruitcosa*, the tree *Magnolia officinalis*, and others may be useful in the therapeutic management of Epstein-Barr virus infection. Extracts from the basidiomycetes *Mycena pura* and *Nidula candida* may be useful in the treatment of leukemia. Com-

pounds isolated from *Streptomyces platensis* may be useful in the treatment of thrombocytopenia. Compounds obtained from the marine sponge *Aplysina archeri* have been reported to inhibit the growth of the feline leukemia virus. Scalarane-type bishomo-sesterterpenes isolated from the marine sponge *Phyllospongia foliascens* have been reported to exhibit cytotoxic, antithrombocytic, and vasodilation activities. It should be noted that a number of natural products are based on the coumarin nucleus and as such may exhibit antithrombotic and antiplatelet activities. A number of blood-sucking animals have small, low-molecular-weight proteins in their salivas that interfere with the clotting of blood and therefore might be of value as potential anticoagulants. *Streptomyces hygroscopicus ascomyceticus* manufactures a macrolide that has been reported to have immunosuppressant activity and may prove to be beneficial in preventing transplant rejection in humans. It is entirely possible that these compounds and others offer sufficient structural diversity, range of biological activities, and differing mechanisms of action that new, safer, and more efficacious drugs to treat blood-based disorders could well burgeon from this library.

A wide variety of natural products are claimed to possess immunosuppressant activity, but it is often difficult to dissect this activity away from associated cytotoxicity [101]. Since the first heart transplant in the late 1960s, medicine has progressed to the point where most organ transplants have become relatively routine procedures. The survival of individuals with transplants is owed in large part to the discovery of the fungal metabolite cyclosporine A in 1970 and its widespread use starting in 1978. Indeed, cyclosporine A has achieved such success that it is currently being evaluated for value in the treatment of Crohn's disease, systemic lupus erythematosus, and rheumatoid arthritis. Research efforts abound in the area of natural products and immunosuppression. A methyl analog of oligomycin F isolated from *Streptomyces ostreogriseus* has been reported to quite effectively suppress B-cell activation and T-cell activation in the presence of mitogens at concentrations comparable to that of cyclosporine A. Concanamycin F first isolated from *Streptomyces diastatochromogenes* in 1992 has been found to possess a wide array of biological activities including immunosuppressive and antiviral activities. The experimental immunosuppressant (+)-discodermolide isolated from the marine sponge *Discodermia dissoluta* exhibits relatively nonspecific immunosuppression, causing the cell cycle to arrest during $G_2$ and M phases. This compound's current primary interest is as a potential antineoplastic agent since it stabilizes microtubules and prevents depolymerization, effectively causing cell cyclic arrest during the metaphase to anaphase transition. This same mode of activity is shared with Taxol (Paclitaxel), the epothilones, eleutherobin, and the sarcodictyins. The didemnins, cyclic peptides, were first isolated from the marine tunicate *Trididemnum solidum* and exhibit immunosuppressive activity through a generalized cytotoxicity mediated by inhibition of progression through the $G_1$ phase of the cell cycle by an unknown mechanism. The trichopolyns I to V from the fungus *Trichoderma polysporum* are

lipopeptides that suppress the proliferation of lymphocytes in the murine allogeneic mixed lymphocyte response assay. Triptolide from the plant *Tripterygium winfordii* demonstrates immunosuppressant activity through the inhibition of IL-2 receptor expression and signal transduction. The novel heteroaromatic compound lymphostin, obtained from *Streptomyces* KY11783 has demonstrated immunosuppressant activity through its potent inhibition of the lymphocyte kinase $p56^{lck}$. Over the last decade, research activities on immunosuppressants of natural product origin have focused on the mechanisms of inhibition of T-cell activation and proliferation. This approach has been fruitful, leading to the generation of significant information about signaling pathways between T cells, greater detail about the roles of T cells in immune function, and the discovery of Tacrolimus (Prograf) from the soil fungus *Streptomyces tsukubaensis*. As immunological research progresses, increasingly more potential targets will be elucidated for immunomodulatory therapeutic intervention. Natural products will undoubtedly provide a sound platform for the delivery of natural-product-based therapeutic agent candidates.

Natural-products-based anticancer drug discovery continues to be an active area of research throughout the world [34, 102, 112, 147]. While cancer incidences and the frequencies of types of cancer may vary from country to country, the most common sites for the development of neoplasia are generally considered to be the breast, colon/rectum, prostate, cervix/uterus, esophagus/stomach, pancreas, liver, lung, urinary bladder, kidney, ovary, oral cavity, and blood (leukemia and non-Hodgkin lymphoma) [147]. Currently, the chemotherapeutic management of these tumors involves a variety of different plant-based chemicals that are either currently in use or in clinical trials and include such drug classes as the vinca alkaloids, lignans, taxanes, stilbenes, flavones, cephalotaxanes, camptothecins, and taxanes. Despite the wide range of organ structure, type, and function, great similarities exist between the organs with regard to the pathogenesis of cancer. As more and more details of the molecular biology of cancer are revealed, more targets will present themselves for possible therapeutic chemical intervention in the growth and development of neoplasms. A somewhat new approach is that of cancer chemoprevention, where chemoprevention is defined as the prevention, delay, or reversal of carcinogenesis [112]. A few of the more promising cancer chemopreventive agents are (compound, plant source, target): brusatol, *Brucea javanica*, differentiation; zapotin, *Casimiroa edulis*, differentiation and apoptosis; apigenin, *Mezoneuron cacullatum*, antimutagenesis; deguelin, *Mundelea sericea*, inhibitor of ornithine decarboxylase; brassinin, *Brassica* spp., inducer of quinine reductase; and resveratrol, *Cassia quinquangulata*, cyclooxygenase inhibitor. A final note with regard to this approach is that it is important to appreciate that the distinction between chemopreventive agent and chemotherapeutic agent can become quite blurred.

A recurrent theme in neoplasia is the alteration of cell cycle control. One therapeutic approach to the treatment of neoplasia is the development of a treatment that would return to normal the altered cell cycle [143]. Cyclin-

dependent kinases (CDKs) control the progression of a cell through its growth cycle. CDKs are regulated through a series of site-specific complex mechanisms, and the components of such mechanisms include activating cyclins and endogenous CDK inhibitors. Processes of such mechanisms involve regulatory phosphorylation. There are natural products such as butyrolactone and staurosporine that are currently known to be able to provide such activity. These compounds and others generated from their platform are adenosine 5′-triphosphate (ATP) site-directed inhibitors and directly antagonize the activity of CDKs. Further research should more fully elucidate the most efficacious endpoint of CDK inhibition and lead to the control of neoplastic growth and possibly even bring about cytostasis or apoptosis.

The introduction of active agents derived from natural sources into the anticancer weaponry has already significantly changed the futures of many individuals afflicted with cancer of many different types. Continued research into natural sources will continue to deliver newer and more promising chemicals and chemical classes of anticancer agents with novel mechanisms of action that will improve survival rates to even higher degrees.

Human immunodeficiency virus infection is a devastating, globally widespread disease that consumes significant health-care dollars in the due course of management of patients [79]. Most of the currently useful anti-HIV agents are nucleosides and are limited in use due to severe toxicity and emerging drug resistance. Natural products, with their broad chemical structural diversity, provide an excellent opportunity to deliver significant therapeutic advances in the treatment of HIV [167]. Many natural products with novel structures have been identified as having anti-HIV activities [79, 167]. Betulinic acid, a triterpenoid isolated from *Syzigium claviflorum*, has been found to contain anti-HIV activity in lymphocytes. The quassinoside glycoside isolated from *Allanthus altissima* has been found to inhibit HIV replication. Artemisinin, isolated from *Artemisia anuua*, is a sesquiterpene lactone that is of special interest because of its novel structure, potent antimalarial activity, and activity against *Pneumocystis carinii*. A novel phorbol ester isolated from *Excoecaria agallocba* has been reported to be a potent inhibitor of HIV-1 reverse transcriptase. Indeed, most of the natural product chemicals that are attracting interest in this area of research are secondary metabolites such as terpenes, phenolics, peptides, alkaloids, and carbohydrates and are also inhibitors of HIV reverse transcriptase. Other target opportunities in the life cycle of the human immunodeficiency virus available for exploitation are: (1) attachment of virus to cell surface, (2) penetration and fusion of the virus with the cell membrane, (3) reverse transcription via reverse transcriptase, (4) integration into the host genome, (5) synthesis of viral proteins including zinc fingers, and (6) processing of viral polypeptide with HIV protease and assembly of viral proteins and DNA into a viral particle, maturation, and extrusion of the mature virus [167].

Infectious viral diseases remain a worldwide problem. Viruses have been resistant to therapy or treatment longer than most other forms of life because

their nature is to depend on the cells that they infect for their multiplication and survival [41]. Such a characteristic has made the development of effective antiviral chemotherapeutic agents difficult. Today there are few effective antivirals available for use. In order to confidently wage the war against viruses, research efforts are now turning to the molecular diversity available from natural products. For the period 1983 to 1994, seven out of 10 synthetic agents approved by the Food and Drug Administration (FDA) for use as antivirals were based on a natural product. These drugs are famciclovir, stavudine, zidovudine, zalcitabine, ganciclovir, sorivudine, and didanosine. The viral genome can be composed of either RNA or DNA and HIV, which was discussed earlier is an RNA containing virus. The general potential targets of antiviral chemotherapy are: (1) attachment of virus to host cell, (2) penetration of the host cell by the virus, (3) viral particle uncoating, release, and transport of viral nucleic acid and transport proteins, (4) nucleic acid polymerase release/activation, (5) translation of mRNA (messenger RNA) to polypeptides (early proteins), (6) transcription of mRNA, (7) replication of nucleic acids, (8) protein synthesis (late proteins), (9) viral polypeptide cleavage into polypeptides necessary for maturation, (10) assembly of viral capsids and precursors, (11) encapsidation of nucleic acid, (12) envelopment, and (13) release. Early antiviral research focused on compounds that inhibited viral DNA synthesis, purine, and pyrimidine nucleoside analogs. Today most current antiviral agents target RNA-based viruses and the inhibition of reverse transcriptase in order to block the transcription of the RNA genome to DNA. Such inhibition would prevent the synthesis of viral mRNA and proteins. Protease inhibitors affect the synthesis of late viral proteins and viral packaging activity. There are no currently available drugs that target early viral protein synthesis. Antiviral compound research has included alkaloids, carbohydrates, chromones, coumarins, flavonoids, lignans, phenolics, quinines, xanthones, phenylpropanoids, tannins, terpenes, steroids, iridoids, thiopenes, polyacetylenes, lactones, butenolides, phospholipids, proteins, peptides, and lectins. While plants have been a common hunting ground, many other sources are now starting to be explored, especially the marine environment. The use of natural products in the field of antiviral research appears to be limited only by the imagination of the researcher.

This review has demonstrated that natural products are indeed viable sources and resources for drug discovery and development [3]. Indeed, without natural products, medicine would be lacking in therapeutic tools in several important clinical areas such as neurodegenerative disease, cardiovascular disease, solid tumors treatment, and immunoinflammatory disease [4, 64, 122]. Furthermore, the continual emergence of new natural product chemical structure skeletons, with interesting biological activities along with the potential for chemical modification and synthesis bode well for the utility of natural products. Finally, the uses of natural products need to be by no means restricted to pharmaceuticals but can also be expanded to agrochemicals. For example, the use of pyrethrins obtained from *Chrysanthemum* spp. as insecti-

cides has been very popular over the years and persists today. Research continues into the use of natural products as pesticides. While the success stories have not been as numerous or spectacular for herbicides as they have been for drugs and pesticides, there have been victories along the way and the future holds strong potential for this field also [40, 99].

## 4 DISCOVERY OF NATURAL PRODUCTS

### Literature Sources

Natural products can come from anywhere. People most commonly think of plants first when talking about natural products, but trees and shrubs can also provide excellent sources of material that could provide the basis of a new therapeutic agent. Animals too, whether highly developed or poorly developed, whether they live on land, sea, or in the air can be excellent sources of natural products. Bacteria, smuts, rusts, yeasts, molds, fungi, and many other forms of what we consider to be primitive life can provide compounds or leads to compounds that can potentially be very useful therapeutic agents. Suffice it to say that natural products can come from any point or level on the phylogenetic tree. When searching for natural products, one should never feel that a form of life is too low, simple, or grotesque to provide a compound of interest. However, before one goes marching out into the woods, sailing out into the sea, climbing the highest mountains, or descending into the deepest caves, it is appropriate to perform a little bit of research, and hence a visit to the library becomes the first step in any search for a natural product. Remember that the use of a natural product as a therapeutic agent requires that one match some particular characteristic of the compound with a disease or condition. This matching process involves a two-tiered process. The first tier can be comprised of a thorough evaluation of the pathophysiologic condition of interest including any pertinent history, etiology, clinical manifestations, biochemistry, clinical chemistry, hematology, physiology, pathology, and therapeutics. With the therapeutic target in mind and a complete understanding of the pathophysiology of the condition, one can then begin a search for a natural product that has some particular characteristic that might suggest that it has utility as a potential therapeutic agent. Alternatively, one could take the approach of observing or finding a particular characteristic of a natural product and then searching for a useful disease or condition to treat with the material. The choice of which path to follow is a personal one, with either selection being equally useful. Admittedly, in the past and not infrequently the search for and investigation of a natural product arose out of serendipity.

There has been an explosion of information in the biomedical sciences over the last 25 years, and attempting to find information about natural products can be a challenging task [39]. Ideally, it is important to know the history, folk-

lore, origin of use, source, chemical structure, availability, method of preparation, pharmacology, toxicology, and therapeutics of any natural product. However, the reality is that many times even for compounds or preparations that have been used for centuries, there are significant gaps in this portfolio of desired information. Nonetheless, a trip to the library can be very useful. It is amazing what research projects individuals have become engaged in over the years. Excellent sources of information on natural products can be readily found on the Internet. Some of these sites available on the Internet are shown in Table 1. Information retrieval services or search engines are also quite useful tools. These, however, can be expensive and require the aid of someone skilled in their use. Some of the more common sources are shown in Table 2. A number of helpful books have been published on natural products over the years, but the preponderance of them dwells on the topic of plants. The more common of these references are shown in Table 3. Finally, periodicals probably represent the most current and timely sources of research and information on natural products. Some of the better and more prominent journals are shown in Table 4.

## Environmental Sources

Myriad opportunities abound throughout nature that can provide natural products with significant therapeutic potential [20, 29, 30, 61, 77, 83]. These opportunities can present themselves from almost any niche of nature and most likely some that have not even yet been discovered.

**TABLE 1 Internet Sites**

| | |
|---|---|
| American Botanical Council<br>*http://www.herbalgram.org* | American Herbalists Guild<br>*http://www.americanherbalistsguild.com/top.htm* |
| Complementary & Alternative Methods<br>*http://www.cancer.org/eprise/main/docroot/eto/eto_5?sitearea=eto* | Dr. Duke's Phytochemical and Ethnobotanical Databases<br>*http://www.ars-grin.gov/duke* |
| Herb Research Foundation<br>*http://www.herbs.org* | Herbal Education Services<br>*http://www.botanicalmedicine.org* |
| Herbal Medicine: Internet Resources: Alternative Medicine<br>*http://www.pitt.edu/~cbw/herb.html* | HerbMed<br>*http://www.herbmed.org/* |
| International Herb Association<br>*http://www.iherb.org* | MEDLINEplus: Alternative Medicine<br>*http://www.nlm.nih.gov/medlineplus/alternativemedicine.html* |
| National Center for Complementary and Alternative Medicine<br>*http://nccam.nih.gov/* | World Health Organization Publications<br>*http://www.who.int/dsa/cat98/trad8.htm* |

Adapted from DerMarderosian and Beutler [39].

**TABLE 2 Information Retrieval Services**

| | |
|---|---|
| BIOSIS<br>*http://www.biosis.org* | Chemical Abstracts Service<br>*http://www.cas.org/* |
| Current Contents and Science Citation Index<br>*http://www.isinet.com/isi* | *Excerpta Botanica. Section A, Taxonomica et Chorologica*, International Association for Plant Taxonomy, G. Fischer, Stuttgart, New York. |
| The Herb Research Foundation<br>*http://www.herbs.org/* | *IPA (International Pharmaceutical Abstracts)*, *http://info.cas.org/ONLINE/DBSS/ipass.html.* Database contains international coverage of pharmacy and health-related literature. |
| *Lynn Index*, Massachusetts College of Pharmacy | *Medicinal and Aromatic Plants Abstracts*, Publications and Information Directorate, Council of Scientific and Industrial Research (CSIR), New Delhi, India. |
| MEDLINE, MEDLARS<br>*http://www.nlm.nih.gov/* | NAPRALERT (NAtural PRoducts ALERT)<br>*http://www.aq.uiuc.edu/~ffh/napra.html* |
| NAPRONET<br>*http://ccl.net/chemistry/resources/ tips/list/NAPRONET/index.shtml* | Poisindex System<br>*infor@mdx.com or http://www.micromedex.com/products/poisindex* |
| Toxicology Information Response Center (TIRC)<br>*http://www.ornl.gov/TechResources/tirc/hmepg.html* | TOXLINE<br>*http://toxnet.nlm.nih.gov/cgi-bin/sis/htmlgen?TOXLINE* |

Adapted from DerMarderosian and Beutler [39].

***Microbes*** The observation of the effects of microbial secondary metabolites on pathogenic fungi and bacteria spawned the antibiotic era [29, 30, 38]. Since its inception, humankind has grown to take for granted the wonders of antibiotics. Indeed, results of antibiotic use were so impressive that compounds of this general type were for the most part the only chemicals used against pathogenic microorganisms. Due to escalating research and development costs, the difficulties in identifying novel structures and the problems in finding new mechanisms of action, the golden era of antibiotics appeared to be meeting its own demise. Many individuals even professed that the use of antibiotics might even be passé, with modern medicine choosing more modern techniques for treatment. However, this same library of antibiotics had, over the same time, also been found to exhibit other biological properties that might be beneficial to humankind. Accordingly, research into the complete biological activity profiles of antibiotics began with the intent of identifying the utility of these compounds for various pharmacological or agrochemical applications. This shift in focus expanded the search for new natural products from microbes, where microbial metabolites might be used to treat diseases other than those caused

**TABLE 3 Books**

| | |
|---|---|
| L. Aikman. *Nature's Healing Arts: From Folk Medicine to Modern Drugs.* Washington, DC: National Geographic Society, 1977. | H. Baslow. *Marine Pharmacology: A Study of Toxins and Other Biologically Active Substances of Marine Origin.* Baltimore: Williams & Wilkins, 1969. |
| J. L. Beal and E. Reinhard, eds. *Natural Products as Medicinal Agents: Plenary Lectures of the International Research Congress on Medicinal Plant Research*, Strasbourg, July 1980. Stuttgart: Hippokrates, c1981. | W. H. Blackwell. *Poisonous and Medicinal. Plants* Englewood Cliffs, NJ: Prentice Hall, 1990. |
| M. Blumenthal, ed. *Herbal Medicine: Expanded Commission E Monographs.* Austin, TX: American Botanical Council, 2000. | M. Blumenthal, ed. *The Complete German Commission E Monographs: Therapeutic Guide to Herbal Medicines.* Austin, TX: American Botanical Council, 1998. |
| M. Bricklin. *The Practical Encyclopedia of Natural Healing*, new rev. ed. Emmaus, PA: Rodale Press, 1983. | *British Herbal Pharmacopoeia.* Great Britain: British Herbal Medicine Association, 1996. |
| W. Bucherl, E. E. Buckley, and V. Deulofeu, eds. *Venomous Animals and Their Venoms*, 3 vols. New York: Academic Press, 1968–71. | M. Castleman. *The Healing Herbs: The Ultimate Guide to the Curative Power of Nature's Medicines.* Emmaus, PA: Rodale Press, 1991. |
| F. Densmore, *How Indians Use Wild Plants for Food, Medicine, and Crafts.* Washington, DC: Government Printing Office, 1928. Reprint, New York: Dover, 1974. | A. H. Der Marderosian and L. E. Liberti. *Natural Products Medicine: A Scientific Guide to Foods, Drugs, Cosmetics.* Philadelphia: G.F. Stickley, 1988. |
| J. A. Duke. *CRC Handbook of Medicinal Herbs.* Boca Raton, FL: CRC Press, 1985. | W. C. Evans, *Trease and Evans' Pharmacognosy*, 14th ed. London: WB Saunders, 1996. |
| C. Facciola. *Cornucopia: A Source Book of Edible Plants.* Vista, CA: Kampong Publications, 1990. | *Foster S. Tyler's Honest Herbal: A Sensible Guide to the Use of Herbs and Related Remedies.* 4th ed. New York: Haworth Herbal Press, 1998. |
| B. W. Halstead. *Poisonous and Venomous Marine Animals of the World*, 2nd rev. ed. Princeton, NJ: Darwin Press, 1988. | G. Henslow. *The Plants of the Bible: Their Ancient and Mediaeval History Popularly Described.* London: Masters, 1906. |
| D. Hoffman. *The Herbal Handbook: A User's Guide to Medical Herbalism.* Rochester, VT: Healing Arts Press, 1998. | D. Hoffman. ed. *The Information Sourcebook of Herbal Medicine.* Freedom, CA: Crossing Press, 1994. |

**TABLE 3** ***Continued***


R. W. Kerr. *Herbalism through the Ages*, 7th ed. San Jose, CA: Supreme Grand Lodge of AMORC, 1980.

J. M. Kingsbury. *Deadly Harvest: A Guide to Common Poisonous Plants.* New York: Holt, Rinehart and Winston, 1965.

J. M. Kingsbury. *Poisonous Plants of the United States and Canada.* Englewood Cliffs, NJ: Prentice-Hall, 1964.

P. Krogsgaard-Larsen, S. B. Christensen, H. Kofod, eds. *Natural Products and Drug Development: Proceedings of the Alfred Benzon Symposium 20 Held at the Premises of the Royal Danish Academy of Sciences and Letters, Copenhagen, 7–11 August 1983.* Copenhagen: Munksgaard, 1984.

K. F. Lampe and M. A. McCann. *AMA Handbook of Poisonous and Injurious Plants.* Chicago: American Medical Association, 1985.

A. Y. Leung. *Encyclopedia of Common Natural Ingredients Used in Food, Drugs, and Cosmetics*, 2nd ed. New York: Wiley, 1996.

W. H. Lewis, and M. P. F. Elvin-Lewis. *Medical Botany: Plants Affecting Man's Health.* New York: Wiley, 1977.

I. E. Liener. ed. *Toxic Constituents of Plant Foodstuffs*, 2nd ed. New York: Academic Press, 1980.

D. J. Mabberly. *The Plant-Book* 2nd ed. Cambridge: Cambridge University Press, 1997.

J. E. Meyer. *The Herbalist*, rev ed. Glenwood, IL: Meyerbooks, 1986.

J. F. Morton. *Atlas of Medicinal Plants of Middle America: Bahamas to Yucaton.* Springfield, IL: C. C. Thomas, 1981.

J. F. Morton. *Major Medicinal Plants: Botany, Culture, and Uses.* Springfield, IL: Thomas, 1977.

P. Ody. *The Complete Medicinal Herbal.* New York: Dorling Kindersley, 1993.

A. Osol and R. Pratt, eds. *The United States Dispensatory*, 27th ed. Philadelphia: Lippincott, 1973.

G. Penso. *Inventory of Medicinal Plants Used in the Different Countries.* Geneva: World Health Organization, 1980.

T. Robinson. *The Organic Constituents of Higher Plants: Their Chemistry and Interrelationships*, 6th ed. North Amherst, MA: Cordus Press, 1991.

F. Rosengarten. *The Book of Spices.* Wynnewood, PA: Livingston, 1969.

P. Schauenberg. *Guide to Medicinal Plants.* New Canaan, CT: Keats, 1977.

J. E. Simon. *Herbs, An Indexed Bibliography, 1971–1980: The Scientific Literature on Selected Herbs, and Aromatic and Medicinal Plants of the Temperate Zone.* Hamden, CT: Shoe String Press, 1984.

D. G. Spoerke. *Herbal Medications.* Santa Barbara, CA: Woodbridge Press, 1990.

**TABLE 3** ***Continued***

| | |
|---|---|
| R. P. Steiner, ed. *Folk Medicine: The Art and the Science.* Washington, DC: American Chemical Society, 1986. | T. Swain, ed. *Plants in the Development of Modern Medicine.* Cambridge, MA: Harvard University Press, 1972. |
| M. Sweet. *Common Edible and Useful Plants of the East and Midwest.* Healdsburg, CA: Naturegraph Publishers, 1975. | M. Sweet. *Common Edible and Useful Plants of the West.* Healdsburg, CA: Naturegraph Publishers, 1976. |
| V. E. Tyler, L. R. Brady, and J. E. Robbers. *Pharmacognosy*, 9th ed. Philadelphia: Lea and Febiger, 1988. | World Health Organization. *WHO Monographs on Selected Medicinal Plants*, Vol. 1. Geneva: World Health Organization, 1999. |
| H. W. Youngken and J .S. Karas. *Common Poisonous Plants of New England.* U.S. Public Health Service Pub. No. 1220. Washington, DC: Government Printing Office, 1964. | M. Zohary. *Plants of the Bible.* New York: Cambridge University Press, 1982. |

Adapted from DerMarderosian and Beutler [39].

by bacteria and fungi. Microorganisms have proven to be an excellent source of novel natural products including polyketide and peptide antibiotics as well as classes of other biologically active compounds [123]. Today, microbial metabolites are used as antineoplastic agents (e.g., mitomycin), immunosuppressive agents (e.g., rapamycin), hypocholesterolemic agents (e.g., pravastatin), enzyme inhibitors (e.g., desferal), antimigraine agents (e.g., ergot alkaloids), herbicides (e.g., bialaphos), antiparasitic agents (e.g., salinomycin), bioinsecticides (e.g., tetranactin), and ruminant growth promoters (e.g., monensin) [38]. It is noteworthy that some of these compounds when originally discovered failed in their development for their original uses as either antibiotics or as agricultural fungicides. Bacteriocins are ribosomally produced antibiotic peptides and proteins that can be subdivided into different categories, lantibiotics, and microcins. Lantibiotics are produced by Gram-positive bacteria and microcins are produced by Gram-negative bacteria. Both lantibiotics and microcins possess an ability to form pores or punch holes in membranes of susceptible microorganisms. This property is of interest to the food industry, as bacteriocins are produced by *Lactococcus* spp., which are used in the preservation of various foodstuffs [123].

The cyanobacterium *Nostoc ellipsosporum* has been found to produce a novel protein (CV-N), which has generated interest because of its viricidal activity and apparent potential as an anti-HIV therapeutic agent. The antiviral activity of this chemical is reported to be mediated through specific interactions with the HIV envelope glycoproteins gp120 and possibly gp41.

**TABLE 4 Periodicals**

| | |
|---|---|
| *American Journal of Natural Medicine* *http://www.impakt.com* | *Botanical Review* *http://www.nybg.org/bsci/spub/botr/frntpg3b.html* |
| *Bulletin on Narcotics* *http://www.odccp.org/bulletin_on_narcotics.html* | *Canadian Journal of Botany* *http://www.nrc.ca/cgi-bin/cisti/journals/rp/rp2_desc_e?cjb* |
| *Canadian Journal of Herbalism* *http://www.herbalists.on.ca/journal/* | *Economic Botany* *http://www.econbot.org* |
| *European Journal of Herbal Medicine* *http://www.ejhm.co.uk/* | *Herb Companion Press* *http://www.interweave.com/* |
| *Herb Quarterly* *http://www.herbquarterly.com* | *International Journal of Aromatherapy* *http://www.harcourt-international.com/journals/ijar* |
| *Journal of Ethnopharmacology* *http://www.elsevier.com/locate/jethpharm* | *Journal of Natural Products* *http://pubs.acs.org/journals/jnprdf/index.html* |
| *Medical Anthropology: Cross Cultural Studies in Health and Illness* *http://www.sfu.ca/medanth* | *Medical Herbalism: A Journal for the Clinical Practitioner* *http://medherb.com/MHHOME.SHTML* |
| *Natural Health* *http://www.naturalhealth1.com* | *Natural Product Letters* *http://www.tandf.co.uk/journals/titles/10575634.html* |
| *Natural Product Reports* *http://www.rsc.org/is/journals/current/npr/nprpub.htm* | *Pharmaceutical Biology* *http://www.szp.swets.nl/szp/frameset.htm* |
| *Phytochemistry: The International Journal of Plant Biochemistry and Molecular Biology, The Journal of the Phytochemical Society of Europe and the Phytochemical Society of North America* *http://www.elsevier.nl/locate/inca/273* | *Phytomedicine: International Journal of Phytotherapy and Phytopharmacology* *http://www.urbanfischer.de/journals* |
| *Phytotherapy Research* *http://www3.interscience.wiley.com/cgi-bin/jtoc?ID=12567* | *Plant Foods for Human Nutrition* *http://www.wkap.nl/jrnltoc.htm/0921-9668/contents* |
| *Planta Medica: Natural Products and Medicinal Plant Research* *http://www.thieme.de/plantamedica/fr_inhalt.html* | *Toxicon: Official Journal of The International Society on Toxicology* *http://www.elsevier.com/locate/toxicon* |
| *Veterinary and Human Toxicology*, American Academy of Veterinary and Comparative Toxicology, Comparative Toxicology Laboratories | *Z Naturforsch* *http://www.ncbi.nlm.nih.gov/entrez/query.fcgi* |

Adapted from DerMarderosian and Beutler [39].

Research has further revealed that CV-N is a new class of antiviral agent because of its unique interaction with envelope glycoproteins.

Biologically active proteins produced from fungi should not be ignored. *Trichoderma viride* produces a polypeptide, alamethicin, that has demonstrated ion-gating activity. Some fungal ribotoxins such as mitogillin have been found to act as specific ribonucleases. The edible mushroom *Rozites caperata* has been found to produce a compound, RC-183, which inhibits herpes simplex virus (HSV-1 and HSV-2) in a murine animal model.

Various compounds isolated from microbes and fungi have gained favor in their utility as tools to investigate such activities as nerve growth and mechanisms of microbial infiltration and pathogenesis. The fruiting body of *Pleurotus ostreatus* has been reported to produce a lectin that demonstrates potent antitumor activity in mice. Similarly, another lectin from *Volvariella volvacea* shows antiproliferative activity against various tumor cell lines via the mediation of a concentration-dependent stimulation of the expression of cyclin kinase inhibitors resulting in cell cycle arrest in the $G_2$/M phase. Many other fungal lectins have been isolated, but complete determinations of their biological activity profiles remain to be completed.

The ingenious application of molecular and cellular biology to detect receptor agonistic and antagonistic activities of compounds has tremendously aided drug research activity. In addition to the many opportunities that exist for microbial secondary metabolites with regard to medicine and agrochemicals, time has also demonstrated the increased need for the development of novel antibiotics because of the development of resistant strains of pathogens, the appearance of new diseases, and the inadequacy or toxicity of current drugs [37]. The diversity of microorganisms is of a staggering quantity, and only an extremely small proportion of bacteria and fungi have been examined for the production of potentially useful secondary metabolites.

***Plants*** The higher plants produce a variety of different types of compounds, including biologically active proteins. Some of these types of compounds are even shared with other organisms, and they include such chemical families as lectins, defensins, cyclotides, and ribosome-inactivating proteins [123]. Ribosome-inactivating proteins are a group of proteins exhibiting a wide spectrum of biological activities, including a ribonucleolytic activity for which the group is named. These compounds can be obtained from *Panax ginseng* and other plants and have been reported to demonstrate antifungal and antiviral activities. Ribosome-inactivating proteins from *Phytolacca americana* have been reported to be active against HIV and from *Saponaria officinalis* to possess antineoplastic activity. Plant antimicrobial peptides comprise another large group of biologically active compounds. This group of compounds can be further subdivided into thionins, defensins, cyclotides, and lectins. Thionins are small proteins that selectively form disulfide bridges with other proteins or form ion channels in membranes. This ability to make membranes more permeable suggests the potential for antimicrobial activity. Defensins are

cysteine-rich peptides that also permeabilize membranes but appear to be very specific in their activity, targeting fungal cell membranes and not mammalian or bacterial cell membranes. Cyclotides are a protein family, whose mechanism of action has not yet been elucidated, but have demonstrated inhibitory activity against HIV-1. Finally, the lectins are proteins that have a noncatalytic domain that binds reversibly to specific carbohydrates. This activity encompasses potentially a wide spectrum of biological activities including antineoplastic activity, immunostimulation activity, immunosuppression activity, antimicrobial activity, and antimicrobial activity.

Higher plants have been over time an extremely popular source of natural products [23, 95, 123, 127, 130]. Since 1961 approximately nine different compounds derived from plants have been approved in the United States as antineoplastic agents [95]. These drugs include vinblastine, vincristine, vinorelbine, etoposide, teniposide, paclitaxel, docetaxel, topotecan, and irinotecan. The mechanisms of action of these compounds ranges from that of tubulin inhibition to the inhibition of the essential DNA enzymes topoisomerase I and topoisomerase II or both toposimoerases I and II.

*Podophyllum peltatum* reportedly has curative properties on the venereal wart, *Condyloma acuminatum*, along with other therapeutic benefits [130]. The active glycoside component of extracts from *P. peltatum* has been found to inhibit mitosis in vitro. Furthermore, derivatives of this compound have been found to be capable of arresting cells in either late S phase or early $G_2$ phase, without inhibiting microtubule assembly. The latex from the plant *Euphorbia lateriflora* has been used both as a purgative and a cure for ringworm [23]. The Pacific yew tree yielded the antineoplastic, paclitaxel [29, 30]. *Newboutonia vellutina*, a Euphorbiaceae, has been reported to have utility both as a parasiticide as well as a treatment for gastric distension [23]. Compounds isolated and identified from this source will undoubtedly continue to make strong contributions to modern therapeutics.

***Insects*** Various polypeptides of interest have been isolated from the venoms of arachnids and arthropods that prey on insects [123]. Indeed a variety of reviews have been published on ion channel toxins from scorpions and specific venom proteins and neurotoxins from arthropod venoms and their effects on the cardiovascular system. The caterpillar *Lonomia achelous* has been reported to be the source of a biologically active protein that causes a coagulopathy that is mediated via specific interactions with Factor V in the coagulation cascade. While this discovery requires further work to evaluate its therapeutic potential, the discovery may open new opportunities in thrombosis research. Insect peptides have been the subject of research into the immune defense system of insects but have not yet been investigated for effects and potential benefit in humans. Compounds from this peptide group include such sources as the termite (*Pseudacanthotermes spiniger*), the mosquito (*Anopheles gambiae*), the moth (*Heliothis viriscens*), and the beetle (*Oryctes rhinoc-*

*eros*). Insect-derived natural products offer another strong potential avenue for the development of future drugs.

***Vertebrates*** Research into a variety of antimicrobial peptides, such as magainins, defensins, cathelicidins, and protegrins generated by vertebrates has over recent time become very popular. Cathelicidin-type peptides are a broad range of antimicrobial proteins that have been isolated from rabbits, mice, sheep, and humans. Cathelicidins are composed of two different domains, the cathelin and antimicrobial domains. The cathelin domain becomes bactericidal after cleavage from the antimicrobial domain. Purportedly, these materials bind to lipopolysaccharide and neutralize its activity. The pit viper, *Bothrops jaracaca*, produces a compound that spurred the synthesis of the angiotensin-converting enzyme (ACE) inhibitors captropril and enalpril [29, 30]. The skin of the poisonous frog, *Epipedobates tricolor*, produces epibatidine. This substance has ultimately led to the creation of a new class of analgesics. Other vertebrates produce compounds that have been the subjects of substantial research, which offer potential opportunities to identify useful compounds in the areas of cardiovascular function, immune function, and central nervous system function.

***Marine Organisms*** The marine environment, arguably the original source of all life, is a rich source of bioactive compounds [14, 45–48, 84, 89, 96, 144, 157]. More than 70 percent of our planet's surface is covered by the oceans, and some experts feel that the potentially available biodiversity on the deep seafloor or coral reefs is greater than that existing in the rainforests [62].

Consider the fact that many marine organisms have soft bodies and lead a sedentary lifestyle, making a chemical system of defense almost essential for survival. Marine organisms have evolved the ability to synthesize such toxic compounds or extract or convert pertinent compounds from other marine microorganisms. Natural products from marine organisms are released into the water and therefore are rapidly diluted, accordingly they must be very potent materials to have the desired end effect. The richly available marine biodiversity that is available to us has to this point only been explored to an extremely limited extent. Furthermore, the primary chemical diversity available from marine organisms is most likely capable of delivering an even greater abundance of secondary metabolites for research use. For all of these reasons it is believed that the natural products that are available from the seas and oceans provide a tremendous opportunity for the discovery of novel therapeutic agents [62].

The first discovery of a marine-based biologically active compound of therapeutic interest was really quite by accident approximately 10 years after the end of the World War II [29, 30]. The C-nucleosides isolated from the Caribbean sponge *Cryptotheca crypta* were found to possess antiviral activity. This discovery eventually led to the development of cytosine arabinoside, a

useful antineoplastic agent. Biologically active marine proteins derived from the venom of marine snails of the *Conus* genus have attracted a significant level of research over the years [123]. These conotoxin peptides interact in a unique fashion with voltage-gated ion channels to induce a wide spectrum of pharmacological effects. Such effects include anesthesia, analgesia, and anticonvulsant activity. The conotoxin ziconotide is currently under review in the United States for use in the treatment of chronic, opiate-resistant pain. According to some estimates, there are most likely approximately 1000 different *Conus* snails. Each snail produces up to approximately 200 different venoms. The broad spectrum of biological activities manifested by each of these venom components multiplied by the number of snails and venom components available suggests significant opportunity for new drugs from the snail alone. The mussel *Mytilis edulis* has been reported to produce antibacterial peptides and cytotoxic lectins. Horseshoe crabs produce a variety of different antibacterial peptides and proteins. Indeed, *Limulus polyphemus* produces an interesting group of antimicrobial peptides referred to as polyphemusins, and a synthetic peptide based on the sequence of polyphemusin II has been reported to strongly inhibit the cytopathic effect of infection with HIV.

It has been reported that the tunicate *Styela clava* produces α-helical antimicrobial peptides called clavanins that are homologous to the magainins produced by certain types of frogs. The marine worm *Cerebratulus lacteus* produces neurotoxic polypeptides, which have been found to contain the ability to make membranes more permeable. Sponges have been found to produce a wide variety of interesting secondary metabolites. For example, lectins isolated from *Chondrilla nucula* have achieved utility in the histochemical labeling of melanoma and breast and thyroid carcinomas. Another compound, the protein mapacalcine, produced by *Cliona vastifica*, has been reported to specifically block non-L-type calcium channels in murine duodenal myocytes, while at the same time not exhibiting any affect on T-type calcium flux or potassium or chloride currents. Various proteins from sponges have been reported to selectively kill human tumor cells. For example, a protein from the sponge *Tethya ingalli* lyses ovarian cancer cells; a glycoprotein from *Pachymatisma johnstonii* has been discovered to inhibit cell growth at the $G_0/G_1$ phase via a unique mechanism in a non-small-cell-bronchopulmonary carcinoma line. Interestingly, this same chemical has been found to have potential as an antiparasitic in the treatment of leishmaniasis, by demonstrating cytotoxic activity against the parasite in the promastigote and amastigote stages of the life cycle. The protein niphatevirin isolated from the sponge *Niphates erecta* has been discovered to inhibit HIV-induced cytopathic effects, cell-to-cell fusion, and syncytium formation via interacting directly with the CD4 cellular receptor. Finally, a compound isolated from the sponge *Microciona prolifera* has been reported to bind to gp120 and resultantly protect T-lymphoblastoid cells from infection with HIV.

The so-called mining of the sea for potential drugs did not start in earnest until the mid-1970s because of a technical inability to effectively gain access

to the biodiversity that exists within the seas and oceans. Despite this slow start, the potential for contribution of new drugs is staggering, a vision that undoubtedly will become reality over the next decade. Examples of such research activity and fulfillment of this future hope are (chemical, source, chemical class, chemical target, therapeutic Indication): AM336 (AMRAD), cone snail, peptide, ion channels, chronic pain; GTS21 (Taiho), nemertine worm, anabaseine-derivative, ion channels, Alzheimer's disease and schizophrenia; LAF389 (Novartis), sponge, amino acid derivative, methioninc aminopeptidase inhibitor, cancer; OAS1000 (OsteoArthritis Sciences), soft coral, diterpene-pentoseglycoside, $PLA_2$ inhibitor, wound healing and inflammation; ILX651 (Ilex Oncology), sea slug, peptide, microtubule-interference, cancer; Cematodin (Knoll), sea slug, peptide, microtubule interference, cancer; Yondelis, sea squirt, isoquinolone, DNA-interactive agent, cancer; Alipidin, sea squirt, cyclic depsipeptide, oxidative stress inducer, cancer; Kahalalide F, sea slug/alga, cyclic depsipeptide, lysosomotropic compound, cancer; KRN7000 (Kirin), sponge, α-galactosylceramide, immunostimulatory agent, cancer; squalamine lactate, shark, aminosteroid, calcium-binding protein antagonist, cancer; IPL512602 Inflazyme/Aventis], sponge, steroid, unknown, inflammation [62].

## 5 ESSENTIAL PHARMACODYNAMICS

Modern science has provided a detailed understanding of the interaction of many therapeutic drugs with biological systems at a biochemical or molecular biological level. The proper use of such information can provide a wealth of tools for the discovery of new drug opportunities by providing a framework to permit comparison of the mechanisms of action, biological activities, therapeutic indices, and the therapeutic potentials of drugs along with the ability to forecast possible problems [7, 87, 121, 131, 134, 137, 159].

### Protein Targets for Drug Action

The actions of drugs can be divided into those occurring at specific sites and those that are nonspecific. Nonspecific effects are typically mediated through a generalized effect in many organs, and the response observed depends upon the distribution of the drug. The response is usually associated with the organ or organs having the highest concentrations of the drug. Specific effects are produced by an interaction of the drug with a specific site or sites either on the cell membrane or inside of the cell. The protein targets for drug action on mammalian cells can be broadly divided into receptors, ion channels, enzymes, transcription factors, and other nonspecific sites of action. The term *target* can also be broadly defined to include such things as microorganisms and cancer cells, but this discussion will be limited to normal mammalian systems or cells. It should be noted that while there are other types of protein that are known

to function as drug targets, it must be appreciated that there exist many drugs whose sites of action have not yet been elucidated in detail. Furthermore, many drugs are known to bind to plasma proteins as well as to various cellular constituents, without producing any obvious physiological effect. Nevertheless the above generalization that most drugs act on one of the five types of protein targets listed above is a reasonable initial working classification system.

## General Principles of Drug Action

There are a variety of different types of drug action. Accordingly, drugs can be classified into specific categories such as agonists, antagonists, partial agonists, inverse agonists, allosteric modulators, enzyme inhibitors or activators, and those having nonspecific action.

Agonists bind to a receptor or site of action and produce a conformational change in the receptor or that site, which mimics the action of the normal physiologic binding ligand. At low concentrations, the activity of the drug is additive with the natural ligand. Drugs can differ in both their affinity or strength of binding and the rate of association and dissociation from the receptor or binding site. The affinity or strength of binding of the drug to the receptor ultimately determines the concentration necessary to produce a response and therefore is directly related to the potency of the drug. For some compounds a maximal response may of necessity invoke the contribution of all available receptors, but for most drugs a maximal response is produced while some receptors remain unoccupied. The presence of spare receptors becomes an important point when considering changes in the numbers of available receptors resulting from adaptive responses occurring in response to either chronic exposure to a drug or the irreversible binding of an antagonist. The rate of binding or dissociation of a drug to a receptor or site of action is generally of little importance in determining the rate of onset or termination of a drug's elicited effect in vivo because such behavior depends mostly on the rate of delivery to and removal from the target organ, in other words, the rates of absorption and elimination of the drug from the body. The effect of changes in the numbers of receptors on the dose–response curve for an agonist depends on the potency of the agonist, receptor occupancy by the agonist, and the efficacy of the agonist.

Antagonists bind to a receptor but do not elicit the necessary conformational change required to produce the normal response–effect. These types of compounds will block access to the receptor or binding site by the normal physiologic ligand. It is important to keep in mind that antagonist-induced effects may only be observable when the normal agonist or ligand is present. The binding of most clinically useful antagonists is both reversible and competitive. Consequently, typically a receptor or site of action blockage can be overcome by increasing the concentration of the natural ligand or another agonist drug. Most antagonists shift the dose–response

curve to the right but do not alter the magnitude of the maximum possible response.

Partial agonists demonstrate both agonist and antagonist activities. For these types of drugs, the activity demonstrated by the drug is a function of the concentration of the natural ligand or agonist. Maximal binding of the partial agonist to a receptor will produce only a submaximal response. This is most likely the result of incomplete amplification of the receptor signal via G proteins. A partial agonist will demonstrate agonist activity at low concentrations of the natural ligand, but the dose–response will not attain maximal activity even when all of the receptors are occupied. Alternatively, at high concentrations of the natural ligand a partial agonist will behave as an antagonist because it will prevent the access of the natural ligand to the receptor, thereby preventing a maximal response.

Inverse agonists act in such a fashion on receptors as to produce a change opposite to that caused by an agonist. While not totally understood, the discovery of this phenomenon has given rise to the theory that receptors exist in equilibrium between active and inactive forms in the absence of an agonist ligand. The presence of an agonist will increase the proportion of active receptors, but the presence of an inverse agonist will shift the balance toward the more inactive receptors, thereby reducing the level of basal activity. Antagonists, when bound to a receptor will block the activity not only of agonists but also of inverse agonists. The mechanism of action of inverse agonists is not well characterized but may involve the destabilization of receptor–G protein coupling. Inverse agonists may preferentially bind to the inactivated form of the receptor, shifting the equilibrium away from the active form. An additional complication to this whole concept is the fact that compounds that are normal antagonists for some tissue receptors act as inverse agonists for the same receptor in a different tissue. The concept of inverse antagonism and its use in the therapeutic utility of drugs remains to be fully understood and better characterized.

Allosteric modulators do not act directly on a ligand/receptor site but may bind elsewhere on the receptor to enhance or decrease the binding of the natural physiologic ligand to the receptor. Some drugs have a target site of action that is an enzyme, and these compounds can effect their action on either the catalytic site or elsewhere on the molecule at an allosteric site. Finally some compounds merely elicit a broad and generalized effect that causes the desired therapeutic outcome. Examples of such activity might be that of osmotic diuretics and their ability to induce diuresis through the general action of osmosis or general anesthetics, which modulate neuronal cell membrane fluidity.

Drug action also involves the demonstration of a number of different important properties. These are specificity, selectivity, potency, and efficacy. Specificity refers to the fact that many drugs act only at one type of receptor (e.g., cholinergic versus noradrenergic). Drugs that are not specific in their action for one type of receptor display a wide variety of not necessarily desirable side effects.

The site of action of a drug may in some cases involve one or more members of a family or group of receptors. In such a situation, where the drug may act preferentially with one member of the group, it is said to be selective in its action. Alternatively, the drug may show a similar affinity for more than one member of a group or family of receptors and is therefore referred to as being nonselective. Drugs may demonstrate a predilection for a particular receptor type or subtype and may bind to a different receptor type or subtype to different degrees. In these situations, it is possible to determine the individual dose–response relationships for each receptor type or subtype. The selectivity of a drug is a measure of the degree of separation of the individual dose–response curves for different receptor types or subtypes. Ultimately, the expression of selectivity is dependent upon the dose and the concentration of the drug at the different receptors. High concentrations of an agonist drug will typically generate a maximal occupancy of all of the involved and available receptor types or subtypes with little to no discrimination or selectivity. High concentrations of an agonist drug can even create a blockade of the receptor. One should always be careful to include the appropriate qualifications when referring to selectivity and nonselectivity because populations of receptors can be related or unrelated to each other. For example, one can speak of the selectivity of a drug for $\alpha$-adrenoreceptors as compared to $\beta$-adrenoceptors or $\beta_1$-adrenoceptors as compared to $\beta_2$-adrenoceptors.

The in vitro potency of a drug is determined by the strength of its binding to a receptor. This is also referred to as the affinity of the drug for the receptor. The more potent a drug is, the lower will be the concentration required to affect its binding to a receptor and to provide a response for an agonist or to block a response for the case of an antagonist. Potencies of different drugs are generally compared or contrasted using ratios of the different doses required to either elicit or block an equivalent response. Dose–response curves are typically S-shaped. If the mechanism of action is identical or very similar for the drugs being compared, the linear or midportions between the lower and upper plateaus of the dose–response curve are commonly found to be parallel. To make comparisons between different points on distinct dose–response curves more obvious and facile, a linearizing transform can be utilized for each sigmoid curve. A popular technique for linearizing data is the logit plot. Potencies are generally not determined by comparing responses at identical or similar doses. This is because with the selection of a single administered dose, inherent differences between individual drugs will cause too large of a difference in drug responses. Rather, relative potencies are determined by an evaluation of the ratio of the doses for different drugs that produce equivalent responses on the respective drug dose–response curves. For example, the doses that cause 50 percent inhibition of the target activity on various dose–response curves could be compared for different drugs. Drugs that demonstrate in vitro the highest affinities for a receptor are generally the most potent. However, the in vivo dose–response relationship will be a function of the delivery of the drug to the site of action. Such delivery is related to the absorption, distribu-

tion, and elimination (pharmacokinetics) of a drug. Therefore, it is important to keep in mind that for a series of related drugs in vivo behavior may not always reflect in vitro receptor binding properties.

The efficacy of a drug is its ability to produce the maximum possible response. For example, drugs can be classified as full agonists or partial agonists. Full agonists produce an increase in response with increase in concentration up to that point at which the maximum possible response is elicited. A partial agonist will produce an increase in response with an increase in concentration, but it cannot produce a maximal possible response.

## Molecular Aspects of Drug Action—Receptors

The activities of most cellular processes are highly controlled in order for the cell to exist under optimum homeostatic conditions not only when at rest but also in response to the myriad of physiological and metabolic demands that are placed upon it. These reflex actions require specific responses appropriate for a given cell type to signals elicited as a result of a change in physiological state. These signals most commonly are specific chemicals that are released into the general circulation or are locally released but then are recognized by a targeted or specific cell. Ultimately the condition of homeostasis is based upon (a) the generation of a chemical signal, (b) the recognition of a chemical signal, and (c) the generation of an appropriate response or cellular change(s).

A chemical signal eventually binds to a specific type of macromolecule of a cell. This binding in and of itself then triggers a cellular response. The chemical signal is referred to as the ligand. The cellular macromolecule is referred to as the receptor. Receptors can be located in different places. They may be located in the cell membrane in order to accept extracellular ligands that cannot cross or only cross with difficulty the cell membrane. Alternatively, they can be located in the cytoplasm, where they will react with lipid-soluble ligands that can pass through the cell membrane. The binding produces a receptor–ligand complex that will then generate an appropriate cellular response. There are three general types of responses: direct, indirect, and second messenger. A direct response might be the inhibition of a specific process. An indirect response might involve the interaction of the receptor–ligand complex with another moiety (e.g., macromolecule) to continue or complete the process. A second-messenger response would involve the production of an additional chemical signal (second messenger), which would ultimately control a cellular process.

There are different structural and functional classifications of receptors, but generally speaking there are just a few functional families whose members share both common mechanisms of action and similarities in molecular structure. There are at least four main types of receptors: types 1 through 4. The classification of receptors is based on molecular structure and the nature of the receptor–effector linkage. Type 1 receptors are typically located in a mem-

brane and are directly coupled to an ion channel. Type 2 receptors are located in a membrane and are coupled by a G protein to an enzyme or channel. Type 3 receptors are located in membranes and are directly coupled to an enzyme. Finally, type 4 receptors are located in the nucleus or cytosol and are coupled via DNA to gene transcription.

Although there are some exceptions, ligands typically bind to receptors in a reversible fashion. Accordingly, the intensity and duration of the response generated from the binding of the ligand to the receptor is a function of the lifetime of the ligand–receptor complex. The interaction(s) between a receptor and a ligand most commonly do not involve the formation of permanent covalent bonds, but rather do involve the formation of weaker and reversible forces, such as ionic bonding, hydrogen bonding, van der Waals forces, and hydrophobic interactions.

Within a physiologic entity, there are myriad possible extracellular and intracellular chemical signals that are produced that can affect multiple different processes. Subsequently, a very important property of a receptor is its specificity or the extent to which a receptor can recognize, discriminate, and respond to only one signal. Some receptors demonstrate a very high degree of specificity and will bind only a single endogenous ligand, while other receptors are less specific and will bind a number of different endogenous ligands. This ability of receptors to recognize, discriminate, and bind to a given ligand depends on the degree and type of interaction between the receptor molecule and specific chemical structural characteristics of the ligand. Chemical structural differences between ligands may be very subtle in nature. However, receptor specificity occurs because the generation of a fully functional receptor–ligand complex requires the formation of reversible binding interaction between various different molecular sites on the ligand and on the receptor in a special three-dimensional spatial relationship. Receptors themselves are proteins folded into a three-dimensional configuration so that the specific arrangement of reversible interaction sites on the receptor itself is consolidated into a very small volume, referred to as the receptor binding site.

As receptors and ligands each have three-dimensional configurations, the ligand must be presented to the receptor in a very specific three-dimensional configuration to bind and produce a functional receptor–ligand complex. Many drugs exist in different stereoisomeric forms, and each of these various stereoisomers may well exhibit different receptor binding behavior and resulting response. Different isomers can be equally active or some active, some partially active and some even toxic. Furthermore, it is important to realize that if a particular drug exists in equal concentrations of two different isomeric forms, only one of which is active and the other of which is inactive, then only 50 percent of the drug mixture is therapeutic. This has led over the years to the recognition of the value of the separation and purification of individual isomers of drugs for development as potential therapeutic agents.

Different types of receptors recognize and bind to different ligands. Furthermore, there may be different subtypes of a given receptor, each of which

recognizes or binds to the same specific ligand but generates different intracellular responses. Various receptor subtypes are often found in different tissues, organs, or distribution patterns in an organ and hence can produce different end effects. The existence of receptor subtypes creates the opportunity for specific drugs to produce extremely selective actions with fewer unwanted side effects.

At any given time, the number of receptors in a cell is not static but rather is a dynamic. There is a high turnover of receptors as they are continuously removed and replaced. Drug treatment itself can either increase the number of receptors, a phenomenon that is called upregulation, or decrease the number of receptors, a phenomenon referred to as downregulation. This potential change in numbers of receptors can be an important aspect of drug treatment and management of a clinical case to achieve a desired therapeutic response.

## Molecular Aspects of Drug Action—Ion Channels

Receptors for several neurotransmitters exist as agonist-regulated, ion-selective channels in the plasma membrane of a cell and are referred to as ligand-gated ion channels. These receptors send their signals by altering a cell's membrane potential or its ionic composition. This group of receptors includes nicotinic cholinergic receptors, γ-aminobutyric acid receptors (subtype A), glycine receptors, aspartate receptors, glutamate receptors, nicotinic cholinergic receptors, and 5-hydroxytryptamine (subtype 3) receptors. These receptors are all multiple subunit proteins. Each protein subunit spans the cell membrane, and the subunits are arranged symmetrically in such a fashion as to form a channel. This channel opens and closes upon proper stimulation as a result of specifically induced molecular structural changes.

While some ion channels are linked to a receptor and open only when the receptor is occupied by an agonist or ligand, there are other types of ion channels, which themselves serve as targets for drug action. This type of interaction can be indirect, involving other intermediates, but the interaction can also be direct with the behavior of the channel being modulated by the binding of a drug directly to a part or parts of the channel protein. The simplest type of ion channel blocking involves a physical chemical barricade of the channel opening by the drug itself. More complex types of direct ion channel blocking involve drug–channel protein interactions. Normally, an ion channel opens in response to membrane depolarization. In these situations, a drug, by binding with ion channel protein alters or modulates the probability of the channel opening. The degree of inhibition or facilitation of the opening of the ion channel is a function of the chemical structure of the drug itself. In this latter case, a drug actually affects the gating of the channel, while in the former case the drug actually blocks penetration of the channel. The modulation of ion channels by drugs is a very important mechanism through which pharmacologic effects are mediated at the cellular level.

## Molecular Aspects of Drug Action—G-Protein-Coupled Systems

There is a large family of receptors that utilize heterotrimeric guanosine 5′-triphosphate (GTP)-binding regulatory proteins. These regulatory proteins, known as G proteins, behave as transducers to send signals to their effector proteins. Ligands for G-protein-coupled receptors include eicosanoids, a variety of lipid signaling molecules, various peptides, different proteins, and a host of biogenic amines. Effectors for G-protein-regulated receptors include adenyl cyclase, phospholipase $C_\beta$, $Ca^{2+}$ currents, rho GTP exchange catalysts, inward-rectifying $K^+$ currents, and phosphatidyl inositol-3-kinase. G-protein receptors occur widely in nature and are widely used drug targets.

G-protein-coupled receptors span the cell membrane and exist as a bundle of seven α helices. A cleft exists in the three-dimensional configuration of these seven α helices, which binds agonists on the extracellular face of the receptor. Alternatively, agonists can also bind to a globular ligand-binding domain located on the extracellular face of the G-protein-coupled receptor. The G proteins themselves bind to the cytoplasmic face of the receptor. These receptors respond to the binding of an agonist by promoting the binding of GTP to the G protein. The binding of the GTP to the G protein activates the G protein, which in turn activates the effector protein. The G protein remains in an activated state until it hydrolyzes the bound GTP to guanosine 5′-diphosphate (GDP). G proteins are composed of an α-GTP-binding subunit, which is specifically recognized by the receptor and an associated dimmer of β and γ subunits. G proteins are defined by α-subunit composition, such as $\alpha_s$, $\alpha_i$, $\alpha_o$, $\alpha_q$, and $\alpha_{13}$. The activation of the $G_\alpha$ subunit by GTP permits the regulation of an effector protein and promotes the release of $G_{\beta\gamma}$ subunits, which in turn regulate their own group of effectors.

There are many different G-protein-coupled receptor types in a cell. Each type is specific for one of many different G proteins. Each $G_\alpha$ subunit can in turn regulate one or more effectors. Therefore, receptors that bind to multiple different ligands can focus their signals through a single G protein. Alternatively, a G-protein-coupled receptor is also capable of sending multiple signals through activation of more than one G protein species. It should be appreciated that receptor G-protein effector systems are intricately versatile and complex networks.

## Molecular Aspects of Drug Action—Receptors as Enzymes

Receptors with inherent enzymatic activity are most commonly cell surface protein kinases. These receptors demonstrate their regulatory activity by phosphorylating various effector proteins at the inner face of the cell membrane. The biochemical reaction of phosphorylation changes the molecular structure, biological properties, and hence the biological activities of an effector or its interactions with other protein molecules. Catalytic activities include tyrosine kinase, tyrosine phosphatase, guanyl cyclase, serine kinase, and threonine

kinase. However, tyrosine residues are the most common substrate. Basic receptor protein kinases are composed of an agonistic-binding domain on the extracellular membrane surface, a single element that spans the cell membrane and a protein kinase domain on the cytoplasmic face of the cell membrane. Other variations exist of this basic model and include oligomerization and/or the elaboration of other regulatory or protein binding domains to the cytoplasmic portion of the receptor. Some protein kinase receptors lack the attached intracellular enzymic domain, and in response to agonist binding, attract, attach, or link to and activate a distinct soluble and mobile protein kinase on the cytoplasmic side of the cell membrane.

There can be further variation on the above structural theme for receptor enzymes. For example, tyrosine phosphatases possess an extracellular enzymatic domain. In other receptor enzymes, the intracellular domain is not a protein kinase, but rather a guanyl cyclase, which synthesizes a second messenger, adenosine-3′,5′-cyclic monophosphate (cAMP). Other variations of this receptor type may also be possible.

Before leaving this category of receptor, it is appropriate to mention second messengers. A variety of signals within a cell are sent via second-messenger pathways. While there are few cytoplasmic second messengers, their release and presence can affect many different activities. Recognized second messengers include cAMP, nitric oxide, diacylglycerol, $Ca^{2+}$, inositol phosphates, and guanosine-3,5′-cyclic monophosphate (cGMP). Second messengers can exert control on each other either directly or indirectly. Directly, they can alter the metabolism of other second messengers. Indirectly, they can share intracellular targets. Such a complex pattern of regulatory activity permits a cell to respond to the presence of an agonist with an integrated and concerted expression of second messengers and responses. Cyclic AMP is the best known of the second messengers and is synthesized by adenyl cyclase under the control of different G-protein-coupled receptors.

## Molecular Aspects of Drug Action—Transcription Factors

There are receptors for steroid hormones, thyroid hormones, retinoids, vitamin D, and other molecules that are soluble DNA binding proteins that regulate the transcription of specific genes. These receptors are part of a larger family of transcription factors that are regulated by phosphorylation, by association with other proteins, binding to metabolites, or binding to regulatory ligands. These receptors exist as homo- and heterodimers, and their actions and activity can also be regulated by higher order oligomerization with other regulatory molecules.

## Molecular Aspects of Drug Action—Other Targets

Enzymes can also serve as targets for drugs. Typically, the drug is a substrate that is structurally related to the normal substrate and acts as a competitive

inhibitor of the enzyme. Another type of enzyme–drug interaction occurs when a drug is a false substrate. In this type of situation, the drug undergoes the enzyme-mediated chemical transformation to form an atypical product, which then undermines the normal metabolic pathway. This altered function of the metabolic pathway occurs as a result of the creation of a substrate that is not usable for the next step in the metabolic process or ultimately generates a partially functional or nonfunctional end product. It should be mentioned that drugs can also be metabolized by enzymes in such a fashion as to either convert the drug into a toxic reactive intermediate causing significant cellular damage or convert the drug from a totally inert moiety (prodrug) into an active compound with therapeutic benefit.

In addition to the sites and mechanisms of action discussed above, drugs may also bind to and either activate or inhibit other specific sites. These could include specific types of cells, cellular organelles, transport proteins, specific ion pump, and the like. A discussion of all of these additional targets is beyond the scope of this chapter.

## Drug Tolerance

An observed decrease in drug response with repeated doses is commonly referred to as the development of tolerance. Tolerance may occur as a result of a decrease in the concentration of the drug at the receptor site or through a decrease in response of the receptor to the same concentration of the drug. Some drugs can stimulate their own metabolism, are therefore eliminated more rapidly with repeated dosage, and accordingly less drug is available to elicit a response. However, the most clinically important and relevant examples of tolerance result from changes in numbers of receptors and subsequent quantitative changes in concentration–response relationships.

Desensitization is a term that is used to describe changes in the dose–response relationship, irrespective of time, arising from a decrease in response of the receptor. Desensitization can result from decreased G-protein coupling, decreased receptor binding affinity, downstream modulation of the initial signal, or the downregulation (decreased numbers) of receptors. This latter process is slow and takes hours to days to become effective. However, extracellular receptors coupled to G proteins can show rapid desensitization within minutes during continued activation that can occur through two different mechanisms. The first of these mechanisms is homologous desensitization, in which enzymes are activated as a result of ligand binding to a receptor–G protein complex. These enzymes include G-protein-coupled receptor kinases (GRKs), which interact with the $\beta\gamma$ subunit of the G protein and inactivate the occupied receptor protein by phosphorylation. The second of these mechanisms is heterologous desensitization, in which the receptor, whether occupied or not, is inactivated through phosphorylation by a cAMP-dependent kinase, which can be switched on by a variety of signals that increase cAMP. Both $\alpha_1$-adrenoceptors and muscarinic receptors, which are

linked to phospholipase C, may also undergo desensitization via receptor phosphorylation, which uncouples the G protein from the receptor. Regardless of the mechanism, ultimately the phosphorylated receptor protein is eventually internalized and subjected to intracellular dephosphorylation before reentering the cytoplasmic membrane. Finally, the downstream modulation of a signal may occur through either a feedback mechanism or through simple depletion of an essential cofactor.

## 6 SCREENING FOR NATURAL PRODUCT ACTVITY

What is a screen? A screen is an assay or biological assay that provides a tool that can be used to test for or establish the presence and level of a target activity in a specific sample. Bioassays in a screening program should be rapid, simple to conduct, relevant, capable of being automated, cost effective, and of the potential to deliver high throughput [15, 16, 69]. Appropriate technology should be used to permit low limits of detection. This last point is important because the concentration of desirable compounds is unknown in each sample and so it behooves one to strive for the lowest possible limit of detection. Screens should also be specific for the molecular or cellular therapeutic target of choice. Appropriate additional discriminatory tests outside of the focus of the chosen target activity, such as cytotoxicity measurement for cell-based assays, or isotype specificity tests for molecular assays, are valuable in that they provide additional information relative to the overall value of a potential hit. Furthermore, data generated from all screens in which samples are tested should be compared so that selective hits can be identified at an early rather than a late stage. Such a combination of specific screens, data comparisons, and discriminatory assays makes possible the earliest selection of the best hits for continued work and success. The screens must work in the presence of the compounds to be tested and accordingly must be compatible with a given molecule's physico-chemical characteristics. Accordingly, natural product screens need to be operational in the presence of solvents and buffered against extremes of pH and ionic strength and should not be affected by the presence of color. Screens should always incorporate the proper use of controls, both positive and negative. Screens should be bidirectional with regard to their output and have a defined and easily interpretable endpoint. Screens need to be capable of delivering quantifiable data. Screens can be designed in such a fashion as to monitor only a single biological activity. However, there is value to the approach of coupling biological screens to evaluate multiple general biological activities in addition to the target activity [38]. This is because a compound that does not provide a target hit may well generate value for itself by demonstrating some other unrelated and unexpected activity. Alternatively, compounds that do provide hits may well expand their value by demonstrating other unexpected types of biological activity. Additional considerations of the construction of screens, screening programs, and screen design are beyond

the scope of this chapter, and the reader is instead referred to an excellent reference on the subject by Gad [54]. What will be addressed here are some of the common features of screens and screening specific to research on the development of natural products as drugs.

Before one can begin to screen, it is important to know for what one wants to screen. Drug discovery begins with basic ideas, ideas relevant to therapeutic targets and sources of compounds [40, 66, 164]. Therapeutic targets can arise from genomics, the molecular cloning of receptors and signaling molecules, a detailed understanding of physiology, biochemistry, and pathology, research into folklore or ethnomedicine, and knowledge of the traditional uses of natural products. Sources of compounds can present from existing chemical libraries, historical compound collections, natural product libraries, combinatorial libraries, rational chemical synthesis, general or targeted literature searches based upon existing knowledge or leads, and antisense oligonucleotides. The simultaneous consideration of all or at least several of these areas is essential to the original design of a screening program.

A totally random approach in the selection of a source can be coupled with mass screening, but such a path is typically not successful. But to be fair, the random approach is more likely to generate compound novelty [41]. Generally, approaches utilizing literature searching, existing chemical libraries and folklore, ethnomedicine, or traditional medicine are the most popular because of their cost effectiveness. It is worth noting that programs and selections based on or incorporating ethnomedicine or foklore are five times more effective in the ultimate generation of leads. The most effective approach is generally considered to be a mixing of as many components of the therapeutic arm with the compound source arm as opposed to the selection of any single aspect.

It is worth noting at this point that a new approach is being used in the field of natural products that leads to the generation of "unnatural natural product compounds" [40]. This approach is termed *biochemical combinatorial chemistry*. In this technique, appropriate secondary metabolic enzymes are isolated from a crude natural product mixture. These enzymes are then used to generate unnatural metabolites, which are isolated and then subjected to a bioassay procedure that couples a bioassay to an analytical procedure permitting structure elucidation.

What does one put into a screen? It is very tempting to consider the purification of natural products into their individual components before embarking on screening activities. However, this is a very economically challenging and financially unrealistic approach [65]. The classical approach is to design a screen or screening program that will permit the use of the assay or assays to provide guidance to successive steps in purification. An advantage of this is that the effectiveness and efficiency of purification can also be simultaneously evaluated at the same time as biological activity is being enhanced. Ultimately, after sufficient purification, a chemical structure can be determined for the active moiety. However, Grabley and Sattler [60] are of the opinion that the

use of cost-effective physico-chemical and chemical screening procedures will facilitate biological screening because of an ability to provide purer extracts if not pure compounds for biological screening.

Some objectives should be kept in mind when preparing natural product extracts for their ultimate introduction into the screening process [70]. First, every attempt should be made to stop ongoing biological processes. Second, steps should be taken to provide chemical stability of the compounds in the extract. Third, efforts need to be made to minimize losses of material. Fourth, sample preparation costs need to be minimized. When trudging through the iterations required to purify mixtures, one should be sure to: (1) focus on the activity of interest, (2) focus on the compound(s) of interest, (3) eliminate nuisance materials such as cell parts, biopolymers, and other compounds not of interest, and (4) be sure to enrich the composition of compounds of interest.

What types of screens, assays, or bioassays should be used? Assays for activity can be performed at a variety of levels ranging from the molecular level to the whole organism. While it is true that the high-throughput screening of synthetic compounds generated by combinatorial methods may be best achieved at the molecular level, Duke et al. [40] are of the opinion that natural product screening should be performed at the highest level possible since more effort per compound has been invested in the discovery of each compound.

Historically, natural products have been subjected to what is termed bioassay-directed isolation. In this approach, a crude natural product mixture is subjected to fractionation and the individual fractions then bioassayed for specific biological activity. This process continues repetitively with comparison of individual fraction assay data to a bioassay database. When the data is shown to match a previously known profile, the process is terminated and the sample is discarded. If the data is shown to provide a new profile, the structure of the compound is determined. However, this approach can lead to the rediscovery of previously known compounds after significant effort has been expended.

Low-molecular-weight natural products from a variety of sources represent unique structural diversity. In order to more adequately and efficiently access this diversity, various new strategies improving targeting and direction have been developed [60]. Modern separation/chemical characterization approaches can eliminate much of this problem by identification of the compounds before they are subjected to bioassay. Indeed the coupling of such techniques to biological screens can improve the quality of the assay result and shorten research and development time frames. These new tandem approaches are termed fractionation-driven bioassays. While biological screening directly correlates to a predefined biological effect, physico-chemical and chemical screens do not. In this latter case, lead selection is based on either physico-chemical properties or chemical reactivities. In both cases, the first step is the chromatographic separation of compounds from the complex source mixture. In the second step, the physico-chemical properties or chemical reactivities of the separated compounds are analyzed. Both of these chemical-based strategies have proven to be of value as auxiliary or

supplemental methods to biological approaches. Data generated from physico-chemical and/or chemical screening is very helpful in the de-replication or early identification and exclusion of known or otherwise unsuitable compounds that occur during high-throughput biological screening programs. Furthermore, the use of physico-chemical and chemical screening will aid in the establishment and building of natural-product-based compound libraries, which could then be used more broadly in testing programs.

There are two general types of new tandem assays. The first of these is referred to as the fractionation-driven bioassay. In this method, a crude natural product mixture is subjected to fractionation and the individual fractions then subjected to nuclear magnetic resonance (NMR) spectroscopy or mass spectrometry/mass spectrometry (MS/MS). The structures of the compounds in the individual fractions are identified; and, if they are known, their biological activity profiles are evaluated from an existing database. If the structures are unknown, then the compounds are subjected to bioassay. In an alternative approach, termed the isolate and assay approach, a crude natural product mixture is subjected to automated fractionation and purification. The individual fractions are then subjected to bioassay. Desirable biological activity serves as the trigger to subject the sample to NMR or MS/MS and ascertain the structure(s). If the material is a novel compound, the structure can be optimized. If the material is a previously known compound, it may well be discarded, depending on its biological activity or toxicological profile.

A broad range of screening technologies are currently available for use in screening for natural-product-based drugs [69]. For molecular targets, such procedures would include generalized solution-phase assays, immobilized substrate assays, scintillation proximity assays, and time-resolved fluorescence assays. For cell based targets, cell signaling, cell communication, cell receptor, and reported gene assays are available. Isolated subcellular systems are also available [156]. Bioassays can, if desired, incorporate lower level organisms, isolated vertebrate organs, or whole animals. In short, systems are limited only by the creativity and design of the screener.

Examples of screening programs can be readily found in the literature. Just to reference a few, Quinn [128, 129] has reported on his efforts on prospecting the biodiversity available in Queensland; Mehta and Pezzuto [112] have published on their program to identify cancer preventive agents from plants; El Sayed [41] has reported on his screening program for antiviral agents; Barrett [5] has written on his program to find novel antifungal agents; Yang and co-workers [167] has reported on his search for anti-HIV compounds from natural sources; and Bindseil and co-workers [9] have published on their experiences on screening with pure compounds.

The emergence of high-throughput screening (HTS) has permitted the rapid screening of extremely large collections of structurally diverse synthetic compounds against a variety of novel and diverse disease targets [19, 128, 129, 150]. However, despite initial hopes that HTS was the final solution for drug discovery, for reasons that will not be discussed here, HTS has not achieved

that distinction. Nevertheless, HTS is still a powerful tool. HTS strategies focus on the ability to screen large libraries of compounds. However, the limiting factor in HTS is the ability to access large numbers of chemically diverse substances. Natural products are undoubtedly the greatest source of structural diversity. Accordingly, HTS of the unparalleled diversity that exists in natural product extracts offers the highest probability for discovery of novel lead compounds and should therefore be viewed as being complementary to compounds generated from combinatorial chemistry alone. The synergistic melding of HTS and natural products has started, and, as it progresses along its development path, exciting new breakthroughs will undoubtedly be presented.

As a compound generates interest through a variety of screening assays and progresses down the drug discovery path, certain questions need to be asked [5]. Is the chemical structure of the compound novel? Is the mechanism of action of the moiety novel or of utility? Is the biological activity of the compound useful? Is the potency of the compound reasonable? Is a proof of concept available? Is chemical modification of the structure possible? Is solubility a problem? Can the material be synthesized on a large scale? The ability to ask and answer these questions effectively early on in the process will be highly predictive of the ultimate success of a particular line of research. While there are no "correct" answers to these questions, as answers will be different depending on the therapeutic indication and other available therapeutic alternatives, they still need to be addressed to provide proper program focus.

Before ending this section, it is important to emphasize the necessity and importance of keeping detailed records as one initiates the screening process [70]. Maintenance of a secure physical inventory with a controlled environment and adequate records is another important detail. The use of bar code identifiers is very desirable for samples and relevant computer programs are readily available. It should be considered to be essential to establish a complete database, which should include the source, source location, isolation details, any relevant taxonomic information (kingdom, genus, and species), and any other relevant information (third-party suppliers of reagents, potential pathogenicity or toxicity, relevance to any international biodiversity treaties), preservation methods, age, position or location within a freezer, and the like. Other considerations in the design and management of a successful screening program could be limitation of access, storage of reference or backup samples, storage of a backup copy of inventory and database, establishment of tracking procedures, writing of standard operating practices (SOP) for sample handling, and temperature alarms.

## 7 ISOLATION AND PURIFICATION OF NATURAL PRODUCTS

Why do scientists working with natural products isolate and purify them? For one of two reasons: (1) to ascertain what the natural product is and (2) to carry

out sufficient experimental work necessary to biologically characterize or profile the compound. It can be quite a sobering experience to look at a flask full of dark-colored, inhomogeneous sludge and liquid and realize that one is going to attempt to isolate just one particular type of molecule from all of the other materials that are present. To put this in perspective, typically the material sought after represents only about 0.0001 percent of the total biomass in the flask [15, 16]. Then, just to make things even more challenging, the desired molecules can also be bound with other materials and molecules present in the mixture, making the desired compound(s) even harder to purify. It is important to keep in mind that the isolation of natural products differs from that of the more prevalent biological macromolecules. This is because natural products are typically secondary metabolites and as such are smaller in size, chemically more diverse in structure, and present in smaller concentrations than the more homogeneous proteins, carbohydrates, lipids, nucleic acids, and the like.

This section will not present a condensed work on separation science and procedures as that is best left to any of the myriad analytical chemistry textbooks. What will be attempted here is to provide sufficient guidance on the isolation and purification of natural products so that proper focus can be assured in the design and implementation of a successful isolation and purification program.

Before initiating an isolation and purification, there are a number of basic questions that need to be asked and answered. First, what are you trying to isolate and purify? There are a number of different possible targets: (1) an unknown compound associated with a particular biological activity, (2) a previously known compound present in a specific organism, (3) a group of compounds within an organism that are all structurally related to each other, (4) all of the metabolites produced by one natural product source that are not produced by another closely related source, or (5) all of the molecules of a particular organism.

Second, why are you trying to isolate this material? While the asking of such a question might appear to be superficially inane, it is important to know why you want something so that you know how much of it you might need. Possible reasons for carrying out an extraction might be: (1) the generation and supply of larger amounts of an already known compound so that more extensive biological testing such as pharmacology and toxicology can be performed on the material, (2) the purification of a small amount of material for initial biological and chemical characterization to be performed, or (3) to purify sufficient material in order to conduct complete structural studies and further biological activity characterization.

Third, what type of purity is desired? If a natural product compound is to be used for biological testing, it is important to know not only the degree of purity of the material but also the nature of the impurities. It needs to be appreciated that the impurities themselves can contribute significantly to any biological activity observed in the screening program. If the material is to be

used in more refined pharmacological or pharmacokinetic testing, then the material should generally be at least 99 percent pure. If, on the other hand, the material is to be used only for chemical characterization, the acceptable level of purity can range from 95 to 99 percent. Such a range of purity will generally be sufficient for the determination and assignation of a complex chemical structure via such techniques as NMR spectroscopy, infrared (IR) spectroscopy, and MS/MS spectrometry. It should be noted that if the compound under consideration is present in a high concentration in the starting material and a standard for that compound already exists, then structural confirmation can be achieved with less pure material and the associated purification scheme will be composed of fewer steps. Depending upon one's goals, varying degrees of purification may be acceptable. X-ray crystallographic studies will demand material of 99.9 percent purity, while detection of the presence or absence of a specific structural feature via analysis of the ultraviolet spectrum may tolerate purities down to a level of 50 percent. An important concept of purification is that the relationship between purity of compound achieved during natural product extraction and the amount of effort expended to achieve such a level of purity is almost exponential in nature. When starting with a crude, complex mixture, it is very easy to eliminate large components of unwanted material. However, as the purity begins to escalate, it can become infinitely more challenging to improve purity levels. For example, the effort required to go from 50 percent purity to 90 percent purity can pale in comparison to the effort required to go from 99.5 to 99.9 percent purity. In concordance with this, it is fair to state that the relationship between purity level and yield are also exponentially related. In a purification scheme no step delivers the desired material in 100 percent yield. Each extraction step results in the loss of material, and when working to attain very high levels of purity, losses can be extreme. While it may be necessary to take only very "centralized" cuts in a purification step, keep in mind that the "tail" cuts can themselves be later subjected to reprocessing.

Fourth, what type of fractionation should be used in the isolation and purification scheme? All separation processes involve the division of a mixture into a number of discrete fractions. This process is called fractionation. Such fractions can be physically separate such as the two phases of a liquid–liquid extraction or they may not be physically separate such as the continuous eluate from a chromatography column. The eluate from a chromatography column can then be artificially divided into fractions via the use of a fraction collector. The method of fractionation depends on the sample and the goals of the separation. Fractions are typically equal in size and can be large or small in volume. The collection of a large number of small fractions improves the probability that each fraction might contain pure compound. However, such an approach requires significant work in the analysis of each fraction. Additionally, this approach may spread the desired compound over so many fractions that if the target molecule(s) was present originally only in low levels, it may prove undetectable in any one of the fractions. Alternatively, if the separation

process is cruder, employing the collection of only a few large volume fractions, a more rapid and facile tracking of the desired compound and its activity is possible.

Fifth, what is the nature of the compound? The answer to this question depends on how much is already known about the compound. General features that are useful at this stage of the project are acid/base properties ($pK_a$, $pK_b$), molecular charge, stability, and solubility (hydrophilicity/lipophilicity). If the target molecule is an unknown moiety, it is very likely that little of this information is known and all of it will have to be determined along the way. If one is isolating a known compound, much of this information will already be available. Finally, if the goal is to isolate a number of secondary metabolites rather than a single molecule, then the value of this step is less important, but an appreciation of the relative values of these parameters can still be useful with regard to understanding the characteristics of the mixture.

Sixth, where is the desired activity localized? Each potential source of a natural product source—whether it is plant, tree, moss, bacteria, vertebrate, invertebrate, insect, terrestrial, or marine based—has components or parts in which the desired activity or compound is present in greater concentration as opposed to other parts in which the compound is present in lesser amounts. To obviate any problems associated with dilution of the compound and its activity, the initial biomass should be selected on the basis of its content of the target biological activity.

Only with the thoughtful provision of answers to the above questions can one have a clear idea of what one is attempting to achieve and how to successfully secure the project goals. It should be obvious that there is no correct or incorrect protocol or standard operating procedure for the isolation and purification of natural products. Indeed, the final method or scheme itself is most likely to vary with the answers to the above questions as well as the natural product source and the specifics of the assays and biological assays that are to be used in the screening program. However, a consult of the literature is essential during the design and construction of an isolation and purification program for any natural product compound [149, 165]. While it is possible that extensive data may have already been published on the compound one is trying to isolate or compounds related to it, it is also entirely possible that nothing is known. Regardless, proper use of the natural products literature can facilitate the effort invested into the design and implementation of a specific isolation and purification program.

A variety of different techniques can be used for the isolation and purification of natural product compounds [15, 16, 82]. These techniques include, but are not limited to, solid-phase extraction [15, 16], high-performance liquid chromatography (HPLC) [15, 16], gradient high-performance liquid chromatography [15, 16], bioautography [165]; thin-layer chromatography (TLC) [15, 16], countercurrent chromatography [2, 49, 50], droplet countercurrent chromatography [74], vacuum column chromatography [165], desalting [149], liquid–liquid chromatography [75], paper chromatography [15, 16], ion

exchange chromatography [15, 16], size exclusion chromatography [73], affinity chromatography [15, 16], acid–base switching technology [42], centrifugal partition chromatography [74], liquid–solid chromatography [15, 16], microwave-assisted extraction [80], pressurized solvent extraction [80], large-scale solvent extraction [42], and supercritical fluid extraction [94, 126]. While the theories along with the relative advantages and disadvantages behind each one of these procedures have not been discussed here, the listing will serve as a catalog of potential techniques available to the researcher. Specific details on any of these procedures can be obtained from any number of books on separation.

A debate still exists as to the timing of isolation and purification in the drug discovery process [65]. It is always tempting to isolate and purify before screening, but understandably this can present a challenge. Classical approaches have used a successful marriage between purification steps and bioassay activity assessment to isolate, identify, and fully characterize natural product compounds. It should be appreciated that with the advances in chromatographic and analytical techniques that have taken place over the last 15 years, the time required to proceed from an initial hit to an identified active compound should take no longer than for the resynthesis and purification of a potential active compound from a combinatorial library. Accordingly, the timing of isolation and purification in the natural product drug research and development timeline should not persist as such a point of contention as previously [9].

## 8 STRUCTURE IDENTIFICATION OF NATURAL PRODUCTS

The chemical structures of natural product compounds are tremendously diverse and can be very elegant in their nature [98, 115, 116, 170]. Such diversity can present a challenge to the analytical or medicinal chemist attempting to unravel the mystery of the chemical structure of an unknown material presented to him or her. However, modern technology has made structure identification simpler and faster. Today, scientists take for granted such techniques as MS, MS/MS, IR, Fourier transform infrared spectroscopy (FTIR), NMR, Fourier transform nuclear magnetic resonance spectroscopy (FTNMR), and others. It is beyond the scope of this chapter to discuss the theory and relative merits of each of these techniques, and the reader is urged to consult appropriate textbooks on analytical chemistry. However, it is worth mentioning that some particularly exciting developments in structure determination pertinent to the area of natural products have come from the field of computer-assisted structure elucidation (CASE). Several computer programs have become available for scientists to use and a number of publications reporting on the utility of this technological advancement have been forthcoming [154]. As has been previously described, the elucidation of the structure of a natural product begins with the collection of a crude material. This material is then

subjected to a series of separation steps, usually involving chromatography, delivering in the end pure compound(s). Finally, a set of spectroscopic and spectrometric experiments are performed on the pure compound to delineate the structural characteristics. Such analysis may even reveal the two-dimensional or three-dimensional structure of the isolated chemical.

Time is money in the drug development business and in order to accelerate activities, the following steps of the structural elucidation process should be considered to be targets amenable to automation: (1) the choice of the smallest group of procedures that is most likely to reveal the unknown structure [NMR, FTNMR, MS, MS/MS, IR, FTIR, ultraviolet (UV), etc.], (2) the acquisition of data from the selected procedures, (3) the analysis of data from the selected procedures, and (4) the use of a computer program to construct the structure from collected spectroscopic and spectrometric data. Historically, structure identification has occurred after purification was complete. However, not infrequently in these cases, structural identification revealed that all of the previous laborious steps of purification had produced a compound that was already known or of an undesirable type. Now, the coupling of liquid chromatography (LC) or high-pressure liquid chromatography (HPLC) with such technologies as MS, MS/MS, and NMR has permitted the construction of devices that allow the injection of a crude sample, separation of the sample using automatically determined optimized conditions, and on the fly spectrometry or spectroscopy followed by CASE for each set of acquired spectra. The identification of unwanted compounds or de-replication should occur as early as possible in the natural product isolation and purification process to avoid the loss of time and funds. The development of coupled techniques has permitted the achievement of that goal. Indeed, coupled techniques such as LC/NMR/MS have now evolved and have potent application in the pharmaceutical field. As the sensitivity of instrumentation continues to improve, the value of these coupled techniques will increase even more.

A number of powerful aids for NMR spectroscopic interpretation have emerged and are now readily available from instrument manufacturers and include programs such as Auralia and AMIX [117]. However, any comprehensive CASE program will be based on a quality structure generator. To this point, only a few high-quality, pure structure generators have been developed. While over the years there have been a variety of programs that have been developed, they have been limited to those researched and developed by small, private groups. Now there are some highly capable programs that are commercially available, such as Assemble (Upstream Solutions GmbH, Zurich, Switzerland) and MOLGEN (http://www.molgen.de). A well-known deterministic CASE system that is often cited in the literature is CISOC-SES, which is now commercially available under the acronym NMR-SAMS (Spectrum Research, Madison, WI, USA). Another deterministic CASE program, COCON has been relatively recently introduced, and several examples demonstrating its value in structural elucidation have been reported [97, 154].

Deterministic algorithms have a limit to the size of the molecule with which they can work. To overcome this, a stochastic structure generator has been published for the computer-assisted structure elucidation of organic molecules [153]. The name of the program is SENECA. This program is written in Java and therefore is platform independent and allows for a simple plug-in mechanism for new spectroscopic data types. Theoretically, many different kinds of properties can be plugged into this system, as long as the property can be reliably calculated from a generated molecule.

Classical CASE systems can at least attempt to provide the two-dimensional structure of an unknown molecule. Now, with the greater exposure of, availability of, capabilities of, and demand for CASE programs, efforts are being made toward incorporation of the ability to confidently determine the three-dimensional structure of unknown molecules [68].

Advancements in chromatography, spectrometry, and spectroscopy together with breakthroughs in the coupling of these technologies are important steps in the production of a fully automated and integrated natural products structure determination instrument, which will provide significant advantage to the early, rapid, and facile identification of new natural-product-based drug opportunities.

## 9 SYNTHESIS OF NATURAL PRODUCTS

Once a natural product compound has been screened for biological activity, isolated, purified, its structure identified, and the pharmacological profile refined, the journey is not over. The molecule may turn out to be too complex in nature and too expensive to be synthesized. Indeed, when compared to a purely man-made synthetic alternative, many times the natural product compound is quickly eliminated from further consideration because of considerations of time and potential costs of synthetic production.

Any given natural product compound may possess unacceptable physicochemical, pharmacodynamic, pharmacokinetic, or bioavailability properties or demonstrate excessive toxicity and will therefore require optimization of its chemical structure. Optimization involves a dissection of the lead molecule and the synthetic addition, removal, replacement, or modification of substituent groups so as to enhance the utility and efficacy of the molecule. The synthesis of a complicated molecule is a very difficult task since every group and atom must be placed in a proper position and with the correct stereochemistry.

Such chemical structure modification or synthesis has been performed over decades by what might be termed more classical means. Indeed, the complete synthesis of natural products has been an area of interest for a long time, and the efforts to produce man-made natural products has provided significant challenge and learning opportunity over the years [21, 43, 53, 103, 115, 116, 125, 133, 151]. Because of the widespread chemical diversity that can be found

in natural products, an ever-expanding collection of fascinating natural product compounds will continue to be presented to chemists for synthesis [162]. If one compares the chemical diversity of man-made synthetic products with the chemical diversity of natural product compounds, it quickly becomes apparent that there are significant qualitative differences between synthetics and natural products [114]. Natural product compounds contain more alcoholic and ether groups, while pure synthetic compounds possess more aromatics, amines, and amides. If one looks at group combinations, there are higher percentages of alcohol/ether, alcohol/ester, arene/alcohol, arene, alcohol, or ether functionalities in natural product compounds when compared to the synthetic compounds. However, pure man-made synthetic compounds are found to have combinations such as arene/amine, amine/amide, or amine/arene/amide in higher frequencies than natural product type of compounds. Finally, natural product compounds are found to more commonly possess bridgehead atoms and contain a greater number of rotatable bonds per molecule, chiral centers per molecule, and rings per molecule than pure man-made synthetic compounds. The importance of these differences is that they reveal and emphasize the complementarity of natural product compounds as a group with man-made synthetic compounds. Despite all of the knowledge and achievements that have been gained and the advances that have been made, classical synthetic organic chemistry will not alone unlock and open the potential of natural products to the pharmaceutical marketplace. Instead, the future lies in the synergistic union of classical organic chemistry with microbiology, biochemistry, combinatorial chemistry, and other fields to provide new synthetic strategies to generate natural-product-based drugs.

The history and specific techniques of combinatorial chemistry are beyond the scope of this chapter, and the reader is referred to appropriate textbooks for discussions on that topic. While it should be recognized that combinatorial chemistry is a perfect match for high-throughput screening because of its ability to produce large numbers of compounds in a short period of time. The promise of combinatorial chemistry to deliver more drug candidates within a shorter period of time has remained unfulfilled. What has been lacking in combinatorial chemistry is the skeletal structural novelty that natural products can provide [1, 12, 13, 59, 63, 111, 119, 120, 155, 158, 161, 163].

Over time, organic synthetic chemists have become interested in enzymes and their potential role in natural product synthesis [85, 109, 136]. Enzymes have great power as catalysts for regiospecific and stereocontrolled synthesis. These biological catalysts are very capable at room temperature of converting inexpensive substrates into value-added products at a significantly high throughput. However, barriers remain to the more expanded use of enzymes in organic synthetic chemistry. The most important of such obstacles includes the inability of enzymes to work with unnatural substrates.

Combinatorial biosynthesis utilizes enzymes from various natural product source biosynthetic pathways to create novel chemical structures [10, 27, 135].

The engineering of polyketide synthases has thus far been the central point of this activity and led to the production of several erythromycin analogs. The end result of such research activity will be the development of more rational and faster methods of production of new compounds for the development of therapeutic agents from natural products.

The research on and screening of natural products today is focusing on many different therapeutic indications. Fermentation broths and plant extracts have done well in delivering leads and genomics and molecular biology have done well in delivering targets. Regardless of the type of compounds involved, improved efficiencies in the design and synthesis of natural-product-based agonists and antagonists will be key to a realization of the full potential of natural products as drugs [1, 139, 140].

## 10 DEVELOPMENT OF NATURAL PRODUCTS

### Regulatory Guidelines and Nonclinical Development

The classical model of drug development is composed of three phases: discovery, development, and marketing [108]. Discovery is the first of these phases and is composed of two essential components, drug discovery and drug design. Development is the second of these phases and is composed of two large components, preclinical studies and clinical studies [30, 56–58, 95, 108, 146]. The development phase, although lengthy, expensive, and time consuming is meant in its design and conduct to cull out undesirable compounds and ensure the safety and efficacy of compounds that successfully run the developmental gauntlet and decrease the hazard and risk of exposure for humans. We have touched on the drug discovery and design in earlier sections of this chapter. In the course of these phases, therapeutic targets have been selected and screens were designed to identify drug candidates. These same candidates were then isolated, purified, identified, and then structurally optimized in consideration of the molecule's biological activity profile. This process of optimization attempts to minimize toxicity and maximize therapeutic value, efficacy, and pharmacokinetic characteristics. Finally, research was performed to elucidate a mechanism of action for the potential drug and animal models identified to establish efficacy of the material. Pharmacology studies, a component of the discovery phase, confirm the efficacy of a potential drug. The availability of analytical methods is essential to establish the exposure of the test system to the drug candidate and the presence of the drug in various tissues and fluids. The absorption, distribution, metabolism, and excretion and elimination of the compound also need to be determined. Human in vitro P-450 studies are necessary to be performed to fully understand the hepatic metabolic pathways. These studies may also herald the potential for serious drug interactions that may have to be addressed during the development cycle.

Safety pharmacology studies are part of the development phase and the preclinical development program but can be performed either as part of the pharmacologic profiling of a drug candidate or as a prelude to the toxicology studies. These types of studies are a preliminary hazard assessment of the test article in key organ systems such as the central nervous system, cardiovascular system, gastrointestinal system, renal system, and pulmonary system and essential in the planning of human clinical trials. For an excellent and detailed discussion of safety pharmacology, the reader is referred to Gad [54].

Preclinical studies begin some of the more rigorous testing that a potential candidate must successfully endure and survive. It needs to be recognized that a decision to proceed with preclinical studies represents a major commitment of time, resources, effort, and money. The major objectives of any preclinical development program should be: (1) development of a Good Manufacturing Practices (GMP) synthetic method to produce the test article, (2) synthesis of a supply of test article that is adequate to permit the performance of all preclinical work and the initiation of clinical studies, (3) the creation of a usable and tolerable preclinical formulation(s) and efficacious clinical formulation(s), (4) complete pharmacologic and pharmacokinetic profiling of the test article, (5) performance of proper toxicology studies to support an Investigational New Drug (IND) application, and (6) construction of a complete and detailed informational platform to permit the recommendation of an initial human dose in phase I studies.

Nonclinical or preclinical studies are the same thing and are composed of studies on drug processes, pharmacology, and toxicology. Drug processes involve the scaling up of synthetic procedures or the bulk preparation of the drug [32, 146]. This means the development of a synthetic method that permits the synthesis of compound at a reasonable cost, in a reasonable amount of time in such a fashion that regulatory and GMP guidelines are satisfied. A major obstacle in the performance of preclinical studies is often the ability to merely "have drug." Drug supply is often overlooked in the drug development program and can cause costly delays. Related to this is the existence of some sort of formulation for the drug candidate that will permit effective dosing in preclinical studies. The lack of proper planning on a preclinical formulation (not the final clinical formulation) is oft overlooked, and this oversight can cause significant problems of delay or confounding toxicity. Also during the preclinical phase of the drug development paradigm, various analytical techniques must be developed and validated in a variety of species and biological milieu, analytical standards defined and made, compound stability profiles generated, and research into the proper human formulation initiated and conducted. Sometimes even radiolabeled drugs will be required to be made for some studies, a difficult, costly, and time-consuming project in and of itself.

Toxicology studies are performed to assess a drug candidate's potential safety. These studies are so critical that there are established procedures called Good Laboratory Practices (GLP), which specify how studies are to be done.

Toxicology studies have a profound quality assurance component, which facilitate inspection and permit validation of the data by a regulatory agency or reviewer. There are numerous different types of toxicology studies and the details of them are beyond the scope of this discussion, but the most comprehensive, detailed, and descriptive source is by Gad [56]. These studies typically include mutagenicity, genotoxicity, and cytotoxicity studies; acute studies; subchronic studies; chronic studies; reproductive and developmental toxicity studies; carcinogenicity studies; immunotoxicity assessments; various worker safety studies; and certain specialized toxicity studies, the conduct of which is determined by compound-specific issues or dose route. Toxicology studies should in their core design determine the target organ(s) of any potential toxicity, characterize the shape of the dose–response curve, ascertain the reversibility of any potential lesion, and facilitate the selection of a human dose in phase I clinical trials. An excellent example of preclinical and clinical development programs for a natural-product-based drug has been reported by Rowinsky et al. on the antineoplastic agent Taxol [141].

When a drug candidate completes the preclinical studies, an IND is filed and clinical trials are initiated. Clinical trials are composed of phases I to IV. Phase I studies represent the first exposure of a drug to humans. These studies examine the effects of single and multiple increasing doses in small numbers of normal and/or patient volunteers. Tolerance to the compound is also evaluated. Basic pharmacokinetic studies are performed in conjunction with the studies to aid in the determination of later dosage regimens and to fully characterize the routes of metabolism, excretion, and elimination and assess the presence and amounts of active and inactive or toxic metabolites. Typically only a small number of volunteers are involved. Phase II studies represent the first time that a drug candidate is tested in humans for efficacy. However, safety and tolerance are still monitored in these studies concurrently. In phase II studies, different dose–range finding studies are performed to optimize the dose of the drug, to maximize its efficacy, and to minimize any compound-associated intolerance. These studies typically involve up to several hundred patients and generally provide the first indications of the potential benefit of the drug as compared to its risk of exposure. The last clinical studies that need to be performed before submitting a complete information package for regulatory approval are phase III studies [New Drug Application (NDA) for filing in the United States]. In these studies, the drug candidate is utilized at the optimum dose, in the target patient population, and in exactly in the same fashion that the drug will be used if eventually approved and marketed. These studies will verify efficacy and safety and detect adverse reactions and contraindications. These are very large studies composed of hundreds to thousands of volunteer patients, depending upon the therapeutic indication. Phase IV studies are postmarketing surveillance studies, which are conducted after a drug has been approved for sale. These studies typically involve adverse reaction reporting, surveys and general sampling, and testing evaluations.

The design of a proper and detailed preclinical program is a challenging exercise, and the successful navigation of the program with its on-time completion should not be taken for granted. Effective project management with a timely resolution to all obstacles that will present themselves along the way can save significant amounts of money. Accordingly, consideration should always be given to the retention of a professional and suitably experienced preclinical project manager. The expenses incurred with the design and implementation of a preclinical development program is more than offset by the avoidance of cost and/or time overruns.

As stated earlier in this chapter, natural products can be interpreted by different people to mean different things. For purposes of this discussion, the term *natural products* refers to drugs of natural product origin and not to botanicals, herbals, or traditional medicine products. The development of botanicals, herbals, dietary supplements, and over-the-counter products is not relevant to this chapter and is covered elsewhere [152]. There are no regulations or guidelines that are specific to the development of therapeutic agents developed from natural products. Indeed, the development of a natural-product-based therapeutic candidate is more than adequately covered in existing governmental guidelines. These guidelines are available from a variety of governmental sources [*www.fda.gov* (U.S. FDA), *www.eudra.org* (European Union), *www.mhw.go.jp/english* (Japan), *www.ifpma.org* (ICH)] or references on preclinical development [51, 55–58, 105, 106]. The most comprehensive and detailed of these references is by Gad [56]. However, it should be noted that according to a draft document entitled "Botanical Drug Products Docket Number: OOD-1932" issued by the Office of Nutritional Products, Labeling and Dietary Supplements (HFS-800) in 2003, if a natural product has been sold in a crude or semipurified form as an over-the-counter drug and a sponsor wishes to file an IND on a moiety purified from such a preparation, then reduced documentation may be acceptable to the FDA.

## Learning from the Mistakes of the Past in the Development of Natural Products

The very first sample of bark from the Pacific yew (*Taxus brevifolia*) was originally collected in 1962 by the U.S. Department of Agriculture as part of a plant screening program established by the Cancer Chemotherapy National Service Center of the National Cancer Institute (NCI) [32]. It was not until 1964 that a positive response was found for the extract in the KB cytotoxicity assay. Taxol was then identified as the active component of the mixture in 1969. Not unlike many other compounds being tested at that time, paclitaxel originally demonstrated moderate in vivo activity against P388 and L1210 murine leukemia models and was not considered to be a promising candidate for further development. However, strong activity demonstrated by Taxol against the B16 melanoma line introduced by the NCI in 1975 caused a reevaluation

of the material. Indeed as a direct result of this additional research, Taxol was recommended as a candidate for preclinical development in 1977. Further work demonstrated strong activity against human tumor xenograft systems and stimulated hope of efficacious performance in the clinic. Taxol's mechanism of action was elucidated in 1979, formulation work completed, and toxicology studies started in 1980 [32, 100]. With the completion of nonclinical studies, approval was given for entry into phase I clinical trials in 1983 [28, 141]. The early clinical trials raised serious issues of toxicity. Indeed, further development of Taxol was almost discontinued. The problems, however, were determined to be related to the poor solubility of Taxol in aqueous systems and the necessity for a high dose, when compared to other antineoplastic agents of the time. The development of a suitable formulation required the use of Cremophor EL, a polyethoxylated castor oil derivative, which created a whole new set of challenges to be addressed [36]. Eventually, a safe clinical regimen was established reducing the incidence of allergic reactions, and Taxol proceeded into phase II clinical trials. However, at this point in development, faith and interest in the success of Taxol began to wane for a variety of reasons, and the priority of Taxol's development was again lowered. Indeed, the support for Taxol decreased to such a level that the production of bulk quantities of Taxol needed to proceed through clinical trials was severely cut back. This position forced the creation of a critical problem when Taxol demonstrated significant antitumor activity in phase II clinical trials in patients afflicted with ovarian cancer. The preliminary reports from clinical trials indicated an approximate 30 percent response rate, with some patients actually achieving remission [110]. Additional clinical trials supported the initial clinical response rate, and accordingly there was a strong demand for more Taxol. Concurrently, it was also discovered that Taxol demonstrated very favorable responses in patients afflicted with metastatic breast cancer [71]. Add to this the findings that favorable clinical responses to the administration of Taxol were also being found in patients afflicted with lung cancer, malignant melanoma, lymphomas, and cancers of the head and neck, and it is easy to appreciate the magnitude of the gap created between the supply of Taxol and the demand for Taxol [32]. Unfortunately, all of this favorable clinical performance was met with not just a poor supply of Taxol but an extremely disappointing inability to produce Taxol on a large scale. The crisis in the supply of paclitaxel was eventually resolved as a result of collaborative efforts between many groups [28, 32]. Such a crash collaborative effort, while admirable, came about as a result of poor planning.

Today, Taxol is viewed as a very important antineoplastic agent on the front line of cancer chemotherapy. Yet, despite its success, it must be sadly admitted that on several occasions during paclitaxel's development, Taxol almost did not make it to market. In an attempt to learn from the mistakes of the past, what lessons can be learned from the Taxol story? These are addressed in the next section when we talk of the future of natural products in drug discovery and development.

## 11 FUTURE OF NATURAL PRODUCTS

Current libraries providing fodder for the search for new therapeutic agents include, for the most part, but are not limited to, synthetic chemical libraries of purified natural products from a variety of different sources [27]. When combinatorial chemistry made its entry onto the drug discovery stage, many people considered this technique to be the ultimate solution for the discovery of potential new therapeutic agents. Accordingly, the perceived value of natural products in the drug discovery process began to pale. While combinatorial chemistry has demonstrated that it is certainly of value in the process of lead optimization, nature itself presents the most diverse and complete source of leads. A relatively recent approach that unites the strengths of combinatorial chemistry and natural product identification is a process referred to as combinatorial biosynthesis [10, 27, 135]. In this technique, the pathway leading to the production of a natural product is identified and the genetic basis of it elucidated. Genetic modifications of the organism are then made causing the production of different biologically active products. These products can then be evaluated for target therapeutic potency in various screens. While this approach has been utilized by some biotechnology companies, widespread acceptance in the pharmaceutical industry has been lacking to this point [27]. However, as acceptance develops, so will the identification of many new potential drug candidates.

To fully capitalize on the extensive biodiversity available to us in natural products, high-throughput screening processes need to be improved upon so that they can provide a higher degree of quantifiable rather than quantal (active versus not active) results, permitting a more adequate description of whether or not any given compound may be considered to be potentially active.

It is also important to realize that one should not rely on just one either in vitro or in vivo screening assay in a discovery program. Obviously, there has to be an economic component to screening programs with some degree of cost containment, but a realistic balance can be achieved, which provides a realistic cost-to-benefit ratio. The discovery and development of paclitaxel is a good example of this concern [32].

The early conduct of studies leading to the elucidation of the mechanism of action of a compound is important. The delineation of a novel mechanism of action can provide considerable impetus to the continued development of a chemical moiety. Irrespective of novelty, the establishment and correlation of a chemical structure with a mechanism of action can permit the categorization of a compound with other chemicals, potential therapeutic agents, and drugs and more rapidly facilitate a compound's potential for therapeutic use.

The efforts to identify potential leads for the treatment of a constellation of maladies and diseases are increasing at a hectic pace and the growth of numbers of actual compound candidates is markedly outpacing the availabil-

ity of screening procedures. In order to accommodate the requisite testing associated with the generation of chemical leads, advances will need to be made in the methods of and approaches to the synthesis of natural products. Critical demands will be placed upon synthetic and medicinal chemists to produce pure chiral products in significant amounts. These products will undoubtedly be the result of complex synthetic pathways, which will incorporate catalysts and biocatalysts, soluble and immobilized elements, and biologically generated as well as man-made moieties. These multistep processes will need to be functional on both a low bench scale as well as a commercial scale. The development of paclitaxel suffered a severe developmental program setback because of an inability to produce adequate supplies at an appropriate time. In association with the development of effective synthetic pathways leading to the production of natural products, attention must also be paid to the required formulation of such products permitting human use, as was well demonstrated with Taxol.

Genetic research will continue to elucidate an ever-increasing number of potential pathophysiologic targets, which can be mated with hopeful novel therapeutic agents. This coupled with the fact that little of nature's natural products archives have really been adequately researched portends great opportunity. How big is this opportunity? Admittedly, there are no sound estimates for the actual total numbers of species of plants, shrubs, trees, insects, fungi, smuts, rusts, marine organisms, and the like that exist on land and in the waters in this world [24]. But, with regard to microorganisms, less than 0.1 percent of the total microorganisms present in the soil to date have been evaluated, simply because they cannot be cultured [27, 138]. Only about 70,000 of 2 to 5 million different fungi are considered to have been identified [27]. Similarly, only about 800,000 of approximately 20 million different insects have been identified.

Structure–activity studies on leads generated from natural products sources combined with computerized graphic model building will become increasingly more prevalent. Such activity in turn will result in the discovery of molecules with optimal activity, improved bioavailability, fewer side effects, and very desirable therapeutic indices [156].

However, abuse of our natural resources will certainly limit our ability to learn from nature [27, 78, 132]. Lack of proper conservation, the expanding pall of pollution, and the wanton destruction of forests, especially the rain forests, will, for example, diminish the opportunity to gain valuable knowledge from a staggering amount of indigenous biodiversity and obviate the possibility to generate a valuable therapeutic agent. Indeed, it is estimated that approximately one-eighth of all plant species are currently near the point of extinction [78]. Alternatively, the highly focused and single-minded research into natural products can in and of itself lead to the depletion of sources of natural products and even the eventual extinction of species [22].

In the hunt for natural products another obstacle that will have to be more commonly addressed is the issue of intellectual property rights [3, 11, 27, 32,

107, 160]. As various biotechnology and pharmaceutical companies attempt to expand their libraries with compounds generated from various biodiversity sources, many countries are beginning to impose significant demands on the use of such raw materials and are limiting access without the expression of adequate consideration to the "landowners" [124]. Of course, ultimately there is the potential for significant economic reward for involved countries, but the various forms of life, which historically have been considered to be free for the taking, are no longer free. The potential conflict between host countries and prospecting organizations will undoubtedly be resolved with the establishment of contingent compensation plans making use of rights and royalties coupled with the attainment of specific milestones. The building of the requisite trust to establish meaningful, lasting, and functional business relationships will require time, patience, education, and respect for the different cultures involved [140]. History will color all negotiations and ethnic and cultural differences will serve to make the process painfully slow.

The search for and the development of natural products will result in the creation of a variety of different types of alliances between industry, government, individuals, universities, and hopefully even foster a spirit of international cooperation [25, 26, 32, 86]. Over the last few years, those involved in the pharmaceutical industry have already seen the increasing numbers of partnerships, alliances, agreements, and other types of relationships forged between large pharmaceutical companies and smaller biotechnology companies, all for the purpose of seizing opportunities of new technology. The collaborations that currently exist in the field of chemotherapy can set an example for other therapeutic areas as well as drug treatment in general. Another pertinent example has been the formation of the International Cooperative Biodiversity Group (ICBG) Program [32]. The ICBG is a collection of academic, industrial, U.S. governmental organizations, and developing countries. This program is also jointly sponsored by the National Science Foundation (NSF), parts of the National Institutes of Health (NIH), the National Cancer Institute (NCI), the Fogarty International Center, the National Institute of Allergy and Infectious Diseases (NIAID), the National Heart, Lung, and Blood Institute (NHLBI), and the National Institute of Mental Health (NIMH). The purpose of this program is to facilitate research and drug discovery from natural sources. This program is of extreme value and promotes the identification, establishment of an inventory, conservation of natural resources, and economic development of financially challenged countries.

Thought should be given to the more expanded establishment of repository programs. The NCI has established a Natural Products Repository (NPR), which is an extremely valuable resource in the search for and discovery of new drugs. The opportunities are there, they just need to be found.

In conclusion, drug discovery can be significantly improved through the use of the knowledge to be gained from research into natural products. However, despite the powerful resource that natural products present to us, the knowledge that is locked therein can only be fully realized with proper management

of the resources, the parallel development of ancillary technologies, and the fostering of open and shared communication.

## REFERENCES


1. Abel, U., Koch, C., Speitling, M., Hansske, F. G. (2002). Modern methods to produce natural-product libraries. *Curr. Opin. Chem. Biol.*, *6*(4), 453–458.
2. Alvi, K. A. (2000). A strategy for rapid identification of novel therapeutic leads from natural products. In S. Cutler and H. Cutler (Eds.), *Biologically Active Natural Products*. CRC, Boca Raton, FL, pp. 185–195.
3. Artuso, A. (1997). Natural product research and the emerging market for biochemical resources. *J. Res. Pharmaceut. Econ.*, *8*(2), 3–23.
4. Banerji, A. (2000). Resurgence of natural product research—A Phoenix act, *PINSA-A. Proc. Ind. Natl. Sci. Acad., Part A: Phys. Sci.*, *66*(3/4), 383–392.
5. Barrett, D. (2002). From natural products to clinically useful antifungals. *Biochim. Biophys. Acta, 1587*, 224–233.
6. Barron, R. L., Vanscoy, G. J. (1993). Natural products and the athlete: Facts and folklore. *Ann. Pharmacother.*, *27*(5), 607–615.
7. Bauer, L. A. (1997). Individualization of drug therapy: Clinical pharmacokinetics and pharmacodynamics, basic concepts of pathophysiology and pharmacotherapy. In J. T. DiPiro, G. R. Matzke, L. M. Posey, R. L. Talbert, B. G. Wells, and G. C. Yee, (Eds.) (L. M. Posey, Sect. Ed.), *Pharmacotherapy, A Pathophysiologic Approach*, 3rd ed. Appleton & Lange, Stamford, CT, pp. 29–48.
8. Bhattaram, V. A., Graefe, U., Kohlert, C., Veit, M., Derendorf, H. (2002). Pharmacokinetics and bioavailability of herbal medicinal products. *Phytomedicine*, *9*(Suppl. 3), 1–33.
9. Bindseil, K. U., Jakupovic, J., Wolf, D., Lavayre, J., Leboul, J., van der Pyl, D. (2001). Pure compound libraries; a new perspective for natural product based drug discovery. *Drug Discov. Today*, *6*(16), 840–847.
10. Borman, S. (1998). Biosynthesis combinatorially. *Chem. Eng. News*, September 14, pp. 29–30.
11. Boyd, M. R. (1996). The position of intellectual property rights in drug discovery and development from natural products. *J. Ethnopharmacol.*, *51*(1–3), 17–25; discussion, 25–27.
12. Breinbauer, R., Vetter, I. R., Waldmann, H. (2002). From protein domains to drug candidates—Natural products as guiding principles in the design and synthesis of compound libraries. *Ang. Chem. Int. Ed.*, *41*(16), 2878–2890.
13. Breinbeuer, R., Manger, M., Scheck, M., Waldmann, H. (2002). Natural product guided compound library development. *Curr. Med. Chem.*, *9*, 2129–2145.
14. Burja, A. M., Banaigs, B., Abou-Mansour, E., Burgess, J. G., Wright, P. C. (2001). Marine cyanobacteria—A prolific source of natural products. *Tetrahedron*, *57*, 9347–9377.
15. Cannell, R. J. P. (1998). Follow-up of natural product isolation. *Methods Biotechnol.*, *4*(Natural Products Isolation), 425–463.

16. Cannell, R. J. P. (1998). How to approach the isolation of a natural product. In R. J. P. Cannell (Ed.), *Methods in Biotechnology*, Vol. 4: *Natural Products Isolation.* Humana, Totowa, NJ, pp. 1–51.
17. Chan, K. (1995). Progress in traditional Chinese medicine. *Trends Pharmacol. Sci.*, *16*(6), 182–187.
18. Chung, I., Kim, Y., Ahn, J., Lee, H., Chen, G., Manji, H. K., Potter, W. Z., Pickar, D. (1995). Pharmacologic profile of natural products used to treat psychotic illnesses. *Psychopharmacol. Bull.*, *31*(1), 139–145.
19. Claeson, P., Bohlin, L. (1997). Some aspects of bioassay methods in natural-product research aimed at drug lead discovery. *Trends Biotechnol.*, *15/17*, 245–248.
20. Clark, A. M. (1996). Natural products as a resource for new drugs. *Pharmaceut. Res.*, *13/8*, 1133–1141.
21. Compostella, F., Lay, L. (2001). Critical surveys covering the year 2000: Total syntheses of natural products. In A. Corbella (Ed.), *Seminars in Organic Synthesis, Summer School, 26th*, Gargano, Italy, June 18–22, 2001. Societa Chimica Italiana, Rome, Italy, pp. 453–581.
22. Concannon, J. A., DeMeo, T. E. (1997). Goldenseal: Facing a hidden crisis. *Endangered Species Bull.*, *22*(6), 10–12.
23. Connolly, J. D. (1997). Natural products from around the world. *Revista Latinoamer. Quim.*, *25*(2), 77–85.
24. Cordell, G. A. (1998). *Nat. Prod. Updates*, *6*(1), 2.
25. Cordell, G. A. (1995). Natural products as medicinal and biological agents: Potentiating the resources of the rain forest. In P. F. Seidel, O. R. Gottlieb, and M. A. Colho Kaplan (Eds.), *Chemistry of the Amazon.* American Chemical Society Symposium Series, No. 588, Washington, DC, pp. 8–18.
26. Cordell, G. A. (1993). Pharmacognosy—New roots for an old science. In F. Z. Atta-ur-Rahman and Basha (Eds.), *Studies in Natural Products Chemistry,* Vol. 13: *Bioactive Natural Products (Part A)*. Elsevier Science, Amsterdam, pp. 629–700.
27. Cordell, G. A. (2002). Recent developments in the study of biologically active natural products. *ACGC Chem. Commun.*, *14*, 31–63.
28. Cragg, G. M., Schepartz S. A., Suffness M., Grever M. R. (1993). The taxol supply crisis. New NCI policies for handling the large-scale production of novel natural product anticancer and anti-HIV agents. *J. Nat. Prod.*, *56*, 1657–1668.
29. Cragg, G. M., Newman, D. J. (2001). Natural product drug discovery in the next millennium. *Pharmaceut. Biol.*, *39*(Suppl.), 8–17.
30. Cragg, G. M., Newman, D. J. (2001). Natural products drug discovery and development at the United Sates National Cancer Institute. In L. Yuan (Ed.), *Drug Discovery and Traditional Chinese Medicine: Science, Regulation, and Globalization*, [International Conference on Traditional Chinese Medicine: Science, Regulation and Globalization], 1st, College Park, MD, August 30–September 2, 2000 (2001) (meeting Date 2000). Kluwer Academic, Hingham, MA, pp. 19–32.
31. Cragg, G. M., Newman, D. J., Snader, K. M. (1997). Natural products in drug discovery and development. *J. Nat. Prod.*, *60*(1), 52–60.
32. Cragg, G. M. (1998). Paclitaxel (Taxol®): A success story with valuable lessons for natural product drug discovery and development. *Med. Res. Rev.*, *5*, 315–331.

33. Crews, D., Willingham, E., Skipper, J. K. (2000). Endocrine disruptors: Present issues, future directions. *Q. Rev. Biol.*, *75*(3), 243–260.
34. Da Rocha, A. B., Lopes, R. M., Schwartsmann, G. (2001). Natural products in anticancer therapy. *Curr. Opin. Pharmacol.*, *1*(4), 364–369.
35. Dahanukar, S. A., Kulkarni, R. A., Rege, N. N. (2000). Pharmacology of medicinal plants and natural products. *Ind. J. Pharmacol.*, *32*, S81–S118.
36. Davignon, J. P., Cradock J. C. (1987). In S. K. Carter and K. Hellman (Eds.), *Principles of Chemotherapy.* McGraw-Hill, New York, pp. 212–250.
37. Demain, A. L., Elander, R. P. (1999). The beta-lactam antibiotics: Past, present, and future. *Antonie Van Leeuwenhoek*, *75*(1/2), 5–19.
38. Demain, A. L. (1998). Microbial natural products: Alive and well in 1998. *Nat. Biotechnol.*, *16*(1), 3–4.
39. DerMarderosian, A., Beutler, J. A. (Eds.), *The Review of Natural Products*, 3rd ed. Facts and Comparisons, St. Louis, MO, pp. 824–826.
40. Duke, S. O., Dayan, F. E., Romagni, J. G., Rimando, A. M. (2000). Natural products as sources of herbicides: Current status and future trends. *Weed Rese.*, *40*(1), 99–111.
41. El Sayed, K. A. (2000). Natural products as antiviral agents. *Stud. Nat. Prod. Chem.*, *24* [Bioactive Natural Products (Part E)], 473–572.
42. Ellis, N. (2001). Extracting APIs from natural products. *Speciality Chem. Mag.*, *21*(6), 10–12.
43. Fallis, A. G. (1999). 1998 Alfred Bader Award Lecture. Tangents and targets: The synthetic highway from natural products to medicine. *Can. J. Chem.*, *77*(2), 159–177.
44. Farnsworth, N. R., Akerele, O., Bingel, A. S., Soejarto, D. D., Guo, Z. (1985). Medicinal plants in therapy. *Bull WHO*, *63*, 965–981.
45. Faulkner, D. J. (1998). Marine natural products. *Nat. Prod. Repts.*, *15*(2), 113–158.
46. Faulkner, D. J. (2000). Highlights of marine natural products chemistry (1972–1999). *Nat. Prod. Repts.*, *17*(1), 1–6.
47. Faulkner, D. J. (2000). Marine natural products. *Nat. Prod. Repts.*, *17*(1), 7–55.
48. Faulkner, D. J. (2002). Marine natural products. *Nat. Prod. Repts.*, *19*(1), 1–48.
49. Foucault, A. P., Chevolot, L. (1998). Counter-current chromatography: Instrumentation, solvent selection and some recent applications to natural product purification. *J. Chromatogr. A*, *808*(1–2), 3–22.
50. Foucault, A. P. (1997). Recent advances in purification of natural products by countercurrent chromatography. In P. Schreier (Ed.), *Natural Products Analysis: Chromatography, Spectroscopy, Biological Testing*, Symposium, Wuerzburg, Germany, September 1997. Vieweg, Wiesbaden, Germany, pp. 13–25.
51. Fuchs, R. L., Thomas, J. A. (Eds.) (2002). *Biotechnology and Safety Assessment*, 3rd ed. Academic Press, San Diego, CA.
52. Fugh-Berman, A., Cott, J. (1999). Dietary supplements and natural products as psychotherapeutic agents. *Psychosomatic Med.*, *61*(5), 712–728.
53. Furstner, A. (1999). Venturing into catalysis based natural product synthesis. *Synlett*, *10*, 1523–1533.

54. Gad, S. C. (2003). Principles of screening and study design. In *Safety Pharmacology in Pharmaceutical Development and Approval.* CRC, Boca Raton, FL, pp. 23–42.
55. Gad, S. C., Chengelis, C. P. (1998). *Acute Toxicology Testing*, 2nd ed. Academic, San Diego, CA.
56. Gad, S. C. (2002). *Drug Safety Evaluation.* Wiley-Interscience, New York.
57. Gad, S. C. (2001). *Regulatory Toxicology*, 2nd ed. Taylor & Francis.
58. Gad, S. C. (1995). *Safety Assessment for Pharmaceuticals.* Van Nostrand Reinhold.
59. Ganesan, A. (2001). Integrating natural product synthesis and combinatorial chemistry. *Pure Appl. Chem.*, *73*(7), 1033–1039.
60. Grabley, S., Sattler, I. (2003). Natural products for lead identification: Nature is a valuable resource for providing tools. In A. Hillisch and R. Hingenfeld (Eds.), *Modern Methods of Drug Discovery.* Birkhäuser Verlag, Switzerland, pp. 87–107.
61. Grabley, S., Thiericke, R., Sattler, I. (2000). Tools for drug discovery: Natural product-based libraries. *Ernst Schering Res. Found. Workshop*, *32*, 217–252.
62. Haefner, B. (2003). Drugs from the deep: Marine natural products as drug candidates. *Drug Discov. Today*, *8*(12), 536.
63. Hall, D. G., Manku, S., Wang, F. (2000). Solution- and solid-phase strategies for the design, synthesis, and screening of libraries based on natural product templates: A comprehensive survey. *J. Combinatorial Chem.*, *3*(2), 125–150.
64. Harvey, A. (2000). Strategies for discovering drugs from previously unexplored natural products. *Drug Discov. Today*, *5*(7), 294–300.
65. Harvey, A. (2001). The continuing value of natural products to drug discovery. *GIT Lab. J.*, *5*(6), 284–285.
66. Harvey, A. L. (1999). Medicines from nature: Are natural products still relevant to drug discovery? *Trends Pharmacol. Sci.*, *20*(5), 196–198.
67. Havsteen, B. (1983). Flavonoids, a class of natural products of high pharmacological potency. *Biochem. Pharmacol.*, *32*(7), 1141–1148.
68. Hemmer, M. C., Gasteiger, J. (2000). Prediction of three-dimensional molecular structures using information from infrared spectra. *Anal. Chim. Acta*, *420*, 145–154.
69. Hill, D. C. (1998). Advanced screening technology and informatics for natural products drug discovery. In A. L. Harvey (Ed.), *Advances in Drug Discovery Techniques.* Wiley, Chichester, pp. 25–37.
70. Hilton, M. D. (2003). Natural products: Discovery and screening. In V. A. Vinci and S. R. Parekh (Eds.), *Handbook of Industrial Cell Cultures.* Humana, Totowa, NJ, pp. 107–136.
71. Holmes, F. A., Walters, R. S., Thierault, R. L., Forman, A. D., Newton, L. K., Raber, M. N., Buzdar, A. U., Frye, D. K., Hortobagyi, G. N. (1991). Phase II trial of taxol, an active drug in the treatment of metastatic breast cancer. *J. Natl. Cancer Inst.,* *83*, 1797–1805.
72. Holt, G. A., Chandra, A. (2002). Herbs in the modern healthcare environment—An overview of uses, legalities, and the role of the healthcare professional. *Clin. Res. Regulatory Affairs (USA)*, *19*, 83–107.

73. Horton, T. (1972). Large-scale gel filtration for purification of natural products. *Am. Lab.*, *4*(5), 83–84, 86–88, 90–91.
74. Hostettmann, K., Hamburger, M., Hostettmann, M., Marston, A. (1991). New developments in the separation of natural products. *Recent Adv. Phytochem.*, *25* (Modern Phytochemical Methods), 1–32.
75. Hostettmann, K., Marston, A. (1990). Liquid-liquid partition chromatography in natural product isolation. *Anal. Chim. Acta*, *236*(1), 63–76.
76. Jarvis, B. B. (2000). The role of natural products in evolution. *Recent Adv. Phytochem.*, *34* (Evolution of Metabolic Pathways), 1–24.
77. Jia, W., Liu, V. J. K., Tang, L. (2002). Chinese herbal drugs for diabetes. *Frontiers Biotechnol. Pharmaceut.*, *3*, 337–358.
78. Johnston, B. A. (1998). Major diversity loss: 1 in 8 plants in global study threatened. *HerbalGram, 43*, 54.
79. Jung, M., Lee, S., Kim, H., Kim, H. (2000). Recent studies on natural products as anti-HIV agents. *Curr. Med. Chem.*, *7*(6), 649–661.
80. Kaufmann, B., Christen, P. (2002). Recent extraction techniques for natural products: Microwave-assisted extraction and pressurized solvent extraction. *Phytochem. Anal.*, *13*(2), 105–113.
81. Kaul, P. N., Joshi, B. S. (2001). Alternative medicine: Herbal drugs and their critical appraisal—Part II. *Prog. Drug Res., 57*, 1–75.
82. Keim, C., Ladisch, M. R. (2000). Bioseparation of natural products. *Prog. Biotechnol.*, *16* (Bioseparation Engineering), 15–20.
83. Kelecom, A. (1999). Chemistry of marine natural products: Yesterday, today and tomorrow. *Anais da Academia Brasileira de Ciencias, 71*(2), 249–263.
84. Kerr, R. G., Kerr, S. S. (1999). Marine natural products as therapeutic agents. *Expert Opin. Therapeutic Patents*, *9*(9), 1207–1222.
85. Khosla, C. (2000). Natural product biosynthesis: A new interface between enzymology and medicine. *J. Org. Chem.*, *65*(24), 8127–8133.
86. Koch, C., Neumann T., Thiericke R., Grabley S. (1999). A central natural product pool—New approach in drug discovery strategies. In S. Grabley (Ed.), *Drug Discovery from Nature*. Hans-Knoll-Institut fur Naturstoff-Forschung Jena, Germany, pp. 51–55.
87. Koeller, J., Yee., G. C. (1997). Molecular biology and biotechnology drugs, basic concepts of pathophysiology and pharmacotherapy. In J. T. DiPiro, G. R. Matzke, L. M. Posey, R. L. Talbert, B. G. Wells, and G. C. Yee (Eds.), (L. M. Posey, Section Editor), *Pharmacotherapy, A Pathophysiologic Approach,* 3rd ed. Appleton & Lange, Stamford, CT, pp. 117–128.
88. Koh, H., Woo, S. (2000). Chinese proprietary medicine in Singapore: Regulatory control of toxic heavy metals and undeclared drugs. *Drug Safety*, *23*(5), 351–362.
89. Konig, G. M., Wright, A. D. (1996). Marine natural products research: Current directions and future potential. *Planta Med.*, *62*(3), 193–211.
90. Kroll, D. J. (2001). Concerns and needs for research in herbal supplement pharmacotherapy and safety. *J. Herbal Pharmacother.*, *1*(2), 3–23.
91. Lake, J. (2000). Natural product-derived treatments of neuropsychiatric disorders: Review of progress and recommendations. *Stud. Nat. Prod. Chem.*, *24*, No. Bioactive Natural Products (Part E), 1093–1137.

92. Lake, J. (2000). Psychotropic medications from natural products: A review of promising research and recommendations. *Alternative Ther. Health Med.*, *6*(3), 36, 39–45, 47–52.
93. Lang, Q., Wai, C. M. (2001). Supercritical fluid extraction in herbal and natural product studies—A practical review. *Talanta*, *53*(4), 771–782.
94. Lang, F., Keller, K., Ihrig, M., Oudtshoorn-Eckard, J., Moller, H., Srinivasan, S., Yu, H. (2001). Biopharmaceutical characterisation of herbal medicinal products. *Pharmazeutische Industrie*, *63*(10), 1005–1010.
95. Lee, K. H. (1999). Anticancer drug design based on plant-derived natural products. *J. Biomed. Sci.*, *6*(4), 236–250.
96. Liberra, K., Lindequist, U. (1995). Marine fungi—A prolific resource of biologically active natural products? *Pharmazie*, *50*(9), 583–588.
97. Lindel, T., Junker, J., Köck, M. (1997). COCON: From NMR correlation data to molecular constitutions. *J. Mol. Model*, *3*, 364–368.
98. Mabry, T. J. (2001). Selected topics from forty years of natural products research: Betalains to flavonoids, antiviral proteins, and neurotoxic nonprotein amino acids. *J. Nat. Products*, *64*(12), 1596–1604.
99. Majetich, G. Recent trends in the use of natural products and their derivatives as potential pharmaceutical agents. In S. J. Cutler and H. G. Cutler (Eds.), *Biologically Active Natural Products.* CRC, Boca Raton, FL, pp. 61–71.
100. Manfredi, J. J., Horwitz, S. B. (1984). Taxol: An antimitotic agent with a new mechanism of action. *Pharmacol. Ther., 25*, 83.
101. Mann, J. (2001). Natural products as immunosuppressive agents. *Nat. Prod. Repts.*, *18*(4), 417–430.
102. Mann, J. (2002). Natural products in cancer chemotherapy: Past, present and future. *Nat. Rev. Cancer*, *2*(2), 143–148.
103. Marko, I. (2001). Natural product synthesis: The art of total synthesis. *Science*, *294*(5548), 1842–1843.
104. Marriott, B. M. (2001). The role of dietary supplements in health. An overview in the United States. *Adv. Exper. Med. Biol.*, *492* (Nutrition and Cancer Prevention), 203–217.
105. Mathieu, M. (1997). *Biologics Development: A Regulatory Overview*, 2nd ed. PAREXEL International Corporation.
106. Mathieu, M., Evans, A. G. (1997). *New Drug Development A Regulatory Overview*, Fourth Edition, PAREXEL International Corporation.
107. Mays, T. D., Mazan, K. D. (1996). Legal issues in sharing the benefits of biodiversity prospecting, *J. Ethnopharmacol.*, *51*(1–3), 93–102; discussion 102-9.
108. McChesney, J. D. (2000). Commercialization of plant-derived natural products as pharmaceuticals: a view from the trenches. In S. J. Cutler and H. G. Cutler (Eds.), *Biologically Active Natural Products.* CRC, Boca Raton, Fla., pp. 253–264.
109. McCoy, M. (1999). BIOCATALYSIS GROWS FOR DRUG SYNTHESIS. *Chem. Eng. News*, January 4, pp. 10–13.
110. McGuire, W. P., Rowinsky, E. K., Rosenheim, N. B., Grumbine, F. C., Ettinger, D. S., Armstong, D. K., Donehower, R. C. (1989). Taxol: a unique antineoplastic agent

with significant activity in advanced ovarian epithelial neoplasms. *Annals of Internal Medicine*, *111*, 273–279.

111. Mehta, G., Singh, V. (2002). Hybrid systems through natural product leads: an approach towards new molecular entities. *Chem. Soc. Rev.*, *31*(6), 324–334.
112. Mehta, R. G., Pezzuto, J. M. (2002). Discovery of cancer preventive agents from natural products: from plants to prevention. *Curr. Oncol. Rep.*, *4*(6), 478–486.
113. Miles, D. H., Nguyen, C. L., Miles, D. H. (1998). Utilization of natural products for treatment of blood diseases. *Curr. Med. Chem.*, *5*(6), 421–440.
114. Müller, H., Brackhagen, O., Brunne, R., Henkel, T., Reichel, F. (2000). Natural products in drug discovery. *Ernst Schering Res. Found. Workshop*, *32*, 205–216.
115. Nakanishi, K. (1999). An historical perspective of natural products chemistry. *Comprehensive Nat. Prod. Chem.*, *8*, xxi–xxxviii.
116. Nakanishi, K. (1999). An historical perspective of natural products chemistry. In S. Ushio (Ed.), *Comprehensive Natural Products Chemistry*, Vol. 1. Elsevier Science B.V., Amsterdam, pp. 23–40.
117. Neidig, K. P., Geyer, M., Gorler, A., Antz, C., Saffrich, R., Beneicke, W., Kalbitzer, H. R. (1995). Aurelia, a program for computer-aided analysis of multidimensional NMR spectra. *J. Biomol. NMR*, *6*, 255–270.
118. Newman, D. J., Cragg, G. M., Snader, K. M., (2000). The influence of natural products upon drug discovery. *Nat. Prod. Repts.*, *17*, 215–234.
119. Nicolaou, K. C., Pfefferkorn, J. A. (2001). Solid phase synthesis of complex natural products and libraries thereof. *Biopolymers*, *60*(3), 171–193.
120. Nielsen, J. (2002). Combinatorial synthesis of natural products. *Curr. Opin. Chem. Biol.*, *6*(3), 297–305.
121. Nies, A. S. (2001). Principles of therapeutics. In J. G. Hardman and L. E. Limbird (Eds.), A. G. Gilman (Consulting Editor), *Goodman & Gilman's The Pharmacological Basis of Therapeutics*, McGraw-Hill, New York, pp. 45–66.
122. Nisbet, L. J., Moore, M. L. (1997). Will natural products remain an important source of drug research for the future? *Curr. Opin. Biotechnol.*, *8*(6), 708–712.
123. O'Keefe, B. R. (2001). Biologically active proteins from natural product extracts. *J. Nat. Prod.*, *64*, 1373–1381.
124. O'Neill, M. J., Lewis, J. A. (1993). The renaissance of plant research in the pharmaceutical industry. In A. D. Kinghorn and M. F. Balandrin (Eds.), *Human Medicinal Agents from Plants*, ACS Symposium No. 534. American Chemical Society, Washington, DC, pp. 48–55.
125. Patchett, A. (2002). 2002 Alfred Burger Award Address in Medicinal Chemistry. Natural products and design: Interrelated approaches in drug discovery. *J. Med. Chem.*, *45*(26), 5609–5616.
126. Puri, R. K. (1998). Supercritical fluid extraction and its applications in natural products. *Ind. J. Nat. Prod.*, *14*(2), 3–23.
127. Qin, G., Xu, R. (1998). Recent advances on bioactive natural products from Chinese medicinal plants. *Med. Res. Rev.*, *18*(6), 375–382.
128. Quinn, R. J. (1999). High-throughput screening in natural product drug discovery in Australia utilising Australia's biodiversity. *Drug Devel. Res.*, *46*(3/4), 250–254.

129. Quinn, R. J. (1999). QPRI's system for screening of natural products. *Proc. Phytochem. Soc. Eur.*, *43* (Bioassay Methods in Natural Product Research and Drug Development), 151–157.
130. Ram, V. J., Kumari, S., (2001). Natural products of plant origin as anticancer agents. *Drug News and Perspectives*, *14*(8), 465–482.
131. Rang, H. P., Dale, M. M., Ritter, J. M., Gardner, P. *Pharmacology.* Churchill Livingstone, New York, NY, 1995.
132. Raven, P. H. (1988). Our diminishing tropical forests. In E. O. Wilson (Ed.), *Bio-Diversity*. National Academy Press, Washington, DC, pp. 119–122.
133. Righi, P. (1999). Critical surveys covering the year 1998: Total syntheses of natural products. In A. Corbella (Ed.), *Seminars in Organic Synthesis, Summer School, 24th*, Gargano, Italy, June 14–18, 1999. Societa Chimica Italiana, Rome, Italy, pp. 401–518.
134. Robinson, D. H., Narducci, W. A., Ueda, C. T. (1997). Drug delivery and administration, basic concepts of pathophysiology and pharmacotherapy. In J. T. DiPiro, G. R. Matzke, L. M. Posey, R. L. Talbert, B. G. Wells, and G. C. Yee (Eds.) (Posey, L.M., Section Editor), *Pharmacotherapy, A Pathophysiologic Approach*, 3rd ed. Appleton & Lange, pp. 49–76.
135. Rodriguez, E., McDaniel, R. (2001). Combinatorial biosynthesis of antimicrobials and other natural products. *Curr. Opin. Microbiol.*, *4*(5), 526–534.
136. Roessner, C. A., Scott, A. I. (1996). Genetically engineered synthesis of natural products: From alkaloids to corrins. *Annu. Rev. Microbiol.*, *50*, 467–490.
137. Ross, E. M., Kenakin, T. P. (2001). Pharmacodynamics: Mechanisms of drug action and the relationship between drug concentration and effect, in J. G. Hardman, and L. E. Limbird (Eds.) (A. G. Gilman, Consulting Editor), *Goodman & Gilman's The Pharmacological Basis of Therapeutics*, 10th ed. McGraw-Hill, New York, pp. 31–44.
138. Rouhi, A. M. (1999). Chemistry from unknown microbes. *Chem. Eng. News, 77*, January 4, pp. 21–22.
139. Rouhi, A. M. (1996). Natural products' scope expanding. IUPAC symposium showcases breadth of approaches chemists use to understand complex molecules from natural sources. *Chem. Eng. News*, *74*(43), 34–43.
140. Rouhi, A. M. (1997). Seeking drugs in natural products. *Chem. Eng. News*, *75*(14), 14–20.
141. Rowinsky, E. K., Cazenave L. A., Donehower R. C. (1990). Taxol: A novel investigational antimicrotubule agent. *J. Natl. Cancer Inst.*, *82*, 1247–1259.
142. Sandsborg, W. N. A., Rolfsen, N. (1999). Natural products in drug discovery and development. *Proc. Phytochem. Soc. Eur. Bioassay Methods Nat. Prod. Res. Drug Devel.*, *43*, 143–149.
143. Sausville, E., Johnson, J., Alley, M., Zaharevitz, D., Senderowicz, A. M. (2000). Inhibition of CDKs as a therapeutic modality. *Ann. N.Y. Acad. Sci.*, *910* (Colorectal Cancer), 207–222.
144. Scheuer, P. J. (1995). Marine natural products research: A look into the dive bag. *J. Nat. Prod. (Lloydia)*, *58*(3), 335–343.
145. Schoental, R. (1965). Toxicology of natural products. *Food Cosmetics Toxicol.*, *3*(4), 609–620.

146. Schuster, B. G. (2001). Demonstrating the validity of natural products as anti-infective drugs. *J. Alternative Complementary Med.*, *7* (Suppl. 1), S73–82.
147. Schwartsmann, G., Ratain, M. J., Cragg, G. M., Wong, J. E., Saijo, N., Parkinson, D. R., Fujiwara, Y., Pazdur, R., Newman, D. J., Dagher, R., DiLeone, L. (2002). Anticancer drug discovery and development throughout the world. *J. Clin. Oncol.* (Sept. 15), *20* (Suppl. 18), 47S–59S.
148. Shapiro, K., Gong, W. C. (2002). Natural products used for diabetes. *J. Am. Pharmaceut. Assoc.*, *42*, 217–226.
149. Shimizu, Y. (1998). Purification of water-soluble natural products. *Methods Biotechnol.*, *4* (Natural Products Isolation), 329–341.
150. Silva, C. J., Brian, P., Peterson, T. (2001). Screening of combinatorial biology libraries for natural products discovery. *Drugs Pharmaceut. Sci.*, *114* (Handbook of Drug Screening), 357–382.
151. Sorensen, E. J. (1999). Selected, recent developments in the chemical synthesis of biologically-active natural products. *Curr. Opin. Drug Discov. Devel.*, *2*(6), 606–630.
152. Spainhour, C. B. (2001). OTC drugs and nutraceuticals. In S. C. Gad (Ed.), *Regulatory Toxicology*, 2nd ed. Taylor & Francis, London, pp. 192–204.
153. Steinbeck, C. (2001). A platform-independent, distributed, and parallel system for computer-assisted structure elucidation in organic chemistry. *J. Chem. Inf. Comput. Sci.*, *41*, 1500–1507.
154. Steinbeck, C. (2001). The automation of natural product structure elucidation. *Curr. Opin. Drug Discov. Devel.*, *4*(3), 338–342.
155. Tsoi, C. J., Khosla, C. (1995). Combinatorial biosynthesis of "unnatural" natural products: The polyketide example. *Chem. Biol.*, *2*(6), 355–362.
156. Vlietinck, A. J., Apers, S. (2001). Biological screening methods in the search for pharmacologically active natural products. In C. Tringali (Ed.), *Bioactive Compounds from Natural Sources.* Taylor & Francis, London, pp. 1–29.
157. Volkman, J. K. (1999). Australasian research on marine natural products: Chemistry, bioactivity and ecology. *Marine Freshwater Res.*, *50*(8), 761–779.
158. Von Dohren, H. (1995). Peptides. *Biotechnology*, *28*, 129–171.
159. Waller, D. G., Renwick, A. G., Hillier, K. (2001). *Medical Pharmacology and Therapeutics.* W. B. Saunders, Harcourt Publishers Limited, Edinburgh.
160. Waterman, P. G. (1998). Natural products for drug discovery: An overview. In A. L. Harvey (Ed.), *Advances in Drug Discovery Techniques.* J Wiley, Chichester, pp. 13–23.
161. Wessjohann, L. A. (2000). Synthesis of natural-product-based compound libraries. *Curr. Opin. Chem. Biol.*, *4*(3), 303–309.
162. Whitehead, R. (1999). Natural product chemistry. *Annu. Repts. Progr. Chem., Sect. B: Organic Chem.*, *95*, 183–205.
163. Wilson, L. J. (2000). Recent advances in solid-phase synthesis of natural products, In K. Burgess (Ed.), *Solid-Phase Organic Synthesis*, J Wiley, New York, pp. 247–267.
164. Woodruff, H. B. (1999). Natural products from microorganisms. An Odyssey revisited. *Actinomycetologica*, *13*(2), 58–67.

165. Wright, A. (1998). Isolation of marine natural products. *Methods Biotechnol.*, *4* (Natural Products Isolation), 365–408.
166. Wrigley, S. K., Chicarelli-Robinson, M. (1997). Natural products research and Pharmaceuticals in the 1990's. *Annu. Repts. Med. Chem.*, *32*, 285–294.
167. Yang, S. S., Cragg, G. M., Newman, D. J., Bader, J. P. (2001). Natural product-based anti-HIV drug discovery and development facilitated by the NCI Development Therapeutics Program. *J. Nat. Prod.*, *64*(2), 265–277.
168. Yao, X., Hu, K., Peng, J., Qiao, S., Qiu, F., Dong, A., Wang, N., Cui, C., You, S., Shao, G., Chen, Y., Xu, S. (1998). Current status and prospects of research on natural medicinal products in China. *Int. Congr. Ser.*, *1157* (Towards Natural Medicine Research in the 21st Century), 445–455.
169. Yu, D., Chen, Y., Liang, X. (2000). Structural chemistry and biological activities of natural products from Chinese herbal medicines—part II. *Res. Commun. Mol. Pathol. Pharmacol.*, *108*(5–6), 393–436.
170. Yu, D., Chen, Y., Liang, X. (2001). Structural chemistry and biological activities of natural products from Chinese herbal medicines. *Emerging Drugs*, *1* (Molecular Aspects of Asian Medicines), 215–321.

# CANCER CELL PROTEOMICS USING MOLECULAR APTAMERS

Weihong Tan, Zehui Cao, Dihua Shangguan, Ying Li, Zhiwen Tang, Prabodhika Mallikratchy, and Hui Chen

*University of Florida*
*Gainesville, Florida*

1 INTRODUCTION
2 MOLECULAR DIAGNOSIS OF CANCERS
3 APTAMERS FOR TUMOR MARKER DETECTION
4 FLUORESCENCE ENERGY TRANSFER FOR APTAMER-BASED PDGF ANALYSIS
5 APTAMERS AS FLUORESCENCE ANISOTRPY PROBES FOR PROTEIN ANALYSIS
6 WHOLE-CELL-BASED APTAMER SELECTION FOR MULTIPLE TUMOR MARKERS
7 CONCLUSION
References

## 1 INTRODUCTION

Tremendous progress has been made in understanding human cancers in recent decades. It is now well accepted that cancer is uncontrolled cell growth resulting from multiple mutated genes. These mutations could come from

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

either an imperfect deoxyribonucleic acid (DNA) copying process in cells or damages to DNA by environmental mutagens. Despite our increasing understanding of cancers, finding a cure for cancers remains a great challenge. One of the important aspects of the cancer treatment is the accurate diagnosis of cancers. As late-stage cancers can only be treated with a very low success rate, early detection of benign tumors is critically important for patient survival.

With recent advances, especially the combination of positron emission tomography (PET) and computed tomography (CT), cancer-imaging techniques are becoming increasingly important in tumor detection [1]. In PET, radioactively linked glucose is injected into the patient. Since tumor cells are more active in growth and division than normal cells, they need more resources such as glucose from blood. Thus the accumulation of radioisotope-labeled glucose in tumor cells can be used to locate the tumor. Combined with CT, PET provides clear images that show the location, size, and shape of the tumor. This technique has been useful not only in tumor detection but also in determining the stage of the cancer, monitoring the progression of the tumor, and guiding cancer treatment. However, one problem with cancer imaging is that it is associated with expensive scanning equipment. The test itself is also expensive. More economical and readily available cancer diagnoses, such as those based on molecular recognition, may be more applicable in many situations.

## 2 MOLECULAR DIAGNOSIS OF CANCERS

The molecular diagnosis of cancers relies on biomarker molecules that are produced in higher than normal levels either directly by tumor cells or by the response of the human body to the presence of cancers. Detection of the biomarkers in a patient's body fluids can serve as the first step in cancer diagnosis and provide critical information to doctors as to whether or not a biopsy is needed. Tumor markers can be proteins or hormones. Some classic tumor markers include α-fetoprotein (AFP), carcinoembryonic antigen (CEA), and prostate-specific antigen (PSA). They are usually not very specific to a particular cancer as the level of one tumor marker can be elevated by more than one type of cancers. Another problem is that the presence of cancer does not necessarily cause a detectable level of tumor markers, especially in the early stages. Extra caution is thus needed in some cases to avoid false negatives.

With a better understanding of the genetic basis of cancers, some biomolecules may emerge as a new class of tumor markers with greater sensitivity and selectivity. The mutated genes in cancer cells usually lead to the expression of new proteins not found in normal cells, the overexpression of certain proteins that promote cell growth and division, such as growth factors and related proteins, or the underexpression of proteins that inhibit tumor cell proliferation, such as the p53 protein. One example of a protein capable of indicating cancers is the human epidermal growth factor receptor 2 (HER2),

which is usually found on cell membranes. HER2 has been shown to be overexpressed in about 25 percent of all breast cancer patients [2]. Therefore, tests for the HER2 protein in tissue samples are recommended for breast cancer diagnosis. Overexpression of many other growth factors, including the insulin-like growth factor-I (IGF-I), the epidermal growth factor (EGF), and the platelet-derived growth factor (PDGF), are also related to tumor progression [3–5]. Sensitive detection of these proteins may serve as a good indication of both the presence and stage of cancers.

## 3 APTAMERS FOR TUMOR MARKER DETECTION

Over the last decades, antibodies have been the preferred recognition agents for protein detection. Their very high affinity and selectivity toward the target molecules ensure sensitive detection of specific proteins even in relatively complex samples. Antibodies for many important proteins have been developed, and some of them are routinely used for the clinical diagnosis of various diseases.

Despite these great successes in molecular diagnosis, antibodies possess some limitations in certain applications. First, the production of antibodies relies on an animal host and is a rather long process. This presents significant challenges for using antibodies in proteomic research, where thousands of proteins that are not well characterized need to be analyzed simultaneously. Second, as a special type of protein, antibodies are also sensitive to features of their surrounding environments, such as pH, temperature, ionic strength, and so forth. Labeling antibodies with fluorescent dyes or the immobilization of antibodies on a solid surface may alter their tertiary structures and thus their binding properties to targets. Furthermore, the long-term storage of antibodies may pose problems in maintaining their activities, which adds to the already high cost of using antibodies for protein probes.

Designed to mimic some of the antibodies' functions, a new class of molecules called aptamers may greatly challenge the role of antibodies in molecular diagnosis in the near future. Aptamers are nucleic acids that have high affinity and selectivity for their target molecules. By using the exponential enrichment (SELEX) process [6, 7] for the systematic evolution of ligands, oligonucleotide sequences can be isolated to recognize virtually any class of molecules [8]. The SELEX process begins with a library of synthesized oligonucleotides usually containing $10^{14}$ to $10^{15}$ random sequences. This library is then incubated with the target molecule of interest under certain conditions. The sequences that interact with and bind to the target molecules are isolated and amplified by PCR for the next round of incubation. This process is repeated until sequences that bind to the target with highest affinity and selectivity are determined. Compared to antibodies, aptamers have similar high affinity and selectivity for proteins [9, 10]. Due to their structural stability, aptamers can withstand harsher experimental conditions than antibodies and can be stored, modified, and immobilized without causing much

degradation. It is also much easier and considerably cheaper to make aptamers using common DNA synthesizers.

Aptamers can be easily and site-specifically engineered with fluorescent labels for real-time protein recognition. By forming intramolecular signal transduction mechanisms, the labeled aptamers can report targets in homogenous solutions, eliminating the need to separate bound complex molecules from unbound molecules, which is often required in antibody-based protein assays. This translates into fast and convenient protein detection that is essential for large-scale high-throughput proteomic research. As an example, the real-time detection of PDGF in various approaches is demonstrated.

## 4 FLUORESCENCE ENERGY TRANSFER FOR APTAMER-BASED PDGF ANALYSIS

A member of a growth factor family, PDGF is involved in regulating cell growth and division. It is composed of a combination of primarily two types of subunits, A and B. Thus PDGF usually exists in three isoforms, PDGF-AA, AB, and BB. Overexpression of PDGF, especially the BB form, has been found in malignant tumors. A DNA aptamer that specifically binds to the PDGF-B chain was previously reported [11]. In physiological conditions, the stable conformation of the aptamer contains two regions of base pairing but is relatively unconstrained. Upon binding with PDGF, the aptamer forms a closely packed structure that includes a three-way helix junction with a conserved single-stranded loop at the junction point that is essential for binding (Fig. 1). Based on this fact, the sensitive detection of PDGF can be achieved using fluores-

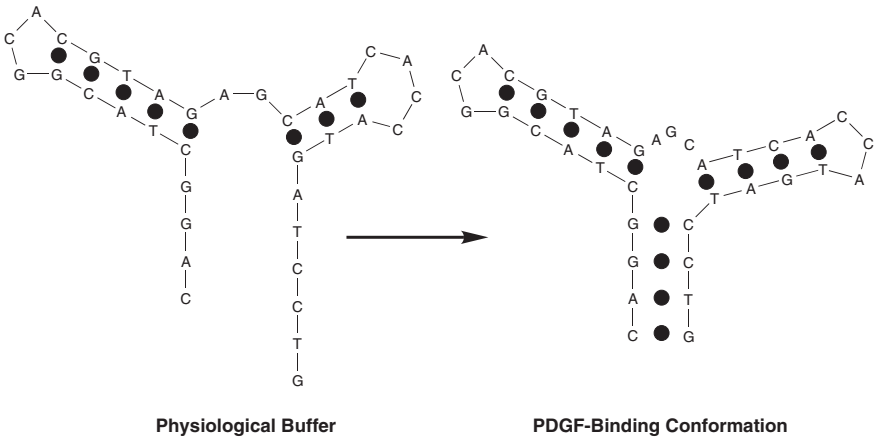


**Figure 1** Conformational change of PDGF aptamer. Binding of PDGF causes the aptamer to change from a relatively unconstrained structure to a tightly packed one containing a three-way helical junction with a conserved single-stranded loop.

cence resonance energy transfer (FRET). The energy transfer between "donor" and "acceptor" molecules with overlapping emission and absorption spectra is highly dependent on the distance between these two molecules. Significant FRET usually happens with a distance less than 10 nm. This forms the basis for a FRET-based PDGF assay using aptamer. When the donor and acceptor are labeled at both ends of the PDGF aptamer (Fig. 2), binding of the aptamer to PDGF will reduce the distance between the two labels to the level at which the energy transfer becomes significant, thus revealing the presence of PDGF. Two types of energy transfer can be utilized. One of them uses a nonfluorescent molecule as the acceptor, which is usually called the quencher. Therefore, no emission of the acceptor will be observed and the only indication of FRET is the quenching of the donor fluorescence. In the other method of conducting FRET, the fluorescent acceptor produces an emission upon energy transfer.

The PDGF aptamer has been turned into a probe specific for PDGF by labeling the 5′ end with fluorescein and the 3′ end with a DABCYL quencher [12]. Due to its resemblance to molecular beacons used for nucleic acid detection, this probe is given the name molecular beacon aptamer (MBA). The addition of PDGF-BB as low as sub-nanomolar causes a detectable reduction of the fluorescence signal in real time as the fluorophore and quencher are

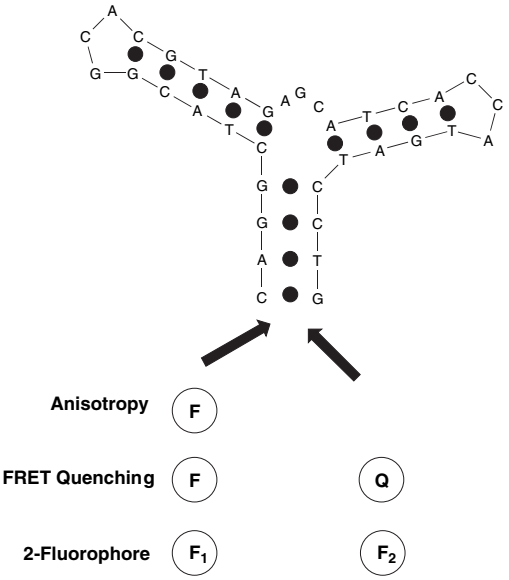


**Figure 2** Engineering a PDGF probe from the PDGF aptamer. By strategic placement of single or double fluorophores (F) and a quencher (Q), the aptamer can be transformed into fluorescence probes based on fluorescence anisotropy, FRET quenching, and 2-fluorophore FRET. Positioning the labels at either end of the aptamer sequence minimizes interference with aptamer–protein binding.

brought together upon binding. This FRET-based assay leads to PDGF detection with not only high sensitivity but also good selectivity. A variety of extracellular proteins and unrelated growth factors are mixed with the MBA and show no noticeable fluorescence change. On the other hand, the other two variants of PDGF, PDGF-AA and PDGF-AB, can also cause a decrease of fluorescence in the MBA, but to a lesser extent due to their lower binding affinity to the aptamer. This difference in binding affinity is applied to construct an assay to selectively analyze variants of PDGF. A sample solution with each of the three isoforms of PDGF undergoes a serial dilution with a physiological buffer. The same amount of MBA is then added to the aliquots. Fluorescence is read in each aliquot by a microtiter plate reader and a dose–response curve is created. The three isoforms have clearly different dose-response curves, with the slopes of the curves decreasing from PDGF-BB to AB and then AA (Fig. 3). The slope change can be attributed to differences in binding affinity to the MBA as more protein with lower affinity is needed to obtain the same MBA quenching level. The variants of the same protein can be differentiated using the MBA-based assay.

The same approach can be used to monitor PDGF-BB in real biological samples such as cell media. Conditioned media from different cell lines, including human breast carcinoma cells HTB-26 and normal murine BALB/3T3 fibroblast cells, are diluted multiple times and incubated with a fixed amount of the MBA in a 96-well microtiter plate to obtain dose–response curves. Similarly, serial dilution of a PDGF-BB in a buffer solution was employed to build another dose–response curve. Comparing these three sets of results, one can see that the curves for pure PDGF-BB and HTB-26 cells have very similar slopes and maximum extents of quenching, indicating something in the cell

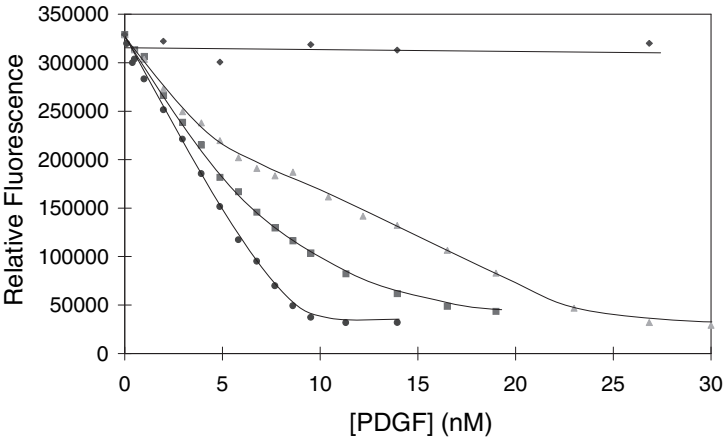


**Figure 3** Dose-response curves of PDGF variants. Fluorescence signals of MBA for serial dilutions of PDGF-AA (triangle), PDGF-AB (square), PDGF-BB (circle), and denatured PDGF-BB (diamond). The concentration of the MBA was 10 nM.

media had almost the same affinity to the aptamer as pure PDGF-BB (Fig. 4). In contrast, the curve for BALB/3T3 cells has a flatter slope and less of a quenching effect. Since the slope of a dose–response curve is independent of the actual concentration of the target, but only related to the affinity between target and probe, it is likely that the BALB/3T3 cells do not produce PDGF-BB actively.

A two-fluorophore-based FRET assay should also be suitable for PDGF detection. Previously, this approach has been used for the detection of human α-thrombin with thrombin aptamer [13]. The aptamer behaves similarly to PDGF aptamer. When bound to the target protein, it takes a G-quadruplex conformation, yielding a more compact structure than in free solution. Two fluorophores with overlapping spectra are labeled on the aptamer. Binding of the aptamer to α-thrombin brings the two dyes into close proximity, thus triggering the energy transfer between the two molecules. Given that the donor emission declines and the acceptor emission increases as a result of the FRET, the ratio of the donor and acceptor emissions is measured instead of the fluorescence of a single dye. A large difference between the two emissions is expected before and after protein binding, leading to more sensitive detection.

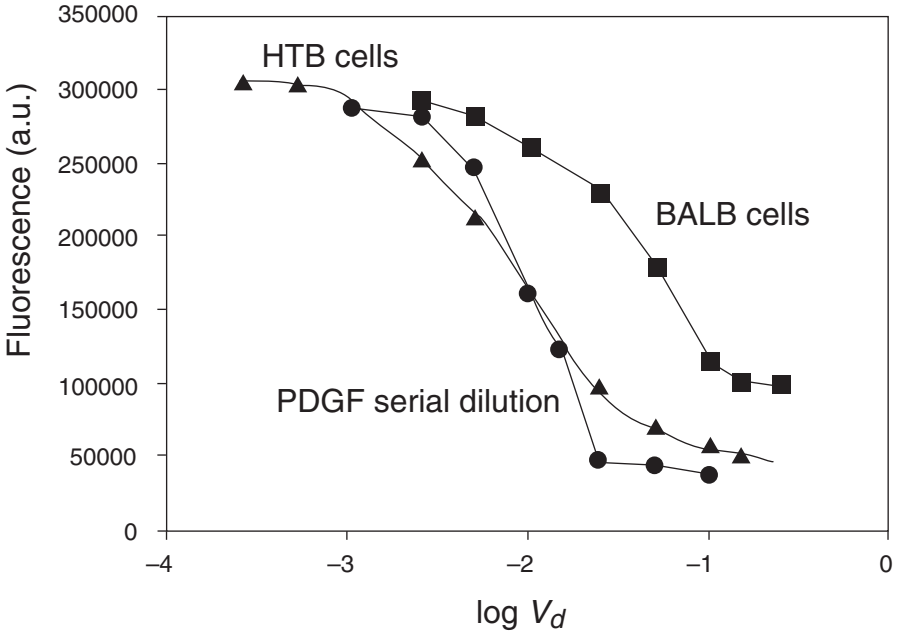


**Figure 4** Cell sample dose–response curves. Fluorescence signals of mixtures of 10 nM of MBA with serial dilutions of protein preparations of conditioned cell media from HTB cells (triangle) and BALB cells (square). Serial dilutions of a 500-nM solution of PDGF-BB (circle) were used as a control. Vd is the dilution factor.

The two-fluorophore assay has a threefold lower detection limit than the fluorophore–quencher assay for α-thrombin. A similar method can be employed to improve the sensitivity of the aptamer-based PDGF assay.

Other reports using different strategies to incorporate FRET into aptamer-based assays have also successfully detected various targets in real time [13–15]. Another alternative approach links ribozyme with aptamer [16]. Binding between the aptamer and the target activates the ribozyme, and a nucleic acid with fluorophore and quencher labels is cleaved, resulting in fluorescence increase.

## 5 APTAMERS AS FLUORESCENCE ANISOTROPY PROBES FOR PROTEIN ANALYSIS

Fluorescence anisotropy (FA) has been widely used for studying biomolecular interactions. A relatively small fluorescent molecule, when bound to a relatively large molecule, will demonstrate an increased anisotropy. The principles behind FA can be explained as follows. When excited by a polarized light, those fluorescent molecules with absorption transition moments oriented along the electric vector of the incident light are preferentially excited. The rotational diffusion of the molecules on the excited state leads to depolarized emission. The extent of the depolarization can be described in terms of anisotropy ($r$). Anisotropy is usually measured using the parallel emission $I_{\parallel}$ and perpendicular emission $I_{\perp}$:

$$r = (I_{\parallel} - I_{\perp}) / (I_{\parallel} + 2 \bullet I_{\perp})$$

Since FA is related to a molecule's ability to rotate in the excited state, the size as well as the molecular weight of the molecule plays an important role in determining this molecule's anisotropy.

The primary advantage of using aptamers as anisotropy probes is their relatively small sizes. Compared to monoclonal antibodies, they are in most cases much smaller than the target proteins. This makes aptamer-based anisotropy probes ideal for protein detection. With the aptamer conveniently labeled with one fluorophore, the anisotropy probe will report binding with a target protein via the increase in anisotropy of the fluorophore. Unlike the FRET-based assays, where the conformational change of the aptamer is essential for target detection, the aptamer anisotropy assay is not as heavily dependent on the structure of the aptamer probe. Therefore, it may be highly useful in applications where understanding of the aptamer structure is limited.

An anisotropy probe has been made by labeling the PDGF aptamer with one fluorescein dye [17]. Adding PDGF-BB to the aptamer results in an increased anisotropy of the aptamer due to the increased overall molecular weight upon complex formation. A detection limit of 2 nM PDGF-BB can be achieved. The selectivity of the assay is as good as that of the FRET-based

assay. This can be attributed to the fact that various other proteins do not affect the anisotropy of the aptamer.

Aptamer-based assays using either FRET or fluorescence anisotropy as a signal transduction mechanism have proven to be ideal for sensitive and selective protein analysis. Their ability to detect targets in real time and homogeneous solutions holds great potential for easy, cheap, and fast applications in clinical diagnosis. As the in vitro selection of aptamers has evolved into an automated process [18], more and more aptamers for proteins, including important cancer biomarkers, are expected to be available in the near future.

## 6 WHOLE-CELL-BASED APTAMER SELECTION FOR MULTIPLE TUMOR MARKERS

Recent advances in proteomic research, especially the emergence of new analytical techniques, have begun to show great impact on cancer diagnosis. Based on the simultaneous detection of multiple protein markers, these new approaches have already demonstrated excellent sensitivity and specificity in cancer diagnosis that are beyond the capability of traditional methods. Evolved from matrix-assisted laser desorption ionization (MALDI) mass spectrometry, a new technique called surface-enhanced laser/desorption ionization (SELDI) is capable of identifying proteins with high throughput. With the help of SELDI, biomarkers for specific cancers have been discovered and successfully applied to real-world cancer diagnosis [19, 20]. The idea of this new method is to analyze the protein content of serum from patients with known cancer by SELDI. The acquired mass spectra are then compared to those obtained from healthy people. An iterative searching algorithm is used to identify the proteins in cancer patient serum that can best differentiate cancer patients from healthy people. Searching the serum from unknown patients for the newly found protein markers can reveal whether the patients have this specific cancer. Cancer patients, including those with early-stage cancer, are correctly diagnosed in most cases.

Despite the initial success, problems associated with the new strategy have begun to come to people's attention [21]. One of the issues is that two different laboratories using the same approach to study the same type of cancer have identified very different sets of protein markers, yet they both have success in cancer diagnosis using their biomarkers. This may indicate that the discovery of biomarkers using SELDI may be highly sensitive to sample preparation and experimental conditions. Therefore, great care needs to be taken in order to achieve comparable results across different laboratories. Furthermore, expensive instruments in this new method may prevent the application of the biomarkers in small clinics.

It is clear from these new advances that multiple protein markers produced by cancers, once identified, can lead to more sensitive and specific cancer diagnosis than using a single tumor marker. Therefore, it is of great importance to

develop panels of probes for simultaneous biomarker detection in real time. One promising strategy for this purpose is to conduct aptamer selection targeted at a complex mixture of samples. In order to be applicable to cancer diagnosis, the use of cancer-cell-related complex protein mixtures as the target for selection is preferred.

The selection of high-affinity ligands for complex targets has been demonstrated both in theory [23] and in experiments for red blood cell membranes [22]. Since then, aptamers have been isolated for various complex systems, mostly cell membranes [24–27]. Since proteins not found or in low abundance on normal cell membranes may be present on tumor cell membranes, those proteins can serve as biomarkers for specific cancers. Whole-cell selection may yield a panel of aptamers that have high affinity and selectivity to the important tumor markers.

The procedural details of whole-cell selection vary between reports, but generally they incorporate the incubation of a random DNA or ribonucleic acid (RNA) library with target cells followed by the controlled elution of the sequences bound to the target cells. These sequences were subsequently amplified by polymerase chain reaction (PCR) and then reincubated with target cells iteratively. After several iterations (6 to 25), the pool of selected oligonucleotides was cloned and sequenced. A comparison of the sequence homology allowed the grouping of the aptamers into families that recognized distinct targets. For the discovery of tumor cell markers, since most abundant species on tumor cell membranes may not be related to cancer, it is advisable to do a counter selection using closely related normal cells. After incubation of the DNA or RNA library with the normal cells, only the unbound nucleic acids will proceed to be incubated with the tumor cells (Fig. 5). In this way, DNA or RNA molecules that bind to proteins that are present on both normal and tumor cells are not selected.

The selected aptamers can be engineered into protein-binding probes based on either FRET or fluorescence anisotropy. These probes can then be made into a kit for clinical diagnosis. Tissue samples from patients are incubated with the aptamer panel and the fluorescence signals can be read by something

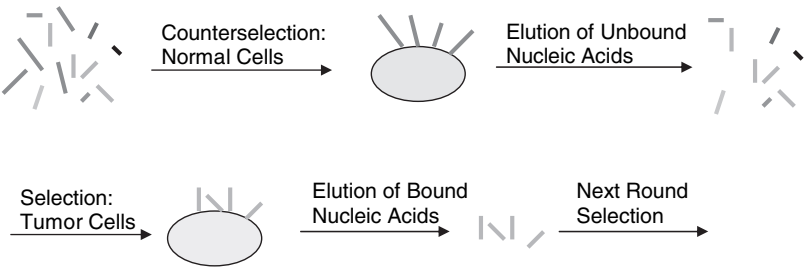


**Figure 5** Schematic of cell-based aptamer selection and counter selection.

as simple as a microtiter plate reader. Positive results from most of the aptamers will be strong evidence for the presence of tumor cells that have been particularly targeted. As a result, cancer patients, including those with early-stage cancer, may be diagnosed with high confidence.

Another way to incorporate the selected aptamer panel into cancer diagnosis is to attach the aptamers with either radioactive or fluorescent labels and inject them into the patient's body. With relatively small molecular weights, aptamers should provide fast tumor penetration and blood clearance, which makes them excellent for high-contrast imaging [28]. A whole-body scanning will reveal the localization of the aptamers inside the body. Coexistence of the aptamers in a particular part of the body should give a strong indication of a tumor in that area. Because multiple tumor binding probes are utilized, the aptamer tests should be superior to traditional antibody tests in terms of specificity and reliability.

## 7 CONCLUSION

Aptamers have the potential to be excellent probes for the recognition of single protein targets as well as for a panel of proteins. Combined with fluorescence-based signal transduction mechanisms, the detection of proteins using aptamers can be both effective and convenient with high sensitivity and selectivity. To take advantage of the capability of aptamers in molecular recognition for cancer diagnosis, it is desirable to employ a set of aptamers to simultaneously analyze valuable tumor markers specific to a particular type of cancer. The whole-cell-based aptamer selection may be a promising new method for the rapid identification of nucleic acid ligands with high affinity for biomarkers on targeted tumor cells. It is reasonable to believe that the evolution of this new approach will eventually have a major impact on clinical cancer diagnosis.

***Acknowledgment*** We thank our colleagues at the University of Florida for their help with this chapter. This work was supported by NIH GM66137 and NSF NIRT EF-0304569.

## REFERENCES

1. Townsend, D. W., Carney, J. P., Yap, J. T., Hall, N. C. (2004). PET/CT today and tomorrow. *J. Nucl. Med.*, 45 (Suppl.) 1, 4S–14S.
2. Slamon, D. J., Godolphin, W., Jones, L. A., et al. (1998). Studies of the Her-2/Neu Proto-oncogene in human-breast and ovarian-cancer. *Science*, *244*(4905), 707–712.
3. Chan, J. M., Stampfer, M. J., Giovannucci, E., et al. (1998). Plasma insulin-like growth factor-I and prostate cancer risk: A prospective study. *Science*, *279*(5350), 563–566.

4. Osborne, C. K., Hamilton, B., Titus, G., Livingston, R. B. (1980). Epidermal growth factor stimulation of human breast cancer cells in culture. *Cancer Res.*, *40*(7), 2361–2366.
5. Ariad, S., Seymour, L., Bezwoda, W. R. (1991). Platelet-derived growth factor (PDGF) in plasma of breast cancer patients: Correlation with stage and rate of progression. *Breast Cancer Res. Treat.*, *20*(1), 11–17.
6. Tuerk, C., Gold, L. (1990). Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science*, *249*(4968), 505–510.
7. Ellington, A. D., Szostak, J. W. (1990). In vitro selection of RNA molecules that bind specific ligands. *Nature*, *346*(6287), 818–822.
8. Jayasena, S. D. (1999). Aptamers: An emerging class of molecules that rival antibodies in diagnostics. *Clin. Chem.*, *45*(9), 1628–1650.
9. Gold, L. (1995). Oligonucleotides as research, diagnostic, and therapeutic agents. *J. Biol. Chem.*, *270*(23), 1,3581–1,3584.
10. Xu, W., Ellington, A. D. (1996). Anti-peptide aptamers recognize amino acid sequence and bind a protein epitope. *Proc. Natl. Acad. Sci. USA*, *93*(15), 7475–7480.
11. Green, L. S., Jellinek, D., Jenison, R., et al. (1996). Inhibitory DNA ligands to platelet-derived growth factor B-chain. *Biochemistry*, *35*(45), 14413–14424.
12. Fang, X., Sen, A., Vicens, M., Tan, W. (2003). Synthetic DNA aptamers to detect protein molecular variants in a high-throughput fluorescence quenching assay. *ChemBioChem.*, *4*(9), 829–834.
13. Li, J. J., Fang, X., Tan, W. (2002). Molecular aptamer beacons for real-time protein recognition. *Biochem. Biophys. Res. Commun.*, *292*(1), 31–40.
14. Hamaguchi, N., Ellington, A., Stanton, M. (2001). Aptamer beacons for the direct detection of proteins. *Anal. Biochem.*, *294*(2), 126–131.
15. Nutiu, R., Li, Y. (2003). Structure-switching signaling aptamers. *J. Am. Chem. Soc.*, *125*(16), 4771–4778.
16. Hartig, J. S., Najafi-Shoushtari, S. H., Grune, I., et al. (2002). Protein-dependent ribozymes report molecular interactions in real time. *Nat. Biotechnol.*, *20*(7), 717–722.
17. Fang, X., Cao, Z., Beck, T., Tan, W. (2001). Molecular aptamer for real-time oncoprotein platelet-derived growth factor monitoring by fluorescence anisotropy. *Anal. Chem.*, *73*(23), 5752–5757.
18. Cox, J. C., Ellington, A. D. (2001). Automated selection of anti-protein aptamers. *Bioorg. Med. Chem.*, *9*(10), 2525–2531.
19. Petricoin, E. F., Ardekani, A. M., Hitt, B. A. et al. (2002). Use of proteomic patterns in serum to identify ovarian cancer. *Lancet*, *359*(9306), 572–577.
20. Petricoin, E. F., Ornstein, D. K., Paweletz, C. P. et al. (2002). Serum proteomic patterns for detection of prostate cancer. *J. Nat. Cancer Inst.*, *94*(20), 1576–1578.
21. Diamandis, E. P. (2003). Point—Proteomic patterns in biological fluids: Do they represent the future of cancer diagnostics? *Clin. Chem.*, *49*(8), 1272–1275.
22. Vant-Hull, B., Payano-Baez, A., Davis, R. H., Gold, L. (1998). The mathematics of SELEX against complex targets. *J. Mol. Biol.*, *278*(3), 579–597.

23. Morris, K. N., Jensen, K. B., Julin, C. M., Weil, M., Gold, L. (1998). High affinity ligands from in vitro selection: Complex targets. *Proc. Natl. Acad. Sci. USA*, *95*(6), 2902–2907.
24. Blank, M., Weinschenk, T., Priemer, M., Schluesener, H. (2001). Systematic evolution of a DNA aptamer binding to rat brain tumor microvessels. Selective targeting of endothelial regulatory protein pigpen. *J. Biol. Chem.*, *276*(19), 1,6464–1,6468.
25. Daniels, D. A., Chen, H., Hicke, B. J., Swiderek, K. M., Gold, L. (2003). A tenascin-C aptamer identified by tumor cell SELEX: Systematic evolution of ligands by exponential enrichment. *Proc. Natl. Acad. Sci. USA*, *100*(26), 1,5416–1,5421.
26. Hicke, B. J., Marion, C., Chang, Y. F., et al. (2001). Tenascin-C aptamers are generated using tumor cells and purified protein. *J. Biol. Chem.*, *276*(52), 4,8644–4,8654.
27. Wang, C., Zhang, M., Yang, G., et al. (2003). Single-stranded DNA aptamers that bind differentiated but not parental cells: Subtractive systematic evolution of ligands by exponential enrichment. *J. Biotechnol.*, *102*(1), 15–22.
28. Cerchia, L., Hamm, J., Libri, D., Tavitian, B., de F, V. (2002). Nucleic acid aptamers in cancer medicine. *FEBS Lett.*, *528*(1–3), 12–16.

# MOLECULAR SIMILARITY METHODS AND QSAR MODELS AS TOOLS FOR VIRTUAL SCREENING

JÜRGEN BAJORATH
*B-IT, International Center for Information Technology*
*Rheinische Friedrich–Wilhelms–University Bonn*
*Bonn, Germany*

1 INTRODUCTION
Virtual Versus High-Throughput Screening
Compound Sources
Ligand- Versus Target-Based Virtual Screening
2 SELECTED SCIENTIFIC CONCEPTS FOR VIRTUAL SCREENING
Molecular descriptors and chemical spaces
Molecular similarity
Quantitative Structure–Activity Relationship Analysis
3 GENERAL LIMITATIONS OF VIRTUAL SCREENING ANALYSIS
Similar Property Principle Revisited
Quantifying Molecular Similarity
Potency and Selectivity Issues
4 SPECIFIC METHODS
Two-Dimensional Substructure Searching
Topological and Shape Representations
Three-Dimensional Pharmacophores
Molecular Fingerprints
Two-Dimensional Versus Three-Dimensional Methods
Clustering Techniques
Partitioning Methods
QSAR Models in Virtual Screening
Drug Likeness

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

5 VIRTUAL SCREENING PERFORMANCE
6 CONCLUSIONS AND OUTLOOK
References

## 1 INTRODUCTION

In pharmaceutical research, computational (in silico, virtual) screening of compound databases has become an almost independent discipline [1, 2], although it is frequently discussed in the context of chemoinformatics [3, 4]. Its basic idea is to computationally screen two- or three-dimensional (2D or 3D) representations of compounds in large databases for molecules having desired biological activity. Similar to biological high-throughput screening (HTS), virtual screening (VS) is basically employed as a hit identification tool [1, 5] and cannot be expected to produce "magic bullets" [1] through some form of "computational magic." As will be discussed in this chapter, the algorithmic basis for many current VS methods is rather stringent, and compound selection criteria are derived in a scientifically rigorous manner.

The increasing attraction of VS approaches in recent years is in part due to the fact that significant computational power has become available cheaply and that many large-scale calculations can be readily carried out on desktop computing environments. Spurred on by such technological advances, a wealth of methods has been and continues to be developed for VS applications. Another reason for the attractiveness of VS methodologies is due to the fact that compound source databases have been dramatically growing in size over the past years. It is no longer uncommon for pharmaceutical compound inventories to contain a million compounds or more, and—based on our own experience—it is easily possible in today's chemistry market to collect 4 to 5 million compounds from vendor catalogs or other sources. The availability of so many compounds puts substantial constraints on hit identification efforts and more or less requires adding computational components to compound selection and screening repertoires. It is, therefore, not too surprising that there appears to be steadily increasing interest in VS approaches in the pharmaceutical industry, even outside "computer-heavy" environments. This chapter covers some major types of contemporary VS methods, in particular, those that are based on the analysis of molecular similarity and quantitative structure–activity relationships (QSAR). Prior to the introduction of specific concepts and methods, some more general aspects of VS and its relation to biological screening are discussed.

## Virtual Versus High-Throughput Screening

Despite many technological advances that continue to be made in the HTS arena toward miniaturization and increasingly higher throughput rates [6], HTS campaigns represent a costly affair for the pharmaceutical industry and often yield fairly limited success [7]. It is true that the costs per HTS experiment have been substantially reduced over the years in many cases, for example, simple enzyme assays that are nowadays often referred to as "penny screens." However, for more complex enzyme-linked immunosorbent assay (ELISA) or cell-based assays, reagent costs alone can easily amount to 50 cents or a dollar per compound well. Thus, a single-dose HTS screen of a compound deck containing 500,000 to a million molecules requires rather significant resources and does not even take into account required secondary assays or other follow-up studies. Thus, it is evident that "screening all compounds on all targets" is generally cost prohibitive. Therefore, preselection of compound sets by VS having a reasonable probability to act against a given target becomes attractive. Moreover, for time-consuming and expensive assays, for example, those depending on long-term culture of specific cell lines, computational analysis might initially replace HTS altogether, and small sets of compounds, perhaps tens to hundreds, are selected for testing in several rounds in order to obtain hit information with a minimum number of experiments.

In this context, it is often misunderstood that VS and HTS are in essence not competing but rather complementary disciplines. Of course, VS might sometimes replace HTS campaigns and vice versa, and virtual and high-throughput screeners may often develop preferences for rather different strategies in drug discovery (e.g., rationalization of compound selection versus increasingly higher throughput). However, this does not take away from the principal—and scientifically attractive—complementarity of VS and HTS, which is emphasized, for example, by the recent introduction of focused or sequential screening schemes [5, 8], as further discussed later on. Thus, as is the case for many research activities in drug discovery, a meaningful integration of experimental and computational screening approaches might often substantially increase the probability of ultimate success [9].

## Compound Sources

Another aspect of VS that is often not well understood and might require some clarification is the source of compounds that are evaluated in silico. Are these compounds always virtual (and thus perhaps of limited synthetic feasibility)? Certainly, compounds are stored on the computer in a virtual format but—more often than not—these compounds are "real." Essentially, three sources of compounds can be distinguished for use in VS calculations. Compound inventories in the pharmaceutical industry are nowadays mostly cataloged and administered in computational compound management systems that provide more or less immediate access for VS. Thus, in these cases,

compounds that have been synthesized and are physically stored and available for testing are a source for both HTS and VS. Virtual access to screening decks also provides an excellent basis for the integration of in silico and biological screening [5], at least if advanced compound handling and plating systems (e.g., for cherry picking of compounds) are available. Compound libraries assembled from public domain sources or vendor catalogs are another common source for VS. Although synthetic routes have been established for these compounds and they should, in principle, be available for acquisition, experience shows that, on average, often only about 50 percent of such public domain compounds can be readily acquired. Thus, for candidate compounds selected from such libraries, resynthesis is often required (making these compound sources "pseudoreal"). Large virtual combinatorial libraries currently represent the only truly theoretical compound source for VS [10]. Here combinatorial libraries are enumerated in silico or stored as reagents or molecular fragments for VS access. The underlying idea is to computationally select sets of compounds from vast chemical spaces as candidates for synthesis, dependent on the particular search problem [10]. In some cases, such virtual source libraries can be very large, literally containing billions of possible combinatorial compounds. The design and storage of libraries of this size would hardly be possible without almost unlimited hardware resources that have become available in recent years.

## Ligand- Versus Target-Based Virtual Screening

Virtual screening techniques can principally be divided into target- and ligand-based approaches. The former critically depend on the availability of accurate three-dimensional structures of therapeutic targets, mostly proteins, and are often referred to as docking (of putative ligands into binding sites) [11, 12]. Structure-based VS attempts to identify active compounds de novo by focusing only on target structure information. By contrast, ligand-based VS depends on the availability of known active molecules, for example, enzyme inhibitors, receptor agonists, or antagonists, that are then used as templates for VS. Here the primary goal is the identification of molecules that exhibit biological activity similar to the templates but belong to different structural classes [1, 5]. This is often necessary to provide a starting point for a new lead optimization project focusing on molecules that are synthetically more accessible than the templates or have lower toxicity. In addition, such "lead hopping" exercises are often also carried out in order to circumvent patent positions of competitors in drug discovery. In practice, target- and ligand-based VS are not applied in a mutually exclusive manner. Whenever target structures and active compounds are available, these methods are used in concert in order to increase the probability of identifying novel hits. Also, the availability of experimental structures of target–ligand complexes makes it possible to closely combine these efforts, for example, by studying candidate compounds in the structural context of template molecules bound to receptor sites.

## 2 SELECTED SCIENTIFIC CONCEPTS FOR VIRTUAL SCREENING

In this section, major scientific concepts are introduced that provide a basis for the development of VS tools. In addition, some general limitations and caveats of VS analysis are also discussed in this section.

### Molecular Descriptors and Chemical Spaces

Essentially all ligand-based VS calculations depend on the use of computationally implemented descriptors of molecular structure, physico-chemcial properties, or pharmacophore features. This is the case because the comparison of molecular characteristics beyond simple structural criteria requires the calculation of property values based on physical or mathematical models. Accordingly, thousands of different molecular descriptors have been reported over the years that often greatly vary in their degree of complexity and sophistication [13, 14]. Table 1 lists a few classes of molecular descriptors that are widely used for compound classification, QSAR, and VS applications. Descriptors are frequently divided into 1D, 2D, or 3D descriptors, dependent on the dimensionality of the molecular representation from which they are calculated [14]. According to this scheme, 1D descriptors are calculated from molecular composition formulas and include bulk properties such as molecular weight. Two-dimensional descriptors include, for example, molecular connectivity indices, graphs, or charge descriptors. Finally, 3D descriptors

**TABLE 1 Desciptor Categories**[a]

| Descriptor Category | Examples |
|---|---|
| Physical properties | Molecular weight $\log P$(o/w) |
| Atom and bond counts | Number of nitrogen atoms |
| | Number of aromatic atoms |
| | Number of rotatable bonds |
| Pharmacophore features | Number of hydrogen bond acceptors |
| | Total van der Waal surface areas of basic atoms |
| Charge descriptors | Total positive partial charge |
| | Dipole moment from partial charges |
| Connectivity and shape descriptors | Kier and Hall molecular shape indices |
| Surface area and volume | Solvent-accessible surface area |
| | Van der Waals volume |

[a] Some type of widely used molecular descriptors are shown and representative examples are provided. It should be noted that categories often overlap and that many descriptors can be assigned to more than one class.

capture properties or structural features that are molecular conformation-dependent such as dipole moment, solvent-accessible surface, or specific pharmacophore arrangements. Furthermore, contributions of different types of descriptors can be combined into rather complex formulations, for example, descriptors combining molecular topology and electronic atom properties [15], descriptors that map features such as hydrophobic or polar character onto molecular surface representations [16], or composite descriptors that are calculated as linear combinations of many single descriptors and their specific contributions [17].

Molecular descriptors are selected for the definition of chemical spaces into which molecules are projected based on their calculated descriptor values and where qualitative or quantitative comparisons are carried out. Following this idea, $n$ selected descriptors form the axes of an $n$-dimensional chemical reference space, and each test molecule is assigned a vector that determines its position in this space, as illustrated in Figure 1. Although underlying scientific concepts and details of many VS methods differ, molecular descriptors are fundamentally important for their application, and it is often critical to identify sets of descriptors that are suitable or preferred for a specific VS application. Automated descriptor selection has been facilitated by the application of machine learning techniques such as genetic algorithms [17, 18] or neural nets [19] or by identifying descriptors having highest information content in a compound collection under study [20].

## Molecular Similarity

The introduction of the concept of molecular similarity has greatly influenced research in chemoinformatics and VS [21, 22]. The computational assessment

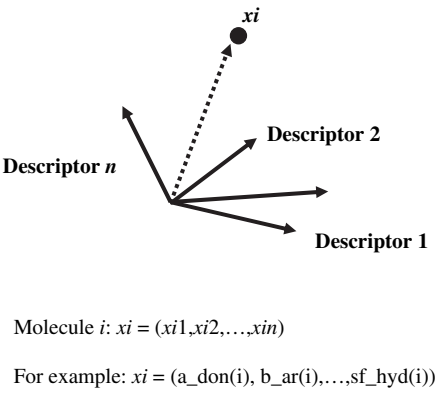


**Figure 1** Vector representation of a molecule in $n$-dimensional descriptor space. In this example, the position of test molecule $i$ is determined by its values of $n$ molecular descriptors used to define the chemical space (e.g., a_don, number of donor atoms; b_ar, number of aromatic bonds; sf_hyd, hydrophobic surface area).

of molecular similarity goes well beyond structural criteria or molecular topology and takes into account diverse molecular properties and, importantly, biological activity. The *similar property principle* states that similar molecules should have similar activity [21]. Despite its generality and exceptions (see below), it captures the essence of molecular similarity analysis and its relevance for ligand-based VS: If one can apply computationally derived models to identify molecular properties that are responsible for specific activity, beyond what one can "see," then it should also be possible, by application of such models, to identify molecules having similar activity. Figure 2 provides a few examples.

In descriptor spaces, molecular similarity can be assessed in qualitative terms (e.g., which molecules map to the same section in space?) or quantitative terms (what is the exact distance between these molecules?). Similarly,

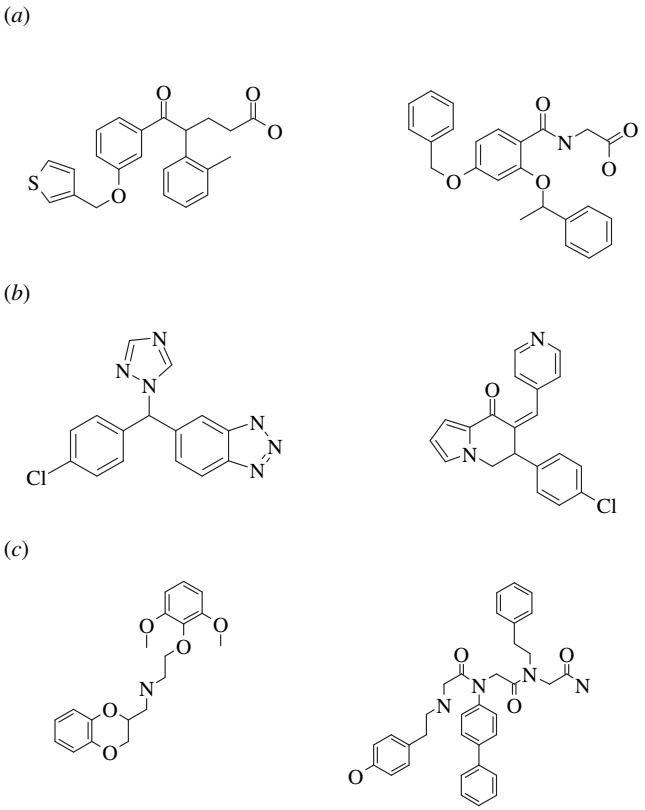


**Figure 2** Remote similarity relationships that were successfully recognized using in-house developed similarity methods: (*a*) endothelin A antagonists, (*b*) aromatase inhibitors, (*c*) α1-adrenoreceptor ligands).

abstract molecular representations can be generated by transforming descriptor settings into bit strings of varying length, and the resulting bit patterns of these so-called fingerprints of molecules (described below in more detail) can be quantitatively compared to assess their degree of similarity. Regardless of the algorithms used, the quantitative evaluation of molecular similarity depends on mathematical measures [22, 23] to calculate the similarity of fingerprint-type representations or distances in chemical reference spaces. Some of the most popular similarity metrics are reported in Table 2.

It should also be noted that the concept of molecular diversity is related to but distinct from molecular similarity, although methods and metrics used for similarity and diversity analysis in part overlap. Major focal points of diversity analysis have included the design of chemically diverse (combinatorial) libraries [24, 25] and the selection of optimally diverse or representative subsets from large compound sources [25, 26]. In diversity analysis, it is typically attempted to evenly populate chemical reference spaces with compounds and design molecules that fulfill these requirements and are thus sufficiently dissimilar from each other. Creating chemical diversity in library design often involves the estimation of diversity distributions from the properties of either reagents or products of chemical reactions [25]. Recently, the design of chemically diverse libraries for screening has become less popular than it used to be in the mid- to late 1990s, and there is a clear trend to generate smaller compound libraries that are focused on therapeutic target classes or specific activities, which then often involves similarity comparison of candidate compounds and known active molecules [27].

**TABLE 2 Similarity Coefficients and Distance Functions**[a]

| | |
|---|---|
| Tanimoto coefficient | $\mathrm{Tc} = n_{ij}/(n_i + n_j - n_{ij})$ |
| Dice coefficient | $\mathrm{Dc} = n_{ij}/(n_i + n_j)/2$ |
| Cosine coefficient | $\mathrm{Cc} = n_{ij}/(n_i n_j)^{1/2}$ |
| Hamming distance | $\mathrm{HD}_{ij} = (n_i + n_j - 2n_{ij})$ |
| Euclidean distance | $\mathrm{ED}_{ij} = (n_i + n_j - 2n_{ij})^{1/2}$ |
| Average distance | $D = \dfrac{\sum_{i=1}^{n}\sum_{j=1}^{n} D_{ij}}{n(n-1)}$ |

[a] Shown are similarity metrics and distance functions that are often used for bit string comparisons. In these formulations, $n_i$ and $n_j$ are the number of bits set on for molecules $i$ and $j$, respectively, and $n_{ij}$ is the number of bits in common to both molecules. $D_{ij}$ is the distance between molecules $i$ and $j$, D the average distance, and $n$ the total number of molecules.

## Quantitative Structure–Activity Relationship Analysis

In QSAR, the relationship between biological activity of a compound and its molecular properties is explored, and these properties are by necessity expressed or modeled using molecular descriptors. Standard QSAR analysis attempts to express biological activity as a linear combination of different descriptors, thus postulating the presence of a linear relationship between activity and relevant molecular properties [28]. Coefficients in these equations assign weights to descriptors according to the calculated importance of their contributions for predicting activity. Typically, a series of analogs with known activity is used as a learning set for model building. Resulting QSAR models are evaluated using cross-validation techniques and well-performing models are used to predict structural modifications that might increase potency. Thus, different from molecular similarity evaluation, standard QSAR analysis is usually limited to congeneric compound sets. In 2D-QSAR, model building is based on 2D representation of molecules and 2D descriptors. In 3D-QSAR, bioactive conformations of test compounds are predicted and conformation-dependent steric and electrostatic fields of these compounds are generated and probed on grid representations with pseudoatoms for favorable interaction sites [29]. Of course, such 3D descriptors can also be used in combination with 1D or 2D descriptors to derive QSAR models. The quality of 3D-QSAR models critically depends, first and foremost, on the correct prediction of bioactive conformations and also on correct alignment or superposition of test compounds (when comparing their fields). In order to address the uncertainties of correctly predicting bioactive conformations, the 4D-QSAR formalism was introduced [30, 31]. Here, rather than deriving models based on a single putative conformation, a conformational ensemble of test molecules is calculated for 3D-QSAR, which in essence adds conformational space as a fourth dimension to the modeling process. The 4D-QSAR approach does not alleviate the alignment problem, and superpositions of test molecules are usually generated and optimized in an automated fashion [32]. Higher dimensional QSAR models, with additional degrees of freedom, can be derived by, for example, implicit or explicit inclusion of binding site elements in the modeling process [33].

Like in molecular similarity analysis, an important question in QSAR is how to find the "best" descriptors for solving a specific structure–activity relationship problem. Experience and intuition are often used in combination with trial-and-error calculations, but QSAR modeling can also be coupled to machine learning techniques such as genetic algorithms or neural nets in order to identify preferred descriptor combinations. Furthermore, it is often required to reduce the number of descriptors and identify those that make the most important contributions to the predictions. Initial 3D-QSAR models typically involve a large number of descriptors and reducing their numbers increases the interpretability of the models and decreases their complexity, which often

renders them less susceptible to calculation errors due to overfitting of data or overlearning (as these effects are often caused by redundancy of descriptor contributions). The elimination of input variables corresponds to a reduction in dimensionality of descriptor reference spaces for establishing QSAR equations. For linear structure–activity relationships, partial least–squares regression is often used, and other descriptor space dimension reduction techniques, as described later on, can also be applied. Neural network methods are increasingly employed in 3D-QSAR, not only for descriptor selection but also for dimension reduction and model simplification. For example, Kohonen self-organizing neural networks or maps, an unsupervised learning technique, project high-dimensional (descriptor) space representations onto 2D maps [34]. In addition to descriptor reduction through self-organizing maps, another attractive feature of neural network methods in 3D-QSAR is that these methods are designed to model nonlinear relationships [35], which presents a departure from the classical QSAR theme. The major drawback of applying neural network methods is their "black box" character, that is, the inability to extract actual models from neural network simulations and interpret them in physical or chemical terms.

## 3 GENERAL LIMITATIONS OF VIRTUAL SCREENING ANALYSIS

At this stage of the discussion, one should also comment on principal limitations of ligand-based VS that more or less apply regardless of the methods used. General limitations of QSAR analysis (e.g., correct prediction of bioactive conformations, alignment problem, overfitting of data and overtraining, linear versus nonlinear structure–activity relationships) also affect QSAR models that are adapted for VS. Therefore, limitations different from QSAR-specific ones will be discussed here, in particular, those that relate to molecular similarity analysis.

### Similar Property Principle Revisited

The postulate that similar molecules should have similar activity is intuitive, although it is, of course, also known that structurally similar molecules do not necessarily need to display equivalent biological activity [36]. The identification of remote similarity relationships (i.e., exploring structural and property differences that are tolerated within a specific activity class) is a major focal point of molecular similarity analysis in pharmaceutical research. In sharp contrast to the basic ideas underlying this approach is a fact well known by medicinal chemists, namely that a single chemical change can render a molecule within a series of active analogs almost or completely inactive. Figure 3 shows an example to further illustrate this point. The possibility of these alternative scenarios, dependent on the type of the search problem under investigation,

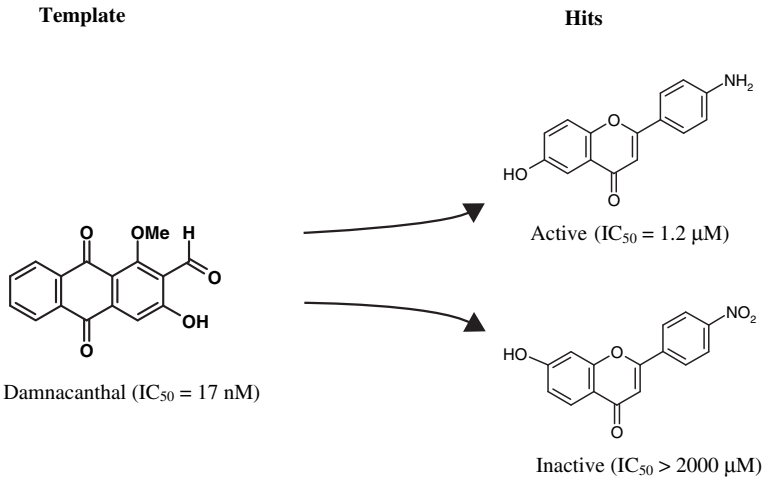


**Figure 3** Similarity paradox—a caveat for virtual screening. Results of a similarity search are shown using damnacanthal (a potent inhibitor of tyrosine kinase $p56^{lck}$) as a template. The search identified a number of molecules sharing the same core structure. However, minor modifications of this scaffold (typical for analogs) render compounds either active or inactive. Two examples are shown. Virtual screening calculations should identify these analogs as similar to damnacanthal. However, if by chance only the inactive analog would have been selected, VS analysis would have apparently failed, although a relevant scaffold was indeed identified. Adapted from [1].

has substantial implication for VS analysis. For example, VS calculations are expected to recognize molecules as similar that share the same core structure. However, if only inactive analogs were selected for testing, VS analysis would have been considered a failure, although a relevant structural class was successfully identified. This conundrum has been termed the *similarity paradox* [1] and also led to the computational study of *latent hits* [37], that is, inactive compounds that can be transformed into active ones by very simple chemical modifications. For VS, this illustrates a potential danger associated with selection of only one or a few representative molecules from a series of compounds recognized as being similar to search templates. Compound selection strategies that generally address these difficulties are yet to be developed, and decisions are usually made on a case-by-case basis, dependent on the specifics of the search problem and the obtained results.

## Quantifying Molecular Similarity

As mentioned above, one of the hallmarks of computational molecular similarity analysis is the ability to quantify similarity relationships, for example, by using metrics shown in Table 2. Similarity is expressed as a numerical value and relationships above a given threshold value are considered similar.

However, this convenient way of measuring similarity has some caveats. As an example, the Tanimoto coefficient (Tc), also shown in Table 2, is perhaps the most popular metric for quantitative comparison of bit string-type descriptors or fingerprints. This coefficient relates the number of bits that are set on in two-bit string representations to each other and quantifies fingerprint overlap. If the strings do not share any bits that are set on, the Tc value is 0 (no similarity) and if the bit settings are identical, Tc is 1 (maximum similarity). The calculation of Tc values for the assessment of molecular similarity has some inherent statistical limitations [38, 39]. More importantly, however, it is a priori not clear which threshold values correspond to "true" similarity. The comparison of studies using different fingerprint-type descriptors and molecular data sets illustrates this problematic issue. Patterson et al. evaluated different types of descriptors in the study of molecular similarity and established the concept of *neighborhood behavior* [40]. A major conclusion was that 85 percent of test compounds that displayed a Tc value of 0.85 in a fingerprint comparison to an active molecule were themselves active [40]. Similar findings were subsequently reported by others [41], and a Tc threshold value of 0.85 has in many instances become a guideline for the detection of biologically relevant similarity. However, in an analysis of various screening data sets, Martin et al. found that at a fingerprint similarity level of 0.85 only about 30 percent of detected compounds were active [42]. These findings are well in accord with observations that different similarity search methods frequently produce different results, dependent on the characteristics of activity classes that are studied [43]. Taken together, these findings demonstrate that results of similarity search calculations (and other VS methods) are generally influenced by the search tools applied, the data sets used, and the compound classes studied. Consequently, it is difficult to define generally applicable similarity threshold values or rules that unambiguously relate calculated and biological similarities to each other, which presents a potential caveat for VS analysis. Therefore, in some cases, Tc threshold values have been optimized for specific fingerprints and compound classes [4].

## Potency and Selectivity Issues

Another limitation of VS has to do with the fact that one of its major goals is the detection of remote similarity relationships, that is, finding different structural classes having similar activity. For these calculations, optimized molecules are usually used as starting points that are typically taken from the patent literature or publications and have potency in the nanomolar range and high selectivity. It is often also attempted to find novel chemotypes as potential backup compounds for clinical candidates. Thus, template molecules are typically the result of a thorough lead optimization process. If VS calculations successfully identify a novel chemotype displaying similar activity, these molecules depart from optimized structural motifs and are therefore usually much less active. Simply put, successful VS analysis is expected to produce hits, not

leads, which provide the starting point for a new medicinal chemistry program and optimization effort [1]. It is fairly typical for VS to produce hits with potency in the low micromolar range and moderate selectivity when highly potent templates are used [1]. This is sometimes not well understood and can lead to unrealistic expectations from VS technology.

## 4 SPECIFIC METHODS

In the following, different VS methods are discussed that have varying degrees of complexity. The first group of methods focuses on similarity searching, which involves quantitative assessment of molecular similarity. This is followed by clustering and partitioning techniques that can either be qualitative or quantitative in nature. Finally, various QSAR models for VS are discussed.

### Two-Dimensional Substructure Searching

Perhaps the simplest form of VS, as we understand it today, is the search for matching substructures in compound databases. Early algorithms for substructural analysis did not utilize molecular descriptors but were based on graph representations where atoms are represented as nodes and bonds as edges [44]. In substructure analysis, comparison of molecular graphs aims at detecting subgraph isomorphism by comparing and matching node positions and edges [44]. This is a computationally expensive procedure and not amenable to screening of very large compound databases. Therefore, substructure search methods used as screening tools today mostly utilize dictionaries of molecular fragment-type descriptors or keys for comparisons and detect the presence or absence of such keys in database molecules [45]. For example, the publicly available 166 MACCS keys from MDL [46] are a popular set of structural fragment-type descriptors. Substructure searching can be directly linked to the analysis of biological activity, for example, by detecting recurrent fragments in active molecules and calculating the statistical significance of this frequency of occurrence relative to their distribution in other database molecules. This type of statistical substructure and property analysis has been implemented in the LeadScope program, which utilizes a dictionary containing more than 20,000 fragments [47]. For VS applications, a major limitation of substructure analysis is the difficulty of finding novel chemotypes, simply because only the presence or absence of predefined core structures can be detected. Thus, scaffold transitions can hardly be accomplished. By contrast, substructure methods are very effective when searching for analogs of active compounds or structurally related molecules.

### Topological and Shape Representations

A number of specialized complex molecular descriptors have been designed that can be used as queries for similarity searching. For example, electro-

topological indices combine molecular topology and electronic properties by capturing the valence state of each atom in a molecule and its electronic structure, taking into account perturbations caused by neighboring atoms [15, 48]. Each atom in a molecule is assigned a numerical E-state value and atom type. The atom types represented in a molecule can be used as a combined query and searched against atom types calculated for database molecules. Molecular comparisons are then carried out by calculating the Euclidian distance between atom type indices in template and test molecules [48]. By contrast, topomer shape descriptors are derived by fragmenting molecules in a defined way, assigning a rule-based conformation to each fragment, and calculating a molecular field for each fragment [49]. Thus, these descriptors capture whole-molecule shape information as a sum of fragment contributions. Molecular shape comparisons are carried out by calculating field differences for all possible fragment-by-fragment comparisons in two molecules. This fragment-based shape-matching algorithm has been shown to be applicable to similarity searching and used to screen large virtual libraries for active molecules [49].

## Three-Dimensional Pharmacophores

Pharmacophores are generally defined as spatial arrangements of atoms or groups in molecules that are responsible for specific interactions with enzymes or receptors and the resulting biological activity. Similar to 2D substructures, pharmacophores can in some cases be derived from 2D molecular representations, but usually pharmacophore models are three dimensional and used for 3D database searching. Pharmacophore analysis has a long history in pharmaceutical research [50]. Early pharmacophore methods captured bond lengths and angles between specified atoms [51] or encoded shape information from atom sphere models [52]. Currently, pharmacophores are mostly represented as molecular (pseudo-) graphs where nodes correspond to pharmacophore points (atoms, groups) and edges to interpoint distances. Different features can be mapped to each point (e.g., hydrogen bond acceptors or aromatic groups) and interpoint distances can be binned in order to generate distance ranges. This structure makes it possible to differentiate similar pharmacophore arrangements from each other, rather than obtaining only yes/no answers when trying to match query and database pharmacophores. Conventionally used 3-point pharmacophores have a few years ago been extended to 4-point pharmacophore models [53]. Figure 4 shows a comparison

**Figure 4** Multiple-point pharmacophore models. The figure illustrates how pharmacophore models are derived from a known or predicted active conformation of a molecule. Atoms, groups, or features are represented as a graph from which three- and four-point pharmacophores are extracted. A four-point pharmacophore consist of four points and six distances (thereby defining a spatial arrangement). In a pharmacophore search, each point could represent several potential features and each distance might be divided into different ranges.

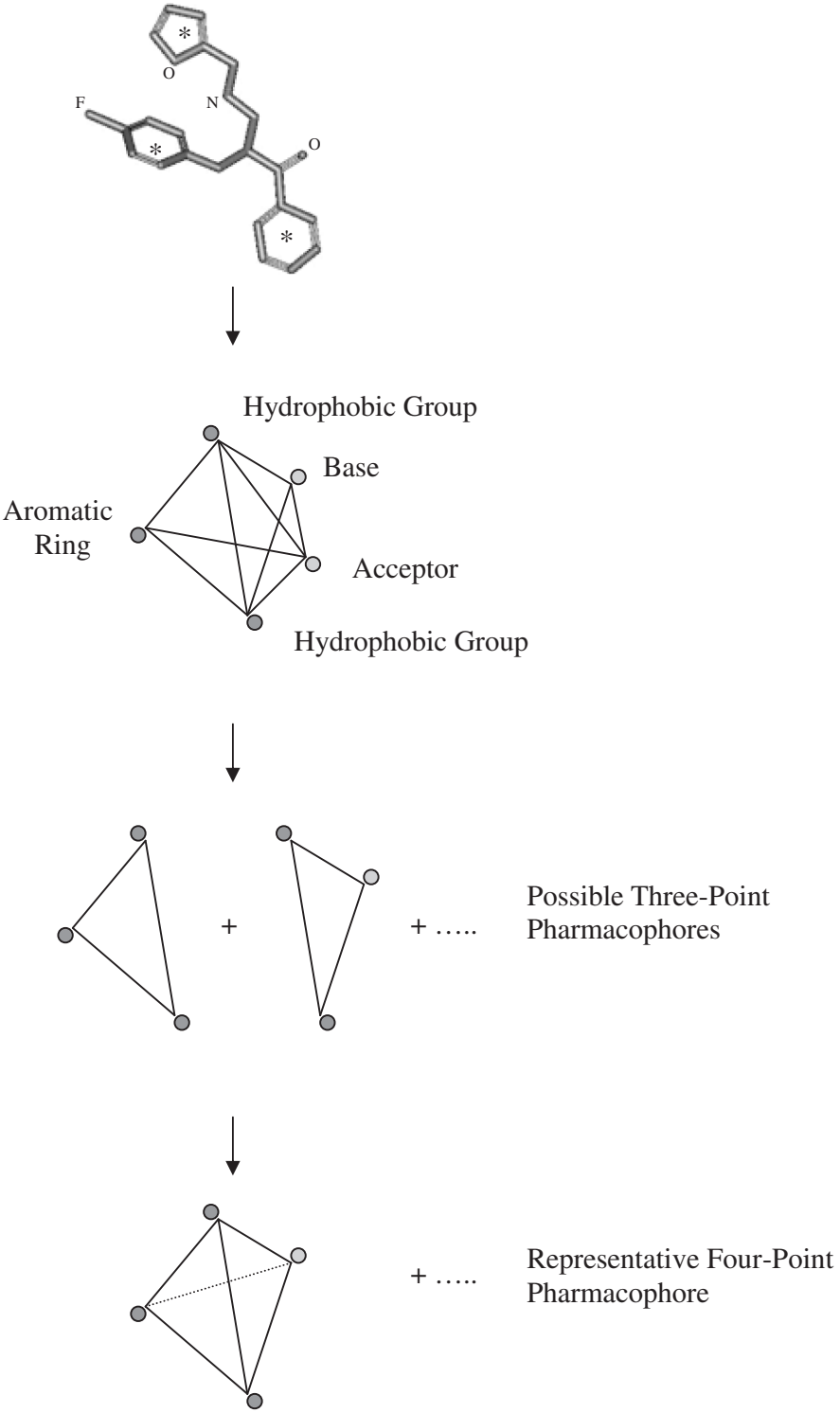
F
O
N
O
Hydrophobic Group
Base
Aromatic Ring
Acceptor
Hydrophobic Group
+
+ …..
Possible Three-Point Pharmacophores
+ …..
Representative Four-Point Pharmacophore

of such 3- and 4-point pharmacophores. The introduction of 4-point pharmacophores has much increased the resolution of pharmacophore searching. For example, given a case where 6 features can be mapped to each point and each distance is divided into 10 ranges, a 3-point model could capture 33,000 potential pharmacophore arrangements whereas a 4-point pharmacophore would be able to monitor about 13 million.

Given the dramatic increase in computational performance in recent years, 3D pharmacophore searching can be readily applied to VS. Database searching for single pharmacophores has two major caveats. First, pharmacophore models typically represent a (nonconfirmed) hypothesis, as the pharmacophores are derived from the analysis of analogs, QSAR modeling, or inspection of crystal structures. Even if X-ray structures of active molecules in complex with their receptors are available, it is often not at all certain that attractive interactions that can be "seen" are indeed determinants of biological activity. Second, similar to the situation in conventional QSAR, it must generally be assumed that computed 3D structures of database molecules are representative of their bioactive conformations, which is probably the largest source of errors in pharmacophore searching. These difficulties have been addressed by the design of pharmacophore fingerprints, as discussed in the following.

## Molecular Fingerprints

Bit string representations of various molecular descriptors are among the most popular methods for similarity searching. As already mentioned, such fingerprints are used in combination with various metrics (the Tanimoto coefficient being the most popular one) to quantify the degree of similarity between template and test molecules. Therefore, fingerprints need to be precalculated for database compounds and then compared to the corresponding fingerprints of query molecules. For VS analysis, fingerprints are attractive tools because bit string comparisons are very fast. The time-limiting step is usually the calculation of fingerprints of database compounds, which is required only once. It is important to note that the design and complexity of fingerprint representations often significantly differs, as further discussed below.

***Two-Dimensional Fingerprints*** In simple fingerprints, each bit position accounts for the presence (if the bit is set on) or absence of a specific feature in a molecule. MACCS keys present an instructive example for this type of fingerprint [46]. If used in a bit string format (rather than as single substructure search queries), MACCS keys consist of a string of 166 bits, each of which accounts for the presence or absence of a specific structural key. BCI fingerprints (often consisting of 1024 bits) are also generated from different families of structural fragments [54], conceptually similar to MACCS keys. Any fragment dictionary-based approach has the intrinsic limitation that only cat-

aloged fragments can be detected in test molecules, which may preclude the recognition of unique structural motifs. However, this limitation can sometimes be overcome by increasingly generalizing molecular fragment types, thus incorporating some degree of fuzziness in the analysis [54].

So-called minifingerprints are other short bit string representations, consisting of about 60 to 200 bits, that combine structural keys and molecular property descriptors specifically selected for their ability to correctly classify compounds according to their biological activity [55]. In these designs, value ranges of property descriptors (e.g., the number of hydrogen bond acceptors in a molecule) are incrementally encoded (e.g., no acceptors, no bit is set on; one or two acceptors, one bit is set on; three or four descriptors, two bits are set on, etc.). A special type of minifingerprint involves a binary transformation of molecular property descriptors based on statistical medians of descriptor value distributions in large compound databases [56]. This means that a bit position reserved for a particular descriptor is set to one if its value is larger than or equal to the median and set to zero if it is smaller. This novel coding scheme has made it possible to encode many property descriptors in a short bit string format. Although each descriptor contribution is monitored at a low level of resolution, simultaneous contributions of many different types of descriptors make these fingerprints very effective in similarity searching [56].

Daylight fingerprints represent a different class of 2D search tools [57]. Here all possible connectivity pathways in a molecule through up to seven bonds are recorded and mapped to a bit string format. Different from the simple designs discussed above (where each position is associated with a specific descriptor or value range), a hashing algorithm is applied to map connectivity pathways to overlapping bit segments. Daylight fingerprints are variable in length, dependent on hashing density, and popular forms consist of 1024 or 2048 bit positions.

***Three-Dimensional Pharmacophore Fingerprints*** In pharmacophore fingerprints, each bit position typically monitors the presence or absence of one possible arrangement of a predefined three- or four-point pharmacophore in a test molecule, and these fingerprints therefore often consist of millions of bits [58–60]. Accordingly, they represent VS tools that are quite different from relatively simple 2D fingerprints. Possible pharmacophore arrangements are calculated based on fast systematic conformational search of test compounds. Pharmacophore fingerprinting elegantly circumvents some limitations of single pharmacophore searching. These fingerprints do not rely on accurate prediction of bioactive conformations and monitor possible spatial arrangements of functional groups at a high level of resolution. The underlying hypothesis is that increasing overlap of possible pharmacophore patterns in template and database compounds correlates with an increasing probability that these molecules share similar activity. Although pharmacophore fingerprinting is computationally demanding, it has become a much applied

technique, given the large computational resources that are widely available by now.

***Fingerprint Profiling and Scaling*** Methods have been introduced to tune fingerprint search calculations toward recognition of specific compound classes. When fingerprints are calculated and averaged for compounds belonging to diverse activity classes, compound class-specific bit patterns can often be observed [61]. Figure 5 shows examples of such fingerprint profiles. Dependent on the particular compound class, some bit positions may always, or almost always, be set on, thus defining consensus bit settings for this class. The application of scaling factors to such consensus bits emphasizes fingerprint profile differences during similarity search calculations and has been shown to lead to a significant increase in search performance [61], as reported in Table 3. Profiling and scaling can be applied to any 2D or 3D fingerprint representation where each bit monitors a specific feature but is much more difficult to apply to hashed or folded bit strings.

## Two-Dimensional Versus Three-Dimensional Methods

At this stage of the discussion, it might be appropriate to comment on the relative performance of 2D and 3D descriptors and methods, a question that continues to be much debated in the literature. In principle, 3D methodologies should outperform 2D approaches, simply because molecules are active in three dimensions. However, difficulties involved in accurately calculating some conformation-dependent 3D properties and predicting bioactive compound conformations often reduce the accuracy of 3D methods to an extent that they are outperformed by simpler (and less error-prone) 2D approaches [4]. Moreover, 2D methods might also have some intrinsic advantages. For example, structural key-type descriptors implicitly encode much chemical information that might otherwise be difficult to explicitly calculate, which may—at least in part—explain their good performance in many applications [62]. Different studies have provided support for the superiority of either 3D [40, 59] or 2D [62, 63] methods. In some cases, successful recognition of remote similarity

**Figure 5** Fingerprint profiles. Shown are representative profiles for two different biological activity classes calculated with a model fingerprint consisting of 61 bit positions (horizontal axis). The vertical axis records the frequency of bit occurrence within an activity class (i.e., 1 means the bit is always set on and 0 means it is never set on). Bits that are mostly or always set on define consensus positions for an activity class. The profiles of different activity classes show some notable differences that can be exploited in similarity searching by applying scale factors to class-specific consensus positions. Scaling increases the probability of identifying similar molecules.

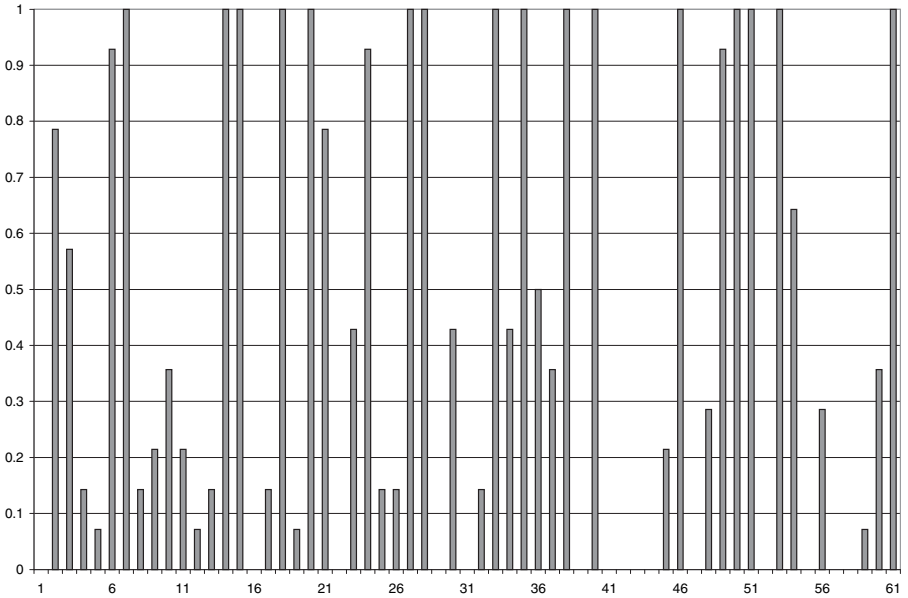
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
1
6
11
16
21
26
31
36
41
46
51
56
61

(*a*)

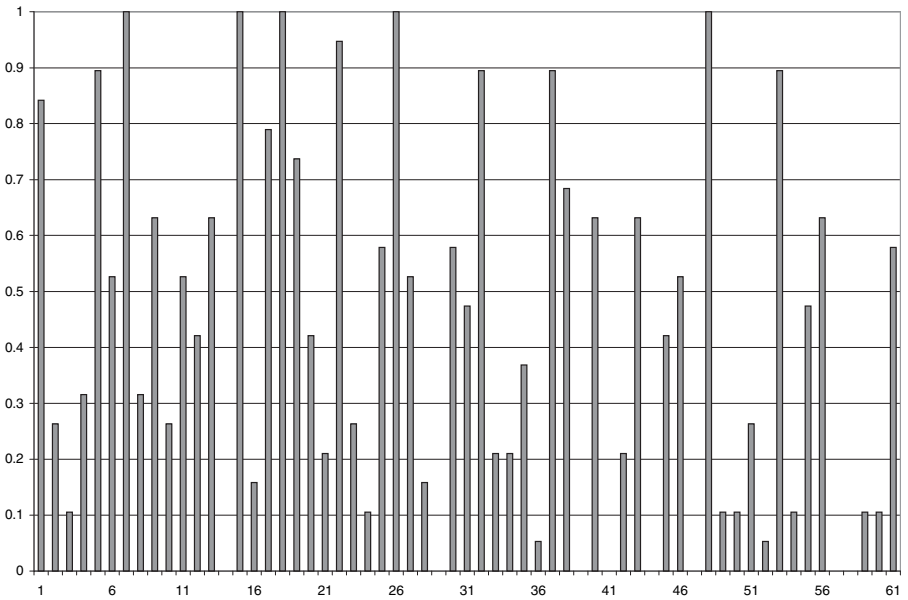
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
1
6
11
16
21
26
31
36
41
46
51
56
61

(*b*)

**TABLE 3 Fingerprint Search Calculations and Scaling**[a]

| Fingerprint | Scaling | Tc | Percent Right | Percent Wrong |
|---|---|---|---|---|
| MACCS | Yes | 0.94 | 47.9 | 0.030 |
| | No | 0.78 | 35.3 | 0.015 |
| SE-MFP | Yes | 0.96 | 46.5 | 0.012 |
| | No | 0.81 | 38.3 | 0.016 |
| MP-MFP | Yes | 0.90 | 58.7 | 0.018 |
| | No | 0.74 | 43.1 | 0.018 |

[a] The table summarizes the results of systematic similarity search calculations in a database containing 175 compounds belonging to 23 diverse activity classes and 5000 randomly selected synthetic compounds. Three fingerprints were tested, MACCS keys and two in-house designed fingerprints, SE-MFP and MP-MFP (described in [55] and [56], respectively). Overall best Tc similarity threshold values are also reported. Under scaling conditions, a scaling factor of 10 was uniformly applied to consensus positions with a bit frequency of at least 95 percent in the fingerprint profile of each activity class (see also Fig. 5). Here Percent Right reports the percentage of correctly identified active compounds averaged over all activity classes and Percent Wrong the average percentage of false positives (including both active and/or background compounds). As can be seen, fingerprint scaling leads to a significant increase in search performance, with MP-MFP producing best results.
Data were taken from [56].

relationships was thought to strictly depend on the application of 3D approaches [60], but these findings could nevertheless be well reproduced using much simpler 2D methods [64]. Simply put, it is probably not possible to claim that either 2D or 3D methods are generally superior since much depends on the specifics of each test case and application [4]. What can be said, however, is that 2D descriptors and methods perform surprisingly well in the analysis of molecular similarity and in many VS situations.

## Clustering Techniques

Cluster analysis of compounds in chemical reference spaces has been applied for more than 20 years and continues to play an important role in pharmaceutical research [65]. The major goal of clustering is to group together sets of compound with similar structures and/or chemical features and distinguish them from others. Such compound classification techniques can assess molecular similarity in a qualitative or quantitative way. Fingerprint calculations can be based on single template molecules, whereas clustering or partitioning techniques become attractive whenever series of known active molecules are available. In a typical VS analysis, one adds these active template molecules to a compound database, finds descriptor solutions and clustering conditions that group active molecules together, and selects other compounds that occur in these clusters (and are thus considered similar) as candidates for testing. In the following, different clustering algorithms will be briefly discussed.

***Hierarchical Versus Nonhierarchical Clustering*** Clustering methods can be divided into two major categories, hierarchical and nonhierarchical approaches [65]. Nonhierarchical clustering groups molecules together into a defined number of clusters, without considering intercluster relationships. Typically, nearest neighbors are calculated for each test molecule and two compounds are included in the same cluster if they share a prespecified minimum number of nearest neighbors. Hierarchical clustering, on the other hand, establishes relationships between initially generated clusters and is further divided into divisive or agglomerative approaches. In hierarchical-divisive clustering, the starting point is one large cluster that is subdivided based on distance calculations and similarity relationships (top-down approach). By contrast, in hierarchical-agglomerative clustering, one begins with singletons that are combined into increasingly larger clusters (bottom-up).

For both hierarchical and nonhierarchical approaches, the optimum clustering level and preferred input parameters are difficult to determine upfront and, therefore, cluster-level selection algorithms must usually be applied to determine conditions giving best results [66]. Such algorithms work, for example, by identifying clustering levels that balance the number of clusters with their average population or density. This can be accomplished by comparing intracluster distances between molecules in all clusters as well as distances between cluster centers. For chemistry applications, the most popular hierarchical and nonhierarchical methods have generally been Ward's (agglomerative) [67] and Jarvis–Patrick [68] clustering, respectively. For compound classification, Ward's clustering has been identified as a method of choice in a number of cases [62].

***Multidomain Clustering*** For compound classification, hierarchical-agglomerative clustering with level selection has been implemented, in addition to self-organizing maps, in a decision-tree-like algorithm combining a variety of simulation techniques [69]. Here the final clustering step of compound sets, typically HTS data, is carried out using structural keys as descriptors, and for acceptable clusters the largest common substructure is determined. Clusters can be scored by calculating the average activity of the compounds they contain. If carried out in an iterative fashion, the algorithm makes it possible to identify small sets of compounds that have common substructures associated with biological activity. Moreover, molecules containing different substructures that modulate activity, so-called multidomain compounds, can appear in different clusters, thus revealing multiple structure–activity relationships [69]. Although this phylogenetic-like tree method was originally developed to process HTS data sets and extract structure–activity relationships from them, it can also be applied to VS by adding sets of known active compounds to databases and analyzing the content of clusters into which these molecules fall.

## Partitioning Methods

In recent years, partitioning algorithms have become increasingly popular for compound classification, analysis of HTS results, and VS applications [70]. One of the major reasons for this popularity, among others, is the computational efficiency of many partitioning techniques. Like clustering, partitioning operates in chemical descriptor spaces. However, regardless of the algorithm, cluster analysis always involves the calculation and comparison of pairwise distances between compounds in reference spaces. Computationally, this is not problematic for small to medium-size compound data sets, but, when databases are large (containing millions of molecules), these pairwise distance calculations become more and more difficult and at some point prohibitive. By contrast, partitioning algorithms typically calculate subdivisions or sections of chemical space into which compounds fall, based on their descriptor coordinates or signature patterns. Thus, partitioning does not require pairwise distance comparisons and can therefore be applied to much larger compound databases than distance-based clustering.

Different classes of partitioning methods have been developed that are amenable to VS. Just as in cluster analysis, known active compounds are added to a source database that is then partitioned, and database compounds that occur in partitions with active molecules are considered as candidates for testing. Different from cluster analysis, partitioning does not produce intercompound distances as a possible quantitative measure of molecular similarity (although such distances can be calculated retroactively). Thus, partitioning is usually more qualitative in nature.

***Cell-Based Partitioning*** In recent years, cell-based partitioning has become one of the most popular analysis techniques in computational medicinal and combinatorial chemistry [70]. Here partitions or cells are generated to evenly segment chemical space by applying binning schemes to each descriptor axis. The dimensionality of the descriptor space and the number of bins per axis ultimately determine the number of cells, which is a major determinant of the outcome of compound classification calculations. Cell-based partitioning is generally carried out in low-dimensional descriptor spaces (e.g., six dimensions) for several reasons. In low-dimensional reference spaces with orthogonal axes, it is possible to limit space distortions due to descriptor correlation effects, avoid large numbers of unpopulated cells, and control cell population density. In addition, low-dimensional spaces can often be further reduced to three dimensions without loosing too much information, which makes it possible to visualize compound distributions, a rather important requirement for many chemical applications. Low-dimensional reference spaces can be constructed by using especially designed complex descriptors that combine different types of contributions (e.g., molecular connectivity and hydrogen bond potential) in an uncorrelated manner [71]. Alternatively, it might be required to reduce the dimensionality of originally defined high-dimensional

descriptor spaces. Dimension reduction can be accomplished by calculating a small number of orthogonal descriptors axes from the original descriptors, for example, as linear combinations of these descriptors [17, 72], or by applying nonlinear functions that extract the most significant (descriptor-dependent) relationships between compounds and project compound distributions into lower-dimensional spaces [73]. Significant relationships or features would include, for example, detected similarity between compounds that share comparable activity.

***Direct Space Methods*** Recent studies have shown that dimension reduction is not necessarily required for effective partitioning if original descriptor spaces are simplified. For example, by use of statistical medians of descriptor value distributions, as mentioned earlier, molecular property descriptors can be transformed into a binary format. Therefore, the median partitioning algorithm was developed that uses one property descriptor per partitioning step to iteratively divide a compound database into subpopulations above and below the median value of this descriptor [74]. According to this scheme, the choice of $n$ descriptors and partitioning steps results in a total of $2^n$ unique partitions, each of which is characterized by a binary partition code. Thus, median partitioning operates in multidimensional, albeit simplified, descriptor spaces but does not involve dimension reduction. Essentially, every molecule is assigned a vector in multidimensional descriptor space with the length along each dimension being either zero or one (unity). As summarized in Table 4,

**TABLE 4 Compound Classification by Cell-Based and Direct Space Partitioning**[a]

| Method | Dim | Bins | %P | P | nP | S | M | nM |
|---|---|---|---|---|---|---|---|---|
| Cell-based partitioning | 6 | 7 | 55.2 | 57 | 175 | 82 | 19 | 60 |
| Median partitioning | 19 | n/a | 63.1 | 69 | 200 | 86 | 22 | 31 |

[a] For these calculations, 317 compounds belonging to 21 different activity classes were added to 2000 randomly selected synthetic molecules. This database was then partitioned using different methods. Column 2 (Dim) reports the dimensionality of the descriptor reference space and column 3 (Bins) the number of intervals on each axis for cell-based partitioning. In direct space partitioning, no binning scheme is applied. As a cell-based partitioning method, partitioning based on principal component analysis was used, as described in [72]. Principal component analysis is one of several dimension reduction methods. Column 4 (%P) reports percentage of active compounds that occur in so-called pure partitions. These partitions exclusively contain compounds sharing the same activity. In activity-based compound classification, this is the desired partitioning result. Column 5 (P) gives the total number of pure partitions, column 6 (nP) the total number of compounds in pure partitions, column 7 (S) the number of singletons, column 8 (M) the number of mixed partitions, and column 9 (nM) the total number of compounds in mixed partitions. Mixed partitions contain molecules having different activities or active and randomly selected molecules (classification errors). As can be seen, median partitioning in a 19-dimensional descriptor space compares favorably with cell-based partitioning in a 6-dimensional space when classifying these compounds according to their biological activity.
Data were taken from [74].

median partitioning can perform as well as or even better than cell-based partitioning in the classification of compounds according to diverse biological activities [74]. In its adaptation for VS of very large compound databases, median partitioning is carried out in a recursive manner [75]. With the aid of machine learning techniques, descriptor combinations are identified that successfully co-partition known active molecules added as "baits" to the database. At each recursion level, database compounds are pooled that occur in the same partitions as successfully co-partitioned baits, other database compounds are discarded, descriptor selection is reinitialized, and the process is repeated until only a small number of database compound remains (as potential candidates for testing). During the course of this analysis, these candidates have consistently been recognized as "similar" to bait molecules because they have always been found in active partitions.

Another recently developed direct space method, termed *dynamic mapping of consensus positions*, combines elements of binary descriptor transformation, partitioning, and bit string methods [76]. Starting from a large pool of binary-transformed descriptors, initially, the maximum number of descriptors is determined that have identical bit settings for all compounds belonging to the studied activity class. This defines a descriptor space where all compounds map exactly to the same (consensus) position. Then, in order to eliminate an increasing number of database compounds and ultimately retain only those that are most similar to the baits, the dimensionality (and resolution) of the descriptor space is increased in subsequent steps. This is accomplished by defining consensus positions that no longer require identical descriptor settings for all test compounds but permit one or more known actives to deviate in one or more descriptor settings from the others. This stepwise dimension extension eliminates most database compounds from activity class-specific consensus positions [76]. Like median partitioning, this approach makes use of simplified descriptor spaces, but it does no longer depend on machine learning techniques for descriptor selection. Moreover, it operates not only in multidimensional descriptor spaces but utilizes dimension extension as an approach to identify compounds that are similar to known actives. A unique feature of this mapping algorithm is that it generates compound class-specific descriptor reference spaces, which distinguishes it from many other classification methods. As a VS tool, dynamic mapping of consensus positions is very fast, permitting the rapid processing of hundreds of descriptors and millions of database compounds. Importantly, the method has achieved promising hit rates in large-scale VS analysis, as reported in Table 5.

***Decision Tree Algorithms*** Another type of statistical partitioning method utilizes decision tree structures as the primary classification tool. The perhaps most popular decision tree method for the analysis of HTS data and VS is recursive partitioning [77]. Initially, learning sets with known active and inactive compounds are processed along a tree structure using two-state structural or topological descriptors. At every branch of the tree, descriptors are used to

**TABLE 5 Virtual Screening by Recursive Median Partitioning and Dynamic Mapping of Consensus Positions**[a]

| | DMC | | RMP | |
|---|---|---|---|---|
| Activity Class | HR (%) | RR (%) | HR (%) | RR (%) |
| I | 26 | 9 | 12 | 18 |
| I | 50 | 18 | 3 | 11 |
| III | 74 | 29 | 3 | 5 |
| IV | 9 | 23 | 21 | 22 |
| V | 26 | 9 | 13 | 40 |

[a] Recursive median partitioning is abbreviated RMP and dynamic mapping of consensus positions is abbreviated DMC. Calculations were carried out on five different activity classes in a database consisting of ~1.3 million compounds. These are: I, benzodiazepines; II, cyclooxygenase-2 inhibitors; III, histamine H3 antagonists; IV, serotonin receptor ligands; V, tyrosine kinase inhibitors. In these VS calculations, between 21 and 61 active compounds were "hidden" in the database (active database compounds) as potential hits, and 10 bait molecules were used in each case. Hit rates (HR) were calculated by dividing the number of correctly identified active database compounds (hits) by the total number of selected database molecules, and recovery rates (RR) by dividing the number of hits by the total number of active database compounds. For both RMP and DMC calculations, hit and recovery rates were significant but DMC produced higher hit rates for four of the five activity classes that were studied.
Data were taken from [76].

successively divide the data set into compound subsets that have or do not have the tested features. If the partitioning analysis is successful, subsets that are significantly enriched with active compounds are obtained in terminal nodes that associate active molecules with specific descriptor settings and pathways. These activity-related descriptor combinations can then be used as predictive models to recursively partition large compound databases and search for novel active molecules [77]. Computationally, recursive partitioning is perhaps the most efficient partitioning method and millions of compounds and very large numbers of descriptors can be rapidly processed. The method has some known limitations. Different from multidomain clustering, as discussed above, active molecules can usually only occur in one terminal node. Furthermore, molecules with similar activity yet significant structural differences (remote similarity relationships) might be separated during the partitioning process and occur in different terminal nodes. In addition, the method is prone to high false-positive rates when predicting active compounds [78]. Therefore, several extensions of the recursive partitioning algorithm have recently been introduced that address some of these issues. For example, the use of multiple trees, so-called recursion forests, permits the application of consensus scoring schemes, which have been shown to produce significant hit rates in VS, while reducing false-positive rates [78]. In addition, tree structures have been developed for recursive partitioning that allow the simultaneous analysis and prediction of more than one property as a function of descriptor set-

tings, for example, specificity in addition to activity [79]. This would make it possible, for example, to process a compound library that has been screened against multiple targets using a single tree structure and address selectivity questions.

## QSAR Models in Virtual Screening

Despite its long-standing history in medicinal chemistry, QSAR techniques have only in recent years been adapted for VS. This might at least in part be due to the fact that QSAR analysis is predominantly applied during lead optimization, rather than hit identification. In essence, however, applying a QSAR model to screen a large compound library does not differ much from using it to analyze or predict a series of analogs. On the other hand, it is also clear that QSAR models are in principle better suited to search for analogs of known leads than for novel chemotypes, which might limit their application in VS. In the context of VS, QSAR overlaps with pharmacophore searching, as many pharmacophore models are based on QSAR, and the boundaries between these approaches are somewhat fluid. Recently, different types of QSAR models have been directly applied to VS.

***Two-Dimensional QSAR Models*** Due to the computational efficiency of 2D descriptor-based models, it is in principle straightforward to use 2D-QSAR models to virtually screen large compound databases. However, not many 2D-QSAR VS studies have been reported thus far. In a recent application, a number of 2D-QSAR models for cyclooxygenase-2 inhibitors were developed using different methodologies (including, e.g., partial least-squares linear regression as well as supervised and unsupervised neural networks) and applied in combination to develop a consensus prediction theme [80]. In this case, one searches for molecules that are consistently favored by several models. These QSAR models and consensus predictions were then applied in a preliminary screen of the database of the National Cancer Institute (~200,000 compounds) and several potentially interesting molecules were found [80].

***Three-Dimensional QSAR Virtual Screens*** Although the derivation of 3D-QSAR models is more complex and time consuming than 2D-QSAR modeling, there are more VS applications of 3D-QSAR, which is likely due to the fact that 3D-QSAR is more closely related to 3D pharmacophore searching. For example, in the discovery of novel VLA-4 antagonists, a 3D pharmacophore was constructed based on known crystallographic and structure–activity data and used to screen a combinatorial library. Compounds identified in this search were then subjected to 3D-QSAR analysis in order to better understand the different activity levels of identified hits [81]. Similarly, a pharmacophore search was combined with 3D-QSAR (comparative molecular field) analysis to identify hits and improve their potency [82]. In this case, the analysis went a step further and computed analogs were also studied by

docking calculations, thus putting high emphasis on the exploitation of structural data for both ligands and receptor [82].

***Four-Dimensional QSAR*** Although computationally more expensive, a multidimensional QSAR formalism was also adapted for in silico screening by including conformational ensembles (of glycogen phopsphorylase inhibitors) in 3D-QSAR analysis [83]. In this study, it was well recognized that screening (or design) of target-focused combinatorial libraries should be a prime application for QSAR-based VS since these libraries should already contain preferred structural motifs. Moreover, the analysis revealed that overfitting of 4D-QSAR models to training data—although usually known to represent a problem for QSAR predictions—might sometimes be beneficial for their application in VS simply because it favors the identification of library compounds that are likely to sample (and probe) molecular regions not well represented in training set molecules. Such findings might substantially aid in the design of second-generation focused libraries.

***Probabilistic Methods*** QSAR-like approaches have also been developed for large-scale property predictions. Binary QSAR uses combinations of molecular descriptors to build relationships between compounds and their activity or other properties of interest [84]. The approach is based on Bayes's theorem, a mathematical formulation determining the probability that a result is due to a specific cause among multiple plausible causes. Binary QSAR makes use of learning sets to construct probability density functions that estimate properties of molecules from their descriptor values. Compared to standard QSAR approaches, this methodology is more qualitative in nature because it ultimately develops predictive models based on a binary formulation of activity that can be applied to large molecular test sets. Each test molecule is assigned a probability to be active and, based on a chosen threshold value, classified as either active or inactive. Similar to recursive partitioning, binary QSAR can be applied to efficiently process large HTS data sets and, using so-derived models, rapidly screen databases for molecules having a high probability to be active. An attractive feature of this probabilistic approach is that the derivation of predictive models does not depend on the availability of highly accurate data sets for learning. Thus, it can usually be applied to process raw HTS data. That binary QSAR calculations are not limited to activity predictions, but are more widely applicable, has been demonstrated, for example, by studies to systematically distinguish between naturally occurring and synthetic molecules [85].

Another conceptually similar but algorithmically distinct probabilistic machine learning technique that is experiencing increasing interest in the context of compound classification and VS is binary kernel discrimination [86]. Similar to binary QSAR, this methodology uses learning sets to calculate the probability of a molecule to be active (by applying kernel functions). However, the approach is currently more constrained in the choice of molec-

ular descriptors and usually limited to two-state descriptors, similar to recursive partitioning.

### Drug Likeness

Better understanding and predicting druglike or leadlike features of candidate compounds during the early phases of drug discovery has become a current focal point in computational medicinal chemistry and also influenced the development of VS methods. Neural network simulations have been extensively applied to build computational models that distinguish drugs from nondrugs [19]. However, as mentioned earlier, such models typically lack interpretability and, consequently, it has also been attempted to build simpler and more intuitive ones. These include decision tree models [87] and, in their simplest form, dictionaries of preferred and nonpreferred functional groups or rules that prioritize desired structural motifs and physico-chemical compound characteristics [88]. For VS, rule-based prediction schemes of druglikeness have become rather attractive and are widely used because they can be added as simple filter functions to screening calculations [5]. In addition, various computational models have been developed for ADME-related properties that can be carried out as virtual screens or added to VS methods. These include, for example, blood–brain barrier penetration [89] or oral availability, for which QSAR models reaching 70 percent prediction accuracy were developed based on a learning set consisting of known drugs [90].

## 5 VIRTUAL SCREENING PERFORMANCE

Having addressed general limitations of VS, one should also discuss how successful in silico screening has been thus far. Like any computational model, VS tools are generally rich in approximations and might frequently fail for technical reasons. Furthermore, as already mentioned earlier, different similarity search methods often produce different results [43], as the calculations are much influenced by specific features of studied compound classes. However, without doubt VS methods work and help to discover novel active molecules. Based on experience gathered over the past few years, novel hits with activity in the micromolar range are identified in about 75 percent of the VS projects carried out in our laboratory. In some cases, in-house developed VS methods, as described herein, were capable of identifying hits in situation where HTS had failed or was not feasible (e.g., because cell-based assays were too complex or time consuming to be formatted for HTS). In two other cases (structurally relatively simple) natural products with specific biological activity were chemically not accessible but could be replaced by *synthetic mimics* based on VS calculations, thus permitting the generation of analogs.

While obtaining desired VS results is far from being a routine procedure, a number of successful case studies have been reported in the literature for

target structure- and ligand-based VS as well as for combinations of different 2D and 3D methodologies [5, 10]. Case studies have been discussed in this chapter in the context of specific methods. Some of these studies—but also others—are capable of illustrating the potential and current performance levels of VS in more general terms. For example, in a recent investigation, HTS and structure-based VS were carried out in parallel and successfully identified diverse inhibitors of tyrosine phosphatase-1B with activities in the low- to mid-micromolar range [91]. In this study, docking calculations ultimately produced significantly higher hit rates than random screening (by up to three orders of magnitude). In ligand-based VS, a combination of 3D pharmacophore search and 3D-QSAR analysis identified VLA-4 antagonists with activity in the sub-micromolar range [81], a potency level that is still an exception rather than the rule for many VS applications. In order to make use of all target- and ligand-related information that has become available, different types of VS methodologies are often combined, which further increases the probability of success. For example, when combining similarity searching, docking, and binary QSAR calculations, five new antagonists of the BCL-xL/Bak interaction were identified with potency in the low-micromolar range (and after testing fewer than 50 candidate molecules) [92]. When new hits were subsequently used to aid in the design of a target-focused library, seven additional active compounds were obtained [92]. Similarly, docking calculations were recently combined with 2D filtering methods and 3D pharmacophore searching in a study that discovered several new inhibitors of carbonic anhydrase [93].

As discussed in detail in this chapter, compound classification techniques (in particular, partitioning) have become increasingly attractive as similarity-based approaches for VS, due to their computational efficiency and success rates. For example, results of benchmark calculations reported for different partitioning methods illustrate this point. For a set of monoamine oxidase inhibitors, recursive partitioning produced a 15-fold improvement in hit rate over random selection [77]. Furthermore, in large-scale VS calculations, recursive median partitioning achieved hit rates between 3 and 21 percent for five different compound classes [75]. Even higher hit rates between 9 and 74 percent were obtained in a parallel study on these classes using dynamic mapping of consensus positions [76], a method conceptually related to median partitioning, as discussed (see Table 5). Although clustering and partitioning methods display significant promise as stand-alone VS tools, these techniques become particularly attractive when used in close combination with biological screening for iterative preselection of compound sets, as already indicated in the introductory sections. This process is often called sequential screening [8]. In these situations, VS has been shown in various independent investigations to increase hit rates by one to two orders of magnitude [5, 8], and this performance level can currently be considered as a standard in the field. For example, in a study targeting a number of cell-based anticancer assays, two rounds of sequential screening produced hit rates of up to 40% [5]. Moreover, an analysis of a number of different two-stage sequential screening investiga-

tion showed that testing of 10 to 20 percent of compounds in screening libraries was generally sufficient to identify 50 to 80 percent of potential hits [5]. These findings suggest that VS analysis often significantly reduces the number of compounds that need to be tested until novel molecules of interest are identified. It is important to note that achieving such efficiency levels does not require dramatic improvements in hit rates. For example, analysis of a large HTS data set (for an undisclosed target) has shown that up to fivefold increase in hit rates could be achieved by application of recursive partitioning [94] and that only ~20 percent of available database compounds needed to be screened in order to identify ~75 percent of potential hits [94]. These findings are well in accord with the numbers reported for two-stage sequential screens, as mentioned above, and have significant practical implications for organizing HTS campaigns.

Clearly, despite their documented promise, VS methods should not only be evaluated based on their performance as (stand-alone) hit identification tools but also based on their ability to complement experimental programs and increase their efficiency.

## 6 CONCLUSIONS AND OUTLOOK

Many molecular similarity methods have matured over the years and substantially contributed to making VS an important component of pharmaceutical research. In addition, novel methodologies continue to be developed at a fast pace to further increase the accuracy and computational efficiency of molecular similarity analysis and VS. Despite its predominant application in lead optimization, QSAR techniques have recently been adapted for VS. Moreover, QSAR-type probabilistic methods have been specifically designed for screening applications. Multidimensional QSAR modeling alleviates problems of accurately predicting bioactive conformations of ligands and will likely be more widely applied in the context of VS. It is also anticipated that ligand- and target-based design techniques will be more closely combined in the furture, as already exemplified by receptor-dependent QSAR modeling and methods for structure-based design of pharmacophores and 3D fingerprints. As far as experimental applications are concerned, currently available data suggest that the integration of VS and HTS is particularly promising, despite some practical constraints that still exist (such as limited availability of sufficiently flexible compound plating and management systems). However, there is little doubt that VS and HTS are highly complementary disciplines.

## REFERENCES

1. Bajorath, J. (2002). Virtual screening: Methods, expectations, and reality. *Curr. Drug Discov., 2*(3), 24–28.

2. Walters, W. P., Stahl, M. T., Murcko, M. A. (1998). Virtual screening—An overview. *Drug Discov. Today, 3*, 160–178.
3. Bajorath, J. (2001). Rational drug discovery revisited: Interfacing experimental programs with bio- and chemo-informatics. *Drug Discov. Today, 6*, 989–995.
4. Bajorath, J. (2001). Selected concepts and investigations in compound classification, molecular descriptor analysis, and virtual screening. *J. Chem. Inf. Comput. Sci., 41*, 233–245.
5. Bajorath, J. (2002). Integration of virtual and high-throughput screening. *Nat. Rev. Drug Discov., 1*, 882–894.
6. Fox, S., Farr-Jones, S., Yund, M. A. (1999). High-throughput screening for drug discovery: Continually transitioning into new technologies. *J. Biomol. Screen., 4*, 183–186.
7. Lahana, R. (1999). How many leads from HTS? *Drug Discov. Today, 4*, 447–448.
8. Engels, M. F. M., Venkatarangan, P. (2001). Smart screening: Approaches to efficient HTS. *Curr. Opin. Drug. Discov. Devel., 4*, 275–283.
9. Owens, J. (2003). Jürgen Bajorath discusses integrating disciplines in the drug discovery industry. *Targets, 2*, 45–47.
10. Green, D. V. (2003). Virtual screening of virtual libraries. *Prog. Med. Chem., 41*, 61–97.
11. Brooijmans, N., Kuntz, I. D. (2003). Molecular recognition and docking algorithms. *Annu. Rev. Biophys. Biomol. Struct., 32*, 335–373.
12. Abagyan, R., Totrov, M. (2001). High-throughput docking for lead generation. *Curr. Opin. Chem. Biol., 5*, 375–382.
13. Livingstone, D. J. (2000). The characterization of chemical structures using molecular properties. A survey. *J. Chem. Inf. Comput. Sci., 40*, 195–209.
14. Xue, L., Bajorath, J. (2000). Molecular descriptors in chemoinformatics, computational combinatorial chemistry, and virtual screening. *Combin. Chem. High Throughput Screen., 3*, 363–372.
15. Hall, L. H., Kier, L. B. (2000). The E-state as the basis for molecular structure space definition and structure similarity. *J. Chem. Inf. Comput. Sci., 40*, 784–791.
16. Labute, P. (2000). A widely applicable set of descriptors. *J. Mol. Graph. Model., 18*, 464–477.
17. Xue, L., Bajorath, J. (2000). Molecular descriptors for effective classification of biologically active compounds based on principal component analysis identified by a genetic algorithm. *J. Chem. Inf. Comput. Sci., 40*, 801–809.
18. Gillet, V. J., Willett, P., Bradshaw, J. (1998). Identification of biological activity profiles using substructural analysis and genetic algorithms. *J. Chem. Inf. Comput. Sci., 38*, 165–179.
19. Sadowski, J. (2000). Optimization of chemical libraries by neural network methods. *Curr. Opin. Chem. Biol., 4*, 280–282.
20. Godden, J. W., Bajorath, J. (2003). An information-theoretic approach to descriptor selection for database profiling and QSAR modeling. *QSAR Comb. Sci., 22*, 487–497.
21. Johnson, M. A., Maggiora, G. M. (Eds) (1990). *Concepts and Applications of Molecular Similarity.* Wiley, New York.

22. Maggiora, G. M., Shanmugasundaram, V. (2004). Molecular similarity measures. In J. Bajorath (Ed.), *Chemoinformatics: Concepts, Methods, and Tools for Drug Discovery*. Humana, Totowa, NJ.
23. Willett, P., Barnard, J. M., Downs, G. M. (1998). Chemical similarity searching. *J. Chem. Inf. Comput. Sci., 38*, 983–996.
24. Martin, E. J., Blaney, J. M., Siani, M. A., Spellmeyer, D. C, Wong, A. K., Moos, W. H. (1995). Measuring diversity: Experimental design of combinatorial libraries for drug discovery. *J. Med. Chem., 38*, 1431–1436.
25. Martin, Y. C. (2001). Diverse viewpoints on computational aspects of molecular diversity. *J. Comb. Chem., 3*, 231–250.
26. Willett, P. (1999). Dissimilarity-based algorithms for selecting structurally diverse sets of compounds. *J. Comput. Biol., 6*, 447–457.
27. Xue, L., Godden, J. W., Stahura, F. L., Bajorath, J. (2001). A dual fingerprint-based metric for the design of focused combinatorial libraries and analogs. *J. Mol. Model., 7*, 125–131.
28. Free, S. M. Jr., Wilson, J. W. (1964). A mathematical contribution to structure-activity studies. *J. Med. Chem., 7*, 395–399.
29. Cramer, R. D. III, Patterson, D. E., Bunce, J. D. (1988). Comparative molecular field analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins. *J. Am. Chem. Soc., 110*, 5959–5967.
30. Hopfinger, A. J. (1980). A QSAR investigation of dihydrofolate reductase inhibition by Baker triazines based upon molecular shape analysis. *J. Am. Chem. Soc., 102*, 7196–7206.
31. Hopfinger, A. J., Wang, S., Tokarski, J. S., Jin, B., Albuquerque, M., Madhav, P. J., Duraiswami, C. (1997). Construction of 3D-QSAR models using the 4D-QSAR analysis formalism. *J. Am. Chem. Soc., 119*, 10509–10524.
32. Kearsley, S. K., Smith, G. M. (1990). An alternative method for the alignment of molecular structures: Maximizing electrostatic and steric overlap. *Tetrahedron Comput. Meth., 3*, 615–633.
33. Vedani, A., Dobler, M. (2002). 5D-QSAR: The key for simulating induced fit? *J. Med. Chem., 45*, 2139–2149.
34. Kohonen, T. (1989). *Self-Organization and Associative Memory.* Springer, Berlin.
35. Tetko, I. V., Kovalishyn, V. V., Livingstone, D. J. (2001). Volume learning algorithm artificial neural networks for 3D-QSAR studies. *J. Med. Chem., 44*, 2411–2420.
36. Kubinyi, H. (1998). Similarity and dissimilarity—A medicinal chemist's view. *Perspect. Drug Discov. Design, 11*, 225–252.
37. Mestres, J., Veeneman, G. H. (2003). Identification of "latent hits" in compound screening collections. *J. Med. Chem., 46*, 3441–3444.
38. Flower, D. R. (1998). On the properties of bit-string measures of chemical similarity. *J. Chem. Inf. Comput. Sci., 38*, 379–386.
39. Godden, J. W., Xue, L., Bajorath, J. (2000). Combinatorial preferences affect molecular similarity/diversity calculations using binary fingerprints and Tanimoto coefficients. *J. Chem. Inf. Comput. Sci., 40*, 163–166.
40. Patterson, D. E., Cramer, R. D. III, Ferguson, A. M., Clark, R. D., Weinberger, L. E. (1996). Neighborhood behavior: A useful concept for validation of "molecular diversity" descriptors. *J. Med. Chem., 49*, 3049–3059.

41. Matter, H. (1997). Selecting optimally diverse compounds from structure databases: A validation study of two-dimensional and three-dimensional descriptors. *J. Med. Chem., 40*, 1219–1229.
42. Martin, Y. C., Kofron, J. L., Traphagen, L. M. (2002). Do structurally similar molecules have similar biological activity? *J. Med. Chem., 45*, 4350–4358.
43. Sheridan, R. P., Kearsley, S. K. (2002). Why do we need so many chemical similarity search methods? *Drug Discov. Today, 7*, 903–911.
44. Barnard, J. M. (1993). Substructure searching methods. Old and new. *J. Chem. Inf. Comput. Sci., 33*, 532–538.
45. Merlot, C., Domine, D., Cleva, C., Church, D. J. (2003). Chemical substructures in drug discovery. *Drug Discov. Today, 8*, 594–602.
46. MACCS keys, MDL Information Systems, San Leandro, CA.
47. Roberts, G., Myatt, G. J., Johnson, W. P., Cross, K. P., Blower, P. (2000). LeadScope: Software for exploring large sets of screening data. *J. Chem. Inf. Comput. Sci., 40*, 1302–1314.
48. Kier, L. B., Hall, L. H. (2001). Database organization and searching with E-state indices. *SAR QSAR Environ. Res., 12*, 55–74.
49. Cramer, R. D. III, Poss, M. A., Hermsmeier, M. A., Caulfield, T. J., Kowala, M. C., Valentine, M. T. (1999). Prospective identification of biologically active structures by topomer similarity searching. *J. Med. Chem., 42*, 3919–3933.
50. Martin, Y. C. (1992). 3D database searching in drug design. *J. Med. Chem., 35*, 2145–2154.
51. Murray-Rust, P., Motherwell, S. (1978). Computer retrieval and analysis of molecular geometry. 1. General principles and methods. *Acta Chryst., 34B*, 2518–2526.
52. Sheridan, P., Venkataragharvan, R. (1987). Designing novel nicotinic agonists by searching a database of molecular shapes. *J. Comput.-Aided Mol. Design, 1*, 243–256.
53. Mason, J. S., Morize, I., Menard, P. R., Cheney, D. L., Hulme, C., Labaudiniere, R. F. (1999). New 4-point pharmacophore method for molecular similarity and diversity applications: Overview over the method and applications, including a novel approach to the design of combinatorial libraries containing privileged substructures. *J. Med. Chem., 42*, 3251–3264.
54. Barnard, J. M., Downs, G. M. (1997). Chemical fragment generation and clustering software. *J. Chem. Inf. Comput. Sci., 37*, 141–142.
55. Xue, L., Godden, J. W., Bajorath, J. (2003). Mini-fingerprints for virtual screening: Design principles and generation of novel prototypes based on information theory. *SAR QSAR Environ. Res., 14*, 27–40.
56. Xue, L., Godden, J. W., Stahura, F. L., Bajorath, J. (2003). Design and evaluation of a molecular fingerprint involving the transformation of property descriptor values into a binary classification scheme. *J. Chem. Inf. Comput. Sci., 43*, 1151–1157.
57. James, C. A., Weininger, D. *Daylight Theory Manual.* Daylight Chemical Information Systems, Irvine, CA.
58. McGregor, M. J., Muskal, S. M. (1999). Pharmacophore fingerprinting. 1. Application to QSAR and focused library design. *J. Chem. Inf. Comput. Sci., 39*, 569–574.

59. Mason, J. S., Cheney, D. L. (2000). Library design and virtual screening using multiple point pharmacophore fingerprints. *Pac. Symp. Biocomput., 5*, 576–587.

60. Bradley, E. K., Beroza, P., Penzotti, J. E., Grootenhuis, P. D. J., Spellmeyer, D. C., Miller, J. L. (2000). A rapid computational method for lead evolution: Description and application to $\alpha_1$-adrenergic antagonists. *J. Med. Chem., 43*, 2770–2774.

61. Xue, L., Godden, J. W., Stahura, F. L., Bajorath, J. (2003). Profile scaling increases the similarity search performance of molecular fingerprints containing numerical descriptors and structural keys. *J. Chem. Inf. Comput. Sci., 43*, 1218–1225.

62. Brown, R. D., Martin, Y. C. (1996). Use of structure-activity data to compare structure-based clustering methods and descriptors for use in compound selection. *J. Chem. Inf. Comput. Sci., 36*, 572–584.

63. Matter, H., Pötter, T. (1999). Comparing 3D pharmacophore triplets and 2D fingerprints for selecting diverse compound subsets. *J. Chem. Inf. Comput. Sci., 39*, 1211–1225.

64. Xue, L., Stahura, F. L., Godden, J. W., Bajorath, J. (2001). Mini-fingerprints detect similar activity of receptor ligands previously recognized only by three-dimensional pharmacophore-based methods. *J. Chem. Inf. Comput. Sci., 41*, 394–401.

65. Willett, P. (1987). *Similarity and Clustering in Chemical Information Systems.* Research Studies Press, Letchworth.

66. Wild, D. J., Blankley, C. J. (2000). Comparison of 2D fingerprint types and hierarchy level selection methods for structural grouping using Ward's clustering. *J. Chem. Inf. Comput. Sci., 40*, 155–162.

67. Ward, J. H. (1963). Hierarchical grouping to optimize an objective function. *J. Am. Stat. Assoc., 58*, 236–244.

68. Jarvis, R. A., Patrick, E. A. (1973). Clustering using a similarity measure based on shared near neighbors. *IEEE Trans. Comput., C22*, 1025–1034.

69. Nicolaou, C. A., Tamura, S. Y., Kelley, B. P., Bassett, S. I., Nutt, R. F. (2002). Analysis of large screening data sets via adaptively grown phylogenetic-like trees. *J. Chem. Inf. Comput. Sci., 42*, 1069–1079.

70. Stahura, F. L., Bajorath, J. (2003). Partitioning methods for the identification of active molecules. *Curr. Med. Chem., 8*, 707–715.

71. Pearlman, R. S., Smith, K. M. (1998). Novel software tools for chemical diversity. *Perspect. Drug Discov. Design, 9*, 339–353.

72. Xue, L., Bajorath, J. (2002). Accurate partitioning of compounds belonging to diverse activity classes. *J. Chem. Inf. Comput. Sci., 42*, 757–764.

73. Agrafiotis, D. K., Lobanov, V. S. (2000). Nonlinear mapping networks. *J. Chem. Inf. Comput. Sci., 40*, 1356–1362.

74. Godden, J. W., Xue, L., Bajorath, J. (2002). Classification of biologically active compounds by median partitioning. *J. Chem. Inf. Comput. Sci., 42*, 1263–1269.

75. Godden, J. W., Furr, J. R., Bajorath, J. (2003). Recursive median partitioning for virtual screening of large databases. *J. Chem. Inf. Comput. Sci., 43*, 182–188.

76. Godden, J. W., Furr, J. R., Xue, L., Stahura, F. L., Bajorath, J. (2004). Molecular similarity analysis and virtual screening by mapping of consensus positions in binary-transformed descriptor spaces with variable dimensionality. *J. Chem. Inf. Comput. Sci., 44*, 21–29.

77. Rusinko, A. III, Farmen, M. W., Lambert, C. G., Brown, P. L., Young, S. S. (1999). Analysis of a large structure/biological activity data set using recursive partitioning. *J. Chem. Inf. Comput. Sci., 39*, 1017–1026.

78. van Rhee, A. M. (2003). Use of recursion forests in the sequential screening process: Consensus selection by multiple recursion trees. *J. Chem. Inf. Comput. Sci., 43*, 941–948.

79. Stockfisch, T. P. (2003). Partially unified multiple property recursive partitioning (PUMP-RP): A new method for predicting and understanding drug selectivity. *J. Chem. Inf. Comput. Sci., 43*, 1608–1613.

80. Baurin, N., Mozziconacci, J.-C., Arnoult, E., Chavatte, P., Marot, C., Morin-Allory, L. (In press). 2D QSAR consensus prediction for high-throughput virtual screening. An application to COX-2 inhibition modeling and screening of the NCI database. *J. Chem. Inf. Comput. Sci.*

81. Singh, J., van Vlijmen, H., Liao, Y., Lee, W. C., Cornebise, M., Harris, M., Shu, I. H., Gill, A., Cuervo, J. H., Abraham, W. M., Adams, S. P. (2002). Identification of potent and novel α4β1 antagonists using in silico screening. *J. Med. Chem., 45*, 2988–2993.

82. Chen, H. F., Dong, X. C., Zen, B. S., Gao, K., Yuan, S. G., Panaye, A., Doucet, J. P., Fan, B. T. (2003). Virtual screening and rational drug design using structure generation system based on 3D-QSAR and docking. *SAR QSAR Environ. Res., 14*, 251–264.

83. Hopfinger, A. J., Reaka, A., Venkatarangan, P., Duca, J. S., Wang, S. (1999). Construction of a virtual high throughput screen by 4D-QSAR analysis: Application to a combinatorial library of glucose inhibitors of glycogen phosphorylase b. *J. Chem. Inf. Comput. Sci., 39*, 1151–1160.

84. Labute, P. (1999). Binary QSAR: A new method for the determination of quantitative structure activity relationships. *Pac. Symp. Biocomput., 4*, 444–455.

85. Stahura, F. L., Godden, J. W., Xue, L., Bajorath, J. (2000). Distinguishing between natural products and synthetic molecules by Shannon descriptor entropy analysis and binary QSAR calculations. *J. Chem. Inf. Comput. Sci., 40*, 1245–1252.

86. Harper, G., Bradshaw, J., Gittin, J. C., Green, D. V. S., Leach, A. R. (2001). Prediction of biological activity for high-throughput screening using binary kernel discrimination. *J. Chem. Inf. Comput. Sci., 41*, 1295–1300.

87. Wagener, M., van Geerestein, V. J. (2000). Potential drugs and non-drugs: Prediction and identification of important structural features. *J. Chem. Inf. Comput. Sci., 40*, 280–292.

88. Muegge, I. (2003). Selection criteria for drug-like compounds. *Med. Res. Rev., 23*, 302–321.

89. van de Waterbeemd, H., Camenisch, G., Folkers, G., Chretien, J. R., Raevsky, O. A. (1998). Estimation of blood-brain barrier crossing of drugs using molecular size and shape and H-bonding descriptors. *J. Drug Target., 6*, 151–165.

90. Yoshida, F., Topliss, J. G. (2000). QSAR model for drug human oral bioavailability. *J. Med. Chem., 43*, 2575–2585.

91. Doman, I. N., McGovern, S. L., Witherbee, B. J., Kasten, T. P., Kurumbail, R., Stallings, W. C., Connolly, D. T., Shoichet, B. K. (2002). Molecular docking and high-throughput screening for novel inhibitors of protein tyrosine phosphatase-1B. *J. Med. Chem., 45*, 2213–2221.

92. Stahura, F. L., Xue, L., Godden, J. W., Bajorath, J. (2002). Methods for compound selection focused on hits and application in drug discovery. *J. Mol. Graph. Model., 20*, 439–446.
93. Grüneberg, S., Stubbs, M. T., Klebe, G. (2002). Successful virtual screening for novel inhibitors of human carbonic anhydrase: Strategy and experimental confirmation. *J. Med. Chem., 45*, 3588–3602.
94. van Rhee, A. M., Stocker, J., Printzenhoff, D., Creech, C., Wagoner, P. K., Spear, K. L. (2001). Retrospective analysis of an experimental high-throughput screening data set by recursive partitioning. *J. Comb. Chem., 3*, 267–277.

# SYSTEMS BIOLOGY: APPLICATIONS IN DRUG DISCOVERY

Sean Ekins, Andrej Bugrim, Yuri Nikolsky, and Tatiana Nikolskaya

*GeneGo*
*St. Joseph, Michigan*

1 DEFINITIONS OF SYSTEMS BIOLOGY
2 MODELING PATHWAYS, CELLS, WHOLE ORGANS, AND DISEASES
3 DATABASES AS A FOUNDATION FOR SYSTEMS BIOLOGY
4 HIGH-THROUGHPUT DATA
DNA Sequence Variations
Gene Expression Profiling
Proteomics
Metabolomics and Metabonomics
5 MINING HIGH-THROUGHPUT DATA
Limitations of Clustering Approaches for Gene Expression Data
Advanced Computational Methods for Interpretation of Microarray Data
6 AVAILABLE RESOURCES FOR BIOLOGICAL PATHWAY ANALYSIS
7 APPLICATIONS OF SYSTEMS BIOLOGY FOR DRUG DISCOVERY
Glaucoma
Breast Cancer
Systems ADME/Tox
8 FUTURE DEVELOPMENT OF PATHWAY ANALYSIS SOFTWARE
Reconstruction of Condition-Specific Molecular and Functional Networks
9 FUTURE OF SYSTEMS BIOLOGY APPLIED TO DRUG DISCOVERY
References

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

## 1 DEFINITIONS OF SYSTEMS BIOLOGY

The last several years have seen a paradigm shift for biology. Large-scale international collaborations using high-throughput biology methods have resulted in the sequencing and annotation of the human genome [1, 2], several other mammalian genomes, and hundreds of other genomes as well. The next addition to our growing knowledge base will occur over 2 to 4 years and result in the genomewide individual sets of polymorphisms. The ribonucleic acid (RNA) microarray-based expression technology has also evolved into an almost routine laboratory technique dominating proteomics, metabolomics [3], and metabonomics [4], which have also made substantial progress and will be described later.

Both government agencies and drug companies have invested heavily in these technologies, partly based on the expectation that they will individually or in combination lead to an increase in efficiency of drug discovery. However, despite the unprecedented technology progression and substantial advancement of human biology, genomics has so far failed to contribute to the drug discovery pipeline. In fact, the number of drugs approved by the Food and Drug Administration (FDA) decreased in recent years, with the average cost of bringing these new drugs to the market exceeding $1 billion including marketing costs and taking between 10 to 12 years [5]. The process of drug discovery is therefore lengthy and ultimately complex with several economic and technology factors contributing to this inefficiency. One problem is the poor utilization of already accumulated experimental data, mostly from preclinical research. Recently, robust statistical solutions for data analysis were introduced by bioinformatics companies, and these have been mainly used for data point clustering, visualization, and comparisons between experimental series. However, such statistics alone are not sufficient for meaningful data mining in the context of disease mechanisms. Most experts in industry and academia agree that only functional analysis of the data based on a fundamental understanding of the human biology can solve this bottleneck. The complexity of our own biology, therefore, requires a systemwide approach to genomic and other molecular data analysis. This brings us to the forefront of systems biology, which focuses on the structure and dynamics of complex life systems and is rapidly becoming a leading approach to the integration and mining of data in human biology.

Systems biology can be briefly defined as the integration of genetic, proteomic, transcriptomic, and metabonomic data using computational methods [4]. The tenet of systems biology has been described as identification of all the genes and proteins in an organism, a parts list, which is certainly insufficient to understand the whole. It is therefore apparent that understanding the assembly of these parts [6, 7] (Fig. 1) is truly the key to understanding life, form, and function. In what appears a sweeping rebuttal to reductionism, systems biology appears the new paradigm for interpretation of the biologically complex data sets we are now increasingly generating [7], with higher

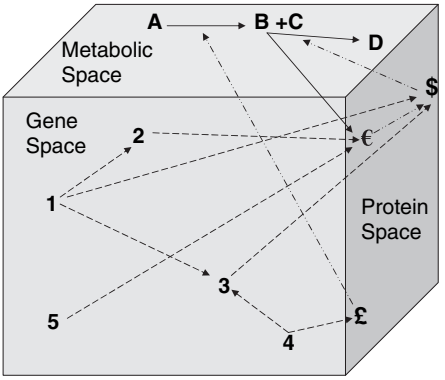


**Figure 1** Schematic to illustrate the complex network of interconnections in the cell between gene space, protein space, and metabolic space. In gene space numbers represent distinct genes with arrows representing connections to other genes or transcription to proteins. In protein space the currency symbols represent unique proteins with arrows representing reactions catalyzed or protein–protein interactions. In metabolic space letters represent molecules that are substrates or metabolites with arrows representing reactions or effects on proteins. (This figure is available in full color at ftp://ftp.wiley.com/public/sci_tech_med/drug_discovery/.)

throughput technologies and the accumulated knowledge on human protein interactions defining certain conditions. Analysis of affected protein networks will allow identification of the sensitive pathways that have high utility as biomarkers for such conditions and new therapeutic targets. This new era for biology has in turn revolutionized the fields of expertise needed from contributors [8] such that engineers and computer scientists are just as likely as molecular biologists to be working on biological projects.

We will describe the various different tools and databases for this type of analysis, their applications, utility, and limitations. The integration of the methods of systems biology with traditional methods of computational biology and computational chemistry represents a powerful approach to combine experimental data, literature information, and predicted properties such that decisions can be made earlier in the development of a drug as to its potential for successful development.

## 2 MODELING PATHWAYS, CELLS, WHOLE ORGANS, AND DISEASES

Due to the use of high-throughput approaches the quality and quantity of available experimental biology observations is no longer a roadblock to understanding the whole cell. The challenge of requiring a data-driven approach for biology pressures computing science as it stands at present to develop advanced integrated tools that organize and structure the existing data for a

complex system to make it more understandable. This is particularly important when we need to use this information for discovery of improved therapeutics and hypothesis testing in silico [9]. Computational models continue to be generated to explain complex cellular signaling pathways, metabolic cascades, and gene regulation [10–12]. These models ultimately resemble complex circuit diagrams, and analogies can be drawn with engineering with respect to feedback, robustness, and fragility [13, 14]. Laws identical to those derived for the electrical system may be obeyed during metabolic reactions [15, 16]. The simulation of these cell circuits with their high level of interconnections and organization is therefore possible in the same way that electrical circuits can be simulated prior to production. This approach may lead to the understanding and replication of experimental results through simulations [17].

Engineering terms are also being used with increasing frequency to describe transcriptional regulatory networks such as the role of hepatocyte nuclear factors (HNF) in control of pancreas and liver gene expression [18]. With the growing database of biological data surrounding metabolic pathways it becomes feasible to reverse engineer such circuit diagrams. One group used genetic programming (previously applied to the automated construction of electrical circuits) to build small metabolic pathways and genetic networks [19]. Such modeling approaches are likely to have a considerable impact on being able to recreate and compute the many simultaneous operations occurring within a cell. Much has been written regarding the fundamentals and difficulties of model construction and analysis [20], and the reader is referred to the studies in this section as well as those based on modeling microbes [21]. The result of such efforts can lead to a second example of modeling in systems biology, namely the "virtual cell" based on a heirachical system of molecules, reactions, and intracellular structure [22–24]. This type of approach enables the construction of a physiological model using biophysical knowledge and the application of physical laws. Once built, these types of models enable the effects of many variables to be studied, provide an understanding of functioning under stress, disease, gene knockout or knockin, and chemical modification [9]. This whole-cell approach has also been used to model the calcium wave in a neuronal cell [22]. The modeling and simulation of integrated biochemical and genetic processes for *Mycoplasma genitalium* has also been accomplished with E-CELL software [25]. All of these types of models represent a tool for iterative hypothesis testing, merging experimental data with computational/bioinformatics approaches. It has been suggested that in cells, molecules may interact with other cellular components in a manner dependent on their molecular properties and probalistic collisions with other reactive molecules [4]. Complex interactions also take place concurrently between endogenous metabolic pathways and the metabolism of xenobiotics requiring a systems biology approach [see later section on systems absorption, distribution, metabolism, excretion, and toxicology (ADME/Tox)]. For example, the need for in silico approaches to predict action on the cardiac cell is at present the most high profile due to the life-threatening consequences of cardiac

arrhythmia [9]. However, drugs themselves seem to rarely bind to a single channel, receptor, or other protein, which makes predictions more difficult. Yet models of the whole heart have been generated to predict action potentials, the spread of activation and enable blood flow calculations [9, 26]. This is just one example of a complex cellular process that could be modeled more completely by understanding whole-cell and whole-organ behavior. It is likely that other whole-organ models will be created in a similar fashion until at some point they can all be integrated. The modeling of disease processes is also possible via *regulatory circuits* generated based on many input parameters that can be used to make predictions and validate therapeutic targets [27, 28]. These types of approaches might provide insights into overlapping molecular processes in disease such as oxidative stress as a trigger for Alzheimer's disease [29] and hypertension [30].

## 3 DATABASES AS A FOUNDATION FOR SYSTEMS BIOLOGY

Metabolism is certainly the most studied area of human biology with over a century of research leading to an understanding of metabolic networks as the major core effector in the living cell. Along with ion channel modulation, changing the concentration of metabolites (including, e.g., neurotransmitters and hormones) is a key method by which the organism acts in response to stimuli. Tissue-specific enzymatic reactions are the endpoints of complex vertical signaling events initiated by membrane receptors and amplified through signal transduction cascades and the ultimate activation of transcription factors. Overall, the interconnected metabolic pathways are the most fundamental network in the body. Gathering and organizing this type of knowledge on cellular pathways is a key first step in creating any system-level representation of a biological system. Since the 1970s, several teams in both industry and academia have systematically studied and organized databases on proteins, enzyme-encoding genes, and metabolic and cell-signaling pathways. Metabolic databases were generated for a wide diversity of organisms to centralize the growing quantities of biological information enabling access to scientists [31] for pathway reconstruction. The following list represents a summary of key databases of interest to those in the systems biology field, including many organism-specific and general-purpose databases for metabolism.

*BIND* Biomolecular Interaction Network Database (*http://www.bind.ca/*), stores full descriptions of molecular interactions, complexes, and pathways for virtually all components of molecular mechanisms and can be used to study networks of interactions, map pathways across taxonomic branches, or to generate information for kinetic simulations [32].

*Biocarta* A database of hundreds of metabolic and signaling pathways that integrates emerging proteomic information and summarizes important

resources providing information for over 120,000 genes from multiple species (*www.biocarta.com*).

*BioCyc* Knowledge Library is the new name for EcoCyc and MetaCyc (*http://metacyc.org/*). It is a collection of pathway/genome databases. Each database in the BioCyc collection describes the genome and metabolic pathways of a single organism, with the exception of the MetaCyc database, which is a reference source on metabolic pathways from many organisms. The MetaCyc metabolic pathway database contains pathways from over 150 different organisms and describes metabolic pathways, reactions, enzymes, and substrate compounds gathered from a variety of literature sources.

*BRENDA* BRaunschweig ENzyme DAtabase (*http://www.brenda.uni-koeln.de/*) is a protein function database including metabolic and enzymatic information [33, 34].

*CSNDB* Cell Signaling Networks Database (*http://geo.nihs.go.jp/csndb/*) is a data and knowledge base for signaling pathways of human cells. It compiles the information on biological molecules, sequences, structures, functions, and biological reactions that transfer the cellular signals. Signaling pathways are compiled as binary relationships of biomolecules represented by graphs drawn automatically [35].

*DIP* Database of Interacting Proteins (*http://dip.doe-mbi.ucla.edu/*) catalogs experimentally determined interactions between proteins curated manually by expert curators and also automatically using computational approaches [36].

*EcoCyc* A database that describes the genome and the biochemical machinery of *Escherichia coli* K12 MG1655 (*http://ecocyc.org/*) and the functions of each of its molecular parts, to facilitate a system-level understanding.

*EMP* Enzymes and Metabolic Pathways (*http://emp.mcs.anl.gov/*) covers all aspects of enzymology and metabolism with about 30,000 records compiled from about 15,000 original journal publications [37]. EMP can also be used for metabolic reconstructions from sequenced genomes.

*ERGO* The ERGO system represents the development of a genome analysis strategy into a multidimensional environment, supporting both automatic and manual genomewide curation and integrating genomic information with biochemical data, literature, and high-throughput analysis into a comprehensive user-friendly network of metabolic and nonmetabolic pathways (*http://ergo.integratedgenomics.com/ERGO/*).

*GenMAPP* Gene MicroArray Pathway Profiler is a computer program designed for viewing and analyzing gene expression data on microarray pathway profiles (MAPPs) representing biological pathways and any other grouping of genes (*http://www.genmapp.org/*). GenMAPP facilitates the analysis of the large amounts of data produced in deoxyribonucleic acid

(DNA) microarray experiments and allows one to visualize gene expression data in a biological context [38].

*GeNet* Regulatory gene network provides information on the functional organization of regulatory gene networks acting at embryogenesis (*http://www.csa.ru/Inst/gorb_dep/inbios/genet/genet.htm*).

*HPRD* Human Protein Reference Database (*http://www.hprd.org/*) is a new resource for human-specific protein information that includes nearly 16,000 protein–protein interactions, 8000 proteins, posttranslational modifications, enzyme–substrate relations, and disease associations. Information is manually curated from the biomedical literature [39, 40].

*HumanCyc* A database that presents metabolic pathways as an organizing framework for the human genome. HumanCyc (*http://humancyc.org/*) provides the user with an extended dimension for functional analysis at the genomic level [41].

*KEGG* Kyoto Encyclopedia of Genes and Genomes (*http://www.genome.ad.jp/kegg/*) is possibly the best known database [42]. KEGG is a database of metabolic and regulatory data for systematic analysis of gene functions, including gene, pathway, and ligand databases enabling the comparison of two genome maps, manipulation of expression maps, graph comparison, and path computation.

*MPW* A collection of more than 3000 metabolic and functional diagrams representing functionality in 211 different organisms (*http://bioresearch.ac.uk/whatsnew/detail/9020325.html*) [43].

*PKR* A database of protein kinase information including tools for structural and computational analyses as well as links to related information maintained by others (*http://www.sdsc.edu/kinases/*) [44].

*SENTRA* A database of sensory signal transduction proteins in mostly microbes; includes information about sensory transduction proteins as well as annotated data from the SwissProt and EMBL databases [45].

*SPAD* The Signaling Pathway Database (*http://www.grt.kyushu-u.ac.jp/eny-doc/*) is used to provide an overview of signal transduction and is compiled in order to describe information on interactions between protein–protein, protein, and DNA as well as information on sequences of DNA and proteins. SPAD is divided to four categories based on extracellular signal molecules (growth factor, cytokine, and hormone) and stress, which initiate the intracellular signaling pathway.

*STKE* Signal Transduction Knowledge Environment (*http://www.stke.org/*) is an Internet-based resource for information on signal transduction.

*SwissProt* A curated protein sequence database with a high level of annotation including the description of the function of a protein, its domain structure, posttranslational modifications, and variants [46, 47].

*TAR-GET* A database of proteins of biotechnological and pharmaceutical interest (*http://www-wit.mcs.anl.gov/target/*). This database was created

to support prioritization of targets for three-dimensional (3D) structural determination.

*TRANSPATH* Comprises data relating to molecules participating in signal transduction and the reactions they undergo, spanning a complex network of interconnected signaling components (*http://193.175.244.148/*) [48].

*Unigene* An experimental system for autopartitioning GenBank sequences into nonredundant gene-orientated clusters with each cluster representing a unique gene (*http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=unigene*). At present there are thought to be approximately 106,219 human genes [49].

*UMBBD* The University of Minnesota biocatalysis/biodegradation database of xenobiotics (*http://umbbd.ahc.umn.edu/*) developed to provide information on microbial enzyme-catalyzed reactions that are important for biotechnology [50, 51].

*WIT2* What is There? (*http://www-unix.mcs.anl.gov/compbio/*) is an integrated system for support of genetic sequence and comparative analysis of sequenced genomes and metabolic reconstructions from the sequence data. This database superceded PUMA—phylogeny of the unicellular organisms (*http:wit.mcs.anl.gov/WIT2/*), a database of metabolic pathways alignments [52].

Combined, these databases provide a high-quality, fairly comprehensive snapshot of human and other organism metabolism or signal transduction that most drug companies and large research centers have selectively licensed. Databases represent one component used for building and testing models of cellular pathways. For example, by integrating genomic and proteomic analyses of a systematically perturbed metabolic network, one group has been able to identify proteins regulated posttranscriptionally in the yeast–galactose–utilization pathway [53]. A similar approach has also been taken to build a genomic regulatory network for development in the sea urchin [54]. Being independent of each other, the databases currently available can be effectively networked using integration modules, such as SRS available from Lion Bioscience (*http://www.lionbioscience.com/solutions/e20472/e20475/index_eng.html*) and BioRS available from Biomax (*http://www.biomax.de/products/f_prod_BioRS.html*). However, neither of these systems can be used as a framework for data mining as they lack key features limiting their utility as follows: The amount of data on pathways, reactions, and interactions captured in these databases is insufficient to be used effectively for deciphering genomewide data sets. According to our own estimates, KEGG covers about 2500 products of human genes that are mostly metabolic enzymes. This represents approximately 15 percent of the human genes with known function. The generation of content for these databases has been slow with mainly voluntary contributions for certain protein families, which does not address

the specific needs of drug discovery. These proteins need to be the most prominent drug targets such as G-protein coupled receptors (GPCRs) and other membrane receptors, kinases, and ion channels important for disease pathways. The collections of pathways in many of these databases (with the exception of ERGO, HumanCyc, HPRD, and Biocarta) are not organism specific, with the data having been collected from many species and different phylla (bacteria, fungi, plants, animals). The human-specific maps in databases like KEGG represent rather generic schemes that are the same for all organisms. The information on human enzymes may also be superimposed onto a network of generic pathways, which ultimately causes many errors and inconsistencies. However, only tissue, cell-type-specific, and expression time-adjusted maps are likely to be relevant, although these are not provided by any of these sources to date. Importantly, the architecture of these databases relies on molecular entities such as genes or proteins, as the elementary objects that are acceptable for isolated proteins and in agreement with the paradigm of *one gene—one protein* relationships as in bacterial systems. With human and other complex eukaryotic systems, this simplistic approach does not provide the functional links between objects, such as when all proteins are involved in the same pathway, or for the expression of all pathway related genes that encode these proteins in the context of the known human disease or disorder. Some of the available database resources provide the pathways in the form of predrawn metabolic maps with no database records for the interactions that constitute them. As a result, users are limited to retrieving one-dimensional information regarding an object, for example, a gene sequence, but not for function-related objects such as all relevant proteins in the pathway or all pathways involved in a disease. The maps themselves are generally static and cannot be supplemented by the users own proprietary data or pathways. With the exception of GenMAPP, few databases provide the tools for the incorporation and visualization of a user's high-throughput data on the pathways and maps. Finally, without links between clinical data and specific elements in the biochemical and cell-signaling networks, models of disease-related cellular pathways are not possible. There have been some novel attempts at using computational software to provide databases of the components of metabolic pathways. One example reproduced radiotracer experiments for substrate product relationships in *E. coli* by mapping the atoms derived from the molecular formulas and computing all pathways between any two given compounds [55]. At present this type of approach appears to have been rarely used.

## 4 HIGH-THROUGHPUT DATA

The dramatic increase in life science research output is largely due to the high-throughput technologies that can now be generated routinely. It is important to understand the limitations of this data as it plays a key role in systems biology modeling.

## DNA Sequence Variations

Since it was first suggested that individual sequence variations can be used as markers for specific traits [56], detection and understanding of individual DNA sequence variability became mainstream in human biology and drug discovery. Specific sequence variations (polymorphisms) and their linear combinations in close vicinity on the chromosome (haplotypes) are known to be associated with predisposition to diseases and differential responses to drugs. Hence these can be used as molecular markers in genetic analysis on an unprecedented scale. A number of types of sequence-based markers were introduced over the last two decades for genetic analysis, including restriction fragment length polymorphisms (RFLPs), variable number tandem repeats (VNTRs), microsatellites, and finally single-nucleotide polymorphisms (SNPs) [57], which became widely used following completion of the draft human genome [2]. SNPs represent an inherited single-nucleotide variation in a given nucleotide in the genome (either in coding or noncoding regions), which is usually biallelic with a minor allele frequency of at least 1 percent [58]. In the human genome the saturation frequency of SNPs is predicted to reach 1 per 200 base pairs (bp), while the current average density of SNPs maps is about 1/1000 bp [59]. SNPs are more abundant in noncoding regions, represent a permanent feature of genomic DNA, are stable, inheritable, and easily detectible either by de novo DNA sequencing or by several methods for in vitro detection of nucleotide mismatches. SNPs can be detected by TaqMan fluorescence probes [60], mass spectrometry of short DNA fragments [61], pyrosequencing [62], Golden Gate assays [63], and resequencing by hybridization [59]. Independently of the detection method, SNP data fundamentally represents the percentage distribution of certain alleles in studied populations. About 6 million human SNPs have been detected of which approximately a quarter are publicly available via the SNP consortium (*http://snp.cshl.org/*), NCBI, and other sources and can be visualized using gene and chromosome browsers in software packages.

## Gene Expression Profiling

***Microarrays*** Complimentary DNA (cDNA) and oligonucleotide microarrays are the most widely used and robust tools for measuring gene expression on a large scale. Microarray technology is based on the fundamental property of complimentarity of genetic code in DNA and RNA [64, 65] and the classic technique of hybridization of nucleic acids, ubiquitous in biology for decades. Since the late 1980s, several technology breakthroughs have boosted the throughput several orders of magnitude and made it possible to evaluate transcription of a genome of any complexity in one experiment. In general, the total RNA is extracted from samples, amplified, labeled with fluorescent nucleotides, and hybridized to arrays with cDNA or oligonucleotide probes. After washing, the array is scanned by laser and visualized by specialized soft-

ware. Oligonucleotide arrays measure the absolute level of each messenger RNA (mRNA), whereas cDNA arrays are typically hybridized with two samples labeled with red dye Cy5 and green dye Cy3 providing relative expression levels for each transcription unit [66]. Although based on the same hybridization reaction, these two array technologies are substantially different in terms of normalization and interpretation and are poorly comparable [67]. Microarrays are typically used for identification of a set of genes with changed level of expression in response to a certain condition(s). The statistical significance of differential expression is then monitored by various statistical tests (described in detail later). Microarray technology is widely applied throughput drug discovery for molecular disease subcategorization [68, 69], prediction of disease recurrence [70], treatment outcome [71], and potential drug toxicity [72–74]. Microarray gene expression profiles are the most reliable, high-throughput, and informative techniques available (excluding, perhaps, genomewide scan by sequencing). Compared with the hundreds to thousands of proteins and metabolites, oligonucleotide arrays may contain over a million distinct features (*http://www.affymetrix.com*). The microarray experiments themselves are run in a highly parallel manner and may contain hundreds to thousands of samples in large clinical studies. The major drawback of gene expression profiling is the distance of transcription from the assembly of active proteins and eventually phenotype (Fig. 2). In any cellular system, the information flow is defined by the main dogma of molecular biology: from DNA to mRNA to proteins. The process is more complicated and includes alternative splicing, RNA maturation, RNA editing, mRNA trafficking, protein trafficing, posttranslational modifications, protein folding, and assembly into active complexes. Eventually, active proteins perform certain cellular functions such as a metabolic transformation of malonyl into acetyl-CoA (Fig. 2), representing one reaction out of thousands of metabolic transformations in the human body regulated at multiple levels from the cell membrane receptors to core metabolism. Ultimately, the interpretation of gene expression profiles would benefit from combination with other available high-throughput data [75].

Although hundreds of large-scale microarray studies have been conducted, the vast majority of the data published is effectively unavailable for public use [76]. The best sources of the available data from microarrays are the Stanford Microarray Database, SMD (*http://genome-www5.stanford.edu/MicroArrat/SMD*), NCBI Gene Expression Omnibus (*http://www.ncbi.nlm.nih.gov/geo*), the database at the Whitehead Institute (*http://www-genome.wi.mit.edu/cgi-bin/cancer/datasets.cgi*), and the NIEHS (*http://dirniehs.nih.gov/microarray/datasets/home-pub.htm*) [66].

***Serial Analysis of Gene Expression*** The serial analysis of gene expression (SAGE) method was developed for simultaneous detection and quantification of the mRNA content of eukaryotic cells [77]. It is based on the synthesis of cDNA from the total cellular mRNA pool, immobilization on streptavidin

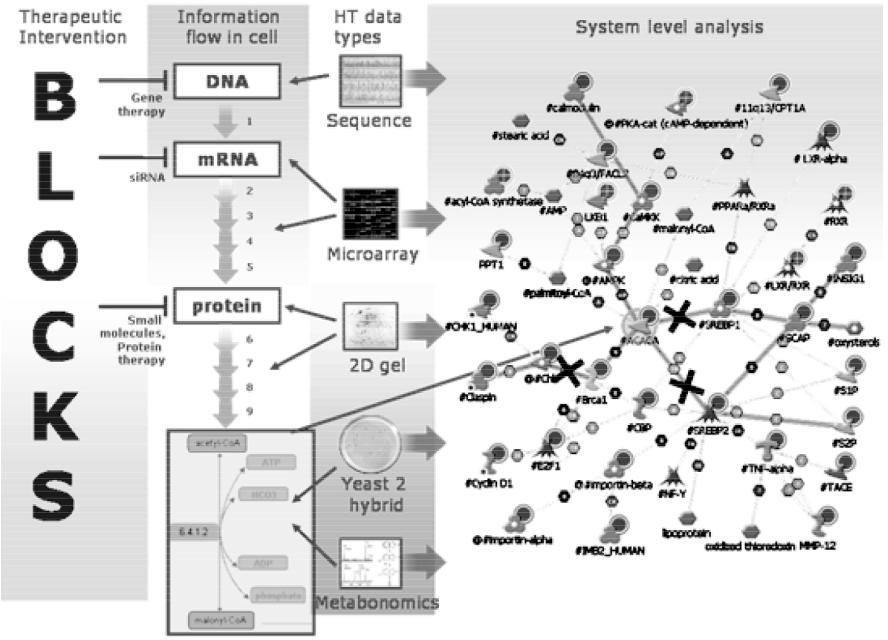


**Figure 2** Utility of high-throughput data for providing information on the functional blocks and use for systems-level analysis. Nine levels of regulation of protein activity in a human cell can be summarized: (1) gene transcription, (2) mRNA processing and editing, (3) mRNA transport from nucleus, (4) mRNA stabilization, (5) protein translation, (6) protein transport, (7) folding and protein stabilization. (8) allosteric modulation, and (9) covalent modification. (This figure is available in full color at ftp://ftp.wiley.com/public/sci_tech_med/drug_discovery/.)

beads, and application of a combination of cloning with two restrictases and polymerase chain reaction (PCR) amplification to come up with hybrid DNA concateners consisting of 3′-end sequences of the expressed cDNAs (tags). The follow-up sequencing of multitag DNA molecules allows identification of the originally expressed genes by querying short tag sequences with sequence databases. The differential expression is determined by the analysis of quantities of gene products under different conditions. The method has two major advantages: It is quantitative and, unlike microarrays, does not require prior sequence information about the expressed gene sequences; however, the method allows recognition of mainly well-annotated genes. The short length of tag sequences (13 to 14 nucleotides) poorly suits sequence alignment and other bioinformatics techniques [78]. Also, the tags are not very specific, as in some cases different genes share the same tag [79]. Although not as widely used as microarrays (especially in pharmaceutical companies), SAGE analysis is inexpensive and can be used for the same applications. Some SAGE experiments are very elaborate and large scale, for instance, when drug sensitivity is tested on multiple lines of tumors and cell lines [80]. SAGE is often

applied in combination with other technologies. Subsets of candidates of differentially expressed genes determined by SAGE can be placed on arrays and used for screening of large patient group [81]. A public SAGE data depository is available at NCBI (*http://www.genome.wi.mit.edu/cgi-bin/cancer/datasets.cgi*).

## Proteomics

The term *proteome* was first coined to define the pool of proteins expressed by the genome [82]. As a field, proteomics deals with quantitative and qualitative multiparallel measurement of protein concentration and/or expression in whole-tissue samples [83]. Measuring protein expression directly is important because possessing a mature mRNA transcript is not sufficient for having a corresponding active protein, due to posttranslational modifications (PTMs), proteolysis, and other dynamic processes causing functional changes [84]. Expression proteomics studies typically involve identification of a set of proteins expressed in a certain tissue under differential conditions and quantification of protein abundance is achieved by a combination of two techniques. Two-dimensional gel electrophoresis (2-DE) allows separation of hundreds of soluble proteins via consecutive application of isoelectric focusing based on a protein's net charge. This is followed by separation based on size in sodium dodecyl sulfate (SDS) polyacrylamide gels [85, 86]. High-abundance proteins can be quantified from the gel using chromatography and fractioning. The method was of relatively limited use until the mid-1990s when advances in mass spectrometry (MS) and bioinformatics enabled the high-throughput identification and quantification of the individual proteins either from the spots on 2D gels or via a combination of chromatography and fractioning [87, 88]. The peptide mass fingerprint approach consists of automated or manual excision of the spots from the gels, protease treatment (using trypsin) of the protein mixture, followed by analysis with matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF-MS) [89]. The peptide composition is then compared to species-specific peptide mass matching [90], and small proteins and those with unknown peptide composition are resolved using tandem MS allowing partial amino acid sequencing [91], followed by bioinformatics sequence alignment against protein sequence databases [87].

Compared to RNA expression analysis and SAGE, proteomics technology suffers from substantial problems including the absence of many important membrane-based receptors and channels that are insoluble proteins and missing from the data. High-abundance proteins such as albumins often obscure low-abundance proteins on 2-DE gels, and their effective separation requires complex prefractioning and other techniques [92, 93]. Proteomics is, however, quite widely used in research and clinical diagnosis of most complex diseases [94–96].

Most cellular processes are performed by assemblies of interacting proteins and proteomics is the only technology that allows experimental analysis of

protein complexes and protein–protein interactions. The experimental arsenal of functional proteomics includes several ligand fishing techniques such as co-immunoprecipitation, sequential affinity tag purification [97], and biochemical interaction analysis–mass spectrometry [98]. Specific antibodies or tag-specific proteins are used as the bait for catching the interacting proteins as prey, followed by complex resolution by MS [84]. In two key studies, hundreds of yeast protein complexes were resolved by these methods, comprehensively covering the whole yeast proteome [99, 100].

***Yeast Two Hybrid*** The most popular method for detection of protein–protein interactions is known as the yeast 2-hybrid system (Y2H [101]). Y2H is a simple genetic test based on modular structure of yeast transcriptional factors Gal 4 and Lex A, which are only active as dimers assembled on palindrome binding sites. Truncated for dimerization domains, the proteins can restore activity when fused to a pair of physically interacting proteins. The activity can be detected using color selection with the *LacZ* reporter or by complementation of auxotrophic mutants *leu2*, *ade2*, and *his3* [102]. The Y2H system is inexpensive and powerful for screening of protein libraries for interacting protein pairs, identification of unknown interacting genes, and detection of weak interactions that could be missed by proteomics methods [103]. This method is also a truly high-throughput method scalable for genomewide protein interactions screens. Some of the most important model organisms with sequenced genomes including yeast [104], fly *Drosophila melanogaster* [105], and worm, *Caenorhabditis elegans* [106] were mapped for interactions in this way. However, the system is known for its high (up to 50 percent) level of false-positive and false-negative interactions [107] due to interacting proteins that are overexpressed, subjected to folding in the foreign environment of yeast cells, and have to be trafficked into the nucleus prior to action. Filtering out false positives can be achieved by cross-referencing with other experimental data and superimposition of YH2 interactions onto networks of annotated pathways [108]. Specialty variations of Y2H have been developed, such as a reverse two-hybrid system for studying molecules disrupting the interactions and the split-ubiquitin system to analyze membrane proteins [102]. The raw Y2H data is therefore qualitative and represents simple tables of condition-specific interacting partners.

***Protein Arrays*** Ever since the introduction of DNA chips in the early 1990s, creation of arrays with grids of active proteins became the holy grail of proteomics research. Unfortunately, the key DNA properties enabling the whole-genome arraying (active primary structure, PCR amplification, ease of labeling, stability, etc.) are not shared with proteins. Arraying of properly folded, nondenatured proteins and the development of detection techniques suitable for proteins of many different functions are nontrivial tasks not easily resolved yet [109]. The first proteins to be successfully used in array development were stable like antibodies [110]. As whole antibodies are large and

inconvenient to work with, the technology was extended to antibody fragments [111], phage display-based expressed variable fragments [112], and oligonucleotides aptamers [113] as the source of content for arraying. MALDI targets investigated with affinity ligands, followed by affinity MS, are usually applied for capturing and detection of proteins of interest. Several techniques have also been developed for attaching proteins to surfaces such as the application of adenylate-treated glass, which reacts with primary amines [114], and tagging proteins with polyhistidine attached to nickel-containing slides [115]. To circumvent the differences in activity assays, the proteins are often arrayed according to functional target class (e.g., kinase arrays), although more recently multifunctional arrays have been developed.

## Metabolomics and Metabonomics

Metabolomics and metabonomics are similar terms defining the profiling, measurement, and analysis of the components of the metabolome representing all low-molecular-weight molecules present in cells in a particular condition. Metabolomics can be defined as a discipline investigating metabolic regulation and fluxes in cells, and metabonomics is a larger scale, systemic determination of biochemical profiles and their regulation in biofluids and tissues [4]. Metabolites represent the endpoint of the organism's response to stimuli, the ultimate outcome of the complex information transfer network, signaling, and biochemical transformations such that metabolic profiling is therefore the most direct measure of physiology [116]. The composition of metabolites in samples can be very heterogeneous, and the measurement procedure often involves prefractioning, selective enrichment, and parallel measurement. The most robust methods in metabolomics are gas chromatography–mass spectrometry [(GC–MS), which resolves 500 to 1000 metabolites [117]], and liquid chromatography–mass spectrometry (LC–MS) [118]. These two methods, applied in combination, allow resolution of a thousand peaks for known organic small molecules and unidentified metabolites [119]. Unknown metabolites can be identified using a higher precision Fourier transformation mass spectrometry (FT-MS) [117]. The third approach is nuclear magnetic resonance (NMR) spectroscopy, applied mostly for detection of high-abundance known metabolites [120] and is particularly useful for metabolic fingerprinting used in clinical studies [121].

Interpretation of metabolic profiles of cells and tissues imposes arguably more problems than any other method described earlier, as metabolic fluxes and profiles cannot be explained by regulation of gene or protein expression alone [122]. The ultimate goal of quantification of all metabolites in a cellular system is currently impossible. This is due to lack of reproducible and robust automated analytical techniques, chemical heterogeneity of metabolites, lack of automated extraction techniques, and low resolution of existing hardware [123]. Moreover, taking into account complex relationships between endogenous and xenobiotic metabolism and the coexistence in the human body of up

to 5000 species of metabolically active microorganisms, the full metabolic profile cannot be interpreted. This requires consideration of probabilistic concepts for determining symbiotic human metabolism [4]. Such complexity requires interpretation at the systems level, via integration of metabolomics with gene expression and proteomics data to unify metabolic, regulatory networks, and infer condition-specific metabolic networks [124], shifting our thinking of metabolic pathways as networks and subsystems [125].

## 5 MINING HIGH-THROUGHPUT DATA

The quantitative leap in experimental throughput required methods for comprehension of global gene expression experiments that have been useful in certain applications. Interactions could, however, be missed if analysis is restricted to expression patterns only. The statistical analysis of microarray data results in gene sets that are classifiers for certain conditions such as the disease type, but they do not explain biological mechanisms and cannot be directly linked to clinical observations or other phenotypes. Visualization-based approaches to high-throughput data lead into model-based approaches to provide broader understanding and insight. The state of the art at present are graphical approaches that link objects with vectors and other graphical annotations that can be queried [126].

### Limitations of Clustering Approaches for Gene Expression Data

Three general statistical procedures are used for clustering of gene expression patterns on arrays including hierarchical clustering [127], *k*-value clustering [128], and self-organizing maps (SOM or Kohonen maps) [129]. The most common method is the hierarchical clustering procedure, which consists of a consecutive pairwise grouping of spots: The closest pairs are replaced by a single data point representing their average. Ultimately, this results in an expression pattern represented as a phylogenetic tree with a branch length corresponding to the degree of similarity between sets. In contrast, *K*-means clustering uses a predetermined K value that is the expected number of elements in the cluster. This approach, therefore, produces an unorganized compilation of clusters. A further approach is Bayesian clustering, which generates multiple recursive partitions into subgroups and assigns Bayesian probabilities to each group. This method requires a strong prior data distribution and is used in combination with other methods [130]. SOMs are better suited for exploratory data analysis as they impose a partial nonrigid, structure on clusters and facilitate visual interpretation [129, 131]. All these statistical methods have been applied to microarray-based disease subclassification by either unsupervised or supervised methods. In unsupervised clustering, the differential expression data is parsed into clusters of points statistically most related without prior knowledge of the phenotypic subcategories [132]. These differ-

ent methods are therefore useful for various applications with each having attendant advantages and disadvantages. As regulation of mRNA expression constitutes only one level of biological control, these levels of complexity multiply in global genomewide experiments such that information can be lost or altered and clustering at the first level of expression can only be indirectly linked to eventual phenotype. This one-dimensional analysis is a limitation in the applicability of expression microarrays for studying complex diseases [6].

## Advanced Computational Methods for Interpretation of Microarray Data

***From Linear Pathways to Complex Biological Networks*** Traditionally, functional organization of a biological system was described in terms of pathways. Pathways were thought of as fairly small linear chains of biochemical reactions or signaling interactions that lead from a defined starting point (e.g., cell surface receptor) to a target effector (e.g., transcriptional factor). Such a description is in part due to the nature of biological research itself that until recently was inherently low throughput, with data scattered in tens of thousands of individual publications. In recent years, however, three major developments have taken place in this area in terms of pathway databases, natural language processing (NLP) algorithms for automatic extraction of pathway information online abstracts, and high-throughput techniques for determining potential protein–protein interactions (described above). Based on the data available from these technologies, it has become clear that organization of intra- and intercellular molecular processes is much more complex. Contrary to the previous model of fairly independent small pathways, it is now evident that known molecular processes can be linked into rather large highly interconnected networks. A recent review has described in considerable detail the theory behind network models, their architecture, and how they can be used to provide insight into the functional organization of the cell [125].

Further analysis of the existing databases on metabolic and signaling pathways shows that if taken together, known reactions and interactions would form a large cluster linked via molecular nodes shared among many processes [133]. If one tries to describe such clusters in terms of linear pathways, then billions of cascades can be generated. For example, an analysis of ~20,000 cell-signaling interactions culled from the literature (obtained from MetaCore, GeneGo Inc.) shows that such a set contains approximately 2 billion five-step pathways.

It is important to note that, taken together, information on molecular processes derived from different sources represents a universe of a putative biological functionality. Only a small fraction of it will be realized in a cell at any given time. Genes, proteins, and metabolites are tightly regulated on many different levels in molecular networks (Fig. 1). For example, only a fraction of genes are realized as protein products in a particular cell type. Additional constraints include spatial separation by subcellular localization, thermody-

namic feasibility of a process in a specific environment, competition among processes for shared molecules and other physical, chemical, and biological limitations. Theoretically, a computational model of a full network can predict the dynamic behavior of a system. At present, however, dynamic models can only be built for a limited number of fairly well studied systems, and selection of which processes to model is limited by unavailable or inaccurate quantitative information for most of the tens of thousands of reactions and interactions. To add to the complexity, unique isoforms of the same protein expressed in different cells may vary dramatically in binding constants and other parameters. The lack of detailed quantitative knowledge is a major impediment for direct dynamical modeling of large molecular networks, hence requiring an alternative approach for modeling.

***Elucidating Pathways and Interactions from the Literature*** There is also a growing interest in elucidating the pathways and protein–protein interactions from the literature. Differentially expressed genes can be clustered into functional groups based on their names or co-occurrence in MEDLINE texts [134, 135]. Different computational methods like hierachial clustering [136] and NLP (reviewed in [137]) have been applied for enhancing the biological relevance of clusters and literature coverage. However, only 30 percent of experimentally verified protein interactions were reported to correspond to pairs from MEDLINE abstracts [138] and false-positive interactions composed about 50 percent of NLP-extracted associations, partly due to poor synonym resolution [134]. A more recent publication used an NLP method, MedScan, to extract 2976 interactions between human proteins from MEDLINE abstracts with a precision of 91 percent for 361 randomly extracted protein interactions [139]. There are a number of databases covering different aspects of protein function that can be used for testing these methods, such as those described earlier. These are generally more biologically relevant sources of information on interacting proteins, particularly when extracted from full-text experimental literature (BIND, INTERACT [140], and DIP). However, these databases have relatively low coverage for human protein interactions. Subsequently, MedScan was able to extract novel (96 percent) information that was not in BIND and DIP [139]. The first attempts at consolidation of two approaches, that is, co-expression clustering and deduction of functional networks from literature and knowledge databases were performed for *E. coli* and yeast [141]. Overall the functional interpretation of expression data is in its infancy, and tools that have been developed have not been applied yet for the analysis of human and other higher eukaryotes. In many cases, there is a conserved gene expression pattern that is shared between species, and this corresponds to the same functional groups of orthologs. Similar genes are therefore regulated in the same manner across species, and these are usually involved in basic processes such as the cell cycle, metabolism, and secretion. New genes involved in these processes can therefore be identified via comparative analysis [142, 143] among the different

species, and alterations in gene expression can be detected by microarray techniques [143].

Since many biological functions are carried out by discrete modules of physically interacting proteins [144], searching for the algorithms useful in identifying such interactive networks has become the key to microarray expression analysis and one of the methods for functional annotation for novel genes successfully demonstrated with yeast [145, 146]. Some algorithms can link the expressed genes with known transcriptional factors via the genomewide identification of DNA binding sites [147] or enable building of condition-specific probabilistic models [148]. The other published studies group the co-expressed genes based on the occurrence of common cis elements in the context of reference knowledge databases of transcriptional factors [149] of well-known metabolic pathways. For instance, MIPS (Munich Information Center for Protein Sequences) functional categories [150] may be used, which take into account pairwise sequence-based conservation between co-expressed proteins [151]. Microarray data is therefore often linked to curated knowledge databases on proteins, genes, and pathways such as SwissProt, KEGG, and others. Although robust and relatively comprehensive, the available analytical methods and databases are not sufficient for coping with the vast complexity of human biology and the elucidation of integrated cellular networks.

***Making Sense of Biological Networks*** Systems biology is hence evolving to analyze all high-throughput experimental data in the context of the global physiology and disease pathology. This approach seeks to identify the networks of cellular pathways and corresponding physically interacting proteins [152]. The combination of all of these data types would be ideal to provide a truly multidimensional understanding of cell function. As direct dynamical modeling is currently limited to well-understood systems, a novel approach for making sense of large networks has been developed in recent years. This approach combines graph theory and statistical methods for finding network modules that represent potential active functional units. Two major directions of research exist in this area. The first represents attempts to understand the fundamental global organization of molecular networks. The second aims at finding potential functional motifs by studying local properties of a network and by combining protein or metabolic networks with data on gene expression for elucidation of condition-specific active modules. One important result achieved in this area is the realization that biological networks of different origins (e.g., metabolic, regulatory, protein interactions, networks for different organisms) share the same global architecture [105, 106, 125, 153]. Namely, these networks are scale free, meaning that distribution of node connectivity obeys the power law: $P(k) \sim k^{-\gamma}$, where $k$ is the number of node links (also called node degree) and $P(k)$ is the fraction of nodes in the network with exactly $k$ links. One consequence of such an architecture is the presence of so-called hub nodes, which are high-degree nodes connected to many low-degree

nodes [125]. Such topological features give networks the property of robustness with removal of even a substantial fraction of nodes leaving the network connected [154]. This robustness may also have far-reaching implications for the selection of appropriate drug targets.

While understanding the global network architecture is an important first step, from the practical standpoint of drug discovery it is even more interesting to find disease-specific functional modules that may contain potential drug targets. This task is a much harder one as any complex network can be divided into subsets in many different ways, potentially generating billions of combinations. Different algorithms and criteria have been used in an attempt to automate parsing of such large networks into modules. One set of methods finds network modules based on network connectivity and various clustering algorithms. Examples include the Monte Carlo optimization method for finding tightly connected clusters of nodes [155], clustering of nodes based on the shortest path length distribution [156], and other graph clustering algorithms [157]. It was shown that many clusters that are found in such a way correspond to either known protein complexes or metabolic pathways [155]. Another way to approach finding functional modules is by analyzing network motifs [158]. Motifs are fairly simple subgraphs that share certain structural features, usually related to potential functionality. For example, feedback or feed-forward loops are typical motifs. In this approach a number of different motifs in the network are calculated and compared with a number of the same motifs in a randomly connected network. Those motifs in which the network is enriched may represent potential functional modules. In several studies performed to date motifs were found in regulatory networks of *E. coli* [159, 160] and in yeast protein interaction networks [161]. The next step in elucidation of functional modules is the analysis of high-throughput molecular data in the context of networks. By performing such analyses one may find mechanisms that are active under specific circumstances, such as disease or drug treatments (discussed in more detail later).

## 6 AVAILABLE RESOURCES FOR BIOLOGICAL PATHWAY ANALYSIS

The functional annotation of the filamentous fungus *Aspergillus nidulans* was to our knowledge one of the first attempts to reconstruct functionality of a complex eukaryotic organism based on the integration of Expressed Sequencing Tags (EST) data (from the Unigene EST collection) with a comprehensive literature-based collection of functional pathways metabolic pathways database (MPW) [43]. The study was based on a method of metabolic reconstruction, proposed for annotating prokaryotic genomes in the computational environment WIT [52, 162]. The method consisted of compiling metabolic blocks (pathways and subsystems) corresponding to the genetic component of an organism and connecting these blocks via intermediates into wire diagrams

or metabolic reconstruction models. Over the past few years a new generation of computational biology software products has emerged for biological pathway construction and analysis such that the following are now commercially available.

**Cell illustrator** (*http://www.gene-networks.com*) is a tool for manual construction of pathways and simulating pathway mechanisms of action of both baseline and abnormal conditions.

**DiMSim** (*http://www.bio.cam.ac.uk/~mw263/bioinformatics.html*) is a methodology for simulating the flow of metabolites through networks of interacting metabolic reactions. This enables views of the reactions involved in metabolic processes in a graph format useful as the basis for simulations. The DiMSim approach is in marked contrast to other metabolic simulation systems that are based on differential equations, it is therefore able to model down to single molecules and cope with the special characteristics of biochemical systems [163].

**Ingenuity pathways analysis** (*http://www.ingenuity.com/*) enables the entry of a gene list from microarray and proteomics experiments that are then evaluated against a knowledge base to produce a gene network.

**MetaCore** (*www.genego.com*) relies on a unique, expert-curated database of pathways annotated with tissue specificity and subcellular localization and containing pathways for over 80 percent of known human genes. This data comes primarily from two sources: the original publications and expression data from the Unigene database. MetaCore also provides links between the pathways (or their elements) and disease states collected from the literature. Currently, there are more than 32,000 disease links classified into 6 major categories from the highest to lowest level verification, namely c*ause*, *manifestations*, *hypothesis*, *animal, treatment*, and *no relation* and provides several different ways of exploring the data. The major components of MetaCore are described in more detail below.

*Metabolic Component* A comprehensive collection of metabolic pathways with interactive maps of the major functional blocks of human metabolism is organized in a three-tier architecture.

*Cell-Signaling Component* Cell-signaling and regulatory pathways important for diseases such as cancer (see later) are the major experimental content with approximately 20,000 signaling interactions assigned to over 6000 molecular objects. Some of these interactions are assigned to classes containing more than one molecule with signaling vertically integrated with other types of cellular processes.

*Pathway Visualization Options* Visualization options in MetaCore include over 300 fixed-curated maps for the major functional blocks of cell signaling and metabolism, color-coded for subcellular localization, and superimposed on the graphical images of the various human tissues (Fig. 3). In addition, there

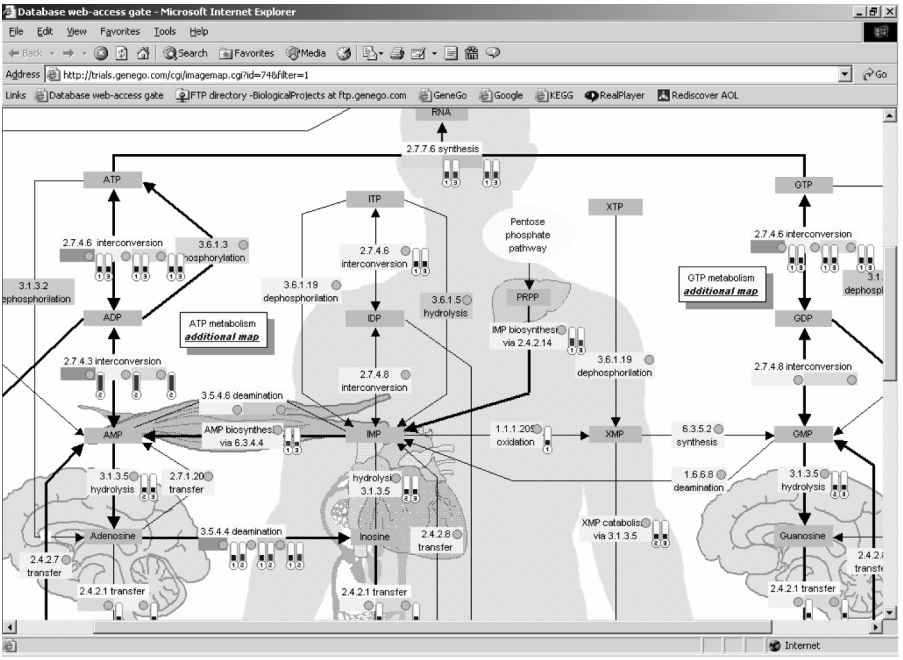


**Figure 3** Interactive map for human purine metabolism from MetaCore. Metabolic pathways are shown as rectangular icons; the enzymes are tagged with EC numbers. The differentially expressed genes are uploaded directly from a microarray experiment and are marked next to the pathways as thermometer objects.

is a pull-down menu with a list of human EST libraries from the Unigene database. Computer-generated signaling and metabolic networks can also be produced that are highly interactive.

*Data Visualization Options* Both the curated maps and computer-generated networks can be used as a framework for visualizing gene expression (thermometer-like icons, Fig. 3) and other types of high-throughput data such as Affymetrix and Agilent microarrays, 2D proteomics data, SAGE, DNA sequence data, and molecular formulas for metabonomics. The visualization tools allow concurrent comparison of multiple experiments, with custom normalization and filtering features.

**PathArt** (*http://www.jubilantbiosys.com/pd.htm*) represents over 600 signaling and metabolic pathways, 5000 molecules and 7500 interactions incorporating data from peer-reviewed articles and patents enabling pathways and critical genes to be identified and compared across physiological diseases and organisms. The data covered is cell, tissue, and organism specific for human and mouse, with coverage of protein–protein interactions their specific domains, motifs, key binding residues, and experimental techniques.

**Pathway Assist** (*http://www.ariadnegenomics.com/products/pathway.html*) is a software tool for pathway creation, expansion, and visualization to allow the interpretation of gene regulation and protein–protein interactions using an NLP method (MedScan, described previously) [164] and can be used with the PathArt database. The software can utilize microarray data, allow data filtering and parsing from many sources, and currently describes more than 100,000 events of regulation, interaction, and modification between proteins, cell processes, and small molecules.

**PathDB** (*http://www.ncgr.org/pathdb/*) is both a data repository and a system for building and visualizing cellular networks targeted for the gene expression, proteomics, and metabolic profiling communities. Uses include finding all pathways and phenotypes associated with genes in a cluster or validating computational predicted associations with known biological data.

**PubGene** (*http://www.pubgene.org/*) can search through the millions of biology-related studies published for the names of genes and proteins and for sequence homologies between them. The articles are then searched for the occurrence of gene pairs. Based on this information, a map of all genes is made, based on the numerical information of gene pair frequencies. When combined with results of microarray analysis, the program can then estimate the importance of the various relationships between genes and proteins, utilizing the biological knowledge for human, mouse, rat, and yeast genes and proteins, from such databases as Medline, RefSeq, and PIR. A number of layers of metadata from various sources are included to allow for flexible information content display depending on the user's preference. PubGene uses in-house dictionaries of gene and protein names and gives the user synonyms resulting from a query [134].

**VLX Biological Modeler** (*http://www.teranode.com/applications/pathway modeling.php*) automatically translates pathway models into mathematical models based on biophysical and biochemical principles.

**The Cognia Molecular System** (*http://www.cognia.com/products_cm_indepth.htm*) comes with a preloaded framework of information parsed from public data sources on over 100,000 proteins, genes, interactions, 250,000 compounds, and with links to over 35 external databases. This content is broad, covering many species and functions, roles, locations, and interactions that can be displayed as networks with NetworkBuilder. The visual output of the NetworkBuilder is filterable and customizable to extract the dynamically built output as well as component attributes within the database.

All of these products have unique underlying proprietary pathway databases. These are either human curated (GeneGo MetaCore, Jubilant PathArt) or are compiled from MEDLINE abstracts using automated text-mining tools (Ingenuity PathwayAssist). Alternatively, they may contain a combination of curated and automatically extracted data (Ariadne Pathway Analyst). Their respective performance obviously varies depending on the flexibility of the database schemas, the quantity and quality of the assembled data, and the algorithms implemented into their data-mining tools. To date there has been

no comparison of the different approaches published, and it is likely that investigators in the industry may require access to multiple software systems to balance the overall strengths and weaknesses of individual methods.

## 7 APPLICATIONS OF SYSTEMS BIOLOGY FOR DRUG DISCOVERY

The following examples describe applications of gene network analysis software and systems reconstruction for particular topics relevant to drug discovery.

### Glaucoma

Glaucoma is a progressive optic neuropathy characterized by structural damage to the optic nerve leading to blindness through loss of retinal ganglion cells. Risk factors in glaucoma include elevated intraocular pressure (IOP), age, race, family history, myopia, and diabetes. The growing body of evidence indicates there is a heightened risk of developing glaucoma among individuals with Alzheimer's disease [165–168]. The pathogenesis of the optic nerve neuropathy in glaucoma is a matter of debate. It is widely accepted that elevated IOP and a variety of factors may all contribute to primary insult of the optic nerve [169–171]. Cellular mechanisms involving vascular insufficiency [172, 173], vasospasm [174, 175], glutamate exitotoxicity [176], neurotoxic cytokine release [177, 178], abnormal metabolism, and autoimmune reaction have all been suggested [170]. There are numerous major factors contributing to the cumulative optic nerve insult and malnutrition in glaucoma (Fig. 4). Ultimately, glaucoma affects an estimated 70 million people worldwide, mostly 40 or older. This represents a fast growing market for pharmaceutical intervention because of the progressively aging population of the Western world. At present there are neither drugs on the market that address the disease mechanism or reliable molecular diagnostics for addressing susceptibility or severity. As is the case with many complex diseases, global gene expression profiling using microarrays is becoming an important powerful tool in glaucoma research.

Over the last 2 years, several groups studied differential gene expression in glaucomous human optic nerve head (ONH) astrocytes [179], monkey [180], and rat [181] disease models as well as response to corticoid hormones in different eye tissues [182, 183] and the response in normal ONH astrocytes to high pressure (a key risk factor associated with glaucoma [184]). In the pioneering work of Hernandes et al. [179] the expression of genes in cultured human astrocytes from glaucomous and normal ONH astrocytes were compared using standard Affymetrix U95Av2 arrays consisting of over 12,000 genes. Approximately 1700 genes were differentially regulated in diseased astrocytes relative to normal ONHs. In 150 of these genes a 5-fold or higher

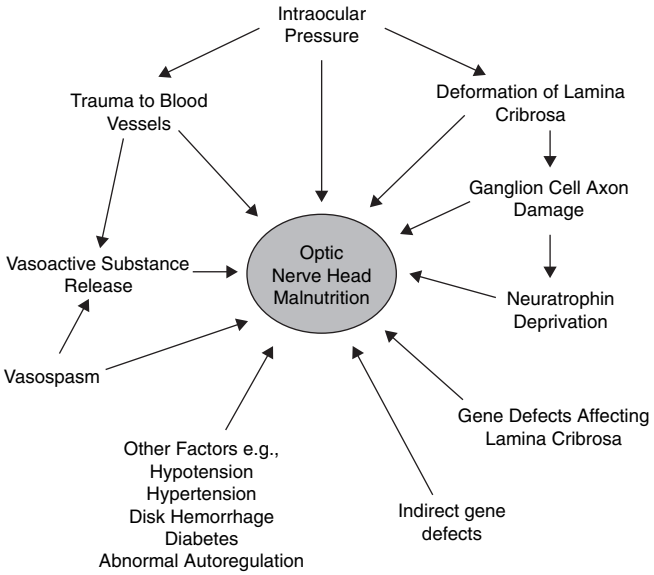


**Figure 4** Some possible causes of initial insult to the optic nerve head in different glaucoma patients [171]. While increased intraocular pressure (IOP) is the most significant risk factor for glaucoma, RGCs cell death caused by optic nerve deformation may provide an explanation for a mechanical cause of glaucoma. While elevated IOP definitely plays a role in structural displacement of the ONH causing cytosceletal alteration, and loss of microtubules in RGC axons impedes retrograde axonal transport [258–260], it is conceivable that it also provides indirect insults via reduced blood flow and reactivation of microglia [178]. The role of vascular factors also is thought to be of significance to optic nerve and RGCs injury [172–175].

difference was noted, and these could be classified according to generic biological functions, including signal transduction, transcription, cell adhesion, proliferation, and metabolism. In the follow-up study [184], the same group interrogated the response of normal cultured astrocytes under increased hydrostatic pressure (HP) using the same molecular techniques and type of array. The expression of 596 genes was altered; 38 genes were upregulated and 24 downregulated over time with a threshold of 1.5 or higher. The 38 genes were also analyzed by hierarchical clustering analysis. The genes were of multiple cellular functions that were indicative of the systemic effect of high pressure on astrocytes. In two studies on animal models, glaucoma was induced in the retina of rat [181] and monkey [180] while differential expression was evaluated using total mRNA from whole retina as probes. The advantage of these models is that the genetic background is the same since one eye remains normal, and disease is induced in the other of the same animal. Sixty-two and 39 genes were differentially expressed in mild and severe glaucoma in monkey, respectively. In rat 81 genes were differentially expressed, although most were

not the same genes. Finally, using dexamethasone the induction of gene expression was studied in cultured human trabecular meshwork (HTM) cells [183] and several other eye tissues [182]. Dexamethasone specifically induces expression of myocilin (TIGR/MYOC) gene in HTM cells and is linked to several types of glaucoma. When comparing these two studies, the results were neither consistent nor conclusive: 30 genes were upregulated over 2-fold in HTM-DEX cells in the Ishibashi study [183] while 249 genes were upregulated 4-fold or greater in the work of Lo et al. [182].

Microarray expression analysis in glaucoma has not as yet been used for subcategorizing different types of glaucoma, identifying prognostic gene lists for predicting disease outcome and response to treatment, or comparing human and animal studies. The approach of using enriched cell-type-specific primary cultures has been undertaken by Hernandez and co-authors [179], and this approach is likely the most promising since it uncovers the contribution of specific retina cell types in the development of neuropathy and allows the reconstruction of cell-specific alterations in molecular pathways implicated in the disease. We have compared the genes expressed in disease tissue (the endpoint of microarray analysis) from four recent studies in glaucoma and showed a lack of consistency between these different sample sets. Less than 5 percent of the genes were the same (on average) between any pair of experiments (Table 1). This lack of consistency in genes identified was somewhat surprising since the disease pathologies appear similar among monkey, rat, and human and because these species share about 99 percent of the same genes. Even more surprising is the inconsistency between expression in normal human astrocytes under high pressure (believed to be the key risk factor in glaucoma) and glaucomatous astrocytes. Since high pressure is believed to be a key risk factor for glaucoma or, at least, lead to the induction of glaucoma, a higher percentage of common genes were expected. These conclusions are based upon signature molecular profiles found when comparing between individuals for other diseases such as in breast cancer [130]. Although there are several potential reasons for such inconsistencies, the major problem is the application of analytical methods to tissues from different genetic backgrounds with different (and uncontrolled) environmental conditions.

We have performed a comprehensive computational gene network analysis of microarray expression in human glaucoma using these previously published data sets [179]. The raw expression data on about 1700 genes differentially regulated in glaucoma astrocytes from 4 glaucoma patients and 4 age-matched normal individuals was loaded in MetaCore. After data normalization and processing, multiple networks were then built using different algorithms and expression threshold ratios. One of many networks is presented on Figure 5. The threshold ratio of 2.5 and higher was set up for this network before we applied the stringent algorithm called *immediate interactions*. This algorithm allows the connection of only proteins (presented as coding genes), which are: (1) experimentally shown to physically interact pairwise and (2) being over- or underexpressed above the threshold ratio. The pathways are highlighted in Figure 5 as solid lines.

**TABLE 1 Overlap Between Sets of Glaucoma-Relevant Differentially Expressed Genes from Recent Studies**

| | Hydrostatic (1.5-fold) [261] | Glaucoma (5-fold) [179] | Rat Retina (2-fold) [181] | Monkey Retina, Mild Glaucoma (1.5-fold) [180] | Monkey Retina, Severe Glaucoma (1.5-fold) [180] |
|---|---|---|---|---|---|
| Hydrostatic (1.5-fold) [261] | | CDC42, FOS, VEGFC, Clusterin, HF1, C1q, C5, BMP4 | VEGFC | | |
| Glaucoma (5-fold) [179] | CDC42, FOS, VEGFC, Clusterin, HF1, C1q, C5, BMP4 | | C1q, c-Jun | C1q, c-Jun | C1q, c-Jun |
| Rat retina (2-fold) [181] | VEGFC | C1q, c-Jun | | C1q, c-Jun, GFAP, ceruloplasmin | C1q, c-Jun, GFAP, ceruloplasmin |
| Monkey retina, mild glaucoma (1.5-fold) [180] | | C1q, c-Jun | C1q, c-Jun, GFAP, ceruloplasmin | | v-akt, chitinase 3-like 1, C4a, MHC class I DR alfa, FB, KIAA0805 |
| Monkey retina, severe glaucoma (1.5-fold) [180] | | C1q, c-Jun | C1q, c-Jun, GFAP, ceruloplasmin | v-akt, chitinase 3-like 1, C4a, MHC class I DR alfa, FB, KIAA0805 | |

Of particular interest are the grossly downregulated pathways of integrins, actin synthesis, actin-mediated signaling, TGF-beta1, NGF, collagen IV biosynthesis, and microtubule polymerization presented on the left hand of the network. Underexpression of some elements of these pathways is in good agreement with the observed phenotypic changes accompanying pathological reactivation and has been shown as a signature for glaucoma astrocytes [179, 185, 186]. Overexpressed pathways (on the right-hand side on the network) predominantly includes Nf-kB and AP-1/c-Fos/c-Jun-activated pathways. Both transcription factors Nf-kB and AP-1/c-Fos/c-Jun are strategically located in the watershed zone separating the up- and downregulated pathways. These genes are subsequently linked to the activation of downstream pathways via activation of transcription factors Rel, c-Myc, HMGY, and CREB, consistent with their known functions in promoting cell survival and amplification [187]. The network also indicates several activating interactions leading to potentially cytotoxic pathways and compounds such as interleukin-8 (IL-8), interleukin-6 (IL-6), amyloid precurson protein (APP)/β-amyloid, nitric oxide synthase (iNOS), and clusterin. The growing body of experimental data links the upregulation of these pathways with glaucoma and RGC death [188]. Separate smaller networks showing interactions within these individual nodes were generated using a smaller subset of genes (Figs. 6*a*–6*c*). Examples of the types of results and interpretation of the data are described next.

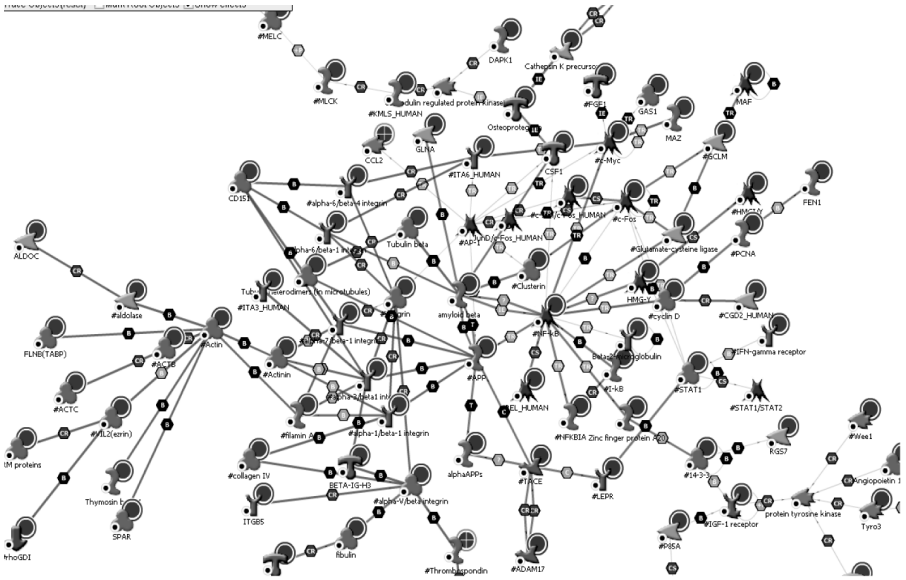


**Figure 5** Functional network for differential gene expression in glaucoma. Pathways are marked in solid lines. Circles on the nodes mark the differentially expressed genes. The expression ratios and the page with gene/protein/interaction annotations are hyperlinked with nodes. (This figure is available in full color at ftp://ftp.wiley.com/public/sci_tech_med/drug_discovery/.)

The central node APP (Fig. 6*a*) links the upregulation of AP-1, NFkB, TGF-α2, protein kinase (PKC), and ADAM17/TACE pathways with the activation of amyloid precursor protein (APP) gene in glaucomatous ONHAs. The increased APP accumulation is likely to translate into the accumulation of neurotoxic β-amyloid peptide promoted by the c-Jun N-terminal kinase/mitogen activated protein kinase (JNK/MAPK) pathway, caspase3, reactive oxygen species (ROS) and β-amyloid itself. Others have reported similar results to these for reactive astrocytes in the central nervous system (CNS) [189]. APP/β-amyloid induces cell motility via binding to integrins and may become neurotoxic via astrocyte-mediated oxidative stress and mitochondrial dysfunction in neurons [190]. The interaction between two major nodes (Fig. 6*b*) also demonstrates the major mechanisms of iNOS induction in reactive astrocytes where it has been demonstrated to be neurotoxic in vitro and in vivo [177, 191]. The IL-1, NFkB, and tumor necrosis factor (TNF-α) pathways have been shown to play a role in iNOS overexpression in ONH [192]. In addition to a characterized role in complement regulation, clusterin has several debatable functions in the eye and CNS. Overexpression of clusterin is also neurotoxic via an astrocyte-mediated mechanism [193] and is highly upregulated in glaucomatous ONHAs possibly contributing to neurotoxicity either alone or in association with β-amyloid peptide (Fig. 6*c*).

The computational gene network analysis illustrates four proteins that play a key role in switching NFkB, transcriptional factor AP1, amyloid precursor protein APP, and misfolded β-amyloid peptide from downregulated pathways to upregulated pathways. Glaucoma may ultimately result from incorrect folding of APP, triggering a cascade of events to culminate in the loss of supportive and detoxification functions of ONH astroglia before finally leading to oxidative stress and apoptosis of retinal ganglion cells. This mechanism is similar to the cell death of neurons in Alzheimer's disease (AD) and was not previously shown experimentally for ONH astrocytes, as it cannot be deduced from a comparison of the lists of differentially regulated gene clusters. Possible parallels between neuron death in AD and death of ganglion neurons in glaucoma has been recently established in experimental studies [194, 195], as has the role of astrocytes as key mediators of β-amyloid-induced death [190]. The potential role of complement activation and clusterin accumulation have also been suggested from animal studies [193, 196] and is important to note that several of the key interactions established in the network analysis were not previously included in MetaCore.

## Breast Cancer

Breast cancer is a complex disease or a group of underlying disease states that involves many tissues and cell subtypes [197]. Currently, the clinical subtyping of breast cancers is limited to indexing based on pathological staging and tumor grading [198, 199] and is carried out by standardized scoring algorithms [200]. These indexes are neither precise nor individualized, and their broad

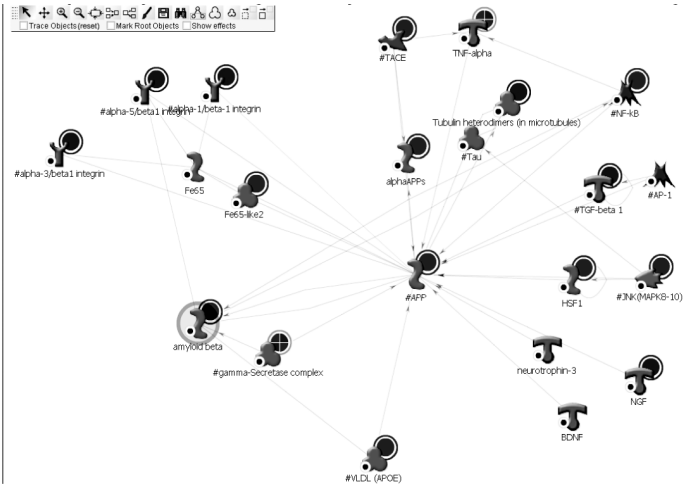
Trace Objects (reset)
Mark Root Objects
Show effects
#TACE
TNF-alpha
#alpha-5/beta1 integrin
#alpha-1/beta-1 integrin
Tubulin heterodimers (in microtubules)
#NF-kB
#alpha-3/beta1 integrin
Fe65
#Tau
alphaAPPs
#TGF-beta 1
#AP-1
Fe65-like2
#APP
HSF1
#JNK(MAPK8-10)
amyloid beta
#gamma-Secretase complex
neurotrophin-3
NGF
BDNF
#VLDL (APOE)

(*a*)

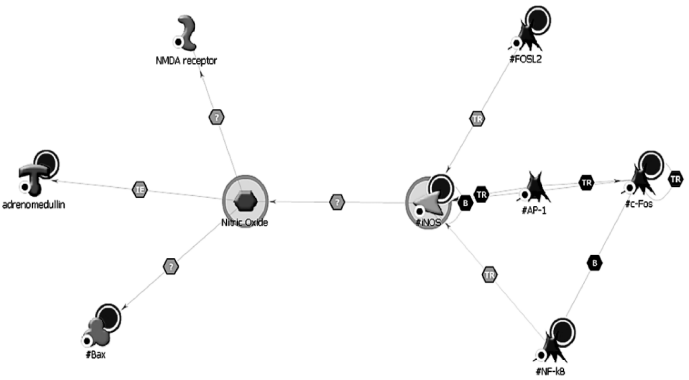
NMDA receptor
#FOSL2
adrenomedullin
Nitric Oxide
#NOS
#AP-1
#c-Fos
#Bax
#NF-kB

(*b*)

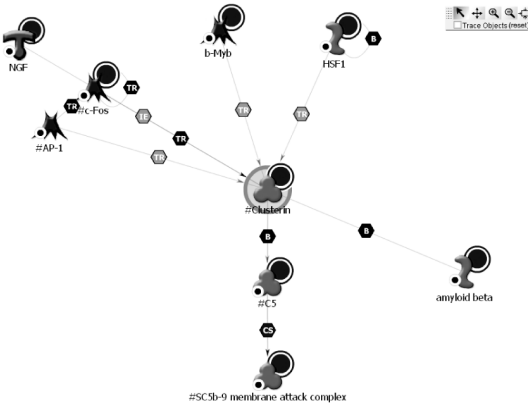
Trace Objects (reset
NGF
b-Myb
HSF1
#c-Fos
#AP-1
#Clusterin
#C5
amyloid beta
#SC5b-9 membrane attack complex

(*c*)

application results in overtreatment of many patients with chemo- and endocrine therapies [70]. Multiple phenotypical factors influence prognosis for breast cancer, including age, histological grade, tumor size, and lymph node and status of estrogen hormone receptor (ER) and *BRCA1* gene expression [201]. No single factor is a sufficient predictor of disease outcome. Global profiling of gene expression has quickly become the method of choice in breast cancer research for subcategorization of breast cancer with respect to ER state [69, 202–204], percentage of cells in the S phase of cell cycle [205], pathological staging and tumor grading [206], and a combination of gene expression/morphology subclasses [68, 207]. Overall, studies (Table 2) have shown microarray expression analysis can separate breast cancers into known and new categories with distinct pathological features (reviewed in [70]). ER-positive and ER-negative tumors can be distinguished by several dozen-strong classifier gene sets [69, 130]. ER-negative tumors can be further subdivided according to the long-term postoperational prognosis [204], and ER-positive tumors according to the patients' ethnic status [201]. More precise subdivision into five groups according to *ERBB2* oncogene overexpression, association with luminal subtypes, and ER overexpression was revealed in a multiyear study [68, 207, 208].

Microarray profiling has also proved useful for predicting development of metastases by detecting the presence of a signature set of differentially expressed genes in primary tumors [69, 130, 209] and for characterization of disease progression and tumor grading [206]. This is based on the hypothesis that cancer samples sharing the same patterns of expression clusters are expected to be similar in terms of pathology, drug sensitivity, and/or other phenotypes. Supervised analysis assumes prior separation of samples into predefined phenotypic classes. The analysis identifies those sets of genes for which expression is the most distinct between the classes and suffers from the flaws described previously [210]. This may explain the high rate of false positives in breast cancer prognosis based on array clustering [70], the exclusion of individual genes with high experimentally determined impact in breast cancer from a set of prognostic marker genes [69, 209], and the inconsistency between classifier gene sets [68]. Although surprisingly no significant difference between distinct pathological stages of the disease were detected in the earlier studies, there were altered expression signatures apparent between different tumor grades [206]. A subset of these genes correlated well with progression from DCIS (ductile carcinoma in situ) to IDC (invasive ductile carcinoma) stages. Microarray analysis has been used for predicting tumor sensitivity to chemotherapy treatment with docetaxel [211], adriamycin, and

◀

**Figure 6** Smaller networks for glaucoma microarray data. (*a*) The APP protein node and its immediate interaction space. (*b*) iNOS cross-activation network in glaucomatous astrocytes. (*c*) Potential role of clusterin in glaucoma pathology of astrocytes. The arrows indicate direction of protein interactions. (This figure is available in full color at ftp://ftp.wiley.com/public/sci_tech_med/drug_discovery/.)

**TABLE 2 Major Gene Expression Studies and Data Sets in Breast Cancer**

| Study | Array | Samples/Cohort | Statistical Analysis | Reference |
|---|---|---|---|---|
| Discrimination of pathological stages | 42,000 cDNAs | 61 IDC, ILC tumors | Hierarchical clustering SAM, PAM | 262 |
| ER status, prognosis, and racial differences | 8,064 cDNAs | 36 primary tumors | Unsupervised PCA | 201 |
| Prediction of ER status, SPF (S phase of cell cycle) | 6,728 cDNAs | 48 primary tumors | Artificial neural networks | 205 |
| Postoperative prognosis<br>For ER(–) cancers | 25,344 cDNAs: NCBI's Unigene | 10 tumors: 5 dead, 5 survived in 5 years | Mann–Whitney test, permutation test | 204 |
| Subcategorization of tumors prognostic signatures | 8,102 cDNAs[a] | 115 malignant tumors | Unsupervised hierarch. clustering, PAM, Kaplan–Meir analysis | 68 |
| Predictors of lymph node Status | 12,000 features U95Av2 (Affy) | 89 primary tumors | Combined stat. tree model: *k*-means clustering, Bayesian classification Tree | 130 |
| Signatures for pathological states, tumor grades | 12,000 cDNAs | 36 tumors of different grades | Linear discriminant analysis | 206 |
| Prediction of therapeutic response to docetaxel | 12,000 features U95Av2 (Aff) | 24 tumors training set | Two-sample *t*-test, global permutation test, BRB | 211 |
| Subclassification, clinical prognosis (metastasis) | 25,000 cDNAs | 117 tumors from <55 years old patients | Unsupervised hierarchical clustering, 3-step supervised Classification | 69 |
| Clinical prognosis based on ER status | 7,129 features HuGeneF1 (Affymetrix) | 49 primary tumors | Binary regression model: SVD, Bayesian analysis | 203 |

[a] In this study, the authors also reanalyzed data from [69, 203].

cyclophosphamide [212], in spite of the very complex overall expression response to these drugs [211]. There are further limitations of these microarray studies as different groups have used unique gene sets on their arrays and no comparison studies or simple tag ID alignments have been performed to date. The selection of patient cohorts, sample size, experimental and analytical techniques all vary greatly between these studies. And as in the case of glaucoma the predictor genes suggested by each study are almost incomparable even when they predict similar outcomes [69, 130].

Not surprisingly, prediction of the disease course including metastasis and recurrence using prognostic signature gene sets is, however, prone to a high rate of false positives. The fundamental van't Veer study [69] had a false-positive rate of 27 percent in 5 years (nonmetastatic patients were incorrectly assigned to the metastatic group). Microarray efficiency in these studies is also questionable with several thousand genes on genomewide arrays displaying statistically significant differential expression versus healthly tissues, but only a few dozen of these genes are in the final sets of classifiers.

We have tested the analysis of computationally derived gene network signatures as a method of elucidating the relationship between gene expression changes induced by a compound and cellular endpoints. The approach is rooted in the idea of modular organization of large-scale networks of biological processes proposed by Lee Hartwell and colleagues [144] in which various types of cellular functionality are provided by relatively small, transient but tightly connected networks of molecules (5 to 25 nodes) engaged in performing specific functions. We analyzed the expression data from a study investigating the effects of 4-hydroxytamoxifen and estrogen treatment in MCF-7 breast cancer cell cultures [213]. After importing the data into MetaCore, filters were applied to retain only those genes that showed the strongest variance in expression (the top 2 percent, or approximately 40 genes) as a response to either treatment. The networks that link these strongly responsive genes to each other were recreated using the direct interactions algorithm allowing only genes from the subset to be present in the network. The result is several small networks of 2 to 15 genes and a handful of nonconnected nodes. The largest subnetworks for 4-hydroxytamoxifen (*a*) and estrogen (*b*) treatments are shown in Figure 7. These networks include mostly cell-cycle-related genes in agreement with the conclusion of hierarchical clustering showing that at the molecular level, both compounds act as agonists of cell proliferation, evoking very similar patterns of gene expression [213]. Unlike clustering, however, these computationally derived gene networks show the absence of cyclin D1 among 4-hydroxytamoxifen-induced genes, resulting in a significant difference in the network topology and overall fingerprint. For estrogen treatment there is a single large network that encompasses all cell-cycle-related genes induced by the treatment while the 4-hydroxytamoxifen network breaks into two smaller clusters [214]. These unique network signatures may be due to 4-hydroxytamoxifen acting as an antagonist of cell proliferation, while estrogen is an agonist [213] with network analysis identifying

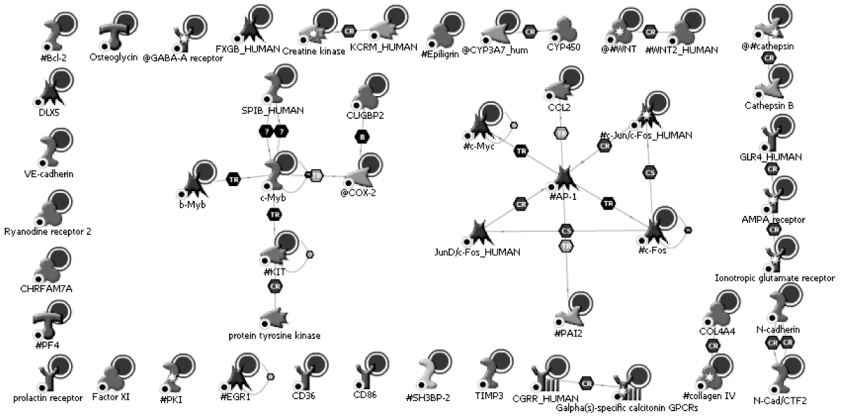


(*a*)

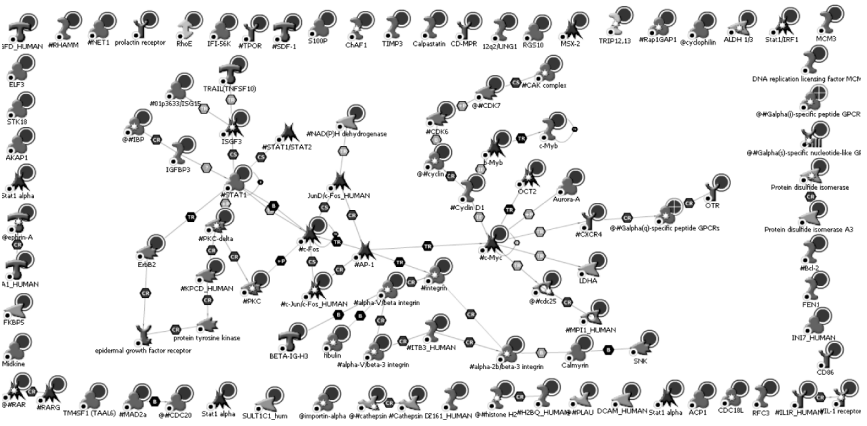


(*b*)

**Figure 7** Clusters of highly responsive genes for the treatment of MC-7 breast cancer cells with (*a*) 4-hydroxytamoxifen and (*b*) estrogen after 24 hr. Both treatments induce similar sets of cell-cycle-related genes, but the topology of the networks is different, correlating with the unique impact of these treatments on cell proliferation. Genes toward the exterior possess fewer interactions than those located centrally. (This figure is available in full color at ftp://ftp.wiley.com/public/sci_tech_med/drug_discovery/.)

cyclin D1 as a responsible element well known as key for the $G_1$/S transition [215, 216].

## Systems ADME/Tox

Understanding or computationally predicting the ADME/Tox properties of molecules earlier can help prevent late-stage clinical failure and has attracted considerable investment for research into developing these methods and

models [217–219]. The majority of xenobiotics undergo metabolism via the cytochrome P450 (CYP) enzymes, which are capable of either inactivating or activating these molecules. To date specific CYP-substrate/inhibitor recognition interactions have been studied extensively and several quantitative structure–activity relationships (QSAR) and pharmacophore models have been built for a limited number of these enzymes in a reductionist manner [220–222]. It is clear the computational models need to be integrated with the technologies described earlier for systems biology [6, 217, 223] including metabolic databases and others containing ADME-associated proteins or pathways such as PharmaGKB [224], the human membrane transporter database [225], the ADME-AP database [226], and the nuclear receptor database [227]. There is considerable complexity that needs to be incorporated such as modeling of a single enzyme's catalysis of one molecule to multiple metabolites [228] that has been shown using commercially available cell-modeling software. This study showed that toxicity of a substrate is likely to be minimized when there are multiple metabolites, and this therefore may represent a biological advantage. To date the complex interconnections where one molecule may be metabolized by different enzymes with different kinetic constants and then the metabolite cascade may follow a unique pathway has not been modeled. In addition the complex kinetics [229] and general protein promiscuity [218] that have been observed in vitro for many of these drug metabolizing enzymes may be an important mechanism for enabling rapid clearance of toxins [230], and this in turn may be understood by using a systems biology approach [231].

Understanding the regulation of these enzymes is also important, with nuclear receptors identified as central to the many proteins involved in drug metabolism and transport. The pregnane X-receptor (PXR) is a transcriptional regulator of CYP3A, human MDR1, CYP2C8/9, CYP2B6, and other genes involved in the transport, metabolism, and biosynthesis of bile acids [232]. A second orphan nuclear receptor, the constitutive androstane receptor (CAR), has approximately 40 percent identity with PXR in the ligand binding domain. CAR accumulates in the nucleus, heterodimerizes with RXR, binds to the two Phenobarbital-responsive elements, and ultimately activates transcription of the CYP2B6 [233] and other genes in a manner similar to how PXR can activate CYP2B6 as well as CYP3A4 [232]. The complex interactions and cross-talk of these nuclear interactions urgently requires computational models that can predict binding to these proteins and enable visualization of connections between different genes and ligands [232, 234].

Global gene expression profiling has a role for evaluation of potential toxicity of drug candidates [235, 236]. Two companies have marketed content databases with patterns of gene expression on whole genome rat arrays in response to drugs and known toxic molecules. The Iconix Drug Matrix product (*http://www.iconixpharm.com/products/products_drugmatrix.html*) represents a collection of gene expression profiles (Amersham rat arrays with 8000 features) from liver of rats treated with about 450 drugs with known toxicologi-

cal properties. GeneLogic used Affymetrix rat arrays in its ToxExpress product (*http://www.genelogic.com/solutions/toxexpress/*). Paradigm Genetics (now known as Icoria) is generating toxicogenomic data in collaboration with NIEHS, which would suggest a large microarray database will be created in the future that may be accessible. The $EDGE^2$ database, a public effort at the University of Wisconsin is striving to create a public database of mouse gene expression profiles after treatment with different toxic molecules [72]. In all of these efforts different clustering techniques are applied to analyze gene expression patterns and extract toxicity signatures.

Published studies using microarrays for toxicology include one group that treated rats with 52 hepatotoxins and used subtractive hybridization to create a final microarray enriched with liver genes, which was then benchmarked with 3 hepatotoxins. This study suggested hundreds of genes were affected by these xenobiotics, some of which had not been previously associated with toxicity of the compounds [237]. For example, phenobarbital treatment in mice affects the expression of 138 genes when assessed using the NIEHS ToxChip microarray, and approximately half of these genes are regulated by CAR [238]. These studies and databases of expression profiles will certainly be a valuable start for understanding the complexities of toxicity.

One product has currently been developed for systems ADME/Tox (MetaDrug, *www.genego.com*) that predicts the major metabolites from an input structure, scores them for the enzymes likely to be involved, and can use microarray data to visualize gene expression of the genes important for toxicology. A comprehensive set of over 3500 substrate–product reactions for 38 human cytochromes and other enzymes has been assembled. This data has been used in the assessment of product–substrate specificity for the CYPs [239], the classification of druglike molecules based on their $K_m$ values for human CYPs [240] and for building models for predicting the rates ($V_{max}$) of N-dealkylation by CYP3A4 and CYP2D6 [241]. This complete combination of predictive and visualization tools with literature data is key to explain metabolism and toxicity for unknown compounds. Ultimately this type of combined approach will enable the prediction of idiosyncratic drug reactions that are often not detected until the drug has been released onto the market [242–244]. Our understanding of some of the factors resulting in drug-induced toxicity has expanded to focus on the molecular mechanisms involved. In particular, there has been considerable focus on hepatotoxicity mediated by drug–drug interactions seen as a major cause of the failure of clinical candidates. We can demonstrate the use of this approach using theophylline, a treatment for asthma. Cimetidine, diltiazem, ciprofloxacin, and enoxacin have all been shown to impair the elimination of theophyline, while phenytoin induces its metabolism. The interactions of all of these ligands with CYPs and other proteins can be visualized with MetaDrug (Fig. 8). Theophyline is shown with connections (as a substrate) to CYP1A1, CYP1A2, CYP1B1, CYP2C9, CYP2D6, CYP2E1, and CYP3A4 [244b, 244c]. The following would inhibit a range of these CYPs: cimetidine (CYP1A2, CYP2C9, CYP2D6, CYP2E1,

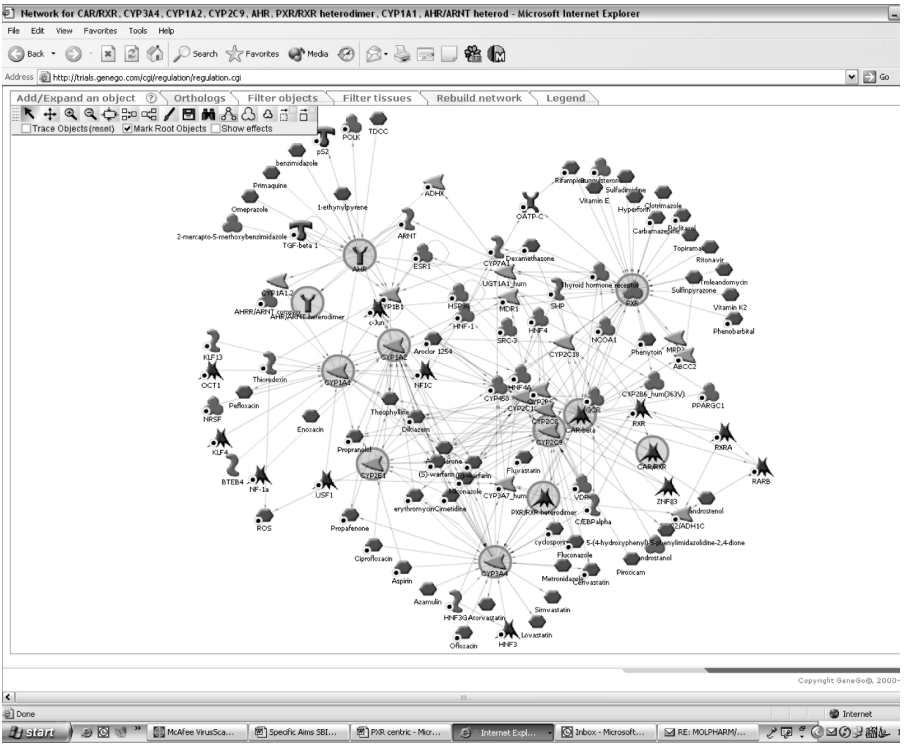


**Figure 8** MetaDrug interaction map of key proteins and ligands linked to theophylline (circled) The network of cell signaling interactions built from the list of genes. This network has been automatically reconstructed, based on the interactions contained in the database. The genes from the original list are encircled. (This figure is available in full color at ftp://ftp.wiley.com/public/sci_tech_med/drug_discovery/.)

CYP3A4 [245]); diltiazem (CYP3A4 [246]); ciprofloxacin (CYP3A4 and CYP1A2 [247]), and enoxacin (CYP1A2 [248]). Phenytoin is a human PXR activator [249], which in turn will increase the levels of CYP3A4 and ultimately increase the clearance of theophylline by this route. Besides showing all these well-known interactions other potential regulatory and signaling pathways can be observed that may be important for further drug interactions with theophylline, which might not have been previously considered. Certainly there is potential for further integration of systems biology approaches into ADME/Tox [250].

## 8 FUTURE DEVELOPMENT OF PATHWAY ANALYSIS SOFTWARE

The merging of currently independent networks for signaling and metabolism and the enrichment of the mammalian network database with interactions

shown for other organisms will be important for the future development of pathway analysis. All currently available genomic and proteomic databases are designed based on molecular entities, such as genes or proteins, as elementary objects [46, 251]. Recently, several public resources started linking their molecular data with the Gene Ontology (GO) "biological process" classification (*www.geneontology.org*) as knowledge of the biological roles of shared proteins is generally transferable across eukaryotes [252]. Such associations are useful in retrieving the subsets of molecular objects related to specific GO categories, but classification in itself does not provide connectivity among processes on the same level of hierarchy. Organizing relevant biological and chemical information around the core entities related to functional processes such as pathways, rather than molecular objects, may be advantageous and will require a novel data model (Fig. 9). Ultimately, a set of flexible many-to-many relationships will link functional and molecular information to more adequately capture the dual nature of biological systems. The physical objects (molecules, ions) and processes (reactions, binding, and transport) are linked as operational units by inputs and outputs exchanging matter and information.

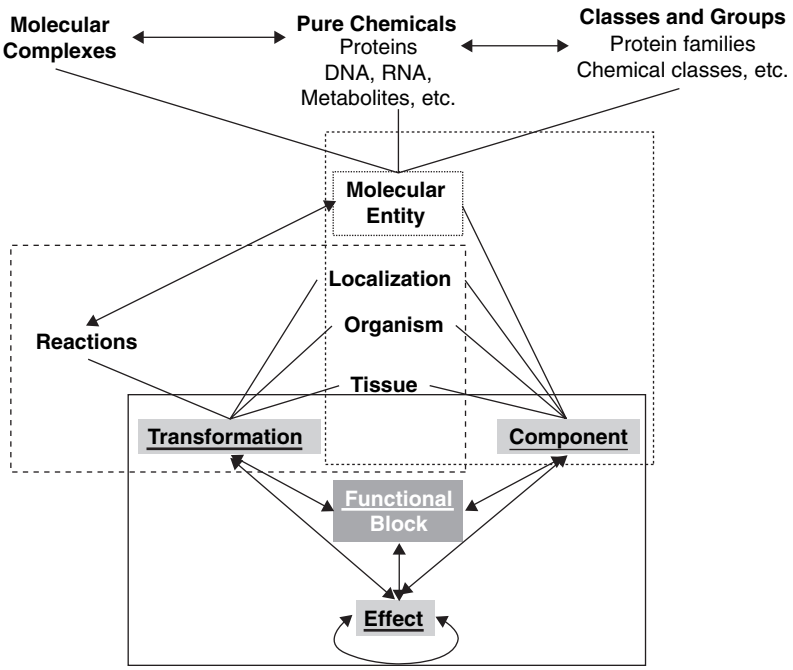


**Figure 9** Proposed future database model. Borders enclose the database entities related to core functional entities: dotted border, entities related to the component; dashed line, entities related to the transformation; solid border, entities related to the functional block.

New algorithms will also be required for analyzing the networks of intracellular processes and the identification of self-consistent functional modules in them. In the living cell, the networks (particularly the signaling ones) are very dynamic and transient as they exist usually for a short time, in particular cellular compartments and under different conditions. Capturing the transient nature of cellular networking may be possible by defining the functional modules and the sets of related subnetworks operating in a concerted way and by developing an automated protocol for reconstruction of pathways and functional blocks involved in diseases and other conditions. This will produce functionally meaningful modules for cell signaling and metabolism. Recently, several systems biology teams developed methods for identification of tightly connected subnetworks as molecular complexes or functional modules [155, 156, 253, 254]. These studies were based on publicly available data on relatively simple yeast and bacterial protein interaction networks. The networks represented undirected interaction graphs with equal weights assigned to all interactions. Moreover, most of the data came from Y2H assays, strongly biased toward physical associations between proteins (such as complex formation), and known for its false-positive rate. The proteins that are related by other means (e.g., enzymes in a metabolic pathway or transient signal transduction interactions) could be missing from such data.

Future network analysis algorithms may also be based on the curated networks of cell signaling and metabolic processes, which contain additional information associated with the nodes and edges of the networks. The changes of expression of two molecules should correspond to the way they are connected in the network in terms of directionality and type of links between molecules. For example, in MetaCore this information describes three types of edges (activation, inhibition, and unspecified), 12 different mechanisms (e.g., phosphorylation, binding, and transcriptional regulation), and types of nodes (e.g., metabolic enzyme, second messenger, ion channel, and transcription factors). The previously developed algorithm [155] could then be used to cluster highly connected nodes found in a large network of protein interactions using an optimization approach. The sets of nodes are selected to maximize the function $Q(m,n) = 2m/[n(n-1)]$, where $m$ is the number of interactions among $n$ nodes. Thus, a cluster of proteins in which every member interacts with every other member will score the highest ($Q = 1$). The Monte Carlo optimization procedure starts with a randomly selected set of connected nodes and proceeds by moving selected nodes along the edges of the graph to maximize $Q$. Previously, this method was found to be effective in recognizing protein complexes and functional modules in the protein interaction network in yeast [155] in which the clusters were scored based on connectivity. Figure 10 shows graphs with the same underlying topology but a different directionality of the edges. In panel (*a*), one can find several directed paths containing 4 nodes. Altogether such paths connect 7 out of 9 nodes of the graph (marked by bold arrows). On panel (*b*) there are no directed paths of length 3. It is clear that graph (*a*) is more likely to represent a functional pathway than graph (*b*). The

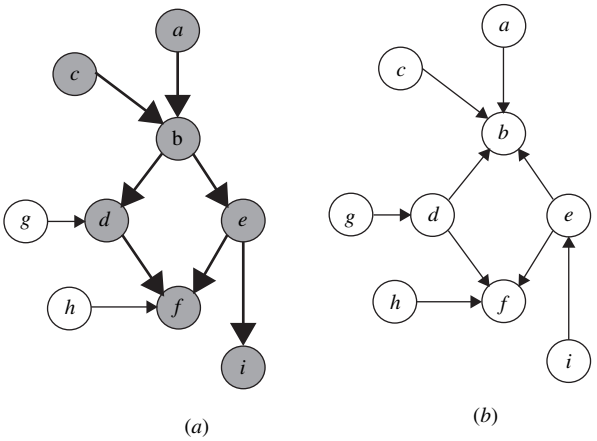


**Figure 10** Schematic to demonstrate two graphs with identical topology but different directionality of edges. Graph (*a*) is more likely to represent a functional module as it contains longer directed paths than graph (*b*).

implementation of this method in a computational algorithm to find subgraphs containing long ordered paths (containing 4 or more nodes) will be valuable for gene network modeling in which the criterion for inclusion of a node will be whether it contributes to a longer directed path or allows new directed paths of sufficient length. For example, if a cluster of nodes *b*, *d*, *e*, *f* in Figure 10*a* were used as a seed, then the addition of node *a* would create a path of length 4. The consecutive inclusion of nodes *c* and *i* would not create longer paths but would create additional paths of length 4. An important way to assess the efficiency of the proposed algorithms is to evaluate the statistical significance of found modules with a random reassignment of the node connections leading to a different distribution of clusters [155]. Tight clusters are highly unlikely ($P < 10^{-5}$) when observing the distribution of clusters in a randomly connected network.

Comparing relatively small functional network modules of 5 to 25 nodes will be computationally tractable and several network comparison algorithms have recently been applied to the analysis of intracellular pathways. Ogata and co-workers [255] have described a heuristic graph comparison algorithm and its application for finding functionally related enzymes clusters (FREC) across 10 species of microorganisms [255]. An extension of this method has also been applied for finding the similarity between protein and gene expression networks [256]. More recently, an algorithm has been described that finds common interaction pathways by global alignment of protein interaction networks from the yeast *Saccharomyces cerevisiae* and a bacterium *Helicobacter pylori* [257]. New techniques that take advantage of annotated human interactions in the databases could be based on this algorithm. The construction of a multiparametric distance function between network modules from

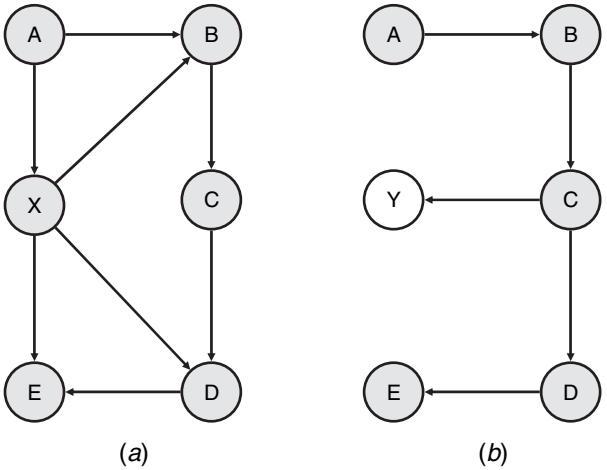


**Figure 11** Schematic to illustrate how two network modules sharing 83 percent molecular composition can have significant topological differences. On network (*a*) the shortest distance between nodes A and E is 2, while on network (*b*) it is 4. The global connectivity of network (*a*) is also higher. (This figure is available in full color at ftp://ftp.wiley.com/public/sci_tech_med/drug_discovery/.)

different samples would be a first step. Then the computation of a matrix of distances among all network modules from different samples would be possible. Once the distances are computed, an appropriate clustering algorithm can be selected to find modules that correlate with the condition of interest. In many cases networks from different samples may have similar but nonidentical compositions but may or may not be substantially different in their topology. For example, the difference in just one node can change network topology as shown in Figure 11. It is important to distinguish between network modules that have similar molecular composition but different topologies. The distance between nodes could also be used such that A and E is 2 in Figure 11*a* and 4 in Figure 11*b*. Other network descriptors may include global connectivity or the connectivity of differentially expressed genes within the module, minimal path lengths connecting them (TopNet, 53), directionality of interactions, or *P* values for the distribution of highly differentiated genes. These approaches all represent unique measures of similarity/dissimilarity between network signatures and could be used to classify them.

## Reconstruction of Condition-Specific Molecular and Functional Networks

The development of programs for the custom selection of functional modules based on the input sets of conditional gene expression data (both mRNA and protein expression) will be another important progression in computational gene network analysis software. Such input-related functional network modules will be relatively small and specific for the particular data set in ques-

tion. In this case the genes or proteins of interest will be connected by the shortest, most direct and functionally most relevant interactions. This approach will rely on the topology of the underlying curated network encoding considerably more information and providing an understanding of affected genes up or downstream. Capturing the gene network will be more useful than lists of genes and will point to potential biomarkers. A final goal would be to use the integrated high-throughput data for the reconstruction of disease-specific molecular and functional networks (described earlier), analyze their topology and behavior, and ultimately rank potential drug targets according to their role in the mechanisms of diseases. This will require testing several different algorithms for reconstructing disease states and other conditions with the likely most promising direction coming from a data-intensive consensus approach. Some data are publicly available from studies with genes that demonstrate a different expression level between normal and disease states (*http://www.broad.mit.edu/resources.html; www.ebi.ac.uk*). This can be used to identify the gene clusters representing the affected functional pathways, and these could be grouped based on the connectivity of the underlying network. For each gene a set of nearest-neighbor genes on the network could be computed before identifying subsets of genes tightly connected either directly to each other or bridged by no more than one gene that is not a member of the original list. For every gene at least one of its nearest neighbors (or the central gene itself) is also a nearest neighbor of at least one other gene from the subset (Fig. 12).

## 9 FUTURE OF SYSTEMS BIOLOGY APPLIED TO DRUG DISCOVERY

We have seen in recent years a move toward generating high-throughput screening data. This has occurred earlier in drug discovery based around molecules of interest, particularly in terms of their physico-chemical properties and their biological activities toward both desirable and undesirable targets in isolation. This data is also being used to derive computational algorithms to predict these properties from molecular structure. Increasingly, higher content biological data is also being generated after cell or animal treatment and the levels of metabolites, genes, and proteins are determined. Combining the reductionist type of approach for a molecule (does it bind or not to a particular protein) with the effect on a whole system (what is the global effect on metabolism, gene expression, and transcription) will be important if we are to understand and predict efficacy and toxicity reliably. The applications of systems biology in drug discovery in the future are likely to use the vast amounts of qualitative and quantitative biological data collated in the various databases and network building algorithms to build predictive signatures for diseases and following treatment of cells or tissues with a myriad of molecules. Such a predictive approach will continue to be developed at the interface of

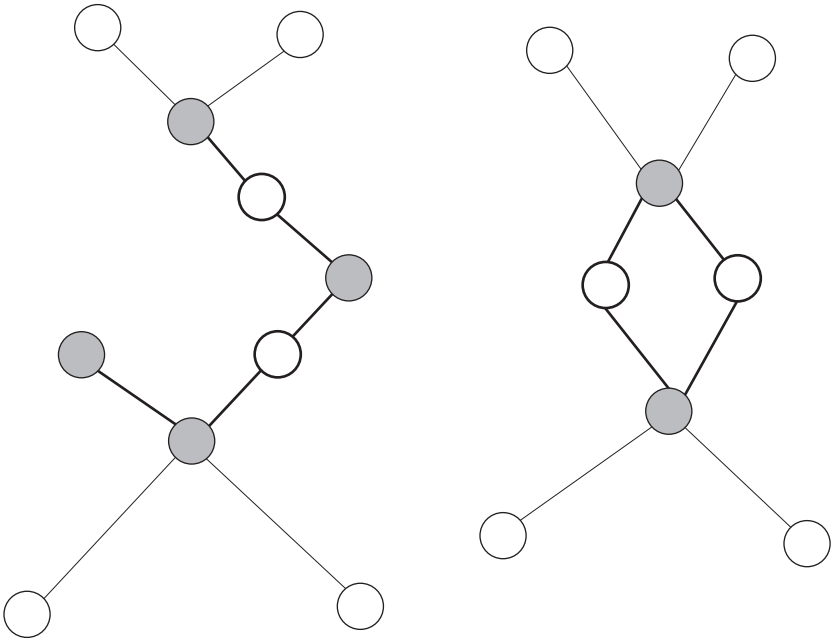

**Figure 12** Schematic to show the identification of tight clusters from differentially expressed genes. Filled circles represent differentially expressed genes from the original set; open circles represent nearest neighbors; bold lines and circles represent genes and edges connecting genes into cluster.

cheminformatics and bioinformatics for application from target selection through clinical data analysis.

***Acknowledgments*** We acknowledge our colleagues at GeneGo, collaborators at Chemical Diversity, Dr. Valery Shestopalov (University of Miami), and NIH grant 1-R43-GM069124–01 "In Silico Assessment of Drug Metabolism and Toxicity."

## REFERENCES

1. Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., Funke, R., Gage, D., Harris, K., Heaford, A., Howland, J., Kann, L., Lehoczky, J., LeVine, R., McEwan, P., McKernan, K., Meldrim, J., Mesirov, J. P., Miranda, C., Morris, W., Naylor, J., Raymond, C., Rosetti, M., Santos, R., Sheridan, A., Sougnez, C., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Wyman, D., Rogers, J., Sulston, J., Ainscough, R., Beck, S., Bentley, D., Burton, J., Clee, C., Carter, N., Coulson, A., Deadman, R., Deloukas, P., Dunham, A., Dunham, I., Durbin, R., French, L., Grafham, D.,

Gregory, S., Hubbard, T., Humphray, S., Hunt, A., Jones, M., Lloyd, C., McMurray, A., Matthews, L., Mercer, S., Milne, S., Mullikin, J. C., Mungall, A., Plumb, R., Ross, M., Shownkeen, R., Sims, S., Waterston, R. H., Wilson, R. K., Hillier, L. W., McPherson, J. D., Marra, M. A., Mardis, E. R., Fulton, L. A., Chinwalla, A. T., Pepin, K. H., Gish, W. R., Chissoe, S. L., Wendl, M. C., Delehaunty, K. D., Miner, T. L., Delehaunty, A., Kramer, J. B., Cook, L. L., Fulton, R. S., Johnson, D. L., Minx, P. J., Clifton, S. W., Hawkins, T., Branscomb, E., Predki, P., Richardson, P., Wenning, S., Slezak, T., Doggett, N., Cheng, J. F., Olsen, A., Lucas, S., Elkin, C., Uberbacher, E., Frazier, M., et al. (2001). Initial sequencing and analysis of the human genome. *Nature*, *409*(6822), 860–921.

2. Venter, J. C., Adams, M. D., Myers, E. W., Li, P. W., Mural, R. J., Sutton, G. G., Smith, H. O., Yandell, M., Evans, C. A., Holt, R. A., Gocayne, J. D., Amanatides, P., Ballew, R. M., Huson, D. H., Wortman, J. R., Zhang, Q., Kodira, C. D., Zheng, X. H., Chen, L., Skupski, M., Subramanian, G., Thomas, P. D., Zhang, J., Gabor Miklos, G. L., Nelson, C., Broder, S., Clark, A. G., Nadeau, J., McKusick, V. A., Zinder, N., Levine, A. J., Roberts, R. J., Simon, M., Slayman, C., Hunkapiller, M., Bolanos, R., Delcher, A., Dew, I., Fasulo, D., Flanigan, M., Florea, L., Halpern, A., Hannenhalli, S., Kravitz, S., Levy, S., Mobarry, C., Reinert, K., Remington, K., Abu-Threideh, J., Beasley, E., Biddick, K., Bonazzi, V., Brandon, R., Cargill, M., Chandramouliswaran, I., Charlab R.., Chaturvedi, K., Deng, Z., Di Francesco, V., Dunn, P., Eilbeck, K., Evangelista, C., Gabrielian, A. E., Gan, W., Ge, W., Gong, F., Gu, Z., Guan, P., Heiman, T. J., Higgins, M. E., Ji, R. R., Ke, Z., Ketchum, K. A., Lai, Z., Lei, Y., Li, Z., Li, J., Liang, Y., Lin, X., Lu, F., Merkulov, G. V., Milshina, N., Moore, H. M., Naik, A. K., Narayan, V. A., Neelam, B., Nusskern, D., Rusch, D. B., Salzberg, S., Shao, W., Shue, B., Sun, J., Wang, Z., Wang, A., Wang, X., Wang, J., Wei, M., Wides, R., Xiao, C., Yan, C., et al. (2001). The sequence of the human genome. *Science*, *291*(5507), 1304–1351.
3. Fiehn, O. (2001). Combining genomics, metabolome analysis, and biochemical modelling to understand metabolic networks. *Compar. Funct. Genom.*, *2*, 155–168.
4. Nicholson, J. K., Wilson, I. D. (2003). Understanding global, systems biology: Metabonomics and the continuum of metabolism. *Natl. Rev. Drug Discov.*, *2*, 668–676.
5. Food and Drug Administration (FDA) (2004). *Innovation Stagnation: Challenge and Opportunity on the Critical Path to New Medicinal Products*, FDA, Washington, DC.
6. Kitano, H. (2002). Computational systems biology. *Nature*, *420*, 206–210.
7. Kitano, H. (2002). Systems biology: A brief overview. *Science*, *295*, 1662–1664.
8. Ideker, T. (2004). Systems biology 101—what you need to know. *Nat. Biotechnol.*, *22*, 473–475.
9. Noble, D., Colatsky, T. J. (2002). A return to rational drug discovery: Computer-based models of cells, organs and systems in drug target identification. *Emerging Ther. Targets, 4*, 39–49.
10. Baxter, D. A., Canavier, C. C., Clark, Jr. J. W., Byrne, J. H. (1999). Computational model for the serotonergic modulation of sensory neorons in Aplysia. *J. Neurophysiol.*, *82*, 2914–2935.

11. Bhalla, U. S. (2002). The chemical organization of signalling interactions. *Bioinformatics*, *18*, 855–863.
12. Kuroda, S., Schwieghofer, N., Kawato, M. (2001). Exploration of signal transduction pathways in cerebellar long-term depression by kinetic simulation. *J. Neurosci.*, *21*, 5693–5702.
13. Csete, M. E., Doyle, J. C. (2002). Reverse engineering of biological complexity. *Science*, *295*, 1664–1669.
14. McAdams, H. H., Arkin, A. P. (2000). Gene regulation: Towards a circuit engineering discipline. *Curr. Biol.*, *10*, R318–R320.
15. Yagil, G. (1976). Isotope exchange in biochemical networks application of an electric circuit technique to a biological reaction system. *J. Theor. Biol.*, *61*, 73–80.
16. Boiteux, A., Busse, H-G. (1989). Circuit analysis of the oscillatory state in glycolysis. *BioSystems*, *22*, 231–240.
17. McAdams, H. H., Shapiro, L. (1995). Circuit simulation of genetic networks. *Science*, *269*, 650–656.
18. Odum, D. T., Zizlsperger, N., Gordon, D. B., Bell, G. W., Rinaldi, N. J., Murray, H. L., Volkert, T. L., Schreiber, J., Rolfe, P. A., Gifford, D. K., Fraenkel, E., Bell, G. I., Young, R. A. (2004). Control of pancreas and liver gene expression by HNF transcription factors. *Science*, *303*, 1378–1381.
19. Koza, J. R., Keane, M. A., Streeter, M. J., Mydlowec, W., Yu, J., Lanza, G. (2003). *Genetic Programming IV: Routine Human-Competitive Machine Intelligence*. Kluwer Academic, Norwell, MA.
20. Wiechert, W. (2002). Modeling and simulation: Tools for metabolic engineering. *J. Biotechnol.*, *94*, 37–63.
21. Covert, M. W., Schilling, C. H., Famili, I., Edwards, J. S., Goryanin, I. I., Selkov, E., Palsson, B. O. (2001). Metabolic modeling of microbial strains in silico. *Trends Biochem. Sci.*, *26*, 179–186.
22. Schaff, J., Fink, C. C., Slepchenko, B., Carson, J. H., Loew, L. M. (1997). A general computational framework for modeling cellular structure and function. *Biophys. J.*, *73*, 1135–1146.
23. Schaff, J., Loew, L. M. The virtual cell. In R. B. Altman (Ed.), *Pacific Symposium on Biocomputing, Mauna Lani, Hawaii, 1999*, pp. 228–239.
24. Fink, C. C., Slepchenko, B., Moraru, I. I., Schaff, J., Watras, J., Loew, L. M. (1999). Morpholical control of inositol-1,4,5-trisphosphate-dependent signals. *J. Cell Biol. 147*, 929–935.
25. Tomita, M., Hashimoto, K., Takahashi, K., Shimizu, T. S., Matsuzaki, Y., Miyoshi, F., Saito, K., Tanida, S., Yugi, K., Ventner, J. C., Hutchison III, C. A. (1999). E-CELL: Software environment for whole-cell simulation. *Bioinformatics*, *15*, 72–84.
26. Noble, D. (2002). Modeling the heart—from genes to cells to the whole organ. *Science*, *295*, 1678–1682.
27. Musante, C. J., Lewis, A. K., Hall, K. (2002). Small- and large-scale biosimulation applied to drug discovery and development. *Drug Discov. Today*, *7*, S192–S196.
28. Bumol, T. F., Watanabe, A. M. (2001). Genetic information, genomic technologies, and the future of drug discovery. *JAMA*, *285*, 551–555.

29. Zhu, X., Raina, A. K., Lee, H. G., Casadesus, G., Smith, M. A., Perry, G. (2004). Oxidative stress signalling in Alzheimer's disease. *Brain Res.*, *1000*, 32–39.

30. Larkin, J. E., Frank, B. C., Gaspard, R. M., Duka, I., Gavras, H., Quakenbush, J. (2004). Cardiac transcriptional response to acute and chronic angiotensin II treatments. *Physiol. Genom*, *18*, 152–166.

31. Gerrard, J. A., Sparrow, A. D., Wells, J. A. (2001). Metabolic databases—what next? *Trends Biochem. Sci.*, *26*, 137–140.

32. Vasquez, A., Flammini, A., Maritan, A., Vespignani, A. (2003). Global protein function prediction from protein-protein interaction networks. *Nat. Biotechnol.*, *21*, 697–700.

33. Schomburg, I., Chang, A., Schomburg, D., (2002) BRENDA, enzyme data and metabolic information. *Nucleic Acids Res.*, *30*, 47–49.

34. Schomburg, I., Chang, A., Hofmann, O., Ebeling, C., Ehrentreich, F., Schomburg, D. (2002). BRENDA: A resource for the enzyme data and metabolic information. *Trends Biochem Sci.*, *27*, 54–56.

35. Takai-Igarashi, T., Kaminuma, T. (1999). A pathway finding system for the cell signaling networks database. *In Silico Biol.*, *1*, 129–146.

36. Xenarios, I., Rice, D. W., Salwinski, L., Baron, M. K., Markotte, E. M., Eisenberg, D. (2000). DIP: The database of interacting proteins. *Nucleic Acids Res.*, *28*, 289–291.

37. Selkov, E., Basmanova, S., Gaasterland, T., Goryanin, I., Gretchin, Y., Maltsev, N., Nenashev, V., Overbeek, R., Panyushkina, E., Pronevitch, L., Selkov, E. J., Yunis, I. (1996). The metabolic pathway collection from EMP: The enzymes and metabolic pathways database. *Nucleic Acids Res.*, *24*, 26–28.

38. Dahlquist, K. D., Salomonis, N., Vranizan, K., Lawlor, S. C., Conklin, B. R., (2002). GenMAPP, a new tool for viewing and analyzing microarray data on biological pathways. *Nat. Genet.*, *31*, 19–20.

39. Peri, S., Navarro, J. D., Kristiansen, T. Z., Amanchy, R., Surendranath, V., Muthusamy, B., Gandhi, T. K., Chandrika, K. N., Deshpande, N., Suresh, S., Rashmi, B. P., Shanker, K., Padma, N., Niranjan, V., Harsha, H. C., Talreja, N., Vrushabendra, B. M., Ramya, M. A., Yatish, A. J., Joy, M., Shivashankar, H. N., Kavitha, M. P., Menezes, M., Choudhury, D. R., Ghosh, N., Saravana, R., Chandran, S., Mohan, S., Jonnalagadda, C. K., Prasad, C. K., Kumar-Sinha, C., Deshpande, K. S., Pandey, A. (2004). Human protein reference database as a discovery resource for proteomics. *Nucleic Acids Res.*, *32*(Database issue), D497–501.

40. Peri, S., Navarro, J. D., Amanchy, R., Kristiansen, T. Z., Jonnalagadda, C. K., Surendranath, V., Niranjan, V., Muthusamy, B., Gandhi, T. K., Gronborg, M., Ibarrola, N., Deshpande, N., Shanker, K., Shivashankar, H. N., Rashmi, B. P., Ramya, M. A., Zhao, Z., Chandrika, K. N., Padma, N., Harsha, H. C., Yatish, A. J., Kavitha, M. P., Menezes, M., Choudhury, D. R., Suresh, S., Ghosh, N., Saravana, R., Chandran, S., Krishna, S., Joy, M., Anand, S. K., Madavan, V., Joseph, A., Wong, G. W., Schiemann, W. P., Constantinescu, S. N., Huang, L., Khosravi-Far, R., Steen, H., Tewari, M., Ghaffari, S., Blobe, G. C., Dang, C. V., Garcia, J. G., Pevsner, J., Jensen, O. N., Roepstorff, P., Deshpande, K. S., Chinnaiyan, A. M., Hamosh, A., Chakravarti, A., Pandey, A. (2003). Development of human protein reference

database as an initial platform for approaching systems biology in humans. *Genome Res.*, *13*(10), 2363–2371.

41. Karp, P. D., Riley, M., Saier, M., Paulsen, I. T., Collado-Vides, J., Paley, S. M., Pellegrini-Toole, A., Bonavides, C., Gama-Castro, S. (2002). The EcoCyc Database. *Nucleic Acids Res.*, *30*, 56–58.
42. Kanehisa, M. A., Goto, S., Kawashima, S., Nakaya, A. (2002). The KEGG databases at GenomeNet. *Nucleic Acids Res.*, *30*, 42–46.
43. Selkov, E. J., Grechkin, Y., Mikhailova, N., Selkov, E. (1998). MPW: The metabolic pathways database. *Nucleic Acids Res.*, *26*, 43–45.
44. Smith, C. M., Shindyalov, I. N., Veretnik, S., Gribskov, M., Taylor, S. S., Ten Eyck, L. F., Bourne, P. E. (1997). The protein kinase resource. *TIBS*, *22*, 444–446.
45. Maltsev, N., Marland, E., Yu, G. X., Bhatnagar, S., Lusk, R. (2002). Sentra, a database of signal transduction proteins. *Nucleic Acids Res.*, *30*, 349–350.
46. Bairoch, A., Apweiler, R. (2000). The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. *Nucleic Acids Res.*, *28*, 45–48.
47. Boeckmann, B., Bairoch, A., Apweiler, R., Blatter M-C., Estreicher, A., Gasteiger, E., Martin, M. J., Michoud, K., O'Donovan, C., Phan, I., Pilbout, S., Schneider, M. (2003). The SWISS-PROT protein knowledge base and its supplement TrEMBL in 2003. *Nucleic Acids Res.*, *31*, 365–370.
48. Schacherer, F., Choi, C., Gotze, U., Krull, M., Pistor, S., Wingender, E. (2001). The TRANSPATH signal transduction database: A knowledge base on signal transduction. *Bioinformatics*, *17*, 1053–1057.
49. Wheeler, D. L., Church, D. M., Federhen, S., Lash, A. E., Madden, T. L., Pontius, J. U., Schuler, G. D., Schriml, L. M., Sequeria, E., Tatusova, T. A., Wagner, L. (2003). Database resources of the national center for biotechnology. *Nucleic Acids Res.*, *31*, 28–33.
50. Ellis, L. B. M., Hou, B. K., Kang, W., Wackett, L. P. (2003). The University of Minnesota Biocatalysis/Biodegradation Database: Post-genomic data mining. *Nucleic Acids Res.*, *31*, 262–265.
51. Ellis, L. B. M., Speedie, S. M., Mcleish, R. (1998). Representing metabolic pathway information : An object orientated approach. *Bioinformatics*, *14*, 803–809.
52. Overbeek, R., Larsen, N., Pusch, G. D., D'Souza, M., Selkov, E. J., Kyrpides, N., Fonstein, M., Maltsev, N., Selkov, E. (2000). WIT: Integrated system for high-throughput genome sequence analysis and metabolic reconstruction. *Nucleic Acids Res.*, *28*, 123–125.
53. Ideker, T., Thorsson, V., Ranish, J. A., Christmas, R., Buhler, J., Eng, J. K., Bumgarner, R., Goodlett, D. R., Aebersold, R., Hood, L. (2001). Integrated genomic and proteomic analyses of a systematically perturbed network. *Science*, *292*, 929–934.
54. Davidson, E. H., Rast, J. P., Oliveri, P., Ransick, A., Calestani, C., Yuh C-H., Minokawa, T., Amore, G., Hinman, V., Arenas-Mena, C., Otim, O., Brown, C. T., Livi, C. B., Lee, P. Y., Revilla, R., Rust, A. G., Pan, Z. J., Schilstra, M. J., Clarke, P. J. C., Arnone, M. I., Rowen, L., Cameron, R. A., McClay, D. R., Hood, L.,

Bolouri, H. (2002). A genomic regulatory network for development. *Science*, *295*, 1669–1678.

55. Arita, M. (2003). In silico atomic tracing by substrate-product relationships in *Escherichia coli* intermediary metabolism. *Genome Res.*, *13*, 2455–2466.
56. Botstein, D., White, R. L., Skolnick, M., Davis, R. W. (1980). Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am. J. Hum. Genet.*, *32*(3), 314–331.
57. Elahi, E., Kumm, J., Ronaghi, M. (2004). Global genetic analysis. *J. Biochem. Mol. Biol.*, *37*(1), 11–27.
58. Botstein, D., Risch, N. (2003). Discovering genotypes underlying human phenotypes: Past successes for mendelian disease, future approaches for complex disease. *Nat. Genet.*, *33* (Suppl.), 228–237.
59. Patil, N., Berno, A. J., Hinds, D. A., Barrett, W. A., Doshi, J. M., Hacker, C. R., Kautzer, C. R., Lee, D. H., Marjoribanks, C., McDonough, D. P., Nguyen, B. T., Norris, M. C., Sheehan, J. B., Shen, N., Stern, D., Stokowski, R. P., Thomas, D. J., Trulson, M. O., Vyas, K. R., Frazer, K. A., Fodor, S. P., Cox, D. R. (2001). Blocks of limited haplotype diversity revealed by high-resolution scanning of human chromosome 21. *Science*, *294*(5547), 1719–1723.
60. Marnellos, G. (2003). High-throughput SNP analysis for genetic association studies. *Curr. Opin. Drug Discov. Devel.*, *6*(3), 317–321.
61. Jurinke, C., van den Boom, D., Cantor, C. R., Koster, H. (2002). Automated genotyping using the DNA MassArray technology. *Methods Mol. Biol.*, *187*, 179–192.
62. Ronaghi, M., Pettersson, B., Uhlen, M., Nyren, P. (1998). PCR-introduced loop structure as primer in DNA sequencing. *Biotechniques*, *25*(5), 876–878, 880–882, 884.
63. Oliphant, A., Barker, D. L., Stuelpnagel, J. R., Chee, M. S. (2002). BeadArray technology: Enabling an accurate, cost-effective approach to high-throughput genotyping. *Biotechniques*, *56–58* (Suppl.), 60–61.
64. Watson, J. D., Crick, F. H. (1953). Genetical implications of the structure of deoxyribonucleic acid. *Nature*, *171*(4361), 964–967.
65. Watson, J. D., Crick, F. H. (1953). Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature*, *171*(4356), 737–738.
66. Butte, A. (2002). The use and analysis of microarray data. *Nat. Rev. Drug Discov.*, *1*(12), 951–960.
67. Kuo, W. P., Jenssen, T. K., Butte, A. J., Ohno-Machado, L., Kohane, I. S. (2002). Analysis of matched mRNA measurements from two different microarray technologies. *Bioinformatics*, *18*(3), 405–412.
68. Sørlie, T., Tibshirani, R., Parker, J., Hastie, T., Marron, J. S., Nobel, A., Deng, S., Johnsen, H., Pesich, R., Geisler, S., Demeter, S., Perov, C. M., Lonning, P. E., Brown, P. O., Borresen-Dale, A. L., Botstein, D. (2003). Repeated observation of breast tumor subtypes in independent gene expression data sets. *Proc. Natl. Acad. Sci. USA*, *100*, 8418–8423.
69. van 't Veer, L. J., Dai, H., van de Vijer, M. J., He, Y. D., van de Vijver, M. J., He, Y. D., Vant Veer, L. J., Dai, H., Hart, A. A., Voskoil, D. W., Schreiber, G. J., Peterse, J. L., Roberts, C., Marton, M. J., Parrish, M., Atsma, D., Nitteveen, A., Glas, A., Delahaye, L., van der Nelde, T., Bartelink, H., Rodenhuis, S., Rutgers, E. T.,

Freind, S. H., Bernards, R. (2002). Gene expression profiling predicts clinical outcome of breast cancer. *Nature*, *415*, 530–536.

70. Cleator, S., Ashworth, A. (2004). Molecular profiling of breast cancer: Clinical implications. *Br. J. Cancer. Res.*, *90*, 1120–1124.
71. Cardoso, F. (2003). Microarray technology and its effect on breast cancer (re)classification and prediction of outcome. *Breast Cancer Res.*, *5*(6), 303–304.
72. Thomas, R. S., Rank, D. R., Penn, S. G., Zastrow, G. M., Hayes, K. R., Pande, K., Glover, E., Silander, T., Craven, M. W., Reddy, J. K., Jovanovich, S. B., Bradfield, C. A. (2001). Identification of toxicologically predictive gene sets using cDNA microarrays. *Mol. Pharmacol.*, *60*, 1189–1194.
73. Waring, J. F., Jolly, R. A., Ciurlionis, R., Lum, P. Y., Praestgaard, J. T., Morfitt, D. C., Buratto, B., Roberts, C., Schadt, E., Ulrich, R. G. (2001). Clustering of hepatotoxins based on mechanism of toxicity using gene expression profiles. *Toxicol. Appl. Pharmacol.*, *175*, 28–42.
74. Gerhold, D., Lu, M., Xu, J., Austin, C., Caskey, C. T., Rushmore, T. (2001). Monitoring expression of genes involved in drug metabolism and toxicology using DNA microarrays. *Physiol. Genomics*, *5*, 161–170.
75. Coen, M., Ruepp, S. U., Lindon, J. C., Nicholson, J. K., Pognan, F., Lenz, E. M., Wilson, I. D. (2004). Integrated application of transcriptomics and metabonomics yields new insight into the toxicity due to paracetamol in the mouse. *J. Pharm. Biomed. Anal.*, *35*(1), 93–105.
76. Perou, C. M. (2001). Show me the data! *Nat. Genet.*, *29*(4), 373.
77. Velculescu, V. E., Zhang, L., Vogelstein, B., Kinzler, K. W. (1995). Serial analysis of gene expression. *Science*, *270*(5235), 484–487.
78. Yamamoto, M., Wakatsuki, T., Hada, A., Ryo, A. (2001). Use of serial analysis of gene expression (SAGE) technology. *J. Immunol. Methods*, *250*(1–2), 45–66.
79. Ryo, A., Kondoh, N., Wakatsuki, T., Hada, A., Yamamoto, N., Yamamoto, M. (2000). A modified serial analysis of gene expression that generates longer sequence tags by nonpalindromic cohesive linker ligation. *Anal. Biochem.*, *277*(1), 160–162.
80. Stein, W. D., Litman, T., Fojo, T., Bates, S. E. (2004). A Serial Analysis of Gene Expression (SAGE) database analysis of chemosensitivity: Comparing solid tumors with cell lines and comparing solid tumors from different tissue origins. *Cancer Res.*, *64*(8), 2805–2816.
81. Nacht, M., Ferguson, A. T., Zhang, W., Petroziello, J. M., Cook, B. P., Gao, Y. H., Maguire, S., Riley, D., Coppola, G., Landes, G. M., Madden, S. L., Sukumar, S. (1999). Combining serial analysis of gene expression and array technologies to identify genes differentially expressed in breast cancer. *Cancer Res.*, *59*(21), 5464–5670.
82. Wilkins, M. R., Sanchez, J. C., Gooley, A. A., Appel, R. D., Humphery-Smith, I., Hochstrasser, D. F., Williams, K. L. (1996). Progress with proteome projects: Why all proteins expressed by a genome should be identified and how to do it. *Biotechnol. Genet. Eng. Rev.*, *13*, 19–50.
83. Anderson, N. L., Anderson, N. G. (1998). Proteome and proteomics: New technologies, new concepts, and new words. *Electrophoresis*, *19*(11), 1853–1861.

84. Klein, J. B., Thongboonkerd, V. (2004). Overview of proteomics. *Contrib. Nephrol.*, *141*, 1–10.
85. O'Farrell, P. H. (1975). High resolution two-dimensional electrophoresis of proteins. *J. Biol. Chem.*, *250*(10), 4007–4721.
86. Klose, J. (1975). Protein mapping by combined isoelectric focusing and electrophoresis of mouse tissues. A novel approach to testing for induced point mutations in mammals. *Humangenetik*, *26*(3), 231–243.
87. Fountoulakis, M. (2000). Two-dimensional electrophoresis. In I. Wilson, C. Poole, and M. Cooke (Eds.), *Encyclopedia of Separation Science, II/Electrophoresis*, Academic, London, pp. 1356–1363.
88. Aebersold, R., Mann, M. (2003). Mass spectrometry-based proteomics. *Nature*, *422*(6928), 198–207.
89. Lahm, H. W., Langen, H. (2000). Mass spectrometry: A tool for the identification of proteins separated by gels. *Electrophoresis*, *21*, 2105–2114.
90. Henzel, W. J., Billeci, T. M., Stults, J. T., Wong, S. C., Grimley, C., Watanabe, C. (1993). Identifying proteins from two-dimensional gels by molecular mass searching of peptide fragments in protein sequence databases. *Proc. Natl. Acad. Sci. USA*, *90*(11), 5011–5015.
91. Mann, M., Wilm, M. (1994). Error-tolerant identification of peptides in sequence databases by peptide sequence tags. *Anal. Chem.*, *66*(24), 4390–4399.
92. Gorg, A., Obermaier, C., Boguth, G., Harder, A., Scheibe, B., Wildgruber, R., Weiss, W. (2000). The current state of two-dimensional electrophoresis with immobilized pH gradients. *Electrophoresis*, *21*(6), 1037–1053.
93. Zuo, X., Speicher, D. W. (2000). A method for global analysis of complex proteomes using sample prefractionation by solution isoelectrofocusing prior to two-dimensional electrophoresis. *Anal. Biochem.*, *284*(2), 266–278.
94. Hanash, S. (2003). Disease proteomics. *Nature*, *422*(6928), 226–232.
95. Zheng, P. P., Kros, J. M., Sillevis-Smitt, P. A., Luider, T. M. (2003). Proteomics in primary brain tumors. *Front Biosci.*, *8*, d451–d463.
96. Fountoulakis, M. (2004). Application of proteomics technologies in the investigation of the brain. *Mass. Spectrom. Rev.*, *23*(4), 231–258.
97. Rigaut, G., Shevchenko, A., Rutz, B., Wilm, M., Mann, M., Seraphin, B. (1999). A generic protein purification method for protein complex characterization and proteome exploration. *Nat. Biotechnol.*, *17*(10), 1030–1032.
98. Nelson, R. W., Nedelkov, D., Tubbs, K. A. (2000). Biosensor chip mass spectrometry: A chip-based proteomics approach. *Electrophoresis*, *21*(6), 1155–1163.
99. Gavin, A. C., Bosche, M., Krause, R., Grandi, P., Marzioch, M., Bauer, A., Schultz, J., Rick, J. M., Michon, A. M., Cruciat, C. M., Remor, M., Hofert, C., Schelder, M., Brajenovic, M., Ruffner, H., Merino, A., Klein, K., Hudak, M., Dickson, D., Rudi, T., Gnau, V., Bauch, A., Bastuck, S., Huhse, B., Leutwein, C., Heurtier, M. A., Copley, R. R., Edelmann, A., Querfurth, E., Rybin, V., Drewes, G., Raida, M., Bouwmeester, T., Bork, P., Seraphin, B., Kuster, B., Neubauer, G., Superti-Furga, G. (2002). Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature*, *415*(6868), 141–147.
100. Ho, Y., Gruhler, A., Heilbut, A., Bader, G. D., Moore, L., Adams, S. L., Millar, A., Taylor, P., Bennett, K., Boutilier, K., Yang, L., Wolting, C., Donaldson, I., Schan-

dorff, S., Shewnarane, J., Vo, M., Taggart, J., Goudreault, M., Muskat, B., Alfarano, C., Dewar, D., Lin, Z., Michalickova, K., Willems, A. R., Sassi, H., Nielsen, P. A., Rasmussen, K. J., Andersen, J. R., Johansen, L. E., Hansen, L. H., Jespersen, H., Podtelejnikov, A., Nielsen, E., Crawford, J., Poulsen, V., Sorensen, B. D., Matthiesen, J., Hendrickson, R. C., Gleeson, F., Pawson, T., Moran, M. F., Durocher, D., Mann, M., Hogue, C. W., Figeys, D., Tyers, M. (2002). Systematic identification of protein complexes in *Saccharomyces cerevisiae* by mass spectrometry. *Nature*, *415*(6868), 180–183.

101. Fields, S., Song, O. (1989). A novel genetic system to detect protein-protein interactions. *Nature*, *340*(6230), 245–246.

102. Cho, S., Park, S. G., Lee do, H., Park, B. C. (2004). Protein-protein interaction networks: From interactions to networks. *J. Biochem. Mol. Biol.*, *37*(1), 45–52.

103. Yang, M., Wu, Z., Fields, S. (1995). Protein-peptide interactions analyzed with yeast two-hybrid system. *Nucleic Acids Res.*, *23*, 1152–1156.

104. Ito, T., Chiba, T., Ozawa, R., Yoshida, M., Hattori, M., Sakaki, Y. (2001). A comprehensive two-hybrid analysis to explore the yeast protein interactome. *Proc. Natl. Acad. Sci. USA*, *98*(8), 4569–4574.

105. Giot, L., Bader, J. S., Brouwer, C., Chaudhuri, A., Kuang, B., Li, Y., Hao, Y. L., Ooi, C. E., Godwin, B., Vitols, E., Vijayadamodar, G., Pochart, P., Machineni, H., Welsh, M., Kong, Y., Zerhusen, B., Malcolm, R., Varrone, Z., Collis, A., Minto, M., Burgess, S., McDaniel, L., Stimpson, E., Spriggs, F., Williams, J., Neurath, K., Ioime, N., Agee, M., Voss, E., Furtak, K., Renzulli, R., Aanensen, N., Carrolla, S., Bickelhaupt, E., Lazovatsky, Y., DaSilva, A., Zhong, J., Stanyon, C. A., Finley, Jr., R. L., White, K. P., Braverman, M., Jarvie, T., Gold, S., Leach, M., Knight, J., Shimkets, R. A., McKenna, M. P., Chant, J., Rothberg, J. M. (2003). A protein interaction map of *Drosophila melanogaster*. *Science*, *302*(5651), 1727–1736.

106. Li, S., Armstrong, C. M., Bertin, N., Ge, H., Milstein, S., Boxem, M., Vidalain, P. O., Han, J. D., Chesneau, A., Hao, T., Goldberg, D. S., Li, N., Martinez, M., Rual, J. F., Lamesch, P., Xu, L., Tewari, M., Wong, S. L., Zhang, L. V., Berriz, G. F., Jacotot, L., Vaglio, P., Reboul, J., Hirozane-Kishikawa, T., Li, Q., Gabel, H. W., Elewa, A., Baumgartner, B., Rose, D. J., Yu, H., Bosak, S., Sequerra, R., Fraser, A., Mango, S. E., Saxton, W. M., Strome, S., Van Den Heuvel, S., Piano, F., Vandenhaute, J., Sardet, C., Gerstein, M., Doucette-Stamm, L., Gunsalus, K. C., Harper, J. W., Cusick, M. E., Roth, F. P., Hill, D. E., Vidal, M. (2004). A map of the interactome network of the metazoan *C. elegans*. *Science*, *303*(5657), 540–543.

107. Mrowka, R., Patzak, A., Herzel, H. (2001). Is there a bias in proteome research? *Genome Res.*, *11*(12), 1971–1973.

108. Jansen, R., Greenbaum, D., Gerstein, M. (2002). Relating whole-genome expression data with protein-protein interactions. *Genome Res.*, *12*(1), 37–46.

109. Lopez, M. F., Pluskal, M. G. (2003). Protein micro- and macroarrays: Digitizing the proteome. *J. Chromatogr.*, *787*, 19–27.

110. Haab, B. B. (2001). Advances in protein microarray technology for protein expression and interaction profiling. *Curr. Opin. Drug Discov. Devel.*, *4*(1), 116–123.

111. Borrebaeck, C. A., Ekstrom, S., Hager, A. C., Nilsson, J., Laurell, T., Marko-Varga, G. (2001). Protein chips based on recombinant antibody fragments: A highly

sensitive approach as detected by mass spectrometry. *Biotechniques*, *30*(5), 1126–1130, 1132.

112. Li, M. (2000). Applications of display technology in protein analysis. *Nat. Biotechnol.*, *18*, 1251–1256.
113. Jenkins, R. E., Pennington, S. R. (2001). Arrays for protein expression profiling: Towards a viable alternative to two-dimensional gel electrophoresis? *Proteomics*, *1*, 13–29.
114. Newman, J. R., Keating, A. E. (2003). Comprehensive identification of human bZIP interactions with coiled-coil arrays. *Science*, *300*(5628), 2097–2101.
115. Zhu, H., Bilgin, M., Bangham, R., Hall, D., Casamayor, A., Bertone, P., Lan, N., Jansen, R., Bidlingmaier, S., Houfek, T., Mitchell, T., Miller, P., Dean, R. A., Gerstein, M., Snyder, M. (2001). Global analysis of protein activities using proteome chips. *Science*, *293*(5537), 2101–2105.
116. Glassbrook, N., Beecher, C., Ryals, J. (2000). Metabolic profiling on the right path. *Nat. Biotechnol.*, *18*(11), 1142–1143.
117. Hall, R., Beale, M., Fiehn, O., Hardy, N., Sumner, L., Bino, R. (2002). Plant metabolomics: The missing link in functional genomics strategies. *Plant Cell*, *14*(7), 1437–1440.
118. Tolstikov, V. V., Fiehn, O. (2002). Analysis of highly polar compounds of plant origin: Combination of hydrophilic interaction chromatography and electrospray ion trap mass spectrometry. *Anal. Biochem.*, *301*(2), 298–307.
119. Stitt, M., Fernie, A. R. (2003). From measurements of metabolites to metabolomics: An "on the fly" perspective illustrated by recent studies of carbon-nitrogen interactions. *Curr. Opin. Biotechnol.*, *14*(2), 136–144.
120. Nicholson, J. K., Lindon, J. C., Holmes, E. (1999). "Metabonomics": Understanding the metabolic responses of living systems to pathophysiological stimuli via multivariate statistical analysis of biological NMR spectroscopic data. *Xenobiotica*, *29*(11), 1181–1189.
121. Lindon, J. C., Nicholson, J. K., Holmes, E., Everette, J. R. (2000). Metabonomics: Metabolic processing studied by NMR spectroscopy of biofluids. *Concepts Magn. Reson.*, *12*, 289–320.
122. ter Kuile, B. H., Westerhoff, H. V. (2001). Transcriptome meets metabolome: Hierarchical and metabolic regulation of the glycolytic pathway. *FEBS Lett.*, *500*(3), 169–171.
123. Goodacre, R., Vaidyanathan, S., Dunn, W. B., Harrigan, G. G., Kell, D. B. (2004). Metabolomics by numbers: Acquiring and understanding global metabolite data. *Trends Biotechnol.*, *22*(5), 245–252.
124. Weckwerth, W., Fiehn, O. (2002). Can we discover novel pathways using metabolomic analysis? *Curr. Opin. Biotechnol.*, *13*(2), 156–160.
125. Barabasi, A-L., Oltvai, Z. N. (2004). Network biology: Understanding the cell's functional organization. *Nature Rev. Genet.*, *5*, 101–113.
126. Gifford, D. K. (2001). Blazing pathways through genetic mountains. *Science*, *293*, 2049–2051.
127. Eisen, M. B., Spellman, P. T., Brown, P. O., Botstein, D. (1998). Cluster analysis and display of genome-wide expression patterns. *Proc. Natl. Acad. Sci. USA*, *95*, 14863–14868.

128. Tavazoie, S., Hughes, J. D., Campbell, M. J., Cho, R. J., Church, G. M. (1999). Systematic determination of genetic network architecture. *Nat. Genet.*, *22*, 281–286.
129. Tamayo, P., Slonim, D., Mesirov, J., Zhu, Q., Kitareewan, S., Dmitrovsky, E., Lander, E. S., Golub, T. R. (1999). Interpreting patterns of gene expression with self-organizing maps: Methods and application to hematopoietic differentiation. *Proc. Natl. Acad. Sci. USA*, *96*, 2907–2912.
130. Huang, E., Cheng, S. H., Dressman, H., Pittman, J., Tsou, M. H., Horng, C. F., Bild, A., Iversen, E. S., Liao, M., West, M., Nevins, J. R., Huan, A. T. (2003). Gene expression predictors of breast cancer outcomes. *Lancet*, *361*, 1590–1596.
131. Kohonen, T. (1997). *Self Organizing Maps*. Springer, Berlin.
132. Quackenbush, J. (2001). Computational analysis of microarray data. *Nat. Rev. Genet.*, *2*, 418–427.
133. Jeong, H., Tombor, B., Albert, R., Oltvai, Z. N., Barabasi, A. L. (2000). The large-scale organization of metabolic networks. *Nature*, *407*(6804), 651–654.
134. Jenssen, T. K., Laegreid, A., Komorowski, J., Hovog, E. (2001). A literature network of human genes for high-throughput analysis of gene expression. *Nat. Genet.*, *28*, 21–28.
135. Chaussabel, D., Sher, A. (2002). Mining microarray expression data by literature profiling. *Genome Biol.*, *3*, 0055.1–0055.16.
136. Raychaudhuri, S., Chang, J. T., Imam, F., Altman, R. B. (2003). The computational analysis of scientific literature to define and recognize gene expression clusters. *Nucleic Acids Res.*, *31*, 4553–4560.
137. Yandell, M. D., Majoros, W. H. (2002). Genomics and natural language processing. *Nat. Rev. Genet.*, *3*, 601–611.
138. Blaschke, K., Valencia, A. (2001). Can bibliographical pointers for known biological data be found automatically? Protein interactions as a case study. *Compar. Funct. Genomics*, *2*, 196–206.
139. Daraselia, N., Yuryev, A., Egorov, S., Novihkova, S., Nikitin, A., Mazo, I. (2003). Extracting human protein interactions from Medline using a full-sentence parser. *Bioinformatics*, *19*.
140. Elbeck, K., Brass, A., Paton, N., Hodgman, C. (1999). INTERACT: An object-oriented protein-protein interactions database. *Proc. Int. Conf. Intel. Syst. Mol. Biol.*, 87–94.
141. Herrgard, M. J., Covert, M. W., Palsson, B. O. (2003). Reconciling gene expression data with known genome-scale regulatory network structures. *Genome Res.*, *13*, 2423–2434.
142. Stuart, J. M., Segal, E., Koller, D., Kim, S. K. (2003). A gene-coexpression network for global discovery of conserved genetic modules. *Science*, *302*, 249–255.
143. Grigoryev, D. N., Ma, S-F., Irizarry, R. A., Ye, S. Q., Quakenbush, J., Garcia, J. G. N. (2004). Orthologous gene-expression profiling in multi-species models: Search for candidate genes. *Genome Biol.*, *5*, R34.1–R34.13.
144. Hartwell, L. H., Hopfield, J. J., Leibler, S., Murray, A. W. (1999). From molecular to modular cell biology. *Nature*, *402*, C47–C52.
145. Wyrick, J. J., Aparico, J. G., Chen, T., Barnett, J. D., Jennings, E. G., Young, R. A., Bell, S. P., Aparicio, O. M. (2001). Genome-wide distribution of ORC and MCM

proteins in *S. cerevisiae*: High-resolution mapping of replication origins. *Science*, *294*, 2357–2360.

146. Wu, L. F., Hughes, T. R., Davierwala, A. P., Robinson, M. D., Stoughton, R., Altschuler, S. J. (2002). Large-scale prediction of *Saccharomyces cerevisiae* gene function using overlapping transcriptional clusters. *Nat. Genet.*, *31*, 255–265.
147. Bar-Joseph, Z., Gerber, G., Simon, I., Gifford, D. K., Jaakola, T. S. (2003). Comparing the continuous representation of time-series expression profiles to identify differentially expressed genes. *Proc. Natl. Acad. Sci. USA*, *100*, 10146–10151.
148. Segal, E., Shapira, M., Regev, A., Pe'er, D., Botstein, D., Koller, D., Freidman, N. (2003). Module networks: Identifying regulatory modules and their condition-specific regulators from gene expression data. *Nat. Genet.*, *34*, 166–176.
149. Matys, V., Fricke, E., Geffers, R., Gossling, E., Haubrock, M., Hehl, R., et al. (2003). TRANSFAC®: Transcriptional regulation, from patterns to profiles. *NAR*, *1*, 374–378.
150. Ihmels, J., Friedlander, G., Bergmann, S., Sarig, O., Ziv, Y., Barkai, N. (2002). Revealing modular organization in the yeast transcriptional network. *Nat. Genet.*, *31*, 370–377.
151. van Noort, V., Snel, B., Huynen, M. A. (2003). Predicting gene function by conserved co-expression. *Trends Genet.*, *19*, 238–242.
152. Michnik, S. W. (2004). Proteomics in living cells. *Drug Discov. Today*, *9*, 262–267.
153. Wagner, A., Fell, D. A. (2001). The small world inside large metabolic networks. *Proc. R. Soc. Lond. B Biol. Sci.*, *268*(1478), 1803–1810.
154. Albert, R., Jeong, H., Barabasi, A. L. (2000). Error and attack tolerance of complex networks. *Nature*, *406*(6794), 378–382.
155. Spirin, V., Mirny, L. A. (2003). Protein complexes and functional modules in molecular networks. *Proc. Natl. Acad. Sci. USA*, *100*, 12123–12128.
156. Rives, A. W., Galitski, T. (2003). Modular organization of cellular networks. *Proc. Natl. Acad. Sci. USA*, *100*, 1128–1133.
157. Pereira-Leal, J. B., Enright, A. J., Ouzounis, C. A. (2004). Detection of functional modules from protein interaction networks. *Proteins*, *54*(1), 49–57.
158. Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D., Alon, U. (2002). Network motifs: Simple building blocks of complex networks. *Science*, *298*, 824–827.
159. Shen-Orr, S. S., Milo, R., Mangan, S., Alon, U. (2002). Network motifs in the transcriptional regulation network of *Escherichia coli*. *Nat. Genet.*, *31*(1), 64–68.
160. Dobrin, R., Beg, Q. K., Barabasi, A. L., Oltvai, Z. N. (2004). Aggregation of topological motifs in the *Escherichia coli* transcriptional regulatory network. *BMC Bioinformatics*, *5*(1), 10.
161. Wuchty, S., Oltvai, Z. N., Barabasi, A. L. (2003). Evolutionary conservation of motif constituents in the yeast protein interaction network. *Nat. Genet.*, *35*(2), 176–179.
162. Selkov, E., Maltsev, N., Olsen, G. J., Overbeek, R., Whitman, W. B. (1997). A reconstruction of the metabolism of *Methanococcus jannaschii* from sequence data. *Gene*, *197*, GC11–26.

163. Xia, X-Q., Wise, M. J. (2003). DiMSim: A discrete-event simulator of metabolic networks. *J. Chem. Inf. Comput. Sci.*, *43*, 1011–1019.

164. Nikitin, A., Egorov, S., Daraselia, N., Mazo, I. (2003). Pathway studio—the analysis and navigation of molecular networks. *Bioinformatics*, *19*, 2155–2157.

165. Vickers, J. C. (1997). The cellular mechanism underlying neuronal degeneration in glaucoma: Parallels with Alzheimer's disease. *Aust. N. Z. J. Opthalmol.*, *25*, 105–109.

166. Janciauskiene, S., Krakau, T. (2001). Alzheimer's peptide: A possible link between glaucoma, exfoliation syndrome and Alzheimer's disease. *Acta Ophthalmol. Scand.*, *79*, 328–329.

167. Bayer, A. U., Ferrari, F. (2002). Severe progression of glaucomatous optic neuropathy in patients with Alzheimer's disease. *Eye*, *16*, 209–212.

168. Parisi, V. (2003). Correlation between morphological and functional retinal impairment in patients affected by ocular hypertension, glaucoma, demyelinating optic neuritis and Alzheimer's disease. *Semin. Opthalmol.*, *18*, 50–57.

169. Quigley, H. A. (1998). Selectivity in glaucoma injury. *Arch. Opthalmol.*, *116*, 396–398.

170. Wax, M. B., Tezel, G. (2002). Neurobiology of glaucomatous optic neuropathy: Diverse cellular events in neurodegeneration and neuroprotection. *Mol. Neurobiol.*, *26*, 45–55.

171. Osborne, N. N., Chidlow, G., Wood, J., Casson, R. (2003). Some current ideas on the pathogenesis and the role of neuroprotection in glaucomatous optic neuropathy. *Eur. J. Opthalmol.*, *13*, S19–S26.

172. Zhao, D. Y., Cioffi, G. A. (2000). Anterior optic nerve microvascular changes in human glaucomatous optic neuropathy. *Eye*, *14*, 445–449.

173. Anderson, D. R. (1999). Introductory comments on blood flow autoregulation in the optic nerve head and vascular risk factors in glaucoma. *Surv. Opthalmol.*, *43* (Suppl. 1), S5–S9.

174. Drance, S., Anderson, D. R., Schulzer, M. (2001). Collaborative Normal-Tension Glaucoma Study Group. Risk factors for progression of visual field abnormalities in normal-tension glaucoma. *Am. J. Ophthalmol.*, *131*, 699–708.

175. Gasser, P., Flammer, J. (1991). Blood-cell velocity in the nailfold capillaries of patients with normal-tension and high-tension glaucoma. *Am. J. Ophthalmol.*, *15*, 585–588.

176. Dreyer, E. B., Zurakowski, D., Schumer, R. A., Podos, S. M., Lipton, S. A. (1996). Elevated glutamate levels in the vitreous body of humans and monkeys with glaucoma. *Arch. Opthalmol.*, *114*, 299–305.

177. Yuan, L., Neufeld, A. H. (2000). Tumor necrosis factor-alpha: A potentially neurodestructive cytokine produced by glia in the human glaucomatous optic nerve head. *Glia*, *32*, 42–50.

178. Morgan, J., Caprioli, J., Koseki, Y. (1999). Nitric oxide mediates excitotoxic and anoxic damage in rat retinal ganglion cells cocultured with astroglia. *Arch. Opthalmol.*, *117*, 1524–1529.

179. Hernandez, M. R., Agapova, O. A., Yang, P., Salvador-Silva, M., Rocard, C. S., Aoi, S. (2002). Differential gene expression in astrocytes from human normal

and glaucomatous optic nerve head analyzed by cDNA microarray. *Glia*, *38*, 45–64.

180. Miyahara, T., Kikuchi, T., Akimoto, M., Kurokawa, T., Shibuki, H., Yoshimura, N. (2003). Gene microarray analysis of experimental glaucomatous retina from cynomologous monkey. *Invest. Ophthalmol. Vis. Sci.*, *44*, 4347–4356.
181. Ahmed, F., Brown, K. M., Stephan, D. A., Morrison, J. C., Johnson, E. C., Tomarev, S. I. (2004). Microarray analysis of changes in mRNA levels in the rat retina after experimental elevation of intraocular pressure. *Invest. Ophthalmol. Vis. Sci.*, *45*, 1247–1258.
182. Lo, W. R., Rowlette, L. L., Caballero, M., Yang, P., Hernandez, M. R., Borras, T. (2003). Tissue differential microarray analysis of dexamethasone induction reveals potential mechanisms of steroid glaucoma. *Invest. Ophthalmol. Vis. Sci.*, *44*, 473–485.
183. Ishibashi, T., Nakazawa, M., Ono, H., Satoh, N., Gojobori, T., Fujiwara, S. (2003). Microarray analysis of embryonic retinoic acid target genes in the ascidian *Ciona intestinalis*. *Dev. Growth Differ.*, *45*, 249–259.
184. Yang, P., Agapova, O., Parker, A., Shannon, W., Pecen, P., Duncan, J., Salvador-Silva, M., Hernandez, M. R. (2004). DNA microarray analysis of gene expression in human optic nerve head astrocytes in response to hydrostatic pressure. *Physiol. Genomics*, *27*, 157–169.
185. Hernandez, M. R. (2000). The optic nerve head in glaucoma: Role of astrocytes in tissue remodeling. *Prog. Retin. Eye Res.*, *19*, 297–321.
186. Swanson, R. A., Ying, W., Kauppinen, T. M. (2004). Astrocyte influences on ischemic neuronal death. *Curr. Mol. Med.*, *4*, 193–205.
187. Shaulian, E., Karin, M. (2002). AP-1 as a regulator of cell life and death. *Cell Biol.*, *4*, E131–E136.
188. Neufeld, A. H., Liu, B. (2003). Glaucomatous optic neuropathy: When glia misbehave. *Neuroscientist*, *9*, 485–495.
189. Abraham, C. R. (2001). Reactive astrocytes and alpha1-antichymotrypsin in Alzheimer's disease. *Neurobiol. Aging*, *22*, 931–936.
190. Abramov, A. Y., Canevari, L., Duchen, M. R. (2003). Changes in intracellular calcium and glutathione in astrocytes as the primary mechanism of amyloid neurotoxicity. *J. Neurosci.*, *23*, 5088–5095.
191. Liu, B., Neufeld, A. H. (2003). Activation of epidermal growth factor receptor signals induction of nitric oxide synthase-2 in human optic nerve head astrocytes in glaucomatous optic neuropathy. *Neurobiol. Dis.*, *13*, 109–123.
192. Yan, Z., Chen, Z., Chen, Z. (2000). Modulation of nitric oxide synthase isoenzymes in reperfused skeletal muscle. *Chin. J. Traumatol.*, *3*, 76–80.
193. Han, B. H., DeMattos, R. B., Dugan, L. L., Kim-Han, J. S., Brendza, R. P., Fryer, J. D., Kierson, M., Cirrito, J., Quick, K., Harmony, J. A., Aronow, B. J., Holtzman, D. M. (2001). Clusterin contributes to caspase-3-independent brain injury following neonatal hypoxia-ischemia. *Nat. Med.*, *7*, 338–343.
194. McKinnon, S. J. (2003). Glaucoma: Ocular Alzheimer's disease? *Front Biosci.*, *8*, S1140–S1156.
195. McKinnon, S. J., Lehman, D. M., Tahzib, N. G., Ransom, N. L., Reitsamer, H. A., Liston, P., LaCasse, E., Li, Q., Korneluk, R. G., Hauswirth, W. W. (2002). Bac-

uloviral IAP repeat-containing-4 protects optic nerve axons in a rat glaucoma model. *Mol. Ther.*, *5*, 780–787.

196. Singhrao, S. K., Neal, J. W., Rushmere, N. K., Morgan, B. P., Gasque, P. (2000). Spontaneous classical pathway activation and deficiency of membrane regulators render human neurons susceptible to complement lysis. *Am. J. Pathol.*, *157*, 905–918.
197. Jeffrey, S. S., Fero, M. J., Borresen-Dale, A. L., Botstein, D. (2002). Expression array technology in the diagnosis and treatment of breast cancer. *Mol. Interv.*, *2*, 101–109.
198. Allred, D. C., Mohsin, S. K., Fuqua, S. W. A. (2001). Histological and biological evolution of human premalignant breast disease. *Endocrinol. Relat. Cancer*, *8*, 47–61.
199. Dalton, L. W., Pinder, S. E., Elston, C. E., Ellis, I. O., Page, D. L., Dupont, W. D., Blamey, R. W. (2000). Histologic grading of breast cancer: Linkage of patient outcome with level of pathologist agreement. *Mod. Pathol.*, *13*, 730–735.
200. Eifel, P., Axelson, J. A., Costa, J., Crowley, J., et al. (2001). NIH Consensus Development Conference Statement: Adjuvant therapy for breast cancer. *J. Natl. Cancer Inst.*, *93*, 979–989.
201. Selaru, F. M., Yin, J., Olaru, A., Mori, Y., Xu, Y., Epstein, S. H., Sato, F., Deacu, E., Wang, S., Sterian, A., Fulton, A., Abraham, J. M., Shibata, D., Baquet, C., Stass, S. A., Meltzer, S. J. (2004). An unsupervised approach to identify molecular phenotypic components influencing breast cancer features. *Cancer Res.*, *64*, 1584–1588.
202. Gruvberger, S., Ringner, M., Chen, Y., Panavally, H., Saal, L. H., Borg, A., Ferno, M., Peterson, C., Meltzer, P. S. (2001). Estrogen receptor status in breast cancer is associated with remarkably distinct gene expression patterns. *Cancer Res.*, *61*, 5979–5984.
203. West, M., Blanchette, C., Dressman, H., Huang, E., Ishida, S., Spang, R., Zuzan, H., Olson, J. A., Jr., Marks, J. R., Nevins, J. R. (2001). Predicting the clinical status of human breast cancer by using gene expression profiles. *Proc. Natl. Acad. Sci. USA*, *98*, 11462–11467.
204. Nagahata, T., Onda, M., Emi, M., Nagai, H., Tsumagari, K., Fujimoto, T., Hirano, A., Sato, T., Nishikawa, K., Akiyama, F., Sakamoto, G., Kasumi, F., Miki, Y., Tanaka, T., Tsunoda, T. (2004). Expression profiling to predict postoperative prognosis for estrogen receptor-negative breast cancers by analysis of 25,344 genes on a cDNA microarray. *Cancer Sci.*, *95*, 218–225.
205. Gruvberger-Saal, S. K., Eden, P., Ringner, M., Badletorp, B., et al. (2004). Predicting continuous values of prognostic markers in breast cancer from microarray gene expression profiles. *Mol. Cancer Ther.*, *3*, 161–168.
206. Ma, X-J., Salunga, R., Tuggle, J. T., Gaudet, J., et al. (2003). Gene expression profiles of human breast cancer progression. *Proc. Natl. Acad. Sci. USA*, *100*, 5974–5979.
207. Sørlie, T., Perou, C. M., Tibshirani, R., Aas, T. (2001). Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc. Natl. Acad. Sci. USA*, *98*, 10869–10874.

208. Perou, C. M., Sørlie, T., Eisen, M. B., et al. (2000). Molecular portraits of human breast cancers. *Nature*, *406*, 747–752.

209. van de Vijer, M. J., He, Y. D., van 'T Veer, L. J., Dai, H., et al. (2002). Gene-expression signature as a predictor of survival in breast cancer. *N. Engl. J. Med.*, *347*, 1999–2009.

210. Schultze, A., Downwards, J. (2001). Navigating gene expression using microarrays—a technology review. *Nat. Cell Biol.*, *3*, E190–E195.

211. Chang, J. C., Wooten, E. C., Tsimelzon, A., Hilsenbeck, S. G., et al. (2003). Gene expression profiling for the prediction of therapeutic response to docetaxel in patients with breast cancer. *Lancet*, *362*, 362–369.

212. Sotiriou, C., Powles, T. J., Dowsett, M., Jazaeri, A. A., et al. (2002). Gene expression profiles derived from fine needle aspiration correlate with response to systemic chemotherapy in breast cancer. *Breast Cancer Res.*, *4*, R5.

213. Hodges, L. C., Cook, J. D., Lobenhofer, E. K., Li, L., Bennett, L., Bushel, P. R., Aldaz, C. M., Afshari, C. A., Walker, C. L. (2003). Tamoxifen functions as a molecular agonist inducing cell cycle-associated genes in breast cancer cells. *Mol. Cancer Res.*, *1*, 300–311.

214. Nikolsky, Y., Ekins, S., Nikolskaya, T., Bugrim, A. (2005). A novel method for generation of signature networks as biomarkers from complex high throughout data. *Tox. Lett.* (in press).

215. Prall, O. W., Rogan, E. M., Musgrove, E. A., Watts, C. K., Sutherland, R. L. (1998). c-Myc or cyclin D1 mimics estrogen effects on cyclin E-Cdk2 activation and cell cycle reentry. *Mol. Cell Biol.*, *18*, 4499–4508.

216. Musgrove, E. A., Sarcevic, B., Sutherland, R. L. (1996). Inducible expression of cyclin D1 in T-47D human breast cancer cell is sufficient for Cdk2 activation and pRB hyperphosphorylation. *J. Cell Biochem.*, *60*, 363–378.

217. Ekins, S., Boulanger, B., Swaan, P. W., Hupcey, M., A. Z. (2002). Towards a new age of virtual ADME/TOX and multidimensional drug discovery. *J. Comput. Aided Mol. Des.*, *16*, 381–401.

218. Ekins, S. (2004). Predicting undesirable drug interactions with promiscuous proteins in silico. *Drug Discov. Today*, *9*, 276–285.

219. Greene, N. (2002). Computer systems for the prediction of toxicity: An update. *Adv. Drug Del. Rev.*, *54*, 417–431.

220. Ekins, S., de Groot, M., Jones, J. P. (2001). Pharmacophore and three dimensional quantitative structure activity relationship methods for modeling cytochrome P450 active sites. *Drug Metab. Dispos.*, *29*, 936–944.

221. Jones, J. P., Mysinger, M., Korzekwa, K. R. (2002). Computational models for cytochrome P450: A predictive electronic model for aromatic oxidation and hydrogen abstraction. *Drug Metab. Dispos.*, *30*, 7–12.

222. Jones, J. P., Korzekwa, K. R. (1996). *Predicting the Rates and Regioselectivity of Reactions Mediated by the P450 Superfamily*, Vol. 272. Academic New York.

223. Bugrim, A., Nikolskaya, T., Nikolsky, Y. (2004). Early prediction of drug metabolism and toxicity: Systems biology approach and modeling. *Drug Discov. Today*, *9*, 127–135.

224. Oliver, D. E., Rubin, D. L., Stuart, J. M., Hewett, M., Klein, T. E., Altman, R. B. (2002). Ontology development for a pharmacogenetics knowledge base. In *Pac Symp. Biocomp.*, pp. 88–99.
225. Yan, Q., Sadee, W. (2000). Human membrane transporter database: A web-accessible relational database for drug transport studies and pharmacogenomics. *AAPS Pharmsci.*, *2*, E20.
226. Sun, L. Z., Ji, Z. L., Chen, X., Wang, J. F., Chen, Y. Z. (2002). ADME-AP: A datbase of ADME associated proteins. *Bioinformatics*, *18*, 1699–1700.
227. Nakata, K., Yukawa, M., Komiyama, N., Nakano, T., Kaminuma, T. (2002). A nuclear receptor database that maps pathways to diseases. *Genome Infomatics*, *13*, 515–516.
228. Cook, D. L., Atkins, W. M. (1997). Enhanced detoxication due to distributive catalysis and toxic thresholds: A kinetic analysis. *Biochemistry*, *36*, 10802–10806.
229. Ekins, S., Stresser, D. M., Williams, J. A. (2003). In vitro and pharmacophore insights into CYP3A enzymes. *Trends Pharmacol. Sci.*, *24*, 191–196.
230. Ekins, S., Ring, B. J., Binkley, S. N., Hall, S. D., Wrighton, S. A. (1998). Autoactivation and activation of cytochrome P450s. *Int. J. Clin. Pharmacol. Therapeut.*, *36*, 642–651.
231. Atkins, W. M. (2004). Implications of the allosteric kinetics of cytochrome P450s. *Drug Discov. Today*, *9*, 478–484.
232. Ekins, S., Mirny, L., Schuetz, E. G. (2002). A ligand-based approach to understanding selectivity of nuclear hormone receptors PXR, CAR, FXR, LXRa and LXRb. *Pharm. Res.*, *19*, 1788–1800.
233. Sueyoshi, T., Kawamato, T., Zelko, I., Honkakoski, P., Negishi, M. (1999). The repressed nuclear receptor CAR responds to phenobarbital in activating the human CYP2B6 gene. *J. Biol. Chem.*, *274*, 6043–6046.
234. Mankowski, D. C., Ekins, S. (2003). Prediction of human drug metabolizing enzyme induction. *Curr. Drug Metab.*, *4*, 381–391.
235. Marton, M. J., DeRisi, J. L., Bennet, H. A., Iyer, V. R., Meyer, M. R., Roberts, C. J., Stoughton, R., Burchard, J., Slade, D., Dai, H., Bassett, D. E. J., Hartwell, L. H., Brown, P. O., Friend, S. H. (1998). Drug target validation and identification of secondary drug target effects using DNA microarrays. *Nat. Med.*, *4*, 1293–1301.
236. Gerhold, D., Jensen, R. V., Gullans, S. R. (2002). Better therapeutics through microarrays. *Nat. Genet.*, *32*, 547–552.
237. Waring, J. F., Cavet, G., Jolly, R. A., McDowell, J., Dai, H., Ciurlionis, R., Zhang, C., Stoughton, R., Lum, P. Y., Ferguson, A., Roberts, C. J. (2003). Development of a DNA Microarray for toxicology based on hepatoxin-regulated sequences. *Env. Health Perspect.*, *111*, 863–870.
238. Ueda, A., Hamadeh, H. K., Webb, H. K., Yamamoto, Y., Sueyoshi, T., Afshari, C. A., Lehmann, J. M., Negishi, M. (2002). Diverse roles of the nuclear orphan receptor CAR in regulating hepatic genes in response to phenobarbital. *Mol. Pharmacol.*, *61*, 1–6.
239. Korolev, D., Balakin, K. V., Nikolsky, Y., Kirillov, E., Ivanenkov, Y. A., Savchuk, N. P., Ivashchenko, A. A., Nikolskaya, T. (2003). Modeling of human cytochrome p450-mediated drug metabolism using unsupervised machine learning approach. *J. Med. Chem.*, *46*(17), 3631–3643.

240. Balakin, K. V., Ekins, S., Bugrim, A., Ivanenkov, Y. A., Korolev, D., Nikolsky, Y., Skorenko, S. A., Ivashchenko, A. A., Savchuk, N. P., Nikolskaya, T. (2004). Computational Kohonen self organizing maps for prediction of binding to human cytochrome P450 3A4. *Drug. Metab. Dispos*, *32*, 1183–1189.

241. Balakin, K. V., Ekins, S., Bugrim, A., Ivanenkov, Y. A., Korolev, D., Nikolsky, Y., Ivashchenko, A. A., Savchuk, N. P., Nikolskaya, T. (2004). Quantitative structure-metabolism relationship of the metabolic N-dealkylation rates. *Drug. Metab. Dispos*, *32*, 1111–1120.

242. Uetrecht, J. (2003). Screening for the potential of a drug candidate to cause idiosyncratic drug reactions. *Drug Discov. Today*, *8*, 832–837.

243. Williams, D. P., Park, B. K. (2003). Idiosyncratic toxicity: The role of toxicophores and bioactivation. *Drug Discov. Today*, *8*, 1044–1050.

244a. Li, A. P. (2002). A review of the common properties of drugs with idiosyncratic hepatotoxicity and the "multiple determinant hypothesis" for the manifestation of idiosyncratic drug toxicity. *Chem. Biol. Interact.*, *142*, 7–23.

244b. Ha, H. R., Chen, J., Freiburghaus, A. U., Follath, F. (1995). Metabolism of theophylline by cDNA-expressed human cytochromes P-450. *Br. J. Clin. Pharmacol.*, *39*, 321–326.

244c. Tjia, J. F., Colbert, J., Back, D. J. (1996). Theophylline metabolism in human liver microsomes: inhibition studies. *J. Pharmacol. Exp. Ther.*, *276*, 912–917.

245. Furuta, S., Kamada, E., Suzuki, T., Sugimoto, T., Kawabata, Y., Shinozaki, Y., Sano, H. (2001). Inhibition of drug metabolism in human liver microsomes by nizatidine, cimetidine and omeprazole. *Xenobiotica*, *31*(1), 1–10.

246. Sutton, D., Butler, A. M., Nadin, L., Murray, M. (1997). Role of CYP3A4 in human hepatic diltiazem N-demethylation: Inhibition of CYP3A4 activity by oxidized diltiazem metabolites. *J. Pharmacol. Exp. Ther.*, *282*(1), 294–300.

247. Loi, C. M., Parker, B. M., Cusack, B. J., Vestal, R. E. (1997). Aging and drug interactions. III. Individual and combined effects of cimetidine and cimetidine and ciprofloxacin on theophylline metabolism in healthy male and female nonsmokers. *J. Pharmacol. Exp. Ther.*, *280*, 627–637.

248. Sorgel. F., Kinzig, M. (1993). Pharmacokinetics of gyrase inhibitors, part 2: Renal and hepatic elimination pathways and drug interactions. *Am. J. Med. 94*(3A), 56S–69S.

249. Luo, G., Cunningham, M., Kim, S., Burn, T., Lin, J., Sinz, M., Hamilton, G., Rizzo, C., Jolley, S., Gilbert, D., Downey, A., Mudra, D., Graham, R., Carroll, K., Xie, J., Madan, A., Parkinson, A., Christ, D., Selling, B., LeCluyse, E., Gan, L. S. (2002). CYP3A4 induction by drugs: Correlation between a pregnane X receptor reporter gene assay and CYP3A4 expression in human hepatocytes. *Drug Metab. Dispos.*, *30*(7), 795–804.

250. Ekins, S., Nikolsky, Y., Nikolskaya, T. (2005). Application of systems biology to absorption, distribution, metabolism, excretion and toxicity. *Trends Pharm. Sci.* (in press).

251. Benson, D. A., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J., Rapp, B. A., Wheeler, D. L. (2002). GenBank. *Nucleic Acids Res.*, *30*, 17–20.

252. Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T., Harris, M. A., Hill, D. P., Issel-

Tarver, L., Kasarskis, A., Lewis, S., Matese, J. C., Richardson, J. E., Ringwald, M., Rubin, G. M., Sherlock, G. (2000). Gene Ontology: Tool for the unification of biology. *Nat. Genet.*, *25*, 25–29.

253. Snel, B., Bork, P., Huynen, M. A. (2002). The identification of functional modules from the genomic association of genes. *Proc. Natl. Acad. Sci. USA*, *99*, 5890–5895.
254. Wolf, D. M., Arkin, A. P. (2003). Motifs modules and games in bacteria. *Curr. Opin. Microbiol.*, *6*, 125–134.
255. Ogata, H., Fujibuchi, W., Goto, S., Kanehisa, M. A. (2000). Heuristic graph comparison algorithm and its application to detect functionally related enzyme clusters. *Nucleic Acids Res.*, *28*, 4021–4028.
256. Nakaya, A., Goto, S., Kanehisa, M. A. (2001). Extraction of correlated gene clusters by multiple graph comparison. *Genome Inform. Ser. Workshop Genome Inform. 12*, 44–53.
257. Kelley, B. P., Sharan, R., Karp, R. M., Sittler, T., Root, D. E., Stockwell, B. R., Ideker, T. (2003). Conserved pathways within bacteria and yeast as revealed by global protein network alignment. *Proc. Natl. Acad. Sci. USA*, *100*, 11394–11399.
258. Jafari, S. S., Nielson, M., Graham, D. I., Maxwell, W. L. (1998). Axonal cytoskeletal changes after nondisruptive axonal injury. II. Intermediate sized axons. *J. Neurotrauma*, *15*, 955–966.
259. Jafari, S. S., Maxwell, W. L., Neilson, M., Graham, D. I. (1997). Axonal cytoskeletal changes after non-disruptive axonal injury. *J. Neurocytol.*, *26*, 207–221.
260. Quigley, H. A., Anderson, D. R. (1977). Distribution of axonal transport blockade by acute intraocular pressure elevation in the primate optic nerve head. *Ophthalmol. Vis. Sci.*, *16*, 640–644.
261. Salvador-Silva, M., Aoi, S., Parker, A., Yang, P., Pecen, P., Hernandez, M. R. (2004). Responses and signaling pathways in human optic nerve head astrocytes exposed to hydrostatic pressure in vitro. *Glia*, *45*, 364–377.
262. Zhao, H., Langerod, A., Ji, Y., Nowels, K. W., Nesland, J. M., Tibshirani, R., Bukholm, I. K., Karesen, R., Botstein, D., Borresen-Dale, A. L., Jeffrey, S. S. (2004). Different gene expression patterns in invasive lobular and ductal carcinomas of the breast. *Mol. Biol. Cell*, *15*, 2523–2536.

# HIGH-THROUGHPUT FLOW CYTOMETRY

Larry A. Sklar, Peter C. Simons, Anna Waller, Sean M. Biggs, Susan M. Young, Marius Olah, Cristian Bologa, Tudor I. Oprea, Eric R. Prossnitz, and Bruce S. Edwards

*University of New Mexico School of Medicine*
*University of New Mexico Health Sciences Center*
*Albuquerque, New Mexico*

1 INTRODUCTION
2 OVERVIEW
GPCR Signaling Pathways as Discovoery Targets
Flow Cytometry Well Suited to Diversity Analysis
Challenge of Throughput
Challenge of Content
Challenge of Informatics
Screening
Technolgy Associated with HyperCyt
3 PROTEIN EXPRESSION AND CHARACTERIZATION
GPCR Constructs
Soluble GPCRs Reconsitute with G Proteins
FPR Assembles with Arrestin
FPR Colocalization with Arrestin in Cells Parallels FPR Assemblies in Solution
Arrestin Regulates Intracellular FPR Traffic
FPR-$G_{i\alpha 2}$ Fusion Protein Functions Physiologically
FPR Tail Assembly (Protein Domain Display)
Proof of Principle for β2AR and Other GPCRs
4 SOLUBLE AND BEAD-BASED ASSEMBLIES
FPR Assemblies Used to Characterize and Validate the Displays

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

β2AR Extends the Soluble GPCR on a Bead Concept
Display of Cell Cycle Proteins Demonstrates Further Generality
5 HIGH-THROUGHPUT FLOW CYTOMETRY TECHNOLOGIES
HyperCyt Exhibits Versatility and Compatibility with High Content
HyperCyt Adds Microfluidic Mixing for Analysis of Soluble Compounds
HyperCyt Cell Sorting Extends Microfluidic Mixing
Sample Carryover in Microfluidic Mixing Systems
Software Development
FPR Ligand Screening Assay
6 VIRTUAL SCRENING
Docking and Scoring
Generation and Maintenance of Chemical Databases for VS
Virtual Screening in the Context of High-Performance Computing
FPR Case Study
7 SUMMARY
References

## 1 INTRODUCTION

This chapter describes the creation of a discovery team in an academic environment. The team is specifically structured to perform four integrated activities: (1) expressing and characterizing membrane proteins of interest in appropriate quantities for analysis, (2) formatting the materials for analysis of cell physiology and molecular interactions in a novel manner, typically through solubilization of the components and display on particles, (3) performing screens of small-molecule activity as well as protein–protein interactions on a novel high-throughput flow cytometric instrument platform, and (4) integrating virtual screening and cheminformatics with activity screening to improve the overall efficiency of the discovery process. The results of the overall process are intended to provide insight into cell physiology, ligand–receptor and protein–protein interactions, the discovery of small molecules as probes for biological systems and, potentially, as leads in drug discovery. The process itself poses an interesting set of challenges and opportunities. The challenges arise in part from the need to produce material in appropriate quantity and with appropriate properties for screening. In addition, the screening needs to be performed in a mode that is targeted rather than brute force. The opportunities come not only from the discoveries that result from the research but are also derived from research into the nature of the discovery process itself and the insights that result from experiments into discovery. The purpose of this chapter is twofold: first, to provide a context for our particular approach to discovery research and, second, to provide specific examples of how the individual components of the process function in our laboratories.

## 2 OVERVIEW

Flow cytometry is proving to be a good match for G-protein-coupled receptor (GPCR) drug discovery, taking advantage of its versatility of assays, its ability to discriminate free and bound molecular complexes, and the incorporation of microfluidic principles to optimize sample delivery. A number of recent developments have suggested that flow cytometry would be a good match for chemical and biological diversity in general. With up to 30,000 human genes that encode proteins, the number of protein species may reach several hundred thousand including splice variants and posttranslational modification. In addition, there are upwards of 1 million natural polymorphisms in the human genome and other mutations that create protein variants. The analysis of protein–protein and protein–ligand interactions has broad-ranging implications for biomedical research.

The GPCRs are the most numerous molecular family in the human genome, representing 600 or more proteins that allow extracellular signals to induce cellular responses. Each GPCR could couple to as many as 1620 G-protein heterotrimers (27α, 5β, and 12γ subunits), relatively few of which are characterized. Signaling then proceeds through a host of effectors. The GPCR signals are terminated through steps involving phosphorylation by a family of kinases, interaction with arrestin family members, and processing inside the cell. While cell expression experiments (e.g., one GPCR and one G-protein heterotrimer) can typically resolve one set of interactions at a time, combinatorial approaches may resolve many interactions simultaneously. Because of the multiplicity of components, there has been little systematic cellular approach to GPCR diversity issues. Flat arrays and the flow cytometric analog, multiplex bead-based approaches, are proposed to meet this challenge [1]. Flow cytometric approaches will prove to be applicable both for understanding assemblies and mechanisms and, when combined with high-throughput (HyperCyt) and virtual screening, in small-molecule discovery. Moreover, these approaches have the potential to replace radioligands and to compete with the sensitivity and throughput of plasmon resonance. The introduction of soluble GPCR molecular assemblies in this arena has the potential of opening the door for other membrane molecules, and other signaling and response pathways. Here, we will focus on GPCR and related proteins (cell- and bead-based assemblies), a robust high-throughput (HT) flow cytometry platform (our HyperCyt), and the integration of virtual screening.

### GPCR Signaling Pathways as Discovery Targets

The GPCRs represent the target of ~50 percent of the prescription drugs on the market. They are involved in virtually every physiological process in the human body, with ligands including light, odorants, amines, peptides, proteins, lipids, and nucleotides. GPCRs contain a common structural motif with seven

transmembrane α helices. With the recent description of the three-dimensional crystal structure of rhodopsin in its inactive state, a greater, though still incomplete, understanding of the functions of this receptor family has been achieved [2, 3]. In addition to the activation of G proteins and separation of their α and βγ subunits (Fig. 1), GPCRs are regulated primarily by kinases, including second-messenger kinases and the family of G-protein-coupled receptor kinases (GRKs) [4]. Following GPCR phosphorylation by GRKs, arrestins associate with GPCRs [5]. The traditional role of arrestins has been to serve as desensitizing agents, preventing further association of GPCRs with G proteins. However, recent studies have demonstrated that arrestins can serve as adapters in the process of receptor internalization as well as scaffolds in the activation of numerous kinase pathways. Interactions between GPCRs and cellular proteins such as adaptins, rab GTPases, phosphatases, and ion channels have also been described [6]. The newly described regulators of G-protein signaling (RGS) proteins are downstream, controlling the kinetics and specificity of GPCR-mediated G-protein signals [7]. The localized tissue distributions and diverse structural elements of RGS proteins make them an intriguing

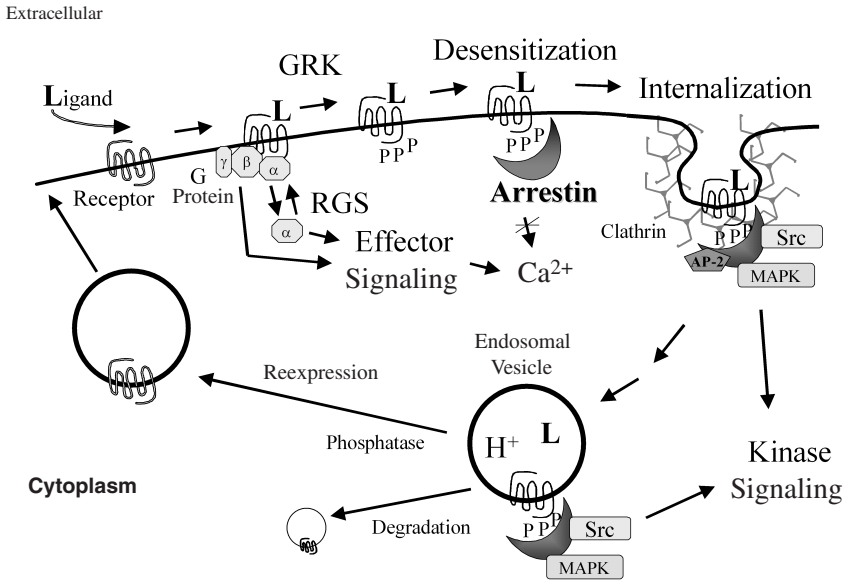


**Figure 1** GPCR activation, termination, and processing. When a GPCR binds a ligand, the α subunit of the G-protein heterotrimer releases bound GDP and binds GTP, then separates from the βγ subunit. The GPCR gets phosphorylated by GRK, and the phosphorylated receptor binds arrestin. Both of these steps can lead to effector signaling. The GPCR is then internalized in an endosomal vesicle, from which it can be degraded or recycled back to the surface of the cell.

target for drug development [8]. Cytometric approaches that we employ for GPCR and related protein–protein interactions are listed in Table 1.

The GPCRs are divided into five major families based on sequence homologies. These include the class A (rhodopsin), class B (secretin-like), class C (metabotropic glutamate), class D (fungal pheromone), and class E [cyclic adenosine monophosphate (cAMP) receptors of dicyostelium]. The major class of physiological significance to humans is the class A group of GPCRs. In this class are the monoamine, peptide, hormone protein, olfactory, prostanoid, nucleotide-like, cannabinoid, platelet-activating factor, gonadotropin-releasing hormone, thyrotropin-releasing hormone, melatonin,

**TABLE 1 Flow Cytometric Approaches to GPCR Physiology and Molecular Assembly**

| Physiology or Assembly | Approach |
|---|---|
| Expression of GPCR or signaling partner on cells [30–33, 98] | Analysis and sorting of transfected cells based on fluorescent ligand, antibody, or GFP fusion protein |
| GPCR physiology [30–33, 35, 36, 98, 99] | Analysis of cell response, GPCR internalization, GPCR desensitization, GPCR recycling |
| Solubilization of GPCR [30–33, 35–37, 98, 100] | Screen detergent for ability to produce receptor that binds to ligand beads or G-protein beads (see below) |
| Isolation and display of proteins [34–38] | Epitope-tagged proteins/epitope recognizing beads |
| Reconstitution of components for assembly analysis | Form complexes of GPCR and partners on particles |
| Assembly and disassembly kinetics [36, 37, 101–105] | Analyze kinetics with subsecond resolution with "rapid mix flow cytometer" (200-ms dead time) |
| GPCR phosphorylation status | Ability of GPCR to bind to arrestin beads |
| Ternary complex formation on G-protein beads [36, 37, 106] | Assembly of GPCR on G-protein beads with agonist binding (Figs. 8*b*, 8*c*, and 8*d*) |
| Display of GPCR ligand on ligand beads [37] | Competition of GPCR on ligand beads for agonist or antagonist binding (Fig. 8*g*) |
| Duplex assay of ligand beads and G-protein beads [37, 38] | Single step discrimination of agonist and antagonist |
| Display of protein domain on beads and multiplex [34] | Proteomic analysis of binding between domains of partners |
| G-protein multiplex [106] | Proteomic analysis of specificity of receptor binding to G proteins; screen for solubilization; screen for ligands; screen for phosphorylation |

viral, lysosphingolipid, LTB4, and related orphan receptors. The largest subfamilies are the monoamine, peptide, and olfactory groups.

We have focused on the formyl peptide receptor (FPR), a member of the chemoattractant/chemokine subfamily of GPCRs involved in leukocyte trafficking and inflammatory responses to bacterial peptides [9–11]. FPR signals through $G_i$ proteins and intracellular calcium. The ligands for FPR are short hydrophobic peptides, typically with an *N*-formyl group on the amino-terminal methionine, followed by two to five hydrophobic amino acids such as leucine, phenylalanine, norleucine, and tyrosine [12]. FPR, with a family of fluorescent ligands of varying affinity used to assess its presence and functional state [12], has been the prototype for the development of most of the assemblies and cell responses in Table 1. More recently we have examined the β2-adrenergic receptor (β2AR), the prototypic monoamine receptor. β2AR signals through $G_s$ and adenylate cyclase. The contrasts between β2AR and FPR present notable opportunities. β2AR internalization depends on arrestin, whereas FPR internalization does not. It has been reported that β2AR specificity for $G_s$ and $G_i$ depend upon its phosphorylation status and ligand. β2AR has a constitutively active mutant that is predicted to assemble with $G_s$ in the absence of ligand.

Currently, there are ~200 GPCRs with known ligands and ~150 orphan GPCRs [13]. There are also ~330 full-length human odorant-like GPCR genes known to date [14]. Since odorant ligands have been found for only a few of the potential odorant receptors, the odorant-like classification and terminology for a novel receptor are defined primarily by sequence homology [15]. Odorant-like receptor genes have been discovered in the olfactory epithelium, as might be expected, but they are also expressed in tissues lacking any known olfactory function, suggesting that some odorant-like GPCRs are not involved in the chemosensory process. Some GPCRs detect pheromones, and in humans, many pheromones are derived from steroid hormones. One olfactory-like GPCR has been shown to be expressed in the prostate (prostate-specific GPCR, or PSGR), and its expression is greatly increased in prostate tumors, suggesting a role in the process of carcinogenesis [16]. Another, GPR30, detects estrogen, previously thought only to activate nuclear receptors. GPR30 was cloned as an orphan receptor and subsequently shown to be activated by estrogen. A recent report identified a testosterone-activated GPCR [17], although the sequence has not yet been reported. As hundreds of cloned GPCRs remain orphans, there is great opportunity in the discovery of their ligands and drugs that modulate their activity.

## Flow Cytometry Well Suited to Diversity Analysis

Biological diversity (numbers of proteins) and chemical diversity (numbers of compounds) present the analytical challenges of throughput, content, and informatics (Table 2). We anticipate that flow cytometry, which we envisioned as playing a role in drug discovery, can play a larger role in meeting

**TABLE 2 Flow Cytometry for Diversity and Discovery Research**

| Technical Capability | Description |
|---|---|
| Sorting and analysis rates | Up to 50,000 events/s; 4.32 billion/day |
| Sample handling (HyperCyt) [19–24, 28, 41–43, 49, 103, 107–109] | Projected to 100,000 endpoint assays/day (100/min), small volume (1 μL); to 30 online mixes/min; to 10 sorted samples/min |
| Assays [18, 20, 21, 30–33, 98, 108] | Cell responses and populations, bead-based assemblies |
| Molecular assemblies [18, 36, 37, 86, 110–114] | Site numbers, reaction rates, $K_d$'s, any two molecules |
| Sensitivity | Attomole to femtomole per sample |
| Number of parameters | Up to 12 parameters per assay |
| Content [35, 103] | Multiplex to 100 assays/sample; 10–20 plex assay/s (864,000/day) |

the challenges of diversity by creating systems for measuring protein–protein interactions and cell responses with high throughput and high content. A flow cytometer is a multiparameter instrument that is normally used in a largely manual mode, working with individual samples. Literally thousands of different assays have been performed with a flow cytometer on cells or particles ranging from single parameter endpoint assays to time-dependent responses and complex assays involving multiple cell populations or simultaneous analysis of multiple responses. Because single particles are illuminated in a small volume and the signal is a pulse above a direct current background, the detection can resolve fluorescence on the particle from the volume around it without a wash step. The parameters include scatter signals and up to 10 or more fluorescence signals depending upon the number of lasers. Multiplexing is a technique that takes advantage of the multiparameter capability of flow cytometry. For example, an assay could use the green signal. The particles could be addressed using the red and orange signals to create a suspension array. Because flow cytometers are accessible to investigators in nearly all research environments, we believe that both the biological and instrumental platforms required for diversity analysis will be within their grasp.

The flow cytometer can be used as a generic platform for molecular assembly by attaching one molecular species to a particle using any number of tagging schemes (epitope tags, biotinylation, etc.). The particles become bright when a second species, labeled with a fluorophore, binds to the first. The labeling schemes include conjugation, fluorescent protein fusions, and the like. Tools are available for performing subsecond kinetic analysis. Our experience suggests that analysis can be extended to pairs of molecules of all types, including proteins, peptides, small molecules, deoxyribonucleic acid (DNA), ribonucleic acid (RNA), lipids, toxins, and so forth. The sensitivity is related to the number of particles, the fluorophores per particle, and the sample volume. It

is practical to sample hundreds of particles in 1-μL volume with tens of thousands of molecules per particle. This translates into attomole quantities, or picogram amounts of a typical protein.

## Challenge of Throughput

For the greatest impact, flow cytometry needs to deliver individual samples (microliter volumes) from multiwell plates at rates to 100,000 samples/day. Our HyperCyt system [18, 19] interfaces a flow cytometer and autosampler (Fig. 2). Commercial alternatives are ~30 times slower due to differences in the way samples are delivered and data is acquired. In HyperCyt, samples are separated from one another by air bubbles and delivered in a continuous stream. Likewise, the data are collected in a continuous stream, a plate representing a single data file, with the contents of each well separated by a gap introduced by the air bubble. The data are analyzed by proprietary software (FCSQuery) developed by Bruce Edwards. HyperCyt is a microsystem where microfluidic dimensions are at work in transporting samples to the flow cytometer.

We have addressed problems of mixing and carryover [20–24]. Mixing in HyperCyt requires breaking laminar flow and can be performed at sample delivery rates >1 sample/s. However, separation of samples by air bubbles in microfluidics leads to two types of carryover, of differing significance. Particle

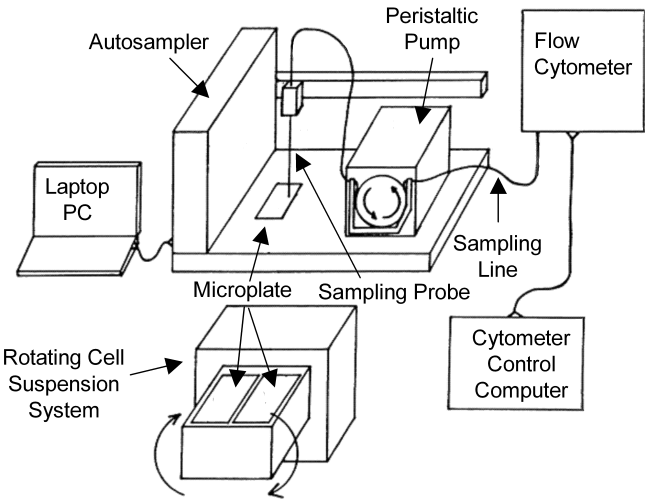


**Figure 2** HyperCyt. A laptop PC controls both an autosampler and the cytometer control computer. Cells in a 96-well plate are kept in suspension with a rotating suspension system, which depends on surface tension to keep the fluid and cells in the wells while the plate is upside down. A peristaltic pump moves fluid and cells from the autosampler probe to the flow cytometer sample tube against the sheath flow pressure.

carryover from sample to sample is minimized by bubbles because particles tend to travel in the middle of the stream. Practically speaking, in an endpoint assay where the assay contents are premixed and developed in the well, a sample is contaminated with ~1 percent of the particles from the prior sample. For higher stringency, a wash plug can be easily added between samples at the cost of reducing throughput. Of greater significance, particularly for online assays, is fluid carryover. In this case, the assay (such as a cell response to a drug compound) develops online, and some of the compound from each sample is left behind as a thin film.

It is worth noting that flow cytometry is used in protein engineering or expression of combinatorial molecular libraries. Here, cells expressing a unique fluorescence phenotype can be analyzed and sorted at rates up to 50,000 per second (4.32 billion separate assays/24 h). We have participated in a project with Novasite Inc. based on identifying protein variants according to cell responses. Here, each sample is mixed online with a compound and analyzed for response in a primary screen using HyperCyt. Sorting (cloning the variants) is performed as a secondary screen with HyperCyt for active compounds. Another drug discovery possibility is the "one-compound, one-bead" format [25]. In this scenario, appropriately labeled target proteins associate with a bead expressing a compound specific to the target. A modes-sized peptide array for nuclear receptors has already been reported [26]. As in the case of protein variants expressed on cells, the particles that become labeled could potentially be sorted at rates to 4.32 billion/day. However, particles ~100 μm diameter are required for the mass spectrometry step of identifying the compound on a particular bead. Sorting of one-bead, one-compound libraries has been achieved with the COPAS technology (www.unionbio.com) at rates to tens of thousands of particles/hour, adequate for screening libraries. This technology is a perfect match for our soluble GPCR-green fluorescent protein (GFP) fusion proteins, and could be used both in screening for active compounds and screening for detergents that keep receptors active.

## Challenge of Content

Content in discovery research can be based on performing multiple measurements at a single time. High content usually requires multiparameter measurements at which flow cytometry excels. With digital signal processing, particle analysis rates in flow cytometry can easily reach 10,000 events/s with particle densities ~$10^4$ particles/μL, and sample delivery ~1 μL/s at 15 psi or less. High content has been achieved by multiplexing 100 distinct bead sets using two fluorescent signals from particle sets labeled with 1 of 10 different levels of brightness for each color (www.luminexcorp.com). Since the particle statistics could allow for 10- to 20-plex assays in each second of 5000 to 10,000 events, overall multiplex throughput for HyperCyt could approach 1 million assays/day. For drug discovery, a library of compounds could be tested against a family of receptors. Thus, a cell mixture could contain 10 sets of cells, each

with a different GPCR, and each with a different address tag based on fluorescent intensity or color combinations. Each compound could be tested for specificity with respect to the members of the GPCR family. For clinical diagnostics, each well may contain a cell sample characterized by a complex assay such as immunophenotyping of blood leukocytes. With 10,000 cells stained with up to 10 fluorescent antibodies for distinct leukocyte subsets, one sample each second would allow dozens of leukocyte subsets to be identified.

Our team has recently considered one-step discrimination of GPCR agonists and antagonists. Using β2AR as the model receptor, we learned how to produce GPCR in soluble form as a GFP fusion protein that binds to G proteins displayed on beads (G-protein beads), and simultaneously to ligands displayed on distinct beads (ligand beads). The G-protein assembly forms in the presence of an agonist; the ligand assembly is inhibited by agonist or antagonist. A biplex of these two assays conducted simultaneously discriminates an agonist and antagonist in a single step. Moreover, a proteomic analysis results from displaying different G proteins as a multiplexed array, each GPCR evaluated for its ability to associate with a particular G protein. Content could be further enhanced by combining a cell response assay with a molecular assembly in the same sample well to evaluate physiology and assembly simultaneously.

## Challenge of Informatics

To fully realize the potential of flow cytometry to meet the challenges of diversity, throughput, and content, informatic systems need to be integrated with the data collection efforts. Our group has provided persuasive evidence for acquiring a multiwell plate of samples as a single data file, where each cluster of events represents the contents of a single well [18–21, 27, 28]. This challenge is extended by multiplexing where each cluster of events represents 10 or more assays collected each second. In addition, the data need to be tracked back to the contents of the well, each well potentially containing a member of a chemical library. The data then need to be output in a format compatible with cheminformatic analysis of the assay result, the activity of the compound.

Since Beilstein and Chemical Abstracts Service have indexed ~15 to 20 million molecules to date, it is reasonable to assume that the medicinal chemistry space does not exceed 100 million compounds. This number is amenable to cheminformatics data handling, with computer capacity doubling every 2 years compared to the 5 to 10 percent annual growth of chemical space (note: this excludes the virtual chemical space). It is safe to assume, therefore, that tools and methodologies developed in cheminformatics will adequately handle large problems over the next 5 to 10 years. Chemical diversity, in a drug discovery sense, is one of the key issues in cheminformatics today. It is widely recognized that computational methods that handle molecular diversity—regardless of the chosen metric—need to operate under the restrictions

imposed by pharmacokinetics (e.g., the desired properties of passive transcellular permeability, appropriate tissue distribution, aqueous solubility, metabolic stability, good excretion, and low toxicity profile, or ADMET), in addition to pharmacodynamics (efficacy, selectivity, potency). With numbers such as $10^{200}$ molecules accessible in the universe (attributed to D. Weininger) and $10^{16}$ chemicals accessible via combinatorial chemistry reactions (attributed to M. Geysen), it is not surprising that lead discovery, picking a needle in the haystack, is daunting.

Chemical vendors began to offer "diverse" chemical libraries in the early 1990s. By 2002, most of these companies had evolved high-quality, leadlike and druglike chemicals. Our in-house database, CADB08 (Commercially Available Data Base version 0.8), started with a collection of offerings from ChemNavigator (www.chemnavigator.com)—which covers over 100 vendors worldwide and offers over 7 million unique structures. Following duplicate removal, elimination of unwanted structures and property filtering, we obtained a database containing approximately 900,000 small, nonpeptide chemical structures. At the other end of the spectrum are efforts to mine the chemical space represented by natural-product-like molecules. This features a relatively high density of stereo centers, leading to structures of increased complexity. ChemBank, an ongoing project organized by the Initiative for Chemical Genomics (ICG) and being built by the Institute for Chemistry and Cell Biology (ICCB) (www.chembank.med.harvard.edu) has accumulated over 300,000 natural-product-like structures, based on the Diversity-oriented Organic Synthesis (DOS) approach [29]. Taken together, these two chemical collections represent a balanced, chemically diverse library that can be used for virtual screening.

## Screening

Drug discovery in the last decade has undergone dramatic changes. Given the relatively low success rate of structure-based drug design in the early 1990s, the pharmaceutical industry reoriented itself to build the *high-throughput screening (HTS) machine*, whereby very large numbers of chemicals were screened for biological activities. Combinatorial chemistry, developed in part to justify the HTS machine, led to the development of large sets of chemicals for screening. At the same time, smaller subsets of libraries are dedicated to high-throughput (HT) nuclear magnetic resonance (NMR) screening where small molecules are tested for binding via NMR experiments, or to HT crystallography where small molecules are co-crystallized with proteins of interest. However, these two approaches are not applied, at an industrial scale, for GPCR targets, as there is only one crystal structure available (bovine rhodopsin). In the postgenomic era, brute force screening seems to be ramping down and is being replaced by targeted screens, on smaller, focused libraries.

These efforts now are likely to start with virtual screening, including docking ligands onto crystal structures. If crystal structures are not available,

such as for most GPCR, ligands are docked into a homology model. An alternative is to perform pharmacophore searches, assuming that enough active molecules are known a priori to develop such pharmacophores. Selected members of a large, virtually screened library are then pooled into a sublibrary for physical screening. The output of the results can be used to select other compounds, guide further synthesis, and as a tool for predicting the ADMET characteristics of the active compounds. An iterative process of virtual and biological screening is within reach of teams that do not have the resources of biotech companies or pharmaceutical companies.

### Technology Associated with HyperCyt

The potential for displaying virtually any molecule or cell response in a format compatible with HyperCyt (sampling rates >1/s with 10- to 20-plex assays) will make flow cytometry a powerful option for the real-time analysis of molecular interactions in drug discovery and diversity research. The strategy to achieve optimal outcome requires four complementary activities. Component 1 expresses and characterizes proteins relevant to signal transduction and termination. Component 2 displays functional proteins in a manner compatible with flow cytometric analysis. Component 3 implements robust HT screening. Component 4 implements virtual screening tools and selects compound libraries for HyperCyt. The resulting technological advances will allow us to define mechanisms of cell activation through GPCRs and to explore the diversity of molecular interactions in GPCR pathways.

## 3 PROTEIN EXPRESSION AND CHARACTERIZATION

There are two major goals of component 1 (protein expression and characterization). The first was to generate and characterize the tools needed to reconstitute signaling complexes on beads in component 2 (soluble and bead-based assemblies). To this end, we have engineered many signal transduction molecules appropriate for reconstitution studies. The second goal was to develop a battery of flow cytometric assays that could be used to evaluate GPCR pathways. Thus, the characterization of these expressed molecules in cells and in soluble systems prior to their display on beads led to significant biological advances.

### GPCR Constructs

A significant fraction of our effort has been directed toward studying complexes between GPCR and G proteins and between GPCR and arrestins. We have generated wild-type, phosphorylation deficient mutants, and chimeric proteins including GPCR-GFP, FPR-$G_i$, FPR-$G_i$-GFP, and GPCR-$G_i$. We have

made receptors with hexahistidine ($his_6$) or Myc tags at the amino terminus to facilitate purification and immobilization.

## Soluble GPCRs Reconstitute with G Proteins

Our earliest studies focused on reconstituting the wild-type FPR with purified G proteins in solution as a precursor to reconstituting the complexes on beads. For these experiments we used a fluorometric assay to determine the dissociation of a fluorescent ligand from the solubilized FPR in detergent (Fig. 3*a*). The ligand dissociation rate differs between FPR coupled to a G protein and FPR alone. By determining the sensitivity of the dissociation of a fluorescent ligand to the presence of guanine nucleotide, which disrupts G-protein coupling, we demonstrated time- and concentration-dependent reconstitution of FPR with both endogenous and exogenous G proteins. Solubilized FPR with bound ligand and Gαi3 protein interacted with an affinity ~$10^{-6}$ M.

## FPR Assembles with Arrestin

We extended our studies of FPR assemblies to those involving the protein arrestin. As arrestin predominantly interacts with phosphorylated FPR, we obtained FPR from cells that had been stimulated with ligand to achieve in vivo phosphorylation. We demonstrated a concentration- and time-dependent

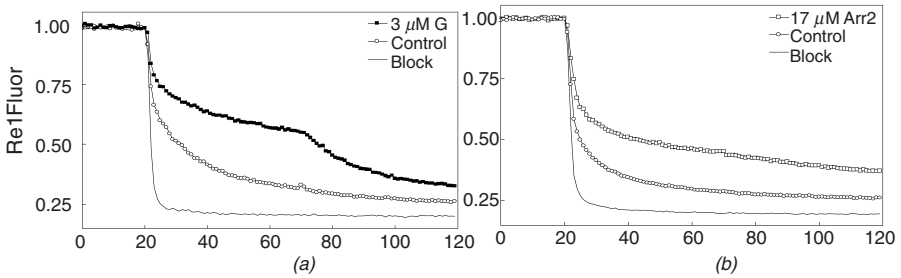


**Figure 3** Fluorescent peptide assay of reconstitution of FPR with (*a*) G-protein and (*b*) arrestin. GTPγS (added at 70 s) distinguishes G-protein coupling, which is sensitive to the guanine nucleotide analog, from arrestin binding, which is insensitive. The data were obtained in a stirred fluorometer cuvet by addition of an antibody that quenches the fluorescence of unbound peptide, added at 20 s. When the receptor is blocked with excess nonfluorescent peptide (block), the peptide is rapidly quenched to about 0.25 of the initial fluorescence of the mixture. When there is only receptor and fluorescent peptide (control), the bound peptide remains fluorescent past 25 s, dissociates exponentially, and is immediately quenched. Both G protein and arrestin increase the affinity of the receptor for fluorescent ligand, resulting in higher binding and slower dissociation. Only the G protein is rapidly converted to an inactive state by the introduction of GTPγS at 70 s, and the fluorescent peptide dissociates quickly again [30].

reconstitution of liganded, phosphorylated FPR with exogenous arrestin-2 and -3 to form a high agonist affinity, nucleotide-insensitive complex with effective concentration $EC_{50}$ values of 0.5 and 0.9 μM, respectively ([30], Fig. 3*b*). Moreover, we demonstrated that the addition of G proteins was unable to alter the ligand dissociation kinetics or induce a GTPγS-sensitive state of the phosphorylated FPR. The properties of phosphorylated FPR were reversible upon phosphatase treatment.

We further investigated the interactions between partially phosphorylated FPR and G proteins and arrestins. For these studies, we used FPR phosphorylation-deficient mutants (ΔA, ΔB, ΔC, ΔD). The ΔA and ΔB mutants contain only four of eight nonoverlapping potential phosphorylation sites, whereas the ΔC and ΔD each contain six of the eight potential phosphorylation sites. We demonstrated that phosphorylation of the wild-type FPR lowers its affinity for G protein, whereas mutant receptors lacking four potential phosphorylation sites (ΔA and ΔB) retain their ability to couple to G protein [31]. Phosphorylated mutant receptors lacking only two potential phosphorylation sites (ΔC and ΔD) did not couple to G protein but did bind arrestins.

## FPR Colocalization with Arrestin in Cells Parallels FPR Assemblies in Solution

To corroborate the assemblies with the physiology of the proteins in vivo, we used confocal fluorescence microscopy to evaluate the interactions between FPR, as assessed by fluorescent ligand, and arrestins, as assessed with arrestin–GFP constructs (Fig. 4). We developed a novel series of red fluorescent ligands to allow simultaneous analysis of both proteins [32]. This approach demonstrated that, whereas stimulated wild-type FPR in whole cells colocalized with arrestin-2, the stimulated GPCR lacking four potential phosphorylation sites displayed no colocalization with arrestin-2 [31]. However, mutant receptors lacking only two potential phosphorylation sites are restored

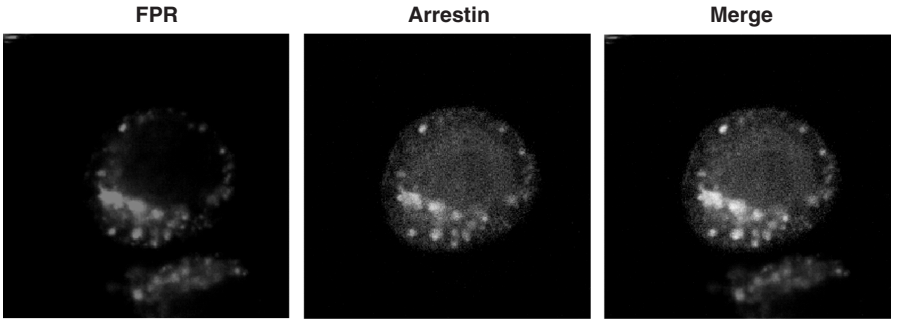


**Figure 4** Colocalization of FPR (originally red ligand) with arrestin (originally green GFP) Following stimulation for 10 min with agonist. The brighter overlap demonstrates colocalization for a large fraction of the FPR [32].

in their ability to colocalize with arrestin-2 in vivo. Thus, we concluded that there is a submaximal level of FPR phosphorylation that simultaneously results in an inhibition of G-protein binding and an induction of arrestin binding. We proposed that phosphorylation alone may be sufficient to desensitize the FPR in vivo, raising the possibility that for certain GPCRs, desensitization may not be the primary function of arrestin.

We have subsequently evaluated mechanisms of arrestin activation and the role of the activation state of arrestin on the formation and properties of the FPR complex. Very little was known about the relationships between the sites of receptor phosphorylation, the resulting affinities of arrestin binding, and the ensuing mechanisms of receptor regulation for any given GPCR. To investigate these interactions, we used an active truncated mutant of arrestin (amino acids 1 to 382) and phosphorylation-deficient mutants of FPR described above. We demonstrated that the phosphorylation status of residues between amino acids 328 and 332 of the FPR is a key determinant that regulates the affinity of FPR for arrestins, whereas the phosphorylation status of residues between amino acids 334 and 339 regulates the affinity of FPR for agonist when arrestin is bound [32]. Again confocal fluorescence microscopy confirmed the receptor binding interactions in vivo. The results indicated that the agonist affinity state of FPR is principally regulated by phosphorylation at specific sites and is not simply a consequence of arrestin binding, contradicting current dogma. Furthermore, these studies were the first to demonstrate that GPCR agonist affinity and the affinity of arrestin binding to the phosphorylated receptor are regulated by distinct receptor phosphorylation sites.

## Arrestin Regulates Intracellular FPR Traffic

Our most recent studies examined the role of arrestins in FPR trafficking in the cell (Fig. 5). Surprisingly, our previous results indicated that FPR internalization does not require arrestin. To further examine roles for arrestins in FPR function, we determined the effects of activating arrestin mutations on ternary complex formation (ligand, FPR, arrestin) and cellular trafficking. One such activating mutation, arrestin-2-3A protein, contains targeted mutations of three hydrophobic residues (I386A, V387A, F388A) in the carboxy-terminus that contribute to basal inactivity in the wild-type protein. Again FPR–arrestin interactions were confirmed in vivo by colocalization studies using confocal microscopy with arrestin–GFP chimeras and a fluorescent FPR agonist. To assess the effects of activated arrestin expression on receptor trafficking, we used a novel quantitative, in vivo flow cytometric assay. We demonstrated that activated, but not wild-type, arrestin expression inhibits FPR internalization [32]. We further demonstrated that arrestin-2-3A–GFP expression inhibits FPR recycling. In contrast to results with FPR, expression of activated arrestin had no effect on either the internalization or recycling rate of β2AR, which internalizes in an arrestin-dependent manner. These

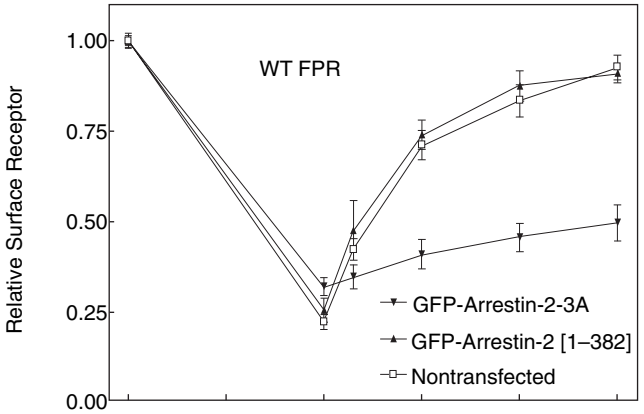


**Figure 5** FPR recycling. Cells expressing FPR were stimulated with ligand to bring surface expression down to about 25 percent in all cases. The ligand was washed away, and the amount of receptor on the cell surface was measured over time by ligand binding. The rate of surface receptor reexpression was inhibited by an activated mutant of arrestin (arrestin-2A:I386A/V387A/F388A), suggesting that arrestin plays a critical role in receptor recycling in this system, rather than in receptor downregulation as in the visual system [32].

results indicate that arrestin is required for recycling of certain GPCRs, such as FPR, and that the activation state of arrestin regulates this process.

## FPR-$G_{i\alpha 2}$ Fusion Protein Functions Physiologically

As suggested above, phosphorylation of wild-type FPR might be sufficient to prevent coupling of GPCRs to G proteins in vivo. We investigated this phenomenon using a chimeric protein between FPR and the $G_{\alpha i2}$ subunit of heterotrimeric G proteins. The chimeric GPCR-G protein approach for component 2 could replace the requirement of providing an exogenous $G_{\alpha}$ subunit for each GPCR display. The functional capabilities of this chimeric protein were evaluated both in vivo, in stably transfected myeloid U937 cells, and in vitro, using our soluble reconstitution system [33]. The chimeric protein existed as a soluble complex containing $\beta\gamma$ subunits. The chimeric protein mobilized intracellular calcium, desensitized normally in response to agonist, and was internalized (Fig. 6) and recycled at rates similar to those of the wild-type FPR. Confocal fluorescence microscopy revealed that internalized chimeric receptors colocalized with arrestin, as well as G protein. Soluble reconstitution experiments revealed that this chimeric receptor, even in its phosphorylated state, existed as a high ligand affinity G-protein complex, only partially prevented through the addition of arrestins. These results indicated that under conditions of high local G-protein concentrations, the phosphorylated form of the FPR is still capable of binding G proteins.

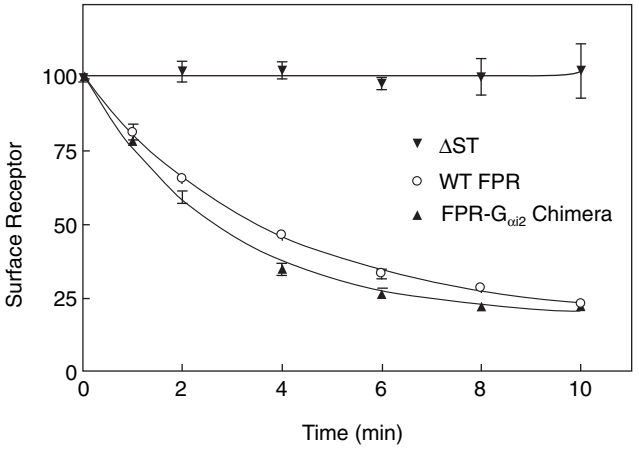


**Figure 6** FPR internalization depends on phosphorylation. Internalization of wild-type (WT) FPR, a mutant lacking C-terminal phosphorylation sites, and a chimeric FPR-$G_{\alpha i2}$ protein were followed by ligand binding.

## FPR Tail Assembly (Protein Domain Display)

Because GPCRs are integral membrane proteins, any assay of their function in solution requires the presence of detergent. As an alternative, we utilized peptides representing the unphosphorylated and phosphorylated FPR carboxyl terminus. These were chemically synthesized (e.g., biotin-MGQDFRE RLIHALPASLERALTEDS(P)TQT(P)SDT(P)TN(P)STLPSAEVELQAK-OH) to be homogeneous and bound to polystyrene beads via a biotin/streptavidin (SA) interaction. In experiments that excluded detergents, we chose fluorescein-conjugated arrestins to examine binding interactions between arrestins and the bead-bound FPR carboxyl terminus. Analyses were performed by flow cytometry (Fig. 7). Arrestin-2 and arrestin-3 bound to the 47mer FPR carboxyl-terminal peptide in a phosphorylation-dependent manner, with $K_d$ values in the micromolar range [34]. Activated arrestin mutants that display phosphorylation-independent binding to intact receptors also bound the bead-bound FPR terminus in a phosphorylation-dependent manner, but with greater affinity than the full-length arrestins, with $K_d$ values in the 5 to 50 nM range. The results suggested that the carboxyl terminus of arrestin is a critical determinant in regulating the binding affinity of arrestin for the phosphorylated domains of GPCRs, and furthermore that non-detergent-based approaches to investigate GPCR assemblies represent a viable alternative to approaches utilizing solubilized holo-GPCRs. To this end we have synthesized a number of carboxy-terminal peptides of varying length, both phosphorylated and not, from GPCRs including rhodopsin, β2AR, and vasopressin receptors. Preliminary data indicate that the minimal length for arrestin binding is between 16 and 21 amino acids. The number of phosphates and the amino acid sequence specificity remain to be characterized.

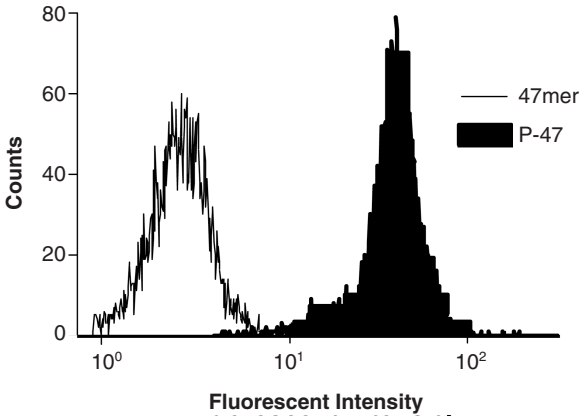


**Figure 7** Protein domain display. Beads were coated with a synthetic peptide representing the 47 C-terminal amino acids of the FPR (phosphorylated or not), and fluoresceinated arrestin was allowed to bind to each of the beads. Beads with the phosphorylated peptide bound more arrestin, as expected from whole protein experiments.

### Proof of Principle for β2AR and Other GPCRs

To extend studies carried out with FPR to other GPCRs, we have engineered a group of GPCR-GFP chimeric proteins. To date, the most important of these has been β2AR. β2AR is arguably the most studied and best understood of all GPCRs with well over 1000 listings in PubMed. Despite the extensive literature concerning β2AR, there are still many fundamental issues regarding its activation and function that remain poorly understood. In addition to reconstitution studies with FPR, similar studies with β2AR have provided significant insights into GPCR function as well as providing a novel screen for agonists and antagonists of this receptor. We have used FPR-GFP and β2AR-GFP chimeric proteins for bead-based assemblies and screening in component 2.

## 4 SOLUBLE AND BEAD-BASED ASSEMBLIES

Component 2 (soluble and bead-based assemblies) requires several elements. These include: (1) solubilizing and characterizing GPCRs produced by component 1, (2) evaluating and implementing particle chemistry for GPCR display, (3) validating physiology of expressed GPCRs by comparing cell and bead results, (4) analyzing quantitative aspects of each assembly ($K_d$, kinetics, etc), and (5) miniaturizing the protocols for each assembly to be compatible with screening by component 3. We have now succeeded in displaying solubilized GPCR complexes on beads in several ways (Fig. 8) and

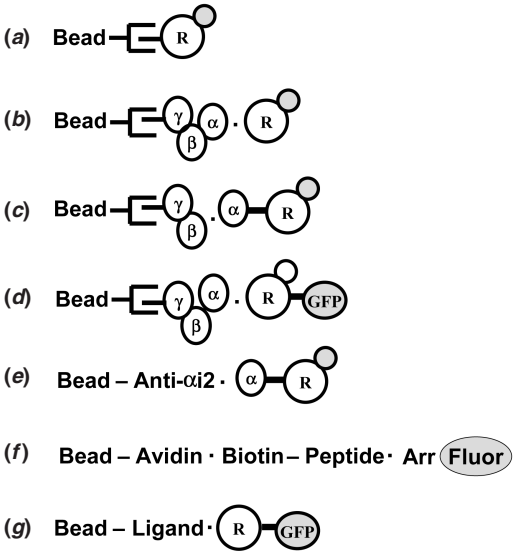


**Figure 8** Representative bead assemblies. The fluorescent moiety in each case is shaded. Dots represent noncovalent interactions. These assemblies are described in the text.

**TABLE 3 Principles for Protein Display on Particles**

| | |
|---|---|
| Bead surface | Hydrophilic preferred (dextran); streptavidin coated (not naked hydrophobic) latex works; we have not tried to coat functionalized latex. |
| Bead stability | Latex or dextran preferred; silica/glass may shatter. |
| Bead chemistry | Dextran compatible with chelated Ni, ligands, GST, fluorescent labels for multiplex. |
| Reagent display | Displaying the most expensive pure protein or ligand is preferred with nonspecific Ni chelate beads (cost efficient). |
| Reagent capture (GSH, biotin) | GSH and SA beads may be useful to capture proteins in a single step from cell lysates and display the protein for assembly and flow cytometry analysis. |

made some observations regarding the principle of molecular display on particles (Table 3).

1. His6-tagged FPR were directly displayed on Ni-chelate beads and detected with fluorescent formyl peptides [35] (Fig. 8*a*).
2. G-protein βγ subunits were displayed using his6-tags on Ni-chelate beads or FLAG tags captured with biotinylated anti-FLAG antibody on SA beads. By adding $G_\alpha$ subunits, FPR and its fluorescent ligand (Fig. 8*b*, shaded ball) were captured [36, 37].

3. His6-tagged G-protein βγ subunits captured the FPR-$G_{\alpha i2}$ fusion protein (Fig. 8*c*), which was detected with fluorescent formyl peptide ligand [36].
4. His6-tagged G-protein βγ subunits, displayed with $G_{\alpha i}$ subunits, captured the GFP-FPR fusion protein in the presence of nonfluorescent ligand (Fig. 8*d*). This assembly has been used for other GFP-GPCR fusion proteins [36, 37].
5. Anti-$G_{\alpha i2}$ antibody was displayed on antibody-capture beads to bind the FPR-$G_{\alpha i2}$ fusion protein, and was detected by fluorescent formyl peptide ligand (Fig. 8*e*). This assembly was used to prove that the FPR and $G_{\alpha}$ subunits were in the fusion protein and serves as a representative alternative to Western blotting [33].
6. Biotinylated GPCR tail peptides were displayed on streptavidin beads for analysis of arrestin assembly and phosphorylation [34] (Fig. 7 and Fig. 8*f*).
7. The β2AR antagonist dihydroalprenolol (DHA) was displayed on dextran beads to capture β2AR-GFP [37] (Fig. 8*g*).
8. FPR and bacteriorhodopsin were reconstituted into lipid bilayers enclosing beads (not shown; unpublished observation).

Taken together, these displays represent an essentially complete proof of principle set for display of GPCRs with signal transduction and termination partners. We will discuss FPR and β2AR in more detail.

## FPR Assemblies Used to Characterize and Validate the Displays

FPR, our first soluble GPCR, served as a prototype because it allowed direct comparison between displays obtained with fluorescent peptide ligand and fusion proteins [36] (Fig. 8*a* to 8*d*). The display was generalizable, demonstrated with two choices of epitope tags on the βγ, two types of beads, three FPR constructs, three $G_{\alpha}$ subunits, and two $G_{\beta\gamma}$ subunits. The results suggested that ~100,000 $G_{\beta\gamma}$ subunits/bead were displayed in appropriate orientation to bind FPR. We evaluated the time and concentration dependence of ternary complex formation (G protein, FPR, and ligand). As a result, affinities of ligand for FPR, ligand–FPR complex for G protein, and FPR-$G_{i\alpha2}$ fusion protein for $G_{\beta\gamma}$, were found to be consistent with comparable assemblies in detergent suspension.

The performance of the assembly was assessed in applications representing their potential to reveal ternary complex mechanisms [36]. First, we showed for a family of ligands that the affinity of FPR for its ligands varied in parallel with the assembly of FPR and G protein. Second, we showed that the beads could be used as a sensor to report the availability of FPR in solution, measuring competition between G protein on the bead and soluble G protein. Third, we showed kinetic discrimination between individual steps

of ternary complex activation and disassembly. We showed that the disassembly of FPR from G protein in the presence of GTPγS and ligand occurs on the subsecond time frame. We also showed that the assembly was sensitive to the presence of RGS proteins (regulator of G protein signaling) in a manner consistent with their known ability to stimulate GTPase activity of the G proteins [36, 38].

## β2AR Extends the Soluble GPCR on a Bead Concept

We solubilized the β2AR-GFP fusion protein and used it to approach a fundamental goal of drug discovery, the discrimination of agonist, antagonist, and a partial agonist, potentially in a single step [37]. For these assemblies we established two novel bead display systems and developed a multiplex approach for simultaneous analysis of the two assemblies, ligand beads (Fig. 8*g*) and G protein beads (Fig. 8*d*) using $G_{s\alpha}$ instead of $G_{i\alpha}$. We developed a ligand on a bead assembly by conjugating an antagonist (DHA) to a dextran bead. The assembly of β2AR-GFP was sensitive to the presence of agonist or antagonist (Fig. 9) and measured ligand affinity. The G-protein assembly (Fig. 10) was sensitive to the presence of agonists and resolved partial and full agonists. The behavior of the complexes on particles was essentially identical to the membrane-bound form of the receptor (Table 4) indicating that the physiological structure of β2AR is retained after solubilization. By modifying one set of beads to fluoresce red or using a different size [38] to resolve the two assays,

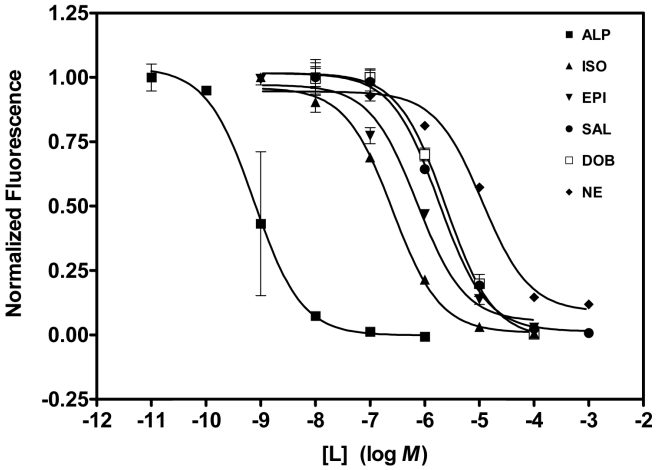


**Figure 9** DHA bead competitive assay. Competition curves were generated as increasing ligand concentrations competed the β2AR-GFP off DHA beads. $K_d$ values for nonfluorescent ligands were calculated from these data. The ligands used were alprenolol (ALP), isoproterenol (ISO), epinephrine (EPI), salbutamol (SAL), dobutamine (DOB), and norepinephrine (NE).

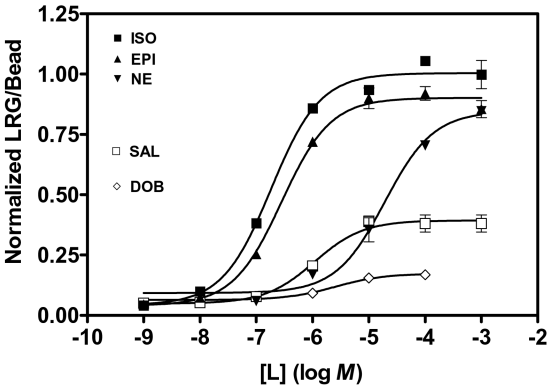


**Figure 10** G-protein beads bind β2AR-GFP in the presence of β2AR agonists. Dose–response curves were generated as increasing agonist concentrations caused the β2AR-GFP to form ARG complexes on the G-protein beads. The ligands used were isoproterenol (ISO), epinephrine (EPI), norepinephrine (NE), salbutamol (SAL), and dobutamine (DOB).

**TABLE 4 β2AR Assembly**[a]

| Ligand | LR $K_d$ (nM) | LRG $EC_{50}$ (nM) | LRG Efficacy |
|---|---|---|---|
| ALP | <1.3 | NA | NA |
| ISO | 220 | 180 | 1.0 |
| EPI | 680 | 280 | 0.9 |
| NE | 19,000 | 19,000 | 0.9 |
| SAL | 2,300 | 4,000 | 0.3 |
| DOB | 2,400 | 2,600 | 0.2 |

[a] Data obtained from three experiments as described for Figure 9, using DHA beads to obtain $K_d$ values, and Figure 10, using G-protein beads to obtain $EC_{50}$ values.

we could discriminate agonist and antagonist in a single step [37]. While a single-step primary screen would resolve agonists and antagonists, a dose–response secondary screen would resolve full and partial agonists. Molecules that block interactions between β2AR and G protein, uniquely, would not be active in the assembly of Fig. 9 but would compete with agonist in Fig. 10. Mathematical modeling of the ternary complex data with the soluble receptor systems is consistent with the idea that agonists promote higher affinity interaction of β2AR with G proteins than do partial agonists.

## Display of Cell Cycle Proteins Demonstrates Further Generality

We collaborated with Dorota Skowyra (St. Louis University) on the system that controls cell cycle in yeast and has extensive homology with mammalian cell cycle control. There are three complexes: a substrate complex including

Sic, a ubiquitin ligase complex, and the proteasome complex that recognizes that ubiquitinated substrate, degrades it, and enables cell cycle progress. We have displayed the GST-tagged ubiquitin ligase through a GSH-bead, and evaluated binding of a GFP–Sic substrate complex [39]. Not only important in cell cycle, this display represents an opportunity to bridge the power of yeast genetics and HyperCyt.

## 5 HIGH-THROUGHPUT FLOW CYTOMETRY TECHNOLOGIES

We have developed two successive generations of HT flow cytometry sampling technology. The first, designated plug flow cytometry, used a reciprocating multiport flow injection valve to execute up to 10 endpoint assays per minute, 4 online mixing experiments per minute, and, in secondary screens, a 15-point concentration gradient of soluble compound in ~2 min [40–43]. The second-generation technology [19], designated HyperCyt, uses a peristaltic pump in combination with an autosampler to boost endpoint assay performance to rates in excess of 1 sample/s (Fig. 2). HyperCyt hardware and software have since been more fully developed and characterized for high content, microfluidic mixing, cell sorting, carryover, and screening.

### HyperCyt Exhibits Versatility and Compatibility with High Content

We have validated cell-based endpoint assays for ligand binding, surface antigen expression, immunophenotyping, and cell adhesion. For example, at analysis rates of 1.5 s/sample a multiparameter (high content) lymphocyte immunophenotyping microassay in 96-well plates provided immunofluorescence results comparable to manual analysis (Fig. 11) [18]. High-content bead assays have been prototyped (not shown). In validation experiments with Terasaki microplates (sixty 10-μL wells per plate), all 60 cell samples were clearly resolved and reproducibly quantified over a 90-s analysis period [18]. The volume of the Terasaki wells is large enough that evaporative loss of fluid is not a problem but small enough that quantity-limited cells and reagents may be used at otherwise prohibitively high concentrations. Terasaki and 96-well "crystallization" microplates also enable a novel yet simple solution to the problem of cell settling. Uniform particle suspensions are maintained for an hour or more by periodically inverting the microplates in a "Ferris wheel" apparatus [18] (Fig. 2). Liquid surface tension prevents sample loss from inverted wells. This microvolume technology led to the potential application of HyperCyt in disease endpoint assays (signal transduction intermediates, cell adhesion) that can efficiently use limited cell numbers available from patient samples.

### HyperCyt Adds Microfluidic Mixing for Analysis of Soluble Compounds

We have described two novel microfluidic mixing approaches for use in conjunction with HyperCyt. These approaches have been characterized with

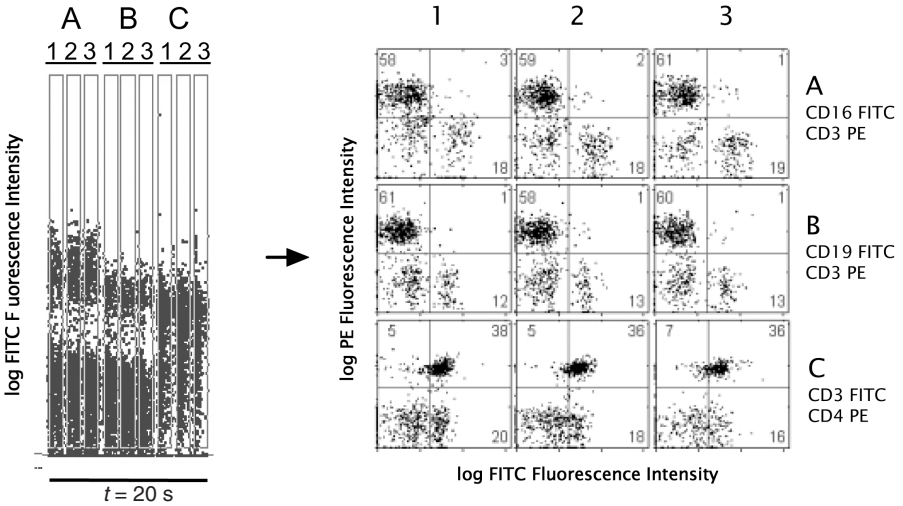


**Figure 11** High-content, high-throughput immunophenotyping with HyperCyt. (*Left*) One parameter output with time. (*Right*) Replotted data, with high-content two-parameter output.

fluorescent cell calcium responses to sampled peptides as well as with a novel bead-based assay [biotin displacement of fluorescein isothiocyanate (FITC)-biotin on SA beads] that provides a fluorescence response analogous in time and kinetics to a cell calcium response [44]. The first approach used the pulsatile delivery of the HyperCyt peristaltic pump to generate mixing of particles with fluid compounds sampled from microplate wells [20]. The second approach uses a 100-μm magnetic "flea" contained within the HyperCyt delivery tubing and a magnetic stirrer [21]. Soluble compounds can be sampled from microplate wells and processed at rates to 10 samples/min [21].

## HyperCyt Cell Sorting Extends Microfluidic Mixing

We have interfaced HyperCyt with a MoFlo flow cytometer and achieved online mixing combined with sorting of responding cells (Fig. 12). A relatively long sample aspiration time (10 s) followed by 3 × 3 s intersample washing enabled screening of FPR responses at a rate of 3 peptide concentrations/min with ~10,000 cells analyzed/concentration. When cells highly responsive to peptide were collected by sorting and expanded in culture for 12 to 30 days, the sort-purified cells showed enhanced sensitivity to low peptide concentrations and more sustained responses to all stimulatory peptide concentrations compared to the unsorted cell population from which they were derived [27]. A major technical advance changed the configuration of a flow nozzle to be compatible with the air bubbles between samples created by HyperCyt.

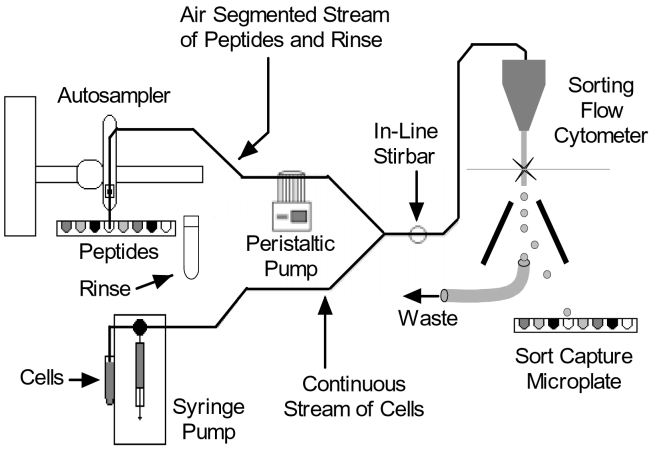


**Figure 12** HyperCyt mixing and sorting to isolate ligand-responsive cells.

## Sample Carryover in Microfluidic Mixing Systems

HyperCyt can deliver fluid samples of soluble compounds to mix with a continuous stream of cells at rates as fast as an autosampler can move from well to well. However, overall sample throughput is limited by unique aspects of intersample fluid carryover associated with the air-bubble segmented fluid flow of HyperCyt. Using a system in which samples are delivered to a spectrofluorometer, it has been possible to isolate and characterize factors important in such fluid carryover [23]. The most critical factor appears to be tubing internal surface area, while tubing junctions are of lesser importance (a contrast to particle carryover). Fluid carryover arises from a thin fluid film left on the inner tubing surface as the sample transits the length of the tube. The film acts as a transport mechanism to hold a small fraction of fluid from the sample as it passes by and deposits it into the subsequent sample.

## Software Development

Software development has been central to progress in the automation and implementation of HyperCyt. FCSQuery, a Visual $C^{++}$ program, was initially created to support the novel concept of time-resolved multisample data collection in which flow cytometry data from up to 96 samples are collected and stored in a single time-resolved data file. The program automatically detects time-resolved data clusters corresponding to individual samples, quantifies a range of statistical parameters for each sample and outputs the summary data to a Microsoft Excel spreadsheet. We are unaware of any commercial flow cytometry software with this capability.

A second Visual C$^{++}$ program, CRFSoft, was created to enable real-time synchronization of the multiple components of the HyperCyt system. This is a multithreaded program by which a laptop computer controls the flow cytometer data acquisition computer, the peristaltic pump, and the autosampler while monitoring fluorescent "feedback" signals directly from the flow cytometer in real time [45]. In the initial implementation of this software, fluorescent beads sampled at the beginning and end of autosampling served as the synchronization signal to initiate and terminate data acquisition.

### FPR Ligand Screening Assay

To validate HyperCyt for compound screening applications, FPR has been mutated to block internalization. The detection reagent is a fluorescein-conjugated peptide, fMLFK-FITC, that binds the FPR mutant with a $K_d$ ~3 nM. The objective is to detect potential anti-inflammatory compounds analogous to the anti-inflammatory drug, cyclosporin H, that block binding of fMLFK-FITC to the FPR. The cells were distributed in a 96-well microplate, incubated with unlabeled blocking peptides, then added with 1.5 nM fMLFK-FITC to quantify free FPR. To determine the limits of assay sensitivity, two blocking peptides with low and intermediate FPR affinity were used between $10^{-7}$ and $10^{-4}$ M: 2Pep (fML, inhibition constant, $K_i$ ~6000 nM) and 3Pep (fMLF, $K_i$ ~170 nM). With HyperCyt sampling at 1 well/s, the resulting 96 time-resolved event clusters (Fig. 13*a*) were analyzed to determine the median fluorescence intensity (MFI) of FPR-bound fMLFK-FITC in each well (Fig. 13*b*). The MFI was 69 ± 10 fluorescence units (FU) in the absence of blocking peptide (Fig. 13*b*, solid line), while a high FPR affinity blocking peptide (4Pep, $K_i$ ~3 nM) reduced it to 6 ± 1 FU (Fig. 13*b* and 13*c*). Blocking was considered significant if the MFI was more than 3 standard deviation (SD) below the mean of unblocked cells (Fig. 13*b*, dashed line) in all 8 replicate wells. This occurred with $10^{-4}$ M 2Pep and $10^{-4}$ to $10^{-6}$ M 3Pep. These results indicate that our screening assay will reliably detect compounds with micromolar $K_i$'s. We have subsequently applied HyperCyt to screen the Prestwick Chemical Library (PCL) and found a single positive compound, sulfinpyrazone, which was previously identified as a member of a series of FPR antagonists (see below). This result confirms our approach with respect to the assay and the technology.

## 6 VIRTUAL SCREENING

Virtual screening [46, 47] is becoming, de facto, a complement to bioactivity screening [48]. We view the ability to perform virtual screening as relying on docking compounds and scoring binding interactions in the context of chemical databases and high-performance computing.

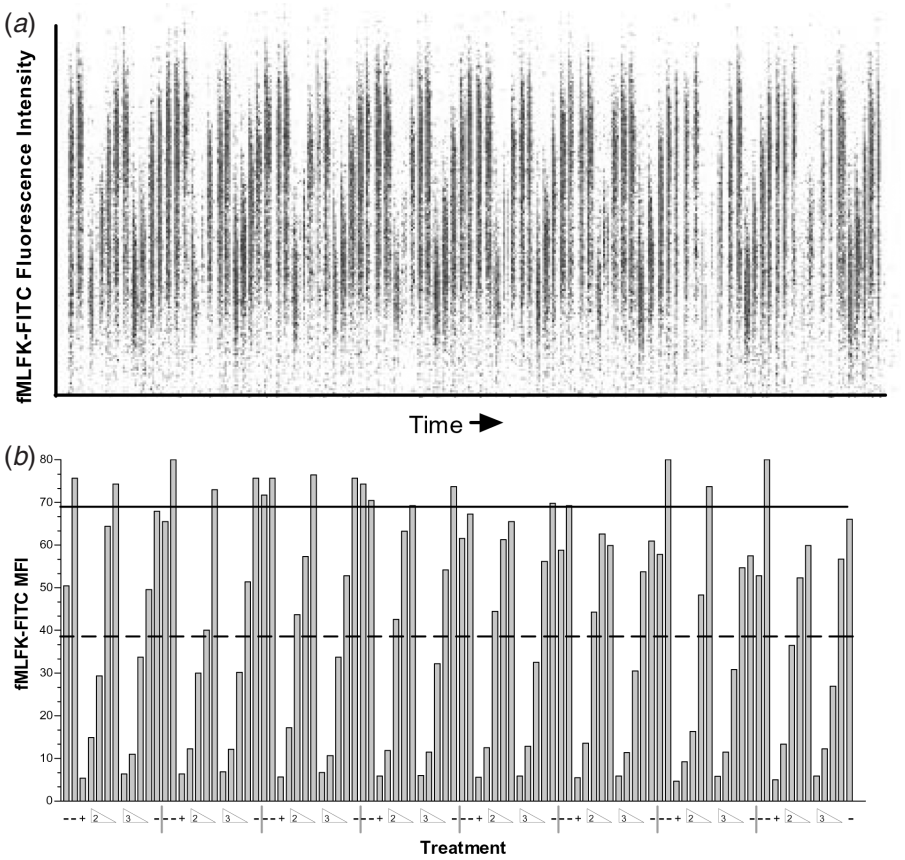


(c)

| Row | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Symbol | - | - | + | 2 | | | | 3 | | | | - |
| Block (M) | None 0 | None 0 | 4Pep $10^{-3}$ | 2Pep $10^{-4}$ | 2Pep $10^{-5}$ | 2Pep $10^{-6}$ | 2Pep $10^{-7}$ | 3Pep $10^{-4}$ | 3Pep $10^{-5}$ | 3Pep $10^{-6}$ | 3Pep $10^{-7}$ | None 0 |
| MFI | 62 | 77 | 6 | 13 | 36 | 56 | 69 | 6 | 12 | 31 | 54 | 68 |
| SD | 9 | 9 | 1 | 2 | 10 | 9 | 7 | 1 | 1 | 2 | 2 | 7 |

**Figure 13** FPR ligand screening assay and performance. Two test peptides with low (2Pep or fML) and intermediate (3Pep or fMLF) FPR binding affinity were assessed for competitive binding activity in the FPR assay against 4Pep-FITC (fMLFK-FITC), using FPR-expressing cells. (*a*) Time-resolved clusters of cell fluorescence data, one cluster for each well, that were produced in conjunction with HyperCyte analysis. (*b*) Plot of the mean fluorescence intensity (MFI) of cells from each of the wells. The lines below the graph delineate each 12-well series. (*c*) Setup of the 12-well series: 2 unblocked controls, 1 blocked control, 4 concentrations of each test peptide, and one last unblocked control. This series was repeated 8 times on a 96-well plate.

## Docking and Scoring

A steroidal scaffold that binds to β-tubulin, an anticancer target previously known to bind taxol and its derivatives [49], was identified by virtually screening the NCI (National Cancer Institute) open database. Given a three-dimensional (3D) structure of the receptor, virtual screening relies on docking and scoring to provide candidates for further analysis. DOCK [50], GOLD [51], FlexX [52], AutoDock [53], and FRED (www.eyesopen.com) are among the most widely used docking programs in the field. Most of these programs can dock libraries of single structures or even multiconformer libraries and are thus suitable for high-throughput searches.

Regardless of the choice of the docking software, the ability to score the ligands in an appropriate manner is limited. There are four major categories of scoring functions. First, knowledge-based methods [54–56] use Boltzman-weighted potentials of mean force, derived from statistical analyses of ligand–receptor interatomic contacts, based on available complexes in the Protein Data Bank (PDB) [57]. For example, SMoG [58–60], Muegge [61, 62], and DrugScore [63, 64] are implementations of this approach. Second, "Master equation" approaches [65] are semiquantitative estimates of the energetic contributions of various interaction types. The Williams approach [66, 67] was targeted at peptide–peptide interactions [68], while Rose devised a scoring function intended for protein targets [69, 70]. Third, regression-based methods [71, 72] take advantage of the available biological activity for training sets of ligand–receptor complexes extracted from the PDB. These include SCORE2 [72], VALIDATE [73], VALIDATEII [74], Pro_Score [75, 76], Jain [77], Horvath [46], and SCORE [78, 79]. Fourth, Poisson–Boltzman equation solvers address electrostatics and incorporate solvent effects, for example, ZAP [80]. Regression-based methods require biological input for parameterization, whereas the other three approaches do not. While perceived as a disadvantage, this type of scoring outperforms the others whenever biological results can capture target-specific information. None of the scoring schemes has been shown to be general, thus consensus scoring [81] has gained popularity. Recently, pharmacokinetic awareness [82] was also integrated into a novel scoring scheme [83] that allows for simultaneous optimization of binding affinity and pharmacokinetic properties.

## Generation and Maintenance of Chemical Databases for VS

WOMBAT (WOrld of Molecular BioAcTivity) is a database that captures chemical structures and biological activities published in the medicinal chemistry literature. Its current release, WOMBAT 2003.2, contains 53,126 entries (47,872 unique structures), totaling 98,662 biological activities on 506 unique targets. Its activity list also includes 236 inactive compounds, 7982 activities "less than" and 159 activities "greater than" entries. WOMBAT 2003.2 contains 2148 different series from 2143 papers published in medicinal chemistry

**TABLE 5 Target Class Distribution in WOMBAT 2003.2**

| Target Class | Compounds | Percentage |
|---|---|---|
| G-protein coupled receptors | 19,839 | 37.3 |
| Ion channels | 6,090 | 11.5 |
| Aspartyl proteases | 2,656 | 5.00 |
| Serine proteases | 2,582 | 4.86 |
| Kinases | 2,111 | 3.97 |
| Transporters | 1,689 | 3.18 |
| Cysteine proteases | 504 | 0.95 |
| Nuclear hormone receptors | 448 | 0.84 |
| Others | 19,233 | 36.2 |

journals between 1975 and 2002 (over 90 percent coming from *Journal of Medicinal Chemistry*). The target class distribution of WOMBAT, a product from Sunset Molecular Discovery LLC (www.sunsetmolecular.com), is given in Table 5. A subset of 21,700 entries from WOMBAT, representing the first 40 percent of the database, was used to benchmark virtual screening tools based on FRED (Fast Rigid Exhaustive Docking), at the UNM High Performance Computing Center (HPC@UNM, www.hpc.unm.edu/).

CADB08 (Commercially Available Data Base version 0.8) contains 877,876 small, nonpeptide chemical structures. These structures, or their derivatives, are potentially available for purchase. CADB08 structures were sorted in the increasing order for rotatable bonds and molecular weight, then converted from SMILES [84] to 3D structures with Omicron [85], the OpenEye (OE) 3D-structure generator. The Omicron-generated structures were then submitted to OMEGA, the OE multiple-conformer generator, to sample conformational space for individual molecules, with an increment of 5 structures per rotor. We applied an energetic cutoff of 5 kcal to ensure that (mostly) room-temperature accessible conformers, allowing a computational error of 3 kcal, are represented for each molecule. Multiple conformers (occupying ~30 GB of disk space) are needed for docking with OE's FRED since this software is geared to rigid structures. The 1-conformer version of CADB08 was used to evaluate the performance of our virtual screening tools.

ChemBank has accumulated over 300,000 natural-product-like structures. CDLDB is the chemical collection from Chemical Diversity Labs, Inc., (CDL), from San Diego, CA (www.chemdiv.com). CDL has been marketing GPCR-focused libraries since 2001. The Prestwick Chemical Library (PCL) is an 880-compound library produced by Prestwick Chemical of Strasbourg, France. PCL contains out-of-patent drugs (85 percent) and known alkaloids (15 percent).

### Virtual Screening in the Context of High-Performance Computing

CADB08 was virtually screened on Los Lobos, a nationally ranked Linux cluster with 256 dual-processor IBM Netfinity 4500 R (Intel Pentium-III 733

MHz) nodes, with Myrinet connectivity and 2 TB of storage space, from HPC@UNM. The OE software suite was ported to the MPI (message parsing interface) by Andrew Pineda (HPC@UNM), who wrote an MPI-based wrapper code in C for this task [86]. We note that virtual screening is considered an "embarassingly parallel" problem. The OE-MPI wrapper code manages output files from serial runs, summarizes the results (e.g., RMS values and scoring from docking runs), and provides a balanced workload across processors (Fig. 14). It handles errors (structures that cannot be docked are ignored without interrupting the virtual screening run), while making use of the high-speed communications network (Myrinet) on Los Lobos. OE-MPI is flexible, since we recently incorporated ROCS (Rapid Overlay of Chemical Structures [85]) into the VS protocol. ROCS is a Gaussian-shape volume overlap filter that limits the number of molecules, or conformers, to be docked by FRED into the binding site by first matching their 3D shape with the binding site. This can speed up the calculation by several orders of magnitude.

The OE-MPI code was benchmarked (Fig. 15) by Andy Pineda, Bryan Shiloff, and Tudor Oprea, against 21,700 molecules from WOMBAT, using the X-ray structures of the ligand-binding domain of the estradiol receptor α (complexed with the agonist estradiol) [87] (PDB: 1A52) and with the antagonist 4-hydroxy-tamoxifen [88] (PDB: 3ERT) using the protocol described in Fig. 16. SMILES input was sent to OMEGA for 1-conformer generation, then to FRED for docking and to FredA for scoring. This protocol recovered all

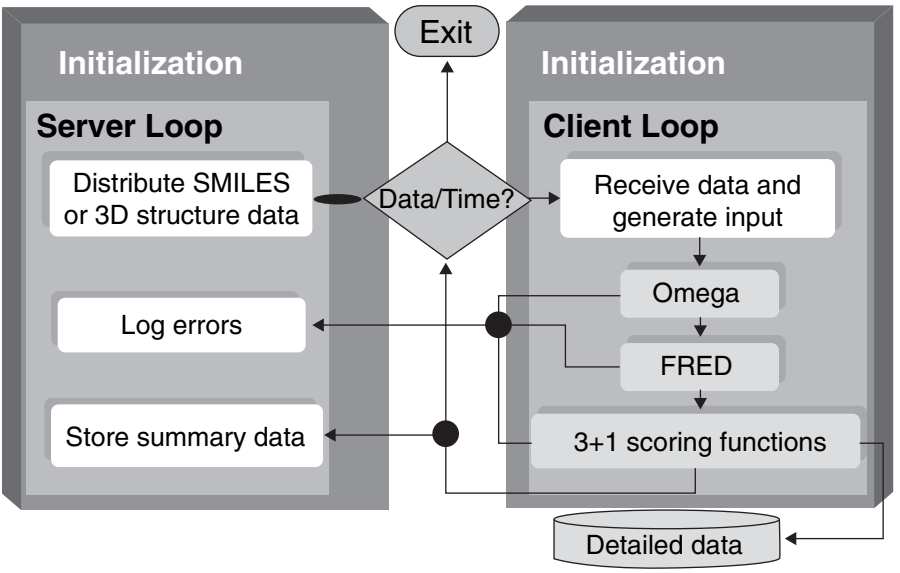


**Figure 14** Client/server architecture for the OE-MPI code used to perform VS on Linux clusters.

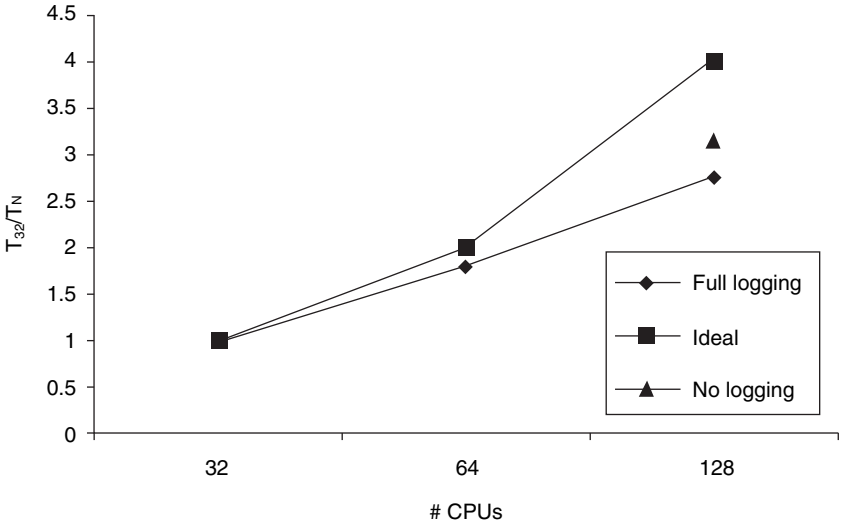


**Figure 15** Evaluation of parallel speed increase for OE-MPI software. 21,000 SMILES from WOMBAT were used as a benchmark. The deviation from ideal is likely due to slow network I/O. Best performance is about 78 percent of ideal. 90+ percent should be attainable with I/O tuning. Best throughput was 67 SMILES per CPU/h (~207,000 molecules/day on 128 processors).

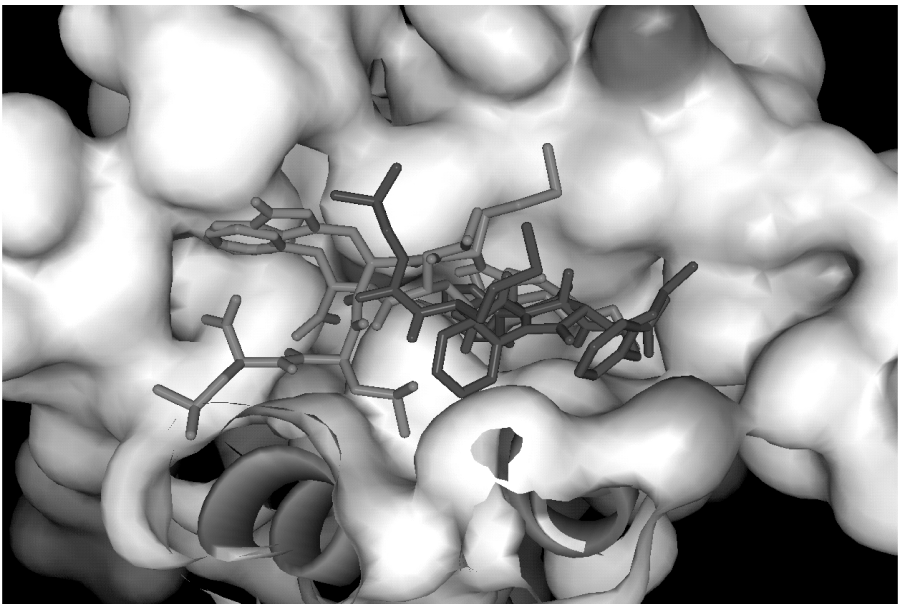

**Figure 16** Binding modes of two formyl peptide agonists (light gray) and two formyl peptide antagonists (dark gray; cyclosporin H and phenylbutanone).

five ER-α antagonists (nM) present in the 21,700 WOMBAT, in the top 10 FredA results, using the consensus scoring method. All 56 known ER-α actives from the 21,700 WOMBAT entries were recovered in the top 5000 molecules. This is not surprising since most of these control molecules are active only at the micromolar level. Random screening would require 21,700 experiments to recover the entire set. Screening <25 percent of the database to recover all actives is considered a good result, given the inaccuracies of the scoring functions. On the down side, FRED docked one conformer of Raloxifene (a potent ER-α antagonist) into the 1A52 structure (the agonist binding site), an artifact due to the high conformational flexibility of Raloxifene. However, Raloxifene was recovered among the top 3 hits of the 3ERT (tamoxifen, ER-α antagonist) docking.

We further evaluated the OE-MPI software on the 1-conformer version of CADB08, using the same protocol as in Fig. 14. The 3 scoring functions from FredA are: chemscore (regression), Gaussian shape score (steric), and screen-score (PMF); the +1 refers to the consensus score. OE-MPI generated 147,000 results per day on 198 processors (~30 molecules/CPU/h on <50 percent of the Los Lobos cluster). We were able to process 315,000 molecules in 2 consecutive runs. The difference between the benchmark results shown in Fig. 15 and the more realistic trials performed with CADB08 is due to increased chemical diversity and size. Given the combined size of CADB08 and ChemBank (1.2 million structures), and the simplifying assumption that each molecule has 50 conformers, we would require 408 CPU days on 198 processors per target at the rate described above. We anticipate that the introduction of the ROCS shape filter will reduce the VS effort by 50 percent or more.

## FPR Case Study

Preliminary results indicate that it is possible to derive homology models for GPCRs and use these models to identify from a chemical database candidates for biological testing. In the absence of co-crystallized inhibitors bound to each target, we are incorporating additional information (e.g., mutation studies made possible by component 1) in our binding site definitions. We are using all available experimental results to further increase the probability that our models are correct and validate these models against amino acid mutation studies. Evidence from cross-linking and mutagenesis suggests that the ligand binding site is located between transmembrane (TM) helices, near the extracellular face of the membrane [89–93]. Sequence alignments of GPCRs have shown that they share a high degree of homology in the TM domains. The analysis by Baldwin on ~500 GPCR sequences identified 36 amino acids that were conserved in >70 percent of GPCRs [94, 95]. Based on these highly conserved residues, we derived several homology models of FPR TM helices and of most of the protein (1 to 337 domain) using as a template the 3D structure of bovine rhodopsin [96]. We refined these models using the

experimental data from cross-linking and mutagenesis studies (not shown). The model that best agrees with cross-linking, mutagenesis, and ligand binding data, including only the 7 TM helices, was chosen for docking and pharmacophore modeling studies. Three agonists and four antagonists were docked in the FPR ligand binding site (Fig. 18) using AUTODOCK 3.05 [53]. The docking results suggested that antagonists have a different binding mode than agonists (Fig. 16), which allowed us to formulate a 3 point (one hydrogen bond, two hydrophobic) pharmacophore hypothesis specific for FPR antagonists.

Sulfinpyrazone, the PCL hit identified by HyperCyt, ranked eight in the virtual screening results (Table 6). Its measured affinity, 24 μM, confirms a 1991 report by Levesque et al. [97]. Our pharmacophore hypothesis, confirmed by sulfinpyrazone, formed the basis for selecting 4325 molecules from CDL collection (except for the GPCR-focused library) as an FPR-targeted subset that has, so far, yielded 68 primary hits from 9 distinct chemical series, and all but one are confirmed antagonists. Of these, 10 have measured $K_i$ values below 10 μM. When comparing the FPR screening results of PCL, a library where no target-specific selection was exercised, with ~0.1 percent hit rate, and those of CDL-based FPR library, where a pharmacophore was used for the selection, with ~1.6 percent hit rate. The primary hits in the FPR library show consistent structure–activity relationships, which allow us to further prioritize the development of compounds with higher activity.

## 7 SUMMARY

We have developed a team for discovery research that has four highly integrated components. In the first component, expertise in cell and molecular biology and biochemistry is coupled to research into physiological pathways that regulate signal transduction and receptor processing. This expertise contributes to the expression of proteins, and the creation of cell- and particle-based assays for screening and protein–protein interactions. In the second component, expertise in biophysics, particle chemistry, fluorescence, flow cytometry, and mathematical modeling is coupled to research into ligand–receptor and protein–protein interactions. This expertise creates particle-based assays for multiplexed analysis and discovery of small molecule activities such as the discrimination of full and partial agonists. In the third component, expertise in flow cytometry hardware, software, and microfluidics allowed for the development of a novel instrument platform compatible with cell- and particle-based analysis. This expertise generated screening capabilities that have high-throughput, high-content, and real-time elements for small-molecule discovery. In the fourth component, expertise in computing and cheminformatics couple with the other three components to improve the efficiency of the small-molecule discovery process through the integration of virtual screening, chemical databases, and high-performance computing.

**TABLE 6 Hits from Virtual Screening of the Prestwick Chemical Library for Binding to the FPR**

| Cpd ID | Structure | Molname | Cpd RANK | ID | Structure | Molname | RANK |
|---|---|---|---|---|---|---|---|
| 278 | 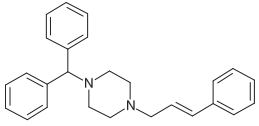 | Cinnarizine | 1 | 312 | 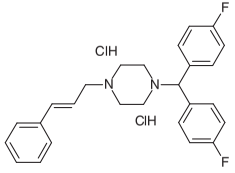 | Flunarizine | 6 |
| 308 | 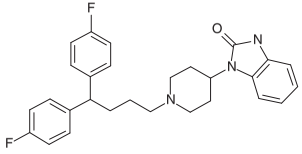 | Pimozide | 2 | 297 | 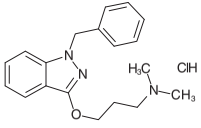 | Benzydamine | 7 |
| 292 | 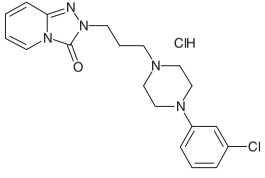 | Trazodone | 3 | 290 | 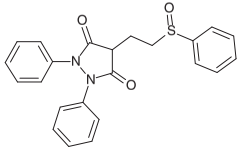 | Sulfinpyrazone | 8 |
| 320 | 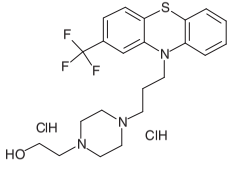 | Fluphenazine | 4 | 293 | 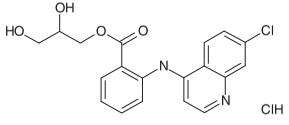 | Glafenine | 9 |
| 295 | 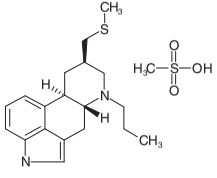 | Pergolide | 5 | 299 | 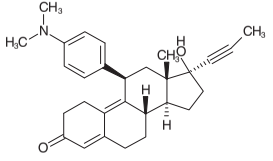 | Mifepristone | 10 |

## REFERENCES


1. Nolan, J. P., Sklar, L. A. (2002). Suspension array technology: Evolution of the flat-array paradigm. *Trends Biotechnol.*, *20*, 9–12.
2. Luecke, H., Schobert, B., Lanyi, J. K., Spudich, E. N., Spudich, J. L. (2001). Crystal structure of sensory rhodopsin II at 2.4 angstroms: Insights into color tuning and transducer interaction. *Science*, *293*, 1499–1503.
3. Vaidehi, N., Floriano, W. B., Trabanino, R., Hall, S. E., Freddolino, P., Choi, E. J., Zamanakos, G., Goddard, W. A., 3rd (2002). Prediction of structure and function of G protein-coupled receptors. *Proc. Nat. Acad. Sci. USA*, *99*, 12622–12627.
4. Penn, R. B., Pronin, A. N., Benovic, J. L. (2000). Regulation of G protein-coupled receptor kinases. *Trends Cardiovasc. Med.*, *10*, 81–89.
5. Ferguson, S. S. (2001). Evolving concepts in G protein-coupled receptor endocytosis: The role in receptor desensitization and signaling. *Pharmacol. Rev.*, *53*, 1–24.
6. Pierce, K. L., Premont, R. T., Lefkowitz, R. J. (2002). Seven-transmembrane receptors. *Nat. Rev. Mol. Cell Biol.*, *3*, 639–650.
7. Ross, E. M., Wilkie, T. M. (2000). GTPase-activating proteins for heterotrimeric G proteins: Regulators of G protein signaling (RGS) and RGS-like proteins. *Ann. Rev. Biochem.*, *69*, 795–827.
8. Neubig, R. R. (2002). Regulators of G protein signaling (RGS proteins): Novel central nervous system drug targets. *J. Peptide Res.*, *60*, 312–316.
9. Sklar, L. A. (1987). Real-time spectroscopic analysis of ligand-receptor dynamics. *Annu. Rev. Biophys. Biophys. Chem.*, *16*, 479–506.
10. Sklar, L. A., Edwards, B. S., Graves, S. W., Nolan, J. P., Prossnitz, E. R. (2002). Flow cytometric analysis of ligand-receptor interactions and molecular assemblies. *Annu. Rev. Biophys. Biomol. Struct.*, *31*, 97–119.
11. Prossnitz, E. R., Sklar, L. A. (2004). Chemoattractant/complement receptors. In W. J. Lennarz and M. D. Lane (ed.), *Encyclopedia of Biological Chemistry*, vol. 1. pp. 425–429. Elsevier Science, San Diego.
12. Vilven, J. C., Domalewski, M., Prossnitz, E. R., Ye, R. D., Muthukumaraswamy, N., Harris, R. B., Freer, R. J., Sklar, L. A. (1998). Strategies for positioning fluorescent probes and crosslinkers on formyl peptide ligands. *J. Receptor Signal Transduct. Res.*, *18*, 187–221.
13. Vanti, W. B., Nguyen, T., Cheng, R., Lynch, K. R., George, S. R., O'Dowd, B. F. (2003). Novel human G-protein-coupled receptors. *Biochem. Biophys. Res. Commun.*, *305*, 67–71.
14. Mombaerts, P. (2001). The human repertoire of odorant receptor genes and pseudogenes. *Annu. Rev. Genom. Hum. Genet.*, *2*, 493–510.
15. Dryer, L., Berghard, A. (1999). Odorant receptors: A plethora of G-protein-coupled receptors. *Trends Pharmacol. Sci.*, *20*, 413–417.
16. Xia, C., Ma, W., Wang, F., Hua, S., Liu, M. (2001). Identification of a prostate-specific G-protein coupled receptor in prostate cancer. *Oncogene*, *20*, 5903–5907.
17. Estrada, M., Espinosa, A., Muller, M., Jaimovich, E. (2003). Testosterone stimulates intracellular calcium release and mitogen-activated protein kinases via a G protein-coupled receptor in skeletal muscle cells. *Endocrinology*, *144*, 3586–3597.

18. Ramirez, S., Aiken, C. T., Andrzejewski, B., Sklar, L. A., Edwards, B. S. (2003). High-throughput flow cytometry: Validation in microvolume bioassays. *Cytometry*, *53A*, 55–65.
19. Kuckuck, F. W., Edwards, B. S., Sklar, L. A. (2001). High throughput flow cytometry. *Cytometry*, *44*, 83–90.
20. Jackson, W. C., Kuckuck, F., Edwards, B. S., Mammoli, A., Gallegos, C. M., Lopez, G. P., Buranda, T., Sklar, L. A. (2002). Mixing small volumes for continuous high-throughput flow cytometry: Performance of a mixing Y and peristaltic sample delivery. *Cytometry*, *47*, 183–191.
21. Jackson, W. C., Bennett, T. A., Edwards, B. S., Prossnitz, E., Lopez, G. P., Sklar, L. A. (2002). Performance of in-line microfluidic mixers in laminar flow for high-throughput flow cytometry. *Biotechniques*, *33*, 220–226.
22. Truesdell, R. A., Vorobieff, P. V., Sklar, L. A., Mammoli, A. A. (2003). Mixing of a continuous flow of two fluids due to unsteady flow. *Phys. Rev. E*, *67*, 066304.
23. Bartsch, J. W., Waller, A., Mammoli, A. A., Buranda, T., Sklar, L. A., Tran, H. D., Edwards, B. S. (2004). An Investigation of liquid carry-over and sample residual for the Hypercyt® sample delivery system. *Anal. Chem.*, *76*, 3810–3817.
24. Truesdell, R., Bartsch, J., Buranda, T., Sklar, L. A., Mammoli, A. (In press). Direct measurement of mixing quality in pulsatile flow micromixer. *Fluids*.
25. Liu, R., Marik, J., Lam, K. S. (2002). A novel peptide-based encoding system for "one-bead one-compound" peptidomimetic and small molecule combinatorial libraries. *J. Am. Chem. Soc.*, *124*, 7678–7680.
26. Iannone, M. A., Consler, T. G., Pearce, K. H., Stimmel, J. B., Parks, D. J., Gray, J. G. (2001). Multiplexed molecular interactions of nuclear receptors using fluorescent microspheres. *Cytometry*, *44*, 326–337.
27. Young, S. M., Curry, M. S., Ransom, J. T., Ballesteros, J., Prossnitz, E. R., Sklar, L. A., Edwards, B. S. (2004). High throughput microfluidic mixing and multi-parametric sorting for bioactive compound screening. *J. Biomol. Screening*, *9*, 103–111.
28. Edwards, B. S., Andrzejewski, B., Sklar, L. A. (In preparation). Automation for Hypercyte®.
29. Schreiber, S. L. (2000). Target-oriented and diversity-oriented organic synthesis in drug discovery. *Science*, *287*, 1964–1969.
30. Key, T. A., Bennett, T. A., Foutz, T. D., Gurevich, V. V., Sklar, L. A., Prossnitz, E. R. (2001). Regulation of formyl peptide receptor agonist affinity by reconstitution with arrestins and heterotrimeric G proteins. *J. Biol. Chem.*, *276*, 49204–49212.
31. Bennett, T. A., Foutz, T. D., Gurevich, V. V., Sklar, L. A., Prossnitz, E. R. (2001). Partial phosphorylation of the N-formyl peptide receptor inhibits G protein association independent of arrestin binding. *J. Biol. Chem.*, *276*, 49195–49203.
32. Key, T. A., Foutz, T. D., Gurevich, V. V., Sklar, L. A., Prossnitz, E. R. (2003). N-formyl peptide receptor phosphorylation domains differentially regulate arrestin and agonist affinity. *J. Biol. Chem.*, *278*, 4041–4047.
33. Shi, M., Bennett, T. A., Cimino, D. F., Maestas, D. C., Foutz, T. D., Gurevich, V. V., Sklar, L. A., Prossnitz, E. R. (2003). Functional capabilities of an N-formyl peptide receptor-G(alpha)(i)(2) fusion protein: Assemblies with G proteins and arrestins. *Biochemistry*, *42*, 7283–7293.

34. Potter, R. M., Key, T. A., Gurevich, V. V., Sklar, L. A., Prossnitz, E. R. (2002). Arrestin variants display differential binding characteristics for the phosphorylated N-formyl peptide receptor carboxyl terminus. *J. Biol. Chem.*, *277*, 8970–8978.
35. Sklar, L. A., Vilven, J., Lynam, E., Neldon, D., Bennett, T. A., Prossnitz, E. (2000). Solubilization and display of G protein-coupled receptors on beads for real-time fluorescence and flow cytometric analysis. *Biotechniques*, *28*, 976–980, 982–985.
36. Simons, P. C., Shi, M., Foutz, T., Cimino, D. F., Lewis, J., Buranda, T., Lim, W. K., Neubig, R. R., McIntire, W. E., Garrison, J., Prossnitz, E., Sklar, L. A. (2003). Ligand-receptor-G-protein molecular assemblies on beads for mechanistic studies and screening by flow cytometry. *Mol. Pharmacol.*, *64*, 1227–1238.
37. Simons, P. C., Biggs, S. M., Waller, A., Foutz, T., Cimino, D. F., Guo, Q., Neubig, R. R., Tang, W. J., Prossnitz, E. R., Sklar, L. A. (2004). Real-time analysis of ternary complex on particles: Direct evidence for partial agonism at the agonist-receptor-G protein complex assembly step of signal transduction. *J. Biol. Chem.*, *279*, 13514–13521.
38. Waller, A., Simons, P., Prossnitz, E. R., Edwards, B. S., Sklar, L. A. (2003). High throughput screening of G-protein coupled receptors via flow cytometry. *Combinatorial Chem. High Throughput Screening*, *6*, 389–397.
39. Deffenbaugh, A. E., Scaglione, K. M., Zhang, L., Moore, J. M., Buranda, T., Sklar, L. A., Skowyra, D. (2003). Release of ubiquitin-charged Cdc34-S-Ub from the RING domain is essential for ubiquitination of the SCF(Cdc4)-bound substrate Sic1. *Cell*, *114*, 611–622.
40. Edwards, B. S., Kuckuck, F., Sklar, L. A. (1999). Plug flow cytometry: An automated coupling device for rapid sequential flow cytometric sample analysis. *Cytometry*, *37*, 156–159.
41. Edwards, B. S., Kuckuck, F. W., Prossnitz, E. R., Okun, A., Ransom, J. T., Sklar, L. A. (2001). Plug flow cytometry extends analytical capabilities in cell adhesion and receptor pharmacology. *Cytometry*, *43*, 211–216.
42. Edwards, B. S., Sklar, L. A. (2001). Plug flow cytometry. In J. P. Robinson (ed.), *Current Protocols in Cytometry*. Wiley, New York, pp. 1.17.11–11.17.10.
43. Edwards, B. S., Kuckuck, F. W., Prossnitz, E. R., Ransom, J. T., Sklar, L. A. (2001). HTPS flow cytometry: A novel platform for automated high throughput drug discovery and characterization. *J. Biomol. Screening*, *6*, 83–90.
44. Buranda, T., Lopez, G. P., Simons, P., Pastuszyn, A., Sklar, L. A. (2001). Detection of epitope-tagged proteins in flow cytometry: Fluorescence resonance energy transfer-based assays on beads with femtomole resolution. *Anal. Biochem.*, *298*, 151–162.
45. Edwards, B. S., Andrzejewski, B., Ramirez, S., Sklar, L. A. (2002). Multi-threaded integration of high throughput flow cytometry autosampling and analysis. *Cytometry, Suppl. 11*, 120.
46. Horvath, D. (1997). A virtual screening approach applied to the search for trypanothione reductase inhibitors. *J. Med. Chem.*, *40*, 2412–2423.
47. Walters, W. P., Stahl, M. T., Murcko, M. A. (1998). Virtual screening—an overview. *Drug Discov. Today*, *3*, 160–178.
48. Mestres, J. (2002). Virtual screening: A real screening complement to high-throughput screening. *Biochem. Soc. Trans.*, *30*, 797–799.

49. Wu, J. H., Batist, G., Zamir, L. O. (2001). Identification of a novel steroid derivative, NSC12983, as a paclitaxel-like tubulin assembly promoter by 3-D virtual screening. *Anticancer Drug Design*, *16*, 129–133.
50. Ewing, T. J., Makino, S., Skillman, A. G., Kuntz, I. D. (2001). DOCK 4.0: Search strategies for automated molecular docking of flexible molecule databases. *J. Comput. Aided Mol. Des. 15*, 411–428.
51. Jones, G., Willett, P., Glen, R. C., Leach, A. R., Taylor, R. (1997). Development and validation of a genetic algorithm for flexible docking. *J. Mol. Biol.*, *267*, 727–748.
52. Kramer, B., Metz, G., Rarey, M., Lengauer, T. (1999). Ligand docking and screening with FlexX. *Med. Chem. Res.*, *9*, 463–478.
53. Osterberg, F., Morris, G. M., Sanner, M. F., Olson, A. J., Goodsell, D. S. (2002). Automated docking to multiple target structures: Incorporation of protein mobility and structural water heterogeneity in AutoDock. *Proteins*, *46*, 34–40.
54. Sippl, M. J. (1993). Boltzmann's principle, knowledge-based mean fields and protein folding. An approach to the computational determination of protein structures. *J. Comput. Aided Mol. Des.*, *7*, 473–501.
55. Sippl, M. J. (1995). Knowledge-based potentials for proteins. *Curr. Opin. Struct. Biol.*, *5*, 229–235.
56. Domingues, F. S., Koppensteiner, W. A., Jaritz, M., Prlic, A., Weichenberger, C., Wiederstein, M., Floeckner, H., Lackner, P., Sippl, M. J. (1999). Sustained performance of knowledge-based potentials in fold recognition. *Proteins, Suppl. 3*, 112–120.
57. Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., Bourne, P. E. (2000). The Protein Data Bank. *Nucleic Acids Res.*, *28*, 235–242.
58. DeWitte, R. S., Shakhnovich, E. I. (1996). SMoG: De Novo design method based on simple, fast, and accurate free energy estimates. 1. Methodology and supporting evidence. *J. Am. Chem. Soc.*, *118*, 11733–11744.
59. DeWitte, R. S., Ishchenko, A. V., Shakhnovich, E. I. (1997). SMoG: De novo design method based on simple, fast, and accurate free energy estimates. 2. Case studies in molecular design. *J. Am. Chem. Soc.*, *119*, 4608–4617.
60. Ishchenko, A. V., Shakhnovich, E. I. (2002). SMall Molecule Growth 2001 (SMoG2001): An improved knowledge-based scoring function for protein-ligand interactions. *J. Med. Chem.*, *45*, 2770–2780.
61. Muegge, I., Martin, Y. C. (1999). A general and fast scoring function for protein-ligand interactions: A simplified potential approach. *J. Med. Chem.*, *42*, 791–804.
62. Muegge, I., Martin, Y. C., Hajduk, P. J., Fesik, S. W. (1999). Evaluation of PMF scoring in docking weak ligands to the FK506 binding protein. *J. Med. Chem.*, *42*, 2498–2503.
63. Gohlke, H., Hendlich, M., Klebe, G. (2000). Knowledge-based scoring function to predict protein-ligand interactions. *J. Mol. Biol.*, *295*, 337–356.
64. Sotriffer, C. A., Gohlke, H., Klebe, G. (2002). Docking into knowledge-based potential fields: A comparative evaluation of DrugScore. *J. Med. Chem.*, *45*, 1967–1970.

65. Ajay, Murcko, M. A. (1995). Computational methods to predict binding free energy in ligand-receptor complexes. *J. Med. Chem.*, *38*, 4953–4967.
66. Williams, D. H., Cox, J. P. L., Doig, A. J., Gardner, M., Gerhard, U., Kaye, P. T., Lal, A. R., Nicholls, I. A., Salter, C. J., Mitchell, R. C. (1991). Toward the semi-quantitative estimation of binding constants—guides for peptide binding in aqueous solution. *J. Am. Chem. Soc.*, *113*, 7020–7030.
67. Williams, D. H., Bardsley, B. (1999). Estimating binding constants—The hydrophobic effect and cooperativity. *Perspect. Drug Discov. Des.*, *17*, 43–59.
68. Mackay, J. P., Gerhard, U., Beauregard, D. A., Maplestone, R. A., Williams, D. H. (1994). Dissection of the contributions toward dimerization of glycopeptide antibiotics. *J. Am. Chem. Soc.*, *116*, 4573–4580.
69. Rose, P. W. (1997). Scoring methods in ligand design. In *Second UCSF Course in Computer-Aided Mol. Design*, San Francisco.
70. Marrone, T. J., Luty, B. A., Rose, P. W. (2000). Discovering high-affinity ligands from the computationally predicted structures and affinities of small molecules bound to a target: A virtual screening approach. *Perspect. Drug Discov. Des.*, *20*, 209–230.
71. Verkhivker, G., Appelt, K., Freer, S. T., Villafranca, J. E. (1995). Empirical free energy calculations of ligand-protein crystallographic complexes. I. Knowledge-based ligand-protein interaction potentials applied to the prediction of human immunodeficiency virus 1 protease binding affinity. *Protein Eng. 8*, 677–691.
72. Bohm, H. J. (1998). Prediction of binding constants of protein ligands: A fast method for the prioritization of hits obtained from de novo design or 3D database search programs. *J. Comput. Aided Mol. Des.*, *12*, 309–323.
73. Head, R. D., Smythe, M. L., Oprea, T. I., Waller, C. L., Green, S. M., Marshall, G. R. (1996). VALIDATE: A new method for the receptor-based prediction of binding affinities of novel ligands. *J. Am. Chem. Soc.*, *118*, 3959–3969.
74. Marshall, G. R., Head, R., Ragno, R. (2000). Affinity prediction: The *sine qua non*. In E. D. Cera (Ed.), *Thermodynamics in Biology*. Oxford University Press, New York, pp. 87–111.
75. Eldridge, M. D., Murray, C. W., Auton, T. R., Paolini, G. V., Mee, R. P. (1997). Empirical scoring functions: I. The development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes. *J. Comput. Aided Mol. Des.*, *11*, 425–445.
76. Murray, C. W., Auton, T. R., Eldridge, M. D. (1998). Empirical scoring functions. II. The testing of an empirical scoring function for the prediction of ligand-receptor binding affinities and the use of Bayesian regression to improve the quality of the model. *J. Comput. Aided Mol. Des.*, *12*, 503–519.
77. Jain, A. N. (1996). Scoring noncovalent protein-ligand interactions: A continuous differentiable function tuned to compute binding affinities. *J. Comput. Aided Mol. Des.*, *10*, 427–440.
78. Wang, R. X., Liu, L., Lai, L. H., Tang, Y. Q. (1998). SCORE: A new empirical method for estimating the binding affinity of a protein-ligand complex. *J. Mol. Modeling*, *4*, 379–394.

79. Wang, R., Lai, L., Wang, S. (2002). Further development and validation of empirical scoring functions for structure-based binding affinity prediction. *J. Comput. Aided Mol. Des.*, *16*, 11–26.
80. Grant, J. A., Pickup, B. T., Nicholls, A. (2001). A smooth permittivity function for Poisson-Boltzmann solvation methods. *J. Computat. Chem.*, *22*, 608–640.
81. Wang, R., Wang, S. (2001). How does consensus scoring work for virtual library screening? An idealized computer experiment. *J. Chem. Inform. Comput. Sci.*, *41*, 1422–1426.
82. Oprea, T. I. (2002). Virtual screening in drug discovery: A viewpoint. *Molecules*, *7*, 51–62.
83. Zamora, I., Oprea, T., Cruciani, G., Pastor, M., Ungell, A. L. (2003). Surface descriptors for protein-ligand affinity prediction. *J. Med. Chem.*, *46*, 25–33.
84. Weininger, D. (1988). SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. *J. Chem. Inform. Comput. Sci.*, *28*, 31–36.
85. Bostrom, J., Greenwood, J. R., Gottfries, J. (2003). Assessing the performance of OMEGA with respect to retrieving bioactive conformations. *J. Mol. Graphics Modelling*, *21*, 449–462.
86. Buranda, T., Sklar, L. A., Lopez, G., Simons, P., Huang, J., Perez-Luna, V. (2001). Fluorescence and FRET-based assays for biomolecules. Pending U.S. Patent and Registered Trademark.
87. Tanenbaum, D. M., Wang, Y., Williams, S. P., Sigler, P. B. (1998). Crystallographic comparison of the estrogen and progesterone receptor's ligand binding domains. *Proc. Nat. Acad. Sci. USA*, *95*, 5998–6003.
88. Shiau, A. K., Barstad, D., Loria, P. M., Cheng, L., Kushner, P. J., Agard, D. A., Greene, G. L. (1998). The structural basis of estrogen receptor/coactivator recognition and the antagonism of this interaction by tamoxifen. *Cell*, *95*, 927–937.
89. Quehenberger, O., Prossnitz, E. R., Cavanagh, S. L., Cochrane, C. G., Ye, R. D. (1993). Multiple domains of the N-formyl peptide receptor are required for high-affinity ligand binding. Construction and analysis of chimeric N-formyl peptide receptors. *J. Biol. Chem.*, *268*, 18167–18175.
90. Quehenberger, O., Pan, Z. K., Prossnitz, E. R., Cavanagh, S. L., Cochrane, C. G., Ye, R. D. (1997). Identification of an N-formyl peptide receptor ligand binding domain by a gain-of-function approach. *Biochem. Biophys. Res. Commun.*, *238*, 377–381.
91. Miettinen, H. M., Mills, J. S., Gripentrog, J. M., Dratz, E. A., Granger, B. L., Jesaitis, A. J. (1997). The ligand binding site of the formyl peptide receptor maps in the transmembrane region. *J. Immunol.*, *159*, 4045–4054.
92. Mills, J. S., Miettinen, H. M., Barnidge, D., Vlases, M. J., Wimer-Mackin, S., Dratz, E. A., Sunner, J., Jesaitis, A. J. (1998). Identification of a ligand binding site in the human neutrophil formyl peptide receptor using a site-specific fluorescent photoaffinity label and mass spectrometry. *J. Biol. Chem.*, *273*, 10428–10435.
93. Mills, J. S., Miettinen, H. M., Cummings, D., Jesaitis, A. J. (2000). Characterization of the binding site on the formyl peptide receptor using three receptor mutants and analogs of Met-Leu-Phe and Met-Met-Trp-Leu-Leu. *J. Biol. Chem.*, *275*, 39012–39017.

94. Baldwin, J. M. (1993). The probable arrangement of the helices in G protein-coupled receptors. *EMBO J.*, *12*, 1693–1703.

95. Baldwin, J. M., Schertler, G. F., Unger, V. M. (1997). An alpha-carbon template for the transmembrane helices in the rhodopsin family of G-protein-coupled receptors. *J. Mol. Biol.*, *272*, 144–164.

96. Palczewski, K., Kumasaka, T., Hori, T., Behnke, C. A., Motoshima, H., Fox, B. A., Le Trong, I., Teller, D. C., Okada, T., Stenkamp, R. E., Yamamoto, M., Miyano, M. (2000). Crystal structure of rhodopsin: A G protein-coupled receptor. *Science*, *289*, 739–745.

97. Levesque, L., Gaudreault, R. C., Marceau, F. (1991). The interaction of 3,5-pyrazolidinedione drugs with receptors for f-Met-Leu-Phe on human neutrophil leukocytes: A study of the structure-activity relationship. *Can. J. Physiol. Pharmacol.*, *69*, 419–425.

98. Bennett, T. A., Key, T. A., Gurevich, V. V., Neubig, R., Prossnitz, E. R., Sklar, L. A. (2001). Real-time analysis of G protein-coupled receptor reconstitution in a solubilized system. *J. Biol. Chem.*, *276*, 22453–22460.

99. Key, T. A., Vines, C. M., Wagener, B. M., Gurevich, V. V., Sklar, L. A., Prossnitz, E. R. (2000). Inhibition of chemoattractant N-formyl peptide receptor internalization by active arrestins. *Traffic.*, *2*, 87–99.

100. Biggs, S., Prossnitz, E., Sklar, L. (In press). Analysis in real-time of the assembly and disassembly of soluble G protein-coupled receptor-G protein complexes. Wiley, New York.

101. Seamer, L., Sklar, L. A. (2001). Time as a flow cytometric parameter. *Methods Cell. Biol.*, *63*, 169–183.

102. Seamer, L. C., Kuckuck, F., Sklar, L. A. (1999). Sheath fluid control to permit stable flow in rapid mix flow cytometry. *Cytometry*, *35*, 75–79.

103. Graves, S. W., Nolan, J. P., Sklar, L. A. (In press). Molecular assemblies, probes and proteomics in flow cytometry. In L. A. Sklar (Ed.), *Flow Cytometry in Biotechnology.* Oxford University Press, New York.

104. Wu, Y., Buranda, T., Lopez, G. P., Prossnitz, E. R., Sklar, L. A. (In preparation). Rapid kinetic instrumentation for GPCR assembly.

105. Graves, S., Nolan, J., Jett, J., Martin, J., Sklar, L. (2002). Nozzle design parameters and their effects on rapid sample delivery in flow cytometry. *Cytometry*, *47*, 127–137.

106. Simons, P. C., Vines, C. M., Key, T. A., Potter, R. M., Sklar, L. A., Prossnitz, E. R. (In press). Analysis of GPCR assemblies by flow cytometry. In L. A. Sklar (Ed.), *Flow Cytometry in Biotechnology*. Oxford University Press, New York.

107. Young, S. M., Curry, M. S., Ransom, J. T., Ballesteros, J. A., Prossnitz, E. R., Sklar, L. A., Edwards, B. S. (2004). High-throughput microfluidic mixing and multiparametric cell sorting for bioactive compound screening. *J. Biomol. Screening*, *9*, 103–111.

108. Sklar, L. A. (2003). High throughput screening for drug discovery by flow cytometry. In *Business Briefings: Future Drug Discovery*. World Markets Research Centre, London, p. 50.

109. Edwards, B. S., Sklar, L. A. (In press). Automation and high throughput. In L. A. Sklar (Ed.), *Flow Cytometry in Biotechnology*. Oxford University Press, New York.
110. Buranda, T., Huang, J. M., Perez-Luna, V. H., Schreyer, B., Sklar, L. A., Lopez, G. P. (2002). Biomolecular recognition on well-characterized beads packed in microfluidic channels. *Anal. Chem.*, *74*, 1149–1156.
111. Piyasena, M., Buranda, T., Huang, J., Wu, Y., Sklar, L. A., and Lopez, G. P. (2004). Near simultaneous and real-time analysis of multiple analytes in affinity microcolumns using defined molecular assemblies on beads, *Anal. Chem.*, *76*, 6266–6273.
112. Fu, Q., Rao, G. V. R., Ista, L. K., Xu, Sklar, L. A., Ward, T. L., Lopez, G. P. (In press). Synthesis of smart mesoporous materials. *MRS Proc.*
113. Buranda, T., Sklar, L. A., Lopez, G. P. (In press). Flow cytometry, beads and microchannels. In L. A. Sklar (Ed.), *Flow Cytometry in Biotechnology*. Oxford University Press, New York.
114. Nolan, J. P., Sklar, L. A. (1998). The emergence of flow cytometry for sensitive, real-time measurements of molecular interactions. *Nat. Biotechnol.*, *16*, 633–638.

# COMBINING NMR SPECTRAL INFORMATION WITH ASSOCIATED STRUCTURAL FEATURES TO FORM COMPUTATIONALLY NONINTENSIVE, RUGGED, AND OBJECTIVE MODELS OF BIOLOGICAL ACTIVITY

RICHARD D. BEGER, DAN A. BUZATU, AND JON G. WILKES
*National Center for Toxicological Research, Food and Drug Administration*
*Jefferson, Arkansas*

1 INTRODUCTION
Nuclear Magnetic Resonance Spectral Information Content
Biological Effect Modeling Based on NMR Spectra and Associated Structural Features
Previous NMR Comparative Spectral Analysis Modeling Approaches
Advantages of Modeling with Simulated Rather Than Experimental NMR Spectra
Adding Structurally Assigned NMR Chemical Shift Information to 2D Templates
Inspiration Based on Extension of Multidimensional NMR Techniques
2 METHODS AND EXAMPLES
Predicted NMR Spectra
Pattern Recognition Methods and Model Development
Linear and Nonlinear Aspects
Model Validation: LOO and LNO
Procedure for Integrating NMR Spectral Information with Molecular Structure

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

CoSCoSA Procedure for Incorporating Molecular Structure into NMR Spectra
Producing a $^{13}C$ and $^{15}N$ Heteronuclear Connectivity Matrix
3 FUTURE IMPROVEMENTS
Commercial Software Products
Improving the Algorithms for Binning Spectral Data
4 CONCLUSIONS
References

## 1 INTRODUCTION

Currently, quantitative structure–activity relationship (QSAR) and structure–activity relationship (SAR) models are used for drug discovery and to form ADME (absorption, distribution, metabolism, and excretion) models [1–4]. Three-dimensional (3D) QSAR models are typically based on physical fields obtained by superimposing each compound as a whole on a 3D grid; SAR modeling simplifies the process by breaking large compounds into secondary structural pieces but does not attempt to provide a physical basis for modeling the biological activity [1, 5–11]. The calculations used in building QSAR models depend on physical constants that vary in significance between structures of different types and may also exhibit nonlinear relationships for models based on a single structural type. Further, the selection of the most appropriate 3D structure for each molecule requires a number of assumptions. By trial and error, one can build a model for a finite number of molecules, adjusting conformational alignments to determine which orientations give the model the best relationship to known biological activities from a training set. However, the same process is much more prone to error when applied to structures for which the biological activity is unknown. The necessary conformational and other assumptions can give results that may be inaccurate, especially when applied for predictive purposes to structures of unknown activity.

This chapter covers an alternative strategy for model building known as quantitative spectrometric data–activity relationships (QSDARs). QSDAR models are based on the correlation between the triangular structure–spectra–activity relationship, whereas the aforementioned 3D QSAR models are based on the structure–activity relationship. 3D QSAR results show that receptor binding of a compound can be modeled successfully, if the model is carefully constructed and validated. These 3D QSAR models are based partly on electrostatics and partly on molecular geometry [5–9]. Mass spectra primarily reflect molecular geometry. Ultraviolet (UV), infrared (IR), and nuclear magnetic resonance (NMR) spectra reflect quantum mechanically constrained components that, like QSAR factors, depend on local electrosta-

tics and geometry. UV, IR, and NMR spectra are all modulated by the principles of quantum mechanics. For the model-building purposes described here, an advantage of NMR is that spectral features (chemical shifts) can be assigned to the specific atom that produces them, whereas this is not generally possible for UV, IR, or mass spectra (MS).

## Nuclear Magnetic Resonance Spectral Information Content

The $^{13}C$ NMR spectrum of a compound contains a pattern of frequencies that correspond directly to the quantum mechanical properties of a carbon nuclear magnetic dipole in a magnetic field. The pattern reflects the local electrostatic environment and electron orbital configuration of each carbon atom. Ab initio quantum mechanical calculations of $^{13}C$ chemical shift tensors in proteins reveal that they are dependent on the three-dimensional structural environment and electrostatic potential [12]. The NMR chemical shift tensor has two terms, a diamagnetic term and a paramagnetic term [13]. The diamagnetic term of an NMR chemical shift tensor is directly related to the electrostatic potential at its nucleus [13]. The paramagnetic term in the NMR chemical shift tensor is dependent upon the orbital configuration [13]. For $^{13}C$ NMR spectra the differences between the diamagnetic and paramagnetic term are very large so the spectral regions for different carbon orbital configurations are generally well-separated from each other.

Typically, $^{13}C$ NMR chemical shifts in the 0 to 100 ppm range are associated with carbon atoms that have $sp^3$ hybrid orbitals, with the more upfield shifts having a positive electrostatic potential (e.g., methyl groups) and the downfield shifts having a more negative electrostatic potential (e.g., ester bonds). $^{13}C$ NMR chemical shifts in the 100 to 220 ppm range are associated with carbon atoms that are $sp^2$ hybridized, with the more upfield shifts having a positive electrostatic potential (e.g., benzyl groups) and the downfield shifts having a more negative electrostatic potential (e.g., carbonyl groups). Superimposed on these basic relationships are effects arising from each molecule's carbon backbone as well as other carbon or heteroatom substituents. The effect of substituents on the $^{13}C$ NMR chemical shifts can be felt from as far as five bonds away or, depending on how the molecule has folded back upon itself, at even greater bond numbers directly through space. Thus, the final disposition of a molecule's NMR spectral features depends on structural and physical elements that, according to SAR theory, also determine the biological activity of the molecule. It follows that patterns can be discovered and a model can be built that correlate a finite training set of NMR spectra with the known biological characteristics of the compounds in the set. Once the patterns are validated, they can then be used for predicting the strength of the same biological activity for other compounds not used in the model-building process. This is the conceptual foundation of QSDAR.

It is possible to build spectrometric data–activity relationship (SDAR) and QSDAR models without associating the pattern of spectral features to the

molecular geometry. Figure 1 shows the QSDAR modeling flowchart including binning, weighting, and validation steps. First, the structures of a set of compounds are generated and their $^{13}C$ NMR spectrum are predicted or found in a database. The set of compounds' $^{13}C$ NMR spectra are saved as a set of ordered pairs: chemical shift frequencies in parts per million and the area under the peak. The area under a specific chemical shift frequency is first normalized to an integer. A nondegenerate frequency is assigned an area of 100, a doubly degenerate frequency (two $^{13}C$ NMR chemical shifts at the same frequency) has an area of 200, and so forth. This initial normalization is done so that: (1) all the spectra have a similar signal-to-noise ratio and (2) line width variations due to differences in NMR instrumental field strengths, shimming, coupling to protons, temperature, pH, or solvent are eliminated. The bins define the number of chemical shift peaks within a ppm range. The bin width used in inputting the $^{13}C$ NMR spectra can be optimized by allowing the spectral window width to vary between 1.0 and 2.0 ppm and determining which width produces the best cross-validated models.

The QSDAR models can be defined using a selection of the most correlated individual bins. Alternatively, the models can use all the spectral bins but with the most significant ones more heavily weighted. A well-known method of producing optimal weighting is by calculating principal components (PCs)

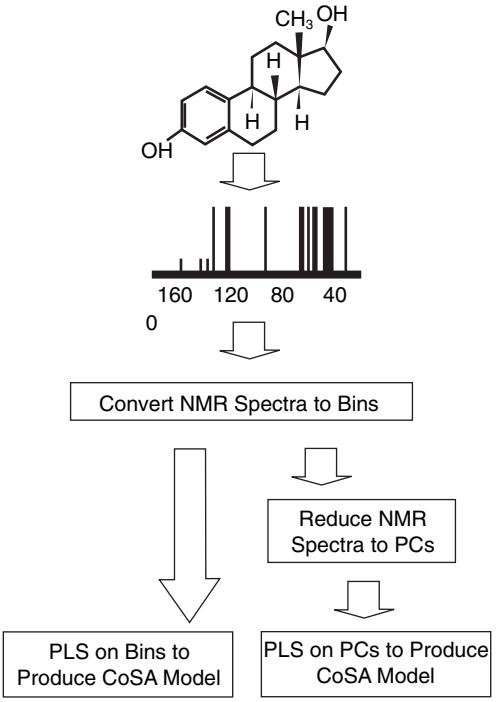


**Figure 1** Procedural flowchart for QSDAR modeling.

of variation. Each PC contains a linear combination of the original bin intensities. The PCs are orthogonal axes or vectors ranked by rotation and translation in decreasing order of their capacity to distinguish differences among the training set spectra. Calculation of PCs produces a better quality model in cases where multiple spectral features interact to define the global SDAR.

## Biological Effect Modeling Based on NMR Spectra and Associated Structural Features

The patterns in a one-dimensional SDAR or QSDAR model reflect the set of electrostatic and orbital relationships within a training set of molecules. However, the spectral features (chemical shifts) are not associated with particular atomic locations within the molecules. Since atoms in different parts of a large molecule may have similar chemical shifts and fill the same spectral bin, it follows that one-dimensional (1D) QSDAR may confound irrelevant with relevant spectral signals. An additional geometric dimension can be added to the spectral data, by indicating which chemical shifts are associated with two adjacent carbon atoms. This would produce a two-dimensional (2D) QSDAR model. Rather than a model in which bins at 220 and 50 ppm are associated with a particular biological effect, a 2D QSDAR model based on the same training set might suggest that the effect occurs only when these two chemical shifts are found on adjacent atoms. This improved basis for pattern definition is conceptually similar to 3D QSAR but with fundamental advantages. Unlike 3D QSAR, 2D QSDAR and analogous higher dimension models described later are developed without a hypothesized docking of each molecule onto a 3D grid, in relation to a known enzymatic binding site, or by comparison to the shape and size of a particular reference molecule. In 2D QSDAR, the second dimension simply correlates intramolecular spatial relations, two atoms at a time. This can be done without defining a standard molecular type with respect to which all other molecules are regarded as variants. Further, the energetically optimal conformation of each molecule interacting with the enzyme or other protein receptor mediating the biological effect does not need to be known. The subjective assumptions involved in 3D QSAR are unnecessary for 2D QSDAR. This provides a philosophically objective basis for modeling. Finally, it avoids the problematic assumption that the energy minimum conformation best represents each molecule's capacity to cause the biological effect. Below we will describe in more detail the modeling quality produced by the basic 1D QSDARs. Improvements obtained by using higher dimensions are then explained and discussed.

## Previous NMR Comparative Spectral Analysis Modeling Approaches

We have used partial least-squares discriminant analysis (PLS-DA) to form a relationship between biological activity and experimental 1D $^{13}C$ NMR spectral data. This modeling technique is called comparative spectral analysis

(CoSA). We have developed an internally cross-validated and externally tested SDAR model of 108 diverse compounds binding to the estrogen receptor [14]. SDAR models segregated the 108 compounds into binding classifications based on relative strength with 20 strong, 15 medium, and 73 weak. Based only on $^{13}$C NMR spectroscopic data, the statistical pattern recognition program with 22 Fisher-weighted PCs used 89.8 percent of the total variance and had a cross-validation of 75.0 percent. The first principal component accounted for 32.5 percent of the cumulative variance.

## Advantages of Modeling with Simulated Rather Than Experimental NMR Spectra

The success of these first SDAR models based on experimental 1D $^{13}$C NMR spectra led to look into the possibility of using predicted $^{13}$C NMR spectra. $^{13}$C NMR chemical shifts have been used to predict and refine chemical structures [15, 16]. Conversely, the chemical structure of a compound has been used to predict its $^{1}$H and $^{13}$C NMR chemical shifts [17–19]. The ability to use predicted over experimental NMR spectra in a QSDAR model is not necessary, but it saves time, money, and can prevent possible toxic exposures. Predicted $^{13}$C NMR data points allow for the spectra to be independent of the solvent used. ACD/Labs CNMR predictor software (ACD/Labs, Toronto, Canada, www.acdlabs.com) was used to predict 1D $^{13}$C NMR spectra that produced robust SDAR models of polychlorinated dibenzodioxins (PCDD) and polychlorinated dibenzofurans (PCDF) binding to aryl hydrocarbon receptor (AhR) [20] and rates of dechlorination for chlorobenzenes, chlorophenols, and chloroanilines [21].

The early success of qualitative SDAR models led us believe that QSDAR was feasible. Multiple linear regression (MLR) statistical techniques were used to form a relationship between predicted 1D $^{13}$C NMR data and biological activity data. We were able to produce reliable QSDAR models of 30 steroids binding to the corticosterone binding globulin [22]. Another group of models quantified the tendency of 50 steroids to bind to the aromatase enzyme [23]. Binding to the aryl hydrocarbon receptor (AhR) was similarly modeled quantitatively for 26 PCDFs, 14 PCDDs, 12 polychlorinated biphenyls (PCBs), or for all of these compounds together [24]. These MLR 1D comparative spectral analysis (CoSA) models using simulated $^{13}$C NMR data yielded higher cross-validated correlations than were seen with comparative molecular field analysis (CoMFA) methods.

## Adding Structurally Assigned NMR Chemical Shift Information to 2D Templates

One limitation with 1D CoSA models is that they lack direct 3D structural information. The most obvious way to combine structural and spectral information is to establish one particular molecule as a best or normal represen-

tative of those causing a particular biological effect. Each carbon atom in this compound's backbone is numbered and all other compounds to be modeled must use the same backbone numbering system. Then, each compound's pattern is defined by the ordered pair (carbon number, chemical shift) rather than by the previously described system of (chemical shift bin number, occupancy number). This means that the chemical shifts have been assigned to the carbon atoms that produced them. The pattern as defined is correlated with the biological activity of each molecule. The resulting model combines structural information with the assigned simulated $^{13}C$ NMR chemical shifts. We have named this 1D SDAR method comparative structurally assigned spectra analysis (CoSASA) to distinguish it from the substantially different, unassigned methods previously described as CoSA. One supposes that the ability to include spatial relationships in SDAR modeling should improve the quality of the results. In fact, when used on the same spectral training set, we observed inferior results by CoSASA compared to CoSA [22, 23]. This unexpected result challenged our understanding of the factors affecting model quality.

## Inspiration Based on Extension of Multidimensional NMR Techniques

Another way to combine structural and spectral information is to express geometrical information in spectral space. In the same way that 2D, 3D, and four-dimensional (4D) NMR experiments use additional spectral dimensions to reduce spectral overlap [25], we conceptualized an analogous way of correlating spectral and structural dimensions into a single multidimensional matrix. We hypothesized that spectra defined in such a matrix would produce improved QSDAR models. We also realized that the matrix definition would not limit compounds to those sharing the same backbone template. Compounds in the same model could have quite dissimilar structures: They could differ in the number and connectivity of carbon atoms, as well as the number and identity of constituents or atoms. Below we describe the way to define a multidimensional QSDAR data structure matrix.

A molecule's 3D connectivity matrix can be built by displaying all the possible carbon-to-carbon connectivities with their assigned carbon NMR chemical shifts on two dimensions and the distance between the two atoms in a third, orthogonal dimension. Figure 2 shows all the atom-to-atom connectivities and the 3D connectivity matrix for 3β-estradiol. The $X$ axis represents the chemical shifts of carbon $i$, and the $Y$ axis represents those of carbon $j$. The $Z$ axis represents the distance between carbon $i$ and carbon $j$ ($r_{ij}$). By representing molecules with this matrix, the subjective superposition of molecules on a template is avoided. Other parameterization needed to develop typical 3D QSAR models is either avoided or minimized. For flexible molecules, some atom-to-atom distances can vary, so representations of multiple conformations in the 3D connectivity matrix format can be accommodated and will be explained later.

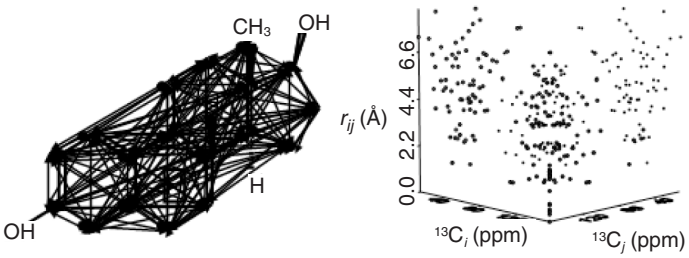


**Figure 2** The 3D connectivity matrix of 3β-estradiol.

The connectivity matrix shown in Figure 2 is symmetrical about the *x-y* diagonal. That is, for every connection between atom *i* and atom *j* the identical relationship is represented across the diagonal at the connection between atom *j* and atom *i*. Along the *X-Y* diagonal at $r_{ij} = 0$ is the 1D $^{13}C$ NMR spectrum of 3β-estradiol. At $r_{ij} \approx 1.4$ Å are the nearest neighbor atom-to-atom connections and at $r_{ij} > 1.4$ Å are all the other distance related atom-to-atom connections.

The information in a 3D connectivity matrix determines a compound's structure, so the information in the matrix that is actually used for a model can be reduced to simplify and accelerate computations. One possible way to reduce the 3D matrix is to cut it into a set, arbitrarily, of four 2D spectral planes. The first 2D plane is the nearest neighbor through-bond connectivity plane. The three other 2D planes are constructed by binning or projecting the actual distances onto the nearest plane. This effectively compresses and greatly simplifies distance information along the *Z* axis. In addition to the bond-connectivity plane, there is a plane for somewhat greater atom-to-atom connections (2.0 Å $< r_{ij} <$ 3.6 Å, perhaps two bond lengths but through space), another for medium range atom-to-atom connections (3.6 Å $< r_{ij} <$ 6.0 Å), and a fourth for long-distance atom-to-atom connections ($r_{ij} >$ 6.0 Å). Figure 3 shows the four 2D planes and all the atom-to-atom connections in each 2D plane for the reduced 3D connectivity matrix of 3β-estradiol. Similarities between the pattern of 2D spectral data associated with the biological activity of the training set compounds and the spectral data for the test compound are detected and used to determine whether the compound is predicted to exhibit the biological activity. We call this procedure comparative structural connectivity spectra analysis (CoSCoSA), to distinguish it from CoSA and CoSASA. We will now compare CoSCoSA modeling to actual or hypothetical NMR multidimensional analytical methods.

Standard NMR instrumental techniques include 2D $^{1}H$-$^{1}H$ correlation spectroscopy (COSY) [26] experiments in which connectivity relationships through three bonds are found for nearest neighbor protons with an off-diagonal cross peak. The $^{1}H$-$^{1}H$ COSY experiment is comparable to a 2D $^{13}C$-$^{13}C$ COSY experiment, though this experiment is not usually practical because

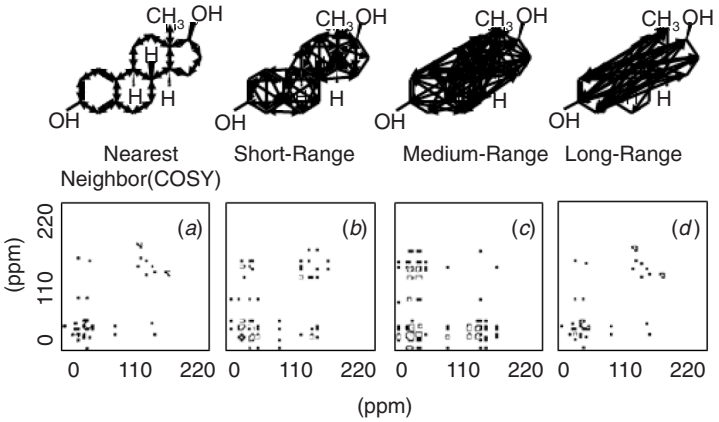


**Figure 3** The 2D connectivity planes from the 3D connectivity matrix of 3β-estradiol.

of the low natural abundance of $^{13}C$. The shortest distance layer of the connectivity matrix for $^{13}C$ 3D QSDAR methods described here corresponds in layout and information content seen in $^{13}C$-$^{13}C$ COSY experiments. Using the structurally assigned predicted spectra and adding the nearest neighbor information as cross peaks produces a simulated $^{13}C$-$^{13}C$ COSY layer in the matrix.

$^{13}C$-$^{13}C$ COSY experiments have similarities to 2D $^{1}H$-$^{13}C$ heteronuclear single quantum correlation (HSQC) [27] and $^{1}H$-$^{13}C$ heteronuclear multiple quantum coherence (HMQC) [28] NMR experiments, both of which represent the connectivity of carbons and their attached protons. In practice, 2D $^{13}C$-$^{13}C$ COSY experiments are seldom run because small molecules are rarely fully $^{13}C$ labeled. Even if molecules are fully labeled with $^{13}C$, through-bond connections are usually obtained directly from other NMR experiments like HCCH or indirectly from $^{1}H$-$^{1}H$ COSY with $^{13}C$-$^{1}H$ HMQC and HMBC [29].

In 3D QSDAR, 2D $^{13}C$-$^{13}C$ distance spectra that contain short-, medium-, and long-distance through-space connectivity spectral patterns are easily simulated using predicted spectra. Since the spectra are predicted based on the compound structures, the atom producing each chemical shift is known. It is then possible to estimate atom-to-atom distances ($r_{ij}$) for each pair of atoms and to associate the $r_{ij}$ values with the corresponding spectral features. The shorter distance relationships in 3D QSDAR are analogous to 2D $^{1}H$-$^{1}H$ nuclear Overhauser effect spectroscopy (NOESY) NMR experiments where correlations through space are found for neighboring protons that are less than 5 Å apart. As with COSY, NOESY spectra are expressed as a matrix with off-diagonal cross peaks [30]. The volumes of the cross peaks in a NOESY experiment are dependent on the distance between the protons, the mixing time of

the experiment, and the number of different nuclear Overhauser effect (NOE) spin diffusion pathways available for dipolar magnetization transfer. The volumes in a NOESY spectrum can be used to solve the three-dimensional structure of a protein by assigning short-range constraints to large volumes, medium-range constraints to medium volumes, and long-range constraints to small volumes [31]. $^{13}C$-$^{13}C$ NOESY experiments are for all practical purposes never executed, again because naturally occurring small compounds are not fully $^{13}C$ labeled. Since it is based on simulated spectra, the corresponding 3D QSDAR matrix is definable and its utility for modeling is practical.

## 2 METHODS AND EXAMPLES

### Predicted NMR Spectra

Simulated $^{13}C$ NMR spectra were determined using the ACD/Labs CNMR predictor software (ACD, Toronto, Canada, www.acdlabs.com). In this and other NMR predictor programs [17–19], NMR spectra are calculated by a substructure similarity technique called hierarchically ordered spherical description of environment (HOSE) [32], which correlates similar substructures with similar NMR chemical shifts. Since some of the compounds were in the predictor software spectral database used for the HOSE calculation, their predicted spectra were equivalent to experimental NMR spectra. Since it used only simulated $^{13}C$ NMR spectra, QSDAR modeling was driven completely within the computer. No experimental data were acquired. The $^{13}C$ NMR chemical shift database used by ACD/Labs CNMR predictor contained 1.7 million chemical shifts from experimental NMR spectra of 140,000 compounds.

### Pattern Recognition Methods and Model Development

A simple description of pattern recognition is the ability to discern a pattern or an analogous set of interrelationships within a set of data (i.e., to model a set of data). This basic concept underlies most scientific discoveries or theories. In a practical sense, pattern recognition is the process of creating a numerical model of a system or endpoint based on descriptors of that process or system. This turns out to be nothing more than mapping or approximating a function that describes the system. What is being modeled can vary greatly. It can be as simple as the amount of correction that needs to be applied to several mechanical parameters of a machine's arm to maintain a certain manufacturing tolerance, or a complex function such as the etiological interaction between several hundred biochemical parameters to produce a tumor.

Once it is established that a set of data contains some distinguishable pattern, a number of examples can be used to model that pattern. This is the

model development component of the pattern recognition process. Examples from within data sets, that is, members of a data set containing unique descriptors, are used to establish models that describe the data sets as accurately as possible. This process is usually referred to as the training or learning stage, and it consists of training the pattern recognition method with repeated examples of input data (descriptors) and associated outcomes or endpoints. Training continues until the training set is modeled to the desired accuracy. A sufficient number of examples from a data set are required to enable the method (any pattern recognition method) to discover the pattern in the data and create an accurate model of that data. The number of examples that must be provided increases as the complexity of the pattern recognition problem modeled increases. In other words, the harder the question being asked is, the more examples that are needed to answer the question. This concept can be demonstrated with the SDAR and QSDAR techniques. For example, a simple SDAR model providing binary yes/no classification results for a biological or physiological endpoint (such as weak or strong receptor binding for a set of drug molecules) can be developed from a data set with as few as 10 examples. However, a QSDAR model of an analogous biological or physiological endpoint for a set of molecules, but producing an accurate quantitative result, usually requires at least 3 to 4 times that number of examples.

Pattern recognition model development can be broadly split into two categories, supervised methods and unsupervised methods. Supervised methods learn by experiencing multiple examples of both input stimuli and their associated outcomes. An example of this type of input–output relationship is the relation between the number of a patient's physiological parameters and his or her outcome after a medical procedure. In this example, the physiological data and the patient's outcome can be measured, and both types of data can be used to develop a model using a supervised method. Generalizations can be produced using this type of information and used to make predictions. There are many examples of supervised methods including principal components, multiple linear regression methods, and many types of artificial neural networks [33–35]. Unsupervised methods establish patterns in data differently compared to supervised methods. Input stimuli for a number of examples are provided to the method in a similar manner to the supervised method, but the outcome is not provided. Associations are established within the data between the examples based on the stimuli, and the result is a clustering of the data according to similarity. A simple example of this type of data structure is a population analysis of a community based on people's occupational parameters (education level, degree, type of job, income level, etc.). Self-organizing maps, decision trees, and *k*-means are some examples of frequently used unsupervised methods [33]. Since the SDAR and QSDAR methods are inherently supervised techniques, the remaining discussion will focus on these types of methods.

### Linear and Nonlinear Aspects

Most pattern recognition methods fall into two main categories, linear and nonlinear methods. Linear methods use linear functions to map a relationship, and nonlinear methods use nonlinear functions to describe patterns. In performing pattern recognition and model development on a data set, it is ideal to use a method that is most aligned with the data structure (i.e., linear method to describe linear function, nonlinear method to describe nonlinear behavior in a data set). Although it is true that both types of methods can be used on both types of data structures, better generalization and predictive performance will be obtained from the nonlinear methods even when modeling data that has only slight nonlinear behavior.

### Model Validation: LOO and LNO

The analysis of a developed model can be done using the leave-one-out (LOO) cross-validation procedure [1]. This procedure consists of excluding each example systematically from a training set, using the remaining examples to develop the model, and then using the model to predict the output property of the excluded example. For example, if a training set consists of a certain number of drug molecules and their respective binding affinities to a particular receptor, the LOO procedure would be used to methodically exclude one molecule and its binding affinity from the training set, and then use the developed model to predict the binding affinity of the excluded molecule. If the LOO procedure is performed $n$ times, where $n$ is the total number of examples in the data set, a cross-validation covariance coefficient can be calculated using the combined predicted results. The cross-validated $r^2$ (termed $q^2$) can be derived from

$$q^2 = 1 - \frac{\text{PRESS}}{\text{SD}}$$

where PRESS is the sum of the differences between the actual and predicted output data for each example during LOO cross-validation, and SD is the sum of the squared deviations between the measured and mean outputs of each example in the training set. We believe that $q^2$ is a more valid measure than $r^2$ for assessing the reliability of a mathematical model intended for predictive applications.

A modified version of the LOO cross-validation procedure is the leave-$N$-out (LNO) cross-validation procedure [36]. The LNO procedure is identical to the LOO procedure, but instead of removing only one example at a time from a data set for validation, $N$ examples are removed from the training set for testing. The number of examples excluded ($N$) can be as many as desired, but typically $N$ is not usually greater than 50 percent of the total data set. LNO is a more powerful procedure than LOO. It is a brute-force assessment of the

robustness and ruggedness of a model, especially when *N* represents a substantial percentage of the total data. It is a procedure that is not typically used, and, when used with good results, it is indicative of an exceptional predictive model.

## Procedure for Integrating NMR Spectral Information with Molecular Structure

Figure 4 shows a flowchart of the procedures that were used to make comparative structurally assigned spectral analysis (CoSASA) models. CoSASA models are produced by using the assigned $^{13}C$ NMR chemical shifts at the 20 positions in the steroid backbone template as shown in Figure 5. CoSASA modeling is analogous to QSAR modeling in which the 3D grid used in QSAR corresponds to the CoSASA's structural template positions. Each CoSASA "bin" represents the chemical shift intensity of a particular carbon atom in the molecule. The structural information is contained in the carbon backbone numbering system and is readily available without expert interpretation of an NMR spectrum because each simulated $^{13}C$ NMR chemical shift is predicted based on an identifiable carbon situated in the molecular structure. For this mode of data representation, it is possible to develop two types of CoSASA QSDAR models, one based on the most correlated individual bins and another on a selection of the PCs calculated from the population levels of all the bins.

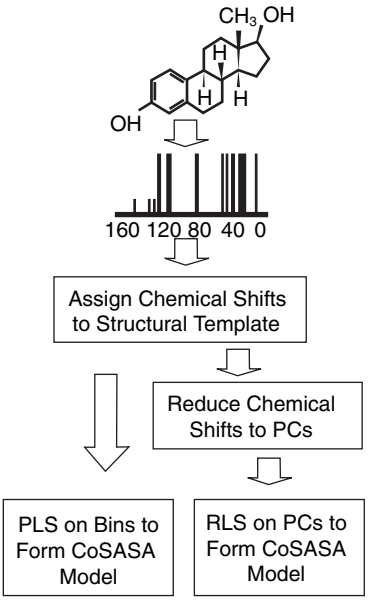


**Figure 4** Procedural flowchart for CoSASA modeling.

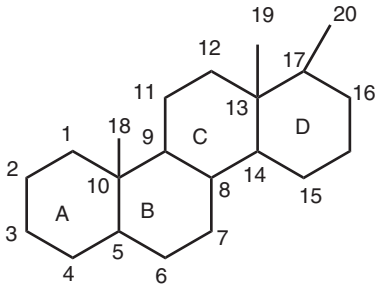


**Figure 5** Numbering scheme of a steroid template for CoSASA modeling.

***CoSASA Model of Corticosterone Binding*** Each of the 30 compounds specified in Table 1 and Figure 6 has known binding affinity to the corticosterone binding globulin. The assigned $^{13}C$ NMR spectra of the 30 steroids were placed in the steroid template positions shown in Figure 5 so MLR statistical analysis could be performed. The CoSASA model of corticosterone binding gave an explained variance ($r^2$) of 0.80 and a LOO cross-validated variance ($q^2$) of 0.73 [22]. The steroid backbone CoSASA model was based on the change in chemical shift frequency of atoms from positions 3, 14, and 20 on the steroid template. Positions 3 and 20 are near the steroid positions that previous QSAR models have shown are the active regions for corticosterone binding [37, 38]. The $q^2$ of 0.73 for the CoSASA model is slightly better than the QSAR $q^2$ of 0.68 but the CoSASA is much easier to build. We did not add more structurally assigned chemical shifts to the model because we wanted to limit the analysis to three variables, so that our CoSASA model would compare in this regard to the QSAR. This was a tactical choice. Without additional effort, more atoms could have been included and trends in $q^2$ as a function of number of bins used indicate that further improvements in predictive accuracy of the CoSASA model were possible.

***CoSASA Model for the Aromatase Enzyme*** Each of the 50 compounds specified in Table 2 and Figure 7 have known binding affinity to the aromatase enzyme. The assigned $^{13}C$ NMR spectra of the 50 steroids were placed in the steroid template positions shown in Figure 5 so MLR statistical analysis could be performed. An aromatase CoSASA model was built by selecting the five most correlated assigned $^{13}C$ NMR chemical shifts [23]. The five $^{13}C$ NMR chemical shifts used in this model were the steroid template positions 3, 9, 6, 7, and 12. Individual $^{13}C$ NMR chemical shifts from atoms in the steroid template positions 3, 9, 6, and 7 all had correlations greater than 0.6 to the relative binding activity of the enzyme. The explained correlation to the binding activity data ($r^2$) of this model is 0.75 and the LOO cross-validated variance ($q^2$) is 0.66. An aromatase CoSASA model using the 5 most correlated principal components had an $r^2$ of 0.75 and the LOO $q^2$ of 0.67. In the 50-compound training set, position 3 showed the highest correlation to aromatase

**TABLE 1 Structures of Corticosteroids Used in QSDAR Models of Corticosteroid Binding Globulin Data [22, 36, 37, 40, 51, 52]**

| Number | Structure[a] | $R_1$ | $R_2$ | $R_3$ | $R_4$ | $R_5$ | $R_6$ | $R_7$ | $R_8$ | $R_9$ | $R_{10}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | SB | OH | H | H | H | OH | H | | | | |
| 2 | SE | OH | OH | H | | | | | | | |
| 3 | SC | =O | H | =O | | | | H | H | H | H |
| 4 | SB | OH | H | H | H | =O | | | | | |
| 5 | SC | =O | OH | $COCH_2OH$ | H | | | H | H | H | H |
| 6 | SC | =O | OH | $COCH_2OH$ | OH | | | H | H | H | H |
| 7 | SC | =O | =O | $COCH_2OH$ | OH | | | | H | H | H |
| 8 | SE | OH | =O | | | | | | | | |
| 9 | SC | =O | H | $COCH_2OH$ | H | | | H | H | H | H |
| 10 | SC | =O | H | $COCH_2OH$ | OH | | | H | H | H | H |
| 11 | SB | =O | | H | H | OH | H | | | | |
| 12 | SD | OH | OH | H | H | | | | | | |
| 13 | SD | OH | OH | H | OH | | | | | | |
| 14 | SD | OH | =O | | H | | | | | | |
| 15 | SB | H | OH | H | H | =O | | | | | |
| 16 | SE | OH | COMe | H | | | | | | | |
| 17 | SE | OH | COMe | OH | | | | | | | |
| 18 | SC | =O | H | COMe | H | | | H | H | H | H |
| 19 | SC | =O | H | COMe | OH | | | H | H | H | H |
| 20 | SC | =O | H | OH | H | | | H | H | H | H |
| 21 | SF | =O | OH | $COCH_2OH$ | OH | | | | | | |
| 22 | SC | =O | OH | $COCH_2OCOMe$ | | | | H | H | H | H |
| 23 | SC | =O | =O | COMe | H | | | | H | H | H |
| 24 | SC | =O | H | $COCH_2OH$ | H | | | OH | H | H | H |
| 25[b] | SC | =O | H | OH | H | | | H | H | H | H |
| 26 | SC | =O | H | COMe | OH | | | H | OH | H | H |
| 27 | SC | =O | H | COMe | H | | | H | Me | H | H |
| 28[b] | SC | =O | H | COMe | H | | | H | H | H | H |
| 29 | SC | =O | OH | $COCH_2OH$ | OH | | | H | H | Me | H |
| 30 | SC | =O | OH | $COCH_2OH$ | OH | | | H | H | Me | F |

[a] Structures according to Figure 6.
[b] H (hydrogen) instead of Me at $C_{10}$ steroid skeleton.

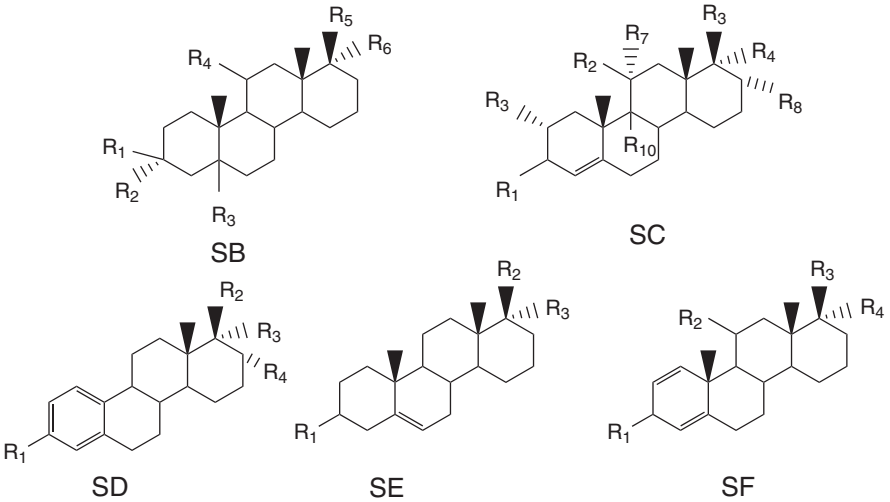


**Figure 6** Corticosteroid binding steroids structures.

binding. The polar electronegative keto group at position 3 (compounds **20**, **23** to **25**, **34** to **50**) was easily detected and correlated to enzyme inhibition when the chemical shift of position 3 was between 198 and 200 ppm. In a previous QSAR model, a correlation was found between a polar electronegative region around positions 2 to 4 and inhibition of the enzyme [7]. The keto group at position 7 (compounds **1** to **8**, **14** to **16**) produced chemical shifts between 200 and 201 ppm and correlated to weak inhibitor activity. As before, a correlation between the chemical shift from the position 7 keto group and weak aromatase inhibition was found in the previous QSAR model [7, 23]. The weak activities of the androst-5-ene compounds were associated with chemical shifts between 117 and 127 ppm from position 6. Again, a similar correlation was found in QSAR models between all the androst-5-ene compounds and lower inhibitor activity to the enzyme [7]. In contrast, the strong correlation of the chemical shift from position 9 with aromatase inhibitor activity was unexpected based on the published QSAR model. The chemical shifts between 55 and 58 ppm in position 9 come from steroids with 6-akyl substitutions. The 6-alkyl substituted steroids are strong inhibitors of the enzyme. It is significant that a correlation around position 9 and inhibitor activity to the enzyme was not reported in the previous QSAR model, whereas the CoSASA QSDAR model based on chemical shifts from 5 atoms correctly showed a high correlation of position 9 to inhibitor activity [23]. The $q^2$ of 0.6 for the CoSASA model is slightly worse than the QSAR $q^2$ of 0.72 but, again, the spectral data model is substantially easier to build. As before, we did not include more structurally assigned chemical shifts in the model because—for purposes of comparison—we wanted to limit the CoSASA analysis to the same number of variables, 5, used in the corresponding QSAR. As with the coticosterone

**TABLE 2 Structures of Steroids Used in QSDAR Models of Binding to Aromatase Enzyme [7, 23, 41]**

| Number | Binding Activity | Structure[a] | $R_1$ | $R_2$ | $R_3$ | $R_4$ | $R_5$ |
|---|---|---|---|---|---|---|---|
| 1 | –2.92 | A | $CH_2OH$ | =O | | | |
| 2 | –3.54 | A | $CH_2OH$ | OH | H | | |
| 3 | –3.00 | A | CHO | =O | | | |
| 4 | –3.26 | A | H | =O | | | |
| 5 | –2.62 | A | Me | OH | H | | |
| 6 | –3.06 | B | $CH_2OH$ | =O | | | |
| 7 | –2.14 | B | CHO | =O | | | |
| 8 | –2.36 | B | H | =O | | | |
| 9 | –1.89 | D | $CH_2OH$ | =O | | H | |
| 10 | –2.88 | D | $CH_2OH$ | OH | H | H | |
| 11 | –2.03 | D | CHO | =O | | H | |
| 12 | –0.97 | D | Me | =O | | H | |
| 13 | –2.93 | D | Me | =O | | Br | |
| 14 | –1.28 | A | Me | =O | | | |
| 15 | –1.23 | B | Me | =O | | | |
| 16 | –2.61 | B | Me | OH | H | | |
| 17 | –2.36 | D | Me | OH | H | H | |
| 18 | –0.65 | F | =O | | | | |
| 19 | –2.19 | F | OH | H | | | |
| 20 | –1.03 | H | H | H | H | | |
| 21 | 0.00 | C | Me | =O | | H | H |
| 22 | 0.46 | C | $CH_2OH$ | =O | | H | H |
| 23 | –0.84 | H | $CH_2OH$ | H | H | | |
| 24 | 0.15 | H | Me | =O | | | |
| 25 | –0.13 | E | =O | | =O | | $CF_2$ |
| 26 | 0.87 | E | =O | | H | H | $CH_2$ |
| 27 | –0.51 | E | OH | H | H | H | $CH_2$ |
| 28 | –1.35 | C | Me | OH | H | H | H |
| 29 | –0.67 | C | $CH_2OH$ | OH | H | H | H |
| 30 | –0.89 | C | $MeC(O)OCH_2$ | =O | | H | H |
| 31 | –0.79 | C | Me | =O | | H | Br |
| 32 | –1.09 | C | Me | =O | | H | H |
| 33 | –1.08 | C | $CF_3$ | =O | | H | H |
| 34 | 0.56 | I | Me | | | | |
| 35 | 0.87 | J | Me | | | | |
| 36 | 1.56 | I | $C_2H_5$ | | | | |
| 37 | 0.94 | J | $C_2H_5$ | | | | |
| 38 | 0.94 | I | $C_3H_7$ | | | | |
| 39 | 0.78 | J | $C_3H_7$ | | | | |
| 40 | 0.65 | I | $C_nH_9$ | | | | |
| 41 | 0.53 | J | $C_4H_9$ | | | | |
| 42 | 0.21 | I | $CH(CH_3)_2$ | | | | |
| 43 | 0.04 | J | $CH(CH_3)_2$ | | | | |
| 44 | –0.04 | I | $C_6H_5$ | | | | |
| 45 | 0.24 | J | $C_6H_5$ | | | | |
| 46 | –0.24 | I | $CH_2C_6H_5$ | | | | |
| 47 | 0.61 | J | $CH_2C_6H_5$ | | | | |
| 48 | 0.91 | I | $CH{=}CH_2$ | | | | |
| 49 | –0.32 | I | C=CH | | | | |
| 50 | 0.96 | G | | | | | |

[a] Structure column refer to Figure 5.7.

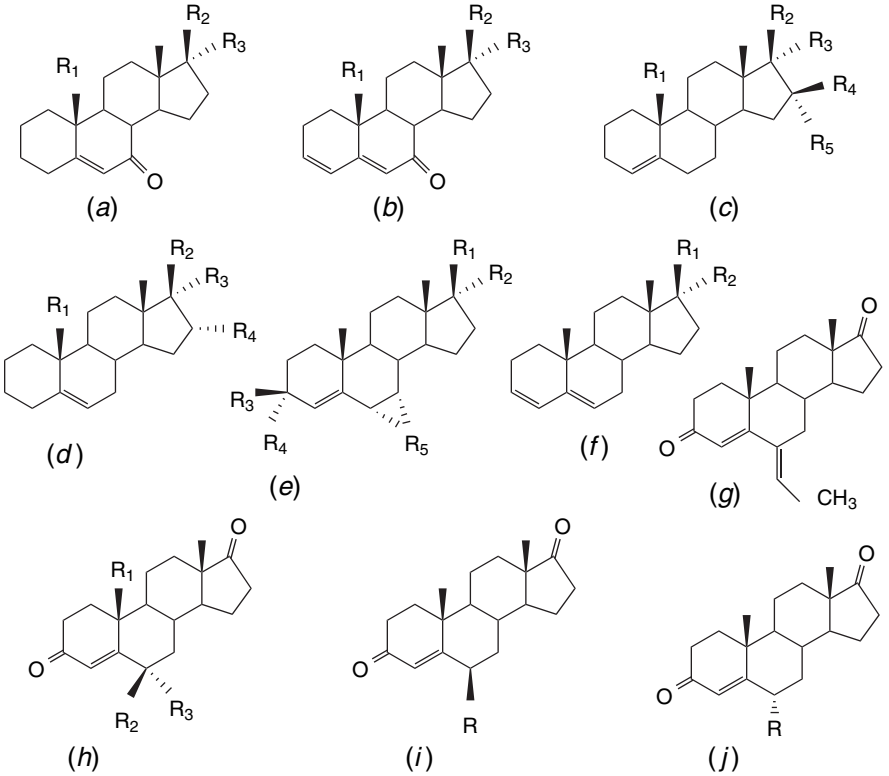


**Figure 7** Aromatase enzyme steroid structures.

binding CoSASA, more atoms could have been added to improve model predictive accuracy.

***CoSASA Model of Aryl Hydrocarbon Receptor Binding*** Figure 8*a* shows the backbone numbering system for 14 polychlorinated dibenzo-*p*-dioxins (PCDD), and Figure 8*b* shows the numbering system for 26 polychlorinated dibenzofurans (PCDF). The assigned $^{13}C$ NMR spectra of the 14 PCDD and 26 PCDF were placed in the template positions shown in Figures 8*a* and 8*b* so MLR statistical analysis could be performed. CoSASA models for PCDD and PCDF binding to the aryl hydrocarbon receptor were built from 12 assigned carbon chemical shifts as previously described [24, 38]. The 2D CoSASA model for the 26 PCDF compounds based on selection of 6 chemical shifts had an $r^2$ of 0.74 and a $q^2$ of 0.70. The $q^2$ of 0.70 for the CoSASA model is slightly worse than the QSAR $q^2$ of 0.72 but is much easier to build. The 2D CoSASA model for the 14 PCDD compounds based on selection of 5 chemical shifts had an $r^2$ of 0.81 and a $q^2$ of 0.53. The $q^2$ of 0.73 for the QSAR model is much better than the CoSASA $q^2$ of 0.53. The fact that the CoSASA

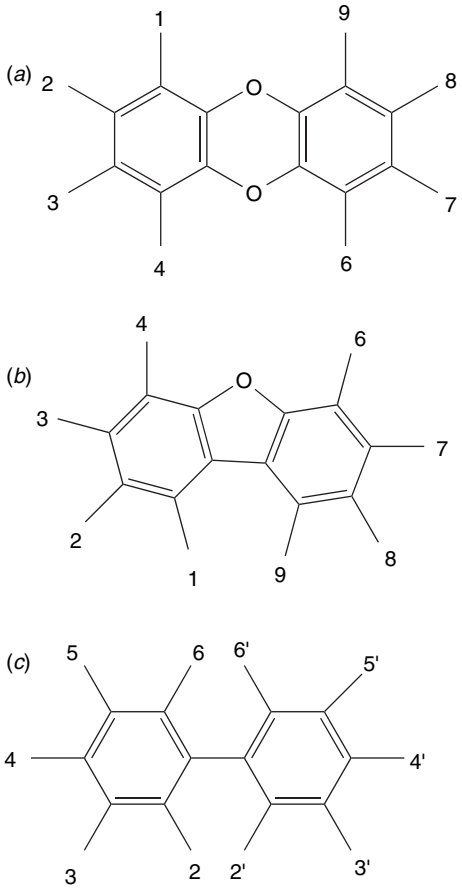


**Figure 8** (*a*) Polychlorinated dibenzo-*p*-dioxins (PCDD). (*b*) Polychlorinated dibenzofurans (PCDF). (*c*) Poluchlorinated biphenyls (PCB).

model was far worse than the QSAR model may result from the large proportion of compounds in the training set that had 2 or 4 axes of symmetry. Symmetry information is lost when atoms included in the model are structurally described with respect to template location only.

## CoSCoSA Procedure for Incorporating Molecular Structure into NMR Spectra

Figure 9 shows the flowchart on the comparative structural connectivity spectral analysis (CoSCoSA) modeling procedure [39–42]. The structures are used to predict the assigned 1D $^{13}C$ NMR spectra in this procedure. The resolution of the 2D spectra was typically reduced to 2.0 ppm in both dimensions

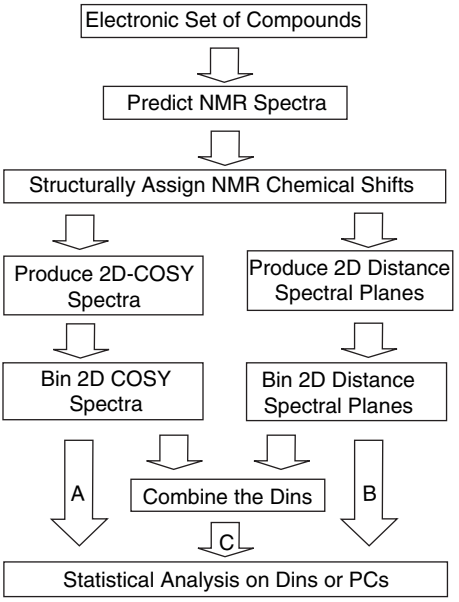


**Figure 9** Procedural flowchart of CoSCoSA models.

to produce multiply populated bins for statistical analysis and to reduce the effects of uncertainties in the simulated spectra. The spectral widths were chosen because of convenience and prior success. The data in each 2D $^{13}C$-$^{13}C$ NMR spectral plane were saved as two-dimensional bins within a particular spectral range and normalized. A 2D bin with a single carbon-to-carbon connectivity was assigned an area of 100, and a bin with two carbon-to-carbon connections in a bin was assigned an area of 200, and so forth. This way of representing bin populations assured that all the carbon-to-carbon connections would have a similar signal-to-noise ratio.

Structural assignments from the predicted $^{13}C$ NMR spectra were used to produce predicted 2D $^{13}C$-$^{13}C$ COSY and theoretical 2D $^{13}C$-$^{13}C$ distance spectra. The arrows in Figure 10*a* show the through-bond neighboring carbon-to-carbon connections of a 3β-estradiol molecular backbone without any side chains. These through-bond carbon-to-carbon connections were used to simulate a 2D $^{13}C$-$^{13}C$ COSY spectrum of the steroid compounds. The arrows in Figures 3*a* to 3*d* show the short, medium, and long through space carbon-to-carbon connections. The atom-to-atom connections are broken into four 2D planes. The bin widths can be set by the model developer. In the case shown in Figure 3, the first plane represented nearest neighbor pairs, the second plane included short-distance atom-to-atom connections ($2.0\,Å < r_{ij} < 3.6\,Å$); the third plane, medium-distance atom-to-atom connections ($3.6\,Å < r_{ij} < 6.0\,Å$); and the fourth plane, all long-distance atom-to-atom connections ($r_{ij} > 6.0\,Å$).

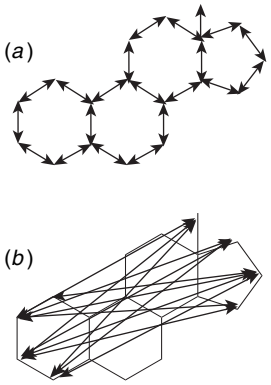


**Figure 10** (*a*) Arrows represent the carbon-to-carbon through nearest-neighbor bond connections for the backbone of a steroid used in the predicted 2D $^{13}C$-$^{13}C$ COSY spectra. (*b*) Arrows represent the carbon-to-carbon long-range space connections used in the theoretical 2D $^{13}C$-$^{13}C$ distance spectra.

***2D CoSCoSA Model of 130 Compounds Binding to the Estrogen Receptor***

Table 3 shows the log of the relative binding activity (RBA) data for 130 structurally diverse compounds used to train 2D CoSCoSA models [38]. The data set was produced at the National Center for Toxicological Research (NCTR) using a competitive estrogen receptor (ER) binding assay with radiolabeled estradiol ($[^3H]E_2$ ) in rat uterine cytosol obtained from ovariectomized uteri of Sprague–Dawley rats [8, 43, 44]. This data set spanned 7 orders in magnitude ranging from a log(RBA) value of –4 for a compound with weak binding affinity to the estrogen receptor, to a log(RBA) of 2 for a compound with strong binding affinity to the estrogen receptor. For a particular molecule, the RBA to the estrogen receptor is defined as 100 times the ratio of the molar concentrations of 17β-estradiol and the competing compound required to decrease the amount of receptor-bound 17β-estradiol by 50 percent. Thus 17β-estradiol had an RBA of 100 and a $\log_{10}$(RBA) of 2.0.

For each of the 130 compounds, the $^{13}C$ 2D $^{13}C$-$^{13}C$ COSY NMR experiment was simulated using the ACD/Labs CNMR predictor software (ACD, Toronto, Canada, www.acdlabs.com). Figures 3*a* and 10*a*, shows the carbon-to-carbon nearest neighbor connections needed to produce the 2D $^{13}C$-$^{13}C$ COSY NMR spectral data to develop a model for 130 diverse compounds whose RBAs to the estrogen receptor are shown in Table 3 [39]. The simulated 2D $^{13}C$-$^{13}C$ COSY NMR spectra were formed by using the NMR spectral assignments obtained from predicted carbon chemical shifts to identify nearest neighboring carbon atoms and establish carbon-to-carbon through-bond connectivity spectral patterns of each compound. The 2D $^{13}C$-$^{13}C$ COSY spectra for all 130 compounds were binned into 2.0 ppm by 2.0 ppm square bins.

All statistical analysis was performed by Statistica software (Statsoft, Tulsa, OK). Because this was a very diverse data set for each CoSCoSA model, we

**TABLE 3 130 Compounds and Training Set Predictions of Estrogen Receptor Binding [35, 39]**

| Compound Name | Compound Number | Exp. log(RBA) | 16 Bin CoSCoSA log(RBA) | 15 Bin + $L_{<7.5Å}$ CoSCoSA log(RBA) |
|---|---|---|---|---|
| Diethylstillbesterol | 1 | 2.60 | 1.44 | 1.43 |
| *meso*-Hexestrol | 2 | 2.48 | 2.86 | 2.70 |
| Ethinyl estradiol | 3 | 2.28 | 1.44 | 1.43 |
| 4-Hydroxyestradiol | 4 | 2.24 | 2.45 | 2.39 |
| 4-Hydroxytamoxifen | 5 | 2.24 | 0.58 | 0.58 |
| 17β-Estradiol | 6 | 2.00 | 2.36 | 1.85 |
| α-Zearalenol | 7 | 1.63 | 0.51 | 0.51 |
| ICI182780 | 8 | 1.57 | 1.08 | 1.12 |
| Dienestrol | 9 | 1.57 | 1.44 | 1.43 |
| α-Zearalanol | 10 | 1.48 | 0.51 | 0.51 |
| 2-Hydroxyestradiol | 11 | 1.47 | 1.26 | 1.32 |
| Diethylstilbestrol monomethyl ether | 12 | 1.31 | 1.44 | 1.43 |
| 3,3′-Dihydroxyhestrol | 13 | 1.19 | 0.75 | 0.60 |
| Droloxifene | 14 | 1.18 | 1.63 | 1.63 |
| Dimethylstibestrol | 15 | 1.16 | 0.04 | 0.15 |
| ICI164384 | 16 | 1.16 | 1.08 | 1.12 |
| Moxestrol | 17 | 1.14 | 1.44 | 1.43 |
| 17-Deoxyestradiol | 18 | 1.14 | 0.62 | 0.79 |
| 2,6-Dimethylhexestrol | 19 | 1.11 | 0.64 | 0.62 |
| Estriol | 20 | 0.99 | 0.62 | 0.79 |
| Monomethyl ether hexestrol | 21 | 0.97 | 0.51 | 0.93 |
| Estrone | 22 | 0.86 | 0.62 | 0.79 |
| p-*meso*-Phenol | 23 | 0.6 | 1.35 | 1.26 |
| 17α-Estradiol | 24 | 0.49 | 1.17 | 0.79 |
| Dihydroxymethoxychlorolefin | 25 | 0.42 | –0.10 | –0.10 |
| Mestranol | 26 | 0.35 | 1.44 | 1.43 |
| Zearalanone | 27 | 0.32 | 0.51 | 0.51 |
| Tamoxifen citrate | 28 | 0.21 | 0.58 | 0.58 |
| Toremifene citrate | 29 | 0.14 | 0.58 | 0.58 |
| α,α-Dimethylbethyl allenolic acid | 30 | –0.02 | –0.04 | –0.02 |
| Coumestrol | 31 | –0.05 | 0.43 | 0.05 |
| 4-Ethyl-7-OH-(p-meoxyphenol) dihydro-1-benzopyran-2-one | 32 | –0.05 | 0.11 | 0.15 |
| Nafoxidine | 33 | –0.14 | 0.58 | 0.58 |
| Clomiphene citrate | 34 | –0.14 | –0.47 | –0.47 |
| 1,3,5-Estratrien-3,6α-17β-triol | 35 | –0.15 | –0.60 | –0.61 |
| β-Zearalanol | 36 | –0.19 | 0.51 | 0.51 |
| 3-OH-estra-1,3,5-trien-16-one | 37 | –0.29 | –0.08 | 0.25 |

**TABLE 3** ***Continued***

| Compound Name | Compound Number | Exp. log (RBA) | 16 Bin CoSCoSA log (RBA) | 15 Bin + $L_{<7.5Å}$ CoSCoSA log (RBA) |
|---|---|---|---|---|
| 3-Deoxyestradiol | 38 | –0.30 | –1.47 | –1.55 |
| 3,6,4′-Trihydroxyflavone | 39 | –0.35 | –0.33 | –0.35 |
| Denistein | 40 | –0.36 | –2.66 | –2.61 |
| 4,4′-Dihroxystilbene | 41 | –0.55 | –0.56 | –0.51 |
| Dihydroxymethoxychlor (HPTE) | 42 | –0.60 | –1.47 | –2.21 |
| Monohydroxymethoxychlorolefin | 43 | –0.63 | –0.10 | –0.10 |
| 2,3,4,5-TetraCl-4′-biphenylol | 44 | –0.64 | –1.61 | –1.56 |
| Norethynodrel | 45 | –0.67 | –2.66 | –2.61 |
| 2,2′,4,4′-Tetrahydroxybenzil | 46 | –0.68 | –0.80 | –0.81 |
| β-Zearalenol | 47 | –0.69 | 0.51 | 0.51 |
| 4,6-Dihydroxyflavone | 48 | –0.82 | –2.07 | –1.95 |
| Equol | 49 | –0.82 | –0.66 | 0.05 |
| Monohydroxymethoxychlor | 50 | –0.89 | –2.07 | –1.95 |
| 3β-Androstanediol | 51 | –0.92 | –2.66 | –2.61 |
| Bisphenol B | 52 | –1.07 | –2.66 | –2.61 |
| Phloretin | 53 | –1.16 | –0.80 | –0.81 |
| Dietheylstilbestrol dimethyl ether | 54 | –1.25 | –0.66 | –0.68 |
| 2′,4,4′-Trihydroxychalcone | 55 | –1.26 | –1.73 | –1.71 |
| 2,5-Dichloro-4′-biphenylol | 56 | –1.44 | –1.61 | –1.56 |
| 4,4′-[1,2-ethanediyl)bisphenol | 57 | –1.44 | –2.66 | –2.61 |
| 17β-Estradiol-16β-OH-16-methyl-3-ether | 58 | –1.48 | –1.34 | –1.34 |
| Aurin | 59 | –1.50 | –0.56 | –0.51 |
| Nordihydroguariareticacid | 60 | –1.51 | –2.66 | –2.61 |
| 4-Nonylphenol | 61 | –1.53 | –1.61 | –1.56 |
| Apigenin | 62 | –1.55 | –2.07 | –1.95 |
| Kaempferol | 63 | –1.61 | –2.66 | –2.61 |
| Daidzein | 64 | –1.65 | –1.61 | –1.56 |
| 3-Methylestriol | 65 | –1.65 | –1.34 | –1.34 |
| 4-Dodecylphenol | 66 | –1.73 | –2.66 | –2.61 |
| 2-Ethylhexyl-4-hydroxybenzoate | 67 | –1.74 | –2.66 | –2.61 |
| 4-*tert*-Octylphenol | 68 | –1.82 | –2.66 | –2.61 |
| Phenolphthalein | 69 | –1.87 | –1.47 | –1.30 |
| Kepone | 70 | –1.89 | –2.66 | –2.61 |
| Heptyl-4-hydroxybenzoate | 71 | –2.09 | –2.66 | –2.61 |
| Bisphenol A | 72 | –2.11 | –2.66 | –2.61 |
| Naringenin | 73 | –2.13 | –2.66 | –2.61 |
| 4-Cl-4′-biphenylol | 74 | –2.18 | –2.66 | –2.61 |
| 3-Deoxyestrone | 75 | –2.2 | –1.47 | –1.55 |
| 4-Octylphenol | 76 | –2.31 | –2.66 | –2.61 |

**TABLE 3** ***Continued***

| Compound Name | Compound Number | Exp. log (RBA) | 16 Bin CoSCoSA log (RBA) | 15 Bin + $L_{<7.5Å}$ CoSCoSA log (RBA) |
|---|---|---|---|---|
| Fisetin | 77 | –2.35 | –2.14 | –2.09 |
| 3′,4′,7-Trihydroxyisoflavone | 78 | –2.35 | –2.66 | –2.61 |
| Biochanin A | 79 | –2.37 | –2.66 | –2.61 |
| 4-OH-chalcone | 80 | –2.43 | –2.66 | –2.61 |
| 4′-OH-chalcone | 81 | –2.43 | –2.66 | –2.61 |
| 2,2′-Methylenebis [4-chlorophenol) | 82 | –2.45 | –2.07 | –1.95 |
| 4,4′-Dihydroxybenzophenone | 83 | –2.46 | –2.66 | –2.61 |
| Benzyl-4-hydroxybenzoate | 84 | –2.54 | –2.66 | –2.61 |
| 2,4-Dihyroxybenzophenone | 85 | –2.61 | –2.66 | –2.61 |
| 4′-Hydroxyflavanone | 86 | –2.65 | –3.15 | –2.96 |
| 3α-Androstanediol | 87 | –2.67 | –2.66 | –2.61 |
| 4-Phenethylphenol | 88 | –2.69 | –2.66 | –2.61 |
| Prunetin | 89 | –2.74 | –2.66 | –2.61 |
| Doisynoestrol | 90 | –2.74 | –2.66 | –2.61 |
| Myricetin | 91 | –2.75 | –2.66 | –2.61 |
| 2-Cl-4-biphenylol | 92 | –2.77 | –3.21 | –2.61 |
| Triphenylethylene | 93 | –2.78 | –2.66 | –2.61 |
| 3′-OH-flavanone | 94 | –2.78 | –3.43 | –3.31 |
| Chalcone | 95 | –2.82 | –2.66 | –2.61 |
| *o*,*p*′,-DDT | 96 | –2.85 | –2.66 | –2.61 |
| 4-Heptyloxyphenol | 97 | –2.88 | –2.66 | –2.61 |
| Dihydrotestosterone | 98 | –2.89 | –2.66 | –2.61 |
| Formononetin | 99 | –2.98 | –2.66 | –2.61 |
| bis-[4-Hydroxyphenyl)methane | 100 | –3.02 | –2.66 | –2.61 |
| *p*-Phenylphenol | 101 | –3.04 | –2.66 | –2.61 |
| 6-Hydroxyflavanone | 102 | –3.05 | –2.14 | –2.09 |
| 4,4′-Sulfonyldiphenol | 103 | –3.07 | –1.47 | –1.30 |
| Butyl-4-hydroxybenzoate | 104 | –3.07 | –2.66 | –2.61 |
| Diphenolic acid | 105 | –3.13 | –2.66 | –2.61 |
| 1,3-Diphenyltetramethyldisiloxane | 106 | –3.16 | –2.66 | –2.61 |
| Propyl-4-hydroxybenzoate | 107 | –3.22 | –2.66 | –2.61 |
| Ethyl-4-hydrobenzoate | 108 | –3.22 | –2.66 | –2.61 |
| Phenol red | 109 | –3.25 | –2.66 | –2.61 |
| 3,3′,5,5′-TetraCl-4,4′-biphenyldiol | 110 | –3.25 | –2.66 | –2.61 |
| 4-*tert*-Amylphenol | 111 | –3.26 | –2.66 | –3.53 |
| Baicalein | 112 | –3.35 | –2.66 | –2.61 |
| Morin | 113 | –3.35 | –2.66 | –2.61 |
| 4-sec-Butylphenol | 114 | –3.37 | –2.07 | –1.95 |
| 4-Cl-3-methylphenol | 115 | –3.38 | –2.66 | –3.53 |
| 6-Hydroxyflavone | 116 | –3.41 | –2.66 | –2.61 |
| 4-Benzyloxyphenol | 117 | –3.44 | –2.66 | –2.61 |

**TABLE 3** ***Continued***

| Compound Name | Compound Number | Exp. log (RBA) | 16 Bin CoSCoSA log (RBA) | 15 Bin + $L_{<7.5Å}$ CoSCoSA log (RBA) |
|---|---|---|---|---|
| 3-Phenylphenol | 118 | –3.44 | –2.14 | –2.09 |
| Methyl-4-hydrobenzoate | 119 | –3.44 | –2.66 | –3.53 |
| 2-*sec*-Butylphenol | 120 | –3.54 | –3.20 | –3.04 |
| 2,4′-Dichlorobiphenyl | 121 | –3.61 | –2.66 | –2.61 |
| 4-*tert*-Butylphenol | 122 | –3.61 | –3.75 | –3.53 |
| 2-Cl-4-methylphenol | 123 | –3.66 | –2.66 | –3.53 |
| Phenolphthalin | 124 | –3.67 | –2.66 | –2.61 |
| 4-Cl-2-methylphenol | 125 | –3.67 | –2.66 | –3.53 |
| 7-Hydroxyflavanone | 126 | –3.73 | –2.66 | –2.61 |
| 3-Ethylphenol | 127 | –3.87 | –2.90 | –3.70 |
| Rutin | 128 | –4.09 | –2.66 | –2.61 |
| 4-Ethylphenol | 129 | –4.17 | –3.75 | –3.53 |
| 4-Methylphenol | 130 | –4.50 | –3.75 | –3.53 |

used forward multiple linear regression (MLR) on a selected subset of spectral bins. After binning all 130 compounds, only 605 from the 7381 bins contained nonzero elements (called *hits*) in them. Of the 605 populated bins, only 337 contained more than one hit. From the remaining 337 multiply populated bins, an increasing number of the mostly highly correlated bins were selected by trial and error and used to construct MLR models until a model was obtained that had an $r^2$ greater than 0.8 and a $F$ value greater than 30. We were able to identify 16 bins this way and produce a model that had an $r^2$ greater than 0.82, the $F$ value was still increasing with the addition of a bin, and the $p$ value of the bin added was significant ($p < 0.05$). We did not select any bins that had less than 2 hits. The reason for this is that a bin with one hit can inappropriately add to the $r^2$ of a model but cannot improve the LOO cross-validation ($q_1^2$). The use of a large number of very small, singly populated bins is the reason that Bursi and co-workers [45] had a high $r^2$ and very low $q_1^2$. To more rigorously test the validity of the CoSCoSA models, two L13O [10 percent of the data excluded) cross-validations were performed on each of the models. In these "leave-multiple-samples-out" experiments, the compounds omitted were varied and the results of the two corresponding experiments were averaged.

Figure 11*a* shows results of a CoSCoSA model based on the MLR analysis of 16 selected 2D bins from the $^{13}C$-$^{13}C$ COSY spectral data. The 16-bin COSY model for the 130 estrogenic compounds had an explained variance $r^2$ of 0.827, a LOO $q_1^2$ of 0.78, and an average L13O cross-validated variance ($q_{13}^2$) of 0.78 ± 0.01. The CoSCoSA model was based on COSY bins 28-12 (bin 1), 72-20 (bin 2), 54-28 (bin 3), 50-38 (bin 4), 64-56 (bin 5), 164-104 (bin

6), 152-108 (bin 7), 156-110 (bin 8), 126-112 (bin 9), 140-112 (bin 10), 142-112 (bin 11), 154-112 (bin 12), 154-114 (bin 13), 156-114 (bin 14), 128-116 (bin 15), and 126-120 (bin 16). All 2.0 ppm bins were written using the format *a-b*, where *a* and *b* are the ppm values corresponding to the two "connected" atoms. The MLR equation for the 16-bin CoSCoSA model is

$$\begin{aligned}\text{Predicted } \log_{10}(\text{RBA}) = {} & 0.00999 * \text{bin } 1 + 0.03173 * \text{bin } 2 + 0.0071 * \text{bin } 3 \\ & + 0.01196 * \text{bin } 4 + 0.02191 * \text{bin } 5 + 0.0093 * \text{bin } 6 \\ & + 0.2329 * \text{bin } 7 + 0.01324 * \text{bin } 8 + 0.00737 * \text{bin } 9 \\ & + 0.02558 * \text{bin } 10 + 0.0392 * \text{bin } 11 + 0.03094 * \text{bin } 12 \\ & - 0.00545 * \text{bin } 13 + 0.00526 * \text{bin } 14 + 0.00298 * \text{bin } 15 \\ & - 0.00768 * \text{bin } 16\end{aligned}$$

All bins had more than three hits except for bins 152-108 and 140-112, each of which had only two hits. The correlation matrix for the 16 bins was calculated and only two sets of bins had a correlation between them greater than 0.5. The greatest average correlation between any bin with the other 17 bins was 0.04 and many of the average correlations were much lower than 0.04. The lack of a large correlation among bins suggests that the resulting patterns were based on essentially orthogonal data. The COSY bin 28-12 was most often associated with the $CH_3$ carbon connected to the $CH_2$ in the ethyl groups in diethylstilbestrol and hexestrol-like compounds. Twelve of the 14 compounds with a COSY hit in 28-12 had a $\log_{10}$(RBA) greater than –0.05. Compounds that populated a COSY bin at 154-112 were most often associated with the 3-carbon position connected to the 2-carbon position in the A-ring of 17β-estradiol like compounds. Nine of the 10 compounds with a COSY hit in bin 154-112 had a $\log_{10}$(RBA) greater than –0.05. Fourteen compounds had a hit in the COSY bin at 128-116. The 128-116 COSY bin was most often associated with the 2- to 3- and 5- to 6-carbon positions in a phenol ring. Twelve of the 14 compounds with a COSY hit in bin 128-116 had a $\log_{10}$(RBA) less than 0.60. The 24 compounds that had a hit or multiple hits in the COSY bin at 156-114 were most often associated with the hydroxylated carbon of a phenol ring connected to its two nearest neighboring carbons. Only 5 of the 24 compounds with a COSY hit in bin 156-114 had $\log_{10}$(RBA) less than –1.65. The 6 compounds that had a COSY bin at 64-56 were most often associated with the 2 carbons between the oxygen ester and the nitrodimethyl of tamoxifen-like compounds. Similar spectral-structural associations could be made for the other COSY bins effectively used in the CoSCoSA models.

Figure 11*b* shows results for the CoSCoSA model that was based on the MLR analysis of 15 selected 2D $^{13}C$-$^{13}C$ COSY bins plus the single distance variable. The distance variable, $L_{<7.5Å}$, was assigned a value of 1 when the maximum distance between nonhydrogen atoms in a compound was less than 7.5 Å (compact) and a value of zero for all other compounds. The $L_{<7.5Å}$ vari-

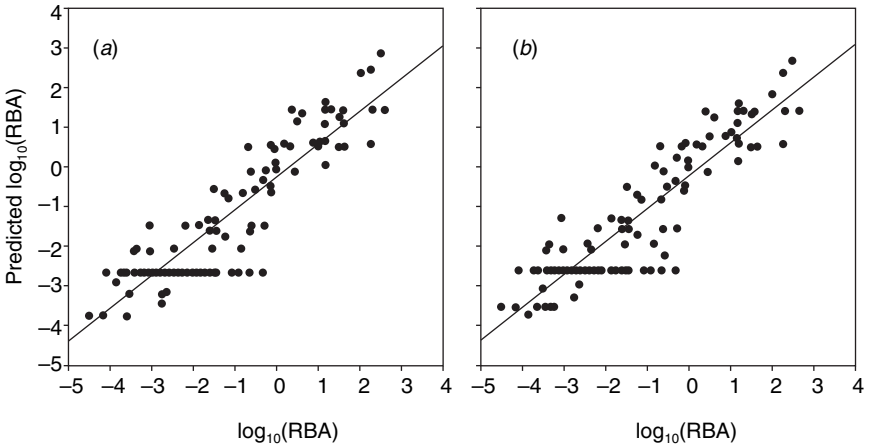


**Figure 11** (*a*) 16 bin 2D-CoSCoSA model of estrogen binding. (*b*) 15 + L < 7.5 2D CoSCoSA model of estrogen binding.

able replaced the COSY bin at 154-114 (bin 13) in the previous 16-bin CoSCoSA model. The MLR equation for the 15-bin-with-$L_{<7.5Å}$ CoSCoSA model is

$$\begin{aligned}\text{Predicted log(RBA)} = {} & 0.00969 * \text{bin } 1 + 0.03122 * \text{bin } 2 + 0.00637 * \text{bin } 3 \\ & + 0.01066 * \text{bin } 4 + 0.02142 * \text{bin } 5 + 0.00902 * \text{bin } 6 \\ & + 0.2263 * \text{bin } 7 + 0.01275 * \text{bin } 8 + 0.00732 * \text{bin } 9 \\ & + 0.02507 * \text{bin } 10 + 0.03934 * \text{bin } 11 + 0.02666 * \text{bin } 12 \\ & + 0.00526 * \text{bin } 14 + 0.00329 * \text{bin } 15 - 0.00701 * \text{bin } 16 \\ & - 0.91773 * L_{<7.5Å}\end{aligned}$$

This 15-bin-with-$L_{<7.5Å}$ model had an $r^2$ of 0.83, a $q_1^2$ of 0.79, and an average $q_{13}^2$ of 0.78 ± 0.01. In this model, the $L_{<7.5Å}$ variable selected 9 compounds, all of which had a log(RBA) lower than –3.26. Smaller, compact molecules tended to bind weakly.

In Figures 11*a* and 11*b*, the line of compounds predicted to have a $\log_{10}$(RBA) of –2.60 is a set of compounds that did not have a hit in any of the 16 bins used to formulate the two CoSCoSA models. There were 56 compounds in the 16-bin CoSCoSA model that did not have any hit in one of the 16 bins. The 15-bin-with-$L_{<7.5Å}$ model had 52 compounds that did not have a hit in any of the 15 bins or $L_{<7.5Å}$. The removal of compounds with no hits from the CoSCoSA models did not change the $r^2$ or $q^2$ of the model by more than 2 percent. Three of the compounds in the 16-bin CoSCoSA model had residuals greater than 2 standard deviations (3β-androstanediol, genistein, and norethynodrel). In the 16-bin CoSCoSA model, only one other compound had

a predicted residual greater than 2 standard deviations and it was 4-hydoxy-tamoxifen. The 15-bin-with-$L_{<7.5Å}$ model had 4 compounds with no hits that had residuals greater than 2 standard deviations. They are the same 3 compounds poorly modeled by the 16-bin CoSCoSA model plus 4,4′-sulfonyldiphenol. There were two compounds in the 15-bin-with-$L_{<7.5Å}$ CoSCoSA model with predicted residuals greater than 2 standard deviations, 4-hydoxy-tamoxifen and dihydroxymethoxychlor (HPTE). Almost all of the compounds with no hits in the 16 bins had experimental $\log_{10}$(RBA) lower than –1.0. The CoSCoSA models did not find a spectral relationship for these weakly binding compounds to the estrogen receptor. Most of the other bins in both CoSCoSA models were used to form a relationship between a spectral bin and binding to the estrogen receptor with a $\log_{10}$(RBA) stronger than –2.60.

To further test the ruggedness of CoSCoSA models of estrogen receptor binding, we predicted the $\log_{10}$(RBA)s of compounds from two published external data sets [46, 47]. The $\log_{10}$(RBA)s from those external data sets possessed a greater variability in binding activity. So, by a previously published method, a set of compounds that had their binding activity determined in all three labs [46, 47] was used to normalize the external data sets to the NCTR data [8, 48]. We then built the CoSCoSA models as before and used the resulting MLR equations to predict the $\log_{10}$(RBA) of the compounds in the test set. We used the published normalized $\log_{10}$(RBA) for 27 compounds from Waller and Kuiper data in our external testing of the CoSCoSA models [8]. However, many of the occupied bins for the new compounds from the external data set did not fall into the original 605 occupied bins. (The original set of bins comprised only 8.2 percent of the 2D COSY spectral plane.) We inferred that, in the different molecular context of the external data sets, NMR chemical shift information was expressed in adjacent but nonincluded bins. NMR chemical shifts exist along a continuum, and the process of binning them for this type of pattern recognition inherently compromises the pattern when it barely misses a selected bin.

To account for this source of confusion with the external data, we tried adding various fractions of "near-miss" signals into each compound's spectrum. With this in mind, we used the CoSCoSA model's MLR equation to predict the normalized log(RBA) of the compounds in the external test set. However, compounds from the external test set with bins that were one bin away (one of 8 bins surrounding a 2D bin) from the original 605 populated bins were modeled using none, one-quarter, and one-half of that bin's intensity in the nearest neighboring bin used in the original CoSCoSA model.

Table 64 shows the predictions for 27 compounds. The first 21 compounds show the predictions for Waller's data set [46] using both the 16-bin and 15-bin-plus-$L_{<7.5Å}$ model of estrogen binding. To make the predictions, we simulated the 2D spectra of the 21 compounds, again using ACD/Labs CNMR predictor software (ACD, Toronto, Canada, www.acdlabs.com). The simulated spectra of the test set were binned into the same 605 bins. However, many of the occupied bins for these compounds did not fall into the original 605 occu-

**TABLE 4 External Test Set Predictions of Estrogen Receptor Binding**[a]

| Name | Normalized log (RBA) | 16-CoSASA | 15+$L_{<7.5Å}$-CoSCoSA | CoMFA |
|---|---|---|---|---|
| 2-*tert*-Butylphenol | –4.55 | –2.66 | –3.53 | –3.83 |
| 3-*tert*-Butylphenol | –4.82 | –1.50 ± 0.64 | –2.34 ± 0.66 | –3.33 |
| 2,4,6,-triCl-4′-biphenylol | –0.16 | –1.61 | –1.56 | –1.60 |
| 2-Cl-4,4′-biphenyldiol | –0.61 | –1.61 | –1.56 | –1.49 |
| 2,6-Dichloro-4′-biphenylol | –1.11 | –1.61 | –1.56 | –2.41 |
| 2,3,5,6,TetraCl-4,4′-biphenyldiol | –2.18 | –1.61 | –1.56 | –0.82 |
| 2,2′,3,3′,6,6′-HexaCl-4-biphenylol | –2.74 | –2.14 | –2.08 | –3.06 |
| 2,2′,3,4′,6,6′-HexaCl-4-biphenylol | –2.60 | –1.61 | –1.56 | –2.48 |
| 2,2′,3,6,6′-PentaCl-4-biphenylol | –1.97 | –1.61 | –1.56 | –3.07 |
| 2,2′5,5′-TetraCl-biphenyl | –2.67 | –2.66 | –2.61 | –2.74 |
| 2,2′,4,4′,5,5′-HeaxCl-biphenyl | –2.83 | –2.66 | –2.61 | –1.52 |
| 2,2′,4,4′,6,6′-HexaCl-biphenyl | –1.87 | –2.66 | –2.61 | –1.83 |
| 2,2′,3,3′,5,5′-HeaxCl-6′-biphenylol | –2.69 | –2.36 | –2.29 | –3.01 |
| 4′-Deoxyindenestrol | –1.37 | 2.89 ± 0.63 | 2.96 ± 0.67 | –0.53 |
| 4′-Deoxyindenestrol | –0.23 | 2.89 ± 0.63 | 2.96 ± 0.67 | 0.111 |
| 5′-Deoxyindenestrol | –0.59 | –0.61 | –0.59 | –1.00 |
| 5′-Deoxyindenestrol | 0.35 | –0.61 | –0.59 | –0.59 |
| Indenestrol A (R) | 1.08 | 3.95 ± 0.64 | 4.01 ± 0.67 | 0.29 |
| Indenestrol A (S) | 2.39 | 3.95 ± 0.64 | 4.01 ± 0.67 | 0.62 |
| R 5020 | –1.81 | –2.45 ± 0.18 | –2.48 ± 0.17 | –0.70 |
| Zearalenone | 0.91 | 0.51 | 0.51 | –0.12 |
| 5-Androstenediol | –0.49 | –2.66 | –2.61 | –0.66 |
| 16a-Bromoestradiol | 1.41 | –0.11 | 0.05 | 0.33 |
| 16-Ketoestradiol | –0.38 | –0.11 | 0.05 | 0.58 |
| 17-epi-Estriol | 0.98 | –0.11 | 0.05 | –0.16 |
| 2-OH-estrone | –0.19 | 1.26 | 1.32 | 0.36 |
| Raloxifene | 1.34 | 0.17 ± 0.63 | 0.20 ± 0.66 | –0.24 |

[a] The plus and minus sign reveals the variation seen when using none and one-half of a bin's intensity in a neighboring bins used to formulate the CoSCoSA model [35, 39].

pied bins (8.2 percent of the 2D COSY spectral plane). Therefore, if the simulated spectra did not fall into one of the original 605 populated bins, we put none, one-quarter, and one-half of the bin's intensity into the neighboring bin or bins used in the CoSCoSA model. We then built the CoSCoSA model as before and used its MLR equation to predict the $\log_{10}$(RBA) of the compounds in the test set. Only 6 of the 27 compounds from Waller and Kuiper's

external data sets had binned COSY chemical shifts that were within one bin of those 16 bins used to formulate a CoSCoSA model. In Table 4, for these 6 compounds, we predicted the $\log_{10}$(RBA) using one-quarter intensity in a neighboring bin and plus or minus the deviation seen when predicting the $\log_{10}$(RBA) when using none and one-half intensity in the neighboring bin used for a CoSCoSA model. For Waller's test set and one quarter of a bin's intensity in neighboring bins, we achieved a $q^2_{pred}$ of 0.50 for the 16-bin CoSCoSA model and a $q^2_{pred}$ 0.57 for the 15-bin-plus-$L_{<7.5Å}$ CoSCoSA model. When using one half of a bin's intensity in a neighboring bin, we got a $q^2_{pred}$ of 0.45 for the 16-bin CoSCoSA model and a $q^2_{pred}$ 0.52 for the 15-bin-plus-$L_{<7.5Å}$ CoSCoSA model. Using none of a bin's intensity in a neighboring bin, we got a $q^2_{pred}$ of 0.55 for the 16-bin CoSCoSA model and a $q^2_{pred}$ of 0.62 for the 15-bin-plus-$L_{<7.5Å}$ CoSCoSA model. A comparative mean field analysis (CoMFA) model had a $q^2_{pred}$ of 0.70 for Waller's normalized test set [8, 46]. When Indenstrol A (R), Indenestrol A (S), 4′deoxyindenestrol (R), and 4′-deoxyindenestrol (S) are removed from Waller's test set and none of a bin's intensity from neighboring bins, a $q^2_{pred}$ of 0.59 for the 16-bin model and a $q^2_{pred}$ 0.74 for the 15-bin-plus-$L_{<7.5Å}$ model are achieved. Further inspection of the predictions for Indenstrol A (R), Indenestrol A (S), 4′deoxyindenestrol (R), and 4′-deoxyindenestrol (S) revealed that one chemical shift prediction that was highly suspect (142 ppm). When we checked the structures used in the prediction of this chemical shift, all the compounds had chemical shifts from 134 to 139 ppm and not 142 ppm. On closer examination of the spectral prediction process, we found that the database structures used as the basis for prediction all had corresponding chemical shifts from 134 to 139 ppm, not 142 ppm. It appears that an error in the spectral prediction process may have contributed to poorer modeling results for these outliers.

***3D-CoSCoSA Model of Corticosterone Binding*** A three-dimensional quantitative spectrometric data–activity relationship (3D QSDAR) model of corticosteroid binding globulin was developed by combining NMR spectral information with structural information to form a 3D connectivity matrix. The 3D connectivity matrix was built by displaying all possible carbon-to-carbon connections with their assigned carbon NMR chemical shifts and distances between the carbons. The matrix was simplified by compressing data into a 2-D $^{13}$C-$^{13}$C COSY plane and selected theoretical 2D $^{13}$C-$^{13}$C distance connectivity spectral slices to model binding for 30 steroids. Not all the atom-to-atom connectivity information is needed to develop a good CoSCoSA model. In the 2D CoSCoAS model of estrogen binding just presented, we used only the nearest neighbor COSY data. It is known for many steroids that the important binding sites are near the 3 and 17 positions of the molecule. Therefore, effective models can be built using only the nearest neighbor information and the long-range connections between areas surrounding positions 3 and 17 that are approximately 7.5 Å apart. To include information around positions 3 and 7, through-space carbon-to-carbon connections that were greater than 6.0 Å

were included to produce a theoretical 2D $^{13}C$-$^{13}C$ distance connectivity spectrum that contains cross peaks, for atom *i* and atom *j* whenever the two carbons were greater than 6.0 Å apart. The 2D $^{13}C$-$^{13}C$ COSY and 2D $^{13}C$-$^{13}C$ distance connectivity spectra are symmetrical across the diagonal and for modeling purposes, only half of each individual spectrum contains all the necessary information. One-dimensional $^{13}C$ NMR spectra were not used in these CoSCoSA models because the 1D chemical shifts are highly correlated to the COSY and long-range cross peaks and do not provide additional information.

Structurally assigned $^{13}C$ NMR spectra of the 30 steroids in Figure 6 and Table 1 were used to produce the predicted 2D $^{13}C$-$^{13}C$ COSY and theoretical 2D $^{13}C$-$^{13}C$ distance spectra [41]. The arrows in Figure 3*a* show the through-bond neighboring carbon-to-carbon connections of a steroid backbone molecule without any side chains. These through-bond carbon-to-carbon connections were used to simulate 2D $^{13}C$-$^{13}C$ COSY spectra of the steroid compounds. The arrows in Figure 10*b* show all through-space carbon-to-carbon connections greater than 6.9 Å for β-estradiol. These through-space carbon-to-carbon connections and any other long-distance connections from side chains were used to produce the theoretical 2D $^{13}C$-$^{13}C$ distance connectivity spectra.

Four CoSCoSA models of binding to the corticosteroid binding globulin were built from the 2D COSY and 2D long-range distance spectra. One model used only the 2D COSY and the other, only the 2D long-range distance spectra. Also, COSY and distance spectra were combined in two different ways, either using the combined raw spectra (3D) or using the combined principal components (PCs) from the COSY and distance PCs calculated separately.

For the first case, 2D $^{13}C$-$^{13}C$ COSY spectra data were reduced to PCs. The PCs were used by forward multiple linear regression to produce the COSY model. Similarly, for the second approach, 2D $^{13}C$-$^{13}C$ distance connectivity data were reduced to PCs. These PCs were then used via forward multiple linear regression to produce a model of the long-distance connectivity data. The third type of model combined the raw 2D $^{13}C$-$^{13}C$ COSY and the 2D $^{13}C$-$^{13}C$ distance connectivity spectra into a single 3D data set before PC extraction [41]. These PCs were then used for forward multiple linear regression to produce the model for the three-dimensional representation of the 2D $^{13}C$-$^{13}C$ COSY and 2D $^{13}C$-$^{13}C$ distance spectra data. The fourth approach used the combined PCs from the individual 2D $^{13}C$-$^{13}C$ COSY and the 2D $^{13}C$-$^{13}C$ distance connectivity models to produce a combined through-bond and through-space CoSCoSA model.

The PCs for all CoSCoSA model were produced by evaluating the connectivity bins with forward multiple linear regression analysis and using only the most correlated PCs. The *F* test for many of the models continued to rise as more PCs were included in modeling. This process can lead to overfitting of the model. To avoid overfitting, we limited the number of PCs to either three or eight. Three reflected the number of independent components used

in previous QSAR models for this endpoint. Eight reflected optimal results for this type of data.

Table 5 contains a comparison of the model performance parameters ($n$, $r^2$, $q^2$, and number of components) for corticosterone binding models based on QSAR [37], HE-state/E-state [52], E-state [52], SOM [53], combination QSAR/E-state [38], 2D CoSASA [22], 1D CoSA [22], the four multiple linear regression CoSCoSA models, and a CoSCoSA PD-ANN model validated two different ways (explained below). All four CoSCoSA models with 8 PCs have a strong correlation ($r^2$) and cross-validated variance ($q^2$) and are favorable when compared to the previous published models of binding to the corticosteroid binding globulin.

Figure 12*a* is a plot of the predicted binding versus experimental binding for the CoSCoSA 2.0-ppm resolution model based on $^{13}$C-$^{13}$C COSY data. A model based on 8 PCs had an explained correlation ($r^2$) of 0.93 and a cross-validated variance ($q^2$) of 0.88, which indicates self-consistency and excellent predictive capability. Figure 12*b* is a plot of the predicted binding versus experimental binding for the CoSCoSA 2.0-ppm resolution model based on $^{13}$C-$^{13}$C distance greater than 6.9 Å. Using 8 PCs, the $r^2$ of this model is 0.89 and the $q^2$ is 0.79. Figure 12*c* is a plot of the predicted binding versus experimental binding for the CoSCoSA 2.0-ppm resolution model based on the combined $^{13}$C-$^{13}$C COSY and $^{13}$C-$^{13}$C distance connectivity PCs. The $r^2$ of this model is 0.96 and the $q^2$ of this model is 0.92, which indicates excellent predictive capability. Finally Figure 12*d* is a plot of the predicted binding

**TABLE 5 Corticosterone Model Performance Parameters n, r², q² and Number of Components**

| Model | $n$ | $r^2$ | $q^2$ | Components |
|---|---|---|---|---|
| QSAR | 31 | 0.72 | 0.68[a] | 3 (PCs) |
| HE state/E-state | 31 | 0.98[a]/0.96[b] | 0.80[a]/0.76[b] | 3[a] (PCs)/5[b] (PCs) |
| E-state | 31 | 0.96[a]/0.96[b] | 0.79[a]/0.67[b] | 3[a](PCs)/4[b] (PCs) |
| SOM | 31 | 0.85 | — | 3 (PCs) |
| QSAR/E-state | 30 | 0.82 | 0.78 | 3 (atoms) |
| 2D CoSASA | 30 | 0.80 | 0.73 | 3 (atoms) |
| 1D CoSA | 30 | 0.80 | 0.78 | 3 (bins) |
| CoSCoSA COSY | 30 | 0.84/0.93 | 0.74/0.88 | 3 (PCs)/8 (PCs) |
| CoSCoSA distance | 30 | 0.66/0.89 | 0.38/0.79 | 3 (PCs)/8 (PCs) |
| CoSCoSA COSY + distance | 30 | 0.84/0.96 | 0.74/0.92 | 3 (PCs)/8 (PCs) |
| 3D[c] CoSCoSA | 30 | 0.78/0.92 | 0.68/0.81 | 3 (PCs)/8 (PCs) |
| PD-ANN (3D[c]) CoSCoSA | 30 | 0.96 | 0.78[d] | 593:198:1 |
| PD-ANN (3D[c]) CoSCoSA | 30 | 0.96 | 0.73[e] | 593:198:1 |

[a] 1.0-Å models.
[b] 2.0-Å models.
[c] 3D is the combination of COSY and distance data before PC extraction.
[d] A L3O $q_3^2$.
[e] A L10O $q_{10}^2$ [22, 36, 37, 40, 51, 52].

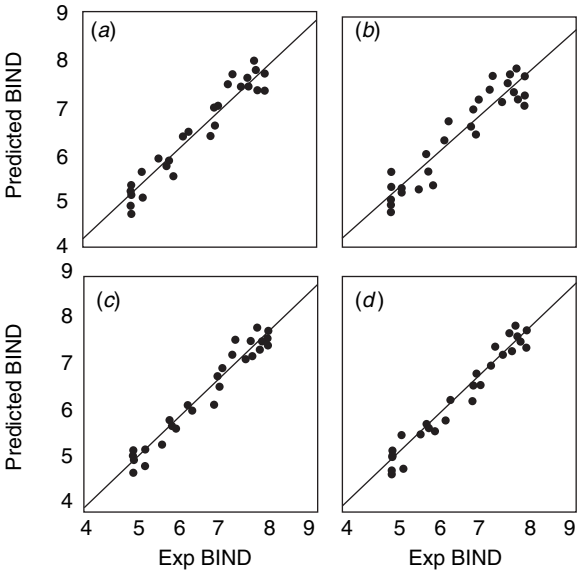


**Figure 12** Plot of the predicted binding versus experimental binding for the corticosterone globulin. (*a*) The $^{13}C$-$^{13}C$ COSY data, (*b*) $^{13}C$-$^{13}C$ distance connectivity data greater than 6.9 Å, (*c*) the combined $^{13}C$-$^{13}C$ COSY and $^{13}C$-$^{13}C$ distance connectivity PCs, (*d*) the combined $^{13}C$-$^{13}C$ COSY and $^{13}C$-$^{13}C$ distance connectivity data before PCs are extracted from the combined data.

versus experimental binding for the CoSCoSA 2.0-ppm resolution model based on the combined $^{13}C$-$^{13}C$ COSY and $^{13}C$-$^{13}C$ distance connectivity spectral data before principal component extraction. The explained correlation ($r^2$) of this model is 0.92 and the cross-validated variance ($q^2$) of this model is 0.81.

***Artificial Neural Network Model of 3D CoSCoSA Model of Corticosterone Binding*** Another model for predicting steroid binding activity to the corticosteroid binding globulin was developed using a parallel distributed-artificial neural network (PD-ANN). The PD-ANN is an algorithm that allows for simultaneous training and testing of multiple neural networks. The drawback to using an ANN is the time required to find an optimal network configuration. Traditionally this search has been performed in serial fashion, one configuration at a time, on one computer at a time. The PD-ANN takes advantage of an Internet-based parallel distribution scheme to perform the optimization task on multiple machines simultaneously. The result is a dramatic decrease in the time required for finding an optimal network configuration. For this study, the ANN consisted of an error back propagating, feed-forward Java code. The back-propagation algorithm in the PD-ANN was based on Rummelhart, Hinton, and Williams' generalized delta rule [49]. It

was developed "in-house" and parallel distributed using JGravity, a Java-based parallel distribution platform [50].

More than 100 varying network configurations were evaluated over a couple of days using four high-speed Pentium personal computers. Two types of cross-validations were performed. In one case, each distributed neural network removed three examples and trained on the rest, rotating through the rest of the data on subsequent training sessions, and thus performing a comprehensive series of L3O cross-validations. In another case, 10 examples were left out, performing a L10O cross-validation.

The configuration consisting of 593 input, 198 hidden, and 1 output nodes provided the best validated results. This roughly corresponds to a hidden layer that is equal to $1/3n$, where $n$ = number of input nodes. This is a common optimal input-hidden node relationship for three-layer, feed-forward neural networks [51]. The best results were obtained when the network was trained for 3400 cycles, with a learning rate of 0.1, and using a sigmoid transfer function. A sigmoid transfer function in the back-propagation algorithm required that the data (input and output) be scaled within a range between 0 and 1. The raw data for this model was actually scaled from 0.1 to 0.9.

All of the spectral bins based on the combination of $^{13}C$-$^{13}C$ COSY and $^{13}C$-$^{13}C$ distance spectra were used to develop the CoSCoSA PD-ANN model. The explained correlation ($r^2$) of this model is 0.96. Two cross-validated variance coefficients were calculated for this model. The leave-3-out ($q_3^2$) procedure yielded a value of 0.78, and the leave-10-out ($q_{10}^2$) procedure yielded a value of 0.73. These numbers illustrate a high degree of consistency and predictive capability for the PD-ANN model. All four CoSCoSA models based on 8 PCs have a $q^2$ greater than the 0.68 seen for the QSAR model. Three of the 4 CoSCoSA models based on 3 PCs have a $q^2$ greater than the 0.68 seen for the QSAR model. The only CoSCoSA model that did not have a $q^2$ greater than 0.68 was the $^{13}C$-$^{13}C$ distance connectivity model based on only 3 PCs. The HE-state and E-state models have a greater $r^2$ than all the QSDAR models, but these models are very computationally intensive with many distance formulas used for every point in the grid. Still, all the 2.0-ppm resolution CoSCoSA models based on 8 PCs have an $r^2$ greater than 0.89 and a $q^2$ greater than 0.79. All the CoSCoSA models with 8 PCs have a predictability that is much better or comparable to the predictability for QSAR, CoSA, CoSASA, HE-state/E-state, and E-state models. Our earlier CoSA QSDAR model started with only 256 spectral bins, a number subsequently reduced to 94 when all columns were removed that consistently had either zero population or only one nonzero entry. An analogous adjustment was made for these 3D models. The 2.0-ppm CoSCoSA models started with 6441 bins, a number then reduced to 271 for the $^{13}C$-$^{13}C$ COSY data and 322 for the $^{13}C$-$^{13}C$ distance connectivity data when all the columns with only zeroes were removed. The models used less than 5 percent of the available 2D connectivity spectral space.

A set of rules was used to combine all bins with only one hit into a populated nearest neighbor bin so that information content from their occupancy

could be retained in the model. When several bins with multiple hits were equally close to the bin with one hit, the one-hit bin was added to the bin with the least number of hits. When all the bins with one hit were treated this way, the remaining bins for 2.0-ppm $^{13}C$-$^{13}C$ COSY modeling were reduced from 271 to 178. Similarly, the number of bins used for 2.0-ppm $^{13}C$-$^{13}C$ distance connectivity modeling was reduced from 322 to 194. Using the new data with treated one-hit bins, the resulting $r^2$ for the $^{13}C$-$^{13}C$ COSY model increased from 0.93 to 0.94 and $q^2$ increased from 0.88 to 0.89. The treated $^{13}C$-$^{13}C$ distance connectivity data produced a model where the $r^2$ increased from 0.89 to 0.91 and $q^2$ from 0.79 to 0.81. Using both the treated $^{13}C$-$^{13}C$ COSY and $^{13}C$-$^{13}C$ distance connectivity PCs for the type-four models, the $r^2$ decreased from 0.96 to 0.95 and $q^2$ from 0.92 to 0.90. In contrast, the adjusted spectra, when $^{13}C$-$^{13}C$ COSY and $^{13}C$-$^{13}C$ distance connectivity raw data were combined before the extraction of PCs, caused the $r^2$ to increase from 0.92 to 0.93 and $q^2$ increased from 0.81 to 0.84. The type-four (PC) approach to spectral combination gave superior results, but the effect of retaining information from the singly populated bins improved type-three (3D) results.

It is interesting to note that a strong predictive model was produced by the PD-ANN using the combined COSY and distance $^{13}C$ spectral data. There was no preprocessing (such as principal components or other types of regression analysis to extract features or bin population adjustments) performed on the data prior to its use. The model used the raw spectral information. The strength of the model is clearly demonstrated with the average $q_{10}^2$ of 0.73. The total data set consisted of only 30 compounds. Thus each time 10 compounds were left out, the data set use to build the model was depleted by 33 percent.

***3D CoSCoSA Model of Binding to the Aromatase Enzyme*** The assigned $^{13}C$ NMR spectra of 50 steroids shown in Figure 7 and Table 2 were used to develop four types of CoSCoSA models of binding to the aromatase enzyme [42]. These four model types paralleled the four CoSCoSA models of the corticosteroid binding globulin: one model using nearest neighbor 2D COSY connectivities, a second model using 2D long-range distance connectivities only, a third model using COSY and long-distance spectra combined using the combined PCs derived by independent PC analysis of the two distance ranges, and the fourth model (3D) using the combined raw spectra as one might expect for a direct 3D representation of each molecule. Methods of statistical analyses also paralleled the earlier work. In this case, each of the four types of CoSCoSA model was built using the five most correlated PCs and, in a variant procedure, rebuilt for comparison using the number of most correlated PCs that produced a maximum result for the $F$ value.

Table 6 contains a comparison of the model performance parameters ($n$, $r^2$, $q^2$, and number of components) used for QSAR [7], 1D CoSA [23], 2D CoSASA [23], and CoSCoSA models [42]. All four CoSCoSA models with PCs have a strong correlation ($r^2$) and cross-validated variance ($q^2$) and are favorable when compared to the previously published models of binding to

**TABLE 6 Aromatase Model Performance Parameters n, $r^2$, and $q^2$ [7, 23, 41]**

| Model | Number of PCs | $r^2$ | $q^2$ |
|---|---|---|---|
| CoMFA | 5 | 0.94 | 0.72 |
| 1D CoSA | 5 | 0.78 | 0.71 |
| 1D CoSA | 5 bins | 0.82 | 0.77 |
| 2D CoSASA | 5 | 0.75 | 0.67 |
| 2D CoSASA | 5 atoms | 0.74 | 0.66 |
| 2D CoSCoSA COSY | 5 | 0.77 | 0.68 |
| 2D CoSCoSA COSY | 9 | 0.89 | 0.89 |
| 2D 6-9 Å Distance | 5 | 0.65 | 0.65 |
| 2D CoSCoSA 6-9 Å Distance | 7 | 0.72 | 0.72 |
| 2D CoSCoSA COSY + Distance | 5 | 0.77 | 0.68 |
| 2D CoSCoSA COSY + Distance | 10 | 0.92 | 0.86 |
| 3D CoSCoSA | 5 | 0.77 | 0.77 |
| 3D CoSCoSA | 8 | 0.87 | 0.83 |

the aromatase enzyme. To further validate these results, the statistical process was tested by randomizing the binding activity data to see whether the CoSCosCA models could develop an invalid association based on meaningless spectral data–aromatase binding correlations. As hoped, excellent statistical correlation occured using actual binding data while the random data produced models having very low to zero correlations.

Figure 13*a* is a plot of the predicted binding versus experimental binding for the 2.0-ppm resolution CoSCoSA model based on $^{13}C$-$^{13}C$ COSY data and 9 PCs. The COSY model was the best model, and using 9 PCs it had an $r^2$ of 0.89 and a $q^2$ of 0.89, which indicates self-consistency and excellent predictive capability. Figure 13*b* is a plot of the predicted binding versus experimental binding for the 2.0-ppm resolution CoSCoSA model based on long-distance spectral data. The $r^2$ of this model was 0.72 and the $q^2$ was 0.72. Figure 13*c* is a plot of the predicted versus experimental binding for the 2.0 ppm resolution (PC) CoSCoSA model based on the combined $^{13}C$-$^{13}C$ COSY and $^{13}C$-$^{13}C$ distance connectivity PCs. The model used 10 PCs and had an $r^2$ of 0.92 and a $q^2$ of 0.86. Figure 13*d* is a plot of the predicted binding versus experimental binding for the 2.0-ppm resolution (3D) CoSCoSA model for $^{13}C$-$^{13}C$ COSY and $^{13}C$-$^{13}C$ distance connectivity spectral data when the two were combined before principal component extraction. The 3D model was based on 8 PCs and had an $r^2$ of 0.87 and a LOO cross-validated $q^2$ of 0.83 [42]. As was found for corticosteroid binding, the PC method of representing distance/chemical shift data associations produces better CoSCoSA models than the 3D approach.

All four CoSCoSA models of aromatase enzyme binding have a $q^2$ greater than the 0.72 seen for the QSAR model [7]. We believe it is valid to compare our models based on more than five PCs to QSAR models based on five components. The CoSCoSA models use spectral data in a "digital" or quantized

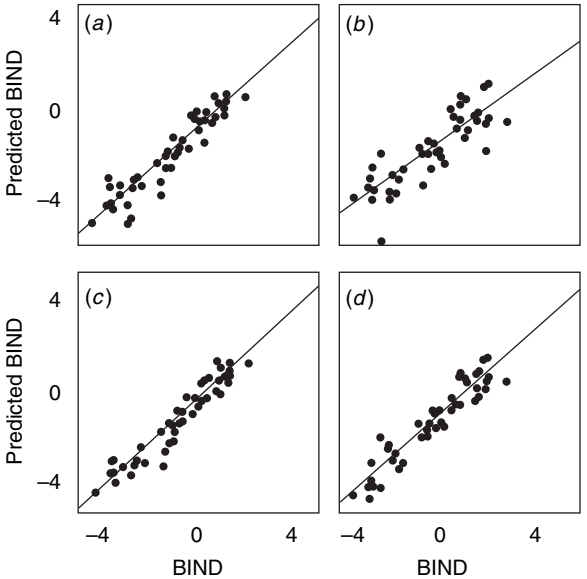


**Figure 13** Plot of the predicted binding versus experimental binding for the aromatase enzyme. (*a*) The $^{13}C$-$^{13}C$ COSY data, (*b*) all the $^{13}C$-$^{13}C$ distance connectivity data greater than 6.9 Å, (*c*) the combined $^{13}C$-$^{13}C$ COSY and $^{13}C$-$^{13}C$ distance connectivity PCs, (*d*) the combined $^{13}C$-$^{13}C$ COSY and $^{13}C$-$^{13}C$ distance connectivity data before PCs are extracted from the combined data.

fashion, whereas QSAR models are based on "analog" or continuum format data. Our CoSCoSA models and QSAR models are based on similar physical phenomena. Both model types have access to electrostatic molecular features. The dynamic range of data used to build the modeling association differs significantly between these two approaches. However, under cross-validation it becomes clear that the CoSCoSA models have a better signal to noise (predictability) when more components are used. The use of an optimal rather than identical number of PCs for comparisons compensates for the data structure difference.

Another significant advantage of QSDAR compared to QSAR models is that the former require fewer assumptions to produce a model. During the calculation of electrostatic potentials in 3D QSAR modeling, assumptions are made for solvent effects, partial charges, and dielectrics. Further, only one structural conformation, typically the gas-phase minimum energy conformation, is assumed to represent each molecule to establish its tendencies with respect to the structure–activity relationship. Each of the QSAR assumptions and approximations are prone to produce significant error. Three-dimensional QSDAR modeling does not rely on those assumptions. $^{13}C$ NMR spectral data inherently reflect a Boltzmann's distribution of structural conformations to produce a "quantum mechanical energy" that represents the average struc-

tural local environment for every carbon atom in the molecule. This average quantum mechanical environment is obtained for a molecule in solution and not derived from gas-phase electrostatic interactions. These facts help explain the superior results of 3D QSDAR models in comparison to 3D QSAR.

Our simpler CoSA aromatase binding model started with 256 spectral bins, which were reduced to 87 when all the columns with only zeroes or with only one nonzero entry were removed. Our 2.0-ppm CoSCoSA model started with 6441 bins (3D overall, 2D in space). When all nonpopulated columns were removed, these were reduced in number to 280 for the $^{13}C$-$^{13}C$ COSY data and 397 for the $^{13}C$-$^{13}C$ distance connectivity data. These CoSCoSA models, for the particular training set and a 2.0-ppm resolution spectral bin width, used less than 6 percent of the available 2D connectivity spectral space.

The 2.0-ppm bins in QSDAR models tend to override errors from using predicted $^{13}C$ NMR data. Using bins considerably wider than typical instrumental resolution changes the information content of raw spectra and, by reducing the number of spectral bins, also reduces the computational intensity of modeling. Both results are desirable for modeling purposes. Instrument-dependent elements (shape, resolution) of the chemical shift peaks do not add artificial, irrelevant variability. At 2.0-ppm resolution, the CoSCoSA models retained enough spectral and structural information to represent the important activity associations. Also, wide spectral bins increase the number that are multiply populated, and so produce models of binding to the aromatase enzyme that have excellent predictive qualities under cross-validation.

***3D CoSCoSA Models of Aryl Hydrocarbon Receptor Binding*** Previously reported log $EC_{50}$ (median effective concentration) binding activity data [24, 54–60] were used to develop models of PCDD, PCDF, and PCB compounds. $^{13}C$ NMR spectra were simulated as before using ACD/Labs CNMR predictor software (ACD, Toronto, Canada, *www.acdlabs.com*) and used for QSDAR both CoSASA and CoSCoSA modeling. For these compound types, no chemical shift peaks were present outside the 107 to 159 ppm range. The use of predicted rather than experimentally measured NMR chemical shifts is not necessary to build the CoSCoSA and CoSASA models, but it saves time, money, and in this case prevents possible toxic exposures.

CoSCoSA and CoSASA models were produced by using the assigned $^{13}C$ NMR chemical shifts at the 12-carbon positions in the PCDF, PCDD, and PCB molecules, as shown in Figure 8. CoSASA models for PCDD and PCDF compounds were built directly from the 12 assigned carbon chemical shifts in each predicted spectrum, as previously described. No binning was involved for CoSASA models.

For 3D CoSCoSA, chemical shifts and atom assignments from the predicted spectra were used to form 2D through-bond COSY spectra and 2D through-space medium-range and long-range spectra. We reduced the resolution of all 2D spectra by populating 2.0-ppm bins. These dimensions were chosen, as before, to multiply populate bins for statistical analysis and to reduce

the effects of uncertainties from the use of simulated spectra. The 2.0-ppm bin spectral width was used successfully for AhR binding CoSA models [24].

CoSCoSA models were built from 2D bins in the plane of nearest neighbor through-bond (COSY) plane, a medium-range through-space distance plane, and a long-range through-space distance plane. The arrows in Figures 14*a*, 14*b*, and 14*c* show the through-bond nearest neighbor carbon-to-carbon connections of PCDF, PCDD, and PCB molecules. The arrows in Figures 14*d* and 14*e* show a selection of medium-range through-space connections, in this case from the 2 and 3 or 7 and 8 anchoring positions to the middle ring carbons in PCDF and PCDD molecules. Likewise for PCB molecules, in Figure 14*f*, we defined anchoring points as the 3, 4, and 5 positions and the 3′, 4′, and 5′ positions, so the "middle-ring" atoms consisted in this case of the two ring-connecting carbons only. The arrows in Figures 14*g* and 14*h* show the long-range through-space connections from the 2 and 3 or 7 and 8 anchoring positions to the opposite ring carbons in PCDF and PCDD molecules. Likewise for PCB molecules, in Figure 14*i*, the long-range connection from the 3, 4, and 5 positions and 3′, 4′, and 5′ positions to the carbons on the opposite ring. Since this long-range connectivity interaction overlapped the

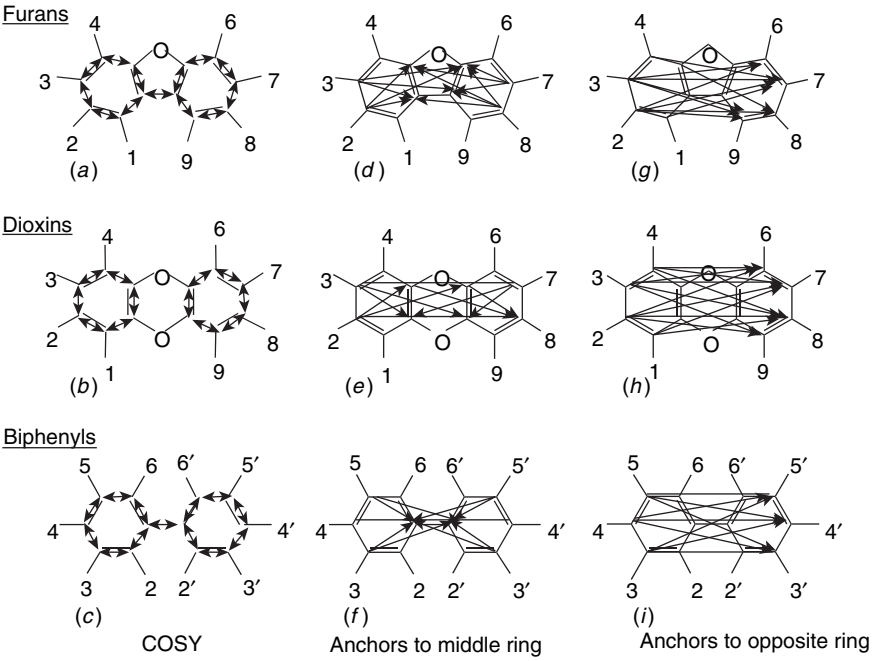


**Figure 14** Arrows represent the 2D $^{13}C$-$^{13}C$ COSY spectra for (*a*) PCDFs, (*b*) PCDDs, and (*c*) PCBs. The arrows represent middle-range 2D $^{13}C$-$^{13}C$ distance spectra for (*d*) PCDFs, (*e*) PCDDs, and (*f*) PCBs. The arrows represent long-range 2D $^{13}C$-$^{13}C$ distance spectra for (*g*) PCDFs, (*h*) PCDDs, and (*i*) PCBs.

two anchoring points for each molecule, we choose only one anchor ring as the "origin" for which all long-range through-space connections originated. CoSCoSA models were built using the nearest neighbor 2D spectral plane only, the two anchoring structural elements through-distance 2D planes, and a combination of through-bond and through-space information. Since there were two anchors for every molecule, we might theoretically have separated the medium-range through-space distance connections from the outer ring anchors to middle of the compound 14 (Figs. 14*d*, 14*e*, and 14*f*) into two separate 2D planes. We did not do so because the training set was small and increasing the number of degrees of freedom with additional spectral dimensions would risk producing an overfitted model. In the 2D COSY, symmetry inherent in the nearest neighbor relationship produces a pair of arrows, only one of which is needed to represent each relationship. Therefore, only half the 2D COSY spectral connections were used. In contrast, the medium-range and long-range through-space distance spectra, defined using anchoring origins do not have the same symmetry and duplication. Therefore, the whole 2D spectral plane is needed and was used in model development.

Statistical analysis was performed by Statistica software (Statsoft, Tulsa, OK). For each CoSCoSA model, we used forward multiple linear regression (MLR) on the bins. We did not include in the models any bins with fewer than 2 hits. For the 26 PCDFs, forward MLR selection of 6 bins from COSY, through-space distance and combined COSY and through-space distance ranges were used to produce the corresponding CoSCoSA models. Forward MLR selection of 3 bins was used for COSY, through-space distance and combined CoSCoSA models of the 14 PCDDs and the 12 PCBs. The CoSCoSA model of all 52 compounds was built by forward MLR selection of 10 bins from the combined COSY and through-space distance-generated 2D spectral planes.

Figures 15*a* to 15*c* show plots of the predicted versus experimental binding for CoSCoSA models based on MLR of through-bond COSY spectral data for PCDFs, PCDDs, and PCBs, respectively. Figures 15*d* to 15*f* show graphs of the predicted versus experimental binding for CoSCoSA models based on MLR of through-space medium-range and long-range spectral data for PCDFs, PCDDs, and PCBs, respectively. Figures 15*g* to 15*i* show the predicted versus experimental binding for CoSCoSA models based on MLR of COSY combined with through-space medium-range and long-range spectral data for PCDFs, PCDDs, and PCBs, respectively. Tables 7 to 9 contain CoSCoSA parameters $n$, $r^2$, $q^2$, $q_4^2$, $F$, and 2D bins for the PCDF, PCDD, and PCB models, respectively [39].

Population of the COSY bin 153-113 was associated with 10 of the 11 PCDFs that had activities weaker than −7.0. Population of COSY bin 155-119 was associated with 12 of 16 compounds that had binding activities stronger than −7.0. Bins associated with strong or weak binding, once they are identified by COSY and other CoSCoSA methods, can be related to the substructure features that gave rise to the chemical shifts. The recognition of important

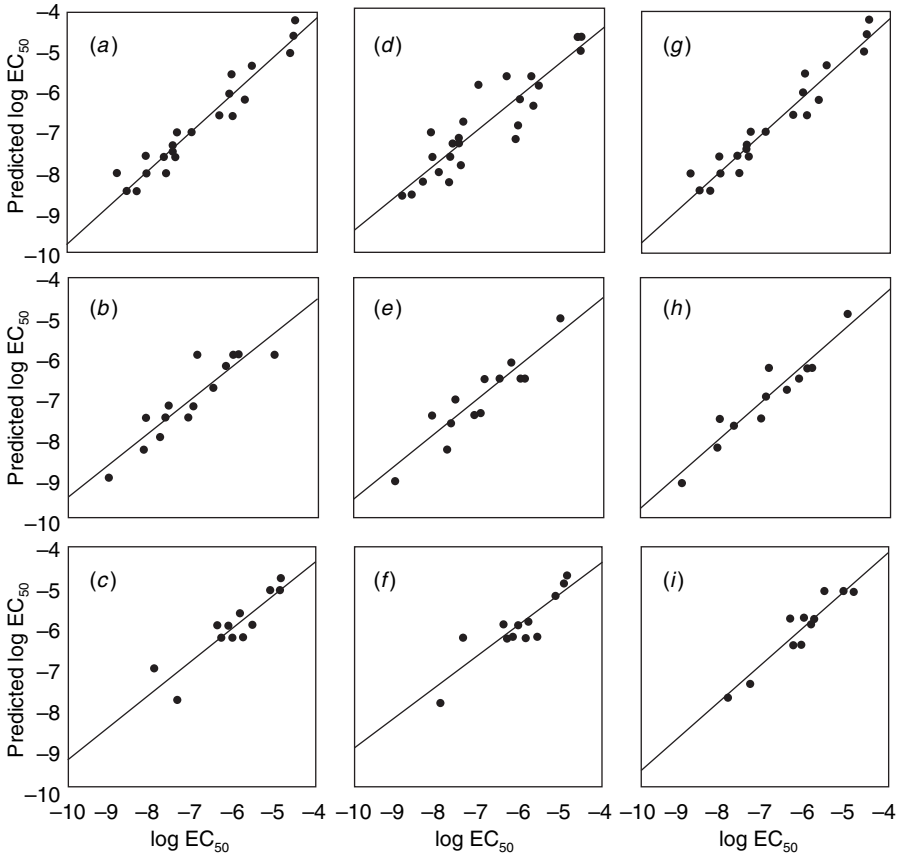


**Figure 15** Graphs of the CoSCoSA models based on 2D $^{13}C$-$^{13}C$ COSY spectra for (*a*) PCDFs, (*b*) PCDDs, and (*c*) PCBs. The graphs of the CoSCoSA models based on 2D medium-range and long-range $^{13}C$-$^{13}C$ distance spectra for (*d*) PCDFs, (*e*) PCDDs, and (*f*) PCBs. The graphs of the CoSCoSA models based on 2D COSY and distance spectra for (*g*) PCDFs, (*h*) PCDDs, and (*i*) PCBs.

substructure features allows the inference of toxic features and other activity mechanisms. In this case, COSY bin 155-119 was populated when the ester oxygen of the furan was connected to the carbon with a chemical shift of 155 ppm. For PCDDs, the COSY bin at 143-123 was populated for all 8 compounds having binding activities stronger than –6.96. In this case, COSY bin 143-123 was populated when the ester oxygen of the dioxin group was connected to the carbon with a chemical shift of 143 ppm. For PCBs, the COSY bin at 137-125 was populated for both compounds having a binding activity stronger than –7.0. For PCDFs, the long-range bin 121-117 marked 4 of the 6 compounds that had activities weaker than –6.0. In the PCDD distance model, bin 127-143 was associated with all 8 compounds having binding activities stronger than –6.96.

The best COSY and through-space model for PCDFs turned out to be exactly the same as the COSY model for PCDF compounds. That is, for the PCDFs, the short-range connections all had better predicative strength than any of the longer distance associations. For PCDDs, the medium through-space bin 127-143 correctly identified all eight PCDD compounds with binding activities stronger than –6.96, as it had done in the less complicated spectral format. The COSY bin at 137-125 correctly identified both PCBs with a binding activity stronger than –7.0.

Figure 16 shows the predicted versus experimental binding for all 52 compounds, of all three structural types, using the combined COSY, medium- and long-range distance spectra. Statistical data for the combined CoSCoSA model are shown in Table 10, along with comparable data based on other methods. The $r^2$ for the PCDFs is 0.87, for PCDDs is 0.84, and for PCBs is 0.75. The CoSCoSA model is derived from ten 2D bins: 3 from COSY spectra, 4 from medium-range spectra, and 3 from long-range spectra. It is significant that each of these ten 2D bins represents peaks of only one compound type. One of the COSY and one of the medium-range bins have hits only for PCDDs. Another pair (a COSY and a medium-range bin) are occupied only for PCBs. The remaining 6 bins have hits only for PCDFs. From the combined COSY with through-space results, we conclude that the CoSCoSA models for PCDF, PCDD, and PCB compounds each contained relevant information sufficient to produce accurate models of binding affinity to the AhR. We also infer that building effective CoSCoSA models for diverse, multistructure-type data sets will require sufficient bins to represent the binding relationships for each of the structural types with a minimum of 2 bins each.

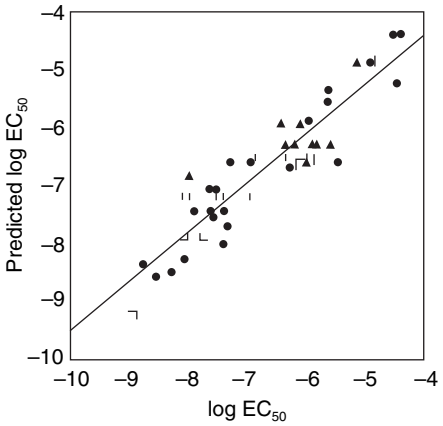


**Figure 16** Plot of the predicted binding versus experimental binding for all 52 compounds using 2D COSY plus medium and long-range spectra. The PCDF compounds are represented by filled circles (●), PCDD compounds shown by open squares (□), and PCB compounds shown with filled triangles (▲).

Tables 7 to 9 address the question of how these CoSCoSA results compare to those from other QSDAR and QSAR modeling approaches. In Table 7 for PCDFs, two of the three CoSCoSA models had higher $r^2$ and $q^2$ values than the corresponding 1D CoSA model. All three CoSCoSA models for PCDFs had higher $r^2$ and $q^2$ than the 2D CoSASA model that associated spectral chemical shifts with structurally assigned locations. Our previous 1-ppm resolution CoSA model based on 5 bins for 26 PCDF compounds had an $r^2$ of 0.82 and $q^2$ of 0.72 [24]. A structural parameter model by Mekenyan and co-workers [61] to produce a 5-component model for 25 PCDF compounds (all 26 PCDF compounds except for 237-trichlorodibenzofuran) that had an $r^2$ of 0.85 and $q^2$ of 0.71. The best model for 39 dibenzofurans proposed by Turner used three infrared EVA components [62]. Since the CoSCoSA models were typically giving $r^2$ above 0.9 and $q^2$ above 0.8, we found a consistent and significant improvement, especially for the prediction accuracy, compared to previous art. In Table 8 for PCDDs, CoSCoSA results were superior to alternative methods in every case but one discussed in MeKenyan [61], which had approximately equivalent quality. In Table 9 for PCBs, the corresponding CoSCoSA models typically exceeded but at

**TABLE 7 26 PCDF Compound Model Performance Parameters n, $r^2$, $q^2$, $q_4^2$, F, and 2D bins[a] [24, 38, 60, 61]**

| Model | $n$ | $r^2$ | $q^2$ | $q_4^2$ | $F$ | 2D Bins |
|---|---|---|---|---|---|---|
| Structural QSAR | 5 | 0.85 | 0.72 | | | |
| CoMFA | 6 PC | 0.85 | 0.72 | | | |
| EVA | 3 PC | 0.96 | 0.73 | | | |
| 1D CoSA | 5 bins | 0.82 | 0.72 | | 18.6 | |
| 2D CoSASA | 6 atoms | 0.74 | 0.70 | | 9.1 | |
| CoSCoSA COSY | 6 bins | 0.92 | 0.84 | 0.84 | 38.7 | C 119-113, C 125-113, C153-113, C 127-119, C 155-119, C 127-125 |
| CoSCoSA Mid + Long | 6 bins | 0.83 | 0.63 | 0.65 | 15.1 | M 127-115, M 125-117, M 119-125, M 125-125, M 113-127, L 121-127 |
| CoSCoSA COSY + Mid + Long | 6 bins | 0.92 | 0.84 | 0.84 | 38.7 | C 119-113, C 125-113, C153-113, C 127-119, C 155-119, C 127-125 |

[a] C stands for COSY, M stands for medium-range, and L stands for long-range spectra.

**TABLE 8 14 PCDD Compound Model Performance Parameters Bin n, $r^2$, $q^2$, $q_2^2$, F, and 2D bins[a] [24, 38, 60, 61]**

| Model | $n$ | $r^2$ | $q^2$ | $q_2^2$ | $F$ | 2D Bins |
|---|---|---|---|---|---|---|
| Structural QSAR | 5 | 0.95 | 0.82 | | | |
| COMFA | 2 PC | 0.88 | 0.73 | | | |
| EVA | 2 PC | 0.85 | 0.48 | | | |
| 1D CoSA | 3 bins | 0.83 | 0.74 | | 16.5 | |
| 2D CoSASA | 5 atoms | 0.81 | 0.53 | | 6.7 | |
| CoSCoSA COSY | 3 bins | 0.83 | 0.75 | 0.74 | 15.9 | C 127-123, C 141-123, C 143-123 |
| CoSCoSA Mid + Long | 3 bins | 0.83 | 0.75 | 0.71 | 16.3 | M 123-141, M 125-141, M 127-141 |
| CoSCoSA COSY + Mid + Long | 3 bins | 0.90 | 0.79 | 0.79 | 16.2 | C 141-123, C 143-123, M 123-141 |

[a] C stands for COSY, M stands for medium-range, and L stands for long-range spectra.

**TABLE 9 12 PCB Compound Model Performance Parameters n, $r^2$, $q^2$, $q_2^2$, F, and 2D Bins[a] [24, 38, 60, 61]**

| Model | $n$ | $r^2$ | $q^2$ | $q_2^2$ | $F$ | 2D Bins |
|---|---|---|---|---|---|---|
| COMFA | 3 PC | 0.87 | 0.49 | | | |
| EVA | 1 PC | 0.72 | 0.16 | | | |
| 1D CoSA | 3 bins | 0.66 | 0.30 | | 5.2 | |
| CoSCoSA COSY | 3 bins | 0.82 | 0.58 | 0.58 | 12.2 | C 135-125, C 127-127, C 133-131 |
| CoSCoSA Mid + Long | 3 bins | 0.77 | 0.66 | 0.47 | 9.1 | M 133-135, L 133-131, L 133-133 |
| CoSCoSA COSY + Mid + Long | 3 bins | 0.91 | 0.80 | 0.80 | 26.3 | C 137-125, M 133-137, M 131-139 |

[a] C stands for COSY, M stands for medium-range, and L stands for long-range spectra.

least equaled all previous methods, particularly for predictions from cross-validation experiments.

Table 10 shows the results for all 52 PCDF, PCDD, and PCB compounds. The combined COSY and through-space CoSCoSA models had a higher $q^2$ than the corresponding 1D CoSA model [39]. Additionally the CoSCoSA model was based on only ten 2D bins, whereas the CoSA model was based on fifteen 1D bins. The previous CoSA model for all 52 compounds had an $r^2$ of 0.87 and a $q^2$ of 0.67 [24], whereas the current CoSCoSA model had an $r^2$ of 0.85 and $q^2$ of 0.73 and $q_4{}^2$ of 0.52. When two outliers are removed from the cross-validation of the 52-compound CoSCoSA model, the $q^2$ and $q_4{}^2$ improve to 0.77. Both outliers occur when a compound has all zeros in every bin except

**TABLE 10 All 52 PCDF, PCDD, and PCB Compound Model Performance Parameters n (parameters used), $r^2$, $q^2$, $q_4^2$, F, and 2D bins[a] [24, 38, 60, 61]**

| Model | $n$ | $r^2$ | $q^2$ | $q_4^2$ | F | 2D Bins |
|---|---|---|---|---|---|---|
| Structural QSAR | 5 | 0.85 | 0.72 | | | |
| CoMFA | 6 PC | 0.88 | 0.71 | | | |
| EVA | 3 PC | 0.87 | 0.68 | | | |
| 1D CoSA | 15 Bins | 0.87 | 0.67 | | 16.6 | |
| CoSCoSA COSY + Mid + Long | 10 Bins | 0.85 | 0.73 | 0.52 | 24.0 | C 121-117, C 141-123, C 137-125, M 119-117, M 129-119, M 133-135, M 123-141, L 119-123, L 119-125, L 121-127 |

[a] C stands for COSY, M stands for medium-range, and L stands for long-range spectra.

for one column and that column has only two bin hits in it. When a column with only one remaining hit in it is used during the LOO or L4O cross-validation process, the linear regression β coefficient can change sign. One needs to be aware of this fact during CoSCoSa modeling. The CoSCoSA composite model for the 52 compounds represents a significant improvement over corresponding results from alternative modeling approaches [24, 61, 62].

## Producing a $^{13}$C and $^{15}$N Heteronuclear Connectivity Matrix

Often, particularly for chemicals with potentially useful pharmaceutical value or toxicity, compound types to be modeled will contain atoms besides carbon, oxygen, and hydrogen. It that case, NMR spectra besides $^{13}$C can be used. The most prominent atom that is both biologically important and for which accurate NMR prediction software is available is $^{15}$N. Other NMR nuclei that could be used for SDAR or QSDAR modeling are $^{17}$O, $^{19}$F, and $^{31}$P. Which of these might be useful would depend on the biological endpoint to model and the availability of an endpoint-characterized training set. Figure 17 shows a typical two-dimensional spectral connectivity layout for a $^{13}$C-$^{15}$N heteronuclear CoSCoSA model. As before for multidimensional $^{13}$C-$^{13}$C models, the symmetry-based data duplicates mean that only half of the spectra in the array are necessary to develop a model.

***$^{13}$C-$^{15}$N Heteronuclear CoSCoSA Model of Antibiotic Cephalosporins***

Figure 18 shows the structures of 17 cephalosporins, widely used antibiotics similar in structure and mode of action to penicillin. There are two nitrogen atoms in the backbone of cephalosporin molecules. These structures were used

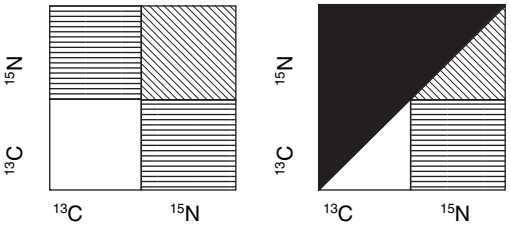


**Figure 17** 2D $^{13}C$-$^{15}N$ heteronuclear connectivity matrix.

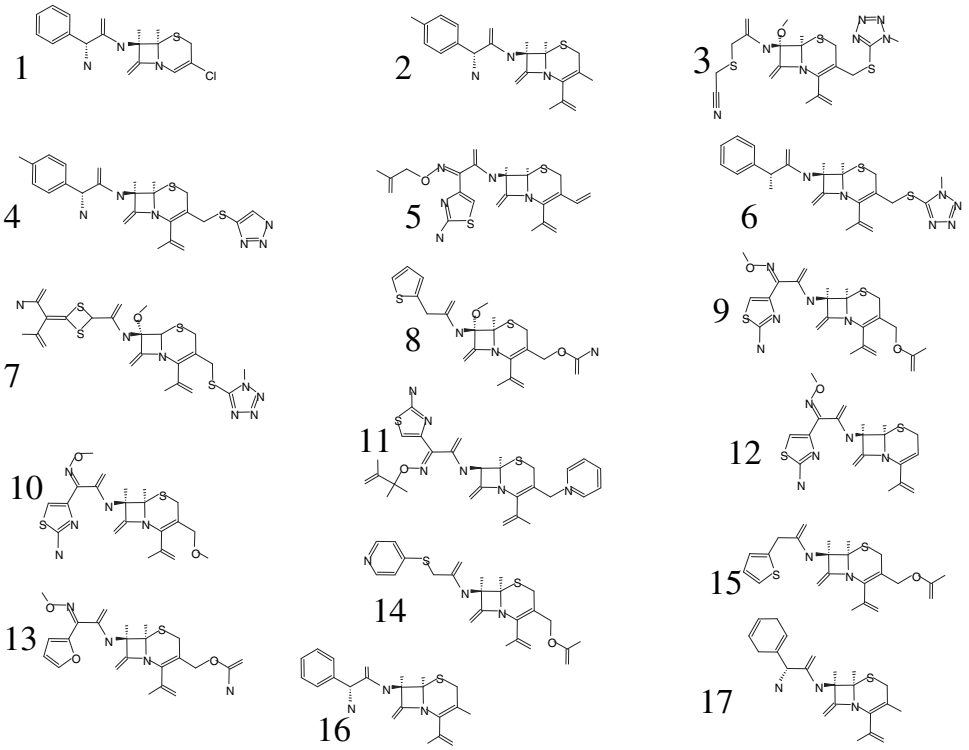


**Figure 18** Celphasporin structures. Numbering of compounds corresponds to Table 11.

as an example to examine the potential of heteronuclear CoSCoSA methods to model the minimum inhibitory concentrations (MIC) of cephalosporin antibiotics, using only the through-bond (COSY-type) carbon-to-carbon and through-bond carbon-to-nitrogen 2D planes. In producing this 2D CoSCoSA model, for technical reasons we defined the endpoint as log(1/MIC) [63].

We used bin sizes of 3.0 × 3.0 ppm for carbon-to-carbon COSY connections and 10.0 × 3.0 ppm for the nitrogen-to-carbon direct connections. Nitrogen chemical shifts were predicted from software available on the ACD/Labs ILAB website (ACD, Toronto, Canada, *www.acdlabs.com*). Carbon shifts were predicted as before. In building the nitrogen-to-carbon connectivity matrix, 700.0 ppm was added to the predicted nitrogen chemical shifts, so that they fell in the range from 300 to 700 ppm. Thus, a single synthetic spectrum could be defined with carbon-to-carbon connectivity bins occupied from 0 to 240 ppm, and nitrogen-to-carbon from 300 to 700 ppm.

Forward regression MLR selected 4 bins from a total of 101 carbon-to-carbon and 48 carbon-to-nitrogen occupied bins. These were (–230 nitrogen, 156 carbon), (–280 nitrogen, 162 carbon), (–230 nitrogen, 168 carbon), and (135 carbon, 24 carbon). The CoSCoSA models were produced as described in previous examples, and the results are shown in Table 11 and Figure 19. Figure 19 shows the results of the cephalosporin CoSCoSA model with a correlation $r^2 = 0.92$, $F$ value = 36.2, $P < .000005$, LOO of $q_1^2 = 0.88$, average L4O of

**TABLE 11 Results of Cephalosporin CoSCoSA Modeling**[a]

| Compounds | Number in Figure 18 | MIC | log(1/MIC) | Leave-1-Out log(1/MIC) Predicted | Leave-4-Out log(1/MIC) Predicted |
|---|---|---|---|---|---|
| Cefaclor | 1 | 8 | –0.90 | –1.22 | –1.23 |
| Cefadroxil | 2 | 8 | –0.90 | –0.90 | –0.90 |
| Cefmetazole | 3 | 16 | –1.20 | –1.15 | –1.15 |
| Cefaperazone | 4 | 16 | –1.20 | –1.15 | –1.13 |
| Cefixime | 5 | 8 | –0.90 | –0.88 | –1.20 |
| Cefamandole | 6 | 16 | –1.20 | –1.15 | –1.15 |
| Cefotetan | 7 | 16 | –1.20 | –1.15 | –1.13 |
| Cefoxitin | 8 | 16 | –1.20 | –1.15 | –1.23 |
| Cefotaxime | 9 | 16 | –1.20 | –1.21 | –1.23 |
| Cefpodoxime | 10 | 2 | –0.30 | –0.32 | –0.32 |
| Ceftazidime | 11 | 16 | –1.20 | –1.21 | –1.23 |
| Ceftizoxime | 12 | 8 | –0.90 | –0.85 | –0.83 |
| Cefuroxime | 13 | 4 | –0.60 | –0.63 | –0.42 |
| Cephapirin | 14 | 8 | –0.90 | –0.90 | –0.90 |
| Cephalothin | 15 | 8 | –0.90 | –0.90 | –0.90 |
| Cephalexin | 16 | 8 | –0.90 | –0.90 | –0.90 |
| Cephradine | 17 | 8 | –0.90 | –0.90 | –0.90 |

[a] MIC = minimal inhibitory concentration [62].

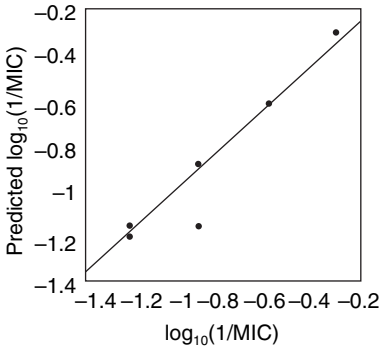


**Figure 19** 2D $^{13}C$-$^{15}N$ heteronuclear CoSCoSA model of celphasporin antibacterial activity.

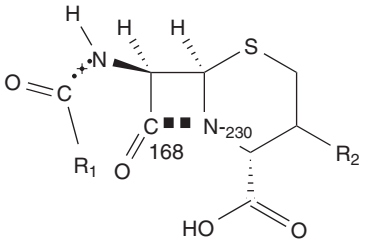


**Figure 20** Structural explanation of bins selected in the 2D $^{13}C$-$^{15}N$ heteronuclear CoSCoSA model of celphasporin antibacterial activity. Dashed bond between carbon 168 and nitrogen (–230) corresponds to bond that undergoes acid hydrolysis and then binds to enzyme transpeptidase and irreversibly inactivates it. Dotted bond between carbon and nitrogen corresponds to bins reflecting carbon (156), nitrogen (–270) and carbon (162), nitrogen (–280) selected bins and may explain the backbone torsional angle dependence of dipeptides binding affinity to transpeptidase.

$q_4^2$ = 0.79, and standard deviation = 0.03. Figure 20 shows the structurally assigned interpretation of the chemical shifts used to formulate the CoSCoSA model of cephalosporins. The chemical shifts at (–230 nitrogen,168 carbon) identifies the carbon-to-nitrogen bond where acid hydrolysis occurs and allows the celphalosporin to bind to bacterial enzyme transpeptidase, irreversibly inactivating the enzyme and stopping growth [64]. The chemical shifts at (–230 nitrogen,156 carbon) and (–280 nitrogen,162 carbon) identify another carbon-to-nitrogen bond in the middle of each celphalosporin compound. Since the enzyme transpeptidase binds to dipeptides, it is reasonable to assume that the other carbon-to-nitrogen bond [which looks like part of an amino acid backbone (carbonyl-to-amide-to-alpha carbon)] is involved via a hydrogen bond to the enzyme. The two different chemical shifts for the same carbon-to-nitrogen bond can represent the bond in two different configurations, two

different electrostatic potentials, or more likely a combination of different configurations and associated electrostatic potentials. It has been shown that peptides interact with the penicillin binding transpeptidase proteins with binding energy dependent on the backbone torsion angles [65]. We mention this because most descriptions of cephalosporin activity mechanisms do not discuss the critical role of the second carbon-to-nitrogen bond. The CoSCoSA model including heteronuclear NMR peaks predicted that both carbon-to-nitrogen bonds contribute significantly to the antibiotic strength of cephalosporin. We tried modeling cephalosporin activity using $^{13}$C COSY data alone, but the models were very poor with respect to all statistical measures of goodness of fit and leave-one-out cross-validation (data not shown). In this case, both $^{15}$N and $^{13}$C chemical shifts were very important in producing a reliable model of biological activity.

***Using Molecular Dynamics of Compounds to Produce a 4D Connectivity Matrix*** A 4D-connectivity matrix can be defined as the sum of an arbitrary number, say 100, 3D connectivity matrices. In the simplified version of this 4D connectivity matrix concept, chemical shifts of atom *i* and atom *j* are assumed not to change as the molecule bends or twists with time, but the distance between atoms *i* and atom *j* would differ as a function of the molecular conformation. For any two atoms, molecular dynamics programs can estimate interatomic distance, a value that may change over some range and for which the percentage of time that the distance is within a certain bin will vary, depending on molecular connections, degrees of freedom, and the like. This concept applied to CoCoSA modeling would treat the distance between atoms as a potential variable rather than as a constant. A score of 100 in a 4D connectivity matrix will represent unvarying distances between two atoms as seen in bonds and also between more distant atoms if the molecules are very rigid. For two atoms in flexible molecules there will be a distribution of distance hits along the *z* axis varying from 1 to some maximum (most probable conformation). For all the possible atom pairs, distance distributions will be Gaussian, or skewed-Gaussian, functions when there is a single maximum distance. When there is more than one maximum, more complex distribution will be seen. In the simplified regime of CoSCoSA distance modeling, the occupancy pattern for a particular molecule and a particular through-space distance bin would be 100 if the full range of possible distances lay within that 2D bin. If 30 percent of the time, the molecular shape were such that the atomic distances lay outside that bin, the bin occupancy would be represented with 70, and the remaining 2D bin(s) would share the remaining 30 points. We shall now discuss why this way of representing molecular conformation characteristics confers a significant advantage when building a CoSCoSA model, particularly for cases in which the training set includes very diverse compound types.

Binding of a ligand to a receptor occurs because of a lowering of Gibbs free energy for the combined ligand–receptor system. As in physical and chemical

systems, lowered Gibbs free energy of biological systems depends on two factors: favorable changes in enthalpy and/or entropy changes. Traditional 3D QSAR approaches are biased to reflect enthalpy changes, which ignore entropy contributions. Classical 3D QSAR models are based on electrostatic and stereospecific patterns in space as they correlate with biological activity. These correlations work because they reflect electrostatic differences and the corresponding changes of enthalpy that occur during binding. These QSAR models neglect entropy changes. Entropy calculations are routinely neglected because they are often misunderstood and because it is not easy to conceive a way in which such phenomena can be reflected in models that are built directly using molecular 3D structural conformations. Modelers typically use only the minimum energy conformer for each compound during model development, and this ignores the changes in molecular conformation that are statistically possible and that certainly occur during the compound's interaction with its enzymatic substrate when binding occurs.

***Using the 4D Connectivity Matrix to Estimate Configurational Entropy***
We have used molecular dynamics of 130 structurally diverse compounds from the previous 2D CoSCoSA model of estrogen binding to estimate the configurational entropy of each compound. A 4D connectivity matrix allows for multiple conformations of a molecule to be displayed. The multiple conformations can be put into an "entropy-like" equation to estimate the effect of configurational entropy [66–68]:

$$S \propto 1/N \sum_{i<j} p_{ij} \ln p_{ij}$$

where $p_{ij}$, a probability that must lie between zero and one, is calculated as the percentage of time the distance between atoms $i$ and $j$ lie within the $r_{ij}$ spatial distance bin, with this percentage then divided by 100 to express the occupancy within the probability range required by the equation.

Representation of entropy effects can and should be simplified, and this is possible by at least two methods. One method is to calculate the average $r_{ij}$ distance for each molecule and express the entropy surrogate using each $p_{ij}$ value as its number of standard deviations from the average. This type of data pretreatment is called autoscaling. The second method, analogous to the one used to simplify 3D connectivity matrices, is to break interatomic distance space into a number of 2D planes. For the 130 compound set, we used four 2D planes with distances of $r_{ij} < 2.5$ Å, 2.5 Å $< r_{ij} < 5.0$ Å, 5.0 Å $< r_{ij} < 7.5$ Å, and 7.5 Å $< r_{ij}$. As mentioned above, $p_{ij}$ is defined as the percentage of time the distance between atoms $i$ and $j$ lies within the constraints of the specified 2D plane. A rigid molecule (or through-bond neighbors) will have distances between atoms that fluctuate very little, and $p_{ij}$ for the occupied bin in those cases will equal to 100/100 = 1.0, with all other $ij$ bins unoccupied. The corresponding configurational entropy will be 0. A flexible molecule will have a distribution of $p_{ij}$ across several 2D planes in the 4D matrix. The definitions

above, applied to the flexible molecule, would be expected to yield a higher configuration entropy.

Molecular dynamics simulations were run using Hyperchem 7.0 (Hypercube Inc., Gainsville, FL) for 130 compounds. First, Polak-Ribiere conjugate gradient energy minimization was executed for 1000 iterations or until a minimized structure was obtained. The temperature was slowly raised to 310 K over 12 ps. (310 K was chosen to estimate an average body temperature, the presumed condition in which conformational variation is relevant to receptor binding.) Then, a 100-ps molecular dynamics trajectory was executed at 310 K and a structure was saved every 1 ps. The 100 structures from the trajectory were used to calculate $p_{ij}$ for every nonhydrogen-to-nonhydrogen atomic pair.

Figure 21 shows the entropy of the unbound free compound versus estrogen receptor log(RBA) for the 130 compounds. The plot clearly shows that the more configurational entropy a compound has, the stronger it binds to the estrogen receptor. This relationship appears to violate the principles of physics, at least with respect to the expectations described two paragraphs above. We understand this counterintuitive relationship to derive from several interacting phenomena. First, the compounds compared here were initially selected for a QSAR model, presumably because they could be effectively modeled by electrostatics (enthalpy). Other compounds for which structure–activity data were available were not included in this QSAR test set. Some earlier projects from the same group had included almost 200 estrogens [43]. We, in turn, selected the same test set to facilitate comparisons of our results to the published QSAR. When one realizes the possible nonrandom nature of the test set compounds, the plot starts to make sense. If, based on enthalpy, a molecule has less capability to bind, it cannot have a large entropy loss upon binding. Likewise, a molecule with substantial enthalpic contribution to receptor binding can maintain binding capacity even with an unfavorable entropy

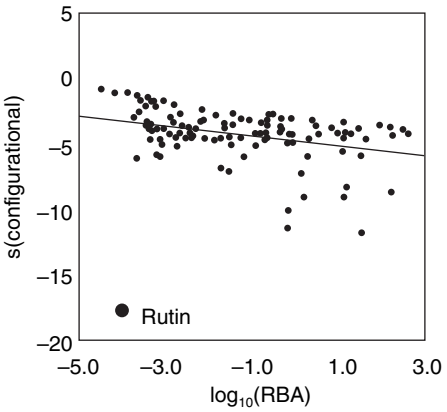


**Figure 21** log(RBA) versus configurational entropy for 130 compounds that bind to the estrogen receptor. Rutin is shown with a larger circle.

term. We believe our apparently contraphysics entropy trend is a second-order consequence of the first-order relative relationships inherent in selection of this set compounds.

In Figure 21, rutin appears as a rather dramatic outlier compared to all other compounds modeled by our 4D QSDAR method. The result might be an inexplicable artifact. All modeling methods have weaknesses and perform better for some compounds than for others. There is a possible explanation for its behavior in the 4D QSDAR model. Experimentally, rutin has very low receptor binding. The rutin molecule is much larger than the others in this 130-member comparison set. Because of its size, it has large conformational entropy in its unbound state and this may be the reason it does not bind strongly to the estrogen receptor. The 3D QSAR would have us believe that the reason rutin does not bind strongly to the estrogen receptor is a result of strong steric hindrance and weak electrostatics. The 2D QSDAR model of estrogen receptor binding for rutin was fairly weak.

When it is bound, rutin's total entropy is probably not decreased very much because the majority of the molecule lies outside the binding site and is still conformationally free. The model attempts to estimate the entropic effect of binding as a function of the free state entropy for each test set molecule. For rutin, the 4D QSDAR model gives a poorer prediction than QSAR, which ignores entropy altogether. For a complete estimation of configurational entropy changes upon binding, one would need to compare the molecular dynamics trajectories of the compounds and the particular enzymatic substrate in both the free and bound states along with solution effects, rather than use only the range of free state conformations of the test set molecules.

The results described here illustrate for the first time how entropic phenomena can be realistically incorporated into QSDAR modeling. We intuit that larger data sets (from which outlier molecules have not been excluded to compensate for methodological limitations) will benefit from representation using entropic parameters. We suppose that, for a training set inclusive of the full range of structural variants, spectral pattern definition using only small molecule structural connectivity and conformational variability can provide significant access to entropy term contributions and improve the quality and utility of biological models. We are also beginning to examine other strategies that would make the 4D QSDAR spectral representation illustrated above less sensitive to the type of error demonstrated for rutin.

## 3 FUTURE IMPROVEMENTS

### Commercial Software Products

ACD/Labs (ACD, Toronto, Canada, *www.acdlabs.com*) is the supplier of the most complete suite of spectroscopy tools allowing both experimental data processing for all general analytical data as well as the prediction of multinuclear 1D and 2D NMR data (H1, C13, N15, F19, P31). ACD/Labs released soft-

ware modules in 2005 that enable both experimental and predicted 1D and 2D spectra to be passed to an add-on module for SDAR and QSDAR modeling. The company also intends to include pattern recognition methods that will correlate the spectra to biological and toxicological activity. Building 1D and 2D SDARs manually is not a complicated process, and the same is true in principle for 3D and 4D variants. However, for technical reasons, the automation of interatomic distance spectral binning, especially when based on molecular dynamic models, will greatly facilitate an otherwise tedious and impractical manual process. It is likely that this capability will be incorporated into a later version of the software.

## Improving the Algorithms for Binning Spectral Data

CoSCoSA modeling is a powerful modeling technique because it uniquely combines quantum mechanical information from the chemical shifts with internal molecular distances. Optimizing the bin size through smart binning techniques, the distance cutoffs, multiple conformations of a structure, multiple spectra, and the number of distance connectivity spectra used in a CoSCoSA model may be needed or helpful for modeling biological, physical, chemical, ADME, or toxicological endpoints.

***Potential Applications*** We have demonstrated various QSDAR models capable of predicting the strength of chemical interactions with biological systems. This kind of prediction is useful for estimating pure chemical toxicity or even risk assessment of environmental contaminants that are complicated mixtures. It is also useful for pharmaceutical lead compound identification and for qualifying new drugs with respect to side effects or unintended toxicity. In that regard, we believe the Food and Drug Administration (FDA) and other drug regulatory authorities could use QSDAR for anticipating problems, to specify the types of toxicity testing required for a new drug application. We have also shown the predictive value of QSDAR methods when the endpoint is essentially chemical (decomposition) but the process is mediated through the enzymatic capabilities of bacteria. This suggests possible QSDAR applicability in relation to bioremediation efforts.

The type of QSDAR method useful for a particular application depends on the amount of data available for the model training set, the variety of compound types to be modeled, the elemental composition of the compounds in the training set, and the predictive accuracy required. One-dimensional and 2D models are easier to set up and less computationally intensive to build and validate than 3D and 4D models. It is possible that some applications, such as drug lead discovery, may benefit from using the simpler methods early in a process and the more elaborate ones after the number of candidates has been reduced and a backbone scaffold has been selected.

The best computational approaches appear to depend less on the particular application and more on the overall size and character of the data set to be examined. When all members of the training set are structurally similar, 3D

QSDAR models can be built from PCs, as was shown for steroids binding to corticosterone binding globulin or aromatase enzyme or from bins, as was shown in cephalosporin antibacterial activity. When the training set molecules are not structurally similar, 3D QSDAR models are best built from selecting bins in the 2D planes of the 3D connectivity matrix by PLS forward regression, as was demonstrated for the 130 compounds interacting with the estrogen receptor. In the latter case, selected bins represented two chemical shifts (energies) separated by a specified distance. The chemical shifts typically signified the atom type in a specific environment. A selected bin in a CoSCoSA model represents two atom types at a specified distance, which are shown by the model to be positively or negatively associated with the biological activity being modeled. The MLR-selected bins can then be used to build a model of biological activity and to understand the mechanism underlying the biological activity. The selected bins can be used as structure-spectral anchors in models for lead identification or lead optimization of new drugs. This approach was shown for toxicity assessment application using PCDD, PCDF, and PCBs binding to the ArH receptor, but there is no reason the same concept cannot be applied to drug discovery.

The CoSCoSA modeling system can be applied to receptor binding systems for which the structure–activity relationship is unknown, a common situation faced by new drug discovery or lead optimization programs in the pharmaceutical industry. Producing QSAR models without detailed structural information is unreliable and is often based on intuition. QSDAR offers, in contrast, a more objective way of building models. Ironically, the use of spectral data–activity relationships allows an unbiased examination of specific structure–activity relationships, which can suggest a more meaningful way to discover antibiotics or other pharmaceuticals. This potential was demonstrated in the heteroatom NMR spectral model of cephalosporin activity.

The inclusion of molecular dynamics entropy information in QSDAR modeling promises access to information not available in QSAR approaches. The example shown illustrates a potential limitation of the technique when the molecular complexity of the training set compounds to be predicted is similar or diverse. If it is not reasonable to build a training set that spans a diverse range, QSAR methods blind to the deleterious effects of entropy may provide superior or complimentary models. One way to design compounds that have stronger binding to a receptor is to lower the configurational entropy in the unbound state while retaining the positive electrostatic contributions and minimal steric hindrances needed for strong binding affinity.

## 4 CONCLUSIONS

Structure and chemical shift information from the 3D connectivity matrix was used to produce very accurate models of biological binding activity. The 3D connectivity matrix uniquely combines quantum mechanical information from

the chemical shifts with nearest neighbor and a compound's internal atom-to-atom distance information. The combined information from COSY and long-range distance connectivity information from the 3D connectivity matrix was used to produce CoSCoSA models that were as accurate or more accurate and reliable than QSAR or E-state models based on separate calculations for electrostatics and steric interactions. The quality of results obtained from CoSCoSA models based on simulated NMR data should improve as improvement occurs in the spectral simulation software and prediction errors are reduced. Three-dimensional QSDAR modeling is a complementary technique that is unique and has widespread application to many informatics challenges. It is not intended to replace 3D QSAR and SAR modeling but to provide computational model developers a sensitive, alternate perspective on biological and toxicological properties. For example, in the case of cephalosporins binding to the transpeptidase enzyme, the CoSCoSA model can give multiple conformation information about binding mechanisms that could not be obtained from 3D QSAR or SAR models. The added advantage of this approach is that it can produce these kinds of results very rapidly.

CoSCoSA modeling is based on the unbiased distance-related distribution of a molecule's NMR spectral features, which directly corresponds to structural elements that, according to structure–activity relationship theory, determine the biological activity of the molecule. The diamagnetic term of $^{13}C$ NMR spectra is directly related to the electrostatic potential and has been shown by QSAR models and physical theory to be directly related to biological activity. The paramagnetic term of $^{13}C$ NMR spectra is related to the type of orbitals around a carbon nucleus. The orbitals around a nucleus have been shown to be directly related to bond energy, local molecular dynamics, local configurational entropy, and biological activity.

Three-dimensional QSDAR models are in a way a combination of conventional SAR and QSAR modeling methods. Instead of using a 3D grid to calculate electrostatic energies, the energies in the form of chemical shifts are put onto the compound itself in all the possible two-atom connections. The diamagnetic term of the chemical shift is directly dependent on the electrostatic energy of a nucleus; the paramagnetic term of the chemical is directly dependent on the type of orbital. So in this sense, 3D QSDAR models split up the electrostatic energies into electron orbital type. The distribution of distances between two atoms is related to molecular shape and configuration entropy. Three-dimensional QSDAR modeling is the only modeling method that can evaluate enthalpy and entropy effects at the same time.

***Acknowledgments*** We thank Kathleen J. Holm, an NCTR summer student who aided in the development of the CoSCoSA models for estrogen receptor binding, cephalosporin antibacterial activity, and estimates of the configurational entropy.

## REFERENCES


1. Cramer, R. D., Bunce, J. D., Patterson, D. E. (1988). Cross-validation, bootstrapping, and partial least squares compared with multiple regression in conventional QSAR studies. *Quant. Struct. Act. Relat.*, *7*, 18–25.
2. Cramer, R. D., Paterson, D. E., Bunce, J. D. (1988). Comparative molecular field analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins. *J. Am. Chem. Soc.*, *110*, 5959–5967.
3. Klopman, G. (1984). Artificial intelligence approach to structure-activity studies. Computer automated structure evaluation of biological activity of organic molecules. *J. Am. Chem. Soc.*, *106*, 7315–7321.
4. Klopman, G. (1992). MULTICASE1. A hierarchial computer automated structure evaluation program. *Quant. Struct. Act. Rel.*, *11*, 176–184.
5. Hansch, C., Leo, A. (1995). *Exploring QSAR—Fundamentals and Applications in Chemistry and Biology.* American Chemical Society, Washington DC.
6. Tong, W., Perkins, R., Xing, L., Welsh, W. J., Sheehan, D. M. (1997). QSAR models for binding of estrogenic compounds to estrogen receptor α and β subtypes. *Endocrinology*, *138*, 4022–4025.
7. Oprea, T. I., Garcia, A. E. (1996). Three-dimensional quantitative structure-activity relationships of steroid aromatase inhibitors. *J. Comput. Aid. Mol. Des.*, *10*, 186–200.
8. Shi, L. M., Fang, H., Tong, W., Wu, J., Perkins, R., Blair, R. M., Branham, W. S., Dial, S. L., Moland, C. L., D. M. Sheehan, D. M. (2001). QSAR models using a large diverse set of estrogens. *J. Chem. Inf. Comput. Sci.*, *41*, 186–195.
9. Katritzky, A. R., Mu, L., Labanov, V. S., Karelson, M. (1996). Correlation of boiling points with molecular structure. 1. A training set of 298 diverse organics and a test set of 9 simple inorganics. *J. Phys. Chem.*, *100*, 10400–10407.
10. Fujita, T., Iwasa, J., Hansch, C. (1964). A new substituent constant, π, derived from partition coefficient. *J. Am. Chem. Soc.*, *86*, 5175–5180.
11. Branbury, S. P. (1995). Quantitative structure-activity relationship and ecological risk assessment: An overview of predictive aquatic toxicology research. *Toxicology*, *25*, 67–89.
12. De Dios, A. C., Pearson, J. G., Oldfield, E. (1993). Secondary and tertiary structural effects on protein NMR chemical shifts: An *ab initio* approach. *Science*, *260*, 1491–1496.
13. Emsley, J. W., Feeney, J., Sutcliffe, L. H. (1965). *High Resolution Nuclear Magnetic Resonance*, Vol. I. Pergamon, Oxford, pp. 1–287.
14. Beger, R. D., Freeman, J., Lay Jr., J., Wilkes, J., Miller, D. W. (2000). $^{13}$C NMR and EI mass spectrometric data-activity relationship (SDAR) model of estrogen receptor binding. *Toxicol. Appl. Pharmacol.*, *169*, 17–25.
15. Beger, R. D., Bolton, P. H. (1997). Protein ϕ and ψ dihedral restraints determined from multidimensional hypersurface correlations of backbone chemical shifts and their use in the determination of protein tertiary structures. *J. Biomol. NMR*, *10*, 129–142.
16. Wishart, D. S., Sykes B. D. (1994). Chemical shifts as a tool for structure determination. *Methods Enzymol.*, *239*, 363–392.

17. Lefebvre B., Williams, A. (2003). $^{13}$C NMR chemical shift prediction: A comparison of methods and a case study analysis of Paclitaxel (TAXOL®). Poster presentation, 44th ENC.
18. Meiler, J., Meusinger, R., Will, M. (2000). Fast determination of 13 C NMR chemical shifts using artificial neural networks. *J. Chem. Inf. Comp. Sci.*, *40*, 1169–1176.
19. Kvasnicka, V. (1991). An application of neural networks in chemistry. Prediction of $^{13}$C NMR chemical shifts. *J. Math. Chem.*, *6*, 63–76.
20. Shade, L., Beger, R. D., Wilkes. J. G. (2003). New computerized method for modeling binding affinities to the aryl hydrocarbon receptor using $^{13}$C NMR spectra. *Environ. Toxicol. Chem.*, *22*, 501–509.
21. Beger, R. D., Freeman, J., Lay Jr., J. O., Wilkes, J. G., Miller, D. W. (2000). Producing $^{13}$C NMR, infrared absorption and EI mass spectrometric data models of the monodechlorination of chlorobenzenes, chlorophenols, and chloroanilines. *J. Chem. Inf. Comput. Sci.*, *40*, 1449–1455.
22. Beger, R. D., Wilkes, J. G. (2001). Developing $^{13}$C NMR quantitative spectrometric data-activity relationship (QSDAR) models of steroid binding to the corticosteroid binding globulin. *J. Comput. Aid. Mol. Des.*, *15*, 659–669.
23. Beger, R. D., Wilkes, J. G. (2001). $^{13}$C NMR quantitative spectrometric data-activity relationship (QSDAR) models to the aromatase enzyme. *J. Chem. Inf. Comput. Sci.*, *41*, 1360–1366.
24. Beger, R. D., Wilkes, J. G. (2001). Models of polychlorinated dibenzodioxins, dibenzofurans, and biphenyls binding affinity to the aryl hydrocarbon receptor developed using $^{13}$C NMR data. *J. Chem. Inf. Comput. Sci.*, *41*, 1322–1329.
25. Bax, A., Grzesiek, S. (1993). Methodological advances in protein NMR. *Acc. Chem. Res.*, *26*, 131–138.
26. Aue, W. P., Bartholdi, E., Ernst, R. R. (1976). Two-dimensional spectroscopy. Application to nuclear magnetic resonance. *J. Chem. Phys.*, *64*, 2229–2246.
27. Bodenhausen, G., Ruben, D. J. (1980). Natural abundance nitrogen-15 NMR by enhanced heteronuclear spectroscopy. *Chem. Phys. Lett.*, *69*, 185–189.
28. Bax, A., Griffey, R. H., Hawkins, B. L. (1983). Sensitivity-enhanced correlation of 15N and 1H chemical shifts in natural-abundance samples via multiple quantum coherence. *J. Am. Chem. Soc.*, *105*, 7188–7190.
29. Bax, A., Summers, M. F. (1986). $^{1}$H and $^{13}$C Assignments from sensitivity-enhanced detection of heteronuclear multiple-bond connectivity by 2D multiple quantum NMR. *J. Am. Chem. Soc.*, *108*, 2093–2094.
30. Kumar, A., Ernst, R. R., Wuthrich, K. (1980). A two-dimensional nuclear Overhauser enhancement (2D NOE) experiment for the elucidation of complete proton-proton cross-relaxation networks in biological macromolecules. *Biochem. Biophys. Res. Comms.*, *95*, 1–6.
31. Clore, G. M., Gronenborn, A. M., Nilges, M., Ryan C. A. (1987). Three-dimensional structure of potato carboxypeptidase inhibitor in solution. A study using nuclear magnetic resonance, distance geometry, and restrained molecular dynamics. *Biochemistry*, *26*, 8012–8023.
32. Bremser, W. (1978). HOSE—A novel substructure code. *Anal. Chim. Acta.*, *103*, 355–365.

33. Livingstone, D. (1995). *Data Analysis for Chemists.* Oxford University Press, New York.
34. Johnson, M., Maggiora, G. M. (1990). *Concepts and Applications of Molecular Similarity.* Wiley, New York.
35. Bishop, C. M. (1996). *Neural Networks for Pattern Recognition.* Oxford University Press, New York.
36. Shi, L. M., Fang, H., Tong, W., Wu, J., Perkins, R., M. Blair, R., Branham, W. S., Dial, S. L., Moland, C. L., Sheehan, D. M. (2001). QSAR models using a large diverse set of estrogens, *J. Chem. Inf. Comput. Sci.*, *41*, 186–195.
37. Good, A. C., So, S. S., Richards, W. G. (1993). Structure-activity relationships from molecular similarity matrices. *J. Med. Chem.*, *36*, 433–438.
38. De Gregorio, C., Kier, L. B., Hall, L. H. (1988). QSAR modeling with electrotopological state indices: Corticosteroids. *J. Comput.-Aided Mol. Design*, *2*, 557–561.
39. Beger, R. D., Buzatu, D., Wilkes, J. G. (2002). Combining NMR spectral and structural data to form models of polychlorinated dibenzodioxins, dibenzofurans, and biphenyls binding to the AhR. *J. Comput.-Aided Mol. Design*, *16*, 727–740.
40. Beger, R. D., Holm, K., Buzatu, D., Wilkes, J. G. (2003). Using simulated 2D $^{13}$C-$^{13}$C NMR spectral data to model a diverse set of estrogens. *Internet Electronic J. Mol. Design*, *2*, 435–453.
41. Beger, R. D., Buzatu, D., Wilkes, J. G., Lay, Jr., J. O. (2002). Developing comparative structural connectivity spectra analysis (CoSCoSA) models of steroid binding to the corticosteroid binding globulin. *J. Chem. Inf. Comput. Sci.*, *42*, 1123–1131.
42. Beger, R. D., Wilkes, J. G. (2002). Comparative structural connectivity spectra analysis (CoSCoSA) models of steroids binding to the aromatase enzyme. *J. Mol. Recognition*, *15*, 154–162.
43. Blair, R. M., Fang, H., Branham, W. S., Hass, B. S., Dial, S. L., Moland, C. L., Tong, W., Shi, L., Perkins, R., Sheehan, D. M. (2000). The estrogen receptor relative binding affinities of 188 natural and xenochemicals: Structural diversity of ligands. *Toxicol. Sci.*, *54*, 138–153.
44. Branham, W. S., Dial, S. L., Moland, C. L., Hass, B. S., Blair, R. M., Fang, H., Shi, L., Tong, W., Perkins, R. G., Sheehan, D. M. (2002). Phytoestrogens and mycoestrogens bind to the rat uterine estrogen receptor, *J. Nutr.*, *132*, 658–664.
45. Bursi, R., Dao, T.,van Wilk, T., de Gooyer, M., Kellenbach, E., Verwer, P. (1999). Comparative spectra analysis (CoSA): Spectra as three-dimensional molecular descriptors for the prediction of biological activities. *J. Chem. Inf. Comput. Sci.*, *39*, 861–867.
46. Waller, C. L., Opera, T. I., Chae, K., Park, H. K., Korach, K. S., Laws, S. C., Wiese, T. E., Kelce, W. R., Gray Jr., L. E. (1996). Ligand-based identification of environmental estrogens, *Chem. Res. Toxicol.*, *9*, 1240–1248.
47. Kuiper, G. G. J. M., Carlsson, B., Grandien, K., Enmark, E., Haggblad, J., Nilsson, S., Gustafsson, J.-A. (1997). Comparison of the ligand binding specificity and transcript tissue distribution of estrogen receptors α and β. *Endocrinology*, *138*, 863–870.
48. Fang, H., Tong, W., Perkins, R., Soto, A., Prechtl, N., Sheehan, D. M. (1998). Quantitative comparison of in vitro assays for estrogenic assays. *Environ. Health Prospect.*, *139*, 723–729.

49. Rumelhart, D. E., McClelland, T. L. (1986). *Parallel Distributed Processing.* Brandford Books/MIT Press, Cambridge, MA.

50. *Jgravity*, version 1.0 beta-1 build-10. Titan Systems, LinCom Division.

51. Devillers, J. (1996). *Neural Networks in QSAR and Drug Design.* Academic, New York.

52. Kellogg, G. E., Kier, L. B., Gaillard, P., Hall, L. H. (1996). E-state fields: Applications to 3D QSAR. *J. Comput.-Aided Mol. Design*, *10*, 513–520.

53. Polanski, J. (1997). The receptor-like neural network for modeling corticosteroid and testosterone binding globulins. *J. Chem. Inf. Comput. Sci.*, *37*, 553–561.

54. Safe, S. (1990). Polychlorinated biphenyls (PCBs), dibenzo-*p*-dioxins (PCDDs), dibenzofurans (PCDFs), and related compounds: Environmental and mechanistic considerations which support the development of toxic equivalency factors (TEFs). *Crit. Rev. Toxicol.*, *21*, 50–88.

55. Bhandiera, S., Sawyer, T., Romkes, M., Zmudzka, B., Safe, L., Mason, G., Keys, B., Safe, S. (1984). Polychlorinated dibenzofurans (PCDFs): Effects of structure on binding to the 2,3,7,8-TDDD cytosolic receptor protein, AHH induction and toxicity. *Toxicology*, *32*, 131–144.

56. Bandiera, S., Safe, S., Okey, A. B. (1982). Binding of polychlorinated biphenyls classified as either phenobarbitone-, 3-methylcholanthrene- or mixed type inducers to cytosolic Ah receptor. *Chem. Biol. Interact.*, *39*, 259–277.

57. Poland, A., Knutson, J. C. (1982). 2,3,7,8-Tetrachlorodibenzo-*p*-dioxin and related halogenated aromatic hydrocarbons: Examination of the mechanism of toxicity. *Annu. Rev. Pharmacol. Toxicol.*, *22*, 517–554.

58. Poland, A., Glover, E., Kende, A. S. (1976). Stereospecific, high affinity binding of 2,3,7,8-tetrachlorodibenzo-*p*-dioxin by hepatic cytosol: Evidence that the binding species is the receptor for induction of aryl hydrocarbon hydroxylase. *J. Biol. Chem.*, *251*, 493–494.

59. Safe, S. (1984). Polychlorinated biphenyls (PCBs) and polybrominated biphenyls (PBBs): Biochemistry, toxicology and mechanism of action. *Crit. Rev. Toxicol.*, *13*, 319–395.

60. Safe, S. H. (1986). Comparative toxicology and mechanism of action of polychlorinated dibenzo-*p*-dioxins and dibenzofurans. *Annu. Rev. Pharmacol. Toxicol.*, *26*, 371–399.

61. Mekemyan, O. G., Veith, G. D., Call, D. J., Ankley, G. T. (1996). A QSAR evaluation of Ah receptor binding of halogenated aromatic xenobiotics. *Environ. Health Perspect.*, *104*, 1302–1310.

62. Turner, D. B., Willett, P., Ferguson, A. M., Heritage, T. (1997). Evaluation of novel infrared range vibration-based descriptor (EVA) for QSAR studies. 1. General application. *J. Comput.-Aided Design*, *11*, 409–422.

63. *Physicians' Desk Reference*, 54th ed. (2000). Medical Economics Company, Montvale, NJ.

64. Joklik, W. K., Willett, H. P., Amos, D. B., Wilfert, C. M. (1988). *Zinsser Microbiolgy*, 19th ed., pp. 128–160.

65. Grail, B. M., Payne, J. W. (2002). Conformational analysis of bacterial cell wall peptides indicates how particular conformations have influenced the evolution of

penicillin-binding proteins, b-lactam antibiotics and antibiotic resistance mechanisms. *J. Mol. Recognit.*, *15*, 113–125.

66. Compadre, R. L. L., Pearlstein, R. A., Hopfinger, A. J., Seydel, J. K. (1987). A quantitative structure-activity relationship analysis of some 4-aminophenyl sulfone antibacterial agents using linear free energy and molecular modeling methods. *J. Med. Chem.*, *30*, 900–906.
67. Pickett, S. D., Sternberg, M. J. E. (1993). Empirical scale of side-chain conformational entropy in protein folding. *J. Mol. Biol.*, *231*, 825–839.
68. Karplus, M., Kushick, J. N. (1981). Method for estimating the configurational entropy of macromolecules. *Macromolecules*, *14*, 325–332.

# USING MICROSOFT EXCEL® AS A LABORATORY DATA MANAGEMENT TOOL

A. Erik Rubin, Mark F. Russo, and William Neil
*Bristol-Myers Squibb*
*New Brunswick, New Jersey*

1 INTRODUCTION
2 EXCEL AS A LABORATORY DATA MANAGEMENT TOOL
Deciding If Excel Is the Right Choice for Your Application
3 TAPPING INTO THE HIDDEN POWER OF EXCEL: VISUAL BASIC FOR APPLICATIONS
Recording a Macro
Fundamentals of VBA and Excel
4 READING, WRITING, AND PRESENTING DATA
Clipboard
Opening and Saving Data Files
Reading and Writing Data Files Using VBA
Importing Data from External Databases
Automating the Generation of Reports
5 SUMMARY AND CONCLUSION

## 1 INTRODUCTION

Few and far between are laboratory scientists who have not faced a data management challenge of some kind in their careers. The chief reason for this is

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

that data are, in fact, the primary deliverables from laboratory experiments. Data management troubles can range from simple tasks such as devising a system to store printouts of high-performance liquid chromatography (HPLC) chromatograms and link them to entries in one's notebook to complicated scenarios such as creating a data storage and visualization application for a high-throughput workflow that operates 24 h a day. With the advent of high-throughput screening (HTS) and high-throughput experimentation (HTE) strategies and techniques and their prevalence in modern laboratories, data collection and storage obstacles are more common than ever. In the midst of this chaotic setting, Excel can be a tremendous asset to someone faced with a data management predicament. The key to success, however, is knowing when Excel is the appropriate tool for the job.

This chapter will explore Excel's utility as a data management tool. We will look at criteria for selecting Excel as a data management solution, examine Visual Basic for Applications (Excel's built-in programming language), discuss common techniques for manipulating data in Excel, and consider strategies for automatic reporting of data.

## 2 EXCEL AS A LABORATORY DATA MANAGEMENT TOOL

### Deciding If Excel Is the Right Choice for Your Application

Consider the following two data management challenges. Can you determine which one is appropriate for Excel?

*Scenario 1* A chemical engineer wishes to measure the rate of filtration of a slurry of product in a reaction solvent during an isolation step. She accomplishes this by directing the filtrate to a collection vessel that sits atop a balance. She plots the mass of the collection vessel (which she reads off the balance display and records in her notebook) as a function of time. The engineer would like to automate this process by having a computer collect the mass from the balance every minute or so, store the data in a table, and plot the results on a graph. She would like to be able to e-mail the results to the other team members.

*Scenario 2* A pharmaceutical company desires a compound management system to facilitate storage and retrieval of its hundreds of thousands of compounds. The data management software should enable tracking of the location of each compound, store information about each compound, relate the usage of each sample back to the original sample entry, and be accessible by hundreds of users through the company's intranet.

You probably guessed correctly. Scenario 1 is indeed the challenge that is appropriate for Excel. How did you choose? What tipped you off? Let us look more closely at the criteria for choosing Excel as a data management solution.

First and foremost, *Excel is not a database management system*. Consequently, rule number one is: If your application requires a database manage-

ment system, Excel is not appropriate. A database program, such as Access or Oracle, is a better choice in that case. In fact, Excel and Access share many things in common. Both allow powerful queries for sorting and filtering, both run sophisticated calculations on stored data, and both store data in tabular format. In addition, both allow the generation of reports in multiple formats, both have a built-in programming language for customizations, and both can connect to external data sources such as files, databases, and instruments. So, how does one decide if the application requires a database such as Access or a spreadsheet such as Excel? There are a few simple guidelines that will help.

A database management system such as Access is appropriate when:

- Your data must be described in multiple tables that are related to each other. That is, you require a relational database.
- You must store a very large amount of data (thousands of entries).
- Your data include large text fields or binary data such as images.
- Many users will be interacting with and updating the data.

Excel is appropriate when:

- You can describe your data in a single table (you do not require a relational database).
- You are primarily performing calculations on, statistical comparisons of, or plotting your data (i.e., your data are primarily numerical).
- Your data set is of reasonable size (no more than 65,536 rows in theory, no more than a few thousand in practice).
- Only a few users will be interacting with and updating the data.

In addition to the technical guidelines above, there are two additional important criteria to consider. First, most scientists have Excel installed on their computers. Second, most scientists know how to use Excel already. Excel is so common in the scientific laboratory, it has been said that if chemists built planes, they'd probably be flown using Excel. User familiarity with the application environment and user interface is of tremendous benefit to an application developer and should be taken advantage of whenever possible.

Revisiting the two hypothetical data management challenges above, it is now clear that scenario 2 is completely inappropriate for Excel. However, closer inspection of scenario 1 reveals a few issues that are not addressed by simply *selecting* Excel as the data management tool. Recall that, in scenario 1, the engineer required that Excel collect data from the balance at predefined time intervals. How will Excel know when and how to get the data from the balance? The better part of this chapter is dedicated to learning how to get data into and out of Excel. The key is knowing how to tap into the hidden power of Excel. That is, learning how to exploit Excel's built-in programming language, Visual Basic for Applications.

## 3 TAPPING INTO THE HIDDEN POWER OF EXCEL: VISUAL BASIC FOR APPLICATIONS

As a Microsoft Excel user, you have undoubtedly come across Visual Basic for Applications (VBA) at some point even though you may not have realized it. Have you ever recorded a macro in Excel or used a macro recorded by someone else? Have you ever installed "software" from an instrument vendor that modifies Excel by adding a new menu item to the menu bar in order to expose some specific functionality? In fact, it is the built-in VBA functionality in Excel that enables both of these scenarios. In this section, we'll learn how.

### Recording a Macro

Getting started with VBA and Excel is actually very easy: you merely create a macro. A macro is simply a routine or sequence of events (mouse clicks, key strokes, etc.) that is *recorded* and then, when the macro is run, played back exactly. Macros are very useful for lengthy tasks in Excel that are needed frequently. Recording a macro is a cinch. To begin, launch Excel and be sure that a blank workbook is open. On the Tools menu, point to Macros, and then click on Record New Macro. If you like, give your macro a name by entering it in the Macro name box. Click OK. You will see a new toolbar pop up that has a VCR-style stop button on it. You will need to click this button when you want to stop recording your macro. Type some numbers into the first five or six cells in column A on the active worksheet. Highlight those cells and then chart them using the Excel Chart Wizard (use the default wizard settings for this example). Click the Stop button to stop recording the macro. Your worksheet should look something like Figure 1. To test the macro, clear the worksheet (including the charts) and, on the Tools menu, point to Macro, then click on Macros. Select your macro from the list box and click Run. Your worksheet should look exactly as it did before clearing it (Fig. 1).

Under the hood, macros are nothing more than VBA subroutines that were written automatically by the Excel macro recorder as you performed the actions. In fact, the Excel macro recorder is the single most powerful instrument available to the Excel VBA programmer for learning how to write code to perform certain tasks. If you do not know how to do something, record a macro and look at the generated code. It is that simple. To view the code generated by the macro recorder, you must open the Excel Visual Basic Editor (VBE). To do this, on the Tools menu, point to Macro, and click on Visual Basic Editor. The VBE looks like, and in many ways is, a completely separate program embedded within Microsoft Excel. With the VBE open, you will see a window titled Project in the upper left-hand corner. This is the Project Explorer. In this window, you will see a tree structure with one or more root entries each called VBAProject (Fig. 2). In parentheses, after "VBAProject," the name of the workbook will be displayed (e.g., "Book2" in this

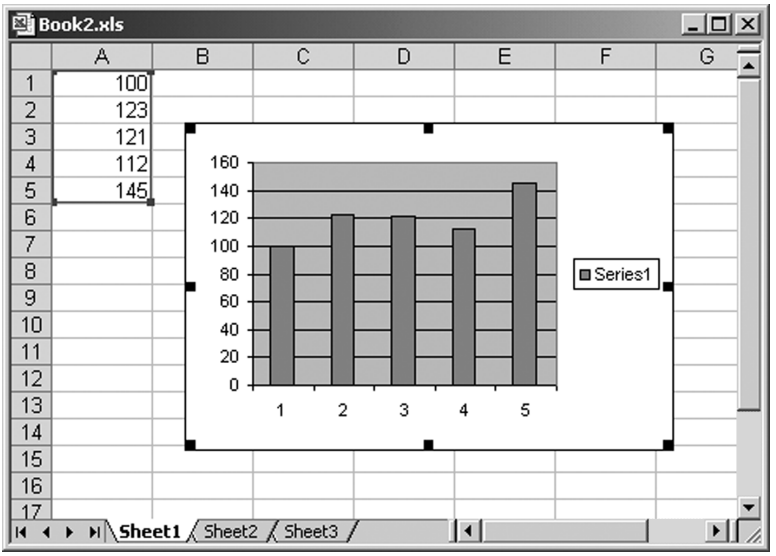


**Figure 1** Macro output.

example). Expand the root node that corresponds to the workbook that you are working with, then expand the node with the name Modules. Double-click on the node called Module1. This will display the module that contains the code for your macro. Your screen should look something like Figure 2. The code that is displayed is raw VBA code. If you have never seen VBA code before, congratulations! You just wrote your first VBA code block without typing a single line. Now that we have seen VBA code, in the next section we will delve into the meaning of this code and learn how one can use VBA to manipulate Excel.

## Fundamentals of VBA and Excel

To really become an expert VBA programmer in Excel, it takes a lot of patience and practice and it helps to have a project or a problem that you are trying to solve. There are many excellent texts written on the subject and new books are popping up continuously. The reader is encouraged to consult these references as his or her education progresses. Although a complete treatment of the interaction between VBA and Excel is neither feasible nor appropriate in this chapter, the authors believe that the key to understanding the power of VBA is appreciating four simple concepts that are fundamental to programming with VBA in Excel. These four concepts are:

- Understanding the structure of a VBA application (where the code is located)

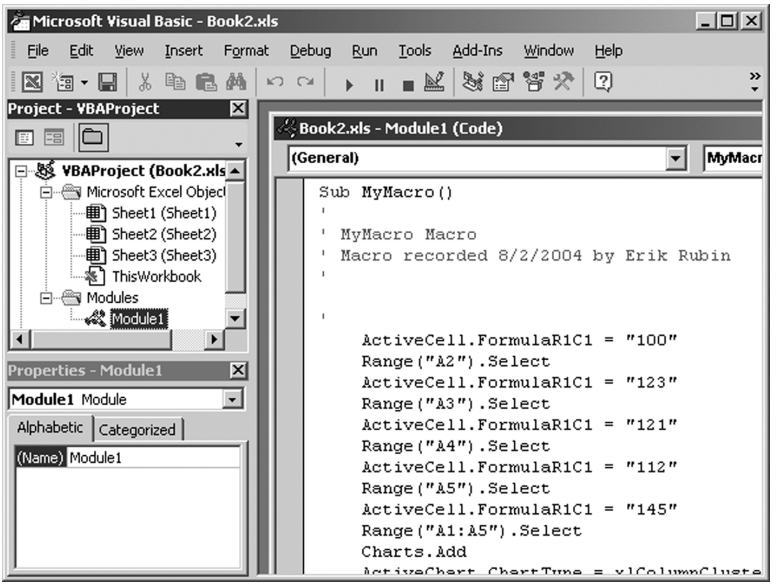


**Figure 2** Visual basic editor (VBE) with macro code showing.

- Basic VBA language syntax (how to write the code)
- Event-driven code execution (when the code will run)
- Interacting with Excel objects (how to manipulate Excel through code)

In this section we will introduce each of these concepts to familiarize the reader. Later, we will address the application of these concepts to simple data management challenges.

***Visual Basic Editor: Elements of a VBA Program*** As you probably have gleaned from the introduction to Excel macros, Excel VBA code is written in the Visual Basic Editor (VBE) and the code is stored in code modules. VBA code modules are associated with an Excel workbook and are actually embedded into the binary ".XLS" file (i.e., they are not stored as separate files). There are four types of VBA code modules in which you can write code. These are (1) standard modules, (2) user forms, (3) class modules, and (4) Microsoft Excel Object modules. Let us look at each of these types of code modules.

1. We have already seen standard modules. The Excel macro recorder added a standard module called Module1 to your VBA project when you recorded a new macro in the example above. To add a standard module to the VBA project yourself, switch to the VBE (ALT+F11) and activate the VBA project to which you want to add a module (to activate the project, click on your VBAProject in the project explorer). On the Insert menu, click on Module. A new module will be added to the Modules

node in the project explorer. A standard module is a common place to store procedures and functions for your VBA program. You can organize your code by using several modules and naming each module appropriately.

2. UserForms allow the VBA programmer to create custom dialog boxes that aid in interfacing with the end user. To add a UserForm to your project, click on UserForm on the Insert menu while in the VBE. A custom UserForm is a great tool when you need to collect information from a user in order to move to the next step in a program. A user form actually consists of two parts: a form designer and a code module. The form designer is where the programmer lays out the look and feel of the form and decides where the controls will be located. The code module is where the programmer writes code to respond to end-user actions on the form (e.g., when the user selects an item from a drop-down or clicks a button). Designing custom user forms is very simple, although we will not discuss them further here.
3. Class modules allow the programmer to create custom objects. A treatment of object-oriented programming is out of the scope of this chapter.
4. Excel Object modules, which are actually class modules, are associated with Excel workbooks and worksheets. There is one module called ThisWorkbook that is associated with each workbook, and each worksheet in the workbook has an associated sheet object module. You cannot add and remove sheet object modules in the traditional way. They are added and removed along with the actual Excel worksheets. You cannot delete the ThisWorkbook module ever. These modules can hold code in a manner similar to a standard module. These modules are especially useful, however, because they allow the programmer to respond to (write code for) specific events associated with the underlying object. For example, you could run initialization code whenever a workbook is opened, or run cleanup code when a workbook is closed. You could also prevent the deletion of a worksheet. We will look more closely at this topic when we address event-driven programming later in this chapter.

***VBA Language Primer*** The Visual Basic for Applications language is like all programming languages at the root. Interim data are stored in variables, code stored in code modules are organized into callable procedures and functions, and the flow of a sequence of code is controlled by standard execution control constructs such as loops and conditional statements. In this part, we will introduce the VBA syntax associated with these core programming elements.

As mentioned, interim data are stored in variables. In VBA, variables are declared using the `Dim` keyword and can hold several different intrinsic data types. For example, to declare a variable of type integer, one would use a code statement similar to the one in Listing 1:

```
Dim intMyInteger As Integer
```

**Listing 1** Variable declaration in VBA.

This statement can occur anywhere in a code module and essentially defines a placeholder for your data called `intMyInteger`. As we will see, variables can be used within procedures and functions to manipulate the data the variables contain. There are several intrinsic data types available to the VBA programmer; Table 1 lists the most common and useful of the lot. When a variable data type listed in Table 1 is found after the `As` keyword in a variable declaration, it defines the type of data that is stored in the variable. For example, the variable `intMyInteger` above could not store the text value "Visual Basic for Applications", which is of type String.

When code is executed in VBA, it is actually executed by calling a procedure. The code is contained within the definition of the procedure. For example, in the macro recording example above, the macro code is contained within a procedure called "MyMacro" (Fig. 2). When the macro is run, the Excel macro execution engine actually searches for and runs the procedure that has the same name as the macro requested. VBA offers a number of ways to define procedures, and we will introduce the two most common: subroutines and functions.

A subroutine is a list of code statements (a code block) that is contained within `Sub` and `End Sub` statements. A subroutine does not return a value to the caller after its execution. In Listing 2, subroutine `MyFirstSubroutine` declares a variable of type String and then sets the value of that variable to "I love hockey". If you were to call this subroutine, it would execute the two statements and then exit uneventfully:

```
Sub MyFirstSubroutine()
     Dim strMyString As String
     strMyString = "I love hockey"
End Sub
```

**Listing 2** A subroutine.

**TABLE 1 Common Intrinsic VBA Data Types**

| Variable Data Type | Data Type Description |
|---|---|
| Boolean | True or False |
| Integer | An integer in the range –32,768 to 32,767 |
| Long | An integer in the range –2,147,483,648 to 2,147,483,647 |
| Double | A floating-point number roughly in the range –1.79769313486232E308 to 1.79769313486232E308 |
| String | A sequence of characters |
| Variant | A data type that can hold any intrinsic data type |

A function, on the other hand, is a code block that is enclosed by `Function` and `End Function` statements. Unlike subroutines, a function returns a value to the caller after it executes. For example, in Listing 3 the function `MyFirstFunction` simply returns the String value "I love hockey too". Note that the function is defined as a data type (String, in this case). This data type is the type associated with the returned value. If no type is indicated, the default type is Variant.

```
Function MyFirstFunction() As String
    MyFirstFunction = "I love hockey too"
End Function
```

**Listing 3** A function.

In addition, one procedure can call another procedure. So, in Listing 4, when `MySecondSubroutine` is executed, `strMyString` will be set to the return value of `MyFirstFunction`, which is, if you recall from Listing 3, "I love hockey too".

```
Sub MySecondSubroutine ()
    Dim strMyString As String
    strMyString = MyFirstFunction()
End Sub
```

**Listing 4** Example of a procedure calling another procedure.

As we learned, one procedure can call another and that procedure can call yet an other procedure, and so on, and so on. But who or what calls the first procedure in the sequence? The answer to that question is, in fact, one of the most important concepts associated with learning how to program VBA in Excel. We will answer that question in the next part of this section, event-driven programming. But before we move on, there are a few more basic VBA syntax topics we must cover.

Subroutines can take arguments. Arguments are variables that are defined right in the declaration of the subroutine and act as a way to pass information to a subroutine when it is called. For example, in Listing 5, function `TimesTwo` defines an argument, called `MyNumber`, of type integer. This function takes the number held in `MyNumber` and doubles it. The function then returns the doubled value to the caller.

```
Function TimesTwo (MyNumber As Integer) As Integer
    TimesTwo = MyNumber * 2
End Function
```

**Listing 5** Passing arguments to procedures.

VBA offers standard methods of code execution control as well. These are code constructs that allow looping (or iteration) and condition evaluation (or

selection). VBA offers two main looping constructs: `For-Next` and `Do-Loop`. Listing 6 demonstrates a `For-Next` loop that increments a variable named `I` from a value of 1 to 5. During each iteration, the value of `I` is added to another variable, `J`. When the code block is finished executing, `J` holds the value 15, which is the result of 1 + 2 + 3 + 4 + 5.

```
Dim I As Integer
Dim J As Integer
For I = 1 to 5
     J = J + I
Next I
```

**Listing 6** A `For-Next` Loop example.

A `Do-Loop`, on the other hand, executes continuously while or until a condition is met. For example, the condition can be checked each time before a code block is executed as in Listing 7.

```
Dim I As Integer
Do While I < 10
     I = I + 1
Loop
```

**Listing 7** A `Do While-Loop` example.

If the condition is verified at the end of each block execution, then the `Until` keyword is placed after the `Loop` keyword as in Listing 8. In fact, `While` and `Until` can be used interchangeably in both examples. It is more readable, however, to use the convention `Do While-Loop` and `Do-Loop Until`.

```
Dim I As Integer
Do
     I = I + 1
Loop Until I < 10
```

**Listing 8** A `Do-Loop Until` example.

In addition to iteration execution control, VBA defines an important conditional construct: `If-Then-Else`. The `If-Then-Else` construct may also include an arbitrary number of conditions by adding `Elseif` statements. For example, consider Listing 9.

```
Dim MyNumber As Integer
MyNumber = 4
If MyNumber = 1 Then
    MsgBox "MyNumber is 1"
Elseif MyNumber = 5 Then
    MsgBox "MyNumber is 5"
Else
    MsgBox "MyNumber is neither 1 nor 5"
End If
```

**Listing 9** An `If-Then-Else` Block.

The code in Listing 9 first declares an integer variable and sets its value to 4. The `If-Then-Else` block that follows opens with a statement that is interpreted as "If the number held in the variable `MyNumber` is equal to 1 then execute the code right underneath me (but before the next condition)." If the initial condition is not met (i.e., the number is not 1), the next condition is evaluated. In this case, the next condition reads "If the number held in the variable `MyNumber` is equal to 5 then execute the code right underneath me." If that condition is not met either, the next condition, `Else`, catches all remaining conditions. The `If-Then-Else` construct ends with the `End If` statement.

***Event-Driven Programming*** Let us return, for a moment, to the VBA Primer part above, where we presented the "chicken and egg" scenario for code execution. We introduced the fact that one procedure can call another and that that procedure can in turn call another procedure. The question we left hanging was "Who calls the first procedure?" In VBA, this question is the same as asking "How do I run my program?" The answer is "by responding to events." We have seen an example of event-driven programming already: running our macro. To run the macro, we know to use the Excel macro player from the Tools menu. Another way to interpret this is that the macro code was run as a response to the event associated with the user selecting to play the macro from the Tools menu. Most of your code in a VBA program will be run in response to events. Some examples of events that you might code for are a user clicking a button, a user selecting a custom menu item, or a user closing or opening a workbook. In short, when and how your code will run will depend on one main thing: the events for which you coded responses.

Event-driven programming is best illustrated with an example. Let us say you wanted to modify Excel's behavior to include the following features for your workbook: When a new worksheet is added to the workbook, (a) the date and time that workbook was opened is printed in cell A1 of the new sheet, and (b) the date and time that the worksheet was created is printed in cell A2 of the new sheet. Solving the problem in this example involves two steps:

1. When the workbook opens, store the current date and time.
2. When a new worksheet is added, print the stored time from (1) in cell A1 of the new sheet and print the current date and time in cell A2.

Coding for these two steps involves responding to the event associated with a workbook opening and the event associated with a new worksheet being added to a workbook. Without programmatic events telling the programmer when these actions occur, it would be impossible to predict when a new sheet was added to the workbook, for example. Let us look at a VBA solution to the problem stated above. To begin, open Excel and switch to the VBE. While in the VBE, activate the VBAProject associated with your workbook by clicking on it in the Project Explorer. From the Insert menu in the VBE, click Module. In the newly created standard module, type the code in Listing 10.

```
Dim WorkbookOpenTime As Variant

Function GetWorkbookOpenTime() As Variant
    GetWorkbookOpenTime = WorkbookOpenTime
End Function

Sub SetWorkbookOpenTime()
    WorkbookOpenTime = Now
End Sub
```

**Listing 10** Part 1 of the code for event-driven programming example.

The code in Listing 10 first declares a variable called `WorkbookOpenTime`. This variable will store the date and time when the workbook was opened. Next, a function called `GetWorkbookOpenTime` is defined, which will return the value of the variable `WorkbookOpenTime`. Finally, a subroutine, `SetWorkbookOpenTime`, is defined that sets `WorkbookOpenTime` to the current date and time. You have probably noticed the use of keyword `Now` in the code above. `Now` is a special function, built into VBA, that returns the current date and time.

To write code for the workbook events that we need for this example, open the `ThisWorkbook` code module by double clicking on it in the Project Explorer. At the top of the code window, you will see two drop-down boxes. From the left drop-down box, choose Workbook. With Workbook selected in the left box, the right drop-down box will list all of the available events that you can code for in Excel. Select Open from this menu. VBA automatically adds code to the module for you. The code is the outline (or stub code) for the event procedure that will be called when the workbook is opened. An event procedure is no different from the subroutines we discussed previously. There is one exception, however: The name of an event procedure is very specific. The name must be in the format `Object_Event`. In the present case,

the event procedure name is `Workbook_Open`. In most cases, as in the present case, Excel writes the shell of the event procedure for you automatically. To add the stub code for the other event that we need to respond to, choose Workbook again from the left drop-down box, then click on NewSheet from the right drop-down box in the `ThisWorkbook` code module. This will add the event procedure shell for the NewSheet event. To respond to an event, all the programmer must do is write code between the `Sub` and `End Sub` statements in the event procedure. For this example, complete the automatically generated event procedure shells as in Listing 11.

```
Private Sub Workbook_Open()
    SetWorkbookOpenTime
End Sub

Private Sub Workbook_NewSheet(ByVal Sh As Object)
    Sh.Range("A1").Value = "Workbook Opened: " & _
         GetWorkbookOpenTime
    Sh.Range("A2").Value = "Sheet Created: " & Now
End Sub
```

**Listing 11** Event procedure code.

That is all that is necessary to code a solution to our problem. In Listing 11, the first event procedure fires when the workbook is opened. The subroutine `SetWorkbookOpenTime` is called, which, if you recall from Listing 10, sets the variable `WorkbookOpenTime` to the current time. The second event procedure fires when any new worksheet is added to the workbook. Note that the event procedure code for the NewSheet event passes an argument, `Sh`, of type Object. This is a reference to an Excel object that represents the new worksheet that was added. This reference is used to place the workbook opened date in cell A1 and the worksheet creation date in cell A2. In the next part, we will explore Excel objects and their use in more detail.

***Excel Objects*** Understanding Excel objects is as essential to programming in Excel as the steering wheel is to driving your car. Fully appreciating Excel objects, however, is not a simple task. The final part of this section is dedicated to presenting Excel objects. In order to properly introduce Excel objects we must first understand the concept of software objects.

Objects are software entities (blocks of code) that contain certain related functionality. The term used to describe the containment of this functionality is *encapsulation*; objects are said to *encapsulate* related functionality. In practice, software objects make their encapsulated functionality available to the programmer through procedures and variables, just like the procedures and variables that we introduced earlier in this chapter. With objects, however, variables are not usually directly accessible and must be read and set through

properties that the object makes available. By convention, the procedures that an object makes available are called methods, and not subroutines or functions. In technical lingo, an object is said to *expose* its methods and properties, and the specific methods and properties that an object exposes is referred to as the object's *interface*. Syntactically, an object's interface is addressed through the so-called *dot notation*. For example, let us take an arbitrary object called `MyInstrument` and assume it exposes one property called `Name` (that returns a string containing the instrument name) and one method called `Start` (that starts the instrument). In VBA code, the method would be called using the syntax `MyInstrument.Start` and the property would be read using syntax similar to `SomeStringVariable = MyInstrument.Name`. In dot notation, the object's name and the property or method name are separated by a period.

Oftentimes, software objects correspond to entities that are readily identifiable. This is the case with many of the objects in Excel. For example, two of the main object types that programmers frequently interact with are the Workbook and Worksheet objects. These objects encapsulate all of the functionality associated with Excel Workbooks and Worksheets. The Workbook and Worksheet objects' interfaces allow the programmer, through VBA, to interact with, manipulate, and modify the Worksheets in an Excel workbook and the Workbook in an Excel file. Excel objects represent a bridge between the VBA core language and the Excel application. For this reason, interacting with Excel objects is the crux of programming VBA in Excel. There are hundreds of Excel objects and they are arranged in a hierarchical representation called the Excel *object model*. The object browser offers an excellent way to inspect the Excel object model and all of the properties and methods that Excel objects expose. To open the object browser, while in the VBE click on Object Browser in the View menu. With the object browser open, select Excel from the top drop-down menu (this will probably initially say ⟨All Libraries⟩ by default). The left pane of the browser lists all of the available Excel objects. The right pane lists all of the available properties and methods exposed by the object selected in the left pane.

Let us explore Excel objects more deeply by investigating the Worksheet object more closely. Every reader of this chapter is likely very familiar with the Excel worksheet concept. An Excel workbook is made up of a collection of Excel worksheets and the worksheets are where data, tables, and formulas are stored. What types of properties and methods can you imagine for the Worksheet object? The Worksheet object has a `Name` property, which changes the text displayed on the worksheet's tab. It has a `Cells` property that represents the cells on the worksheet. We will look at the `Cells` property when we address the Excel Range object later in this section. In fact, the Excel Worksheet object exposes 53 properties to the VBA programmer. Would it surprise you to learn that, included in its 27 methods, the Worksheet object exposes a `Delete` method that removes the worksheet from the workbook and a `CheckSpelling` method that verifies the spelling of words located in

the worksheet's cells? In actuality, most of the Worksheet object's interface should be familiar to the everyday Excel user because it facilitates everyday worksheet tasks programmatically.

Let us return to the last example of the previous section on event-driven programming. In Listing 11, we wrote code to respond to the Workbook event, NewSheet, that fires when a new worksheet is added to the Excel workbook. In the event procedure, we made use of the Worksheet object, `Sh`, which is passed to the programmer as an argument when the event procedure is called by Excel. `Sh`, in this case is a variable that references a Worksheet object. The specific worksheet object that it references is the one that was just added to the workbook. We used the worksheet's `Range` property to place data in the newly added worksheet's cells. This code illustrates another feature of objects: An object's property can be an object too. In the code, we access the Worksheet object's `Range` property. This property returns an object of type Range (an object we will discus shortly), and it, in turn, exposes a property called `Value` that gives us access to the contents of the specific range or cell. Syntactically, this is the origin of the "dot-dot" notation that we see in the event procedure code in Listing 11.

When it comes to data management, there is no Excel object more important than the Range object. This is, of course, because the whole business of data management using Excel is getting data into and out of cells, and an Excel Range object is the object model representation of a cell in a worksheet. In fact, a Range object can be any collection of cells in a worksheet. It can be contiguous or disjoint. It can represent an entire row or an entire column. A range can even represent a random selection of cells on a worksheet. This may seem complicated at first. It may seem as though, unlike the Workbook and Worksheet objects, the Range object is not associated with any readily identifiable Excel component. At closer inspection, however, one sees that a Range object can be anything that you, as an Excel user, can select on a worksheet. For example, if you click on cell A1, you have just selected a range that comprises one cell. If you hold down the Ctrl key and click on cell B6, you now have selected a range that comprises two noncontiguous cells. In you click on the "H" column header, you selected a range that comprises an entire column. So, in fact, a Range object is identifiable after all.

Using a Range object associated with a worksheet is quite simple. A Worksheet exposes two key properties that return Range objects when they are called. These are the `Cells` and the `Range` properties. The key difference between these two properties is the format of the arguments that they take. The `Cells` property accepts two integers that refer to the row and the column of the cell that you would like to reference. For example, to obtain a reference to Cell B1, you could use the VBA code in the Listing 12.

```
Dim MyRange As Range
Set MyRange = ActiveSheet.Cells(1,2)
```

**Listing 12** Using the Cells property of the Worksheet object.

Listing 12 highlights one of the most frustrating elements of learning how to program in VBA as a beginner. As a reader, you have just read through thousands of words introducing you to VBA syntax and Excel objects, but in these two lines of code in Listing 12, there are three things that have not been introduced anywhere in this chapter. The reality is that mastering VBA and Excel takes time, patience, and experience, and no single chapter can do the subject justice. Specifically, the new material presented here are the use of a Range type in a variable declaration, the use of the `Set` keyword, and the use of `ActiveSheet`. Just like the declaration of a variable of a type intrinsic to VBA (Table 1), a variable can be declared as any of the Excel objects in the object model. One can look at Excel objects as data types *intrinsic* to Excel. For example, to declare a variable of type Worksheet, one need only type `Dim wsht As Worksheet`. In the example above, `MyRange` is a variable of type Range. There is one difference between variables that are Excel object types and variables that are intrinsic data types. The difference is the use of the `Set` keyword during assignment. For example consider the code in Listing 13:

```
Dim J as Integer
J = 1000
```

**Listing 13** Standard assignment to a VBA variable.

The above code is syntactically correct and will run with no problem. However, the code in Listing 14 would result in a error:

```
Dim R as Range
R = ActiveSheet.Cells(1,2)
```

**Listing 14** Incorrect assignment of an object to a variable.

The difference is that assignment to object variables requires the `Set` keyword. So, the correct syntax for the above code is shown in Listing 15:

```
Dim R as Range
Set R = ActiveSheet.Cells(1,2)
```

**Listing 15** Using the `Set` keyword to assign an object to a variable.

Finally, `ActiveSheet` is a globally available reference to a Worksheet object that represents the active worksheet. The active worksheet is the worksheet in the foreground of the active workbook. That is, the worksheet currently being "worked on" by the Excel user.

Getting back to the Range object, we see that the code `ActiveSheet.Cells(1,2)` returns a Range that represents the cell in row 1, column 2 of the active worksheet. That is, the code returns the `Range` object corresponding to cell B1. The `Range` property of the Worksheet object oper-

ates a little differently. In its most common syntax, the Range property accepts a string with A1-style notation as the argument. For Excel users, this is a familiar syntax because it is the same as the syntax used to reference cells in Excel formulas. For example, the following code would be used to reference the six cells in the Excel range A1:B3:

```
Set R = ActiveSheet.Range("A1:B3")
```

**Listing 16** Using the Range property of the Worksheet object.

Like all objects, the Range object exposes properties and methods. The most useful Range properties to the beginner, however, are the `Value` and `Formula` properties. The `Value` property allows the programmer to read and write the contents of a cell represented by a Range. For example, the code in Listing 17 sets the contents of the cell A3 to "I just set this programmatically".

```
ActiveSheet.Range("A3").Value = _
            "I just set this programmatically"
```

**Listing 17** Putting data into an Excel cell.

Similarly, the code in Listing 18 gets the contents of cell C4 and stores the result in a variable called `MyVar`:

```
Dim MyVar as Variant
MyVar = ActiveSheet.Cells(4,3).Value
```

**Listing 18** Getting data out of an Excel cell.

The Formula property allows the programmer to set and get a `cells` formula through code. For example, the code in Listing 19 places the number 10 in cell A3 and then assigns the Excel formula "=A3/2" to cell A4. The result is displayed as the number 10 in cell A3 and the number 5 in cell A4.

```
ActiveSheet.Range("A3") = 10
ActiveSheet.Cells(4,1).Formula = "=A3/2"
```

**Listing 19** Using the Formula property of the Range object.

In this section, we introduced the hidden power of Excel: Visual Basic for Applications. We began with macro recording, traveled through basic syntax, looked at event-driven programming, and culminated with Excel objects, where we learned how to get data into and out of Excel cells. The heart of data management programming in Excel is getting data into and out of Excel cells, and the remainder of this chapter will explore how we can build on what was introduced in this section to address more intricate data management challenges.

## 4 READING, WRITING, AND PRESENTING DATA

Excel has many powerful tools for managing and manipulating data. But, before these tools can be applied, your data must be moved into the cells of an Excel worksheet. This task is relatively straightforward for small, conveniently formatted data sets. Unfortunately, it can be very difficult for larger data sets that are stored in a format not recognized by Excel. After completing the desired data manipulation, frequently it is necessary to insert the data into a worksheet, display it graphically, or move the data out of Excel, into a file, another program, or a reporting tool. In this section we will describe how to move data into and out of Excel, and how to present it.

### Clipboard

The simplest way to move data into and out of Excel is through the clipboard, using the copy and paste functions. When using your favorite program, if you highlight text, an image, or some other data, and then type Ctrl+C or select Copy from the Edit menu, the data you have highlighted are copied to the Windows clipboard. The format of the data on the clipboard will vary. Indeed, many times multiple formats are stored on the clipboard in order to maximize the number of ways in which the data can be pasted. For example, when you highlight and copy data from the cells of an Excel worksheet, the data are copied to the clipboard in a surprising number of formats. You can see this by copying data from cells in Excel, then selecting Paste Special from Microsoft Word's Edit menu. Multiple formats are listed in the Paste Special dialog that is displayed, including various formatted and unformatted textual options, and even the option to paste the data as an image. The more you understand about the formats used by various programs to write data to the clipboard, the better you will be able to transfer your data using the clipboard.

For exchanging data with Excel, the most useful clipboard format option is the basic Unformatted Text format. Understanding this will help you appreciate the result of pasting data copied from an Excel worksheet into a plain text editor, such as Notepad, or how to format data in a text editor to ensure that it is copied and pasted properly into Excel. The format used by Excel for its Unformatted Text paste option is simple: Data items in the cells of a row are separated by tab characters, and rows are terminated with two characters in sequence—a carriage return followed by a linefeed. You can see this by copying data from cells in Excel and pasting into a Microsoft Word document using the Paste Special option of the Edit menu. Once pasted, your data will likely be arranged in a nice tabular format. To see the special tab and line terminating symbols in Word, select the paragraph symbol (¶) from Word's Standard toolbar. Tab characters will be displayed as right-pointing arrows (→) and the carriage-return+linefeed sequence of two characters will be displayed as the paragraph symbol (¶). To prove that the data are in a natural format for Excel, select the text in Word and copy it to the clipboard. Return to Excel,

highlight an empty cell and paste it. This time there is no need to use Paste Special. Your data will be separated automatically into the cells of your worksheet.

Other options are available for when the data are not in the preferred tab-delimited Excel format. Let us assume you have a microplate reader that saves measured data to a text file as eight rows, with each row containing 12 comma-delimited numbers. For example, consider the file in Figure 3, which has been opened in the Notepad text editor. If you select and copy the contents of the file, and paste them into Excel, you'll end up with something that looks like Figure 4. Excel will not automatically split the numerical values among the cells of the worksheet. Instead, each row in the file will be pasted into the cells of the first column of the worksheet. To split the data into adjacent cells, select the data in the first column and choose the Text to Columns option from the Data menu in Excel. From the wizard that is displayed after this option is selected, choose Delimited from the first step and click the Next button. On the second step of the wizard check Command and Space from Delimiters, and the "Treat consecutive delimiters as one" option (see Fig. 5). The Comma option splits each row at the comma characters. The Space and "Treat

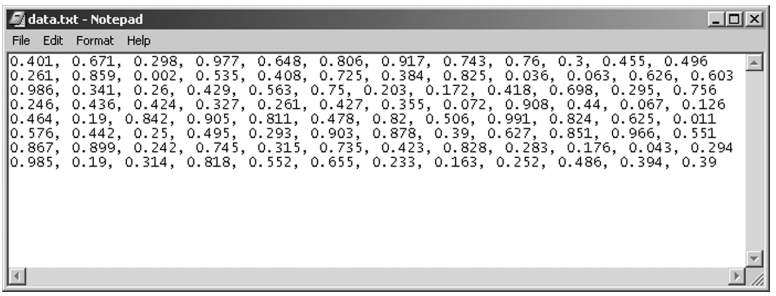


**Figure 3** Typical data file generated by a laboratory instrument.

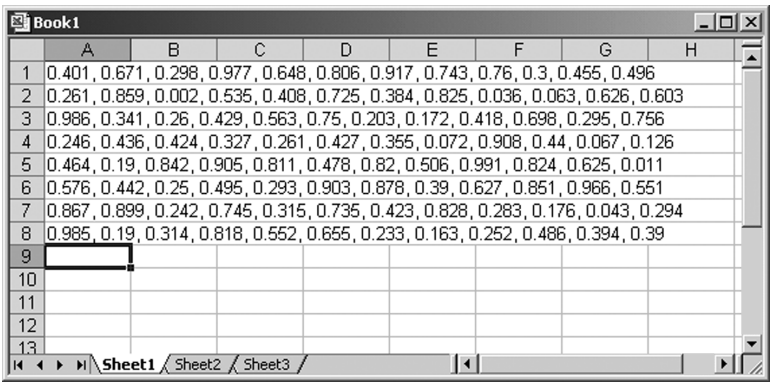


**Figure 4** Comma-delimited data is pasted into the first cell of each row.

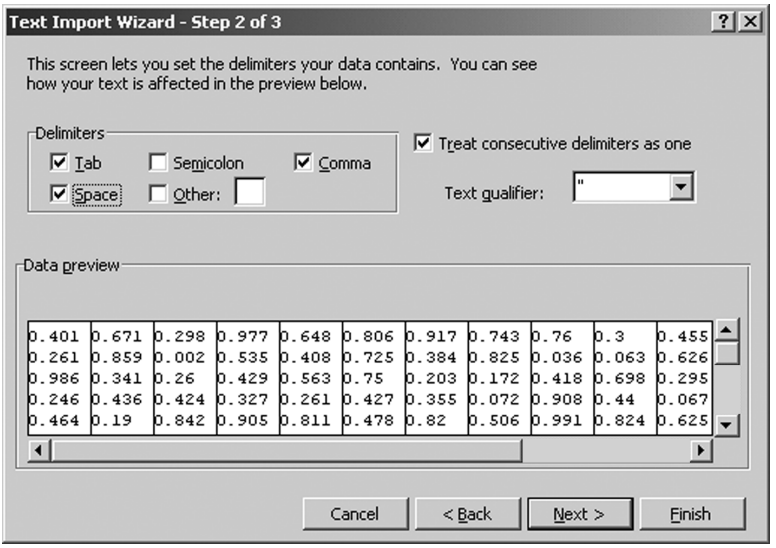


**Figure 5** Second step of the Text to Columns wizard allows column dividers to be selected.

consecutive delimiters as one" options effectively eliminate the extraneous space that follows each comma in each row. Clicking the Finish button will cause the data to be split into the adjacent cells of a worksheet. If you paste data into the worksheet once again before Excel is closed, the same "Text to Columns . . ." options will be applied. If the copied data are formatted similarly, each line will be split into adjacent cells automatically without the need to select the "Text to Columns . . ." menu option. Once Excel is closed, the options are reset to their initial values and will need to be selected again the next time Excel is opened.

## Opening and Saving Data Files

***Text Import Wizard*** Some laboratory instruments have the ability to save measurements as native Excel files. It is much more common for an instrument to store measured data in plain text files. A data file may contain much more than a conveniently formatted table of numbers, but often, at some location in the file, the data are stored in a simple tabular format. Rather than open the file in a text editor and copy the data to the clipboard, Excel makes it possible to open the file directly into a worksheet. When attempting to open a text file, Excel recognizes that the file is plain text and displays the Text Import Wizard. Figure 6 is the dialog that will be displayed when you attempt to open the previous data.txt (Fig. 3) file directly in Excel using the Open menu option of the File menu. Note that this wizard is almost identical to the Text to Columns wizard. The Text Import Wizard gives you a couple of extra

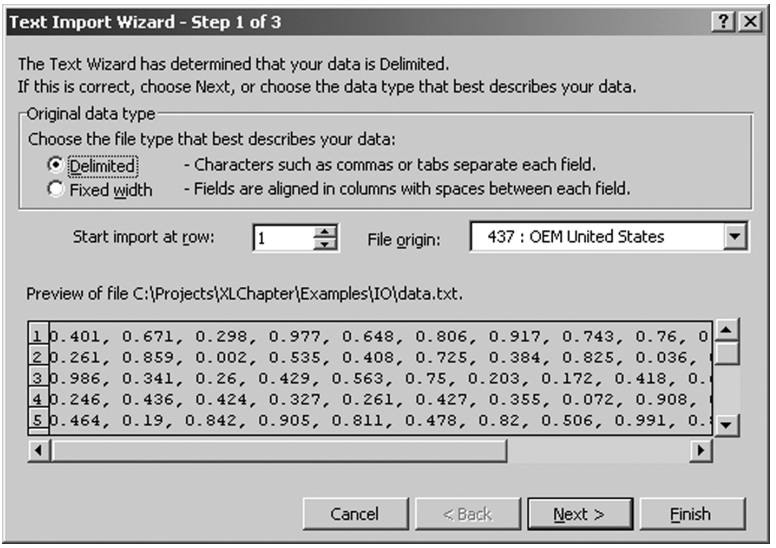


**Figure 6** First step of the Text Import Wizard.

options, including the ability to skip rows in the file, which can be handy when you would like to skip over file header information. The Text Import Wizard process for splitting data into the cells of your worksheet is identical to the Text to Columns Wizard. This is a convenient option when data are stored in a file in a simple tabular format.

***Saving Text and CSV Files*** Excel also provides an easy way to save data in a worksheet to plain text files rather than in Excel's native binary file format. Choosing the Save As option from the File menu will display the "Save As" dialog. In addition to specifying the name and location of the file to save, the file type can be selected. If the "Text (tab delimited)" option is selected, the data in the worksheet will be saved to a plain text file, where the data in each row is stored as a row in the text file, with data items separated by tab characters. The "CSV (comma delimited)" option in the "Save As" dialog will save the data in a similar format, with the only difference being that the data items are separated by commas.

## Reading and Writing Data Files Using VBA

When plain text instrument files are written in a format that is not amenable to copy-and-paste or the Import Wizard, we must do the heavy lifting associated with opening the file, reading its contents, and transferring the data into the cells of a worksheet. In this section we will cover some of the more common techniques for opening, closing, reading, and writing data files with VBA and Excel.

***Selecting Files to Open or Save*** The first step to accessing a file is to identify the file's name. In rare cases the file name will be fixed and can be hard-coded into a program. More commonly the user must be prompted to select or enter the file name when the VBA script runs. Excel provides two built-in functions to accomplish this; both are accessed through the Excel Application object.

The `GetSaveAsFileName` method of the Application object will display a standard Windows file dialog box and allow the user to navigate the file system and either select or enter a file name into which data can be written. `GetSaveAsFileName` takes a number of optional arguments, including `InitialFileName`, `FileFilter`, and `Title`. The `InitialFileName` argument is used to set the file name that will be displayed when the dialog is first opened. It is handy to set this to the most commonly used file name in order to save a few keystrokes. The `Title` parameter is a string that will be displayed as the caption of the dialog. It is common to set this argument to give the user of your script a hint on the kind of file that is expected. The `FileFilter` argument is a string that specifies one or more file name filters to be used by the dialog when selecting which files to be displayed. The function returns the selected file name, or if the user clicks the Cancel button, it returns False.

The subroutine in Listing 20 illustrates the use of `GetSaveAsFilename`. Note that we are using the line continuation character (_) in order to split a long command over several lines. In this example the `vrnFileName` variable is declared as type Variant. We use a Variant because the `GetSaveAsFilename` method can return multiple data types: When the user makes a selection, a String is returned representing the complete file path, otherwise the Boolean value False is returned when the dialog is cancelled.

```
Sub WriteAFile()

    Dim vrnFileName As Variant

    ' Ask the user for the file name
    vrnFileName = Application.GetSaveAsFilename( _
         InitialFileName:="data.csv", _
         FileFilter:= _
         "CSV Files (*.csv),*.csv, All Files (*.*),*.*", _
         Title:="Select Data File")
    ' Exit if the dialog is cancelled
    If vrnFileName = False Then Exit Sub

    ' Open and write the file here
End Sub
```

**Listing 20** A subroutine illustrating the `GetSaveAsFilename` method of Application.

The `GetOpenFilename` method of the Application object displays a dialog that allows the user to enter or select a file that is intended to be opened. The method is called in a manner very similar to `GetSaveAsFilename`. One difference between the two methods is that `GetOpenFilename` does not accept the `InitialFileName` argument.

***File Access Types*** In VBA, there are three ways to access data in a file. These are called sequential, random, and binary access. The main difference among these file access types is the way that data are read and written. Random and binary file access types usually are used to manipulate data stored in files in a form other than plain text. When a file is opened for random access, data must be stored in the file as fixed-size records; these records must be read and written as complete units. Binary file access permits individual bytes to be read and written. A common feature shared by these two file access types is that the data stored in the file can be accessed and manipulated in random order.

When a file is opened for sequential access, the data in the file can be read or written in a sequential manner only, that is, from the beginning of the file to the end. Unlike the other file access types, it is not possible to move randomly about a file opened for sequential access. Even though the sequential access type is much more limiting, it is by far the more common of the three access types because the sequential access type offers more convenient options for reading and writing plain text file data. We will consider only the sequential file access type in this section.

***Opening and Closing Sequential Access Files*** Before data in a file can be accessed, the file must be opened. This is accomplished with the `Open` statement in VBA. The `Open` statement takes a number of arguments, including the name of the file to open, the file mode, and a unique file number.

In VBA unique numbers (integers) are used for nearly all of its native file-related commands. These numbers can take on any value between 1 and 511 but cannot be reused for a second file until the first file is closed. File numbers are assigned when a file is opened with the `Open` statement. It is possible to specify unused file numbers for the files that you open, but it is far easier to use a built-in VBA function called `FreeFile` to select available file numbers automatically. `FreeFile` is guaranteed to return a number that is not currently being used for another file.

The sequential file access type comes in three flavors, or modes: Input, Output, and Append. When opening a file with the Input mode keyword, you only will be able to read data from the file. Opening a file with the Output mode keyword will erase the file and allow you to start writing from the beginning. The Append mode keyword will open the file for writing but not erase its contents. Instead, writing will begin immediately after the end of the file. In other words, you will be appending to the file when you write.

When one is finished reading or writing a file, it must be closed. Closing a file is simply a matter of executing the `Close` statement, optionally followed by the pound sign (#) and the file number.

Listing 21 is a complete VBA subroutine called `ReadDataFromAFile`. It illustrates all the code necessary for selecting and opening a file for input, and closing it when finished. The `intFileNum` variable is declared as an integer and is assigned a new file number using the `FreeFile` statement. The `vrnFileName` variable is declared to hold the response from the `GetOpenFilename` function. The response will be either the complete file path containing the data to be read, or False if the dialog was cancelled.

The `Open` statement takes a number of arguments. The file name is given first as the `vrnFileName` variable, the `For Input` portion of the statement indicates that the file will be read, and the file number is assigned using the `As #intFileNum` portion of the statement. The file is closed at the end of the subroutine with the `Close` statement. The specific file to be closed is indicated by specifying the `#intFileNum` argument. Leaving out the argument will cause all open files to be closed.

```
Sub ReadDataFromAFile()
    Dim intFileNum As Integer ' To hold the file number
    Dim vrnFileName As Variant ' To hold the file name

    intFileNum = FreeFile ' Get a file number

    ' Ask the user for the file name
    vrnFileName = Application.GetOpenFilename( _
         FileFilter:= _
         "CSV Files (*.csv),*.csv, All Files (*.*),*.*", _
         Title:="Read Data File")

    ' Exit if the dialog is cancelled
    If vrnFileName = False Then Exit Sub

    ' Open the file for Input
    Open vrnFileName For Input As #intFileNum

    ' Read data here

    ' Close the file before exiting
    Close #intFileNum
End Sub
```

**Listing 21** Sample code for selecting a file name, opening the file, and closing it.

This subroutine can be tested by entering it into any code window of Excel's Visual Basic Environment. To run the code, move the cursor to a location within the subroutine and click the Run toolbar button (the blue right-pointing arrow on the standard toolbar of the VBE), or press the F5 key. The code will prompt for a file name, open the selected file, close the file, and exit in an uneventful manner. If you enter a file that does not exist, the dialog will not close. VBA will not create a file automatically when opening for Input. Conversely, when you open a file using the Output or Append modes, VBA will automatically create the file if it does not exist already.

In this section we have discussed how to open and close a file using the various sequential file access modes. In the next section we will discuss how to read and write these files.

***Reading and Writing Sequential Access Files*** The most common way to write data to a file is to use the `Print #` statement. The first argument required by `Print #` is the number of the file used in the `Open` statement. Following the file number argument is a sequence of data items to be printed to the file. The number of data items is not fixed; essentially any number of items can be written with `Print #`.

Listing 22 includes a subroutine named `LogMessage`, which is a simple but powerful utility for keeping a log file. This subroutine can be very handy when there is a need to monitor the processing of a large amount of data or the progress of a complex program. Arguments passed to the `LogMessage` subroutine include the message to be written to the log file, followed by the path that identifies the location of the file. Because an `Open For Append` statement will create a file automatically if it does not exist, creating and maintaining a log file is as simple as calling the `LogMessage` subroutine.

In the body of `LogMessage`, the log file is opened, the message is written, and the file is closed. Because the file is opened For Append, the message will be added to the end of the file without erasing previously written data contained in the file. Note that the actual string written to the file is more than just the message that was passed to the subroutine. Concatenated to the front of the message is the current date and time followed by a colon.

The second subroutine in Listing 22 called `TestLog` simply shows how to call `LogMessage`. To run a test, enter both subroutines into any code module in the VBE, move your cursor to a point within the `TestLog` function and hit the F5 key. Check for the log.txt file at the specified location on your hard drive.

```
Sub LogMessage(strMessage As String, _
                strLogFile As String)

    Dim intFileNum As Integer

    intFileNum = FreeFile
    Open strLogFile For Append As #intFileNum
    Print #intFileNum, Now & ": " & strMessage
    Close #intFileNum

End Sub

Sub TestLog()
    LogMessage "Computation complete", "C:\log.txt"
End Sub
```

**Listing 22** A simple subroutine for logging to a file, and a subroutine to test it.

Another statement that can be used to write data to a file is the `Write #` statement. Arguments expected by `Write #` are identical to `Print #`. The main difference between these two statements is that `Write #` will always write data to a file in a format that guarantees the data can be read in again by VBA. For example, all strings written using `Write #` are surrounded with double quotes (“), and dates are surrounded with hash symbols (#). Sometimes, this behavior is exactly what we want. Other times, when the file is to be read by humans, or another program that does not understand these special symbols, using `Write #` will only introduce unnecessary complexity.

```
Sub TestWriteStatement()
    Dim vrnFileName As Variant    ' File name
    Dim intFileNum As Integer     ' File number
    Dim strWell As String         ' Name of the well
    Dim blnRgntAdded As Boolean   ' True if some
                                  ' reagent was added
    Dim dblMeasurement As Double  ' Measurement taken
    Dim datTime As Date           ' Date-time of
                                  ' measurement

    strWell = "A01" ' Assign test data

    blnRgntAdded = True
    dblMeasurement = 1.234
    datTime = Now
```

**Listing 23** Testing the `Write #` statement.

```
    ' Get file name
    vrnFileName = Application.GetSaveAsFilename( _
         InitialFileName:="test.txt", _
         FileFilter:= _
         "Text Files (*.txt),*.txt,All Files (*.*),*.*", _
         Title:="Test Write Statement")

    ' Exit if cancelled
    If vrnFileName = False Then Exit Sub

    intFileNum = FreeFile ' Set file number

    ' Write data to the file
    Open vrnFileName For Output As #intFileNum
    Write #intFileNum, strWell, blnRgntAdded, _
         dblMeasurement, datTime

    Close #intFileNum ' Close the file
End Sub
```

**Listing 23** *Continued*

After running the `TestWriteStatement` subroutine in Listing 23, the contents of the newly written file will look similar to the following:

```
"A01",#TRUE#,1.234,#2004-10-24 08:12:50#
```

The complement to `Write #` is the `Input #` statement. In addition to the file number, the `Input #` statement takes a list of variables that will be assigned to the values that are read from the file. The data in the file being read must be in the proper format—the same format used by the `Write #` statement.

Enter the code in Listing 24 into a code module in the VBE. Before running it, make sure your Immediate Window is displayed by selecting the View | Immediate Window menu option in the VBE. This is necessary in order to see the output that will be printed by the `Debug.Print` statement, which is included in the listing. When ready, run your subroutine as usual. Select the same file that was written by `TestWriteStatement` subroutine. The data written to that file will be read and printed in the Immediate Window.

```
Sub TestInputStatement()
    Dim vrnFileName As Variant   ' File name
    Dim intFileNum As Integer    ' File number
    Dim strWell As String        ' Name of the well
    Dim blnRgntAdded As Boolean  ' True if some
                                 ' reagent was added
```

**Listing 24** Testing the `Input #` statement.

```
    Dim dblMeasurement As Double ' Measurement taken
    Dim datTime As Date          ' Date-time of
                                 ' measurement

    ' Get file name
    vrnFileName = Application.GetOpenFilename( _
         FileFilter:= _
         "Text Files (*.txt),*.txt, All Files (*.*),*.*", _
         Title:="Test Input Statement")

    ' Exit if cancelled
    If vrnFileName = False Then Exit Sub

    ' Set file number and open
    intFileNum = FreeFile
    Open vrnFileName For Input As #intFileNum

    ' Read data, close file and print
    Input #intFileNum, strWell, blnRgntAdded, _
         dblMeasurement, datTime
    Close #intFileNum
    Debug.Print strWell, blnRgntAdded, _
         dblMeasurement, datTime
End Sub
```

**Listing 24** *Continued*

The code in Listing 23 and Listing 24 write and read a single line of four data items. In practice, it is more likely that you will write many lines of data using a loop construct. For example, you may need to write or read 96 lines, one for each well of a microtiter plate.

When data are written to a file in a format that is not easily read using the `Input #` statement, a little more work is required. The entire line of data must be read as a single string first, and then the individual data items extracted from the string. We will discuss string manipulation in the next section. For now, let us look at how an entire line of data can be read using the `Line Input #` statement. Listing 25 illustrates how to use `Line Input #`. It is similar to `Input #`, with only a single string argument to hold the data that is read.

If you have a file named C:\test2.txt that contains the following line of data:

```
1.23; 3.45; 5.67; 7.89
```

running the code in Listing 25 will print that line to the Immediate Window exactly as it appears in the file. The data values, which are separated by semicolons, are not assigned to individual variables, as was the case with `Input #`

in Listing 24. Since the `Input #` statement expects data values to be separated by commas, using it on this file would result in a runtime error. Instead, the entire string of characters on the first line is read and assigned to the `strLine` string variable:

```
Sub TestLineInputStatement()
    Dim strFileName As String ' File name
    Dim intFileNum As Integer ' File number
    Dim strLine As String     ' Entire line of
                              ' data

    strFileName = "C:\test2.txt" ' Set file name

    ' Set file number and open file
    intFileNum = FreeFile
    Open strFileName For Input As #intFileNum

    'Read the entire line and close the file
    Line Input #intFileNum, strLine
    Close #intFileNum

    ' Print in the Immediate window
    Debug.Print strLine

End Sub
```

**Listing 25 Reading an entire line from a file using the `Line Input #` statement.**

***Manipulating and Parsing Data from Strings*** To extract data from strings that are read from a file, the characters that represent the data items must be dissected from the string and possibly converted to the proper VBA data types, for example, integers or double precision numbers. This dissection process is called *parsing*. VBA includes many functions for manipulating strings, which can be used to parse the desired data from a string. Table 2 lists several VBA functions that are useful for parsing strings. A short description is included with each function. For more detailed information, refer to the VBA Help documentation, which can be displayed from the Help menu in the VBE.

***An Example*** Let us consider a realistic example. Assume we have a file, the first few lines of which appear as in Figure 7. This is a sample data file written by an instrument that takes a single measurement from each well of a microtiter plate. Unfortunately, due to the way in which the instrument operates, the file starts with an unknown number of blank lines. The actual plate data section is initiated with a line that begins with the string "Plate:", and is followed by a bar code that has been assigned to the plate. The well

**TABLE 2 VBA Functions Useful for Parsing Strings**

| Function | Description |
|---|---|
| InStr | Search one string looking for a specified substring contained within it. If found, return the position of the substring. |
| InStrRev | Similar to InStr, only search for the substring starting from the end of the string and proceeding back to the beginning. |
| Replace | Search a string, replacing all occurrences of one substring with another substring. Return the substituted string. |
| Split | Split a string on a specified delimiting substring, returning an array of delimited strings. |
| Join | Join an array of strings and return a single string. Optionally insert a specified delimiter between each string element in the joined string. |
| Trim | Remove all spaces from the beginning and end of a string and return the result. |
| Left | Return a substring composed of the specified number of characters at beginning of a string. |
| Right | Return a substring composed of the specified number of characters at end of a string. |
| Mid | Return a substring composed of a specified number of characters starting from a specified position within a string. |
| IsNumeric | Returns True if the argument (possibly a string) can be converted to a number. |
| CInt | Convert the argument (possibly a string) to an integer. |
| CLong | Convert the argument (possibly a string) to a long integer. |
| CDbl | Convert the argument (possibly a string) to a double precision number. |
| CDate | Convert the argument (possibly a string) to a date. |

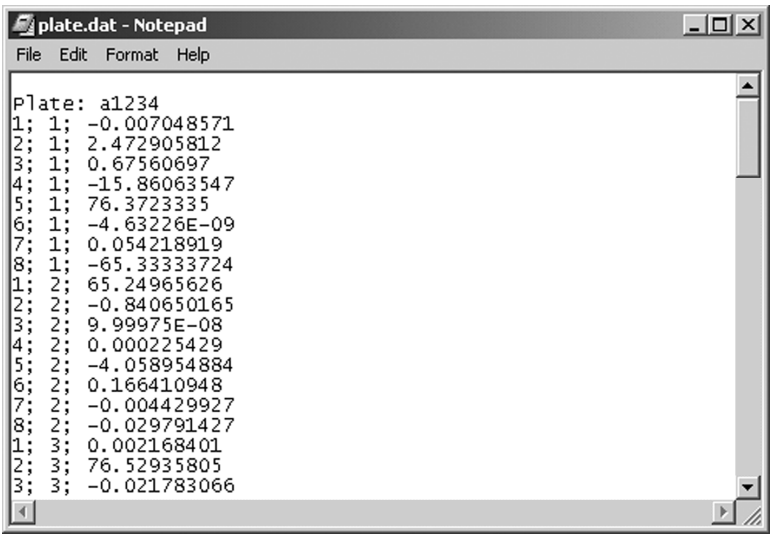


**Figure 7** Sample Plate Data File.

measurements are formatted as three columns. The first column corresponds to the well row, the second column to the well column, and the last column to the measured well data. The challenge is to read the file and parse the data into the cells of an Excel spreadsheet. The data should be displayed in the spreadsheet in an 8 by 12 block of cells rather than in columns. The layout should correspond to that of a microtiter plate in a landscape orientation. In the following example we tackle the problem using the following three steps:

1. Finding the data section in the file
2. Reading and parsing the data from the file
3. Converting the strings to numbers.

In the next section we'll discuss how to transfer the data to the cells of a worksheet, and to format the worksheet to enhance the data presentation.

Listing 26 provides a complete subroutine for parsing, converting, and displaying all data in the Immediate Window. Subroutine comments help identify exactly what is being done. Note the use of several functions listed in Table 2.

After variables are declared, a file is selected and opened. First, all blank lines are read, trimmed of white space, tested, and discarded. The first non-blank line is checked for the proper format, and the bar code is removed using the `Mid` and `Trim` functions. Immediately following, 96 lines of data are read and split on the semicolon character into the `arrLine` string array using the `Split` function. The row and column designation are converted to Integers using the `CInt` conversion function. The measurements are converted to Doubles using the `CDbl` function and stored in the proper location of the `arrData` array. The subroutine completes with a couple of nested loops that print the data in the array to the Immediate Window.

```
Sub ParsePlateData1()
    ' Parse data into an array
    Dim vrnFileName As Variant ' File name
    Dim intFileNum As Integer  ' File number
    Dim strLine As String      ' Line read
                               ' from the file
    Dim strBarcode As String   ' Parsed bar code
    Dim arrLine() As String    ' Array to hold result
                               ' of Split command
    Dim intRow As Integer      ' Current row
    Dim intCol As Integer      ' Current column
    ' Array to hold parsed plate data
    Dim arrData(1 To 8, 1 To 12) As Double
```

**Listing 26** Parse plate data from a file.

```
' Loop counters
Dim I As Integer, J As Integer

' Select the file
vrnFileName = Application.GetOpenFilename( _
     FileFilter:= _
     "Text Files (*.dat),*.dat, All Files (*.*),*.*", _
     Title:="Select Data File")
' Exit if cancelled
If vrnFileName = False Then Exit Sub
' Get file number and open
intFileNum = FreeFile
Open vrnFileName For Input As #intFileNum

Line Input #intFileNum, strLine ' Read line
strLine = Trim(strLine) ' Trim whitespace

' Discard any blank lines
Do While strLine = ""
     Line Input #intFileNum, strLine
     strLine = Trim(strLine)
Loop

' Parse barcode
If Left(strLine, 6) = "Plate:" Then
     strBarcode = Trim(Mid(strLine, 7))
Else
     Debug.Print "Invalid file format"
     Exit Sub
End If

' Parse data
For I = 1 To 96 ' Loop over all Lines
     ' Read line
     Line Input #intFileNum, strLine
     ' Split on semicolon
     arrLine = Split(strLine, ";")
     ' Convert row ...
     intRow = CInt(arrLine(0))
     ' Convert column ...
     intCol = CInt(arrLine(1))
```

**Listing 26** *Continued*

```
            ' Convert well data ...
            arrData(intRow, intCol) = _
                CDbl(arrLine(2))
        Next I

        Close #intFileNum ' Close file

        For I = 1 To 8 ' Print data
            For J = 1 To 11
                Debug.Print arrData(I, J);
            Next J
            Debug.Print arrData(I, J)
        Next I

    End Sub
```

**Listing 26** *Continued*

## Importing Data from External Databases

Another way to move data into Excel is to import it from an external database. A common practice of an organization is to store large amounts of diverse data in centralized databases. These data can form the basis of calculations that you may need to perform. Fortunately, Excel makes it easy to import data into a Worksheet from a centralized database.

***Understanding Databases*** Databases come in many flavors and speak many languages. Perhaps the most common database type is the relational database. Microsoft Access, SQL Server, Oracle, and other databases are examples. A relational database is a collection of tables, where each table consists of a two-dimensional array of data organized as rows (records) and columns (fields). Tables 3 and 4 present data that make up two sample tables in a hypothetical relational database. The sample tables are named Compound_Data and Inventory_Data, respectively. Compound_Data holds data that describe compounds in a compound store. Inventory_Data holds data that describe the current inventory of each compound.

The two sample tables have a common field called Compound_ID. The data in this field are used to relate the records in the two tables, hence the name *relational* database. It is possible to query the tables in this sample database in order to select all inventory information from compounds with a molecular weight greater than 500. This requires an inspection of data in both tables, which is accomplished by *joining* the two tables on the common Compound_ID field values.

Another way to get a sense for the important elements of a relational database is to compare them to similar elements in Excel. Table 5 presents such

**TABLE 3 Sample Table Named Compound_Data**

| Compound_ID | Name | MW |
|---|---|---|
| 100001 | Compound1 | 504.07 |
| 100002 | Compound2 | 576.81 |
| 100003 | Compound3 | 331.49 |
| 100004 | Compound4 | 371.09 |
| 100005 | Compound5 | 454.52 |
| 100006 | Compound6 | 488.36 |
| 100007 | Compound7 | 405.04 |
| 100008 | Compound8 | 568.20 |

**TABLE 4 Sample Table Named Inventory_Data**

| Compound_ID | Barcode | Tare_Weight | Current_Weight |
|---|---|---|---|
| 100001 | 9999901 | 8.01 | 27.74 |
| 100001 | 9999902 | 8.57 | 14.90 |
| 100001 | 9999903 | 8.27 | 15.84 |
| 100002 | 9999904 | 8.11 | 28.02 |
| 100003 | 9999905 | 8.03 | 27.25 |
| 100004 | 9999906 | 8.82 | 11.05 |
| 100005 | 9999907 | 8.74 | 18.37 |
| 100006 | 9999908 | 8.74 | 17.48 |
| 100006 | 9999909 | 8.91 | 14.42 |
| 100007 | 9999910 | 8.03 | 21.05 |

**TABLE 5 Similar Excel and Relational Database Elements**

| Excel | Relational Database |
|---|---|
| Workbook | Database files |
| Worksheet | Table |
| Column | Field |
| Row | Record |

a comparison. Even though we can make this comparison, it is important to remember that Excel is not relational in nature; it is not possible to join data in the columns of Worksheets using standard Excel capabilities.

***Accessing Databases using Excel and ODBC*** The multitude of languages spoken by the various database engines can make importing data a complex endeavor. Fortunately, there is a common language through which many database engines can be accessed. This common language is called Open DataBase Connectivity (ODBC). An ODBC driver is software that accepts a command

in the common ODBC language and translates it into a database-specific form. Data obtained from a database-specific command is repackaged and returned in a common format. Almost all major relational database engines provide an ODBC driver. In fact, there are even ODBC drivers that expose data in properly formatted nonrelational databases, such as Excel Workbooks and flat files.

Early in this chapter, we stated that Excel is not a database management system. We used this statement as an argument against using Excel in circumstance that call for a database management system. Nevertheless, for illustration purposes we will show how to set up Excel as a simple database and access it through an ODBC connection. Due to the common language of ODBC, the method for importing data is identical for more powerful relational database engines.

Creating an Excel database is simple. As an example, open up a new Excel Workbook and rename one Worksheet by double clicking on its tab and entering Compound_Data. Enter data into the cells of this Worksheet, as shown in Figure 8. Do not forget to enter the field names in the first row. Rename another Worksheet to Inventory_Data and enter the data shown in Figure 9. Save the Workbook with the file name InvDatabase.xls.

The only remaining step necessary to turn the data in this Workbook into an ODBC-accessible database is to name the ranges in each Worksheet that contain table data. Highlight the range of cells containing data to be treated as a single table and enter the table name in the Name box, which is just to the left of the formula bar above the Worksheet area (See the circled element in Fig. 10). Once the name is entered, make sure to hit the Enter key, otherwise the name will not be assigned to the range. Set the name of the data range on the first sheet to Compound_Data, and the data range on the second sheet to Inventory_Data. Resave and close the modified Workbook.

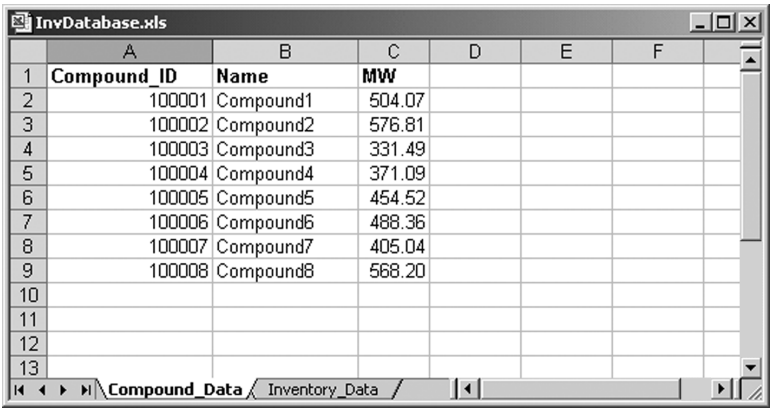

InvDatabase.xls

| | A | B | C | D | E | F |
|---|---|---|---|---|---|---|
| 1 | Compound_ID | Name | MW | | | |
| 2 | 100001 | Compound1 | 504.07 | | | |
| 3 | 100002 | Compound2 | 576.81 | | | |
| 4 | 100003 | Compound3 | 331.49 | | | |
| 5 | 100004 | Compound4 | 371.09 | | | |
| 6 | 100005 | Compound5 | 454.52 | | | |
| 7 | 100006 | Compound6 | 488.36 | | | |
| 8 | 100007 | Compound7 | 405.04 | | | |
| 9 | 100008 | Compound8 | 568.20 | | | |
| 10 | | | | | | |
| 11 | | | | | | |
| 12 | | | | | | |
| 13 | | | | | | |

Compound_Data / Inventory_Data

**Figure 8** Compound_Data worksheet.

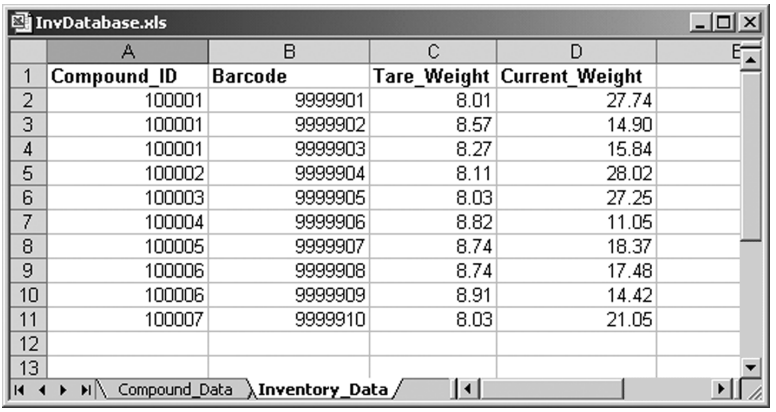

| | A | B | C | D |
|---|---|---|---|---|
| 1 | Compound_ID | Barcode | Tare_Weight | Current_Weight |
| 2 | 100001 | 9999901 | 8.01 | 27.74 |
| 3 | 100001 | 9999902 | 8.57 | 14.90 |
| 4 | 100001 | 9999903 | 8.27 | 15.84 |
| 5 | 100002 | 9999904 | 8.11 | 28.02 |
| 6 | 100003 | 9999905 | 8.03 | 27.25 |
| 7 | 100004 | 9999906 | 8.82 | 11.05 |
| 8 | 100005 | 9999907 | 8.74 | 18.37 |
| 9 | 100006 | 9999908 | 8.74 | 17.48 |
| 10 | 100006 | 9999909 | 8.91 | 14.42 |
| 11 | 100007 | 9999910 | 8.03 | 21.05 |
| 12 | | | | |
| 13 | | | | |

**Figure 9** Inventory Data worksheet.

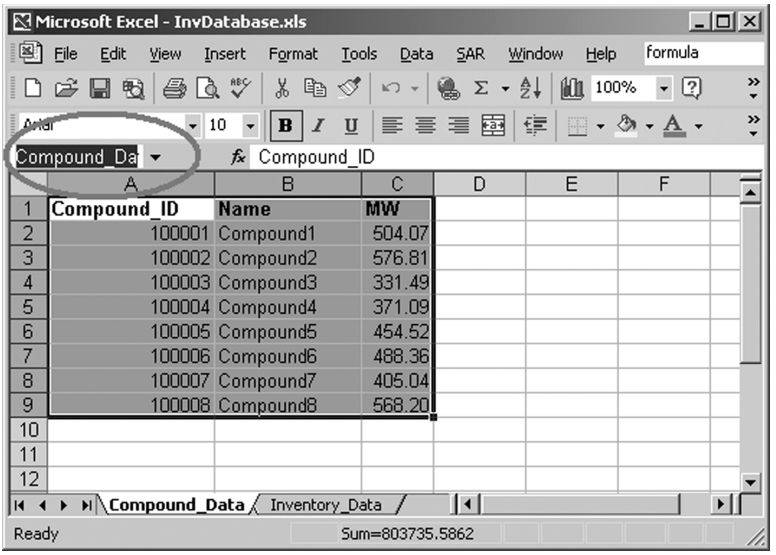

| | A | B | C |
|---|---|---|---|
| 1 | Compound_ID | Name | MW |
| 2 | 100001 | Compound1 | 504.07 |
| 3 | 100002 | Compound2 | 576.81 |
| 4 | 100003 | Compound3 | 331.49 |
| 5 | 100004 | Compound4 | 371.09 |
| 6 | 100005 | Compound5 | 454.52 |
| 7 | 100006 | Compound6 | 488.36 |
| 8 | 100007 | Compound7 | 405.04 |
| 9 | 100008 | Compound8 | 568.20 |
| 10 | | | |
| 11 | | | |
| 12 | | | |

**Figure 10** Creating an ODBC-accessible table in Excel by naming a range of cells.

***Importing Data into Excel*** Data can be imported right into an Excel Worksheet from an external database using menu options. To see this in action, open a new Workbook and select the Data | Import External Data | Import Data . . . menu option. You will see the Select Data Source dialog, similar to that shown in Figure 11. The dialog will show all the available data sources, including *Data Source Names* (DSN). A DSN contains all the information necessary for Excel to make an ODBC connection with an external data source.

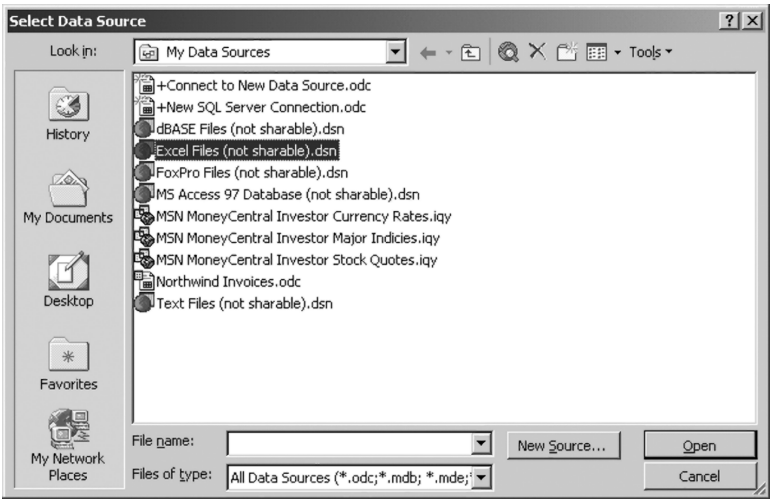


**Figure 11** Select Data Source Dialog.

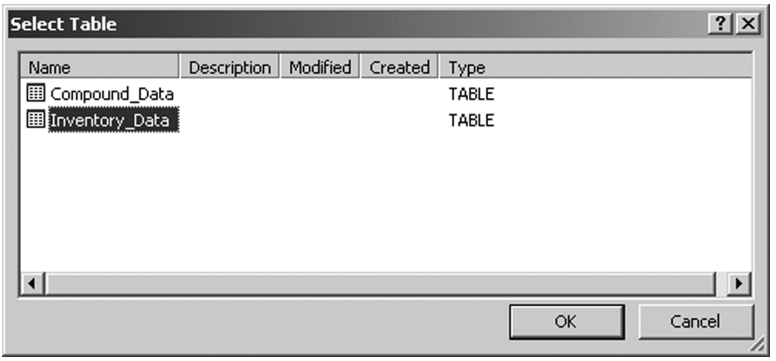


**Figure 12** Tables in the InvDatabase.xls Excel database.

Excel is distributed with several preconfigured DSNs, including one for Excel itself.

Select the Excel Files DSN and click the Open button. Navigate to and select the InvDatabase.xls Excel file, and click OK. You should see a dialog that shows all tables in the Excel database (Fig. 12). Select one of the tables and click OK to view the Import Data dialog. If you click OK again, all the data from this table will be imported into your new Workbook. Instead, if you click the Edit Query. . . button from the Import Data dialog, Excel will display the Query Wizard from a helper application called Microsoft Query. The Microsoft Excel installation package includes Microsoft Query, but it is not installed by default. You may need to install this application from your Excel installation CDs before continuing.

Follow the Query Wizard prompts. Select different fields from the available tables and filter the records to match your specified criteria. In Figure 13 we are restricting records from the Inventory_Data table to those with a Compound_ID field value of 100001. After proceeding through all the remaining Query Wizard prompts, the filtered records can be returned to your Excel Worksheet, as shown in Figure 14.

In fact, using the Query Wizard, it is possible to join data in multiple tables of the external Excel database, even though Excel is not relational. It is the Excel ODBC driver that imparts the relational capabilities on top of the Excel database—truly a nice added benefit.

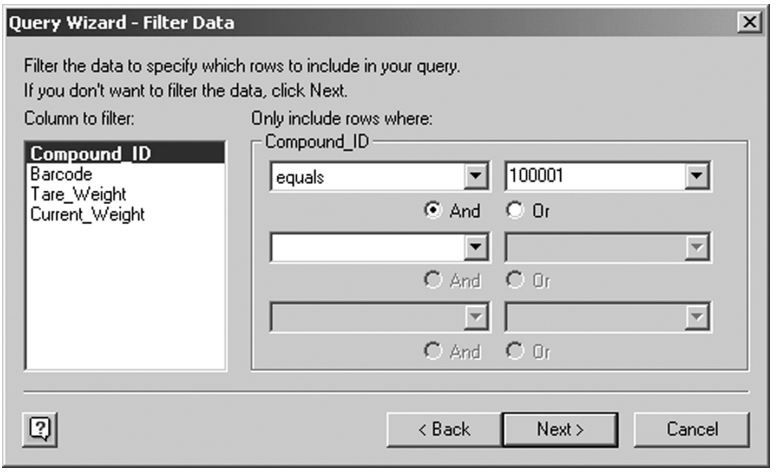


**Figure 13** Filtering records of the Inventory_Data table to those with a Compound_ID of 100001.

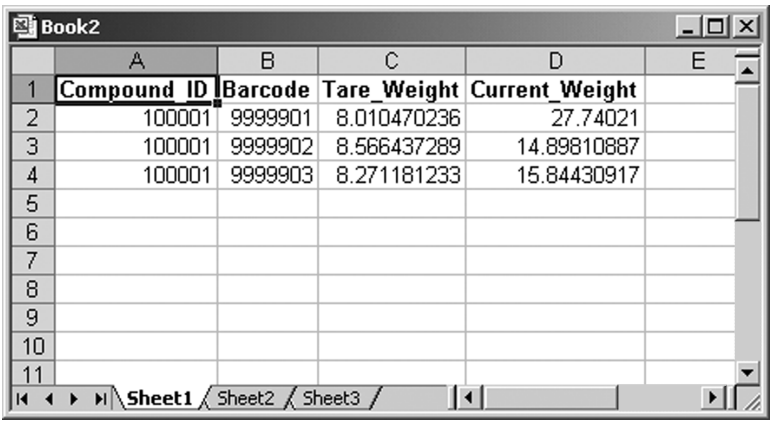

Book2

| | A | B | C | D | E |
|---|---|---|---|---|---|
| 1 | Compound_ID | Barcode | Tare_Weight | Current_Weight | |
| 2 | 100001 | 9999901 | 8.010470236 | 27.74021 | |
| 3 | 100001 | 9999902 | 8.566437289 | 14.89810887 | |
| 4 | 100001 | 9999903 | 8.271181233 | 15.84430917 | |
| 5 | | | | | |
| 6 | | | | | |
| 7 | | | | | |
| 8 | | | | | |
| 9 | | | | | |
| 10 | | | | | |
| 11 | | | | | |

Sheet1 Sheet2 Sheet3

**Figure 14** Records returned from the external database import.

***Importing Data from an External Database with VBA*** We can make use of the Macro Recorder to learn how to automate the external data import process using VBA. By repeating the process of the previous section with the Macro Recorder activated, you will generate a subroutine that looks like the one in Listing 27. This listing has been edited slightly to make it more readable.

Listing 27 shows that three important strings were generated: a connection string, a path to a DSN, and a Structured Query Language (SQL) query. The connection string includes the pertinent information to locate the database. In this case, it is our InvDatabase.xls file stored in the root of the C drive. The DSN string is a path to the appropriate DSN file. In this case it points to the Excel File DSN installed with Excel.

The last string is the SQL query that tells the ODBC driver how to remove the data from the database. SQL is a standard language for manipulating data in a relational database. There are four basic statements used to manipulate data in a database using SQL.

SELECT—retrieves data
UPDATE—modifies existing data
INSERT—adds data
DELETE—removes data

The particular SQL query generated in this example is a SELECT statement. The basic syntax for a SELECT statement is as follows:

SELECT [Field Name(s)] FROM [Table Name(s)] WHERE [Criteria]

This requests that the data in [Field Name(s)] be returned from the tables in [Table Name(s)], subject to the conditions in [Criteria]. It is beyond the scope of this chapter to describe SQL in detail. A search of the Internet will reveal many good sources of information about SQL. Modifications of the SQL query in Listing 27 will allow you to customize the data import process, and rerun it as necessary.

```
Sub Macro1()
    Dim strConn As String
    Dim strSQL As String
    Dim strDSN As String

    ' Connection string
    strConn = _
        "ODBC;DSN=Excel Files;DBQ=C:\InvDatabase.xls;"
```

**Listing 27** An automatically generated VBA Macro to select data from an external database.

```
    strConn = strConn & _
         "DefaultDir=C:\;DriverId=790;"
    strConn = strConn & _
         "MaxBufferSize=2048;PageTimeout=5;"

    ' DSN File
    strDSN = _
         "C:\Program Files\Common Files\ODBC\Data Sources\"
    strDSN = strDSN & "Excel Files (not sharable).dsn"

    ' SQL Query
    strSQL = "SELECT Inventory_Data.Compound_ID, "
    strSQL = strSQL & "Inventory_Data.Barcode, "
    strSQL = strSQL & "Inventory_Data.Tare_Weight, "
    strSQL = strSQL & "Inventory_Data.Current_Weight "
    strSQL = strSQL & "FROM `C:\InvDatabase`.Inventory_Data "
    strSQL = strSQL & "Inventory_Data "
    strSQL = strSQL & _
         "WHERE (Inventory_Data.Compound_ID=100001)"

    With ActiveSheet.QueryTables.Add( _
          Connection:=strConn, Destination:=Range("A1"))
         .CommandText = Array(strSQL)
         .Name = "Excel Files (not sharable)"
         .FieldNames = True
         .RowNumbers = False
         .FillAdjacentFormulas = False
         .PreserveFormatting = True
         .RefreshOnFileOpen = False
         .BackgroundQuery = True
         .RefreshStyle = xlInsertDeleteCells
         .SavePassword = True
         .SaveData = True
         .AdjustColumnWidth = True
         .RefreshPeriod = 0
         .PreserveColumnInfo = True
         .SourceConnectionFile = strDSN
         .Refresh BackgroundQuery:=False
    End With
End Sub
```

**Listing 27** *Continued*

***Setting up an Oracle ODBC connection*** As mentioned, ODBC can be used to access other databases, such a Microsoft Access or Oracle. It is only a matter of having the appropriate drivers installed and setting up the proper DSN. The following shows how to setup an ODBC connection for Oracle.

From the Windows Task Bar, select Start | Settings | Control Panel | Administrative Tools | Data Sources (ODBC). Click the Add . . . button and select an ODBC driver. In this example we will use the driver named Microsoft ODBC for Oracle (Fig. 15). Click Finish and complete the installation.

After this step you have the option to create an associated DSN. In the Create New Data Source dialog, enter a name, a description, your user name, and the Oracle server information (Fig. 16). If you are unsure of the proper parameters, try contacting your database administrator.

From this point you can access data in the Oracle database in a manner identical to the way in which data was accessed in the external Excel database. Make sure to use the Macro Recorder to regenerate the VBA in order to capture the proper connection string and DSN.

## Automating the Generation of Reports

Reports are generated in Excel by inserting data into the cells of a Worksheet and formatting the cells to create an informative and pleasing presentation. In a previous section we saw how to read and write the value of a cell using the Range object's `Value` property. The next step is to take a closer look at

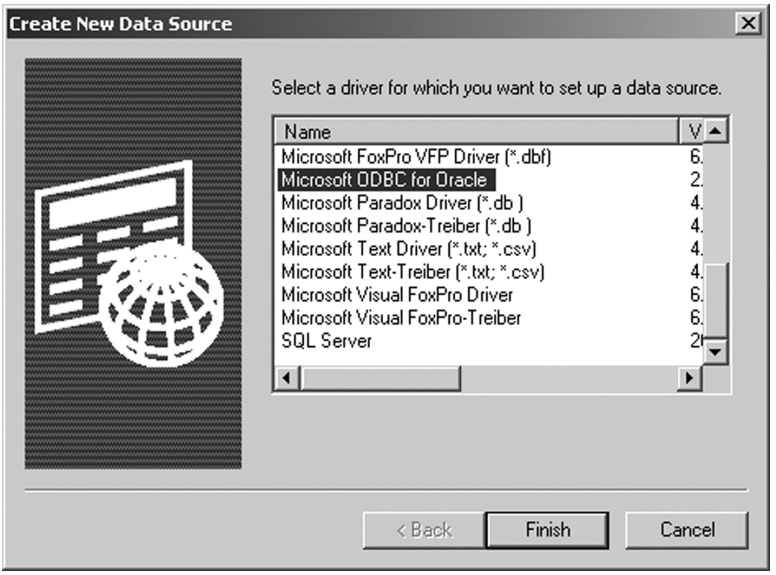


**Figure 15** Creating a new ODBC Data Source for Oracle.

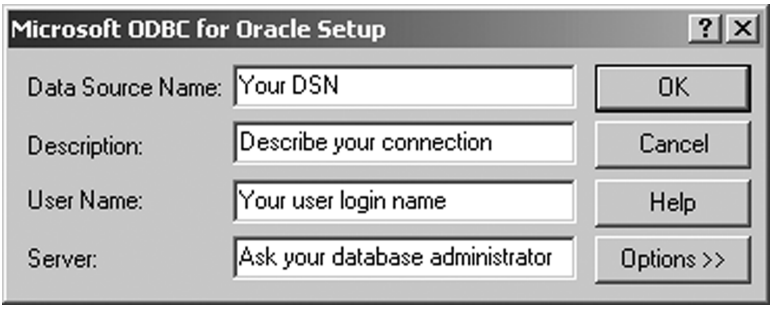


**Figure 16** Configuring the Oracle DSN.

how to change a cell's appearance, and how to use that to convey important information about the data being reported.

***Formatting Worksheets*** Everything about the way a Worksheet is formatted can be changed programmatically using VBA. This includes the way cell values are formatted, their alignment, and the font used, as well as cell borders and colors. This topic is much too large to cover in any detail here. Instead, we will investigate a few useful topics, and recommend that the interested reader make use of the Macro Recorder as a way to discovery how to programmatically change the appearance of a Worksheet.

When dealing with data that are associated with a microtiter plate or rack of tubes, it helps to display those data in a format that mimics the layout of the plate of tubes. Such displays are natural in Excel since the rectangular array of containers maps nicely onto a rectangular array of cells. The display can be enhanced by formatting the cells that delineate the plate or rack data, and marking the rows and columns with appropriate labels.

We are interested in formatting an 8 by 12 range of cells to hold an array of data associated with the wells of a 96-well microtiter plate. The goal is to outline the range with a thick border, to label columns above with the numbers "1" through "12" centered on the column, and the rows on the left with the letters "A" through "H," right justified. We want to encapsulate the formatting commands in a separate VBA subroutine so that it is possible to format any 8 by 12 Range as a plate. By taking this approach, we will be able to reuse the formatting function on any Range of cells. To get started, a macro was recorded while formatting a range of cells in accordance with our specifications. When finished, the generated subroutine was edited to make it more efficient and reusable—Listing 28 is the result.

Unlike a generated macro, first note that the subroutine in Listing 28 takes a Range object as an argument. This was added to the generated macro and represents the Range to be formatted. We make use of the With statement in the subroutine since all commands will be applied to the same Range.

The first formatting step is on the borders. Notice that a Range has a subordinate collection of Border objects appropriately called Borders. Each Border object itself has `LineStyle` and `Weight` properties. Each of the four Border objects in the `Borders` collection is accessed using a predefined constant, then their `LineStyle` and `Weight` properties are set appropriately. Note that we did not have to look up these objects, properties, or constants, since the Macro Recorder generated the code for us with all the appropriate keywords.

Following border formatting, we set the format of the numbers displayed within the Range. The Range object has a `NumberFormat` property, which takes a string that defines the format. Again, the Macro Recorder identified the property to set and its proper value. We only added a little editing.

Setting up column and row labels took a bit more work. The Macro Recorder gave us the `HorizontalAlignment` property and constants, but we added the loops and cell access using the `Cells` method of the Range object. One thing to note about the indexes of the `Cells` method is the ability to reach cells that fall outside the Range. The `FormatPlateDisplay` subroutine takes a Range argument that defines the 8 by 12 array of cells. To generate column and row labels we need to format cells in the row above the Range as well as cells in the column to the left of the Range. The column label format loop uses the loop index to set column values in the `Cells` method, but the row index is set to "0." The `Cells` method provides access to the cells in the parent Range starting at index "1." An index of "0" can be used to access cells just prior to the parent Range. In the column label format loop the row index of "0" provides access to the row above the parent Range that defines the plate display. Similarly, the row label format loop uses a column index of "0," which provides access to the column to the left of the parent Range. The final trick to note is the use of the `Chr$` function in the row label format loop. `Chr$` takes an integer and returns the corresponding ASCII encoded character. A capital "A" has an ASCII value of 65, with subsequent capital letters following in numeric order. We use this fact to generate the letters "A" through "H" by adding 64 to the row number, and passing this to the `Chr$` function.

To test the `FormatPlateDisplay` function we called it with an argument of `Sheet1.Range("B2:M9")`. The result is displayed in Figure 17.

```
Sub FormatPlateDisplay(Plate As Range)
    ' Format an 8x12 Range as a plate display
    Dim I As Integer

    With Plate

        ' Draw a thick border around the plate Range
        .Borders(xlEdgeLeft).LineStyle = xlContinuous
```

**Listing 28** Subroutine to format a Range as a plate display.

```
        .Borders(xlEdgeLeft).Weight = xlThick
        .Borders(xlEdgeTop).LineStyle = xlContinuous
        .Borders(xlEdgeTop).Weight = xlThick
        .Borders(xlEdgeBottom).LineStyle = xlContinuous
        .Borders(xlEdgeBottom).Weight = xlThick
        .Borders(xlEdgeRight).LineStyle = xlContinuous
        .Borders(xlEdgeRight).Weight = xlThick

        ' Set number format for all cells in Range
        .NumberFormat = "0.00"

        ' Set column labels
        For I = 1 To 12
            .Cells(0, I).Value = I
            .Cells(0, I).HorizontalAlignment = xlCenter
        Next I

        ' Set row labels
        For I = 1 To 8
            .Cells(I, 0).Value = Chr$(I + 64)
            .Cells(I, 0).HorizontalAlignment = xlRight
        Next I

    End With

End Sub
```

**Listing 28** *Continued*

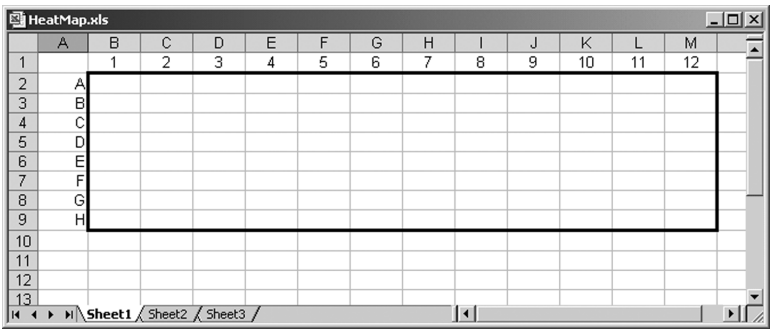


**Figure 17** Result from calling the FormatPlateDisplay subroutine on Range B2:M9.

***Generating a Heat Map Display*** One simple but powerful way to display data is through the use of a "heat map." In a heat map display, data values are rendered graphically using varying shades of two colors. The lower half of a numerical range is represented using shades that vary from a "cool" color to a neutral one, and the upper half of the range from the neutral color to "hot" color. In a sense, the colors indicate a level of heat associated with each data value. We will generate heat maps that use blue as the cool color, white as the neutral color, and red as the hot color.

To generate a heat map in Excel, the problem that we must solve is one of changing the background color of a cell based on the cell's value. It is possible do this in a rudimentary manner through the use of a feature in Excel called Conditional Formatting. Conditional Formats can be set for each cell through Excel's standard menu options. A maximum of three formats can be specified along with the conditions that must be satisfied in order for the associated format to be applied to the cell. Unfortunately, this is insufficient for generating a heat map because many more color shades and conditions can be required. VBA must be used to achieve the enhanced functionality necessary for generating a heat map.

Once again, by using the Macro Recorder to change a cell's background color, we learn that a Range object contains another object called `Interior`, which itself has a property called `ColorIndex`. Excel has a set of colors that can be used on various objects throughout its display. These colors are identified with an integer called a `ColorIndex`. Setting the `ColorIndex` property of the Range's `Interior` object will cause the background color of a cell to change to the one identified by the index value. With this in mind, we selected color indexes for five shades of blue, five shades of red, and white.

The HeatMap subroutine in Listing 29 is a complete procedure for inserting data values into the cells of a worksheet and overlying that with a heat map display. The subroutine takes two arguments: a Range object that represents the 8 by 12 block of cells on which to create the heat map and an 8 by 12 array of data values.

The first step of the `HeatMap` subroutine is to build an array of the color indexes identified for the heat map display. These color indexes are stored in an array called `arrClrIdx`. The color index to be used for a cell is determined by translating a cell's data value into an array index of the `arrClrIdx` array. The next two steps of the `HeatMap` subroutine are to set the data range that maps into the color gradient, and to call the `FormatPlateDisplay` subroutine given in Listing 28 in order to initialize the plate display. Following all initialization, a doubly nested loop over all data in the provided data array inserts the raw data value into the associated cell and then maps the data value into an index in the color index array, and uses that index to set the background color of the cell using the ColorIndex property of the Cell's Interior object.

In Listing 26 we parsed data from a file and loaded it into an array. The loaded data were printed to the Immediate Window. We can now modify this

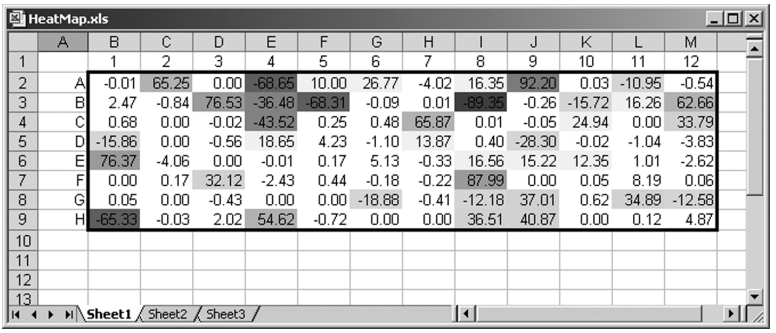

HeatMap.xls

| | A | B | C | D | E | F | G | H | I | J | K | L | M |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| 2 | A | -0.01 | 65.25 | 0.00 | -68.65 | 10.00 | 26.77 | -4.02 | 16.35 | 92.20 | 0.03 | -10.95 | -0.54 |
| 3 | B | 2.47 | -0.84 | 76.53 | -36.48 | -68.31 | -0.09 | 0.01 | -89.35 | -0.26 | -15.72 | 16.26 | 62.66 |
| 4 | C | 0.68 | 0.00 | -0.02 | -43.52 | 0.25 | 0.48 | 65.87 | 0.01 | -0.05 | 24.94 | 0.00 | 33.79 |
| 5 | D | -15.86 | 0.00 | -0.56 | 18.65 | 4.23 | -1.10 | 13.87 | 0.40 | -28.30 | -0.02 | -1.04 | -3.83 |
| 6 | E | 76.37 | -4.06 | 0.00 | -0.01 | 0.17 | 5.13 | -0.33 | 16.56 | 15.22 | 12.35 | 1.01 | -2.62 |
| 7 | F | 0.00 | 0.17 | 32.12 | -2.43 | 0.44 | -0.18 | -0.22 | 87.99 | 0.00 | 0.05 | 8.19 | 0.06 |
| 8 | G | 0.05 | 0.00 | -0.43 | 0.00 | 0.00 | -18.88 | -0.41 | -12.18 | 37.01 | 0.62 | 34.89 | -12.58 |
| 9 | H | -65.33 | -0.03 | 2.02 | 54.62 | -0.72 | 0.00 | 0.00 | 36.51 | 40.87 | 0.00 | 0.12 | 4.87 |
| 10 | | | | | | | | | | | | | |
| 11 | | | | | | | | | | | | | |
| 12 | | | | | | | | | | | | | |
| 13 | | | | | | | | | | | | | |

Sheet1 Sheet2 Sheet3

**Figure 18** Heat Map display generated from Listing 30.

example to display the data as a Heat Map. Listing 30 modifies Listing 26 by removing the print statements and calling the `HeatMap` subroutine with appropriate arguments. Figure 18 illustrates the result of running the code in Listing 30 and selecting an appropriate data file.

```
Sub HeatMap(Plate As Range, arrData() As Double)
    Dim I As Integer ' Loop counter
    Dim J As Integer ' Loop counter
    Dim dblMax As Double ' Maximum value in range
    Dim dblMin As Double ' Minimum value in range
    Dim intVal As Integer ' Color index
    Dim arrClrIdx(10) As Integer ' Color index array

    ' Set color indexes that define color gradient
    arrClrIdx(0) = 11 ' Blue (coolest)
    arrClrIdx(1) = 32
    arrClrIdx(2) = 41
    arrClrIdx(3) = 33
    arrClrIdx(4) = 34
    arrClrIdx(5) = 2  ' White
    arrClrIdx(6) = 27
    arrClrIdx(7) = 44
    arrClrIdx(8) = 45
    arrClrIdx(9) = 46
    arrClrIdx(10) = 3 ' Red (hottest)

    ' Set data range to map into color gradient
    dblMin = -100
    dblMax = 100
    ' Call the sheet formatting routine
```

**Listing 29** Heat Map generation subroutine.

```
    FormatPlateDisplay Plate
    ' Insert data into cells, and color
    ' cells based on data Values
    For I = 1 To 8
        For J = 1 To 12
            ' Display data value in cell
            Plate.Cells(I, J).Value = _
                arrData(I, J)
            ' Map value into color index array
            intVal = CInt( _
                10# * (arrData(I, J) - dblMin) / _
                (dblMax - dblMin))
            ' Set cell background color
            Plate.Cells(I, J).Interior.ColorIndex = _
                arrClrIdx(intVal)
        Next J
    Next I
End Sub
```

**Listing 29** *Continued*

```
Sub ParsePlateData()
    ' Parse data into an array
    ' and display as a heat map
    Dim vrnFileName As Variant ' File name
    Dim intFileNum As Integer ' File number
    ' Line read from the file
    Dim strLine As String
    Dim strBarcode As String ' Parsed bar code
    ' Result of Split Command
    Dim arrLine() As String
    Dim intRow As Integer ' Current row
    Dim intCol As Integer ' Current column
    ' Parsed plate data
    Dim arrData(1 To 8, 1 To 12) _
        As Double
    ' Loop counters
    Dim I As Integer, J As Integer

    ' Select the file
    vrnFileName = Application.GetOpenFilename( _
        FileFilter:= _
```

**Listing 30** Generate a Heat Map display using data loaded from a file.

```
                "Text Files (*.dat),*.dat, " & _
                " All Files (*.*),*.*", _
                Title:="Select Data File")
    ' Exit if cancelled
    If vrnFileName = False Then Exit Sub

    intFileNum = FreeFile ' Get file number and open
    Open vrnFileName For Input As #intFileNum
    Line Input #intFileNum, strLine ' Read line

    strLine = Trim(strLine) ' Trim whitespace
    Do While strLine = "" ' Discard any blank Lines
         Line Input #intFileNum, strLine
         strLine = Trim(strLine)
    Loop
    ' Parse bar code
    If Left(strLine, 6) = "Plate:" Then
         strBarcode = Trim(Mid(strLine, 7))
    Else
         Debug.Print "Invalid file format"
         Exit Sub
    End If
    ' Parse data
    For I = 1 To 96 ' Loop over all Lines
         Line Input #intFileNum, strLine ' Read line
         ' Split on semicolon
         arrLine = Split(strLine, ";")
         ' Convert row ...
         intRow = CInt(arrLine(0))
         ' Convert column ...
         intCol = CInt(arrLine(1))
         ' Convert well data ...
         arrData(intRow, intCol) = _
              CDbl(arrLine(2))
    Next I
    Close #intFileNum ' Close file

    ' Generate the HeatMap
    HeatMap Sheet1.Range("B2:M9"), arrData
End Sub
```

**Listing 30** *Continued*

## 5 SUMMARY AND CONCLUSION

In this chapter we only have touched upon the power and utility of the pairing between Excel and the VBA programming language. Even though a complete discussion of the features of Excel and VBA would span several volumes and many thousands of pages, it should be clear that Excel is a vastly underutilized tool for managing data generated in a life sciences laboratory. There is an underappreciation for the wide diversity of data management tasks that can be handled by Excel and VBA.

It should also be clear that Excel is not a panacea; there are many laboratory data management situations for which Excel is not suited. As VBA and the functionality behind Excel are mastered, there is a tendency to attempt to use it to solve problems that require more sophisticated tools, such as a full-blown multiuser enterprise data management system, a document workflow system, or a collection of highly refined data mining utilities. Excel can be an excellent prototyping tool for these situations, but the skilled laboratory data manager will recognize when the time has come to benefit from more powerful tools.

For several years now Microsoft has been putting their efforts into the creation and enhancement of a new software development technology called .Net. The .Net Framework allows different programming languages and libraries to work together seamlessly. Microsoft Excel programming has been brought under the .Net Framework through a technology called Visual Studio for Applications (VSA). VSA will allow Excel scripting to be performed from one of the many .Net languages. Nevertheless, Microsoft has recognized that VBA is a powerful tool and that there are millions of lines of VBA code that have been written for Excel, and are in daily use. Given this reality, Microsoft has stated that it will continue to support VBA in Excel and other applications into the future.

As a powerful, full featured, and fully integrated programming environment for Excel, VBA remains a preferred choice for building Excel-based applications for managing laboratory data.

# AGE OF REGULATION

SANDY WEINBERG
*Fast Trak BioDefense, GE Healthcare*
*Atlantic, Georgia*

GERALD J. WHARTENBY
*Regulatory Consultant for GE Healthcare*
*Philadelphia, Pennsylvania*

1 INTRODUCTION
2 ORIGINS OF THE FDA
3 21 CFR PART 11
4 RISK ASSESSMENT
5 RISK ASSESSMENT AND REGULATION
6 REGULATION IN DRUG DISCOVERY: THE FUTURE

## 1 INTRODUCTION

In examining regulations, often it is easy to become inundated with the complexity of existing rules and regulations. Like being in a maze, we travel through the twist and turns of current Food and Drug Administration (FDA) regulations often leading us in directions where a clear path is difficult to find. The number, import, and intricacy of the U.S. FDA regulations, requirements, guidelines, and draft documents are often so overwhelming as to make it difficult to see the underlying ideology that directs so many FDA measures.

As we travel this maze of regulations, there are three underlying precepts that together provide a foundation for and an understanding of all the myriad

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

of FDA procedures: (1) proximal causality, (2) risk assessment, and (3) self-regulation. Having a clear understanding of these three precepts can help one understand future FDA actions.

*Proximal causality* is probably the most basic of FDA regulatory models. Let us consider a consumer who buys a product clearly labeled as containing dehydrated corn. The consumer has an expectation that the box contains pure corn kernels that are free from contaminants such as mice or rat droppings and also free from insect parts. Public health and consumer confidence are maintained as long as the label is accurate and the product is in fact free of contaminants. Consumers also need the assurance that this final product was processed properly from the beginning.

To ensure that this really does occur, regulators are authorized to investigate the packing plant itself to be certain that it is free of infestation. If the packing plant processes raw materials to eliminate those contaminants, regulators need to focus their attention on the cleanliness and operation of the processing procedure. It would also be necessary to inspect and regulate any and all equipment used in the process as well as the records of operation and maintenance of that equipment. And in the computerized world we live in, the computer systems too are subject to the eye of regulation.

Regulatory attention focuses over time on the earlier, inner steps of groundwork and progression, always moving backward to more proximal causalities of safety threats and associated problems. In part, that movement is a function of satisfaction with the more surface levels of control; in part, it is a function of deeper understanding of the ultimate causes of problems. Arguably an underlying causality of the trend derives from the reality that once investigators have achieved a satisfactory level of control for a surface issue they have the time and resources to dig deeper. It is clear that the trend to move backward to proximal causality is a response to consumer demand—standards of cleanliness are continually raised in response to public understanding. The same trend applies to the performance of drugs and their side effects, the reliability and therapeutic value of medical devices (heart pacemakers, dialysis machines), the purity of biologics (vaccines, blood products, etc.), and the quality and safety of all the other FDA regulated products (such as X-ray machines and microwave ovens) and their processes.

As we move through the regulatory maze, there is an ever-mounting cost to ensure the safety and quality of these various products. But does the cost of such regulations ever become a limiting factor?

Human blood provides an interesting example. While most blood transfusion donors are unpaid, there is a cost associated with the collection. As the screening cost rises, the resulting expense rises proportionately. To put it simply, if it cost $1.00 to collect a unit of blood, and only one in five units pass the screening process, the cost of a usable unit of blood now becomes $5.00. If the public health and safety is significantly increased, then the dollar factor is well worth it. But, if the screening provides only minimal value, do we not then need to look at a second method of regulation—risk assessment. It is

clear that over the history of the FDA no regulated product is absolutely "risk free." So we must look at "benefits vs. risks."

*Risk assessment* is the second regulatory trend today. Simply put—it is the balancing of increased safety against the limited effect of the cost of implementing that safety factor. Having its origin in the medical device area, the concept of risk analysis considers the probability and potential severity of an adverse patient reaction, the overall performance of the system, and the process of manufacturing. Regulatory scrutiny is then assigned, and supporting evidence collected, in proportion to the degree of risk inherent in a given situation. Remembering that no regulated product is risk free and so consideration is given to a product with an eye toward risk vs. benefits especially in products used in life-threatening situations.

*Self-regulation* in the pharma industries is the ability to police oneself. The FDA has taken the consistent position that the industry is self-regulated and that it is the agency's job to oversee and police this self-regulation.

All companies and products falling under the wide umbrella of the FDA are subject to internal quality assurance requirements. Even if a product performed flawlessly, without any side effects, purity defects, or adverse reactions, the manufacturer would be justly criticized if it did not have in place a quality assurance program. It is the goal of the FDA and its industry partners to continually improve and oversee the minimization of potential problems.

Working in conjunction with the proximity causality and risk assessment trends, this emphasis on self-regulation results in a continuing shift of attention backwards in the laboratory and manufacturing processes. Users of laboratory information management systems (LIMS) now routinely investigate their software suppliers since the internal quality of an automated laboratory is dependent increasingly on the accuracy of the software that manages and interprets experimental results. In much the same way manufacturers extend their quality assurance to reach vendors of raw materials, of manufacturers equipment, and software control systems. The self-regulation, tempered by a risk assessment to put priority on areas of greatest potential danger, forces a proximal causality shift.

It is clear that we live in a world of regulation. Promoting public health and protecting consumers have placed us on the road to ensuring that effective and safe products reach the market place. All in all we are concerned with the quality of life expected by each member of our society. But how did all this come about? Let us take a step back and look into the history of the FDA.

## 2 ORIGINS OF THE FDA

It seems that the quality of life has been a major concern of people from the beginning of time. What they ate, how it was prepared, how they lived, concern over medical care and obviously the medications and drugs they took, and, of

course, the medical devices that were recommended and used have been scrutinized over the history of humankind.

It has become a trend over this history of humans to look for help from authorities to ensure not only the safety of these devices and drugs but also the purity of products consumed and the effectiveness of the products as well.

How a product is prepared and how it is packaged and labeled are questions frequently asked over time. Is the product safe? All this has eventually laid the groundwork for future regulation laws. The cry to reform was loud and clear.

As far back as 1202, King John of England proclaimed the first English food law, the Assize of Bread, which prohibited adulteration of bread with such ingredients as ground peas or beans. By 1820, eleven physicians met in Washington, D.C., and established the U.S. Pharmacopeia, the first compendium of standard drugs for the United States. And it was President Lincoln who appointed a chemist, Charles M. Wetherill, to serve in the new Department of Agriculture, the beginning of the Bureau of Chemistry and the forerunner of the Food and Drug Administration. Since President Lincoln's time, there have been 100 major developments in the history of the FDA.

From the 1902 Biologics Control Act (passed to ensure purity and safety of serums, vaccines, and similar products used to prevent and treat humans) to the Food and Drugs Act (1906) passed to deal with misbranded and adulterated foods, drinks, and drugs to the 1938 Federal Food, Drug and Cosmetic Act (requiring new drugs to be shown safe before marketing) down to today where concern is questioned over temporary tattoos, each development has continued to lay the foundation of regulatory laws.

Today we question the members of Congress over the importation of drugs that may not have passed all the checks and balances in place here and ask ourselves how effective they are and is the risk worth taking based on the price of the drugs. The questions continue but the fundamental mission of the FDA is well defined.

The FDA's mission is clear:

1. To promote the public health by promptly and efficiently reviewing clinical research and taking appropriate action on the marketing of regulated products in a timely manner.
2. With respect to such products, protect the public health by ensuring that foods are safe, wholesome, sanitary, and properly labeled; human and veterinary drugs are safe and effective; there is reasonable assurance of the safety and effectiveness of devices intended for human use; cosmetics are safe and properly labeled; and public health and safety are protected from electronic product radiation.
3. Participate through appropriate processes with representatives of other countries to reduce the burden of regulation, harmonize regulatory requirements, and achieve appropriate reciprocal arrangements.

It is quite clear that the history of the FDA is rich, varied, and interconnected with the major political reformist trends of the last century. Buried in the detail of significant dates, regulatory revisions, and personalities are five aspects of the regulatory structure:

1. Accuracy of labeling
2. Growth
3. Oppositional
4. Speed the approval
5. Risk assessment

What is in a product, how those ingredients affect humans, what a product may be called, and how effective it is have laid a foundation of labeling accuracy at the very heart of the FDA's labeling concerns. The FDA has been most concerned with making certain that consumers as well as health-care professionals are provided with accurate information on consumer products.

In an article—"The Rise and Fall of Federal Food Standards in the United States: The Case of the Peanut Butter and Jelly Sandwich," Society for the Social History of Medicine, Spring Conference, 1999, Aberdeen, Scotland—Suzanne White Junod writes: "In 1906, Congress passed the U.S. Pure Food and Drugs Act. It was one of the first consumer protection acts passed in the United States." Dr. Junod further stated:

> For an act propelled into law through a focus on Food, the law was surprisingly and seriously flawed in its food provision. The first flaw was apparent even before the law was passed. Offshoots of the food industry (principally manufactures of rectified or blended whiskey and chemical preservatives including formaldehyde, borax, copper salts, salicylated, saccharin, and sodium benzoates) defeated a provision which would have allowed the Government to set standards for food products.

She continued:

> The second flaw was not quite so obvious until 1920s brought a heyday in inventive advertising, and the Great Depression of the 1930s created a market for cheap, inferior products. And this is where jelly first enters my sandwich standards analogy.

Concerned over strict food standards—these food industries inserted a so-called "Distinctive Name Proviso" into law. Dr. Junod stated in the article that this distinctive name proviso permitted the marketing of foods that would have otherwise been illegal under the 1906 act. She further stated in the case of jam and jelly, they would have been considered adulterated or misbranded under the law since they had so little fruit. Beautiful food dye hues, artificial pectin, and grass seeds, accompanied by expensive yet tasteful packaging, and promotion through clever advertising, all created a new kind of fabricated

product. This product was given a fanciful and distinctive yet meaningless name—BRED-SPREAD. This product typifies the kind of inferior product that began to gain a foothold in the U.S. marketplace beginning in the 1920s.

Consumers were faced with a problem in labeling appearing on market products. Could one depend on the label to tell one the quality of food and how much of a certain product was really in the can?

But it was the FDA in 1933 that started the process of rewriting the 1906 Pure Food and Drugs Act in its efforts to rescue a potentially harmful industry.

The FDA writes: "Consumers rely on product labels to know what the product is and how it is to be used. The agency regulates what's on these labels to ensure that they are truthful and that they provide useable information that helps consumers make healthy, safe decisions when using the product" (FDA presentation entitled "Protecting Consumer", Protecting Public Health—PPT Slide 13 of 34).

From fake medical devices to almost jelly to patent medicines that claimed to heal just about everything, the FDA has continually launched a truth-in-labeling campaign to protect consumers from false claims to not only the purity of the product but to the effectiveness and safety of the product. Accuracy of labeling emerged as a significant FDA historical trend.

The second aspect of the regulation structure is "growth." Whether or not there has been a general growth of "big government" or of the quality of regulations is a matter of ongoing political debate. It is obvious that the growth of the number and variety of products and companies subject to FDA regulations is a matter of record. Through the history of the FDA we see an agency that constantly needs to respond to the challenges presented to the consumer by highly complex industries often producing highly complex products. The development of biotechnology, the discovery of new classes of pharmaceuticals, and the invention of new medical devices have all swollen the U.S. Pharmacopoeia in and lists of potential treatments.

The results of increasing demand and all but static supply (of regulations and regulatory budgets) has resulted in increasing long review times, decreasing common field investigation and inspection, and a need to find ways of rationing regulators energies without compromising public safety and health, a complex problem in itself.

But yet another aspect that gives rise to a regulatory pattern is the relationship of the product industry to the FDA. Is it an oppositional relationship by choice or history? Does a history of false claims from the early 1920s of certain product manufacturers to cure just about anything and the need to protect the consumer give rise to an oppositional relationship?

Have a sequence of events led to an oppositional atmosphere or has the development of the FDA over its history simply led it to the status of being the "watchdog?"

It is a fact that the FDA is the consumer *watchdog*. It stands in opposition to industry excesses, cut corners, and of course false claims.

The need of consumers for quick access to new products and, of course, scandals about unsafe or ineffective treatments has laid the groundwork for yet another aspect of the regulatory structure—speed the approval.

The problem of consumer demand for rapid release of promising and safe cures and treatments without regard to the science of testing and review represents another historical trend in the FDA. Whenever a new disease receives significant public attention, the FDA feels the pressure to accelerate the approval process. In recent years the tendency has been seen in human immunodeficiency virsus/acquired immunodeficiency syndrome (HIV/AIDS) treatments, in terrorist-response vaccines and treatments, in Alzheimer's treatments, and in severe acute respiratory syndrome (SARS) diagnostics. The call to speed up the process is heard. But it is obvious to the industry that the FDA represents only a minor portion of the lengthy approval process that certainly can take up to 10 years.

We would rush to judgment to find fault only with the FDA. It is the range of quality and safety studies that take time, and while the FDA does have responsibilities to review and analyze those studies, blaming the FDA for the time required for approval and acceptance is nonproductive.

These four aspects have produced the final trend that is only recently receiving widespread attention. The first four trends force a rationing of regulatory effort and argue for focusing those limited efforts on areas of greatest quality concern to bring safe and effective products to market as rapidly as possible. The current course to use "risk assessment" as the rationing determinant is the result of those pressures.

The potential ultimate effect of adverse events, measuring both probability and severity, is used in risk assessment as a rationing factor.

## 3 21 CFR PART 11

Part 11 of section 21 of the Code of Federal Regulations (CFR) is an evolutionary result of the historical trends affecting the Food and Drug Administration. Part 11 deals with electronic records, with the systems that produce those records in automated laboratories, clinical testing environments, and manufacturing facilities, and with the electronic approvals and signatures of those records. The regulation includes specifications for the validation and testing of systems, for archiving and retrieval of records, and for training and standard operating support of user personnel.

In the proximal causality trend, Part 11 represents the regulation of a tool used in support of Good Manufacturing Practices (GMP), Good Laboratory Practices (GLP), and Good Clinical Practices (GCP) guidelines. In the risk assessment trend, Part 11 is clearly subject to a rationing of regulatory energy and evidence in accordance with the probability and severity of adverse events. Conventionally in the self-regulation trend, Part 11 requires that an organization establish standards for quality and prove conformity to those

standards. In the development of the regulatory world we live in, Part 11 conforms with the initial focus on labeling (here, the accurate description of system requirements), with the growth of regulation (in response to the growth of automation), with the independence from and opposition to industry, imposing standards of quality and testing on the computer industry, with attempts at speeding the time to market (spearheaded by automating the review process), and with the use of risk assessment as the rationing factor, previously discussed in the blending of historical and regulatory trends.

In the remainder of this chapter you will find the six steps of conducting a risk assessment, a check list that outlines the broad scope of 21 CFR Part 11 and a drug discovery regulation checklist.

## 4 RISK ASSESSMENT

Risk assessment is the other side of benefits analysis. It examines the probability and severity of "negative" outcomes of a process, applicable to drug discovery, utilization of a pacemaker, manufacture of a drug, or any other process. However, risk assessment has a unique role: it can help determine the amount of compliance evidence required for acceptance. How closely the FDA will examine and how much time, energy, and expense should be invested in proving compliance will be determined by a risk assessment of the drug discovery process (or any other process).

To conduct a risk assessment, integrate these six steps into the process of planning and implementing a compliance strategy:

1. Utilizing the written requirement documentation for a specified system, device, computerized equipment, component, or application, identify the desired performance under each of the defined requirement application.
2. Utilizing historical experiences with the specified system, device, computerized equipment, component, or application, and/or utilizing historical experiences related to predicate systems, devices, computerized equipment, components, or applications, and/or utilizing industry standards related to the specified system, device, computerized equipment, component, or application, determine the alternate undesired performances for each defined requirement application.
3. Determine the probability of each occurrence performance, for both the desired performance and the undesired performances for each defined requirement. Probability of occurrence can be calculated utilizing historical experiences with the specified system, device, computerized equipment, component, or application, and/or utilizing historical experiences related to predicate systems, devices, computerized equipment, components, or applications, and/or utilizing industry standards related to the specified system, device, computerized equipment, component, or application.

4. Analyze each of the undesired performances to characterize the severity of that performance in terms of risk to human life, health, and/or safety.
   a. "High severity" is generally defined as loss of life, substantial loss of quality of life, and/or substantial disabling effect.
   b. "Medium severity" is generally defined as compromise of quality of life and/or some disabling effect.
   c. "Low severity" is generally defined as little or no effect on quality of life or on normal life activities.
5. For each undesired result calculate the risk according to the following grid:

| SEVERITY | PROBABILITY | RISK |
|---|---|---|
| High severity | High probability | High risk |
| High severity | Medium probability | High risk |
| High severity | Low probability | Medium risk |
| Medium severity | High probability | Medium risk |
| Medium severity | Medium probability | Medium risk |
| Medium severity | Low probability | Low risk |
| Low severity | High probability | Medium risk |
| Low severity | Medium probability | Low risk |
| Low severity | Low probability | Low risk |

6. Apply the table results to the validation protocol or policy to determine the appropriate level of testing and validation.

## 5 RISK ASSESSMENT AND REGULATION

Risk assessment provides a measure of degree of regulation in the process of regulating the drug discovery and development. Although regulation is applied to all aspects of the development process, two caveats mitigate that reality.

First, regulation of drug discovery and development is applied only retrospectively. That is, the end product and process are subject to FDA requirements ONLY IF the decision is made to move the discovered substance forward in the developmental process. Lines and paths that are abandoned or halted are not subject to regulatory review. That restriction may be of little practical value since all paths are considered potentially successful or they would not have been initiated. Most organizations, therefore, opt to integrate regulatory controls in anticipation of possible success.

The second caveat has a greater strategic and tactical value. The practical focus of the FDA is upon the tools utilized in discovery. While an educator may try to assess reading skills to determine a student's readiness to learn,

regulatory attention focuses on the tools to assess the process. The student's reading skills and the discovery tools are both elements in the process and indicators for potential problems.

Since the drug discovery and development process is largely based upon the use of complex systems-automated laboratory equipment, molecular modeling software, statistical analysis systems, and so forth, those tools tend to be computers. The very nature of computers causes the most regulatory concern.

Computer systems are complex, based upon internalized and sometimes obscured decision rules. Computers are flexible, allowing multiple applications with appropriate changes in internal directions (programs). They can overwrite the changes made to those decision rules, directions, and even databases, often leaving little or no markers indicating changes were made.

These characteristics, coupled with the industry's increased reliance on the use of computer systems, have led to a focusing of a significant portion of regulatory energy on the automated tools of the field. In the drug discovery arena where reliance on computers is particularly high, the major regulatory focus, mitigated by risk assessment, has been on the systems that collect, analyze, manipulate, and report. The result of that focus is the newest of the FDA regulations, 21 CFR Part 11, Electronic Signatures and Electronic Archives.

Originally intended to define rules for accepting electronic signatures in lieu of human (paper) signatures on documents, Part 11 has been expanded to define virtually all aspects of computer control and of the regulation of computer systems. The requirement includes standards for documentation of the validation (testing and managerial control) of systems; of the training and operating procedure assistance to be provided to users; of security, change control, and disaster recovery assurances, etc. Here is a checklist that outlines the broad scope of 21 CFR Part 11:

- System is used in support of drug-related research, laboratory analysis, clinical research, manufacturing, production, and/or tracking.
- Record and/or signature system has been subjected to an appropriate and thorough system validation audit.
- Audit:
  - Conducted within 24 months
  - Conducted by independent or outside expert
  - Included review of:
    - Testing documentation
    - Development documentation
    - Standard Operating Procedure (SOP) documentation
  - Included inspection of operating environment
  - Trail review
    - Archive
- Change control

- Archive
- Disaster recovery
- Use
- Training
- System validation documentation collected
  - Evidence of requirements
  - Design approvals
  - Testing
  - Implementation
- Validation
  - Protocol
  - Team credentials
- Development documentation
  - Requirements/design document
  - Trace matrix
- Standard operating procedures
  - Use
  - Training
  - Change control
  - Archive
  - Disaster recovery
  - Audit trail review
- Testing
  - Boarder cases
  - Norm cases
  - Code review
- System Inventory
  - Hardware
  - Software
- Records are retained:
  - For appropriate length of time. Generally 10 years or two generations over treatment duration
  - In machine readable form
  - In human readable form
  - In heat proof, fire proof, flood-protected environment
- Records
  - Are appropriately labeled
  - Can be restored in reasonable (generally 72 h) time
- Procedures are in place to restrict access to data and records to appropriately authorized persons.

- Operation checks of system have been designed to assure appropriate functioning of hardware and software
- Audit trails
  - Preferably electronic and protected
  - Alternately manual and carefully monitored
  - Built into system to detect and identify data changed, including
  - Tracking of time
  - Date of change
  - Change agent
  - Reason for authorized change
- Password/password
- Password/key
- Password/biological
- Electronic signatures utilized only in system with:
  - Dual-level unique identifier authorizations
  - Internal procedures to assure that approved documents have not been modified (without authorization) from specified date and time
- Time system
  - Zulu time
  - Greenwich mean time (GMT)
  - Location-affixed time
  - Single time zone
  - Date system
  - U.S. (dd/mm/yy)
  - International (mm/dd/yy)
- Methodology has been implemented to assure the validity of input data
  - Might include
    - Dual confirmation of input
    - Use of check digits
    - Internal norm confirmations
    - Other techniques
- Systems users and administrators have:
  - Received appropriate regulatory and functional training
  - Ready and constant access to appropriately comprehensive, clear, applicable, timely, and management-approved SOPs
- All aspects of the electronic records and electronic signature systems in place have been designed to provide a level of security and control equal to or exceeding the equivalent controls inherent to manual (paper) systems.

The Part 11 requirement provides guidance for control of the system tools used in the drug development process. It serves as an outline of the FDA scrutiny of drug discovery and development. Evidence of compliance represents an insurance policy providing confidence that, should a drug move along the pipeline to a New Drug Application (NDA), the Food and Drug Administration will accept supporting data from the early stages of the development process.

## 6 REGULATION IN DRUG DISCOVERY: THE FUTURE

The regulation of drug discovery is an evolutionary process, balancing the public need for *access* to effective treatments and cures against the need to assure the *safety* of those emerging drugs. The U.S. Food and Drug Administration and equivalent agencies worldwide continually cope with that delicate balance, comparing the Hippocratic dictum "First, do no harm" with the demands of disease and injury victims and their families who plead "try something!"

The balance is currently maintained through three underlying principles of drug discovery regulation. First, the trend to *proximal causality* continually shifts the regulatory focus to the primal events in the pharmaceutical chain. In the future that focus will fall increasingly on the reliability of molecular models, of preclinical testing, and other early-stage procedures and computerized equipment used in the first stages of the drug discovery process.

Second, emerging awareness of the need to avoid the unnecessary and ineffective cost of excess regulation of products and procedures that represent little or no real safety threat has led to the growing utilization of a *risk assessment* in determining the appropriate depth of regulatory focus. The trend is expanding and generalizing: Risk assessment will stand as the most significant regulatory innovation of the decade. In many cases in which early drug discovery activities are so far removed from final production of drug product as to represent little real danger, the risk assessment will balance the trend to proximal causality to product a rationally limited regulatory involvement.

Finally, emerging as dominate in the self-canceling balance of proximal causality and risk, will be the continuing trend toward *self-regulation*. Expect the FDA to serve as a supra quality assurance unit periodically and remotely overseeing the drug discovery process to monitor the self-regulation of that process of drug discovery organizations themselves. The primary responsibility to establish appropriate and risk-assessed standards, to conform to those standards, to document that conformance, and to audit the process will lie with the pharma industries.

In such an evolving world the drug discovery process can be expected to appropriately balance public safety against development, leading ultimately to public assess to new drugs. It is an optimistic and realistic model.

**Drug Discovery Regulation Checklist**

- ***Risk assessment***
  - Determination of nondesired alternate outcomes and results of the drug discovery operation
  - Calculation of the approximate probability of those occurrences, utilizing historical logs and/or results of predicate device operations
  - Determination of the potential severity (in terms of direct threat to human health and safety) of those occurrences
  - Resulting categorization of the operation as low, high, or medium risk
    - If risk is determined to be low, other steps can be:
    - Mitigated (limited testing in validation)
    - Eliminated (the vendor audit)
    - Reduced (the PQ)
- **Validation of the operation** including:
  - **System validation** of all automated components, including:
    - Documentation of system requirements and design
    - Development of a trace matrix
    - Development and exercise of appropriate test scripts
    - Review of the code itself
    - Analysis of the SOPs to assure:
    - Appropriate use
    - Archive
    - Disaster recovery
    - Change control
    - Training
- **Process validation**, focusing on possible contamination of
  - Media, equipment, and final product, including but not limited to:
    - Bacterial contamination (both Gram positive and Gram negative)
    - Viral contamination
    - Material contamination
- **Part 11 audit** emphasizes:
  - An audit trail to track all data changes
  - If electronic signatures are in use:
    - Dual confirmation of identity
    - Locking of document after signature
    - Unambiguous time/date stamp
- **Archive** (electronic and human readable) of all files:
  - Control of data accuracy
  - Appropriate training

- **Audit of system vendor** either by:
  - Independent "expert witness" certifying compliance with appropriate regulations or
  - The end user organization
    - Key elements:
    - The criteria utilized
    - The credibility of the auditor
- **Installation qualification (IQ)**, using preestablished standards to assure appropriate initial implementation of the system or systems. The IQ may be conducted by the vendor, by the end-user organization, or by a combination of both.
- **Initial and periodic calibration**, in accordance with a metrology plan appropriate to the specific system or systems in use. Some systems are self-calibrating; a few others do not require recalibration after initial installation
- **Operational Qualification (OQ) and Performance Qualification (PQ)**, sometimes performed separately and (in some circumstances and systems) combined. The OQ assures the system is ready for use; the PQ assures that it is appropriately in use.
  - **Problem report**: A system for reporting (and reviewing) any encountered malfunctions, necessary changes, or other problems.

# SIMULTANEOUS SCREENING OF MULTIPLE CELL LINES USING THE CELLCARD SYSTEM

Oren E. Beske

*Vitra Bioscience, Inc.*
*Mountain View, California*

1 INTRODUCTION 354
2 CELLCARD TECHNOLOGY 356
3 GENERAL CONSIDERATIONS FOR MULTIPLEXING CELL-BASED ASSAYS 357
4 GENERAL PROTOCOL FOR USING THE CELLCARD SYSTEM 359
Experiment Design and Plate Layout 359
CellCard Preparation 360
Tissue Culture 360
CellCard Mixing and Dispensing into 96-Well Format 360
Compound Addition and Assay Protocol Considerations 360
CellCard Reading 361
Image Analysis and Data Visualization 361
5 THEORETICAL EXAMPLE: APPLICATION OF CELLULAR ARRAYS TO THE Identification and Profiling of Cell Type Selective Antimitotic Compounds 362
6 SPECIFIC CASE STUDIES USING THE CELLCARD SYSTEM 363
7 MULTIPLEXED DOSE CURVE ANALYSIS USING THE TUNEL ASSAY 363
8 MULTIPLEXED ANTIPROLIFERATION SCREEN; IDENTIFYING CELL TYPE SELECTIVE Compounds 367
9 SUMMARY OF KEY FEATURES OF CELLCARD SYSTEM 370
References 371

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

## 1 INTRODUCTION

Historically, biochemical assays have vastly outnumbered cell-based assays in the screening environment of the drug discovery process. The popularity of Biochemical assays has been driven by their robustness, reproducibility, and inexpensive nature. As a result, biochemical assays are used in many steps of the drug discovery process including primary screening, lead identification, lead optimization, and the determination of chemical structure–activity relationships (SAR). Although robust and relatively easy to implement, many biochemical assays do not address important biological issues, that is, the complex cell biology surrounding most targets. The desire for more biologically relevant data from primary high-throughput screening (HTS) as well as downstream processes (i.e., lead identification and optimization) has resulted in a dramatic increase in the development of technologies specifically designed for cell-based assays and the use of cell-based assays in the drug discovery process (*http://www.hitechbiz.com*).

Traditionally, cell-based assays were slow, lacked reproducibility, and were prohibitively expensive. However, as our understanding of cell biology increases and new technologies emerge, cell-based screens are becoming faster, cheaper, more reproducible, and increasingly powerful. This is evident in the vast number of new biosensors, assays, and detection systems being developed for cell-based HTS [1–5]. In addition, the use of genetic tags, such as fluorescent proteins, have enabled scientist to develop unique assays that require no additional staining and provide a means to acquire kinetic as well as localization data (*http://www.bioimage.com*). Novel dyes (e.g., DiBAC, Fluo-3) have brought complicated cell biological processes to simple bench-top readers [6–9]. These bench-top readers encompass flexible new detection technologies that enable scientists to robustly acquire data from cell-based assays performed on a single microscope slide or in the context of a fully automated HTS facility.

Concomitant with these advances in cell biology and assay development there has been tremendous strides forward in the field of cellular imaging technologies. Recent years have seen multiple imagers brought to market (*http://www.cellomics.com*, *http://www.amershambiosciences.com*, *http://www.atto.com*, *http://www.moleculardevices.com*, *http://www.evotectechnologies.com*) with the ability to scan microtiter plates gathering multiple parameters from each well. These imaging technologies can be separated into two general categories: microimagers and macroimagers.

Macroimagers have the ability to image an entire microtiter plate with a single exposure. Since this approach acquires an image of the entire plate, macroimagers often have the ability to work with 96-, 384-, and 1536-well microtiter plates without having to significantly reconfigure the instrument. One example of this is the fluorecscence imaging plate reader, or FLIPR [7, 10, 11]. The instrument was developed to acquire quantitative kinetic measurements of calcium flux. It images the entire plate with a single exposure

and therefore has a very fast read time, allowing for high-throughput operations. However, since these macroimagers acquire very low magnification images, they obtain average measurements of a cell population in much the same way photomultiplier tube (PMT) based readers do. As a result, it is not possible to identify heterogeneity within the cellular population, nor is it possible to gather subcellular information.

Microimaging technologies (*http://www.cellomics.com*, *http://www.evotechtechnologies.com*, *http://www.moleculardevices.com*) employ higher magnification approaches resulting in the ability to gain intracellular resolution. Typically, acquisition protocols can be highly customized ranging from 2× to 20× magnifications, 1 to 10 images per well, and 1 to 4 fluorescent labels. Microimagers have slower plate read times and provide fundamentally different information than macroimagers. While macroimagers enable rapid imaging of an entire plate, they are restricted to simple population-based intensity measurements from the well. Microimagers, by virtue of their ability to gain intracellular spatial resolution, allow scientist to analyze complex intracellular events such as nuclear translocation, nuclear morphology, and endosome formation [12–14]. The quantification of these cell biological phenotypes is gained through the use of sophisticated image analysis algorithms. These algorithms act upon the images and translate them into numbers that accurately represent the biology of interest. For example, using a single nuclear dye, these algorithms can determine nuclear size, density, condensation, fragmentation, and the like. In addition, the cellular samples can be stained with multiple fluorochromes in order to analyze multiple biological parameters from each well. In doing such, one can analyze multiple aspects of a complex biological process within a single well. This approach, often referred to as high content screening (HCS), has gained increasing popularity in recent years as it enables more information to be gathered from each microtiter well [15, 16].

The advent of DNA (deoxyribonucleic acid) arrays enabled scientists to gather more than one data point from a single experiment or microtiter well. This approach, widely used in the analysis of gene expression, has expanded to the analysis of the genome, proteins, and cells [17–20]. This parallel analysis of multiple analytes (often referred to as multiplexing) provides the researcher with the magnitude of an effect (the signal) as well as its selectivity (the comparison of the signal across analytes). By using multiplexing technologies, the data points come from a single reaction and therefore can be confidently compared to one another. That is, pipetting and multiple other sources of error cannot be the cause of any selectivity identified. Finally, multiplexing is a means to miniaturization. That is, since multiple data points are obtained from a single reaction, the amount of sample used per data point is greatly reduced, making it a very attractive approach when reagents are expensive or there is a rare sample to be analyzed, that is, patient primary cells. Multiplexing technologies decrease the cost as well as the reagent consumption for each data point.

Vitra Bioscience, Inc. (*http://www.vitrabio.com*) has developed a multiplexing technology that enables cell lines to be robustly multiplexed. This platform technology is called the CellCard system. The CellCard technology merges the ability to gather multiparametric intracellular data afforded by microimaging with the ability to simultaneously analyze a compounds potency and cell-type selectivity. This is achieved because the CellCard technology enables multiple cell lines to be assayed (multiplexed) in a single microtiter well [21, 22]. (add Beske 2004) This chapter describes the CellCard technology and how it can be applied to the determination of cell-type selectivity of bioactive molecules.

## 2 CELLCARD TECHNOLOGY

The CellCard technology is centered on an encoded microcarrier approach to multiplexing. By associating each cell type to a uniquely encoded microcarrier (CellCard), these cells can be uniquely identified and their response measured. The CellCards (Fig. 1*a*) are flat, rectangular, and contain a colored barcode. The encoding strategy employs two colored bands (small arrow) on either side of a clear data readout section (large arrow). This physical separation of the data from the encoding strategy ensures that there is no cross-talk or limitation of one imposed on the other. Therefore, the encoding bands do not limit the number of fluorophores that can be used to measure cellular parameters. The size and aspect ratio of the CellCards allow for approximately 80 to 100 of them to be automatically deposited in a single well of a 96-well microtiter plate (Fig. 1*b*). To perform a multiplex experiment using Cell-

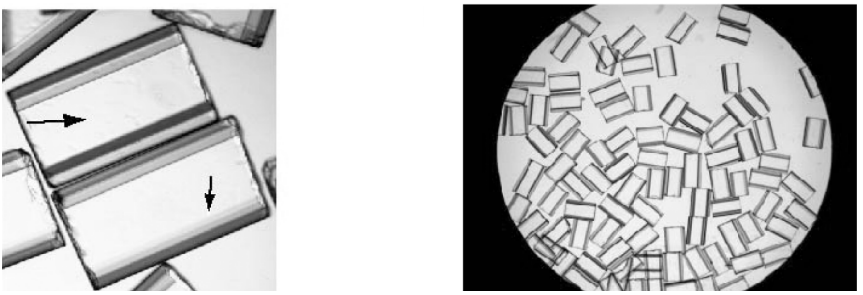

(*a*) (*b*)

**Figure 1** (*a*) The CellCards are flat and rectangular in aspect ratio. They are functionally separated into two parts: A colored barcode (small arrow) and an optically clear data readout section (large arrow). (*b*) The current implementation of the CellCard system is in 96-well microtiter plate format. This image shows multiple CellCard codes dispensed into a single well. There are approximately 80 CellCards in this well.

Cards is relatively simple (Scheme 1) and will be discussed in more detail later in the chapter. Briefly, adherent cells are grown on the surface of the CellCards with each cell type being uniquely associated with a code. That is, cell type 1 is grown on code 1 while cell type 2 is grown on code 2. After all of the cell types have adhered to the CellCards, they are mixed and then dispensed into a 96-well microtiter plate (Fig. 1*b*). In doing so, an aliquot of the encoded cell mixture is placed in each of the wells, generating a multiplexed assays in each well. Compounds are then added to the wells, incubated for the appropriate amount of time, and the cells stained for the biological parameters of interest. Finally, the CellCard array is decoded, and the cellular data captured and analyzed via the CellCard reader and associated software.

## 3 GENERAL CONSIDERATIONS FOR MULTIPLEXING CELL-BASED ASSAYS

Since the separate cell types to be analyzed are placed in a single well, there are some unique considerations that must be taken into account when deciding which cell types and assays are appropriate to be adapted to CellCards in a multiplexed format.

The living cells will be sharing a common extracellular environment during the compound incubation time. Therefore, one factor to take into consideration when choosing cell lines is the media conditions in which the incubation will take place. For example, if the cell lines to be assayed are all stable transfectants (i.e., overexpressing targets of interest) selected in Chinese hamster ovary (CHO) cells with a neomycin resistance marker, this media concern will be trivial since they are already being grown using common media conditions. On the other hand, if the cell lines are tumor cell lines derived from different tissues of origin (as presented later in this chapter) the situation is slightly different. If the cell lines all grow in a common media (i.e., RPMI +10 percent serum), there is no media concern. However, if a subset of the cell lines have been growing in different media, there may be a concern that common media-growth conditions are not being used. There are multiple possible solutions to remedy this issue. One solution is to adapt the cell lines all to a common media. This can be accomplished by passaging them for approximately 2 weeks (the time required is cell type and media type dependent) in the new media. Then, once adapted, the cell line should be recharacterized with respect to the relevant biology (i.e., doubling time or signaling ability) for the assay. If the biology is within a reasonable range for the screen, this cell line can now easily be multiplexed within the common media of the other cell types. This approach may not be appropriate for some cell types depending on the assay. Therefore, one may want to continue growing the cell lines in their independent media and only combine them into a common media for the duration of the compound incubation. Although this gets around the necessity to adapt

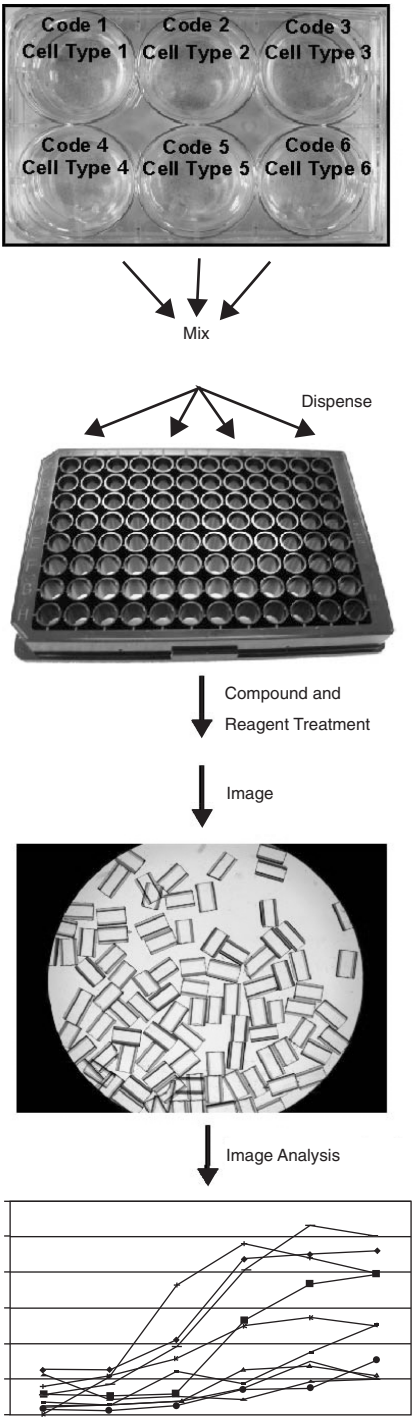
Code 1
Cell Type 1
Code 2
Cell Type 2
Code 3
Cell Type 3
Code 4
Cell Type 4
Code 5
Cell Type 5
Code 6
Cell Type 6
Mix
Dispense
Compound and
Reagent Treatment
Image
Image Analysis

the cell line to a new growing condition, it could bring about other concerns when the cell line is exposed to the common media for the compound incubation time. Whatever approach is taken to address this issue, the new assay conditions and biological consequences must be characterized and be within an adequate range before performing the experiment. If this is not done, it may be difficult to interpret the data, since the new assay conditions may introduce undesired or uncharacterized effects on the biological parameters being measured.

The assay of choice and compound incubation conditions should also be appropriately chosen in order to gain the most robust biological information. With all of the cell types in a single well it is advantageous if their biological responses, as determined by the parameters being assayed, occur with similar kinetics. For example, take the case where cellular proliferation is being measured and one is screening for a compound that acts at a specific step in the cell cycle. Since the screen is designed to find inhibitors of a specific cell cycle event, the cell lines should have similar doubling times. In that way, the compound incubation time should be chosen that ensures that each of the cell types have transited through the cell cycle in order to allow the compound to exert its effect. For example, assume the compound incubation time is set at 48 h and the multiplexed experiment contains a cell line with a 24-h doubling time (fast growing) and another with a 72-h doubling time (slow growing). In this scenario some cells from the slow-growing cell line may not have experienced the appropriate molecular events of the cell cycle for the compound to exert its effects, while the faster growing cell line will have, on average, transited that critical step twice. In such a case, any lack of antiproliferative activity seen against the slow-growing line would be difficult to interpret. It could be due to the lack of cell cycle events or a lack of potency against that cell line.

As with the implementation of any new technology or assay, the appropriate assay development should be done in order to ensure that the assay conditions provide the optimal dynamic range and reproducibility.

## 4 GENERAL PROTOCOL FOR USING THE CELLCARD SYSTEM

### Experiment Design and Plate Layout

A core software component of the CellCard system is the Experiment Manager application. This application provides a simple, easy-to-use interface

◄

**Scheme 1** The CellCard protocol starts with growing different cell types on unique codes in 6-well format. After mixing the cell cards and the attached cells they are dispensed into 96-well format. The wells are scanned after compound addition and assay reagents are added. Image analysis algorithms decode the cellular array and extract the biologically relevant data from the images.

to design and manage all aspects of the multiplexed experiment to be performed. It stores information (i.e., fluorochromes, analysis parameters, number of plates, etc.) about the assay being run in order to appropriately configure the CellCard dispenser and reader. In addition, other experimental details, for example, this software allows the user to log cellular and compound information (i.e., lot number, concentration), that are useful for downstream data analysis can also be managed with this software tool.

### CellCard Preparation

In preparation for a CellCard experiment, the CellCards are first dispensed into 6-well plates. Before adding the CellCards, disposable ladles are placed at the bottom of the 6-well plate to aid in their removal prior to mixing them and their transfer into a microtiter plate (see below). Then, the appropriate amount, determined by the number of wells to be assayed, of CellCards are added to the ladles and dispersed into a monolayer.

### Tissue Culture

The adherent cell lines to be multiplexed are maintained using standard tissue culture protocols in the media of choice. The cells should be split and accurately counted prior to plating them on the CellCards. Typically, 1 to $5 \times 10^5$ cells are seeded into the wells containing the CellCards. A range of seeding densities should be tested to ensure the proper confluence of cells for the particular assay that is to be performed. The cells should be allowed to adhere (under standard incubation conditions) for an appropriate amount of time to ensure robust adhesion to the particles. This is cell type specific and can range from 5 to 18 h. We recommend allowing the cells to adhere during an overnight incubation.

### CellCard Mixing and Dispensing into 96-Well Format

After the cells have adhered to the CellCards, they are transferred into a 15-mL conical tube for mixing. A single 360° inversion is sufficient to mix the CellCards. The conical tube and the appropriate number of 96-well plates are then loaded onto the robotic CellCard dispenser. A detailed protocol for dispensing the CellCard mixture into 96-well plate format is provided in the CellCard users manual. After the CellCards have been dispensed into 96-well format, the plates are placed into cell culture incubation conditions for recovery prior to compound addition.

### Compound Addition and Assay Protocol Considerations

Remove the 96-well plates containing the CellCards and attached cells from the incubator. Then, using a wand aspirator or a plate washer, remove media

from the wells to ensure that the same residual volume is present in all wells. We recommend leaving 50 μL of residual media in each well. Add the appropriate amount of compound to each well and mix. We typically add 50 μL of a 2× solution of the compound. After the compounds have been added to each of the wells, return the plates to the cell culture incubator for the appropriate incubation time. This incubation time is assay dependent and can range from 30 min for a nuclear translocation assay to multiple days for a proliferation assay.

In general, only minor alterations in standard assay protocols need to be made when using the CellCard system. For example, incubation times with and concentrations of staining reagents remain the same with respect to an assay run on plates without CellCards. We recommend that all washes be done as gently as possible to limit CellCard agitation and possible sloughing off of cells from the particles. In addition, we recommend that the wells are not aspirated dry. We typically leave ~20 μL in each well at the end of a wash to minimize the potential removal of CellCards from the wells.

## CellCard Reading

Before scaning the plates with the CellCard reader, the appropriate plate description file (generated in the experiment designer application) must be loaded. This will configure the reader to address the appropriate wells and acquire images of the appropriate assay-specific fluorochromes.

After the microtiter plate has been placed in the reader (either manually or via a robot), the image acquisition parameters must be set. The parameters for acquiring the bright-field images will be automatically set by the reader without user intervention. This will ensure that the CellCard codes will posses the proper hue's in the final image and robust code recognition. The default parameters for fluorescent image acquisition will be read in from the plate description file. However, these typically need to be adjusted by the user to ensure optimal image quality. A detailed protocol for this is provided in the user's manual. The acquisition parameters are set in such a way as to ensure that the images can be efficiently used by the analysis algorithms and accurately represents the biological readouts that are to be measured. If there are multiple plates in the experiment, these acquisition settings are set as the default for all subsequent plates to be imaged during the same session. Once the acquisition parameters are set, the plate is scanned by the imager and the images saved to the appropriate directories designated in the design of the experiment.

## Image Analysis and Data Visualization

To extract the cell type specific data, the images acquired by the CellCard reader must be processed by a series of image analysis algorithms. The software will read in the assay-specific default image analysis parameters and

proceed with the analysis. It is recommended that to run the analysis on at least two wells, that is, positive and negative controls, before analyzing the entire experiment. The CellCard system manual provides a detailed guide for how to accurately set the proper image analysis parameters.

The data tables generated by the CellCard analysis software are saved as a tab-delimited flat files. This data file can then be opened and visualized in the Vitra Bioscience data visualization software or other standard data visualization applications (e.g., Spotfire). In addition, the flat file structure is a standard one allowing the data to be parsed and uploaded into an enterprise database for further analysis and storage.

## 5 THEORETICAL EXAMPLE: APPLICATION OF CELLULAR ARRAYS TO THE IDENTIFICATION AND PROFILING OF CELL TYPE SELECTIVE ANTIMITOTIC COMPOUNDS

The development of antitumor drugs necessitates the determination of a compounds' potency as well as cell-type selectivity. The ability for 10 cell lines to be robustly assayed using the CellCard system enables both of these key pieces of information to be gathered simultaneously from a single microtiter well.

In addition, the multiplexing of such experiments results in a significant decrease in the time to gather data. Most proliferation assays require 48- to 72-h incubation times. Therefore, each experimental cycle from seeding cells to gathering data lasts roughly a week. Suppose there are 10 cell lines that are to be used to profile compound activity, and there are resources to run 30 plates a week. If these compounds are to be analyzed in a triplicate dose curve format (at 4 compounds and 1 cell type per plate) this means that the allocated resources can only accommodate12 compounds per week. However, if the 10 cell lines were multiplexed, the throughput would increase 10-fold. This would result in an overall throughput of 120 compounds per week. This would shorten the time required to gather the data on 120 compounds by a full month!

The choice of cell lines included in this type of profiling will determine the biological specificity being assayed. For example, one possible goal might be to determine the ability of a compound to inhibit proliferation across a series of cell lines from different tissues. The utility of this approach has been shown by the National Cancer Institute (NCI) in its compound screening efforts (*http://dtp.nci.nih.gov/index.html*). To this end multiple cell lines representing different tissues of origin would be grown on the cell cards, mixed, and assayed in each well. For example, by including two cell lines from each of the following tissues—colon, breast, lung, ovary, and prostate—one could be profiling the compounds ability to inhibit proliferation across the various tissue-derived cell lines.

Alternatively, the ultimate goal might be to determine the pathway or target selective effects of a compound. The validation of this approach was shown in

the identification of a target selective compound that selectively inhibited k-Ras transformed cells [23]. This could be expanded to include other targets as well. For example, one could generate isogenic cell lines that differ in their genotypes at a single specific locus. One cell line might harbor a p53 mutation known to be associated with cancers while another may be a p53 knockout. In the same way, other cell lines could be generated to readout other pathways. When these are screened along with their parental cell type (the one used initially to generate the isogenic strains) compounds would be assayed for their ability to elicit a target or pathway selective response.

The choice of cell types to be used in the multiplex experiment is important to be certain. One must carefully decide on the exact biological question and the nature of the selectivity that is to be determined. In addition, it is critical that the kinetics of the biological response to be measured is similar for all cell types. That is, if the assay is designed to measure proliferation, it is advantageous to match doubling times of the cell types to be multiplexed. Once chosen, the combination of cells and the assay chosen to implement using the CellCard system will provide the user with the potency and selectivity information desired.

## 6 SPECIFIC CASE STUDIES USING THE CELLCARD SYSTEM

To demonstrate the feasibility of how to use this novel technology to profile compounds, we will present two case studies. In the first, the TUNEL assay was used to analyze the proapoptotic and antiproliferative effects of staurosporine across multiple cell lines. By comparing the dose-dependent effects on these two parameters, interesting cell type specific responses were identified. In the second example, a small set of compounds were analyzed in a single-dose format to simulate a screening scenario. This experiment was designed to identify antiproliferative compounds with cell type selectivity. In addition to measuring proliferation, this multiparametric assay also measures BrdU incorporation, a measure of cell cycle activity. By using these two parameters, the screen was able to identify cell type selective antiproliferative compounds using a single well.

## 7 MULTIPLEXED DOSE CURVE ANALYSIS USING THE TUNEL ASSAY

In this experiment six cell lines were simultaneously assayed and analyzed for their response to the potent proapoptotic compound staurosporine. The assay used was the TUNEL (Terminal Uracil Nick-End Label) assay, which enzymatically labels double-stranded DNA breaks, a hallmark of late-stage apoptosis. Analysis of the cellular staining from this procedure results in an apoptotic index parameter. This measurement has been normalized to a

no-drug control and represents a fold increase over background. In addition to this measure, a DNA intercalating dye is used to label all nuclei and therefore serves not only to normalize the TUNEL data to cellular number but also to measure cell number as a measure of growth. For the growth parameter the data was normalized using the NCI paradigm (for an in-depth discussion please visit the NCI website at *http://dtp.nci.nih.gov/docs/compare/compare_methodology.html#specon*). Briefly, this approach results in a relative growth index such that a value of one indicates a cytostatic environment. A relative growth of greater than one represents proliferation, and numbers less than one indicate cell loss or cytotoxicity.

Figure 2 demonstrates how these two parameters, apoptotic index and relative growth, can be robustly derived from the same set of images. The cell line shown in Figure 2 (HT29s) demonstrates how this compound induces growth arrest and a concomitant induction of apoptosis.

However, this only represents one of the six cell lines that were present in each of the wells. Figure 3*a* shows the apopotic index and relative growth data for each of the cell lines. Immediately evident is the cell type specific differences that can be seen in these graphs. The magnitude of the responses, the slopes of the lines, the concentrations that yield maximal response all vary from cell type to cell type. A subset of these cell lines are shown in Figure 3*b*.

These graphs (Fig. 3*b*) show the response of A549, ADR-RES, and MCF7 cells. Note that despite the fact that the $ED_{50s}$ for the apoptotic response are nearly identical for each of the cell types, the magnitude of the response is significantly diminished for MCF7 cells. A simple analysis of $ED_{50}$ values would not have revealed this difference. Interestingly, even though the A549 and ADR-RES cells show superimposable responses when graphed for their apoptotic response, their relative growth profiles are different. The A549 cell proliferation curve is left shifted from the ADR-RES cells at the lower concentrations. This demonstrates that A549 proliferation is more sensitive to staurosporine than ADR-RES. The comparison of these multiple parameters in different cell lines enabled us to identify these significant cell type specific differences. The biological mechanisms underlying these differences are unknown. However, it demonstrates the power and utility of the CellCard system for comparing multiple parameters with the unequivocal knowledge that each data point is gathered from identical experimental conditions.

**Figure 2** Multiparametric analysis of the TUNEL assay. (*a*) This graph demonstrates the apoptotic response of HT29 cells to staurosporine. The data has been normalized to represent a fold increase in apoptotic signal over background. (*b*) Analysis of the nuclear stain provides a growth index. The growth index has been normalized to a time zero control (NCI paradigm) such that no relative growth, a cytostatic situation, equals one. This relative growth plot displays the antiproliferative activity of staurosporine toward HT29 cells.

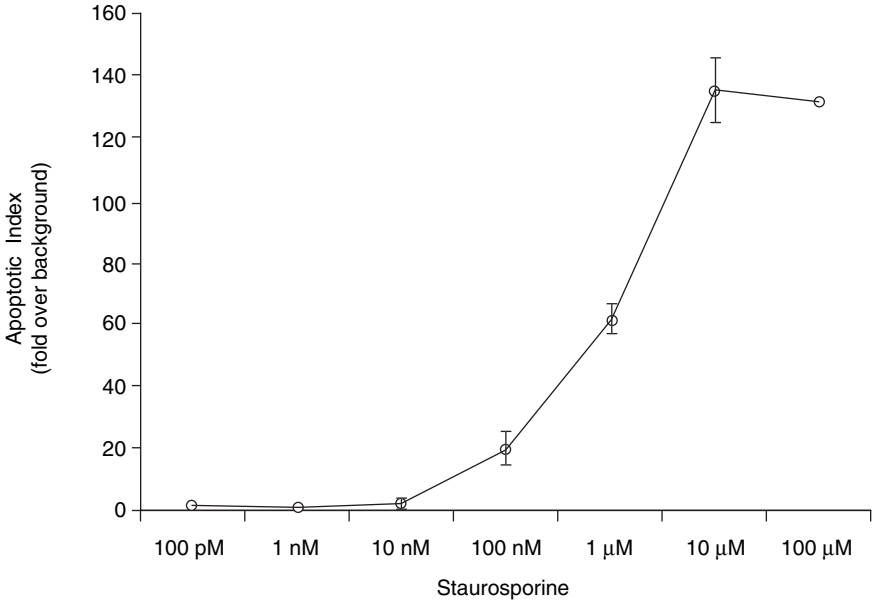
160
140
120
100
80
60
40
20
0
Apoptotic Index
(fold over background)
100 pM
1 nM
10 nM
100 nM
1 μM
10 μM
100 μM
Staurosporine

(*a*)

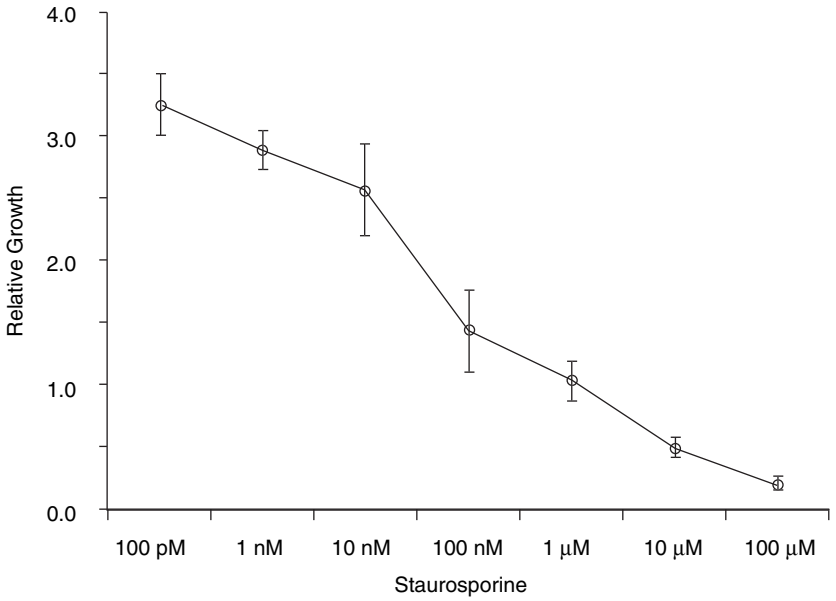
4.0
3.0
2.0
1.0
0.0
Relative Growth
100 pM
1 nM
10 nM
100 nM
1 μM
10 μM
100 μM
Staurosporine

(*b*)

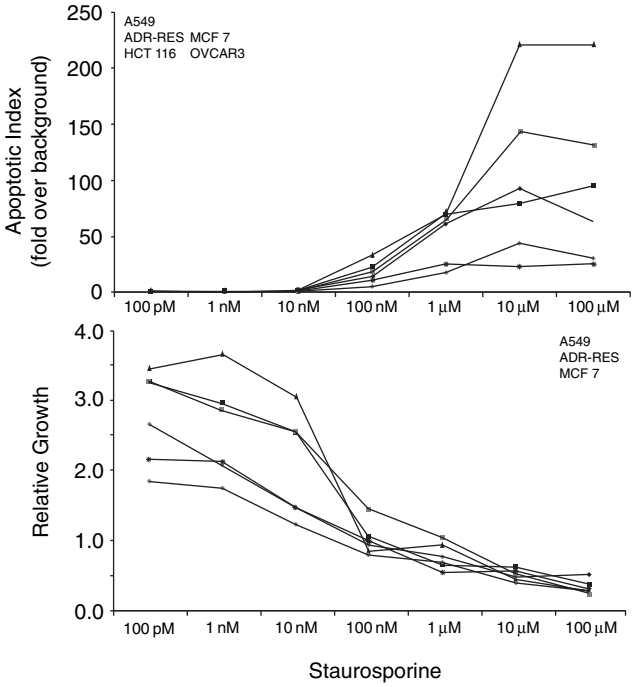


(*a*)

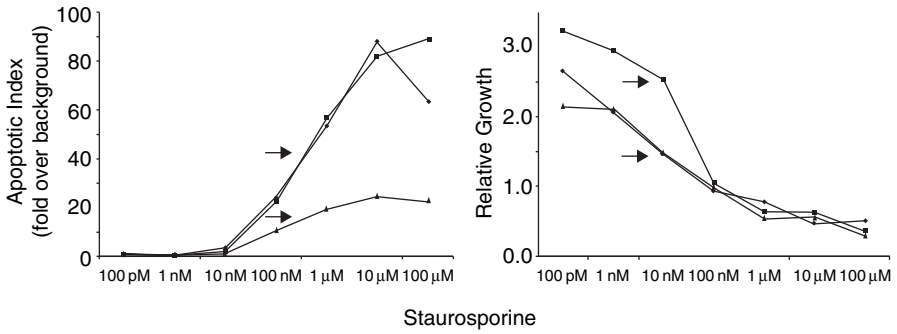


(*b*)

**Figure 3** Multiplexed dose curve analysis of staurosporine using the TUNEL assay. (*a*) Six cell lines (A549—blue, ADR-RES—pink, HT29—aqua, MCF7—maroon, and OVCAR3—brown) were simultaneously assayed for their apoptotic and proliferative responses to increasing concentrations of staurosporine. (*b*) Three of the cell lines presented in (*a*) are shown for a more direct comparison of the parameters. Comparison of the parameters shows that even though the $ED_{50}$ values for the apoptotic responses are identical for these cell lines (red arrows) their growth responses differ (blue and black arrows). Notably, all the data from each dose presented was generated from a single well. Therefore, all of this data in this figure was generated from seven wells.

## 8 MULTIPLEXED ANTIPROLIFERATION SCREEN; IDENTIFYING CELL TYPE SELECTIVE COMPOUNDS

A subset of the NCI's structural diversity set of compounds was screened for antiproliferative activity. The 10 cell lines used in this experiment came from the ATCC and represent tumors from varying tissues of origin (Table 1). Since these cell lines have a cell cycle time of approximately 20 to 30 h, a compound incubation time of 48 h was chosen. Experiments with camptothecin and taxol, known antiproliferative compounds, were performed to ensure that the incubation time and other assay parameters were appropriate (data not shown).

Each plate had a layout in which columns 2 to 11 of the 96-well plate were used for the compounds to be screened. Columns 1 and 12 contained positive and negative control conditions. The details of this layout are presented in Figure 4. The maroon squares represent negative control wells that were treated with 1 percent DMSO (dimethyl sulfoxide, the compound solvent). The green squares represent the positive control wells that were treated with 1 μM camptothecin. The other wells in columns 1 and 12 represent other controls not relevant to this discussion. The test compounds were screened at a concentration of 1 μM in 1 percent DMSO. After a 48-h incubation the cells were labeled with BrdU and stained according to protocol provided in the CellCard system manual. The plates were then imaged and analyzed using the CellCard system.

This assay provides the user with two key biological parameters for analysis; the first is proliferation. As part of the protocol, the nuclei are stained, and therefore this fluorochrome can be used to assess cell number and therefore the amount of cellular proliferation. In addition, the assay measures BrdU incorporation, a general measure of DNA polymerase activity. This activity is usually associated with the DNA synthesis phase (S-pahse) of the cell cycle,

**TABLE 1 The Cell Types and Tissues of Origin Used in a Screen for Cell Type Selective Anti-Proliferative Compounds**

| Cell Type | Tissue of Origin |
|---|---|
| A549 | Lung |
| Adr-Res | Unknown |
| HCT116 | Colon |
| HT29 | Colon |
| M14 | Skin |
| MCF7 | Breast |
| OVCAR3 | Ovary |
| OVCAR5 | Ovary |
| SKMEL2 | Skin |
| SKMEL28 | Skin |

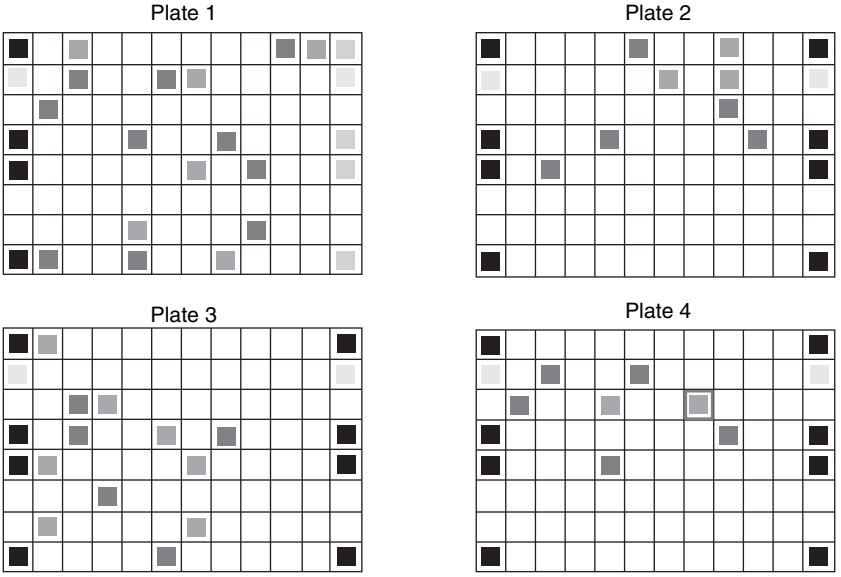


**Figure 4** Summary of the plate layout and results from a compound screen designed to identify cell type selective antiproliferative compounds. The maroon squares represent negative control wells while the green squares are wells treated with 1 μM camptothecin, a known antiproliferative compound. The library compounds were dispensed in columns 2 to 11 at a final concentration on 1 μM. The blue squares represent nonselective antiproliferatives, whereas the red squares represent wells with cell type selective antiproliferative activity. The data from well C8 on plate 4 is shown in Figure 5.

and therefore quantification of BrdU incorporation is returned as an S-phase index. Since each cell type has different absolute values for these measures, the data was normalized to the no-drug (negative) controls. The normalization routine results in a value that is the percentage of the no-drug control. Therefore, these controls represent 100 percent activity. An example of these control values for plate 4 are shown on the left side of Figure 5.

In Figure 4 the red and blue squares show the hits that were identified on these plates. A hit was defined as any compound that inhibited proliferation of at least one cell type by at least 50 percent. The blue squares were nonselective compounds. That is, they inhibited the proliferation of all cell types by 50 percent or more. On the other hand, the red squares represent those compounds whose antiproliferative activity showed some cell type selectivity. In order to be selective the compound was required to have at least a twofold differential effect between at least two cell types. In addition, the cell type least effected by the compound had to show less than 50 percent inhibition (greater than 50 percent of control proliferation). As is evident from this figure, using the CellCard system compounds were simultaneously analyzed for their

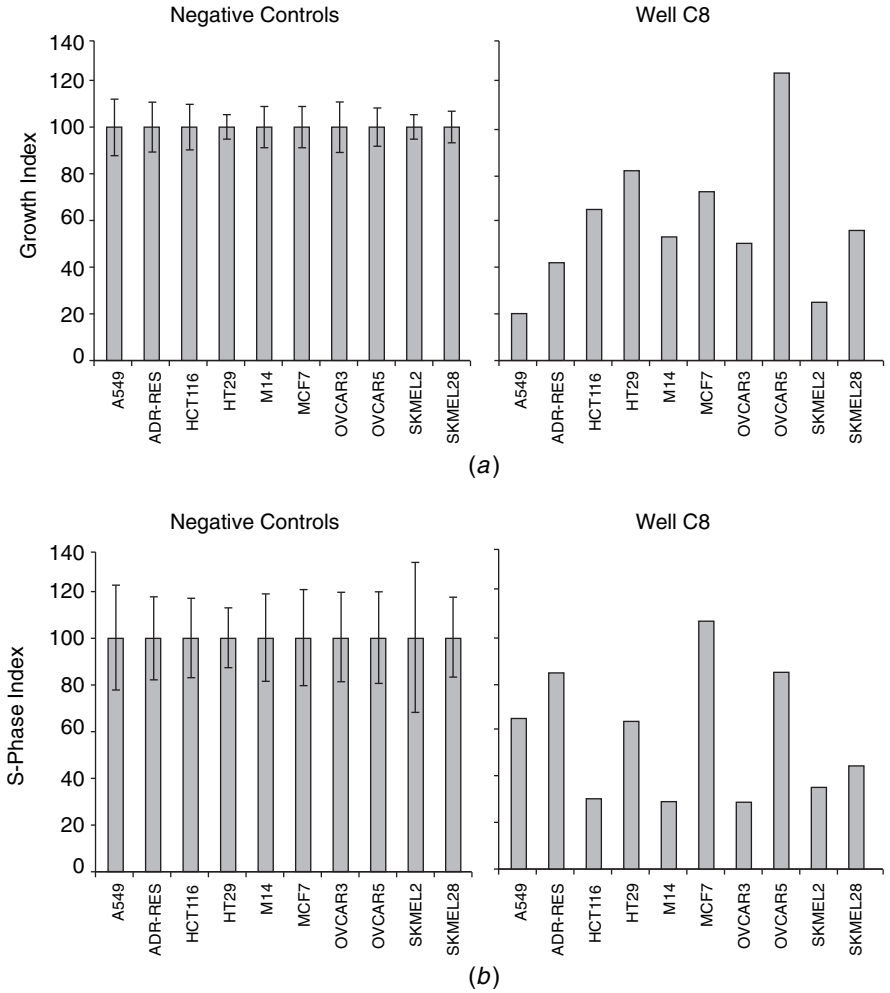


**Figure 5** Single-well compound profiling. (*a*) Graphs of the growth index parameter. The graph on the left represents the average and standard deviations of the negative control wells on plate 4. The graph on the right shows the cell type sepcific profile of activity derived from well C8 on plate 4. (*b*) Graphs of S-phase index. As above, the graph on the left represents that data from the negative controls on plate 4. Similarly, the graph on the right shows the cell cycle (S-phase index) data generated from well C8 on plate 4.

potency and selectivity. This resulted in the identification of cell type selective antiproliferative compounds. Most importantly, the negative control wells showed very reproducible results on these plates.

The data from a single well, C8, in plate 4 is shown in Figure 5. The graph in Figure 5*a* demonstrates the variability of the antiproliferative response

across the cell lines in this well and demonstrates the selectivity of the compound. To verify these results, the BrdU incorporation data was also analyzed. Keep in mind, both parameters are obtained from a single well and, therefore, the data from one parameter is directly comparable to the other. Further analysis shows that the BrdU data closely correlates with the proliferation data. For example, the OVCAR5 cell line displays no inhibition of proliferation or decrease in S-phase index when exposed to this compound. On the other hand, both the proliferation and S-phase index of the SKMEL2 cell line were significantly inhibited by this compound. Therefore, the use of the S-phase index as a confirmatory parameter served to provide a second data point on which to confirm the hits. When applied to a larger scale screening environment, this type of confirmatory analysis from the same cellular population should serve to significantly lower false-positive rates and provide a means to identify potential false negatives.

This screen yielded 20 unique, biologically relevant data points from each microtiter well. The direct comparison of these data points enables the identification of unique cell type specific difference and stimulates the design of unique screens. For example, hit compounds can be defined based on their profile across multiple parameters as well as cell types without calling into question the well-to-well, plate-to-plate, and/or day-to-day variability of the data. The data across cell types is generated within a single well in identical reaction conditions with a concomitant savings in reagents and time. Furthermore, depending on the cell lines chosen for the screen, biological mechanism of action could be screened for as well.

## 9 SUMMARY OF KEY FEATURES OF CELLCARD SYSTEM

The CellCard system developed by Vitra Bioscience provides a novel platform for cell-based assays. At the core of the platform are encoded particles referred to as CellCards. By growing individual cell types on CellCards with unique codes, the platform enables multiple cell types to be assayed simultaneously in 96-well format. This results in a miniaturization of the assays by decreasing the overall number of cells, reagents, and time required for each data point. This is especially valuable when rare reagents (i.e., primary cells, natural product compounds) are at stake. In this case, they can be used more efficiently and more data gathered from a single reagent lot. The case studies presented here demonstrate how comparing multiple parametric measurements across multiple cell types is a powerful way to profile compounds and identify cell type specific differences. Taken together, the data can be confidently compared knowing that some sources of variability (i.e., day-to-day) cannot be responsible for cell type specific differences identified.

Moving forward, the CellCard technology will continue to drive novel screen designs and compound profiling strategies.

## REFERENCES


1. Gonzalez, J. E., Negulescu, P. A. (1998). Intracellular detection assays for high-throughput screening. *Curr. Opin. Biotechnol.*, *9*(6), 624–631.
2. Hertzberg, R. P., Pope, A. J. (2000). High-throughput screening: New technology for the 21st century. *Curr. Opin. Chem. Biol.*, *4*(4), 445–451.
3. Zaccolo, M., De Giorgi, F., et al. (2000). A genetically encoded, fluorescent indicator for cyclic AMP in living cells. *Nat. Cell. Biol.*, *2*(1), 25–29.
4. Durick, K., Negulescu, P. (2001). Cellular biosensors for drug discovery. *Biosens. Bioelectron.*, *16*(7–8), 587–592.
5. Numann, R., Negulescu, P. A. (2001). High-throughput screening strategies for cardiac ion channels. *Trends Cardiovasc. Med.*, *11*(2), 54–59.
6. Tretyn, A., Kado, R. T., et al. (1997). Loading and localization of Fluo-3 and Fluo-3/AM calcium indicators in sinapis alba root tissue. *Folia Histochem. Cytobiol.*, *35*(1), 41–51.
7. Smart, D., Jerman, J. C., et al. (2001). Characterisation using FLIPR of human vanilloid VR1 receptor pharmacology. *Eur. J. Pharmacol.*, *417*(1–2), 51–58.
8. Whiteaker, K. L., Gopalakrishnan, S. M., et al. (2001). Validation of FLIPR membrane potential dye for high throughput screening of potassium channel modulators. *J. Biomol. Screen.*, *6*(5), 305–312.
9. Baxter, D. F., Kirk, M., et al. (2002). A novel membrane potential-sensitive fluorescent dye improves cell-based assays for ion channels. *J. Biomol. Screen.*, *7*(1), 79–85.
10. Sullivan, E., Tucker, E. M., et al. (1999). Measurement of [Ca2+] using the fluorometric imaging plate reader (FLIPR). *Methods Mol. Biol.*, *114*, 125–133.
11. Sharif, N. A., Kelly, C. R., et al. (2003). Bimatoprost (Lumigan((R))) is an agonist at the cloned human ocular FP prostaglandin receptor: Real-time FLIPR-based intracellular Ca(2+) mobilization studies. *Prostaglandins Leukot. Essent. Fatty Acids*, *68*(1), 27–33.
12. Ding, G. J., Fischer, P. A., et al. (1998). Characterization and quantitation of NF-kappaB nuclear translocation induced by interleukin-1 and tumor necrosis factor-alpha. Development and use of a high capacity fluorescence cytometric system. *J. Biol. Chem.*, *273*(44), 28897–28905.
13. Conway, B. R., Minor, L. K., et al. (1999). Quantification of G-protein coupled receptor internatilization using G-protein coupled receptor-green fluorescent protein conjugates with the ArrayScantrade mark high-content screening system. *J. Biomol. Screen.*, *4*(2), 75–86.
14. Conway, B. R., Demarest, K. T. (2002). The use of biosensors to study GPCR function: Applications for high-content screening. *Receptors Channels*, *8*(5–6), 331–341.
15. Kapur, R. (2002). Fluorescence imaging and engineered biosensors: Functional and activity-based sensing using high content screening. *Ann. NY Acad. Sci.*, *961*, 196–197.
16. Dove, A. (2003). Screening for content—The evolution of high throughput. *Nat. Biotechnol.*, *21*(8), 859–864.

17. DeRisi, J., Penland, L., et al. (1996). Use of a cDNA microarray to analyse gene expression patterns in human cancer. *Nat. Genet.*, *14*(4), 457–460.
18. Ziauddin, J., Sabatini, D. M. (2001). Microarrays of cells expressing defined cDNAs. *Nature*, *411*(6833), 107–110.
19. Eickhoff, H., Konthur, Z., et al. (2002). Protein array technology: The tool to bridge genomics and proteomics. *Adv. Biochem. Eng. Biotechnol.*, *77*, 103–112.
20. Wu, R. Z., Bailey, S. N., et al. (2002). Cell-biological applications of transfected-cell microarrays. *Trends Cell. Biol.*, *12*(10), 485–488.
21. Beske, O. A. G. S. (2002). High-throughput cell analysis using multiplexed array technologies. *Drug Discov. Today*, *7*(18), S131–S135.
22. Beske (2004).
23. Torrance, C. J., Agrawal, V., et al. (2001). Use of isogenic human cancer cells for high-throughput screening and drug discovery. *Nat. Biotechnol.*, *19*(10), 940–945.

# PROTEIN X-RAY CRYSTALLOGRAPHY IN DRUG DISCOVERY

PETER NOLLERT, MICHAEL D. FEESE, BART L. STAKER, AND HIDONG KIM

*deCODE biostructures*
*Bainbride Island, Washington*

1 INTRODUCTION
2 CRYSTALLIZATION
Background
Formats
Membrane Proteins
Crystallization Factors
Observation and Documentation
Crystallization Formulation Screens
Cryopreservation for Cryocrystallography
Crystal-Based Drug Discovery
Derivatization for Anomalous Diffraction Experiments
Crystallization Data Sources
Protein Crystallization Demonstration Experiment
3 X-RAY DIFFRACTION EXPERIMENT
Diffraction Theory
X-ray Diffraction by Macromolecule Crystals
Fourier Synthesis in X-ray Crystallography
Bragg's Law and the Angular Dependence of X-ray Diffraction
Electron Clouds and Thermal Motion
X-ray Diffraction Data Collection in Practice
Home Laboratory Diffraction Data Collection
Synchrotron Diffraction Data Collection
4 X-RAY CRYSTAL STRUCTURE DETERMINATION
Phase Problem

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

Structure Factor 405
Heavy Atom Replacement Methods 406
Multiple Isomorphous Replacement 407
Anomalous Dispersion Methods 412
Molecular Replacement 416
5 GENERATION AND ANALYSIS OF STRUCTURAL MODELS 420
Aspects of Crystallographic Models of Macromolecules 420
Building the Initial Model 422
Refinement and Analysis of Structural Models 425
Analysis and Preparation of Structural Models 432
Crystallography–Drug Discovery Interface 433
6 EXAMPLES FOR THE USE OF X-RAY CRYSTALLOGRAPHY IN DRUG DISCOVERY 434
Lead Optimization—Structure-Based Drug Design 436
Antistructures 441
Protein Therapeutics 442
In silico Screening Based on Crystallographic Structural Models 442
Crystallographic Screening 443
Crystallographic Fragment Screening 444
Site-Directed Leads via Fragment Tethering 445
Structural Genomics 446
7 LIMITATIONS AND CHALLENGES OF X-RAY CRYSTALLOGRAPHY IN THE DRUG DISCOVERY PROCESS 447
References 450

## 1 INTRODUCTION

The fundamental goal of applying protein X-ray crystallography to drug discovery is to increase its speed, quality, and its rate of success while reducing the cost. In conjunction with molecular biology, protein X-ray crystallography forms a seamless interface between target and lead discovery. Nucleic acid sequences from genetic studies provide the entry point for protein expression, the first step in crystallographic projects. The product of the crystallographic endeavor, the *structure*, is the accurate description of protein and ligand atom positions in three-dimensional space. Starting in the 1980s, protein X-ray crystallography has impacted the drug discovery process, primarily in the stage of lead optimization. Recently, however, by means of innovation and integration, protein crystallography has been transformed into an enabling technology now covering several stages of the drug discovery process (Fig. 1). Besides lead optimization, protein X-ray crystallography has affected drug discovery in (a) the identification of new drug targets, (b) the understanding of molecular target mechanisms, and (c) the discovery of new lead compounds.

The integration of protein X-ray crystallography into the early discovery process by joining it with chemical synthesis and assaying has become a very

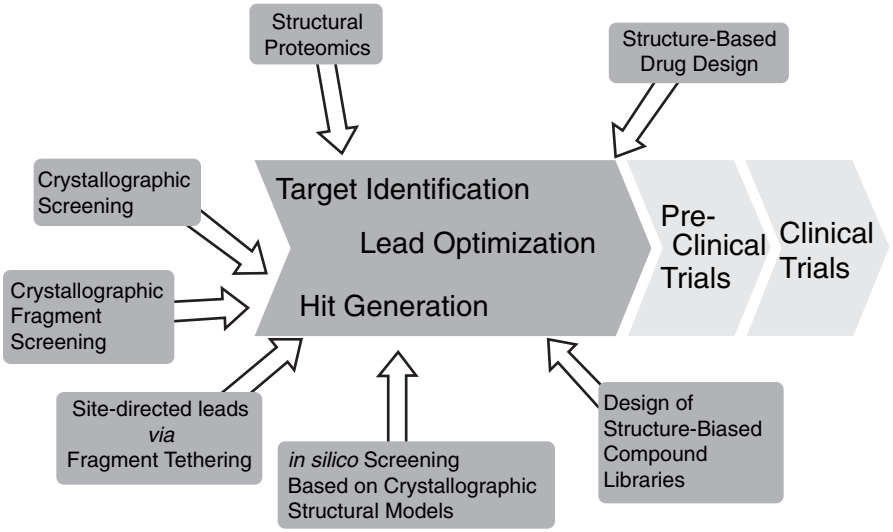


**Figure 1** Impact of protein X-ray crystallography on and use of structural information for drug discovery purposes.

effective tool for lead optimization. Chemical synthesis can be iteratively directed toward more promising compounds and away from undesirable compounds based on insight into the interaction of ligands with the target protein. Thus, structure-based lead optimization has positively affected potency, selectivity profile, or pharmacokinetic properties of drug candidates. The application of protein X-ray crystallography impacts lead discovery by methods such as (a) X-ray crystallographic screening, (b) fragment screening, and (c) site-directed fragment tethering. Furthermore, structural information is being used to design and test target structure-based compound libraries (e.g., compound libraries specific for kinases) and to screen ligand compounds in silico with the goal of enriching useful compounds in virtual libraries prior to high-throughput assay-based screening.

*X-ray crystallographic screening* [1] combines the steps of lead identification, structural assessment, and subsequent lead optimization by structure determination of crystals that are soaked in compound cocktails. Ligands with highest affinities are selected by the crystal, and they are identified crystallographically. The utility of this strategy is exemplified in a section below.

*Fragment screening* is a variation of this theme and serves as a tool for assembling new lead compounds from small druglike fragments. Initially, libraries of leadlike fragments are formulated and co-crystallized or soaked in crystals. The resulting structures serve as a basis for combining novel chemical leads. The advantage over conventional assay-based screening is that very low affinity fragments with novel structures can be found.

*Site-directed lead discovery by fragment tethering* [2, 3] adds an additional layer of complexity to fragment screening, allowing the discovery of low-

molecular-weight ligands. The strategy is based on the covalent modification of a target protein at a particular site on its surface via thiol-chemistry and the mass-spectrometric detection of weakly binding ligand precursors. Crystallography provides the tool to observe the binding mode and to direct the chemical synthesis of fused analogs, the latter having potentiated affinity.

*Structure-based drug design* consists of iterative processes aimed at the optimization of existing leads. The process includes the choice and the determination of a target structure, choosing a method for lead discovery and, crucially, the evaluation of drug leads [4]. A schematic description of the iterative process involved in structure-based drug design is shown in a later section (Fig. 22) [4]. The success of structure-based lead optimization is measured by improvements of affinity, specificity, or ADME properties.

*Structure-based compound libraries* can be obtained by so-called in silico screening. Crystallographic models are used to enrich compound libraries prior to conventional high-throughput screening [5]. Ligand structures are "docked" in silico into the binding pockets of protein structures, and the resulting modeled complexes often identify the correct ligand binding mode. More importantly, sets of accordingly treated compounds can be ranked by applying scoring functions that estimate ligand affinity, allowing the enrichment of actual chemical compound libraries with potentially good binders.

*Structural genomics* efforts seek to determine protein structures on the genome scale. The availability of many X-ray crystallographic structures is expected to deeply impact the discovery of new drugs and targets. Anticipated benefits are, for example, the discovery of new targets by assignment of functions to orphan targets, the improved quality of homology models for difficult targets, and the use of *antitargets* to decrease ligand cross-reactivity.

This chapter seeks to familiarize scientists with X-ray protein crystallographic techniques and their exemplified application for the purpose of drug discovery. Outlined are the individual steps that are required for determining X-ray crystallographic structures of proteins and protein–ligand complexes. First, the generation of crystals and their use in the X-ray diffraction experiment is described. Then state-of-the-art methods for crystal structure determination, model generation, and their refinement are reviewed, followed by an account of various ways of analyzing the resulting structural models. Several prominent examples are presented where X-ray crystallography has aided drug discovery. Finally, the limitations of the method are discussed.

## 2 CRYSTALLIZATION

Crystallography is the science of crystals. The word *crystallography*, a composite, derives from the Greek words *crustallos* or κρυστταλος and *graphos* or γραφος for *writing*. The word *crustallos* meaning "frozen" or "clear ice," captures several macroscopic crystal properties of these fascinating materials. Besides being transparent and solid, crystals are often facetted and, most

importantly, have a high degree of internal order. They consist of regular three-dimensional lattices of molecules (see later section about the fundamentals of X-ray diffraction) and protein molecules with or without their bound drug compounds. Protein crystals consist of ordered protein and usually a sizeable fraction of disordered solvent, water. How are such crystals obtained?

Although there are no fixed rules as to how to grow protein crystals, the advent and combination of molecular biology, affinity tag purification tools, and laboratory liquid dispensing automation has helped to define generalized crystal growth procedures. Every protein, however, is different and a certain method and condition that works well for one particular protein is usually not applicable to the crystallization of other proteins. Furthermore, it is not possible to predict crystallization conditions on the basis of sequence. Therefore, the de novo crystallization of a particular protein poses a significant challenge.

While crystallographers have used systematic approaches to search for crystallization conditions, in most cases, preformulated screening kits are used (Table 1). Once a crystallization hit has been identified in a crystallization scouting trial, these initial conditions are often refined in order to grow crystals that are suitable for X-ray diffraction experiments. On the other hand, many protein crystallization conditions have been reported in the scientific literature and in databases (see below) and are therefore much simpler to carry out. Finally, comprehensive and practical advice on crystallization experiments can be found in Bergfors [6], McPherson [7, 8], and in Ducruix and Giegé [9].

## Background

Protein crystallization occurs in three stages, nucleation, growth, and cessation of growth [10, 11]. Proteins crystallize from supersaturated solutions where the concentration exceeds their equilibrium solubility. The state of supersaturation depends on many factors such as the concentration and nature of the protein in question but also on salts and other components of the solution. It is this state of supersaturation that needs to be created in a crystallization experiment—the so-called crystallization setup—for crystal nuclei to form and crystals to grow (Fig. 2). The underlying principle is to alter the properties of the solvent (water), disrupt the interaction of water molecules with protein molecules, and increase the attractive interaction among protein molecules. This is generally done with (a) salts, (b) organic solvents, or (c) polymer compounds. When salts dissolve in water, their ions become hydrated and capture water molecules, the latter of which become unavailable to interaction with proteins. Therefore, the addition of salts such as ammonium sulfate may be used to favor protein–protein interactions, to form crystallization nuclei, and support crystal growth, thus *salting out* the protein.

Intriguingly, the opposite procedure, *salting in*, the removal of ions from a protein solution may also be used for the crystallization of proteins. Ions balance protein surface charges, and, once they are removed, protein molecules can balance their electrostatics by interacting with each other [12].

**TABLE 1 Selection of Vendors of Crystallography-Related Products**

| Category | Items | Vendor / Company |
|---|---|---|
| Crystallization tools | Crystal screening formulations, various crystallography supplies | Hampton Research |
| | Crystal screening formulations, crystallization plates | Emerald BioSystems |
| | Crystal screening formulations, | Jena Bioscience |
| | Microfluidic crystallization devices | Fluidigm |
| | Crystallization plates | Nextal |
| | Crystallization plates | Corning |
| | Crystallization plates | Greiner |
| Crystallization trial robotics | CrystalMiner (database application for crystallization trials) | Emerald BioSystems |
| | CrysTel (integrated incubation and imaging system) | |
| | CrystalMonitor (automated crystallization imaging workstation) | |
| | MatrixMaker (automated solution formulation robot for the formulation of crystallization screening kits) | |
| | RoboFill (automated crystallization) | RoboDesign |
| | Odyssey (crystallization incubation and imaging system) | |
| | RoboMicroscope II (automated crystallization imaging workstation) | |
| | CrystalScore (automated crystallization imaging workstation) | Diversified Scientific |
| | 925 PC Workstation (crystallization setup automation) | Gilson |
| | Crystal Farm (integrated incubation and imaging system) | Discovery Partners |
| | Rock Maker (software application for crystallization trials) | International |
| | Crystallization automation | Douglas instruments |
| X-ray diffraction instrumentation | X-ray diffraction instrumentation, detectors, optics and detectors, robotic sample changing system | Bruker |
| | X-ray diffraction instrumentation, detectors, optics, and detectors, robotic sample changing system | Mar Research |
| | X-ray diffraction instrumentation, detectors, optics, and detectors, robotic sample changing system | Rigaku MSC |
| | X-ray optics | Osmic |

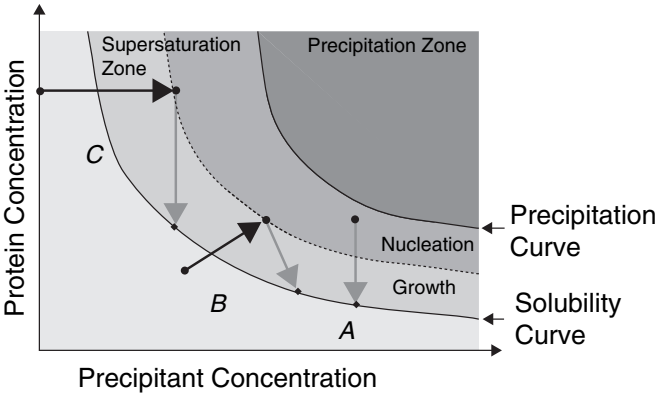


**Figure 2** Schematic phase diagram of protein crystallization in the salting-out regime. Crystal nucleation is a critical phenomenon that may occur only in a certain area of the supersaturation zone and crystals may grow under conditions of supersaturation once nuclei have formed. (*A*) Pathway of a batch-type crystallization experiment. By mixing protein with the precipitant solution, the protein becomes supersaturated. Crystal nuclei form and crystals grow until the protein concentration in solution is saturated. (*B*) Pathway of a vapor diffusion-type crystallization experiment. The slow concentration process—via vapor diffusion—that follows mixing of the protein with the precipitant solution causes the protein to become supersaturated. The vapor diffusion process causes a concomitant increase in precipitant concentration thus extending the crystal growth process. (*C*) Pathway of a dialysis-type crystallization experiment. As the precipitant diffuses into the protein chamber, a state of protein supersaturation is reached. Once nuclei have formed, protein crystals may grow as long as the protein concentration remains supersaturated.

Organic solvents also bind water molecules, but in addition they lower the dielectric constant in the crystallization solution, thus enhancing electrostatic interactions between protein molecules. Finally, crowding agents such as hydrophilic polymers compete with protein molecules for hydration. This arises directly from water binding and indirectly via excluded-volume effects.

The mechanism of crystal nucleation is poorly understood. Once a protein solution is supersaturated and nuclei are present, crystals may grow on preformed crystal faces by deposition into lattice positions via diffusion and convection [13]. The crystalline state of matter is generally considered the thermodynamically most stable state, and the crystallization process can therefore be understood in terms of free energy minimization. Indeed, a lowering of the free energy by 12 to 25 kJ/mol has been measured for protein crystallization processes [14]. The last process in crystallization, the cessation of growth, may occur for a number of reasons such as limited protein supply or the poisoning of crystal growth surfaces by contaminations and crystalline defects.

## Formats

A multitude of crystallization methods and formats are available, with batch and vapor diffusion being the most popular ones (Fig. 3). Since the quantity of protein sample is often limited, microcrystallization methods have been devised for batch, vapor diffusion, and other, more exotic, crystallization methods. They can all be carried out manually using appropriate plasticware (Table 1) or by appropriate liquid dispensing instrumentation such as multichannel pipettors or dispensing robots. Comparative studies of protein crystallization by vapor diffusion and microbatch techniques yielded a similar effectiveness of the two methods [15]. Some methods, however, tend to succeed for certain targets while other methods are more useful for different protein targets.

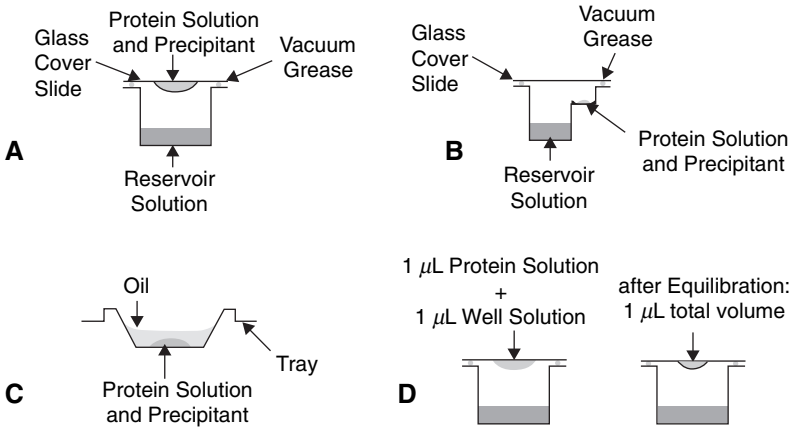


**Figure 3** Schematic depiction of typical formats for vapor diffusion and batch crystallization experiments. (*a*) Hanging drop format of a vapor diffusion crystallization experiment. A small drop containing the protein and precipitant solution is applied onto a glass slide. The glass slide is attached to the crystallization chamber, the bottom of which is filled with the so-called reservoir solution. The concentrations of volatile components in the drop and reservoir solutions equilibrate over time through the vapor phase. Crystals form in the hanging drop. (*b*) Sitting drop format of a vapor diffusion crystallization experiment. A small drop containing the protein and precipitant solution are placed into a well. The crystallization chamber may be sealed with a glass cover slide or with transparent tape. Well and reservoir communicate via the vapor phase. Crystals form in the sitting drop. (*c*) "Under oil" format of a batch crystallization experiment. A small drop containing protein and precipitant solution is placed into a crystallization well. A layer of oil prevents dehydration. Crystals form in the drop. (*d*) Setup and equilibration of a vapor diffusion crystallization experiment with the hanging drop format. A 2-μL hanging drop (consisting of 1-μL protein solution and 1-μL precipitant solution from the reservoir) equilibrates via the vapor phase. Water transport occurs from the hanging drop to the reservoir because the hydrostatic pressure in the hanging drop is initially lower than in the reservoir solution. At equilibrium the hanging drop has a volume of 1 μL, with protein and precipitant concentration similar to that of the initially separated components.

*Batch-type crystallization setups* are conceptually simple experiments. Here, a protein solution is combined with a precipitant solution and crystals form within this solution. In order to prevent dehydration, oils may be added to seal off the crystallization experiment [16, 17] (Fig. 3*c*). Popular crystallization trays for microbatch crystallizations are so-called Terazaki plates, which provide small wells that can be used to hold final trial volumes of less than 1 μL [15] and 5 to 10 μL oil (silicon oil, paraffin oil, or mixtures). Oddly, the state of supersaturation is reached by dilution of the protein solution with a precipitant solution (*A* in Fig. 2). Therefore starting protein concentrations are often required to be chosen higher for this crystallization method than for the other methods.

*Vapor diffusion crystallization experiments* can be carried out in sitting drop or in hanging drop format. While the hanging drop arrangement is less prone to crystals sticking to a solid surface, it is less cumbersome to set up sitting drop crystallizations. In a hanging drop crystallization experiment, the protein solution is combined with the precipitant solution on a glass cover slide, inverted and, attached to a well containing the reservoir solution (Fig. 3*a*). Prior to the attachment, the rim of the well is beaded with vacuum grease in order to prevent dehydration of the crystallization chamber. The sitting drop format is very similar. Specialized crystallization plates are used that provide a platform to hold the drop (Fig. 3*b*). The chamber may be sealed off with transparent tape or with vacuum grease and a glass cover slide.

In vapor diffusion experiments, the precipitating solution is usually filled into the reservoir. Once protein and precipitation solution are combined and the crystallization chamber is sealed, the crystallization drop and the reservoir solution communicate via the vapor phase. At a typical 1:1 mixing ratio, the concentration of precipitant in the crystallization drop is only half of that in the reservoir solution (Fig. 3*d*). During the course of equilibration, the concentration of the precipitant increases to that in the reservoir, and, concomitantly, the drop volume decreases by about half (*B* in Fig. 2). The kinetics of this equilibration process and the pathway to supersaturation depend on the type of precipitant, the crystal chamber and drop geometry, and temperature. Crystals may form during or after equilibration.

*Protein crystallization by dialysis* is somewhat less popular but provides a high degree of control in the crystallization process [18]. First, a protein solution is filled into a small chamber that is closed by a dialysis membrane. Then a precipitant solution is brought into contact with that membrane and an exchange of small molecules and ions can occur. Exchange of proteins and other macromolecules are excluded by the dialysis membrane. Thus, salt concentrations can be increased slowly without *overshooting* into the precipitation zone (*C* in Fig. 2).

## Membrane Proteins

About half of all drug targets constitute membrane proteins [19], proteins that are inserted in cellular membranes where they support critical functions such

as solute transport, energy conversion, and signal transduction. Unfortunately, not a single structure of a transmembrane protein of pharmaceutical relevance is available to date. This is a result of the inherent difficulties of working with membrane proteins in the stages of expression, purification, and crystallization. There is, however, no principle reason for such proteins to systematically fail crystallization attempts. Indeed, many structures of membrane proteins have been determined once concerted efforts were taken, including structures of homologous proteins of the most prominent class of transmembrane protein drug targets, GPCRs (G-protein-coupled receptors). Several methods have been devised for the crystallization of this class of proteins, including the use of protein–detergent complexes, bicelles, and lipidic cubic phases [20, 21]. The latter method employs a lipid matrix as a crystallization facilitator, and the practicalities are described in detail by Nollert et al. [22].

## Crystallization Factors

What are the factors that affect the crystal nucleation and growth processes? Of course, the quality and purity of the protein sample is of great importance since impurities may interact and disturb crystal packing. Fundamentally, the protein sequence, its organization into domains, and oligomerization state are decisive factors. The purity of proteins is usually judged in a semiquantitative way by sodium dodecyl sulfate (SDS) polyacrylamide gel electrophoresis (PAGE) and Coomassie brilliant blue staining. Purity levels above 90 to 95 percent are considered acceptable. For some proteins, however, crystallization can be a purification method, and crystals of minor contaminants have been grown from mixtures of many proteins. Particle homogeneity, that is, the uniformity of the oligomeric state is also an important factor, usually characterized by dynamic light scattering or by size exclusion chromatography (SEC). The latter chromatography is often applied as a final so-called "polishing step" in protein purification. Prior to use, protein samples are often centrifuged or filtrated to remove particulate contaminants. In order to prevent degradation, protease inhibitors and antimicrobials such as sodium azide or potassium cyanide (0.02 percent final concentration) are sometimes added. The concentration of the protein and the precipitation agent used have a tremendous effect on crystallization (Fig. 2). It is advisable to start crystallization experiments at high protein concentrations (greater than 10 mg/mL) and lower these once precipitate is detected in crystallization setups (Fig. 4*b*). Conversely, the protein and precipitant concentration may need to be increased if the drop remains clear (Fig. 4*a*).

Recombinant proteins are often expressed with tag fusions to assure high expression yield, solubility, protection from proteolysis, improved folding, and simple purification via affinity chromatography [23]. Their presence can either aid or impede crystallization, depending on the nature of the fusion, linker, and host protein. Several crystal structures have been obtained with their uncleaved large fusion proteins (including maltose-binding protein, thioredoxin, and glutathione-*S*-transferase) intact, where short three to five amino

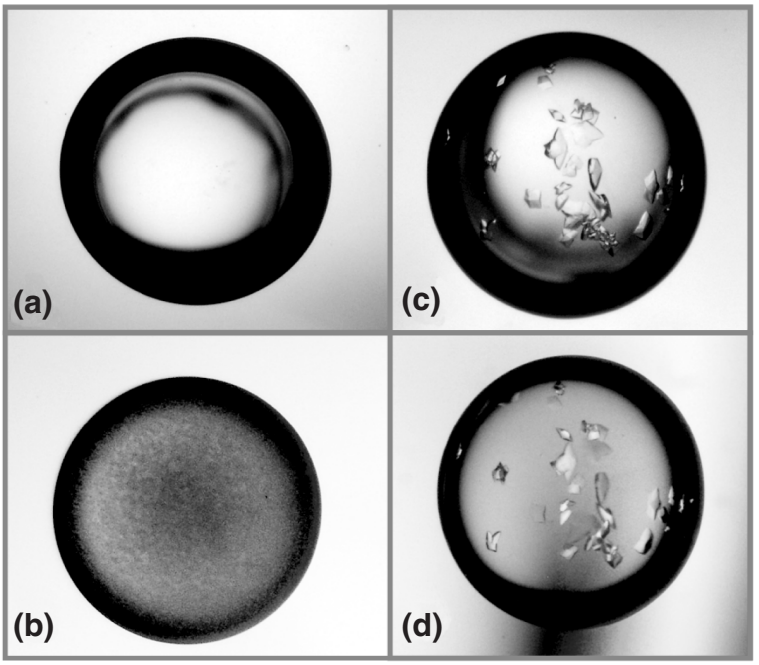


**Figure 4** Images of hanging drop crystallization experiments. (*a*) Clear 1-μL drop at the outset of the crystallization experiment. (*b*) Precipitate. (*c*) Crystals of lysozyme inside a hanging drop. (*d*) Hanging drop with birefringent lysozyme crystals, imaged under cross-polarization setting. (This figure is available in full color at ftp://ftp.wiley.com/public/sci_tech_med/drug_discovery/.)

acid linkers were employed [24]. Alternatively, these tags may be removed with specific proteolytic enzymes that cleave appropriately engineered linker sites between the tag and the host protein. Once the protein is purified and concentrated, it may be stored via rapid freezing in liquid nitrogen. A cooling procedure proved beneficial that employs protein solution volumes below 50 μL and 0.2 mL ultrathin-walled polymerase chain reaction (PCR) tubes [25]. Fast thawing to room temperature was critical in order to prevent precipitation.

Besides precipitant type and concentration, further factors that affect crystallization include temperature, buffer type and concentration, the presence of additives, crystallization format, geometry, and other environmental parameters. It is not possible to screen all of these factors systematically. Several hundred crystallization experiments are therefore usually carried out varying the temperature (4 and 20°C) and formulations of precipitating agents, while all other parameters are kept constant. Subsequent fine screening may then be accomplished by systematically screening other factors. Rather exotic factors such as electric and magnetic fields have been identified to affect crystal quality. Their systematic use in the crystallography laboratory, however, is limited.

### Observation and Documentation

Since it is often not known at the outset of crystallization trials how long it takes for protein crystals to grow, crystallization experiments are evaluated multiple times, days to months after setup. Crystallization experiments are routinely inspected visually with the aid of dissecting microscopes. These are operated under transmission illumination and often a pseudo-dark-field illumination option or crossed polarizers are used to increase crystal detection (Fig. 4). Microcrystals with dimensions of less than 5 μm may be detected with such instruments. The outcome of crystallization experiments may be classified according to preset schemes, for example, "clear," "precipitate," "crystal," or "other," while crystals may be described on the basis of their shape (the crystal *habit*), size, and growth features. Taking images of crystals and crystallization setups helps documenting the progress of crystallization. Specialized software, such as Emerald Biosystem's CrystalMiner is available to assist in keeping track and in documenting crystallization trials. Its use in combination with automated imaging workstations such as CrystalMonitor is a very powerful tool for efficiently managing crystallization experiments. Although several algorithms are currently being developed by different groups to automate crystal detection, the optical inspection by crystallographers remains the main inspection route to date.

Intriguingly, the morphology of crystals need not correlate with their utility for X-ray diffraction purposes. Generally, large (>100 μm) and faceted protein crystals yield higher resolution diffraction data than small (<50 μm) crystals. Anecdotal evidence indicates though that small crystals may be of higher quality and some "ugly-looking" crystals diffract better than well-shaped ones. The gold standard in characterizing protein crystals is always a diffraction experiment.

### Crystallization Formulation Screens

Crystallizing a given protein for the first time is analogous to searching for a needle in a haystack. Screening crystallization conditions is a multidimensional search because many factors affect crystallization (see below). Two approaches have been used to tackle this challenge: (a) systematic assessment of the protein's solubility and precipitation behavior and (b) screening with preformulated screening kits. The latter is very popular since it can yield crystals results without the need for individualized treatment for every target protein.

Crystallization screening kits are commercially available, many of them in 96-well format (Table 1). These contain sets of formulations, some of which have a proven track record to aid in the crystallization of proteins [26]. The formulations often consist of three components: salt, precipitating agent, and a pH setting buffer. Due to the numerous possible permutations, a so-called sparse matrix is selected consisting of a subset of all possible combinations [27]. Recently Majeed et al. [28] designed a *synergy screen* to overcome some

of the limitations of these screens by combining multiple precipitation agents. If crystal hits are not obtained with the initial setup, follow-up experiments should be set up with lowered protein or precipitant concentrations where precipitate was observed and increased protein concentrations where clear drops were observed. In cases where no crystals are obtained from such trials, experimenters often switch to a different screening kit, or many screening kits are set up in parallel at the outset. Due to the stochastic nature of screening in general, it is impossible to know a priori how many crystallization experiments need to be set up in order to obtain a crystallization hit. Based on statistical analysis and a 1 to 2 percent likelihood for obtaining initial hits for typical proteins [29], it is estimated that screening 228 or 459 random crystallization conditions provides a 99 percent likelihood of observing at least one crystallization hit. The software program CRYSTOOL allows creating such random screens.

The systematic assessment of protein precipitation and solubility behavior is an iterative process. A particular precipitation agent is chosen and combined with the protein in batch-type crystallization experiments to map out a phase diagram similar to that shown in Figure 2. This is repeated for other precipitation agents if hits are not produced. Popular precipitation agents are ammonium sulfate, polyethylene glycols, Na/K phosphate, sodium chloride, and MPD [30].

Both crystallization approaches, systematic assessment or random screening, may provide either crystals suitable for X-ray diffraction experiments or poor hits that require refinement of crystallization conditions. In the latter case, fine screens varying one component concentration at a time are prepared in order to optimize crystal growth. In addition, crystal seeding experiments can be set up in which small or micro crystals serve as nucleation points for larger crystals useful for X-ray diffraction data collection [6].

Some of these different crystallization methodologies have been shown to work better for one protein than for others. Therefore, in de novo crystallization projects the type of crystallization method applied may be a useful parameter to test.

## Cryopreservation for Cryocrystallography

In order to expose protein crystals to a beam of X-rays, individual crystals are either mounted within a glass capillary or they are "fished" with a filament loop and cooled to low temperatures, for example, that of liquid nitrogen [31, 32], cooled propane, ethane, $CCl_3F$, or $BrCF_3$. A gaseous nitrogen jet is the most popular method for cooling crystals and allows X-ray diffraction experiments to be carried out at temperatures around 100 K. This procedure reduces X-ray radiation damage and reduces protein motion, both increasing the diffraction data quality.

Water crystals diffract X-rays and produce characteristic powder diffraction rings, and, thus, their formation during the crystal cooling process should

be avoided. This can be done by flash-cooling, which transforms liquid water into a glassy material. Water glass formation can be aided by the addition of so-called cryoprotectants such as glycerol, methyl-pentendiole, or trehalose. A popular strategy is to first prepare different cryoprotectant:precipitant solution mixing ratios, test their diffraction properties when frozen, and use the solution with the lowest cryoprotectant content that does not yield ice rings. Alternatively, ice formation on the protein crystal during the cooling process may be suppressed with high concentrations of certain salts. Lithium formate, chloride, and other highly soluble salts, so-called cryosalts, have been shown to suppress the formation of water ice upon cooling from ambient termperature to around 100 K [33]. Cyroprotectants may be avoided altogether by stripping surface-associated water from crystals by pulling them through oil, for example, perfluoropolyether, paratone-N, and paratone-N/mineral oil mixes [34].

## Crystal-Based Drug Discovery

There are two ways to prepare crystals of target proteins with their bound small-molecule ligands. One is *co-crystallization* where the ligand–protein complex is crystallized, the other is *soaking* of pre-formed apo-crystals with a solution containing ligand molecules. While soaking may be very economical since a single crystallization setup can provide up to hundreds of usable crystals, not all ligands can be soaked in, with solubility issues posing a common problem. In co-crystallization experiments, at equilibrium two protein species are present, the apo-protein (no ligand bound) and the ligand-bound form. The relative fraction of these two forms is determined by the affinity constant $K_a$, the concentration of the ligand and the concentration of the protein. At an excess of ligand concentration for a simple single binding site model, the equilibrium binding is given by

$$[PL] = \frac{[P]_T K_a [L]}{1 + K_a [L]} \tag{1}$$

with [PL] the concentration of the protein–ligand complex, $[P]_T$ the total protein concentration, and [L] the concentration of the free ligand. In many cases, it is difficult to estimate the proper concentrations in a crystallization setup. The goal is, of course, to populate most protein molecules with their ligand. As a rule of thumb, the concentration of a ligand to be used in a crystallization setup should exceed that of the protein by more than a factor of ten. A typical protein sample for crystallization is around 10 mg/mL protein. Depending on the molecular weight of the protein, the protein concentration will be around 250 μM. The ligand concentration should also exceed the dissociation constant of the ligand (which is the reciprocal association constant) by more than a factor of 10. Thus, if possible, a ligand concentration of more than 1 mM is often used.

*Co-crystallization* requires the formation of the ligand–protein complex prior to the crystallization process. This may be achieved simply by mixing of a protein solution with the ligand solution. In some cases where the off-rate of the ligand–protein complex is very low, a purification of the complex by size exclusion chromatography is feasible. It is evident that adding the small molecule, and possibly an organic solvent, usually DMSO (dimethylsulfoxide), and the formation of two protein species (apo-protein and ligand–protein complex) may alter the crystallization regime and thus the crystal quality or crystallization success. Therefore, co-crystallizations are frequently optimized for a particular ligand.

The ligand binding event often triggers subtle conformational transitions toward compacted protein structures, preventing the formation of those crystal contacts that are present in the respective apo-crystals. Alternatively, ligand binding may lead to the formation of crystals with a different packing arrangement and space group. This so-called ligand-depended crystal polymorphism [35] can be used as a tool for (a) binding or (b) the identification of initial crystallization conditions. The latter may be employed practically by screening a set of ligands as additives in co-crystallization experiments. As the formed ligand–protein complexes differ slightly in their conformations, different crystal contacts and packing arrangements are being sampled, thus increasing the chance of successful crystallization. Once a crystallization condition is identified, it may be used as a starting point for crystallizations with further ligands. This strategy is supported by the many cases where crystals of protein–ligand complexes have been obtained more readily than those of the respective apo-forms. Indeed, apo-forms of many ligand binding proteins have never been crystallized at all. It must be noted, however, that a particular crystallization condition that readily yields crystals of the apo-crystal form may not necessarily produce any crystals of ligand-bound forms. Most troubling, a given crystallization condition may not be compatible with binding of the ligand to the protein. Such a case may be envisaged where a protein crystallizes at low pH but does not bind its ligand at that pH.

*Soaking* compounds into pre-formed crystals is very popular because it short-cuts the possibly tedious process of identifying optimized co-crystallization conditions with a ligand. Soaking is performed by either (a) transferring the crystal into a solution consisting of precipitant solution and the ligand compound or by (b) addition of the ligand compound directly to the crystal-containing crystallization drop. It is difficult to estimate the preferred incubation time since the formation of the ligand–protein complex depends on the on-rate, diffusion of the compound within the crystal, and the component concentrations. In order to render hydrophobic drug leads soluble in aqueous solutions, organic solvents, detergents, or cyclodextrins have been employed.

The ligand compound or the solubilizing agent may have detrimental effects on the X-ray diffraction quality of crystals. In many cases visible cracks form upon soaking of ligands into apo-crystals. Evidently, substantial ligand-induced conformational transitions occur that may not be compatible with the

original packing of the apo-crystal form. In such cases, co-crystallization experiments are warranted. Even though the formation of visible cracks renders crystals unusable for diffraction purposes, the effect may, for special cases, be used as an indicator for ligand binding [36].

Both soaking and co-crystallization may be employed for crystal-based drug discovery. The locations and binding modes of lead compounds and the specific interactions of the small-molecule atoms with those of the protein provide unique insight that can be used in several ways (see Section 6). Furthermore, cocktails of structurally diverse compounds or fractions of fragment libraries have been co-crystallized or soaked into apo-crystals and the ensuing crystal structures displayed high-affinity ligands [1], thus forming the basis of *crystallographic screening*.

### Derivatization for Anomalous Diffraction Experiments

In order to obtain crystallographic phases crystals may be derivatized and used to collect anomalous X-ray diffraction data. In cases where proteins are expressed recombinantly, the protein may be labeled in vivo with selenium, substituting all sulfur-containing methionine residues by seleno-methionine. This method has become very popular due to the availability of tunable X-ray sources at synchrotrons. Chemical derivatization with heavy metals is a soaking or co-crystallization-based method. Cysteine, histidine, and methionine residues may react with heavy atoms such as mercury, gold, platinum, and iridium, whereas glutamate and aspartate can be complexed with lanthanides and actinides such as uranium and samarium [37]. Popular derivatizing reagents are $K_2PtCl_4$, $KAu(CN)_2$, $Hg(CH_3COO)_2$, $Pt(NH_3)_2Cl_2$, and $HgCl_2$ [38]. The binding of a particular heavy-metal compound to a protein may be screened prior to the crystallization and diffraction experiment by a simple native gel shift assay. Once a suitable reagent has been identified, the protocols for binding of heavy-metal compounds to proteins is similar to the binding of small-molecule compounds to form protein–drug complexes.

### Crystallization Data Sources

More than 25,000 protein crystal structures have been deposited in the publicly accessible Protein Data Bank (PDB) (http://www.rcsb.org/pdb/). The crystallization conditions are available within the PDB files under section REMARK 280. As an example the crystallization conditions for hen egg lysozyme given in the structure report 3LZT are shown below:

```
REMARK 280 CRYSTALLIZATION CONDITIONS: BATCH METHOD USED. 1% PROTEIN
REMARK 280 SOLUTION IN 100MM SODIUM ACETATE PH 4.5-4.6. SODIUM
REMARK 280 NITRATE ADDED TO A CONCENTRATION OF 20MGS/ML. CRYSTALS
REMARK 280 GROWN AT ROOM TEMPERATURE.
```

Here, the batch crystallization method was used where a 10-mg/mL solution of hen egg white lysozyme in 100 mM sodium acetate buffer at pH 4.5 to 4.6 was combined with a precipitant solution containing a 20-mg/mL sodium nitrate solution. Scientific publications describing protein structures usually list crystallization conditions in their methods section. A specialized protein crystallization periodical does not exist. The International Union of Crystallography (IUCr), however, publishes a monthly journal *Acta Crystallographica D—Biological Crystallography* that hosts a section on crystallization papers, where crystallization methods, conditions, and technical advances in crystallization methodology are reported.

Finally, the crystallization database BMCD (Biological Macromolecule Crystallization Database), set up by Garry Gilliland, is a useful resource for finding crystallization conditions of previously crystallized proteins [30, 39]. Version 2.0 of the BMCD is available on the Web at http://wwwbmcd.nist.gov:8080/bmcd/bmcd.html.

### Protein Crystallization Demonstration Experiment

Crystals of hen egg white lysozyme may be grown readily for demonstration purposes. A detailed procedure is given below for the crystallization of lysozyme according to the sitting drop vapor diffusion format.

1. Prepare a 100-mg/mL lysozyme solution by weighing out 100 mg of lyophilized chicken egg white lysozyme [Sigma (L7651)] into a tube and adding 1 mL of 50 mM sodium acetate at pH 4.5.
2. Prepare the precipitant solution; 30 percent (w/v) MPEG [mono-methyl polyethylene glycol 5000, Sigma (M7268)], 1 M sodium chloride, 50 mM sodium acetate at pH 4.5.
3. Fill 200 μL of the precipitant solution into the reservoir.
4. Pipet 2 μL from the reservoir solution into the crystallization well.
5. Add 2 μL of the lysozyme solution to the crystallization.
6. Crystals similar to those shown in Figure 4 appear within 1 h.

## 3 X-RAY DIFFRACTION EXPERIMENT

The information required to determine the three-dimensional crystal structure of a protein can be extracted from the X-ray diffraction data on a crystal of the target protein. In this section, basic diffraction theory and diffraction data collection practice will be discussed.

### Diffraction Theory

Images of objects with dimensions ranging in size from a single cell to macroscopic scale are readily obtainable by visible light photography or microscopy.

Since the dimensions of the sample objects in light photography and microscopy are roughly the same order of magnitude, or larger, compared with the radiation wavelength being used, the light radiation is scattered by the object. The scattered radiation is then refocused by lens systems to give an image of the object.

To obtain images of individual molecules with dimensions on the order of $10^{-7}$–$10^{-9}$ m, radiation of wavelength much shorter than that of visible light, such as X-rays, is required. The atoms within an individual molecule will scatter X-rays, but there are no lens systems that can refocus the scattered X-rays to readily give an image of the target molecule. It is possible, however, to reconstruct a model of the target molecule by mathematical calculations on the X-rays scattered by the target molecule.

X-ray diffraction by protein crystals provides the data used in protein X-ray crystallographic structure determinations. Diffraction is the phenomenon whereby radiation is scattered by objects of dimensions comparable to the radiation wavelength. When the dimensions of the scattering objects and the incident radiation are comparable, the scattering objects act as new sources of the radiation and reemit the radiation in all directions (Fig. 5).

The radiation scattered by the diffracting objects are waves and, therefore, have both an amplitude and a phase component. Any given scattered wave will be to some degree in or out of phase with some other scattered wave. When two or more scattered waves are in phase with each other, there will be constructive interference, giving rise to diffraction maxima. When the scattered waves are out of phase with each other, there will be destructive interference, giving rise to extinctions of no intensity (Fig. 6). The resulting diffraction image will therefore have a distinct pattern, depending on the shape, multiplicity, and arrangement of the scattering objects.

Figure 7 shows examples of simple diffracting arrays. The samples in the first column could be holes, and the samples in the second column could be corresponding diffraction patterns from light passing through the array of holes. One thing to note is that the dimensions in the scattering array and the dimensions in the diffraction pattern are inversely proportional. In Figure 7, the widely spaced holes result in closely spaced diffraction maxima, while the closely spaced holes result in widely spaced diffraction maxima. These two-dimensional arrays of scattering objects result in two-dimensional diffraction patterns.

## X-ray Diffraction by Macromolecule Crystals

The scattering of X-rays by protein crystals and their resulting diffraction patterns are completely analogous to the preceding examples of diffraction by slits and holes (Fig. 7). The spacings between atoms (bond lengths) in the protein crystal and the wavelengths of the X-rays are both on the order of $10^{-10}$ m (1 Å). Protein crystals will therefore scatter X-rays to give diffraction

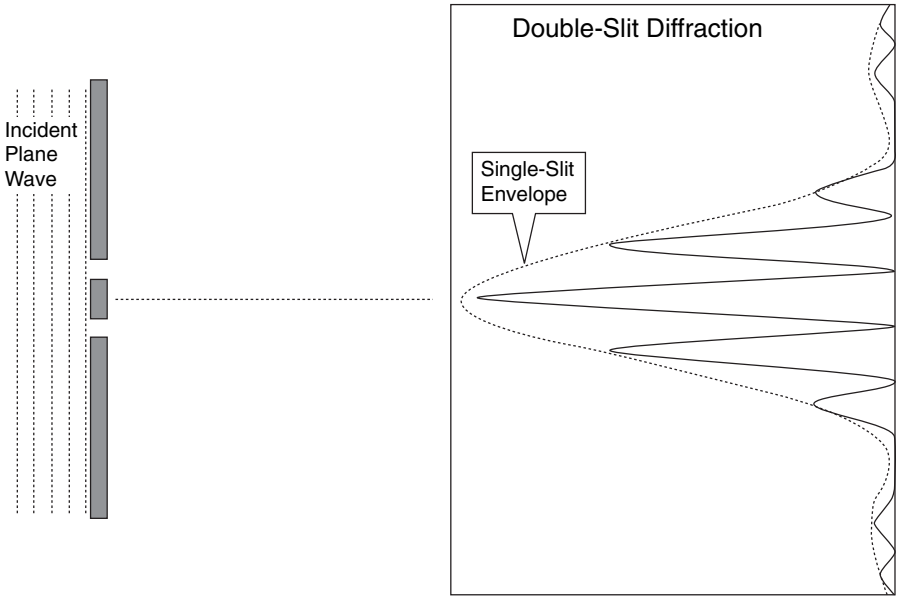


**Figure 5** Diffraction of waves passing through two slits [40]. As radiation passes through a slit of width comparable to the radiation wavelength, the slit scatters the radiation in all directions. In this example, each of the two slits on the left act as emitters of the incident radiation. The radiation emitted by the two slits will interfere, resulting in the solid-line diffraction pattern on the right. (The relative heights of the peaks in the diffraction pattern correspond to intensity.) Interfering waves which do not have a large angular deviation from the incident radiation, are largely in phase with each other resulting in intensity maxima. As the angular deviation of the scattered waves from the incident radiation increases, the scattered waves are less in phase with each other, resulting in the general intensity falloff as one moves away from the center of the diffraction pattern. Shown in broken line is the diffraction pattern of a single slit.

images. In a protein X-ray crystallography experiment, a crystal of the target protein is placed in an intense X-ray beam of a particular wavelength $\lambda$. The protein crystal is a three-dimensional array of diffracting objects (the atoms and their electrons), which scatters the incident X-rays to produce a three-dimensional diffraction pattern. The three-dimensional diffraction pattern is typically recorded in two-dimensional sections by a detection plate. Figure 8 shows sample protein crystal diffraction images. Each of the dark spots in the diffraction images represents a diffracted X-ray whose intensities are maxima of constructive interference between scattered waves. In a typical protein diffraction data collection, tens or hundreds of thousands of such diffraction spots are collected from a single crystal.

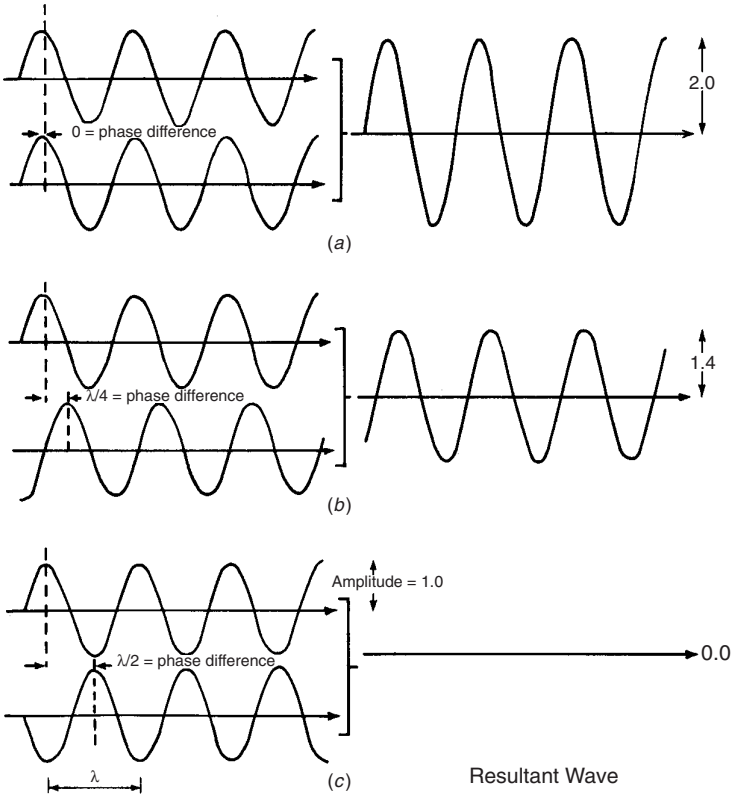


**Figure 6** Constructive and destructive interference of waves [41]. The resultant wave produced by two or more interfering component waves is obtained by adding their amplitudes at corresponding points. (*a*) Total constructive interference. The two component waves are completely in phase. The maximum amplitude of the resultant wave is double that of each component wave. (*b*) Partial constructive interference. The two component waves are 90° out of phase. The resultant wave has an amplitude 1.4× the amplitude of each component wave. (*c*) Total destructive interference. The two component waves are 180° out of phase. The two component waves cancel each other out, and the resultant wave has zero amplitude.

In the analysis of the X-ray diffraction pattern, each of the spots (also referred to as reflections) is designated as $F_{hkl}$, wherein the set of three integers, *hkl*, are called Miller indices. The Miller indices give the position of the reflection relative to the orientation of the crystal during the data collection.

## Fourier Synthesis in X-ray Crystallography

Since each of the spots in the diffraction pattern represents a wave (the diffracted X-rays), Fourier synthesis of these diffracted waves can be used to

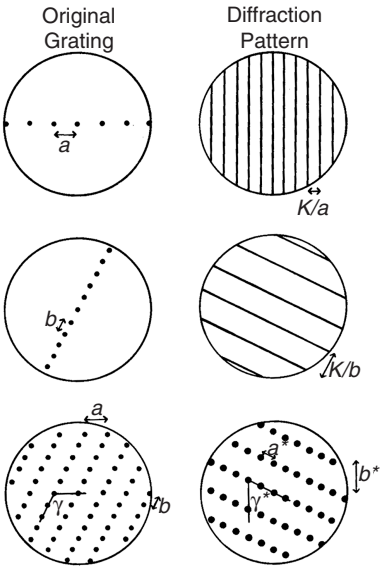


**Figure 7** Schematic representations of diffracting arrays (left column) and their resultant diffraction patterns (right column). In the diffraction arrays, the black dots represent holes. In the diffracting arrays, the black lines and black dots represent intensity maxima of the diffracted radiation. Note that the sampling in the diffraction pattern is along the direction of the diffracting array. Comparing the top and middle examples shows that the sampling regions are farther apart in the diffraction pattern of the middle example, due to the closer spacing of the holes in the diffracting array. The bottom example shows a two-dimensional diffracting array. Its diffraction pattern shows discreet sampling in two directions, resulting in spots of intensity maxima in the diffraction pattern, as opposed to lines in the other two examples [41].

reconstruct the diffracting array, namely the electron density of the protein in the crystal. The atoms of the protein, known from its amino or nucleic acid sequence, are then built into the reconstructed electron density to give the three-dimensional structure of the target protein.

Fourier syntheses are mathematical calculations used to reconstruct any regularly repeating pattern, regardless of complexity, via the summation of relatively simple sine and cosine curves [Eq. (2)].

$$f(x) = a_0 + \sum_h (a_h \cos 2\pi hx + b_h \sin 2\pi hx) \qquad h = 1, 2, 3, \ldots, n \tag{2}$$

As a very rudimentary example of a Fourier synthesis, consider the square wave in Figure 9. At first glance, it may appear impossible to reconstruct this function, especially its sharp corners, from smoothly curved sines and cosines. The summation of only four terms in the Fourier series, however, results in a function that is clearly beginning to take on the shape of the square

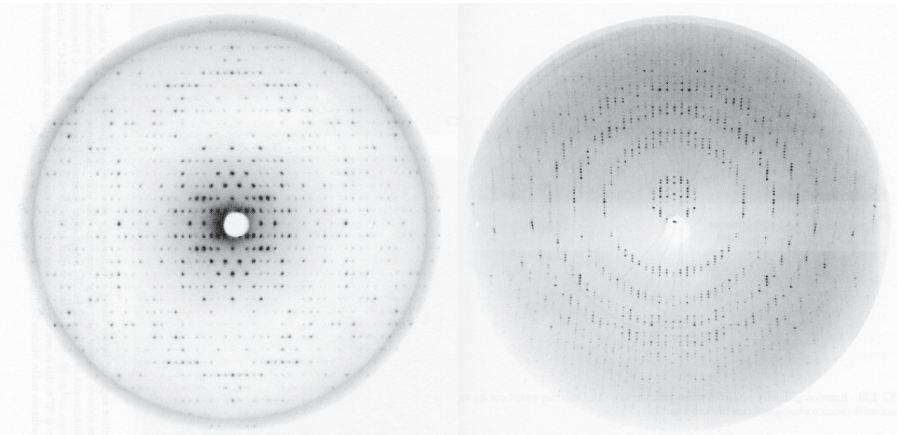

**Figure 8** Representative diffraction patterns used in protein X-ray crystal structure determination. Each of the spots in these diffraction patterns is generated by diffracted X-rays. Note the general tendency for the intensities of the diffraction spots to decrease as one moves farther away from the center of the diffraction pattern.

wave. The addition of more Fourier terms will produce a resultant function that looks increasingly like the target square wave.

In protein X-ray crystallography, the "wave" that one tries to reconstruct is the electron density within the protein crystal. The electron density in the crystal is a regularly repeating three-dimensional function. The wave components [the sines and cosines in Eq. (2)] used to reconstruct the electron density of the target macromolecule are the individual spots of the X-ray diffraction pattern. Equation (3) gives the electron density of the target protein at a point $(x, y, z)$:

$$\rho(x, y, z) = (1/V)\sum_{hkl} |F_{hkl}| \cos[2\pi(hx + ky + lz) - \alpha(hkl)] \tag{3}$$

In Equation 3, the electron density at a three-dimensional coordinate position $(x, y, z)$, $\rho(x, y, z)$, is expressed as the Fourier summation of all of the spots in the diffraction image. The diffraction spots, $F_{hkl}$, are waves, which have amplitude and phase components. Their amplitude components are expressed as $|F_{hkl}|$. Their phase components are the $\alpha(hkl)$ terms in Eq. 3, where $\alpha(hkl)$ is the angular phase of $F_{hkl}$. The (1/V) factor is the reciprocal of the unit cell volume of the sample crystal. The Fourier summation of the electron density, Eq. 3, appears to lack the sine terms in the general Fourier summation, Eq. 2. Due to Friedel's Law (Section 4) generally holding true in protein diffraction, certain pairs of spots have the same intensity but opposite phase angles and the sine terms in Eq. 2 cancel out.

Only the amplitude components of the $F_{hkl}$'s are directly measurable by the X-ray detection hardware used in the diffraction experiment. The phase components are determined by mathematical procedures that are discussed in the next Section 4. The important thing to note is that in Eq. (3), as was the

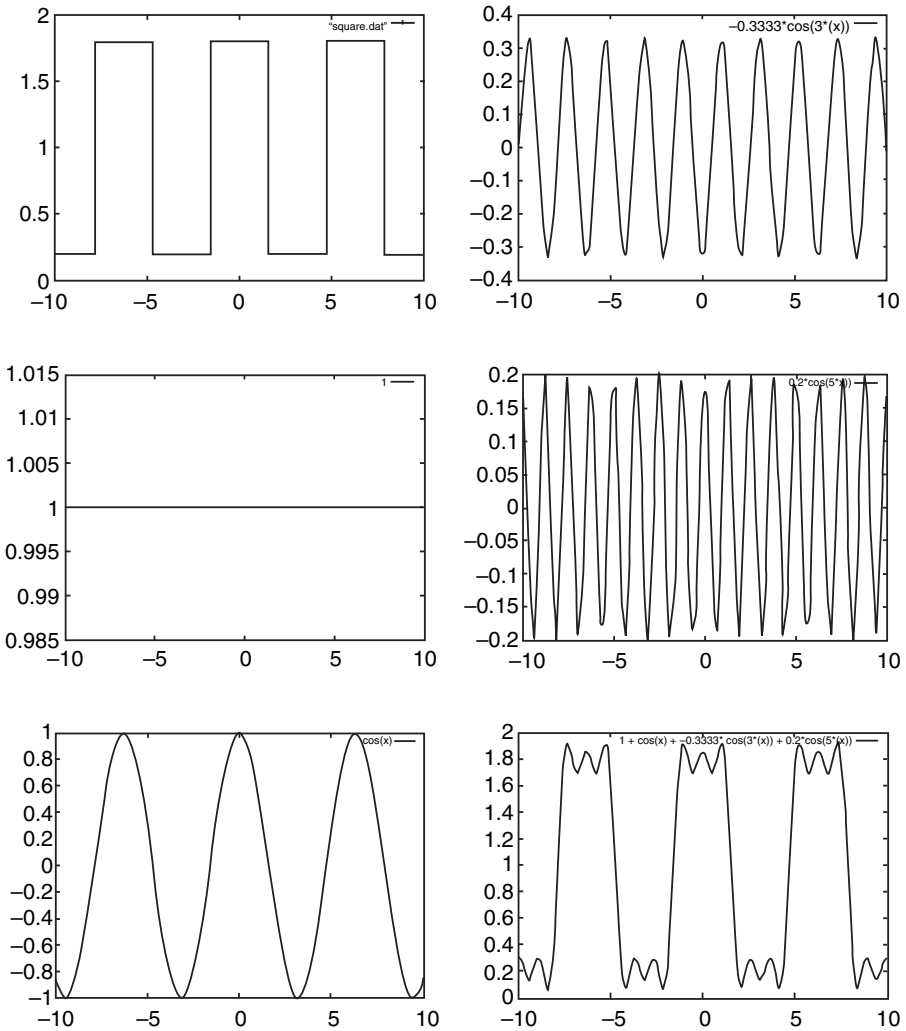


**Figure 9** Reconstruction of a square-wave by Fourier synthesis. The target square-wave function is shown in the upper left. The summation of four cosine functions [the constant function is cos(0)] result in the function in the lower right, which is beginning to take on the step shape of the target square wave.

case for the example of the square wave (Fig. 9), the reconstructed electron density $\rho(x, y, z)$ becomes a more accurate representation of the true electron density as more terms $F_{hkl}$ are added to the summation.

As seen in Figure 8, the X-ray diffraction pattern contains certain symmetry. The symmetry of the diffraction pattern is dictated by the internal symmetry of the protein crystal, more specifically, the symmetry of the unit cell of the protein crystal. The unit cell of a crystal can be described as a box with edges of length $a$, $b$, $c$ and angles between these edges of $\alpha$, $\beta$, $\gamma$. The unit cell

of a crystal is the smallest portion of the crystal that can be used to generate the entire crystal by whole unit cell translations along the $a, b, c$ cell edges. For protein crystals, the unit cell edges will be on the order of 50 to 400 Å. The unit cell is analogous to the scattering objects in Figure 7. A protein crystal with a large unit cell will produce a diffraction pattern with close spacings between the diffraction maxima, while a protein crystal with a small unit cell will produce a diffraction pattern with far spacings between the diffraction maxima. Because of this inverse relationship between unit cell size and diffraction pattern spacing, the diffraction pattern is called the reciprocal lattice of the actual crystal lattice.

Within each unit cell, there can be multiple copies of the protein molecule. These copies within a unit cell are related to all of the other copies in the unit cell by mathematical symmetry operations, such as rotations or translations. The smallest portion of the unit cell that can be used to generate the entire unit cell by the symmetry operations is called the asymmetric unit. The expression for the electron density [Eq. (3)] refers to an entire unit cell. In practice, since the contents of the unit cell can be generated by the mathematical symmetry operations, the X-ray crystallographic determination of protein structures generally refers to a single asymmetric unit of the crystal.

## Bragg's Law and the Angular Dependence of X-ray Diffraction

The somewhat complicated phenomenon of diffraction by the three-dimensional array of a protein crystal can be conceptually simplified by treating diffraction as the reflection of the incident X-ray beam from planes within the crystal. These planes are planes of electrons in the crystal, with the electrons being concentrated at the atoms. The incident X-rays can be thought of as reflecting off of these planes, resulting in the spots in the diffraction pattern where the X-rays strike the detection plate. The diffraction spots are therefore termed *reflections*. The geometric formulation of diffraction as reflections from planes is shown in Figure 10.

In Figure 10, the incident X-rays of wavelength $\lambda$ are represented by **1** and **2** and are reflected by planes $P_1$ and $P_2$ in the crystal with an interplanar separation $d$, resulting in the reflected X-rays **1′** and **2′**. The incident X-rays **1** and **2** make an angle $\theta$ with the planes $P_1$ and $P_2$. If the reflected X-rays **1′** and **2′** are to result in a beam of maximum intensity, the X-rays represented by **1** and **2**, and **1′** and **2′** must be in phase. In this geometric construction, for **1′** and **2′** to be in phase to produce an intensity maximum, the extra distance traveled by **2** and **2′** compared with the distance traveled by **1** and **1′** must be an integral number of wavelengths $\lambda$. Thus,

$$ACB = 2AC = n\lambda \tag{4}$$

where $n$ is an integer. Since $AC/d = \sin\theta$, Eq. (4) can be rewritten as

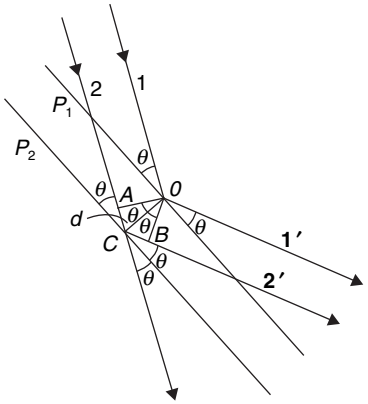


**Figure 10** Geometric construction for the Bragg theory of diffraction [42]. Parallel planes of electrons in the crystal are represented by $P_1$ and $P_2$. The incident X-rays are represented by rays 1 and 2. Diffraction is treated as "reflection" of rays 1 and 2 off of the planes $P_1$ and $P_2$, resulting in the diffracted waves represented by rays **1′** and **2′**.

$$n\lambda = 2d\sin\theta \tag{5}$$

Equation (5) is Bragg's law, which was derived by Sir William H. Bragg and his son Sir William. L. Bragg, who were awarded the 1915 Nobel Prize in physics for their work in X-ray crystallography. The interplanar spacing $d$ is the resolution of the particular diffracted wave. The resolution can be thought of as the resolvable distance between planes that "reflect" X-rays and lead to diffraction. It can be seen from Bragg's law that for constant $n$, diffracted X-rays with a greater angular divergence from the incident X-rays, those with high θ and farther away from the center of the diffraction pattern, result from reflections off of planes with small interplanar spacings $d$. Therefore, those reflections in the diffraction pattern with small interplanar spacings $d$ and large θ are the high-resolution reflections.

The Bragg construction of diffraction also provides a description of the Miller indices *hkl* assigned to a particular reflection. The reflection $F_{hkl}$ can be considered to be the resultant reflection off of the planes of constant spacing intersecting the unit cell edges $a$, $b$, and $c$, respectively $h$, $k$, and $l$ times. A low-resolution reflection will have low integer values for *hkl*. For example, the reflection $F_{100}$ is the reflection off the plane intersecting the $a$ cell edge once, and never intersecting the $b$ and $c$ cell edges. This is the plane defined by the $b$ and $c$ cell edges. The spacing of this plane is the $a$ unit cell edge. For a typical protein crystal, the resolution of the $F_{100}$ reflection, $d$ in Eq. (5), will be 50 to 400 Å. As the Miller indices *hkl* for a reflection increase, the spacings of the reflecting planes become smaller within the finite volume of the crystal's unit cell. Reflections $F_{hkl}$ with higher integer values of *hkl* are therefore the higher resolution reflections in an X-ray diffraction data set.

## Electron Clouds and Thermal Motion

Equation (5) gives only the positions of the diffracted X-rays in the diffraction pattern, and no indication of the intensity of the diffracted X-rays. The intensities of the reflections in the diffraction pattern are highly dependent on the content of the crystal and the angle θ of the diffracted X-ray. Because the electron clouds that scatter the X-rays have a finite volume, as θ increases, X-rays scattered by one part of the electron cloud are increasingly out of phase with X-rays scattered by another part of the electron cloud. Thus, higher resolution reflections with higher θ are naturally weaker in intensity than lower resolution reflections.

Thermal motion will also affect the intensities of scattered X-rays. The intensity fall-off shown in Figure 11*a* assumes that the carbon atom is stationary. The atoms in the crystal, however, will always be vibrating about their equilibrium points to some degree. This thermal motion effectively increases the volume of the electron cloud and further weakens the intensities of the higher resolution (high θ) reflections. Thermal motion is reflected in the exponential term $\exp[-B(\sin^2\theta)/\lambda^2]$, where $B$ is a measure of the thermal motion of the atom. If $B = 0$, there is no thermal motion, and the exponential term goes to 1. In reality, $B > 0$. Figure 11*b* shows the effect of thermal motion on the intensities of scattered X-rays. When $B = 0$, the angular dependence of the diffracted X-ray intensities is the same as in Figure 12. When $B$ increases, the intensities of the diffracted reflections are further weakened as θ increases.

Figure 8 shows the effects of the angular dependence of scattering on actual diffraction patterns for protein crystals. One can see that there is a

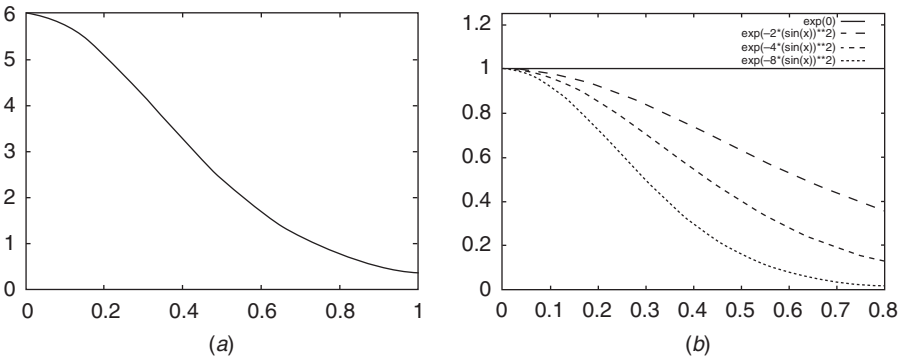


**Figure 11** (*a*) Scattering factor of carbon as a function of angular deviation of the diffracted wave from the incident X-ray beam. The scattering factor $f_C$, plotted along the $y$ axis, is a measure of the intensity of the diffracted wave at a particular angular deviation, $(\sin\theta)/\lambda$ (plotted along the $x$ axis) from the undiffracted X-ray beam. (*b*) Effects of thermal motion on the angular dependence of X-ray scattering. The thermal motion factor $\exp[-B(\sin^2\theta)/\lambda^2]$ for values of $B$, ranging from 0 to 8. Atoms with higher $B$ values have greater thermal motion. When $B = 0$, the function is constant and does not affect the natural intensity falloff shown in (*a*). As $B$ increases, the intensity falloff shown in (*a*) becomes more rapid as $(\sin\theta)/\lambda$ increases.

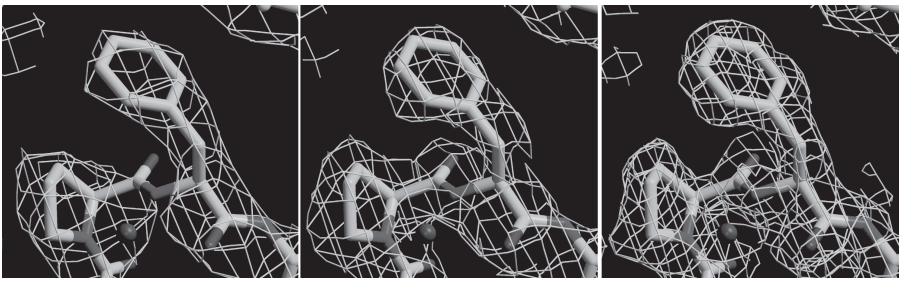

**Figure 12** Electron density maps calculated for a protein at increasing resolutions: *(Left)* 3.5 Å, *(middle)* 2.8 Å, and *(right)* 2.1 Å. The electron density is shown in cyan chicken wire. Electron density is usually displayed as chicken wire contour lines around the protein model. The chicken wire contours correspond to ordered electron density in the crystal structure, and should therefore be superimposable on the model of the protein. (This figure is available in full color at ftp://ftp.wiley.com/public/sci_tech_med/drug_discovery/.)

general tendency for the intensities of the spots to weaken as their positions move farther from the center of the image. For any given protein crystal, the higher resolution reflections are more difficult to collect, if not unattainable, compared with the lower resolution reflections.

Similar to the example of the square wave reconstruction in Figure 9, adding more, higher resolution, reflection terms to the electron density Fourier transform [Eq. (1)], gives a more accurate representation of the electron density for the target protein. Figure 12 shows the effects of using higher resolution reflections in the calculation of electron density maps. At lower resolutions, the electron density gives the general shape of the target protein's electron density. At higher resolutions, the electron density reveals more details of the target protein structure.

## X-ray Diffraction Data Collection in Practice

The collection of protein X-ray diffraction data requires three basic hardware components: X-ray source, goniometer, and X-ray detector. There are numerous configurations of these components and additional X-ray optical components that can be used. Figure 13 shows a schematic of an X-ray diffraction data collection system. An intense, monochromatic X-ray beam produced by an X-ray source strikes the sample protein crystal. The X-rays diffracted by the crystal are measured by an X-ray detector, on the opposite side of the crystal as the X-ray source. The protein crystal is mounted on a goniometer, which allows for very precise angular adjustments of the crystal position. During a data collection, the crystal will be rotated about one or more axes by the goniometer. Usually, the crystal is rotated about the $\phi$ axis, which is

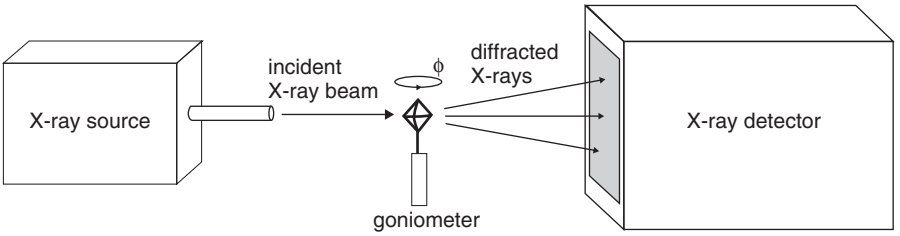


**Figure 13** Schematic of the basic components of a protein X-ray diffraction data collection experiment. An X-ray source produces an intense X-ray beam. The sample protein crystal is mounted on a goniometer and placed in the X-ray beam. The goniometer allows precise angular positioning of the crystal. During data collection, the crystal is generally rotated about the ϕ axis. The diffracted X-rays are recorded by an X-ray detector.

normal to the incident X-ray beam. The X-ray detector can be positioned at an angular offset 2θ relative to the incident X-ray beam to collect higher resolution reflections [higher 2θ, Eq. (5)]. In Figure 13, the 2θ offset is 0°, that is, the incident X-ray beam, crystal, and X-ray detector are colinear.

Modern detectors are generally two-dimensional plates that record many diffraction spots simultaneously in a single X-ray exposure (Fig. 8). Data-processing software determines the crystal symmetry and unit cell parameters for the diffraction data, based on the pattern of the diffraction spots. The software then predicts positions on the detector where the reflections should fall and measures the intensities and positions of the reflections (Fig. 14). The measured intensities of the reflections are then converted to amplitudes, which are roughly the square root of the intensities.

The quality of the processed diffraction data can be gauged by certain statistical indicators. One of the most popular measures of data quality is $R_{\text{sym}}$. As discussed above, the protein X-ray diffraction pattern has certain symmetry. Reflections with Miller indices *hkl* are crystallographically constrained to have the same intensity as certain other reflections with different Miller indices due to the symmetry of the crystal. Therefore, while many total reflections, sometimes several hundreds of thousands, may be collected in an X-ray diffraction data set, the set of unique reflections is far less. A 2- to 20-fold excess, or redundancy, of reflections is usually collected in a protein X-ray diffraction data set. In theory, the reflection *hkl* and all of its symmetry related mates should have the same intensity. In practice, however, there will be differences in the measured intensities of these reflections. $R_{\text{sym}}$ is the measure of how similar the intensities of these reflections are. The expression for $R_{\text{sym}}$ is

$$R_{sym} = \sum_{hkl} [|I - \langle I \rangle| / I] \tag{6}$$

In Eq. (6), $I$ is the intensity of a reflection with Miller indices *hkl*. $<I>$ is the average intensity of that reflection and all of its symmetry-related reflections. In an ideal data set where all of the symmetry-related reflections have

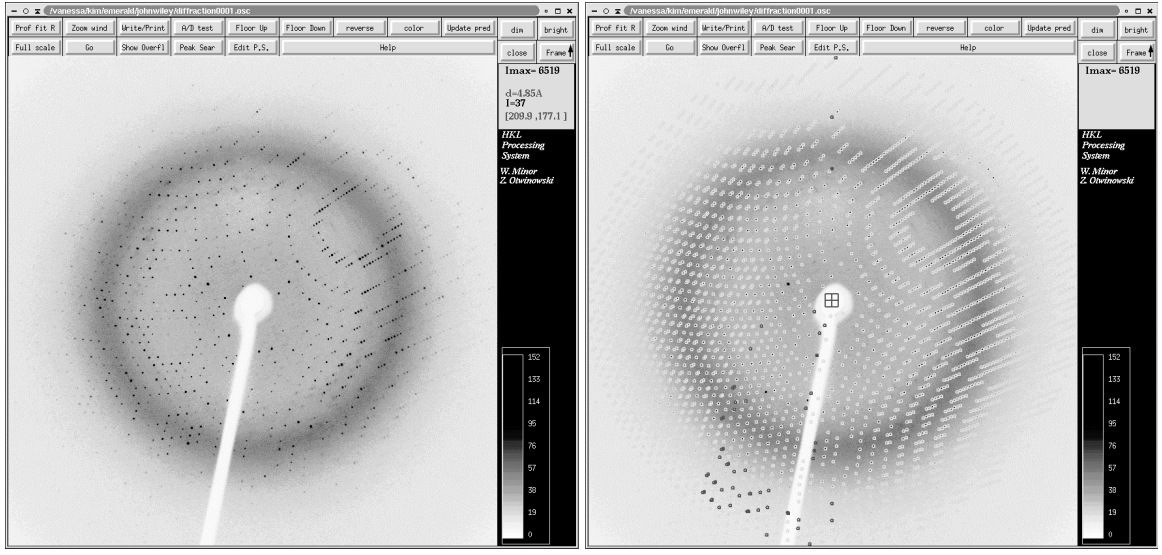


**Figure 14** An example of diffraction data collection and processing. *(Left)* A protein crystal diffraction image. *(Right)* The reflections of the protein diffraction image with prediction circles calculated by the data-processing software. The prediction circles should coincide with the reflections. Images from the diffraction data-processing software HKL2000 [43].

identical measured intensities, the numerator in Eq. (6) is zero, and $R_{sym}$ is zero. In practice, typical protein X-ray diffraction data sets have $R_{sym}$'s of 0.05 to 0.15 over the entire resolution range of the data, with $R_{sym}$ increasing for higher resolution reflections. As previously mentioned, higher resolution reflections generally have weaker intensity and are more difficult to measure accurately. In general, a lower $R_{sym}$ is considered an indication of a better quality data set.

One of the problems with using $R_{sym}$ as an indicator of data quality is that $R_{sym}$ is highly dependent on the redundancy of the data (the number of times a reflection is recorded). As more measurements are made for a reflection and its symmetry-related mates, $R_{sym}$ increases [44], indicating deterioration of the data set. This is statistically counterintuitive since the quality of the data must improve with increased sampling (higher redundancy). To address this statistical flaw in $R_{sym}$, the indicator $R_{mrgd}$, which accounts for data redundancy, has recently been proposed [44]. Still, $R_{sym}$ is more quoted as an indicator of data quality for protein X-ray diffraction data.

Another indicator of data quality that is statistically more robust than $R_{sym}$ is $I/\sigma$, the signal-to-noise ratio of the measured intensity for a reflection. One of the most important aspects of $I/\sigma$ is that it is more reliable than $R_{sym}$ in defining the resolution limit of the data. The resolution limit is a critical descriptor of a protein diffraction data set. The resolution limit indicates how much detail should be expected in the electron density maps (Fig. 12) and also guides model refinement strategies (discussed below in Section 5). The resolution limit is sometimes taken as the resolution at which $R_{sym}$ exceeds a certain threshold value, chosen by the experimenter. Typical threshold values for $R_{sym}$ in setting the resolution limits of protein X-ray diffraction data range from 0.40 to 0.50. The experimenter can, however, intentionally or unintentionally manipulate $R_{sym}$ to lower (better) values by such practices as reducing the number of measured reflections or rejecting reflections with weak intensity. In contrast, $I/\sigma$ sets the resolution limit since by definition, it is the signal to noise ratio of the measurements and is less prone to experimenter manipulation [45]. An $I/\sigma$ of 2 is often taken as the resolution limit of a protein X-ray diffraction data set. In most cases, though, both $R_{sym}$ and $I/\sigma$ are quoted for diffraction data in the primary protein crystallography literature.

## Home Laboratory Diffraction Data Collection

The basic configuration shown in Figure 13 can be installed in any home laboratory and is now common throughout protein crystallography groups. There are several vendors offering home laboratory X-ray diffraction hardware. The purchase price of a complete home laboratory X-ray diffraction system is around US$500,000.

Most home laboratory systems use a copper rotating anode as the X-ray source. In such an X-ray source, an intense electron beam produced by heating a metal filament, the cathode, strikes a rotating cylinder of copper, the anode.

The electron beam excites the copper atoms in the anode, which emit X-ray radiation. The electronic energy spacing of copper is such that the emitted X-rays have $\lambda = 1.54$ Å (Cu $K_\alpha$ radiation). According to Bragg's law [Eq. (5)], the highest resolution diffraction data attainable with Cu $K_\alpha$ radiation is around 0.8 Å [42]. This maximum resolution is very rarely attained with most protein crystals, primarily due to crystalline imperfections. As shown in Figure 12, however, much more modest resolution limits are sufficient to confidently determine protein crystal structures.

The X-ray detector in a typical home laboratory system is an imaging plate coated with a phosphorescent material. The diffracted X-rays impinging on the imaging plate excite the phosphors on the plate. The plate is then scanned with a laser, and the phosphors release the stored energy from the X-rays as visible light [46]. The released light is then analyzed for the positions and intensities of the diffracted X-rays in the original image. Typically, a crystal is exposed to the incident X-ray beam for 5 to 30 min for one diffraction image. The readout of the image plate takes about 2 to 5 min. On some home laboratory systems, there are two or more imaging plates so that one plate can be exposed while the other plate is read, resulting in no latency for plate readout. The goniometer then reorients the crystal to the next angular position, and the next image is collected. For protein crystals, 90 to 180 images are typically collected in a complete data set.

Not shown in Figure 13 since it is not strictly required for diffraction is a cryostat. Protein crystals are very delicate and are subject to rapid radiation damage from the incident X-ray beam, most likely from the formation of free radicals in the crystal. Radiation damage continually degrades the diffraction resolution of a crystal. To prevent radiation damage, the sample crystal is mounted on the goniometer so that the crystal is in a constant jet of nitrogen gas near liquid nitrogen temperature. While such cooling of the crystal does not eliminate radiation damage, it usually retards radiation damage enough to allow collection of complete diffraction data from a single sample crystal. The use of cryostats in diffraction data collection is now practically universal in protein crystallography.

## Synchrotron Diffraction Data Collection

The convenience of a home laboratory X-ray diffraction system is often mitigated by the relatively weak intensity of the incident X-ray beam. While the diffraction resolution limit of a Cu anode X-ray diffraction system is about 0.8 Å, protein crystals are usually very weakly diffracting, and diffraction resolution to 3.0 Å or better is a reasonable expectation with home laboratory X-ray sources.

To obtain higher resolution X-ray diffraction from protein crystals, data are often collected at synchrotron beamlines. Synchrotrons are large circular particle accelerators in which charged particles travel at relativistic speeds in an orbital path. Charged particles subjected to an acceleration emit radiation.

This radiation is emitted in the direction of travel of the charged particle if it had not been subjected to the acceleration.

In a synchrotron, electrons travel in a circular orbit. A centripetal acceleration is therefore placed on the electrons, which results in radiation being emitted tangential to the circular path of the electrons. The emitted radiation is then directed into experimental stations built as spurs off of the main synchrotron ring. Since the emitted radiation spans the electromagnetic spectrum, many different types of experimental stations can be built at a single synchrotron, each station selecting different wavelengths from the emitted radiation, such as ultraviolet or X-rays.

Synchrotron beamlines are capable of several orders of magnitude greater X-ray flux than home laboratory X-ray diffraction systems. Since the intensity of the incident X-ray beam is much greater, high-resolution reflections, which are barely measurable on home laboratory systems, can have useful intensities when collected at a synchrotron. The same protein crystal can therefore give much higher resolution diffraction data at a synchrotron.

In addition to high intensity, the wavelength of the incident X-ray beam at a synchrotron can be tuned. On a typical home laboratory Cu anode X-ray source, the wavelength of the incident radiation is fixed at 1.54 Å. At a synchrotron, through the use of various optical elements, the wavelength of the incident X-ray beam entering an experimental station can be tuned over a range of several angstroms. The application of such wavelength tuning is critical in anomalous dispersion diffraction data collection, which will be discussed in the next section.

Overall, the components of a synchrotron X-ray diffraction system are as shown in Figure 13. The much greater intensity of the synchrotron incident X-ray beam, however, results in changes in the practice of data collection compared with the home laboratory X-ray system. Data collection is much faster at a synchrotron. A data set taking 2 to 3 days to collect on a home laboratory system can be collected in an hour at a synchrotron. Individual exposures requiring 5 to 30 min on a home laboratory system generally take 5 to 60 s at a synchrotron.

The rapid data collection possible at synchrotrons allows for different hardware compared with a home laboratory system. A major difference is in the type of X-ray detector. Most synchrotron beamlines are equipped with charge-coupled device (CCD) detectors. CCDs are light-sensitive integrated circuit chips. In a CCD X-ray detector, the diffracted X-rays impinge on a large phosphorescent face plate. The X-rays are converted to visible light by the phosphors. The visible light is then directed through an optical taper to demagnify the light down to the CCD chip [47]. One of the great advantages of using CCD detectors instead of imaging plates on a synchrotron beamline is the faster readout times of the CCD detectors, which are on the order of a few seconds instead of minutes, as with an imaging plate. With well-diffracting protein crystals, it is possible to collect a complete diffraction data set at a synchrotron in less than half an hour [48].

Another hardware becoming more common at synchrotron beamlines is the robotic sample mounter. These devices allow multiple crystals, on the order of 100, to be loaded into a magazine, which is generally filled with liquid nitrogen to keep the crystals frozen. The crystals are then automatically mounted and retrieved from the goniometer without the need for human intervention [49, 50].

Synchrotrons are massive public works projects, with physical scale on the order of kilometers, built by governmental agencies and costing around $1 billion. Some examples are the Advanced Photon Source at Argonne National Laboratory in Argonne, Illinois, and the SPring-8 synchrotron in Hyogo, Japan. While synchrotrons were originally built for high-energy particle physics experiments, their use in protein crystallography is growing rapidly. There are currently on the order of 50 to 100 synchrotron beamlines worldwide dedicated to protein crystallography, with new ones being commissioned [51].

## 4 X-RAY CRYSTAL STRUCTURE DETERMINATION

The previous section discussed general diffraction theory and diffraction data collection. This chapter will give an overview of determining the protein crystal structure from the diffraction data.

### Phase Problem

As described in the previous Section 3, the electron density of a protein, and therefore its three-dimensional structure, can be reconstructed by the summation of sine and cosine terms. Such a summation is called a Fourier synthesis. The sine and cosine terms are the collected diffraction data, the reflections in the diffraction images (Fig. 8). These reflections are the X-rays diffracted by the crystal. Since these X-rays are waves, they have both amplitude and phase components.

The amplitude components of the diffraction data can be measured directly by the X-ray detector, which measures the intensities of the diffraction spots. The phase components, however, cannot be measured directly. This inability to directly measure the phase in diffraction data is referred to as the *phase problem*. In this chapter, various methods for determining phases for protein X-ray diffraction data will be discussed.

### Structure Factor

In a diffraction image, each of the reflections identified by the Miller indices *hkl* are described mathematically by a structure factor $F_{hkl}$. The structure factor is the resultant vector of all of the waves diffracted in the direction of the spot *hkl* by all of the atoms in the crystal. The structure factor corresponding to a spot *hkl* is therefore the vector sum of the structure factors of each atom in the crystal. The structure factor of an individual atom in the crystal can be expressed as:

$$f_j = |f_j| e^{i\delta j} \tag{7}$$

where $|f_j|$ is the magnitude of the structure factor, which can be measured directly, and $e^{i\delta j}$ is the phase, which cannot be measured directly. ($i$ is the imaginary number square root of –1.) $\delta j$ is the phase angle of the atom.

The phase of a structure factor $F_{hkl}$ is the phase difference of the resultant diffraction of all of the atoms in the crystal unit cell with respect to the origin of the unit cell. In radians, the phase of each atom, δ, can be expressed as [42]:

$$\delta = 2\pi(hx + ky + lz) \tag{8}$$

In this expression $h$, $k$, and $l$ are the Miller indices, and $x$, $y$, and $z$ describe the position of the atom in fractional coordinates of the unit cell. Since the structure factor $F_{hkl}$ for a reflection with Miller indices $hkl$ is the sum of all of the individual atom structure factors, the structure factor $F_{hkl}$ can be expressed as:

$$F_{hkl} = \sum |f_j| e^{i\delta j} \tag{9}$$

or

$$F_{hkl} = |F_{hkl}| e^{2\pi i(hx+ky+lz)} \tag{10}$$

where $|F_{hkl}|$ is the directly measurable magnitude term, and $e^{2\pi i(hx + ky + lz)}$ is the phase term that cannot be directly measured. The structure factors are then summed as in Eq. (3) to calculate the electron density.

Although the phase of a reflection $F_{hkl}$ cannot be directly measured by the X-ray detector, Eq. (8) shows that the phase is defined by the geometric coordinate terms $hx$, $ky$, and $lz$. Thus, by determining the positions of atoms within the crystal, the phases can be calculated. The remainder of this chapter will discuss methods for determining atomic positions to calculate phases for the diffraction data, and ultimately determine the three-dimensional crystal structures of proteins.

## Heavy Atom Replacement Methods

Heavy atom replacement methods are used to determine structures of proteins for which no partial structures or reasonable approximations of the structures are available. These methods are suitable for determining the structures of proteins with new or unknown folds.

The determination of phases in crystallography is essentially the determination of one or more atomic positions within the crystal. While a direct visualization of specimen positions, like in visible or electron microscopy, is not possible in X-ray diffraction, atomic positions can be extracted from the phaseless diffraction data.

### Multiple Isomorphous Replacement

As stated previously, the Fourier transform for the electron density using X-ray diffraction data [Eq. (3)] requires the phases of the diffracted X-rays, which in turn require positional information about the atoms in the crystal. Another Fourier transform calculated from the X-ray diffraction data, the Patterson function, does not require any phase or positional information. The Patterson function is expressed as:

$$P_u = (1/a)\sum_h |F|^2 \cos(2\pi hu) \tag{11}$$

The Patterson function is mapped onto a coordinate system proportional to the direct unit cell dimensions of the crystal with the *xyz* coordinates of the direct unit cell converted to *uvw* in the Patterson map. In Eq. (11) for the Patterson function, the value of the Patterson function, in this case along the *u* axis, is a function of the *h* Miller index for a particular reflection and the square of the amplitude, $|F|^2$, of the reflection. Since the amplitudes of the reflections are experimentally measured, the Patterson function can be calculated directly from the diffraction data. Whereas in the electron density Fourier transform [Eq. (3)] the maxima correspond to atomic positions, the maxima in the Patterson function correspond to interatomic vectors. Furthermore, the intensity of the Patterson peaks is weighted by the atomic masses of the two atoms at the ends of the interatomic vectors.

In a protein consisting of nearly all carbon, nitrogen, and oxygen atoms (hydrogen atoms are generally not visible in protein X-ray diffraction), the Patterson function of interatomic vectors is practically uninterpretable. This is because all of the atoms are of similar mass so all of the peaks in the Patterson map are of similar value. If, however, there were a heavy atom in the crystal, such as mercury or lead, which had a much higher atomic mass than the carbons, nitrogens, and oxygens, the interatomic vectors between these heavy atoms would have a much higher value and would clearly stand out in the Patterson map.

In order to calculate interpretable Patterson functions, protein crystals can be soaked in micromolar to millimolar concentrations of heavy atom solutions to incorporate the heavy atoms into the protein crystals [37]. Some of the most common heavy atoms used for protein crystallography phasing are gold, lead, mercury, and platinum. These metals generally bind to the protein at sulfur-containing residues. As implied by the name *isomorphous replacement*, in soaking protein crystals with heavy atoms, it is important that the soaking procedure does not result in a significant change in the unit cell dimensions of the crystal. A crystal whose unit cell changes significantly upon heavy atom soaking will not be useful in the structure determination. In practice, soaking relatively fragile protein crystals in heavy atom solutions often results in damage to the crystal. The damage can range from simple reduction of dif-

fraction resolution, which crystals can still be useful, to complete extinguishing of diffraction. It is not uncommon to screen tens or hundreds of different heavy atoms and soaking conditions to get a single heavy-atom-soaked protein crystal that diffracts and maintains the unit cell dimensions of the unsoaked protein crystal.

Due to the symmetry of the crystals, there are generally multiple copies of the protein molecule in the unit cell of the crystal, with one or more copies of the protein in the asymmetric unit of the crystal's unit cell. The heavy atoms will bind to each of the multiple protein atoms in the unit cell. The heavy atoms will therefore occupy all of the asymmetric units in the crystal. The mathematical symmetry operations relating each asymmetric unit to another result in the maxima in the Patterson map occurring at certain sections of the *uvw* Patterson space. Determining the atom positions from Patterson maps can be done manually by relatively simple algebraic equations. There are now, however, numerous software suites that automatically determine heavy atom positions from Patterson maps, and the modern protein crystallographer rarely, if ever, determines heavy atom positions manually.

X-ray diffraction data collected from a native protein crystal lacking heavy atoms (data set P) and diffraction data from a crystal of the same protein containing heavy atoms (data set PH) will differ in the measured intensities of their reflections solely by the contribution of the heavy atoms in data set PH. These differences in intensities can be used to calculate an isomorphous difference Patterson map, which will show peaks corresponding to the interatomic vectors between the heavy atoms (Fig. 15).

The isomorphous difference Patterson function calculated from data sets P and PH is expressed as:

$$P_u = (1/a)\sum_h |F_{\mathrm{P}} - F_{\mathrm{PH}}|^2 \cos(2\pi hu) \tag{12}$$

Compare this expression for the isomorphous difference Patterson with that of the general Patterson function [Eq. (11)]. In the isomorphous difference Patterson, the amplitude term is the difference between the amplitudes in the native protein and heavy-atom-derivatized protein data sets. The amplitude differences due to the presence of heavy atoms in data set PH will show up as strong peaks, allowing interpretation of the isomorphous difference Patterson maps.

The coordinates of the heavy atoms calculated from the positions of the Patterson peaks essentially determine the heavy atom substructure of data set PH. The amplitudes of the reflections from the heavy-atom-only substructure of data set PH are easily calculated by the appropriate summation of heavy atom scattering factors, which are listed in standard tables. The phases of the reflections from the heavy-atom-only substructure are calculated from Eq. (8). Once the heavy atom substructure has been determined, phases can be determined for the protein diffraction data.

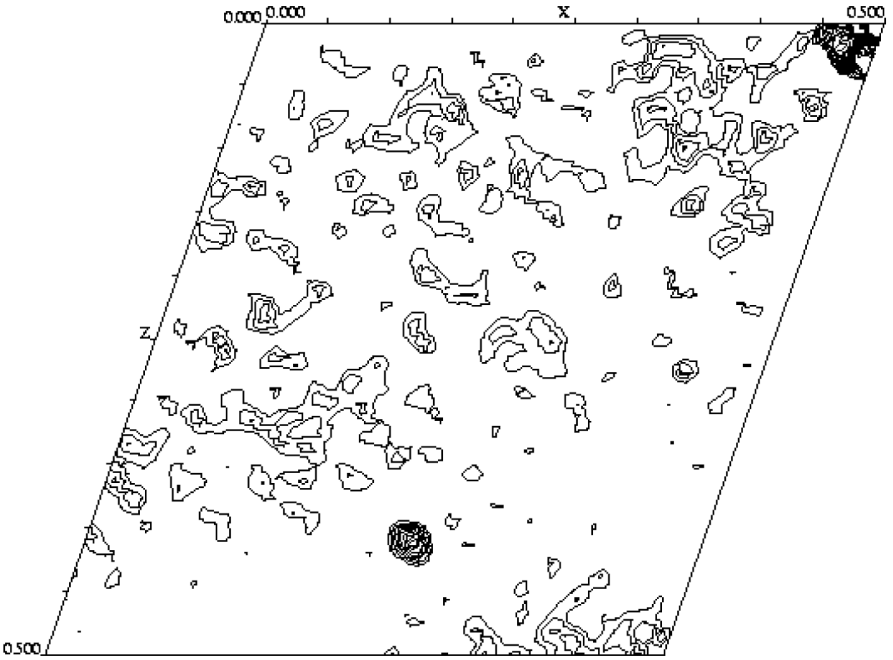


**Figure 15** An isomorphous difference Patterson map calculated according to Eq. (12). The peak at approximately $X = 0.25$, $Z = 0.40$ corresponds to the interatomic vector between the heavy atoms in the heavy-atom-derivatized diffraction data set.

The general concept of phasing X-ray diffraction data with heavy atom replacement can be illustrated with Argand diagrams. Since each reflection $F_{hkl}$ has an amplitude and phase component, the reflections can be represented as vectors in an Argand diagram (Fig. 16). The amplitude of the vector is determined from the measured intensity of the reflection $F_{hkl}$. The phase of $F_{hkl}$ cannot be measured, so the phase angle in the Argand diagram is indeterminate from 0–2π radians.

The data sets P and PH can now be plotted on Argand diagrams to determine phases for the data set P. The only experimentally measurable difference between the diffraction data sets P and PH are the intensities of the reflections, which are due to the presence of the heavy atoms in data set PH. For a given reflection $F_{hkl}$, the superposition of the Argand diagrams for data set P and PH, with data set PH offset by the amplitude and phase of the heavy-atom-only substructure calculated for data set PH, shows the intersection of the P and PH phase circles at two points (Fig. 16). These two points of intersection are possible phases for the reflection $F_{hkl}$. Given only these two possible points, a good choice for the phase is the mean of the two points. Choosing the mean from just two points, however, can lead to a very large phase error. To break the phase ambiguity, diffraction data can be used from a second

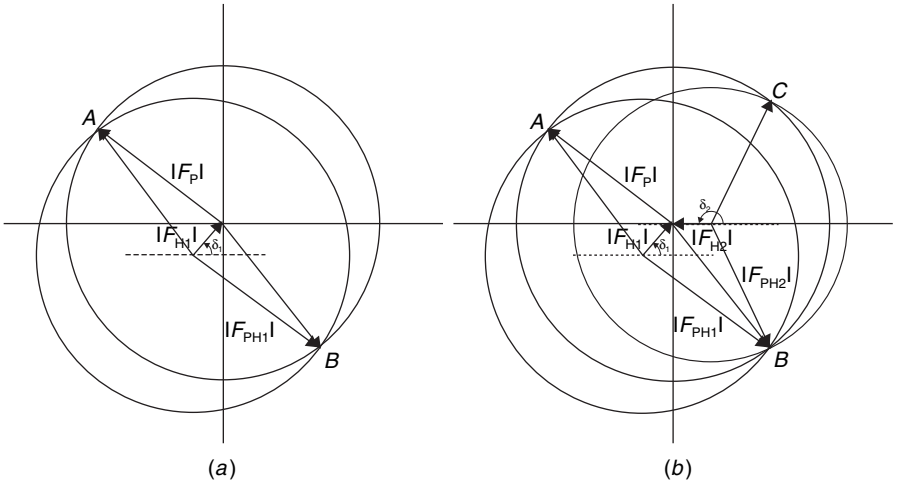


**Figure 16** Argand diagrams illustrating phase determination. (*a*) The measured reflection $F_P$ with Miller indices *hkl* from the protein crystal diffraction data is plotted as a vector with magnitude $|F_P|$. No phase information is present in the diffraction data, so the phase of the reflection is indeterminate from 0 to $2\pi$ radians as indicated by the light circle. The same reflection with Miller indices *hkl* from the heavy-atom-derivatized diffraction data is plotted as a vector with magnitude $|F_{PH1}|$, shown as the heavy circle. The reflection $F_{PH1}$ is plotted offset from the reflection $F_P$ by the magnitude, $|F_{H1}|$, and phase angle, $\delta_1$, of the heavy-atom-only substructure, as determined from isomorphous Patterson maps. The points *A* and *B* where the light and heavy circles intersect are two possible phase angles for the protein reflection $F_P$. This *A* or *B* ambiguity is a graphical representation of the phase problem. (*b*) Resolution of the phase problem using a second heavy atom derivative. The reflections $F_P$ and $F_{PH1}$ are plotted as in (*a*). In addition, the reflection with the same Miller indices *hkl* from a second heavy-atom-derivatized diffraction data set is plotted as a vector with magnitude $|F_{PH2}|$. The reflection $F_{PH2}$ is plotted offset from the reflection $F_P$ by the magnitude, $|F_{H2}|$, and phase angle, $\delta_2$, of the heavy-atom-only substructure in the second heavy-atom-derivative data set, as determined from isomorphous Patterson maps. The heavy circle from the second heavy atom data set also intersects the light circle of the protein-only diffraction data set at two points, *B* and *C*. One of these points, *B*, is common between the two possible phases indicated by the first heavy atom derivative and is the correct phase angle for the reflection $F_P$.

heavy-atom-derivatized crystal, data set PH2. After determining the amplitudes and phases of the heavy-atom-only substructure of data set PH2 by Patterson methods as described for data set PH, the reflection $F_{hkl}$ from data set PH2 can be superimposed on the Argand diagrams for data sets P and PH (Fig. 16). With two, or more, heavy atom data sets, the correct phase is the point at which all of the phase circles intersect. Because of the requirement for more than one heavy atom derivative diffraction data set, this method is called multiple isomorphous replacement (MIR). Once the phases are deter-

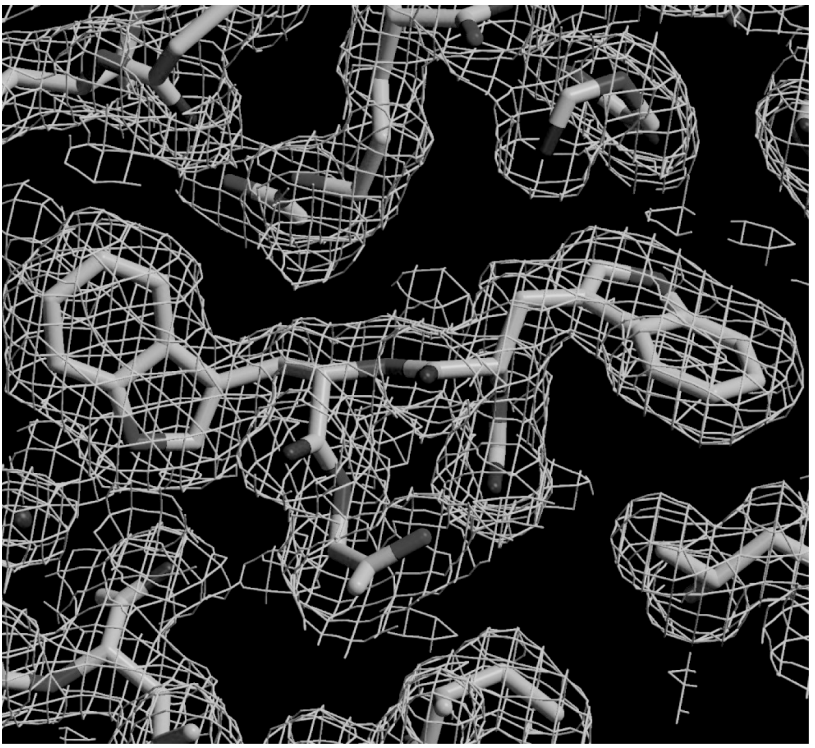

**Figure 17** A $2F_o$–$F_c$ electron density map calculated according to Eq. (13). (This figure is available in full color at ftp://ftp.wiley.com/public/sci_tech_med/drug_discovery/.)

mined for the measured reflections, incorporating these phases into the Fourier summation for the electron density [Eq. (3)] will give an electron density map, into which the model of the protein can be built.

After determination of the phases and building in the protein model, electron density maps are generally calculated by Fourier transforms using coefficients $2F_o - F_c$. The explicit expression for the electron density map calculation is [52]:

$$\rho(x, y, z) = (1/V)\sum_{hkl}(2|F_o| - |F_c|)\cos[2\pi(hx + ky + lz) - \alpha(hkl)] \qquad (13)$$

In Eq. (13), $|F_o|$ and $|F_c|$ are, respectively, the amplitudes of the reflections from the experimental diffraction data and the structure model built into the electron density, and $\alpha$ is the calculated phase for the reflection with Miller indices *hkl*. This type of electron density map shows the electron density of the calculated model, and the difference electron density of the target protein structure and the calculated model [52]. An example of an electron density map calculated according to Eq. (13) is shown in Figure 17. The electron density is usually represented in chicken-wire contours into which the protein model can be built.

In practice, successful interpretation and application of MIR data to determine a protein crystal structure is not as straightforward as may appear in Figure 17. In MIR, it is assumed that the measured intensity differences between the native protein data set P and heavy atom data sets PH*n* are due solely to the presence of the heavy atoms in PH*n*. These differences in intensity are on the order of a few percent, so great accuracy and care is required in the measurement and processing of the diffraction data. Also, the unit cell parameters between the P and PH*n* crystals must not deviate significantly. It is not uncommon for a heavy atom soak of a protein crystal to distort the crystal and change the unit cell parameters. Such heavy-atom-derivatized crystals are then of little value in the structure determination. Because of its relatively stringent experimental requirements, MIR is becoming less and less popular compared with other methods, described below. Still, the general concepts of phase determination by MIR carry through to the other methods.

## Anomalous Dispersion Methods

***Friedel's Law*** As mentioned previously, X-ray diffraction can be treated as reflections of X-rays according to Bragg's law [Eq. (5)]. In protein X-ray crystallography, the X-rays can be thought to reflect off of planes of electrons (or atoms) in the crystal. Friedel's law [Eq. (14)] states that the intensities of the reflection $F_{hkl}$ and the reflection $F_{\bar{h}\bar{k}\bar{l}}$ are equal $(\bar{h} = -h)$:

$$I_{\bar{h}\bar{k}\bar{l}} = I_{\bar{h}\bar{k}\bar{l}} \tag{14}$$

That is, the intensities of the reflections with opposite Miller indices are equal. Applied to the reflection treatment of diffraction, Friedel's law simply states that the reflection intensity off of the front side of a plane and the back side of the same plane are equal. This assumption is valid when collecting diffraction data from native protein crystals that contain no heavy atoms, and even heavy-atom-derivatized crystals used in MIR under certain wavelengths of incident radiation.

***Anomalous Dispersion*** Anomalous dispersion is the phenomenon whereby certain atoms under certain incident radiation wavelengths absorb some of the incident radiation, instead of simply scattering the radiation, as in normal diffraction. Typical anomalous scatterers used in protein X-ray crystallography are heavy atoms such as Hg, Au, and Pb, many of which are also used in MIR. Anomalous dispersion results in the breaking of Friedel's law. As a result, within a single diffraction data set, there are differences between the measured intensities of reflections that would otherwise be the same in the absence of anomalous scattering. These intensity differences can be used as in MIR to determine substructures of the anomalous scatterers in the crystal, which are then used to phase the diffraction data and determine the structures of the protein.

The main experimental requirement for anomalous dispersion X-ray diffraction data collection is similar to that of MIR, namely a protein crystal derivatized with an anomalous scatterer. Such derivatization can be done by soaking the protein crystal in a solution of an anomalous scatterer. In addition to the classic heavy metals used in MIR, halides can also give excellent anomalous diffraction data for protein structure determination [53].

The derivatization of crystals with an anomalous scatterer can also be done at the molecular biology level. Selenium is an excellent anomalous scatterer at wavelengths suitable for protein crystallography. Proteins can be expressed in bacteria and insect cells in a medium containing the amino acid Se-Met instead of Met, where the sulfur in Met is replaced with selenium. Crystals of the Se-Met protein then automatically contain an anomalous scatterer. Instead of soaking exogenous heavy atoms into a protein crystal and risking damage to the crystal, incorporation of an anomalous scatterer directly into the amino acid sequence of a protein reduces the crystal sample preparation steps and may preserve the diffraction power of the crystal.

For a given anomalous scatterer, the strength of the anomalous dispersion diffraction signal is dependent on the wavelength of the incident radiation. Therefore, another experimental requirement for anomalous X-ray diffraction data collection, which is not a requirement for MIR, is an incident radiation source tuned to the wavelength at which the anomalous dispersion signal is measurable for the anomalous scatterer in the crystal. Before the widespread use of synchrotron radiation, anomalous dispersion diffraction data were difficult to collect. Most protein diffraction data were collected on laboratory X-ray sources with fixed wavelength, usually Cu $K_\alpha$, with $\lambda = 1.54$ Å. This wavelength was not optimized to measure the largest anomalous dispersion signal for the heavy atoms typically used as anomalous scatterers. Modern synchrotron beamlines now allow very precise tuning of the incident radiation to maximize the anomalous signal, making anomalous diffraction data collection routine.

***Structure Determination by Anomalous Dispersion*** As discussed in the previous section on MIR, the unambiguous determination of phases from diffraction data generally requires the measurement of three different diffraction data sets, from the native protein crystal and at least two different heavy-atom-derivatized crystals. If one heavy-atom-derivatized crystal is available, and the heavy atom is also an anomalous scatterer, the anomalous signal from the heavy atom can be used to break the phase ambiguity associated with a single isomorphous data set (Fig. 16).

In a diffraction data set lacking any anomalous scatterers, there are numerous reflections, Bijvoet mates, which have the same measured intensity due to Friedels law [Eq. (14)]. If anomalous dispersion is present, the Bijvoet pair of reflections $F_{hkl}$ and $F_{\bar{h}\bar{k}\bar{l}}$ does not have the same intensity. These differences in the measured intensities result in different intensity measurements for the reflections $F_{hkl}$ and $F_{\bar{h}\bar{k}\bar{l}}$, whereas in the nonanomalous dispersion isomorphous

replacement case these two reflections would have the same measured intensity. A single anomalous scatterer-derivatized crystal can thus give two contributions to the Argand diagram, for the reflections $F_{hkl}$ and $F_{\bar{h}\bar{k}\bar{l}}$.

As in MIR, in anomalous dispersion methods, the positions of the anomalous scatterers are calculated from Patterson maps. The expression for an anomalous Patterson is

$$P_u = (1/a)\sum_h |F_+ - F_-|^2 \cos(2\pi hu) \tag{15}$$

Compare this expression with that for the isomorphous difference Patterson [Eq. (12)]. In the anomalous Patterson expression, $|F_+ - F_-|^2$ is the square of the difference in amplitude between the reflection $F_+$ and its Bijvoet mate $F_-$, the amplitude difference resulting from anomalous dispersion. The positions of the anomalous scatterers used in phasing the diffraction data are determined from anomalous Patterson maps (Fig. 18). As in the isomorphous difference Patterson (Fig. 15), the peaks in the anomalous Patterson correspond to the interatomic vectors between the anomalous scatterers in the anomalous diffraction data set. Analysis of the anomalous Patterson peaks can determine the anomalous scatterer-only substructure in the anomalous diffraction data set. In modern protein crystallography, direct or ab initio methods, of which the mathematics will not be discussed here, are often used to determine positions from Patterson maps.

Combined with the data from the native crystal, the phase ambiguity can be broken using just a single anomalous scatterer derivative (Fig. 16). The two different measured reflections $F_+$ and $F_-$ in the anomalous diffraction data

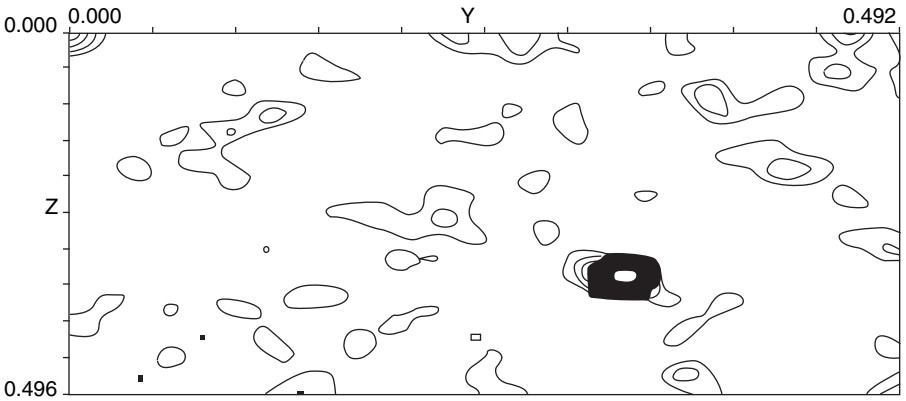


**Figure 18** An anomalous Patterson map calculated according to Eq. (15). The major peak occurring at approximately $Y = 0.33$, $Z = 0.33$ corresponds to the interatomic vector between the anomalous scatterer atoms in the anomalous scatterer-derivatized diffraction data set.

set correspond to the reflections $F_{PH1}$ and $F_{PH2}$ in Figure 16. This method of using a single anomalous scatterer-derivatized crystal for phasing is called single isomorphous replacement with anomalous scattering (SIRAS).

Since the differences in measured intensities due to anomalous dispersion are small, on the order of 3 to 5 percent of the total measured intensity [54], anomalous dispersion data must be measured and processed with great accuracy and care. Diffraction data can be collected from additional derivative crystals prepared with different anomalous scatterers. The anomalous dispersion diffraction data from these additional derivative crystals can then be combined with the native protein and first-derivative data to give additional phasing information. Such use of multiple anomalous scatterer-derivatized crystals is called multiple isomorphous replacement with anomalous scattering (MIRAS).

Protein diffraction data can also be phased by purely anomalous dispersion data, without the use of diffraction data from an underivatized protein crystal. One of the earliest widespread uses of purely anomalous dispersion diffraction data in protein X-ray crystallography was multiple-wavelength anomalous dispersion (MAD). In MAD experiments, multiple data sets are collected from a single crystal derivatized with an anomalous scatterer, with each data set collected at different wavelengths. These wavelengths are at or near the wavelength that gives the maximum anomalous dispersion signal for the anomalous scatterer. At each wavelength, there will be more or less anomalous signal from the anomalous scatterer, which will result in changes in measured intensity for a particular reflection between the data sets.

The MAD diffraction experiments are entirely analogous to MIR diffraction experiments. Whereas in MIR diffraction experiments the wavelength remains fixed and multiple different heavy atom derivatives are used to collect data sets with differences in measured intensities, MAD diffraction experiments maintain the anomalous scatterer and use multiple different wavelengths to collect data sets with differences in measured intensities. MAD experiments can be considered as in situ isomorphous replacements where physics, rather than chemistry, is used to produce the change in scattering intensity at the site [55]. Once the positions of the anomalous scatterers are located, usually by Patterson methods, the phases for the reflections are determined as in the MIR case.

In addition to MAD, there is also single-wavelength anomalous dispersion (SAD) for phasing protein diffraction data. In SAD, X-ray diffraction data are collected from a single crystal derivatized with an anomalous scatterer at a single wavelength. The intensities of the reflections $F_{hkl}$ and $F_{\bar{h}\bar{k}\bar{l}}$ are measured separately, essentially giving two data sets from a single crystal. As shown in Figure 16*a*, the Argand diagram analysis of only two diffraction data sets leads to a phase ambiguity. For a SAD diffraction data set, the two different amplitudes for the Bijvoet pair of reflections $F_+$ and $F_-$ correspond to the reflections $F_P$ and $F_{PH1}$ in Figure 16*a*. A native diffraction data set collected from an isomorphous crystal lacking the anomalous scatterer can be used to

break the phase ambiguity, which is the SIRAS technique described above. If, however, no additional isomorphous data sets are available, the correct phases can still be derived from purely SAD data.

Modern SAD protein structure determinations use probabilistic methods to determine initial phases and their reliability [53]. Once a set of phases are determined for a diffraction data set, an electron density map can be calculated [Eq. (13)]. The electron density maps can then be modified according to reasonable assumptions. A common electron density modification technique is solvent flattening. In the crystal, the interstitial regions between the protein molecules are occupied by solvent, which is usually disordered. This disordered solvent region should be relatively featureless, compared with the protein. By iteratively smoothing the electron density in the solvent region, the electron density features in the protein region will become enhanced until they become interpretable. This method of iterative electron density modification to determine protein structures from pure SAD data is called iterative single-wavelength anomalous scattering (ISAS) [56]. In general, the steps involved in determining protein structures from pure SAD data are determining the positions of the anomalous scatterers, determination of initial phases, and electron density modification until the electron density maps become interpretable.

For the crystal structure determination of proteins with unknown folds, anomalous dispersion techniques such as MAD and SAD are now much more popular than pure isomorphous replacement techniques such as MIR. One of the major advantages of anomalous dispersion over isomorphous replacement techniques is that anomalous dispersion data sets can all be collected from a single crystal. This avoids the problem of nonisomorphism, which can make MIR data collection and interpretation difficult. In MIR, multiple crystals derivatized with multiple different heavy atoms are used for data collection. Each heavy atom and its soaking conditions can have different effects on the crystal, sometimes causing significant distortions of the crystal unit cell. MIR data collected from crystals whose unit cells are significantly different from that of the native protein crystal (nonisomorphous data sets) are of little value for phasing. With MAD and SAD experiments, using cryo-data collection techniques, a single crystal derivatized with an anomalous scatterer can be used to collect all of the diffraction data sets. All of the data sets are therefore isomorphous since all of the data were collected from a single crystal.

## Molecular Replacement

The isomorphous replacement and anomalous dispersion methods are required for determination of protein structures of unknown folds. These methods assume no prior knowledge of the target protein structure and involve the initial determination of a heavy atom or anomalous scatterer substructure. These substructures are then used to calculate phases for the protein

diffraction data. If, however, there is reasonably accurate prior knowledge of the protein structure, molecular replacement (MR) methods can be used to determine the crystal structure of the target protein.

In MR structure determinations, the experimental diffraction data are probed with a search model that is assumed to be a good structural homolog of the target protein. The best search model would be the structure of the target protein itself. Structures of proteins with high sequence homology are also good search models in MR. In general, search models with more than 50 percent sequence identity to the target protein should give a correct structure of the target protein by MR [54]. Species isomorphs of a target protein, which generally share about 75 percent or greater sequence identity, often work for MR structure determinations.

As with isomorphous replacement and anomalous dispersion methods, the correct atomic positions must be determined in order to calculate phases for the diffraction data. In isomorphous replacement and anomalous dispersion methods, novel substructures are used as the atomic positions for initial phasing. In MR, the structure of the isolated target protein is already known or assumed to be very similar to that of a known protein structure. The MR calculations place the known structure, the search model, in the correct crystal lattice according to the experimental diffraction data. In a sense, isomorphous replacement and anomalous dispersion methods determine "protein" structures in that these methods assume no prior knowledge of the target protein structure. MR methods, in contrast, determine "crystal" structures in that the entire structure of the target protein is already known, and the calculations are merely (but not always so simply) fitting the known structure to the crystal lattice of the diffraction data.

Although there are many different algorithms and software packages for MR calculations, the basic computations are illustrated in Figure 19. The experimental diffraction data contains information about the structure and crystal lattice arrangement of the target protein. A suitable search model is first rotated so that its orientation is the same as the target protein in the experimental crystal with respect to a certain coordinate system. The rotation is usually expressed as a set of three Eulerian angles $\alpha$, $\beta$, and $\gamma$. In this rotation search, the Patterson vectors for the various orientations of the search model are compared with those of the target protein in the diffraction data. As described previously, the Patterson vectors are the interatomic vectors in the protein in the crystal. One can imagine two types of interatomic vectors, intramolecular and intermolecular, the intramolecular vectors generally being shorter than the intermolecular vectors. The intramolecular vectors occur entirely within a single protein molecule, and their orientations define the orientation of the protein. In the rotational search in MR, the shorter intramolecular Patterson vectors in the search model are compared with those of the target protein in the diffraction data. The search model is then rotated until the intramolecular vectors of the search model and target protein have high correlation.

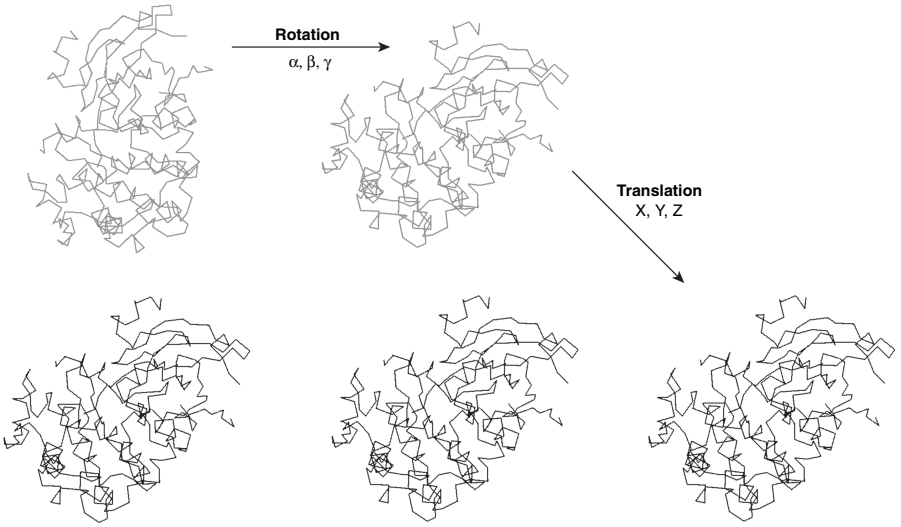


**Figure 19** Schematic of an MR structure determination. (left) The target protein structure is drawn in black lines. The search model protein structure in an arbitrary orientation and position with respect to the target protein is drawn in red lines. (middle) The search model is rotated so that its orientation corresponds to that of the target protein structure as indicated by the intramolecular Patterson vectors in the experimental diffraction data. (right) The search model is translated so that it is correctly placed in the unit cell of the experimental diffraction data. (This figure is available in full color at ftp://ftp.wiley.com/public/sci_tech_med/drug_discovery/.)

After the rotational search, the correct placement of the oriented search model in the target protein crystal unit cell is determined by a translational search. In the translational search, the oriented search model is translated within the unit cell of the crystal until the crystal packing of the search model matches that of the experimental diffraction data. There are various methods for the translational search. In one method, the longer intermolecular Patterson vectors are compared with those of the experimental data at each translation of the search model. The highest correlation between the sets of intermolecular Patterson vectors is generally the correct solution for the crystal structure of the target protein.

Compared with isomorphous replacement and anomalous dispersion methods, crystal structure determination by MR has much less experimental requirements. No special derivatization of the sample crystal with heavy atoms or anomalous scatterers is required. Only a single data set from the crystal of the target protein is required. Also, no particular radiation wavelength is required for data collection. The ability to use diffraction data collected at any wavelength makes structure determination by MR readily accessible since all of the necessary diffraction data can be collected on a standard laboratory X-ray source (assuming adequate diffraction power of the crystals).

At this point, one may wonder why go through the effort of determining the "crystal" structure of the target protein by MR if the "protein" structure

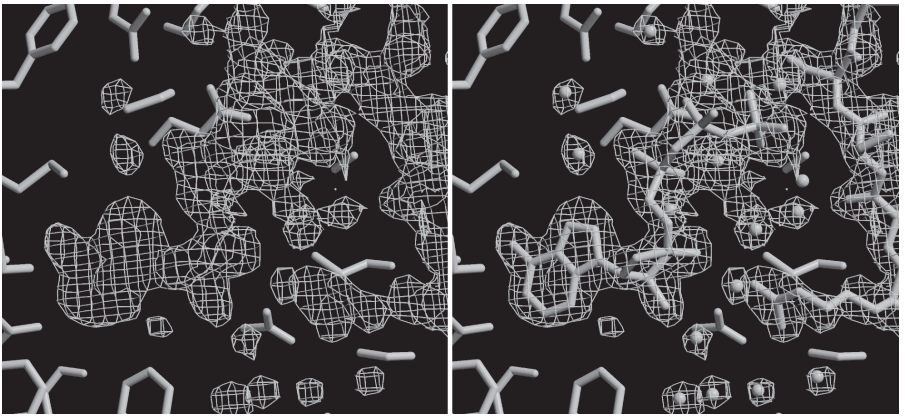

**Figure 20** $F_o$–$F_c$ electron density map calculated from an MR structure determination. *(Left)* The structure of a protein–ligand complex has been determined by MR using the protein-only structure as the search model. The $F_o$–$F_c$ electron density map calculated from the protein-only MR solution reveals electron density (blue chicken wire contours) which does not correspond to the search model structure (yellow bonds). *(Right)* The structure of the bound ligand is built into the difference electron density in the $F_o$–$F_c$ electron density map. Additional protein residues not present in the search model and bound waters are also built into the difference electron density. (This figure is available in full color at ftp://ftp.wiley.com/public/sci_tech_med/drug_discovery/.)

is already known or assumed to be that of the search model. MR crystal structure determination is required because the search model may not be complete. Although the overall structure of the target protein may already be known, the search model will lack certain structural features of the experimental data, such as amino acid mutations, insertions, and deletions, and the binding modes of ligands bound to the target protein. These additional features are not incorporated into the search model during MR calculations. The electron densities of these additional structures can only be calculated after determining the correct crystal structure of the target protein, and using phases determined from the correct placement of the protein in the crystal.

Once the search model has been correctly rotated and translated with respect to the diffraction data, difference electron density maps, using Fourier coefficients $F_o - F_c$, reveal structures that were not part of the search model. The explicit expression for this type of electron density map is

$$\rho(x, y, z) = (1/V)\sum_{hkl}(|F_o| - |F_c|)\cos[2\pi(hx + ky + lz) - \alpha(hkl)] \qquad (16)$$

Compare this expression with that of the $2F_o - F_c$ electron density map [Eq. (13)]. The $F_o - F_c$ will show electron density in regions where the search model and the target protein structures differ and no electron density where the two structures are the same. Figure 20 illustrates the MR structure determination of a protein–compound complex. The experimental diffraction

data were collected from a crystal of the protein–compound complex. A protein-only search model was used in the MR structure determination. The $F_o - F_c$ electron density map reveals the electron density of bound ligand, which was not present in the search model. Regions of the map where the search model and the target complex structure are the same are relatively featureless.

MR structure determination is now the crystallographic computation engine for X-ray structure-based drug design. Protein–compound complex structures are determined by MR from diffraction data collected from crystals of protein–compound complexes, either co-crystallized from protein–compound complex samples or having the compound soaked into protein crystals. Using MR structure determination, difference electron density maps can be calculated within minutes of collecting and processing the diffraction data to reveal the binding mode of the compound. Often, the protein–compound complex crystals are isomorphous with previous protein–compound complexes and the apo-protein crystals. In such cases, the MR solution is a trivial zero rotation and zero translation transformation of the search model (the target protein structure), and the model can proceed directly to structure refinement, which will be discussed in following sections.

All of the protein diffraction data phasing methods described above are now handled by automated software packages. These software include CNX (CNS), CCP4, PHASES, SOLVE, SHARP, SHELX, and EPMR, among others. Due to the great advances in the automation of crystallographic computing and computing power in general, the phase problem, which was once the central task of protein structure determination, is now less of a concern in protein X-ray crystallography.

## 5 GENERATION AND ANALYSIS OF STRUCTURAL MODELS

### Aspects of Crystallographic Models of Macromolecules

The end result of crystallographic structure determination is the generation of an accurate three-dimensional model of the contents of the asymmetric unit of the crystal. In the simplest terms this model consists of the *spatial (xyz) coordinates* and elemental type for each atom. In most cases a great deal more information is included explicitly in the model or implied by convention or context. Protein Data Bank (PDB) file format coordinate records for each atom typically contain *xyz* coordinates, thermal motion parameters (*B factor* or temperature factor), and relative *occupancy* (0.0 to 1.0). B factors represent the thermal vibration of each atom (see previous section). Individual atom thermal motion can be modeled either as isotropic or anisotropic motion. In the isotropic case the atom is modeled as if it vibrates with equal amplitude in three orthogonal directions, producing a spherical time-averaged distribution of electron density. Isotropic thermal motion can be represented

by a single B-factor parameter that is proportional to the mean-square amplitude of vibration. In the anisotropic case the atom is modeled with vibrations of differing amplitude in three orthogonal directions, producing an ellipsoidal time-averaged distribution of electron density. Complete description of anisotropic thermal motion requires six parameters ($U_{11}$, $U_{22}$, $U_{33}$, $U_{12}$, $U_{13}$, $U_{23}$) that are expressed as the mean-square amplitude of vibration in angstroms. These parameters serve to define the orientation of the axes of the ellipsoid and the mean-square amplitude of vibration along each axis. Atomic resolution data (1.0 Å resolution or better) is required to properly refine anisotropic B factors.

The last parameter, occupancy, determines the fraction of unit cells within the crystal that contain the given atom or molecule. This parameter is typically applied to ligands, ordered water molecules, and other ions and small molecules that are often observed bound to specific sites in proteins and other macromolecules. Because protein–protein contacts form the lattice interactions that make up the crystal, the atoms of the protein region of the crystal are in general constrained to have occupancies of 1.0. Exceptions to this are instances in which multiple conformations for individual residues or regions of the peptide chain can be discerned in the electron density maps. PDB format ATOM records for models with isotropic B factors contain at least 10 entries (atom number, atom name, residue type, chain name, residue number, X, Y, Z, B, Occ), but the PDB format allows a number of additional entries such as segment identifiers (equivalent to chain name) used by the XPLOR/CNS suite [57] and atom types and charges used by Refmac [58]:

```
ATOM 22 OH   TYR A 5 -42.861  -17.317  103.842  1.00  75.00  O
ATOM 23 N    ILE A 6 -35.203  -14.922  106.691  1.00  32.68  N
ATOM 24 CA   ILE A 6 -34.025  -14.142  106.955  1.00  31.42  C
ATOM 25 C    ILE A 6 -33.743  -13.217  105.782  1.00  30.29  C
ATOM 26 O    ILE A 6 -33.841  -13.620  104.644  1.00  51.93  O
ATOM 27 CB   ILE A 6 -32.833  -14.967  107.379  1.00  39.55  C
ATOM 28 CG1  ILE A 6 -32.291  -14.302  108.607  1.00  42.56  C
ATOM 29 CG2  ILE A 6 -31.732  -14.979  106.344  1.00  32.26  C
ATOM 30 CD1  ILE A 6 -31.813  -15.334  109.585  1.00  75.00  C
```

PDB format atom definitions for models with anisotropic B factors require two records. The first is a standard ATOM record with an equivalent B factor calculated from an average of $U_{11}$, $U_{22}$, and $U_{33}$. The second ANISOU card lists $U_{11}$, $U_{22}$, $U_{33}$, $U_{12}$, $U_{13}$, and $U_{23}$ sequentially, but scaled by a factor of $10^4$:

```
ATOM   107 N  GLY 13  12.681 37.302  -25.211 1.000  15.56      N
ANISOU 107 N  GLY 13 2406   1892    1614     198    519    -328  N
ATOM   108 CA GLY 13  11.982 37.996  -26.241 1.000  16.92      C
ANISOU 108 CA GLY 13 2748   2004    1679     -21   155     -419  C
ATOM   109 C  GLY 13  11.678 39.447  -26.008 1.000  15.73      C
```

```
ANISOU 109 C  GLY 13 2555   1955    1468     87   357      -109  C
ATOM   110 O  GLY 13  11.444 40.201  -26.971 1.000  20.93      O
ANISOU 110 O  GLY 13 3837   2505    1611     164  -121     189  O
ATOM   111 N  ASN 14  11.608 39.863  -24.755 1.000  13.68      N
ANISOU 111 N  ASN 14 2059   1674    1462     27   244      -96  N
```

Bonding and connectivity between atoms and residues is not in general defined explicitly in PDB format, but in most cases it is implicit in the relationship between atom names, residue names, residue numbers, and chain names. However, it is possible to explicitly define bonding and some nonbonding interactions with CONECT (general), LINK (general), SSBOND (disulfide bond), HYDBOND (hydrogen bond), and SLTBRG (salt bridge) records.

A complete description of the PDB format and all allowed records can be found at the Research Consortium for Structural Biology Web page (*www.rcsb.org*). In addition to the individual atom properties and connectivity, other parameters introduced during refinement such as noncrystallographic symmetry restraints, bulk solvent corrections, and TLS groups are important components of the final model. These aspects of model refinement will be discussed in more detail under "Refinement and Analysis of Structural Models."

## Building the Initial Model

Building a de novo model of a macromolecule into initial electron density maps is often an arduous task. There are a number of software packages in common use (Quanta, XtalView, O, TurboFrodo, Coot) with more or less sophisticated tools to aid in interactive model building. The efficiency of the process depends on the quality of phases and the resolution of the map, the sophistication of the model-building software employed, and the crystallographer's familiarity with the molecule of interest as well as many general features of protein and nucleic acid structure.

Starting with only the primary amino acid sequence, a great deal can be learned about a given protein using freely available online database and analysis resources. The homology search engine BLAST (Basic Local Alignment Search Tool [59]), available from the National Center for Biotechnology Information (NCBI, *www.ncbi.nlm.nih.gov*), allows rapid identification of proteins with homologous regions of sequence, some of which might have known structures in the RCSB Protein Data Bank that can serve as templates for model building. Multiple sequence alignments of a homologous family of proteins will often reveal highly conserved residues that are crucial for structure or activity. JPRED [60] from the Barton group at the University of Dundee provides multiple sequence alignment with consensus secondary structure prediction. NCBI and the ExPASy (Expert Protein Analysis System) proteomics server of the Swiss Institute of Bioinformatics both provide excellent suites of

tools for primary structure analysis. These tools allow identification of known amino acid motifs, such as substrate or cofactor binding motifs (e.g., the Rossman fold, P-loop motif, iron-sulfur clusters) and protein domains with known homologous structures that might be represented in the sequence of interest.

The SCOP ([61]; *scop.mrc-lmb.cam.ac.uk/scop/*) and CATH ([62]; *http://www.biochem.ucl.ac.uk/bsm/cath/*) online databases provide classification of protein structure with examples of each represented fold. An excellent resource for learning the basics of protein structure from peptide conformations to common folds is *Introduction to Protein Structure* by Carl Branden and John Tooze (Garland Science, UK). Knowledge of simple common aspects of tertiary structure such as the common helix-helix knobs-in-holes packing angles, the left-handed twist of β sheets and the right-handed twist of individual β strands, and the ways in which loop regions most commonly connect secondary structures will assist greatly in interpretation of blurry regions of electron density and recognition of secondary structure elements at low resolution. Another valuable reference is the classic study by P.N. Lewis, F.A. Momany, and H.A. Scheraga [63]. This study details the peptide conformations of the tight turns and chain reversals often encountered at the surfaces of proteins.

The main objectives of initial model building are to find the correct trace for the peptide chain and to make the correct sequence assignment for each residue. Before beginning actual model building, it is advisable to examine a number of electron density maps calculated to different resolutions and with different weighing [figure of merit or $\sigma(A)$] and density modification schemes (solvent flipping or flattening, histogram matching, or skeletonization). Less accurate phases for the highest resolution data might create noise that makes interpretation of the maps difficult. Initial maps calculated at lower resolution might reveal the trace of the peptide chain more clearly. The percentage of solvent content used in solvent flattening calculations can have a dramatic impact on the quality of the map, and the best estimate of the real solvent content will not necessarily produce the most interpretable electron density maps. Several values above and below the best estimate of the true value should be tried.

At this early stage density skeletonization, or "bones," calculations are a tremendous aid in determining the quality of electron density maps. The overall connectedness of the electron density can be gauged and secondary structure elements can often be recognized. Skeletonization calculations produce traces through the highest electron density points in the map. In well-phased electron density maps the bones will clearly reveal the trace of the peptide chain. In a poorly phased map the bones will be fragmented. Some arbitrary adjustment of the parameters used in the skeletonization calculations is often required to obtain the best results. If the map has been phased experimentally by isomorphous replacement or anomalous dispersion methods, it can be helpful to overlay the positions of the phasing sites onto the electron

density. Heavy atoms are likely to bind to only a few reactive side chains [37], and if phases are derived from seleno-methionine anomalous dispersion data, the phasing positions will instantly reveal the location of all methionine residues in the structure. This information can dramatically speed up the initial chain tracing and sequence assignment. In addition, for heavy atom derivatives the maps should be examined for severe rippling in the vicinity of the phasing sites that indicates improper refinement of occupancy or B factor for the heavy atom. Adjusting these parameters until the rippling is minimized can make valuable improvements in the phases and thus in the overall quality of the maps.

Once suitable maps are obtained and optimal skeletonizations have been produced, chain tracing can begin. The most basic method of chain tracing is "baton" building in which a $C_\alpha$-only trace is built interactively by sequentially adding about 3.5-Å-long baton segments. This is relatively straightforward with a map that is good enough to approximately locate the $C_\alpha$ positions by following the bones and observing the side-chain branch points. Some modern model-building software packages (e.g., QUANTA) have automated chain-tracing features that can build a $C_\alpha$ trace rapidly or even identify and automatically build secondary structure elements. It is also possible to use real-space rotation and translation searches to place models of secondary structures or even entire domains into initial electron density maps (FFFear, ACORN). These techniques can save a great deal of time by fitting large sections of standard secondary structure or approximately correct tertiary structure that can then be corrected interactively or by refinement software. Some modern model-building software packages are equipped with libraries of loop, turn, and random coil segments taken from known protein structures that can be screened rapidly for a good fit to difficult regions of electron density. As a general rule, segments that cannot be built with confidence should be left out of the initial model. Phase improvement by refinement of correctly built structure will improve the overall map and gradually allow interpretation of the difficult regions.

Once at least a partial chain trace is in hand, sequence assignment can begin. As mentioned above, phasing positions of heavy or anomalous atoms can occasionally be used as starting points for sequence assignment. Identification of the active site or bound ligands will aid in locating residues known to be involved in catalysis or binding. Aromatic residues such as tryptophan, tyrosine, and phenylalanine are often easily recognized because they are large and are most often located in the interior of proteins where their conformational entropy is low and the quality of the electron density is good. Because solvent flipping and flattening masks are not perfect, these procedures sometimes degrade the quality of electron density for surface loops and side chains. Disulfide bonds can often be recognized by strong connecting electron density. In fact, the electron density of disulfide bonds is often strong enough that it confuses early chain-tracing attempts, particularly automated procedures. If data has been collected using a long wavelength, such as Cu $K_\alpha$,

an anomalous difference map might reveal the locations of many sulfur atoms from methionine and cystine residues.

After sequence assignment a full-atom model of the peptide chain must be constructed and fit to the electron density. It is best to use a model-building software package that does automated full-atom model construction starting from a sequence-assigned $C_\alpha$ trace. The quality of auto-built models is often low, requiring manual adjustment to fit the electron density via additional interactive model building. Real space fitting, refinement, and regularization protocols allow this to be done rapidly. These algorithms can use either the electron density map contours (grid methods) or gradients calculated from these contours in least-squares algorithms to provide the best fitting position and conformation of residues or segments of peptide chain. Peptide conformations are fit either by iteratively adjusting or searching atom positions and torsions, by searching rotamer and conformation libraries, or by Monte Carlo methods. Grid-based methods have the smallest radius of convergence and are most appropriate for fitting single residues or side chains. Gradient methods have a larger radius of converge (several angstroms) and are more generally applicable, particularly in combination with Monte Carlo methods, for fitting longer segments of peptide chain. Because the radius of convergence of standard reciprocal space refinement techniques is small, it is important to obtain a reasonably good fit of atom positions in the initial model building.

## Refinement and Analysis of Structural Models

Refinement is an iterative process that modifies the parameters of the model in a rational way to agree with the structure factor amplitudes derived from the observed diffraction intensities. This requires an error function that gauges the agreement between the structure factors predicted by the model and those derived from the observed data as well as an algorithm that alters the parameters of the model in such a way as to improve this agreement while maintaining a model that is both chemically and physically reasonable. Upon each iteration the model and the calculated phases improve, yielding better electron density maps that will allow the model to be extended into previously uninterpretable regions. These new electron density maps reveal more clearly the electron density for other well-ordered molecules such as waters of solvation, ions, and ligands. Unless extremely accurate experimental phases are obtained (e.g., from high-resolution MAD data), refinement is essential for the complete interpretation of the data.

***Refinement Methods*** The error function used in protein refinement is typically either of the "least squares" or "maximum likelihood" type. Only the most cursory description of the mathematics of refinement can be presented here.

The least-squares method uses the sum-squared difference between the observed structure factors and those calculated from the model:

$$Q = \sum_{hkl} w_{hkl} (|F_o(hkl)| - k|F_c(hkl)|)^2 \tag{17}$$

where $w(hkl)$ is a weight, often resolution dependent in macromolecular refinement, and $k$ is a scaling constant that allows meaningful comparison between $|F_o|$ and $|F_c|$. The summation is over all $m$ $|F_o|$, which constitutes a set of $m$ equations. Each corresponding $|F_c|$ is a function of the $n$ parameters of the model, which constitute $n$ unknown parameters. The sum-square error function can then be represented as a system of equations in matrix form. For a given set of starting model parameters, the system of equations can be solved by iterative methods using a truncated Taylor expansion to represent $|F_c|$ as a function of the $n$ parameters of the model and simplified block-diagonal least-squares matrices in which all terms that are not highly correlated (such as B factor and occupancy, if occupancy is being refined) are set to zero. Because of these approximations, multiple iterations, each corresponding to small shifts in the values of each model parameter, are required to reach convergence.

Maximum-likelihood methods are statistical and derive from Bayes's theorem. Bayes's theorem is expressed by the equation

$$P(A) * P(B \mid A) = P(B) * P(A \mid B) \tag{18}$$

In words, Eq. (18) states: (the probability of $A$) times (the probability of $B$ assuming that $A$ is true) is equal to (the probability of $B$) times (the probability of $A$ assuming that $B$ is true). The corresponding expression for crystallographic refinement is

$$P(F_o) * P(\text{model} \mid |F_o|) = P(\text{model}) * P(|F_o| \mid \text{model}) \tag{19}$$

It is more convenient to introduce $F_c$ as a function of the model parameters and write

$$P(F_o) * P(|F_c| \mid |F_o|) = P(|F_c|) * P(|F_o| \mid |F_c|) \tag{20}$$

However, the *prior probability* $P(|F_c|)$ and the *normalizing factor* $P(F_o)$ are constants and can be omitted without altering the essence of the relationship. The remaining terms are the *posterior probability* $P(|F_c| \mid |F_o|)$ and the *likelihood* $P(|F_o| \mid |F_c|)$. In the absence of the constants $P(|F_c|)$ and $P(F_o)$ this last term cannot be properly designated as a probability and is recast as the *likelihood* $\boldsymbol{L}(|F_o| \mid |F_c|)$ leaving the expression

$$P(|F_c| \mid |F_o|) \propto L(|F_o| \mid |F_c|) \tag{21}$$

The total likelihood is expressed as the joint probability for all $F_o$:

$$L_{\text{total}} = \prod_{hkl} P(|F_{hkl,o}| \mid |F_{hkl,c}|) \tag{22}$$

but in practice the maximum-likelihood error function is often the negative log of the likelihood

$$-\log L_{\text{total}} = -\sum_{hkl} \log[P(|F_{hkl,\text{o}}| \,|\, |F_{hkl,\text{c}}|)] \tag{23}$$

The precise functions used to calculate the likelihood are quite complex and will not be presented here.

***Restraints and Parameterization in Refinement*** As described above, the relationship between observed diffraction data and the parameters of the model could be represented as a system of $m$ equations ($F_o$) and $n$ unknowns (parameters of the model that determine $F_c$). If the ratio of $m/n \geq 1$, the system is said to be determined or overdetermined and will have either one or infinitely many solutions. If the ratio of $m/n$ is $<1$, the system is underdetermined and will have infinitely many solutions. In principle, this system of equations should have a solution corresponding to the exact structure within the crystal. In practice, because of experimental errors in the measurement of $F_o$, resolution limits of the data due to defects in the crystals, and limitations of our theoretical models, it is rarely possible to obtain a solution that defines the model precisely. In cases where the data are of atomic resolution (about 1.0 Å or better), it is possible to refine individual atoms according to their $xyz$ coordinates, occupancy, and B factors (isotropic or anisotropic) and obtain an acceptable solution. The system of equations is determined or overdetermined with respect to these parameters.

When the data are not of atomic resolution, the system of equations is underdetermined with respect to the parameters of the model. An acceptable solution cannot be obtained unless additional restraints (or constraints) are added to the model in order to increase the effective number of observations that determine the model. Most of these restraints bear upon molecular geometry. Prior knowledge of the structure and chemistry of organic molecules from many sources (not the least of which are atomic resolution structures of organic compounds including peptides, amino acids, and nucleotides) allows the design of libraries of ideal values and standard deviations for the van der Waals radii, bond lengths, angles, dihedral angles, and planarity (of peptide bonds, aromatic rings, carboxylates, and amides) of the molecular structures of proteins and other macromolecules. These ideal values and standard deviations are applied as restraints that are designed to maintain a chemically and physically reasonable model during refinement. These restraints take the form of new terms added to the refinement error functions.

Generating proper library definitions and restraints for novel ligands is not a trivial problem, and an accurate definition is a prerequisite for correct interpretation of the data. It requires complete knowledge of the chemical structure, bonding, hybridization, protonation, ionization, and chirality. However, most common representations of molecules are highly simplified abstractions that do not properly describe the true nature of the chemical species involved.

In some cases only careful experiments or a high-resolution crystal structure of the compound will resolve uncertainties. In the absence of such data some guesses must be made. The standard two-dimensional chemical structure representation of the anticancer drug topotecan (Hycamptin) is given in Figure 21 as an example. It is not necessarily obvious from the two-dimensional representation of the structure that N3 of the tertiary amine substituent is $sp^3$ hydridized and could be protonated depending upon the pH, C10 is $sp^3$ hybridized and is not part of the conjugated ring system, or that the pyridone ring (D) is aromatic. This information was determined partly from standard chemical knowledge and partly from the crystal structure of a related compound, camptothecin iodoacetate [64].

Hydrogen atoms are usually excluded as part of the refined model unless the resolution is exceptionally high. This is due to their small contribution to X-ray diffraction (only one electron) and their large contribution to the total number of parameters in the model. Some refinement packages allow the model to be refined with riding hydrogens (hydrogens placed in standard positions with standard bond lengths and angles) added to the model. The hydrogens are not refined but do have an effect on refinement, somewhat in $F_c$, but mainly in the van der Waals terms, which can lead to improved distributions of torsion angles.

Further restraints can be added to take advantage of correlated vibrational properties. Simple B-factor restraints account for the physical reality that the vibrational amplitudes of covalently bonded atoms can be correlated. A recent advance in model parameterization has been the introduction of TLS groups in refinement [65]. TLS groups are substructures of the model (sometimes entire domains) that have correlated vibrational properties. These correlated vibrations are described by three tensors that designate the translational (T), librational (oscillation) (L), and screw (S) components of the motion. TLS groups account for an entire continuum of conformations of part of the model in a parsimonious manner with respect to the number of parameters required.

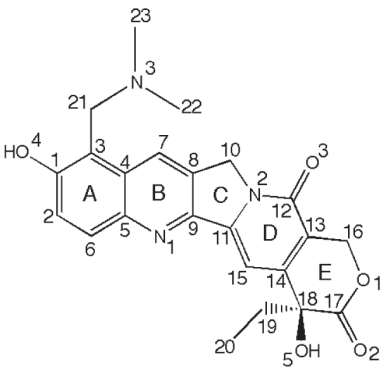


**Figure 21** Standard two-dimensional chemical structure representation of topotecan.

The greatest difficulty here is the definition of the TLS groups. Unless the data are of good enough quality and high enough resolution to properly refine anisotropic B factors, it is difficult to determine which regions of the structure have correlated vibrations (and in this case TLS groups will not be necessary). In general educated estimates and trial-and-error must guide the choice of TLS groups. An example of TLS group definition with the refined T, L, and S tensors is shown below.

```
REMARK   7 TLS DETAILS
REMARK   7 NUMBER OF TLS GROUPS :   2
REMARK   7
REMARK   7 TLS GROUP :    1
REMARK   7  NUMBER OF COMPONENTS GROUP :   1
REMARK   7  COMPONENTS     C SSSEQI   TO C SSSEQI
REMARK   7  RESIDUE RANGE :   A   865     A 1110
REMARK   7  ORIGIN FOR THE GROUP (A): 45.6660 11.3419 41.2873
REMARK   7  T TENSOR
REMARK   7   T11:   0.0157 T22:   0.0641
REMARK   7   T33:   0.0294 T12:   0.0162
REMARK   7   T13:   0.0031 T23:   0.0101
REMARK   7  L TENSOR
REMARK   7   L11:   0.8473 L22:   1.0405
REMARK   7   L33:   1.1744 L12:   0.3695
REMARK   7   L13: -0.0069 L23:   0.3502
REMARK   7  S TENSOR
REMARK   7   S11: -0.0245 S12: -0.0044 S13:   0.0036
REMARK   7   S21: -0.0046 S22:   0.0421 S23: -0.0484
REMARK   7   S31: -0.1460 S32:   0.0620 S33: -0.0176
```

If more than one chemically identical molecule is present in the asymmetric unit, improper or noncrystallographic symmetry (NCS) restraints (or constraints) can be applied that force the conformations of the molecules to be similar (or identical). If constraints are used, the molecules are forced to be identical and the effective number of the constrained parameters of the model is reduced by a factor of $n$ for $n$-fold NCS. If restraints are used, the reduction in the effective number of parameters is less clear. Noncrystallographic symmetry is only useful in cases were the data are of poor quality or rather low resolution (about 3.0 Å or worse). In general the packing interactions of the NCS-related molecules will be different, and real differences in conformation of both the main chain and side chains can be expected and should not be ignored unless the data is of insufficient quality to observe them.

The parameters of the model discussed so far bear only upon the structure of the ordered phase of the crystal, but the disordered phase (solvent region) of the crystal, which can be 30 to 80 percent of the unit cell volume, also contributes to the intensity of $F_o$, particularly at low resolution. A suitable model for this contribution is required and can make significant improvement in the difference electron density of unbuilt regions of the model. The solvent X-ray

scattering in essence introduces a contrast that significantly decreases the observed diffraction intensities relative to what would be observed if the disordered region of the crystal were vacuum. The time-average electron density in the solvent region of the crystal is constant and can be represented by the parameter $K_{sol}$, the ratio between the average solvent electron density and the average protein electron density. A second parameter $B_{sol}$ is effectively a thermal factor that is used in a resolution-dependent exponential scaling term: Because different regions of the solvent do not scatter in phase, the intensity of solvent scattering is very weak at high resolution but quite strong at low resolution. The most common bulk solvent scaling function used is based upon Babinet's principle [66], which states that the Fourier transform of a mask and its complement have structure factors of the same amplitude but phases of opposite sign. For this method only the two parameters $K_{sol}$ and $B_{sol}$ are introduced to the model and only the amplitudes of the structure factors are changed. This function is used by most current refinement packages. A second method is the mask method in which the phases for the solvent scattering are calculated from a grid of dummy atoms filling the solvent region of the model. For this method $K_{sol}$ and $B_{sol}$ are also determined, but both the amplitudes and the phases of the structure factors are altered. This method is used by the XPLOR/CNS [57] refinement package.

***Weighting and Judging Progress of Refinement*** The model restraint terms and the X-ray diffraction terms of the refinement error function both act upon the fundamental parameters of the model (*xyz*, B, occupancy). An overall weighting factor balances the contributions of these opposing aspects of the error function. Each restraint term is also given a weight that can be adjusted to give the desired final root mean square (rms) deviations from the ideal (library) values. For example, in least-squares refinement the error function with geometric restraints can be represented as:

$$\begin{aligned} Q = {} & W_{\text{X-ray}}\left[\sum\nolimits_{hkl} w_{hkl}(|F_{\text{o}}(hkl)| - k|F_{\text{c}}(hkl)|)^2\right] \\ & + \sum\nolimits_{\text{distance}} w_{\text{dist}}(D_{\text{ideal}} - D_{\text{model}})^2 \\ & + \sum\nolimits_{\text{angle}} w_{\text{angle}}(\theta_{\text{ideal}} - \theta_{\text{model}})^2 \\ & + \sum\nolimits_{\text{torsion}} w_{\text{torsion}}(\chi_{\text{ideal}} - \chi_{\text{model}})^2 \end{aligned} \tag{24}$$

and so on for all restraint terms

Proper choice of weights is essential for obtaining the best fit to the observed data while maintaining a chemically and physically reasonable model. The agreement between the model and the observed data is judged by the residual or $R_{\text{factor}}$ defined as:

$$R = \sum \|F_{\text{o}}| - k|F_{\text{c}}\| \,/\, \sum |F_{\text{o}}| \tag{25}$$

The $R_{\text{factor}}$ gives a normalized sum-error between the observed and calculated structure factors and is often expressed as a percentage. As the model comes to give a better representation of the observed data, $k|F_c|$ will approach the value of $|F_o|$ and the numerator in Eq. (25) will decrease. Thus the overall $R_{\text{factor}}$ will decrease as refinement progresses. The $R_{\text{factor}}$ is the most commonly sited metric of correctness of fit for refinement, but many others including correlation coefficients, figures of merit, $\sigma(A)$, and some variants of the $R_{\text{factor}}$ are also often calculated and output by refinement software. Both least-squares and maximum-likelihood methods allow estimations of the standard uncertainties of the *xyz* coordinates, occupancies, and B factors. Lower standard uncertainties should correlate with improvements in $R_{\text{factor}}$ and the other metrics. All refinement programs will output a list of rms deviations from ideality for all restraint terms. These data are a crucial guide for achieving a properly refined model and for judging the overall X-ray vs. geometry weight. In addition to the overall statistics for restraints, refinement programs will also list specific bonds, angles, torsions, and the like that deviate significantly (greater than three standard deviations at least) from the ideal values. These outliers often indicate regions of the model that require further interactive adjustment before refinement will converge.

In addition to judging the progress of refinement toward the best-fit model for the data, the refinement results must be used to gauge the appropriateness of the parameterization of the model. If the observation/parameter ratio is low, it is possible to overfit the model and produce a refined structure that appears to fit the data much better than it actually does. This is analogous to fitting a two-exponential decay with a three-exponential equation: The fit might very well be better, but the model parameters obtained will not represent the true nature of the data. An obvious case of this problem would be refinement with anisotropic B factors against data to no better than about 2.5 Å. The extra parameters will allow an apparent better fit to the data, but they will not represent true individual atom anisotropic thermal motion because it is not properly represented in the data.

In most cases the choice of model parameterization requires more subtle judgment, especially for restraints and weights. The free $R_{\text{factor}}$ ($R_{\text{free}}$) was introduced as a means of validating the choice of model parameterization. $R_{\text{free}}$ is a standard $R_{\text{factor}}$, but it is calculated against a small subset of the data (5 to 10 percent) that is set aside and not used in model refinement. The agreement between these data and the model, therefore, provides an unbiased gauge of the appropriateness of the model parameterization. If the model parameters are appropriate, the $R_{\text{factor}}$ and $R_{\text{free}}$ will both decrease as refinement progresses and the agreement between the model and the working data is improved. If the data are being overfit the $R_{\text{free}}$ remains static or increases even though the $R_{\text{factor}}$ continues to decrease. The standard $R_{\text{free}}$ is a limited application of the notion of the jackknife test in which data are binned randomly and the model fit many times. Each time the model is fit with a different bin omitted and the result tested for agreement against the omitted data. This prevents the inter-

pretation of the refinement being skewed by an unfortunate initial choice of test data for $R_{free}$ calculations.

## Analysis and Preparation of Structural Models

No refined crystallographic model is without errors and even the best refined model is unlikely to be an ideal starting place for computational work without some further modification. A number of protein structure validation software packages are available (ProCheck, WhatIf) to assist in locating and annotating errors. A first step is to make certain that the model represents the expected sequence. It is not uncommon to encounter residues in the model (commonly surface-exposed residues such as lysine, arginine, glutamate, and glutamine) that are represented as alanine because the side chains could not be distinguished in the electron density maps. In other cases, the residue might be represented with the proper name but have missing atoms. In high-resolution structures some residues might be modeled in several alternative conformations. For most computational modeling only one of the conformations can be chosen. Last but not least, inverted chiral centers must be identified and corrected. Some of this information could be included in the header of a standard PDB file. In addition, a PDB file typically represents the contents of the asymmetric unit of the crystal. In some cases the biologically relevant molecule could be only part of the asymmetric unit or it could span several asymmetric units. If the biologically relevant molecule does span more than one asymmetric unit, then the appropriate matrices for generating the complete molecule are most often included in the header of the PDB file.

Primary literature reporting protein X-ray crystal structures usually include a table, often called Table 1, summarizing the X-ray diffraction data. Results for the structural model refinement are also sometimes included in this table. A typical Table 1 is shown on page 433.

Geometric refinement statistics are a good indication of the overall quality of the model, but they can also mask specific instances of significant deviation from reasonable values. Geometric outliers (bonds, angles, dihedrals, and planes that deviate significantly from ideal values) are the first indication of problem areas of the structure. Any restraint that differs by more than 10 $\sigma$ from that of the ideal value should be examined. A second crucial indicator of appropriate geometry is the Ramachandran plot [67]. The Ramachandram plot is a plot of the $\varphi$ ($C-N-C_{\alpha}-C$) and $\psi$ ($N-C_{\alpha}-C-N$) torsion angles of the peptide chain for each residue in the model. Each type of regular secondary structure has characteristic peptide chain geometry that corresponds to a particular region of the Ramachandran plot. In addition, steric hindrance restricts the peptide torsions to certain regions of the Ramachandran plot. When peptide torsions are observed outside of these regions, it is a clear indication of a problem in the model. One of the most common problems is a peptide flip in which the carbonyl function of the peptide chain should be pointed roughly in the opposite direction from its orientation in the model.

| | |
|---|---|
| X-ray source | Cu rotating anode |
| X-ray detector | R-AXIS IV |
| Wavelength (Å) | 1.54 |
| Detector distance (mm) | 250 |
| $\Delta\phi$ per image (°) | 1 |
| Space group | $P2_12_12_1$ |
| Unit cell | |
| *a* (Å) | 62.7 |
| *b* (Å) | 64.2 |
| *c* (Å) | 107.7 |
| Total reflections | 231,651 |
| Scaling resolution (Å) | 50–2.87 |
| Unique reflections | 10,463 |
| Overall completeness (%) | 99.9 |
| Last shell completeness (%) | 99.9 |
| Overall $R_{sym}$ | 0.050 |
| Last shell $R_{sym}$ | 0.512 |
| Overall $I/\sigma$ | 22.9 |
| Last shell $I/\sigma$ | 2.4 |

Correcting the error and briefly refining or regularizing the model should bring the torsions back into the allowed regions. It should be noted that one rare but real feature of protein structure called the $\gamma$ turn [68] can place residues into an apparenlty strongly disallowed region of the Ramachandran plot ($\varphi$ = ~75°, $\psi$ = ~–60°). Cis-peptides, almost always prolines, also have a distinctive signature in a Ramachandran plot. Prolines are particularly restricted due to the covalent bond between N and $C_\delta$, while glycine, due to its lack of a side chain, has almost no restrictions on its peptide torsions.

In preparation for computational work particular attention should be paid to the protonation state and hydrogen bonding interactions in the model. The precise side-chain rotamers of histidine, glutamine, asparagine, and arginine residues define their hydrogen possibilities. However, the correct rotamers are not always clear in the electron density maps, and often there is not a great deal of attention paid to the appropriateness of possible hydrogen bonding interactions during refinement. In many cases the correct arrangement is not clear or several possible arrangements seem likely, and a best guess must be made.

## Crystallography–Drug Discovery Interface

Protein crystal structures provide essential information for docking algorithms, ligand-growing algorithms, and for generation of structure-based pharmacophore models. Analysis of protein structure models could also reveal novel or unexpected binding sites (or even accessory pockets to expected binding sites) that might not be exploited by conventional SAR. This information could then be used to guide synthesis of new inhibitors that take full

advantage of the potential for favorable interactions with the target protein. Many algorithms have been developed for active site and potential binding site identification based on electrostatic, hydrogen bonding, and hydrophobic interaction potentials. Other algorithms analyze surface topology for grooves and clefts that could accommodate small-molecule ligands of certain shapes.

X-ray crystallography is the only experimental technique that provides detailed knowledge of the structure of ligand binding sites in proteins. Knowledge of binding site structure opens the door to physical interpretation of structure–activity relationships among inhibitor compounds, and so it is an invaluable aid in drug discovery and design. The greatest value in crystallography comes from crystal structures of actual protein–ligand complexes. These crystal structures reveal the precise binding mode of compounds or even entire classes of compounds as well as revealing consequent conformational changes in the protein. In many cases the ligand-bound conformation of the protein is a much more useful target for computational drug design techniques than is the unbound conformation. Docking and ligand-growing algorithms, scoring functions, and molecular mechanics calculations are most often exquisitely sensitive to small changes in atomic coordinates of the binding site, as are real molecular interactions. Thus unbound conformations, homology models (unless derived from very highly homologous crystal structures), and other theoretical models of active sites are generally of less value in detailed computational modeling because they do not precisely represent the active site structure to which the ligands actually bind. Approximate models of binding sites can be of use in the generation of structure-based pharmacophore models, but the detailed structure of the actual ligand binding site is certainly more appropriate in most cases.

## 6 EXAMPLES FOR THE USE OF X-RAY CRYSTALLOGRAPHY IN DRUG DISCOVERY

Structure-based drug design is an integral tool in lead optimization of drug discovery programs in all major drug companies. In most cases the structure of the target protein with and without small molecules is investigated with X-ray crystallographic techniques, while nuclear magnetic resonance (NMR) techniques are frequently used to determine specific interactions of small molecules with their target molecules. The recent technical improvements seen in many areas of X-ray crystallography allow researchers to determine crystallographic structures of their target protein in complex with a new ligand within hours to days at moderate cost. This capability helps accelerating the work cycle of lead optimization, where new leads are synthesized and precise interactions on the atom level are being determined, to time scales of less than a single week. The work flow of structure-based drug design including target structure based in silico screening and lead optimization is lined out in Figure 22. In recent years fundamentally new applications such as crystallographic

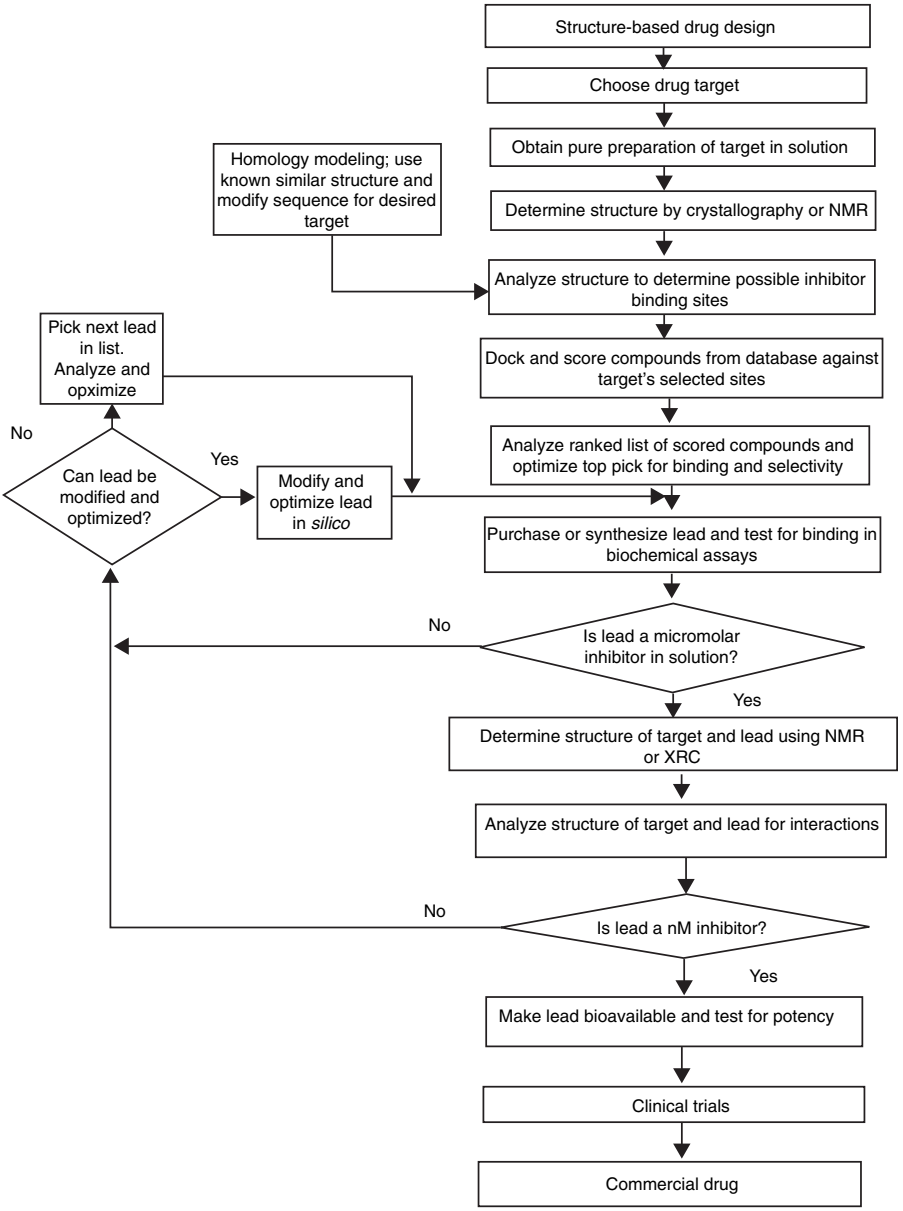


**Figure 22** Schematic work flow of structure-based drug design. (Figure taken from Anderson [4].)

screening have been developed that apply X-ray crystallography as a versatile and instructive tool at different stages of the drug discovery process. Thus, X-ray crystallography can be applied, with different intentions, throughout the entire lead discovery and optimization process. The impact of structure-guided drug design on clinical agents is well documented [69, 70]. The following examples document the use of X-ray crystallography from its "classic" application in lead optimization to various specific applications in drug discovery (Fig. 1).

## Lead Optimization—Structure-Based Drug Design

The process of incrementally changing a compound based on the analysis of its three-dimensional protein–drug complex is generally referred to as *structure-based drug design* (SBDD) [71]. The SBDD process is usually organized in a circular work flow (Fig. 22) and often driven by a collaboration of several groups within an organization that are proficient in several research disciplines. At least two groups are required, one responsible for chemical synthesis and the other for X-ray crystallographic analysis. Often additional researchers in biology, pharmacology, and modeling are involved. They supplement a drug discovery effort with testing of efficacy and toxicity of newly designed and synthesized compounds. The SBDD cycle often starts with a lead compound originating from a natural product source or one that has been identified by screening real or virtual compound libraries. A structure of the protein target in complex with the lead compound initiates the SBDD cycle. Chemical modifications are proposed based on the analysis of this structure, and new compounds are synthesized from such chemical hypotheses. The most important design factors are (a) a perfect complementary geometric fit of the ligand with that of the binding site, (b) low-energy conformations of ligand and protein, (c) complementary electrostatic potentials, (d) the formation of ionic or hydrogen bonds between functional groups, and (e) maximization of the hydrophobic contact area at the ligand–target interface. Since the binding of a small molecule to a protein surface is a complex phenomenon, specific predictions of the effect of certain changes of the ligand molecule are fraught with uncertainty. Generally, affinities increase with an increase of the hydrophobic interface. The free energy contribution of H bonds pays tribute to the balance of solvation energies and that of the formed or broken H bond. Even seemingly small alterations of single-ligand functionalities may have unexpected consequences. Therefore, newly synthesized compounds need to be tested for efficacy and toxicity, and new ligand–target complex structures need to be determined. A thorough analysis of these results stimulates a new round of SBDD. Generally, computer-aided design is lead by the need to optimize or maintain protein binding as well as enhance efficacy and decrease toxicity or other undesirable effects. Eventually new compounds are synthesized

and the SBDD cycle is repeated until an acceptable compound is discovered or the project is canceled.

Arguably the first marketed drug whose generation was significantly impacted by X-ray crystallography is the angiotensin-converting enzyme (ACE) inhibitor Captopril from Bristol Myers Squibb. Curiously though, structures of the actual target apo-ACE or those of ACE in complex with inhibitors were not available. Instead, in the 1970s investigators used the X-ray crystallographic structure of the closely related zinc protease bovine carboxypeptidase A in complex with benzylsuccinate to create a homology model of ACE. Using this homology model and the structure of the inhibitor complex investigators rationalized the synthesis of new drug compounds, eventually creating Captopril (Fig. 23), a highly selective compound with tight binding affinity and low side effects. Captopril was approved by the Food and Drug Administration (FDA) and marketed in 1981. Today, Captopril has undergone the usual lifecycle of approved drugs and is available as a generic.

The first successful drug to reach clinical use that was based on an SBDD development cycle is claimed by Merck Pharmaceuticals. Trusopt (generic dorzolamide) is an inhibitor of carbonic anhydrase II. The discovery of Trusopt was stimulated by X-ray crystallographic structures of carbonic anhydrase II in complex with acetazolamide and other sulfonamides. Rational design originating from the zolamide scaffold assisted in the discovery of inhibitor compounds with activities increased by three orders of magnitude (Fig. 24).

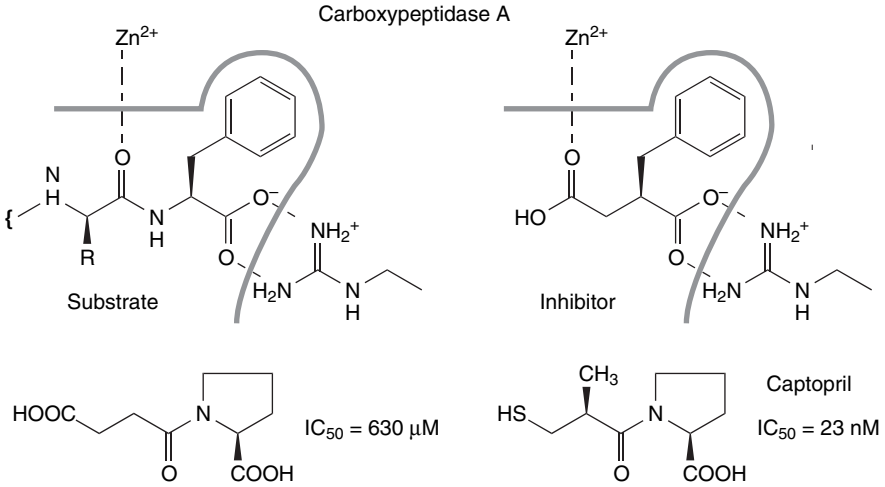


**Figure 23** X-ray crystallographic data of carboxypeptidase A shows that a guanidinium group in an arginine residue and a zinc ion are crucial for complexing the natural substrate (top left) and the inhibitor compound benzylsuccinate (top right). The optimization of the first ACE lead *N*-succinoyl-prolin (bottom left) to Captopril (bottom right) was aided by rationalization of the binding interaction in the homologous enzyme. (Taken with permission from Kubinyi [72]).

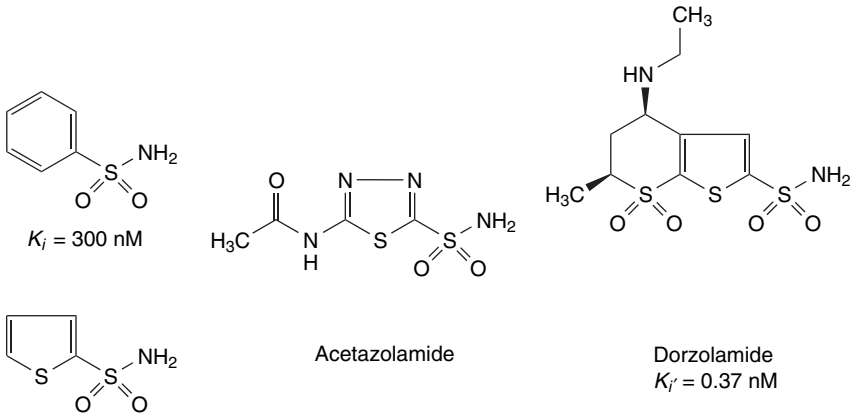


**Figure 24** Carbonic anhydrase II inhibitors. Structures of carbonic anydrase in complex with moderately active compounds led to the discovery of the highly active compound dorzolamide.

***HIV Protease Inhibitors*** The most profound impact that X-ray crystallography has had on current clinical practice and public health was the discovery of the human immune virus (HIV) protease inhibitors. HIV protease is a crucial enzyme important for replication of HIV. The HIV protease processes two of the three gene products of HIV into active enzymes and is required for viral reproduction. Inhibition of the enzyme can prevent HIV replication. Pharmaceutical companies had a strong background in the development of protease inhibitors, as many therapeutic targets are also proteases. The first X-ray crystallographic structure of HIV protease was published in 1989 and, within 8 years four separate compounds from four different companies, each developed using SBDD, were approved by the FDA for HIV treatment. The convergence of several unique circumstances aided this remarkable achievement: The expertise in pharmaceutical companies working on aspartyl proteases coupled with an intense public interest and generous financial support by the U.S. government provided the extraordinary context. Crucially, HIV protease is an enzyme amenable to facile co-crystallization with drug compounds, and the availability of appropriate structural information at a timely point in several drug discovery projects allowed a concentrated SBDD focus.

**Figure 25** Pathway of the discovery of Mozenavir. Dupont's initial HIV protease inhibitors (compound 7) resulted from a computer-based three-dimensional search of known chemical entities using a pharmacophore hypothesis (top left). Scaffold entities 8, 9, and 10 were used as a base compound to add functional groups on the P1 and P2 pseudosymmetric compounds. The lower half of the panel shows the final compound DMP-450 (Mozenavir) with distances in Å to protein atoms of the two molecules of HIV protease in the crystal structure.

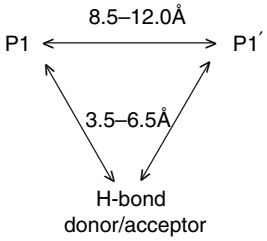


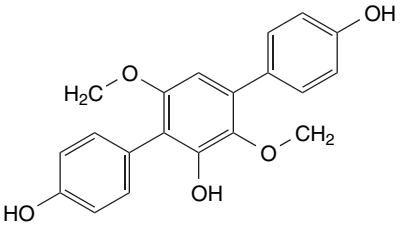


7
(DuPont Merck)

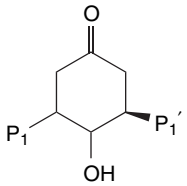


8

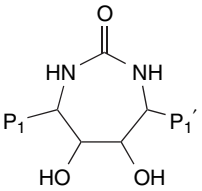


9

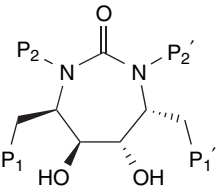


10

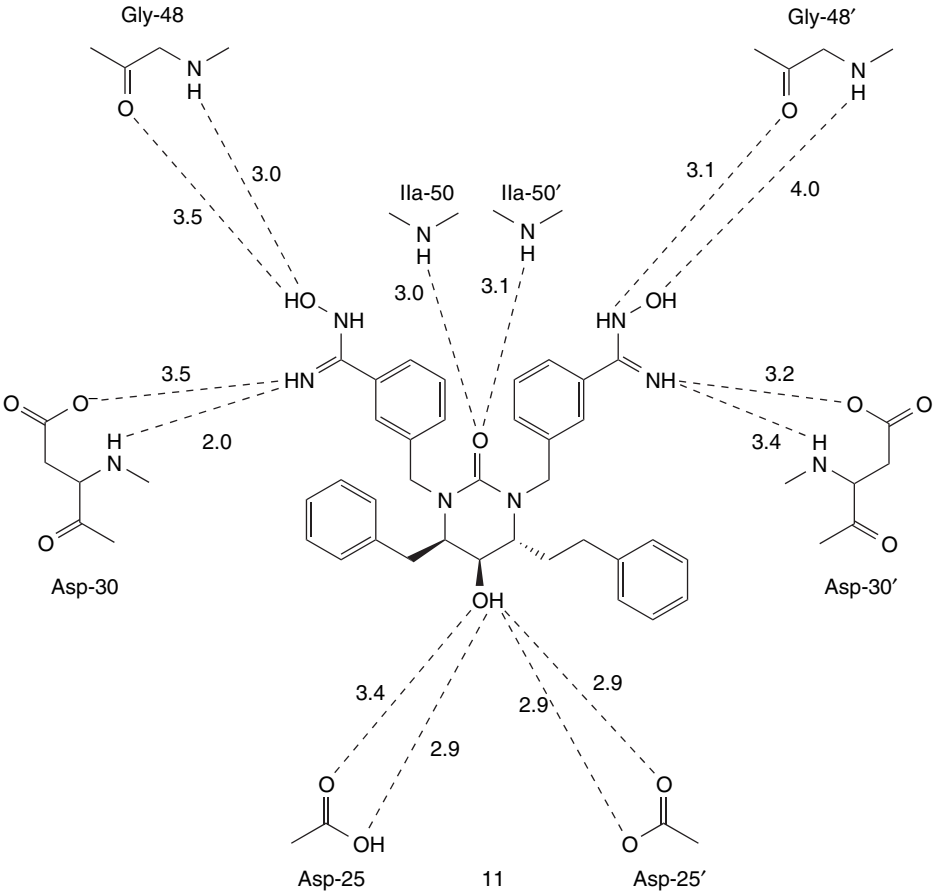


11
(DuPont Merck)
$K_i$ = 0.018 nM

The discovery path of Mozenavir, an HIV protease inhibitor, nicely illustrates the SBDD process. Dupont began its HIV protease investigation using a pharmacophore hypothesis based on the X-ray crystallographic structure of the apo-enzyme HIV protease. This pharmacophore consisted of two lipophilic groups separated by 8.5 to 12 Å and coupled with one hydrogen bond donor/acceptors at a distance of 3.5 to 6.5 Å (Fig. 25). Extensive searches of the Cambridge Structural Database, composed of three-dimensional structures of small-molecule compounds, revealed a potential lead compound. The initial lead was pseudosymmetrical, with the two ends of the molecule being virtual mirror images of each other. This property was significant because it was known that HIV protease functions as an obligatory dimer. From the crystal structure it was known that the interface of the two proteins is composed of identical residues from each protease molecule, suggesting that inhibitor compounds may contain symmetrical moieties. Co-crystal structures with initial lead compounds revealed an important feature of the potential inhibitor binding site. One crucial observation was the realization that the methoxy groups of compound 7 (Fig. 25) could replace an ordered water molecule. Different scaffold compounds were then chosen to add methoxy- as well as other functional groups. Iterative SBDD cycles continued until several molecules were identified that had nanomolar affinities and favorable pharmacokinetic properties. Clinical development of the final compound, DMP450 (Mozenavir) was turned over to Triangle Pharmaceuticals and stopped in 2002 due to Mozenavir's side effects [70].

***Protein Kinases*** Protein kinases are therapeutic targets for a variety of diseases. More than 500 kinases have been identified in the human genome, they are enzymes dependent on adenosine 5′-triphosphate (ATP) that phosphorylate other proteins. Most kinases act in cell-signaling pathways, phosphorylating other signaling proteins whose activity is then either turned on or turned off as a result of the attached negatively charged phosphate group. The catalytic domains of most kinases are structurally conserved; however, their mechanisms of regulation are distinctly different. The catalytic domains are composed of a bi-lobed structure consisting of a helical domain and a β-strand domain (Fig. 26). ATP is bound at the interface of the two domains, and this binding site has been deemed an attractive site for designing ATP competitive inhibitors that block ATP binding, thus preventing phosphorylation. Unfortunately, the ATP binding site is highly conserved in most kinases, thus rational design faces serious challenges in creating inhibitors selective for a specific kinase with little inhibition of other closely related kinases. SBDD of kinase inhibitors is popular because of the low experimental barriers. Conserved and unique residues can be identified in the inhibitor binding site and rational design attempts to maximize interactions with unique residues in the inhibitor binding site while minimizing interactions with conserved residues. The underlying hypothesis is that specificity can be created as interactions with the inhibitor are increased with residues specific for the kinase in question while cross-reactivity decreases for other kinases.

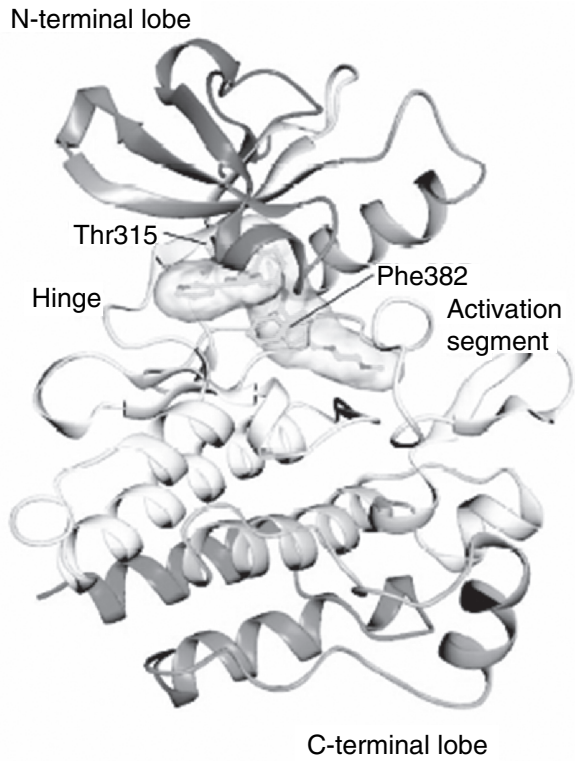


**Figure 26** Overview of the structure of a kinase with an inhibitor bound. Kinases offer the possibility to design inhibitors based on stabilization of inactive conformations. (This figure is available in full color at ftp://ftp.wiley.com/public/sci_tech_med/drug_discovery/.)

## Antistructures

As discussed, the main applications of X-ray crystallography in drug discovery and optimization projects are based on the analysis of the target and the interactions of targets with their ligands. The design efforts are aimed at strengthening the resulting complexes. The opposite approach, however, weakening the interaction of drug leads with some proteins may be used to one's advantage. The interacting proteins are not the actual targets but those that cause detrimental effects in drug efficacy. Such proteins are sometimes called *antitargets* and may be related enzymes with similar substrate binding pockets but with very different function, such as kinases or phosphatases.

Generally, weak binding of small-molecule drugs to serum albumin and detoxification proteins such as cytochrome P450 is a desired property. P450 enzymes are involved in the oxidative metabolization of most drugs and are often the source of drug-related side effects or their toxicity. Several P450 structures are available [73] and may be used for in silico docking studies and the published crystallization methods may be used to grow crystals for soaking or co-crystallization studies. The goal of such projects is to increase lead effi-

cacy by defining lead modifications that weaken the interaction with anti-target proteins while maintaining the affinity to the target protein. It would be highly desirable to extend further this negative template design, that is, by including membrane-bound drug transporters, to modulate, based on molecular structural insight, the critical processes of absorption, distribution, metabolism, excretion, and toxicity. These possible applications clearly demonstrate that X-ray crystallography is headed toward effecting later stage discovery and early stage drug development projects.

Human serum albumins consist of three domains with six rather promiscuous ligand binding sites. Many drug leads bind to this protein causing a serious problem for lead discovery. In an instructive example, the feasibility of its "design away" approach was demonstrated for diflunisal, a nonsteroidal anti-inflammatory cyclooxygenase inhibitor. Note that 99 percent of diflunisal in serum is unavailable to the target due to binding to human serum albumin. This requires high doses of up to 250 mg diflunisal to be administered, causing gastrointestinal irritation as a serious side effect. In a structure-based drug design effort diflunisal analogs were synthesized that were deemed to bind less efficiently to HSA-III (a human serum albumin subdomain). Several compounds were generated that exhibited more than 100-fold reduced binding to HSA-III (with only 10-fold reduction in affinity for full-length albumin). Significantly, several of these compounds maintained their activity against the actual target, cyclooxygenase-2 [74].

## Protein Therapeutics

Protein therapeutics such as EPO and insulin are the hallmark of this fast-growing class of drugs. Hardly any new optimization program excludes the use of X-ray crystallography. Based on natural products, protein therapeutics are directly amenable to X-ray crystallographic investigation and subsequent redesign. Classes of protein or peptide therapeutics are monoclonal antibodies, cytokines, enzymes, and viral fusion inhibitors. In a landmark study Ewert et al. [75] used X-ray crystallographic structure-based antibody engineering to aid in the identification of residues that improve unsatisfactory antibody properties [75]. In a different case, an HIV entry inhibitor was designed based on CD4. The 27-amino-acid CD4 mimic interacted with gp120 and is bound to HIV particles with CD4-like affinity. This mini-CD4 is a prototype HIV-1 inhibitor and a potential component for vaccine formulations [76].

## In silico Screening Based on Crystallographic Structural Models

X-ray crystallographic structures of proteins may be used to preselect a small number of compounds from compound libraries. Different types of computer-based algorithms such as docking [77] may be employed to predict the formation of ligand–protein complexes with single compounds or with entire compound libraries ("virtual screening" or "in silico screening"). This computational approach is particularly advantageous for targets where target struc-

tural information is readily available at no cost (i.e., in the public domain). A side-by-side comparison of assay-based high-throughput screening (HTS) with such virtual screening on the same target protein has been described by Doman et al. [5]. Numerous structures of PTP1B (tyrosine phosphatase 1 B), a diabetes type II drug target, are available from the Protein Data Bank. Molecular docking of ca. 235,000 compounds into the closed, ligand-removed structure of PTP1B yielded 365 high-scoring candidates that were tested for inhibition in enzyme-based assays. Of these candidates 34.8 percent inhibited PTP1B with $IC_{50}$ values below 100 μM, representing an enrichment of 1700-fold (as compared to random screening). Conventional high-throughput screening on the other hand, yielded only 85 hits with $IC_{50}$ values below 100 μM out of 400,000 tested molecules (corresponding to a 0.021 percent hit rate). Interestingly, the hit lists were rather different and the hits generated by molecular docking appeared more druglike than the HTS hits, suggesting that the two screening techniques may be used in a complementary way. Indeed, the integration of virtual and high-throughput screening is judged to be a promising approach in modern lead discovery projects [78].

## Crystallographic Screening

X-ray crystallographic screening is a modern combination of lead identification, immediate X-ray crystallographic structural evaluation, and subsequent lead optimization on crystals that are soaked with ligand mixtures. Ingeniously, the binding capability of some proteins in crystals is employed to "fish" and present tight binders. Those ligands with the highest affinity are identified and selected for further rounds of optimiziation [79]. Compared to conventional high-throughput screening, crystallographic screening (a) yields hits with activities in the mM to 30 μM range, (b) yields hits with evidently defined binding interactions, and (c) involves only minimal hit-to-lead synthetic chemistry efforts because follow-up libraries can be focused using detailed information from the crystal structure. Nienaber et al. [1] demonstrated that crystallographic screening can be performed in a rapid, efficient, and high-throughput fashion. They established the utility of the iterative process by discovering 8-aminopyrimidyl-2-aminoquinoline, a new class of anticancer urokinase inhibitors (Fig. 27). Initially, 61 compounds were divided into 9 separate mixtures with 6 to 8 compounds each. Care was taken to distribute into a particular cocktail those compounds that had the greatest degree of structural diversity in order to facilitate subsequent ligand identification based on the shape of $F_o$-$F_c$ electron density maps. Nine urokinase crystals were soaked with individual cocktails, X-ray diffraction data was generated, and the resulting electron density maps were inspected. They showed the shape and orientation of compounds (Fig. 27 *a*). In one case two binders were present. Here, the removal of the prominent compound and resoaking allowed to identify the electron density of the second ligand. In the subsequent lead optimization step previous structure–activity relation ship (SAR) data was

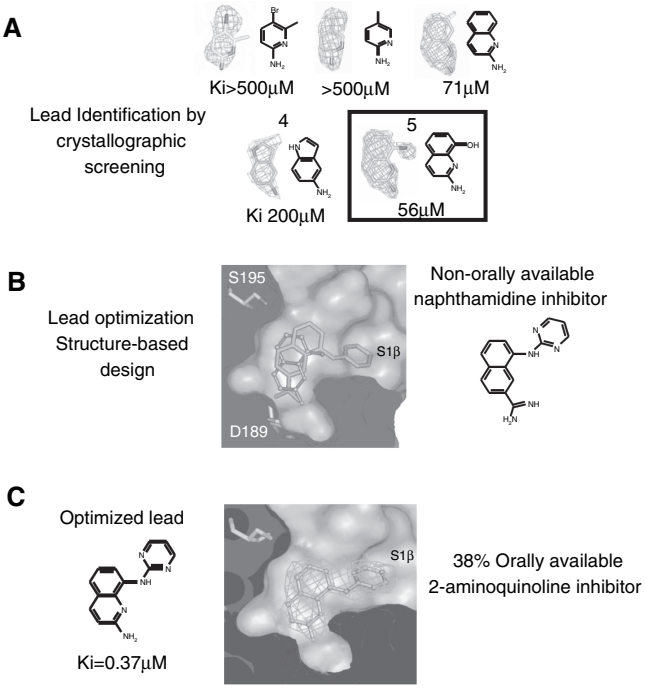


**Figure 27** Urokinase lead identification via crystallographic screening and optimization [1]. (*a*) Initial $F_o$–$F_c$ electron density maps for ligands that were identified from compound cocktail-soaked urokinase crystals. (*b*) Crystal structures of 8-aminopyrimidyl-2-naphtamidine (orange) and a 2-aminoquinoline lead (blue). (*c*) Structure and $2F_o$–$F_c$ electron density map for the optimized lead compound 8-aminopyrimidyl-2-aminoquinoline. (This figure is available in full color at ftp://ftp.wiley.com/public/sci_tech_med/drug_discovery/.)

included and lead to the development of 8-aminopyrimidyl-2-aminoquinoline, a ligand with a ca. 100-fold increased inhibitor potency ($K_i$ = 0.37 μM) and a 38 percent oral availability, as determined by in vivo pharmacokinetic tests. This type of process is capable of identifying weaker binding ligands (1 mM) and is applicable where apo-crystals are available and tolerate soaking. Crystallographic screening may also be used to facilitate the validation of new targets, the development of assays and assist in assigning biochemical function to orphan targets.

## Crystallographic Fragment Screening

A variation of this theme is crystallographic fragment screening. Here crystals are soaked with cocktails that contain small druglike fragments rather than complete leadlike compounds. Once several fragments are identified crystal-

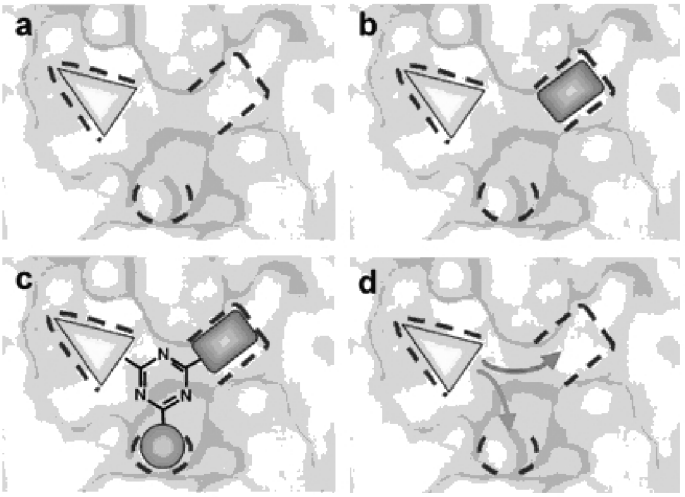


**Figure 28** Schematic crystallographic fragment screening. Once fragments are identified (*a*, *b*) they can be joined (*c*) resulting in a leadlike compound or fragments may be developed along the lines of conventional structure-based drug design. (Figure taken from Blundell et al., 2002 [80].) (This figure is available in full color at ftp://ftp.wiley.com/public/sci_tech_med/drug_discovery/.)

lographically, they can be developed into new lead compounds (Fig. 28). Curiously, low-affinity small fragments that bind adjacent binding pockets can be joined and result in a larger molecule with increased affinity. Typically fragment libraries consisting of only a few hundred to a thousand compounds are screened. Crystallographic fragment screening is a new and promising technology employed by several biotechnology companies; however, specific examples for drug discovery have not yet been published in the scientific literature.

Rees et al.[81] discuss 25 examples for the successful application of the fragment-based lead discovery approach, some of them aided by crystallographic screening. They also formulate a "Rule of three" in which the average fragment is characterized as (a) having a mass of less than 300 Da, (b) having less than or equal to 3 hydrogen bond donors, (c) having less than or equal to 3 hydrogen bond acceptors, and (d) having a $c \log P$ of 3. In addition, the number of rotatable bonds was on average less than or equal to 3 and the polar surface area was about 60 $Å^2$.

## Site-Directed Leads via Fragment Tethering

An additional layer of complexity is added by generating site-directed leads via fragment tethering ([2]; Fig. 29). In a first step target proteins are covalently modified at a particular site on the surface. Mass-spectrometric detection allows the identification of weakly binding ligand precursors. In a second

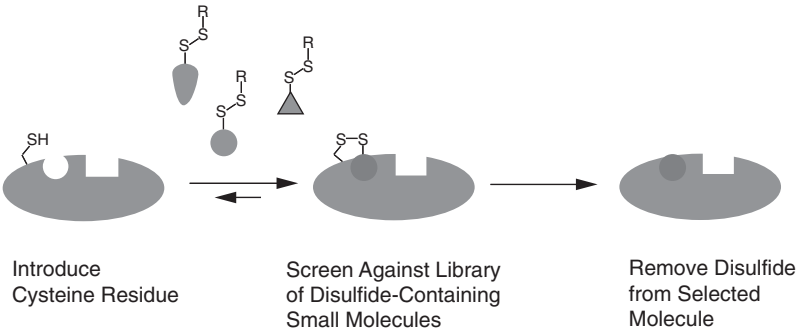


**Figure 29** Schematic illustration of the fragment tethering approach.

step X-ray crystallography provides the tool to observe the interaction of the precursor with the protein and helps directing the chemical synthesis of fused analogs with potentiated affinity. This strategy was used to generate a potent inhibitor for the anticancer target thymidylate synthase [3]. Although thymidilate synthase contains an active site cystein, this reactive group can be introduced by surface mutatenesis (native amino acid to Cys). A library of 1200 disulfide-containing compounds was screened in pools of up to 100 compounds. Several disulfide adducts were detected by mass spectrometry. One of the selected compounds was investigated further: The inhibition constant $K_i$ of the tether-free analog *N*-tosyl-D-proline was 1.1 mM. Subsequent crystallographic structure determination of the tymidilate synthase adduct and lead optimization improved the affinity over 3000-fold. This approach has been refined to extended tethereing where the first identified tethered compound serves as an anchor for the next fragment [82]. The tethering approach enables the nucleation of drug design efforts at specific sites on protein surfaces.

## Structural Genomics

Current genome-wide *structural genomics* programs aim to determine representative structures for all proteins. The goal is to infer biological function from similar structures of known function [83]. Once representative structures are obtained, structural homology models of all members of each protein family can be built using these templates. This information is useful in selecting and classifying drug targets because the crystallographic structure holds information regarding protein function. The assignment of function by comparison on DNA (deoxyribonucleic acid) level fails for some proteins that have different sequence but similar fold. Furthermore, structures of proteins with bound natural ligands may expose their mechanism of action and allow researchers to act on this insight. In a proof-of-principle experiment

Zarembinski et al. [84] demonstrated that the function of a gene can indeed be assigned by X-ray crystallographic structure determination. They reported the 1.7-Å resolution structure of MJ0577, a protein form the hyperthermophile *Methanococcus jannaschii*. Unexpectedly, a bound ATP molecule was identified from the electron density map. Therefore, the protein was deduced to work as an ATPase or an ATP-mediated molecular switch. The latter function was confirmed by subsequent biochemical experiments. Similarly, a thymidilate synthase complementing protein was discovered from a structural genomics project by structure-based functional analysis, leading to its classification as an antibacterial drug target [85].

At first sight, this approach may seem to be a digression, but it addresses a grave problem in protein X-ray crystallography on new targets: There is no guarantee for obtaining a structural result. Current success rates for generating structures of novel protein targets are below 50 percent. However, once the scope of targets is increased, for example, including homologous proteins from various organisms, the probability of successful generation of X-ray crystallographic structures increases dramatically.

The number of potential drug targets has increased from about 500 in the mid-1990s to several thousand possible targets these days. It is the goal of structural genomics programs to provide structures of these new drug targets, even those of not yet identified targets or those proteins that may be used in the future for expedient homology model-building purposes once homologous proteins have been identified as useful drug targets. The formation of the Structural Genomics Consortium and its financial support by Glaxo-Smith-Kline pays tribute to the value of this approach. That consortium's goal is to determine 350 X-ray crystallographic structures of proteins directly related to human health, including proteins associated with cancer, neurological, and infectious diseases, within 3 years, starting from 2003.

The abundance of protein structural information has fueled large-scale bioinformatics approaches, aiding the drug discovery process at many stages. One particular application could be the improvement of the final drug product profile by systematic biasing drug lead optimization away from all known structures by treating their entirety as *antitargets*.

## 7 LIMITATIONS AND CHALLENGES OF X-RAY CRYSTALLOGRAPHY IN THE DRUG DISCOVERY PROCESS

Despite the outstanding track record of X-ray crystallography as an excellent tool for drug discovery purpose, its use has several limitations. The most serious ones result from (a) problems arising from protein flexibility, (b) shortcomings in computational methods to use structural information for the proper quantification of the energetics of protein–ligand interactions, and (c) difficulties in applying the crystallographic method to challenging targets such as membrane proteins.

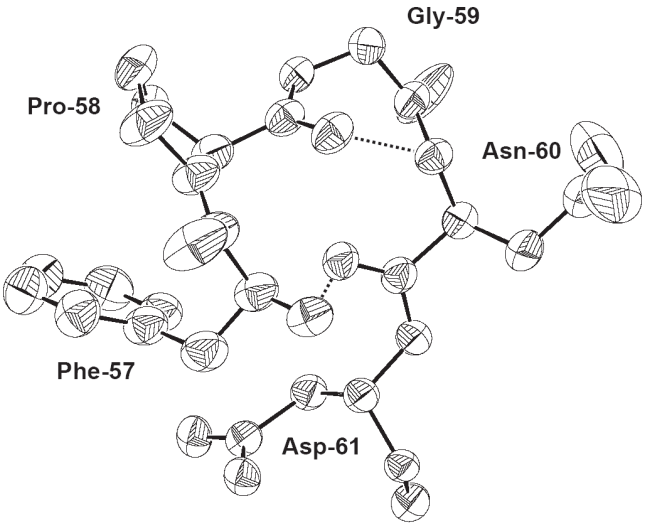


**Figure 30** Elipsoidal representation of atom positions in a loop region in a serine protease depicting, for example, the high flexibility of the amide group of Asn-60 with respect to backbone atoms. Anisotropic displacement parameters can be calculated from atomic resolution data giving an impression of the amplitude and the direction of the mean displacement. (Figure taken from Schneider [86].)

Proteins are not as rigid as their visual representations in models suggest. Both proteins and small molecules are indeed flexible entities with thermal fluctuations, including bond vibrations in the picosecond range to slower nanosecond range side-chain rotations to very slow conformational changes that take place in time frames of micro to milliseconds. The amplitudes of the corresponding positional fluctuations are significantly large to affect ligand binding. Conversely, X-ray crystallographic structure describes protein structures as static, based on an apparent ensemble average at 100 K. The positional variations of atoms due to flexibility and crystal packing effects in standard resolution structures (1.5-to 3-Å resolution) is usually described by a single parameter, the B factor, the latter of which being a weak predictor of structural flexibility. Only few ultra-high-resolution structures of small proteins have allowed resolving alternative side-chain occupancies and anisotropic B-factor refinement [86] (Fig. 30).

The current standard, using a single model to describe instrinsically heterogenous proteins, is an oversimplification that may lead to inaccurate models (Fig. 31). This shortcoming may be resolved by generating and refining an ensemble for alternative structures [87]. For interleukin 1β such models differed by as much as 1 Å between models.

There are numerous examples that demonstrate that the protein structure of a ligand bound form differs significantly from that of the ligand-free form or to that bound to a different ligand [88]. Apparently the ligand selects and

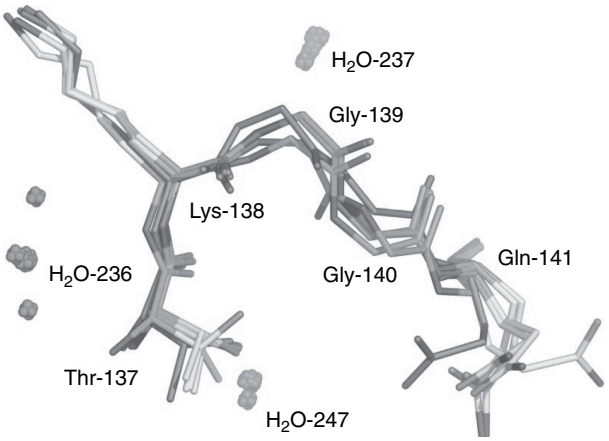


**Figure 31** Structural heterogeneity in human interleukin1β. The ensemble of models displays considerable backbone variability, disordered side chains, and multiple locations of water molecules. The models were obtained from the same set of 2.3 Å resolution X-ray diffraction data, and refined to similar levels. (Image taken from DePristo et al. [87].) (This figure is available in full color at ftp://ftp.wiley.com/public/sci_tech_med/drug_discovery/.)

stabilizes a particular conformational state of the pool of many low-energy states that proteins can exist in at thermal equilibrium. In the process of SBDD this causes a high degree of unpredictability, making the method less useful. However, some proteins are less flexible, and their conformation hardly changes when ligands bind. These are the targets that are particularly susceptible to conventional SBDD efforts. There are only a few solutions to this fundamental predicament [89], notably the computationally intensive approach to treat the protein as a flexible entity and the tethering discovery approach. Understanding this limitation, however, may serve as the best antidote against the overuse of this tool.

Sanders et al. [90] point out a serious shortcoming of crystallographic screening. They described the discovery of competitive inhibitors for dihydroneopterin aldolase via crystallographic screening and demonstrated that several compounds with $IC_{50}$ around 1 μM were negative in crystal soaking experiments. Apparently the conformational shift associated with the binding of these missed compounds did not allow association to the protein in the preformed crystal.

The deficiencies of current computational methods to properly quantify the interactions of proteins with ligands is one of the consequences of molecular flexibility. But even more fundamentally, our current understanding of the energetics of ligand–protein interaction and hence their proper quantification by scoring functions is limited [91]. A weak point remains, for instance, in the description of entropic terms for binding interactions, although progress is being made and an energetic penalty of 10 to 30 kcal/mol is estimated for protein reorganization due to binding [92].

Finally, due to major technical bottlenecks X-ray crystallography cannot be applied to all drug discovery programs. The reason for this is the high uncertainty involved in obtaining any useful results with a given allocation of resources. The uncertainty is caused by the many risk-fraught steps involved in the crystallographic endeavor. Modern high-throughput crystallographic technologies seek to reduce this risk and have in some cases provided de novo structural information within less than 2 months at reasonable cost. This is not a given, though, since the outcome of a particular X-ray crystallographic project is unpredictable. This weakness reduces the applicability of X-ray crystallography to about half of all discovery projects in major pharmaceutical companies.

At the time of this writing, almost 30,000 protein structures were available from the PDB, an impressive record of the methodology's performance and significance. However, many structures for the presently ca. 500 proteins targeted by current drugs are not available because they have not been crystallized and their crystallographic structures have not been determined. This is particularly bothersome for membrane proteins, the latter of which represent half of the proteins that are targeted by today's marketed drugs [19]. Some 45 percent of molecular targets of known drugs are G-protein-coupled receptors (GPCR), proteins that have been notoriously difficult to deal with in the stages of overexpression, purification, and crystallization. Indeed, to date only one structure of such a protein has been determined, that of bovine rhodopsin, and several structures of similar bacterial homologs with seven transmembrane helices. It is anticipated that committed efforts in crystallographic structure determination projects can surmount the experimental barriers and yield structural information of this highly important protein class. The current state of SBDD on such valuable GPCR targets is reminiscent of the situation in the early 1970s when homology models based on carboxypeptidase A were successfully applied in the "design" of ACE inhibitors. Indeed, the first "virtually" discovered compound targeting a predicted structure of the serotonin receptor (a GPCR) entered phase 1 trials in 2004 and, Abbott's drug Atrasentan, targeting the endothelin-A receptor was optimized with the help of a homology model of the endothelin-A receptor (also a GPCR; [93]).

***Acknowledgments*** We are greatly indebted to Dr. Wendy Sanderson and Dr. Ehmke Pohl for thoroughly reading and revising parts of this chapter.

## REFERENCES

1. Nienaber, V. L., Richardson, P. L., Klighofer, V., Bouska, J. J., Giranda, V. L., Greer, J. (2000). Discovering novel ligands for macromolecules using X-ray crystallographic screening. *Nat. Biotechnol., 18*, 1105–1108.
2. Erlanson, D., Wells, J., Braisted, A. (2004). Tethering: Fragment-based drug discovery. *Annu. Rev. Biophys. Biomol. Struct., 33*, 199–223.

3. Erlanson, D. A., Braisted, A. C., Raphael, D. R., Randal, M., Stroud, R. M., Gordon, E. M., Wells, J. A. (2000). Site-directed ligand discovery. *PNAS*, *19*, 9367–9372.
4. Anderson, A. C. (2003). The process of structure-based drug discovery. *Chem. Biol.*, *10*, 787–797.
5. Doman, T. N., McGovern, S. L., Witherbee, B. J., Kasten, T. P., Kurumbail, R., Stallings, W. C., Conolly, D. T., Shoichet, B. K. (2002). Molecular docking and high-throughput screening for novel inhibitors of protein tyrosine phosphatase-1B. *J. Med. Chem.*, *45*, 2213–2221.
6. Bergfors, T. M. (1999). *Protein Crystallization, Techniques, Strategies and Tips; a Lab Manual*. International University Line, La Jolla, CA.
7. McPherson, A. (1989). *Preparation and Analysis of Protein Crystals*. Krieger Publishing.
8. McPherson, A. (1998). *Crystallization of Biological Macromolecules*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
9. Ducruix, A., Giegé, G. (1992). *Crystallization of Nucleic Acids and Proteins, a Practical Approach*. IRL Press.
10. Weber, P. C. (1991). Physical principles of protein crystallizaiton. *Adv. Protein Chem., 41*, 1–36.
11. Durbin, S. D., Feher, G. (1996). Protein crystallization. *Annu. Rev. Phys. Chem.*, *47*, 171–204.
12. McPherson, A. (1990). Current approaches to macromolecular crystallization. *Eur. J. Biochem.*, *189*, 1–23.
13. Chernov, A. A. (2003). Protein crystals and their growth. *J. Struct. Biol.*, *142*, 3–21.
14. Drenth, J., Haas, C. (1992). Protein crystals and their stability. *J. Cryst. Growth*, *122*, 107–109.
15. Chayen, N. E. (1998). Comparative studies of protein crystallizaiton by vapour-diffusion and microbatch technique. *Acta Crystallogr.*, D*54*, 8–15.
16. Chayen, N. E., Shaw Stewart, P. D., Blow, D. M. (1992). Microbatch crystallization under oil—A new technique allowing many small-volume crystallization trials. *J. Cryst. Growth*, *122*, 176–180.
17. Chayen N. E. (1997). The role of oil in macromolecular crystallization. *Structure*, *5*, 1269–1274.
18. Zeppezauer, M. (1971). In W. B. Jakoby (Ed.), M*ethods in Enzymology*, Vol 22. Academic, New York and London.
19. Drews, J. (2000). Drug discovery: A historical perspective. *Science*, *287*, 1960–1964.
20. Iwata, S., (Ed.) (2003). *Methods and Results in Crystallization of Membrane Proteins*. International University Line, Biotechnology Series.
21. Hunte, C. C., von Jagow, G., Schagger, H (2003). *Membrane Protein Purification and Crystallization: A Practical Guide.* Academic, San Diego.
22. Nollert, P., Navarro, J., Landau, E. M. (2001). Crystallization of membrane proteins *in cubo*. *Methods Enzymol.*, *343*, 183–199.
23. Terpe, K. (2003). Overview of tag protein fusions: From molecular and biochemical fundamentals to commercial systems. *Appl. Microbiol. Biotechnol.*, *60*, 523–533.
24. Smyth, D. R., Mroziewicz, M. K., McGrath, W. J., Listwan, P., Kobe, B. (2003). Crystal structures of fusion proteins with large-affinity tags. *Protein Sci.*, *12*, 1313–1322.

25. Deng, J., Davies, D. R., Wisedchaisri, G., Wu, M., Hol, W. G. J., Mehlin, C. (2004). An improved protocol for rapid freezing of protein samples for long-term storage. *Acta Cryslallogr. D60*, 203–204.
26. Page, R., Grzechnik, S. K., Canaves, J., Spraggon, G., Kreusch, A., Kuhn, P., Stevens, R. C., Lesley, S. (2003). Shotgun crystallization strategy for structural genomics: An optimized two-tiered crystallization screen against the *Thermatoga maritima* proteome. *Acta Crystallogr. D*, *D59*, 1028–1037.
27. Jancarik, J., Kim, S.-H. (1991). Spare matrix sampling: A screening method for crystallization of proteins. *J. Appl. Crystallogr.*, *24*, 409–411.
28. Majeed, S., Ofek, G., Belachew, A., Huang, C., Zhou, T., Kwong, P. D. (2003). Enhancing protein crystallization through precipitant synergy. *Structure*, *11*, 1–20.
29. Segelke, B. W. (2001). Efficiency analysis of sampling protocols used in protein crystallization screening. *J. Cryst. Growth*, *232*, 553–562.
30. Gilliland, G. L., Tung, M., Ladner, J. (1996). The Biological Macromolecule Crystallization Database and NASA Protein Crystal Growth Archive. *J. Res. Natl. Inst. Stand. Technol.*, *101*, 309–320.
31. Garman, E. F., Schneider, T. R. (1997). Macromolecular cryocrystallography. *J. Appl. Crystallogr.*, *30*, 211–237.
32. Parkin, S., Hope, H. (1998). Macromolecular cryocrystallography: Cooling, mounting, storage and transportation of crystals. *J. Appl. Crystallogr.*, *31*, 945–953.
33. Rubinson, K. A., Ladner, J. E., Tordova, M., Gilliland, G. L. (2000). Cryosalts: Suppression of ice formation in macromolecular crystallography. *Acta Crystallogr.*, *D56*, 996–1001.
34. Riboldi-Tunncliffe, A., Hilgenfeld, R. (1999). Cryocrystallography with oil—An old idea revived. *J. Appl. Crystallogr.*, *32*, 1003–1005.
35. Carter, C. W., Jr., Doublie, S., Coleman, D. E. (1994). Quantitative analysis of crystal growth. Tryptophanyl-tRNA synthetase crystal polymorphism and its relationship to catalysis. *J. Mol. Biol.*, *238*(3), 346–365.
36. Stewart, L., Clark, R., Behnke, C. (2002). High-thoughput crystallization and structure determination in drug discovery. *Drug Discov. Today*, *7*(3), 187–196.
37. Petsko, G. A. (1985). Preparation of isomorphous heavy-atom derivatives, *Methods Enzymol.*, *114*, 147–156.
38. Rould, M. (1997). Screening for heavy-atom derivatives and obtaining accurate isomorphous differences. Methods Enzymol., *276*, 112–147.
39. Gilliland, G. L. (1988). A biological macromolecule crystallization database: A basis for a crystallization strategy. *J. Cryst. Growth*, *90*, 51–59.
40. HyperPhyscis web site: http://hyperphysics.phy-astr.gsu.edu/hbase/phyopt/mulslid.html#c4.
41. Glusker, J. P., Trueblood, K. N. (1972). *Crystal Structure Analysis: A Primer*. Oxford University Press, New York.
42. Stout, G. H., Jensen, L. H. (1989). *X-ray Structure Determination: A Practical Guide*. Wiley, New York, p. 202.
43. Otwinowski, Z., Minor, W. (1997). Processing of oscillation data collected in oscillation mode. *Methods Enzymol.*, *276*, 307–326.

44. Diedrichs, K., Karplus, K. A. (2002). Improved R-factors for diffraction in macromolecular crystallography. *Nat. Struct. Biol.*, *4*, 269–275.
45. Gewirth, D. (2003). *The HKL Manual*. pp. 89–90.
46. Fuji web site: http://home.fujifilm.com/products/science/ip/principle.html.
47. ADSC web site: http://www.adsc-xray.com/products.html.
48. Walsh, M. A., Dementieva, I., Evans, G., Sanishvili, R., Joachimiak, A. (1999). Taking MAD to the extreme: Ultrafast protein structure determination. *Acta Crystallogra.*, *D55*, 1168–1173.
49. Cohen, A. E., Ellis, P. J., Miller, M. D., Deacon, A. M., Phizackerley, R. P. (2002). An automated system to mount cryo-cooled protein crystals on a synchrotron beamline, using compact sample cassettes and a small-scale robot. *J. Appl. Crystallogr.*, *35*, 720–726.
50. Pohl, E., Ristau, U., Gehrmann, T., Jahn, D., Robrahn, B., Malthan, D., Dobler, H., Hermes, C. (2004). Automation of the EMBL Hamburg protein crystallography beamline BW7B. *J. Synchrotron Radiation*, *11*, 1–6.
51. ALS web site: http://www-als.lbl.gov/als/synchrotron_sources.html.
52. Drenth, J. (1994). *Principles of Protein X-ray Crystallography*. Springer-Verlag, New York, pp. 282–283.
53. Dauter, Z., Dauter, M., Dodson, E. (2002). Jolly SAD. *Acta Crystallogr.*, *D58*, 494–506.
54. McRee, D. (1993). *Practical Protein Crystallography*. Academic, San Diego, p. 145.
55. Hendrickson, W. A (1991). Determination of macromolecular structures from anomalous diffraction of synchrotron radiation. *Science*, *254*, 51–58.
56. Wang, B. C. (1985). Resolution of phase ambiguity in macromolecular crystallography. *Methods Enzymol.*, *115*, 90–112.
57. Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., Warren, G. L. (1998). Crystallography and NMR System: A new software suite for macromolecular structure determination. *Acta Crystallogr.*, *D54*, 905–921.
58. Murshudov, G. N., Vagin, A. A., Dodson, E. J. (1997). Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr*, *D53*, 240–255.
59. Altschul, S. F., Gish, W., Miller, W., Myers, E. W., Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.*, *215*, 403–410.
60. Cuff, J. A., Clamp, M. E., Siddiqui, A. S., Finlay, M, Barton, G. J. (1998). Jpred: A consensus secondary structure prediction server. *Bioinformatics*, *14*, 892–893.
61. Murzin, A. G., Brenner, S. E., Hubbard, T., Chothia, C. (1995). SCOP: A structural classification of proteins database for the investigation of sequences and structures. *J. Mol. Biol.*, *247*, 536–540.
62. Orengo, C. A., Michie, A. D., Jones, S., Jones, D. T., Swindells, M. B., Thornton, J. M. (1997). CATH: A hierarchic classification of protein domain structures. *Structure*, *5*(8), 1093–1108.
63. Lewis, P. N., Momany, F. A., Scheraga, H. A. (1973). Chain reversals in proteins. *Biochim. Biophys. Acta*, *303*, 211–229.

64. McPhail, A. T., Sim, G. A. (1968). The structure of camptothecin: X-ray analysis of camptothecin iodoacetate. *J. Chem. Soc. (B)*, *384*, 923–928.
65. Winn, M. D., Isupov, M. N., Murshudov, G. N. (2001). Use of TLS parameters to model anisotropic displacements in macromolecular refinement. *Acta Crystallogr.*, *D57*, 122–133.
66. Moews, P. C., Kretsinger, R. H. (1975). Refinement of the structure of carp muscle calcium-binding parvalbumin by model building and difference Fourier analysis. *J. Mol. Biol.*, *91*, 201–228.
67. Ramachandran, G. N., Ramakrishnan, C., Sasisekharan, V. (1963). Stereochemistry of polypeptide chain configurations. *J. Mol. Biol.*, *7*, 95–99.
68. Printz, M. P., Nemethy, G. (1972). The γ turn, a possible folded conformation of the peptide chain. Comparison with the β turn. *Macromolecules*, *5*, 755–758.
69. Hardy, L. W., Malikayil, A. (2003). The impact of structure-guided drug design on clinical agents. *Curr. Drug Discov.*, 15–20.
70. Klebe, G. (2000). Recent developments in structure-based drug design. *J. Mol. Med.*, *78*, 269–281.
71. Veerapandian, P. (Ed.) (1997). *Structure-Based Drug Design.* Marcel Dekker, New York.
72. Kubinyi, H. (1999). *J. Receptor Signal Transduction Res.*, *19*, 15–39.
73. Williams, P. A., Cosme, J., Vinkovic, D. M., Ward, A., Angove, H. C., Day, P. J., Vonrhein, C., Tickle, I. J., Jhoi, H. (2004). Crystal structures of human cytochrome P450 3A4 bound to metyrapone and progesterone. *Science*, *305*, 683–686.
74. Mao, H., Hajduk, P. J., Craig, R., Bell, R., Bore, T., Fesik, S. W. (2001). Rational design of diflunisal analogues with reduced affinity for human serum albumin. *J. Am. Chem. Soc.*, *123*, 10429–10435.
75. Ewert, S., Honegger, A., Plückthun, A. (2003). Structure-based improvement of the biophysical properties of immunoglobulin VH domains with a generalizable approach. *Biochemistry*, *42*, 1517–1528.
76. Stricher, M. L. F., Misse, D., Sironi, F., Pugniere, M., Barthe, I., Magne, X., Roumestand, C., Menez, A., Lusso, P. (2003). Rational design of a CD4 mimic that inhibits HIV-1 entry neutralization epitopes. *Nat. Biotechnol.*, *21*, 71–76.
77. Taylor, R. D., Jewsbury, P. J., Essex, J. W. (2002). A review of protein-small molecule docking methods. *J. Computer-Aided Mol. Des.*, *16*, 151–166.
78. Bajorath, J. (2002). Integration of virtual and high-throughput screening. *Nat. Rev. Drug Discov.*, *1*, 882–894.
79. Carr, R., Jothi, H., et al. (2002). Structure-based screening of low affinity compounds. *Drug Discov. Today*, *7*, 522–527.
80. Blundell, T. L., Jhoti, H., Abell, C. (2002). High-throughput crystallography for lead discovery in drug design. *Nature Reviews*, *1*, 45–54.
81. Rees, D. C., Congreve, M., Murray, C. W., Carr, R. (2004). Fragment-based lead discovery. *Nat. Rev.*, *3*, 660–672.
82. Erlanson, D. A., Lam, J. W., Wiesmann, C., Luong, T. N., Simmons, R. L., DeLano, W. L., Choong, I. C., Burdett, M. T., Flanagan, W. M., Lee, D., Gordon, E. M., O'Brien, T. (2003). In situ assembly of enzyme inhibitors using extended tethering. *Nat. Biotechnol.*, *21*(3), 308–314.

83. Zhang, Kim (2003). Overview of structural genomics: From structure to function. *Curr. Opin. Chem. Biol.*, *7*, 28.
84. Zarembinski, T. I., Hung, L.-W., Mueller-Dieckmann, H.-J., Kim, K.-K., Yokota, H., Kim, R., Kim, S.-H. (1998). Structure-based assignment of the biochemical function of a hypothetical protein: A test case of structural genomics. *PNAS*, *95*(26), 15189–15193.
85. Mathews, I. I., Deacon, A. M., Canaves, J. M., McMullan, D., et al. (2003). Functional analysis of substrate and cofactor complex structures of a thymidylate synthase complementing protein. *Structure*, *11*, 677–690.
86. Schneider, T. R. (1996). What can we learn from anisotropic temperature factors? In E. Dodson, M. Moore, A. Ralph, and S. Bailey (Eds.), *Proceedings of the CCP4 Study Weekend.* SERC Daresbury Laboratory, Daresbury, UK, pp. 133–144.
87. De Pristo, M. A., de Bakker, P. I. W., Blundell, T. L. (2004). Heterogeneity and inaccuracy in protein structures solved by X-Ray crystallograph. *Structure*, *12*, 831–838.
88. Davis, A., Teague, S. (1999). Hydrogen bonding, hydrophobic interactions, and failure of the rigid receptor hypothesis. *Angew. Chem. Int. Ed. Engl*,. *38*, 736–749.
89. Teague, S. J. (2003). Implications of protein flexibility for drug discovery. *Nat. Rev. Drug Discov.*, *2*, 527–541.
90. Sanders, W. J., Nienaber, V. L., Lerner, C. G., McCall, J. O., Merrick, S. M., Swanson, S. J., Harlan, J. E., Stoll, V. S., Stamper, G. F., Betz, S. F., Condroski, K. R., Meadows, R. P., Severin, J. M., Walter, K. A., Magdalinos, P., Jakob, C. G., Wagner, R., Beutel, B. A. (2004). Discovery of potent inhibitors of dihydroneopterin aldolase using CrystaLEAD high-throughput X-ray crystallographic screening and structure-directed lead optimization. *J. Med. Chem.*, *47*, 1709–1718.
91. Reddy, M. R., Erion, M. D. (Eds.) (2004). *Free Energy Calculations in Rational Drug Design.* Kluwer Academic.
92. Verkhiveker, G. M., Bouzida, D., Gehlhaar, D. K., Rejto, P. A., Freer, S. T., Rose, P. W. (2002). Complexity and simplicity of ligand-macromolecule interactions: The energy landscape perspective. *Curr. Opin. Struct. Biol.*, *12*, 197–203.
93. Thiel, K. A. (2004). Structure-aided drug design's next generation. *Nat. Biotechnol.*, *22*(5), 513–519.

# BIOLOGICAL AND CHEMISTRY ASSAYS AVAILABLE DURING DRUG DISCOVERY AND DEVELOPABILITY ASSESSMENT

DUANE B. LAKINGS
*Drug Safety Evaluation Consulting, Inc.*
*Elgin, Texas*

1 INTRODUCTION
2 DISCOVERY ASSAYS
In Vitro Pharmacology Assays
Analytical Chemistry Assays
In Vivo Pharmacology Assays
3 DEVELOPABILITY ASSAYS
Additional Pharmacology
Formulation Development Assays
Drug Delivery Assays
Bioanalytical Chemistry Assays
Pharmacokinetic Assays
Drug Metabolism Assays
Toxicology Assays
4 CONCLUSIONS

## 1 INTRODUCTION

The discovery and development of a novel therapeutic agent, whether a small organic molecule (commonly referred to as novel chemical entity, or NCE) or

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

a macromolecule (such as a natural protein, modified protein, polypeptide, or oligonucleotide) requires scientific expertise from a number of different disciplines and an enormous amount of time (at least 5 to 6 years if everything goes smoothly but more realistically 10 to 12 years) and money (hundreds of millions of dollars). While humans may be the ultimate test species to ascertain the safety, efficacy, and pharmacokinetics of a potential new therapeutic agent, research studies conducted during the discovery phase in in vitro systems and in animal models can determine whether a discovery lead or group of leads:

1. Has the desired pharmacological profile (i.e., biological activity) for mediating a human disease process.
2. Has the necessary developability properties (e.g., sufficient aqueous solubility for delivery across membranes and metabolic stability or acceptable metabolism) to be a successful drug candidate.
3. Does not have a toxicity profile that could cause adverse experiences in humans at pharmacological doses.

At present, the drug discovery process consists of two distinct phases, like the two faces on a coin. On one side is the drug discovery *face* where researchers are charged with quickly finding new leads using protein, enzyme, or other targets that can be antagonized or agonized to possibly mediate a human disease or disorder. Some of these targets have been well characterized, but others have only recently been identified using genomic and proteomic research results. Compounds previously synthesized, combinatorial chemistry libraries, and/or newly prepared compounds (now commonly prepared using a rational drug design paradigm that models the active site of the target with the chemical structure to obtain a "better fit" and thus greater biological activity) are tested against the target using a high-throughput system (HTS) that can evaluate thousands of compounds in a short period of time. The compounds with the highest biological activity are designated *hits*. These hits are usually structurally modified and then retested in the HTS pharmacology screen to identify the pharmacophore (or the part of the chemical structure responsible for activity) and to determine if the addition or change of other chemical moieties increases or decreases the biological activity. This process is termed the structure–activity relationship (SAR) assessment. The modified hits with the highest biological activity in the in vitro pharmacology screen are designated *leads*. In years past, the lead with the highest activity in the screen was selected as the preclinical drug candidate. However, most of these leads have limited druglike properties, primarily because the active site of the target is lipophilic, and thus the most active compounds are also highly lipophilic (a property that causes limited aqueous solubility that in turn limits the ability of the compound to be delivered to the active site after administration to animal models or humans). Thus, most pharmaceutical and biotechnology firms now conduct some additional studies on these discovery leads.

On the other side of the coin is the lead optimization or developability assessment face where researchers conduct studies to determine if the new leads have the necessary druglike properties to reach the pharmacological site of action in sufficient concentration and for a sufficient duration to effectively interact with the in vivo target. These researchers use the expression *fail early, fail often* to define their charge of "killing", or removing from further consideration those discovery leads that do not have the desired characteristics or attributes for further development. This goal of fail early, fail often should be reached, whenever possible, before the initiation of clinical trial testing and preferably even before starting the definitive preclinical studies needed to support first-in-human clinical studies. Failing during expensive and time-consuming clinical trials should occur as infrequently as possible, especially for small biotechnology companies that have limited resources and where a clinical failure may result in having to close the doors.

Historically, for every 100 compounds screened for toxicity and biological activity in animal models, only 1 has the necessary pharmacology and safety profiles for evaluation in humans. Of those compounds tested in humans, only about 1 in 5 is successfully brought to the marketplace. This poor rate of success has been attributed to a number of factors, a primary one being that in vitro systems and animal models are not truly predictive of biological activity and/or safety in humans. The problem, however, may be that insufficient time and resources are put into first discovering the lead and then characterizing the lead or group of leads in pharmacology, drug metabolism, and toxicology models to first select the "optimal" compound (i.e., the lead with the desired druglike attributes needed for successful development) and then to critically evaluate the results to ensure that a compound with developability problems does not enter into preclinical and clinical development until those demerits are resolved. Instead of a rush from the first sign of biological activity in an in vitro pharmacology test to clinical trail evaluation, careful design, conduct, and interpretation of additional pharmacology, pharmacokinetic, and toxicology evaluations will detect "loser" compounds much earlier. This "weeding out of losers" will allow precious time and resources to be devoted to identifying discovery leads with a better chance than 1 in 5 of successfully completing clinical studies and becoming a marketed therapeutic agent. Being able to identify the 499 losers of the 500 discovery leads that have a potentially desirable biological activity as early as possible in the development process will save substantial time and resources.

Thus, one side of the drug discovery coin is to find as many novel leads as possible, and the other side is to identify the lead with the best chance of success and to "kill" as many of the other leads as possible. On both sides of the coin, researchers rely on various biological and chemical assays to evaluate the pharmacological activity and druglike properties of discovery leads.

This chapter summarizes some of these in vitro and in vivo assays available to and employed by researchers during the drug discovery and developability

assessment phases of the discovery and development processes. The information presented should help researchers identify what assay systems are available and how these assay systems might be used to assist in first finding discovery hits and then in designing experiments to evaluate and optimize discovery leads as they progress through the drug discovery process and before they enter the preclinical research studies required to support regulatory agency submissions for first-in-human clinical trials.

## 2 DISCOVERY ASSAYS

The drug discovery process has undergone enormous changes during the past few years. After years of first synthesizing individual compounds, then purifying and obtaining physical and chemical characterization of the new chemicals, and finally testing the NCEs in in vitro and in vivo pharmacology models of a particular human disease, the pharmaceutical and biotechnology industry has embraced rationally designed combinatorial chemistry as a way to generate large numbers of organic compounds for evaluation. Other sources of compounds for evaluation include previously synthesized compounds, with some pharmaceutical companies having more than 100,000 unique chemical structures in their repositories. Using computer models of the active site of a target, *virtual* compounds with the appropriate structure to fit into and bind with the active site can be identified. Once these chemical structures have been determined, organic chemists synthesize series of compounds around these *structures*. These compounds are evaluated first in in vitro and then in in vivo pharmacological assays.

### In Vitro Pharmacology Assays

Existing compounds or rationally designed, combinatorial chemistry-generated libraries are screened for biological activity, usually in an in vitro system in which a known biochemical process, which is thought to mimic a human disease or disorder, is agonized or antagonized. The increased number of compounds or mixtures of compounds to be tested has required pharmacologists to devise novel techniques or assays to screen rapidly for biological activity. Because the amount of material available from combinatorial chemistry syntheses is small (i.e., usually microgram quantities), miniaturization of the biological assay system is a necessary first requirement. The large number of compounds makes automation of the assay the second requirement. The use of first 96-well, then 384-well, and now larger titer plates and robotics to add the appropriate small amounts of combinatorial chemistry libraries and reagents, including the material that generates the signal for a positive result, provides such a system for rapid assessment of biological activity for a large number of compounds. Those assay wells that elicit a positive or negative

response, depending on the biological test being conducted, are identified and if necessary, the compound(s) of interest is isolated and identified.

These in vitro pharmacology assays are commonly disease-target specific and are designed using one of two primary approaches. The first approach uses only the target, either isolated from cells or tissues that express the target or by incorporating the genetic code for the target into a cell system and, after expression, isolating the target. The target is then combined with known factors needed for activity and a detection system. These known factors might be an energy source, such as adenosine triphosphate (ATP), or a coenzyme, such as nicotinamide-adenine dinucleotide (NAD). Detection systems are usually target specific and are frequently a known substrate for the target that is or is not affected in the presence of an "active" discovery compound. The amount of change in the substrate or generation of the target–substrate product is a measure of the activity of the discovery compound being tested. The target may be bound to the titer plate or free in solution. Using robotics, aliquots of solutions containing the discovery compounds to be evaluated are added. Usually three or four concentrations, but sometimes more, of each discovery compound are tested. To describe the activity–concentration curves needed for evaluating and comparing the biological activities of a number of discovery leads, at least four levels of each compound need to be tested over a fairly wide concentration range. Since many of these discovery compounds have only limited aqueous solubility, a common practice is to prepare high-concentration solutions in dimethyl sulfoxide (DMSO) or another organic solvent, such as ethanol, and then dilute this solution with an aqueous buffer, such as phosphate buffered saline (PBS), to obtain the desired concentration range for testing. Even with this practice, some discovery compounds have insufficient solubility for effective testing against the target. The advantages of this approach are that the assay system is relatively simple, stable, and uniform across titer plates, and thus a large number of discovery compounds can be evaluated and compared in a short period of time. Disadvantages include that the discovery compounds do not have to be delivered (i.e., cross one or more membranes) to reach the target and that biotransformation products (i.e., metabolites) of the discovery compounds are not tested for activity.

The second approach used for in vitro pharmacology activity screening is to have the target expressed in a cell system. The "best" cell system would be the one where the target is expressed in the body and where the discovery compound will need to be delivered to be pharmacologically effective. This requires knowledge of where and how the target interacts with other proteins and systems to control or mediate a biological process, which when not functioning properly leads to a human disease or disorder. For "old" targets (i.e., targets that have been used for years for discovering new drugs), this functional knowledge has been mostly defined. However, for "novel" targets (i.e., targets that are being discovered through genomic and proteomic efforts), this

knowledge is still being generated, and many of these novel targets may not be "drugable," which means that interaction with the target will not effectively medicate a disease process or that interaction with the target will cause undesirable effects in tissues or organs that express the target but are not involved in the disease process. Another approach used for a cell-based in vitro pharmacology assay is to incorporate the genetic code for the target, and possibly a detection system, into a cell, such as a bacteria or mammalian cell. The cell then expresses the target and the ability of discovery leads to interact with the target can be determined. While this cell-based approach has the advantage of requiring the discovery compounds to be delivered across at least one membrane, the cell line expressing the target may not be representative of the disease process to be mediated or contain the other proteins or enzymes involved with the target. Even so, an in vitro pharmacology assay in a cell system is considered more "robust" than an assay where the target is "free" in the assay system. Another advantage of a cell-based system, as will be discussed later in this chapter, is that information on the potential toxicity of a discovery compound can also be obtained by evaluating sufficiently high concentrations that may produce a toxic event, such as cell death.

Whichever approach is used, the in vitro pharmacology assay needs to be carefully characterized and, when possible, validated. Characterization includes determining what cofactors and other reagents are required for the target to be "active" and where in the body (i.e., which tissues and organs) the target is expressed. Validation means that agonizing or antagonizing the target in an in vivo animal model of the disease or, better yet, in humans produces the desired effect of mediating a disease process or disorder. Most of the newly defined targets have not yet been fully characterized or validated. However, many academic and pharmaceutical company laboratories are busy with these endeavors and will hopefully soon have a number of the novel targets sufficiently characterized so that discovery compounds that interact with these targets can be further evaluated in animal models of human diseases or disorders or in humans. In any case, those discovery compounds with the desired biological activity against a target in an in vitro pharmacology assay are designated hits and are evaluated further.

## Analytical Chemistry Assays

Once the initial in vitro pharmacology assay results are available, the hits with the desired biological activity will be studied more intensively, usually in an in vivo animal model (discussed below) of the disease indication. Before these additional pharmacology assays can be conducted, a sufficient quantity of each discovery hit to be tested further will need to be synthesized, purified, identified, and characterized. The techniques employed for the synthesis and purification of the discovery hits are outside the scope of this chapter on assays; however, the identification and characterization of the discovery hits requires the use of analytical chemistry assays. For identification, the assay techniques

most commonly employed are mass spectrometry (MS) and nuclear magnetic resonance (NMR) spectroscopy. MS can usually provide molecular weight (MW) information on the hits, and the MS fragmentation pattern can be used to determine key structural aspects of the molecules. The MW of the hits confirms the elemental composition (i.e., the number of carbons, hydrogens, nitrogens, oxygens, and other elements in the chemical structure) of the compounds that were synthesized. NMR spectra provide information on how the various moieties within the chemical structure interact, and, along with the MS fragmentation pattern, they are used to show that the compounds prepared are those with the in vitro biological activity. A common problem at this stage of the discovery process is that the hits being evaluated have poor purity, sometimes 50 percent or less, and thus the biological activity observed may not be only from the hits but also from pharmacologically active impurities or degradation products. Thus, each of the hits to be considered further should be characterized for purity, which requires an analytical chemistry assay that can separate a hit from impurities and/or degradants and that can detect not only the hit but also any impurity and/or degradant. Analytical chemistry techniques employed for determining the purity of hits are chromatographic assays such as thin-layer chromatography (TLC), high-performance liquid chromatography (HPLC), and for macromolecule leads, size exclusion chromatography and sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS–PAGE). The most common detection system for each of these assays is ultraviolet (UV) detection at a relatively low wavelength (i.e., 214 nm, since most organic compounds absorb light at this wavelength). Given the chemical structure of an analyte (i.e., a discovery hit), an experienced analytical chemist can usually select the appropriate assay method and the initial method parameter settings. Parameters that will need to be considered include the chromatographic column type (reversed–phase, normal phase, ion exchange, size exclusion) and size, the mobile phase and its composition (organic phase, buffer, and the percentage of each), the elution technique (isocratic or gradient), and the detection system (UV, refractive index, MS). The discovery lead is dissolved in the mobile phase and is injected. The chromatogram should contain one major peak (i.e., the discovery hit) and possibly a number of minor peaks, which may be impurities from the synthetic procedure. If more than one major peak is detected, further evaluations are immediately necessary since the compound of interest may be "contaminated" with another compound or may have partially degraded. Which of these major components has the desired biological activity will need to be assessed before additional testing is conducted.

If a discovery hit has one or more chiral centers, a stereospecific analytical chemistry method with the ability to separate all of the enantiomers will be necessary. Chiral chromatography or derivatization to form diastereoisomers (compounds with two asymmetric centers) are techniques commonly used. If a single enantiomer is being considered for further development, conversion to the other enantiomer is considered instability with the formed enantiomer

being a degradation product. If possible, the enantiomers should be separated (using analytical chemistry techniques) or synthesized and tested in the in vitro pharmacological assay as soon as possible. If one enantiomer has more biological active than the other, further studies should be conducted on the "more active" enantiomer. If the enantiomers have similar biological activity, development of a racemic mixture is a possibility. However, regulatory authorities recommend that the pharmacology, pharmacokinetic, and toxicology profiles of the enantiomers be separately determined and the results be used to justify the development of a racemic mixture over a single enantiomer. Thus, evaluations on the enantiomers are needed as early as possible and should start with the in vitro pharmacology assay.

## In Vivo Pharmacology Assays

After a sufficient quantity of the active compounds "discovered" during the in vitro pharmacology screening process have been prepared and characterized, additional studies are conducted to more fully evaluate the biological activity of these hits. An important requirement for this additional testing is having or developing an animal model, or in vivo pharmacological assay, that correlates to, or mimics, a disease or disorder in humans. Developing these animal models can be time consuming and expensive and is often complicated by the fact that the model may not simulate the disease as manifested in humans. Many important human diseases, including psychoses, depression, Alzheimer's disease, AIDS (acquired immunodeficient syndrome), and many cancers, do not yet have predictive animal models. In addition, most targets being discovered using genomic and proteomic efforts do not yet have an in vivo pharmacological model.

A necessary first step in the development of an in vivo pharmacological assay is to determine where in the body [i.e., the organ(s) or tissue(s) or particular cell types within an organ or tissue] the target is expressed and how interaction with the target will mediate the disease process. If a target is also expressed in organs or tissues that are not associated with the disease or disorder, interactions with the target in these organs/tissues may produce undesirable effects. Thus, targets that are widespread in the body may not be as "drugable" as a target that is expressed only in organ/tissue of interest [i.e., a brain neuron for a central nervous system (CNS) disorder]. Another important aspect is whether the target is within a cell and, if so, where within the cell; has the active site on the cell surface; or is excreted by the cell and thus is in extracelluar space. Targets that are intercellular may be more difficult to reach and may be involved in more than one biochemical pathway or cascade. Inhibiting or activating such a target may result in the desired effect but may also adversely affect other processes and thus lead to undesirable responses. Targets that are on cell surfaces are usually readily assessable and frequently are involved in a single process, such as providing a signal for controlling an intercellular reaction or assisting in the transport of

an ion or biochemical through the cell wall. Targets that are in extracelluar space are also readily assessable but may have a wider distribution throughout the body. Information or where and why a target is expressed is critical for the successful definition of an animal model of a human disease or disorder.

After the animal model has been characterized, various dose levels of the hit(s), identified from screens of combinatorial chemistry libraries or other sources, are administered to the test species and a dose–response curve is generated. The most commonly used endpoint is the dose that provides a 50 percent effective dose response (the $ED_{50}$). Structural analogs of the hit(s) are frequently synthesized and tested in the same model to generate a family of curves with varying biological potencies or $ED_{50}$'s. These structural analogs may contain the addition or deletion of polar or nonpolar moieties that enhance or decrease the biological activity of the hit. Structural changes that enhance the biological activity of a hit are continued until a compound with the desired pharmacological profile or biological potency is obtained. This hit is now designated a lead. Frequently, a number of leads, each with a unique chemical structure, are identified.

At times, companies select the lead with the greatest biological activity in the in vivo pharmacological assay for further development. However, at this point in the drug discovery process, many unanswered questions still exist. Some of these concerns are:

1. Does the animal model used to identify the leads reflect, in most or all aspects, the disease in humans?
2. Is the delivery or extent of exposure of all the leads similar so that the generated dose–response curves accurately predict the compound with the greatest in vivo potency?
3. Are the route and frequency of administration and the formulation used to dose the pharmacological animal model similar to those proposed for preclinical animal safety studies and for human safety and efficacy evaluations?
4. Do the leads have similar pharmacokinetic and drug metabolism profiles so that the durations of exposure to the pharmacologically active compounds are similar?
5. Do any of the leads produce unacceptable toxicity in organs or physiological systems not involved in the desired pharmacology of the leads?

Before the more formal and definitive preclinical development begins, attempts should be made to answer as many of these questions as possible. The following section discusses some of the assays utilized in drug developability experiments that can be conducted relatively quickly and cheaply to ascertain whether a lead has the necessary biopharmaceutical, or druglike, characteristics needed for further development.

## 3 DEVELOPABILITY ASSAYS

Completed drug discovery research studies indicate that a compound (now designated a discovery lead) has sufficient in vitro and in vivo biological activity to mediate a disease process and thus has potential to become a human therapeutic agent. Is this lead now ready to be transferred from the discovery area to the preclinical development group? Should additional, nondefinitive research experiments be conducted to more fully characterize the properties of the lead? If more studies are considered necessary, what experiments should be conducted? If additional studies are to be performed, what types of biological and chemical assays are available to support these efforts?

Frequently, more than one discovery lead has the desired biological activity in the in vivo pharmacological assay. When that is the case, which of these leads should be selected for entry into the more formal preclinical drug development process? What criteria, other than biological potency, should be used for selecting this "optimal" lead from a group of leads? Can a series of research experiments (i.e., lead optimization studies) be identified for determining which of the leads have the druglike attributes considered necessary for successful development and which of the leads have undesirable properties or demerits that would be suggestive of potential development problems?

This section describes some of the biological and chemical research assays that could, and in most cases should, be used to evaluate the potential of a discovery lead to become a developmental candidate or to select the optimal lead from a group of discovery leads. Figure 1 shows where these developability experiments fit into the drug discovery and development process. These nondefinitive developability studies may also uncover problems that have to be resolved before the definitive preclinical development studies, required to support a regulatory agency submission for first-in-human clinical trials, are

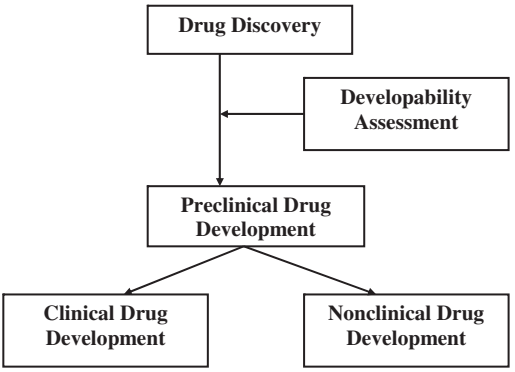


**Figure 1** Where developability assessment experiments fit into the drug discovery and development process.

started and before the clinical protocols to evaluate the human safety, efficacy, and pharmacokinetics of a drug candidate are designed.

Before the preclinical drug development program is begun and because each discovery program is company and compound specific, a number of questions should be asked and answered so that the developability assessment research studies (both types and designs and whether or not appropriate assays are available) may be planned effectively and the timelines for their completion can be determined. These questions include, but are not limited to:

1. What is (are) the human disease indication(s) or disorder(s) to be studied in the clinic?
2. What are the proposed route and frequency of dosing for human clinical trails?
3. What is the estimated pharmacologic active substance (i.e., the discovery lead or an active metabolite) concentration in physiological fluids and how long should that concentration be maintained so that the desired biological response can be obtained?
4. What, if any, are the biological markers to monitor toxicity or therapeutic effectiveness in to-be-conducted preclinical, nonclinical, and clinical studies?

Additional questions need to be considered, including whether a lead can be synthesized and purified in sufficient quantity to support a development program, how to document and validate these processes, how to characterize and document the identity of and impurities present in the drug substance and the proposed drug product, and how to determine the stability of the drug substance and the proposed drug product. While each of these processes requires assays, these questions are outside the scope of this discussion in this chapter.

At least eight, and possibly more, scientific disciplines are involved in the developability characterization of a discovery lead or group of leads, and each of these disciplines utilize different assays, some biological and other chemical. As shown in Figure 2, these disciplines are in vivo pharmacology, analytical chemistry, solubility and stability, drug delivery, bioanalytical chemistry, animal pharmacokinetics, drug metabolism, and toxicology. The following sections provide some summary information on the assays for each of these scientific disciplines. The discussion is mostly for a single discovery lead; if more than one lead is being evaluated to select the lead with the most druglike properties, which depends on a variety of factors such as disease indication, route and frequency of dosing, and so forth, for further development, similar experiments could be conducted on each lead. Normally, these developability assessment experiments are conducted in an iterative fashion with the attributes considered most important evaluated first. Generally, in vitro evaluations,

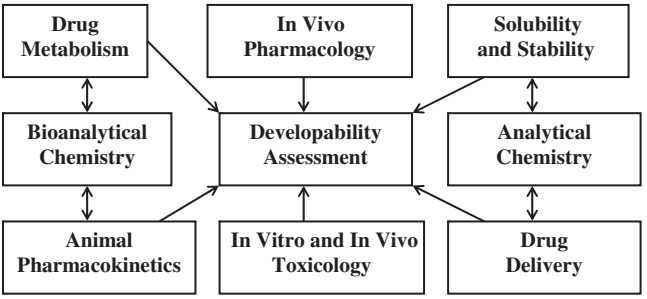


**Figure 2** Scientific disciplines involved in developability assessment.

which are less costly (both in terms of resources and time), are conducted before in vivo studies. If a lead or none of the leads in a group of leads has the desired profile for a given attribute, additional discovery leads should be identified and tested, like the SAR approach discussed earlier. Once a lead or a number of leads from a group has an acceptable profile for each attribute being evaluated, that lead or group of leads moves into the next stage of developability assessment experiments. This lead-optimization iterative process should be continued until a lead with the desired properties has been identified. At times, more than one lead will have the desired, or at least an acceptable, developability profile. When that pleasant occurrence happens, a primary lead can be designed with the other leads serving as backup candidates.

## Additional Pharmacology

Preliminary pharmacology evaluations using in vitro or animal model assays developed and characterized during the discovery process will have shown that a lead interacts with a biological process suggestive of human therapeutic benefit. Depending on the design and extent of these early studies, additional pharmacology studies or assays may be needed to further characterize the dose, or physiologic fluid concentration, response curve using the proposed clinical route and frequency of administration. If possible, these pharmacology studies should be conducted in both males and females to show that the biological response is not gender specific. If the indication is a disease of age, older or younger animals, as appropriate, should be evaluated. The $ED_{50}$ dose should be determined and that value divided into the no-observable-adverse-effect-level (NOAEL) dose in the same animal species, described later in the section on toxicology, estimates a therapeutic ratio (TR) or index (TI). If the TR is one or less, a lead will most likely elicit adverse effects in addition to the beneficial response. Unless the lead is for the treatment of a life-threatening disease, such as AIDS, some cancers, or certain CNS indications, a low TR is a warning sign that the lead may not have the necessary properties for further development. A TR of 5, preferably 10 or higher, indicates that

a lead will most likely produce a pharmacological response before causing dose-limiting toxicities.

If possible, these developability pharmacology studies should be conducted with dosing to steady state unless the frequency of dosing to be used in clinical trials is as a single-dose or a few dose therapeutic. The number of doses required to reach steady state depends on the lead's pharmacokinetic profile (discussed later in the section on preliminary pharmacokinetics) in the same animal model. These multiple-dose studies provide information on the frequency of dosing necessary to maximize the biological response. This is particularly important for compounds that inhibit an enzymatic system or are effective only during certain phases of the cell cycle.

Another important aspect that should be considered when conducting these pharmacology evaluations is the identification and characterization of potential biological markers that may be predictive of therapeutic benefit. For some targets, markers may already be known, and, if so, assays for these markers should be adapted and incorporated into the studies. However, for many targets, predictive biological markers are not yet available. If the interaction of the lead with a target inhibits or enhances the formation of a particular biochemical, the concentration of that biochemical may be indicative of therapeutic response. Developing and utilizing assays for this marker first in animal models and then in human clinical studies may define a surrogate marker for tracking disease response or progress.

These pharmacology evaluations assist in the selection of dose levels and dosing regimens for preliminary and definitive toxicology studies and for initial human pharmacology or phase 1 safety and tolerance clinical trials. If the effective pharmacologic dose is unknown, underdosing and achieving no therapeutic response or overdosing and being unable to define an NOAEL dose are undesirable possibilities. In such cases, the development of a potential beneficial therapeutic agent could be inappropriately discontinued.

## Formulation Development Assays

Nonclinical formulation definition of a lead is not usually studied in detail during the transition from discovery to preclinical development. The experiments necessary to define an acceptable formulation depend on the proposed clinical route of administration and usually require substantial quantities (i.e., milligram or gram amounts) of the lead. For a compound to be administered by intravenous injection or infusion, the formulation needs to be compatible with blood so that the compound does not precipitate when administered and has minimal local toxicity. Leads that are highly lipophilic or have limited aqueous solubility are the most likely compounds to have these types of problems. A low extent of, or a high variability in, absorption can cause problems for leads administered by other routes, such as oral, subcutaneous, or dermal. For compounds that are poorly absorbed, the amount reaching the site of action may be insufficient to elicit or to maintain a desired pharmacologic

response. If the absorption is variable and the TR is low, a toxic response may be observed in some, but not all, animals or later in some humans. Experiments conducted by the author have shown that when the extent of absorption is 50 percent or more and the variability of absorption is less than 50 percent of the amount absorbed, which gives a 25 to 75 percent range of absorption, a lead has an acceptable bioavailability for further development. For leads with an extent of absorption less than 25 percent of the administered dose or a variability of absorption of more than 100 percent of the amount absorbed, other formulations with absorption enhancers or solubilizers might be evaluated to improve the drug delivery profile. If improvement in the drug delivery profile is not possible, the chances that a lead with low, variable absorption will become a therapeutic product are greatly reduced. The candidacy of such a compound should be carefully considered.

The analytical chemistry assay defined for characterizing a discovery hit should be further developed for the quantification of a lead in nonclinical formulations and should predict whether the compound degrades from the time of preparation to the time of dosing. An assay with this ability is called a stability-indicating assay. The physical and chemical properties of the lead usually suggest a technique (heat, light, pH) that can degrade the compound. Samples stored under nondegrading and degrading conditions are assayed by the stability-indicating assay. If the lead is not stable, formulation excipients possibly can be added to prevent the degradation or the lead can be maintained under conditions that provide sufficient stability for testing in animal models. However, when a lead has limited stability, either alone or in nonclinical formulations, the development of a clinical formulation with a sufficient shelf life for marketing is problematic. Again the candidacy of such a compound should be carefully considered.

For proteins and other large molecules, degradation may include changes in the secondary or tertiary structure, provided that rearrangement back to the original, biologically active structure does not occur. Stability-indicating assays for macromolecules should assess structural changes that cause reductions in biological activity. However, a protein may have a number of amino acids removed from one or both ends of the molecule and still retain some, and possibly all, biological activity relative to the intact protein. This chemically modified peptide may have different drug delivery and pharmacokinetic profiles or be more toxic than the parent protein. Structural modifications may not be apparent if biological activity alone is used to determine the amount of the macromolecule in a formulation. Thus, experiments to demonstrate that an assay method is stability indicating need to be carefully designed and conducted. For macromolecules, a specific chemical assay, such as HPLC or enzyme-linked immunosorbent assay (ELISA), and a biological potency assay, such as the in vitro pharmacological assay, may be necessary to determine the concentration and stability of a compound in a formulation.

The stability-indicating assay should be used to determine the amount of the lead in nonclinical formulations used for dosing animals in in vivo phar-

macology, preliminary pharmacokinetic, drug metabolism, and toxicology studies. For single-dose studies, samples from each formulation at each dose level can be collected before dosing and after the completion of dosing. For multiple-dose studies, samples from each formulation used can be collected before the first dose and after the last dose or at selected other times. Results from these analyses ensure that the formulations contained the desired amount of the lead, that the concentrations did not change during the dosing period, and that the test species received the appropriate dose levels.

Another application of a stability-indicating assay is to evaluate the solubility profile of the discovery lead under the various conditions that the lead may experience during dosing and delivery to the site of pharmacological action. As mentioned earlier, a lead needs to have some aqueous solubility in order the effectively cross membranes. A lead to be orally administered will be subjected to high acidic conditions in the stomach and then to the more neutral conditions of the upper gastrointestinal (GI) tract before being absorbed into systemic circulation. Thus, stability of the lead is necessary under both acidic and neutral conditions. For most other routes of delivery, a lead will be subjected to neutral (i.e., physiological) to slightly basic conditions. The analytical chemistry assay developed earlier is usually capable of evaluating the solubility of a lead over the range of pHs. With slight modification, the robotic system used for the in vitro pharmacology assay can be adapted for the generation of samples for solubility assessments. Using the concentration of the lead that provides the desired biological activity as a starting point, the desired solubility range can be estimated. For a lead with high potency (i.e., pg/mL or ng/mL), aqueous solubility at the milligram/milliliter level may not be necessary. A high-concentration solution, frequently in an organic solvent such as DMSO, of the lead is prepared and small aliquots, usually microliter quantities, are added to aqueous solutions that contain a buffer, commonly phosphate buffer saline (PBS), and are adjusted in pH to give the desired pH range. After appropriate mixing, the samples are centrifuged or filtered to remove any precipitate (i.e., unsolubilized lead), and aliquots of the supernatant are assayed using the quantitative analytical chemistry method. The concentrations of the lead in the samples are used to prepare a solubility profile for that lead. Discovery leads that have poor aqueous solubility under conditions expected to be encountered during dosing and delivery to the site of pharmacological activity usually make poor drug candidates.

## Drug Delivery Assays

Without an acceptable nonclinical formulation, the extent and variability of delivery may make interpretation of results from developability studies meaningless and prevent the continued development of a potentially useful therapeutic agent. The best formulation is of little or no use if the lead is not effectively delivered to the site of biologic action. One of the primary reasons discovery leads are not successful drug candidates is limited or insufficient

transport across various membranes from the site of dosing to the site of activity. Only compounds administered intravenously to mediate a disease indication expressed in the cardiovascular system do not have to cross at least one membrane in order to reach the site of action. Thus, assessments of a lead's ability to cross membranes should be conducted as early as possible. Many pharmaceutical companies use delivery potential as a key indicator for whether or not a discovery lead should move into preclinical development.

Since most membranes are lipophilic in nature, a lead has to have some lipophilic characteristics in order to effectively diffuse into and across the membrane. However, in order to reach the membrane, the lead has to be dissolved in the surrounding media, which is aqueous. Thus, the lead also needs some hydrophilic properties in order to have sufficient aqueous solubility to be transported to the membrane. An estimate of a lead's ability to have both the lipophilic and hydrophilic characteristics necessary for effective delivery, primarily from the GI tract, can be determined from the chemical structure of the compound and using what is commonly called Lipinski's rules of five, which are four rules with cut-off numbers that are 5 or multiples of 5. These rules are:

1. A molecular weight of less than 500 daltons.
2. A $\log P$ (octanol–water coefficient) of less than 5.
3. Hydrogen bond donors (sum of hydroxyl and amine groups) less than 5.
4. Hydrogen bond receptors (sum of nitrogen and oxygen atoms) less than 10.

While these rules may be somewhat predictive of a lead's ability to cross membranes, not all compounds having the desired characteristics are orally absorbed or effectively transported across membranes. Thus, laboratory experiments, which need characterized assays, are required to determine if a lead will effectively be delivered to the site of biologic action.

A number of in vitro models or assays are available to evaluate the delivery potential of a lead. For a lead to be administered orally, the Caco-2 model is most commonly employed. Other in vitro systems are available to evaluate transport across other membrane types, including the blood-brain barrier (BBB) for CNS disorders, lung for pulmonary delivery, skin for topical delivery or to treat dermal disorders, and eye for ocular delivery. Each of these assay systems is summarized below. The analytical chemistry assay discussed earlier is usually capable of quantifying the lead in samples generated by these in vitro delivery models.

For determining the potential intestinal absorption and secretion of discovery leads, the most commonly used in vitro assay system employs Caco-2 cells, which were originally derived from a human colorectal carcinoma. When cultured on semipermeable membranes, Caco-2 cells differentiate into a highly functionalized epithelial barrier with remarkable morphological and bio-

chemical similarity to the small-intestinal epithelium. The disappearance of a discovery lead in the reservoir above the confluent cell layer and the permeable membrane and the appearance of the lead and/or degradants or metabolites on the other side are measured using an analytical chemistry technique. Permeability coefficients obtained from Caco-2 cell transport studies have been shown to correlate with human intestinal absorption. Caco-2 assessments are commonly conducted in 24-well assay systems and can be configured for high-throughput systems.

A 96-well technique using a lipid membrane has been shown to have excellent correlation with other in vitro absorption systems (e.g., Caco-2 cells) and with in vivo results. When the assessment of the delivery potential of a large number of discovery leads, or even discovery hits, is needed, this membrane assay system may have application.

Whether attempting to assess the ability of discovery lead(s) for a CNS indication to reach the site of action or evaluating the potential toxicity of a lead(s) that should not be extensively delivered to the brain, assessment of transport across the BBB is a necessary first assessment. Primary microvascular cells from bovine brain tissue, which express the appropriate tight junctions and have the P-glycoprotein transport characteristics of the BBB, are commonly used for studying BBB permeability. The assay system and analytical chemistry techniques used for Caco-2 studies of GI permeability are employed.

Pulmonary delivery is a route of administration that has a number of advantages over oral administration. A discovery lead(s) administered to the lung can reach systemic circulation without having to cross the GI mucosa and without having to get past the liver, where many compounds are highly metabolized (i.e., have high first-pass metabolism). To evaluate the potential of a discovery lead for delivery through the lung, an in vitro assay system is available. Cultured, normal, human-derived tracheal/bronchial epithelial cells form a pseudostratified, highly differentiated model that closely resembles the epithelial tissue of the respiratory tract. Histological cross sections of both the in vitro tissue and a normal human bronchiole show a pseudostratified mucociliary phenotype. Numerous microvilli and cilia on the apical surface are present and confirm the presence of tight junctions. Cultures grown on cell culture inserts at the air–liquid interface provide gas-phase exposure of volatile materials and allow the measurement of transepithelial permeability for inhaled drug delivery studies on discovery leads.

Another potential route of delivery is dermal, and an in vitro assay system is available for evaluating the ability of a discovery lead(s) to penetrate through the skin and thus to reach systemic circulation. Normal human epidermal keratinocytes cultured on a permeable membrane produce a stratified, highly differentiated, organotypic tissue model of the human epidermis. The cultured tissue consists of metabolically and mitotically active cells organized into basal, spinous, and granular layers along a multilayered stratum corneum. The skin construct has an air-liquid interface and exhibits in vivo-like morphological and growth characteristics that allows discovery leads to be directly

applied to the surface of the tissue. This in vitro system approximates the barrier of normal human skin. The topical mode of application of the lead mimics the route of human exposure for some disease indications. Of course, treatment of some skin disorders does not require delivery of a therapeutic agent across the skin and into the body. When that is the case, substantial delivery into systemic circulation may be considered a developability demerit since only local exposure is desired and necessary to effectively treat the disorder. The in vitro skin assay may then be used to predict which discovery leads have less delivery across the skin and thus be predictable of lower exposure to potential organs or tissues of toxicity.

If the human disease or disorder to be evaluated is expressed in the eye, the ability of a discovery lead(s) to be delivered to the site of biological activity can be evaluated using an in vitro assay that mimic the eye. A three-dimensional in vitro tissue construct models the human corneal epithelium. Normal, human-derived epidermal keratinocytes cultured on a permeable membrane form a stratified, squamous multilayered epithelium similar to that of the cornea. This in vitro construct has an air-liquid interface and exhibits in vivo-like morphological and growth characteristics allowing discovery leads to be directly applied to the surface of the tissue and to approximate in vivo conditions. Analysis of a discovery lead(s) concentration on the liquid side of the system provides a measure of delivery into the eye.

## Bioanalytical Chemistry Assays

If not already available (which is common for newly discovered leads that may mediate a novel disease target), a bioanalytical chemistry method or assay needs to be defined and characterized for the quantification of the lead(s) in physiological fluids. This assay can then support experiments in some of the other scientific disciplines involved in assessing the developability of the lead and, after appropriate validation, the preclinical, nonclinical, and clinical development of a selected drug candidate. For preliminary studies, a bioanalytical chemistry assay should be characterized to demonstrate the range of reliable results, the lower and upper limits of quantification, specificity, accuracy, and precision. In addition, evaluations to determine the matrix to be used (blood, plasma, serum) should be conducted, and, once selected, the stability of the lead in that matrix type should be determined.

The first step in developing and characterizing a bioanalytical chemistry assay is to select the analytical technique. For an analyte with a molecular weight of less than 1000, instrumental methods such as HPLC coupled to an MS or liquid chromatography/mass spectrometry (LC/MS) or to a tandem or LC/MS/MS (a very sensitive and specific technique, and the most common method employed today by the pharmaceutical industry), or HPLC or gas chromatography (GC) with a variety of detectors, including UV, fluorescence, flame ionization (FID), and electron capture (EC), may be used. A macromolecular lead may require an ELISA or radioimmunoassay (RIA) method,

both of which require the generation of antibodies for capture and detection of the macromolecule. Generation of these antibodies may not be possible in the timeline required for supporting developability assessment experiments. If the in vitro pharmacology assay has been sufficiently characterized (as discussed below), this assay may have application as a bioanalytical chemistry assay. Samples of the lead prepared in assay diluent and in a physiological fluid and over a large concentration range should be analyzed to show that the technique produces an appropriate signal to detect the analyte and to determine the potential interference caused by the matrix, which is termed assay specificity. The ability to quantify a lead in a physiological fluid may depend on the matrix. For example, serum is a poor choice when the analyte interacts with clotting factors and thus may be coprecipitated when blood is allowed to clot to produce serum. The matrix that gives the best recovery and has the least interference when the compound is added to whole blood should be selected.

The specificity of an assay evaluates the potential interferences from physiological matrix components in specimens from the different animal species to be used in pharmacology and toxicology experiments. Samples from each species to be studied are analyzed neat (with no added analyte or lead) and fortified with known amounts of the lead, and the results are calculated using a standard curve prepared in assay diluent. The response obtained from the neat samples indicate the level of interference from each matrix, and the calculated amounts in the fortified samples show the difference in absolute recovery from the matrix compared with the analyte in buffer. If the absolute recovery is low, that is, less than 50 percent, and/or highly variable, that is, greater than 25 percent, the assay may not have the desired characteristics to quantify the lead in collected physiological fluid specimens. Additional development should be expended on such a method so that the assay will provide reliable results that can be used to evaluate the pharmacokinetic profile of the lead.

Acceptable results from the above experiments suggest whether a bioanalytical chemistry method should be able to quantify a lead in a physiological matrix. The range and reliability of quantification of the assay are assessed through the preparation and analysis of standard curves, prepared in either diluted or undiluted matrix, and multiple samples fortified at two or more concentrations and commonly referred to as quality control (QC) samples. The standard curve responses that can be described by a mathematical equation (linear, quadratic, sigmoidal) define the range of reliable results. The lower limit of quantification (LLQ) is the lowest signal that can be accurately measured above background and should not be confused with the limit of detection (LOD), which is the lowest level that can be detected. The upper limit of quantification (ULQ) is the highest signal that can be defined by the response curve. The results from the analysis of QC samples, usually three or more at each concentration, provide information on precision, defined as the ability to obtain similar calculated concentrations from samples containing the same

amounts of analyte, and accuracy, which is the ability to predict the actual concentration of the analyte in a sample.

The ability to measure a lead in a physiological fluid is not useful if the analyte is unstable during collection, processing, storage, or sample preparation. Thus, a nondefinitive stability study should be conducted to ensure that the compound does not degrade in blood during processing to obtain plasma or serum, during the time (hours, days, and weeks) and under the conditions (room temperature, refrigerated, frozen at –20 or –80°C) that collected specimens may be stored until analyzed, and during sample preparation, which includes thawing of frozen samples. The design of stability experiments is usually analyte and program specific. The results ensure that measured concentrations in unknown specimens reflect the amount of lead present at the time of collection.

Successful completion of the above experiments will characterize the assay method for use in evaluating a lead in animal models. If multiple leads are being evaluated, an assay capable of quantifying multiple leads using the same assay conditions should be the goal. If a lead is selected for further development, the method will need to be validated for each matrix and for each species before being used to support definitive toxicology, drug metabolism, and pharmacokinetic studies.

## Pharmacokinetic Assays

The first animal pharmacokinetic (PK) study confirms that the bioanalytical chemistry assay is useful in characterizing the absorption and disposition profiles of a lead. The animal species for this study is usually the same as used in in vivo pharmacology evaluations, most likely a rodent. A study design or assay for a lead that has pharmacologic activity when administered orally to rats may consist of dosing at least two rats with intravenous (IV) bolus injections at a dose level between 25 and 50 percent of the pharmacologically active dose and dosing at least two rats orally (PO) at the pharmacologically active dose. Serial blood samples, collected from each rat and processed to obtain the desired physiological fluid, are analyzed by the developed and characterized bioanalytical chemistry assay. The number of samples and the collection times are program specific and are usually determined by the "desired" PK profile for the lead, which in turn is defined by the desired interaction of the lead with the target. Commonly, 10 samples are collected from each test species. After IV dosing, the collection times might be 0 (predose), 10, 20, 30, 45, 60, 90, 120, 240, and 480 min postdose. For the PO doses, the collection time series could be 0 (predose), 15, 30, 45, 60, 90, 120, 180, 240, and 480 min postdose. The goal of the specimen collection time series should be to have a sufficient number of samples available for defining important PK parameters, such as distribution and disposition rate constants and their associated half-lives, concentration maximum ($C_{max}$), and time ($T_{max}$) to $C_{max}$ after non-IV dosing, and area under the concentration–time curve (AUC). The physiological fluid concen-

tration of the lead versus time profiles after IV dosing provide preliminary information on the distribution and disposition kinetics of the lead. These IV results certify that the assay method is useful for quantifying the lead in specimens obtained from animals, predict the concentration ranges that can be expected in animal specimens at pharmacological and toxicological doses, and assist in determining the sampling times to be used in more definitive animal PK experiments. The plasma concentration versus time profiles after PO dosing provide preliminary information on the absorption kinetics and the absolute bioavailability of the lead. The design of additional animal PK studies depends on the results of the preliminary animal PK study, the theoretic kinetic profile needed to produce the desired pharmacology response, and the results from preliminary toxicology experiments (discussed below).

For most drug development programs, toxicology studies in two or more species are necessary. In this case, preliminary animal PK studies should be conducted in each species projected for use in animal safety studies. If differences in delivery or disposition exist between species and result in an enhanced or decreased toxicology profile, the PK profiles may explain, in part, the different toxicology profiles. If possible, physiological fluid specimens should be obtained from animals in the preliminary toxicology studies to determine the extent and uniformity of exposure, which is termed toxicokinetics (TK) over the evaluated dose range. Normally, three or four specimens from each animal are sufficient for TK evaluations, but this level of sampling may not be possible for all studies. A single specimen at one collection time can be obtained from one or two animals in a dose group, and the other animals in that dose group can be sampled at other times. Analyses of these specimens provide data on the extent of exposure but not on the uniformity of exposure within a dose group. For multiple-dose studies, specimens are usually obtained after the first dose and after the last, or next to last, dose. The results provide information on possible changes in exposure and on the accumulation potential of the lead and can be used to design multiple-dose animal PK and other development studies. If the change in disposition or accumulation is substantial, modification of the dosing regimen may be necessary to obtain the desired concentration profile after dosing to steady state.

## Drug Metabolism Assays

The number and design of drug metabolism (DM) studies needed to characterize the fate of a lead in the body depend on the results from preliminary animal PK and toxicology studies. Commonly, the results from these in vivo experiments are not available during discovery, and developability assessments and in vitro DM assays are utilized to determine the metabolic stability and/or the extent of metabolism of a lead and to compare the extent of metabolism among various species, including humans. These in vitro experiments can be conducted in a variety of assay systems, including CYP450 isozymes (the enzymes responsible for most oxidative metabolism of drugs),

microsomes, hepatocytes, or liver slices. Since hepatocytes contain both phase 1 (oxidative, hydrolysis, and reduction) and phase 2 (conjugation) metabolism systems and can be relatively easily obtained from pharmacology and toxicology animal species and from human donors, many researchers select this model for the first assessment of metabolism. If the results from hepatocytes show extensive metabolism, additional in vitro experiments are usually conducted first in microsomes to ascertain if oxidative metabolism is present and then in isolated CYP450 isozymes to determine which enzyme, or enzymes, is responsible. Extensive metabolism is not necessarily a death knell for a discovery lead. If rapid clearance from the body is a desired developability attribute for effectively treating a disease indication, rapid metabolism to inactive metabolites may be advantageous. However, for most disease indications, extensive metabolism may prevent delivery of a pharmacologically active substance to the site of biological action in sufficient concentration to produce the desired response. Thus, a lead that is extensively metabolized may not be a successful drug candidate.

Another reason for conducting in vitro metabolism studies early is to determine if species differences are present. Evaluating metabolism in the pharmacological and proposed toxicological animal species and in humans assists in selecting the species that are similar, at least in metabolism, to humans for definitive toxicology studies. If an animal species has limited metabolism while humans may have extensive metabolism, pharmacological and/or toxicological metabolites may be responsible for some, or all, of the biological activity or adverse effects in humans, and these responses would not be observed in the animal model. Conversely, if an animal species has extensive or different metabolism compared to humans, the pharmacological and/or toxicological profiles in that species would probably not be predictive of efficacy and/or safety in humans.

If desired, which is sometimes the case when metabolism is extensive, the metabolites generated from in vitro systems can be isolated and identified. After preparation of sufficient quantities for additional testing, these metabolites can be evaluated for pharmacological and/or toxicological potential. This author, like many drug discovery and development researchers, has found metabolites with equal or greater biological activity when compared to the parent compound. At times, these pharmacologically activity metabolites have more druglike attributes than the parent and can be developed either as a replacement of the parent or as a second-generation drug candidate.

## Toxicology Assays

Toxicology studies or assays are conducted to define the safety profile of a compound and include definition of the NOAEL, maximum tolerated dose or MTD, potential organs of toxicity, and potential biochemical markers to detect and track toxic events. Most discovery leads that do not become therapeutic products have unacceptable toxicity in animals and/or humans. Before the

definitive toxicology studies needed to support submission for first-in-human testing are initiated, a number of in vitro and animal experiments can be conducted to characterize the potential toxicity of a discovery lead. These early toxicology evaluations are usually conducted in the same species as used in pharmacology evaluations. As mentioned earlier, the lowest dose that has no toxicity, or an acceptable level of toxicity, is compared with the dose that gives the desired pharmacological response in the same animal species to obtain a TR or TI for that species.

A toxicology program to obtain toxicological characterization of a discovery lead should be accomplished through close interaction with the efforts of other scientists conducting other developability experiments. Before drug safety studies are conducted, a sufficient quantity of the lead should be available and characterized, using appropriately characterized analytical chemistry assays, so that testing is conducted with a known compound. If the lead requires formulation before dosing, the formulation should be the same for each study. If a change in the synthesis, purification, or formulation is necessary to improve the biopharmaceutical properties of the lead or the drug delivery profile, then some of the early toxicology studies should be repeated with the new formulation to determine whether the safety profile has been altered. These early safety studies do not need to be, but in many cases are, conducted in compliance with Good Laboratory Practice (GLP) regulations. However, these experiments should be designed and conducted as close as possible to the processes used for definitive GLP toxicology studies. Then, the results will be scientifically defensible and useful in predicting the toxicity expected from the GLP studies. Examples of the early toxicology studies or safety assays needed to characterize a lead include the following.

***In Vitro Toxicology*** When a number of discovery leads has been identified and needs to be further evaluated to select the optimal lead for further evaluations, the potential for toxic effects may be determined using in vitro assays, such as cell-based systems or microarrays. By incubating various concentrations of the leads with cells, such as the pharmacological target cell or hepatocytes, and measuring an adverse effect, such as cell death or change in cell function, the compounds can be stratified as to toxicological potential. Since nephrotoxicity is second only to hepatotoxicity as a major problem associated with adverse experiences during clinical trials, both hepatocytes and primary human kidney cells are commonly used in in vitro toxicology evaluations. Similarly, microarrays that have systems considered predictive of toxic events can be used to determine which leads "turn on" or "turn off" these systems after a lead has been incubated with cells or administered to an animal model and appropriate samples collected. Defining the appropriate samples and when to collect them can be difficult. If the wrong samples are collected or the right samples are collected at the wrong times, the results may not be predictive of the toxicological profile of the discovery lead(s). While most, if not all, toxicologists think these in vitro systems cannot be used to predict toxicology in

animal models or humans, the results may be useful in evaluating a group of discovery leads to determine which lead may have a more acceptable profile compared to the others.

An alternate approach for an in vitro toxicological assessment is to use a biological marker that is considered predictive of toxic events. One possibility for such a marker is glutathione S-transferases (GSTs), which are important phase 2 drug-metabolizing enzymes. GSTs are found in high concentrations in several organs of key toxicological importance, such as the kidney and liver, and GST expression is induced in the presence of many toxic substances. GSTs are rapidly and readily released in the event of toxicological insult, thus making GSTs very sensitive biomarkers of organ damage.

Discovery leads are incubated with hepatocytes from various species. After incubation and centrifugation to remove the cells, the supernatants are assayed for GST content using already available and characterized assays. If appropriately designed, in vitro metabolism studies conducted with hepatocytes could be the source of these samples. Since hepatocytes contain both phase 1 and phase 2 metabolism systems, this in vitro toxicological assessment provides information not only about the safety profile of a discovery lead but also on that of its metabolites. The concentration range evaluated should include a level near the projected pharmacologically active level, a level at 5 to 10 times the projected pharmacologically active level, and a level near the maximum solubility of the lead (determined during solubility assessments).

***In Vivo Toxicology*** To evaluate the qualitative and quantitative single-dose toxicity of a drug candidate, a single dose at a number of dose levels is administered by the proposed clinical route and the animals are observed for 14 days after dosing. The acute study is not an $LD_{50}$ (dose that is lethal to 50 percent of test species) study, which is not needed for overall risk assessment according to an ICH guideline. This ICH guideline suggests that the drug candidate dose levels include at least one that produces pharmacological activity and one that causes overt evidence of major or life-threatening toxicity and that a vehicle control group is included. The acute toxicity study should evaluate both the IV route and the intended clinical route of administration, unless the clinical route is IV. The studies are usually conducted in two relevant mammalian species, one of which is a nonrodent, and unless scientifically unjustifiable (i.e., a gender-specific human disease or disorder), should use equal numbers of males and females for each species evaluated. The test species is observed for 14 days after dosing and, as with all toxicology studies, all signs of toxicity with time of onset, duration of symptoms, and reversibility are recorded. Also, the time to first observations of lethality is recorded. Gross necropsies are performed on all animals sacrificed moribund, found dead, or terminated after 14 days of observation and the results are presented by group. An evaluation of results should include all observations made and a discussion of the toxicological findings and their implications to humans, taking into

account the pharmacology of the lead, the proposed human therapeutic use and dose, and experience with related drugs. The highest no-toxic-effect dose and the highest nonlethal dose are noted. A similar design, without the 14-day observation period, can be employed to evaluate the potential toxicology profile of a lead or a group of leads.

Organs or tissues that are the targets of toxicity may be identified during a full histology workup of animals in each dose group and from the results obtained from the analyses of clinical chemistry samples. However, during developability assessment, a full histology assessment may not be necessary and a selected set of organs or tissues (i.e., those considered most likely to be organs of toxicity such as the liver, kidney, lung, heart, brain) may be evaluated. If possible, investigations into the biochemical mechanism of identified toxicity should be initiated. Results from these experiments can provide insight into potential toxicities for a lead or a class of leads in definitive toxicology studies, identify biological markers that predict a toxic event, and suggest conditions in human patients where administration of a drug candidate is contraindicated. If results from the early toxicology studies show that a lead has an unacceptable level of toxicity, the development candidacy of such a compound should be carefully considered.

If desired, the safety of a lead can be further assessed by conducting safety pharmacology and/or genotoxicity studies. These studies, which are to be completed prior to the initiation of human clinical testing, are more commonly conducted after selection of a drug candidate. However, if some discovery leads are considered "equal" after other developability assessments have been completed, the results from safety pharmacology or genotoxicity studies may be able to identify the "optimal" lead or determine that some leads do not have the desired profile and should not be selected as the drug candidate.

## 4 CONCLUSIONS

During drug discovery and developability assessment, many biological and chemical assays are needed for evaluating the biological potency of compounds and for characterizing the attributes and demerits of discovery leads. Generally, two different assay set types, like the two faces of a coin, are necessary. On one side, the discovery side, assays need to be defined and/or developed to evaluate thousands of compounds quickly and to determine which, if any, of these compounds have the ability to antagonize or agonize a target that is thought to play an important role in a human disease process or disorder. Pharmacology, both in vitro and in vivo, and analytical chemistry assays used to support this activity need to be designed with this goal in mind.

On the other side or face of the coin are developability assessment assays that are used to evaluate whether or not a discovery lead, or one of a group of leads, has the desired attributes considered necessary for successful devel-

opment and does not have a major demerit that may be predictive of development problems.

In this author's opinion, many discovery leads are transferred to the preclinical development process with insufficient characterization to assess their development potential. This lack of knowledge usually results in poorly designed experiments that are not data productive and that, in many cases, have to be repeated when the drug candidate shows unexpected toxicity, low and variable delivery, instability or solubility problems, or unacceptable pharmacokinetic and drug metabolism profiles. In all too many cases, the recognition of these problem areas results in the termination of development for a potentially useful therapeutic agent. At best, the problems encountered cause a delay, at times substantial, in the development of a candidate while additional studies are conducted to elucidate the causes of the problems and minimize their effect. Then, the definitive development experiments have to be repeated.

The developability experiments, with their associated assays, can more fully characterize a discovery lead before the compound enters the definitive preclinical, then the nonclinical and clinical, drug development processes. The experimental designs and assays could also be used, with minor modifications, to evaluate a group of discovery leads with acceptable biological potency and thus to select the lead with the best characteristics, that is, the most druglike attributes and the least demerits, for further development.

If these developability experiments are completed as part of the transition from discovery to development, discovery leads that do not have the characteristics necessary to become therapeutic agents can be identified early and prevented (i.e., the fail early, fail often approach) from entering the development process. Analogs of a lead with unacceptable characteristics can be evaluated to find a development candidate that has more optimal properties. In addition, the results from the developability studies will allow the preclinical development studies to be designed and conducted in a timely, cost-efficient manner and thus most likely allow the candidate to have an earlier entry into the clinic.

# STRATEGIES AND METHODS IN MONITORING AND TARGETING PROTEIN–PROTEIN INTERACTIONS

Arianna Loregian and Giorgio Palù
*University of Padova*
*Padova, Italy*

1 INTRODUCTION
Importance of Studying Protein–Protein Interactions
2 METHODS FOR THE STUDY OF PROTEIN–PROTEIN INTERACTIONS
Two-Hybrid System
Peptide Display Technologies
Protein Microassays
Protein Mass Spectrometry
Computational Methods of Analysis of Protein–Protein Interactions
3 DISCOVERY OF INHIBITORS OF PROTEIN–PROTEIN INTERACTIONS
Two-Hybrid System Variants
Other in Vivo Genetic Selection Systems
Inhibitors from Phage Display
Microarrays and Small Molecules
References

## 1 INTRODUCTION

### Importance of Studying Protein–Protein Interactions

Protein–protein interactions are intrinsic to virtually every cellular process ranging from cell cycle control, DNA (deoxyribonucleic acid) replication,

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

transcription, splicing, and translation, to intermediary metabolism, secretion, formation of cellular macrostructures and enzymatic complexes. Apart from a variety of stable protein–protein interactions, there are a plethora of transient protein–protein interactions that control and regulate a large number of cellular processes. All modifications of proteins involve such transient protein–protein interactions. Indeed, glycosyl transferases, acyl transferases, kinases, phosphatases, and proteases interact only transiently with their protein substrates. Such protein-modifying enzymes encompass a large number of fundamental processes such as signal transduction, cell growth, and metabolic pathways.

Since protein–protein interactions play a role in nearly all events that take place in a cell, information on the function of an unknown protein can be obtained by investigating its interaction with other proteins whose functions are already known. Thus, if the function of one protein is known, then the function of its binding partner is likely to be related. This concept has been called "guilt by association" and allows the researcher to employ a relatively small number of functionally characterized proteins and to quickly assign functions to their uncharacterized binding partners.

Moreover, alteration of protein–protein interactions is known to contribute to many diseases. As an example, tumor-forming viruses cause uncontrolled proliferation of the host cell by dissociating important protein–protein interactions between regulatory proteins of the cell cycle [1]. Hence, the manipulation of protein–protein interactions that contribute to disease is a potential therapeutic strategy. The contact surfaces of the protein complexes have unique structure and properties, and they are more conservative in comparison with the active site of enzymes. So they represent prospective targets for a new generation of drugs. During the last decade, numerous investigations were undertaken to find or design small molecules that block protein dimerization or heterologous protein–protein interactions [2].

To date, a variety of genetic and biochemical methods exist for studying protein–protein interactions and identifying inhibitors of such interactions. This chapter describes recent developments in proteomic research. In more detail, the first part of this chapter focuses on technologies recently developed in protein interaction investigation, that is, yeast two-hybrid screens, phage display, protein microarray technology, two-dimensional electrophoresis coupled to mass spectroscopy, and so forth. Different strategies are compared; problems that are encountered in studying protein–protein interactions, solutions to these problems, and advantages and limitations of various methods and techniques are also discussed.

The second part presents recent approaches to identify and characterize new inhibitory molecules that act by disrupting biologically relevant protein–protein interactions. In particular, screening strategies that employ variants of the technologies reviewed in the first part of this chapter are discussed.

## 2 METHODS FOR THE STUDY OF PROTEIN–PROTEIN INTERACTIONS

### Two-Hybrid System

***Principle of Yeast Two-Hybrid System*** The yeast two-hybrid system was originally developed by Fields and Song as a genetic assay to detect protein–protein interactions in a cellular setting [3]. It takes advantage of the finding that many eukaryotic transcription activators have at least two distinct functional domains, one that directs binding to a promoter DNA sequence and one that activates transcription [4, 5]. This fact was demonstrated by exchanging DNA binding domains and activation domains from one transcription factor to another while retaining its function. For example, it was shown that the activation domain of yeast GAL4 protein could be fused to the DNA-binding domain of *Escherichia coli* LexA to create a functional hybrid transcription activator in yeast [6].

In the classical yeast two-hybrid approach, a *bait* is constructed by fusing a protein X to the DNA binding domain (DBD) derived from a transcription factor and a *prey* is created by fusing a protein Y to the activation domain (AD) of a transcription factor. The bait and prey fusions are coexpressed in yeast, where the interaction of X and Y leads to the reconstitution of a functional transcription factor and induces the expression of a reporter gene(s) integrated in the region downstream of the DBD binding sites (Fig. 1). Commonly, auxotrophic markers (e.g., *HIS3* and *LEU2*) that can be selected for are used as reporter genes in combination with the *lacZ* gene encoding bacterial β-galactosidase. *LEU2* and *HIS3* allow selection of positive interactions by monitoring growth on selective plates lacking leucine or histidine, respectively, whereas *lacZ* expression can be easily measured using a colorimetric assay. Thus, the yeast two-hybrid system is devised to identify genes encoding proteins that physically interact with a given protein in vivo. The technique enables not only the characterization of known interaction couples but also the identification of new interacting partners and even embodies the technological means to manipulate protein–protein interactions.

The yeast two-hybrid system presents a number of major advantages over alternative biochemical and genetic approaches. First of all, it relies on an assay performed in vivo (in the yeast host cell), and thus it is not affected by the artificial conditions of in vitro assays. Second, it allows very high numbers of coding sequences to be assayed in a relatively simple experiment. In fact, an appealing feature of yeast two-hybrid system is the minimal requirements to start a screening. Only the complementary DNA (cDNA), full-length or even partial, of the gene of interest is required, in contrast to sometimes high quantities of purified proteins or good-quality antibodies needed in classical biochemical approaches. In addition, one of the major advantages of this system is that the identification of an interacting protein implies that at the same time the corresponding gene is cloned.

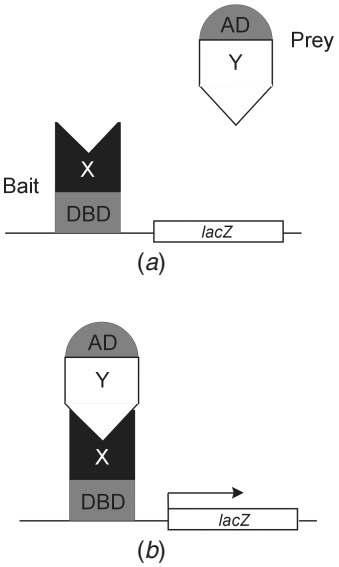


**Figure 1** Principle of the yeast two-hybrid system. (*a*) Two chimeras, one (Bait) containing the DNA binding domain (DBD) fused to a protein X, and one (Prey) containing an activation domain (AD) fused to a protein Y, are cotransfected into an appropriate host strain. (*b*) If X and Y interact, the DBD and AD are brought into proximity and can activate transcription of a reporter gene(s) (here *lacZ*).

Weak and transient interactions, often the most interesting, can be more easily detected in two-hybrid assays since the genetic reporter gene strategy results in a significant amplification. By doing semiquantitative experiments, one can even interpret affinities from two-hybrid experiments. An exhaustive study of the sensitivity of the two-hybrid system with regard to the affinity of the protein–protein interaction has demonstrated that the strength of interaction, as predicted by the two-hybrid approach, usually correlates with that determined in vitro, allowing discrimination of low-, intermediate-, and high-affinity interactions [7]. The GAL4-based two-hybrid system can detect protein–protein interactions with a dissociation constant ($K_d$) of $10^{-6}$ M [8].

Since the emergence of the two-hybrid approach in 1989, a number of improvements have been developed that have increased its applicability. The two-hybrid system was initially predicted to be limited to the analysis of cytoplasmic proteins, as extracellular proteins are often *N*-glycosylated and contain disulfide bonds, both of which do not occur in the yeast nucleus [9]. However, several successes were reported with transmembrane receptors. Appropriate extracellular receptor–ligand interactions were demonstrated for the growth hormone, prolactin and growth-hormone-releasing receptors [10, 11]. Proteins of length as small as 8 to 10 residues [12] and as large as 778

residues [13] have been investigated in a two-hybrid study. Hydrophobic domains may affect expression but may be less problematic when expressed as a small percentage of the overall protein and when contained within the protein itself.

It should be noted, however, that the two-hybrid system cannot be used for all protein–protein interaction studies. For several experimental reasons, some proteins are not suited for this approach. Limitations and drawbacks of the two-hybrid system are listed in the next paragraphs.

First of all, a potential risk regards the use of artificially made fusion proteins. The fusion might change the actual conformation of the bait and/or prey and consequently alter its functionality. This misfolding might result in a limited activity or in the inaccessibility of binding sites. The best way to check the correct conformation of the protein of interest is to test its interaction with a known binding partner. This will only work if both fusion proteins are folded properly. Another way to circumvent the problem could be the reciprocal transfer of proteins, that is, switching the protein from the DBD fusion to the AD fusion and vice versa.

Since the readout of the two-hybrid system makes use of a transcriptional event, one of the most crucial initial experiments is to test whether the bait protein is able to activate transcription on its own. This can be a problem with transcriptional activators that naturally contain domains that activate the reporter genes when fused to the DBD. In addition, many proteins other than transcription factors have been shown to activate transcription when artificially fused to DBD [14]. Both classes of proteins are referred as *self-activators*. When dealing with self-activators, several approaches can be undertaken. First, the expression level of the DBD-X fusion can be decreased by using centromeric vectors and/or weaker promoters. Second, when using *HIS3* as a reporter gene, the concentration of 3-aminotriazole (3AT) can be increased to elevate the growth threshold of the host strain. The drug 3AT acts as a competitive inhibitor of the *HIS3*-encoded enzyme [8]. Under these conditions, it is expected that, even though the DBD-X fusion activates *HIS3* transcription to some extent, the DBD-X/AD-Y interaction leads to more *HIS3* expression to overcome the growth threshold imposed by 3AT [15]. However, these two strategies are not always successful, especially for strong self-activators. In this case, a "switched" two-hybrid system can be used [16], wherein the bait protein is fused to the activation domain (AD-X) and a DBD-Y library is screened.

One of the most ambiguous drawbacks is that the two-hybrid system makes use of yeast, *Saccharomyces cerevisiae*, as a host. The use of yeast can be seen as an advantage since yeast is closer to higher eukaryotics than those systems based on bacterial hosts or in vitro experiments. However, some interactions depend upon posttranslational modifications that do not, or inappropriately, occur in yeast. Such modifications are frequent and include phosphorylation, some types of glycosylation, and the formation of disulfide bridges. Some new versions of the two-hybrid system, however, try to overcome this problem by

coexpressing the enzyme responsible for the posttranslational modification. Several plasmids have been designed to allow conditional expression of such a "third" partner. In addition to performing posttranslational modifications of one or both of the two interacting proteins, this third protein can directly contribute to the formation of a trimeric complex [17]. For instance, it has been shown that a tyrosine phosphorylation-dependent interaction could be detected when the corresponding kinase was coexpressed in yeast cells [18]. These different variations that involve third partners as native proteins, in the absence of any fused domains, are referred to as *tri-brid systems*.

Another limitation of the two-hybrid system is that both the DBD-X and the AD-Y fusions need to be translocated into the yeast nucleus. Thus, some two-hybrid vectors encode DBD and AD fused to a nuclear localization signal (NLS) to target the fusion proteins to the nucleus. However, proper localization can represent a major difficulty, especially when dealing with membrane-anchored proteins, extracellular proteins, or proteins that contain strong targeting signals. For this reason, some variations of the system have been designed for those proteins that are not active in the yeast nucleus, such as the *ubiquitin-based split-protein sensor* (USPS) system [19] and the *Sos recruitment system* (SRS) [20].

In addition, some proteins might become toxic upon expression in yeast. A number of proteins, such as homeobox gene products or cyclins are indeed toxic when expressed into the yeast nucleus. The use of an inducible promoter might circumvent the problem. Other proteins might proteolyze essential yeast proteins or the DBD and AD fusions.

It should also be kept in mind that it is impossible to exclude the possibility that a third protein Z is bridging the two interacting partners, and therefore a two-hybrid screen does not necessarily identify direct binary interactions. For example, some DBD-X fusions (e.g., DBD-lamin) are capable of allowing activation by the AD-Y fusions independent of a direct contact between X and Y. In addition, indirect interactions have been reported where an endogenous yeast protein acts as a bridge [21]. Nevertheless, the detection of such indirect protein–protein interactions can still be considered as an indication of a potential functional link between two proteins.

Finally, one of the major concerns for the yeast two-hybrid system are the so-called false positives, which actually include two different categories, namely technical and biological ones. The technical false positive is an apparent two-hybrid interaction that does not rely upon the assembly of two-hybrid proteins. Since the screen is based on the transcriptional activation of reporter genes, any event leading to an increase in the rate of transcription might be misinterpreted as a positive DBD-X/AD-Y interaction. The biological false positive means a bona fide two-hybrid interaction with no physiological relevance. Due to the so-called time/space constraints, it is potentially possible that two proteins, although able to interact, are never in close proximity to each other within the cell. The two proteins could be expressed in different cell types, or, even when present in the same cell, they could be

localized in different intracellular compartments. Moreover, interacting proteins can be expressed at different times during embryogenesis or during homeostasis (e.g., at different points in the cell cycle). So once two interacting partners are selected, the biological significance of this interaction has to be investigated.

Most recently, modified two-hybrid strategies have been designed to increase the specificity in order to limit the above problems of false positives. Several strategies have been proposed. First, multiple reporter genes were designed for which the corresponding promoters are unrelated, containing different initiation sites and TATA boxes. Since many false positives of the two-hybrid method are promoter context dependent, the specificity of the assay is increased by testing the phenotypes conferred by different reporter genes in the same cell [22, 23]. Second, the expression level of the two fusion proteins was reduced by using truncated versions of the alcohol dehydrogenase 1 (ADH1) promoter and centromeric vectors, which are maintained at low copies in the yeast [24]. Third, mating strategies were introduced that enable parallel screening with many different baits [25, 26].

The quality of a two-hybrid screen can also be evaluated in terms of the number of expected interactions that were not detected. These are often referred to as *false negatives*. In many cases, the reasons for the lack of recovery of expected interactions are unknown, but several possibilities can be speculated. For example, the instability and/or the improper folding of a DBD-X or AD-Y fusion protein could affect its transcriptional activity. Alternatively, particular fusion proteins might be toxic and affect the viability of the host cells. In addition, the gene encoding an interacting protein may simply not be represented in the library due to a low complexity of the library or a bias in the representation of that clone.

Despite these and other limitations, the power of the yeast two-hybrid system is so enormous that it has quickly become the most frequently used assay to detect novel protein–protein interactions. A recent publication estimates that more than 50 percent of all interactions described in the literature have been detected using the yeast two-hybrid system [27].

***Reagents*** Several comparable versions of yeast two-hybrid system are available. All systems share common elements. All use (1) a plasmid that directs the synthesis of a *bait*, that is, a known protein that is brought to DNA by being fused to a DBD, (2) a plasmid that directs the synthesis of a *prey(s)*, that is, a protein(s) fused to an AD and eventually other useful moieties, and finally (3) one or more reporter genes (*reporters*) with upstream binding sites for the bait. The versions differ in their specifics, and these details can be relevant to their successful use. The following paragraphs provide an overview of the most common yeast two-hybrid system reagents.

*Target and Bait Vectors* When performing a two-hybrid screen, the first decision to be made is the choice of the most appropriate vector system. A large

number of different DBD- and AD-containing vectors are available. A list of the most commonly used plasmids in the yeast two-hybrid system together with their peculiar features is reported in Table 1. In addition, Parent et al. [28] have compiled an extensive list of yeast cloning vectors.

The most extensively used DBD and AD vectors are GAL4-based, probably because they were the first commercially available. The GAL4-based DBD vectors contain a portion of the yeast GAL4 protein [8, 29, 30]. This portion, encoded by residues 1 to 147, is sufficient to bind tightly to appropriate DNA binding sites, directs dimerization, and localizes fused proteins to the nucleus; it also contains a domain that weakly activates transcription from mammalian cell extracts in vitro, and it is thus conceivable that this domain may increase transcription resulting from weakly interacting proteins. The GAL4-based AD vectors usually contain the GAL4 activation domain II (aa 768–881).

Alternative systems make use of the DBD of the bacterial LexA protein and the AD of VP16 or the so-called B42AD. In LexA-based two-hybrid systems, the DBD is provided by the entire *E. coli* LexA protein [31, 32]. LexA carriers a dimerization domain at its C-terminus [33] and normally functions as a repressor of SOS genes in *E. coli* by binding LexA operator sequences that are part of the promoter [34]. When used in the yeast two-hybrid system, the LexA protein does not act as a repressor because the LexA operators are integrated upstream of the minimal promoter and coding region of the reporter gene. In yeast, LexA and most LexA derivatives enter the nucleus but are not necessarily nuclear localized. Both systems, GAL4 and LexA, have advantages and drawbacks that make the choice quite difficult. Since the LexA- and the GAL4-based two-hybrid systems have different properties, it is not unreasonable to imagine that some interactions might be detected differently in the two systems. Trying both will increase the chance of success.

Activation domain vectors can differ not only in the activation domain they carry but also in whether they contain other useful moieties such as epitopes tags and/or nuclear localization sequences [8, 29, 31, 32]. Some AD vectors, for example, pACT2, encode a hemaglutinin (HA) epitope tag (YPYDVPDYA in single-letter code), fused in frame to the activation domain [35]. This allows the protein to be detected with commercially available anti-HA antibodies in Western blot analysis. Some ADs are stronger than others. Although strong ADs should allow detection of weaker interactions, their expression can also be toxic to the cell due to poorly understood effects, either by subtraction of cofactors needed for transcription of other genes or by toxic effects that manifest when strong ADs are brought to DNA. Thus, it is possible that strong ADs may prevent detection of some interactions. AD fusion proteins also differ in whether they are expressed conditionally [31] or constitutively [8, 29]. Conditional expression allows the transcription phenotypes obtained in selections for interactors to be ascribed to the synthesis of the fusion protein, thus reducing the number of false-positive cells that grow because their reporters are aberrantly transcribed.

**TABLE 1 Overview of Most Commonly Used Two-Hybrid Vectors**

| Plasmid Name | Selection Marker: In Yeast | Selection Marker: In *E. coli* | Functional Domain | Promoter | Comments |
|---|---|---|---|---|---|
| *LexA-based Plasmids* | | | | | |
| pBTM116 | *TRP1* | Ap | LexA | ADH1 (truncated) | Basic plasmid used for cloning bait |
| pEG202 | *HIS3* | Ap | LexA | ADH1 (full length) | Basic plasmid used for cloning bait |
| pJK202 | *HIS3* | Ap | LexA + NLS | ADH1 (full length) | Same as pEG202 but incorporates an NLS between LexA and polylinker; used to ensure nuclear translocation of bait |
| pNLexA | *HIS3* | Ap | LexA | | ADH promoter expresses polylinker followed by LexA; used with baits where N-terminal residues must remain unblocked |
| pGilda | *HIS3* | Ap | LexA | GAL1 (full length), inducible promoter, centromeric vector | Used for baits whose continuous presence is toxic |
| pEE2021 | *HIS3* | Ap | LexA | | An intergrating form of pEG202 that can be targeted in *HIS3* following digestion with *Kpn*I; used when lower expression levels of baits are required |
| pMW101 | *HIS3* | Cm | LexA | ADH1 (full length) | Like pEG202, but with different antibiotic resistance marker |
| pMW103 | *HIS3* | Km | LexA | ADH1 (full length) | Like pEG202, but with different antibiotic resistance marker |

**TABLE 1** ***Continued***

| Plasmid Name | Selection Marker: In Yeast | Selection Marker: In *E. coli* | Functional Domain | Promoter | Comments |
|---|---|---|---|---|---|
| pHybLex/Zeo | Zeocin | Zeocin | LexA | ADH1 (truncated) | Bait cloning vector with Zeocin marker; compatible with all two-hybrid systems |
| *GAL4 DBD-based Plasmids* | | | | | |
| pMA424 | *HIS3* | Ap | GAL4 DBD | Original vector, 12 kb | Basic plasmid used for cloning bait |
| pGBT9 | *TRP1* | Ap | GAL4 DBD | ADH1 (truncated) | |
| pAS1 | *TRP1* | Ap | GAL4 DBD + HA | ADH1 (full length), CYH2 | Incorporates a HA epitope for detection of baits |
| pAS2 | *TRP1* | Ap | GAL4 DBD + HA | ADH1 (full length), CYH2 | Incorporates a HA epitope for detection of baits |
| pAS2-1 | *TRP1* | Ap | GAL4 DBD | ADH1 (full length), CYH2 | |
| *GAL4 AD-based Plasmids* | | | | | |
| pGAD2F | *LEU2* | Ap | GAL4 AD | Original vector, 13 kb | Basic plasmid used for cloning prey |
| pGAD424 | *LEU2* | Ap | GAL4 AD | ADH1 (truncated) | Basic plasmid used for cloning prey |
| pGAD10 | *LEU2* | Ap | GAL4 AD | ADH1 (truncated) | Basic plasmid used for cloning prey |
| pJG4-5 | *TRP1* | Ap | GAL4 AD + NLS + HA | GAL1 | GAL1 promoter expresses GAL4 AD fused to NLS and HA; used to express cDNA libraries |

| | | | | | |
|---|---|---|---|---|---|
| pJG4-5I | *TRP1* | Ap | GAL4 AD + NLS + HA | GAL1 | An integrating form of pJG4-5 that can be targeted in *TRP1* following digestion with *Bsu*36I; used when lower expression levels of preys are required |
| pMW102 | *TRP1* | Km | GAL4 AD + NLS + HA | GAL1 | Like pJG4-5, but with different antibiotic resistance marker |
| pMW104 | *TRP1* | Cm | GAL4 AD + NLS + HA | GAL1 | Like pJG4-5, but with different antibiotic resistance marker |
| pACT1 | *LEU2* | Ap | GAL4 AD | | |
| pACT2 | *LEU2* | Ap | GAL4 AD + HA | ADH1 (truncated), medium expression | Incorporates a HA epitope for detection of preys |
| *Others* | | | | | |
| pB42AD | *TRP1* | Ap | B42 + SV40 NLS + HA | GAL1 (full length), inducible promoter | GAL1 promoter expresses B42AD, SV40 NLS, HA epitope, polylinker |
| pYESTrp | *TRP1* | Ap | V5 epitope + SV40 NLS + B42 | GAL1 (full length), inducible promoter | GAL1 promoter expresses SV40 NLS, B42AD, V5 epitope, polylinker; contains f1 ori and T7 promoter/flanking site; used to express cDNA libraries |
| pSD-10 | *URA3* | Ap | VP16AD | | |

A key feature to be considered when choosing a vector for a successful screening is the promoter that regulates the expression level of the hybrid proteins. The 1500-bp full-length ADH1 promoter, which normally drives the expression of the metabolic enzyme alcohol dehydrogenase 1, allows high-level expression of sequences under its control. Expression from this promoter is maximal during logarithmic growth of the yeast cells and becomes repressed in late log phase by ethanol accumulation in the medium. Instead of this full-length promoter, some two-hybrid vectors, for example, pGAD424 and pGBT9, contain a truncated 410-bp ADH1 promoter. Expression from this promoter leads to low or very low levels of fusion protein expression [36]. The choice of the expression vector might be also influenced by the nature of the target. If the target is expected to interfere with the endogenous yeast metabolism, a lower expression might be beneficial. However, if the expression of the fusion protein needs to be assayed (e.g., by Western blot analysis), a higher expression level is more convenient.

The expression level depends not only on the promoter strength but also on the copy number of the plasmid. In most commonly used two-hybrid plasmids, the origin of replication is the 2μ origin. The 2μ plasmids are maintained stable and at high copy numbers (50 to 100 copies per cell) in yeast and function solely for their own replication [37]. It should be noted that in the context of a reverse two-hybrid system (see Reverse Two-Hybrid System), the expression levels of the fusion proteins should be maintained as low as possible since "background" interactions, which are more likely to occur at high protein expression levels, will kill the yeast cell. Therefore, the vectors used in the reverse two-hybrid systems make use of low-copy, centromeric vectors. Also when using toxic proteins, the use of centromeric plasmids in forward two-hybrid screens can be most helpful.

Finally, other characteristics that might influence the choice of a vector are of a more practical nature. In all vector systems care must be taken to maintain the proper reading frame when creating the fusion proteins. Thus, an important parameter to be considered when choosing a vector is the compatibility of the multiple cloning site. Another consideration concerns the marker gene(s). The most commonly used markers in yeast are genes encoding amino acid biosynthetic enzymes. Most two-hybrid yeast strains have a lesion in either *TRP1*, *LEU2*, *HIS3*, *ADE2*, and/or *URA3*, which enables selection for cells that were transformed with plasmids that carry the corresponding gene by growth in the absence of the appropriate amino acid(s). Alternatively, in a recently introduced vector (pHybLex/Zeo) zeocin is used as a selection marker.

*Reporter Genes and Host Yeast Strains* Reporter genes differ in the phenotypes they confer. Positive selection markers, such as *lacZ*, are widely used and provide colorimetric readout after an enzymatic assay. In the alternative, prototrophic markers (*LEU2* or *HIS3*) can be used and enable positive selection on media deficient for the specific amino acid encoded by the reporter gene,

wherein a two-hybrid interaction results in cell growth. Reporter genes resulting in a sensitivity phenotype, such as G418 [38], *CYH2* [39], or *CAN1* [40], allow counter selection, wherein a productive protein–protein interaction abrogates cell growth (see Reverse Two-Hybrid System). In this case, yeast growth selection can be utilized to select mutations, proteins, peptides, or small molecules that dissociate target protein–protein interactions.

Many two-hybrid systems commonly use dual reporter systems to confirm the identification of a protein–protein interaction and/or putative novel binding partners isolated via a screen. For example, after an initial genetic selection based on growth assays, a secondary screen is performed with a second and different reporter gene such as *lacZ* to allow higher specificity [8, 31, 32]. This strategy allows one to address the question of the large numbers of transformants that need to be screened to properly survey the complexity of cDNA libraries.

Reporters also differ in the number and affinity of upstream binding sites for the bait and in the position of these sites relative to the transcription start point [31]. In the GAL4-based two-hybrid system, either the intact GAL1 upstream activating sequence (UAS), which contains four GAL4 binding sites, or an artificially constructed UAS consisting of three copies of the 17-mer consensus binding sequence is used. To avoid interference by endogenous GAL4 and GAL80 proteins, the yeast host strains used in the GAL4-based two-hybrid system must carry deletions both of the *GAL4* and *GAL80* genes. Due to the absence of these two proteins, the yeast cells grow more slowly as compared to the wild-type yeast. The use of bacterial LexA overcomes this problem. In LexA-based two-hybrid systems, expression of the reporter gene is under the control of six to eight copies of the LexA operator sequence and a minimal promoter.

Reporter genes can be integrated into the genome or reside on a plasmid. The disadvantage of having another plasmid containing the reporter gene, and thus the need for an additional auxotrophic marker, is compensated by several advantages. One of the major advantages of having the *lacZ* reporter gene on a high-copy-number plasmid is that weak signals are greatly amplified, which makes it possible to assay β-galactosidase activity directly on the selection plate by including X-gal in the medium. This avoids tedious replica and/or filter lift assays.

Finally, reporter genes differ in the number of molecules of the reporter gene product necessary to score the phenotype. These differences affect the strength of the protein–protein interactions that the system can detect. A list of the most commonly used yeast strains and reporter genes is given in Table 2.

*Library Choice* The source of DNA for the library is a key parameter for a successful screen. It is always a good idea to start with a library prepared from a tissue in which the target protein is known to be biologically relevant. A list of laboratory-made or commercially available cDNA or genomic libraries

**TABLE 2 Overview of Most Commonly Used Yeast Strains and Their Reporter Genes**

| Strain | Reporter Gene(s) | Reporter Gene Regulated by | Origin of UAS |
|---|---|---|---|
| *GAL4-Based* | | | |
| Y187 | *lacZ* | GAL4 | GAL1 (= 4 × $UAS_{G\ 17\text{-}mer}$) |
| SFY526 | *lacZ* | GAL4 | GAL1 (= 4 × $UAS_{G\ 17\text{-}mer}$) |
| H7Fc | *lacZ,* | GAL4 | 3 × $UAS_{G\ 17\text{-}mer}$ |
| | *HIS3* | GAL4 | GAL1 (= 4 × $UAS_{G\ 17\text{-}mer}$) |
| YRG-2 | *lacZ,* | GAL4 | 3 × $UAS_{G\ 17\text{-}mer}$ |
| | *HIS3* | GAL4 | GAL1 (= 4 × $UAS_{G\ 17\text{-}mer}$) |
| CG-1945 | *lacZ,* | GAL4 | 3 × $UAS_{G\ 17\text{-}mer}$ |
| | *HIS3* | GAL4 | GAL1 (= 4 × $UAS_{G\ 17\text{-}mer}$) |
| Y190 | *lacZ,* | GAL4 | GAL1 (= 4 × $UAS_{G\ 17\text{-}mer}$) |
| | *HIS3* | GAL4 | GAL1 (= 4 × $UAS_{G\ 17\text{-}mer}$) |
| *LexA-Based* | | | |
| EGY48 | *LEU2* | 6 × LexA op | |
| EGY191 | *LEU2* | 2 × LexA op | |
| L40 | *HIS3,* | 4 × LexA op | |
| | *lacZ* | 8 × LexA op | |

cloned in AD vectors is reported in Table 3. In most screens, cDNAs were derived from random- or oligo(dT)-primed RNA (ribonucleic acid). It should be kept in mind that, in contrast to genomic libraries, the relative representation of each cDNA closely reflects the endogenous expression level of the corresponding gene. Thus, interesting interacting proteins might be underrepresented, if their RNA is expressed at low levels in the cell. One solution is the use of normalized libraries. The process of normalizing cDNA libraries consists of reducing the representation of highly expressed cDNAs [41]. In addition, the choice of a random-primed versus an oligo(dT)-primed cDNA library can significantly influence the results of the screen. Discrete protein domains are more likely to be screened with random-primed libraries, while clones encoding nearly full-length proteins are enriched when oligo(dT)-primed libraries are used. In contrast, the complexity of genomic libraries is directly correlated to the size of the genome and to the number of independent clones that compose the library. For organisms that possess compact genomes, with small intergenic sequences and few introns, screening a genomic library instead of a cDNA library could be advantageous.

When screening libraries, a good representation is crucial. To screen a mammalian cDNA library until saturation, more than 5 to 10 × $10^6$ yeast transformants need to be screened. In classical two-hybrid library preparations only one out of six fused cDNAs is in the correct frame, increasing the total number

**TABLE 3 Overview of Laboratory-Made or Commercially Available cDNA or Genomic Libraries Cloned in AD Vectors**

| Host | Source | Vector (AD) | Supplier |
|---|---|---|---|
| Human Adult | 293 (kidney) cells, cDNA | pACT2 (GAL4 AD) | Clontech |
| | HaCat (keratinocyte) cells, cDNA | pACT2 (GAL4 AD) | Clontech |
| | HeLa (cervical carcinoma) cells, exponentially growing, cDNA | pJG4-5 (GAL4 AD) | R. Brent |
| | HeLa cells, cDNA | pPC86 (GAL4 AD) | Invitrogen |
| | HeLa cells, cDNA | pYESTrp (B42) | Invitrogen |
| | HeLa cells, cDNA | pGADT7-Rec (GAL4 AD) | Clontech |
| | WI-38 cells (lung fibroblast), serum starved, cDNA | pJG4-5 (GAL4 AD) | R. Brent |
| | Jurkat cells (T-cell leukemia), exponentially growing, cDNA | pJG4-5 (GAL4 AD) | R. Brent |
| | Aorta, cDNA | pACT2 (GAL4 AD) | Clontech |
| | Bladder, cDNA | pYESTrp2 (B42) | Invitrogen |
| | Bone marrow, cDNA | pACT2 (GAL4 AD) | Clontech |
| | Brain, cDNA | pEXP-AD502 (GAL4 AD) | Invitrogen |
| | Brain, cDNA | pYESTrp (B42) | Invitrogen |
| | Brain, cDNA | pACT2 (GAL4 AD) | Clontech |
| | Breast, cDNA | pYESTrp2 (B42) | Invitrogen |
| | Breast tumor, cDNA | pYESTrp2 (B42) | Invitrogen |
| | Chondrocyte, cDNA | pACT2 (GAL4 AD) | Clontech |
| | Colon tumor, cDNA | pYESTrp2 (B42) | Invitrogen |
| | Erythroleukemia, cDNA | pACT2 (GAL4 AD) | Clontech |
| | Heart, cDNA | pPC86 (GAL4 AD) | Invitrogen |
| | Heart, cDNA | pEXP-AD502 (GAL4 AD) | Invitrogen |
| | Heart, cDNA | pACT2 (GAL4 AD) | Clontech |
| | Keratinocyte foreskin, cDNA | pGAD10 (GAL4 AD) | Clontech |
| | Kidney, cDNA | pYESTrp2 (B42) | Invitrogen |

**TABLE 3** ***Continued***

| Host | Source | Vector (AD) | Supplier |
|---|---|---|---|
| | Kidney, cDNA | pACT2 (GAL4 AD) | Clontech |
| | Leukocyte, cDNA | pACT2 (GAL4 AD) | Clontech |
| | Liver, cDNA | pJG4-5 (GAL4 AD) | J. Pugh |
| | Liver, cDNA | pPC86 (GAL4 AD) | Invitrogen |
| | Liver, cDNA | pYESTrp (B42) | Invitrogen |
| | Liver, cDNA | pACT2 (GAL4 AD) | Clontech |
| | Lung, cDNA | pPC86 (GAL4 AD) | Invitrogen |
| | Lung, cDNA | pYESTrp2 (B42) | Invitrogen |
| | Lung tumor, cDNA | pYESTrp2 (B42) | Invitrogen |
| | Lymph node, cDNA | pACT2 (GAL4 AD) | Clontech |
| | Lymphocyte, EBV-transformed peripheral blood lymphocytes, B-cell population, Ig+, cDNA | pACT2 (GAL4 AD) | Clontech |
| | Mammary gland, cDNA | pACT2 (GAL4 AD) | Clontech |
| | Ovary, cDNA | pYESTrp2 (B42) | Invitrogen |
| | Ovary, cDNA | pACT2 (GAL4 AD) | Clontech |
| | Pancreas, cDNA | pACT2 (GAL4 AD) | Clontech |
| | Placenta, cDNA | pYESTrp2 (B42) | Invitrogen |
| | Placenta, cDNA | pGADT7-Rec (GAL4 AD) | Clontech |
| | Prostate, cDNA | pPC86 (GAL4 AD) | Invitrogen |
| | Prostate, cDNA | pYESTrp2 (B42) | Invitrogen |
| | Prostate, cDNA | pACT2 (GAL4 AD) | Clontech |
| | Prostate, adenocarcinoma, cDNA | pPC86 (GAL4 AD) | Invitrogen |
| | Prostate, leiomyosarcoma, cDNA | pPC86 (GAL4 AD) | Invitrogen |
| | Skeletal muscle, cDNA | pPC86 (GAL4 AD) | Invitrogen |
| | Skeletal muscle, cDNA | pACT2 (GAL4 AD) | Clontech |
| | Spleen, cDNA | pPC86 (GAL4 AD) | Invitrogen |
| | Spleen, cDNA | pYESTrp2 (B42) | Invitrogen |

| | | | |
|---|---|---|---|
| | Spleen, cDNA | pACT2 (GAL4 AD) | Clontech |
| | Testes, cDNA | pYESTrp (B42) | Invitrogen |
| | Testes, cDNA | pACT2 (GAL4 AD) | Clontech |
| | Thymus, cDNA | pACT2 (GAL4 AD) | Clontech |
| Human Fetus | Brain, 22 weeks, cDNA | pJG4-5 (GAL4 AD) | R. Brent |
| | Brain, cDNA | pPC86 (GAL4 AD) | Invitrogen |
| | Brain, cDNA | pEXP-AD502 (GAL4 AD) | Invitrogen |
| | Brain, cDNA | pACT2 (GAL4 AD) | Clontech |
| | Kidney, cDNA | pACT2 (GAL4 AD) | Clontech |
| | Liver, cDNA | pYESTrp (B42) | Invitrogen |
| | Liver, cDNA | pACT2 (GAL4 AD) | Clontech |
| Mouse | $CD4^+$ T cells, cDNA | pJG4-5 (GAL4 AD) | V. Prasad |
| | Activated lymph nodes, cDNA | pPC86 (GAL4 AD) | Invitrogen |
| | Brain, cDNA | pPC86 (GAL4 AD) | Invitrogen |
| | Brain, cDNA | pACT2 (GAL4 AD) | Clontech |
| | 7-day embryo, cDNA | pACT2 (GAL4 AD) | Clontech |
| | 8.5-day embryo, cDNA | pPC86 (GAL4 AD) | Invitrogen |
| | 10.5-day embryo, cDNA | pPC86 (GAL4 AD) | Invitrogen |
| | 11-day Embryo, cDNA | pACT2 (GAL4 AD) | Clontech |
| | 17-day Embryos, cDNA | pACT2 (GAL4 AD) | Clontech |
| | Embryonic fibroblasts, cDNA | pACT2 (GAL4 AD) | Clontech |
| | Kidney, cDNA | pACT2 (GAL4 AD) | Clontech |
| | Liver, cDNA | pPC86 (GAL4 AD) | Invitrogen |
| | Liver, cDNA | pACT2 (GAL4 AD) | Clontech |
| | Lymphoma, cDNA | pACT2 (GAL4 AD) | Clontech |
| | Testis, cDNA | pACT2 (GAL4 AD) | Clontech |
| Hamster | CHO (Chinese hamster ovary) cells, exponentially growing, cDNA | pJG4-5 (GAL4 AD) | V. Prasad |

**TABLE 3** ***Continued***

| Host | Source | Vector (AD) | Supplier |
|---|---|---|---|
| Rat | Brain, cDNA | pACT2 (GAL4 AD) | Clontech |
| | Brain, cDNA | pPC86 (GAL4 AD) | Invitrogen |
| | Liver, cDNA | pACT2 (GAL4 AD) | Clontech |
| | Liver, cDNA | pPC86 (GAL4 AD) | Invitrogen |
| | Lung, cDNA | pACT2 (GAL4 AD) | Clontech |
| Fish | Zebrafish, 1-month *Brachidanio rerio*, cDNA | pGAD10 (GAL4 AD) | Clontech |
| *Xenopus laevis* | Unfertilized oocyte, ages 1–3 years, injected with gonadotropin | pACT2 (GAL4 AD) | Clontech |
| *Drosophila melanogaster* | Adult whole, Canton S strain, cDNA | pACT2 (GAL4 AD) | Clontech |
| | Disc, cDNA | pJG4-5 (GAL4 AD) | R. Brent |
| | 0- to 12-h embryos, cDNA | pJG4-5 (GAL4 AD) | R. Brent |
| | 24-h embryos, cDNA | pACT2 (GAL4 AD) | Clontech |
| | Ovary, cDNA | pJG4-5 (GAL4 AD) | R. Brent |
| *Caenorhabditis elegans* | cDNA | pPC86 (GAL4 AD) | Invitrogen |
| *Saccharomyces cerevisiae* | Strain S228C, genomic DNA | pJG4-5 (GAL4 AD) | R. Brent |
| *Schizosaccharomyces pombe* | Strain SP972, exponentially growing, cDNA | pGAD GH (GAL4 AD) | Clontech |

of independent clones to be screened to over a million. Making directional libraries of a relevant tissue or cell type could be a solution.

Screening of libraries selects for optimized interactions. Many isolates may not represent full-length cDNA. Indeed, it has been shown that protein subdomains can interact better than full-length proteins, probably reflecting domain function during folding of the protein. The best way to circumvent this problem is probably to clone only full-length cDNAs in the correct open reading frame. Although extremely time consuming, this approach was taken to establish the complete yeast protein linkage map (see Whole Genome Approaches Using the Two-Hybrid System).

For the library plasmid there is a major difference in the promoter used between the LexA and GAL4 system. While in the most commonly used plasmids of the GAL4 system, fusion proteins are weakly and constitutively expressed, they are cloned behind a stronger but inducible promoter in the LexA system. In one of the last versions of library plasmid used in the GAL4 system, pACT2, a truncated weakly expressing version of ADH1 is used.

Important is also the relative strength of the AD, related to its ability to initiate transcription. Both VP16 and the AD of GAL4 are strong activators making the system more sensitive. This might be needed for the detection of weak interactions but inevitably results in higher backgrounds. Therefore, the use of B42AD, which has intermediate transactivation activity, might be beneficial in conducting a screening.

***New Developments*** Since the original description of the yeast two-hybrid system 16 years ago, a number of "variations on a theme" based on the initial idea have been described. In one set of variations, the classical configuration of the two-hybrid fusion proteins was modified to expand the range of possible protein–protein interactions that could be analyzed. For example, systems were developed to detect interactions that involve transcriptional activators or repressors, and membrane or extracellular proteins. In another set of variations, systems were developed to detect trimeric interactions or macromolecular interactions where a third protein activates or modifies one of the fused proteins and facilitates the protein–protein interaction (*three-hybrid systems*). Finally, the original concept was turned upside down and *reverse two-hybrid systems* were developed to identify mutations, peptides, or small molecules that dissociate protein–protein interactions. The next paragraphs describe the first two types of two-hybrid system variants; the reverse two-hybrid systems are discussed in the second part of the chapter. An additional review on progress and variations in two-hybrid and three-hybrid technologies is available [42].

*Improvements of Classical Yeast Two-Hybrid System* Refinements of vectors for use in the yeast two-hybrid system have been reviewed by Roder et al. [43]. A new host strain has been recently developed that is extremely sensitive to weak interactions and eliminates nearly all false positives using simple

plate assays. This yeast strain contains three reporter genes: *HIS3*, *ADE2*, and *LacZ*, each under control of a different promoter (GAL1, GAL2, and GAL7, respectively) that responds to the same activator, GAL4 [23]. In another system, the target protein is expressed as a fusion with the DBD domain of the human estrogen receptor (ER) in a yeast strain containing an integrated *URA3* reporter gene driven by one or three ER response elements (EREs). In contrast to the *HIS3*- and *LEU2*-based systems, it can be assayed quantitatively by determining the activity of the reporter gene product orotidine-5′-monophosphate decarboxylase in cell-free extracts, and also allows negative selection by using 5-fluoroorotic acid [44].

*Two-Bait Systems* The last few years have seen the advent of *two-bait systems* in which different baits are bound to DNA upstream of different reporters. These systems have been used to identify proteins that interact with different domains of a protein (Snf1) [45], different alleles of a protein (Ras) [46], and to identify mutant proteins that differentially bind to two known binding partners of a wild-type protein (Ste5) [47]. Combining data from these systems with data from classical two-hybrid systems likely can help dissect multimeric protein complexes [46]. In fact, two-bait systems, like other systems in which a third protein is expressed, can detect interactions that depend on modification or bridging by a third protein [48].

*Sos Recruitment System* A recent variation of the yeast two-hybrid system, termed *Sos recruitment system* (SRS), is based on the Ras pathway in yeast and has the advantage of bypassing the reconstitution of a transcription factor [20]. The SRS is based on the observation that the mammalian guanosine 5′-diphosphate–guanosine 5′-triphosphate (GDP–GTP) exchange factor (GEF) hSos can only activate Ras when hSos is localized to the plasma membrane. In mammalian cells, this occurs by recruitment to the cytoplasmatic tail of activated growth factor receptors. The yeast *S. cerevisiae* requires a functional Ras signaling pathway for cell viability. A yeast strain containing a point mutation in the yeast GEF, *cdc25-2*, shows temperature-sensitive growth, but expression of hSos artificially targeted to the membrane by myristoylation or farnesylation can restore growth at nonpermissive temperature. Thus, the SRS uses the requirement of hSos to be recruited to the membrane to rescue growth as a method to detect protein–protein interactions. The target protein is fused to the GEF domain of hSos (GEF-X), and the partner protein or cDNA to be screened is fused to the Src membrane-anchoring (MA) domain (MA-Y). When these fusion proteins are coexpressed in a *cdc25-2* yeast strain, the cells grow at the nonpermissive temperature only if the two hybrid proteins interact; in this case, the GEF-X/MA-Y interaction allows recruitment of hSos to the membrane and rescue of the yeast *cdc25-2* mutation (Fig. 2).

The SRS may be the method of choice in studying interactions that involve transcriptional activators or repressors because it is not based on a transcriptional readout. Initially, the SRS was tested using the interacting AP-1 factors

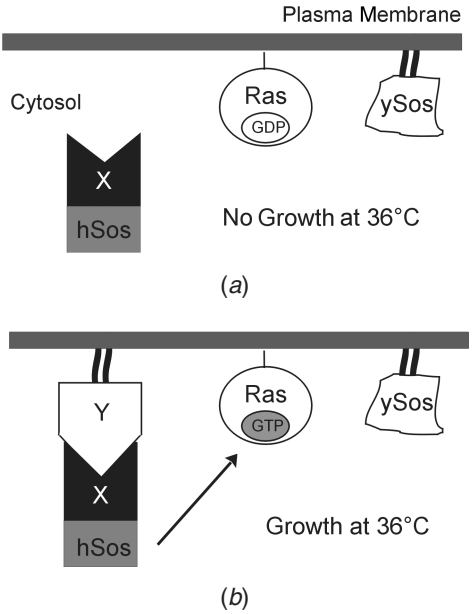


**Figure 2** Principle of the SRS system. (*a*) In yeast cells carrying the *cdc25-2* mutation the yeast Ras guanyl exchange factor *cdc25* (ySos) is inactivated at nonpermissive temperature (36°C), leading to growth arrest. A bait X that is fused to the human Ras guanyl exchange factor Sos (hSos) cannot overcome the growth defect because the hSos-X fusion protein is located in the cytoplasm. (*b*) If an interacting protein Y is localized to the plasma membrane by means of a myristoylation signal, the interaction between X and Y recruits hSos-x to the membrane where it stimulates GDP exchange on Ras, thereby circumventing the *cdc25-2* mutation. Consequently, the yeast cells will grow at 36°C.

c-Jun and c-Fos, which are difficult to study in a conventional yeast two-hybrid system since both factors are strong self-activators [20, 49]. This system has been then successfully applied for identification of BRCA1 interacting partners [50]. In addition, certain proteins may function more physiologically when expressed in the cytoplasm rather than in the nucleus. For example, the SRS could be more suitable for examining interactions between proteins that require modification by cytoplasmic or membrane-associated enzymes.

However, a potential limitation of the SRS-based screening procedure is the isolation of false positives encoding Ras. In fact, mammalian Ras can bypass the requirement for a functional Ras GEF and represents a predictable false positive in this system. A possible solution to the problem is based on the observation that overexpression of GTPase activating protein (GAP) can suppress the bypass of the Cdc25 function by Ras [51]. However, it remains to be determined whether this improvement indeed significantly decreases the number of false positives.

*The USPS System* The *ubiquitin-based split-protein sensor* (USPS) system, originally developed by Johnsson and Varshavsky [19], relies on the properties of the ubiquitin protein. The ubiquitin protein consists of 76 amino acids and is highly conserved between all eukaryotes [52]. Its primary function is to act as a "tag" for degradation by being attached to proteins through ubiquitin conjugating enzymes. Once a protein becomes tagged with several ubiquitin moieties, it is transported to the 26S proteasome, where ubiquitin-specific proteases (UBPs) cut the peptide bond at a double glycine motif in the junction between the attached ubiquitins and the target protein. The released ubiquitin moieties are then recycled back to the cytoplasm, whereas the target protein is degraded by the 26S proteasome [53].

The USPS system takes advantage of the highly specific cleavage mediated by the UBPs. The expression of a fusion protein consisting of ubiquitin and a C-terminally attached reporter in yeast results in fast and complete cleavage by UBPs within minutes [19]. The folded structure of ubiquitin is crucial to the recognition and subsequent cleavage events by UBPs: expression of an N-terminal ubiquitin moiety carrying a point mutation in a hydrophobic core residue (NubG) together with the C-terminal ubiquitin moiety (Cub) in the same yeast cell does not result in cleavage by UBPs anymore, presumably because the partially unfolded NubG moiety does not recognize and bind to the Cub moiety. As the UBPs do not recognize Cub alone, no cleavage of the attached reporter takes place. In order to exploit the ubiquitin moieties as reporters for detecting protein–protein interactions, two binding partners, X and Y, are fused to NubG and Cub, respectively. Upon interaction of X and Y, the NubG and Cub moieties are brought into close proximity, resulting in a partial refolding of NubG, followed by reassociation of NubG and Cub into what the authors termed *split-ubiquitin*. Split-ubiquitin represents a good substrate for UBPs and, therefore, the attached reporter is cleaved off (Fig. 3). In the original description of the system, the reporter protein used was human dihydrofolate reductase (DHFR) expressed as a fusion to Cub, and cleavage of such a hybrid was used as an indication for X/Y interaction.

As the formation of split-ubiquitin and the subsequent cleavage by UBPs do not depend on any particular intracellular localization of the proteins, the split-ubiquitin system is suitable for the study of membrane proteins. This system has been successfully used to detect interaction among a variety of proteins, including sucrose transporters [54], yeast oligosaccharyl transferases [55], transmembrane proteins normally present in the endoplasmic reticulum [56, 57], and proteins involved in viral replication in *Arabidopsis* [58]. This two-hybrid variant was also proven capable of reconstituting the well-characterized homodimerization of the yeast Gcn4p leucine zipper domain. However, this strategy has not yet been adapted for the identification of novel interacting proteins.

*Protein Three-Hybrid Systems* The protein three-hybrid systems represent an additional adaptation of the yeast two-hybrid system that allows investi-

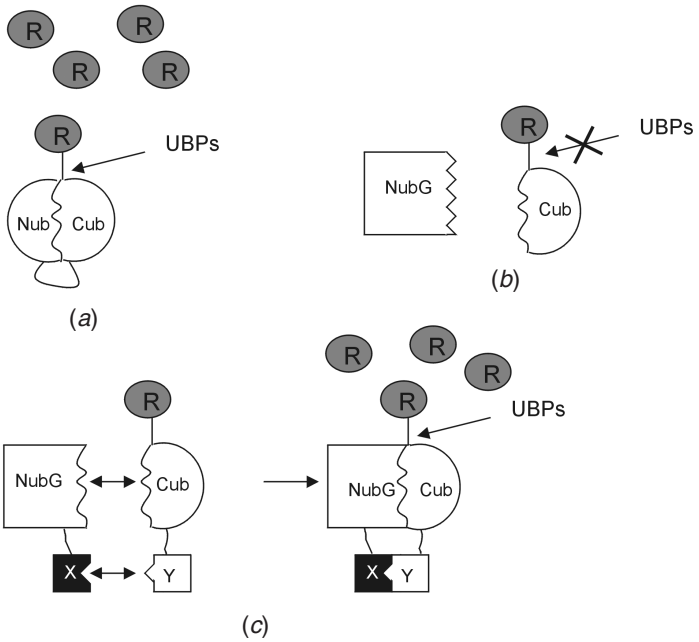


**Figure 3** The ubiquitin-based split-protein system (USPS). (*a*) Native ubiquitin is recognized by UBPs, which cleave at a double-glycine motif located at the carboxy terminus of ubiquitin. A reporter protein (R) that is fused to the C-terminus of ubiquitin is released upon cleavage. (*b*) Cleavage at an extended loop separates ubiquitin into a C-terminal (Cub) and an N-terminal (Nub) domain. The introduction of a point mutation into Nub (NubG) causes its partial misfolding and therefore abolishes the binding of Nub to Cub. As UBPs do not recognize the isolated Cub domain, the reporter R is not cleaved from the C-terminus of Cub. (*c*) A protein X is fused to NubG and its interacting partner Y is fused to Cub. The interaction of X and Y brings NubG and Cub together, allowing the partial refolding of NubG and leading to their reassociation into split-ubiquitin. Like native ubiquitin, split-ubiquitin is recognized by UBPs, leading to the cleavage and release of the reporter R.

gation of ternary protein complexes. In fact, stable interaction of proteins X and Y may rely on the presence of a third protein Z. Protein Z either mediates the interaction or induces a conformational change in one of the proteins (e.g., X) so that it promotes interaction with another protein Y. One of these systems was employed by Zhang and Lauter [48] to investigate intracellular associations of proteins during initiation of the epidermal growth factor receptor (EGFR) signal transduction cascade. Within the yeast cells, the EGFR cytoplasmic domain was expressed as a DBD fusion protein and Sos as an AD fusion protein. To identify new interacting proteins, a mouse brain library fused to the SV40 T-antigen NLS was coexpressed from a third vector in conjunction with the EGFR–DBD and the Sos–AD hybrid proteins, leading to the identification of the mouse Grb2 protein [48]. Recently, another variant of

this system was proposed by Tirode et al. [17]. This version uses both LexA and GAL4 plasmids and is based on the addition of a Met25 promoter for the third partner. This third partner is cloned on the same plasmid that already encodes the GAL4 and LexA hybrid proteins. By doing this, a selection marker is saved. Transcription of the third protein is repressed upon addition of methionine in the culture medium and provides an elegant way to perform negative controls [17].

An alternative use of a third protein coexpressed with the two-hybrid fusion proteins is exploiting the third protein's ability to modify or activate one of the fused proteins and facilitate the protein–protein interaction. As an example, the kinase three-hybrid system termed *tri-brid system* can detect protein–protein interactions that depend on posttranslational modifications. In fact, one of the limitations of the classical version of the yeast two-hybrid system is the lack of some crucial posttranslational modifications (such as tyrosine phosphorylation) in *S. cerevisiae* [59]. This was solved by introduction of a third component, that is, a cytosolic tyrosine kinase, which then phosphorylated substrates in the yeast cell. In principle, it should be possible to incorporate any posttranslational or allosteric regulation that is desired. However, care must be taken that the posttranslational modification is properly performed in yeast [60].

Finally, a yeast multiprotein system that allows simple and rapid detection of interactions among up to four proteins was recently developed by Loregian et al. [61]. This system, which further expands the three-hybrid system to include a fourth protein, could be most useful to dissect multiprotein complexes.

*Mammalian Two-Hybrid* Although most two-hybrid systems use yeast, some mammalian variants do exist. In one, the gene of interest is expressed as a fusion with the GAL4 DBD. The second gene is expressed as a fusion with VP16 AD. A third vector, pG5CAT, contains the chloramphenicol acetyltransferase (*CAT*) reporter gene downstream of a GAL4 responsive UAS. These three vectors are cotransfected into a mammalian cell line. Interaction between proteins X and Y is assayed by measuring *CAT* gene expression by any standard method [62]. In another version, interaction of proteins fused to VP16 with a GAL4-derived bait drives expression of hygromycin B phosphotransferase, *CAT*, or CD4 cell surface antigen [63].

***New Applications of the Yeast Two-Hybrid System*** During the past decade, together with the evolution of the technique, new applications have emerged. Apart from the molecular dissection of known interactions and the identification of new interacting partners, several new problems can now be addressed by means of the two-hybrid technology.

*Interaction Suppression* Interaction suppression makes use of the two-hybrid technique to evaluate the biological relevance of an interaction between two

proteins. First, mutations that affect interaction of a target protein (A) to its partner (B) are identified. Once a mutant version of A is found, the influence of this mutation on the phenotype can be studied. However, if one finds mutants of protein A that do not bind protein B, one cannot exclude that also the interaction of protein A with an unknown protein C is abolished and accounts for the change in phenotype. Therefore, in a second step one has to search for binding suppressors, that is, a mutant version of B that is able to interact with the mutated version of A and restores the altered phenotype. This approach was taken to study the Ras/Raf pathway [64–66].

*Protease Trap* This system relies on the nuclear localization of the hybrid protein [67]. A functional transcription factor is fused to a domain that prevents it from translocating into the nucleus. If a protease site is cloned in between the transcription factor and this domain, it is possibile to search for a protease that cuts off the antinuclear localization domain enabling nuclear localization of the transcription factor and finally resulting in reporter gene expression. This approach could also be applied to screen for target sequences for a known protease by cloning random sequences in between the two functional domains.

*Whole Genome Approaches Using the Yeast Two-Hybrid System* One of the most spectacular and ambitious applications of the two-hybrid system consists of establishing so-called protein linkage maps (PLMs). These maps consist of all possible protein interactions that occur during the entire life span of a cell. PLMs might provide information on new functions of known proteins by discovery of unexpected interactions, identify cross-connections between cellular pathways, result in the identification of novel drug targets, and gain more insight in the overall cell system. Since it is a genetic system, the yeast two-hybrid system is well suited to high-throughput applications such as the identification of interactions that take place between all proteins expressed in a given cell or organism.

Currently, two approaches are being used to generate PLMs. In the so-called *matrix approach*, or *array approach*, a set of open reading frames (ORFs) is amplified using the polymerase chain reaction (PCR), cloned as both bait and prey (i.e., as a fusion to a DBD and as a fusion to an AD), and then expressed into isogenic yeast strains of opposite mating type. The yeast strain expressing a DBD fusion is then mated with an array of clones each expressing a different AD fusion [25]. Since yeast mating is a very efficient process, numerous combinations of DBD-X/AD-Y can be assayed simultaneously. Practically, this task is carried out by robots that transfer aliquots from a lawn of cells expressing one DBD fusion to arrays of cells each expressing a different AD fusion. This procedure is repeated for each strain expressing a DBD fusion, until all DBD fusions have been mated with the entire AD array. Positive interactions are selected through the ability of diploid yeast colonies containing an interacting fusion pair to grow on selective media. In order to

eliminate the false positives arising from such an approach, the experiments are performed in duplicate, and only interactors found in both experiments are considered to be true positives. An advantage of the matrix approach is that it rapidly becomes clear which locations produce false positives, confirming that the system is working properly; if a particular AD hybrid in the array binds all DBD fusions, it most likely represents a false positive and should thus be discarded. This approach is nevertheless restricted by several limitations. First, only known or predicted proteins can be tested, and this is an obvious key problem in genome-wide projects. Second, the matrix approach enables the use of a single set of growth conditions, which precludes the possibility of using different selective conditions for each specific interaction. For example, the 3AT concentration cannot be adjusted to account for weak self-activation of some baits. Third, the use of full-length proteins for both the DBD-X and the AD-Y fusions might prevent the identification of several interactions due to various intrinsic problems such as folding, degradation, and toxicity.

A second approach in genome-wide yeast two-hybrid analysis is the so-called *library screening approach* in which DBD fusions are screened against complex libraries containing AD fusions of full-length ORFs or ORF fragments. In contrast to the matrix approach, this method does not separate the different AD fusions on an array. Instead, the library is divided into pools, and each yeast strain expressing a DBD fusion is mated with a library pool. Then, diploid cells containing an interacting protein pair are selected. The library screening approach is more sensitive than the matrix approach since it uses not only full-length ORFs but also random ORFs fragments. Often, a protein–protein interaction can be detected using fragments of the proteins of interest but not the full-length proteins, as they do not fold properly when expressed in yeast or are degraded [68]. The use of fragments often circumventes these problems. On the other hand, library screens are much more expensive and time consuming than matrix screens since they require the analysis of larger numbers of clones. In addition, the library plasmids encoding AD hybrids have to be isolated and sequenced from all selected clones in order to identify the interacting partners.

In the past few years, several systematic two-hybrid projects have been undertaken to analyze protein–protein interactions at a global level [68–72]. These are summarized in Table 4, together with relevant information about the type of cell or organism that was investigated, the method that was used to perform the screen, and the number of protein–protein interactions that were identified. The first PMLs were reported in 1994 for *Drosophila melanogaster* [73], followed by similar screenings for T7 bacteriophage [74] and a subset of yeast proteins involved in messenger RNA (mRNA) splicing [26]. Likewise, Walhout et al. [68] and Boulton et al. [71] mapped protein–protein interaction networks involved in *Caenorhabditis elegans* vulval development and DNA-damaging response components, respectively. Finally, Rain et al. [72] have recently built a large-scale protein–protein inter-

action map of the human gastric bacterial pathogen *Helicobacter pylori,* whose genome encodes 1590 putative ORFs [75]. A total of 261 baits were used to detect 1524 interactions, resulting in a protein interaction map covering much of the proteome.

The most comprehensive large-scale screening approaches reported to date focus on the yeast *S. cerevisiae.* Two groups studied all 6131 annotated yeast ORFs using both matrix and library screening approaches [69, 70, 76]. Initially, Ito et al. [76] performed a large-scale matrix screen using 159 ORFs that were cloned as DBD and AD fusions. They analyzed 430 matings (representing 10 percent of all possible combinations between DBD and AD fusions) and identified 175 interactions, of which 163 had not previously been reported. They recently completed their systematic approach, identifying a total of 841 interactions [70]. The second group performed both matrix and library screenings [69]. Using the matrix method, 192 ORFs were created as DBD fusions and then mated with the 6131 yeast ORFs fused to the AD, resulting in 281 protein–protein interactions. For the library screen, a library was made by pooling all 5345 AD-fused ORFs. These were then mated separately to the same 5345 ORFs fused to the DBD, yielding a total of 692 protein–protein interactions.

A surprising observation was made when comparing the results of Ito et al. [76] and Uetz et al. [69]. Despite the fact that both groups used the same yeast ORFs in their experiments, only 20 percent of all interactions detected in the two screens overlapped [70]. The reasons for this small overlap are difficult to explain, but most likely the differences are due to the different experimental settings. For example, the different DBD and AD plasmids used by the two groups may have affected the folding and expression level of the proteins, the use of PCR to amplify the yeast ORFs may have introduced mutations that abolish interactions, and, most importantly, the stringency of selection may have been different, eliminating interactions seen by one group from the other group's results. In summary, the large-scale screenings carried out so far indicate that even when using exhaustive library screens that potentially cover all interactions in a genome, it is still difficult to estimate what percentage of protein–protein interactions that occur in a cell or organism under investigation are actually identified in such screens. The small overlap between data sets of Ito et al. [76] and Uetz et al. [69] suggests that even within the subset of protein–protein interactions that can be detected using the yeast two-hybrid system, such screenings are still far from being exhaustive.

## Peptide Display Technologies

Surface display of peptide and proteins is another approach to detect and characterize protein–protein interactions. The advantage of peptide display technology over other research tools and its great potential have been demonstrated by successful applications in diverse fields of biomedical research, including drug discovery [77–79].

**TABLE 4 Overview of Genome-Wide Protein–Protein Interaction Studies**

| Organism | Predicted ORFs | Sample | Screen Method | Number of Interactions | References |
|---|---|---|---|---|---|
| T7 bacteriophage | 55 | Proteome | Matrix screen + Library screen | 25 | 1 |
| Vaccinia virus | 266 | Proteome | Matrix screen | 37 | 2 |
| Hepatitis C virus | 10 | Proteome | Matrix screen + Library screen | 15 | 3 |
| *H. pylori* | 1,590 | Subset of proteome | Library screen | 1,524 | 4 |
| *S. cerevisiae* | 6,131 | Nuclear proteins involved in pre-mRNA splicing pathway | Library screen | 170 | 5 |
| | | RNA polymerase III subunits | Library screen | 20 | 6 |
| | | SM-like proteins | Library screen | 229 | 7 |
| | | Proteome | Library screen | 841 | 8 |
| | | Subset of proteome | Matrix screen | 281 | 9 |
| | | | Library screen | 692 | |
| *C. elegans* | 19,293 | Proteins involved in vulval development | Matrix screen | 11 | 10 |
| | | | Library screen | 148 | |
| | | Proteins involved in DNA damage response | Matrix screen | 26 | 11 |
| | | | Library screen | 165 | |
| | | 26S proteasome subunits | Matrix screen | 17 | 12 |
| | | | Library screen | 138 | |
| *D. melanogaster* | 13,600 | Cell-cycle regulatory proteins | Matrix Screen | 19 | 13 |

1. Bartel, P. L., Roecklein, J. A., SenGupta, D., Fields, S. (1996). A protein linkage map of *Escherichia coli* bacteriophage T7. *Nat. Genet.*, *12*, 72–77.
2. McCraith, S., Holtzman, T., Moss, B., Fields, S. (2000). Genome-wide analysis of vaccinia virus protein-protein interactions. *Proc. Nat. Acad. Sci. USA*, *97*, 4879–4884.
3. Flajolet, M., Rotondo, G., Daviet, L., Bergametti, F., Inchauspe, G., Tiollais, P., Transy, C., Legrain, P. (2000). A genomic approach of the hepatitis C virus generates a protein interaction map. *Gene*, *242*, 369–379.
4. Rain, J. C., Selig, L., De Reuse, H., Battaglia, V., Reverdy, C., Simon, S., Lenzen, G., Petel, F., Wojcik, J., Schachter, V., Chemama, Y., Labigne, A., Legrain, P. (2001). The protein-protein interaction map of *Helicobacter pylori*. *Nature*, *409*, 211–215.
5. Fromont-Racine, M., Rain, J. C., Legrain, P. (1997). Toward a functional analysis of the yeast genome through exhaustive two-hybrid screens. *Nat. Genet.*, *16*, 277–282.
6. Flores, A., Briand, J. F., Gadal, O., Andrau, J. C., Rubbi, L., Van Mullem, V., Boschiero, C., Goussot, M., Marck, C., Carles, C., Thuriaux, P., Sentenac, A., Werner, M. (1999). A protein-protein interaction map of yeast RNA polymerase III. *Proc. Nat. Acad. Sci. USA*, *96*, 7815–7820.
7. Fromont-Racine, M., Mayes, A. E., Brunet-Simon, A., Rain, J. C., Colley, A., Dix, I., Decourty, L., Joly, N., Ricard, F., Beggs, J. D., Legrain, P. (2000). Genome-wide protein interaction screens reveal functional networks involving Sm-like proteins. *Yeast*, *17*, 95–110.
8. Ito, T., Chiba, T., Ozawa, R., Yoshida, M., Hattori, M., Sakaki, Y. (2001). A comprehensive two-hybrid analysis to explore the yeast protein interactome. *Proc. Nat. Acad. Sci. USA*, *98*, 4569–4574.
9. Uetz, P., Giot, L., Cagney, G., Mansfield, T. A., Judson, R. S., Knight, J. R., Lockshon, D., Narayan, V., Srinivasan, M., Pochart, P., Qureshi-Emili, A., Li, Y., Godwin, B., Conover, D., Kalbfleisch, T., Vijayadamodar, G., Yang, M., Johnston, M., Fields, S., Rothberg, J. M. (2000). A comprehensive analysis of protein-protein interactions in *Saccharomyces cerevisiae*. *Nature*, *403*, 623–627.
10. Walhout, A. J., Sordella, R., Lu, X., Hartley, J. L., Temple, G. F., Brasch, M. A., Thierry-Mieg, N., Vidal, M. (2000). Protein interaction mapping in *C. elegans* using proteins involved in vulval development. *Science*, *287*, 116–122.
11. Boulton, S. J., Gartner, A., Reboul, J., Vaglio, P., Dyson, N., Hill, D. E., Vidal, M. (2002). Combined functional genomic maps of the *C. elegans* DNA damage response. *Science*, *295*, 127–131.
12. Davy, A., Bello, P., Thierry-Mieg, N., Vaglio, P., Hitti, J., Doucette-Stamm, L., Thierry-Mieg, D., Reboul, J., Boulton, S., Walhout, A.J., Coux, O., Vidal, M. (2001). A protein-protein interaction map of the *Caenorhabditis elegans* 26S proteasome. *EMBO Reps.*, *2*, 821–828.
13. Finley, R. L., Jr., Brent, R. (1994). Interaction mating reveals binary and ternary connections between *Drosophila* cell cycle regulators. *Proc. Nat. Acad. Sci. USA*, *91*, 12980–12984.

***Phage Display*** Phage display is the most widely adopted molecular display technique. Smith first demonstrated in 1985 [80] that an *E. coli* phage can express a fusion protein bearing a foreign peptide on its surface and suggested that libraries of recombinant phages could be constructed and screened to identify proteins that bind to a specific partner. In general, phage display usually refers to an in vitro selection technique in which a peptide or a protein is genetically fused to a coat protein of a filamentous phage, such as M13, fd, and f1, resulting in display of the fused protein to the surface of the virion, while the DNA encoding the fusion resides within the phage.

Protein sequences are usually fused to the amino terminus of either the major coat protein, pVIII (about 3000 copies per phage), or a tail protein, pIII (5 copies per phage). Thus, a foreign DNA sequence inserted in these genes results in multiple copies of the fusion protein displayed by the phage. This is termed *polyvalent display*. In general, polyvalent display is limited to small peptides because larger inserts interfere with the function of the coat proteins, and the phage becomes poorly infective. In polyvalent display, the peptides are either displayed in an unconstrained form or in a constrained form by introducing flanking cysteine residues. The latter peptides are much less flexible, and peptides selected from constrained libraries have quite often higher affinities than those selected from unconstrained libraries. The minor coat protein pVI has been also used to display foreign peptide sequences fused at its C-terminus, a system that can be more suitable when a free C-terminal end is necessary for proper activity [81].

Next, a phage library is enriched for members that bind an immobilized target by a procedure called *panning* (Fig. 4) in which the phage library is incubated with the protein of interest bound to a plastic dish or well. After washing to remove nonspecifically bound phages, specifically bound phages are eluted either by competition with free target or by more stringent washes. The library is then amplified in *E. coli*, and the selection is repeated to enrich for phages possessing peptides specifically binding to the target protein. Multiple rounds of selection make this approach very efficient. Finally, the DNA is extracted from individual viral plaques and sequenced to determine the sequence of the peptides that bind to the target protein.

A major advance in phage display came with the development of the *monovalent systems* in which the coat protein fusion is expressed from a phagemid, and a helper phage supplies a large excess of the wild-type coat protein [82, 83]. Therefore, the phages are functional because the recombinant protein forms only a small amount of the total coat protein. The vast majority (>99 percent) of the population of phage virions displays either one or no copies of the fusion protein on their surface. Such phages can accommodate up to 100 kDa of heterologous protein without any significant decrease of phage infectivity. In addition, monovalent phage display avoids potential avidity effects observed with polyvalent display, in which the phage can attach to the absorbent at multiple points.

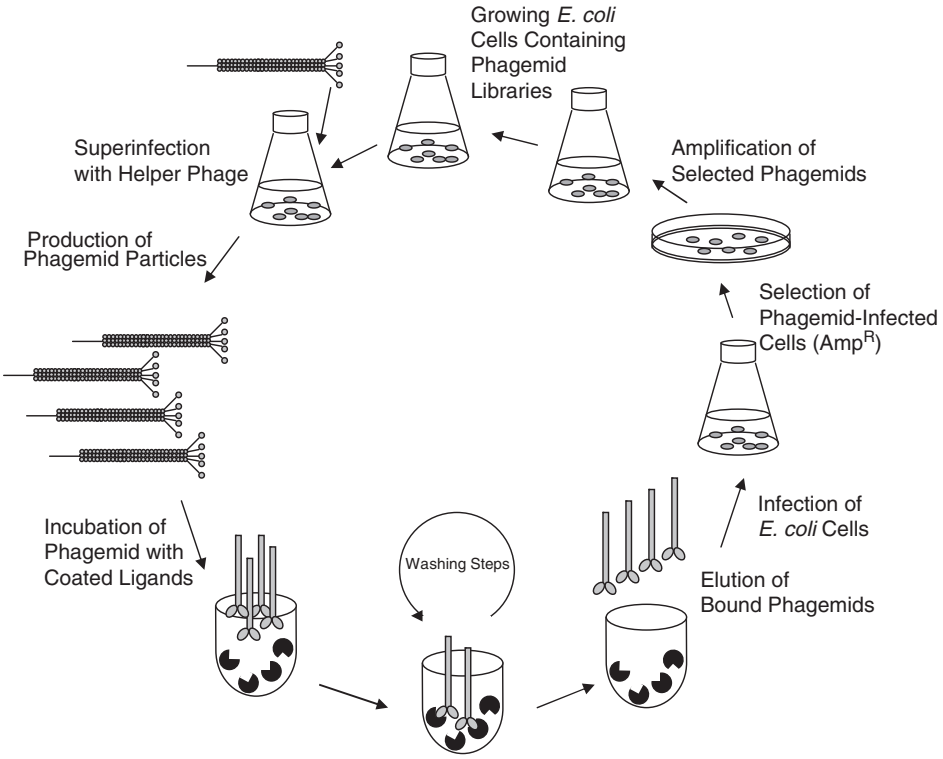


**Figure 4** Scheme of the panning cycle. Phage particles displaying fusion proteins of peptide variants are produced by infecting phagemid libraries, i.e., *E. coli* cells carrying the phagemid as a plasmid DNA, with helper phage (*upper left*). These particles are incubated with a target of interest that has been immobilized on a plate or bead (*lower left*). Particles with low affinity for the target are washed away (*bottom*). The remaining particles are eluted by disrupting the binding interactions between the phage and target, e.g., with acid, and then used to infect fresh *E. coli* cells; clones are selected for antibiotic resistance (*right half*). The enriched clones are now ready to be packaged. The process is then repeated, resulting in stepwise enrichment of the phage pool in favor of the tightest binding sequences. After multiple rounds of selection/amplification, individual clones are characterized by DNA sequencing.

Consequently, phages were used for the display not only of short peptides but also of functional proteins such as antibody fragments, hormones, enzymes, and enzyme inhibitors, as well as for the selection of specific phage on the basis of functional interactions (antibody–antigen, hormone–hormone receptor, enzyme–enzyme inhibitor, etc.). Antibodies have been displayed on phages in the form of scFv or Fab′ fragments using either the phage or the phagemid system. In the latter case, either the VH-CH1 or VL-CL chain is fused to pIII while the other chain is expressed without pIII fusion. Thus, phage display has

been applied successfully not only to epitope mapping but also to identification of critical amino acids responsible for protein–protein interactions and to isolate leads for the discovery of new therapeutics (see the second part of this chapter).

Phage display presents several advantages for the study of protein–protein interactions. It is a rapid procedure and has been shown to be widely applicable. There are, however, limitations to the technique. In early examples of phage display, the length of the peptides displayed on bacteriophage virions was usually restricted to 6 to 15 amino acids because larger polypeptides (>10 residues for pVIII display) compromised phage infectivity and thus could not be efficiently displayed. The development of phagemid display systems solved this problem by fusing the peptides to an additional coat protein encoded by a phagemid vector and subsequently infecting with a helper phage [82]. Other intrinsic restrictions of display systems based on filamentous phage include their inability to display protein sequences toxic to bacterial cells and the fact that the fusion protein must be secreted across the bacterial membrane. An alternative system to alleviate the toxicity problem is represented by the use of bacteriophage λ, in which the synthesis of any toxic protein can be repressed during the lysogenic state and induced for a very short time just before cell lysis [84]. Furthermore, protein secretion is not required for phage λ display because this phage assembles intracellularly. In addition, λ-based systems offer the advantage of efficient construction and maintenance of very large libraries. Two λ-phage display systems have recently been described. One approach entails fusion to the V gene, which encodes a tail fiber, at the end corresponding to the C-terminus [85]. Alternatively, peptide sequences can be fused to the amino terminus of the D protein on the phage capsid [84]. Other phage vectors wherein displayed protein sequences are not secreted are the bacteriophage T4, where polypeptides can be exposed by fusion with the C-terminus of fibritin, a protein of the bacteriophage neck [86], and the bacteriophage T7 (see Reagents).

Other disadvantages of the phage display include the use of a bacterial host, which may preclude the correct folding or modification of some proteins, and the fact that all phage-encoded proteins are fusion proteins, which may limit the activity or the accessibility for binding of certain proteins.

Finally, one should keep in mind that, since the phage display libraries consist of fully randomized peptides displayed on phage, a binding partner identified in a particular panning experiment will not necessarily correspond to a "natural" ligand for the target. The panning procedure iteratively selects for those peptides that best bind the target under the panning conditions in vitro, without regard to the biological role of the target in vivo. For certain targets, such as antibodies with linear epitopes, the selected sequence will in all likelihood correspond to that region of the antigen recognized by the antibody. For targets that bind to large surfaces of a protein, or discontinuous regions of the primary sequence, the selected sequences are less likely to resemble the "natural" ligand. As a result, caution should be taken if one is

planning on using the DNA corresponding to the selected sequences as probes when trying to clone any natural ligand proteins. In contrast, cDNA expression libraries are limited to natural proteins and, as a result, are much more likely to yield the native ligand for the target protein.

***Other Peptide Display Systems*** In the past few years, there have been numerous developments in the peptide display technology to make it applicable to a variety of protein–protein and protein–peptide interactions.

*Bacterial Display Systems* A significant restriction of phage display is the unpredictable expression bias against some eukaryotic protein sequences expressed in *E. coli*, because incorporation of any protein fusion into phage particles depends upon the ability of *E. coli* to express that protein sequence in soluble form. A number of alternatives to phage display have been developed for the detection of protein–protein interactions in bacteria. One approach to physically link a protein to its gene is by fusion with the *lac* repressor [87]. Other methods of isolating peptides by attaching the encoded protein to its gene include affinity selection of nascent proteins translated on polysomes [88] and expression on the surface of bacteria [89]. Proteins or peptides have been displayed on the surface of *E. coli,* fused either to outer membrane proteins [90] or flagella [91]. However, like phage display, *E. coli* display is subjected to potential library bias due to expression in a prokaryotic host.

*Eukaryotic Display Systems* In contrast to *E. coli*, yeast cells possess eukaryotic posttranslational processing enzymes and secretory machinery. Thus, a display system using *S. cerevisiae* as the host organism should overcome the library biases due to expression in a prokaryotic host. The yeast surface display system uses the α-agglutin mating adhesion receptor, which consists of two domains, Aga1 and Aga2 [92], to display recombinant protein on the surface of *S. cerevisiae*. Peptide sequences are fused to the carboxyl terminus of Aga2, and the Aga2 fusion is disulfide bonded to Aga1, which is in turn covalently linked to the yeast cell wall by phosphatidyl inositol glycan linkages. Therefore, each product of a DNA library is tethered to the surface of the yeast cell wall in a manner that makes it accessible to macromolecular recognition without steric hindrance from cell wall components. This could be of significant advantage when displaying peptides as candidate molecules for the disruption of protein–protein interactions (see the second part of this chapter).

A mammalian peptide display system also exists, wherein proteins of interest are targeted and anchored to the cell surface by cloning their genes in frame with an N-terminal secretion signal and the C-terminal transmembrane anchoring domain of platelet-derived growth factor receptor (see Reagents). In addition, a variety of eukaryotic viruses have been also proposed for the display of foreign peptides. Examples of these are the tobacco mosaic virus, in which *Plasmodium malariae* epitopes were displayed both on the surface

of a loop region and at the C-terminus of the coat protein [93], and human rhinovirus 14, on whose surface a library of variants of the V3 loop of human immunodeficiency virus type 1 (HIV-1) gp120 has been displayed [94]. The eukaryotic display of peptide sequences has been successfully achieved also using the baculovirus *Autographa californica*, through expression of foreign peptide sequences as coat protein fusions [95].

*In Vitro Display Technologies* Some new display technologies have recently emerged that operate entirely in vitro and use PCR rather than cells to amplify genetic material [79]. For example, one of these systems relies on the unique properties of a replication initiator protein, P2A, encoded by the A gene of *E. coli* bacteriophage P2. This protein covalently attaches to the same molecule from which it has been expressed. By constructing a peptide library from random sequences linked to the A gene, the expressed P2A-peptide fusion moieties are directly tagged with the peptide coding sequence. These in vitro display technologies, by circumventing the need to introduce DNA into a host cell, could allow the construction of libraries that are several orders of magnitude larger (potentially up to $10^{15}$ molecules) than those obtainable when using whole cells or viruses for peptide display. Moreover, the linkage between the peptide sequence and its encoding nucleic acid leads to several advantages, including a greater control over binding conditions to the target molecule and the ease with which PCR-based mutagenesis and recombination can be performed.

The choice of the most appropriate expression system for the display of peptide libraries ultimately depends on the protein–protein interaction of interest. There would be no advantage, for example, in using a eukaryotic display system for the display of short sequence libraries when phage systems have proven so effective because of their simplicity, minimal cost, and ease of manipulation. On the other hand, eukaryotic display systems could be required for longer peptide sequences whose functional display needs proper posttranslational modifications.

***Reagents*** In the last few years there has been significant progress in developing vectors for peptide display. In early versions of phage display, the foreign DNA was inserted into the phage genome between the N-terminal domains of gene III needed for infection and the C-terminal domain required for morphogenesis. In subsequent versions, the fusion site was moved to the N-terminus, which improved infectivity of the phage. Based on these findings, libraries displaying peptides of various length instead of gene fragments were created by various groups and successfully applied to the identification of protein ligands and epitope mapping.

Another important reason for improving the vectors for display of cDNA libraries was that fusion to either pIII or pVIII requires maintenance of the correct reading frame at both the junction with the signal sequence and with the C-terminus of the phage coat protein. Therefore, full-length cDNAs cannot

be used as they contain stop codons. To overcome this problem, vectors have been developed that only require a single fusion at the end of the cDNA encoding the amino terminus [81, 96].

One of the currently most used versions is the Ph.D. display system, which enables the display of custom peptide libraries on the surface of bacteriophage M13 as coat protein fusions, creating a physical linkage between each displayed peptide and its encoding DNA sequence. This system utilizes the display vector M13KE, which is an M13 derivative with cloning sites engineered for N-terminal pIII fusion, resulting in a valency of 5 displayed peptides per phage. The use of a phage vector, rather then a phagemid, simplifies the intermediate amplification steps since neither antibiotic selection nor helper phage superinfection are required. However, displayed proteins longer than 20 to 30 residues have a deleterious effect on phage infectivity.

Large display libraries can be easily made using the T7Select vector. Peptides up to 50 residues can be displayed in large copy numbers (415 per phage); in the alternative, proteins up to 1200 amino acids can be displayed in low copy numbers (0.1 to 1 per phage) or midcopy numbers (10 per phage). Displayed peptides and proteins do not need to be exported through the periplasm and the cell membrane, as phage assembly takes place inside the *E. coli* cell and mature phages are realesed by cell lysis. Moreover, T7 grows rapidly, which saves time during cloning/screening procedures, and is extremely stable, expanding the variety of agents that can be used in panning.

The FliTrx Random Display Library system employs the same basic route of expression-fusion protein display as does phage display, but without the phage. The vehicle of choice is the FliTrx vector, which positions a diverse library of peptides in a flagellin (*fliC*)-thioredoxine (*trxA*) fusion protein (FliTrx). The recombinant protein is exported and assembled into partially functional flagella and displayed on the bacterial cell surface in a conformationally constrained form due to insertion of the peptide into the active-site loop of the thioredoxine protein. Peptides inserted in this loop have both their N- and C-terminus tethered by the rigid and stable tertiary fold of the thioredoxine molecule. With this system, phage infection and isolation steps are eliminated.

Finally, the pDisplay is a mammalian expression vector that is designed to target recombinant proteins to the surface of mammalian cells. Displayed proteins can be analyzed for their ability to interact with known or putative ligands. Proteins are targeted and anchored to the cell surface by cloning the gene of interest in frame with pDisplay's unique N-terminal cell surface targeting signal and the C-terminal transmembrane anchoring domain of platelet-derived growth factor receptor.

## Protein Microarrays

Another approach for the high-throughput characterization of protein–protein interactions that has been recently developed exploits the protein

microarray technology [97]. Here, proteins are expressed, purified, and attached to the surface of a microarray in a way that preserves their folding and their ability to specifically bind other proteins. Initially, protein chips were utilized to identify yeast genes encoding specific biochemical activities [98]. Recently, Zhu et al. [99] applied this technology to the study of protein–protein interactions.

Advances in microfabrication [100] and surface chemistry methods [101] have all allowed the development of several protein microarray formats, including the commercially available ProteinChips, functional protein arrays, antibody arrays, peptide arrays, interaction arrays, and surface plasmon resonance (SPR)-based arrays.

Apart from protein–protein interactions, protein microarrays can also be used to study other binding activities, such as protein–nucleic acid, protein–small molecule, and protein–drug interactions. Although these are all in vitro assays, the advantage is that the experimental conditions can be well controlled. For example, different cofactors or inhibitors can be included in the assays to vary the stringency of the binding reaction. Another advantage is that these assays are not biased toward abundant proteins. In addition, with adequate detection methods, protein microarrays can be used to identify the protein substrates of various enzymes such as proteases, methyl transferases, protein kinases, and phosphatases [102]. Finally, protein microarrays can be used to identify in vivo posttranslational modifications by probing for specific modifications, such as phosphorylation or glycosylation, using antibodies or lectins, respectively. Protein microarrays also have the potential advantage of allowing analysis of the kinetics of protein–protein interactions via real-time detection methods.

In summary, protein microarrays can be used to globally analyze the activities of proteins including their interaction with proteins, nucleic acids, lipids, carbohydrates, and small molecules. Because of its versatile and miniaturized nature, microarray technology is expected to enormously flourish in the near future. Different types of commonly used protein microarrays are listed in Table 5 and described in the following sections.

*ProteinChips* In the ProteinChip technology, proteins are exposed to chips with different surface features in parallel so that sets of proteins with common properties (hydrophilic, hydrophobic, charged, etc.) will attach to a particular type of surface. After washing to remove unbound proteins, proteins of interest can be enriched on the chip surface by selective washing, and/or protease digestion [103]. ProteinChips require very small sample amounts and can be used directly with biological fluids to provide information on posttranslational modifications, protein structure, and other properties [103, 104]. Currently, ProteinChips have been used for a wide variety of applications including the identification of collagen binding proteins in *Lactobacillus* species [105] and the quantitation of prostate-specific membrane antigen levels [106].

*Functional Protein Microarrays* Another approach for the exhaustive characterization of the biochemical activities of proteins of interest is using functional protein microarrays. In this technique, sets of proteins or even an entire proteome is overexpressed, purified, distributed in an array format, and then assayed for specific activities. These arrays can be used not only to screen for biochemical activities of interest but also to analyze posttranslational modifications and to detect binding to proteins, antibodies, small molecules, and drugs. The latter use has potentially powerful applications in drug discovery.

There are several types of functional protein arrays (reviewed in [102]): nanowells, which are miniature wells; thick absorbent surfaces, such as hydrogels; and solid surface supports, such as glass slides. Nanowell arrays typically contain wells 1 mm or less in diameter. The wells can be made of plastic material, such as polydimethylsiloxane (PDMS), or alternatively can be etched in glass. The nanowells have the advantges of compartmentalizing samples and of reducing evaporation. Using the nanowell arrays, Zhu et al. [98] characterized the kinase–substrate specificity of almost all (119 of 122) yeast kinases using 17 different substrates. The substrates were first covalently attached to the surface of the nanowells, and then the protein kinases were incubated with the substrates and [$^{33}$P]ATP (adenosine 5′-triphosphate). After washing away the kinases and unincorporated ATP, the nanowell chips were analyzed for phosphorylated substrates. Not only known kinase–substrate interactions were detected but also many novel activities were identified.

Glass microscope slides are also widely used for functional protein microarrays, as they are compatible with many commercial scanners. MacBeath and Schreiber [107] used this format to detect antibody–antigen interactions, protein kinase activities, and protein binding to small molecules. The major limitation in functional protein microarrays has been the preparation of proteins to analyze. This requires high-throughput expression and purification methods that can yield a large number of functionally active proteins. This problem was recently overcome (see High-Throughput Protein Production).

*Antibody Microarrays* In antibody microarrays, antibodies are usually spotted onto a glass slide; then the arrays are incubated with a cell lysate or serum to allow the antigens present in the sample to bind to their cognate antibodies. The bound antigen is detected either by using radioactively labeled or fluorescently tagged proteins or by using a secondary antibody against each antigen of interest.

Low-density antibody arrays have been created that quantitate the levels of several proteins in serum [108, 109] and blood [110]. Recently, Sreekumar and colleagues [111] spotted 146 different antibodies on high-density glass arrays to characterize the differential expression of a number of antigens in LoVo colon carcinoma cells. They found that radiation treatment of the cells upregulated the levels of many interesting proteins, including tumor necrosis

**TABLE 5 Overview of Current Protein Microarrays**

| Support | Binding | Advantages | Disadvantages | References |
|---|---|---|---|---|
| Ni-NTA-coated | Affinity binding | Low background; protein binding in uniform orientation; specific, strong, and high-density protein binding | Proteins have to be Hisx6 tagged | 1 |
| Avidin-coated | Affinity binding | Low background; protein binding in uniform orientation; specific, strong, and high-density protein binding | Proteins have to be biotinylated | 2 |
| DNA/RNA-coated | Hybridization | Low background; protein binding in uniform orientation; specific, strong, and high-density protein binding | Difficult production of labeled proteins | 3 |
| Nitrocellulose | Adsorption | High capacity of protein binding, no requirement of protein modification | High background, nonspecific adsorption, low-density protein binding | 4, 5 |
| Polylysine-coated | Adsorption | No requirement of protein modification | Nonspecific binding | 6 |
| Polyvinylidene fluoride | Adsorption | High capacity of protein binding, no requirement of protein modification | Random orientation of surface-bound proteins | 7, 8 |
| 3D gel pad and agarose thin film | Diffusion | No requirement of protein modification, high capacity of protein binding | Difficult to make, not commercially available | 9, 10 |
| Epoxy-activated | Cross-linking | Strong and high-density protein binding, high-resolution detection methods | Protein binding in random orientation | 11 |
| Aldehyde-activated | Cross-linking | Strong and high-density protein binding, high-resolution detection | Protein binding in random orientation | 1, 12 |

| | | | | |
|---|---|---|---|---|
| Gold-coated silicon | Cross-linking | Low background; specific, strong, and high- density protein binding; well suited for SPR and mass spectroscopy detection | Protein binding in random orientation; difficult to make, not commercially available | 13, 14 |
| Polydimethylsiloxane nanowells | Cross-linking | Specific, strong, and high-density protein binding, useful for biochemical assays | Protein binding in random orientation | 11 |


1. Zhu, H., Bilgin, M., Bangham, R., Hall, D., Casamayor, A., Bertone, P., Lan, N., Jansen, R., Bidlingmaier, S., Houfek, T., Mitchell, T., Miller, P., Dean, R. A., Gerstein, M., Snyder, M. (2001). Global analysis of protein activities using proteome chips. *Science*, *293*, 2101–2105.
2. Rowe, C. A., Tender, L. M., Feldstein, M. J., Golden, J. P., Scruggs, S. B., MacCraith, B. D., Cras, J. J., Ligler, F. S. (1999). Array biosensor for simultaneous identification of bacterial, viral, and protein analytes. *Anal. Chem.*, *71*, 3846–3852.
3. Winssinger, N., Ficarro, S., Schultz, P. G., Harris, J. L. (2002). Profiling protein function with small molecule microarrays. *Proc. Nat. Acad. Sci. USA*, *99*, 11139–11144.
4. Joos, T. O., Schrenk, M., Hopfl, P., Kroger, K., Chowdhury, U., Stoll, D., Schorner, D., Durr, M., Herick, K., Rupp, S., Sohn, K., Hammerle, H. (2000). A microarray enzyme-linked immunosorbent assay for autoimmune diagnostics. *Electrophoresis*, *21*, 2641–2650.
5. Ge, H. (2000). UPA, a universal protein array system for quantitative detection of protein-protein, protein-DNA, protein-RNA and protein-ligand interactions. *Nucleic Acids Res.*, *28*, e3.
6. Haab, B. B., Dunham, M. J., Brown, P. O. (2001). Protein microarrays for highly parallel detection and quantitation of specific proteins and antibodies in complex solutions. *Genome Biol.*, *2*, RESEARCH0004.
7. Cahill, D. J. (2001). Protein and antibody arrays and their medical applications. *J. Immunol. Methods*, *250*, 81–91.
8. Walter, G., Bussow, K., Cahill, D., Lueking, A., Lehrach, H. (2000). Protein arrays for gene expression and molecular interaction screening. *Curr. Opin. Microbiol.*, *3*, 298–302.
9. Guschin, D., Yershov, G., Zaslavsky, A., Gemmell, A., Shick, V., Proudnikov, D., Arenkov, P., Mirzabekov, A. (1997). Manual manufacturing of oligonucleotide, DNA, and protein microchips. *Anal. Biochem.*, *250*, 203–211.
10. Afanassiev, V., Hanemann, V., Wolfl, S. (2000). Preparation of DNA and protein micro arrays on glass slides coated with an agarose film. *Nucleic Acids Res.*, *28*, E66.
11. Zhu, H., Klemic, J. F., Chang, S., Bertone, P., Casamayor, A., Klemic, K. G., Smith, D., Gerstein, M., Reed, M. A., Snyder, M. (2000). Analysis of yeast protein kinases using protein chips. *Nat. Genet.*, *26*, 283–289.
12. MacBeath, G., Schreiber, S.L. (2000). Printing proteins as microarrays for high-throughput function determination. *Science*, *289*, 1760–1763.
13. Houseman, B. T., Huh, J. H., Kron, S. J., Mrksich, M. (2002). Peptide chips for the quantitative evaluation of protein kinase activity. *Nat. Biotechnol.*, *20*, 270–274.
14. Bieri, C., Ernst, O. P., Heyse, S., Hofmann, K. P., Vogel, H. (1999). Micropatterned immobilization of a G protein-coupled receptor and direct detection of G protein activation. *Nat. Biotechnol.*, *17*, 1105–1108.

factor-related ligand, DNA fragmentation factors 40 and 45, and p53, and downregulated the levels of other proteins.

The major problem with antibody arrays is antibody specificity. Haab et al. [112] evaluated the reactivity of 115 antibodies with their cognate antigens. Protein microarrays containing either immobilized antibody or immobilized antigen were probed with antigens or antibodies, respectively. Only 30 percent of the antibody–antigen pairs showed specific binding, indicating that most antibodies are not suitable for quantitative detection. Nevertheless, for those antibodies that are specific, quantitation of antigen levels in a complex mixture could be obtained. In fact, in a later study they demonstrated that antibody microarrays could be used to obtain serum profiles [113].

*SPR-Based Arrays* Surface plasmon resonance (SPR) has been shown to be a versatile detection tool to analyze the kinetics of protein–protein interactions over a wide range of affinities, molecular weights, and binding rates [114–116]. Although the commercially available SPR chips are not yet high-throughput, Myszka and Rich [117] recently described a sensor surface with 64 individual immobilization sites in a single flow cell. Alternatively, Sapsford and colleagues [118] developed an antibody array biosensor and studied the kinetics of antigen–antibody interactions. Surface plasmon resonance–biomolecular interaction analysis (SPR–BIA) is another chip-based method that has been used for kinetic and thermodynamic studies of protein–protein interactions. Finally, in an adaptation of biosensor technology, soluble recombinant insulinlike growth factor receptor domains have been isolated on a chip to analyze protein–protein interactions [119]. This allowed the characterization of the binding of protein domains to other proteins and also provided information about the function of nonbinding domains.

*Peptide Microarrays* A modified version of protein microarrays is peptide microarrays, which can be used as potential ligands and as substrates for enzymatic activites when probing with proteins or other molecules. In one recent study, Houseman and colleagues [120] immobilized 9-mer peptide kinase substrates on a gold-coated glass surface to form a high-density peptide microarray and analyzed the phosphorylation of the peptides using fluorescence, SPR, and phosphoimaging. They could also quantitatively evaluate the effect of three kinase inhibitors. This study demonstrated the usefulness of coupling peptide chips with various detection methods to quantitatively analyze the dynamics of enzyme–substrate interactions, which could have potential applications in drug discovery.

*Interaction Arrays* In the field of protein–protein interaction assays, dot-blot filter arrays have been used to screen for specific interactions of immobilized proteins with other proteins. For example, Ge [121] has described a protein array for quantitative detection of protein binding to a variety of proteins, nucleic acids, and small molecules. A set of 48 proteins were spotted onto

nitrocellulose filters and a $^{32}$P-labeled glutathione *S*-transferase (GST) fusion of the human protein p52, known to be a transcriptional activator, was incubated with the array. He detected binding to nucleolin, which has been reported to have several functions, including ribosomal RNA processing and regulation of cell doubling time in cancer cells. In addition, interactions of DNA, RNA, or low-molecular-weight ligands with the immobilized molecules were shown. Such arrays could be miniaturized and therefore have the potential to be created in a microarray format.

Recent work by Zhu et al. [99] demonstrated the extraordinary power of array-based technology for proteomic studies. They cloned ~94 percent (>5800 of 6131) of the yeast ORFs in a yeast expression vector that expresses the proteins as N-terminally GST-Hisx6 double-tagged fusions and developed a high-throughput protein purification method using the GST tags. The purified proteins were then attached to Ni-NTA-coated glass slides using the Hisx6 tags. These microarrays could be utilized to characterize protein–protein interactions on a genome-wide scale. Using calmodulin as a model protein to probe the arrays, many known interactions could be confirmed and a set of novel binding proteins was detected. Analysis of the sequences of these proteins revealed the presence of a binding motif and therefore demonstrated the specificity of the binding interaction. Moreover, experiments aimed at detecting protein–lipid interactions demonstrated that this technology also enables the identification of proteins able to bind low-molecular-weight molecules. This suggests that an entire proteome can be immobilized onto a glass surface to directly screen for interactions with proteins and small molecules and therefore opens the possibility to examine an entire proteome directly for protein–drug interactions.

### *Reagents and Methods*

*High-Throughput Protein Production* A major limiting factor influencing the development of protein arrays is the availability of adequate protein reagents. Proteins are more challenging to prepare for the microarray format than DNA. For successful implementation on chips, proteins need to be chemically stable to the derivatization and retain activity. Protein functionality is often dependent on the state of proteins, such as subcellular localization, posttranslational modifications, and binding to other proteins. Nevertheless, in recent years many research groups and companies have contributed enormous effort in developing high-throughput protein purification methods. The commonly used methods for preparing protein reagents include cloning as fusions to protein tags such as Hisx6 or GST, expression in a suitable host, and high-throughput purification [122]. High-throughput methods have been used to purify proteins from various host cells, including *E. coli*, yeast, insects, and human cell lines, allowing the preparation of microarrays containing over 100 proteins and even an entire proteome [99, 102, 122–128].

Alternatively, proteins can be produced using cell-free expression systems. For example, Keefe and Szostak [129] developed an mRNA display system in which each protein is in vitro translated and covalently linked through its C-terminus to the 3′ end of its coding mRNA. More interestingly, He and Taussig [130] have reported a method for in vitro translation of Hisx6-tagged proteins directly onto Ni-NTA-coated surfaces via a protein in situ array (PISA), enabling single-chain antibodies to be arrayed for high-throughput immunoassays. This method has the advantage of synthesizing in situ proteins that may not be amenable to deposition from solution or to cell-based expression (e.g., insoluble proteins).

Antibodies used for protein microarrays can derive from polyclonal sera or can be hybridoma-derived mAbs, recombinant antibodies, or antibodies selected from phage display libraries.

Finally, while most approaches entail the attachment of purified proteins to solid supports, some approaches adopt on-chip synthesis of probes. These include the high-density peptide chemical synthesis approach. The so-called SPOT technology of immobilized peptides is being developed for a number of surfaces including chips and membranes to study several activities, including protein–protein interactions [131]. Membranes are made functional by chemical derivatization followed by spacer/linker attachment. Peptides are then created by standard Fmoc peptide synthesis. Such a synthetic strategy enables the use of peptides containing unnatural residues.

*Supports* Due to the susceptibility of protein activity to derivatization and immobilization chemistries, considerable effort has been directed toward optimizing the solid surface of protein arrays. It is important that the support retains proteins in an active state at high densities, can be printed in such a fashion that the proteins remain in a moisturized environment, and is compatible with most commercial arrayers and scanners. Soft substrates such as polyvinylidene fluoride (PVDF), polystyrene, and nitrocellulose membranes, which have been used to attach proteins in traditional biochemical assays (e.g., immunoblot and phage display), are often not suitable to protein microarrays [121, 122, 132]. These surfaces often do not allow high protein density, as the spotted material may spread on the surface, and/or they may not allow optimal signal-to-noise ratios [123, 127, 128, 133]. These problems have been overcome using other materials that have low fluorescence background and thus are compatible with most assays. The nature of the surface chemistry varies and includes porous acrylamide gel, membranes, plain glass, polymeric microwells, nanowells, and three-dimensional (3D) surface structures (Table 5).

Polyacrylamide gel and agarose microarrays, patterned on a glass surface, have been created by Guschin et al. [134] and Afanassiev et al. [135], respectively. Because both acrylamide and agarose tend to form hydrophilic and highly porous matrices, proteins can readily diffuse and therefore need to be immobilized by cross-linking to reactive ligands present in the matrix. Analytes are then added to these 3D arrays to perform the assays [134]. Because of the formation of 3D matrixes on the glass surface, the binding capacity is

higher than that on a 2D surface; moreover, the moisturized environment prevents protein denaturation and thereby helps keep proteins in their active state. In addition to the sophisticated processes required to fabricate such 3D matrixes, the major disadvantage of the 3D arrays is that it is more difficult to change buffers and recover trapped molecules from these microarrays [102]. In contrast to 3D arrays, Zhu et al. [98] created an open nanowell structure on a PDMS surface supported by standard glass slides. The nanowells significantly reduce evaporation and minimize cross-contamination and background. Thanks to the open nanowell structure, different reagents and buffers can be sequentially added, which is essential for some biochemical assays. In addition, bound molecules can be easily recovered from the nanowells. Moreover, by covering the nanowells with gold, mass spectrometry and SPR analyses can be performed. The major disadvantage of this technology is that specialized equipment is required to load the nanowells at high density. Many researchers now directly array proteins and antibodies/antigens onto plain glass slides [98, 99, 107, 112, 136, 137]. To keep proteins in a wet environment during the arraying process, glycerol (30 to 40 percent) is added to the sample buffer and the spotting is performed in a moisturized environment [99, 107].

An additional consideration concerns the optimization of the chemistry used to attach the proteins to the support, as certain proteins are more prone to nonspecific adsorption and loss of functionality upon exposure to solid phases. To immobilize proteins to a solid substrate, the surface of the substrate has to be modified to achieve the maximum binding capacity. A convenient method is to coat the glass surface with a thin nitrocellulose membrane or poly-lysine such that proteins can be passively adsorbed to the modified surface through nonspecific interactions [121, 136, 138]. The attached proteins lay on the surface in random orientation and can be washed off under stringent washing conditions. Moreover, the background level is usually higher because of nonspecific attachment.

To achieve more specific and stronger protein binding, several groups have created reactive surfaces on glass that can covalently cross-link to proteins [98, 99, 107]. Generally, a self-assembled monolayer (SAM) is formed using a bifunctional silane cross-linker, which has one functional group that reacts with the hydroxyl groups on glass surface and another free one that can either directly react with primary amine groups of proteins (i.e., aldehyde or epoxy groups) or can be further chemically modified to reach maximum binding specificity [107, 139]. Gold-coated glass surface is another variation [120, 140]. To form a SAM on a gold surface, bifunctional thio-alkylene is usually utilized, which has a SH group that reacts with gold and another free one that reacts with target molecules. The advantage of using gold-coated surface is that mass spectrometry and SPR can eventually be used as detection methods to identify the bound molecules or to monitor the dynamics of the reactions, respectively [120, 140, 141]. This approach provides the opportunity to study dynamics of biochemical reactions in a high-throughput fashion and thus has great potential in drug discovery and biomedical research [141].

In the above covalent cross-linking approaches, because the reactive groups are also present in the side chains of proteins, it is conceivable that the proteins bind to the surface in a random fashion, which may alter the native conformation of proteins, reduce the activity of proteins, or make them inaccessible to interacting partners. Perhaps the best means of protein attachment is via highly specific affinity interactions [98, 133]. Proteins fused with a high-affinity tag at their N- or C-terminus are linked to the surface of the chip through this tag, and, thus, all of the attached proteins should orient uniformly away from the surface [99]. Using this method, immobilized proteins are more likely to remain in their native conformation, while the analytes have easier access to the active sites of proteins. This approach was first successfully demonstrated in attaching 5800 fusion proteins containing a Hisx6 tag onto a Ni-NTA-coated glass slide [99] (see, above, Interaction Arrays). Other affinity methods such as glutathione/GST and phosphonate/serine esterase cutinase ligand/protein tags have been also proposed [142].

*Protein Delivery Systems (Arrayers)* Arrays are created by spotting proteins onto the solid surface in an organized high-density format followed by immobilization to create a stable probe. Although a 96-format dot-blot instrument has been used to create low-density protein arrays on filters [121, 123, 138], high-density protein microarrays (>30,000 spots per slide) can be obtained using robotic printing tools ("arrayers"). Considerable expertise in this field has been achieved from work with DNA microarrays using photolithography, ink jet and contact printing, liquid dispensing, and piezoelectrics [143]. All of these methodologies have advantages and disadvantages related to the accuracy of dispensing and the compatibility with high-throughput arraying of proteins.

The contact printing arrayers deposit subnanoliter sample volumes directly to the surface using tiny pins with or without capillary slots. Because these contact printing robots cannot align their pins to the prefabricated structures and need to touch the surface, noncontact robotic printers, which use inkjet technology, were used to deliver nanoliter to picoliter protein droplets to nanowells [98] and polyacrylamide gel packets [144]. Recently, electrospray deposition technology was applied to deliver dry proteins to a dextran-grafted surface [145]. This technology further decreased the spot size from ~150 to ~30 μm.

Other successful reports of protein arraying include a system for automated high-density protein delivery onto PVDF filters from liquid bacterial cultures [127]. Use of a thin (250 μm) transfer stamp enabled high spotting density (4,800 spots per slide) and detection of 10 pg of protein.

Finally, a group from Harvard University [107] has recently reported a method for generating high-density (16,000 spots/cm$^2$) protein arrays onto chemically derivatized (aldehyde-containing silane) glass slides in low (nanoliter) volume. The proteins were deposited onto the array in the presence of glycerol to avoid loss of functionality due to dehydration. The approach is sug-

gested where it is intended to study protein–protein interactions or to carry out protein substrate identification.

*Detection Methods* The detection devices used in protein microarrays include charge-coupled device (CCD) cameras or laser scanners with confocal detection optics. Fluorescence detection methods are generally the preferred detection method because they are safe, simple, and extremely sensitive and can have very high resolution [128, 133]. Typically, a chip is either directly probed with a fluorescent molecule (e.g., fluorescently labeled protein) or in two steps by first using a tagged probe (e.g., biotin), which can then be detected using a fluorescently labeled affinity reagent (e.g., streptavidin). Some researchers made use of fluorescent cyanine dyes (e.g., Cy5) on antibody arrays, detecting human myeloma proteins [146]. Cyanine dyes were also used for a recent study of protein–protein interactions [147]. Another fluorescent labeling method is rolling circle amplification (RCA), which is extremely sensitive [148]. However, other detection methods have been proposed. For example, ELISA (enzyme-linked immunosorbent assay) was first used to detect proteins for both glass arrays [134] and filter arrays [124, 136]. Zhu et al. [98] and Ge [121] have used radioisotope labeling to study kinase–substrate interactions in nanowells, and protein–protein, protein–DNA, and protein–drug interactions on filter arrays, respectively.

Because labeling molecules can sometimes affect protein activity and are restricted to the available detection channels, nonlabeling methods have advantages as a direct detection approach for protein microarrays. As an example, the surface-enhanced laser desorption and ionization (SELDI) technology adapted mass spectrometry as a readout system to analyze differential protein expression on spot arrays [149, 150]. Protein extracts are incubated on different spots of the same adsorptive surface chemistry (e.g., cation–anion exchange material, hydrophobic surface, etc.). After washing away unbound proteins, the whole variety of nonspecifically captured target proteins can be analyzed using SELDI mass spectroscopy.

Label-free SPR-based detection systems can also be used [114, 151]. Intrinsic Bioprobes have linked a SPR-based biosensor to MALDI-TOF (matrix-assisted laser desorption ionization–time of flight) mass spectrometry, designed as a chip-based proteomic tool for providing information about the function and structure of proteins [101]. This allows the real-time analysis of protein interactions with a sensitivity in the low femtomolar range.

Time-resolved fluorimetry (TRF) offers the advantage of greater sensitivity due to a lower background signal [152]. This is achieved by the use of pulsed excitation and time-gated detection. An adaptation of an immunoassay recently used antibodies to detect protein on a solid surface. The antibodies were covalently attached to a unique oligonucleotide primer. This novel approach permitted a rolling circle amplification reaction for increased sensitivity by PCR amplification of the associated DNA tag [137, 153, 154].

Atomic force microscopy (AFM) is a powerful detection technology capable of detecting protein interactions at the level of a single molecule [155]. Protiveris is developing protein biosensors, based on chemomechanical actuation of silicon microcantilevers in a principle similar to AFM. This unique technology should provide very high sensitivity without the need for derivatization. In a recent example applied to a biological system, microcantilever-based arrays have been used to detect prostate-specific antigens [156].

## Protein Mass Spectrometry

For organisms with known genomes, recent advances in protein mass spectrometry (MS) have vastly facilitated the dissection of interactions in protein complexes [157]. Proteins and tryptic peptides from these complexes can be analyzed by MALDI–TOF [157], sequences deducted from their mass, and the sequences compared with a database of predicted proteins encoded by the organism's genome. If mass alone cannot predict the exact sequence, fragmentation methods (nanoelectrospray tandem mass spectrometry) can be used to produce stretches of up to 16 residues of sequence from femtomolar amounts of protein fragments [158].

So far, such methods have been used to identify caspase 8 as an interactor with immunoprecipitated CD95 [159] and to characterize the protein complement of a number of multiprotein complexes including the yeast spindle pole body [160], the yeast spliceosomal U1 small nuclear ribonucleoprotein (snRNP) [161], the yeast anaphase promoting complex [162], the spliceosome [163], and *trans*-Golgi network-derived transport vesicles [164].

Clearly, mass spectrometric methods have tremendous potential. As its sensitivity and ease of use increase, MS will come to complement biological methods for detecting and analyzing protein interactions and may eventually supplant them. However, other than its considerable cost and the high level of technical sophistication required, the most significant limitation of MS for analysis of protein interactions is that the protein complexes first need to be isolated by physical methods such as electrophoresis [158]. Many proteins will not be detected because, for example, they are too small, large, scarce, alkaline, or acidic to be analyzed by two-dimensional gel analysis, or because the interaction is too weak and transient to allow affinity purification.

### *Reagents and Methods*

*Separation and Isolation of Proteins by 2D PAGE* An established technique for protein separation is two-dimensional polyacrylamide gel electrophoresis (2D PAGE), which separates proteins based on two properties: their size and their isoelectric point. Two-dimensional PAGE is capable of profiling many thousands of proteins on a single matrix with exquisite resolution, separating isoforms differing by a single posttranslational modification. Posttranslational modifications detectable by 2D PAGE include deamidation [165], phosphorylation [166], and glycosylation [167, 168]. Thus, 2D PAGE is an open assay

approach, which allows the separation and detection of proteins from a wide variety of sources without the need for any prior knowledge of function.

In recent years, there have been several technical advances that have led to improvement in the quality of 2D PAGE separations. Most notably, these include the introduction of immobilized pH gradients (IPGs), which provide better reproducibility and increased resolution [169–172]. Originally, the pH gradient for isoelectric focusing in 2D PAGE was created by carrier ampholytes. These are mixtures of a few hundred different homologs of amphoteric buffers, synthesized together in one reaction flask. The mixtures contain buffers with isoelectric points evenly distributed over a wide pH spectrum (i.e., from pH 3 to pH 10), and they have high buffering power at their isoelectric points. When an electric field is applied, they begin to migrate according to their charges toward the anode or the cathode, and automatically form stable pH gradients. The pH gradient works well when protein separation is performed in native conditions. However, in order to obtain a high-resolution 2D protein separation, electrophoresis is usually performed under denaturing conditions, which destabilizes the pH gradient and increases the running time.

Today, with the development of IPGs [173], the pH gradients in these gels are prepared by copolymerizing acrylamide monomers with acrylamide derivatives containing carboxylic and tertiary amino groups. Because the buffering groups that form the gradient are fixed, the pH gradient cannot drift and is not influenced by the sample composition. The use of IPGs has allowed many methodical innovations for 2D PAGE. Premanufactured gel strips and instruments are available as commercial products. The immobilized strips are of various lengths and cover various pH ranges, including broad pH range (pH 3 to 10), narrower ranges (pH 4 to 7 and pH 6 to 11), and even a range of a single pH unit increment. The method can be applied to different experimental purposes, and the separation can be improved by using larger gel format and narrow pH range immobilized strips.

The ability of 2D PAGE to resolve thousands of proteins in one gel makes it an unbeatable technology, and it is expected to remain in use for another decade. However, although a powerful technique, there are limitations to 2D PAGE-based approaches. First, it is rather labor intensive and requires a high level of skill. Moreover, 2D PAGE has proven resistant to automation [174, 175] and image analysis remains a bottleneck. Second, without prior enrichment the approach can lack the sensitivity of other assay formats, such as ELISA [176, 177]. Thus, scarce proteins either cannot or can hardly be detected on 2D gels. Although this problem can be circumvented by increasing the sample amount, there may be a risk of overloading the system and reducing the resolution [177]. Finally, proteins with extreme pH (below 3 or above 10) are not separated [169], but focused as vertical lines on both sides. In addition, many large or hydrophobic proteins do not enter the gel during the first dimension and are therefore refractory to separation and quantitation by 2D PAGE [178, 179].

*Methods of Protein Detection and Image Analysis* There are many ways to detect proteins in 2D gels. The most common methods are Coomassie blue staining [180, 181] and silver staining [182, 183]. Other methods, including [$^{35}$S]-Met or $^{14}$C radiolabeling [184, 185], zinc imidazole [186, 187], colloidal gold [188, 189], India ink [97, 190], ponceau S, and amido black [181], can also be used in different applications to achieve better sensitivity in particular cases. However, there are some drawbacks. For example, glycoproteins are not stained by Coomassie blue [180], and many organic dyes are unsuitable for protein detection on PVDF membranes if samples are to be used for MALDI-TOF.

After 2D electrophoresis and protein visualization by staining, images of gels are digitized for computer analysis by an image scanner or fluorescent scanner and are then subjected to analysis by special image analysis software (either PDQUEST from BioRad or ImageMaster from Amersham Pharmacia Biotech). The 2D patterns are very complex, and special software tools are required to find differentially expressed proteins in a series of gels, such as posttranslational modified proteins or up- and downregulated proteins. Image spots on the gels are initially detected, manually edited, and then matched.

*LC-MS-Based Approaches* An alternative to 2D PAGE for the separation of proteins is liquid chromatography (LC). Reverse-phase high-pressure liquid chromatography (RP-HPLC) has the advantage of using a liquid phase compatible with direct injection into a mass spectrometer (LC-MS) for identification. Unfortunately, proteins are not generically amenable to high-quality separation by RP-HPLC. This problem has been solved by digesting the proteins to peptides with a protease, usually trypsin. The resulting peptides behave in a more predictable manner than the intact proteins and can be readily separated by RP-HPLC. This approach has meant that analysis of protein mixtures by multidimensional LC in conjunction with MS is becoming feasible [191]. To date, the LC-MS approach has been successfully used to characterize protein complexes [191, 192] and to specifically look at protein–protein interactions [193].

*Protein Identification Using Mass Spectrometry* After resolving the protein mixtures, the next step is protein identification. Proteins can usually be identified by peptide mass fingerprinting (PMF) [187, 194–196] and database searching. In PMF, the peptide masses of unknown proteins are compared to the predicted masses of peptides from the theoretical digestion of proteins in a database. First, the protein spots are excised and digested with a protease (e.g., trypsin or chymotrypsin). The digest is then applied onto a sample plate and coated with matrix. The matrix is typically a small energy-absorbing molecule such as α-cyano-4-hydroxycinnamic acid or 2,5-dihydroxybenzoic acid. The analyte is spotted, together with the matrix, on the sample plate and allowed to evaporate, resulting in the formation of crystals. The plate is then put into the MALDI-TOF mass spectrometer, and the laser is automatically

targeted to specific places on the plate, and peptide mass spectra are then obtained. If the protein is digested with trypsin, the trypsin autolytic fragment peaks can serve as internal standards for mass calibration. Several software packages are available to perform database matching such as Protein Prospector [197], ProFound [198], and MASCOT [199] (see Table 6). The more numbers of peptides match to a protein in database, the more likely the protein identification is. The advantage of using PMF is that the protein identification process is fast and user-friendly. But the success of the method can be compromised by several factors: (1) mass redundancy of peptides with the same masses but different amino acid compositions can cause ambiguity in protein identification; (2) PMF cannot identify posttranslationally modified peptides since such information is not present in the database; (3) PMF cannot analyze samples containing a mixture of proteins since they generate mixtures of peptides after the digestion; and (4) insufficient peptides are obtained in the PMF to submit to the database search, that is, there is insufficient data to identify the unknown protein.

When such problems occur, it is necessary to use postsource decay (PSD) to confirm the result. Because PSD can deduce the amino acid sequence of peptides from normal, posttranslationally modified, or novel proteins of interest, this can significantly increase the accuracy of the protein identification and sometimes leads to the discovery of new proteins. With peptide sequencing, the amino acid sequence of unknown peptides can be obtained and then used to search the database to identify the protein from which it was derived. Unlike PMF, PSD can be used for gels containing more than one protein. This advantage greatly improves the protein identification process since protein bands from one-dimensional (1D) gel can be identified whether homogenous or not. The PMF data can be supplemented with partial amino acid sequence information along the result of database search, so an unsuccessful search of the protein database with the PMF data may be reversed with an additional partial sequence. The drawback of PSD is that it is not user-friendly since the process is not easily automated. As a result, MS analysis and database searching takes considerable time and must be performed by an experienced operator.

*Affinity Tagging and Mass Spectroscopy* Affinity purification has long been used to study interaction of several proteins in a multiprotein complex. Several methods have been used for isolating protein complexes, of which the most common is the use of GST fusions [200] where the protein of interest whose binding partners need to be determined is expressed as a fusion protein with a cleavable GST tag and immobilized to a solid support. The immobilized protein is incubated with a cell extract containing the target proteins so that multiprotein complexes can be formed. The protein complexes are then cleaved off from the GST tag, purified, and the bound proteins identified by 2D PAGE and MS [200].

Because no a priori knowledge of the interacting proteins is needed, this approach can be used to identify novel interactions between known proteins

**TABLE 6 Publicly Available Database Matching Tools for Protein Identification from Mass Spectrometry Data**

| Name | URL |
|---|---|
| Mascot | *http://www.matrixscience.com* |
| Mowse | *http://www.hgmp.mrc.ac.uk/Bioinformatics/Webapp/mowse* |
| MS-Tag | *http://www.prospector.ucsf.edu* |
| Pepsea | *http://www.pepsea.protana.com* |
| ProFound | *http://www.proteometrics.com* |
| Protein Prospector | *http://www.prospector.ucsf.edu/* |
| SEQUEST | *http://www.fields.scripps.edu/sequest* |
| Sonar | *http://www.proteometrics.com* |

in a cellular pathway or new interacting partners. The purification of a protein complex not only helps to identify its components but also provides information on the relationship between apparently unrelated cellular pathways [200]. A crucial requirement for the characterization of protein complexes is high affinity between the tagged protein and the other complex elements to ensure that components of the protein complex are not lost during the purification process [201].

Using this approach, the human spliceosome multiprotein complex has been isolated using biotinylated RNA as a bait and analyzed with a 2D gel to identify 19 novel factors [163]. Other protein complexes that have been characterized using this approach include proteins found in the yeast nuclear-pore complex and the GroEL chaperonin system [202].

The high-throughput potential of this approach has been demonstrated by two recent reports on the systematic characterization of protein complexes in yeast [203, 204]. These two high-throughput studies differed in their experimental design. In one study, Ho et al. [204] constructed 725 inducible FLAG epitope–tagged fusions, which were overexpressed and purified together with their associated proteins. This study identified more than 3000 protein–protein interactions involving 1578 yeast proteins. In another study, Gavin et al. [203] integrated a tandem affinity purification (TAP) tag cassette by homologous recombination at the 3′ end of each ORF and used haploid growing cells. The advantages of this strategy are that in the resulting yeast strains there is no competition for the "untagged" protein and that expression is under the control of the endogenous promoters. After TAP and sodium dodecyl sulfate (SDS)–PAGE, peptide mass fingerprinting by MALDI-MS was used to identify the proteins. Starting with a set of more than 500 chromosomal tagged genes, Gavin et al. [203] were able to purify and subsequently characterize 232 protein complexes encompassing 1440 yeast proteins. The differences in the results of these two studies likely reflect the fact that one study used endogenous protein levels, whereas the other used overexpressed proteins. The overexpressed proteins are likely to interact with more proteins, but they may also

associate with proteins that they do not normally bind and thereby yield false positives.

Compared with the yeast two-hybrid approach and protein microarrays, affinity purification coupled to MS detection offers some major advantages but also has some drawbacks. On the whole, this approach is more physiological because protein–protein interactions are analyzed in vivo, rather than reconstituted ex vivo or in vitro. The approach is not restricted to one particular cell type or organism. Importantly, it also reveals which proteins and isoforms belong to a given cellular proteome. Whereas in the yeast two-hybrid system both binding partners are expressed as fusion proteins, in the affinity purification approach only one component of the complex is expressed as a fusion protein, reducing the probability of steric interference. On the other hand, a clear advantage of the two-hybrid system and of protein microarrays is that these strategies are economical and more easily amenable to automation for high-throughput approaches. They also are better suited for the detection of transient or unstable interactions. It is important to note that the yeast two-hybrid system and protein microarrays yield information about binary physical interactions, which are not necessarily mapped by affinity purification approaches and vice versa. Therefore, all methods are highly complementary and should be applied in combination to produce a comprehensive protein–protein interaction map of a given cell.

## Computational Methods of Analysis of Protein–Protein Interactions

Computational methods play an important role at all stages of the characterization of protein–protein interactions. They are used to predict putative interactions, to validate the results of high-throughput interaction screens, and to analyze the protein networks inferred from interaction databases.

Publicly accessible databases of protein–protein interactions are most useful for the analysis of protein interaction data. Several databases that are currently available (Table 7) provide access to both experimental data and the results of different computational methods of inference. For example, molecular interaction (MINT) is a database designed to store data on functional interactions between proteins. Beyond cataloging binary complexes, MINT was conceived to store other types of functional interactions, including enzymatic modifications of one of the interacting partners. Presently, MINT stores 4568 interactions, 782 of which are indirect or genetic interactions [205]. The database of interacting proteins (DIP) documents experimentally determined protein–protein interactions. At the moment, it contains 10,432 interactions, 90 percent of which are from high-throughput experiments in microorganisms [27]. Entries describing interactions among proteins from mammalian proteomes are approximately 750. BIND (Biomolecular Interaction Network Database) has a somewhat larger scope and provides information about bimolecular interactions, complexes, and pathways [206]. BIND contains 5939 interactions, approximately 300 of which describe interactions among

**TABLE 7 Databases of Protein–Protein Interactions**

| Database | URL | Comments | References |
|---|---|---|---|
| BIND | *http://www.bind.ca* | Interaction database that archives biomolecular interaction, complex and pathway information | 1 |
| DIP | *http://dip.doe-mbi.ucla.edu* | Collections of experimentally determined protein–protein interactions | 2 |
| InterDOM | *http://InterDom.lit.org.sg* | This database focuses on providing supporting evidence for the detected protein interactions based on putative protein domain interactions | 3 |
| LiveDIP | *http://dip.doc-mbi.ucla.edu/ldip.html* | Extension of DIP providing access to information on functional states of protein complexes | 4 |
| MINT | *http://cbm.bio.uniroma2.it/mint* | | 5 |
| MIPS | *http://mips.gsf.de* | *S. cerevisiae* specific; also provides information on genetic interactions | 6 |
| PREDICTOME | *http://predictome.bu.edu* | Compilation of functional link predictions with experimental, genome-scale yeast two-hybrid data | 7 |

| | | | |
|---|---|---|---|
| STRING | *http://www.bork.embl-heidelberg.de/STRING* | Compilations of functional link predictions based on gene proximity, common evolutionary history (phylogenetic profiles), and domain fusion events | 8 |
| The GRID | *http://biodata.mshri.on.ca/grid/servlet/Index* | Compilation of BIND, MIPS, and several genome-scale datasets; *S. cerevisiae* specific | 9 |


1. Bader, G. D., Donaldson, I., Wolting, C., Ouellette, B. F., Pawson, T., Hogue, C. W. (2001). BIND—The Biomolecular Interaction Network Database. *Nucleic Acids Res.*, *29*, 242–245.
2. Xenarios, I., Salwinski, L., Duan, X. J., Higney, P., Kim, S. M., Eisenberg, D. (2002). DIP, the Database of Interacting Proteins: A research tool for studying cellular networks of protein interactions. *Nucleic Acids Res.*, *30*, 303–305.
3. Ng, S. K., Zhang, Z., Tan, S. H., Lin, K. (2003). InterDom: A database of putative interacting protein domains for validating predicted protein interactions and complexes. *Nucleic Acids Res.*, *31*, 251–254.
4. Duan, X. J., Xenarios, I., Eisenberg, D. (2002). Describing biological protein interactions in terms of protein states and state transitions: The LiveDIP database. *Mol. Cell. Proteomics*, *1*, 104–116.
5. Zanzoni, A., Montecchi-Palazzi, L., Quondam, M., Ausiello, G., Helmer-Citterich, M., Cesareni, G. (2002). MINT: A Molecular INTeraction database. *FEBS Lett.*, *513*, 135–140.
6. Mewes, H. W., Frishman, D., Guldener, U., Mannhaupt, G., Mayer, K., Mokrejs, M., Morgenstern, B., Munsterkotter, M., Rudd, S., Weil, B. (2002). MIPS: A database for genomes and protein sequences. *Nucleic Acids Res.*, *30*, 31–34.
7. Mellor, J. C., Yanai, I., Clodfelter, K. H., Mintseris, J., DeLisi, C. (2002). Predictome: A database of putative functional links between proteins. *Nucleic Acids Res.*, *30*, 306–309.
8. von Mering, C., Huynen, M., Jaeggi, D., Schmidt, S., Bork, P., Snel, B. (2003). STRING: A database of predicted functional associations between proteins. *Nucleic Acids Res.*, *31*, 258–261.
9. Breitkreutz, B. J., Stark, C., Tyers, M. (2003). The GRID: The General Repository for Interaction Datasets. *Genome Biol.*, *4*, R23.

mammalian proteins. Even if these databases are still largely incomplete, they have provided bioinformaticians and biologists in general with a unique and easily searchable depository of interaction information.

## 3 DISCOVERY OF INHIBITORS OF PROTEIN–PROTEIN INTERACTIONS

Protein–protein interfaces are considered to be new prospective drug targets. Indeed, given the ubiquitous nature of protein–protein interactions, and the knowledge that inappropriate protein–protein binding can lead to disease, it is not surprising that protein–protein interactions have attracted the attention of scientists in the pharmaceutical industry and elsewhere who are interested in producing inhibitors for use as biochemical tools or therapeutic agents. There are ample examples in the literature of the use of antibodies, dominant negative proteins, or peptides to inhibit particular protein–protein assemblies. In contrast, there are only a few examples of small "druglike" molecules that disrupt protein–protein interactions and can act inside cells, but the number of successes is rapidly growing [207, 208].

A variety of strategies can be adopted to identify and characterize compounds that disrupt protein–protein interactions. Here, only recent advances in screening strategies that employ variants of the technologies reviewed in the first part of this chapter are discussed.

### Two-Hybrid System Variants

Variants of the classical yeast two-hybrid system known as *reverse two-hybrid* [1, 209] and *split-hybrid* [210] systems have been recently developed for identifying peptides and/or small molecules that dissociate protein–protein interactions. These extensions of the yeast two-hybrid system from a basic tool of exploratory research to a tool for drug discovery take advantage of several key aspects of yeast. *S. cerevisiae* provides a robust organism that can easily be manipulated by molecular techniques, the complete genome information is available, reagents are inexpensive, and assays are easily configured to automation for high-throughput screen design. On the other hand, a primary constraint of these permutations of the two-hybrid system may be the permeability of yeast to inhibitor molecules. Therefore, the applicability of these systems seems to be limited to small molecules and, in any case, preliminary studies should be performed to evaluate the permeability properties of each molecule. If standard yeast strains prove relatively impermeable to the desired molecule, this barrier may be circumvented by utilizing yeast permeability mutants such as *PDR5*, *erg6*, or *SNQ2* [211], although the viability of such strains may present additional complications. Alternatively, inhibitory peptides dissociating a target protein–protein interaction could be identified by a *reverse three-hybrid system,* wherein yeast cells expressing the two interacting

fusion proteins are co-transformed with a third vector expressing the peptide library in a nonfused form or fused to the SV40 T-antigen NLS, by analogy to three-hybrid systems that we and others recently proposed to detect ternary protein–protein interactions [48, 61].

***Reverse Two-Hybrid Systems*** The reverse two-hybrid system is an "upside-down" version of the classical yeast two-hybrid system in which the interaction between target proteins causes the activation of a toxic or lethal marker gene. All reverse two-hybrid systems make use of yeast strains in which productive protein–protein interactions increase the expression of a counter-selectable marker that is toxic under particular conditions. Under these conditions, dissociation of an interaction provides a selective advantage, thereby facilitating detection. A few growing yeast colonies in which hybrids fail to interact can be identified among millions of nongrowing colonies expressing interacting proteins.

Several reverse two-hybrid systems have been proposed. The first reverse two-hybrid system uses a yeast strain that is resistant to cycloheximide due to the presence of a mutant *CYH2* gene [39]. This strain also contains the wild-type *CYH2* allele under the transcriptional control of the GAL1 promoter. The *CYH2* gene encodes the L29 ribosomal protein and confers cycloheximide sensitivity [212] and is dominant over the *cyh2* allele, which produces cycloheximide resistance. Using this reporter gene, a productive protein–protein interaction of GAL4-based fusion proteins drives expression of the *CYH2* gene, and the yeast strain becomes sensitive to cycloheximide, resulting in abrogation of cell growth when cells are plated on selective media containing cycloheximide. Disruption of the protein–protein interaction results in a failure of *CYH2* gene transcription, and cells retain the cycloheximide-resistant phenotype; thus, the strain is able to grow on media containing cycloheximide.

Another reverse two-hybrid system makes use of the counterselectable marker *URA3* (Fig. 5). The *URA3* gene encodes an enzyme, orotidine-5′-phosphate decarboxylase [213], involved in uracil biosynthesis, that confers uracil prototrophy allowing cell growth on uracil-deficient media. The same enzyme can also catalyze the conversion of a nontoxic analog, 5-fluoroorotic acid (5-FOA), into a toxic product, 5-fluorouracil [214]. This conveniently allows both positive and negative growth selection, on medium lacking uracil or on medium containing 5-FOA, respectively. Hence, mutations that prevent an interaction can be selected from large libraries or randomly generated alleles. Similarly, molecules that dissociate or prevent an interaction could be selected from large libraries of peptides or compounds [209, 215].

In the past few years, the reverse two-hybrid system has been applied successfully to the selection of dissociative molecules. In the first description of this approach, Vidal et al. [209] demonstrated that peptides corresponding to portions of adenovirus E1A protein were able to disrupt the pRB/E2F1 interaction. In another study, the interaction between the activin receptor R1 and

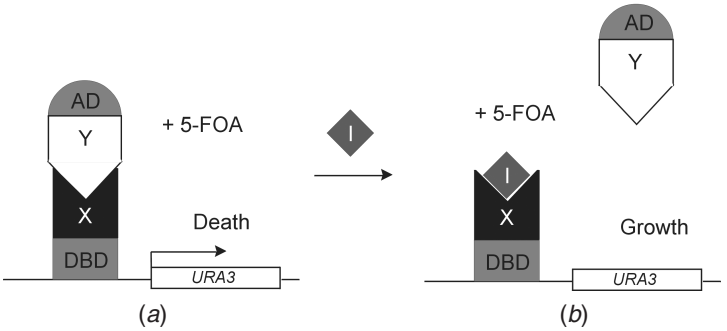


**Figure 5** Reverse two-hybrid system designed to identify small-molecule inhibitors of protein–protein interactions. (*a*) Interaction of protein X, fused to the DNA binding domain (DBD), with protein Y, fused to the activating domain (AD) of a transcription factor, recruits the AD proximal to a promoter containing DBD binding sites upstream of an *URA3* reporter gene. This reconstitutes a functional transcriptional activator that in turn induces the synthesis of Ura3p, leading to yeast death in medium containing the pro-toxin 5-fluoroorotic acid (5-FOA). (*b*) If a small molecule (I, inhibitor) disrupts the intracellular protein–protein interaction, proximity of the AD to the promoter is removed, and transcription of the *URA3* gene is abolished, allowing cell survival in the presence of 5-FOA.

the immunophilin protein FKBP12 was shown to be dissociated by nanomolar concentrations of the small molecule FK506 [216]. Recently, a novel compound, WAY141520, which inhibits the interaction between β3 and α1B subunits of N-type calcium channels, has been selected from a collection of ~150,000 molecules [217]. This study demonstrated that small-molecule inhibitors can indeed be identified from large libraries using the reverse two-hybrid system.

***Split-Hybrid System*** An alternative approach to inhibition studies of protein–protein interactions, which has been termed the *split-hybrid system*, uses the *E. coli* tetracycline repressor (TetR) and operator as a coupled two-step sequential reporter gene system [210]. In this system, expression of TetR is induced following protein–protein interaction. TetR consequently binds to a *tetR* operator-regulated *HIS3* gene, repressing its transcription (Fig. 6). In these conditions, cell growth is inhibited on media deficient in histidine. Dissociation of the target protein–protein interaction therefore prevents *tetR* transcription, which in turn restores *HIS3* synthesis and allows cells to grow in the absence of histidine. The system can be modulated by addition of tetracycline, which relieves the repression in a dose-dependent manner, to selection media. The system's utility was demonstrated by testing interactions of cyclic adenosine 5′-monophosphate (cAMP) response element binding protein (CREB) and its coactivator protein CREB binding protein (CBP). The interaction of CREB and CBP depends on phosphorylation of a critical

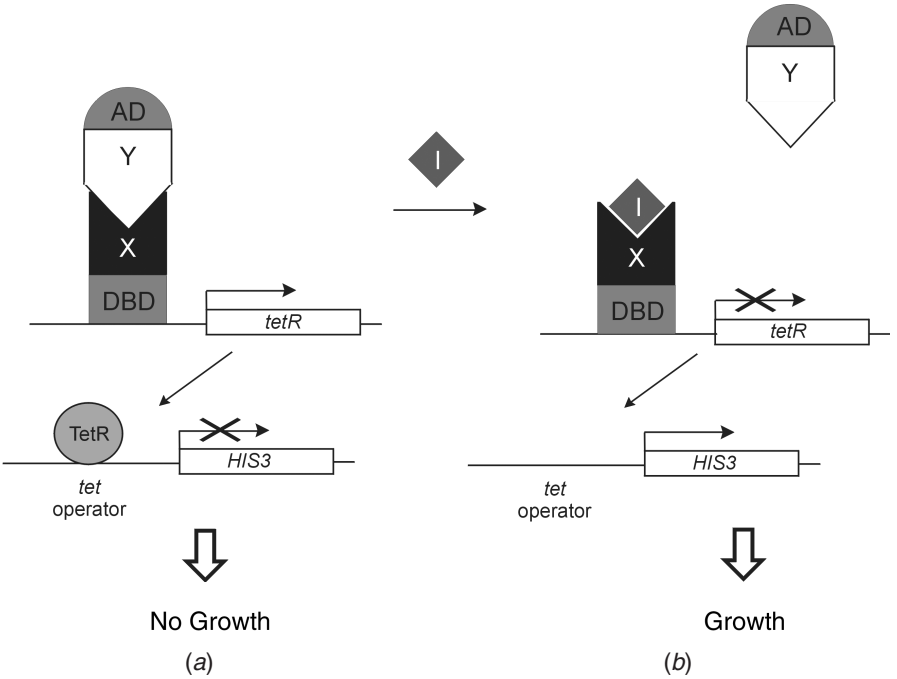


**Figure 6** General schematic of the yeast split-hybrid system. (*a*) The interaction between one protein (X) fused to a DNA binding domain (DBD) and another (Y) fused to a transcriptional activation domain (AD) promotes the expression of tetracycline repressor (TetR), which binds to *tet* operator upstream of the *HIS3* reporter gene, preventing the growth of yeast cells in histidine-deficient media. (*b*) The disruption of the X–Y interaction by an inhibitor molecule (I) prevents expression of TetR. This, in turn, allows expression of *HIS3* reporter gene and hence growth of yeast cells in histidine-deficient media.

serine residue in CREB. Coexpression of wild-type CREB and CBP fusion proteins results in the activation of the *tetR* gene and production of TetR protein. TetR protein binds to the *tet* operator of the sequential reporter gene and effectively prevents *HIS3* gene expression. No cell growth was observed on histidine-deficient media. Prevention of the CREB/CBP interaction via mutation of the critical serine, or other disruptive mutations, resulted in histidine prototrophy, since the *tetR* gene product is not produced and hence the *HIS3* gene is not repressed [210]. In addition to mutagenesis studies, the split-hybrid system could allow the screening of compounds that dissociate relevant protein–protein interactions.

***Disruption of Protein–Protein Interactions by Peptide Aptamers*** In a third approach, Geyer et al. [218] used a modified yeast two-hybrid system to identify peptide aptamers that potentially disrupted protein–protein interactions.

Peptide aptamers are proteins that contain a conformationally constrained peptide region of variable sequence displayed from a scaffold, for example, *E. coli* thioredoxin. Geyer et al. expressed aptamers at high levels in cells engineered to express the bait and prey proteins at low concentrations and identified disrupted interactions by decrease of the positive signal. Using this system, they selected peptide aptamers that overcame the cell cycle arrest in *S. cerevisiae* [218].

## Other in Vivo Genetic Selection Systems

Approaches conceptually similar to the above yeast two-hybrid system variants, but based on prokaryotic cells, do exist. One of these systems relies on the properties of the repressor protein cI of bacteriophage λ [219]. The cI repressor protein binds to its operator as a homodimer. Each monomer has two distinct domains, a C-terminal dimerization domain and an N-terminal DNA binding domain. In the system proposed by Park and Raines [219], a protein of interest (X) is fused to the N-terminal DNA binding domain of cI (NcI) to create a hybrid repressor (NcI-X). The dimerization of protein X reconstitutes a functional repressor that is able to bind DNA and to repress the transcription of *tet* and *lacZ* reporter genes (Fig. 7). The reporter genes are under control of the $\lambda P_R$ promoter in a reporter plasmid that also drives the expression of NcI-X. Thus, *E. coli* cells transformed with the reporter plasmid show a $LacZ^{-}Tet^{S}$ phenotype. After cotransformation with a plasmid expressing a peptide library, bacteria expressing dissociative peptides show a $LacZ^{+}Tet^{R}$ phenotype, growing on media containing tetracycline and forming blue colonies. The power of this approach has been recently demonstrated by selecting nine-residue peptides from a combinatorial library that dissociate the HIV-1 protease dimer [219]. Even though Park and Raines [219] used this genetic selection method for targeting a homodimer, they suggest that the

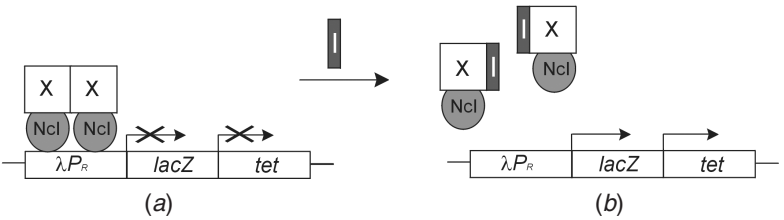


**Figure 7** Genetic selection for inhibitors of a homodimeric protein–protein interaction. (*a*) The dimer of a hybrid λ repressor, NcI-X, composed of the N-terminal DNA binding domain of λ repressor protein (NcI) and a protein (X) able to dimerize, turns off *lacZ* and *tet* reporter genes by binding the $\lambda P_R$ promoter. (*b*) The expression of *lacZ* and *tet* reporter genes is turned on when an inhibitor (I) dissociates the NcI-X homodimer. In these conditions, cells expressing the *tet* and *lacZ* reporter genes survive and form blue colonies in the presence of tetracycline and X-Gal.

same approach could be applicable to identify molecules that disrupt a heterodimeric protein complex, as the assembly of effective heterodimeric hybrid repressor proteins has been reported [220]. Moreover, although Park and Raines panned a peptide library, this method may also be used to select small-molecule inhibitors.

### Inhibitors from Phage Display

Phage display is another technology that can produce specific binding agents for any target molecule and is especially well suited for identifying peptides that bind specific protein domains and dissociate protein complexes. Display of peptides on the surface of bacteriophage virions has become an important means of generating peptide libraries. The very large size of either random libraries or pool of individual variants of a single sequence that can be generated means that complex mixtures can be screened. In fact, display of a repertoire of different peptide sequences on phages can enable the simultaneous analysis of $>10^9$ molecules in a single assay. The problems intrinsic to analysis of small amounts of a selected chemical are thus overcome by the simple and economical process of viral replication. Moreover, high-throughput technology for automated phage display selection has been also developed using picking-spotting robots [221].

Thus, phage display has been applied successfully not only to identify critical amino acids responsible for protein–protein interactions but also to isolate leads for the discovery of new therapeutics [78]. Several antagonist of protein–protein interactions have been discovered using phage libraries. These include peptides blocking IL-1$\alpha$ binding to type I IL-1 receptor [222], insulin-like growth factor 1 binding to its regulatory binding protein IGFBP-1 [223], the angiogenesis factor VEGF binding to its cell-surface receptor KDR [224, 225], IgG Fc binding to staphylococcal protein A [226], and HIV-1 gp120 binding to CD4 cell receptor [227]. All of these peptides are between 10 and 20 residues in length, still a bit large to transform easily into small-molecule drug candidates. They are also quite hydrophobic and tend to cover a large surface on the target protein [225, 226]. In any case, display technology can provide a complementary tool with combinatorial chemistry, by identifying lead peptides that can be the basis for the design of small molecules that disrupt specific protein–protein interactions.

### Microarrays and Small Molecules

Recently, several groups have proposed methods for identifying small-molecule inhibitors using a microarray format. MacBeath and Schreiber [107] immobilized small organic compounds from a combinatorial chemistry library to create a high-density small-molecule microarray. Single resin beads from combinatorial synthesis were placed in 96-well plates, and the organic molecules were chemically released from the beads. The organic molecules were

diluted, spotted, and covalently attached on derivatized glass slides. These microarrays were then incubated with fluorescently labeled target proteins to identify new ligands. Similarly, Winssinger et al. [228] constructed a library of small molecules by tethering them to a peptide nucleic acid (PNA) tag. These PNA tags provide the basis for the structure of the corresponding small molecules and immobilize them at specific sites on the chip. The study also showed that the immobilized small molecules could withstand stringent washing conditions. As a test case, these arrays were used to identify a small-molecule inhibitor of a caspase. As a third example, Kuruvilla et al. [229] constructed a small-molecule microarray containing a collection of 3780 structurally complex 1,3-dioxane compounds to dissect the function of a yeast protein, Ure2, which is a central regulator of the nitrogen metabolic pathway. The library of small molecules was synthesized with a technology platform based on one bead–one stock solution and parsed out to form the small-molecule microarray [230, 231]. The microarray was then probed with fluorescently labeled Ure2. One compound (uretupamine) was identified as specifically inhibiting Ure2 function in a subsequent reporter assay. An analysis of gene expression profiles determined that uretupamine inhibits a particular function of Ure2 without affecting other functions of the protein [232].

The microarray format used in these approaches enables high-throughput screening with minimum amounts of small molecules. This strategy is therefore expected not only to be useful to select ligands of every protein of interest in an entire proteome but also to allow parallel high-throughput screening of small molecules that can disrupt relevant protein–protein interactions, which may greatly facilitate the development of pharmaceutical agents.

## REFERENCES

1. Vidal, M., Endoh, H. (1999). Prospects for drug screening using the reverse two-hybrid system. *Trends Biotechnol., 17*, 374–381.
2. Loregian, A., Marsden, H. S., Palù, G. (2002). Protein–protein interactions as targets for antiviral chemotherapy. *Rev. Med. Virol. 12*, 239–262.
3. Fields, S., Song, O. (1989). A novel genetic system to detect protein–protein interactions. *Nature, 340*, 245–246.
4. Hope, I. A., Struhl, K. (1986). Functional dissection of a eukaryotic transcriptional activator protein, GCN4 of yeast. *Cell, 46*, 885–894.
5. Keegan, L., Gill, G., Ptashne, M. (1986). Separation of DNA binding from the transcription-activating function of a eukaryotic regulatory protein. *Science, 231*, 699–704.
6. Brent, R., Ptashne, M. (1985). A eukaryotic transcriptional activator bearing the DNA specificity of a prokaryotic repressor. *Cell, 43*, 729–736.
7. Estojak, J., Brent, R., Golemis, E. A. (1995). Correlation of two-hybrid affinity data with in vitro measurements. *Mol. Cell. Biol., 15*, 5820–5829.

8. Durfee, T., Becherer, K., Chen, P. L., Yeh, S. H., Yang, Y., Kilburn, A. E., Lee, W. H., Elledge, S. J. (1993). The retinoblastoma protein associates with the protein phosphatase type 1 catalytic subunit. *Genes Devel., 7*, 555–569.
9. Fields, S., Sternglanz, R. (1994). The two-hybrid system: An assay for protein–protein interactions. *Trends Genet., 10*, 286–292.
10. Young, K. H., Ozenberger, B. A. (1995). Investigation of ligand binding to members of the cytokine receptor family within a microbial system. *Ann. N.Y. Acad. Sci., 766*, 279–281.
11. Kajkowski, E. M., Price, L. A., Pausch, M. H., Young, K. H., Ozenberger, B. A. (1997). Investigation of growth hormone releasing hormone receptor structure and activity using yeast expression technologies. *J. Receptor Signal Transduc. Res., 17*, 293–303.
12. Heery, D. M., Kalkhoven, E., Hoare, S., Parker, M. G. (1997). A signature motif in transcriptional co-activators mediates binding to nuclear receptors. *Nature, 387*, 733–736.
13. Loregian, A., Bortolozzo, K., Boso, S., Sapino, B., Betti, M., Biasolo, M. A., Caputo, A., Palù, G. (2003). The Sp1 transcription factor does not directly interact with the HIV-1 Tat protein. *J. Cell. Physiol., 196*, 251–257.
14. Ma, J., Ptashne, M. (1987). A new class of yeast transcriptional activators. *Cell, 51*, 113–119.
15. Yasugi, T., Benson, J. D., Sakai, H., Vidal, M., Howley, P. M. (1997). Mapping and characterization of the interaction domains of human papillomavirus type 16 E1 and E2 proteins. *J. Virol., 71*, 891–899.
16. Du, W., Vidal, M., Xie, J. E., Dyson, N. (1996). RBF, a novel RB-related gene that regulates E2F activity and interacts with cyclin E in Drosophila. *Genes Devel., 10*, 1206–1218.
17. Tirode, F., Malaguti, C., Romero, F., Attar, R., Camonis, J., Egly, J. M. (1997). A conditionally expressed third partner stabilizes or prevents the formation of a transcriptional activator in a three-hybrid system. *J. Biol. Chem., 272*, 22995–22999.
18. Osborne, M. A., Dalton, S., Kochan, J. P. (1995). The yeast tribrid system—genetic detection of trans-phosphorylated ITAM-SH2-interactions. *Bio/Technology, 13*, 1474–1478.
19. Johnsson, N., Varshavsky, A. (1994). Split ubiquitin as a sensor of protein interactions in vivo. *Proc. Natl. Acad. Sci. USA, 91*, 10340–10344.
20. Aronheim, A., Zandi, E., Hennemann, H., Elledge, S. J., Karin, M. (1997). Isolation of an AP-1 repressor by a novel method for detecting protein–protein interactions. *Mol. Cell. Biol., 17*, 3094–3102.
21. Neville, M., Stutz, F., Lee, L., Davis, L. I., Rosbash, M. (1997). The importin-beta family member Crm1p bridges the interaction between Rev and the nuclear pore complex during nuclear export. *Curr. Biol., 7*, 767–775.
22. Bartel, P., Chien, C. T., Sternglanz, R., Fields, S. (1993). Elimination of false positives that arise in using the two-hybrid system. *Biotechniques, 14*, 920–924.
23. James, P., Halladay, J., Craig, E. A. (1996). Genomic libraries and a host strain designed for highly efficient two-hybrid selection in yeast. *Genetics, 144*, 1425–1436.

24. Chevray, P. M., Nathans, D. (1992). Protein interaction cloning in yeast: Identification of mammalian proteins that react with the leucine zipper of Jun. *Proc. Natl. Acad. Sci. USA, 89*, 5789–5793.
25. Bendixen, C., Gangloff, S., Rothstein, R. (1994). A yeast mating-selection scheme for detection of protein–protein interactions. *Nucleic Acids Res., 22*, 1778–1779.
26. Fromont-Racine, M., Rain, J. C., Legrain, P. (1997). Toward a functional analysis of the yeast genome through exhaustive two-hybrid screens. *Nat. Genet., 16*, 277–282.
27. Xenarios, I., Fernandez, E., Salwinski, L., Duan, X. J., Thompson, M. J., Marcotte, E. M., Eisenberg, D. (2001). DIP: The Database of Interacting Proteins: 2001 update. *Nucleic Acids Res., 29*, 239–241.
28. Parent, S. A., Fenimore, C. M., Bostian, K. A. (1985). Vector systems for the expression, analysis and cloning of DNA sequences in *S. cerevisiae*. *Yeast, 1*, 83–138.
29. Chien, C. T., Bartel, P. L., Sternglanz, R., Fields, S. (1991). The two-hybrid system: A method to identify and clone genes for proteins that interact with a protein of interest. *Proc. Natl. Acad. Sci. USA, 88*, 9578–9582.
30. Harper, J. W., Adami, G. R., Wei, N., Keyomarsi, K., Elledge, S. J. (1993). The p21 Cdk-interacting protein Cip1 is a potent inhibitor of G1 cyclin-dependent kinases. *Cell, 75*, 805–816.
31. Gyuris, J., Golemis, E., Chertkov, H., Brent, R. (1993). Cdi1, a human G1 and S phase protein phosphatase that associates with Cdk2. *Cell, 75*, 791–803.
32. Vojtek, A. B., Hollenberg, S. M., Cooper, J. A. (1993). Mammalian Ras interacts directly with the serine/threonine kinase Raf. *Cell, 74*, 205–214.
33. Brent, R. (1982). Regulation and autoregulation by lexA protein. *Biochimie, 64*, 565–569.
34. Ebina, Y., Takahara, Y., Kishi, F., Nakazawa, A., Brent, R. (1983). LexA protein is a repressor of the colicin E1 gene. *J. Biol. Chem., 258*, 13258–13261.
35. Green, N., Alexander, H., Olson, A., Alexander, S., Shinnick, T. M., Sutcliffe, J. G., Lerner, R. A. (1982). Immunogenic structure of the influenza virus hemagglutinin. *Cell, 28*, 477–487.
36. Legrain, P., Dokhelar, M. C., Transy, C. (1994). Detection of protein–protein interactions using different vectors in the two-hybrid system. *Nucleic Acids Res., 22*, 3241–3242.
37. Fradet, Y., Tardif, M., Parent-Vaugeois, C. (1985). [The use of multiparameter flow cytometry in the detection and evaluation of human bladder tumors]. *Union Medicale du Canada, 114*, 778–780.
38. Hadfield, C., Jordan, B. E., Mount, R. C., Pretorius, G. H., Burak, E. (1990). G418-resistance as a dominant marker and reporter for gene expression in *Saccharomyces cerevisiae*. *Curr. Genet., 18*, 303–313.
39. Leanna, C. A., Hannink, M. (1996). The reverse two-hybrid system: A genetic scheme for selection against specific protein/protein interactions. *Nucleic Acids Res., 24*, 3341–3347.
40. Ahmad, M., Bussey, H. (1986). Yeast arginine permease: Nucleotide sequence of the *CAN1* gene. *Curr. Genet., 10*, 587–592.

41. Soares, M. B., Bonaldo, M. F., Jelene, P., Su, L., Lawton, L., Efstratiadis, A. (1994). Construction and characterization of a normalized cDNA library. *Proc. Natl. Acad. Sci. USA, 91*, 9228–9232.
42. Drees, B. L. (1999). Progress and variations in two-hybrid and three-hybrid technologies. *Curr. Opin. Chem. Biol., 3*, 64–70.
43. Roder, K. H., Wolf, S. S., Schweizer, M. (1996). Refinement of vectors for use in the yeast two-hybrid system. *Anal. Biochem., 241*, 260–262.
44. Le Douarin, B., Pierrat, B., vom Baur, E., Chambon, P., Losson, R. (1995). A new version of the two-hybrid assay for detection of protein–protein interactions. *Nucleic Acids Res., 23*, 876–878.
45. Jiang, R., Carlson, M. (1996). Glucose regulates protein interactions within the yeast SNF1 protein kinase complex. *Genes Devel., 10*, 3105–3115.
46. Xu, C. W., Mendelsohn, A. R., Brent, R. (1997). Cells that register logical relationships among proteins. *Proc. Natl. Acad. Sci. USA, 94*, 12473–12478.
47. Inouye, C., Dhillon, N., Durfee, T., Zambryski, P. C., Thorner, J. (1997). Mutational analysis of STE5 in the yeast *Saccharomyces cerevisiae*: Application of a differential interaction trap assay for examining protein–protein interactions. *Genetics, 147*, 479–492.
48. Zhang, J., Lautar, S. (1996). A yeast three-hybrid method to clone ternary protein complex components. *Anal. Biochem., 242*, 68–72.
49. Struhl, K. (1988). The JUN oncoprotein, a vertebrate transcription factor, activates transcription in yeast. *Nature, 332*, 649–650.
50. Yu, X., Wu, L. C., Bowcock, A. M., Aronheim, A., Baer, R. (1998). The C-terminal (BRCT) domains of BRCA1 interact in vivo with CtIP, a protein implicated in the CtBP pathway of transcriptional repression. *J. Biol. Chem., 273*, 25388–25392.
51. Aronheim, A. (1997). Improved efficiency sos recruitment system: Expression of the mammalian GAP reduces isolation of Ras GTPase false positives. *Nucleic Acids Res., 25*, 3373–3374.
52. Goldstein, G., Scheid, M., Hammerling, U., Schlesinger, D. H., Niall, H. D., Boyse, E. A. (1975). Isolation of a polypeptide that has lymphocyte-differentiating properties and is probably represented universally in living cells. *Proc. Natl. Acad. Sci. USA, 72*, 11–15.
53. Hershko, A., Ciechanover, A. (1992). The ubiquitin system for protein degradation. *Annu. Rev. Biochem., 61*, 761–807.
54. Reinders, A., Schulze, W., Kuhn, C., Barker, L., Schulz, A., Ward, J. M., Frommer, W. B. (2002). Protein–protein interactions between sucrose transporters of different affinities colocalized in the same enucleate sieve element. *Plant Cell, 14*, 1567–1577.
55. Stagljar, I., Korostensky, C., Johnsson, N., te Heesen, S. (1998). A genetic system based on split-ubiquitin for the analysis of interactions between membrane proteins *in vivo*. *Proc. Natl. Acad. Sci. USA, 95*, 5187–5192.
56. Wang, B., Nguyen, M., Breckenridge, D. G., Stojanovic, M., Clemons, P. A., Kuppig, S., Shore, G. C. (2003). Uncleaved BAP31 in association with A4 protein at the endoplasmic reticulum is an inhibitor of Fas-initiated release of cytochrome c from mitochondria. *J. Biol. Chem., 278*, 14461–14468.

57. Wittke, S., Dunnwald, M., Albertsen, M., Johnsson, N. (2002). Recognition of a subset of signal sequences by Ssh1p, a Sec61p-related protein in the membrane of endoplasmic reticulum of yeast *Saccharomyces cerevisiae*. *Mol. Biol. Cell, 13*, 2223–2232.
58. Tsujimoto, Y., Numaga, T., Ohshima, K., Yano, M. A., Ohsawa, R., Goto, D. B., Naito, S., Ishikawa, M. (2003). *Arabidopsis* TOBAMOVIRUS MULTIPLICATION (TOM) 2 locus encodes a transmembrane protein that interacts with TOM1. *EMBO J., 22*, 335–343.
59. Lim, M. Y., Dailey, D., Martin, G. S., Thorner, J. (1993). Yeast MCK1 protein kinase autophosphorylates at tyrosine and serine but phosphorylates exogenous substrates at serine and threonine. *J. Biol. Chem., 268*, 21155–21164.
60. Dixon, E. P., Johnstone, E. M., Liu, X., Little, S. P. (1997). An inverse mammalian two-hybrid system for beta secretase and other proteases. *Anal. Biochem., 249*, 239–241.
61. Loregian, A., Bortolozzo, K., Boso, S., Caputo, A., Palù, G. (2003). Interaction of Sp1 transcription factor with HIV-1 Tat protein: Looking for cellular partners. *FEBS Lett., 543*, 61–65.
62. Buchert, M., Schneider, S., Adams, M. T., Hefti, H. P., Moelling, K., Hovens, C. M. (1997). Useful vectors for the two-hybrid system in mammalian cells. *Biotechniques, 23*, 396–398, 400, 402.
63. Luo, Y., Batalao, A., Zhou, H., Zhu, L. (1997). Mammalian two-hybrid system: A complementary approach to the yeast two-hybrid system. *Biotechniques, 22*, 350–352.
64. Van Aelst, L., White, M. A., Wigler, M. H. (1994). Ras partners. *Cold Spring Harbor Symp. Quantitative Biol., 59*, 181–186.
65. White, M. A., Nicolette, C., Minden, A., Polverino, A., Van Aelst, L., Karin, M., Wigler, M. H. (1995). Multiple Ras functions can contribute to mammalian cell transformation. *Cell, 80*, 533–541.
66. Khosravi-Far, R., White, M. A., Westwick, J. K., Solski, P. A., Chrzanowska-Wodnicka, M., Van Aelst, L., Wigler, M. H., Der, C. J. (1996). Oncogenic Ras activation of Raf/mitogen-activated protein kinase-independent pathways is sufficient to cause tumorigenic transformation. *Mol. Cell. Biol., 16*, 3923–3933.
67. van Criekinge, W., Cornelis, S., Van De Craen, M., Vandenabeele, P., Fiers, W., Beyaert, R. (1999). GAL4 is a substrate for caspases: Implications for two-hybrid screening and other GAL4-based assays. *Mol. Cell. Biol. Res. Communs., 1*, 158–161.
68. Walhout, A. J., Sordella, R., Lu, X., Hartley, J. L., Temple, G. F., Brasch, M. A., Thierry-Mieg, N., Vidal, M. (2000). Protein interaction mapping in *C. elegans* using proteins involved in vulval development. *Science, 287*, 116–122.
69. Uetz, P., Giot, L., Cagney, G., Mansfield, T. A., Judson, R. S., Knight, J. R., Lockshon, D., Narayan, V., Srinivasan, M., Pochart, P., Qureshi-Emili, A., Li, Y., Godwin, B., Conover, D., Kalbfleisch, T., Vijayadamodar, G., Yang, M., Johnston, M., Fields, S., Rothberg, J. M. (2000). A comprehensive analysis of protein–protein interactions in *Saccharomyces cerevisiae*. *Nature, 403*, 623–627.
70. Ito, T., Chiba, T., Ozawa, R., Yoshida, M., Hattori, M., Sakaki, Y. (2001). A comprehensive two-hybrid analysis to explore the yeast protein interactome. *Proc. Natl. Acad. Sci. USA, 98*, 4569–4574.

71. Boulton, S. J., Gartner, A., Reboul, J., Vaglio, P., Dyson, N., Hill, D. E., Vidal, M. (2002). Combined functional genomic maps of the *C. elegans* DNA damage response. *Science, 295*, 127–131.
72. Rain, J. C., Selig, L., De Reuse, H., Battaglia, V., Reverdy, C., Simon, S., Lenzen, G., Petel, F., Wojcik, J., Schachter, V., Chemama, Y., Labigne, A., Legrain, P. (2001). The protein–protein interaction map of *Helicobacter pylori*. *Nature, 409*, 211–215.
73. Finley, R. L., Jr., Brent, R. (1994). Interaction mating reveals binary and ternary connections between *Drosophila* cell cycle regulators. *Proc. Natl. Acad. Sci. USA, 91*, 12980–12984.
74. Bartel, P. L., Roecklein, J. A., SenGupta, D., Fields, S. (1996). A protein linkage map of *Escherichia coli* bacteriophage T7. *Nat. Genet., 12*, 72–77.
75. Tomb, J. F., White, O., Kerlavage, A. R., Clayton, R. A., Sutton, G. G., Fleischmann, R. D., Ketchum, K. A., Klenk, H. P., Gill, S., Dougherty, B. A., Nelson, K., Quackenbush, J., Zhou, L., Kirkness, E. F., Peterson, S., Loftus, B., Richardson, D., Dodson, R., Khalak, H. G., Glodek, A., McKenney, K., Fitzegerald, L. M., Lee, N., Adams, M. D., Venter, J. C., et al. (1997). The complete genome sequence of the gastric pathogen *Helicobacter pylori*. *Nature, 388*, 539–547.
76. Ito, T., Tashiro, K., Muta, S., Ozawa, R., Chiba, T., Nishizawa, M., Yamamoto, K., Kuhara, S., Sakaki, Y. (2000). Toward a protein–protein interaction map of the budding yeast: A comprehensive system to examine two-hybrid interactions in all possible combinations between the yeast proteins. *Proc. Natl. Acad. Sci. USA, 97*, 1143–1147.
77. Johnsson, K., Ge, L. (1999). Phage display of combinatorial peptide and protein libraries and their applications in biology and chemistry. *Curr. Top. Microbiol. Immunol., 243*, 87–105.
78. Lowman, H. B. (1997). Bacteriophage display and discovery of peptide leads for drug development. *Annu. Rev. Biophys. Biomol. Struct., 26*, 401–424.
79. FitzGerald, K. (2000). *In vitro* display technologies—new tools for drug discovery. *Drug Discov. Today, 5*, 253–258.
80. Smith, G. P. (1985). Filamentous fusion phage: Novel expression vectors that display cloned antigens on the virion surface. *Science, 228*, 1315–1317.
81. Jespers, L. S., Messens, J. H., De Keyser, A., Eeckhout, D., Van den Brande, I., Gansemans, Y. G., Lauwereys, M. J., Vlasuk, G. P., Stanssens, P. E. (1995). Surface expression and ligand-based selection of cDNAs fused to filamentous phage gene VI. *Bio/Technology, 13*, 378–382.
82. Bass, S., Greene, R., Wells, J. A. (1990). Hormone phage: An enrichment method for variant proteins with altered binding properties. *Proteins, 8*, 309–314.
83. Lowman, H. B., Bass, S. H., Simpson, N., Wells, J. A. (1991). Selecting high-affinity binding proteins by monovalent phage display. *Biochemistry, 30*, 10832–10838.
84. Sternberg, N., Hoess, R. H. (1995). Display of peptides and proteins on the surface of bacteriophage lambda. *Proc. Natl. Acad. Sci. USA, 92*, 1609–1613.
85. Maruyama, I. N., Maruyama, H. I., Brenner, S. (1994). Lambda foo: A lambda phage vector for the expression of foreign proteins. *Proc. Natl. Acad. Sci. USA, 91*, 8273–8277.
86. Efimov, V. P., Nepluev, I. V., Mesyanzhinov, V. V. (1995). Bacteriophage T4 as a surface display vector. *Virus Genes, 10*, 173–177.

87. Cull, M. G., Miller, J. F., Schatz, P. J. (1992). Screening for receptor ligands using large libraries of peptides linked to the C terminus of the lac repressor. *Proc. Natl. Acad. Sci. USA, 89*, 1865–1869.
88. Mattheakis, L. C., Bhatt, R. R., Dower, W. J. (1994). An *in vitro* polysome display system for identifying ligands from very large peptide libraries. *Proc. Natl. Acad. Sci. USA, 91*, 9022–9026.
89. Francisco, J. A., Campbell, R., Iverson, B. L., Georgiou, G. (1993). Production and fluorescence-activated cell sorting of *Escherichia coli* expressing a functional antibody fragment on the external surface. *Proc. Natl. Acad. Sci. USA, 90*, 10444–10448.
90. Lang, H. (2000). Outer membrane proteins as surface display systems. *Int. J. Med. Microbiol., 290*, 579–585.
91. Westerlund-Wikstrom, B. (2000). Peptide display on bacterial flagella: Principles and applications. *Int. J. Med. Microbiol., 290*, 223–230.
92. Boder, E. T., Wittrup, K. D. (1997). Yeast surface display for screening combinatorial polypeptide libraries. *Nat. Biotechnol., 15*, 553–557.
93. Turpen, T. H., Reinl, S. J., Charoenvit, Y., Hoffman, S. L., Fallarme, V., Grill, L. K. (1995). Malarial epitopes expressed on the surface of recombinant tobacco mosaic virus. *Bio/Technology, 13*, 53–57.
94. Resnick, D. A., Smith, A. D., Gesiler, S. C., Zhang, A., Arnold, E., Arnold, G. F. (1995). Chimeras from a human rhinovirus 14-human immunodeficiency virus type 1 (HIV-1) V3 loop seroprevalence library induce neutralizing responses against HIV-1. *J. Virol., 69*, 2406–2411.
95. Grabherr, R., Ernst, W. (2001). The baculovirus expression system as a tool for generating diversity by viral surface display. *Combinatorial Chem. High Throughput Screening, 4*, 185–192.
96. Crameri, R., Suter, M. (1993). Display of biologically active proteins on the surface of filamentous phages: A cDNA cloning system for selection of functional gene products linked to the genetic information responsible for their production. *Gene, 137*, 69–75.
97. Li, K. W., Geraerts, W. P., van Elk, R., Joosse, J. (1989). Quantification of proteins in the subnanogram and nanogram range: Comparison of the AuroDye, FerriDye, and India ink staining methods. *Anal. Biochem., 182*, 44–47.
98. Zhu, H., Klemic, J. F., Chang, S., Bertone, P., Casamayor, A., Klemic, K. G., Smith, D., Gerstein, M., Reed, M. A., Snyder, M. (2000). Analysis of yeast protein kinases using protein chips. *Nat. Genet., 26*, 283–289.
99. Zhu, H., Bilgin, M., Bangham, R., Hall, D., Casamayor, A., Bertone, P., Lan, N., Jansen, R., Bidlingmaier, S., Houfek, T., Mitchell, T., Miller, P., Dean, R. A., Gerstein, M., Snyder, M. (2001). Global analysis of protein activities using proteome chips. *Science, 293*, 2101–2105.
100. Folch, A., Toner, M. (2000). Microengineering of cellular interactions. *Annu. Rev. Biomed. Eng., 2*, 227–256.
101. Nelson, R. W., Nedelkov, D., Tubbs, K. A. (2000). Biosensor chip mass spectrometry: A chip-based proteomics approach. *Electrophoresis, 21*, 1155–1163.
102. Zhu, H., Snyder, M. (2001). Protein arrays and microarrays. *Curr. Opin. Chem. Biol., 5*, 40–45.

103. Fung, E. T., Thulasiraman, V., Weinberger, S. R., Dalmasso, E. A. (2001). Protein biochips for differential profiling. *Curr. Opin. Biotechnol., 12*, 65–69.

104. Merchant, M., Weinberger, S. R. (2000). Recent advancements in surface-enhanced laser desorption/ionization-time of flight-mass spectrometry. *Electrophoresis, 21*, 1164–1177.

105. Howard, J. C., Heinemann, C., Thatcher, B. J., Martin, B., Gan, B. S., Reid, G. (2000). Identification of collagen-binding proteins in *Lactobacillus* spp. with surface-enhanced laser desorption/ionization-time of flight ProteinChip technology. *Appl. Environ. Microbiol., 66*, 4396–4400.

106. Xiao, Z., Jiang, X., Beckett, M. L., Wright, G. L., Jr. (2000). Generation of a baculovirus recombinant prostate-specific membrane antigen and its use in the development of a novel protein biochip quantitative immunoassay. *Protein Expression and Purification, 19*, 12–21.

107. MacBeath, G., Schreiber, S. L. (2000). Printing proteins as microarrays for high-throughput function determination. *Science, 289*, 1760–1763.

108. Moody, M. D., Van Arsdell, S. W., Murphy, K. P., Orencole, S. F., Burns, C. (2001). Array-based ELISAs for high-throughput analysis of human cytokines. *Biotechniques, 31*, 186–190, 192–184.

109. Huang, R. P., Huang, R., Fan, Y., Lin, Y. (2001). Simultaneous detection of multiple cytokines from conditioned media and patient's sera by an antibody-based protein array system. *Anal. Biochem., 294*, 55–62.

110. Wiese, R., Belosludtsev, Y., Powdrill, T., Thompson, P., Hogan, M. (2001). Simultaneous multianalyte ELISA performed on a microarray platform. *Clin. Chem., 47*, 1451–1457.

111. Sreekumar, A., Nyati, M. K., Varambally, S., Barrette, T. R., Ghosh, D., Lawrence, T. S., Chinnaiyan, A. M. (2001). Profiling of cancer cells using protein microarrays: Discovery of novel radiation-regulated proteins. *Cancer Res., 61*, 7585–7593.

112. Haab, B. B., Dunham, M. J., Brown, P. O. (2001). Protein microarrays for highly parallel detection and quantitation of specific proteins and antibodies in complex solutions. *Genome Biol., 2*, RESEARCH0004.

113. Miller, J. C., Butler, E. B., Teh, B. S., Haab, B. B. (2001). The application of protein microarrays to serum diagnostics: Prostate cancer as a test case. *Disease Markers, 17*, 225–234.

114. McDonnell, J. M. (2001). Surface plasmon resonance: Towards an understanding of the mechanisms of biological molecular recognition. *Curr. Opin. Chem. Biol., 5*, 572–577.

115. Nieba, L., Nieba-Axmann, S. E., Persson, A., Hamalainen, M., Edebratt, F., Hansson, A., Lidholm, J., Magnusson, K., Karlsson, A. F., Pluckthun, A. (1997). BIACORE analysis of histidine-tagged proteins using a chelating NTA sensor chip. *Anal. Biochem., 252*, 217–228.

116. Salamon, Z., Brown, M. F., Tollin, G. (1999). Plasmon resonance spectroscopy: Probing molecular interactions within membranes. *Trends Biochem. Sci., 24*, 213–219.

117. Myszka, D. G., Rich, R. L. (2000). Implementing surface plasmon resonance biosensors in drug discovery. *Pharm. Sci. Technol. Today, 3*, 310–317.

118. Sapsford, K. E., Liron, Z., Shubin, Y. S., Ligler, F. S. (2001). Kinetics of antigen binding to arrays of antibodies in different sized spots. *Anal. Chem., 73*, 5518–5524.
119. Linnell, J., Groeger, G., Hassan, A. B. (2001). Real time kinetics of insulin-like growth factor II (IGF-II) interaction with the IGF-II/mannose 6-phosphate receptor: The effects of domain 13 and pH. *J. Biol. Chem., 276*, 23986–23991.
120. Houseman, B. T., Huh, J. H., Kron, S. J., Mrksich, M. (2002). Peptide chips for the quantitative evaluation of protein kinase activity. *Nat. Biotechnol., 20*, 270–274.
121. Ge, H. (2000). UPA, a universal protein array system for quantitative detection of protein–protein, protein-DNA, protein-RNA and protein-ligand interactions. *Nucleic Acids Res., 28*, e3.
122. Cahill, D. J. (2001). Protein and antibody arrays and their medical applications. *J. Immunol. Methods, 250*, 81–91.
123. Stoll, D., Templin, M. F., Schrenk, M., Traub, P. C., Vohringer, C. F., Joos, T. O. (2002). Protein microarray technology. *Frontiers Biosci., 7*, c13–32.
124. Bussow, K., Cahill, D., Nietfeld, W., Bancroft, D., Scherzinger, E., Lehrach, H., Walter, G. (1998). A method for global protein expression and antibody screening on high-density filters of an arrayed cDNA library. *Nucleic Acids Res., 26*, 5007–5008.
125. Braun, P., Hu, Y., Shen, B., Halleck, A., Koundinya, M., Harlow, E., LaBaer, J. (2002). Proteome-scale purification of human proteins from bacteria. *Proc. Natl. Acad. Sci. USA, 99*, 2654–2659.
126. Albala, J. S., Franke, K., McConnell, I. R., Pak, K. L., Folta, P. A., Rubinfeld, B., Davies, A. H., Lennon, G. G., Clark, R. (2000). From genes to proteins: High-throughput expression and purification of the human proteome. *J. Cell. Biochem., 80*, 187–191.
127. Lueking, A., Horn, M., Eickhoff, H., Bussow, K., Lehrach, H., Walter, G. (1999). Protein microarrays for gene expression and antibody screening. *Anal. Biochem., 270*, 103–111.
128. Haab, B. B. (2001). Advances in protein microarray technology for protein expression and interaction profiling. *Curr. Opin. Drug Discov. Devel., 4*, 116–123.
129. Keefe, A. D., Szostak, J. W. (2001). Functional proteins from a random-sequence library. *Nature, 410*, 715–718.
130. He, M., Taussig, M. J. (2001). Single step generation of protein arrays from DNA by cell-free expression and *in situ* immobilisation (PISA method). *Nucleic Acids Res., 29*, E73–73.
131. Reineke, U., Volkmer-Engert, R., Schneider-Mergener, J. (2001). Applications of peptide arrays prepared by the SPOT-technology. *Curr. Opin. Biotechnol., 12*, 59–64.
132. Walter, G., Bussow, K., Cahill, D., Lueking, A., Lehrach, H. (2000). Protein arrays for gene expression and molecular interaction screening. *Curr. Opin. Microbiol., 3*, 298–302.
133. Templin, M. F., Stoll, D., Schrenk, M., Traub, P. C., Vohringer, C. F., Joos, T. O. (2002). Protein microarray technology. *Trends Biotechnol., 20*, 160–166.
134. Guschin, D., Yershov, G., Zaslavsky, A., Gemmell, A., Shick, V., Proudnikov, D., Arenkov, P., Mirzabekov, A. (1997). Manual manufacturing of oligonucleotide, DNA, and protein microchips. *Anal. Biochem., 250*, 203–211.

135. Afanassiev, V., Hanemann, V., Wolfl, S. (2000). Preparation of DNA and protein micro arrays on glass slides coated with an agarose film. *Nucleic Acids Res., 28*, E66.

136. Joos, T. O., Schrenk, M., Hopfl, P., Kroger, K., Chowdhury, U., Stoll, D., Schorner, D., Durr, M., Herick, K., Rupp, S., Sohn, K., Hammerle, H. (2000). A microarray enzyme-linked immunosorbent assay for autoimmune diagnostics. *Electrophoresis, 21*, 2641–2650.

137. Schweitzer, B., Kingsmore, S. F. (2002). Measuring proteins on microarrays. *Curr. Opin. Biotechnol., 13*, 14–19.

138. de Wildt, R. M., Mundy, C. R., Gorick, B. D., Tomlinson, I. M. (2000). Antibody arrays for high-throughput screening of antibody-antigen interactions. *Nat. Biotechnol., 18*, 989–994.

139. Heyse, S., Vogel, H., Sanger, M., Sigrist, H. (1995). Covalent attachment of functionalized lipid bilayers to planar waveguides for measuring protein binding to biomimetic membranes. *Protein Sci., 4*, 2532–2544.

140. Bieri, C., Ernst, O. P., Heyse, S., Hofmann, K. P., Vogel, H. (1999). Micropatterned immobilization of a G protein-coupled receptor and direct detection of G protein activation. *Nat. Biotechnol., 17*, 1105–1108.

141. Rich, R. L., Day, Y. S., Morton, T. A., Myszka, D. G. (2001). High-resolution and high-throughput protocols for measuring drug/human serum albumin interactions using BIACORE. *Anal. Biochem., 296*, 197–207.

142. Hodneland, C. D., Lee, Y. S., Min, D. H., Mrksich, M. (2002). Selective immobilization of proteins to self-assembled monolayers presenting active site-directed capture ligands. *Proc. Natl. Acad. Sci. USA, 99*, 5048–5052.

143. Lin, S. C., Tseng, F. G., Huang, H. M., Huang, C. Y., Chieng, C. C. (2001). Micro-sized 2D protein arrays immobilized by micro-stamps and micro-wells for disease diagnosis and drug screening. *Fres. J. Anal. Chem., 371*, 202–208.

144. Arenkov, P., Kukhtin, A., Gemmell, A., Voloshchuk, S., Chupeeva, V., Mirzabekov, A. (2000). Protein microchips: Use for immunoassay and enzymatic reactions. *Anal. Biochem., 278*, 123–131.

145. Avseenko, N. V., Morozova, T. Y., Ataullakhanov, F. I., Morozov, V. N. (2002). Immunoassay with multicomponent protein microarrays fabricated by electrospray deposition. *Anal. Chem., 74*, 927–933.

146. Silzel, J. W., Cercek, B., Dodson, C., Tsay, T., Obremski, R. J. (1998). Mass-sensing, multianalyte microarray immunoassay with imaging detection. *Clin. Chem., 44*, 2036–2043.

147. Marx, J. (2000). Medicine. DNA arrays reveal cancer in its many forms. *Science, 289*, 1670–1672.

148. Lizardi, P. M., Huang, X., Zhu, Z., Bray-Ward, P., Thomas, D. C., Ward, D. C. (1998). Mutation detection and single-molecule counting using isothermal rolling-circle amplification. *Nat. Genet., 19*, 225–232.

149. Weinberger, S. R., Morris, T. S., Pawlak, M. (2000). Recent trends in protein biochip technology. *Pharmacogenomics, 1*, 395–416.

150. Davies, H., Lomas, L., Austen, B. (1999). Profiling of amyloid beta peptide variants using SELDI Protein Chip arrays. *Biotechniques, 27*, 1258–1261.

151. Mullett, W. M., Lai, E. P., Yeung, J. M. (2000). Surface plasmon resonance-based immunoassays. *Methods, 22*, 77–91.

152. Luo, L. Y., Diamandis, E. P. (2000). Preliminary examination of time-resolved fluorometry for protein array applications. *Luminescence, 15*, 409–413.

153. Schweitzer, B., Roberts, S., Grimwade, B., Shao, W., Wang, M., Fu, Q., Shu, Q., Laroche, I., Zhou, Z., Tchernev, V. T., Christiansen, J., Velleca, M., Kingsmore, S. F. (2002). Multiplexed protein profiling on microarrays by rolling-circle amplification. *Nat. Biotechnol., 20*, 359–365.

154. Schweitzer, B., Wiltshire, S., Lambert, J., O'Malley, S., Kukanskis, K., Zhu, Z., Kingsmore, S. F., Lizardi, P. M., Ward, D. C. (2000). Inaugural article: Immunoassays with rolling circle DNA amplification: A versatile platform for ultrasensitive antigen detection. *Proc. Natl. Acad. Sci. USA, 97*, 10113–10119.

155. Stolz, M., Stoffler, D., Aebi, U., Goldsbury, C. (2000). Monitoring biomolecular interactions by time-lapse atomic force microscopy. *J. Struct. Biol., 131*, 171–180.

156. Wu, G., Datar, R. H., Hansen, K. M., Thundat, T., Cote, R. J., Majumdar, A. (2001). Bioassay of prostate-specific antigen (PSA) using microcantilevers. *Nat. Biotechnol., 19*, 856–860.

157. Andersen, J. S., Svensson, B., Roepstorff, P. (1996). Electrospray ionization and matrix assisted laser desorption/ionization mass spectrometry: Powerful analytical tools in recombinant protein chemistry. *Nat. Biotechnol., 14*, 449–457.

158. Jensen, O. N., Wilm, M., Shevchenko, A., Mann, M. (1999). Sample preparation methods for mass spectrometric peptide mapping directly from 2-DE gels. *Methods Mol. Biol., 112*, 513–530.

159. Muzio, M., Chinnaiyan, A. M., Kischkel, F. C., O'Rourke, K., Shevchenko, A., Ni, J., Scaffidi, C., Bretz, J. D., Zhang, M., Gentz, R., Mann, M., Krammer, P. H., Peter, M. E., Dixit, V. M. (1996). FLICE, a novel FADD-homologous ICE/CED-3-like protease, is recruited to the CD95 (Fas/APO-1) death-inducing signaling complex. *Cell, 85*, 817–827.

160. Wigge, P. A., Jensen, O. N., Holmes, S., Soues, S., Mann, M., Kilmartin, J. V. (1998). Analysis of the *Saccharomyces* spindle pole by matrix-assisted laser desorption/ionization (MALDI) mass spectrometry. *J. Cell Biol., 141*, 967–977.

161. Neubauer, G., Gottschalk, A., Fabrizio, P., Seraphin, B., Luhrmann, R., Mann, M. (1997). Identification of the proteins of the yeast U1 small nuclear ribonucleoprotein complex by mass spectrometry. *Proc. Natl. Acad. Sci. USA, 94*, 385–390.

162. Zachariae, W., Shevchenko, A., Andrews, P. D., Ciosk, R., Galova, M., Stark, M. J., Mann, M., Nasmyth, K. (1998). Mass spectrometric analysis of the anaphase-promoting complex from yeast: Identification of a subunit related to cullins. *Science, 279*, 1216–1219.

163. Neubauer, G., King, A., Rappsilber, J., Calvio, C., Watson, M., Ajuh, P., Sleeman, J., Lamond, A., Mann, M. (1998). Mass spectrometry and EST-database searching allows characterization of the multi-protein spliceosome complex. *Nat. Genet., 20*, 46–50.

164. Shevchenko, A., Keller, P., Scheiffele, P., Mann, M., Simons, K. (1997). Identification of components of trans-Golgi network-derived transport vesicles and detergent-insoluble complexes by nanoelectrospray tandem mass spectrometry. *Electrophoresis, 18*, 2591–2600.

165. Sarioglu, H., Lottspeich, F., Walk, T., Jung, G., Eckerskorn, C. (2000). Deamidation as a widespread phenomenon in two-dimensional polyacrylamide gel electrophoresis of human blood plasma proteins. *Electrophoresis, 21*, 2209–2218.
166. Kaufmann, H., Bailey, J. E., Fussenegger, M. (2001). Use of antibodies for detection of phosphorylated proteins separated by two-dimensional gel electrophoresis. *Proteomics, 1*, 194–199.
167. Taniguchi, N., Ekuni, A., Ko, J. H., Miyoshi, E., Ikeda, Y., Ihara, Y., Nishikawa, A., Honke, K., Takahashi, M. (2001). A glycomic approach to the identification and characterization of glycoprotein function in cells transfected with glycosyltransferase genes. *Proteomics, 1*, 239–247.
168. Fivaz, M., Vilbois, F., Pasquali, C., van der Goot, F. G. (2000). Analysis of glycosyl phosphatidylinositol-anchored proteins by two-dimensional gel electrophoresis. *Electrophoresis, 21*, 3351–3356.
169. Gorg, A., Obermaier, C., Boguth, G., Harder, A., Scheibe, B., Wildgruber, R., Weiss, W. (2000). The current state of two-dimensional electrophoresis with immobilized pH gradients. *Electrophoresis, 21*, 1037–1053.
170. Gorg, A., Obermaier, C., Boguth, G., Weiss, W. (1999). Recent developments in two-dimensional gel electrophoresis with immobilized pH gradients: Wide pH gradients up to pH 12, longer separation distances and simplified procedures. *Electrophoresis, 20*, 712–717.
171. Hoving, S., Voshol, H., van Oostrum, J. (2000). Towards high performance two-dimensional gel electrophoresis using ultrazoom gels. *Electrophoresis, 21*, 2617–2621.
172. Wildgruber, R., Harder, A., Obermaier, C., Boguth, G., Weiss, W., Fey, S. J., Larsen, P. M., Gorg, A. (2000). Towards higher resolution: Two-dimensional electrophoresis of *Saccharomyces cerevisiae* proteins using overlapping narrow immobilized pH gradients. *Electrophoresis, 21*, 2610–2616.
173. Bjellqvist, B., Ek, K., Righetti, P. G., Gianazza, E., Gorg, A., Westermeier, R., Postel, W. (1982). Isoelectric focusing in immobilized pH gradients: Principle, methodology and some applications. *J. Biochem. Biophys. Methods, 6*, 317–339.
174. Lopez, M. F. (2000). Better approaches to finding the needle in a haystack: Optimizing proteome analysis through automation. *Electrophoresis, 21*, 1082–1093.
175. Quadroni, M., James, P. (1999). Proteomics and automation. *Electrophoresis, 20*, 664–677.
176. Gygi, S. P., Corthals, G. L., Zhang, Y., Rochon, Y., Aebersold, R. (2000). Evaluation of two-dimensional gel electrophoresis-based proteome analysis technology. *Proc. Natl. Acad. Sci. USA, 97*, 9390–9395.
177. Corthals, G. L., Wasinger, V. C., Hochstrasser, D. F., Sanchez, J. C. (2000). The dynamic range of protein expression: A challenge for proteomic research. *Electrophoresis, 21*, 1104–1115.
178. Santoni, V., Kieffer, S., Desclaux, D., Masson, F., Rabilloud, T. (2000). Membrane proteomics: Use of additive main effects with multiplicative interaction model to classify plasma membrane proteins according to their solubility and electrophoretic properties. *Electrophoresis, 21*, 3329–3344.

179. Molloy, M. P., Phadke, N. D., Maddock, J. R., Andrews, P. C. (2001). Two-dimensional electrophoresis and peptide mass fingerprinting of bacterial outer membrane proteins. *Electrophoresis, 22*, 1686–1696.

180. Goldberg, H. A., Domenicucci, C., Pringle, G. A., Sodek, J. (1988). Mineral-binding proteoglycans of fetal porcine calvarial bone. *J. Biol. Chem., 263*, 12092–12101.

181. Sanchez, J. C., Ravier, F., Pasquali, C., Frutiger, S., Paquet, N., Bjellqvist, B., Hochstrasser, D. F., Hughes, G. J. (1992). Improving the detection of proteins after transfer to polyvinylidene difluoride membranes. *Electrophoresis, 13*, 715–717.

182. Rabilloud, T. (1992). A comparison between low background silver diammine and silver nitrate protein stains. *Electrophoresis, 13*, 429–439.

183. Hochstrasser, D. F., Merril, C. R. (1988). "Catalysts" for polyacrylamide gel polymerization and detection of proteins by silver staining. *Appl. Theor. Electrophor., 1*, 35–40.

184. Garrels, J. I., Franza, B. R., Jr. (1989). The REF52 protein database. Methods of database construction and analysis using the QUEST system and characterizations of protein patterns from proliferating and quiescent REF52 cells. *J. Biol. Chem., 264*, 5283–5298.

185. Latham, K. E., Garrels, J. I., Solter, D. (1993). Two-dimensional gel analysis of protein synthesis. *Methods Enzymol., 225*, 473–489.

186. Ortiz, M. L., Calero, M., Fernandez Patron, C., Patron, C. F., Castellanos, L., Mendez, E. (1992). Imidazole-SDS-Zn reverse staining of proteins in gels containing or not SDS and microsequence of individual unmodified electroblotted proteins. *FEBS Lett., 296*, 300–304.

187. James, P., Quadroni, M., Carafoli, E., Gonnet, G. (1993). Protein identification by mass profile fingerprinting. *Biochem. Biophys. Res. Commun., 195*, 58–64.

188. Yamaguchi, K., Asakawa, H. (1988). Preparation of colloidal gold for staining proteins electrotransferred onto nitrocellulose membranes. *Anal. Biochem., 172*, 104–107.

189. Eckerskorn, C., Strupat, K., Karas, M., Hillenkamp, F., Lottspeich, F. (1992). Mass spectrometric analysis of blotted proteins after gel electrophoretic separation by matrix-assisted laser desorption/ionization. *Electrophoresis, 13*, 664–665.

190. Hughes, J. H., Mack, K., Hamparian, V. V. (1988). India ink staining of proteins on nylon and hydrophobic membranes. *Anal. Biochem., 173*, 18–25.

191. Link, A. J., Eng, J., Schieltz, D. M., Carmack, E., Mize, G. J., Morris, D. R., Garvik, B. M., Yates, J. R., 3rd. (1999). Direct analysis of protein complexes using mass spectrometry. *Nat. Biotechnol., 17*, 676–682.

192. Washburn, M. P., Wolters, D., Yates, J. R., 3rd. (2001). Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat. Biotechnol., 19*, 242–247.

193. Muller, D. R., Schindler, P., Towbin, H., Wirth, U., Voshol, H., Hoving, S., Steinmetz, M. O. (2001). Isotope-tagged cross-linking reagents. A new tool in mass spectrometric protein interaction analysis. *Anal. Chem., 73*, 1927–1934.

194. Jensen, O. N., Podtelejnikov, A. V., Mann, M. (1997). Identification of the components of simple protein mixtures by high-accuracy peptide mass mapping and database searching. *Anal. Chem., 69*, 4741–4750.

195. Mann, M., Hojrup, P., Roepstorff, P. (1993). Use of mass spectrometric molecular weight information to identify proteins in sequence databases. *Biol. Mass Spectrom., 22*, 338–345.

196. Yates, J. R., 3rd, Speicher, S., Griffin, P. R., Hunkapiller, T. (1993). Peptide mass maps: A highly informative approach to protein identification. *Anal. Biochem., 214*, 397–408.

197. Clauser, K. R., Baker, P., Burlingame, A. L. (1999). Role of accurate mass measurement (+/– 10 ppm) in protein identification strategies employing MS or MS/MS and database searching. *Anal. Chem., 71*, 2871–2882.

198. Zhang, W., Chait, B. T. (2000). ProFound: An expert system for protein identification using mass spectrometric peptide mapping information. *Anal. Chem., 72*, 2482–2489.

199. Perkins, D. N., Pappin, D. J., Creasy, D. M., Cottrell, J. S. (1999). Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis, 20*, 3551–3567.

200. Pandey, A., Mann, M. (2000). Proteomics to study genes and genomes. *Nature, 405*, 837–846.

201. Naaby-Hansen, S., Waterfield, M. D., Cramer, R. (2001). Proteomics-post-genomic cartography to understand gene function. *Trends Pharmacol. Sci., 22*, 376–384.

202. Blobel, G., Wozniak, R. W. (2000). Proteomics for the pore. *Nature, 403*, 835–836.

203. Gavin, A. C., Bosche, M., Krause, R., Grandi, P., Marzioch, M., Bauer, A., Schultz, J., Rick, J. M., Michon, A. M., Cruciat, C. M., Remor, M., Hofert, C., Schelder, M., Brajenovic, M., Ruffner, H., Merino, A., Klein, K., Hudak, M., Dickson, D., Rudi, T., Gnau, V., Bauch, A., Bastuck, S., Huhse, B., Leutwein, C., Heurtier, M. A., Copley, R. R., Edelmann, A., Querfurth, E., Rybin, V., Drewes, G., Raida, M., Bouwmeester, T., Bork, P., Seraphin, B., Kuster, B., Neubauer, G., Superti-Furga, G. (2002). Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature, 415*, 141–147.

204. Ho, Y., Gruhler, A., Heilbut, A., Bader, G. D., Moore, L., Adams, S. L., Millar, A., Taylor, P., Bennett, K., Boutilier, K., Yang, L., Wolting, C., Donaldson, I., Schandorff, S., Shewnarane, J., Vo, M., Taggart, J., Goudreault, M., Muskat, B., Alfarano, C., Dewar, D., Lin, Z., Michalickova, K., Willems, A. R., Sassi, H., Nielsen, P. A., Rasmussen, K. J., Andersen, J. R., Johansen, L. E., Hansen, L. H., Jespersen, H., Podtelejnikov, A., Nielsen, E., Crawford, J., Poulsen, V., Sorensen, B. D., Matthiesen, J., Hendrickson, R. C., Gleeson, F., Pawson, T., Moran, M. F., Durocher, D., Mann, M., Hogue, C. W., Figeys, D., Tyers, M. (2002). Systematic identification of protein complexes in *Saccharomyces cerevisiae* by mass spectrometry. *Nature, 415*, 180–183.

205. Zanzoni, A., Montecchi-Palazzi, L., Quondam, M., Ausiello, G., Helmer-Citterich, M., Cesareni, G. (2002). MINT: A Molecular INTeraction database. *FEBS Lett., 513*, 135–140.

206. Bader, G. D., Donaldson, I., Wolting, C., Ouellette, B. F., Pawson, T., Hogue, C. W. (2001). BIND—The Biomolecular Interaction Network Database. *Nucleic Acids Res., 29*, 242–245.

207. Cochran, A. G. (2000). Antagonists of protein–protein interactions. *Chem. Biol., 7*, R85–94.

208. Toogood, P. L. (2002). Inhibition of protein–protein association by small molecules: Approaches and progress. *J. Med. Chem., 45*, 1543–1558.

209. Vidal, M., Brachmann, R. K., Fattaey, A., Harlow, E., Boeke, J. D. (1996). Reverse two-hybrid and one-hybrid systems to detect dissociation of protein–protein and DNA-protein interactions. *Proc. Natl. Acad. Sci. USA, 93*, 10315–10320.

210. Shih, H. M., Goldman, P. S., DeMaggio, A. J., Hollenberg, S. M., Goodman, R. H., Hoekstra, M. F. (1996). A positive genetic selection for disrupting protein-protein interactions: Identification of CREB mutations that prevent association with the coactivator CBP. *Proc. Natl. Acad. Sci. USA, 93*, 13896–13901.

211. Shah, N., Klausner, R. D. (1993). Brefeldin A reversibly inhibits secretion in *Saccharomyces cerevisiae*. *J. Biol. Chem., 268*, 5345–5348.

212. Kaufer, N. F., Fried, H. M., Schwindinger, W. F., Jasin, M., Warner, J. R. (1983). Cycloheximide resistance in yeast: The gene and its protein. *Nucleic Acids Res., 11*, 3123–3135.

213. Rose, M., Botstein, D. (1983). Structure and function of the yeast URA3 gene. Differentially regulated expression of hybrid beta-galactosidase from overlapping coding sequences in yeast. *J. Mol. Biol., 170*, 883–904.

214. Boeke, J. D., LaCroute, F., Fink, G. R. (1984). A positive selection for mutants lacking orotidine-5′-phosphate decarboxylase activity in yeast: 5-Fluoro-orotic acid resistance. *Mol. Gen. Genet., 197*, 345–346.

215. Vidal, M., Braun, P., Chen, E., Boeke, J. D., Harlow, E. (1996). Genetic characterization of a mammalian protein–protein interaction domain by using a yeast reverse two-hybrid system. *Proc. Natl. Acad. Sci. USA, 93*, 10321–10326.

216. Huang, J., Schreiber, S. L. (1997). A yeast genetic system for selecting small molecule inhibitors of protein–protein interactions in nanodroplets. *Proc. Natl. Acad. Sci. USA, 94*, 13396–13401.

217. Young, K., Lin, S., Sun, L., Lee, E., Modi, M., Hellings, S., Husbands, M., Ozenberger, B., Franco, R. (1998). Identification of a calcium channel modulator using a high throughput yeast two-hybrid screen. *Nat. Biotechnol., 16*, 946–950.

218. Geyer, C. R., Colman-Lerner, A., Brent, R. (1999). "Mutagenesis" by peptide aptamers identifies genetic network members and pathway connections. *Proc. Natl. Acad. Sci. USA, 96*, 8567–8572.

219. Park, S. H., Raines, R. T. (2000). Genetic selection for dissociative inhibitors of designated protein–protein interactions. *Nat. Biotechnol., 18*, 847–851.

220. Hu, J. C. (1995). Repressor fusions as a tool to study protein–protein interactions. *Structure, 3*, 431–433.

221. Walter, G., Konthur, Z., Lehrach, H. (2001). High-throughput screening of surface displayed gene products. *Combinatorial Chem. High Throughput Screening, 4*, 193–205.

222. Yanofsky, S. D., Baldwin, D. N., Butler, J. H., Holden, F. R., Jacobs, J. W., Balasubramanian, P., Chinn, J. P., Cwirla, S. E., Peters-Bhatt, E., Whitehorn, E. A., Tate, E. H., Akeson, A., Bowlin, T. L., Dower, W. J., Barrett, R. W. (1996). High affinity type I interleukin 1 receptor antagonists discovered by screening recombinant peptide libraries. *Proc. Natl. Acad. Sci. USA, 93*, 7381–7386.

223. Lowman, H. B., Chen, Y. M., Skelton, N. J., Mortensen, D. L., Tomlinson, E. E., Sadick, M. D., Robinson, I. C., Clark, R. G. (1998). Molecular mimics of insulin-

like growth factor 1 (IGF-1) for inhibiting IGF-1: IGF-binding protein interactions. *Biochemistry, 37*, 8870–8878.

224. Fairbrother, W. J., Christinger, H. W., Cochran, A. G., Fuh, G., Keenan, C. J., Quan, C., Shriver, S. K., Tom, J. Y., Wells, J. A., Cunningham, B. C. (1998). Novel peptides selected to bind vascular endothelial growth factor target the receptor-binding site. *Biochemistry, 37*, 17754–17764.
225. Wiesmann, C., Christinger, H. W., Cochran, A. G., Cunningham, B. C., Fairbrother, W. J., Keenan, C. J., Meng, G., de Vos, A. M. (1998). Crystal structure of the complex between VEGF and a receptor-blocking peptide. *Biochemistry, 37*, 17765–17772.
226. DeLano, W. L., Ultsch, M. H., de Vos, A. M., Wells, J. A. (2000). Convergent solutions to binding at a protein–protein interface. *Science, 287*, 1279–1283.
227. Ferrer, M., Harrison, S. C. (1999). Peptide ligands to human immunodeficiency virus type 1 gp120 identified from phage display libraries. *J. Virol., 73*, 5795–5802.
228. Winssinger, N., Ficarro, S., Schultz, P. G., Harris, J. L. (2002). Profiling protein function with small molecule microarrays. *Proc. Natl. Acad. Sci. USA, 99*, 11139–11144.
229. Kuruvilla, F. G., Shamji, A. F., Sternson, S. M., Hergenrother, P. J., Schreiber, S. L. (2002). Dissecting glucose signalling with diversity-oriented synthesis and small-molecule microarrays. *Nature, 416*, 653–657.
230. Blackwell, H. E., Perez, L., Stavenger, R. A., Tallarico, J. A., Cope Eatough, E., Foley, M. A., Schreiber, S. L. (2001). A one-bead, one-stock solution approach to chemical genetics: Part 1. *Chem. Biol., 8*, 1167–1182.
231. Clemons, P. A., Koehler, A. N., Wagner, B. K., Sprigings, T. G., Spring, D. R., King, R. W., Schreiber, S. L., Foley, M. A. (2001). A one-bead, one-stock solution approach to chemical genetics: Part 2. *Chem. Biol., 8*, 1183–1195.
232. Chen, J. (2002). Protein modulators made to order. *Chem. Biol., 9*, 543–544.

# HIGH-THROUGHPUT SCREENING: EVOLUTION OF TECHNOLOGY AND METHODS

MARTYN N. BANKS, ANGELA M. CACACE, JONATHAN O'CONNELL, AND JOHN G. HOUSTON

*Bristol Myers Squibb Co.*
*Pharmaceutical Research Institute*
*Wallingford, Connecticut*

1 INTRODUCTION
2 EVOLUTION AND INTEGRATION OF HT PLATFORMS INTO THE DISCOVERY PROCESS
Evolution of Instrumentation and Screening Formats
Miniaturization
Evolution of Cellular HTS Assays
Evolution of HTS Technologies into the Drug Discovery Process
Quality Control in the HTS Process
3 HTS AUTOMATED SYSTEMS: TECHNOLOGY AND PROCESS
Plate Design
Compound Management
Liquid Handling
Accuracy and Precision
Assay Detection
Automation
Data Analysis
4 ADVANCES IN HIGH-THROUGHPUT BIOASSAY TECHNOLOGY
Overview of HTS Assay Formats
Reagent Production
Designing Bioassays for HTS

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

Fluorescent Methods in HTS
Cell-Based Assay Design
Additional Advances in Cellular Assay Formats
References

## 1 INTRODUCTION

Over the last 20 years high-throughput screening (HTS) has become a successful, reliable component of the drug discovery process. The evolution of science and technology surrounding HTS approaches has progressed rapidly over this same time period: from its origins in low-throughput testing of natural product extracts for antimicrobial activity to today's heavily automated, industrialized drug discovery process that enables the screening of millions of compounds against a range of complex, biological targets. During this time, we have seen an array of unique technology solutions and waves of standardization come to the fore to create the process we now recognize and use.

So, what is the best description of HTS today? To a large extent, it can still be defined by the scale of its operation and its reliance on serendipity for a successful outcome. A process by which hundreds of thousands to millions of compounds are tested for activity against disease targets of interest with the goal of identifying truly active, progressible "hits" [1].

Creative insights and smart ideas have significantly improved our ability to identify a greater number and quality of hits, especially in the last 5 years, as each of the core components that constitute the lead discovery process have been systematically enhanced.

Compound collections, assay design techniques, HTS technology platforms, and data capture/analysis systems are the essential core components of the lead discovery process, and each will be discussed in detail in this chapter. Figure 1 describes a typical lead discovery process highlighting these key functions.

Compound collections are typically constructed from a variety of different sources such as in-house "legacy" compounds, external acquisitions, combinatorial libraries, and natural product samples. The science around improving the "diversity" and quality of the compounds in such a collection has been significantly enriched since the mid-1980s. At that time a paucity of knowledge and experience created compound collections of poor quality and dubious lineage. Today, compound collections are recognized as major assets within a companies discovery arsenal and are treated as such. Compounds are now typically stored in custom-designed, fully automated warehouses where compounds are formatted for screening, monitored for quality, and instantly retrievable for testing.

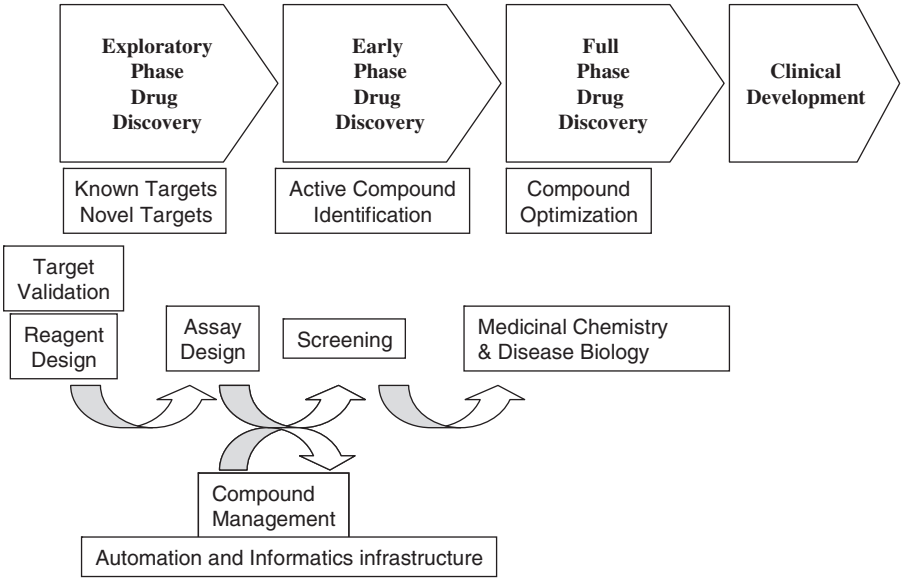


**Figure 1** Overview of the drug discovery process.

Judicious pruning and regrowth of the screening deck can greatly enhance the quality and diversity of the collection and significantly improve the success rates of the HTS campaigns (Bristol Myers Squibb in-house experience).

Design of a physiologically relevant and pharmacologically accurate assay is also a crucial part of the HTS process and can materially affect the successful outcome of a screening campaign. Converting a target of interest into a bioassay suitable for scale-up to HTS can be a rate-limiting step in the lead discovery process. The complex matrix of parameters that have to be taken into account in the design of an assay can include choice of assay mode (agonist or antagonist), availability of standard protocols, reagent availability, and scalability. More specific design elements, once a basic format and protocol have been chosen, will include optimization of key assay parameters such as pH, temperature, time or cell number, type of media, and the like. Today an array of assay design choices are available to the HTS specialist that are tried and trusted in the HTS environment.

Just as the choices for assay design have improved over the years, so has the range of HTS technology platforms used to run HTS campaigns. This area of HTS support can trace its roots to the days of ad hoc user groups sharing knowledge about automated solutions for labor-intensive tasks such as compound dispensing or reagent transfers. Nowadays a full range of automated equipment is available to the HTS specialist from the relatively simple, such

as modular workstations that can carry out plate-to-plate transfers, or reagent dispensing, or have detection capability, to the more complex, such as the multitasking, turnkey robotic arm, and track systems.

Some companies (e.g., GlaxoSmithKline) have taken the automation route to possibly its ultimate endpoint with the construction of drug discovery factories where the process of drug discovery is completely housed in a custom-built, fully automated building. For a basic description of the types of automation approaches available to the typical HTS lab, one can read Bojanic et al. [2].

The final major component of the lead discovery process is data management operations. For capturing the millions of individual data points emanating from a screening run to their visual display, detailed analysis, and interpretation, a whole host of in-house and commercially available software packages have now been developed.

High-throughput screening as a discovery process has benefited greatly from significant advances in automation, microtiter plate design, and bioassay techniques, as well as taking advantage of custom-designed compound and data management systems. It has evolved into a fully integrated, seamless process that the majority of pharmaceutical and biotechnology companies rely on to generate lead compounds to enable their drug discovery programs.

The HTS story brings together a plethora of different technologies, skill sets, and insights from a range of industries. This chapter will attempt to review some of the main features of the HTS platform that are essential for the effective functioning of the drug discovery process as well as illuminating some of the evolutionary milestones in HTS techniques over the last 20 years.

The twin demands of increasing the capacity to test compounds while maintaining, or even reducing, costs, has encouraged screening groups and the HTS support industry to come up with ever-changing screening formats utilizing state-of-the-art automation and miniaturization techniques. However, the development of an array of high-throughput screening formats starting from simple bacterial lawn assays, through spot and wash techniques, to homogeneous assay formats, each brought with them a set of technology problems that needed to be solved. Even today HTS scientists still need to consider a range of process issues such as how to get compounds to the assay, which type of assay to use, how to store compounds in readily available sets, guaranteeing the purity and stability of stored compounds, finding rapid and reliable ways to dispense compounds and reagents, development of sensitive detection devices to measure assay products, effective data capture and analysis tools to deal with the volume of data being generated. Each of these problems and questions will be touched on in this chapter.

This chapter will review the variety of screening plate formats that have been developed from the 96-well plate to the high-density 3456 plates and the implications of their usage on the supporting technologies and assay designs. Automation support from compound and reagent handling as well as compound storage architecture and reagent production will be discussed. The

range of bioassay techniques now available for use by HTS scientists will be summarized.

## 2 EVOLUTION AND INTEGRATION OF HT PLATFORMS INTO THE DISCOVERY PROCESS

The real impetus for HTS as we know it today came in the late 1970s. Even though the first prototypical microtiter plate was described in 1954 [3], the ubiquitous 96-well plate became popular with the increasing use of enzyme-linked immunosorbent assays (ELISA) during the late 1970s [4, 5]. The strengthening of three technology streams—reaction vessels, methods to reproducibly deliver small volumes of reagents, and methods of measuring the endpoint of an assay (absorbance, fluorescence, and luminescence)—has driven the evolution of HTS over the last 20 years.

Three other major developments have also been drivers for HTS evolution. The development of precise nonseparation bioassays, the ability to organize and distribute large numbers of chemicals [6], and the analysis of large volumes of data were critical for building the HTS infrastructure. The integration of these technologies from chemistry, biology, engineering, and information technology are the critical elements in performing a cost-effective and timely high-throughput screen [2]. We will describe the developments that led to the process of HTS including new instrumentation, homogeneous bioassays, assay miniaturization, cell-based assays, and strategies to deal with the large number of compounds identified in HTS (Fig. 2).

### Evolution of Instrumentation and Screening Formats

This evolution of instrumentation and screening formats is best described by example of the evolving methodologies used to identify compounds that affect

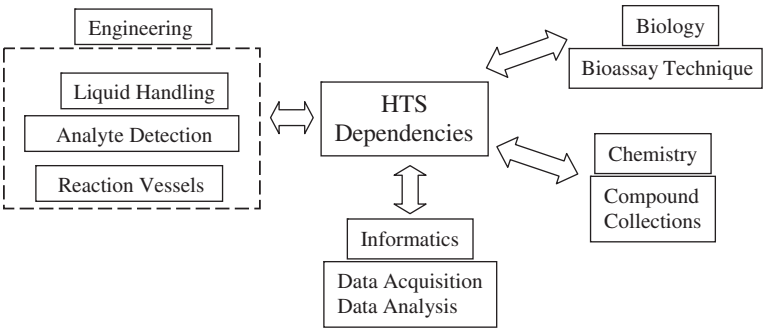


**Figure 2** HTS evolution has been dependent on the development of a whole range of different disciplines.

the activity of tyrosine kinases. Tyrosine kinases are critical for effective second-messenger signaling in many different cell types, and aberrant activity has been demonstrated in a variety of oncogenic states. More recently, anticancer drugs that modulate tyrosine kinase activity have shown efficacy in the clinic [7].

One of the first screening methods for identifying compounds that affect tyrosine kinase activity involves incubating enzyme, radiolabeled [$^{32}$P]–adenosine triphospate (ATP), and substrate in a tube. The phosphorylated substrate was then spotted onto P81 phosphocellulose paper. After washing to remove residual [$^{32}$P]-ATP, the amount of phosphorylated substrate was quantified in a $\gamma$ counter [8]. This method was very tedious and low throughput. Improvements in throughput were made by running the assay in 96-well microtiter plates, spotting reaction products onto a large sheet of P81 paper using a 12-channel pipette, and then washing the entire sheet to remove radiolabeled ATP. The dried paper was then exposed to photographic film, and spots that produce lower levels of exposure were indicative of tyrosine kinase inhibition. To determine the degree of inhibition, individual zones corresponding to each well of the microtiter plate were cut from the P81 paper and placed separately into a scintillation counter for Cherenkov counting. Alternatively, the film could be semiquantitatively assessed based on the intensity of the developed spot or zone. This approach was further refined by the introduction of the phosphoimagers and gas flow imagers [9] in the mid- to late 1980s. These innovations provided a means of quantitation that eliminated the need to cut up individual squares.

The next important step in the evolution of screening formats was the use of immunoassays to quantitate levels of phosphotyrosine [10]. In this screening method, a synthetic tyrosine kinase substrate, for example, PolyGlu-Tyr, was adsorbed onto the surface of a microtiter well. Unbound material was removed using a 96-well plate washer, and enzyme, ATP, and putative inhibitors were incubated in the substrate-coated well. After a brief incubation, the contents of the well were removed and the phosphorylated substrate was detected by a phosphotyrosine-specific antibody using stardard ELISA techniques. This new format offered significant advantages; it was nonisotopic, it increased parallel operations, and used the plate readers and washers that had been developed by engineers to support the growing ELISA market. However, the method also had significant disadvantages; the assay still involved multiple steps and washing the microtiter plates was tedious. Today, when it is commonplace to screen 1 million compounds, this format would involve 10,400 plates and hundreds of liters of wash fluids. The reader should be aware that there were significant scientific issues with this methodology, specifically the nonphysiological nature of the substrate and its solid-phase disposition in the assay.

Importantly during this period, new technologies were emerging that allowed the development of a filtration assay. Harvesters that took the contents of 96-well microtiter plates through a filter mat containing 96 zones in a

6 × 16 array were developed based on a concept designed by Warner and Potter [11, 12]. The filter mat was automatically washed as part of the harvesting action and was then ready for quantitation. A range of substrates could now be used, including protein substrates that could be acid precipitated and the radiolabeled product captured on a glass fiber mat. In addition, new scintillation counters were developed that could read the incorporated radioactivity of the entire filter. The new counters used 6 photomultipliers that significantly reduced the time required to measure the whole filter mat.

A significant improvement in high-throughput screening came with the introduction of the homogeneous assay, where the reaction could be measured without a separation stage. This was critical to HTS evolution because it allowed higher precision, a more rapid throughput, and led the way for further miniaturization. One early problem with homogeneous assays was the susceptibility to give false-positive results, for example, as in scintillation proximity assays (SPA) [13]. In a scintillation proximity assay, beads containing scintillant are coated with substrate and incubated with the tyrosine kinase, radiolabeled [$^{33}$P]-ATP and putative inhibitors. Compounds that interfere with tyrosine kinase activity decrease the amount of radiolabeled [$^{33}$P] in proximity to the bead containing the scintillant and subsequently decrease the light output. At that time, the SPA beads used emit light at 420 nm; however, a significant number of colored compounds will absorb light in this part of the spectrum. By reducing the light output, this data could be misinterpreted as inhibition of tyrosine kinase activity. Concomitant with the emergence of SPA was the development of the 96-well scintillation counter (Topcount, Packard Instruments, USA, and Microbeta, Wallac Oy, Finland), a necessary prerequisite for HTS because this new instrument accepted the 96-well plate directly and had the ability to quench correct for colored compounds to a certain extent.

## Miniaturization

Until the early 1990s the 96-well plate was the standard format for HTS, however, the development of new assay technologies and pressure from HTS laboratories facilitated the design of new microplate formats. The size of the wells was decreased to yield 384 wells that had the same footprint as a 96-well plate. This reduced the volume of assays by a factor of 4. Instrumentation companies rapidly adapted to this change with liquid-handling devices that had better spatial resolution and were capable of delivering smaller volumes. Detection devices were also improved to give higher sensitivity and higher throughput.

The late 1990s brought further miniaturization and changes in plate design with higher density plates such as 864-well plates from Affymax, 9600-well plates from Dupont Pharmaceuticals, and 3456-well plates from Aurora Biosciences. The use of plates with higher well densities required significant adaptation of liquid handling and detection technologies. These changes in plate

design posed significant problems not seen in the evolution of 96-well to 384-well platforms. For that transition, plate readers for absorbance and fluorescence measurements were dependent on photomultiplier tubes that required significant read times. A 96-well scintillation counter with a mask could be adapted to allow each well to be read in a 384-well assay, however, the read time increased by a factor of 4, which compromised throughput. Measurement of 1536-well plate assays was improbable using photomultiplier technology. The implementation of charged coupled devices (CCD) that through a telecentric lens could image the whole plate in one measurement allowed measurement of assays using higher density plates [14]. This imaging innovation also had other advantages. For example, the same instrument could be used for luminescence, fluorescence, and fluorescent polarization readouts from any microtiter plate format. In addition, the CCD chip has a light detection sensitivity that is more red shifted than photomultiplier tubes and therefore could better accommodate a wide range of fluorescent dyes. New SPA beads were developed to work with CDD imagers that shifted the wavelength of the emitted light to 620 nm; this in turn reduced the false-positive interference from colored compounds. Today, we have automated assay platforms that can routinely run low-volume assays (5 to 10 μL) with high precision and process the evaluation of millions of compounds in 2 to 3 weeks. Advancements in bioassay technology, instrumentation, control software, and efficient processes have now provided the drug hunter with very powerful tools [2].

## Evolution of Cellular HTS Assays

Cell-based assays for HTS first emerged in the mid-1980s. These assays were known as *black box screens* where growth factors, cytokines, or hormones produced a change in growth of cells, but it was not clear how these signaling events culminated in a cellular response. The early measurements of cellular proliferation used [$^3$H]-thymidine incorporation as an indicator of deoxyribonucleic acid (DNA) synthesis [15–17]. It was not until the late 1980s that redox dyes (3-[4,5-dimethylthiazol-2-yl]-2,5 diphenyltetrazolium bromide), which measure the oxidative potential of the mitochondria as an indicator of cell proliferation, became available for the HTS scientist. This enabled the development of homogeneous proliferation assays that were amenable to automated systems and could be measured using simple plate readers [18, 19]. Advances in molecular biology were critical to the engineering of mammalian cells that could express targets, cofactors, and reporters for cell-based HTS [20]. The earliest reporter assays, described in 1982, included chloramphenicol acetyltransferase and β-galactosidase, and these were rapidly followed by luciferase, aequorin, and green fluorescent proteins (from *A. victoria*), and secreted alkaline phosphatase [21–26]. Reporter assays enabled the study of complex cellular signaling pathways that could be modulated pharmacologically. For luciferase-based assays, luminometers with integrated pipetting are required to measure flash luminescence [27]. The advance of

"glow" luminescence enabled conventional microplate scintillation counter technology to capture and measure light from these assays [28]. Reporters such as β-lactamase and renilla luciferase, which measure real-time changes in gene expression in living cells, have also been developed [29–31]. In addition, there are synthetic versions of click beetle luciferase genes for dual-color measurements in multiplexed reporter assays allowing multiple signaling pathways to be interrogated as one time [32]. Another major advance in cell-based screening came with the ability to directly measure changes in the levels of second messengers (calcium, cyclic adenosine monophosphate (cAMP), inositol-3-phosphate (IP3), and mitogen activated protein (MAP) kinase) within the cell in an automated screening environment [33–36]. Advances in automated fluorescence microscopy enabled more complex measurements within cells such as phenotypic changes, intracellular protein translocation, and intracellular posttranslational modifications [37, 38]. Until recently with the development of instruments like Patch Xpress, patch clamp for single-cell electrophysiological measurements eluded automation [39]. Advances such as those mentioned above have rapidly increased the number of cell-based screens and also allow the screener to run assays against targets in the appropriate cellular context.

## Evolution of HTS Technologies into the Drug Discovery Process

The developments discussed here describe an HTS technology base capable of assaying a wide range of in vitro and cellular targets, which has resulted in the identification of many active compounds than can be progressed into medicinal chemistry programs. To prioritize which compounds progress in the drug discovery process, medicinal chemists consider several factors, including potential liabilities, compounds selectivity, drug–drug interaction problems, drug metabolism, and absorption or permeability issues.

Figure 3 shows a typical flow of activity through a traditional drug discovery program. Primary assays are performed to give feedback on a medicinal chemistry hypothesis; secondary assays are then performed to provide more detailed data. This then is followed by selectivity information and then liability information. At each step, it is typical to reduce the number of compounds being progressed. This serial cycle then repeats as the chemistry hypothesis evolves. Compression of this cycle improves the efficiency of drug discovery. One way to achieve this increase in efficiency is to use parallel approaches, such that data on primary bioassay, selectivity, and potential liability can be assessed together by the medicinal chemist (Fig. 4). This requires many assays to be run simultaneously but on fewer compounds than a typical HTS.

One major difference between the HTS operation and the parallel lead optimization process is the way that compounds are moved through the process. For HTS a large collection of compounds is assembled, usually divided into a series of master compound plates, all of the same format, and all

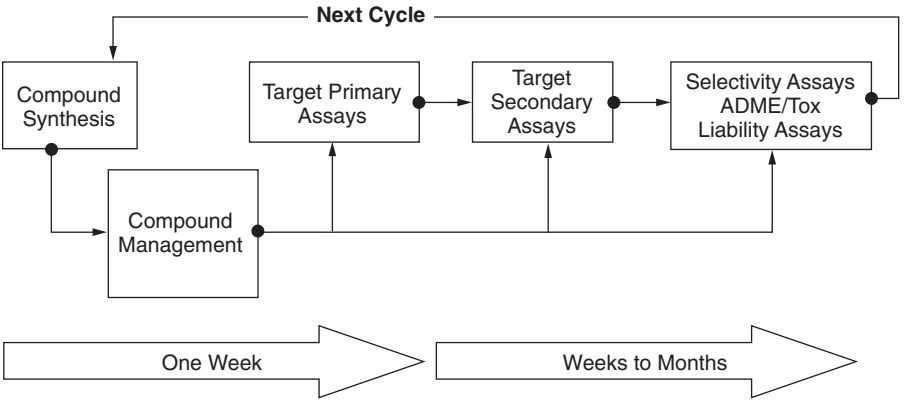


**Figure 3** Serial workflow diagram that is typical of many drug discovery programs.

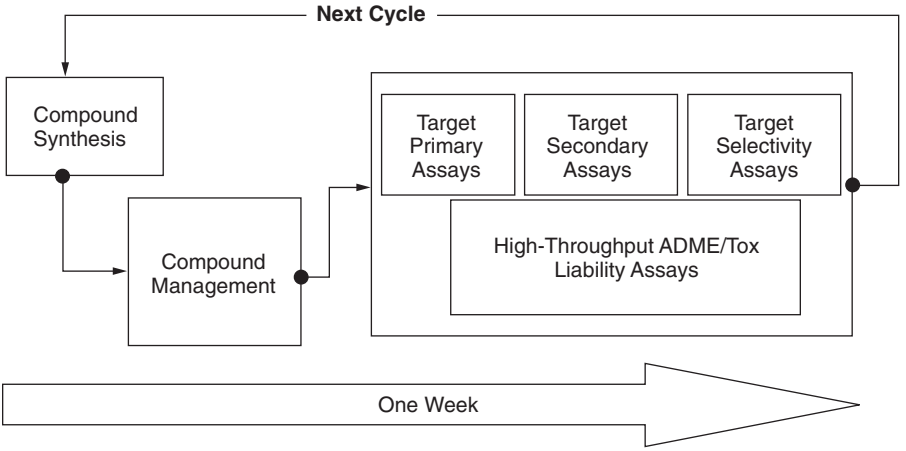


**Figure 4** Parallel workflow diagram demonstrates the advantage of the process infrastructure that has been developed for high-throughput screening.

screened for bioactivity against a target. For lead optimization a small number of compounds arrive from a series of different programs. These compounds then have to be routed to a range of different assay suites that may exist in different assay format, for example, 96-well, 384-well, or higher. For a given drug discovery program, the type of bioassay changes as it progresses toward late-stage drug discovery and new assays complement or replace older ones. Also, critical compounds may need to be checked against a broad range of assays, rather than just the program-specific assays. This process is summarized in Figure 5.

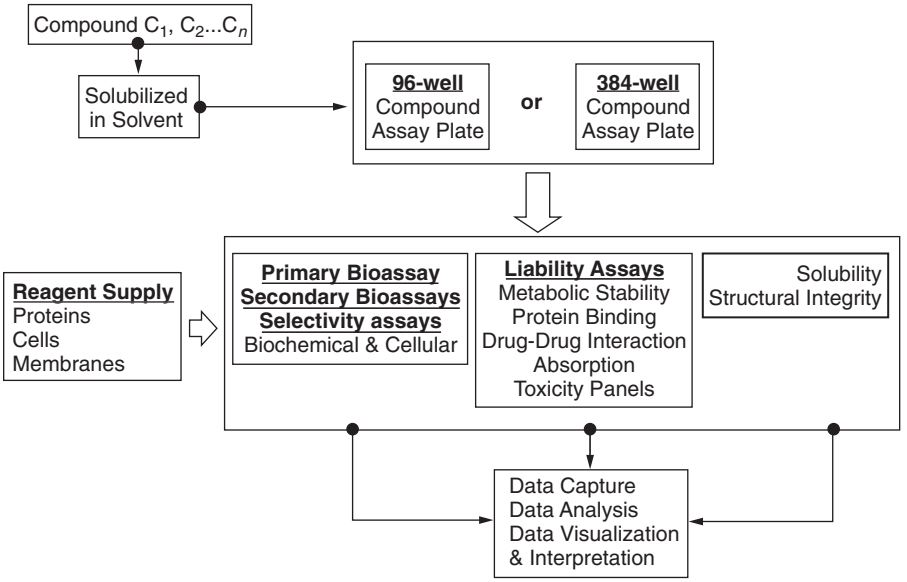


**Figure 5** Detailed flowchart of the compound distribution, reagent supply, assay, and data analysis that needs to be synchronized to achieve an efficient parallel lead optimization process.

## Quality Control in the HTS Process

The quality of the data from HTS assays is dependent upon the supply of robust reagents, and this is an integral part of the process. Some reagents such as proteins and membranes can be stockpiled and quality controlled off-line. For cell-based assays, cells may need to be provided at the time of the assay. New automated cell culture machines greatly facilitate this process (e.g., SelecT from the Automation Partnership, UK).

Assay accuracy and precision is critical to the success of HTS strategies and is constantly monitored through the use of standard compounds and blank plates. Hardware quality control requires preventative maintenance and daily calibration. By integrating quality control ranges in the robotic control software, the assay performance can automatically monitor and alert scientists to any issues in prosecuting the assay. By successfully completing the quality control limits set on the assay, the data then moves directly to databases that intelligently process the raw data. An example can be found in drug concentration response analysis where automated curve-fitting routines are embedded into data analysis software and performed automatically. Here curves are fit upon comparison to a range of pharmacological rules. These rules determine whether there is a reasonable curve that fits the rules applied to minimal or maximal activity or slope. If there is a reasonable pharmacological fit of the data, it then streams into a data-reporting information technology (IT) tool.

The scientist is alerted when data points fail to fit using the IT tool, and the responsible individual can make the necessary judgment call to repeat the experiment.

For selectivity or liability one might want to test the compound at a concentration significantly higher than that used in the primary bioassay. Therefore, it is important to understand the solubility of the compound under the assay conditions. From our own experience, a compound's liability can be overlooked due to poor solubility in the liability or selectivity assay, and these false negatives can misdirect a drug discovery program. Inclusion of structural integrity analysis of the compound being tested into the screening process is also an important quality control parameter. Structural integrity information confirms the identity of the compound tested in the bioassay and also verifies that the compound has not degraded during the process. In the past 20 years, we have witnessed a dramatic change in the way that bioassays are performed from both a scientific and technology perspective, and this has been driven by multidisciplinary teams involved in engineering, biology, and chemistry.

## 3 HTS AUTOMATED SYSTEMS: TECHNOLOGY AND PROCESS

The last 10 years have seen a significant increase in the size of pharmaceutical compound collections from hundreds of thousands to millions. With the advent of HTS in the early 1990s, a typical lead discovery organization would have been expected to screen between 10 and 20,000 compounds through 15 targets a year. Today, many companies can screen 1 to 5 million compounds through 20 to 100 targets per year with each campaign lasting only a matter of weeks [40].

The areas that will be discussed here with relevance to screening large numbers of compounds are: advances in plate design, compound management, liquid handling, assay detection, automation for HTS, and data analysis.

### Plate Design

One of the most significant advances in HTS technology over the last 15 years has been the ability to reduce assay volume and increase microtiter plate well density. In the mid-1990s, the predominant assay format was the 96-well plate [41]. By the late 1990s to the early 2000s, assay formats moved to much higher well-density plates (Table 1) [41–44]. Advances in plate design required a combination of enhanced liquid handling technology to deliver smaller volumes and enhanced reader technology to measure assay output from >96-well plates. The evolution into the 384-well plate was dependent upon the ability of liquid handling equipment and assay detectors to manage both the 96- and 384-well plate. For example, instrumentation with 96 tips can pipette in 4 quadrants to a 384-well plate and likewise for the transition from 384-well

**TABLE 1 Assay Formats Including Plate Density and Reaction Volumes**

| Plate Format | Well Layout | Max Volume (μL) | Typical Assay Volume (μL) |
|---|---|---|---|
| 96 | 12 × 8 | 250 | 100 |
| 384 | 24 × 16 | 80 | 50 |
| 1536 | 48 × 32 | 10 | 5 |
| 3456 | 72 × 48 | 2.2 | 2 |

**TABLE 2 HTS Assay Plate Types**

| | | |
|---|---|---|
| Clear | Absorbance assays | Clear light path for light to pass through |
| White solid | Luminescence | Reflective surface to maximize light output |
| White, clear bottom | Luminescence | Reflective but allows bottom read |
| Black solid | Fluorescence | Minimizes background |
| Black, clear bottom | Fluorescence | Low background but allows bottom read |

plate to 1536-well plate (4 × 384). The ultra low volume 3456-well plate (36 × 96) was produced by Aurora, when the 96-well plate was still the predominant assay format for HTS and was considered a more revolutionary plate development process. The principles of assay design remain the same regardless of plate format. In reality, the success of an HTS assay depends on the ability to dispense liquids in the volume ranges indicated in Table 1. Other critical factors to consider when adopting a higher density plate format for HTS include length of detection in terms of instrument read times and changes in surface area to volume ratios. In the case of fluorescence assay detection, the read times vary, and this is dependent on the choice of instrument, for example, a single device based on a photomultiplier tube (PMT) could take up to 15 min to read an entire 1536-well plate on a well-by-well basis. This means that well number 1536 will have been incubated for 15 min longer than well number 1 unless the assay is stopped and has a stable signal. Finally, the plastic binding surface is larger in ratio to total volume in higher density formats. While this has not been reported to be an issue, in our hands, this surface-to-volume ratio will have an effect if the assay reagents are particularly sticky or hydrophobic. Hydrophilic coated plates can help minimize this effect. The number and variety of different plate types that are available is quite significant within each of these formats. Table 2 describes the range of plate types that are currently available for HTS.

There are also a number of plate coatings that are available for a multitude of bioassay applications. There are too many to discuss here, but they include antibody coatings to bind proteins, scintillant coatings for specific radiochem-

ical applications, poly-D-Lysine, and collagen for cellular adhesion and the previously mentioned low binding plates.

Three different plastics are used for microtiter plates throughout the HTS process. The first is polypropylene, which is preferred for compound storage because it is solvent tolerant and has low binding properties for compounds. The second is polystyrene, which is the preferred assay plate due to rigidity for automation, flexibility to be colored (as described in Table 2), and use for a variety of assay technologies. The third is cyclo-olephin copolymer (COC), which is a newer plate type with potential for compound storage since it has lower binding properties for compounds and has excellent rigidity as well as optical clarity.

## Compound Management

As discussed above, the compound collection of most pharmaceutical companies has grown significantly and collections are now in the 1 to 5 million compound range [45]. This presents challenges for both storage and distribution. A number of key considerations are discussed below.

***Composition*** Historically, the pharmaceutical compound collection has typically included legacy compounds that a company has made for discovery programs over many years. The main issue with these legacy compounds is that this type of collection lacks diversity since it was made to target specific proteins. Examples of this from our own experience include work that generated many $\beta$-lactams for antimicrobial work and a large number of steroids in the late 1980s. Still, today a large number of compounds within screening decks are made up of $\beta$-lactams and steroids for exactly this reason. Consequently, there has been a focus within pharmaceutical companies to increase the number of compounds within screening decks to increase chemical diversity and remove compounds that are unlikely to progress to lead compounds.

In HTS, it is important to screen a chemically diverse set of compounds to find novel leads. However, measurement of diversity has been, and still is, a very contentious issue. In order to have an optimal screening deck that covers all permutations of chemical diversity, a compound collection would have to be on the order of $10^{60}$ compounds [46]. Clearly, it is not possible to screen such a large number of compounds, so efforts are made to sample this diversity and to populate the collection with compounds that have potential for progression into the drug discovery process. This is no trivial task, and yet it is critical to the success of HTS because any HTS campaign can only be as successful as the compounds it tests. The success of an HTS campaign is measured by the number of novel chemotypes that progress into full medicinal chemistry programs, where the definition of a chemotype is a compound or set of compounds that have a unique structure or chemical composition.

There are many companies that build libraries of compounds and sell them to pharmaceutical companies to populate their compound collections. This is

a popular approach since the cost of synthesizing compounds is high from a personel perspective. However, it can also mean that the same library is being tested by many companies, often against the same targets. In this case, the chance of finding a compound that is different from a competitor is low, and so this shifts the emphasis to the speed with which a company can discover and progress a compound first. Additionally, the use of combinatorial libraries can mean that chemical diversity within a library is poor since they tend to be built on common scaffolds. However, if these libraries contain active compounds based on similar chemical scaffolds, then some level of structure–activity relationships (SAR) can be derived very rapidly from the HTS.

Another approach to organizing a compound collection is to group compounds into categories of potential activity, that is, kinase or G-protein-coupled receptor (GPCR) activity. This enables a smaller subset of compounds to be tested against targets that are either very early in the discovery process or considered too challenging (i.e., cost prohibitive, limited reagents, etc.) to screen the entire compound collection [47, 48].

***Wet or Dry Compound Storage*** Wet or dry refers to storing compounds as solids or dissolved in solvent. The generic standard for supplying compound to a bioassay is in dimethyl sulfoxide (DMSO). This is the standard for solubilizing the majority of compounds in a screening collection, but there are a number of significant issues relating to storage of compounds in a universal solvent such as DMSO [49]. First, a number of compounds are not soluble in DMSO. Thus, the predicted concentration of compound will never be attained in the bioassay. Second, DMSO is very hydroscopic, so, when exposed to the atmosphere, it adsorbs moisture and will continue to do so until it reaches equilibrium at about 70 percent DMSO, 30 percent $H_2O$. Now the compound is diluted, and solubility under these aqueous conditions may be adversely effected, both of which alter the effective compound concentration in the bioassay. Third, DMSO can dissolve certain plate materials and these may have an adverse effect on the bioassay. Finally, DMSO is reactive with some compounds, especially in the presence of oxygen, thereby effecting the integrity of compounds in storage.

There have been many different solutions to address the above issues that vary in complexity with the ultimate goal of maintaining a completely dry environment for the storage and distribution of compounds. Recent literature [49] suggests that it is best to store compounds in DMSO in an inert, dry atmosphere at room temperature, thus avoiding freeze–thaw cycles. The investment to achieve such storage conditions may be significant, especially when the size of the compound collection is considered. For example, Figure 6 shows a photograph of the Haystack liquid store at Bristol-Myers Squibb that is two stories high. Not only must the environment where the compounds are stored be controlled, but the environment where compounds are dissolved, transferred to and from the store, replicated for screening, and stored for screening must also be controlled. Given the cost of full environmental control of

**Figure 6** Photo of Haystack (The Technology Partnership, Cambridge, UK). The haystack compound storage system at Bristol-Myers Squibb showing the "wet" tube store.

their compound collections, some companies have taken the route of making up compounds in either 90 percent or 70 percent DMSO from the outset and then using less extensive environmental control. While less expensive, this compromise will reduce the aqueous solubility of certain compounds.

Long-term storage of compounds is optimal when compounds are stored as dry samples and preferably at low temperature. Thus, most companies have a dry-compound store, often in tubes as well as a wet store. At periodic intervals, samples from the dry store are taken to replenish the wet store in DMSO.

***Compound Storage in Plate Formats*** Compound storage in plate formats should reflect the most typical HTS usage plate format. Storing compounds in 384-well format is most efficient if screens are run in a 384-well format. When preparing a compound plate for a bioassay, the replication cycle for compound transfer is from one plate to a daughter plate (or many identical daughter plates), rather than four 96-well plates to one 384-well plate. However, this may not be the most efficient process when individual compound samples

are required since an entire plate would need to be removed just to pick an individual compound. Consequently, multiple formats are used to store compounds.

For example, the Haystack system at Bristol-Myers Squibb in Connecticut (Fig. 6) has three components. The first is a dry store where compounds are stored as powders or films. The second is a tube store that stores compounds in 100 percent DMSO in sealed microtubes. These microtubes are made from the dry tubes by adding volatile solvent to the dry tubes, transferring the required volume out, and then drying the sample within the original tube back down for storage. These microtubes are used to generate master plates for HTS and for the selection of individual compounds. The master plates for HTS are then stored in the third area of Haystack where they can be retrieved to make assay plates for HTS.

***Replication for HTS*** As discussed above, storage temperature and environmental conditions will have a significant effect on the transfer of compounds into screening plates, otherwise known as the compound replication process, so only generalities are discussed here. The key points to consider are that freeze–thaw cycles should be minimized as should DMSO exposure to the atmosphere. Figure 7 represents a typical compound replication process.

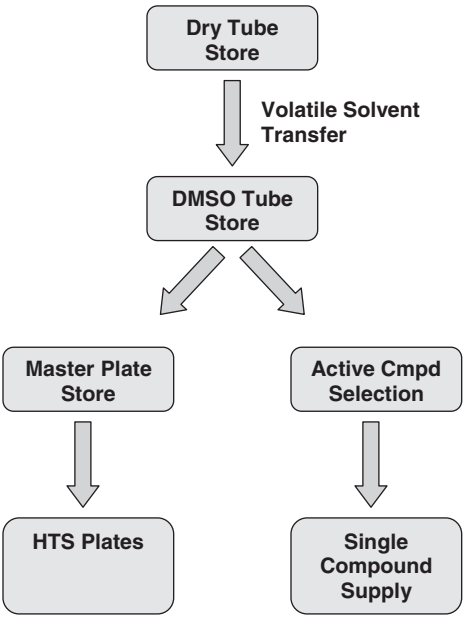


**Figure 7** Compound storage. Flowchart representing the typical process for compound storage and retrieval using the Haystack automated store at Bristol Myers Squibb Co.

Dry compound from a tube store is used to make a number of master plates in DMSO. Periodically, a master plate is used to make a number of screening plates ready for bioassays.

The important factors to consider are the tolerance of the assay to DMSO and the concentration of compound. A number of biochemical assays can tolerate higher DMSO concentrations in the 2 to 5 percent range that enables plates to be used directly, but cellular assays rarely tolerate higher than 1 percent DMSO and inevitably require an intermediate dilution unless nanoliter dispensing is available (see next section). Typical compound concentrations for screening are in the 10 μM range and typical compound storage concentrations are in the 2 mM range.

***Nanoliter Compound Dispensing*** As discussed above, some assays have a low tolerance for DMSO. This means that as the volume of the bioassay decreases when moving from a 384-well to 1536-well plate, the volume of DMSO to be added needs to decrease proportionally. For example, if a typical 1536-well assay volume is 5 μL, to attain a DMSO concentration of 1 percent without using an intermediate aqueous dilution, a compound addition of 50 nL is required [50, 51].

The value of avoiding an intermediate aqueous dilution during screening is illustrated in Figure 8. Here, $IC_{50}$ determinations for 1000 compounds were performed in a standard enzyme assay via two routes. The first using an inter-

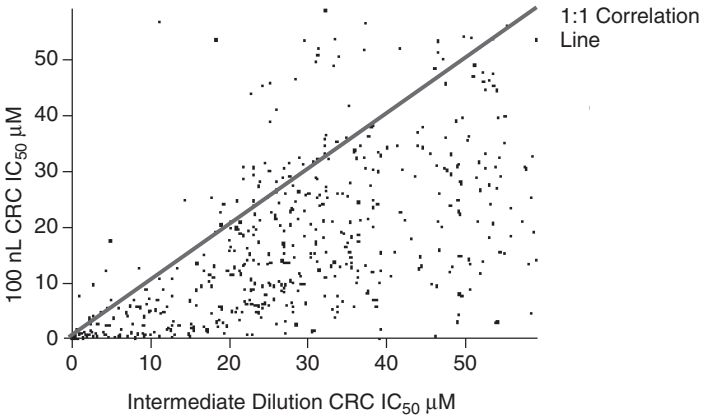


**Figure 8** Correlation of intermediate vs. 100% DMSO compound transfers where the graph represents the correlation of 1000 $IC_{50}$ determinations performed for an enzyme-based target via two routes. (1) Making up compound in 100% DMSO, creating a 3-fold, 10-point serial dilution in 100% DMSO. Performing a 10-fold dilution in water and transferring 1 μL to the assay plate. Represented on the *x* axis. (2) Making up the compound in 100% DMSO, creating a 3-fold, 10-point serial dilution in 100% DMSO. Transferring 100 μL direct to the assay plate. Represented on the *y* axis.

mediate aqueous dilution and the second using a 100-nL dispense of compound in 100 percent DMSO directly to the assay plate. As can be seen, the majority of compounds appear more potent via the 100-nL method. The enzyme used in this assay had a reasonably hydrophobic binding pocket, and a number of the active compounds tested were hydrophobic in nature. The difference in potency occurs because of the decrease in solubility of hydrophobic compounds in aqueous solution. Thus, if the compound is at a high concentration in 100 percent DMSO, it is more likely to drop out of solution when it is diluted to a low concentration in aqueous buffer.

It is only in recent years that the technology has been made available to dispense compounds in the low nanoliter range [50]. A significant advance came in 2003 with the onset of acoustic dispensing. This technology uses ultrasound to eject nanoliter droplets from the surface of liquid in a plate that contains compounds and fires the droplets into an inverted destination plate above. This liquid handling technique is very precise and accurate, and, because this dispense method is noncontact, no washing pipette tips are involved.

***Closed-Loop Screening*** Closed-loop screening is a term used to describe the process where an assay is run continuously from initial screening to $IC_{50}$ determination. The initial screen assays a large collection of compounds at a single compound concentration, and then active compounds are confirmed on that same system from the same source compound plate. Critical to this process is the ability to carry and store a large number of compound plates on the screening platform. Thus, when the assay is running, HTS assay plates can be prepared from these on-line compound plates. The clear advantage in this approach is that reconfirmed positive compounds are tested continuously from the original plates stored on the assay system, and the dependency on a compound management department to select these compounds from a large store is circumvented. This benefit has become more important as the size of compound collections has increased, for example, if a compound collection is 1 million compounds and of these 1 percent are active, then 10,000 compounds will need to be supplied to a confirmatory assay.

## Liquid Handling

The key challenge for either the bench scientist or the automated screening platform is reliably dispensing all components of the assay with sufficient accuracy and precision. Working in the larger plate volumes such as 100 μL for the 96-well plates means that individual assay components need to be dispensed in the volume range of 10 to 50 μL. As the well density of the plate increases, the volume of each well decreases, and accurate liquid dispensing becomes a more significant challenge. Table 3 summarizes the different types of liquid handling devices that are available, along with their working volumes. However, before getting into detail here, it is important to consider

**TABLE 3 Summary of Liquid Handling Types**[a]

| Type | Plates | Volume Range (μL) |
|---|---|---|
| 8- to 16-tip variable span | 96, 384 (some 1536) | 0.5–100 |
| 8- to 16-tip fixed span | 96, 384 (some 1536) | 0.5–100 |
| 96 head, tips | 96, 384 | 2–100 |
| 384-head, tips | 384, 1536 | 0.5–50 |
| Peristaltic pump | 96, 384 | 5–200 |
| Solenoid valve | 96, 384 1536, 3456 | 0.2–10 |
| Pin tools | 96, 384 1536, 3456 | 0.005–0.1 |
| Acoustic | 96, 384 1536, 3456 | 0.002–0.2 |

[a] This table represents typical averages only.

the parameters that define the liquid handling performance: accuracy and precision.

## Accuracy and Precision

Liquid handling needs to be both precise and accurate. The term *accuracy* simply refers to the question: "Does this instrument dispense the absolute volume that has been requested?" Thus if 10 μL were to be dispensed and the average of all dispenses was 10 μL, then instruments would have excellent accuracy. The variation in dispense volume is more a question of precision. Consequently, accuracy is often represented as percent change from nominal. For example, an accuracy of 10 percent on a 10 μL dispense means that all dispenses will be between 9 and 11 μL. For the majority of instrumentation used in HTS, it is expected that they will have an accuracy of <10 percent of nominal.

Precision is a measure of how variable a dispense is and takes into account the standard deviation of all dispenses. Precision is typically expressed in terms of the coefficient of variation expressed as a percentage (%CV). It is standard practice to aim for liquid handler performances in the with %CV values of <5 percent.

It is imperative to continuously monitor the performance of an instrument to prevent waste of valuable reagents and data. Both precision and accuracy are essential to the understanding of instrument performance. For example (Table 4), analyzing the accuracy alone would indicate that either instrument 1 or 3 performed the best, since in each case the average dispense is 10 μL. If one measures precision alone, it would indicate that instrument 2 performed best since it has the lowest %CV, indicating that all dispenses are very similar. Close interpretation of the data, however, suggests that instrument 1 performed with variable dispense. Instruments 2 and 3 are more difficult to differentiate as 2 gives precise data but over dispenses, yet 3 exhibits more variation but is more accurate. It is also important to test the instrument with

**TABLE 4 Dispense Volumes for Specific Instrument[a]**

| | Instrument 1 | Instrument 2 | Instrument 3 |
|---|---|---|---|
| Dispense volumes (μL) | 8.1 | 10.3 | 9.1 |
| | 9.5 | 10.2 | 11 |
| | 9.1 | 10.2 | 9.5 |
| | 1.6 | 10.2 | 9.4 |
| | 12 | 9.9 | 10.5 |
| | 8.8 | 10.1 | 10 |
| | 10.2 | 10 | 10.7 |
| | 11.2 | 9.9 | 11 |
| | 12.6 | 10.3 | 9.2 |
| | 13.1 | 10.1 | 10.4 |
| | 9.9 | 10.4 | 10.2 |
| | 14 | 10.4 | 9.2 |
| Average dispense | 10.0 | 10.2 | 10.0 |
| % CV | 32.2 | 1.7 | 7.1 |
| Number outside 10% | 7.0 | 0.0 | 2.0 |

[a] Shows the actual volume dispensed when a 10-μL volume was requested over 12 replicates. This data is used to analyze precision and accuracy.

a range of buffers and reagents. The precision and accuracy will vary significantly with the reagents that are used. For example, glycerol may stabilize proteins but it increases viscosity. Bovine serum albumin (BSA) may also stabilize the assay reagents but it is liable to foam as do detergents. Additionally, it is very important to examine if the liquid handler creates a pattern following dispensing that could indicate that a pipette tip is clogged or not functioning. For example, if an instrument has 384 tips, each well is dispensed independently into a 384-well plate, but in a 1536-well plate each tip dispenses to 4 adjacent wells so it is important to look for differences in blocks of 4. If an instrument has 8 tips that dispense in a linear pattern across a plate, then differences between rows or dispense patterns within a row may occur. For example, the Labsystems MultiDrop dispenses in the pattern indicated in Figure 9. Panel 1 demonstrates an obvious dispense pattern where the columns show a low initial dispense. Panel 2 shows a typical dispense with no patterns and a good level of precision. Finally, the liquid handling instruments are validated as part of transitioning a bioassay from the laboratory bench to an automated system. This final test ensures that the instrument will function with precision and accuracy during routine use with a variety of assay reagents including liquids of varying viscosities, beads, and cells.

## Assay Detection

In HTS there are a number of assay technologies that use a range of detection methods. This list includes absorbance, fluorescence, fluorescence reso-

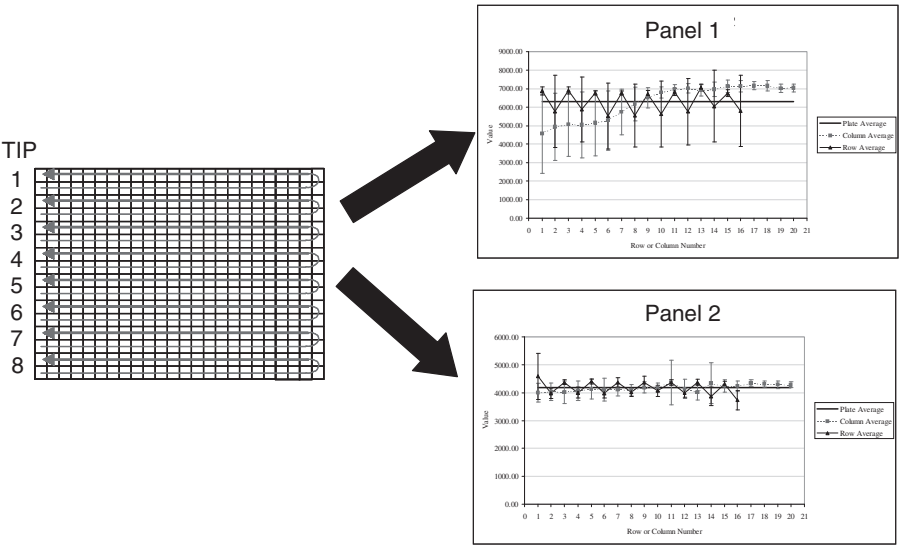


**Figure 9** Dispense patterns from an 8-tip device in a 384-well plate shows the dispense pattern on the left, starting in row B and dispensing across the plates and back down row A. Panel 1 shows a significant pattern when row and column averages are plotted. Panel B demonstrates a typical "good" dispense from the same instrument.

nance energy transfer (FRET), homogeneous time resolved fluorescence (HTRF), flash luminescence, glow luminescence, SPA, and fluorescence life time (see Section 4). There have been significant advances in the field of assay detection over the last 5 years [2]. This has primarily been as a result of the development of multimodality imaging technology. As has been mentioned previously in this chapter, the majority of detection methodologies have traditionally been based around photomultiplier tube (PMT) technology. This has the advantage of being very sensitive but realistically means that each PMT can measure one well at a time. This was not an issue when the majority of HTS assays were based in the 96-well plate, but as the well density of plates has increased, the limits of PMT technology have been exceeded. For example, if an SPA assay were being run in a 384-well plate, it takes about 1 min to acquire enough signal from each well via a PMT to get acceptable assay data. Certain scintillation counters have as many as 12 PMT tubes, so they can measure up to 12 wells at once, but this still translates to about 40 min to read a single 384-well plate. For other technologies such as fluorescence, FRET, and HTRF, the throughput limits of PMT-based readers were not reached until assays were executed in 1536-well plates.

Imaging technology uses highly cooled CCD chips and telecentric lenses to measure signal from entire plates at one time [14]. The highly specialized cameras are cooled to temperatures as low as –103°C to reduce electronic noise in the image and increase sensitivity. Combining this with excitation

methods to enable illumination of the entire plate, read times have been reduced to as little as 2 s for most fluorescent technologies and 3 min for SPA. This was achieved independently of plate format.

As the complexity of reader technology has increased, so has the cost. Consequently, the trend is toward multimodality instruments that enable a number of technologies to be read on a single instrument. However, there are some technologies that require dedicated instrumentation due to their specific needs. Examples of these include flash luminescence, which requires integrated liquid handling to dispense reagents at the exact same time as the image of the plate is being taken to capture the burst of light from a very rapid response. There are also ion-channel-related assays that measure the flux of ions within cells, and these assays have dedicated systems with online pipetting such as the FLIPR (Molecular Devices, California).

## Automation

As the number of plates assayed increased, the desire to fully automate the process also increased. By the mid-1990s, the typical HTS lab had a significant automation infrastructure to run screens. This automation consists of a robot arm to move the plate between various liquid handlers, incubators, and detectors usually no different to the ones that the HTS scientist would use at the bench. To tie all this together is a piece of software called a scheduler that controls both the robot arm and the individual devices attached to the system. Automating the HTS process does offer some clear advantages in efficiency and productivity [42, 52, 53]. First, the scheduler enables each plate to be treated identically to the one that went before it so that incubation times are identical from plate to plate. By contrast, manual screening tends to batch process plates, which means that plate 1 often differs in incubation times from the last plate in a batch. Second, the robot is able to run 24 h a day and 7 days a week if desired. Finally, the lab scientist is freed up to process data, design other assays, and perform auxiliary tasks while the robot is running the assay.

While there have been few revolutionary advances in HTS robotics over the last 15 years, they are significantly more robust and mechanically precise than earlier robotic systems. The major advances have typically come from the new liquid handling or detection instrumentation that is integrated on these automated systems. Advances in these peripheral systems have greatly increased the rate at which plates are processed through the individual steps. Now the rate limiting step is the speed of the robot arm. One solution is that robotic developers have linked robot arms together to increase efficiency. An example of this is the Dimension 4 system (CRS Thermo, Canada), which has a belt to move plates up and down to the system and then separate robot arms for each peripheral instrument integrated onto the automated system.

***Scheduling Software*** Scheduling software can be described as dynamic, static, or perpetual. The definition of each varies but is generally considered

as follows. In the case of a static scheduler, the software predetermines every move that the robot is to make for each step of the assay and the number of plates required. This is based on preprogrammed timing for each individual move of the robotic arm. The benefit is that the robotic operator knows exactly what the robot will do, when each plate will be processed, and what the overall timing or variation for the process will be. The downside is that it is not possible to change once the schedule has started, it is inflexible to pauses or errors that may occur during a screening process on the automated system and no conditional steps can be programmed. In the case of the dynamic scheduler, the software calculates each move in real time and no moves are predetermined. The advantage is that priorities can be assigned to certain tasks, conditional steps can be programmed, for example, error recovery is more flexible. The downside is that it is more complex to program with a strong reliance on a well-determined cycle time for each plate to prevent the occurrence of deadlocks where the robot will grind to a halt because it physically cannot complete a move because equipment is occupied. Lastly, the perpetual scheduler calculates each step in real time but also adapts and generates each move as the process continues. Examples of such applications are crystal growth and cell culture where incubation times are not known and the next step is not initiated until a specific criteria has been reached indicating the robot should perform a specific function.

***Reagent Management*** Reagent stability during continuous HTS is a critical factor in generating accurate and precise data over time. It is critical to empirically determine the stability of all the reagents to find optimal conditions, typically overnight or for about 16 h. The standard approach to extend the life of a reagent on a system is to cool it to 4°C. This is easily achievable by the use of vessels chilled with water jackets, peltier devices, or refrigerators.

There are a number of additional factors that have to be considered in addition to temperature when automating the dispensation of biological reagents. These include nonspecific binding to tubing or the vessel, temperature variations over time, stirring, and shear force to name a few. The binding of proteins to the surface of vessels, tubing, and tips is a significant issue, but these can be minimized with the use of detergents, carrier proteins, and the like. From an automation perspective, if the dispenser carries reagents through tubing and that reagent sits in the tubing for the period of a cycle time or 5 min. This creates a number of issues: (1) reagent binding to the tubing, (2) temperature variation between reagent that sat in the tube and reagent that is drawn from the storage vessel, and (3) the potential for reagent settling in the tubing. These issues are usually best solved by using empty and prime cycles to empty the tubing back to the storage vessel between dispenses.

## Data Analysis

Given the number of compounds screened in today's HTS labs, it is obvious that a huge amount of data is generated. The days of manually processing data

were left behind a number of years ago. HTS labs now have fully integrated data systems that automatically extract data from robotic systems, process it, apply quality control (QC) filters, upload to databases, and report the analyzed data. Approaches to this vary significantly between companies, some preferring to buy commercial software and others preferring to write their own custom software; however, the principles remain relatively similar.

***Data Capture and QC*** Once a plate is read on either a robotic instrument or a stand-alone detector, the data should be immediately fed to a database where it can be analyzed. This prevents the user from having to manually move around vast numbers of data files with the potential for introducing error. Analysis of the data may simply involve normalizing the data to generate a percent inhibition relative to controls or it may involve a more complex calculation such as a rate determination, relative to previous reads of the same plate.

Once the plate has been analyzed, it must then pass predetermined QC parameters. Typically, this involves the statistic $Z'$ (Fig. 10, Table 5) [54].

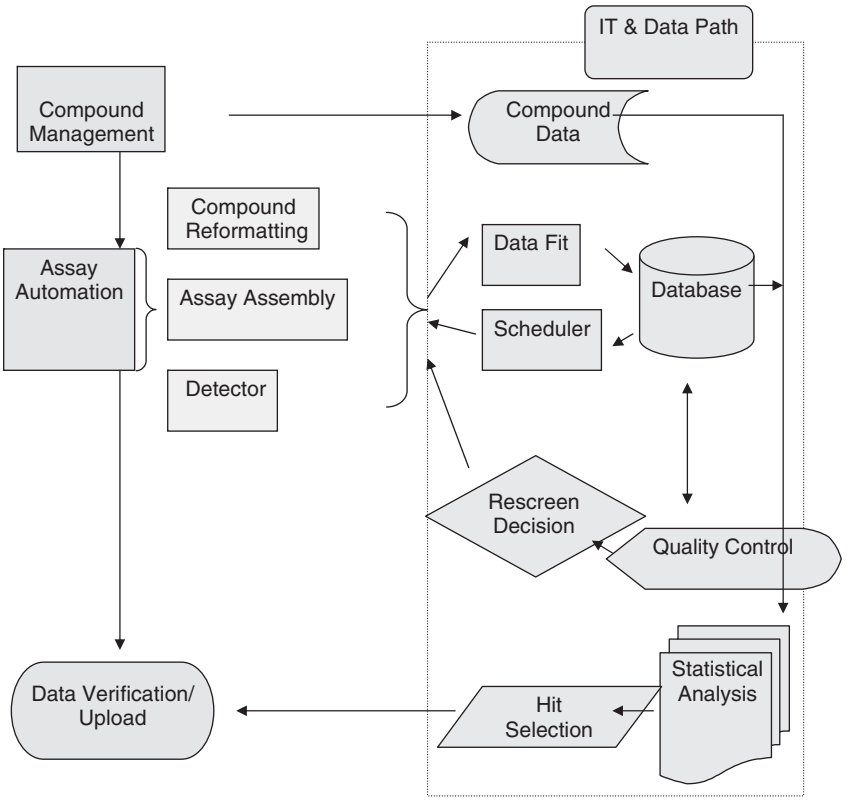


**Figure 10** Informatics architecture that support a high-throughput screening operation. [Adapted from *DDT*, Vol, 6, No. 12 (2001).]

**TABLE 5 Statistical Parameters**

| Statistic | Formula |
|---|---|
| The *mean* is a measure of the central value of a set of data. | $\bar{x} = \frac{1}{b}\sum_{i=1xi}^{b}$ |
| The *standard deviation* measures the variation present in the data. | $s = \sqrt{\frac{1}{b-1}\sum_{i=1}^{b}(x_i - \bar{x})^2}$ |
| The *CV* expresses the standard deviation of the data as a percentage of the mean. | $\mathrm{CV} = \frac{s}{x} \times 100$ |
| The *signal window* measures the distance between the distributions of the totals and the blanks. | $\frac{(\bar{x}_3 - 3 * s_3) - (\bar{x}_1 + 3 * s_1)}{\sqrt{\frac{s_1^2 + s_3^2}{2}}}$ |
| The *Z* statistic measures the signal window as a fraction of the distance between the means of the distribution | $1 - \frac{3 * s_3 + 3 * s_1}{\bar{x}_3 - \bar{x}_1}.$ |
| The *signal-to-noise* (*S/N*) ratio can be considered the "signal strength" of the screen divided by the average variability of the screen. | $\mathrm{S/N} = \frac{(\bar{x}_3 - \bar{x}_1)}{\sqrt{s_3^2 + s_1^2}}$ |
| The *signal-to-background* ratio (S/B) | $\mathrm{S/B} = \frac{\bar{x}_3}{x_1}$ |

Software will allow the user to determine appropriate QC parameters, such as $Z' > 0.5$, and then determine a process path for that plate depending on the result. The data automatically progresses to corporate databases if the QC criteria are met. Data that fails to meet these criteria are flagged, and, if the failure continues to repeat for the next plates, automated screening is halted.

***$IC_{50}$ analysis*** In today's HTS environment, the number of $IC_{50}$ curves generated is in the thousands. This can take days to process manually, so many companies have developed software that automatically processes and analyzes $IC_{50}$ data. In our experience, with software written in-house at Bristol Myers Squibb (BMS), we have reduced a day's data analysis time to about 15 min of work. The impact of this has been significant since it has enabled many more $IC_{50}$ determinations to be performed.

***Hit Selection*** The output of an HTS is a set of compounds that show a broad range of activity. A key step when performing an HTS is determining which compounds are active. These compounds are commonly referred to as a hits. For example, when screening an assay to look for compounds that inhibit a particular enzyme, a wide range of inhibition values will be obtained. This is highlighted in Figure 11. In order to identify which compounds are active, a cut-off threshold must be selected. There are a number of critical factors that need to be considered when selecting a threshold.

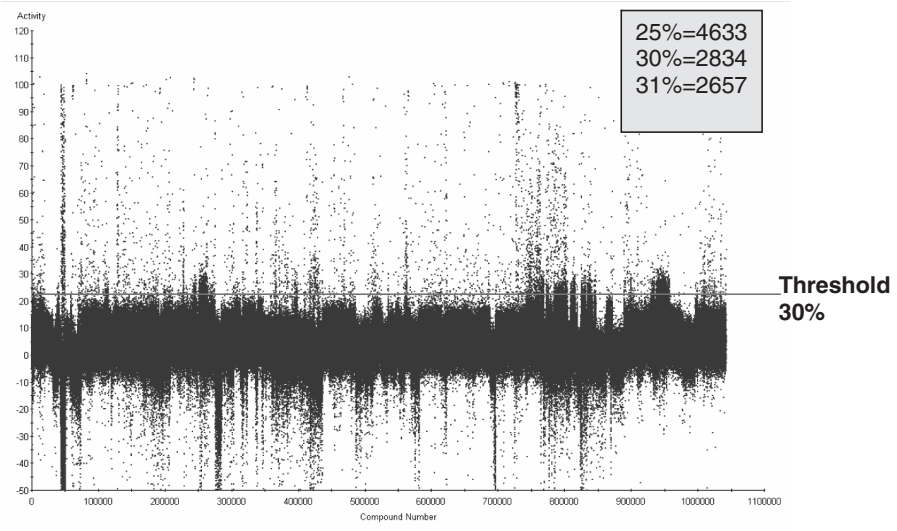


(*a*)

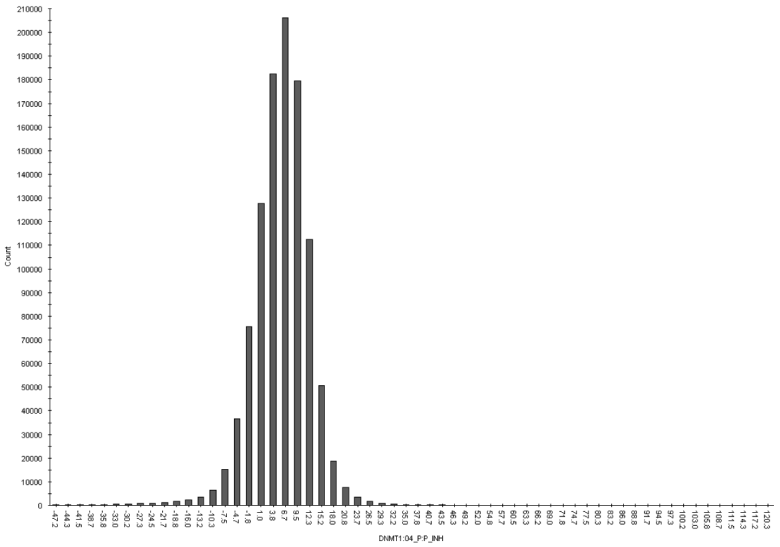


(*b*)

**Figure 11** (*a*) Scatter plot and (*b*) histogram showing activity for over 1 million compounds in a single assay. The graph in (*a*) shows the activity of over 1 million compounds tested in a methyl transferase biochemical assay. Each compound was tested in single point at 10 μM. Illustrated on the graph is the threshold used to select hits. Any compound falling above this threshold was called a primary hit. The inset box gives the number of compounds that fall above a range of threshold values. (*b*) Histogram illustrating the same data.

First, the quality of the data must meet predetermined criteria as described previously. Second, the range of activity is usually a distribution, with the median close to 0 percent inhibition (Fig. 11). The spread of this distribution helps to determine where the threshold should be set, which is typically outside of three standard deviations from the mean. Thus, an active compound will have a discrete activity that separates it from the activity of the vast majority of compounds tested. In the example given in Figure 11, the majority of compounds have an activity of less than 30 percent inhibition, and, therefore, this was designated the cut-off threshold. Finally, from a pragmatic perspective, the number of hits generated needs to be within the range that the compound management department can accommodate in terms of picking compounds for further evaluation.

## 4 ADVANCES IN HIGH-THROUGHPUT BIOASSAY TECHNOLOGY

The field of high-throughput screening has evolved in large part due to the need to reduce the cost (money and people) of screening large sets of compounds (1 million or larger) against molecular targets of therapeutic interest. This movement away from biological assays that are performed in test tubes to assays that are performed in multiwell dishes drives the necessity to examine assay biotechnologies that are robust, homogenous, and amenable to integration on automated systems. The true advances came with technologies that moved away from complex multiple-step assays with tube transfers to homogenous assays where steps could now be eliminated but the target biology remains intact. This movement to higher density homogeneous formats has occurred concurrently with advances in detector technology, liquid handling, robotic integration tools, and informatics infrastructure.

The scale and complexity of the screening approach taken can largely depend upon the resources and technological infrastructure available to the organization. In companies with limited budgets, multiple screening formats and approaches may be constrained. However, in those companies with large R&D budgets multiple approaches to screening a well-validated target can be implemented. A bioassay for a HTS campaign is generally decided upon by a team of people including disease biologists, medicinal chemists, and HTS professionals. This team defines the goals for the screen. These goals include understanding the pharmacological outcome (i.e., agonist or antagonist) to be achieved, knowledge surrounding the target type and how to assay for its activity in a biologically relevant system, and finally bioassay connectivity to an appropriate preclinical model of disease. The HTS bioassay chosen must be robust and reproducible from day to day and demonstrate low intra- and interplate variability. There is also the choice of whether to approach the target from a biochemical or cell-based assay. For a complete discussion on the debate between biochemical or cell-based assay approaches to drug discovery,

see references 55 to 58. The method selected to meet the screening objectives will most likely be a method that is homogeneous and does not require any complex filtration or multiple removal steps.

## Overview of HTS Assay Formats

In general the choice of high-throughput assay format is guided by the goals of the screen and also practical availability of reagents. Some of the pragmatic issues that enter into how to screen a target relate to (1) amount of protein available or cellular supply, (2) substrate requirements, (3) pharmacological relevance of the cellular system, (4) capacity to follow-up the target, and (5) cost to screen a given compound collection. Pharmaceutical targets can be screened using either cell-based or biochemical assays. Cell-based assays include assays that capture binding events, measurement of changes in cellular metabolism, and physiological changes in intracellular second messengers, regulation of intracellular enzymes, and modulation of intracellular signal components. While cell-based assays provide context for pharmacological targets in their native environment, the outcome of these assays can be less straightforward then biochemical assays. Some examples of cell-based assays include the following: calcium flux, voltage-sensitive dyes, reporter assays, whole-cell binding, cAMP detection, physiological changes, protein translocation, and enzymatic activation. Table 6 summarizes cell-based assays and detection methods.

Biochemical assays are generally preferred for intracellular targets including cytoplasmic enzymes. A well-optimized biochemical assay will, in general, demonstrate less data scatter than a cell-based approach. In addition, the presence of a single target in the well makes the follow-up process less complex. The compounds identified from a biochemical screen do not have to be cell permeable nor do they need to demonstrate significant potency. Therefore, they may actually be more chemically diverse than those identified from a cellular screen. Biochemical assays can be separated into two categories: separation based (product is measured after isolation from starting material) and homogeneous (no separation required). Enzymatic assays can be run as kinetic or endpoint assays, and basic mechanistic considerations will be discussed below. Examples of biochemical assays include kinase, protease, metabolic enzymes, and enzymes that modify nucleotides. Table 7 summarizes biochemical approaches to HT bioassays for these targets along with detection methods.

## Reagent Production

Over the last several years, the demand for cell-based assays has increased dramatically. This has been driven by the requirement to support hit identification, lead evaluation in preclinical model systems, and assays to identify compound liabilities. Multiple approaches can be taken to develop cell-based reagents for drug discovery and include transiently transfected cell populations and stable cell line development [59]. Robust transient transfec-

**TABLE 6 Summary of Cell-Based Assays and Detection Methods for HTS**[a]

| Assay Type | Application | Detection Method |
|---|---|---|
| Proliferation assays | Identify cytotoxic compounds | Alamar blue (FLINT) MTT/MTS (Absorption) TUNEL (TRF) |
| Reporter assays | Signal transduction pathways and direct NHR assays | Luciferase (luminescence) β-lactamase (HTRF) |
| Second-messenger assays | cAMP, IP3, MAP kinase, etc. | TRF, FP, α screen, enzyme complementation, etc. |
| Ion flux assays | Intracellular changes in calcium or other ions<br>Measurements of changes in membrane | Fluo-4 (FLINT), Aequorin (flash luminescence), membrane potential dyes (FLINT) |
| High content screening assays | Intracellular events, physiological or morphologic changes, protein translocation or posttranslational modification, multiparametric cytotoxicity, RNA/DNA content, receptor | Automated fluorescence microscopy (varying quality depending on optical configuration, excitation and emission detectors) |

[a] This is not a comprehensive list.

**TABLE 7 Summary of Biochemical Assays and Detection Methods for HTS**[a]

| Assay Type | Application |
|---|---|
| Time-resolved fluorescence | Kinases, second messengers, proteases, helicases, protein–protein interactions, etc. |
| Fluorescence polarization | Kinases, phosphatases, receptor–ligand binding, protein–protein interactions, etc. |
| Enzyme complementation | cAMP, IP3, kinases, other enzymes, protein–protein interactions, etc. |
| Luminescence | Enzymes that utilize ATP, etc. |
| Scintillation proximity assays | Receptor–ligand binding, enymatic assays, protein–protein interactions |

[a] This is not a comprehensive list.

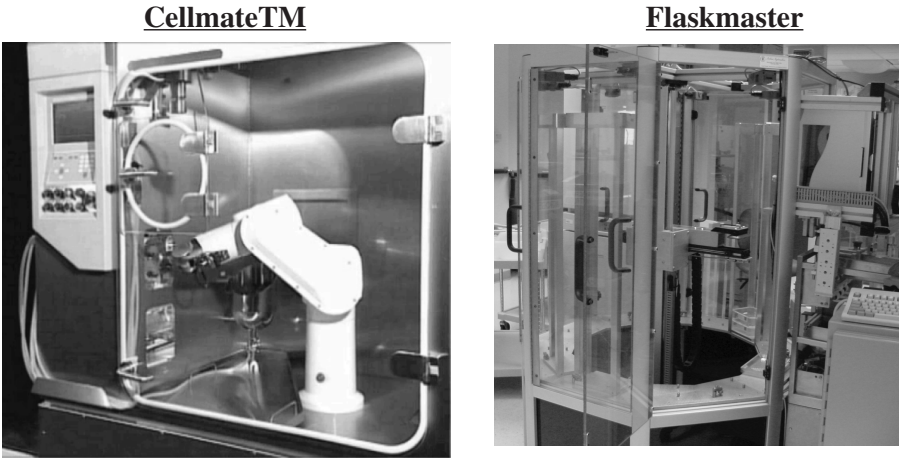


**Figure 12** Automated infrastructure for bulk production of cellular reagents for HTS. The Cellmate (The Technology Partnership, Cambridge, UK). The Cellmate can reform bulk production of up to 500+ flasks, 1–2 cell lines/day and cell harvest by scraping. Unattended operation of the Cellmate is made possible with the addition of the Flaskmaster (BMS) to feed flasks into and out of the instrument.

tion technologies including lipid-based, viral, and electroporation methods have been developed and are used to support in vitro screening assays. Automated solutions have been applied to scale-up these reagents (Cellmate/Flaskmaster; Fig. 12) [60]. During screening, quality control associated with these transfection methods can be monitored through the use of the fluorescent proteins with spectral properties that do not interfere with the corresponding assay formats. Stable cell line development can occur in parallel with the generation of these transient transfected cell populations. Flow cytometry has become an integral component in the process of stable cell line development and is used to monitor, sort, and assess the pharmacology of cell-based targets relative to their corresponding expression levels (epitope tagging, IRES GFP constructs, fluorescent physiological indicators, etc.). Cell line selection for screening requires the ability to process multiple cell clones in parallel using a variety of assay formats in parallel.

The logistics of managing the "just-in-time" need for mammalian cells with active screening efforts requires careful planning and coordination. For example, at BMS, requests for cell line development, maintenance, transfer, mycoplasma testing, FACS analysis, and cryogenic preservation can be made using web-based requesting tools similar to the tools used for compound requesting. A cell-tracking database allows screening groups to track resources, quality control, and document issues that develop pertaining to particular cell lines. Finally, integration of the SelecT fully automated cell culture system allows for the maintenance, quality control, and supply of cells for screening purposes. This system is equipped with scheduling software that

**Figure 13** The SelecT (The Technology Partnership, Cambridge, U.K.). A fully automated cell culture system performs routine maintenance of up to 50 cell lines per day. It manages growth, incubation, and routine passaging of multiple cell lines with unattended operation. The SelecT performs cell counting, viability measurements, and cell plating into microtiter plates.

allows multiple cell lines to be processed as necessary for screening (see Fig. 13). In addition, the system has added consistency to cell-based assays, reduced cell waste, and allowed staff to have continuous access to cells. Many groups responsible for the production of reagents have been moving to bulk production and cryogenic preservation of cell-based reagents for screening; thus, decoupling the just-in-time need for living cells for HTS assays. In this way, one quality-controlled batch of cells can be used throughout the screening campaign. These approaches improve data quality and consistency in HTS.

Biochemical reagents including purified proteins, membrane vesicles, and the like, can be produced in a variety of expression systems including *Escherichia coli*, pichia, insect cells, and mammalian cells. The advent of recombination-based cloning systems allows for the rapid generation of multiple expression constructs in parallel. The capacity for purification of protein from multiple expression systems and parallel purification schemes is very labor intensive and requires quality control analyses to ensure that proteins are purified to near homogeneity, have not formed aggregates. The scale up process is not linear with respect to protein production when compared to the original

growth conditions in a test tube, therefore, both growth conditions and purification processes need to be optimized further for bulk reagent production.

## Designing Bioassays for HTS

When approaching an HTS campaign to identify small-molecule modulators of targets for drug discovery, one must consider the pharmacological outcome as well as the quality of the data produced from the assay. The importance of data quality relates to statistical validation in HTS and this was previously discussed in Section 3. Briefly, a $Z'$ of >0.5 [54] indicates that there is acceptable statistical variation of the data surrounding the maximal signal and the background or minimal signal in the assay, therefore, the assay has the potential to reliably identify active compounds from inactive compounds. The balance is choosing an assay that exhibits sufficient signal, is mechanistically sound, can be run rapidly, and is cost effective.

***Mechanistic Considerations for Biochemical Assays*** Important considerations for the design of enzymatic assays for HTS include the rate at which an enzyme catalyzes conversion of substrate to product, the enzyme concentration, and substrate concentration. For a complete review of mechanistic enzymology, see Ref. 61. Kinetic assays determine the slope/rate of the enzymatic reaction, require multiple measurements over time, exhibit excellent correlation over time for known actives, and are less prone to compound interference. If one performs an enzymatic reaction as an endpoint assay, inhibitors may act to slow the enzyme rate but will be missed in an endpoint HTS because the reaction product is measured at a time point after the enzymatic reaction has run to completion. Another important experimental variable for enzymatic HTS assay design is the enzyme concentration that has a linear effect on reaction velocity. If one chooses an enzyme concentration that is too high, it may minimize the apparent inhibitor effect ($IC_{50} = 1/2$ [Enz]). Competitive enzyme inhibitors compete with substrate for binding to the free enzyme. In this case, assays run at high substrate concentration, inhibit the formation of the enzyme/inhibitor complex, and are less likely to identify competitive inhibitors. Therefore, it is recommended that screening assays be run at substrate concentrations below the $K_m$. A second class of inhibitor binds exclusively to the enzyme/substrate complex, known as uncompetitive inhibitors, and these are more likely to be detected at high substrate concentrations above $K_m$. A third type is a mixed inhibitor, and these bind both the free enzyme and the enzyme/substrate complex. These are less sensitive to substrate concentration. A good understanding of how a particular inhibitor modality may act under the physiological conditions that are appropriate for the target pharmacology provides a significant advantage in designing the appropriate biochemical assay for the target. When the modality of inhibition is unknown, it is always best practice to run the assay where $K_m$ equals substrate concentration. The use of natural substrates is recommended for HTS assays as this more closely approximates the in vivo situation. The use of full-

length enzymes as opposed to activated catalytic domains is strongly recommended as there may be differences in catalysis by the full-length enzyme in vivo. Order of reagent addition may also be an important consideration during an HTS assay. This will depend on the stability of reagents and the mechanism of the reaction. Addition of a substrate to an enzyme may help stabilize the enzyme or stabilizing agents can be added (BSA, detergents, glycerol, etc.). Running an enzymatic assay under equilibrium conditions is important, and so inhibitors that are "slow binders" may require a longer preincubation time. It is generally recommended that compounds be incubated for a period of between 5 and 15 min.

***Ligand–Receptor Binding Interactions*** Although radiometric assays are beginning to decrease with the growing use of fluorescence assay technologies, radiometric assays still comprise between 20 and 50 percent of all screens performed [62]. In the future, label-free methods to measure biophysical interactions may become standard, and these methods include optical and dielectric methods and acoustic biosensors. Indeed innovative cellular assay technologies have been developed based on radio-frequency spectrometry and bioimpedance measurements [63]. Receptor-mediated signal transduction events can be linked to bioimpedance changes by performing dielectric spectroscopy of cells across a spectrum of frequencies (1 KHz to 110 MHz) [63]. The advantage of this technology is that a single assay technique can be used to monitor $G_i$, $G_s$, and $G_q$ without the need of fluorescent labels, tagged proteins, or promiscuous G proteins.

The use of radiolabeled ligands in filtration assays has historically been used to measure high-affinity binding interactions for cell surface receptors and has been replaced by SPA [64–66]. SPA is the current binding assay format of choice for the GPCR (7 transmembrane receptors) HTS assay. The need for separation has been eliminated using a homogeneous technique that relies on the excitation of a scintillant incorporated into microbeads or plates upon binding of the radiolabeled ligand to a receptor immobilized on the surface of a lectin-coated surface of the solid support. Radioactive ligand in close proximity to the receptor results in the emission of light. Free radioligand in solution is too distant from the scintillant, and the β particle energy is dissipated into the aqueous environment. Major advances in this technology over the past 5 years include the development of beads that emit light at 615 nm in contrast to the traditional SPA bead, which emits light around 420 nm. This longer wavelength is well suited for detection by the CCD cameras in modern imagers and results ultimately in fewer problems with color quench by test compounds. These systems provide the added advantage of sensitivity.

When designing an SPA HTS assay, there are many assay conditions that need to be optimized. A multiparameter experiment is performed to determine an acceptable signal window (see Section 3). In such a matrixed experiment, the buffer conditions, amount of receptor, bead concentration, and

ligand concentrations are co-varied to guide the designer to choose conditions for the HTS assay such that it is statistically robust (see discussion of $Z'$ in Section 3). The $K_d$ is then determined following optimization of bead settling/equilibration time, and specific assay conditions have been determined. The $K_d$ is defined as the concentration of ligand required at which half the receptor molecules are bound at equilibrium. This variable can then be approximated by titrating a fixed receptor population with increasing concentrations of hot ligand and measuring the fraction of occupied receptor. The $K_d$ can only be determined accurately under equilibrium conditions. Generally, if an HTS assay is run at equilibrium, there will be agreement between the HTS binding assay and the functional cell-based assays. An example of correlation of the $K_d$ of a ligand binding to a receptor using standard filtration methods versus the apparent $K_d$ determined empirically in a SPA is shown in Figure 14. The reversibility of radiolabeled ligand binding and discrimination of specific binding from nonspecific binding or background signal can then be determined by mixing a fixed concentration of receptor with primary ligand to establish the receptor : ligand concentration. The $K_a$ for the unlabeled compound should equal the $K_d$ unless the label has had an effect on the binding, and this would be demonstrated in a typical competitive displacement curve. The plateau at high competitive ligand concentrations generally represents the increase in nonspecific binding signal in the assay. One may choose to further evaluate a nonspecific signal by using other competitive ligands of diverse mechanism and structure. But, generally, in screening one is looking for compounds that displace the radioligand. The signal in a radioligand SPA is determined in part by the concentration of ligand used in the assay. Assays run at up to 10 times the ligand concentration tend to demonstrate maximal signal intensity. The problems associated with HTS assays that utilize such high ligand concentrations include high nonspecific background signal and the

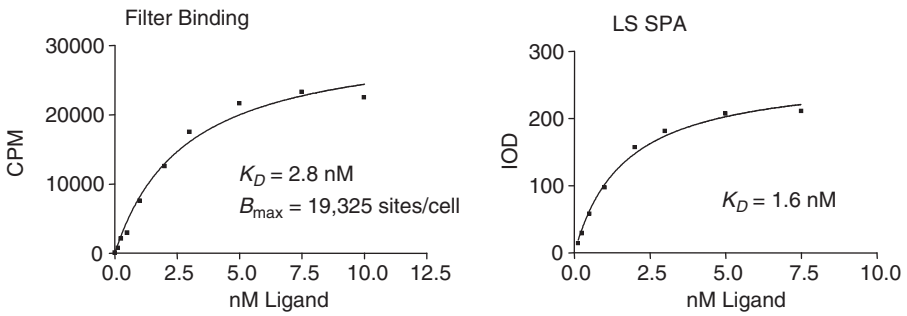


**Figure 14** Shows ligand saturation binding analysis of the same GPCR using either standard filter binding or Leadseeker SPA. In this analysis the $K_D$ values are within a twofold of one another and in agreement with internal historical data. This also allows the HT screener to report apparent $K_i$ values for competitive compounds identified during HTS.

limited ability to detect compounds with lower affinity for the receptor. Obviously, the goal would be to run the assay at or close to the $K_d$ such that the assay is sensitive. Since the goal is to identify compounds that exert their activity under normal physiological conditions, it is important to compare $K_d$ values of ligands in the HTS assay compared to a more physiological assay. For a review of receptor–ligand interaction see Kanakin et al. [67]. Additional mechanistic considerations need to be made for functional assays for GPCRs. A complete discussion of orthosteric versus allosteric modulation of GPCRs as well as an inverse, partial, and full agonism and mechanisms of reversible or insurmountable antagonism are beyond the scope of this chapter [68].

## Fluorescent Methods in HTS

Many of the methods used to identify small-molecule modulators of drug targets involve fluorescence. Being that fluorescence-based detection bioassay technologies tend to be homogeneous and sensitive, they are very common choices for assaying a variety of target types. Fluorescent methods include fluorescence intensity (FLINT), fluorescence polarization (FP), fluorescence resonance energy transfer (FRET), fluorescent proteins (GFP and RFP), time-resolved fluorescence (TRF), subcellular imaging (SCI), fluorescence lifetime (FL), and fluorescent methods associated with detecting single molecules like fluorescence correlation spectroscopy (FCS) and one- and two-dimensional fluorescence intensity distribution analysis (FIDA) [69, 70]. FP is the quantitation of single or complex biomolecule rotation and the polarization of the emitted fluorescence. Generally, the fluorescent indicator is linked to an antigen or a ligand, which then is bound by an antibody or receptor thereby causing a large mass change. Polarization values represent the rotation rates or the related mathematical parameter, anisotropy. This widely used method is volume-independent and can be used in high-density, low-volume HTS formats [71].

Moving away from radiation-dependent energy transfer along with improvements in dyes and detection technologies has reinvigorated an interest in FRET-based assays where fluorophore pairs with overlapping emission and excitation spectrum exchange energy as donor and acceptor fluors. This assay format is suitable for interaction of proteins and cell-based assay formats; for example, antibody—antigen interactions, protease substrate that separates the donor and acceptor fluors. The FRET principle can be applied with long lifetime lanthanide-based fluorescent dyes ($Eu^{3+}$ cryptate) as a homogeneous time-resolved fluorescence, or HTRF, that enable discrimination between fluorophore and signals form interfering assay substituents like compounds [72]. Proprietary HTRF reagents have been designed to assay receptor activation, proteases, kinases, and protein–protein interactions under commercial names such as Trace from CisBio and LANCE from Perkin-Elmer [69].

Fluorescence lifetime assays use time-domain measurements whereby fluorophores are excited by short, high-frequency laser pulses, and the lifetime

of a fluorophore is determined. Detectors of FL have been developed though technically challenging (Tecan, Molecular Devices, and Evotec). The full benefit of this technology remains to be seen as the magnitude of changes in FL are difficult to predict and are empirically determined.

Fluorescent methods for HTS have recently achieved single-molecule detection that are very sensitive, information rich, and assayed in reduced reaction volumes using confocal optics where de-convoluted one- and two-dimensional photon count distributions provide data at the molecular level. Readout parameters include polarization, inhibition, molecular brightness, and concentration of multiple assay components [69]. For HTS the most powerful application of FCS is FIDA, enabling direct measure of ligand binding states applied to receptor–ligand interactions [73].

***Fluorescent Methods in Cell-Based Bioassays*** Cell-based functional assays using fluorescent probes initially were developed using organic small molecules. These are highly fluorescent and have been used to monitor changes in real-time gene expression in reporter assays, measuring local changes in ion concentrations, membrane potential, pH, and other physiological changes [55, 74, 75]. Caged fluorescent calcium-sensitive dyes such as Fluo-4 are loaded as membrane-permeable esters that are then cleaved by intracellular esterases and exhibit altered quantum yield upon calcium binding. An example of Fluo-4 measurement of calcium flux in response to a cognate ligand binding to a $G_q$-coupled 7 TM receptor is shown in Figure 15. These changes in intracellular calcium can be measured using a FLIPR (for more detail see the following section). Other cell-based and FRET-based dyes have been developed to detect changes in membrane potential as well as reporter assays that measure changes in real-time gene expression of the reporter gene β-lactamase [76–79].

Attempts to resolve intracellular events that occur using fluorescent-labeled ribonucleic acid (RNA), DNA, or proteins have allowed tracking of movement or changes in intracellular localization triggered by signaling cascades or physiologic changes in organelles. Now proteins can be fused to GFP or RFP to take advantage of their fluorescent properties to examine spatial and temporal resolution using fluorescence microscopy, and this subcellular imaging has become known as high content screening (HCS) [80, 81].

## Cell-Based Assay Design

When designing a cell-based assay multiple considerations need to be assessed at the outset. The first consideration is evaluation of the cell line to be used prior to expressing the target: Can enough cells be produced to support the screening effort, do the cells senesce over time, and so forth? Other important variables for cell line selection at the outset of assay design include the growth characteristics of the given cell type, whether the cells are prone to changes in phenotype or morphology, and will these changes have the potential to impact

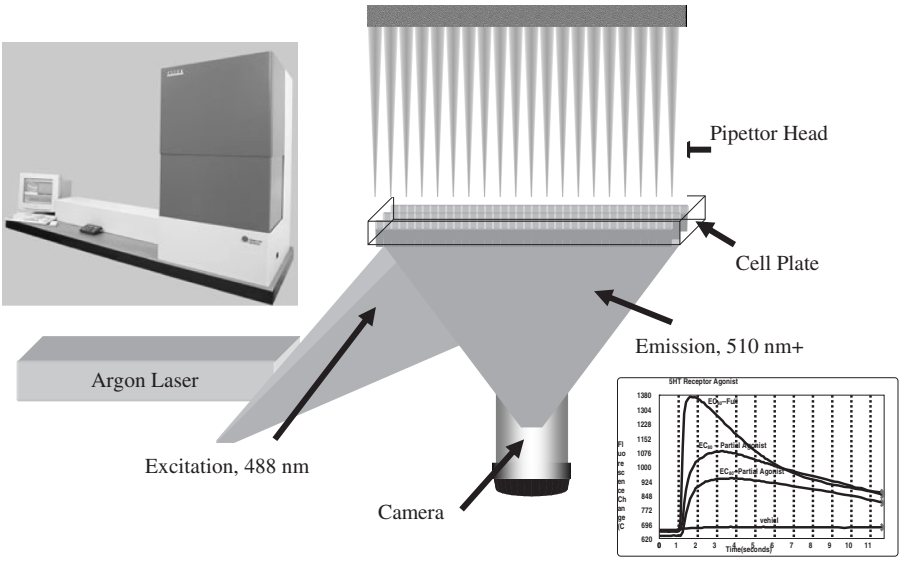


**Figure 15** Depicts a FLIPR (Molecular Devices) fluorometric imaging system that is useful for quantitating kinetic changes in intracellular fluorescence. The inset FLIPR trace shows a range of G protein coupled receptor agonist responses; full and partial efficacy as compared to vehicle control.

the outcome of the assay? Characterization of target expression (either endogenous or exogenous) as well as any critical cofactors that may need to be expressed in the cell line to ensure the assay will perform (i.e., Gα subunits, reporter constructs, enzyme subunits, etc.) over time is also critical. For example, one key question could be: Is expression consistent from passage to passage? Once the cell line has been characterized for routine maintenance and expression, it is then expanded for assay design.

The cell culture optimization within the microtiter plates is an important part of assay design for cell-based screening. Cell number, assay media including serum, and other additives are monitored and tested along with the assay components. As a key example, the critical variables for designing a functional 7TM-R assay will be discussed. There are multiple parameters such as cellular dissociation conditions, cellular dispensing, cell density in the well, and media components in the well that are critical to optimal cell plating for a FLIPR calcium flux assay. In this assay, we measure changes in intracellular calcium upon ligand binding to a $G_q$-coupled 7TM-R. After plating, the next day the calcium indicator, Fluo-4 AM, is added to the cells in a HEPES-buffered saline solution. The composition of the buffer for certain cell-based assays is critical to the overall performance, that is, ionic concentration and composition can be critical for performing calcium flux assays or measuring membrane potential. For a review of membrane permeability and membrane potential see [82]. If the media needs to be removed and aspirated prior to addition of the fluorescent indicator, aspiration tip heights need to be adjusted

on the liquid handling device to ensure that the monolayer of cells is intact. The concentration and choice of fluorescent calcium indicator needs to assessed in the assay. Finally, the agonist and antagonist pharmacology can be assessed in the assay following optimization of liquid handling on the FLIPR including pipet height, addition volumes. For a schematic representation of the FLIPR and inset agonist kinetic analysis see Figure 15.

## Additional Advances in Cellular Assay Formats

Ion channels have long eluded pharmaceutical research because of the difficulties developing functional assays that adequately measure changes in membrane potential in response to ligands or currents. Fluorescent dyes have been developed that can be used to measure changes in membrane potential. When combined with fluorescent plate readers that contain on-line pipettors, rapid changes in fluorescence in response to ion channel openers and blockers can be measured [83, 84]. Although these surrogate assays for ion channels allow for high-throughput hit identification, they do not replace single-cell patch recordings. Many companies have developed automated methods to perform electrophysiological measurements on cells, and these include Robocyte, OpusXpress, IonWorks, and Patch Express [39, 85]. These advances will allow higher throughput ion channel drug discovery and liability measurements to become a reality and enable therapeutics to be developed against this difficult target class.

Other cellular assay formats use reporter-based assays that measure the effects of compounds that modulate signal transduction cascades. Reporters such as β-lactamase, luciferase, and nitroreductase are commonly useful in functional cell-based screening for a variety of target types including receptors, enzymes, protein–protein, and nuclear hormone receptors. Development of click beetle luciferase reporters now enable multiple pathways to be interrogated within a single cell [32].

Enzyme fragment complementation is another flexible cell-based assay format that is based on high-affinity complementation of a small fragment of β-galactosidase (Pro-label) to an inactive deletion mutant of the enzyme (enzyme acceptor) and generates a signal via hydrolysis of a substrate [87]. This assay format has enabled direct measurement of a wide variety of second messengers and signal transducers within cells including cAMP, IP3, MAP kinase, and others. This method can also be useful for measuring protein–protein interactions within cells [87].

With advances in automated fluorescence microscopy came the evolution of high content screening (HCS). HCS has made possible complex measurements in cellular morphology and physiology, changes in intracellular protein translocation and posttranslational modifications, measurements of receptor internalization, and multiparameter cytotoxicity measurements. For a detailed review of HCS applications and instrumentation see reference 38.

The evolution of HTS bioassay technologies has greatly enhanced the quality of data produced from screening large compound collections, amount

of information added to the compounds following screening, and ultimately improved decision making by medicinal chemistry to facilitate the drug discovery process.

## REFERENCES


1. Houston, J. G., Banks, M. N. (2003). High throughput screening for lead discovery. In D. J. Abraham (Ed.), *Burger's Medicinal Chemistry and Drug Discovery*, Vol. 2. Wiley, New York, pp. 37–69.
2. Bojanic, D., Keighley, W. W., Russell, M. J., Wood, T. P. (1997). Factors for the successful integration of assays, equipment, robotics, and software. In J. P. Devlin (Ed.), High Throughput Screening: the Discovery of Bioactive Substances. Marcel Dekker, New York, pp. 493–508.
3. Takatsy, G. (1955). The use of spiral loops in serological and virological methods. *Acta Microbiol. Hung.*, *3*, 191–202.
4. Engvall, E., Perlmann, P. (1971). Enzyme lined immunoassay (ELISA) quantitative assay of immunoglobulin G. *Immunochemistry*, *8*, 871–874.
5. Voller A., Bidwell, D. E. (1985). Enzyme immunoassays. In W. P. Collins (Ed.), *Alternative Immunoassays*. Wiley, New York, pp. 77–86.
6. Wood, T., Keghley, W. (2004). Automation sample supply for high throughput screening. *Eur. Pharm. Rev.*, *9*, 68–73.
7. Jarmaat, M. L., Giaccone, G. (2003). Small molecule epidermal growth factor receptor tyrosine kinase inhibitors. *Oncologist*, *8*, 576–586.
8. Foulkes, J. G., Chow, M., Gorka, C., Frackelton, A. R., Baltimimore, D. (1985). Purification and characterization of a protein tyrosine kinase encoded by the Abelson murine leukemia virus. *J. Biol. Chem.*, *260*, 8070–8077.
9. Filthut, H. (1989). A new detector for radiochromatography and radio-labelled multisample distribution. *J. Planas Chrom. Modern TLC*, *3*, 198–202.
10. Schraag, B., Staal, G. E. J., Adriaansen-Slot, S. S., Salden, M., Rijkesen, G. (1993). Standardization of an enzyme-linked immunosorbent assay for the determination of tyrosine kinase activity. *Anal. Biochem.*, *211*, 233–239.
11. Potter, C. G., Warner, G., Yrjonen, T., Soine, E. (1986). A liquid scintillation counter specifically designed for samples deposited on a flat matrix. *Phys. Med. Biol.*, *31*, 361–369.
12. Warner, G. T., Potter, C. G., Yrjonen, T., Soinl, E. (1985). A new design for a liquid scintillation counter for micro samples using flat bed geometry. *Int. J. Appl. Radiat.* Isot., *36*, 819–821.
13. Cook, N. D. (1996). Scintillation proximity assay a versatile high throughput screening technology. *Drug Discov. Today*, *1*, 287–294.
14. Ramm, P. (1999). Imaging systems in assay screening. *Drug Discov. Today*, *4*, 401–410.
15. Jakob, W., Heder, G., Halle, W. (1981). DNA synthesis in cultures of calf aortic endothelial cells treated with corpus luteum extract: Its inhibition by $^{3}$H-thymidine. *Exp. Pathol.*, *20*, 41–50.

16. Hoy, C. A., Lewis, E. D., Schimke, R. T. (1990). Perturbation of DNA replication and cell cycle progression by commonly used [$^3$H] thymidine labeling protocols. *Mol. Cell. Biol.*, *10*, 1584–1592.
17. Hu, V. W., Black, G. E., Torres-Duarte, A., Abramson, F. P. (2002). $^3$H-thymidine is a defective tool with which to measure rates of DNA synthesis. *FASEB J.*, *16*, 1456–1457.
18. Borenfreund, E., Babich, H., Martin-Alguacil, N. (1988). Comparisons of two in vitro cytotoxicity assays—The natural red (NR) and tetrazolium MTT tests. *Toxicol. in Vitro*, *2*, 1–6.
19. Mosmann, T. (1983). Rapid colorimetric assay for cellular growth and survival: Application to proliferation and cytotoxicity assays. *J. Immunol. Methods*, *65*, 55–63.
20. Maxwell, I. H., Maxwell, F. (1988). Electroporation of mammalian cells with firefly luciferase expression differ markedly among cell types. *DNA*, *7*, 557.
21. Gorman, C., Moffat, L. F., Howard, B. H. (1982). Recombinant genomes which express chloramphenicol acetyltransferase in mammalian cells. *Mol. Cell. Biol.*, *2*, 1044.
22. An, G., Kidaka, K., Siminovitch, K. (1982). Expression of bacterial β-galactosidase in animal cells. *Mol. Cell. Biol.*, *2*, 1628.
23. Bronstein, I., Fortin, J., Stanley, E., Stewart, G. S. A. B., Kricka, L. J. (1994). Chemiluminescent and bioluminescent reporter gene assays. *Anal. Biochem.*, *219*, 169.
24. Pazzagali, M., Devine, J. H., Peterson, D. O., Baldwin, T. O. (1992). Use of bacterial and firefly luciferases as reporter genes in DEAE-dextran-mediated transfection of mammalian cells. *Anal. Biochem.*, *204*, 315.
25. Fisher, M., Harbron, S., Rabin, B. R. (1995). An amplified chemiluminescent assay for the detection of alkaline phosphatase. *Anal. Biochem.*, *227*, 73.
26. Chalfie, M., Tu, Y., Euskirchen, G., Ward, W. W., Prasher, D. C. (1994). Green fluorescent protein as a marker for gene expression. *Science*, *263*, 802.
27. Kolb, A., Roelant, C., Scheirer, W. (1994). Reagents and instruments for luminescence assays in drug discovery. In A. K. Campbell, L. J. Kricka, and P. E. Stanley (Eds.), *Bioluminiescence and Chemiluminescence*. Wiley, Chichester, p. 357.
28. Stanley, P. E. (1989). A review of bioluminescent ATP techniques in rapid microbiology. *J. Biolum. ChemiLumin.*, *4*, 375–380.
29. Yang, Y., Thomason, D. B. (1993). An easily synthesized photolyzeable luciferase substrate for *in vivo* luciferase activity measurement. *BioTechniques*, *15*, 848.
30. Tanahashi, H., Ito, T., Inoye, S., Tsuji, F. I., Sakaki, Y. (1990). Photoprotein aequorin: Use as a reporter enzyme in studying gene expression in mammalian cells. *Gene*, *96*, 249.
31. Zlokarnik, G. (2000). Fusions to beta-lactamase as a reporter for gene expression in live mammalian cells. *Methods Enzymol.*, *326*, 221–244.
32. Wood, K. V., Lam, A. Y., Seliger, H. H., McElroy, W. D. (1989). Complementary DNA coding click beetle luciferases can elicit bioluminescence of different colors. *Science*, *233*, 700.
33. Gabriel, D., Verner, M., Pfrifer, M. J., Daser, B., Tenzillon, L., Bouhelal, R. (2003). HTS technologies for direct cyclic AMP measurement. *Assay Drug Rev.*, *1*(2), 291–302.

34. Berridge, M. J., Downes, C. P., Hanley, M. R. (1982). Lithium amplifies agonist dependent phosphotidylinositol in brain and salivary glands. *Biochem. J.*, *206*, 587–595.

35. Wilson, D. B., Connell, T. M., Bross, T. E., Majerus, P. W., Sterman, W. R. Tyter, A. N. Rubin, L. J., Brown, J. E. (1985). Isolation and characterization of the inositol cyclic phosphate products of polyphosphoinositide cleavage by phospholipase C. Physiological effects in permeabilized platelets and Limulus photoreceptor cells. *J. Biol. Chem.*, *260*, 13496–13501.

36. Gutkind, J. S. (1998). Cell growth control by G-protein-coupled receptors from signal transduction to signal integration. *Oncogene*, *17*, 1331–1342.

37. Beske, O. E., Goldband, S. (2002). HT cell analysis using multiplexed assay technologies. *DDT*, *7*(18, Suppl.), S131–S135.

38. Zamanova, L., Schenk, A., Valler, M. J., Nienhaus, U., Heilker, R. (2003). Confocal optics microscopy for biochemical and cellular HTS. *Drug Discov. Today*, 8(23), 1085–1093.

39. *The PatchXpress Users Newsletter,* Vol. II (March 2004). Axon Instrument, Mol. Devices., CA, USA.

40. Houston, J. G., Banks, M. (1997). The chemical-biological interface: Developments in automated and miniaturized screening technology. *Curr. Opin. Biotechnol.*, *8*, 734–740.

41. Burbaum, J. J. (2000). The evolution of miniaturized well plates. *J. Biomol. Screen.*, *5*, 5–8.

42. Banks, M. (2000). Approaches to high throughput toxicity screening. In C. K. Goldfard and W. Pucell (Eds.), *Automation and Technology for HTS in Drug Development.* Yaler and Francis, New York, pp. 9–29.

43. Wolcke, J., Ullman, D. (2001). Miniaturized HTS technologies. *Drug Discov. Today*, *6*, 637–646.

44. Mere, L., Bennert, T., Coassin, P., England, P., Hamman, B., Rink, T., Zimmerman, S., Negulescu, P. (1999). Miniaturized FRET assays and microfluidics: Key components for ultra-high-throughput screening. *Drug Discov. Today*, *4*, 363–369.

45. Spencer, P. A. The challenges of managing a compound collection. *Eur. Pharmaceut. Rev.*, *19*, 51–75.

46. Bohacek, R., McMartin, C., Guida, W. (1996). The art and practice of structure-based drug design: A molecular modeling perspective. *Med. Results Rev.*, *16*, 3–50.

47. Miller, M. A. (2002). Chemical database techniques in drug discovery. *Nat. Rev. Drug Discov.*, *1*, 220–227.

48. Mestres, J. (2002). Virtual screening: A real screening complement to high throughput screening. *Biochem. Soc. Trans.*, *30*, 797–799.

49. Cheng, X., Hochlowski, J., Tang, H., Hepp, D., Beckner, C., Kantor, S., Smitt, R. (2003). Studies on reposotory compound stability in DMSO under various conditions. *J. Biomol. Screen.*, *8*, 292–304.

50. Comley, J. (2002). Continued miniaturization of assay technologies drives market for nanolitre dispensing. *Drug Discov. World*, *5*(3), 43–53.

51. Comley, J. (2002). Nanolitre dispensing: On the point of delivery. *Drug Discovery World*, *3*(3), 43–46.

52. Oldenburg, K. R. (1999). Automation basics: Robotics vs workstations. *J. Biomol. Screen.*, *4*, 53–54.
53. Wildey, M. J., Homon, C. A., Hutchins, B. (1999). Allegro™: Moving the bar upwards. *J. Biomol. Screen.*, *4*, 115–118.
54. Zhang, J. H., Chung, T. D. Y., Oldenburg, K. R. (1999). A simple statistical parameter for use in evaluation and validation of high throughput screening assays. *J. Biomol. Screen.*, *4*, 67–73.
55. Johnston, P. A. (2002). Cellular platforms for HTS: Three case studies. *Drug Discov. Today*, *7*(6), 353–363.
56. Pagliaro, L., Praestegaard, M. (2001). Transfected cell lines as tools for high throughput screening: A call for standards. *J. Biomol. Screen.*, *6*(3), 133–136.
57. Moore, K., Rees, S. (2001). Cell-based versus isolated target screening: How lucky do you feel? *J. Biomol. Screen.*, *6*(2), 69–74.
58. Gonzalez, J. E., Negulescu, P. A. (1998). Intracellular detection assays for high-throughput screening. *Curr. Opin. Biotechnol.*, *9*, 624–631.
59. Pagliaro, L., Praestegaard, M. (2001). Transfected cell lines as tools for high throughput screening. *Call for Standards*, *6*(3), 133–135.
60. Bernard, C., Conners, D., Barber, L., Jayachandra, S., Bullen, A., Cacace, A. (2004). Adjunct automation to the cellmate cell culture robot., *J.A.L.A.*, *9*(4), 209–217.
61. Copeland, R. A. (2003). Mechanistic considerations in HTS. *Anal. Biochem.*, *320*, 1–12.
62. Hertzberg, R. P., Pope, A. J. (2000). High-throughput screening: new technology for the 21st century. *Curr. Opin. Chem. Biol.*, *4*, 445–451.
63. Ciambrone, G. J., Liu, V. F., Lin, D. C., McGuinness, R. P., Leung, G. K., Pitchford, S. (2004). Cellular dielectric spectroscopy: A powerful new approach to label-free cellular analysis. *J. Biomol. Screen.*, *9*(6), 467–480.
64. Major, J. (1999). What is the future of HTS? *J. Screen.*, *4*, 119.
65. Bosworth, N., Tower, P. (1989). Scintillation proximity assay. Nature, *341*, 167–168.
66. Sittampalam, G., Kahl, S., Janzen, Q. (1997). HTS: Advances in assay technologies. *Curr. Opin. Chem. Biol.*, *1*, 384–391.
67. Kenakin, T. (2004). Principles: receptor theory in pharmacology. *Trends Pharmacol. Sci.*, *25*(4), 186–192.
68. Kenakin, T. (2004). G Protein coupled receptors and allosteric machines. *Receptors and Channels.*, *10*(2), 51–60.
69. Gribbon, P., Sewing, A. (2003). Fluorescent readouts in HTS: No gain without pain? *Drug Discov. Today*, *8*(22), 1035–1043.
70. Eggeling, C., Brand, L., Ullmann, D., Jager, S. (2003). Highly sensitive fluorescence detection technology currently available for HTS. *Drug Discov. Today*, *8*(14).
71. Turconi, S., Shea, K., Ashman, S. (2001). Real experiences of uHTS: A prototypic 1536 well fluorescence anisotropy-based uHTS screen and application of well-level quality control procedures. *J. Biomol. Screen.*, *6*, 275–290.
72. Bazin, H., Trinquet, E., Mathis, G. (2002). Time resolved amplification of cryptate emission: A versatile technology to trace biomolecular interactions. *J. Biotechnol.*, *82*, 233–250.

73. Scheel, A. A., Funsch, B., Busch, M., Gradl, G., Pschoff, J., Lohse, M. J. (2001). Receptor-ligand interactions studied with homogeneous fluorescence-based assays suitable for miniaturized screening. *J. Biomol. Screen.*, *6*, 11–18.
74. Kunapoli, P., Ransom, R., Murphy, K. L., Pettibone, D., Kerby, J., Grimwood, S., Zuck, P., Hodder, P., Lacson, R., Hoffman, I., Inglese, J., Strulovici, B. (2003). Development of an intact cell reporter gene beta-lactamase assay for G protein-coupled receptors for high throughput screening. *Anal. Biochem.*, *314*, 16–29.
75. Chambers, C., Smith, F., Williams, C., Marlos, S., Liu, Z. H., Hayer, P., Ciaramella, G., Keighley, W., Gribbon, P., Sewing, A. (2003). Measuring intracellular calcium fluxes in high throuthput mode. *Comb. Chem. High Throughput Screen.*, *6*, 355–362.
76. Baxter, D. F., Kirk, M., Garcia, A. F., Raimond, A., Holmquist, M. H., Flint, K. K., Bojanic, D., Distefano, P. S., Curtis, R. (2002). A novel membrane potential-sensitive fluorescent dye improves cell-based assays for ion channels. *J. Biomol. Screen.*, *7*, 79–85.
77. Farinas, J., Chow, A. W., Nada, H. G. (2001). A microfluidic device for measuring cellular membrane potential. *Anal. Biochem.*, *295*, 138–142.
78. Gonzalez, J. E., Maher, M. P. (2002). Cellular fluorescent indicators and voltage/ion probe reader (VIPR™): Tools for ion channel and receptor drug discovery. *Receptors Channels*, *8*, 283–295.
79. Zlokarnick, G., Negulescu, P.A., Knapp, T., Mere, L., Burres, N., Feng, L., Witney, M., Roemer, K., Tsien, R.Y. (1998). Quantitation of transcriptional and clonal selection of single living cells with beta lactamase as reporter. *Science*, *279*, 84–88.
80. Chiesa, A., Rapizzi, E., Tosella, V., Pinton, P., de Virglio, M., Fogarty, K. E., Rizzuto, R. (2001). Recombinant aequorin and green fluorescent protein as valuable tools in the study of cell signaling. *Biochem. J.*, *355*, 1–12.
81. Lippincott-Schwartz, J., Patterson, G. H. (2003). Development and use of fluorescent protein markers in living cells. *Science*, *300*, 87–90.
82. Falconer, M., Smith, F., Surah-Narwal, S., Congrave, G., Liu, Z., Hayter, P., Ciarmella, G., Keighley, W., Haddock, P., Waldron, G., Sewing, A. (2002). High-throughput screening for ion channel modulators. *J. Biomol. Screen.*, *7*, 460–465.
83 Cooper, M. A. (2003). Biosensor profiling of molecular interactions in pharmacology. *Curr. Opin. Pharmacol.*, *3*, 557–562.
84. Tang, W., Kang, J., Wu, X., Ranme, D., Shen, H., Li, Z., Dunnington, D., Garyantes, T. (2001). Development and evaluation of high throughput functional assay methods for hERG potassium channel. *J. Biomol. Screen.*, *6*, 325–331.
85. Schroader, K., Neagle, B., Trezise, D. J., Worley, J. (2003). IonWorks HT: A new high throughput electrophysiology measurement platform. *J. Biomol. Screen.*, *8*, 50–64.
86. Scheirer, W. (1997). Reporter gene assay applications. In J. P. Devlin (Ed.), *HTS: The Discovery of Bioactive Substances*. Marcel Dekker, New York, pp. 401–412.
87. Eglen, R. M., Pribilla, I. (2004). High-throughput screening using label-free technologies. *J. Biomol. Screen.*, *9*(6), 465–466.

# METAL-ENHANCED FLUORESCENCE: APPLICATION TO HIGH-THROUGHPUT SCREENING AND DRUG DISCOVERY

Kadir Aslan,[1] Ignacy Gryczynski,[2] Joanna Malicka,[2] Joseph R. Lakowicz,[2] and Chris D. Geddes[1,2]

[1]*Institute of Fluorescence, University of Maryland Biotechnology Institute Baltimore, Maryland*

[2]*Center for Fluorescence Spectroscopy, University of Maryland School of Medicine Baltimore, Maryland*

1 METAL–FLUOROPHORE INTERACTIONS

2 METAL SUBSTRATES FOR METAL-ENHANCED FLUORESCENCE
Silver Island Films for MEF
Silver Colloid Films for MEF
Anisotopic Silver Nanostructures for MEF
Laser-Deposited Silver for MEF
Electrochemically Deposited Silver for MEF
Electroplating of Silver for MEF
Roughened Silver Electrodes for MEF
Silver Fractal-Like Structures on Glass Substrates for MEF

3 METAL-ENHANCED FLUORESCENCE SENSING: APPLICATION TO SENSING IN PLATE WELLS
Enhanced DNA/RNA Detection Using MEF (Hybridization Assays)
Enhanced DNA Labels
Over-labeled Proteins as Ultrabright Probes
Enhanced and Selective Excitation Using Multiphoton Excitation and Metallic Nanoparticles
Enhanced Energy Transfer on Silver Surfaces

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

Ratiometric Sensing on Surfaces
MEF Solution-Based Assays
Directional Emission
4 ROLE OF MEF IN HTS AND DRUG DISCOVERY
5 CLOSING REMARKS
References

## 1 METAL-FLUOROPHORE INTERACTIONS

Fluorescence detection is the basis of most assays in high-throughput screening and drug discovery today. Although fluorescence is a highly sensitive technique, there is always a need for reduced detection limits and/or small copy-number detection. Fluorophore detectability is usually limited by autofluorescence of the samples and/or the photostability of the fluorophores. In an effort to overcome these problems, we have been investigating the use of metallic nanostructures to favorably modify the spectral properties of fluorophores [1–4]. We refer to the use of fluorophore–metal interactions as radiative decay engineering (RDE) or metal-enhanced fluorescence (MEF).

In the present review, we describe the effects of different silver nanostructures, which were prepared by various methods in our laboratories, on the emission intensity and photostability of fluorophores that are frequently used in many biological assays. The silver nanostructures consist of subwavelength-size nanoparticles of silver deposited on inert substrates. We show that proximity to silver nanostructures results in a preferential increase in intensity of low-quantum-yield fluorophores and that the lifetimes decrease as the intensities increase. We subsequently discuss the use of MEF for its applications in high-throughput screening and drug discovery.

It is well known that the presence of a nearby metal surface (m) can favorably increase the radiative decay rate by addition of a new rate $\Gamma_m$ (Fig. 1*a*) [1, 2]. The metallic surface can cause Förster-like quenching with a rate ($k_m$), can concentrate the incident field ($E_m$), and can, importantly for sensing applications, increase the radiative decay rate ($\Gamma_m$). These new phenomena typically occur at different distances from the metal surface as depicted in Figure 2. As the value of $\Gamma_m$ increases, that is, the spontaneous rate at which a fluorophore emits photons, the quantum yield increases while the lifetime decreases. In order to illustrate this point, we calculated the lifetime and quantum yield for fluorophores with an assumed natural lifetime $\tau_N = 10\,\text{ns}$, $\Gamma = 10^8\,\text{s}^{-1}$ and various values for the nonradiative decay rates and quantum yields. The values of $k_{nr}$ varied from 0 to $9.9 \times 10^7\,\text{s}^{-1}$, resulting in quantum yields from 1.0 to 0.01. Suppose now the metal results in increasing values of $\Gamma_m$. Since $\Gamma_m$ is a rate process returning the fluorophore to the ground state, the lifetime decreases as $\Gamma_m$ becomes comparable and larger than $\Gamma$ (Fig. 3*a*).

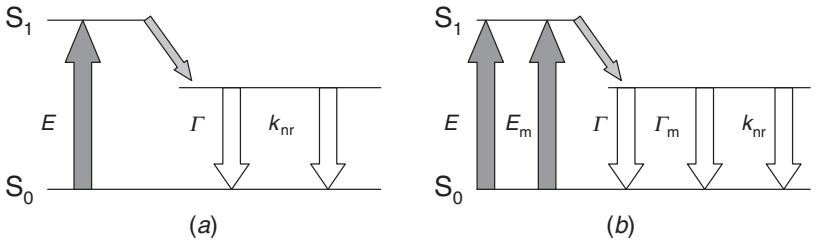


**Figure 1** Classical Jablonski diagram for (*a*) the free space condition and (*b*) the modified form in the presence of metallic particles, islands, colloids, or silver nanostructures. $E$ excitation, $E_m$, metal-enhanced excitation rate; $\Gamma_m$, radiative rate in the presence of metal. For our studies, we do not consider the effects of metals on $k_{nr}$.

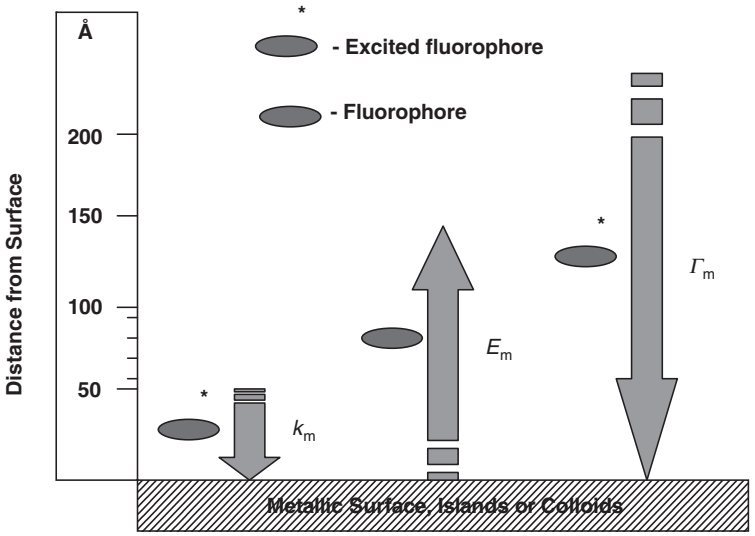


**Figure 2** Predicted distance dependencies for a metallic surface on the transitions of a fluorophore. $\Gamma_m$ modifications typically occur at distances larger than 50 Å from the metal surface.

In contrast, as $\Gamma_m$ / $\Gamma$ increases, $Q_m$ increases but no change is observed for fluorophores where $Q_0 = 1$ (Fig. 3*b*).

As a result of these calculations, we have predicted that the metallic surfaces can create new unique fluorophores with increased quantum yields and shorter lifetimes. Figure 4 illustrates that the presence of a metal surface within close proximity of a fluorophore with low quantum yield ($Q_0 = 0.01$) increases its quantum yield ~10-fold, resulting in brighter emission, while reducing its lifetime 10-fold, resulting in an enhanced photostability of the

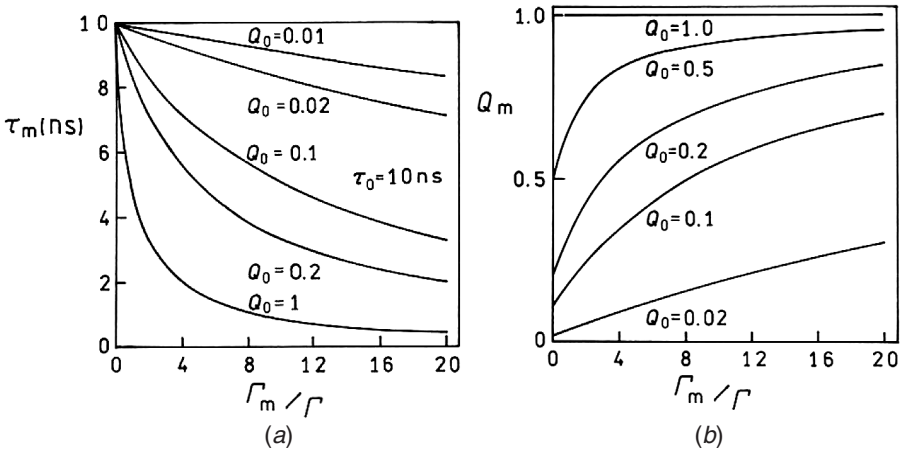


**Figure 3** Effect of an increase in radiative decay rate ($\Gamma_m$ / $\Gamma$) on the lifetime and quantum yield of a fluorophore.

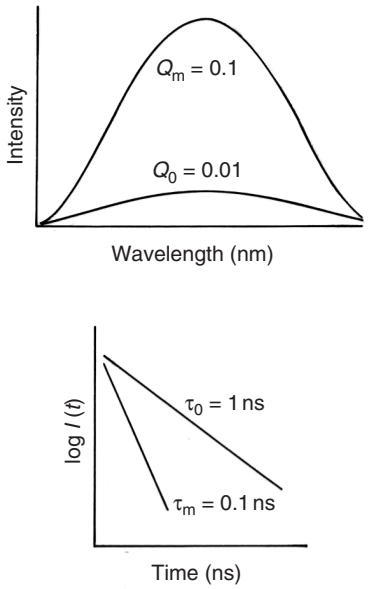


**Figure 4** Metallic surfaces can create unique fluorophores with high quantum yields and short lifetimes. These are likely to find multifarious applications in metal-enhanced fluorescence sensing.

fluorophore due to spending less time in an excited state, that is, less time for oxidation and/or other processes.

We speculated that the properties of the fluorophores with different quantum yields would be affected differently by the presence of nearby metal due to the difference in radiative rate modifications, with a view to understanding metal–fluorophore interactions for sensing applications. We tested this hypothesis by investigating the changes in emission properties of two fluorophores with similar absorption/emission spectra but with different quantum yield (Rhodamine B, quantum yields, $Q_0 = 0.48$ and Rose Bengal, $Q_0 = 0.02$). Figure 5 summarizes our observations with these fluorophores. The emission from Rhodamine B on silver island films (SiFs) was ≈20 percent higher as compared to the emission from the glass side. On the other hand, emission from Rose Bengal on SiFs was ≈5-fold higher than those on the glass side. The emission spectra in Figure 5 do not solely demonstrate an increase in the radiative decay rate. However, the demonstration of such an increase can be done by additional intensity decay (lifetime) measurements. Figure 6

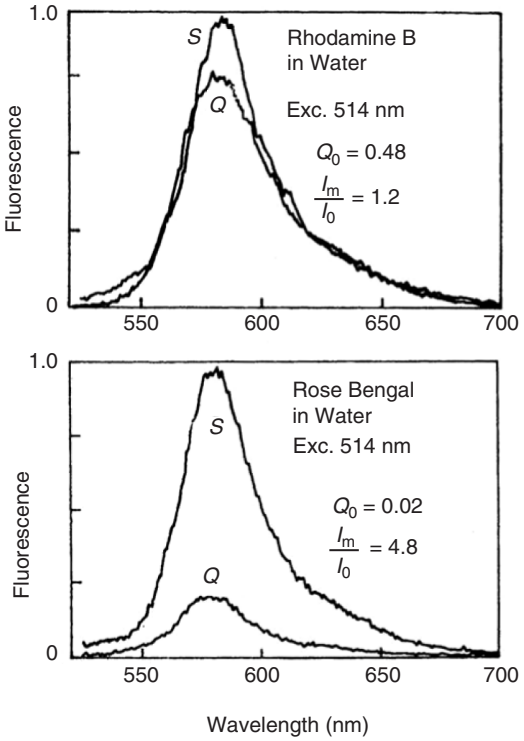


**Figure 5** Effect of silver island films (SiFs) on the emission spectra of (*a*) Rhodamine B and (*b*) Rose Bengal. These probes were chosen to demonstrate metal-enhanced fluorescence as they both have similar spectral characteristics but different quantum yields. Adapted from Ref. 2.

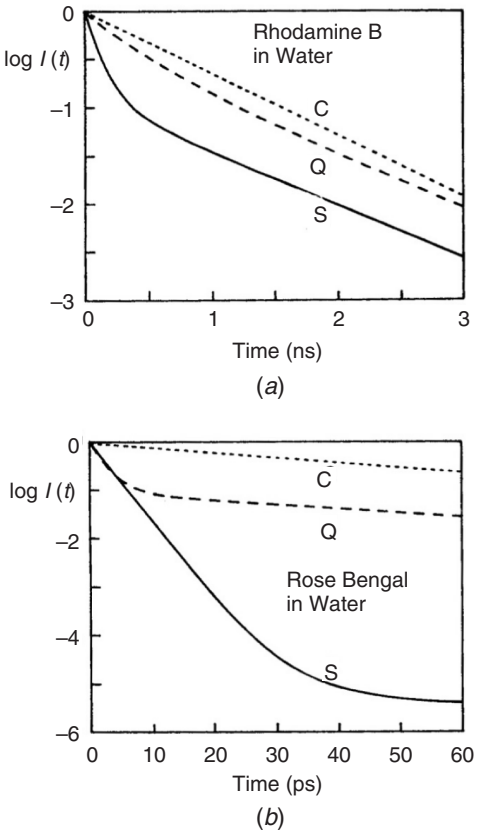


**Figure 6** Intensity decays for (*a*) Rhodamine B and (*b*) Rose Bengal in cuvette (C), between quartz slides (Q), and between silvered slides (S). Adapted from Ref. 2.

shows that the intensity decay is more rapid near SiFs. We interpret this decrease in lifetime as due to the interactions of the fluorophore with the metal surface. Hence an increase in intensity accompanied by a reduction in lifetime can only be explored by a radiative rate modification; compare Figures 1 and 2 and the following equations:

$$Q_{\mathrm{m}} = \frac{\Gamma + \Gamma_{\mathrm{m}}}{\Gamma + \Gamma_{\mathrm{m}} + k_{\mathrm{nr}}}$$

$$\tau_{\mathrm{m}} = \frac{1}{\Gamma + \Gamma_{\mathrm{m}} + k_{\mathrm{nr}}}$$

The results from Rhodamine B and Rose Bengal (Figs. 5 and 6) were consistent with our expectations that the presence of metal increases the emission

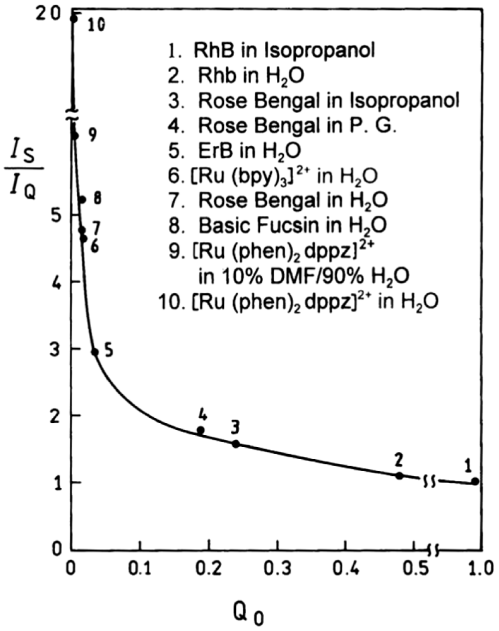


**Figure 7** Effect of silver island films on the quantum yields of both high and low quantum yield fluorophores. $I_S$, intensity between silvered plates, $I_Q$, intensity between quartz plates (control sample). Adapted from Ref. 2.

intensity (quantum yield) and decreases the lifetime of the fluorophores. However, one could be concerned with possible artifacts due to dye binding to the surfaces or other unknown effects. For this reason, we examined a number of additional fluorophores between uncoated quartz plates and between silver island films. In all cases, the emission was more intense for the solution between the silver islands. For example, $[Ru(bpy)_3]$ and $[Ru(phen)_2dppz]$ have quantum yields near 0.02 and 0.001, respectively. A larger enhancement was found for $[Ru(phen)_2dppz]$ than for $[Ru(bpy)_3]$ (Fig. 7). The enhancements for 10 different fluorophore solutions are shown in Figure 7. In all cases, a lower bulk-phase quantum yield results in larger enhancements for samples between silver island films. We rationalized that the simple enhancement factor is $\approx 1/Q_0$, where $Q_0$ is the quantum yield (intensity) in the absence of metal.

## 2 METAL SUBSTRATES FOR METAL-ENHANCED FLUORESCENCE

The effects of metals in close proximity to fluorophores have been predicted theoretically [5–7] and are analogous to the increases in Raman signals

observed for surface-enhanced Raman scattering (SERS), *except* that the effects on fluorescence are due to through-space interactions, which do not require molecular contact between the fluorophores and the metal. For use in medical and biotechnology applications, such as diagnostic or microfluidic devices, it would be useful to obtain metal-enhanced fluorescence (MEF) at desired locations in the measurement device, that is, *MEF on demand*. We subsequently investigated various methods for the preparation of silver nanostructures, and used indocyanine green (ICG) and fluorescein-labeled proteins and oligonucleotides to test our new potential sensing platforms, noting the widespread use of ICG and fluorescein in medical applications due their Food and Drug Administration (FDA) approval for use in humans.

## Silver Island Films for MEF

In recent years we have been reporting our observations on the favorable effects of silver nanoparticles deposited randomly on glass substrates (silver island films, SiFs) for increasing the intensities and photostability of fluorophores, particularly those with low quantum yields [2–4, 8]. A typical absorption spectrum of SiFs, which display absorption maximum near 430 nm, is given in Figure 8*a*. Figure 8*b* also shows the schematic for a fluorophore between two SiFs (note that the configuration of slides were used for the experiments throughout this work, unless otherwise indicated). We have investigated the effects of SiFs on the properties of ICG, when bound to human serum albumin (HSA). The emission spectra of ICG–HSA bound to quartz and SiFs are shown in Figure 8*b*. The intensity of ICG is increased approximately 10-fold on the SiFs as compared with quartz. The spectral shape is similar both on quartz and SiFs. We found the same amount of increase in the emission of ICG whether the surfaces were coated with HSA, which already contained bound ICG, or if the surfaces were first coated with HSA followed by exposure to a dilute solution of ICG. From ongoing studies of albumin-coated surfaces, we estimated that the same amount of HSA binds to each surface, with the difference in binding being less than a factor of 2. Hence the observed increase in the intensity on SiFs is not due to the increased ICG–HSA binding but rather due to a change in the quantum yield and/or rate of excitation of ICG near metallic silver nanoparticles [8].

## Silver Colloid Films for MEF

In contrast to silver island films, the preparation of colloidal suspensions of silver yields homogeneously sized spherical silver particles. An advantage of a colloidal suspension is the possibility of injection for medical imaging (silver and gold colloids are widely used in medicine), such as in retinal angiography, which could widely benefit from an enhanced ICG and fluorescein quantum yield and photostability. Thus, we also investigated the effects of spherical silver particles on the emission of ICG. The sample geometry was similar to

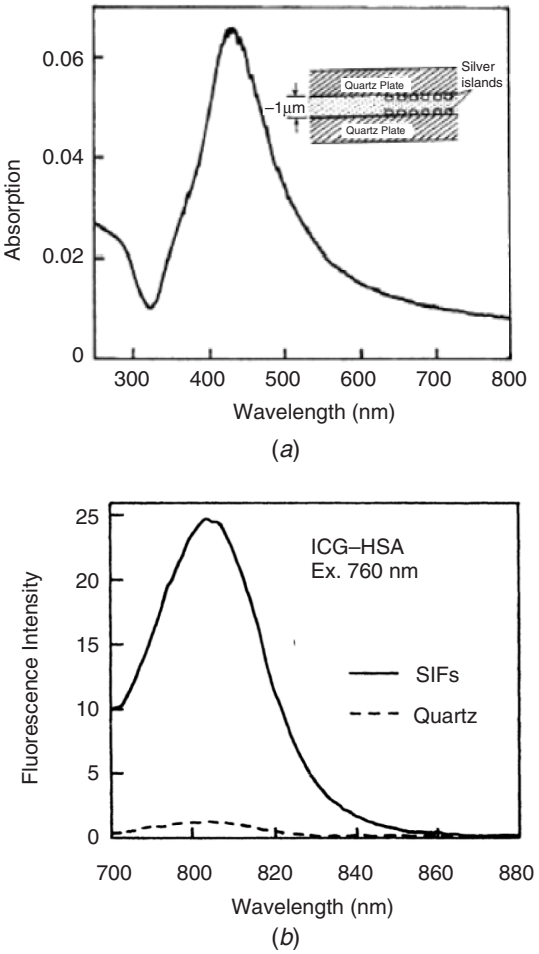


**Figure 8** Typical absorption (*a*) spectrum silver island films (SiFs). Schematic for a fluorophore solution between two SiFs (inset). (*b*) Fluorescence spectra of ICG–HSA on SiFs and on quartz. Adapted from Ref. 8.

that of SiFs, except the spherical silver particles were prepared separately and were immobilized to 1-amino-3-propylethoxy silane (APS)-coated glass slides by immersing the glass in a solution of the particles (Fig. 9*a*). Figure 9 shows an absorption spectrum, typical of our colloid-coated APS glass slides. The absorption peak centered near 430 nm is typical of colloidal silver particles with subwavelength dimensions. An atomic force microscopy (AFM) image of silver-colloid-coated glass slides shows that the size of the silvercolloids was smaller than 50 nm with partly aggregated sections (Fig. 9*b*). The surfaces were incubated with ICG–HSA to obtain a monolayer surface coating. The emission spectra showed a ≈ 30-fold larger intensity on the surfaces coated with

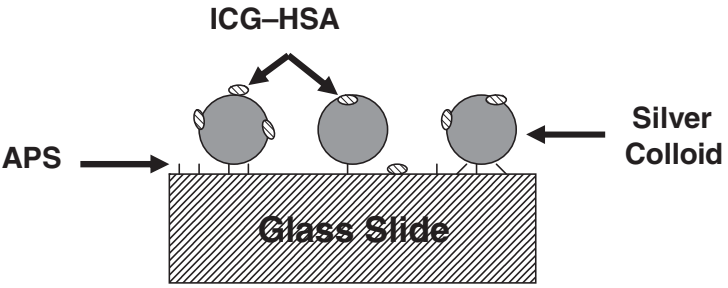


(*a*)

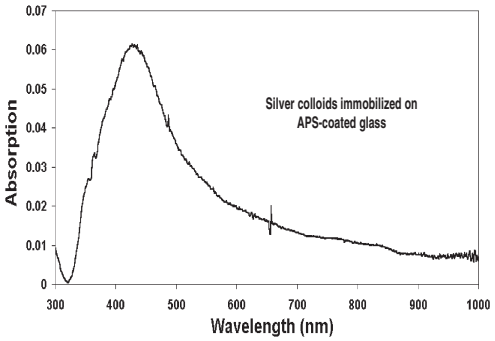


(*b*)

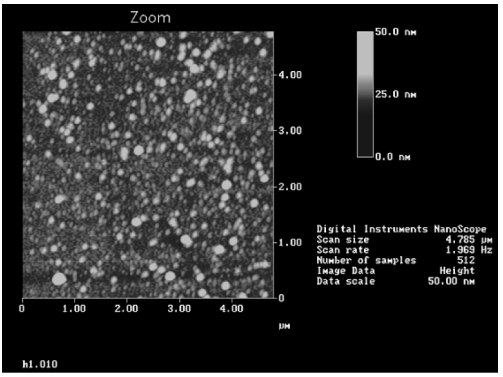


(*c*)

**Figure 9** (*a*) Glass surface geometry. (*b*) APS is used to functionalize the surface of the glass with amine groups, which readily bind silver colloids, absorption spectrum of silver colloids on APS-coated glass, and (*c*) AFM image of a silver colloid coated glass. APS; 1-amino-3-propylethoxy silane. Adapted from Ref. 9.

silver colloids (Fig. 10) [9]. The fluorescence intensities typically increased with the concentration of APS used to treat the cleaned surfaces (Fig. 11*a*), which also appeared to correlate with the optical density of silver colloids at 430 nm (not shown). We also measured the lifetimes of ICG on both surfaces and observed a significant reduction of the lifetimes on the silver colloids (Fig. 11*b*), providing additional evidence that the increase in intensity is in fact

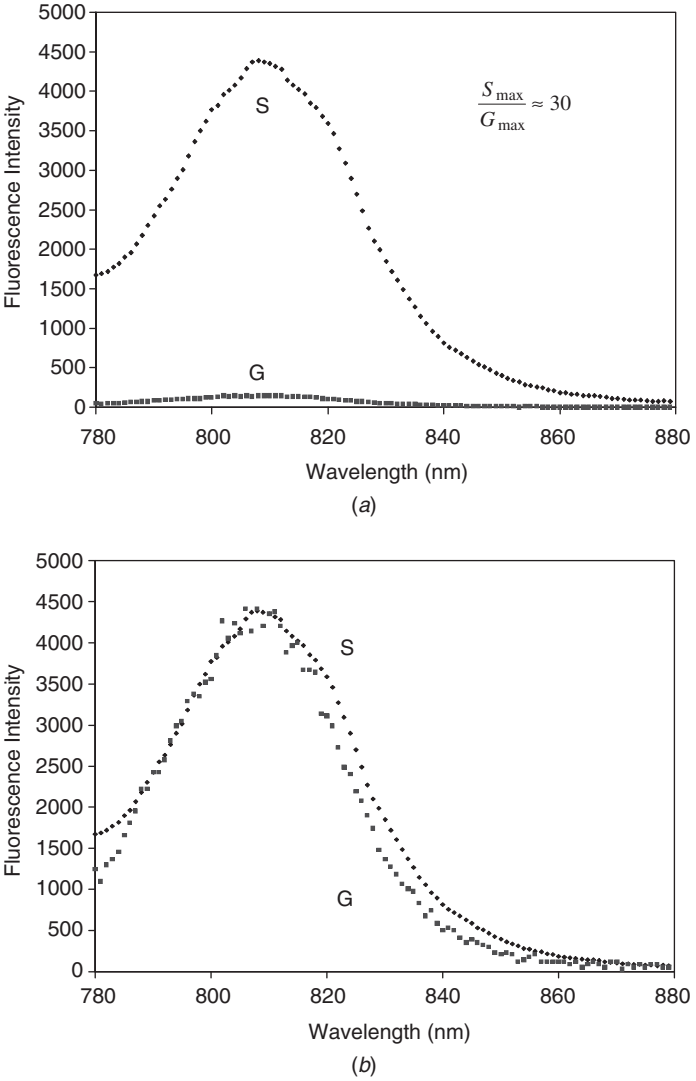


**Figure 10** (*a*) Fluorescence intensity of ICG–HSA coated glass, *G*, and above silver colloids, *S*, Ex = 760 nm. (*b*) Fluorescence intensities normalized to the intensity on silver colloids. Adapted from Ref. 9.

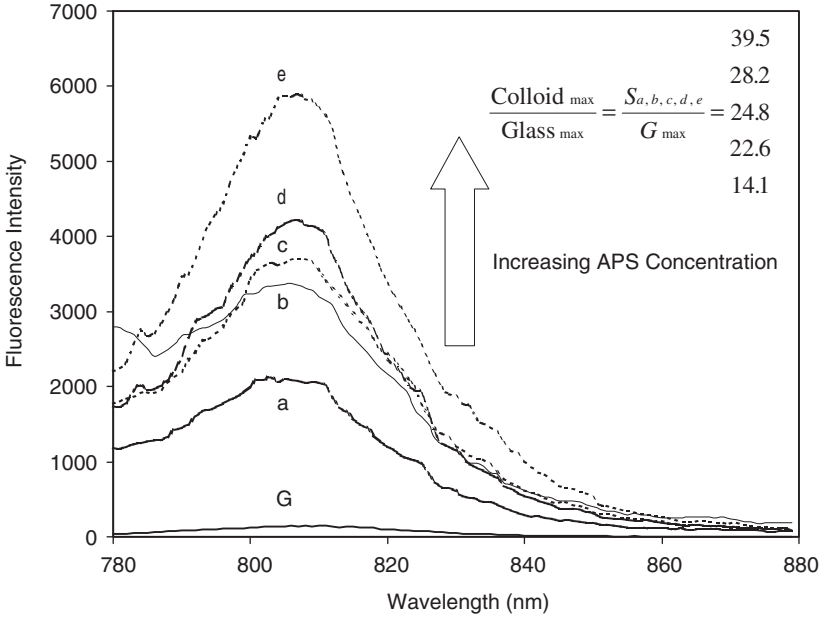
7000
6000
5000
4000
3000
2000
1000
0
Fluorescence Intensity
780
800
820
840
860
880
Wavelength (nm)
e
d
c
b
a
G
39.5
28.2
24.8
22.6
14.1
Colloid max / Glass max = Sa,b,c,d,e / G max
Increasing APS Concentration

(*a*)

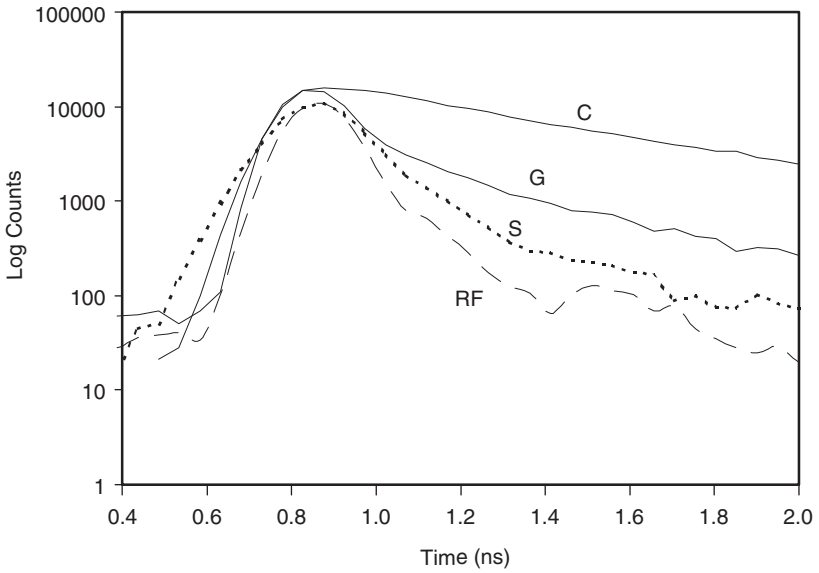
100000
10000
1000
100
10
1
Log Counts
0.4
0.6
0.8
1.0
1.2
1.4
1.6
1.8
2.0
Time (ns)
C
G
S
RF

(*b*)

due to a modification of the radiative decay rate, $\Gamma_m$, by the close proximity to silver colloids [9].

## Anisotropic Silver Nanostructures for MEF

The optical properties of nonspherical particles have been attracting many researchers since the early twentieth century. Figure 12*a* shows typical absorption spectra of silver colloids, which can be calculated for the small particle limit ($r << \lambda$) from the complex dielectric constant of the metal [10, 11]. The absorption spectra are dependent on the shape of the particles as well as the dielectric constant of the surrounding medium, with elongated spheroids displaying longer absorption wavelengths.

Several groups have considered the effects of metallic spheroids on the spectral properties of nearby fluorophores [12–16]. A typical model is shown in Figure 12*b* for a prolate spheroid with an aspect ratio of $a/b$. The particle is assumed to be a metallic ellipsoid with a fluorophore positioned near the particle. The fluorophore is located outside the particle at a distance $r$ from the center of the spheroid and a distance $d$ from the surface. The fluorophore is located on the major axis and can be oriented parallel or perpendicular to the metallic surface. The presence of a metallic particle can have dramatic effects on the radiative decay rate of a nearby fluorophore. Figure 12*c* shows the radiative rates expected for a fluorophore at various distances from the surface of a silver particle and for different orientations of the fluorophore transition moment. The most remarkable effect is for a fluorophore perpendicular to the surface of a spheroid with $a/b = 1.75$. In this case the radiative rate can be enhanced by a factor of 1000-fold or greater. The effect is much smaller for a sphere ($a/b = 1.0$) and much smaller for a more elongated spheroid ($a/b = 3.0$) when the optical transition is not in resonance.

We have developed a methodology for depositing silver nanorods with controlled loadings onto glass substrates. The absorption spectra of silver nanorods deposited on glass substrates are shown in Figure 13*a*. Silver nanorods display two distinct surface plasmon peaks; transverse and longitudinal, which typically appear at ≈420 and ≈650 nm, respectively. In our experiments, the longitudinal surface plasmon peak shifted and increased in absorbance as more nanorods are deposited on the surface of the substrates.

**Figure 11** Fluorescence intensity of ICG–HSA on silver colloids as a function of increased APS used. Cleaned glass slides were initially soaked in: (a) 0.1, (b) 0.25, (c) 0.5, (d) 1.0, and (e) 1.25 percent (v/v) APS solution for 4 h, washed and soaked in colloid solution for 4 days. (*a*) G—Fluorescence intensity of ICG–HSA deposited on 0.5 percent APS covered glass. ICG coverage for the glass controls, i.e., for the different percent APS-coated glass slides, was approximately constant. (*b*) Complex intensity decays of ICG–HSA in a cuvette (buffer), *C*, on glass slides, *G*, and silver colloids, *S*. *RF*—Instrumental response function. Adapted from Ref. 9.

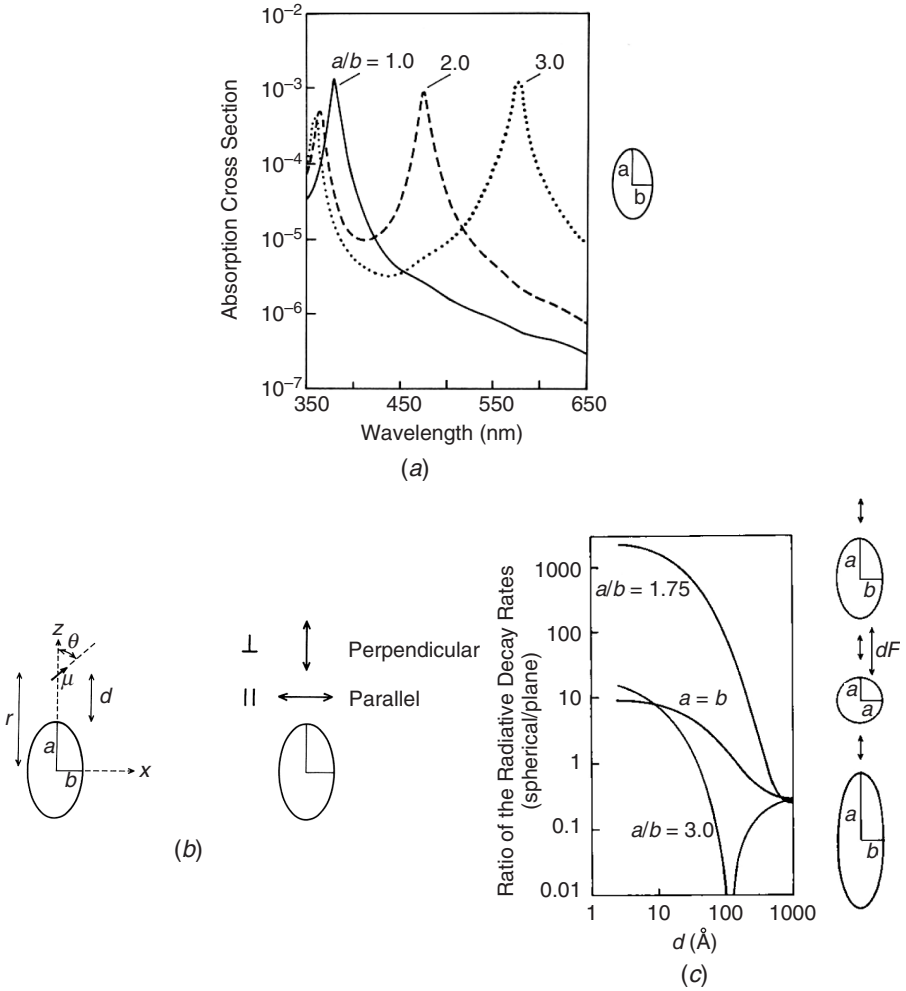


**Figure 12** (*a*)Calculated absorption cross section of silver nanostructures. (*b*, *c*) Effect of a metal spheroid on the radiative decay rate. Adapted from Ref. 1.

In parallel to these measurements, we have observed an increase in the size of the nanorods (by AFM, data not shown). In order to compare the extent of enhancement of fluorescence with respect to the extent of loading of silver nanorods deposited on the surface, we have arbitrarily chosen the value of absorbance at 650 nm as a means of loading of the nanorods on the surface. This is because the 650-nm band is solely attributed to the longitudinal absorbance of the nanorods.

Figure 13(*b*) shows the fluorescence emission intensity of ICG–HSA measured from both glass and on silver nanorods, and the enhancement factor

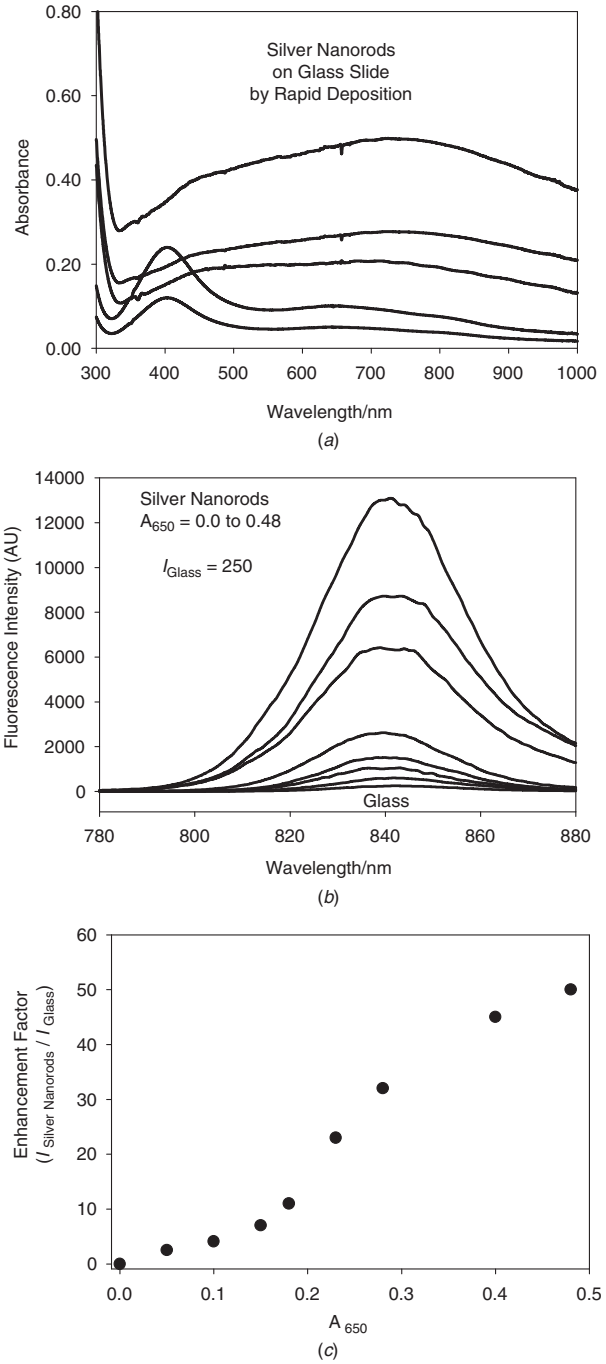


**Figure 13** (*a*)Absorption spectra of silver nanorods, (*b*) emission spectra of ICG–HSA on silver nanorods, and (*c*) enhancement factor versus the absorbance.

versus the loading density of silver nanorods. We have obtained up to 50-fold enhancement in emission of ICG on silver nanorods when compared to the emission on glass (Fig. 13*c*). We also measured the lifetime of ICG on both glass and silver nanorods and observed a significant reduction in the lifetime of ICG on silver nanorods, providing the evidence that an increased emission is due to radiative rate modifications, $\Gamma_m$.

## Laser-Deposited Silver for MEF

It would be useful to obtain MEF at desired locations in a measurement device for use in medical and biotechnology applications, such as diagnostic or microfluidic devices, or even plate wells. While a variety of methods could be used, we reasoned that the light-directed deposition of silver would be widely applicable. In a typical preparation of light-induced deposition of silver on glass slides; a 180-μL solution of silver colloids (optical density = 0.1) was syringed between the glass microscope slide and the plastic cover slip, which created a microsample chamber 0.5 mm thick (Fig. 14). The sample chamber was irradiated by a HeCd laser beam, with a power of ~8 mW, which was collimated but defocused using a 10× microscope objective [numerical aperture (NA)] 0.40, to provide illumination over a 0.5-mm-diameter spot [17]. This resulted in deposition of silver on the surface of the glass substrate.

We examined the emission spectrum of ICG–HSA when bound to illuminated or nonilluminated regions of the APS-treated slides. For APS-treated slides (Fig. 15) the intensity of ICG was increased about sevenfold in the

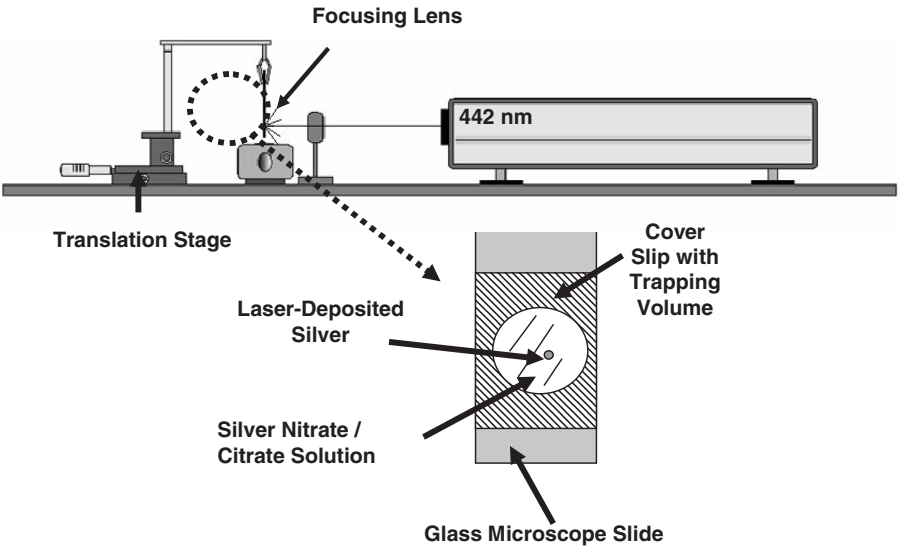


**Figure 14** Experimental setup for light-induced deposition of silver on APS-coated glass microscope slides. Adapted from Ref. 17.

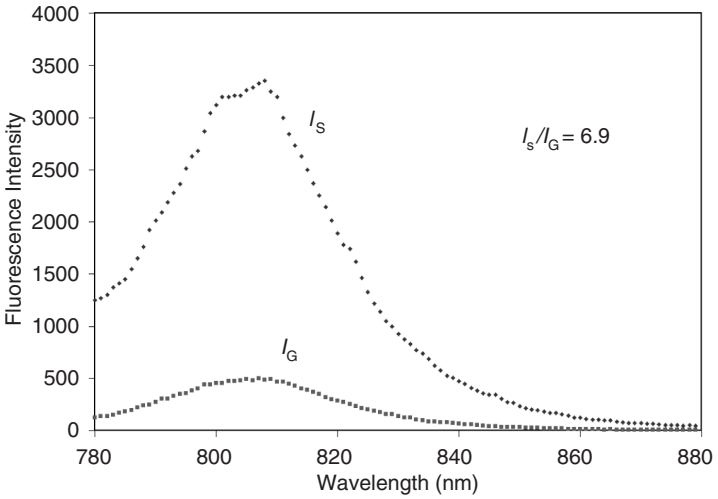


**Figure 15** Fluorescence spectra of ICG–HSA on glass and on light-deposited silver. $I_S$, intensity on silver, $I_G$, intensity on glass. Adapted from Ref. 17.

regions with laser-deposited silver. We examined the photostability of ICG–HSA when bound to glass or laser-deposited silver. We reasoned that ICG molecules with shortened lifetimes should also be more photostable because there is less time for photochemical processes to occur. The intensity of ICG–HSA was recorded with continuous illumination at 760 nm. When excited with the same incident power, the fluorescence intensities, when considered on the same intensity scale, decreased somewhat more rapidly on the silver (Fig. 16*a*). Alternatively, one can consider the photostability of ICG when the incident intensity is adjusted to result in the same initial signal intensities on silver and glass. In this case (Fig. 16*b*) photobleaching is slower on the silver surfaces. The fact that the photobleaching is not accelerated for ICG on silver indicates that the increased intensities on silver *are likely not due* to an increased rate of excitation.

## Electrochemically Deposited Silver for MEF

One of our methods for silver deposition onto glass/quartz substrates was to pass a controlled current between two silver electrodes ($9 \times 35 \times 0.1$ mm, separated by 10 mm) in pure water under illumination with laser (Fig. 17*a*) [18]. For the production of silver colloids, a simple constant-current generator circuit (60 μA) was used. After 30 min of current flow, a clear glass microscope slide was positioned within the cuvette (no chemical glass surface modifications) and was illuminated (HeCd, 442 nm). We observed silver deposition on the glass microscope slide, the amount depending on the illumination time. Simultaneous electrolysis and 442-nm laser illumination resulted in the

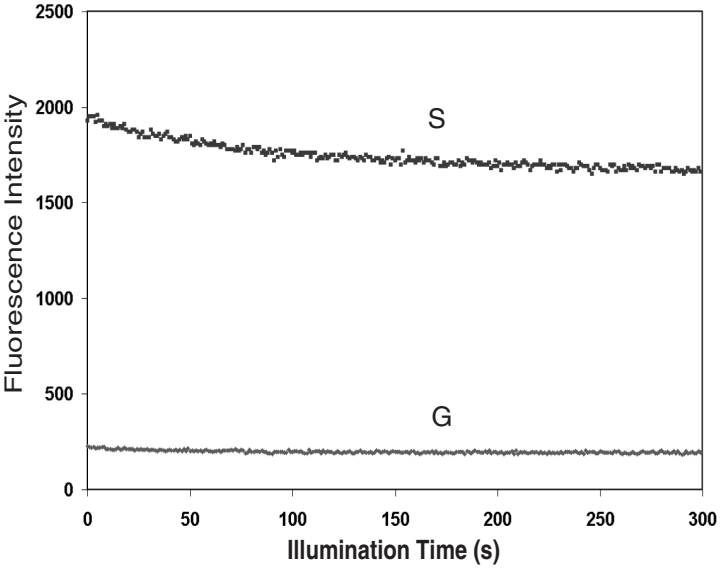


(*a*)

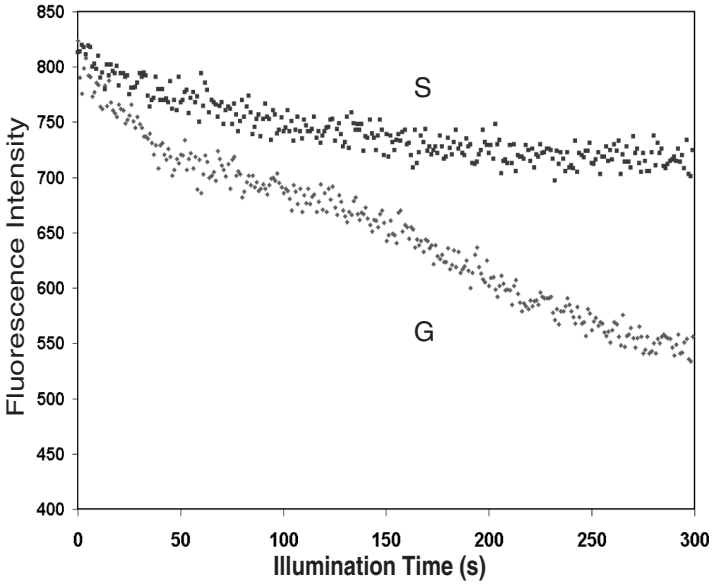


(*b*)

**Figure 16** Photostability of ICG–HSA on (G) glass, and (**S**) laser-deposited silver, measured with the same excitation power at (*a*) 760 nm, and (*b*) with the laser power at 760 nm adjusted for the same initial fluorescence intensity. Laser-deposited samples were made by focusing 442-nm laser light onto APS-coated glass slides immersed in a $AgNO_3$ citrate solution for 15 min. The OD of the sample was ~0.3. Adapted from Ref. 17.

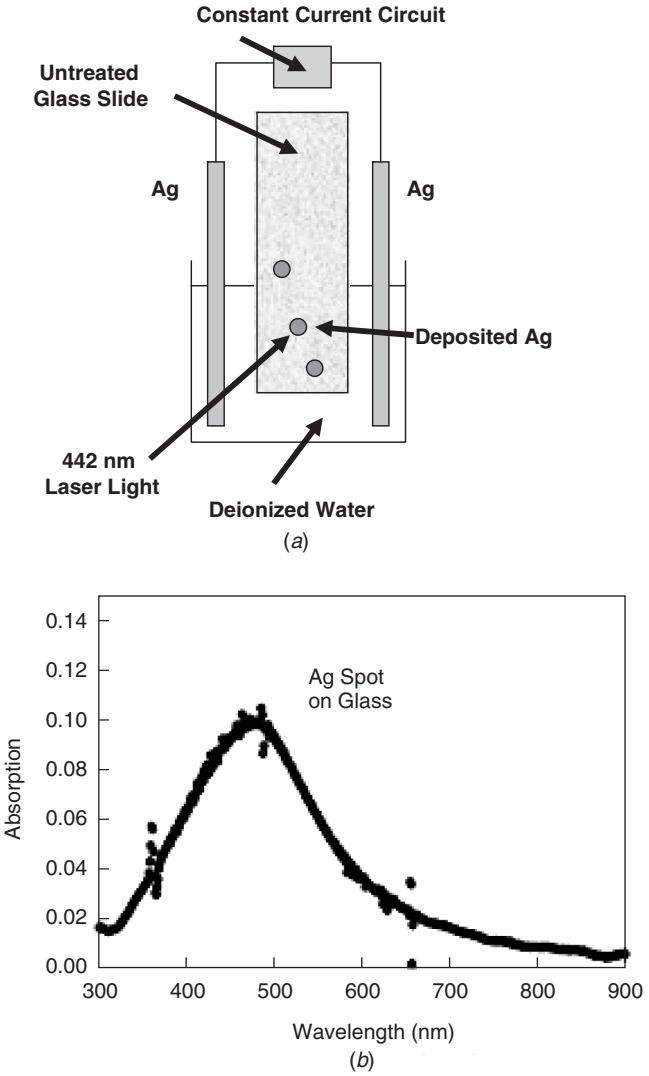


**Figure 17** Light-deposited silver produced electrochemically. (*a*) Constant current circuit and, (*b*) absorbance spectrum of silver spot on glass. Adapted from Ref. 18.

deposition of metallic silver in the targeted illuminated region of the glass substrate. An absorption spectrum of the deposited silver is shown in Figure 17*b*. A single absorption band is present on glass indicating that the silver particles are somewhat spherical. We examined ICG–HSA when coated on glass (G) or silver particles (S). The emission intensity was increased about 18-fold on the silver particles (Fig. 18). We found a dramatic increase

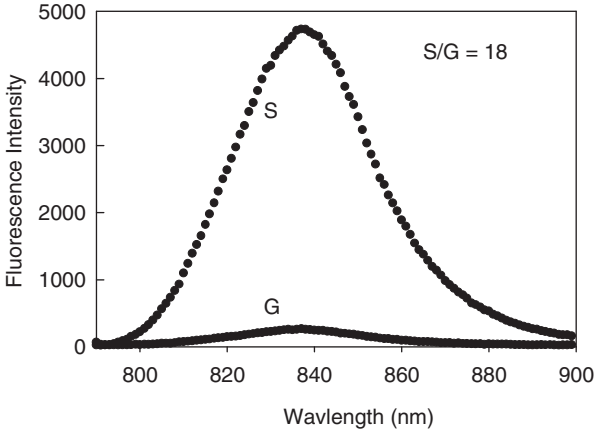


**Figure 18** Fluorescence spectra of ICG–HSA on glass (G) and on light-deposited silver (S). Adapted from Ref. 18.

in the photostability near the silver particles (Fig. 19). This very encouraging result indicates that a much higher signal can be obtained from each fluorophore prior to photodestruction, and that more photons can be obtained per fluorophore before the ICG on the silver eventually degrades [18].

## Electroplating of Silver for MEF

In this method, the silver cathode electrode (as shown in the scheme in Fig. 17) was replaced with an ITO-coated glass electrode (Fig. 20*a*). The current was again 60 μA. After a short period of time, silver readily deposited on the ITO surface (no laser illumination), the extent of which was again dependent on the exposure time. Two maxima were found on the ITO absorption spectrum, which eventually formed one large broad band. This suggests that the particles are elongated and display both transverse and longitudinal resonances (not shown). Enhanced fluorescence emission from ICG–HSA was also found for silver particles on ITO (Fig. 20*b*), and there also appeared to be a small blue shift on silvered ITO [18].

## Roughened Silver Electrodes for MEF

In a typical preparation, silver electrodes are placed 10 mm apart in deionized water (Fig. 21) [19]. A constant current of 60 μA is supplied across the two electrodes for 10 min by a constant-current generator. Figure 22 shows the time-dependent growth of the silver nanostructures on the silver cathode. In comparison, the anode was relatively unperturbed. Three electrodes were coated with ICG–HSA and studied; the roughened cathode, the anode, and an unroughened electrode. Essentially no emission was seen from ICG–HSA on

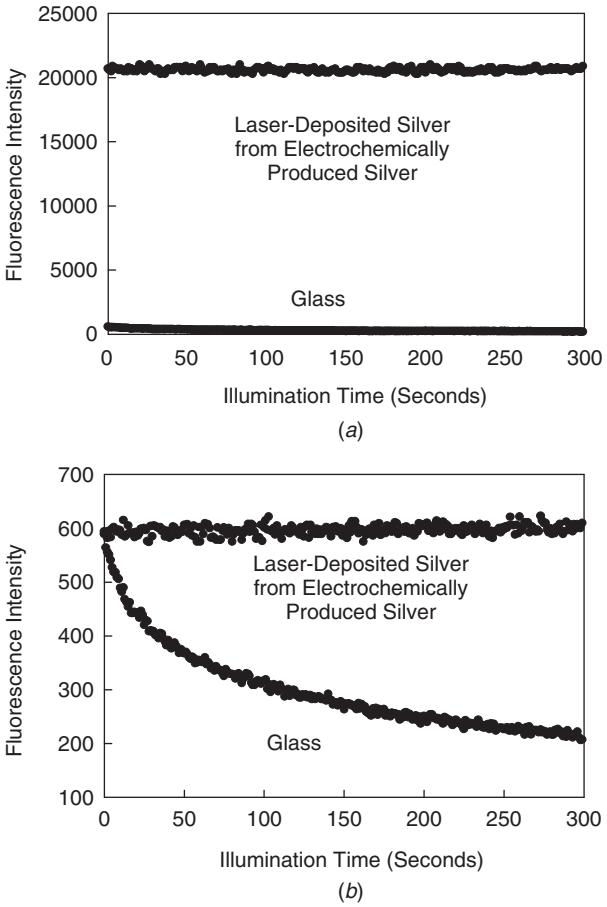


**Figure 19** Photostability of ICG–HSA on glass and laser-deposited silver produced via electrolysis (*a*) measured using the same excitation power at 760 nm and (*b*) with power *adjusted* to give the same initial fluorescence intensities. In all the measurements, vertically polarized excitation was used, while the fluorescence emission was observed at the magic angle, that is, 54.7°. Adapted from Ref. 18.

an unroughened, bright silver surface (Ag in Fig. 23*a*, the control). However, a dramatically larger signal was observed on the roughened cathode, and a somewhat smaller signal was observed on the anode. In all our experiments we typically found that the roughened silver cathode was ~20 to 100-fold more fluorescent than the unroughened control Ag electrode. In comparison the Ag anode was 5 to 50 times more fluorescent than the Ag control. When we increased the time for roughening to over 1 h, the intensities of both electrodes after coating with ICG–HSA were essentially the same, but still 50-fold more fluorescent than the unroughened Ag control. The emission spectra on the two electrodes probably had the same emission maximum, where the slight shift

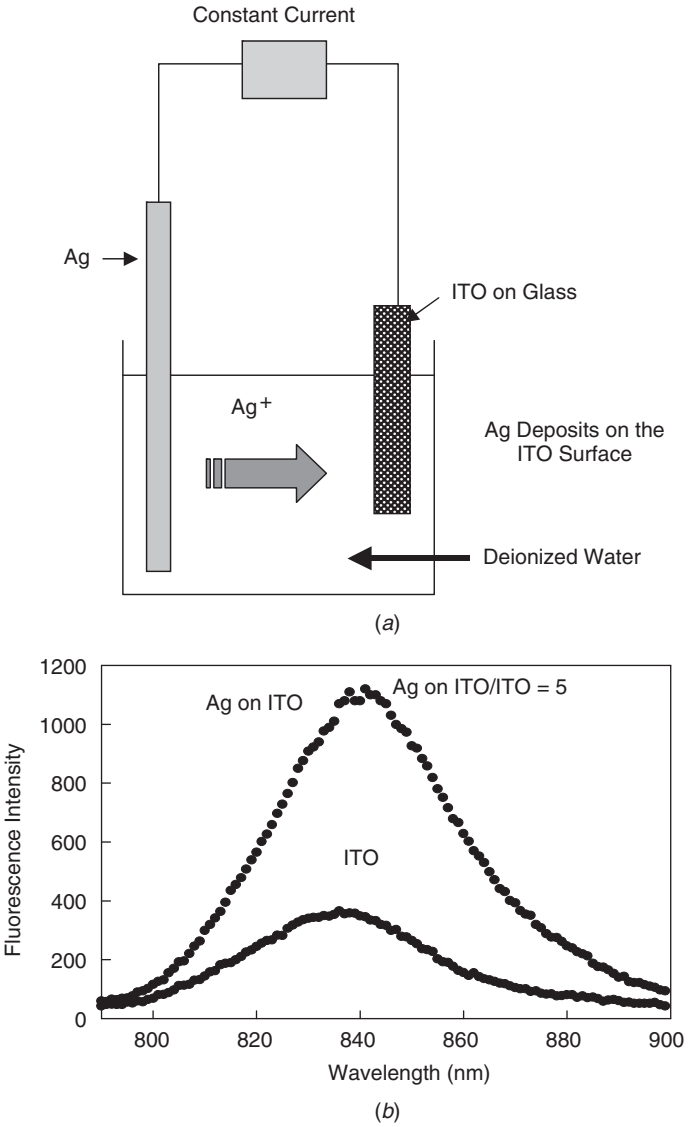


**Figure 20** (*a*) Electroplating of silver on substrates and (*b*), fluorescence emission spectra of ICG–HSA on ITO and silver deposited on ITO. Adapted from Ref. 18.

seen in Figure 23*a* is thought to be due to the filters used to reject the scattered light. It should be noted that the amount of material coated on both surfaces was approximately the same, and the effect was not due to an increased surface area and therefore increased protein coverage on the roughened surface [19].

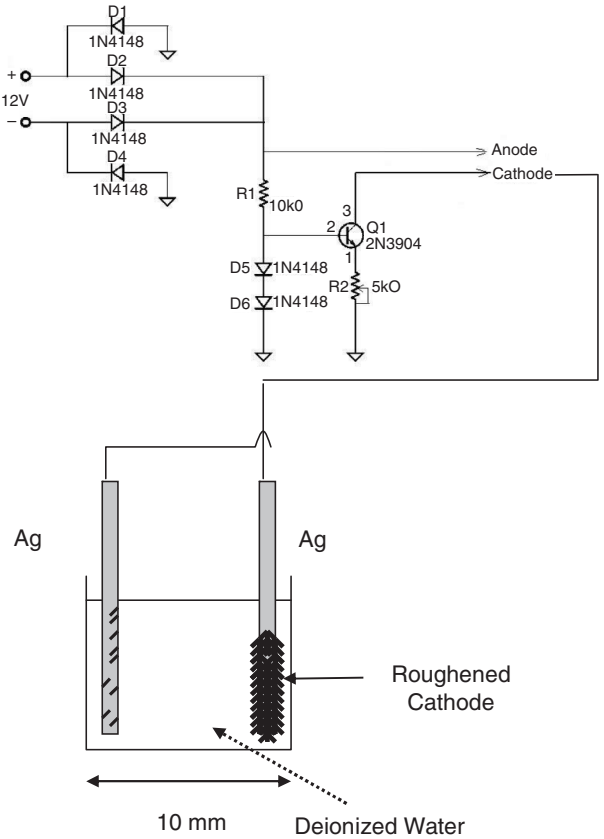


**Figure 21** Experimental setup for the production of roughened silver electrodes. Adapted from Ref. 18.

To place our findings in context with the enhanced Raman signals observed with chloride-dipped electrodes, we examined the Ag cathode after dipping in $10^{-4}$ M NaCl for 1 h. In contrast to SERS electrodes, we observed a decrease in fluorescence intensity for ICG–HSA chloride-dipped electrodes (Fig. 23*b*), confirming our hypothesis of MEF being a through-space phenomenon as compared to SERS, which is known to be due to a contact interaction.

## Silver Fractal-Like Structures on Glass Substrates for MEF

Using a similar method to produce roughened silver electrodes and electroplated silver, fractal-like silver structures were generated on glass using two silver electrodes held between two glass microscope slides (Fig. 24*a*) [20]. A direct current of 10 μA was passed between the electrodes for about 10 min, during which the voltage started near 5 V and decreased to 2 V. During the

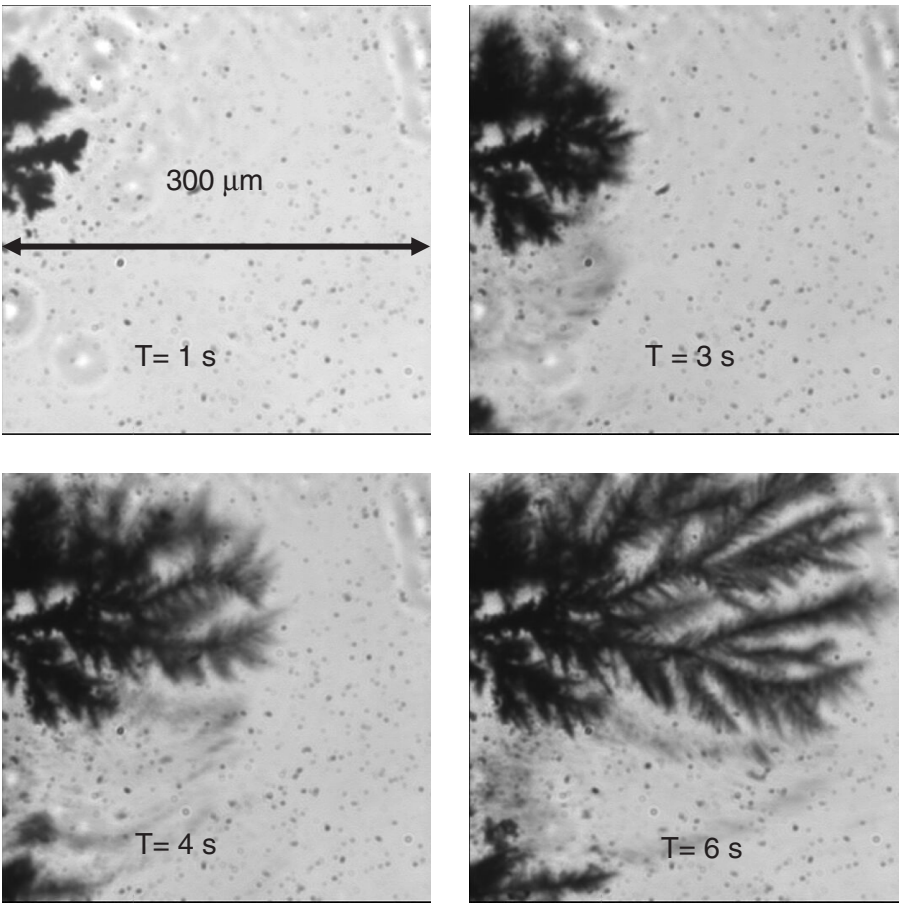


**Figure 22** Fractal-like silver growth on the silver cathode as a function of time, visualized using transmitted light. This structure was characteristic of the whole electrode. Adapted from Ref. 19.

current flow, fractal silver structures grew on the cathode and then on the glass near the cathode (Fig. 24), thus producing silver nanostructures on glass, as compared to those grown on silver electrodes as described in the previous section. Similarly to the silver electrodes, the structures grew rapidly but appeared to twist as they grew. Following passage of the current, the silver structures on glass were coated with FITC–HSA overnight.

For the FITC–HSA-coated fractal silver surfaces on glass, we were able to measure a fluorescence image very similar to that of the fractal silver surface alone (bright-field image) using the same apparatus (Fig. 25). Interestingly, regions of high and low fluorescence intensity were observed (Fig. 25*b*). This

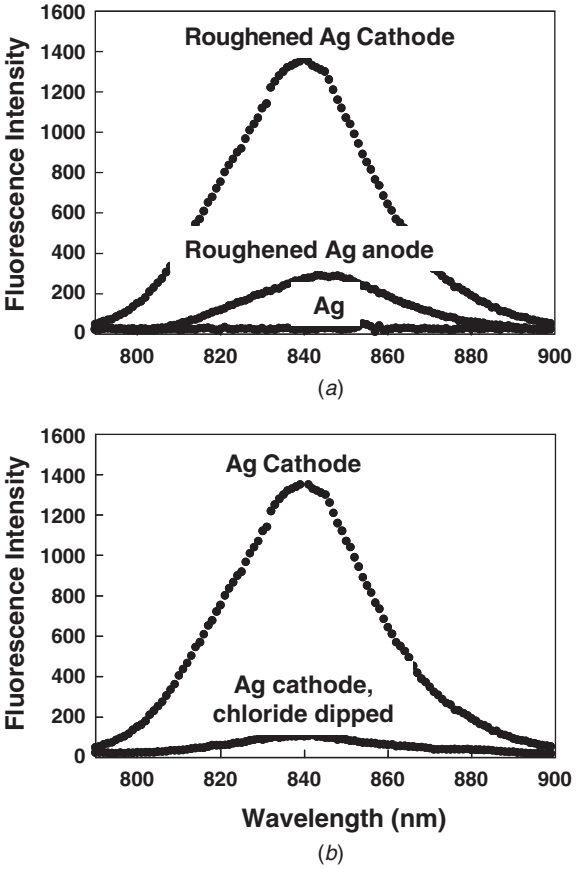


**Figure 23** Fluorescence emission spectra of ICG–HSA on (*a*) roughened silver electrodes and (*b*) silver cathode. Adapted from Ref. 19.

result is roughly consistent with recent SERS data, which showed the presence of intense signals that appeared to be located between clusters of particles [21, 22]. Figure 26 represents the intensity of FITC–HSA on silvered glass. The emission intensities range from 100-fold (position 6) to 600-fold (position 1) greater than the signal from FITC–HSA on unsilvered glass, that is, the control sample. We recognized that some of the increase in intensity of FITC–HSA could be due to binding of more FITC–HSA to silver structures with large surface areas. We note that the fluorescein is not quenched on the surface, probably because the size of an HSA molecule positions the fluorescein about 40 Å from the surface, which is near the distance for maximal radiative rate and therefore fluorescence enhancement; compare Figure 2.

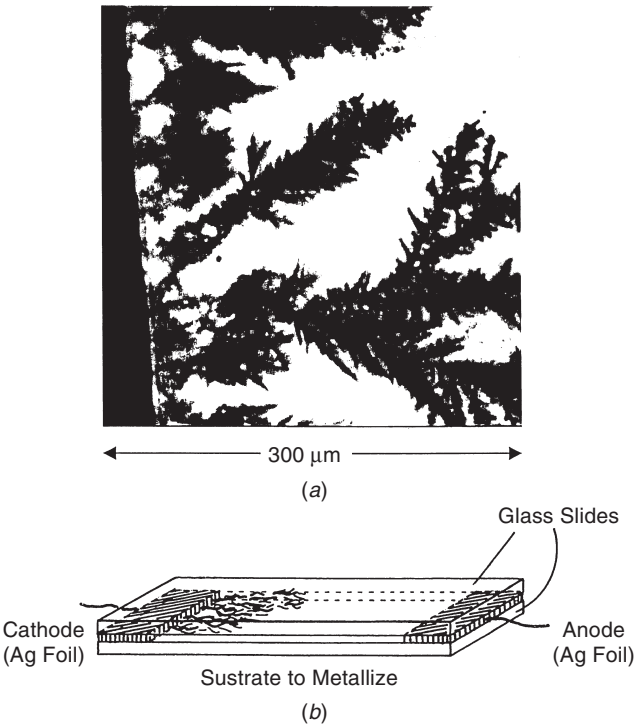


**Figure 24** (*a*) Bright-field image of fractal-like silver surfaces on glass. (*b*) Configuration for creation of fractal-like silver surfaces on glass. Adapted from Ref 20.

We also studied the photostability of FITC on the fractal silver surface and for comparison on silver island films and uncoated quartz. Although the relative photobleaching is higher on fractal silver, the increased rate of photobleaching is less than the increase in intensity (Fig. 27). From the areas under these curves we estimate ≈16-fold and ≈160-fold more photons can be detected from the FITC–HSA on SiFs or fractal silver, respectively, relative to quartz, before photobleaching [20].

As we have shown, metallic silver can be deposited on a variety of surfaces by a variety of methods. The deposited silver is still useful for metal-enhanced fluorescence, and indeed some deposits seem better for overall fluorescence intensity enhancements, for example, silver fractals. Given the versatility of these deposition techniques, we can readily envisage their widespread use in metal-enhanced fluorescence sensing applications and indeed in drug discovery. In the following sections we show some actual enhanced fluorescence sensing schemes that are a direct result of metal-enhanced fluorescence and our knowledge gained from studying different silvered surfaces.

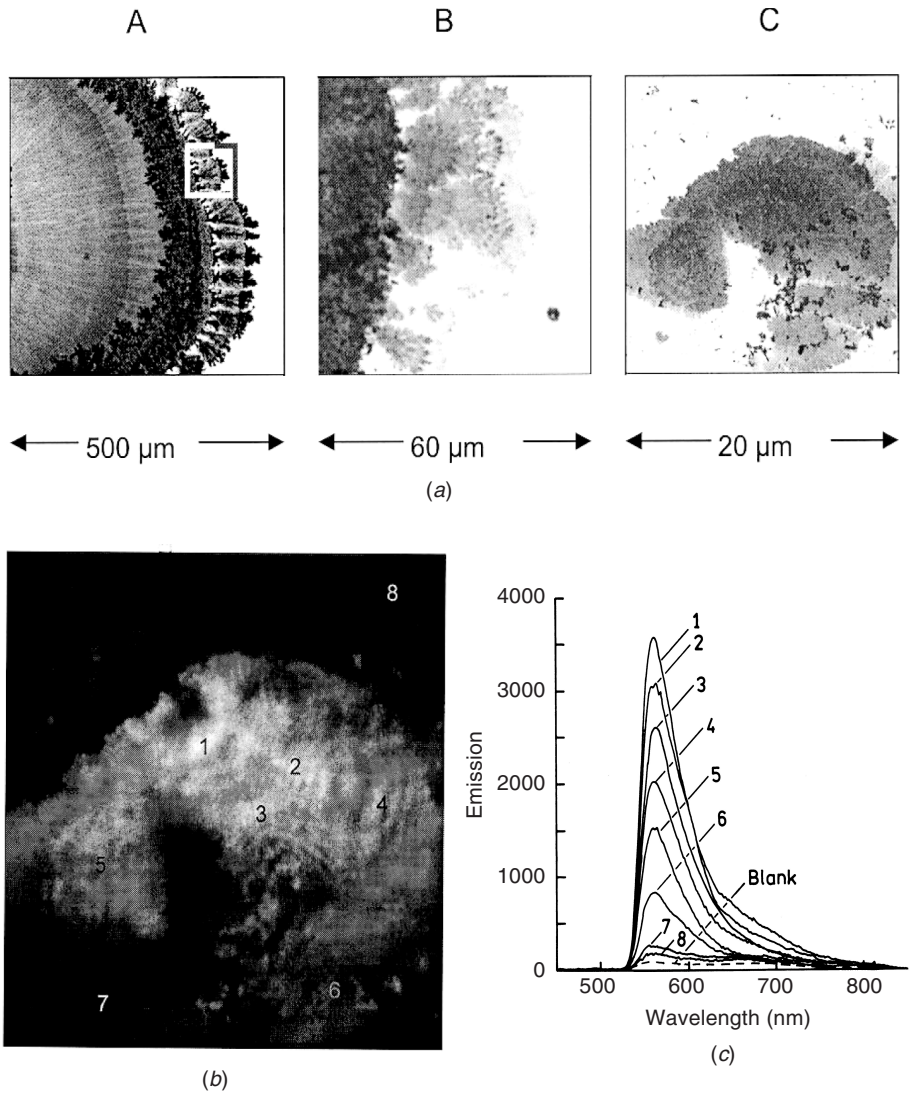


**Figure 25** (*a*) Silver nanostructures deposited on glass during electroplating (A). Panels B and C are consecutive magnification of the marked area on panel A; bright-field image. (*b*) Fluorescence image of FITC–HSA deposited on the silver structure above and (*c*) the emission spectra of the numbered areas shown on the right. Adapted from Ref. 20.

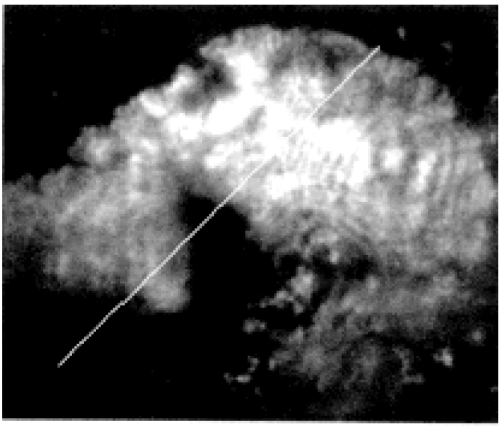

(*a*)

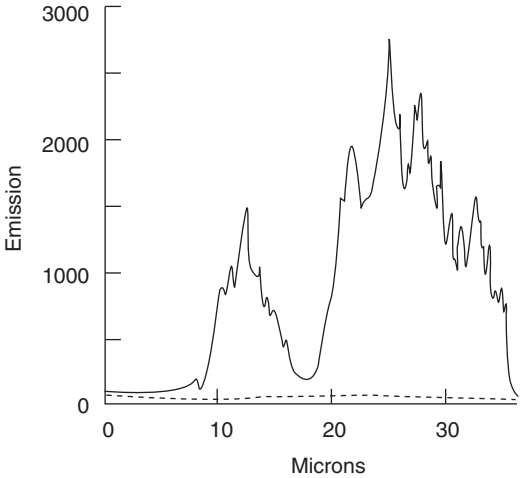


(*b*)

**Figure 26** (*a*) Diagonal scan of the emission intensity of FITC–HSA. (*b*) Dashed line is the intensity observed across a line of equivalent length across unsilvered glass. Adapted from Ref. 20.

## 3 METAL-ENHANCED FLUORESCENCE SENSING: APPLICATION TO SENSING IN PLATE WELLS

### Enhanced DNA/RNA Detection Using MEF (Hybridization Assays)

Detection of deoxyribonucleic acid (DNA) hybridization is the basis of a wide range of biotechnology and diagnostic applications, such as gene chips [23, 24], polymerase chain reaction (PCR) [25, 26], and for fluorescence in situ hybridization [27]. In all these applications increased sensitivity is desirable,

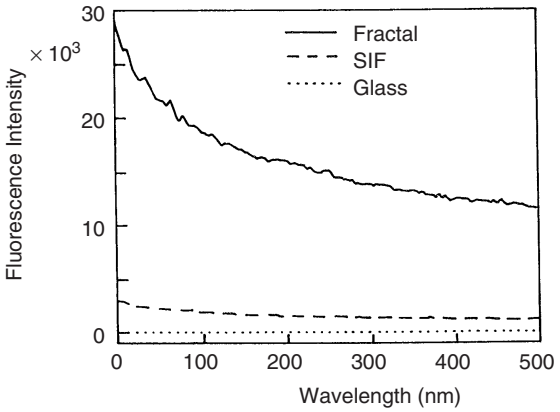


**Figure 27** Photostability of FITC–HSA on silver fractals, SiFs, and glass. Adapted from Ref. 20.

particularly for detection of a small number of copies of biohazard agents. Also, it would be valuable to have a general approach to detect the changes in fluorescence intensity upon hybridization. In general, the detectability of a fluorophore is determined by two factors, the extent of *background emission* from the sample, and the *photostability* of the fluorophore. A highly photostable fluorophore can undergo about $10^6$ excitation–relaxation cycles prior to photobleaching, yielding about $10^3$ to $10^4$ measured photons per fluorophore [28, 29]. Background emission from the samples can easily overwhelm weak emission signals.

In a recent report, we described an approach that should provide a readily measurable change in fluorescence intensity in DNA hybridization formats [30]. Our approach increases the intensity relative to the background and increases the number of detected photons per fluorophore molecule by a factor of 10-fold or more. Figure 28*a* shows the sequence and structure of the oligomers used in these experiments. The thiolated oligonucleotide ss-DNA-SH was used as the capture sequence, which bound spontaneously to the silver particles. The sample containing the silver-bound DNA was positioned in a fluorometer (geometry in Fig. 29) followed by addition of ss-Fl-DNA, which is an amount approximately equal to the amount of silver-bound capture DNA.

The fluorescence intensity began to increase immediately upon mixing and leveled off after about 20 min (Fig. 29*a*). We believe this increase in intensity is due to localization of ss-Fl-DNA near the silver particles by hybridization with the capture oligomer (Fig. 28*c*). In control experiments, we hybridized ss-DNA-SH with ss-Fl-DNA prior to deposition on silver particles. We found a similar 12-fold increase in intensity upon immobilization on silver as compared to an equivalent amount of ds-Fl-DNA-SH in solution.

5'-TCC ACA CAC CAC TGG CCA TCT TC-3'-SH

***ss-DNA-SH***

**Fl**-3'-AGG TGT GTG GTG ACC GGT AGA AG-5'

***ss-Fl-DNA***

5'-TCC ACA CAC CAC TGG CCA TCT TC-3'-SH
**Fl**-3'-AGG TGT GTG GTG ACC GGT AGA AG-5'

***ds-Fl-DNA-SH***

(*a*)

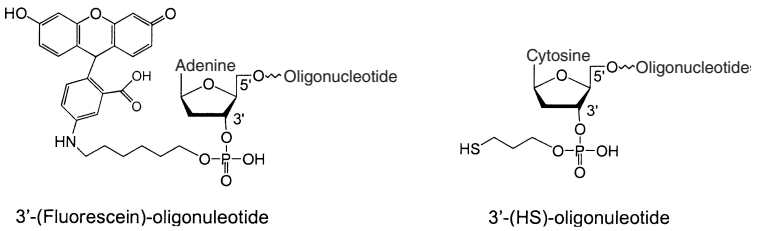


3'-(Fluorescein)-oligonuleotide

3'-(HS)-oligonuleotide

(*b*)

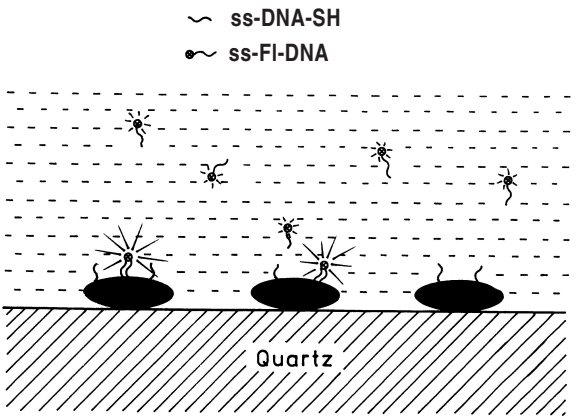


(*c*)

**Figure 28** Structures of DNA oligomers. The lower panel shows a schematic of the oligomers bound to silver particles. Adapted from Ref. 30.

We examined the emission spectra of ss-Fl-DNA before and after hybridization to form ds-Fl-DNA-SH (Fig. 30). The fluorescence intensity was found to be 12-fold higher for the bound form. This dramatic increase can be seen visually by the photographs in Figure 30*b*. There was no detectable shift in the emission spectra. The intensity increase was reversed by melting the DNA at 80°C (Fig. 31) and increased once again upon cooling and presumed rehybridization. The intensity did not recover completely upon

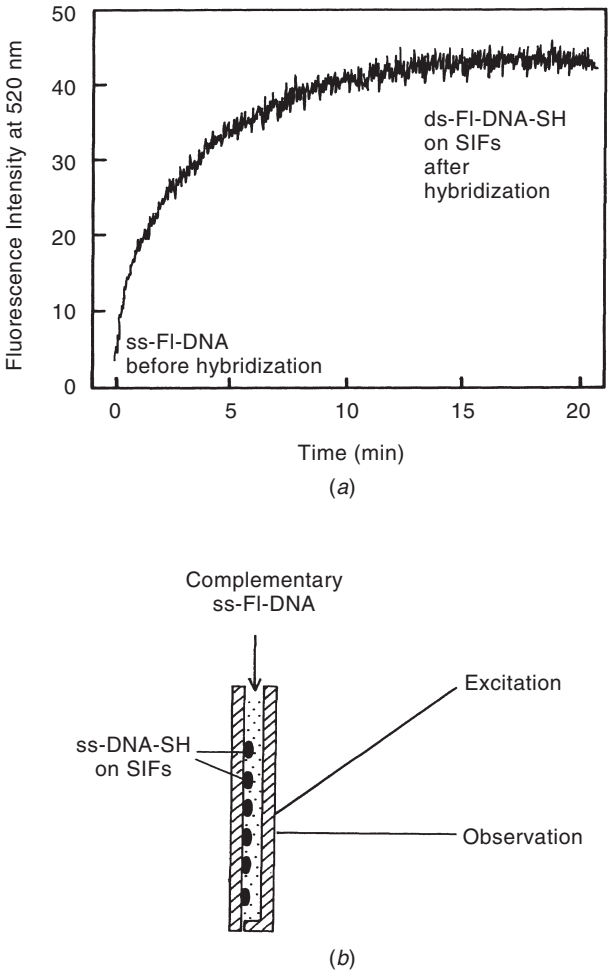


**Figure 29** (*a*) Time-dependent hybridization of ss-Fl-DNA to ss-DNA-SH. (*b*) Lower panel shows the sample configuration. Adapted from Ref. 30.

slow cooling, which may be due to loss of capture DNA from the silver surfaces.

Figure 32 shows the frequency-domain intensity decays of the single-stranded fluorescein-labeled oligomer in solution and the double-stranded oligomer when bound to silver particles. The lifetime is dramatically shortened for the silver-bound oligomer, which strongly supports our conclusion that the intensity increase is due to localization of the fluorophore near the silver surfaces. In control experiments, we found that emissions of fluorescein in the single- and double-stranded oligos were similar to within 10 percent. The double-stranded form displayed an approximate 10 percent smaller intensity.

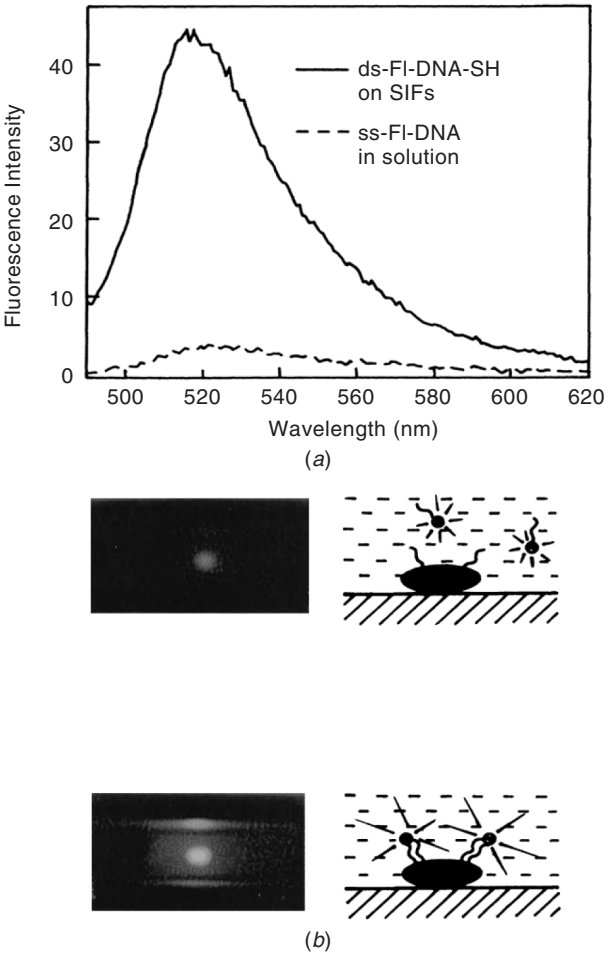


**Figure 30** (*a*) Emission spectra of ss-Fl-DNA in solution (dashed line) and bound (solid line) to silver particles. Roughly the same number of molecules of ss-Fl-DNA and ds-Fl-DNA-SH was in the illuminated area. (*b*) The lower panel shows photographs of ss-Fl-DNA in solution (top) and ds-Fl-DNA-SH on SiFs (bottom). Adapted from Ref. 30.

Hence the differences in intensity and lifetime between the solution and silver-bound forms are not due to effects of hybridization on the fluorescein probe. It is interesting to note that there is no detectable 4-ns component for the sample with silver particles, indicating that all the emission is due to the silver-bound DNA.

The differences in lifetime of fluorescein between the solution and silver-bound form suggested an alternative approach to measuring hybridization. The emission phase angle and modulation measurements could be useful in

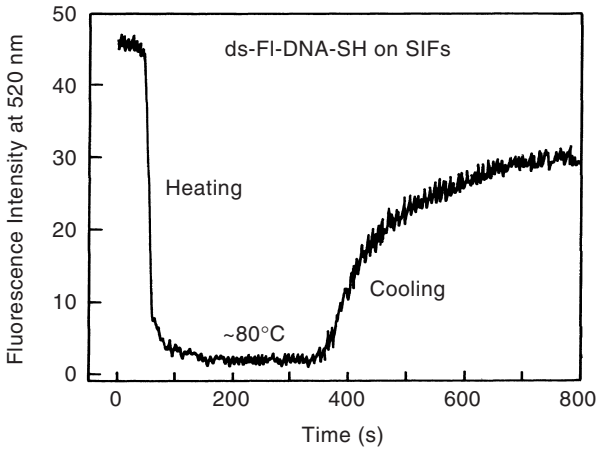


**Figure 31** Thermal melting of ds-Fl-DNA-SH on a SiFs near 80°C and rehybridization of Fl-DNA-SH on SiFs. Adapted from Ref. 30.

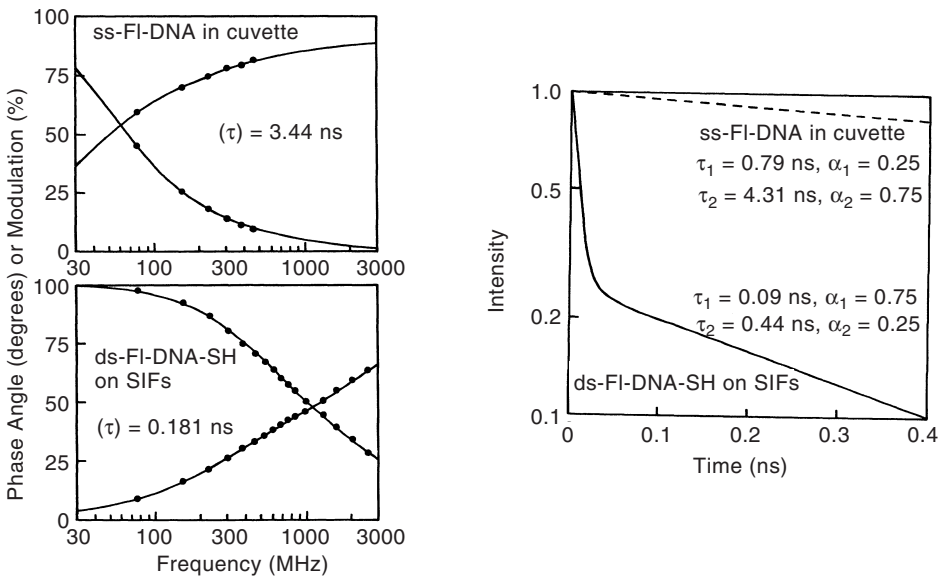


**Figure 32** Intensity decays of ss-Fl-DNA in solution and ds-Fl-DNA-SH on SiFs measured in (*a*) the frequency-domain and (*b*) reconstructed in the time domain. Adapted from Ref. 30.

detecting DNA hybridization; since these values depend on the fluorescence lifetime, they were expected to change upon hybridization. These measurements revealed a rapid decrease in phase angle and increase in modulation following addition of ss-Fl-DNA to the silver-bound capture oligomer (Fig. 33). These changes are due to the decrease in lifetime upon binding to the capture oligomers on the silver particles. The changes in phase and modulation (Fig. 33) occur somewhat more rapidly than the change in intensity (Fig. 29). This difference occurs because the phase and modulation are intensity-weighted parameters. It is important to note that phase angle and modulation measurements are mostly independent of total intensity, to within the

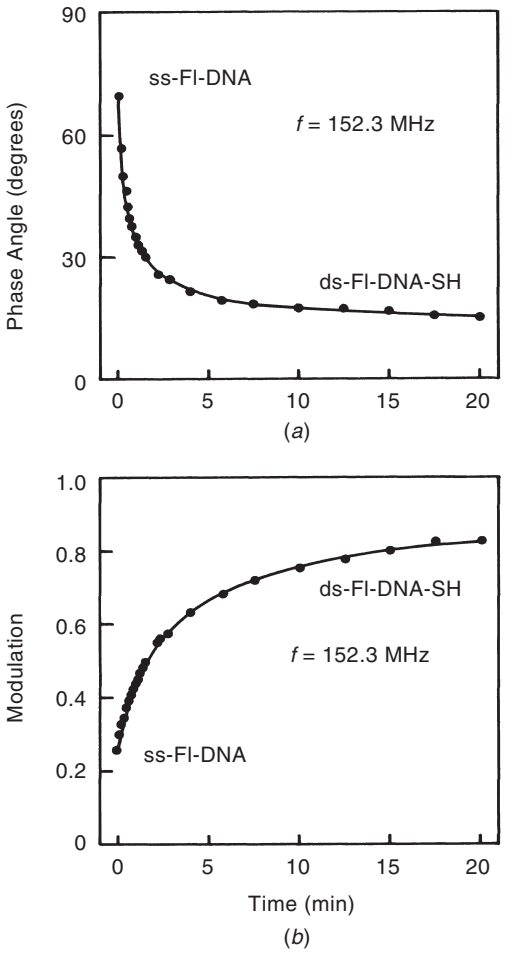


**Figure 33** (*a*) Phase angle and (*b*) modulation of ss-Fl-DNA upon hybridization to silver-bound ss-DNA-SH. The light modulation frequency and detection frequency were 152.3 MHz. Adapted from Ref. 30.

limitations of the instrumentation and the extent of background fluorescence in the sample [31]. This is important because the extent of hybridization can be measured using optical components such as optical fibers where the intensity may vary, or in microwell plates, where the well-to-well intensity may vary due to the plate or adsorbing species in the sample.

We examined the intensity of Fl-DNA-SH on the silver surfaces with continuous illumination (Fig. 34). As a comparison we examined a similar biotinylated oligomer bound to a silvered quartz slide using biotinylated bovine serum albumin and avidin. We found that fluorescein in our samples deposited on silver particles photobleaches more slowly than when deposited on a protein monolayer. It should be noted that with the protein monolayer the Fl-DNA was uniformly deposited on the entire silvered slide, that is, also between silver particles. The lifetimes measured for the sample on the protein monolayer contained a long, ≈4-ns component. This component indicates that there is some fluorescein bound to the glass surface between the silver particles.

## Enhanced DNA Labels

We also investigated the effects of SiFs on the emission spectral properties of the fluorophore-labeled DNA. Emission spectra of Cy3-DNA and Cy5-DNA on glass and on SiFs are shown in Figure 35. The emission intensity is increased two- to threefold between SiFs as compared to between the quartz slides for Cy3-DNA and Cy5-DNA, respectively. The slightly larger increase in emission intensity for Cy5-DNA compared to Cy3-DNA is consistent with the results

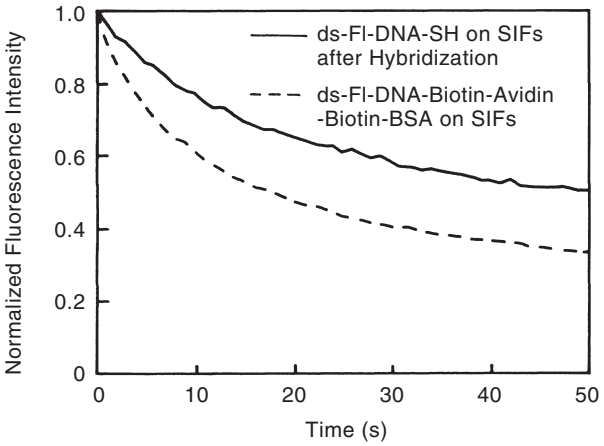


**Figure 34** Photostability of ds-Fl-DNA-SH bound to SiFs (solid line). For comparison we show the photostability of a similar biotinylated oligomer bound to a protein monolayer of BSA–biotin–avidin uniformly deposited on SiFs (dashed line). In this case some of Fl-DNA–biotin are distant from and not affected by the silver particles. Adapted from Ref. 30.

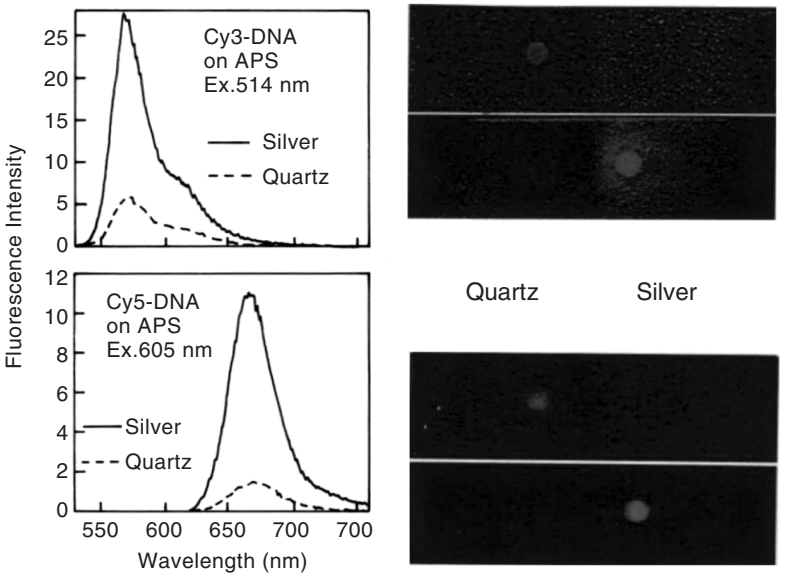


**Figure 35** Emission spectra of Cy3-DNA (*top left*) and Cy5-DNA (*bottom left*) between quartz plates with and without SiFs. Photographs of the corresponding fluorophores. The emission from Cy3 on quartz (left side of picture top right) is very weak, as compared to silver side (right side of the picture). The emission from Cy5 on quartz (left side of the picture bottom right) is very weak, as compared to silver side (right side of the picture). Clearly, the presence of silver results in dramatic increases in observed fluorescence intensities. Adapted from Ref. 32.

where larger enhancements are observed with low quantum yield fluorophores. Figure 35 also shows the photographs of the labeled oligomers on quartz and on SiFs. The emission from the labeled-DNA on quartz is almost invisible but is brightly visible on the SiFs. This difference in intensity is due to an increase in the photonic mode density near the fluorophore, which in turn results in an increase in the radiative decay rate and quantum yield of the fluorophores. We note that the photographs are taken through emission filters, and the increase in emission intensity is not due to an increased excitation scatter from the silvered plates [32].

Photostability of the labeled DNA was studied by measuring the emission intensity during continuous illumination at a laser power of 20 mW (Fig. 36). The intensity initially dropped rapidly, but became more constant at longer illumination times. Although not a quantitative result, examination of these plots visually suggests slower photobleaching at longer times in the presence of silver particles compared with quartz slides.

Another widely used fluorophore in DNA detection is 1,19-[1,3-propanediylbis [(dimethylimino)-3,1-propanediyl]]bis[4-[(3-methyl-2(3H)-

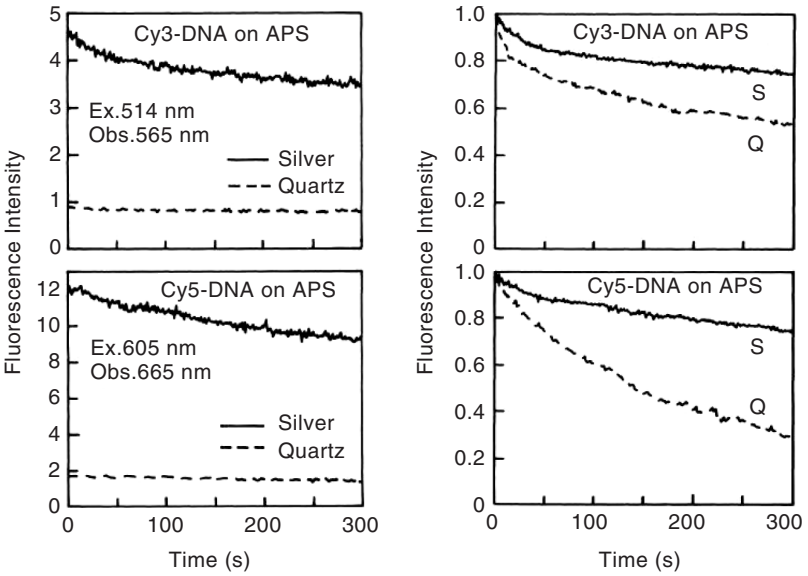


**Figure 36** Photostability of Cy3-DNA and Cy5-DNA between quartz plates with (S) and without (Q) silver island films. The laser power was 20 mW. Adapted from Ref. 32.

benzoxazolylidene) methyl]]-quinolinum tetraiodide (YOYO-1) [33–35]. This fluorophore contains multiple positive charges, binds strongly to DNA, and displays almost no fluorescence in water. As a result, it is useful for the observation of DNA in electrophoretic gels. We measured the emission spectra of YOYO-1–DNA when bound to quartz or silver via the protein layers (Fig. 37). The emission intensity is 15-fold higher on the SiFs than on quartz [36]. This dramatic intensity increase can be seen in the upper panels, which are photographs of equal amounts of YOYO-1–DNA spotted on a slide.

## Over-labeled Proteins as Ultrabright Probes

Fluorescein is one of the most widely used probes in immunoassays or immunostaining of biological specimens with specific antibodies. However, fluorescein is prone to self-quenching, which is due to Forster resonance energy transfer between nearby fluorescein molecules (homo-transfer) [37]. As a result, the intensity of labeled protein does not increase with higher increased extents of labeling, but actually decreases. Figure 38 shows the spectral properties of FITC–HSA with molar labeling ratios ($L$) ranging from 1-to-1 ($L = 1$) to 1-to-9 ($L = 9$) (the samples had the same optical density at 490 nm). The relative intensity decreased progressively with increased labeling. The insert in Figure 38 shows the intensities normalized to the same amount of protein, so that the relative fluorescein concentration increases ninefold along

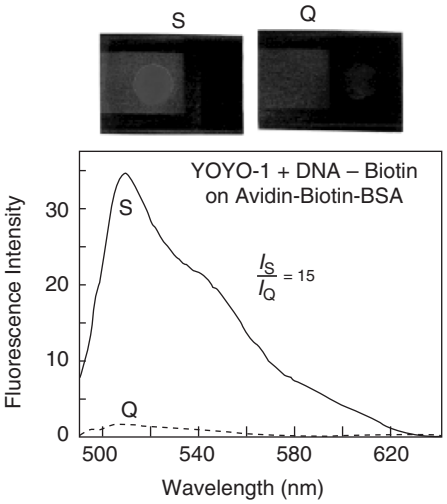


**Figure 37** Emission spectra of YOYO-1-labeled DNA bound to the quartz (Q) and silver (S) surfaces. The upper panels show a real-color photograph of labeled DNA spotted on the silver (left) and quartz (right) surfaces. Adapted from Ref. 36.

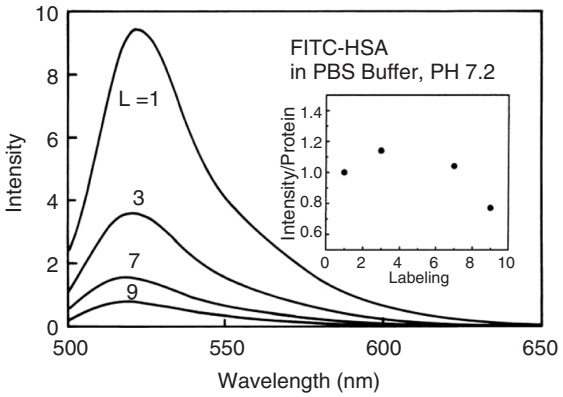


**Figure 38** Dependence of emission intensity FITC–HSA on the degree of labeling. Adapted from Ref. 38.

the $x$ axis. It is important to note that the intensity per labeled protein molecule does not increase and in fact decreases, as the labeling ratio is increased from 1 to 9.

However, we found that the self-quenching could be largely eliminated by the close proximity to SiFs [38], as can be seen from the emission spectra for labeling ratios of 1 and 7 (Fig. 39), and from the dependence of the intensity on the extent of labeling (Fig. 40). We speculate that

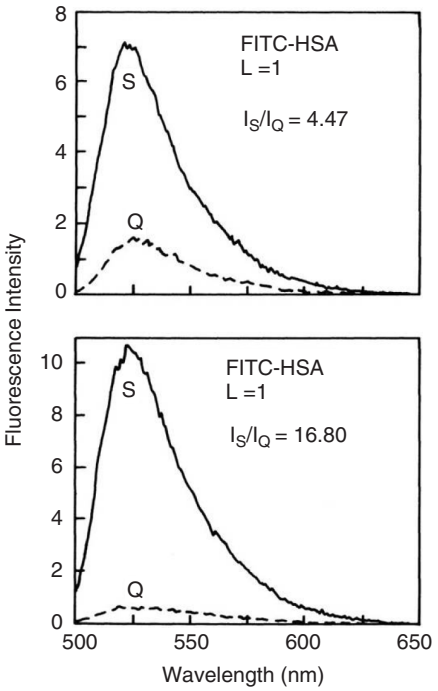


**Figure 39** Emission spectra of FITC–HSA with different degrees of labeling on quartz (Q) and on SiFs (S). Silver island films *release* the quenched fluorescence of overlabeled proteins. Adapted from Ref. 38.

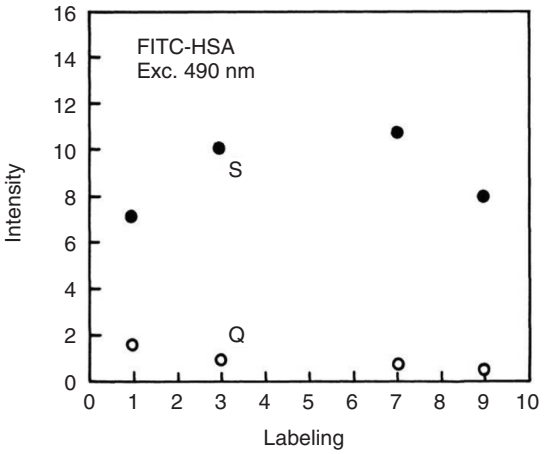


**Figure 40** Emission intensity of FITC–HSA at 520 nm vs. different degrees of labeling on quartz and on SiFs (S). Adapted from Ref. 38.

the *decrease* in self-quenching is due to an increase in the rate of radiative decay, $\Gamma_m$.

It is informative to consider how silver-enhanced fluorescence, particularly of a heavily labeled sample, can be used for improved assays and sensing. Figure 41 shows emission spectra of a quartz plate coated with FITC–HSA to which we adjusted the concentration of Rhodamine B (0.25 μM) to result in an approximate 1.5-fold larger Rhodamine B intensity. In this example, one can consider the Rhodamine B to be simple autofluorescence or any other interference signal. When the same conditions are used for FITC–HSA on silver with $L = 1$, the fluorescein emission is now 2- to 3-fold higher than that of Rhodamine B (Fig. 41*b*). When using the heavily labeled sample ($L = 7$), the fluorescein emission becomes more dominant (Fig. 41*c*). This suggests that the use of overlabeled proteins increases signal intensity and suppresses the unwanted background autofluorescence.

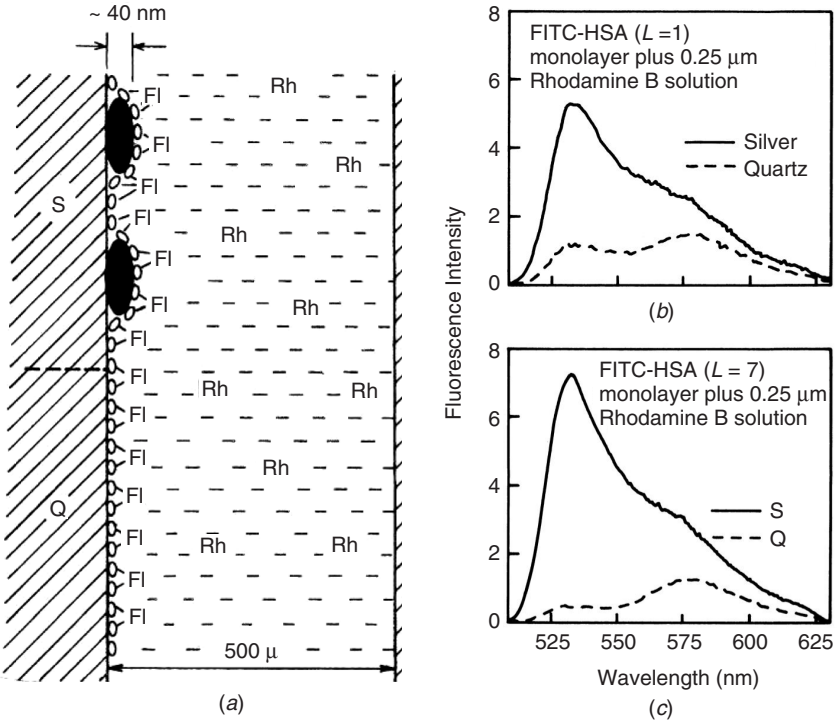


**Figure 41** Emission spectra of a monolayer of (*b*) FITC–HSA $L = 1$ and (*c*) $L = 7$ containing 0.25 mM Rhodamine B between the quartz plates (Q) or one SIF (S). (*a*) Schematic of the sample with bound fluorescein and free Rhodamine B. Adapted from Ref. 38.

## Enhanced and Selective Excitation Using Multiphoton Excitation and Metallic Nanoparticles

For multiphoton excitation, most of the excitation occurs at the focal point of the excitation, where the local intensity is the highest. We reasoned that this property might provide us with the opportunity to obtain enhanced fluorescence from fluorophores on the localized regions of our silver island films.

Recently, we have reported our findings on the *enhanced* and *localized* multiphoton excitation of Rhodamine B (RhB) fluorescence near metallic silver islands [39]. Figure 42 shows the emission spectra of $10^{-4}$ M RhB between silver island films with 2-photon excitation at 852 nm. The increase in fluorescence emission intensity for RhB molecules adjacent to metallic silver islands

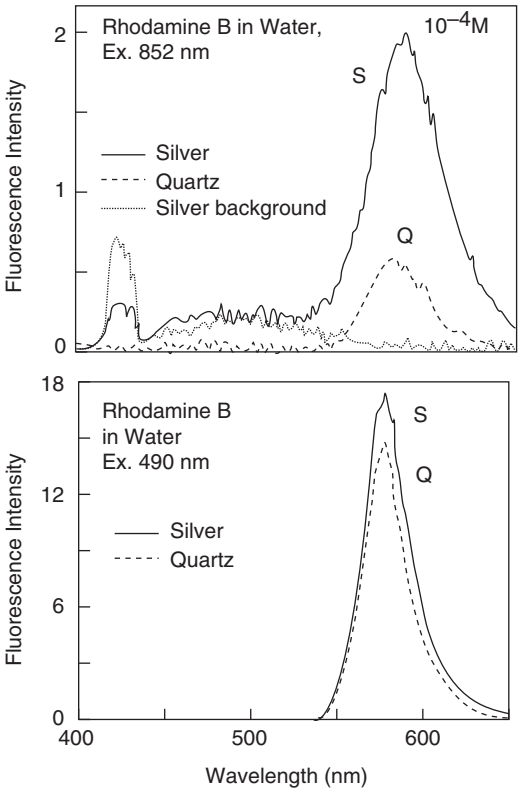


**Figure 42** (*a*) Emission spectra of $10^{-4}$ M RhB between silver island films (S) with 2-photon excitation at 852 nm (2) from a Tsunami mode-locked Ti : sapphire laser, 80 MHz repetition rate, 90 fs pulse and about 0.5 W average power. Also shown are the emission spectra observed from uncoated quartz slides (Q), and silver islands alone without RhB. (*b*) RhB between silver islands, S(-), or quartz plates, Q(. . .), with 1-photon excitation at 490 nm. Adapted from Ref. 39.

is accompanied by a reduction in lifetime, compared to that observed using 1-photon excitation.

Given the high quantum yield of RhB ($Q_0$ = 0.48), these results can be explained by the metallic particles significantly increasing the $E_m$ around the RhB molecules. Moreover, given the sample geometry and the absence of any notable increase in emission intensity using 1-photon excitation, as well the fact that the 1-photon mean lifetime remained essentially unchanged both in the presence and absence of silver, suggests that enhanced 2-photon excitation is *localized* to regions in close proximity to the silver islands. That is, a much more dramatic enhancement is possible for multiphoton excitation.

For a 2-photon absorption process the rate of excitation is proportional to the square of the incident intensity. This suggests that 2-photon excitation could be enhanced by a factor of $3.8 \times 10^8$. Such an enhancement in the excitation rate is thought to provide selective excitation of fluorophores *near to* metal islands or colloids, even if the solution contains a considerable concentration of other fluorophores that could undergo 2-photon excitation at the same wavelength but are more distant from the metals surface (Fig. 43). This interpretation is borne out by the fact that given the overwhelming excess of high quantum yield RhB in this sample geometry (96 percent of solution is too distant for *any* fluorophore–metal effect), the fluorescence lifetime is still shorter than that typically observed for bulk solution RhB, in the absence of metal. Also, for our samples we found that for 1-photon excitation, the photostability of RhB was not affected by the presence or absence of silver islands (Fig. 44*b*). However, for 2-photon excitation, an increased photostability was observed for RhB in the presence of silver islands (Fig. 44*a*). These results are consistent with shorter lifetime observed for RhB between silver islands and with our interpretation that multiphoton excitation occurs preferentially near to the silver islands.

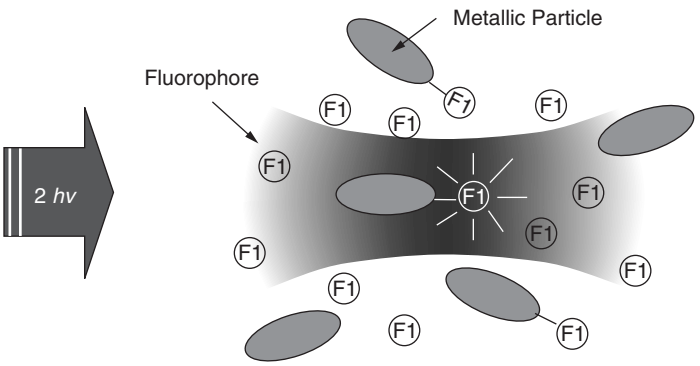


**Figure 43** Pictorial representation of the preferential multiphoton excitation of fluorophores in close proximity to metal, in the presence of free fluorophore, Fl. Adapted from Ref. 39.

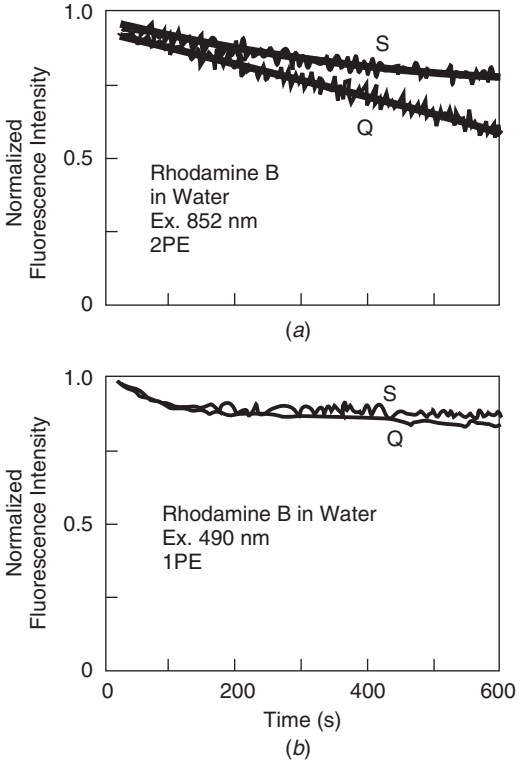


**Figure 44** (*a*) Photostability of RhB between quartz slides (Q), and silver island films (S), with 2-photon excitation at 852 nm, and (*b*) with 1-photon excitation at 490 nm. The 490-nm excitation was from an argon ion laser attenuated to about 10 mW. Fluorescence was observed at 580 nm. Adapted from Ref. 39.

## Enhanced Energy Transfer on Silver Surfaces

In a recent report, we have described the effects of metallic particles on fluorescence resonance energy transfer (RET) between donors and acceptors covalently bound to DNA [40]. Theoretical studies have predicted increased rates of energy transfer over distances as large as 700 Å near silver particles of appropriate size and shape [41, 42]. To test this prediction, we prepared a double helical DNA oligomer, 23 base pairs long, with a donor and acceptor placed at opposite sides about 75 Å apart. Because the Forster distance ($R_0$) value is near 50 Å, little energy transfer is expected under free-space conditions. We used steady-state and time-resolved fluorescence to determine the effects of silver island films on RET between the widely spaced donor–acceptor pairs. We examined the donor-labeled and acceptor- labeled oligomers separately, in the absence of RET, to determine the effects of the silver particles. Figure 45*a* shows the emission spectra of 7-amino-3-((((propyl)amino)car-

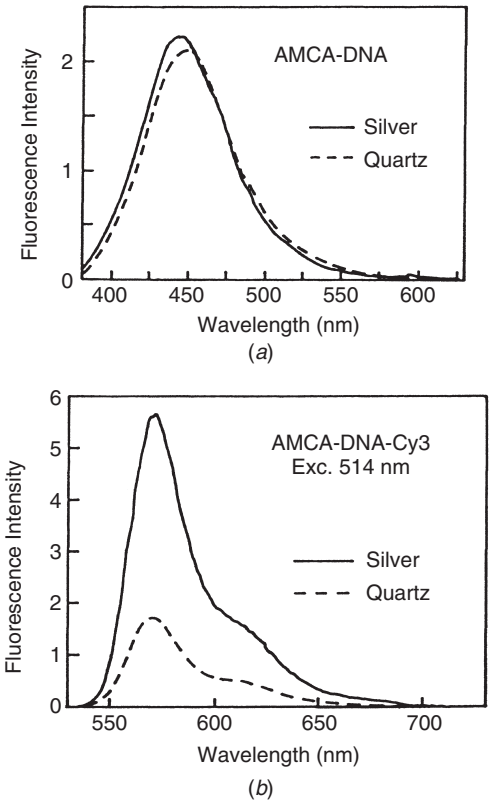


**Figure 45** Emission spectra of (*a*) the donor AMCA-DNA between quartz plates (– – –) and between silver island films (— — —). (*b*) Emission spectra of the directly excited acceptor in AMCA-DNA-Cy3 between quartz plates (– – –) and between silver island films (— — —). Adapted from Ref. 40.

bonyl)methyl)-4-methyl coumarin-6-sulfonic acid (AMCA)-DNA between uncoated quartz plates and between silver island films. Figure 45*b* also shows the emission spectra of the donor–acceptor pair AMCA-DNA-Cy3 with no excitation of the donor and direct excitation of the acceptor at 514 nm. The intensity of the donor-alone AMCA-DNA was essentially unchanged by the silver particles. In contrast the intensity of DNA-Cy3 was increased several-fold. The different effects are consistent with the effects expected for high and low quantum yield fluorophores. It is not possible to increase a quantum yield above unity, and the larger increases in quantum yield (observed fluorescence intensity) are obtained for lower quantum yield fluorophores.

We subsequently examined the emission spectra of the donor- and acceptor-labeled DNA between quartz plates and silver island films (Fig. 46). An increase in the acceptor emission is seen in the donor-normalized spectra (Fig. 46*c*). It is difficult to judge the extent to which the increased acceptor emis-

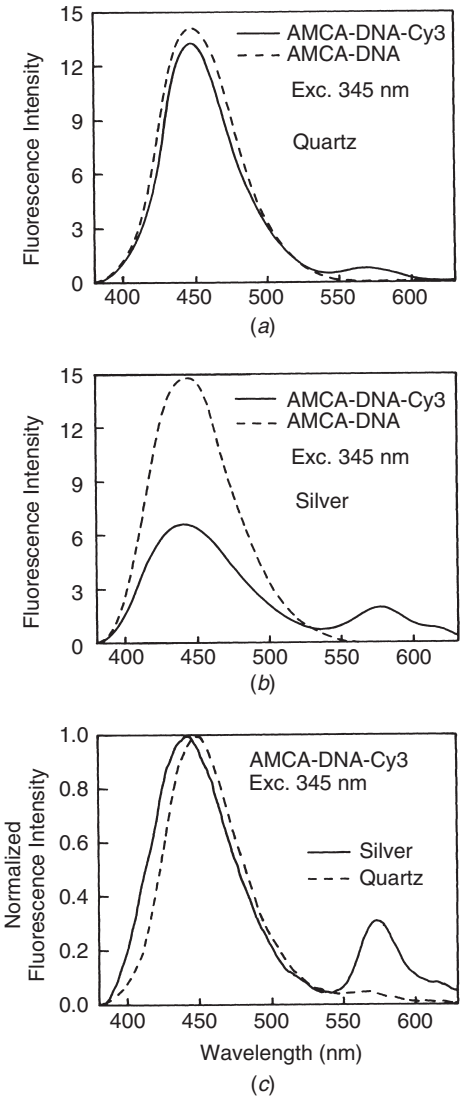


**Figure 46** Emission spectra of AMCA-DNA donor and AMCA-DNA-Cy3 the donor–acceptor pair on (*a*) quartz and (*b*) silver. (*c*) Shows normalized emission spectra of AMCA-DNACy3 donor–acceptor between quartz (– –) and between silver island films (—). Adapted from Ref. 40.

sion is due to increased energy transfer or due to the increase intensity of Cy3 shown in Figure 45. Nonetheless, it was clear that much of the Cy3 emission from AMCA-DNA-Cy3 between the silver islands is due to RET.

Our analysis of the frequency-domain intensity decays of the AMCA-DNA-Cy3 suggested that $R_0$ contributed to the 30 percent of the donor emis-

sion (not shown). The 30 percent fraction of the strongly affected donor–acceptor pairs is in disagreement with the estimated 4 percent volume of the sample, which is close to these silver particles. To clarify this discrepancy we repeated our experiments with different sample geometry, using one SiFs instead of two SiFs. That is, one side of the sandwich was a SiFs and the other an unsilvered quartz plate. This was accomplished by coating different slides with one-third or two-thirds of the area with silver, so that there were regions of the sample between two quartz plates, between one quartz and one SiFs, or between two silver island films (Fig. 47*a*). When there was only one SiFs, we found *no* increase in energy transfer (Fig. 47) and *no* decrease in the donor lifetime.

It is difficult to anticipate the future uses of long-range resonance energy transfer because all present RET assays have been designed to position the donors and acceptors within the upper range of Forster distances near 50 Å. One possibility for metal-enhanced RET is detection of target sequences with larger numbers of base pairs (Fig. 48). Shorter D–A distances can be detected between quartz and larger distances between two silvered plates. One can also imagine the use of induced long-range RET for analysis of chromosomes with fluorescence in situ hybridization (FISH). As currently performed, the emission spectra of the FISH samples reflects the location of specific sequences,

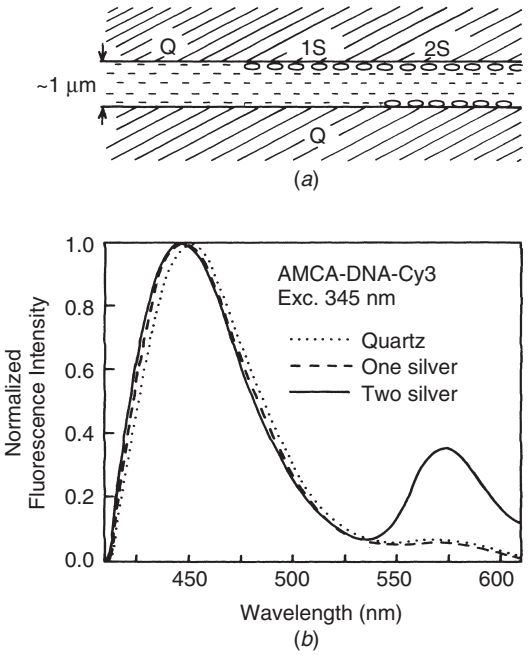


**Figure 47** Normalized emission spectra of AMCA-DNA-Cy3 recorded on quartz (Q), one silvered slide (1 S), and two silvered slides (2 S). Adapted from Ref. 40.

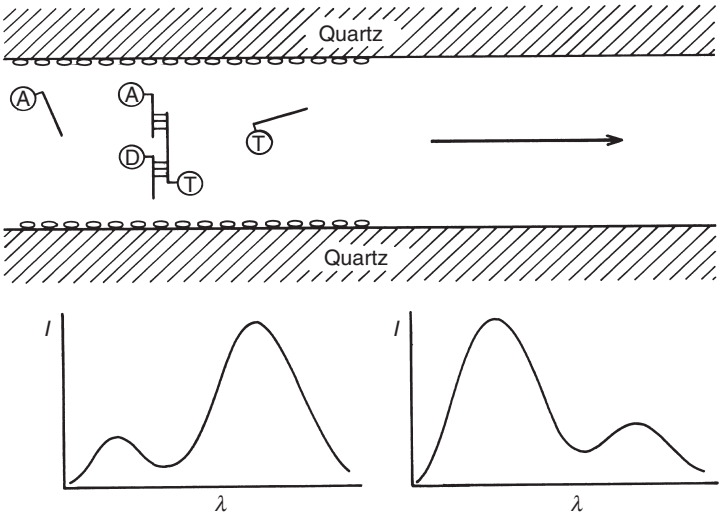


**Figure 48** Detection of DNA sequences using long-range RET on a silver particle surface. D, A, and T are the donor, acceptor, and target nucleotides, respectively.

and RET does not normally occur between the labeled oligonucleotides used in these hybridizations. This situation, however, may change for labeled chromosomes; especially if placing a solution in a microcavity-type system results in RET over hundreds of angstroms. Further studies by our laboratories are under way.

## Ratiometric Sensing on Surfaces

We have also developed new methodologies for attaching well-known pH probes (Dextran-SNAFL-2, Molecular Probes) onto silver island films. We hypothesized that SiFs could increase the sensitivity of these probes based on our observations with other fluorophores. We have developed two different methods to immobilize pH probes on to glass substrates (for control samples) and silver island films.

Immobilization of dextran-SNAFL-2 onto glass substrates is done by the following procedure: Precleaned glass substrates were soaked in a 2 percent v/v solution of 3-(glycidoxypropyl) trimethoxysilane (GOPTES) in ethanol for 2 h. Then, the GOPTES-coated glass substrates were removed from the solution and rinsed several times with ethanol. In order to remove unbound deposited materials, the substrates were additionally placed in ethanol in the ultrasonic bath for 30 sec. This procedure was repeated at least four more times. After drying the substrates in a stream of dry nitrogen, they were placed in a basic solution of dextran/SNAFL-2 conjugate and incubated for 20 h. Then, the glass substrates were removed from dextran solution and rinsed with deionized water, and dried in a stream of dry nitrogen (Fig. 49).

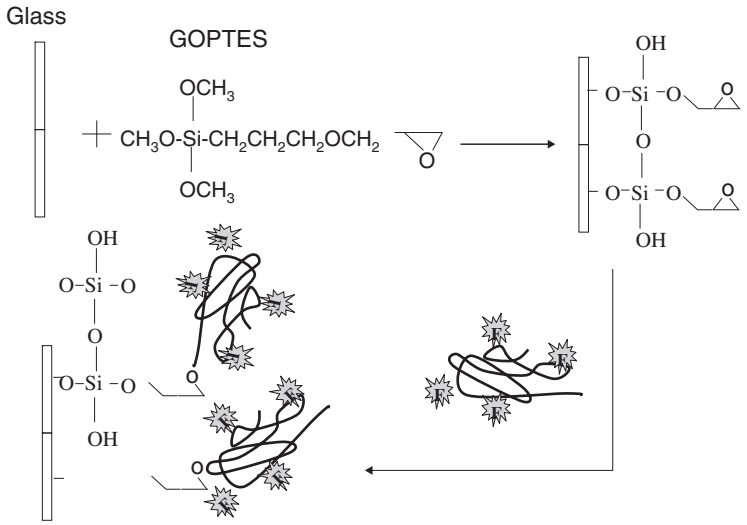


**Figure 49** Experimental procedure for immobilization of pH and Ca probes onto glass slides.

Immobilization of dextran-SNAFL-2 onto SiFs is done by the following procedure as depicted in Figure 50. SiFs were prepared according to the procedure as reported in the literature. Then, the SiFs were soaked in a 2 percent v/v solution of 3-(aminopropyl) triethoxysilane (APS) in ethanol for 2 h. The APS-coated SiFs were removed from the solution and rinsed several times with ethanol to remove the unbound APS. APS-coated SiFs were subsequently soaked in a 2 percent v/v solution of GOPTES in ethanol for 2 h. The unbound GOPTES were removed by rinsing the substrates with ethanol several times. After drying the substrates in a stream of dry nitrogen, they were placed in a basic solution of dextran/SNAFL-2 conjugate and incubated for 20 h. The glass substrates were then removed from the dextran solution and rinsed with deionized water and then dried in a stream of dry nitrogen.

SNAFLs (seminaftofluoresceins) and SNARFs (seminaftorhodafluors) are classes of pH probes with wavelength ratiometric pH sensing capabilities. The acid form of the probe displays a spectrum with an emission peak at 552 nm and base form at 635 nm (Fig. 51). Large spectral shifts between both forms of the probe readily allows for convenient pH sensing using the intensity ratio at two emission wavelengths.

We have thus measured fluorescence emission intensity of dextran-SNAFL-2, which is immobilized onto glass and SiFs, and we determined the enhancement factor for both the acid form (pH 6.8) and the base form (pH 8.9). The acid form of SNAFL-2 displayed a larger enhancement (average of 43.4) than that of the base form (average of 31.1). One explanation of this observation is that the acid form of SNAFL-2 (which is protonated in excited state) inter-

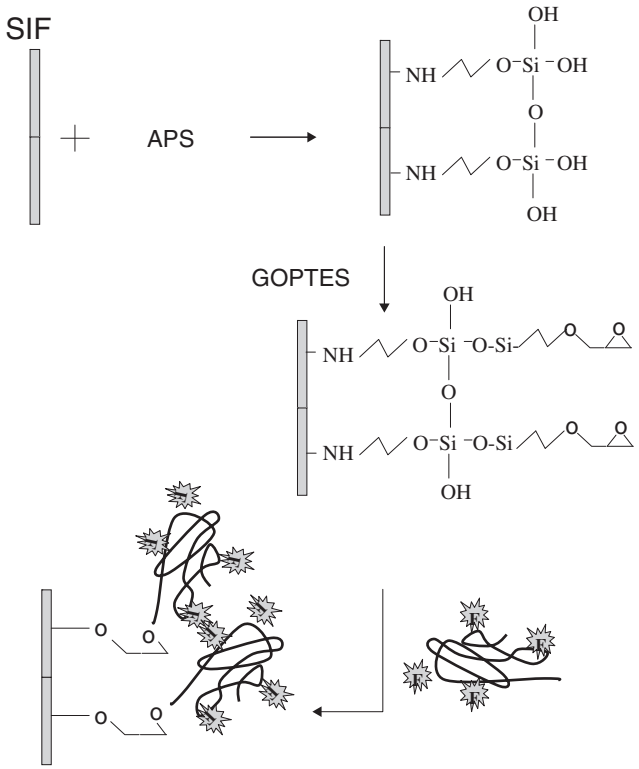


**Figure 50** Experimental procedure for immobilization of pH and Ca probes onto SiFs.

acts more strongly with the silver nanoparticles, as its energy is closer to a nanoparticle resonance (absorbance peak for SiFs is about 440 nm), than that of base form.

We have determined that pH sensitivity of SNAFL-2 is retained, if not better, on SiFs (Fig. 52), which is a similar ratiometric response obtained from both when SNAFL-2 is in solution or immobilized onto SiFs. Figure 53 shows the photostability of SNAFL-2 immobilized onto glass and SiFs. When SNAFL-2 is immobilized onto the glass surface, it photobleaches quickly, but photobleaches substantially slower on SiFs. High-intensity enhancement and improved photostability are desired features of probes for chemical sensing and cellular imaging, allowing the use of reduced probe concentrations (less toxicity, less problem with probe loading), or the use of a lower excitation intensity (less damage to cells). Additionally, the short data acquisition times provides for less exposure damage to cells. Another advantage of a higher sample intensity is the ability to use low-power excitation sources such as LEDs (light-emitting diodes) or laser diodes, substantially reducing any instrumentation costs.

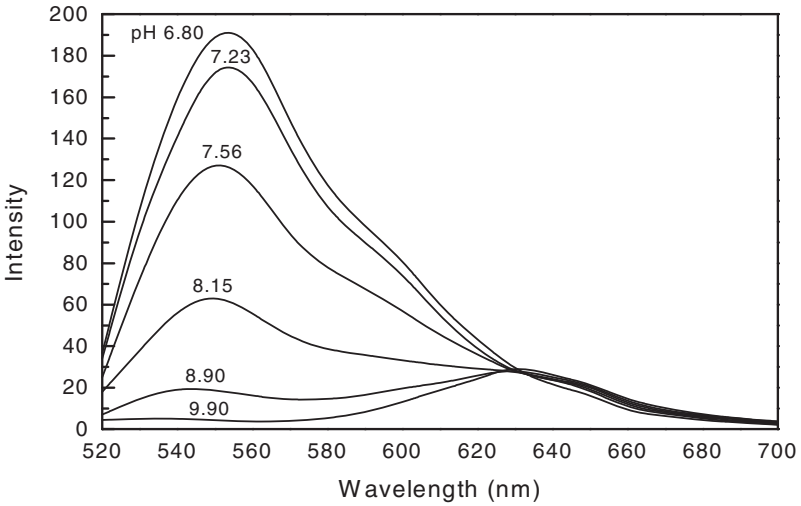


**Figure 51** Emission spectra of SNAFL 2-Dextran in various pH solutions (PBS buffer) with 130 mM KCl. The excitation wavelength was 488 nm and the temperature was 23°C.

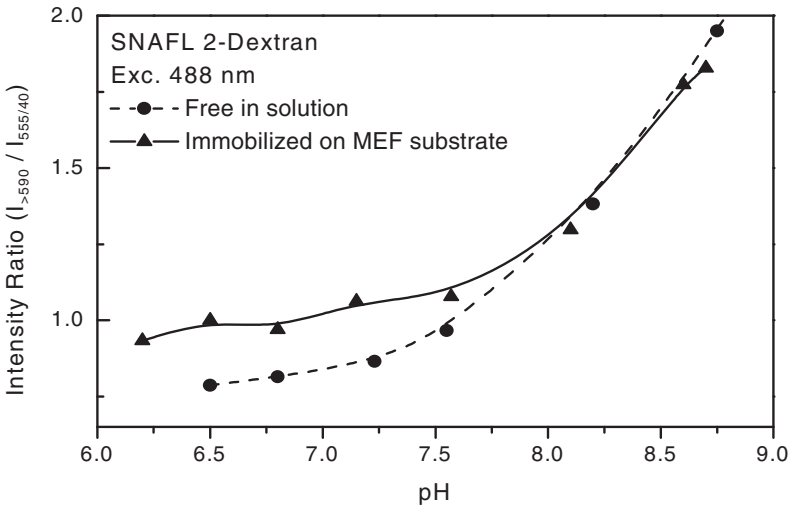


**Figure 52** pH sensitivity of SNAFL-2 is retained in the presence of silver nanoparticles. The intensity on MEF substrate is about 30 to 40 times brighter than intensity of probe immobilized on bare glass slide, demonstrating enhanced ratiometric sensitivity.

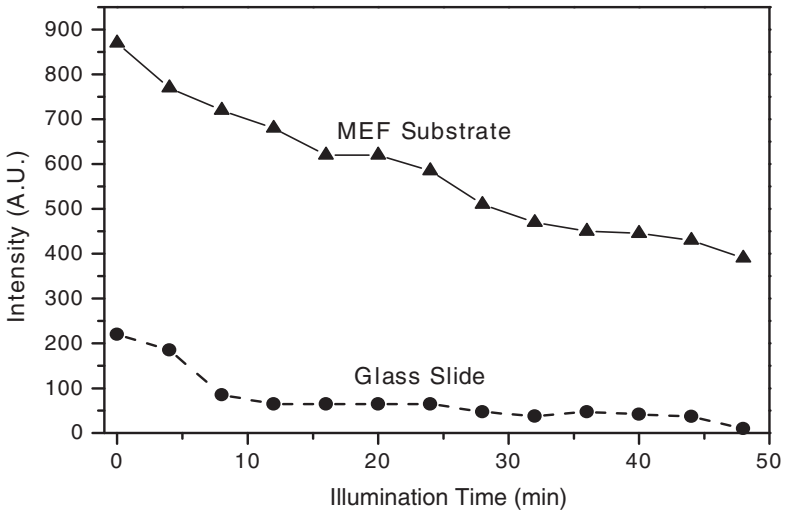


**Figure 53** Photostability of SNAFL 2-Dextran (pH 6.3) immobilized on the MEF substrate and glass slide. The illumination conditions were the same for each slide: Green LED (535 nm, 0.5 mW), microscope objective 20× and CCD camera was used as the intensity detector.

## MEF Solution-Based Assays

In order to illustrate the usefulness of MEF in solution-based biosensing applications, including high-throughput screening (HTS) plate wells, we have prepared silica-coated silver spheres in suspension. In this regard, silver spheres were prepared by using the following procedure: 2 mL of 1.16 mM trisodium citrate solution was added dropwise to a heated (95°C) aqueous solution of 0.65 mM of $AgNO_3$ while stirring. The mixture was kept heated for 10 min, and then it was cooled to room temperature. This procedure yields silver spheres with sizes in the range of 30 to 80 nm. The surface of the silver spheres were modified with APS in ethanol. The APS-coated silver spheres were resuspended in predetermined amount of water and $NH_4OH$. Then, a solution of tetraethylorthosilicate (TEOS) was rapidly injected, the reaction continuing for a period of time to control thickness of the $SiO_2$ coating layer. The amount of TEOS was calculated based on the total area of silver spheres and the desired shell thickness, assuming complete conversion of TEOS to silica (Fig. 54). Figure 54*b* shows a typical transmission electron microscopy (TEM) image of the silica-coated silver spheres.

In order to demonstrate that the silica-coated silver spheres can result in metal-enhanced fluorescence, we have utilized the well-known biotin–streptavidin interactions. These interactions occur in relatively fast reaction times (20 min) and one of the strongest biological interactions found in nature (dissociation constant, $K_D = 10^{-15}$ M). For this purpose, we have further modi-

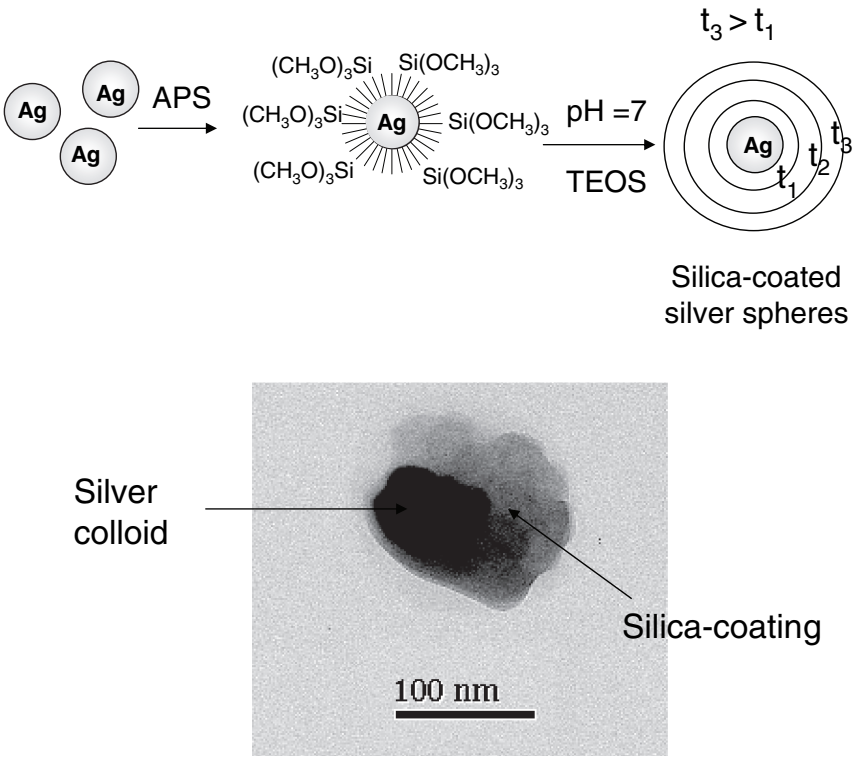


**Figure 54** (*a*) Experimental procedure for coating of silver spheres with silica, and (*b*) a representative TEM image of silica-coated silver spheres.

fied the surface of the silica-coated silver spheres with bovine serum albumin (BSA)–biotin. Then, we have aggregated the biotinylated-silver spheres with Cy3-labeled streptavidin (Fig. 55). Figure 56 shows the fluorescence emission intensities recorded from the aggregated biotinylated-silver spheres (no label) and from the nonaggregated biotinylated-silver spheres (control samples, C1 and C2), respectively. Aggregation of the labeled biotinylated-silver spheres resulted in approximately 3- to 5-fold higher fluorescence intensity than the nonaggregate system or the aggregated system with no label. Interestingly, the Cy3 lifetime was also reduced in accordance with a radiative rate modification.

## Directional Emission

We have recently reported a new method for fluorescence detection that promises to increase sensitivity by up to 1000-fold [43]. Our new method

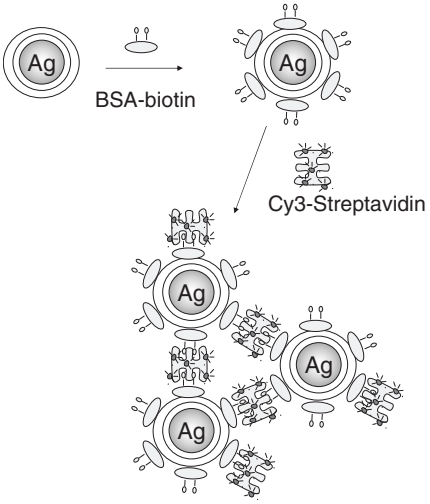


**Figure 55** Experimental procedure for coating of silver spheres with biotinylated-BSA and aggregation of silica-coated silver spheres with Cy3-labeled streptavidin.

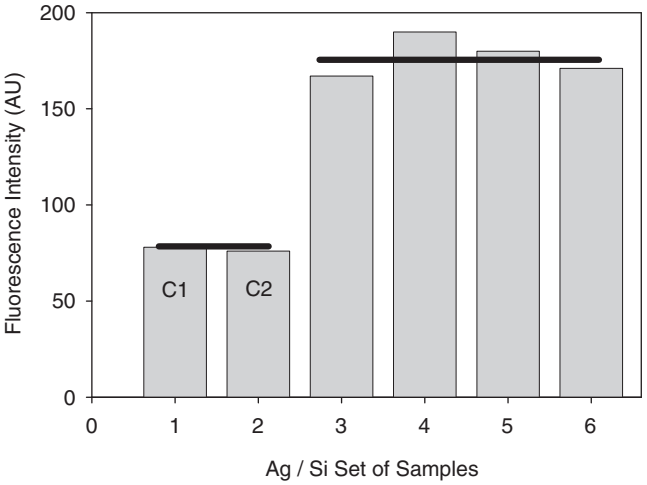


**Figure 56** Fluorescence emission intensity recorded from the aggregated silica-coated silver spheres (samples 3–6, four separate samples) and control samples (samples 1–2).

depends on the coupling of excited fluorophores with the surface plasmons present in thin metal films, typically silver, gold, or aluminum. This new phenomenon, surface plasmon-coupled emission (SPCE) occurs for fluorophores 20 to 250 nm from the metal surface, where the emission occurs over a narrow angular distribution, converting normally isotropic emission into easily col-

lected directional emission. The interaction is *independent* on the mode of excitation, that is, it does not require evanescent wave or surface plasmon excitation. With typical optical components, the collection efficiency is 1 percent or less, however, our new approach promises to couple up to 50 percent of the emission from unoriented and appropriately distanced fluorophores. We believe our new findings offer both unique and exciting possibilities for high-sensitivity fluorescence detection, as distal fluorescence (autofluorescence) from fluorophores or species from the metal surface only weakly couples. In addition SPCE is highly polarized, and autofluorescence can be further discriminated against by collecting only the polarized component or the light emanating at the appropriate angle. Further, different emission wavelengths couple at different angles allowing spectral discrimination without additional dispersive optics.

In our preliminary studies of SPCE, a continuous 50-nm-thick silver film on a glass substrate was spin coated with varying thicknesses of sulforhodamine 101 (S101) doped PVA [43]. Figure 57*a* shows the configuration that was used for SPCE studies. This combined sample was positioned on a precise rotary stage, which allows excitation and observation at any desired angle relative to the vertical axis along the cylinder. Two modes of excitation were considered in our studies (Fig. 57*b*).

We have found that the mode of excitation does not matter for SPCE [43–45]. That is, an excited fluorophore should couple with the surface plasmons whether it is excited by the surface plasmon evanescent field or directly using the reverse Kretschmann (RK) configuration. Figure 58 shows the dependence of the emission intensity on observation angle with RK excitation, noting that the incident field cannot induce surface plasmons [1]. On the prism side (back side B) of the sample, the emission is very sharply distributed at ±47° or ±50° from the normal axis for both the 15- (top) and 30-nm (bottom) films, respectively. This small difference in angle is due to the thickness of the PVA film. The intensity observed on the front (F) side of the sample was much lower and not sharply distributed at any particular angle, as illustrated in Figure 58. Higher emission intensity is observed on the front side of the thicker sample (bottom), which is consistent with lower efficiency coupling into the surface plasmon for fluorophores more distal from the metal surface.

We additionally examined the polarization of SPCE and found that the emission is polarized in the plane of incidence (*p*-polarized) irrespective of the polarization of the excitation. This polarization of the directional emission is an interesting find and proves that it is due to coupling of the surface plasmons, and that the polarization of the SPCE is independent of the polarization of the normal incidence excitation. This suggests that the emission dipoles parallel to the plane of incidence couple into the surface plasmon, and dipoles perpendicular to the plane of incidence do not result in SPCE, or at least display only very weak coupling [43, 45]. Figure 59 shows the emission for S101 in PVA observed with a hemispherical prism and RK excitation, where the hemispherical prism allows for the SPCE cone to be observed.

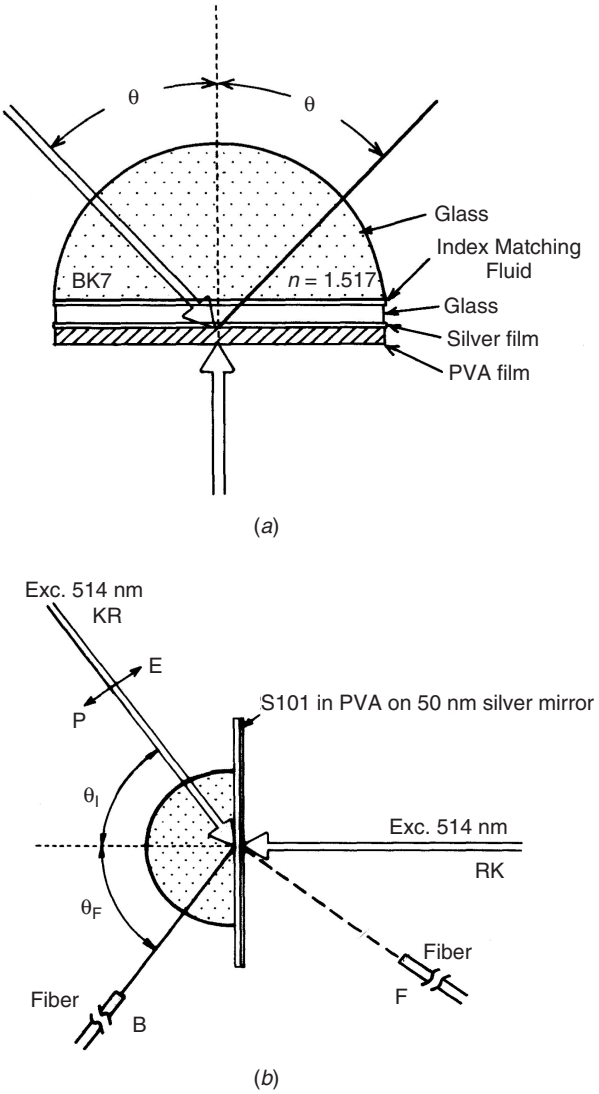


**Figure 57** (*a*) Configuration of hemicylinder and spin-coated PVA-fluorophore slide, and (*b*) experimental geometry for the measurement of free space emission (F) and SPCE (B) with the Kretschmann (KR) and reverse Kretschmann (RK) configuration. Adapted from Ref. 43.

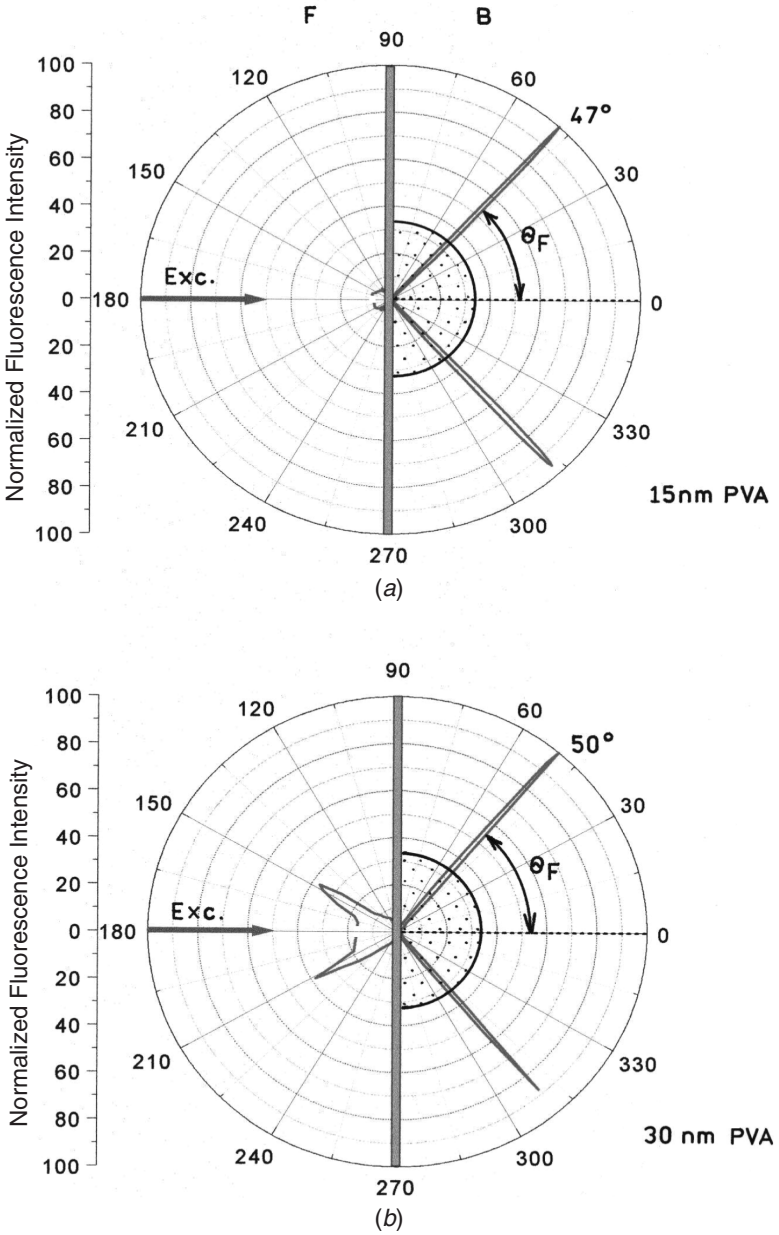


**Figure 58** Angular dependence of S101 emission excited using the reverse Kretschmann configuration. The PVA thickness was approximately (*a*) 15 nm and (*b*) 30 nm. Adapted from Ref. 43.

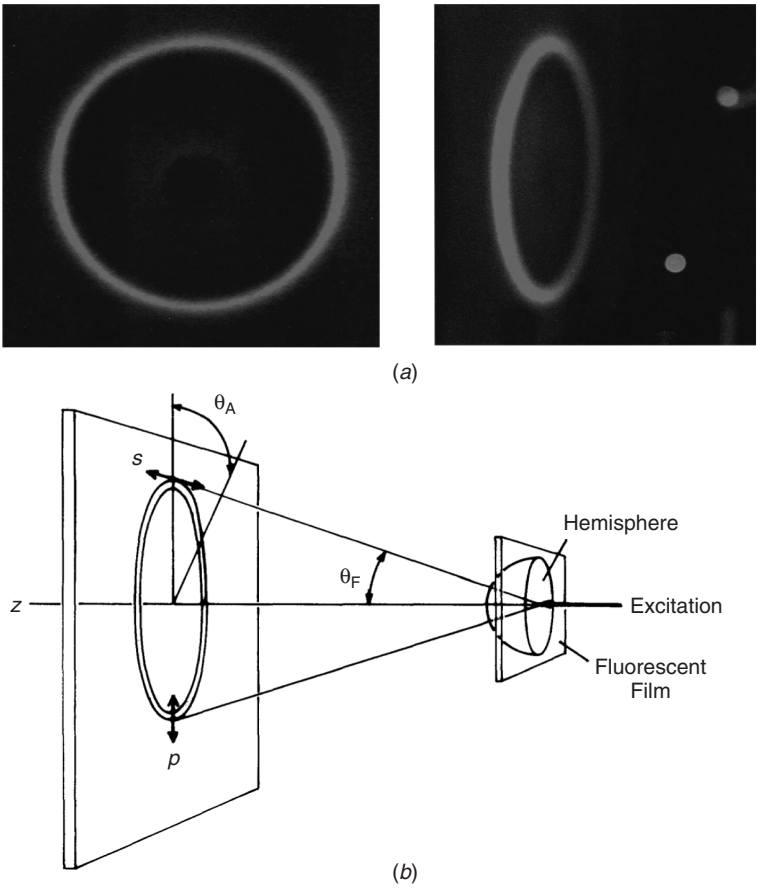


**Figure 59** (*a*) Emission for S101 in PVA observed with a hemispherical prism and RK excitation and (*b*) the cone of emission with a hemispherical prism. The emission was incident on tracing paper and photographed through a LWP 550 filter. Adapted from Ref. 43.

We also have found that SPCE strongly depends on wavelength. This suggests that fluorophores with different emission maxima will display SPCE at different angles. Subsequently, Figure 60 shows the SPCE from a mixture of fluorophores using RK excitation and a hemispherical prism, with 532-nm excitation. We recorded the emission spectra at different observation angles where the spectra are clearly distinct at each angle (not shown), with the shorter wavelengths occurring for larger angles [45]. This remarkable find suggests the potential for multifluorophore surface assays featuring the intrinsic spectral resolution of SPCE.

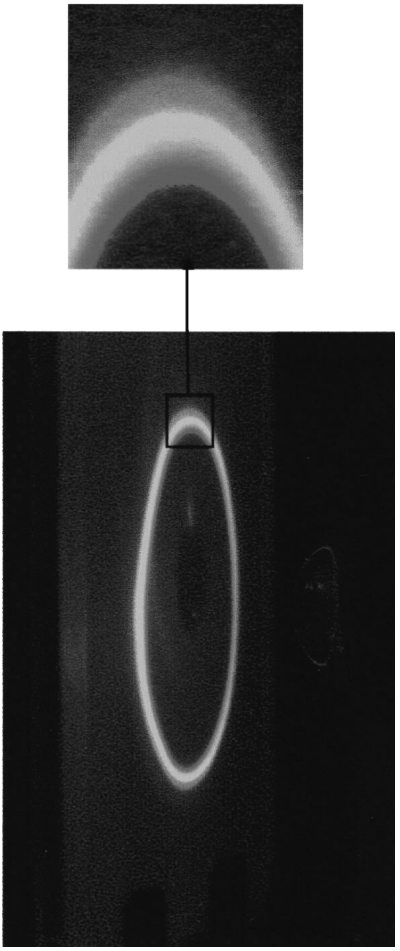

**Figure 60** Photograph of SPCE from a mixture of S101, Rhodamine 123, and Pyridine 2, using RK excitation and a hemispherical prism. Initially the fluorophore concentrations were optimized to give similar free space fluorescence intensities. Adapted from Ref. 43.

## 4 ROLE OF MEF IN HTS AND DRUG DISCOVERY

Metal-enhanced fluorescence appears to be most suitable to the fluorescence assays used in drug discovery and DNA analysis. As we have shown, silver can be readily deposited on glass or polymer substrates by a variety of methods. Silver colloids are easily prepared and can be attached to surfaces functionalized with amine or sulfhydryl groups. One can imagine the bottom of multiwell plates or DNA arrays being coated with silver particles. A variety of new assay formats are possible. Assays could be based on the lightning rod effect,

that is, $E_m$ modifications. The biochemical affinity interactions could bring the fluorophore close to the metal surface, for localized excitation, eliminating the washing steps. Another approach could be to use low quantum yield fluorophores, and the increased quantum yield of fluorophores brought in close proximity with the metal (Fig. 61). These effects might be coupled with another remarkable property of metal–fluorophore interactions. If the metal is close to a semitransparent metallic surface, the emission can couple into the metal and become directional rather than isotropic (Fig. 62), offering the possibility of a new high-sensitivity detection system for HTS wells.

Metal-enhanced fluorescence is not limited to single metallic particles. Theory predicted that the lightening rod effect is much stronger between two metallic spheres than for isolated spheres. If the biochemical affinity reaction brings particles into closer proximity, then excitation fields may be substantially increased in the spaces between particles (Fig. 63). Multiphoton excitation is also known to be increased near metallic surfaces and may be even more efficient between metal particles. Also, proximity to the particles can result in long-range energy transfer (Fig. 64), which would allow selective detection of macromolecular complexes. Scattering from corrugated surfaces

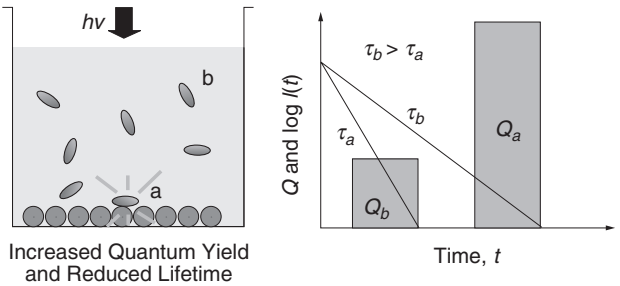


**Figure 61** Potential uses of metal-enhanced fluorescence in drug discovery based on local increases in quantum yield. Adapted from Ref. 46.

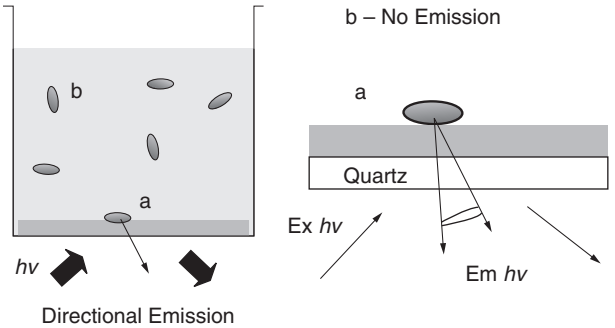


**Figure 62** Potential uses of metal-enhanced fluorescence in drug discovery based on directional emission. Adapted from Ref. 46.

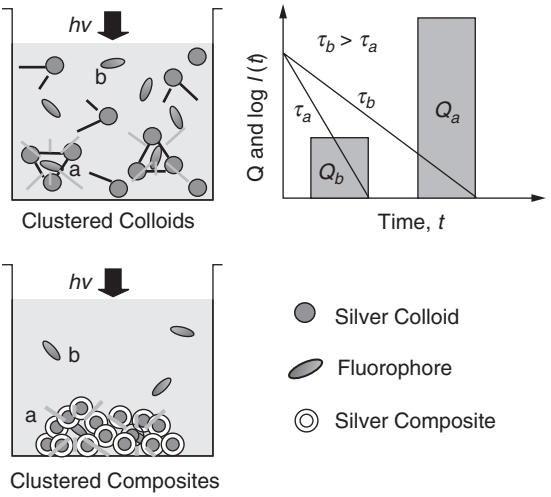


**Figure 63** Potential uses of metal-enhanced fluorescence based on colloid clustering. Adapted from Ref. 46.

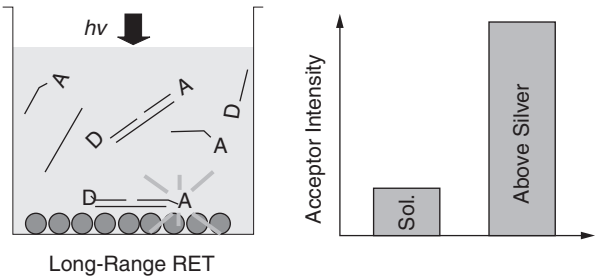


**Figure 64** Potential uses of metal-enhanced fluorescence based or long range RET. Adapted from Ref. 46.

could also be used to increase the detection efficiency of fluorophores (Fig. 65).

## 5 CLOSING REMARKS

In this invited review chapter, we have summarized the favorable effects of fluorophores in close proximity to metallic nanostructures. These favorable effects include increased fluorescence intensities (increased quantum yields), increased probe photostability (reduced fluorescence lifetimes), and increased rates of excitation and energy transfer.

Metallic silver can be deposited by a variety of methods, depending on the sensing application, on demand if required, and using fairly simple apparatus

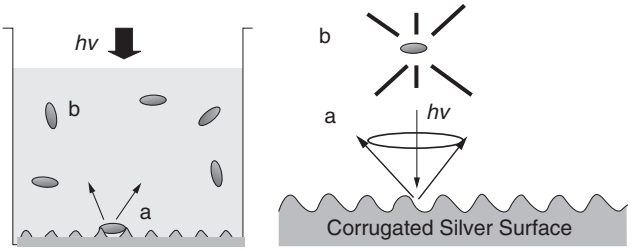


**Figure 65** Potential uses of metal-enhanced fluorescence in drug discovery based on corrugated surfaces. Adapted from Ref. 46.

and biologically inert chemicals. The metal–fluorophore effects we have observed to date offer unique perspectives in fluorescence sensing, providing for improved background suppression, increased detection limits, and localized excitation near to silver nanostructures. While our results to date are for the most part preliminary, we believe MEF will find many applications in sensing, especially in high-throughput screening and in drug discovery.

***Acknowledgments*** Work was supported by the NIH, National Center for Research Resources, RR-01889. Financial support to J. R. Lakowicz, C. D. Geddes, and I. Gryczynski from University of Maryland Biotechnology Institute (UMBI) is also acknowledged.

## ACRONYMS AND SYMBOLS

| | |
|---|---|
| $\Gamma$ | radiative decay rate in the absence of metal |
| $\Gamma_m$ | radiative decay rate in the presence of metal |
| APS | 1-amino-3-propylethoxy silane |
| $E$ | excitation rate |
| $E_m$ | metal-enhanced excitation rate |
| HAS | human serum albumin |
| ICG | indocyanine green |
| KR | Kretschmann |
| $k_m$ | quenching rate in the presence of metal |
| $k_{nr}$ | nonradiative rates |
| MEF | metal-enhanced fluorescence |
| $Q_0$ | quantum yield |
| $r$ | particle diameter |
| RET | fluorescence resonance energy transfer |
| RK | reverse Kretschmann |
| SERS | surface-enhanced raman scattering |
| SiFs | silver island films |
| SPCE | surface plasmon-coupled emission |

## REFERENCES


1. Lakowicz, J. R. (2001). Radiative decay engineering: Biophysical and biomedical applications. *Anal. Biochem.*, *298*, 1–24.
2. Lakowicz, J. R., Shen, Y., D'Auria, S., Malicka, J., Fang, J., Gryczynski, Z., Gryczynski, I. (2002). Radiative decay engineering 2. Effects of silver island films on fluorescence intensity, lifetimes, and resonance energy transfer. *Anal. Biochem.*, *301*, 261–277.
3. Lakowicz, J. R., Shen, Y., Gryczynski, Z., D'Auria, S., Gryczynski, I. (2001). Intrinsic fluorescence from DNA can be enhanced by metallic particles. *Biochem. Biophys. Res. Commun.*, *286*, 875–879.
4. Gryczynski, I., Malicka, J., Shen, Y., Gryczynski, Z., Lakowicz, J. R. (2002). Multiphoton excitation of fluorescence near metallic particles: Enhanced and localized excitation. *J. Phys. Chem. B*, *106*, 2191–2195.
5. Gersten, J., Nitzan, A. (1981). Spectroscopic properties of molecules interacting with small dielectric particles. *J. Chem. Phys.*, *75*, 1139–1152.
6. Weitz, D. A., Garoff, S., Gersten, J. I., Nitzan, A. (1983). The enhancement of Raman scattering, resonance Raman scattering and fluorescence from molecules absorbed on a rough silver surface. *J. Chem. Phys.*, *78*(9), 5324–5338.
7. Kummerlen, J., Leitner, A., Brunner, H., Aussenegg, F. R., Wokaun, A. (1983). *Mol. Phys.*, *80*, 1031.
8. Malicka, J., Gryczynski, I., Geddes, C. D., Lakowicz, J. R. (2003). Metal-enhanced emission from indocyanine green: A new approach to in vivo imaging. *J. Biomed. Opt.*, *8*(3), 472–478.
9. Geddes, C. D., Cao, H., Gryczynski, I., Gryczynski, Z., Fang, J., Lakowicz, J. R. (2003). Metal-enhanced fluorescence due to silver colloids on a planar surface: Potential applications of indocyanine green to in vivo imaging. *J. Phys. Chem. A*, *107*, 3443–3449.
10. Link, S., El-Sayed, M. A. (2000). Shape and size dependence of radiative, nonradiative and photothermal properties of gold nanocrystals. *Int. Rev. Phys. Chem.*, *19*, 409–453.
11. Kreibig, U., Vollmer, M., Toennies, J. P. (1995). *Optical Properties of Metal Clusters*. Springer-Verlag, Berlin.
12. Gersten, J., Nitzan, A. (1981). Spectroscopic properties of molecules interacting with small dielectric particles. *J. Chem. Phys.*, *75*, 1139–1152.
13. Chew, H. (1987). Transition rates of atoms near spherical surfaces. *J. Chem. Phys.*, *87*, 1355–1360.
14. Philpott, M. R. (1975). Effect of surface plasmons on transitions in molecules. *J. Chem. Phys.*, *62*(5), 1812–1817.
15. Chance, R. R., Prock, A., Silbey, R. (1978). Molecular fluorescence and energy transfer near interfaces. *Adv. Chem. Phys.*, *37*, 1–65.
16. Weitz, D. A., Garoff, S., Gersten, J. I., Nitzan, A. (1983). The enhancement of Raman scattering, resonance Raman scattering and fluorescence from molecules absorbed on a rough silver surface. *J. Chem. Phys.*, *78*(9), 5324–5338.
17. Geddes, C. D., Parfenov, A., Lakowicz, J. R. (2003). Photodeposition of silver can result in metal-enhanced fluorescence. *Appl. Spectrosc.*, *57*(5), 526–531.

18. Geddes, C. D., Parfenov, A., Roll, D., Fang, J., Lakowicz, J. R. (2003). Electrochemical and laser deposition of silver for use in metal-enhanced fluorescence. *Langmuir*, *19*(15), 6236–6241.
19. Geddes, C. D., Parfenov, A., Roll, D., Gryczynski, I., Malicka, J., Lakowicz, J. R. (2003). Silver fractal-like structures for metal-enhanced fluorescence: Enhanced fluorescence intensities and increased probe photostabilities. *J. Fluorescence*, *13*(3), 267–276.
20. Parfenov, A., Gryczynski, I., Malicka, J., Geddes, C. D., Lakowicz, J. R. (2003). Enhanced fluorescence from fluorophores on fractal silver surfaces. *J. Phys. Chem. B*, *107*(34), 8829–8833.
21. Michaels, A. M., Jiang, J., Brus, L. (2000). Ag nanocrystal junctions as the site for surface-enhanced Raman scattering of single rhodamine 6G molecules. *J. Phys. Chem. B*, *104*, 11965–11971.
22. Michaels, A. M., Nirmal, M., Brus, L. E. (1999). Surface enhanced Raman spectroscopy of individual rhodamine 6G molecules on large Ag nanocrystals. *J. Am. Chem. Soc.*, *121*, 9932–9939.
23. Brown, P. O., Botstein, D. (1999). Exploring the new world of the genome with DNA microarrays. *Nat. Genet. Suppl.*, *21*, 33–37.
24. Schena, M., Heller, R. A., Theriault, T. P., Konrad, K., Lachenmeier, E., Davis, R. W. (1998). Microarrays: Biotechnology's discovery platform for functional genomics. *TIBTECH*, *16*, 301–306.
25. Komurian-Pradel, F., Paranhos-Bacala, G., Sodoyer, M., Chevallier, P., Mandrand, B., Lotteau, V., Andre, P. (2001). Quantitation of HCV RNA using real-time PCR and fluorimetry. *J. Virol. Methods*, *95*, 111–119.
26. Walker, N. J. (2002). A technique whose time has come. *Science*, *296*, 557–559.
27. Difilippantonio, M. J., Ried, T. (2003). Technicolor genome analysis. In J. R. Lakowicz (Ed.), *Topics in Fluorescence Spectroscopy*, Vol. 7: *DNA Technology*. Kluwer Academic/Plenum, New York, pp. 291–316.
28. Soper, S. A., Nutter, H. L., Keller, R. A., Davis, L. M., Shera, E. B. (1993). The photophysical constants of several fluorescent dyes pertaining to ultrasensitive fluorescence spectroscopy. *Photochem. Photobiol.*, *57*, 972–977.
29. Amrbose, W. P., Goodwin, P. M., Jett, J. H., VanOrden, A., Werner, J. H., Keller, R. A. (1999). Single molecule fluorescence spectroscopy at ambient temperature. *Chem. Rev.*, *99*, 2929–2956.
30. Malicka, J., Gryczynski, I., Lakowicz, J. R. (2003). DNA hybridization assays using metal-enhanced fluorescence. *Biochem. Biophys. Res. Commun.*, *306*, 213–218.
31. Szmacinski, H., Lakowicz, J. R. (1994). Lifetime-based sensing. In J. R. Lakowicz (Ed.), *Topics in Fluorescence Spectroscopy*. Plenum, New York, pp. 295–334.
32. Lakowicz, J. R., Malicka, J., Gryczynski, I. (2003). Silver particles enhance emission of fluorescent DNA oligomers. *BioTechniques*, *34*, 62–68.
33. Rye, H., Yue, S., Wemmer, D., Quesada, M. A., Haugland, R. P., Mathies, R. A., Glazer, A. N. (1992). Stable fluorescent complexes of double-stranded DNA with bis-intercalating asymmetric cyanine dyes: Properties of applications. *Nucleic Acids Res.*, *20*, 2803–2812.

34. Benson, S. C., Mathies, R. A., Glazer, A. N. (1993). Heterodimeric DNA binding dyes designed for energy transfer: Stability and applications of the DNA complexes. *Nucleic Acids Res.*, *21*, 5720–5726.
35. Benson, S. C., Zeng, Z., Glazer, A. N. (1995). Fluorescence energy transfer cyanine heterodimers with high affinity for double-stranded DNA. *Anal. Biochem.*, *231*, 247–255.
36. Lakowicz, J. R., Malicka, J., Gryczynski, I. (2003). Increased intensities of YOYO-1-labeled DNA oligomers near silver particles. *Photochem. Photobiol.*, *77*(6), 604–608.
37. Forster, Th. (1948). Intermolecular energy migration and fluorescence. *Ann. Phys.*, *2,* 55–75.
38. Lakowicz, J. R., Malicka, J., D'Auria, S., Gryczynski, I. (2003). Release of the self-quenching of fluorescence near silver metallic surfaces. *Anal. Biochem.*, *320*, 13–20.
39. Lakowicz, J. R., Gryczynski, I., Malicka, J., Gryczynski, Z., Geddes, C. D. (2002). Enhanced and localized multiphoton excited fluorescence near metallic silver islands: Metallic islands can increase probe photostability. *J. Fluorescence*, *12*, 299–302.
40. Malicka, J., Gryczynski, I., Fang, J., Kusba, J., Lakowicz, J. R. (2003). Increased resonance energy transfer between fluorophores bound to DNA in proximity to metallic silver particles. *Anal. Biochem.*, *315*, 160–169.
41. Hua, X. M., Gersten, J. I., Nitzan, A. (1985). Theory of energy transfer between molecules near solid state particles. *J. Chem. Phys.*, *83*, 3650–3659.
42. Gersten, J. I., Nitzan, A. (1984). Accelerated energy transfer between molecules near a solid particle. *Chem. Phys. Lett.*, *104*(1), 31–37.
43. Gryczynski, I., Malicka, J., Gryczynski, Z., Lakowicz, J. R. (2004). Radiative decay engineering 4. Experimental studies of surface plasmon-coupled directional emission. *Anal. Biochem.*, *324*, 170–182.
44. Liebermann, T., Knoll, W. (2000). Surface plasmon field-enhanced fluorescence spectroscopy. *Colloid Surf.*, *171*, 115–130.
45. Lakowicz, J. R. (2004). Radiative decay engineering 3: Surface plasmon-coupled directional emission. *Anal. Biochem.*, *324*, 153–169.
46. Geddes, C. D., Gryczynski, I., Malicka, J., Gryczynski, Z., Lakowicz, J. R. (2003). Metal-enhanced fluorescence: Potential applicatiosn in HTS. *Comb. Chem. High Throughput Screen.*, *6*, 109–117.

# METHODS FOR THE DESIGN AND ANALYSIS OF REPLICATE-EXPERIMENT STUDIES TO ESTABLISH ASSAY REPRODUCIBILITY AND THE EQUIVALENCE OF TWO POTENCY ASSAYS

Brian J. Eastwood, Amy K. Chesterfield, Mary C. Wolff, and Christian C. Felder

*Eli Lilly and Company*
*Indianapolis, Indiana*

1 INTRODUCTION
2 BIOMOLECULAR SCREENING ASSAY CONCENTRATION–RESPONSE MODELS AND OUTCOMES
Affinity Models and Parameter Estimates
Muscarinic $M_1$ Receptor GTP$\gamma^{35}$S Assay
3 ESTIMATING POTENCY PRECISION
Replicate-Experiment Protocol
Statistical Model for Replicate-Experiment Protocol
Replicate-Experiment Outcomes
Model Parameter and Outcome Estimates
Acceptance Criteria
Example Data
Methods for Interlaboratory Comparisons

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

4 STATISTICAL PROPERTIES OF ESTIMATES AND ACCEPTANCE CRITERIA
Data-Generating Model
Properties of Estimated Outcomes
Properties of Acceptance Criteria
References

## 1 INTRODUCTION

In pharmaceutical drug discovery molecules are evaluated in biochemical assays to ascertain the potency of the molecule acting on the proposed biological target. Initially molecules are typically run at a single concentration in a high-throughput screen. Then molecules exhibiting substantial activity are run again in a concentration–response assay to estimate the potency of the molecule for the biological target. From this leads are identified, and then new chemical entities are developed to improve the potency against the biological target and further enhance other properties such as pharmacokinetics and margin of safety. These new chemical entities are often assessed directly in the concentration–response assay to develop and refine the structure–activity relationship.

Statistical parameters estimated in validation studies to establish the reliability of single-concentration assays, such as the Signal Window and Z′-Factor, are well established [1–3]. Eastwood et al. [4] showed that these statistics do not ensure that potency estimates from concentration–response assays will be sufficiently reliable for structure–activity relationship use, and proposed the minimum significant ratio (MSR) as a statistical parameter to estimate the precision of potency estimates from concentration–response assays. They also proposed a standard experimental protocol to evaluate the precision of estimates derived from concentration response assays. In this chapter the statistical models describing the data generated from these protocols are developed and the statistical properties of the outcomes described in Eastwood et al. [4] are examined. This chapter is organized as follows: Section 2 of "Pharmacokinetics and Toxicokinetics in Drug Safety Evaluation" contains a description of a model commonly used to describe biochemical concentration–response assays in pharmaceutical drug discovery. Concentration–response data from a human $M_1$ muscarinic [$^{35}$S]guanosine-5′*o*-(3-thio)triphosphate (GTP$\gamma^{35}$S) assay is described and fit by this model. Section 3 of "Pharmacokinetics and Toxicokinetics in Drug Safety Evaluation" contains a description of the replicate-experiment protocols proposed in Eastwood et al. [4], a statistical model for data derived from these experiments, and acceptance criteria for concentration–response assays. The methods are applied to the GTP$\gamma^{35}$S assay to estimate the precision of the potency estimates. Methods are proposed to extend the replicate-experiment protocol to cover transferring an assay to a new laboratory. Section 4 of "Phar-

macokinetics and Toxicokinetics in Drug Safety Evaluation" contains statistical distribution theory describing the estimators and outcomes and simulation studies further describing the properties of the parameters.

## 2 BIOMOLECULAR SCREENING ASSAY CONCENTRATION–RESPONSE MODELS AND OUTCOMES

### Affinity Models and Parameter Estimates

See Kenakin [5] for a general discussion of affinity models. For a concentration–response assay of a stimulatory, agonist, or activation nature a mechanistic model commonly used to fit the data is the four-parameter logistic (4PL) model:

$$z(c) = b + \frac{t - b}{1 + (\mathrm{EC}_{50}/c)^s} \tag{1}$$

where $c$ is the concentration of the compound, $z$ is the assay response, $b$ and $s$ are the bottom and slope (nuisance) parameters, $\mathrm{EC}_{50}$ is the concentration giving a half-maximal response, and $t$ is the top or relative efficacy parameter. The $\mathrm{EC}_{50}$ is the affinity, or potency, estimate for such data. Usually the data have been normalized to separately measured responses representing no stimulus and maximal stimulus of a reference agonist. There are slight differences in the models for receptor binding and inhibitory assays. In receptor binding assays $t = 100$, $b = 0$, and $s = -1$ are theoretical values when the data have been normalized in this manner, and $s = -1$ holds for responses that have not been normalized as well. For inhibitory assays normalized in the manner described above, $t = 100$, and for some types of competitive inhibition, $b = 0$. For both receptor binding and inhibition assays the potency (half-maximal response parameter) is denoted as the $\mathrm{IC}_{50}$.

There are many different forms of the 4PL equation and a commonly used form in discovery research, notably by the software package Graphpad/Prism [6], is

$$z(c) = b + \frac{t - b}{1 + 10\wedge[(\log \mathrm{EC}_{50} - \log(c))s]} \tag{2}$$

where $\log(c)$ is the base 10 logarithm of the concentration and ^ means exponent.

The $\mathrm{EC}_{50}$ parameter estimate is right-skewed, whereas the $\log \mathrm{EC}_{50}$ parameter estimate is nearly symmetric [7]. Both $\mathrm{EC}_{50}$ and $\log \mathrm{EC}_{50}$ estimates are approximately normally distributed [8], but since the $\log \mathrm{EC}_{50}$ estimate is less skewed, it will be better approximated by a normal distribution in small sample sizes [9].

## Muscarinic $M_1$ Receptor GTP$\gamma^{35}$S Assay

A replicate-experiment protocol was carried out for a human $M_1$ muscarinic receptor GTP$\gamma^{35}$S binding assay. Details on the assay protocol are as follows: GTP$\gamma^{35}$S binding was determined using the antibody capture scintillation proximity method [10]. Briefly, 100 μL homogenized membrane preparations from CHO cells over expressing human $M_1$ muscarinic receptors (20 to 4 fmol/well; Perkin-Elmer Life Sciences, Boston, MA) were incubated for 30 min with 50 μL compound (11-point half-log serial dilutions starting at 100 μM concentration) in a 96-well format. Fifty microliters of GTP$\gamma^{35}$S (500 pM final concentration; Perkin-Elmer Life Sciences, Boston, MA) were added per well and incubated for 30 min. Following incubation, membranes were solubilized by addition of 20 μL 3 percent Nonidet P40 (0.27 percent final concentration; Roche, Indianapolis, IN) per well and incubated for 30 min. Twenty microliters rabbit polyclonal antibodies to $G\alpha_{q/11}$ (1/400 final concentration; Covance, Denver, PA) were added per well and incubated for 60 min. Anti-rabbit SPA PVT antibody binding beads (1.25 mg/well final concentration; Amersham Life Sciences, Piscataway, NJ) were suspended in 20-mL assay buffer and 50 μL were added per well and incubated for 3 h. Assay plates were then centrifuged for 10 min at 1000 *g*, and radioactivity was counted for 1 min per well on a Wallac Trilux MicroBeta plate counter (Perkin-Elmer Life Sciences, Boston, MA). All incubations were performed at room temperature in a 20 mM N-(2-hydroxyethyl)piperazine-N′-(2-ethanesulfonic acid) (HEPES), 100 mM NaCl, 5 mM $MgCl_2$, pH 7.4 assay buffer.

All data were analyzed using GraphPad/Prism [11] to determine concentration–response curves by fitting data using a 4PL model. Responses for test compounds were normalized to 100 μM oxotremorine-M.

Figure 1 contains concentration–response data and least-squares fits of the 4PL equation for acetylcholine (Fig. 1*a*) and carbachol (Fig. 1*b*). The $\log EC_{50}$ estimates can be transformed to linear estimates by the delta method [12]. For this model the formulas are

$$\widehat{EC_{50}} = 10\wedge(\widehat{\log EC_{50}}) \quad \text{and} \quad SE(\widehat{EC_{50}}) = SE\,(\widehat{\log EC_{50}})\widehat{EC_{50}}\ln(10) \qquad (3)$$

for the standard error (SE). Using these formulas gives $EC_{50}$ estimates and standard errors for acetylcholine and carbachol of 276.7 (40.1) nM and 2871 (398) nM. Note that the standard error of carbachol is nearly 10-fold higher than the standard error of acetylcholine, which is consistent with the lognormal distribution assumption of $EC_{50}$ parameter estimates.

Based on the point estimates acetylcholine is 10.4-fold more potent than carbachol. However, the statistical variation of the assay is currently unknown, and hence it is necessary to determine the assay variation to know if this ratio of potencies is within normal variation limits in the assay or if this represents a real potency difference between the two molecules. The standard errors of the parameter estimates can be used, but, as discussed in Eastwood et al. [4], these are unreliable as the data-generating process does not satisfy the statis-

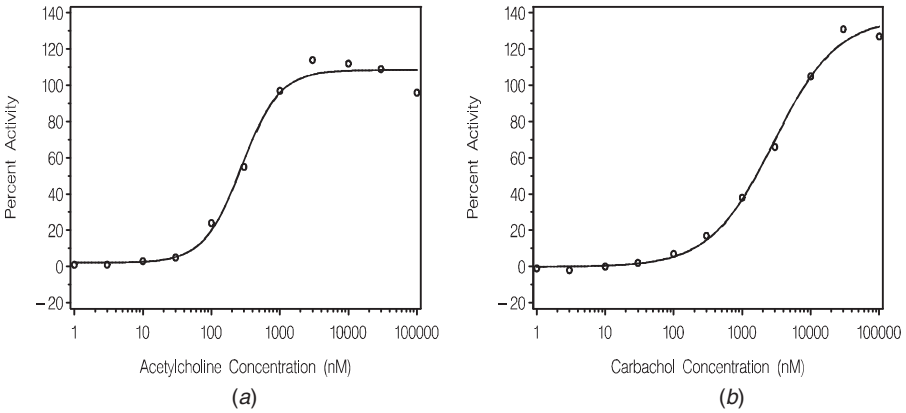


**Figure 1** Concentration–response curves of (*a*) acetylcholine and (*b*) carbachol in the GTP$\gamma^{35}$S hM1 receptor assay. Four-parameter logistic estimates (standard errors) for acetylcholine are $b = 2.1$ (3.39), $t = 108.4$ (3.31), $s = 0.96$ (0.11), and log $EC_{50} = 2.44$ (0.063), and for carbachol are $b = -0.2$ (2.10), $t = 137.3$ (5.17), $s = 0.94$ (0.10), and log $EC_{50} = 0.46$ (0.060).

tical assumptions underlying the least-squares model, and estimates are often derived on a small number of data points (in this case 11 points are in the concentration–response curves). Eastwood et al. [4] proposed using a replicate-experiment protocol to estimate assay precision. In Sections 3 of "Pharmacokinetics and Toxicokinetics in Drug Safety Evaluation" and 15.4 we develop the statistical model for the replicate-experiment protocol and examine the properties of the parameter estimates.

## 3 ESTIMATING POTENCY PRECISION

### Replicate-Experiment Protocol

The precision and accuracy of the biological parameter estimates can not be assured, or even estimated, from validation methods such as Signal Window or Z′-factor employed to validate the single concentration assays [4]. Methods to formally examine the precision have not been standardized in drug discovery. Eastwood et al. [4] proposed a set of experiments based on well-known clinical methods described in Bland and Altman [13]. The protocol, briefly summarized from [4], is to evaluate a set of molecules in two independent runs of the assay and examine the agreement in the resulting potency estimates. Note that we assume experimental iterations require changes across days, and therefore throughout this chapter we use the terms *run* and *day* synonymously. More important than the time lapse between runs is the assumption of independence. By that we mean two complete iterations of the entire assay process including fresh preparation of buffers and other experimental materials. This protocol then is a refinement of an interobserver agreement study where the "observers" are two runs from the same study.

Methods for examining agreement between two measurements are numerous. See Carrasco and Jover [14], for a summary and classification of these methods. The methods used here are derived from those described by Bland and Altman [13]. Using the classification scheme in Carrasco and Jover [14], the techniques described here are a disaggregate approach as they evaluate each component of the measurement model.

## Statistical Model for Replicate-Experiment Protocol

We first describe a statistical model for the replicate-experiment protocol of two runs of each compound within a single laboratory. Let $X_t$ and $Y_t$ be the log potencies obtained for run 1 and run 2, respectively, of compound $t$, for $t = 1, \ldots, n$. The statistical model we use is to assume $X_t = \mu_t + e_t$ and $Y_t = \mu_t + f_t$, where $\mu_t$ is the true log potency, and $e_t$ and $f_t$ are the measurement errors for run 1 and run 2, respectively. We assume further that $\mu_t \sim (\mu, \sigma_C^2)$, $e_t \sim N(\mu_X, \sigma^2)$, and $f_t \sim N(\mu_Y, \sigma^2)$. Note that no assumption is made about the distribution of the compound's true values other than that each compound has a possibly different potency and that the potency distribution does have a finite mean and variance.

We do assume that the measurement errors are independent, identically and normally distributed, independent across days, and that the measurement standard deviation does not vary by day. The normal distribution assumption is reasonable here as the log potency is a parameter from a nonlinear regression model and hence has an approximate normal distribution [8]. The standard deviation, $\sigma$, will typically not be equal to the standard error reported by the nonlinear regression fitting program. See [4] for a discussion of this point.

Each day may have a different mean. An alternative to the latter assumption is to assume that there is a run-day variance component, and that $\mu_X$ and $\mu_Y$ are realizations from this distribution. In the simulation work to evaluate sample size and acceptance criterion properties, we make this assumption and examine the effect of interday variation on the properties of the estimators. However, for analysis this assumption is not necessary, and instead we consider the day effects to be fixed, or alternatively, conditional on the day realizations.

## Replicate-Experiment Outcomes

Eastwood et al. [4] identify three outcomes for the replicate-experiment protocol: the minimum significant ratio (MSR), the mean ratio (MR), and limits of agreement (LsA). Their derivations are as follows.

On any particular day, let $LP_s$ and $LP_t$ be the log potencies for two molecules $s$ and $t$. Then under the same assumptions stated for the replicate-experiment protocol, the difference in log potencies is distributed as $LP_s - LP_t \sim N(\mu_s - \mu_t, 2\sigma^2)$. Therefore to test the null hypothesis $H_0$: $\mu_s = \mu_t$, the test sta-

tistic $T = \mathrm{LP}_s - \mathrm{LP}_t$ has a $N(0, 2\sigma^2)$ distribution under the null hypothesis. Thus if $Z_{1-\alpha/2}$ is the $(1 - \alpha/2)$th percentile from a normal distribution, one would reject the null hypothesis if $|T| > Z_{1-\alpha/2}\sqrt{2}\sigma$, or alternatively if the ratio of the lower potency (i.e., higher numerical value) to the higher potency estimate exceeded $10\wedge(Z_{1-\alpha/2}\sqrt{2}\sigma)$. Therefore, $10\wedge(Z_{1-\alpha/2}\sqrt{2}\sigma)$ is the minimum significant ratio for level $\alpha$, or $\mathrm{MSR}_\alpha$. Taking $\alpha$ to be 5 percent, and rounding to one significant digit gives the formula $\mathrm{MSR} = 10\wedge(2\sqrt{2}\sigma)$. Throughout this chapter, unless denoted otherwise, MSR is determined using $Z_{1-\alpha/2} = 2$.

The mean ratio (MR) is the average ratio across the compound set between the two days. To see this, note that

$$\mathrm{MR} = (10\wedge\mu_X)/(10\wedge\mu_Y) = 10\wedge(\mu_X - \mu_Y)$$

and that

$$\left(\frac{1}{n}\right)\sum E[X_t - Y_t] = \mu_X - \mu_Y$$

The limits of agreement (LsA) are as defined by Bland and Altman [13]. In terms of population parameters on the log scale, they are can be written as

$$\mathrm{LsA}_\alpha = 10\wedge\left[(\mu_X - \mu_Y) \pm Z_{1-\alpha/2}\sigma_d\right]$$

where $\sigma_d = \sqrt{2}\sigma$ is the standard deviation of the paired difference. Translating to the linear scale the results are

$$\mathrm{LsA}_\alpha \triangleq (\mathrm{LsA}_{L,\alpha}, \mathrm{LsA}_{U,\alpha}) = (\mathrm{MR}/\mathrm{MSR}_\alpha, \mathrm{MR} \times \mathrm{MSR}_\alpha)$$

A key consideration in the selection of these outcomes is that all are independent of $\{\mu_t: t = 1, \ldots, n\}$, that is, they do not depend on the actual potency of compounds chosen. In practice, it would be appropriate to choose a set with a large range in potencies from a generalizability perspective, but theoretically the outcomes are independent of the potency distribution.

## Model Parameter and Outcome Estimates

Let $(x_t, y_t)$, $t = 1, \ldots, n$ be the observed log potencies for the $n$ compounds, and let $d_t$ and $m_t$ be difference and mean log potency across the two runs for the $n$ compounds. Bland and Altman [13] recommend plotting $m_t$ versus $d_t$. However, we prefer to plot $r_t = 10\wedge d_t$ versus $gm_t = 10\wedge m_t$ with each plotted on a logarithmic scale as potency ratios and values are more familiar to our biologists and chemists than log potencies. Each run's average error, $\mu_x$ and $\mu_y$, is not individually estimable, but their difference, $\mu_x - \mu_y$, is estimated by $\bar{x} - \bar{y}$.

The average potency ratio, $\bar{r} = \bar{x}/\bar{y}$ = MR, is plotted as a horizontal line on the ratio-geometric mean scatter plot.

The data are naturally paired, and so under the null hypothesis $H_0$: $\mu_X = \mu_Y$ $\sqrt{n/2}\log(\mathrm{MR})/\sigma$ has a $t$ distribution with $n - 1$ degrees of freedom. Let $s_d$ be the sample standard deviation of $d_t$ the difference in log potencies. Then $\hat{\sigma} = s_d/\sqrt{2}$ is the usual unbiased estimate of $\sigma$. If $t_{0.975,n-1}$ is the 97.5th percentile from a distribution with $n - 1$ degrees of freedom and if $M = 10\wedge(t_{0.975,n-1}\, s_d/\sqrt{n})$, then a 95 percent confidence interval for the MR is given by $(\mathrm{MR}_L, \mathrm{MR}_U) = (\widehat{\mathrm{MR}}/M, \widehat{\mathrm{MR}} \times M)$. The null hypothesis is rejected if the confidence interval does not contain the value 1.0.

The mean ratio limits do not describe a region that contains the data but instead a region that contains the true mean ratio MR. As Bland and Altman [13] show, this interval is not relevant for describing individual agreement, and they recommend plotting ± 2 standard deviations as the limits of agreement, which corresponds approximately to $\alpha = 0.05$. Thus the estimated MSR is $\widehat{\mathrm{MSR}} = 10\wedge(2s_d)$, and the sample limits of agreement are $\widehat{\mathrm{LsA}} \triangleq (\widehat{\mathrm{LsA}}_L, \widehat{\mathrm{LsA}}_U) = (\widehat{\mathrm{MR}}/\widehat{\mathrm{MSR}}, \widehat{\mathrm{MR}} \times \widehat{\mathrm{MSR}})$.

Note that $\hat{\mu}_t = (x_t + y_t)/2$, $t = 1, \ldots, n$ and $\hat{\mu} = (\hat{\mu}_1 + \ldots \hat{\mu}_n)/n$ are the compound estimates and overall compound mean. However, these parameters are not of primary interest in the replicate-experiment protocol. In fact, unless the sample of molecules tested is a random sample from the population, $\hat{\mu}$ will be a biased estimator of $\mu$.

## Acceptance Criteria

See Eastwood et al. [4] for a discussion of MSR requirements from pharmacological validation and medicinal chemistry structure–activity relationship perspectives.

The MSR can be directly estimated from the replicate-experiment protocol. Sufficient evidence pertaining to the across-run variation can only be obtained over several runs of the assay, which is beyond the scope of most preproduction evaluation efforts and not discussed here. In the replicate-experiment protocol, across-run variation impacts the magnitude of the MR, and therefore of the LsA, and so preliminary decisions can be made based on either the MR or LsA outcomes. Simulation evidence presented in Section 4 of "Pharmacokinetics and Toxicokinetics in Drug Safety Evaluation" suggests that a sample size of 20 to 30 compounds works reasonably well for criteria designed to reject assays if the assay variation, taking into account across-run variation, exceeds a threefold potency ratio.

## Example Data

A replicate-experiment protocol was conducted using 24 compounds for the muscarinic human $M_1$ GTP$\gamma^{35}$S assay. Each compound was run on two sep-

**TABLE 1 Four-Parameter Logistic Regression Parameter Estimates for Run 1 of GTPγ$^{35}$S M1 Agonist Assay**

| Compound | *b* Est (SE) | *t* Est (SE) | *s* Est (SE) | log $EC_{50}$ Est (SE) |
|---|---|---|---|---|
| C1 | –5.3 (4.29) | 110.6 (3.79) | 1.19 (0.23) | 2.30 (0.077) |
| C2 | 2.1 (3.39) | 108.4 (3.31) | 1.57 (0.32) | 2.44 (0.063) |
| C3 | –4.7 (1.24) | 80.0 (3.13) | 0.96 (0.11) | 3.43 (0.074) |
| C4 | –0.2 (2.10) | 137.3 (5.17) | 0.94 (0.10) | 3.46 (0.060) |
| C5 | 0.3 (1.11) | 91.2 (2.34) | 0.87 (0.07) | 3.28 (0.045) |
| C6 | 2.5 (1.83) | 38.2 (0.82) | 1.18 (0.21) | 1.53 (0.107) |
| C7 | –3.3 (2.68) | 74.0 (3.24) | 0.75 (0.12) | 2.68 (0.095) |
| C8 | –4.9 (2.26) | 94.1 (10.98) | 0.58 (0.10) | 3.38 (0.188) |
| C9 | –0.6 (2.99) | 106.8 (2.56) | 1.15 (0.16) | 2.27 (0.058) |
| C10 | –2.6 (1.04) | 88.2 (2.09) | 1.05 (0.09) | 3.41 (0.035) |
| C11 | –2.9 (0.53) | 58.6 (0.86) | 1.16 (0.07) | 3.14 (0.024) |
| C12 | 0.7 (1.41) | 114.0 (2.41) | 1.03 (0.08) | 3.15 (0.038) |
| C13 | 1.3 (3.23) | 62.9 (2.02) | 1.30 (0.30) | 1.81 (0.090) |
| C14 | 3.0 (1.01) | 84.1 (1.93) | 1.38 (0.14) | 2.42 (0.036) |
| C15 | 1.3 (1.27) | 98.9 (2.34) | 1.09 (0.10) | 2.26 (0.040) |
| C16 | 4.5 (3.88) | 51.7 (4.61) | 1.18 (0.56) | 1.74 (0.203) |
| C17 | –7.0 (4.21) | 93.1 (3.67) | 0.84 (0.15) | 2.34 (0.096) |
| C18 | –1.6 (1.11) | 25.5 (0.79) | 0.99 (0.17) | 2.07 (0.082) |
| C19 | –1.5 (0.78) | 32.3 (1.06) | 1.51 (0.25) | 3.01 (0.054) |
| C20 | 0.3 (1.46) | 22.1 (2.02) | 1.28 (0.56) | 2.98 (0.172) |
| C21 | 0.9 (1.61) | 37.3 (1.83) | 0.95 (0.22) | 2.64 (0.113) |
| C22 | 1.0 (1.44) | 29.5 (1.24) | 1.05 (0.24) | 2.13 (0.104) |
| C23 | 0.1 (2.42) | 58.9 (1.92) | 1.19 (0.24) | 2.16 (0.081) |
| C24 | –1.1 (0.92) | 30.1 (1.27) | 0.92 (0.15) | 2.87 (0.084) |

arate days. Tables 1 and 2 contain the complete 4PL model parameter estimates for run 1 and run 2, respectively, calculated using Eq. (2). Table 3 contains the $EC_{50}$ parameter estimates and standard errors for both runs, calculated using the delta method of Eq. (3). Acetylcholine is compound C2 and carbachol is compound C4. Figure 2 is the plot of the potency ratios versus the geometric mean potency. There does not appear to be any systematic trend between agreement and geometric mean potency, and all ratios appear inside the LsA estimates, which are 0.68 to 1.55. The MSR estimate is 1.51. There appears to be little evidence of interrun variation as the MR estimate is 1.02 with 95 percent confidence interval ±10 percent ($M = 1.1$).

To return to the issue at the end of Section 2 of "Pharmacokinetics and Toxicokinetics in Drug Safety Evaluation", the potency ratio of carbachol to acetylcholine was 10.4 in run 1, 9.51 in run 2, and both are well above the MSR estimate of 1.51. This analysis confirms that in this assay there is a real potency difference between acetylcholine and carbachol.

**TABLE 2 Four-Parameter Logistic Regression Parameter Estimates for Run 2 of GTPγ$^{35}$S M1 Agonist Assay**

| Compound | *b* Est (SE) | *t* Est (SE) | *s* Est (SE) | log $EC_{50}$ Est (SE) |
|---|---|---|---|---|
| C1 | –2.7 (2.33) | 102.5 (2.01) | 1.15 (0.13) | 2.28 (0.046) |
| C2 | 1.7 (3.35) | 98.5 (3.38) | 1.43 (0.30) | 2.49 (0.073) |
| C3 | 1.9 (0.91) | 78.8 (2.05) | 0.97 (0.08) | 3.46 (0.040) |
| C4 | 4.0 (0.53) | 116.7 (1.29) | 0.97 (0.03) | 3.47 (0.018) |
| C5 | 4.5 (0.91) | 77.8 (1.51) | 1.06 (0.09) | 3.13 (0.037) |
| C6 | 8.7 (0.90) | 35.2 (0.48) | 1.51 (0.23) | 1.48 (0.048) |
| C7 | 4.8 (1.63) | 67.7 (1.71) | 1.17 (0.17) | 2.55 (0.060) |
| C8 | 5.1 (2.53) | 84.9 (6.86) | 0.68 (0.14) | 3.39 (0.152) |
| C9 | –2.5 (2.60) | 104.4 (2.10) | 1.25 (0.15) | 2.34 (0.046) |
| C10 | 0.7 (0.95) | 92.5 (1.84) | 1.17 (0.09) | 3.34 (0.032) |
| C11 | 0.4 (1.60) | 66.4 (2.78) | 1.17 (0.20) | 3.23 (0.071) |
| C12 | 3.3 (1.86) | 111.1 (2.46) | 1.15 (0.12) | 3.05 (0.042) |
| C13 | 3.9 (1.59) | 71.6 (0.95) | 0.99 (0.09) | 1.88 (0.043) |
| C14 | 5.2 (1.56) | 90.8 (2.19) | 1.74 (0.23) | 2.43 (0.036) |
| C15 | 2.1 (1.32) | 99.2 (1.76) | 1.66 (0.16) | 2.37 (0.027) |
| C16 | 1.5 (1.78) | 58.2 (1.14) | 1.17 (0.14) | 1.55 (0.049) |
| C17 | –3.1 (1.68) | 103.2 (1.38) | 1.06 (0.08) | 2.23 (0.033) |
| C18 | –1.2 (1.80) | 35.3 (1.41) | 0.85 (0.17) | 2.22 (0.107) |
| C19 | 0.9 (0.87) | 43.5 (1.30) | 1.31 (0.19) | 3.10 (0.053) |
| C20 | 4.0 (1.82) | 30.2 (2.39) | 2.11 (1.45) | 3.00 (0.133) |
| C21 | 1.7 (0.98) | 48.0 (1.09) | 1.22 (0.15) | 2.63 (0.050) |
| C22 | 2.9 (1.45) | 35.6 (1.13) | 1.24 (0.27) | 2.13 (0.086) |
| C23 | 2.6 (1.84) | 70.4 (1.67) | 1.01 (0.12) | 2.17 (0.057) |
| C24 | –0.2 (3.00) | 44.6 (3.92) | 0.58 (0.16) | 2.72 (0.191) |

## Methods for Interlaboratory Comparisons

Often the follow-up work is conducted in a different laboratory than the screening laboratory. In those instances it is necessary to validate that the protocols in the two labs yield similar results. To examine this property, the replicate-experiment protocol is modified so that the screening laboratory run evaluates the compound set a single time, and the confirmation laboratory conducts the replicate-experiment protocol as outlined above. In addition to the comparisons of the two runs in the confirmation laboratory, the screening laboratory's run will be compared to the first run of the confirmation laboratory.

This is a "minimalist" form of an interlaboratory comparison study and is based on the assumption that the screening laboratory's assay has previously been evaluated and in production, whereas the confirmation laboratory is just initiating the evaluation process. In situations that are different from this scenario, more elaborate evaluation protocols may be possible and/or

**TABLE 3 $EC_{50}$ Estimates for Runs 1 and 2 of GTP$\gamma^{35}$S M1 Agonist Assay**

| | Run 1 | | Run 2 | |
|---|---|---|---|---|
| Compound | $EC_{50}$ | (SE) | $EC_{50}$ | (SE) |
| C1 | 201.4 | (35.8) | 191.4 | (20.4) |
| C2 | 276.7 | (40.1) | 308.3 | (51.6) |
| C3 | 2697.7 | (459.4) | 2904.0 | (268.3) |
| C4 | 2870.8 | (397.9) | 2930.9 | (122.2) |
| C5 | 1923.1 | (198.4) | 1345.9 | (114.8) |
| C6 | 33.8 | (8.4) | 30.4 | (3.3) |
| C7 | 481.9 | (105.1) | 354.0 | (48.9) |
| C8 | 2404.4 | (1040.3) | 2426.6 | (849.9) |
| C9 | 186.2 | (24.9) | 216.8 | (23.0) |
| C10 | 2546.8 | (207.5) | 2208.0 | (163.9) |
| C11 | 1383.6 | (78.0) | 1690.4 | (276.4) |
| C12 | 1412.5 | (124.1) | 1127.2 | (110.0) |
| C13 | 64.4 | (13.4) | 76.2 | (7.5) |
| C14 | 263.6 | (21.7) | 269.2 | (22.3) |
| C15 | 182.0 | (17.0) | 232.8 | (14.5) |
| C16 | 54.3 | (25.4) | 35.5 | (4.0) |
| C17 | 218.3 | (48.2) | 171.4 | (13.1) |
| C18 | 116.4 | (21.9) | 164.8 | (40.5) |
| C19 | 1025.7 | (127.6) | 1253.1 | (153.6) |
| C20 | 948.4 | (374.5) | 1006.9 | (308.4) |
| C21 | 434.5 | (113.2) | 427.6 | (49.1) |
| C22 | 134.6 | (32.3) | 136.1 | (26.9) |
| C23 | 144.5 | (27.1) | 146.6 | (19.2) |
| C24 | 737.9 | (142.7) | 528.4 | (232.7) |

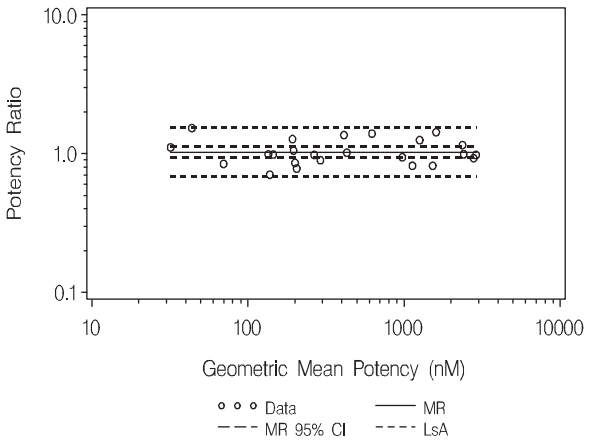


**Figure 2** Potency ratio versus geometric mean potency results from the replicate-experiment study for the GTP$\gamma^{35}$S hM1 agonist assay. $\widehat{\text{MSR}} = 1.51$, $\widehat{\text{MR}} = 1.02$ [95% confidence interval = 0.94–1.12, $p = 0.57$], and $\widehat{\text{LsA}} = (0.68, 1.55)$.

appropriate. For example, if the monitoring history of the production laboratory is very extensive, then an extension to these methods has been proposed to compare assay variations across laboratories [15].

If we change the notation so that $X_t$ and $Y_t$ are the log potency estimates from laboratories 1 and 2 for compound $t, t = 1, \ldots, n$, then a model to describe this situation would be $X_t = \mu_t + l_1 + e_t$ and $Y_t = \mu_t + l_2 + f_t$. Assumptions about $\mu_t$, $e_t$, and $f_t$ are as before. The parameters $l_1$ and $l_2$ are fixed-effect parameters representing the difference in average log potency as measured by laboratories 1 and 2 versus "true" log potency of the molecule. An assumption we are making is that the laboratory bias is constant across all molecules.

Note that "interlaboratory" comparison methods may also be used to compare two different protocols within the same laboratory. These intralab comparisons could include personnel differences, equipment differences, cell line changes, or assay differences based on a cloned cell line versus a native tissue assay.

Under this model the minimum significant ratio, $\mathrm{MRS}_\alpha = 10\text{^}(Z_{1-\alpha/2}\sqrt{2}\sigma)$ is defined as before. However, the mean ratio is now $\mathrm{MR} = 10\text{^}(l_1 - l_2 + \mu_X - \mu_Y)$. Note that none of the four parameters, $l_1$, $l_2$, $\mu_X$, and $\mu_Y$, is individually estimable, neither is $l_1$–$l_2$ nor $\mu_X$–$\mu_Y$. The limits of agreement are $\mathrm{LsA}_\alpha = (\mathrm{MR}/\mathrm{MSR}_\alpha, \mathrm{MR}\times\mathrm{MSR}_\alpha)$ as before.

The outcomes are estimated as per the replicate-experiment protocol.

## 4 STATISTICAL PROPERTIES OF ESTIMATES AND ACCEPTANCE CRITERIA

### Data-Generating Model

The model of the data-generating process is a slight generalization of that developed in Section 3 of "Pharmacokinetics and Toxicokinetics in Drug Safety Evaluation". We now assume that on any run $r$ the log potency of molecule $t$ is given by $Z_{rt} = \mu_t + \rho_r + \varepsilon_{rt}$, for $r = 1, 2$ and $t = 1, \ldots, n$, where $\mu_t$ is the true log potency of compound $t$, $\rho_r$ is a random run (day) shift of all molecules, and $\varepsilon_{rt}$ is the random measurement error. We assume $\varepsilon_{rt} \sim N(0,\sigma^2)$, $\rho_r \sim N(0,\sigma_R{}^2)$, and that $\rho_r$ and $\varepsilon_{rt}$ are statistically independent. The model of Section 3 of "Pharmacokinetics and Toxicokinetics in Drug Safety Evaluation" is just a special case of this model, where we let $\mu_X$ and $\mu_Y$ be two specific runs and consider them fixed throughout the replicate experiment. Then $e_t = \mu_X + \varepsilon_{1t}$ and $f_t = \mu_Y + \varepsilon_{2t}$.

We make this generalization because run effects are not controllable and molecules are usually compared across runs instead of within runs. A statistically significant difference between potency estimates of two molecules then holds if the ratio of the two potencies (weaker to stronger) exceeds the minimum significant ratio that takes into account both within- and across-run variation. Using the same arguments as in Section 3 of "Pharmacokinetics and Toxicokinetics in Drug Safety Evaluation", the critical value is $\mathrm{OvrMSR}_\alpha = 10\text{^}\left[Z_{1-\alpha/2}\sqrt{2(\sigma_R^2 + \sigma^2)}\right]$, and as with the MSR, $Z_{1-\alpha/2} = 2$ unless

explicitly defined otherwise. In this section we will examine the statistical properties of $\widehat{\text{MSR}}$, $\widehat{\text{MR}}$, and $\widehat{\text{LsA}}$ for various values of OvrMSR, MSR, and $n$. Note that because $\sigma_R^2 \geq 0$ OvrMSR ≥ MSR, which imposes limits on the simulation boundaries.

## Properties of Estimated Outcomes

In this section we look at the statistical properties of $\widehat{\text{MSR}}$, $\widehat{\text{MR}}$, and $\widehat{\text{LsA}}$ under the assumption that the log potency estimates are normally distributed. We will examine their properties through both probability distribution theory and Monte Carlo simulation. We will examine cases for MSR varying from 1 to 4, and for OvrMSR varying from 1 to 8, subject to the constraint that OvrMSR ≥ MSR (which must hold by definition). Since none of the estimates depends upon the actual potencies of the molecules, the log potencies are left unspecified. The sample size is varied from 10 to 40 compounds. All simulation data sets are generated and estimators calculated within a data step using Version 8.2 of the SAS system [16]. Plots are produced by Proc Plot Version 8 of SAS/Graph [17]. Simulations are conducted using 5000 Monte Carlo repetitions.

Under the data-generating model of the previous section, if we condition on $\mu_X$ and $\mu_Y$, then $T = \sqrt{n}(\bar{X} - \bar{Y})/s_d$ has a noncentral $t_{n-1,\delta}$ distribution with noncentrality parameter $\delta = \sqrt{n}\,(\mu_X - \mu_Y)/(2\sigma)$. As mentioned in Section 3 of "Pharmacokinetics and Toxicokinetics in Drug Safety Evaluation", under the null hypothesis, $H_0$: $\mu_X = \mu_Y$, $T$ has a central $t$ distribution and can be used to test that null hypothesis. However, under the data-generating model from the beginning of Section 4 of "Pharmacokinetics and Toxicokinetics in Drug Safety Evaluation", this hypothesis is false unless $\sigma_R^2 = 0$, and therefore this hypothesis is of little interest. In fact, testing it may lead to counterintuitive conclusions, as the noncentrality parameter $\delta$ is a decreasing function of the assay measurement error $\sigma$, hence of the MSR. Consequently, the more precise an assay is within each run, the more likely one will conclude there is a statistically significant difference between two runs of the assay.

The phrase "conditioning on $\mu_X$ and $\mu_Y$" means that we consider the two runs as fixed and test a hypothesis about those two particular runs. But, if we consider the two days as just representative runs of the assay production process, then the two run day means are themselves random variables and not fixed values. Under this "unconditional" analysis, the statistic $T$ does not have a $t$ distribution and does not have any well-characterized probability distribution. Therefore to examine this distribution we used Monte Carlo simulation.

Figure 3 shows the effect of OvrMSR, MSR, and $n$ on the distribution of $\widehat{\text{MR}}$. Figure 3*a* shows box plots of the distribution of MR for different combinations of OvrMSR and MSR, using a sample size of 25 compounds. For all combinations the median $\widehat{\text{MR}}$ is 1.0, but as OvrMSR increases the variability and skewness of $\widehat{\text{MR}}$ increase. However, for a fixed

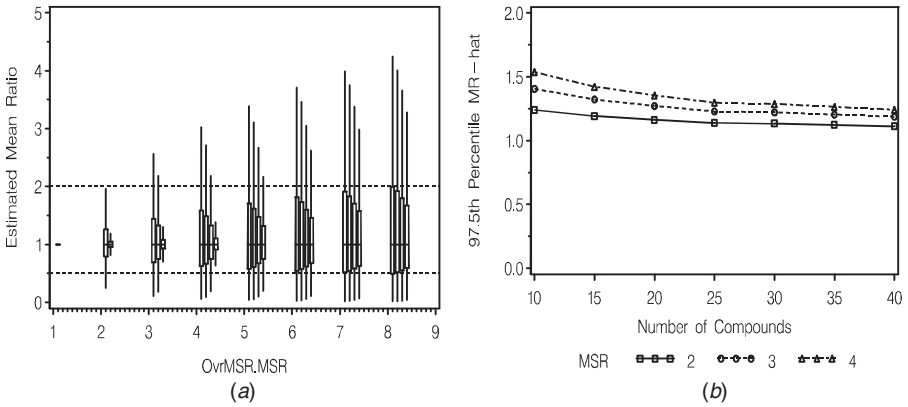


**Figure 3** (*a*) Box plot of distribution of $\widehat{\mathrm{MR}}$ versus OvrMSR.MSR (i.e., whole number indicates the OvrMSR, decimal indicates the MSR) using 25 compounds. (*b*) 97.5th percentile of $\widehat{\mathrm{MR}}$ versus number of compounds by MSR when OvrMSR = MSR.

value of OvrMSR the variability of $\widehat{\mathrm{MR}}$ decreases as MSR increases, and the distribution becomes more symmetric. These changes occur because $\mathrm{SE}(\overline{X}-\overline{Y})=\sqrt{2}(\sigma_R^2+\sigma^2/n)$, and hence holding OvrMSR fixed while increasing MSR effectively lowers $\sigma_R^2$ and increases $\sigma^2$. Figure 3*b* shows the insensitivity of the distribution of $\widehat{\mathrm{MR}}$ to the sample size. When $\sigma_R^2 = 0$ (i.e., OvrMSR = MSR), sample size has its greatest effect, specifically a slight decrease in the variability of $\widehat{\mathrm{MR}}$ as the sample size increases. The effect is even smaller at positive values of across-run variation.

Figure 4 shows the effect of OvrMSR, MSR, and $n$ on the precision of $\widehat{\mathrm{MR}}$. Since a 95 percent confidence interval for MR is obtained by multiplying and dividing $\widehat{\mathrm{MR}}$ by $M$, where $M$ is as defined in Section 3 of "Pharmacokinetics and Toxicokinetics in Drug Safety Evaluation", we examine precision by examining the distribution of $M$. Figure 4*a* shows the distribution of $M$ versus both OvrMSR and MSR for a fixed sample size of 25. Note that the confidence interval precision does not depend upon OvrMSR, but the precision does depend upon the actual value of MSR. The lack of a relationship between $M$ and OvrMSR is predicted from the distributional theory, as is the increasing relationship between $M$ and MSR. Figure 4*b* shows the effect of sample size on confidence interval precision. With less than a 5 percent error rate, if MSR ≤ 4, then a sample size of 25 compounds is sufficient to ensure that $\widehat{\mathrm{MR}}$ is within 50 percent of the true value. Note that the 97.5th percentile of $M$ is an increasing function of MSR, not because the median value is increasing but instead because the variability and skewness of the distribution are increasing as MSR increases, as shown in Figure 4*a*.

Figure 5 shows the effect of OvrMSR, MSR, and $n$ on the power to reject the null hypothesis $H_0$: $\mu_X = \mu_Y$. Figure 5*a* shows for a fixed sample size of 25 compounds that power increases as OvrMSR increases provided the MSR

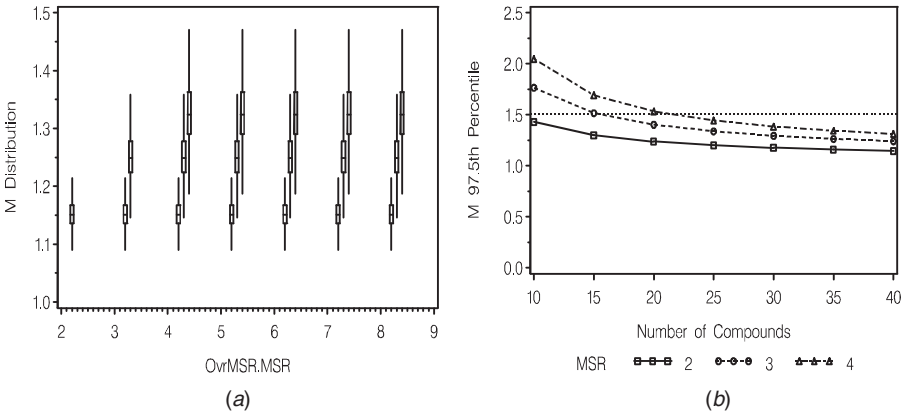


**Figure 4** (*a*) Box plot of distribution of 95% confidence interval multiplier, *M*, for MR versus OvrMSR.MSR (i.e., whole number indicates the OvrMSR, decimal indicates the MSR) using 25 compounds. (*b*) 97.5th percentile of *M* versus number of compounds by MSR when OvrMSR = MSR.

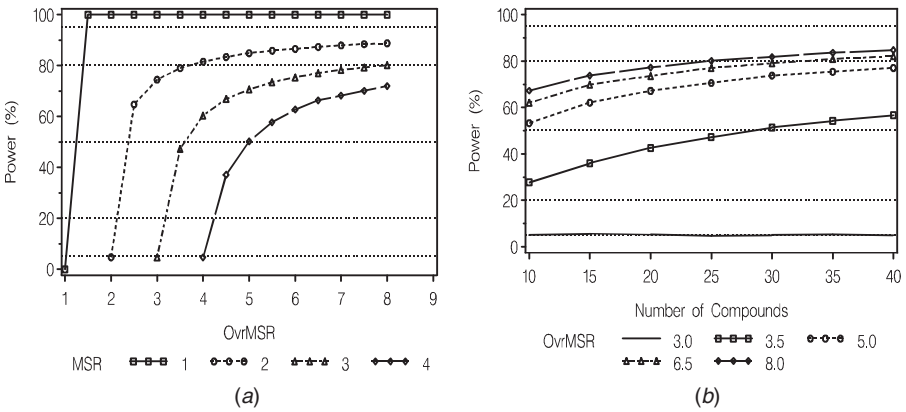


**Figure 5** (*a*) Power to reject null hypothesis $H_0$: $\mu_X = \mu_Y$ versus OvrMSR by MSR for a sample of size 25 compounds. (*b*) Power versus number of compounds by OvrMSR when MSR = 3.

is fixed. However, it also shows that for a fixed OvrMSR power increases as MSR decreases. Hence the lower the measurement error, the more likely one will conclude there is across-run variation, that is, the better the assay is within-run, the more likely it will detect across-run variation. Figure 5*b* shows the power versus sample size relationship for various values of OvrMSR, holding MSR fixed at a value of 3. A sample size of 25 will have between 50 and 80 percent power to detect across-run variation, whereas a sample of size 10 will not exceed 70 percent power at any likely value of OvrMSR. Note, however,

that across-run variation is usually present, and that because of the decreasing function of power versus MSR (holding the other values fixed) a straight hypothesis test is easily misinterpreted and not recommended.

The distribution of MSR parallels that of the precision of $\widehat{\mathrm{MR}}$. Figure 6 shows the effect of OvrMSR, $\widehat{\mathrm{MSR}}$, and $n$ on the distribution of $\widehat{\mathrm{MSR}}$. Figure 6*a* illustrates that the distribution of MSR is independent of OvrMSR for a sample size of 25 compounds. However, both the average value and variance of $\widehat{\mathrm{MSR}}$ are increasing functions of the true MSR. Note that the distribution of MSR is approximately symmetric for all MSR $\leq 4$ at this sample size. Figure 6*b* shows the value of the 97.5th percentile of $\widehat{\mathrm{MSR}}$ versus the sample size for MSR = 2, 3, and 4. The results show that for 25 compounds $\widehat{\mathrm{MSR}}$ will be within 50 percent of its true value, whereas a sample size of 20 or less will fail to achieve this objective.

Finally, we examine properties of the limits of agreement. These limits are intended to provide an interval that contains the upper and lower bounds on potency ratios if two estimates are obtained on the same compound. Figure 7 examines the coverage properties of LsA as a function of OvrMSR, MSR, and $n$. Figure 7*a* illustrates the probability that the LsA interval will not contain all the points, which varies from 25 percent at $n = 10$ compounds to over 90 percent at $n = 40$ compounds. Note that except for the degenerate case of no assay variation (MSR = 1 or OvrMSR = 1), the coverage probability is independent of the assay variation, both within and across run. Figure 7*b* shows the probability that more than 5 percent of the sample will lie outside the limits of agreement (i.e., more than 1 compound for $n < 30$ and more than 2 compounds for $n$ between 30 and 40). As in Figure 7*a*, except for degenerate cases, this probability depends only upon the sample size. At $n = 10$ compounds there

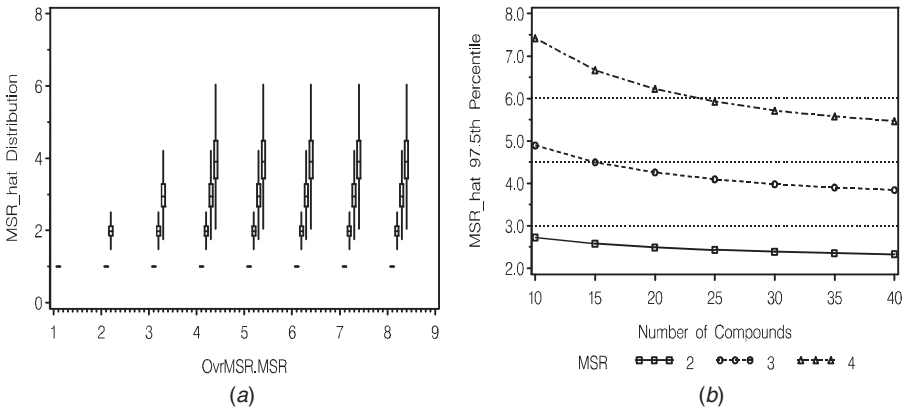


**Figure 6** (*a*) Box plot of distribution of $\widehat{\mathrm{MSR}}$ versus OvrMSR.MSR (i.e., whole number indicates the OvrMSR, decimal indicates the MSR) using 25 compounds. (*b*) 97.5th percentile of $\widehat{\mathrm{MSR}}$ versus number of compounds by MSR when OvrMSR = 8 (note the relationship holds for any value of OvrMSR > MSR).

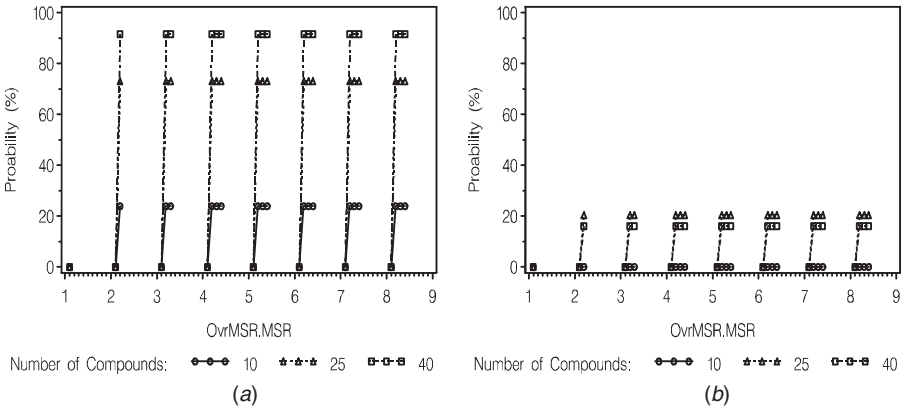


**Figure 7** (*a*) Probability (%) that $\widehat{\text{LsA}}$ does not contain all tested compounds versus OvrMSR.MSR (i.e., whole number indicates the OvrMSR, decimal indicates the MSR) by 10, 25, and 40 compounds. (*b*) Probability (%) that more than 5% of sample (1 compound for $n < 30$ and 2 compounds for $30 \le n \le 40$) lies outside $\widehat{\text{LsA}}$ versus OvrMSR.MSR by 10, 25, and 40 compounds.

is a negligible probability that increases to about 20 percent at $n = 40$ compounds. This latter probability is higher than the nominal rate that results from basing the interval on an estimated value of the assay variation that is not included in the nominal error rate.

Figure 8 examines the distribution of the more extreme value of the limits of agreement. This is defined as the value "furthest" away from 1.0, which is equal to $\widehat{\text{LsA}}_{\max} = \max\{1/\widehat{\text{LsA}}_L, \widehat{\text{LsA}}_U\}$. For a sample size of 25 compounds Figure 8*a* shows box plots of the distribution of $\widehat{\text{LsA}}_{\max}$. Note the median, variation about the median, and skewness are all increasing functions of both OvrMSR and MSR. For MSR = 3, Figure 8*b* shows the relationship of both OvrMSR and $n$. $\text{LsA}_{\max}$ is an increasing function of the former, and relatively independent of the latter.

## Properties of Acceptance Criteria

In Section 3 of "Pharmacokinetics and Toxicokinetics in Drug Safety Evaluation" three outcomes were defined from the replicate-experiment protocol: the MSR, MR, and LsA. Each of these outcomes is positively correlated with one or both components of assay variation, and in the preceding section we examined the sampling distributions of each. In this section we look at the statistical properties of the following three rejection rules or tests:

$$\widehat{\text{MSR}} > C \tag{4}$$

$$\widehat{\text{MR}} > C \tag{5}$$

$$\widehat{\text{LsA}} \not\subset [1/C, C] \tag{6}$$

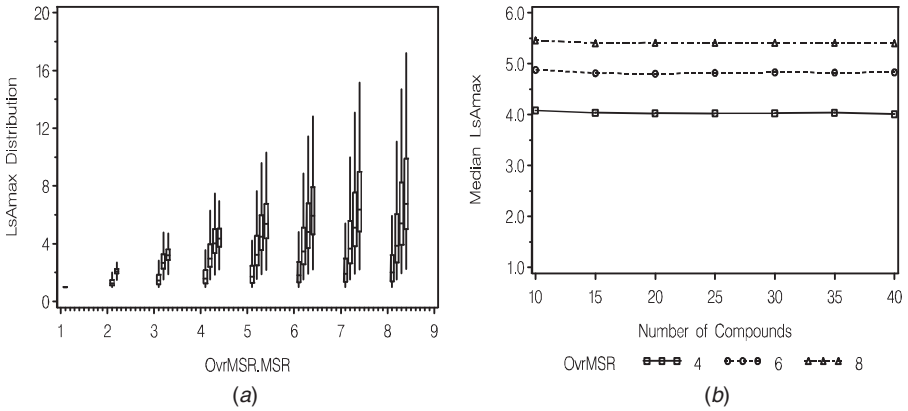


**Figure 8** (*a*) Box plot of distribution of $\widehat{\text{LsA}}_{\max}$ versus OvrMSR.MSR (i.e., whole number indicates the OvrMSR, decimal indicates the MSR) using 25 compounds. (*b*) 97.5th percentile of $\widehat{\text{LsA}}_{\max}$ versus number of compounds by OvrMSR when MSR = 3.

where the critical value, $C$, is greater than one in each case (and is usually different in the three tests). Note that in the case of Eq. (6), an equivalent statement to that above is to reject the assay if $\widehat{\text{LsA}}_{\max} > C$. Throughout this section we estimate the probability of rejecting an assay, defined as the power of the test. The setting of the critical value has to be, of course, related to the biological and chemical requirements of the assay. Throughout this section we assume that the objectives are to reject assays if OvrMSR ≥ 3.

We showed in the previous section that $\widehat{\text{MSR}}$ is estimated to within approximately 50 percent, positively correlated to within-run assay variation, and independent of across-run assay variation. Figure 9 shows the effect of OvrMSR, MSR, $C$, and $n$ on power. Figure 9*a* shows that for $n = 25$ compounds and $C = 3$, power is independent of OvrMSR (except for the degenerate case of OvrMSR = 0). For MSR = 2, power is less than 1 percent, is just less than 50 percent when MSR = 3, and is over 90 percent for MSR ≥ 4. Figure 9*b* shows the effect of sample size, critical value, and MSR on power when OvrMSR > MSR. The results indicate setting C = 2.5 will result in high rejection rates of potentially good assays (i.e., when MSR = 2), especially when combined with a low sample size ($n = 10$). Conversely, setting $C = 3.5$ results in less than 80 percent power to reject the assay when MSR = 4 at all sample sizes. Small sample sizes ($n < 15$) only have adequate power for low critical values, and a sample size of 20 to 30 is required to have high power to reject assays whose MSR = 4. The conclusion from this analysis is that Eq. (4) alone will only suffice in those cases where across-run assay variation is negligible, but in cases where the within-assay variation exceeds 4.0, this criterion will have high probability of rejecting the assay.

We showed in Figure 4*b* that $\widehat{\text{MR}}$ seldom exceeds 1.5 in cases of no across-run assay variation. So in Figure 10 we examine the power about the value

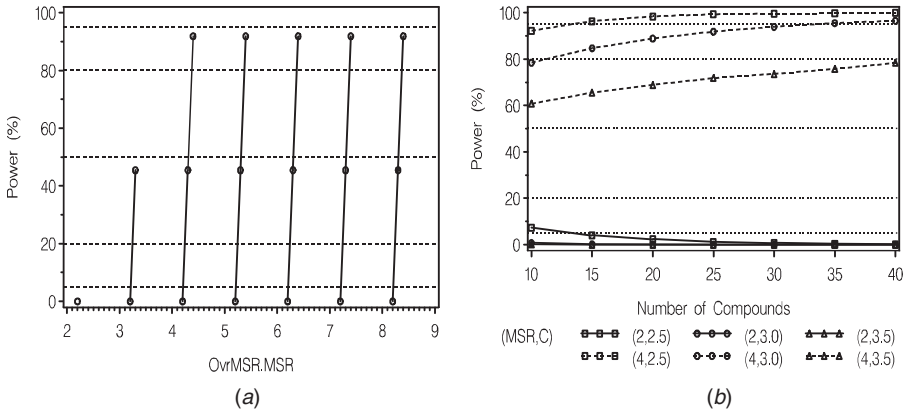


**Figure 9** (*a*) Power (%) that $\widehat{\text{MSR}} > 3$ versus OvrMSR.MSR (i.e., whole number indicates the OvrMSR, decimal indicates the MSR) for a sample of size 25 compounds. (*b*) Power versus number of compounds by MSR and *C* when OvrMSR > MSR.

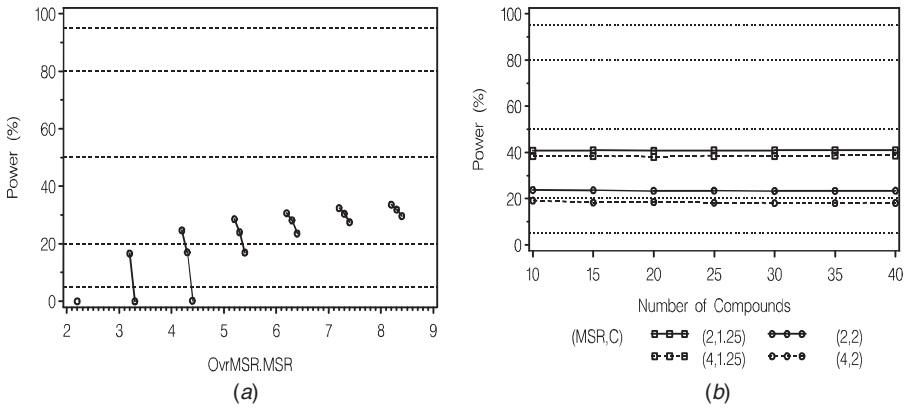


**Figure 10** (*a*) Power (%) that $\widehat{\text{MR}} > 1.5$ versus OvrMSR.MSR (i.e., whole number indicates the OvrMSR, decimal indicates the MSR) for a sample of size 25 compounds. (*b*) Power versus number of compounds by MSR and *C* when OvrMSR = 8.

$C = 1.5$. Figure 10*a* shows the effect of OvrMSR and MSR on power for $n = 25$ compounds and $C = 1.5$. The power is a very flat function of OvrMSR, and power remains below 40 percent for all values examined. Moreover, for a fixed OvrMSR, power decreases as MSR increases. Figure 10*b* shows the effect of *n*, MSR, and *C* on power for the extreme case of OvrMSR = 8. Note that power is invariant to sample size with this test, power decreases as MSR increases as noted previously, and that even a critical value of $C = 1.25$ only achieves about 40 percent power. However, the combination of MSR = 2 and OvrMSR = 3 will fail this test about more than 30 percent of the time, and

hence setting a low threshold is not recommended. The conclusion from this analysis is that test 2 [Eq. (5)] has inadequate power.

From Figure 8*a* we showed that for OvrMSR = 3 median $\widehat{\mathrm{LsA}}_{\max}$ values were between 2 and 3, depending upon the value of MSR, and from Figure 8*b* we showed that median $\widehat{\mathrm{LsA}}_{\max}$ values exceed 4 in the set MSR = 3 and OvrMSR > 4. Thus, in Figure 11 we examine the power of test 3 [Eq. (6)] around a critical value of $C = 3$. Figure 11*a* shows the effect of OvrMSR and MSR on power for $n = 25$ compounds. For OvrMSR ≥ 4 over 80 percent power is obtained in all cases except where MSR = 2. Note that in all such cases the power in test 3 [Eq. (6)] is significantly enhanced over the power in test 1 [Eq. (4)]. However, note that for OvrMSR = 3 power may exceed 50 percent in those cases where the across-run component of variation is negligible. Figure 11*b* shows the effect of $n$, $C$, and MSR on power when OvrMSR = 8. At MSR = 4 high power (>90 percent) is maintained for all sample sizes and choices of $C$. At MSR = 2 adequate power (>80 percent) is obtained only through setting a low threshold ($C = 2.5$). The conclusion from this analysis is that test 3 [Eq. (6)] has excellent power to detect high OvrMSR except in those cases of low MSR.

The overall recommendation is that rejection rules be based on a combination of tests 1 and 3, [Eqs. (4) and (6)] and that test 2 [Eq. (5)] should be avoided. Note that if the same critical value is used in both tests 1 and 3 [Eqs. (4) and (6)], then rejection in test 1 [Eq. (4)] implies rejection in test 3 [Eq. (6)]. Therefore, to avoid redundancy, the critical value for test 3 [Eq. (6)] should be set higher than that of test 1 [Eq. (4)], but the optimal settings cannot be determined without reference to the chemistry and

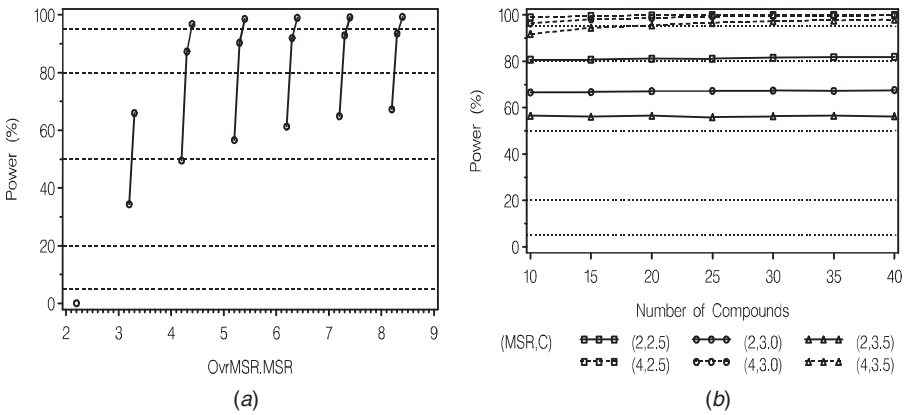


**Figure 11** (*a*) Power (%) that $\widehat{\mathrm{LsA}}_{\max} > 3$ versus OvrMSR.MSR (i.e., whole number indicates the OvrMSR, decimal indicates the MSR) for a sample of size 25 compounds. (*b*) Power versus number of compounds by MSR and $C$ when OvrMSR = 8.

biology requirements. Consequently, final settings are beyond the scope of this chapter.

***Acknowledgment*** The authors wish to acknowledge numerous discussions with Viswanath Devanarayan, Mark Farmen, Barry Sawyer, and Philip Iversen that have helped refine the replicate-experiment protocol since it was conceived.

## REFERENCES

1. Sittampalam, G. S., Iversen, P. W., Boadt, J. A., Kahl, S. D., Bright, S., Zock, J. M., Janzen, W. P., Lister, M. D. (1997). Design of signal windows in high throughput screening assays for drug discovery. *J. Biomol. Screen., 2*, 159–169.
2. Zhang, J.-H., Chung, T. D. Y., Oldenburg, K. R. (1999). A simple statistical parameter for use in evaluation and validation of high throughput screening assays. *J. Biomol. Screen., 4*, 67–73.
3. Taylor, P. B., Stewart, F. P., Dunnington, D. J., Quinn, S.T., Schulz, C. K., Vaidya, K. S., Kurali, E., Tonia, R. L., Xiong, W. C., Sherrill, T. P., Snider, J. S., Terpstra, N. D., Hertzberg, R. P. (2000). Automated assay optimization with integrated statistics and smart robotics. *J. Biomol. Screen., 5*, 213–225.
4. Eastwood, B. J., Farmen, M. W., Iversen, P. W, Craft, T. J., Smallwood, J. K., Garbison, K. E., Delapp, N. W., Smith, G. F. (Submitted for publication). The minimum significant ratio: A statistical parameter to characterize the reproducibility of potency estimates from concentration-response assays. *J. Biomol. Screen.*
5. Kenakin, T. (1997). *Pharmacologic Analysis of Drug-Receptor Interaction,* 3rd ed. Lippencott-Raven, Philadelphia, pp. 242–289.
6. Molulsky, H., Christopoulos, A. (2003). *Fitting Models to Biological Data Using Linear and Nonlinear Regression: A Practical Guide to Curvefitting.* Graphpad Software, San Diego, CA, pp. 259–260.
7. Ratkowsky, D. A., Reedy, T. J. (1986). Choosing near-linear parameters in the four-parameter logistic model for radioligand and related assays. *Biometrics, 42*, 575–582.
8. Gallant, A. R. (1987). *Nonlinear Statistical Models.* Wiley, New York, pp. 16–18.
9. Serfling, R. J. (1980). *Approximation Theorems in Mathematical Statistics.* Wiley, New York, p. 33.
10. DeLapp, N. W., McKinzie, J. H., Sawyer, B. D., Vandergriff, A., Falcone, J., McClure, D., Felder, C. C. (1999). Determination of [$^{35}$S] guanosine-5′-*O*-(3-thio) triphosphate binding mediated by cholinergic muscarinic receptors in membranes from Chinese hamster ovary cells and rat striatum using an anti-G protein scintillation proximity assay. *J. Pharmacol. Exper. Therapeut., 289*(2), 946–955.
11. Motulsky, H. (2003). *Graphpad/Prism® User's Guide, Version 4*. Graphpad Software, San Diego, CA.
12. Bishop, Y. M. M., Feinberg, S. E., Holland, P. W. (1975). *Discrete Multivariate Analysis: Theory and Practice.* MIT Press, Cambridge, MA, pp. 486–487.

13. Bland, J. M., Altman, D. G. (1986). Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet, I*, 307–310.
14. Carrasco, J. L., Jover, L. (2003). Estimating the generalized concordance correlation coefficient through variance components. *Biometrics, 59*, 849–858.
15. Eastwood, B. J. (2001). Comparing two measurement devices: Review and extensions to estimate new device variability. *Computing Science and Statistics*, *33*, 462–471.
16. SAS Institute (1999). *SAS ® Language Reference: Dictionary, Version 8*. SAS Institute, Cary, NC, pp. 226–626.
17. SAS Institute (1999). *SAS/GRAPH® Software: Reference, Version 8*. SAS Institute, Cary, NC, pp. 801–859.

# COUPLED LUMINESCENT METHODS IN DRUG DISCOVERY: 3-MIN ASSAYS FOR CYTOTOXICITY AND PHOSPHATASE ACTIVITY

Michael J. Corey
*Chromos*
*Seattle, Washington*

Robert J. Kinders

1 INTRODUCTION
Today's Assay Requirements
Fluorescence versus Luminance
Scintillation Proximity Assays
Coupled Enzyme Assays
Coupled Luminescent Assays as Alternatives to Fluorescent Assays
Chemiluminescent Assays
Coupled Bioluminescent Assay Methods
Measurement of Cytotoxicity by Release Assays
Measurement of Phosphatase Activity
2 THE DEATHTRAK 3-MIN CYTOTOXICITY ASSAY
Introduction to DeathTRAK
Modes of Use of DeathTRAK in Drug Discovery
Summary of Cytotoxicity/Proliferation Methods
3 PHOSTRAK: A VERY RAPID, HOMOGENEOUS GENERAL PHOSPHATASE ASSAY
Measurement of Phosphatase Activities
Introduction to PhosTRAK

*The methods described herein are covered by US Patent#6811990 and are licensed to Cell Technology, Inc., Mountain View, CA (www.celltechnology.com).

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

Applicability of PhosTRAK Method to Various Types of Phosphatases
Representative PhosTRAK Results with Diverse Phosphatases
Limitations Due to Complexity and Thermodynamic Considerations
PhosTRAK Summary
4 COUPLED LUMINESCENT ASSAYS IN PRACTICE: SPECIFICATIONS AND LIMITATIONS
Assay Specifications
Backward Reactions
5 SUMMARY AND CONCLUSIONS
6 FORMULATIONS
Reaction Master Mixes (Cocktails) and Dilution Buffers
PhosTRAK Formulations
Lytic Formulations
References

## 1 INTRODUCTION

### Today's Assay Requirements

The prototypical high-throughput screening (HTS) operation in today's drug discovery environment is directed to identification of modulators of enzyme activities or receptor binding. There was a time when the inhibition constant of a single transition-state analog was determined by a graduate student nervously keeping atmospheric $CO_2$ out of his vessel as he titrated evolved acid with excruciating care, and *receptor binding assays* referred to a postdoc grinding up a mouse organ to mix with her precious radiolabeled ligand. There was such a time—but today's practitioner of lead identification in a drug discovery enterprise feels little nostalgia for it. Current demands on discovery groups require that $10^5$ to $10^6$ or more molecules be tested in a single experiment, quickly, accurately, and at acceptable expense. The various revolutions in biotechnology of the past 30 years have largely given way to evolutionary changes, favoring faster, smaller, operationally simpler, and more information-intensive assay methods. Developments in instrumentation that have contributed to this process are outside of the current subject matter, but biochemical manipulations may also be applied, ideally leading to ever more quantifiable photons from fewer molecular events in a shorter time frame. This process is most readily apparent in the shift from chromogenic and other traditional methods to fluorescent, luminescent, and scintillation proximity assays, a trend that will accelerate as assay volumes enter the submicroliter range.

Coupled luminescent assays are only one class of an expanding array of sensitive methods. At present, coupled luminescent methods are underutilized, largely due to the lack of widespread understanding of their value or even of their existence. What follows is a brief treatment of several alternative means of generating many photons from small samples in a short time

interval. To understand why coupled luminescent methods are the most efficient and sensitive available in addressing certain types of problems, while they are marginal or unsuitable for others, some insight into the nature, theory, and practice of alternative techniques is needed.

## Fluorescence versus Luminance

From the scientist's viewpoint, nature has bestowed powerful gifts on the phenomenon of fluorescence, including the associated techniques of fluorescence quenching, environmentally sensitive spectral shifts that are subject to synthetic modification, fluorescence polarization (FP), and fluorescence resonance energy transfer (FRET). Some of these technologies are already widely disseminated, such as the use of FP in drug immunoassays (as in the Abbott TDx). It is safe to predict that for rapid, specific, and exquisitely sensitive measurements of intermolecular association events, fluorescent technologies will dominate the field for at least several decades. There are also many clever fluorescent methods of measuring enzyme activities, but enzyme catalysis is intrinsically different from binding processes in that enzyme-catalyzed reactions are almost always observed indirectly (i.e., by their products). To put it another way, in studying enzyme activity and inhibition, one frequently deals with the concentrations of particular molecules, rather than the presence or absence of molecular complexes. This important difference represents an opportunity for luminescent assay methods to play a role, since a variety of means of coupling the concentrations of specific molecules to light emission are known, and more are sure to be developed. Possible reasons for preferring a luminescent assay are: (1) The sensitivity can be even more exquisite than that of fluorimetry [1, 2]; (2) there is no need to develop or attach a fluorophore; (3) in many cases the readout is almost immediate, which is not the case with many fluorescent binding assays; (4) no expensive lamp is needed; (5) coupled luminescent assays can be made operationally simpler; and (6) for some reactions luminance may be the only available approach to a rapid, homogeneous format.

## Scintillation Proximity Assays

A thorough discussion of the pros and cons of scintillation proximity assays (SPAs) in drug development HTS is beyond the scope of this chapter, but certain aspects of SPA that may cause one to choose or reject this option in favor of fluorescent or luminescent assays can be briefly described. Assays employing radioisotopes have generally been operationally complex, even apart from the exposure and disposal problems, because a labor-intensive separation of "bound" from "unbound" isotopically labeled molecules has been necessary. The SPA concept avoids this difficulty and promotes radioactive assays to the realm of homogeneous assays suitable for HTS, by providing a means of distinguishing bound from unbound in situ. This is accomplished by including a solid-phase matrix (usually a bead) containing both the target

(unlabeled) molecule and a chemical scintillant. Binding of the labeled compound brings the radioisotope into proximity with the scintillant, leading to a luminance signal upon disintegration of a bound, but not an unbound, atom of the label. Because many ligands can be appropriately labeled and many targets can be successfully immobilized, this assay method is quite general. One finds in reviewing the literature that a high proportion of SPA articles report development of new applications. Thus SPA represents an approach that is quite likely to provide a successful, homogeneous method, even if one knows relatively little about the system and possesses only a conventional scintillation counter (or luminometer). However, the disadvantages of SPA compared to a fluorescent or luminescent method are important. Although technically homogeneous, in that no physical separations are required, successful application of SPA often requires multiple addition and incubation steps, some of which may be prolonged and involve shaking or other manipulations. For example, an SPA for inorganic phosphate has a very good limit of detection of ~$10^{-7}$ M phosphate but requires 2 h and multiple steps and is a single-point assay [3]. An HTS method of identifying protein tyrosine phosphatase 1B inhibitors makes clever use of a catalytically inactive active-site mutant to turn the problem into a binding assay, but the method employs seven reagent additions and three shaking incubations, illustrating both the flexibility and the operational difficulties attending SPA [4]. The entire Raf/MEK/ERK signal transduction pathway has been reconstituted in vitro and studied by a homogeneous method employing SPA, greatly improving the chances of identifying modulators of the pathway; however, the method takes 8 h, including a 6-h settling step [5]. Finally, there is some question as to whether SPAs provide the same quality of information in terms of sensitivity and dose–response curves as competing fluorescent methods [6]. In short, SPAs are an excellent way of turning apparently intractable screening problems into homogeneous assays; but they are unlikely to be the method of choice if one-step or two-step fluorescent or luminescent alternatives are available.

## Coupled Enzyme Assays

Coupled enzyme assays have a distinguished history. The "standard" clinical and research assays for many enzymes are coupled assays, often with spectrophotometric readouts. In a case that may be taken as typical of the thought processes involved in development of such an assay, researchers investigating antibiotic function were faced with the problem of detecting D-alanine and found a chromogenic answer [7, 8]. While this solution is effective in an academic research setting, it does not provide the throughput needed by drug discovery groups and is limited by the extinction coefficient of the molecule being measured. Another group subsequently developed an assay in which horseradish peroxidase (HRP) was generated in a coupled assay and yielded a fluorescent signal [9]. However, luminescent HRP substrates are also available

with greater sensitivity [10], and this may eventually prove to be a further stage in the evolution of D-alanine assays.

## Coupled Luminescent Assays as Alternatives to Fluorescent Assays

The issue of coupling of enzyme activities is in a sense independent of the fluorescence/luminance choice; however, in practice, fluorescent and coupled luminescent readouts often represent competing alternatives. This is because, although there are many fortuitous exceptions, most enzyme-catalyzed reactions and intermolecular associations do not yield conveniently robust and distinguishable fluorescence changes or luminance signals on their own. Thus the problem of detecting the molecular event often reduces itself to the question of how to cause the product(s) of the event to release photons (or inhibit their release). Introduction of a fluorophore that is perturbed by the event is one method; coupling of the products to light generation is another. Advocates of luminance technology could present theoretical reasons in support of a claim for greater sensitivity, including the following: (1) Luminance detection requires no lamp, so that the sample chamber can be darker, lowering the background; (2) unlike fluorescent methods, luminance detection requires no monochromator or bandpass filter to block the exciting wavelengths, so that every photon contributes to the signal; (3) luminescence is inherently isotropic, allowing the detector to cover as great a spherical arc as the engineering permits; this is not true of fluorescence, which requires angular separation of the detector from the incident light. The theoretical advantages in sensitivity appear to be borne out in practice, at least in some cases (for examples and analyses, see [1, 2]). Fluorescence has inherent counterbalancing advantages, including the possibility of pumping virtually any desired level of excitation energy into the system, and the more refined approach of using modern synthetic chemistry to tune the fluorophore(s) to precise specifications. It is hoped that the methods and approaches presented herein will help the reader to imagine how both possibilities might contribute to his or her particular goals.

## Chemiluminescent Assays

In modern parlance *bioluminescence* refers to light production by an enzymatic system (such as a luciferase) or other light-emitting protein (such as aequorin [11]), while *chemiluminescence* is photon emission by a (usually) small molecule in the absence of a catalyst. For a number of enzymes it is possible to design substrates that, upon catalysis, yield a product that can emit a photon. Thus the primary reaction generates light, and no coupling enzyme is required. This method has led to significant improvements in detection capabilities [12]. However, the limitations of the method in the HTS environment are often significant: (1) The enzyme substrate is unnatural, and inhibition effects are therefore hard to interpret; (2) in a related point, the $K_m$ of a chemiluminescent substrate is frequently much greater (i.e., the binding is much

weaker) than that of the natural substrate, allowing competition by weak competitive inhibitors that are not useful as lead compounds; (3) luminescent substrates often have special chemical, solvent, temperature, or pH requirements necessitating extra steps in the assay process; (4) the requirements mentioned under (3) may make the reaction incompatible with the biochemical system under study; and (5) chemiluminescent substrates are not available for many reactions.

## Coupled Bioluminescent Assay Methods

The two major types of luciferase used in coupled reactions are bacterial luciferases, whose light emission is coupled to oxidation of reduced nicotinamide adenine dinucleotide (NADH), and beetle luciferases, which use the free energy of hydrolysis of adenosine triphosphate (ATP) to generate light. Systems employing aequorin as the luminescent protein can also be engineered to yield signals in response to the activities of certain enzymes involved in $Ca^{2+}$ mobilization and signaling [13], including the very important class of G-protein-coupled receptors (GPCRs). These systems are of considerable interest in drug discovery. In general, there is little overlap between the $Ca^{2+}$ mobilization processes that are subject to these methods and the phosphate-dependent reaction series described herein, although a coupled luminescent assay for the guanosine triphophatase (GTPase) activity of GPCRs is not inconceivable.

***Coupled Methods Employing Bacterial Luciferases*** Bacterial luciferases [14] have yielded a number of useful coupled assay methods, including means of quantifying a range of small molecules with excellent sensitivity (ethanol [15], folate [16], glucose [17], lactic acid [18], bile acids [19]). Some reactions, including the reactions involving glyceraldehyde-3-phosphate dehydrogenase discussed below, allow a choice of either bacterial or beetle luciferases as coupling enzymes. The author's decision to concentrate on firefly luciferase has been based primarily on practical considerations, including the poor aqueous solubility of the long-chain aldehyde substrates of the bacterial enzyme, which may require a separate addition and mixing step; the tendency of the bacterial signal to decay more rapidly; the existence of stabilized and engineered systems employing the firefly enzyme; and the fact that the firefly enzyme has been more fully characterized. Moreover, methods of ATP detection are more likely to be generally applicable in drug discovery since the enzymes of critical interest that are involved in phosphate metabolism outnumber those that generate NAD(P)H.

***Coupled Assays with Firefly Luciferase*** Many in vivo reactions create ATP, and many others that are coupled to hydrolysis of ATP in vivo can be made to run backwards in vitro. In principle, the reaction rate of any enzyme involved in these pathways can be monitored by coupling production (or

consumption) of ATP to light generation by firefly luciferase. In practice, many of the potential coupled systems are unworkable in vitro, but a number of enzymatic reactions or reaction series have been combined with luciferase, some of which are discussed below. As with the chemiluminescent assays, some of the methods are intended for detection of small molecules, and others are adaptable to detection of either small molecules or enzymatic activities, depending on what is omitted, that is, what is the limiting reagent. See [20] for examples of very early uses of coupled methods with firefly luciferase for detection of small molecules.

*Role of Coupled ATP/Luciferase Reactions in Drug Discovery* Thousands of enzymes are necessary components of metabolism, but a large proportion of the enzymes of phosphate metabolism are specifically involved in information flow. As a result, many of them not only play essential roles in life-sustaining pathways but are also vital elements of reaction series that are subject to *information failure*. Many phosphorylation and dephosphorylation reactions are critical mediators of major human health problems such as cancer, diabetes, autoimmunity, and inflammation. These features make these enzymes highly attractive targets for drug development, based on the philosophy that subtle perturbations of the body's signaling systems, rather than gross, "brute-force" interventions in metabolic flows, are the best route to highly specific and effective drugs. Of the enzymes that are currently the major targets of HTS, only proteases are *not* directly involved in phosphate metabolism. The other major groups, including kinases, phosphatases, deoxyribonucleic acid (DNA) polymerases and many nucleases, phosphodiesterases, and the whole generic category of "ATPases," can be made to change the concentration of ATP in a reaction vessel and can thus yield a luminance signal via luciferase. Still other enzymes can be coupled to phosphate metabolism by the inclusion of "bridging" enzymes. If we restrict ourselves to the desirable trait of yielding a *positive* signal (i.e., an increase in luminance when the enzyme is active), it might appear that we have lost most kinases and polymerases for thermodynamic reasons, but this is not the case (see below). In short, over half of the assays carried out in today's drug discovery endeavors (on a frequency basis) are adaptable to ATP/luciferase-coupled systems.

*Advantages of Coupled ATP/Luciferase Systems* Many of the advantages of these systems can be defined in terms of what is *not* required. Typically a coupled luminescent assay employing the firefly luciferase can be developed as a one-step assay in which a reagent cocktail is added and measurement begins immediately (although multistep assays may be preferred for particular applications). This represents a significant savings in operational complexity, leading to lower labor, equipment, and materials costs and fewer failure modes. Because the assays are generally very fast, instrument utilization is minimized; in effect, more work can be done with fewer instruments and less space. This effect will be increasingly important as phased-array luminometers

become more common. These instruments can read entire plates in 2 s or less, but this speed is largely wasted if the biochemistry of the assay requires prolonged incubations or multiple steps. Coupled luminescent assays are frequently the only means of taking full advantage of the capabilities of these instruments. Generation of ATP by the reaction system also provides a continuous readout. This yields the advantages of (1) flexibility in data reduction, (2) the possibility of prolonged incubation to achieve the desired sensitivity, and (3) the opportunity for a second reading if the initial reading is somehow missed. The components of these coupled assays are often less expensive than the fluorophores, labeled peptides, and antibodies used in other assay methods; even using 96-well plates, it is quite realistic to expect reagent costs of $0.01 to $0.05 per well or less, and this cost should scale down with effective miniaturization. The homogeneous format also eliminates the need for washing and separation steps and reduces generation of hazardous waste. Finally, recent developments in stabilization of firefly luciferase and other methods of enhancing and prolonging luminescent signals will continue to add value to these methods.

*Adaptation of Coupled Luminescent Assays to High-Throughput Screening*
Because of their operational simplicity, it can be relatively simple to adapt a coupled luminescent assay to use in HTS, and the sensitivity usually allows reduction to a denser plate format. However, there are also special requirements that must be met. In using firefly luciferase, some consideration must be given to its stability. If available, one can employ a stabilized luciferase; otherwise, reductants are beneficial, and it may be necessary to keep the luciferase cold and protect it from light until use, depending on the wait time. Another issue that does not arise as frequently in other kinds of assays is the purity of certain assay components. Coupled assays that yield ATP require an ATP precursor, such as adenosine diphosphate (ADP), adenosine monophosphate (AMP), or cyclic AMP (cAMP), as a substrate. If the substrate is contaminated with ATP, even in small amounts, there will be a significant background signal. ATP can generally be separated from these other components by the use of any of several methods [22–24]. Similarly, commercial enzyme preparations are often contaminated with other enzymes. If one of the contaminating enzymes is the substance under test (as in the DeathTRAK assay; see below), then the user must purify the enzyme, inactivate the troublesome contaminant in a way that will not cause subsequent inhibition during the screening process, or accept a dynamic background signal.

Fortunately, these background sources are usually very constant, and as a result they generally have surprisingly small effects on the quality of the results, including the $Z$ values. However, to achieve this predictability, the user must be sensitive to such issues as the order of reagent addition. For example, if ADP is present in the reagent cocktail, and one of the coupling enzymes is slightly contaminated with one or more of the enzymes capable of transferring phosphate moieties among nucleotides, then mixing of the ADP with the

coupling enzymes should occur only very shortly before assay initiation; otherwise a high background must be expected. Finally, scientists involved in HTS are used to the idea that many fluorophores and other reagents must be protected from light, but a researcher developing or employing a coupled luminescent assay should try to develop a new way of thinking about light. In these assay methods, light is not merely a readout or a damaging form of radiation; it is also potentially a *reagent*. As such, it can drive a "backward" coupled reaction if care is not taken to avoid this possibility, and the result can be an undesirable *lag phase* in the reaction that can lower the quality of the data and degrade throughput (see Section 16.4).

*Limitations and Considerations* There are other considerations that may influence the choice of a coupled luminescent assay. Primary among these is the appropriate pairing of a screening strategy with the objectives of a study. As a simple example, suppose one is screening for a novel antibiotic for use against a specific bacterium, using midlog cultures. If one is willing to accept any effect that reduces the bacterial count, including bacteriostasis, wall or membrane lysis, metabolic poisons, and genotoxic effects, then it is hard to improve on a simple turbidometric growth assay (although the sensitivity of a coupled luminescent enzyme-release assay may be an advantage in working with early-phase cultures). However, if the researcher is interested in mechanism, or wishes to distinguish killers from bacteriostatic agents, or wishes to measure lysis and proliferation separately, then a more sophisticated approach may be required. An example from mammalian systems may also be warranted. Suppose a worker wishes to screen a completely random library for druglike characteristics against an enzyme target. The primary assay should be simple from a biochemical point of view since this will help to alleviate the problem of systemically generated false hits. This may be a good case for an SPA approach since these systems are often less sensitive to interferents. Moreover, since one has little a priori information about a given "hit," the secondary screening process should be rigorous and information intensive. Now suppose that one is screening with a *targeted* library, that is, a library of entities with at least one common structural element that is known to be selective for the desired enzyme. Biochemical simplicity is now less important as an assay parameter since a much lower proportion of false hits is to be expected. Analytical sensitivity and precision are now driving attributes of the primary screening method, while the secondary screening method should probe selectivity, including both false-hit modes of the primary screen and biochemical or biological characteristics that are independent of the readout of the primary. A very rapid, ultrasensitive coupled luminescent assay may be the best choice for the primary method.

The application of these principles to the use of coupled luminescent assays leads to two conclusions: (1) The more one knows about one's library, the more appropriate it may be to choose a rapid, ultrasensitive assay involving relatively complex biochemistry; and (2) if a coupled method is used with a

random library, then the secondary screening method should be directed to the elements of the system that are most likely to yield false positives (such as downstream signal-generating enzymes).

*Examples of Coupled Luminescent Assays Employing Firefly Luciferase* See [25] for an excellent source of theory, methods, and practical information about firefly luciferase, including coupled assays. The number of assay types in coupled-luminescent category is still small compared to the potential breadth of the field, but expansion is accelerating for at least two reasons: (1) As mentioned above, the drive to smaller samples and faster instrumentation places an increasingly great premium on rapid, ultrasensitive methods; and (2) the availability of thermostable luciferases adds considerably to the flexibility of these methods. Examples provided include assays for kinases and DNA polymerases, as well as the glyceraldehyde-3-phosphate dehydrogenase release cytotoxicity assay and phosphate/phosphatase assay, which are the experimental foci of this chapter.

KINASES Kinases are the second most important molecular target in HTS, after GPCRs. Promega sells an HTS kinase assay kit (Kinase-Glo) of which the readout is luciferase detection of the consumption of ATP by the kinase. The kit represents a successful approach, but coupled systems of this type are inherently burdened by the "negative-signal" characteristic. In other words, the kinase yields a *negative* luminance readout, and inhibition of the kinase *increases* the readout. Although in some cases negative-signal systems prove to be the only or best alternative, the problems of such methods are not merely the counterintuitive nature of the scheme. In such a scheme, a significant amount of the ATP must be consumed to generate a signal; otherwise one is faced by the statistical pitfalls of analyzing a small difference between two large numbers. But if significant amounts of ATP are consumed, it is easy to lose the "initial-rate" assumption that is dear to enzymologists. In practice, this means that conclusions drawn about kinetic parameters (such as inhibition constants) may not be as reliable as if they were measured differently. However, the most important drawback of a negative-signal system is one that may be invisible to the user: How does one interpret an absence of light? Did the kinase do its job and consume the ATP, indicating that the tested compound is inactive, or was this the next Captopril or Viagra, but one that yielded no signal in the assay because it adventitiously blocked the luciferase system as well? This problem is usually avoided in positive-signal systems, which are more likely to generate false-positive hits than false negatives. The user must judge the likelihood and potential cost of each type of failure mode in deciding whether to use a negative-signal assay.

The general impression is that a kinase is an enzyme that transfers a phosphate group from ATP to a target substrate molecule. However, kinases, like other enzymes, are not exempt from the principle of microscopic reversibility, and, with no special considerations in play (such as off-gassing or further

conversion of a product), kinase-catalyzed reactions are readily reversible in practice. This means that they can generate ATP, which means that their activity can be measured with firefly luciferase. Some scientists without specific enzymology training are uncomfortable with the idea of screening so-called reverse reactions; however, thermodynamic theory is absolutely clear on the point that if a molecule inhibits the forward reaction, it must inhibit the reverse reaction, and to the same extent. In any case, the *forward vs. reverse* distinction is itself not always clear-cut. (Is there a kinetic difference between an "alcohol dehydrogenase" and an "aldehyde reductase"?) Kinases were so named because the first described kinases activated proteins by phosphorylating them, thereby catabolizing ATP; but what of creatine kinase? This enzyme is very proficient at *synthesizing* ATP, and in fact the reaction is thermodynamically downhill in this direction under physiological conditions. A method of measuring creatine kinase activity by luciferase detection of evolved ATP was described some 25 years ago [26]. The original method is clever but not startling since the reaction is thermodynamically favorable. The idea was subsequently broadened in a more recent patent application to include reactions in which ATP synthesis is thermodynamically disfavored [27]. This category embraces nearly all kinases, although it does not mean that all kinases can be assayed in this way; however, the apparently unfavorable $\Delta G$ of the reactions successfully handled implies that the thermodynamic issue in itself is not a major barrier to development of "backward" kinase reaction schemes. Nevertheless, thermodynamics may still prove to be an important limitation on what can be achieved in other systems, especially those in which natural hydrolysis of ATP is not coupled to an energy-storing process that can be readily reversed in vitro. In any case, the contrast between the convenient, homogeneous one-step nature of this assay and the relatively slow, clumsy HTS methods in current use for kinase screening indicates that coupled luminescent methods have a promising future in this area.

PHOSPHOROLYSIS Scientists at Promega Corporation have developed a coupled luminescent assay (READIT) for the activity of nucleic acid polymerases, with extensions to detection of specific sequences, including single nucleotide polymorphisms [28–30]. This is a highly promising area in which coupled luminescent assays are likely to surpass competing methods in efficiency and accuracy. Given the central importance of nucleic acids in biology, the authors believe that the use of this technology is likely to expand greatly in the next few years. General awareness of coupled luminescent methods may be raised as a side effect.

PHOSPHODIESTERASES AND BEYOND Although "biologicals" are making impressive inroads in terms of investigational new drug (IND) submissions, much of drug discovery can still be regarded as a process of characterizing small molecules and their effects. Since this is not far removed from the process of *detecting* small molecules, synergies may be achieved and methods

borrowed from seemingly distant fields for use in drug discovery. Given the wide range of enzymes involved in phosphate metabolism, a little imagination can readily yield a new idea or application that appears far removed from the initial objective. A single hypothetical example originating in the world of food science, where detection of caffeine is of interest, is sufficient to demonstrate the principle. Caffeine inhibits a cAMP-dependent phosphodiesterase (PDE); cAMP is a product of the nominal reaction of adenylate cyclase, an enzyme that ordinarily consumes ATP. However, it should be possible to employ mass-action principles by supplying cAMP and pyrophosphate to the adenylate cyclase in the presence of PDE. If caffeine is not present, the PDE will exhaust the cAMP and depress the luminescent signal. However, if the PDE is inhibited by caffeine, the cAMP will give rise to light production instead. (Here a negative-signal assay design is presented, despite the potential disadvantages cited above.) Caffeine itself is not of great current interest in drug discovery, but other PDE inhibitors are of major significance [31–33]. Many such schemes are possible.

## Measurement of Cytotoxicity by Release Assays

Cytotoxicity has been measured by monitoring the release of various molecules from damaged cells for several decades. A modern example of such a method is the coupled fluorescent lactate dehydrogenase (LDH) release assay CytoTox-ONE from Promega. This method is operationally very convenient, has moderate sensitivity [limit of detection (LoD) ~200 nucleated mammalian cells], is moderately rapid (10 min), and is quite expensive, listing at >$25.00 per 96-well plate. Another approach is to transfect target cells with an enzyme that can then yield a fluorescent or luminescent signal upon cell lysis. Two enzymes that have been employed in this way are firefly luciferase itself and B-galactosidase [34]. The method obviously requires a separate transfection step, but this can be done in a bulk fashion and adds little to the cost or labor required. Unfortunately, neither luciferase nor B-galactosidase is well suited to the purpose; luciferase has a short half-life under test conditions, while B-galactosidase is not readily released into the assay medium by membrane rupture alone and requires additional steps for a successful assay. Release of naturally present alkaline phosphatase can be monitored by means of a luminescent substrate, but only a few types of cells release measurable quantities of this enzyme on lysis [35].

The important glycolytic enzyme glyceraldehyde-3-phosphate dehydrogenase (G3PDH) has also been used for decades to measure cytotoxicity by spectrophotometric means [36]. The original coupled luminescent assay method for G3PDH was published in 1997 [37]. This method was extremely sensitive, with an LoD of 0.03 nucleated mammalian cell or 1 rabbit erythrocyte. However, it was impractical for use in HTS, not only because three separate transfers were involved but also because it was not a homogeneous method:

the Reagent cocktail was not adapted for use with live cells, which had to be removed by centrifugation or filtering to avoid a false signal due to leakage from these cells. The G3PDH assay method known as DeathTRAK was developed to address these problems [38].

### Measurement of Phosphatase Activity

Currently the most widespread methods of assessing phosphatase activity in drug discovery programs are the use of radioactive phosphorus tracers (especially with SPA) and antibody-based detection of phosphorylated or dephosphorylated target molecules. The coupled luminescent method proposed herein, PhosTRAK [38], is a very rapid and sensitive alternative. For the phosphatases tested to date, the assay is truly a one-step method: All reaction components, the phosphatase, and modulators are mixed and the readout is obtained as the phosphatase reaction proceeds (see Section 16.3).

## 2 DEATHTRAK 3-MIN CYTOTOXICITY ASSAY

### Introduction to DeathTRAK

DeathTRAK is a homogeneous, one-step assay (see Fig. 1 for a depiction of the biochemical scheme of DeathTRAK). The limit of detection is well under one nucleated mammalian cell. (Means of achieving still greater sensitivity are discussed in Section 16.4.) As an essential glycolytic enzyme, G3PDH is abundantly present in all known cells, and the DeathTRAK method is therefore applicable to quantification of any cell type. The two major modes of use are in assessment of membrane rupture or damage and in measurement of total biomass. It should also be noted that like other release methods, DeathTRAK is poorly suited to detection of intracellular events, including fatal events, that do not affect membrane integrity. For example, release assays are not generally the first choice for detecting apoptosis, although they may well be useful in distinguishing apoptosis from necrosis if used in combination with a metabolic assay that is sensitive to intracellular events.

Three enzymatic reactions are depicted in the biochemical scheme of DeathTRAK presented in Figure 1. G3PDH is the test enzyme. In the normal flow of glycolysis, the free energy associated with the reducing capacity of the C–H bond of carbon-1 of glyceraldehyde-3-phosphate is captured in the formation of ATP from ADP and inorganic phosphate by the sequential reactions catalyzed by G3PDH and phosphoglycerokinase (PGK). The reaction series is exergonic, and the components are all readily available. Firefly luciferase and the small molecules needed for its activity are included in the reaction cocktail. The coupled reaction series produces light immediately upon mixing, assuming there is a source of G3PDH.

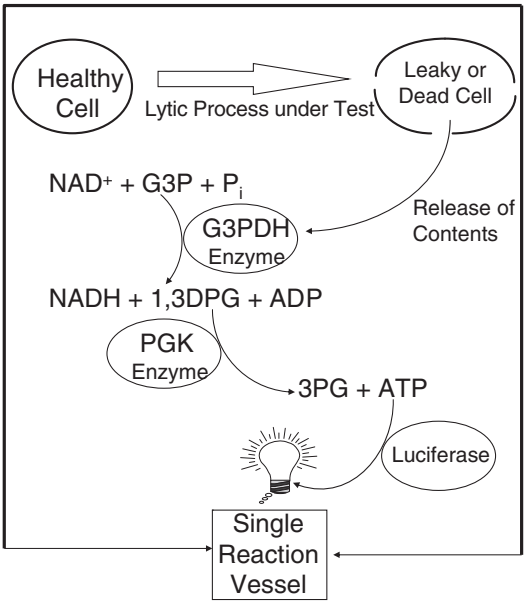


**Figure 1** Biochemical scheme of DeathTRAK homogeneous cytotoxicity assay. $NAD^+$, nicotinamide adenine dinucleotide (oxidized form); G3P, glyceraldehyde-3-phosphate; $P_i$, $PO_4^{3-}$; G3PDH, glyceraldehyde-3-phosphate dehydrogenase; NADH, nicotinamide adenine dinucleotide (reduced form); 1,3DPG, 1,3-diphosphoglycerate; ADP, adenosine triphosphate; PGK, phosphoglycerokinase; 3PG, 3-phosphoglycerate; ATP, adenosine triphosphate.

## Modes of Use of DeathTRAK in Drug Discovery

***Applications of DeathTRAK to Measurement of Cytotoxic Processes*** Many death modes lead to membrane rupture, including natural killer (NK), lymphokine-activated killer (LAK), cytotoxic T lymphocytes (CTL), and complement attack, pore-formers, detergents, various other antibiotics, and osmolysis. Numerous molecules that modulate these processes are under consideration as drug candidates, and other aspects of the processes await investigation by drug discovery groups. Thus the potential applications of DeathTRAK in screening for drugs that modulate natural cytotoxicity are widespread. DeathTRAK and the earlier glyceraldehyde-3-phosphate dehydrogenase/phosphoglycerokinase/luciferase (GPL) assay were originally developed to measure complement activity [39]. Figure 2 shows the results of a complement assay using DeathTRAK, in which the effects of an anti-Factor I antibody were assessed. Figure 3 shows the results of an assay of cytotoxic T lymphocytes (CTL), selected and expanded by the proprietary Rapid Expansion Method (CellExSys, Inc.; [40] and references therein), using the same assay protocol. The DeathTRAK method appears to have major advantages

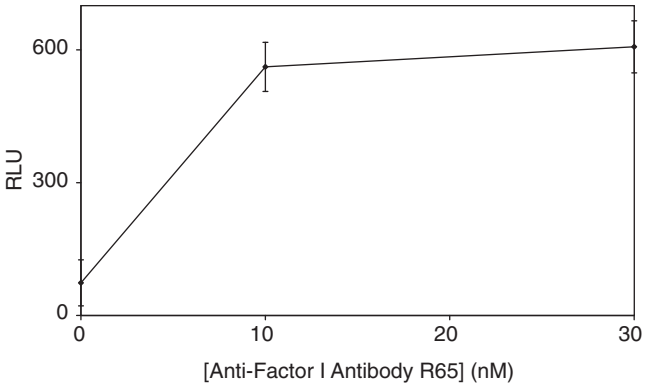


**Figure 2** Enhancement of complement-mediated lysis of PC-3 prostate cancer cell line measured by DeathTRAK. Human complement serum (Sigma) was preincubated with the indicated concentrations of rat monoclonal antibody R65 (Alidex, Inc., Redmond, WA) against human complement Factor I, following by complement assays as previously described [39].

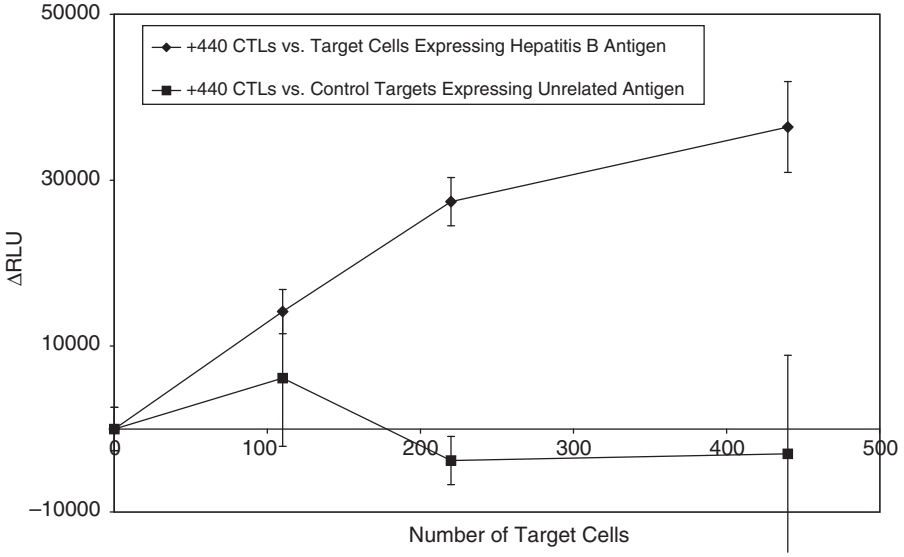


**Figure 3** Measurement of cytotoxic T lymphocyte activity against autologous cells expressing specific hepatitis B antigen. CTLs were grown and isolated by the rapid expansion method, which is proprietary to CellExSys, Inc. and is described in [40]. Targets were EBV-immortalized lymphocytes from the same patient, infected with a vaccinia vector containing a recombinant gene for an HBV-specific antigen. Control targets were infected with an unrelated antigen. The infection process induces a certain degree of leakiness, which accounts for most of the scatter seen in the control reactions. Cells at 10× were incubated for 4 h at 37°C, whereupon 1/10 of the reaction volume (5 μL) was transferred to 45 μL of DeathTRAK cocktail (Section 16.6) and read on a TopCount scintillation counter. Reported data are the differences between the initial read and a subsequent read at 2.33 min.

over the standard $^{51}$Cr-release assay in assessment of CTL activity, some of which were enumerated above.

***Biomass Measurement (Competitive Comparison)*** Nearly all assays commonly referred to as *proliferation assays* actually measure either total biomass or integrated metabolism. Currently, the most important true biomass assays used in drug discovery are turbidometry and measurements of ATP release. As mentioned above, turbidometric methods are very good ways of quantifying bacteria and certain other microbes under controlled circumstances. However, they yield no mechanistic information and do not distinguish between healthy, morbid, and dead cells. This lack of specificity is more serious in dealing with mammalian cells, especially since many types are adherent or semiadherent, and cell debris may interfere with the turbidometric signal.

*ATP Release Assays* ATP release is a much more sensitive way of measuring numbers of live cells than turbidometry, but since ATP has a short half-life outside of living cells, these are actually *viability* assays. Using this method, it is possible to obtain separate signals for prokaryotes and eukaryotes by using so-called somatic cell lytic agents, which do not lyse prokaryotes, followed by either lytic detergents or solutions specific for the type of prokaryote under study (see below).

The advantages of ATP release assays are important [41]. The method is rapid, relatively sensitive, and linear over a very wide range. It also has the important advantage relative to more complex release assays that it has been well characterized in a number of systems. However, the disadvantages are also sufficient to motivate a researcher to choose a different assay under some circumstances. The rate of ATP release from various cell types is surprisingly variable, possibly because so much of the compound is trapped inside of various intracellular organelles; multiple membrane systems must therefore be ruptured before a consistent readout can be obtained. Thus the speed and convenience of the method are good but not always excellent. The sensitivity is also approximately 30- to 100-fold poorer than that available from coupled enzyme systems. However, the most important drawback is the inherent statistical flaw in using these assays for measurements of cytotoxicity, in that for sensitive detection of cell death, one must subtract two large numbers, each with its associated error, to obtain the signal strength. Finally, many of the commercial ATP release kits are expensive, a significant limitation in the HTS setting.

*Metabolic Assays* The other important type of *proliferation* assay is really a method of measuring metabolic rate. These assays include, for example, the MTT, XTT, WST-1, and Alamar Blue techniques. (The BrdU-incorporation assay [42, 43] as currently practiced does not qualify as a high-throughput method.) All of these methods rely on chemical reactions that take place in intact mitochondria. When mitochondrial transmembrane redox potential is

lost, the reactions slow or stop. Thus the coupling between the signal and the viability of the cells under study is quite good. However, these assays are inherently flawed in several ways [44]. First, the signal obtained represents the *integral* of metabolism. If the researcher performs a 2.5-h experiment, and all the cells suddenly die at 2.3 h, the signal will look almost the same as if they were still alive. (Cytotoxicity can easily exhibit this type of nonlinear time dependence if one is studying a highly cooperative system such as complement.) Moreover, some fairly broad categories of compounds can cause direct changes in the redox signal in the absence of cytotoxic effects [45, 46], and the test reagents are believed to interfere with metabolism in some cases. The assays are also ill-suited to measurements of cytotoxicity for the same statistical reasons associated with ATP release assays.

*DeathTRAK in Biomass Measurement* A rapid, ultrasensitive assay for total biomass may be an attractive alternative to a metabolism-based viability assay or cell-counting procedure. Speed, accuracy, and easy adaptability to automation are possible reasons. G3PDH can be released from cells and measured in a one-step procedure. Because the enzyme is both catalytically efficient and abundantly present, a signal can generally be obtained in 30 s or less, although the researcher may decide on a 3-min assay to attain the specified sensitivity. In fact the assay can be continued until the desired sensitivity is achieved. Figure 4 shows the results of 3-min assays of dilutions of detergent-permeabilized cells. Data reduction can be accomplished by automated linear fitting of the time-dependent readouts; however, single-point measurements after a fixed length of time are also quite valid. This is illustrated in Figures 2 and 5, in which the linear-fit method, using all data collected over 20 min, can be compared with a single readout at 2.6 min (see Fig. 2). The two graphs yield the same information, and this is generally the case except at very low signal strength.

*DeathTRAK Cytotoxicity/Proliferation Dual Mode* Unlike ATP, the G3PDH molecule is relatively stable in a cell culture supernatant and can be stabilized further. Because of this, it is possible, following a challenge such as a potential antibiotic or candidate cancer drug, to measure both cytotoxicity (as represented by total membrane lysis or damage) and live biomass in the same sample, simply by adding detergent and remeasuring. This is an alternative to viability/cytotoxicity methods reported in the past, which generally have the same drawbacks as the independent assays that are being combined (e.g., [3-(4,5-dimethylthiazol-2-yl)]-5-(3-carboxymethoxyphenyl)-2-(4-sulfophenyl)-2H-tetrazolium (MTS) reduction and LDH release [47]). The dual-mode DeathTRAK process requires 6 min and is depicted schematically in Figure 6. Figure 7 shows an example of the results obtained with this dual-mode *cytotoxicity/biomass* or *cytotoxicity/proliferation* assay. In this case the assay method was used to measure the dose that is lethal to 50 percent of test subjects ($LD_{50}$) (cytotoxicity) of a detergent (Nonidet P-40), which was also used

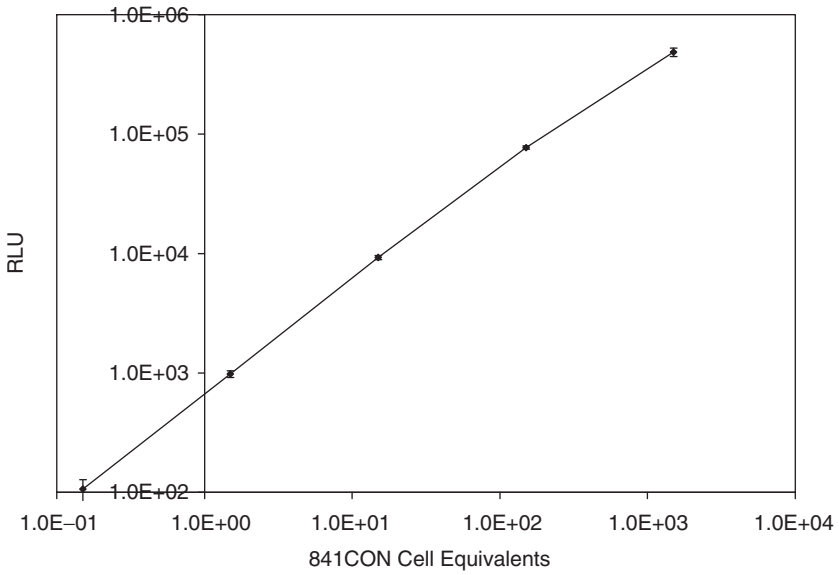


**Figure 4** Linear response of DeathTRAK in cell counting. 841CON (a normal human colonic epithelium-derived cell line) cells were counted, adjusted to $3 \times 10^6$/mL, killed by addition of 0.2% Nonidet P-40, and diluted to yield the indicated cell equivalents in 5-μL aliquots, which were mixed (in triplicate) with 45 μL of DeathTRAK cocktail (Section 16.6). A two-point readout is shown, consisting of the difference between the reading at 2.7 min and the initial reading. The background rate has been subtracted from all points. Error bars (standard deviations) are shown but are small. The 0.15-cell point reading was approximately 50 percent higher than background and was statistically distinguishable from background by *T* test ($p$ = .0014), with a *Z* factor of 0.13. The *Z* factor of the 1.5-cell point was 0.78.

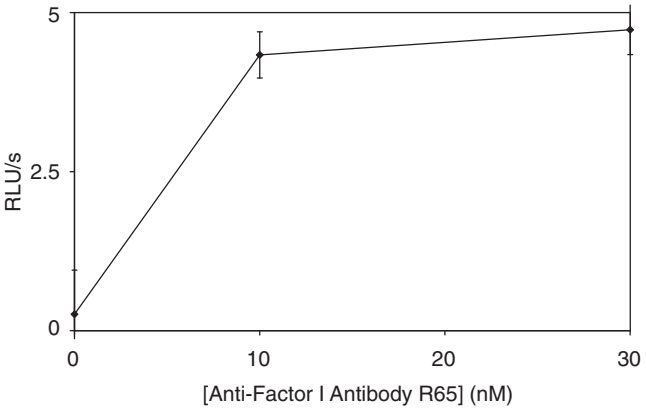


**Figure 5** Antibody-mediated effects on complement lysis of PC-3 cells measured by one-step DeathTRAK assay (linear-fit analysis). The same experiment as in Figure B is depicted. Linear fits of data gathered up to 20 min after initiation are graphed.

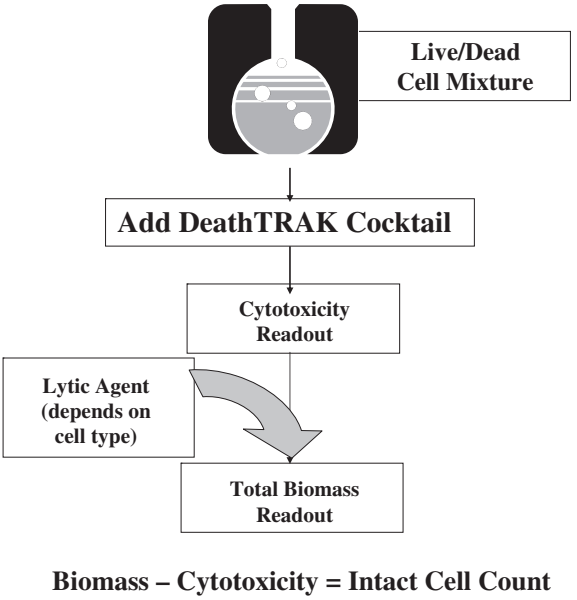


**Figure 6** Scheme of the DeathTRAK cytotoxicity/proliferation dual mode. The data obtained directly are the cytotoxicity measurement (i.e., enzyme released due to the process under test) and the biomass or total cell count (i.e., the sum of the cytotoxicity measurement and the viability readout at the end of the experiment). The viability readout is obtained by subtracting the first value from the second.

as the total lysis agent for the proliferation measurement. One might expect some evidence of interference between the two modes when the same substance is used as both test agent and assay reagent, but fortunately, although DeathTRAK is inhibited about 45 percent by 0.2% Nonidet P-40, this inhibition effect is nearly independent of detergent concentration above 0.2 percent, and very little interference was seen. Thus the "proliferation" signal is virtually constant, reflecting the constant total number of cells seeded into the wells. The calculated $LD_{50}$ obtained from the cytotoxicity readout of the experiment is in agreement with the known critical Micelle concentration of Nonidet P-40. Equivalent results were obtained for the HL-60, PC-3, and T-24 cells lines, except that killing T-24 required a significantly higher concentration of Nonidet P-40 (~2.5-fold greater). In a library-screening paradigm, one can obtain both cytotoxicity (representing direct lysis or membrane damage) and proliferation data (represent static or growth inhibitory effects) from a single plate in the same fashion.

### CYTOTOXICITY/PROLIFERATION MEASUREMENTS OF GRAM-NEGATIVE BACTERIA

Gram-negative prokaryotes may be studied in the same way. Figures 8 and 9 depict respective cytotoxicity and biomass measurements of *Escherichia coli* strain K1 (generously provided by Dr. Craig Rubens of Children's

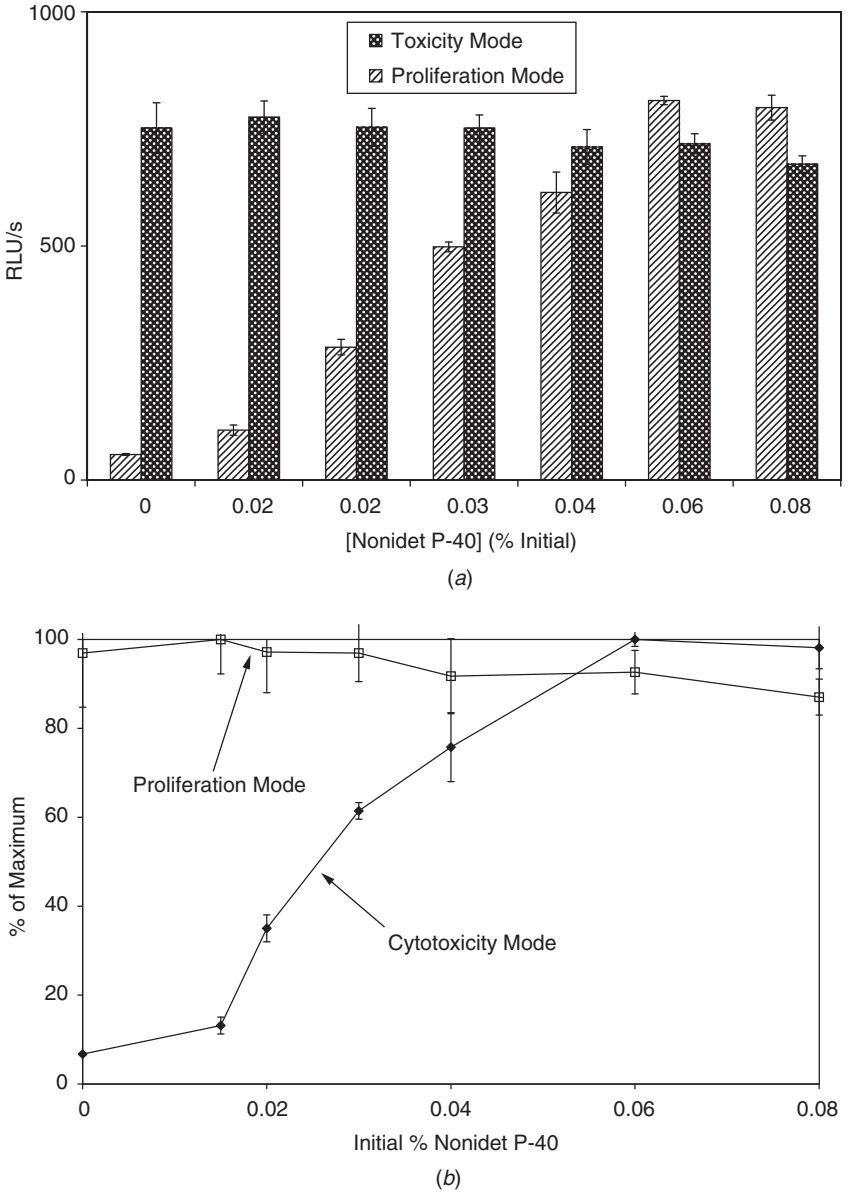


**Figure 7** Cytotoxicity/proliferation mode with 841CON cells. (*a*) The detergent Nonidet P-40 acted as both the toxin and the total lysis reagent. Detergent was added at the indicated concentration to 1000 841CON cells that had grown overnight in 100 μL IMDM, whereupon standard DeathTRAK cocktail (100 μL) was added and luminance was read for 3 min; 10 μL of 0.42% Nonidet P-40 was then added to each well and the proliferation measurement was taken for 3 min. The data reported are the linear fits from 1 to 3 min of each data set. All time-linear correlation coefficients were >0.999. Note that the displayed errors (standard deviations) incorporate both assay scatter and errors in plating the cells. (*b*) A simple visual presentation, similar to a screening report, of the respective effects on toxicity and proliferation of the evaluated drug (Nonidet P-40) in the assay.

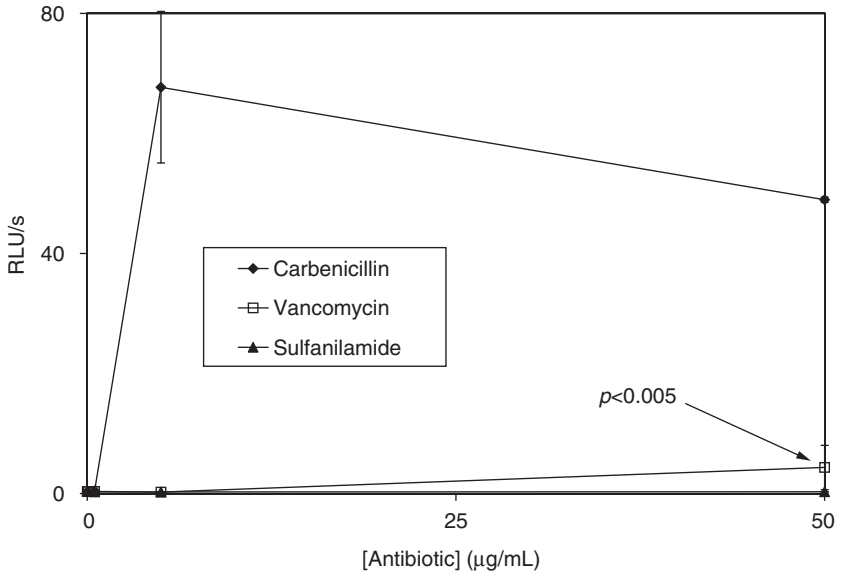


**Figure 8** DeathTRAK cytotoxicity measurements of the effects of three antibiotics on *E. coli* strain K1. Bacteria were grown to midlog phase, washed into PBS, and resuspended to an $A_{600}$ of 1.549 (~$2.18 \times 10^8$/mL). These were further diluted to a nominal $2 \times 10^5$/mL, whereupon a 10 percent volume of 10 mM dithiothreitol/1 percent PICGUWS (a protease inhibitor; Sigma) was added to the cells; 55 µL of cells were distributed to each well and antibiotics or vehicle (PBS) were added to the indicated final concentrations (reactions in duplicate). After 3 h, 45 µL of the standard DeathTRAK cocktail was added and the luminance was read for 2.5 min. The linear fit is reported (correlations were 0.99 or greater except for one, 0.97).

Hospital and Regional Medical Center, Seattle, Washington) challenged by carbenicillin, vancomycin, and sulfanilamide antibiotics. An estimate of the $LD_{50}$ of carbenicillin can be made from the data shown in Figure 9, but no mechanistic information is available; however, the cytotoxicity wing of the same assay shows that carbenicillin causes a loss of cell wall integrity as part of the killing process. Vancomycin has only a small effect on viability of this strain under these conditions but could be identified as a hit in the cytotoxicity wing, since the release data, though much lower than the carbenicillin data, are significantly different from zero (Fig. 8). As expected, sulfanilamide was not toxic to *E. coli* K1 under these conditions.

CYTOTOXICITY/PROLIFERATION MEASUREMENTS OF GRAM-POSITIVE BACTERIA

Similar data are shown for a Gram-positive organism in Figures 10 and 11. The proliferation/viability panel shows that all three antibiotics were effective against this group-A streptococcus (also provided by Dr. Craig Rubens), but the $LD_{50}$ are distinct and are easily estimated with considerable precision

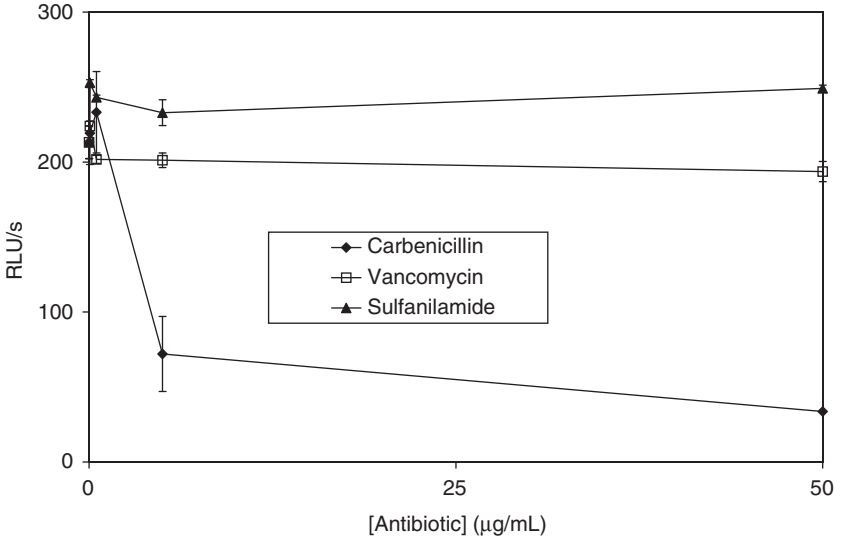


**Figure 9** DeathTRAK proliferation/viability measurements of the effects of three antibiotics on *E. coli* K1: 10 µL of $G^-$ lytic cocktail (Section 16.6) was added to the reactions shown in Figure 8 and the luminance was read for a further 2.5 min. Minimum linear correlation: 0.997.

from the data obtained in this 3-min assay. The toxicity results for the Gram-positives are more complex, and here it should be noted that Gram-positives are "leakier" than Gram-negatives in DeathTRAK assays. Thus, the toxicity data for Gram-positives amount to a complex average of killing effects, cytostatic effects, and leakage effects. It might be concluded that such a seriously convoluted readout is not useful, but when the data are combined with the viability observations, there may be mechanistic information available that is not apparent from either alone. One interpretation of the toxicity data is that gentamicin kills most of the cells (thereby reducing total synthesis of the G3PDH test enzyme and lowering the overall signal) and makes the rest somewhat leaky; carbenicillin makes them very leaky; and vancomycin is the most effective killer but causes no leakiness beyond the natural rate. Whether or not this sort of the analysis is worthwhile in a given system is a question that must be answered by the individual researcher, but in any case, an antibiotic screening strategy in which baseline readouts for cytotoxicity and viability are established during the development work, and significant anomalies in *either* type of readout are investigated further, is likely to miss fewer useful molecules than methods employing single readouts. Note that if both kinds of data are desired in a very rapid assay, it is necessary to lyse the cells very quickly between the two measurements. This can be accomplished for both the Gram-

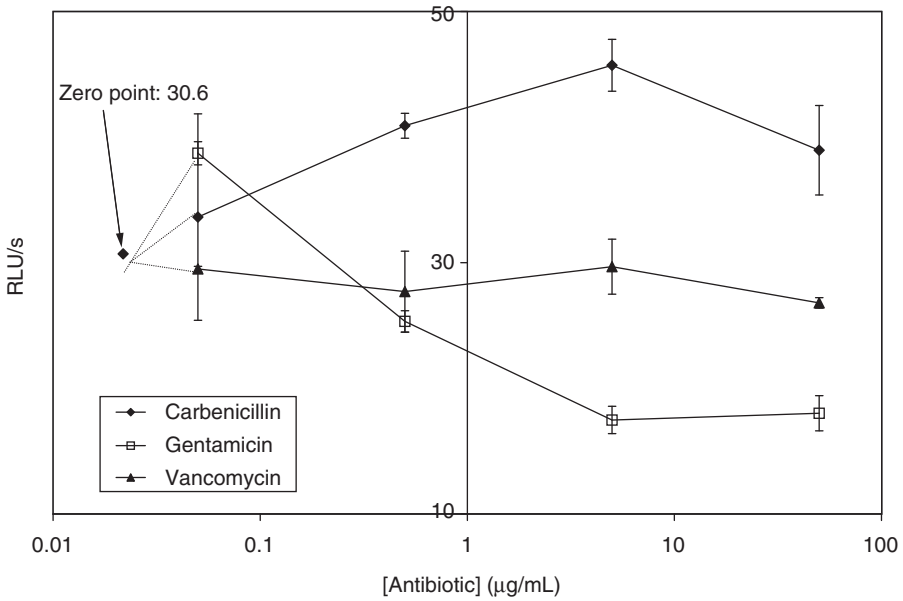


**Figure 10** Cytotoxicity readout of group-A streptococcus challenged with three antibiotics, measured by DeathTRAK. Group-A Streptococcus (GAS) were grown overnight in THY medium, washed 2X with THY, and resuspended at approximately $4 \times 10^6$/mL in the same medium; 50-μL aliquots of the cells were transferred to microtiter wells and antibiotic or vehicle (PBS) was added in 5 μL (duplicates). GAS were then incubated for 90 min at 37°C with 240 rpm shaking. The DeathTRAK cocktail (45 μL) was then added directly and luminance was read for 4.3 min; linear fits are reported. Toxicity readout is a sum of natural leakiness of this organism and leakiness due to membrane damage. Error bars are standard deviations.

negatives and Gram-positives studied in these experiments, but the suggested lytic formulations are different (see Section 16.6), and there is no certainty that the same formulations will be effective with other species or strains. Different bacteria, and perhaps even different strains of the same species, may be resistant to lysis by particular detergents and/or pore formers, and new lytic formulations may be needed when such organisms are employed as targets. This stands in contrast to the situation with mammalian cells in which all cell types tested were very rapidly lysed by 0.2% Nonidet P-40.

## Summary of Cytotoxicity/Proliferation Methods

Apart from speed, the major parameters influencing the choice of a cytotoxicity assay for use in HTS in a drug discovery paradigm include sensitivity, ease

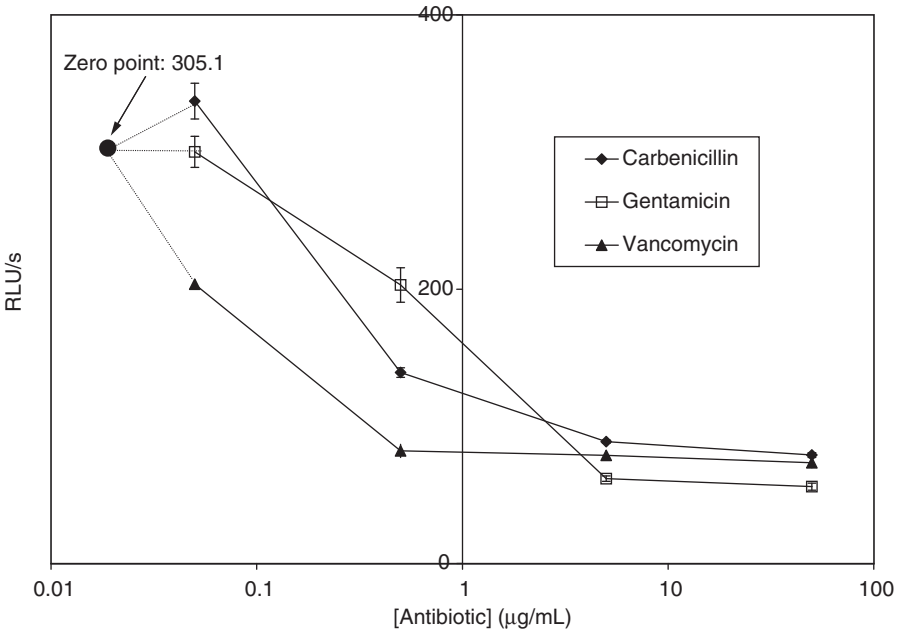


**Figure 11** Proliferation/viability readout of group-A streptococcus challenged with three antibiotics, measured by DeathTRAK: 10 µL of $G^+$ lytic cocktail (Section 16.6) was added to the reactions shown in Figure 10 and the luminance was read for a further 2.3 min; the linear fit is reported. Error bars (standard deviations) are displayed but most are small.

of operation and adaptability to HTS, cost, information quality, and statistical characteristics. From many of these viewpoints, a coupled luminescent assay such as DeathTRAK has a number of clear advantages. However, in managing such a program, the extent of validation and historical acceptance of a procedure also comes into play, and these considerations must be weighed against the perceived benefits.

## 3 PHOSTRAK: A VERY RAPID, HOMOGENEOUS GENERAL PHOSPHATASE ASSAY

### Measurement of Phosphatase Activities

The reaction catalyzed by phosphatases is not the simple reverse of the kinase reaction since no ATP is involved in the phosphatase reaction; nevertheless, in the biological sense, phosphatases may be considered as the dephosphorylating counterpart of kinases, and their importance in biological signal transduction is comparable, if not equal, to that of kinases. The search for mod-

ulators of phosphatase activities with druglike characteristics is an intensive and accelerating part of current drug discovery efforts [48, 49]. A substantial portion of the screening is carried out by methods involving specific antibody recognition of reactants or products [49, 50]. While there are various levels of sophistication associated with these methods, ranging all the way from immunoprecipitation [51] to advanced FRET and FP techniques (some of which employ novel high-affinity ligands instead of or in addition to antibodies [49]), these methods are generally slow because the association process is slow, frequently require separations and/or multiple addition, incubation, or mixing steps (although FP assays are often operationally simple), and are often expensive because of the antibodies and chemical conjugates involved. One alternative method that is not available with kinases (without additional coupling enzymes) is direct detection of liberated inorganic phosphate. This possibility is exploited in the malachite green assay [52]. This method has the advantage of using natural substrates, and the materials are inexpensive, but the method is a fairly slow, multistep process. Finally, a rapid chemiluminescent assay has been developed [53]; this assay yields speed and sensitivity, but the major disadvantage is that this technique employs a substrate that is so different from the natural substrates of most phosphatases that any inhibition data gathered would be highly suspect, simply because for most phosphatases the $K_m$ of the unnatural substrate is so poor that the reaction is too easy to inhibit. Moreover, the method requires multiple steps. Thus there is still a need for a rapid, sensitive, homogeneous assay that can make use of natural or quasi-natural substrates.

## Introduction to PhosTRAK

PhosTRAK is a coupled luminescent method of measuring the concentration of free phosphate, and, by extension, of phosphatase activity [38]. The biochemical scheme of PhosTRAK is related to that of DeathTRAK (compare Figs. 1 and 12), and similar speed and sensitivity are achieved; however, performance of PhosTRAK is distinct in a number of respects. Phosphate-free buffers must be used since free phosphate (hereafter, $P_i$) is the limiting reagent. A more significant difference is that DeathTRAK is a small set of related methods (depending on the cell type), whereas PhosTRAK is essentially a template on which a specific phosphatase assay can be developed.

## Applicability of PhosTRAK Method to Various Types of Phosphatases

Phosphatases are a large class of enzymes with varying characteristics; some exhibit metal dependence, while others, especially the class known as acid phosphatases, have pH activity profiles that are significantly different from those of most enzymes. Protein phosphatases have been recognized as critical

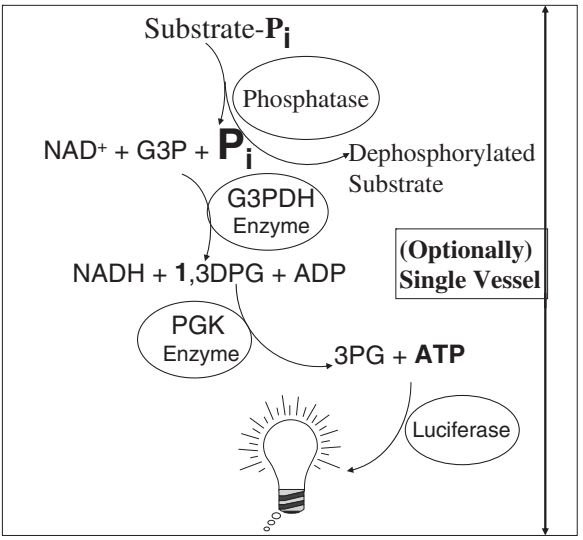


**Figure 12** Biochemical scheme of PhosTRAK. $NAD^+$, nicotinamide adenine dinucleotide (oxidized form); G3P, glyceraldehyde-3-phosphate; $P_i$, $PO_4^{3-}$; G3PDH, glyceraldehyde-3-phosphate dehydrogenase; NADH, nicotinamide adenine dinucleotide (reduced form); 1,3DPG, 1,3-diphosphoglycerate; ADP, adenosine triphosphate; PGK, phosphoglycerokinase; 3PG, 3-phosphoglycerate; ATP, adenosine triphosphate.

mediators of signaling pathways for some time (see [54–58] for recent reviews of aspects of this field), but many important phosphatases act on small molecules, such as inositol and its derivatives [59, 60] or phosphatidic acids [61, 62]. It is evident from work described herein and by analogy with other assay methods that a fairly small range of substrates is sufficient to handle a large subset of the protein phosphatases [e.g., Thr-Arg-Asp-Ile-Tyr-($PO_3$)-Glu-Thr-Asp-Tyr-($PO_3$)-Tyr-($PO_3$)-Arg-Lys from the insulin receptor [63] and/or the Fischer substrate Glu-Asn-Asp-Tyr-($PO_3$)-Ile-Asn-Ala-Ser-Leu [64] should be useful with most protein tyrosine phosphatases], but other phosphatases require individual treatment, including custom-synthesized substrates in some cases. Experiments and common sense suggest that the great majority of phosphatase activities can be measured in one-step reactions using a PhosTRAK-based scheme, although many will require some initial development work. However, there are likely to be a number of phosphatases (acid and alkaline phosphatases are obvious examples) that require separate reaction conditions for the hydrolysis and detection steps, leading to two-step methods, and there may also be phosphatases for which coupled luminescent schemes are simply unworkable or inappropriate.

## Representative PhosTRAK Results with Diverse Phosphatases

Figures 13 to 16 present some of the data that have been obtained to date with PhosTRAK. The *E. coli* λ phosphatase [65] was assayed with unmodified casein as a substrate (Fig. 13). Casein has enough natural phosphorylation to provide an excellent, low-cost substrate for many protein phosphatases. While casein is not the natural substrate of these phosphatases, it is arguably much superior to synthetic, nonproteinaceous substrates as an analog of natural protein substrates. The leukocyte antigen-related phosphatase [66] was assayed with the Fischer phosphopeptide substrate [64]. The sensitivity is excellent, but no saturation was seen in this concentration regime (Fig. 14). The T-Cell Protein Tyrosine Phosphatase [67, 68] (TCPTP, Sigma) was also successfully assayed in the same system (data not shown). Note that leukocyte antigen-related phosphatase (LARP) and TCPTP were assayed in identical buffers, while the λ phosphatase required only the addition of manganese to this buffer. Calcineurin, however, is a very unusual enzyme, with complex kinetics and cofactor requirements [69]. It was chosen for study primarily because it was thought to represent a special challenge, and, indeed, development of a single-step assay of calcineurin required a modest amount of additional work, and the 3-min goal in a very low-cost assay was not achieved, evidently because this was insufficient time for the autoactivation

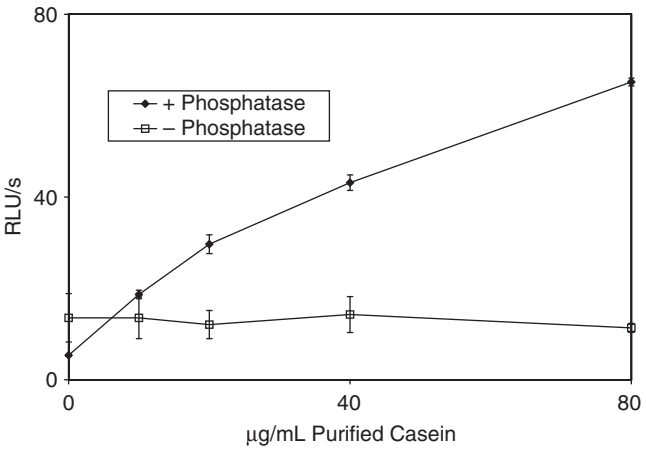


**Figure 13** *E. coli* λ-phosphatase activity measured by PhosTRAK. λ-Phosphatase was assayed with purified but unmodified α-casein (Sigma) as substrate. Casein was dissolved in $H_2O$ at 400 μg/mL and added in 1/10 of the final assay volume to yield the indicated concentrations. The PhosTRAK cocktail (Section 16.6) with or without λ phosphatase was made 2 mM in $MnCl_2$. Substrate was added in 1/10 volume and luminance was read. Data from 2 to 6 min were taken for analysis; linear fits are reported.

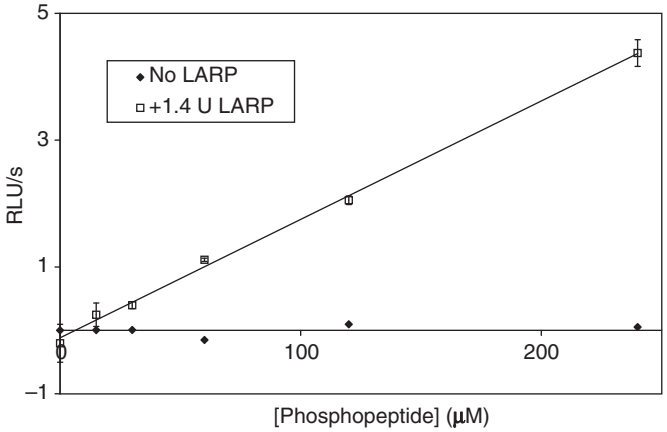


**Figure 14** Substrate dependence of LAR phosphatase/Fischer substrate reaction measured by PhosTRAK. Measurements were taken with or without 1.4 units LARP (1 unit hydrolyzes 1.0 nmol of *p*-nitrophenylphosphate per minute at 30°C, pH 7.0). The substrate was dissolved at 3 mM in DMSO, and the DMSO concentration was equalized across all reactions at 4 percent; 45 μL of PhosTRAK cocktail with or without 31.1 units/mL LARP were aliquoted into each assay well, the substrate was added in 5 μL, and the luminance was read for 3 min. The background rate without enzyme and substrate was subtracted from the data.

process to act. However, a strong signal was evident at 12 min (Fig. 15), using an amount of calcineurin corresponding to approximately 6 percent of the quantity used in the CalBiochem assay based on malachite green. Assays performed in the presence of the calcineurin autoinhibitory peptide [70] required preincubation for the autoactivation and inhibition processes, but the enzyme activity was still measured in a 3-min endpoint reaction (Fig. 16).

## Limitations Due to Complexity and Thermodynamic Considerations

Although it is possible in principle to couple virtually any reaction to ATP or NADH production by using a sufficiently large and complex system, the bounds of pragmatism enter the assay design process in several ways. If numerous enzymes are present, the difficulties of interpretation are multiplied. Some enzymes may have conflicting requirements for small molecules, pH, cofactors, ionic strength, or other environmental factors. Thermodynamic principles impose other significant restrictions on the possibilities. For example, numerous reactions and reaction series involving ATP produce free phosphate or pyrophosphate, and these reactions can be reversed to yield a direct assay for free phosphate; however, the G3PDH-PGK system is one of a limited number of simple reaction series in which ATP is produced from free phosphate

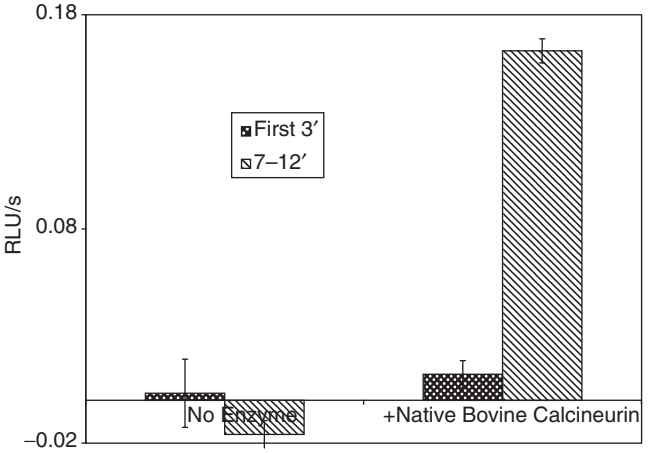


**Figure 15** PhosTRAK measurement of inhibition of calcineurin by autoinhibitory peptide. Recombinant human calcineurin Aα and B coexpressed subunits (protein phosphatase 2B) were assayed in a buffer containing calmodulin, $MnCl_2$, and other components as recommended by the supplier (CalBiochem, San Diego, CA) with 150 μM RII peptide (DLDVPIPGRFDRRV($PO_3$)SVAAE) in the presence of 0, 10, or 30 μM autoinhibitory peptide (AIP) in a 25-μL volume in triplicate. The enzyme with or without AIP was incubated without substrate for 15 min at room temperature. RII peptide was then added and the reaction was transferred to 30°C for 67 min; 5 μL of the reaction, containing 2.2 units of calcineurin (the unit is defined as the amount of enzyme that releases 1 pmol of $P_i$ per minute from the RII peptide at 30°C, pH 7.4), was then transferred to a microtiter well with 45 μL of the standard PhosTRAK cocktail, except that the amounts of PGK and G3PDH enzymes were 50 and 66 percent greater, respectively. Luminance was read after 3.02 min. No enzyme background was subtracted.

*exothermically*, and this appears to be a key element of the ease and flexibility of PhosTRAK. Alternative systems involving substrate-level phosphorylation are also possible, and some pathways suggest the possibility of other coupled luminescent enzyme assays with applications in bacteriology and antibiotic screening [71].

## PhosTRAK Summary

There remains considerable development and validation work to do for PhosTRAK. The availability of ultrapure reagents would likely improve the dynamic range by an order of magnitude at each end. Extensive inhibitor studies would clarify the importance of spurious inhibitory modes due to the presence of multiple enzymes and confirm or refute the hypothesis that targeted libraries are likely to perform best with this assay. The breadth of applications of coupled luminescent assays such as PhosTRAK will gradually

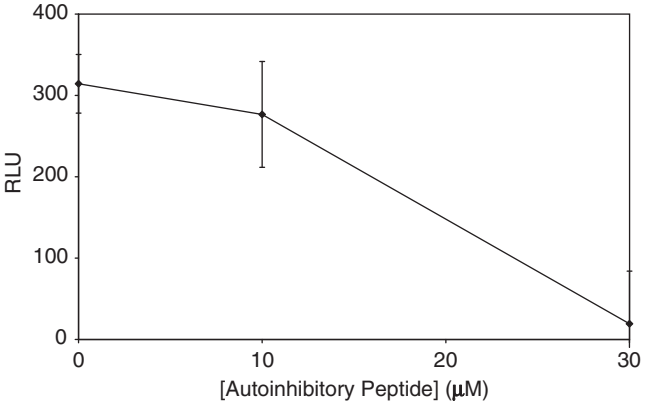


**Figure 16** Autoactivation of native bovine calcineurin observed by one-step PhosTRAK assay. Forty units of enzyme were incubated in duplicate with a combination of the calcineurin reaction and PhosTRAK cocktail modified as described in the legend to Figure 15. Luminance was measured every 20 s for 12 min. Linear fits of the first 3 min and the interval from 7 to 12 min are shown.

become apparent as additional formulations permit one-step reactions with a broader range of enzymes. In still other coupled systems, the phosphatase reaction may merely be an intermediate coupling step downstream from a different enzyme of interest.

## 4 COUPLED LUMINESCENT ASSAYS IN PRACTICE: SPECIFICATIONS AND LIMITATIONS

### Assay Specifications

Specification parameters of the coupled luminescent assays described herein include speed, sensitivity, operational simplicity, $Z$ values, and cost.

***Speed and Linear Response*** With the exception of the 12-min calcineurin assay described above, the one-step assays described herein can be accomplished in 3 min from initiation. If a single reading is to be taken, a G3PDH standard curve should be included in the experiment to assure that the readings are within the linear range of the response. If biochemical saturation is observed (i.e., a signal outside the linear range of a G3PDH standard curve), the data from 0 to 2 min may be usable, if they have been captured; alternatively, the samples can be diluted or the reaction cocktail adjusted appropriately (usually by adding ADP). Given that signal generation begins immediately in most cases, if a very rapid luminescence reader is available, it may be possible to develop an assay with a plate turnover under 30 s, although a sophisticated injection scheme would be required.

***Sensitivity***

*DeathTRAK Sensitivity/Limit of Detection* The DeathTRAK assay as described, using vendor-supplied materials without further purification, is capable of distinguishing the G3PDH released by 0.15 nucleated mammalian cell equivalent from the background signal. An acceptable $Z$ value is achieved at about one nucleated mammalian cell.

*PhosTRAK Sensitivity/Limit of Detection* PhosTRAK as described, using vendor-supplied materials without further purification, can distinguish 3 pmol of $P_i$ from the background signal.

***Operational Characteristics*** Both DeathTRAK and PhosTRAK are fundamentally one-step assays, in which the reaction under study occurs simultaneously with the measurement, and only a single reagent addition is required. However, this does not imply that no other steps are needed under any circumstances. For example, for cytotoxicity measurements by DeathTRAK, the usual practice is to initiate the (potentially) cytotoxic process, wait for an effect, and then add the reaction cocktail and read. In the case of PhosTRAK, true one-step protocols appear to be possible for many or most phosphatases, but there are exceptions, as noted. A related issue is *how* to add the cocktail. If all the components of the coupled reaction are stored together while the test reactions wait for equipment, or for other plates to be read, undesirable side reactions can occur (as well as the backward reaction; see below). Usually it is possible to inject the critical enzymes separately from the other components or to include an essential factor in the test well instead of the reagent reservoir; if not, the reaction cocktail can be mixed just before the run. Another matter is whether to keep components cold. Obviously, this depends on their stability and how long they will have to be stored before use. Again, with a little thought given to protocol design, it may be possible to keep critical reagents cold in a small volume that will have little effect on the temperature in the well after injection. In any case, temperature equilibration is fairly rapid inside a typical luminometer, and as a result, temperature artifacts with significant consequences are relatively rare.

***Z Values*** The $Z$ (or $Z'$) value or factor [72] is often reported as a measure of assay "quality." Calculation of the $Z$ factor [Eq. (1)] takes into account both the signal strength and the scatter of the positive and background signals, while properly ignoring the absolute background level, essentially conveying the ability of the assay as performed to distinguish signal variations from noise:

$$Z = 1 - \frac{3(\mathrm{STDEV}_{\mathrm{positive}} + \mathrm{STDEV}_{\mathrm{negative}})}{\mathrm{AVG(positive\ control)} - \mathrm{AVG(negative\ control)}} \tag{1}$$

where STDEV is the standard deviation, AVG is the arithmetic mean, and "positive control" refers to the maximum signal obtained without inhibition.

Of interest is the fact that the absolute level of the background signal does not enter into the calculation of the *Z* factor; only the variabilities and the difference between positive and negative signals matter. The implication of this for coupled luminescent systems is that limited contamination of assay components by reagents under test (such as ATP in the DeathTRAK system, and either ATP or $P_i$ in the PhosTRAK system) has little or no effect on the *Z* factor since the contamination introduces a significant but *constant* background signal. In a sense this corresponds to reality in that what is important is the statistical distinction between signal and background, rather than the absolute level of either. However, this can be carried too far: Some sources of noise scale with the background signal strength, and a very high background level can cause biochemical and/or detector saturation or even unwanted side reactions.

The *Z* factor is useful in that it informs the researcher as to whether a desired degree of change in a phenomenon will be detectable at a high confidence level. However, the *Z* factor has inherent limitations and is sometimes misapplied. The *Z* factor describes the performance of an assay in a given formulation, but, since error modes typically observed in HTS often do not scale with signal strength, it is often possible to "buy" a high *Z* factor by simply using more of the target agent; this, however, requires the use of more of the test compounds as well to make use of the increased amount of target. Moreover, the value of a high *Z* factor is frequently overestimated: Strong changes in signal are usually the most important, and a *Z* factor of 0.5 or greater is generally quite adequate for detection of these interactions. Rather than the magnitude of the *Z* factor, the *cost* of achieving an acceptable *Z* factor is usually a better basis for decision making. Conversely, when reagents are inexpensive, higher *Z* factors may permit identification of less pronounced effects that may aid in deconvoluting structure–function relationships or provide lead compounds subject to improvements. The assays described herein have been optimized for sensitivity, rapidity, and low cost; however, the very nature of one-step assays permits fewer opportunities for the introduction of measurement errors, and the *Z* values of these and other operationally simple assays often reflect this simplicity. Representative *Z* values obtained were as follows: Figure 5, 0.77; Figures 7*a* and 7*b*, 0.87; Figure 8, 0.97; Figure 9, 0.82. Figure 10 exhibits *Z* values of <0.5 because the leakiness of the Gram-positive organisms leads to scatter in the control; however, in Figure 11, the biomass data, an average *Z* value of $0.77 \pm 0.16$, is obtained over all the runs at 0.5 to 50 μg/mL of the antibiotics. The *Z* value of the λ phosphatase assay (Fig. 13) is 0.90. In the LARP assay (Fig. 14), *Z* values of $0.60 \pm 0.21$ were obtained at 15 to 120 μM substrate. Despite the low signal strength, the *Z* value of the one-step calcineurin assay depicted in Figure 15 is 0.80, reflecting the excellent separation between signal and background. Typical *Z* values of DeathTRAK and PhosTRAK assays using HTS conditions are 0.6 to 0.9.

***Cost*** Even at full retail prices, the cost of these assays per 45-μL aliquot is less than $0.05. Most of this is due to the luciferase, which can be cloned and expressed, or purified, or obtained in bulk.

## Backward Reactions

In reaction series in which time-dependent accumulation of ATP is an essential component of the signal, care must be taken to avoid artifactual fluctuations in the ATP concentration due to backward reactions of luciferase. These reactions can occur if all three products of the canonical luciferase reaction are present, that is, AMP, pyrophosphate, and light. The kind of problem this can create is depicted in Figure 17. Here the components of the DeathTRAK reaction series have been mixed and intentionally incubated in the presence of room light for several minutes before initiation of the test reaction. Because there is a small amount of G3PDH contaminating the PGK preparation, the test reaction can proceed slowly even in the absence of added G3PDH, yielding products of the luciferase reaction. A pseudo-steady-state develops in which the concentration of ATP is maintained at a slightly elevated level by the action of the backwards luciferase reaction on these products, together with available photons from ambient light. When the reaction is initiated by adding the sample, the microtiter plate is simultaneously transferred to the dark chamber of the luminometer. This alters the pseudo-steady-state concentration of ATP since high levels of light are no longer available and the backwards reaction cannot occur. The concentration of ATP slowly drops until a new "dark pseudo-steady-state" is achieved, at which point the response to the sample can be measured accurately. As Figure 17 illustrates, the delay can be as long as 10 min. One half of the plate was covered with an opaque plate seal during the preincubation; this reduced but did not eliminate the lag phase. The solution is simple in theory and usually, in practice: Once the full reagent cocktail has been mixed, it should be protected from light.

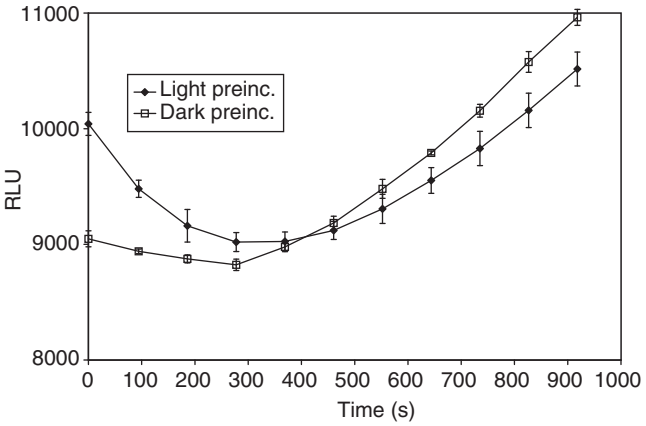


**Figure 17** Demonstration of light-dependent lag phase of DeathTRAK. Standard DeathTRAK assays using detergent-lysed Raji cell (in triplicate) were preincubated under a fluorescent lamp for 10 min and immediately transferred to luminometer chamber for reading. Covering half of the plate with an opaque plate seal reduced but did not eliminate the lag phase.

Another approach is to supply the elements of the cocktail in two separate injections, or provide a portion of it with the sample, so that the test reaction cannot occur before initiation. The time-linear correlation coefficients reported herein, many greater than 0.999, were obtained either by protecting the cocktails from light or by adding the PGK component immediately before assay initiation.

## 5 SUMMARY AND CONCLUSIONS

Coupled luminescent technology has wide potential. Two novel assay methods based on this technology are described in detail herein, but the range of potential applications is great and has been little explored to date. As these examples make clear, the practitioner of coupled luminescent assays can typically expect very rapid and sensitive readouts with minimal sample and reagent usage. However, there are idiosyncrasies and limitations inherent to the method, many of which are associated with the presence of multiple catalytic activities; only the use of proper controls and secondary screens can assure correct interpretation of the results. It is hoped that this review will aid the drug discovery researcher in deciding whether and how coupled luminescent techniques can advance the goals of his or her program.

## 6 FORMULATIONS

### Reaction Master Mixes (Cocktails) and Dilution Buffers

***DeathTRAK Standard Cocktail*** The standard cocktail is used for measurement of both eukaryotes and prokaryotes. The cocktail may be made as specified, using aliquots of the 4XGP cocktail, which may be stored at −70°C for up to 2 years. Alternatively, the nonlabile cocktail (below) may be stored for up to 1 year.

| | |
|---|---|
| IMDM mammalian growth medium | 31.2% |
| Phosphate-buffered saline (PBS) | 16.9% |
| ATP Assay Diluent (Sigma) | 38.1% |
| 4XGP cocktail (see below) | 8.5% |
| ATP Assay Cocktail (Sigma) | 4.2% |
| Phosphoglycerokinase* | 0.11 units/mL (diluted with PGK diluent, below) |
| Dithiothreitol | 0.0004 M final |
| ADP | 3 μM final* |

*In working with some systems, the ADP and phosphoglycerokinase concentrations may be adjusted upward to improve sensitivity, especially if the static

background level and rate of increase (respectively) are not important concerns. If extreme sensitivity and a very wide dynamic range are needed, it may be fruitful to make or obtain an ADP preparation that is substantially free of ATP [22–24] and eliminate or inactivate the G3PDH contamination in the PGK preparation. Unit definition for PGK: "One unit will convert 1.0 μmol of 1,3-diphosphoglycerate to 3-phosphoglycerate per min at pH 6.9 at 25°C."

***4XGP Cocktail***

80% 5X PGK diluent (below)
0.004 M nicotinamide adenine dinucleotide (oxidized form)
0.012 M glyceraldehyde-3-phosphate
0.004 M dithiothreitol
QS to volume with $H_2O$

***5X PGK Diluent***

0.5 M triethanolamine
0.25 M $NaH_2PO_4$
0.005 M EDTA (ethylenediaminetetraacetic acid)
0.5 mg/mL bovine serum albumin (fraction V)

***Nonlabile Cocktail*** In the alternative composition, which is easier to reconstitute but has a shorter shelf life, all the reagents except the ATP assay mix, phosphoglycerokinase, and ADP are mixed together and stored at −70°C in suitable aliquots. This method has the advantage that the ADP solution can be made fresh to adjust to the ATP contamination level of a particular lot, or to take advantage of purified ADP if available.

***1X G3PDH Diluent***

To 200 parts 5X PGK diluent add:
1 part 1 M dithiothreitol
799 parts $H_2O$

**PhosTRAK Formulations**

***PhosTRAK Standard Cocktail***

0.05 M Tris-HCl (final) (pH 7.4)
38% ATP assay diluent
4.2% ATP assay mix
PGK = $4.5 \times 10^{-4}$ units/mL
G3PDH = $4.9 \times 10^{-4}$ units/mL
8.1% Phosphate-free 4XGP cocktail (below)

QS to volume with $H_2O$
(Add shortly before assay in minimal volume)
6 μM ADP
3 μM G3P

Unit definition for G3PDH: "One unit will reduce 1.0 μmol of 3-phosphoglycerate to D-glyceraldehyde-3-phosphate per minute in a coupled system with 3-phosphoglyceric phosphokinase at pH 7.6 at 25°C."

***Phosphate-Free GP Cocktail***

83.5% 5× phosphate-free PGK diluent
0.004 M dithiothreitol
0.0042 M $NAD^+$
QS to volume with $H_2O$

***Phosphate-Free 5X PGK Diluent***

0.5 M triethanolamine
0.25 M Tris pH 7.4
0.005 M EDTA
0.5 mg/mL bovine serum albumin (fraction V)
Titrate to pH 7.3 with HCl; QS to volume with $H_2O$

***Phosphate-Free Nonlabile Cocktail (NLC)*** An NLC can also be made up for the PhosTRAK reaction, as follows. However, if long-term storage is intended, it should be kept in mind that G3P is more stable at mildly acidic pH. Hydrolyzed G3P will, of course, contribute to the background rate of the PhosTRAK signal. G3P can be omitted from the NLC and added just prior to the assay, or the pH and buffer strength of the NLC can be adjusted to suit the storage requirements.

0.05 M Tris-HCl (final) (pH 7.4)
38% ATP assay diluent
8.1% Phosphate-free 4XGP cocktail (below)
6 μM ADP
3 μM G3P
0.001 M dithiothreitol
QS to volume with $H_2O$
PGK, G3PDH, ATP assay mix, and (optionally) G3P are added shortly before the run.

### Lytic Formulations

Although these formulations were found to be effective with the cell types and strains specified, it is clear that some strains are resistant to certain detergents and/or antibiotics (such as polymyxin B). In the case of mammalian cells, no cell type tested has proven to be resistant to Nonidet P-40, but the possibility that such a cell type exists still remains. Thus the lytic formulation under consideration must be carefully tested with the cells to be used for screening.

***Mammalian Cells and Gram-Positive Prokaryotes*** For mammalian cells and Gram-positive prokaryotes, add the lytic agent in 10 percent of the culture volume to achieve the following final concentrations (before addition of DeathTRAK cocktail):

0.2% Nonidet P-40 or Igepal
0.001 M dithiothreitol

In the modes of use tested and envisioned, the assay cocktail is added immediately after the lytic agent; thus no pH buffering is required in the lytic agent. However, if a contemplated application involves prolonged incubation in the presence of the lytic agent, it may be desirable to buffer the lytic agent to prevent inactivation of G3PDH. The use of protease inhibitors may also be indicated in such cases.

***Gram-Negative Prokaryotes*** For Gram-negatives, the lytic agent is made up in PBS and added in 10 percent of the culture volume to yield final concentrations (prior to addition of the DeathTRAK cocktail) as follows:

3000 units/mL polymyxin B
25 mg/mL chicken egg-white lysozyme

## ABBREVIATIONS

| | |
|---|---|
| 3PG | 3-phosphoglycerate |
| 4XGP | 4X glyceraldehyde-3-phosphate dehydrogenase/ phosphoglycerokinase cocktail |
| ADP | adenosine diphosphate |
| AMP | adenosine monophosphate |
| ATP | adenosine triphosphate |
| BrdU | bromodeoxyuridine |
| cAMP | cyclic adenosine monophosphate |
| CTL | cytotoxic T lymphocyte |
| DTT | dithiothreitol |

| | |
|---|---|
| FP | fluorescence polarization |
| FRET | fluorescence resonance energy transfer |
| G3P | glyceraldehyde-3-phosphate |
| G3PDH | glyceraldehyde-3-phosphate dehydrogenase |
| GPCR | G-protein coupled receptor |
| GPL | glyceraldehyde-3-phosphate dehydrogenase/ phosphoglycerokinase/luciferase |
| HTS | high-throughput screening |
| IND | investigational new drug |
| $K_m$ | Michaelis constant |
| LAK | lymphokine-activated killer |
| LARP | leukocyte antigen-related phosphatase |
| LDH | lactate dehydrogenase |
| MTT | 3-[4,5-dimethylthiazol-2-yl]-2,5-diphenyltetrazolium bromide |
| $NAD^+$ | nicotinamide adenine dinucleotide (oxidized form) |
| NADH | nicotinamide adenine dinucleotide (reduced form) |
| NADPH | nicotinamide adenine dinucleotide phosphate (reduced form) |
| NK | natural killer |
| NLC | nonlabile cocktail |
| PBS | phosphate-buffered saline |
| PDE | phosphodiesterase |
| PGK | phosphoglycerokinase |
| $P_i$ | inorganic phosphate = $PO_4^{3-}$ |
| RLU | relative luminance units |
| SPA | scintillation proximity assay |
| TCPTP | T-cell protein tyrosine phosphatase |
| WST-1 | 2-(4-iodophenyl)-3-(4-nitrophenyl)-5-(2,4-disulfophenyl)-2H-tetrazolium |
| XTT | 2,3-Bis(2-methoxy-4-nitro-5-sulfophenyl)-2H-tetrazolium-5-carboxanilide |

## REFERENCES


1. Berthold, F. (1990). In K. Van Dyke and R. Van Dyke (Eds.), *Luminescence Immunoassay and Molecular Applications*. CRC Press, Boca Raton, FL, pp. 11–25.
2. Khalil, O. S. (2002). In J. E. Butler (Ed.), *Immunochemistry of Solid Phase Immunoassay*. CRC Press, Boca Raton, FL, pp. 67–83.
3. Jeffery, J. A., Sharom, J. R., Fazekas, M., Rudd, P., Welchner, E., Thauvette, L., White, P. W. (2002). An ATPase assay using scintillation proximity beads for high-throughput screening or kinetic analysis. *Anal. Biochem.*, *304*, 55–62.
4. Skorey, K. I., Kennedy, B. P., Friesen, R. W., Ramachandran, C. (2001). Development of a robust scintillation proximity assay for protein tyrosine phosphatase 1B using the catalytically inactive (C215S) mutant. *Anal. Biochem.*, *291*, 269–278.

5. McDonald, O. B., Chen, W. J., Ellis, B., Hoffman, C., Overton, L., Rink, M., Smith, A., Marshall, C. J., Wood, E. R. (1999). A scintillation proximity assay for the Raf/MEK/ERK kinase cascade: High-throughput screening and identification of selective enzyme inhibitors. *Anal. Biochem.*, *268*, 318–329.
6. Sills, M. A., Weiss, D., Pham, Q., Schweitzer, R., Wu, X., Wu, J. J. (2002). Comparison of assay technologies for a tyrosine kinase assay generates different results in high throughput screening. *J. Biomol. Screen.*, *7*, 191–214.
7. Schindler, P. W., Konig, W., Chatterjee, S., Ganguli, B. N. (1986). Improved screening for beta-lactam antibiotics. A sensitive, high-throughput assay using DD-carboxypeptidase and a novel chromophore-labeled substrate. *J. Antibiotics (Tokyo)*, *39*, 53–57.
8. Voss, K., Galensa, R. (2000). Determination of L- and D-amino acids in foodstuffs by coupling of high-performance liquid chromatography with enzyme reactors. *Amino Acids*, *18*, 339–352.
9. Gutheil, W. G., Stefanova, M. E., Nicholas, R. A. (2000). Fluorescent coupled enzyme assays for D-alanine: Application to penicillin-binding protein and vancomycin activity assays. *Anal. Biochem.*, *287*, 196–202.
10. Sudhaharan, T., Reddy, A. R. (1999). A bifunctional luminogenic substrate for two luminescent enzymes: Firefly luciferase and horseradish peroxidase. *Anal. Biochem.*, *271*, 159–167.
11. Ross, W. N. (1989). Changes in intracellular calcium during neuron activity. *Annu. Revi. Physiol.*, *51*, 491–506.
12. Li, F., Zhang, C., Guo, X., Feng, W. (2003). Chemiluminescence detection in HPLC and CE for pharmaceutical and biomedical analysis. *Biomed. Chromatogr.*, *17*, 96–105.
13. Dupriez, V. J., Maes, K., Le Poul, E., Burgeon, E., Detheux, M. (2002). Aequorin-based functional assays for G-protein-coupled receptors, ion channels, and tyrosine kinase receptors. *Receptors and Channels*, *8*, 319–330.
14. Hastings, J. W., Gibson, Q. H. (1963). Intermediates in the bioluminescent oxidation of reduced flavin mononucleotide. *J. Biol. Chem.*, *238*, 2537–2554.
15. Anderstam, B., Gutierrez, A., Lundin, A., Alvestrand, A. (1998). A luminometric assay for determination of ethanol in microdialysates. *Scand. J. Clin. Lab. Invest.*, *58*, 89–96.
16. Huang, W., Feltus, A., Witkowski, A., Daunert, S. (1996). Homogeneous bioluminescence competitive binding assay for folate based on a coupled glucose-6-phosphate dehydrogenase—Bacterial luciferase enzyme system. *Anal. Chem.*, *68*, 1646–1650.
17. Wieland, E., Wilder-Smith, E., Kather, H. (1986). Automatic bioluminescent glucose determination using commercially available reagent kits coupled to the bacterial NAD(P)H-linked luciferase system. *J. Clin. Chem. Clin. Biochem.*, *24*, 399–403.
18. Robrish, S. A., Curtis, M. A., Sharer, S. A., Bowen, W. H. (1984). The analysis of picomole amounts of L(+)- and D(−)-lactic acid in samples of dental plaque using bacterial luciferase. *Anal. Biochem.*, *136*, 503–508.
19. Schoelmerich, J., Hinkley, J. E., Macdonald, I. A., Hofmann, A. F., DeLuca, M. (1983). A bioluminescent assay for 12-alpha-hydroxy bile acids using immobilized enzymes. *Anal. Biochem.*, *133*, 244–250.

20. Nygaard, S. F., Rorth, M. (1969). An enzymatic assay of 2,3-diphosphoglycerate in blood. *Scand. J. Clin. Lab. Invest.*, *24*, 399–403.

21. Loos, J. A., Prins, H. K. (1971). A mechanized system for the determination of ATP + ADP, 2,3-diphosphoglycerate, glucose 1,6-diphosphate and lactate in small amounts of blood cells. *Circulation*, *43*, 1141–1146.

22. de Korte, D., Haverkort, W. A., van Gennip, A. H., Roos, D. (1985). Nucleotide profiles of normal human blood cells determined by high-performance liquid chromatography. *Anal. Biochem.*, *147*, 197–209.

23. Crescentini, G., Stocchi, V. (1984). Fast reversed-phase high-performance liquid chromatographic determination of nucleotides in red blood cells. *J. Chromatogr.*, *290*, 393–399.

24. Rochette-Egly, C., Kempf, J., Egly, J. M. (1979). A new chromatographic method using immobilized acriflavin for measuring cyclic AMP in cells prelabeled with radioactive adenine. *J. Cyclic Nucleotide Res.*, *5*, 397–406.

25. Lundin, A. (2000). Use of firefly luciferase in ATP-related assays of biomass, enzymes, and metabolites. *Methods Enzymol.*, *305*, 346–401.

26. Lundin, A., Gerhardt, W., Lindberg, K., Lovgren, T., Nordlander, R., Nyquist, O., Styrelius, I. (1979). Determination of creatine kinase B subunit activity by continuous monitoring of ATP—A new bioluminescent technique applied to clinical chemistry. *Clin. Biochem.*, *12*, 214–215.

27. Welch, A. R. (2000). High throughput assay method for enzymes which metabolically hydrolyze nucleoside triphosphates and an assay system therefore. Patent No. WO 00/18950, assigned to Diagnon Corporation.

28. Learish, R. D., Shultz, J., Ho, S., Bulleit, R. F. (2002). Small-scale telomere repeat sequence content assay using pyrophosphorolysis coupled with ATP detection. *Biotechniques*, *33*, 1349–1353.

29. Rhodes, R. B., Lewis, K., Shultz, J., Huber, S., Voelkerding, K. V., Leonard, D. G., Tsongalis, G. J., Kephart, D. D. (2001). Analysis of the factor V Leiden mutation using the READIT™ assay. *Mol. Diagnostics*, *6*, 55–61.

30. Tsongalis, G. J., Rainey, B. J., Hodges, K. A. (2001). READIT: A novel technology used in the interrogation of nucleic acid sequences for single-nucleotide polymorphisms. *Exper. Mol. Pathol.*, *71*, 222–225.

31. Rocco, P. R., Momesso, D. P., Figueira, R. C., Ferreira, H. C., Cadete, R. A., Legora-Machado, A., Koatz, V. L., Lima, L. M., Barreiro, E. J., Zin, W. A. (2003). Therapeutic potential of a new phosphodiesterase inhibitor in acute lung injury. *Eur. Respir. J.*, *22*, 20–27.

32. Reffelmann, T., Kloner, R. A. (2003). Therapeutic potential of phosphodiesterase 5 inhibition for cardiovascular disease. *Circulation*, *108*, 239–244.

33. Boolell, M., Allen, M. J., Ballard, S. A., Gepi-Attee, S., Muirhead, G. J., Naylor, A. M., Osterloh, I. H., Gingell, C. (1996). Sildenafil: An orally active type 5 cyclic GMP-specific phosphodiesterase inhibitor for the treatment of penile erectile dysfunction. *Int. J. Impotence Res.*, *8*, 47–52.

34. Schafer, H., Schafer, A., Kiderlen, A. F., Masihi, K. N., Burger, R. (1997). A highly sensitive cytotoxicity assay based on the release of reporter enzymes, from stably transfected cell lines. *J. Immunol. Methods*, *204*, 89–98.

35. Kasatori, N., Urayama, T., Mori, T., Ishikawa, F. (1994). Cytotoxicity test based on luminescent assay of alkaline phosphatase released from target cells [in Japanese]. *Rinsho Byori*, *42*, 1050–1054.
36. Rich, G. T., Dawson, A. P., Pryor, J. S. (1984). Glyceraldehyde-3-phosphate dehydrogenase release from erythrocytes during haemolysis. No evidence for substantial binding of the enzyme to the membrane in the intact cell. *Biochem. J.*, *221*, 197–202.
37. Corey, M. J., Kinders, R. J., Brown, L. G., Vessella, R. L. (1997) A very sensitive coupled luminescent assay for cytotoxicity and complement-mediated lysis. *J. Immunol. Methods*, *207*, 43–51.
38. Corey, M. J., Kinders, R. J. (2002). Methods and compositions for coupled luminescent assays. U. S. patent application No. 10/071,350, assigned to M. J. Corey.
39. Corey, M. J., Kinders, R. J., Poduje, C. M., Bruce, C. L., Rowley, H., Brown, L. G., Hass, G. M., Vessella, R. L. (2000). Mechanistic studies of the effects of anti-factor H antibodies on complement-mediated lysis. *J. Biol. Chem.*, *275*, 12917–12925.
40. Hardwick, A., McMillen, D., Martinez, J., Austin, A., Posey, A., Ave-Teel, C., Maples, P., Schneider, S. (2003). Clinical-scale production of antigen-specific T cells directed against hepatitis B virus. *Bioproc. J.*, *2*, 27–31.
41. Slater, K. (2001). Cytotoxicity tests for high-throughput drug discovery. *Curr. Opin. Biotechnol.*, *12*, 70–74.
42. Crissman, H. A., Steinkamp, J. A. (1987). A new method for rapid and sensitive detection of bromodeoxyuridine in DNA-replicating cells. *Exper. Cell Res.*, *173*, 256–261.
43. Maybaum, J., Kott, M. G., Johnson, N. J., Ensminger, W. D., Stetson, P. L. (1987). Analysis of bromodeoxyuridine incorporation into DNA: Comparison of gas chromatographic/mass spectrometric, CsCl gradient sedimentation, and specific radioactivity methods. *Anal. Biochem.*, *161*, 164–171.
44. Niu, Q., Zhao, C., Jing, Z. (2001). An evaluation of the colorimetric assays based on enzymatic reactions used in the measurement of human natural cytotoxicity. *J. Immunol. Methods*, *251*, 11–19.
45. Bruggisser, R., von Daeniken, K., Jundt, G., Schaffner, W., Tullberg-Reinert, H. (2002). Interference of plant extracts, phytoestrogens and antioxidants with the MTT tetrazolium assay. *Planta Med.*, *68*, 445–448.
46. Natarajan, M., Mohan, S., Martinez, B. R., Meltz, M. L., Herman, T. S. (2000). Antioxidant compounds interfere with the 3-[4,5-dimethylthiazol-2-yl]-2,5-diphenyltetrazolium bromide cytotoxicity assay. *Cancer Detect. Prevent.*, *24*, 405–414.
47. Wong, J. K., Kennedy, P. R., Belcher, S. M. (2001). Simplified serum- and steroid-free culture conditions for high-throughput viability analysis of primary cultures of cerebellar granule neurons. *J. Neurosci. Methods*, *110*, 45–55.
48. Doman, T. N., McGovern, S. L., Witherbee, B. J., Kasten, T. P., Kurumbail, R., Stallings, W. C., Connolly, D. T., Shoichet, B. K. (2002). Molecular docking and high-throughput screening for novel inhibitors of protein tyrosine phosphatase-1B. *J. Med. Chem.*, *45*, 2213–2221.

49. Parker, G. J., Law, T. L., Lenoch, F. J., Bolger, R. E. (2000). Development of high throughput screening assays using fluorescence polarization: Nuclear receptor-ligand-binding and kinase/phosphatase assays. *J. Biomol. Screen.*, *5*, 77–88.
50. Mendoza, L. G., McQuary, P., Mongan, A., Gangadharan, R., Brignac, S., Eggers, M. (1999). High-throughput microarray-based enzyme-linked immunosorbent assay (ELISA). *Biotechniques*, *27*, 778–780, 782–786, 788.
51. Agazie, Y. M., Hayman, M. J. (2003). Development of an efficient "substrate-trapping" mutant of Src homology phosphotyrosine phosphatase 2 and identification of the epidermal growth factor receptor, Gab1, and three other proteins as target substrates. *J. Biol. Chem.*, *278*, 13952–13958.
52. Baykov, A. A., Evtushenko, O. A., Avaeva, S. M. (1988). A malachite green procedure for orthophosphate determination and its use in alkaline phosphatase-based enzyme immunoassay. *Anal. Biochem.*, *171*, 266–270.
53. Bronstein, I. Y. (2000). Method of detecting a substance using enzymatically-induced decomposition of dioxetanes. U.S. Patent No. RE36,536, assigned to Tropix, Inc.
54. Charbonneau, H., Tonks, N. K. (1992). 1002 Protein phosphatases? *Annu. Rev. Cell Biol.*, *8*, 463–493.
55. Tonks, N. K. (2003). PTP1B: From the sidelines to the front lines! *FEBS Lett.*, *546*, 140–148.
56. Neel, B. G., Gu, H., Pao, L. (2003). The 'Shp'ing news: SH2 domain-containing tyrosine phosphatases in cell signaling. *Trends Biochem. Sci.*, *28*, 284–293.
57. Van Hoof, C., Goris, J. (2003). Phosphatases in apoptosis: To be or not to be, PP2A is in the heart of the question. *Biochim. Biophys. Acta*, *1640*, 97–104.
58. Irie-Sasaki, J., Sasaki, T., Penninger, J. M. (2003). CD45 regulated signaling pathways. *Curr. Top. Med. Chem.*, *3*, 783–796.
59. Atack, J. R. (1996). Inositol monophosphatase, the putative therapeutic target for lithium. *Brain Res. Rev.*, *22*, 183–190.
60. Jiang, H., Harris, M. B., Rothman, P. (2000). IL-4/IL-13 signaling beyond JAK/STAT. *J. Allergy Clin. Immunol.*, *105*, 1063–1070.
61. Sciorra, V. A., Morris, A. J. (2002). Roles for lipid phosphate phosphatases in regulation of cellular signaling. *Biochim. Biophys. Acta*, *1582*, 45–51.
62. Kanoh, H., Sakane, F., Imai, S., Wada, I. (1993). Diacylglycerol kinase and phosphatidic acid phosphatase—Enzymes metabolizing lipid second messengers. *Cell Signaling*, *5*, 495–503.
63. Lee, J. P., Cho, H., Bannwarth, W., Kitas, E. A., Walsh, C. T. (1992). NMR analysis of regioselectivity in dephosphorylation of a triphosphotyrosyl dodecapeptide autophosphorylation site of the insulin receptor by a catalytic fragment of LAR phosphotyrosine phosphatase. *Protein Sci.*, *1*, 1353–1362.
64. Daum, G., Solca, F., Diltz, C. D., Zhao, Z., Cool, D. E., Fischer, E. H. (1993). A general peptide substrate for protein tyrosine phosphatases. *Anal. Biochem.*, *211*, 50–54.
65. Zhuo, S., Clemens, J. C., Hakes, D. J., Barford, D., Dixon, J. E. (1993). Expression, purification, crystallization, and biochemical characterization of a recombinant protein phosphatase. *J. Biol. Chem.*, *268*, 17754–17761.

66. Streuli, M., Krueger, N. X., Tsai, A. Y., Saito, H. (1988). A family of receptor-linked protein tyrosine phosphatases in humans and *Drosophila. Proc. Natl. Acad. Sci. USA*, *86*, 8698–8702.
67. Ibarra-Sanchez, M. J., Simoncic, P. D., Nestel, F. R., Duplay, P., Lapp, W. S., Tremblay, M. L. (2000). The T-cell protein tyrosine phosphatase. *Semin. Immunol.*, *12*, 379–386.
68. Hao, L., Tiganis, T., Tonks, N. K., Charbonneau, H. (1997). The noncatalytic C-terminal segment of the T cell protein tyrosine phosphatase regulates activity via an intramolecular mechanism. *J. Biol. Chem.*, *272*, 29322–29329.
69. Martin, B., Pallen, C. J., Wang, J. H., Graves, D. J. (1985). Use of fluorinated tyrosine phosphates to probe the substrate specificity of the low molecular weight phosphatase activity of calcineurin. *J. Biol. Chem.*, *260*, 14932–14937.
70. Perrino, B. A. (1999). Regulation of calcineurin phosphatase activity by its auto-inhibitory domain. *Arch. Biochem. Biophys.*, *372*, 159–165.
71. Kapatral, V., Bina, X., Chakrabarty, A. M. (2000). Succinyl coenzyme A synthetase of *Pseudomonas aeruginosa* with a broad specificity for nucleoside triphosphate (NTP) synthesis modulates specificity for NTP synthesis by the 12-kilodalton form. *J. Bacteriol.*, *182*, 1333–1339.
72. Zhang, J. H., Chung, T. D., Oldenburg, K. R. (1999). A simple statistical parameter for use in evaluation and validation of high throughput screening assays. *J. Biomol. Screen.*, 4, 67–73.

# DESIGN AND PHARMACEUTICAL APPLICATIONS OF PRODRUGS

Tomi Järvinen
*University of Kuopio*
*Kuopio, Finland*

Jarkko Rautio
*University of Kuopio*
*Kuopio, Finland*

Mar Masson
*University of Iceland*
*Reykjavik, Iceland*

Thorsteinn Loftsson
*University of Iceland*
*Reykjavik, Iceland*

1 INTRODUCTION
2 DEFINITION OF THE CONCEPTS
Prodrug Concept
Double-Prodrug Concept
Soft Drug Concept
Chemical Delivery System Concept
3 REQUIREMENTS FOR CLINICALLY USEFUL PRODRUGS
4 BIOREVERSIBLE PRODRUG STRUCTURES FOR THE MOST COMMON FUNCTIONAL GROUPS
Prodrugs for Carboxyl Groups
Prodrugs for Phosphate and Phosphonate Groups
Prodrugs for Hydroxyl Groups
Prodrugs for Amines and Amides

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

5 PHARMACEUTICAL APPLICATIONS OF THE PRODRUG STRATEGY
Improved Drug Absorption
Improved Aqueous Solubility
Prolonged Duration of Action
Improved Drug Targeting
Improved Formulation and Delivery of Peptides
Reduced Side Effects
Macromolecular Prodrugs
References

## 1 INTRODUCTION

The process of drug discovery has been changed considerably during the last decade. Modern drug discovery technologies such as computer-aided drug design, combinatorial chemistry, and high-throughput pharmacological screens have provided large numbers of highly potent and selective drug candidates. However, optimal physico-chemical and biopharmaceutical requirements for the formulation and delivery of these drugs are not selected by these procedures, and thus the new drug candidates often have poor "druglike" properties and a high risk of failure in subsequent preclinical or clinical development [1]. Thus, both drug formulation and delivery must be carefully considered throughout the drug design process.

There are various approaches that can be applied to overcome poor properties of drug formulation and delivery. Some of these problems can be overcome by dosage form design. Another approach is to develop various drug analogs, but the risk in this approach is the irreversible modification of bioactive center(s) and subsequent loss of therapeutic effect. An important strategy approach to improve poor drug formulation and delivery properties is prodrug technology. Historically, the term *prodrug* was first introduced about 40 years ago to describe compounds that undergo biotransformation prior to exhibiting pharmacological effects [2]. In the 1970s and 1980s, the prodrug approach was primarily applied to improve physico-chemical, biopharmaceutical, and/or pharmacokinetic properties of drugs that already were on the market. Currently, prodrug design and research should be more involved in the lead optimization step during the drug discovery process (Fig. 1).

The prodrug approach is a valuable tool to further optimize potent structures and to solve potential formulation or delivery problems. Currently, about 5 percent of all worldwide approved drugs can be classified as prodrugs, and this number is still increasing [3]. Despite these high numbers, understanding the full potential of the prodrug approach in modern drug discovery and development has only just begun, and many prodrug innovations remain to be discovered.

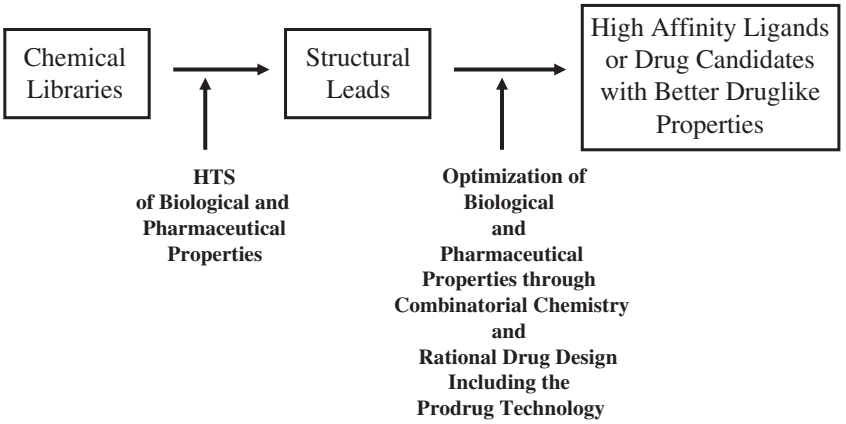


**Figure 1** Modern drug discovery paradigm (modified from lectures by Prof. Valentino Stella and Prof. Ronald Borchardt).

## 2 DEFINITION OF THE CONCEPTS

### Prodrug Concept

Prodrugs are pharmacologically inactive derivatives of drug molecules that require a chemical or enzymatic transformation in order to release the active parent drug in vivo, prior to exerting a pharmacological effect (Fig. 2) [4]. Prodrugs have also been called reversible or bioreversible derivatives, latentiated drugs, or biolabile drug–carrier conjugates, but the term *prodrug* is now the standardized term. In most cases, prodrugs are simple chemical derivatives that are one or two chemical or enzymatic steps away from the parent drug. In certain cases, a prodrug may consist of two pharmacologically active drugs that are coupled together in a single molecule, so that each drug acts as a promoiety for the other drug. Such derivatives are called co-drugs [5].

The major aim in designing prodrugs is to overcome limitations of a parent drug that would otherwise hinder its clinical use. Drug analogs are also used to overcome such limitations (Fig. 2), but facile modifications of the functional group(s) or the bioactive center(s) may result in unacceptable changes in the desired pharmacological profile, compared to the parent drug, which is not a problem in the case of prodrugs.

### Double-Prodrug Concept

Various drawbacks of prodrug derivatives can be overcome by preparing a prodrug from another prodrug, the result being a double prodrug (pro-

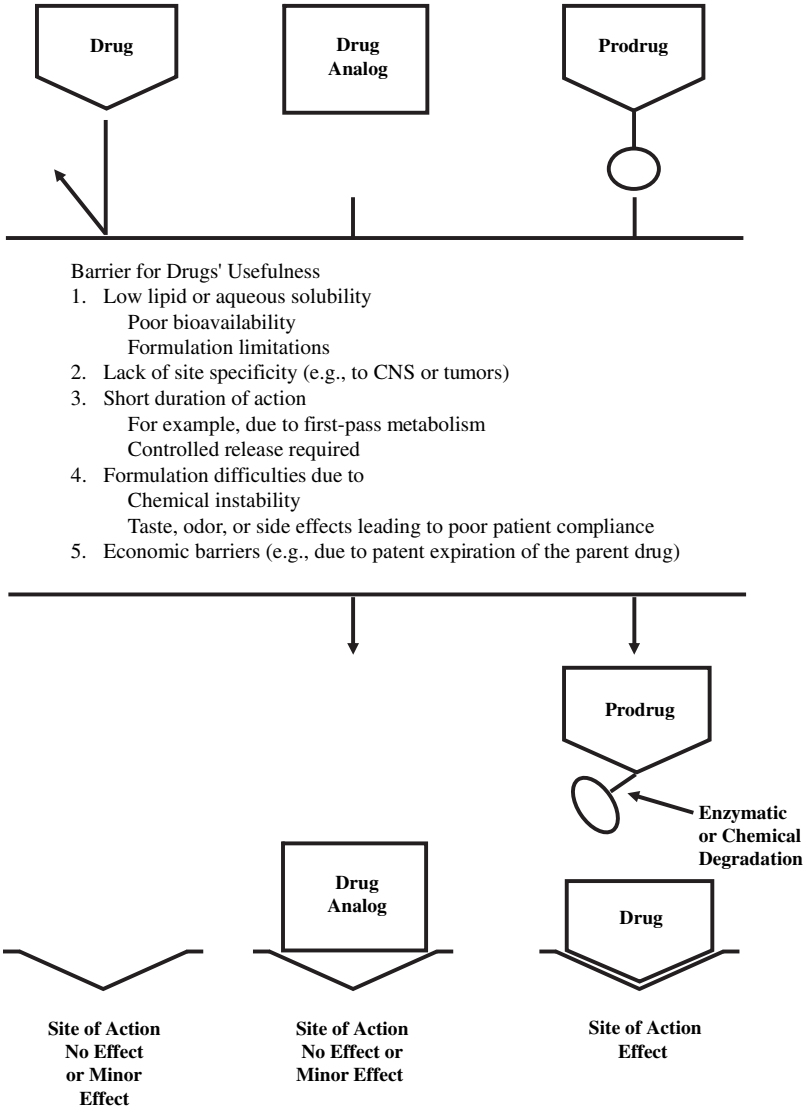


**Figure 2** Basic principle of the prodrug concept.

prodrug). For example, stability problems of a prodrug, especially in solutions, can sometimes be solved by preparing a double prodrug that requires enzymatic hydrolysis before spontaneous chemical release of the parent drug (Fig. 3). The double-prodrug concept has been applied, for example, in development of fosphenytoin [6], prodrugs of 9-[(*R*)-2-(phosphonomethoxy)propyl]adenine (tenofovir) [7], and pilocarpine prodrugs [8]. The term *double prodrug* should not be confused with the term *bi-functional*

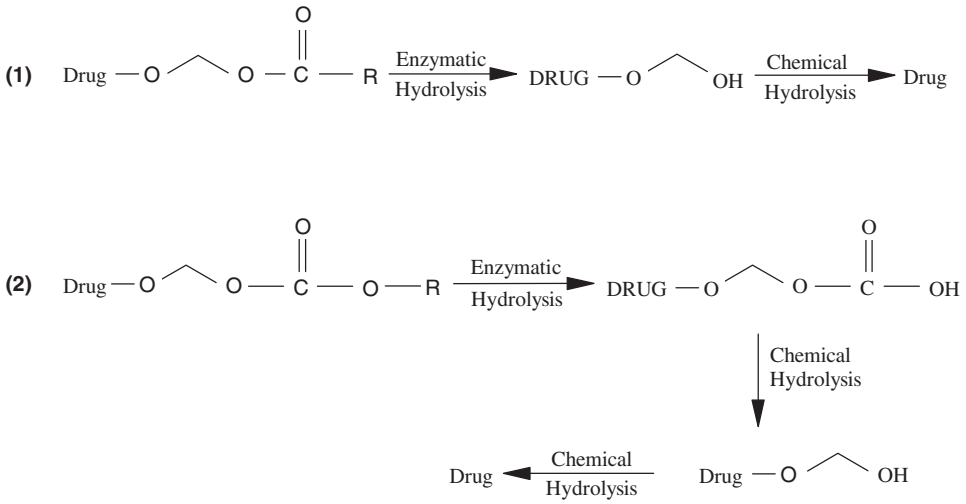


**Figure 3** General mechanism for the bioconversion of (1) acyloxymethyl and (2) alkoxycarbonyloxymethyl double prodrug structures.

*prodrug*, which indicates a prodrug that has been modified at two functional groups on the parent molecule.

## Soft Drug Concept

Soft drugs, unlike prodrugs, are active drugs that are designed to undergo a predicable and controllable deactivation in vivo after achieving their therapeutic role. Soft drugs are active agents that have a metabolically weak structure built into the molecule. After exerting their therapeutic effect at the desired site, soft drugs are rapidly metabolized to form inactive metabolites, which in turn are rapidly metabolized to prevent unwanted pharmacological activity or toxicity. The concept of soft drugs was introduced in the 1970s, and since then various classes of soft drugs have been developed such as soft analgetics, soft antimicrobials, soft β-blockers, and soft corticosteroids [9–11]. Soft drugs are most commonly designed for topical (e.g., ophthalmic, nasal, dermal, and pulmonary) applications. Some of the commercially available soft drugs are, for example, Brevibloc (soft β-blocker esmonol), Ultiva (soft opioid analgetic remifentanyl), and Alrex/Lotemax (soft corticosteroid loteprednol etabonate). If the prodrug approach is applied to a soft drug, the resulting product is known as a *pro-soft drug* [12].

## Chemical Delivery System Concept

A chemical delivery system (CDS) is an inactive drug derivative that undergoes several predictable enzymatic transformations via inactive intermediates

and finally delivers the active drug to the site of action. In an ideal case, the active drug molecule is released from its CDS only in the target organ or tissue, and thus CDS remains inactive elsewhere in the body and is susceptible to rapid elimination [10]. The main focus of CDS drug design has been to target a variety of drugs to the central nervous system, and applications for ophthalmic use already exist. In some applications it can be argued whether or not the appropriate term is prodrug or CDS. Nowadays the term *retrometabolic drug design* is often used to include both soft drug design and chemical delivery systems.

## 3 REQUIREMENTS FOR CLINICALLY USEFUL PRODRUGS

The major goal in designing a prodrug is to overcome various physico-chemical, biopharmaceutical, and/or pharmacokinetic problems that may be associated with a parent drug molecule, which otherwise would be of limited clinical use. The prodrug approach should also be considered when other dosage forms from the active compound are considered. For example, the prodrug approach is extremely useful in developing orally active compounds from an intravenous (i.v.) active compound. The common barriers to drug formulation and delivery most often addressed by the prodrug approach are:

1. Low aqueous solubility

   Prevents the development of aqueous-based formulations (e.g., solutions, drops, infusions).

   Leads to both rate-limited dissolution and variable oral bioavailability.
2. Low lipid solubility

   Results in low membrane permeation and low oral bioavailability.

   Hinders delivery across barriers such as cornea and skin (e.g., topical drug delivery).

   Limits the design of lipid-based formulations.
3. Short duration of action (due to rapid elimination from the body or strong first-pass metabolism)

   Necessitates frequent administration of a drug that often leads to poor patient compliance.
4. Lack of site specificity (e.g., poor brain, tumor, or colon targeting)

   May lead to undesirable systemic effects.
5. Poor taste or odor

   May lead to poor patient compliance.
6. Side effects

   May lead to safety concerns.

   May also lead to poor patient compliance.

7. Economic barriers
   For example; patent and product lifetime may be prolonged by prodrugs with improved formulation and/or drug delivery properties.

A successful prodrug design requires that the reasons for developing prodrugs are clearly defined and understood. The following points must be evaluated in order to confirm that the prodrug concept may be clinically useful:

1. Identification of the suitable functional groups on the parent molecule.
2. Identification of the precise drug formulation or delivery problem.
3. Identification of the physico-chemical/pharmaceutical properties required in order to overcome the specific problem.
4. Selection of a bioreversible prodrug structure that has the proper physico-chemical/pharmaceutical properties to overcome the problem.
5. The selected prodrug-structure can be economically synthesized.

The requirements for a clinically useful prodrug are dependent on dosage form and administration route. In general, prodrugs should fulfill at least most of the following criteria:

1. Synthesis and purification methods for prodrug should be feasible.
2. Adequate chemical stability
   Adequate shelf life in the final formulation.
   Prodrug will not degrade before reaching the site of action.
3. Adequate aqueous solubility
   Allowing dissolution and aqueous-based formulation development.
4. Optimal lipophilicity, to confirm efficient permeation across biological membranes
   $\log P$ value of 2 is considered to be optimal for gastrointestinal (GI) absorption, for example.
4. The prodrug must release the parent drug in vivo at a reasonable rate
   before, during, or after absorption or at the specific site of action.
5. The prodrug should be nontoxic and tasteless, and the promoiety released after bioactivation must also be nontoxic.

Both solubility and permeability are considered fundamental parameters for oral drug absorption. A biopharmaceutical classification scheme (BSC) has been proposed to categorize drugs into four groups, based on their aqueous solubility and permeability properties (Fig. 4) [13]. Drug molecules that belong to class I are the most promising, from the bioavailability and drug delivery point of view. If drug molecules in classes II to IV have suitable functional groups for prodrug derivatization, the prodrug approach can be applied in order to change molecules from classes II to IV to class I (Fig. 4).

**Biopharmaceutics Classification System**

| | |
|---|---|
| **Class I**<br>**Highly soluble**<br>**Highly permeable** | **Class II**<br>**Poorly soluble**<br>**Highly permeable** |
| **Class III**<br>**Highly soluble**<br>**Poorly permeable** | **Class IV**<br>**Poorly soluble**<br>**Poorly permeable** |

**Effect of Prodrugs on Biopharmaceutics Classification**

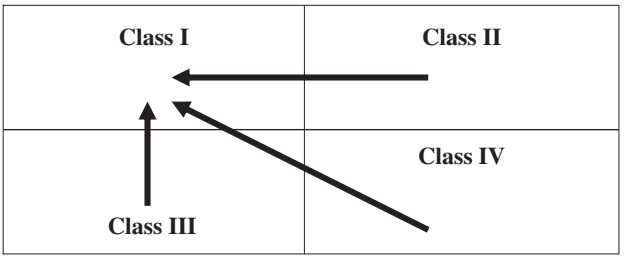


**Figure 4** Biopharmaceutic drug classes and the effect of prodrugs on biopharmaceutics classification.

## 4 BIOREVERSIBLE PRODRUG STRUCTURES FOR THE MOST COMMON FUNCTIONAL GROUPS

Prodrugs are commonly formed by linking a promoiety to a nucleophilic carboxyl, phosphate, hydroxyl, or amino group on the parent drug molecule via a labile linkage. The labile linkage is usually designed to be stable in a pharmaceutical formulation and be readily cleaved by enzymes in vivo to release the parent drug after administration. This labile linkage can also be susceptible to nonenzymatic chemical cleavage. The release of the parent drug can then be triggered by a pH shift from a relatively acidic (or basic) formulation to a physiological pH of 7.4 in vivo. However, the utility of the latter approach is somewhat limited by the difficulty of achieving sufficient stability in the pharmaceutical formulation [14].

Key enzymes that are most commonly targeted for the cleavage of a labile prodrug linkage include various esterases and alkaline phosphatases, which are ubiquitous in mucosal membranes, blood, liver, and other organs [15]. A variety of other enzymes [16], such as aminoacylases, cystein conjugate β-lysase, γ-glutamyltransferases, dipeptidases, amino- and carboxypeotidases, oxoprolinase, β-glucronidase [17], and azoreductase [18] can also be targeted for the bioconversion of a prodrug. However, this type of enzymatic activity

is not ubiquitous in all tissues, and prodrugs converted by these enzymes are, therefore, more suitable for organ-targeted drug delivery.

Many older drugs, as well as some new ones, are prodrugs that do not contain promoieties as defined above. For example, nabumetone undergoes a rapid liver oxidase catalyzed biotansformation to the bioactive form (Fig. 5) [19]. Similarly, various antiviral and anticancer nucleosidic drugs are transformed to their corresponding bioactive monophosphates and triphosphates in vivo in a kinase-mediated step. In such cases, these in vivo metabolic transformations are sometimes difficult to predict from in vitro investigations, and the metabolic processing of these drugs can vary considerably between and within animal species. The discussion in this chapter will, therefore, be limited to the former type of prodrugs, where a well-defined promoiety is linked to a suitable nucleophilic group on a parent drug molecule.

## Prodrugs for Carboxyl Groups

In general, the p$K_a$ of a drug carboxyl group is in the range of 2.5 to 5.5. In the upper gastric fluids (pH 1.3), the carboxyl group will be neutral, but mainly ionized in the small intestine and other physiological fluids (pH 7.4). Ionization does not necessarily prevent drug absorption, especially in the case of oral dosing, because at least some part of the drug will be in the uncharged form. However, masking the carboxyl group can reduce some side effects, such as local irritation. Passive absorption can also be enhanced by forming neutral lipophilic prodrugs.

The most common prodrugs are those requiring a hydrolytic bioconversion that is mediated by an enzymatic reaction. Therefore, prodrugs derived from carboxyl groups are commonly obtained by formation of esters [20], which are substrates for various esterases found throughout the blood, intestinal mucosa, liver, and other organs [21]. Moreover, it is possible to develop ester prodrugs that have higly variable aqueous solubility or lipophilicity characteristics by carefully selecting the prodrug moiety. Examples of promoieties for the carboxyl group are shown in Fig. 6.

***Alkyl and Aryl Ester Prodrugs*** Prodrugs of angiotensin-converting enzyme (ACE) inhibitors are some of the most successful prodrugs in clinical use today

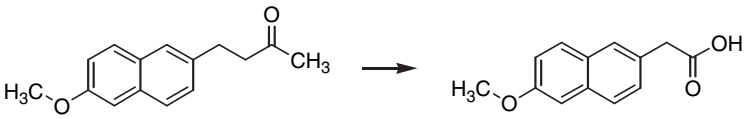


Nabumetone

Bioactive Form:
6-methoxy-2-naphthylacetic acid (6-MNA)

**Figure 5** Bioconversion of nabumetone to the bioactive form.

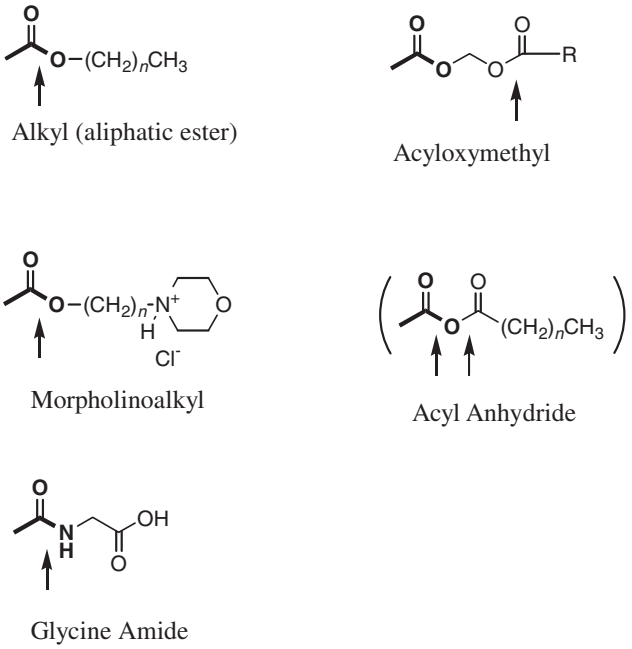


**Figure 6** Examples of promoieties for carboxyl groups. The arrows show the biolabile linkages.

[16]. Various ester derivatives have been investigated to improve the oral bioavailability of enalaprilate. While the proline carboxyl esters convert to diketopiperazine in an intramolecular cyclization reaction that is chemically unstable [22], esters on the *N*-carboxyalkyl group are chemically more stable and absorbed well after oral administration. A prototype of this class of prodrugs is enalapril, which is a monoethyl ester of the ACE inhibitor enalaprilate (Fig. 7). Other ACE inhibitors that have been introduced since enalapril, such as benzapril and ramipril, are also mono ethyl esters of diacids.

The main disadvantage of simple alkyl ester prodrugs is that they will tend to reduce aqueous solubility, relative to the parent drug. Since aqueous solubility and dissolution rates can be a limiting factor in drug absorption, low aqueous solubility can be addressed by introducing an amino group on the alkyl chain. A series of morphinoalkyl (ethyl, propyl, butyl) esters of indomethacin and naproxen [23] have been investigated in vitro and in vivo. These prodrugs were freely soluble in simulated gastric fluid and aqueous pH 7.4 phosphate buffer. The octanol/pH 7.4 phosphate buffer distribution coeffcient was increased by 40- to 1500-fold through prodrug formation, relative to the parent drug, whereas the octanol/simulated gastric fluid (pH 1.3) distribution coefficient decreased 3- to 20-fold. The esters were relatively stable at pH 7.4 with a half-life of 10 h, but they were rapidly converted to the parent

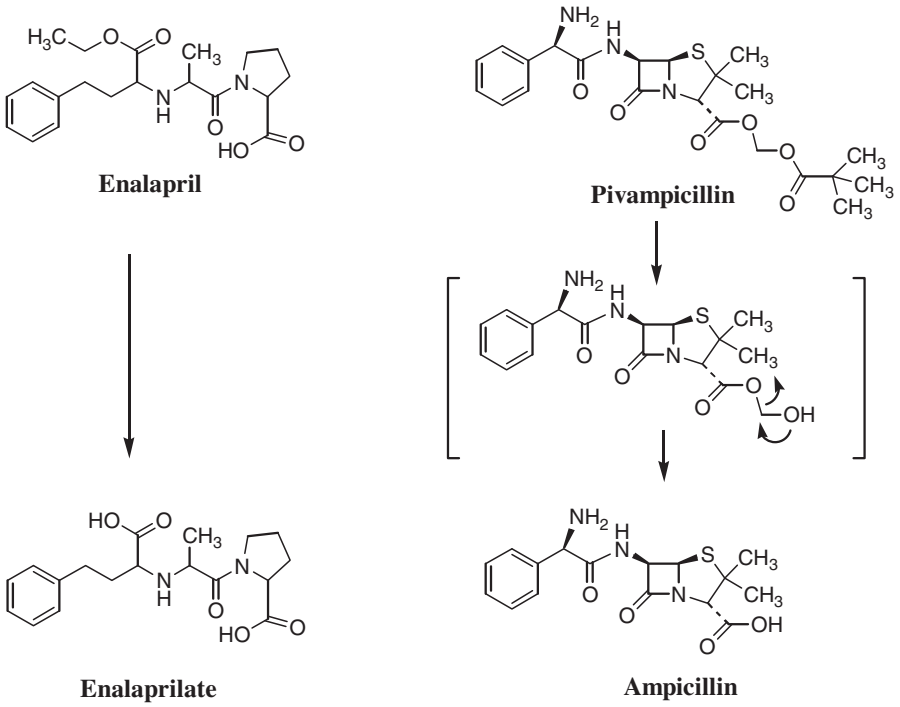


**Figure 7** Examples of prodrugs with modification on the carboxyl groups and their bioconversion.

drug in rat plasma with half-life of 1.2 to 31 min. These prodrugs improved the bioavailability of the parent drug only slightly, but gastric mucosal injury was significantly reduced.

***Acyloxyalkyl Prodrugs*** Bioconversion of simple alkyl esters may sometimes be inefficient in humans, due to steric hindrance around the carboxyl group on the drug molecule, which limits the therapeutic potential. α-Acyloxymethyl derivatives are double esters, from which the terminal ester moiety is first removed through an enzyme-catalyzed hydrolysis, with the formation of a highly unstable hydroxymethyl ester intermediate that subsequently releases both formaldehyde and the parent drug. In α-acyloxyalkyl esters that contain a longer alkyl chain (e.g., ethyl, propyl, butyl), hydrolysis may take place at the carbonyl of the parent drug, rather than the carbonyl of the promoiety [24].

As an example, various acyloxyalkyl esters of β-lactam antibiotics are well absorbed and efficiently hydrolyzed by esterases in vivo and in vitro, while prodrugs that are simple esters of the C-3 thiazolidine carboxyl group are ineffective in humans, due to limited hydrolysis [25]. The pivaloyloxymethyl derivative of ampicillin has been clinically introduced as pivampicillin (Fig. 7). More examples of ampicillin prodrugs are discussed in Section 5 of "Coupled

Luminescent Methods in Drug Discovery: 3-Min Assays for Cytotoxicity and Phosphatase Activity".

***Amides*** In general, amides are more stable toward hydrolysis than esters, but various amides formed with amino acids are in vivo substrates for endopeptidase [21]. For example, glycine amide prodrugs of ketoprofen and naproxen have been evaluated for efficacy in mice and gastrointestinal irritation in rats [26]. Hydrolysis in serum is relatively slow, but these prodrugs exhibit similar anti-inflammatory activity as the parent drugs. Such prodrugs contain a free carboxylic acid group, and the physico-chemical properties are, therefore, not significantly altered with the prodrug modification. However, gastrointestinal irritation is still somewhat reduced with the glycine amide prodrugs.

## Prodrugs for Phosphate and Phosphonate Groups

The phosphate and phosphonate groups are negatively charged zwitterions at nearly all physiological pH values, making them very polar and poorly absorbed from the GI tract. Moreover, drugs containing these groups also tend to have a low volume of distribution, and they are, therefore, subject to rapid and efficient renal clearance [27]. Prodrug strategies for phosphate and phosphonate functional groups usually mask the charge by chemical derivation to obtain neutral and better absorbed prodrugs (Fig. 8).

***Prodrugs for Phosphate Groups*** Phosphate prodrugs are primarily prodrugs for 5′-*O*-monophosphates nucleoside analogs. The unmodified 5′-*O*-monophosphate esters have very limited oral bioavailability, due to poor absorption and rapid dephosphorylation in the gastrointestinal tract [27]. Lipophilic dialkyl ester prodrugs are often well absorbed, but their bioconversion is sometimes slow and inefficient. For example, simple dialkyl phosphate esters (5′-*O*-monophosphate) of the antiviral drug zidovudine (AZT) have no activity toward inhibiting the human immunodeficiency virus (HIV) replication [28] because of poor bioconversion to AZT.

The in vivo enzymatic conversion of phosphate esters is catalyzed by phosphodiesterases and phosphonodiesterases [27]. The phosphate ester linkage to the nucleside is liable to both chemical and enzymatic cleavage. Phosphate ester prodrugs tend, therefore, to be converted to the parent nucleoside rather than the monophosphate nucleotide. One way to circumvent this problem is to form homo- or hetero-dinucleoside phosphate derivatives of biologically active nucleoside analogs. Hydrolysis of the ester bond produces one molecule of nucleoside and one molecule of the nucleotide. These two products should then be synergistically active in the cell. The 5′ → 5′ homo- and hetero-dinucleoside phosphates of the antiviral nucleoside analogs AZT, ddC, and ddI (Fig. 9), have been investigated for activity against HIV [27, 29]. However, the hydrolysis of these compounds was inefficient and, thus, these prodrugs exhibited lower activity than the parent drugs.

For Phosphate Groups:

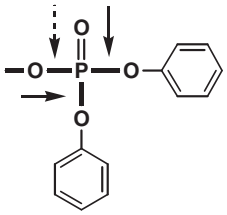

Diphenyl

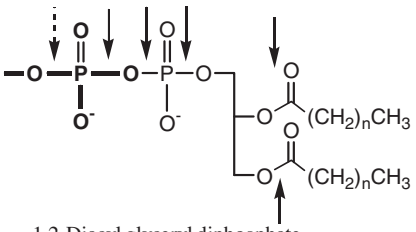

1,2-Diacyl glyceryl diphosphate

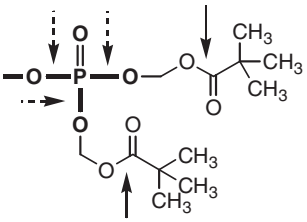

Di(pivaloyloxymethyl)

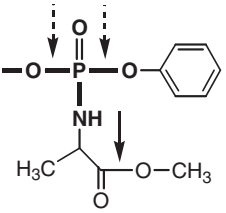

Phenyl methoxyalaninyl

For Phosphonate Groups:

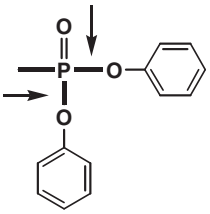

Diphenyl

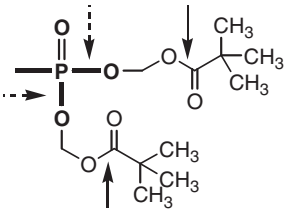

Di(pivaloyloxymethyl)

**Figure 8** Examples of promoieties for phosphate and phosphonate groups. The arrows show the biolabile linkages.

An attempt to better control the bioconversion of the prodrug is an acyloxymethyl phosphate ester, from which the acyloxy group is cleaved by esterases, and the resulting hydroxymethyl ester derivative is rapidly eliminated to release formaldehyde and the phosphate group. For example, from di(pivaloyloxymethyl) phosphate esters the first pivaloyloxy group is efficiently removed, whereas the hydrolytic cleavage of the second acyloxymethyl ester group can be considerably slower, due to lower affinity of the ionized compound for the enzyme [27]. A mechanistic study for the enzymatic conversion of di(pivaloyloxymethyl) AZT monophosphate (AZTMP) has shown that the hydrolytic cleavage of the first pivaloyloxmethyl group is catalyzed by esterases, whereas the second group is mainly removed through hydrolysis of the phosphate ester bond by phosphodiesterases [30].

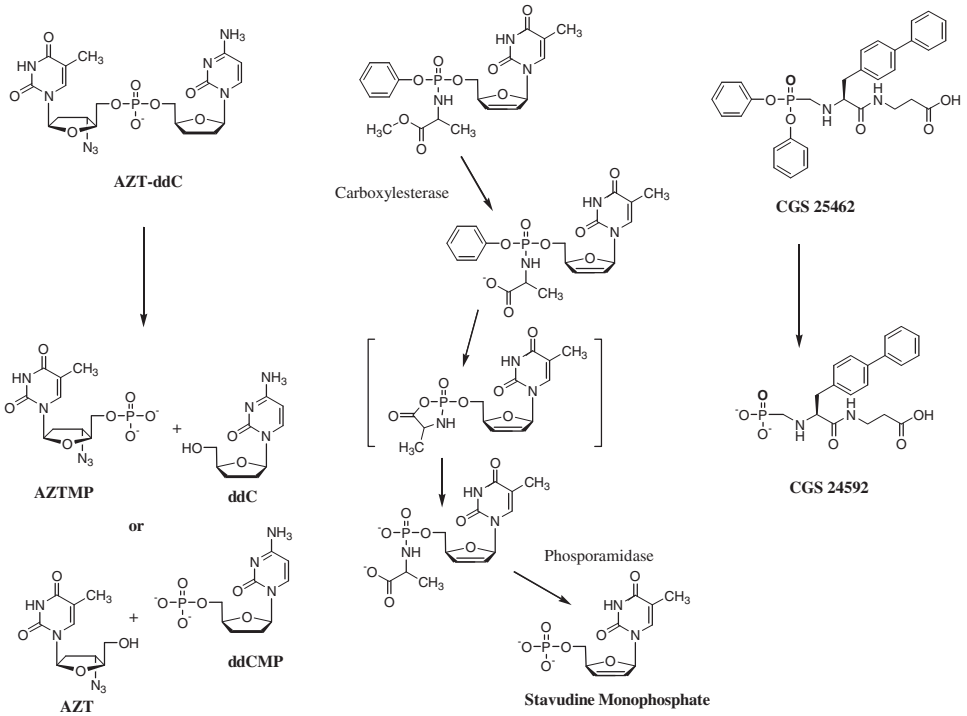


**Figure 9** Examples of prodrugs for phosphates and phosphonates and their bioconversion.

Diphosphate 1,2-diacyl glycerol derivatives have been investigated as prodrugs for nucleoside monophosphates. In this case, the phospholipid prodrug is targeted to the cellular membrane, and monophosphate is released in a phosphatase-catalyzed hydrolysis reaction. For example, acyclovir diphosphate dimyrstoylglycerol can be effective against the HIV virus in thymidine kinase deficient cell lines, whereas acyclovir and acyclovir monophosphate are ineffective [31].

A study of the degradation pathway for stavudine-5′-(phenyl methoxy-L-alaninyl) phosphate (Fig. 9) showed that the methoxy ester is first cleaved in a esterase-catalyzed reaction, followed by a spontaneous chemical hydrolysis of the aryl phosphoester, which is intramolecularly catalyzed by the free carboxylic acid group. The amino acid moiety is then efficiently removed in a reaction catalyzed by phosphoroamidases [32]. This prodrug was equally effective against HIV in normal and kinase-deficient cell lines. Because the bioconversion is dependent on the chemical structure of the amino acid and the nucleoside moiety, the methoxy-L-valinyl derivatives are stable toward enzymatic hydrolysis and have very little efficacy against HIV.

***Prodrugs for Phosphonate Groups*** The C-P bond linking a phosphonate group to the parent drug structure is not subject to enzymatic hydrolysis, and the selection of suitable promoieties is, therefore, less problematic than in the case of phosphate groups.

The most simple prodrug approach for increasing the lipophilicity of phosphonates is by esterification of the phosphonic OH groups. Several alkyl and aryl esters of phosphonates have been made, but they usually suffer from their high stability in biological media. In a series of dialkyl prodrugs, 9-[2-(phosphonomethoxy)ethoxy]adenine bioconversion to the parent drug depended on the chemical nature of the alkyl chain [33]. As the size of the alkyl chain increased, the prodrugs broke down more efficiently to the monoester, but the monoester failed to release the parent phosphonate. Similar results were obtained from a study of dialkyl prodrugs of 9-[2-(phosphonomethoxy)ethyl]adenine (PMEA) [34].

A series of dialkyl and diaryl prodrugs were studied in order to increase oral bioavailability of the neutral endopeptidase inhibitor CGS 24592 [35]. The diethyl and dicyclohexyl prodrugs were resistant to hydrolysis, whereas the diphenyl (Fig. 9) and di(acyloxyalkyl) derivatives were efficiently hydrolyzed in human plasma. Moreover, they significantly increased the bioavailability from <3 percent in rats.

Bis(acyloxymethyl) esters have been developed to overcome the hydrolytic resistance that has been observed with alkyl and aryl esters phosphonates. Investigations of prodrugs to increase the bioavailability of bisphosphonate antiosteoporosis drugs have shown that tri- and di(pivaloyloxymethyl) prodrugs of clodronic acid rapidly degrade in liver homogenates to release clodronic acid in quantitative amounts, whereas the benzoyloxypropyl esters were not subject to either enzymatic or chemical hydrolysis [36]. Monoamine derivatives were rapidly hydrolyzed at pH 7.4, but the reaction was not subject to enzymatic catalysis [37]. This suggests that the monoamine derivatives could be useful intermediates in the design of enzymatically labile prodrugs.

## Prodrugs for Hydroxyl Groups

Prodrugs are commonly derived from existing hydroxyl groups on the parent drug molecule, usually by ester formation (Fig. 10). The esters are typically formed by a condensation reaction between the hydroxyl group and a carboxylic acid or by reaction with an acid halide. This condensation is commonly carried out by using a coupling agent, such as 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide and a catalyst such as 4-dimethylaminopyridine (DMAP). Both aliphatic and aromatic esters can be substrates for nonspecific esterases, which are widely distributed throughout various tissues and blood serum. Phosphate esters are substrates for phosphatases found in blood and as intestinal-brush border-bound enzymes.

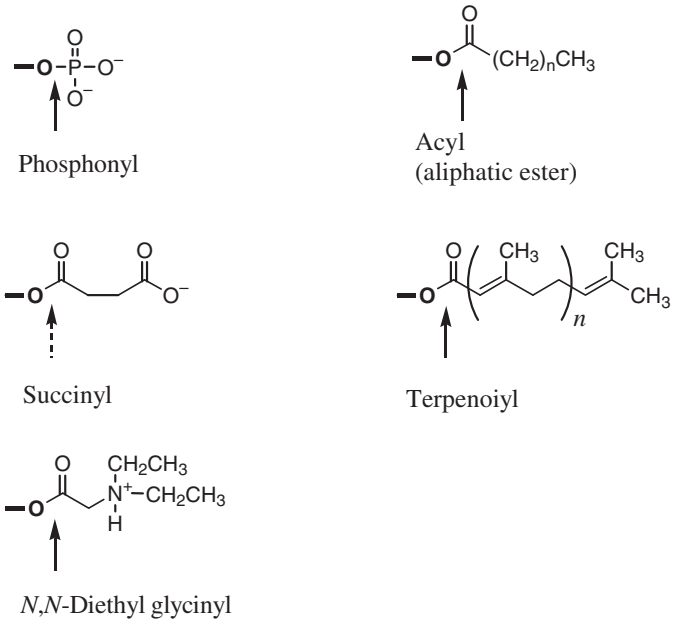


**Figure 10** Examples of promoieties for hydroxyl groups. The arrows indicate the biolabile linkages.

***Water-Soluble Prodrugs*** Water-soluble prodrugs are commonly obtained as phosphates, succinates, or amino acid esters of the hydroxyl group. For example, succinate esters and phosphate esters of corticosterioids have been used for intravenous administration. In general, the succinate esters have a limited chemical stability [20], due to intramolecular hydrolytic catalysis of the ester bonds by the neighboring carboxyl group [38]. These compounds are, therefore, marketed as a freeze-dried powder, which is dissolved just prior to administration. Due to anionic charges, the succinate esters are relatively poor substrates for carboxylesterases. Therefore, they are only slowly converted after intravenous administration and partially eliminated unchanged in urine [39, 40].

Phosphate esters are more stable toward chemical hydrolysis and can be formulated as aqueous solutions with an acceptable shelf life [20]. Phosphate esters of corticosteroids, such as methylpredinsolone (Fig. 11) [38] are rapidly and quantitatively converted to the parent drug in vivo.

Amino acid esters, which are substrates for endopeptidases in vivo [21], are also considered as suitable promoieties that improve aqueous solubility. The synthesis of these prodrugs involves either condensation with *N*-protected amino acids or reaction of an amine with an α-halogenacetate ester of the

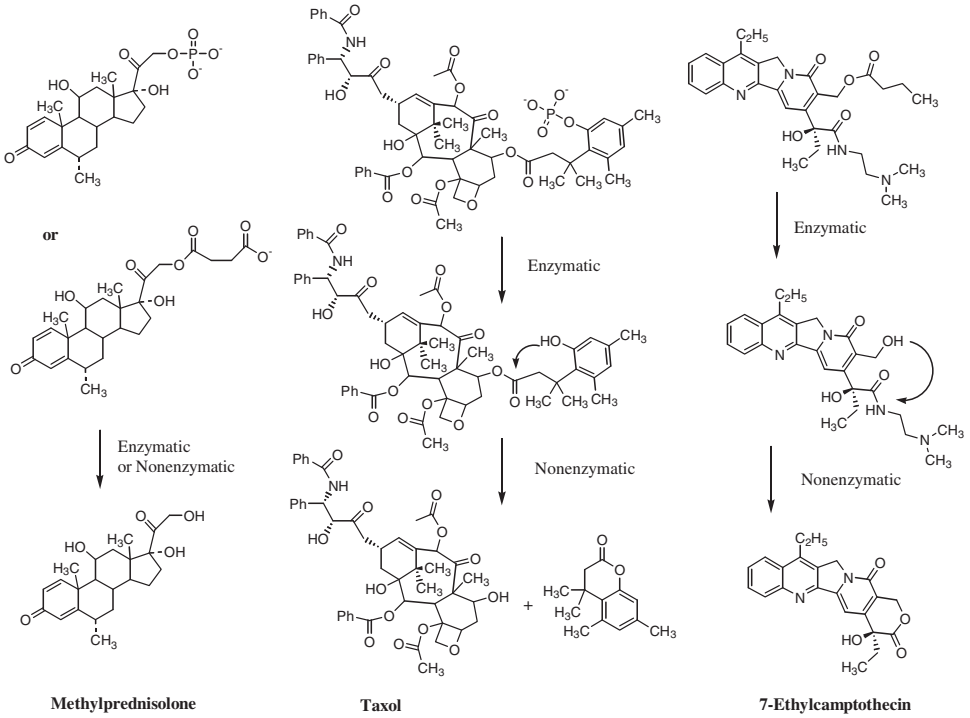


**Figure 11** Prodrugs derived from hydroxyl groups and their bioconversion.

parent drug. Amino acid derivatives such as *N*-propyl-, *N*,*N*-dimethyl, and diethtyl glycine esters, 4-morpholinoacetate, and methyl-1-pirperazinoacetate [41, 42] were freely water-soluble and their half-lives for bioconversion were less than 30 min. However, their chemical stability was limited. In general, α-amino acid esters tend to be unstable in aqueous solution due to an electron-withdrawing effect of the protonated amino group [20].

Hydrophilic corticosteroid prodrugs can also be obtained by forming a glycoside derivative from the primary hydroxyl group. Glycosides can be relatively stable in the small intestine but easily bioconverted by bacterial glycosidases in the colon. This type of prodrug, for example, dexamethasone-β-D-glucuronide [17], is therefore intended for colon-specific delivery.

***Lipophilic Prodrugs*** Lipophilic produgs of drugs containing a hydroxyl group can be obtained by esterification with a carboxylic acid that has a non-polar side chain. The purpose of producing a more lipophilic prodrug is to enhance both absorption and transmembrane delivery of polar drugs. For example, various lipophilic linear aliphatic esters of metronidazole have been investigated for dermal delivery. Enzymatic conversion in human plasma and skin homogenates is faster with long-chain aliphatic esters, such as caproyl

esters, than with shorter chains, especially acetyl and propionyl esters. In vitro investigations have shown that metronidazole acetate can penetrate through excised mouse skin and is partially metabolized to metronidazole [43].

***Promoieties to Overcome Steric Hindrance*** Steric hindrance in a drug structure can affect the enzymatic lability of esters formed at the hydroxyl group. In such cases, the lack of enzymatic lability can be addressed by introducing a suitable spacer between the modified hydroxyl group and the ester function. The ester is then cleaved in an enzymatic reaction, followed by a rapid intramolecular chemical reaction of the spacer group to release the parent drug.

Taxol phosphates have, for example, been investigated as prodrugs for the poorly soluble anticancer drug taxol [44]. Simple phosphate esters were more soluble than the parent drug, but they were also stable toward alkaline phosphatase in plasma and consequently ineffective in vivo. The steric hindrance around the phosphate ester was reduced by introducing an oxyphenylpropionate spacer group (Fig. 11). These phosphate esters were readily cleaved in purified alkaline phosphatase preparations, and the remaining hydroxyphenylpropionate group was eliminated through an intramolecular lactonization reaction. However, these prodrugs were not converted in animal plasma, apparently due to tight protein binding, and consequently had no in vivo activity.

***Prodrugs of Lactones*** Prodrugs of lactones can be obtained by forming an ester or an amide with the hydroxyl group in the open ring of a lactone. When incubated in plasma, they revert to the parent drug in a two-step process, with initial enzymatic degradation of the ester or amide on the hydroxyl group followed by spontaneous ring closure to the lactone. This approach has been used for water-soluble prodrugs of camptothecin derivatives (Fig. 11), where an amide is formed from the carboxyl group of the open lactam and an enzyme-labile acyl moiety is attached to the hydroxyl group [45, 46].

## Prodrugs for Amines and Amides

Nucleophilic amino groups are present in many drugs, and they can be found as either primary, secondary, or tertiary aliphatic or aromatic amines. They can also be part of resonance-stabilized systems, such as heterocyclic aromatic groups, amides, imides, or guanidines. Depending on structure, the $pK_a$ of a protonized basic amino group can vary from less than 2 to more than 11. Amino groups in resonance-stabilized systems can also act as weak acids. Acidic amino groups will generally be neutral at physiological pH values, and the charge of a moderately basic amino groups is not necessarily a barrier to drug absorption. However, amino groups are often subjected to modification, especially if readily esterifiable hydroxyl or carboxyl groups are not present in the drug

structure [14]. Various promoieties have been investigated for amines, with particular consideration toward the nature of the amino group (Fig. 12).

**N-*Acyl Prodrugs*** The *N*-acylation of a basic amino group gives an amide, which has limited utility in prodrug design due to their relative in vivo stability [14]. However, certain amides can be substrates for specific enzymes in vivo. For example, the γ-glutamyl derivatives of dopamine are cleaved by γ-glutamyl transferases in the kidneys and have been considered for kidney-targeted drug delivery [14, 47]. Biolabile prodrugs can, however, sometimes be obtained by acylation of relatively acidic amides or imides. For example, the half-life of the $N^3$-acetyl derivative of 5-fluorouracil (Fig. 13) is 40 min at pH 7.4 (37°C) and only about 4 min in 80 percent human plasma [14].

**N-*Mannich Bases*** *N*-Mannich bases have been considered as chemically reversible prodrugs for NH-acidic compounds (e.g., amides, imides) and both aliphatic and aromatic amines. *N*-Mannich bases decompose in aqueous, alkaline, and slightly acidic solutions. However, they are relatively stable in acidic formulations, and the release of the parent drug is, therefore, triggered by a shift to higher pH values after administration. For example, the *N*-Mannich base formed between salicylamide and piperidine is relatively stable below pH 4, whereas the half-life for hydrolysis at pH 7.4 (37°C) is 14 min [48]. The *N*-Mannich base promoiety can also be utilized for water-soluble prodrugs of drugs with NH-acidic amino (amides) groups. For example, various *N*-Mannich bases of carbamazepine (Fig. 13) have been developed as water-soluble prodrugs for parenteral administration [49].

**N-*α-Acyloxymethyl and N-Phosphoroxymethyl Derivatives*** Probably the most successful and common approach to prodrugs from amines is to form *N*-α-acyloxymethyl or *N*-phosphoroxymethyl derivatives. The bioconversion of these prodrugs undergoes a two-step process; first an enzymatic cleavage of the ester bond, catalyzed by esterases or phosphatases, followed by a spontaneous chemical decomposition of the *N*-hydroxymethyl intermediate, which releases formaldehyde and the parent drug. Useful prodrugs cannot be obtained by *N*-α-acyloxymethylation of primary or secondary amines due to the extreme lability of such derivatives [14]. This approach is, therefore, limited to acidic imides or amides. For example, various such promoieties have been considered for the sparingly soluble and poorly absorbed anticonvulsant phenytoin. The water-soluble *N*-phosphoroxymethyl derivative of fosphenytoin is now marketed as Cerebyx and is further discussed in Section 5 of "Coupled Luminescent Methods in Drug Discovery: 3-Min Assays for Cytotoxicity and Phosphatase Activity". The *N*-α-acyloxymethyl and *N*-phosphoryloxymethyl promoieties have also been used to obtain either water-soluble or lipophilic prodrugs of theophylline (Fig. 13) [50], allopurinol [51], 6-mercaptopurine [52], and 5-fluorouracil [53].

Promoieties for Basic Amino Groups:

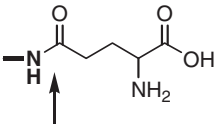


*N*-γ-Glutamyl

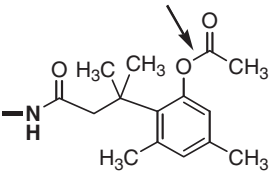


*N*-3(2'-Acetoxy-4',6'dimethylphenyl) 3,3-Propionyl

Promoieties for Acidic Amino Groups:

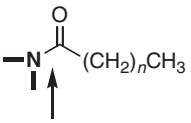


*N*-Acyl

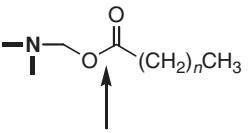


*N*-Acyloxymethyl

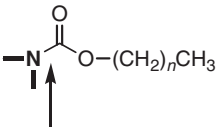


*N*-Acyloxycarbonyl

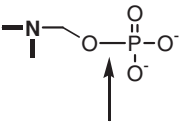


*N*-Phosphoroxymethyl

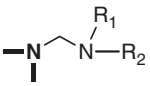


*N*-Mannich base

**Figure 12** Examples of promoieties for amino groups. The arrows indicate the biolabile linkages.

## 5 PHARMACEUTICAL APPLICATIONS OF THE PRODRUG STRATEGY

### Improved Drug Absorption

***Oral Drug Delivery*** For orally administered drugs, the intestinal epithelia serves as physical barrier by forming tight cellular junctions that restrict paracellular drug delivery into the systemic circulation. Consequently, efficacious drugs must possess either optimal physico-chemical properties (e.g., lipophilic-

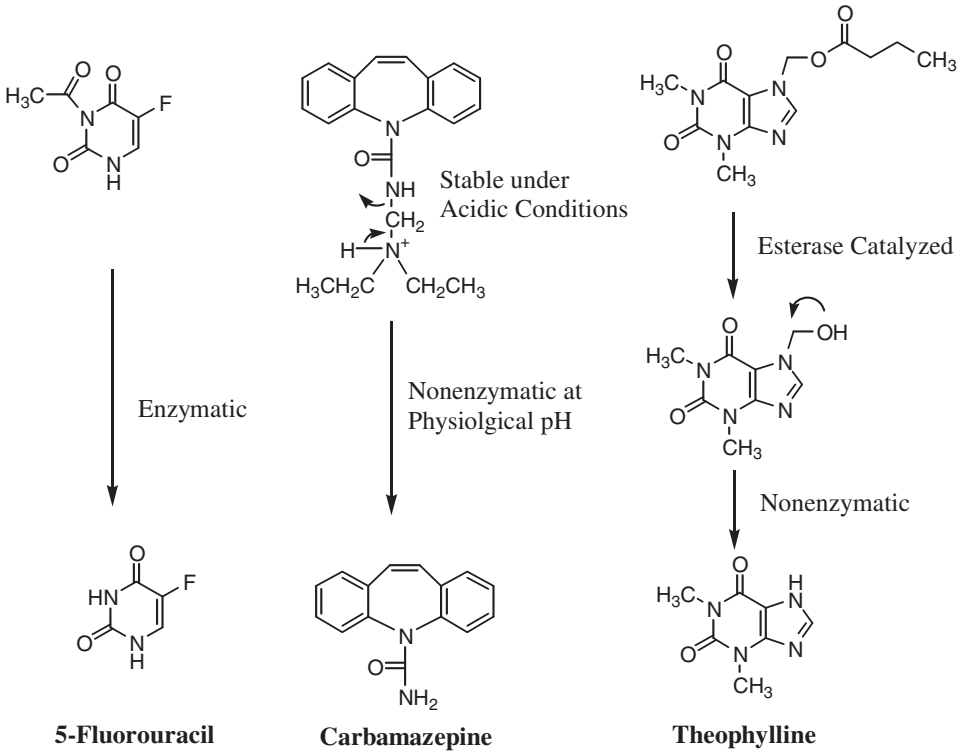


**Figure 13** Prodrugs derived from amino and amide groups and their bioconversion.

ity, sufficient molecular weight, ionic charge, hydrogen-bonding potential, or structural conformation) in order to passively diffuse via the transcellular route, or the drug must have structural features that allow it to be taken up by one of the endogenous transport mechanisms present at the intestinal epithelium [54]. Other prerequisites for a drug's effective oral absorption are sufficient aqueous solubility at the absorption site and adequate chemical stability in the GI tract or sufficient stability during first-pass metabolism.

*Lipophilic Prodrugs* Since nonfacilitated and largely nonspecific passive transcellular transport is the most common absorption route in oral drug delivery, lipophilicity can be increased to achieve better diffusion across the lipoidal biological membranes, and thus better oral bioavailability. This approach has been one of the major research areas of prodrug research. Well-known and successful examples of oral prodrugs are various lipophilic ampicillin derivatives. Ampicillin is highly polar and in a zwitterionic form in the pH range of the gastrointestinal tract and has an oral bioavailability of only 30 to 40 percent. A transient increase in its lipophilicity by esterification resulted in

clinically useful ester prodrugs of ampicillin (Fig. 14), namely pivampicillin (pivaloylmethyl ester), bacampicillin (ethoxycarbonylethyl ester), and talampicillin (phtalidyl ester), which are all well-absorbed and rapidly hydrolyzed to ampicillin either during or after absorption [25].

More recent examples, for example, tenofovir disoproxil, adefovir dipivoxil, ximelagatran, and BIBR 1048, show further proof of the concept. In particular, the acyclic nucleoside phosphonate 9-[2-(phosphonylmethoxy) propyl]adenine (PMPA, tenofovir) (Fig. 15) has potent and selective activity against HIV. However, high hydrophilicity of the phosphonic acid moiety has been postulated to account for its poor oral bioavailability (less than 0.1 percent in Caco-2 cell monolayer model). On the basis of its favorable in vitro chemical and enzymatic degradation properties, and octanol/water partition coefficient [$\log P$ (pH 6.5) = 1.3], bis(isopropyloxycarbonyloxymethyl)PMPA [bis(POC)-PMPA, tenofovir disoproxil] was selected for further development [55]. In clinical studies, tenofovir disproxil has been well tolerated, with an oral bioavailability in fasted patients of approximately 25 percent [56]. Similarly, the lipophilic bis[(pivaloyloxy)methyl] ester prodrug adefovir (PMEA), a close hydrophilic analog of tenofovir (Fig. 15), enhanced the oral bioavailability up to 30 to 40 percent, compared to only 12 percent when adefovir alone was administered in HIV-infected patients [57]. Recently, both tenofovir disoproxil and adefovir dipivoxil have been approved for the market. Furthermore, lipophilic prodrugs such as ximelagatran and BIBR 1048 (Fig. 16) may be the first examples of orally administered thrombin inhibitors [58].

*Carrier-Mediated Transport* This means of transport is particularly important where drugs or prodrugs are either polar or charged, and passive transcellular absorption is neglible. Therefore, the targeting of intestinal epithelial membrane transporters has become an attractive approach for improving the oral bioavailability of poorly absorbed drugs. By using a bioreversible prodrug derivatization strategy, a drug with low membrane permeability is converted

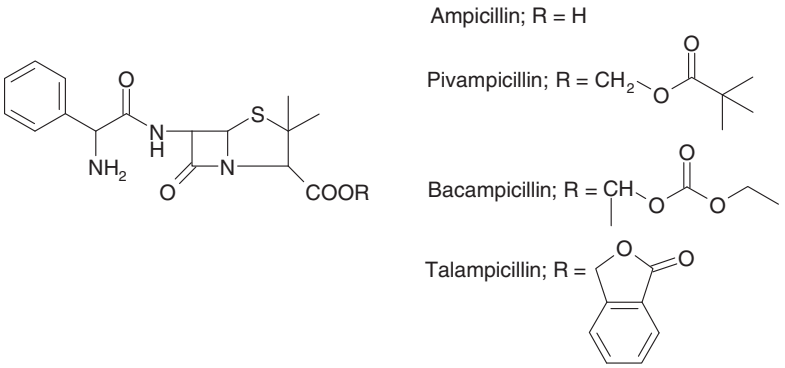


**Figure 14** Prodrugs of ampicillin.

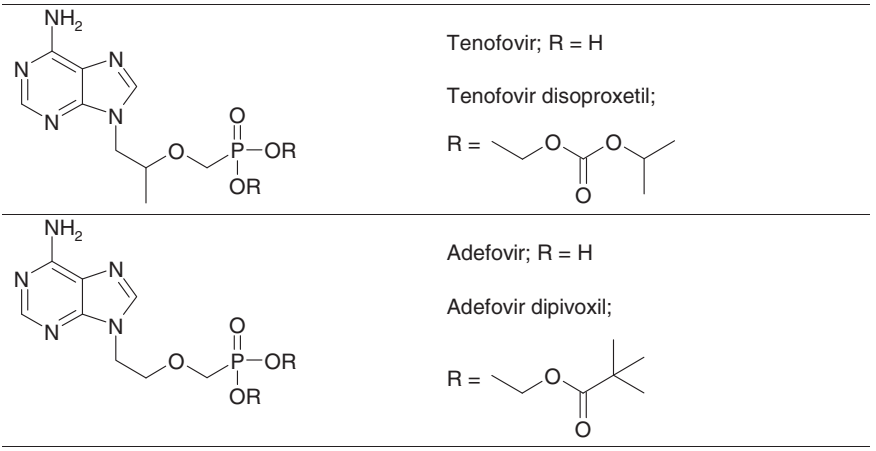


**Figure 15** Prodrugs of tenofovir and adefovir.

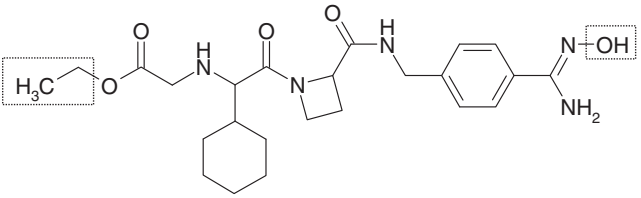

Ximelagatran (promoieties marked with the rectangle)

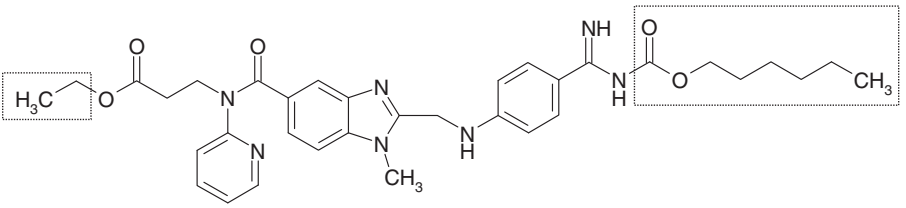

BIBR 1048 (promoieties marked with the rectangle)

**Figure 16** Prodrugs of thrombin inhibitors.

to a prodrug that is a substrate for the endogenous transport system. The nutrient transport systems have mainly been investigated and utilized for the enhancement of intestinal absorption, and these primarily include peptide, amino acid, bile acid, and glucose transporters. Other systems seem to have narrow structural requirements that currently make them less relevant for this purpose. Among the membrane transporters, the peptide transport mechanism appears to be most promising as it is widely distributed throughout the small intestine and shows low substrate specificity, together with high transport capacity [54, 59, 60]. Therefore, in addition to endogenous peptides, various

drugs can be recognized as substrates by peptide transporters, even drugs that do not include amino acid residues.

A good example of a prodrug that exploits intestinal transporters is valacyclovir, the L-valyl ester of acyclovir (Fig. 17). Acyclovir is a specific and selective inhibitor of viral herpes replication and is neither highly lipid nor aqueous soluble, with limited and variable oral bioavailability (15 to 21 percent) that decreases with increasing doses [61]. Conversely, valacyclovir has an oral bioavailability of three to five times higher than that of acyclovir itself. Moreover, valacyclovir is rapidly and extensively hydrolyzed to acyclovir after oral administration [62, 63]. Studies of its transport mechanisms have demonstrated that the improved oral bioavailability is due to its active transport by the human intestinal peptide transporter (PEPT1) [64]. Similarly, biolabile L-valyl ester prodrugs of AZT and ganciclovir afforded 3- to 10-fold better intestinal permeability than their parent drugs, and their membrane transport was mediated predominantly by the peptide transporter [65, 66], although they do not possess an amino acid residue in their structures.

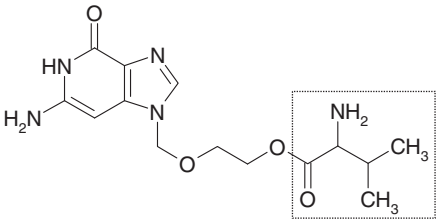


**Figure 17** Valacyclovir (promoiety is marked by the rectangle).

*Others* For poorly water-soluble drugs, dissolution in the GI tract can be a rate-limiting step for intestinal absorption, leading to poor and variable oral bioavailability. For example, milproxifene phosphate has a higher aqueous solubility than its insoluble parent drug, milproxifene, which results in a nearly 10-fold increase in permeation across the Caco-2 cell monolayer, due to its better dissolution properties [67]. Increasing aqueous solubility by the prodrug approach is more thoroughly discussed below.

Due to its great instability in acidic environment, as well as its hydrophilic character, the semisynthetic β-lactam antibiotic carbenicillin has been only used parenterally in clinical practice. Its facilitated oral absorption was achieved by bioreversible esterification of the side-chain carboxyl group, which resulted not only in improved lipophilicity but also in greater acid stability [68]. Carbenecillin is released from the prodrug carindacillin with relative ease by enzymatic hydrolysis, following absorption.

***Dermal Drug Delivery*** The skin, and in particular its most superficial layer, the stratum corneum, provides a unique and impressive resistance to molecular transport in either direction. Most existing drugs do not possess the

requisite physico-chemical properties for efficient percutaneous absorption, and therapeutic levels of these agents cannot be achieved in either tissue or blood. It has been demonstrated in numerous studies that the ideal properties for a drug molecule to readily diffuse across the skin are both adequate hydrophilicity and lipophilicity (i.e., adequate water and lipid solubility) and an optimal partition coefficient [69]. Especially the partition coefficient seems to be crucially important in establishing a high initial drug concentration within the superficial skin layers. These optimal features can often be achieved by the prodrug approach, which is a very efficient means of manipulating drug–skin and drug–vehicle interactions. Prodrugs have been studied in both topical drug delivery (local treatment preferred) and transdermal drug delivery (systemic treatment preferred) [70, 71]. The principle of dermal and transdermal prodrug delivery is illustrated in Figure 18.

*Topical Drug Delivery* Naproxen is a nonsteroidal anti-inflammatory drug (NSAID) with both analgesic and antipyretic properties. It is suitable for the treatment of both topical and some soft-tissue disorders. Unfortunately, its topical administration in humans shows a bioavailability of only 1 to 2 percent, which may be attributed to a very polar and ionizable carboxylic acid group. Various naproxen prodrugs (Table 1) were studied in attempts to increase the topical penetration of naproxen by transiently masking the carboxylic acid by esterification [24, 73–75]. Acyloxyalkyl, aminoacyloxyalkyl, and aminoalkyl esters of naproxen readily hydrolyzed to naproxen in vitro and possessed highly variable aqueous solubilities and lipophilicities. The highly lipophilic acyloxyalkyl esters did not enhance the dermal permeation of naproxen across

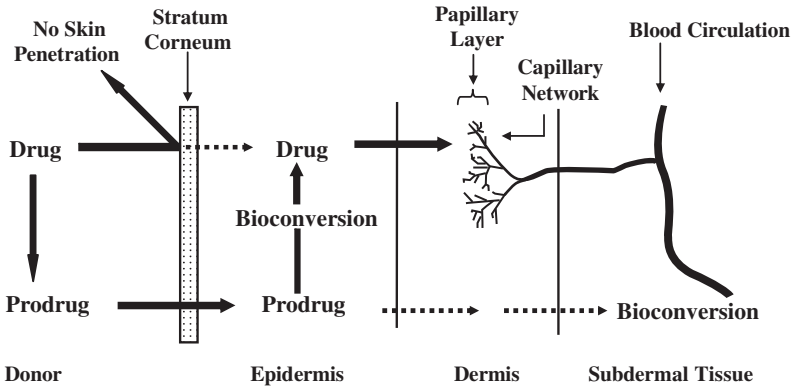


**Figure 18** An illustration of the dermal and transdermal prodrug concept (modified from Ref. 72). In topical (i.e., dermal) drug delivery, bioconversion of the prodrug into the parent drug must take place by enzymes in the skin immediately after application, but when a systemic effect is desired (transdermal drug delivery), release of the parent drug can take place in the cutaneous circulation as well.

**TABLE 1 Apparent Partition Coefficients, Aqueous Solubilities, and Steady-State Fluxes of Naproxen and Its Prodrugs**

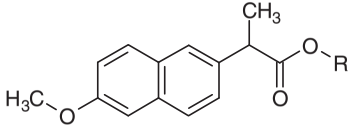


| Compound | R | Log $P_{app}$[a] pH 7.4 | Aqueous Solubility pH 7.4 (mM) | Flux[b] (nmol/cm$^2$·h) |
|---|---|---|---|---|
| Naproxen | H | 0.30 ± 0.03 | 101.9 ± 1.3 | 6.5 ± 0.6 |
| Acetyloxy-ethyl ester | $-(CH_2)_2-O-CO-CH_3$ | 3.50 ± 0.03 | 0.060 ± 0.003 | 0.4 ± 0.1 |
| Methyl-piperazinyl-acyloxyethyl ester | 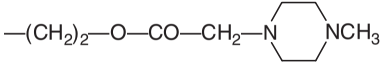 | 1.16 ± 0.08 | 4.1 ± 0.4 | 24.6 ± 1.0[c] |
| Methyl-piperazinyl-ethyl ester | 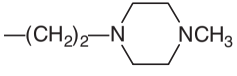 | 2.29 ± 0.02 | 0.37 ± 0.03 | 58.5 ± 13.0[c] |
| Methyl-piperazinyl-butyl ester | 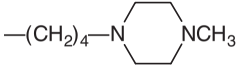 | 2.44 ± 0.09 | 1.13 ± 0.09 | 27.7 ± 1.8[c] |

[a] $P_{app}$ is the apparent partition coefficient between 1-octanol and phosphate buffer (0.16 M) at room temperature.
[b] The delivery of total naproxen species through excised human skin in vitro in isotonic phosphate buffer (0.05 M, pH 7.4) at 37°C.
[c] $p < .05$ compared to naproxen (ANOVA, Fisher's PLSD test).

the human skin in vitro, most probably due to their very low aqueous solubilities. In contrast, members in a series of aminoacyloxyalkyl esters (amino acids as the promoiety) and aminoalkyl esters had a combination of adequate aqueous solubility and lipophilicity, and gave significant in vitro improvements in the skin permeation of naproxen (Table 1). Moreover, the prodrugs hydrolyzed simultaneously during permeation, which is of great interest to move forward in the development of topical dosage forms.

It is also interesting note that at pH 7.4, where the highest flux values of both naproxen and prodrugs were achieved, significantly lower concentrations of prodrugs were used in the donor compartment than that of naproxen to achieve dermal transport. The permeability coefficients, which, unlike flux, are independent of donor concentration, were higher (up to 998-fold) for almost all prodrugs when compared to that of naproxen at pH 7.4. However, at pH 5.0 only two prodrugs exhibited higher permeability coefficients than naproxen, despite their flux values being lower than those for any other studied compound. Because the permeability of the studied compounds increased from aqueous solutions as the solubility of the compounds in these solutions decreased, the high permeability coefficient of these prodrugs can be explained by their very low aqueous solubility and decreased concentration in the applied vehicle, when compared to other studied compounds. Therefore, the flux, which measures the mass of material transported through the skin, may be a more relevant therapeutic parameter than the permeability coefficient.

*Transdermal Drug Delivery* Propranolol, an adrenergic β-blocker, is a widely used and clinically effective cardiovascular agent that is indicated in the treatment of angina pectoris, hypertension, and cardiac arrhythmia. However, low and variable bioavailability follows the oral admistration of propranolol, which may be attributed to extensive stereoselective hepatic first-pass metabolism. Because the transdermal route of administration allows for systemic drug absorption while eliminating hepatic first-pass effects, the relatively impermeable propranolol was made more lipophilic by the formation of diastereomeric ester prodrugs [76]. Subsequent in vitro diffusion studies across rat skin demonstrated a 3- to 12-fold higher flux value associated with propranolol than with the prodrugs. On the other hand, the permeability coefficient was higher for all prodrugs, due to their lower solubilities in the donor phase, compared to that of propranolol. Based on the hydrolysis data in conjunction with the flux values, it could be concluded that drugs having both optimum partitioning characteristics and resistance to enzymatic hydrolysis during permeation were able to afford the highest transdermal permeation.

In summary, the design of dermal prodrugs (topical or transdermal) not only includes the optimization of physico-chemical properties (lipid and aqueous solubility) but also considers the evaluation of enzymatic degradation properties within permeated membranes. When local treatment is

preferred, the active parent drug should be released during permeation or immediately afterwards. When systemic treatment is preferred, the parent drug can be released into the systemic circulation as well.

***Ophthalmic Drug Delivery*** In general, the site of action for ophthalmic drugs is located inside the eye. Thus, topical delivery of eyedrops into the lower cul-de-sac of the eye is the most common method of drug treatment in ocular disease. After topical administration, an aqueous eyedrop solution mixes with tear fluid to be dispersed over the eye surface. However, various precorneal factors (i.e., drainage of instilled solution, noncorneal absorption, induced lacrimation) limit ocular absorption by shortening the cornea contact time of applied drugs. These factors, and the corneal barrier itself, limit penetration of a topically administered ophthalmic drug. As a result, only a few percent of the applied dose is delivered into the intraocular tissues, the major part (50 to 99 percent) being absorbed into the systemic circulation (Fig. 19), where it can cause various unwanted side effects.

The cornea is generally considered to be a major pathway for the ocular permeation of topically applied drugs [77]. Compared to many other epithelial tissues (e.g., bronchial, intestinal, nasal, tracheal), the corneal epithelium is relatively impermeable but less than the stratum corneum of the skin [78]. For corneal permeation, lipophilicity of the drug seems to be the most important property, and both parabolic [79] and sigmoidal [80] curves have been

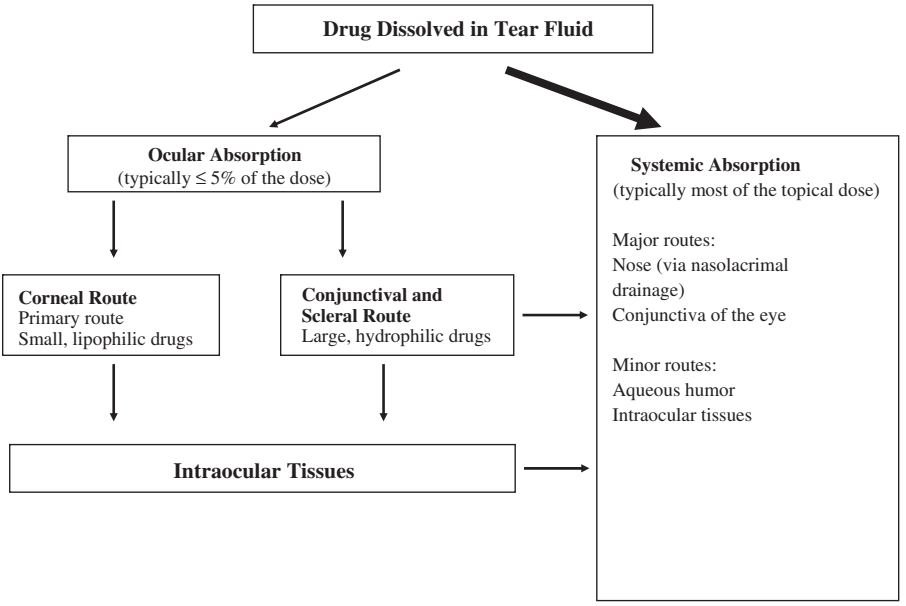


**Figure 19** Main absorption routes of a topically applied ophthalmic drug.

used to describe the influence of drug lipophilicity on corneal drug permeation. The optimum apparent partition coefficient ($P_{app}$; octanol/pH 7.4 buffer) for corneal drug absorption is in the range of 100 to 1000 [81, 82], which is consistent with the lipophilic nature of the corneal epithelium [83]. Aqueous solubility is another drug property important for the efficacy of ophthalmic drug delivery. The surface of the eye is constantly being cleaned and moistened by the aqueous tear fluid. Thus, it is difficult for drug molecules to be absorbed by the corneal epithelium unless they are soluble in the tear film. In addition, the drug's water solubility must be sufficient to enable the formulation of aqueous eyedrops. The dilemma is that an ideal potential ophthalmic drug should be simultaneously both water soluble and lipid soluble, and only a few molecules can fulfill these severe criteria. Because of this fact, various pharmaceutical technologies, such as prodrugs and cyclodextrins, have been applied to improve the physico-chemical properties of ophthalmic drugs.

Ocular absorption of a drug can be substantially enhanced by increasing its lipophilicity (hydrophilic compounds) or aqueous solubility (poorly water-soluble compounds). Prodrug derivatization, as a method to achieve these objectives, is a potential and a very useful approach that was introduced to ophthalmology about 20 years ago when the ocular absorption of epinephrine was substantially improved by this approach.

Dipivalyl epinephrine (dipivefrine) is a dipivalic acid diester of epinephrine (Fig. 20) that releases epinephrine after corneal absorption [84, 85]. Dipivefrine penetrates the human cornea 17 times faster than epinephrine [86] due to the fact that dipivefrine is 600 times more lipophilic (at pH 7.2) than epinephrine [87]. Consequently, a smaller topical dose of dipivefrine achieves a similar therapeutic effect in the eye [88] and the smaller prodrug dose results in decreased systemic levels of epinephrine. Compared to the conventional 2 percent epinephrine hydrochloride eyedrop, a 0.1 percent dipivefrine eyedrop results in an only slightly less effective IOP lowering effect, while side effects are greatly reduced [89].

Prostaglandin analogs represent a new class of active ocular hypotensive agents. Commercially available latanoprost, travoprost, and unoprostone (Fig. 21) are lipophilic isopropyl ester prodrugs that rapidly hydrolyze in the cornea to the biologically active prostaglandin [90, 91, and references cited therein]. The present lipophilic prodrugs achieve efficient ocular absorption and make the topical administration possible with decreased systemic side effects. Various lipophilic prodrugs have been studied in order to enhance ophthalmic absorption of drugs such as pilocarpine [8, 92], timolol [93], and tilisolol [94].

Another possible prodrug strategy for enhancing drug absorption is a prodrug design based on the structures of membrane transporters. However, very little is known about the presence of various transporters on the corneal epithelium. For example, hydrophilic L-valyl ester of acyclovir is threefold more permeable across the intact rabbit cornea than the parent drug [95]. This enhanced permeability suggests the presence of a carrier-mediated transport

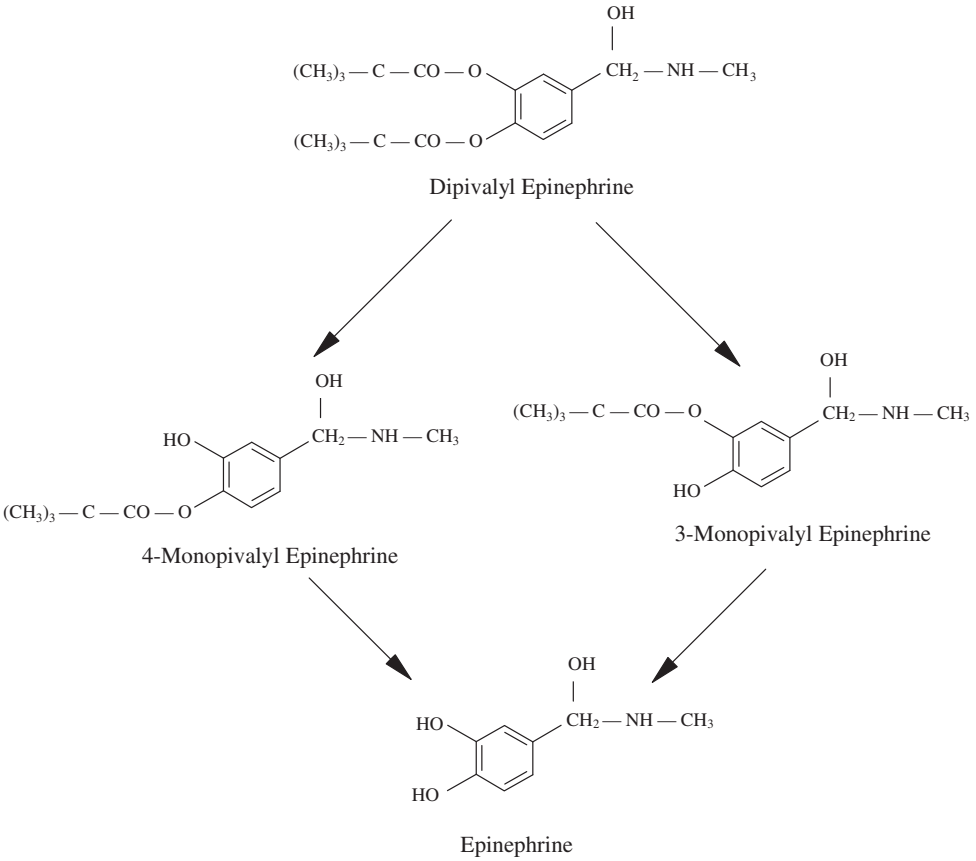


**Figure 20** Bioconversion of dipivalyl epinephrine.

mechanism that is specific for peptides on the cornea and may be targeted to improve the ocular absorption of poorly absorbed drugs.

***Nasal Drug Delivery*** Systemic drug delivery via the nasal cavity circumvents the gastrointestinal degradation and first-pass elimination associated with oral drug delivery. Moreover, the nasal route can be an attractive alternative to parenteral administration. In the healthy nose, epithelium is covered by a 2- to 4-mm layer of aqueous mucus, which is renewed approximately every 10 min. This aqueous mucus layer is one of three physical barriers for drug absorption by the nasal cavity. The other two barriers are the lipophilic epithelium lining the nasal cavity and an effective enzymatic pathway, both of which result in presystemic drug degradation. The enzymatic barrier in nasal administration is, however, not as efficient as in drug metabolism in the gastro-

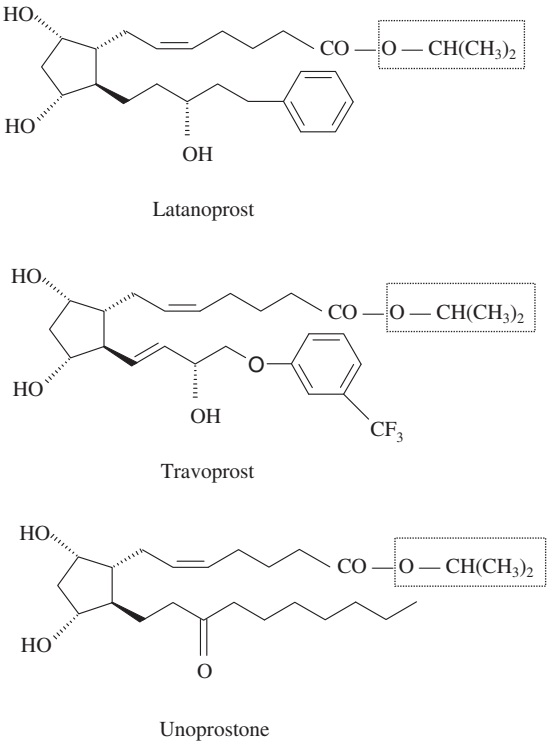


**Figure 21** Chemical structures of latanoprost, travoprost, and unoprostone (promoieties are marked by the rectangle).

intestinal tract. Drug absorption from the nasal cavity is predominately passive. Small hydrophilic drugs are absorbed via passive diffusion through aqueous channels, whereas lipophilic drugs are absorbed via the transcellular route. Drug molecules are rapidly absorbed from the nasal cavity, resulting in plasma profiles that are frequently comparable to those obtained after intravenous delivery. However, the absorption rates decrease with increasing molecular weight and fall off sharply when the weight exceeds 1000 daltons [96, 97].

*Water-Soluble Prodrugs* Enhancement of nasal drug delivery can be obtained by the formation of water-soluble derivatives of lipophilic drugs. Such derivatives will not only increase the amount of dissolved drug in the mucus, resulting in enhanced drug diffusion through the aqueous mucus barrier, but also allows the dose to be dissolved in a small volume (i.e., less than 200 μL) of an aqueous nasal vehicle. Testosterone, for example, undergoes extensive first-pass metabolism when given orally. The drug is completely absorbed in rats after nasal administration, but it is difficult to formulate a sufficient drug dosage (about 3 mg) for intranasal delivery in humans due to its

low aqueous solubility (0.01 mg/mL). Through the formation of a water-soluble prodrug ester, testosterone 17β-*N, N*-dimethylglycinate hydrochloride, the solubility was increased to over 100 mg/mL. The resulting bioavailability of testosterone after nasal administration of the prodrug to rats was reported to be nearly 100 percent [98].

The nasal administration of benzodiazepines is another example. Benzodiazepines are generally well, but slowly, absorbed from the gastrointestinal tract, reaching maximum serum concentrations about 2 h after oral administration. Benzodiazepines are used to treat panic attacks and seizures and as sedatives before anesthesia and surgery, where they are administered intravenously for rapid onset of drug action. As in the case of testosterone, the nasal administration of benzodiazepines is hampered by their low aqueous solubility. It is possible to increase their solubility by opening the diazepine ring of the benzodiazepine molecule; that is, by the formation of a prodrug. For benzodiazepines, such as alprazolam, midazolam, and triazolam, this ring opening is completely reversible (Fig. 22). In human serum in vitro at 37°C, the half-life for the first-order rate constant for the ring-closing reaction was estimated to be less than 2 min for both alprazolam and midazolam [99]. An aqueous nasal formulation containing the midazolam prodrug had an absolute bioavailability of 64 to 73 percent [99, 100].

*Lipophilic Prodrugs* Lipophilic prodrugs are thought to be absorbed from the nasal cavity via the transcellular route and, thus, their absorption is strongly affected by their lipophilicity. Acyclovir is a hydrophilic antiviral agent that does not readily permeate biological membranes [101]. The nasal absorption of acyclovir was enhanced significantly by forming lipophilic prodrug esters of the parent drug, but the results indicate that enzymatic hydrolysis of the esters prior to absorption (i.e., by the enzymatic barrier) reduced the amount of drug absorbed (Table 2) [102, 103]. Further studies

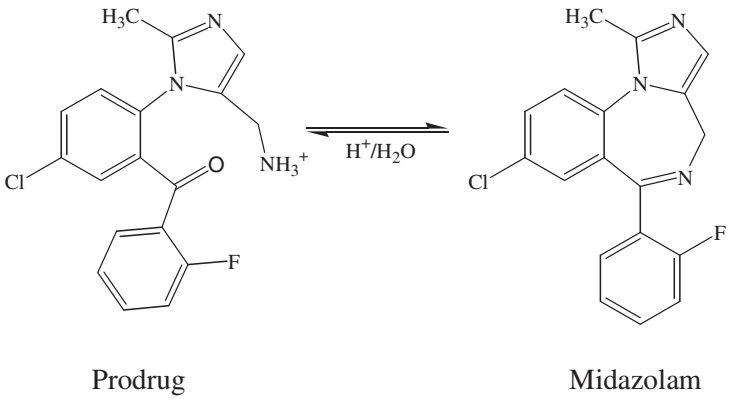


**Figure 22** The prodrug–midazolam equilibrium.

**TABLE 2 Structure of Acyclovir and Acyclovir Ester Prodrugs, Their Aqueous Solubilities, Partition Coefficients, Half-lives in Rat Plasma, and Percentages of Absorption in a Rat Nasal Perfusion Model [102, 103]**

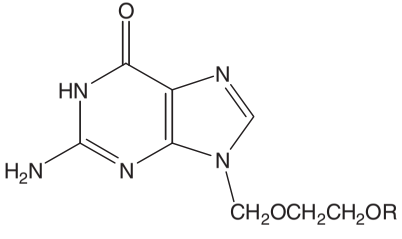


| R = | Solubility[a] (mM) | PC[b] | $t_{1/2}$ (min) | Percent Uptake in 90 min |
|---|---|---|---|---|
| H | 11.2 | 0.06 | — | 0.0 |
| $-C(=O)-(CH_2)_2CH_3$ | 4.6 | 0.83 | 25.6 | 9.7 |
| $-C(=O)-C(CH_3)_3$ | 1.5 | 2.01 | 105.0 | 14.8 |
| $-C(=O)-(CH_2)_3CH_3$ | 1.6 | 2.35 | 16.2 | 17.2 |
| $-C(=O)-(CH_2)_4CH_3$ | 0.7 | 8.58 | 3.6 | 30.1 |

[a] Solubility in pH 7.4 isotonic phosphate buffer at 37°C.
[b] Partition coefficient between 1-octanol and pH 7.4 isotonic phosphate buffer at 37°C.

in rat nasal mucosal homogenate showed that the rate and extent of prodrug degradation assumes a direct relationship with the acyl side-chain length. Lengthening of the linear acyl chain from three to eight carbons resulted in more than 600-fold increase in first-order degradation rate constants. Esters with branched acyl chains were more resistant to nasal carboxylesterase-mediated cleavage, and, thus, the branched esters not only increased the lipophilicity of acyclovir but also minimized mucosal hydrolysis, resulting in better drug absorption from the nasal cavity.

*Enhanced Brain Delivery* The anti-Parkinson drug levodopa is a zwitterion and, thus, both hydrophilic and water insoluble. Both lipophilicity and water solubility of the drug was increased through the formation of alkyl prodrug esters. Nasal administration of the prodrugs resulted in rapid and complete absorption into the systemic circulation and improved dopamine delivery to the central nervous system (CNS) when compared to intravenous administration of the prodrug. These results indicate that the prodrugs were delivered from the nasal cavity to the cerebral spinal fluid via the olfactory epithelium [104, 105].

*Transport-Mediated Absorption* Some amino acids have been shown to be absorbed via active transport from the nasal cavity and can be used as transport moieties for the active absorption of drugs [106]. For example, studies have indicated that an L-aspartate β-ester prodrug of acyclovir improved the nasal absorption of acyclovir by an active transport system [107].

*Prodrugs of Peptides* Both peptides and protein drugs are susceptible to rapid enzymatic degradation. For example, the aminopeptidase-catalyzed hydrolysis of Leu-enkephalin and Met-enkephalin is responsible for the rapid deactivation of these peptides at various mucosal sites. Formation of 4-imidazolidinone derivatives of enkephalins almost totally prevented aminopeptidase cleavage of the drugs. These derivatives were readily converted to the parent peptides by spontaneous hydrolysis [108]. In has been suggested that such prodrugs can enhance the bioavailability of peptides after nasal administration [109].

## Improved Aqueous Solubility

The efficient aqueous solubility of a drug molecule is a prerequisite for the preparation of aqueous-based solutions for drug delivery. Moreover, aqueous solubility is also one of the key physico-chemical properties in the oral drug absorption process. Two prodrug approaches have been introduced to enhance the aqueous solubility of a poorly soluble drug: (1) decreasing the parent drug's melting point by prodrug derivatization and/or (2) introducing an ionizable/polar promoiety to the parent drug, with the latter being the more common approach [15]. The most commonly utilized ionizable and, therefore, solubilizing prodrug structures are presented in Table 3.

**TABLE 3 Examples of Water-Soluble Prodrug Structures for Poorly Water-Soluble Drugs**

| Prodrug Derivative | Molecular Structure |
|---|---|
| Phosphate esters | Drug—O—$PO_3^{-2}$ |
| α-Amino acid esters and amides (from alcohols X = O; from amines X = NH; R = alkyl/aryl group) | 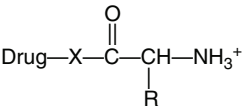 |
| Amino acid amides (from carboxylic acids; R = alkyl/aryl group) | 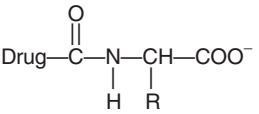 |
| Hemiesters (typically $n$ = 2–4) | 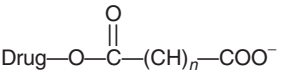 |
| Aminoalkyl esters (R = alkane group; $R_1$ and $R_2$ = H and/or alkane group, or heterocycle) | 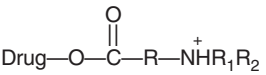 |
| Trialkylammonium (quaternary ammonium alkyl) esters (R = alkane group; $R_1$, $R_2$, and $R_3$ = alkyl groups) | 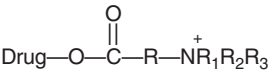 |
| Glucose derivatives (as an example glucuronide) | 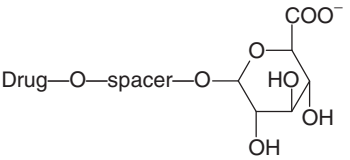 |
| Sulfate esters | Drug—O—$SO_3^-$ |

***Phosphates*** A number of phosphoric acid esters have been developed as potential water-soluble prodrugs. Many are marketed as injectable dosage forms while only a few are used in oral dosage forms [67]. The phosphate promoiety can be introduced either directly or via various spacer groups to the parent drug molecule that contains a hydroxyl or various amino functional groups, where it is enzymatically cleaved with relative ease by endogenous alkaline phosphatases.

*Etoposide Phosphate* A representative example of a phosphate prodrug is etoposide phosphate, which is currently in clinical use. Due to its poor aqueous

solubility and chemical instability, etoposide requires a complex IV formulation and must be diluted to a concentration of less than 0.4 mg/mL to avoid precipitation. Monophosphorylation of the phenolic group of etoposide increases the aqueous solubility up to 20 mg/mL and, furthermore, improves its chemical stability. After IV administration to patients, rapid and quantitative bioconversion of etoposide phosphate to etoposide was observed. The efficacy of the phosphate prodrug also proved to be bioequivalent to etoposide, with analogous therapeutic effects [110].

*Fosphenytoin* Phenytoin is a sparingly water-soluble (20 to 25 μg/mL) weakly acidic ($pK_a = 8.3$) drug. In comparison, fosphenytoin [(phenytoin-3-yl)methyl phosphate] is a phosphonooxymethyl prodrug of phenytoin that has high aqueous solubility (140 mg/mL), which also provides good chemical stability within neutral and slightly basic pH ranges. The bioconversion process involves a phosphatase-catalyzed dephosphorylation, followed by spontaneous decomposition of the hydroxymethyl intermediate, eventually releasing the parent drug and formaldehyde (Fig. 23). In epileptic patients, both intra-

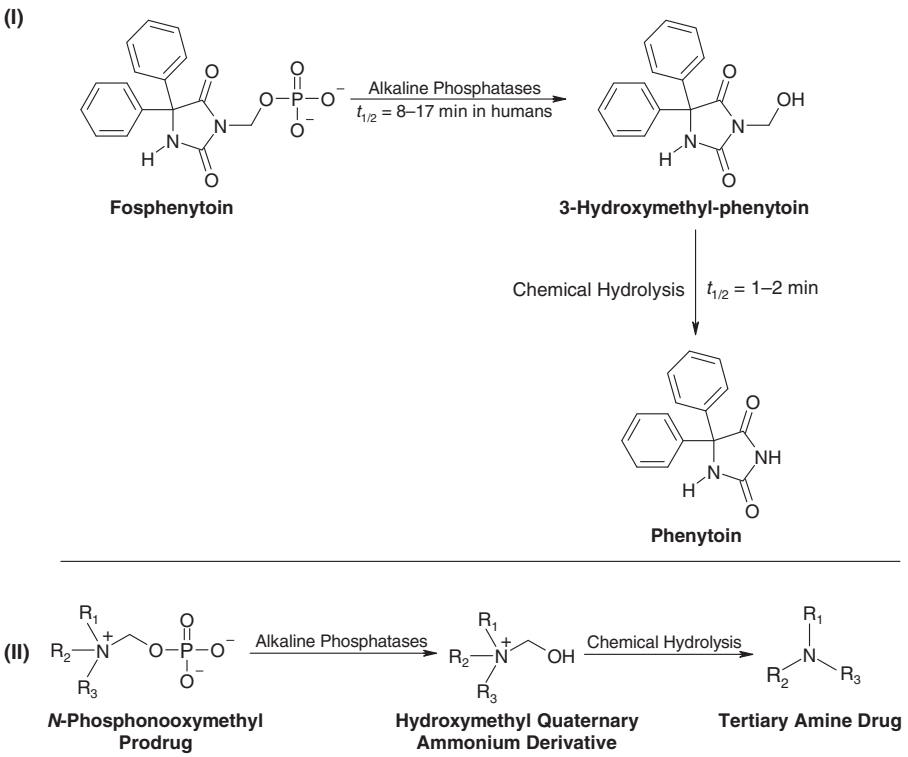


**Figure 23** Bioconversion of (I) fosphenytoin and (II) *N*-phosphonooxymethyl prodrugs.

muscular and intravenous administrations of fosphenytoin led to its rapid and quantitatively bioconversion to phenytoin with half-lives ranging from 8 to 17 min. Moreover, the blood levels of phenytoin following intramuscular administration of fosphenytoin were far superior to those achieved from phenytoin sodium alone. The pharmacokinetics, pharmacodynamics, and safety properties of fosphenytoin have been thoroughly reviewed by Stella [6]. The same prodrug strategy has been extended to various drugs containing a tertiary amino group, which were then derivatized to *N*-phosphonooxymethyl prodrugs [111–113]. Again, highly water-soluble prodrugs were substrates for alkaline phosphatase that cleaved the phosphoric acid monoester bond, followed by spontaneous in vivo chemical hydrolysis to the tertiary amine and formaldehyde (Fig. 23).

***Other Prodrug Structures*** A number of hemiester prodrugs (mainly hemisuccinate esters) have been studied as water-soluble prodrugs [114–116]. These derivatives usually improve aqueous solubility of the parent drug, but the limited amount of published information indicates that hemiester prodrugs generally possess limited chemical stability in aqueous solutions and give a slow and incomplete hydrolysis to the parent drug in human plasma. In contrast, their bioconversion appears relatively fast in the presence of hepatic hydrolases.

Amino acid prodrugs have been evaluated as water-soluble derivatives of various alcohols [42, 117] and amines [118, 119], as the corresponding ester or amide prodrugs. In addition to good aqueous solubility, the use of amino acids as a promoiety provides rapid and quantitative bioconversion by ubiquitous esterases and/or peptidases. While amino acid ester prodrugs show usually relatively rapid chemical hydrolysis, which precludes the development of solution formulations, amino acid amide prodrugs, in general, exhibit greater chemical stability while still maintaining enzymatic susceptibility.

## Prolonged Duration of Action

In general, prodrugs have been designed to prolong the duration of a particular drug action by two means: (1) inactive prodrug is delivered into the body at a sustained rate, where it rapidly releases the parent drug, or (2) after administration the prodrug releases the parent drug over a prolonged and relatively well-defined period of time. Primary goals for the prolongation of drug action can be an improvement in patient compliance by eliminating the need for frequent dosing, and by reducing side effects through a steady drug concentration.

The slow release for very lipophilic prodrug derivatives of several steroids (e.g., testosterone, nandrolone) and neuroleptics (e.g., fluphenazine, flupenthixol, haloperidol) from the site of intramuscular injection results in a prolonged duration of action. Once released into the tissue fluid or blood, prodrugs are rapidly bioconverted in most cases with no attenuation of their

therapeutic action. In the case of the neuroleptic fluphenazine (Fig. 24), the onset of action generally appears between 24 to 72 h after injection, and then continues for 1 to 8 weeks with an average duration of 3 to 4 weeks [120].

The prodrug approach has also been evaluated to obtain sustained drug action as a means of prolonging release of the parent drug. Longer duration of drug action has been achieved, for example, with a prodrug of insulin [121], which prolongs a normoglucemic profile in diabetic rats, due to a slow bioconversion of insulin from (9-fluorenylmethoxycarbonyl-$SO_3H)_3$-derivatized insulin, and with a bisdimethylcarbamate prodrug (bambuterol) of the bronchodilator terbutaline [122, 123], which after oral administration is predominantly bioconverted to terbutaline outside the lungs via a cascade of hydrolysis and oxidation reactions (Fig. 25). Moreover, prodrugs of various antiglaucoma drugs such as pilocarpine, timolol, and epinephrine have a slow bioconversion to their parent drugs, achieving a longer duration of action that is, however, closely related to improved ocular permeation. A prodrug of epinephrine was described above.

## Improved Drug Targeting

Targeted or site-specific drug delivery is one of the ultimate goals in controlled drug delivery. In prodrug design, targeting can be achieved by site-directed drug delivery or site-specific drug bioactivation (Table 4). With site-directed drug delivery, such as localized drug delivery (e.g., dermal and ocular drug delivery), the bioreversible prodrug is selectively or primarily transported, as an intact prodrug, to the site of drug action. On the other hand, with site-specific bioactivation, the prodrug can be widely distributed all over the body but undergoes bioactivation and exerts activity only at the desired site. The most common prodrug applications in targeted drug delivery are described as follows. In many prodrug applications, either site-directed drug delivery or site-specific bioactivation has been combined in order to achieve successful targeted drug delivery.

***Tumor Targeting*** The physiological barriers of solid tumors, which are aggregated masses of cells caused by changes or mutation in genes that control

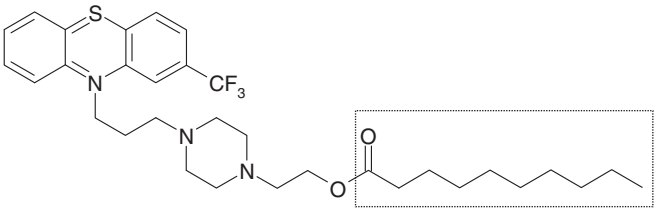


**Figure 24** Fluphenazine decanoate (lipophilic promoiety marked by the rectangle).

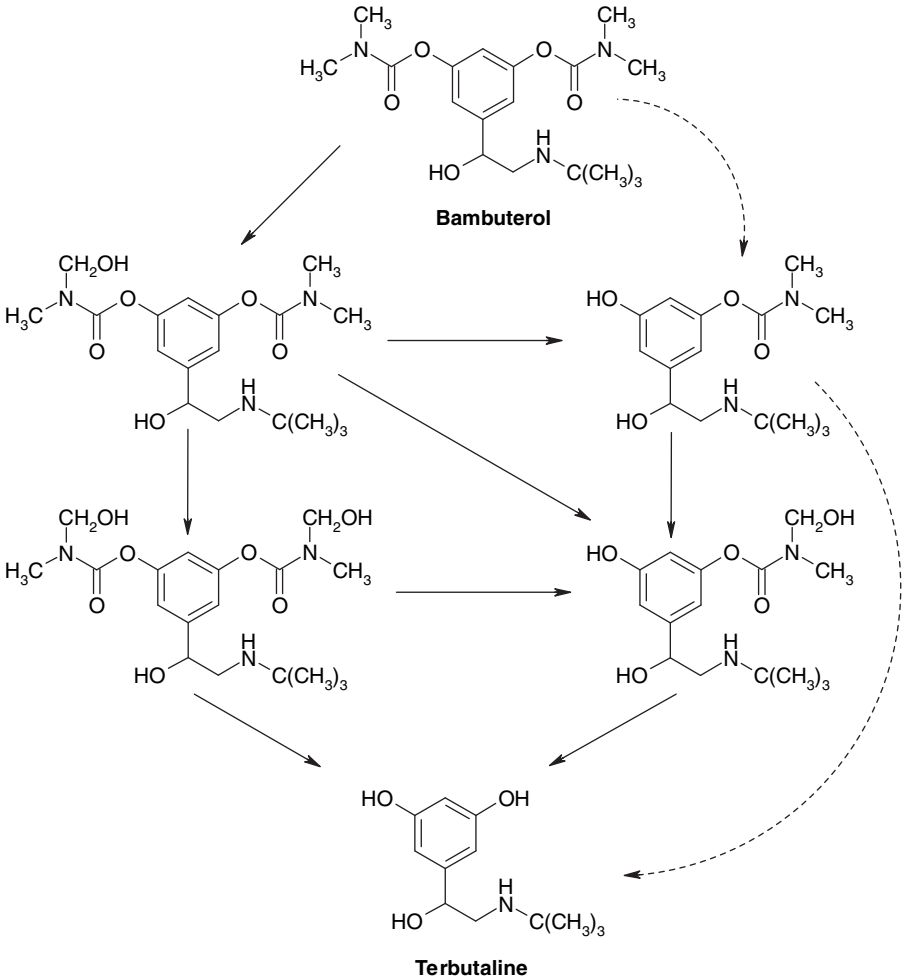


**Figure 25** Bioconversion of bambuterol to terbutaline via hydrolytic (broken arrows) and oxidative (solid arrows) pathways.

healthy cell division, considerably restrict cancer chemotherapy. Compared with the normal, ordered vasculature of healthy tissues, blood vessels in tumors are highly abnormal and have distended capillaries with leaky walls and sluggish flow, leading to inconsistent drug delivery. Furthermore, the increased intratumoral interstitial pressure limits the penetration of drugs into tumors. Highly irregular blood flow results in localized hypoxia, which subsequently leads to tumor resistance to certain anticancer agents and to radiotherapy as well. However, increased bioreductive enzyme expression is an adaptive strategy that solid tumors use to detoxify anticancer drugs [124–126], and the hypoxic environment further contributes to the increased reductase activity of

**TABLE 4 Prodrug Approaches to Improve Drug Targeting**

| Site-Directed Drug Delivery | Site-Specific Bioactivation |
|---|---|
| Prodrug is selectively or primarily transported to the site of drug action:<br>• Localized drug delivery (e.g., dermal and ocular drug delivery)<br>• Prodrug improves drug delivery to the specific tissue (e.g., improved CNS drug delivery) | Prodrug is widely distributed throughout the body:<br>• Bioactivation takes place only at the desired site of drug action |

tumors. Due to the high proliferation rates of tumor cells, in addition to bioreductive activity, the levels of other enzymes such as acid phosphatase, uridine phosphorylase, β-glucuronidase, and γ-glutamyl transpeptidase are often elevated in these cells and have been exploited in targeted prodrug–tumor delivery [124]. Recent prodrug approaches have been actively applied to achieve very precise and direct effects at the tumor site of action, with minimal effects to the rest of the body. Therefore, an effective prodrug must be selectively cleaved to the active drug by specific activating enzymes at the site of action. For tumor drug delivery, the first studied prodrug-activating enzymes were endogenous to specific tumor types (e.g., sarcoma 180, NK lymphoma, Walker carcinoma 256, Novikoff hepatoma), such as β-glucuronidase and aryl sulfatase, and a NAD(P)H dehydrogenase isozyme that sensitized tumors to aniline mustard and 5-aziridinyl-2,4-dinitrobenzamide, respectively. To expand the range of tumors susceptible to this mode of therapy, prodrug-activating enzymes were initially targeted to tumor cells by using antibodies and more recently genes. These approaches are referred to as ADEPT (antibody-directed enzyme prodrug therapy) and GDEPT (gene-directed enzyme prodrug therapy), VDEPT (virus-directed enzyme prodrug therapy), or GPAT (gene–prodrug activation therapy), respectively. The latter approach is also known as "suicide gene therapy." Some specific examples of hypoxia-selective prodrugs, and both ADEPT and GDEPT technologies in targeted tumor delivery are described in the following section.

*Hypoxia-Selective Prodrugs* Despite highly irregular blood flow to the solid tumor, with the development of oxygen-depleted areas (hypoxia) that result in poor drug delivery, hypoxic cells also present a major therapeutic opportunity for tumor targeting via bioreductive prodrugs. The substantial selectivity of hypoxia-selective drugs depends upon the presence of overexpressed reductase enzymes in tumor cells that reduce bioreductive prodrugs to active cytotoxic radicals under hypoxia. In normal cells, under aerobic conditions, the reduced radical is oxidized to the nontoxic prodrug with the concomitant production of a superoxide radical (Fig. 26). There are three main classes of hypoxia-selective prodrugs that are in clinical use or are being developed for

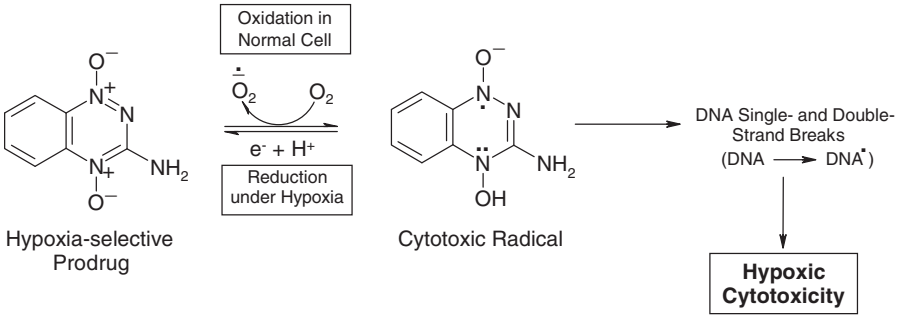


**Figure 26** Schematic presentation of bioreduction for a hypoxia-selective prodrug (tirapazamine, TPZ) to its cytotoxic free radical, causing preferential toxicity to hypoxic cells.

clinical use: quinone-based compounds, nitroimidazoles, and *N*-oxides [124–127].

Among the quinones of clinical interest that act as hypoxia-selective prodrugs are mitomycin C, its *N*-methyl derivative porfiromycin, and aziridinylquinone EO-9. All are structurally similar and are converted upon reductive metabolism to semiquinones and hydroquinones, which probably alkylate nucleophiles [e.g., DNA (deoxyribonucleic acid)] through covalent binding. Although EO-9 demonstrated excellent antitumor activity in hypoxic tumor cells and in a solid tumor animal model, it failed in clinical trials due to its short half-life (rapid elimination) and poor distribution. Moreover, mitomycin C has demonstrated only modest selectivity toward hypoxic cells and its clinical utility is also limited because of severe side effects. Despite promising preclinical data, and an acceptable toxicity profile, porfiromycin is considered to be inferior to mitomycin C.

The most extensively studied compounds in a second class of hypoxia-selective prodrugs are 2-nitroimidazoles, of which RSU1069 has been shown to be highly efficient both in vitro and in vivo. RSU1069 containing an aziridinyl alkylating group in the side chain (dual function) showed increased hypoxia selectivity compared to its nonalkylating nitroimidazole analogs and was bioconverted into a bifunctional alkylating drug by reduction. In further drug development, its bromoethyl prodrug RB6145 was shown to form RSU1069 rapidly under physiological conditions and have less systemic toxicity and improved pharmacokinetic properties compared to RSU1069. Clinical trials of RB6145 were, however, aborted due to irreversible cytotoxicity toward retinal cells.

The *N*-oxide, tirapazamide (TPZ), is the first drug introduced into clinical use as a selective hypoxia-activated prodrug. TPZ undergoes 1-electron bioreduction to monodeoxygenated toxic products (Fig. 26) in the presence of cytochrome P450, NADPH : cytochrome P450 oxidoreductase, xanthine oxidase, and aldehyde dehydrogenase. Under hypoxia, the formed oxidizing

radical leads to DNA single- and double-strand breaks, which kills the tumor cell.

*Antibody-Directed Enzyme Prodrug Therapy (ADEDT)* In the ADEPT design (Fig. 27), prodrug-activating enzymes are targeted toward human tumor xenografts by conjugating them to tumor-selective monoclonal antibodies. An antitumor antibody is covalently conjugated to an enzyme not normally present in the extracellular fluid or on cell membranes (e.g., bacterial or viral enzymes), and these conjugates are then localized in the tumor via systemic administration. After allowing for the conjugate to clear from the blood, a prodrug is administered. The activation of a prodrug to form a chemotherapeutic drug then takes place by reacting with the pretargeted enzyme, which was delivered onto the tumor surface. ADEPT has progressed into several phase I clinical trials (e.g., CMDA prodrug with a murine monoclonal antibody linked to the bacterial enzyme carboxypeptidase) and has shown evidence of tumor response. Although giving high tumor to blood ratios of antibody–enzyme conjugate, the targeting systems require the administration of a clearing antibody in addition to the antibody–enzyme conjugate. Moreover, the paucity of tumor-specific antigens has limited the applicability of ADEPT. This limitation has in turn led to the development of GDEPT, in which vectors are used to deliver suicide genes that encode specific prodrug-activating enzymes for tumor cells. The ADEPT approach has been reviewed recently [128, 129].

*Gene-Directed Enzyme Prodrug Therapy (GDEPT)* In GDEPT (Fig. 27), an inactive prodrug can be activated to release a cytotoxic drug by an enzyme that has been delivered via a gene to the tumor for expression. GPAT is a

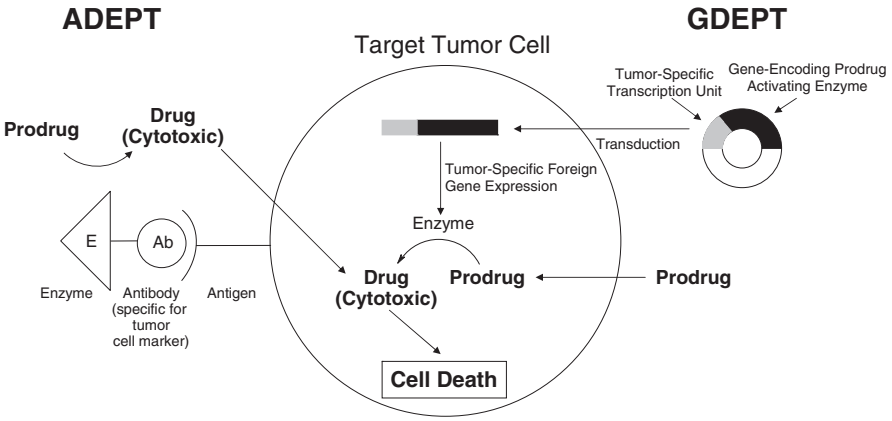


**Figure 27** Outline of antibody-directed enzyme prodrug therapy (ADEPT) and gene-directed enzyme prodrug therapy (GDEPT).

variation of GDEPT, which uses known transcriptional differences between normal and tumor cells to achieve the selective expression of a prodrug-metabolizing enzyme [128]. The enzymes used are from nonmammalian origin, such as viral tyrosine kinase, bacterial cytosine deaminase, carboxypeptidase G2, purine nucleoside phosphorylase, and nitroreductase or enzymes of human origin that are absent or are expressed at only low concentrations in tumor cells, such as deoxycytidine kinase and cytochrome P450. A number of GDEPT approaches have been described in the past decade [130, 131]. A prototypical example of this approach is centered on inserting a thymine kinase gene into the herpes simplex virus (HSV-TK), followed by combination treatment with the prodrug ganciclovir (GCV). Another widely studied example is the bacterial gene cytosine deaminase (CD), which sensitizes tumor cells to the antifungal drug, 5-fluorocytosine (5-FC). As a result 5-FC is bioconverted to an anticancer drug, 5-fluorouracil (5-FU). Both of these approaches have been tested in clinical trials; HSV-TK-GCV against, for example, mesothelioma, ovarian cancer, glioblastoma, breast cancer, melanoma, multiple myeloma, and astrocytoma, and CD-5-FC has been used against colon and breast carcinomas. In addition to these classical examples, many other enzyme–prodrug combinations have been explored for GDEPT [132, 133]. Despite promising in vitro and in vivo results, the antitumor effect in clinical trials has not proved promising, primarily due to difficulty in achieving selective gene delivery to a sufficient number of tumor cells.

***Brain Targeting*** According to prevailing consensus only small, nonionic, and lipid-soluble molecules can diffuse across the blood–brain barrier (BBB) from the systemic circulation, whereas larger, more water-soluble, ionic molecules do not readily cross the BBB. There are exceptions, however, and some large, polar molecules can be transported into the brain via endogenous transporters. Therefore, the strategies used to improve drug delivery and targeting to the brain utilize passive drug uptake for lipophilic prodrugs and also more sophisticated approaches, such as chemical drug delivery system and carrier-mediated transport.

*Lipophilic Prodrugs* A series of aliphatic and aromatic lipophilic ester prodrugs of the anticancer agent chlorambucil have been developed to increase efficacy in the treatment of brain tumors [134, 135]. While short-chain aliphatic ester (e.g., methyl and propyl) and the aromatic esters (e.g., phenylmethyl and phenylethyl) were hydrolyzed too rapidly in rat and human plasma in vitro, long-chain aliphatic esters (hexyl and octyl) had more favorable bioconversion rates. However, as lipophilic prodrugs, they were highly bound to proteins and did not readily enter the brain. The sterically more hindered chlorambucil-tertiary butyl ester was subsequently developed in an attempt to maintain the desired slower bioconversion rate while having reduced lipophilicity and protein binding. Significant amounts of chlorambucil-tertiary butyl ester did enter the rat brain, with a 35-fold greater increase compared to an equimolar

dose of chlorambucil [135]. However, it exhibited poor bioconversion to chlorambucil in brain tissue and despite an enhanced brain delivery ratio, chlorambucil-tertiary butyl ester did not demonstrate anticancer activity superior to the equimolar administration of chlorambucil. Thus, increased lipophilicity over the parent drug alone does not ensure improved drug efficacy. While enhanced lipophilicity may improve permeation across the BBB, it also tends to increase both uptake into other tissues, causing an increased tissue burden and interaction with plasma proteins. Moreover, both bioconversion selectivity (serum vs. brain) and rate of bioconversion in the targeted tissue should be taken into account when designing bioreversible prodrugs for brain targeting.

*Chemical Drug Delivery System (CDS)* The principle of CDS, in addition to providing access to the brain by increasing the lipophilicity of a drug, exploits specific properties of the BBB to lock drugs in the brain on arrival and prevent them from recrossing the BBB. Linking an active drug molecule to a bioremovable lipophilic targetor moiety, 1,4-dihydro-*N*-methylnicotinic acid (dihydrotrigonelline), for example, results in a derivative that readily distributes throughout the body and brain after administration due to its lipophilic character. Once inside the brain parenchyma, the lipophilic dihydrotrigonelline is enzymatically oxidized to the ionic quaternary salt via the ubiquitous $NAD(P)H \leftrightarrow NAD(P)^+$ coenzyme system (Fig. 28). The acquisition of charge has a dual effect: accelerating the rate of systemic elimination of this derivative and capturing the ionic drug-targetor inside the brain. Subsequently, slow release of the drug from the targetor can result in a sustained and brain-specific release of free active drug, and the targetor is readily removed from the brain by active processes. The CDS has been explored with a wide variety of hydroxy- and amino-containing drugs [136, 137], and considerably increased brain targeting has been achieved, for example, zidovudine (AZT), ganciclovir, benzylbenicillin, and estradiol.

Among all CDSs, estradiol-CDS is in the most advanced stage of investigation [137]. Estradiol is a lipophilic drug with an octanol/water partition coefficient ($\log P$) of 3.3, and derivatization with 1,4-dihydrotrigonelline as a targetor further increases lipophilicity ($\log P = 4.5$), thus enabling better transport across the BBB. Oxidation of the targetor moiety leads to a more ionic and less lipophilic form, with a $\log P$ value of only –0.14. A study in rats found that the concentration of estradiol in rat brain was elevated four to five times longer after estradiol-CDS administration than after estradiol treatment. Moreover, clinical evaluations suggested a potent central effect with only slight elevations in systemic estrogen levels.

*Carrier-Mediated Transport* Carrier-mediated drug delivery takes advantage of facilitative endogenous transport systems that are present in brain endothelial cells. A number of carrier transport systems in the BBB are responsible for brain uptake of nutrients and their analogs, such as glucose, amino acids,

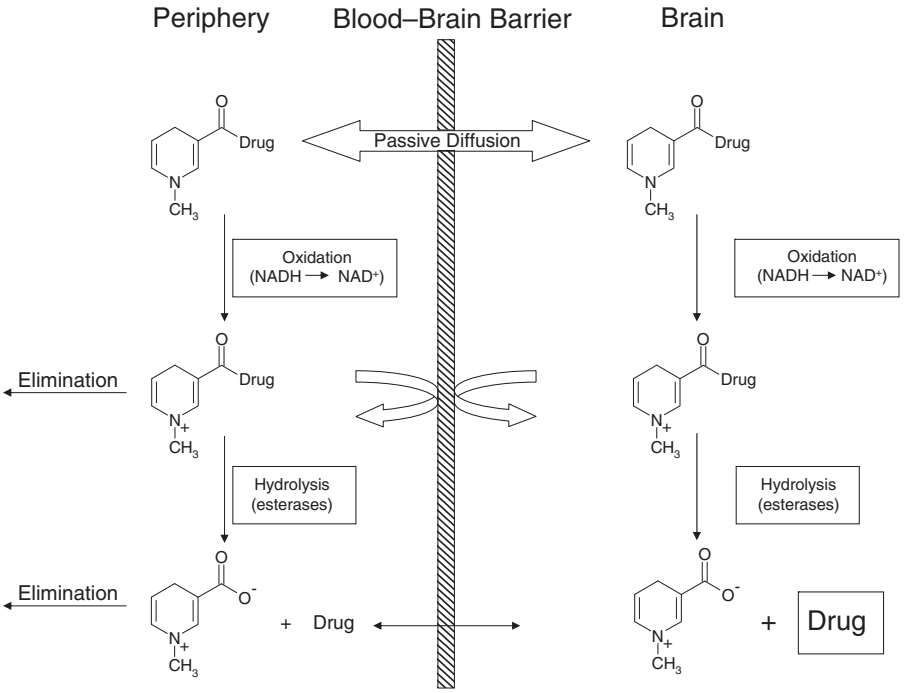


**Figure 28** Principle of the chemical drug delivery system (CDS).

choline, vitamins, low-density lipoprotein (LDL), and nucleosides in the systemic circulation [138, 139]. Of the nutrient transport systems present at the BBB, the carriers for glucose and large neutral amino acids (LAT) have been estimated to have a sufficiently high transport capacity that hold promise for significant drug delivery to the brain. From these two systems, the glucose carrier is very restrictive in its requirements, and only those molecules closely resembling D-glucose (e.g., D-mannose, D-galactose) are transported. By contrast, the large neutral amino acids transport (LAT) system has the highest maximum transport rate and is less specific for its substrates and, thus, is capable of transporting numerous endogenous and exogenous amino acids of great structural variety. A classic example of the LAT-mediated transport is the anti-Parkinson's agent dopamine. Dopamine itself is poorly transported across the BBB, while its precursor molecule L-dopa is transported into the brain by the LAT system. Once in brain, L-dopa is decarboxylated to dopamine. Other amino acid mimetic drugs in clinical use that enter the brain predominantly via LAT-mediated transport are α-methyldopa (a centrally acting antihypertensive) and gabapentin (an antiepileptic).

In an analogous way to LAT-mediated transport, the glucose transporter was employed to facilitate the penetration of glucose prodrugs of chlorambucil into the brain [140]. All glucose-conjugated prodrugs were able to inhibit

the transport of [$^{14}$C]glucose by the glucose transporter (GLUT1) isoform in a concentration-dependent manner and displayed an inhibition curve similar to that of the inhibition by glucose itself. Especially the 6-*O*-glucose ester derivative of chlorambucil showed activity at a 160-fold lower concentration than chlorambucil itself. Similarly, glucose and galactose derivatives of the anticolvulsant agent nipecotic acid [141], and the potential neuroprotective agent 7-chlorokynurenic acid [142], have been explored as a possible means of improving brain delivery of the active parent drug by providing suitable substrates for active membrane transporters. Moreover, ascorbic acid derivatives of nipecotic acid were shown to competitively inhibit the ascorbate transporter SVCT2, which is expressed in the choroid plexus [143]. In vivo studies with rats showed increased significance in latency for the appearance of pentylentetrazol-induced tonic convulsions than nipecotic acid, demonstrating better absorption compared to the nonconjugated parent drug.

***Colon Targeting*** To achieve successful colon targeting, a drug needs to be protected from absorption and degradation in the upper gastrointestinal tract and then released into the proximal colon. The best known prodrugs that target the colon are azo-bond derivatives of 5-aminosalicylic acid (5-ASA). These prodrugs are formed by attaching a hydrophilic promoiety to 5-ASA and are absorbed from the upper gastrointestinal tract to a very limited extent due to their polar nature, but they are converted to their constituent entities by azoreductases produced by anaerobic colonic bacteria. Carrier molecules that have been used include, for example, sulfapyridine in sulfasalazine, 4-amino benzoyl-β-alanine in balsalazine, and *p*-aminohippurate (4-amino benzoyl glycine) in ipsalazine (Fig. 29) [144, 145]. The first prodrug introduced for this purpose, sulfasalazine, causes side effects due to the formation of sulfapyridine, from the degradation of the prodrug. Balsalazine has been found to be therapeutically more effective in clinical trials with fewer side effects. However, in spite of promising pharmacokinetic results, ipsalazine has not developed further. Olsalazine, a dimer of two 5-ASA molecules linked by an azo-bond, has shown promising results in clinical trials (Fig. 29).

Similarly, hydrophilic glycoside and glucuronide prodrugs of drugs such as dexamethasone, prednisolone, budesonide, naloxone, and nalmefene, which release the active parent drug by bacterial glycosidases and glucuronidases, have been studied for their colon-targeting potential [144, 145]. In vivo studies in rats revealed that nearly 60 percent of an oral 21-β-D-glucosidic prodrug of dexamethasone reached the colon in the form of the free drug. In the case of prednisolone-β-D-glucoside, 15 percent of the given dose reached the colon. When unmodified steroids were administered orally, they were absorbed from the small intestine and less that 1 percent of dose reached the colon.

***Kidney Targeting*** Renal-specific drug targeting may be an attractive approach under the following conditions: when drugs reaching the kidney cause unwanted extrarenal effects, or the intrarenal transport of a drug is not optimal in relation to the target cell within the kidney, or when a drug is

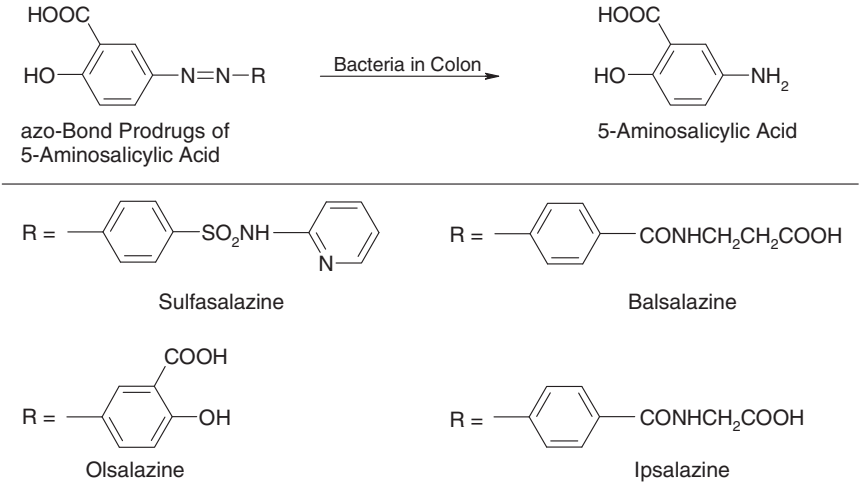


**Figure 29** Azo-bound prodrugs of 5-aminosalicylic acid (5-ASA) and their bioconversion.

inactivated before reaching the site of action in the kidneys, or when pathological conditions such as abnormalities in glomerular filtration, tubular secretion, or when the occurrence of proteinuria can affect the normal renal distribution of a drug.

Most research on targeted drug delivery to the kidney has focused on the development of amino acid prodrugs [146]. The strategy is based upon the relative high concentration of bioconverting enzymes in the proximal tubular cell, either cytosolic enzymes, such as L-amino acid decarboxylase, β-lyase, and *N*-acetyl transferase, or enzymes at the brush border of the proximal tubule, such as γ-glutamyl transpeptidase. These prodrugs are designed to be activated by one or two enzymes in order to release the active parent drug. For example, dopamine showed a high renal specificity following the intraperitoneal administration in mice of the double prodrug, γ-glutamyl-L-dopa [147]. This double prodrug first requires γ-glutamyl transpeptidase-catalyzed cleavage of the γ-glutamyl linkage, and consequently the resulting L-dopa is decarboxylated to dopamine by L-amino acid decarboxylase. Targeted delivery to the kidney has also been achieved with *N*-acetyl-γ-glutamyl prodrugs of sulfamethoxazole [148]. On the other hand, L-γ-glutamyl-sulfamethoxazole, which requires only γ-glutamyl transpeptidase for its activation, did not show renal selectivity [149]. Up to now, results of the prodrug approach in renal targeting are still rather disappointing and without a clear therapeutic impact.

## Improved Formulation and Delivery of Peptides

Bioactive peptides have become an important class of drugs with the recent advantages in biotechnology, peptide synthesis, and modern screening tech-

nologies. Although numerous bioactive peptides and proteins have been identified, they are usually limited to parenteral use, due to unfavorable biopharmaceutical properties. The hydrolysis of peptides is catalyzed by a variety of digestive and mucosal peptidases and proteases with broad substrate specificities. This enzymatic activity is a significant barrier for the absorption of peptides, and they are often extensively degraded before and after entering the systemic circulation. Therapeutic peptides tend to be relatively hydrophilic, which limits their penetration through the cellular membrane. Thus, they are not absorbed via the transcellular route, but primarily via a less efficient paracellular route or by cellular endocytosis. Efflux system activity also limits the absorption of peptides and peptidomimetics. The oral bioavailability of peptides is, therefore, typically less than 1 to 2 percent [150]. Systematic clearance is also rapid and the in vivo half-life is generally less than 30 min [150, 151].

Limited stability of peptides can often be addressed by designing peptidomimetic drugs with modified peptide backbones [152]. However, the clinical development peptidomimetics is still hampered by inadequate bioavailability, due to limited absorption and low aqueous solubility.

The usual objectives for producing a prodrug from a peptide is to improve both enzymatic stability and membrane permeation properties. Prodrugs for improved solubility have also been investigated. The C-terminal carboxyl and N-terminal amino groups are commonly selected for modification, but the amino acid side chains and the peptide backbone are also targeted. Many innovative prodrugs have been reported, but the development of clinically successful peptide prodrugs still remains a challenge. These prodrugs must sufficiently resist enzymatic degradation and be relatively lipophilic to allow absorption, but at the same time they must be sufficiently labile to release the parent peptide prior to systematic clearance.

***Prodrugs to Improve Aqueous Solubility*** Prodrugs that increase the aqueous solubility of the immunosuppressive hydrophobic cyclic undecapeptide cyclosporin A have been developed, based on the *N,O*-acyl migration reaction [153] (Fig. 30). The *O*-acyl-peptide isomers can be formed under mild acidic conditions, but they will revert back to the normal *N*-acyl form under physiological conditions. The *O*-acyl isocyclosporin A is quantitatively converted back to cyclosporin A under simulated physiological conditions, but the half-life for this reaction is more than 4 h at pH 7.4 (37°C) [154], which is considered to be too slow as a useful peptide prodrug [153].

The aqueous solubilities of the *O*-acyl prodrugs of the HIV protease inhibitors KNI-720, KNI-272, and KNI-727 (Fig. 30) are more than 4000-fold higher than that of the parent drugs [155, 156]. These prodrugs are rapidly ($t_{1/2}$ < 1 min) and quantitatively converted back to the parent drugs under simulated physiological conditions (pH 7.4, 37°C).

***Prodrugs for Improved Enzymatic Stability*** Endopeptidases are important barriers against the absorption of peptides such as ACTH, oxytocin, various

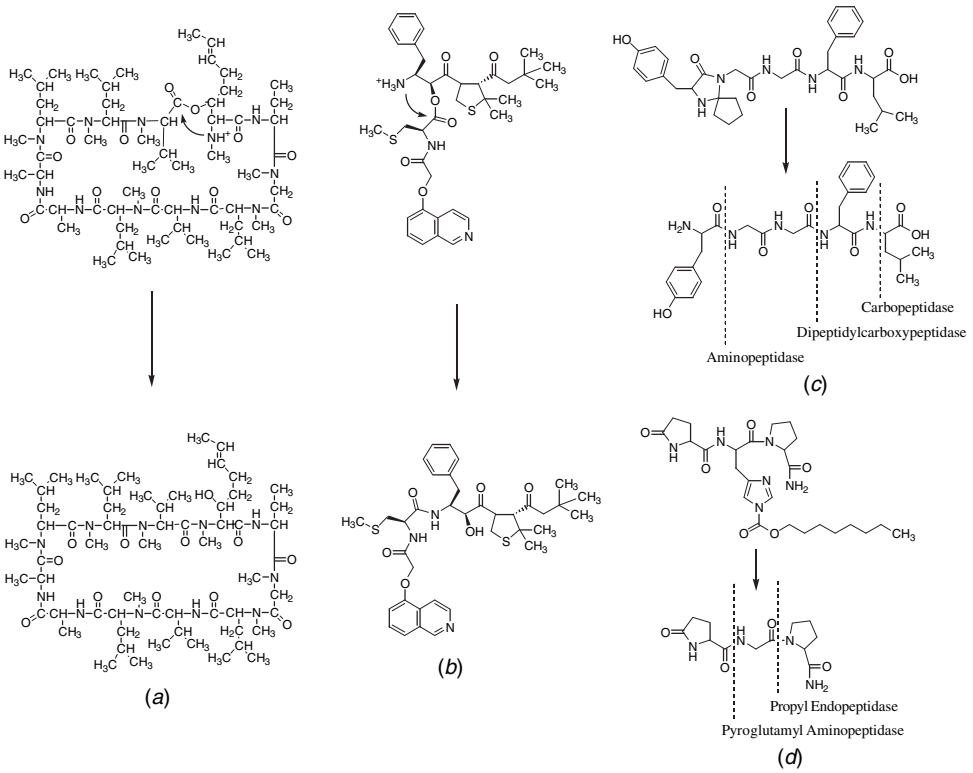


**Figure 30** Prodrugs of (*a*) cyclosporin A and (*b*) KNI-727 based on the *O*,*N*-acyl migration reaction, the (*c*) 4-imidazolidone prodrug of Leu-enkephaline and (*d*) *N*-octyloxycabonyl prodrug of TRH.

rennin inhibitors, and enkephalins [157]. Prodrugs derived from both C-terminal and N-terminal residues can protect a peptide from the degradation catalyzed by endopeptidases. For example, derivation of the C-terminal carboxylic group of the pseudopeptide rennin inhibitor SR42128 with a highly lipophilic 1,3-dipalmytoyl acetyl glycerol promoiety prevents α-chymotrypsin-catalyzed degradation [158].

An α-aminoamide moiety appears in a large number of peptides. These peptides can be condensed with aldehydes or ketones to form 4-imidazolidone prodrugs that will protect the parent peptide from an aminopeptidase-catalyzed degradation. These prodrugs are spontaneously hydrolyzed in aqueous solutions. The rate of hydrolysis depends on the structure of the parent peptide and the carbonyl component [157]. For example, the 4-imidazolidone derivatives of Leu-enkephalin that are formed with either acetaldehyde, acetone, or cyclopentanone (Fig. 30) decompose at pH 7.4 (37°C) with half-lives of 30, 10.9, and 3.1 h, respectively [157].

Side-chain modification has also been tried to protect against enzyme-catalyzed degradation. This approach has, for example, been used for prodrugs of the CNS-active thyrotropin-releasing hormone (TRH). Although this peptide has been suggested for the treatment of various conditions, its clinical utility is limited by a short plasma half-life (6 to 8 min), due to rapid enzymatic degradation, as well as limited transport across the blood–brain barrier [159].

Modifications on the imidazole group of the histidine residue of TRH to form *N*-alkoxycarbonyl prodrugs greatly increased stability toward TRH-specific pyroglutamyl aminopeptidases found in serum, and the half-life in plasma was also improved [157, 159]. Although the lipophilicity increased significantly with the octyloxycarbonyl prodrug (Fig. 30), this did not result in any improvement in penetration through rabbit and rat intestinal tissue or Caco-2 cell monolayers. This lack of penetration enhancement was explained as a result of enzymatic cleavage by carboxypeptidase and nonspecific esterase.

Prodrugs can also be obtained by forming an ester with hydroxyl groups on amino acid moieties. The *O*-pivaloyl ester of the hydroxyl group on tyrosine of the antidiuretic peptide desmopressin is an example of such a prodrug. This prodrug is more stable toward α-chymotrypsin-catalyzed degradation than the parent peptide [160].

***Prodrug for Increased Membrane Permeation*** Bioreversible cyclization of the peptide backbone is now one of the most promising approaches for peptide prodrugs with improved enzymatic stability and membrane permeability. The cyclization is achieved by linking the N-terminal amino group and the C-terminal carboxyl groups together through an enzyme-labile promoiety. Cyclic-acyloxyalkyl-based, phenylpropionic acid, or coumarine-based peptide prodrugs have been investigated (Fig. 31). These promoieties are cleaved in a slow esterase-catalyzed step that is followed by a rapid chemical reaction that releases the parent peptide [151]. Lacking terminal residues, these prodrugs are not substrates for endopeptidase. Lipophilicity is also increased, due to the lipophilic nature of the promoieties by masking the charge of terminal residues, due to the fact that the cyclic prodrugs exist in unique solution structures, some of which contain a high degree of intramolecular hydrogen bonding [161]. However, the effect of cyclization on permeation varies significantly with the structure of the peptide and promoiety. Lipophilic cyclic peptides are absorbed via the transcellular route, but the net flux is often limited by substrate properties for the polarized efflux systems (e. g., MRP and P-glycoprotein) [151]. For example, the cyclic phenylpropionic acid prodrugs of the opioid peptides Leu-enkephalin and DADLE exhibited 1680- and 77-fold increases in flux through Caco-2 cell monolayers, respectively, compared to those of the parent peptides. In contrast, very little permeation was observed with the cyclic acyloxyalkyl prodrugs of these peptides. This difference was mainly explained by the fact that only the acyloxyalkyl prodrugs were substrates for the efflux systems in Caco-2 cells [162, 163].

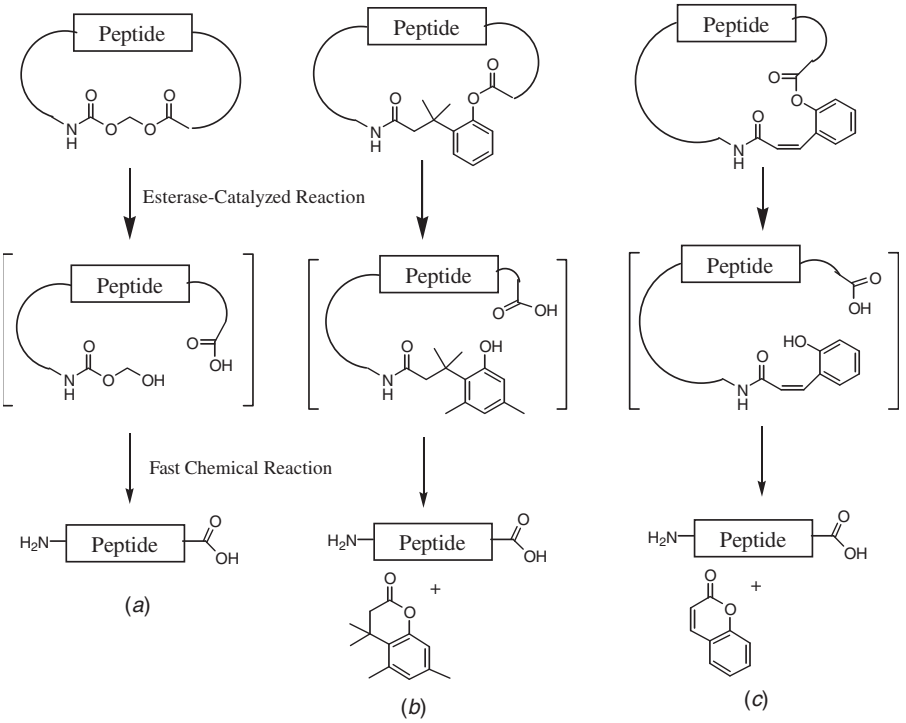


**Figure 31** The bioconversion of (*a*) cyclic acyloxyalkyl, (*b*) phenylpropionic acid, and (*c*) coumarine-based peptide prodrugs.

## Reduced Side Effects

Prodrugs have been utilized to minimize both local and systemic side effects. In general, improved drug targeting results in decreased side effects. Prolonged release of an active drug from a prodrug derivative also decreases side effects due to decreased plasma or specific tissue peak concentrations. In addition, the utilization of topical drug administration with prodrugs, instead of systemic administration, may decrease systemic side effects, or water-soluble prodrugs may allow IV administration without pain or irritation at an injection site. The pain or irritation at an injection site may be due to irritable formulation (low or high pH, co-solvents) or probable precipitation of the drug molecule at the injection site.

Fosphenytoin is an excellent example of how the prodrug approach can be used to develop an IV dosage form that decreases both side effects and local irritation [6]. Sodium phenytoin for IV and IM (intramuscular) administration was previously formulated as a very alkaline solution (pH near 12) that contained 10 percent ethanol and 40 percent propylene glycol. This formulation must be given slowly and minimally diluted to avoid precipitation and serious side effects. At least, the formulation only results in local tissue irritation.

Fosphenytoin, a water-soluble and bioreversible prodrug of phenytoin, increased the aqueous solubility of phenytoin by about 4000-fold at physiological pH and allows preparation of an aqueous and well-tolerated IV formulation.

Many topically applied ophthalmic drugs are absorbed into the systemic circulation, giving rise to possible systemic side effects. The general prodrug strategy is to improve the corneal absorption of a drug molecule by increasing lipophilicity, thus reducing the installed dose, systemic absorption, and side effects. For example, lipophilic prodrugs of timolol [164] and epinephrine [89] have been developed to decrease systemic side effects via improved corneal absorption.

The GI irritation resulting from NSAIDs is thought to be due to two different mechanisms: a direct contact of acidic drug on the GI mucosal and an inhibition of prostaglandin biosynthesis. Recent studies have confirmed that this GI side effect from NSAIDs are mainly a result of inhibiting COX-1, which regulates prostaglandin synthesis for normal cell activity. Thus, alternative formulations, administration routes, and prodrugs should not be able to eliminate all these side effects. However, direct contact of the NSAIDs has been shown to produce GI lesions, and, thus, the temporary masking of the carboxylic moiety has been proposed as a possible method to reduce or even abolish GI side effects [165, 166]. Thus, most oral prodrugs of NSAIDs have been developed to overcome GI-tract irritation, such as ibuprofen [167], indomethacin [168], ketoprofen [169], and naproxen [170].

## Macromolecular Prodrugs

A variety of biological and synthetic macromolecules have been conjugated either directly or with various spacer groups to drug molecules by means of a biodegradable bond to make either macromolecular prodrugs or polymeric

**TABLE 5 Macromolecular Promoieties for Prodrug Design**

| Macromolecular Promoiety | Reference |
|---|---|
| Proteins | |
| albumin | 172 |
| antibodies | 129, 172 |
| Polysaccharides | |
| dextrans | 173 |
| Polyamino acids | |
| polyaspartamide | 174, 175 |
| polyglutamic acid | 176 |
| Miscellaneous | |
| *N*-(2-hydroxypropyl)methacrylamide (HPMA) | 177 |
| Styrene-maleic acid anhydride copolymer (SMA) | 178, 179 |
| Polyethylene glycol (PEG) | 180–182 |

prodrugs. The macromolecular prodrugs offer several advantages over other drug delivery methods such as: (1) improved drug targeting, (2) improved aqueous solubility, (3) extended duration of action, and (4) improved chemical stability. An increased understanding of the biological factors most relevant to macromolecular prodrugs has enabled the design of such therapeutics that are sufficiently interesting to justify clinical testing, and several macromolecular prodrugs have already entered clinical trials [171] as anticancer agents. Low-molecular-weight prodrugs are the primary subject of the present chapter and, therefore, a more detailed description of some of the most commonly utilized macromolecules as promoieties can be found in the references listed in Table 5.

## REFERENCES


1. Venkatesh, S., Lipper, R. A. (2000). Role of the development scientist in the compound lead selection and optimization. *J. Pharm. Sci.*, *89*, 145–154.
2. Alber, A. (1958). Chemical aspects of selective toxicity. *Nature*, *16*, 421–423.
3. Krise, J. P., Stella, V. J. (2003). Prodrugs: A year of renewed interest in an old concept. *AAPS Newsmag.*, *16*, 25.
4. Stella, V. J., Charman, W. N. A., Naringrekar, V. H. (1985). Prodrugs. Do they have advantages in clinical practice? *Drugs*, *29*, 455–473.
5. Leppänen, J., Huuskonen, J., Nevalainen, T., Gynther, J., Taipale, H., Järvinen, T. (2002). Design and synthesis of a novel L-dopa-entacapone codrugs. *J. Med. Chem., 45*, 1379–1382.
6. Stella, V. J. (1996). A case for prodrugs: Fosphenytoin. *Adv. Drug Delivery Rev.*, *19*, 311–330.
7. Shaw, J. P., Sueoko, C. M., Oliyai, R., Lee, W. A., Arimilli, M. N., Kim, C. U., Cundy, K. C. (1997). Metabolism and pharmacokinetics of novel oral prodrugs of 9-[(*R*)-2-(phosphonomethoxy)propyl]adenine (PMPA) in dogs. *Pharm. Res.*, *14*, 1824–1829.
8. Bundgaard, H., Falch, E., Larsen, C., Mosher, G. L., Mikkelson, T. J. (1986). Pilocarpine prodrugs II: Synthesis, stability, bioconversion, and physicochemical properties of sequentially labile pilocarpine acid diesters. *J. Pharm. Sci.*, *75*, 775–783.
9. Thorsteinsson, T., Loftsson, T., Masson, M. (2003). Soft antibacterial agents. *Curr. Med. Chem.*, *10*, 1129–1136.
10. Bodor, N., Buchwald, P. (1997). Drug targeting via retrometabolic approaches. *Pharmacol. Ther.*, *76*, 1–27.
11. Bodor, N., Buchwald, P. (2000). Soft drug design: General principals and recent applications. *Med. Res. Rev.*, *20*, 58–101.
12. Bodor, N. (1985). Prodrugs versus soft drugs. In H. Bundgaard (Ed.), *Design of Prodrugs*. Elsevier Science, Amsterdam, pp. 333–354.
13. Amidon, G. L., Lennernas, H., Shah, V. P., Crison, J. R. (1995). A theoretical basis for a bipharmaceutic classification: The correlation of in vitro drug product dissolution and in vivo bioavailability. *Pharm. Res.*, *12*, 413–420.

14. Bundgaard, H. (1985). Formation of prodrugs of amines, amides, ureides, and imides. *Methods Enzymol.*, *112*, 347–359.
15. Fleisher, D., Bong, R., Stewart, B. H. (1996). Improved oral drug delivery: Solubility limitations overcome by the use of prodrugs. *Adv. Drug Delivery Rev.*, *19*, 115–130.
16. Taylor, M. D. (1996). Improved passive oral drug delivery via prodrugs. *Adv. Drug Delivery Rev.*, *19*, 131–148.
17. Haeberlin, B., Rubas, W., Nolen, H. W., Friend, D. R. (1993). In-vitro evaluation of dexamethasone-beta-D-glucuronide for colon-specific drug-delivery. *Pharm. Res.*, *10*, 1553–1562.
18. Chourasia, M. K., Jain, S. K. (2003). Pharmaceutical approaches to colon targeted drug delivery systems. *J. Pharm. Pharm. Sci.*, *6*, 33–66.
19. Cabri, W., Di Fabio, R. (2000). *From Bench to Market*. Oxford University Press, Oxford, p. 148.
20. Prokai, L., Prokai-Tatarai, K. (1999). In P. K. Gupta and G. A. Brazeau (Eds.), *Prodrugs. Injectable Drug Development. Techniques to Reduce Pain and Irritation*. Interpharm, Denver, CO, pp. 267–306.
21. Sinkula, A. A., Yalkowsky, S. H. (1975). Rationale for design of biologically reversible drug derivatives—prodrugs. *J. Pharm. Sci.*, *64*, 181–210.
22. Menard, J., Patchett, A. A. (2001). Angiotensin-converting enzyme inhibitors. *Adv. Protein Chem.*, *56*, 13–75.
23. Tammara, V. K., Narurkar, M. M., Crider, A. M., Khan, M. A. (1993). Synthesis and evaluation of morpholinoalkyl ester prodrugs of indomethacin and naproxen. *Pharm. Res.*, *10*, 1191–1199.
24. Rautio, J., Taipale, H., Gynther, J., Vepsäläinen, J., Nevalainen, T., Järvinen, T. (1998). *In vitro* evaluation of acyloxyalkyl esters as dermal prodrugs of ketoprofen and naproxen. *J. Pharm. Sci.*, *87*, 1622–1628.
25. Ferres, H. (1983). Pro-drugs of β-lactam antibiotics. *Drugs of Today*, *19*, 499–538.
26. Shanbhag, V. R., Crider, A. M., Gokhale, R., Harpalani, A., Dick, R. M. (1992). Ester and amide prodrugs of ibuprofen and naproxen—Synthesis, antiinflammatory activity, and gastrointestinal toxicity. *J. Pharm. Sci.*, *81*, 149–154.
27. Krise, J. P., Stella, V. J. (1996). Prodrugs of phosphates, phosphonates, and phosphinates. *Adv. Drug Delivery Rev.*, *19*, 287–310.
28. McGuigan, C., Turner, S., Nicholls, S. R., Oconnor, T. J., Kinchington, D. (1994). Haloalkyl phosphate derivatives of act as inhibitors of HIV—Studies in the phosphate region. *Antiviral Chem. Chemother.*, *5*, 162–168.
29. Schott, H., Ludwig, P. S., Immelmann, A., Schwendener, R. A. (1999). Synthesis and *in vitro* anti-HIV activities of amphiphilic heterodinucleoside phosphate derivatives containing the 2′,3′-dideoxynucleosides ddC, AZT and ddI. *Eur. J. Med. Chem.*, *34*, 343–352.
30. Pompon, A., Lefebvre, I., Imbach, J. L., Kahn, S., Farquhar, D. (1994). Decomposition pathways of the mono-(pivaloyl-oxymethyl) and bis(pivaloyl-oxymethyl) esters of azidothymidine 5′-monophosphate in cell extract and in tissue-culture medium—An application of the online Isrp-cleaning HPLC technique. *Antiviral Chem. Chemother.*, *5*, 91–98.

31. Hostetler, K. Y., Parker, S., Sridhar, C. N., Martin, M. J., Li, J. L., Stuhmiller, L. M., Vanwijk, G. M. T., Vandenbosch, H., Gardner, M. F., Aldern, K. A., Richman, D. D. (1993). Acyclovir diphosphate dimyristoylglycerol—A phospholipid prodrug with activity against acyclovir-resistant herpes-simplex virus. *Proc. Nat. Acad. Sci. USA*, *90*, 11,835–11,839.

32. Saboulard, D., Naesens, L., Cahard, D., Salgado, A., Pathirana, R., Velazquez, S., McGuigan, C., De Clercq, E., Balzarini, J. (1999). Characterization of the activation pathway of phosphoramidate triester prodrugs of stavudine and zidovudine. *Mol. Pharmacol.*, *56*, 693–704.

33. Serafinowska, H. T., Ashton, R. J., Bailey, S., Harnden, M. R., Jackson, S. M., Sutton, D. (1995). Synthesis and in vivo evaluation of prodrugs of 9-[2-(phosphonomethoxy)ethoxy]adenine. *J. Med. Chem.*, *38*, 1372–1379.

34. Starrett, J. E. Jr, Tortolani, D. R., Russell, J., Hitchcock, M. J., Whiterock, V., Martin, J. C., Mansuri, M. M. (1994). Synthesis, oral bioavailability determination, and in vitro evaluation of prodrugs of the antiviral agent 9-[2-(phosphonomethoxy)ethyl]adenine (PMEA). *J. Med. Chem.*, *37*, 1857–1864.

35. Delombaert, S., Erion, M. D., Tan, J., Blanchard, L., Elchehabi, L., Ghai, R. D., Sakane, Y., Berry, C., Trapani, A. J. (1994). *N*-Phosphonomethyl dipeptides and their phosphonate prodrugs, a new-generation of neutral endopeptidase (Nep, Ec-3.4.24.11) inhibitors. *J. Med. Chem.*, *37*, 498–511.

36. Niemi, R., Vepsäläinen, J., Taipale, H., Järvinen, T. (1999). Bisphosphonate prodrugs: Synthesis and in vitro evaluation of novel acyloxyalkyl esters of clodronic acid. *J. Med. Chem.*, *42*, 5053–5058.

37. Niemi, R., Pennanen, H., Vepsäläinen, J., Taipale, H., Järvinen, T. (1998). Bisphosphonate prodrugs: Synthesis and in vitro evaluation of novel partial amides of clodronic acid. *Int. J. Pharm.*, *174*, 111–115.

38. Larsen, C., Johansen, M. (1987). Macromolecular prodrugs. 4. Kinetics of hydrolysis of metronidazole monosuccinate dextran ester conjugates in aqueous-solution and in plasma—Sequential release of metronidazole from the conjugates at physiological pH. *Int. J. Pharm.*, *35*, 39–45.

39. Mollmann, H., Rohdewald, P., Barth, J., Mollmann, C., Verho, M., Derendorf, H. (1988). Comparative pharmacokinetics of methylprednisolone phosphate and hemisuccinate in high-doses. *Pharm. Res.*, *5*, 509–513.

40. Kong, A. N., Jusko, W. J. (1991). Disposition of methylprednisolone and its sodium succinate prodrug in vivo and in perfused liver of rats—Nonlinear and sequential 1st-pass elimination. *J. Pharm. Sci.*, *80*, 409–415.

41. Bundgaard, H., Larsen, C., Thorbek, P. (1984). Prodrugs as drug delivery systems. 26. Preparation and enzymatic-hydrolysis of various water-soluble amino-acid esters of metronidazole. *Int. J. Pharm.*, *18*, 67–77.

42. Mahfouz, N. M., Hassan, M. A. (2001). Synthesis, chemical and enzymatic hydrolysis, and bioavailability evaluation in rabbits of metronidazole amino acid ester prodrugs with enhanced water solubility. *J. Pharm. Pharm.*, *53*, 841–848.

43. Másson, M., Thorsteinsson, T., Sigurdsson, T. H., Loftsson, T. (2000). Lipophilic metronidazole derivatives and their absoprtion through hairless mouse skin. *Die Pharmazie*, *55*, 369–371.

44. Ueda, Y., Matiskella, J. D., Mikkilineni, A. B., Farina, V., Knipe, J. O., Rose, W. C., Casazza, A. M., Vyas, D. M. (1995). Novel, water-soluble phosphate derivatives of 2′-ethoxy carbonylpaclitaxel as potential prodrugs of paclitaxel—Synthesis and antitumor evaluation. *Bioorganic Med. Chem. Lett.*, *5*, 247–252.
45. Sawada, S., Yaegashi, T., Furuta, T., Yokokura, T., Miyasaka, T. (1993). Chemical modification of an antitumor alkaloid, 20(*S*)-camptothecin–E-lactone ring-modified water-soluble derivatives of 7-ethylcamptothecin. *Chem. Pharm. Bull.*, *41*, 310–313.
46. Yaegashi, T., Sawada, S., Nagata, H., Furuta, T., Yokokura, T., Miyasaka, T. (1994). Synthesis and antitumor-activity of 20(*S*)-camptothecin derivatives—A-ring-substituted 7-ethylcamptothecins and their E-ring-modified water-soluble derivatives. *Chem. Pharm. Bull.*, *42*, 2518–2525.
47. Bodor, N., Sloan, K. B., Higuchi, T., Sasahara, K. (1977). Improved delivery through biological-membranes. 4. Prodrugs of L-dopa. *J. Med. Chem.*, *20*, 1435–1445.
48. Johansen, M., Bundgaard, H. (1980). Pro-drugs as drug delivery systems. 13. Kinetics of decomposition of *N*-Mannich bases of salicylamide and assessment of their suitability as possible pro-drugs for amines. *Int. J. Pharm.*, *7*, 119–127.
49. Bundgaard, H., Johansen, M., Stella, V., Cortese, M. (1982). Pro-drugs as drug delivery systems. 21. Preparation, physicochemical properties and bioavailability of a novel water-soluble pro-drug type for carbamazepine. *Int. J. Pharm.*, *10*, 181–192.
50. Redden, R., Douglas, J. E., Burke, M. J., Horrobin, D. F. (1998). *In vitro* hydrolysis of polyunsaturated fatty acid *N*-acyloxymethyl derivatives of theophylline. *Int. J. Pharm.*, *165*, 87.
51. Bundgaard, H., Falch, E., Jensen, E. (1989). A novel solution-stable, water-soluble prodrug type for drugs containing a hydroxyl or an NH-acidic group. *J. Med. Chem.*, *32*, 2503–2507.
52. Sloan, K. B., Hashida, M., Alexander, J., Bodor, N., Higuchi, T. (1983). Prodrugs of 6-thiopurines—Enhanced delivery through the skin. *J. Pharm. Sci.*, *72*, 372–378.
53. Mollgaard, B., Hoelgaard, A., Bundgaard, H. (1982). Pro-drugs as drug delivery systems. 23. Improved dermal delivery of 5-fluorouracil through human-skin via *N*-acyloxymethyl pro-drug derivatives. *Int. J. Pharm.*, *12*, 153–162.
54. Yang, C., Tirucherai, G. S., Mitra, A. K. (2001). Prodrug based optimal drug delivery via membrane transporter/receptor. *Expert Opin. Biol. Ther.*, *1*, 159–175.
55. Shaw, J. P., Sueoka, C. M., Oliyai, R., Lee, W. A., Arimilli, M. N., Kim, C. U., Cundy, K. C. (1997). Metabolism and pharmacokinetics of novel oral prodrugs of 9-[(*R*)-2-(phosphonomethoxy)propyl]adenine (PMPA) in dogs. *Pharm. Res.*, *14*, 1824–1829.
56. Gallant, J. E., Deresinski, S. (2003). Tenofovir disoproxil fumarate. *Clin. Infect. Dis.*, *37*, 944–950.
57. Cundy, K. C., Barditch-Crovo, P., Walker, R. E., Collier, A. C., Ebeling, D., Toole, J., Jaffe, H. S. (1995). Clinical pharmacokinetics of adefovir in human immunodeficiency virus type 1-infected patients. *Antimicrob. Agents Chemother.*, *39*, 2401–2405.

58. Gustafsson, D. (2003). Oral direct thrombin inhibitors in clinical development. *J. Internal Med.*, *254*, 322–334.
59. Han, H. K., Amidon, G. L. (2000). Targeted prodrug design to optimize drug delivery. *AAPS PharmSci*, *2*, E6.
60. Steffansen, B., Nielsen, C. U., Brodin, B., Eriksson, A. H., Andersen, R., Frokjaer, S. (2004). Intestinal solute carriers: An overview of trends and strategies for improving oral drug absorption. *Eur. J. Pharm. Sci.*, *21*, 3–16.
61. de Miranda, P., Good, S. S. (1992). Species differences in the metabolism and disposition of antiviral nucleoside analogues: 1. Acyclovir. *Antiviral Chem. Chemother.*, *3*, 1–8.
62. Weller, S., Blum, M. R., Doucette, M., Burnette, T., Cederberg, D. M., de Miranda, P., Smiley, M. L. (1993). Pharmacokinetics of the acyclovir pro-drug valaciclovir after escalating single- and multiple-dose administration to normal volunteers. *Clin. Pharm. Ther.*, *54*, 595–605.
63. Beutner, K. R., Friedman, D. J., Forszpaniak, C., Andersen, P. L., Wood, M. J. (1995). Valaciclovir compared with acyclovir for improved therapy for herpes zoster in immunocompetent adults. *Antimicrob. Agents Chemother.*, *39*, 1546–1553.
64. Guo, A., Hu, P., Balimane, P. V., Leibach, F. H., Sinko, P. J. (1999). Interactions of a nonpeptidic drug, valacyclovir, with the human intestinal peptide transporter (hPEPT1) expressed in a mammalian cell line. *J. Pharmcol. Exper. Ther.*, *289*, 448–454.
65. Han, H., de Vrueh, R. L., Rhie, J. K., Covitz, K. M., Smith, P. L., Lee, C. P., Oh, D. M., Sadee, W., Amidon, G. L. (1998). 5′-Amino acid esters of antiviral nucleosides, acyclovir, and AZT are absorbed by the intestinal PEPT1 peptide transporter. *Pharm. Res.*, *15*, 1154–1159.
66. Sugawara, M., Huang, W., Fei, Y. J., Leibach, F. H., Ganapathy, V., Ganapathy, M. E. (2000). Transport of valganciclovir, a ganciclovir prodrug, via peptide transporters PEPT1 and PEPT2. *J. Pharm. Sci.*, *89*, 781–789.
67. Heimbach, T., Oh, D. M., Li, L. Y., Forsberg, M., Savolainen, J., Leppanen, J., Matsunaga, Y., Flynn, G., Fleisher, D. (2003). Absorption rate limit considerations for oral phosphate prodrugs. *Pharm. Res.*, *20*, 848–856.
68. Li, Y. H., Ito, K., Tsuda, Y., Kohda, R., Yamada, H., Itoh, T. (1999). Mechanism of intestinal absorption of an orally active beta-lactam prodrug: Uptake and transport of carindacillin in Caco-2 cells. *J. Pharmcol. Exper. Ther.*, *290*, 958–964.
69. Sloan, K. B., Wasdo, S. (2003). Designing for topical delivery: Prodrugs can make the difference. *Med. Res. Rev.*, *23*, 763–793.
70. Sloan, K. B. (1989). Prodrugs for dermal delivery. *Adv. Drug Delivery Rev.*, *3*, 67–101.
71. Sloan, K. B. (1992). Functional group considerations in the development of prodrugs approaches to solving topical delivery problems. In K. B. Sloan (Ed.), *Prodrugs; Topical and Ocular Drug Delivery*. Marcel Dekker, New York, pp. 17–116.
72. Chien, Y. W. (1983). Logics of transdermal controlled drug administration. *Drug Devel. Ind. Pharm.*, *9*, 497–520.

73. Rautio, J., Nevalainen, T., Taipale, H., Vepsäläinen, J., Gynther, J., Pedersen, T., Järvinen, T. (1999). Synthesis and in vitro evaluation of aminoacyloxyalkyl esters of 2-(6-methoxy-2-naphthyl)propionic acid as novel naproxen prodrugs for dermal drug delivery. *Pharm. Res.*, *16*, 1172–1178.

74. Rautio, J., Nevalainen, T., Taipale, H., Vepsäläinen, J., Gynther, J., Laine, K., Järvinen, T. (2000). Synthesis and in vitro evaluation of novel morpholinyl- and methylpiperazinylacyloxyalkyl prodrugs of 2-(6-methoxy-2-naphthyl)propionic acid (naproxen) for topical drug delivery. *J. Med. Chem.*, *43*, 1489–1494.

75. Rautio, J., Nevalainen, T., Taipale, H., Vepsäläinen, J., Gynther, J., Laine, K., Järvinen, T. (2000). Piperazinylalkyl prodrugs of naproxen improve in vitro skin permeation. *Eur. J. Pharm. Sci.*, *11*, 157–163.

76. Udata, C., Tirucherai, G., Mitra, A. K. (1999). Synthesis, stereoselective enzymatic hydrolysis, and skin permeation of diastereomeric propranolol ester prodrugs. *J. Pharm. Sci.*, *88*, 544–550.

77. Doane, M. G., Jensen, A. D., Dohlman, C. H. (1978). Penetration routes of topically applied eye medications. *Am. J. Ophthalmol.*, *85*, 383–386.

78. Rojanasakul, Y., Wang, L-Y., Bhat, M., Glover, D. D., Malanga, C. J., Ma, K. H. (1992). The transport barrier of epithelia: A comparative study on membrane permeability and charge selectivity in the rabbit. *Pharm. Res.*, *9*, 1029–1034.

79. Chien, D-S., Sasaki, H., Bundgaard, H., Buur, A., Lee, V. H. L. (1991). Role of enzymatic lability in the corneal and conjunctival penetration on timolol ester prodrugs in the pigmented rabbit. *Pharm. Res.*, *8*, 728–733.

80. Wang, W., Sasaki, H., Chien, D.-S., Lee, V. H., L. (1991). Lipophilicity influence on conjunctival drug penetration in the pigmented rabbit: A comparison with corneal penetration. *Curr. Eye Res.*, *10*, 571–579.

81. Schoenwald, R. D., Huang, H.-S. (1983). Corneal penetration behavior of beta-blocking agents I: Physicochemical factors. *J. Pharm. Sci.*, *72*, 1266–1272.

82. Schoenwald, R. D., Ward, R. L. (1978). Relationship between steroid permeability across excised rabbit cornea and octanol-water partition coefficients. *J. Pharm. Sci.*, *67*, 787–789.

83. Sasaki, H., Yamamura, M., Mukai, T., Nishida, K., Nakamura, J., Nakashima, M., Ichikawa, M. (1999). Enhancement of ocular drug penetration. *Crit. Rev. Ther. Drug Carrier Syst.*, *16*, 85–146.

84. Hussain, A., Truelove, J. E. (1976). Prodrug approaches to enhancement of physicochemical properties of drugs IV: Novel epinephrine prodrug. *J. Pharm. Sci.*, *65*, 1510–1512.

85. Anderson, J. A. (1980). Systemic absorption of topical ocularly applied epinephrine and dipivefrin. *Arch. Ophthalmol.*, *98*, 350–353.

86. Mandell, A. I., Stentz, F. (1978). Dipivalyl epinephrine: A new pro-drug in the treatment of glaucoma. *Ophthalmology*, *85*, 268–275.

87. Wei, C., Anderson, J. A., Leopold, I. (1978). Ocular absorption and metabolism of topically applied epinephrine and a dipivalyl ester of epinephrine. *Investigative Ophthalmol. Vis. Sci.*, *17*, 315–321.

88. Kaback, M. B., Podos, S. M., Harbin, Jr. T. S., Mandell, A., Becker, B. (1976). The effects of dipivalyl epinephrine on the eye. *Am. J. Ophthalmol.*, *81*, 768–772.

89. Kohn, A. N., Moss, A. P., Hargett, N. A., Ritch, R., Smith, H., Podos, S. M. (1979). Clinical comparison of dipivalyl epinephrine and epinephrine in the treatment of glaucoma. *Am. J. Ophthalmol.*, *87*, 196–201.
90. Susanna, R., Chew, P., Kitazawa, Y. (2002). Current status of prostaglandin therapy; Latanoprost and unoprostone. *Surv. Ophthalmol.*, *47*, S97–S104.
91. Netlans, P. A., Landry, T., Sullivan, E. K., Andrew, R., Silver, L., Weiner, A., Mallick, S., Dickerson, J., Bergamini, W. V. W., Robertson, S. M., Davis, A. D. (2001). Travoprost compared with latanoprost and timolol in patients with open-angle glaucoma or ocular hypertension. *Am. J. Ophthalmol.*, *132*, 472–484.
92. Suhonen, P., Järvinen, T., Lehmussaari, K., Reunamäki, T., Urtti A. (1996). Rate control of ocular pilocarpine delivery with bispilocarpic acid diesters. *Int. J. Pharm.*, 127, 85–94.
93. Chang, S.-C., Bundgaard, H., Buur, A., Lee, V. H. L. (1988). Low dose *O*-butyryl timolol improves the therapeutic index of timolol in the pigmented rabbit. *Investigative Ophthalmol. Vis. Sci.*, *29*, 626–629.
94. Sasaki, H., Igarashi, Y., Nishida, K., Nakamura, J. (1993). Ocular delivery of the β-blocker, tilisolol, through the prodrug approach. *Int. J. Pharm.*, *93*, 49–60.
95. Anand, B. S., Mitra, A. K. (2002). Mechanism of corneal permeation of L-valyl ester of acyclovir: Targeting the oligopeptide transporter on the rabbit cornea. *Pharm. Res.*, *19*, 1194–1202.
96. Washington, N., Washington, C., Wilson, C. G. (2001). Nasal drug delivery. In *Physiological Pharmaceutics. Barriers to Drug Absorption*, 2nd ed., Taylor & Francis, London, pp. 199–220.
97. Pezron, I., Tirucherai, G. S., Duvvuri, S., Mitra, A. K. (2002). Prodrug strategies in nasal drug delivery. *Expert Opin. Ther. Patents*, *12*, 331–340.
98. Hussain, A. A., Al-Bayatti, A. A., Dakkuri, A., Okochi, K., Hussain, M. A. (2002). Testosterone 17β-*N*,*N*-dimethylglycinate hydrochloride: A prodrug with a potential for nasal delivery of testosterone. *J. Pharm. Sci.*, *91*, 785–789.
99. Loftsson, T., Gudmundsdottir, H., Sigurjonsdottir, J. F., Sigurdsson, H. H., Sigfusson, S. D., Masson, M., Stefansson, E. (2001). Cyclodextrin solubilization of benzodiazepines: Formulation of midazolam nasal spray. *Int. J. Pharm.*, *212*, 29–40.
100. Gudmundsdottir, H., Sigurjonsdottir, J. F., Masson, M., Fjalldal, O., Stefansson, E., Loftsson, T. (2001). Intranasal administration of midazolam in a cyclodextrin based formulation: Bioavailability and clinical evaluation in humans. *Pharmazie*, *56*, 963–966.
101. Cox, D. S., Raje, S., Gao, H. L., Salama, N. N., Eddington, N. D. (2002). Enhanced permeability of molecular weight markers and poorly bioavailable compounds across Caco-2 cell monolayers using the absorption enhancer, zonula occludens toxin. *Pharm. Res.*, *19*, 1680–1688.
102. Shao, Z., Hoffman, A. J., Mitra, A. K. (1994). Biodegradation characteristics of acyclovir 2′-esters by respiratory carboxylesterases: Implications in prodrug design for intranasal and pulmonary drug delivery. *Int. J. Pharm.*, *112*, 181–190.
103. Shao, Z., Park, G.-B., Krishnamoorthy, R., Mitra, A. K. (1994). The physicochemical properties, plasma enzymatic hydrolysis, and nasal absorption of acyclovir and its 2′-ester prodrugs. *Pharm. Res.*, *11*, 237–242.

104. Hussain, A. A. (1998). Intranasal drug delivery. *Adv. Drug Delivery Rev.*, *29*, 39–49.
105. Kao, H. D., Traboulsi, A., Itoh, S., Dittert, L., Hussain, A. (2000). Enhancement of systemic and CNS specific delivery of L-dopa by the nasal administration of its water soluble prodrugs. *Pharm. Res.*, *17*, 978–984.
106. Yang, C., Mitra, A. K. (2001). Nasal absorption of tyrosine-linked model compounds. *J. Pharm. Sci.*, *90*, 340–347.
107. Yang, C., Gao, H., Mitra, A. K. (2001). Chemical stability, enzymatic hydrolysis, and nasal uptake of amino acid ester prodrugs of acyclovir. *J. Pharm. Sci.*, *90*, 617–624.
108. Rasmussen, G. J., Bundgaard, H. (1991). Prodrugs of peptides. 15. 4-Imidazolidinone prodrug derivatives of enkephalins to prevent aminopeptidase-catalyzed metabolism in plasma and absorptive mucosae. *Int. J. Pharm.*, *76*, 113–122.
109. Tirucherai, G. S., Pezron, I., Mitra, A. K. (2002). Novel approaches to nasal delivery of peptides and proteins. *S. T.P. Pharma Sci.*, *12*, 3–12.
110. Witterland, A. H. I., Koks, C. H. W., Beijnen, J. H. (1996). Etoposide phosphate, the water soluble prodrug of etoposide. *Pharm. World Sci.*, *18*, 163–170.
111. Krise, J. P., Narisawa, S., Stella, V. J. (1999). A novel prodrug approach for tertiary amines. 2. Physicochemical and *in vitro* enzymatic evaluation of selected *N*-phosphonooxymethyl prodrugs. *J. Pharm. Sci.*, *88*, 922–927.
112. Krise, J. P., Charman, W. N., Charman, S. A., Stella, V. J. (1999). A novel prodrug approach for tertiary amines. 3. *In vivo* evaluation of two *N*-phosphonooxymethyl prodrugs in rats and dogs. *J. Pharm. Sci.*, *88*, 928–932.
113. Krise, J. P., Zygmunt, J., Georg, G. I., Stella, V. J. (1999). Novel prodrug approach for tertiary amines: Synthesis and preliminary evaluation of *N*-phosphonooxymethyl prodrugs. *J. Med. Chem.*, *42*, 3094–3100.
114. Larsen, C., Kurtzhals, P., Johansen, M. (1988). Kinetics of regeneration of metronidazole from hemiesters of maleic acid, succinic acid and glutaric acid in aqueous buffer, human plasma and pig liver homogenate. *Int. J. Pharm.*, *41*, 121–129.
115. Robinson, R. P., Reiter, L. A., Barth, W. E., Campeta, A. M., Cooper, K., Cronin, B. J., Destito, R., Donahue, K. M., Falkner, F. C., Fiese, E. F., Johnson, D. L., Kuperman, A. V., Liston, T. E., Malloy, D., Martin, J. J., Mitchell, D. Y., Rusek, F. W., Shamblin, S. L., Wright, C. F. (1996). Discovery of the hemifumarate and (alpha-L-alanyloxy)methyl ether as prodrugs of an antirheumatic oxindole: Prodrugs for the enolic OH group. *J. Med. Chem.*, *39*, 10–18.
116. Hendler, S. S., Sanchez, R. A., Zielinski, J. (2001). Water soluble pro-drugs of propofol. U. S. Patent 6,254,853 B1.
117. Hudkins, R. L., Iqbal, M., Park, C. H., Goldstein, J., Herman, J. L., Shek, E., Murakata, C., Mallamo, J. P. (1998). Prodrug esters of the indolocarbazole CEP-751 (KT-6587). *Bioorganic Med. Chem. Lett.*, *8*, 1873–1876.
118. Bradshaw, T. D., Chua, M. S., Browne, H. L., Trapani, V., Sausville, E. A., Stevens, M. F. (2002). *In vitro* evaluation of amino acid prodrugs of novel antitumour 2-(4-amino-3-methylphenyl)benzothiazoles. *Br. J. Cancer*, *86*, 1348–1354.
119. Nam, N. H., Kim, Y., You, Y. J., Hong, D. H., Kim, H. M., Ahn, B. Z. (2003). Water soluble prodrugs of the antitumor agent 3-(3-amino-4-methoxy)phenyl]-2-(3,4,5-trimethoxyphenyl)cyclopent-2-ene-1-one. *Bioorganic Med. Chem.*, *11*, 1021–1029.

120. Marder, S. R., Aravagiri, M., Wirshing, W. C., Wirshing, D. A., Lebell, M., Mintz, J. (2002). Fluphenazine plasma level monitoring for patients receiving fluphenazine decanoate. *Schizophr. Res.*, *53*, 25–30.
121. Shechter, Y., Tsubery, H., Fridkin, M. (2003). Suspensions of pro-drug insulin greatly prolong normoglycemic patterns in diabetic rats. *Biochem. Biophys. Res. Commun.*, *307*, 315–321.
122. Svensson, L. A., Tunek, A. (1988). The design and bioactivation of presystemically stable prodrugs. *Drug Metabolism Rev.*, *19*, 165–194.
123. Tunek, A., Levin, E., Svensson, L. A. (1988). Hydrolysis of 3*H*-bambuterol, a carbamate prodrug of terbutaline, in blood from humans and laboratory animals *in vitro*. *Biochem. Pharmcol.*, *37*, 3867–3876.
124. Sinhababu, A. K., Thakker, D. R. (1996). Prodrugs of anticancer agents. *Adv. Drug Delivery Rev.*, *19*, 241–273.
125. Brown, J. M., Giaccia, A. J. (1998). The unique physiology of solid tumors: Opportunities (and problems) for cancer therapy. *Cancer Res.*, *58*, 408–416.
126. Naylor, M. A., Thomson, P. (2001). Recent advances in bioreductive drug targeting. *Mini Rev. Med. Chem.*, *1*, 17–29.
127. Wardman, P., Dennis, M. F., Everett, S. A., Patel, K. B., Stratford, M. R., Tracy, M. (1995). Radicals from one-electron reduction of nitro compounds, aromatic *N*-oxides and quinones: The kinetic basis for hypoxia-selective, bioreductive drugs. *Biochem. Soc. Symp.*, *61*, 171–194.
128. Xu, G., McLeod, H. L. (2001). Strategies for enzyme/prodrug cancer therapy. *Clin. Cancer Res.*, *7*, 3314–3324.
129. Jung, M. (2001). Antibody directed enzyme prodrug therapy (ADEPT) and related approaches for anticancer therapy. *Mini Rev. Med. Chem.*, *1*, 399–407.
130. Aghi, M., Hochberg, F., Breakefield, X. O. (2000). Prodrug activation enzymes in cancer gene therapy. *J. Gene Med.*, *2*, 148–164.
131. Chen, L., Waxman, D. J. (2002). Cytochrome P450 gene-directed enzyme prodrug therapy (GDEPT) for cancer. *Curr. Pharm. Des.*, *8*, 1405–1416.
132. Niculescu-Duvaz, I., Spooner, R., Marais, R., Springer, C. J. (1998). Gene-directed enzyme prodrug therapy. *Bioconjugate Chem.*, *9*, 4–22.
133. Springer, C. J., Niculescu-Duvaz, I. (2000). Approaches to gene-directed enzyme prodrug therapy (GDEPT). *Adv. Exper. Med. Biol.*, *465*, 403–409.
134. Genka, S., Deutsch, J., Shetty, U. H., Stahle, P. L., John, V., Lieberburg, I. M., Ali-Osman, F., Rapoport, S. I, Greig, N. H. (1993). Development of lipophilic anticancer agents for the treatment of brain tumors by the esterification of water-soluble chlorambucil. *Clin. Exper. Metastasis*, *11*, 131–140.
135. Greig, N. H., Genka, S., Daly, E. M., Sweeney, D. J., Rapoport, S. I. (1990). Physicochemical and pharmacokinetic parameters of seven lipophilic chlorambucil esters designed for brain penetration. *Cancer Chemother. Pharmcol.*, *25*, 311–319.
136. Bodor, N., Buchwald, P. (1999). Recent advances in the brain targeting of neuropharmaceuticals by chemical delivery systems. *Adv. Drug Delivery Rev.*, *36*, 229–254.
137. Bodor, N., Buchwald, P. (2002). Barriers to remember: Brain-targeting chemical delivery systems and Alzheimer´s disease. *Drug Discov. Today*, *7*, 766–774.

138. Tsuji, A., Tamai, I. I. (1999). Carrier-mediated or specialized transport of drugs across the blood-brain barrier. *Adv. Drug Delivery Rev.*, *5*, 277–290.

139. Tamai, I., Tsuji, A.(2000). Transporter-mediated permeation of drugs across the blood-brain barrier. *J. Pharm. Sci.*, *89*, 1371–1388.

140. Halmos, T., Santarromana, M., Antonakis, K., Scherman, D. (1996). Synthesis of glucose-chlorambucil derivatives and their recognition by the human GLUT1 glucose transporter. *Eur. J. Pharmcol.*, *318*, 477–484.

141. Bonina, F. P., Arenare, L., Palagiano, F., Saija, A., Nava, F., Trombetta, D., de Caprariis, P. (1999). Synthesis, stability, and pharmacological evaluation of nipecotic acid prodrugs. *J. Pharm. Sci.*, *88*, 561–567.

142. Battaglia, G., La Russa, M., Bruno, V., Arenare, L., Ippolito, R., Copani, A., Bonina, F., Nicoletti, F. (2000). Systemically administered D-glucose conjugates of 7-chlorokynurenic acid are centrally available and exert anticonvulsant activity in rodents. *Brain Res.*, *860*, 149–156.

143. Manfredini, S., Pavan, B., Vertuani, S., Scaglianti, M., Compagnone, D., Biondi, C., Scatturin, A., Tanganelli, S., Ferraro, L., Prasad, P., Dalpiaz, A. (2002). Design, synthesis and activity of ascorbic acid prodrugs of nipecotic, kynurenic and diclophenamic acids, liable to increase neurotropic activity. *J. Med. Chem.*, *45*, 559–562.

144. Kearney, A. S. (1996). Prodrugs and targeted drug delivery. *Adv. Drug Delivery Rev.*, *19*, 225–239.

145. Chourasia, M. K., Jain, S. K. (2003). Pharmaceutical approaches to colon targeted drug delivery systems. *J. Pharm. Pharm. Sci.*, *6*, 33–66.

146. Haas, M., Moolenaar, F., Meijer, D. K., de Zeeuw, D. (2002). Specific drug delivery to the kidney. *Cardiovasc. Drugs Ther.*, *16*, 489–496.

147. Wilk, S., Mizoguchi, H., Orlowski, M. (1978). γ-Glutamyl dopa: A kidney-specific dopamine precursor. *J. Pharmcol. Exper. Ther.*, *206*, 227–232.

148. Orlowski, M., Mizoguchi, H., Wilk, S. (1980). *N*-acyl-gamma-glutamyl derivatives of sulfamethoxazole as models of kidney-selective prodrugs. *J. Pharmcol. Exper. Ther.*, *212*, 167–172.

149. Lee, M. R. (1990). Five years' experience with gamma-L-glutamyl-L-dopa: A relatively renally specific dopaminergic prodrug in man. *J. Autonomic Pharmcol.*, *10*, 103–108.

150. Gangwar, S., Pauletti, G. M., Wang, B. H., Siahaan, T. J., Stella, V. J., Borchardt, R. T. (1997). Prodrug strategies to enhance the intestinal absorption of peptides. *Drug Discov. Today*, *2*, 148–155.

151. Borchardt, R. T. (1999). Optimizing oral absorption of peptides using prodrug strategies. *J. Controlled Release*, *62*, 231–238.

152. Ripka, A. S., Rich, D. H. (1998). Peptidomimetic design. *Curr. Opin. Chem. Biol.*, *2*, 441–452.

153. Oliyai, R. (1996). Prodrugs of peptides and peptidomimetics for improved formulation and delivery. *Adv. Drug Delivery Rev.*, *19*, 275–286.

154. Oliyai, R., Stella, V. J. (1992). Kinetics and mechanism of isomerization of cyclosporine-A. *Pharm. Res.*, *9*, 617–622.

155. Hamada, Y., Matsumoto, H., Kimura, T., Hayashi, Y., Kiso, Y. (2003). Effect of the acyl groups on O → N acyl migration in the water-soluble prodrugs of HIV-1 protease inhibitor. *Bioorganic Med. Chem. Lett.*, *13*, 2727–2730.

156. Hamada, Y., Ohtake, J., Sohma, Y., Kimura, T., Hayashi, Y., Kiso, Y. (2002). New water-soluble prodrugs of HIV protease inhibitors based on O → N intramolecular acyl migration. *Bioorganic Med. Chem.*, *10*, 4155–4167.
157. Bundgaard, H. (1992). The utility of the prodrug approach to improve peptide absorption. *J. Controlled Release*, *21*, 63–72.
158. Delie, F., Couvreur, P., Nisato, D., Michel, J. B., Puisieux, F., Letourneux, Y. (1994). Synthesis and *in-vitro* study of a diglyceride prodrug of a peptide. *Pharm. Res.*, *11*, 1082–1087.
159. Friis, G. J. (2000). Peptide and protein derivatives. In S. Frokjaer and L. Hovgaard (Eds.), *Pharmaceutical. Formulation Development of Peptides and Proteins*. Taylor & Francis, London, pp. 206–219.
160. Kahns, A. H., Buur, A., Bundgaard, H. (1993). Prodrugs of peptides. 18. Synthesis and evaluation of various esters of desmopressin (Ddavp). *Pharm. Res.*, *10*, 68–74.
161. Gudmundsson, O. S., Jois, S. D. S., Vander Velde, D. G., Siahaan, T. J., Wang, B., Borchardt, R. T. (1999). The effect of conformation on the membrane permeation of coumarinic acid-, and phenylpropionic acid-based cyclic prodrugs of opioid peptides. *J. Peptide Res.*, *53*, 383–392.
162. Bak, A., Gudmundsson, O. S., Friis, G. J., Siahaan, T. J., Borchardt, R. T. (1999). Acyloxyalkoxy-based cyclic prodrugs of opioid peptides: Evaluation of the chemical and enzymatic stability as well as their transport properties across Caco-2 cell monolayers. *Pharm. Res.*, *16*, 24–29.
163. Gudmundsson, O. S., Nimkar, K., Gangwar, S., Siahaan, T., Borchardt, R. T. (1999). Phenylpropionic acid-based cyclic prodrugs of opioid peptides that exhibit metabolic stability to peptidases and excellent cellular permeation. *Pharm. Res.*, *16*, 16–23.
164. Chang, S.-C., Bundgaard, H., Buur, A., Lee, V. H. L. (1987). Improved corneal penetration of timolol by prodrugs as a means to reduce systemic drug load. *Investigative Ophthalmol. Vis. Sci.*, *29*, 487–491.
165. Mahmud, T., Rafi, S. S., Scott, D. L., Wriilesworth, J. M., Bjarnason, I. (1996). Nonsteroidal antiiflammatory drugs and uncoupling of mitochondrial oxidative phosphorylation. *Arthritis Rheum.*, *39*, 1998–2003.
166. Ogiso, T., Iwaki, M., Tanino, T., Nagai, T., Ueda, Y., Muraoka, O., Tanabe, G. (1996). Pharmacokinetics of indomethacin ester prodrugs: Gastrointestinal and hepatic toxicity and the hydrolytic capacity of various tissues in rats. *Biol. Pharm. Bull.*, *19*, 1178–1183.
167. Samara, E., Avnir, D., Ladkani, D., Bialer, M. (1995). Pharmacokinetic analysis of diethylcarbonate prodrugs of ibuprofen and naproxen. *Biopharm. Drug Disposition*, *16*, 201–210.
168. Bonina, F., Trombetta, D., Borzi, A., De Pasquale, A., Saija, A. (1997). 1-Ethylazacycloalkan-2-one indomethacin esters as new oral prodrugs: Chemical stability, enzymatic hydrolysis, anti-inflammatory activity and gastrointestinal toxicity. *Int. J. Pharm.*, *156*, 245–250.
169. Jaksic, P., Mlinaric-Majerski, K., Zorc, B., Dumic, M. (1996). Macromolecular prodrugs. VI. Kinetic study of poly[α,β-(*N*-2-hydroxyethyl-DL-aspartamide]-ketoprofen hydrolysis. *Int. J. Pharm.*, *135*, 177–182.
170. Mahfouz, N. M., Omar, F. A., Aboul-Fadl, T. (1999). Cyclic amide derivatives as potential prodrugs II: *N*-hydroxymethylsuccinimide-/istatin esters of some

NSAIDs as prodrugs with an improved therapeutic index. *Eur. J. Med. Chem.*, *34*, 551–562.

171. Duncan, R., Gac-Breton, S., Keane, R., Musila, R., Sat, Y. N., Satchi, R., Searle, F. (2001). Polymer-drug conjugates, PDEPT and PELT: Basic principles for design and transfer from the laboratory to clinic. *J. Controlled Release*, *74*, 135–146.
172. Bickel, U., Yoshikawa, T., Pardridge, W. M. (2001). Delivery of peptides and proteins through the blood-brain barrier. *Adv. Drug Delivery Rev.*, *46*, 247–279.
173. Mehvar, R. (2000). Dextrans for targeted and sustained delivery of therapeutic and imaging agents. *J. Controlled Release*, *69*, 1–25.
174. Lovrek, M., Zorc, B., Boneschans, B., Butula, I. (2000). Macromolecular prodrugs. VIII. Synthesis of polymer-gemfibrozil conjugates. *Int. J. Pharm.*, *200*, 59–66.
175. van der Merwe, T., Boneschans, B., Zorc, B., Breytenbach, J., Zovko, M. (2002). Macromolecular prodrugs: X. Kinetics of fenoprofen release from PHEA-fenoprofen conjugate. *Int. J. Pharm.*, *241*, 223–230.
176. Li, C. (2002). Poly(L-glutamic acid)—Anticancer drug conjugates. *Adv. Drug Delivery Rev.*, *54*, 695–713.
177. Kopecek, J., Kopeckova, P., Minko, T., Lu, Z. (2000). HPMA copolymer-anticancer drug conjugates: Design, activity, and mechanism of action. *Eur. J. Pharm. Biopharm.*, *50*, 61–81.
178. Maeda, H., Sawa, T., Konno, T. (2001). Mechanism of tumor-targeted delivery of macromolecular drugs, including the EPR effect in solid tumor and clinical overview of the prototype polymeric drug SMANCS. *J. Controlled Release*, *74*, 47–61.
179. Maeda, H. (2001). SMANCS and polymer-conjugated macromolecular drugs: Advantages in cancer chemotherapy. *Adv. Drug Delivery Rev.*, *46*, 169–185.
180. Greenwald, R. B., Conover, C. D., Choe, Y. H. (2000). Poly(ethylene glycol) conjugated drugs and prodrugs: A comprehensive review. *Crit. Rev. Ther. Drug Carrier Sys.*, *17*, 101–161.
181. Greenwald, R. B. (2001). PEG drugs: An overview. *J. Controlled Release*, *74*, 159–171.
182. Greenwald, R. B., Choe, Y. H., McGuire, J., Conover, C. D. (2003). Effective drug delivery by PEGylated drug conjugates. *Adv. Drug Delivery Rev.*, *55*, 217–250.

# GABA AND GLUTAMATE RECEPTOR LIGANDS AND THEIR THERAPEUTIC POTENTIAL IN CNS DISORDERS

Ulf Madsen, Hans Bräuner-Osborne,
Jeremy R. Greenwood, Tommy N. Johansen,
Povl Krogsgaard-Larsen, Tommy Liljefors,
Mogens Nielsen, and Bente Frølund

*Danish University of Pharmaceutical Sciences*
*Copenhagen, Denmark*

1 THERAPEUTIC PROSPECTS FOR GABA AND GLUTAMATE RECEPTOR LIGANDS: GENERAL ASPECTS
GABA and Glutamate Receptor Ligands as Potential Drugs in Neurological Diseases
Central Nervous System Ischemia
Alzheimer's Disease
Other Diseases and Disease Conditions
2 GABA: INHIBITORY NEUROTRANSMITTER
Classification of the GABA System and Therapeutic Targets
GABA Biosynthesis and Metabolism
GABA Reuptake
The $GABA_A$ Receptor Complex
$GABA_B$ Receptor Ligands
$GABA_C$ Receptor Agonists and Antagonists
3 GLUTAMATE: EXCITATORY NEUROTRANSMITTER
Classification of Glutamic Acid Receptors
NMDA Receptor Ligands
AMPA Receptor Ligands

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

KA Receptor Agonists and Antagonists
Structure of Ionotropic Glutamate Receptors
Metabotropic Glutamate Receptor Ligands
Structure of Metabotropic Glutamate Receptors
References

## 1 THERAPEUTIC PROSPECTS FOR GABA AND GLUTAMATE RECEPTOR LIGANDS: GENERAL ASPECTS

The complex functions of the mammalian central nervous system (CNS) are primarily determined by two superior systems: (1) excitation by the major excitatory amino acid transmitter, (*S*)-glutamic acid (Glu), which depolarizes neurons via a large number of receptor subtypes, and (2) inhibition by 4-aminobutyric acid (GABA), which hyperpolarizes neurons via multiple receptor subtypes. Although the roles of amino acid neurotransmitters in the CNS are far from being fully mapped out, accumulating evidence from a variety of neurochemical and electrophysiological studies seems to indicate that a majority, perhaps all, central neurons are under excitatory and inhibitory control by Glu and GABA [1–4].

It is interesting to note that the brain and spinal cord functions are essentially determined by two simple amino acids. However, the transmission processes mediated by Glu and GABA are very complex and apparently highly regulated. The functions and interactions of the amino acid neurotransmitter systems are further complicated by the fact that Glu is the biosynthetic precursor of GABA. Although GABA predominantly hyperpolarizes neurons, depolarizing actions of GABA are also observed, primarily at the early stages of development of the brain. In addition, there is evidence to suggest that a third mechanism may play a fundamental role in the function of the brain, namely disinhibition. The indirect neuronal excitation implies synaptic contact between two inhibitory neurons. This indirect excitatory mechanism in the CNS has never been unequivocally proven or disproven, but a number of apparently paradoxical observations have been explained on the basis of disinhibition.

Both Glu and GABA operate through ionotropic as well as metabotropic receptors. Whereas the former classes of receptors, iGluRs and $GABA_{A,C}$, mediate fast neurotransmission, the metabotropic receptors, mGluRs and $GABA_B$, are involved in slow neurotransmission processes. All of these subclasses of receptors are heterogeneous and may be involved in a variety of CNS disorders and disease conditions. Thus, all subtypes of Glu and GABA receptors are potential therapeutic targets [5–7].

In light of the ubiquitous presence of these amino acid receptors in the CNS, it is not straightforward to design Glu and GABA receptor ligands as

drugs with selective effects on diseases involving degenerating or derailed amino acid neurotransmission in restricted brain areas. Glu and GABA receptor subtypes are, however, unevenly distributed in the CNS, and this fact opens up the prospect of developing receptor ligands as drugs with a regioselective effect on particular brain disorders [1–7].

There are numerous examples of therapeutic agents acting as agonists or antagonists at metabotropic (G-protein-coupled) receptors in other CNS neurotransmitter systems, suggesting that from a drug development point of view the metabotropic $GABA_B$ receptors and the mGluRs would be particularly attractive. On the other hand, the ionotropic $GABA_{A,C}$ receptors and the iGluRs, which mediate fast neurotransmission, represent complex challenges as drug targets. Thus, full antagonists or agonists at ionotropic Glu or GABA receptors are less likely to be useful therapeutic agents. Whereas the former type of ligands may cause severe CNS disturbances via blockade of fundamental neurotransmission processes, Glu agonists are likely to exert neurotoxic effects in the human CNS, as observed in numerous animal models. Full $GABA_A$ agonists, which show effective receptor desensitization in model studies, may also desensitize receptors in human brain tissue after systemic administration and thus provoke undesirable "functional antagonism." These issues make partial agonists targeting ionotropic Glu or GABA receptors more attractive as therapeutic agents. Medicinal chemists are faced not only with the challenges of designing partial agonists with optimal levels of efficacy for each disease with malfunctioning ionotropic amino acid receptor mechanisms, but also of developing principles for the design of such ligands.

Modulation of ionotropic amino acid receptors by compounds interacting noncompetitively with the receptors via binding to physiological modulatory sites or nonphysiological binding sites at receptor protein interfaces are likely to play increasingly important roles as therapeutic agents. A notable example of this category of drugs is the benzodiazepines, which interact with binding sites of no apparent physiological function, but located at protein interfaces in heteropentameric $GABA_A$ receptor complexes.

## GABA and Glutamate Receptor Ligands as Potential Drugs in Neurological Diseases

It is well known that Glu can act as a neurotoxin, especially when energy supply is compromised. This has given rise to the proposal that injury to neurons in many neurological disorders may be caused, at least in part, by overstimulation of Glu receptors (excitotoxicity) [8]. At least to some extent, this may involve reversal of Glu transporters. Such neurological conditions range from acute insults such as stroke, trauma, and epilepsy, to chronic neurodegenerative states such as Huntington's disease, Parkinson's disease, AIDS (acquired immunodeficiency syndrome) dementia, and amyotrophic lateral sclerosis [1, 8].

## Central Nervous System Ischemia

Stroke was the first clinical indication considered for Glu receptor antagonists because the evidence for an etiological role of excitotoxic mechanisms is the most convincing. A very large number of studies have been performed with iGluR antagonists in experimentally induced ischemic stroke in laboratory animals, and, in most cases, these model experiments have revealed neuroprotective effects elicited by different subclasses of competitive and noncompetitive iGluR antagonists.

Interpretation of these results and extrapolation of the observations to the human clinic are, however, complicated by the different effects of iGluR antagonists in models of focal and global ischemia, and the quite pronounced influence on the results of the experimental condition. A number of iGluR antagonists entered early clinical trials for stroke, and, with different patterns and degrees of severity, these compounds have produced severe and frequently unacceptable side effects such as psychotomimetic effects, respiratory depression, and cardiovascular dysregulation.

These predominantly negative clinical results seem to emphasize that blockade of iGluRs by full antagonists has limited therapeutic potential. In principle, low-efficacy iGluR agonists that thus predominantly antagonize the receptors may have some therapeutic potential. Alternatively, functional iGluR partial agonism, established by coadministration of an iGluR antagonist and an iGluR agonist at a fixed ratio [9] may find therapeutical use. Although difficult to establish as a practical approach, functional partial agonism does, in principle, present a flexible approach to therapies based on the concept of partial agonism.

Little is known about the prospects for mGluR ligands, antagonists or perhaps agonists, in ischemia, but there generally is a growing interest in this heterogeneous group of receptors as therapeutic targets.

## Alzheimer's Disease

Alzheimer's disease (AD) is a neurodegenerative disorder characterized by cell loss, pathological changes in neuronal transmission, and the formation of senile plaques and neurofibrillary tangles. The cholinergic transmitter system is particularly vulnerable to degeneration, and loss of cholinergic pathways in the CNS may largely explain the cognitive impairments and memory deficiencies that are prominent symptoms of AD patients [1].

In the progression of the neurodegenerative processes in the brains of AD patients, not only cholinergic neurons but also other transmitter systems, including Glu and GABA neurons, are increasingly affected. It has been proposed that the β-amyloid content of the senile plaques renders neighboring neurons vulnerable to excitotoxicity, which is likely to contribute to the multifactorial disease mechanism in AD.

Whereas neither full Glu receptor agonists nor antagonists appear to have obvious therapeutic prospects in AD, partial Glu agonists or functional partial

Glu agonists with appropriately balanced agonist/antagonist profiles may, at least in principle, hold interest as neuroprotective agents or drugs for symptomatic treatment of AD patients.

Since Glu neurotransmitter systems, like cholinergic neuronal pathways, play a key role in learning and memory functions, various GluR ligands have been studied in animal models of these CNS functions. There has recently been a particular pharmacological interest in compounds showing positive noncompetitive modulatory effects at iGluRs. Such compounds have been shown to improve learning and memory in model systems, but, obviously, potentiation of iGluR functions may provoke excitotoxicity, emphasizing that therapeutic targeting of iGluRs to AD is a difficult balancing trick.

The results of studies on different $GABA_A$ receptor ligands in animal models relevant to learning and memory seem to support GABAergic therapeutic approaches in AD. Whereas administration of $GABA_A$ agonists impairs learning and memory in animals via inhibition of cholinergic pathways, memory enhancement is observed after administration of a $GABA_A$ antagonist. Similarly, compounds showing negative noncompetitive modulatory effects at $GABA_A$ receptors enhance performance in learning and memory tasks in animal models.

The lack of positive and particularly negative effects of high-efficacy partial $GABA_A$ agonists in AD patients is of interest and may support the view that low-efficacy partial $GABA_A$ agonists hold clinical potential in AD.

## Other Diseases and Disease Conditions

As the result of extensive neurobiological, molecular pharmacological, animal behavioral, and clinical studies, a steadily increasing number of potential therapeutic targets among the Glu and GABA neurotransmitter systems are being identified. Compounds potentiating $GABA_A$ receptor functions, including benzodiazepines, and agents that increase GABA levels via inhibition of the GABA-metabolizing enzyme, GABA-transaminase (GABA-T), or GABA reuptake systems all show antiepileptic effects. Different competitive and noncompetitive Glu receptor antagonists are currently undergoing clinical trials in epilepsy [1, 2].

Growing evidence suggests involvement of Glu transmitter systems in schizophrenia [1, 10, 11]. It is possible that schizophrenia is a reflection of a pattern of imbalances in the CNS and not simply a unidirectional change in neurotransmission. Several lines of research suggest that Glu neurotransmitter mechanisms contribute to the pathophysiology of schizophrenia. Based on extensive animal behavioral studies, drugs that target mGluRs may be considered more attractive than iGluR ligands therapeutically, as well as regards side effect. Undoubtedly animal behavioral and clinical studies will steadily grow in this area.

The number of diseases and disease conditions potentially susceptible to Glu and GABA receptor ligand therapeutic interventions is increasing. Dif-

ferent GluR ligands have been subjected to intensive studies in various models of pain, and a high-efficacy partial $GABA_A$ agonist has been shown in the human clinic to possess potent nonopioid analgesic effects [2]. This latter category of compounds also is of major clinical interest as agents capable of reestablishing normal sleep architecture in patients suffering from sleep disorder [2, 12].

## 2 GABA: INHIBITORY NEUROTRANSMITTER

### Classification of the GABA System and Therapeutic Targets

The GABA system is the major inhibitory neurotransmitter system in the CNS, and GABAergic mechanisms are directly involved in a variety of physiological and behavioral processes and in many neurological illnesses. The GABA neurotransmitter system includes a number of synaptic processes (Fig. 1), such as the various enzymes and mechanisms involved in synthesis, release, and metabolism of GABA, the transmembrane GABA transporters, and the GABA receptors ($GABA_A$, $GABA_B$, and $GABA_C$). Each of these entities presents a potential drug target, but so far only the GABA receptors, the GABA transporters, and the GABA-metabolizing enzyme GABA-T have proven therapeutically useful.

Although many drugs interacting with the GABA neurotransmitter system have been clinically helpful in the treatment of anxiety, epilepsy, sleep

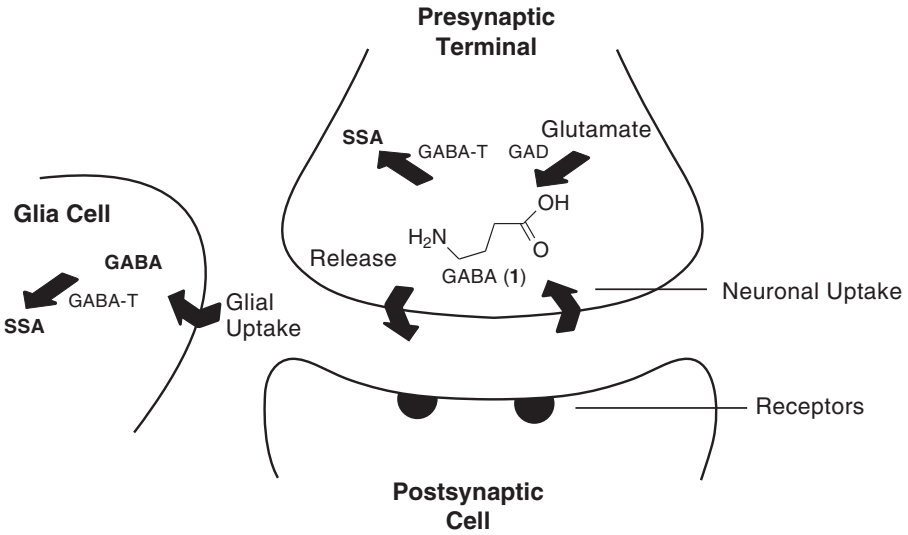


**Figure 1** Schematic illustration of the biochemical pathways, transport mechanisms, and receptors at a GABA-operated synapse. GAD: Glutamate decarboxylase; GABA-T: GABA transferase; SSA: Succinic acid semialdehyde.

disorders, alcohol withdrawal, and in the induction and maintenance of anesthesia, there are still problems that need to be solved. Therefore, major efforts are underway to improve the therapeutic spectrum of the presently available drugs, reduce their side effects, and discover new therapeutic tools and targets.

## GABA Biosynthesis and Metabolism

The main pathway of GABA synthesis is the decarboxylation of Glu, catalyzed by Glu decarboxylase (GAD), which employs pyridoxal-5′-phosphate (PLP) as a cofactor. GABA (**1**) is released into synaptic clefts by depolarization of presynaptic neurons and is rapidly removed from the synapses by reuptake into both glia and presynaptic nerve terminals, where it is initially catabolized by transamination into succinic semialdehyde. The major pathway of its degradation is via transamination with α-ketoglutarate, catalyzed by the PLP-dependent enzyme GABA-T. This enzymatic degradation takes place within GABAergic neurons, as well as in surrounding glia cells. It has been shown that inhibition of the activity of GABA-T results in an increase in available GABA, which makes it a therapeutic target in, for example, neurological disorders associated with abnormally low levels of GABA [13–15].

***Inhibitors of GABA Metabolism*** Based on the knowledge of the mechanism of inactivation of GABA-T, a number of inhibitors of this enzyme have been developed [16–20]. These compounds are typically analogs of GABA, containing appropriate, notably unsaturated, substituents at the γ-carbon adjacent to the γ-amino group. Via a deprotonation at the γ-carbon, the substituents are converted by GABA-T into electrophiles, which react with nucleophilic groups at or near the active site of the enzyme and thereby inactivate the enzyme irreversibly [19]. The receptor protein nucleophile that participates in this reaction is known to be Lys329, which links PLP to the active site [21].

Although GABA-T, like other PLP-dependent enzymes, does not show strict stereospecificity with respect to inactivation by mechanism-based inactivators, such inhibitors do react with the enzyme in a stereoselective manner. Thus the *S* forms of the GABA-T inhibitors Vigabatrin (**2**) and its fluoromethyl derivative (**3**) are more active as GABA-T inactivators than the respective *R*-isomers [22] (Fig. 2). These observations are in good agreement with results obtained from modeling studies of the inhibitors using the reported crystal structure of the enzyme [21]. The α-carboxylate of the inhibitor–PLP complex interacts with Arg192, which positions the *S*-proton of the γ-carbon appropriately for abstraction by Lys329.

The therapeutic potential of the GABA metabolic system is illustrated by Vigabatrin (**2**), which is utilized in the treatment of epilepsy [23], and is in clinical trials for treatment of drug addiction [24–26]. Unfortunately, the high risk of irreversible visual field defects, frequently associated with Vigabatrin

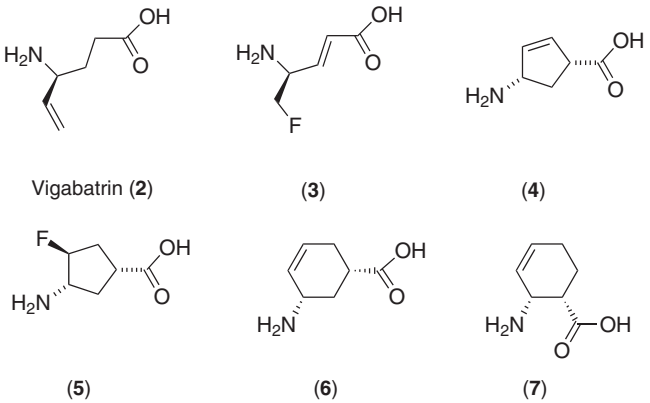


**Figure 2** Structures of some GABA-T inhibitors, compound **4** being inactive.

treatment, continues to cause concern [27, 28]. However, a recent study on rats has revealed that low-dose Vigabatrin maintains therapeutic efficacy for the treatment of cocaine addiction with reduced risk for emergence of visual defects [29].

In order to optimize the inactivator–enzyme interaction, conformationally rigid analogs of Vigabatrin and 4-amino-5-halopentanoic acid have been developed [20, 30]. Unlike the cyclopentene analog (**4**) of Vigabatrin, which does not inactivate GABA-T, the saturated 5-fluoro analog (**5**) exhibits inactivation properties similar to the corresponding open-chain analogs.

The apperance of the crystal structure of GABA-T [21] initiated the development of a series of cyclohexene analogs of Vigabatrin that possess slightly increased flexibility about the double bond [16, 17]. The analogs **6** and **7** irreversibly inhibit GABA-T, whereas the corresponding *trans*-isomers are competitive reversible inhibitors of the enzyme. Once bound, the proton of the *trans*-isomer, which needs to be removed for the reaction to take place, is not oriented toward the active site (Lys 329). Thus, no irreversible reaction takes place [17]. Interestingly, similar effects on drug addiction have been observed for competitive reversible GABA-T inhibitors [25, 31].

## GABA Reuptake

The reuptake of GABA into surrounding neurons and glial cells is affected by selective GABA transporters that belong to the superfamily of sodium- and chloride-dependent neurotransmitter transporters. Four different transporters have been identified displaying regionally distinct distributions within the CNS [32]. The four subtypes have been cloned from different species by different research groups leading to different acronyms, bringing the numbering of the transporters into confusion [32]. Whereas the homologous transporters cloned

from rat, mouse, and human have been numbered GAT-1, GAT1, and GAT-1, respectively, the mouse GAT3 and GAT4 correspond to the rat and human GAT-2 and GAT-3. The betaine/GABA transporter from humans (BGT-1) corresponds to the mouse GAT2. Based on studies of the regional distribution of GABA transporters, GAT1 seems to be the most abundant in the CNS, and it appears to be predominantly localized in neurons. GAT3 and GAT4 seem to be preferentally expressed in glial cells, while GAT2 is equally expressed in neurons and in glial cells [32–38].

Inhibition of GABA uptake has been shown to enhance the synaptic efficacy of GABA and to increase the extracellular concentration of GABA in the brain [39, 40]. The anticonvulsant effects seen in different seizure models following the administration of GABA uptake inhibitors substantiates the functional importance of these observations. In addition to their anticonvulsant activities, in vivo and in vitro studies indicate that GABAergic agents, including reuptake inhibitors, possess neuroprotective [41, 42], cognition-enhancing [43], and pain-relieving properties [44].

It has been established that the transport systems show different substrate specificities, which makes selective pharmacological manipulation of the GABA transport system possible [45]. For the purpose of stimulating GABA neurotransmission, the most effective strategy for such therapeutic intervention seems to be a selective blockade of glial GABA uptake, in order to elevate the level of GABA taken up by the neuronal carrier.

***Inhibitors of GABA Reuptake*** The classical GABA transport inhibitors nipecotic acid (**8**) and guvacine (**9**) are effective inhibitors of both neuronal and glial GABA uptake [46–49] (Fig. 3). Introducing small substituents at the amino group of nipecotic acid (**8**) or guvacine (**9**) results in compounds with reduced affinity for GABA transport carriers [50], whereas large lipophilic derivatives such as *N*-(4,4-diphenyl-3-butenyl)nipecotic acid (**12**) are much more potent than the parent compounds [51]. By contrast with nipecotic acid (**8**) and guvacine (**9**), such lipophilic compounds are able to cross the blood–brain barrier (BBB) and have been shown to be potent anticonvulsants in animal models. This development has led to the compound Tiagabine (**13**), now marketed as a therapeutic adjunct for the treatment of epilepsy [23]. Due to nonoptimal pharmacokinetic profile of Tiagabine (**13**) [52] a large number of lipophilic analogs of nipecotic acid (**8**) have been synthesized, that include an electronegative moiety in the carbon chain (e.g., compounds **14** and **15**) [53–57]. Although some of these analogs have turned out to be more potent than Tiagabine (**13**), a new therapeutic candidate apparently has not emerged from these studies.

Although effective as an anticonvulsant, Tiagabine (**13**) does not display glial selectivity and has been shown to be GAT-1 selective [58]. By the developing *exo*-THPO (**10**) and analogs thereof, a certain degree of selectivity for the glial transport system has been achieved [59]. *N*-Me-*exo*-THPO (**11**)

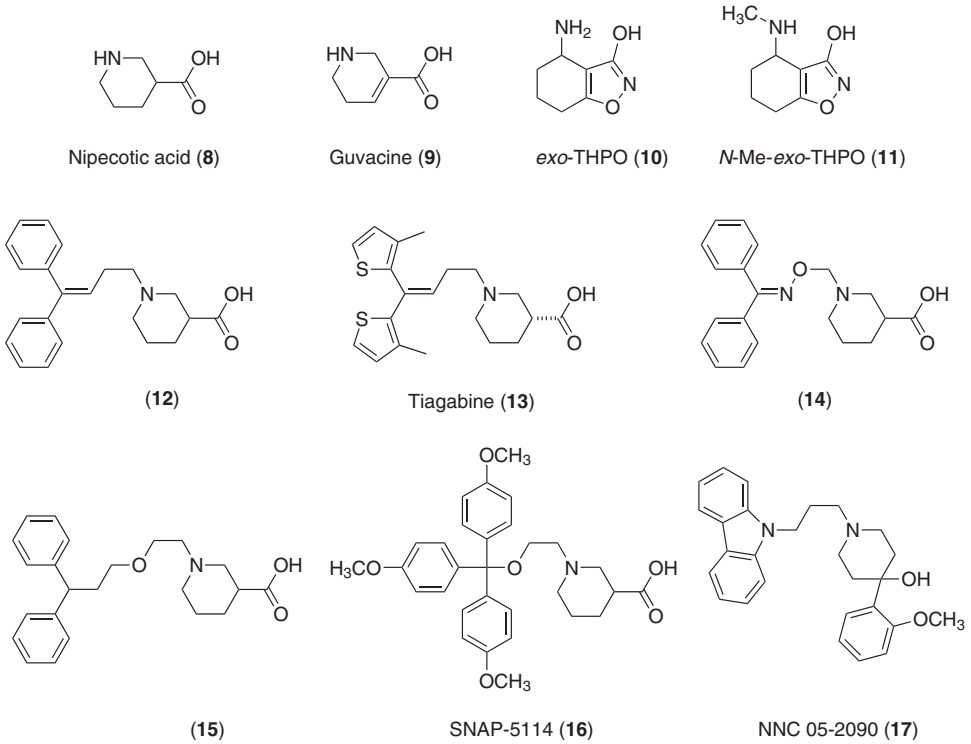


**Figure 3** Structures of some GABA uptake inhibitors.

exhibits more than 10-fold higher inhibitory potency for glial transport than for neuronal transport of GABA [59]. Although Tiagabine (**13**) is approximately 1000-fold more potent than *exo*-THPO (**10**) and its *N*-alkylated analogs as an inhibitor of GABA uptake in cultured neuronal cells and in GAT1 expressing cells, the anticonvulsant potencies of *exo*-THPO (**10**) and its analogs have been shown to be only 3 to 7-fold less than that of Tiagabine (**13**) [58]. Thus, the anticonvulsant efficacies of *exo*-THPO analogs seem to correlate better with their ability to inhibit glial rather than neuronal GABA transport [58]. Although glia selective, both *exo*-THPO analogs and liphophilic analogs of nipecotic acid and guvacine interact selectively with GAT1. Very few inhibitors of GABA transport have been shown to display selectivity for other GABA transporters (GAT2, -3, or -4). SNAP-5114 (**16**) [60, 61] has been reported to inhibit GAT2-4, and a group of compounds exemplified by NNC 05-2090 (**17**) seem to inhibit GAT2 and GAT4 [62]. Further development of subtype-selective compounds would be useful for studying the pharmacological and physiological importance of the GAT subtypes and may help understand the different anticonvulsant profiles observed for different inhibitors [63].

## The $GABA_A$ Receptor Complex

The $GABA_A$ receptor is a member of the superfamily of ligand-gated ion channels (Fig. 4). The structure and function of this group of receptors displays a high degree of complexity, which has been the subject of a number of excellent reviews over the last decade [5, 64–70]. The $GABA_A$ receptor is an oligomeric protein that forms a complex containing a considerable number of separate but allosterically interacting binding sites (Fig. 4*b*). This is reflected in the structural diversity of compounds acting at $GABA_A$ receptors, including important drugs such as benzodiazepines, barbiturates, and neurosteroids. The heterogenity of the $GABA_A$ receptor in the brain is pronounced. The exact number of different receptor subtypes is not known, but so far a total of 19 different subunits belonging to several subunit classes have been cloned ($\alpha$1-6, $\beta$1-3, $\gamma$1-3, $\delta$, $\varepsilon$, $\theta$, $\pi$, and $\rho$1-3). $GABA_A$ receptors are built up as pentameric assemblies of different families of receptor subunits, as illustrated in Figure 4*a*. It is believed that the most abundant receptor subtype consists of two $\alpha_1$, two $\beta_2$, and one $\gamma_2$ subunit in a well-defined structural arrangement forming a pore/channel in the cell membrane that is permeable to chloride ions.

Site-directed mutagenesis studies have shown that the binding site(s) for GABA, $GABA_A$ agonists, and competitive antagonists are located at the interface between $\alpha$ and $\beta$ subunits, whereas the binding site(s) for benzodiazepines are located at the interface between $\alpha$ and $\gamma$ subunits.

The $GABA_A$ receptors display different pharmacology and regional distribution depending on their subunit combination suggesting that $GABA_A$ receptor ligands with subunit-specific effects may provide drugs with improved clinical utility.

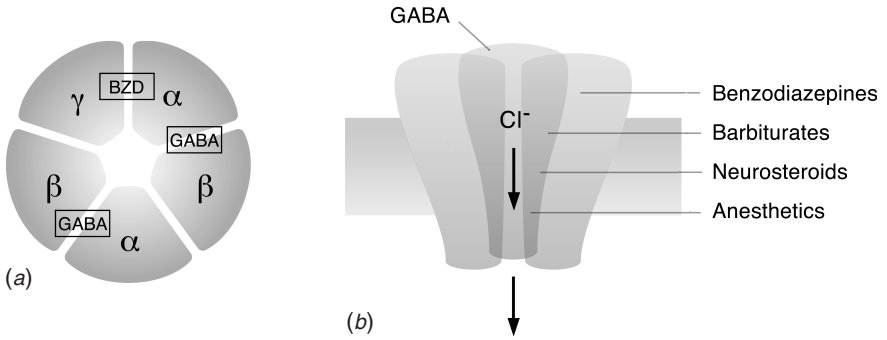


**Figure 4** (*a*) Schematic model of the pentameric structure of the $GABA_A$ receptor complex indicating the position of the binding sites for GABA and benzodiazepines (BZD). (*b*) Schematic illustration of the $GABA_A$ receptor complex indicating the chloride ion channel and additional binding sites.

***The GABA$_A$ Receptor Agonists and Antagonists*** In spite of the very distinct and specific structural requirements for GABA$_A$ receptor agonists, conformational restriction of different parts of the GABA (**1**) molecule and bioisosteric replacement of the carboxyl group have led to a series of specific GABA$_A$ agonists. Some of these compounds have played a key role in the pharmacological characterization of GABA$_A$ receptors.

Muscimol (**18**) (Fig. 5), a constituent of the mushroom *Amanita muscaria*, has been highly useful in the investigation of the GABA$_A$ receptors [48]. However, the fact that muscimol (**18**) is nonselective, being both a GABA$_A$ receptor agonist [71, 72] and a substrate for GABA-T [73], and moreover a neurotoxin, prevents its therapeutic application. Because the 3-isoxazolol moiety has proved to be efficient as a bioisosteric replacement for the carboxyl group of GABA, muscimol (**18**) has been extensively used as a lead structure in the development of GABA$_A$ receptor ligands.

The conformationallly restricted bicylic analog THIP (**19**), a specific GABA$_A$ agonist [74], has been shown to be devoid of the neurotoxic properties of muscimol (**18**) and metabolically stable, in contrast to muscimol (**18**). THIP (**19**) shows nonopioid analgesic effects [75] and is in clinical trials as a novel hypnotic agent. Unlike GABA, muscimol (**18**) and THIP (**19**) are capable of penetrating the BBB after systemic administration, in spite of their zwitterionic structures [76].

Based on the structure of THIP (**19**), a series of monoheterocyclic GABA$_A$ agonists has been developed, including isonipecotic acid (**20**) and isoguvacine (**21**) [74, 77]. THIP (**19**) and the equipotent analog isoguvacine (**21**) are now used as standard GABA$_A$ agonists.

Bicuculline (**22**) [78] (Fig. 6) and bicuculline methochloride (BMC) [79] are classical GABA$_A$ antagonists. SR 95531(**23**) belongs to a group of arylaminopyridazine analogs that show potent and selective competitive GABA$_A$ receptor antagonist effect [80, 81]. SR 95531(**23**) is now used as a standard GABA$_A$ antagonist.

The nonannulated THIP analogs, 4-PIOL (**24**) [82], the corresponding sulfur analog thio-4-PIOL (**25**) and **26** [83] represent a group of low-efficacy GABA$_A$ agonists that predominantly show GABA$_A$ antagonist profiles [83–85]. As reported for other GABA$_A$ antagonists, the functional consequences of the subunit composition of the GABA$_A$ receptors on thio-4-PIOL (**25**) pharmacology are negligible [86, 87]. This contrasts with the effects

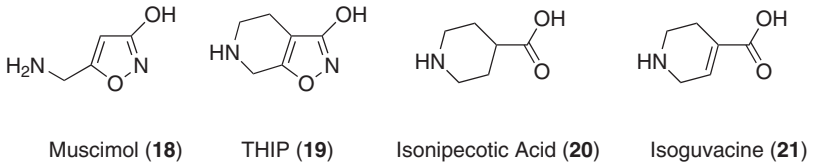


**Figure 5** Structures of some GABA$_A$ receptor agonists.

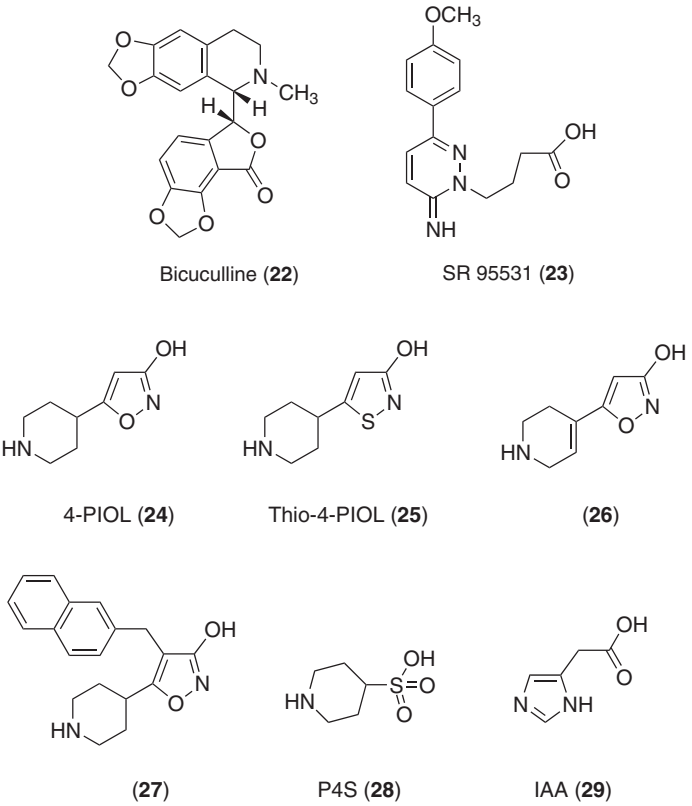


**Figure 6** Structures of some $GABA_A$ receptor antagonists (**22**, **23**, and **27**) and some partial $GABA_A$ receptor agonists (**24–26**, **28**, and **29**).

observed on the pharmacology of $GABA_A$ agonists [86]. Importantly, unlike full and highly efficacious partial $GABA_A$ agonists, repeated administration of 4-PIOL (**24**) has been reported to not cause significant desensitization of $GABA_A$ receptors [85, 88].

Based on the structure of 4-PIOL (**24**), a series of potent and selective competitive $GABA_A$ antagonists has been developed, as exemplified by compound **27**, with potencies in the low nanomolar range (see also the following section) [89]. Very little information has been obtained regarding direct acting full $GABA_A$ receptor agonist or antagonists in clinical studies. Therapeutic use of $GABA_A$ receptor agonists and antagonists that can interact with all $GABA_A$ receptor subtypes in the CNS may be associated with side effects, such as tolerance and convulsions. In this regard, partial agonists with various levels of efficacy, displaying either predominantly agonist or antagonist profiles, could be of therapeutic relevance, as exemplified by THIP.

THIP (**19**) and the heterocylic GABA isosteres piperidine-4-sulfonic acid (P4S, **28**) and imidazole-4-acetic acid (IAA, **29**) show the characteristics of

partial agonists [90]. Binding studies have shown only relatively low selectivity of $GABA_A$ agonists for $GABA_A$ receptors containing different α or β subunits [87]. By contrast, electrophysiological investigations using heteromeric ($\alpha_x\beta_3\gamma_2$) human $GABA_A$ receptors expressed in *Xenopus* oocytes have revealed that the relative efficacy as well as the potency displayed by these compounds is highly dependent on the receptor subunit combination [86]. The patterns of $GABA_A$ receptor subunit dependence of the maximal responses of THIP (**19**) and IAA (**29**) are qualitatively the same as that seen for P4S (**28**) (Fig. 7). Agonist potencies are frequently highest at receptors containing $\alpha_5$ or especially $\alpha_6$ subunits and generally lowest at $\alpha_4$-containing receptors. Coexpression of different α subunits in combination with β and γ subunits has been shown to result in unpredictable changes in both efficacy and potency of $GABA_A$ receptor agonists [88, 90, 91].

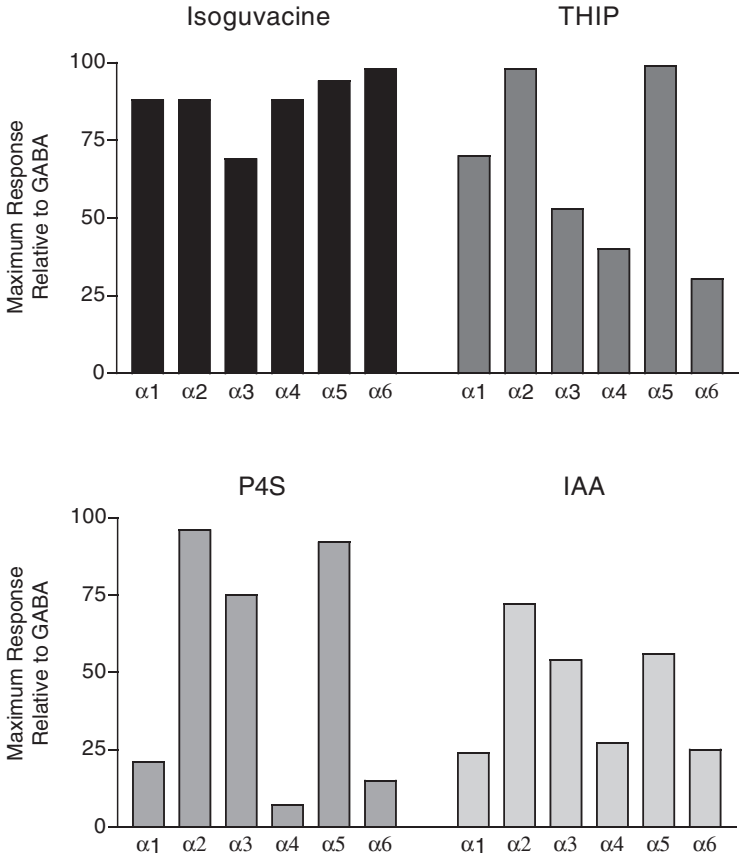


**Figure 7** Subunit-dependent efficacy of isoguvacine (**21**), THIP (**19**), P4S (**28**), and IAA (**29**) on $\alpha_x\beta_3\gamma_{2S}$ containing $GABA_A$ receptors expressed in oocytes. Maximum response at every receptor combination was determined using 3 mM GABA [86].

Another example of the consequences of the variation of subunits is the effect of substituting of a δ subunit for the $\gamma_2$ subunit in a receptor complex containing $\alpha_4\beta_3$ subunits. THIP (**19**) behaves as a partial agonist at $\alpha_4\beta_3\gamma_2$-containing receptors, while the compound behaves as a superagonist at $\alpha_4\beta_3\delta$-containing receptors, with a maximum response 60 percent larger than that of GABA, suggesting that GABA acts as a partial agonist at the $\alpha_4\beta_3\delta$ receptor subtype [92, 93]. Interestingly, δ-containing receptors show a 50-fold higher affinity for GABA than γ-containing receptors [94]. The $\alpha_4$ and the δ subunits are co-localized in the thalamus and hippocampus [95, 96], and some of these receptors are presumed to be extrasynaptic [97]. This restricted expression and putative extrasynaptic location makes them potentially interesting as drug targets.

It has been shown that the hypnotic effect of THIP (**19**) is elicited by a mechanism different from those of benzodiazepines and ethanol [98]. Most likely, this effect is mediated by benzodiazepine-insensitive $\alpha_4$-containing receptors. These receptor subtypes are located in regions distinctly different from the most abundant $\alpha_1$-containing receptors that are supposed to be responsible for the effects of currently used hypnotics [99]. These observations have led to the suggestion that THIP (**19**) has a major effect on $\alpha_4$- and δ-containing extrasynaptic receptors [98].

Despite the fact that the amino acid residues contributing to the binding site of GABA appear to be conserved within the α and β subunits [69], the functional consequences of binding are highly dependent on the subunit composition of individual $GABA_A$ receptors. Moreover, the pharmacological response seems to be determined not only by α and β subunits but by the interaction of all subunits present in the receptor complex. The determinants of these changes in pharmacological profiles that results in functional selectivity are still not clarified, making the development of compounds with a predefined receptor subtype-selectivity profile difficult.

***Receptor and Pharmacophore Models of the GABA Site of the $GABA_A$ Receptor Complex*** There is currently no experimentally determined three-dimensional structure of the $GABA_A$ receptor complex. Thus, the study of the structure–activity relationship (SAR) and the design of novel $GABA_A$ ligands cannot be based on experimental structural information about the receptor protein. The $GABA_A$ receptors belong to the same superfamily as the nicotinic acetylcholine receptors, and a soluble homopentameric acetylcholine binding protein (AChBP) from the snail *Lymnaea stagnalis* has been isolated [100]. This protein is weakly homologous to the extracellular ligand-binding domain of the nicotinic acetylcholine receptor. The protein has been successfully crystallized and its three-dimensional structure has been solved by X-ray crystallography [101]. The AChBP structure has been used for homology building of the extracellular ligand-binding domain of $GABA_A$ receptors [102–104]. The receptor models obtained may be used as tools for the design of mutagenesis experiments. However, it should be noted that the AChBP

protein only shares approximately 18 percent sequence identity with $GABA_A$ receptor subunits, making sequence alignment and homology modeling questionable. Thus, given the low level of sequence identity, it is not straightforward to use these models for the prediction of ligand affinity or SAR studies.

In the absence of an experimentally determined three-dimensional (3D) structure for the $GABA_A$ receptor, rationalization of SARs and design of novel ligands may be attempted by the use of pharmacophore models. Such a model is developed on the basis of analysis of the molecular properties of known ligands and their receptor binding data [105, 106]. A well-developed pharmacophore model may be employed to design new ligands and/or to search databases for new compounds that are compatible with the model [105, 106].

Most attempts to develop pharmacophore models for ligands binding to the GABA site of $GABA_A$ receptors have used the assumption that the 3-isoxazolol rings in, for example, muscimol (**18**), THIP (**19**), and 4-PIOL (**24**) (Figs. 5 and 6) are located in very similar regions in the binding cavity and thus may be superimposed in a pharmacophore model. Furthermore, it has been assumed that the relative spatial locations of the nitrogen atom and the 3-isoxazolol ring in THIP (**19**) define the bioactive conformation of muscimol (**18**). Based on these assumptions, a number of molecular modeling studies have been carried out in order to develop pharmacophore models for $GABA_A$ receptor agonists and competitive antagonists [107, 108].

However, it has not been possible to rationalize crucial experimental data by using models in which the 3-isoxazolol rings of muscimol (**18**), THIP (**19**), and 4-PIOL (**24**) are superimposed. For instance, it has been shown that replacement of the 3-isoxazolol ring in these molecules by an isothiazolol ring (e.g., compound **25**) has significantly different effects on the receptor affinity [72, 83, 109]. This indicates that the 3-isoxazolol rings in the three classes of compounds do not occupy identical positions in the receptor cavity. Furthermore, high-level quantum chemical ab initio calculations indicate that the assumed THIP-like bioactive conformation of muscimol has a very high conformational energy [110], which is not compatible with its high affinity. In addition, the introduction of a methyl group in the 4-position of the 3-isoxazolol ring of muscimol (**18**) severely inhibits interaction with the $GABA_A$ receptor recognition site [111]. This effect probably reflects steric repulsion between the methyl group and the receptor, which makes a direct superimposition of the 3-isoxazolol rings of muscimol (**18**) and THIP (**19**) highly questionable.

The equipotent activities of 4-PIOL (**24**) and compound **26** [83] clearly demonstrates that the distance between the nitrogen atom and the 3-isoxazolol ring in the bioactive conformation of 4-PIOL (**24**) is significantly larger than it is in THIP (**19**) and muscimol (**18**). A pharmacophore model that overcomes these problems has been developed [89, 112, 113] by noting that an arginine residue is a suitable binding partner to the carboxylate group of the

endogenous ligand GABA, as well as to the 3-isoxazolol anions of muscimol (**18**), THIP (**19**), and 4-PIOL (**24**). High-level quantum chemical ab initio calculations show that a bidentate interaction between the arginine side chain and the ligand is compatible with observed SARs [114]. The involvement of an arginine residue in the binding of muscimol (**18**) has been demonstrated by site-directed mutagenesis studies [115–117]. The flexibility of the arginine side chain makes it feasible to accommodate muscimol (**18**) and 4-PIOL (**24**) in the same binding pocket, by assuming that the side chain adopts different conformations when binding to the two ligands, as illustrated in Figure 8.

Figure 9 displays the deduced binding modes of muscimol (**18**), THIP (**19**), and 4-PIOL (**24**). It should be noted that the 3-isoxazolol rings of the three compounds do not overlap, which is in agreement with experimental evidence (see above). The black tetrahedrons in Figure 9 denote sterically "forbidden" regions. The pharmacophore model indicates that the 4-position in 4-PIOL (**24**) (marked by an arrow) does not correspond to the "forbidden" 4-position in muscimol (**18**) (see above). Thus, by contrast with muscimol

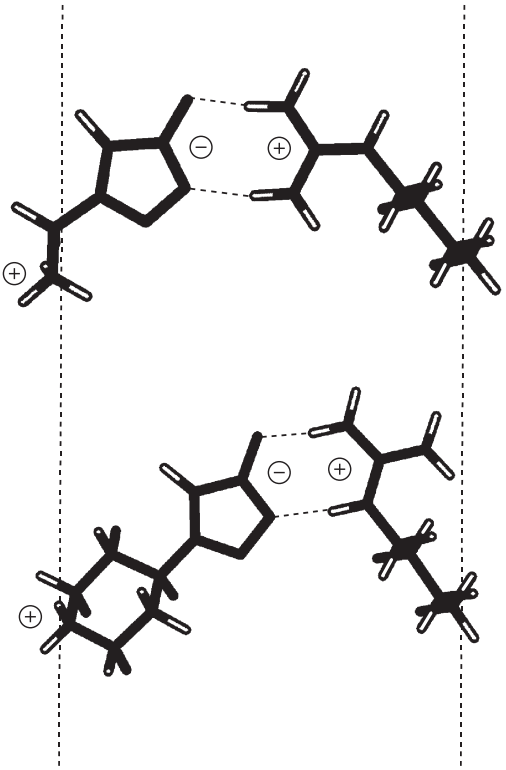

**Figure 8** Proposed binding mode of muscimol (**18**) and 4-PIOL (**24**) to the $GABA_A$ receptor employing two different conformations of an arginine side chain.

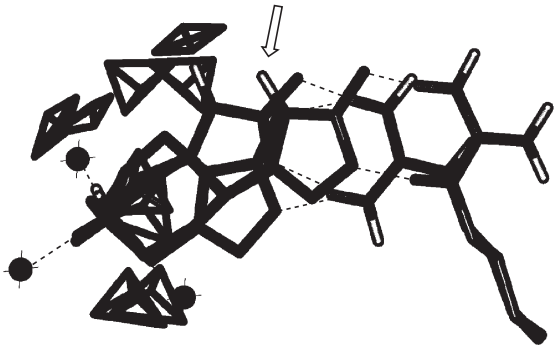

**Figure 9** Pharmacophore model of $GABA_A$ receptor agonists showing the proposed binding modes of muscimol (**18**), 4-PIOL (**24**), and THIP (**19**). Dashed lines indicate hydrogen bond interactions. The cyan-colored tetrahedrons indicate positions of methyl groups in $GABA_A$ agonists that cause strong steric repulsion with the receptor. The arrow points at the 4-isoxazolyl position in 4-PIOL (**24**). The magenta-colored spheres indicate sites with which the ammonium group in muscimol (**18**) interacts via hydrogen bonds. Hydrogens on carbon have been omitted for clarity, except for hydrogens at the 4-isoxazolyl position of muscimol and 4-PIOL.

derivatives, substitution at the 4-position of 4-PIOL (**24**) may be allowed according to the model. This prediction has resulted in the successful design of a number of 4-substituted 4-PIOL analogs with alkyl, phenylalkyl, diphenylalkyl, and naphthylalkyl substituents, demonstrating the presence of a receptor pocket of considerable dimensions in the vicinity of the 4-position of 4-PIOL (**24**) [89, 113].

Compound **27** displays the highest affinity of the 4-substituted 4-PIOLs with an increase in affinity by a factor of about 200 compared to 4-PIOL (**24**). Compound **27**, fitted to the pharmacophore model, is shown in Figure 10.

***Ligands for the Benzodiazepine Site and Other Allosteric Sites of the $GABA_A$ Receptor Complex*** The $GABA_A$ receptor complex is the target of a large number of structurally diverse compounds, some of which are pharmacologically active and used clinically. Agents that increase chloride ion flux directly or indirectly through the $GABA_A$ receptor have sedative, anxiolytic, anaesthetic, hypnotic, and/or anticonvulsive action. Compounds that decrease chloride ion current may induce convulsions, anxiety, or increase cognitive functions. These modulatory effects are produced by compounds such as benzodiazepines, ethanol, general anesthetics, barbiturates, and neuroactive steroids via a wide range of distinct allosteric binding sites within the pentameric, proteinous receptor complex [5, 69, 118, 119].

The benzodiazepine binding site was defined in 1977 as the binding site for 1,4-benzodiazepines [e.g., diazepam (**30**) (Fig. 11)] in a rat membrane suspension. Since then a wide variety of structurally diverse nonbenzodiazepine

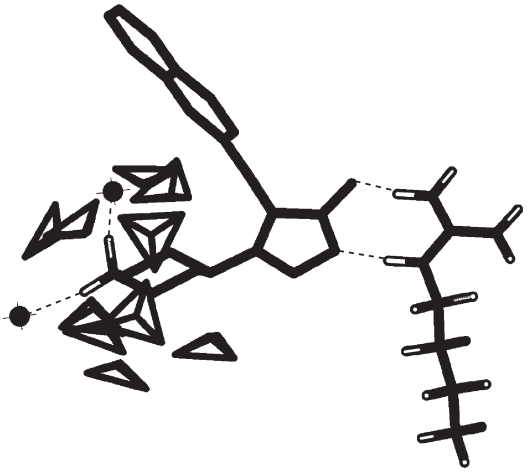

**Figure 10** Compound **27** fitted to the $GABA_A$ pharmacophore model. Hydrogens on carbon have been omitted for clarity.

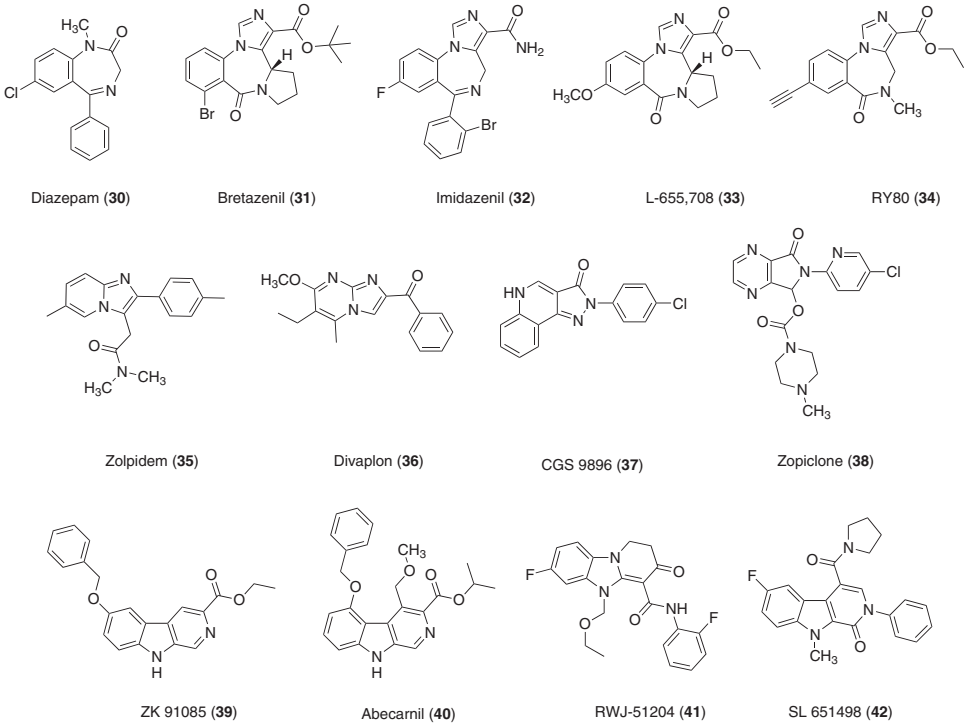


**Figure 11** Structures of some ligands for the benzodiazepine binding site.

compounds, as illustrated in Figures 11 and 12, have been found to interact with this binding site. These classes of substances include: imidazopyridines [Zolpidem (**35**)], imidazopyrimidines [Divaplon (**36**)], pyrazoloquinolinones [CGS 9896 (**37**)], cyclopyrrolones [Zopiclone (**38**)], β-carbolines [Abecarnil (**40**)], pyrazolopyrimidine [Zaleplon (**44**)], and triazolopyridazines [L-838,417 (**45**)] [67, 120]. The benzodiazepine binding site lies within the $GABA_A$ receptor complex, through which the various ligands exert their effects.

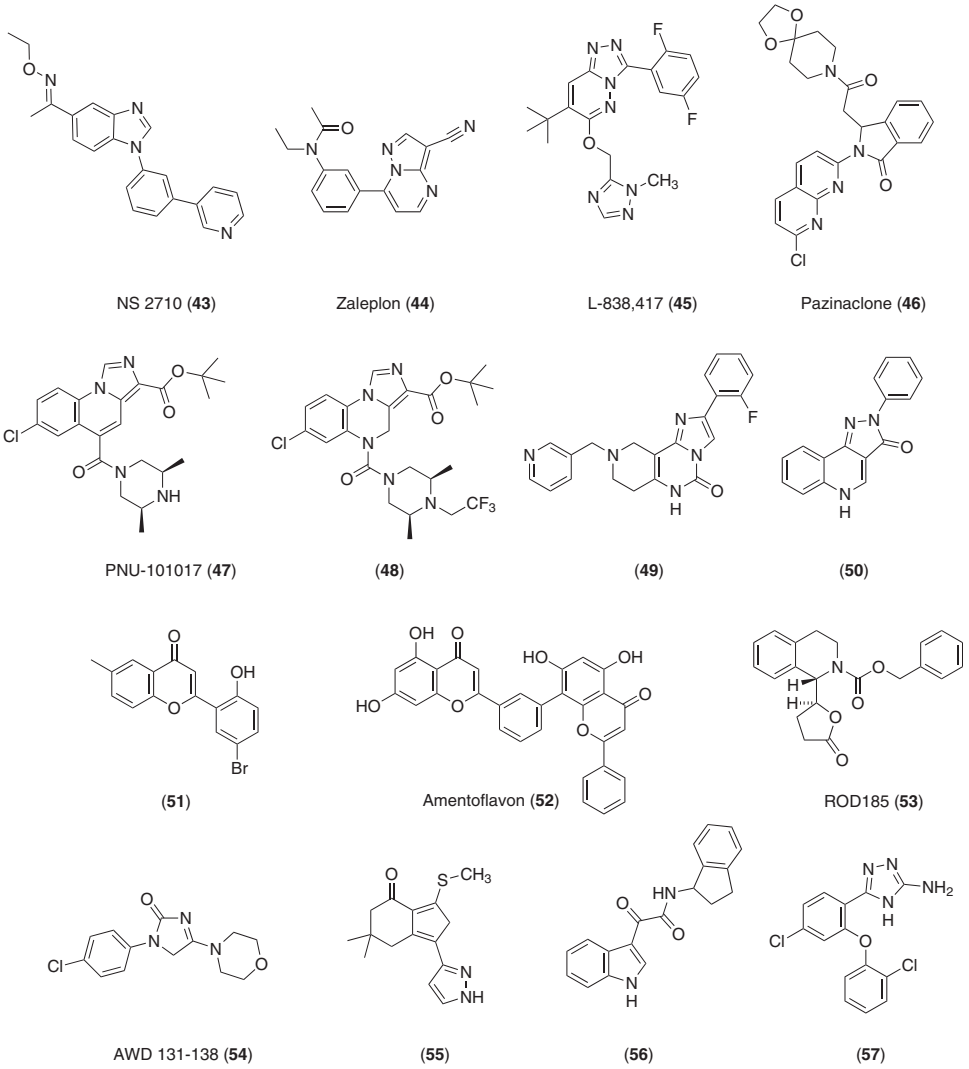


**Figure 12** Structures of some ligands for the benzodiazepine binding site.

Knowledge of the multiplicity and complexity of the benzodiazepine binding sites has increased dramatically over time, from one site, to two sites, to an unknown number of different regionally distinct subtypes in the mammalian brain. The complexity has been further enhanced by the discovery of the agonist–antagonist–inverse agonist principle [121].

Studies of benzodiazepine sites at $GABA_A$ receptor complexes are performed using (a) radioactive labeled ligands in vitro or in vivo (see, e.g., [122–124]); (b) by fluorescence based methods, including recombinant $GABA_A$ receptor complexes, which have been made cation ($Ca^{2+}$) permeable via specific point mutations in subunits [92, 125, 126]; (c) by $^{36}Cl$-flow measurements [127]; or (d) by electrophysiological techniques. Genetically modified mice (so-called knockin mice) have been used to differentiate the various pharmacological effects of agonists on the benzodiazepine binding site [99, 128]. The biological materials used for assaying benzodiazepine binding sites include: brain membrane preparations and brain slices; recombinant receptors [human or rat complementary deoxyribonucleic acid (cDNA) or messenger ribonucleic acid (mRNA)]; any combination of $\alpha_x$ ($x = 1, 2, 3, 5$), $\beta_y$ ($y = 1, 2, 3$), and $\gamma_2$ expressed in *Xenopus* oocytes, HEK 293 cells, CHO cells, $S_f9$ cells, or by construction of stable $GABA_A$ receptor-expressing cell lines.

The clinical effects of nonselective benzodiazepines such as diazepam (**30**) include sedation, muscle relaxation, anxiolysis, and memory impairment. Furthermore, the drugs can induce tolerance and dependence. Using the above-mentioned tools for studying the $GABA_A$ receptor complex, the benzodiazepine pharmacology of different $GABA_A$ subtypes is emerging, and the observed clinical effects of, for instances, diazepam (**30**) are being partly dissected. The $\alpha$ subunit seems to be a determinant for the pharmacology displayed by benzodiazepines. The sedative effect is suggested to be mediated primarily by $\alpha_1$-containing $GABA_A$ receptors, whereas $\alpha_2$-containing receptors seem to be important for different anxiolytic effects [99, 128–130]. The very distinct distribution of $\alpha_5$-containing $GABA_A$ receptors, being limited primarily to the hippocampus, which is known to be of importance in learning and memory, makes this group of receptors interesting as therapeutic targets. These observations have been used in the development of novel compounds with more specific actions.

In the following discussion selected ligands for the benzodiazepine site at the $GABA_A$ receptor complex are listed (see Figs. 11 and 12):

*Bretazenil* (**31**) and *Imidazenil* (**32**) are high-affinity partial agonists of the benzodiazepine binding sites (nonselective) [131]. *L-655,708* (**33**) is a partial inverse agonist. The substance shows higher affinity for $\alpha_5$-containing receptors than for $\alpha_1$-containing receptors [132]. *RY 80* (**34**) shows about 50 times higher affinity for $\alpha_5$-containing receptors than for $\alpha_1$-, $\alpha_4$-, or $\alpha_6$-containing receptors [133]. *ZK 91085* (**39**) is a β-carboline derivative with high affinity for the benzodiazepine binding sites. The substance stimulates $GABA_A$ receptor function both via the benzodiazepine site and the Loreclezole site (see below)

[134]. *RWJ-51204* (**41**) is described as a high-affinity partial agonist. The compound shows more anxiolytic activity and less sedation in mice and monkeys but not in rats [135]. *SL 651498* (**42**) shows stronger binding affinity for $\alpha_1$- and $\alpha_2$-containing receptors compared to $\alpha_5$-containing receptors. The compound is a full agonist at $\alpha_2$- and $\alpha_3$-containing receptors, while it has a partial agonist profile at $\alpha_1$- and $\alpha_5$-containing receptors [136].

*NS 2710 (ME 3127)* (**43**) is described as possessing functional subtype selectivity and anxiolytic activity [137]. *Zaleplon* (**44**) is a nonbenzodiazepine hypnotic drug. It has a lower $EC_{50}$ value in potentiating GABA effects at $\alpha_1$-containing receptors compared to $\alpha_2$-, $\alpha_3$-, and $\alpha_5$-containing receptors [138]. *L-838,417* (**45**) has anxiolytic action. The efficacy of the substance shows subtype selectivity: enhanced responses at $\alpha_2$-, $\alpha_3$-, and $\alpha_5$-containing receptors, while no modulation of the GABA response is found at $\alpha_1$-containing receptors [99]. The substance is devoid of the sedative and amnesic effects usually described for agonists of the benzodiazepine binding site. *Pazinaclone (DN 2327)* (**46**) is reported to be a partial agonist, although the substance shows the same adverse effects as full agonists [139]. *PNU-101017* (**47**) has high affinity for the benzodiazepine binding sites and shows partial agonist-like inhibition of induced convulsions in mice, as well as a weak anxiolytic effect. Besides the benzodiazepine binding site, the compound may act via a second low-affinity site at $GABA_A$ receptors producing a negative allosteric effect, whereas the trifluoroethyl analog **48**, produces a positive allosteric effect [140, 141]. *Derivatives of imidazopyridopyrimidine* (e.g., **49**) show high affinity and some selectivity toward $GABA_A$ receptor subtypes. One derivative is reported to be an $\alpha_2/\alpha_3$-selective partial agonist showing in vivo activity [142]. *Derivatives of pyrazoloquinolinones* (e.g., **50**) have been found in several studies to be very potent ligands for benzodiazepine sites [143]. Naturally occurring and synthetic *flavone derivatives* (**51**, **52**) are inhibitors of benzodiazepine binding in vitro [144, 145]. Amentoflavone (**52**), a biflavonoid, was identified as the first nonnitrogenous substance showing high affinity for benzodiazepine binding sites [146]. *ROD 185* (**53**) is reported to bind to benzodiazepine sites as well as to a novel allosteric modulatory site at the $GABA_A$ receptor complex [147]. *AWD 131–138* (**54**) is a low-affinity partial agonist. The compound has anticonvulsant and anxiolytic properties in rodents [148]. *Derivatives of benzothiophenone* (e.g., **55**) show higher affinity for $\alpha_5$-containing receptors compared to $\alpha_1$-, $\alpha_2$-, and $\alpha_3$-containing receptors. Some analogs have inverse agonist activity at $\alpha_5$-containing receptors; enhancement of cognitive performance is seen in rat behavioral models [149]. *Derivatives of N-(indol-3-yl glyoxyl)amine* (e.g., **56**) bind with higher affinity to $\alpha_1$-containing receptors than for $\alpha_2$-, $\alpha_3$-, and $\alpha_5$-containing receptors [150, 151]. *Derivatives of phenyltriazole* (e.g., **57**) show anticonvulsant and sedative effects, although less potent than diazepam in animal studies [152].

Apart from the benzodiazepine binding site, a number of $GABA_A$ sites exist that recognize specific chemical patterns. In the following discussion, selected

ligands are listed for some of these nonbenzodiazepine allosteric sites at the $GABA_A$ receptor complex (see Fig. 13):

*Barbiturates* [e.g., pentobarbital (**58**)], *neurosteroids*, and *neuroactive steroids* [e.g., α-THDOC (**59**) and alphaxolone (**60**)] are selective allosteric modulators of $GABA_A$ receptors [68, 153–155]. Although many actions of the barbiturates and the steroids are similar, they are believed to interact with distinct binding sites within the $GABA_A$ receptor complex. *Anesthetics* generally have low affinity for $GABA_A$ receptor complexes, but defined amino acid mutations in α and β subunits have been found to be critical for the action of general anesthetics [156]. As an example Enflurane (**61**) potentiates GABA currents in $\alpha_3\beta_3\gamma_2$ subunit containing receptors [157]. *Loreclezole* (**62**) is an anticonvulsive compound that potentiates $GABA_A$ receptor function, depending on the β subunit: $\beta_2$- or $\beta_3$-containing receptors have much higher affinity for Loreclezole (**62**) compared to $\beta_1$-containing receptors [158]. α-*EMTBL* (**63**) has no affinity for the benzodiazepine binding sites or the GABA binding site; it shows a profile similar to that of Loreclezole (**62**) [159]. *Etomidate* (**64**) is an anxiolytic compound that potentiates GABA-induced currents depending on the type of β subunit present in the receptor complex, as seen for Loreclezole (**62**) [160, 161]. *Etifoxin* (**65**) is an anxiolytic substance, which potentiates GABA-induced currents, depending on the type of β subunit present in the receptor complex as seen for Loreclezole (**62**) [162]. *Mefenamic acid* (**66**) has anticonvulsant and proconvulsant activity. The substance potentiates $GABA_A$ receptor function in vitro. It exhibits a profile comparable to Loreclezole (**62**) [163]. *MDL 26,479 (Suritozole)* (**67**) is reported to be active in animal models that are predictive of cognitive enhancement and antidepressant activity. The compound has been suggested to act at the $GABA_A$ receptor complex at a site different from the benzodiazepine binding site [164]. *Tracazolate* (**68**) possesses anxiolytic and anticonvulsant activity. The compound potentiates GABA-induced currents in recombinant receptors [165].

*Propofol* (**69**) activates $GABA_A$ receptors in the absence of GABA [166]. *SB-205384* (**70**) potentiates GABA-induced currents. The substance prolongs the half-life of the decay of current after GABA removal. Its effect is selective for $\alpha_3$-containing receptors [167]. *Retigabine* (**71**) is an antiepileptic drug. The substance interacts allosterically with the $GABA_A$ receptor complex, possibly via a novel site coupled to the GABA binding site. The compound is also a potassium channel opener [168]. *Fluoxetine* (**72**) is reported to affect $GABA_A$ receptor function via a novel modulatory site. The substance increases the response to subthreshold concentrations of GABA but does not alter the maximum response. The $\alpha_5$-containing receptors show less response compared to other α-containing receptors [169]. *Clozapine* (**73**) is an antipsychotic drug. The compound is reported to inhibit GABA-evoked chloride ion currents at $\alpha_1\beta_2\gamma_{2L}$ $GABA_A$ receptors expressed in *Xenopus* oocytes [170].

A wide range of different ligands are known to block the chloride channel [2]. These compounds could be defined as negative allosteric modulators of

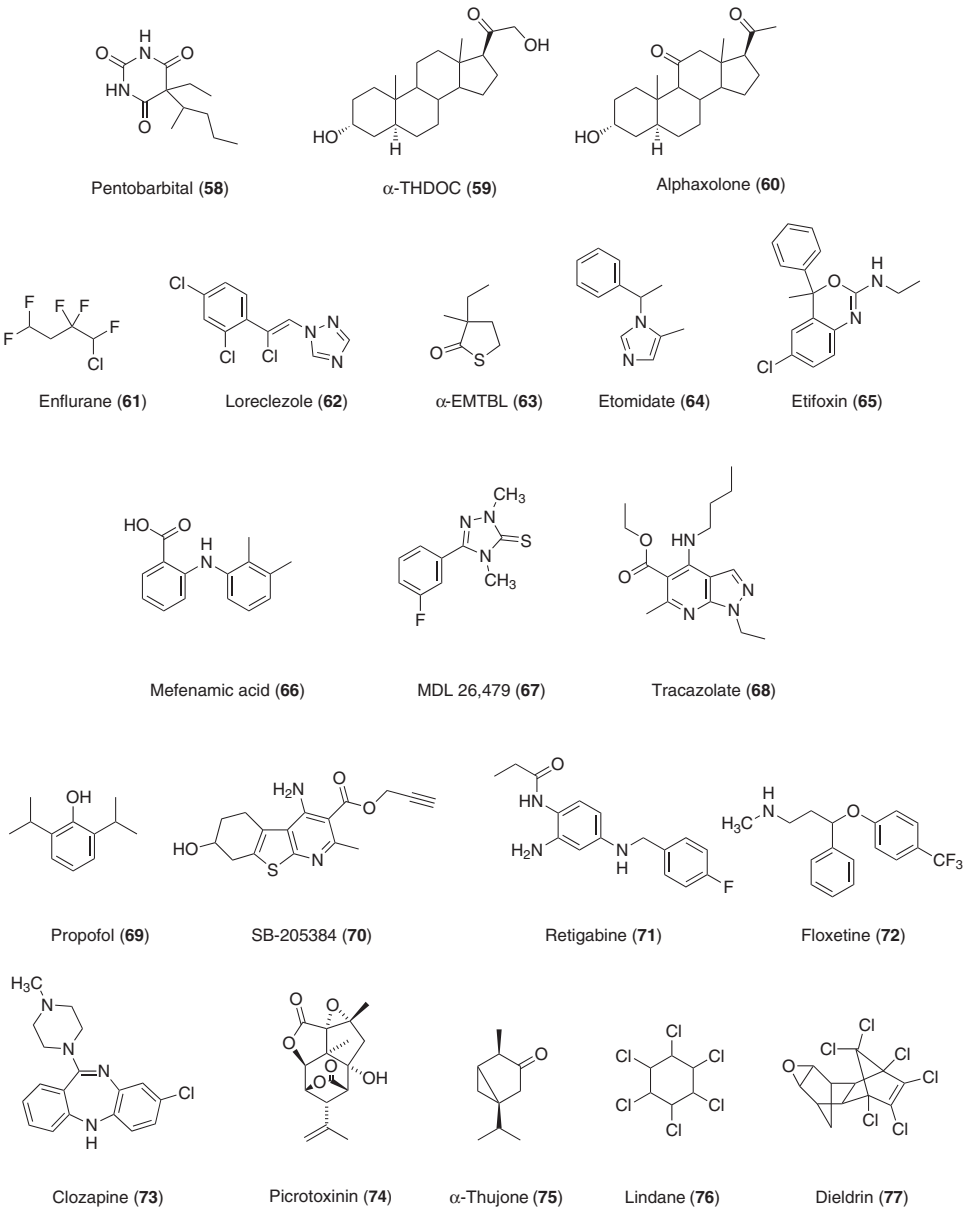


**Figure 13** Structures of some ligands for nonbenzodiazepine allosteric binding sites at the $GABA_A$ receptor.

$GABA_A$ receptors (or noncompetitive $GABA_A$ receptor antagonists). *Picrotoxinin* (**74**) is a potent convulsant. α-*Thujone* (**75**), which is an ingredient of absinthe, has been found to interact with the $GABA_A$ receptor complex via the picrotoxin binding site [171]. *Lindane* (**76**) and *Dieldrin* (**77**) are insecticides known to inhibit $GABA_A$ receptor function [172].

***Receptor and Pharmacophore Models of the Benzodiazepine Site of the $GABA_A$ Receptor Complex*** The X-ray crystallographic structure of a soluble homopentameric AChBP [101] described in the previous section has been employed via homology models in attempts to obtain structural information about the benzodiazepine binding site [102, 103]. The amino acid residues forming the benzodiazepine binding site have been predicted, and a mechanism for the mediation of positive cooperativity at the benzodiazepine site has been proposed [103]. As mentioned earlier the low level of sequence identity between AChBP and the corresponding parts of $GABA_A$ receptors necessitates extensive validation of models before any conclusions regarding their usefulness can be drawn.

A number of pharmacophore models of ligand binding at the benzodiazepine site of the $GABA_A$ receptor complex have been reported [144, 173–179]. The most comprehensive and well-validated pharmacophore model reported so far is that of Cook and co-workers [174, 175]. The model is based on 136 different ligands from 10 structurally different classes of compounds. In developing the model it was assumed that agonists, antagonists, and inverse agonists share the same binding pocket. The elements of the pharmacophore model are shown in Figure 14. H1 and A2 denote hydrogen bond donor and acceptor sites, respectively, whereas H2/A3 is a bifunctional hydrogen bond donor/acceptor site. L1, L2, L3, and $L_{Di}$ are lipophilic regions and S1, S2, and S3 are regions of steric repulsive ligand-receptor interactions (receptor-essential volumes). As shown in Figure 14, 1,4-benzodiazepines such as diazepam (**30**) are proposed to bind in a different binding mode than the more planar ligands such as the pyrazoloquinolone CGS 9896 (**37**). Diazepam and CGS 9896 both interact with H1, H2/A3, L1, and L2 sites, but the L3 site is only occupied by 1,4-benzodiazepines, and the $L_{Di}$ site only by compounds similar to CGS 9896.

This model has been widely used for the design of novel ligands [180, 181] and for SAR analyses [182–184]. The validity and usefulness of the model have been further demonstrated by an extensive study of flavone derivatives [144, 178]. It should be noted that flavones were not included in the original development of the model. As shown in Figure 15, these studies have resulted in an extension of the model built by Cook and co-workers [174, 175], by the identification of two new regions of steric ligand–receptor repulsions, S4 and S5, and the proposal of a large region (denoted "channel" in Fig. 15) in which a wide variety of substituents may be accommodated. This region is proposed to be partly lipophilic, with possibilities for hydrogen bonding. The extension of this region may be a channel in which larger substituents are only partly

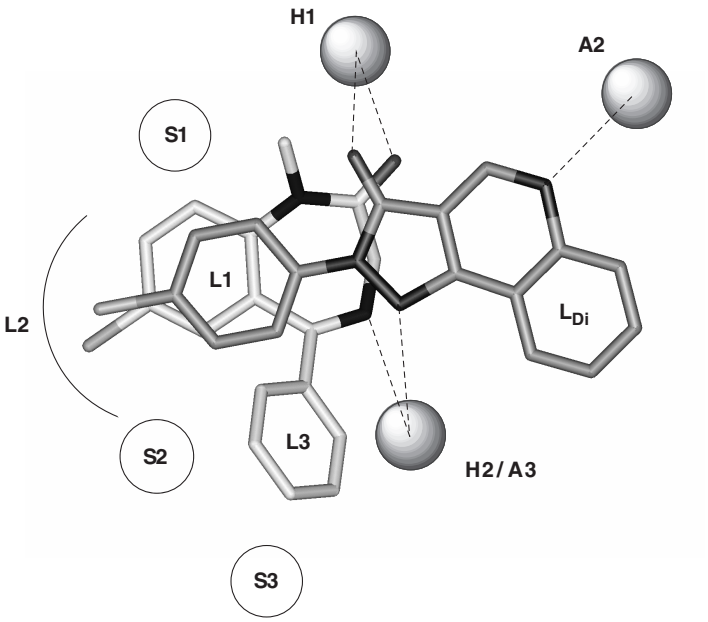


**Figure 14** Pharmacophore model developed by Cook and co-workers displaying the proposed binding modes of diazepam (**30**) and CGS 9896 (**37**). Dashed lines indicate hydrogen bonds. Hydrogens on carbon are omitted for clarity.

desolvated. Such a channel may exist at the interface between an α and a γ subunit in the $GABA_A$ receptor where the benzodiazepine binding site is most probably located.

By using the model, a number of high-affinity flavone derivatives have been designed. The highest affinity flavonoid (**51**) is shown fitted to the model in Figure 15 ($K_i$ = 0.9 nM, [$^3$H]-Ro 15–1788 binding to rat cortical membrane preparations) [178]. Flavone itself displays an affinity of 4200 nM. Thus, the use of the pharmacophore model has resulted in an increase in the affinity by a factor of 4700 via the introduction of three properly positioned substituents.

The pharmacophore model displayed in Figure 15 has been converted into a search query and used for database searching with the Catalyst software suite (Accelrys, Inc.). A number of hits have been identified, and these may be used as interesting starting points for the design of novel ligands for benzodiazepine sites [185].

Current developments of pharmacophore models for benzodiazepine sites focus on models for specific receptor subtypes [133, 186, 187]. Most of this work has been performed in the context of the pharmacophore model developed by Cook and co-workers [174, 175], described above. As more data for subtype-selective compounds become available, such pharmacophore models will undoubtedly be further developed to a level where they can be effectively

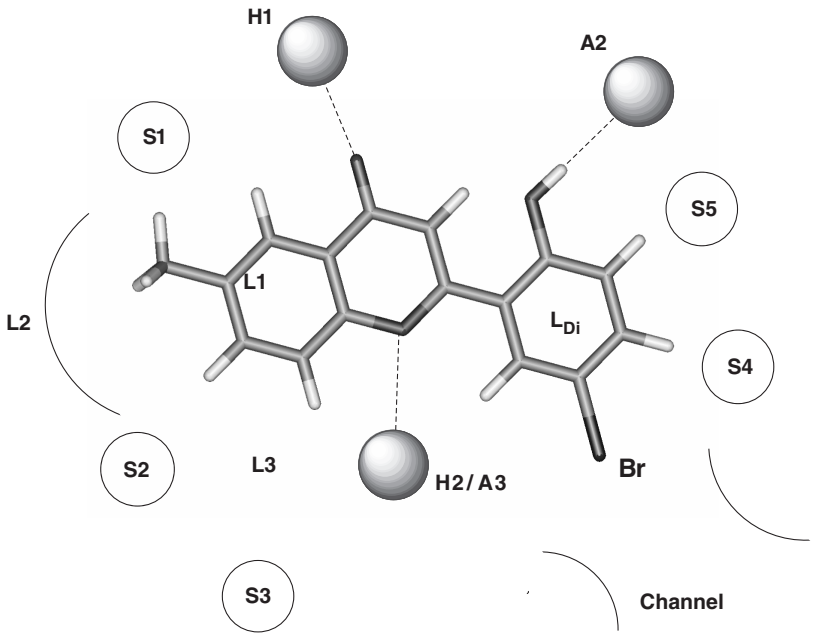


**Figure 15** Proposed binding mode of 5′-bromo-2′-hydroxy-6-methylflavone (**51**) in the extended pharmacophore model. Dashed lines indicate hydrogen bonds.

used for the design of novel subtype-selective ligands. It may also be within reach to employ pharmacophore models to elucidate the molecular properties involved in the actions of agonism and inverse agonism [174, 188, 189].

## $GABA_B$ Receptor Ligands

***The $GABA_B$ Receptor*** $GABA_B$ receptors were first identified in the early 1980s on the basis of pharmacological responses to the agonist baclofen (**78**) and insensitivity to the $GABA_A$ antagonist bicuculline (**22**) (Fig. 6) [190, 191], but resisted cloning until the late 1990s. $GABA_B$ receptors belong to the family of G-protein-coupled receptors (GPCR), which upon activation cause a decrease in calcium and an increase in potassium membrane conductance and inhibit cyclic adenosine 5′-monophosphate (cAMP) formation. The response is thus inhibitory and leads to, for example, hyperpolarization and decreased neurotransmitter release [192]. Specifically, the receptor belongs to the family C of GPCRs, with similarity to the mGluRs (see Metabotropic glutamate Receptor Ligands), which are characterized by a large extracellular amino-terminal domain (ATD) and a 7-transmembrane (7TM) domain (Fig. 16) [193].

Cloning revealed that the receptors consist of two subtypes, $GABA_{B(1)}$ and $GABA_{B(2)}$, which form a heterodimeric receptor through a coiled-coil inter-

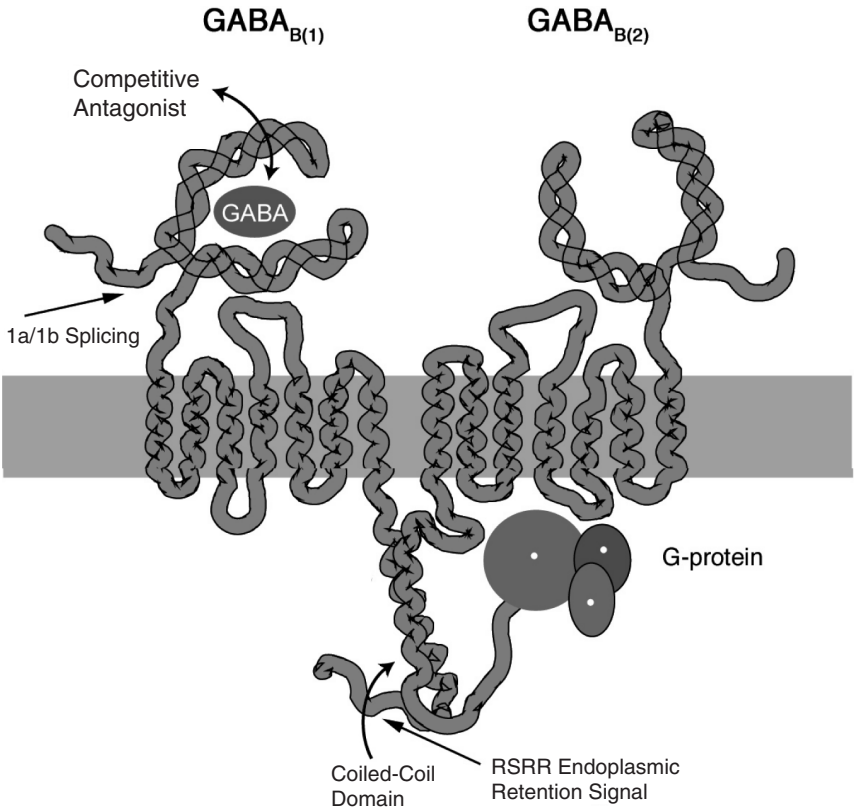


**Figure 16** Schematic illustration of the $GABA_{B(1)}$ and $GABA_{B(2)}$ subunits forming the heterodimeric $GABA_B$ receptor. GABA (**1**) binds in the cleft of the amino-terminal domain of $GABA_{B(1)}$, which leads to G-protein activation through the $GABA_{B(2)}$ subunit. The localization of the coiled-coil domain, the RSRR endoplasmic retention signal and the $GABA_{B(1a)/(1b)}$ splice site has been noted. Adapted from [194].

action in the intracellular carboxy-terminus (Fig. 16) [195–198]. The roles of the individual subunits in the receptor complex are just beginning to emerge. $GABA_{B(1)}$ contains the orthosteric ligand-binding site [199, 200] and is retained in the endoplasmatic reticulum (ER) in the absence of $GABA_{B(2)}$ [195–198]. It has been shown that $GABA_{B(1)}$ features an RSRR motif adjacent to the α-helical coil, which is responsible for ER retention by an as yet unknown mechanism [201–203]. Formation of the coiled-coil interaction by heteromerization appears to mask the RSRR motif, allowing release of the $GABA_{B(1)}$ subunit from the ER, and surface trafficking of the heteromeric complex. However, $GABA_{B(2)}$ is not merely a chaperone that ensures proper trafficking of the receptor complex to the plasma membrane. Data from several studies indicates that this subunit is mainly responsible for coupling to the G-protein [204–208].

In spite of intense research, the $GABA_{B(1,2)}$ heterodimer remains the only functional $GABA_B$ receptor described so far. Since it has become virtually axiomatic that neurotransmitter receptor systems are composed of pharmacologically distinct subgroups, it was expected that $GABA_B$ receptors would form a family of related genes. Therefore, the discovery that $GABA_B$ receptors arise from the expression of only two genes comes as a surprise and significantly limits the therapeutic potential of $GABA_B$ selective ligands. This is despite the fact that pharmacological differences have been observed in vivo and in vitro, which suggested the existence of multiple receptor subtypes [192].

***$GABA_B$ Receptor Agonists*** As already mentioned, the $GABA_B$ receptor is selectively activated by the GABA (**1**) analog baclofen (**78**), of which the *R*-form is the active enantiomer. Baclofen (**78**) was developed as a lipophilic derivative of GABA (**1**), in an attempt to enhance the BBB penetrability of the endogenous ligand. Since 1972, racemic baclofen (**78**) has been marketed for the treatment of spasticity due to its muscle-relaxant effects [209]. This action appears largely due to a reduction in neurotransmitter release onto motoneurons in the ventral horn of the spinal cord. $GABA_B$ receptor agonists also represent attractive drug targets in the pharmacotherapy of various neurological and psychiatric disorders, including neuropathic pain, anxiety, depression, absence epilepsy, and drug addiction. However, the (side) effects of agonists, principally sedation, tolerance, and muscle relaxation, limit their utility for the treatment of many of these diseases. Unfortunately, the lack of $GABA_B$ receptor subtypes makes the therapeutic prospects of agonists rather dim.

Although a number of $GABA_B$ receptor agonists have been synthesized, the number of selective and potent agonists for the $GABA_B$ receptor is limited, and (*R*)-baclofen still remains one of the more potent and selective agonists for the $GABA_B$ receptor. So far, the most successful strategy toward increasing potency has been bioisosteric replacement of the carboxyl group of GABA (**1**) with a phosphinic acid moiety. Thus, the phosphinic acid analog of GABA (**1**), CGP27492 (**79**), is the most potent $GABA_B$ agonist reported to date, being approximately 10-fold more potent than GABA (**1**) [210].

***Competitive $GABA_B$ Receptor Antagonists*** The first $GABA_B$ receptor antagonists to be reported were phaclofen (**80**) and saclofen (**81**) (Fig. 17), the respective phosphonic acid and sulfonic acid bioisosteric analogs of baclofen (**78**) [211, 212]. 2-OH-Saclofen (**82**) is the most potent antagonist obtained using these scaffolds [213], however, and with potencies in the mid-micromolar range there was room for improvement. Furthermore, these compounds were unable to penetrate the BBB, which made in vivo studies difficult.

In an attempt to improve the pharmacology of the $GABA_B$ agonist CGP27492 (**79**), a new series of selective and potent $GABA_B$ antagonists,

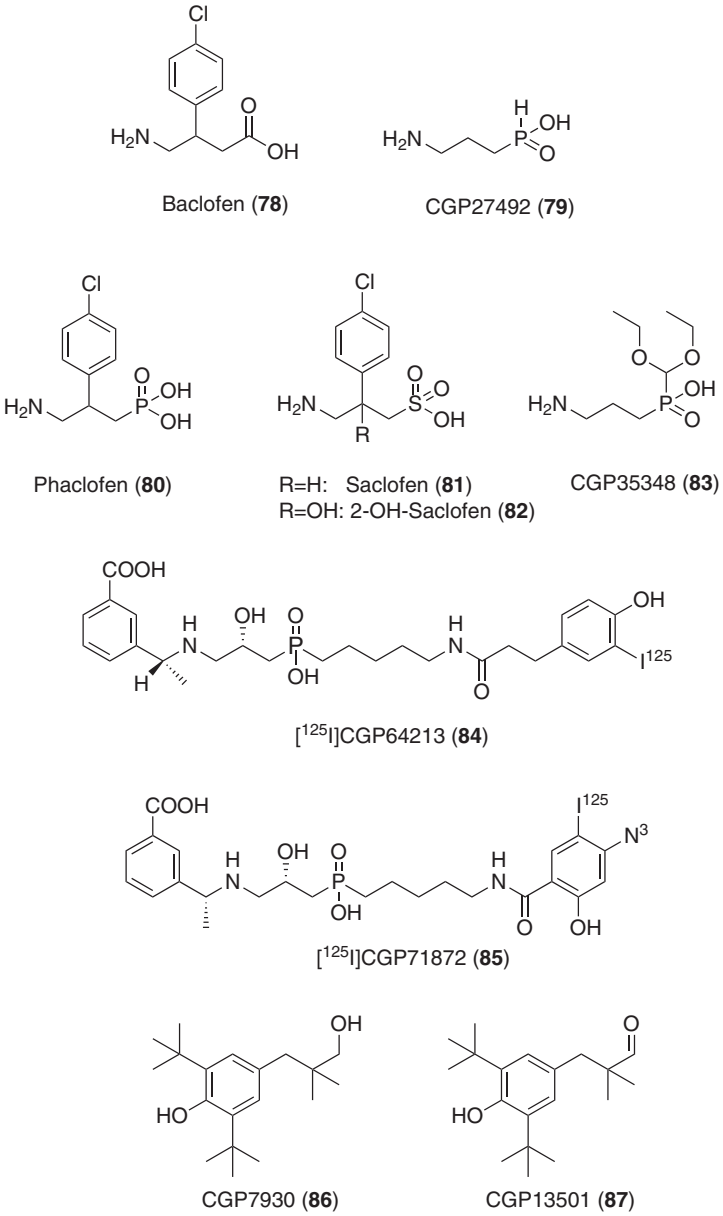


**Figure 17** Chemical structures of ligands used to characterize and clone $GABA_B$ receptors. The recently identified positive allosteric modulators CGP7930 (**86**) and CGP13501 (**87**) are expected to broaden the spectrum of therapeutic applications for $GABA_B$ drugs.

capable of penetrating the BBB after systemic administration, were discovered [214]. It was found that substituents larger than methyl conferred antagonism, compounds such as CGP35348 (**83**) displaying low-micromolar potency. However, it was not possible to break the nanomolar barrier with this series of compounds, which was finally achieved by substituting the amino group of the γ-aminopropylphosphinic acids with branched benzyl substituents. In this series, many sub- to low-nanomolar compounds have been reported, including the radioligands [$^{125}$I]CGP64213 (**84**) and [$^{125}$I]CGP71872 (**85**), which were crucial tools for the functional cloning of the receptor [215].

The development of antagonists capable of penetrating the BBB has allowed in vivo analysis of the therapeutic potentials of such ligands. Accordingly, it has been suggested, based on animal studies, that $GABA_B$ antagonists could improve cognitive processes and be beneficial against epilepsy and depression. However, further studies are needed to validate these effects, and again severe side effects seem difficult to avoid due to the absence of receptor subtypes.

***Positive Allosteric Modulators of $GABA_B$ Receptors*** $Ca^{2+}$ allosterically modulates the potency of GABA (**1**) but not baclofen (**78**), causing a potentiation of GABA (**1**) responses [216–218]. The $EC_{50}$ for $Ca^{2+}$ potentiation is 37 μM, and it is thus likely that the $Ca^{2+}$ site on the $GABA_B$ receptors is saturated under normal physiological conditions [217]. However, it is possible that $Ca^{2+}$ modulation plays a role in pathophysiological situations characterized by a significant drop in extracellular $Ca^{2+}$. A series of substituted phenols with positive allosteric modulator effect has been described [219]. These compounds, CGP7930 (**86**) and CGP13501 (**87**) (Fig. 17), are small organic molecules and thus attractive from a drug development perspective. Like $Ca^{2+}$, these compounds do not activate the receptor directly but increase the potency and affinity of classical $GABA_B$ agonists including GABA (**1**) [219]. The compounds do not displace binding of radiolabeled orthosteric ligands and therefore do not bind to the GABA binding site. So far the allosteric binding site has not been localized and characterized, but certain positive allosteric modulators at mGluR1 and calcium-sensing receptors have been shown to bind within the 7TM domain [220, 221]. This suggests that CGP7930 (**86**) and CGP13501 (**87**) also bind to the 7TM domain of $GABA_{B(1)}$ and/or $GABA_{B(2)}$, but other binding sites have not been ruled out [222]. Due to their dependence on GABA (**1**) for activity, the novel allosteric modulators are expected to show an improved side effect profile compared to $GABA_B$ agonists, which act indiscriminately on $GABA_B$ receptors. Unlike agonists, these drugs should not induce tolerance because they do not cause prolonged receptor activation leading to desensitization and internalization. If allosteric modulators exhibit the therapeutic potential of baclofen (**78**), while being devoid of major side effects and muscle relaxant properties, they will considerably broaden the spectrum of therapeutic applications of $GABA_B$ ligands.

## $GABA_C$ Receptor Agonists and Antagonists

Shortly after the identification of the $GABA_B$ receptor, a third class of GABA receptors, the $GABA_C$ receptors, was discovered [223]. These receptors are not blocked by bicuculline (**22**) nor modulated by benzodiazepines or barbiturates, which typically affect $GABA_A$ receptors [223, 224]. By contrast with $GABA_B$ receptors, $GABA_C$ receptors are not activated by baclofen (**78**) or inhibited by phaclofen (**80**) [223].

$GABA_C$ receptors are the most incompletely characterized of the GABA receptors, but in the past decade major progress toward understanding the molecular nature of $GABA_C$ receptors has been made [225, 226]. To date three different subtypes of $GABA_C$ receptors have been identified in humans ($\rho_{1-3}$) [227–229], where the $\rho_1$ subunit is predominantly located in the retina, the $\rho_2$ subunit is found in most brain regions [230], and the $\rho_3$ subunit is found in the hippocampus [231]. The membrane topology is assumed to be very similar to that of $GABA_A$ receptors, but $GABA_C$ receptors are believed to consist of only ρ subunits ($\rho_{1-3}$). The three isoforms of the ρ subunits can assemble into homomeric or pseudoheteromeric chloride channels showing a pharmacology different from the $GABA_A$ receptors [68, 232].

CACA (**88**) has been the key compound in identifying the $GABA_C$ receptors [223, 233] (Fig. 18). This compound is a moderately potent partial $GABA_C$ agonist and is inactive at $GABA_A$ receptors. The *trans*-isomer TACA (**89**) shows no preference for neither of the two receptor classes [233]. CAMP (**90**) is a selective $GABA_C$ agonist that is inactive at $GABA_A$ receptors [234]. TPMPA (**91**) was the first selective antagonist for $GABA_C$ receptors to be developed and is at least 100 times more potent as an antagonist at $GABA_C$ receptors than at $GABA_A$ receptors [235, 236]. Recently, the isosterically

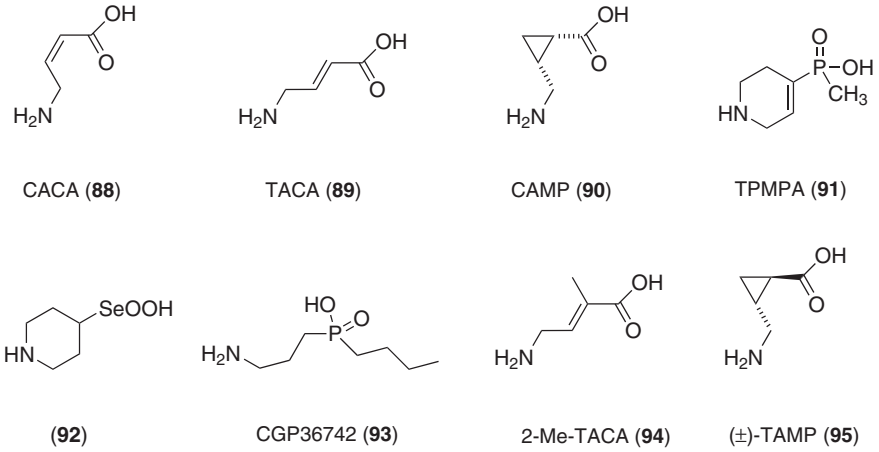


**Figure 18** Structures of some ligands for the $GABA_C$ receptor.

derived seleninic acid analog of isonipecotic acid, compound **92**, has been reported to be more potent than TPMPA (**91**) at $GABA_C$ $\rho_1$ receptors and to display a higher selectivity than TPMPA (**86**) for $GABA_C$ receptors relative to $GABA_A$ receptors [237].

Interestingly, the $GABA_C$ receptors share several ligands with the $GABA_A$ and $GABA_B$ receptors. Most $GABA_A$ agonists seem to have some agonist/antagonist action at $GABA_C$ receptors, exemplified by THIP (**19**) and P4S (**28**), which are partial agonists at $GABA_A$ receptors and competitive antagonists at $GABA_C$ receptors [238]. Muscimol (**2**), isoguvacine (**21**), and IAA (**29**) act as partial $GABA_C$ receptor agonists [238, 239]. Several phosphinic acid analogs of GABA, such as CGP36742 (**93**), have been identified as $GABA_C$ antagonists [240]. This group of compounds was originally developed as $GABA_B$ receptor antagonists [210, 214].

Certain ligands have been shown to act selectively or preferentially at specific ρ subunits. As an example, TPMPA (**91**) is eight times more potent at $\rho_1$ than at $\rho_2$ $GABA_C$ receptors expressed in *Xenopus* oocytes [241] and half as potent at recombinant $\rho_3$ $GABA_C$ receptors than at $\rho_1$ [242]. 2-Methyl-TACA (**94**) is a competitive antagonist at human $\rho_1$ $GABA_C$ receptors, a partial agonist at human $\rho_2$ $GABA_C$ receptors, but without effect at rat $\rho_3$ $GABA_C$ receptors [241], whereas (±)-TAMP (**95**) is a potent antagonist at rat $\rho_3$ $GABA_C$ receptors and a partial agonist at $\rho_1$ and $\rho_2$ $GABA_C$ receptors [242–244].

Based on studies in the retina, where $GABA_C$ receptors predominate, it has been established that this class of receptors plays a unique functional role in retinal signal processing [245, 246]. To clarify the functional relevance of the different $GABA_C$ receptor subtypes, the emerging pharmacological differences between recombinant $\rho_1$, $\rho_2$, and $\rho_3$ are of significant importance.

## 3 GLUTAMATE: EXCITATORY NEUROTRANSMITTER

### Classification of Glutamic Acid Receptors

Glu operates, as described in Section 1 of "Design and Pharmaceutical Applications of Prodrugs", via two main receptor classes, iGluRs and mGluRs, each consisting of three groups of receptor subtypes [1, 2]. The three heterogeneous groups of iGluRs are named after selective agonists, *N*-methyl-D-aspartic acid (NMDA), 2-amino-3-(3-hydroxy-5-methyl-4-isoxazolyl)propionic acid (AMPA), and kainic acid (KA) receptors. Native iGluRs assemble as homo- or heterotetramers, from subunits within each receptor type: NMDA receptors from NR1, NR2A-D, and NR3A,B; AMPA receptors from GluR1-4; and KA receptors from GluR5-7 and KA1,2 (Fig. 19) [247, 248]. At many central synapses, Glu activates a mixed population of NMDA and AMPA/KA receptors. Fast excitatory transmission is mainly mediated by AMPA/KA receptors, since the NMDA receptor ion channel is blocked by $Mg^{2+}$ at normal resting potential (see next section).

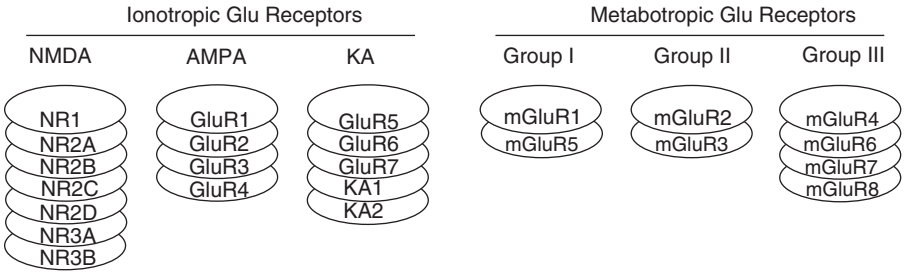


**Figure 19** Schematic illustration of Glu receptor multiplicity. Each circle represents a receptor subunit or a receptor subtype for the ionotropic and metabotropic receptors.

The mGluRs lead to slower and longer lasting neuronal effects and modulate the activity of iGluRs. mGluRs are divided into groups I, II, and III based on pharmacology, signal transduction mechanisms, and sequence homology [2, 248]. Group I mGluRs stimulate phospholipase C and include the cloned subtypes mGluR1,5. Groups II and III mGluRs both inhibit cAMP formation and consist of the subtypes mGluR2,3 and mGluR4,6-8, respectively (Fig. 19). mGluRs assemble as dimeric structures and probably as homomers [249]. It is now generally agreed that both iGluRs and mGluRs play important roles in the healthy as well as the diseased CNS and that all subtypes of these receptors are potential targets for therapeutic intervention in a number of diseases (see Section 1 of "Design and Pharmaceutical Applications of Prodrugs").

## NMDA Receptor Ligands

The NMDA receptors include a number of different binding sites and thus several potential targets for therapeutic attack. NMDA receptors are cation channels that flux $Na^+$, $K^+$, and $Ca^{2+}$, of which the latter is implicated in the neurotoxicity observed after excessive receptor stimulation. NMDA receptors are unique among ligand-gated ion channels in that they require two different agonists for activation, Glu (**96**) (Fig. 20) and glycine (**127**) (Fig. 23), and at the same time membrane depolarization in order to relieve a blockade by $Mg^{2+}$ [247, 248]. Polyamines such as spermine and spermidine modulate the activity of NMDA receptors in a biphasic manner [2, 249]. At micromolar concentrations polyamines enhance the activity of the ion channel, whereas at higher, probably nonphysiological, concentrations the ion channels are blocked in a voltage-dependent manner. Furthermore, zinc, redox modulators, and NO have been shown to inhibit the NMDA receptors via allosteric sites [250].

In the normal brain, NMDA receptors are fundamental to development and function because of their involvement in synaptic plasticity and neuronal signaling processes, including mechanisms of learning and memory. Furthermore,

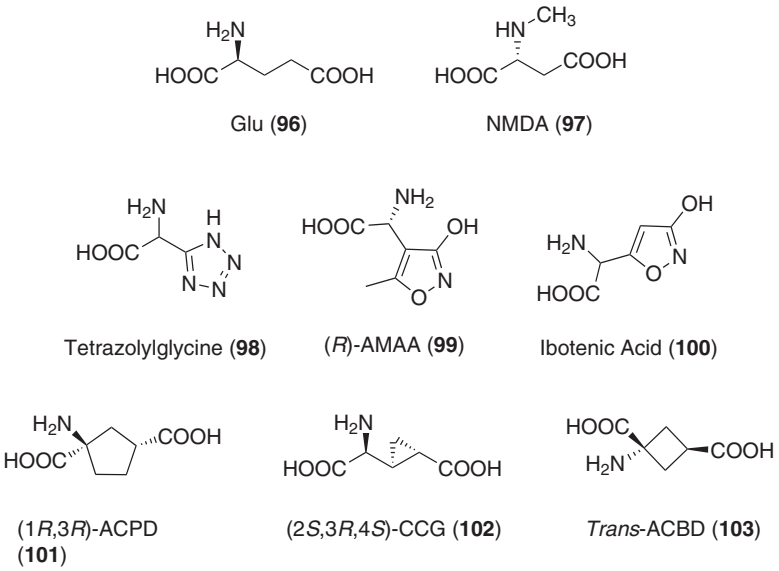


**Figure 20** Structures of Glu (**96**) and some NMDA receptor agonists.

NMDA receptor-induced neurotoxicity is intimately involved in a number of neuronal disorders as previously described (see Section 1 of "Design and Pharmaceutical Applications of Prodrugs").

Functional NMDA receptors are heteromers, typically consisting of NR1 and NR2 subunits [251, 252] and probably possessing a tetrameric structure [247], although this has not been thoroughly investigated as in the case of AMPA receptors (see below). The Glu binding site is located on NR2 subunits and the glycine binding site on NR1 subunits [251]. A combination of NR1 and NR3 subunits expressed in *Xenopus* oocytes form receptors that are activated by glycine alone [251], although the existence or physiological significance of such receptors in the CNS is not known.

Recently, it has been shown that NMDA receptors have a direct protein–protein interaction with the G-protein-coupled dopamine receptor subtype D1 [253]. The study of NMDA receptor signaling mechanisms and the search for therapeutic agents are still areas of active research. In the following sections ligands for the major sites located at the NMDA receptor complex will be described.

***NMDA Receptor Agonists*** NMDA receptors are unusual among Glu recognition sites in as much as the majority of the more potent ligands, both agonists and antagonists, posses an *R*-configuration about the α-amino acid center. NMDA (**97**) itself represents the only known agonist in which *N*-methylation does not lead to reduced affinity [254]. Other potent NMDA agonists have been developed, particularly by replacement of the distal acidic group and/or by conformational restriction of the three essential functional groups, namely

the α-amino group, the α-carboxyl group, and the ω-acidic moiety. Among the potent NMDA agonists with a different distal acidic group are tetrazolylglycine [255] (**98**) and (*R*)-AMAA [256] (**99**), exemplifying two widely used carboxyl group bioisosteric groups, the tetrazole and the 3-isoxazolol, respectively. (*R*)-AMAA (**99**) and other Glu ligands have been developed using the naturally occurring neurotoxin ibotenic acid (**100**) as a lead [257]. Ibotenic acid (**100**) is, apart from being a potent NMDA agonist, a potent agonist of some mGluR subtypes and a somewhat weaker agonist at other Glu receptor types.

Conformational restriction of the Glu backbone leading to Glu analogs with a folded conformation has led to a number of potent and selective NMDA agonists [258, 259]. (1*R*,3*R*)-ACPD (**101**) [260], (2*S*,3*R*,4*S*)-CCG (**102**) [261], and *trans*-ACBD (**103**) [262] are examples of such carbocyclic acidic amino acids, exemplifying that the higher activity does not always reside with the *R* form.

***Competitive NMDA Receptor Antagonists*** A large number of potent and selective competitive NMDA antagonists (Fig. 21) have been developed over the past two decades, and the availability of these compounds has greatly facilitated studies of the physiological and pathophysiological roles of NMDA receptors [2]. The distance between the two acidic groups in NMDA antagonists is typically one or three C–C bonds longer than in Glu. Many potent ligands have successfully been developed using ω-phosphonic acid analogs such as (*R*)-APV (**104**) as lead structures [263]. Combination of an ω-

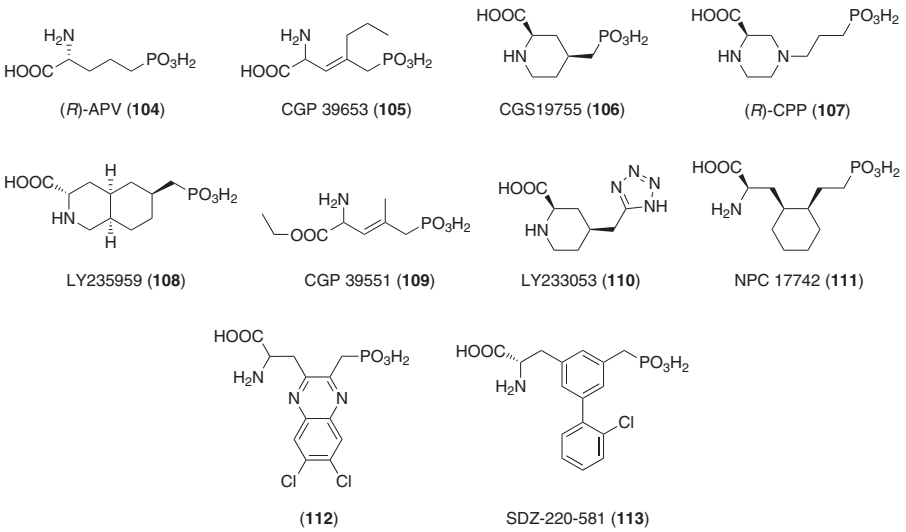


**Figure 21** Structures of some competitive NMDA receptor antagonists.

phosphonate group, a long carbon backbone, and conformational restriction has led to different series of potent antagonists.

Conformational restriction has been achieved by use of double bonds (CGP 39653 (**105**) [264]) and ring systems (CGS19755 (**106**) [265] and (*R*)-CPP (**107**) [266]), or bicyclic structures (LY235959 (**108**) [267]). These antagonists have shown very effective neuroprotective properties in various in vitro models. However, many of these compounds suffer from poor BBB penetration. The esterified analog CGP 39551 (**109**) [264], which is a prodrug, shows good oral bioavailability and anticonvulsant activity. LY233053 (**110**) [268] represents another class of antagonist with a tetrazole ring as the terminal acidic group. Substitution of the tetrazole for a phosphono group has limited effect on in vitro activity and shows improved bioavailability.

The carbocyclic analog NPC 17742 (**111**) [269] is active after systemic administration. The quinoxalin-alanine analog **112** [270] shows high antagonist potency with concomitant affinity for the glycine site. Related to the latter two analogs is a series of phenylalanine-based antagonists, represented by SDZ-220-581 (**113**) [271], and the ability of **113** to penetrate the BBB may be explained by increased lipophilicity, though active transport has been suggested [272, 273].

***Uncompetitive and Noncompetitive NMDA Receptor Antagonists*** The dissociative anesthetics PCP (**114**) and ketamine (**115**) block the NMDA receptor channel in a use-dependent manner [274] (Fig. 22). Thus, initial agonist activation of the channel is a prerequisite in order for such uncompetive antagonists to gain access to the binding site, which is situated within the ion channel. The antagonists eventually become trapped within the ion channel, and this may result in very slow kinetics. MK-801 [275] (**116**) has been developed as a very effective uncompetitive NMDA antagonist and has been extensively investigated to probe the therapeutic utility of such compounds, notably for the treatment of ischemic insults such as stroke [276]. MK-801 (**116**) and related high-affinity ligands have, however, shown severe side effects, including psychotomimetic effects, neuronal vacuolization, and impairment of learning and memory [277, 278]. Ligands with lower affinity, such as memantine (**117**) [279], dextromethorphan (**118**) [280], and remacemide (**119**) [281] have shown improved therapeutic indexes. Memantine (**117**) is being used for the treatment of AD and Parkinson's disease [279, 282] and may also have potential in the treatment of AIDS dementia [283]. The fast kinetics and low affinity of memantine (**117**) compared to MK-801 (**116**) may explain the absence of severe side effects. Remacemide (**119**) also shows a favorable side effect profile as compared to other uncompetitive antagonists and may have potential in the treatment of Huntington's chorea and Parkinson's disease [281, 284].

Two substituted aminoamides with low-affinity uncompetitive NMDA antagonist activities have been described, namely the ethynyl-substituted analog **120** [285] of the clinical antidepressant agent, milnacipran, and

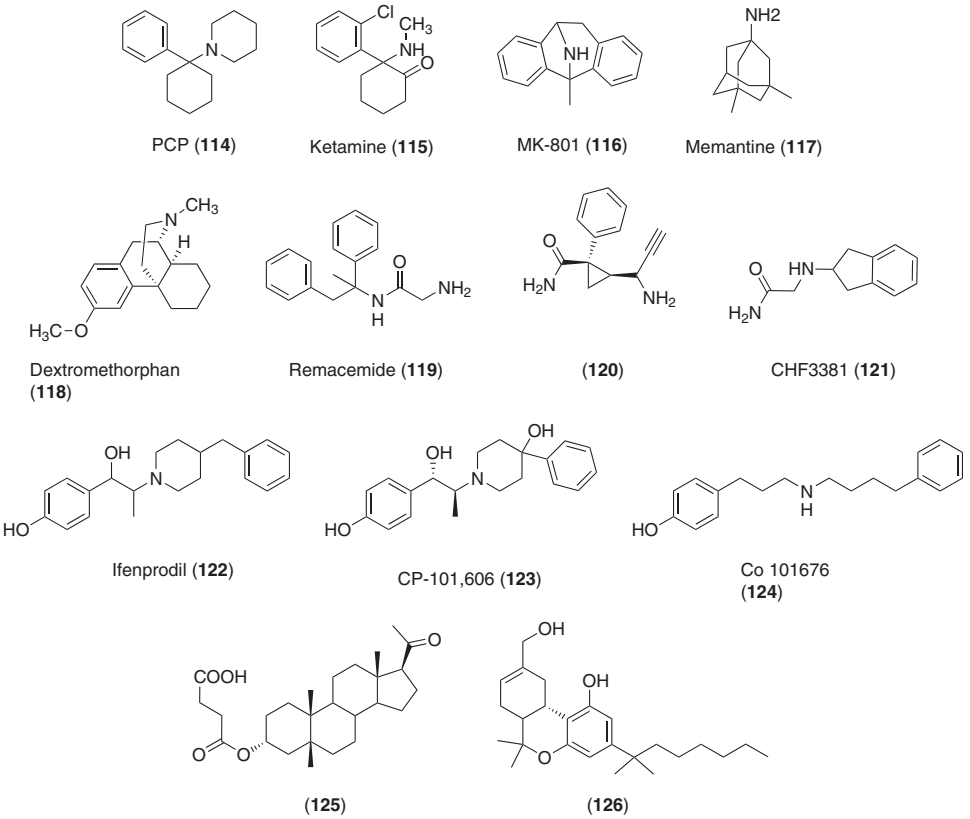


**Figure 22** Structures of some uncompetitive (**114–121**) and noncompetitive (**122–126**) NMDA receptor antagonists.

CHF3381 [286] (**121**). Compound **120** also is a potent serotonin uptake inhibitor, and like the parent compound milnacipran, which shows no serious side effects and good BBB penetration, these analogs may represent a new class of antidepressants [285]. CHF3381 (**121**) has been studied in animal models and shows potential as a neuroprotective agent with antiepileptic and antinociceptive activity [286–288].

Several compounds with noncompetitive activity at NMDA receptors have been described, and these compounds most likely do not bind to the same site as the uncompetitive ligands. Ifenprodil (**122**) and CP-101,606 (**123**) represent an important series of noncompetitive NMDA receptor antagonists. These compounds are active in ischemia models and as anticonvulsants and antinociceptive agents [289, 290]. Early clinical trials with these analogs have been disappointing due to unwanted effects. The side effect profiles do, however, seem to be significantly improved as compared to, for example, competitive NMDA antagonists [291]. Co 101676 (**124**) represents a series of structurally

simplified ifenprodil analogs [292]. The sterol hemisuccinate (**125**) has been shown to be a noncompetitive NMDA antagonist and to be protective against NMDA-induced seizures in mice [293]. The cannabinoid dexanabinol [294] (**126**) shows noncompetitive NMDA antagonism and significant neuroprotective effects.

***Glycine Co-agonist Site*** The excitatory co-agonist site for glycine (**127**) (Fig. 23) at the NMDA receptor is named the glycine$_B$ receptor. This receptor site is different from the inhibitory glycine receptors found primarily in the spinal cord of the mammalian CNS, where glycine activates strychnine-sensitive ionotropic receptors named glycine$_A$ receptors [295]. Glycine$_B$

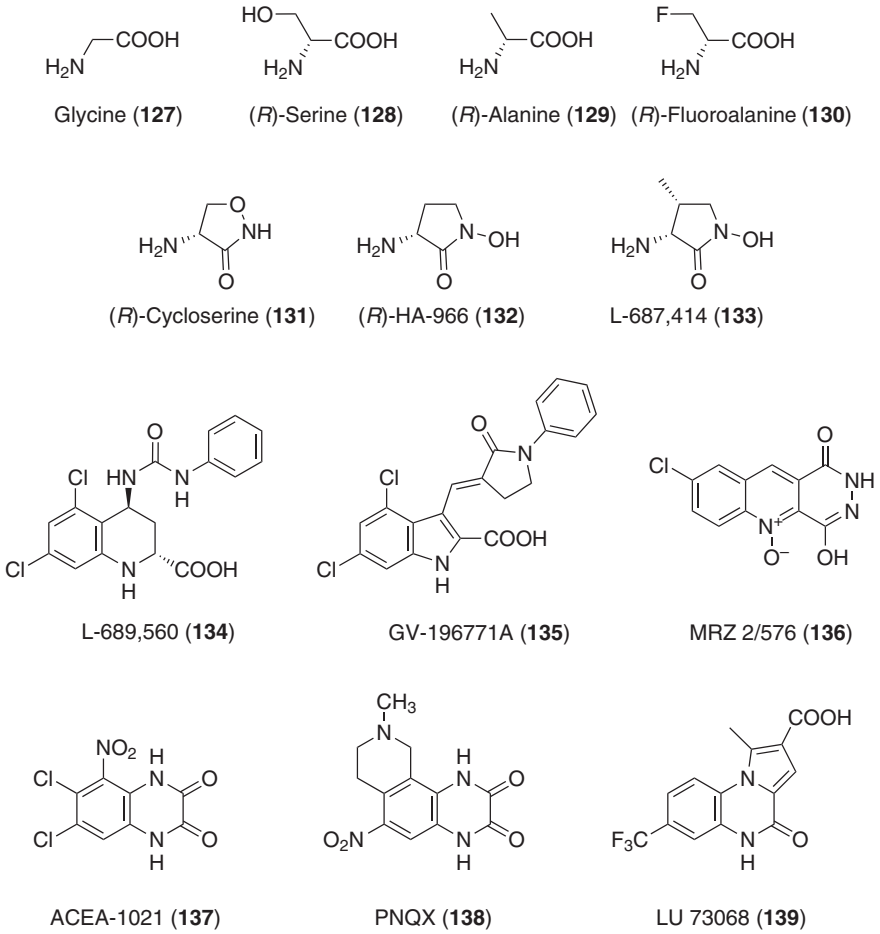


**Figure 23** Structures of glycine (**127**) and some compounds showing agonist (**128–130**), partial agonist (**131–133**), or antagonist (**134–139**) activity at the glycine$_B$ receptor site at the NMDA receptor complex.

receptors seem to modulate the level of activity at NMDA receptors. A certain concentration level of glycine is always present in the synapse. Thus, Glu is activating the NMDA receptors, whereas the level of glycine can modulate this activity and possibly control receptor desensitization [296]. (*R*)-Serine (**128**) is a potential endogeneous agonist at glycine$_B$ receptors [297, 298], and other small α-amino acids such as (*R*)-alanine (**129**) and (*S*)-fluoroalanine (**130**) enantioselectively bind to and activate glycine$_B$ receptors [299].

Limited success of competitive NMDA receptor antagonists as therapeutic agents has focused attention on the glycine$_B$ site. (*R*)-Cycloserine (**131**) and (*R*)-HA-966 (**132**) are partial glycine agonists, capable of penetrating the BBB after systemic administration. These pyrrolidine analogs have become useful tools for studying the in vivo role of the glycine$_B$ site [300, 301]. (*R*)-Cycloserine (**131**) has shown promising effects in the treatment of schizophrenia and AD [251, 302]. The methylated analog L-687,414 (**133**) is more potent as an antagonist than the parent compound, (*R*)-HA-966 (**132**) [303]. L-687,414 (**133**) has neuroprotective properties in animal models, and it does not prevent hippocampal LTP, which is often seen after administration of full antagonists. This indicates that partial agonists may have therapeutic advantages as compared to full antagonists in terms of fewer side effects [304]. A number of glycine$_B$ antagonists have also been developed. L-689,560 (**134**) [305] displays high potency and is derived from the endogenous compound kynurenic acid, the first glycine$_B$ antagonist reported. Further development of such structures has led to GV196771A [306] (**135**). The potent pyridophtalazindione MRZ 2/576 (**136**) represents another class of glycine$_B$ receptor antagonist that exhibits neuroprotective and antinociceptive properties in animal models [307, 308]. Quinoxalinediones such as ACEA-1021 [309] (**137**) and PNQX [310] (**138**) show selective antagonist activity at the glycine$_B$ site (see below). Other quinoxalinediones as LU 73068 (**139**) have shown mixed glycine$_B$ site and AMPA/KA receptor affinities and produce potent anticonvulsive effects [311].

***Subtype-Selective NMDA Ligands and Their Therapeutic Potential*** A number of competitive, uncompetitive, noncompetitive NMDA antagonists and glycine$_B$ site antagonists have shown promise in various animal models. However, when these compounds have been administered to humans, severe side effects have limited or prevented their therapeutic use. A better understanding of the heterogeneity of NMDA receptors may, however, provide new therapeutic possibilities. Subtype-selective agents could potentially lack such side effects.

Noncompetitive antagonists such as ifenprodil (**122**), CP-101,606 (**123**), and Co 101676 (**124**) have received much attention because of their subtype selectivity and their mode of action, specifically their activity-dependent blockade of NR2B-containing NMDA receptors [291, 312]. Ifenprodil (**122**) inhibits NR1/NR2B receptors at least 150-fold more than NR1/NR2C or NR1/NR2D receptors [313].

NMDA receptor blocking agents, such as (*R*)-CPP (**107**), ketamine (**115**), and dextromethorphan (**118**) have shown antinociceptive activity in chronic neurogenic pain models as well as in humans [280, 314, 315]. Unfortunately, the effects of the compounds are often seen at doses causing the typical side effects described earlier for NMDA receptor antagonists. The future development of analgesic agents may depend on the identification of subtype-selective agents, and much interest is focused on NR2B selective NMDA antagonists [291, 316]. NMDA blocking agents also show anxiolytic properties. However, it should be emphasized that even modest cognitive side effects are unlikely to be acceptable for this clinical application.

The affinity of $glycine_B$ receptor ligands is dependent on the presence of the NR2 subtype, although the glycine binding site is located on the NR1 subunit [317]. GV196771A (**135**) shows some selectivity for heteromers containing NR2A/2B subunits in combination with NR1, and has shown activity in animal models of neuropathic pain [306].

Certain uncompetitive antagonists such as MK-801 (**116**) (slower kinetics), dextromethorphan (**118**), and memantine (**117**) have shown some NR2C subtype selectivity, although there is some controversy about the selectivity of memantine (**117**) as an NMDA receptor ligand [279, 318].

## AMPA Receptor Ligands

***AMPA Receptor Agonists*** A large number of selective and potent AMPA receptor agonists have been developed by substitution of a heterocyclic bioisosteric group for the distal carboxylate group of Glu. For example, the heterocycles 1,2,4-oxadiazole-3,5-dione, 3-isoxazolol, and uracil, as represented by quisqualic acid (**140**), AMPA (**144**), and (*S*)-willardiine (**141**), respectively, have been incorporated into numerous AMPA receptor agonists (Fig. 24). The natural product quisqualic acid (**140**) was the first agonist in use for pharmacological characterization of AMPA receptors, but due to nonselective action it was later replaced by AMPA (**144**). Analogs of AMPA containing aliphatic, aromatic, or heteroaromatic substituents at the 5-position of the isoxazole ring have been synthesized and pharmacologically characterized [319–321]. Replacement of the 5-methyl group of AMPA by small alkyl groups, such as ethyl or trifluoromethyl, resulted in compounds with potencies in the same range as that of AMPA and that show similar intrinsic activity [322, 323]. Substitution of the methyl group of AMPA for hydrogen resulted in demethyl-AMPA (**145**), which has 300 times lower agonist potency in the rat cortical wedge assay system, but only 10 times lower affinity [321]. This apparent loss of agonist activity of demethyl-AMPA in the rat cortical wedge may be explained by an observed affinity for the cloned Glu transporters EAAT1 and EAAT2, though not for EAAT3 [324]. Thus, demethyl-AMPA may be a substrate for the former two transporters and may be removed from the synaptic cleft. Increasing the size of the 5-alkyl substituent of AMPA to three or more carbon atoms markedly reduces agonist potency [321, 325].

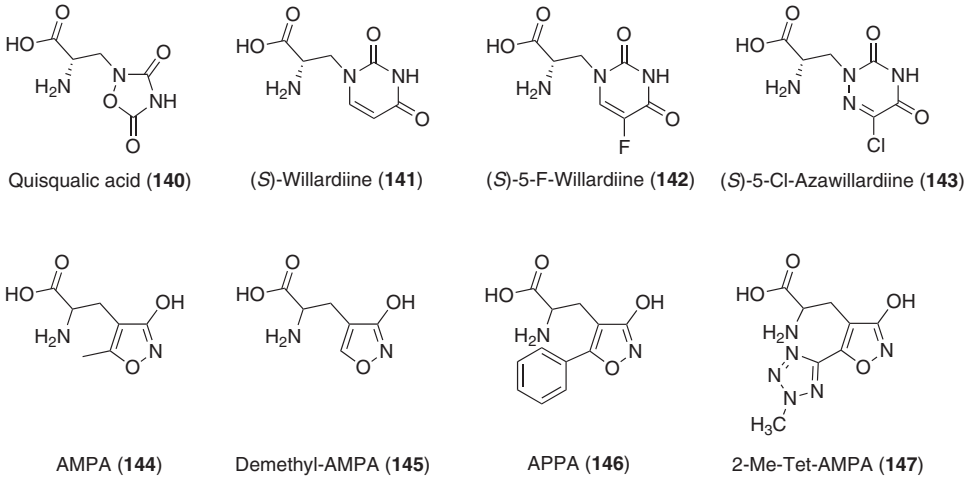


**Figure 24** Structures of AMPA (**144**) and some AMPA receptor agonists.

Whereas all alkyl-substituted AMPA analogs synthesized so far have shown either full agonism or inactivity at AMPA receptors, the 5-phenyl-substituted analog of AMPA, APPA (**146**), shows the characteristics of a weak partial AMPA receptor agonist, with an efficacy of approximately 60 percent relative to that of AMPA [326]. Quite surprisingly, (*S*)-APPA turned out to be a full AMPA receptor agonist, while (*R*)-APPA proved to be a weak competitive AMPA receptor antagonist [327]. In order to elucidate the SAR of these 5-substituted AMPA-analogs further, a number of 5-membered heteroaromatic substituents have been introduced, including an unsubstituted 5-tetrazolyl (Tet-AMPA), 1-methyl-5-tetrazolyl (1-Me-Tet-AMPA), and 2-methyl-5-tetrazolyl (2-Me-Tet-AMPA) (**147**) [319]. Whereas Tet-AMPA and 1-Me-Tet-AMPA are devoid of agonist activity at AMPA receptors, 2-Me-Tet-AMPA (**147**) proved to be the most potent AMPA receptor agonist synthesized so far, at least 10 times more potent than AMPA [319]. This suggests that steric as well as electronic properties play important roles for the binding of ligands to the receptor.

Electrophysiological studies of the enantiomers of 2-Me-Tet-AMPA (**147**) have shown that the excitatory properties reside exclusively with the *S*-enantiomer [328]. Compared to AMPA (**144**), which does not distinguish between the individual cloned AMPA receptors [328, 329], (*S*)-2-Me-Tet-AMPA shows some selectivity for homomers of GluR3 and GluR4 over receptors constructed from GluR1 or GluR1/GluR2, in electrophysiological studies in *Xenopus* oocytes.

Uracil and 6-azauracil are both effective bioisosteres of the distal carboxylate group of Glu. The natural product (*S*)-willardiine (**141**), in which the $pK_a$ of the uracil moiety is considerably higher than that of the distal carboxylate group of Glu [330], showed low, but measurable, affinity for both AMPA and

KA receptors and induces a rapidly desensitizing response in hippocampal neurons [331]. Introduction of electron-withdrawing substituents in the 5-position of the uracil moiety, increasing the acidity of the heterocycle markedly, improves the receptor affinity and has provided some of the most potent and subtype-selective AMPA and KA receptor agonists yet identified [332]. SAR studies on 5-halogen-substituted willardiines have shown that AMPA receptor affinity and potency in hippocampal neurons were dependent on the size and the electron-withdrawing effect of the substituent [331–335]. (*S*)-5-F-Willardiine (**142**) (Fig. 24) is the most potent agonist in this series, whereas willardiine analogs with sterically larger substituents apparently were not so well accommodated by AMPA receptors [332]. 5-Halogen-substituted 6-azawillardiines showed an SAR somewhat similar to that of the 5-substituted willardiines with (*S*)-5-Cl-6-azawillardiine (**143**) (Fig. 24) possessing highest AMPA receptor affinity [332]. Furthermore, (*S*)-5-F-willardiine (**142**) and (*S*)-willardiine (**141**) induce strongly desensitizing agonist responses at AMPA receptors, whereas willardiine analogs containing larger substituents, such as (*S*)-5-I-willardiine (**176**) (Fig. 29), produce much weaker desensitization comparable to that of KA [331]. At cloned AMPA receptors, the willardiine series of compounds show remarkable subtype selectivity, having 10 to 20 times higher affinity for GluR1 or GluR2 than GluR4 [332]. By contrast, the 6-azawillardiine analogs, like AMPA, did not distinguish significantly between the individual AMPA receptor subtypes [332].

The isoxazole-based Glu homolog (*S*)-Br-HIBO (**148**) also shows AMPA receptor subtype selectivity, preferring GluR1 over GluR3 in receptor binding and functional assays [336] (Fig. 25). The structural basis for this selectivity was first modeled and confirmed by mutation studies [337], followed by

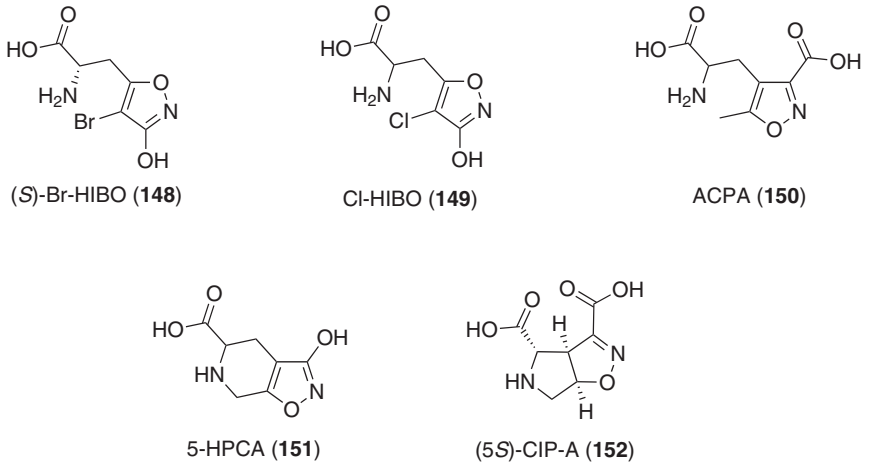


**Figure 25** Structures of AMPA receptor selective Glu homologs (**148–150**) and AMPA receptor selective bicyclic Glu analogs (**150**, **151**).

an X-ray crystallographic characterization [338]. As for the willardiines, the selectivity of (*S*)-Br-HIBO (**148**) is intimately related to preferential hydrogen bonding of the distal anion of (*S*)-Br-HIBO (**148**) through a water molecule to a nonconserved tyrosine residue of GluR2 [335, 337, 338]. A bromo-chloro replacement gave the racemic compound Cl-HIBO (**149**), which was recently reported to be a strongly desensitizing full agonist at GluR1, showing increased receptor selectivity ratios in electrophysiology experiments, ranging from 275 to 1600 for GluR1/2 over GluR3/4 [339]. Replacing the 3-isoxazolol group of AMPA by a 3-carboxyisoxazole unit gives the Glu homolog ACPA (**150**), which is a selective AMPA receptor agonist that is more potent than AMPA on cortical neurons as well as in *Xenopus* oocytes expressing cloned AMPA receptors [340, 341]. The potent excitatory AMPA receptor activity of ACPA (**150**) has been shown to reside with the *S*-enantiomer [342].

Although the 3-isoxazolol and the uracil ring systems have been the most widely used bioisosteres for the distal carboxyl group of Glu in the design of AMPA receptor agonists, other acidic heterocyclic units have proven useful as well. As mentioned in the beginning of this section, AMPA receptors were originally identified pharmacologically by using the naturally occurring Glu analog quisqualic acid (**140**) [343, 344] (Fig. 24). This 1,2,4-oxadiazole-3,5-dione analog of Glu was later shown to possess other Glu receptor activities, in particular potent mGluR agonist effects. As a consequence, quisqualic acid (**140**) has been replaced by the highly selective compounds AMPA (**144**) [345, 346] and (*S*)-F-willardiine (**142**) [347], as standard agonist at AMPA receptors. Conformational restriction of the skeleton of Glu has played an important role in the design of selective GluR ligands. However, only few structurally rigid AMPA receptor-selective Glu analogs have been reported. One such example is the cyclized analog of AMPA, 5-HPCA (**151**) [76] (Fig. 25), which recently has been resolved [348].

Interestingly, the pharmacological effects of 5-HPCA (**151**) reside exclusively with the *R*-enantiomer, in striking contrast to the usual stereoselectivity trend among AMPA receptor agonists. To examine the structural basis of the observed enantiopharmacology, (*R*)- and (*S*)-5-HPCA have been docked into the GluR2 ligand binding site. These studies showed that due to the rigid nature of the compound (*R*)-5-HPCA, but not (*S*)-5-HPCA, can present almost all of the required pharmacophore elements to the receptor [348]. Whereas (*R*)-5-HPCA selectively displaces [$^3$H]AMPA binding, the 3-carboxyisoxazolineproline, CIP-A, a bicyclic Glu analog, displays affinity for both AMPA and KA receptors [349]. This dual affinity resides exclusively with the 5*S*-isomer, (5*S*)-CIP-A (**152**) (Fig. 25) [350].

***Competitive and Noncompetitive AMPA Receptor Antagonists*** Early pharmacological studies on AMPA and KA receptors were hampered by the lack of selective and potent antagonists. The discovery of the quinoxaline-2,3-diones CNQX (**153**) and DNQX (**154**) was a breakthrough since these com-

pounds are quite potent antagonists, although nonselective [351] (Fig. 26). Subsequently, the more potent analog NBQX (**155**) was shown to be neuroprotective in cerebral ischemia and to have improved AMPA receptor selectivity compared to CNQX (**153**) [352]. However, NBQX (**155**) failed in clinical trials because of nephrotoxicity due to a limited aqueous solubility but nonetheless has become a valuable tool for research. DNQX (**154**) has played a key role in elucidating the binding mode of competitive antagonists, as it was the first antagonist co-crystallized with the GluR2 ligand-binding domain [353] (see Structure of Ionotropic Glutamate Receptors). Attempts to improve the aqueous solubility of such antagonists without losing activity at AMPA receptors, by introducing appropriate polar substituents onto the quinoxaline-2,3-dione ring system, have been highly successful and have resulted in very potent AMPA receptor antagonists, as exemplified by ZK200775 (**156**) [354] and compound **157** [355].

Another series of potent and selective competitive AMPA receptor antagonists based on the isantin oxime skeleton includes NS 1209 (**158**), which shows long-lasting neuroprotection in animal models of ischemia and an increased aqueous solubility compared to NBQX (**155**) [356] (Fig. 26). In recent years, several series of imidazo[1,2-*a*]indeno[1,2-*e*]pyrazin-based com-

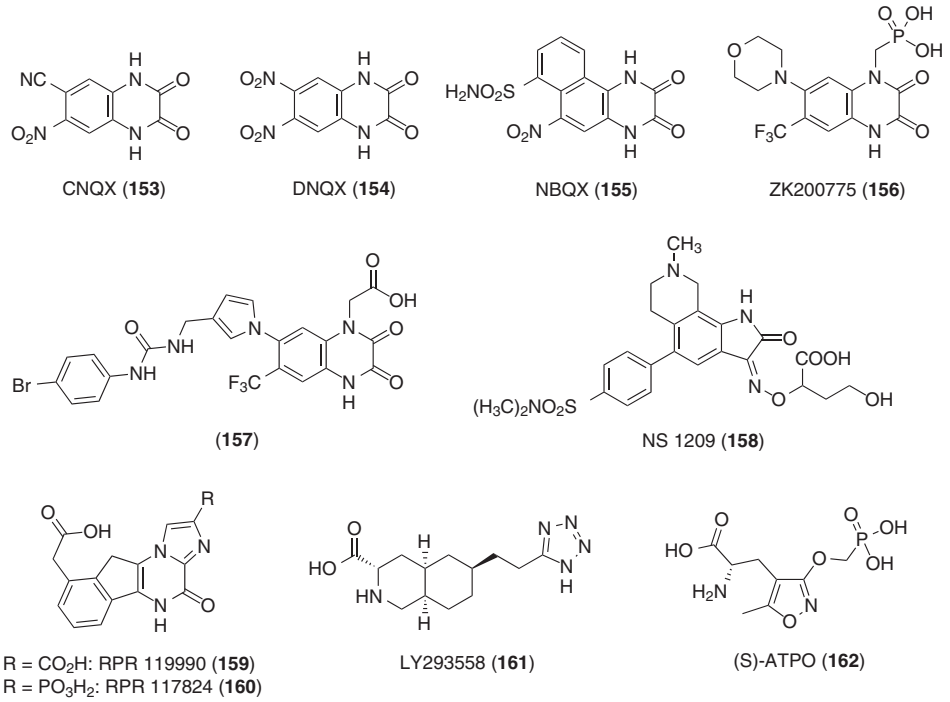


**Figure 26** Structures of some competitive AMPA receptor antagonists.

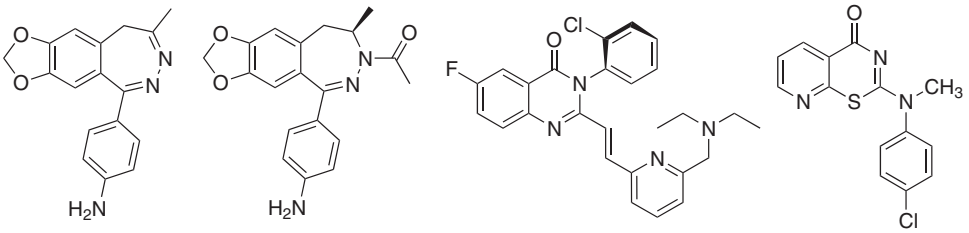


**Figure 27** Structures of some noncompetitive AMPA receptor antagonists.

petitive antagonists have been identified [357–360]. The initial compounds displayed high, although nonselective, affinity for AMPA and NMDA receptors, but through extensive SAR studies, highly potent and selective compounds such as RPR 119990 (**159**) and RPR 117824 (**160**) have been identified. However, a drawback of this series of compounds is their lack of oral activity [361]. At least two classes of amino-acid-containing compounds, based on decahydroisoquinoline-3-carboxylic acid [362] and AMPA [363], have been found to be competitive AMPA receptor antagonists. LY293558 (**161**), a member of the former class, is systemically active, although it shows significant antagonist effects at KA receptors in addition to its potent AMPA receptor blocking effects [329, 364, 365]. The AMPA receptor antagonist (*S*)-ATPO (**162**), which was designed using AMPA as a lead structure, has a carbon backbone longer than that which normally confers AMPA receptor agonism [366].

The 2,3-benzodiazepines, such as GYKI 52466 (**163**) and Talampanel (**164**), represent a class of noncompetitive AMPA receptor antagonists that have enabled effective pharmacological separation of AMPA and KA receptor-mediated events [367] (Fig. 27). These compounds appear to bind to sites distinct from the agonist recognition site, and are thus negative allosteric modulators [368, 369]. Talampanel (**164**), currently under clinical development as a treatment for multiple sclerosis, epilepsy, and Parkinson's disease [367], may inhibit AMPA receptor function even in the presence of high levels of Glu [367, 370]. A number of structurally similar potent noncompetitive AMPA receptor antagonists with anticonvulsant activity, typified by CP-465,022 (**165**) and YM928 (**166**), have been identified [371, 372]. CP-465,022 (**165**), the (+)-atropisomer of CP-392,110, interacts stereoselectively with a binding site identical to that of the 2,3-benzodiazepines [371, 373].

***Modulatory Agents at AMPA Receptors*** The agonist-induced desensitization of AMPA receptors can be markedly inhibited by a number of structurally dissimilar AMPA receptor potentiators known as AMPA-kines (Fig. 28), including aniracetam (**167**), cyclothiazide (**168**), and in particular CX-516 (**169**), which has been shown to improve memory function in aged rats [374].

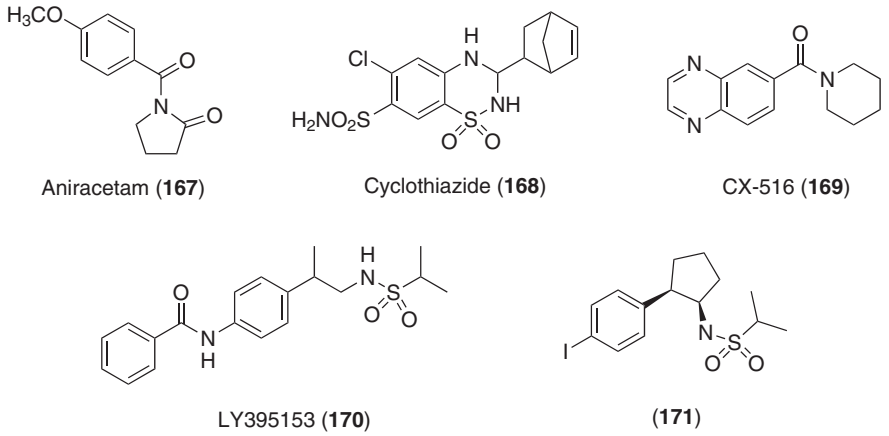


**Figure 28** Structures of some positive modulators of AMPA receptors.

These AMPA-kines positively modulate ion flux via stabilization of receptor subunit interface contacts and subsequent reduction in the degree of desensitization [375]. A series of more potent arylpropylsulfonamide-based AMPA-kines has been identified, including LY395153 (**170**) [376]. [$^3$H]LY395153 labels a modulatory site on AMPA receptors with nanomolar affinity but remains unclear whether all of the structurally diverse AMPA-kines share the same binding site [377]. Incorporation of the sulfonamidopropyl moiety of the arylpropylsulfonamide potentiators into a 2-substituted 1-sulfonamidocyclopentane ring has provided compound **171**, the most effective AMPA-kine yet described [378].

## KA Receptor Agonists and Antagonists

The pharmacology and pathophysiology of KA receptors are far less well understood than for AMPA receptors. However, identification of selective agonists and competitive antagonists has developed the field of KA receptor research, especially over the last 5 years, and has provided insight into roles of these receptors in the CNS. For a number of years, KA (**172**) and domoic acid (**173**) have been used as standard KA receptor agonists despite their activities at AMPA receptors, characterized by nondesensitizing responses at these receptors (Fig. 29). More selective KA receptor agonist activities have been described for (*S*)-ATPA (**174**) [329, 379], (*S*)-thio-ATPA (**175**) [380], and (*S*)-5-I-willardiine (**176**) [332, 381]. These compounds exhibit some selectivity among low-affinity KA receptor subtypes, namely a preferance for GluR5 over GluR6 [332, 379, 380]. Studies on isolated dorsal roots showed that ATPA and (*S*)-5-I-willardiine (**176**) are, respectively, 10- and 100-fold more potent than KA [381]. Whereas (*S*)-ATPA (**174**) and KA (**172**) induce strongly desensitizing responses in *Xenopus* oocytes expressing homomeric KA receptors of

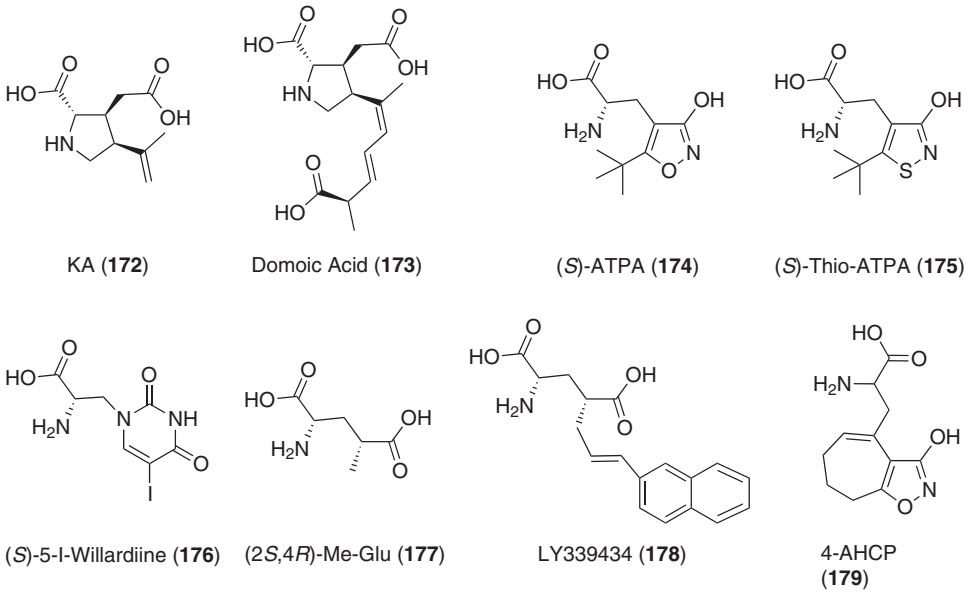


**Figure 29** Structures of KA (**172**) and some KA receptor agonists.

the GluR5 subtype, (*S*)-thio-ATPA (**175**) produce only partially desensitizing responses [380]. (*S*)-ATPA (**174**), (*S*)-thio-ATPA (**175**), and (*S*)-5-I-willardiine (**176**) are structurally related to potent AMPA agonists discussed in earlier sections, illustrating that the structural characteristics required for activation of GluR1-4 and GluR5 receptors are very similar. However, the presence of the relatively bulky and lipophilic *tert*-butyl- or iodo-substituents of these compounds is apparently the major determinant of the observed receptor selectivity. The ligand-binding domain of GluR2 has recently been crystallized in complex with (*S*)-ATPA (**174**) [382] (see next section).

Among the four possible stereoisomers of the 4-methyl-substituted analog of Glu, only the 2*S*,4*R*-isomer (**177**) shows selectivity for KA receptors, with a binding affinity comparable to that of KA. Replacement of the 4-methyl group of (2*S*,4*R*)-Me-Glu (**177**), a ligand that shows limited selectivity between GluR5 and GluR6, by a range of bulky, unsaturated substitutents containing alkyl, aryl, or heteroaryl groups [383–386] has yielded a number of interesting GluR5 receptor-selective compounds including LY339434 (**178**). LY339434 shows approximately a 100-fold selectivity for GluR5 over GluR6 and no affinity for GluR1, 2, or 4 receptors [383].

The conformationally restricted 3-isoxazolol-containing Glu homolog, 4-AHCP (**179**), was initially demonstrated to be a very potent AMPA agonist at rat spinal neurons [387]. In more recent studies, the excitatory effect of 4-AHCP (**179**) has been shown to reside with the *S*-enantiomer, which is equipotent with (*S*)-ATPA (**174**) and 35 to 115 times more potent at GluR5 than at cloned AMPA receptors or at GluR6/KA2 [388].

Whereas a large number of selective competitive AMPA receptor antagonists have been identified, particularly over the last decade, only a few selective KA receptor antagonists have been reported. Most of these KA receptor-preferring antagonists are structurally related to AMPA receptor antagonists and show at least some affinity toward AMPA receptors further to that KA receptor affinity, reflecting that similar structural features are required for activation and blockade of AMPA and some KA receptors, in particular GluR5.

One of the first reported KA receptor-preferring antagonists was the isantin oxime, NS 102 (**180**), which shows some selectivity toward low-affinity [$^3$H]KA sites [389] as well as antagonist effect at homomeric GluR6 [390] (Fig. 30). However, low aqueous solubility has limited the use of NS 102 (**180**) as a pharmacological tool. A number of decahydroisoquinoline-based acidic amino acids, including LY382884 (**181**), has been characterized as competitive GluR5-selective antagonists that exhibit antinociceptive effects [391]. In the course of SAR studies using this decahydroisoquinoline skeleton, excellent receptor selectivity been observed for compound **182**, which shows approximately 1000-fold higher affinity for GluR5 than for GluR1–4, GluR6, or GluR7. In addition, a group of quinoxalinedione-derived antagonists structurally related to the AMPA receptor antagonist **157** (Fig. 26), and with high affinity for GluR5, have been reported [392]. One example is BSF 91594 (**183**), which shows more than 50-fold selectivity for GluR5 over other KA receptor subtypes or AMPA receptors.

Most recently, a series of arylureidobenzoic acids have been reported as the first compounds with noncompetitive antagonist activity at GluR5 [393]. The most potent ligands, exemplified by compound **184**, exhibit more than 50-fold selectivity for GluR5 over GluR6 or the AMPA receptor subtypes. This com-

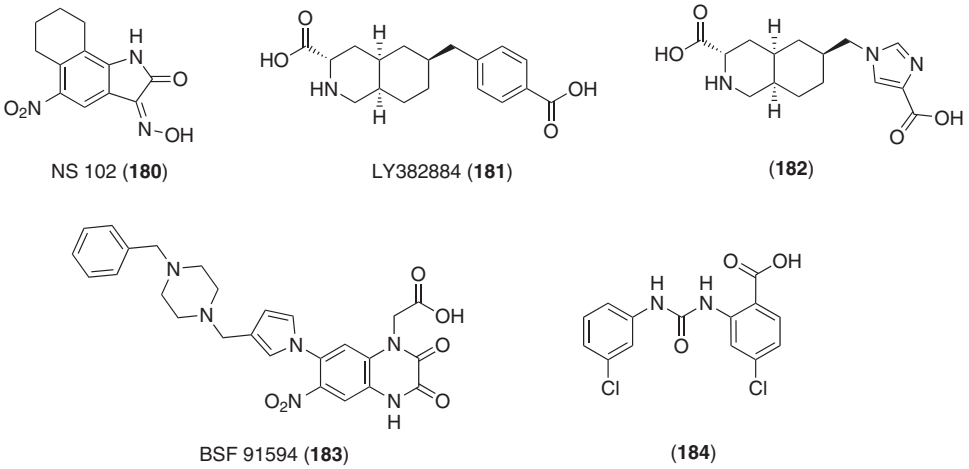


**Figure 30** Structures of some competitive KA receptor antagonists (**180–183**) and a noncompetitive KA receptor antagonist (**184**).

pound may prove valuable to further pharmacological research and as a lead in the search of therapeutic glutamatergic agents.

## Structure of Ionotropic Glutamate Receptors

iGluR subunits may be considered as modular constructions of three distinct domains (Fig. 31): (1) a transmembrane segment consisting of three transmembrane helices and a reentrant loop, distantly related to $K^+$ channels such as KcsA, forming the walls of the pore; (2) an extracellular ligand binding domain (S1S2) composed of two lobes connected by a hinge ("venus flytrap"), distantly related to glutamine binding protein (QBP); and (3) an N-terminal domain structurally related to periplasmic binding proteins such as leucine–isoleucine–valine binding protein (LIVBP), and which appears to play a role in oligomerization [394]. It is currently believed that iGluRs assemble as dimers-of-dimers, although the evidence is weaker for NMDA receptors than for non-NMDA receptors [251]. Numerous studies, including electro-

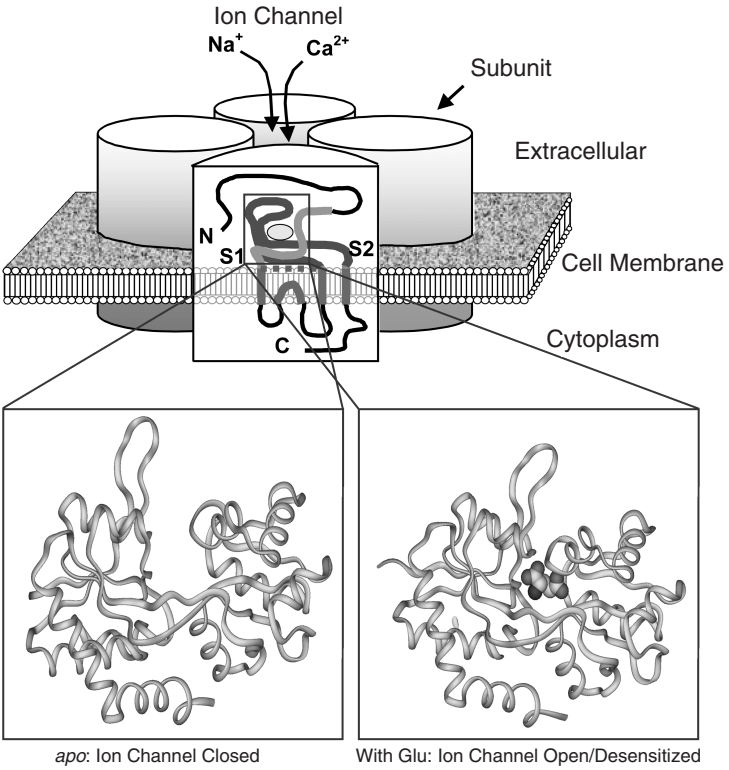


**Figure 31** Illustration of the tetrameric topology of AMPA receptors (*top*), showing *apo* and agonized state of the ligand binding core (*below*).

physiology [395, 396], electron microscopy [397], and X-ray crystallography [335] now support a tetrameric structure for the AMPA receptor that resembles some $K^+$ channels (Fig. 31).

Our understanding of the competitive binding site of iGluRs has been greatly enhanced by the development of a soluble construct of S1S2 that possesses many of the properties of the intact receptor [398]. Experimentation with linkers of various lengths connecting the two half-domains led to the crystallization of GluR2-S1S2, first with the partial agonist KA (**172**) [399], then with a wide variety of ligands as well as various point mutations. A bacterial homolog known as GluR0 has also been crystallized, and recently, a construct of the glycine-binding NMDA subunit, NR1. The currently available crystal structures are listed in Table 1.

As can be seen from Table 1, different competitive ligands (agonists, partial agonists, antagonists) induce different conformational changes in GluR2-S1S2. Parallel structural studies on NMDA receptors, long the subject of pharmacophore modeling [291], have now commenced. Structure-based design, docking, and homology modeling can be performed, using the structural repertoire presented by these crystallographic snapshots to take account of receptor flexibility [339, 388, 406, 407]. While the molecular dynamics of the GluR2 ligand binding core have been probed, as well as that of the related GlnBP, simulations have as yet not been able to reproduce the process of domain closure or opening, possibly because of the relatively long simulation times required [408, 409]. At GluR2, and presumably at other receptors in this class with highly conserved binding sites, Glu binds with a partially folded conformation [353]. The binding mode of Glu at Glu-sensitive NMDA receptors has not yet been demonstrated, although on the basis of pharmacophore modeling, it is at least as folded as at AMPA receptors. The amino acid moieties of agonists at GluR2 appear to be zwitterionized and are involved in ion-pair interactions—the ligand α-carboxylate with R485 in domain 1 and the ammonium with E705 in domain 2. By contrast, the terminal acidic moieties of Glu and AMPA agonists, while apparently ionized at the receptor, are not involved in ion pairing but rather accept a number of hydrogen bonds from receptor hydroxy and amide groups, as well as participating in a water-mediated hydrogen bond network. The receptor shows substantial flexibility in the distal region, with peptide side chains and water molecules rearranging to accomodate the ligand. For example, the 3-isoxazolol moiety of AMPA (**144**) does not bind in a similar position to Glu's ω-carboxylate, whereas that of 2-Me-Tet-AMPA (**147**) does (Fig. 32) [338, 353].

The structural details of the binding modes of two distinct classes of AMPA receptor antagonists are known: the unionized planar quinoxalinedione DNQX [353] and the tri-ionized AMPA analog (*S*)-ATPO (**162**) [338]. Neither of these ligands fully exploit the domain-open antagonized binding site, and neither come into close contact with nonconserved residues (Fig. 33). There is therefore substantial scope for the development of subtype-specific non-NMDA receptor antagonists.

**TABLE 1 List of Available Crystal Structures of iGluR Binding Domains in Complex with Various Ligands**

| Receptor | Ligand | Ligand Class | Domain Closure | Reference |
|---|---|---|---|---|
| NR1 (NMDA) | Glycine (**127**) | Endogenous agonist | | 251 |
| | (*R*)-Serine (**128**) | Endogenous agonist | | 251 |
| | (*R*)-Cycloserine (**131**) | Agonist | | 251 |
| | Dichlorokynurenic acid | Antagonist | | 251 |
| GluR2 (AMPA) | Glu (**96**) | Endogenous agonist | 20° | 353 |
| | AMPA (**144**) | Agonist | 20° | 353 |
| | Demethyl-AMPA (**145**) | Agonist | 20° | 400 |
| | 2-Me-Tet-AMPA (**147**) | Agonist | 21° | 338 |
| | (*S*)-ATPA (**174**) | Agonist (GluR5 selective) | 21° | 382 |
| | Thio-ATPA (**175**) | Agonist (GluR5 selective) | 18° | 401 |
| | Br-HIBO (**148**) | Agonist (GluR1,2 selective) | 17° | 338 |
| | ACPA (**150**) | Agonist | 20° | 338 |
| | Quisqualic acid (**140**) | Agonist | 20° | 402 |
| | (*S*)-Willardiine (**141**) | Agonist (GluR1,2 selective) | 17° | 334, 335 |
| | (*S*)-F-Willardiine (**142**) | Agonist (GluR1,2 selective) | 16° | 334, 335 |
| | (*S*)-Br-Willardiine | Agonist | 15° | 334, 335 |
| | (*S*)-I-Willardiine (**176**) | Agonist (GluR5 selective) | 11° | 334, 335 |
| | KA (**172**) | Partial agonist | 12° | 399 |

| | | | | |
|---|---|---|---|---|
| | DNQX (**154**) | Antagonist | 5° | 353 |
| | (*S*)-ATPO (**162**) | Antagonist | 2–5° | 403 |
| | *Apo* | (None) | 0° | 353 |
| GluR2 (Y702F) | ACPA (**150**), (*S*)-Br-HIBO (**148**) | | | 338 |
| GluR2 (L650T) | AMPA (**144**), quisqualic acid (**140**), KA (**172**) | | | 404 |
| GluR2 (L650T, L483Y) | AMPA (**144**) | | | 404 |
| GluR2 (L483Y) | AMPA (**144**), DNQX (**154**) | | | 375 |
| GluR2 (N754D) | KA (**172**) | | | 375 |
| GluR2 (N754S) | Glu (**96**) + cyclothiazide (**196**) | Allosteric modulator (desensitization blocker) | | 375 |
| GluR0 | Glu (**96**), (*R*)-serine (**128**), *apo* | | | 405 |

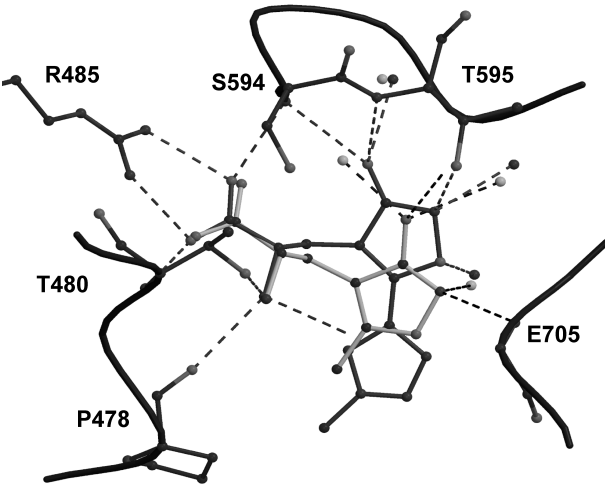


**Figure 32** Full agonists AMPA (**144**) and 2-Me-Tet-AMPA (**147**) co-crystallized with GluR2-S1S2. Note the difference in the positioning of the isoxazole rings and the differences in accompanying water molecules and hydrogen bond network.

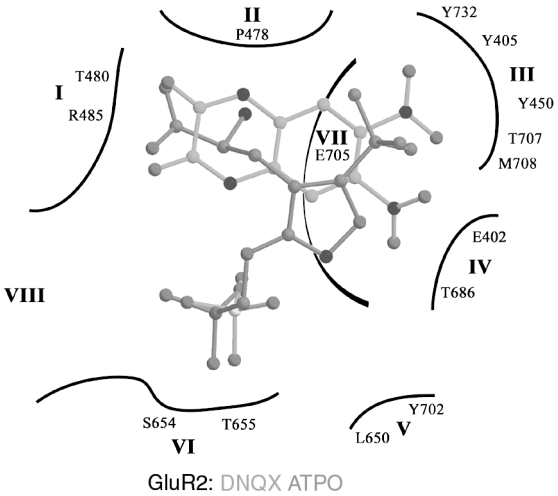


**Figure 33** Competitive antagonists DNQX (**154**) and ATPO (**162**) co-crystallized with GluR2-S1S2, showing different binding modes of the two classes of antagonists. Regions III–VI show some variations between AMPA and KA receptors, while only region V varies among AMPA receptor subtypes.

The high homology among closely related iGluR subunits (see Table 2), for example, GluR1-7, allows models to be built from the X-ray structures of GluR2 with confidence [337, 339, 388, 406, 410]. While the competitive binding sites of GluR1,2 are essentially identical, as are GluR3,4, the single nonconserved residue between these two subclasses of AMPA receptors, Y/F702, has been open to exploitation, even though it only interacts with ligands via a water molecule. Thus, a series of selective GluR1,2 ligands has been designed and synthesized, and the basis of selectivity is well understood [339, 388]. The deliniation between GluR5 [which weakly binds both AMPA (**144**) and KA (**172**)] and the AMPA-preferring GluR1-4 is sharper; L/V650 and M/S708 make direct steric contact with the ligand. GluR5-selective ligands such as (*S*)-4-AHCP (**179**) and (*S*)-thio-ATPA (**175**) exploit the altered shape of the binding cavity [388]. The development of specific agonists and antagonists for other subtypes is ongoing, a task made easier by such detailed structural information.

As with many ion channels, iGluRs exhibit desensitization, that is, that following rapid activation and opening of the channel, the receptor undergoes a conformational change on a millisecond timescale, characterized by low conductance even in the continued presence of agonist [247]. In general, full agonists cause rapid and almost complete desensitization, whereas partial agonists are weakly desensitizing. Much is now known about the transmission of free energy of binding of agonists and partial agonists to domain closure, channel opening, and subsequent desensitization, even though we lack atomic resolution structural detail of the full-length receptor [334]. Compounds modulating desensitization are of particular interest as therapeutic agents, and great inroads have been made toward unravelling the structural mechanism of desensitization [375]. While the S1S2 construct used in crystallization experiments has generally been of the *flop* splice variant of GluR2, some of the properties of the *flip* isoform have been investigated by introducing the key mutation N754S, known to confer sensitivity to the desensitization blocker cyclothiazide (**168**) [411].

This crystal structure (Fig. 34) reveals that cyclothiazide (**168**), and presumably other allosteric modulators of AMPA receptors, affect desensitization by binding to and stabilizing the interface between two subunit monomers [375]. It is likely that other AMPA receptor modulators, both positive and negative, act via this interface—a promising target for structure-based drug design.

## Metabotropic Glutamate Receptor Ligands

The cloning of the mGluRs and the evidence that has subsequently emerged on their potential utility as drug targets in a variety of neurological disorders has encouraged medicinal chemists to design ligands targeted at the mGluRs. Over the first decade since mGluRs were cloned, our knowledge of the SAR of these receptors has increased dramatically and led to the discovery of new

**TABLE 2 Binding Site Homology of Selected iGluRs**[a]

| | 402 | 450 | 478 | 485 | 650 | 654 | | 686 | 702 | 705 | | 723 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| NR1 | I **C** T | K **F** G | A **P** L **T** I | E **R** A | T **V** K | Q S **S** | **V** D | A **I** Q | A **F** I | W **D** S A | **V** L | G **F** G |
| GluR2 | L **E** S | K **Y** G | A **P** L **T** I | V **R** E | T **L** D | S G **S** | **T** K | T **T** A | A **Y** L | L **E** S T | **M** N | G **Y** G |
| GluR4 | M **E** S | K **Y** G | A **P** L **T** I | V **R** E | T **L** D | S G **S** | **T** K | T **T** A | A **F** L | L **E** S T | **M** N | G **Y** G |
| GluR5 | L **E** E | K **Y** G | A **P** L **T** I | V **R** E | A **V** R | D G **S** | **T** M | N **S** D | A **L** L | M **E** S T | **S** I | G **Y** G |
| GluR6 | L **E** E | K **Y** G | A **P** L **T** I | V **R** E | A **V** E | D G **A** | **T** M | S **N** E | A **F** L | M **E** S T | **T** I | G **Y** G |
| KA1 | L **E** N | V **Y** G | A **G** L **T** I | E **R** E | T **I** H | G G **S** | **S** M | S **T** E | A **F** L | L **E** S T | **M** N | G **Y** G |
| δ2 | L **E** E | K **Y** G | S **A** L **T** I | D **R** E | T **V** L | D S **A** | **V** Y | E **S** Q | A **F** V | W **D** A A | **V** L | G **Y** G |
| Function | "Lock" | Pocket | $-NH_3^+$ | $-COO^-$ | Pocket | Distal | anion | "Lock" | Pocket | $-NH_3^+$ | Pocket | Pocket |

[a] Numbering according to GluR2. Residues in bold are known to interact directly with ligands or line the binding site.

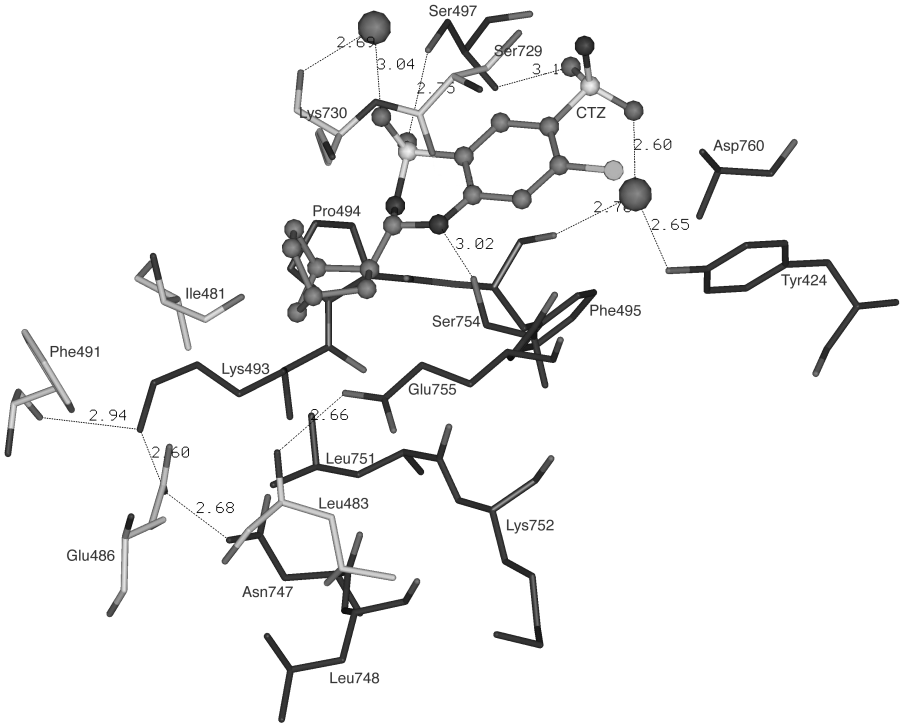


**Figure 34** Cyclothiazide (**168**) bound to the (N754S)GluR2-S1S2 dimer interface. Interfacial residues colored black or cyan according to monomer of origin.

potent and selective ligands [412]. The vast majority of these ligands are Glu analogs incorporating a glycine moiety and a distal acidic function, but positive and negative allosteric modulators with no structural similarity to Glu have also been discovered. In the following sections we will describe some of the principles of drug design that have been employed by medicinal chemists in order to design ligands with increased receptor selectivity and potency.

### *Metabotropic Glutamate Receptor Agonists*

*Conformationally Constraint Analogs* The first agonist to show selectivity for mGluRs over to iGluRs was (1*S*,3*R*)-ACPD (**185**) [413]. When tested on the eight cloned mGluRs it acts rather nonselectively [414–420], but it has been used extensively as a template for design of new mGluR ligands, some of which are shown in Figure 35. Within these series of compounds it is interesting to note the profound changes in receptor selectivity, potency, and efficacy caused by relatively small structural changes. Introduction of a nitrogen atom in the C4 position of **185** gives (2*R*,4*R*)-APDC (**186**), which displays an increased potency for group II receptors compared to the parent compound while losing

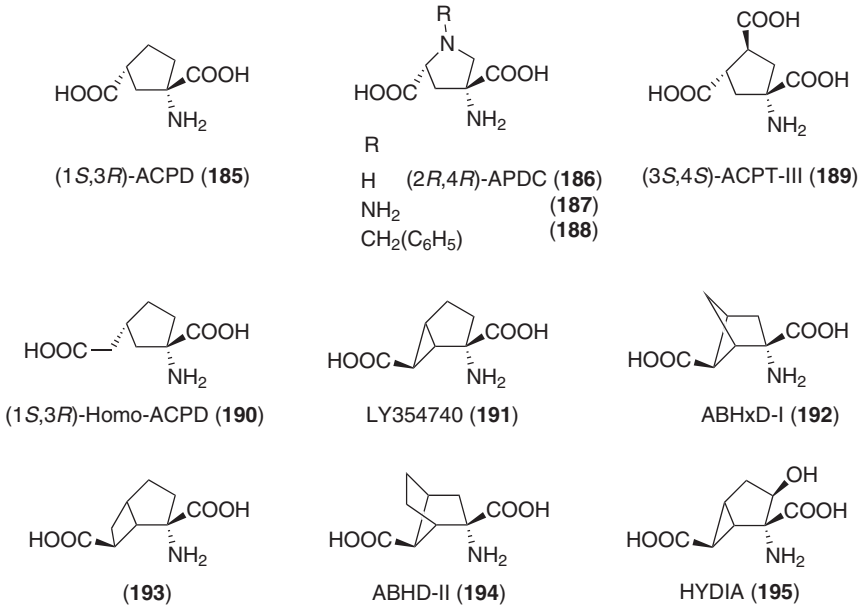


**Figure 35** Structures of some mGluR agonists.

affinity for group I and III receptors [421–423]. N1-substitutions of **186** have been reported to cause significantly reduced potency at group II receptors [423–425]. Interestingly, however, 1-amino-APDC (**187**) acts as a partial group II agonist [425], and 1-benzyl-APDC (**188**) shifts selectivity from mGluR2 and mGluR3 toward mGluR6 [423]. Similarly, addition of a carboxylic acid moiety at the corresponding C4-position of **185** to give (3*S*,4*S*)-ACPT-III (**189**) provides agonist potency at mGluR4 comparable to Glu, as well as weak antagonist effects at mGluR1 and mGluR2 [426].

In 1997 we reported that (1*S*,3*R*)-homo-ACPD (**190**), the homolog of **185**, possesses increased selectivity for mGluR2 as compared to mGluR1 and mGluR4, albeit with lower potency and efficacy than the parent compound [427]. At the same time, a dramatic 10,000-fold increase in potency at group II receptors was reported by further restraining the conformation of **190** by introducing a single bond between the side chain and the 5-position in the cyclopropane ring to give LY354740 (**191**) [421, 428, 429]. Compound **191** displays low-nanomolar agonist potency at mGluR2 and mGluR3, low-micromolar agonist potency at mGluR6 and mGluR8, while showing no activity at the remaining mGlu receptors.

The synthesis and pharmacology of ABHxD-I (**192**) has been reported [430]. Interestingly, this compound displays agonist potency comparable to Glu at mGluR1 to mGluR6; thus the compound is less selective than the parent compound **185**. The observation has been of key importance in developing early models of the mGluR binding site. Compound **192** is quite a rigid

molecule, which adopts a conformation corresponding to the extended anti–anti conformation of Glu. The observation that the compound is a potent agonist on all three mGluR groups led the authors to suggest that Glu adopts the same extended conformation at all three receptor groups, and that group selectivity is thus not a consequence of different conformations but rather a consequence of other factors such as steric hindrance [430]. This hypothesis has later been confirmed by several more extensive pharmacophore models [431–433]. One notable example of these sharp steric requirements can be seen by comparing highly selective group II agonist **191** with the nonselective agonist **192**. Both compounds prefer conformations corresponding to the extended anti–anti conformation of Glu [423, 430]. However, overlaying the compounds with Glu (all in the anti–anti conformation) shows that **191** does occupy extra volume compared to **192** [430]. Accordingly, it has been suggested that group II receptors can accept this extra volume, whereas group I and III receptors cannot [430]. The unique pharmacological profile of **191** is further highlighted by other close analogs such as compound **193** and ABHD-II (**194**), which are inactive and weak group I/II agonists, respectively [434, 435].

The 2-(carboxycyclopropyl)glycines are also conformationally restricted analogs of Glu, albeit with more flexibility than the bicyclic compounds described above. In agreement with the pharmacophore models, it is also the isomers [in particular (2*S*,3*R*,4*S*)-CCG, **102**] (Fig. 20) with the extended conformation that are agonists at mGluRs [436]. Also in agreement with the findings from pharmacophore modeling showing that it is the fully extended conformation of Glu that recognizes mGluR1, mGluR2, and mGluR4, (2*S*,3*R*,4*S*)-CCG (**102**) has been shown to display potent agonist activity at members of all three mGluR groups, albeit with some preference for the group II receptors [436, 437]. The high potency and group II selectivity of **191** has inspired the synthesis of analogs, a few of which are shown in Figure 35. The 3-hydroxy analog of **191**, HYDIA (**195**), has been reported [438]. It is interesting to note that such a modest change in the parent compound converts the compound from an agonist into a potent group II antagonist.

*Extension of Backbone Chain Length* Based on the observation that (*S*)-2-aminoadipic acid (**196**) [439–442], the homolog of Glu, displays increased mGluR selectivity compared to Glu, we have further studied the effects of backbone extension (Fig. 36). (*S*)-2-Aminoadipic acid (**196**) selectively activates mGluR2 and mGluR6, whereas it has no effect on mGluR1, mGluR4, or mGluR5. Working along these lines led to the discovery that (*S*)-homo-AMPA (**197**) is a specific agonist at mGluR6 with no activity at mGluR1-5 or mGluR7 [440, 441]. Since mGluR4, mGluR6, and mGluR7 all belong to group III, it was surprising to find that (*S*)-2-aminoadipic acid (**196**) and (*S*)-homo-AMPA (**197**) only activate mGluR6 within this group. It is also interesting to note the dramatic pharmacological difference between AMPA (**144**) and its homolog **197** at AMPA receptors. As described previously, the former is a

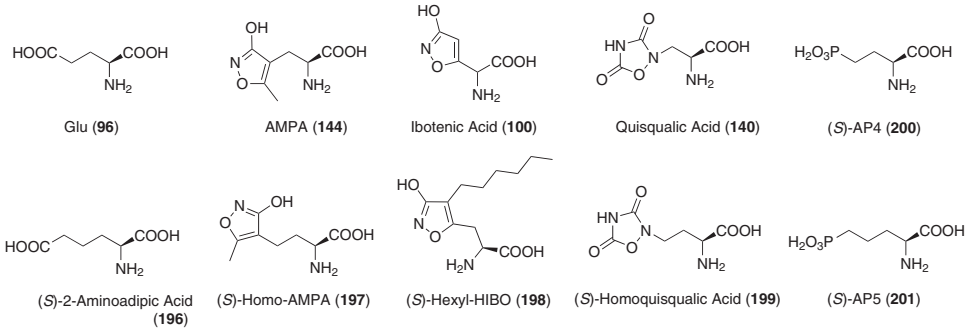


**Figure 36** Structures of some Glu analogs showing effects at iGluRs and/or at mGluRs (*upper row*) and the corresponding homologs that interact preferentially with mGluRs (*lower row*).

highly selective agonist, whereas the latter is completely inactive at AMPA receptors [441]. In addition to their previously described effects as AMPA receptor agonists, ibotenic acid (**100**) and quisqualic acid (**140**) were shown in the mid-1980s to stimulate IP accumulation indicative of metabotropic effects [443]. The cloning of mGluRs confirmed these results by showing that both compounds activate group I receptors [417, 444, 445]. Whereas quisqualic acid (**140**) is more potent than Glu at group I mGluRs, ibotenic acid (**100**) is less potent.

Furthermore, the latter compound also shows activity at group II receptors [442]. The pharmacological effects of the homologs of ibotenic acid (**100**) and quisqualic acid (**140**) has been reported [442]. Interestingly, (*S*)-Br-HIBO (**148**) (Fig. 25) is a selective (albeit weak) antagonist at mGluR1 and mGluR5. Introducing more lipophilic/bulky side chains in the 4-position increases the activity of the compounds as group I antagonists. HIBO analogs with small side chains also activate AMPA receptors, but in contrast to the mGluR1,5 antagonism, this activity diminishes with increasing size of the side chain. This trend peaks at six carbons providing (*S*)-hexyl-HIBO (**198**), which is a fairly potent group I antagonist devoid of activity at AMPA receptors [446]. (*S*)-Homoquisqualic acid (**199**) also produces effects quite different from the parent compound **140**. On mGluR5 (*S*)-homoquisqualic acid (**199**) is also an agonist albeit with 500-fold lower activity than quisqualic acid (**140**), whereas the homolog displays surprising antagonism at mGluR1 [447]. Furthermore, (*S*)-homoquisqualic acid (**199**) is an agonist at mGluR2 in contrast to the parent compound [447].

(*S*)-AP4 (**200**) is also a classical mGluR agonist that in the early 1980s was shown to reduce evoked release of Glu via a presynaptic mechanism [448]. Cloning of group III receptors has shown that (*S*)-AP4 (**200**) is an agonist at these receptors, some 10-fold more potent than Glu, and that many of the effects of this ligand in the CNS can be ascribed to activation of group III

receptors [416, 418, 419, 449]. Again, the homolog, (*S*)-AP5 (**201**), shows a rather different pharmacological profile from the parent compound. (*S*)-AP5 (**201**) also activates group III receptors, but with a markedly lower potency [447, 450]. However, it also antagonizes mGluR2, unlike the parent compound **200** [447, 451].

Taken together, the results from the homologs clearly demonstrate that the distance between the glycine moiety and the distal acidic group is of vital importance for the pharmacological activity. As pointed out, the selectivity profile of the homologs are often quite different from the parent compounds. Furthermore, in several cases, agonists have been converted into antagonists when the backbone chain length is increased. While the effects are not predictable at a glance, the results show that chain length is a parameter that is worth taking into consideration when analogs of newly identified mGluR "hits" are being designed.

*Bioisosteres* We have already described a number of compounds in which the distal carboxylic acid of Glu has been replaced by an acidic bioisostere (see Fig. 36). Compounds such as quisqualic acid (**140**) (bioisosteric group: 1,2,4-oxadiazol-3,5-dione), ibotenic acid (**100**), and (*S*)-homo-AMPA (**197**) (bioisosteric group: 3-isoxazolol) and (*S*)-AP4 (**200**) (bioisosteric group: phosphonic acid) show that all three groups of mGluRs are capable of accepting bioisosteres to replace the distal carboxyl group. The increased receptor selectivity shown by these compounds compared to Glu also demonstrates that the region of the receptor, in which the distal carboxylic acid binds, differs among the receptor subtypes. Thus, further research into new bioisosteres may be a fruitful path to new subtype-selective mGluR ligands. This argument is further substantiated by group I selective agonists employing phenol as the distal acidic group (see Fig. 37). For example, (*S*)-3,5-DHPG (**202**) is a selective agonist at mGluR1 and mGluR5 [452–454], equipotent with Glu on both group I subtypes.

### *Competitive Metabotropic Glutamate Receptor Antagonists*

*Phenylglycines* One of the first potent mGluR antagonists to be reported was (*S*)-4CPG (**203**), which antagonizes (1*S*,3*R*)-ACPD-induced (**185**) IP formation in cerebral cortical slices [455]. Likewise, (*S*)-4CPG (**203**) has been shown to antagonize IP formation by mGluR1, and furthermore, the compound seems to be an agonist [453] or antagonist [454] at mGluR2 with no effect on mGluR4. (*S*)-4CPG (**203**) has been used extensively as a template for designing further potent and selective antagonists at mGluR1. Whereas it has been debated whether (*S*)-4CPG (**203**) is an agonist or antagonist at mGluR2 [453, 454], it is generally agreed that (*S*)-4C3HPG (**204**) is an mGluR1 antagonist/mGluR2 agonist and that the α-methylated analog, (*S*)-M4CPG (**207**), is an antagonist at both subtypes [453, 454]. It has been shown that the antagonist potency is increased by methylation at the 2-position of the phenyl ring. Thus (+)-4C2MPG (**205**) and 4C3H2MPG (**206**) are both approximately five-

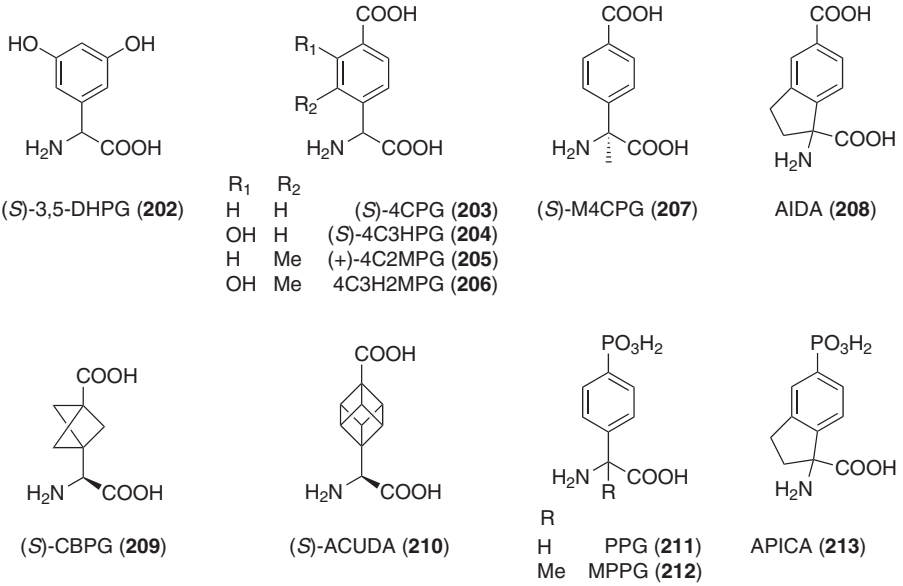


**Figure 37** Structures of some mGluR ligands.

fold more potent than the nonmethylated parent compounds [456]. AIDA (**208**), a compound that can be viewed as a cyclized analog of either **207** or **205**, has also been shown to possess mGluR1 antagonist activity, albeit with lower potency than the noncyclic compounds [457, 458]. The 4-carboxyphenylglycines discussed above show mGluR1 antagonist potencies in the 5- to 200-μM range. Interestingly, the potencies are somewhat affected by the agonist used to determine the functional potency [459]. Furthermore, it is notable that the 4-carboxyphenylglycines show selectivity for the mGluR1 subtype with nil or weak activities at the closely related mGluR5 subtype [442, 458–460]. One exception to this rule is the α-thioxanthylmethyl analog LY393675 (**223**), which has been shown to be equipotent at mGluR1 and mGluR5, with low-micromolar potency [461]. Likewise, (*S*)-hexyl-HIBO (**198**) is equipotent as an antagonist at mGluR1 and mGluR5 [446].

Based on modeling studies, it has been suggested that the antagonist action of the 4-carboxyphenylglycines is not caused by the phenyl ring in this series of compounds, but rather by the coplanar arrangement of the glycine moiety and the distal acidic group caused by the flat nonflexible phenyl ring [462]. In order to test this hypothesis, (*S*)-CBPG (**209**) and (*S*)-ACUDA (**210**) were synthesized [463, 464]. Both of these compounds retain the coplanar arrangement of the glycine moiety and the distal acidic group, and in agreement with their hypothesis, both compounds proved to be moderately potent mGluR1 antagonists with nil or weak activity at mGluR2, mGluR4, and mGluR5 [463, 464].

The phosphonic acid bioisostere of **203**, (PPG, **211**), is a potent group III agonist [465]. Based on the results from the 4-carboxyphenylglycines, the phosphonic acid bioisostere of **207**, MPPG (**212**) [466], and that of **208**, APICA (**213**) [467] have been designed. By analogy with results showing that α-methylation of **203** converts the compound from being an agonist into an antagonist at mGluR2 [267], **212** is an antagonist at group III mGluRs [467–469]. Compound **212** is slightly more potent as an antagonist at mGluR2, which is rather surprising given that neither (*S*)-AP4 (**200**) nor PPG (**211**) have any appreciable effect on this receptor subtype [467–469]. APICA (**213**) shows a pharmacological profile quite similar to that of MPPG (**212**), being slightly less potent as an mGluR2 antagonist than the parent compound [467].

α-*Substitutions*. As eluded to above, α-methylation has been widely used to derive antagonists from agonists. Some of these analogs are shown in Figure 38. Maintaining the selectivity profiles as of their antagonistic parent compounds, MAP4 (**214**) and MCCGI (**215**) antagonize mGluR2 and mGluR4, respectively, albeit with significantly reduced antagonist potency compared to the parent agonist [469, 470]. As mentioned above, α-methylation of the potent and selective group III agonist PPG (**211**), results in the mixed group II/III antagonist MPPG (**212**) [467–469]. Another example of swhich in group selectivity and loss of potency obtained by α-methylation is (*S*)-MetQUIS (**216**), which is a moderately potent antagonist at mGluR2 and a very weak antagonist at mGluR1 and mGluR5 [471]. Although some important compounds have been obtained, when the results of α-methylation are viewed as a whole, this does not seem to be a generally viable strategy for designing antagonists.

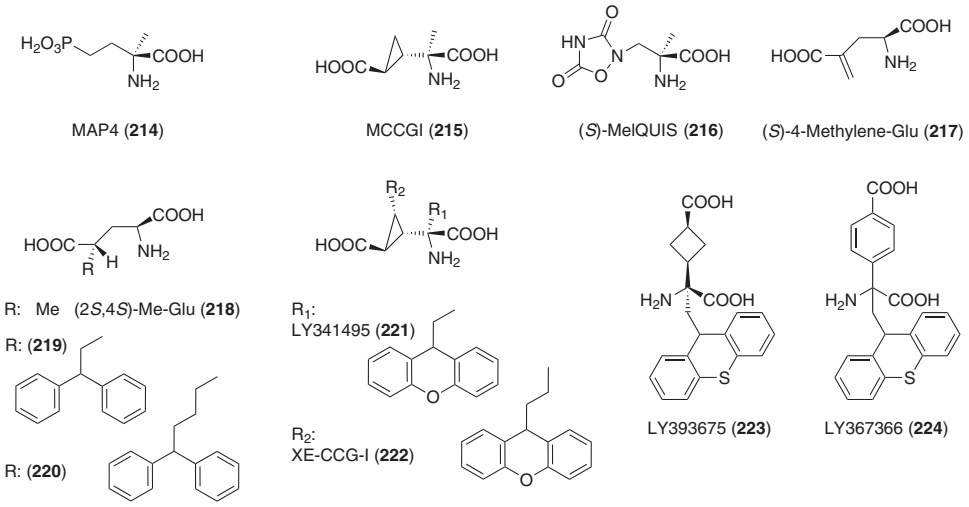


**Figure 38** Structure of some mGluR ligands.

Substituting agonists with bulky, lipophilic side chains has been a much more successful approach to the design of potent antagonists. Two of the pioneering compounds in this class are 4-substituted analogs of Glu such as **219** and **220**. Both compounds are specific antagonists for mGluR2 and mGluR3 with potencies in the low-micromolar range [472, 473]. Interestingly, compounds with small substituents in the same position, such as (2*S*,4*S*)-Me-Glu (**218**) or (*S*)-4-methylene-Glu (**217**), are more potent agonists at mGluR2 than Glu, with some activity at mGluR1 but without appreciable activity at mGluR4 [474]. Thus by increasing the bulk and lipophilicity at the 4-position of these fly-swatter substituents, the selectivity for group II is retained, and even increased, but the compounds are converted from agonists to antagonists. More importantly, by contrast with α-methylation, the potency of the antagonists is retained compared to the parent agonist compounds.

These results have inspired the design of a number of new potent antagonists with fly-swatter side chains. MCCGI (**215**), the methylated analog of **102** (Fig. 25), was first substituted with fly-swatters in the 2-position. It was found that the most potent compound was obtained with a xanthylmethyl subsituent conferring LY341495 (**221**), which displays antagonist activity at group II receptors in the low-nanomolar range [475, 476]. However, like the parent compound (2*S*,3*R*,4*S*)-CCG (**102**), LY341495 (**221**) also shows affinity for other subtypes, especially mGluR8 [477]. The synthesis and pharmacology of analogs with fly-swatter substituents at the 3′-position were recently reported. In this series, the compound with the xanthylethyl substituent, XE-CCG-I (**222**) was found to be the most potent antagonist, with potencies at mGluR2 and mGluR3 comparable to **221** [478]. Finally, the fly-swatter principle has been successfully applied to 3-carboxycyclobutylglycine, for example, LY393675 (**223**) [479] and to 4CPG (**203**), for example, LY367366 (**224**) [461]. Compounds **223** and **224**, respectively, show submicromolar and low-micromolar potency at both group I subtypes, which for the latter is quite remarkable given that most other phenylglycine analogs prefer the mGluR1 subtype [442, 458–461].

It can be concluded that in their antagonized state receptors from all three mGluR groups can accommodate quite large and lipophilic side chains in a variety of positions. Furthermore, compared with small α substituents such as methyl groups, which most often confer antagonists with reduced potency, the large fly-swatter substitutents in most cases confer antagonism and with increased potency. Thus the fly-swatter principle seems to be an attractive strategy for obtaining potent subtype-selective antagonists.

### *Allosteric Modulators of Metabotropic Glutamate Receptors*

*Noncompetitive Antagonists* In 1996, CPCCOEt (**225**), a non-amino-acid compound with no structural similarity with Glu, was reported as a group I selective antagonist [480]. Since then, it has been shown that the compound is a selective mGluR1 noncompetitive antagonist with no mGluR5 activity,

acting at the 7TM region rather than the agonist-binding ATD [481–483]. A number of other non-amino-acid mGluR antagonists have been discovered (see Fig. 39). BAY36-7620 (**226**) and EM-TBPC (**227**) are also mGluR1-specific antagonists acting at the 7TM domain with nanomolar potencies [484, 485]. The enol ether Ro 64-5229 (**228**) shows submicromolar noncompetitive antagonist potency at mGluR2 with no activity at mGluR1, mGluR3, mGluR4, mGluR5, NMDA, or AMPA receptors [486, 487].

SIB-1893 (**229**) and MPEP (**230**) have been reported [488, 489] to be selective, noncompetitive antagonists with nanomolar potencies at mGluR5, MPEP (**230**) being the most potent [488, 489]. Like CPCCOEt (**225**), MPEP (**230**) has been shown to act at the 7TM region rather than the agonist-binding ATD [490]. Recently, it was reported that SIB-1893 (**229**) and MPEP (**230**) also act as positive allosteric modulators of mGluR4, albeit with somewhat lower

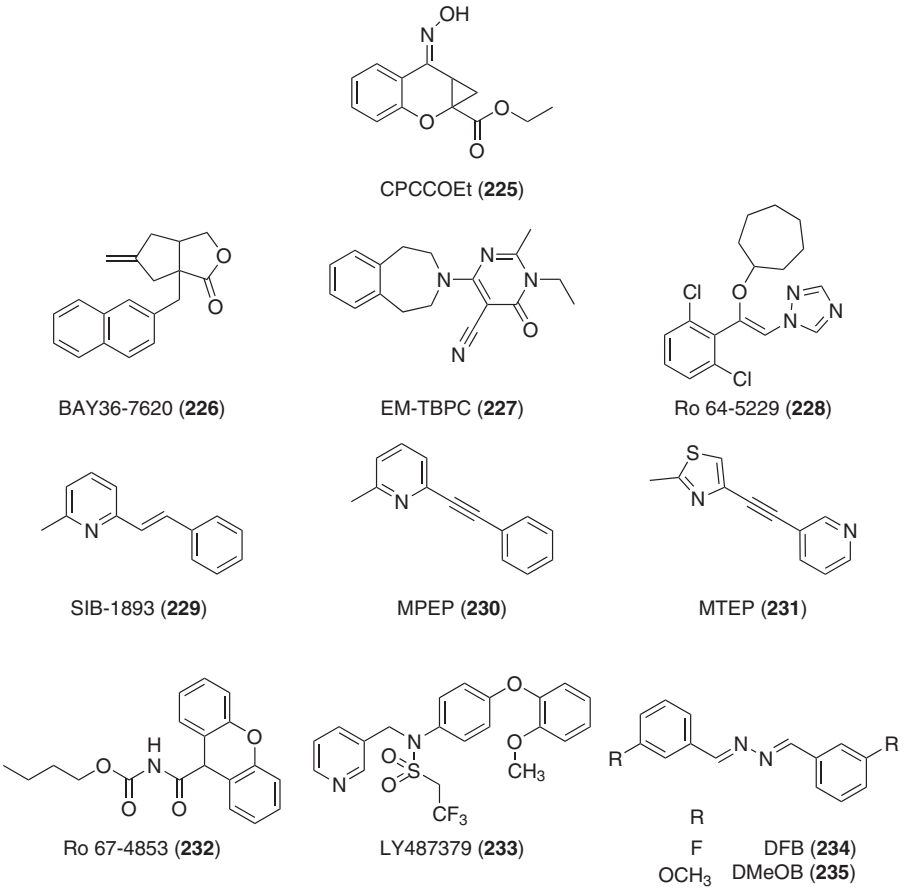


**Figure 39** Structures of some noncompetitive mGluR antagonists and positive allosteric modulators.

potency than the antagonism at mGluR5 [491]. Furthermore, MPEP (**230**) antagonizes NMDA receptors with low-micromolar potency [492], which has led to the design of the analog MTEP (**231**), which is slightly more potent than **230** as an antagonist at mGluR5 and shows no NMDA antagonist activity [492].

*Positive Allosteric Modulators* As already mentioned in the previous section, SIB-1893 (**229**) and MPEP (**230**) act as positive allosteric modulators at mGluR4. The allosteric effect is dependent upon Glu activation, and the compounds are thus unable to activate the mGluR4 receptor directly. Instead, the compounds enhance the response mediated by Glu, causing a left-ward shift of concentration–response curves and an increase in the maximum response [491]. Other compounds with similar pharmacological profiles at other mGluR subtypes have also been reported, including the mGluR1-specific positive allosteric modulator Ro 67-4853 (**232**) [220], the mGluR2-specific compound LY487379 (**233**) [493], and the mGluR5-specific compound DFB (**234**) [494]. Interestingly, DMeOB (**235**), which is a close analog of DFB, acts as a noncompetitive antagonist at mGluR5 [494], which together with the dual effects of SIB-1893 (**229**) and MPEP (**230**) at mGluR4 and mGluR5, illustrates the delicate balance of the activation mechanism of this class of G-protein-coupled receptors.

The allosteric modulators described in the preceding paragraphs have been discovered by high-throughput screening of compound libraries. It is interesting to note the diverse structures that have come out of these screens and that the compounds are highly selective even within the mGluR groups I, II, and III. As has been demonstrated for several of the compounds, most if not all of the noncompetitive antagonists act outside the agonist-binding pocket located in the ATD, which could very well be the explanation for the high degree of subtype selectivity. Given that the binding of the endogenous agonist Glu is of immense importance for proper receptor function, it is not surprising that this domain has been more highly conserved during evolution than the rest of the receptor. Likewise, it is not surprising that ligands acting outside the agonist-binding pocket show a higher degree of receptor selectivity than ligands that are structurally similar with Glu and act within the pocket. With this in mind, high-throughput screening is a useful step in the discovery of new subtype-specific ligands.

***Metabotropic Glutamate Receptors as Putative Drug Targets*** As previously described, group I mGluRs stimulate phopholipase C and thus lead to cell excitation, whereas group II and III mGluRs inhibit adenylate cyclase, and thus cause cell inhibition. It has also been described that Glu is involved in many neurodegenerative, psychiatric, and cognitive disorders characterized by either hyper- or hypofunction of the glutamatergic system. Together, these observations have led to the pursuit of group I antagonists and group II/III agonists as potential therapeutics for disorders characterized by glutamater-

gic hyperfunction and likewise group I agonists and group II/III antagonists for disorders characterized by glutamatergic hyperfunction. Testing these ideas was initially hampered by a lack of selective, potent ligands, but the recent development of such compounds [e.g., LY354740 (**191**) and the allosteric modulators described in the previous section] has greatly enhanced our repertoire and confirmed that the mGluRs do indeed hold promise as novel therapeutic targets. It is beyond the scope of this chapter to describe these results in detail, and the reader is thus referred to several recent articles that review the preclinical results of mGluR ligands as drug candidates [495–498].

## Structure of Metabotropic Glutamate Receptors

The current state of knowledge regarding the structural detail of mGluRs resembles that of iGluRs. Through recent breakthroughs in X-ray crystallography, we have obtained a series of snapshots of the extracellular binding domain of one of the subtypes at atomic resolution, mGluR1 [499, 500]. mGluRs belong to family C of the 7TM GPCRs, and at present, insight into the membrane-spanning region is limited to comparison with crystal structures such as that of rhodopsin, a family A GPCR [222]. There is strong evidence that mGluRs operate as homodimers, covalently bound via a disulfide bridge [501]. mGluRs are distinguished by the presence of a large extracellular N-terminal superdomain, which is homologous to a class of periplasmic binding proteins, such as L-arabinose binding protein [502]. Note the two large globular domains (I and II) connected by a hinge region and separated from the 7TM domain by a conserved cysteine-rich loop (Fig. 40).

While crystallization of the full-length receptor is unlikely in the near future, the first successful crystallization of an extracellular domain was reported in 2000, specifically mGluR1 in the presence and absence of the endogenous ligand Glu [398]. Of particular interest is the fact that according to these structures, the monomers can undergo a conformational change, or 70° "twist" with respect to one another. Furthermore, Glu induces domain closure of only one of the superdomains of the dimer, even though the ligand is bound to both ("closed–open", Fig. 41). Thus this crystal structure gives ligand design hints for both competitive agonists and antagonists and also suggests a role for the dimer interface. More recently [500], mGluR1 was co-crystallized with an antagonist, (*S*)-M4CPG (**207**), which confirms that competitive mGluR antagonists stabilize the "open–open" untwisted state. In addition, mGluR1 was co-crystallized in the presence of the positive modulator $Gd^{3+}$ and Glu, revealing a "closed–closed" twisted state. $Gd^{3+}$ appears to stabilize the twisted state via coordination to residues in the dimer interface, known previously to accommodate an engineered negatively modulatory $Zn^{2+}$ site [503]. This leads to the following picture of mGluR activation (Fig. 41), where the free energy of ligand binding induces domain closure and twisting,

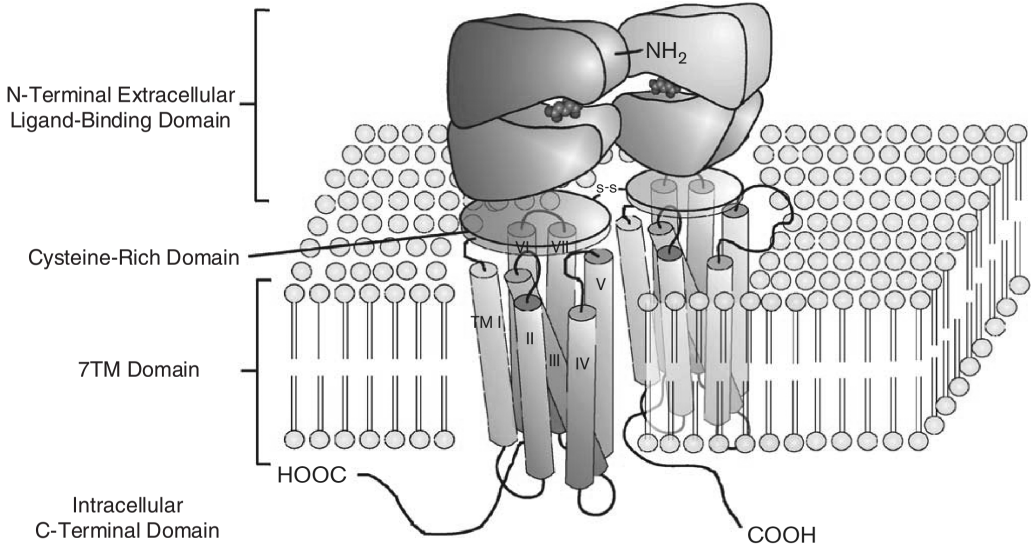


**Figure 40** Illustration of the topology of the mGluR dimer. The structure of the N-terminal super-domain has been established by X-ray crystallography for mGluR1, whereas the 7TM region has been homology modeled based on the structure of rhodopsin (illustration from [495]).

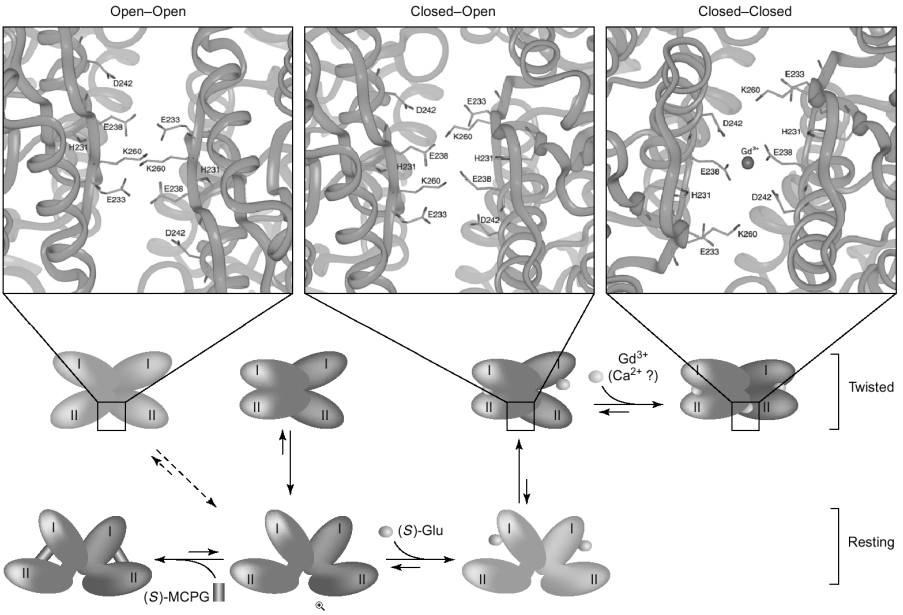


**Figure 41** Illustration of the induced twisting and domain closure of the extracellular portion of mGluR1, presumably coupled to signal transduction via the 7TM domain. The changes in the dimer interface are highlighted above

which is then mechanically transferred via the 7TM domain to the G-protein by as yet unknown transformations.

Although the binding site sequence homology between mGluR1 and the other mGluRs is lower than that seen among iGluRs (Table 3), a number of attempts have been made to model other members of the family with a view to structure-based design and explanations of subtype selectivity. Such models have been particularly useful in conjunction with site-directed mutagenesis and subsequent pharmacological assays [504–507]. Whereas the distal but not the α-carboxylate of Glu forms an ion pair with the domain-closed iGluRs, the situation at mGluR is reversed: The distal carboxylate binds to a positively charged hydrophilic receptor region. As with iGluRs, binding is mediated by water molecules involved in hydrogen bond networks, which may be disrupted or displaced by other ligands.

While Glu binds to mGluRs in an extended conformation, there may be minor differences in the binding mode between subgroups, especially in the positioning and orientation of the distal carboxylate [508]. Note that while the residues that bind the α-amino acid are highly conserved, there is very little conservation about the distal carboxylate (Fig. 42). The binding site differences between the three groups of mGluRs may be exploited relatively easily for the design of subtype-specific ligands. Within members of a given group,

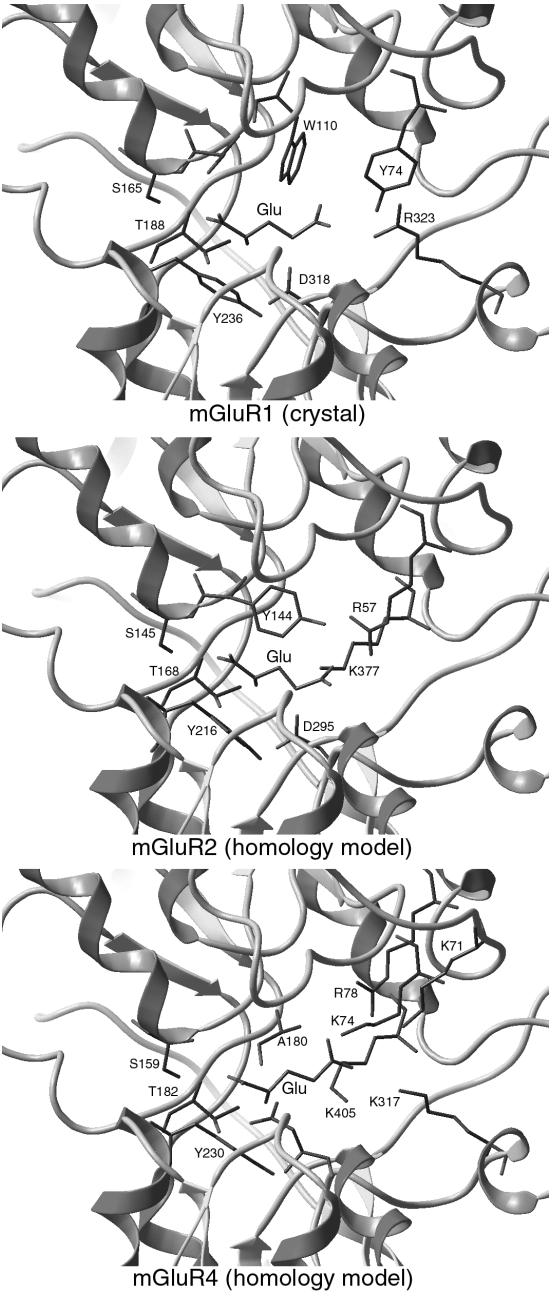


**Figure 42** Domain-closed binding sites of mGluR1, mGluR2, and mGluR4 as representatives of groups I, II, and III mGluRs. mGluR2 and mGluR4 have been modeled from mGluR1 ([504, 508] and M. B. Hermit and J. R. Greenwood, unpublished data).

**TABLE 3 Binding Site Homology of mGluRs**[a]

| Residue | 74 | 78 | 110 | 163 | 164 | 165 | 166 | 186 | 187 | 188 | 189 | 208 | 211 | 236 | 292 | 293 | 318 | 319 | 320 | 323 | 409 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| mGluR1 | **Y** | **R** | W | G | S | S | S | **S** | A | **T** | S | **D** | Q | **Y** | **E** | **G** | **D** | G | W | **R** | **K** |
| mGluR5 | Y | R | W | G | S | S | S | S | A | T | S | D | Q | Y | E | G | D | G | W | R | K |
| mGluR2 | R | R | S | S | Y | S | D | A | S | T | S | D | Q | Y | R | S | D | G | W | L | K |
| mGluR3 | R | R | S | S | Y | S | S | A | S | T | S | D | Q | Y | R | S | D | G | W | Q | K |
| mGluR4 | K | R | S | S | G | S | S | A | S | T | A | D | Q | Y | N | E | D | S | W | K | K |
| mGluR6 | Q | R | S | S | A | S | S | A | S | T | A | D | Q | Y | N | E | D | S | W | K | K |
| mGluR7 | N | R | S | S | G | S | S | A | S | T | A | D | Q | Y | N | D | D | S | W | K | K |
| mGluR8 | K | R | S | A | A | S | S | A | S | T | A | D | Q | Y | N | E | D | S | W | K | K |

[a] Residues within 6 Å of Glu bound to mGluR1 listed. Residues in bold known to interact with Glu in mGluR1. Numbering according to mGluR1 (M. B. Hermit, unpublished data).

binding site conservation renders large selectivity differences difficult to obtain for competitive ligands.

## REFERENCES


1. Bräuner-Osborne, H., Egebjerg, J., Nielsen, E. Ø., Madsen, U., Krogsgaard-Larsen, P. (2000). Ligands for glutamate receptors: Design and therapeutic prospects. *J. Med. Chem.*, *43*, 2609–2645.
2. Egebjerg, J., Schousboe, A., Krogsgaard-Larsen, P. (2002). *Glutamate and GABA Receptors and Transporters.* Taylor & Francis, London.
3. Barnard, E. A. (1992). Receptor classes and the transmitter-gated ion channels. *Trends. Biochem. Sci.*, *17*, 368–374.
4. Monaghan, D. T., Wenthold, R. J. (1997). *The Ionotropic Glutamate Receptors*. Humana Press, Totowa.
5. Johnston, G. A. (1996). GABAA receptor pharmacology. *Pharmacol. Ther.*, *69*, 173–198.
6. Krogsgaard-Larsen, P., Frølund, B., Jørgensen, F. S., Schousboe, A. (1994). $GABA_A$ receptor agonists, partial agonists and antagonists. Design and therapeutic prospects. *J. Med. Chem.*, *37*, 2489–2505.
7. Zefirova, O. N., Zefirov, N. S. (2000). Physiologically active compounds interacting with glutamate receptors. *Russ J. Org. Chem.*, *36*, 1231–1258.
8. Carlsson, M., Carlsson, A. (1990). Interactions between glutamatergic and monoaminergic systems within the basal ganglia—implications for schizophrenia and Parkinson's disease. *Trends Neurosci.*, *13*, 272–276.
9. Ebert, B., Madsen, U., Søby, K. K., Krogsgaard-Larsen, P. (1996). Functional partial agonism at ionotropic excitatory amino acid receptors. *Neurochem. Int.*, *29*, 309–316.
10. Dean, B., Hussain, T., Hayes, W., Scarr, E., Kitsoulis, S., Hill, C., Opeskin, K., Copolov, D. L. (1999). Changes in serotonin2A and GABA(A) receptors in schizophrenia: Studies on the human dorsolateral prefrontal cortex. *J. Neurochem.*, *72*, 1593–1599.
11. Ulas, J., Cotman, C. W. (1993). Excitatory amino acid receptors in schizophrenia. *Schizophrenia Bull.*, *19*, 105–117.
12. Lancel, M., Faulhaber, J. (1996). The GABAA agonist THIP (gaboxadol) increases non-REM sleep and enhances delta activity in the rat. *Neuroreport*, *7*, 2241–2245.
13. Sherif, F., Ahmed, S. S., Eriksson, L. (1995). Brain aminobutyrate aminotransferase activity in Alzheimer's disease. *Neurodegeneration*, *4*, 114–115.
14. Sherif, F. M., Ahmed, S. S. (1995). Basic aspects of GABA-transaminase in neuropsychiatric disorders. *Clin. Biochem.*, *28*, 145–154.
15. Jones-Davis, D. M., Macdonald, R. L. (2003). GABA(A) receptor function and pharmacology in epilepsy and status epilepticus. *Curr. Opin. Pharmacol.*, *3*, 12–18.
16. Choi, S., Silverman, R. B. (2002). Inactivation and inhibition of gamma-aminobutyric acid aminotransferase by conformationally restricted vigabatrin analogues. *J. Med. Chem.*, *45*, 4531–4539.

17. Choi, S., Storici, P., Schirmer, T., Silverman, R. B. (2002). Design of a conformationally restricted analogue of the antiepilepsy drug Vigabatrin that directs its mechanism of inactivation of gamma-aminobutyric acid aminotransferase. *J. Am. Chem. Soc.*, *124*, 1620–1624.
18. Johnson, T. R., Silverman, R. B. (1999). Syntheses of (*Z*)- and (*E*)-4-amino-2-(trifluoromethyl)-2-butenoic acid and their inactivation of gamma-aminobutyric acid aminotransferase. *Bioorg. Med. Chem.*, *7*, 1625–1636.
19. Nanavati, S. M., Silverman, R. B. (1989). Design of potential anticonvulsant agents: Mechanistic classification of GABA aminotransferase inactivators. *J. Med. Chem.*, *32*, 2413–2421.
20. Qiu, J., Silverman, R. B. (2000). A new class of conformationally rigid analogues of 4-amino-5-halopentanoic acids, potent inactivators of gamma-aminobutyric acid aminotransferase. *J. Med. Chem.*, *43*, 706–720.
21. Storici, P., Capitani, G., De Biase, D., Moser, M., John, R. A., Jansonius, J. N., Schirmer, T. (1999). Crystal structure of GABA-aminotransferase, a target for antiepileptic drug therapy. *Biochemistry*, *38*, 8628–8634.
22. Bouclier, M., Jung, M. J., Lippert, B. (1979). Stereochemistry of reactions catalysed by mammalian-brain L-glutamate 1-carboxy-lyase and 4-aminobutyrate: 2-Oxoglutarate aminotransferase. *Eur. J. Biochem.*, *98*, 363–368.
23. Bialer, M., Johannessen, S. I., Kupferberg, H. J., Levy, R. H., Loiseau, P., Perucca, E. (2002). Progress report on new antiepileptic drugs: A summary of the Sixth Eilat Conference (EILAT VI). *Epilepsy Res.*, *51*, 31–71.
24. Dewey, S. L., Morgan, A. E., Ashby, C. R., Jr., Horan, B., Kushner, S. A., Logan, J., Volkow, N. D., Fowler, J. S., Gardner, E. L., Brodie, J. D. (1998). A novel strategy for the treatment of cocaine addiction. *Synapse*, *30*, 119–129.
25. Gerasimov, M. R., Schiffer, W. K., Brodie, J. D., Lennon, I. C., Taylor, S. J., Dewey, S. L. (2000). Gamma-aminobutyric acid mimetic drugs differentially inhibit the dopaminergic response to cocaine. *Eur. J. Pharmacol.*, *395*, 129–135.
26. Stromberg, M. F., Mackler, S. A., Volpicelli, J. R., O'Brien, C. P., Dewey, S. L. (2001). The effect of gamma-vinyl-GABA on the consumption of concurrently available oral cocaine and ethanol in the rat. *Pharmacol. Biochem. Behav.*, *68*, 291–299.
27. Paul, S. R., Krauss, G. L., Miller, N. R., Medura, M. T., Miller, T. A., Johnson, M. A. (2001). Visual function is stable in patients who continue long-term vigabatrin therapy: Implications for clinical decision making. *Epilepsia*, *42*, 525–530.
28. Nousiainen, I., Mantyjarvi, M., Kalviainen, R. (2001). No reversion in vigabatrin-associated visual field defects. *Neurology*, *57*, 1916–1917.
29. Schiffer, W. K., Marsteller, D., Dewey, S. L. (2003). Sub-chronic low dose gamma-vinyl GABA (vigabatrin) inhibits cocaine-induced increases in nucleus accumbens dopamine. *Psychopharmacol. (Berl.)*, *168*, 339–343.
30. Qiu, J., Pingsterhaus, J. M., Silverman, R. B. (1999). Inhibition and substrate activity of conformationally rigid vigabatrin analogues with gamma-aminobutyric acid aminotransferase. *J. Med. Chem.*, *42*, 4725–4728.
31. Ashby, C. R., Jr., Paul, M., Gardner, E. L., Gerasimov, M. R., Dewey, S. L., Lennon, I. C., Taylor, S. J. (2002). Systemic administration of 1*R*,4*S*-4-amino-cyclopent-2-

ene-carboxylic acid, a reversible inhibitor of GABA transaminase, blocks expression of conditioned place preference to cocaine and nicotine in rats. *Synapse*, *44*, 61–63.

32. Borden, L. A. (1996). GABA transporter heterogeneity: Pharmacology and cellular localization. *Neurochem. Int.*, *29*, 335–356.
33. Ribak, C. E., Tong, W. M., Brecha, N. C. (1996). Astrocytic processes compensate for the apparent lack of GABA transporters in the axon terminals of cerebellar Purkinje cells. *Anat. Embryol. (Berl.)*, *194*, 379–390.
34. Ribak, C. E., Tong, W. M., Brecha, N. C. (1996). GABA plasma membrane transporters, GAT-1 and GAT-3, display different distributions in the rat hippocampus. *J. Comp. Neurol.*, *367*, 595–606.
35. Johnson, J., Chen, T. K., Rickman, D. W., Evans, C., Brecha, N. C. (1996). Multiple gamma-aminobutyric acid plasma membrane transporters (GAT-1, GAT-2, GAT-3) in the rat retina. *J. Comp. Neurol.*, *375*, 212–224.
36. Yang, C. Y., Brecha, N. C., Tsao, E. (1997). Immunocytochemical localization of gamma-aminobutyric acid plasma membrane transporters in the tiger salamander retina. *J. Comp. Neurol.*, *389*, 117–126.
37. De Biasi, S., Vitellaro-Zuccarello, L., Brecha, N. C. (1998). Immunoreactivity for the GABA transporter-1 and GABA transporter-3 is restricted to astrocytes in the rat thalamus. A light and electron-microscopic immunolocalization. *Neuroscience*, *83*, 815–828.
38. Minelli, A., DeBiasi, S., Brecha, N. C., Zuccarello, L. V., Conti, F. (1996). GAT-3, a high-affinity GABA plasma membrane transporter, is localized to astrocytic processes, and it is not confined to the vicinity of GABAergic synapses in the cerebral cortex. *J. Neurosci.*, *16*, 6255–6264.
39. Ebert, U., Krnjevic, K. (1990). Systemic CI-966, a new gamma-aminobutyric acid uptake blocker, enhances gamma-aminobutyric acid action in CA1 pyramidal layer in situ. *Can. J. Physiol. Pharmacol.*, *68*, 1194–1199.
40. Richards, D. A., Bowery, N. G. (1996). Comparative effects of the GABA uptake inhibitors, tiagabine and NNC-711, on extracellular GABA levels in the rat ventrolateral thalamus. *Neurochem. Res.*, *21*, 135–140.
41. Leker, R. R., Neufeld, M. Y. (2003). Anti-epileptic drugs as possible neuroprotectants in cerebral ischemia. *Brain Res. Rev.*, *42*, 187–203.
42. Rekling, J. C. (2003). Neuroprotective effects of anticonvulsants in rat hippocampal slice cultures exposed to oxygen/glucose deprivation. *Neurosci. Lett.*, *335*, 167–170.
43. O'Connell, A. W., Fox, G. B., Kjoller, C., Gallagher, H. C., Murphy, K. J., Kelly, J., Regan, C. M. (2001). Anti-ischemic and cognition-enhancing properties of NNC-711, a gamma-aminobutyric acid reuptake inhibitor. *Eur. J. Pharmacol.*, *424*, 37–44.
44. Backonja, M. M. (2002). Use of anticonvulsants for treatment of neuropathic pain. *Neurology*, *59*, S14–S17.
45. Krogsgaard-Larsen, P., Frølund, B., Frydenvang, K. (2000). GABA uptake inhibitors. Design, molecular pharmacology and therapeutic aspects. *Curr. Pharm. Des.*, *6*, 1193–1209.

46. Johnston, G. A., Curtis, D. R., Beart, P. M., Game, C. J., McCulloch, R. M., Twitchin, B. (1975). Cis- and trans-4-aminocrotonic acid as GABA analogues of restricted conformation. *J. Neurochem.*, *24*, 157–160.
47. Johnston, G. A. R., Krogsgaard-Larsen, P., Stephanson, A. L., Twitchin, B. (1976). Inhibition of the uptake of GABA and related amino acids in rat brain slices by the optical isomers of nipecotic acid. *J. Neurochem.*, *26*, 1029–1032.
48. Krogsgaard-Larsen, P., Johnston, G. A. R. (1975). Inhibition of GABA uptake in rat brain slices by nipecotic acid, various isoxazoles and related compounds. *J. Neurochem.*, *25*, 797–802.
49. Schousboe, A., Thorbek, P., Hertz, L., Krogsgaard-Larsen, P. (1979). Effects of GABA analogues of restricted conformation on GABA transport in astrocytes and brain cortex slices and on GABA receptor binding. *J. Neurochem.*, *33*, 181–189.
50. Krogsgaard-Larsen, P., Falch, E., Larsson, O. M., Schousboe, A. (1987). GABA uptake inhibitors: Relevance to antiepileptic drug research. *Epilepsy Res.*, *1*, 77–93.
51. Ali, F. E., Bondinell, W. E., Dandridge, P. A., Frazee, J. S., Garvey, E., Girard, G. R., Kaiser, C., Ku, T. W., Lafferty, J. J., Moonsammy, G. I., Oh, H. J., Rush, J. A., Setler, P. E., Stringer, O. D., Venslavsky, J. W., Volpe, B. W., Yunger, L. M., Zirkle, C. L. (1985). Orally active and potent inhibitors of "gamma"-aminobutyric acid uptake. *J. Med. Chem.*, *28*, 653–660.
52. Genton, P., Guerrini, R., Perucca, E. (2001). Tiagabine in clinical practice. *Epilepsia*, *42* (Suppl. 3), 42–45.
53. Andersen, K. E., Lau, J., Lundt, B. F., Petersen, H., Huusfeldt, P. O., Suzdak, P. D., Swedberg, M. D. (2001). Synthesis of novel GABA uptake inhibitors. Part 6: Preparation and evaluation of N-Omega asymmetrically substituted nipecotic acid derivatives. *Bioorg. Med. Chem.*, *9*, 2773–2785.
54. Andersen, K. E., Sorensen, J. L., Huusfeldt, P. O., Knutsen, L. J., Lau, J., Lundt, B. F., Petersen, H., Suzdak, P. D., Swedberg, M. D. (1999). Synthesis of novel GABA uptake inhibitors. 4. Bioisosteric transformation and successive optimization of known GABA uptake inhibitors leading to a series of potent anticonvulsant drug candidates. *J. Med. Chem.*, *42*, 4281–4291.
55. Andersen, K. E., Sorensen, J. L., Lau, J., Lundt, B. F., Petersen, H., Huusfeldt, P. O., Suzdak, P. D., Swedberg, M. D. (2001). Synthesis of novel gamma-aminobutyric acid (GABA) uptake inhibitors. 5. (1) Preparation and structure-activity studies of tricyclic analogues of known GABA uptake inhibitors. *J. Med. Chem.*, *44*, 2152–2163.
56. Knutsen, L. J., Andersen, K. E., Lau, J., Lundt, B. F., Henry, R. F., Morton, H. E., Naerum, L., Petersen, H., Stephensen, H., Suzdak, P. D., Swedberg, M. D., Thomsen, C., Sorensen, P. O. (1999). Synthesis of novel GABA uptake inhibitors. 3. Diaryloxime and diarylvinyl ether derivatives of nipecotic acid and guvacine as anticonvulsant agents. *J. Med. Chem.*, *42*, 3447–3462.
57. Falch, E., Krogsgaard-Larsen, P. (1989). GABA uptake inhibitors containing mono- and diarylmethoxyalkyl N-substituents. *Drug. Des. Delivery.*, *4*, 205–215.
58. White, H. S., Sarup, A., Bolvig, T., Kristensen, A. S., Petersen, G., Nelson, N., Pickering, D. S., Larsson, O. M., Frølund, B., Krogsgaard-Larsen, P., Schousboe, A. (2002). Correlation between anticonvulsant activity and inhibitory action on glial

gamma-aminobutyric acid uptake of the highly selective mouse gamma-aminobutyric acid transporter 1 inhibitor 3-hydroxy-4-amino-4,5,6,7-tetrahydro-1,2-benzisoxazole and its N-alkylated analogs. *J. Pharmacol. Exp. Ther.*, *302*, 636–644.

59. Falch, E., Perregaard, J., Frølund, B., Søkilde, B., Buur, A., Hansen, L. M., Frydenvang, K., Brehm, L., Bolvig, T., Larsson, O. M., Sanchez, C., White, H. S., Schousboe, A., Krogsgaard-Larsen, P. (1999). Selective inhibitors of glial GABA uptake: Synthesis, absolute stereochemistry and pharmacology of the enantiomers of 3-hydroxy-4-amino-4,5,6,7-tetrahydro-1,2-benzisoxazole (exo-THPO) and analogues. *J. Med. Chem.*, *42*, 5402–5414.
60. Borden, L. A., Dhar, T. G., Smith, K. E., Branchek, T. A., Gluchowski, C., Weinshank, R. L. (1994). Cloning of the human homologue of the GABA transporter GAT-3 and identification of a novel inhibitor with selectivity for this site. *Receptors Channels*, *2*, 207–213.
61. Dhar, T. G., Borden, L. A., Tyagarajan, S., Smith, K. E., Branchek, T. A., Weinshank, R. L., Gluchowski, C. (1994). Design, synthesis and evaluation of substituted triarylnipecotic acid derivatives as GABA uptake inhibitors: Identification of a ligand with moderate affinity and selectivity for the cloned human GABA transporter GAT-3. *J. Med. Chem.*, *37*, 2334–2342.
62. Thomsen, C., Sørensen, P. O., Egebjerg, J. (1997). 1-(3-(9*H*-carbazol-9-yl)-1-propyl)-4-(2-methoxyphenyl)-4-piperidinol, a novel subtype selective inhibitor of the mouse type II GABA-transporter. *Br. J. Pharmacol.*, *120*, 983–985.
63. Dalby, N. O. (2000). GABA-level increasing and anticonvulsant effects of three different GABA uptake inhibitors. *Neuropharmacology*, *39*, 2399–2407.
64. Macdonald, R. L., Olsen, R. W. (1994). GABAA receptor channels. *Annu. Rev. Neurosci.*, *17*, 569–602.
65. Sieghart, W. (1995). Structure and pharmacology of "gamma"-aminobutyric acidA receptor subtypes. *Pharmacol. Rev.*, *47*, 181–234.
66. Doble, A., Martin, I. L. (1996). *The $GABA_A$/Benzodiazepine Receptor as a Target for Psychoactive Drugs*. Springer-Verlag, Heidelberg.
67. Barnard, E. A., Skolnick, P., Olsen, R. W., Mohler, H., Sieghart, W., Biggio, G., Braestrup, C., Bateson, A. N., Langer, S. Z. (1998). International Union of Pharmacology. XV. Subtypes of gamma-aminobutyric $acid_A$ receptors: classification on the basis of subunit structure and receptor function. *Pharmacol. Rev.*, *50*, 291–313.
68. Chebib, M., Johnston, G. A. R. (2000). GABA-Activated ligand gated ion channels: Medicinal chemistry and molecular biology. *J. Med. Chem.*, *43*, 1427–1447.
69. Korpi, E. R., Grunder, G., Luddens, H. (2002). Drug interactions at GABA(A) receptors. *Prog. Neurobiol.*, *67*, 113–159.
70. Frølund, B., Ebert, B., Kristiansen, U., Liljefors, T., Krogsgaard-Larsen, P. (2002). $GABA_A$ Receptor ligands and their therapeutic potentials. *Curr. Med. Chem.*, *2*, 817–832.
71. Krogsgaard-Larsen, P., Brehm, L., Schaumburg, K. (1981). Muscimol, a psychoactive constituent of *Amanita muscaria*, as a medicinal chemical model structure. *Acta. Chem. Scand. Ser B*, *B35*, 311–324.

72. Krogsgaard-Larsen, P., Hjeds, H., Curtis, D. R., Lodge, D., Johnston, G. A. R. (1979). Dihydromuscimol, thiomuscimol and related heterocyclic compounds as GABA analogues. *J. Neurochem.*, *32*, 1717–1724.
73. Fowler, L. J., Lovell, D. H., John, R. A. (1983). Reaction of muscimol with 4-aminobutyrate aminotransferase. *J. Neurochem.*, *41*, 1751–1754.
74. Krogsgaard-Larsen, P., Johnston, G. A. R., Lodge, D., Curtis, D. R. (1977). A new class of GABA agonist. *Nature*, *268*, 53–55.
75. Kendall, D. A., Browner, M., Enna, S. J. (1982). Comparison of the antinociceptive effect of gamma-aminobutyric acid (GABA) agonists: Evidence for a cholinergic involvement. *J. Pharmacol. Exp. Ther.*, *220*, 482–487.
76. Krogsgaard-Larsen, P., Brehm, L., Johansen, J. S., Vinzents, P., Lauridsen, J., Curtis, D. R. (1985). Synthesis and structure-activity studies on excitatory amino acids structurally related to ibotenic acid. *J. Med. Chem.*, *28*, 673–679.
77. Krogsgaard-Larsen, P., Roldskov-Christiansen, T. (1979). GABA agonists. Synthesis and structure-activity studies on analogs of isoguvacine and THIP. *Eur. J. Med. Chem. Chim. Ther.*, *14*, 157–164.
78. Curtis, D. R., Duggan, A. W., Felix, D., Johnston, G. A. R. (1970). GABA, bicuculline and central inhibition. *Nature*, *226*, 1222–1224.
79. Johnston, G. A. R., Beart, P. M., Curtis, D. R., Game, C. J. A., McCulloch, R. M., MacLachlan, R. M. (1972). Bicuculline methochloride as a GABA antagonist. *Nature (New Biol.)*, *240*, 219–220.
80. Wermuth, C. G., Biziére, K. (1986). Pyridazinyl-GABA derivatives: A new class of synthetic $GABA_A$ antagonists. *Trends Pharmacol. Sci.*, *7*, 421–424.
81. Wermuth, C. G., Bourguignon, J.-J., Schlewer, G., Gies, J.-P., Schoenfelder, A., Melikian, A., Bouchet, M.-J., Chantreux, D., Molimard, J.-C., Heaulme, M., Chambon, J.-P., Biziere, K. (1987). Synthesis and structure-activity relationships of a series of aminopyridazine derivatives of γ-aminobutyric acid acting as selective GABA-A antagonists. *J. Med. Chem.*, *30*, 239–249.
82. Byberg, J. R., Labouta, I. M., Falch, E., Hjeds, H., Krogsgaard-Larsen, P., Curtis, D. R., Gynther, B. D. (1987). Synthesis and biological activity of a GABA-A agonist which has no effect on benzodiazepine binding and structurally related glycine antagonists. *Drug. Des. Delivery.*, *1*, 261–274.
83. Frølund, B., Kristiansen, U., Brehm, L., Hansen, A. B., Krogsgaard-Larsen, P., Falch, E. (1995). Partial $GABA_A$ receptor agonists. Synthesis and in vitro pharmacology of a series of nonannulated analogs of 4,5,6,7-tetrahydroisoxazolo[4,5-*c*]pyridin-3-ol (THIP). *J. Med. Chem.*, *38*, 3287–3296.
84. Mortensen, M., Frølund, B., Jørgensen, A. T., Liljefors, T., Krogsgaard-Larsen, P., Ebert, B. (2002). Activity of novel 4-PIOL analogues at human $\alpha_1\beta_2\gamma_{2\sigma}$ $GABA_A$ receptors—correlation with hydrophobicity. *Eur. J. Pharmacol.*, *451*, 125–132.
85. Kristiansen, U., Lambert, J. D. C., Falch, E., Krogsgaard-Larsen, P. (1991). Electrophysiological studies of the $GABA_A$ receptor ligand, 4-PIOL, on cultured hippocampal neurones. *Br. J. Pharmacol.*, *104*, 85–90.
86. Ebert, B., Mortensen, M., Thompson, S. A., Kehler, J., Wafford, K. A., Krogsgaard-Larsen, P. (2001). Bioisosteric determinants for subtype selectivity of ligands for heteromeric $GABA_A$ receptors. *Bioorg. Med. Chem. Lett.*, *11*, 1573–1577.

87. Ebert, B., Thompson, S. A., Suonatsou, K., McKernan, R., Krogsgaard-Larsen, P., Wafford, K. A. (1997). Differences in agonist/antagonist binding affinity and receptor transduction using recombinant human gamma-aminobutyric acid type A receptors. *Mol. Pharmacol.*, *52*, 1150–1156.
88. Hansen, S. L., Ebert, B., Fjalland, B., Kristiansen, U. (2001). Effects of $GABA_A$ receptor partial agonists in primary cultures of cerebellar granule neurons and cerebral cortical neurons reflect different receptor subunit compositions. *Br. J. Pharmacol.*, *133*, 539–549.
89. Frølund, B., Jørgensen, A. T., Tagmose, L., Stensbøl, T. B., Vestergaard, H. T., Engblom, C., Kristiansen, U., Sanchez, C., Krogsgaard-Larsen, P., Liljefors, T. (2002). A novel class of potent 4-arylalkyl substituted 3-isoxazolol $GABA_A$ antagonists: synthesis, pharmacology and molecular modeling. *J. Med. Chem.*, *45*, 2454–2468.
90. Ebert, B., Wafford, K. A., Whiting, P. J., Krogsgaard-Larsen, P., Kemp, J. A. (1994). Molecular pharmacology of γ-aminobutyric acid type A receptor agonists and partial agonists in oocytes injected with different α, β and γ receptor subunit combinations. *Mol. Pharmacol.*, *46*, 957–963.
91. Verdoorn, T. A. (1994). Formation of heteromeric gamma-aminobutyric acid type A receptors containing two different alpha subunits. *Mol. Pharmacol.*, *45*, 475–480.
92. Adkins, C. E., Pillai, G. V., Kerby, J., Bonnert, T. P., Haldon, C., McKernan, R. M., Gonzalez, J. E., Oades, K., Whiting, P. J., Simpson, P. B. (2001). Alpha4beta3delta GABA(A) receptors characterized by fluorescence resonance energy transfer-derived measurements of membrane potential. *J. Biol. Chem.*, *276*, 38934–38939.
93. Brown, N., Kerby, J., Bonnert, T. P., Whiting, P. J., Wafford, K. A. (2002). Pharmacological characterization of a novel cell line expressing human alpha(4)beta(3)delta GABA(A) receptors. *Br. J. Pharmacol.*, *136*, 965–974.
94. Saxena, N. C., Macdonald, R. L. (1996). Properties of putative cerebellar gamma-aminobutyric acid A receptor isoforms. *Mol. Pharmacol.*, *49*, 567–579.
95. Pirker, S., Schwarzer, C., Wieselthaler, A., Sieghart, W., Sperk, G. (2000). GABA(A) receptors: immunocytochemical distribution of 13 subunits in the adult rat brain. *Neuroscience*, *101*, 815–850.
96. Sur, C., Farrar, S. J., Kerby, J., Whiting, P. J., Atack, J. R., McKernan, R. M. (1999). Preferential coassembly of α4 and δ subunits of the gamma-aminobutyric $acid_A$ receptor in rat thalamus. *Mol. Pharmacol.*, *56*, 110–115.
97. Nusser, Z., Mody, I. (2002). Selective modulation of tonic and phasic inhibitions in dentate gyrus granule cells. *J. Neurophysiol.*, *87*, 2624–2628.
98. Storustovu, S., Ebert, B. (2003). Gaboxadol: In vitro interaction studies with benzodiazepines and ethanol suggest functional selectivity. *Eur. J. Pharmacol.*, *467*, 49–56.
99. McKernan, R. M., Rosahl, T. W., Reynolds, D. S., Sur, C., Wafford, K. A., Atack, J. R., Farrar, S., Myers, J., Cook, G., Ferris, P., Garrett, L., Bristow, L., Marshall, G., Macaulay, A., Brown, N., Howell, O., Moore, K. W., Carling, R. W., Street, L. J., Castro, J. L., Ragan, C. I., Dawson, G. R., Whiting, P. J. (2000). Sedative but not anxiolytic properties of benzodiazepines are mediated by the $GABA_A$ receptor α1 subtype. *Nat. Neurosci.*, *3*, 587–592.

100. Smit, A. B., Syed, N. I., Schaap, D., van Minnen, J., Klumperman, J., Kits, K. S., Lodder, H., van der Schors, R. C., van Elk, R., Sorgedrager, B., Brejc, K., Sixma, T. K., Geraerts, W. P. (2001). A glia-derived acetylcholine-binding protein that modulates synaptic transmission. *Nature*, *411*, 261–268.

101. Brejc, K., van Dijk, W. J., Klaassen, R. V., Schuurmans, M., van Der Oost, J., Smit, A. B., Sixma, T. K. (2001). Crystal structure of an ACh-binding protein reveals the ligand-binding domain of nicotinic receptors. *Nature*, *411*, 269–276.

102. Cromer, B. A., Morton, C. J., Parker, M. W. (2002). Anxiety over GABA(A) receptor structure relieved by AChBP. *Trends Biochem. Sci.*, *27*, 280–287.

103. Trudell, J. R. (2002). Unique assignment of inter-subunit association in $GABA_A$ $\alpha_1 \beta_3 \gamma_2$ receptors by molecular modelling. *Biochim. Biophys. Acta*, *1565*, 91–96.

104. Ernst, M., Brauchart, D., Boresch, S., Sieghart, W. (2003). comparative modeling of GABA(A) receptors: Limits, insights, future developments. *Neuroscience*, *119*, 933–943.

105. Liljefors, T., Pettersson, I. (2002). Computer-aided development and use of three-dimensional pharmacophore models. In P. Krogsgaard-Larsen, T. Liljefors, and U. Madsen (Eds.), *Textbook of Drug Design and Discovery*. Taylor and Francis, pp. 86–116, London.

106. Güner, O. F. (2000). *Pharmacophore Perception, Development, and Use in Drug Design.* International University Line, La Jolla.

107. Galvez-Ruano, E., Aprison, M. J., Robertson, D. H., Lipkowitz, K. B. (1995). Identifying agonistic and antagonistic mechanisms operative at the GABA receptor. *J. Neurosci. Res.*, *42*, 666–673.

108. Buur, J. R. B., Hjeds, H., Krogsgaard-Larsen, P., Jørgensen, F. S. (1993). Conformational analysis and molecular modelling of a partial GABAA agonist and a glycine antagonist related to the GABAA agonist, THIP. *Drug Des. Discov.*, *10*, 213–229.

109. Krogsgaard-Larsen, P., Mikkelsen, H., Jacobsen, P., Falch, E., Curtis, D. R., Peet, M. J., Leah, J. D. (1983). 4,5,6,7-Tetrahydroisothiazolo[5,4-*c*]pyridin-3-ol and related analogues of THIP. Synthesis and biological activity. *J. Med. Chem.*, *26*, 895–900.

110. Brehm, L., Frydenvang, K., Hansen, L. M., Norrby, P.-O., Krogsgaard-Larsen, P., Liljefors, T. (1997). Structural features of muscimol, a potent $GABA_A$ receptor agonist. Crystal structure and quantum chemical ab initio calculations. *Struct. Chem.*, *8*, 443–451.

111. Krogsgaard-Larsen, P., Johnston, G. A. R. (1978). Structure-activity studies on the inhibition of GABA binding to rat brain membranes by muscimol and related compounds. *J. Neurochem.*, *30*, 1377–1382.

112. Tagmose, L. (2000). A pharmacophore model for $GABA_A$ receptor agonists. Ph.D. Thesis, Royal Danish School of Pharmacy, Copenhagen.

113. Frølund, B., Tagmose, L., Liljefors, T., Stensbøl, T. B., Engblom, C., Kristiansen, U., Krogsgaard-Larsen, P. (2000). A novel class of potent 3-isoxazolol $GABA_A$ antagonists: Design, synthesis, and pharmacology. *J. Med. Chem.*, *43*, 4930–4933.

114. Tagmose, L., Hansen, L. M., Norrby, P.-O., Liljefors, T. (2000). Differences in agonist binding pattern for the $GABA_A$ and the AMPA receptors illustrated by high-level ab initio calculations. In K. Gundertofte and F. S. Jørgensen (Eds.),

*Molecular Modelling and Prediction of Bioactivity*. Kluwer Academic/Plenum Publishers, pp. 365–366, New York.

115. Westh-Hansen, S. E., Witt, M. R., Dekermendjian, K., Liljefors, T., Rasmussen, P. B., Nielsen, M. (1999). Arginine residue 120 of the human $GABA_A$ receptor $\alpha_1$, subunit is essential for GABA binding and chloride ion current gating. *NeuroReport*, *10*, 2417–2421.
116. Boileau, A. J., Evers, A. R., Davis, A. F., Czajkowski, C. (1999). Mapping the agonist binding site of the GABAA receptor: Evidence for a beta-strand. *J. Neurosci.*, *19*, 4847–4854.
117. Hartvig, L., Lükensmejer, B., Liljefors, T., Dekermendjian, K. (2000). Two conserved arginines in the extracellular N-terminal domain of the GABA(A) receptor alpha(5) subunit are crucial for receptor function. *J. Neurochem.*, *75*, 1746–1753.
118. Nutt, D. J., Malizia, A. L. (2001). New insights into the role of the GABA(A)-benzodiazepine receptor in psychiatric disorder. *Br. J. Psychiatry*, *179*, 390–396.
119. Mohler, H., Fritschy, J. M., Rudolph, U. (2002). A new benzodiazepine pharmacology. *J. Pharmacol. Exp. Ther.*, *300*, 2–8.
120. Atack, J. R. (2003). Anxioselective compounds acting at the GABAA receptor benzodiazepine binding site. *Curr. Drug. Target CNS Neurol. Disord.*, *2*, 213–232.
121. Braestrup, C., Schmiechen, R., Neef, G., Nielsen, M., Petersen, E. N. (1982). Interaction of convulsive ligands with benzodiazepine receptors. *Science*, *216*, 1241–1243.
122. Arbilla, S., Allen, J., Wick, A., Langer, S. Z. (1986). High affinity [$^3$H]zolpidem binding in the rat brain: An imidazopyridine with agonist properties at central benzodiazepine receptors. *Eur. J. Pharmacol.*, *130*, 257–263.
123. Katsifis, A., Mardon, K., Mattner, F., Loc'h, C., McPhee, M. E., Dikic, B., Kassiou, M., Ridley, D. D. (2003). Pharmacological evaluation of (*S*)-8-[(123)I]iodobretazenil: A radioligand for in vivo studies of central benzodiazepine receptors. *Nucl. Med. Biol.*, *30*, 191–198.
124. Quirk, K., Blurton, P., Fletcher, S., Leeson, P., Tang, F., Mellilo, D., Ragan, C. I., McKernan, R. M. (1996). [$^3$H]L-655,708, a novel ligand selective for the benzodiazepine site of GABAA receptors which contain the alpha 5 subunit. *Neuropharmacology*, *35*, 1331–1335.
125. Farrar, S. J., Whiting, P. J., Bonnert, T. P., McKernan, R. M. (1999). Stoichiometry of a ligand-gated ion channel determined by fluorescence energy transfer. *J. Biol. Chem.*, *274*, 10100–10104.
126. Jensen, M. L., Timmermann, D. B., Johansen, T. H., Schousboe, A., Varming, T., Ahring, P. K. (2002). The beta subunit determines the ion selectivity of the GABAA receptor. *J. Biol. Chem.*, *277*, 41438–41447.
127. Smith, A. J., Alder, L., Silk, J., Adkins, C., Fletcher, A. E., Scales, T., Kerby, J., Marshall, G., Wafford, K. A., McKernan, R. M., Atack, J. R. (2001). Effect of alpha subunit on allosteric modulation of ion channel function in stably expressed human recombinant gamma-aminobutyric acid(A) receptors determined using (36)Cl ion flux. *Mol. Pharmacol.*, *59*, 1108–1118.
128. Rudolph, U., Crestani, F., Benke, D., Brunig, I., Benson, J. A., Fritschy, J. M., Martin, J. R., Bluethmann, H., Mohler, H. (1999). Benzodiazepine actions

mediated by specific gamma-aminobutyric acid$_A$ receptor subtypes. *Nature*, *401*, 796–800.

129. Crestani, F., Martin, J. R., Mohler, H., Rudolph, U. (2000). Mechanism of action of the hypnotic zolpidem in vivo. *Br. J. Pharmacol.*, *131*, 1251–1254.
130. Low, K., Crestani, F., Keist, R., Benke, D., Brunig, I., Benson, J. A., Fritschy, J. M., Rulicke, T., Bluethmann, H., Mohler, H., Rudolph, U. (2000). Molecular and neuronal substrate for the selective attenuation of anxiety. *Science*, *290*, 131–134.
131. Costa, E., Auta, J., Grayson, D. R., Matsumoto, K., Pappas, G. D., Zhang, X., Guidotti, A. (2002). GABAA receptors and benzodiazepines: A role for dendritic resident subunit mRNAs. *Neuropharmacology*, *43*, 925–937.
132. Navarro, J. F., Buron, E., Martin-Lopez, M. (2002). Anxiogenic-like activity of L-655,708, a selective ligand for the benzodiazepine site of GABA(A) receptors which contain the alpha-5 subunit, in the elevated plus-maze test. *Prog. Neuropsychopharmacol. Biol. Psychiatry*, *26*, 1389–1392.
133. Huang, Q., He, X., Ma, C., Liu, R., Yu, S., Dayer, C. A., Wenger, G. R., McKernan, R., Cook, J. M. (2000). Pharmacophore/receptor models for GABA(A)/BzR subtypes ($\alpha1\beta3\gamma2$, $\alpha5\beta3\gamma2$, and $\alpha6\beta3\gamma2$) via a comprehensive ligand-mapping approach. *J. Med. Chem.*, *43*, 71–95.
134. Thomet, U., Baur, R., Scholze, P., Sieghart, W., Sigel, E. (1999). Dual mode of stimulation by the beta-carboline ZK 91085 of recombinant GABA(A) receptor currents: Molecular determinants affecting its action. *Br. J. Pharmacol.*, *127*, 1231–1239.
135. Dubinsky, B., Vaidya, A. H., Rosenthal, D. I., Hochman, C., Crooke, J. J., DeLuca, S., DeVine, A., Cheo-Isaacs, C. T., Carter, A. R., Jordan, A. D., Reitz, A. B., Shank, R. P. (2002). 5-Ethoxymethyl-7-fluoro-3-oxo-1,2,3,5-Tetrahydrobenzo[4,5] imidazo[1,2*a*]pyridine-4-*N*-(2-fluorophenyl)carboxamide (RWJ-51204), a new nonbenzodiazepine anxiolytic. *J. Pharmacol. Exp. Ther.*, *303*, 777–790.
136. Griebel, G., Perrault, G., Simiand, J., Cohen, C., Granger, P., Depoortere, H., Francon, D., Avenet, P., Schoemaker, H., Evanno, Y., Sevrin, M., George, P., Scatton, B. (2003). SL651498, a GABAA receptor agonist with subtype-selective efficacy, as a potential treatment for generalized anxiety disorder and muscle spasms. *CNS Drug. Rev.*, *9*, 3–20.
137. Hood, S. D., Argyropoulos, S. V., Nutt, D. J. (2000). Agents in development for anxiety disorders. *CNS Drugs*, *6*, 421–431.
138. Sanna, E., Busonero, F., Talani, G., Carta, M., Massa, F., Peis, M., Maciocco, E., Biggio, G. (2002). Comparison of the effects of zaleplon, zolpidem, and triazolam at various GABA(A) receptor subtypes. *Eur. J. Pharmacol.*, *451*, 103–110.
139. Inui, Y., Yamamoto, M., Awasaki, Y., Nishida, N. (1996). A nonbenzodiazepine partial agonist, *S*-(+)-DN-2327, has minimal physical dependence-producing liability, but shows cross-dependence on barbital in rats. *Prog. Neuropsychopharmacol. Biol. Psychiatry*, *20*, 1197–1211.
140. Tang, A. H., Franklin, S. R., Carter, D. B., Sethy, V. H., Needham, L. M., Jacobsen, E. J., Von Voigtlander, P. F. (1997). Anxiolytic-like effects of PNU-101017, a partial agonist at the benzodiazepine receptor. *Psychopharmacol. (Berl.)*, *131*, 255–263.
141. Im, H. K., Im, W. B., Pregenzer, J. F., Stratman, N. C., VonVoigtlander, P. F., Jacobsen, E. J. (1998). Two imidazoquinoxaline ligands for the benzodiazepine site

sharing a second low affinity site on rat GABA(A) receptors but with the opposite functionality. *Br. J. Pharmacol.*, *123*, 1490–1494.

142. Albaugh, P. A., Marshall, L., Gregory, J., White, G., Hutchison, A., Ross, P. C., Gallagher, D. W., Tallman, J. F., Crago, M., Cassella, J. V. (2002). Synthesis and biological evaluation of 7,8,9,10-tetrahydroimidazo[1,2-*c*]pyrido[3,4-e]pyrimdin-5 (6*H*)-ones as functionally selective ligands of the benzodiazepine receptor site on the GABA(A) receptor. *J. Med. Chem.*, *45*, 5043–5051.
143. Savini, L., Chiasserini, L., Pellerano, C., Biggio, G., Maciocco, E., Serra, M., Cinone, N., Carrieri, A., Altomare, C., Carotti, A. (2001). High affinity central benzodiazepine receptor ligands. Part 2: Quantitative structure-activity relationships and comparative molecular field analysis of pyrazolo[4,3-*c*]quinolin-3-ones. *Bioorg. Med. Chem.*, *9*, 431–444.
144. Dekermendjian, K., Kahnberg, P., Witt, M.-R., Sterner, O., Nielsen, M., Liljefors, T. (1999). Structure-activity relationships and molecular modelling analysis of flavonoids binding to the benzodiazepine site of the rat brain GABAA receptor complex. *J. Med. Chem.*, *42*, 4343–4350.
145. Marder, M., Paladini, A. C. (2002). GABA(A)-receptor ligands of flavonoid structure. *Curr. Top. Med. Chem.*, *2*, 853–867.
146. Nielsen, M., Frokjaer, S., Braestrup, C. (1988). High affinity of the naturally-occurring biflavonoid, amentoflavon, to brain benzodiazepine receptors in vitro. *Biochem. Pharmacol.*, *37*, 3285–3287.
147. Sigel, E., Baur, R., Furtmueller, R., Razet, R., Dodd, R. H., Sieghart, W. (2001). Differential cross talk of ROD compounds with the benzodiazepine binding site. *Mol. Pharmacol.*, *59*, 1470–1477.
148. Yasar, S., Bergman, J., Munzar, P., Redhi, G., Tober, C., Knebel, N., Zschiesche, M., Paronis, C. (2003). Evaluation of the novel antiepileptic drug, AWD 131–138, for benzodiazepine-like discriminative stimulus and reinforcing effects in squirrel monkeys. *Eur. J. Pharmacol.*, *465*, 257–265.
149. Chambers, M. S., Atack, J. R., Broughton, H. B., Collinson, N., Cook, S., Dawson, G. R., Hobbs, S. C., Marshall, G., Maubach, K. A., Pillai, G. V., Reeve, A. J., MacLeod, A. M. (2003). Identification of a novel, selective GABA(A) alpha5 receptor inverse agonist which enhances cognition. *J. Med. Chem.*, *46*, 2227–2240.
150. Collins, I., Davey, W. B., Rowley, M., Quirk, K., Bromidge, F. A., McKernan, R. M., Thompson, S. A., Wafford, K. A. (2000). *N*-(indol-3-ylglyoxylyl)piperidines: High affinity agonists of human GABA-A receptors containing the alpha1 subunit. *Bioorg. Med. Chem. Lett.*, *10*, 1381–1384.
151. Primofiore, G., Settimo, F. D., Taliani, S., Marini, A. M., Novellino, E., Greco, G., Lavecchia, A., Besnard, F., Trincavelli, L., Costa, B., Martini, C. (2001). Novel *N*-(arylalkyl)indol-3-ylglyoxylylamides targeted as ligands of the benzodiazepine receptor: Synthesis, biological evaluation, and molecular modeling analysis of the structure-activity relationships. *J. Med. Chem.*, *44*, 2286–2297.
152. Akbarzadeh, T., Tabatabai, S. A., Khoshnoud, M. J., Shafaghi, B., Shafiee, A. (2003). Design and synthesis of 4*H*-3-(2-phenoxy)phenyl-1,2,4-triazole derivatives as benzodiazepine receptor agonists. *Bioorg. Med. Chem.*, *11*, 769–773.
153. Covey, D. F., Evers, A. S., Mennerick, S., Zorumski, C. F., Purdy, R. H. (2001). Recent developments in structure-activity relationships for steroid modulators of GABA(A) receptors. *Brain Res. Brain Res. Rev.*, *37*, 91–97.

154. Lambert, J. J., Belelli, D., Harney, S. C., Peters, J. A., Frenguelli, B. G. (2001). Modulation of native and recombinant GABA(A) receptors by endogenous and synthetic neuroactive steroids. *Brain Res. Rev.*, *37*, 68–80.

155. Lambert, J. J., Belelli, D., Hill-Venning, C., Peters, J. A. (1995). Neurosteroids and GABAA receptor function. *Trends Pharmacol. Sci.*, *16*, 295–303.

156. Krasowski, M. D., Harrison, N. L. (1999). General anaesthetic actions on ligand-gated ion channels. *Cell. Mol. Life Sci.*, *55*, 1278–1303.

157. Siegwart, R., Jurd, R., Rudolph, U. (2002). Molecular determinants for the action of general anesthetics at recombinant alpha(2)beta(3)gamma(2)gamma-aminobutyric acid(A) receptors. *J. Neurochem.*, *80*, 140–148.

158. Wingrove, P. B., Wafford, K. A., Bain, C., Whiting, P. J. (1994). The modulatory action of loreclezole at the gamma-aminobutyric acid type A receptor is determined by a single amino acid in the beta 2 and beta 3 subunit. *Proc. Natl. Acad. Sci. USA*, *91*, 4569–4573.

159. El Hadri, A., Abouabdellah, A., Thomet, U., Baur, R., Furtmuller, R., Sigel, E., Sieghart, W., Dodd, R. H. (2002). N-Substituted 4-amino-3,3-dipropyl-2(3*H*)-furanones: New positive allosteric modulators of the GABA(A) receptor sharing electrophysiological properties with the anticonvulsant loreclezole. *J. Med. Chem.*, *45*, 2824–2831.

160. Belelli, D., Lambert, J. J., Peters, J. A., Wafford, K., Whiting, P. J. (1997). The interaction of the general anesthetic etomidate with the gamma-aminobutyric acid type A receptor is influenced by a single amino acid. *Proc. Natl. Acad. Sci. USA*, *94*, 11031–11036.

161. Belelli, D., Muntoni, A. L., Merrywest, S. D., Gentet, L. J., Casula, A., Callachan, H., Madau, P., Gemmell, D. K., Hamilton, N. M., Lambert, J. J., Sillar, K. T., Peters, J. A. (2003). The in vitro and in vivo enantioselectivity of etomidate implicates the GABAA receptor in general anaesthesia. *Neuropharmacology*, *45*, 57–71.

162. Hamon, A., Morel, A., Hue, B., Verleye, M., Gillardin, J. M. (2003). The modulatory effects of the anxiolytic etifoxine on GABA(A) receptors are mediated by the beta subunit. *Neuropharmacology*, *45*, 293–303.

163. Halliwell, R. F., Thomas, P., Patten, D., James, C. H., Martinez-Torres, A., Miledi, R., Smart, T. G. (1999). Subunit-selective modulation of GABAA receptors by the non-steroidal anti-inflammatory agent, mefenamic acid. *Eur. J. Neurosci.*, *11*, 2897–2905.

164. Miller, J. A., Dudley, M. W., Kehne, J. H., Sorensen, S. M., Kane, J. M. (1992). MDL 26,479: A potential cognition enhancer with benzodiazepine inverse agonist-like properties. *Br. J. Pharmacol.*, *107*, 78–86.

165. Thompson, S. A., Wingrove, P. B., Connelly, L., Whiting, P. J., Wafford, K. A. (2002). Tracazolate reveals a novel type of allosteric interaction with recombinant gamma-aminobutyric acid(A) receptors. *Mol. Pharmacol.*, *61*, 861–869.

166. Krasowski, M. D., Hong, X., Hopfinger, A. J., Harrison, N. L. (2002). 4D-QSAR analysis of a set of propofol analogues: Mapping binding sites for an anesthetic phenol on the GABA(A) receptor. *J. Med. Chem.*, *45*, 3210–3221.

167. Meadows, H. J., Kumar, C. S., Pritchett, D. B., Blackburn, T. P., Benham, C. D. (1998). SB-205384: A GABA(A) receptor modulator with novel mechanism of action that shows subunit selectivity. *Br. J. Pharmacol.*, *123*, 1253–1259.

168. van Rijn, C. M., Willems-van Bree, E. (2003). Synergy between retigabine and GABA in modulating the convulsant site of the GABAA receptor complex. *Eur. J. Pharmacol.*, *464*, 95–100.

169. Robinson, R. T., Drafts, B. C., Fisher, J. L. (2003). Fluoxetine increases GABA(A) receptor activity through a novel modulatory site. *J. Pharmacol. Exp. Ther.*, *304*, 978–984.

170. Asproni, B., Pau, A., Bitti, M., Melosu, M., Cerri, R., Dazzi, L., Seu, E., Maciocco, E., Sanna, E., Busonero, F., Talani, G., Pusceddu, L., Altomare, C., Trapani, G., Biggio, G. (2002). Synthesis and pharmacological evaluation of 1-[(1,2-diphenyl-1*H*-4-imidazolyl)methyl]-4-phenylpiperazines with clozapine-like mixed activities at dopamine D(2), serotonin, and GABA(A) receptors. *J. Med. Chem.*, *45*, 4655–4668.

171. Hold, K. M., Sirisoma, N. S., Ikeda, T., Narahashi, T., Casida, J. E. (2000). Alpha-thujone (the active component of absinthe): Gamma-aminobutyric acid type A receptor modulation and metabolic detoxification. *Proc. Natl. Acad. Sci. USA*, *97*, 3826–3831.

172. Casida, J. E. (1993). Insecticide action at the GABA-gated chloride channel: Recognition, progress, and prospects. *Arch. Insect. Biochem. Physiol.*, *22*, 13–23.

173. Borea, P. A., Gilli, G., Bertolasi, V., Ferretti, V. (1987). Stereochemical features controlling binding and intrinsic activity properties of benzodiazepine-receptor ligands. *Mol. Pharmacol.*, *31*, 334–344.

174. Zhang, W., Koehler, K. F., Zhang, P., Cook, J. M. (1995). Development of a comprehensive pharmacophore model for the benzodiazepine receptor. *Drug. Des. Discov.*, *12*, 193–248.

175. He, X., Zhang, C., Cook, J. M. (2001). Model of the BzR binding site: Correlation of the data from site-directed mutagenesis and the pharmacophore/receptor model. *Med. Chem. Res.*, *10*, 269–308.

176. Hong, X., Hopfinger, A. J. (2003). 3D-pharmacophores of flavonoid binding at the benzodiazepine GABA(A) receptor site using 4D-QSAR analysis. *J. Chem. Inf. Comput. Sci.*, *43*, 324–336.

177. Huang, X., Liu, T., Gu, J., Luo, X., Ji, R., Cao, Y., Xue, H., Wong, J. T., Wong, B. L., Pei, G., Jiang, H., Chen, K. (2001). 3D-QSAR model of flavonoids binding at benzodiazepine site in GABAA receptors. *J. Med. Chem.*, *44*, 1883–1891.

178. Kahnberg, P., Lager, E., Rosenberg, C., Schougaard, J., Camet, L., Sterner, O., Ostergaard Nielsen, E., Nielsen, M., Liljefors, T. (2002). Refinement and evaluation of a pharmacophore model for flavone derivatives binding to the benzodiazepine site of the GABA(A) receptor. *J. Med. Chem.*, *45*, 4188–4201.

179. Marder, M., Estiu, G., Blanch, L. B., Viola, H., Wasowski, C., Medina, J. H., Paladini, A. C. (2001). Molecular modeling and QSAR analysis of the interaction of flavone derivatives with the benzodiazepine binding site of the GABA(A) receptor complex. *Bioorg. Med. Chem.*, *9*, 323–335.

180. Diaz-Arauzo, H., Evoniuk, G. E., Skolnick, P., Cook, J. M. (1991). The agonist pharmacophore of the benzodiazepine receptor. Synthesis of a selective anticonvulsant/anxiolytic. *J. Med. Chem.*, *34*, 1754–1756.

181. Diaz-Arauzo, H., Koehler, K. F., Hagen, T. J., Cook, J. M. (1991). Synthetic and computer assisted analysis of the pharmacophore for agonists at benzodiazepine receptors. *Life Sci.*, *49*, 207–216.

182. Cox, E. D., Diaz-Arauzo, H., Huang, Q., Reddy, M. S., Ma, C., Harris, B., McKernan, R., Skolnick, P., Cook, J. M. (1998). Synthesis and evaluation of analogues of the partial agonist 6-(propyloxy)-4-(methoxymethyl)-beta-carboline-3-carboxylic acid ethyl ester (6-PBC) and the full agonist 6-(benzyloxy)-4-(methoxymethyl)-beta-carboline-3-carboxylic acid ethyl ester (Zk 93423) at wild type and recombinant GABAA receptors. *J. Med. Chem.*, *41*, 2537–2552.

183. Da Settimo, A., Primofiore, G., Da Settimo, F., Marini, A. M., Novellino, E., Greco, G., Gesi, M., Martini, C., Giannaccini, G., Lucacchini, A. (1998). *N′*-Phenylindol-3-ylglyoxylohydrazide derivatives: Synthesis, structure-activity relationships, molecular modeling studies, and pharmacological action on brain benzodiazepine receptors. *J. Med. Chem.*, *41*, 3821–3830.

184. Da Settimo, A., Primofiore, G., Da Settimo, F., Marini, A. M., Novellino, E., Greco, G., Martini, C., Giannaccini, G., Lucacchini, A. (1996). Synthesis, structure-activity relationships, and molecular modeling studies of *N*-(indol-3-ylglyoxylyl)benzylamine derivatives acting at the benzodiazepine receptor. *J. Med. Chem.*, *39*, 5083–5091.

185. Kahnberg, P., Howard, M. H., Liljefors, T., Nielsen, M., Nielsen, E.ø., Sterner, O., Pettersson, I. (2004). The use of a pharmacophore model for identification of novel ligands for the benzodiazepine binding site of the $GABA_A$ receptor *J. Mol. Graph. Model.*, *23*, 253–261.

186. He, X., Huang, Q., Ma, C., Yu, S., McKernan, R., Cook, J. M. (2000). Pharmacophore/receptor models for GABA(A)/BzR $\alpha2\beta3\gamma2$, $\alpha3\beta3\gamma2$ and $\alpha4\beta3\gamma2$ recombinant subtypes. Included volume analysis and comparison to $\alpha1\beta3\gamma2$, $\alpha5\beta3\gamma2$, and $\alpha6\beta3\gamma2$ subtypes. *Drug. Des. Discov.*, *17*, 131–171.

187. Filizola, M., Harris, D. L., Loew, G. H. (2000). Development of a 3D pharmacophore for nonspecific ligand recognition of $\alpha1$, $\alpha2$, $\alpha3$, $\alpha5$, and $\alpha6$ containing GABA(A)/benzodiazepine receptors. *Bioorg. Med. Chem.*, *8*, 1799–1807.

188. Allen, M. S., LaLoggia, A. J., Dorn, L. J., Martin, M. J., Costantino, G., Hagen, T. J., Koehler, K. F., Skolnick, P., Cook, J. M. (1992). Predictive binding of beta-carboline inverse agonists and antagonists via the CoMFA/GOLPE approach. *J. Med. Chem.*, *35*, 4001–4010.

189. Allen, M. S., Tan, Y. C., Trudell, M. L., Narayanan, K., Schindler, L. R., Martin, M. J., Schultz, C., Hagen, T. J., Koehler, K. F., Codding, P. W., et al. (1990). Synthetic and computer-assisted analyses of the pharmacophore for the benzodiazepine receptor inverse agonist site. *J. Med. Chem.*, *33*, 2343–2357.

190. Bowery, N. G., Hill, D. R., Hudson, A. L., Doble, A., Middlemiss, D. N., Shaw, J., Turnbull, M. (1980). (−)-Baclofen decreases neurotransmitter release in the mammalian CNS by an action at a novel GABA receptor. *Nature*, *283*, 92–94.

191. Hill, D. R., Bowery, N. G. (1981). $^3$H-Baclofen and $^3$H-GABA bind to bicuculline-insensitive $GABA_B$ sites in rat brain. *Nature*, *290*, 149–152.

192. Bowery, N. G., Bettler, B., Froestl, W., Gallagher, J. P., Marshall, F., Raiteri, M., Bonner, T. I., Enna, S. J. (2002). International Union of Pharmacology. XXXIII.

Mammalian gamma-aminobutyric acid(B) receptors: Structure and function. *Pharmacol. Rev.*, *54*, 247–264.

193. Kaupmann, K., Huggel, K., Heid, J., Flor, P. J., Bischoff, S., Mickel, S. J., McMaster, G., Angst, C., Bittiger, H., Froestl, W., Bettler, B. (1997). Expression cloning of GABA(B) receptors uncovers similarity to metabotropic glutamate receptors. *Nature*, *386*, 239–246.
194. Marshall, F. H., Jones, K. A., Kaupmann, K., Bettler, B. (1999). GABAB receptors—the first 7TM heterodimers. *Trends Pharmacol. Sci.*, *20*, 396–399.
195. Kaupmann, K., Malitschek, B., Schuler, V., Heid, J., Froestl, W., Beck, P., Mosbacher, J., Bischoff, S., Kulik, A., Shigemoto, R., Karschin, A., Bettler, B. (1998). GABA(B)-receptor subtypes assemble into functional heteromeric complexes. *Nature*, *396*, 683–687.
196. Jones, K. A., Borowsky, B., Tamm, J. A., Craig, D. A., Durkin, M. M., Dai, M., Yao, W. J., Johnson, M., Gunwaldsen, C., Huang, L. Y., Tang, C., Shen, Q., Salon, J. A., Morse, K., Laz, T., Smith, K. E., Nagarathnam, D., Noble, S. A., Branchek, T. A., Gerald, C. (1998). GABA(B) receptors function as a heteromeric assembly of the subunits GABA(B)R1 and GABA(B)R2. *Nature*, *396*, 674–679.
197. White, J. H., Wise, A., Main, M. J., Green, A., Fraser, N. J., Disney, G. H., Barnes, A. A., Emson, P., Foord, S. M., Marshall, F. H. (1998). Heterodimerization is required for the formation of a functional GABA(B) receptor. *Nature*, *396*, 679–682.
198. Kuner, R., Kohr, G., Grunewald, S., Eisenhardt, G., Bach, A., Kornau, H. C. (1999). Role of heteromer formation in GABAB receptor function. *Science*, *283*, 74–77.
199. Malitschek, B., Schweizer, C., Keir, M., Heid, J., Froestl, W., Mosbacher, J., Kuhn, R., Henley, J., Joly, C., Pin, J. P., Kaupmann, K., Bettler, B. (1999). The N-terminal domain of gamma-aminobutyric acid(B) receptors is sufficient to specify agonist and antagonist binding. *Mol. Pharmacol.*, *56*, 448–454.
200. Kniazeff, J., Galvez, T., Labesse, G., Pin, J. P. (2002). No ligand binding in the GB2 subunit of the GABA(B) receptor is required for activation and allosteric interaction between the subunits. *J. Neurosci.*, *22*, 7352–7361.
201. Margeta-Mitrovic, M., Jan, Y. N., Jan, L. Y. (2000). A trafficking checkpoint controls GABA(B) receptor heterodimerization. *Neuron*, *27*, 97–106.
202. Pagano, A., Rovelli, G., Mosbacher, J., Lohmann, T., Duthey, B., Stauffer, D., Ristig, D., Schuler, V., Meigel, I., Lampert, C., Stein, T., Prezeau, L., Blahos, J., Pin, J., Froestl, W., Kuhn, R., Heid, J., Kaupmann, K., Bettler, B. (2001). C-terminal interaction is essential for surface trafficking but not for heteromeric assembly of GABA(b) receptors. *J. Neurosci.*, *21*, 1189–1202.
203. Calver, A. R., Robbins, M. J., Cosio, C., Rice, S. Q., Babbs, A. J., Hirst, W. D., Boyfield, I., Wood, M. D., Russell, R. B., Price, G. W., Couve, A., Moss, S. J., Pangalos, M. N. (2001). The C-terminal domains of the GABA(b) receptor subunits mediate intracellular trafficking but are not required for receptor signaling. *J. Neurosci.*, *21*, 1203–1210.
204. Galvez, T., Duthey, B., Kniazeff, J., Blahos, J., Rovelli, G., Bettler, B., Prezeau, L., Pin, J. P. (2001). Allosteric interactions between GB1 and GB2 subunits are required for optimal GABA(B) receptor function. *EMBO J.*, *20*, 2152–2159.

205. Margeta-Mitrovic, M., Jan, Y. N., Jan, L. Y. (2001). Function of GB1 and GB2 subunits in G protein coupling of GABA(B) receptors. *Proc. Natl. Acad. Sci. USA*, *98*, 14649–14654.

206. Havlickova, M., Prezeau, L., Duthey, B., Bettler, B., Pin, J. P., Blahos, J. (2002). The intracellular loops of the GB2 subunit are crucial for G-protein coupling of the heteromeric gamma-aminobutyrate B receptor. *Mol. Pharmacol.*, *62*, 343–350.

207. Duthey, B., Caudron, S., Perroy, J., Bettler, B., Fagni, L., Pin, J. P., Prezeau, L. (2002). A single subunit (GB2) is required for G-protein activation by the heterodimeric GABA(B) receptor. *J. Biol. Chem.*, *277*, 3236–3241.

208. Robbins, M. J., Calver, A. R., Filippov, A. K., Hirst, W. D., Russell, R. B., Wood, M. D., Nasir, S., Couve, A., Brown, D. A., Moss, S. J., Pangalos, M. N. (2001). GABA(B2) is essential for G-protein coupling of the GABA(B) receptor heterodimer. *J. Neurosci.*, *21*, 8043–8052.

209. Campbell, S. K., Almeida, G. L., Penn, R. D., Corcos, D. M. (1995). The effects of intrathecally administered baclofen on function in patients with spasticity. *Phys. Ther.*, *75*, 352–362.

210. Froestl, W., Mickel, S. J., Hall, R. G., von Sprecher, G., Strub, D., Baumann, P. A., Brugger, F., Gentsch, C., Jaekel, J., Olpe, H. R., Rihs, G., Vassout, A., Waldmeier, P. C., Bittiger, H. (1995). Phosphinic acid analogues of GABA. 1. New potent and selective GABAB agonists. *J. Med. Chem.*, *38*, 3297–3312.

211. Kerr, D. I., Ong, J., Prager, R. H., Gynther, B. D., Curtis, D. R. (1987). Phaclofen: A peripheral and central baclofen antagonist. *Brain Res.*, *405*, 150–154.

212. Kerr, D. I., Ong, J., Johnston, G. A., Abbenante, J., Prager, R. H. (1989). Antagonism at GABAB receptors by saclofen and related sulphonic analogues of baclofen and GABA. *Neurosci. Lett.*, *107*, 239–244.

213. Kerr, D. I., Ong, J., Johnston, G. A., Abbenante, J., Prager, R. H. (1988). 2-Hydroxysaclofen: An improved antagonist at central and peripheral GABAB receptors. *Neurosci. Lett.*, *92*, 92–96.

214. Froestl, W., Mickel, S. J., von Sprecher, G., Diel, P. J., Hall, R. G., Maier, L., Strub, D., Melillo, V., Baumann, P. A., Bernasconi, R., Gentsch, C., Hauser, K., Jaekel, J., Karlsson, G., Klebs, K., Maltre, L., Mondadori, C., Olpe, H. R., Walmeier, P. C., Bittiger, H. (1995). Phosphinic acid analogues of GABA. 2. Selective, orally active GABAB antagonists. *J. Med. Chem.*, *38*, 3313–3331.

215. Froestl, W., Bettler, B., Bittiger, H., Heid, J., Kaupmann, K., Mickel, S. J., Strub, D. (2003). Ligands for expression cloning and isolation of GABA(B) receptors. *Farmaco*, *58*, 173–183.

216. Jensen, A. A., Madsen, B. E., Krogsgaard-Larsen, P., Bräuner-Osborne, H. (2001). Pharmacological characterization of homobaclofen on wild type and mutant GABA(B)1b receptors coexpressed with the GABA(B)2 receptor. *Eur. J. Pharmacol.*, *417*, 177–180.

217. Galvez, T., Urwyler, S., Prezeau, L., Mosbacher, J., Joly, C., Malitschek, B., Heid, J., Brabet, I., Froestl, W., Bettler, B., Kaupmann, K., Pin, J. P. (2000). Ca(2+) requirement for high-affinity gamma-aminobutyric acid (GABA) binding at GABA(B) receptors: Involvement of serine 269 of the GABA(B)R1 subunit. *Mol. Pharmacol.*, *57*, 419–426.

218. Pin, J. P., Parmentier, M. L., Prezeau, L. (2001). Positive allosteric modulators for gamma-aminobutyric acid(B) receptors open new routes for the development of drugs targeting family 3 G-protein-coupled receptors. *Mol. Pharmacol.*, *60*, 881–884.
219. Urwyler, S., Mosbacher, J., Lingenhoehl, K., Heid, J., Hofstetter, K., Froestl, W., Bettler, B., Kaupmann, K. (2001). Positive allosteric modulation of native and recombinant gamma-aminobutyric acid(B) receptors by 2,6-di-*tert*-butyl-4-(3-hydroxy-2,2-dimethyl-propyl)-phenol (CGP7930) and its aldehyde analog CGP13501. *Mol. Pharmacol.*, *60*, 963–971.
220. Knoflach, F., Mutel, V., Jolidon, S., Kew, J. N., Malherbe, P., Vieira, E., Wichmann, J., Kemp, J. A. (2001). Positive allosteric modulators of metabotropic glutamate 1 receptor: Characterization, mechanism of action, and binding site. *Proc. Natl. Acad. Sci. USA*, *98*, 13402–13407.
221. Hauache, O. M., Hu, J., Ray, K., Xie, R., Jacobson, K. A., Spiegel, A. M. (2000). Effects of a calcimimetic compound and naturally activating mutations on the human Ca2+ receptor and on Ca2+ receptor/metabotropic glutamate chimeric receptors. *Endocrinology*, *141*, 4156–4163.
222. Jensen, A. A., Greenwood, J. R., Bräuner-Osborne, H. (2002). The dance of the clams: Twists and turns in the family C GPCR homodimer. *Trends Pharmacol. Sci.*, *23*, 491–493.
223. Drew, C. A., Johnston, G. A., Weatherby, R. P. (1984). Bicuculline-insensitive GABA receptors: Studies on the binding of (–)-baclofen to rat cerebellar membranes. *Neurosci. Lett.*, *52*, 317–321.
224. Shimada, S., Cutting, G., Uhl, G. R. (1992). Gamma-aminobutyric acid A or C receptor? Gamma-aminobutyric acid rho 1 receptor RNA induces bicuculline-, barbiturate-, and benzodiazepine-insensitive gamma-aminobutyric acid responses in *Xenopus* oocytes. *Mol. Pharmacol.*, *41*, 683–687.
225. Enz, R. (2001). GABA(C) receptors: A molecular view. *Biol. Chem.*, *382*, 1111–1122.
226. Zhang, D., Pan, Z. H., Awobuluyi, M., Lipton, S. A. (2001). Structure and function of GABA(C) receptors: A comparison of native versus recombinant receptors. *Trends Pharmacol. Sci.*, *22*, 121–132.
227. Cutting, G. R., Curristin, S., Zoghbi, H., O'Hara, B., Seldin, M. F., Uhl, G. R. (1992). Identification of a putative gamma-aminobutyric acid (GABA) receptor subunit rho2 cDNA and colocalization of the genes encoding rho2 (GABRR2) and rho1 (GABRR1) to human chromosome 6q14–q21 and mouse chromosome 4. *Genomics*, *12*, 801–806.
228. Cutting, G. R., Lu, L., O'Hara, B. F., Kasch, L. M., Montrose-Rafizadeh, C., Donovan, D. M., Shimada, S., Antonarakis, S. E., Guggino, W. B., Uhl, G. R., et al. (1991). Cloning of the gamma-aminobutyric acid (GABA) rho 1 cDNA: A GABA receptor subunit highly expressed in the retina. *Proc. Natl. Acad. Sci. USA*, *88*, 2673–2677.
229. Bailey, M. E., Albrecht, B. E., Johnson, K. J., Darlison, M. G. (1999). Genetic linkage and radiation hybrid mapping of the three human GABA(C) receptor rho subunit genes: GABRR1, GABRR2 and GABRR3. *Biochim. Biophys. Acta*, *1447*, 307–312.

230. Enz, R., Cutting, G. R. (1999). GABAC receptor rho subunits are heterogeneously expressed in the human CNS and form homo- and heterooligomers with distinct physical properties. *Eur. J. Neurosci.*, *11*, 41–50.

231. Boue-Grabot, E., Roudbaraki, M., Bascles, L., Tramu, G., Bloch, B., Garret, M. (1998). Expression of GABA receptor rho subunits in rat brain. *J. Neurochem.*, *70*, 899–907.

232. Johnston, G. A. (2002). Medicinal chemistry and molecular pharmacology of GABA(C) receptors. *Curr. Top. Med. Chem.*, *2*, 903–913.

233. Johnston, G. A. R., Krogsgaard-Larsen, P., Stephanson, A. (1975). Betel nut constituents as inhibitors of gamma-aminobutyric acid uptake. *Nature (London)*, *258*, 627–628.

234. Duke, R. K., Chebib, M., Balcar, V. J., Allan, R. D., Mewett, K. N., Johnston, G. A. (2000). (+)- and (–)-*cis*-2-Aminomethylcyclopropanecarboxylic acids show opposite pharmacology at recombinant rho(1) and rho(2) GABA(C) receptors. *J. Neurochem.*, *75*, 2602–2610.

235. Murata, Y., Woodward, R. M., Miledi, R., Overman, L. E. (1996). The first selective antagonist for a GABAC receptor. *Bioorg. Med. Chem. Lett.*, *6*, 2073–2076.

236. Ragozzino, D., Woodward, R. M., Murata, Y., Eusebi, F., Overman, L. E., Miledi, R. (1996). Design and in vitro pharmacology of a selective gamma-aminobutyric acidC receptor antagonist. *Mol. Pharmacol.*, *50*, 1024–1030.

237. Krehan, D., Frølund, B., Krogsgaard-Larsen, P., Kehler, J., Johnston, G. A., Chebib, M. (2003). Phosphinic, phosphonic and seleninic acid bioisosteres of isonipecotic acid as novel and selective GABA(C) receptor antagonists. *Neurochem. Int.*, *42*, 561–565.

238. Woodward, R. M., Polenzani, L., Miledi, R. (1993). Characterization of bicuculline/baclofen-insensitive (rho-like) gamma-aminobutyric acid receptors expressed in *Xenopus* oocytes. II. Pharmacology of gamma-aminobutyric acidA and gamma-aminobutyric acidB receptor agonists and antagonists. *Mol. Pharmacol.*, *43*, 609–625.

239. Qian, H., Dowling, J. E. (1993). Novel GABA responses from rod-driven retinal horizontal cells. *Nature*, *361*, 162–164.

240. Chebib, M., Vandenberg, R. J., Froestl, W., Johnston, G. A. (1997). Unsaturated phosphinic analogues of gamma-aminobutyric acid as GABA(C) receptor antagonists. *Eur. J. Pharmacol.*, *329*, 223–229.

241. Chebib, M., Mewett, K. N., Johnston, G. A. (1998). GABA(C) receptor antagonists differentiate between human ρ1 and ρ2 receptors expressed in *Xenopus* oocytes. *Eur. J. Pharmacol.*, *357*, 227–234.

242. Vien, J., Duke, R. K., Mewett, K. N., Johnston, G. A., Shingai, R., Chebib, M. (2002). *trans*-4-Amino-2-methylbut-2-enoic acid (2-MeTACA) and (+/–)-*trans*-2-aminomethylcyclopropanecarboxylic acid ((+/–)-TAMP) can differentiate rat ρ3 from human ρ1 and ρ2 recombinant GABA(C) receptors. *Br. J. Pharmacol.*, *135*, 883–890.

243. Kusama, T., Spivak, C. E., Whiting, P., Dawson, V. L., Schaeffer, J. C., Uhl, G. R. (1993). Pharmacology of GABA ρ 1 and GABA α/β receptors expressed in *Xenopus* oocytes and COS cells. *Br. J. Pharmacol.*, *109*, 200–206.

244. Kusama, T., Wang, T. L., Guggino, W. B., Cutting, G. R., Uhl, G. R. (1993). GABA ρ 2 receptor pharmacological profile: GABA recognition site similarities to ρ 1. *Eur. J. Pharmacol.*, *245*, 83–84.

245. McGillem, G. S., Rotolo, T. C., Dacheux, R. F. (2000). GABA responses of rod bipolar cells in rabbit retinal slices. *Vis. Neurosci.*, *17*, 381–389.

246. Matsui, K., Hasegawa, J., Tachibana, M. (2001). Modulation of excitatory synaptic transmission by GABA(C) receptor-mediated feedback in the mouse inner retina. *J. Neurophysiol.*, *86*, 2285–2298.

247. Dingledine, R., Borges, K., Bowie, D., Traynelis, S. F. (1999). The glutamate receptor ion channels. *Pharmacol. Rev.*, *51*, 7–61.

248. Hollmann, M., Heinemann, S. (1994). Cloned glutamate receptors. *Annu. Rev. Neurosci.*, *17*, 31–108.

249. Ransom, R. W., Stec, N. L. (1988). Cooperative modulation of [$^3$H]MK-801 binding to the *N*-methyl-d-aspartate receptor ion channel complex by l-glutamate, glycine and polyamines. *J. Neurochem.*, *51*, 830–836.

250. Lynch, D. R., Guttmann, R. P. (2002). Excitotoxicity: Perspectives based on *N*-methyl-d-aspartate receptor subtypes. *J. Pharmacol. Exp. Ther.*, *300*, 717–723.

251. Furukawa, H., Gouaux, E. (2003). Mechanisms of activation, inhibition and specificity: Crystal structures of the NMDA receptor NR1 ligand-binding core. *EMBO J.*, *22*, 2873–2885.

252. Tikhonova, I. G., Baskin, II, Palyulin, V. A., Zefirov, N. S., Bachurin, S. O. (2002). Structural basis for understanding structure-activity relationships for the glutamate binding site of the NMDA receptor. *J. Med. Chem.*, *45*, 3836–3843.

253. Lee, F. J., Xue, S., Pei, L., Vukusic, B., Chery, N., Wang, Y., Wang, Y. T., Niznik, H. B., Yu, X. M., Liu, F. (2002). Dual regulation of NMDA receptor functions by direct protein-protein interactions with the dopamine D1 receptor. *Cell*, *111*, 219–230.

254. Collingridge, G. L., Watkins, J. C. (1994). *The NMDA Receptor.* Oxford University Press, Oxford.

255. Lunn, W. H. W., Schoepp, D. D., Calligaro, D. O., Vasileff, R. T., Heinz, L. J., Salhoff, C. R., O'Malley, P. J. (1992). dl-Tetrazol-5-ylglycine, a highly potent NMDA agonist: Its synthesis and NMDA receptor efficacy. *J. Med. Chem.*, *35*, 4608–4612.

256. Madsen, U., Frydenvang, K., Ebert, B., Johansen, T. N., Brehm, L., Krogsgaard-Larsen, P. (1996). NMDA receptor agonists. Resolution, absolute stereochemistry and pharmacology of the enantiomers of 2-amino-2-(3-hydroxy-5-methyl-4-isoxazolyl)acetic acid (AMAA). *J. Med. Chem.*, *39*, 183–190.

257. Krogsgaard-Larsen, P., Ebert, B., Lund, T. M., Bräuner-Osborne, H., Sløk, F. A., Johansen, T. N., Brehm, L., Madsen, U. (1996). Design of excitatory amino acid receptor agonists, partial agonists and antagonists: Ibotenic acid as a key lead structure. *Eur. J. Med. Chem.*, *31*, 515–537.

258. Kyle, D. J., Patch, R. J., Karbon, E. W., Ferkany, J. W. (1992). NMDA receptors: Heterogeneity and agonism. In P. Krogsgaard-Larsen and J. J. Hansen (Eds.), *Excitatory Amino Acid Receptors: Design of Agonists and Antagonists.* Ellis Horwood, pp. 121–161, Chichester.

259. Jane, D. E., Olverman, H. J., Watkins, J. C. (1994). Agonists and competitive antagonists: Structure-activity and molecular modelling studies. In G. L.

Collingridge, J. C. Watkins (Eds.), *The NMDA Receptor.* Oxford University Press, pp. 31–104, Oxford.

260. Curry, K., Peet, M. J., Magnuson, D. S. K., McLennan, H. (1988). Synthesis, resolution, and absolute configuration of the isomers of the neuronal excitant 1-amino-1,3-cyclopentanedicarboxylic acid. *J. Med. Chem.*, *31*, 864–867.
261. Shinozaki, H., Ishida, M., Shimamoto, K., Ohfune, Y. (1989). A conformationally restricted analogue of l-glutamate, the (2*S*, 3*R*,4*S*) isomer of l-α-(carboxycyclopropyl)glycine, activates the NMDA-type receptor more markedly than NMDA in the isolated rat spinal cord. *Brain Res.*, *480*, 355–359.
262. Allan, R. D., Hanrahan, J. R., Hambley, T. W., Johnston, G. A. R., Mewett, K. N., Mitrovic, A. D. (1990). Synthesis and activity of a potent *N*-methyl-d-aspartic acid agonist, *trans*-1-aminocyclobutane-1,3-dicarboxylic acid, and related phosphonic and carboxylic acids. *J. Med. Chem.*, *33*, 2905–2915.
263. Ornstein, P. L., Klimkowski, V. J. (1992). Competitive NMDA receptor antagonists. In P. Krogsgaard-Larsen and J. J. Hansen (Eds.), *Excitatory Amino Acid Receptors: Design of Agonists and Antagonists.* Ellis Horwood, pp. 183–200, Chichester.
264. Fagg, G. E., Olpe, H.-R., Pozza, M. F., Baud, J., Steinmann, M., Schmutz, M., Portet, C., Baumann, P., Thedinga, K., Bittiger, H., Allgeier, H., Heckendorn, R., Angst, C., Brundish, D., Dingwall, J. G. (1990). CGP 37849 and CGP 39551: Novel and potent *N*-methyl-d-aspartate receptor antagonists with oral activity. *Br. J. Pharmacol.*, *99*, 791–797.
265. Lehmann, J., Hutchison, A. J., McPherson, S. E., Mondadori, C., Schmutz, M., Sinton, C. M., Tsai, C., Murphy, D. E., Steel, D. J., Williams, M. (1988). CGS 19755, a selective and competitive *N*-methyl-d-aspartate-type excitatory amino acid receptor antagonist. *J. Pharmacol. Exp. Ther.*, *246*, 65–75.
266. Lehmann, J., Schneider, J., McPherson, S., Murphy, D. E., Bernard, P., Tsai, C., Bennett, D. A., Pastor, G., Steele, D. J., Boehm, C., Cheney, D. L., Liebermann, J. M., Williams, M., Wood, P. L. (1987). CPP, a selective *N*-methyl-d-aspartate (NMDA)-type receptor antagonist: Characterization *in vitro* and *in vivo*. *J. Pharmacol. Exp. Ther.*, *240*, 737–746.
267. Hutchison, A. J., Williams, M., Angst, C., de Jesus, R., Blanchard, L., Jackson, R. H., Wilusz, E. J., Murphy, D. E., Bernard, P. S., Schneider, J., Campbell, T., Guida, W., Sills, M. A. (1989). 4-(Phosphonoalkyl)- and 4-(phosphonoalkenyl)-2-piperidinecarboxylic acids: Synthesis, activity at *N*-methyl-d-aspartic acid receptors, and anticonvulsant activity. *J. Med. Chem.*, *32*, 2171–2178.
268. Ornstein, P. L., Schoepp, D. D., Arnold, M. B., Leander, D., Lodge, D., Paschal, J. W., Elzey, T. (1991). 4-(Tetrazolylalkyl)piperidine-2-carboxylic acids. Potent and selective *N*-methyl-d-aspartic acid receptor antagonists with a short duration of action. *J. Med. Chem.*, *34*, 90–97.
269. Ferkany, J. W., Hamilton, G. S., Patch, R. J., Huang, Z., Borosky, S. A., Bednar, D. L., Jones, B. E., Zubrowski, R., Willetts, J., Karbon, E. W. (1993). Pharmacological profile of NPC 17742 [2*R*,4*R*,5*S*-(2-amino-4,5-(1, 2-cyclohexyl)-7-phosphonoheptanoic acid)], a potent, selective and competitive *N*-methyl-d-aspartate receptor antagonist. *J. Pharmacol. Exp. Ther.*, *264*, 256–264.
270. Baudy, R. B., Greenblatt, L. P., Jirkovsky, I. L., Conklin, M., Russo, R. J., Bramlett, D. R., Emrey, T. A., Simmonds, J. T., Kowal, D. M., Stein, R. P., Tasse, R. P. (1993). Potent quinoxaline-spaced phosphono alpha-amino acids of the AP-6

type as competitive NMDA antagonists: Synthesis and biological evaluation. *J. Med. Chem.*, *36*, 331–342.

271. Urwyler, S., Laurie, D., Lowe, D. A., Meier, C. L., Muller, W. (1996). Biphenyl-derivatives of 2-amino-7-phosphonoheptanoic acid, a novel class of potent competitive *N*-methyl-d-aspartate receptor antagonist—I. Pharmacological characterization in vitro. *Neuropharmacology*, *35*, 643–654.
272. Urwyler, S., Campbell, E., Fricker, G., Jenner, P., Lemaire, M., McAllister, K. H., Müller, W. (1996). Biphenyl-derivatives of 2-amino-7-phosphonoheptanoic acid, a novel class of potent competitive *N*-methyl-d-aspartate receptor antagonist—II. Pharmacological characterization in vivo. *Neuropharmacology*, *35*, 655–669.
273. Li, J. H., Bigge, C. F., Williamson, R. M., Borosky, S. A., Vartanian, M. G., Ortwine, D. F. (1995). Potent, orally active, competitive *N*-methyl-d-aspartate (NMDA) receptor antagonists are substrates for a neutral amino acid uptake system in Chinese hamster ovary cells. *J. Med. Chem.*, *38*, 1955–1965.
274. Anis, N. A., Berry, S. C., Burton, N. R., Lodge, D. (1983). The dissociative anaesthetics, ketamine and phencyclidine, selectively reduce excitation of central mammalian neurones by *N*-methyl-aspartate. *Br. J. Pharmacol.*, *79*, 565–575.
275. Wong, E. H. F., Kemp, J. A., Priestly, T., Knight, A. R., Woodruff, G. N., Iversen, L. L. (1986). The anticonvulsant MK-801 is a potent *N*-methyl-d-aspartate antagonist. *Proc. Natl. Acad. Sci. USA*, *83*, 7104–7108.
276. Buchan, A. M. (1990). Do NMDA antagonists protect against cerebral ischemia: Are clinical trials warranted? *Cerebrovasc. Brain Metab. Rev.*, *2*, 1–26.
277. Herrling, P. L. (1997). *Excitatory Amino Acids: Clinical Results with Antagonists*. Academic Press, London.
278. Iversen, L. L., Kemp, J. A. (1994). Non-competitive NMDA antagonists as drugs. In G. L. Collingridge and J. C. Watkins (Eds.), *The NMDA Receptor.* Oxford University Press, Oxford.
279. Parsons, C. G., Danysz, W., Quack, G. (1999). Memantine is a clinically well tolerated *N*-methyl-d-aspartate (NMDA) receptor antagonist—a review of preclinical data. *Neuropharmacology*, *38*, 735–767.
280. Nelson, K. A., Park, K. M., Robinovitz, E., Tsigos, C., Max, M. B. (1997). High-dose oral dextromethorphan versus placebo in painful diabetic neuropathy and postherpetic neuralgia. *Neurology*, *48*, 1212–1218.
281. Greenamyre, J. T., Eller, R. V., Zhang, Z., Ovadia, A., Kurlan, R., Gash, D. M. (1994). Antiparkinsonian effects of remacemide hydrochloride, a glutamate antagonist, in rodent and primate models of Parkinson's disease. *Ann. Neurol.*, *35*, 655–661.
282. Rabey, J. M., Nissipeanu, P., Korczyn, A. D. (1992). Efficacy of memantine, an NMDA receptor antagonist, in the treatment of Parkinson's disease. *J. Neural. Transm. Park. Dis. Dement. Sect.*, *4*, 277–282.
283. Lipton, S. A. (1994). HIV-related neuronal injury. Potential therapeutic intervention with calcium channel antagonists, NMDA antagonists. *Mol. Neurobiol.*, *8*, 181–196.
284. Kieburtz, K., Feigin, A., McDermott, M., Como, P., Abwender, D., Zimmerman, C., Hickey, C., Orme, C., Claude, K., Sotack, J., Greenamyre, J. T., Dunn, C.,

Shoulson, I. (1996). A controlled trial of remacemide hydrochloride in Huntington's disease. *Mov. Disord.*, *11*, 273–277.

285. Shuto, S., Ono, S., Imoto, H., Yoshii, K., Matsuda, A. (1998). Synthesis and biological activity of conformationally restricted analogues of milnacipran: (1*S*, 2*R*)-1-phenyl-2-[(*R*)-1-amino-2-propynyl]-*N*,*N*-diethylcyclopropanecarboxamide is a novel class of NMDA receptor channel blocker. *J. Med. Chem.*, *41*, 3507–3514.
286. Villetti, G., Bregola, G., Bassani, F., Bergamaschi, M., Rondelli, I., Pietra, C., Simonato, M. (2001). Preclinical evaluation of CHF3381 as a novel antiepileptic agent. *Neuropharmacology*, *40*, 866–878.
287. Villetti, G., Bergamaschi, M., Bassani, F., Bolzoni, P. T., Maiorino, M., Pietra, C., Rondelli, I., Chamiot-Clerc, P., Simonato, M., Barbieri, M. (2003). Antinociceptive activity of the *N*-methyl-d-aspartate receptor antagonist *N*-(2-indanyl)-glycinamide hydrochloride (CHF3381) in experimental models of inflammatory and neuropathic pain. *J. Pharmacol. Exp. Ther.*, *306*, 804–814.
288. Tarral, A., Dostert, P., Guillevic, Y., Fabbri, L., Rondelli, I., Mariotti, F., Imbimbo, B. P. (2003). Safety, pharmacokinetics, and pharmacodynamics of CHF 3381, a novel *N*-methyl-d-aspartate antagonist, after single oral doses in healthy subjects. *J. Clin. Pharmacol.*, *43*, 901–911.
289. Gotti, B., Duverger, D., Bertin, J., Carter, C., Dupont, R., Frost, J., Gaudilliere, B., MacKenzie, E. T., Rousseau, J., Scatton, B., Wick, A. (1988). Ifenprodil and SL 82.0715 as cerebral anti-ischemic agents. I. Evidence for efficacy in models of focal cerebral ischemia. *J. Pharmacol. Exp. Ther.*, *247*, 1211–1221.
290. Araki, T., Kogure, K., Nishioka, K. (1990). Comparative neuroprotective effects of pentobarbital, vinpocetine, flunarizine and ifenprodil on ischemic neuronal damage in the gerbil hippocampus. *Res. Exp. Med. (Berl.)*, *190*, 19–23.
291. Chenard, B. D., Menniti, F. S. (1999). Antagonists selective for NMDA receptors containing the NR2B subunit. *Curr. Pharm. Des.*, *5*, 381–404.
292. Tamiz, A. P., Whittemore, E. R., Zhou, Z. L., Huang, J. C., Drewe, J. A., Chen, J. C., Cai, S. X., Weber, E., Woodward, R. M., Keana, J. F. (1998). Structure-activity relationships for a series of bis(phenylalkyl)amines: Potent subtype-selective inhibitors of *N*-methyl-d-aspartate receptors. *J. Med. Chem.*, *41*, 3499–3506.
293. Weaver, C. E., Marek, P., Park-Chung, M., Tam, S. W., Farb, D. H. (1997). Neuroprotective activity of a new class of steroidal inhibitors of the *N*-methyl-D-aspartate receptor. *Proc. Natl. Acad. Sci. USA*, *94*, 10450–10454.
294. Feigenbaum, J. J., Bergmann, F., Richmond, S. A., Mechoulam, R., Nadler, V., Kloog, Y., Sokolovsky, M. (1989). Nonpsychotropic cannabinoid acts as a functional *N*-methyl-D-aspartate receptor blocker. *Proc. Natl. Acad. Sci. USA*, *86*, 9584–9587.
295. Stephenson, F. A., Turner, A. J. (1998). *Amino Acid Neurotransmission*. Portland Press, London.
296. McBain, C. J., Mayer, M. L. (1994). *N*-methyl-D-aspartic acid receptor structure and function. *Physiol. Rev.*, *74*, 723–760.
297. Wolosker, H., Blackshaw, S., Snyder, S. H. (1999). Serine racemase: A glial enzyme synthesizing D-serine to regulate glutamate-*N*-methyl-D-aspartate neurotransmission. *Proc. Natl. Acad. Sci. USA*, *96*, 13409–13414.

298. Mothet, J. P., Parent, A. T., Wolosker, H., Brady, R. O., Jr., Linden, D. J., Ferris, C. D., Rogawski, M. A., Snyder, S. H. (2000). D-Serine is an endogenous ligand for the glycine site of the *N*-methyl-D-aspartate receptor. *Proc. Natl. Acad. Sci. USA*, *97*, 4926–4931.

299. McBain, C. J., Kleckner, N. W., Wyrick, S., Dingledine, R. (1989). Structural requirements for activation of the glycine coagonist site of *N*-methyl-D-aspartate receptors expressed in *Xenopus* oocytes. *Mol. Pharmacol.*, *36*, 556–565.

300. Kemp, J. A., Leeson, P. D. (1993). The glycine site of the NMDA receptor—five years on. *Trends Pharmacol. Sci.*, *14*, 20–25.

301. Leeson, P. D., Iversen, L. L. (1994). The glycine site on the NMDA receptor: Structure-activity relationships and therapeutic potential. *J. Med. Chem.*, *37*, 4053–4067.

302. Schwartz, B. L., Hashtroudi, S., Herting, R. L., Schwartz, P., Deutsch, S. I. (1996). D-Cycloserine enhances implicit memory in Alzheimer patients. *Neurology*, *46*, 420–424.

303. Leeson, P. D., Williams, B. J., Baker, R., Ladduwahetty, T., Moore, K. W., Rowley, M. (1990). Effects of five-membered ring conformation on bioreceptor recognition: Identification of 3*R*-amino-1-hydroxy-4*R*-methylpyrrolidin-2-one (L-687,414) as a potent glycine/*N*-methyl-D-aspartate receptor antagonist. *J. Chem. Soc. Chem. Commun.*, 1578–1580.

304. Priestley, T., Marshall, G. R., Hill, R. G., Kemp, J. A. (1998). L-687,414, a low efficacy NMDA receptor glycine site partial agonist in vitro, does not prevent hippocampal LTP in vivo at plasma levels known to be neuroprotective. *Br. J. Pharmacol.*, *124*, 1767–1773.

305. Leeson, P. D., Carling, R. W., Moore, K. W., Moseley, A. M., Smith, J. D., Stevenson, G., Chan, T., Baker, R., Foster, A. C., Grimwood, S., Kemp, J. A., Marshall, G. R., Hoogsteen, K. (1992). 4-Amino-2-carboxytetrahydroquinolines. Structure-activity relationships for antagonism at the glycine site of the NMDA receptor. *J. Med. Chem.*, *35*, 1954–1968.

306. Carignani, C., Ugolini, A., Pinnola, V., Belardetti, F., Trist, D. G., Corsi, M. (1998). NMDA receptor subunit characterization of the glycine site antagonist GV196771A and its action on the spinal cord wind-up. *Naunyn Schmiedebergs Arch. Pharmacol.*, *358*, P1119.

307. Wenk, G. L., Baker, L. M., Stoehr, J. D., Hauss-Wegrzyniak, B., Danysz, W. (1998). Neuroprotection by novel antagonists at the NMDA receptor channel and glycineB sites. *Eur. J. Pharmacol.*, *347*, 183–187.

308. Williams, M., Kowaluk, E. A., Arneric, S. P. (1999). Emerging molecular approaches to pain therapy. *J. Med. Chem.*, *42*, 1481–1500.

309. Cai, S. X., Kher, S. M., Zhou, Z. L., Ilyin, V., Espitia, S. A., Tran, M., Hawkinson, J. E., Woodward, R. M., Weber, E., Keana, J. F. (1997). Structure-activity relationships of alkyl- and alkoxy-substituted 1,4-dihydroquinoxaline-2,3-diones: Potent and systemically active antagonists for the glycine site of the NMDA receptor. *J. Med. Chem.*, *40*, 730–738.

310. Nikam, S. S., Cordon, J. J., Ortwine, D. F., Heimbach, T. H., Blackburn, A. C., Vartanian, M. G., Nelson, C. B., Schwarz, R. D., Boxer, P. A., Rafferty, M. F. (1999). Design and synthesis of novel quinoxaline-2,3-dione AMPA/GlyN receptor antagonists: Amino acid derivatives. *J. Med. Chem.*, *42*, 2266–2271.

311. Potschka, H., Loscher, W., Wlaz, P., Behl, B., Hofmann, H. P., Treiber, H. J., Szabo, L. (1998). LU 73068, a new non-NMDA and glycine/NMDA receptor antagonist: Pharmacological characterization and comparison with NBQX and L-701,324 in the kindling model of epilepsy. *Br. J. Pharmacol.*, *125*, 1258–1266.
312. Mott, D. D., Doherty, J. J., Zhang, S., Washburn, M. S., Fendley, M. J., Lyuboslavsky, P., Traynelis, S. F., Dingledine, R. (1998). Phenylethanolamines inhibit NMDA receptors by enhancing proton inhibition. *Nat. Neurosci.*, *1*, 659–667.
313. Williams, K. (1993). Ifenprodil discriminates subtypes of the *N*-methyl-D-aspartate receptor: Selectivity and mechanisms at recombinant heteromeric receptors. *Mol. Pharmacol.*, *44*, 851–859.
314. Kristensen, J. D., Svensson, B., Gordh, T., Jr. (1992). The NMDA-receptor antagonist CPP abolishes neurogenic "wind-up pain" after intrathecal administration in humans. *Pain*, *51*, 249–253.
315. Eide, P. K., Jorum, E., Stubhaug, A., Bremnes, J., Breivik, H. (1994). Relief of postherpetic neuralgia with the *N*-methyl-D-aspartic acid receptor antagonist ketamine: A double-blind, cross-over comparison with morphine and placebo. *Pain*, *58*, 347–354.
316. Marino, M., Valenti, O., Conn, P. J. (2003). Glutamate receptors and Parkinson's disease: Opportunities for intervention. *Drugs Aging*, *20*, 377–397.
317. Buller, A. L., Larson, H. C., Schneider, B. E., Beaton, J. A., Morrisett, R. A., Monaghan, D. T. (1994). The molecular basis of NMDA receptor subtypes: Native receptor diversity is predicted by subunit composition. *J. Neurosci.*, *14*, 5471–5484.
318. Monaghan, D. T., Larsen, H. (1997). NR1 and NR2 subunit contributions to *N*-methyl-D-aspartate receptor channel blocker pharmacology. *J. Pharmacol. Exp. Ther.*, *280*, 614–620.
319. Bang-Andersen, B., Lenz, S. M., Skjærbæk, N., Søby, K. K., Hansen, H. O., Ebert, B., Bøgesø, K. P., Krogsgaard-Larsen, P. (1997). Heteroaryl analogues of AMPA. Synthesis and quantitative structure-activity relationships. *J. Med. Chem.*, *40*, 2831–2842.
320. Falch, E., Brehm, L., Mikkelsen, I., Johansen, T. N., Skjærbæk, N., Nielsen, B., Stensbøl, T. B., Ebert, B., Krogsgaard-Larsen, P. (1998). Heteroaryl analogues of AMPA. 2. Synthesis, absolute stereochemistry, photochemistry, and structure-activity relationships. *J. Med. Chem.*, *41*, 2513–2523.
321. Sløk, F. A., Ebert, B., Lang, Y., Krogsgaard-Larsen, P., Lenz, S. M., Madsen, U. (1997). Excitatory amino-acid receptor agonists. Synthesis and pharmacology of analogues of 2-amino-3-(3-hydroxy-5-methylisoxazol-4-yl)propionic acid. *Eur. J. Med. Chem.*, *32*, 329–338.
322. Madsen, U., Frølund, B., Lund, T. M., Ebert, B., Krogsgaard-Larsen, P. (1993). Design, synthesis and pharmacology of model compounds for indirect elucidation of the topography of AMPA receptor sites. *Eur. J. Med. Chem.*, *28*, 791–800.
323. Madsen, U., Ebert, B., Krogsgaard-Larsen, P., Wong, E. H. F. (1992). Synthesis and pharmacology of (*RS*)-2-amino-3-(3-hydroxy-5- trifluoromethyl-4-isoxazolyl)propionic acid, a potent AMPA receptor agonist. *Eur. J. Med. Chem.*, *27*, 479–484.
324. Stensbøl, T. B., Uhlmann, P., Morel, S., Eriksen, B. L., Felding, J., Kromann, H., Hermit, M. B., Greenwood, J. R., Bräuner-Osborne, H., Madsen, U., Junager, F., Krogsgaard-Larsen, P., Begtrup, M., Vedsø, P. (2002). Novel 1-hydroxyazole

bioisosteres of glutamic acid. Synthesis, protolytic properties, and pharmacology. *J. Med. Chem.*, *45*, 19–31.

325. Lauridsen, J., Honoré, T., Krogsgaard-Larsen, P. (1985). Ibotenic acid analogues. Synthesis, molecular flexibility, and *in vitro* activity of agonists and antagonists at central glutamic acid receptors. *J. Med. Chem.*, *28*, 668–672.
326. Christensen, I. T., Reinhardt, A., Nielsen, B., Ebert, B., Madsen, U., Nielsen, E. Ø., Brehm, L., Krogsgaard-Larsen, P. (1989). Excitatory amino acid agonists and partial agonists. *Drug Des. Del.*, *5*, 57–71.
327. Ebert, B., Lenz, S. M., Brehm, L., Bregnedal, P., Hansen, J. J., Frederiksen, K., Bøgesø, K. P., Krogsgaard-Larsen, P. (1994). Resolution, absolute stereochemistry, and pharmacology of the *S*-(+)- and *R*-(−)-isomers of the apparent partial AMPA receptor agonist (*R,S*)-2-amino-3-(3-hydroxy-5-phenylisoxazol-4-yl)propionic acid [(*R,S*)-APPA]. *J. Med. Chem.*, *37*, 878–884.
328. Vogensen, S. B., Jensen, H. S., Stensbøl, T. B., Frydenvang, K., Bang-Andersen, B., Johansen, T. N., Egebjerg, J., Krogsgaard-Larsen, P. (2000). Resolution, configurational assignment, and enantiopharmacology of 2-amino-3-[3-hydroxy-5-(2-methyl-2*H*-tetrazol-5-yl)isoxazol-4-yl]propionic acid, a potent GluR3- and GluR4-preferring AMPA receptor agonist. *Chirality*, *12*, 705–713.
329. Clarke, V. R. J., Ballyk, B. A., Hoo, K. H., Mandelzys, A., Pellizzari, A., Bath, C. P., Thomas, J., Sharpe, E. F., Davies, C. H., Ornstein, P. L., Schoepp, D. D., Kamboj, R. K., Collingridge, G. L., Lodge, D., Bleakman, D. (1997). A hippocampal GluR5 kainate receptor regulating inhibitory synaptic transmission. *Nature*, *389*, 599–602.
330. Hill, R. A., Wallace, L. J., Miller, D. D., Weinstein, D. M., Shams, G., Tai, H., Layer, R. T., Willins, D., Uretsky, N. J., Danthi, S. N. (1997). Structure-activity studies for alpha-amino-3-hydroxy-5-methyl-4-isoxazolepropanoic acid receptors: Acidic hydroxyphenylalanines. *J. Med. Chem.*, *40*, 3182–3191.
331. Patneau, D. K., Mayer, M. L., Jane, D. E., Watkins, J. C. (1992). Activation and desensitization of AMPA/kainate receptors by novel derivatives of willardiine. *J. Neurosci.*, *12*, 595–606.
332. Jane, D. E., Hoo, K., Kamboj, R., Deverill, M., Bleakman, D., Mandelzys, A. (1997). Synthesis of willardiine and 6-azawillardiine analogs: Pharmacological characterization on cloned homomeric human AMPA and kainate receptor subtypes. *J. Med. Chem.*, *40*, 3645–3650.
333. Hawkins, L. M., Beaver, K. M., Jane, D. E., Taylor, P. M., Sunter, D. C., Roberts, P. J. (1995). Binding of the new radioligand (*S*)-[$^3$H]AMPA to rat brain synaptic membranes: Effects of a series of structural analogues of the non-NMDA receptor agonist willardiine. *Neuropharmacology*, *34*, 405–410.
334. Jin, R., Gouaux, E. (2003). Probing the function, conformational plasticity, and dimer-dimer contacts of the GluR2 ligand-binding core: Studies of 5-substituted willardiines and GluR2 S1S2 in the crystal. *Biochemistry*, *42*, 5201–5213.
335. Jin, R., Banke, T. G., Mayer, M. L., Traynelis, S. F., Gouaux, E. (2003). Structural basis for partial agonist action at ionotropic glutamate receptors. *Nat. Neurosci.*, *6*, 803–810.
336. Coquelle, T., Christensen, J. K., Banke, T. G., Madsen, U., Schousboe, A., Pickering, D. S. (2000). Agonist discrimination between AMPA receptor subtypes. *Neuroreport*, *11*, 2643–2648.

337. Banke, T. G., Greenwood, J. R., Christensen, J. K., Liljefors, T., Traynelis, S. F., Schousboe, A., Pickering, D. S. (2001). Identification of amino acid residues in GluR1 responsible for ligand binding and desensitization. *J. Neurosci.*, *21*, 3052–3062.

338. Hogner, A., Kastrup, J. S., Jin, R., Liljefors, T., Mayer, M. L., Egebjerg, J., Larsen, I. K., Gouaux, E. (2002). Structural basis for AMPA receptor activation and ligand selectivity: Crystal structures of five agonist complexes with the GluR2 ligand-binding core. *J. Mol. Biol.*, *322*, 93–109.

339. Bjerrum, E. J., Kristensen, A. S., Pickering, D. S., Greenwood, J. R., Nielsen, B., Liljefors, T., Schousboe, A., Bräuner-Osborne, H., Madsen, U. (2003). Design, synthesis, and pharmacology of a highly subtype-selective GluR1/2 agonist, (*RS*)-2-amino-3-(4-chloro-3-hydroxy-5-isoxazolyl)propionic acid (Cl–HIBO). *J. Med. Chem.*, *46*, 2246–2249.

340. Madsen, U., Wong, E. H. F. (1992). Heterocyclic excitatory amino acids. Synthesis and biological activity of novel analogues of AMPA. *J. Med. Chem.*, *35*, 107–111.

341. Wahl, P., Madsen, U., Banke, T., Krogsgaard-Larsen, P., Schousboe, A. (1996). Different characteristics of AMPA receptor agonists acting at AMPA receptors expressed in *Xenopus* oocytes. *Eur. J. Pharmacol.*, *308*, 211–218.

342. Johansen, T. N., Stensbøl, T. B., Nielsen, B., Vogensen, S. B., Frydenvang, K., Sløk, F. A., Bräuner-Osborne, H., Madsen, U., Krogsgaard-Larsen, P. (2001). Resolution, configurational assignment and enantiopharmacology at glutamate receptors of 2-amino-3-(3-carboxy-5-methyl-4-isoxazolyl)propionic acid (ACPA) and demethyl-ACPA. *Chirality*, *13*, 523–532.

343. Watkins, J. C., Evans, R. H. (1981). Excitatory amino acid transmitters. *Annu. Rev. Pharmacol. Toxicol.*, *21*, 165–204.

344. Monaghan, D. T., Bridges, R. J., Cotman, C. W. (1989). The excitatory amino acid receptors. *Annu. Rev. Pharmacol. Toxicol.*, *29*, 365–402.

345. Krogsgaard-Larsen, P., Honore, T., Hansen, J. J., Curtis, D. R., Lodge, D. (1980). New class of glutamate agonist structurally related to ibotenic acid. *Nature*, *284*, 64–66.

346. Watkins, J. C., Krogsgaard-Larsen, P., Honoré, T. (1990). Structure-activity relationships in the development of excitatory amino acid receptor agonists and competitive antagonists. *Trends Pharmacol. Sci.*, *11*, 25–33.

347. Bleakman, D., Lodge, D. (1998). Neuropharmacology of AMPA and kainate receptors. *Neuropharmacology*, *37*, 1187–1204.

348. Vogensen, S. B., Greenwood, J. R., Varming, A. R., Brehm, L., Pickering, D. S., Nielsen, B., Liljefors, T., Clausen, R. P., Johansen, T. N., Krogsgaard-Larsen, P. (2004). A stereochemical anomaly: The cyclised (*R*)-AMPA analogue (*R*)-3-hydroxy-4,5,6,7-tetrahydroisoxazolo[5,4-*c*]pyridine-5-carboxylic acid [(*R*)-5-HPCA] resembles (*S*)-AMPA at glutamate receptors. *Org. Biomol. Chem.*, *2*, 206–213.

349. Conti, P., De Amici, M., De Sarro, G., Stensbøl, T. B., Bräuner-Osborne, H., Madsen, U., De Micheli, C. (1998). Synthesis and pharmacology of a new AMPA-kainate receptor agonist with potent convulsant activity. *J. Med. Chem.*, *41*, 3759–3762.

350. Conti, P., De Amici, M., De Sarro, G., Rizzo, M., Stensbøl, T. B., Bräuner-Osborne, H., Madsen, U., Toma, L., De Micheli, C. (1999). Synthesis and enantiopharmacology of new AMPA-kainate receptor agonists. *J. Med. Chem.*, *42*, 4099–4107.

351. Honoré, T., Davies, S. N., Drejer, J., Fletcher, E. J., Jacobsen, P., Lodge, D., Nielsen, F. E. (1988). Quinoxalinediones: Potent competitive non-NMDA glutamate receptor antagonists. *Science*, *241*, 701–703.

352. Sheardown, M. J., Nielsen, E. Ø., Hansen, A. J., Jacobsen, P., Honoré, T. (1990). 2,3-Dihydroxy-6-nitro-7-sulfamoyl-benzo(*F*)quinoxaline. A neuroprotectant for cerebral ischemia. *Science*, *247*, 571–574.

353. Armstrong, N., Gouaux, E. (2000). Mechanisms for activation and antagonism of an AMPA-sensitive glutamate receptor: Crystal structures of the GluR2 ligand binding core. *Neuron*, *28*, 165–181.

354. Turski, L., Huth, A., Sheardown, M., McDonald, F., Neuhaus, R., Schneider, H. H., Dirnagl, U., Wiegand, F., Jacobsen, P., Ottow, E. (1998). ZK200775: A phosphonate quinoxalinedione AMPA antagonist for neuroprotection in stroke and trauma. *Proc. Natl. Acad. Sci. USA*, *95*, 10960–10965.

355. Lubisch, W., Behl, B., Hofmann, H. P. (1997). Pyrrolylquinoxalinediones: The importance of pyrrolic substitution on AMPA receptor binding. *Bioorg. Med. Chem. Lett.*, *7*, 1101–1106.

356. Nielsen, E. Ø., Varming, T., Mathiesen, C., Jensen, L. H., Møller, A., Gouliaev, A. H., Watjen, F., Drejer, J. (1999). SPD 502: A water-soluble and in vivo long-lasting AMPA antagonist with neuroprotective activity. *J. Pharmacol. Exp. Ther.*, *289*, 1492–1501.

357. Mignani, S., Bohme, G. A., Birraux, G., Boireau, A., Jimonet, P., Damour, D., Genevois-Borella, A., Debono, M. W., Pratt, J., Vuilhorgne, M., Wahl, F., Stutzmann, J. M. (2002). 9-Carboxymethyl-5*H*,10*H*-imidazo[1,2-*a*]indeno[1,2-*e*]pyrazin-4-one-2-carbocylic acid (RPR117824): Selective anticonvulsive and neuroprotective AMPA antagonist. *Bioorg. Med. Chem.*, *10*, 1627–1637.

358. Stutzmann, J. M., Bohme, G. A., Boireau, A., Damour, D., Debono, M. W., Genevois-Borella, A., Jimonet, P., Pratt, J., Randle, J. C., Ribeill, Y., Vuilhorgne, M., Mignani, S. (2001). Synthesis of anticonvulsive AMPA antagonists: 4-Oxo-10-substituted-imidazo[1,2-*e*]pyrazin-2-carboxylic acid derivatives. *Bioorg. Med. Chem. Lett.*, *11*, 1205–1210.

359. Jimonet, P., Bohme, G. A., Bouquerel, J., Boireau, A., Damour, D., Debono, M. W., Genevois-Borella, A., Hardy, J. C., Hubert, P., Manfre, F., Nemecek, P., Pratt, J., Randle, J. C., Ribeill, Y., Stutzmann, J. M., Vuilhorgne, M., Mignani, S. (2001). Bioisosteres of 9-carboxymethyl-4-oxo-imidazo[1,2-*a*]indeno-[1,2-*e*]pyrazin-2-carboxylic acid derivatives. Progress towards selective, potent in vivo AMPA antagonists with longer durations of action. *Bioorg. Med. Chem. Lett.*, *11*, 127–132.

360. Stutzmann, J. M., Bohme, G. A., Boireau, A., Damour, D., Debono, M. W., Genevois-Borella, A., Imperato, A., Jimonet, P., Pratt, J., Randle, J. C., Ribeill, Y., Vuilhorgne, M., Mignani, S. (2000). 4,10-Dihydro-4-oxo-4*H*-imidazo[1,2-*a*]indeno[1,2-*e*]pyrazin-2-carboxylic acid derivatives: Highly potent and selective AMPA receptors antagonists with in vivo activity. *Bioorg. Med. Chem. Lett.*, *10*, 1133–1137.

361. Canton, T., Bohme, G. A., Boireau, A., Bordier, F., Mignani, S., Jimonet, P., Jahn, G., Alavijeh, M., Stygall, J., Roberts, S., Brealey, C., Vuilhorgne, M., Debono, M. W., Le Guern, S., Laville, M., Briet, D., Roux, M., Stutzmann, J. M., Pratt, J. (2001). RPR 119990, a novel alpha-amino-3-hydroxy-5-methyl-4-

isoxazolepropionic acid antagonist: Synthesis, pharmacological properties, and activity in an animal model of amyotrophic lateral sclerosis. *J. Pharmacol. Exp. Ther.*, *299*, 314–322.

362. Ornstein, P. L., Arnold, M. B., Allen, N. K., Bleisch, T., Borromeo, P. S., Lugar, C. W., Leander, J. D., Lodge, D., Schoepp, D. D. (1996). Structure-activity studies of 6-(tetrazolylalkyl)-substituted decahydroisoquinoline-3-carboxylic acid AMPA receptor antagonists. 1. Effects of stereochemistry, chain length, and chain substitution. *J. Med. Chem.*, *39*, 2219–2231.
363. Madsen, U., Bang-Andersen, B., Brehm, L., Christensen, I. T., Ebert, B., Kristoffersen, I. T., Lang, Y., Krogsgaard-Larsen, P. (1996). Synthesis and pharmacology of highly selective carboxy and phosphono isoxazole amino acid AMPA receptor antagonists. *J. Med. Chem.*, *39*, 1682–1691.
364. Ornstein, P. L., Arnold, M. B., Augenstein, N. K., Lodge, D., Leander, J. D., Schoepp, D. D. (1993). (3*SR*,4*aRS*,6*RS*,8*aRS*)-6-[2-(1*H*-tetrazol-5-yl)ethyl] decahydroisoquinoline-3-carboxylic acid: A structurally novel, systemically active, competitive AMPA receptor antagonist. *J. Med. Chem.*, *36*, 2046–2048.
365. Bleakman, R., Schoepp, D. D., Ballyk, B., Bufton, H., Sharpe, E. F., Thomas, K., Ornstein, P. L., Kamboj, R. K. (1996). Pharmacological discrimination of GluR5 and GluR6 kainate receptor subtypes by (3*S*,4*aR*,6*R*,8*aR*)-6-[2-(1(2)*H*-tetrazole-5-yl)ethyl]decahydroisdoquinoline-3-carboxylic acid. *Mol. Pharmacol.*, *49*, 581–585.
366. Møller, E. H., Egebjerg, J., Brehm, L., Stensbøl, T. B., Johansen, T. N., Madsen, U., Krogsgaard-Larsen, P. (1999). Resolution, absolute stereochemistry, and enantiopharmacology of the GluR1–4 and GluR5 antagonist 2-amino-3-[5-*tert*-butyl-3-(phosphonomethoxy)-4-isoxazolyl]propionic acid. *Chirality*, *11*, 752–759.
367. Sólyom, S., Tarnawa, I. (2002). Non-competitive AMPA antagonists of 2,3-benzodiazepine type. *Curr. Pharm. Des.*, *8*, 913–939.
368. Donevan, S. D., Rogawski, M. A. (1993). GYKI 52466, a 2,3-benzodiazepine, is a highly selective, noncompetitive antagonist of AMPA/kainate receptor responses. *Neuron*, *10*, 51–59.
369. Zorumski, C. F., Yamada, K. A., Price, M. T., Olney, J. W. (1993). A benzodiazepine recognition site is associated with the non-NMDA glutamate receptor. *Neuron*, *10*, 61–67.
370. Anderson, B. A., Harn, N. K., Hansen, M. M., Harkness, A. R., Lodge, D., Leander, J. D. (1999). Synthesis and anticonvulsant activity of 3-aryl-5*H*-2,3-benzodiazepine AMPA antagonists. *Bioorg. Med. Chem. Lett.*, *9*, 1953–1956.
371. Welch, W. M., Ewing, F. E., Huang, J., Menniti, F. S., Pagnozzi, M. J., Kelly, K., Seymour, P. A., Guanowsky, V., Guhan, S., Guinn, M. R., Critchett, D., Lazzaro, J., Ganong, A. H., DeVries, K. M., Staigers, T. L., Chenard, B. L. (2001). Atropisomeric quinazolin-4-one derivatives are potent noncompetitive alpha-amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid (AMPA) receptor antagonists. *Bioorg. Med. Chem. Lett.*, *11*, 177–181.
372. Ohno, K., Tsutsumi, R., Matsumoto, N., Yamashita, H., Amada, Y., Shishikura, J., Yatsugi, H. I., Okada, M., Sakamoto, S., Yamaguchi, T. (2003). Functional characterization of YM928, a novel moncompetitive alpha-amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid (AMPA) receptor antagonist. *J. Pharmacol. Exp. Ther.*, *306*, 66–72.

373. Menniti, F. S., Chenard, B. L., Collins, M. B., Ducat, M. F., Elliott, M. L., Ewing, F. E., Huang, J. I., Kelly, K. A., Lazzaro, J. T., Pagnozzi, M. J., Weeks, J. L., Welch, W. M., White, W. F. (2000). Characterization of the binding site for a novel class of noncompetitive alpha-amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid receptor antagonists. *Mol. Pharmacol.*, *58*, 1310–1317.
374. Granger, R., Deadwyler, S., Davis, M., Moskovitz, B., Kessler, M., Rogers, G., Lynch, G. (1996). Facilitation of glutamate receptors reverses an age-associated memory impairment in rats. *Synapse*, *22*, 332–337.
375. Sun, Y., Olson, R., Horning, M., Armstrong, N., Mayer, M., Gouaux, E. (2002). Mechanism of glutamate receptor desensitization. *Nature*, *417*, 245–253.
376. Zarrinmayeh, H., Bleakman, D., Gates, M. R., Yu, H., Zimmerman, D. M., Ornstein, P. L., McKennon, T., Arnold, M. B., Wheeler, W. J., Skolnick, P. (2001). [$^3$H]*N*-2-(4-(*N*-benzamido)phenyl)propyl-2-propanesulfonamide: A novel AMPA receptor potentiator and radioligand. *J. Med. Chem.*, *44*, 302–304.
377. Linden, A. M., Yu, H., Zarrinmayeh, H., Wheeler, W. J., Skolnick, P. (2001). Binding of an AMPA receptor potentiator ([$^3$H]LY395153) to native and recombinant AMPA receptors. *Neuropharmacology*, *40*, 1010–1018.
378. Shepherd, T. A., Aikins, J. A., Bleakman, D., Cantrell, B. E., Rearick, J. P., Simon, R. L., Smith, E. C., Stephenson, G. A., Zimmerman, D. M., Mandelzys, A., Jarvie, K. R., Ho, K., Deverill, M., Kamboj, R. K. (2002). Design and synthesis of a novel series of 1,2-disubstituted cyclopentanes as small, potent potentiators of 2-amino-3-(3-hydroxy-5- methyl-isoxazol-4-yl)propanoic acid (AMPA) receptors. *J. Med. Chem.*, *45*, 2101–2111.
379. Stensbøl, T. B., Borre, L., Johansen, T. N., Egebjerg, J., Madsen, U., Ebert, B., Krogsgaard-Larsen, P. (1999). Resolution, absolute stereochemistry and molecular pharmacology of the enantiomers of ATPA. *Eur. J. Pharmacol.*, *380*, 153–162.
380. Stensbøl, T. B., Jensen, H. S., Nielsen, B., Johansen, T. N., Egebjerg, J., Frydenvang, K., Krogsgaard-Larsen, P. (2001). Stereochemistry and molecular pharmacology of (*S*)-thio-ATPA, a new potent and selective GluR5 agonist. *Eur. J. Pharmacol.*, *411*, 245–253.
381. Thomas, N. K., Hawkins, L. M., Miller, J. C., Troop, H. M., Roberts, P. J., Jane, D. E. (1998). Pharmacological differentiation of kainate receptors on neonatal rat spinal motoneurones and dorsal roots. *Neuropharmacology*, *37*, 1223–1237.
382. Lunn, M. L., Hogner, A., Stensbøl, T. B., Gouaux, E., Egebjerg, J., Kastrup, J. S. (2003). Three-dimensional structure of the ligand-binding core of GluR2 in complex with the agonist (*S*)-ATPA: Implications for receptor subunit selectivity. *J. Med. Chem.*, *46*, 872–875.
383. Small, B., Thomas, J., Kemp, M., Hoo, K., Ballyk, B., Deverill, M., Ogden, A. M., Rubio, A., Pedregal, C., Bleakman, D. (1998). LY339434, a GluR5 kainate receptor agonist. *Neuropharmacology*, *37*, 1261–1267.
384. Pedregal, C., Collado, I., Escribano, A., Ezquerra, J., Dominguez, C., Mateo, A. I., Rubio, A., Baker, S. R., Goldsworthy, J., Kamboj, R. K., Ballyk, B. A., Hoo, K., Bleakman, D. (2000). 4-Alkyl- and 4-cinnamylglutamic acid analogues are potent GluR5 kainate receptor agonists. *J. Med. Chem.*, *43*, 1958–1968.
385. Bunch, L., Johansen, T. H., Bräuner-Osborne, H., Stensbøl, T. B., Johansen, T. N., Krogsgaard-Larsen, P., Madsen, U. (2001). Synthesis and receptor binding affinity of new selective GluR5 ligands. *Bioorg. Med. Chem.*, *9*, 875–879.

386. Valgeirsson, J., Christensen, J. K., Kristensen, A. S., Pickering, D. S., Nielsen, B., Fischer, C. H., Bräuner-Osborne, H., Nielsen, E. ø., Krogsgaard-Larsen, P., Madsen, U. (2003). Synthesis and in vitro pharmacology at AMPA and kainate preferring glutamate receptors of 4-heteroarylmethylidene glutamate analogues. *Bioorg. Med. Chem.*, *11*, 4341–4349.
387. Krogsgaard-Larsen, P., Nielsen, E. Ø., Curtis, D. R. (1984). Ibotenic acid analogues. Synthesis and biological *in vitro* activity of conformationally restricted agonists at excitatory amino acid receptors. *J. Med. Chem.*, *27*, 585–591.
388. Brehm, L., Greenwood, J. R., Hansen, K. B., Nielsen, B., Egebjerg, J., Stensbøl, T. B., Bräuner-Osborne, H., Sløk, F. A., Kronborg, T. T., Krogsgaard-Larsen, P. (2003). (*S*)-2-Amino-3-(3-hydroxy-7,8-dihydro-6*H*-cyclohepta[*d*]isoxazol-4-yl)propionic acid, a potent and selective agonist at the GluR5 subtype of ionotropic glutamate receptors. Synthesis, modeling, and molecular pharmacology. *J. Med. Chem.*, *46*, 1350–1358.
389. Johansen, T. H., Drejer, J., Wätjen, F., Nielsen, E. ø. (1993). A novel non-NMDA receptor antagonist shows selective displacement of low-affinity [$^3$H]kainate binding. *Eur. J. Pharmacol. Mol. Pharmacol. Sect. 246*, 195–204.
390. Verdoorn, T. A., Johansen, T. H., Drejer, J., Nielsen, E. ø. (1994). Selective block of recombinant GluR6 receptors by NS-102, a novel non-NMDA receptor antagonist. *Eur. J. Pharmacol.*, *269*, 43–49.
391. O'Neill, M. J., Bond, A., Ornstein, P. L., Ward, M. A., Hicks, C. A., Hoo, K., Bleakman, D., Lodge, D. (1998). Decahydroisoquinolines: Novel competitive AMPA/kainate antagonists with neuroprotective effects in global cerebral ischaemia. *Neuropharmacology*, *37*, 1211–1222.
392. Lubisch, W., Behl, B., Henn, C., Hofmann, H. P., Reeb, J., Regner, F., Vierling, M. (2002). Pyrrolylquinoxalinediones carrying a piperazine residue represent highly potent and selective ligands to the homomeric kainate receptor GluR5. *Bioorg. Med. Chem. Lett.*, *12*, 2113–2116.
393. Valgeirsson, J., Nielsen, E. Ø., Peters, D., Varming, T., Mathiesen, C., Kristensen, A. S., Madsen, U. (2003). 2-Arylureidobenzoic acids: Selective noncompetitive antagonists for the homomeric kainate receptor subtype GluR5. *J. Med. Chem.*, *46*, 5834–5843.
394. Madden, D. R. (2002). The structure and function of glutamate receptor ion channels. *Nat. Rev. Neurosci.*, *3*, 91–101.
395. Rosenmund, C., Stern-Bach, Y., Stevens, C. F. (1998). The tetrameric structure of a glutamate receptor channel. *Science*, *280*, 1596–1599.
396. Ayalon, G., Stern-Bach, Y. (2001). Functional assembly of AMPA and kainate receptors is mediated by several discrete protein–protein interactions. *Neuron*, *31*, 103–113.
397. Safferling, M., Tichelaar, W., Kummerle, G., Jouppila, A., Kuusinen, A., Keinanen, K., Madden, D. R. (2001). First images of a glutamate receptor ion channel: Oligomeric state and molecular dimensions of GluRB homomers. *Biochemistry*, *40*, 13948–13953.
398. Kuusinen, A., Arvola, M., Keinanen, K. (1995). Molecular dissection of the agonist binding site of an AMPA receptor. *EMBO J.*, *14*, 6327–6332.
399. Armstrong, N., Sun, Y., Chen, G. Q., Gouaux, E. (1998). Structure of a glutamate-receptor ligand-binding core in complex with kainate. *Nature*, *395*, 913–917.

400. Kasper, C., Lunn, M. L., Liljefors, T., Gouaux, E., Egebjerg, J., Kastrup, J. S. (2002). GluR2 ligand-binding core complexes: Importance of the isoxazolol moiety and 5-substituent for the binding mode of AMPA-type agonists. *FEBS Lett.*, *531*, 173–178.
401. Lunn, M. L. (2001). From gene to crystal structure—a study of the ligand binding domain of AMPA and kainate receptors. Ph.D. Thesis, Royal Danish School of Pharmacy.
402. Jin, R., Horning, M., Mayer, M. L., Gouaux, E. (2002). Mechanism of activation and selectivity in a ligand-gated ion channel: Structural and functional studies of GluR2 and quisqualate. *Biochemistry*, *41*, 15635–15643.
403. Hogner, A., Greenwood, J. R., Liljefors, T., Lunn, M. L., Egebjerg, J., Larsen, I. K., Gouaux, E., Kastrup, J. S. (2003). Competitive antagonism of AMPA receptors by ligands of different classes: Crystal structure of ATPO bound to the GluR2 ligand-binding core, in comparison with DNQX. *J. Med. Chem.*, *46*, 214–221.
404. Armstrong, N., Mayer, M., Gouaux, E. (2003). Tuning activation of the AMPA-sensitive GluR2 ion channel by genetic adjustment of agonist-induced conformational changes. *Proc. Natl. Acad. Sci. USA*, *100*, 5736–5741.
405. Chen, G. Q., Cui, C., Mayer, M. L., Gouaux, E. (1999). Functional characterization of a potassium-selective prokaryotic glutamate receptor. *Nature*, *402*, 817–821.
406. Campiani, G., Morelli, E., Nacci, V., Fattorusso, C., Ramunno, A., Novellino, E., Greenwood, J., Liljefors, T., Griffiths, R., Sinclair, C., Reavy, H., Kristensen, A. S., Pickering, D. S., Schousboe, A., Cagnotto, A., Fumagalli, E., Mennini, T. (2001). Characterization of the 1*H*-cyclopentapyrimidine-2,4(1*H*,3*H*)-dione derivative (*S*)-CPW399 as a novel, potent, and subtype-selective AMPA receptor full agonist with partial desensitization properties. *J. Med. Chem.*, *44*, 4501–4504.
407. Johansen, T. N., Greenwood, J. R., Frydenvang, K., Madsen, U., Krogsgaard-Larsen, P. (2003). Stereostructure-activity studies on agonists at the AMPA and kainate subtypes of ionotropic glutamate receptors. *Chirality*, *15*, 167–179.
408. Arinaminpathy, Y., Sansom, M. S., Biggin, P. C. (2002). Molecular dynamics simulations of the ligand-binding domain of the ionotropic glutamate receptor GluR2. *Biophys. J.*, *82*, 676–683.
409. Pang, A., Arinaminpathy, Y., Sansom, M. S., Biggin, P. C. (2003). Interdomain dynamics and ligand binding: Molecular dynamics simulations of glutamine binding protein. *FEBS Lett.*, *550*, 168–174.
410. Pentikäinen, O. T., Settimo, L., Keinanen, K., Johnson, M. S. (2003). Selective agonist binding of (*S*)-amino-3-(3-hydroxy-5-methyl-4-isoxazolyl)propionic acid (AMPA) and (2*S*-(2*a*,3*b*,4*b*))-2-carboxy-4-(1-methylethenyl)3-pyrrolidine-acetic acid (kainate) receptors: A molecular modeling study. *Biochem. Pharmacol.*, *66*, 2413–2425.
411. Partin, K. M., Bowie, D., Mayer, M. L. (1995). Structural determinants of allosteric regulation in alternatively spliced AMPA receptors. *Neuron*, *14*, 833–843.
412. Augelli-Szafran, C. E., Schwarz, R. D. (2003). Metabotropic glutamate receptors: Agonists, antagonists and allosteric modulators. *Annu. Rep. Med. Chem.*, *38*, 21–30.
413. Irving, A. J., Schofield, J. G., Watkins, J. C., Sunter, D. C., Collingridge, G. L. (1990). 1*S*,3*R*-ACPD stimulates and L-AP3 blocks $Ca^{2+}$ mobilization in rat cerebellar neurons. *Eur. J. Pharmacol.*, *186*, 363–365.

414. Aramori, I., Nakanishi, S. (1992). Signal transduction and pharmacological characteristics of a metabotropic glutamate receptor, mGluR1, in transfected CHO cells. *Neuron*, *8*, 757–765.

415. Tanabe, Y., Masu, M., Ishii, T., Shigemoto, R., Nakanishi, S. (1992). A family of metabotropic glutamate receptors. *Neuron*, *8*, 169–179.

416. Tanabe, Y., Nomura, A., Masu, M., Shigemoto, R., Mizuno, N., Nakanishi, S. (1993). Signal transduction, pharmacological properties, and expression patterns of two rat metabotropic glutamate receptors, mGluR3 and mGluR4. *J. Neurosci.*, *13*, 1372–1378.

417. Abe, T., Sugihara, H., Nawa, H., Shigemoto, R., Mizuno, N., Nakanishi, S. (1992). Molecular characterization of a novel metabotropic glutamate receptor mGluR5 coupled to inositol phosphate/$Ca^{2+}$ signal transduction. *J. Biol Chem.*, *267*, 13361–13368.

418. Nakajima, Y., Iwakabe, H., Akazawa, C., Nawa, H., Shigemoto, R., Mizuno, N., Nakanishi, S. (1993). Molecular characterization of a novel retinal metabotropic glutamate receptor mGluR6 with a high agonist selectivity for L-2-amino-4-phosphonobutyrate. *J. Biol Chem.*, *268*, 11868–11873.

419. Okamoto, N., Hori, S., Akazawa, C., Hayashi, Y., Shigemoto, R., Mizuno, N., Nakanishi, S. (1994). Molecular characterization of a new metabotropic glutamate receptor mGluR7 coupled to inhibitory cyclic AMP signal transduction. *J. Biol. Chem.*, *269*, 1231–1236.

420. Saugstad, J. A., Kinzie, J. M., Shinohara, M. M., Segerson, T. P., Westbrook, G. L. (1997). Cloning and expression of rat metabotropic glutamate receptor 8 reveals a distinct pharmacological profile. *Mol. Pharmacol.*, *51*, 119–125.

421. Monn, J. A., Valli, M. J., Massey, S. M., Hansen, M. M., Kress, T. J., Wepsiec, J. P., Harkness, A. R., Grutsch, J. L., Jr., Wright, R. A., Johnson, B. G., Andis, S. L., Kingston, A., Tomlinson, R., Lewis, R., Griffey, K. R., Tizzano, J. P., Schoepp, D. D. (1999). Synthesis, pharmacological characterization, and molecular modeling of heterobicyclic amino acids related to (+)-2-aminobicyclo[3.1.0]hexane-2,6-dicarboxylic acid (LY354740): Identification of two new potent, selective, and systemically active agonists for group II metabotropic glutamate receptors. *J. Med. Chem.*, *42*, 1027–1040.

422. Schoepp, D. D., Salhoff, C. R., Wright, R. A., Johnson, B. G., Burnett, J. P., Mayne, N. G., Belagaje, R., Wu, S., Monn, J. A. (1996). The novel metabotropic glutamate receptor agonist 2*R*,4*R*-APDC potentiates stimulation of phosphoinositide hydrolysis in the rat hippocampus by 3,5-dihydroxyphenylglycine: Evidence for a synergistic interaction between group 1 and group 2 receptors. *Neuropharmacology*, *35*, 1661–1672.

423. Tückmantel, W., Kozikowski, A. P., Wang, S., Pshenichkin, S., Wroblewski, J. T. (1997). Synthesis, molecular modeling, and biology of the 1-benzyl derivative of APDC—an apparent mGluR6 selective ligand. *Bioorg. Med. Chem. Lett.*, *7*, 601–606.

424. Valli, M. J., Schoepp, D. D., Wright, R. A., Johnson, B. G., Kingston, A. E., Tomlinson, R., Monn, J. A. (1998). Synthesis and metabotropic glutamate receptor antagonist activity of N1-substituted analogs of 2*R*,4*R*-4-aminopyrrolidine-2,4-dicarboxylic acid. *Bioorg. Med. Chem. Lett.*, *8*, 1985–1990.

425. Kozikowski, A. P., Araldi, G. L., Tückmantel, W., Pshenichkin, S., Surina, E., Wroblewski, J. T. (1999). 1-Amino-APDC, a partial agonist of group II metabotropic glutamate receptors with neuroprotective properties. *Bioorg. Med. Chem. Lett.*, *9*, 1721–1726.

426. Acher, F. C., Tellier, F. J., Azerad, R., Brabet, I. N., Fagni, L., Pin, J. P. (1997). Synthesis and pharmacological characterization of aminocyclopentanetricarboxylic acids: New tools to discriminate between metabotropic glutamate receptor subtypes. *J. Med. Chem.*, *40*, 3119–3129.

427. Bräuner-Osborne, H., Madsen, U., Mikiciuk-Olasik, E., Curry, K. (1997). New analogues of ACPD with selective activity for group II metabotropic glutamate receptors. *Eur. J. Pharmacol.*, *332*, 327–331.

428. Monn, J. A., Valli, M. J., Massey, S. M., Wright, R. A., Salhoff, C. R., Johnson, B. G., Howe, T., Alt, C. A., Rhodes, G. A., Robey, R. L., Griffey, K. R., Tizzano, J. P., Kallman, M. J., Helton, D. R., Schoepp, D. D. (1997). Design, synthesis and pharmacological characterization of (+)-2-aminobicyclo[3.1.0]hexane-2,6-dicarboxylic acid (LY354740): A potent, selective and orally active group 2 metabotropic glutamate receptor agonist possessing anticonvulsant and anxiolytic properties. *J. Med. Chem.*, *40*, 528–537.

429. Schoepp, D. D., Johnson, B. G., Wright, R. A., Salhoff, C. R., Mayne, N. G., Wu, S., Cockerham, S. L., Burnett, J. P., Belegaje, R., Bleakman, D., Monn, J. A. (1997). LY354740 is a potent and highly selective group II metabotropic glutamate receptor agonist in cells expressing human glutamate receptors. *Neuropharmacology*, *36*, 1–11.

430. Kozikowski, A. P., Steensma, D., Araldi, G. L., Tückmantel, W., Wang, S., Pshenichkin, S., Surina, E., Wroblewski, J. T. (1998). Synthesis and biology of the conformationally restricted ACPD analogue, 2-aminobicyclo[2.1.1]hexane-2,5-dicarboxylic acid-I, a potent mGluR agonist. *J. Med. Chem.*, *41*, 1641–1650.

431. Jullian, N., Brabet, I., Pin, J. P., Acher, F. C. (1999). Agonist selectivity of mGluR1 and mGluR2 metabotropic receptors: A different environment but similar recognition of an extended glutamate conformation. *J. Med. Chem.*, *42*, 1546–1555.

432. Costantino, G., Macchiarulo, A., Pellicciari, R. (1999). Pharmacophore models of group I and group II metabotropic glutamate receptor agonists. Analysis of conformational, steric, and topological parameters affecting potency and selectivity. *J. Med. Chem.*, *42*, 2816–2827.

433. Bessis, A.-S., Jullian, N., Coudert, E., Pin, J.-P., Acher, F. (1999). Extended glutamate activates metabotropic glutamate receptor types 1, 2 and 4: Selective features at mGluR4 binding site. *Neuropharmacology*, *38*, 1543–1551.

434. Kozikowski, A. P., Araldi, G. L., Flippen-Anderson, J., George, C., Pshenichkin, S., Surina, E., Wroblewski, J. (1998). Synthesis and metabotropic glutamate receptor activity of a 2-aminobicyclo[3.2.0]heptane-2,5-dicarboxylic acid, a molecule possessing an extended glutamate conformation. *Bioorg. Med. Chem. Lett.*, *8*, 925–930.

435. Tellier, F., Acher, F., Brabet, I., Pin, J. P., Azerad, R. (1998). Aminobicyclo[2.2.1]heptane dicarboxylic acids (ABHD), rigid analogs of ACPD and glutamic acid: Synthesis and pharmacological activity on metabotropic receptors mGluR1 and mGluR2. *Bioorg. Med. Chem.*, *6*, 195–208.

436. Hayashi, Y., Tanabe, Y., Aramori, I., Masu, M., Shimamoto, K., Ohfune, Y., Nakanishi, S. (1992). Agonist analysis of 2-(carboxycyclopropyl)glycine isomers for cloned metabotropic glutamate receptor subtypes expressed in Chinese hamster ovary cells. *Br. J. Pharmacol.*, *107*, 539–543.

437. Brabet, I., Parmentier, M. L., De Colle, C., Bockaert, J., Acher, F., Pin, J. P. (1998). Comparative effect of L-CCG-I, DCG-IV and gamma-carboxy-L-glutamate on all cloned metabotropic glutamate receptor subtypes. *Neuropharmacology*, *37*, 1043–1051.

438. Adam, G., Hennig, M., Kolczewski, S., Ohresser, S., Wichmann, J., Woltering, T., Mutel, V. (1999). Synthesis and receptor pharmacology of HYDIA; a new group II mGlu receptor antagonist. *Neuropharmacology*, *38*, A1.

439. Thomsen, C., Hansen, L., Suzdak, P. D. (1994). L-Glutamate uptake inhibitors may stimulate phosphoinositide hydrolysis in baby hamster kidney cells expressing mGluR1a via heteroexchange with L-glutamate without direct activation of mGluR1a. *J. Neurochem.*, *63*, 2038–2047.

440. Bräuner-Osborne, H., Sløk, F. A., Skjærbæk, N., Ebert, B., Sekiyama, N., Nakanishi, S., Krogsgaard-Larsen, P. (1996). A new highly selective metabotropic excitatory amino acid agonist: 2-amino-4-(3-hydroxy-5-methylisoxazol-4-yl)butyric acid. *J. Med. Chem.*, *39*, 3188–3194.

441. Ahmadian, H., Nielsen, B., Bräuner-Osborne, H., Johansen, T. N., Stensbøl, T. B., Sløk, F. A., Sekiyama, N., Nakanishi, S., Krogsgaard-Larsen, P., Madsen, U. (1997). (*S*)-Homo-AMPA, a specific agonist at the mGlu6 subtype of metabotropic glutamic acid receptors. *J. Med. Chem.*, *40*, 3700–3705.

442. Bräuner-Osborne, H., Nielsen, B., Krogsgaard-Larsen, P. (1998). Molecular pharmacology of homologues of ibotenic acid at cloned metabotropic glutamic acid receptors. *Eur. J. Pharmacol.*, *350*, 311–316.

443. Nicoletti, F., Meek, J. L., Iadarola, M. J., Chuang, D. M., Roth, B. L., Costa, E. (1986). Coupling of inositol phospholipid metabolism with excitatory amino acid recognition sites in rat hippocampus. *J. Neurochem.*, *46*, 40–46.

444. Masu, M., Tanabe, Y., Tsuchida, K., Shigemoto, R., Nakanishi, S. (1991). Sequence and expression of a metabotropic glutamate receptor. *Nature*, *349*, 760–765.

445. Houamed, K. M., Kuijper, J. L., Gilbert, T. L., Haldeman, B. A., O'Hara, P. J., Mulvihill, E. R., Almers, W., Hagen, F. S. (1991). Cloning, expression, and gene structure of a G protein-coupled glutamate receptor from rat brain. *Science*, *252*, 1318–1321.

446. Madsen, U., Bräuner-Osborne, H., Frydenvang, K., Hvene, L., Johansen, T. N., Nielsen, B., Sánchez, C., Stensbøl, T. B., Bischoff, F., Krogsgaard-Larsen, P. (2001). Synthesis and pharmacology of 3-isoxazolol amino acids as selective antagonists at group I metabotropic glutamate receptors. *J. Med. Chem.*, *44*, 1051–1059.

447. Bräuner-Osborne, H., Krogsgaard-Larsen, P. (1998). Pharmacology of (*S*)-homoquisqualic acid and (*S*)-2-amino-5-phosphonopentanoic acid [(*S*)-AP5] at cloned metabotropic glutamate receptors. *Br. J. Pharmacol.*, *123*, 269–274.

448. Koerner, J. F., Cotman, C. W. (1981). Micromolar L-2-amino-4-phosphonobutyric acid selectively inhibits perforant path synapses from lateral entorhinal cortex. *Brain Res.*, *216*, 192–198.

449. Duvoisin, R. M., Zhang, C. X., Ramonell, K. (1995). A novel metabotropic glutamate receptor expressed in the retina and olfactory bulb. *J. Neurosci.*, *15*, 3075–3083.

450. Johansen, P. A., Chase, L. A., Sinor, A. D., Koerner, J. F., Johnson, R. L., Robinson, M. B. (1995). Type 4a metabotropic glutamate receptor: Identification of new potent agonists and differentiation from the L-(+)-2-amino-4-phosphonobutanoic acid-sensitive receptor in the lateral perforant pathway in rats. *Mol. Pharmacol.*, *48*, 140–149.

451. Cartmell, J., Adam, G., Chaboz, S., Henningsen, R., Kemp, J. A., Klingelschmidt, A., Metzler, V., Monsma, F., Schaffhauser, H., Wichmann, J., Mutel, V. (1998). Characterization of [$^3$H]-(2*S*,2′*R*,3′*R*)-2-(2′,3′-dicarboxy-cyclopropyl)glycine ([$^3$H]-DCG-IV) binding to metabotropic $mGlu_2$ receptor-transfected cell membranes. *Br. J. Pharmacol.*, *123*, 497–504.

452. Ito, I., Kohda, A., Tanabe, S., Hirose, E., Hayashi, M., Mitsunaga, S., Sugiyama, H. (1992). 3,5-Dihydroxyphenyl-glycine: A potent agonist of metabotropic glutamate receptors. *Neuroreport*, *3*, 1013–1016.

453. Hayashi, Y., Sekiyama, N., Nakanishi, S., Jane, D. E., Sunter, D. C., Birse, E. F., Udvarhelyi, P. M., Watkins, J. C. (1994). Analysis of agonist and antagonist activities of phenylglycine derivatives for different cloned metabotropic glutamate receptor subtypes. *J. Neurosci.*, *14*, 3370–3377.

454. Thomsen, C., Boel, E., Suzdak, P. D. (1994). Actions of phenylglycine analogs at subtypes of the metabotropic glutamate receptor family. *Eur. J. Pharmacol.*, *267*, 77–84.

455. Eaton, S. A., Jane, D. E., Jones, P. L., Porter, R. H., Pook, P. C., Sunter, D. C., Udvarhelyi, P. M., Roberts, P. J., Salt, T. E., Watkins, J. C. (1993). Competitive antagonism at metabotropic glutamate receptors by (*S*)-4-carboxyphenylglycine and (*RS*)-alpha-methyl-4-carboxyphenylglycine. *Eur. J. Pharmacol.*, *244*, 195–197.

456. Clark, B. P., Baker, S. R., Goldsworth, J., Harris, J. R., Kingston, A. E. (1997). (+)-2-Methyl-4-carboxyphenylglycine (LY367385) selectively antagonises metabotropic glutamate mGluR1 receptors. *Bioorg. Med. Chem. Lett.*, *7*, 2777–2780.

457. Pellicciari, R., Luneia, R., Costantino, G., Marinozzi, M., Natalini, B., Jakobsen, P., Kanstrup, A., Lombardi, G., Moroni, F., Thomsen, C. (1995). 1-Aminoindan-1,5-dicarboxylic acid: A novel antagonist at phospholipase C-linked metabotropic glutamate receptors. *J. Med. Chem.*, *38*, 3717–3719.

458. Moroni, F., Lombardi, G., Thomsen, C., Leonardi, P., Attucci, S., Peruginelli, F., Torregrossa, S. A., Pellegrini-Giampietro, D. E., Luneia, R., Pellicciari, R. (1997). Pharmacological characterization of 1-aminoindan-1,5-dicarboxylic acid, a potent mGluR1 antagonist. *J. Pharmacol. Exp. Ther.*, *281*, 721–729.

459. Brabet, I., Mary, S., Bockaert, J., Pin, J.-P. (1995). Phenylglycine derivatives discriminate between mGluR1- and mGluR5-mediated responses. *Neuropharmacology*, *34*, 895–903.

460. Kingston, A. E., Burnett, J. P., Mayne, N. G., Lodge, D. (1995). Pharmacological analysis of 4-carboxyphenylglycine derivatives: Comparison of effects on mGluR1α and mGluR5a subtypes. *Neuropharmacology*, *34*, 887–894.

461. Clark, B. P., Harris, J. R., Kingston, A. E., McManus, D. (1998). Alpha-substituted phenylglycines as group I metabotropic glutamate receptor antagonists. *XVth Eur. Int. Symp. Med. Chem.*, 183.

462. Costantino, G., Pellicciari, R. (1996). Homology modeling of metabotropic glutamate receptors. (mGluRs) structural motifs affecting binding modes and pharmacological profile of mGluR1 agonists and competitive antagonists. *J. Med. Chem.*, *39*, 3998–4006.

463. Pellicciari, R., Raimondo, M., Marinozzi, M., Natalini, B., Costantino, G., Thomsen, C. (1996). (*S*)-(+)-2-(3′-Carboxybicyclo[1.1.1]pentyl)glycine, a structurally new group i metabotropic glutamate receptor antagonist. *J. Med. Chem.*, *39*, 2874–2876.

464. Pellicciari, R., Costantino, G., Giovagnoni, E., Mattoli, L., Brabet, I., Pin, J. P. (1998). Synthesis and preliminary evaluation of (*S*)-2-(4′-carboxycubyl)glycine, a new selective mGluR1 antagonist. *Bioorg. Med. Chem. Lett.*, *8*, 1569–1574.

465. Gasparini, F., Bruno, V., Battaglia, G., Lukic, S., Leonhardt, T., Inderbitzin, W., Laurie, D., Sommer, B., Varney, M. A., Hess, S. D., Johnson, E. C., Kuhn, R., Urwyler, S., Sauer, D., Portet, C., Schmutz, M., Nicoletti, F., Flor, P. J. (1999). (*RS*)-4-Phosphonophenylglycine, a potent and selective group III metabotropic glutamate receptor agonist, is anticonvulsive and neuroprotective in vivo. *J. Pharmacol. Exp. Ther.*, *289*, 1678–1687.

466. Jane, D. E., Pittaway, K., Sunter, D. C., Thomas, N. K., Watkins, J. C. (1995). New phenylglycine derivatives with potent and selective antagonist activity at presynaptic glutamate receptors in neonatal rat spinal cord. *Neuropharmacology*, *34*, 851–856.

467. Ma, D., Tian, H., Sun, H., Kozikowski, A. P., Pshenichkin, S., Wroblewski, J. T. (1997). Synthesis and biological activity of cyclic analogues of MPPG and MCPG as metabotropic glutamate receptor antagonists. *Bioorg. Med. Chem. Lett.*, *7*, 1195–1198.

468. Thomsen, C., Bruno, V., Nicoletti, F., Marinozzi, M., Pellicciari, R. (1996). (2*S*,1′*S*,2′*S*,3′*R*)-2-(2′-Carboxy-3′-phenylcyclopropyl)glycine, a potent and selective antagonist of type 2 metabotropic glutamate receptors. *Mol. Pharmacol.*, *50*, 6–9.

469. Gomeza, J., Mary, S., Brabet, I., Parmentier, M. L., Restituito, S., Bockaert, J., Pin, J.-P. (1996). Coupling of metabotropic glutamate receptors 2 and 4 to Gα15, Gα16, and chimeric G αq/i proteins: Characterization of new antagonists. *Mol. Pharmacol.*, *50*, 923–930.

470. Johansen, P. A., Robinson, M. B. (1995). Identification of 2-amino-2-methyl-4-phosphonobutanoic acid as an antagonist at the mGlu(4a) receptor. *Eur. J. Pharmacol.*, *290*, R1–R3.

471. Kozikowski, A. P., Steensma, D., Varasi, M., Pshenichkin, S., Surina, E., Wroblewski, J. T. (1998). α-Substituted quisqualic acid analogs: New metabotropic glutamate receptor group II selective antagonists. *Bioorg. Med. Chem. Lett.*, *8*, 447–452.

472. Wermuth, C. G., Mann, A., Schoenfelder, A., Wright, R. A., Johnson, B. G., Burnett, J. P., Mayne, N. G., Schoepp, D. D. (1996). (2*S*,4*S*)-2-Amino-4-(4,4-diphenylbut-1-yl)-pentane-1,5-dioic acid: A potent and selective antagonist for metabotropic glutamate receptors negatively linked to adenylate cyclase. *J. Med. Chem.*, *39*, 814–816.

473. Escribano, A., Ezquerra, J., Pedregal, C., Rubio, A., Yruretagoyena, B., Baker, S. R., Wright, R. A., Johnson, B. G., Schoepp, D. D. (1998). (2S,4S)-amino-4-(2,2-

diphenylethyl)pentanedioic acid selective group 2 metabotropic glutamate receptor antagonist. *Bioorg. Med. Chem. Lett.*, *8*, 765–770.

474. Bräuner-Osborne, H., Nielsen, B., Stensbøl, T. B., Johansen, T. N., Skjærbæk, N., Krogsgaard-Larsen, P. (1997). Molecular pharmacology of 4-substituted glutamic acid analogues at ionotropic and metabotropic excitatory amino acid receptors. *Eur. J. Pharmacol.*, *335*, R1–R3.
475. Ornstein, P. L., Bleisch, T. J., Arnold, M. B., Wright, R. A., Johnson, B. G., Schoepp, D. D. (1998). 2-Substituted (2*SR*)-2-amino-2-((1*SR*,2*SR*)-2-carboxycycloprop-1-yl)glycines as potent and selective antagonists of group II metabotropic glutamate receptors. 1. Effects of alkyl, arylalkyl, and diarylalkyl substitution. *J. Med. Chem.*, *41*, 346–357.
476. Kingston, A. E., Ornstein, P. L., Wright, R. A., Johnson, B. G., Mayne, N. G., Burnett, J. P., Belagaje, R., Wu, S., Schoepp, D. D. (1998). LY341495 is a nanomolar potent and selective antagonist of group II metabotropic glutamate receptors. *Neuropharmacology*, *37*, 1–12.
477. Wu, S., Wright, R. A., Rockey, P. K., Burgett, S. G., Arnold, J. S., Rosteck, P. R., Jr., Johnson, B. G., Schoepp, D. D., Belagaje, R. M. (1998). Group III human metabotropic glutamate receptors 4, 7 and 8: Molecular cloning, functional expression, and comparison of pharmacological properties in RGT cells. *Brain. Res. Mol. Brain. Res.*, *53*, 88–97.
478. Pellicciari, R., Costantino, G., Marinozzi, M., Macchiarulo, A., Amori, L., Josef Flor, P., Gasparini, F., Kuhn, R., Urwyler, S. (2001). Design, synthesis and preliminary evaluation of novel 3′-substituted carboxycyclopropylglycines as antagonists at group 2 metabotropic glutamate receptors. *Bioorg. Med. Chem. Lett.*, *11*, 3179–3182.
479. Chen, Y., Bacon, G., Sher, E., Clark, B. P., Kallman, M. J., Wright, R. A., Johnson, B. G., Schoepp, D. D., Kingston, A. E. (2000). Evaluation of the activity of a novel metabotropic glutamate receptor antagonist (+/–)-2-amino-2-(3-*cis* and *trans*-carboxycyclobutyl-3-(9-thioxanthyl)propionic acid) in the in vitro neonatal spinal cord and in an in vivo pain model. *Neuroscience*, *95*, 787–793.
480. Annoura, H., Fukunaga, A., Uesugi, M., Tatsuoka, T., Horikawa, Y. (1996). A novel class of antagonists for metabotropic glutamate receptors, 7-(hydroxyimino)cyclopropan[*b*]chromen-1*a*-carboxylates. *Bioorg. Med. Chem. Lett.*, *6*, 763–766.
481. Hermans, E., Nahorski, S. R., Challiss, R. A. (1998). Reversible and noncompetitive antagonist profile of CPCCOEt at the human type 1α metabotropic glutamate receptor. *Neuropharmacology*, *37*, 1645–1647.
482. Litschig, S., Gasparini, F., Rüegg, D., Stoehr, N., Flor, P. J., Vranesic, I., Prézeau, L., Pin, J.-P., Thomsen, C., Kuhn, R. (1999). CPCCOEt, a noncompetitive metabotropic glutamate receptor 1 antagonist, inhibits receptor signaling without affecting glutamate binding. *Mol. Pharmacol.*, *55*, 453–461.
483. Bräuner-Osborne, H., Jensen, A. A., Krogsgaard-Larsen, P. (1999). Interaction of CPCCOEt with a chimeric $mGlu_{1b}$ and calcium sensing receptor. *Neuroreport*, *10*, 3923–3925.
484. Carroll, F. Y., Stolle, A., Beart, P. M., Voerste, A., Brabet, I., Mauler, F., Joly, C., Antonicek, H., Bockaert, J., Müller, T., Pin, J. P., Prézeau, L. (2001). BAY36-7620: A potent non-competitive $mGlu_1$ receptor antagonist with inverse agonist activity. *Mol. Pharmacol.*, *59*, 965–973.

485. Malherbe, P., Kratochwil, N., Knoflach, F., Zenner, M. T., Kew, J. N., Kratzeisen, C., Maerki, H. P., Adam, G., Mutel, V. (2003). Mutational analysis and molecular modeling of the allosteric binding site of a novel, selective, noncompetitive antagonist of the metabotropic glutamate 1 receptor. *J. Biol Chem.*, *278*, 8340–8347.

486. Kolczewski, S., Adam, G., Stadler, H., Mutel, V., Wichmann, J., Woltering, T. (1999). Synthesis of heterocyclic enol ethers and their use as group 2 metabotropic glutamate receptor antagonists. *Bioorg. Med. Chem. Lett.*, *9*, 2173–2176.

487. Adam, G., Kolczewski, S., Ohresser, S., Wichmann, J., Woltering, T., Mutel, V. (1999). Synthesis, structure-activity relationship and receptor pharmacology of new non competitive, non-amino acid, mGlu2 receptor selective antagonists. *Neuropharmacology*, *10*, A1.

488. Varney, M. A., Cosford, N. D., Jachec, C., Rao, S. P., Sacaan, A., Lin, F. F., Bleicher, L., Santori, E. M., Flor, P. J., Allgeier, H., Gasparini, F., Kuhn, R., Hess, S. D., Veli elebi, G., Johnson, E. C. (1999). SIB-1757 and SIB-1893: Selective, noncompetitive antagonists of metabotropic glutamate receptor type 5. *J. Pharmacol. Exp. Ther.*, *290*, 170–181.

489. Gasparini, F., Lingenhöhl, K., Stoehr, N., Flor, P. J., Heinrich, M., Vranesic, I., Biollaz, M., Allgeier, H., Heckendorn, R., Urwyler, S., Varney, M. A., Johnson, E. C., Hess, S. D., Rao, S. P., Sacaan, A. I., Santori, E. M., Veliçelebi, G., Kuhn, R. (1999). 2-Methyl-6-(phenylethynyl)-pyridine (MPEP), a potent, selective and systemically active $mGlu_5$ receptor antagonist. *Neuropharmacology*, *38*, 1493–1503.

490. Pagano, A., Ruegg, D., Litschig, S., Stoehr, N., Stierlin, C., Heinrich, M., Floersheim, P., Prezeau, L., Carroll, F., Pin, J. P., Cambria, A., Vranesic, I., Flor, P. J., Gasparini, F., Kuhn, R. (2000). The non-competitive antagonists 2-methyl-6-(phenylethynyl)pyridine and 7-hydroxyiminocyclopropan[*b*]chromen-1*a*-carboxylic acid ethyl ester interact with overlapping binding pockets in the transmembrane region of group I metabotropic glutamate receptors. *J. Biol Chem.*, *275*, 33750–33758.

491. Mathiesen, J. M., Svendsen, N., Bräuner-Osborne, H., Thomsen, C., Ramirez, M. T. (2003). Positive allosteric modulation of the human metabotropic glutamate receptor 4 (hmGluR4) by SIB-1893 and MPEP. *Br. J. Pharmacol.*, *138*, 1026–1030.

492. Cosford, N. D., Tehrani, L., Roppe, J., Schweiger, E., Smith, N. D., Anderson, J., Bristow, L., Brodkin, J., Jiang, X., McDonald, I., Rao, S., Washburn, M., Varney, M. A. (2003). 3-[(2-Methyl-1,3-thiazol-4-yl)ethynyl]-pyridine: A potent and highly selective metabotropic glutamate subtype 5 receptor antagonist with anxiolytic activity. *J. Med. Chem.*, *46*, 204–206.

493. Schaffhauser, H., Rowe, B. A., Morales, S., Chavez-Noriega, L. E., Yin, R., Jachec, C., Rao, S. P., Bain, G., Pinkerton, A. B., Vernier, J. M., Bristow, L. J., Varney, M. A., Daggett, L. P. (2003). Pharmacological characterization and identification of amino acids involved in the positive modulation of metabotropic glutamate receptor subtype 2. *Mol. Pharmacol.*, *64*, 798–810.

494. O'Brien, J. A., Lemaire, W., Chen, T. B., Chang, R. S., Jacobson, M. A., Ha, S. N., Lindsley, C. W., Schaffhauser, H. J., Sur, C., Pettibone, D. J., Conn, P. J., Williams, D. L., Jr. (2003). A family of highly selective allosteric modulators

of the metabotropic glutamate receptor subtype 5. *Mol. Pharmacol.*, *64*, 731–740.

495. Spooren, W., Ballard, T., Gasparini, F., Amalric, M., Mutel, V., Schreiber, R. (2003). Insight into the function of Group I and Group II metabotropic glutamate (mGlu) receptors: Behavioural characterization and implications for the treatment of CNS disorders. *Behav. Pharmacol.*, *14*, 257–277.
496. Gasparini, F., Kuhn, R., Pin, J. P. (2002). Allosteric modulators of group I metabotropic glutamate receptors: Novel subtype-selective ligands and therapeutic perspectives. *Curr. Opin. Pharmacol.*, *2*, 43–49.
497. Moldrich, R. X., Chapman, A. G., De Sarro, G., Meldrum, B. S. (2003). Glutamate metabotropic receptors as targets for drug therapy in epilepsy. *Eur. J. Pharmacol.*, *476*, 3–16.
498. Flor, P. J., Battaglia, G., Nicoletti, F., Gasparini, F., Bruno, V. (2002). Neuroprotective activity of metabotropic glutamate receptor ligands. *Adv. Exp. Med. Biol.*, *513*, 197–223.
499. Kunishima, N., Shimada, Y., Tsuji, Y., Sato, T., Yamamoto, M., Kumasaka, T., Nakanishi, S., Jingami, H., Morikawa, K. (2000). Structural basis of glutamate recognition by a dimeric metabotropic glutamate receptor. *Nature*, *407*, 971–977.
500. Tsuchiya, D., Kunishima, N., Kamiya, N., Jingami, H., Morikawa, K. (2002). Structural views of the ligand-binding cores of a metabotropic glutamate receptor complexed with an antagonist and both glutamate and Gd3+. *Proc. Natl. Acad. Sci. USA*, *99*, 2660–2665.
501. Romano, C., Yang, W. L., O'Malley, K. L. (1996). Metabotropic glutamate receptor 5 is a disulfide-linked dimer. *J. Biol Chem.*, *271*, 28612-28616.
502. Costantino, G., Macchiarulo, A., Pellicciari, R. (2001). Homology model of the closed, functionally active, form of the amino terminal domain of mGlur1. *Bioorg. Med. Chem.*, *9*, 847–852.
503. Jensen, A. A., Sheppard, P. O., Jensen, L. B., O'Hara, P. J., Bräuner-Osborne, H. (2001). Construction of a high affinity zinc binding site in the metabotropic glutamate receptor mGluR1: Noncompetitive antagonism originating from the amino-terminal domain of a family C G-protein-coupled receptor. *J. Biol Chem.*, *276*, 10110–10118.
504. Clausen, R. P., Bräuner-Osborne, H., Greenwood, J. R., Hermit, M. B., Stensbøl, T. B., Nielsen, B., Krogsgaard-Larsen, P. (2002). Selective agonists at group II metabotropic glutamate receptors: Synthesis, stereochemistry, and molecular pharmacology of (*S*)- and (*R*)-2-amino-4-(4-hydroxy[1,2,5]thiadiazol-3-yl)butyric acid. *J. Med. Chem.*, *45*, 4240–4245.
505. Bessis, A. S., Rondard, P., Gaven, F., Brabet, I., Triballeau, N., Prezeau, L., Acher, F., Pin, J. P. (2002). Closure of the Venus flytrap module of mGlu8 receptor and the activation process: Insights from mutations converting antagonists into agonists. *Proc. Natl. Acad. Sci. USA*, *99*, 11097–11102.
506. Sato, T., Shimada, Y., Nagasawa, N., Nakanishi, S., Jingami, H. (2003). Amino acid mutagenesis of the ligand binding site and the dimer interface of the metabotropic glutamate receptor 1. Identification of crucial residues for setting the activated state. *J. Biol Chem.*, *278*, 4314–4321.

507. Yao, Y., Pattabiraman, N., Michne, W. F., Huang, X. P., Hampson, D. R. (2003). Molecular modeling and mutagenesis of the ligand-binding pocket of the mGlu3 subtype of metabotropic glutamate receptor. *J. Neurochem.*, *86*, 947–957.
508. Hermit, M. B., Greenwood, J. R., Nielsen, B., Bunch, L., Jørgensen, C. G., Vestergaard, H. T., Stensbøl, T. B., Sanchez, C., Krogsgaard-Larsen, P., Madsen, U., Bräuner-Osborne, H. (2004). Ibotenic acid and thioibotenic acid: A remarkable difference in activity at group III metabotropic glutamate receptors. *Eur. J. Pharmacol.*, *486*, 241–250.

# CARDIAC SARCOLEMMAL ATP-SENSITIVE POTASSIUM CHANNEL ANTAGONISTS: NOVEL ISCHEMIA-SELECTIVE ANTIARRHYTHMIC AGENTS

George E. Billman

*The Ohio State University*
*Columbus, Ohio*

1 INTRODUCTION
2 EFFECTS OF MYOCARDIAL ISCHEMIA ON EXTRACELLULAR POTASSIUM
3 EFFECT OF EXTRACELLULAR POTASSIUM ON VENTRICULAR ARRHYTHMIAS
4 EFFECT OF ATP-SENSITIVE POTASSIUM CHANNEL ANTAGONISTS ON VENTRICULAR ARRHYTHMIAS
Nonselective ATP-Sensitive Potassium Channel Antagonists
Selective ATP-Sensitive Potassium Channel Antagonists
Proarrhythmic Effects of ATP-Sensitive Potassium Channel Agonists
5 SUMMARY
References

## 1 INTRODUCTION

Sudden cardiac death (defined as unexpected death from cardiac causes that occurs within 1 h of the onset of symptoms [1]) remains the leading cause of

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

death in industrialized countries, accounting for between 300,000 and 500,000 deaths each year in the United States [2–5]. Holter monitoring studies reveal that these sudden deaths most frequently (up to 93 percent) resulted from ventricular tachyarrhythmias [6–8]. Despite the enormity of this problem, the development of safe and effective antiarrhythmic agents remains elusive. In fact, several initially promising antiarrhythmic drugs have actually been shown to increase, rather than to decrease, the risk for arrhythmic death in patients recovering from myocardial infarction. For example, the now infamous Cardiac Arrhythmia Suppression Trial (CAST study [9]) demonstrated that, although class I antiarrhythmic drugs (i.e., drugs that block sodium channels) effectively suppressed premature ventricular contractions, some of these compounds (flecainide and encainide) increased the risk for arrhythmic cardiac death. In a similar manner, many class III antiarrhythmic drugs (drugs that prolong refractory period, most likely via modulation of potassium channels) have been shown to prolong QT interval, to promote the life-threatening tachyarrhythmia torsade de pointes (i.e., polymorphic ventricular tachycardia in which the QRS waves seem to "twist" around the baseline), and to increase cardiac mortality in some patient populations [10, 11]. Unfortunately, only a few drugs have been clinically proven to reduce cardiac mortality in high-risk patients, such as patients recovering from myocardial infarction. To date, only β-adrenergic receptor antagonists and amiodarone, a class III antiarrhythmic drug that also blocks β-adrenergic receptors, have been shown to reduce sudden cardiac death [4, 12–15]. Even the best currently available therapies have not been completely successful in the suppression of malignant arrhythmias, and they also frequently exhibit untoward side effects. For example, mortality following myocardial infarction remains high among patients with substantial ventricular dysfunction, even with β-adrenergic receptor antagonist therapy. The 1-year mortality is 10 percent or higher, with sudden death accounting for approximately one-third of the deaths in these high-risk patients [16]. In a similar manner, the long-term use of amiodarone is limited due to adverse side effects, including pulmonary fibrous and thyroid toxicity [124]. Clearly, the "ideal" antiarrhythmic drug has yet to be discovered.

The discouraging clinical experience with most antiarrhythmic drugs indicates that these agents may attack the wrong therapeutic target. Although these drugs were effective against some types of arrhythmias, they clearly failed to prevent those arrhythmias that culminate in death. In order to develop agents that will reduce sudden death, one must first identify those factors that trigger the malignant arrhythmias. As previously noted, ventricular tachyarrhythmias are responsible for sudden cardiac death. Only a minority of these patients had a known history of heart disease prior to the collapse, yet up to 90 percent of all sudden death patients were subsequently shown to have underlying coronary artery disease [2]. Thus, it is likely that myocardial ischemia plays a crucial role in the induction of the lethal arrhythmias in these patients. If this hypothesis is correct, then it is crucial to identify the factor or factors that render the ischemic myocardium vulnerable to arrhythmia for-

mation. Once identified, it should then be possible to develop therapeutic interventions that correct these ischemically induced changes in cardiac electrical stability and, thereby, prevent sudden death. The most effective antiarrhythmic agents would be those that target the ischemic myocardium with little or no action on the normal (i.e., nonischemic) cardiac tissue. As a consequence of this selectivity, one would expect that these drugs should also have a low propensity for proarrhythmic events.

It is well established that myocardial ischemia provokes abnormalities in the biochemical homeostasis of individual cardiac cells. These intracellular changes culminate in the disruption of cellular electrophysiologic properties, and as a result life-threatening alterations in cardiac rhythm, such as ventricular fibrillation, frequently occur. A number of different chemical substances have been proposed as possible causative factors in the genesis of ventricular fibrillation during myocardial ischemia, including catecholamines, amphiphilic products of lipid metabolism, various peptides, cytosolic calcium accumulation, and increases in extracellular potassium [17–19]. An accumulating body of evidence suggests that the activation of specific potassium channels, the adenosine triphosphate (ATP)-sensitive potassium channels, leads to potassium efflux from the ischemic cells, thereby inducing the electrophysiologic changes that are ultimately responsible for the formation of lethal cardiac arrhythmias [20–22]. As such, drugs that selectively inhibit these channels should prove to be particularly effective in the prevention of sudden cardiac death. Therefore, this chapter will first discuss the relationship between ischemically induced alterations in extracellular potassium and arrhythmia formation and then evaluate the antiarrhythmic potential of ATP-sensitive potassium channel antagonists, drugs that may act selectively on the ischemic myocardium.

## 2 EFFECTS OF MYOCARDIAL ISCHEMIA ON EXTRACELLULAR POTASSIUM

Myocardial ischemia provokes both rapid increases in extracellular potassium and reductions in action potential duration. Harris and co-workers [23, 24] were the first to show that extracellular potassium rises dramatically after coronary artery ligation, correlating with the onset of ventricular arrhythmias. They further demonstrated that intracoronary injections of potassium chloride provoked electrocardiographic (ECG) changes and triggered ventricular arrhythmias similar to those induced by myocardial ischemia [23, 24]. They proposed that changes in extracellular potassium represented a major factor in the development of malignant arrhythmias during ischemia. Subsequent studies that used ion-selective electrodes to measure potassium activity directly have confirmed these earlier observations [25–27]. Extracellular potassium increased within the first 15 s and reached a plateau within 5 to 10 min after the interruption of coronary perfusion [25–28]. Furthermore,

regional differences or inhomogeneities of potassium accumulation were recorded, accompanied by corresponding differences in ventricular electrical activity [25, 26, 28]. This extracellular potassium accumulation resulted primarily from increases in potassium efflux rather than from decreased potassium influx due to inhibition of the $Na^+/K^+$-ATPase [30–32].

A growing body of evidence demonstrates that the ischemically induced potassium accumulation and the corresponding reductions in action potential duration result, primarily, from the opening of ATP-sensitive potassium channels [33]. Using the patch clamp technique, Trube and Hescheler [34] were the first to record single ATP-sensitive potassium channel activity. Noma [35] and Trube and Hescheler [34] further demonstrated that reductions in cellular ATP induced by cyanide exposure evoked an outward potassium current. They proposed that the activation of an ATP-sensitive potassium channel might be responsible for the reductions in action potential duration induced by hypoxia. Several studies further implicated the activation of this current in the changes in cardiac action potential and extracellular potassium accumulation during myocardial ischemia [36–46]. The ATP-sensitive potassium channel inhibitor, glibenclamide, for example, has been shown either to attenuate or to abolish reductions in action potential duration in hypoxic myocytes [37, 38], isolated cardiac tissue [41–43, 45, 46], and regionally or globally ischemic hearts [36, 40, 47]. This sulfonylurea drug also reduced extracellular potassium accumulation induced by ischemia [39, 43–45]. Conversely, ATP-sensitive potassium channel agonists (pinacidil, cromakalim) exacerbated ischemically induced reductions in action potential duration, as well as promoted extracellular potassium accumulation [36–38, 40–43, 47–50].

However, it must be emphasized that the ATP-sensitive potassium channel is activated only at low ATP concentrations with half-maximum suppression of channel opening at 20 to 100 μM [34, 35, 51, 52], yet intracellular concentrations are normally much higher (5 to 10 mM). Furthermore, cytosolic ATP levels remain in the millimolar range for the first 10 min of hypoxia, well after potassium accumulation begins [32, 53] Therefore, the role that this channel plays in the response to myocardial ischemia can be questioned [33]. A number of studies provide an explanation for this apparent paradox. Cardiac tissue contains a very high density of ATP-sensitive potassium channels. As a consequence, only a small increase in the open-state probability (<1% of maximum) was required to shorten action potential duration during ischemia [54–56]. Indeed, Faivre and Findlay [54] found, in guinea pig myocytes, that the opening of only 30 channels (less than 1 percent of the population) provoked a 50 percent reduction in action potential duration. They concluded that "physiologically relevant activity of the $K_{ATP}$ channel in cardiac membrane is confined to a very small percentage of the possible cell $K_{ATP}$ current, and thus intracellular ATP would not have to fall very far before the opening of $K_{ATP}$ channels would influence cardiac excitability" [54]. In a similar manner, Deutsch et al. [57] correlated potassium current during hypoxia in an intact rabbit papillary muscle. They showed that during hypoxia, a significant short-

ening of action potential duration (blocked by glibenclamide) occurred when tissue ATP levels fell by approximately 25 percent. They also concluded that only modest changes in cellular ATP were required to induce major changes in cardiac electrical properties.

The activation of the ATP-sensitive potassium channel may also contribute significantly to the S-T segment changes associated with myocardial ischemia. In anesthetized open-chest dogs [58] the intracoronary injection of the ATP-sensitive potassium channel opener, pinacidil, elicited elevations in the S-T segment very similar to those induced by myocardial ischemia. Abnormal electrocardiographic (T wave) changes suggestive of alterations of repolarization have also been reported in up to 30 percent of patients taking pinacidil as an antihypertensive medication [59]. Conversely, glibenclamide attenuated the S-T segment elevations induced by the occlusion of the left anterior descending coronary artery [60]. Similar results were obtained in conscious dogs. Billman et al. [61] demonstrated that either glibenclamide or the cardioselective ATP-sensitive potassium channel antagonist HMR 1098 attenuated ischemically induced S-T segment changes. They further reported that these drugs prevented ischemically induced increases in the descending portion of the T wave (an index of the transmural dispersion of repolarization [62]). Finally, myocardial ischemia failed to alter the ECG of mice in which the *Kir 6.2* gene (the gene responsible for the pore forming subunit of the cardiac ATP-sensitive potassium channel [63]) had been removed [64]. The large changes in the ECG that were induced by ligation of the left coronary artery in the wild-type control mice could be suppressed by prior treatment with the cardioselective ATP-sensitive potassium channel antagonist HMR 1098 [64]. It therefore seems likely that the activation of the ATP-sensitive potassium channel contributes significantly to alterations in cardiac electrical stability induced by myocardial ischemia, leading to the formation of malignant arrhythmias.

## 3 EFFECT OF EXTRACELLULAR POTASSIUM ON VENTRICULAR RHYTHM

Cardiac arrhythmias can result from either abnormalities in impulse generation or impulse conduction [65]. The extracellular accumulation of potassium induced by myocardial ischemia can provoke these perturbations in cardiac electrical activity. Elevations in extracellular potassium promote the depolarization of the tissue surrounding the ischemic regions, as noted above. The flow of this injury current (electrotonic current flow between ischemic and normal cells) has been implicated as a potential cause for the initiation of premature ventricular beats. Under normal conditions ventricular cells do not display a spontaneous rhythm. However, an automatic rhythm can be produced when the cells are partially depolarized [65]. Coronel et al. [29], in fact, demonstrated an increased excitability of normal tissue near the border of the

ischemia, a region of the heart in which extracellular potassium concentrations were also increased.

Changes in action potential duration induced by alterations in potassium efflux can also provoke abnormalities of impulse conduction. As noted above, increased potassium efflux from ischemic tissue triggers a reduction in action potential duration. A major factor contributing to ventricular fibrillation, particularly during myocardial ischemia, is a dispersion or inhomogeneity of the refractory period that is related to regional differences in action potential duration [65]. This allows for the fragmentation of impulse conduction during ensuing beats. As previously noted, the activation of the ATP-sensitive potassium channel produced large reductions in action potential duration [36–43, 45–47], which are inhibited by glibenclamide [36–38, 40–42, 47] but exacerbated by ATP-sensitive potassium channel agonists [43, 48, 66–68]. A differential ATP sensitivity of the ATP-dependent potassium channel has also been reported between endocardial and epicardial cells during ischemia such that smaller reductions in ATP were necessary to activate potassium channels located in epicardial tissue [69]. As a result, an inhomogeneity in extracellular potassium and refractory period, as well as a gradient in action potential duration, was recorded between the epicardial and endocardial tissue [69–71]. Nonuniformities in refractory period could set the stage for the formation of irregular reentrant pathways and ventricular arrhythmias. In a similar manner, the ATP-sensitive potassium channel antagonist glibenclamide was shown to attenuate ischemically induced reductions in the refractory period [61, 72]. In contrast, activation of the ATP-sensitive potassium current with pinacidil elicited a marked dispersion of repolarization and refractory period between the epicardium and endocardium, leading to the development of extrasystoles [73]. These effects could be abolished by glibenclamide [73]. More recently, Coromilas et al. [74] demonstrated that the pinacidil restored excitability in the tissue adjacent to the ischemic region (i.e., the epicardial border zone). This reactivation of formerly inexcitable tissue led to the formation of reentrant circuits and the induction of ventricular tachycardia. Thus, the activation of the ATP-sensitive potassium channel could contribute significantly to the induction of malignant arrhythmias by changing impulse generation (depolarization-induced changes in automaticity), conduction (refractory period dispersion), or a combination of both.

## 4 EFFECT OF ATP-SENSITIVE POTASSIUM CHANNEL ANTAGONISTS ON VENTRICULAR ARRHYTHMIAS

As noted above, it is now generally accepted that the activation of the ATP-sensitive potassium channel during myocardial ischemia provokes a potassium efflux and reductions in action potential duration that lead to dispersion of repolarization. Since heterogeneity of repolarization plays a crucial role in the induction of ventricular fibrillation, drugs that prevent ATP-sensitive potas-

sium channel activation should be particularly effective in the suppression of malignant arrhythmias induced by ischemia.

## Nonselective ATP-Sensitive Potassium Channel Antagonists

The sulfonylurea drug glibenclamide (Fig. 1) prevented hypoxia-induced reductions in action potential duration in single cells, isolated hearts, and intact animals [36, 37, 40–43, 45–47, 75–79]. This drug also attenuated the ischemically induced changes in the S-T segment in intact anesthetized or unanesthetized animals [60, 61]. Therefore, one would predict that glibenclamide should also prevent arrhythmias that arise as a consequence of these ischemically induced changes in cardiac electrical properties.

Glibenclamide prevented arrhythmias induced by ischemia in a variety of experimental models (Table 1). For example, this drug abolished ventricular arrhythmias induced by ischemia in isolated hearts [40, 80–83]. Furthermore, both glibenclamide and glimepride reduced blood glucose and decreased the incidence of irreversible ventricular fibrillation induced by reperfusion (after a 6-min coronary occlusion) in anesthetized rats [84, 140]. Glibenclamide also reduced the incidence of life-threatening arrhythmias induced by coronary artery ligation in the conscious rat [85] and anesthetized rabbit [86], as well as improved survival in anesthetized rats during ischemia and reperfusion [87]. This drug also inhibited arrhythmias associated with the intracellular calcium overload induced by ouabain [140, 144]. Billman et al. [61, 88] further demonstrated that glibenclamide prevented ventricular fibrillation induced by the combination of acute ischemia during submaximal exercise in animals previously shown to be susceptible to sudden death. It should be noted, however, that in these animals glibenclamide significantly reduced the exercise, and reactive hyperemia induced increases in coronary blood flow, as well as depressed ventricular function (large reductions in left ventricular *dP*/*dt* maximum) [61, 88]. In the isolated working rabbit heart, glibenclamide provoked an immediate decrease in coronary blood flow reducing forward flow to zero [68]. Thus, glibenclamide has potent vasoconstrictor effects due to the inhibiton of ATP-sensitive potassium channels located on the coronary vascular smooth muscle.

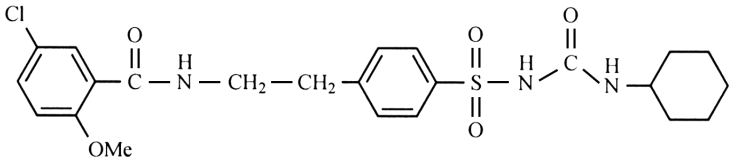


Glibenclamide

**Figure 1** Chemical structure of glibenclamide.

**TABLE 1 Effect of Nonselective ATP-Sensitive Potassium Channel Antagonist on Ischemic Cardiac Arrhythmias—Preclinical Studies**[a]

| Model | Result | Reference |
|---|---|---|
| Rabbit heart | ↓ VF | Pogasta et al. [81] |
| Rat heart | ↓ VF | Wolleben et al. [83] |
| Rat heart | ↓ VF | Kantor et al. [40] |
| Dog and rabbit hearts | No Effect | Smallwood et al. [47] |
| Guinea pig heart | ↓ VF Duration | Gwilt et al. [80] |
| Anesthetized rat | ↓ VF | Zhang et al. [140] |
| Rat heart | ↓ VF Duration | Bril et al. [141] |
| Guinea pig tissue | ↓ Re-entrant PVCs | Pasnani and Ferrier [142] |
| Conscious dog | ↓ VF | Billman et al. [88] |
| Rat heart | ↓ VF | Tosaki et al. [82] |
| Rat heart | No Effect | Rees and Curtis [143] |
| Conscious rat | ↓ VF | Lepran et al. [85] |
| Anesthetized rat | ↓ VF | Baczko et al. [87] |
| Anesthetized rabbit | ↓ VF | Barrett and Walker [86] |
| Conscious dog | ↓ VF | Billman et al. [61] |
| Anesthetized rat | ↓ VF | El Reyani et al. [84] |

[a] VF, ventricular fibrillation; PVCs, premature ventricular contractions (activations).

There are a few clinical studies that illustrate the antiarrhythmic potential of ATP-sensitive potassium channels (Table 2, 89–92]. Cacciapuoti et al. [89] found that glibenclamide significantly reduced the frequency and severity of ventricular arrhythmias recorded during transient ischemia in non-insulin-dependent diabetic patients with coronary artery disease. This drug, however, did not affect nonischemic arrhythmias nor did it change the number or the length of the ischemic episodes. Glibenclamide significantly reduced the incidence of ventricular fibrillation in non-insulin-dependent patients with acute myocardial infarction [92]. Recently, the effects of sulfonylurea drugs on the incidence of ventricular arrhtyhmias (24-h Holter monitoring) were evaluated in nondiabetic and diabetic patients with decompensated congestive heart failure [93]. These authors found that diabetic patients receiving sulfonylurea agents (glibenclamide or glipizide) exhibited a significantly lower incidence of both repetitive ventricular beats and runs of nonsustained ventricular tachycardia as compared to either nondiabetic patients or diabetic patients who did not receive sulfonlyurea drugs. It is interesting to note that glibenclamide also reduced the S-T segment elevation and prevented ventricular fibrillation induced by chest impact (a baseball delivered at 30 mph to the chest of anesthetized swine [94]). The authors concluded that "selective $K_{ATP}$ channel activation may be a pivotal mechanism in sudden death resulting from low-energy chest wall trauma" (i.e., Commotio Cordis) "in young people during sporting activities" [94].

**TABLE 2 Effect of Nonselective ATP-Sensitive Potassium Channel Antagonists on Cardiac Arrhythmias—Clinical Studies[a]**

| Patient Population | Result | Reference |
|---|---|---|
| Non-insulin-dependent diabetics treated with glibenclamide ($n = 19$) | Significant reduction in the frequency of PVCs and nonsustained VT | Cacciapuoti et al. [89] |
| Non-insulin-dependent diabetics ($n = 232, 106$ treated with glibenclamide, 126 with other hypoglycemic drugs) and nondiabetic patients all with acute myocardial infarction | Incidence of VF or sustained VT reduced in the glibencalimide-treated patients. Cardiovascular mortality highest in nonglibenclamide-treated diabetic patients | Lomuscio et al. [92]<br>Lomuscio and Fiorentini [91] |
| Retrospective analysis of acute myocardial infarction patients ($n = 5, 715, 745$ diabetic patients) | VF rates similar in nondiabetic (11.0%) and diabetic patients taking glibenclamide (11.8%) but much higher in diabetic patients taking gliclazide (18%) or insulin (22.8%) | Davis et al. [90] |
| Diabetic patients with decompensated heart failure ($n = 207$) | Significant reduction in incidence of repetitive ventricular beats and nonsustained VT in patients treated with either gilbenclamide or glipizide | Aronson et al. [93] |

[a] VF, ventricular fibrillation; VT, ventricular tachycardia; PVCs, premature ventricular contractions (activations).

It is important to emphasis that glibenclamide is not selective for cardiac tissue. As noted above, this drug can profoundly reduce coronary blood flow via actions on vascular smooth muscle and can also promote insulin release, thereby provoking hypoglycemia [61]. These noncardiac actions would limit the antiarrhythmic potential of glibenclamide in the clinic. Cardioselective compounds should have fewer side effects and would therefore provide a better therapeutic option than the nonselective ATP-sensitive potassium channel antagonist glibenclamide.

## Selective ATP-Sensitive Potassium Channel Antagonists

Several different ATP-sensitive potassium channel subtypes have been identified (Table 3). The ATP-sensitive potassium channel consists of a pore-forming subunit coupled to a sulfonylurea receptor [95–99, 139]. The functional channel forms as a hetero-octomer composed of a tetramer of the

**TABLE 3 Possible ATP-Sensitive Potassium Channel Subtypes: Combinations of the Pore Forming Units (Kir 6.1 or Kir 6.2) and the Sulfonylurea Receptors (SUR1, SUR2A, and SUR2B)[a]**

| Channel | Tissue | Effect of Activation | Effect of Inhibition | Reference |
|---|---|---|---|---|
| Kir6.1/SUR1 | Mitochondria (?) | Ischemic preconditioning | Prevents ischemic preconditioning | Liu et al. [100] |
| Kir6.1/SUR2B | Vascular smooth muscle and coronary endothelial cells | Vasodilation and increase blood low | Prevent vasodilation, promote vasoconstriciton | Suszuki et al. [63]<br>Schnizler et al. [145] |
| Kir6.2/SUR1 | Pancreatic β-cells | Insulin secretion | | Aguilar-Bryan et al. [146] |
| Kir6.2/SUR1 | Neurons | Reduce excitability or neurotransmitter release | Increase excitiability or neurotransmitter release | Liss and Roeper [147] |
| Kir6.2/SUR2A | Cardiac muscle | Decrease action potential duration | Prevent decrease in action potential duration | Inagaki et al [97] |
| Kir6.2/SUR2A | Skeletal muscle | Prevent glucose uptake—fatigue | Glucose uptake | Chutkow et al. [148]<br>Gong et al. [149] |
| Kir6.2/SUR2B | Nonvascular smooth muscle | Muscle relaxation (e.g., urinary bladder tone) | Prevent muscle relaxation | Gopalakrishnan et al. [150] |

[a] Note that neither specific actions nor a target tissue for Kir6.1/SUR2A have yet been identified.

pore and four sulfonyl receptor subunits. At present, two different pore-forming subunits have been identified, both of which produce an inward rectifier potassium current (Kir 6.1 and Kir 6.2) [64, 100]. Three different sulfonylurea receptor subtypes have been isolated; SUR1 (on pancreatic islet cells), SUR2A (on cardiac tissue), and SUR2B (on vascular smooth muscle) [96, 99, 101, 102]. Thus, six different potassium channel pore and sulfonylurea receptor combinations are possible (Table 3). Suzuki et al. [63] and Manning-Fox et al. [98] recently demonstrated that Kir 6.2 and Kir 6.1 were required for cardiac and vascular smooth muscle ATP-sensitive potassium channel activity, respectively. They concluded that Kir 6.2/SUR2A most likely forms the cardiac cell membrane ATP-sensitive potassium channel, while Kir 6.1/SUR2B is located on vascular smooth muscle. In a similar manner, Lui et al. [100] reported that the actitivy to mitochondrial ATP-sensitive potassium channels was most closely mimicked by the Kir 6.1/SUR1 subtype, an conclusion that has not yet been confirmed. Given this apparent tissue specificity and the limited number of possible channel–receptor pairings, it should be possible to develop compounds that selectively inhibit (or activate) a particular ATP-sensitive potassium channel subtype. In particular, a drug that selectively blocks the Kir 6.2/SUR2A subtype should prevent ischemically induced changes in cardiac electrical properties (e.g., reductions in action potential duration) and thereby prevent arrhythmias without the untoward side effects noted for the nonselective ATP-sensitive channel antagonist glibenclamide.

***HMR 1883, a Cardioselective ATP-Potassium Channel Antagonists***
The sulfonylthiourea drug (1-[5-[2–5-chloro-*o*-ansamide)ethyl]2-methoxyphenyl]sulfonyl]-3-methylthiourea) HMR 1883 (Fig. 2) and its sodium salt HMR 1098 were recently developed to block the cardiac ATP-sensitive potassium channel [103]. HMR 1883 inhibited the sarcolemmal cardiac ATP-sensitive potassium channel activated by the channel opener rilmakalin at a much lower concentration (guinea pig myocytes $IC_{50} = 0.6 - 2.2\,\mu M$) than was required to promote insulin release (9- to 50-fold higher concentration was required to block rat pancreatic insulinoma, RIN m5F, cells) [104]. In a similar manner, HMR 1098 reversed the action potential shortening induced by rilmakalim in human cardiomycytes ($IC_{50} = 0.42 \pm 0.008\,\mu M$) and proved to be more potent in acidic condtions ($IC_{50} = 0.24 \pm 0.009\,\mu M$) as would occur during myocardial ischemia [105]. Indeed, this drug attenuated hypoxia-induced shortening of cardiac action potential in rat and guinea pig tissue [104]. In contrast to glibenclamide, HMR 1883 did not alter hypoxia-induced increases in coronary blood flow in Langendorff perfused guinea pig hearts [104, 106]. HMR 1098 also inhibited Rb+ efflux through Kir 6.2/SUR2A channels expressed in HEK293 cells ($IC_{50} = 181\,nM$), demonstrating that this system may be used to screen for compounds with a high affinity for the cardiac sarcolemmal channel [106]. Significantly, HMR 1883 did not alter flavoprotein autofluorescence, an index of mitochondrial redox state [107]. In

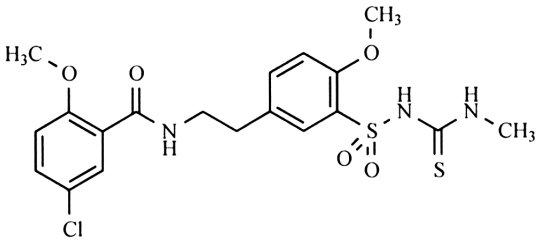


HMR 1883

**Figure 2** Chemical structure of HMR 1883.

contrast, 5-hydroxydecanoic acid, a selective blocker of mitochondrial channels, completely inhibited the flavoprotein fluorescence [108]. In agreement with these in vitro findings, HMR 1883 did not prevent the cardioprotective effects induced by ischemic preconditioning in either rat [109] or rabbit [110]. Accumulating evidence suggests that the activation of mitochondrial ATP-sensitive potassium channels plays a crucial role in the mechanical protection that results from ischemic preconditioning [111–113]. Recently, Liu et al. [100] demonstrated the mitochondrial ATP-sensitive potassium channel most closely resembles the Kir 6.1/SUR1 subtype. Thus, HMR 1883 can inhibit cardiac membrane ATP-sensitive potassium channels with miminal effects on mitochondrial channels.

HMR 1883 also prevented ischemically induced changes in the S-T segment in anesthetized mice [64], anesthetized swine [114], or conscious dogs [115]. A similar response was also noted for glibenclamide but not for 5-hydroxydecanoic acid [116]. In the conscious dogs with healed myocardial infarctions both HMR 1883 and glibenclamide prevented ischemically induced reductions in effective refractory period [61]. Furthermore, HMR 1883 significantly reduced monophasic action potential shortening induced by coronary artery occlusion in anesthetized pigs [117] and prevented an ischemically induced dispersion of ventricular repolarization in isolated Langendorff perfused rabbit hearts [118].

As would be predicted based upon these electrophysiological actions, HMR 1883/1098 prevented ischemically induced arrhtyhmias in a number of models (Table 4). For example, HMR 1883 reduced cardiac mortality in anesthetized pigs [113, 119] and prevented ventricular fibrillation induced by myocardial ischemia and reperfusion in rats [120]. HMR 1098 also significantly reduced arrhythmias induced by programmed electrical stimulation in dogs with healed myocardial infarctions without altering blood glucose levels [121]. In a similar manner, both glibenclamide and HMR 1883 [61], but not 5-hydoxydecanoic acid [116], significantly reduced the incidence of ventricular fibrillation induced by myocardial ischemia in conscious dogs with healed anterior wall myocardial infarctions. HMR 1098, but not 5-hydoxydecanoic acid, also prevented ventricular fibrillation induced by the ATP-sensitive

**TABLE 4 Effect of Cardioselective ATP-Sensitive Potassium Channel Antagonists on Ischemic Cardiac Arrhythmias**[a]

| Model | Result | Reference |
|---|---|---|
| Anesthetized pig | HMR 1883, ↓ VF | Bohn et al. [119] |
| Conscious dog | HMR 1883, ↓ VF | Billman et al. [61] |
| Anesthetized pig | HMR 1883, ↓ VF | Wirth et al. [114] |
| Anesthetized rat | HMR 1883, ↓ VF | Wirth et al. [120] |
| Anesthetized dog | HMR 1098, ↓ arrhythmias induced by programmed electrical stimulation dogs with healed myocardial infarctions | Zhu et al. [121] |
| Rabbit heart | HMR 1098 ↓ VF by programmed electrical stimulation | Behrens et al. [118] |
| Rabbit heart | HMR 1098, ↓ VF induced by pinadicil | Fischbach et al. [122] |
| Conscious dog | HMR 1402, ↓ VF | Billman et al. [123] |

[a] VF, ventricular fibrillation.

potassium channel agonist pinacidil in isolated rabbit hearts subjected to hypoxic perfusion [122]. These data strongly suggest that the selective opening of sarcolemmal ATP-sensitive potassium channels during ischemia provokes ventricular fibrillation while activation of the mitochodrial channels does not. Since, as noted above, ischemic preconditioning may result at least in part from the activation of mitochondrial channels, it may be possible to develop drugs that selectively activate these channels. However, nonspecific ATP-sensitive potassium channels should be avoided due the enhanced risk for malignant arrhythmias that would result from the activation of the sarcolemmal channels.

In contrast to glibenclamide, HMR 1883 did not alter plasma insulin or blood glucose levels in these animals [61]. Furthermore, glibenclamide but not HMR 1883 significantly reduced exercise-induced increase in mean cornary blood flow and provoked large reductions in left ventricular $dP/dt$ maximum (both at rest and during exercise) [61]. Thus, HMR 1883 appears to act selectively on the cardiac cell membrane ATP-sensitive potassium channel and thereby prevents ischemically induced arrhythmias with minimal effects on either pancreatic or smooth muscle channels.

***HMR 1402, a Cardioselective ATP-Sensitive Potassium Antagonists*** As noted above, at least six different $K_{ATP}$ channels are possible. Recent evidence suggests that the Kir 6.2/SUR2A combination is restricted to cardiac muscle [63, 98]. Thus, substances that preferentially inhibit this channel should display selectivity for cardiac tissue. Recently, a second cardioselective ATP-sensitive

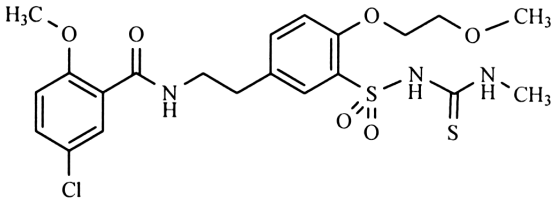


HMR 1402

**Figure 3** Chemical structure of HMR 1402.

potassium channel antagonist, HMR 1402 (Fig. 3), 1-[[5-[2(5-chloro-*o*-anisamido)ethyl]-β-methoxyethoxyphenyl]sulfonyl]-3-methylthioura has been developed [99]. This compound is structural similar to HMR 1883/1098. As was noted for HMR 1883/1098, HMR 1402 had no significant effects on action potential duration ($APD_{90}$), the resting potential, the amplitude of the phase 1 of the action potential, or on the upstroke velocity [123]. These data suggest that HMR 1402 did not affect potassium channels (i.e., $I_{K1}$, $I_K$, and $I_{Ks}$) or sodium channel at rest. This conclusion was directly confirmed by patch-clamp experiments performed in either rat or guinea pig ventricular myocytes [123]. In contrast, HMR 1402 potently blocked the rilmakalim-activated $K_{ATP}$ channels in both guinea pig papillary muscles and rat ventricular myocytes [123]. At an external pH of 6.0, this inhibition was approximately 6.1 times more potent than that reported for HMR 1883 ($IC_{50}$ for HMR 1402: 98 nM [123] versus $IC_{50}$ for HMR 1883: 0.6 μM [104]). Similarly, hypoxia consistently elicited a marked reduction in $APD_{90}$ that was potently antagonized by HMR 1402 [106, 123]. This inhibition of the hypoxia-induced shortening of the action potential duration was more potent for HMR 1402 than has been previously reported for HMR 1883 [104]. For example, 0.5 μM of HMR 1883 had no significant effect [104], while 0.3 μM of HMR 1402 produced a significant inhibition of reductions in $APD_{90}$ induced by hypoxia [123]. Thus, one may conclude that HMR 1402 is more potent in blocking rilmakalim-activated and hypoxia-activated $K_{ATP}$ channels than HMR 1883.

In contrast, HMR 1402 exhibited slightly more potent inhibition of vascular smooth and pancreatic β-cells than had been previously reported for HMR 1883/1098. For example, HMR 1402, in contrast to HMR 1883/1098, reduced hypoxia-induced increases in coronary flow at low concentrations [104, 123]. However, the same dose of glibenclamide (10 μM) provoked much larger reductions in coronary flow under normoxic conditions as well as hypoxic conditions [104]. As such, in isolated guinea pig hearts, HMR 1402 was more potent in inhibiting hypoxia-induced vasodilation than HMR 1883 but was still much less potent than glibenclamide. In a similar manner, HMR 1402 only partially inhibited the effect of diazoxide on the pancreatic β-cell membrane potential ($IC_{50} = 3.9$ μM [123]). This inhibition was somewhat more potent than that of HMR 1883 ($IC_{50}$ approximately 20 μM) [98, 104] but considerably less

than that of glibenclamide ($IC_{50}$ = 9.3 nM) [104]. HMR 1402, like HMR 1883/1098, did not alter either plasma insulin or blood glucose levels in conscious dogs [123]. This was in marked contrast to the pronounced hypoglycemia and the increase in plasma insulin provoked by glibenclamide [61, 123] These in vitro and in vivo data strongly suggest that HMR 1402 acts preferentially on cardiac $K_{ATP}$ channels but with slightly less selectively than HMR 1883/1098.

As one might predict, HMR 1402 was found to reduce the incidence of ventricular fibrillation induced by myocardial ischemia, protecting seven of eight animals tested [123]. This protection was very similar to that noted for both HMR 1883 and glibenclamide [61, 88]. However, it is important to emphasize that, in contrast to the actions of either HMR 1402 or HMR 1883, glibenclamide significantly reduced both exercise and reactive hyperemia-induced increases in coronary blood flow, as well as depressed ventricular function (large reductions in left ventricular $dP/dt$ maximum) in animals [61, 88]. Therefore, nonselective $K_{ATP}$ channel antagonist may protect against ischemic arrhythmias but not without potentially significant adverse side effects.

## Proarrhythmic Effects of ATP-Sensitive Potassium Channel Agonists

In contrast to ATP-sensitive potassium channel antagonists, channel agonists induced arrhythmias during ischemia in both isolated hearts and intact animals [83, 125, 127]. The ATP-sensitive potassium channel agonist pinacidil facilitated ventricular fibrillation during myocardial ischemia in isolated rat [83], guinea pig [127] or rabbit [125] hearts with reduced potassium levels. Di Diego and Antzelevitch [73] found that the activation of the ATP-sensitive potassium channel with pinacidil caused marked inhomogeneities of refractory period, which provoked extrasystoles. These extrasystoles were blocked by glibenclamide in strips of isolated canine myocardium. They concluded that this dispersion of repolarization and refractory period formed a substrate for reentry. Indeed, the ATP-sensitive potassium channel opener chromakalim reduced effective refractory period and increased vulnerability for reentrant arrhythmias [67]. Furthermore, this drug also increased interventricular dispersion of refractory period and induced ventricular fibrillation in 5 of 12 isolated rabbit hearts under normoxic conditions [68]. Chi et al. [126] further demonstrated that pinacidil increased the frequency of ventricular fibrillation during myocardial ischemia in unanesthetized dogs. This drug had no effect on arrhythmias induced by electrical stimulation in normally perfused tissue [126]. More recently, Coromilas et al. [74] demonstrated that the pinacidil restored excitability in the tissue adjacent to the ischemic region (i.e., the epicardial border zone). This reactivation of formerly inexcitable tissue led to the formation of reentrant circuits and the induction of ventricular tachycardia.

Although proarrhythmic events have not been associated with ATP-sensitive potassium channel agonists in the clinical setting [128, 129], these

drugs can alter ECG parameters that have been linked to arrhythmia formation. Indeed, abnormal T-wave changes suggestive of alteration in cardiac repolarization were reported in 30 percent of the hypertensive patients taking pinacidil [59]. In the presence of ischemia, these cardiac electrophysiological actions of nonselective ATP-sensitive potassium channel agonists could become more pronounced, increasing the propensity for life-threatening arrhythmias. However, nicorandil reduced the number of episodes of nonsustained ventricular tachycardia in patients with unstable angina [130]. This drug also reduced the episodes of transient ischemia. Thus, it is likely that the antiarrhythmic effects of this drug probably resulted secondarily from reductions in ischemia rather than a direct action on the ventricular myocardium. Remme and Wilde [131] concluded that since clinical studies of ATP-sensitive potassium channel agonists have not reported any major adverse effects on cardiac rhythm, the proarrhythmic potential of these agents may be "overestimated."

In contrast to these nonselective agonists, mitochondrial antagonists may lack proarrhythmic effects. For example, the novel mitochondrial agonist, BMS-191095 [132], reduced myocardial infarction size and preserved mechanical function during ischemia/reperfusion without adversely affecting cardiac electrophysiological parameters or the induction of arrhythmias [133].

When considered together, these reports indicate that the activation of sarcolemmal ATP-sensitive potassium channels promote arrhythmias, particularly in the setting of acute myocardial ischemia. In contrast, the selective activation of mitochondrial ATP-sensitive potassium channels contribute to ischemic preconditioning and thereby reduce the mechanical dysfunction induced by ischemia without affecting cardiac electrical stability. Therefore, the "ideal" modulator of ATP-sensitive potassium channels would both inhibit sarcolemmal channels and increase the activation (open) of mitochondrial channels in cardiac tissue.

## 5 SUMMARY

The activation of cardiac cell membrane ATP-sensitive potassium channels during myocardial ischemia promotes potassium efflux, reductions in action potential duration, and inhomogeneities in repolarization, thereby creating a substrate for reentrant arrhythmias. Drugs that block this channel should be particularly effective antiarrhythmic agents. Indeed, it is interesting to note that many currently available antiarrhythmic drugs, including verapamil, mibefradil, quinidine, lidocaine, and amiodarone, have also been reported to inhibit ATP-sensitive potassium channels at therapeutic concentrations [134–137]. Of particular note, amiodarone selectively inhibited sarcolemmal ATP-sensitive potassium channels without affecting mitochondrial channels [138]. Therefore, the inhibition of the ATP-sensitive potassium channel may be required for antiarrhythmic actions during myocardial ischemia. HMR

1883 (or its sodium salt HMR 1098) or HMR 1402 selectively blocks cardiac sarcolemmal ATP-sensitive potassium channels. As such, these drugs attenuate ischemically induced changes in cardiac electrical properties, thereby preventing malignant arrhythmias without the untoward effects of other drugs. Since, as noted above, the ATP-sensitive potassium channel only becomes active as ATP levels fall, these drugs have the added advantage that they would have effects *only* on ischemic tissue with little or no effect noted on normal tissue. Thus, selective antagonists of the cardiac cell surface ATP-sensitive potassium channel may represent the first truly ischemia-selective antiarrhythmic medications and should be free of the proarrhythmic actions that have plagued many of the currently available antiarrhythmic drugs.

## REFERENCES


1. Torp-Pedersen, C., Kober, L., Elming, H., Burchart, H. (1997). Classification of sudden and arrhythmic death. *PACE*, *20*, 2545–2552.
2. Abildstrom, S. Z., Kobler, L., Torp-Pedersen, C. (1999). Epidemiology of arrhythmic and sudden death in the chronic phase of ischemic heart disease. *Cardiac Electrophysiol. Rev.*, *3*, 177–179.
3. Gillum, R. (1989). Sudden coronary death in the United States 1980–1985. *Circulation*, *79*, 756–765.
4. Zipes, D. P., Wellens, H. J. (1998). Sudden cardiac death. *Circulation*, *98*, 2334–2351.
5. Zheng, Z-J., Croft, J. B., Giles, W. H., Mensah, G. A. (2001). Sudden cardiac death in the United States, 1989 to 1998. *Circulation*, *104*, 2158–2163.
6. Bayes de Luna, A., Coumel, P., LeClercq, J. F. (1989). Ambulatory sudden cardiac death: Mechanisms of production of fatal arrhythmia on the basis of data from 157 cases. *Am. Heart. J.*, *117*, 151–159.
7. Hinkle, L. E. J., Thaler, H. T. (1982). Clinical classification of cardiac deaths. *Circulation*, *65*, 457–464.
8. Greene, H. L. (1990). Sudden arrhythmic cardiac death: Mechanisms, resuscitation and classification: The Seattle perspective. *Am. J. Cardiol.*, *65*, 4B–12B.
9. Echt, D. S., Liebson, P. R., Mitchell, L. B., Peters, R. W., Obiasmanno, D., Barker, A. H., Arensberg, D., Baker, A., Freedman, L., Greene, H. L., Hunter, M. L., Richardson, D. W. (1991). Mortality and morbidity in patients receiving encainide, flecainide, or placebo. *N. Engl. J. Med.*, *324*, 782–788.
10. Sager, P. T. (1999). New advances in class III antiarrhythmic drug therapy. *Curr. Opin. Cardiol.*, *14*, 15–23.
11. Waldo, A. L., Camm, A. J., de Ruyter, H., Friedman, P. L., MacNeil, D. J., Pauls, J. F., Pitt, B., Pratt, C. M., Schwartz, P. J., Veltri, E. P., for the SWORD Investigators. (1996). Effect of *d*-sotalol on mortality in patients with left ventricular dysfunction after recent and remote myocardial infarction. *Lancet*, *348*, 7–12.
12. Amiodarone Trials Meta-Analysis Investigators. (1997). Effects of prophylactic amiodarone on mortality after acute myocardial infarction and in congestive heart

failure: Meta-analysis of individual data from 6500 patients in randomized trials. *Lancet*, *350*, 1417–1424.

13. Held, P., Yusuf, S. (1989). Early intravenous beta-blockade in acute myocardial infarction. *Cardiology*, *76*, 132–143.
14. Held, P. H., Yusuf, S. (1993). Effects of beta-blockers and $Ca^{2+}$ channel blockers in acute myocardial infarction. *Eur. Heart. J.*, *14*(Suppl. F), 18–25.
15. Kendall, M. J., Lynch, K. P., Hjalmarson, A., Kjekshus, J. (1995). β-Blockers and sudden cardiac death. *Ann. Intern. Med.*, *123*, 353–367.
16. Buxton, A. E., Lee, K. L., Fisher, J. D., Josephson, M. E., Prystowsky, E. N., Hafley, G. (1999). A randomized study of the prevention of sudden death in patients with coronary artery disease. *N. Engl. J. Med.*, *341*, 1882–1890.
17. Opie, L. H., Nathan, D., Lubbe, W. F. (1979). Biochemical aspects of arrhythmogenesis. *Am. J. Cardiol.*, *43*, 131–148.
18. Curtis, M. J., Pugsley, M. K., Walker, M. J. A. (1993). Endogenous chemical mediators of ventricular arrythmias in ischaemic heart disease. *Cardiovasc. Res.*, *27*, 703–719.
19. Billman, G. E. (1991). The antiarrhythmic and antifibrillatory effects of calcium antagonists. *J. Cardiovasc. Pharmacol.*, *18*(Suppl. 10), S107–S117.
20. Billman, G. E. (1994). The role of the ATP-sensitive $K^+$ channel in $K^+$ accumulation and cardiac arrhythmias during myocardial ischemia. *Cardiovasc. Res.*, *28*, 762–769.
21. Wilde, A. A. M. (1993). Role of ATP-sensitive $K^+$ channel current in ischemic arrhythmias. *Cardiovasc. Drugs Ther.*, *7*, 521–526.
22. Wilde, A. M. M., Aksnes, G. (1995). Myocardial potassium loss and cell depolarization in ischaemia and hypoxia. *Cardiovasc. Res.*, *29*, 1–15.
23. Harris, A. S. (1966). Potassium and experimental coronary occlusion. *Am. Heart. J.*, *71*, 797–802.
24. Harris, A. S., Bisteni, A., Russell, R. A., Brigham, J. C., Firestone, J. E. (1954). Excitory factors in ventricular tachycardia resulting from myocardial ischemia: Potassium a major excitant. *Science*, *119*, 200–203.
25. Coronel, R., Fiolet, JW., Wilms-Schopman, F. J., Schaapherder, A. F., Johnson, T. A., Gettes, L. S., Janse, M. J. (1988). Distribution of extracellular potassium and its relation to electrophysiologic changes during acute myocardial ischemia in the isolated perfused porcine heart. *Circulation*, *77*, 1125–1138.
26. Hill, J. L., Gettes, L. S. (1980). Effect of acute coronary artery occlusion on local myocardial extracellular $K^+$ activity in swine. *Circulation*, *61*, 768–778.
27. Kleber, A. G. (1984). Extracellular potassium accumulation in acute myocardial ischemia. *J. Mol. Cell. Cardiol.*, *16*, 389–394.
28. Johnson, T. A., Engle, C. L., Boyd, L. M., Koch, G. G., Gwinn, M., Gettes, L. S. (1991). Magnitude and time course of extracellular potassium inhomogeneities during acute ischemia in pigs. Effect of verapamil. *Circulation*, *83*, 622–634.
29. Coronel, R., Wilms-Schopman, F. J., Opthof, T., van Capelle, F. J., Janse, M. J. (1991). Injury current and gradients of diastolic stimulation threshold, TQ potential, and extracellular potassium concentration during acute regional ischemia in the isolated perfused pig heart. *Circ. Res.*, *68*, 1241–1249.

30. Kleber, A. G. (1983). Resting membrane potential, extracellular potassium activity, and intracellular sodium activity during acute global ischemia in isolated perfused guinea pig hearts. *Circ. Res.*, *52*, 442–450.
31. Rau, E. E., Shine, K. I., Langer, G. A. (1977). Potassium exchange and mechanical performance in anoxic mammalian myocardium. *Am. J. Physiol. Heart Circ. Physiol.*, *232*, H85–H94.
32. Shine, K. I., Douglas, A. M., Ricchiuti, N. (1977). $^{42}$K exchange during myocardial ischemia. *Am. J. Physiol. Heart Circ. Physiol.*, *232*, H564–H570.
33. Wilde, A. M. M. (1998). ATP and the role of $I_{KATP}$ during acute myocardial ischaemia: Controversies revive. *Cardiovasc. Res.*, *35*, 181–183.
34. Trube, G., Hescheler, J. (1984). Inward-rectifying channels in isolated patches of the heart cell membrane: ATP-dependence and comparison with cell-attached patches. *Pfugers Arch.*, *401*, 178–184.
35. Noma, A. (1983). ATP-regulated $K^+$ channels in cardiac muscle. *Nature*, *305*, 147–148.
36. Bekheit, S. S., Restivo, M., Boutjdir, M., Henkin, R., Gooyandeh, K., Assadi, M., Khatib, S., Gough, W. B., El-Sherif, N. (1990). Effects of glyburide on ischemia-induced changes in extracellular potassium and local myocardial activation: A potential new approach to the management of ischemia-induced malignant ventricular arrhythmias. *Am. Heart. J.*, *119*, 1025–1033.
37. Benndorf, K., Friedrich, M., Hirche, H. (1991). Anoxia opens ATP regulated K channels in isolated heart cells of the guinea pig. *Pflugers Arch.*, *419*, 108–110.
38. Fosset, M., De Weille, J. R., Green, R. D., Schmid-Antomarchi, H., Lazdunski, M. (1988). Antidiabetic sulfonylureas control action potential properties in heart cells via high affinity receptors that are linked to ATP-dependent $K^+$ channels. *J. Biol. Chem.*, *263*, 7933–7936.
39. Hicks, M. N., Cobbe, S. M. (1991). Effect of glibenclamide on extracellular potassium accumulation and the electrophysiological changes during myocardial ischaemia in the arterially perfused interventricular septum of rabbit. *Cardiovasc. Res.*, *25*, 407–413.
40. Kantor, P. F., Coetzee, W. A., Carmeliet, E. E., Dennis, S. C., Opie, L. H. (1990). Reduction of ischemic $K^+$ loss and arrhythmias in rat hearts: Effect of glibenclamide, a sulfonylurea. *Circ. Res.*, *66*, 478–485.
41. Nakaya, H., Takeda, Y., Tohse, N., Kanno, M. (1991). Effects of ATP-sensitive $K^+$ channel blockers on the action potential shortening in hypoxic and ischaemic myocardium. *Br. J. Pharmacol.*, *103*, 1019–1026.
42. Ruiz-Petrich, E., Leblanc, N., deLorenzi, F., Allard, Y., Schanne, O. F. (1992). Effects of $K^+$ channel blockers on the action potential of hypoxic rabbit myocardium. *Br. J. Pharmacol.*, *106*, 924–930.
43. Vanheel, B., De Hemptinne, A. (1992). Influence of KATP channel modulation on net potassium efflux from ischaemic mammalian cardiac tissue. *Cardiovasc. Res.*, *26*, 1030–1039.
44. Venkatesh, N., Lamp, S. T., Weiss, J. N. (1991). Sulfonylureas, ATP-sensitive $K^+$ channels, and cellular $K^+$ loss during hypoxia, ischemia, and metabolic inhibition in mammalian ventricle. *Circ. Res.*, *69*, 623–637.

45. Weiss, J. N., Venkatesh, N., Lamp, S. T. (1992). ATP-sensitive $K^+$ channels and cellular $K^+$ loss in hypoxic and ischaemic mammalian ventricle. *J. Physiol. (Lond.)*, *447*, 649–673.
46. Wilde, A. A., Escande, D., Schumacher, C. A., Thuringer, D., Mestre, M., Fiolet, J. W., Janse, M. J. (1990). Potassium accumulation in the globally ischemic mammalian heart. A role for the ATP-sensitive potassium channel. *Circ. Res.*, *67*, 835–843.
47. Smallwood, J. K., Ertel, P. J., Steinberg, M. I. (1990). Modification by glibenclamide of the electrophysiological consequences of myocardial ischaemia in dogs and rabbits. *Naunyn Schmiedebergs Arch. Pharmacol.*, *342*, 214–220.
48. Edwards, G., Weston, A. H. (1993). The pharmacology of ATP-sensitive potassium channels. *Annu. Rev. Pharmacol. Toxicol.*, *33*, 597–637.
49. Mitani, A., Kinoshita, K., Fukamachi, K., Sakamoto, M., Kurisu, K., Tsuruhara, Y., Fukumura, F., Nakashima, A., Tokunaga, K. (1991). Effects of glibenclamide and nicorandil on cardiac function during ischemia and reperfusion in isolated perfused rat hearts. *Am. J. Physiol. Heart Circ. Physiol.*, *261*, H1864–H1871.
50. Venkatesh, N., Stuart, J. S., Lamp, S. T., Alexander, L. D., Weiss, J. N. (1992). Activation of ATP-sensitive $K^+$ channels by cromakalim: Effects on cellular $K^+$ loss and cardiac function in ischemic and reperfused mammalian ventricle. *Circ. Res.*, *71*, 1324–1333.
51. Findlay, I. (1988). ATP4- and ATP.Mg inhibit the ATP-sensitive K+ channel of rat ventricular myocytes. *Pflugers Arch.*, *412*, 37–41.
52. Nichols, C. G., Lederer, W. J. (1990). The regulation of ATP-sensitive $K^+$ channel activity in intact and permeabilized rat ventricular myocytes. *J. Physiol. (Lond.)*, *423*, 91–110.
53. Rovetto, M. J., Whitmer, J. T., Neely, J. R. (1973). Comparison of the effects of anoxia and whole heart ischemia on carbohydrate utilization in isolated working rat hearts. *Circ. Res.*, *32*, 699–711.
54. Faivre, J. F., Findlay, I. (1990). Action potential duration and activation of ATP-sensitive potassium current in isolated guinea-pig ventricular myocytes. *Biochim. Biophys. Acta*, *1029*, 167–172.
55. Nichols, C. G., Lederer, W. J. (1991). Adenosine triphosphate-sensitive potassium channels in the cardiovascular system. *Am. J. Physiol. Heart Circ. Physiol.*, *261*, H1675–H1686.
56. Nichols, C. G., Ripoll, C., Lederer, W. J. (1991). ATP-sensitive potassium channel modulation of the guinea pig ventricular action potential and contraction. *Circ. Res.*, *68*, 280–287.
57. Deutsch, N., Klitzner, T. S., Lamp, S. T., Weiss, J. N. (1991). Activation of cardiac ATP-sensitive $K^+$ current during hypoxia: Correlation with tissue ATP levels. *Am. J. Physiol. Heart Circ. Physiol.*, *261*, H671–H676.
58. Kubota, I., Yamaki, M., Shibata, T., Ikeno, E., Hosoya, Y., Tomoike, H. (1993). Role of ATP-sensitive $K^+$ channel on ECG ST segment elevation during a bout of myocardial ischemia. A study on epicardial mapping in dogs. *Circulation*, *88*, 1845–1851.
59. Goldberg, M. R. (1988). Clinical pharmacology of pinacidil, a prototype for drugs that affect potassium channels. *J. Cardiovasc. Pharmacol.*, *12*(Suppl. 2), S41–S47.

60. Kondo, T., Kubota, I.,Tachibana, H., Yamaki, M., Tomoike, H. (1996). Glibenclamide attenuates peaked T wave in early phase of myocardial ischemia. *Cardiovasc. Res.*, *31*, 683–687.
61. Billman, G. E., Englert, H. C., Schoelkens, B. A. (1998). HMR 1883, a novel cardioselective inhibitor of the ATP-sensitive potassium channel; Part II: Effects on susceptibility to ventricular fibrillation induced by myocardial ischemia in conscious dogs. *J. Pharmacol. Exp. Ther.*, *286*, 1465–1473.
62. Yan, G. X., Antzelevitch, C. (1998). Cellular basis for the normal T wave and the electrocardiographic manifestations of the long-QT syndrome. *Circulation*, *98*, 1928–1936.
63. Suzuki, M., Li, R. A., Miki, T., Uemura, H., Sakamoto, N., Ohmoto-Sekine, Y., Tamagawa, M., Ogura, T., Seino, S., Marban, E., Nakaya, H. (2001). Functional roles of cardiac and vascular ATP-sensitive potassium channels clarified by Kir6.2-knockout mice. *Circ. Res.*, *88*, 570–577.
64. Li, R. A., Leppo, M., Miki, T., Seino, S., Marban, E. (2000). Molecular basis of electrocardiographic ST-segment elevation. *Circ. Res.*, *87*, 837–839.
65. Wit, A. L., Janse, M. J. (1989). Electrophysiological mechanisms of ventricular arrhythmias resulting from myocardial ischemia and infarction. *Physiol. Rev.*, *69*, 1049–1169.
66. Krause, E., Englert, H., Gögelein, H. (1995). Adenosine triphosphate-dependent $K^+$ currents activated by metabolic inhibition in rat ventricular myocytes differ from those elicited by the channel opener rilmakalim. *Pflugers Arch.*, *429*, 625–635.
67. Uchida, T., Yashima, M., Gotoh, M., Qu, Z., Garfinkel, A., Weiss, J. N., Fishbein, M. C., Mandel, W. J., Chen, P. S., Karagueuzian, H. S. (1999). Mechanism of acceleration of functional reentry in the ventricle: Effects of ATP-sensitive potassium channel opener. *Circulation*, *99*, 704–712.
68. Wolk, R., Cobbe, S. M., Kane, K. A., Hicks, M. N. (1999). Relevance of inter- and intraventricular electrical dispersion to arrhythmogenesis in normal and ischaemic rabbit myocardium: A study with cromakalim, 5-hydroxydecanoate and glibenclamide. *J. Cardiovasc. Pharmacol.*, *33*, 323–334.
69. Furukawa, T., Kimura, S., Furukawa, N., Bassett, A. L., Myerburg, R. J. (1992). Potassium rectifier currents differ in myocytes of endocardial and epicardial origin. *Circ. Res.*, *70*, 91–103.
70. Gilmour, R. F. Jr, Zipes, D. P. (1980). Different electrophysiological responses of canine endocardium and epicardium to combined hyperkalemia, hypoxia, and acidosis. *Circ. Res.*, *46*, 814–825.
71. Kimura, S., Bassett, A. L., Kohya, T., Kozlovskis, P. L., Myerburg, R. J. (1986). Simultaneous recording of action potentials from endocardium and epicardium during ischemia in the isolated cat ventricle: Relation of temporal electrophysiologic heterogeneities to arrhythmias. *Circulation*, *74*, 401–409.
72. Tweedie, D., Henderson, C., Kane, K. (1993). Glibenclamide, but not class III drugs, prevents ischaemic shortening of the refractory period in guinea-pig hearts. *Eur. J. Pharmacol.*, *240*, 251–257.
73. Di Diego, J. M., Antzelevitch, C. (1993). Pinacidil-induced electrical heterogeneity and extrasystolic activity in canine ventricular tissues. Does activation of

ATP-regulated potassium current promote phase 2 reentry? *Circulation*, *88*, 1177–1189.

74. Coromilas, J., Costeas, C., Deruyter, B., Dillon, S. M., Peters, N. S., Wit, A. L. (2002). Effects of pinacidil on electrophysiological properties of epicardial border zone of healing canine infarcts: Possible effects of $K_{ATP}$ channel activation. *Circulation*, *105*, 2309–2317.
75. Bellemin-Baurreau, J., Poizot, A., Hicks, P. E., Rochette, L., Armstrong, J. M. (1994). Effects of ATP-dependent $K^+$ channel modulators on an ischemia-reperfusion rabbit isolated heart model with programmed electrical stimulation. *Eur. J. Pharmacol.*, *256*, 115–124.
76. Dhein, S., Pejman, P., Krusemann, K. (2000). Effects of the $I_{K.ATP}$ blockers glibenclamide and HMR1883 on cardiac electrophysiology during ischemia and reperfusion. *Eur. J. Pharmacol.*, *398*, 273–284.
77. Hamada, K., Yamazaki, J., Nagao, T. (1998). Shortening of monophasic action potential duration during hyperkalemia and myocardial ischemia in anesthetized dogs. *Jpn. J. Pharmacol.*, *76*, 149–154.
78. Koumi, S. I., Martin, R. L., Sato, R. (1997). Alterations in ATP-sensitive potassium channel sensitivity to ATP in failing human hearts. *Am. J. Physiol. Heart Circ. Physiol.*, *272*, H1656–H1665.
79. MacKenzie, I., Saville, V. L., Waterfall, J. F. (1993). Differential class III and glibenclamide effects on action potential duration in guinea-pig papillary muscle during normoxia and hypoxia/ischaemia. *Br. J. Pharmacol.*, *110*, 531–538.
80. Gwilt, M., Henderson, C. G., Orme, J., Rourke, J. D. (1992). Effects of drugs on ventricular fibrillation and ischaemic $K^+$ loss in a model of ischaemia in perfused guinea-pig hearts in vitro. *Eur. J. Pharmacol.*, *220*, 231–236.
81. Pogatsa, G., Koltai, M. Z., Balkanyi, I., Devai, I., Kiss, V., Koszeghy, A. (1988). The effect of various hypoglycaemic sulphonylureas on the cardiotoxicity of glycosides and arrhythmogenic activity due to myocardial ischaemia. *Acta Physiol. Hung.*, *71*, 243–250.
82. Tosaki, A., Engelman, D. T., Engelman, R. M., Das, D. K. (1995). Diabetes and ATP-sensitive potassium channel openers and blockers in isolated ischemic/reperfused hearts. *J. Pharmacol. Exp. Ther.*, *275*, 1115–1123.
83. Wolleben, C. D., Sanguinetti, M. C., Siegl, P. K. (1989). Influence of ATP-sensitive potassium channel modulators on ischemia-induced fibrillation in isolated rat hearts. *J. Mol. Cell. Cardiol.*, *21*, 783–788.
84. El Reyani, N. E., Bozdogan, O., Baczko, I., Lepran, I., Papp, J. G. (1999). Comparison of the efficacy of glibenclamide and glimepiride in reperfusion-induced arrhythmias in rats. *Eur. J. Pharmacol.*, *365*, 187–192.
85. Lepran, I., Baczko, I., Varro, A., Papp, J. G. (1996). ATP-sensitive potassium channel modulators: Both pinacidil and glibenclamide produce antiarrhythmic activity during acute myocardial infarction in conscious rats. *J. Pharmacol. Exp. Ther.*, *277*, 1215–1220.
86. Barrett, T. D., Walker, M. J. (1998). Glibenclamide does not prevent action potential shortening induced by ischemia in anesthetized rabbits but reduces ischemia-induced arrhythmias. *J. Mol. Cell. Cardiol.*, *30*, 999–1008.
87. Baczko, I., Lepran, I., Papp, J. G. (1997). KATP channel modulators increase survival rate during coronary occlusion-reperfusion in anaesthetized rats. *Eur. J. Pharmacol.*, *324*, 77–83.

88. Billman, G. E., Avendano, C. E., Halliwill, J. R., Burroughs, J. M. (1993). The effects of the ATP-dependent potassium channel antagonist glyburide on coronary blood flow and susceptibility to ventricular fibrillation. *J. Cardiovasc. Pharmacol.*, *21*, 197–204.

89. Cacciapuoti, F., Spiezia, R., Bianchi, M., Lama, D., D'Avino, M., Varricchio, M. (1991). Effectiveness of glibenclamide on myocardial ischemic ventricular arrhythmias in non-insulin dependent diabetes mellitus. *Am. J. Cardiol.*, *67*, 843–847.

90. Davis, T. M., Parsons, R. W., Broadhurst, R. J., Hobbs, M. S., Jamrozik, K. (1998). Arrhythmias and mortality after myocardial infarction in diabetic patients. Relationship to diabetes treatment. *Diabetes Care*, *21*, 637–640.

91. Lomuscio, A., Fiorentini, C. (1996). Influence of oral antidiabetic treatment on electrocardiac alterations induced by myocardial infarction. *Diabetes Res. Clin. Pract.*, *31*(Suppl), S21–S26.

92. Lomuscio, A., Vergani, D., Marano, L., Castagnone, M., Fiorentini, C. (1994). Effects of glibenclamide on ventricular fibrillation in non-insulin-dependent diabetics with acute myocardial infarction. *Coron. Artery Dis.*, *5*, 767–771.

93. Aronson, D., Mittleman, M. A., Burger, A. J. (2003). Effects of sulfonylurea agents and adenosine triphosphate dependent potassium channel antagonists on ventricular arrhythmias in patients with decompensated heart failure. *PACE*, *26*, 1254–1261.

94. Link, M. S., Wang, P. J., VanderBrink, B. A., Avelar, E., Pandian, N. G., Maron, B. J., Estes, N. A. (1999). Selective activation of the $K^+_{ATP}$ channel is a mechanism by which sudden death is produced by low-energy chest-wall impact (Commotio cordis). *Circulation*, *100*, 413–418.

95. Gögelein, H. (2001). Inhibition of cardiac ATP-dependent potassium channels by sulfonylurea drugs. *Curr. Opin. Invest. Drugs.*, *2*, 71–80.

96. Gögelein, H., Hartung, J., Englert, H. C. (1999). Molecular basis, pharmacology and physiological role of cardiac $K_{ATP}$ channels. *Cell Physiol. Biochem.*, *9*, 227–241.

97. Inagaki, N., Gonoi, T., Clement, J. P., Namba, N., Inazawa, J., Gonzalez, G., Aguilar-Bryan, L., Seino, S., Bryan, J. (1995). Reconstitution of $I_{KATP}$: An inward rectifier subunit plus the sulfonylurea receptor. *Science*, *270*, 1166–1170.

98. Manning-Fox, J. E. M., Kanji, H. D., French, R. J., Light, P. E. (2002). Cardioselectivity of the sulphonylurea HMR 1098: Studies on native and recombinant cardiac and pancreatic K-ATP channels. *Br. J. Pharmacol.*, *135*, 480–488.

99. Englert, H. C., Heitsch, H., Gerlach, U., Knieps, S. (2003). Blockers of the ATP-sensitive potassium channel SURA/Kir6.2: A new approach to prevent sudden cardiac death. *Curr. Med. Chem.*, *1*, 253–271.

100. Liu, Y., Ren, G., O'Rourke, B., Marban, E., Seharaseyon, J. (2001). Pharmacological comparison of native mitochondrial $K_{ATP}$ channels with molecularly defined surface $K_{ATP}$ channels. *Mol. Pharmacol.*, *59*, 225–230.

101. Gribble, F. M., Tucker, S. J., Seino, S., Ashcroft, F. M. (1998). Tissue specificity of sulfonylureas: Studies on cloned cardiac and beta-cell $K_{ATP}$ channels. *Diabetes*, *47*, 1412–1418.

102. Reimann, F., Proks, P., Ashcroft, F. M. (2001). Effects of mitiglinide (S 21403) on Kir6.2/SUR1, Kir6.2/SUR2A and Kir6.2/SUR2B types of ATP-sensitive potassium channel. *Br. J. Pharmacol.*, *132*, 1542–1548.

103. Englert, H. C., Gerlach, U., Goegelein, H., Hartung, J., Heitsch, H., Mania, D., Scheidler, S. (2001). Cardioselective K (ATP) channel blockers derived from a new series of *m*-anisamidoethylbenzenesulfonylthioureas. *J. Med. Chem.*, *44*, 1085–1098.

104. Gögelein, H., Hartung, J., Englert, H. C., Scholkens, B. A. (1998). HMR 1883, a novel cardioselective inhibitor of the ATP-sensitive potassium channel. Part I: Effects on cardiomyocytes, coronary flow and pancreatic beta-cells. *J. Pharmacol. Exp. Ther.*, *286*, 1453–1464.

105. Kaab, S., Zwermann, L., Barth, A., Hinterseer, M, Englert, H. C., Gögelein, H. R., Nabauer, M. (2003). Selective block of sarcolemmal IKATP in human cardiomycytes using HMR 1098. *Cardiovasc. Drug. Ther.*, *17*, 435–441.

106. Weyermann, A., Vollert, H., Busch, A. E., Bleich, M., Gögelein, H. (2004). Inhibitors of ATP-sensitive potassium channels in guinea pig isolated ischemic hearts. *Naunyn Schmiedebergs Arch. Pharmacol.*, *369*, 374–381.

107. Sato, T. (1999). Signaling in late preconditioning: Involvement of mitochondrial $K_{ATP}$ channels. *Circ. Res.*, *85*, 1113–1114.

108. Liu, Y., Sato, T., O'Rourke, B., Marban, E. (1998). Mitochondrial ATP-dependent potassium channels: Novel effectors of cardioprotection? *Circulation*, *97*, 2463–2469.

109. Fryer, R. M., Eells, J. T., Hsu, A. K., Henry, M. M., Gross, G. J. (2000). Ischemic preconditioning in rats: Role of mitochondrial $K_{ATP}$ channel in preservation of mitochondrial function. *Am. J. Physiol. Heart. Circ. Physiol.*, *278*, H305–H312.

110. Jung, O., Englert, H. C., Jung, W., Gögelein, H., Scholkens, B. A., Busch, A. E., Linz, W. (2000). The $K_{ATP}$ channel blocker HMR 1883 does not abolish the benefit of ischemic preconditioning on myocardial infarct mass in anesthetized rabbits. *Naunyn Schmiedebergs Arch. Pharmacol.*, *361*, 445–451.

111. Gross, G. J., Fryer, R. M. (1999). Sarcolemmal versus mitochondrial ATP-sensitive $K^+$ channels and myocardial preconditioning. *Circ. Res.*, *84*, 973–979.

112. O'Rourke, B. (2000). Myocardial $K_{ATP}$ channels in preconditioning. *Circ. Res.*, *87*, 845–855.

113. Grover, G. J., Garlid, K. D. (2000). ATP-sensitive potassium channels: A review of their cardioprotective pharmacology. *J. Mol. Cell. Cardiol.*, *32*, 677–695.

114. Wirth, K. J., Rosenstein, B., Uhde, J., Englert, H. C., Busch, A. E., Scholkens, B. A. (1999). ATP-sensitive potassium channel blocker HMR 1883 reduces mortality and ischemia-associated electrocardiographic changes in pigs with coronary occlusion. *J. Pharmacol. Exp. Ther.*, *291*, 474–481.

115. Billman, G. E., Houle, M. S., Englert, H. C., Goegelein, H. (1999). Ischemically-induced changes in the T-wave and susceptibility to sudden death: Evidence that activation of the ATP-sensitive potassium channel may contribute to ventricular fibrillation. *Circulation*, *100*(Suppl. I), 1–52. 1999.

116. Billman, G. E., Englert, H. C., Goegelein, H., Busch, A. (2001). Selective sarcolemmal, but not selective mitochondrial, ATP-sensitive potassium channel antagonists prevent ventricular fibrillation induced by ischaemia. *Eur. Heart. J.*, *22*(Abstract Suppl.), 246.

117. Wirth, K. J., Uhde, J., Rosenstein, B., Englert, H. C., Gögelein, H., Scholkens, B. A., Busch, A. E. (2000). $K_{ATP}$ channel blocker HMR 1883 reduces monophasic

action potential shortening during coronary ischemia in anesthetised pigs. *Naunyn Schmiedebergs Arch. Pharmacol.*, *361*, 155–160.

118. Behrens, S., Zabel, M., Janssen, A., Barbierato, M., Schultheiss, H. P. (2001). Influence of a new ATP-sensitive potassium-channel antagonist (HMR 1098) on ventricular fibrillation inducibility during myocardial ischemia. *Eur. Heart. J.*, *22*(Abstract Suppl.), 546.
119. Bohn, H., Englert, H. C., Schoelkens, B. A. (1998). The $K_{ATP}$ channel blocker HMR 1883 attenuates the effects of ischemia on MAP duration and improves survival during LAD occlusion in anaesthetized pig. *Br. J. Pharmacol.*, *124*, 23P.
120. Wirth, K. J., Klaus, E., Englert, H. G., Scholkens, B. A., Linz, W. (1999). HMR 1883, a cardioselective K (ATP) channel blocker, inhibits ischaemia- and reperfusion-induced ventricular fibrillation in rats. *Naunyn Schmiedebergs Arch. Pharmacol.*, *360*, 295–300.
121. Zhu, B. M., Miyamoto, S., Nagawa, Y., Wajima, T., Hashimoto, K. (2003). Effect of sarcolemmal K-ATP blocker HMR 1098 on arrhythmias induced by programmed electrical stimulation in canine old myocardial infarction model: Comparison with glibenclamide. *J. Pharmacol. Sci.*, *93*, 106–113.
122. Fischbach, P. S., White, A., Barrett, T. D., Lucchesi, B. R. (2004). Risk of ventricular proarrhythmia with the selective opening of the myocardial sarcolemmal versus mitochondrial ATP-gated potassium channel. *J. Pharmacol. Exp. Ther.*, *309*, 554–559.
123. Billman, G. E., Houle, M. S., Englert, H. C., Gögelein, H. (2004). Effects of a novel cardioselective ATP-sensitive potassium channel antagonist 1-[[5-[2(5-chloro-*o*-anisamido)ethyl]-β-methoxyethoxyphenyl]sulfonyl]-3-methylthioura, sodium salt (HMR 1402), on susceptibility to ventricular fibrillation induced by myocardial ischemia: In vitro and in vivo studies. *J. Pharmacol. Exp. Ther.*, *309*, 182–192.
124. Nattel, S. (2000). Class III drugs: Amiodarone, bertylium, ibutilide, and sotalol. In D. P. Zipes, J. Jalife (Eds), *Cardiac Electrophysiology from Cell to Bedside*, 3rd ed. W. B. Saunders, Philadelphia, pp. 921–932.
125. Chi, L., Black, S. C., Kuo, P. I., Fagbemi, S. O., Lucchesi, B. R. (1993). Actions of pinacidil at a reduced potassium concentration—A direct cardiac effect possibly involving the ATP-sensitive potassium channel. *J. Cardiovasc. Pharmacol.*, *21*, 179–190.
126. Chi, L., Uprichard, A. C. G., Lucchesi, B. R. (1990). Profibrillatory actions of pinacidil in a conscious canine model of sudden coronary death. *J. Cardiovasc. Pharmacol.*, *15*, 452–464.
127. D'Alonzo, A. J., Zhu, J. L., Darbenzio, R. B., Dorso, C. R., Grover, G. J. (1998). Proarrhythmic effects of pinacidil are partially mediated through enhancement of catecholamine release in isloated perfused guinea-pig hearts. *J. Mol. Cell. Cardiol.*, *30*, 415–423.
128. Friedel, H. A., Brogden, R. N. (1990). Pinacidil—A review of its pharmacodynamic and pharmacokinetic properties and therapeutic potential in the treatment of hypertension. *Drugs*, *39*, 929–967.
129. Krumenacker, M., Roland, E. (1992). Clinical profile of nicorandil—An overview of its hemodynamic properties and therapeutic efficacy. *J. Cardiovasc. Pharmacol.*, *20*(Suppl. 3), S93–S102.

130. Patel, D. J., Purcell, H. J., Fox, K. M. (1999). Cardioprotection by opening of the $K_{ATP}$ channel in unstable angina. *Eur. Heart J.*, *20*, 51–57.

131. Remme, C. A., Wilde, A. M. M. (2000). $K_{ATP}$ channel opener, myocardial ischemia and arrhythmias—Should the electrophysiologist worry? *Cardiovasc. Drug Ther.*, *14*, 17–22.

132. Grover, G. J., D'Alonzo, A. J., Garlid, K. D., Bajgar, R., Lodge, N. J., Sleph, P. G., Darbenzio, R. B., Hess, T. A., Smith, M. A., Paucke, P., Atwal, K. S. (2001). Pharmacologic characterization of BMS-191095, a mitochondrial $K_{ATP}$ opener with no peripheral vasodilator or cardiac action potential shortening activity. *J. Phamacol. Exp. Ther.*, *297*, 1184–1192.

133. Grover, G. J., D'Alonzo, A. J., Darbenzio, R. B., Parhaqm, C. S., Hess, T. A., Bathala, M. S. (2002). In vivo characterization of the mitochondrial selective $K_{ATP}$ opener (3*R*)-*trans*-4-(4-chlorophenyl)-*N*-(1*H*-imidazol-2-ylmethyl)dimethyl-2H-1-benzopyran-6-carbonitril monohydrochloride (BMS-191095): Cardioprotective, hemodynamic, and electrophysiological effects. *J. Pharmacol. Exp. Ther.*, *303*, 132–140.

134. Colatsky, T. J., Follmer, C. H., Starmer, C. F. (1990). Channel specificity in antiarrhythmic drug action. Mechanism of potassium channel block and its role in suppressing and aggravating cardiac arrhythmias. *Circulation*, *82*, 2235–2242.

135. Haworth, R. A., Goknur, A. B., Berkoff, H. A. (1989). Inhibition of ATP-sensitive potassium channels of adult rat heart cells by antiarrhythmic drugs. *Circ. Res.*, *65*, 1157–1160.

136. Holmes, D. S., Sun, Z. Q., Porter, L. M., Bernstein, N. E., Chinitz, L. A., Artman, M., Coetzee, W. A. (2000). Amiodarone inhibits cardiac ATP-sensitive potassium channels. *J. Cardiovasc. Electrophysiol.*, *11*, 1152–1158.

137. Olschewski, A., Brau, M. E., Olschewski, H., Hempelmann,G., Vogel, W. (1996). ATP-dependent potassium channel in rat cardiomyocytes is blocked by lidocaine. Possible impact on the antiarrhythmic action of lidocaine. *Circulation*, *93*, 656–659.

138. Sato, T., Takizawa, T., Saito, T., Kobayashi, S., Hara Nakaya, H. (2003). Amiodarone inhibits sarcolemmal but not mitchondrial K-ATP channels in guinea pig ventricular cells. *J. Pharmacol. Exp. Ther.*, *307*, 955–960.

139. Gribble, F. M., Reimann, F. (2003). Sulphonylurea action revisited: The post-cloning era. *Diabetologia*, *46*, 875–891.

140. Zhang, H. L., Li, Y. S., Fu, S. X., Yang, X. P. (1991). Effects of glibenclamide and tolbutamine on ischemia-induced and ouabain-induced arrhythmias and membrane potentials of ventricular myocardium from rat and guinea-pig. *Acta Pharmacol. Sinica*, *12*, 398–402.

141. Bril, A., Laville, M-P, Gout, B. (1992). Effects of glibenclamide on ventricular arrhythmias and cardiac function in ischaemia and reperfusion in isolated rat heart. *Cardiovasc. Res.*, *26*, 1069–1076.

142. Pasnani, J. S., Ferrier, G. R. (1992). Differential effects of glyburide on premature beats and ventricular tachycardia in an isloated tissue model of ischemia and reperfusion. *J. Pharmacol. Exp. Ther.*, *262*, 1076–1084.

143. Rees, S. A., Curtis, M. J. (1995). Pharmacological analysis in rat of the role of the ATP-sensitive potassium channels as a potential target for antifibrillatory

intervention in acute myocardial-ischemia. *J. Cardiovasc. Pharmacol.*, *26*, 280–288.

144. Yazar, A. Polat, G., Levant, A., Kaygusuz, A., Camdeviren, H., Buyukafsar, K. (2002). Effects of glibenclamide, metaformin and insulin on the incidence and latency of death by ouabian-induced arrhythmias. *Pharmacol. Res.*, *45*, 183–187.
145. Schnitzler, M. M. Y., Derst, C., Daut, J., Preisig-Muller, R. (2000). ATP-sensitive potassium channels in capillaries isloated from guinea-pig heart. *J. Physiol. (Lond.)*, *525*, 307–317.
146. Aguilar-Bryan, L., Nichols, C. G., Wechsler, S. W., Clement, J. P., Boyd, A. E., Gonzales, G., Herrera-Sosa, H., Nguy, K., Bryan, J., Nelson D. (1995). Cloning of the beta cell high affinity sulfonylurea receptor: A regulator of insulin secretion. *Science*, *268*, 423–426.
147. Liss, B., Roeper, J. (2001). Molecular physiology of neuronal K-ATP channels. *Mol. Membr. Biol.*, *18*, 117–127.
148. Chutkow, W. A., Samuel, V., Hansen, P. A., Pu, J., Valdivia, C. R., Makielshi, J. C., Burant, C. F. (2001). Disruption of Sur2-containg $K_{ATP}$ channels enhances insulin-stimulated glucose uptake in skeletal muscle. *Proc. Natl. Acad. Sci. (USA)*, *98*, 11760–11764.
149. Gong, B., Miki, T., Seino, S., Renaud, J. M. (2000). A $K_{ATP}$ channel deficiency affects resting tension, not contractile force, during fatigue in skeletal muscle. *Am. J. Physiol.*, *279*, C1351–1358.
150. Gopalakrishnan, M., Whiteaker, K. L., Molinari, E. J., Davis-Taber, R., Scott, V. E. S., Shieh, C. C., Buckner, S. A., Milicic, I., Cain, J. C., Postl, S., Sullivan, J. P., Brioni, J. D. (1999). Characterization of the ATP-sensitive potassium channels (K-ATP) expressed in guinea pig bladder smooth muscle cells. *J. Pharmacol. Exp. Ther.*, *289*, 551–558.

# FACTORS INFLUENCING THE EFFICACY OF MEDIATOR-SPECIFIC ANTI-INFLAMMATORY, GLUCOCORTICOID, AND ANTICOAGULANT THERAPIES FOR SEPSIS

Peter C. Minneci,[1,2] Katherine J. Deans,[1,2] Michael Haley[1]
Xizhong Cui,[1] Charles Natanson,[1] and Peter Q. Eichacker,[1]

[1]*National Institutes of Health, Bethesda, Maryland*
[2]*Massachusetts General Hospital, Boston, Massachusetts*

1 INTRODUCTION
2 MEDIATOR-SPECIFIC ANTI-INFLAMMATORY THERAPIES IN SEPSIS
Role of Inflammatory Mediators in the Pathophysiology of Sepsis
Mediator-Specific Anti-inflammatory Agent Trials
Factors Influencing the Efficacy of Anti-inflammatory Agents in Sepsis
3 GLUCOCORTICOIDS IN SEPSIS
Glucocorticoid Trials
Factors Influencing the Effects of Glucocorticoids
Comparison of Glucocorticoid Versus Mediator-Specific Anti-inflammatory Agents
4 ANTICOAGULANTS IN SEPSIS
Role of the Coagulation Pathways in Sepsis
Clinical Experience with the Anticoagulant Agents
Factors Influencing the Efficacy of Anticoagulant Agents
5 CONCLUSIONS
References

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

## 1 INTRODUCTION

Sepsis is a major cause of mortality for hospitalized patients worldwide with a mortality rate estimated to be between 30 and 60 percent [1, 2]. During sepsis, cardiovascular dysfunction is produced by bacterial infection with the release of bacterial toxins and the production of harmful host mediators. Septic shock is characterized by a hyperdynamic state including increased cardiac output, decreased systemic vascular resistance, and compromised tissue perfusion secondary to hypotension from loss of vascular integrity and maldistribution of blood flow within the microcirculation. Treatment goals have conventionally included maintenance of blood pressure and identification and eradication of the source of infection. An additional goal of treatment, which is still in evolution, is to interrupt maladaptive responses by the host that contribute to the pathogenesis of this lethal syndrome.

A large amount of evidence now supports the concept that septic shock relates in part to excessive activation of host inflammation and inappropriate activation of coagulation [3–7]. During severe infection, local host defenses, are overcome with the release of microbes and microbial toxins into the intravascular space with rapid activation of systemic host defenses, including the complement, coagulation, and kallikrein-kinin systems and several different types of inflammatory cells. This activation leads to the release of other proinflammatory mediators, including cytokines and oxygen radicals, with further amplification of the immune response. Platelet activation and the formation of microvascular thrombi lead to localized tissue ischemia and other inflammatory stimuli. The host produces endogenous anti-inflammatory and anticoagulant responses to counter this inflammatory activation. However, in some patients, these counterregulatory responses are insufficient, and inflammation and coagulation continue with vasodilatation and tissue injury progressing to shock, organ damage, and death. Interruption of these pathophysiologic events causing septic shock has been the focus of much drug research and development over the past 30 years.

Early attempts to limit excessive inflammation during sepsis with high-dose glucocorticoids were unsuccessful. Because this failure was attributed in part to the nonspecific anti-inflammatory effects of steroids, agents were developed to selectively inhibit specific host inflammatory mediators believed to be the most harmful during sepsis. More recently, based on growing evidence that sepsis may be associated with a state of relative adrenal insufficiency, doses of glucocorticoids much lower than those originally tested have been studied. Finally, to counter activation of the coagulation system during sepsis, anticoagulant agents have been developed and are now under study. To date only low-dose steroids and activated protein C, an anticoagulant, have shown encouraging results in subgroups of patients. However, growing experience in this field now demonstrates that there are important factors capable of influencing the efficacy of each of these different therapies. Accounting for these factors may greatly improve this therapeutic approach. This chapter will

outline the rationale for targeting inflammatory and coagulant pathways during sepsis, review the clinical experience with mediator-specific anti-inflammatory agents, steroids, and anticoagulants, and highlight factors that growing evidence suggests may influence their efficacy.

## 2 MEDIATOR-SPECIFIC ANTI-INFLAMMATORY THERAPIES IN SEPSIS

### Role of Inflammatory Mediators in the Pathophysiology of Sepsis

Evidence of several types supports the role of individual inflammatory mediators in the pathogenesis of septic shock. Many of the manifestations of sepsis can be produced in normal animals and humans by administering purified host inflammatory mediators. Intravenous infusion of either recombinant tumor necrosis factor (TNF)-α or interleukin (IL)-1 to normal animals or humans produces septic physiology characterized by fever, tachycardia, vasodilatation, hypotension, myocardial dysfunction, hemoconcentration, and metabolic acidosis [3, 8–13]. Furthermore, when administered in high doses, these agents are lethal in animal models [14]. The role of the host inflammatory response in the development of septic shock is further supported by the demonstration of increased circulating levels of inflammatory mediators in both animal models of sepsis and septic patients. During meningococcemia, a very lethal form of sepsis, high levels of TNF or other inflammatory mediators were associated with death [15]. Finally, the involvement of the inflammatory response in the development of septic shock is supported by evidence from animal models of sepsis in which treatment with selective anti-inflammatory mediators prevents the physiologic changes associated with sepsis and improves survival [16–31].

### Mediator-Specific Anti-inflammatory Agent Trials

Several different types of mediator-specific anti-inflammatory agents have now been tested in preclinical and clinical trials. Each of these agents was developed to inhibit the effects of a specific proinflammatory mediator thought to play a role in the development of septic shock [3, 5, 32–34]. Agents that have been tested include IL-1 receptor antagonists, antibradykinin agents, anti-TNF antibodies, soluble TNF receptors, platelet activating factor receptor antagonists, platelet activating factor acetylhydrolases, and antiprostaglandin agents. Despite being highly beneficial in preclinical trials, none of these agents demonstrated a significant benefit in their individual clinical trials [35–61] (Table 1). However, a small but significant treatment effect of the mediator-specific anti-inflammatory agents was demonstrated in a recent meta-analysis that combined all of these clinical trials ($n$ = 27) as a group [62] (Fig. 21). The treatment mortality rate was 35.5 percent compared to a control mortality rate of 37.5 percent. The odds ratio of survival with the mediator-specific anti-inflammatory agents was 1.09 (95 percent confidence

**TABLE 1 Clinical Trials of Mediator-Specific Anti-inflammatory Agents**[a]

| Class of Agent | Mechanism of Action | Number of Clinical Trials | Odds Ratio of Survival [62, 111] (95% Confidence Interval) |
|---|---|---|---|
| Anti-TNF mAb | Antibody-specific inactivation of circulating TNF | 9 | 1.10 (0.97–1.25) |
| Soluble TNF receptors | Receptor-specific inactivation of circulating TNF | 3 | 0.95 (0.78–1.16) |
| IL-1 receptor antagonists | Receptor-specific inactivation of IL-1 | 3 | 1.18 (0.97–1.44) |
| Prostaglandin antagonists | Cyclooxygenase inhibition of prostaglandin production | 3 | 1.22 (0.78–1.58) |
| Bradykinin antagonists | Inhibition of the kallikrein/kinin cascade | 2 | 0.91 (0.65–1.27) |
| PAF antagonists | Receptor-specific inactivation or enzyme degradation of PAF | 7 | 1.10 (0.93–1.30) |

[a] Legend: TNF-mAb, tumor necrosis factor monoclonal antibody; PAF, platelet activating factor; IL-1: Interleukin-1.

interval (CI) 1.01 to 1.18, $p = .03$). Hence, these agents resulted in an absolute risk reduction of 2 percent and a relative risk reduction of 7 percent. The improvement in the odds ratio of survival in this meta-analysis supports the involvement of the systemic inflammatory response in the development of the sepsis syndrome. However, it can be seen in this meta-analysis that the larger individual clinical trials (>250 patients) more accurately captured the treatment effect of these agents compared to the smaller individual clinical trials (<250 patients) (Fig. 1). Based on this analysis, an individual trial of any one of these mediator-specific agents would require approximately 6000 patients to demonstrate a statistically significant treatment effect.

## Factors Influencing the Efficacy of Anti-inflammatory Agents in Sepsis

Clinical trials of the mediator-specific anti-inflammatory agents in septic patients did not demonstrate the significant treatment effects that were shown

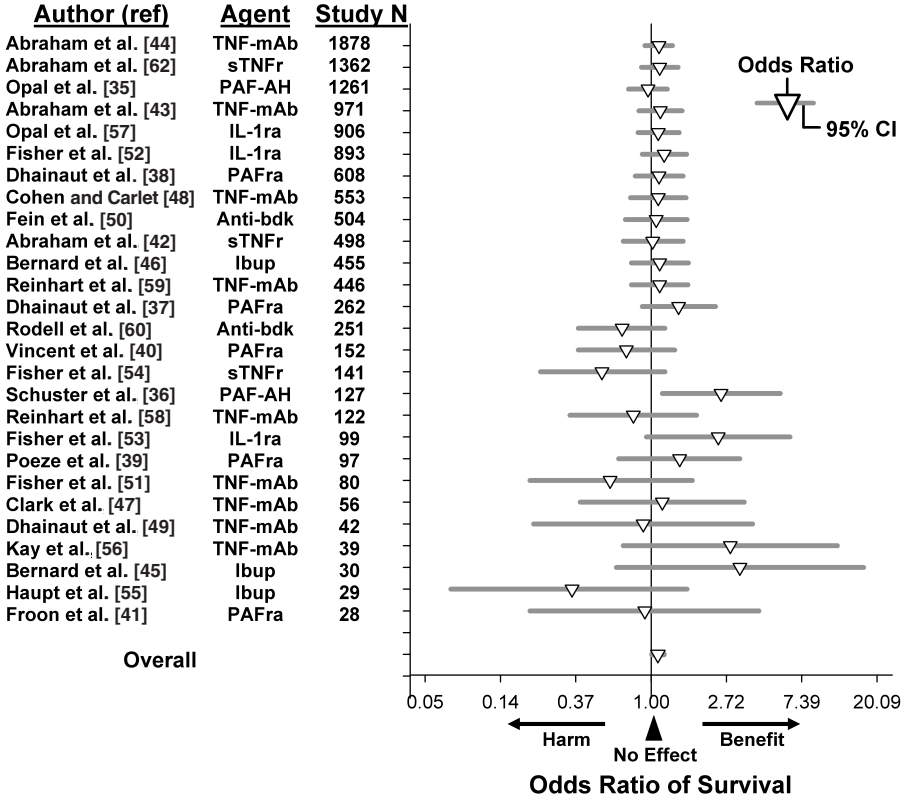


**Figure 1** Odds ratio of survival for clinical trials of mediator-specific anti-inflammatory agents: The odds ratio of survival for each clinical trial of an anti-inflammatory agent is shown as the inverted triangle with the solid line representing its 95% confidence interval [35–61]. The trials are arranged by the number of patients in each trial. The treatment effect becomes more consistent as the size of the trial increases. Overall, there is a significant improvement in survival with treatment ($p = .03$). Legend: TNF-mAb, tumor necrosis factor monoclonal antibody; sTNFr, soluble tumor necrosis factor receptor; IL-1ra, interleukin-1 receptor antagonist; PAFra, platelet activating factor receptor antagonist; PAF-AH, platelet activating factor acetylhydrolase; Ibup, ibuprofen; Anti-bdk, anti bradykinin agent.

in the preclinical trials that tested them. Potential reasons discussed to explain these disparate results have included differences in the type or route of infection. It has also been suggested that the underlying risk of death associated with sepsis may influence the treatment effect of anti-inflammatory agents [63].

A related series of studies recently assessed the influence of several of these factors on the efficacy of the anti-inflammatory agents in sepsis [64]. These studies included an analysis of past published preclinical and clinical trials as

well as prospective experiments. Using metaregression analysis, the investigators examined the effect of these agents in relationship to changes in the type, severity, or site of infection. Of these factors, severity of infection and its associated risk of death, as measured by control group mortality rate, was found to profoundly influence the treatment effect of several different mediator-specific anti-inflammatory agents. Although very beneficial when the risk of death was great, these agents became less beneficial and were potentially harmful as this risk decreased (Fig. 2). This relationship is important because published preclinical trials of mediator-specific anti-inflammatory agents were performed with a very high risk of death (median control mortality rate 88 percent) as compared to the clinical trials that were performed at a significantly lower risk of death (median control mortality rate 41 percent) (Fig. 2). Hence, the preclinical trials of these agents were carried out at a risk of death level where these agents are very beneficial compared to the clinical trials, which had a level of risk at which these agents would have little benefit.

Furthermore, predicted levels of efficacy of these agents in clinical trials based on the demonstrated relationship between severity of illness and treatment effect in the preclinical trials would have been similar to the actual effects that were reported in the clinical trials of these agents. This suggests that the small treatment effects of these agents detected in their clinical trials could have been predicted if the relationship between treatment effect and

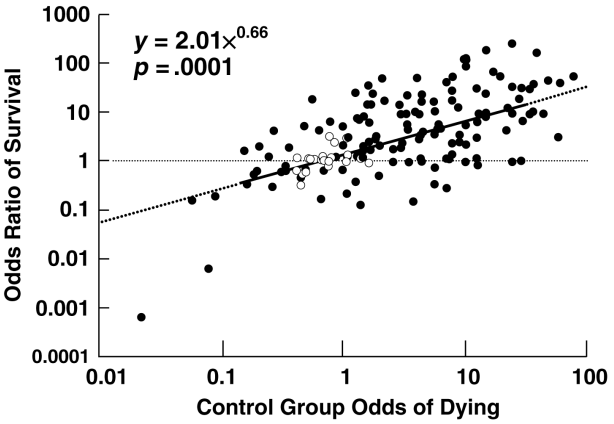


**Figure 2** Relationship between treatment effect of mediator-specific anti-inflammatory agents and severity of illness. A weighted regression line representing the relationship between control odds of dying and the odds ratio of survival with treatment for published and prospective animal studies (closed circles) [64] and in the clinical trials (open circles) of mediator-specific anti-inflammatory agents [35–61]. There is a significant relationship ($p$ = .0001) between the treatment effect of these agents and the control odds of dying. These agents are beneficial at a high risk of death and have no effect or are harmful at a low risk.

risk of death had been taken into account. To be beneficial clinically, these agents may have to be directed to those patients with a high risk of death.

## 3 GLUCOCORTICOIDS IN SEPSIS

### Glucocorticoid Trials

Glucocorticoids were the first type of anti-inflammatory agent tested in the treatment of sepsis. Steroids have been shown to have anti-inflammatory properties and to improve vascular reactivity [65]. As with other anti-inflammatory agents, administration of glucocorticoids in animal models of sepsis led to improved outcome [28, 66–71]. However, early clinical trials using short courses of very high doses of glucocorticoids to block excessive inflammation demonstrated varying efficacy [72–79]. Three meta-analyses of the prospective randomized trials of high-dose glucocorticoids were performed [5, 80, 81]. Two of these analyses demonstrated no beneficial effects of steroid therapy[80, 81], and the third meta-analysis demonstrated harmful effects of steroid therapy after exclusion of a single outlying study (odds ratio of survival of 0.70, 95 percent confidence interval 0.556 to 0.91, $p = .008$) [5] (Fig. 3). The harmful effects of steroids in these trials may have been due to an increased incidence and severity of secondary infections caused by the immunosuppressive effects of high-dose steroid therapy [74, 78, 79].

More recent studies have demonstrated that relative adrenal suppression and adrenal hyporesponsiveness occur in septic patients and correlate with increased mortality [82–84]. The results of these studies led to renewed interest in the potential benefits of steroid therapy in the treatment of septic shock. Five recent studies were performed using lower glucocorticoid doses, termed either *stress* or *physiologic* doses, in septic patients who required persistent vasopressor support [85–89]. Each of these trials demonstrated beneficial effects of steroids on survival and/or shock reversal. A meta-analysis combining the results of these trials demonstrated a consistent and significant beneficial effect of steroids on survival (odds ratio, 95 percent confidence interval, $p$ value) (1.52, 1.03 to 2.27, $p = .036$) and shock reversal (4.79, 2.07 to 11.11, $p = .001$). [90] (Fig. 3). Furthermore, analysis of the three trials that used an ACTH stimulation test to classify patients as either "responders" (normal adrenal function) or "nonresponders" (relative adrenal insufficiency), revealed similar beneficial effects of steroids on survival and shock reversal in both groups [90].

### Factors Influencing the Effects of Glucocorticoids

Analysis of all of the randomized controlled clinical sepsis trials of steroids demonstrated that their effects were dose dependent with high doses being harmful and low doses being beneficial ($p = .004$) [90]. In addition, compared

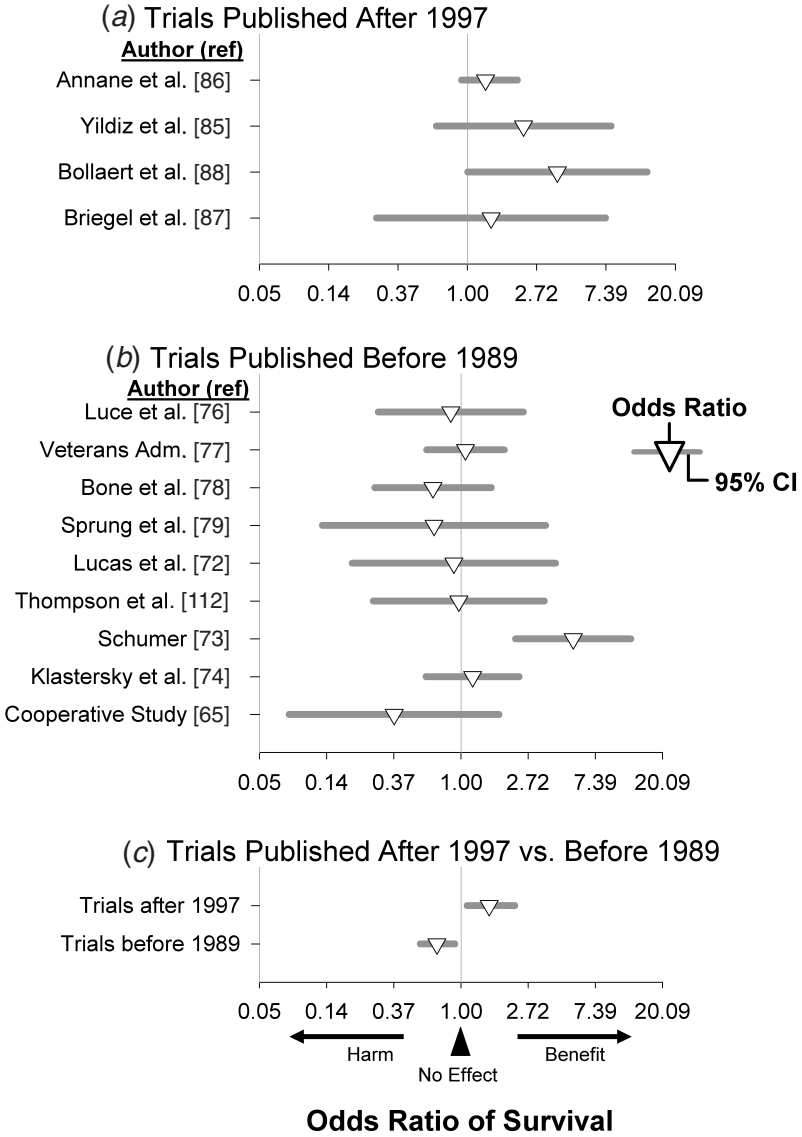
(a) Trials Published After 1997
Author (ref)
Annane et al. [86]
Yildiz et al. [85]
Bollaert et al. [88]
Briegel et al. [87]
0.05 0.14 0.37 1.00 2.72 7.39 20.09
(b) Trials Published Before 1989
Author (ref)
Luce et al. [76]
Veterans Adm. [77]
Bone et al. [78]
Sprung et al. [79]
Lucas et al. [72]
Thompson et al. [112]
Schumer [73]
Klastersky et al. [74]
Cooperative Study [65]
Odds Ratio
95% CI
0.05 0.14 0.37 1.00 2.72 7.39 20.09
(c) Trials Published After 1997 vs. Before 1989
Trials after 1997
Trials before 1989
0.05 0.14 0.37 1.00 2.72 7.39 20.09
Harm
No Effect
Benefit
Odds Ratio of Survival

to the earlier trials of high-dose steroid therapy, the more recent trials of low-dose steroids were performed in sicker patients with higher control mortality rates (mean, $p$ value) (57 vs. 34 percent, $p = .06$) and who were more likely to be on vasopressors (100 vs. 65 percent, $p = .03$). Furthermore treatment in the more recent trials was begun later in the course of the illness (median, $p$ value) (23 vs. <2 h, $p = .02$) and continued for a longer period of time (6 vs. 1 day, $p = .004$) [90]. Therefore, in contrast to high-dose steroids, which decreased survival rates during sepsis, physiologic dose steroids reversed shock and increased survival in patients with established vasopressor-dependent septic shock. However, other factors that differed between earlier and later trials may have also altered the effects of these agents.

## Comparison of Glucocorticoid Versus Mediator-Specific Anti-inflammatory Agents

Several reasons may explain why physiologic dose steroids appeared more beneficial than mediator-specific anti-inflammatory agents in the treatment of sepsis. First, steroids may not only limit inflammatory injury but also reverse a state of relative adrenal insufficiency [88]. Second, although a significant linear relationship between treatment effect and control mortality was not demonstrated among the clinical steroid trials, it is worth noting that the average control mortality rate in the beneficial low-dose glucocorticoid trials was >55 percent. This control mortality rate with steroids was higher than in clinical trials testing mediator-specific anti-inflammatory agents, possibly because patients were only enrolled if they had persistent vasopressor requirement in the former but not the latter. The need for vasopressors in patients receiving steroids may have been a marker for an increased risk of death. As a result, the beneficial glucocorticoid trials were performed in patients with risks of death consistent with those where mediator-specific agents would also have been beneficial [64]. Therefore, the differing effects of these agents may

**Figure 3** Odds ratio of survival for clinical trials of glucocorticoids. The odds ratio of survival (open triangles) and 95% confidence intervals (horizontal lines) with glucocorticoid therapy in sepsis trials. In the four recently published sepsis trials of (*a*) low-dose steroids, the effects of steroids were similar, and (*c*) when combined in a meta-analysis, there was a significant improvement in the odds ratio of survival ($p = .036$) [85–88]. The effects of steroids on the odds ratio of survival in the 9 sepsis trials of high-dose steroids published before 1989 were variable; however, with exclusion of one trial [73], which was a significant outlier, the effects of steroids in the other (*b*) eight trials on the odds ratio of survival were consistent [65, 72, 74, 76–79, 112]. When these eight trials were combined in a meta-analysis, there was a significant decrease in the odds ratio of survival with steroid therapy during sepsis ($p = .008$). Thus, (*c*) the effects of steroids on survival rates in the recent low-dose sepsis trials are significantly different than those in the older trials of high-dose steroids.

be explained by differences in the severity of illness of the patients studied and their underlying risks of death.

## 4 ANTI-COAGULANTS IN SEPSIS

### Role of the Coagulation Pathways in Sepsis

Activation of the coagulation system during sepsis is supported by the frequent occurrence of disseminated intravascular coagulation and the presence of intravascular thrombi at autopsy in septic patients [6, 7]. During sepsis, increased levels of activated coagulation factors, tissue factor and tissue factor pathway inhibitor, and decreased fibrinogen levels have been documented [6]. The intrinsic and extrinsic coagulation pathways are activated by lipopolysaccharides in the bacterial cell wall, which can stimulate tissue factor production. The subsequent formation of thrombin can further activate both procoagulant and anticoagulant pathways. In addition, coagulation system activation potentiates the host inflammatory response through endothelial cell activation with increased cytokine production and leukocyte adhesion [6, 91–93]. These coagulation abnormalities correlate with the development of organ dysfunction and are associated with increased mortality [94, 95]. Based on these interactions and associations, anticoagulants, including antithrombin-III, tissue factor pathway inhibitor, and activated protein C, have been developed and studied as treatments for severe sepsis and septic shock.

### Clinical Experience with the Anticoagulant Agents

Based on successful preclinical trials, there have been 12 randomized controlled clinical trials of anticoagulant agents in septic patients, 8 of antithrombin-III (AT-III), 2 of tissue factor pathway inhibitor (TFPI), and 2 of activated protein C (APC) [96–107]. The overall odds ratio of survival with treatment in the 2 APC trials was 1.31 (95 percent CI 1.07 to 1.61, $p = .009$), in the 2 TFPI trials was 1.01 (95 percent CI 0.84 to 1.23, $p = .86$), and in the 8 AT-III trials was 1.03 (95 percent CI 0.88 to 1.20, $p = .72$). The treatment effects in all of these trials agents were consistent, and there was an overall trend toward a small improvement in survival with anticoagulant therapies (odds ratio of survival 1.09, 95 percent CI 0.98 to 1.21, $p = .098$) (Fig. 4). Furthermore, each of these agents was associated with significant increases in risk of bleeding [94, 96, 106].

### Factors Influencing the Efficacy of Anti-Coagulant Agents

As with both mediator-specific and glucocorticoid anti-inflammatory agents, severity of infection and its associated risk of death may also alter the effects of anticoagulants. Of all the anti-inflammatory and anticoagulant agents now

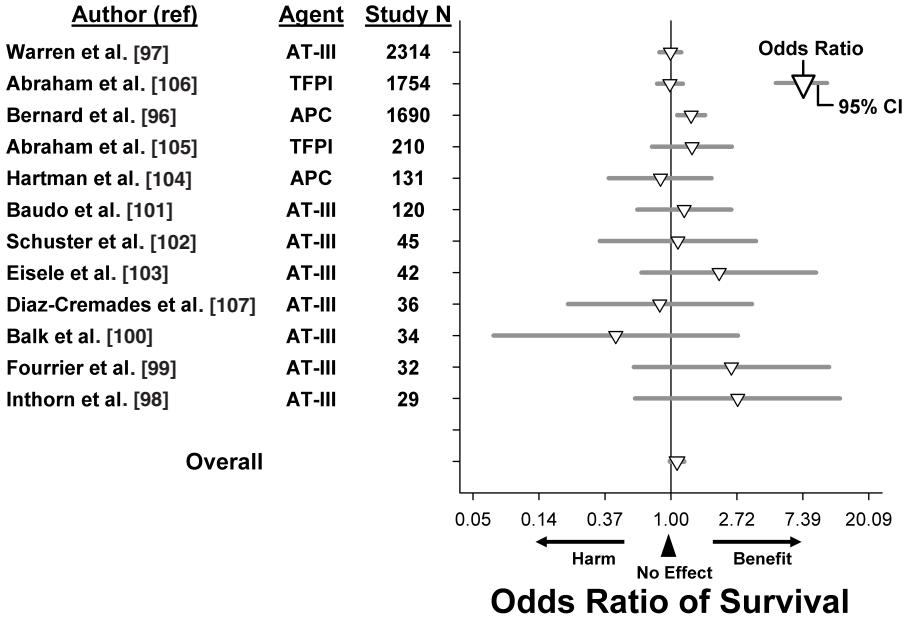


**Figure 4** Odds ratio of survival for clinical trials of the anticoagulant agents. The odds ratio of survival for each clinical trial of an anticoagulant is shown as the inverted triangle with the solid line representing its 95 percent confidence interval [96–107]. Overall, there is a trend toward improvement in survival with treatment ($p = .10$). APC, activated protein C; AT-III, antithrombin III; TFPI, tissue factor pathway inhibitor.

tested in sepsis, only APC was shown to have a significant beneficial effect on survival in a single phase III trial [96]. In a trial of 1690 patients, APC, which has both anti-inflammatory and antithrombotic properties, significantly ($p = .005$) reduced mortality rates from 30.8 percent in controls to 24.7 percent in treated patients [96]. In this trial, however, APC demonstrated a relationship between treatment effect and risk of death that was similar to the one previously noted in preclinical and clinical trials testing mediator-specific anti-inflammatory agents [64]. It was most beneficial in patients with a high risk of death and less beneficial and on the side of harm in patients with a low risk of death (Table 2). Because of this relationship, the use of APC was limited by the Food and Drug Administration (FDA) to patients with severe sepsis (sepsis associated with acute organ dysfunction) and a high risk of death. Furthermore, the FDA requested that the manufacturer of APC perform phase IV trials and clarify the agent's effects in patients with a low risk of death. That trial has reportedly been terminated after enrollment of only 2000 of the 11,000 anticipated subjects for futility.

The significant relationship between treatment effect and severity of illness in the phase III trial of APC suggests that this factor could alter the efficacy

**TABLE 2 Results of Phase III Trial of Activated Protein C [96] Stratified by Severity of Illness[a]**

| APACHE II Score | Control Mortality (%) | Odds Ratio of Survival |
|---|---|---|
| 3–19 | 15 | 0.77 |
| 20–24 | 26 | 1.19 |
| 25–29 | 36 | 1.82 |
| 30–52 | 49 | 1.56 |

[a] Legend: APACHE, acute physiology, age, and chronic health evaluation.

of the other anticoagulant agents. In fact an earlier meta-analysis of 10 of these prospective trials and 1 retrospective suggested that such a relationship might exist [94]. However, with addition of the most recent phase III trial testing TFPI and one other prospective trial and the exclusion of the retrospective trial, this relationship was no longer significant. Furthermore, analysis of subgroups within the phase III trials testing TFPI and antithrombin III failed to show a relationship between risk of death and the efficacy of these. Additional analysis of the large phase III trials testing APC, AT-III, and TFPI suggests that another factor, concurrent treatment with heparin, may have confounded the results of these trials and, in the case of the latter two, potentially masked the influence of severity of sepsis [96, 97, 106]. In each of these trials, the experimental study agents were less beneficial in patients who received concurrent heparin therapy than in those that did not [108] (Fig. 5). This relationship was consistent across all three trials but was strongest in the APC trial. Although there are several potential reasons for these results, one is that potential beneficial anticoagulant effects of heparin may have negated those of the study drugs [108, 109]. Consistent with this possibility, the patients in the placebo groups of these trials that received heparin therapy had a better outcome than patients in the placebo groups not receiving heparin (Fig. 6). Furthermore, in a phase III trial examining the effect of prophylactic low-molecular-weight heparin in critically ill patients, many of whom were likely septic, heparin therapy was reported to significantly improve clinical outcome [110]. Therefore, heparin, which could have its own beneficial effects on survival in critically ill patients, may have decreased the ability of the anticoagulant agents to improve survival.

## 5 CONCLUSIONS

Newer treatment agents for sepsis directed at the host's response during infection have had differing survival effects that appear dependent, at least in part, on several different factors. Mediator-specific anti-inflammatory agents have been shown overall to have a small beneficial effect on survival. However, the

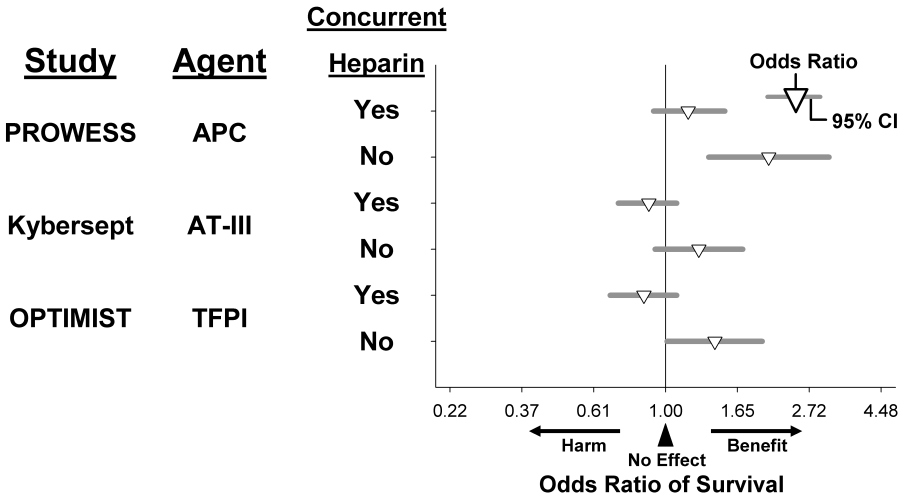


**Figure 5** Effect of concurrent heparin therapy on the odds ratio of survival in the anticoagulant trials. The odds ratio of survival (triangle) and 95 percent confidence intervals (horizontal lines) for the large phase III trials of the three different anticoagulant agents (APC, AT-III, TFPI) in patients who either were or were not receiving concurrent heparin [96, 97, 106]. The treatment effect of each study agent was less beneficial or harmful in patients receiving heparin compared to patients not receiving heparin.

treatment effects of these agents are likely related to the patients' risk of death. In patients at a higher risk of death, these agents may be more beneficial, and in those at lower risk they appear ineffective or potentially harmful. The disparate treatment effects of these agents in clinical and preclinical trials can be explained by this relationship.

The effects of glucocorticoids in sepsis are dose dependent with high doses having a harmful effect and low doses being beneficial. Analysis of the glucocorticoid trials suggests that control mortality rates were higher in the beneficial trials and lower in the harmful ones. While these are consistent with the relationship between severity of illness and treatment effect noted for mediator-specific anti-inflammatory, a significant relationship was not found in the steroid trials. This may be secondary to an overwhelming influence of drug dose on treatment effect.

As a class anticoagulant agents did not have a beneficial effect on survival and did not demonstrate a relationship between risk of death and treatment effect. However, activated protein C did demonstrate a beneficial effect on survival with a relationship between treatment effect and severity of illness that was very similar to the mediator-specific anti-inflammatory agents. Furthermore, the treatment effect of the anticoagulant agents may have been confounded by concurrent heparin therapy. This potential interaction will require further study.

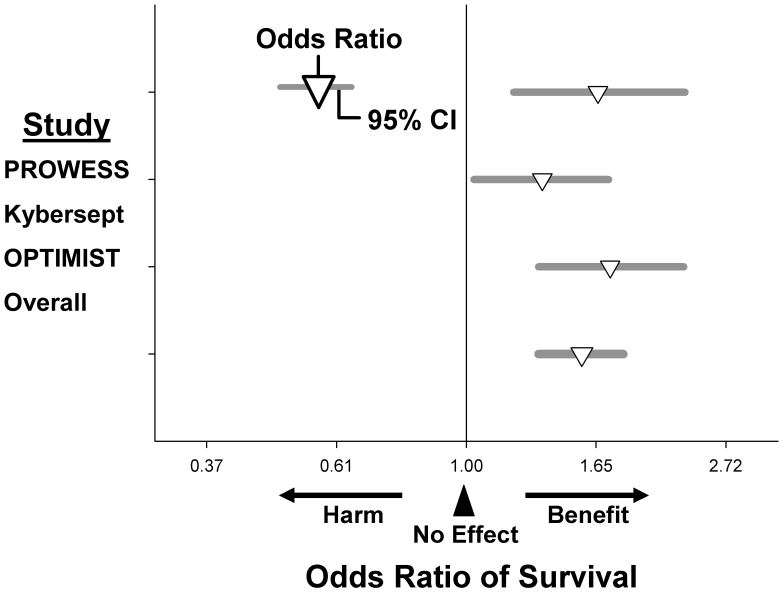


**Figure 6** Efficacy of heparin in the placebo groups of the anticoagulant trials. The odds ratio of survival (triangle) with heparin treatment and 95 percent confidence intervals (horizontal lines) in each of the placebo groups of the phase III trials of APC (PROWESS trial), AT-III (Kybersept trial ), and TFPI (OPTIMIST trial ) [96, 97, 106]. Heparin treatment was beneficial in all placebo groups and resulted in an overall improvement in the odds ratio of survival ($p < .0001$).

Our evolving clinical experience with novel therapies for sepsis, including mediator-specific anti-inflammatory agents, glucocorticoids, and anticoagulant agents, suggests that several factors may alter their effects. Based on this knowledge, these agents may only be beneficial in particular subgroups of patients and therefore require further testing in targeted populations. Development of future agents for sepsis should include careful preclinical and clinical testing that account for factors such as the severity of illness, the dose-dependent effects of the therapies under study, and the influence of concurrent therapies with competing effects.

## REFERENCES

1. Center for Disease Control: Increase in national hospital discharge survey rates for septicemia—United States, 1979–1987. *MMWR*, 1990, *39*, 31–34.
2. Angus, D. C., Linde-Zwirble, W. T., Lidicker, J., Clermont, G., Carcillo, J., Pinsky, M. R. (2001). Epidemiology of severe sepsis in the United States: Analysis of incidence, outcome, and associated costs of care. *Crit. Care Med.*, *29*, 1303–1310.

3. Natanson, C., Hoffman, W. D., Suffredini, A. F., Eichacker, P. Q., Danner, R. L. (1994). Selected treatment strategies for septic shock based on proposed mechanisms of pathogenesis. *Ann. Intern. Med.*, *120*, 771–783.
4. Quezado, Z. M., Banks, S. M., Natanson, C. (1995). New strategies for combatting sepsis: The magic bullets missed the mark . . . but the search continues. *Trends Biotechnol.*, *13*, 56–63.
5. Zeni, F., Freeman, B., Natanson, C. (1997). Anti-inflammatory therapies to treat sepsis and septic shock: A reassessment. *Crit. Care Med.*, *25*, 1095–1100.
6. Camerota, A. J., Creasey, A. A., Patla, V., Larkin, V. A., Fink, M. P. (1998). Delayed treatment with recombinant human tissue factor pathway inhibitor improves survival in rabbits with gram-negative peritonitis. *J. Infect. Dis.*, *177*, 668–676.
7. Park, C. T., Creasey, A. A., Wright, S. D. (1997). Tissue factor pathway inhibitor blocks cellular effects of endotoxin by binding to endotoxin and interfering with transfer to CD14. *Blood*, *89*, 4268–4274.
8. Suffredini, A. F., Fromm, R. E., Parker, M. M., Brenner, M., Kovacs, J. A., Wesley, R. A., Parrillo, J. E. (1989). The cardiovascular response of normal humans to the administration of endotoxin. *N. Engl. J. Med.*, *321*, 280–287.
9. Alexander, R. B., Ponniah, S., Hasday, J. and Hebel, J. R. (1998). Elevated levels of proinflammatory cytokines in the semen of patients with chronic prostatitis/chronic pelvic pain syndrome. *Urology*, *52*, 744–749.
10. Natanson, C., Eichenholz, P. W., Danner, R. L., Eichacker, P. Q., Hoffman, W. D., Kuo, G. C., Banks, S. M., MacVittie, T. J., Parrillo, J. E. (1989). Endotoxin and tumor necrosis factor challenges in dogs simulate the cardiovascular profile of human septic shock. *J. Exp. Med.*, *169*, 823–832.
11. Waage, A., Espevik, T. (1988). Interleukin 1 potentiates the lethal effect of tumor necrosis factor alpha/cachectin in mice. *J. Exp. Med.*, *167*, 1987–1992.
12. Tracey, K. J., Beutler, B., Lowry, S. F., Merryweather, J., Wolpe, S., Milsark, I. W., Hariri, R. J., Fahey, T. J., 3rd, Zentella, A., Albert, J. D., et al. (1986). Shock and tissue injury induced by recombinant human cachectin. *Science*, *234*, 470–474.
13. Okusawa, S., Gelfand, J. A., Ikejima, T., Connolly, R. J., Dinarello, C. A. (1988). Interleukin 1 induces a shock-like state in rabbits. Synergism with tumor necrosis factor and the effect of cyclooxygenase inhibition. *J. Clin. Invest.*, *81*, 1162–1172.
14. Eichacker, P. Q., Hoffman, W. D., Farese, A., Banks, S. M., Kuo, G. C., MacVittie, T. J., Natanson, C. (1991). TNF but not IL-1 in dogs causes lethal lung injury and multiple organ dysfunction similar to human sepsis. *J. Appl. Physiol.*, *71*, 1979–1989.
15. Waage, A., Halstensen, A., Espevik, T. (1997). Association between tumor necrosis factor in serum and fatal outcome in patients with meningococcal disease. *Lancet*, *1*, 355–357.
16. Ohlsson, K., Bjork, P., Bergenfeldt, M., Hageman, R., Thompson, R. C. (1990). Interleukin-1 receptor antagonist reduces mortality from endotoxin shock. *Nature*, *348*, 550–552.
17. Wakabayashi, G., Gelfand, J. A., Burke, J. F., Thompson, R. C., Dinarello, C. A. (1991). A specific receptor antagonist for interleukin 1 prevents *Escherichia coli*-induced shock in rabbits. *Faseb J.*, *5*, 338–343.

18. Fischer, E., Marano, M. A., Van Zee, K. J., Rock, C. S., Hawes, A. S., Thompson, W. A., DeForge, L., Kenney, J. S., Remick, D. G., Bloedow, D. C., et al. (1992). Interleukin-1 receptor blockade improves survival and hemodynamic performance in *Escherichia coli* septic shock, but fails to alter host responses to sublethal endotoxemia. *J. Clin. Invest.*, *89*, 1551–1557.
19. Beutler, B., Milsark, I. W., Cerami, A. C. (1985). Passive immunization against cachectin/tumor necrosis factor protects mice from lethal effect of endotoxin. *Science*, *229*, 869–871.
20. Suitters, A. J., Foulkes, R., Opal, S. M., Palardy, J. E., Emtage, J. S., Rolfe, M., Stephens, S., Morgan, A., Holt, A. R., Chaplin, L. C., et al. (1994). Differential effect of isotype on efficacy of anti-tumor necrosis factor alpha chimeric antibodies in experimental septic shock. *J. Exp. Med.*, *179*, 849–856.
21. Bagby, G. J., Plessala, K. J., Wilson, L. A., Thompson, J. J., Nelson, S. (1991). Divergent efficacy of antibody to tumor necrosis factor-alpha in intravascular and peritonitis models of sepsis. *J. Infect. Dis.*, *163*, 83–88.
22. Mathison, J. C., Wolfson, E., Ulevitch, R. J. (1988). Participation of tumor necrosis factor in the mediation of gram negative bacterial lipopolysaccharide-induced injury in rabbits. *J. Clin. Invest.*, *81*, 1925–1937.
23. Fiedler, V. B., Loof, I., Sander, E., Voehringer, V., Galanos, C., Fournel, M. A. (1992). Monoclonal antibody to tumor necrosis factor–alpha prevents lethal endotoxin sepsis in adult rhesus monkeys. *J. Lab. Clin. Med.*, *120*, 574–588.
24. Emerson, T. E., Jr., Lindsey, D. C., Jesmok, G. J., Duerr, M. L., Fournel, M. A. (1992). Efficacy of monoclonal antibody against tumor necrosis factor alpha in an endotoxemic baboon model. *Circ. Shock*, *38*, 75–84.
25. Eskandari, M. K., Bolgos, G., Miller, C., Nguyen, D. T., DeForge, L. E., Remick, D. G. (1992). Anti-tumor necrosis factor antibody therapy fails to prevent lethality after cecal ligation and puncture or endotoxemia. *J. Immunol.*, *148*, 2724–2730.
26. Silva, A. T., Bayston, K. F., Cohen, J. (1990). Prophylactic and therapeutic effects of a monoclonal antibody to tumor necrosis factor-alpha in experimental gram-negative shock. *J. Infect. Dis.*, *162*, 421–427.
27. Jesmok, G., Lindsey, C., Duerr, M., Fournel, M., Emerson, T., Jr. (1992). Efficacy of monoclonal antibody against human recombinant tumor necrosis factor in *E. coli*-challenged swine. *Am. J. Pathol.*, *141*, 1197–1207.
28. Hinshaw, L. B., Archer, L. T., Beller-Todd, B. K., Coalson, J. J., Flournoy, D. J., Passey, R., Benjamin, B., White, G. L. (1980). Survival of primates in LD100 septic shock following steroid/antibiotic therapy. *J. Surg. Res.*, *28*, 151–170.
29. Hinshaw, L. B., Emerson, T. E., Jr., Taylor, F. B., Jr., Chang, A. C., Duerr, M., Peer, G. T., Flournoy, D. J., White, G. L., Kosanke, S. D., Murray, C. K., et al. (1992). Lethal *Staphylococcus aureus*-induced shock in primates: Prevention of death with anti-TNF antibody. *J. Trauma*, *33*, 568–573.
30. Tracey, K. J., Fong, Y., Hesse, D. G., Manogue, K. R., Lee, A. T., Kuo, G. C., Lowry, S. F., Cerami, A. (1987). Anti-cachectin/TNF monoclonal antibodies prevent septic shock during lethal bacteraemia. *Nature*, *330,* 662–664.
31. Opal, S. M., Cross, A. S., Kelly, N. M., Sadoff, J. C., Bodmer, M. W., Palardy, J. E., Victor, G. H. (1990). Efficacy of a monoclonal antibody directed against tumor necrosis factor in protecting neutropenic rats from lethal infection with *Pseudomonas aeruginosa*. *J. Infect. Dis.*, *161*, 1148–1152.

32. Natanson, C., Danner, R. L., Reilly, J. M., Doerfler, M. L., Hoffman, W. D., Akin, G. L., Hosseini, J. M., Banks, S. M., Elin, R. J., MacVittie, T. J., et al. (1990). Antibiotics versus cardiovascular support in a canine model of human septic shock. *Am. J. Physiol.*, *259*, H1440–H1447.
33. Wheeler, A. P., Bernard, G. R. (1999). Treating patients with severe sepsis. *N. Engl. J. Med.*, *340*, 207–214.
34. Parrillo, J. E. (1993). Pathogenetic mechanisms of septic shock. *N. Engl. J. Med.*, *328*, 1471–1477.
35. Opal, S., Laterre, P. F., Abraham, E., Francois, B., Wittebole, X., Lowry, S., Dhainaut, J. F., Warren, B., Dugernier, T., Lopez, A., Sanchez, M., Demeyer, I., Jauregui, L., Lorente, J. A., McGee, W., Reinhart, K., Kljucar, S., Souza, S., Pribble, J. (2004). Recombinant human platelet-activating factor acetylhydrolase for treatment of severe sepsis: Results of a phase III, multicenter, randomized, double-blind, placebo-controlled, clinical trial. *Crit. Care Med.*, *32*, 332–341.
36. Schuster, D. P., Metzler, M., Opal, S., Lowry, S., Balk, R., Abraham, E., Levy, H., Slotman, G., Coyne, E., Souza, S., Pribble, J. (2003). Recombinant platelet-activating factor acetylhydrolase to prevent acute respiratory distress syndrome and mortality in severe sepsis: Phase IIb, multicenter, randomized, placebo-controlled, clinical trial. *Crit. Care Med.*, *31*, 1612–1619.
37. Dhainaut, J. F., Tenaillon, A., Le Tulzo, Y., Schlemmer, B., Solet, J. P., Wolff, M., Holzapfel, L., Zeni, F., Dreyfuss, D., Mira, J. P., et al. (1994). Platelet-activating factor receptor antagonist BN 52021 in the treatment of severe sepsis: A randomized, double-blind, placebo-controlled, multicenter clinical trial. BN 52021 Sepsis Study Group. *Crit. Care Med.*, *22*, 1720–1728.
38. Dhainaut, J. F., Tenaillon, A., Hemmer, M., Damas, P., Le Tulzo, Y., Radermacher, P., Schaller, M. D., Sollet, J. P., Wolff, M., Holzapfel, L., Zeni, F., Vedrinne, J. M., de Vathaire, F., Gourlay, M. L., Guinot, P., Mira, J. P. (1998). Confirmatory platelet-activating factor receptor antagonist trial in patients with severe gram-negative bacterial sepsis: A phase III, randomized, double-blind, placebo-controlled, multicenter trial. BN 52021 Sepsis Investigator Group. *Crit. Care Med.*, *26*, 1963–1971.
39. Poeze, M., Froon, A. H., Ramsay, G., Buurman, W. A., Greve, J. W. (2000). Decreased organ failure in patients with severe SIRS and septic shock treated with the platelet-activating factor antagonist TCV-309: A prospective, multicenter, double-blind, randomized phase II trial. TCV-309 Septic Shock Study Group. *Shock*, *14*, 421–428.
40. Vincent, J. L., Spapen, H., Bakker, J., Webster, N. R., Curtis, L. (2000). Phase II multicenter clinical study of the platelet-activating factor receptor antagonist BB-882 in the treatment of sepsis. *Crit. Care Med.*, *28*, 638–642.
41. Froon, A. M., Greve, J. W., Buurman, W. A., van der Linden, C. J., Langemeijer, H. J., Ulrich, C., Bourgeois, M. (1996). Treatment with the platelet-activating factor antagonist TCV-309 in patients with severe systemic inflammatory response syndrome: A prospective, multi-center, double-blind, randomized phase II trial. *Shock*, *5*, 313–319.
42. Abraham, E., Glauser, M. P., Butler, T., Garbino, J., Gelmont, D., Laterre, P. F., Kudsk, K., Bruining, H. A., Otto, C., Tobin, E., Zwingelstein, C., Lesslauer, W., Leighton, A. (1997). p55 Tumor necrosis factor receptor fusion protein in

the treatment of patients with severe sepsis and septic shock. A randomized controlled multicenter trial. Ro 45-2081 Study Group. *JAMA*, *277*, 1531–1538.

43. Abraham, E., Wunderink, R., Silverman, H., Perl, T. M., Nasraway, S., Levy, H., Bone, R., Wenzel, R. P., Balk, R., Allred, R., et al. (1995). Efficacy and safety of monoclonal antibody to human tumor necrosis factor alpha in patients with sepsis syndrome. A randomized, controlled, double-blind, multicenter clinical trial. TNF-alpha MAb Sepsis Study Group. *JAMA*, *273*, 934–941.
44. Abraham, E., Anzueto, A., Gutierrez, G., Tessler, S., San Pedro, G., Wunderink, R., Dal Nogare, A., Nasraway, S., Berman, S., Cooney, R., Levy, H., Baughman, R., Rumbak, M., Light, R. B., Poole, L., Allred, R., Constant, J., Pennington, J., Porter, S. (1998). Double-blind randomised controlled trial of monoclonal antibody to human tumour necrosis factor in treatment of septic shock. NORASEPT II Study Group. *Lancet*, *351*, 929–933.
45. Bernard, G. R., Reines, H. D., Halushka, P. V., Higgins, S. B., Metz, C. A., Swindell, B. B., Wright, P. E., Watts, F. L., Vrbanac, J. J. (1991). Prostacyclin and thromboxane A2 formation is increased in human sepsis syndrome. Effects of cyclooxygenase inhibition. *Am. Rev. Respir. Dis.*, *144*, 1095–1101.
46. Bernard, G. R., Wheeler, A. P., Russell, J. A., Schein, R., Summer, W. R., Steinberg, K. P., Fulkerson, W. J., Wright, P. E., Christman, B. W., Dupont, W. D., Higgins, S. B., Swindell, B. B. (1997). The effects of ibuprofen on the physiology and survival of patients with sepsis. The Ibuprofen in Sepsis Study Group. *N. Engl. J. Med.*, *336*, 912–918.
47. Clark, M. A., Plank, L. D., Connolly, A. B., Streat, S. J., Hill, A. A., Gupta, R., Monk, D. N., Shenkin, A., Hill, G. L. (1998). Effect of a chimeric antibody to tumor necrosis factor-alpha on cytokine and physiologic responses in patients with severe sepsis—a randomized, clinical trial. *Crit. Care Med.*, *26*, 1650–1659.
48. Cohen, J., Carlet, J. (1996). INTERSEPT: An international, multicenter, placebo-controlled trial of monoclonal antibody to human tumor necrosis factor-alpha in patients with sepsis. International Sepsis Trial Study Group. *Crit. Care Med.*, *24*, 1431–1440.
49. Dhainaut, J. F., Vincent, J. L., Richard, C., Lejeune, P., Martin, C., Fierobe, L., Stephens, S., Ney, U. M., Sopwith, M. (1995). CDP571, a humanized antibody to human tumor necrosis factor-alpha: Safety, pharmacokinetics, immune response, and influence of the antibody on cytokine concentrations in patients with septic shock. CPD571 Sepsis Study Group. *Crit. Care Med.*, *23*, 1461–1469.
50. Fein, A. M., Bernard, G. R., Criner, G. J., Fletcher, E. C., Good, J. T., Jr., Knaus, W. A., Levy, H., Matuschak, G. M., Shanies, H. M., Taylor, R. W., Rodell, T. C. (1997). Treatment of severe systemic inflammatory response syndrome and sepsis with a novel bradykinin antagonist, deltibant (CP-0127). Results of a randomized, double-blind, placebo-controlled trial. CP-0127 SIRS and Sepsis Study Group. *JAMA*, *277*, 482–487.
51. Fisher, C. J., Jr., Opal, S. M., Dhainaut, J. F., Stephens, S., Zimmerman, J. L., Nightingale, P., Harris, S. J., Schein, R. M., Panacek, E. A., Vincent, J. L., et al. (1993). Influence of an anti-tumor necrosis factor monoclonal antibody on cytokine levels in patients with sepsis. The CB0006 Sepsis Syndrome Study Group. *Crit. Care Med.*, *21*, 318–327.

52. Fisher, C. J., Jr., Dhainaut, J. F., Opal, S. M., Pribble, J. P., Balk, R. A., Slotman, G. J., Iberti, T. J., Rackow, E. C., Shapiro, M. J., Greenman, R. L., et al. (1994). Recombinant human interleukin 1 receptor antagonist in the treatment of patients with sepsis syndrome. Results from a randomized, double-blind, placebo-controlled trial. Phase III rhIL-1ra Sepsis Syndrome Study Group. *JAMA*, *271*, 1836–1843.
53. Fisher, C. J., Jr., Slotman, G. J., Opal, S. M., Pribble, J. P., Bone, R. C., Emmanuel, G., Ng, D., Bloedow, D. C., Catalano, M. A. (1994). Initial evaluation of human recombinant interleukin-1 receptor antagonist in the treatment of sepsis syndrome: A randomized, open-label, placebo-controlled multicenter trial. The IL-1RA Sepsis Syndrome Study Group. *Crit. Care Med.*, *22*, 12–21.
54. Fisher, C. J., Jr., Agosti, J. M., Opal, S. M., Lowry, S. F., Balk, R. A., Sadoff, J. C., Abraham, E., Schein, R. M., Benjamin, E. (1996). Treatment of septic shock with the tumor necrosis factor receptor: Fc fusion protein. The Soluble TNF Receptor Sepsis Study Group. *N. Engl. J. Med.*, *334*, 1697–1702.
55. Haupt, M. T., Jastremski, M. S., Clemmer, T. P., Metz, C. A., Goris, G. B. (1991). Effect of ibuprofen in patients with severe sepsis: A randomized, double-blind, multicenter study. The Ibuprofen Study Group. *Crit. Care Med.*, *19*, 1339–1347.
56. Kay, C. (1996). Paper presented at Cambridge Health Institute's Designing better drugs and clinical trials for sepsis/SIRS: Reducing mortality to patients and suppliers. Washington, DC.
57. Opal, S. M., Fisher, C. J., Jr., Dhainaut, J. F., Vincent, J. L., Brase, R., Lowry, S. F., Sadoff, J. C., Slotman, G. J., Levy, H., Balk, R. A., Shelly, M. P., Pribble, J. P., LaBrecque, J. F., Lookabaugh, J., Donovan, H., Dubin, H., Baughman, R., Norman, J., DeMaria, E., Matzel, K., Abraham, E., Seneff, M. (1997). Confirmatory interleukin-1 receptor antagonist trial in severe sepsis: A phase III, randomized, double-blind, placebo-controlled, multicenter trial. The Interleukin-1 Receptor Antagonist Sepsis Investigator Group. *Crit. Care Med.*, *25*, 1115–1124.
58. Reinhart, K., Wiegand-Lohnert, C., Grimminger, F., Kaul, M., Withington, S., Treacher, D., Eckart, J., Willatts, S., Bouza, C., Krausch, D., Stockenhuber, F., Eiselstein, J., Daum, L., Kempeni, J. (1996). Assessment of the safety and efficacy of the monoclonal anti-tumor necrosis factor antibody-fragment, MAK 195F, in patients with sepsis and septic shock: A multicenter, randomized, placebo-controlled, dose-ranging study. *Crit. Care Med.*, *24*, 733–742.
59. Reinhart, K., Menges, T., Gardlund, B., Harm Zwaveling, J., Smithes, M., Vincent, J. L., Tellado, J. M., Salgado-Remigio, A., Zimlichman, R., Withington, S., Tschaikowsky, K., Brase, R., Damas, P., Kupper, H., Kempeni, J., Eiselstein, J., Kaul, M. (2001). Randomized, placebo-controlled trial of the anti-tumor necrosis factor antibody fragment afelimomab in hyperinflammatory response during severe sepsis: The RAMSES Study. *Crit. Care Med.*, *29*, 765–769.
60. Rodell, T. C., Scharschmidt, L., Knaus, W. A. (1995). CP-0127 SIRS and Sepsis Study Group: Results of a multi-center randomized, placebo-controlled trial of CP-0127, a novel bradykinin antagonist, in patients with SIRS and sepsis. *Shock*, *3*, 60.
61. Abraham, E., Laterre, P. F., Garbino, J., Pingleton, S., Butler, T., Dugernier, T., Margolis, B., Kudsk, K., Zimmerli, W., Anderson, P., Reynaert, M., Lew, D., Lesslauer, W., Passe, S., Cooper, P., Burdeska, A., Modi, M., Leighton, A., Salgo, M., Van der Auwera, P. (2001). Lenercept (p55 tumor necrosis factor receptor fusion

protein) in severe sepsis and early septic shock: A randomized, double-blind, placebo-controlled, multicenter phase III trial with 1,342 patients. *Crit. Care Med.*, *29*, 503–510.

62. Minneci, P. C., Deans, K. J., Banks, S. M., Eichacker, P. Q., Natanson, C. (2004). Should we continue to target the platelet-activating factor pathway in septic patients? *Crit. Care Med.*, *32*, 585–588.
63. Knaus, W. A., Harrell, F. E., Jr., LaBrecque, J. F., Wagner, D. P., Pribble, J. P., Draper, E. A., Fisher, C. J., Jr., Soll, L. (1996). Use of predicted risk of mortality to evaluate the efficacy of anticytokine therapy in sepsis. The rhIL-1ra Phase III Sepsis Syndrome Study Group. *Crit. Care Med.*, *24*, 46–56.
64. Eichacker, P. Q., Parent, C., Kalil, A., Esposito, C., Cui, X., Banks, S. M., Gerstenberger, E. P., Fitz, Y., Danner, R. L., Natanson, C. (2002). Risk and Efficacy of Antiinflammatory Agents: Retrospective and confirmatory studies of sepsis. *Am. J. Respir. Crit. Care Med.*
65. Cooperative Study Group (1963). The effectiveness of hydrocortisone in the management of severe infection. *JAMA*, *183*, 462–465.
66. Fabian, T. C., Patterson, R. (1982). Steroid therapy in septic shock. Survival studies in a laboratory model. *Am. Surg.*, *48*, 614–617.
67. Hinshaw, L. B., Beller-Todd, B. K., Archer, L. T., Benjamin, B., Flournoy, D. J., Passey, R., Wilson, M. F. (1981). Effectiveness of steroid/antibiotic treatment in primates administered LD100 *Escherichia coli*. *Ann. Surg.*, *194*, 51–56.
68. Hinshaw, L. B., Archer, L. T., Beller-Todd, B. K., Benjamin, B., Flournoy, D. J., Passey, R. (1981). Survival of primates in lethal septic shock following delayed treatment with steroid. *Circ. Shock*, *8*, 291–300.
69. Hinshaw, L. B., Beller, B. K., Archer, L. T., Flournoy, D. J., White, G. L., Phillips, R. W. (1979). Recovery from lethal *Escherichia coli* shock in dogs. *Surg. Gynecol. Obstet.*, *149*, 545–553.
70. White, G. L., Archer, L. T., Beller, B. K., Hinshaw, L. B. (1978). Increased survival with methylprednisolone treatment in canine endotoxin shock. *J. Surg. Res.*, *25*, 357–364.
71. Beller, B. K., Archer, L. T., Passey, R. B., Flournoy, D. J., Hinshaw, L. B. (1983). Effectiveness of modified steroid-antibiotic therapies for lethal sepsis in the dog. *Arch. Surg.*, *118*, 1293–1299.
72. Lucas, C. E., Ledgerwood, A. M. (1984). The cardiopulmonary response to massive doses of steroids in patients with septic shock. *Arch. Surg.*, *119*, 537–541.
73. Schumer, W. (1976). Steroids in the treatment of clinical septic shock. *Ann. Surg.*, *184*, 333–341.
74. Klastersky, J., Cappel, R., Debusscher, L. (1971). Effectiveness of betamethasone in management of severe infections. A double-blind study. *N. Engl. J. Med.*, *284*, 1248–1250.
75. Bennett, I. L., Finland, M., Hamborger, M., Kass, E. H., Lepper, M., Waisbren, B. A. (1963). The effectiveness of hydrocortisone in the management of severe infection. *JAMA*, *183*, 462–465.
76. Luce, J. M., Montgomery, A. B., Marks, J. D., Turner, J., Metz, C. A., Murray, J. F. (1988). Ineffectiveness of high-dose methylprednisolone in preventing parenchymal lung injury and improving mortality in patients with septic shock. *Am. Rev. Respir. Dis.*, *138*, 62–68.

77. Veterans Administration Cooperative Study Group (1987). Effect of high-dose glucocorticoid therapy on mortality in patients with clinical signs of systemic sepsis. The Veterans Administration Systemic Sepsis Cooperative Study Group. *N. Engl. J. Med.*, *317*, 659–665.
78. Bone, R. C., Fisher, C. J., Jr., Clemmer, T. P., Slotman, G. J., Metz, C. A., Balk, R. A. (1987). A controlled clinical trial of high-dose methylprednisolone in the treatment of severe sepsis and septic shock. *N. Engl. J. Med.*, *317*, 653–658.
79. Sprung, C. L., Caralis, P. V., Marcial, E. H., Pierce, M., Gelbard, M. A., Long, W. M., Duncan, R. C., Tendler, M. D., Karpf, M. (1984). The effects of high-dose corticosteroids in patients with septic shock. A prospective, controlled study. *N. Engl. J. Med.*, *311*, 1137–1143.
80. Cronin, L., Cook, D. J., Carlet, J., Heyland, D. K., King, D., Lansang, M. A., Fisher, C. J., Jr. (1995). Corticosteroid treatment for sepsis: A critical appraisal and meta-analysis of the literature. *Crit. Care Med.*, *23*, 1430–1439.
81. Lefering, R., Neugebauer, E. A. (1995). Steroid controversy in sepsis and septic shock: A meta-analysis. *Crit. Care Med.*, *23*, 1294–1303.
82. Rothwell, P. M., Udwadia, Z. F., Lawler, P. G. (1991). Cortisol response to corticotropin and survival in septic shock. *Lancet*, *337*, 582–583.
83. Moran, J. L., Chapman, M. J., O'Fathartaigh, M. S., Peisach, A. R., Pannall, P. R., Leppard, P. (1994). Hypocortisolaemia and adrenocortical responsiveness at onset of septic shock. *Intens. Care Med.*, *20*, 489–495.
84. Soni, A., Pepper, G. M., Wyrwinski, P. M., Ramirez, N. E., Simon, R., Pina, T., Gruenspan, H., Vaca, C. E. (1995). Adrenal insufficiency occurring during septic shock: Incidence, outcome, and relationship to peripheral cytokine levels. *Am. J. Med.*, *98*, 266–271.
85. Yildiz, O., Doganay, M., Aygen, B., Guven, M., Keleutimur, F., Tutuu, A. (2002). Physiological-dose steroid therapy in sepsis [ISRCTN36253388]. *Crit. Care*, *6*, 251–259.
86. Annane, D., Sebille, V., Charpentier, C., Bollaert, P. E., Francois, B., Korach, J., Capellier, G., Cohen, Y., Azoulay, E., Troche, G., Chaumet-Riffaut, P., Bellisant, E. (2002). Effect of treatment with low doses of hydrocortisone and fludrocortisone on mortality in patients with septic shock. *JAMA*, *288*, 862–871.
87. Briegel, J., Forst, H., Haller, M., Schelling, G., Kilger, E., Kuprat, G., Hemmer, B., Hummel, T., Lenhart, A., Heyduck, M., Stoll, C., Peter, K. (1999). Stress doses of hydrocortisone reverse hyperdynamic septic shock: A prospective, randomized, double-blind, single-center study. *Crit. Care Med.*, *27*, 723–732.
88. Bollaert, P. E., Charpentier, C., Levy, B., Debouverie, M., Audibert, G., Larcan, A. (1998). Reversal of late septic shock with supraphysiologic doses of hydrocortisone. *Crit. Care Med.*, *26*, 645–650.
89. Chawla, K., Kupfer, Y., Goldman, I., Tessler, S. (1999). Hydrocortisone reverses refractory shock. *Crit. Care Med.*, *27*, A33 (abstract).
90. Minneci, P. C., Deans, K. J., Banks, S. M., Eichacker, P. Q., Natanson, C. (2004). Dose dependent effects of steroids on survival rates and shock during sepsis: A meta-analysis. *Ann. Intern. Med.*
91. Schaub, R. G., Simmons, C. A., Koets, M. H., Romano, P. J., 2nd, Stewart, G. J. (1984). Early events in the formation of a venous thrombus following local trauma and stasis. *Lab. Invest.*, *51*, 218–224.

92. Coughlin, S. R. (1994). Thrombin receptor function and cardiovascular disease. *Trends Cardiovasc. Med.*, *4*, 77–83.

93. Lorant, D. E., Patel, K. D., McIntyre, T. M., McEver, R. P., Prescott, S. M., Zimmerman, G. A. (1991). Coexpression of GMP-140 and PAF by endothelium stimulated by histamine or thrombin: A juxtacrine system for adhesion and activation of neutrophils. *J. Cell. Biol.*, *115*, 223–234.

94. Freeman, B. D., Zehnbauer, B. A., Buchman, T. G. (2003). A meta-analysis of controlled trials of anti-coagulant therapies in patients with sepsis. *Shock*, *20*, 5–9.

95. Kidokoro, A., Iba, T., Fukunaga, M., Yagi, Y. (1996). Alterations in coagulation and fibrinolysis during sepsis. *Shock*, *5*, 223–228.

96. Bernard, G. R., Vincent, J. L., Laterre, P. F., LaRosa, S. P., Dhainaut, J. F., Lopez-Rodriguez, A., Steingrub, J. S., Garber, G. E., Helterbrand, J. D., Ely, E. W., Fisher, C. J., Jr. (2001). Efficacy and safety of recombinant human activated protein C for severe sepsis. *N. Engl. J. Med.*, *344*, 699–709.

97. Warren, B. L., Eid, A., Singer, P., Pillay, S. S., Carl, P., Novak, I., Chalupa, P., Atherstone, A., Penzes, I., Kubler, A., Knaub, S., Keinecke, H. O., Heinrichs, H., Schindel, F., Juers, M., Bone, R. C., Opal, S. M. (2001). Caring for the critically ill patient. High-dose antithrombin III in severe sepsis: A randomized controlled trial. *JAMA*, *286*, 1869–1878.

98. Inthorn, D., Hoffmann, J. N., Hartl, W. H., Muhlbayer, D., Jochum, M. (1997). Antithrombin III supplementation in severe sepsis: Beneficial effects on organ dysfunction. *Shock*, *8*, 328–334.

99. Fourrier, F., Chopin, C., Huart, J. J., Runge, I., Caron, C., Goudemand, J. (1993). Double-blind, placebo-controlled trial of antithrombin III concentrates in septic shock with disseminated intravascular coagulation. *Chest*, *104*, 882–888.

100. Balk, R. A., Bedrosian, C. L. M., Baughman, R., Eisele, B., Keinecke, H. O., Bone, R. C. (1995). Prospective, double-blind, placebo-controlled, trial of ATIII substitution in sepsis. *Intens. Care Med.*, *21*, S17.

101. Baudo, F., Caimi, T. M., de Cataldo, F., Ravizza, A., Arlati, S., Casella, G., Carugo, D., Palareti, G., Legnani, C., Ridolfi, L., Rossi, R., D'Angelo, A., Crippa, L., Giudici, D., Gallioli, G., Wolfler, A., Calori, G. (1998). Antithrombin III (ATIII) replacement therapy in patients with sepsis and/or postsurgical complications: A controlled double-blind, randomized, multicenter study. *Intens. Care Med.*, *24*, 336–342.

102. Schuster, H. P., Eisele, B., Keinecke, H. O., Heinrichs, H., Mescheder, A., Knaub, S. (1997). S-AT III Study: Antithrombin III in patients with sepsis. *Intens. Care Med.*, *23*(Suppl. 1), S76.

103. Eisele, B., Lamy, M., Thijs, L. G., Keinecke, H. O., Schuster, H. P., Matthias, F. R., Fourrier, F., Heinrichs, H., Delvos, U. (1998). Antithrombin III in patients with severe sepsis. A randomized, placebo-controlled, double-blind multicenter trial plus a meta-analysis on all randomized, placebo-controlled, double-blind trials with antithrombin III in severe sepsis. *Intens. Care Med.*, *24*, 663–672.

104. Hartment, D. L., Bernard, G. R., Rosenfeld, B. A., Helterbrand, J. D., Yan, S. B., Fisher, C. J., Jr. (1998). Recombinant human activated protein C (rhAPC) improves coagulation abnormailites associated with severe sepsis. *Intens. Care Med.*, *24*(Suppl. 1), S77.

105. Abraham, E., Reinhart, K., Svoboda, P., Seibert, A., Olthoff, D., Dal Nogare, A., Postier, R., Hempelmann, G., Butler, T., Martin, E., Zwingelstein, C., Percell, S., Shu, V., Leighton, A., Creasey, A. A. (2001). Assessment of the safety of recombinant tissue factor pathway inhibitor in patients with severe sepsis: A multicenter, randomized, placebo-controlled, single-blind, dose escalation study. *Crit. Care Med.*, *29*, 2081–2089.
106. Abraham, E., Reinhart, K., Opal, S., Demeyer, I., Doig, C., Rodriguez, A. L., Beale, R., Svoboda, P., Laterre, P. F., Simon, S., Light, B., Spapen, H., Stone, J., Seibert, A., Peckelsen, C., De Deyne, C., Postier, R., Pettila, V., Artigas, A., Percell, S. R., Shu, V., Zwingelstein, C., Tobias, J., Poole, L., Stolzenbach, J. C., Creasey, A. A. (2003). Efficacy and safety of tifacogin (recombinant tissue factor pathway inhibitor) in severe sepsis: A randomized controlled trial. *JAMA*, *290*, 238–247.
107. Diaz-Cremades, J. M., Lorenzo, R., Sanchez, M., Moreno, M. J., Alsar, M. J., Bosch, J. M., Fajardo, L., Gonzalez, D., Guerrero, D. (1994). Use of antithrombin III in critical patients. *Intens. Care Med.*, *20*, 577–580.
108. Haley, M., Cui, X., Minneci, P. C., Deans, K. J., Natanson, C., Eichacker, P. Q. (2004). Recombinant human activated protein C in sepsis: Assessing its clinical use. *Am. J. Med. Sci.*, *324*, 215–219.
109. Davidson, B. L., Geerts, W. H., Lensing, A. W. (2002). Low-dose heparin for severe sepsis. *N. Engl. J. Med.*, *347*, 1036–1037.
110. Goldhaber, S. Z. (2003). PREVENT (PRospective EValuation of Dalteparin Efficacy in Immobilized PatieNts Trial). Presented at the Society of Critical Care Medicine Critical Care Congress in San Antonio, TX.
111. Sevransky, J., Natanson, C. (1999). Published clinical trials in sepsis: An update. *Sepsis*, *2*, 11–19.
112. Thompson, W. L., Gurley, H. T., Lutz, B. A., Jackson, D. L., Kyols, L. K., Morris, I. A. (1976). Inefficacy of glucocorticoids in shock (double-blind study). *Clin. Res.*, *24*, 258A (abstract).

# COMBINATORIAL CHEMISTRY IN THE DRUG DISCOVERY PROCESS

NATHAN T. ROSS, BRIAN R. MCNAUGHTON, AND BENJAMIN L. MILLER
*University of Rochester*
*Rochester, New York*

1 INTRODUCTION TO COMBINATORIAL CHEMISTRY
Introduction
Origins of Combinatorial Chemistry
Solid-Phase Synthesis of Biopolymers
2 TYPES OF LIBRARY SYNTHESIS
Early Examples of Parallel Synthesis
Split-Pool Synthesis
On-Bead Screening
3 WHO WON? IDENTIFYING HITS FROM SCREENS OF SPLIT-POOL LIBRARIES
Recursive Deconvolution
Tagging Methods
Binary Encoding
Infrared Coded Resins
Radio-Frequency Tagging and the Irori Corporation
4 OTHER COMBINATORIAL SYNTHETIC TECHNIQUES
Photolithographic Synthetic Methods
High-Throughput Methods
5 TAG-FREE METHODOLOGY
Dynamic Combinatorial Chemistry and Allied Methods
Click Chemistry and in Situ Click Chemistry
Thiol-Based in Situ Assembly of Macromolecular Ligands
Structure–Activity Relationship by Nuclear Magnetic Resonance

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

6 LIBRARY DESIGN ISSUES
7 CONCLUSIONS
References

## 1 INTRODUCTION TO COMBINATORIAL CHEMISTRY

### Introduction

The global pharmaceutical industry stands as a testament to the power of fundamental research in chemistry and biology and the impact such efforts can have on human health and human history. Many of the medicines we now take for granted—from penicillin (**1**) and quinine (**2**) to more modern drugs such as Taxol (**3**) and vancomycin (**4**) are the products of drug discovery efforts rooted in "traditional" medicinal chemistry. As discussed in detail throughout this text, such efforts arose initially based on a desire to understand and improve the physiological activity of natural products: compounds isolated from natural sources such as fungi, plants, and marine sponges. A highly simplified (or perhaps simplistic) view of traditional medicinal chemistry is that it is a linear endeavor: A compound is synthesized, or a natural product modified, based on an initial hypothesis of activity. After carrying out a bioassay of the new compound, the hypothesis is modified, another new compound is synthesized, and so on until a structure with the desired properties (receptor affinity, binding selectivity, and pharmacokinetic properties) is finally obtained. In many cases, this can be an exceptionally long process: The road from initial isolation to approval of the Food and Drug Administration (FDA) for Taxol was roughly 30 years [1]! Without exception, it is a remarkably expensive and uncertain process. As has been noted by others: "For every 10,000 or more compounds synthesized each year in an average pharmaceutical company, less than one makes it to the market. Any method that allows the pharmaceutical chemist to increase the likelihood of synthesizing an active analogue or to increase his ability to find, or even design, novel leads is of enormous commercial interest" [2].

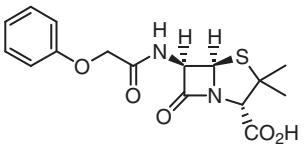


**1**

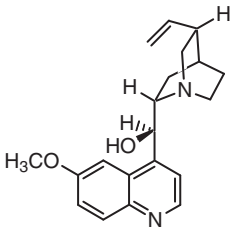


**2**

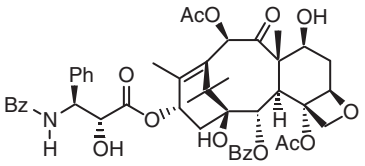


3

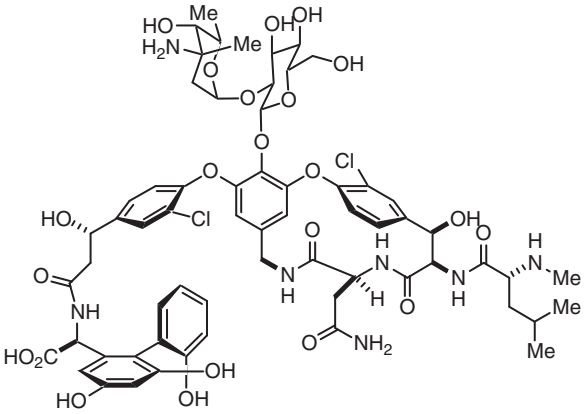


4

Although this statement was written in an account describing research on computer-aided drug design, it can easily be applied to combinatorial chemistry. Broken down into segments, the drug discovery process typically involves target identification, target validation, lead identification, and lead optimization [3]. In a little over a decade, combinatorial chemistry has evolved as an attempt to significantly shorten the lead identification and lead optimization portions of drug development. Although most precisely defined as the exhaustive recombination of sets of reagents (following the mathematical definition of combinatorial), combinatorial chemistry has become synonymous with any process that allows for the synthesis of arbitrarily large sets of different compounds, known as libraries, simultaneously.

## Origins of Combinatorial Chemistry

Combinatorial chemistry largely arose out of two paradigm-altering ideas: solid-supported synthesis and molecular biology. Molecular biology has been critical both because of the significantly greater range of targets it has made available and because of the recognition that much of nature is in essence a "combinatorial chemistry experiment." In addition to obvious parallels at the genetic level, combinatorial strategies can even be seen in naturally occurring small-molecule "libraries" of biologically active compounds. For example, the pupa of the squash beetle *Epilachna borealis* secretes a combinatorial library of macrocyclic polyamines as a defensive strategy (Scheme 1) [4]. Solid-supported synthesis, developed initially as a tool primarily for biochemists and molecular biologists, has become ubiquitous in the field. Because of this, we will describe it in some detail. Several other parallel developments, including high-throughput screening, automation (robotics), and computer-aided molecular design, have also significantly impacted the development of

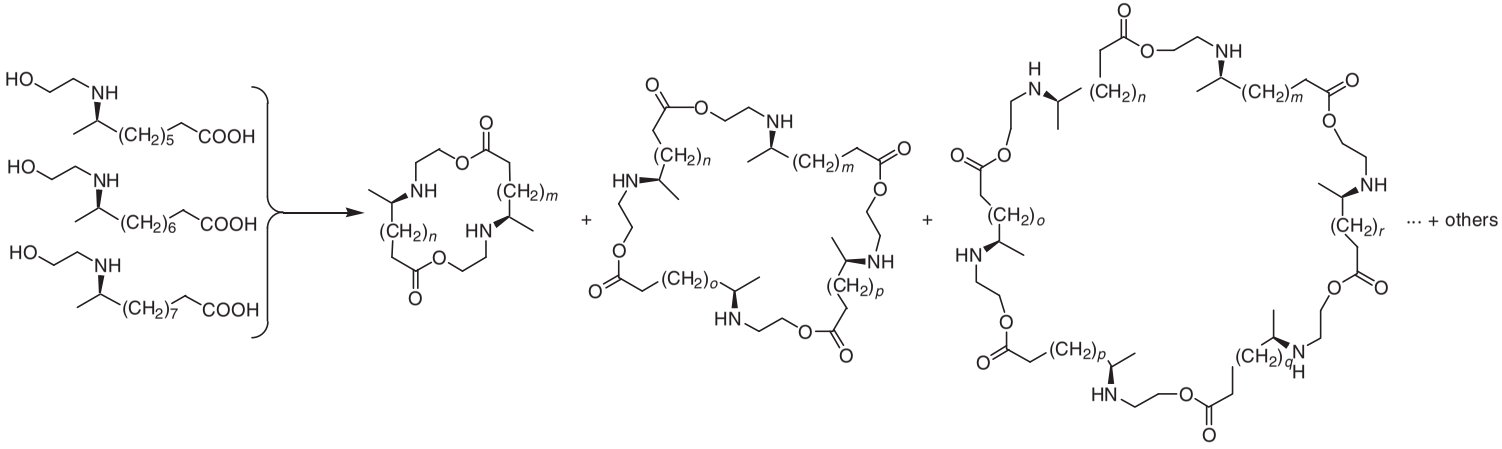


**Scheme 1** Beetle-synthesized combinatorial library of defensive compounds.

combinatorial chemistry. We should note, however, that this review is by no means intended to be comprehensive. The combinatorial chemistry literature has grown explosively over the past few years, and we have undoubtedly neglected many important contributions. We refer those readers interested in more detailed coverage of the field to the many monographs now available [5], as well as to the primary literature.

## Solid-Phase Synthesis of Biopolymers

Solid-phase synthesis emerged from pioneering efforts by Merrifield and co-workers in the 1960s targeted at simplifying the synthesis of longer peptides (four or more amino acids at that time) [6]. Merrifield's work was driven by two primary assumptions. First, attachment of a peptide to a solid support during synthesis was anticipated to significantly improve reaction yields, since excess reagents could be used to drive reactions to completion. Second, unreacted materials remaining in solution at the end of each synthetic step could be subsequently filtered away, simplifying isolation of the final product. In execution, an amino acid with a protected amino terminus was first covalently attached to functionalized polystyrene resin via its carboxy terminus (Scheme 2). Next, the amino terminus was deprotected and reacted with the carboxy terminus of a second amino acid, forming a peptide bond. After the last amino acid was attached, treatment of the resin with a strong acid cleaved the peptide from the polymer matrix. Merrifield showed that this process was highly reproducible, and when the quantities of reactants were optimized, very high reaction yields could be obtained.

Merrifield found that making this process effective requires first selecting a resin made from a functionalized, inert chemical polymer. Although several materials were initially tested, polystyrene cross-linked with divinylbenzene proved to be the most advantageous. This was due to its insolubility in organic and aqueous solvents, its ability to be functionalized via standard aromatic chemistry, and its porous nature (**5**). Polystyrene resin beads are also able to swell to different degrees in different organic solvents, allowing for some control of reagent penetration into the bead depending on the specific solvent employed [7].

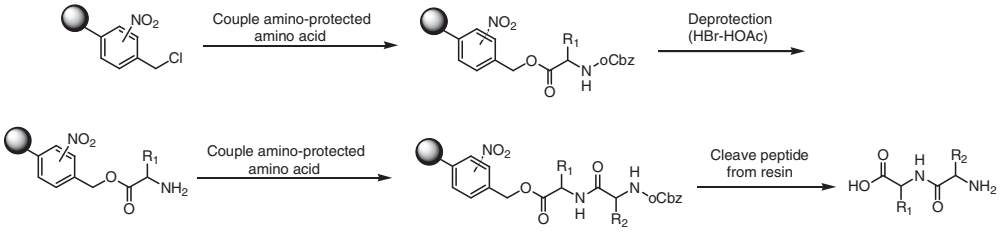


**Scheme 2** Solid-phase peptide synthesis [6].

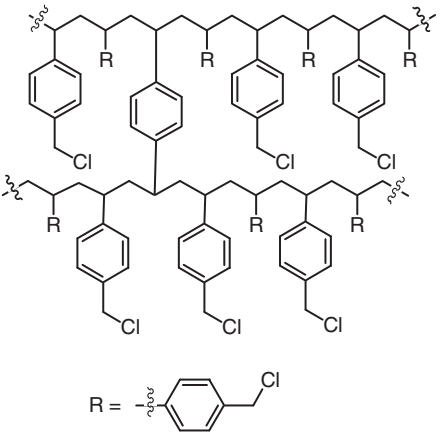


**5**

The resin most closely associated with Merrifield's work is chloromethyl polystyrene-co-divinylbenzene. Providing an initial proof of concept, Merrifield and co-workers synthesized the tetrapeptide Leu-Ala-Gly-Val, using a scheme analogous to that shown in Scheme 2. Following cleavage of the peptide from resin, analysis proved that it was in fact the desired product. Moreover, the purity of the crude material obtained following resin cleavage was high enough to indicate that couplings were very nearly quantitative. These initial results validated Merrifield's hypotheses and set the stage for more complex syntheses.

As effective and efficient as Merrifield's new solid-phase synthesis was, it was not completely without problems. The conditions required to cleave the carbobenzoxy (Cbz) protecting group, HBr-HOAc, were so acidic that some of the peptide was simultaneously cleaved from the resin beads. Initial attempts to circumvent this problem involved modification of the chloromethyl resin by nitration or bromination to increase its acid stability. When this still failed to completely prevent cleavage of material from the resin during deprotection, subsequent syntheses substituted the significantly more acid-labile *t*-butoxycarbonyl (*t*-BOC) group for Cbz. Use of the *t*-BOC protecting group allowed the Merrifield group to synthesize a nonapeptide, the plasma kinin bradykinin (**6**), in just eight days and in an overall yield of 68 percent [8]. This synthesis represented an important advance in the field for several reasons. First, it incorporated several modifications to the initial protocols, which both streamlined the synthetic process and made it more compatible with the synthesis of longer peptides. Second, the bradykinin synthesis demonstrated that the solid-phase methodology was compatible with reactive side chains (in this case, serine and arginine). However, what is perhaps most significant about the synthesis of bradykinin was that it was the first small molecule synthesized on solid-phase resin using automation [9]. The ability to automate solid-phase synthesis was one of the first goals that Merrifield set

for his laboratory at the outset of their work, and it had been accomplished here.

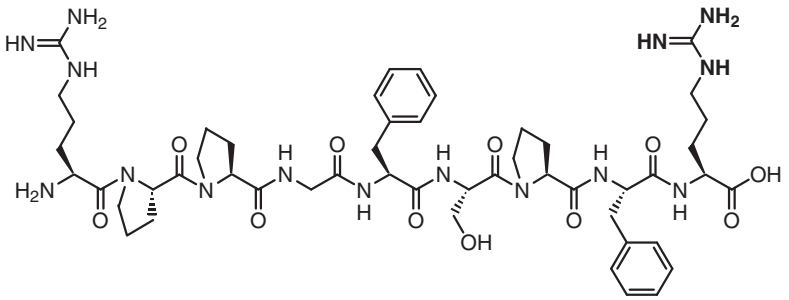

L-ARG-L-PRO-L-PRO-GLY-L-PHE-L-SER-L-PRO-L-PHE-L-ARG

Bradykinin, (**6**)

Another landmark synthesis from the Merrifield laboratory was that of valinomycin (**7**), a cyclic dodecadepsipeptide derived from D/L valine, D-α-hydroxyisovaleric acid, and L-lactic acid. Representing a significant leap in complexity over the tetrapeptide structure described just 5 years previously, the synthesis of valinomycin featured solution-phase production of four building blocks via esterification reactions. These were subsequently incorporated into the solid-phase synthesis as in Merrifield's previous efforts.

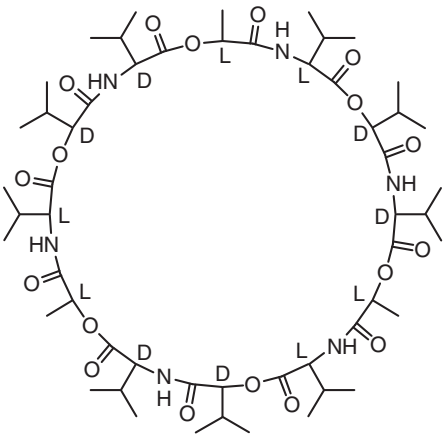

Valinomycin (**7**)

It only took a few years after Merrifield's first publication describing solid-phase synthesis of peptides for others to apply the new methodology to their fields of study. In 1965, Letsinger and Mahadevan (who had also described an alternative strategy for solid-phase peptide synthesis shortly after Merrifield) published a report describing the solid-phase synthesis of deoxycytidylyl-

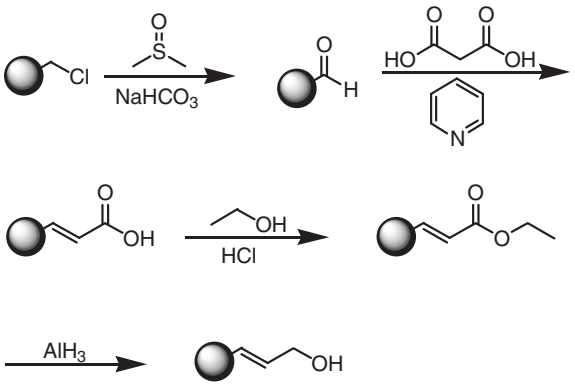


**Scheme 3**

(3′-5′)-thymidine on polystyrene resin [10]. Much like peptide solid-phase synthesis, purification during synthesis was much easier than traditional solution-phase methods. However, unlike Merrifield's peptide synthesis, oligonucleotide synthesis was not very rapid due to more than one step in the synthesis requiring over 2 days of reaction time. Subsequent advances by other authors (vide infra) have enabled solid-phase oligonucleotide synthesis to become the highly optimized, highly mechanized, and inexpensive process it is today.

Several years later, in 1971, Frechet and Schuerch developed a protocol to modify Merrifield's chloromethylated polystyrene resin (Scheme 3), making the first solid-phase synthesis of an oligosaccharide possible [11]. Oligosaccharides present a significantly greater synthetic challenge than oligonucleotide or polypeptide synthesis, in part due to the elaborate protecting schemes that are required to direct reactions to the appropriate hydroxyl group. First, the sugar monomers must have a "stable" protecting group on all of the hydroxyl groups that will not be participating in forming the new glycosidic bond. Second, the hydroxyl that is going to be participating in the new glycosidic bond must have a comparatively labile protecting group, which can be easily cleaved without cleaving the stable protecting group or cleaving the saccharide from the resin itself. This last reason is why Frechet and Schuerch [11] had to modifiy Merrifield's chloromethylated resin to make it stable to conditions that would remove both the "labile" and "stable" protecting groups. Scheme 4 outlines the general procedure for the synthesis of oligosaccharides on solid phase. Solid- and solution-phase oligosaccharide synthesis has continued to be an area of significant interest. Notable advances include work from the Danishefsky [12], Wong [13], and Seeberger [14] groups. Seeberger in particular has been successful in adapting standard peptide synthesis instrumentation to the production of complex oligosaccharides.

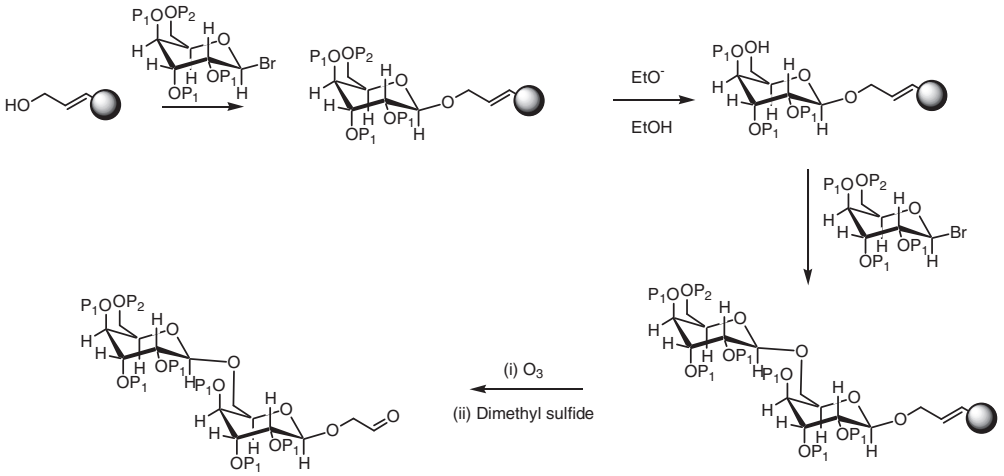


**Scheme 4** Solid-phase disaccharide synthesis [11].

On the shoulders of the pioneering work just described, researchers have used solid-phase chemistry to advance research in fields spanning organic synthesis and chemical biology. A particularly notable extension of Merrifield's peptide synthesis was reported by Kent and co-workers in their construction of backbone engineered human immunodeficiency virus (HIV) protease on solid support [15]. Protein engineering via site-directed mutagenesis is a standard, widely used technique. However, such modifications are necessarily limited to the incorporation of genetically coded (natural) amino acids. Cellular methods for the incorporation of nonnatural amino acids into proteins exist but are generally highly labor intensive [16]. Conversely, solid-phase synthesis provides a means to incorporate nonnatural amino acids, β-amino acids, peptoids (oligomers of *N*-alkyl glycine), and fully nonpeptide segments into the growing peptide chain.

In the Kent group's approach, HIV-1 protease was prepared through the ligation of synthetic peptide segments, corresponding to the two halves of the HIV-1 protease monomer [15]. The resulting active HIV-1 protease was fused through a thioester replacement at the natural peptide bond $Gly^{51}$–$Gly^{52}$, a region sensitive to mutational changes (Scheme 5). Furthermore, this chemospecific ligation allowed fusion of the two peptides without the use of side-chain protective groups.

Along similar lines, the development and advancement of solid supported oligonucleotide synthesis has provided researchers a vehicle to synthesize and perturb desired deoxyribonucleic acid (DNA) sequences in a facile manner. In their landmark report, Khorana and co-workers cited the use of solid support nucleotide synthesis to prepare a 77 base-pair gene, corresponding to yeast alanine transfer ribonucleic acid (tRNA) [17]. The synthetic scheme

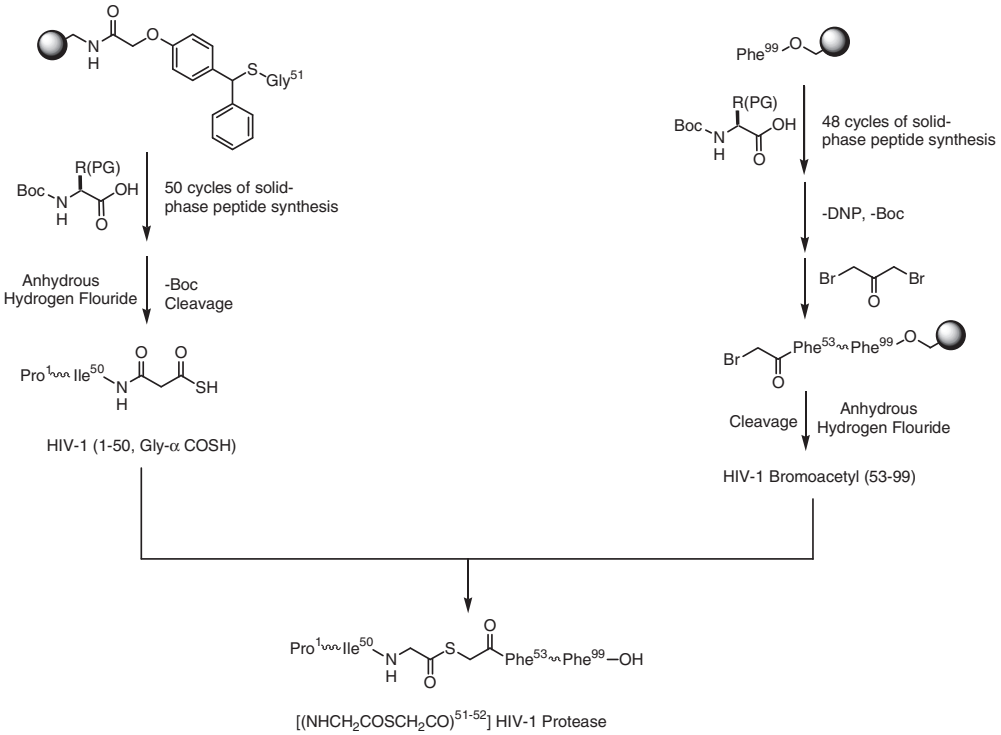


**Scheme 5** Solid-phase synthesis of backbone-engineered HIV-1 protease.

called for solid support chemical synthesis of 15 oligonucleotide segments, containing 3′ and 5′ hydroxyl groups terminating each end, respectively. Phosphorylation of the 5′ hydroxyl group was then performed with the use of $T_4$ polynucleotide kinase, followed by annealing and ligase-induced head-to-tail joining of the appropriate segments, as a function of bihelical complexation. Khorana's group followed this work with the synthesis of a biologically functional *Escherichia coli* tyrosine tRNA suppressor gene, consisting of 202 base pairs [18]. Since these reports, numerous other groups have continued research pertaining to the synthesis of biologically functional genes including human insulin A and B chains [19] and the human leudocyte $\alpha_1$-interferon gene, containing 514 base pairs [20].

In parallel with the development of solid-phase methods for biopolymer synthesis, a small but dedicated cadre of researchers pursued the solid-phase synthesis of more traditionally "organic" molecules. Because of space limitations, we will not discuss these developments in detail. Reviews of this work are recommended, however, since they set the stage for the explosion in solid-phase organic synthetic methods that occurred later [21, 22].

## 2 TYPES OF LIBRARY SYNTHESIS

### Early Examples of Parallel Synthesis

Before the advent of combinatorial chemistry, there existed no means to rapidly generate a wide variety of small molecules with the potential to bind molecular targets. The bulk of druglike molecules were natural products or synthetic analogs, which typically required laborious efforts to synthesize or derive from undifferentiated mixtures. In the late 1970s, the advances in solid-phase synthesis led to the idea that one might be able to make large numbers of compounds simultaneously. If one could also then speed up the screening process, this might provide a completely new paradigm for compound production and analysis. One of the first examples of this new approach came in the early 1980s, when Richard Houghten used his "tea bag" method to make over 500 peptides in a matter of days [23]. Houghton's mesh pouch, which resembled a tea bag (Fig. 1), was dipped in solutions of amino acid and coupling reagents. The pouches were transferred from one solution to the next until the desired peptides were made. At roughly the same time, Ronald Frank and his colleagues were working on developing a method to combinatorially synthesize oligonucleotides on cellulose disks, by a technique that they called the filter disk method [24]. After initial functionalization of the cellulose, nucleotides were sequentially added by dipping the disks in reaction vials con-

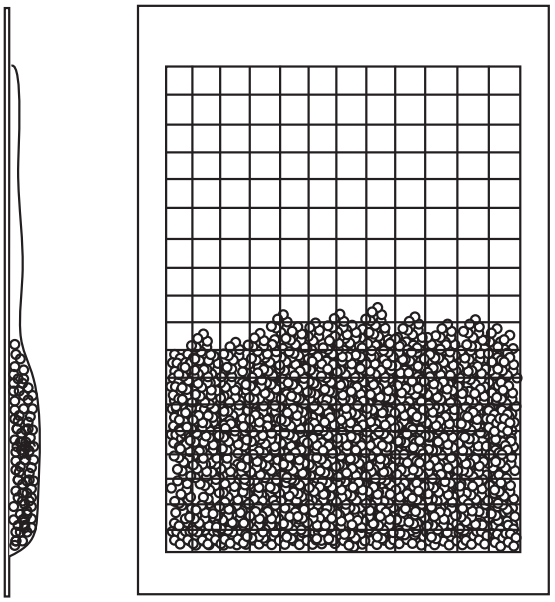

**Figure 1** Houghten "tea bag".

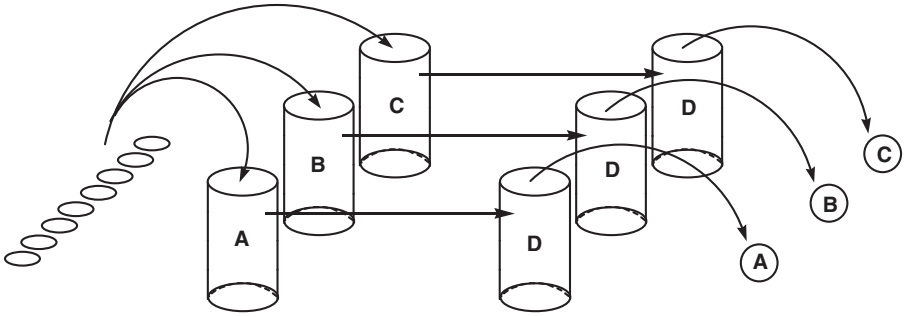


**Figure 2** Filter disk method. Individual cellulose disks are separately treated with reagents (A, B, C) in reaction vessels and may undergo subsequent reaction under identical conditions (D) or undergo further diversification.

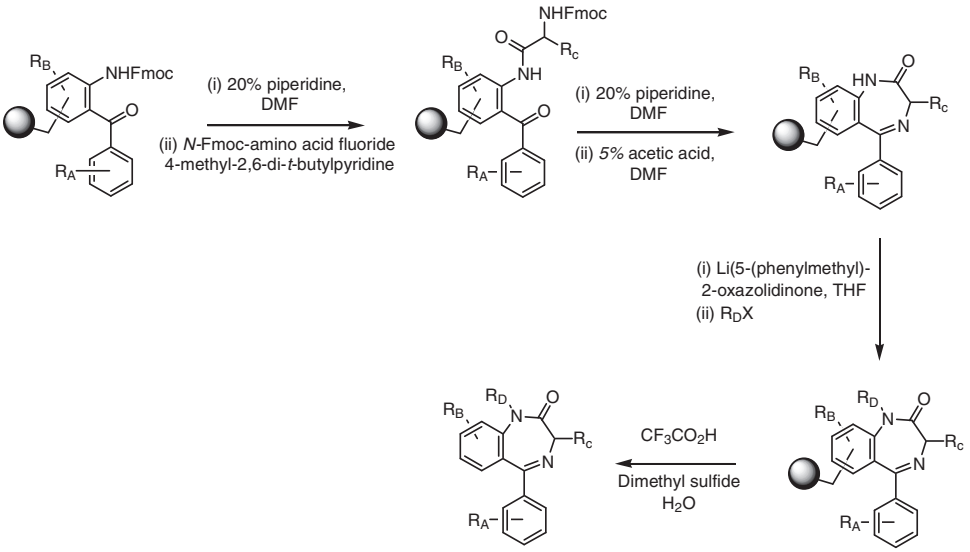


**Scheme 6** 1,4-Benzodiazepine synthesis [25].

taining the desired nucleotide (Fig. 2). Both of these methods fall into the category of parallel synthesis since mixing of reaction products was not possible. The solid support is also critical here: Although highly parallel solution-phase reactions have emerged as an important area of combinatorial synthesis, most early efforts (and a significant fraction of the current art) relied on such solid-supported synthesis. Parallel synthesis of "druglike" small molecules was then a logical extension of these efforts. One relatively early report, by Bunin and Ellman, provided an indication of the potential power of this methodology via the solid-phase synthesis of a series of 1,4-benzodiazepines (Scheme 6) [25].

## Split-Pool Synthesis

As one might imagine, the complexity and logistical overhead associated with parallel synthesis increases significantly as the size of the target library increases. For the production of very large (up to several million compounds) libraries, the split-pool approach (also known as "one bead, one compound") is frequently used. Split-pool synthesis, developed in parallel by the Furka and Lam laboratories, relies (as its name implies) on the ability to iteratively *split* resin into separate reaction vessels, perform a desired synthetic step, then *mix* the beads together again following completion of the reaction [26]. Split-pool is a "true combinatorial" technique in the mathematical sense of the word, in that exhaustive recombination of components takes place. As shown schematically in Figure 3 for the production of a simple 27-compound library via three synthetic steps, one starts with a solid support, usually polystyrene or polystyrene-graft-polyethylene glycol (PEG) resin beads, and splits them into a predetermined number of reaction vessels. The number of vessels corresponds to the number of different ways that the library will be diversified at the first synthetic step. After the first reaction, in which each vessel is treated with a different reagent, the pools are brought back together. The beads are then mixed as efficiently as possible, and split into equivalent groups of beads (or reaction vessels) again. The number of pools again corresponds to the number of different couplings to be done at that synthetic step. Each of these pools gets a unique residue coupled to it, and they are mixed back together into one large pool. This process is repeated until the synthesis of the library is complete. Each "split" step in the synthesis is often described as a *diversification step*.

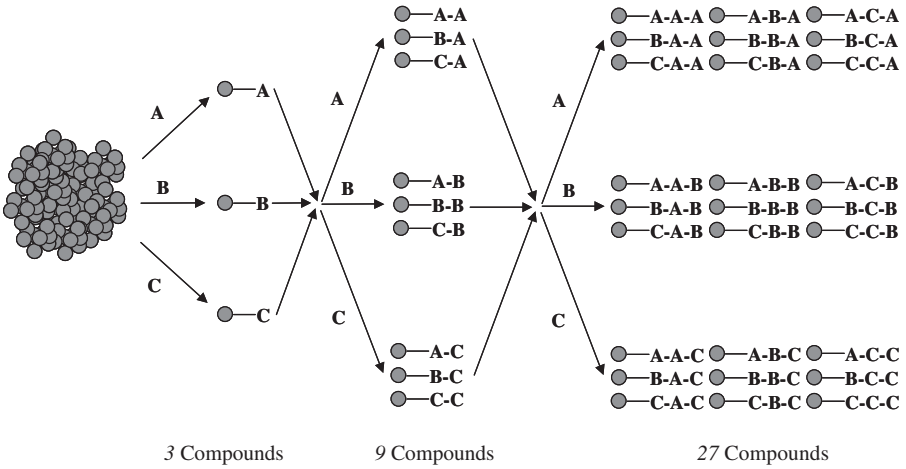


**Figure 3** Split-pool synthesis of a 27-compound library on resin beads, via iterative reaction with reagents A, B, and C.

Split-pool synthesis is a powerful and efficient tool for creating arbitrarily large libraries quickly. Two important features of this synthetic strategy deserve mention. First, an attractive aspect of the split-pool approach is that each individual resin bead carries many copies of a single compound, and not a mixture of different compounds (assuming all reactions go to completion with no side products). This allows for on-bead screening to be used to determine if a ligand will bind to a given target, and circumvents the need for analyzing and purifying thousands (or millions) of compounds before screening. Second, split-pool methodology, with rare exceptions (see description of radio frequency tagging below), does not provide final compounds in a format that allows for purification prior to screening. Furthermore, following a split-pool scheme, each bead usually does not carry a readily discernible history with it, requiring a secondary method for identification of the compound it carries.

### On-Bead Screening

Once one has synthesized a library on solid phase using either multiple parallel or split-mix methods, the next challenge is to identify the compound (or compounds) in the library that have the desired receptor-binding or enzyme-inhibiting activity. For libraries derived from multiple parallel synthesis, this typically involves cleavage of the compound from the bead and screening using one of several high-throughput techniques. For split-pool libraries, cleavage of individual compounds from millions of beads is generally not practical (although there have been some efforts in that direction; vide infra). In these cases, on-bead screening is generally employed. Shown schematically in Figure 4, on-bead screening involves treatment of the full library with a solution of a labeled target receptor. After an incubation period followed by rinsing away excess receptor, visualization according to the particular label used (e.g., fluorescence for a rhodamine-tagged receptor, or treatment with streptavidin-alkaline phosphatase for a biotin-tagged receptor) allows for those beads binding the target to be readily visualized. Of course, control experiments (analogous to the "blocking" step of a Western blot) are essential to provide confidence in the "hit".

## 3 WHO WON? IDENTIFYING HITS FROM SCREENS OF SPLIT-POOL LIBRARIES

Once a hit is identified from a split-pool library, how is the structure of the compound determined? As we have already described, individual beads generally do not carry their synthetic history with them. For oligonucleotide and polypeptide libraries, this is not a problem. Oligonucleotides can be identified by subjecting individual resin beads to polymerase chain reaction (PCR)

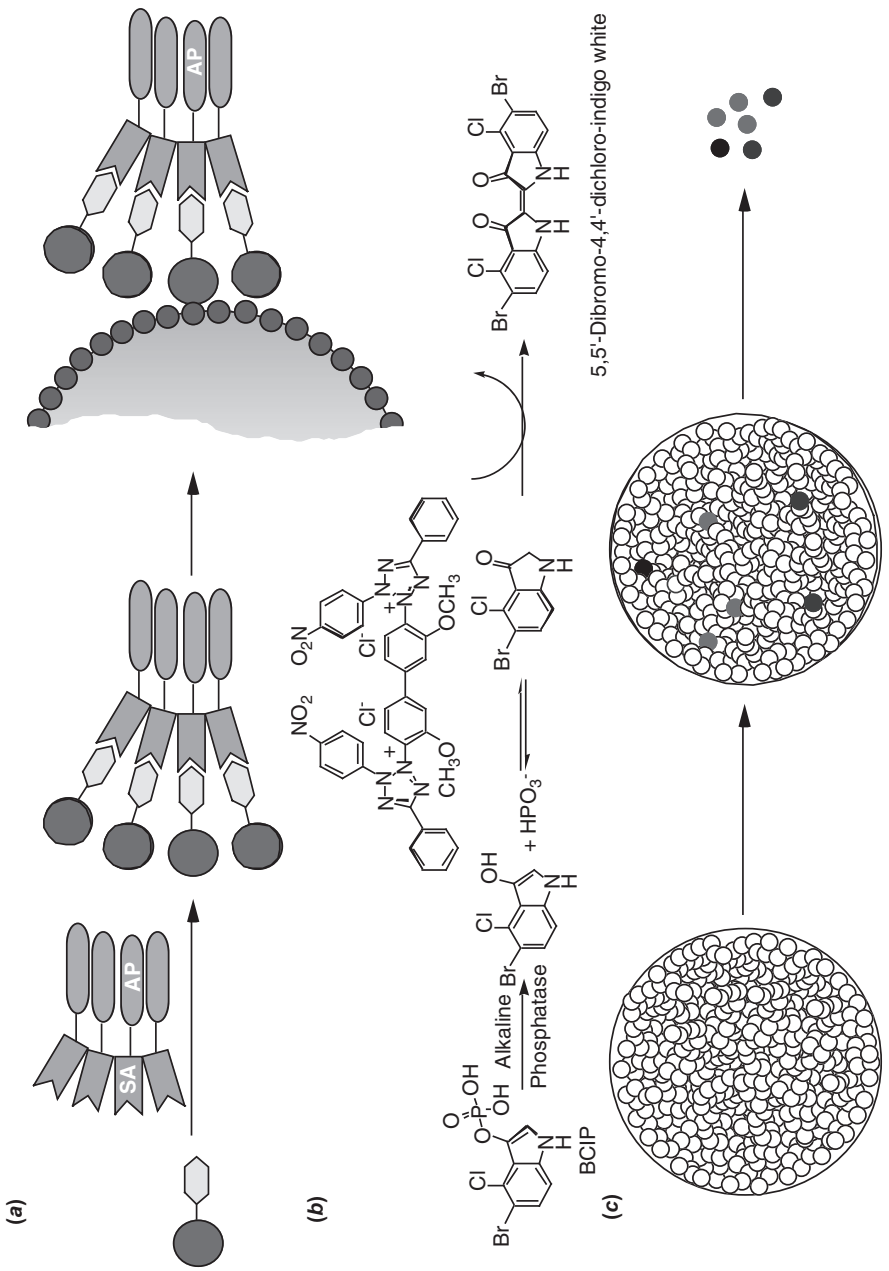


**Figure 4** On-bead screening. (*a*) Treatment of a biotin-tagged protein with tetrameric streptavidin–alkaline phosphatase, followed by interaction with a resin bead. (*b*) Visualization reaction catalyzed by alkaline phosphatase, leading to precipitation of a purple dye on the target bead. (*c*) Staining of hits following deposition of alkaline phosphatase reaction product.

amplification, and often polypeptides can be identified by standard sequencing methods (i.e., Edman degradation) carried out on single beads [27] or via ladder synthesis coupled with mass spectrometry [28]. In rare instances, direct identification of "active" library members from single beads by mass spectrometry is possible, particularly if some reagents incorporate specific isotopes designed to enhance their identification [29, 30]. However, most often it is necessary to "deconvolute" large libraries via an iterative procedure or incorporate a readily identifiable surrogate "tag" for each reagent into every step of the synthesis.

## Recursive Deconvolution

One of the earliest methods for determining the identity of "active" compounds from mixtures of synthetic library members, termed recursive deconvolution, was first described by Sydney Brenner and co-workers in November 1994 [31]. Recursive deconvolution actually starts during library synthesis and goes hand in hand with split-pool synthesis. After doing the first split and coupling of the split-pool synthesis, an aliquot of each batch of resin is removed and set aside for later deconvolution. The nonaliquotted beads are carried on through the next diversification step of the split-pool synthesis. After the second coupling, aliquots from each batch are again taken aside and saved. This process is continued throughout the synthesis, up until the last varied position (also known as a diversification step). At the last varied position the resin is split into individual batches for coupling just as before. However, after this coupling has taken place, the resin in not pooled back together. The rest of the synthesis is carried out in individual batches. On-bead screening is conducted on the individual batches of resin corresponding to the residue at the last varied position, and the batch yielding the most hits is determined to have the best residue at the last varied position. That "best" residue is then coupled to the resin that was aliquotted out after the coupling of the residue just prior to the last varied residue. The synthesis of the molecules is completed in separate batches again. However, this time the batches correspond to the individual residues at the second to last varied position. Screening is again done on-bead, and the batch that yields the most hits is determined to be the one with the best residue at the second to last varied residue. This process continues until all of the varied positions are deconvoluted.

Although recursive deconvolution requires a significant investment of time and material, it still possesses several advantages over other identification methods. For instance, no chemical entities are coupled to the solid-phase resin other than those that make up the library members, eliminating the possibility for chemical tags to produce false positives or false negatives in screening. Also, recursive deconvolution does not require library analysis by high-performance liquid chromatography (HPLC), ultraviolet (UV), infrared (IR), or mass spectrometry.

## Tagging Methods

If one does not wish to use a deconvolution strategy, an alternative is to add an additional, readily detectable reagent at each step of the synthesis that tags each reagent used [32]. Initial examples of this strategy included tagging peptide libraries with oligonucleotides [33, 34]. After selection of a bead as "active," PCR amplification of the oligonucleotide on the bead provided the code corresponding to the peptide. However, for a variety of reasons (including reagent cost and compatibility), other tagging methods were sought.

## Binary Encoding

One of the most innovative tagging protocols was first described by Still and co-workers in 1993 [35, 36]. In their initial report, a series of photolabile, gas chromatograph detectable, chemical tags was employed to deduce amino acid sequences in a 117,649-member peptide-based library. Still's method operates on a binary basis where, for example, three tags can be used to specifically designate one of seven peptide building blocks. For instance, a binary library readout of 101 would indicate tag 1 was present at position 1, tag 2 was not present at position 2, and tag 3 was present at position 3. Three binary positions, therefore, can account for seven unique library members (001, 010, 011, 100, 101, 110, 111). In their initial report, Still and co-workers used 18 tags ($3 \times 6$) to identify 117,649 individual library members [35, 36]. To do this, at each step in the synthesis a series of three unique tags was coupled to less than 1 percent of the total attachment sites on each bead of resin. The residue encoded by those three tags was then coupled to the additional 99+ percent of the resin's functionalized surface. This was repeated until the synthesis was complete, at which point the library was screened on bead. Once a colorimetric readout had been observed, the bead of interest was removed from the rest of the beads. It was then subjected to ultraviolet light, which cleaved the tags from the resin. The tags, each of which has a unique retention time on a gas chromatograph, were then analyzed, and the sequence of the interacting peptide deduced [36].

Figure 5 shows an example of encoding 31 reagents at a diversification step with 6 tags. The chromatographic output of "011010" in this case would indicate that reagent 24 had been incorporated into the compound undergoing analysis on the selected bead. Importantly, this methodology is applicable to libraries of essentially *any* type of compound, as long as the chemistry employed does not modify the tags themselves. Subsequent publication of a set of tags able to undergo carbenoid insertion into the resin bead itself obviated the need for resin beads (and subsequent synthetic steps) to carry amino groups [37].

Several other tagging strategies have been described, including cleavable HPLC-separable tags [38], tags that provide specific IR signatures [39], and even simple tagging schemes based on one-dimensional ordering of beads on

| | 1B | 1C | 1D | 2B | 2D | 3B |
|---|---|---|---|---|---|---|
| 1 | | | | | | 1 |
| 2 | | | | | 1 | |
| 3 | | | | | 1 | 1 |
| 4 | | | | 1 | | |
| 5 | | | | 1 | | 1 |
| 6 | | | | 1 | 1 | |
| 7 | | | | 1 | 1 | 1 |
| 8 | | | 1 | | | |
| 9 | | | 1 | | | 1 |
| 10 | | | 1 | | 1 | |
| 11 | | | 1 | | 1 | 1 |
| 12 | | | 1 | 1 | | |
| 13 | | | 1 | 1 | | 1 |
| 14 | | | 1 | 1 | 1 | |
| 15 | | 1 | | | | |
| 16 | | 1 | | | | 1 |
| 17 | | 1 | | | 1 | |
| 18 | | 1 | | | 1 | 1 |
| 19 | | 1 | | 1 | | |
| 20 | | 1 | | 1 | | 1 |
| 21 | | 1 | | 1 | 1 | |
| 22 | | 1 | 1 | | | |
| 23 | | 1 | 1 | | | 1 |
| 24 | | 1 | 1 | | 1 | |
| 25 | | 1 | 1 | 1 | | |
| 26 | 1 | | | | | |
| 27 | 1 | | | | | 1 |
| 28 | 1 | | | | 1 | |
| 29 | 1 | | | | 1 | 1 |
| 30 | 1 | | | 1 | | |
| 31 | 1 | | 1 | | | |

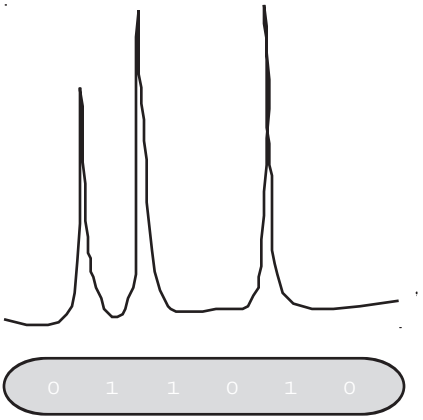


**Figure 5** Binary encoding.

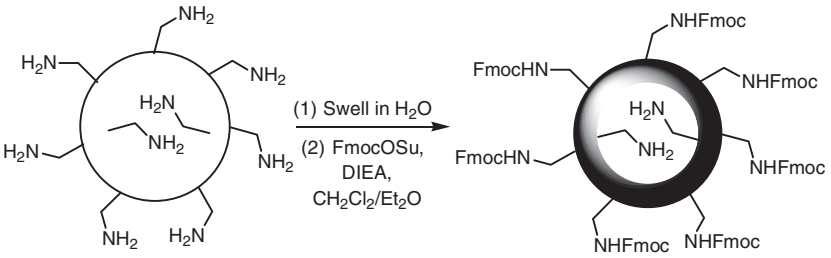


**Scheme 7** Topological segregation of reaction functionality.

a string [40, 41]. A particularly interesting recent variation of tagging has been reported by the Lam group [42]. In this "biphasic" approach (Scheme 7), amino terminated PEG–polystyrene graft polymer beads are first swelled in water, then treated with a solution of 9-fluorenylmethoyloxycarbonyl-*N*-hydroxysuccinimide (FmocOSu) and base in ether/dichloromethane. Because the ether/dichloromethane mixture is unable to penetrate into the water-saturated interior of the bead, Fmoc protection of only the outer layer of beads occurs. *Topological segregation* of reactive groups on the bead subsequently allows for selective tagging of the interior of the bead, while library members are displayed on the external surface. In principle, this strategy has the advantage of elminating interference (and potential "false positives") by the tag in on-bead screening.

## Infrared Coded Resins

An exceptionally innovative strategy for library encoding is the use of the resin bead itself as a means of increasing deconvolution efficiency [43, 44]. Fenniri and co-workers [43, 44] have developed a method to slightly alter the copolymer composition of the Merrifield resin to provide unique bands in the IR and Raman spectrum. These chemically inert resin modifiers do not affect the reactivity of the resin, nor do they have an appreciable affect on resin swelling properties. Also, no library members need to be altered to incorporate tags for deconvolution, as in other earlier methods. All of the resin's surface is therefore available for library synthesis, while providing a spectroscopic readout of which type of resin bead on which the library has been built. This readout, which Fenniri has termed a *barcode* (the similarity to the barcodes on consumer goods is immediately apparent from Fig. 6), can be used to effectively "tag" the first position in a library synthesis and has been speculated to be able to be adapted for complete library deconvolution [43]. Considerable effort has been invested by the Fenniri group in demonstrating that the barcode resins have similar reactivity profiles to standard commercial resins, and that the presence of the synthetic ligand on the bead does not interfere with reading the barcode [44].

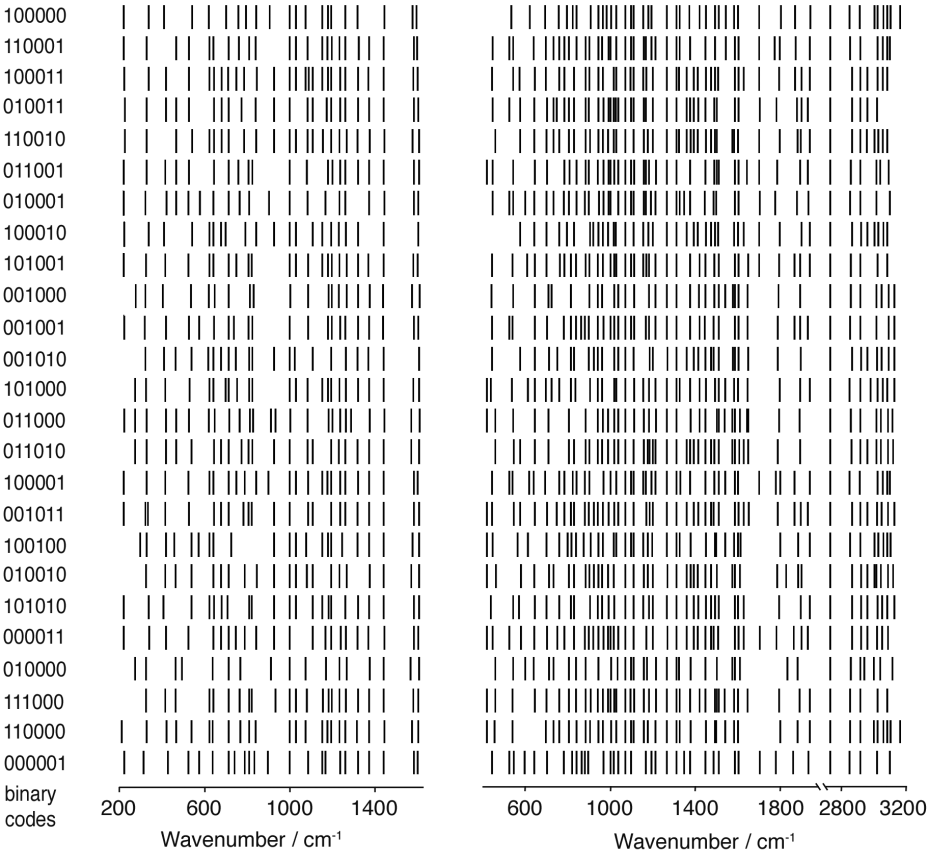


**Figure 6** IR absorption bands used to generate a resin "barcode" [43].

## Radio-Frequency Tagging and the Irori Corporation

In 1995, the Armstrong and Nicolaou groups simultaneously published an elegant solution to the problem of combinatorial library tagging [45, 46]. Much like laboratory animals and consumer goods are now tracked, both groups employed radio-frequency (RF) encoding to track the synthesis history of groups of beads. Effectively, a solution-permeable container (typically an inert material such as Teflon) of beads containing an RF chip (Fig. 7) is prepared. At each step in the synthesis, the IR code of each container is recorded, allowing for its entire synthetic history to be recovered following library screening. Importantly, the process can be automated, and using highly miniaturized systems even applied to the synthesis of moderately large split-pool libraries.

Further miniaturization (and greater library diversity) has been demonstrated by the Irori Corporation, using optical scanning of resin containers labeled with a black and white indexing grid. As a demonstration of this tech-

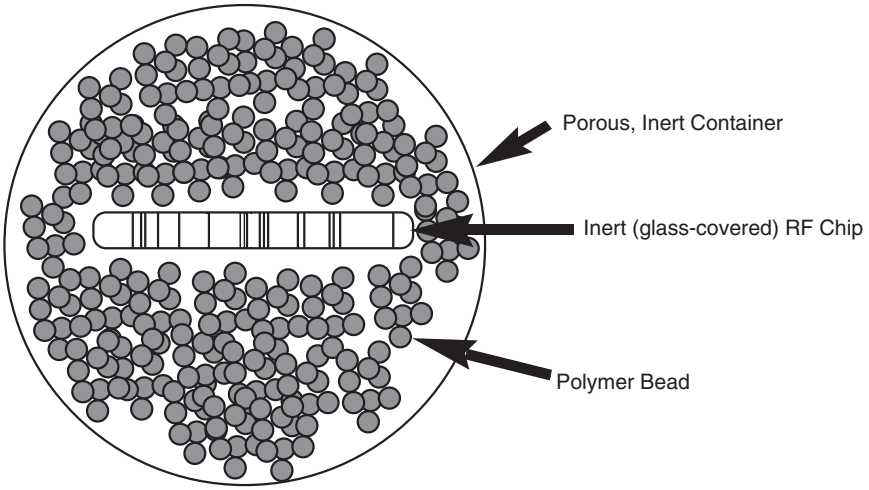


**Figure 7** Radio-frequency encoding.

nology, designated the "NanoKan" system, Nicolaou and co-workers synthesized a 10,000-member library of benzopyrans in NanoKans using split-pool synthesis (Scheme 8) [47]. Given that NanoKans are large relative to the size of individual resin beads, the synthesis of very large (>10,000-compound) libraries is probably not practical using this system. However, the highly automated nature of the encoding process, ability to produce relatively large (milligram) quantities of compounds, and rapid decoding attributes make this an attractive solution for those laboratories able to afford the material and equipment costs.

## 4 OTHER COMBINATORIAL SYNTHETIC TECHNIQUES

### Photolithographic Synthetic Methods

Like its evolution out of early efforts in the solid-phase synthesis of biopolymers, combinatorial chemistry has likewise coevolved with high-throughput screening. As previously mentioned, combinatorial libraries are often screened as single compounds arrayed in 96-, 384-, or 1024-well plates, with even higher densities on the horizon. The increasing miniaturization is in part driven by the relatively small amount of compound available from a single resin bead, but also out of a desire to limit the amount of the target biomolecule required for any single assay. One area of research that has revolutionized genetic analysis, and that is beginning to make its way into combinatorial chemistry, is the idea of the microarray. In terms of the density of different spatially addressable compounds that can be produced per unit area, the champion methodology here is undoubtedly the photolithographic process first

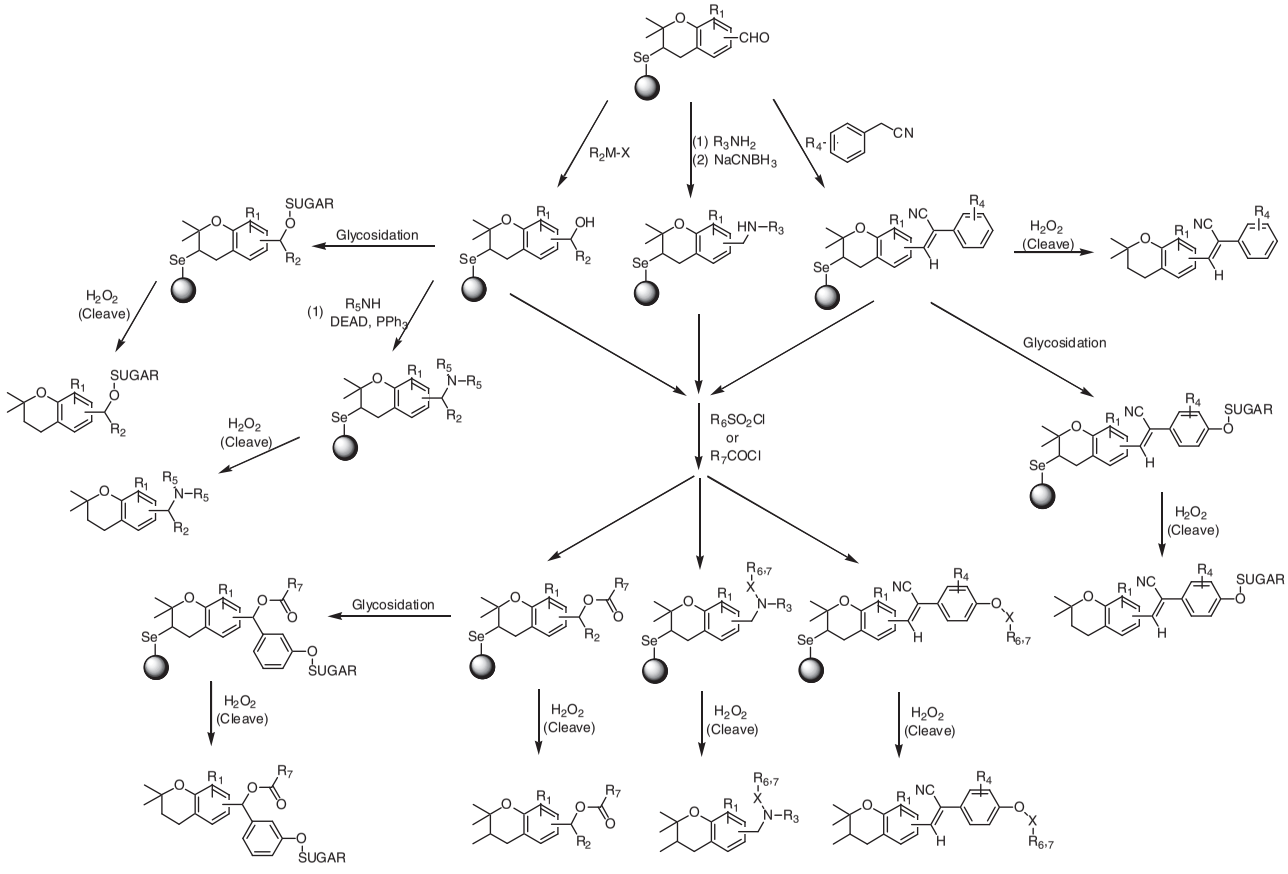


**Scheme 8** Nicolaou benzopyran library.

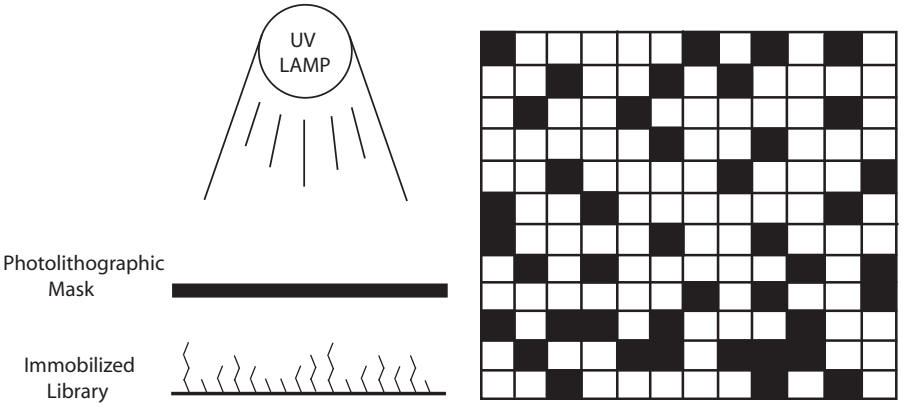


**Figure 8** Light-directed synthesis (*left*) via a photolithographic mask (*right*).

described in 1991 by Fodor and co-workers [48]. Taking a page from the microelectronics industry, photolithographic synthesis of microarrays relies on the availability of photodeprotectable reagents: protected monomers that will become an array of oligomers (bio- or otherwise), or, potentially, protected reagents that can form an array of small molecules. Beginning with a substrate (glass or a thin gold film) derivatized with a photocleavable group, light is shined through a photolithographic mask to selectively deprotect particular regions of the substrate (Fig. 8). Following UV deprotection of an area of the chip, a library building block is coupled to all of the available, deprotected, sites. Changing the photolithographic mask subsequently allows for other areas of the chip to undergo deprotection. The process is repeated until all of the desired library members are generated. On chip screening of labeled targets can then easily be performed, and no deconvolution of binders is necessary because the patterns of the photolithographic masks used to create the immobilized library are recorded during synthesis. Initially developed primarily as a method for peptide synthesis, photolithographic synthesis of oligonucleotides as pursued by the genomics company Affymetrix has become a major industry. In principle, such a method could also be applied to the synthesis of druglike molecules; however, to our knowledge, this has yet to be demonstrated successfully. A recent report by Kodakek and co-workers describing the photolithographic synthesis of arrays of peptoids further illustrates the potential power of this method [49].

## High-Throughput Methods

An alternative to the photolithographic technique is the pin-based spotting method developed by Khodursky and co-workers [50]. Although arrays produced using this method are not capable of yielding the spot density of Affymetrix chips, they can be produced using equipment that is significantly

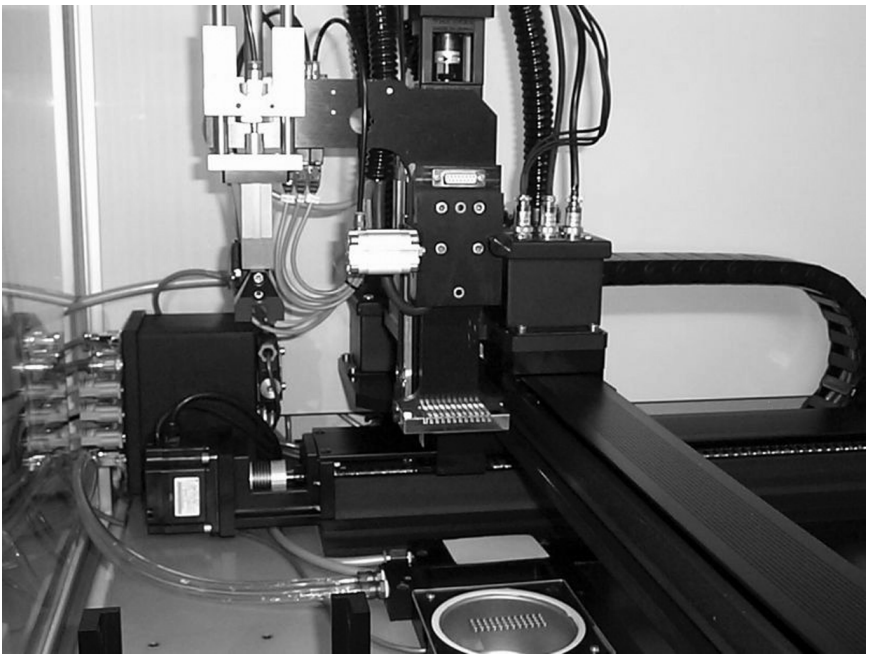

**Figure 9** Commercial microarrayer. The print head, capable of holding 384 pins, is just below the center of the photograph.

less complex and expensive. Still at the early stages of being applied to small-molecule screening, spotted arrays have already shown considerable promise in the rapid screening of libraries. The arrayer itself is effectively a three-axis robot equipped with a print head bearing precisely machined, steel quill pens (or "pins") (Fig. 9). Computer control of the arrayer allows for precise deposition of nanoliter-sized droplets on a glass slide, producing an array of potentially many thousands of spots.

One early demonstration of the utility of spotted small-molecule arrays was reported by Schreiber and Coworkers [51, 52]. Following synthesis of a split-pool library on *macrobeads* (large resin beads capable of carrying greater amounts of each individual library member), resin beads were individually deposited into 384-well plates. In-plate cleavage of library members was carried out, followed by evaporative removal of the solvent. The residue remaining was taken up in dimethyl sulfoxide (DMSO), and spotted onto glass slides using an arrayer analogous to that described by Yanofsky [50]. Subsequent screening using a fluorescence-tagged protein provided ready identification of those compounds on the array with the desired binding ability.

The use of macrobeads in combination with microtiter plate cleavage allows for the generation of stock solutions of the library members. Therefore, libraries can be screened many times, against many different ligands. However,

as with all other split-pool strategies, some form of tagging is necessary to provide a positive identification of library hits. Use of small-molecule microarrays is clearly a technique that is on the rise in academia. Recent reports have included a description of small-molecule arrays targeting profiling caspase activation [53], small-molecule ligands for human IgG [54], and arrays of synthetic oligosaccharides [55].

## 5 TAG-FREE METHODOLOGY

### Dynamic Combinatorial Chemistry and Allied Methods

In addition to the methods described above, combinatorial chemistry has spawned several research efforts designed to mimic nature's methods of deriving compounds with a desired property. In some cases, the goal is to mimic evolution: selecting and amplifying high-affinity binders from dynamic pools of chemical structures. In other cases the goal is to mimic biosynthesis: templating the construction of a molecule from a library of starting materials allowed to interact with a binding pocket. For some of these new techniques, receptor-templated synthesis *and* evolution are coupled.

Evolution is a central tenet of modern biology. The phenotype of an organism is, of course, a reflection of the constant process of mutation, selection, and amplification operating on a vast scale at the molecular (genetic) level (Fig. 10). In the mid-1990s, several research groups, including ours, began reporting proof-of-principle studies of a conceptually new direction in combinatorial chemistry. This field of dynamic combinatorial chemistry (DCC) seeks to mimic what nature so elegantly achieves. In a dynamic combinatorial library, molecular building blocks react in a reversible fashion, affording a continuous interconversion of all possible library members.

Dynamic combinatorial libraries (DCLs) are most often constructed using covalently linked functional groups formed under reversible conditions, such as esters, imines, hydrazones, borates, disulfides, and others [56]. Libraries constructed through hydrogen bonding networks, ionic bond formation, or coordination to a metal center have also been reported [57]. In the dynamic approach, an equilibrating system containing various numbers of reactive

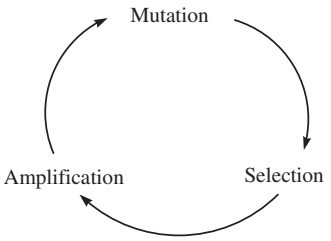


**Figure 10** Darwinian evolution.

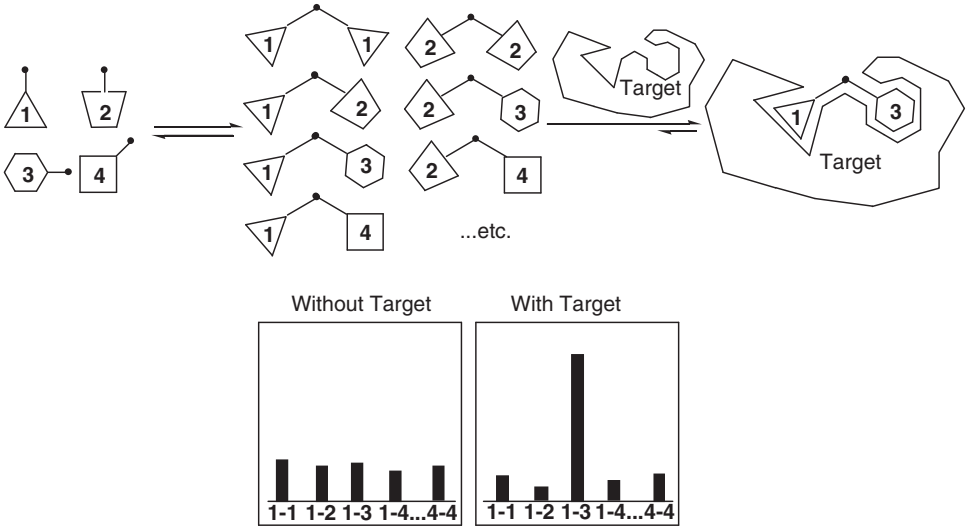


**Figure 11** Target-directed selection and amplification of chosen library members from a dynamic equilibrium.

building blocks provides a broad pool of interconverting compounds. For example, a mixture containing ($c$) aldehydes and ($d$) amines theoretically generates ($c \times d$) imines. When exposed to a target receptor, library members undergo a selection based on binding affinity. Because formation of each library member from its constituent components is reversible (i.e., allowing "mutation" to occur), LeChatelier's principle requires that the equilibrium shift in favor of those compounds "sequestered" by binding to the target. In principle, this means that the highest affinity binder will be subject to the greatest amplification. Selected member(s) can then be monitored by enhancement of mole fraction via HPLC and/or mass spectrometry (MS) or by other suitable methods (Fig. 11).

Several research groups have published proof-of-concept illustrations for the selection and amplification of individual DCL species based on target affinity [58]. Eliseev and co-workers describe the selection and amplification of *Z*/*Z* di-carboxylate ligand upon addition of a guanidinium group, from a photochemically induced *Z*/*Z*, *E*/*E*, and *E*/*Z* equilibrium [59]. Orientation of the carboxylate moieties directly influences the ability to bind the target. As shown in Scheme 9, only the *Z*/*Z* compound exhibits a conformation that affords hydrogen bonding between both carboxylates to the arginine center. After several cycles of photochemical equilibrarion and affinity-chromatographic removal of the *Z*/*Z* isomer, HPLC analysis revealed a dramatic and selective enhancement in the mole fraction of the *Z*/*Z* isomer [59].

A fundamental concern when planning a DCL is ensuring the production of a library exhibiting unbiased representation of all library members. With

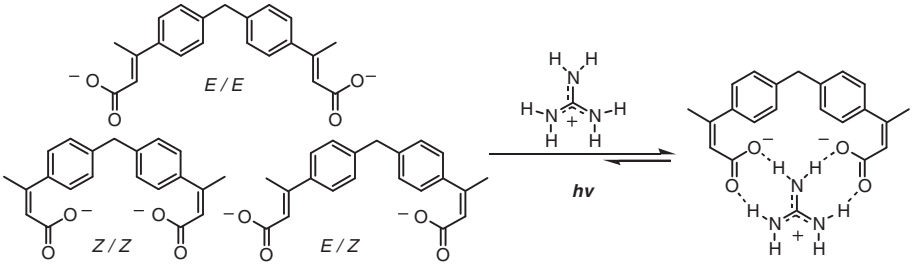


**Scheme 9** Eliseev's selection and amplification of a DCL member based on guanidinium affinity.

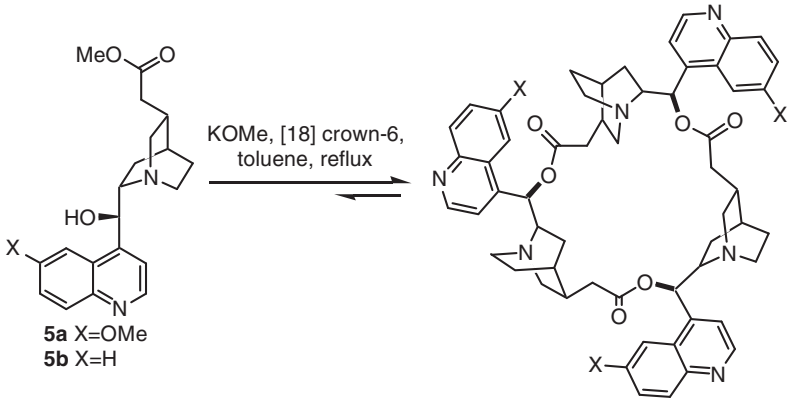


**Scheme 10** Quantitative formation of the cyclic trimer of cinchonidine derived hydroxyl ester **5**.

library selection under thermodynamic control, intrinsic properties of starting materials and library products can affect the observed selection as well as binding to the target. Significant differences in reactivity between individual library components, or self-selection within those components, can result in a DCL containing a limited number of members constituting the majority of the molecular pool, while other theoretical library members are underrepresented or entirely absent.

In several studies, this self-selection property has been explored explicitly. Such an example has been published by Rowan et al. [60] in which a cinchonidine-derived hydroxyl ester (**5**) was reacted with potassium methoxide under thermodynamic conditions, which theoretically could generate numerous cyclic and polymeric compounds. Instead, only the cyclic trimer was obtained in quantitative yield (Scheme 10) [60]. From the perspective of lead discovery, this research demonstrates the importance of assuring a diverse representation of library members when planning a DCL. In general,

researchers have had success in constructing DCLs using a variety of reactions that undergo rapid reversion, including transesterification.

In situ target-induced selection and amplification makes DCC an ideal technique for lead discovery of compounds that bind to biological macromolecules. Because of the in situ amplification technique, reversible reactions chosen must be amenable to physiological conditions, or a bi-phasic or compartmental approach must be used. For example, researchers have shown that thiols undergo facile disulfide exchange under mild conditions in aqueous media [61]; however, when using proteins as a target, component exchange with cysteine residues is a concern. In general, numerous other reversible reactions including imination, hydrazone formation, and olefin metathesis have been employed to prepare DCLs in aqueous media [56]. When using reactions not amenable to a physiological environment, researchers have cited the use of compartmentalization through a semipermeable membrane through which library members formed in an organic phase can pass, and subsequently encounter the chosen target in an aqueous phase [62].

Because of its potential for building up binding structures in a modular fashion, as well as the intended parallel with genetic evolution, DCC would seem to be an ideal method for the identification of new compounds targeting nucleic acids. We have described the selection and amplification of equilibrating metal coordinated salicylaldimine and salicylamide complexes in the development of DNA [63] and RNA [64] ligands, respectively. In part, these efforts were an expansion of several other groups' reports that *nonlabile* organotransition metal complexes could bind DNA and RNA with interesting results [65].

In our initial report, we described the first utilization of immobilized DNA to drive the selection and amplification of ligand(s) from a dynamic library of zinc(II) complexes. Six salicylaldimine "monomers" (**7** to **12**) were reacted with excess zinc(II) chloride, providing a library of 36 unique bis(salicylaldiminato)zinc complexes (**6**) (Fig. 12). Given the ability of zinc(II) to bind a broad range of functionality with variable coordination numbers and geometries, it is likely the total library diversity was much greater. Two independent libraries, one consisting of compounds **7** to **12** and the other containing compounds **7** to **12** and excess zinc(II) chloride were incubated in affinity columns constructed of ds(T·A) cellulose. Following a 2-h incubation of the libraries on resin, solutions were eluted and lyophilized. Analysis of the mixtures was performed by hydrolysis of the complexes with trifluoroacetic acid, followed by derivatization of the amines with excess 2-naphthoyl chloride (to allow for UV detection) and separation/structure elucidation by reverse-phase HPLC. Comparing relative abundance of monomers present after elution provided information about which components bound dsDNA, that is, less monomer present after elution represents binding of that monomer to the immobilized dsDNA.

After multiple analyses and control experiments, monomer **8** proved to be most strongly retained by the dsDNA affinity column in the presence of

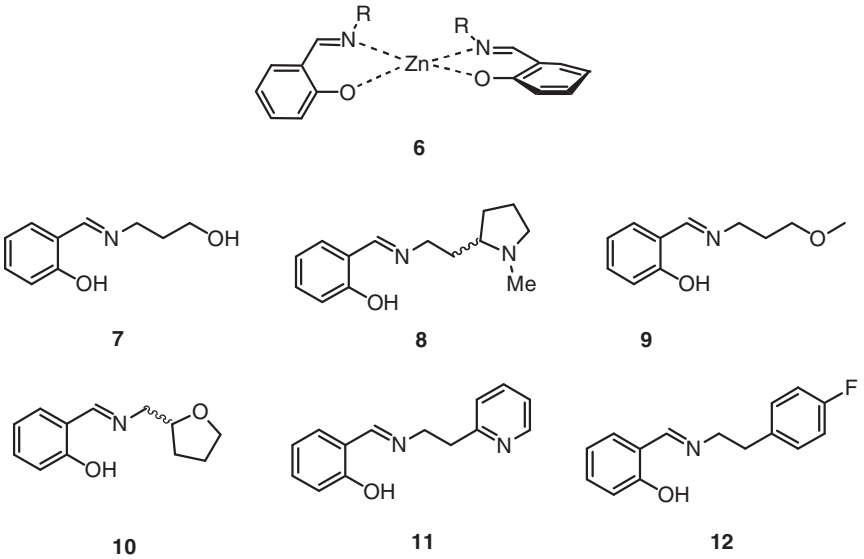


**Figure 12** Salicylaldimine monomers (**7** to **12**) and hypothetical zinc(II)-mediated bis(salicylaldiminato) complex.

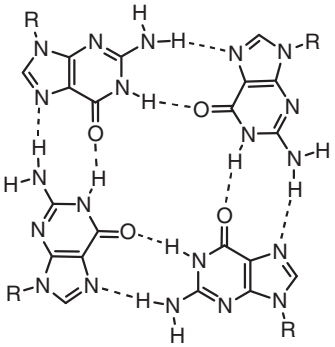

**Figure 13** Hydrogen bonding network in the guanine tetrol motif.

zinc(II). Subsequent UV titration experiments provided a dissociation constant of 1.1 μM, significantly stronger than other possible bis(salicylaldiminato)zinc complexes tested.

Balasubramanian et al. recently reported using a dynamic combinatorial strategy in an effort to discover ligands capable of selectively binding guanine quadruplex DNA [66]. The tertiary structure of guanine quadruplex structure has shown interesting activity involving the regulation of telomere length and exhibits various motifs that differentiate it from B-DNA. For example, the terminal guanine tetrad provides potential for hydrophobic or π-π ligand association (Fig. 13). In previous work, this group demonstrated that the

tetrapeptide sequence F-R-H-R exhibits selective binding to the quadruplex motif [67].

Based on this study, Balasubramanian and co-workers [66] subsequently used disulfide exchange to construct a dynamic library using a thiol-terminated F-R-H-R tetrapeptide **P** and a thiol-containing acridone derivative **A**, proposed to associate with the hydrophobic guanidine tetrol. Equilibrium was established in the presence of glutathione to mediate disulfide exchange between (**A**) and (**P**), as well as diversify the library. Upon equilibration, a biotinylated sequence 5′-biotin$(GTTAGG)_5$ was incubated in the dynamic library. After target-induced reequilibration, disulfide exchange was halted by acidification of the medium, followed by removal of the ligand–target complex through immobilization on streptavidin-coated beads. Subsequent heat-induced denaturation allowed for the release and characterization of selected ligands (Scheme 11). From the nine possible library members, introduction of the guanine quadruplex increased the molar concentration of the hetero complex **A–P** and the peptide dimer **P–P** 400 and 500 percent, respectively. This dynamic approach yielded the first report of a short peptide sequence that displayed selective binding to guanine quadruplex DNA.

Lehn and co-workers [68] have demonstrated that a DCL approach can be used to probe the binding site of the plant lectin Conconavalin A (Con A). Lehn et al. employed acylhydrazone formation through the intermolecular condensation of various hydrazines and aldehydes, reversible in a pH range of 3 to 7. Six monosaccharide benzaldehydes and nine core component hydrazines were used, offering one, two, or three hydrazine reactive sites, thereby producing both monovalent and multivalent DCL members (Scheme 12).

Complete equilibration of the 15 library components afforded a dynamic library containing 474 members. Initial screening was performed using an enzyme-linked lectin assay (ELLA) [69]. In the ELLA assay, microtiter plates were coated with yeast mannan (a known substrate of Con A). Peroxidase-

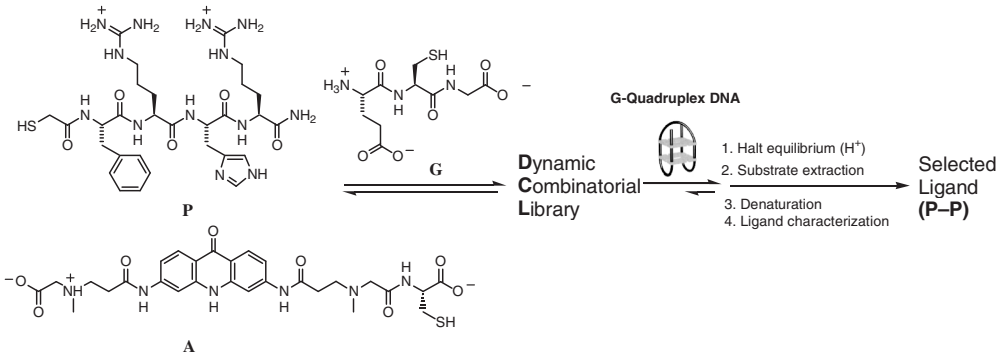


**Scheme 11** Balasubramanian's dynamic approach toward G-quadruplex DNA ligands.

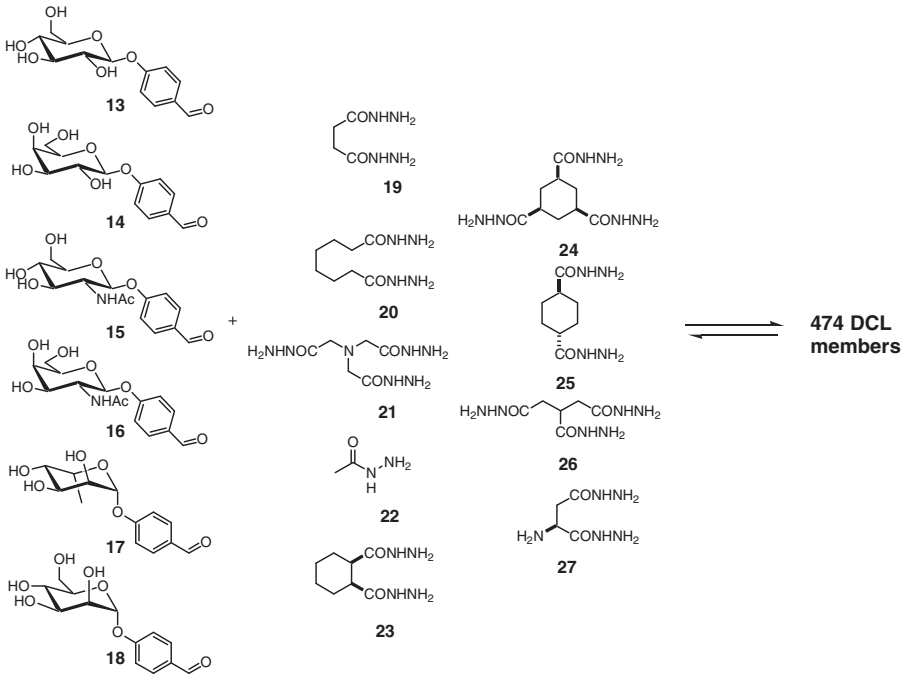


**Scheme 12** Dynamic approach toward Concavalin A ligand [68].

labeled Con A was then introduced in the presence of the library. Binding of Con A by library members was then monitored as a function inhibition activity of the enzyme. Initial results indicated more than one inhibitor of enzyme activity, that is, various members displayed similar binding coefficients. Using a dynamic deconvolution method, 15 sublibraries were made, each omitting one component, thereby excluding any library members consisting of that component [70]. As a result, sublibraries that showed a decrease in inhibitory effect illustrate the importance of that missing component, and therefore any library members containing them. After full analysis, Lehn and co-workers [68] found that the mannose carbohydrate **18** was the most active saccharide, and the two trivalent hydrazines **21** and **26** were the most active linkers. From libraries containing only those constituents, the trimannose compound **(18)$_3$-21** was found to be the best inhibitor. Comparison to methyl-α-D-mannoside yielded an estimated $IC_{50}$ value of 22 μm, a 36-fold increase over the monosaccharide.

As evidenced by the above example, larger DCLs can produce various library members that have similar receptor binding affinity, making it difficult to quantify amplification of a single library member. Kazlauskus et al. have reported the use of irreversible destruction of unbound or weakly bound ligands, thereby making analysis of selected ligands in a large molecular pool an easier task [71]. In their study, numerous aryl sulfonamides were prepared

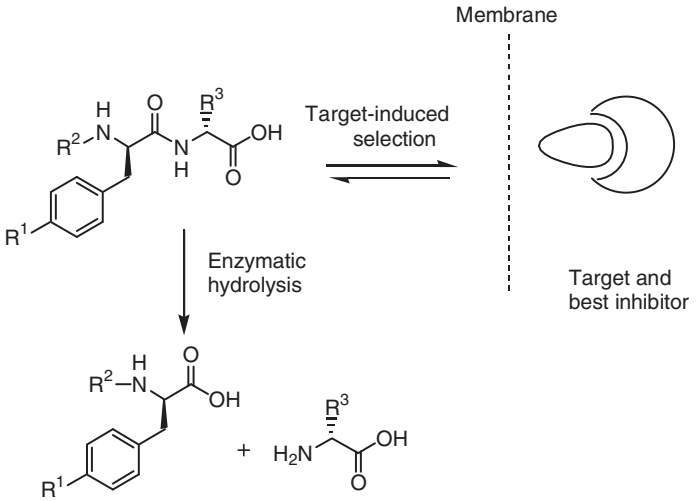


**Scheme 13** Kazlauskus' destruction/amplification approach.

and screened against carbonic anhydrase. After subjection to the target, the protease *pronase* was added, which enzymatically hydrolysized unbound or weakly bound ligands, while maintaining the integrity of ligands with high target affinity, allowing for subsequent analysis (Scheme 13).

Recent reports by Severin [72] and by Sanders and co-workers [73] place the idea of target-mediated selection and amplification of high-affinity binders from dynamic combinatorial libraries on firmer mathematical footing. In essence, both researchers note that the library member amplified to the greatest extent is not necessarily that with the highest affinity to the target, if the DCL experiment is not designed properly. This potential difficulty decreases proportionately as the size of the DCL increases.

While dynamic combinatorial chemistry has yet to produce a major success (in terms of a clinical lead compound) in the same sense that other techniques have, several groups have provided proof of concept experiments indicating that substrate-induced selection can afford detectable amplification of "active" molecules within a library. Recent reports by an industrial group led by Eliseev targeting neuraminidase have demonstrated that DCC does have the potential to be more than just an academic endeavor [74]. We anticipate that the dynamic combinatorial approach will prove to be a novel resource for lead discovery and will continue to provide fundamental insight on aspects of molecular recognition.

## Click Chemistry and in Situ Click Chemistry

As we have already described, the solid-phase approach to combinatorial synthesis has survived and prospered because of two distinct features. First, solid-

phase synthesis allows for the use of excess reagents, driving otherwise moderate reactions toward completion. Second, solid-phase reactions allow for product purification through filtration, bypassing more complex chromatographic procedures. However, despite these attributes, solid-phase chemistry can suffer severe limitations as a tool for the pharmaceutical industry. For example, driving reactivity with the use of excess reagents in the mobile phase becomes wasteful with respect to both reagents and solvent. Second, the use of a traceless linker [75] is often required when preparing drug candidates on solid support, complicating synthetic design. Third, the ability to gain diversity relies on consistent reactivity in the presence of various functionality. Finally, very large scale solid-phase synthesis is prohibitively cumbersone and expensive. Dynamic combinatorial chemistry is one response to these issues; other researchers have recognized that solution-phase multiparallel synthesis using near perfect reactions represents an attractive alternative. One version of this idea has been explored in detail by Sharpless and co-workers, who use the term *click chemistry* to describe the design of libraries relying exclusively on reactions that are exceptionally high yielding and functionality tolerant [76]. *Click reactions* result from a high thermodynamic driving force ($\Delta G \leq -20$ kcal/mol) and thereby afford selective reactivity in the presence of various moieties.

In a typical click chemistry reaction sequence, building blocks are pieced together using well-studied reactions on readily available starting materials. For example, olefins can be extensively decorated through the use of cycloadditions such as 1,3-dipolar and the Diels–Alder reaction, nucleophilic opening of strained rings, dihydroxylation, and aziridination, while carbonyl moieties can be modified via hydrazone formation or other "irreversible" nucleophilic addition. While an argument can be made that click chemistry is limited by its definition (the use of only a few perfect reactions), it has been shown that great diversity can be achieved when using only a few reaction types [77]. As an example, the generation of *spring-loaded* cyclic electrophiles such as epoxides or aziridines, and subsequent nucleophilic ring-opening reactions from olefinic starting material offers a diverse library of molecules (Scheme 14). Other possibilities, including olefin heterocyclization and subsequent nucleophilic ring-opening reactions, are dependable, stereospecific, often highly regioselective, can be performed either in water or in the absence of solvent, and provide the product in nearly quantitative yield.

Because of these attractive aspects, click chemistry is increasingly being used for drug discovery and biomedical research applications, in both academic and industrial laboratories [78]. For example, Fokin, Sharpless, and Wong have reported the use of click chemistry in the formation of an inhibitor of α-1,3-fucosyltransferase (Fuc-T), an enzyme that catalyzes the final glycosylation in the biosynthesis of many important saccharides [79]. The key step in this transformation is a manganese(II)-catalyzed nucleophilic displacement of L-fucose from guanosine diphosphate β-L-fucose (GDP-fucose) from a glycoconjugate acceptor (Fig. 14).

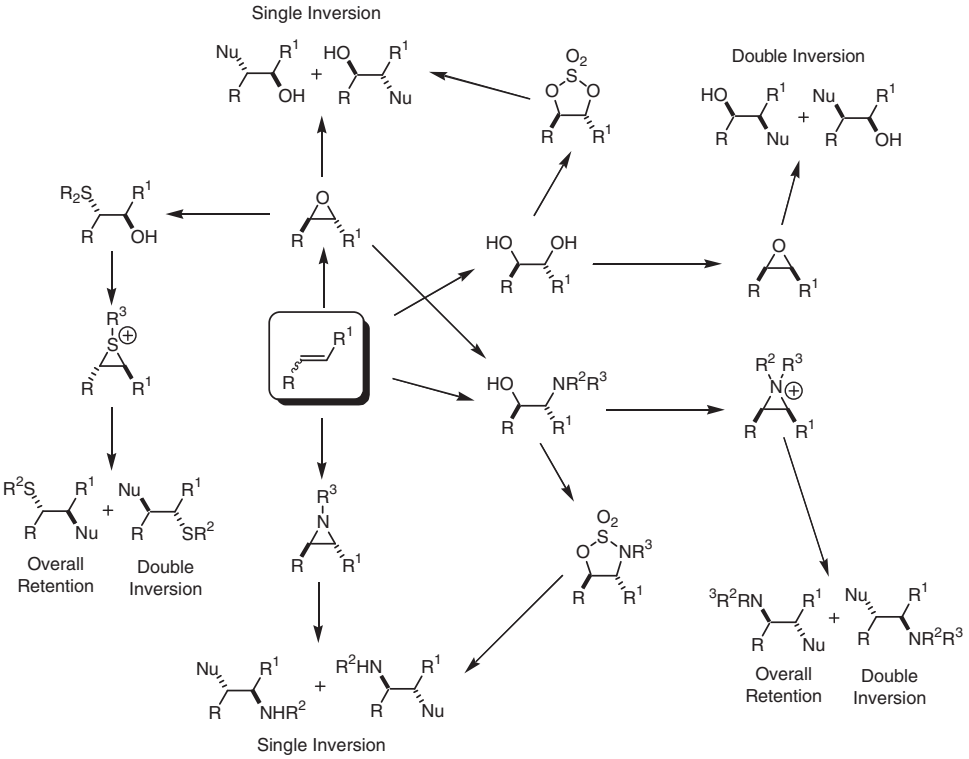


**Scheme 14** Generation of diverse components through the formation of cyclic electrophiles from olefins and their oxidation products.

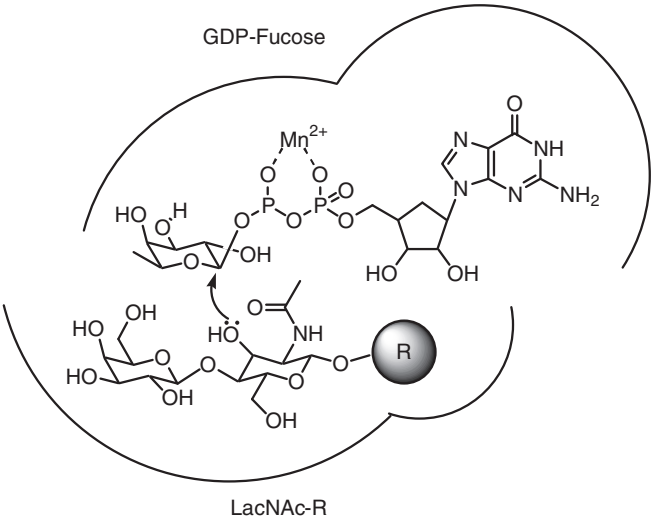


**Figure 14** Fucosyltransferase-catalyzed transfer of β-lucose to acceptor glycoconjugate.

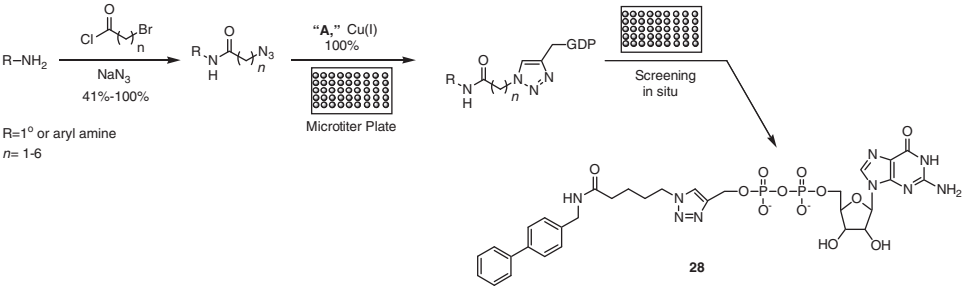


**Scheme 15** Triazole synthesis in microtiter plate and subsequent in situ fucosyltransferase inhibitor detection.

The majority of the binding affinity for Fuc-T lies in the GDP moiety; however, it has been demonstrated that interaction of aromatic or aliphatic functionality in the neighboring hydrophobic pocket can increase affinity 70-fold. Therefore, a library of aromatic and aliphatic GDP-triazole candidates were prepared and screened for inhibition. The triazole formation is accomplished using Huisgen's Cu(I)-catalyzed 1,3-dipolar cycloaddition between an alkyne and azide [80]. This reaction was chosen because it proceeds in water, produces near quantitative yield, and is effective in the presence of vast functionality, granting a "click" seal of approval. Thus, 85 azide compounds were synthesized utilizing various hydrophobic groups and tether lengths of 2 to 6 carbons (Scheme 15).

After completion of triazole formation, compounds were screened for inhibitory activity in situ with the use of a pyruvate kinase/lactate dehydrogenase-coupled enzyme assay. The library was screened against human α-1,3-fucosyltransferase VI (Fuc-T VI), where compound **28** was found to show the highest binding affinity. Steady-state kinetic studies show compound **28** is a competitive inhibitor against GDP-fucose, with a $K_{i(\mathrm{comp})} = 62$ nM, making this compound the first nanomolar inhibitor of Fuc-Ts. Compound **28** was then tested for inhibitory activity against other glycosytransfersases and nucleotide binding enzymes. Fokin, Sharpless, and Wong [79] concluded that compound **28** is highly selective toward Fuc-T VI, showing weak inhibition of α-Fuc-T III and α-Fuc-T VI, and no inhibition of galactosyltransferases α-1,3-GalT and β-1,4-GalT. Weak inhibition was observed against guanylate kinase, and no inhibition was detected against pyruvate kinase. The above scenario is a prime example of the use of a click reaction to create moderately sized libraries in near quantitative yield, from which screens can be conducted without the need for purification of compounds.

The initial discovery by Mock and co-workers of a dramatic enhancement in the reaction rate of the Huisgen 1,3-dipolar cycloaddition of azides and alkynes, a relatively slow reaction under "normal" conditions, displayed for the

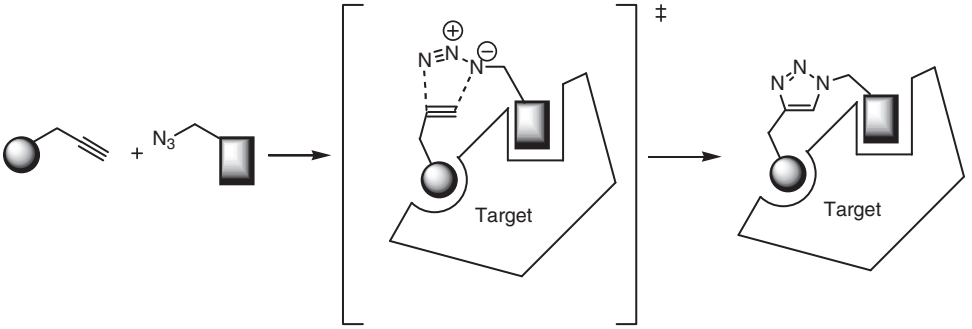


**Scheme 16** In situ target-catalyzed ligand formation using Cu(I)-catalyzed triazole formation.

first time the use of biomacromolecules as reaction vessels [81]. Rate acceleration is presumably brought about through the binding of reaction components to a biomacromolecule, thereby providing a unique microenvironment (Scheme 16). This finding has prompted researchers to investigate an in situ approach to ligand synthesis. Click reactions, benign to physiological conditions, afford the ability to create small-molecule ligands in the presence of their biological substrate, without the potential for side reactions with biomolecular functionality such as amino acid side chains.

Sharpless and co-workers [82] have reported the use of triazole formation, a favored click reaction, as a method for irreversible bond formation in the presence of a desired target. In their approach, Sharpless et al. targeted acetylcholine esterase (AChE), an essential component in neurotransmitter function in the central and peripheral nervous systems. The enzyme displays two neighboring binding sites, located within a narrow gorge approximately 20 Å in depth [83]. Both binding sites have known ligands exhibiting tacrine and phenanthridinium motifs [84]. Using these known ligand motifs of differing linker lengths, decorated with alkynyl and azide functionality, triazole formation was allowed to proceed in the presence of AChE. Because the rate of triazole formation at room temperature in the absence of enzyme is negligible, any detected bi-dentate triazole formed is done so solely from enzyme-assisted rate acceleration. After sufficient incubation in the presence of *Electrophorus* AChE, the mixture was analyzed via desorption–ionization mass spectrometry (DIOS) [85]. Of the 98 potential bivalent ligands, including 34 regioisomeric pairs of *syn* and *anti* tacrine/phenanthridinium triazoles, and 15 regioisomeric pairs of tacrine/tacrine triazoles, an observable amount of only one bidentate ligand was found, TZ2-PA6 (Scheme 17).

Preparation of the selected triazole in the absence of *Electrophorus* AChE provided authentic material in high yield in approximateley 1 : 1 *syn* : *anti* ratio. Regioisomers were separated by HPLC and individually analysized. HPLC comparison shows that the enzyme exclusively produces the *syn* triazole

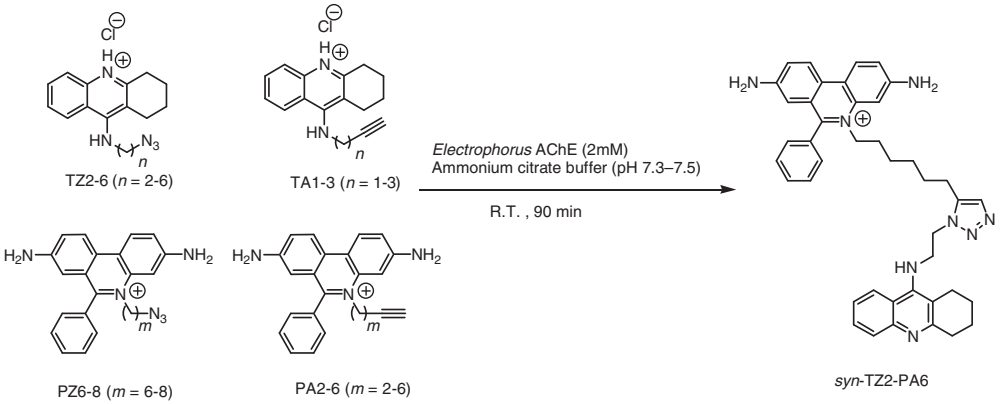


**Scheme 17** Azide (Z), acetylene (A), tacrine (T), and phenanthridinium (P). Substrate-induced selection of AChE.

isomer. It was later found that the *syn* isomer has a $K_d$ of 77 to 410 fM, 140 times stronger than the *anti* isomer.

Click chemistry continues to provide researchers facile ways to create diverse libraries from which biological systems can be perturbed and information regarding molecular recognition can be gained. Conceptually, click chemistry also provides motivation for the discovery of new high-yielding, selective, and functionality-tolerant synthetic methodology.

## Thiol-Based in Situ Assembly of Macromolecular Ligands

The above sections outlined studies in which biological macromolecules were used as reaction vessels, inducing rate acceleration in the formation of high-affinity ligands, or amplification of high-affinity ligands from dynamic libraries. In a related effort, Sunesis Pharmaceuticals has recently described a "tethering" technique that uses a native or engineered cysteine thiol present in close proximity to a macromolecule's active site(s) to covalently retain a relatively low-affinity ligand, from which a multidentate high-affity ligand can be synthesized in situ [86]. After subsequent thiol deprotection, a variety of disulfides are incubated in situ in the presence of a suitable reducing agent such as 2-mercaptoethanol. Like the dymanic combinatorial model, disulfides containing chemical architecture compatible with neighboring binding sites preferentially undergo disulfide exchange with a ligand's covalently attached thiol, thereby capturing a multidentate lead compound that can be elucidated through mass spectrometry (Fig. 15).

In their initial report, Sunesis researchers focused on the enzyme thymidylate synthase, (TM), an essential component in the biosynthetic pathway of dTMP and dUMP, an interesting target for antifungal and anticancer studies [87]. *E. coli* TM contains a cysteine residing in the active site, as well as four

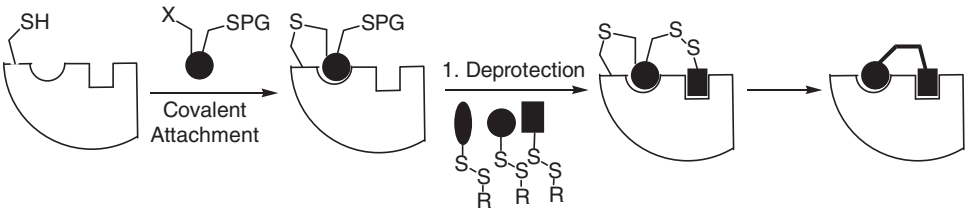


**Figure 15** In situ assembly of high-affinity ligands through covalent capture.

cysteine residues buried within the protein. Prior research suggested the active-site cysteine residue (C146) to be most susceptible to modification. This was confirmed by incubating both wild-type and mutant TM (C146S) in the disulfide cystamine. It was observed that the wild-type enzyme selectively formed disulfide likage at the active site cysteine [87]. Therefore, disulfide libraries were constructed using solid support methodology to couple mono-BOC-cystamine to readily available carboxylic acids, sulfonyl chlorides, isocyanates, isothiocyanates, as well as oxime, and N-tosyl-D-proline-based disulfide libraries. Upon cleavage and sufficient purification, disulfide library members were dissolved in DMSO to a final concentration of 50 or 100 nM. Typical experimental protocol then calls for the addition of 1 μL DMSO solution, containing 8 to 15 disulfide library members to be incubated in 49 μL of approximately 15 μM protein suspended in 25 mM potassium phosphate/1 mM 2-mercaptoethanol buffer (pH 7.5). After sufficient incubation (30 min) the solution is analyzed via HPLC/MS, therefore providing structure elucidation for attached ligands (with respect to their atomic mass of the ligand). Upon screening 1200 library members, HPLC/MS analysis showed a preferential capture of *N*-tosyl-D-proline analogs over other library members. Quantitative structure–activity relationship (SAR) studies show a fair amount of flexibility is afforded to the tosyl phenyl ring, as *t*-Butyl can readily replace methyl moiety without great loss of affinity. Proline cyclic architecture appears to be important in retaining affinity, replacing this moiety with phenylalanine, phenylglycine, or pyrrole results in loss of affinity. Lastly, the phenylsulfonamide scaffold appears essential, as methyl proline is observed as a high-affinity ligand (Fig. 16). Recently, Sunesis has used the tether-based approach to discovered lead compounds for various protein active sites and small-molecule inhibitors of protein–protein interaction such as cytokine interleukin-2 and the IL-2α receptor [88].

## Structure–Activity Relationship by Nuclear Magnetic Resonance

Structural analysis of proteins by nuclear magnetic resonance (NMR) spectroscopy continues to be an extensive area of research [89]. More recently, NMR has proven to be a valuable tool for ligand discovery as well.

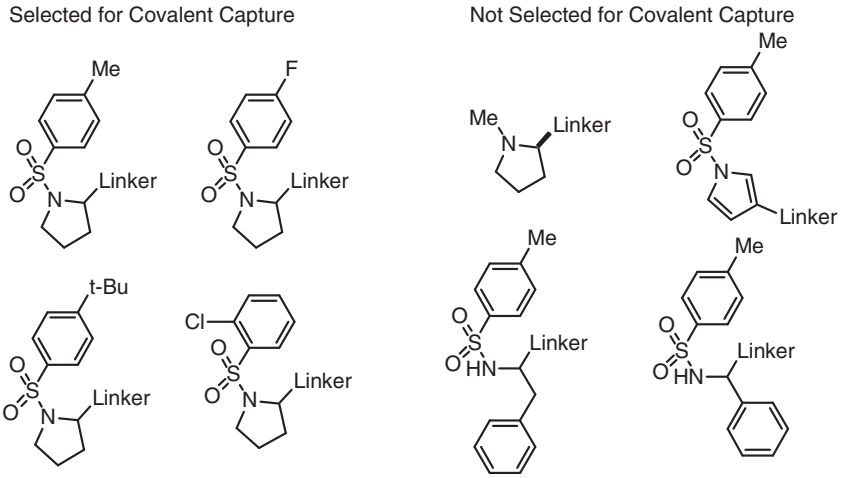


**Figure 16** Covalent capture of ligands by tethering and SAR analysis.

*S*tructure–*a*ctivity *r*elationships by NMR, or simply SAR by NMR, developed by Fesik and co-workers, makes use of the well-known ability to detect the binding of ligands to various sites of a protein through $^{15}$N or $^{1}$H amide chemical shift changes in two-dimensional $^{15}$N-heteronuclear single quantum correlation ($^{15}$N-HSQC) spectra [90].

Using $^{15}$N-labeled protein, spectra can be compiled in a short period of time, allowing numerous ligands to be screened quickly. Once small-molecule ligands are discovered, affinity to the target binding site is optimized through the synthesis and screening of analogous compounds. Relatively low-affinity small-molecule ligand(s) for neighboring binding sites are then found and optimized through identical techniques. Proximal small molecules can be linked synthetically, producing a multidentate ligand with enhanced affinity (Fig. 17).

In their initial study, Fesik and co-workers used the SAR by NMR approach to derive a high-affinity ligand for the well-studied FK506 binding protein (FKBP) [90]. FKBP was initially screened against a library from which many compounds were found to bind with dissociation constants in the millimolar and micromolar range. A trimethoxyphenyl pipecolinic acid derivative, **29**, showed the highest affinity with a $K_d$ of 2.0 μM. The binding site of this small molecule was found to be similar to that of similar moiety within the FK506 structure, and therefore an effort was made to find small-molecule receptors for nearby binding sites. Compounds were then screened in the presence of $^{15}$N-FKBP, saturated in **29**. Those studies provided a small molecule that bound to FKBP with an affinity of $K_d = 0.8$ mM, as determined from the changes in $^{15}$N-HSQC spectrum. Optimization of this low-affinity ligand

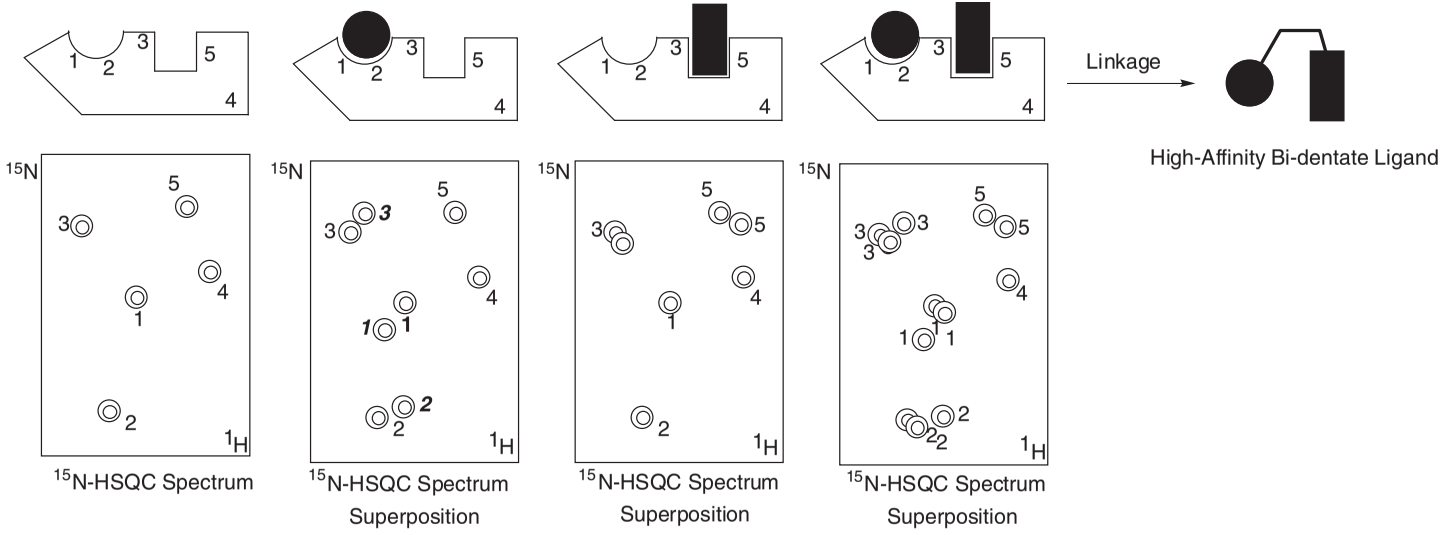


**Figure 17** $^{15}$N-HSQC analysis and development of a multicomponent ligand.

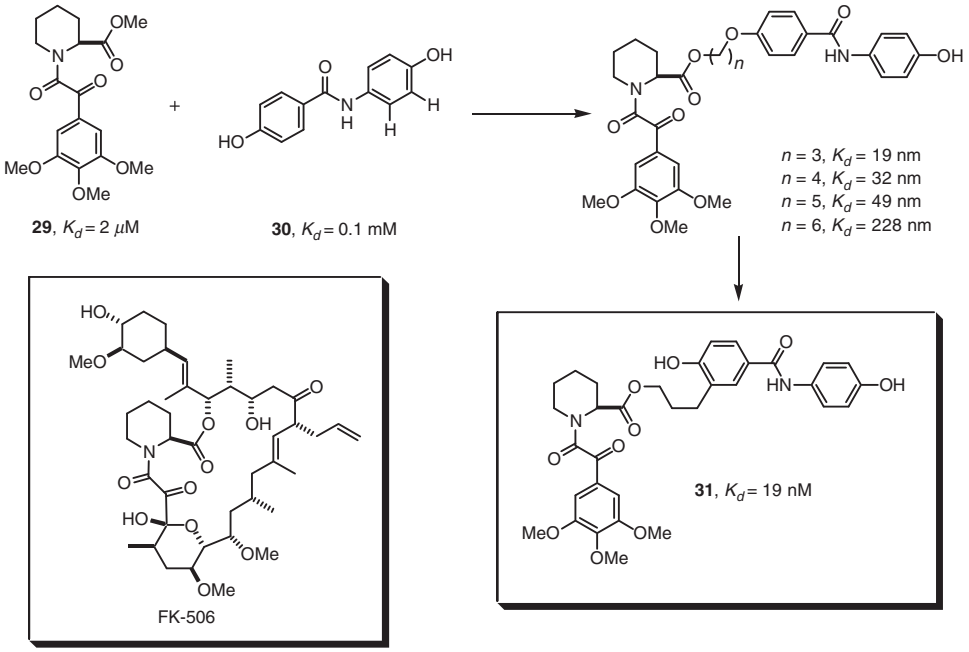


**Scheme 18** Development of high-affinity ligand for FK506 binding protein (FKBP).

provided compound **30**, which bound with a $K_d = 0.1$ mM. Subsequent analysis showed molecules **29** and **30** bound FKBP in sufficient proximity to allow for a chemical linkage to be made. Again, SAR by NMR experiments were performed using $^{15}$N-HSQC spectroscopy, connecting the two low-molecular-weight pieces with linkers of various lengths. One molecule (**31**) incorporating a linker of three carbons provided a ligand with an affinity of $K_d = 19$ nM (Scheme 18).

This novel approach toward high-affinity ligand discovery has recently been used by numerous research groups both in academia and industry to develop inhibitors for biological substrates such as stromelysin [91], the single-chain hepatitis C virus NS3 protease/NA4A cofactor [92], and human papillomavirus E2 protein [93]. As high-throughput NMR systems (including, e.g., HPLC-NMR) capable of very fast two-dimensional spectral acquisition become more widely available, we can anticipate that this technique will find broad usage.

## 6 LIBRARY DESIGN ISSUES

As described above, early combinatorial chemistry efforts grew initially out of solid-phase synthesis studies on biopolymers. Consequently, much of the early

history of combinatorial chemistry is focused on biopolymers and related compounds that could be formed using reactions closely analogous to those used in biopolymer synthesis. Although these approaches allowed the synthesis of large *numbers* of compounds, the actual *structural diversity* of such libraries was somewhat limited. Ellman's demonstration of the combinatorial synthesis of benzodiazepines (Scheme 6), along with several contemporary efforts, both demonstrated that combinatorial chemistry could be applied to the production of "druglike" molecules and led to the concept of the "decorated scaffold" as a means of providing conformationally rigid combinatorial libraries. While such libraries provided access to a conceptually new type of library structure, structural diversity within the library itself was still somewhat limited. In part because of these limitations, it was recognized that molecular weight and hydrophobicity of library compounds were straying beyond what might be considered optimal to lead to successful drugs. This observation was codified by Lipinski and co-workers into the "rule of 5," stating that a potential reason for the relatively low hit-to-lead success rate of many early combinatorial chemistry and high-throughput screening efforts was that libraries strayed beyond the structural norms (molecular weight less than 500, fewer than 5 hydrogen-bond donating groups, fewer than 10 hydrogen-bond accepting groups, log $P$ over 5) for the vast majority of commercial drugs [94]. Of course, one can also view the rule of 5 as a challenge for developing new compounds that might successfully break it, but the primary lesson was that it pays to think carefully about library design, structure space, and the overall goal (drug development, in this case) of the combinatorial/high-throughput screening effort.

In recent years, two parallel developments (one computational and one conceptual) have set the stage for improving library design. On the computational side [95], researchers in academia and industry have devoted considerable effort to the development of tools capable of describing *structural diversity* and *structural similarity* in mathematical terms [96, 97]. Most chemists are capable of describing the similarities and differences among small groups of compounds, but rigorous analysis of large libraries is a significantly more difficult task. For example, describing a 1000-compound library in terms of 4 simple descriptors (molecular weight, number of H-bond acceptors, number of H-bond donors, number of chiral centers) yields a four-dimensional graph with 1000 points! For complex libraries, more precise descriptors such as BCUT values [98, 99] or four-point pharmacophore fingerprints [100, 101] can be used along side computerized reagent selection methods in the planning stages of a library. In combination with computer-aided drug design methodologies, an area of significant interest is the "virtual" screening of libraries generated in silico.

In parallel with advances in computational methodology, the conceptual advance of *diversity-oriented synthesis* is also beginning to influence the manner in which combinatorial libraries are designed and executed [102]. Diversity-oriented synthesis (and the related concept of *function-oriented syn-*

*thesis* [103]) is in many respects a response to target-oriented synthesis. Target-oriented synthesis is perhaps the most familiar mode of thinking for organic chemists: Given a particular natural product or med-chem lead structure, how does one bring simple commercially available starting materials and the vast body of reaction knowledge available to the development of an efficient method for its construction? Target-oriented synthesis relies on retrosynthetic analysis (thinking about chemical synthesis in reverse). In contrast, diversity-oriented synthesis focuses on the development of pathways for the construction of collections of conformationally restricted molecules with complex, diversified structures, and dense functionality. Diversity-oriented synthesis is forward thinking in that reactions are chosen not based on their ability to bring the chemist closer to a single-target structure but rather contribute diversity (via an ability to accommodate multiple input reagents) and complexity (via an ability to form multiple bonds).

An example of a diversity-oriented synthesis developed by the Schreiber group is shown in Scheme 19 [104]. Ugi four-component condensation is used to set up an intramolecular Diels–Alder cycloaddition. Subsequent "bidirectional" amide alkylation with allyl bromide, followed by a tandem

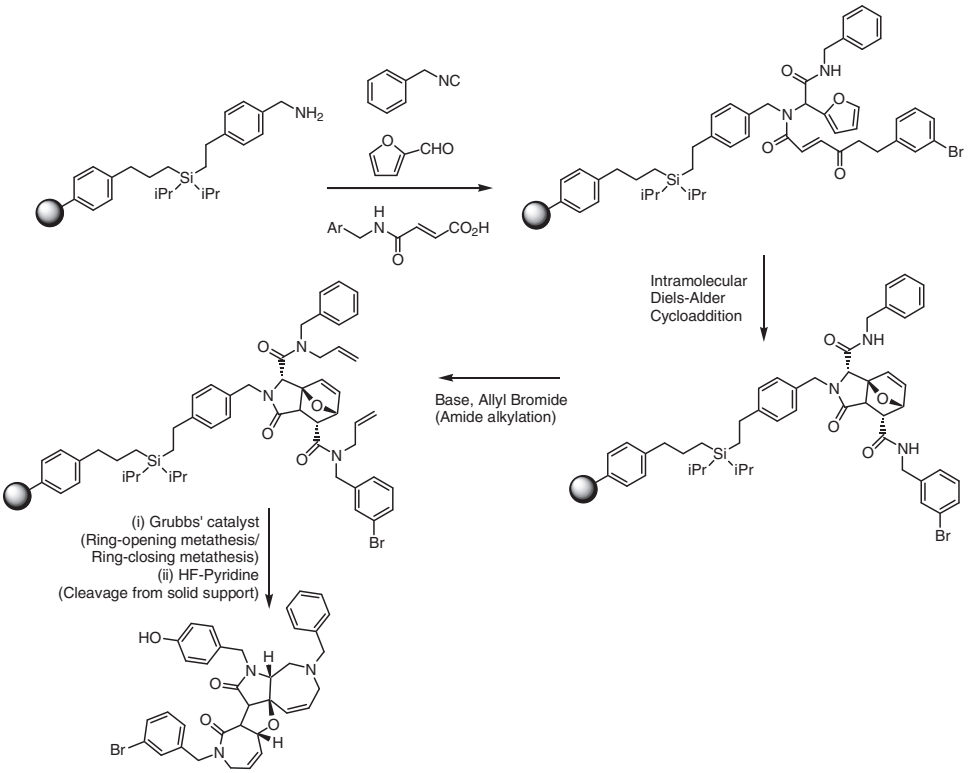


**Scheme 19** Diversity-oriented synthesis [104].

ring-opening metathesis/ring-closing metathesis reaction, provides the final product. Cleavage from the resin bead is then accomplished by treatment with HF-pyridine. Further diversification is possible via elaboration of the bromoarene or phenolic groups.

## 7 CONCLUSIONS

In a relatively short period of time, combinatorial chemistry has become a widely used tool for synthetic chemists looking to streamline lead identification and for those needing novel structures to drive a plethora of studies in chemistry, biology, materials science, and beyond. Combinatorial chemistry has proven to be a valuable method for streamlining the development of new catalysts [105, 106] and new types of arrayed sensors [107]. We can anticipate that combinatorial techniques will be an important part of every chemist's repertoire for the foreseeable future.

## REFERENCES

1. Rowinsky, E. K., Cazenave, L. A., Donehower, R. C. (1990). Taxon: A novel investigational antimicrotubule agent. *J Ntl. Cancer Inst.*, *82*, 1247–1259.
2. Sheridan, R. P., Venkataraghavan, R. (1987). New methods in computer-aided drug design. *Acc. Chem. Res.*, *20*, 322–329.
3. Crafford, C. (2003). The drug discovery gridiron. *Modern Drug Discov.*, 25–26.
4. Schroder, F. C., Farmer, J. J., Attygalle, A. B., Smedley, S. R., Eisner, T., Meinwald, J. (1998). Combinatorial chemistry in insects: A library of defensive macrocyclic polyamines. *Science*, *281*, 428–431.
5. Examples include: (a) Nicolaou, K. C., Hanko, R., Hartwig, W. (Eds.) (2002). *Handbook of Combinatorial Chemistry.* Wiley-VCH. (b) Gordon, E. M., Kerwin, J. F., Jr. (Ed.) (1998). *Combinatorial Chemistry and Molecular Diversity in Drug Discovery.* Wiley-Liss, New York. (c) Czarnik, A. W., DeWitt, S. H. (1997). *A Practical Guide to Combinatorial Chemistry*. American Chemical Society, Washington, DC.
6. Merrifield, R. B. (1963). Solid phase peptide synthesis. I. The synthesis of a tetrapeptide. *J. Am. Chem. Soc.*, *85*, 2149–2154.
7. Hodge, P. (1997). Polymer-supported organic reactions: What takes place in the beads? *Chem. Soc. Rev.*, *26*, 417–424.
8. Merrifield, R. B. (1964). Solid phase peptide synthesis. III. An improved synthesis of bradykinin. *Biochemistry*, *3*, 1385–1390.
9. Merrifield, R. B. (1965). Automated synthesis of peptides. *Science*, *150*, 178–185.
10. Letsinger, R. L., Mahadevan, V. (1965). Oligonucleotide synthesis on a polymer support. *J. Am. Chem. Soc.*, *87*, 3526–3527.
11. Frechet, J. M., Schuerch, C. (1971). Solid-phase synthesis of oligosaccharides. I. Preparation of the solid support. Poly[*p*-(1-propen-3-ol-1-yl)styrene]. *J. Am. Chem. Soc.*, *93*, 492–496.

12. Dudkin, V. Y., Orlova, M., Geng, X., Mandal, M., Olson, W. C., Danishefsky, S. J. (2004). Toward fully synthetic carbohydrate-based HIV antigen design: On the critical role of bivalency. *J. Am. Chem. Soc.*, *126*, 9560–9562.
13. Ritter, T. K., Mong, K.-K. T., Liu, H., Nakatani, T., Wong, C.-H. (2003). A programmable one-pot oligosaccharide synthesis for diversifying the sugar domains of natural products: A case study of vancomycin. *Angewante Chem. Int. Ed.*, *42*, 4657–4660.
14. Love, K. R., Seeberger, P. H. (2004). Automated solid-phase synthesis of protected tumor-associated antigen and blood group determinant oligosaccharides. *Angewante Chem. Int. Ed.*, *43*, 602–605.
15. Schnolzer, M., Kent, S. (1992). Constructing proteins by dovetailing unprotected synthetic peptides: Backbone-engineered HIV protease. *Science*, *256*, 221–225.
16. Zhang, Z., Alfonta, L., Tian, F., Bursulaya, B., Uryu, S., King, D. S., Schultz, P. G. (2004). Selective incorporation of 5-hydroxytryptophan into proteins in mammalian cells. *Proc. Natl. Acad. Sci. USA*, *101*, 8882–8887.
17. Holley, R., Apgar, J., Everett, G., Madison, J., Marquette, M., Merrill, S., Penswick, J., Zamir, A. (1965). Structure of a ribonucleic acid. *Science*, *147*, 1462–1465.
18. Belagaje, R., Brown, E., Gait, M., Khoana, H., Norris, K. (1979). Total synthesis of a tyrosine suppressor transfer RNA gene. XIII. Synthesis of deoxyribopolynucleotide segments corresponding to the nucleotide sequence –1 to –29 in the promoter region. *J. Biol. Chem.*, 5754–5763.
19. Goeddel, D., Kleid, D., Bolivar, F. (1979). Expression in *Escherichia coli* of chemically synthesized genes for human insulin. *Proc. Natl. Acad. Sci. USA*, *76*, 106–110.
20. Edge, M., Greene, A., Heathcliffe, G., Meacock, P., Schuch, W., Scalon, D., Atkinson, T., Newton, C., Markham, A. (1981). Synthesis of human leudocyte α1-interon gene. *Nature*, *292*, 756–760.
21. Hodge, P. (1997). Polymer-supported organic reactions: What takes place in the beads? *Chem. Soc. Rev.*, *26*, 417–424.
22. Frechet, J. M. J. (1981). Synthesis and applications of organic polymers as supports and protecting groups. *Tetrahedron*, *37*, 663–683.
23. Houghten, R. A. (1985). General method for the rapid solid-phase synthesis of large numbers of peptides: Specificity of antigen-antibody interaction at the level of individual amino acids. *Proc. Natl. Acad. Sci.*, *82*, 5131–5135.
24. Frank, R., Heikens, W., Heisterberg-Moutsis, G., Blocker, H. (1983). A new general approach for the simultaneous chemical synthesis of large numbers of oligonucleotides: Segmental solid supports. *Nucleic Acids Res.*, *11*, 4365–4377.
25. Bunin, B. A., Ellman, J. A. (1992). A general and expedient method for the solid-phase synthesis of 1,4-benzodiazepine derivatives. *J. Am. Chem. Soc.*, *114*, 10997–10998.
26. Lam, K. S., Salmon, S. E., Hersh, E. M., Hruby, V. J., Kazmierski, W. M., Knapp, R. J. (1991). A new type of synthetic peptide library for identifying ligand-binding activity. *Nature*, *354*, 82–84.
27. Yan, B. (2002). Single bead analysis in combinatorial chemistry. *Curr. Opin. Chem. Biol.*, *6*, 328–332.
28. Chait, B. T., Wang, R., Beavis, R. C., Kent, S. B. H. (1993). Protein ladder sequencing. *Science*, *262*, 89–92.

29. Lane, S. J., Pipe, A. (2000). Single bead and hard tag decoding using accurate isotopic difference target analysis-encoded combinatorial libraries. *Rapid Commun. Mass Spectrom.*, *14*, 782–793.
30. Geysen, H. M., Wagner, C. D., Bodnar, W. M., Markworth, C. J., Parke, G. J., Schoenen, F. J., Wagner, D. S., Kinder, D. S. (1996). Isotope or mass encoding of combinatorial libraries. *Chem. Biol.*, *3*, 679–688.
31. Erb, E., Janda, K. D., Brenner, S. (1994). Recursive deconvolution of combinatorial chemical libraries. *Proc. Natl. Acad. Sci.*, *91*, 11422–11426.
32. Brenner, S., Lerner, R. A. (1992). Encoded combinatorial chemistry. *Proc. Natl. Acad. Sci. USA*, *89*, 5381–5383.
33. Needels, M. C., Jones, D. G., Tate, E. H., Heinkel, G. L., Kochersperger, L. M., Dower, W. J., Barrett, R. W., Gallop, M. A. (1993). Generation and screening of an oligonucleotide-encoded synthetic peptide library. *Proc. Natl. Acad. Sci. USA*, *90*, 10700–10704.
34. Debaene, F., Mejias, L., Harris, J. L., Winssinger, N. (2004). Synthesis of a PNA-encoded cysteine protease inhibitor library. *Tetrahedron*, *60*, 8677–8690.
35. Ohlmeyer, M. H. J., Swanson, R. N., Dillard, L. W., Reader, J. C., Asouline, G., Kobayashi, R., Wigler, M., Still, W. C. (1993). Complex synthetic chemical libraries indexed with molecular tags. *Proc. Natl. Acad. Sci.*, *90*, 10922–10926.
36. Nestler, H. P., Bartlett, P. A., Still, W. C. (1994). A general method for molecular tagging of encoded combinatorial chemistry libraries. *J. Organic Chem.*, *59*, 4723–4724.
37. Nestler, H. P., Bartlett, P. A., Still, W. C. (1994). A general method for molecular tagging of encoded combinatorial chemistry libraries. *J. Organic Chem.*, *59*, 4723–4724.
38. Ni, Z.-J., Maclean, D., Holmes, C. P., Murphy, M. M., Ruhland, B., Jacobs, J. W., Gordon, E. M., Gallop, M. A. (1996). Versatile approach to encoding combinatorial organic syntheses using chemically robust secondary amine tags. *J. Med. Chem.*, *39*, 1601–1608.
39. Rahman, S. S., Busby, D. J., Lee, D. C. (1998). Infrared and Raman spectra of a single resin bead for analysis of solid-phase reactions and use in encoding combinatorial libraries. *J. Organic Chem.*, *63*, 6196–6199.
40. Krchnak, V., Padera, V. (2003). The encore technique: A novel approach to directed split-and-pool combinatorial synthesis. *Methods Enzymol.*, *369*, 112–124.
41. Shukla, R., Sasaki, Y., Krchnak, V., Smith, B. D. (2004). Identification of synthetic phosphatidylserine translocases from a combinatorial library prepared by directed split-and-pool synthesis. *J. Comb. Chem.*, *6*, 703–709.
42. (1) Wang, X., Zhang, J., Song, A., Lebrilla, C. B., Lam, K. S. (2004). Encoding method for OBOC small molecule libraries using a biphasic approach for ladder-synthesis of coding tags. *J. Am. Chem. Soc.*, *126*, 5740–5749. (2) Liu, R., Marik, J., Lam, K. S. (2002). A novel peptide-based encoding system for "one-bead one-compound" peptidomimetic and small molecule combinatorial libraries. *J. Am. Chem. Soc.*, *124*, 7678–7680.
43. Fenniri, H., Ding, L., Ribbe, A. E., Zyrianov, Y. (2001). Barcoded resins: A new concept for polymer-supported combinatorial library self-deconvolution. *J. Am. Chem. Soc.*, *123*, 8151–8152.

44. Fenniri, H., Chun, S., Ding, L., Zyrianov, Y., Hallenga, K. (2003). Preparation, physical properties, on-bead binding assay and spectroscopic reliability of 25 barcoded polystyrene-poly(ethylene glycol) graft copolymers. *J. Am. Chem. Soc.*, *125*, 10546–10560.
45. Moran, E. J., Sarshar, S., Cargill, J. F., Shahbaz, M. M., Lio, A., Mjalli, A. M. M., Armstrong, R. W. (1995). Radio frequency tag encoded combinatorial library method for the discovery of tripeptide-substituted cinnamic acid inhibitors of the protein tyrosine phosphatase PTP1B. *J. Am. Chem. Soc.*, *117*, 10787–10788.
46. Nicolaou, K. C., Xiao, X.-Y., Parandoosh, Z., Senyei, A., Nova, M. P. (1995). Radiofrequency encoded combinatorial chemistry. *Angewante Chem. Int. Ed.*, *34*, 2289–2291.
47. Nicolaou, K. C., Pfefferkorn, J. A., Mitchell, H. J., Roecker, A. J., Barluenga, S., Cao, G.-Q., Affleck, R. L., Lillig, J. E. (2000). Natural product-like combinatorial libraries based on privileged structures. 2. Construction of a 10,000-membered benzopyran library by directed split-and-pool chemistry using Nanokans and optical encoding. *J. Am. Chem. Soc.*, *122*, 9954–9967.
48. Fodor, S. P. A., Read, J. L., Pirrung, M. C., Stryer, L., Liu, A. T., Solas, D. (1991). Light-directed, spatially addressable parallel chemical synthesis. *Science*, *251*, 767.
49. Li, S., Bowerman, D., Marthandan, N., Klyza, S., Luebke, K. J., Garner, H. R., Kodadek, T. (2004). Photolithographic synthesis of peptoids. *J. Am. Chem. Soc.*, *126*, 4088–4089.
50. Khodursky, A. B., Peter, B. J., Cozzzarelli, N. R., Botstein, D., Brown, P. O., Yanofsky, C. (2000). DNA microarray analysis of gene expression in response to physiological and genetic changes that affect tryptophan metabolism in *Escherichia coli*. *Proc. Natl. Acad. Sci. USA*, *97*, 12170–12175.
51. Blackwell, H. E., Perez, L., Stavenger, R. A., Tallarico, J. A., Eatough, E. C., Foley, M. A., Schreiber, S. L. (2001). A one-bead, one-stock solution approach to chemical genetics: Part 1. *Chem. Biol.*, *8*, 1167–1182.
52. Clemons, P. A., Koehler, A. N., Wagner, B. K., Sprigings, T. G., Spring, D. R., King, R. W., Schreiber, S. L., Foley, M. A. (2001). A one-bead, one-stock solution approach to chemical genetics: Part 2. *Chem. Biol.*, *8*, 1183–1195.
53. Winssinger, N., Ficarro, S., Schultz, P. G., Harris, J. L. (2002). Profiling protein function with small molecule microarrays. *Proc. Natl. Acad. Sci. USA*, *99*, 11139–11144.
54. Uttamchandani, M., Walsh, D. P., Khersonsky, S. M., Huang, X., Yao, S. Q., Chang, Y.-T. (In press). Microarrays of tagged combinatorial triazine libraries in the discovery of small-molecule ligands of human IgG. *J. Comb. Chem.*
55. Adams, E. W., Ratner, D. M., Bokesch, H. R., McMahon, J. B., O'Keefe, B. R., Seeberger, P. H. (2004). Oligosaccharide and glycoprotein microarrays as tools in HIV glycobiology: Glycan-dependent gp120/protein interactions. *Chem. Biol.*, *11*, 875–881.
56. For representative examples see: (1) Miller, B., Karan, C. (2001). RNA-selective coordination complexes identified via dynamic combinatorial chemistry. *J. Am. Chem. Soc.*, *123*, 7455–7456. (2) Hochgurtel, M., Biesinger, R., Kroth, H., Piecha, D., Hofmann, M., Krause, S., Schaaf, O., Nicolau, C., Eliseev, A. V. (2003). Ketones as builing blocks for dynamic combinatorial libraries: Highly active neuraminidase inhibitors generated via selection pressure of the biological target. *J.*

*Med. Chem.*, *46*, 356–358. (3) Otto, S., Furlan, R., Sanders, J. (2002). Selection and amplification of hosts from dynamic combinatorial libraries of macrocyclic disulfides. *Science*, *297*, 590–593.

57. Huc, I., Krishe, M. J., Funeriu, D. P., Lehn, J. M. (1999). Dynamic combinatorial chemistry: Target H-bonding directed assembly of receptors based on bipyridine-metal complexes. *Eur. J. Inorganic Chem.*, 1415–1420.
58. Berl, V., Huc, I., Lehn, J., DeCaian, A., Fisher, J. (1999). Induced fit selection of a barbiturate receptor from a dynamic structural and conformational/configurational library. *Eur. J. Organic Chem.*, *11*, 3089–3094.
59. Eliseev, A., Nelen, M. (1997). Use of molecular recognition to drive chemical evolution: Controlling the composition of an equilibrating mixture of simple arginine receptors. *J. Am. Chem. Soc.*, *119*, 1147–1148.
60. Rowan, S., Hamilton, D., Brady, P., Sanders, J. (1997). Automated recognition, sorting, and covalent self-assembly by predisposed building blocks in a mixture. *J. Am. Chem. Soc.*, *119*, 2578–2579.
61. Whitesides, G., Lees, W. (1993). Equilibrium constants for thiol-disulfide interchange reactions: A coherent, corrected set. *J. Organic Chem.*, *58*, 642–647.
62. Miller, B., Klekota, B. (2003). Biphasic reaction vessel and method of use. 6,599,754, July 29.
63. Klekota, B., Hammond, M., Miller, B. (1997). Generation of novel DNA-binding compounds by selection and amplification from self-assembled libraries. *Tetrahedr. Lett.*, *38*, 8639–8642.
64. Karan, C., Miller, B. (2001). RNA-selective coordination complexes identified via dynamic combinatorial chemistry. *J. Am. Chem. Soc.*, *123*, 7455–7456.
65. For an early example, see: Pyle, A. M., Long, E. C., Barton, J. K. (1989). Shape-selective targeting of DNA by phenanthrenequinone diimine rhodium-III photocleaving agents. *J. Am. Chem. Soc.*, *111*, 4520–4522.
66. Balasubramanian, S., Whitney, A., Ladame, S. (2004). Templated ligand assembly by using G-quadruplex DNA and dynamic covalent chemistry. *Angewante Chem. Int. Ed.*, *43*, 1143–1146.
67. Schouten, J. A., Ladame, S., Mason, S. J., Cooper, M. A., Balasubramanian, S. (2003). G-quadruplex-specific peptide-hemicyanine ligands by partial combinatorial selection. *J. Am. Chem. Soc.*, *125*, 5594–5595.
68. Lehn, J., Ramström, O., Lohmann, S., Bunyapaiboonsri, T. (2004). Dynamic combinatorial carbohydrate libraries: Probing the binding site of concanavalin A lectin. *J. Eur. Chem.*, *10*, 1711–1715.
69. Page, D., Roy, R. (1997). Synthesis and biological properties of mannosylated starbust poly(amidoamine) dendrimers. *Bioconjugate Chem.*, *8*, 714–723.
70. Lehn, J., Bunyapaiboonsri, T., Ramström, O., Lohmann, S., Peng, L., Goeldner, M. (2001). Dynamic deconvolution of a pre-equilibrated dynamic combinatorial library of acetylcholinesterase inhibitors. *ChemBioChem*, *2*, 438–444.
71. Kazlauskus, R., Cheeseman, J., Corbett, A., Shu, R., Croteau, J., Gleason, J. (2002). Amplification of screening sensitivity through selective destruction: Theory and screening of a library of carbanic anhydrase inhibitors. *J. Am. Chem. Soc.*, *124*, 5692–5701.
72. Severin, K. (2004). The advantage of being virtual—Target-induced adaptation and selection in dynamic combinatorial libraries. *Eur. J. Chem.*, *10*, 2565–2580.

73. Corbett, P. T., Otto, S., Sanders, J. K. M. (2004). Correlation between host–guest binding and host amplification in simulated dynamic combinatorial libraries. *Eur. J. Chem.*, *10*, 3139–3143.
74. Hochguertel, M., Kroth, H., Piecha, D., Hofmann, M. W., Nicolau, C., Krause, S., Schaaf, O., Sonnenmoser, G., Eliseev, A. V. (2002). Target-induced formation of neuraminidase inhibitors from in vitro virtual combinatorial libraries. *Proc. Natl. Acad. Sci. USA*, *99*, 3382–3387.
75. James, I. W. (1999). Linkers for solid phase organic synthesis. *Tetrahedron*, *55*, 4855–4956.
76. Kolb, H., Finn, M., Sharpless, K. (2001). Click chemistry. *Angewante Chem. Int. Ed.*, *40*, 2004–2021.
77. Taylor, S., Taylor, A., Screiber, S. (2004). Synthetic strategy toward skeletal diversity via solid-supported, otherwise unstable reactive intermediates. *Angewante Chem. Int. Ed.*, *43*, 1681–1685.
78. Kolb, H., Sharpless, B. (2003). The growing impact of click chemistry on drug discovery. *Drug Discov. Today*, *8*, 1128–1137.
79. Lee, L., Mitchell, M., Huang, S., Fokin, V., Sharpless, B., Wong, C. (2003). A potent and highly selective inhibitor of human α-1,3-fucosytransferase via click chemistry. *J. Am. Chem. Soc.*, *125*, 9588–9589.
80. Huisgen, R. (1984). *1,3-Dipolar Cycloaddition Chemistry*. Wiley, New York, pp. 1–176.
81. Mock, W., Irra, T., Wepseic, J., Manimaran, T. (1983). Cycloaddition induced by cucurbituril. A case of Pauling principle catalysis. *J. Organic Chem.*, *48*, 3619–3620.
82. Lewis, W., Green, L., Grynszpan, F., Radic, Z., Carlier, P., Talyor, P., Finn, M., Sharpless, B. (2002). Click chemistry in situ: Acetylcholinesterase as a reaction vessel for the selective assembly of a femtomolar inhibitor from an array of building blocks. *Angewante Chem. Int. Ed.*, *41*, 1053–1057.
83. Sussman, J., Harel, M., Frowlow, F., Oefner, C., Goldman, A., Toker, L., Silman, I. (1991). Atomic structure of acetycholinesterase from *Torpedo californica*: A prototype acetylcholine-binding protein. *Science*, *253*, 872–879.
84. Radic, Z., Taylor, P. (2001). Interaction kinetics of reversible inhibitors and substrates with acetylcholinesterase and its fasciculin 2 complex. *J. Biol. Chem.*, *276*, 4622–4633.
85. Wei, J., Burkiak, J., Siuzdak, G. (1999). Desorption–ionization mass spectrometry on porous silicon. *Nature*, *401*, 243–246.
86. Erlanson, D., Lam, J., Wiesmann, C., Luong, T., Simmons, R., DeLano, W., Choong, I., Burdett, M., Flanagan, M., Lee, D., Gordon, E., O'Brien, T. (2003). *In situ* assembly of enzyme inhibitors using extended tethering. *Nat. Biotechnol.*, *21*, 308–314.
87. Erlanson, D., Braisted, A., Raphael, D., Rangal, M., Stroud, R., Gordon, E., Wells, J. (2000). Site-directed ligand discovery. *Proc. Natl. Acad. Sci. USA*, *97*, 9367–9372.
88. Erlanson, D., McDowell, R., O'Brien, T. (2004). Fragment based drug discovery. *J. Med. Chem.*, *47*, 3464–3482.
89. Homans, S. (2004). NMR spectroscopy tools for structure-aided drug design. *Angewante Chem. Int. Ed.*, *43*, 290–300.
90. Shuker, S., Hajdek, P., Meadows, R., Fesik, S. (1996). Discovering high-affinity ligands for proteins: SAR by NMR. *Science*, *274*, 1531–1534.

91. Shuker, S., Hajdek, P., Meadows, R., Nettesheim, D., Olejniczak, E., Sheppard, G., Steinman, D., Carrera, G., Marcotte, P., Severin, J., Walter, K., Smith, H., Gubbins, E., Simmer, R., Holzman, T., Morgan, D., Davidsen, S., Summers, J., Fesik, S. (1997). Discovery of potent nonpeptide inhibitors of stromelysin using SAR by NMR. *J. Am. Chem. Soc.*, *119*, 5818–5827.
92. Wyss, D., Arasappan, A., Senior, M., Wang, Y., Beyer, B., Njoroge, G., McCoy, M. (2004). Non-peptidic small molecule inhibitors of the single-chain hepatitis C virus NS3 protease/NS4A cofactor complex discovered by structure-based NMR screening. *J. Med. Chem.*, *47*, 2486–2498.
93. Shuker, S., Hajdek, P., Dinges, J., Miknis, G., Merlock, M., Middleton, T., Kempf, D., Egan, D., Walter, K., Robins, T., Holtzman, T., Fesik, S. (1997). NMR based discovery of lead inhibitors that block DNA binding of the human papillomavirus E2 protein. *J. Med. Chem.*, *40*, 3144–3150.
94. Lipinski, C. A., Lombardo, F., Dominy, B. W., Feeney, P. J. (1997). Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Delivery Rev.*, *23*, 3–25.
95. Evensen, E., Eksterowicz, J. E., Stanton, R. V., Oshiro, C., Grootenhuis, P. D. J., Bradley, E. K. (2003). Comparing performance of computational tools for combinatorial library design. *J. Med. Chem.*, *46*, 5125–5128.
96. Ghose, A. K., Viswanadhan, V. N. (Ed.) (2001). *Combinatorial Library Design and Evaluation.* Marcel Dekker, New York.
97. Martin, Y. C. (2001). Diverse viewpoints on computational aspects of molecular diversity. *J. Comb. Chem.*, *3*, 231–250.
98. Pirard, B., Pickett, S. D. (2000). Classification of kinase inhibitors using BCUT descriptors. *J. Chem. Inf. Comput. Sci.*, *40*, 1431–1440.
99. Gao, H. (2001). Application of BCUT metrics and genetic algorithm in binary QSAR analysis. *J. Chem. Inf. Comput. Sci.*, *41*, 402–407.
100. Lyne, P. D., Kenny, P. W., Cosgrove, D. A., Deng, C., Zabludoff, S., Ashwell, S., Wendoloski, J. J. (2004). Identification of compounds with nanomolar binding affinity for checkpoint kinase-1 using knowledge-based virtual screening. *J. Med. Chem.*, *47*, 1962–1968.
101. Manallack, D. T., Pitt, W. R., Gancia, E., Montana, J. G., Livingstone, D. J., Ford, M. G., Whitley, D. C. (2002). Selecting screening candidates for kinase and G protein-coupled receptor targets using neural networks. *J. Chem. Inf. Comput. Sci.*, *42*, 1256–1262.
102. Schreiber, S. L. (2000). Target-oriented and diversity-oriented organic synthesis in drug discovery. *Science*, *287*, 1964–1969.
103. Wender, P. A., Baryza, J. L., Brenner, S. E., Clarke, M. O., Craske, M. L., Horan, J. C., Meyer, T. (2004). Function oriented synthesis: The design, synthesis, PKC binding and translocation activity of a new bryostatin analog. *Curr. Drug Discov. Techn.*, *1*, 1–11.
104. Lee, D., Sello, J. K., Schreiber, S. L. (2000). Pairwise use of complexity-generating reactions in diversity-oriented organic synthesis. *Organic Lett.*, *2*, 709–712.
105. Francis, M. B., Jacobsen, E. N. (1999). Discovery of novel catalysts for alkene epoxidation from metal-binding combinatorial libraries. *Angewante Chem. Int. Ed.*, *38*, 937–941.

106. Garbacia, S., Touzani, R., Lavastre, O. (2004). Image analysis as a quantitative screening test in combinatorial catalysis: Discovery of an unexpected ruthenium-based catalyst for the Sonogashira reaction. *J. Comb. Chem.*, *6*, 297–300.
107. McCleskey, S. C., Griffin, M. J., Schneider, S. E., McDevitt, J. T., Anslyn, E. V. (2003). Differential receptors create patterns diagnostic for ATP and GTP. *J. Am. Chem. Soc.*, *125*, 1114–1115.

# HERBAL MEDICINES AND ANIMAL MODELS OF GASTROINTESTINAL DISEASES

C.H. Cho
*The University of Hong Kong, Hong Kong, China*

J.K.S. Ko
*School of Chinese Medicine, The Hong Kong Baptist University, Hong Kong, China*

1 EXPERIMENTAL MODELS OF GASTRIC LESIONS AND ULCERATION
Acute Gastric Mucosal Injury
Evaluation of Acute Gastric Lesions
Chronic Gastric Ulcer
Acute and Chronic Duodenal Ulcers
2 EXPERIMENTAL MODELS OF INFLAMMATORY BOWEL DISEASE (IBD)
Trinitrobenzene Sulfonic Acid/Ethanol
Dintrobezene Sulfonic Acid/Ethanol
Dextran Sulfate Sodium
Assessment of Colitis Severity
3 ANIMAL MODELS OF GASTROINTESTINAL CANCER
Gastric Cancer Models
Colorectal Cancer Models
4 HERBAL MEDICINES AS POTENTIAL THERAPEUTIC AGENTS FOR ULCERS AND MALIGNANT DISEASE IN THE GASTROINTESTINAL TRACT
General
Polysaccharides in General
Polysaccharides from Mushroom and Seaweed

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

Polysaccharides from Other Plants
Phenolic Compounds and Terpenoids from Medicinal Plants
References

## 1 EXPERIMENTAL MODELS OF GASTRIC LESIONS AND ULCERATION

Animal models of gastric and duodenal ulcers are the key in studying the pathogenesis of gastrointestinal (GI) diseases and testing antiulcer agents. They provide essential tools to screen herbal medicines for their antiulcer and anticancer effects. These models ideally should mimic human diseases, be easy to produce, and reproducible to allow continuous replication in various laboratories. Tables 1 and 2 summarize the methodologies for these animal models.

### Acute Gastric Mucosal Injury

These animal models are chemically or psychologically induced and cause rapidly developing diffuse surface epithelial injury in the gastric mucosa. This is evident by multiple hemorrhagic erosions. The lesions are usually induced by intragastric administration of chemicals (e.g., concentrated ethanol, HCl, or nonsteroidal anti-inflammatory drugs) or by psychological stress (e.g., cold restraint) in fasted rats. Among these models, ethanol and nonsteroidal anti-inflammatory drugs are probably the most widely used for testing

**TABLE 1 Models of Mucosal Damage (Acute and Chronic) in the Stomach**[a]

| Damaging Agents | Dose and Route of Administration | Animals | Type of Damage | |
|---|---|---|---|---|
| | | | Acute | Chronic |
| Ethanol | 50–100%, 1 mL, IG | Rat | Erosion | — |
| HCl | 0.5–0.5 N, 1 mL, IG | Rat | Erosion | — |
| Aspirin (acidified with 0.2 N HCl) | 50–100 mg/kg, IG | Rat, mouse | Erosion to ulcer | — |
| Indomethacin | 10–50 mg/kg, IG or SC | Rat | Erosion to ulcer | — |
| Stress | Restraint + cold (4°C) for 2 h | Rat | Erosion to ulcer | — |
| Acetic acid | Local or topical application of acetic acid (dosage in text) | Rat | — | Ulcer |

[a] All animals should be fasted at least 18 h before induction of gastric damage. IG = intragastric administration; SC = subcutaneous injection.

**TABLE 2 Models of Duodenal Ulcers**[a]

| Damaging Agents | Dose and Route of Administration | Species |
|---|---|---|
| Secretogogues (histamine, carbachol, pentagastrin) | Carbachol 0.5 μg/kg, SC for 24–48 hr<br>+<br>Histamine 50 μg/kg/min, SC for 24–48 h | Rat |
| | Carbachol 0.5 μg/kg/min, SC for 24–48 h<br>+<br>Pentagastrin 2 μg/kg/min, SC for 24–48 h | Rat |
| | Histamine 90 μg/kg, SC × 8, 30 min apart | Guinea pig |
| Cysteamine HCl | 280–300 mg/kg, IG × 3 (e.g., 9 a.m., 12 noon, 3 pm) or 400 mg/kg, IG × 2 | Rat or mouse |

[a] IG = intragastric administration; SC = subcutaneous injection.

antiulcer agents in animals. Most acute erosions in the stomach induced by these methods heal rapidly in a few days without scar formation. These lesions do not reulcerate spontaneously, an enigmatic feature of ulcer disease in humans.

### Evaluation of Acute Gastric Lesions

Lesions can be evaluated semiquantitively in the gastric mucosa by using a grade (e.g., scale 0 to 3) or quantitatively by planimetry of the hemorrhagic areas [1]. Since the lesions are usually uniform in width and depth, some investigators measure the length of hemorrhagic erosions. Indirect assessment of mucosal lesions is also possible by measuring the amount of blood loss into the gastric lumen either as $^{51}$Cr-labeled red blood cells [2] or as the concentration of hemoglobin [3]. Biochemical measurement of released enzyme has also been used as the endpoint of evaluation [4].

### Chronic Gastric Ulcer

The acetic acid method, although artificial, represents one of the few reproducible models of chronic gastric ulcer. There are several methods of applying the acetic acid to the stomach, the most popular involving injection of a small volume (0.05 mL) of 20 to 30 percent (v/v) acetic acid into the submucosal layer through the serosa of the anterior wall of the glandular stomach in fasted rats. The site of injection is readily recognized by the localized swelling and blanching of the serosa at the injection site. Some modifications consist of topical application of 0.06 mL of 100 percent acetic acid into a mold over the serosa and allowing it to remain for 60 s [5], or luminal application of 60 percent (v/v) of acetic acid (0.12 mL) with the help of a syringe and a ring clamped between the anterior and posterior walls of the stomach for 45 s and withdrawn through the same syringe [6]. All these methods produce an ulcer of unified size after 24 h and that can be maintained for at least 14 days. The ulcer can be easily quantified by measuring the diameter or the area of the

ulcer crater. With the acetic acid model, histological evaluation is essential to assess the degree of necrosis, inflammation, epithelial regeneration, and proliferation of fibrous connective tissue, as well as angiogenesis at the ulcer margin and base.

### Acute and Chronic Duodenal Ulcers

The duodenal ulcer models fall into two groups: those produced by secretogogues that induce multiple, irregular ulcers and those caused by cysteamine-like agents that induce solitary or kissing ulcers (Table 2). Hydrochloric acid seems to be essential for duodenal ulceration. Experimental duodenal ulcers can be induced in rats and guinea pigs by giving secretogogues such as histamine, cholinergic drugs, and pentagastrin [7, 8], which are similar to Zollinger–Ellison syndrome or other human hypersecretory states. The severity of the ulcer may be evaluated semiquantitatively with a scale of 0 to 3, where 0 = normal mucosa, 1 = erosion, 2 = deep ulcer involving the muscularis propria, or 3 = penetrating ulcer (into liver or pancreas). The intensity of the ulcer process and the degree of inflammation and healing should be evaluated by light microscopy.

The duodenal ulcers induced in rats by cysteamine and its derivatives are very similar to those seen in humans, on the basis of functional and morphologic criteria [9]. Ulcers can be induced by this method in both rats and mice. It should be emphasized that animals should not be fasted because food restriction increases the toxicity of the ulcer-inducing agent. The ulcers develop mainly on the antimesenteric or anterior wall of the duodenum. They frequently perforate or penetrate into the liver or pancreas. These ulcers are sharply demarcated craters with a necrotic center. Measurement of the largest diameter of the ulcer can be used to calculate the ulcer area. The same 0 to 3 scale as for ulcers induced by secretogogues provides a semiquantitative evaluation of ulcer severity.

## 2 EXPERIMENTAL MODELS OF INFLAMMATORY BOWEL DISEASE (IBD)

Ample evidence suggests that the mechanisms underlying initiation, progression, and chronicity of IBD are related to the immune system, genetic susceptibility, microbiologic aspects, and the environment. In order to address all these issues, an ideal experimental model of IBD is needed for studying the early events, dissection of the interactions among different inflammatory components, and identification of immunological processes and genes involved. However, to date, no such single and ideal IBD model exists that exactly reflects the complexity of these multifactorial disorders. Nevertheless, there are many experimental models of IBD that work differentially to simulate its various forms, such as ulcerative colitis (UC) and Crohn's disease (CD) in humans [10]. Despite inherent limitation in their similarity to human forms of IBD, studies

of experimental colitis models have advanced our understanding of its pathogenesis [10, 11]. They also provide opportunities to evaluate potential therapeutic agents for this disease in the colon [12]. Current IBD models include chemical- and hapten-induced colitis, which are used in the current experimental settings to assess the efficacy of pharmacological agents in animals.

### Trinitrobenzene Sulfonic Acid/Ethanol

2,4,6-Trinitrobenzene sulfonic acid (TNBS)/ethanol-induced colitis is the most commonly used animal model in various laboratories [11] and was initially established by Morris and co-workers in 1989 [13]. Intrarectal administration of 30 mg TNBS in ethanol (50 percent, v/v) vehicle is injected slowly through a polythene catheter via the rectum into the lumen of the colon, 8 cm proximal to the anus. This results in acute inflammation with ulcers that evolves into chronic inflammation of the distal colon in rats [13] and mice [14]. This approach is based on the hypotheses that in CD, an increase in mucosal permeability results in entry to the lamina propria of a luminal antigen that is not adequately cleared by the mucosal immune system. TNBS is a hapten that, when coupled to a substance of high molecular weight (such as tissue proteins), elicits immunological responses. Ethanol is known as a "barrier breaker" that causes widespread acute mucosal damage when instilled into the distal colon of rats at a concentration of at least 30 percent. The combined administration of TNBS and ethanol results in a severe, transmural, granulomatous inflammation of the distal colon. Once induced, healing is continuous after the initial injury and there are no relapses. The duration of inflammation lasts for at least 8 weeks [10]. It has been well characterized and its macroscopic and microscopic features share many of the histopathological and clinical features of human CD [11, 15, 16].

### Dinitrobenzene Sulfonic Acid/Ethanol

2,4-Dinitrobenzene sulfonic acid (DNBS), a hapten molecule, was developed to produce inflammation closely resembling human UC [16]. Its action is based on the hypersensitivity of intestinal mucosa toward antigen that also occurs in IBD development. Again in this model, 50 percent ethanol is used to remove the protective mucus coating together with 30 mg DNBS given as an enema inserted intrarectally into an anesthetized rat about 8 cm proximal to the anus to induce colitis. In the DNBS model, numerous colonic lesion and other characteristics of UC can be found. Although the inflammatory responses in the DNBS model are similar to those induced by other haptens, such as TNBS [13], granulomatous lesions are not found.

### Dextran Sulfate Sodium

Dextran sulfate sodium (DSS), a sulfated polysaccharide of molecular size of 54,000, can be used to induce acute and chronic colitis in rats and mice. This

model is highly reproducible and resembles human UC [17]. To induce acute colitis, animals are given 3 to 10 percent DSS in drinking water for 7 to 9 days. By the end of the treatment period, animals show signs of diarrhea and gross rectal bleeding. Colon length is reduced and its weight is increased indicating that an inflammatory type of edema occurs. In addition, histological tissue damage includes multiple mucosal erosion and crypt abscesses. For chronic inflammatory studies, animals are treated with 5 percent DSS in drinking water for 7 days followed by a rest period of 7 to 10 days. This cycle of treatment is repeated several times to produce a relapse/remission pattern of inflammation in the colon [18]. In this repeated process of inflammatory injury, cell turnover is rapid and massive, and genetic damage may be caused by random mutation. Faulty genes may appear and as a result dysplasia appears in the mucosa, a clear precursor to malignancy. Low-grade and high-grade dysplasia is accompanied by adenoma followed by carcinoma in the colon [19]. This is a typical inflammation–adenoma–carcinoma model in animals, which is useful for screening both anti-inflammatory and anticancer agents in animals.

## Assessment of Colitis Severity

***Morphological Assessment*** To measure mucosal damage in TNBS and DNBS models, the distal colon is removed, opened, and rinsed thoroughly in ice-cold normal saline. The lesion area is recorded and then measured using a grid. In the assessment of the inflammatory response, 8 cm of the distal colon or the whole colon is weighed. The ratio of colon to body weight is used to assess the degree of colon edema. The whole colon length is also measured to indicate the severity of colonic inflammation [19, 20].

***Histopathological Assessment*** In the DSS colitis model, a longitudinal section of colonic tissue is obtained and fixed in 10 percent formalin. The tissue is processed and embedded in a paraffin block, which is then sectioned at 5 μm and stained with the hematoxylin-eosin to assess damage. The severity of damage is graded according to the system described by Okayasu and co-workers [17], in which 0 = normal; 1 = focal inflammatory cell infiltration including polymorphonuclear leukocytes; 2 = inflammatory cell infiltration, gland dropout, and crypt abscess; and 3 = mucosal ulceration.

In the inflammation produced by TNBS and DNBS, tissue processing and staining are the same as for the DSS model. Microscopic damage is measured as follows: 0 = colon with normal architecture; 1 = damage limited to surface epithelium; 2 = focal ulceration limited to the mucosa; 3 = focal and transmural ulceration and inflammation; 4 = extensive transmural lesion bordered by normal mucosa; and 5 = extensive transmural lesion involving the entire section [21].

***Biochemical Assessment*** Colonic tissues are collected and immediately frozen in liquid nitrogen and stored at –70°C until determination for

myeloperoxidase (MPO) activity according to the method described by Suzuki and coworkers [22]. The supernatant of a homogenate is mixed with hydrogen peroxide and 3,3′,5,5′-tetramethylbenzidine prepared in sodium phosphate buffer (pH 5.4). The final mixture is measured at 450 nm using a spectrophotometer with horseradish as standard. The value of MPO activity is represented as U/g tissue. This enzyme is also taken as an index of neutrophil infiltration.

## 3 ANIMAL MODELS OF GASTROINTESTINAL CANCER

### Gastric Cancer Models

***Human Gastric Cancer Cell Implantation Model*** Direct injection of human gastric cancer cells into the gastric wall is a recently adopted method designed to induce tumor growth in situ in athymic nude mice (Table 3). In this method, a population of $10^6$ gastric cancer cells (either KKLS, MKN-28, or TMK-1) are being injected into the stomach wall in anesthetized animals. Sizable tumors can be notified after a few weeks to a month depending on the carcinogenic potential of the human gastric cancer cell lines [23–25].

***Chemical Induction Model*** Chemical carcinogens may exert their effect directly on contact at the point of application or may require biochemical activation by the host before becoming carcinogenic. Many of these activation reactions are controlled by both endogenous or genetic factors and exogenous environmental elements. Some carcinogens cause cancer in animal models after a single dose, whereas other chemicals require multiple exposures. The incidence and latent period before overt cancer is found may be a function of all these variables. Carcinogens are distinct from other drugs and chemicals

**TABLE 3 Models of Gastric Cancer**

| Type of Model | Cell Types or Carcinogens | Animals | Method | Duration of Induction |
|---|---|---|---|---|
| Cancer cell implantation model | Human gastric cancer cells (KKLS, MKN-28, TMK-1) | Athymic nude mice | Direct injection of cells ($10^6$) into the gastric wall | Few weeks |
| Chemical induction model | *N*-methyl-*N*-nitrisourea (MNU) | Balb/c mice | 240 ppm in drinking water | 50 weeks |
| | *N*-methyl-*N*′-nitro-*N*-nitrosoguanidine (MNNG) | Wistar rats | 25 μg/ml in drinking water | 52 weeks |

in that they may not cause acute toxic effects and their adverse action may require a long latent period. In animals, certain carcinogens given in large quantity may require a relatively short period of time. Some examples of these are illustrated in Table 3 and are discussed below.

Balb/c mice are given 240 ppm *N*-methyl-*N*-nitrosourea (MNU) in drinking water daily during alternative weeks for a total of 10 weeks and killed at the end of the 50th week. More than 70 percent survive and about 80 percent have adenocarcinoma of both differentiated and undifferentiated types [26].

Male Wistar rats are treated with 25 μg/mL *N*-methyl-*N′*-nitro-*N*-nitrosoguanidine (MNNG) in tap water for 52 weeks. About 60 percent of the animals develop differentiated adenocarcinomas in the glandular stomach at the antral site [27]. Higher concentrations of MNNG produce higher incidences of cancer [28].

## Colorectal Cancer Models

***Inflammation-Associated Model*** Experimental studies indicate that dextran sulfate sodium in drinking water either at a concentration of 5 percent (w/v) for 63 days or 180 days or combined with exposure to cigarette smoke generate dysplasia and/or adenocarcinoma in about 30 to 80 percent of the mice [17, 19, 29, 30]. This ulcerative colitis-associated colonic adenoma formation in mice provides an animal model closer to humans in whom with inflammation-associated neoplasia is one of the major pathways in the development of colorectal cancer. This animal model is also useful to study the relationship between colorectal cancer and cigarette smoking, which is one of the major risk factors for colorectal cancer in humans [19, 30].

***Genetically Modified Model*** Mice with mutation of the *APC* gene or defects in the *TCTβ* and *p53* genes, have been developed to generate high percentages of animals with multiple intestinal adenomas or nodular masses in the cecum to the proximal colon by 4 months, respectively [31, 32].

***Chemical Induction Models*** Cancers of the small bowel and colon can be induced by several carcinogens depending on route of administration (Table 4). The most effective carcinogens appear to be the natural product cycasin and the synthetic analogs 1,2-dimethylhydrazine (DMH) and azoxymethane (AOM) [33]. In rats, AOM can be given by subcutaneous (SC) injection once weekly for 2 weeks at a dose of 15 mg/kg body weight. Rats are killed 8, 23, and 38 weeks after the last dose of AOM to assess tumor growth and invasiveness. Most of the colon tumors obtained at week 38 are invasive or non-invasive adenocarcinomas [34].

The DMH requires metabolic activation, and intestinal microflora may play a role in the metabolism of this colon carcinogen. Both mice and rats are commonly used studying colon cancer induced by DMH. In mice, 30 mg/kg DMH is injected SC once weekly for 6 weeks [35]. However, in rats, a smaller dose (25 mg/kg) is administered once a week for 16 weeks [36]. They are killed

**TABLE 4 Models for Carcinogenesis in Large Bowel**[a]

| Chemical | Animal | Route | Organ Involved |
|---|---|---|---|
| Cycasin | Rat | Oral | Large bowel |
| 1,2-Dimethylhydrazine | Rat, mouse, hamster | Oral, SC, IP, IR | Small and large bowel |
| Azoxymethane | Rat, mouse | Oral, SC, IR | Small and large intestine |
| *N*-methyl-*N′*-nitro-*N*-nitroguanidine | Rat | IR | Large intestine |

[a] Routes of administration: oral (OR), subcutaneous (SC), intraperitoneal (IP), and intrarectal (IR).

5 to 12 weeks after the last dose of the carcinogen for histopathological study. The DMH-induced colon tumors are very close to human colon cancer with regard to morphology, pattern of growth, and clinical manifestations.

The direct alkylating agent, MNNG, which does not require metabolic activation, is a potent topical carcinogen. Intrarectal administration of MNNG at a dose of 1 to 3 mg/rat/week for 20 weeks induces colon tumors in 100 percent of rats [37]. This is the most reliable model for the topical and selective production of tumors in the distal colon and rectum. The use of rats, especially the Fischer (F344) strain, seems to be appropriate because rat colons have light and electron microscopic morphologies as well as histochemical properties that are quite similar to those of humans.

## 4 HERBAL MEDICINES AS POTENTIAL THERAPEUTIC AGENTS FOR ULCERS AND MALIGNANT DISEASES IN THE GASTROINTESTINAL TRACT

The experimental ulcer models in animals discussed above provide a platform for studying the mechanisms of action and for screening potential antiulcer and anticancer agents derived from different herbal medicines. In fact, a number of herbal medicines are known to protect against a variety of GI disorders including those in the upper and lower GI tract described above. These medicines contain various chemical structures, have different properties, and are derived from numerous sources. In this chapter we focus on three major types of chemical compounds from various kinds of herbs: polysaccharides, terpenoids, and polyphenols. Their implications for different GI disorders and their mechanisms of action are discussed.

### General

The three biological processes that play key roles in maintaining the integrity and balancing the growth of tissues are cell proliferation, apoptosis, and angiogenesis. The degrees of ulcer healing and tumor growth are greatly influenced

by the levels of these biological activities. We focus on how the different herbal compounds can modulate these processes to promote ulcer healing and inhibit tumor growth in the GI tract. Many herbal medicines contain a large amount of polysaccharides and thousands of phenolic and terpenoid compounds. In this chapter, perspectives on the application of polysaccharides, terpenoids, and phenolic compounds from herbal extracts of different medicinal plants that have demonstrated antiulcer and anticarcinogenic properties in the GI tract are discussed. In addition, some novel herbs that possess anti-inflammatory and/or immunomodulatory actions and have antitumor potential are introduced.

## Polysaccharides in General

Polysaccharides from various herbal medicines have been well studied for their immunomodulatory action and antitumor effect. These actions and effects may be due to stimulation of the phagocytic activity and anti-inflammatory activity. They also increase hematopoiesis in the bone marrow and promote the proliferation of several types of hematopoietic precursor cells. These findings suggest that polysaccharides of plant origins have significant pharmacological actions on somatic cells with high proliferative capacity. Recently, increasing evidence implicates that plant polysaccharides have protective and repairing activities in the GI tract similar to those from the cells in this system; they also share similar biological properties.

There are various sources of polysaccharides with different compositions. Table 5 summarizes the sources of polysaccharides and their pharmacological actions on the GI tract. These polysaccharides have different molecular structures and are mostly sulfated molecules with uronic acid as part of the

**TABLE 5 Sources of Polysaccharides and Their Pharmacological Actions on the Gastrointestinal Tract and Liver**

| Sources | Pharmacological Actions | References |
|---|---|---|
| *Angelica sinensis* | Antiulcer and antitumor in the stomach and liver | 55, 59 |
| | Promote gastric cell migration and proliferation | 56, 57 58 |
| | Antihepatic damage | 59 |
| *Asparagus racemosus* | Antioxidant in hepatic cells | 60 |
| *Aloe vera* | Anti-inflammatory bowel disease | 62 |
| *Bupleurum falcatum* | Antiulcer in the stomach | 52, 53 |
| *Basidiomycetes* mushroom | Antitumor in the stomach and colon | 47, 48 |
| Panax ginseng | Antiulcer in the stomach | 54 |
| *Ulva lactuca* (green seaweed) | Inhibits colon cancer cell proliferation | 40 |
| *Cladosiphon fucoidan* | Gastric protection and *H. pylori* detachment from mucin | 51, 52 |

residue. The sulfated polysaccharides, which contribute to the active components in extracts, produce a variety of pharmacological actions [38–40]. Desulfation abolished their inhibitory action on the proliferation of arterial smooth muscle cells [41]. Some of the actions of these compounds are partially dependent on the size of the polysaccharide molecules and the length of the polysaccharides determining the effectiveness of drugs in different biological systems [42, 43]. Experimental evidence shows that the β-1,3 or β-1,6 linked polysaccharides are responsible for the different pharmacological actions on vascular and tumor cells, while their analog, the α-1,4 linkage can be digested by the saccharidases to monomers in the GI tract [44, 45]. Indeed, the β-glucan receptors were first identified on human monocytes as phagocytic receptors that initiate phagocytosis [46].

## Polysaccharides from Mushroom and Seaweed

The orally bioactive glucans and proteoglycans isolated from mushrooms are currently the most promising class of immunoceuticals. They are capable of augmenting the key pathways of host immunity. Higher *Basidiomycetes* mushrooms have been used in folk medicine throughout the world since ancient times. It has been known for many years that certain mushrooms of higher *Basidiomycetes* origin are effective against cancers of the stomach and esophagus [47, 48]. However, the active components responsible for these activities are largely unknown. Several antitumor polysaccharides such as hetero-β-glucans and their protein complexes have been isolated from medicinal mushrooms. Using standard methods of fractionation and purification of polysaccharides, Chihara isolated a water-soluble antitumor polysaccharide from the fruiting bodies of *Lentinus edodes* in 1978 [49]. This was found to be a 1,3-β-D-glucan with 1,6-β-D-glucopyranoside branches. Clinical data suggest that these lentinan polysaccharides can prolong the life span of patients with advanced and recurrent stomach and colorectal cancer, while displaying little toxicity [44]. Lentinan's antitumor activity is stronger than that of polysaccharides from other fungi or higher plants.

Two proteoglycans from *Coriolus versicolor*—the polysaccharides-K (PSK) and polysaccharides-peptide (PSP) have been demonstrated to extend survival by 5 years or beyond in patients with cancers of the stomach, colon, and esophagus. PSP significantly improves quality of life, provides substantial pain relief, and enhances immune status in a majority of patients with cancers of the stomach, esophagus, and other organs. The uses and the mechanisms of action of these mushroom glucans and proteoglycans for GI cancer were reviewed by Kidd in 2000 [50]. The major action is to enhance anticancer immunity by increasing the numbers of tissue and blood-borne phagocytic cells. Both PSK and PSP may function as antigenic stimuli to enhance cytotoxic killer activity directed at the tumor.

Seaweed provides another source of polysaccharides. The green seaweed *Ulvans lactuc* contains sulfated polysaccharides that can inhibit a colonic

epithelial cancer cell line but have no effect on normal colonocytes. It has been reported that sulfated polysaccharides from *Cladosiphon fucoidan* confer gastric protection and are associated with *Helicobacter pylori* detachment from gastric mucin [51, 52]. These data suggested that they may be useful for patients with *H. pylori*–positive gastric ulcers by reducing the risk of gastric cancer.

## Polysaccharides from Other Plants

The acidic polysaccharide fraction from the roots of *Bulpeurum falcatum*, irrespective of the route of administration, inhibits the formation of gastric lesions induced by necrotizing agents such as ethanol and HCl-ethanol [53]. The protection is due to antisecretory activity in the stomach, stimulation of the defensive mechanism through mucus secretion, and a radical scavenging effect in the gastric mucosa. However, such protection was independent of prostaglandin [53, 54]. Similarly, acidic polysaccharides from the leaves of Panax ginseng also inhibit the formation of gastric lesions through the same mechanisms [55].

Recent studies showed that a crude extract from the roots of *Angelica sinensis* mainly consisting of polysaccharides, prevents neutrophil-dependent ulcers, such as indomethacin- and ethanol-induced damage in rat stomachs [56]. It also enhances ulcer healing in the gastric mucosa, perhaps through its anti-inflammatory action. In normal rat gastric epithelial cells, the same type of extract stimulates cell proliferation and migration [56–58], the two major types of biological processes responsible for ulcer healing. Furthermore, it was demonstrated that the action is mediated in part through stimulation of the expression of epidermal growth factor and *c*-myc at the ulcer site followed by activation of ornithine decarboxylase activity [57, 58]. The protection extends to the liver damage provoked by paracetamol and carbon tetrachloride. This effect is partially mediated through the inhibition of nitric oxide synthase in hepatocytes [59]. Polysaccharides from the same herb also inhibit the invasion and metastasis of hepatocellular carcinoma cells in vitro [60]. In addition to the polysaccharides from *A. sinensis*, other natural polysaccharides from plants have protective effects on the liver. For example, a purified aqueous fraction from *Asparagus racemosus* shows potent antioxidant properties in mitochondrial membranes isolated from rat liver cells [61] and an aqueous extract from the seeds of *Celosia argentea*, containing an acidic polysaccharide, protects against chemically and immunologically induced hepatitis [62]. Recently, an aloe vera gel, a mucilaginous aqueous extract, was shown to have therapeutic potential in inflammatory bowel disease in humans [63].

## Phenolic Compounds and Terpenoids from Medicinal Plants

***Curcumin and Phenolic Compounds*** Numerous phenolic substances present in fruits or medicinal plants have been found to possess potential

cancer chemopreventive activities [64]. One of these is curcumin, a yellow pigment isolated from the rhizome of *Curcuma longa* L., Zingiberaceae, which is extensively used for imparting color and flavor to food. Although it has been used as a chemopreventive agent in Asia for years, investigations of its mode of action remain limited. Turmeric, the powdered form of *C. longa*, is a traditional medicinal remedy that exerts an antioxidant effect with antimutagenic activity [65]. One of the likely reasons for its action is the phenolic nature of its active ingredient curcumin (diferuloylmethane) and the other curcuminoid derivatives in the plant extract, such as feruloyl-*p*-hydroxy-cinamoylmethane and *bis*-(*p*-hydroxy-cinamoyl) methane [66]. Among these, curcumin has been extensively studied as a chemopreventive agent against cancers in various models, including its cytotoxicity to human HCT-15 and HT-29 colon cancer cells [67]. Curcumin is known to have immunomodulatory functions. Studies revealed upregulation of intestinal mucosal $CD4^+$ T cells and B cells by curcumin, which could be essential for its tumor-inhibitory activity [68].

Chemopreventive activity of synthetic curcumin had been observed in colon carcinogenesis models [69]. In addition, it had demonstrated anticarcinogenic effects on papillomas and squamous cell carcinomas of the forestomach, as well as on adenomas and adenocarcinomas of the duodenum in mice [70]. Inhibition of cell proliferation and induction of apoptosis via the Fas signaling pathway had been suggested to be the events responsible for such effects [71]. Given its specific mode of action, some suggested a correlation with the modulation of arachidonic acid metabolism [72, 73], while others had proposed the involvement of nuclear transcriptional factors like AP-1 (activator protein-1) and nuclear factor-kappaB (NF-κB) [74]. Recent data suggested that IκBα, the inhibitory unit of NF-κB, could be cleaved by caspase-3, followed by subsequent NF-κB activation and eventually promote tumor cell death [75]. Other than the transcriptional deactivation pathway, curcumin had also demonstrated an inhibitory action on intracellular signals from protein kinase C [76]. All these phenomena imply that inhibition of colon carcinogenesis by curcumin involves modulation at different cellular and molecular levels. As a matter of fact, the steps in deregulating signaling pathways and/or directly modulating gene transcription could theoretically be targeted as potential sites for chemopreventive intervention. *Curcuma* extract can be administered safely to colorectal cancer patients at doses of up to 2.2 g daily, which is equivalent to about 180 mg of curcumin [77]. With a clear understanding of the precise mechanistic pathways of the chemopreventive actions exerted by the curcuminoid compounds, *Curcuma* extract and curcumin will certainly play a permissive role in the development of target-oriented agents against colon carcinogenesis and the pathogenesis of other GI cancers.

***Licorice and Its Constituents Terpenoids and Polyphenols*** During the search for new chemopreventive agents, some plant triterpenoids have been found to show anticarcinogenic properties. Licorice root (*radix Glycyrrhizae glabra*) is one of the oldest and most frequently used botanicals in Chinese medicine. In

the United States and many Asian countries, licorice products are mostly used as flavoring and sweetening agents in the food and tobacco industries. Cytotoxic and radical scavenging activities of *G. radix* extract have been reported [78]. Essential constituents of licorice root include triterpenoids, such as glycyrrhizin and its aglycone glycyrrhizic acid, along with various polyphenols and polysaccharides [79]. Among these, glycyrrhizic acid (making up 4 to 20 percent of the root) consists of a glucuronic acid moiety attached to a steroid-like triterpenoid glycyrrhetinic acid, which could be responsible for the antioxidation and anti-inflammatory effects (being an inhibitor of both lipoxygenase and cyclooxygenase). The activation of glucuronidation suggests a possible role in the induction of xenobiotic detoxification in the liver. Besides its mild estrogenic effects, glycyrrhizin also encourages the production of hormones such as hydrocortisone and inhibits cyclooxygenase-2 (COX-2) activity, which may further explain its anti-inflammatory and anticarcinogenic properties.

*Glycyrrhizae radix* is the active ingredient in many Chinese/oriental herbal complex formulations. Experimental studies of these decoctions indicated significant potential to inhibit deoxyribonucleic acid (DNA)–synthesizing enzyme activity in colonic carcinomas [80] and to prevent formation of aberrant crypt foci in a colonic carcinogenesis model [81]. A phytopharmaceutical containing *Glycyrrhizae* extract (Revitonil tablets) had demonstrated strong immunostimulating potential [82]. Both licorice root and glycyrrhizin modulate immune activity by inducing interferon and interleukin production in mice [83, 84]. Glycyrrhizin has been reported to stimulate host resistance against solid tumors as well as to inhibit colon tumor cell activity [85]. However, glycyrrhizin is contraindicated in patients with hypertension, heart failure, renal diseases, and liver cirrhosis. Deglycyrrhinized licorice is currently under investigation for its safety and potential clinical use.

Other than glycyrrhizin, about 300 different polyphenols (making up 1 to 5 percent of the root) also possess antioxidative properties and have antitumor potential [86]. Many of them are flavonoids and isoflavonoids, which have weaker estrogenic activity than the triterpenoids. Glabridine, a major isoflavan in licorice root, possesses some estrogen-like activities and a strong antiproliferative action in human breast cancer cells [87]. Similar effects can also be produced by the chalcone isoliquiritigenin, despite its smaller quantity in the licorice root than other flavonoids and triterpenoids [88]. In fact, the tumor growth inhibition produced by isoliquiritigenin seems to be associated with its potential in inducing cell cycle arrest [89]. Besides growth inhibition, isoliquiritigenin also possesses a potent lipoxygenase (LOX) inhibitory effect, rather than COX inhibition [90]. Another possible antitumor action of isoliquiritigenin is its antitube formation effect, which contributes much to the antiangiogenic action of licorice root extract [91]. Alternatively, glabridine and the isoflavene glabrene had both demonstrated activity against *H. pylori*. This could contribute at least in part, to the gastroprotection of the whole extract of licorice root, which contains chemopreventive agents useful against peptic ulcer and gastric cancer [92]. Despite there not being much direct evidences

that nonglycyrrhizin polyphenols from *Glycyrrhizae* extract can exert antitumor action on GI cancers, the fact that isoliquiritigenin can significantly suppress carcinogen-induced aberrant crypt focus formation in mouse colon [93] is at least an indication of their chemopreventive potential against intestinal cancers.

***Other Herbs with Terpenoids*** *Tripterygium wilfordii Hook F* is a Chinese medicinal herb that belongs to the Celastraceae family. Historically, extracts of *T. wilfordii* were used in traditional Chinese medicine and later found to be effective in the treatment of inflammatory/autoimmune disorders such as rheumatoid arthritis [94]. The polysaccharide moiety from *Tripterygium* extract has exhibited profound immunosuppressive properties, inhibiting tumor necrosis factor-α (TNF-α) and cell adhesion molecule expression in human monocytes [95]. It is now known that some diterpenoid components in *T. wilfordii* such as triptolide exert their anti-inflammatory and immunosuppressant effects by inhibiting cytokine production by T lymphocytes [96, 97]. PG490-88 (14-succinyl triptolide sodium salt) is a semisynthetic compound derived from the diterpene triepoxide, or triptolide (PG490), that comes from the *T. wilfordii* extract. PG490-88 alone caused a dose-dependent inhibition of COLO 205 colorectal tumor growth as well as tumor regression. Moreover, PG490 blocked the induction of NF-κB by TNF-α and p53 transcriptional activity induced by chemotherapy, which implicated its synergistic action with other DNA-damaging chemotherapeutic drugs [98]. PG490-88 is now in phase I clinical trials for patients with solid tumors. A recent study had shown that PG490 inhibits metastasis of solid tumors from the stomach [99]. All these findings indeed illustrate the potential of *T. wilfordii* and its diterpenoid component triptolide in treating GI cancers.

The root barks of the Chinese tree *Pseudolarix kaempferi* (Tujingpi), also called Golden Larch Bark in the Western world, contains active medicinal compounds called pseudolaric acids. It is traditionally used in Chinese medicine to treat fungal infections. Pseudolaric acids A, B, C, and D are novel diterpenes extracted from *P. kaempferi.* Among these, pseudolaric acid B is the major antifungal constituent while it also demonstrates cytotoxic action against colorectal cancer cells [100]. In addition, other novel triterpene lactones isolated from the herb also possessed cytotoxicity against human colon cancer cells [101]. Recently, it was reported that pseudolaric acid B can act as a ligand to activate the nuclear peroxisome proliferator activated receptors (PPAR) α, γ, and δ [102]. Both gene and protein expressions of the nuclear hormone receptor PPARγ were upregulated in rodent colon tumors and in certain human colon cancer cells [103]. Since PPARγ is a ligand-modulated transcription factor, it may provide a novel target for chemopreventive strategies for colorectal cancer. The ability of PPARγ activation to concomitantly inhibit COX-2 expression and induce apoptosis in colon cancer cells illustrates its crucial role in anticarcinogenesis [104]. A similar consequence of PPARγ activation was also observed in human gastric cancer cells through induction

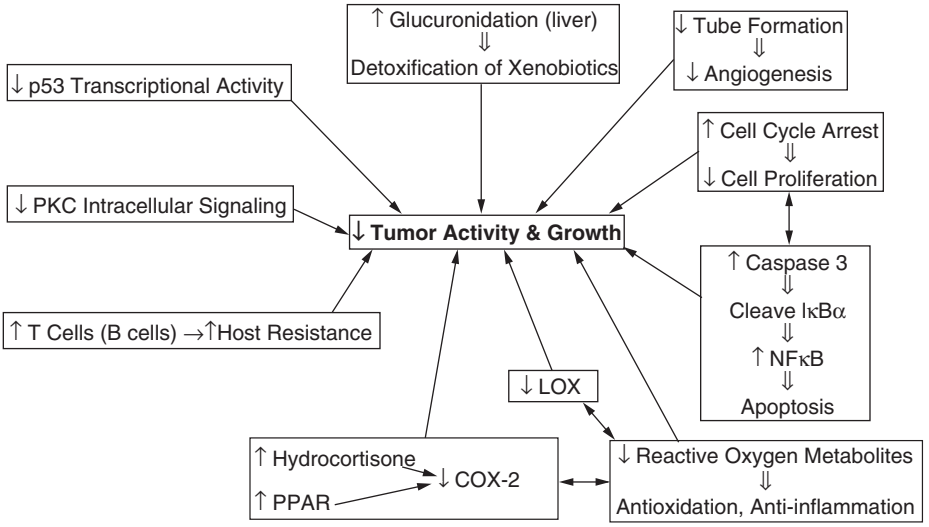


**Figure 1** Proposed mechanistic actions of terpenes and phenolic compounds.

of apoptosis together with G1 cell cycle arrest [105]. In this respect, it appears that *P. kaempferi* and its terpenoid components should be a target for active investigation for the development of effective chemotherapeutic strategies in the twenty-first century. The proposed mechanistic actions for the different terpenoids and phenolic compounds reported in this chapter are summarized in Figure 1.

## REFERENCES

1. Szabo, S., Trier, J. S., Brown, A., Schnoor, J., Homan, H. D., Bradford, J. C. (1985). A quantitative method for assessing the extent of experimental gastric erosions and ulcer. *J. Pharmacol. Methods*, *13*, 59–66.
2. Domschke, S., Domschke, W. (1984). Gastroduodenal damage due to drugs, alcohol and smoking. *Clin. Gastroenterol.*, *13*, 405–436.
3. Lichtenberger, L. M., Graziani, L. A., Dial, E. J., Butler, B. D., Hills, B. A. (1983). Role of surface active phospholipases in gastric cytoprotection. *Science*, *219*, 1327–1329.
4. Whittle, B. J. R., Steel, G. (1985). Evaluation of rat gastric mucosa by a prostaglandin analogue using cellular enzyme marker and histological techniques. *Gastroenterology*, *88*, 315–327.
5. Szabo, S., Cho, C. H. (1988). Animal models for studying the role of eicosanoids in peptic ulcer disease. In K. Hillier (Ed.), *Eicosanoids and the Gastrointestinal Tract*. MTP Press, Lancaster, UK, pp. 75–102.

6. Tsukimi, Y., Okabe, S. (1994). Validity of kissing gastric ulcers induced in rats for screening of antiulcer drugs. *J. Gastroenterol. Hepatol.*, *9*, S60–S65.
7. Robert, A., Stout, T. C., Dale, J. E. (1970). Production by secretogogues of duodenal ulcers in the rat. *Gastroenterology*, *59*, 95–102.
8. Cho, C. H., Pfeiffer, C. J. (1981). Gastrointestinal ulceration in the guinea pig in response to dimaprit, histamine and $H_1$- and $H_2$-blocking agents. *Digest. Dis. Sci.*, *26*, 306–311.
9. Szabo, S. (1978). Animal model of human disease. Duodenal ulcer disease. Animal model: Cysteamine-induced acute and chronic duodenal ulcer in the rat. *Am. J. Pathol.*, *93*, 273–276.
10. Elson, C. O., Sartor, R. B., Tennyson, G. S., Riddell, R. H. (1995). Experimental models of inflammatory bowel disease. *Gastroenterology*, *109*, 1344–1368.
11. Kim, H. S., Berstad, A. (1992). Experimental colitis in animal models. *Scand. J. Gastroenterol.*, *27*, 529–537.
12. Sartor, R. B. (1997). How relevant to human inflammatory bowel diseases are current animal models of intestinal inflammation? *Aliment. Pharmacol. Ther.*, *11*(Suppl. 3), 89–97.
13. Morris, G. P., Beck, P. L., Herridge, M. S., Depew, W. T., Szewczuk, M. R., Wallace, J. L. (1989). Hapten-induced model of chronic inflammation and ulceration in the rat colon. *Gastroenterology*, *96*, 795–804.
14. Chin, K. W., Barrett, K. E. (1994). Mast cells are not essential to inflammation in murine model of colitis. *Digest. Dis. Sci.*, *39*, 513–525.
15. Yamada, Y., Marshall, S., Specian, R. D., Grisham, M. B. (1992). A comparative analysis of two models of colitis in rats. *Gastroenterology*, *102*, 1524–1534.
16. Hawkins, J. V., Emmel, E. L., Feuer, J. J., Nedelman, M. A., Harvey, C. J., Kennedy, A. R., Lichtenstein, G. R., Billings, P. C. (1997). Protease activity in a hapten-induced model of ulcerative colitis in rats. *Digest. Dis. Sci.*, *42*, 1969–1980.
17. Okayasu, I., Hatakeyama, S., Yamada, M., Ohkusa, T., Inagaki, Y., Nakaya, R. (1990). A novel method in the induction of reliable experimental acute and chronic ulcerative colitis in mice. *Gastroenterology*, *98*, 694–702.
18. Copper, H. S., Murphy, S., Kido, K., Yoshitake, H., Flanigan, A. (2000). Dysplasia and cancer in the dextran sulfate sodium mouse colitis model. Relevance to colitis-associated neoplasia in the human: A study of histopathology, B-catenin and p53 expression and the role of inflammation. *Carcinogenesis*, *21*, 757–768.
19. Liu, E. S. L., Ye, Y. N., Shin, V. Y., Yuen, S. T., Leung, S. Y., Wong, B. C. Y., Cho, C. H. (2003). Cigarette smoke exposure increase ulcerative colitis-associated colonic adenoma formation in mice. *Carcinogenesis*, *24*, 1407–1413.
20. Guo, X., Wang, W. P., Cho, C. H. (1999). Involvement of neutrophils and free radicals in the potentiating effects of passive cigarette smoking on inflammatory bowel disease in rats. *Gastroenterology*, *117*, 884–892.
21. Wallace, J. L., Keenan, C. M. (1990). An orally active inhibitor of leukotriene synthesis accelerates healing in a rat model of colitis. *Am. J. Physiol.*, *258*, G527–G537.
22. Suzuki, K., Ota, H., Sasagawa, S., Sakatani, T., Fijikura, T. (1983). Assay method for myeloperoxidase in human polymorphonuclear leukocytes. *Anal. Biochem.*, *132*, 345–352.

23. Jung, Y. D., Mansfield, P. F., Akagi, M., Takeda, A., Liu, W., Bucana, C. D., Kicklin, D. J., Ellis, L. M. (2002). Effects of combination anti-vascular endothelial growth factor receptor and anti-epidermal growth factor receptor therapies on the growth of gastric cancer in a nude mouse model. *Eur. J. Cancer*, *38*, 1133–1140.
24. Shin, V. Y., Wu, W. K. K., Liu, E. S. L., Ye, Y. N., So, W. H. L., Koo, M. W. L., Liu, E. S. L., Luo, J. C., Cho, C. H. (2004). Nicotine promotes gastric tumor growth and neovascularization by activating extracellular regulated signal kinase and cyclooxygenase-2. *Carcinogenesis*, *25*, 2487–2495.
25. Takahashi, Y., Mai, M., Nishioka, K. (2000). α-Difluoromethylornithine induces apoptosis as well as anti-angiogenesis in the inhibition of tumor growth and metastasis in human gastric cancer model. *Int. J. Cancer*, *85*, 243–247.
26. Yamachika, T., Nakanishi, H., Inada, K. I., Tsukamoto, T., Shimizu, N., Kobayashi, K., Fukushima, S., Tatematsu, M. (1998). *N*-methyl-*N*-nitrosourea concentration-dependent, total intake-dependent, induction of adenocarcinomas in the glandular stomach of Balb/c mice. *Jpn. J. Cancer Res.*, *89*, 385–391.
27. Yano, H., Tatsuta, M., Iishi, H., Baba, M., Sakai, N., Uedo, N. (1999). Attenuation by *d*-limonene of sodium chloride-enhanced gastric carcinogenesis induced by *N*-methyl-*N*′-nitro-*N*-nitrosoguanidine in Wistar rats. *Int. J. Cancer*, *82*, 665–668.
28. Uedo, N., Tatsuta, M., Bara, M., Hirasawa, R., Iishi, H., Yano, H., Sakai, N., Uehara, H., Nakaizumi, A. (1999). Inhibition by rat C-erbB-2/neu antisense oligonucleotide of gastric carcinogenesis induced by *N*-methyl-*N*′-nitrosoguanidine in Wistar rats. *Int. J. Cancer*, *83*, 670–673.
29. Yamada, M., Ohkusa, T., Okayasu, I. (1992). Occurrence of dysplasia and adenocarcinoma after experimental chronic ulcerative colitis in hamsters induced by dextran sulphate sodium. *Gut*, *33*, 1521–1527.
30. Ye, Y. N., Liu, E. S. L., Shin, V. Y., Wu, W. K. K., Cho, C. H. (2004). Contributory role of 5-lipoxygenase and its association with angiogenesis in the promotion of inflammation-associated colonic tumorigenesis by cigarette smoking. *Toxicology*, *203*, 179–188.
31. Kado, S., Uchida, K., Funabashi, H., Iwata, S., Nagata, Y., Ando, M., Onoue, M. (2001). Intestinal microflora are necessary for development of spontaneous adenocarcinoma of the large intestine in T-cell receptor β chain and p53 double-knockout mice. *Cancer Res.*, *61*, 2395–2398.
32. Rao, C. V., Cooma, I., Rodriguez, J. G. R., Simi, B., El-Bayoumy, K., Reddy, B. S. (2000). Chemoprevention of familial adenomatous polyposis development in the $APC^{min}$ mouse model by 1,4 phenylene *bis* (methylene) selenocyanate. *Carcinogenesis*, *21*, 617–621.
33. Bralow, S. P., Weisburger, J. H. (1976). Experimental carcinogenesis in the digestive organs. *Clin. Gastroenterol.*, *5*, 527–542.
34. Rao, C. V., Hirose, Y., Indranie, C., Reddy, B. S. (2001). Modulation of experimental colon tumorigenesis by types and amounts of dietary fatty acids. *Cancer Res.*, *61*, 1927–1933.
35. Symolon, H., Schmelz, E. M., Dillehay, D. L., Merrill Jr., A. H. (2004). Dietary soy sphingolipids suppress tumorigenesis and gene expression in 1,2-dimethyldrazine-treated CG1 mice and $Apc^{Min/+}$ mice[1,2]. *J. Nutr.*, *134*, 1157–1161.
36. Nakaji, S., Ishiguro, S., Iwane, S., Ohta, M., Sugawara, K., Sakamoto, J., Fukuda, S. (2004). The prevention of colon carcinogenesis in rats by dietary cellulose is

greater than the promotive effect of dietary lard as assessed by repeated endoscopic observation. *J. Nutr.*, *134*, 935–939.

37. Narisawa, T., Magadia, N. E., Weisburger, J. H., Wynder, E. L. (1974). Promoting effect of bile acids on colon carcinogenesis after intrarectal instillation of *N*-methyl-*N′*-nitro-*N*-nitrosoguanidine in rats. *J. Natl. Cancer Inst.*, *53*, 1093–1097.
38. Benitz, W. E., Lessler, D. S., Coulson, J. D., Bernfield, M. (1986). Heparin inhibits proliferation of fetal vascular smooth muscle cells in the absence of platelet-derived growth factor. *J. Cell. Physiol.*, *127*, 1–7.
39. Klein-Soyer, C., Beretz, A., Cazenave, J. P., Wittendorp-Rechenmann, E., Vonesch, J. L., Rechenmann, R. V., Driot, F., Maffrand, J. P. (1989). Sulfated polysaccharides modulate effects of acidic and basic fibroblast growth factors on repair of injured confluent human vascular endothelium. *Arterioscler. Thromb. Vasc. Biol.*, *9*, 147–153.
40. Kaeffer, B., Benard, C., Lahaye, M., Blottiere, H. M., Cherbut, C. (1999). Biological properties of ulvan, a new source of green seaweed sulfated polysaccharides on cultured normal and cancerous colonic epithelial cells. *Planta Med.*, *65*, 527–531.
41. Vischer, P., Buddecke, E. (1991). Different action of heparin and fucoidan on arterial smooth muscle cell proliferation and thrombospondin and fibronectin metabolism. *Eur. J. Cell Biol.*, *56*, 407–414.
42. Sudhalter, J., Folkman, J., Svah, C. M., Bergendal, K., D'Amore, P. A. (1989). Importance of size, sulfation, and anticoagulant activity in the potentiation of acidic fibroblast growth factor by heparin. *J. Biol. Chem.*, *264*, 6892–6897.
43. Logeart, D., Prigent-Richard, S., Boisson-Vidal C., Chaubet, F., Durand, P., Jozefonvicz, J., Letourneur, D. (1997). Fucans, sulfated polysaccharides extracted from brown seaweeds, inhibit vascular smooth muscle cell proliferation. II. Degradation and molecular weight effect. *Eur. J. Cell Biol.*, *74*, 385–390.
44. Chihara, G. (1992). Recent progress in immunopharmacology and therapeutic effects of polysaccharides. *Devel. Biol. Stand.*, *77*, 191–197.
45. Miao, H. Q., Ishai-Michaeli, R., Peretz, T., Vlodavsky, I. (1995). Laminarin sulfate mimics the effects of heparin on smooth muscle cell proliferation and basic fibroblast growth factor-receptor binding and mitogenic activity. *J. Cell. Physiol.*, *164*, 482–490.
46. Czop, J. K., Kay, J. (1991). Isolation and characterization of β-glucan receptors on human mononuclear phagocytes. *J. Exper. Med.*, *173*, 1511–1520.
47. Yang, Q. Y., Jong, S. C. (1989). Medicinal mushroom in China. *Mushroom Sci.*, *12*, 631–643.
48. Wasser, S. P., Weis, A. L. (1999). Therapeutic effects of substances occurring in higher Basidiomycetes mushroom: A modern perspective. *Crit. Rev. Immunol.*, *19*, 65–96.
49. Chihara, G. (1978). Antitumor and immunological properties of polysaccharides from fungal origin. *Mushroom Sci.*, *9*, 797–814.
50. Kidd, P. M. (2000). The use of mushroom glucans and proteoglycans in cancer treatment. *Alt. Med. Rev.*, *5*, 4–27.
51. Shibata, H., Kimura-Takagi, I., Nagaoka, M., Hashimoto, S., Aiyyama, R., Ueyama, S., Yokokura, T. (2000). Properties of fucoidan from *Cladosiphon okamuranus tokida* in gastric mucosal protection. *Biofactors*, *11*, 235–245.

52. Shibata, H., Iimuro, M., Uchiya, N., Kawamori, T., Nagaoka, M., Ueyama, S., Hashimoto, S., Yokokura, T., Sugimura, T., Wakabay, K. (2003). Preventive effects of *Cladosiphon fucidan* against *Helicobacter pylori* infection in Mongolian gerbils. *Helicobacter*, *8*, 59–65.
53. Sun, X. B., Matsumoto, T., Yamada, H. (1991). Effects of a polysaccharide fraction from the roots of *Bupleurum falcatum L.* on experimental gastric ulcer models in rats and mice. *J. Pharm. Pharmacol.*, *43*, 699–704.
54. Yamada, H. (1995). Structure and pharmacological activity of pectic polysaccharides from the roots of *Bulpleurum falcatum L. Folia Pharmacol. Jpn.*, *106*, 229–237.
55. Sun, X. B., Matsumoto, T., Yamada, H. (1992). Anti-ulcer activity and mode of action of the polysaccharide fraction from the leaves of Panax ginseng. *Planta Med.*, *58*, 432–435.
56. Cho, C. H., Mei, Q. B., Shange, P., Lee, S. S., So, H. L., Guo, X., Li, Y. (2000). Study of the gastrointestinal protective effects of polysaccharides from *Angelica sinensis* in rats. *Planta Med.*, *66*, 348–351.
57. Ye, Y. N., Koo, W. W. L., Li, Y., Matsui, H., Cho, C. H. (2001). *Angelica sinensis* modulates migration and proliferation of gastric epithelial cells. *Life Sci.*, *68*, 961–968.
58. Ye, Y. N., Liu, E. S. L., Shin, V. Y. Cho, C. H. (2001). A mechanistic study of proliferation induced by *Angelica sinensis* in a normal gastric epithelial cell line. *Biochem. Pharmacol.*, *61*, 1439–1448.
59. Ye, Y. N., Li, Y., So, H. L., Cho, C. H. (2001). Protective effects of polysaccharide-enriched fraction from *Angelica sinensis* on hepatic injury. *Life Sci.*, *69*, 637–646.
60. Shang, P., Qian, A. R., Yang, T. H., Jia, M., Mei, Q. B., Cho, C. H., Zhao, W. M., Chen, Z. N. (2003). Experimental study of anti-tumor effects of polysaccharides from *Angelica sinensis. World J. Gastroenterol.*, *9*, 1963–1967.
61. Jayashree, P. K., Krutin, K. B., Thomas, P. A. D. (2000). Antioxidant properties of *Asparagus racemosus* against damage induced by γ-radiation in rat liver mitochondria. *J. Ethnopharmacol.*, *71*, 425–435.
62. Hase, K., Kadota, S., Basnet, P., Takahashi, T., Namba, T. (1996). Protective effect of celosian, an acidic polysaccharide, on chemically and immunologically induced liver injuries. *Biol. Pharm. Bull.*, *19*, 567–572.
63. Langmead, L., Feakins, R. M., Goldthorpe, S., Holt, H., Tsironi, E., De Silva, A., Jewell, D. P., Rampton, D. S. (2004). Randomized, double-blind, placebo-controlled trial of oral aloe vera gel for active ulcerative colitis. *Aliment. Pharmacol. Ther.*, *19*, 739–747.
64. Surh, Y-J. (1999). Molecular mechanisms of chemopreventive effects of selected dietary and medicinal phenolic substances. *Mutat. Res.*, *428*, 305–327.
65. Conney, A. H., Lou, Y. R., Xie, J. G., Osawa, T., Newmark, H. L., Liu, Y., Chang, R. L., Huang, M. T. (1997). Some perspectives on dietary inhibition of carcinogenesis: Studies with curcumin and tea. *Proc. Soc. Exper. Biol. Med.*, *216*, 234–245.
66. Anto, R. J., George, J., Babu, K. V. D., Rajasekharan, K. N., Kuttan, R. (1996). Antimutagenic and anticarcinogenic activity of natural and synthetic curcuminoids. *Mutat. Res.*, *370*, 127–131.

67. Syr, W-R., Shen, C-C., Don, M-J., Ou, J-C., Lee, G-H., Sun, C-M. (1998). Cytotoxicity of curcuminoids and some novel compounds from *Curcuma zedoaria*. *J. Nat. Prod.*, *61*, 1531–1534.
68. Churchill, M., Chadburn, A., Bilinski, R. T., Bertagnolli, M. M. (2000). Inhibition of intestinal tumors by curcumin is associated with changes in the intestinal immune cell profile. *J. Surg. Res.*, *89*, 169–175.
69. Kawamori, T., Lubet, R., Steele, V. E., Kellof, G. J., Kaskey, R. B., Rao, C. V., Reddy, B. S. (1999). Chemopreventive effect of curcumin, a naturally occurring anti-inflammatory agent, during the promotion/progression stages of colon cancer. *Cancer Res.*, *59*, 597–601.
70. Huang, M-T., Lou, Y-R., Ma, W., Newmark, H. L., Reuhl, K. R., Conney, A. H. (1994). Inhibitory effects of dietary curcumin on forestomach, duodenum, and colon carcinogenesis in mice. *Cancer Res.*, *54*, 5841–5847.
71. Moragoda, L., Jaszewski, R., Majumdar, A. P. N. (2001). Curcumin induced modulation of cell cycle and apoptosis in gastric and colon cancer cells. *Anticancer Res.*, *21*, 873–878.
72. Plummer, S. M., Holloway, K. A., Manson, M. M., Munks, R. J. L., Kaptein, A., Farrow, S., Howells, L. (1999). Inhibition of cyclo-oxygenase 2 expression in colon cells by the chemopreventive agent curcumin involves inhibition of NF-κB activation via the NIK/IKK signaling complex. *Oncogene*, *18*, 6013–6020.
73. Rao, C. V., Rivenson, A., Simi, B., Reddy, B. S. (1995). Chemoprevention of colon carcinogenesis by dietary curcumin, a naturally occurring plant phenolic compound. *Cancer Res.*, *55*, 259–266.
74. Baeuerle, P. A., Henkek, T. (1994). Function and activation of NF-kappa B in the immune system. *Annu. Rev. Immunol.*, *12*, 141–179.
75. Reuther, J. Y., Baldwin, A. S. Jr. (1999). Apoptosis promotes a caspase-induced amino-terminal truncation of I-kappa-B-alpha that functions as a stable inhibitor of NF-kappa B. *J. Biol. Chem.*, *274*, 20664–20670.
76. Liu, J. Y., Lin, S. J., Lin, J. K. (1993). Inhibitory effects of curcumin on protein kinase C activity induced by 12-*O*-tetradecanoyl-phorbol-13-acetate in NIH 3T3 cells. *Carcinogenesis*, *14*, 857–861.
77. Sharma, R. A., McLelland, H. R., Hill, K. A., Ireson, C. R., Euden, S. A., Manson, M. M., Pirmohamed, M., Marnett, L. J., Gescher, A. J., Steward, W. P. (2001). Pharmacodynamic and pharmacokinetic study of oral curcuma extract in patients with colorectal cancer. *Clin. Cancer Res.*, *7*, 1894–1900.
78. Nemoto, Y., Satoh, K., Toriizuka, K., Hirai, Y., Tobe, T., Sakagami, H., Nakashima, H., Ida, Y. (2002). Cytotoxic and radical scavenging activity of blended herbal extracts. *In Vivo*, *16*, 327–332.
79. Wang, Z. Y., Nixon, D. W. (2001). Licorice and cancer. *Nutr. Cancer*, *39*, 1–11.
80. Sakamoto, S., Mori, T., Sawaki, K., Kawachi, Y., Kuwa, K., Kudo, H., Suzuki, S., Sugiura, Y., Kasahara, N., Nagasawa, H. (1993). Effects of kampo (Japanese herbal) medicine "sho-saiko-to" on DNA-synthesizing enzyme activity in 1,2-dimethylhydrazine-induced colonic carcinomas in rats. *Planta Med.*, *59*, 152–154.
81. Yoo, B. H., Lee, B. H., Kim, J. S., Kim, N. J., Kim, S. H., Ryu, K. W. (2001). Effects of Shikunshito-Kamiho on fecal enzymes and formation of aberrant crypt foci induced by 1,2-dimethylhydrazine. *Biol. Pharm. Bull.*, *24*, 638–642.

82. Wagner, H., Jurcic, K. (2002). Immunological studies of Revitonil, a phytopharmaceutical containing *Echinacea purpurea* and *Glycyrrhiza glabra* root extract. *Phytomedicine*, *9*, 390–397.

83. Tomoda, M., Shimizu, N., Kanari, M., Gonda, R., Arai, S., Okuda, Y. (1990). Characterization of two polysaccharides having activity on the reticuloendothelial system from the root of *Glycyrrhiza uralensis*. *Chem. Pharm. Bull.*, *38*, 1667–1671.

84. Dai, J. H., Iwatani, Y., Ishida, T., Terunuma, H., Kasai, H., Iwakula, Y., Fujiwara, H., Ito, M. (2001). Glycyrrhizin enhances interleukin-12 production in peritoneal macrophages. *Immunology*, *103*, 235–243.

85. Chung, J. G., Chang, H. L., Lin, W. C., Wang, H. H., Yeh, C. C., Hung, C. F., Li, Y. C. (2000). Inhibition of *N*-acetyltransferase activity and DNA–2-aminofluorene adducts by glycyrrhizic acid in human colon tumour cells. *Food Chem. Toxicol.*, *38*, 163–172.

86. Rafi, M. M., Vastano, B. C., Zhu, N., Ho, C. T., Ghai, G., Rosen, R. T., Gallo, M. A., DiPaola, R. S. (2002). Novel polyphenol molecule isolated from licorice root (*Glycyrrhiza glabra*) induces apoptosis, G2/M cell cycle arrest, and Bcl-2 phosphorylation in tumor cell lines. *J. Agric. Food Chem.*, *50*, 677–684.

87. Tamir, S., Eizenberg, M., Somjen, D., Stern, N., Shelach, R., Kaye, A., Vaya, J. (2000). Estrogenic and antiproliferative properties of glabridin from licorice in human breast cancer cells. *Cancer Res.*, *60*, 5704–5709.

88. Maggiolini, M., Statti, G., Vivacqua, A., Gabriele, S., Rago, V., Loizzo, M., Menichini, F., Amdo, S. (2002). Estrogenic and antiproliferative activities of isoliquiritigenin in MCF7 breast cancer cells. *J. Steroid Biochem. Mol. Biol.*, *82*, 315–322.

89. Kanazawa, M., Satomi, Y., Mizutani, Y., Ukimura, O., Kawauchi, A., Sakai, T., Baba, M., Okuyama, T., Nishino, H., Miki, T. (2003). Isoliquiritigenin inhibits the growth of prostate cancer. *Eur. Urol.*, *43*, 580–586.

90. Yamamoto, S., Aizu, E., Jiang, H., Nakadate, T., Kiyoto, I., Wang, J. C., Kato, R. (1991). The potent anti-tumor-promoting agent isoliquiritigenin. *Carcinogenesis*, *12*, 317–321.

91. Kobayashi, S., Miyamoto, T., Kimura, I., Kimura, M. (1995). Inhibitory effect of isoliquiritin, a compound in licorice root, on angiogenesis in vivo and tube formation in vitro. *Biol. Pharm. Bull.*, *18*, 1382–1386.

92. Fukai, T., Marumo, A., Kaitou, K., Kanda, T., Terada, S., Nomura, T. (2002). Anti-*Helicobacter pylori* flavonoids from licorice extract. *Life Sci.*, *71*, 1449–1463.

93. Baba, M., Asano, R., Takigami, I., Takahashi, T., Ohmura, M., Okada, Y., Sugimoto, H., Arika, T., Nishino, H., Okuyama, T. (2002). Studies on cancer chemoprevention by traditional folk medicines XXV. Inhibitory effect of isoliquiritigenin on azoxymethane-induced murine colon aberrant crypt focus formation and carcinogenesis. *Biol. Pharm. Bull.*, *25*, 247–250.

94. Tao, X., Lipsky, P. E. (2000). The Chinese anti-inflammatory and immunosuppressive herbal remedy *Tripterygium wilfordii* Hook F. *Rheum. Dis. Clin. N. Am.*, *26*, 29–50.

95. Luk, J. M., Lai, W., Tam, P., Koo, M. W. L. (2000). Suppression of cytokine production and cell adhesion molecule expression in human monocytic cell line THP-1 by *Tripterygium wilfordii* polysaccharide moiety. *Life Sci.*, *67*, 155–163.

96. Chen, B. J. (2001). Triptolide, a novel immunosuppressive and anti-inflammatory agent purified from a Chinese herb *Tripterygium wilfordii* Hook F. *Leukem. Lymph.*, *42*, 253–265.
97. Qiu, D., Zhao, G., Aoki, Y., Shi, L., Uyei, A., Nazarian, S., Ng, J. C., Kao, P. N. (1999). Immunosuppressant PG490 (Triptolide) inhibits T-cell interleukin-2 expression at the level of purine-box/nuclear factor of activated T-cells and NF-κB transcriptional activation. *J. Biol. Chem.*, *274*, 13443–13450.
98. Fidler, J. M., Li, K., Chung, C., Wei, K., Ross, J. A., Gao, M., Rosen, G. D. (2003). PG490-88, a derivative of triptolide, causes tumor regression and sensitizes tumors to chemotherapy. *Mol. Cancer Ther.*, *2*, 855–862.
99. Jiang, X. H., Wong, B. C., Lin, M. C., Zhu, G. H., Kung, H. F., Jiang, S. H., Yang, D., Lam, S. K. (2001). Functional p53 is required for triptolide-induced apoptosis and AP-1 and nuclear factor-κB activation in gastric cancer cells. *Oncogene*, *20*, 8009–8018.
100. Pang, D. J., Li, Z. L., Hu, C. Q., Chen, K., Chang, J. J., Lee, K. H. (1990). The cytotoxic principles of *Pseudolarix kaempferi*: Pseudolaric acid-A and -B and related derivatives. *Planta Med.*, *56*, 383–385.
101. Chen, G. F., Li, Z. L., Pan, D. J., Tang, C. M., He, X., Xu, G. Y., Chen, K., Lee, K. H. (1993). The isolation and structural elucidation of four novel triterpene lactones, pseudolarolides A, B, C, and D, from *Pseudolarix kaempferi*. *J. Nat. Prod.*, *56*, 1114–1122.
102. Jaradat, M. S., Noonan, D. J., Wu, B., Avery, M. A., Feller, D. R., Jardat, M. S. (2002). Pseudolaric acid analogs as a new class of peroxisome proliferators-activated receptor agonists. *Planta Med.*, *68*, 667–671.
103. DuBois, R. N., Gupta, R., Brockman, J., Reddy, B. S., Krakow, S. L., Lazar, M. A. (1998). The nuclear eicosanoid receptor, PPARγ, is aberrantly expressed in colonic cancers. *Carcinogenesis*, *19*, 49–53.
104. Yang, W-L., Frucht, H. (2001). Activation of the PPAR pathway induces apoptosis and COX-2 inhibition in HT-29 human colon cancer cells. *Carcinogenesis*, *22*, 1379–1383.
105. Sato, H., Ishihara, S., Kawashima, K., Moriyama, N., Suetsugu, H., Kazumori, H., Okuyama, T., Rumi, M. A. K., Fukuda, R., Nagasue, N., Kinoshita, Y. (2000). Expression of peroxisome proliferator-activated receptor (PPAR)γ in gastric cancer and inhibitory effects of PPARγ agonists. *Br. J. Cancer*, *83*, 1394–1400.

# ENDOCRINE AND METABOLIC AGENTS

Brian L. Furman
*University of Strathclyde, Strathclyde Institute for Biomedical Sciences Glasgow, Scotland*

1 HYPOTHALAMIC HORMONES, THEIR ANALOGS, AND ANTAGONISTS
Thyrotrophin-Releasing Hormone
Corticotrophin-Releasing Hormone
Growth-Hormone-Releasing-Hormone
Gonadotrophin-Releasing Hormone
Somatostatin
2 ANTERIOR PITUITARY HORMONES
Growth Hormone (Somatotropin)
Follicle-Stimulating Hormone
Luteinizing Hormone
3 NEUROHYPOPHYSEAL HORMONES
Vasopressin
Oxytocin
4 THYROID HORMONES AND DRUGS MODULATING THYROID FUNCTION
Iodide
Thionamide Drugs
Potassium Perchlorate
5 DRUGS AFFECTING BONE METABOLISM
Parathyroid Hormone
Calcitonin
Bisphosphonates
Mithramycin
Calcimetics

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

6 ADRENAL CORTICOSTEROIDS
Glucocorticoids
Inhibitors of Glucocorticoid Synthesis
Aldosterone and Other Mineralocorticoids
Mineralocorticoid Antagonists
7 ANDROGENS
Androgen Receptors
Anabolic Steroids
Antiandrogens
5α-Reductase Inhibitors
8 ESTROGENS
Estrogen Receptors
Estrogen Agonists
Antagonists at Estrogen Receptors
Selective Estrogen Receptor Modulators (SERMs)
Aromatase Inhibitors
9 PROGESTINS
Agonists at Progesterone Receptors
Progesterone Receptors
10 TREATMENT OF DIABETES MELLITUS
Insulin
Biguanides
Thiazolidinediones
Sulfonylureas
Meglitinide Analogs
α-Glucosidase Inhibitors
Glucagon-like Peptide-1 (GLP-1) and Glucose-Dependent
Insulinotropic Polypeptide (GIP)
Exendin-4
Glucagon
11 ANTIOBESITY AGENTS
Orlistat
References

This chapter addresses the properties of endocrine hormones, their receptors, hormone analogs acting as agonists and antagonists, and the potential therapeutic applications. The general approach has been to classify the hormones according to their main, or classical, anatomical origin, although the disease classification is used in the case of hormones relevant to diabetes mellitus and obesity. The information is referenced using reviews or, where, appropriate, primary literature.

## 1 HYPOTHALAMIC HORMONES, THEIR ANALOGS AND ANTAGONISTS

Several peptide hormones are synthesized in the hypothalamus. These include oxytocin and vasopressin, which are synthesized in the paraventricular and supraoptic nuclei of the hypothalamus and stored in the posterior lobe of the pituitary gland, to which they are transported along axons in association with neurophysins. Oxytocin and vasopressin will be discussed under Neurohypophyseal Hormones (see later). The other hormones of the hypothalamus have a key role in the regulation of the secretory activity of the anterior pituitary. These are growth-hormone-releasing hormone (GHRH; somatorelin), thyrotropin-releasing hormone (TRH; protirelin), corticotrophin-releasing hormone (CRH), and gonadotropin-releasing hormone (GnRH; gonadorelin) and are secreted into the median eminence from where they are transported to the anterior lobe of the pituitary along long portal blood vessels. There is considerable evidence that TRH, GHRH, and CRH not only stimulate secretion of anterior pituitary hormones but also have trophic effects on their target pituitary cells [1]. The hypothalamic hormones are normally secreted in a pulsatile manner. This pulsatile nature of secretion is physiologically important, and alteration in the pulse frequency changes the secretory response of the anterior pituitary, this being best studied in relation to GnRH [2]. These peptides and/or their synthetic analogs are used in clinical practice either as diagnostic or therapeutic agents [3]. The actions of these hormones, however, extend beyond their effects on the secretion of the hormones of the anterior pituitary. Their potential therapeutic applications, and those of their analogs and antagonists, similarly may be unrelated to the originally identified physiological roles of the hormones.

### Thyrotrophin-Releasing Hormone

Thyrotrophin-releasing hormone (TRH; protirelin; thyroliberin) is the major hypothalamic regulator of the secretion of thyrotrophin [4] and therefore ultimately of the thyroid hormones thyroxine and triiodothyronine. However, a large amount of TRH is found in other regions of the brain, where it may modulate other neurotransmitters, with which it may be colocalized, or even function as a neurotransmitter itself [5]. Many actions other than stimulation of thyrotrophin secretion have been described for TRH and attributed to centrally mediated effects. These include increased arousal/wakefulness, improvements in memory and learning, locomotor activation, antidepressant activity, anticonvulsant activity, antinociceptive activity, reduction in food and water intake, increased gastric acid secretion, increased gastrointestinal contractility, and complex cardiovascular actions [5]. TRH is also found in the periphery, especially in the gastrointestinal and genitourinary tracts [5]. It is inactivated by a selective TRH-degrading ectoenzyme (pyroglutamyl aminopeptidase II) found on the surface of neurons, which appears to provide

the only mechanism for inactivating this peptide. The TRH breakdown product cyclo(His-Pro) (histidyl proline diketopiperazine) has both agonist and antagonist activity at the TRH receptor.

***Thyrotropin-Releasing Hormone Receptors*** TRH acts through two G-protein-coupled receptors, TRH-R1 and TRH-R2, which show different distributions in the brain and periphral tissues but which show identical affinities for TRH and its analogs [6]. These receptors are coupled through $G\alpha_q/G\alpha_{11}$ to the phospholipase C-activated hydrolysis of phosphatidylinositol.

The widespread actions suggest that there may be many interesting potential therapeutic applications and the pharmacology of TRH is in its infancy. Orally active analogs (e.g., taltirelin; CG-3703) have been produced and shown to increase wakefuleness in a canine model of narcolepsy and to be neuroprotective [7, 8].

## Corticotrophin-Releasing Hormone

Corticotropin-releasing hormone, also known as corticotropin-releasing factor (CRF) is a hypothalamic hormone that acts on the anterior pituitary to stimulate the secretion of corticotropin, thereby playing a key role in regulating the synthetic/secretory activity of the adrenal cortex [9]. CRF is also widely distributed in the central nervous system and in the periphery, and its functions probably extend beyond its role in stimulating corticotropin secretion. For example, its release activates the sympathetic nervous system, and thus the hormone is pivotal to the entire stress response. In addition to regulating the stress response, CRF and related peptides have a wide variety of actions in the central nervous system and in the periphery, possibly regulating anxiety, mood, feeding, inflammation, gastric emptying, and blood pressure [10]. There is considerable evidence for a link between hyperactive CRF pathways and depression and anxiety. At least three, additional, naturally occurring, related mammalian peptides are now known. These are urocortin, stresscopin/urocortin III (SCP/UCNIII), and stresscopin-related peptide/urocortin II (SRP/UCNII). Urocortin is also distributed throughout the central nervous system and in the periphery, but its central distribution is more restricted than that of CRF, being expressed strongly in the Edinger–Westphal nucleus and the lateral superior olive, where CRF is not expressed [11]. Urocortin is also found in the hippocampus, basal ganglia, medial septum, paraventricular nucleus, and the lateral hypothalamus [12].

***Corticotropin-Releasing Factor Receptors*** Two main CRF receptors have been identified, $CRF_1$ and $CRF_2$, for which the endogenous ligands show different selectivities. CRF has a much higher affinity (13 to 17-fold) for the $CRF_1$ receptor relative to the $CRF_2$ receptor [10]. There are at least two subtypes of the $CRF_2$ receptor, $CRF_{2A}$ and $CRF_{2B}$ [10]. Both CRF receptors are G-protein-coupled, seven-transmembrane receptors, signaling mainly through $G_s$ to

increase cyclic adenosine monophosphate (cAMP) but also through $G_q$, activating the phospholipase C pathway [10].

***Actions of CRF and Related Peptides*** Numerous studies have characterized the role of CRF in coordinating the behavioral and cardiovascular responses to stress [12, 13]. Central, but not systemic, administration of CRF produces behavioral activation (e.g., increased locomotor activity) in a range of species. It suppresses food intake and enhances behavioral responses to stress [12]. Urocortin also elevates plasma corticotrophin when administered into the rat brain, and stress results in increased urocortin messenger ribonucleic acid (mRNA) expression in the hypothalamus and midbrain, suggesting a role for urocortin in regulating the stress response [12].

The cardiovascular effects of CRF depend on its route of administration. Intracerebroventricular administration in all species produces increases in blood pressure, heart rate, and cardiac output mediated by activation of the sympathetic nervous system [13]. On the other hand, a species-dependent hypotensive effect is observed on peripheral injection [13]. CRF, in high doses in the human, appears to have vasodilator effects [13], and this may contribute to hypotension, causing syncope. The cardiovascular actions of peripherally administered urocortin are more marked than those of CRF. In rats and mice, urocortin produces a long-lasting hypotensive effect, mediated through vasodilatation, and an increase in heart rate [13], although in the sheep, urocortin increases blood pressure due to a pronounced increase in cardiac output, resulting from a direct effect in increasing cardiac contractility. The peripheral cardiovascular actions of urocortin appear to be mediated by the $CRF_2$ receptor [13]. Urocortin is also cardioprotective against ischemia, an effect mediated through the $CRF_2$ receptor [11]. Like CRF, urocortin has been shown to have a number of central and peripheral actions. Thus in various species it suppresses appetite, delays gastric emptying, increases locomotor activity, and is anxiogenic [11, 13].

Although an anti-inflammatory effect of CRF would be anticipated through release of adrenal glucocorticoids, CRF also has direct effects in augmenting proinflammatory cytokine (TNF$\alpha$, IL-1$\beta$) production from macrophages [14]. Urocortin also exerts direct proinflammatory effects and produced mast cell degranulation and an increase in vascular permeability in rat skin [15].

Effects on nonvascular smooth muscle may also be important. Incubation of human myometrial cells with CRF, but not urocortin II or urocortin III, for 8 to 16 h significantly induced mRNA and protein expression of constitutive but not inducible nitric oxide synthase (NOS) isoforms. This action resulted in increased activity of soluble guanylyl cyclase. CRF also caused acute activation of the membrane-bound guanylyl cyclase. These effects appeared to be mediated via the $CRF_1$ receptor [16].

***Therapeutic Applications*** The only current use of CRF is in the differential diagnosis of Cushing's syndrome/Cushing's disease and other adrenal disor-

ders [17]. Patients with pituitary Cushing's disease show an exaggerated increase in plasma corticotrophin and cortisol concentrations in response to CRF injection, whereas those with Cushing's syndrome of adrenal origin, or with the ectopic adrenocorticotropic hormone (ACTH) syndrome, show no response. In normal humans, the disappearance curve of ovine CRF from plasma was biexponential, with a plasma half-life of ~6 to 12 min for the fast component and ~55 to 75 min for the slow component [18, 19]. The long half-life appears to be associated with a prolonged biological effect.

***Corticotrophin-Releasing Factor Antagonists*** There is enormous interest in developing selective antagonists at the $CRF_1$ receptor because of their potential as anxiolytics and antidepressants.

*Antalarmin* is a selective antagonist at the $CRF_1$ receptor. It appears to be highly selective for the $CRF_1$ receptor and potently ($K_i$ 1.3 to 1.9 nM) competes for binding with $^{125}$I-labeled ovine CRF in tissues expressing predominantly $CRF_1$ receptors and does not displace CRF in the heart, which expresses predominantly $CRF_2$ receptors [20].

Chronic administration of antalarmin to rats produced effects consistent with antagonism of the hypothalamic-pituitary-adrenal (HPA) axis, that is, reductions in plasma concentrations of corticotropin and corticosterone and signs of chronic understimulation of the adrenal cortex (reduced width and vascularization of the zona fasciculata; increased number of apoptotic cells in the zona fasciculata) [22]. It also inhibited CRF-induced corticotropin release [20]. Antalarmin inhibited restraint-stress-induced gastric ulceration in rats [23] and ameliorated the behavioural changes produced in the chronic mild stress model of depression in mice [21].

Consistent with direct proinflammatory effects of endogenous CRF via the $CRF_1$ receptor, antalarmin significantly inhibited carageenin-induced subcutaneous inflammation in rats [20] and adjuvant arthritis in rats [24] and also prolonged survival in a mouse model of endotoxin shock [14]. It also antagonized the endogenous CRF-like peptide urocortin in producing an increased vascular permeability in rat skin [15].

*R121919* is a nonpeptide, water-soluble pyrrolopyrimidine that binds with high affinity to human $CRF_1$ receptors [10]. It is an antagonist at $CRF_1$ receptors and was shown to be antidepressant in a clinical study [25].

*CP-154,526* is a nonpeptide antagonist at $CRF_1$ receptors is closely related to antalarmin. It antagonizes CRH- and stress-induced neuroendocrine and behavioral effects and has also been shown to be antidepressant in a rat model [26].

*Antagonists at the $CRF_2$ Receptors* The peptide antisauvagine-30 is more than 300-fold selective for the $CRF_2$ receptor compared with the $CRF_1$ receptor. Currently, there are no nonpeptide antagonists at the $CRF_2$ receptor [27].

### Growth-Hormone-Releasing Hormone

Growth-hormone-releasing hormone (GHRH; somatorelin) is a potent stimulator of the secretion of somatotropin. GHRH acts via a G-protein-coupled receptor to stimulate the synthesis and secretion of GH via activating adenylyl cyclase and increasing cAMP [28]. It also potently activates the MAP kinase pathway in pituitary somatotrophs. A synthetic analog, sermorelin, which has an amino acid composition identical to the N-terminal 29 amino acids of natural GHRH, is used in the diagnosis of abnormalities of GH secretion. A number of peptide antagonists of GHRH (e.g., MZ-5-156; JV-1-36) have been synthesised and have been shown not only to suppress GH secretion but also to inhibit the growth of a number of tumors in vivo and in vitro [29].

### Gonadotrophin-Releasing Hormone

Gonadotrophin-releasing hormone (GnRH; gonadorelin) is secreted by the hypothalamus and regulates the secretion of the pituitary gonadotrophins follicle-stimulating hormone (FSH) and luteinizing hormone (LH) [30]. These hormones in turn regulate the secretory and gametogenic functions of the ovaries and testes. Administration of GnRH produces a rapid rise in the plasma concentrations of both FSH and LH. Several synthetic analogs (triptorelin, goserelin, histrelin, buserelin, deslorelin, leuprorelin, and nafarelin) have been developed. These are also agonists at the GnRH receptor, which is a G-protein-coupled ($G_{q/11}$) receptor but are more potent than the native hormone and are more resistant to proteolysis. For example, goserelin and buserelin are 20 times more potent than GnRH, while histrelin is 150 to 200 times more potent than the native hormone [3, 31]. Like GnRH they bind with high affinity to the GnRH receptor on anterior pituitary cells (e.g., [32]) and acutely stimulate the secretion of LH and FSH. However, prolonged, continuous exposure leads to inhibition of secretion by desensitization of the pituitary gonadotropes [33, 34]. There is evidence for direct effects on the ovaries [35] and on tumor cells [36, 37]. High concentrations ($10^{-4}$ M) of triptorelin had a direct inhibitory effect on the growth of an actively proliferating androgen responsive-prostatic cell line, an effect that was inhibited by the GnRH-receptor antagonist cetrorelix, while low concentrations ($10^{-7}$ M) promoted growth [37].

A second form of GnRH has been identified in the brain, peripheral nervous system and other tissues and is referred to as GnRH-II. There is a distinct receptor for this hormone, which is distributed in the brain similarly to GnRH-II, especially in areas associated with sexual arousal. It is also found in reproductive tissues. Its function is unknown. GnRH-II stimulates gonadotrophin secretion through the GnRH-I receptor [38].

***Therapeutic Applications*** Although GnRH agonists acutely stimulate the secretion of gonadotropins and have been used to induce ovulation, their main

therapeutic uses depend on the downregulation of the GnRH receptor function on the pituitary gonadotropes that occurs during continuous administration, with the resultant inhibition of gonadotropin secretion and, therefore, the inhibition of the follicular development in the female and inhibition of gonadal steroid hormone production in both sexes [39]. This is seen both in the human and in experimental animals. For example, a single injection of microencapsulated histrelin [10 to 300 μg/kg subcutaneous (SC)] induced a dose-dependent disruption of normal estrous cyclical activity in rats, resulting in persistent diestrous-like vaginal cytology. Desensitization was preceded by stimulation; LH secretion was maximal within 8 h and then gradually declined, remaining at diestrous levels from days 7 to 28 [40]. This desensitization forms the basis of their current and potential clinical applications in the treatment of endometriosis, uterine leiomyoma (fibroids), malignant neoplasms (especially prostatic carcinoma), central precocious puberty, and in assisted reproduction protocols. Deslorelin is used currently primarily in veterinary practice to induce ovulation in mares and cattle [41]. GnRH agonists have been proposed as potential contraceptive agents, although this use remains experimental. Because they are peptides, they cannot be given by mouth, but some (e.g., buserelin, nafarelin) are available as nasal sprays. Predictably, women may experience menopausal-like symptoms (hot flushes, increased sweating, vaginal dryness, dyspareunia, loss of libido) some weeks after starting treatment. Withdrawal bleeding may occur during the first few weeks of treatment and breakthrough bleeding may occur during prolonged treatment. Prolonged treatment for several months (as in the treatment of endometriosis) leads to significant loss of bone density [42]. Men being treated for prostatic carcinoma may experience dangerous, and sometimes painful, disease flare and bone pain at the initiation of treatment, as a consequence of the agonist action producing a transient increase in androgen secretion. This may be minimized by the initial use of an antiandrogen, such as flutamide or cyproterone [43].

***Gonadotropin-Releasing Hormone Antagonists*** While the continuous administration of GnRH agonists is highly effective in downregulating of the GnRH receptor function on the pituitary gonadotropes, these agents have the major, although predictable, disadvantage of initially stimulating the receptors, this being responsible for the initial disease flare seen in men being treated for prostatic carcinoma. Thus, peptide antagonists of GnRH have been developed. Early attempts to produce effective antagonists resulted in compounds that produced marked histamine release, as a result of mast cell degranulation [44, 45]. The antagonists that have been developed include cetrorelix and ganirelix, which are both currently available for clinical use, as well as antide, degarelix, abarelix, and antarelix. Unlike the GnRH agonists these compounds are devoid of any initial stimulatory actions. Because these agents are peptides, they still require to be administered parenterally. Potent, selective nonpeptide antagonists are under development,

some of which (CMPD1, NBI-42902, and TAK-013) are intended for oral administration [38].

## Somatostatin

Somatostatin is a naturally occurring peptide found throughout the central and peripheral nervous systems, the D cells of the pancreatic islets and gastrointestinal tract, and in many other structures. It inhibits the secretion of growth hormone (GH; somatotropin), insulin, glucagon, and gut hormones and is generally inhibitory to gastrointestinal motility and exocrine secretion [46]. Somatostatin also has a pronounced antiproliferative effect and promotes cellular apoptosis. It exists as two active forms—somatostatin-14 and somatostatin-28. Its secretion is regulated by many humoral, neural, and nutrient factors [46].

***Somatostatin Receptors*** The actions of somatostatin are mediated via high-affinity, membrane-bound, G-protein-coupled somatostatin receptors (SSTRs) of which five subtypes have been identified (SSTR1 to SSTR5). These subtypes have different tissue distribution and show somewhat different binding affinities (in the nanomolar range) for somatostatin-14 and somatostatin-28 but very marked differences in affinities for somatostatin analogs such as lanreotide and octreotide [46].

The intracellular signaling pathways mediating the effects of somatostatin is complex. All five receptors are negatively coupled to adenylyl cyclase, and the inhibitory actions of somatostatin on hormone secretion are mediated via inhibition of cyclic-AMP formation, activation of potassium channels, reduction in intracellular $Ca^{2+}$, and activation of the serine/threonine protein phosphatase calcineurin [46]. The antiproliferative effects of somatostatin may be mediated via activation of protein tyrosine phosphatases (SHP-2).

***Somatostatin Analogs*** Two synthetic somatostatin analogs are currently available for clinical use, lanreotide and octreotide, while a third, vapreotide, is undergoing clinical evaluation. Both bind with high affinity to somatostatin receptors (SSTRs), showing high selectivity for SSTR2 and SSTR5 relative to other SSTRs [46]. In clonal (CHO)-K1 cells stably expressing hSSTRs1–5 the binding affinities of lanreotide, expressed as $IC_{50}$ values, are 0.75 nM at SSTR2 and 5.2 nM at SSTR5 ($IC_{50}$), compared with 2330 nM at SSTR1, 107 nM at SSTR3, and 2100 nM at SSTR4. Its $EC_{50}$ in inhibiting growth hormone secretion from primary human fetal pituitary cultures was 2.3 nM [47]. The affinities of octreotide for these receptors in the same system were similar.

Several nonpeptide agonists have been synthesized. Some of these have markedly different binding affinities for different somatostatin receptors. Thus L-797, 591 shows very high affinity for the SSTR1 receptor ($IC_{50}$ 1.4 nM), with 120-fold selectivity for this receptor compared with, SSTR4, 1300-fold

selectivity compared with SSTR2, 1600-fold selectivity compared with SSTR3, and 2570-fold selectivity compared with SSTR5 [46]. Others in the same series showed high selectivity for SSTR2 (L-779, 976), SSTR3 (L-796, 778), and SSTR4 (L-803, 087). L-817, 818 shows very high affinity for SSTR5, with 8-fold selectivity compared with SSTR1 and about 130 to 200-fold selectivity compared with SSTR2, SSTR3, and SSTR4. These agents are now being used to obtain information about the somatostatin receptor subtypes that mediate the many effects of this hormone.

***Somatostatin Receptor Antagonists*** Peptides have been developed that are antagonists at somatostatin receptors. CYN-154806 is a potent octapeptide antagonist at the SST2 receptor, with at least 100-fold lower potencies at other receptor subtypes [48]. Others peptides have been developed with high selectivity for SSTR3 [49].

SRA880 was the first selective nonpeptide antagonist at the SSTR1 receptor to be developed. It is a competitive antagonist at the SSTR1 receptor with much lower affinities for the other receptor subtypes [50]. Similarly, nonpeptide (BN81644 and BN81674) antagonists have been developed with high affinity for the SSTR3 receptor [51].

The further development of potent and selective agonists and antagonists at the somatostatin receptor will enable a full exploration of the role of somatostatin and in developing novel therapies. Although it is clearly of enormous value to have highly selective agonists, there is also an argument for developing agonists with a broad spectrum of receptor activity, as they may be useful in neuroendocrine tumors expressing multiple somatostatin receptors.

***Therapeutic Applications*** Lanreotide and octreotide are used to treat acromegaly and neuroendocrine tumors, especially symptoms associated with carcinoid tumors with features of carcinoid syndrome, VIPomas, glucagonomas, and the Zollinger-Ellison syndrome. They have also been used for many other conditions including the control of bleeding of esophageal varices, diarrhea due to acquired immunodeficiency syndrome (AIDS) or chemotherapy, and breast cancer. In patients with acromegaly, lanreotide provides consistent suppression of GH levels to <5 mU/L in approximately 50 to 65 percent of all cases [52]. Most neuroendocrine gastrointestinal tumors express somatostatin receptors, and these somatostatin analogs produce clinical improvement mediated via a direct inhibitory effect on hormone production from the tumors in 30 to 70 percent of patients. Induction of tumor cell apoptosis has been described when high doses of lanreotide (12 mg/day) are administered. Eventually, however, all patients show escape from the therapy with regard both to hormonal production and tumor growth [53]. Radiolabeled lanreotide and octreotide may be used for the diagnosis and treatment of neuroendocrine tumors [52]. Somatotostatin-receptor binding peptides have been developed containing a chelator to enable radiolabeling and an apoptosis-inducing peptide [54].

## 2 ANTERIOR PITUITARY HORMONES

The anterior pituitary synthesizes and secretes a number of hormones, many of which are directly involved in the regulation of other endocrine glands. These include thyrotrophin (thyroid-stimulating hormone; TSH), corticotrophin (adrenocorticotrophic hormone; ACTH), and the two gonadotrophins, follicle-stimulating hormone (FSH) and luteinizing homone (LH). The other major hormones of the anterior pituitary are growth hormone (GH; somatotropin) and prolactin. Anterior pituitary hormones, or their human recombinant analogs, are used where there is a deficiency in the particular anterior pituitary hormone (e.g., somatotropin deficiency) or where an enhanced activity of the endogenous hormone is required to produce a therapeutic action.

### Growth Hormone (Somatotropin)

Growth hormone is secreted in a pulsatile manner from the anterior pituitary gland. Its secretion is regulated by a balance between the actions of the stimulatory GHRH and the inhibitory somatostatin. Moreover, a third hormone, ghrelin, which is abundantly expressed in the stomach and present in the hypothalamus is also a potent stimulator of GH secretion. About 45 percent of the hormone in the circulation is bound to a binding protein, which is similar to the extracellular domain of the growth hormone receptor and which markedly (10-fold) prolongs the half-life of the circulating hormone. Its effects are mediated by a specific GH receptor, one molecule of somatropin binding to two GH receptors, with the formation of a ligand-occupied receptor dimer [55]. Its actions are numerous and the GH receptor is expressed ubiquitously. Actions include stimulation of lipolysis, increased amino acid transport into muscle cells, chondrocyte clonal expansion, phosphorus retention by the kidney, increased calcium and phosphate absorption from the small intestine, increased synthesis of 1,25 dihydroxyvitamin $D_3$, and promotion of linear growth in bones. It has direct actions on adipocytes (increasing lipolysis) and on hepatocytes (increasing gluconeogenesis).

Many, but not all, of the actions of somatotropin are mediated by two factors, now known as insulin-like growth factor I (IGF-I, somatomedin C) and insulin-like growth factor II (IGF-II, somatomedin A) [56]. These factors are probably produced in all tissues, although the liver is probably the major source of circulating hormones. The IGFs are structurally related to proinsulin and act through similar receptor mechanisms, although having their own, distinct high-affinity receptors [56]. There are six high-affinity serum binding proteins that modulate the actions of the IGFs, either increasing or decreasing their effects [57]. The IGFs play a key role in preadolescent growth, produce rapid metabolic effects, and have long-term, autocrine growth-promoting actions [58]. IGF-I also regulates the secretion of somatotropin by a negative feedback effect.

***Therapeutic Applications*** Somatropin (human recombinant growth hormone) is used before epiphyseal fusion for the treatment of short stature due to decreased or absent secretion of endogenous hormone or due to gonadal dysgenesis. The response is monitored by determination of the serum concentration of IGF-I. It is also used to treat wasting associated with AIDS and severe illness. Because of its anabolic actions, it is prone to abuse by athletes. Overdose will induce symptoms of acromegaly.

*Pegvisomant—An Antagonist at the Growth Hormone Receptor* Pegvisomant is a novel recombinant, polypeptide GH analog with nine mutations, eight of which increase the affinity for the GH receptor. However, the mutation at gly(120) to Arg(120) prevents dimerization of the receptor and results in GH antagonism. The mutated protein is covalently bound to polyethylene glycol, reducing renal clearance and prolonging its biological half-life [59]. It suppresses IGF-I secretion in normal subjects [59] and, importantly, in patients with acromegaly [60]. Pegvisomant suppresses serum IGF-I, producing normalization and improvement in symptoms in around 97 percent of patients with acromegaly. In terms of IGF-I normalization, it is the most effective treatment to date [61].

## Follicle-Stimulating Hormone

One of two gonadotrophins (the other being luteinizing hormone, or LH) secreted by the anterior pituitary, FSH is a glycoprotein, the α chain of which is identical to that of LH, the receptor selectivity being conferred by the β chain. Follitrophin is recombinant follicle-stimulating hormone. Two follitrophins are available (follitrophin α and follitrophin β), which differ slightly in their carbohydrate structures. Its actions are those of FSH, which stimulates the development, maturation, and secretory activity of the follicles in women and spermatogenesis in men.

***Follicle-Stimulating Hormone Receptors*** Follitrophin produces its effects by activating the FSH receptor, a G-protein-coupled receptor with a large extracellular domain. Activation of the receptor leads to increased cAMP formation [62]. Although currently there are no readily available FSH receptor antagonists, a selective antagonist, (7-[4-[Bis-(2-carbamoyl-ethyl)-amino]-6-chloro-(1,3,5)-triazin-2-ylamino)-4-hydroxy-3-(4-methoxy-phenylazo)-naphthalene]-2-sulfonic acid), at the human and rat FSH receptor has been developed that prevents FSH-induced cAMP accumulation and steroidogenesis in vitro, as well as inhibiting ovulation in rats [63].

*Therapeutic Applications* The main use of FSH is in the treatment of female infertility, either where ovulation does not occur spontaneously, including women with polycystic ovarian disease, or to produce controlled ovarian hyperstimulation to induce the development of multiple follicles in medically

assisted reproduction (e.g., in vitro fertilization and embryo transfer) [64]. An ovarian hyperstimulation syndrome may occur in ~8.6 percent of patients receiving recombinant FSH [65]. It may also be effective in male infertility in patients with severe isolated hypogonadotropic hypogonadism, in whom spermatogenesis could be achieved after 18 months of treatment with follitrophin, preceded by 6 months of treatment with hCG (LH) [66].

### Luteinizing Hormone

Luteinizing hormone is the second of the two gonadotrophins secreted by the anterior pituitary. It is a glycoprotein, the α chain of which is identical to that of FSH, the receptor selectivity being conferred by the β chain. Human chorionic gonadotrophin (which mimics luteinizing hormone) has been widely used, but human recombinant LH is now available. Its main use is in combination with FSH in the treatment of female infertility, either where ovulation does not occur spontaneously, including women with polycystic ovarian disease, or to produce controlled ovarian hyperstimulation to induce the development of multiple follicles in medically assisted reproduction (e.g., in vitro fertilization and embryo transfer) [3].

The LH receptor is a G-protein-coupled receptor, stimulation of which activates adenylyl cyclase, increasing levels of cAMP, as well as activating phospholipase C [67]. In the latter respect the LH receptor differs from the FSH receptor. Chorionic gonadotrophin has high affinity for the LH receptor but is much less active at the FSH receptor. There do not appear to be any LH receptor antagonists.

## 3 NEUROHYPOHYSEAL HORMONES

The neurohypophysis releases oxytocin and vasopressin (arginine vasopressin in most mammals), which are cyclic nonapeptides synthesized in the paraventricular and supraoptic nuclei of the hypothalamus, from where they are transported along neurons for storage in, or release from, nerve terminals in the posterior lobe of the pituitary. They are transported in neurosecretory vesicles with their carrier proteins neurophysin I and neurophysin II. Both are powerful smooth-muscle contractants, but they have distinctive individual physiology and their actions are mediated via different receptors. Like other hypothalamic hormones, their biology extends far beyond their classical endocrine actions.

### Vasopressin

The classical endocrine functions of vasopressin are the facilitation of water reabsorption by the kidney and the contraction of vascular smooth muscle cells, vasopressin being one of the most potent circulating pressor peptides

[68]. Vasopressin is now also recognized, along with CRF, as a physiological regulator of corticotropin secretion. It has numerous nonendocrine actions and acts as a neurotransmitter in the central nervous system [69]. In the latter context it can modulate social behavior in a variety of species [70] and influences memory [71]. Vasopressin increases the release of von Willebrand factor and Factor VIII from the vascular endothelium.

Vasopressin has been implicated in congestive heart failure and in depressive and anxiety disorders. Thus it has become increasingly important to understand the receptors mediating the effects of vasopressin and therefore to develop selective agonists and antagonists at the various receptors.

***Receptors Mediating the Effects of Vasopressin*** Three main receptors have been identified, referred to as $V_{1a}$, $V_{1b}$, and $V_2$ receptors. The receptors show marked species variability, such that, for example, agonists that are highly selective for the human form of a particular subtype may be nonselective and much less potent at the rat form. Vasopressin receptors are G-protein-coupled receptors, but the different receptors are coupled through different G proteins. Thus the $V_2$ receptor is coupled through $G_s$ and activates adenylyl cyclase, increasing the production of cAMP. On the other hand, the $V_{1a}$ and $V_{1b}$ receptors are coupled through $G_{q/11}$, and its activation results in increased hydrolysis of phosphatidylinositol [72].

The $V_2$ vasopressin receptor is present in the collecting duct of the kidney, where it mediates vasopressin-induced antidiuresis, maintaining normal plasma osmolality, blood volume, and blood pressure. Activation of this receptor results in the incorporation of the water channel aquaporin-2 into the apical membrane of collecting duct cells [73]. Contraction of vascular smooth muscle cells is mediated via the $V_{1a}$ vasopressin receptor and stimulation of corticotropin secretion by the $V_{1b}$ receptors [69]. Stimulation of the release of factor-VIII-related antigen and von Willebrand factor from vascular epithelium is mediated by $V_2$ receptors.

***Agonists and Antagonists at Vasopressin Receptors*** d[D-3-Pal2]VP (Deamino[D-3-(pyridyl)Ala2,Arg8]VP) is a full agonist at the pituitary in stimulating the secretion of corticotrophin [74]. It is a $V_{1b}$ vasopressin receptor agonist, a weak $V_2$ vasopressin receptor agonist, and a weak $V_{1a}$ vasopressin receptor antagonist in the rat [75].

F-180, a vasopressin (VP) structural analog, shows high binding affinity ($K_i = -11$ nM) for the human and bovine $V_{(1a)}$ receptor subtype and is selective for this receptor subtype, at which it is potent in stimulating the accumulation of inositol phosphate [76]. It shows more than 400-fold selectivity for the $V_{1a}$ compared with $V_{1b}$ receptors and even higher selectivity relative to the $V_2$ receptor. At the oxytocin receptor it behaves as an antagonist. However, it has only weak affinity for the rat $V_{1a}$ receptor. F-180 is the first selective $V_{1a}$ agonist described for human and bovine vasopressin receptors.

*Chlorpropamide*, the antidiabetic sulfonylurea, is known to augment vasopressin-mediated antidiuresis, and it has been used to treat cranial diabetes insipidus but is ineffective in patients with nephrogenic diabetes insipidus. It was shown to be a weak inverse agonist for the $V_2$ vasopressin renal receptor and to upregulate vasopressin receptors both in vivo and in vitro incubation [77].

*SSR149415* was the first nonpeptide, orally active drug with nanomolar, selective affinity for mammalian $V_{1b}$ receptors. It exhibits potent antagonist properties both in vitro and in vivo, antagonizing effects of vasopressin in increasing intracellular [$Ca^{2+}$], promoting platelet aggregation, stimulating vascular smooth muscle cell proliferation, producing hypertension, and causing coronary vasospasm. It displays potent anxiolytic and antidepressant-like activities [78].

*SR121463* is a highly selective, orally active $V_2$ receptor antagonist and possesses powerful oral aquaretic properties in various animal species and in humans. SR121463 is well-tolerated and dose dependently increases urine output and decreases urine osmolality. Other orally active $V_2$ receptor antagonists are OPC-31260, OPC-41061 (tolvaptan), VPA-985 (lixivaptan), SR121463, VP-343, and FR-161282. YM-087 (conivaptan), JTV-605, and CL-385004 are mixed $V_{1a}/V_2$ receptor antagonists [79].

***Therapeutic Applications*** Vasopressin itself and its analog dDAVP are used for the treatment of central diabetes insipidus, which results from inadequate secretion of the antidiuretic hormone. Peptide analogs have variable bioavailability even via the intranasal route and orally active, nonpeptide agonist agents, such as OPC51803, are needed.

Another use of vasopressin analogs is in acute variceal hemorrhage, in which condition terlipressin (triglycyl lysine vasopressin), the synthetic analog of lysine vasopressin (the porcine vasopressin), is frequently used. DDAVP (desmopressin-1-deamino, 8-D-arginine vasopressin—dDAVP), a selective $V_2$ agonist, is used widely to treat von Willebrand disease, hemophilia A, and several platelet disorders. Its actions in these conditions may be mediated by stimulating the release of Factor-VIII-related antigen from vascular epithelium and Factor VIII coagulant activity from liver and other unidentified sites, in addition to stimulating the release of von Willebrand factor and tissue plasminogen activator [80, 81].

As indicated above, selective and orally active compounds are now available. These will provide new information about the pathophysiological role of vasopressin, as well as providing new therapies. $V_2$ receptor blockade could be of interest in several water-retaining diseases such as the syndrome of inappropriate antidiuretic hormone secretion (SIADH), liver cirrhosis, and congestive heart failure. Nonpeptide vasopressin $V_2$ receptor antagonists have shown beneficial effects in rat, canine, and porcine models of congestive heart failure and have reached phase II and phase IIIa stages of clinical development, in which they were shown to improve fluid status, osmotic balance, and

hemodynamic status [82]. The anxiolytic and antidepressant properties of $V_{1b}$ receptor antagonists suggest that this class of drugs has potential in the treatment of stress-related disorders, anxiety, and depression [78].

## Oxytocin

The classical effects of oxytocin are stimulation of uterine contractility, particularly in the later stages of pregnancy and the ejection of milk from the lactating mammary gland in response to suckling. However, oxytocin receptors are very widely distributed in the brain and in peripheral tissues and numerous other actions have been described including anxiolytic effects, induction of maternal behavior, satiety, stimulation of insulin and glucagon secretion, natriuresis, impairment of memory, control of reproductive behavior, analgesia, and central control of the cardiovascular system. It has been implicated in penile erection [83] and seminal ejaculation [84]. Not all of these actions are seen in all species and their physiological importance remains to be established.

***Receptors Mediating the Effects of Oxytocin*** Unlike vasopressin, there appears to be only one oxytocin receptor, which is a G-protein-coupled receptor coupled to $G_{q/11}$ and stimulating phospholipase C, resulting in the generation of inositol trisphosphate and diacylglycerol [85]. Oxytocin has also been shown to induce desensitization of the oxytocin receptor, which occurs in a time span that would be relevant to clinical use [85]. However, the overall response to oxytocin (OT) is influenced by the hormonal milieu, in particular the estrogen/progesterone ratio [86]. Thus estrogens stimulate OT receptor expression in uterus, pituitary, kidney, and the ventromedial nucleus of the hypothalamus. On the other hand, the very high concentrations of progesterone produced by the placenta during pregnancy may block oxytocin receptor signaling in the uterus through a nongenomic mechanism, and this may contribute to uterine quiescence [85].

***Agonists and Antagonists at the Oxytocin Receptor*** [$Thr^4$,$Gly^7$]OT is a potent oxytocic agent that is a useful pharmacological tool [87]. It is very selective for the rat, but not the human, oxytocin receptor, having no vasopressor or antidiuretic activity. It has high affinity for hippocampal oxytocin receptors and a very low affinity for central vasopressin and peripheral $V_{1a}$, $V_{1b}$, and $V_2$ vasopressin receptors.

Atosiban is a synthetic analog of oxytocin and acts as a competitive antagonist at the oxytocin receptor. Blockade of this receptor prevents oxytocin from stimulating inositol triphosphate production and inhibits influx of calcium into myometrial cells. The oxytocin-driven release of prostaglandins is also prevented [88].

***Therapeutic Applications of Oxytocin and Its Antagonists*** Oxytocin is principally used to induce or augment labor and in the treatment or prevention of

postpartum hemorrhage. The antagonist atosiban is used in the management of preterm labor (labor occurring before 37 weeks of gestation), where it can be used to delay imminent preterm labor from 24 to 33 weeks of gestation, thereby improving neonatal survival. It may be as effective as $\beta_2$-agonists, such as ritodrine, but is associated with fewer cardiovascular side effects [89, 90]. Myometrial sensitivity to oxytocin increases with increasing gestational age, as the receptor is progressively upregulated and atosiban is much more effective in later pregnancy [91].

## 4 THYROID HORMONES AND DRUGS MODULATING THYROID FUNCTION

Thyroxine ($T_4$) and tri-iodothyronine (liothyronine, $T_3$) are synthesized in the follicular epithelial cells from tyrosine residues found in the thyroid protein thyroglobulin. Iodide transport is a key step in the biosynthesis of $T_4$ and $T_3$, and it is actively transported into thyroid follicle cells by a mechanism that selectively concentrates iodide against a large concentration gradient [92]. The transporter co-transports $Na^+$ and iodide and is referred to as the sodium/iodide symporter [92]. Tyrosine residues are iodinated to form mono- and di-iodotyrosines, which then undergo oxidative coupling to form the active hormones, which are then stored as part of the thyroglobulin molecule within the thyroid follicle. Thryoglobulin, containing the attached thyroid hormones, is secreted at the apical surface of the thyrocytes into the follicle lumen, where it is stored. Thyroid hormone synthesis and secretion are controlled by the anterior pituitary, which secretes thyrotropin under the influence of the hypothalamic hormone thyrotropin-releasing hormone (protirelin; TRH).

$T_4$ is the main secretory product of the thyroid follicle cells and is converted to the active hormone, $T_3$ in the tissues under the influence of microsomal deiodinase enzymes which remove the iodine from the "outer" ring (5′-deiodination). Removal of the iodine from the "inner," tyrosyl ring (5-deiodination) produces an inactive substance known as reverse tri-iodothyronine ($rT_3$) [94]. $T_4$ is highly (>99 percent) bound to plasma protein (mostly to thyroid-binding globulin and transthyretin), with only a small amount of free hormone ($FT_4$) being available for biological activity.

The liver is a major site of metabolism, although deiodination to $T_3$ probably occurs in most tissues [93]. In the presence of normal thyroid function, the plasma half-life of thyroxine is about 6 to 7 days, while that of $T_3$ is about 1 to 2 days.

***Receptors Mediating the Effects of Thyroxine and Tri-iodothyronine*** The classical effects of $T_4$ in maintaining the metabolic activity of tissues by inducing the production of key enzymes are generally accepted to be mediated via specific nuclear thyroid hormone receptors, which bind the active $T_3$ and of which there are two main types, TR$\alpha$ and TR$\beta$ [94].

Around 30 genes have been described that are regulated by thyroid hormones, including sarcoplasmic reticulum calcium adenosinetriphosphatase (ATPase), α-myosin heavy chain, malic enzyme [95], and uncoupling proteins [96]. However, as seen with steroids (see later), there is evidence for rapid, nongenomic actions of thyroid hormones mediated via nonnuclear, cell surface, G-protein-coupled receptors. These actions can be produced by $T_4$ without conversion to $T_3$. Some of these actions seem to be mediated via the mitogen-activated protein kinase (MAP kinase) cascade and are concerned with modulating the actions of growth factors and certain cytokines [97]. A potent, nongenomic action in activating cardiac myocyte sodium channels was demonstrated [98], and this action was shared by $T_4$ and $T_3$, as well as some other iodinated tyrosines, but not by $rT_3$. There is also evidence for direct, nongenomic effects of thyroid hormones on mitochondrial function, via mitochondrial receptors [99].

***Therapeutic Applications*** Thyroxine is used in all diseases associated with clinical hypothyroidism, where it will reverse the signs and symptoms of hypothyroidism. The dose is titrated for individual patients depending on clinical response and measurement of plasma-free thyroid hormone concentration. Although active orally, absorption is variable and it is given by injection in myxedema coma.

The adverse effects of thyroxine and tri-iodothyronine are identical to the symptoms of thyrotoxicosis (anginal pain, cardiac arrhythmias, palpitations, tachycardia, tachypnoea, diarrhea, vomiting, tremors, restlessness, excitability, insomnia, flushing, sweating, excessive weight loss, and headache).

***Thyroid Hormone Analogs*** $T_3$ lowers serum cholesterol but is not used to treat hypercholesterolemia because of its effect in producing tachycardia and cardiac dysrrhythmias. Similarly it exerts a positive inotropic effect, but its ability to produce tachycardia in the same doses precludes its use in heart failure. Analogs of thyroid hormones have been synthesized in an attempt to exploit either the positive inotropic or cholesterol-lowering effects but without adverse effects on the heart. For example, CGS 23425 reduced low-density lipoprotein (LDL) cholesterol by 44 percent in hypercholesterolemic rats but had no effect on atrial contractility and force at doses up to 1000-fold greater [100]. On the other hand, the thyroid hormone analog DITPA (3,5-diiodothyropropionic acid) improved systolic as well as diastolic function through effects on cardiac muscle and is a promising potential new treatment for heart failure [101]. Recent work has provided a molecular basis for the differences between the pharmacological profiles of these two analogs. While DITPA had affinity for both thyroid receptors, CGS 23425 was shown to bind selectively to the β form. This difference was reflected in differences in gene regulation by the two compounds, with DITPA and $T_3$, but not CGS 23425, upregulating genes encoding cardiac contractile proteins, sarcoplasmic reticulum calcium

ATPase, and several proteins of mitochondrial oxidative phosphorylation [102].

## Iodide

Iodide, which is a fundamental constituent of $T_4$ and $T_3$, has complex effects on the thyroid, where it can inhibit the synthesis and secretion of $T_4$ and $T_3$, as well as reducing the cellularity and vascularity of the gland in Graves' disease. It may also produce hyperthyroidism in certain circumstances (Jod–Basedow effect) [103].

In normal individuals excess iodide inhibits thyroid hormone synthesis (acute Wolff–Chaikoff effect), this being mediated probably by the formation of iodopeptide(s) that temporarily inhibit thyroid peroxidase (TPO) mRNA and protein synthesis and, therefore, thyroglobulin iodinations. After a few days, the organification of intrathyroidal iodide resumes and the normal synthesis of $T_4$ and $T_3$ returns ("escape" from the Wolff–Chaikoff effect). This occurs through downregulation of the sodium iodine symporter (NIS), thus decreasing the intrathyroidal inorganic iodine concentration and permitting the TPO–$H_2O_2$ system to resume normal activity. However, in the newborn and the fetus, and in euthyroid patients with autoimmune thyroiditis, this escape phenomenon is not achieved and hypothyroidism results [104, 105]. The ability of iodide to reduce thyroid cellularity and vascularity in patients with Graves' disease is not clearly understood, but iodide was shown to cause apoptosis in rat and porcine thyroid [106, 107].

Iodide is used in the treatment of Graves' disease to prepare for thyroid surgery. It induces euthyroidism and reduces the size and vascularity of the gland. It is administered for 8 to 10 days before surgery, after the patient is rendered euthyroid using an antithyroid drug such as methimazole [108]. Iodide also protects the thyroid against radiation-induced damage due to accidental exposure to radioactive iodine. In adults with normal thyroid function, ingestion of 100 mg of iodide just before exposure to radioactive iodine blocks at least 95 percent of the thyroid uptake of radioactivity [109].

## Thionamide Drugs

The thionamides carbimazole, methimazole, methylthiouracil, and propylthiouracil inhibit the thyroid peroxidases (TPO) involved in catalyzing the iodination of tyrosine residues in thyroglobulin and the oxidative coupling of iodinated tyrosines. This inhibition is competitively antagonized by iodide at low, but not higher, drug concentrations [110]. It has also been suggested that carbimazole and methimazole reduces the autoimmunity underlying Graves' disease, since successful treatment is associated with suppression of thyroid receptor autoantibodies [111]. Carbimazole is rapidly converted to the active compound methimazole [112]. Propylthiouracil additionally inhibits the D1 isoform of the deiodinase enzyme (iodothyronine 5′-deiodinase) that catalyzes

the conversion of thyroxine ($T_4$) to the active tri-iodothyronine ($T_3$) in most peripheral tissues. This may contribute to the more rapid onset of its therapeutic effect compared with that of methimazole. It also inhibits the 5-deiodination of $T_3$, which is an inactivating step; theoretically this could counter the reduced formation of $T_3$ [113], but there does not appear to be any clinical evidence for this. Apart from the effects on deiodinase, the actions and adverse effects of propythiouracil are broadly similar to those of carbimazole and methimazole. Methimazole or propylthiouracil are useful for the production of experimental hypothyroidism in the rat. For example, 0.03 percent methimazole in the drinking water for 4 weeks reduced serum free T4 concentrations by 92 percent [114].

***Therapeutic Applications*** Carbimazole is used in Europe and the United Kingdom, whereas methimazole is used in the United States for the treatment of Graves' disease and other forms of hyperthyroidism. Propylthiouracil is also used. They may be used alone or, in the treatment of Graves' disease, a "block and replace" regime may be used, in which large doses of antithyroid drugs are given continuously, along with thyroxine to prevent iatrogenic hypothyroidism. The rationale for this is partly the belief that large doses of antithyroid drugs suppress the underlying autoimmunity. However, there is no evidence that such regimens are more effective than others [115].

These drugs are generally well tolerated, the major adverse effects appearing to have an immune origin. The most dangerous adverse effect is agranulocytosis, which occurs in 0.1 to 0.5 percent of patients [116]. Agranulocytosis has been suggested to occur almost exclusively during the first 10 weeks of treatment and is probably related to the drug dose [117]. Recombinant human granulocyte colony-stimulating factor (rhG-CSF) may shorten the recovery time in patients with methimazole-induced agranulocytosis. Hepatotoxicity is another rare, but acknowledged, adverse effect [118], with propylthiouracil being more likely than other antithyroid drugs to produce liver injury, which may be asymptomatic [119].

## Potassium Perchlorate

Potassium perchlorate competitively inhibits the active iodide transport mechanism, the sodium/iodide symporter, in the thyroid, which has the capacity to selectively concentrate iodide against a large concentration gradient [92]. Iodide transport is a key step in the biosynthesis of the thyroid hormones thyroxine ($T_4$) and tri-iodothyronine (liothyronine, $T_3$). Although not now used for the treatment of Graves' disease because it may cause aplastic anemia, it is used to treat thyrotoxicosis associated with the antiarrhythmic agent amiodorone, which is an iodine-rich compound with complex effects on thyroid function. Amiodarone-induced thyrotoxicosis occurs in 2 to 12 percent of patients receiving this drug and is due to iodine-induced excessive thyroid

hormone synthesis in patients with nodular goiter or latent Graves' disease or to a thyroid-destructive process caused by amiodarone or iodine [92].

## 5 DRUGS AFFECTING BONE METABOLISM

The prevalence of osteoporosis is high affecting, for example, around 17 percent of postmenopausal women in the United States and up to 30 percent of women older than 65. Given the increasing number of aged people, this presents a major problem because of the attendant morbidity and mortality due to fractures of the hip, vertebrae, and distal forearm. Hip fracture is associated with 20 percent mortality and 50 percent permanent loss in function. Thus there is great interest in developing safe and effective therapies to prevent osteoporosis. Agents affecting bone metabolism comprise a heterogeneous group of drugs. Their main applications are in the treatment and prevention of osteoporosis associated with the menopause or with the use of glucocorticoids [120, 121], Paget's disease [122], and in the treatment of hypercalcemia of malignancy [123].

### Parathyroid Hormone

Parathyroid hormone (PTH), the main secretory product of the parathyroid gland, plays a key role in calcium homeostasis and in the regulation of bone turnover. It is secreted in response to a reduction in extracellular calcium and produces increased resorption of calcium from bone, increased renal reabsorption of calcium and increased renal excretion of phosphate. It also stimulates the formation of calcitriol (1,25-dihydroxycholecalciferol; 1,25 dihydroxy-$D_3$) in the kidney. While the effects of continuous, high levels of parathyroid hormone lead to a net bone loss, there is strong evidence for an anabolic effect of parathyroid hormone on bone. Its secretion in response to a reduction of calcium is controlled by a calcium sensing receptor on the cell membrane of the parathyroid chief cells. This is a G-protein-coupled receptor, activation of which leads to increases in phosphatidylinositol turnover and sustained increases in intracellular calcium and an inhibition of adenylyl cyclase and of PTH secretion [124].

The effects of PTH on bone are exerted through alterations in the formation and activity of the osteclasts and osteoblasts. The regulation of the formation of osteoclasts, the bone resorbing cells, is achieved through a balance between the production and action on osteoclast-precursor cells of two substances osteoprotegerin and RANKL secreted by osteoblasts. RANKL, a member of the TNF (Tumor Nevosis Factor) superfamily, normally binds to its receptor RANK on the osteoclast-precursor cells to promote their development into osteoclasts.

Osteoprotegerin (OPG) is the natural inhibitor of RANKL, binding to RANKL and preventing its interaction with RANK, thereby inhibiting osteo-

clastogenesis [125]. The osteoclastogenic activity of PTH occurs primarily by suppression of OPG gene expression in early osteoblasts and elevation of RANKL gene expression in mature osteoblasts [126].

Parathyroid hormone exerts its effects through binding to a class 2 G-protein-coupled receptor that coupled through $G_s$ activates adenylyl cyclase, increasing cAMP and activating PKA (protein kinase A) [127]. It is also coupled through $G_q$ to phospholipase C, resulting in increased phosphatidyl inositol breakdown, increasing the formation of inositol trisphosphate and diacylglycerol [128]. This receptor also binds PTH-related protein (PTHrP). There are now known to be at least two PTH receptors, PTH-1R and PTH-2R. PTH-2R has only 51 percent homology with PTH-1R but recognizes PTH. However, it does not bind to PTHrP. The natural ligand for this receptor is thought to be TIP39 (tubero-infundibular peptide of 39 aa).

***Anabolic Actions of PTH on Bone*** The anabolic actions of PTH when administered intermittently can be attributed to stimulation of osteoblast differentiation and reduction in osteoblast apotosis [129]. During continuous administration, the effects on osteoclast formation predominates, leading to a net loss of bone. The anabolic effect of PTH is shared by PTHrP and by N-terminal PTH fragments, including recombinant human PTH(1–34), now known as teriparatide [130].

## Calcitonin

The thyroid also secretes calcitonin (CT) from the parafollicular or C cells, which are found among the thyroid follicles in mammals. Calcitonin lowers serum calcium concentrations and the secretion of calcitonin is regulated by the extracellular [$Ca^{2+}$]. Calcitonin is extremely potent (femtomolar concentrations) in inhibiting the resorptive function of mature osteoclasts [131] and is used in Paget's disease of bone, postmenopausal osteoporosis, hypercalcemia, and bone pain due to malignant neoplasms.

***Calcitonin Receptors*** The CT receptor is a G-protein-coupled receptor, which signals through multiple heterotrimeric G proteins, $G_s$, $G_i$, and $G_q$, leading to the activation of adenylyl cyclase and phospholipase C. Adenylyl cyclase and phospholipase C appear to mediate different components of the responses to CT. For example, cAMP-dependent mechanisms lead to reduced osteoclast motility, while protein-kinase-C-dependent events may mediate CT-induced retraction of the cell. CT also activates the Erk1/2 pathway, which may play an important role in mediating the effects of CT on osteoclasts [131].

## Bisphosphonates

This class of agents includes alendronate, etidronate, pamidronate, ibandronate, risedronate, clodronate, tiludronate, and zoledronate, which are var-

iously used as the free acid or as the sodium salts as inhibitors of bone resorption. They are analogs of pyrophosphate, which has been long-known to inhibit calcium phosphate dissolution in vitro [132]. There are two classes of bisphosphonates; the aminobisphosphonates (e.g., pamindronate, ibendronate, risedronate, zoledronate) potently and selectively inhibit the mevalonate pathway by inhibition of farnesyl diphosphate synthase, inhibiting the conversion of geranylpyrophosphate to farnesylpyrophosphate, with a resultant depletion of geranylgeranylpyrophosphate and the inhibition of protein prenylation, including the prenylation of small GTPases. This results in osteoclast apoptosis and inhibition of bone resorption [132–134]. On the other hand, the non-aminobisphosphonates (clodronate, etidronate, tiludronate) do not inhibit farnesyl diphosphate synthase but become incorporated into the phosphate chain of ATP-containing molecules, rendering them nonhydrolysable and cytotoxic to osteoclasts [132, 133].

***Therapeutic Applications*** Bisphosphonates are used widely to prevent/treat loss of bone due to glucocorticoid treatment, postmenopausal estrogen loss or to malignant disease, and in treating hypercalcemia of malignancy and Paget's disease of the bone. High doses of aminobisphosphonates are very effective in hypercalcemia of malignancy and in reducing the risk of skeletal morbidity and fractures in patients with breast cancer or multiple myeloma, with a low incidence of adverse effects [135]. There is interest in using bisphosphonates to target other drugs (e.g., antineoplastic agents, anti-inflammatory drugs) to bone because of their high affinity for this tissue [136]. Bisphosphonates are poorly absorbed from the gastrointestinal tract, with an oral bioavailability of around 1 to 2 percent, but accumulate in bone, from which their clearance is very slow.

## Mithramycin

Mithramycin is an antineoplastic antibiotic that complexes with deoxyribonucleic acid (DNA) in the presence of divalent cations such as magnesium. It has also been used in Paget's disease. Mithramycin has a direct inhibitory effect on bone resorption and is an mRNA synthesis inhibitor [137]. It acts directly on bone, as evidenced by its concentration-dependent (0.1 to 100 nM) inhibition of osteoclastic bone resorption in the in vitro bone slice assay, with a maximum inhibition of ~66 percent [138]. Mithramycin is used to control the hypercalcemia and hypercalciuria of malignancy and to reduce bone turnover in Paget's disease. Because of its marked toxicity, it is generally used only when other measures have failed.

## Calcimimetics

Calcimimetics comprise a new class of drug acting at the parathyroid cell calcium sensing receptor (CaSR) to increase its sensitivity to extracellular

calcium, which will lead to a decrease in circulating PTH levels. This will result in a reduction in serum calcium concentration, which makes these drugs promising as treatments for hyperparathyroidism and parathyroid carcinoma. There are two classes of calcimimetics. Class I calcimimetics are agonists that do not require the presence of calcium and include several organic and inorganic polyvalent cations such as gadolinium, neomycin, and spermine. These agents lack selectivity for the CaSR and are not used therapeutically to lower serum calcium. On the other hand, class II calcimimetics are low-molecular-weight, orally active phenylalkylamine compounds, which are ineffective in the absence of calcium. R568 (*N*-(3-[2-chlorophenyl]propyl)-(R)-α-methyl-3-methoxybenzylamine) is the most extensively studied type II calcimimetic compound [124]. Others being investigated include cinacalcet (AMG073) and NPS 568.

## 6 ADRENAL CORTICOSTEROIDS

### Glucocorticoids

The main naturally occurring glucocorticoid in the human is cortisol (hydrocortisone), whereas in rodents corticosterone fulfills that role. Glucocorticoids are secreted primarily by the zona fasciculata of the adrenal cortex. They have numerous actions including stimulating hepatic glycogen synthesis, promoting gluconeogenesis, impairing peripheral glucose utilization, and exerting immunomodulatory and anti-inflammatory effects. Many of their actions are exerted at the genome level, effects being observed within 30 min to 18 h. Additionally, glucocorticoids produce rapid effects observed within seconds or minutes; these actions are described as nongenomic, as their onset time is far too rapid to require altered protein synthesis.

***Glucocorticoid Receptors*** Glucocorticoids are lipophilic and produce their classical effects by stimulating glucocorticoid receptors, which are members of the nuclear receptor superfamily that includes receptors for sex hormones such as androgens, estrogens and progestins, mineralocorticoids, thyroid hormones, and vitamin D. There are two types of corticosteroid receptors. Type I is the mineralocorticoid receptor, which also binds glucocorticoids and is expressed primarily in the kidney, colon, and salivary and sweat glands, as well as in the hippocampus [139]. The type II receptor is the glucocorticoid receptor (GR), which is expressed ubiquitously as two major splice variants GRα and GRβ. GRα mediates the classical effects of glucocorticoids and the function of GRβ is unclear, although there is evidence that it can act as a dominant negative inhibitor of GRα. Changes in the ratio of GRα:GRβ forms of the receptor could explain relative glucocorticoid resistance in inflammatory diseases [139]. Glucocorticoids enter cells and interact with cytoplasmic GRα, which are bound to heat shock proteins (hsp90). Release of these chaperone

molecules and phosphorylation of GRα result in homodimerization of the receptor–ligand complex, which migrates to the cell nucleus. Domains within the N-terminal region of the receptor interact with either negative or positive glucocorticoid response elements (GREs) in promoter regions of particular genes with consequential initiation or repression of transcription. This mechanism results in the induction of a number of proteins that may contribute to the anti-inflammatory effect of glucocorticoids, including MAP kinase phosphatase-1 (MKP-1), glucocorticoid-induced leucine zipper (GILZ), and annexin I (Anx-A1, formerly known as lipocortin-1) (Fig. 1).

Another important mechanism of glucocorticoid action involves indirect inhibition of transcription by interaction of the ligand-activated GRα with

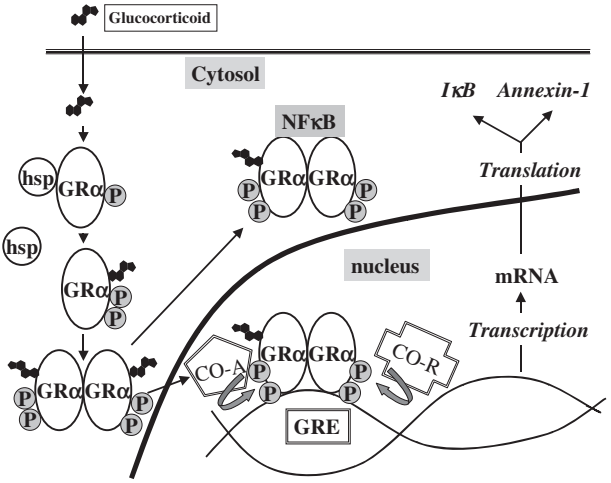


**Figure 1** Schematic diagram showing some mechanisms of action of glucocorticoids [139, 140]. Glucocorticoids enter the cell and interact with cytoplasmic GRα, which are bound to heat shock proteins (hsp), the function of which with other molecules is to maintain the GR in a conformation that allows ligand binding to occur. Release of these molecules and hyperphosphorylation of GRα result in homodimerization of the receptor–ligand complex, which migrates to the cell nucleus. There is also evidence for a role for hsp in the trafficking of the ligand-bound GR to the nucleus. Domains within the N-terminal region of the receptor interact with either negative or positive glucocorticoid response elements (GREs) in promoter regions of particular genes with consequential initiation or repression of transcription. The ligand-activated receptor on binding to DNA may also recruit coactivator (CO-A) or co-repressor (CO-R) protein complexes that enhance the effect of the glucocorticoid receptor in activating or repressing transcription. Activation of transcription results in the induction of a number of proteins that may contribute to the anti-inflammatory effect of glucocorticoids, including annexin 1 and IκB, which is a protein normally bound to the transcription factor NF-κB and which prevents it from translocating to the nucleus. The ligand-activated GRα may also interact directly with transcription factors such as NF-κB and AP-1, preventing them from binding to DNA and effecting transcription.

transcription factors such as NF-κB and AP-1. This interaction occurs in the cytoplasm (Fig. 1). NF-κB plays a key role in inflammation by activating the transcription of pivotal inflammatory cytokines such as TNFα, IL-1β, and GM-CSF, and therefore antagonism of this transcription factor contributes importantly to the anti-inflammatory effect of glucocorticoids [139].

***Mechanisms Underlying Nongenomic Activities of Glucocorticoids*** It is unclear if the rapid glucocorticoid actions, which do not directly affect the transcription or translation of inflammatory genes, involve membrane-bound glucocorticoid receptors, or are mediated via classical cytoplasmic receptors without their required nuclear translocation or interaction with transcription factors [140].

***Therapeutic Applictions*** Glucocorticoids are used to replace endogenous cortisol in adrenal insufficiency, for suppression of systemic and local inflammatory and allergic disorders (rheumatoid arthritis, asthma, allergic rhinitis, contact and atopic dermatitis, allergic conjunctivitis, and ulcerative colitis), in the treatment of shock, congenital adrenal hyperplasia, cerebral edema, nausea and vomiting associated with chemotherapy, psoriasis, and in the palliative management of leukemias and lymphomas. Hydrocortisone was the first glucocorticoid to be used, and many compounds have been introduced that have different potencies and oral bioavailabilities. These include prednisolone, methylprednisolone, betamethasone, dexamethasone, deflazocort, and triamcinolone, which may be used systemically, and aclometasone, beclometasone, clobetasol, diflocortolone, fluocinonide, fluocortolone, halobetasol, halcinonide, and mometasone, which are used topically. Beclometasone, fluticasone, budesonide, and mometasone can also be used by inhalation in the treatment of asthma. Most of the adverse effects of glucocortocoids are predictable from the physiological actions of cortisol and are related to their metabolic, immunomodulatory, and anti-inflammatory effects. These include adrenal suppression as a result of negative feedback actions on the hypothalamus and anterior pituitary, stunted growth in children, elevation of blood glucose, muscle wasting, impaired wound healing, skin atrophy, bruising, increased susceptibility to infection, and osteoporosis with associated risk of vertebral and long-bone fractures. These adverse effects markedly limit the use of glucocorticoids. The adverse effects are miminized by using topical/local treatment instead of systemic treatment, although topical and inhaled steroids can be absorbed in sufficient quantities to cause systemic effects, including adrenal suppression.

***Attempts to Dissociate the Anti-inflammatory and Adverse Effects of Glucocorticoids*** A new class of drugs, described as *dissociated glucocorticoids*, is under development. These compounds are characterized by a separation of the transactivating effects resulting from the activation of GR from those effects that repress transcription. Dissociated steroids potently inhibit

AP-1 or NF-κB, resulting in reduction of the formation of inflammatory mediators (e.g., IL-1β) but produce only weak transactivation via GREs. This should, for example, eliminate the metabolic effects of glucocorticoids. RU 24858 showed a promising profile of anti-inflammatory activity both in vitro (inhibition of lipopolysacchoride-induced IL-1β production) and in vivo (inhibition of cotton pellet granuloma; thymolysis), although it still possessed transactivational properties and exerted osteoporotic activites in vivo [140]. On the other hand, a nonsteroidal GR agonist, ZK216348, suppressed inflammation in rodent skin inflammation models, without elevating blood glucose [140]. This is a potentially very exciting line of research that may yield glucocorticoid-like anti-inflammatory agents with a much enhanced safety profile. However, it should be noted that ZK 216348 still suppressed the HPA axis.

As indicated above, glucocorticoids produce part of their inflammatory effect through increasing the formation of Anx-A1; and recombinant Anx-A1 or derived peptides can mimic many of the anti-inflammatory actions of glucocorticoids both in vitro and in animal models of acute and chronic inflammation. Their effects include inhibition of leukocyte adhesion and migration, induction of apoptosis in inflammatory cells, suppression of the production of inflammatory mediators by inhibition of $PLA_2$ (phospholipase $A_2$), COX-2 (cyclooxygenase-2), and iNOS (inducible nitric oxide synthase) and the induction of the anti-inflammatory cytokine IL-10 [140]. Development of analogs of these peptides may yield useful anti-inflammatory drugs.

***Glucocorticoid Antagonists*** Glucocorticoid antagonists have been synthesized. RU486 (mifepristone) potently inhibits the glucocorticoid activitity of dexamethasone and has high affinity for the glucocorticoid and progesterone receptors, displays partial agonist activity at the glucocorticoid receptors in some tissues, and behaves as a competitive antagonist in others.

## Inhibitors of Corticosteroid Synthesis

Corticosteroids (chiefly cortisol and aldosterone) are synthesized from cholesterol via pregnenolone, utilizing a number of cytochrome P450 enzymes [141]. Thus, P450 cholesterol side-chain cleavage enzyme (CYP11A) catalyses the conversion of cholesterol to pregnenolone, which is then converted to progesterone under the influence of 3β-hydroxysteroid dehydrogenase. Progesterone may be 17-hydroxylated to 17α-hydroxyprogesterone (using steroid 17α-hydroxylase, CYP17) giving rise to cortisol via 11-deoxycortisol, the conversion of which to cortisol is catalyzed by steroid 11β-hydroxylase (CYP11B1). Alternatively, progesterone may be converted to deoxycorticosterone under the influence of steroid 21-hydroxylase (CYP21), giving rise to aldosterone (via corticosterone and 18-hydroxycorticosterone) under the influence of aldosterone synthase (CYP11B2). Several drugs inhibit one or more of the enzymes involved in these pathways resulting in inhibition of corticosteroid synthesis.

***Therapeutic Applications*** Inhibitors of corticosteroid synthesis are used in the treatment of Cushing's syndrome and of certain cancers. Metyrapone, which inhibits steroid 11β-hydroxylase, is also used in the differential diagnosis of corticotrophin-dependent Cushing's syndrome.

*Aminoglutethimide* is an inhibitor of steroid synthesis in the adrenal cortex, where it inhibits a number of cytochrome P450-mediated steps, including cholesterol side-chain cleavage enzyme. It also inhibits the cytochrome P450 aromatase enzyme responsible for the conversion of androgens to estrogens [142]—see later.

*Mitotane*, a metabolite of the insecticide DDT, inhibits adrenal steroid synthesis and is cytotoxic to adrenal cortical cells. It is used in the treatment of inoperable adrenal tumors and in some patients with Cushing's disease. The mechanism of action of mitotane remains unclear, but part of its action may be due to inhibition of cyclic-AMP generation in response to corticotrophin [143]. It may inhibit other steps in steroidogenesis, including inhibition of 11 β-hydroxylase [144]. It is also cytotoxic to adrenal cortical cells [145, 146]; at least in dog adrenocortical cells there is no clear separation of the concentration–response curves in terms of inhibition of steroid production and reduced cell viability, each being evident at 3 μM and probably maximal at 100 μM [146].

*Trilostane* is a selective, reversible, competitive inhibitor of 3β-hydroxysteroid dehydrogenase, a key enzyme involved in the synthesis of steroids. It inhibits the conversion of pregnenolone to progesterone, therefore inhibiting the synthesis of progesterone and therefore of the adrenal, gonadal, and placental steroids derived from progesterone, including cortisol, aldosterone, androgens, and estrogens. It does not inhibit the synthesis of cortisol from progesterone, 17-hydroxyprogesterone or 11-deoxycortisol in guinea pig isolated adrenal cortical cells [147, 148]. Trilostane may be used in the treatment of Cushing's syndrome [149], although responses may be variable. It has also found use in the treatment of breast cancer in postmenopausal women [150] who have experienced a relapse after initial treatment with estrogen receptor antagonists. Trilostane was shown to be an effective alternative to mifepristone for the midterm termination of pregnancy, when administered prior to misoprostil [151].

The antifungal agent *ketoconazole* is also a potent inhibitor of steroid biosynthesis, by inhibiting steroid 11β-hydroxylase and also, more potently, the conversion of 17α-hydroxyprogesterone to androstenedione under the influence of 17α-hydroxylase [152]. It has been used in Cushing's syndrome. *Etomidate*, an anesthetic agent, is a potent inhibitor of 11β-hydroxylase [153], which limits its use to induction, rather than maintenance of anesthesia and prevents its use in long-term sedation in traumatized patients [154].

## Aldosterone and Other Mineralocorticoids

Aldosterone is the endogenous mineralocorticoid and is of fundamental importance in the maintenance of water and electrolyte balance, producing

sodium and water retention and potassium excretion. It regulates sodium absorption in the distal convoluted tubule of the kidney and has analogous actions at other epithelial sites such as the colon [155]. Aldosterone is synthesized primarily in the zona glomerulosa of the adrenal cortex, which unlike other regions of this gland expresses aldosterone synthase (CYP11B2) [156]. There is now much evidence that aldosterone synthase is expressed in cardiac and vascular tissue and that aldosterone is synthesized in these tissues [157]. Aldosterone secretion is regulated primarily by the renin-angiotensin system through angiotensin II, which acts directly on the zona glomerulosa to stimulate hormone synthesis [156]. Potassium also stimulates aldosterone synthesis, through a direct action on glomerulosa cells. Although cortisol has high affinity for the mineralocorticoid receptor, it normally lacks significant mineralocorticoid activity because it is rapidly oxidized by the high affinity, type 2, 11β-hydroxysteroid dehydrogenase expressed in high levels in the kidney and colon [158]. Mutations in the gene for this enzyme (*HSD11B2* gene) result in the syndrome of apparent mineralocorticoid excess due to excessive activation of the mineralocorticoid receptor by cortisol. Other endogenously produced compounds with mineralocorticoid activity are 11-deoxycorticosterone, 18-hydroxydeoxycorticosterone, corticosterone, and 19-nordeoxycorticocosterone [141]. Apart from its fundamentally important physiological role, aldosterone is clearly important in pathophysiology, where it makes an important contribution in cardiac remodeling in heart failure [159] and in various forms of hypertension. It may also contribute to endothelial dysfunction, reduced fibrinolysis, and cardiac arrhythmias.

The only mineralocorticoid in general use is fludrocortisone, which is a synthetic, orally active 9-α fluorinated analog of cortisol and that at normal therapeutic doses has only mineralocorticoid activity. Its mineralocorticoid potency relative to that of cortisol has been attributed to its resistance to 11-β-oxidation [160] by type 2, 11β-hydroxysteroid dehydrogenase. Its main use is in replacement therapy in adrenal cortical insufficiency, and it is the drug of first choice in the treatment of orthostatic hypotension [161].

***Receptors Mediating the Effects of Aldosterone*** Aldosterone acts at the mineralocorticoid receptor (MR) and its classical effects are exerted through modulation of the transcription rates of various genes, including those encoding subunits of the epithelial sodium channel (ENaC) and the $Na^+K^+$ATPase [162]. It is now widely accepted that aldosterone exerts rapid nongenomic actions [163], which may be mediated by the classical MR or by other, as yet ill-defined receptors. The role of membrane-bound aldosterone receptors is uncertain, but there is evidence for their existence and for their resistance to classical aldosterone MR antagonists [164].

## Mineralocorticoid Antagonists

Mineralocorticoid antagonists prevent the action of the aldosterone at the MR. In view of the pathophysiological role of aldosterone, these drugs have

a potentially major value in the treatment of heart failure. A major randomized trial (RALES—Randomized ALdactone Evaluation Study) showed that addition of the aldosterone antagonist spironolactone to standard therapy in patients with congestive heart failure reduced mortality by 30 percent [165, 166]. This effect was reproducible in a rat model of heart failure, in which administration of another aldosterone antagonist, canrenone (a metabolite of spironolactone), attenuated left ventricular remodelling and improved cardiac function [167]. Spironolactone produces adverse effects including gynecomastia, breast tenderness, menstrual irregularities, and impotence. Other aldosterone antagonists have been developed including mexrenone and eplerenone. These drugs have similar affinity to spironolactone for the mineralocorticoid receptor but had a much reduced affinity for androgen and progesterone receptors and have markedly reduced antiandrogenic and progestagenic effects compared to spironolactone in animals [168] and humans [169, 170].

## 7 ANDROGENS

The principal natural androgens are testosterone and dihydrotestosterone, which is produced from testosterone in tissues expressing 5α-reductase activity. Testosterone is secreted primarily by the testes and significant amounts, primarily as the precursor dehydroepiandrosterone, originate in the adrenal cortex, which is the main source of androgens in the female. Androgens play a key role in male reproduction, being responsible for the development and maintenance of the reproductive ducts (epididymis, vas deferens, ejaculatory ducts), glands (prostate, seminal vesicles, bulbourethral glands), and the external genitalia. They are also required for spermatogenesis and the development and maintenance of the male secondary sexual characteristics. Androgens play a key role in the masculinization of external genitalia (function of dihydrotestosterone) and the development of the epididymis, vas deferens, and seminal vesicles from the Wolffian duct (function of testosterone) in the male fetus. The therapeutic applications of androgens are currently quite limited and they are used primarily for androgen-replacement therapy in hypogonadal males.

### Androgen Receptors

There appears to be a single androgen receptor, which is a member of the nuclear receptor superfamily, typically having an N-terminal domain implicated in activation of transcription, a DNA binding domain and a ligand binding domain. Like glucocorticoid receptors, binding of the ligand causes the cytoplasmic receptor to dissociate from heat shock proteins leading to dimerization and phosphorylation of the receptor and its translocation to the nucleus. Within the nucleus the receptor binds to the androgen response

element (ARE). This results in recruitment of hormone- and cell-specific coactivators and activation of transcription [171].

The androgen receptor is widely distributed and, in addition to the male reproductive tract, it is expressed in many tissues including the female external genitalia, skeletal muscle, skin, sebaceous glands, hair follicles, and the central nervous system.

## Anabolic Steroids

The anabolic action of the main male sex hormone testosterone is well known and this property is retained in synthetic derivatives such as nandrolone (19-nortestosterone). Synthetic, 17α-alkylated compounds (oxymethalone, stanozolol, oxandrolone, danazol) are orally active and, in animals, show selectivity for anabolic versus classical androgenic activity, although the evidence for such selectivity in the human is unclear. These agents are used for their anabolic properties in several disease states [172]. All can produce concomitant androgenic side effects and the 17α-alkylated compounds may produce hepatotoxicity. The anabolic actions explain their abuse by athletes who often consume much greater than therapeutic doses, in an attempt to increase muscle mass. Whether performance is actually increased by these agents is uncertain, although a randomized, double-blind, placebo-controlled trial showed that a weekly, large dose of testosterone enanthate increased muscle strength, especially when combined with exercise [173]. Large doses of these drugs have been shown to upregulate androgen receptors, and this may contribute to their anabolic effects [174].

Anabolic steroids act at the androgen receptor. As there is only one androgen receptor, differences in effects in different tissues remain unexplained. Reduced androgenicity of nandrolone is explained by the reduction in its activity by 5α-reductase. Thus it has reduced activity in tissues expressing high levels of this enzyme and full activity in tissues that express low levels of, or no, 5α-reductase [175, 176]. There is also evidence that they act as antiglucocorticoids by competing for binding at the glucocorticoid receptor [177] and thus may exert an anticatabolic effect.

There is great interest in separating the anabolic and androgenic effects of androgens. Although there is only one androgen receptor, the effect of the androgen–receptor combination may vary in a cell-context-specific way because of the different expression of coactivators in different tissues. Moreover, binding of different agents may result in particular receptor conformations that recruit different coactivators. Thus, by analogy with SERMs (see below), the concept of selective androgen receptor modulators (SARMs) has been proposed [178]. Some evidence to support this possibility is derived from the development of a nonsteroidal androgenic agonist agent showing increased efficacy and potency for anabolic effects (levator ani weight) relative to androgenic effects (weight of prostate/seminal vesicles) in castrated rats [179]. Although the tissue selectivity was not very marked, this may indicate

the feasibility of this approach for separating androgenic and anabolic properties.

## Antiandrogens

Compounds interfering with the actions of androgens may be steroidal (cyproterone acetate; mifepristone) or nonsteroidal [flutamide; hydroxyflutamide (the active metabolite of flutamide); nilutamide; bicalutamide]. While the nonsteroidal agents are selective for the androgen receptor, the steroidal compounds display partial agonist activity at this receptor and also antagonist activity at progesterone and glucocorticoid receptors [180]. Binding of antagonists to the androgen receptor prevents the receptor conformation change required to effect the activation of transcription [180].

Nilutamide, flutamide, and bicalutamide are used alone, or in combination with gonadorelin analogs, in the treatment of advanced prostate cancer. These drugs have also had limited evaluation in the management of benign prostatic hypertrophy and also to treat hirsutism and androgenic alopecia in women with polycystic ovarian disease [181]. When used in men, common, predictable adverse effects are gynecomastia, possibly accompanied by galactorrhea, decreased libido, and inhibition of spermatogenesis.

## 5α-Reductase Inhibitors

The conversion of testosterone to dihydrotestosterone, which has a much higher affinity for the androgen receptor, is an important activating mechanism. This conversion is catalyzed by the enzyme 5α-reductase, which is expressed in two isoforms, 5α-reductase type 1 and 5α-reductase type 2, which have only ~50 percent homology and which have different enzymic properties and tissue distribution [182]. The normal prostate, genital skin, seminal vesicles, dermal papilla, and developing fetal male external genitalia predominantly express type 2, whereas the type 1 isoform is the predominant form in established prostate cancer, the scalp, nongenital skin, and the sebaceous gland [182]. There are also marked species differences, which lead to marked differences in the potencies and selectivities of inhibitors, for example, when tested against human enzymes compared with rat enzymes. Thus there is enormous potential for developing type-selective inhibitors that will have applications in benign prostatic hyperplasia (BPH), prostatic cancer, androgenic alopecia, hirsutism, and acne. Finasteride is an azasteroid drug that is selective for the type 2 isoenzyme and that is marketed for the treatment of BPH, in which it causes a marked reduction in serum dihydrotestosterone concentrations, accompanied by a 15 to 25 percent reduction in gland size and reduction in urinary symptoms. Duasteride is a mixed type 1/type 2 inhibitor. Many nonsteroidal compounds, with high selectivity for one or other isoforms are under development [182].

## 8 ESTROGENS

The estrogens are secreted primarily by the gonads with smaller amounts originating in the adrenal cortex. The main physiological estrogen is estradiol-17β. Estrogens play key roles in the development and maintenance of the adult, female reproductive tract, and female secondary sexual characteristics but also have important effects in nonreproductive tissues, particularly bone, the cardiovascular system, and the brain. Estrogens are also implicated in breast and uterine tumor progression.

### Estrogen Receptors

Like receptors for glucocorticoids, androgens, and progestins, the estrogen receptor (ER) belongs to the superfamily of nuclear receptors mediating their effects through alteration in transcription. The estrogen receptor, like the glucocorticoid receptor, is bound to heat shock proteins in the resting state, these being displaced upon hormone binding. The activated receptor dimerizes and interacts with specific DNA estrogen response elements. The conformation of the ligand-bound receptor depends on whether the ligand is an agonist or an antagonist. Binding of a full agonist displaces co-repressor proteins and recruits coactivator proteins that then result in gene transcription. On the other hand antagonist binding favors continued interaction with co-repressors and no transcription occurs. Other drugs, such as tamoxifen, a selective estrogen receptor modulator (SERM), promote a conformation that allows partial interaction with both co-repressors and coactivators. The level of different coactivators and co-repressors may be cell-type dependent, which may explain cell-context-dependent activities of different ligands at the estrogen receptor [183].

Two estrogen receptors are now known—ERα and ERβ. These are both nuclear receptors that, although having considerable homology at the DNA binding domains and therefore similar binding affinities for estrogen response elements, have relatively low overall sequence homology of ~47 percent [184, 185]. These receptors show different tissue distribution. For example, ERα is predominant in uterus, prostate stroma, ovarian thecal cells, and bone, whereas ERβ is expressed predominantly in colon, ovarian granulosa cells, bone marrow, and brain dorsal raphe [185].

Estrogens, like other steroids, can also exert rapid (within minutes), nongenomic effects, some of which are prevented by estrogen antagonists but others of which are not and appear to involve cAMP-mediated mechanisms [186]. For example, the rapid vasodilator effect of estradiol is prevented by the estrogen antagonist ICI 182, 780 and appears to be mediated by endothelial-derived nitric oxide produced under the influence of eNOS activated via a MAP kinase-dependent pathway through ERα [187].

### Estrogen Agonists

The main natural estrogen is estradiol-17β but other natural estrogens include estriol and estrone. The main synthetic estrogens in use are ethinyl estradiol and mestranol, which is inactive until metabolized to ethinyl estradiol. Stilbestrol is a potent, nonsteroidal, synthetic estrogen agonist. Until recently, there were no agonists or antagonists selective for ERα compared with ERβ receptors or vice versa. However, propyl pyrazole triol [PPT; 4-propyl-1,3,5-Tris (4-hydroxyphenyl) pyrazole] appears to have selective agonist activity at Erα and diarylpropiol nitrile appears to have selective ERβ agonist activity [188]. These drugs should facilitate the characterization of responses mediated via the two different estrogen receptors.

### Antagonists at Estrogen Receptors

Clomiphene and ICI 182, 780 behave as pure antagonists at the estrogen receptor. An ERα-selective antagonist, methyl-piperidino-pyrazole, has also been described [188].

### Selective Estrogen Receptor Modulators (SERMs)

Tamoxifen was originally considered as an estrogen antagonist and, accordingly, is widely used in the treatment of estrogen-dependent breast cancer. While clearly acting as an antagonist in the breast, tamoxifen functions as an agonist in bone, uterus, and the cardiovascular system [183]. Raloxifene, on the other hand, shows no agonist effects in the uterus but is an agonist in bone [183]. Because of this behavior as agonists in some tissues and antagonists in others, these drugs and others (e.g., idoxifene, droloxifene, torimifene, droloxifene, ospemifene, lasofoxifene, and arzoxiphene) are referred to as SERMs (selective estrogen receptor modulators).

### Aromatase Inhibitors

Aromatase is responsible for the last stage in the synthesis of estrogens, either from testosterone to give estradiol or from the androgenic testosterone precursor androstanedione to give estrone, which can then be converted into estriol or estradiol. There is much interest in developing potent selective inhibitors of this enzyme, as these have application in the treatment of estrogen-dependent conditions such as breast cancer [189] and endometriosis [190] and in promoting ovulation by removing the estrogen-mediated negative feedback inhibition of gonadotrophin secretion [191]. The first agent to be used clinically was aminoglutethamide (see earlier) but is nonselective, inhibiting several steps in steroid biosynthesis and also inducing aromatase activity [192].

This induction may explain the failure of aminoglutethimide therapy in some cancer patients. Steroidal inhibitors, such as the androstanedione analogs

exemstane and formestane, bind irreversibly to the catalytic site of the enzyme. They are administered parenterally and have androgenic activity. Nonsteroidal, imidazole compounds (anastrozole, letrozole) bind to the cytochrome P450 site of the enzyme. They can be administered orally and have no androgenic activity [189].

***Therapeutic Applications of Agents Acting at Estrogen Receptors*** Agonists are used extensively in hormone replacement therapy following gonadal failure or to replace estrogen in postmenopausal women suffering from menopausal symptoms (hot flushes, vaginal dryness, postmenopausal osteoporosis). They may be used alone or together with a progestin, which is indicated in postmenopausal women with an intact uterus, in order to minimize the risk of endometrial cancer associated with estrogenic stimulation of the uterus. While women benefit hugely from the relief of menopausal symptoms afforded by estrogen replacement therapy and clearly have a reduced risk of osteoporosis (as evidenced by hip fractures) and colorectal cancer, there is controversy concerning whether or not long-term hormone replacement therapy in the menopause is associated with an increased risk of stroke and coronary heart disease, as suggested by a large, randomized trial [193]. Certainly there is no evidence to support the estrogen-mediated protection from coronary artery disease that would be predicted from the very low prevalence of myocardial infarction in premenopausal women.

The other major use is in pharmacological contraception, where they are administered along with a progestin, resulting in failure of ovulation as a result of suppression of the hypothalamic secretion of GnRH and inhibition of gonadotrophin secretion. The progestin component also prevents estrogen-mediated endometrial development and changes in cervical mucus.

The major use of antagonists is in treating cancer of the breast. The SERM tamoxifen is the main drug used for this purpose. SERMs such as raloxifene, which are devoid of uterine stimulant activity but retain agonist activity in bone, are used to treat postmenopausal osteoporosis. Additionally, estrogen antagonists have potential in the treatment of endometriosis.

## 9 PROGESTINS

The natural progestin progesterone is secreted from the corpus luteum during the luteal phase of the menstrual cycle, where it prevents further endometrial proliferation and stimulates endometrial gland branching and secretion. Progesterone plays a key role in mammary gland development and in maintaining pregnancy. In pregnancy, the placenta is the major source of progesterone after the first trimester.

***Progesterone Receptors*** The progesterone receptor is a typical member of the superfamily of nuclear receptors. There are two isoforms of the progesterone receptor—PR-A and PR-B—and these show different tissue expres-

sion and function. For example, in the mouse uterus, where progesterone inhibits the proliferative effect of estrogen, it appears that PR-A mediates the inhibitory response, while PR-B may actually stimulate proliferation [194]. Both PR isoforms are expressed in the mouse mammary gland and although both may be required for mammary gland development, ablation of PR-A in knockout mice did not affect the ability of PR-B to mediate the effect of progesterone in producing ductal side branching and development of the alveoli, suggesting the predominant role of PR-B in this tissue [194]. As with glucocorticoids and estrogens, progestins can produce nongenomic effects. For example, progesterone rapidly (within 2 to 5 min) stimulates insulin secretion in isolated islets and perfused pancreas and reduces glucose uptake by adipocytes [195]. It also inhibits proliferation in human transformed cell lines from the uterine cervix, cells that lack steroid receptors and in which this effect of the hormone is not blocked by progesterone antagonists [196].

## Agonists at Progesterone Receptors

Many compounds are available with agonist activity at progesterone receptors. Their pharmacology is complex because many, either directly or through their metabolites, bind to other steroid receptors resulting, for example, in estrogenic activity in some (norethisterone, lynestenol, norethindorel), androgenic (norethisterone, lynestenol, levonorgestrel, gestodene) and glucocorticoid (chlormadinone acetate, megestrol acetate, medroxyprogesterone acetate), and antimineralocorticoid (drospirenone, gestodene) activities [197]. It is therefore very difficult to make general statements about the pharmacology of these agents.

## Progesterone Antagonists

The best known antagonist of progesterone is mifepristone (RU 486), which is used to terminate pregnancy and which may be used for postcoital contraception. Antagonists at the progesterone receptor also have contraceptive potential as a consequence of their ability to prevent endometrial maturation. By analogy with SERMs, there is interest in developing selective progestin receptor modulators.

***Therapeutic Applications of Progestins*** Progestins are widely used in pharmacological contraception, where they are used either alone or in combination with the estrogens ethinyl estradiol or mestranol or along with estrogens in hormone replacement therapy. The controversy concerning whether or not long-term hormone replacement therapy in the menopause is associated with an increased risk of stroke and coronary heart disease, as suggested by the Women's Health Initiative trial [193], may partly relate to the use in that study of the progestin medroxyprogesterone acetate, which has weak androgenic activity.

## 10 TREATMENT OF DIABETES MELLITUS

The prevalence of diabetes mellitus, especially of Type 2 diabetes, is increasing, a world total of >300 million being predicted by 2030. This high prevalence combined with the associated increased mortality and morbidity, primarily as a result of macrovascular disease and microvascular long-term complications make diabetes mellitus a major health problem. Overwhelming evidence, especially from the DCCT study [198] for Type 1 diabetes and the UKPDS study [199] for Type 2 diabetes suggests that good metabolic control would markedly reduce mortality and morbidity.

While Type 1 diabetes mellitus is due to autoimmune destruction of the β cells and the consequent severe insulin deficiency, it is now broadly accepted that Type 2 diabetes results from both peripheral insulin insensitivity and impaired insulin secretion [200]. There is progressive islet β-cell failure in patients with Type 2 diabetes and a reduction in the β-cell mass [201]. Thus as well as developing drugs that improve target tissue responsiveness to insulin, the islet β cell remains an important target for the development of drugs for treating Type 2 diabetes. Drugs are required that both enhance insulin secretion in response to normal, physiological, meal-related nutrient stimuli and that prevent the progressive loss of islet β-cell mass. Augmentation of meal-related insulin secretion by drugs should be glucose dependent. This will ensure that increased insulin secretion occurs when required, the corollary being that insulin secretion would remain at a basal rate between meals, avoiding hyperinsulinemia and thus hypoglycemia, as seen with sulphonylureas, and the potential cardiovascular risks associated with chronic hyperinsulinemia and insulin resistance [202, 203]. Physiological, meal-related insulin secretion occurs rapidly in response to absorbed nutrients, primarily glucose. Glucose stimulates insulin secretion following its transport into and metabolism in the β cell, generating ATP, which closes $K_{ATP}$ channels, resulting in depolarization and calcium influx (Fig. 2). The effect of glucose on insulin secretion can be amplified by activation of both adenylyl cyclase and phospholipase C catalyzed pathways [204]. The early rapid insulin secretory response to glucose is important in maintaining normal blood glucose, and it is widely accepted that loss of this early-phase insulin secretion is an important and early event in the development of Type 2 diabetes [205].

### Insulin

Insulin is used to replace the deficient hormone in the treatment of diabetes, and currently there is no alternative therapy for Type 1 diabetes. Insulin is also used in the treatment of Type 2 diabetes when this cannot be adequately controlled by orally active antidiabetic drugs. The aim of treatment using insulin is to maintain euglycemia (a plasma glucose level of 4 to 7 mmol/L) without causing hypoglycemia. However, good control is difficult to achieve because of the difficulty of administering insulin in a way that mimics physiological

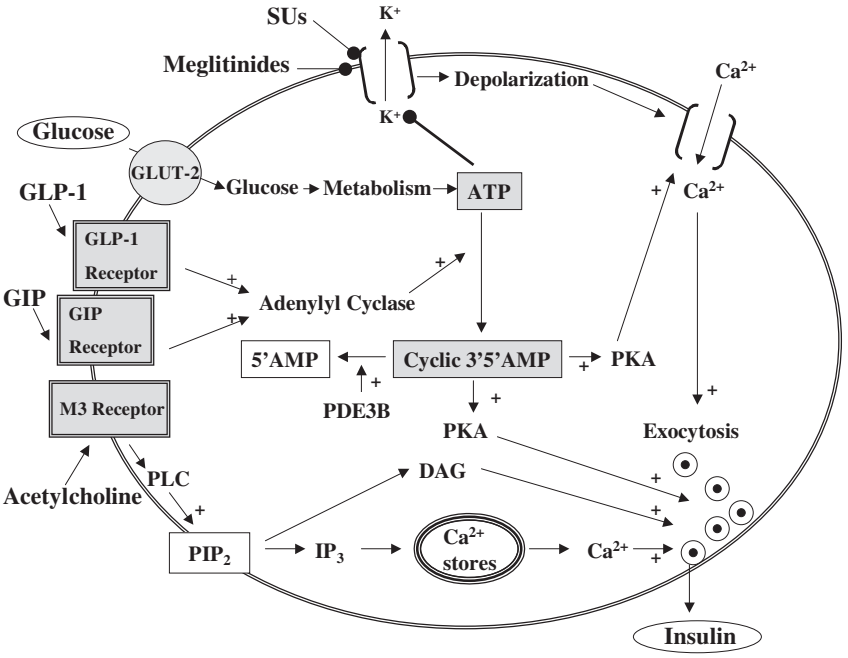


**Figure 2** Schematic illustration of glucose-induced insulin secretion and its amplification mechanisms. Glucose is transported into the pancreatic islet β cell using the transporter GLUT-2. It is metabolized producing ATP. The resultant increase in the ATP : ADP ratio inhibits the plasma membrane $K_{ATP}$ channel, blocking potassium efflux and resulting in depolarization of the β-cell membrane. Depolarization activates L-type voltage-sensitive calcium channels, allowing influx of $Ca^{2+}$ ions, which trigger exocytosis. The sulfonylureas (SUs) and meglitinide analogs also block the $K_{ATP}$ channel by binding to its SUR1 subunit. The effect of glucose is amplified by the second-messenger cyclic-3′5′AMP, inositol 1,4,5-trisphosphate ($IP_3$), and diacylglycerol (DAG). Cyclic-3′5′AMP is produced from ATP as a consequence of activation of adenylyl cyclase, primarily by incretin peptides glucagon-like peptide-1 (GLP-1) and glucose-dependent insulinotropic polypeptide (GIP) secreted by the intestines in response to nutrients. The effect of cyclic-3′5′AMP is terminated by the phosphodiesterase isoform PDE3B. Cyclic-3′5′AMP activates protein kinase A (PKA), which acts at a number of sites (releasing $Ca^{2+}$ from intracellular stores, enhancing $Ca^{2+}$ through L-type channels, and augmenting the effects of $Ca^{2+}$ on exocytosis) to augment the effects of glucose and also produces effects independent of PKA. $IP_3$ and DAG are produced from the membrane lipid phosphatidyl inositol 4,5-bisphosphate ($PIP_2$) by the activation of phospholipase C (PLC). Vagally released acetylcholine acting through the muscarinic $M_3$ cholinoceptor is probably the major physiological activator of this mechanism. $IP_3$ activates the release of $Ca^{2+}$ from intracellular stores and DAG activates protein kinase C (PKC), which amplifies the effect of $Ca^{2+}$ in promoting exocytosis (a + associated with → denotes activation/stimulation; • denotes inhibition).

insulin secretion, with rapid peaks during and immediately following a meal and low, basal concentrations between meals.

Insulin preparations are now largely based upon human insulin prepared by enzymic modification of porcine insulin, by chemical combination of the A and B chains produced using genetically modified bacteria, or from a precursor produced by yeast modified by recombinant DNA technology.

Insulin analogs, with different onsets and durations of action, have been produced. In solution, soluble insulin forms hexamers, which must dissociate for the absorption of insulin from a subcutaneous injection site. If amino acid substitutions are made in the insulin B chain (B28 proline replace by lysine and B29 lysine by proline yielding lispro insulin, or B28 proline by aspartate yielding insulin aspart), the tendency to form hexamers is diminished and this increases bioavailability so that the analogs are absorbed much more rapidly [206]. On the other hand, elongation of the C-terminal end of the B chain of the insulin molecule with two arginine residues and substitution of A21 asparagine by glycine elevates the isoelectric point of insulin from pH 5.4 toward neutral. This analog, insulin glargine, crystallizes in the subcutaneous injection site, providing a very long delay in absorption. Long-acting insulin analogs injected once daily may be combined with the rapid onset, short-acting analogs injected with each meal, which should provide a much more physiological plasma insulin profile than could be achieved using conventional insulin preparations [207].

***Mechanism of Action of Insulin*** The insulin receptor is a tetrameric protein belonging to the subfamily of receptor tyrosine kinases. Binding of insulin to the α subunit activates the tyrosine kinase activity of the β subunit, leading to receptor autophosphorylation and also to tyrosine phosphorylation of the insulin receptor substrate (IRS) adaptor proteins. The phosphorylation of tyrosine sites on IRS allow its interaction with SH2 (Src-homology-2) proteins, including Grb2, which leads to activation of MAP kinase and alterations in protein synthesis and the p85 regulatory subunit of PI-3-kinase (phosphatidylinositol 3-kinase). PI-3 kinase plays a pivotal role in insulin action in enhancing glucose uptake by mobilization of GLUT-4 glucose transporters in skeletal muscle and adipose tissue and, via PKD1 [PtdIns(3,4,5) $P_3$-dependent protein kinase], in activating PKB (protein kinase B), which itself inactivates GSK3 (glycogen synthase kinase 3). Inactivation of GSK3 leads to activation of glycogen synthase and of the initiation factor eIF2B, resulting, respectively, in increased glycogen synthesis and increased protein synthesis [208–210] (Fig. 3).

Understanding the insulin receptor-mediated signaling may allow the development of small, nonpeptide molecules that activate the signaling pathways downstream of the receptor. For example, this could be achieved by a molecule that mimicked PtdIns(3,4,5) $P_3$ in enabling PDK1 to activate PKB. Similarly, a drug that inhibited GSK3 would also mimic effects of insulin [208].

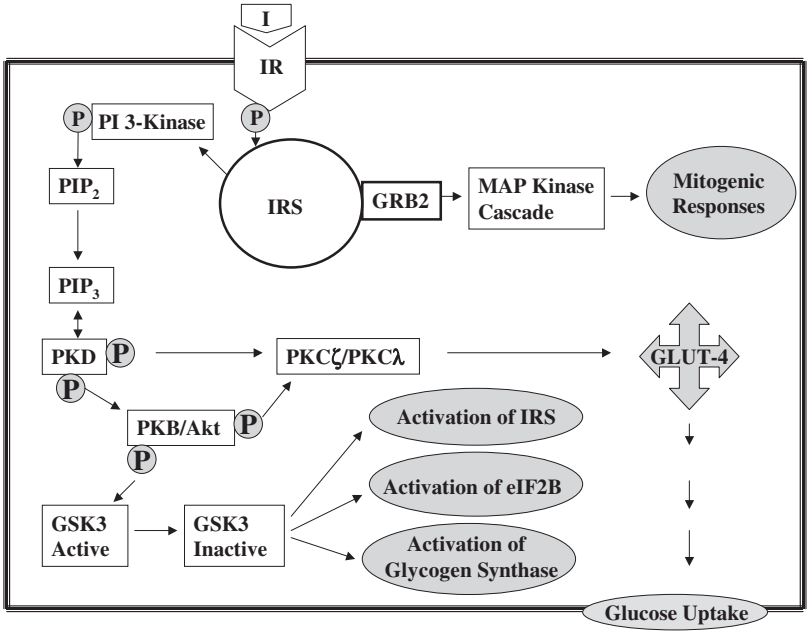


**Figure 3** Mechanism of action of insulin [208–210]. Insulin (I) binds to the α subunits of the insulin receptor (IR), resulting in activation of the tyrosine kinase of the β subunit. Autotyrosine phosphorylation of the receptor results in its interaction with and phosphorylation of insulin receptor substrates (IRS). Phosphorylated sites on IRS interact with various proteins containing Src homology (SH) domains. These include phosphatidylinositol 3-kinase (PI 3-kinase) and growth factor receptor binding protein 2 (Grb2). Interaction with Grb2 results in activation of the MAP kinase cascade and hence mitogenic responses. Activation of PI 3-kinase leads to phosphorylation of the membrane lipid phosphatidyl inositol 4,5-bisphosphate ($PIP_2$) producing phosphatidyl inositol (3,4,5)-trisphosphate ($PIP_3$). $PIP_3$ interacts with protein kinase B (PKB/akt), recruiting it to the membrane and allowing its phosphorylation and activation by phosphatidyl inositol (3,4,5)-trisphosphate-dependent protein kinase (PDK). PKB phosphorylates a number of proteins, including glycogen synthase kinase 3 (GSK3), the activity of which normally inhibits the activity of several enzymes including glycogen synthase, eukaryotic initiation factor 2B (eIF2B), and IRS. Phosphorylation of GSK3 inactivates the enzyme increasing the activity of these enzymes. Thus synthesis of glycogen and protein are activated and insulin signaling through IRS is augmented. PDK (possibly through PKB) also phosphorylates atypical protein kinase C (PKCξ/λ), which activates the trafficking of the glucose transporter (GLUT-4) to the plasma membrane, thereby stimulating glucose uptake (P→ denotes phosphorylation).

## Biguanides

The biguanides are derivatives of the compound biguanide (guanylguanidine) that exert a blood glucose-lowering effect in Type 2 diabetes mellitus. Metformin is now the only biguanide in clinical use [211, 212], following the withdrawal of phenformin and buformin in most countries in the late 1970s due to a high incidence of associated lactic acidosis. Metformin, which has a much lower risk of lactic acidosis, is used widely in the treatment of Type 2 diabetes. Biguanides do not increase plasma insulin concentrations and do not cause severe hypoglycemia; hence they are regarded as antihyperglycemic (rather than hypoglycemic) agents. The United Kingdom Prospective Diabetes Study (UKPDS), a large, randomized, controlled trial over >10 years in Type 2 diabetic patients, noted that patients who received metformin as initial oral antidiabetic drug therapy had lower morbidity and mortality from macrovascular disease [213]. Since glycemic control achieved with metformin was similar to that achieved with other antidiabetic agents (sulfonylureas and insulin) studied in the trial, it is possible that additional effects of metformin may be involved. These could include an improved lipid profile and increased fibrinolysis, which could be related to the fact that metformin does not elevate, and may reduce, serum concentrations of insulin.

Metformin increases insulin sensitivity in liver and muscle, suppressing hepatic gluconeogenesis and glycogenolysis and enhancing insulin-stimulated glucose uptake into skeletal muscle. It also exerts an insulin-independent suppression of fatty acid oxidation and a reduction in hypertriglyceridemia. These effects collectively reduce insulin resistance and glucotoxicity in Type 2 diabetes [214]. There is considerable evidence that AMP-activated protein kinase (AMPK) is an important target for metformin, which activates this enzyme in vitro and in vivo [215]. Activation of AMPK produces very similar effects to those of metformin, decreasing hepatic glucose production and gluconeogenesis, inhibiting lipolysis, lipogenesis, and lipogenic gene expression, and increasing skeletal muscle glucose uptake [216]. In addition to its widespread use in Type 2 diabetes mellitus, metformin is increasingly being used to treat the polycystic ovary syndrome (PCOS), where it leads to a decrease in serum insulin and androgen levels, as well as improved in ovulatory function [217]. It may also be useful in obesity and a randomized double-blind placebo-controlled trial in nondiabetic, obese, hyperinsulinemic adolescents on a low-calorie diet showed metformin to produce weight loss, a decrease in body fat, reduced hyperinsulinemia, and enhanced insulin sensitivity compared to placebo [218].

## Thiazolidinediones

Thiazolidinediones (TZDs), also termed *glitazones*, are used as antidiabetic agents for the treatment of Type 2 diabetes mellitus. Several have been developed and rosiglitazone and pioglitazone are in widespread use, although trogli-

tazone was withdrawn soon after marketing because of idiosyncratic hepatotoxicity. They appear to lower blood glucose by increasing the sensitivity to insulin, which can be demonstrated both in animals and humans. For example, in patients with Type 2 diabetes, troglitazone produced a 30 percent increase in insulin-stimulated glucose disposal [219].

***Receptors for Thiazolidinediones*** Thiazolidinediones act as agonists of the nuclear peroxisome proliferator-activated receptor-γ (PPARγ) at which they bind with very high affinity. PPARγ are nuclear receptors belonging to a superfamily of transcription factors (PPARα, PPARδ, and PPARγ) and that regulate gene transcription in response to small lipophilic ligands, prostaglandin J2 being one natural agonist at this receptor. PPARγ are most strongly expressed in adipose tissue and weakly expressed in liver and skeletal muscle [219–222]. The receptor is found as a heterodimer with the retinoid receptor (RXR), which, in the resting state, binds a co-repressor complex that inhibits transcription. Binding of a thiazolidinedione results in dissociation of the co-repressor complex and recruitment of coactivators, leading to increased transcription.

The mechanisms underlying the increase insulin sensitivity in response to activation of PPARγ are complex and may involve both direct effects on muscle and indirect effects mediated by actions on adipose tissue. Thus activation of PPARγ in adipose tissue leads to increased release of adipocyte-derived hormones that increase insulin sensitivity (adiponectin) and a decrease in the release of those that decrease insulin sensitivity (resistin, TNFα, IL-6). Leptin production is reduced, although this cannot explain an increase in insulin sensitivity. Activation of PPARγ diminishes the release of free fatty acids from adipocytes, which itself will augment insulin sensitivity in muscle, due to the well-known effect of free fatty acids in inhibiting insulin-stimulated glucose uptake by skeletal muscle [219–222]. Paradoxically, mice in which PPARγ are partially deleted (PPARγ +/–) are more insulin sensitive than wild-type animals and have less adipose tissue and put on less weight on a high-fat diet. This raises speculation that PPARγ antagonists may be of value in treating obesity and Type 2 diabetes. There is therefore now interest in developing, by analogy with SERMs, selective PPARγ modulators (SPPARMs) that have mixed agonist/antagonist properties [222].

***Therapeutic Applications*** The TZDs are used alone or in combination with sulfonylureas or meglitinide analogs as oral antidiabetic agents in the treatment of Type 2 diabetes mellitus [223]. They are ineffective in the absence of insulin and therefore are not used in Type 1 diabetes. The drugs promote weight gain, which is potentially important disadvantage when treating Type 2 diabetes. This is partly due to an increase in the number of adipocytes and is predictable from the known ability of PPARγ activation to promote adipocyte differentiation. Thus, as discussed above, the development of PPARγ antagonists and of SPPARMs is a valid objective and compounds of this type

(SR-202—PPARγ antagonist; MCC555—SPPARM), which are antidiabetic and which prevent weight gain in experimentally obese, hyperglycaemic animals are available.

## Sulfonylureas

Sulfonylureas have been classified as so-called *first-generation* agents (tolbutamide, chlorpropamide, tolazamide, and acetohexamide) and *second-generation* agents (glibenclamide, glipizide, gliclazide, glimepiride, and gliquidone). The second-generation agents are significantly more potent, although there is no clear evidence of superiority of the second-generation drugs in terms of therapeutic effectiveness. The major differences among the sulfonylureas, apart from potency, are pharmacokinetic. Some agents are long-acting (chlorpropamide, glibenclamide) while others have much shorter durations of action (tolbutamide, gliquidone). Sulfonylureas are extensively bound to plasma proteins and are metabolized mainly in the liver to compounds that may be inactive in the case of some agents (e.g., gliquidone) or active in the case of others (e.g., acetohexamide, tolazamide). Tolbutamide, glibenclamide, glipizide, and glimepiride are metabolized by CYP2C9 and genetic polymorphisms of this enzyme may markedly influence their pharmacokinetics [224, 225].

A prolonged controversy resulted from a study suggesting that tolbutamide might produce an increase in cardiovascular mortality [226]. However, the UK Prospective Diabetes Study Group showed no increased cardiovascular mortality with sulfonylureas in a large, randomized, controlled trial over 10 years in which blood glucose was intensively controlled [213]. Indeed, sulfonylureas significantly reduced diabetic microvascular endpoints. It must be pointed out, however, that the UKPDS used different sulfonylureas from the one (tolbutamide) used in the UGDP [226] study.

***Receptors for Sulfonylureas*** Sulfonylureas work primarily by stimulating insulin secretion following binding to a high-affinity sulfonylurea receptor subunit (SUR) of the β-cell potassium-sensitive ATP channel ($K_{ATP}$ channel). Binding results in blocking K efflux through the $K_{IR}6.2$ channel, depolarization of the β cell, opening of voltage-sensitive $Ca^{2+}$ channels, influx of $Ca^{2+}$, and insulin secretion [227]. There are different types of SUR (SUR1, SUR2A, SUR2B), the expression of which differs among different tissues. Thus the β-cell potassium-sensitive $K_{ATP}$ channel contains SUR1, whereas skeletal and cardiac muscle contain SUR2A. Different sulfonylureas have different affinities for the various SURs. For example, tolbutamide and gliclazide are significantly more potent in blocking SUR1 (expressed with $K_{IR}6.2$ in *Xenopus* oocytes) relative to SUR2A, whereas glibenclamide and glimepiride are equipotent at both [228], although other findings [229] showed glibenclamide to bind with much higher affinity to SUR1, relative to SUR2. There has been much debate about the role of extrapancreatic effects of sulfonylureas in their metabolic actions. Extrapancreatic effects probably result, at least in part

indirectly from the improved blood glucose control, because chronic hyperglycemia may itself produce insulin resistance [230]. Nevertheless, in vitro effects have been demonstrated on both skeletal muscle and adipose tissue [231]. Clearly, there are $K_{ATP}$ channels in many peripheral tissues, including skeletal muscle [228]. Moreover, disruption of the SUR2A containing $K_{ATP}$ channels in skeletal-muscle-enhanced glucose uptake [232]. Thus extrapancreatic effects in stimulating glucose uptake by skeletal muscle might relate to block of SUR2A-containing $K_{ATP}$ channels. Another intriguing question is whether or not block of heart, vascular smooth muscle, or cardiac muscle $K_{ATP}$ channels, which contain SUR2, might be manifest as some form of detrimental action on these systems.

***Therapeutic Applications*** Sulfonylureas are widely used in the treatment of Type 2 diabetes. Because they work primarily through stimulating the secretion of insulin, they are of no use in the treatment of established Type 1 diabetes, in which there is extensive autoimmune destruction of the β cells. Patients who are older than 40 years and have had diabetes for less than 5 years are most likely to respond. After 10 years only about 50 percent of patients remain satisfactorily controlled by sulfonylureas. This is referred to as secondary failure and may be attributable to β-cell failure and/or dietary noncompliance [233]. There is evidence that improved control can be achieved by combining insulin with sulfonylureas [234].

Although their pharmacology is broadly similar, individual drugs apparently have unique properties. There is evidence that gliclazide has antiplatelet and antioxidant activity independent of its effects on glycemic control [235]. These effects have been attributed to the azabicyclo ring, which is unique to gliclazide among the sulfonylureas. It remains to be seen in long-term, large-scale, randomized, controlled trials if these effects confer additional benefits on patients with Type 2 diabetes, in terms of a reduction in diabetic complications and/or reduced mortality.

The main adverse effect of sulfonylureas is hypoglycemia. Pharmacokinetic considerations are important in determining which agents are more likely to produce hypoglycemia as an adverse effect. Chlorpropamide and glibenclamide are more likely to produce hypoglycemia than are shorter acting drugs such as tolbutamide, glipizide, gliclazide, and gliquidone, especially when renal insufficiency is present [236].

## Meglitinide Analogs

It is widely accepted that loss of early-phase insulin secretion is an important and early event in the development of Type 2 diabetes [205]. Thus agents restoring or replacing this phase of insulin release have a strong theoretical basis for treating this condition. It was shown that a compound, later named meglitinide, that was closely related to the nonsulfonylurea moiety of glibenclamide augmented insulin secretion through blockade of $K_{ATP}$ channel and

was hypoglycemic [237]. Two meglitinide analogs, repaglinide and nateglinide, are currently used for treating patients with Type 2 diabetes mellitus, mitiglinide being in clinical development. For a variety of reasons, these agents are less likely than sulfonylureas to produce hypoglycemia as adverse effect doses [205, 238]. Nateglinide and repaglinide are rapidly absorbed from the gastrointestinal tract. They are also rapidly eliminated through extensive hepatic metabolism (by CYP2C9 and CYP3A4) [205, 239, 240]. Moreover, they appear to show glucose dependency in stimulating insulin secretion and may, at least for nateglinide, dissociate more quickly from the receptor and allow more rapid $K_{ATP}$ channel reactivation [241]. Because of their short action, they are administered before each meal, allowing a more physiological pattern of insulin secretion, and reducing the problem of hypoglycemia caused by missed meals.

***Mechanism of Action*** The mechanism of action of meglitinide analogs is similar to that of the sulfonylureas. They block the β-cell $K_{ATP}$ channel by combining with its SUR1 (sulfonylurea receptor) subunit. However, there is evidence for differences among sulfonylureas, repaglinide and nateglinide, in their interaction with the receptor. Meglitinide binds with similar affinities to the SUR1, SUR2A, and SUR2B receptors (transiently expressed in COS-7 cells), which are subunits of different $K_{ATP}$ channels [229]. Other studies in electrophysiological experiments showed that mitiglinide was more potent in blocking SUR1 relative to SUR2, although repaglinide was equipotent in blocking both [228]. The short action of nateglinide may be partly related to the rapid reactivation of the β-cell $K_{ATP}$ channel following removal of the drug [241]. Site-directed mutagenesis studies indicate that nateglinide binds to a similar site to glibenclamide but a different site from repaglinide, another meglitinide analog [242]. Nevertheless, rat islets desensitized by prolonged exposure to sulfonylureas were still capable of secreting insulin in response to nateglinide [241].

## α-Glucosidase Inhibitors

α-Glucosidases are intestinal brush border enzymes responsible for the hydrolysis of oligosaccharides such as sucrose to glucose; they have relative selectivity for sucrase. Inhibition of their hydrolysis prevents carbohydrate absorption as disaccharides are poorly absorbed. Three drugs, acarbose, voglibose, and miglitol, which are competitive inhibitors of α-glucosidase are currently in use. As they are reversible inhibitors, carbohydrate absorption is delayed, not prevented [243]. Inhibitors of these enzymes thus lower postprandial glucose concentrations and are adjuncts for the treatment of diabetes mellitus. Thus these agents can smooth the daily blood glucose fluctuations in patients with diabetes mellitus and reduce glycosylated hemoglobin concentrations. When used as an adjunct to insulin and diet, they also lower postprandial blood glucose in Type 1 diabetes and produce a modest reduction in

glycosylated hemoglobin [244]. They are also effective in Type 2 diabetes when used alongside diet alone or when coadministered with sulfonylureas, metformin, or insulin [245]. A clinical trial suggested that acarbose might prevent Type 2 diabetes [246], although this finding has been challenged.

## Glucagon-Like Peptide-1 (GLP-1) and Glucose-Dependent Insulinotropic Polypeptide (GIP)

The GLP-1, secreted by intestinal L cells, and GIP, secreted by intestinal K cells, are known as incretins and are secreted from the small intestine in response to the presence of glucose and other nutrients in the gut. These hormones augment glucose-induced insulin secretion in a glucose-dependent manner through activating adenylyl cyclase, leading to an increase in islet β-cell cyclic-AMP [247–249]. Their importance is clearly illustrated by the marked glucose intolerance and impairment of insulin secretion seen in mice lacking receptors for both hormones [250]. GLP-1 is also a potent inhibitor of glucagon secretion [251]. GLP-1 and GIP stimulate insulin gene transcription, pancreatic islet cell proliferation and β-cell replication, promote differentiation of pancreatic ductal cells into glucagon- and insulin-producing cells through a PDX-1 dependent pathway, and inhibit β-cell apoptosis [247–249]. They also suppress gastric emptying. GLP-1 was found to improve blood glucose in experimental animals and in patients with Type 2 diabetes, leading to the idea that GLP-1 and GIP might form the basis for novel treatments. The use of the native hormones is severely limited by their very short duration of action, as both are rapidly degraded by the enzyme dipeptidyl peptidase IV [DPPIV]. Indeed the N-terminally truncated products of this enzymic degradation [GLP-1(6–36) and GIP(3–42)] are antagonists of the respective hormones. Two approaches have been used to address this problem. First, attempts are being made to find analogs of the hormones that are resistant to DPPIV-mediated degradation. At least one stable analog (liraglutide-NN 2211) of GLP-1 is in clinical trial. Moreover, exenatide, a synthetic form of exendin-4, is undergoing clinical evaluation (see Exendin-4). Analogs of GIP stable to DPPIV are being studied in experimental animals, and several of these have been shown be insulinotropic and to be antihyperglycemic [249]. The second approach is to use inhibitors of DPPIV to prolong the effect of the endogenous incretins. Several orally active DPPIV inhibitors have been developed (e.g., NVPDPP728 and P32/98 NVP-LAF237) and have shown effectiveness in experimental and clinical diabetes [252, 253].

## Exendin-4

Exendin-4 is a 39-amino-acid peptide isolated from the venom of the gila monster lizard *Heloderma suspectum* and is a member of the glucagon superfamily of peptide hormones [254]. At the N-terminus the replacement of His7 and Ala8 by His7 and Gly8 makes exendin-4 resistant to cleavage by the DPP-

IV, which probably explains its long duration of action compared with GLP-1 [255]. It is structurally related to GLP-1 and interacts with the GLP-1 receptor, elevating cyclic-AMP in the islet β cell and augmenting glucose-induced insulin secretion [256]. It also increases insulin sensitivity and is effective in lowering blood glucose in several animal models of diabetes [257] and in humans [258, 259]. Like GLP-1 it increases β-cell replication and neogenesis resulting in increased β-cell mass [260], reduces gastric emptying, and diminishes calorific intake [258].

### Glucagon

Glucagon is a polypeptide hormone secreted by the A cells (α cells) of the pancreatic islets and is an important hormone in preventing a fall in the blood glucose concentration [261]. It is secreted in response to hypoglycemia, and there is strong evidence for a major autonomic component in this response [261]. By activating adenylyl cyclase, glucagon stimulates glycogenolysis and gluconeogenesis in hepatocytes [262] and lipolysis in adipose tissue [263]. Glucagon also relaxes smooth muscle through an increase in cAMP [264].

***Glucagon Receptors*** Glucagon mediates its effects by the glucagon receptor, a member of the G-protein-coupled superfamily of seven-helical transmembrane receptors (GPCRs) [265]. It is coupled through $G_s$ to adenylyl cyclase increasing cAMP, but its ability to increase intracellular $Ca^{2+}$ in hepatocytes probably additionally involves activation of the PLC/inositol phosphate pathway [265].

***Therapeutic Applications*** The main use of glucagon is in the treatment of acute insulin-induced hypoglycemia, where the patient has lost consciousness and cannot take oral glucose. It is ineffective in starvation or where liver glycogen stores are depleted. Because of the role of glucagon in the metabolic disturbances of diabetes mellitus [266], there may be a role for antagonists at the glucagon receptor in treating diabetes and both peptide and nonpeptide antagonists are being developed [267–270]. The nonpeptide agent Bay 27-9955 is orally active and was shown in the human to antagonize glucagon in elevating plasma glucose and increasing glucose production [268]. NNC 25-2504 is another, very potent (picomolar to low nanomolar binding affinity) nonpeptide, stable, orally active antagonist [271].

## 11 ANTIOBESITY AGENTS

Obesity is a major risk factor for Type 2 diabetes mellitus, dyslipidemia, myocardial infarction, stroke, and cancers of the breast and colon. A 5 to 10 percent reduction in body weight is associated with a clinically significant improvement in associated morbidity. However, it is a difficult condition to

manage and sustained weight loss requires not only effective drugs but also concomitant adjustments to lifestyle and diet. Very few drugs are currently available and the anorexigenic drugs fenfluramine and dexfenfluramine, which have a central effect mediated through promoting neuronal release of serotonin and inhibiting its reuptake [272], were withdrawn in 1997 because of association with pulmonary hypertension and cardiac valve abnormalities [273]. The only appetite suppressant currently available is sibutramine, which blocks the neuronal uptake of serotonin and norepinephrine [274].

The concept of distinct feeding and satiety centers, previously regarded, respectively, as the lateral hypothalamic area and the ventromedial hypothalamus, has been superseded by the idea of distinct neuronal populations that express particular anorexigenic or orexigenic peptides and activation or inhibition of which switch feeding on or off according to the nutritional status [275]. Appetite, satiety, and feeding behavior are integrated by a number of neuropeptides in the hypothalamus. The orexigenic peptides include neuropeptide Y, expressed in neurons largely originating in the arcuate nucleus (NPY—probably acting through the Y1 and Y5 receptors), the orexins (primarily orexin-A), melanin-concentrating hormone (MCH), and agouti-related protein (Agrp—an antagonist of the melanocortin MC3 and MC4 receptors and expressed in NPY neurons). Lipid endocannabinoids may also play an important role in maintaining appetite. On the other hand, the hypothalamic anorectic peptides include α-melanocyte stimulating hormone (αMSH—acting through MC3 and MC4 receptors) and cocaine and amphetamine-related transcript (CART) [275, 276]. The hypothalamic neuronal circuits modifying feeding behavior are markedly influenced by a number of nutrient and hormonal signals from the periphery. Glucose is a satiety signal, and there are glucose-sensitive neurons in the hypothalamus, which are inhibited by NPY and activated by orexins. Gold-thioglucose-induced damage to these neurons results in hyperphagia and obesity [275]. Insulin acting as a satiety signal probably modulates these neurons. Cholecystokinin (CCK), pancreatic polypeptide, and PYY 3-36 and the preproglucagon-derived peptides glucagon-like peptide 1 (GLP-1) and oxyntomodulin, produced in the small intestine, are also anorectic [275, 276]. On the other hand, ghrelin from the stomach is orexigenic. In broad terms, the orexigenic peptides activate the NPY/Agouti-related peptide neuronal systems and inhibit the anorexogenic neurons, whereas the anorexigenic peptides do the opposite [275, 276] (Fig. 4). All of these peptides and their receptors are potential targets for the development of antiobesity agents. Many agents are under development, although none has yet reached the market.

The discovery that adipose tissue is an active endocrine tissue, secreting hormones (e.g., resistin) that influence the sensitivity to insulin as well as food intake (e.g., leptin, which acts as an anorexigenic signal, stimulating melanocortin neurons) and that this tissue is a major target for the thiazolidinedione-agonists at the PPARγ (peroxisome proliferator-γ) activated receptors [277] has revealed more potential targets. Mitochondrial uncoupling

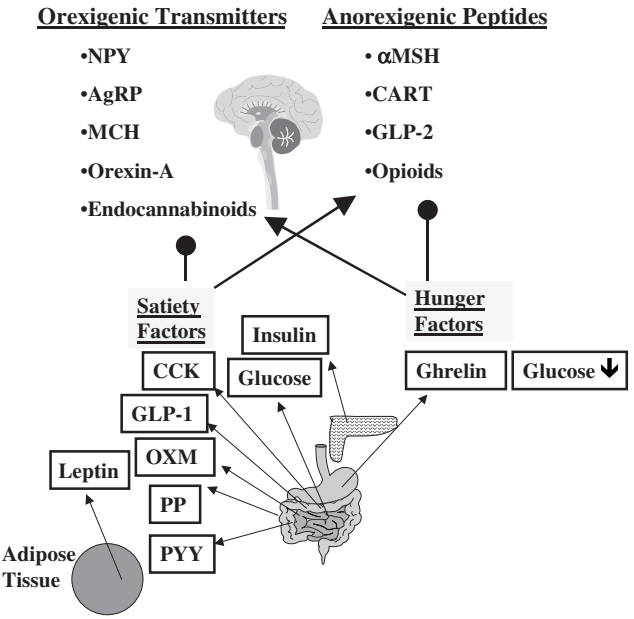


**Figure 4** Regulation of appetite and feeding behavior [275–277]. Distinct neuronal populations express particular anorexigenic or orexigenic neurotransmitter peptides (or lipids). The activity of these neurons is influenced by signals from the periphery, especially the gut. Signals activating feeding increase the activity of orexigenic circuits and concomitantly inhibit anorexigenic neuronal activity, and vice versa for satiety signals. NPY, neuropeptide Y; AgRP, agouti-related protein; MCH, melanin-concentrating hormone; αMSH, α-melanocyte-stimulating hormone; CART, cocaine and amphetamine-regulated transcript; GLP-2, glucagons-like peptide-2; CCK, cholecystokinin; GLP-1, glucagons-like peptide-1; OXM, oxyntomodulin.

proteins in muscle and adipose tissue are also potential targets for new antiobesity agents [278] including agonists at $\beta_3$-adrenoceptors [279].

## Orlistat

Orlistat is an irreversible and highly selective inhibitor of pancreatic and gastric lipase and therefore inhibits the hydrolysis of triglyceride (triacylglycerol) in the gut lumen. It binds covalently to the serine residue on the active site [280]. It shows around 1000-fold selectivity for lipase compared with amylase, trypsin, or chymotrypsin. As triacylglycerol hydrolysis is a prerequisite for intestinal fat absorption, lipase inhibition leads to reduced fat absorption and therefore to decreased availability of fat within the body. This is the basis for the widespread use of orlistat in the treatment of obesity [280, 281]. Orlistat is highly selective for lipase and has no significant effect in inhibiting

amylase, trypsin, or chymotrypsin [280]. Its very poor gastrointestinal absorption (around 1 percent) results in very low plasma concentrations and therefore absence of systemic adverse effects [280].

***Therapeutic Applications*** Orlistat is used for the treatment of obesity in patients who have demonstrated their ability to lose weight as a result of dietary control and increased physical activity. Weight loss after 12 months averages 8 to 10 percent when used with an appropriate hypocaloric diet.

Plasma levels of fat-soluble vitamins are reduced [281], but the potential clinical consequences of this has not been evident. Studies in volunteers have shown a reduced plasma cholecystokinin response to meals, which may suggest an increased risk of gallstones, especially in obese patients who are already at greater risk [281]. Gastrointestinal effects (flatus, oily spotting, fecal urgency, steatorrhea) are quite common and are predictable for a drug that prevents fat absorption.

***Acknowledgment*** I am very grateful to Pat Owen for her editorial assistance.

## REFERENCES

1. Levy, A. (2002). Physiological implications of pituitary trophic activity. *J. Endocrinol.*, *174*, 147–155.
2. Marshall, J. C., Eagleson, C. A., McCartney, C. R. (2001). Hypothalamic dysfunction. *Mol. Cell. Endocrinol.*, *183*, 29–32.
3. Parker, K. L., Schimmer, B. P. (2001). Pituitary hormones and their hypothalamic releasing factors. In J. G. Hardman, L. E. Limbird, A. Goodman Gilman (Eds), *Goodman & Gilman's the Pharmacological Basis of Therapeutics*, 10th ed. McGraw-Hill, New York, pp. 1541–1562.
4. Persani, L. (1998). Hypothalamic thyrotropin-releasing hormone and thyrotropin biological activity. *Thyroid*, *8*, 941–946.
5. Nillni, E. A., Sevarino, K. A. (1999). The biology of pro-thyrotropin-releasing hormone-derived peptides. *Endocr. Rev.*, *20*, 599–648.
6. Sun, Y., Lu, X., Gershengorn, M. C. (2003). Thyrotropin-releasing hormone receptors—Similarities and differences. *J. Mol. Endocrinol.*, *30*, 87–97.
7. Nishino, S., Arrigoni, J., Shelton, J., Kanbayashi, T., Dement, W. C., Mignot, E. (1997). Effects of thyrotropin-releasing hormone and its analogs on daytime sleepiness and cataplexy in canine narcolepsy. *J. Neurosci.*, *17*, 6401–6408.
8. Urayama, A., Yamada, S., Kimura, R., Zhang, J., Watanabe, Y. (2002). Neuroprotective effect and brain receptor binding of taltirelin, a novel thyrotropin-releasing hormone (TRH) analogue, in transient forebrain ischemia of C57BL/6J mice. *Life Sci.*, *72*, 601–607.
9. Vale, W., Spiess, J., Rivier, C., Rivier, J. (1981). Characterization of a 41-residue ovine hypothalamic peptide that stimulates secretion of corticotropin and beta-endorphin. *Science*, *213*, 1394–1397.

10. Dautzenberg, F. M., Hauger, R. L. (2002). The CRF peptide family and their receptors: Yet more partners discovered. *Trends Pharmacol. Sci.*, *23*, 71–77.
11. Latchman, D. S. (2002). Urocortin. *Int. J. Biochem. Cell. Biol.*, *34*, 907–910.
12. Smagin, G. N., Heinrichs, S. C., Dunn, A. J. (2001). The role of CRH in behavioral responses to stress. *Peptides*, *22*, 713–724.
13. Parkes, D. G., Weisinger, R. S., May, C. N. (2001). Cardiovascular actions of CRH and urocortin: An update. *Peptides*, *22*, 821–827.
14. Agelaki, S., Tsatsanis, C., Gravanis, A., Margioris, A. N. (2002). Corticotropin-releasing hormone augments pro-inflammatory cytokine production from macrophages in vitro and in lipopolysaccharide-induced endotoxin shock in mice. *Infect. Immun.*, *70*, 6068–6074.
15. Singh, L. K., Boucher, W., Pang, X., Letourneau, R., Seretakis, D., Green, M., Theoharides, T. C. (1999). Potent mast cell degranulation and vascular permeability triggered by urocortin through activation of corticotropin-releasing hormone receptors. *J. Pharmacol. Exp. Ther.*, *288*, 1349–1356.
16. Aggelidou, E., Hillhouse, E. W., Grammatopoulos, D. K. (2002). Up-regulation of nitric oxide synthase and modulation of the guanylate cyclase activity by corticotropin-releasing hormone but not urocortin II or urocortin III in cultured human pregnant myometrial cells. *Proc. Natl. Acad. Sci. USA*, *99*, 3300–3305.
17. Newell-Price, J., Trainer, P., Besser, M., Grossman, A. (1998). The diagnosis and differential diagnosis of Cushing's syndrome and pseudo-Cushing's states. *Endocr. Rev.*, *19*, 647–672.
18. Nicholson, W. E., DeCherney, G. S., Jackson, R. V., DeBold, C. R., Uderman, H., Alexander, A. N., Rivier, J., Vale, W., Orth, D. N. (1983). Plasma distribution, disappearance half-time, metabolic clearance rate, and degradation of synthetic ovine corticotropin-releasing factor in man. *J. Clin. Endocrinol. Metab.*, *57*, 1263–1269.
19. Schulte, H. M., Chrousos, G. P., Booth, J. D., Oldfield, E. H., Gold, P. W., Cutler, G. B. Jr, Loriaux, D. L. (1984). Corticotropin-releasing factor: Pharmacokinetics in man. *J. Clin. Endocrinol. Metab.*, *58*, 192–196.
20. Webster, E. L., Lewis, D. B., Torpy, D. J., Zachman, E. K., Rice, K. C., Chrousos, G. P. (1996). In vivo and in vitro characterization of antalarmin, a nonpeptide corticotropin-releasing hormone (CRH) receptor antagonist: Suppression of pituitary ACTH release and peripheral inflammation. *Endocrinology*, *137*, 5747–5750.
21. Ducottet, C., Griebel, G., Belzung, C. (2003). Effects of the selective nonpeptide corticotropin-releasing factor receptor 1 antagonist antalarmin in the chronic mild stress model of depression in mice. *Prog. Neuropsychopharmacol. Biol. Psychiatry*, *27*, 625–631.
22. Bornstein, S. R., Webster, E. L., Torpy, D. J., Richman, S. J., Mitsiades, N., Igel, M., Lewis, D. B., Rice, K. C., Joost, H. G., Tsokos, M., Chrousos, G. P. (1998). Chronic effects of a nonpeptide corticotropin-releasing hormone type I receptor antagonist on pituitary-adrenal function, body weight, and metabolic regulation. *Endocrinology*, *139*, 1546–1555.
23. Gabry, K. E., Chrousos, G. P., Rice, K. C., Mostafa, R. M., Sternberg, E., Negrao, A. B., Webster, E. L., McCann, S. M., Gold, P. W. (2002). Marked suppression of

gastric ulcerogenesis and intestinal responses to stress by a novel class of drugs. *Mol. Psychiatry*, *7*, 474–483.

24. Webster, E. L., Barrientos, R. M., Contoreggi, C., Isaac, M. G., Ligier, S., Gabry, K. E., Chrousos, G. P., McCarthy, E. F., Rice, K. C., Gol, P. W., Sternberg, E. M. (2002). Corticotropin releasing hormone (CRH) antagonist attenuates adjuvant induced arthritis: Role of CRH in peripheral inflammation. *J. Rheumatol.*, *29*, 1252–1261.
25. Zobel, A. W., Nickel, T., Kunzel, H. E., Ackl, N., Sonntag, A., Ising, M., Holsboer, F. (2000). Effects of the high-affinity corticotropin-releasing hormone receptor 1 antagonist R121919 in major depression: The first 20 patients treated. *J. Psychiatr. Res.*, *34*, 171–181.
26. Seymour, P. A., Schmidt, A. W., Schulz, D. W. (2003). The pharmacology of CP-154,526, a non-peptide antagonist of the CRH1 receptor: A review. *CNS Drug Rev.*, *9*, 57–96.
27. Mackay, K. B., Stiefel, T. H., Ling, N., Foster, A. C. (2003). Effects of a selective agonist and antagonist of CRF2 receptors on cardiovascular function in the rat. *Eur. J. Pharmacol.*, *469*, 111–115.
28. Mayo, K. E., Miller, T., DeAlmeida, V., Godfrey, P., Zheng, J., Cunha, S. R. (2000). Regulation of the pituitary somatotroph cell by GHRH and its receptor. *Recent Prog. Horm. Res.*, *55*, 237–266.
29. Camanni, F., Ghigo, E., Arvat, E. (1998). Growth hormone-releasing peptides and their analogs. *Front. Neuroendocrinol.*, *19*, 47–72.
30. Bremner, W. J., Huhtaniemi, I., Amory, J. K. (2001). Pituitary gonadotrophins and their disorders. In K. L. Becker (Ed.), *Principles and Practice of Endocrinology and Metabolism*, Vol. 3. Lippincott Williams & Wilkins, Philadelphia, pp. 170–177.
31. Barradell, L. B., McTavish, D. (1993). Histrelin. A review of its pharmacological properties and therapeutic role in central precocious puberty. *Drugs*, *45*, 570–588.
32. Grosse, R., Schmid, A., Schoneberg, T., Herrlich, A., Muhn, P., Schultz, G., Gudermann, T. (2000). Gonadotropin-releasing hormone receptor initiates multiple signaling pathways by exclusively coupling to G(q/11) proteins. *J. Biol. Chem.*, *275*, 9193–9200.
33. Ortmann, O., Diedrich, K. (1999). Pituitary and extrapituitary actions of gonadotrophin-releasing hormone and its analogues. *Hum. Reprod.*, *14*(Suppl. 1), 194–206.
34. Surrey, E. S., Schoolcraft, W. B. (2000). Evaluating strategies for improving ovarian response of the poor responder undergoing assisted reproductive techniques. *Fertil. Steril.*, *73*, 667–676.
35. Yano, T., Yano, N., Matsumi, H., Morita, Y., Tsutsumi, O., Schally, A. V., Taketani, Y. (1997). Effect of luteinizing hormone-releasing hormone analogs on the rat ovarian follicle development. *Horm. Res.*, *48*(Suppl. 3), 35–41.
36. McArdle, C. A., Franklin, J., Green, L., Hislop, J. N. (2002). Signalling, cycling and desensitisation of gonadotrophin-releasing hormone receptors. *J. Endocrinol.*, *173*, 1–11.
37. Ravenna, L., Salvatori, L., Morrone, S., Lubrano, C., Cardillo, M. R., Sciarra, F., Frati, L., Di Silverio, F., Petrangeli, E. (2000). Effects of triptorelin, a

gonadotropin-releasing hormone agonist, on the human prostatic cell lines PC3 and LNCaP. *J. Androl.*, *21*, 549–557.

38. Herbst, K. L. (2003). Gonadotropin-releasing hormone antagonists. *Curr. Opin. Pharmacol.*, *3*, 660–666.
39. Brogden, R. N., Buckley, M. M., Ward, A. (1990). Buserelin. A review of its pharmacodynamic and pharmacokinetic properties, and clinical profile. *Drugs*, *39*, 399–437.
40. Gunnet, J. W., Demarest, K. T., Hahn, D. W., Ericson, E., McGuire, J. L. (1991). Effects of a sustained release formulation of the gonadotrophin-releasing hormone agonist histrelin on serum concentrations of gonadotrophins and oestradiol, and ovarian LH/human chorionic gonadotrophin receptor content in the rat. *J. Endocrinol.*, *131*, 211–218.
41. Johnson, C. A., Thompson, D. L. Jr., Cartmill, J. A. (2002). Pituitary responsiveness to GnRH in mares following deslorelin acetate implantation to hasten ovulation. *J. Anim. Sci.*, *80*, 2681–2687.
42. Dawood, M. Y. (1994). Hormonal therapies for endometriosis: Implications for bone metabolism. *Acta. Obstet. Gynecol. Scand. Suppl.*, *159*, 22–34.
43. Cook, T., Sheridan, W. P. (2000). Development of GnRH antagonists for prostate cancer: New approaches to treatment. *Oncologist*, *5*, 162–168.
44. Lee, C. H., Van Antwerp, D., Hedley, L., Nestor, J. J. Jr., Vickery, B. H. (1989). Comparative studies on the hypotensive effect of LHRH antagonists in anesthetized rats. *Life Sci.*, *45*, 697–702.
45. Rabinovici, J., Rothman, P., Monroe, S. E., Nerenberg, C., Affe, R. B. (1992). Endocrine effects and pharmacokinetic characteristics of a potent new gonadotropin-releasing hormone antagonist (Ganirelix) with minimal histamine-releasing properties: Studies in postmenopausal women. *J. Clin. Endocrinol. Metab.*, *75*, 1220–1225.
46. Patel, Y. C. (1999). Somatostatin and its receptor family. *Front. Neuroendocrinol.*, *20*, 157–198.
47. Shimon, I., Yan, X., Taylor, J. E., Weiss, M. H., Culler, M. D., Melmed, S. (1997). Somatostatin receptor (SSTR) subtype-selective analogues differentially suppress in vitro growth hormone and prolactin in human pituitary adenomas. Novel potential therapy for functional pituitary tumors. *J. Clin. Invest.*, *100*, 2386–2392.
48. Feniuk, W., Jarvie, E., Luo, J., Humphrey, P. P. (2000). Selective somatostatin sst(2) receptor blockade with the novel cyclic octapeptide, CYN-154806. *Neuropharmacology*, *39*, 1443–1450.
49. Reubi, J. C., Schaer, J. C., Wenger, S., Hoeger, C., Erchegyi, J., Waser, B., Rivier, J. (2000). SST3-selective potent peptidic somatostatin receptor antagonists. *Proc. Natl. Acad. Sci. USA*, *97*, 13973–13978.
50. Hoyer, D., Nunn, C., Hannon, J., Schoeffter, P., Feuerbach, D., Schuepbach, E., Langenegger, D., Bouhelal, R., Hurth, K., Neumann, P., Troxler, T., Pfaeffli, P. (2004). SRA880, in vitro characterization of the first non-peptide somatostatin sst(1) receptor antagonist. *Neurosci. Lett.*, *361*, 132–135.
51. Poitout, L., Roubert, P., Contour-Galcera, M. O., Moinet, C., Lannoy, J., Pommier, J., Plas, P., Bigg, D., Thurieau, C. (2001). Identification of potent non-peptide somatostatin antagonists with sst(3) selectivity. *J. Med. Chem.*, *44*, 2990–3000.

52. Stewart, P. M., James, R. A. (1999). The future of somatostatin analogue therapy. *Baillieres Best Pract. Res. Clin. Endocrinol. Metab.*, *13*, 409–418.
53. Eriksson, B., Oberg, K. (1999). Summing up 15 years of somatostatin analog therapy in neuroendocrine tumors: Future outlook. *Ann. Oncol.*, *10*(Suppl. 2), S31–S38.
54. Bernard, B., Capello, A., van Hagen, M., Breeman, W., Srinivasan, A., Schmidt, M., Erion, J., van Gameren, A., Krenning, E., de Jong, M. (2004). Radiolabeled RGD-DTPA-Tyr3-octreotate for receptor-targeted radionuclide therapy. *Cancer Biother. Radiopharm.*, *19*, 173–180.
55. Herrington, J., Carter-Su, C. (2001). Signaling pathways activated by the growth hormone receptor. *Trends Endocrinol. Metab.*, *12*, 252–257.
56. LeRoith, D., Bondy, C., Yakar, S., Liu, J. L., Butler, A. (2001). The somatomedin hypothesis: 2001. *Endocr. Rev.*, *22*, 53–74.
57. Baxter, R. C. (2000). Insulin-like growth factor (IGF)-binding proteins: Interactions with IGFs and intrinsic bioactivities. *Am. J. Physiol. Endocrinol. Metab.*, *278*, E967–E976.
58. O'Dell, S. D., Day, I. N. (1998). Insulin-like growth factor II (IGF-II). *Int. J. Biochem. Cell. Biol.*, *30*, 767–771.
59. Thorner, M. O., Strasburger, C. J., Wu, Z., Straume, M., Bidlingmaier, M., Pezzoli, S. S., Zib, K., Scarlett, J. C., Bennett, W. F. (1999). Growth hormone (GH) receptor blockade with a PEG-modified GH (B2036-PEG) lowers serum insulin-like growth factor-1 but does not acutely stimulate serum GH. *J. Clin. Endocrinol. Metab.*, *84*, 2098–2103.
60. Van der Lely, A. J., Hutson, R. K., Trainer, P. J., Besser, G. M., Barkan, A. L., Katznelson, L., Klibanski, A., Herman-Bonert, V., Melmed, S., Vance, M. L., Freda, P. U., Stewart, P. M., Friend, K. E., Clemmons, D. R., Johannsson, G., Stavrou, S., Cook, D. M., Phillips, L. S., Strasburger, C. J., Hackett, S., Zib, K. A., Davis, R. J., Scarlett, J. C., Thorner, M. O. (2001). Long-term treatment of acromegaly with pegvisomant, a growth hormone receptor antagonist. *Lancet*, *358*, 1745–1749.
61. Parkinson, C., Trainer, P. J. (2001). The place of pegvisomant in the management of acromegaly. *Expert Opin. Investig. Drugs*, *10*, 1725–1735.
62. Simoni, M., Gromoll, J., Nieschlag, E. (1997). The follicle-stimulating hormone receptor: Biochemistry, molecular biology, physiology, and pathophysiology. *Endocr. Rev.*, *6*, 739–773.
63. Arey, B. J., Deecher, D. C., Shen, E. S., Stevis, P. E., Meade, E. H. Jr, Wrobel, J., Frail, D. E., Lopez, F. J. (2002). Identification and characterization of a selective, nonpeptide follicle-stimulating hormone receptor antagonist. *Endocrinology*, *143*, 3822–3829.
64. Daya, S. (2002). Updated meta-analysis of recombinant follicle-stimulating hormone (FSH) versus urinary FSH for ovarian stimulation in assisted reproduction. *Fertil. Steril.*, *77*, 711–714.
65. Lenton, E., Soltan, A., Hewitt, J., Thomson, A., Davies, W., Ashraf, N., Sharma, V., Jenner, L., Ledger, W., McVeigh, E. (2000). Induction of ovulation in women undergoing assisted reproductive techniques: Recombinant human FSH (follitropin alpha) versus highly purified urinary FSH (urofollitropin HP). *Hum. Reprod.*, *15*, 1021–1027.

66. Bouloux, P., Warne, D. W., Loumaye, E. (2002). Efficacy and safety of recombinant human follicle-stimulating hormone in men with isolated hypogonadotropic hypogonadism. *Fertil. Steril.*, *77*, 270–273.
67. Ascoli, M., Fanelli, F., Segaloff, D. L. (2002). The lutropin/choriogonadotropin receptor, a 2002 perspective. *Endocr. Rev.*, *23*, 141–174.
68. Abboud, F. M., Floras, J. S., Aylward, P. E., Guo, G. B., Gupta, B. N., Schmid, P. G. (1990). Role of vasopressin in cardiovascular and blood pressure regulation. *Blood Vessels*, *27*, 106–115.
69. Barberis, C., Tribollet, E. (1996). Vasopressin and oxytocin receptors in the central nervous system. *Crit. Rev. Neurobiol.*, *10*, 119–154.
70. Rose, J. D., Moore, F. L. (2002). Behavioral neuroendocrinology of vasotocin and vasopressin and the sensorimotor processing hypothesis. *Front. Neuroendocrinol.*, *23*, 317–341.
71. Alescio-Lautier, B., Paban, V., Soumireu-Mourat, B. (2000). Neuromodulation of memory in the hippocampus by vasopressin. *Eur. J. Pharmacol.*, *405*, 63–72.
72. Birnbaumer, M. (2000). Vasopressin receptors. *Trends Endocrinol. Metab.*, *11*, 406–410.
73. Knepper, M. A. (1998). Long-term regulation of urinary concentrating capacity. *Am. J. Physiol.*, *275*, F332–F333.
74. Schwartz, J., Derdowska, I., Sobocinska, M., Kupryszewski, G. (1991). A potent new synthetic analog of vasopressin with relative agonist specificity for the pituitary. *Endocrinology*, *129*, 1107–1109.
75. Derick, S., Cheng, L. L., Voirol, J., Stoev, S., Giacomini, M., Wo, N. C., Szeto, H. H., Ben Mimoun, M., Andres, M., Gaillard, R. C., Guillon, G., Manning, M. (2002). [1-Deamino-4-cyclohexylalanine]arginine vasopressin: A potent and specific agonist for vasopressin $V_{1b}$ receptors. *Endocrinology*, *143*, 4655–4664.
76. Andres, M., Trueba, M., Guillon, G. (2002). Pharmacological characterization of F-180: A selective human V(1a) vasopressin receptor agonist of high affinity. *Br. J. Pharmacol.*, *135*, 1828–1836.
77. Durr, J. A., Hensen, J., Ehnis, T., Blankenship, M. S. (2000). Chlorpropamide upregulates antidiuretic hormone receptors and unmasks constitutive receptor signaling. *Am. J. Physiol. Renal Physiol.*, *278*, F799–F808.
78. Griebe, G., Simiand, J., Stemmelin, J., Gal, C. S., Steinberg, R. (2003). The vasopressin V1b receptor as a therapeutic target in stress-related disorders. *Curr. Drug Targets CNS Neurol. Disord.*, *2*, 191–200.
79. Serradeil-Le Gal, C., Wagnon, J., Valette, G., Garcia, G., Pasca, M., Maffrand, J. P., Le Fur, G. (2002). Nonpeptide vasopressin receptor antagonists: Development of selective and orally active V1a, V2 and V1b receptor ligands. *Prog. Brain Res.*, *139*, 197–210.
80. Mannucci, P. M. (2001). How I treat patients with von Willebrand disease. *Blood*, *97*, 1915–1919.
81. Kaufmann, J. E., Vischer, U. M. (2003). Cellular mechanisms of the hemostatic effects of desmopressin (DDAVP). *J. Thromb. Haemost.*, *1*, 682–689.
82. Thibonnier, M. (2003). Vasopressin receptor antagonists in heart failure. *Curr. Opin. Pharmacol.*, *3*, 683–687.

83. Andersson, K. E. (2001). Oxytocin-containing neurons in the PVN project to spinal autonomic nuclei. *Pharmacol. Rev.*, *53*, 417–450.
84. Filippi, S., Vignozzi, L., Vannelli, G. B., Ledda, F., Forti, G., Maggi, M. J. (2003). Role of oxytocin in the ejaculatory process. *Endocrinol. Invest.*, *26*(3, Suppl.), 82–86.
85. Gimpl, G., Fahrenholz, F. (2001). The oxytocin receptor system: Structure, function, and regulation. *Physiol. Rev.*, *81*, 629–683.
86. Zingg, H. H., Laporte, S. A. (2003). The oxytocin receptor. *Trends Endocrinol. Metab.*, *14*, 222–227.
87. Lowbridge, J., Manning, M., Haldar, J, Sawyer, W. H. (1977). Synthesis and some pharmacological properties of [4-threonine, 7-glycine]oxytocin, [1-(L-2-hydroxy-3-mercaptopropanoic acid), 4-threonine, 7-glycine]oxytocin (hydroxy[Thr4, Gly7]oxytocin), and [7-glycine]oxytocin, peptides with high oxytocic-antidiuretic selectivity. *J. Med. Chem.*, *20*, 120–123.
88. Lamont, R. F. (2003). The development and introduction of anti-oxytocic tocolytics. *Br. J. Obstet. Gynaecol.*, *110*, 108–112.
89. World-wide Atosiban Versus Beta-agonists Study Group (2001). Effectiveness and safety of the oxytocin antagonist atosiban versus beta-adrenergic agents in the treatment of preterm labor. *Br. J. Obstet. Gynaecol.*, *108*, 133–142.
90. Goodwin, T. M., Valenzuela, G. J., Silver, H., Creasy, G., and the atosiban study group (1996). Dose ranging study of the oxytocin antagonist atosiban in the treatment of preterm labor. *Obstet. Gynecol.*, *88*, 331–336.
91. Fuchs, A. R., Fuchs, F., Husslein, P., Soloff, M. S. (1984). Oxytocin receptors in the human uterus during pregnancy and parturition. *Am. J. Obstet. Gynecol.*, *150*, 734–741.
92. Wolff, J. (1998). Perchlorate and the thyroid gland. *Pharmacol. Rev.*, *50*, 89–105.
93. Hernandez, A., St Germain, D. L. (1997). Selenodeiodinases and their role in thyroid hormone activation and inactivation. *Curr. Opin. Endocrinol. Diabetes*, *4*, 333–340.
94. Brent, G. A. (1994). The molecular basis of thyroid hormone action. *N. Engl. J. Med.*, *331*, 847–853.
95. Yen, P. M. (2001). Physiological and molecular basis of thyroid hormone action. *Physiol. Rev.*, *81*, 1097–1142.
96. Barbe, P., Larrouy, D., Boulanger, C., Chevillotte, E., Viguerie, N., Thalamas, C., Trastoy, M. O., Roques, M., Vidal, H., Langin, D. (2001). Triiodothyronine-mediated up-regulation of UCP2 and UCP3 mRNA expression in human skeletal muscle without coordinated induction of mitochondrial respiratory chain genes. *FASEB J.*, *15*, 13–15.
97. Davis, P. J., Tillmann, H. C., Davis, F. B., Wehling, M. (2002). Comparison of the mechanisms of nongenomic actions of thyroid hormone and steroid hormones. *J. Endocrinol. Invest.*, *25*, 377–388.
98. Huang, C. J., Geller, H. M., Green, W. L., Craelius, W. (1999). Acute effects of thyroid hormone analogs on sodium currents in neonatal rat myocytes. *J. Mol. Cell. Cardiol.*, *31*, 881–893.
99. Wrutniak-Cabello, C., Casas, F., Cabello, G. (2001). Thyroid hormone action in mitochondria. *J. Mol. Endocrinol.*, *26*, 67–77.

100. Taylor, A. H., Stephan, Z. F., Steele, R. E., Wong, N. C. (1997). Beneficial effects of a novel thyromimetic on lipoprotein metabolism. *Mol. Pharmacol.*, *52*, 542–547.
101. Spooner, P. H., Morkin, E., Goldman, S. (1999). Thyroid hormones and thyroid hormone analogues in the treatment of heart failure. *Coron. Artery Dis.*, *10*, 395–399.
102. Adamson, C. R., Maitra, N., Bahl, J. J., Greer, K., Klewar, S., Hoying, J., Morkin, E. J. (2004). Regulation of gene expression in cardiomyocytes by thyroid hormone and thyroid hormone analogs, 3,5-diiodothyropropionic acid and CGS 23425. *Pharmacol. Exp. Ther.*, *311*, 164–171.
103. Stanbury, J. B., Ermans, A. E., Bourdoux, P., Todd, C., Oken, E., Tonglet, R., Vidor, G., Braverman, L. E., Medeiros-Neto, G. (1998). Iodine-induced hyperthyroidism: Occurrence and epidemiology. *Thyroid*, *8*, 83–100.
104. Eng, P. H., Cardona, G. R., Fang, S. L., Previti, M., Alex, S., Carrasco, N., Chin, W. W., Braverman, L. E. (1999). Escape from the acute Wolff-Chaikoff effect is associated with a decrease in thyroid sodium/iodide symporter messenger ribonucleic acid and protein. *Endocrinology*, *140*, 3404–3410.
105. Markou, K., Georgopoulos, N., Kyriazopoulou, V., Vagenakis, A. G. (2001). Iodine-induced hypothyroidism. *Thyroid*, *11*, 501–510.
106. Burikhanov, R. B., Matsuzaki, S. (2000). Excess iodine induces apoptosis in the thyroid of goitrogen-pretreated rats in vivo. *Thyroid*, *10*, 123–129.
107. Langer, R., Burzler, C., Bechtner, G., Gartner, R. (2003). Influence of iodide and iodolactones on thyroid apoptosis. Evidence that apoptosis induced by iodide is mediated by iodolactones in intact porcine thyroid follicles. *Exp. Clin. Endocrinol. Diabetes*, *111*, 325–329.
108. Kaplan, E. L. (2001). Surgery of the thyroid gland. In K. Becker (Ed.), *Principles and Practice of Endocrinology and Metabolism*, 3rd ed. Lippincott, Williams & Wilkins, Philadelphia, pp. 440–445.
109. Verger, P., Aurengo, A., Geoffroy, B., Le Guen, B. (2001). Iodine kinetics and effectiveness of stable iodine prophylaxis after intake of radioactive iodine: A review. *Thyroid*, *11*, 353–360.
110. Taurog, A. (1976). The mechanism of action of the thioureylene antithyroid drugs. *Endocrinology*, *98*, 1031–1046.
111. Peakman, M., Hussain, M., Cundy, T., Vergani, D. (1989). Increased activated T-lymphocytes and normal thyrotropin receptor antibody levels in Graves' disease in long-term remission. *J. Clin. Lab. Immunol.*, *30*, 1–5.
112. Nakashima, T., Taurog, A. (1979). Rapid conversion of carbimazole to methimazole in serum; evidence for an enzymatic mechanism. *Clin. Endocrinol.*, *10*, 637–648.
113. Heinen, E., Herrmann, J., Mosny, D., Moreno, F., Teschke, R., Kruskemper, H. L. (1981). Inhibition of peripheral deiodination of 3, 5, 3′-triiodothyronine: An adverse effect of propylthiouracil in the treatment of T3-thyrotoxicosis. *J. Endocrinol. Invest.*, *4*, 331–334.
114. Zhang, L., Parratt, J. R., Beastall, G. H., Pyne, N. J., Furman, B. L. (2002). Streptozotocin diabetes protects against arrhythmias in rat isolated hearts: Role of hypothyroidism. *Eur. J. Pharmacol.*, *435*, 269–276.
115. Lucas, A., Salinas, I., Rius, F., Pizarro, E., Granada, M. L., Foz, M., Sanmarti, A. (1997). Medical therapy of Graves' disease: Does thyroxine prevent recurrence of hyperthyroidism? *J. Clin. Endocrinol. Metab.*, *82*, 2410–2413.

116. Bartalena, L., Bogazzi, F., Martino, E. (1996). Adverse effects of thyroid hormone preparations and antithyroid drugs. *Drug Saf.*, *15*, 53–63.
117. Meyer-Gessner, M., Benker, G., Lederbogen, S., Olbricht, T., Reinwein, D. (1994). Antithyroid drug-induced agranulocytosis: Clinical experience with ten patients treated at one institution and review of the literature. *J. Endocrinol. Invest.*, *17*, 29–36.
118. Woeber, K. A. (2002). Methimazole-induced hepatotoxicity. *Endocr. Pract.*, *8*, 222–224.
119. Liaw, Y. F., Huang, M. J., Fan, K. D., Li, K. L., Wu, S. S., Chen,T. J. (1993). Hepatic injury during propylthiouracil therapy in patients with hyperthyroidism. A cohort study. *Ann. Intern. Med.*, *118*, 424–428.
120. Schneider, D. L. (2000). New therapies for the prevention and treatment of osteoporosis. *Curr. Opin. Endocrinol. Diabetes*, *7*, 310–319.
121. Canalis, E. (2000). Glucocorticoid-induced osteoporosis. *Curr. Opin. Endocrinol. Diabetes*, *7*, 320–324.
122. Siris, E. S. (1999). Goals of treatment for Paget's disease of bone. *J. Bone. Miner. Res.*, *14*, 49–52.
123. Body, J. J. (2000). Current and future directions in medical therapy: Hypercalcemia. *Cancer*, *88*(12, Suppl.), 3054–3058.
124. Cohen, A., Silverberg, S. J. (2002). Calcimimetics: Therapeutic potential in hyperparathyroidism. *Curr. Opin. Pharmacol.*, *2*, 734–739.
125. Zaidi, M., Blair, H. C., Moonga, B. S., Abe, E., Huang, C. L. (2003). Osteoclastogenesis, bone resorption, and osteoclast-based therapeutics. *J. Bone Miner. Res.*, *18*, 599–609.
126. Huang, J. C., Sakata, T., Pfleger, L. L., Bencsik, M., Halloran, B. P., Bikle, D. D., Nissenson, R. A. (2004). PTH differentially regulates expression of RANKL and OPG. *J. Bone Miner. Res.*, *19*, 235–244.
127. Gardella, T. J., Juppner, H. (2001). Molecular properties of the PTH/PTHrP receptor. *Trends Endocrinol. Metab.*, *12*, 210–217.
128. Mannstadt, M., Juppner, H., Gardella, T. J. (1999). Receptors for PTH and PTHrP: Their biological importance and functional properties. *Am. J. Physiol.*, *277*(5, Pt. 2), F665–F675.
129. Hock, J. M., Krishnan, V., Onyia, J. E., Bidwell, J. P., Milas, J., Stanislaus, D. (2001). Osteoblast apoptosis and bone turnover. *J. Bone Miner. Res.*, *16*, 975–984.
130. Fox, J. (2002). Developments in parathyroid hormone and related peptides as bone-formation agents. *Curr. Opin. Pharmacol.*, *2*, 338–344.
131. Zaidi, M., Inzerillo, A. M., Moonga, B. S., Bevis, P. J., Huang, C. L. (2002). Forty years of calcitonin—Where are we now? A tribute to the work of Iain Macintyre, FRS. *Bone*, *30*, 655–663.
132. Fleisch, H. (2002). Development of bisphosphonates. *Breast Cancer Res.*, *4*, 30–34.
133. Rogers, M. J., Gordon, S., Benford, H. L., Coxon, F. P., Luckman, S. P., Monkkonen, J., Frith, J. C. (2000). Cellular and molecular mechanisms of action of bisphosphonates. *Cancer*, *88*(12, Suppl.), 2961–2978.
134. Frith, J. C., Rogers, M. J. (2003). Antagonistic effects of different classes of bisphosphonates in osteoclasts and macrophages in vitro. *J. Bone Miner. Res.*, *18*, 204–212.

135. Ross, J. R., Saunders, Y., Edmonds, P. M., Patel, S., Wonderling, D., Normand, C., Broadley, K. (2004). A systematic review of the role of bisphosphonates in metastatic disease. *Health Technol. Assess.*, *8*, 1–176.

136. Uladag, H. (2002). Bisphosphonates as a foundation of drug delivery to bone. *Curr. Pharm. Des.*, *8*, 1929–1944.

137. Kiang, D. T., Loken, M. K., Kennedy, B. J. (1979). Mechanism of the hypocalcemic effect of mithramycin. *J. Clin. Endocrinol. Metab.*, *48*, 341–344.

138. Hall, T. J., Schaeublin, M., Chambers, T. J. (1993). The majority of osteoclasts require mRNA and protein synthesis for bone resorption in vitro. *Biochem. Biophys. Res. Commun.*, *195*, 1245–1253.

139. Umland, S. P., Schleimer, R. P., Johnston, S. L. (2002). Review of the molecular and cellular mechanisms of action of glucocorticoids for use in asthma. *Pulm. Pharmacol. Ther.*, *15*, 35–50.

140. Goulding, N. J. (2004). The molecular complexity of glucocorticoid actions in inflammation: A four ring circus. *Curr. Opin. Pharmacol.*, *4*, 629–636.

141. White, P. C. (2001). Synthesis and metabolism of corticosteroids. In K. L. Becker (Ed.), *Principles and Practice of Endocrinology and Metabolism*, 3rd ed. Lippincott, Williams & Wilkins, Philadelphia, pp. 704–714.

142. Santen, R. J., Misbin, R. I. (1981). Aminoglutethimide: Review of pharmacology and clinical use. *Pharmacotherapy*, *1*, 95–120.

143. Lacroix, M., Hontela, A. (2003). The organochlorine *o,p′*-DDD disrupts the adrenal steroidogenic signaling pathway in rainbow trout (*Oncorhynchus mykiss*). *Toxicol. Appl. Pharmacol.*, *190*, 197–205.

144. Touitou, Y., Bogdan, A., Auzeby, A., Dommergues, J. P. (1979). Glucocorticoid and mineralocorticoid pathways in two adrenocortical carcinomas: Comparison of the effects of *o,p′*-dichlorodiphenyldichloroethane, aminoglutethimide and 2-*p*-aminophenyl-2-phenylethylamine in vitro. *J. Endocrinol.*, *82*, 87–94.

145. Fang, V. S. (1979). Cytotoxic activity of 1-(*o*-chlorophenyl)-1-(*p*-chlorophenyl)-2,2-dichloroethane (mitotane) and its analogs on feminizing adrenal neoplastic cells in culture. *Cancer Res.*, *39*, 139–145.

146. Morishita, K., Okumura, H., Ito, N., Takahashi, N. (2001). Primary culture system of adrenocortical cells from dogs to evaluate direct effects of chemicals on steroidogenesis. *Toxicology*, *165*, 171–178.

147. Lambert, F., Corcelle-Cerf, F., Lammerant, J., Kolanowski, J. (1984). On the specificity of the inhibitory effect of trilostane and aminoglutethimide on adrenocortical steroidogenesis in guinea pig. *Mol. Cell. Endocrinol.*, *37*, 115–120.

148. Lambert, A., Frost, J., Mitchell, R., Robertson, W. R. (1986). On the assessment of the in vitro biopotency and site(s) of action of drugs affecting adrenal steroidogenesis. *Ann. Clin. Biochem.*, *23*, 225–229.

149. Semple, C. G., Beastall, G. H., Gray, C. E., Thomson, J. A. (1983). Trilostane in the management of Cushing's syndrome. *Acta Endocrinol. (Copenh.)*, *102*, 107–110.

150. Ingle, J. N., Krook, J. E., Schaid, D. J., Everson, L. K., Mailliard, J. A., Long, H. J., McCormack, G. W. (1990). Evaluation of trilostane plus hydrocortisone in women with metastatic breast cancer and prior hormonal therapy exposure. *Am. J. Clin. Oncol.*, *13*, 93–97.

151. le Roux, P. A., Tregoning, S. K., Zinn, P. M., van der Spuy, Z. M. (2002). Inhibition of progesterone secretion with trilostane for mid-trimester termination of pregnancy: Randomized controlled trials. *Hum. Reprod.*, *17*, 1483–1489.

152. Weber, M. M., Lang, J., Abedinpour, F., Zeilberger, K., Adelmann, B., Engelhardt, D. (1993). Different inhibitory effect of etomidate and ketoconazole on the human adrenal steroid biosynthesis. *Clin. Investig.*, *71*, 933–938.

153. Dorr, H. G., Kuhnle, U., Holthausen, H., Bidlingmaier, F., Knorr, D. (1984). Etomidate: A selective adrenocortical 11 beta-hydroxylase inhibitor. *Klin. Wochenschr.*, *62*, 1011–1013.

154. Preziosi, P., Vacca, M. (1998). Adrenocortical suppression and other endocrine effects of etomidate. *Life Sci.*, *42*, 477–489.

155. Fuller, P. J. (1997). Aldosterone's effects and mechanism of action. *Curr. Opin. Endocrinol. Diabetes*, *4*, 218–224.

156. Rainey, W. E., White, P. (1998). Functional adrenal zonation and regulation of aldosterone biosynthesis. *Curr. Opin. Endocrinol. Diabetes*, *5*, 175–182.

157. Slight, S. H., Joseph, J., Ganjam, V. K., Weber, K. T. (1999). Extra-adrenal mineralocorticoids and cardiovascular tissue. *J. Mol. Cell. Cardiol.*, *31*, 1175–1184.

158. Ferrari, P., Krozowski, Z. (2000). Role of the 11beta-hydroxysteroid dehydrogenase type 2 in blood pressure regulation. *Kidney Int.*, *57*, 1374–1381.

159. Delcayre, C., Swynghedauw, B. (2002). Molecular mechanisms of myocardial remodeling. The role of aldosterone. *J. Mol. Cell. Cardiol.*, *34*, 1577–1584.

160. Oelkers, W., Buchen, S., Diederich, S., Krain, J., Muhme, S., Schoneshofer, M. (1994). Impaired renal 11 beta-oxidation of 9 alpha-fluorocortisol: An explanation for its mineralocorticoid potency. *J. Clin. Endocrinol. Metab.*, *78*, 928–932.

161. Oldenburg, O., Kribben, A., Baumgart, D., Philipp, T., Erbel, R., Cohen, M. V. (2002). Treatment of orthostatic hypotension. *Curr. Opin. Pharmacol.*, *2*, 740–747.

162. Cole, T. J., Pierce, D. (2001). Mineralocorticoid target genes. *Curr. Opin. Endocrinol. Diabetes*, *8*, 118–123.

163. Funder, J. W. (2001). Non-genomic actions of aldosterone: Role in hypertension. *Curr. Opin. Nephrol. Hypertens.*, *10*, 227–230.

164. Christ, M., Wehling, M. (1999). Rapid actions of aldosterone: Lymphocytes, vascular smooth muscle and endothelial cells. *Steroids*, *64*, 35–41.

165. Pitt, B., Zannad, F., Remme, W. J., Cody, R., Castaigne, A., Perez, A., Palensky, J., Wittes, J. (1999). The effect of spironolactone on morbidity and mortality in patients with severe heart failure. *N. Engl. J. Med.*, *341*, 709–717.

166. Soberman, J. E., Weber, K. T. (2000). Spironolactone in congestive heart failure. *Curr. Hypertens. Rep.*, *2*, 451–456.

167. Cittadini, A., Monti, M. G., Isgaard, J., Casaburi, C., Stromer, H., Di Gianni, A., Serpico, R. Saldamarco, L., Vanasia, M., Sacca, L. (2003). Aldosterone receptor blockade improves left ventricular remodeling and increases ventricular fibrillation threshold in experimental heart failure. *Cardiovasc. Res.*, *58*, 555–564.

168. de Gasparo, M., Joss, U., Ramjoue, H. P., Whitebread, S. E., Haenni, H., Schenkel, L., Kraehenbuehl, C., Biollaz, M., Grob, J., Schmidlin, J. (1987). Three new epoxy-spirolactone derivatives: Characterization in vivo and in vitro. *J. Pharmacol. Exp. Ther.*, *240*, 650–656.

169. McMahon, E. G. (2001). Recent studies with eplerenone, a novel, selective aldosterone receptor antagonist. *Curr. Opin. Pharmacol.*, *1*, 190–196.
170. Pitt, B., Remme, W., Zannad, F., Neaton, J., Martinez, F., Roniker, B., Bittman, R., Hurley, S., Kleiman, J., Gatlin, M. (2003). Eplerenone Post-Acute Myocardial Infarction Heart Failure Efficacy and Survival Study Investigators. Eplerenone, a selective aldosterone blocker, in patients with left ventricular dysfunction after myocardial infarction. *N. Engl. J. Med.*, *348*, 1309–1321.
171. Gobinet, J., Poujol, N., Sultan, Ch. (2002). Molecular action of androgens. *Mol. Cell. Endocrinol.*, *198*, 15–24.
172. Shahidi, N. T. (2001). A review of the chemistry, biological action, and clinical applications of anabolic-androgenic steroids. *Clin. Ther.*, *23*, 1355–1390.
173. Bhasin, S., Storer, T. W., Berman, N., Yarasheski, K. E., Clevenger, B., Phillips, J., Lee, W. P., Bunnell, T. J., Casaburi, R. (1997). Testosterone replacement increases fat-free mass and muscle size in hypogonadal men. *J. Clin. Endocrinol. Metab.*, *82*, 407–413.
174. Kuhn, C. M. (2002). Anabolic steroids. *Recent Prog. Horm. Res.*, *57*, 411–434.
175. Bergink, E. W., Janssen, P. S., Turpijn, E. W., van der Vies, J. (1985). Comparison of the receptor binding properties of nandrolone and testosterone under in vitro and in vivo conditions. *J. Steroid Biochem.*, *22*, 831–836.
176. Sundaram, K., Kumar, N., Monder, C., Bardin, C. W. (1995). Different patterns of metabolism determine the relative anabolic activity of 19-norandrogens. *J. Steroid Biochem. Mol. Biol.*, *53*, 253–257.
177. Zhao, J., Bauman, W. A., Huang, R., Caplan, A. J., Cardozo, C. (2004). Oxandrolone blocks glucocorticoid signaling in an androgen receptor-dependent manner. *Steroids*, *69*, 357–366.
178. Negro-Vilar, A. (1999). Selective androgen receptor modulators (SARMs): A novel approach to androgen therapy for the new millennium. *J. Clin. Endocrinol. Metab.*, *84*, 3459–3466.
179. Yin, D., Gao, W., Kearbey, J. D., Xu, H., Chung, K., He, Y., Marhefka, C. A., Veverka, K. A., Miller, D. D., Dalton, J. T. (2003). Pharmacodynamics of selective androgen receptor modulators. *J. Pharmacol. Exp. Ther.*, *304*, 1334–1340.
180. Berrevoets, C. A., Umar, A., Brinkmann, A. O. (2002). Antiandrogens: Selective androgen receptor modulators. *Mol. Cell. Endocrinol.*, *198*, 97–103.
181. Rittmaster, R. S. (1999). Antiandrogen treatment of polycystic ovary syndrome. *Endocrinol. Metab. Clin. North Am.*, *28*, 409–421.
182. Occhiato, E. G., Guarna, A., Danza, G., Serio, M. (2004). Selective non-steroidal inhibitors of 5 alpha-reductase type 1. *J. Steroid Biochem. Mol. Biol.*, *88*, 1–16.
183. McDonnell, D. P., Connor, C. E., Wijayaratne, A., Chang, C. Y., Norris, J. D. (2002). Definition of the molecular and cellular mechanisms underlying the tissue-selective agonist/antagonist activities of selective estrogen receptor modulators. *Recent Prog. Horm. Res.*, *57*, 295–316.
184. Ruff, M., Gangloff, M., Wurtz, J. M., Moras, D. (2000). Estrogen receptor transcription and transactivation: Structure-function relationship in DNA- and ligand-binding domains of estrogen receptors. *Breast Cancer Res.*, *2*, 353–359.
185. Gustafsson, J. A. (2003). What pharmacologists can learn from recent advances in estrogen signalling. *Trends Pharmacol. Sci.*, *24*, 479–485.

186. Segars, J. H., Drigger, P. H. (2002). Estrogen action and cytoplasmic signaling cascades. Part I: Membrane-associated signaling complexes. *Trends Endocrinol. Metab*, *13*, 349–354.

187. Chen, Z., Yuhanna, I. S., Galcheva-Gargova, Z., Karas, R. H., Mendelsohn, M. E., Shaul, P. W. (1999). Estrogen receptor alpha mediates the nongenomic activation of endothelial nitric oxide synthase by estrogen. *J. Clin. Invest.*, *103*, 401–406.

188. Harrington, W. R., Sheng, S., Barnett, D. H., Petz, L. N., Katzenellenbogen, J. A., Katzenellenbogen, B. S. (2003). Activities of estrogen receptor alpha- and beta-selective ligands at diverse estrogen responsive gene sites mediating transactivation or transrepression. *Mol. Cell. Endocrinol.*, *206*, 13–22.

189. Campos, S. M. (2004). Aromatase inhibitors for breast cancer in postmenopausal women. *Oncologist*, *9*, 126–136.

190. Zeitoun, K. M., Bulun, S. E. (1999). Aromatase: A key molecule in the pathophysiology of endometriosis and a therapeutic target. *Fertil. Steril.*, *72*, 961–969.

191. Mitwally, M. F., Casper, R. F. (2002). Aromatase inhibition for ovarian stimulation: Future avenues for infertility management. *Curr. Opin. Obstet. Gynecol.*, *14*, 255–263.

192. Kao, Y. C., Okubo, T., Sun, X. Z., Chen, S. (1999). Induction of aromatase expression by aminoglutethimide, an aromatase inhibitor that is used to treat breast cancer in postmenopausal women. *Anticancer Res.*, *19*, 2049–2056.

193. Anderson, G. L., Limacher, M., Assaf, A. R., Bassford, T., Beresford, S. A., Black, H., Bonds, D., Brunner, R., Brzyski, R., Caan, B., Chlebowski, R., Curb, D., Gass, M., Hays, J., Heiss, G., Hendrix, S., Howard, B. V., Hsia, J., Hubbell, A., Jackson, R., Johnson, K. C., Judd, H., Kotchen, J. M., Kuller, L., LaCroix, A. Z., Lane, D., Langer, R. D., Lasser, N., Lewis, C. E., Manson, J., Margolis, K., Ockene, J., O'Sullivan, M. J., Phillips, L., Prentice, R. L., Ritenbaugh, C., Robbins, J., Rossouw, J. E., Sarto, G., Stefanick, M. L., Van Horn, L., Wactawski-Wende, J., Wallace, R., Wassertheil-Smoller, S. (2004). Women's Health Initiative Steering Committee. Effects of conjugated equine estrogen in postmenopausal women with hysterectomy: The Women's Health Initiative randomized controlled trial. *JAMA*, *291*, 1701–1712.

194. Conneely, O. M., Mulac-Jericevic, B., DeMayo, F., Lydon, J. P., O'Malley, B. W. (2002). Reproductive functions of progesterone receptors. *Recent Prog. Horm. Res.*, *57*, 339–355.

195. Sutter-Dub, M. T. (2002). Rapid non-genomic and genomic responses to progestogens, estrogens, and glucocorticoids in the endocrine pancreatic B cell, the adipocyte and other cell types. *Steroids*, *67*, 77–93.

196. Sager, G., Orbo, A., Jaeger, R., Engstrom, C. (2003). Non-genomic effects of progestins—Inhibition of cell growth and increased intracellular levels of cyclic nucleotides. *J. Steroid Biochem. Mol. Biol.*, *84*, 1–8.

197. Schindler, A. E., Campagnoli, C., Druckmann, R., Huber, J., Pasqualini, J. R., Schweppe, K. W., Thijssen, J. H. (2003). Classification and pharmacology of progestins. *Maturitas*, *46*(Suppl. 1), S7–S16.

198. The Diabetes Control and Complications Trial Research Group (1993). The effect of intensive treatment of diabetes on the development and progression of long-term complications in insulin-dependent diabetes mellitus. *N. Engl. J. Med.*, 329, 977–986.

199. Turner, R. C. (1998). The U.K. Prospective Diabetes Study. A review. *Diabetes Care*, *21*(Suppl. 3), C35–38.

200. Kahn, S. E. (2003). The relative contribution of insulin resistance annd beta cell dysfunction to the pathophysiology of Type 2 diabetes. *Diabetologia*, *46*, 3–19.

201. Porte, D. Jr., Kahn, S. E. (2001). Beta-cell dysfunction and failure in type 2 diabetes: Potential mechanisms. *Diabetes*, *50*(Suppl. 1), S160–163.

202. Juhan-Vague, I., Alessi, M. C. (1997). PAI-1, obesity, insulin resistance and risk of cardiovascular events. *Thromb. Haemost.*, *78*, 656–660.

203. Bastard, J. P., Pieroni, L., Hainque, B. (2000). Relationship between plasma plasminogen activator inhibitor 1 and insulin resistance. *Diabetes Metab. Res. Rev.*, *16*, 192–201.

204. Howell, S. L., Jones, P. M., Persaud, S. J. (1994). Regulation of insulin secretion: The role of second messengers. *Diabetologia*, *37*(Suppl. 2), S30–35.

205. Dornhorst, A. (2001). Insulinotropic meglitinide analogues. *Lancet*, *358*, 1709–1716.

206. Vajo, Z., Duckworth, W. C. (2000). Genetically engineered insulin analogs: Diabetes in the new millenium. *Pharmacol. Rev.*, *52*, 1–9.

207. Owens, D. R. (2002). Opimising glycaemic control—The role of long-acting insulin analogues in basal insulin therapy. *Br. J. Diabetes Cardiovasc. Dis.*, *2*, 403–407.

208. Alessi, D. R. (2001). Discovery of PDK1, one of the missing links in insulin signal transduction. Colworth Medal Lecture. *Biochem Soc. Trans.*, *29*(Pt. 2), 1–14.

209. Saltiel, A. R., Kahn, C. R. (2001). Insulin signalling and the regulation of glucose and lipid metabolism. *Nature*, *414*, 799–806.

210. Khan, A. H., Pessin, J. E. (2002). Insulin regulation of glucose uptake: A complex interplay of intracellular signalling pathways. *Diabetologia*, *45*, 1475–1483.

211. Schafer, G. (1983). Biguanides: A review of history, pharmacodynamics and therapy. *Diabetes Metab. Rev.*, *9*, 148–163.

212. Bailey, C. J. (1992). Biguanides and NIDDM. *Diabetes Care*, *15*, 755–772.

213. UKPDS. (1998). Intensive blood glucose control with sulfonylureas or insulin compared with conventional treatment and risk of complications in patients with type 2 diabetes (UKPDS 33). UK Prospective Diabetes Study Group. *Lancet*, *352*, 837–853.

214. Radziuk, J., Bailey, C. J., Wiernsperger, N. F., Yudkin, J. S. (2003). Metformin and its liver targets in the treatment of type 2 diabetes. *Curr. Drug Targets Immune Endocr. Metab. Disord.*, *3*, 151–169.

215. Zhou, G., Myers, R., Li, Y., Chen, Y., Shen, X., Fenyk-Melody, J., Wu, M., Ventre, J., Doebber, T., Fujii, N., Musi, N., Hirshman, M. F., Goodyear, L. J., Moller, D. E. (2001). Role of AMP-activated protein kinase in mechanism of metformin action. *J. Clin. Invest.*, *108*, 1167–1174.

216. Hardie, D. G. (2003). Minireview: The AMP-activated protein kinase cascade: The key sensor of cellular energy status. *Endocrinology*, *144*, 5179–5183.

217. Seli, E., Duleba, A. J. (2004). Treatment of PCOS with metformin and other insulin-sensitizing agents. *Curr. Diabetes Rep.*, *4*, 69–75.

218. Kay, J. P., Alemzadeh, R., Langley, G., D'Angelo, L., Smith, P., Holshouser, S. (2001). Beneficial effects of metformin in normoglycemic morbidly obese adolescents. *Metabolism*, *50*, 1457–1461.
219. Olefsky, J. M., Saltiel, A. R. (2000). PPAR gamma and the treatment of insulin resistance. *Trends Endocrinol. Metab.*, *11*, 362–368.
220. Desvergne, B., Wahli, W. (1999). Peroxisome proliferator-activated receptors: Nuclear control of metabolism. *Endocr. Rev.*, *20*, 649–688.
221. Auwerx, J. (1999). PPAR-γ, the ultimate thrifty gene. *Diabetologia*, *42*, 1033–1049.
222. Rangwala, S. M., Lazar, M. A. (2004). Peroxisome proliferator-activated receptor gamma in diabetes and metabolism. *Trends Pharmacol. Sci.*, *25*, 331–336.
223. Krentz, A. J., Bailey, C. J., Melander, A. (2000). Thiazolidinediones for type 2 diabetes. New agents reduce insulin resistance but need long term clinical trials. *BMJ*, *321*, 252–253.
224. Niemi, M., Backman, J. T., Neuvonen, M., Neuvonen, P. J., Kivsito, K. T. (2001). Effects of rifampicin on the pharmacokinetics and pharmacodynamics of glyburide and glipizide. *Clin. Pharmacol. Ther.*, *69*, 400–406.
225. Niemi, M., Cascorbi, I., Timm, R., Kroemer, H. K., Neuvonen, P. J., Kivisto, K. T. (2002). Glyburide and glimepiride pharmacokinetics in subjects with different CYP2C9 genotypes. *Clin. Pharmacol. Ther.*, *72*, 326–332.
226. Meinert, C. L., Knatterud, G. L., Prout, T. E., Klimt, C. R. (1970). A study of the effects of hypoglycemic agents on vascular complications in patients with adult-onset diabetes. II. Mortality results. *Diabetes*, *19*(Suppl.), 789–830.
227. Aguilar-Bryan, L., Bryan, J. (1999). Molecular biology of adenosine triphosphate-sensitive potassium channels. *Endocr. Rev.*, *20*, 101–135.
228. Gribble, F. M. Reimann, F. (2003). Sulphonylurea action revisited: The post-cloning era. *Diabetologia*, *46*, 875–891.
229. Meyer, M., Chudziak, F., Schwanstecher, C., Schwanstecher, M., Panten, U. (1999). Structural requirements of sulphonylureas and analogues for interaction with sulphonylurea receptor subtypes. *Br. J. Pharmacol.*, *128*, 27–34.
230. Kurowski, T. G., Lin, Y., Luo, Z., Tsichlis, P. N., Buse, M. G., Heydrick, S. J., Ruderman, N. B. (1999). Hyperglycemia inhibits insulin activation of Akt/protein kinase B but not phosphatidylinositol 3-kinase in rat skeletal muscle. *Diabetes*, *48*, 658–663.
231. Muller, G., Geisen, K. (1996). Characterization of the molecular mode of action of the sulfonylurea, glimepiride, at adipocytes. *Horm. Metab. Res.*, *28*, 469–487.
232. Chutkow, W. A., Samuel, V., Hansen, P. A., Pu, J., Valdivia, C. R., Makielski, J. C., Burant, C. F. (2001). Disruption of Sur2-containing K(ATP) channels enhances insulin-stimulated glucose uptake in skeletal muscle. *Proc. Natl. Acad. Sci. USA*, *98*, 11760–11764.
233. Goldfine, A. B., Maratos-Flier, E. (2001). Oral agents for the treatment of type 2 diabetes. In K. L. Becker (Ed.), *Principles and Practice of Endocrinology and Metabolism*, 3rd ed. Lippincott, Williams & Wilkins, Philadelphia pp. 1344–1348.
234. Riddle, M. C., Schneider, J. (1998). Beginning insulin treatment of obese patients with evening 70/30 insulin plus glimepiride versus insulin alone. Glimepiride Combination Group. *Diabetes Care*, *21*, 1052–1057.

235. Jennings, P. E. (2000). Vascular benefits of gliclazide beyond glycemic control. *Metabolism*, *49*(10, Suppl. 2), 17–20.

236. Harrower, A. D. (1996). Pharmacokinetics of oral antihyperglycaemic agents in patients with renal insufficiency. *Clin. Pharmacokinet.*, *31*, 111–119.

237. Henquin, J. C. (1990). Established, unsuspected and novel pharmacological insulin secretagogues. In C. J. and Bailey, P. R. Flatt (Eds), *New Antidiabetic Drugs*, 1st ed. Smith Gordon and Company, London, pp. 93–106.

238. Hanefeld, M., Bouter, K. P., Dickinson, S., Guitard, C. (2000). Rapid and short-acting mealtime insulin secretion with nateglinide controls both prandial and mean glycemia. *Diabetes Care*, *23*, 202–207.

239. Weaver, M. L., Orwig, B. A., Rodriguez, L. C., Graham, E. D., Chin, J. A., Shapiro, M. J., McLeod, J. F., Mangold, J. B. (2001). Pharmacokinetics and metabolism of nateglinide in humans. *Drug Metab. Dispos.*, *29*(4, Pt. 1), 415–421.

240. Hatorp, V. (2002). Clinical pharmacokinetics and pharmacodynamics of repaglinide. *Clin. Pharmacokinet.*, *41*, 471–483.

241. Hu, S., Wang, S., Fanelli, B., Bell, P. A., Dunning, B. E., Geisse, S., Schmitz, R., Boettcher, B. R. (2000). Pancreatic β-cell $K_{ATP}$ channel activity and membrane-binding studies with nateglinide: A comparison with sulfonylureas and repaglinide. *J. Pharmacol. Exp. Ther.*, *293*, 444–452.

242. Hansen, A. M., Christensen, I. T., Hansen, J. B., Carr, R. D., Ashcroft, F. M., Wahl, P. (2002). Differential interactions of nateglinide and repaglinide on the human beta-cell sulphonylurea receptor 1. *Diabetes*, *51*, 2789–2795.

243. Salvatore, T., Giugliano, D. (1996). Pharmacokinetic-pharmacodynamic relationships of acarbose. *Clin. Pharmacokinet.*, *30*, 94–106.

244. Hollander, P., Pi-Sunyer, X., Coniff, R. F. (1997). Acarbose in the treatment of type 1 diabetes. *Diabetes Care*, *20*, 248–253.

245. Holman, R. R., Cull, C. A., Turner, R. C. (1999). A randomized double blind trial of acarbose in type 2 diabetes shows improved glycemic control over 3 years (U.K. Prospective Diabtes Study 44). *Diabetes Care*, *22*, 960–964.

246. Chiasson, J. L., Josse, R. G., Gomis, R., Hanefeld, M., Karasik, A., Laakso, M. (2002). STOP-NIDDM Trail Research Group. Acarbose for prevention of type 2 diabetes mellitus: The STOP-NIDDM randomised trial. *Lancet*, *359*, 2072–2077.

247. Doyle, M. E., Egan, J. M. (2001). Glucagon-like peptide-1. *Recent Prog. Horm. Res.*, *56*, 377–399.

248. Drucker, D. J. (2001). Minireview: The glucagon-like peptides. *Endocrinology*, *142*, 521–527.

249. Gault, V. A., Flatt, P. R., O'Harte, F. P. (2003). Glucose-dependent insulinotropic polypeptide analogues and their therapeutic potential for the treatment of obesity-diabetes. *Biochem. Biophys. Res. Commun.*, *308*, 207–213.

250. Preitner, F., Ibberson, M., Franklin, I., Binnert, C., Pende, M., Gjinovci, A., Hansotia, T., Drucker, D. J., Wollheim, C., Burcelin, R., Thorens, B. (2004). Gluco-incretins control insulin secretion at multiple levels as revealed in mice lacking GLP-1 and GIP receptors. *J. Clin. Invest.*, *113*(4), 635–645.

251. Creutzfeldt, W. O., Kleine, N., Willms, B., Orskov, C., Holst, J. J., Nauck, M. A. (1996). Glucagonostatic actions and reduction of fasting hyperglycemia by exoge-

nous glucagon-like peptide I(7-36) amide in type I diabetic patients. *Diabetes Care*, *19*, 580–586.

252. Ahren, B. (2003). Gut peptides and type 2 diabetes mellitus treatment. *Curr. Diabetes Rep.*, *3*, 365–372.
253. Wiedeman, P. E., Trevillyan, J. M. (2003). Dipeptidyl peptidase IV inhibitors for the treatment of impaired glucose tolerance and type 2 diabetes. *Curr. Opin. Investig. Drugs*, *4*, 412–420.
254. Eng, J. (1992). Exendin peptides. *Mt. Sinai J. Med.*, *59*, 147–149.
255. Drucker, D. J. (1998). Glucagon-like peptides. *Diabetes*, *47*, 159–169.
256. Goke, R., Fehmann, H. C., Linn, T., Schmidt, H., Krause, M., Eng, J., Goke, B. (1993). Exendin-4 is a high potency agonist and truncated exendin-(9-39)-amide an antagonist at the glucagon-like peptide 1-(7-36)-amide receptor of insulin secreting beta-cells. *J. Biol. Chem.*, *268*, 19650–19655.
257. Young, A. A., Gedulin, B. R., Bhavsar, S., Bodkin, N., Jodka, C., Hansen, B., Denaro, M. (1999). Glucose-lowering and insulin-sensitizing actions of exendin-4: Studies in obese diabetic (ob/ob, db/db) mice, diabetic fatty Zucker rats, and diabetic rhesus monkeys (*Macaca mulatta*). *Diabetes*, *48*, 1026–1034.
258. Edwards, C. M., Stanley, S. A., Davis, R., Brynes, A. E., Frost, G. S., Seal, L. J., Ghatei, M. A., Bloom, S. R. (2001). Exendin-4 reduces fasting and postprandial glucose and decreases energy intake in healthy volunteers. *Am. J. Physiol. Endocrinol. Metab.*, *281*, E155–E161.
259. Egan, J. M., Clocquet, A. R., Elahi, D. (2002). The insulinotropic effect of acute exendin-4 administered to humans: Comparison of nondiabetic state to type 2 diabetes. *J. Clin. Endocrinol. Metab.*, *87*, 1282–1290.
260. Xu, G., Stoffers, D. A., Habener, J. F., Bonner-Weir, S. (1999). Exendin-4 stimulates both beta-cell replication and neogenesis, resulting in increased beta-cell mass and improved glucose tolerance in diabetic rats. *Diabetes*, *48*, 2270–2276.
261. Taborsky, G. J. Jr, Ahren, B., Havel, P. J. (1998). Autonomic mediation of glucagon secretion during hypoglycemia: Implications for impaired alpha-cell responses in type 1 diabetes. *Diabetes*, *47*, 995–1005.
262. Exton, J. H. (1987). Mechanisms of hormonal regulation of hepatic glucose metabolism. *Diabetes Metab. Rev.*, *3*, 163–183.
263. Schade, D. S., Woodside, W., Eaton, R. P. (1979). The role of glucagon in the regulation of plasma lipids. *Metabolism*, *28*, 874–886.
264. Gu, Z., Jensen, R. T., Maton, P. N. (1992). A primary role for protein kinase A in smooth muscle relaxation induced by adrenergic agonists and neuropeptides. *Am. J. Physiol.*, *263*(3, Pt. 1), G360–364.
265. Mayo, K. E., Miller, L. J., Bataille, D., Dalle, S., Goke, B., Thorens, B., Drucker, D. J. (2003). International Union of Pharmacology. XXXV. The glucagon receptor family. *Pharmacol. Rev.*, *55*, 167–194.
266. Unger, R. H. (1978). Role of glucagon in the pathogenesis of diabetes: The status of the controversy. *Metabolism*, *27*, 1691–1709.
267. Madsen, P., Knudsen, L. B., Wiberg, F. C., Carr, R. D. (1998). Discovery and structure-activity relationship of the first non-peptide competitive human glucagon receptor antagonists. *J. Med. Chem.*, *41*, 5150–5157.

268. Petersen, K. F., Sullivan, J. T. (2001). Effects of a novel glucagon receptor antagonist (Bay 27-9955) on glucagon-stimulated glucose production in humans. *Diabetologia*, *44*, 2018–2024.
269. Ladouceur, G. H., Cook, J. H., Hertzog, D. L., Jones, J. H., Hundertmark, T., Korpusik, M., Lease, T. G., Livingston, J. N., MacDougall, M. L., Osterhout, M. H., Phelan, K., Romero, R. H., Schoen, W. R., Shao, C., Smith, R. A. (2002). Integration of optimized substituent patterns to produce highly potent 4-aryl-pyridine glucagon receptor antagonists. *Bioorg. Med. Chem. Lett.*, *12*, 3421–3424.
270. Ying, J., Ahn, J. M., Jacobsen, N. E., Brown, M. F., Hruby, V. J. (2003). NMR solution structure of the glucagon antagonist [desHis1, desPhe6, Glu9]glucagon amide in the presence of perdeuterated dodecylphosphocholine micelles. *Biochemistry*, *42*, 2825–2835.
271. Madsen, P., Ling, A., Plewe, M., Sams, C. K., Knudsen, L. B., Sidelmann, U. G., Ynddal, L., Brand, C. L., Andersen, B., Murphy, D., Teng, M., Truesdale, L., Kiel, D., May, J., Kuki, A., Shi, S., Johnson, M. D., Teston, K. A., Feng, J., Lakis, J., Anderes, K., Gregor, V., Lau, J. (2002). Optimization of alkylidene hydrazide based human glucagon receptor antagonists. Discovery of the highly potent and orally available 3-cyano-4-hydroxybenzoic acid [1-(2,3,5,6-tetramethylbenzyl)-*1H*-indol-4-ylmethylene]hydrazide. *J. Med. Chem.*, *45*, 5755–5775.
272. Rothman, R. B., Baumann, M. H. (2002). Serotonin releasing agents. Neurochemical, therapeutic and adverse effects. *Pharmacol. Biochem. Behav.*, *71*, 825–836.
273. Gross, S. B., Lepor, N. E. (2000). Anorexigen-related cardiopulmonary toxicity. *Rev. Cardiovasc. Med.*, *1*, 80–89.
274. Finer, N. (2002). Sibutramine: Its mode of action and efficacy. *Int. J. Obes. Relat. Metab. Disord*, *26*(Suppl. 4), S29–S33.
275. Williams, G., Bing, C., Cai, X. J., Harrold, J. A., King, P. J., Liu, X. H. (2001). The hypothalamus and the control of energy homeostasis: Different circuits, different purposes. *Physiol. Behav.*, *74*(4–5), 683–701.
276. Druce, M. R., Small, C. J., Bloom, S. R. (2004). Minireview: Gut peptides regulating satiety. *Endocrinology*, *145*, 2660–2665.
277. Steppan, C. M., Lazar, M. A. (2002). Resistin and obesity-associated insulin resistance. *Trends Endocrinol. Metab.*, *13*, 18–23.
278. Crowley, V., Vidal-Puig, A. J. (2001). Mitochondrial uncoupling proteins (UCPs) and obesity. *Nutr. Metab. Cardiovasc. Dis.*, *11*, 70–75.
279. Weyer, C., Gautier, J. F., Danforth, E. Jr. (1999). Development of beta 3-adrenoceptor agonists for the treatment of obesity and diabetes—An update. *Diabetes Metab.*, *25*(1), 11–21.
280. Ballinger, A., Peikin, S. R. (2002). Orlistat: Its current status as an anti-obesity drug. *Eur. J. Pharmacol.*, *440*(2–3), 109–117.
281. Clapham, J. C., Arch, J. R., Tadayyon, M. (2001). Anti-obesity drugs: A critical review of current therapies and future opportunities. *Pharmacol. Ther.*, *89*, 81–121.

# RESPIRATORY VIRUSES

PAUL D. OLIVO
*Apath, LLC,*
*St. Louis, Missouri*

1 INTRODUCTION
2 SEASONAL PATTERNS
3 PUBLIC HEALTH AND ECONOMIC IMPACT
4 CONTROL OF VIRAL RESPIRATORY INFECTIONS
5 SPECIFIC VIRUS PATHOGENS
Picornaviridae
Coronaviridae
Paramyxoviridae
Orthomyxoviridae
Bunyaviridae
Adenoviridae
Herpesviridae
6 MAJOR CLINICAL SYNDROMES
Common Cold
Pharyngitis
Laryngitis
Acute Laryngotracheobronchitis (Croup)
Tracheitisand Tracheobronchitis
Bronchiolitis
Pneumonia
7 DIAGNOSIS OF VIRAL RESPIRATORY INFECTION
8 EXISTING ANTIVIRAL THERAPY
9 CHALLENGES OF ANTIVIRAL THERAPY FOR RESPIRATORY VIRUS INFECTIONS
10 STRATEGIES FOR DEVELOPMENT OF ANTIVIRAL THERAPY

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

11 DEVELOPMENT OF ANTIVIRALS AGAINST RESPIRATORY VIRUSES
12 ANIMAL MODELS
13 SUMMARY
References

## 1 INTRODUCTION

The respiratory system is very susceptible to microbial invasion. Despite the numerous defense mechanisms that the body uses to counteract microbial invasion, respiratory tract infections are among the most common human maladies. Viruses cause disease by one of several mechanisms. They cause cell death, organ malfunction, or elicit an inflammatory reaction that can account for many of the symptoms and pathology of infection. Pneumonia, or infection within lung tissue, is one of the leading causes of mortality in the world, and viruses contribute either directly or indirectly to a significant fraction of cases of pneumonia.

Viruses are submicroscopic entities of between 20 and 300 nm in diameter. They consist of a nucleic acid genome, which can be either RNA (ribonucleic acid) or DNA (deoxyribonucleic acid), and a protein coat called a capsid. Some viruses also are enveloped in a lipid membrane. Viruses lack metabolic organelles and thus have no ability to generate energy and can only replicate within living cells. Viruses, therefore, exist on the border between inanimate chemistry and life forms.

Based on their overall structure and genetic organization, animal viruses are classified into dozens of different families. Classification is not based on disease because there is no relationship between the type of virus and the diseases they cause. Several hundred antigenically distinct viruses from at least seven virus families are associated with sporadic or epidemic respiratory infections in children and adults [1].

## 2 SEASONAL PATTERNS

Many viral respiratory diseases are associated with a seasonal variation in incidence. Influenza virus and respiratory syncytial virus (RSV) infections occur predominantly in winter months with an overlap of peak periods of incidence [2]. Rhinoviruses are isolated throughout the year but their peak incidence is spring and fall [3]. Adenoviruses and herpesviruses in general do not show seasonal variation with the exception of varicella-zoster virus (chicken pox), which is more common in late winter and early spring.

## 3 PUBLIC HEALTH AND ECONOMIC IMPACT

Respiratory viral infections have a major impact on health. The rate of respiratory viral infections has been calculated based on surveys to be over 85 illnesses per 100 persons per year in the United States, making acute viral respiratory infections among the most common illnesses experienced by people within the United States and probably worldwide [4]. Between a fourth and a half of these illnesses lead people to seek medical attention and cause a significant amount of lost work and school days [5]. The overall burden of disease related to respiratory virus infections, including mortality, hospitalizations, medical costs (asthma, etc.), and societal costs (i.e., lost productivity, etc.) is immense. The resulting negative economic impact of viral respiratory disease in the United States is huge.

Mortality due to acute respiratory infection in otherwise healthy children and adults is rare except for influenza during epidemics. However, acute respiratory infections (viral and bacterial) are the leading cause of morbidity and mortality in children, with several million deaths worldwide and 100,000 deaths per year in the United States. In the 2003/2004 season there were 121 influenza-related deaths in children [6]. In developing countries acute respiratory infections are a major cause of childhood mortality. Measles, for example, which has been successfully controlled by vaccination in developed countries, remains among the major cause of mortality worldwide [7].

## 4 CONTROL OF VIRAL RESPIRATORY INFECTIONS

There are four fundamental modalities that are used to control respiratory infections within a population.

1. Public health measures
2. Vaccination
3. Antiviral therapy
4. Personal hygiene and etiquette

No one of these modalities alone can be counted upon to ensure that society can minimize the negative impact of these diseases on the population. However, particular modalities can predominate in importance for certain viruses. For example, during the 2002/2003 outbreak of severe acute respiratory syndrome (SARS), public health measures, and specifically rapid identification and isolation of infected patients, were critical to the successful efforts to control the outbreak [8]. For influenza, a combination of public health measures and vaccination are the most important control vehicles. Antiviral therapy has not played a major role in the overall management of respiratory infections. Hopefully this will change in the future. Personal hygiene is very

important, and the lack of appropriate hygiene enhances person-to-person transmission of many respiratory viruses. It is particularly important for individuals with respiratory virus infections such as influenza to minimize transmission to others by avoiding school and work during the time that they are coughing and sneezing. If this is not possible, a face mask should be worn and close contact should be avoided or minimized.

## 5 SPECIFIC VIRUS PATHOGENS

Viruses of many types (enveloped/nonenveloped; DNA/RNA, etc.) cause respiratory tract infections, and there is no relationship between type of disease and the nature of the viral structure or genome. In up to 30 percent of lower respiratory infections no pathogen is identified. Although this is in large part due to diagnostic difficulties, this statistic suggests that many new viruses remain to be discovered. In fact, in the past several decades a new human virus has been discovered approximately every 2 years. Viruses from seven families cause most respiratory tract infections. A brief overview of each family and several of the specific viruses that cause respiratory infections is provided. For a more detailed description of these viruses and their replication cycles, the reader is directed to other chapters in this text and various review articles and books [9, 10].

### Picornaviridae

Picornaviridae are nonenveloped, positive-sense RNA viruses that include several viruses that cause respiratory infections including rhinoviruses and enteroviruses. Rhinoviruses are the largest genus of the Picornaviridae family and the single most common cause of the common cold. Rhinoviruses have also been increasingly recognized as a cause of hospitalizations related to acute respiratory tract infections in the United States. Since their discovery in 1956, over 100 different serotypes have been identified. This antigenic variability contributes to the high number of rhinovirus infections that individuals experience during their lifetime.

The rhinovirus virion structure has been extensively studied, and the knowledge gained from this has been put to practical use in the development of a potent rhinovirus inhibitor (see below). The rhinovirus genome consists of a single strand of RNA of 7.2 kilobases that codes for a single polyprotein that is proteolytically cleaved by 2 viral proteases during and following translation into 11 proteins. This proteolysis is a key target of strategies to develop antirhinovirus drugs. Nevertheless, despite the prospect for new drugs, treatment of rhinovirus infections will be difficult. Studies with human volunteers have shown that the incubation period of rhinovirus infection of the upper respiratory tract is as short as 12 h [11, 12]. This makes the window of therapeutic opportunity quite short.

## Coronaviridae

The Coronaviridae family is a large family of enveloped, positive-sense RNA viruses with a distinct crown-shaped (Latin, *corona* = crown) virion that is created by projections of the virus spike protein. At about 29 kilobases, the RNA genome of members of this family is the largest genome among the animal RNA viruses. The family includes two genera that contain many human and animal viruses. The genus *Coronavirus*, which consists of three antigenically distinct groups, contains all the known human pathogens in this family. There are four human coronaviruses that have been well characterized: human coronavirus 229E (HCoV-229E), HCoV-OC43, SARS-associated coronavirus (SARS-CoV), and the recently described coronavirus, HCoV-NL63. HCoV-229E and HCoV-OC43 are the prototype strains of group 1 and group 2, respectively, and, along with numerous other strains, are a major cause of upper respiratory tract disease. HCoV-NL63 was recently isolated from a 7-month-old child suffering from bronchiolitis and conjunctivitis [13]. The complete genome sequence of HCoV-NL63 indicates that this virus is not a recombinant but rather a new group 1 coronavirus. Screening of clinical specimens from individuals suffering from respiratory illness identified seven additional HCoV-NL63-infected individuals, indicating that the virus is widely spread within the human population [13].

Severe acute respiratory syndrome is a viral respiratory illness caused by SARS-associated coronavirus (SARS-CoV) [14]. It was first described following cases in 2003 in the Guandong Province of southern China and gained international attention following an outbreak that spread to numerous countries via a businessman who traveled to a Hong Kong hotel. The disease has an insidious onset and a long incubation period (4 to 14 days). Symptoms progress slowly but steadily in the first few weeks. Fever, cough, and severe shortness of breath are the major symptoms and diarrhea occurs in up to a quarter of cases. There are no definitive laboratory findings but lymphopenia (low lymphocyte count in blood) and increased lactate dehydrogenase (LDH) are frequent. The virus can be grown from numerous samples but most routinely from respiratory secretions and stool. Of course, it is not advisable for clinical laboratories to attempt to culture the virus unless they are equipped with special biosafety containment facilities. Although SARS-CoV is highly contagious during the disease, fortunately it is not contagious prior to symptoms or following recovery. This greatly facilitates control of spread of the disease and makes isolation of cases an effective public health measure. As is true of most severe viral respiratory diseases, treatment is supportive. Therapy with corticosteroids with or without ribavirin has been used without apparent efficacy [8].

## Paramyxoviridae

The Paramyxovidae family includes a large number of enveloped, negative-sense RNA viruses, many of which cause respiratory tract infections. These

include measles virus, respiratory syncytial virus (RSV), parainfluenza virus (PIV), and human metapneumovirus (HMPV). Measles virus causes a highly contagious childhood exanthem (rash syndrome) that has become rare in the United States since universal vaccination was instituted in the 1960s. Malnourished children are susceptible to serious measles pneumonia, which remains a major cause of childhood morbidity and mortality in the developing world. The focus of measles control is on maximizing the prevalence of vaccination. Measles is in fact an eradicable disease like smallpox and polio.

By the age of 2 virtually every child has been infected with RSV [15]. It is the leading cause of serious lower respiratory tract infection in infants and children [2]. Infants with lung or cardiac diseases are at increased risk for RSV infection, and bone marrow transplant recipients are susceptible to life-threatening interstitial pneumonia due to RSV [16, 17]. More recently, RSV has been recognized as an important cause of lower respiratory tract infection in the elderly [18, 19]. Infection early in life is associated with an increased likelihood of reactive airway disease later, possibly related to the predilection for RSV to induce the mechanisms involved in the development of asthma and allergy in children [20–22].

Despite significant effort, therapy of severe RSV infections is suboptimal and an effective vaccine remains elusive [15]. Ribavirin has reasonably antiviral activity against RSV and is used clinically. However, its use has declined because of side effects and difficultly of administration. There are significant ongoing efforts to develop an effective RSV vaccine and more effective antiviral drugs. Because each has its drawbacks and relative merits, it is likely that both vaccines and antivirals will have a place in the control of RSV-related disease.

Palivizumab (neutralizing monoclonal antibody) (Synagis) and respiratory syncytial virus immune globulin intravenous (RSV-IGIV) are used prophylactically to prevent severe lower respiratory tract infections caused by RSV in high-risk infants such as very young children with chronic lung disease and preterm infants [23, 24]. Recently, RSV immunoprophylaxis has been also recommended for infants and children with congenital heart disease [25]. Such immunoprophylaxis has been shown to be effective in that monthly administration of palivizumab during the RSV season results in a 45 to 55 percent decrease in the rate of hospitalization attributable to RSV. Other measures have been proposed to reduce the risk of RSV infection in high-risk infants such as avoidance of child-care centers during the RSV season [25].

Parainfluenza viruses types 1, 2, 3, and 4 (PIV-1, PIV-2, PIV-3, and PIV-4) were isolated about 50 years ago from children with lower respiratory tract infections. PIV-3 is the most prevalent, especially in very young children, but most children are infected with all three viruses by 5 years of age. Whereas RSV is overall more likely to be associated with severe lower respiratory tract infection (LRT) than PIV, the overlap of the clinical spectrum between the two viruses is large. Reinfection of children and adults with RSV and PIV are common because immunity is short-lived. Prior infection, however, may

reduce the severity of subsequent infections. There is no specific antiviral, vaccine, or immunoprophylaxis for PIV.

In 2001 Osterhaus and colleagues isolated a new virus from patients with acute respiratory tract infections [26]. The virus, named human metapneumovirus (HMPV), is a paramyxovirus virus related to avian pneumoviruses. Since that time investigators throughout the world have isolated HMPV from many patients with acute respiratory tract infection. Various studies have shown that HMPV is only occasionally isolated from asymptomatic children but accounts for a significant percentage of upper respiratory tract infections. Recently, it was shown to be a significant cause of lower respiratory tract infections during the first year of life [27]. Overall, HMPV causes a spectrum of disease very similar to RSV [27].

## Orthomyxoviridae

Influenza viruses are members of the large Orthomyxoviridae family of viruses. Orthomyxoviruses are enveloped, negative-sense RNA viruses with genomes that are composed of multiple RNA molecules called segments. Influenza A and B viruses cause a devastating amount of human morbidity and mortality throughout the world. Influenza A viruses are classified based on the type of hemagglutinin (H1 to H15) and neuraminidase (N1 to N9) they contain. Combinations of these two surface glycoproteins define each subtype. Outbreaks of human disease have been limited to combinations of H1, H2, or H3, together with N1 or N2. However, all subtypes have been found in viruses isolated from aquatic birds, which are the natural reservoir for influenza viruses. Hemagglutinin binds to sialic acid-containing receptors on respiratory epithelial cells. Neutralizing antibody to influenza is directed to hemagglutinin and thus it is the most critical immunogen in vaccines.

Influenza viruses, like all RNA viruses, have a penchant for rapid mutation. New strains emerge each year as a result of mutations in surface glycoproteins, a process referred to as antigenic drift. A major change in the hemagglutinin and neuraminidase combination is referred to as antigenic shift. This occurs via a reassortment of the viral genome segments when two influenza viruses infect the same host and genomic segments are exchanged. Pandemic outbreaks occur when reassortment involves human and animal influenza viruses that have not recently circulated in the population. Transmission of avian strains containing H5, H7, or H9 directly to humans followed by adaptation to human-to-human transmission may be another source of pandemic strains [27a]. Worldwide pandemics such as occurred in 1918, 1958, and 1968 have received much attention due to the large number of deaths that occurred during those years. However the number of deaths that have occurred due to influenza viruses during interpandemic years is still substantial. Each winter 10 to 20 percent of the U.S. population is infected with influenza, and there are tens of thousands of deaths and over 100,000 hospitalizations [29].

Influenza virus in healthy persons is usually a tracheobronchitis. Pneumonia and other complications are rare. Serious complications from influenza virus infection occur more frequently in pregnant women, the very young, the very old, and among persons with chronic diseases or immunodeficiency. Pneumonia, both primary viral and secondary bacterial, can lead to respiratory failure that is the most common cause of death related to influenza [30]. Myocarditis, pericarditis, and rhabdomyolysis are rare complications of influenza virus infection. The mechanism of these pathological complications is not clear but may be due to cyokines rather than a direct result of invasion of heart or muscle tissue [31].

Two types of influenza vaccines have been approved for use. The formaldehyde-inactivated intramuscular vaccine has been in use for several decades and the live, attenuated intranasal vaccine was approved in 2003. Both vaccines are trivalent in that they consist of two strains of influenza A virus and one strain of influenza B virus. The strains used in the vaccines are based on the recommendation of a Food and Drug Administration (FDA) Advisory Panel that meets each February, and its decision is based on worldwide surveillance data.

Two neuraminidase inhibitors were recently approved as antiviral therapy for influenza (see below). They are clearly effective at reducing symptoms especially if taken early in the course of the disease. They are very expensive, and their impact on management of influenza will depend on more rapid and point-of-care diagnostic methods to avoid overuse. Recently, it was shown that postexposure prophylaxis of household contacts of those with influenza reduces the secondary spread of influenza in families when the initial household case is treated [32].

## Bunyaviridae

The Bunyaviridae is a large family of enveloped, negative-sense, segmented RNA viruses that contains five genera. Only one of these genera, *Hantavirus*, contains viruses that cause respiratory disease in humans. A number of hantaviruses are known to cause a recently recognized severe respiratory syndrome called Hantavirus cardiopulmonary syndrome or Hantavirus pulmonary syndrome (HPS) [33, 34]. Hantaviruses are distinct from other members of the Bunyaviridae family in that they are not transmitted by arthropod vectors but are rather transmitted directly from aerosolized excreta of rodents, which are chronically and asymptomatically infected. HPS was first recognized after an outbreak in 1993 in the Four Corners region at the intersection of New Mexico, Arizona, Colorado, and Utah. A cluster of fatal cases of pneumonia in young adults was eventually linked to a newly discovered *Hantavirus* called Sin Nombre virus (SNV), which was found to be enzootic in the deer mouse (*Peromyscus maniculatis*). Since then, retrospective studies have identified various hantaviruses as the cause of previously unrecognized outbreaks and a number of subsequent outbreaks have been described

throughout the Western Hemisphere. Fortunately person-to-person transmission from HPS patients does not occur, but in a large outbreak in Argentina caused by the Andes hantavirus, there was evidence of person-to-person transmission [35]. In the United Sates there have been over 300 cases of HPS with a case fatality rate of 30 to 40. Cases occur throughout the year with most cases in spring and summer. The illness usually begins with a fever, myalgias, and cough. After several days to a week these symptoms progress to shortness of breath, a rapid onset of pulmonary edema with hypoxia, and, in some cases, shock related to poor cardiac output and arrythmias. Death can occur quickly after pulmonary symptoms develop, especially if mechanical ventilation is not instituted rapidly. Ribavirin, which has antiviral activity in cell culture and appears to demonstrate efficacy in animal models, unfortunately did not look promising in an uncontrolled clinical trial [36].

## Adenoviridae

Adenoviruses are ubiquitous nonenveloped DNA viruses that cause nonserious infections in immunocompetent individuals. Transplantation recipients and other immunocompromised patients, however, are susceptible to serious adenovirus respiratory tract infections including pneumonia. Treatment options are limited to ribavirin or cidofovir, each of which is associated with significant toxicity.

## Herpesviridae

The Herpesviridae family consists of many large enveloped DNA viruses that include eight distinct types of viruses that cause human disease. Despite similarity in structure and genetic organization, the herpesviruses have a wide spectrum of biological behavior and cause a wide spectrum of disease. The name (Latin, *herpes* = creep) is an anachronism related to the skin eruptions that occur with infection by several members of the family. Herpesviruses are not among the most common causes of viral respiratory tract infections. Many herpesviruses, however, can infect the respiratory tract, and in particular clinical settings several herpesviruses can cause serious pneumonia [37].

Varicella-zoster virus (VZV) causes both chicken pox (a primary infection in children) and shingles (reactivation from a latent infection) that occurs primarily after the fourth or fifth decade of life. Adults who escaped infection with VZV during childhood are at risk for a serious VZV infection as an adult. Primary infection with VZV in adults is frequently associated with a severe interstitial pneumonia that can be life-threatening. The live attenuated VZV vaccine may reduce the number of susceptible adults, but it is unclear whether the vaccine provides the life-long immunity that is associated with a natural infection. VZV, although less susceptible to acyclovir than herpes simplex 1 and 2 viruses, is inhibited by acyclovir and its derivatives.

Another serious pneumonia caused by a herpesvirus is the often fatal human cytomegalovirus (HCMV) interstitial pneumonia. HCMV generally causes an asymptomatic infection of immunocompetent individuals but is a major problem among immunocompromised patients such as those with acquired immunodeficiency syndrome (AIDS), patients undergoing cancer chemotherapy, and, most especially, organ and bone marrow transplant patients. Improved therapy for AIDS, more selective cancer therapy, and better agents for controlling graft rejection have all reduced the severity and prevalence of HCMV-associated infections, but it remains a significant clinical problem. Several different drugs are available to treat HCMV infections, but resistance to these has been an ongoing issue.

## 6 MAJOR CLINICAL SYNDROMES

Infection of the respiratory tract can be conveniently divided into upper and lower respiratory tract infections, or URI and LRI, respectively. URI and LRI can be further characterized by certain clinical syndromes that are associated with infection of particular anatomical regions of the respiratory tract. Below is a brief overview of these syndromes. Table 1 includes a list of most of the viruses associated with each of these syndromes. It must be borne in mind, however, that respiratory infections are often a continuum in which multiple anatomical sites are usually involved in the infection.

### Common Cold

The term *cold* refers to an acute, self-limited syndrome of the upper respiratory tract especially the nasal passages, paranasal sinuses, and the pharynx. Colds are one of the most common illnesses and are caused by a variety of RNA viruses. Adults average 2 to 3 colds per year and children 6 to 8 colds per year [38]. Symptoms include rhinitis (runny nose, nasal stuffiness, sneezing, etc.) and pharyngitis (sore throat) often associated with chills and fever. Other less frequent symptoms include cough, hoarseness, headache, fatigue, and malaise. Findings on physical exam include nasal discharge and pharyngeal erythema but are generally nonspecific. Colds are generally self-limited, lasting approximately 9 to 10 days. Rhinoviruses cause 40 to 50 percent of colds, and the remaining are caused by a variety of RNA viruses including coronaviruses, enteroviruses, PIV, RSV, and influenza A and B viruses.

Transmission of common cold viruses occurs by direct inoculation of virus into the upper respiratory tract and infection of ciliated nasal epithelial cells [39]. Infiltration of polymorphonuclear leukocytes then follows probably as a result of cytokine production. The number of infected cells is not high, and there is little cellular damage that is directly due to viral replication. Most of the symptoms are attributable to the inflammatory mediators triggered by the virus. Interestingly, in contrast to allergic rhinitis, histamine does not appear

**TABLE 1 Association of Specific Viruses with Respiratory Syndromes**

| Viruses | Common Cold | Pharyngitis | Tracheobronchitis | Croup[m] | Bronchiolitis[n] | Pneumonia[p] |
|---|---|---|---|---|---|---|
| Influenza[a] | Infrequent[j] | Common | Common | Common | Common | Common |
| RSV[b] | Infrequent | Infrequent | Common | Major | Common[o] | Common |
| PIV1-3[c] | Infrequent | Common | Common | Common | Major | Infrequent |
| HMPV[d] | Infrequent | Common | Common | Common | Common | Infrequent |
| Measles | | | Infrequent | Infrequent | | Infrequent[q] |
| Rhinovirus | Major | | | | | Infrequent |
| Enterovirus | Common | Common | | | Infrequent | Infrequent |
| Coronavirus | Major | Infrequent | | | Infrequent | Infrequent |
| HIV[e] | Infrequent[k] | | | | | |
| Adenovirus | | Infrequent | Infrequent | Infrequent | Infrequent | Infrequent |
| VZV[f] | | | | | | Infrequent |
| CMV[g] | Infrequent[l] | | | | | Infrequent[r] |
| HSV[h] | Infrequent | | | | Infrequent | Infrequent |
| EBV[I] | Infrequent[l] | | | | | Infrequent |

[a] Influenza viruses A and B. The relative importance varies from year to year.
[b] Respiratory syncytial virus.
[c] Parainfluenza viruses types 1, 2, and 3. For purpose of this table differences among the three types are not considered.
[d] Human metapneumovirus.
[e] Human immunodeficiency virus.
[f] Varicella-zoster virus.
[g] Cytomegalovirus.
[h] Herpes simplex virus.
[i] Epstein–Barr virus.
[j] Infrequent as an overall cause, but may be a common cause in certain clinical circumstances; see specific footnotes.
[k] As a component of primary HIV infection.
[l] As a component of infectious mononucleosis.
[m] Acute laryngotracheobronchitis of children.
[n] In children.
[o] RSV is a major cause of croup in infants.
[p] In immunocompetent patients. For immunocompromised patients see specific footnotes.
[q] Measles is a common cause of pneumonia in malnourished patients in the developing world.
[r] CMV is a common cause of pneumonia in organ transplant recipients.

to play a significant role in the pathophysiology of colds [40, 41]. Complications of colds include secondary bacterial infections of the paranasal sinuses and middle ear and exacerbations of asthma and emphysema. Viruses that cause the common cold are also implicated in otitis media (inner ear inflammation) especially in children [42, 43].

### Pharyngitis

Inflammation of the pharynx, or pharyngitis, is commonly referred to as a sore throat. Viruses cause about one third of cases of acute pharygitis, bacteria cause another one third, and the remaining one third are of unknown etiology. Viral pharygitis primarily occurs as part of a larger symptom complex such as the common cold, influenza, infectious mononucleosis, and primary infection with human immunodeficiency virus (HIV) [44].

### Laryngitis

Acute laryngitis, or inflammation of the vocal cords, is commonly seen in primary care settings. It generally presents as hoarseness and a painful cough. As with pharyngitis, it often accompanies other URI symptoms such as cold symptoms. All the major respiratory viruses have been implicated in the etiology of acute laryngiits.

### Acute Laryngotracheobronchitis (Croup)

Croup is a viral infection of the respiratory tract in young children that results in inflammation of the subglottic region. It leads to dyspnea (shortness of breath), a so-called seal's bark cough, and inspiratory sounds referred to as stridulous notes or simply as stridor. This characteristic sound is pathognomonic to the caregiver familiar with this syndrome. The origin of the term *croup* is not certain, but it is thought to have been derived from an Anglo-Saxon term that meant to cry out in a shrill voice. Although the term *croup* has been used for a number of respiratory infections associated with compromised airflow such as diphtheria, today it is used exclusively for viral laryngotracheobronchitis. Croup is relatively common in young children, with most cases occurring between 3 months and 3 years of age. PIV, most commonly PIV-3, and RSV are the most commonly implicated pathogens.

### Tracheitis and Tracheobronchitis

Tracheitis rarely occurs as an isolated infection and is usually associated with URI and/or bronchitis. It is usually associated with severe pain in the chest due to upper airway damage and inflammation. Infection of the trachea and large bronchi leads to impaired mucociliary clearance mechanisms that, in turn, predisposes to bacterial secondary infection [45].

## Bronchiolitis

Bronchiolitis is an acute lower respiratory tract viral infection of young children that is characterized by acute onset of wheezing associated with cough and respiratory distress. The vast majority of cases are caused by RSV or PIV. Other respiratory viruses account for the remainder of cases, although a small percentage is associated with the bacterium *Mycoplasma pneumonia*. Viral replication in the bronchial epithelial cells precipitates an inflammatory reaction with infiltration of neutrophils, basophils, and eosinophils and release of inflammatory mediators. Inflammatory changes such as edema and cellular necrosis with sloughing lead to obstruction of small airways, which accounts for the wheezing and hyperinflation of the lungs. Wheezing occurs commonly, but the inflammatory process precipitated in virus-induced bronchiolitis is not the same as occurs in allergic reaction or asthma. Distinguishing viral bronchiolitis from the first episode of asthma or from the many other causes of wheezing requires clinical acumen. Therapy is primarily supportive with oxygen, and, if necessary, mechanical ventilation for hospitalized infants. Bronchodilators and corticosteroids are used frequently, but their efficacy has not been proven. Ribavirin administered by aerosol is reserved for those at high risk for severe infection. Most normal children recover fully from bronchiolitis, but there is an increased risk of subsequent lower respiratory tract disease and reactive airway disease especially in those with bronchiolitis early in life.

## Pneumonia

Pneumonia is an infection of the lung parenchyma that leads to inflammatory changes that compromise alveolar gas exchange. Obviously, pneumonia can be the most serious of the respiratory tract infections caused by viruses because it can lead to hypoxia that, if persistent, may result in widespread organ dysfunction. There is a wide spectrum of clinical presentations of pneumonia depending on the immune status of the patient and the specific viral pathogen. Serious viral pneumonia in healthy older children and adults is relatively uncommon. Virtually all the respiratory viruses have been associated with pneumonia in healthy patients, but the morbidity is generally low. The SARS outbreak was particularly unusual and worrisome because many of the fatalities occurred in healthy young adults. Another exception is the hantavirus pulmonary syndrome, which has a high mortality rate in young adults. Fortunately, the number of cases to date has been small. Influenza virus causes pneumonia especially during significant pandemics. One of the most serious consequences of influenza is secondary bacterial pneumonia. Any healthy adult who did not have chicken pox as a child, or who was not vaccinated, is at risk for varicella pneumonia, which can be serious.

The very young, the elderly, those with chronic cardiac disease or pulmonary disease, and immunocompromised patients are at higher risk for certain types of viral pneumonia that can be associated with significant mor-

bidity and mortality. RSV bronchopneumonia can be very serious in preterm infants. In highly immunocompromised patients, such as bone marrow transplant recipients, RSV pneumonia is associated with a high rate of mortality. Human cytomegalovirus is another frequent cause of life-threatening pneumonia in organ transplant patients. Virtually the entire spectrum of human viral pathogens has been described in association with pneumonia in severely immunocompromised patients [46, 47].

The pathogenesis of viral pneumonia is complex and has differential features depending on the host and virus. Pneumonia may occur in a situation where the virus enters the body via the respiratory tract and remains localized, it may extend from the lungs into a systemic viral infection, or it may have entered the lungs as a consequence of a systemic infection. Obviously the pathogenesis will vary with each of these scenarios. In general, viral pneumonia involves a combination of viral-mediated cell death of respiratory bronchiolar epithelial cells, hypersecretion of mucus by mucus-secreting cells, and inflammatory changes such as infiltration of neutrophils and monocytes, edema, and focal hemorrhage. Alveoli can show marked changes including capillary edema, hemorrhage, and formation of hyaline membranes, which explains the poor gas exchange that develops. Viral pneumonia is a diffuse process that generally involves the entire lung and on radiographs appears as a diffuse reticular pattern. This is in contrast to bacterial pneumonia, which is often limited to individual lobes of the lung. The relative prominence of the various pathological changes varies with the virus, the host, and the severity of the pneumonia. In HPS, for example, the most dramatic process is severe alveolar capillary leak that leads to significant fluid buildup in the air sacs. This in turn causes a serious respiratory distress-like syndrome characterized by dramatic hypoxia.

## 7 DIAGNOSIS OF VIRAL RESPIRATORY INFECTION

The clinical presentation of any viral infection of the respiratory tract does not enable a specific diagnosis. There are, in general, no pathognomonic signs and symptoms that distinguish viral respiratory tract infections from bacterial infections or from other causes. The pattern of signs, symptoms, the radiographic (X-ray) pattern, and routine laboratory results can promote a high index of suspicion but do not lead to a specific etiology. Upper respiratory tract infections are characterized by rhinorrhea (runny nose), sneezing, cough, and sore throat with or without fever. Lower respiratory tract infections are most often characterized by a nonproductive (i.e., dry) cough and shortness of breath, associated with generalized symptoms of fever, malaise, and myalgias. The physical findings include particular auscultatory (stethoscope) sounds such as rales, crackles, and wheezing. Cyanosis (blue-colored skin of the digits, lips, etc.) may be observed when hypoxemia is present. For certain viruses such as measles, there may be a characteristic rash.

The history and the clinical setting are critical factors in the diagnostic evaluation. Where the patient lives, whether there have been other cases, whether there has been known contact with an individual already diagnosed, whether the infection occurs during the season when certain viral infections are prevalent, for example, all have a bearing on the likelihood of a particular diagnosis. In addition, the history combined with the clinical presentation will help direct the diagnostic effort.

Diagnostic tests for viral diseases fall in to four categories: serology, antigen detection, viral culture, and nucleic acid testing. Serology involves testing the patient for antibodies to a specific agent. It generally involves measuring acute (during symptoms) and convalescent (after resolution) antibody titers. A rise in titer (usually fourfold) is consistent with recent infection but obviously only provides a retrospective diagnosis. Recently, measurement of specific IgM antibodies has allowed the diagnosis to be made in the acute setting. IgM tests are not available for all pathogens and are of variable sensitivity and specificity. Antigen detection has become the modality of choice for rapid diagnosis of acute viral infections. The most common method used involves obtaining cells from an appropriate site such as a scraping of the posterior pharynx or a bronchoalveolar lavage specimen. Cell-associated specific viral antigens are then detected by direct immunofluoresence, which involves the use of a fluorescently labeled monoclonal antibody directed against a viral protein. Such direct fluorescent antibody (DFA) tests are available for RSV, PIV, adenovirus, and influenza and have high specificity (>90 percent) and reasonably good sensitivity (70 to 90 percent). Viral culture is the most definitive modality for making the diagnosis, but when performed by traditional methods can take 3 to 10 days. Newer methods of culture, especially when combined with DFA, can reduce the time to detect positive cultures to 1 to 3 days. Nucleic acid detection (NAD) allows for the detection of even trace amounts of viral genome in patient samples. The most commonly used methods involve amplification of the viral nucleic acid by polymerase chain reaction (PCR) either directly for DNA viruses or following reverse transcription for RNA viruses. Detection of the amplified nucleic acid can be performed by a variety of methods, but fluorescent-based real-time detection methods are supplanting older methods. The cost and complexity of these tests, however, has limited their routine use.

A critical issue for any diagnostic evaluation is having a clear idea of what tests are appropriate to order. Use of any of the specific tests requires a certain index of suspicion because ordering a panel of tests is prohibitively expensive as a routine mode of operating. Even viral culture is better performed when certain agents are considered in advance because the type of cell lines and the conditions of culturing can affect the detection of particular viruses. The ultimate goal of viral diagnostic testing is the use of multiplexing, that is, the ability to detect many agents from the same specimen using one test. Further supporting this approach is the fact that up to 15 percent of respiratory tract infections are caused by more than one agent (mixed infection). Effective antiviral

therapy requires effective diagnostic modalities. Rapid, point-of-care, inexpensive, multiplexed, sensitive, and specific diagnostic tests will enable antiviral therapy to be more effective, especially for acute viral diseases such as most viral respiratory infections.

## 8 EXISTING ANTIVIRAL THERAPY

Significant progress has been made in the development of antiviral drugs. Just 30 years ago, there were no FDA-approved antiviral drugs. Today, there are over 40 such drugs [48]. Still, the applicability of these drugs continues to be limited. More than half the FDA-approved antiviral drugs, for example, are for use in treating infections caused by human immunodeficiency virus (HIV), and most of the remaining drugs are for use in treating herpesvirus infections. The first oral antiviral was a drug for herpes, acyclovir, developed by the laboratory of Gertrude B. Elion at Burroughs-Wellcome in the 1960s. Acyclovir is a nucleoside analog that fits into the binding pocket of herpes DNA polymerase and terminates the enzyme's growing nucleic acid chain. High-dose acyclovir is efficacious for VZV pneumonia. Another nucleoside analog, ganciclovir, as well as a nonnucleoside DNA polymerase inhibitor, foscarnet, are available to control HCMV pneumonia in immunocompromised patients. Their distinct toxicity and resistance profiles make them a useful pair during the management of these difficult pneumonias.

There are only a handful of small-molecule antiviral drugs available for treatment of the RNA viruses that cause the majority of respiratory tract infections. Consequently, for many RNA viruses, there are only limited, if any, therapeutic options available. Two drugs, amantadine (Symmetrel) and its derivative rimantadine (Flumadine), have been available for many years for prophylactic use in patients at high risk for influenza virus such as residents of nursing homes [49, 50]. These drugs interfere with proton transport of the viral transmembrane protein, M2, but are only effective against influenza A virus.

Recently, two drugs, oseltamivir (Tamiflu) and zanamivir (Relenza), that are inhibitors of the influenza neuraminidase enzyme were approved for treatment of influenza A and B [28, 51]. As expected, their efficacy depends on initiation of therapy as soon as possible. They have been shown to be effective when used prophylactically during an outbreak [32]. This so-called post-exposure prophylaxis may be the most effective use of these drugs. Neither the M2 or neuraminidase inhibitors have been studied in cases of severe pneumonia.

Ribavirin is the only available antiviral drug for treatment of other respiratory viruses [52–56]. It has been used in a small-particle aerosol formulation to treat severe bronchiolitis caused by RSV. The efficacy of ribavirin in RSV bronchopneumonia is controversial, and its use has been limited to treatment of high-risk infants. The increased use of prophylactic monoclonal antibody has contributed to the reduced use of ribavirin. There is limited data on the

efficacy of ribavirin in other viral respiratory diseases. It may be efficacious for PIV and measles infections, but it has not shown to be of value in HPS [36]. There remains the potential for combining ribavirin with other antivirals. For example, ribavirin has limited antiviral effect on hepatitis C virus (HCV) but definitely adds to the therapeutic potency of interferon-α. Whether this approach can be successfully applied to respiratory viral infections remains to be seen. Combination of biological agents such as mAbs and small-molecule antiviral drugs is also an intriguing possibility. The concept of using neutralizing monoclonal anti-RSV antibodies to decrease the long-term morbidity associated with RSV infection in children by reducing RSV replication has also been proposed [57].

## 9 CHALLENGES OF ANTIVIRAL THERAPY FOR RESPIRATORY VIRUS INFECTIONS

There are many difficulties associated with the development of antiviral therapies. The manner in which viruses replicate and their close association with cellular processes during their life cycle makes finding agents with a selective effect on the virus difficult. In addition, there are often multiple strains, serotypes, and sequence variants of any given virus (quasi-species). This inherent variation is related to the high frequency of mutation of viruses, which leads to the rapid selection of viruses that are drug resistant.

Another difficulty is the uncertain correlation between the level of viral replication and the extent of clinical disease. Generally, it is safe to say that viral replication is necessary for most viruses to cause disease, but the amount and duration of viral replication may not correlate temporally with the degree of pathology. Therefore, a therapeutic intervention that focuses on inhibition of viral replication may be inadequate to ameliorate either symptoms or pathology unless it is instituted either before infection or within a short period after infection. As an example, despite the fact that influenza is characterized by significant virus replication in respiratory epithelial cells, there is now growing evidence that the cytokines produced early at the site of infection mediate many of the clinical and pathological manifestations [58]. These cytokines include interferon-α (IFN-α), tumor necrosis factor-α (TNF-α), interleukin-1 (IL-1), interleukin-6 (IL-6), interleukin-8 (IL-8), and various chemokines. Studies of respiratory secretions of experimentally infected humans have shown that peak cytokine levels directly correlated with virus replication and disease. However, cytokines are clearly not the only factor contributing to disease, and they are also likely to be essential for resolution of the infection [58].

Both immunologic and nonimmunologic factors have been implicated in the pathogenesis of RSV-induced disease. RSV bronchiolitis has been proposed to result from production of Th2-type cytokines, although this clearly cannot be the total explanation. Interferon-γ may contribute to RSV-induced

wheezing, possibly through induction of leukotriene release. Pulmonary surfactant has recently been thought to be an important factor in RSV-mediated severe bronchiolitis. Airway damage and hyperreactivity induced during RSV infection can alter long-term lung function and may lead to chronic asthma. While the mechanisms underlying hyperreactivity of the airways are unclear, in a mouse model airway hyperreactivity response was associated with inappropriate production of cytokines, particularly IL-13 [59]. Chemokines have been also thought to be an important contributor to the symptoms that arise during RSV infection, but their role has not been fully elucidated. Mice treated with anti-RANTES (CCL5) antibody demonstrated significant decreases in airway hyperreactivity, suggesting that RANTES (CCL5) is an important mediator of the pathophysiological responses seen in RSV infection [60].

Inflammation and mucus overproduction are partially responsible for RSV-induced disease in infants. In other studies using a murine model, it was shown treatment with anti-CXCR2 antibody resulted in decreased airway hyperreactivity and decreased mucus in the bronchoalveolar lavage fluid. Studies with CXCR2 (–/–) *knockout* mice also demonstrated the important of CXC chemokines in RSV-induced mucus production and airway hyperreactivity [61]. There is also evidence that the nuclear transcription factor NF-κB controls the expression of many genes involved in the inflammatory and immunomodulatory processes that occur during RSV infection. As a possible central activator of inflammatory and innate immune responses to RSV, NF-κB may be a potential target [62]. Additional therapeutic strategies will be promoted by better understanding of the pathogenesis of RSV-induced disease and other viral respiratory tract infections [63].

## 10 STRATEGIES FOR DEVELOPMENT OF ANTIVIRAL THERAPY

There are a variety of approaches that can be applied to the development of antiviral therapies. Cytokines, cytokine inhibitors, monoclonal antibodies directed against the virus, and RNA interference (RNAi)-based agents are some of the modalities under development [64]. Small molecules that inhibit viral replication generally receive the most attention. Such inhibitors are usually focused on viral targets such as viral enzymes (e.g., RNA polymerase, protease, helicase, etc.) or other viral proteins (e.g., fusion proteins). Small molecules that target viral RNA are a promising possibility but have received limited attention [65].

The replication of viruses can be described in terms of a series of key processes that occur sequentially starting with viral attachment to the cell and ending with the exit of new progeny virus from the cell. Each of these events involves the interaction of essential viral proteins with cellular machinery, and each specific activity involved in this process is a potential target of an

inhibitory molecule. The best targets are those that are the most genetically and biochemically tractable. Virally encoded proteins that have an easily measurable activity fit this criterion the best. Reverse genetics can be used to prove that the particular viral protein is required for viral replication, which validates it as a target for an effective antiviral compound. This relatively straightforward validation process contrasts sharply with other areas of drug discovery in which target validation is often one of the most challenging tasks.

The replication of all viruses depends on various viral enzymes and cofactors encoded by the viral genome. Many of these enzymes are therefore excellent targets for antiviral compounds. For RNA viruses these enzymes include RNA-dependent RNA polymerases and their cofactors that replicate the viral genome and synthesize messenger RNA (mRNA), RNA helicases that unwind duplex regions of RNA during replication, and proteases that are involved in the proteolytic processing of viral proteins to their active form. For DNA viruses DNA polymerases and DNA helicases are key targets. Inhibitors that target the virus–cellular interaction at the cellular membrane also represent an effective strategy. Such inhibitors can target either the viral attachment protein or the cellular receptor.

In addition to viral proteins, cellular proteins that are critical for the replication of particular viruses may represent valuable targets for antiviral development. Cellular enzymes that have been shown to be potential targets of antiviral compounds include inosine monophosphate dehydrogenase (IMPDH), *S*-adenosylhomocysteine (SAH) hydrogenase, cytidine triphosphate (CTP) synthetase, and α-glucosidase [66, 67]. It is generally thought that inhibitors of host enzymes would be associated with more toxicity than inhibitors of viral targets. This assumption is not necessarily true for several reasons. First, much drug-related toxicity is unrelated to the mechanism of therapeutic action. Second, the vast majority of medicinals target cellular enzymes and most are safe even when taken chronically. Third, many antiviral drugs that target viral enzymes are associated with significant toxicity (e.g., HIV protease inhibitors). Finally, cellular enzymes are likely to exist that are essential for viral replication, but their activity may not be essential for health of the cell because of redundancy. Such enzymes would, of course, be excellent targets for antiviral drugs.

Screening programs to identify compounds with antiviral activity involve two general methods: *targeted* screening and *unbiased* screening. In the targeted approach, a particular biochemical target is chosen, and then inhibitors of the chosen target are screened. The chosen target is typically an enzyme or a receptor that is known or thought to be essential to viral replication. In the alternative approach, unbiased screening, inhibitors of viral replication are sought without a priori concern for the target. Unbiased screening necessarily involves the use of cell culture for virus replication. This is due, in part, to the fact that cellular targets are likely targets of many antiviral agents. Although cell-based screening has been used successfully throughout the drug discovery field, it has historically been problematic when screening for anti-

viral compounds because it requires inoculation of an infectious virus onto the cells, which complicates the process of screening large libraries of compounds.

Partial viral replication systems have been developed to circumvent the problems associated with cell-based cultures using whole viral systems. In partial viral systems, viral genomes lacking one or more genetic elements essential for complete replication are used to accomplish viral genomic replication without producing the infectious virus. This is particularly important for viral agents such as SARS-CoV, which is classified as biohazard level 3 (BL-3) and for BL-4 agents such as Ebola virus. A screening process that utilizes these incomplete viral genomes can identify inhibitors of any biochemical pathway involved in viral genome replication, transcription, and translation. This allows for screening with respect to multiple possible targets. These targets do not have to be known, thus making the screening process unbiased. In addition, the targets are prevalidated, given that inhibition of RNA replication is directly relevant to the viral disease process. Screening with partial viral replication systems additionally is advantageous because complex viral replication pathways can be easily monitored by measuring levels of viral RNA or expression of a reporter gene (e.g., luciferase, β-galactosidase, secreted alkaline phosphatase, green fluorescent protein, etc.) present in the replicon or defective genome. The major drawback of unbiased, cell-based screening, whether it involves infectious virus or partial genomes, is that the target of an identified antiviral molecule is not immediately known. Efforts to identify the target and ultimately determine the mechanism must be undertaken to enable the process of optimizing the lead compound.

## 11 DEVELOPMENT OF ANTIVIRALS AGAINST RESPIRATORY VIRUSES

The approval of two inhibitors of the influenza virus neuraminidase (NA), oseltamivir (Tamiflu) and zanamivir (Relenza), has opened the door for the development of small-molecule inhibitors of other respiratory viruses. Unfortunately, the door remains only partially open and progress has been slow for a variety of reasons. Another potent NA inhibitor being developed by BioCryst (Birmingham, AL) failed to demonstrate significant efficacy in phase III trials. Many other NA inhibitors have been described that show potent in vitro activity against influenza virus. Whether they will be developed is unclear. In addition to neuraminidase inhibitors, compounds are under development that target other influenza virus replication processes. Interaction of the influenza virus with the cell membrane is an important process during replication of most enveloped viruses. Ongoing work has identified a number of compounds that inhibit interaction of the viral envelop with the cellular membrane. A polyoxometalate (PM523) is a potent inhibitor of influenza, respiratory syncytial virus, and measles virus and acts by inhibiting fusion between the viral envelope and the cellular membrane. Resistance to PM523

mapped to sequences that encoded amino acids in hemagglutinin (HA) that are in the interface of the trimers of HA. Another potential target of anti-influenza virus is the mRNA capping process. Cap formation of influenza mRNA utilizes the 5′-mGpppXm of host mRNA. Inhibitors of cap formation have been described that act on the PB2 viral protein, which is a component of the viral RNA polymerase complex. A metabolic product of ribavirn, 1,2,4-triazole carboxamide (T-CONH2) is inhibitory for influenza replication in vitro and in vivo [68]. Endonuclease inhibitors also have been described.

Pleconaril [Picovir (ViroPharma, Exton, PA)], which binds to the capsid of many picornaviruses including rhinoviruses, is an inhibitor of virus uncoating [69, 70]. Clinical trials showed efficacy against the common cold, but unfortunately it was not approved by the FDA, which sought additional studies on drug interactions [71]. Cell culture-based transdominant genetic techniques were used to discover a peptide inhibitor of human rhinovirus that acts on intercellular adhesion molecule 1 (ICAM-1) [72, 73]. Inhibitors of other rhinovirus targets such as the 3C (e.g., AG7088) and 2A proteases and the RNA polymerase are also under development [74–77].

The increasing recognition of the significant disease burden caused by RSV has stimulated activity in anti-RSV discovery. At one time as many as 20 anti-RSV R&D programs were ongoing in the pharmaceutical/biotechnology industry. Unfortunately, this activity has not borne fruit and has not been sustained. Drug discovery efforts against RSV have primarily focused on fusion inhibitors that target the RSV F protein. One such fusion inhibitor (V-14637; ViroPharma) showed potent activity in preclinical studies and underwent early clinical trials in an inhalation formulation, but its development was suspended by the company. Other fusion inhibitors have been described that have potent anti-RSV activity in cell culture and in animal models. Unfortunately, clinical development is not being pursued. Recently, representatives from Arrow Therapeutics (London, GB) have described an inhibitor (A60444) that apparently targets the nucleocapsid (N) protein. Its biochemical mechanism of action has not been described. It is undergoing phase I clinical trials.

In addition to small molecules other approaches to antiviral therapy are being studied. These include antisense oligonucleotides and RNA interference [64, 78–80]. Other investigators have searched for RSV adhesion receptors on mammalian cells as possible antiviral targets. Soluble human lectins, complex polysaccharides, and a low-molecular-weight selectin antagonist, TBC1269, were used to characterize and isolate the RSV receptor on a human epithelial cell line (Hep2 cells). Selectin antagonists, fucoidan and TBC 1269, inhibit RSV infection both in vitro and in a mouse model of infection. Annexin II has been proposed as a potential RSV receptor on Hep2 cells. Recombinant annexin II binds to RSV G protein, heparin, and plasminogen, and the binding is inhibited by a selectin antagonist, TBC1269. These workers have suggested that inhibitors of annexin II could have potential in treating RSV infection [81].

There has been significant effort to develop an effective therapy for SARS-CoV. In a cynomolgus macaque model of SARS, pegylated interferon showed some promise as a potential treatment [82]. Calpain and β-D-$N^4$-hydroxycytodine have been shown to be inhibitors of SARS-CoV, but their clinical potential is unclear [83]. Rational design based on X-ray crystal structure has been used to design drugs for SARS-CoV 3CL protease.

## 12 ANIMAL MODELS

Animal models are an important component of antiviral development because they provide an indication of in vivo efficacy prior to initiating human clinical trials. However, there are few animal models for human respiratory viruses and most of those that exist are suboptimal. Mouse models have many advantages for testing antivirals. The available genomic information, transgenic strains, immunodeficient strains, immunologic reagents, and the like make murine models of infection highly tractable. Unfortunately, mice are not fully permissive for many human respiratory viruses, and infection of mice with respiratory viruses does not always mimic the relevant human disease [84].

Significant effort has been devoted to the development of animal models of RSV infection using mice, cotton rats, guinea pigs, ferrets, hamsters, calves, and nonhuman primates [85, 86]. These models have been used to test vaccines and to determine the mechanism of vaccine-enhanced disease seen with a formalin-inactivated RSV vaccine [87]. They have also provided a means for testing efficacy and safety of prophylactic and therapeutic immunotherapy [57]. The value of these animal models for testing small-molecule therapeutic drugs, however, is less clear.

The cotton rat (*Sigmodon hispidus*) has been promoted as a useful model for several respiratory viruses including influenza, parainfluenza virus, and RSV [88–90]. After inoculation of the nares with RSV, animals develop pulmonary pathology. Unfortunately, the cotton rat is only "semipermissive" for RSV in that the recoverable virus never exceeds the innoculum. Nonhuman primates such as African green monkeys and Rhesus macaques have been used for studies of vaccines, monoclonal antibodies, and antivirals. The high cost and increasing lack of availability are major drawbacks to the use of primate models. Animal models must be used with the realization that RSV disease in humans has many pathophysiological features. The clinical manifestations depend upon age, genetic makeup, immunologic status, and underlying disease, and there is no single animal model that duplicates all the various forms of RSV disease [86].

A small-animal model is important for the development of new therapies for rhinoviruses. However, existing mouse models of infection are difficult to work with, and until recently mouse cell lines were thought to be generally nonpermissive for HRV replication in vitro. Recently, a virus of the minor-receptor group, HRV1B, was shown to be able to infect and replicate in a

mouse respiratory epithelial cell line (LA-4) more efficiently than in a mouse fibroblast cell line (L) [91]. In contrast the major-receptor group virus, HRV16, requires human intercellular adhesion molecule-1 (ICAM-1) for cell entry and therefore cannot infect LA-4 cells. However, transfection of in-vitro-transcribed HRV16 RNA resulted in the replication of viral RNA and production of infectious virus. Furthermore, these workers showed that expression of a chimeric ICAM-1 molecule, comprising mouse ICAM-1 with extracellular domains 1 and 2 replaced by the equivalent human domains, rendered the otherwise nonpermissive mouse respiratory epithelial cell line susceptible to entry and efficient replication of HRV16. These observations suggest that the development of mouse models of respiratory tract infection by major as well as minor group HRV are possible [91].

A significant effort has been applied to the development of a SARS-CoV animal model. A cynomolgus macaque model of SARS-CoV has been used by one group to test antiviral agents [82]. Murine models have also been studied, but their relevance to human infection is uncertain.

## 13 SUMMARY

There are significant challenges associated with the development of antiviral therapy for respiratory virus infections. These challenges fall into four broad categories: (1) virological, (2) pathophysiological, (3) clinical, and (4) economical. Viruses are obligate intracellular pathogens, and their intimate association with cells during their life cycles makes them very difficult to inhibit compared with extracellular or free-living pathogens such as most bacteria. Another problem is the high rate of mutation of viruses, especially RNA viruses. Thus they have a high predeliction for developing resistance to antiviral agents.

Another problematic issue is that the majority of the symptoms associated with viral respiratory tract infections are triggered by the virus rather than directly mediated by viral replication. After initial infection, respiratory viruses replicate rapidly, and the body responds with a strong innate and acquired immune response that triggers symptoms that can last weeks. Therefore, in order to have an impact on disease symptoms and pathology, therapy, even with an extremely potent inhibitor of viral replication, needs to be initiated within 24 to 48 h of onset of symptoms. This is obviously a very narrow treatment window. From the clinical standpoint this issue is problematic as well. Symptoms are not seen during the incubation period, and there is often a delay in clinical suspicion.

This situation results ultimately in a delay in definitive diagnosis. Consideration of therapy of viral respiratory tract infections, therefore, is not likely to be initiated in a timely manner. Finally, for a variety of reasons antiviral therapies are generally expensive, and, in the existing medical economic structure, an expensive antiviral therapy is not likely to be used presumptively prior

to when a definitive diagnosis has been made. All of these issues make it difficult to start therapy early at the time it is likely to be the most effective.

In summary, there has been significant progress made in the development of antiviral drugs during the past two decades. Most of the progress, however, has been made in the treatment of chronic viral diseases such as HIV-associated AIDS. Treatment of acute viral disease, including most respiratory viral infections, is fraught with many difficulties. Nevertheless, despite these difficulties, we are poised to enter a new era in antiviral therapy. Respiratory viruses cause an immense amount of human morbidity and mortality, and antiviral therapy is an important weapon to control these agents. The development of antiviral drugs to treat respiratory virus infections will be one of the great challenges of the twenty-first century.

## REFERENCES

1. Mackie, P. L. (2003). The classification of viruses infecting the respiratory tract. *Paediatr. Respir. Rev.*, *4*, 84–90.
2. Centers for Disease Control and Prevention (CDC) (1999). Update: respiratory syncytial virus activity–United States, 1997–98 season. *JAMA*, *281*, 127.
3. Arruda, E., Pitkaranta, A., Witek, T. J., Jr., Doyle, C. A., Hayden, F. G. (1997). Frequency and natural history of rhinovirus infections in adults during autumn. *J. Clin. Microbiol.*, *35*, 2864–2868.
4. Department of Health and Human Services (DHHS) (1992). *Current Estimates from the National Health Interview Survey, 1992*. DHHS, Hyattsville, MD, pp. 94–1517.
5. Monto, A. S., Sullivan, K. M. (1993). Acute respiratory illness in the community. Frequency of illness and the agents involved. *Epidemiol. Infect.*, *110*, 145–160.
6. Gerberding, J. L., Morgan, J. G., Shepard, J. A., Kradin, R. L. (2004). Case records of the Massachusetts General Hospital. Weekly clinicopathological exercises. Case 9-2004. An 18-year-old man with respiratory symptoms and shock. *N. Engl. J. Med.*, *350*, 1236–1247.
7. Clements, C. J., Strassburg, M., Cutts, F. T., Torel, C. (1992). The epidemiology of measles. *World Health Stat. Q.*, *45*, 285–291.
8. Wenzel, R. P., Edmond, M. B. (2003). Managing SARS amidst uncertainty. *N. Engl. J. Med.*, *348*, 1947–1948.
9. Richman, R. D., Whitely, R. J., Hayden, F. G. (Eds.) (2002). *Clinical Virology*. ASM Press, Washington, DC.
10. Flint, S. J., Enquist, L. W., Racaniello, V. R., Skalka, A. M. (2003). *Principles of Virology*. ASM Press, Washington, DC.
11. Harris, J. M. 2nd, Gwaltney, J. M., Jr. (1996). Incubation periods of experimental rhinovirus infection and illness. *Clin. Infect. Dis.*, *23*, 1287–1290.
12. Gwaltney, J. M., Jr., Hendley, J. O., Patrie, J. T. (2003). Symptom severity patterns in experimental common colds and their usefulness in timing onset of illness in natural colds. *Clin. Infect. Dis.*, *36*, 714–723.

13. Van Der Hoek, L., Pyrc, K., Jebbink, M. F., Vermeulen-Oost, W., Berkhout, R. J., Wolthers, K. C., Wertheim-Van Dillen, P. M., Kaandorp, J., Spaargaren, J., Berkhout, B. (2004). Identification of a new human coronavirus. *Nat. Med.*, *10*, 368–373.
14. Vijayanand, P., Wilkins, E., Woodhead, M. (2004). Severe acute respiratory syndrome (SARS): A review. *Clin. Med.*, *4*, 152–160.
15. Hall, C. B. (1999). Respiratory syncytial virus: A continuing culprit and conundrum. *J. Pediatr.*, *135*, 2–7.
16. Hertz, M. I., Englund, J. A., Snover, D., Bitterman, P. B., McGlave, P. B. (1989). Respiratory syncytial virus-induced acute lung injury in adult patients with bone marrow transplants: A clinical approach and review of the literature. *Medicine (Baltimore)*, *68*, 269–281.
17. Krinzman, S., Basgoz, N., Kradin, R., Shepard, J. A., Flieder, D. B., Wright, C. D., Wain, J. C., Ginns, L. C. (1998). Respiratory syncytial virus-associated infections in adult recipients of solid organ transplants. *J. Heart Lung Transplant*, *17*, 202–210.
18. Falsey, A. R. (1998). Respiratory syncytial virus infection in older persons. *Vaccine*, *16*, 1775–1778.
19. Han, L. L., Alexander, J. P., Anderson, L. J. (1999). Respiratory syncytial virus pneumonia among the elderly: An assessment of disease burden. *J. Infect. Dis.*, *179*, 25–30.
20. Sigurs, N., Bjarnason, R., Sigurbergsson, F., Kjellman, B. (2000). Respiratory syncytial virus bronchiolitis in infancy is an important risk factor for asthma and allergy at age 7. *Am. J. Respir. Crit. Care Med.*, *161*, 1501–1507.
21. Kneyber, M. C. J., Steyerberg, E. W., de Groot, R., Moll, H. A. (2000). Long-term effects of respiratory syncytial virus (RSV) bronchiolitis in infants and young children: A quantitative review. *Acta Paediatr.*, *89*, 654–660.
22. Papadopoulos, N. G., Psarras, S., Manoussakis, E., Saxoni-Papageorgiou, P. (2003). The role of respiratory viruses in the origin and exacerbations of asthma. *Curr. Opin. Allergy Clin. Immunol.*, *3*, 39–44.
23. Atkins, J. T., Karimi, P., Morris, B. H., McDavid, G., Shim, S. (2000). Prophylaxis for respiratory syncytial virus with respiratory syncytial virus-immunoglobulin intravenous among preterm infants of thirty-two weeks gestation and less: Reduction in incidence, severity of illness and cost. *Pediatr. Infect. Dis. J.*, *19*, 138–143.
24. Carter, B. S. (2000). Palivizumab (synagis): Counting "costs" and values. *Pediatrics*, *106*, 1168–1169.
25. Meissner, H. C., Long, S. S. (2003). Revised indications for the use of palivizumab and respiratory syncytial virus immune globulin intravenous for the prevention of respiratory syncytial virus infections. *Pediatrics*, *112*, 1447–1452.
26. van den Hoogen, B. G., de Jong, J. C., Groen, J., Kuiken, T., de Groot, R., Fouchier, R. A., Osterhaus, A. D. (2001). A newly discovered human pneumovirus isolated from young children with respiratory tract disease. *Nat. Med.*, *7*, 719–724.
27. Williams, J. V., Harris, P. A., Tollefson, S. J., Halburnt-Rush, L. L., Pingsterhaus, J. M., Edwards, K. M., Wright, P. F., Crowe, J. E., Jr. (2004). Human metapneumovirus and lower respiratory tract disease in otherwise healthy infants and children. *N. Engl. J. Med.*, *350*, 443–450.

27a. Nicholson, K. G., Wood, J. M., Zambon, M. (2003). Influenza. *Lancet*, *362*, 1733–1745.

28. Nicholson, K. G., Aoki, F. Y., Osterhaus, A. D., Trottier, S., Carewicz, O., Mercier, C. H., Rode, A., Kinnersley, N., Ward, P. (2000). Efficacy and safety of oseltamivir in treatment of acute influenza: A randomised controlled trial. Neuraminidase Inhibitor Flu Treatment Investigator Group. *Lancet*, *355*, 1845–1850.

29. Thompson, W. W., Shay, D. K., Weintraub, E., Brammer, L., Cox, N., Anderson, L. J., Fukuda, K. (2003). Mortality associated with influenza and respiratory syncytial virus in the United States. *JAMA*, *289*, 179–186.

30. Purcell, K., Fergie, J. (2004). Concurrent serious bacterial infections in 912 infants and children hospitalized for treatment of respiratory syncytial virus lower respiratory tract infection. *Pediatr. Infect. Dis. J.*, *23*, 267–269.

31. Hayden, F. G., Fritz, R., Lobo, M. C., Alvord, W., Strober, W., Straus, S. E. (1998). Local and systemic cytokine responses during experimental human influenza A virus infection. Relation to symptom formation and host defense. *J. Clin. Invest.*, *101*, 643–649.

32. Hayden, F. G., Belshe, R., Villanueva, C., Lanno, R., Hughes, C., Small, I., Dutkowski, R., Ward, P., Carr, J. (2004). Management of influenza in households: A prospective, randomized comparison of oseltamivir treatment with or without postexposure prophylaxis. *J. Infect. Dis.*, *189*, 440–449.

33. Prochoda, K., Mostow, S. R., Greenberg, K. (1993). Hantavirus-associated acute respiratory failure. *N. Engl. J. Med.*, *329*, 1744.

34. Fabbri, M., Maslow, M. J. (2001). Hantavirus pulmonary syndrome in the United States. *Curr. Infect. Dis. Rep.*, *3*, 258–265.

35. Khan, A. S., Young, J. C. (2001). Hantavirus pulmonary syndrome: At the crossroads. *Curr. Opin. Infect. Dis.*, *14*, 205–209.

36. Chapman, L. E., Mertz, G. J., Peters, C. J., Jolson, H. M., Khan, A. S., Ksiazek, T. G., Koster, F. T., Baum, K. F., Rollin, P. E., Pavia, A. T., Holman, R. C., Christenson, J. C., Rubin, P. J., Behrman, R. E., Bell, L. J., Simpson, G. L., Sadek, R. F. (1999). Intravenous ribavirin for hantavirus pulmonary syndrome: Safety and tolerance during 1 year of open-label experience. Ribavirin Study Group. *Antivir. Ther.*, *4*, 211–219.

37. Greenberg, S. B. (1994). Respiratory herpesvirus infections. An overview. *Chest*, *106*, 1S–2S.

38. Gwaltney, J. M., Jr., Hendley, J. O., Simon, G., Jordan, W. S., Jr. (1966). Rhinovirus infections in an industrial population. I. The occurrence of illness. *N. Engl. J. Med.*, *275*, 1261–1268.

39. Arruda, E., Boyle, T. R., Winther, B., Pevear, D. C., Gwaltney, J. M., Jr., Hayden, F. G. (1995). Localization of human rhinovirus replication in the upper respiratory tract by in situ hybridization. *J. Infect. Dis.*, *171*, 1329–1333.

40. Naclerio, R. M., Proud, D., Kagey-Sobotka, A., Lichtenstein, L. M., Hendley, J. O., Gwaltney, J. M., Jr. (1988). Is histamine responsible for the symptoms of rhinovirus colds? A look at the inflammatory mediators following infection. *Pediatr. Infect. Dis. J.*, *7*, 218–222.

41. Naclerio, R. M., Proud, D., Lichtenstein, L. M., Kagey-Sobotka, A., Hendley, J. O., Sorrentino, J., Gwaltney, J. M. (1988). Kinins are generated during experimental rhinovirus colds. *J. Infect. Dis.*, *157*, 133–142.
42. Pitkaranta, A., Virolainen, A., Jero, J., Arruda, E., Hayden, F. G. (1998). Detection of rhinovirus, respiratory syncytial virus, and coronavirus infections in acute otitis media by reverse transcriptase polymerase chain reaction. *Pediatrics*, *102*, 291–295.
43. Heikkinen, T., Chonmaitree, T. (2003). Importance of respiratory viruses in acute otitis media. *Clin. Microbiol. Rev.*, *16*, 230–241.
44. Valle, S. L. (1987). Febrile pharyngitis as the primary sign of HIV infection in a cluster of cases linked by sexual contact. *Scand. J. Infect. Dis.*, *19*, 13–17.
45. Nugent, K. M., Pesanti, E. L. (1983). Tracheal function during influenza infections. *Infect. Immun.*, *42*, 1102–1108.
46. Ison, M. G., Hayden, F. G. (2002). Viral infections in immunocompromised patients: What's new with respiratory viruses? *Curr. Opin. Infect. Dis.*, *15*, 355–367.
47. Garbino, J., Gerbase, M. W., Wunderli, W., Kolarova, L., Nicod, L. P., Rochat, T., Kaiser, L. (2004). Respiratory viruses and severe lower respiratory tract complications in hospitalized patients. *Chest*, *125*, 1033–1039.
48. De Clercq, E. (2001). Antiviral drugs: Current state of the art. *J. Clin. Virol.*, *22*, 73–89.
49. Nicholson, K. G. (1996). Use of antivirals in influenza in the elderly: Prophylaxis and therapy. *Gerontology*, *42*, 280–289.
50. Nicholson, K. G., Wiselka, M. J. (1991). Amantadine for influenza A. *BMJ*, *302*, 425–426.
51. Hayden, F. G., Osterhaus, A. D., Treanor, J. J., Fleming, D. M., Aoki, F. Y., Nicholson, K. G., Bohnen, A. M., Hirst, H. M., Keene, O., Wightman, K. (1997). Efficacy and safety of the neuraminidase inhibitor zanamivir in the treatment of influenzavirus infections. GG167 Influenza Study Group. *N. Engl. J. Med.*, *337*, 874–880.
52. Randolph, A. G., Wang, E. E. (1996). Ribavirin for respiratory syncytial virus lower respiratory tract infection. A systematic overview. *Arch. Pediatr. Adolesc. Med.*, *150*, 942–947.
53. Randolph, A. G., Wang, E. E. (2000). Ribavirin for respiratory syncytial virus infection of the lower respiratory tract. *Cochrane Database Syst. Rev.*, CD000181.
54. Sidwell, R. W., Robins, R. K., Hillyard, I. W. (1979). Ribavirin: An antiviral agent. *Pharmacol. Ther.*, *6*, 123–146.
55. Snell, N. J. (2001). Ribavirin—Current status of a broad spectrum antiviral agent. *Expert Opin. Pharmacother.*, *2*, 1317–1324.
56. Tam, R. C., Lau, J. Y., Hong, Z. (2001). Mechanisms of action of ribavirin in antiviral therapies. *Antivir. Chem. Chemother.*, *12*, 261–272.
57. Mejias, A., Chavez-Bueno, S., Rios, A. M., Saavedra-Lozano, J., Fonseca Aten, M., Hatfield, J., Kapur, P., Gomez, A. M., Jafri, H. S., Ramilo, O. (2004). Anti-respiratory syncytial virus (RSV) neutralizing antibody decreases lung inflammation, airway obstruction, and airway hyperresponsiveness in a murine RSV model. *Antimicrob. Agents Chemother.*, *48*, 1811–1822.

58. Van Reeth, K. (2000). Cytokines in the pathogenesis of influenza. *Vet. Microbiol.*, *74*, 109–116.

59. Tekkanat, K. K., Maassab, H. F., Cho, D. S., Lai, J. J., John, A., Berlin, A., Kaplan, M. H., Lukacs, N. W. (2001). IL-13-induced airway hyperreactivity during respiratory syncytial virus infection is STAT6 dependent. *J. Immunol.*, *166*, 3542–3548.

60. Tekkanat, K. K., Maassab, H., Miller, A., Berlin, A. A., Kunkel, S. L., Lukacs, N. W. (2002). RANTES (CCL5) production during primary respiratory syncytial virus infection exacerbates airway disease. *Eur. J. Immunol.*, *32*, 3276–3284.

61. Miller, A. L., Strieter, R. M., Gruber, A. D., Ho, S. B., Lukacs, N. W. (2003). CXCR2 regulates respiratory syncytial virus-induced airway hyperreactivity and mucus overproduction. *J. Immunol.*, *170*, 3348–3356.

62. Haeberle, H. A., Takizawa, R., Casola, A., Brasier, A. R., Dieterich, H. J., Van Rooijen, N., Gatalica, Z., Garofalo, R. P. (2002). Respiratory syncytial virus-induced activation of nuclear factor-kappaB in the lung involves alveolar macrophages and toll-like receptor 4-dependent pathways. *J. Infect. Dis.*, *186*, 1199–1206.

63. van Schaik, S. M., Welliver, R. C., Kimpen, J. L. (2000). Novel pathways in the pathogenesis of respiratory syncytial virus disease. *Pediatr. Pulmonol.*, *30*, 131–138.

64. Ge, Q., McManus, M. T., Nguyen, T., Shen, C. H., Sharp, P. A., Eisen, H. N., Chen, J. (2003). RNA interference of influenza virus production by directly targeting mRNA for degradation and indirectly inhibiting all viral RNA transcription. *Proc. Natl. Acad. Sci. USA*, *100*, 2718–2723.

65. Hwang, S., Tamilarasu, N., Kibler, K., Cao, H., Ali, A., Ping, Y. H., Jeang, K. T., Rana, T. M. (2003). Discovery of a small molecule Tat-trans-activation-responsive RNA antagonist that potently inhibits human immunodeficiency virus-1 replication. *J. Biol. Chem.*, *278*, 39092–39103.

66. De Clercq, E. (2002). Strategies in the design of antiviral drugs. *Nat. Rev.*, *Drug Discov.*, *1*, 13–25.

67. Leyssen, P., Charlier, N., Paeshuyse, J., De Clercq, E., Neyts, J. (2003). Prospects for antiviral therapy. *Adv. Virus Res.*, *61*, 511–553.

68. Shigeta, S. (2001). Targets of anti-influenza chemotherapy other than neuraminidase and proton pump. *Antivir. Chem. Chemother.*, *12* (Suppl. 1), 179–188.

69. Rogers, J. M., Diana, G. D., McKinlay, M. A. (1999). Pleconaril. A broad spectrum antipicornaviral agent. *Adv. Exp. Med. Biol.*, *458*, 69–76.

70. Romero, J. R. (2001). Pleconaril: A novel antipicornaviral drug. *Expert Opin. Investig. Drugs*, *10*, 369–379.

71. Hayden, F. G., Herrington, D. T., Coats, T. L., Kim, K., Cooper, E. C., Villano, S. A., Liu, S., Hudson, S., Pevear, D. C., Collett, M., McKinlay, M. (2003). Efficacy and safety of oral pleconaril for treatment of colds due to picornaviruses in adults: Results of 2 double-blind, randomized, placebo-controlled trials. *Clin. Infect. Dis.*, *36*, 1523–1532.

72. Poritz, M. A., Malmstrom, S., Schmitt, A., Kim, M. K., Zharkikh, L., Kamb, A., Teng, D. H. (2003). Isolation of a peptide inhibitor of human rhinovirus. *Virology*, *313*, 170–183.

73. Turner, R. B., Wecker, M. T., Pohl, G., Witek, T. J., McNally, E., St. George, R., Winther, B., Hayden, F. G. (1999). Efficacy of tremacamra, a soluble intercellular adhesion molecule 1, for experimental rhinovirus infection: A randomized clinical trial. *JAMA*, *281*, 1797–1804.

74. Reich, S. H., Johnson, T., Wallace, M. B., Kephart, S. E., Fuhrman, S. A., Worland, S. T., Matthews, D. A., Hendrickson, T. F., Chan, F., Meador, J., 3rd, Ferre, R. A., Brown, E. L., DeLisle, D. M., Patick, A. K., Binford, S. L., Ford, C. E. (2000). Substituted benzamide inhibitors of human rhinovirus 3C protease: Structure-based design, synthesis, and biological evaluation. *J. Med. Chem.*, *43*, 1670–1683.

75. Patick, A. K., Binford, S. L., Brothers, M. A., Jackson, R. L., Ford, C. E., Diem, M. D., Maldonado, F., Dragovich, P. S., Zhou, R., Prins, T. J., Fuhrman, S. A., Meador, J. W., Zalman, L. S., Matthews, D. A., Worland, S. T. (1999). In vitro antiviral activity of AG7088, a potent inhibitor of human rhinovirus 3C protease. *Antimicrob. Agents Chemother.*, *43*, 2444–2450.

76. Murray, M. A., Janc, J. W., Venkatraman, S., Babe, L. M. (2001). Peptidyl diazomethyl ketones inhibit the human rhinovirus 3C protease: Effect on virus yield by partial block of P3 polyprotein processing. *Antivir. Chem. Chemother.*, *12*, 273–281.

77. Wang, Q. M., Johnson, R. B., Jungheim, L. N., Cohen, J. D., Villarreal, E. C. (1998). Dual inhibition of human rhinovirus 2A and 3C proteases by homophthalimides. *Antimicrob. Agents Chemother.*, *42*, 916–920.

78. Player, M. R., Barnard, D. L., Torrence, P. F. (1998). Potent inhibition of respiratory syncytial virus replication using a 2-5A-antisense chimera targeted to signals within the virus genomic RNA. *Proc. Natl. Acad. Sci. USA*, *95*, 8874–8879.

79. Torrence, P. F., Powell, L. D. (2002). The quest for an efficacious antiviral for respiratory syncytial virus. *Antivir. Chem. Chemother.*, *13*, 325–344.

80. Joost Haasnoot, P. C., Cupac, D., Berkhout, B. (2003). Inhibition of virus replication by RNA interference. *J. Biomed. Sci.*, *10*, 607–616.

81. Malhotra, R., Ward, M., Bright, H., Priest, R., Foster, M. R., Hurle, M., Blair, E., Bird, M. (2003). Isolation and characterisation of potential respiratory syncytial virus receptor(s) on epithelial cells. *Microbes Infect.*, *5*, 123–133.

82. Haagmans, B. L., Kuiken, T., Martina, B. E., Fouchier, R. A., Rimmelzwaan, G. F., Van Amerongen, G., Van Riel, D., De Jong, T., Itamura, S., Chan, K. H., Tashiro, M., Osterhaus, A. D. (2004). Pegylated interferon-alpha protects type 1 pneumocytes against SARS coronavirus infection in macaques. *Natl. Med.*, *10*, 290–293.

83. Barnard, D. L., Hubbard, V. D., Burton, J., Smee, D. F., Morrey, J. D., Otto, M. J., Sidwell, R. W. (2004). Inhibition of severe acute respiratory syndrome-associated coronavirus (SARSCoV) by calpain inhibitors and beta-D-N4-hydroxycytidine. *Antivir. Chem. Chemother.*, *15*, 15–22.

84. Cook, P. M., Eglin, R. P., Easton, A. J. (1998). Pathogenesis of pneumovirus infections in mice: Detection of pneumonia virus of mice and human respiratory syncytial virus mRNA in lungs of infected mice by in situ hybridization. *J. Gen. Virol.*, *79*, 2411–2417.

85. Graham, B. S., Perkins, M. D., Wright, P. F., Karzon, D. T. (1988). Primary respiratory syncytial virus infection in mice. *J. Med. Virol.*, *26*, 153–162.

86. Byrd, L. G., Prince, G. A. (1997). Animal models of respiratory syncytial virus infection. *Clin. Infect. Dis.*, *25*, 1363–1368.
87. Peebles, R. S., Jr., Sheller, J. R., Collins, R. D., Jarzecka, K., Mitchell, D. B., Graham, B. S. (2000). Respiratory syncytial virus (RSV)-induced airway hyperresponsiveness in allergically sensitized mice is inhibited by live RSV and exacerbated by formalin-inactivated RSV. *J. Infect. Dis.*, *182*, 671–677.
88. Ottolini, M. G., Porter, D. D., Hemming, V. G., Hensen, S. A., Sami, I. R., Prince, G. A. (1996). Semi-permissive replication and functional aspects of the immune response in a cotton rat model of human parainfluenza virus type 3 infection. *J. Gen. Virol.*, *77*, 1739–1743.
89. Blanco, J. C., Pletneva, L., Boukhvalova, M., Richardson, J. Y., Harris, K. A., Prince, G. A. (2004). The cotton rat: An underutilized animal model for human infectious diseases can now be exploited using specific reagents to cytokines, chemokines, and interferons. *J. Interferon Cytokine Res.*, *24*, 21–28.
90. Ottolini, M., Blanco, J., Porter, D., Peterson, L., Curtis, S., Prince, G. (2003). Combination anti-inflammatory and antiviral therapy of influenza in a cotton rat model. *Pediatr. Pulmonol.*, *36*, 290–294.
91. Tuthill, T. J., Papadopoulos, N. G., Jourdan, P., Challinor, L. J., Sharp, N. A., Plumpton, C., Shah, K., Barnard, S., Dash, L., Burnet, J., Killington, R. A., Rowlands, D. J., Clarke, N. J., Blair, E. D., Johnston, S. L. (2003). Mouse respiratory epithelial cells support efficient replication of human rhinovirus. *J. Gen. Virol.*, *84*, 2829–2836.

# STRATEGIES IN THE DESIGN OF ANTIVIRAL DRUGS

Erik De Clercq and Piet Herdewijn
*Rega Institute for Medical Research, Katholieke Universiteit Leuven*
*Leuven, Belgium*

1 INTRODUCTION
2 VIRUS ADSORPTION INIBITORS
3 VIRUS–CELL FUSION INHIBITORS
4 INHIBITORS OF VIRAL DNA OR RNA SYNTHESIS
Viral DNA Polymerase Inhibitors
Reverse Transcriptase Inhibitors
Acyclic Nucleoside Phosphonates
Inhibitors of Processes Associated with Viral RNA Synthesis
5 VIRAL PROTEASE INHIBITORS
6 VIRAL NEURAMINIDASE INHIBITORS
7 IMP DEHYDROGENASE INHIBITORS
8 *S*-ADENOSYLHOMOCYSTEINE HYDROLASE INHIBITORS
9 CONCLUSION
References

## 1 INTRODUCTION

Effective vaccines have led, or may lead, to the eradication of important viral pathogens such as smallpox, polio, measles, mumps, and rubella. But other viral diseases, particularly human immunodeficiency virus (HIV) and hepatitis C

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

virus (HCV), have so far proven intractable to the vaccine approach. The need for effective antiviral drugs is further emphasized by the lack of vaccines for most respiratory tract virus infections [adeno, rhino, parainfluenza, and respiratory syncytial virus (RSV)], the widely occurring human papilloma viruses (HPV), herpesviruses [herpes simplex virus types 1 and 2 (HSV-1, -2), varicella-zoster virus (VZV), Epstein–Barr virus (EBV), cytomegalovirus (CMV), and human herpesviruses types 6, 7, and 8 (HHV-6, -7, -8)], and the vast array of hemorrhagic fever viruses. Although vaccines have been developed for hepatitis B virus (HBV) and influenza (type A and B), they have not been able to eliminate the need for effective chemotherapeutic agents.

Many new antiviral drugs have been licensed in recent years (Table 1), most of which are used for the treatment of HIV infections. Of the current armamentarium [1] of over 30 drugs, 19 are anti-HIV, 5 are anti-CMV, 5 are anti-HSV and anti-VZV, 1 is anti-RSV, 3 are anti-hepatitis B, 2 are anti-hepatitis C, and 2 are anti-influenza. But there is considerable room for improvement, as these compounds are not invariably efficacious or well tolerated. The emergence of drug–virus resistance and drug-related side effects are among the mainstay reasons for further refinement of antiviral drug design and development.

In antiviral drug design and targeting, one could, in principle, aim at either viral proteins (enzymes) or cellular proteins (enzymes). The first approach is likely to yield more specific, less toxic compounds with a narrow spectrum of antiviral activity and a higher likelihood of drug resistance development, whereas the second approach may afford antiviral compounds with a broader activity spectrum, less chance of resistance development but higher likelihood for toxicity. Both routes should be worth exploring, the preferred route being dictated by both the nature of the virus and the targets that the virus or its host cell have to offer.

As exemplified for HIV (Fig. 1), the viral life cycle encompasses a number of critical steps starting from the attachment of the virus to the cell and finishing by the release of the progeny virions from the cell. While the "normal" cytolytic viruses replicate their genome and express their genes autonomously, that is, independently from the host cell metabolism, the replicative cycle of retroviruses gets closely associated with the host cell, in

**Figure 1** The viral life cycle, as exemplified by the human immunodeficiency virus. Viral life cycles have several specific steps, many of which are targets for antiviral drugs. After virus adsorption, enveloped viruses enter the cell by virus–cell fusion. For HIV, which is a retrovirus with an RNA genome, replication of the genome occurs after reverse transcription and integration into the host–cell chromosome. For DNA viruses such as herpesviruses, the genome is replicated by a viral DNA polymerase. After transcription to RNA, translation and proteolytic processing of the precursor polypeptide, viral proteins assemble at the cell membrane, where they bud to release new virions.

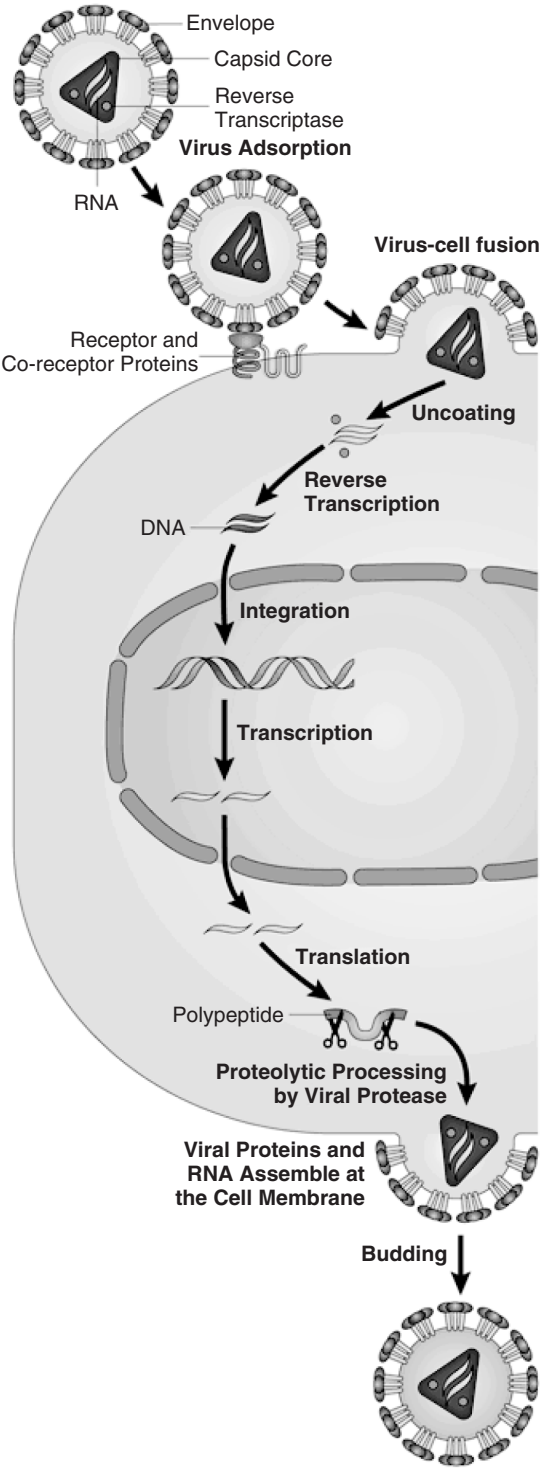
Envelope
Capsid Core
Reverse Transcriptase
Virus Adsorption
RNA
Virus-cell fusion
Receptor and Co-receptor Proteins
Uncoating
Reverse Transcription
DNA
Integration
Transcription
Translation
Polypeptide
Proteolytic Processing by Viral Protease
Viral Proteins and RNA Assemble at the Cell Membrane
Budding

**TABLE 1 Approaches to Antiviral Drug Design That Have Led to Licensed Drugs**

| Approach | Target Virus(es) | Compounds Approved | Selected Compounds in (pre)Clinical Development for the Indicated Target Virus |
|---|---|---|---|
| Virus adsorption inhibitors | HIV, HSV, CMV, RSV, and other enveloped viruses | | Polysulfates, polysulfonates, polycarboxylates, polyoxometalates, chicoric acid, zintevir, cosalane derivatives, cyanovirin, negatively charged albumins |
| Virus–cell fusion inhibitors | HIV, RSV, and other paramyxoviruses | Enfuvirtide | HIV: AMD3100 and SCH-C derivatives |
| Viral DNA polymerase inhibitors | Herpesviruses (HSV-1, -2, VZV, CMV, EBV, HHV-6, -7, -8) | Acyclovir, valaciclovir, ganciclovir, valganciclovir, penciclovir, famciclovir, brivudin,[a] foscarnet, fomivirsen | Bicyclic furopyrimidine nucleoside and nonnucleoside analogs, A5021, cyclohexenylguanine<br>Helicase-primase inhibitors: BILS 179BS and BAY 57-1293 |
| Reverse transcriptase inhibitors | HIV | NRTIs (nucleoside reverse transcriptase inhibitors): zidovudine, didanosine, zalcitabine, stavudine, lamivudine,[b] abacavir | Emtricitabine, amdoxovir, BCH-10652, racivir, reverset, elvucitabine, alovudine |
| | | NNRTIs (nonnucleoside reverse transcriptase inhibitors): nevirapine, delavirdine, efavirenz | Capravirine, dapivirine, UC781, DPC083, SJ-3366, (+)-calanolide A |

| | | | |
|---|---|---|---|
| Acyclic nucleoside phosphonates (NtRTIs (nucleotide reverse transcriptase inhibitors) | DNA viruses (polyoma-, papilloma-, herpes-, adeno-, and poxviruses), HIV, HBV | CMV: cidofovir<br>HBV: adefovir dipivoxil<br>HIV: tenofovir disoproxil fumarate<br>Disoproxil fumarate | |
| Inhibitors of processes associated with viral RNA synthesis | HIV<br>HCV | | HIV: integrase inhibitors<br>HCV: NS5B polymerase inhibitors<br>HCV: NS3 helicase inhibitors |
| Viral protease inhibitors | HIV, herpesviruses, rhinoviruses, HCV | HIV: saquinavir, ritonavir, indinavir, nelfinavir, a mprenavir, lopinavir | HIV: atazanavir, mozenavir, tipranavir<br>Human rhinovirus: AG7088<br>HCV: NS3 protease inhibitors |
| Viral neuraminidase inhibitors | Influenza A and B virus | Zanamivir, oseltamivir[c] | RWJ-270201, A-192558, A-315675 |
| IMP dehydrogenase inhibitors | HCV, RSV | Ribavirin[d] | Mycophenolic acid, EICAR, VX-497 |
| *S*-adenosylhomocysteine hydrolase inhibitors | (–)RNA hemorrhagic fever viruses (e.g., Ebola) | | 3-Deazaneplanocin A |

[a] Brivudin is approved in certain European countries (i.e., Germany).
[b] Lamivudine is also approved for the treatment of HBV.
[c] In addition to zanamivir and oseltamivir, amantadine and rimantadine have been approved as anti-influenza drugs, but these compounds are targeted at the viral uncoating process, not the viral neuraminidase.
[d] For the treatment of HCV infections ribavirin is used in combination with interferon-α.

that, subsequent to the reverse transcription (ribonucleic acid → deoxyribonucleic acid (RNA → DNA)] step, the resulting proviral DNA becomes integrated into the cellular genome and then follows the "classical" transcription and translation processes. Here, we focus primarily on approaches targeted at specific processes in viral infection (Fig. 1), including virus adsorption, virus–cell fusion, viral DNA or RNA synthesis (and cellular components associated therewith), and viral enzymes (protease, neuraminidase). Two cellular enzymatic processes, namely IMP dehydrogenase and *S*-adenosylhomocysteine hydrolase, could also be envisaged as targets for certain classes of antiviral agents, and will therefore be addressed as well.

## 2 VIRUS ADSORPTION INHIBITORS

Numerous polyanionic compounds [e.g., polysulfates such as polyvinylalcohol sulfate, polysulfonates such as polyvinylsulfonate (Fig. 2), polycarboxylates, polynucleotides such as zintevir, polyoxometalates, negatively charged albumins] have been shown to inhibit HIV replication by preventing virus attachment (adsorption) to the cell surface. All these negatively charged polymers may be expected to interact with the positively charged amino acids in the V3 loop (which is rich in R and K residues) of the HIV glycoprotein gp120. In doing so [2], the polyanions shield the V3 loop and thus hamper the binding of the HIV virions with heparin sulfate, the primary binding site at the cell surface before a more specific binding can occur with the CD4 receptor on $CD4^+$ cells.

Heparin sulfate is widely expressed on animal cells and is involved in virus–cell binding for a broad array of enveloped viruses, including HSV [3] and dengue virus [4]. So, polysulfates, polysulfonates, and other polyanionic substances that interfere with the target cell binding of these enveloped viruses may be effective in the treatment and prophylaxis of such infections. In the management of HIV infections, polyanionic substances may have a major role as vaginal microbicides, as, when applied in an appropriate formulation, they may successfully prevent sexual transmission of HIV infection. Moreover, these polyanions are not only active against HIV but also HSV and other sexually transmitted disease (STD) pathogens such as *Neisseria gonorrhoeae* and *Chlamydia trachomatis* [5].

Although polyanions may have multiple sites of interaction (i.e., virus adsorption, reverse transcriptase, integrase), virus attachment to the cells would be the preferred target from a therapeutic viewpoint, as it is the first opportunity to curtail the viral life cycle, and the polyanions do not need to enter the cells (which would be problematic). The interaction of polyanionic substances at this level can also be considered specific, since repeated passage of HIV in the presence of polyanions can lead to resistance mediated by mutations in the envelope glycoprotein gp120, particularly in the V3 loop (K269E,

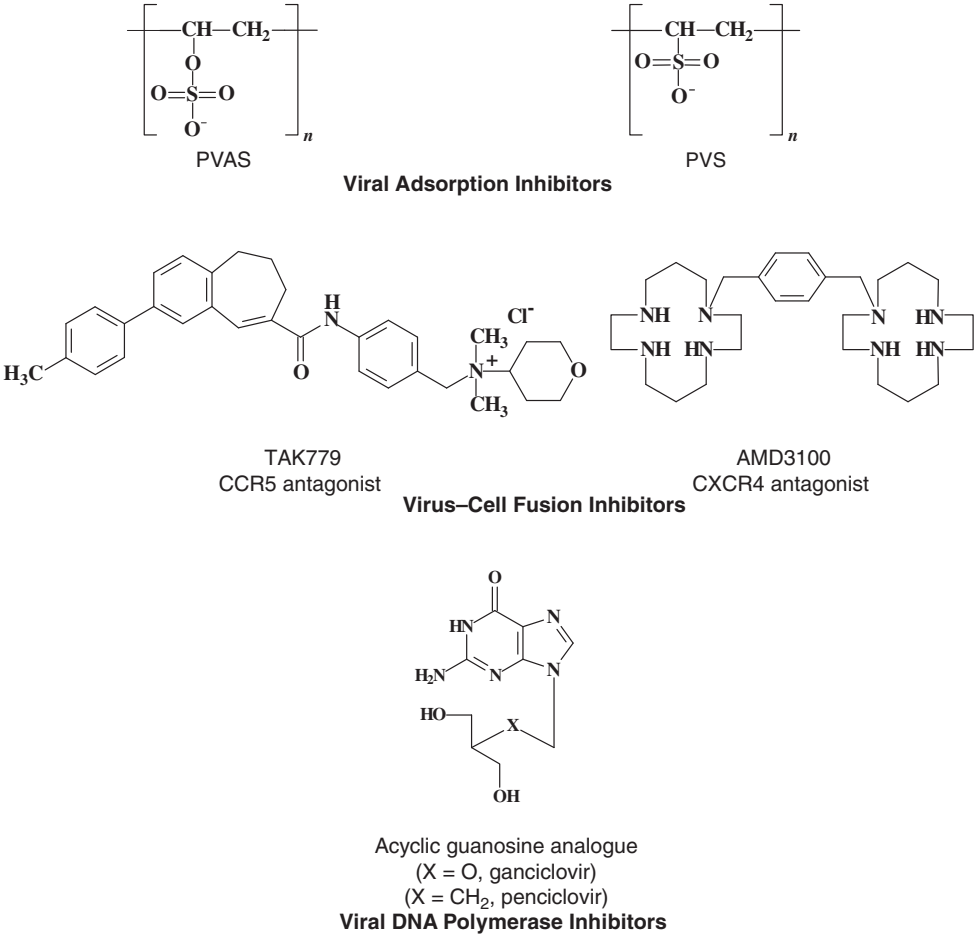


**Figure 2** Basic (skeletal) pharmacophores of the classes of antiviral agents described in this review. (IMP, inosine monophosphate; NRTI, nucleoside reverse transcriptase inhibitor; NNRTI, nonnucleoside reverse transcriptase inhibitor; PVS, polyvinylsulfonate; PVAS, polyvinylalcohol sulfate; SAH, *S*-adenosylhomocysteine.)

Q278H, N293D), as originally shown for dextran sulfate [6] and subsequently for zintevir [7] and negatively charged albumins [8].

## 3 VIRUS–CELL FUSION INHIBITORS

Enveloped viruses, as a rule, enter their host cells by fusion between the viral envelope and cellular plasma membrane (Fig. 1). This fusion process is basically similar for different enveloped virus families (i.e., retro-, paramyxo-, her-

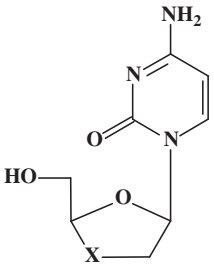


NRTI
(X = $CHN_3$, AZT)

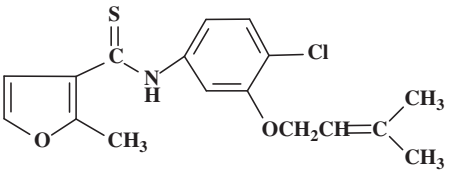


NNRTI (UC781)

**Reverse Transcriptase Inhibitors**

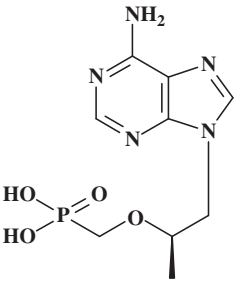


Tenofovir
**Acyclic Nucleoside Phosphonate**

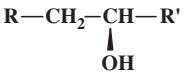


Peptidomimetic

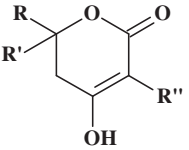


Nonpeptidomimetic

**Protease Inhibitors**

**Figure 2** *Continued*

pesviruses), but for HIV it is preceded by the interaction of the viral glycoprotein gp120 with its co-receptor (CXCR4 for T-tropic or X4 HIV strains, CCR5 for M-tropic or R5 strains). CXCR4 and CCR5 normally act as the receptors for the CXC chemokine SDF-1 (stromal-cell derived factor), CC chemokines RANTES (regulated upon activation, normal T-cell expressed and secreted), and MIP-1α and -1β (macrophage inflammatory proteins), respectively. The fact that CXCR4 and CCR5, quite incidentally, are used by

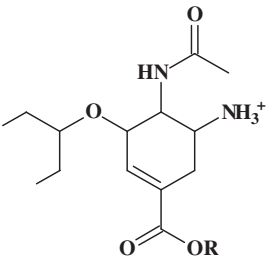


Oseltamivir (R = $CH_2CH_3$)

**Neuraminidase Inhibitor**

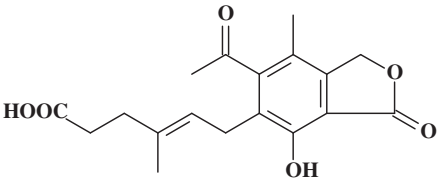


Mycophenolic Acid

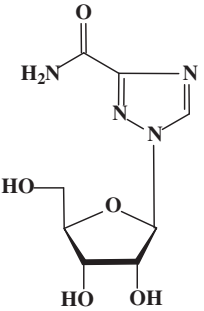


Ribavirin

**IMP Dehydrogenase Inhibitors**

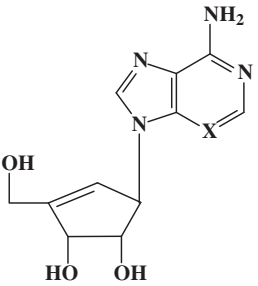


X = N or CH

**SAH Hydrolase Inhibitors**

**Figure 2** *Continued*

HIV as co-receptors to enter the cells has prompted the search for CXCR4 and CCR5 antagonists, which, through blockade of the corresponding co-receptor, may be expected to block HIV entry into the cells.

This has now been demonstrated with a number of compounds, the most prominent among the CXCR4 antagonists being the bicyclam AMD3100 (Fig. 2) [9, 10] and the best documented among the CCR5 antagonists being

TAK-779 (Fig. 2) [11, 12]. An extensive SAR study of phenylenebis(methylene) linked bis-tetraazamacrocycles that inhibit HIV replication has been described. Depending on the substitution of the phenylenebis(methylene) linker, submicromolar anti-HIV activity was exhibited by analogs bearing macrocycles of 12 to 14 ring members. Identical macrocyclic rings are not required for activity. However, substitution of an acyclic polyamine equivalent for one of the cyclam rings resulted in a substantial reduction of anti-HIV potency, clearly establishing the importance of the constrained macrocyclic structure. The parasubstituted phenylene gave a more active compound than the metasubstituted phenylene. The antiviral activity of the bicyclam analogs appeared to be relativly insensitive to the presence of electron-withdrawing or electron-donating substituents on the phenyl ring. The activity of pyridine-linked bicyclams is dependent on the way the heteroaromatic linker is connecting the cyclam rings. For example, 2,6- and 3,5-pyridine linked bicyclams (**3** and **4**) are more potent than the 2,5- and 2,4-substituted pyridine-linked compounds (**5** and **6**). The position of the four amino groups of the tetraazamacrocyclic ring is important. The Py[iso-14]ane $N_4$ compound (**7**) is less active than the parent compound, and introduction of a pyridine group results in a further reduction in biological activity (**8**). However, single pyridine insertion in the concept of an inversed macrocyclic framework (**9**) exhibited anti-HIV-1 potency surpassing that of the saturated, aliphatic counterparts (Fig. 3) [13–15].

The site of interaction of TAK-779 with the transmembrane helices of CCR5 has been mapped (Fig. 4) [12], and, likewise, crucial amino acid residues involved in the binding of AMD3100 to its receptor CXCR4 have been identified [16]. AMD3100 acts on the CXCR4 receptor through binding to $Asp^{171}$ in TM-IV and $Asp^{262}$ in TM-VI, with each of its cyclam moieties. This is demonstrated in the molecular model of the main ligand binding pocket of the CXCR4 receptor wherein AMD3100 was manually docked in a manner that respects the above interactions. AMD3100 could prevent the receptor from changing into an active conformation (Fig. 5).

Recently, a new CCR5 antagonist, SCH-C (SCH 351125), was announced as an orally bioavailable inhibitor of M-tropic R5 strains, capable of suppressing R5 HIV-1 infection both in vitro and in vivo (SCID-hu Thy/Liv mice) [21]. The clinical potential of the CXCR4 and CCR5 antagonists in the management of HIV infections remains to be proven. To ensure maximal coverage (both X4 and R5 strains), dual CXCR4/CCR5 antagonists should be developed or single CCR5 and CXCR4 antagonists should be combined.

The interaction of gp120 with its co-receptor (CCR5 or CXCR4) triggers a series of conformational changes in the gp120/gp41 complex, ultimately leading to the formation of a "trimer-of-hairpins" structure in gp41, which is a bundle of six α-helices: three α-helices formed by the COOH-terminal regions packed in an antiparallel manner with three α-helices formed by the $NH_2$-terminal regions. The fusion peptide region, located at the extreme $NH_2$-terminus, will insert into the cellular membrane, while the COOH-terminal

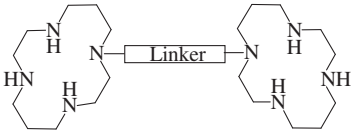


| | | $EC_{50}$ (μM) | $CC_{50}$ (μM) |
|---|---|---|---|
| **1** | | 0.0042 | > 421 |
| **2** | 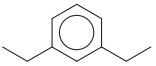 | 0.0337 | > 421 |
| **3** | 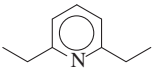 | 0.025 | > 405 |
| **4** | 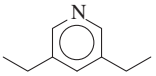 | 0.032 | > 395 |
| **5** | 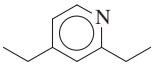 | 16.367 | 17 |
| **6** | 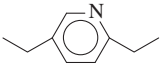 | 0.908 | 18 |
| **7** | 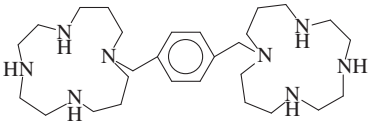 | 0.034 | > 421 |
| **8** | 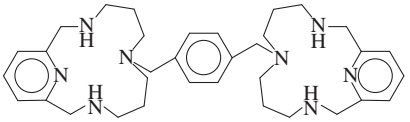 | 0.534 | 199 |
| **9** | 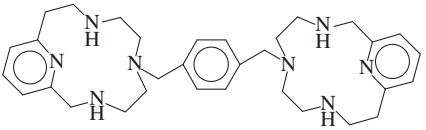 | 0.008 | 194 |

**Figure 3** Anti-HIV-1 ($III_B$)data for a series of bis-azamacrocycles showing the importance of the linker and the constitution of the cyclam moiety.

region remains anchored in the viral envelope: in this sense the trimer-of-hairpins motif brings the two membranes together, and agents that interfere with the formation of the gp41 trimer-of-hairpins structure may be expected to inhibit the fusion process [18].

Several constructs have been designed to interfere with the gp41-mediated fusion process: the so-called 5-helix that binds the COOH-terminal region of

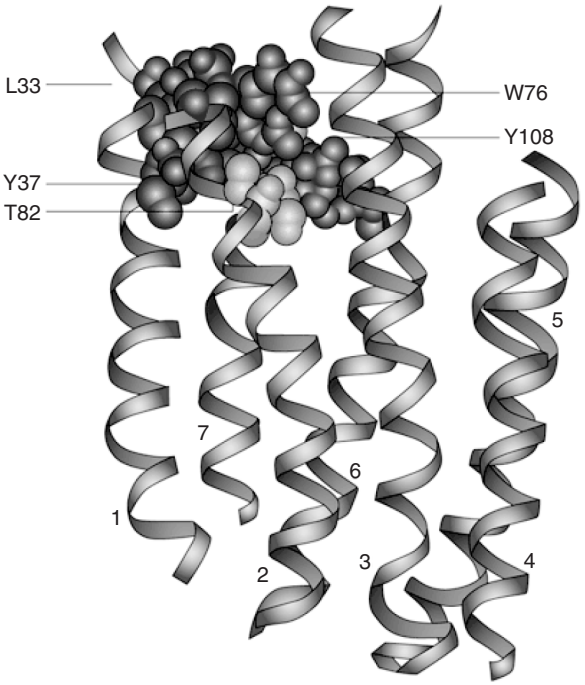


**Figure 4** Interactions of antiviral drugs with their protein targets. A structural model of CCR5 complexed with TAK-779 (Fig. 2) viewed from within the plane of the membrane [12]. The indicated cluster of amino acids in the TAK779 binding site includes several aromatic residues (Y37, W76, Y108) that might form favorable interactions with the aromatic rings of TAK779.

gp41 [22], D-peptide inhibitors that dock into the pocket formed by the α-helices of gp41 [23], and T-20 (enfuvirtide, previously called DP-178, a synthetic 36-amino-acid peptide corresponding to residues 127 to 162 of the ectodomain of gp41). Enfuvirtide disrupts the conformational changes associated with membrane fusion. T-20 has proved effective in reducing the plasma HIV levels in humans, thus providing the proof-of-concept that viral entry can be successfully blocked in vivo [24]. Disadvantages are that enfuvirtide needs a twice-daily subcutaneous injection for delivery and that its manufacturing process is expensive and complicated. Worldwide clinical studies have indicated that enfuvirtide, in combination with three to five other antiretroviral drugs, is able to effect an incremental decrease in viral load [25, 26].

Insight into the HIV fusion process should help in designing fusion inhibitors for other viruses as well, as trimer-of-hairpins motifs may also be predicted [16] for other virus families, including paramyxoviridae (parainfluenza, measles, respiratory syncytial virus). In fact, peptides (like T-20) for each of these paramyxoviruses have been shown to block viral fusion [27].

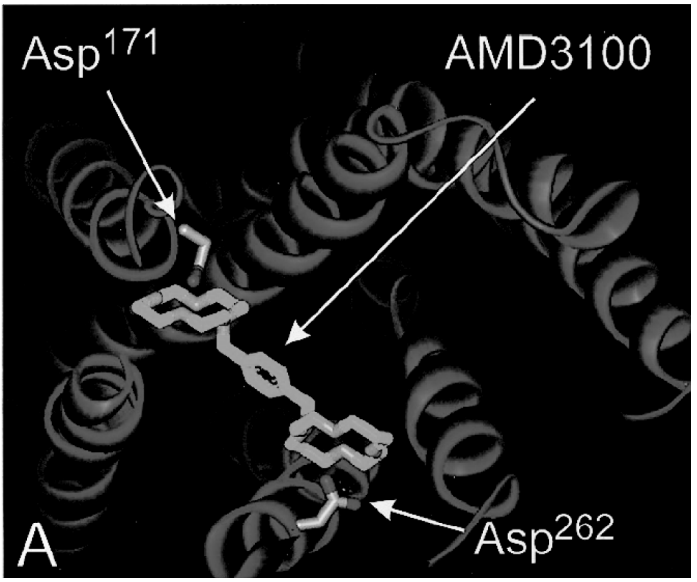


**Figure 5** Molecular model of the main ligand binding pocket of the CXCR4 receptor with AMD3100. The receptor model is built over the rhodospin model of Palczewski et al. [17]. The conformation of AMD3100 is based on structural requirements of high antiviral effect of AMD3100 [18, 19] and the crystallographic X-ray structure of 6,6′-spiro-bis(1,4,8,11-tetraazacyclotetradecane)-dinickel(II)tetraperchlorate [20] obtained from the Cambridge Structural Database.

Also, a cobalt-chelating complex (CTC-96) that has been shown to inhibit infection by herpesviruses (i.e., HSV) through blocking fusion [28] may be expected to possess an extended antiviral activity spectrum, given the premise that enveloped viruses belonging to different virus families share an analogous process of membrane fusion.

## 4 INHIBITORS OF VIRAL DNA OR RNA SYNTHESIS

### Viral DNA Polymerase Inhibitors

Being DNA viruses, the herpesviruses heavily rely for their replication on their own genome-encoded DNA polymerase. In contrast with the retroviruses (Fig. 1), herpesviruses do not have a reverse transcription step in their replicative cycle, which means that their genome can be replicated by the viral DNA polymerase after the latter has been expressed in the virus-infected cell. All the antiviral agents that are currently available for the treatment of herpesvirus (i.e., HSV-1, HSV-2, VZV, CMV) infections are nucleoside analogs: They belong to either the class of the acyclic guanosine analogs (i.e., acyclovir,

penciclovir, ganciclovir, and their oral prodrug forms valaciclovir, famciclovir, and valganciclovir, respectively) or thymidine analogs (i.e., brivudin) (Fig. 2 and Table 1). All these compounds are targeted at the viral DNA polymerase, but before they can interact with viral DNA synthesis, they need to be phosphorylated intracellularly to the triphosphate form. The first (and, for brivudin, also the second) phosphorylation step is ensured by the HSV- or VZV-encoded thymidine kinase, or CMV-encoded protein kinase and is thus confined to the virus-infected cells, which explains the specific antiviral action of the established antiherpetic compounds. Subsequent phosphorylations are performed by cellular kinases. In their triphosphate form, the nucleoside analogs then interact with the viral DNA polymerase as either a competitive inhibitor or an alternate substrate with respect to the natural substrate (i.e., dGTP for the guanosine analogs, dTTP for the thymidine analogs). If the acyclic guanosine analogs act as alternate substrates, their incorporation prevents further chain elongation (Fig. 6*a*).

An important issue with respect to antiherpesvirus (HSV-1, HSV-2, VZV) agents is their intracellular phosphorylation. Thymidine kinase (TK) is a crucial enzyme in the salvage pathway of thymidine 5′-triphosphate, which is a precursor of the thymidine incorporation into DNA. The substrate specificity of human TK is essentially limited to thymidine. The viral enzyme has a broader substrate specificity than the human enzyme and also guanine nucleoside analogs are accepted as substrates. The viral TK is important for the pathogenicity of the virus in humans, and HSV TK is a potential target for drugs that might prevent reactivation of HSV when used prophylactically.

Modified nucleosides have a much higher affinity for HSV-1 TK than for cellular TK, so that phosphorylation may occur preferentially in virus-infected cells. As mentioned before, the antiherpes nucleoside analogs can be classified in guanosine analogs and thymidine analogs. Acyclovir is the prototype of the guanine series. Acyclovir has high selectivity and *quasi*-negligible side effects, but it has rather poor efficacy. Although the triphosphate of acyclovir is an excellent inhibitor of DNA synthesis because of its strong chain-terminating effect, acyclovir itself is a poor substrate for TK. Analogs providing good binding (low $K_m/K_{cat}$) to both viral TK and viral DNA polymerase, while retaining low toxicity and high specificity for the viral TK over the cellular TK, would lead to significantly improved chemotherapy.

The prototype of a pyrimidine nucleoside with antiherpes activity is 5-(*E*)-(2-bromovinyl)-2′-deoxyuridine (BVDU, brivudin). BVDU is phosphorylated preferentially by the HSV-1 encoded dThd/dCyd kinase to the 5′-monophosphate and the 5′-diphosphate; the latter is further phosphorylated to the 5′-triphosphate by cellular kinase(s). BVDU can be incorporated into the DNA. Thus, a common characteristic for all antiherpesvirus (HSV, VZV) nucleosides is that they have to be phosphorylated by the virus-specific kinase in infected cells. The modes of binding of the guanine and of the pyrimidine nucleosides to the viral TK are different, which has been demonstrated by co-crystallization experiments [29]. The guanine moiety of acyclovir is bound

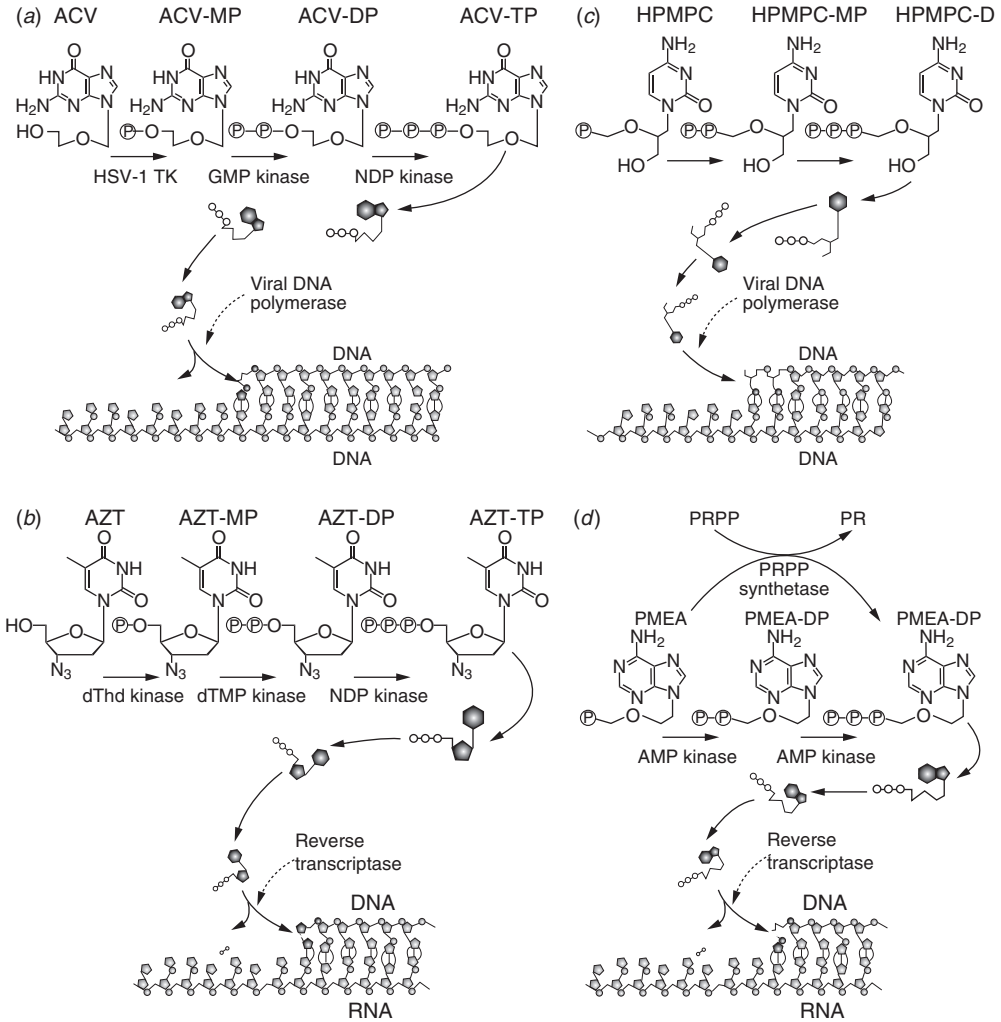


**Figure 6** Examples of antiviral nucleoside analogs acting by a chain termination mechanism. (*a*) Acyclovir (ACV) targets viral DNA polymerases such as the herpesvirus DNA polymerase. Before it can interact with viral DNA synthesis, it needs to be phosphorylated intracellularly, in three steps, to the triphosphate form. The first phosphorylation step is ensured by the HSV-encoded thymidine kinase (TK), and is thus confined to the virus-infected cells. (*b*) Azidothymidine (AZT) targets the HIV reverse transcriptase and also needs to be phosphorylated, in three steps, to the triphospate form before it can interfere with reverse transcription. (*c*) Cidofovir (HPMPC), an acyclic nucleotide analog, which can be viewed as an acyclic nucleoside analog extended by a phosphonate moiety, targets viral DNA polymerases, and is active against DNA viruses whether or not they encode for a specific viral thymidine kinase. In contrast with acyclovir and azidothymidine, cidofovir requires only two phosphorylations to be converted to the active (triphosphate) form. (*d*) Adefovir (PMEA)—also an acyclic nucleoside phosphonate—is active against retroviruses and hepadnaviruses; like cidofovir, adefovir needs only two phosphorylations and can thus bypass the nucleoside kinase reaction that limits the activity of dideoxynucleoside analogs such as AZT.

with hydrogen bond pairing being made with Gln-125 via the 1-NH and 6-carbonyl group; also Gln-125 is close enough for a possible further hydrogen bond with the 2-amino group. The 6-carbonyl group of acyclovir is close enough to the guanidinium group of Arg-176 to form a hydrogen bond. The hydroxyl group of acyclovir, lying close to the position occupied by the 5′-OH of dTh and interacts with the side chains of Arg-163 and Glu-83 and a nearby water (Fig. 7).

BVDU shows a mode of binding to HSV-TK, which is similar to that of deoxythymidine, as highlighted by hydrogen bonding and nonpolar interactions. The deoxyribose makes hydrogen bond interactions via its 3′-OH with Tyr-101 and Glu-225, via its 5′-OH with Glu-83, and via its 5′-OH and a water molecule with Arg-163. The base makes van der Waals contacts with Met-128 and Ile-100 on one side, and with Tyr-172 on the other, and makes pairwise hydrogen bond interactions via its 4-carbonyl and 3-NH groups with the amide of Gln-125.

The 2-carbonyl group of BVDU is hydrogen bonded to two water molecules, which, in turn, interact with the guanidium group of Arg-176, the hydroxyl group of Tyr-101, and the side-chain carbonyl of Gln-125. The bulky 5-substituent of BVDU occupies the deep space available in the neighborhood of residues Trp-88, Tyr-132, Arg-163, and Ala-167. This accommodation is at the expense of a relocation of the side chain of Tyr-132, shifted away from the ligand to make room for the substituent group (Fig. 8).

Thus, the larger guanine group of acyclovir occupies an environment defined by the same residues as the pyrimidine ring of BVDU, and the volume occupied by guanine is approximately coplanar with that occupied by uracil. However, the guanine is located much closer to Tyr-101 and Arg-176, and direct hydrogen bonding is observed in the case of acyclovir. Hydrogen bonding with Gln-125 involves a conformational shift of the side chain, as well as 180° rotation of the amide.

However, the determination of the way a molecule binds in the active site of an enzyme is only the first step in the drug discovery process. Analysis of the structural factors influencing the affinity of the compounds for the enzyme should be followed by calculations of binding energies in function of conformation, which may lead to a quantitative explanation of activity in function of binding strength. This model may then be used for further drug discovery. These studies have been performed with a series of 5-substituted 2′-deoxyuridine substrates of the HSV-1 TK [30]. More than 80 percent of the total interaction energy of thymidine comes from the interaction with only six residues, namely the four hydrogen bond partners Glu-83, Glu-225, Gln-125, and Tyr-101 and residues Tyr-172 and Met-128, which sandwich the nucleoside base by means of stacking interactions. Substitution of the 5-methyl group of thymidine by a bulkier, unsaturated group such as a thiophene has no significant effect on the major binding interactions involving the base and sugar moieties. The 5-substituent is positioned in a small pocket and the accessible portion of this pocket is mainly formed by Tyr-132 (a small displacement of

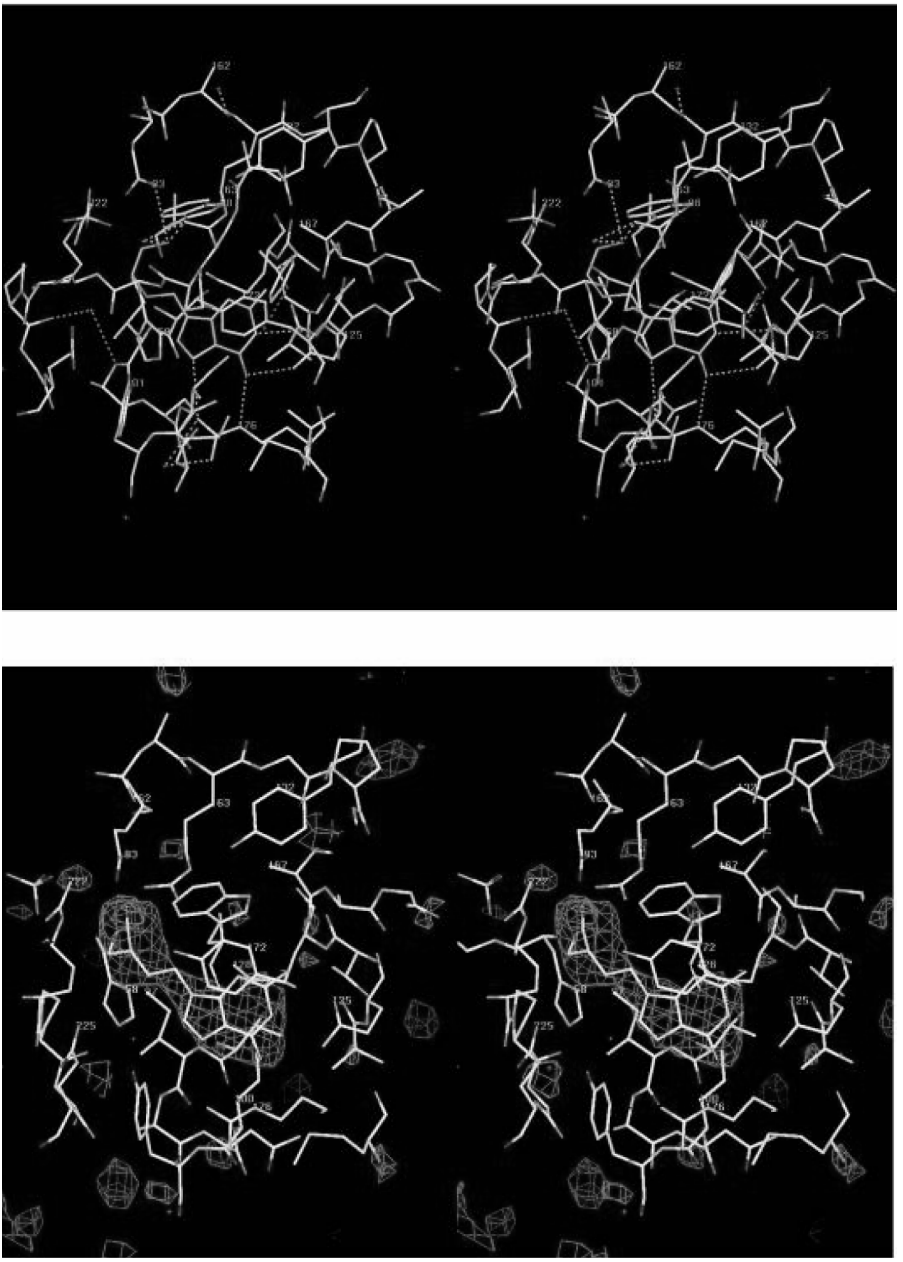

**Figure 7** *Top*: Stereo view of the binding of aciclovir to TK (molecule I), showing active-site residues and intermolecular hydrogen bonding. *Bottom*: Stereo view of the binding of penciclovir to TK (molecule I) superimposed on difference Fourier map contoured at 3σ density.

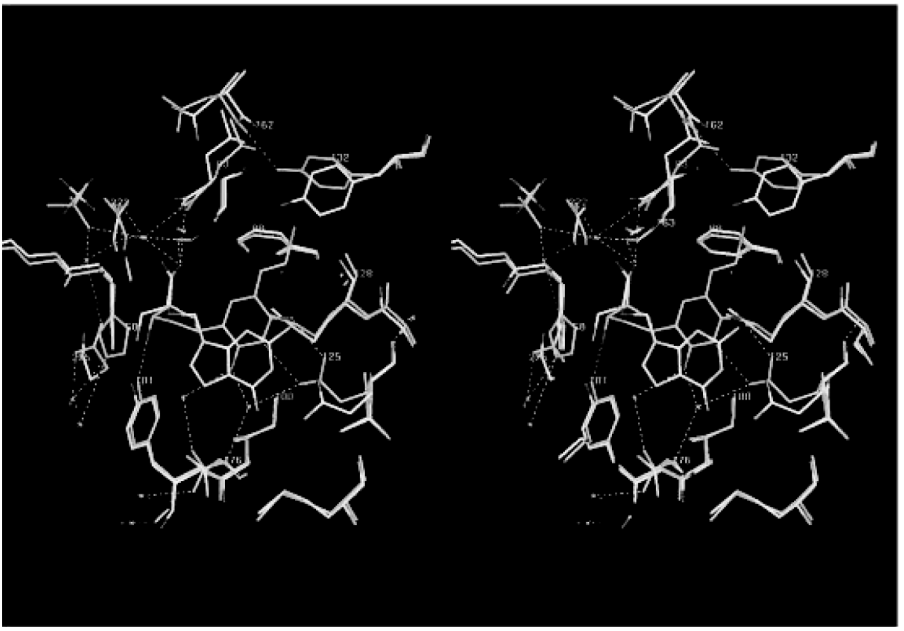

**Figure 8** Stereo view of the active site of the TK:BVDU complex (molecule I) overlaid on the active site of the TK/ganciclovir complex after alignment of enzyme molecules. Intermolecular hydrogen bonding is shown for the TK/BVDU complex.

Tyr-132 is needed to enlarge this pocket to accommodate the bulky substituent). The observed affinity differences of 18 different 5-substituted 2′-deoxyuridines was rationalized based on a binding energy–affinity relationship study. Interaction energies between substrate and enzyme, under the form of nonbonded van der Waals and electrostatic forces, were calculated for each complex and for each experimental setup to provide a measure of the tightness of binding. In order to account for possible differences in hydrophobicity between the compounds, free energies of solvation for each of the 18 substrates were also calculated. The differences in solvation-free energy between the 18 compounds can be largely attributed to the different substitution patterns of position 5. The best correlation was found for the case where the dielectric constant is treated as a distance-dependent function and where only the residues positioned within 3 Å of the substrate were allowed to relax. The least-squares equation for this particular case can be given by

$$\ln(\mathrm{IC}_{50}) = 0.177\ \mathrm{IE} + \Delta G_{\mathrm{sol}} RT + 51.9$$

in which IE stands for interaction energy in kJ/mol, $\Delta G_{\mathrm{sol}}$ is the solvation free energy of the substrate in kJ/mol, $RT$ is expressed in kJ/mol, and $\mathrm{IC}_{50}$ is

expressed in μM. A plot of this equation is given in Fig. 9. It is clear that the affinity for this series of 5-substituted 2′-deoxy uridine substrates can be improved either by increasing the side chain's interaction energy or by increasing the side chain's hydrophobicity (Fig. 10 and Fig. 11).

Is there room for improvement? As the acyclic guanosine analogs are not readily taken up orally, their oral bioavailability has been improved by formulating them as prodrug forms (valaciclovir, famciclovir, and valganciclovir). The success obtained with acyclovir, valaciclovir, and famciclovir in the treatment of HSV and VZV infections has impeded further progress in this area. Yet, brivudin, which is considerably more potent than acyclovir and penciclovir as an anti-VZV agent, represents an important alternative for the treatment of VZV infections. Although brivudin is active in the nanomolar concentration range against VZV replication, its potency can still be superseded by bicyclic furopyrimidine nucleoside analogs bearing a long alkyl or alkylaryl side chain attached to the furane ring [35, 36]. The mechanism of action of these exquisitely potent and selective anti-VZV agents remains to be elucidated, although there is no doubt that their specificity for VZV is governed by the virus-encoded thymidine kinase.

Within the class of the guanosine analogs, several new congeners have been described, namely A-5021 [37, 38] and the D- and L-enantiomers of cyclohexenylguanine [39]. As far as it has been determined, these compounds seem to have an activity spectrum and mode of action similar to that of acyclovir, but further studies seem warranted to verify whether these new guanosine analogs may be endowed with an extended spectrum of activity [i.e., against HHV-6, -7, and -8, which are not (particularly) sensitive to acyclovir or brivudin] or increased in vivo efficacy, or improved pharmacokinetics, or other properties that make them worth pursuing.

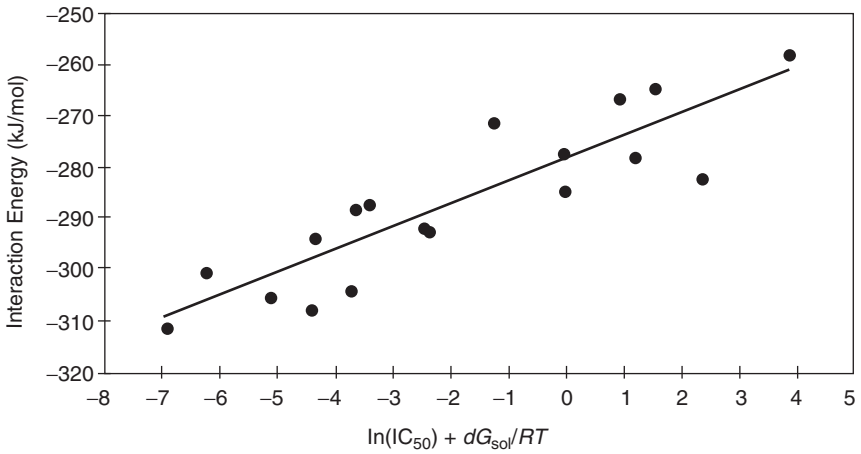


**Figure 9** Plot of $\ln(IC_{50}) + \Delta G_{sol}/RT$ vs. interaction energy.

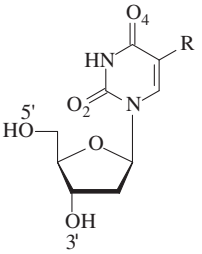


| | -R | $IC_{50}$ (μM) | | -R | $IC_{50}$ (μM) |
|---|---|---|---|---|---|
| **T** | —Me | 1.0 | **9** | 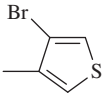 | 143 |
| **1** | 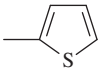 | 2.4 | **10** | 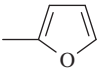 | 2.9 |
| **2** | 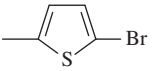 | 3.5 | **11** | 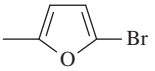 | 51 |
| **3** | 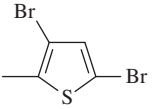 | 62 | **12** | 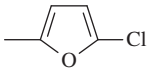 | 28 |
| **4** | 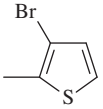 | 61 | **13** | 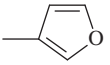 | 1.5 |
| **5** | 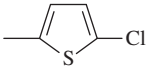 | 2.3 | **14** | 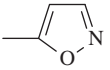 | 36 |
| **6** | 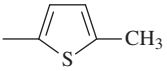 | 2.3 | **15** | 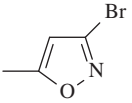 | 34 |
| **7** | 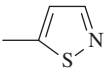 | 3.1 | **16** | 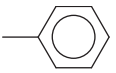 | 102 |
| **8** | 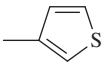 | 4.0 | **17** | 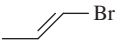 | 0.3 |

**Figure 10** The compounds with their numbering label and experimental affinity for HSV-1 thymidine kinase. Affinity data were obtained from the literature [31–33] and are the concentrations required to inhibit thymidine phosphorylation by 50 percent against radiolabeled thymidine used at a concentration of 1 μM. All affinities were measured under the same conditions, and highly similar $K_m$ values (0.25 μM) were found in all cases [34].

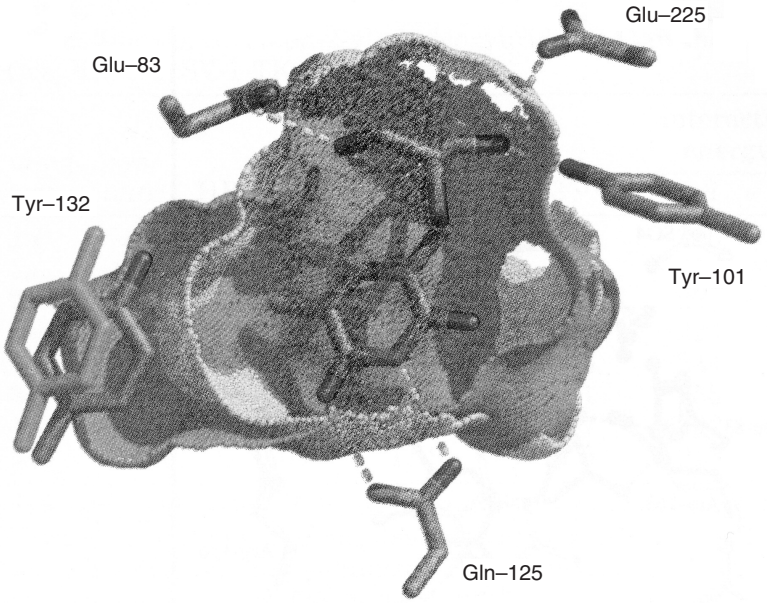


**Figure 11** Only a minor small position shift of Tyr-132 is needed to enlarge the pocket of the 5-methyl group of thymidine so that it becomes large enough to occupy furane or thiophene groups. The bottom part of the solvent-accessible surface of the pocket is as found in the crystal structure; the new pocket after a small position shift of Tyr-132 is also shown.

In addition to the viral DNA polymerase, the HSV DNA helicase–RNA primase, an enzyme composed of the HSV gene products UL5, UL8, and UL52 that in terms of the viral DNA synthesis process precedes the DNA polymerase activity, has been identified as an attractive target for the design of new inhibitors of HSV-1 and HSV-2 infection: BILS 179BS and BAY 57-1293 are two examples of such helicase [40, 41]. These compounds exhibited remarkable in vivo efficacy in animal models of HSV-1 and HSV-2 infection, combined with excellent oral bioavailability. The antiviral activity of BAY 57–1293 was quoted as superior to all compounds currently used to treat HSV infections [42]. These data obviously validate the use of helicase–primase inhibitors for the treatment of HSV infections but leave us with the question of whether this strategy would also work with herpesviruses other than HSV or with DNA viruses at large.

## Reverse Transcriptase Inhibitors

As is evident from Figure 1, the reverse transcriptase plays an essential role in the replicative cycle of retroviruses such as HIV, as it ensures the formation of proviral DNA that will then be integrated into the host cell genome and

passed onto all the progeny cells resulting from the once infected parent cell. The substrate (dNTP) binding site of HIV reverse transcriptase (RT) has proven to be an attractive target for nucleosidic HIV inhibitors: No less than six nucleoside analogs—zidovudine (azidothymidine, AZT), didanosine (dideoxyinosine), zalcitabine (dideoxycytidine), stavudine (didehydrodideoxythymidine, d4T), lamivudine (3′-thiadideoxycytidine), and abacavir—have been licensed as anti-HIV drugs (Fig. 2, Table 1), and several others such as emtricitabine [43], amdoxovir [44], BCH-10652 [(±)-2′-deoxy-3′-oxa-4′-thiocytidine, dOTC) [45], 5-fluoro-substituted dOTC [46], 5-fluoro-substituted d4C (reverset) [47], its L-counterpart (elvucitabine, ACH-126443) [48], and alovudine (MIV-310, the 3′-fluoro counterpart of AZT) [49] are in (advanced) clinical development. All these dideoxynucleoside analogs act according to a similar "recipe": As exemplified for AZT (Fig. 6*b*), they must be phosphorylated intracellularly consecutively by three cellular kinases, a nucleoside kinase, a nucleoside 5′-monophosphate kinase, and a nucleoside 5′-diphosphate kinase, to the corresponding 5′-triphosphate derivative, before the latter can interact, as a chain terminator, with the reverse transcription (RNA → DNA) reaction. One of the mechanisms by which resistance to AZT may arise is through removal of the chain-terminating residue, a kind of repair reaction involving pyrophosphorolysis, that may be regarded as the opposite of the reverse transcriptase reaction. Not all chain terminators are readily removed, for example, the acyclic nucleoside phosphonate derivative tenofovir (PMPA; Fig. 2) is not (see below), and, in this sense, PMPA should be less prone to resistance development than the regular nucleoside analogs.

Lamivudine is different from the other anti-HIV nucleosides because it has the L-configuration. Natural nucleosides and most of their antiviral analogs possess the D-configuration. The biological activity of a nucleoside is dependent on the nature of the base moiety: Thus the aglycon moiety at position C1 is considered as absolutely necessary. As the nucleosides have to be phosphorylated to become active, the presence of a hydroxylmethyl group at position C4 is likewise a prerequisite for biological activity. Thus, the configuration of the C1 and C4 atoms of a nucleoside is of crucial importance. The difference between a D-nucleoside and an L-nucleoside in the dideoxy series is in fact not very pronounced. When the base moiety and the hydroxymethyl group are considered as reference points, the ring oxygen function is situated on the backside in D-nucleosides and on the front side in L-nucleosides, opposite to the C2–C3 bond (Fig. 12) [50]. D- and L-dideoxynucleosides overlap very well. Therefore, it is not surprising that antiviral activity has been found in the series of L-dideoxyribonucleoside (e.g., L-5-dideoxyribocytosine, L-fluorodideoxycytosine, L-oxathiolane series). These are mainly cytosine nucleosides, which may be a reflection of the kinases involved not being very selective. Indeed, similar antiviral activity of D- and L-dideoxynucleosides is not a general rule, as demonstrated by carbovir, where the antiviral activity is more dependent on the stereochemistry of the carbocyclic ring. Abacavir can be distinguished from the other anti-HIV nucleosides because it is a carbocyclic

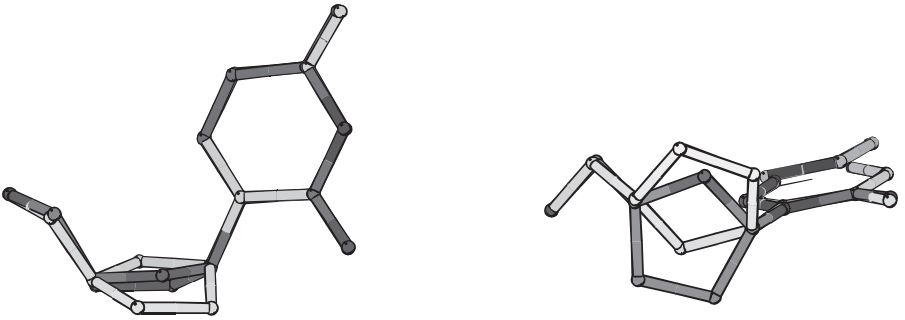

**Figure 12** Molecular overlap of D- and L-dideoxynucleosides showing the subtile difference in sugar orientation between both optical antipodes.

nucleoside analog. It can be considered as a prodrug of the phosphates of (–)carbovir, and a correlation has been demonstrated between carbovir triphosphate formation and anti-HIV activity. Abacavir, however, does not show the same toxicity as carbovir. It allows the nucleoside carbovir, which caused problems in animals, to be bypassed. The pharmacokinetic, distribution, and toxicity profile of abacavir is distinct from and improved over that of (–)carbovir. Structurally, abacavir belongs to the carbocyclic nucleosides, which means that the ring oxygen atom is replaced by a methylene group. As a result, the anomeric center is removed, as is the stereoelectronic influence of the ring oxygen atom. The structure–activity relationship (SAR) of carbocyclic nucleosides is therefore different from that of normal nucleosides.

The first phosphorylation step that converts the 2′,3′-dideoxynucleoside analogs to their 5′-monophosphate can be considered as the bottleneck in the overall metabolic pathway leading to the formation of the active 5′-triphosphate metabolites. If certain dideoxynucleoside analogs (e.g., 2′,3′-dideoxyuridine) are not active against HIV under conditions where others are, this stems from their poor, or lack of, phosphorylation at the nucleoside kinase level. Therefore, attempts have been made at constructing prodrugs of 2′,3′-dideoxynucleoside 5′-monophosphate that deliver the 5′-monophosphate derivatives on cellular uptake, which can then be converted to the corresponding 5′-di- and 5′-triphosphate derivatives. This approach thus bypasses the initial kinase dependency and has been concretized by the design of the phosphoramidate [51,52] and cyclosaligenyl [53,54] prodrugs of d4T 5′-monophosphate. Both prodrugs were found to efficiently deliver within the cells d4T 5′-monophosphate, which, after conversion to its 5′-triphosphate, afforded anti-HIV activity under conditions where the nucleoside (due to inefficient phosphorylation) did not. It remains to be established whether this nucleoside kinase bypass strategy also yields increased antiviral efficacy in vivo.

All the aforementioned 2′,3′-dideoxynucleoside analogs, in their 5′-triphosphate form, act as competitive substrates/inhibitors with respect to the natural substrates (dNTPs) at the catalytic site of HIV RT, and, as HBV uses a similar

RT in its life cycle, all these compounds may be expected to inhibit HBV replication as well. This premise has been borne out particularly for lamivudine, which is currently licensed for the treatment of chronic HBV infections. Such an extended activity spectrum cannot be anticipated for a second class of RT inhibitors, referred to as NNRTIs (for nonnucleoside reverse transcriptase inhibitors), which interact with an allosteric, nonsubstrate binding ("pocket") site on the RT of HIV-1. This binding pocket is located at about 10 Å distance from the substrate-binding site (Fig. 13) [55]. This pocket does not exist in unliganded RT and does not occur in RTs other than HIV-1 RT, or, if it does, only the HIV-1 RT pocket offers the required allowances for interactions with the NNRTIs: that is, stacking interactions with the aromatic amino acids Y181, Y188, W229, and Y318; electrostatic interactions with K101, K103, and E138; van der Waals interactions with L100, V106, Y181, G190, W229, L234, and Y318; and hydrogen bonding with the main-chain peptide bonds [56]. A model for the interaction of a representative NNRTI, the thiocarboxanilide UC781 (Fig. 2), with HIV RT is shown in Fig. 14 [57].

The NNRTIs are notorious for rapidly eliciting virus–drug resistance resulting from mutations at amino acid residues that surround the NNRTI-binding site. In the clinic, the most prominent mutations engendering resistance to NNRTIs are the K103N and Y181C mutations. While, at present, only three NNRTIs (i.e., nevirapine, delavirdine, and efavirenz) have been formally licensed, and several others, that is, UC781 [57], capravirine (S-1153, AG 1549) [58], dapivirine (TMC 125) [59], DPC083 [60], SJ-3366 [61], and ( + )-calanolide [62], are in clinical development, it is obvious that in the future design of new NNRTIs, not only potency and safety but also resilience to drug resistance mutations should come into play [63]. It is noteworthy that some amino acids, such as W229 and Y318 that surround the NNRTI binding site, do not seem apt to mutate, or if they do, they lead to a "suicidal" loss of RT activity [64]. Such immutable amino acids may be prime targets for the rational design of new NNRTIs.

## Acyclic Nucleoside Phosphonates

The acyclic nucleoside phosphonates can be conceived as acyclic nucleoside analogs extended by a phosphonate moiety. The phosphonate group is equi-

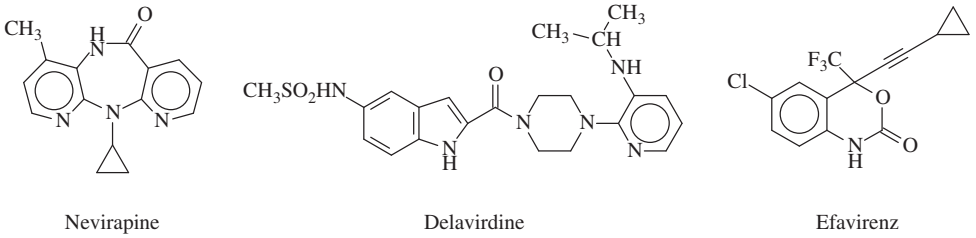


**Figure 13** Structure of nonnucleoside reverse transcriptase inhibitors.

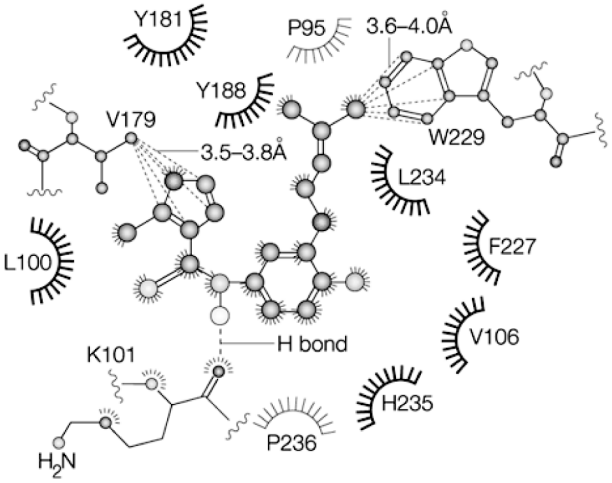


**Figure 14** Interaction of HIV-1 RT with UC781. Features stabilizing the complex between the human immunodeficiency virus 1 (HIV-1) reverse transcriptase (RT) and the nonnucleoside reverse transcriptase inhibitor UC781 (Fig. 2). The hydrogen bond with K101, and the two methyl group–aromatic ring interactions are shown explicitly. Other main hydrophobic contacts are shown with bold lines; minor ones are shown with faint lines [53]. Standard CPK coloring is used.

valent to a phosphate group, but unlike phosphate, phosphonate can no longer be cleaved through esterases that would normally convert nucleoside monophosphates back to their nucleoside form. Consequently, acyclic nucleoside phosphonates may be expected to show a broadened antiviral activity spectrum as compared to that of the acyclic nucleoside analogs (acyclovir, . . .) and dideoxynucleoside analogs (zidovudine, . . .). On the one hand, they should be active against those DNA viruses that do not encode for a specific viral thymidine kinase (TK) or protein kinase (PK) or have become resistant to the nucleoside analogs through TK or PK deficiency. On the other hand, they should also be able to bypass the nucleoside kinase reaction that limits the activity of the dideoxynucleoside analogs against retroviruses (HIV) and hepadnaviruses (HBV).

These premises were fulfilled on both scores, albeit by different types of acyclic nucleoside phosphonates: cidofovir (HPMPC), as to the broad-spectrum activity against DNA viruses; and adefovir (PMEA) and tenofovir (PMPA, Fig. 2), as to the activity against retro- and hepadnaviruses.

Although their eventual activity spectrum is different, both types of compounds share a common strategy in their mode of action: They both need two (instead of three) phosphorylation steps to be converted to their active (diphosphorylated) metabolites, which then act as chain terminators in the DNA polymerase reaction [HPMPC (Fig. 6*c*)] or reverse transcriptase reaction [PMEA, PMPA (Fig. 6*d*)]. For HPMPC to shut down viral DNA syn-

thesis, the incorporation of two consecutive HPMPC units is required [65], whereas for PMEA one such incorporation suffices [66]. In both cases, the acyclic nucleotides remain stably incorporated, presumably because the phosphonate group prevents repair enzymes excising these nucleotides.

The "era" of the acyclic nucleoside phosphonates started with the description of the broad-spectrum anti-DNA virus activity of the adenine derivative HPMPA [67]. These nucleotide analogs are isopolar and sterically similar to the natural nucleotides. It is rather surprising that none of the furanose nucleosides with a 5′-phosphonate moiety exhibit appreciable antiviral activity, which might be partly due to inefficient cellular uptake, partly due to diminished affinity for the metabolizing or target enzymes. The acyclic nucleotide analogs, however, demonstrate antiviral activity, although the structural window with respect to biological activity is rather narrow. Potent antiviral activity has been found mainly in the HPMP (hydroxypropoxymethylphosphonate) series and PME (phosphonomethoxyethyl) series. The hydroxyl group in the HPMP series may be replaced by a fluorine or hydrogen atom. The PME pyrimidine derivatives do not show inhibitory effect on viral multiplication. Phosphonomethylether nucleosides have unusual pharmacokinetic properties. They are very slow to pass biological membranes presumably due to the negatively charged phosphonate moiety. This charge also entails the low oral bioavailability of the phosphonates. Nucleosides phosphonates are not taken up into cells by the nucleoside transport mechanism. There limitations can successfully be addressed by utilizing a prodrug strategy, that is, pivaloyloxymethyl, isopropyloxycarbonyloxymethyl, or hexadexyloxypropyl esters. When chemically or enzymatically deacylated, the hydroxyalkyl phosphonate quickly decomposes to release the phosphonic acid and the corresponding aldehyde or ketone. Based on stability, solubility, and improved oral bioavailability, the bis(POC) derivative of tenofovir and the bis(POM) derivative of adefovir were selected for development and have now been approved for the treatment of HIV and HBV infections, respectively.

The cytosine compound, cidofovir (HPMPC), which appeared to be less harmful to the host in preliminary toxicity experiments, was developed as an antiviral drug [68] and approved for clinical use in the treatment of CMV retinitis in AIDS (acquired immunodeficiency syndrome) patients. Cidofovir also holds great potential for the treatment of several other DNA virus infections, namely TK-deficient HSV and VZV infections that are resistant to acyclovir (or brivudin), and herpesvirus infections at large (EBV, HHV-6, HHV-7, HHV-8), human papillomavirus (HPV) infections [i.e., pharyngeal, esophageal, and laryngeal papillomatosis, plantar and genital warts (condylomata acuminata), and cervical intraepithelial neoplasia], polyomavirus infections (i.e., progressive multifocal leukoencephalopathy), adenovirus infections (i.e., epidemic keratoconjunctivitis), poxvirus infections [i.e., smallpox, monkeypox, cowpox, orf, molluscum contagiosum (for the role of cidofovir in the treatment of poxvirus infections, see also Refs. 69–71)]. Adefovir and tenofovir, the two other protagonists from the acyclic nucleoside phosphonate

group, have in the mean time progressed, both in their oral prodrug forms, adefovir dipivoxil and tenofovir disoproxil (with fumarate added as salt), to final licensing for the treatment of HBV and HIV infections, respectively.

In contrast to all other antiviral drugs, acyclic nucleoside phosphonates have a particularly long intracellular half-life (one to several days), thus allowing infrequent dosing (only once daily for adefovir and tenofovir, or once weekly or every other week for cidofovir), and do not easily lead to resistance, even after prolonged treatment (for more than one or two years). No drug metabolic interactions are known for the acyclic nucleoside phosphonates, which means that they can readily be added onto any drug (combination) regimen, as has been particularly shown for tenofovir in the treatment of HIV infections.

## Inhibitors of Processes Associated with Viral RNA Synthesis

Gene expression (i.e., transcription to viral RNA) of the genome of retroviruses such as HIV is not possible without integration of the proviral DNA into the host chromosome (Fig. 1). Thus the enzyme involved—integrase—has been considered as an attractive target for chemotherapeutic intervention. Numerous integrase inhibitors have been described [72, 73]; of those none, however, with sufficient specificity to be further pursued as an integrase-targeted drug. The problem with integrase inhibitors is that, while they could be effective in an enzyme-based assay, their anti-HIV activity in cell culture may be masked by cytotoxicity. And even if selective anti-HIV activity in cell culture is noted, caution should be exercised in unconditionally attributing this activity to inhibition of the integration process, as the compounds concerned may well owe their anti-HIV activity to an action targeted elsewhere. This has proven to be the case for the anionic compounds zintevir [7] and L-chicoric acid [74], two integrase inhibitors that owe their anti-HIV activity primarily to an interaction with the viral envelope gp120, and thus fall in the category of the polyanionic inhibitors of virus adsorption (see Section 2 of "Statistics in Pharmaceutical Safety Assessment"). Up till now, the only compounds that qualify as genuine integrase inhibitors are diketo acid derivatives (i.e., L-731,988 and L-708,906) [75–77] and pyranodipyrimidine (PDP) derivatives (i.e., V-165) [78]. These compounds were found to inhibit HIV-1 replication in cell culture, on the one hand, and to inhibit the strand transfer function of the integrase (the other catalytic function of the enzyme being endonucleolytic cleavage of the terminal dinucleotide GT from the 3′-end), on the other hand. These two events could be causally linked as mutations in the HIV-1 integrase conferred resistance to the inhibitory effects of the compounds on both strand transfer and HIV-1 infectivity [75]. Clinical trials with the diketo acid derivatives (S-1360 and L-870180) have recently been initiated.

At the transcription level, HIV gene expression may be inhibited by compounds that interact with cellular factors that bind to the long terminal repeat

(LTR) promoter and that are needed for basal-level transcription, such as NF-κB inhibitors [79], but greater specificity may be expected from those compounds that specifically inhibit the transactivation of the HIV LTR promoter by the viral trans-acting transactivator (Tat) protein. The Tat protein interacts specifically with a responsive element, called TAR, located at the beginning of the viral messenger RNA (mRNA) transcribed from the LTR promotor, thereby enhancing ("transactivating") the transcription process. Several compounds have been described as inhibitors of the transcription process, for example, fluoroquinolines [80] and bistriazoloacridone derivatives such as temacrazine [81]. The latter was found to block HIV-1 RNA transcription starting from the HIV LTR promoter without interfering with the transcription of any cellular genes. The peptide analog CGP64222, which is structurally reminiscent of the amino acid 48–56 sequence RKKRRQRRR of the Tat protein, was originally designed to act as a Tat antagonist [82]. However, although CGP64222 is able to interact with the Tat-driven transcription process, its anti-HIV activity in cell culture is mediated primarily by an interaction with CXCR4, the co-receptor for X4 HIV strains [83].

Viral RNA transcription could also be affected by targeting cyclin-dependent kinases (cdks), which are required for the replication of many viruses (including HIV). Indeed, flavopiridol, a typical inhibitor of cdks (particularly cdk9, which is involved in the Tat-driven transcription process), has proven to be effective in blocking HIV infectivity [84].

Capping and methylation of HIV pre-mRNAs are coupled to the elongation by polymerase II. Binding of the capping enzyme and cap methyltransferase to polymerase II depends on phosphorylation of its carboxyl-terminal domain. It has been recently demonstrated that the co-transcriptional capping of HIV mRNA is stimulated by the Tat protein, consequently to its binding to the capping enzyme [85]. These findings implicate capping as an elongation checkpoint critical to HIV gene expression and thus corroborate earlier observations that *S*-adenosylmethionine-dependent methylations play an important role in the Tat-dependent transactivation of transcription from the LTR promoter [86]. They also offer an explanation for the inhibitory effects observed with *S*-adenosylhomocysteine hydrolase inhibitors (see Section 8 of "Statistics in Pharmaceutical Safety Assessment") on Tat-dependent transactivation and HIV replication [86].

One of the virus infections in the greatest need of antiviral therapy is HCV, and, here, two specific enzymatic functions associated with viral RNA synthesis could be envisaged as targets for the design of new antiviral agents, namely the nonstructural protein 3 (NS3)-associated helicase and the nonstructural protein 5B (NS5B) RNA-dependent RNA polymerase. Crystal structures of both enzymes are available [87–89], and both enzymatic activities have been characterized in sufficient detail [90, 91] to facilitate the development of effective HCV chemotherapeutics. For the helicase, there is no precedent, but for the RNA polymerase there is, and the experience gathered from the studies with the HIV RT

inhibitors may be of paramount importance when targeting the HCV RNA polymerase, especially if, as it appears [91], this enzyme shows similar kinetics to the HIV RT. Meanwhile, the first indications of compounds targeted at the NS5B polymerase (i.e., the benzimidazole derivative JTK-003) have appeared [89].

## 5 VIRAL PROTEASE INHIBITORS

Viral proteases play a critical role in the life cycle of many different viruses, including retroviruses such as HIV, herpesviruses, picornaviruses (rhino), and flaviviruses such as HCV, and hence viral proteases have been favored as targets for antiviral agents [92]. Their role is essentially based upon the cleavage of a newly expressed precursor polyprotein into smaller, mature viral proteins, termed functional (if endowed with enzymatic activity) or structural (if being part of the virion structure). For example, in HIV replication, HIV protease cleaves the gag and gag-pol precursor proteins to the structural proteins (p17, p24, p9, p7) and functional proteins (protease, reverse transcriptase/RNase H, and integrase). HIV protease inhibitors have been tailored to the peptidic linkages (e.g., F-P, F-L, and F-T) in the gag and gag-pol precursor proteins that are cleaved by the protease, and have been extensively modeled in the active site of the enzyme. All protease inhibitors that are currently licensed for the treatment of HIV infection, namely saquinavir, ritonavir, indinavir, nelfinavir, amprenavir, and lopinavir, share the same structural determinant (Fig. 2), that is, an hydroxyethylene core (instead of the normal peptidic linkage) that makes them nonscissile, "peptidomimetic," substrate analogs for the HIV protease. The hydroxyethylene hallmark is also present in the azadipeptide atazanavir (BMS-232632) [93], expected to be licensed soon for the treatment of HIV infection.

The HIV protease belongs to the aspartyl enzymes and has a dimeric structure (Fig. 15) [94]. The monomeric units, both identical 99-residue polypeptides, each contribute symmetrically to the catalytic site, and the enzyme is a $C_2$-symmetric dimer. The protease is unique in its specificity, being able to hydrolyze almost any peptide bond, given an optimal distribution of amino acids in $P_4$ through $P_{4'}$. The hallmark of the aspartyl proteases is the presence of two catalytic β-carboxyl groups from the side chains of two aspartyl residues brought into close proximity by the protein fold. In the case of the dimeric HIV-1 protease, each monomer chain donates a single Asp residue at position 25 in the polypeptide chain ($Asp_{25}$ and $Asp_{25'}$). One of the Asp is more acidic ($pK_a$ 3.4 to 3.7) than the other Asp ($pK_a$ 5.5 to 6.5). The HIV protease participates in general acid–general base catalysis where one of the catalytic Asp is protonated and the other is not. The nucleophile in catalysis is, most probably, an activated water molecule. The mechanism is given is Scheme 1. The amide hydrate is considered as the transition state. This mechanism is relevant for inhibitor design, that is, making HIV-1 protease inhibitors as transition-state mimics (nonhydrolysable isosters with tetrahedral geometry).

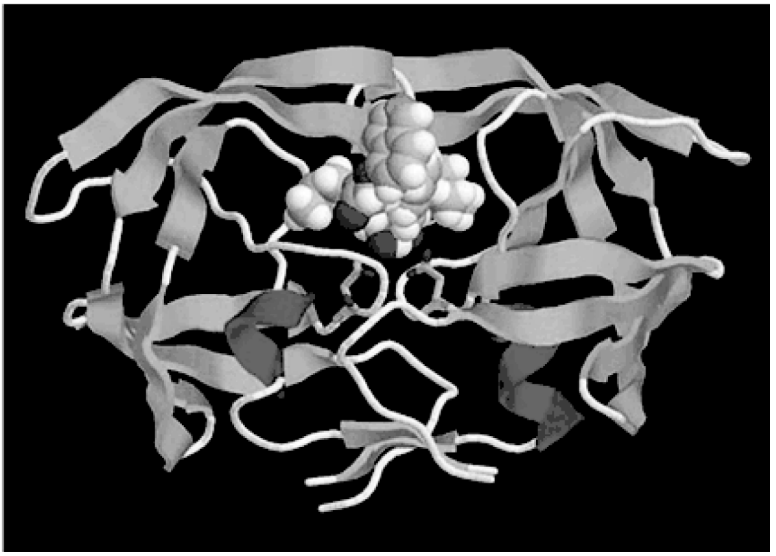

**Figure 15** Interaction of HIV protease with KNI272. Ribbon diagram of human immunodefiency virus (HIV) protease complexed with the peptidomimetic protease inhibitor KNI272; derived from the crystal structure [94]. The inhibitor is shown as a space-filling model, and the two active-site aspartic acids are shown as sticks; both have standard CPK coloring.

In the enzyme, the catalytic triads, $Asp_{25}$-$Thr_{26}$-$Gly_{27}$, and $Asp_{25'}$-$Thr_{26'}$-$Gly_{27'}$ lie on the floor of the cavity with the catalytic $Asp_{25}$ and $Asp_{25'}$ in near coplanar orientation. The ceiling of the catalytic site is composed of a pair of six-amino-acid segments (-$Ile_{47}$-Gly-Gly-Ile-Gly-$Gly_{52}$ . . .) and of the corresponding region 47′–52′, in the other monomer, which forms the flap. The flaps are flexible structures and they play an important role in substrate and inhibitor binding. Upon binding the inhibitors, the HIV-1 protease undergoes a conformational change, where the flaps move from an open conformation to a conformation where the ligand is embraced. A water molecule with approximate tetrahedral coordination bridges the $P_2$ and $P_{1'}$ carbonyl groups of the inhibitor to the amide groups of $Ile_{50}$ and $Ile_{50'}$ of the flaps via hydrogen bonds. The hydroxyethylene group of the transition-state inhibitor is normally aligned with the catalytic carboxylates of $Asp_{25}$ and $Asp_{25'}$.

The core structure for tight binding of an inhibitor consists of four side chains corresponding to the substrate residues $P_2$–$P_{2'}$. In the first protease inhibitor that has reached the market, saquinavir, it was observed that the hydroxyethylamine transition-state mimetic was readily adapted to the $P_{1'}$ proline containing substrate sequence and that the activity of the inhibitor could be markedly improved in potency when the amino acid proline at the $P_{1'}$ site was replaced by the (*S*,*S*,*S*)-decahydro-isoquinoline-3-carbonyl group.

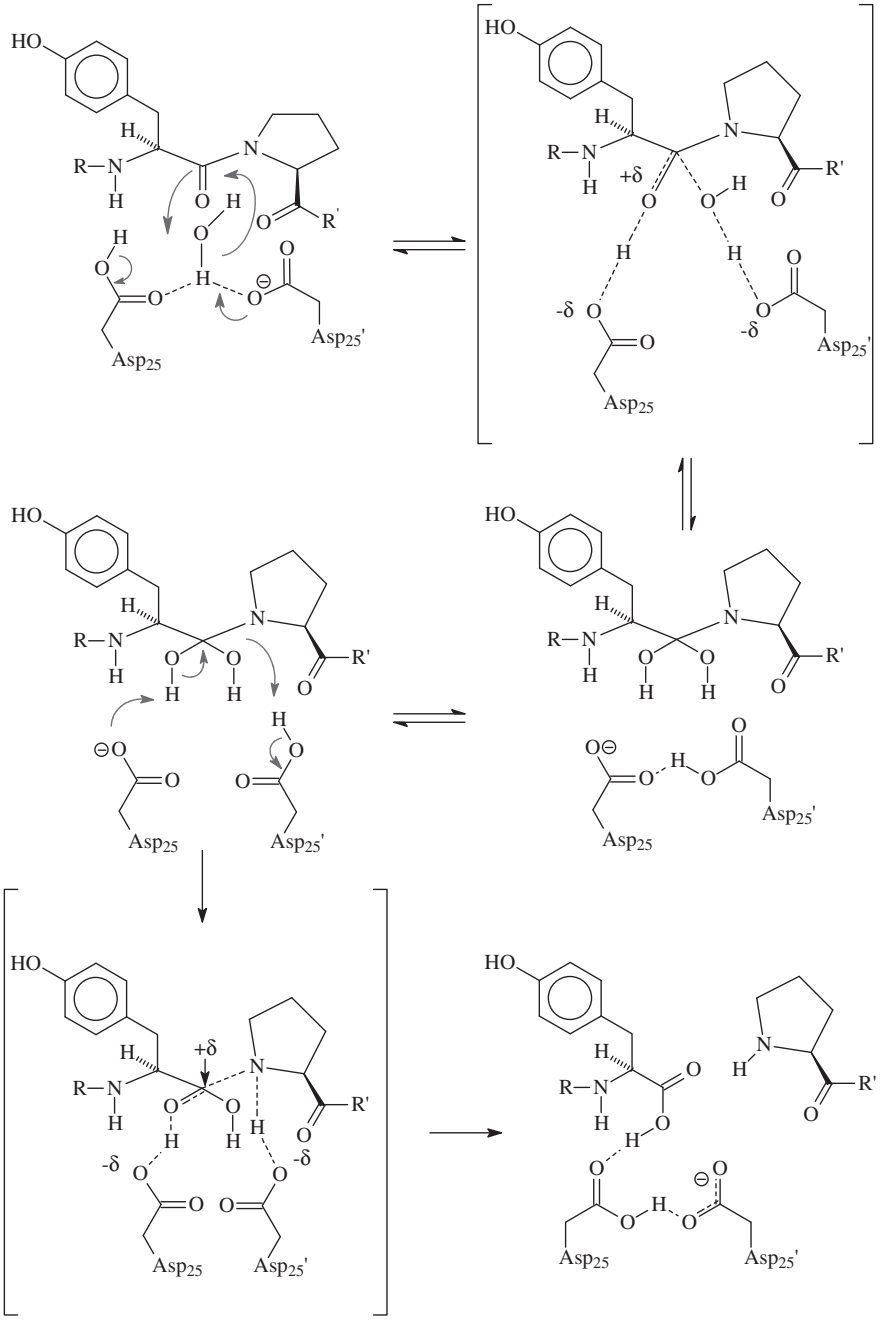


**Scheme 1** Proposed chemical mechanism of the peptidolytic reaction of HIV-1 protease based on kinetic and structural data. Proposed hydrogen bonds are indicated by dashed lines and are based in part on molecular modeling of the enzyme-intermediate complex [96].

An interesting example of de novo drug design is the cyclic urea mozenavir (DMP-450) as a nonpeptidic inhibitor of HIV-1 protease [95], where the structural water linking the flap residues $Ile_{50}$ and $Ile_{50'}$ is displaced (Fig. 16). This is an energetically favorable process. The carbonyl group replacing the structural water molecule is part of a cyclic urea because this function is an excellent hydrogen bond acceptor. The diol moiety is hydrogen binding to $Asp_{25}$ and $Asp_{25'}$. The other substituent ensures high complementarity with the enzyme subsites. However, the virus was able to mount considerable resistance against this compound, which was not further developed, also because of poor oral bioavailability in humans and highly variable blood levels. Mozenavir is, perhaps, one of the nicest examples of rational drug design, attesting that the most "rational" approaches may fail at the end because of (unpredictable) lack of clinical relevance.

The HIV protease inhibitors can be designed using a stepwise approach of lead optimization. This is demonstrated here by the discovery of saquinavir. The discovery of saquinavir started with the identification of a protected dipeptide as a weak enzyme inhibitor. An increase in potency was obtained by introduction of a P2 asparagine residue (Fig. 17). The replacement of

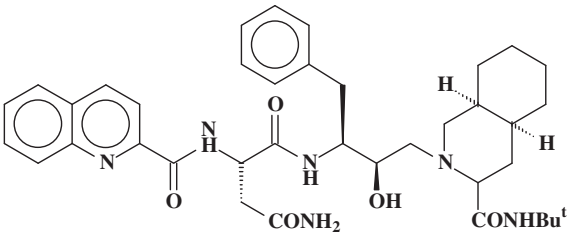


Saquinavir

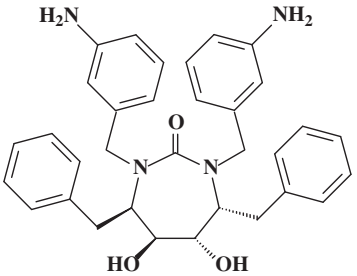


Mozenavir

**Figure 16** Structure of saquinavir and DPM-323. Saquinavir has been developed based on a systematic optimization of a core structure. DPM-323 is the result of a rational design approach, but the compound never reached the market because of failure in late-stage development.

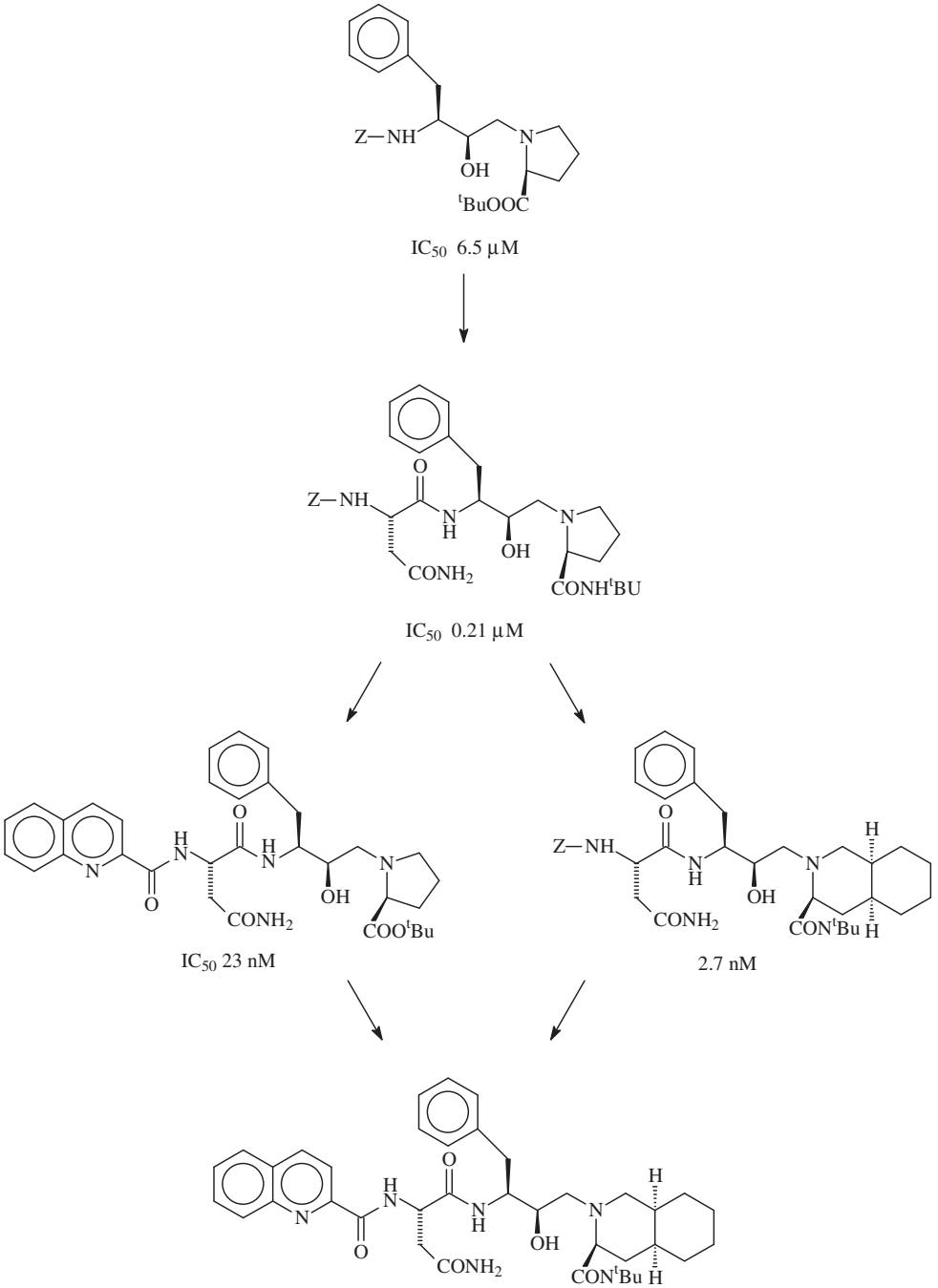


**Figure 17** Lead optimization of a dipeptide protease inhibitor leading to the discovery of saquinavir.

the Z group (in the C-terminal $^{t}$Butylester analog) by a 2-quinolinyl group gave a compound that inhibited HIV-1 at nanomolar concentrations (~20 nM). A systematic investigation of proline replacement led to the discovery of the decahydroisoquinoline substituent as optimal C-terminal ($IC_{50}$: 2.7 nM). An increase in potency was obtained by combining both substituents, leading to the synthesis of saquinavir, which inhibits both HIV-1 and HIV-2 with $IC_{50}$ values <1 nM. The compound has a low but sufficient oral biovailability.

The HIV protease inhibitors (PIs) have proven to be valuable therapeutics in drug combination schedules [HAART (highly active antiretroviral therapy)] with NRTIs and NNRTIs, in the treatment of HIV infections. Yet, they are met by compounding factors such as difficulties in drug adherence, drug–drug interactions, overlapping resistance patterns, and long-term side effects such as lipodystrophy and cardiovascular and metabolic disturbances (i.e., diabetes). This has prompted the search for new, nonpeptidic inhibitors of HIV protease, with (instead of the peptidomimetic hydroxyethylene core) cyclic urea, 4-hydroxycoumarin, L-mannaric acid, or 4-hydroxy-5,6-dihydro-2-pyrone (Fig. 2) [as in tipranavir (PNU-140690)] as the central scaffold [97, 98]. Such compounds should show little, if any, cross-resistance with the peptidomimetic inhibitors [99]; their in vivo efficacy, pharmacokinetics, and short- and long-term safety profiles remain to be established.

Whether the PI approach would be as successful for tackling herpes-, picorna-, and flaviviruses, as it turned out to be for HIV, remains to be seen. While the HIV protease is an "aspartate" protease, herpesvirus proteases belong to the "serine" proteases with SHH as the catalytic triad [100]. Several nonpeptidic inhibitors of herpesviral protease (i.e., CMV protease, which is also referred to as assemblin because of its role in the CMV assembly process) have been described, for example, thieno[2,3-d]oxazinones [101], aryl hydroxylamine derivatives [102], monobactams [103], pyrrolidine-5,5-*trans*-lactams [104], and 1,4-dihydroxynaphthalene and naphthoquinones [105]. While a useful exercise at targeting the herpesviral protease, all these efforts should be viewed as a prelude to further investigations on the in vitro and in vivo inhibitory effects of these compounds on virus replication.

Further advanced is the structure-assisted design of mechanism-based, irreversible, inhibitors of human rhinovirus 3C protease (a "cysteine" protease that is involved in the proteolytic cleavage of the viral precursor polyprotein to both capsid and functional proteins required for RNA replication). This work has yielded a wealth of compounds with potent activity against multiple rhinovirus serotypes [106–111]. Of the human rhinovirus 3C protease inhibitors, AG7088, which was shown to inhibit rhinovirus replication even when added late in the virus life cycle [112], and ruprintivir [113] have proceeded to clinical trials for the prevention and treatment of the common cold.

The HCV protease is a "serine" protease that is encoded by the nonstructural NS3 protease domain and is responsible for the proteolytic cleavage of

the nonstructural NS3, NS4A, NS4B, NS5A, and NS5B from the viral polyprotein (the NS4A protein then binds to the NS3 protein and enhances its proteolytic activity). The HCV NS3-4A protease is remarkably similar to the pestiviral NS3-4A protease [as found in BVDV (bovine viral diarrhea virus)] [114] and has been intensively pursued as a target for the design of inhibitors. Again, as for the herpesviral serine protease, several inhibitors, both peptide-based [115, 116], nonpeptidic [117], and macrocyclic [89], of the HCV NS3-4A protease have been identified, but as there is no cell culture assay available for HCV, their activity against HCV infectivity could not be assessed. Given the similarities of the HCV and BVDV NS3-4A proteases, it would seem justifiable to evaluate putative HCV protease inhibitors for their activity against BVDV replication, for which cell culture assay systems have been established. In addition, HCV protease inhibitors, as well as HCV helicase and polymerase could and should also be evaluated in the subgenomic replicon system [118].

## 6 VIRAL NEURAMINIDASE INHIBITORS

Influenza virus (both A and B) has adopted a unique replication strategy in using one of its surface glycoproteins, hemagglutinin, to bind to the target cell receptor [containing a terminal sialic acid (*N*-acetylneuraminic acid, NANA)], and the other surface glycoprotein, neuraminidase, to leave the cell after the viral replicative cycle has been completed. The viral neuraminidase is thus needed for the elution of the newly formed virus particles from the cells. In addition, the viral neuraminidase may also promote viral movement through the respiratory tract mucus, thus enhancing viral infectivity. Therefore, the influenza viral neuraminidase has been envisaged as a suitable target for the design of specific inhibitors. Neuraminidase (sialidase) is a tetrameric glycoprotein with a subunit molecular weight of 60 kDa. The enzyme is a glycohydrolase cleaving α-ketosidically linked terminal *N*-acetyl-D-neuraminic acid residues on the mucosal sialoglycoproteins (Fig. 18).

An interesting observation is that 18 of the residues in the active site of influenza A and B sialidase sequences that have been characterized, are conserved. Fifteen of the conserved residues are charged. Eight of the strain-invariant active-site residues are positioned to make direct contact with *N*-acetyl-D-neuraminic acid bound into the catalytic site, with a further tier of 10 residues that appears to establish a structural scaffold for the catalytic site. A stabilized sialosyl cation transition-state intermediate is involved when influenza virus sialidase catalyzes the release of *N*-acetyl-D-neuraminic acid from α-sialosides. *N*-acetyl-D-neuraminic acid (as an anomeric mixture) inhibits influenza A virus neuraminidase with a $K_i$ of $10^{-3}$ M. More potent inhibitors are those based on the unsaturated transition-state analogs with a glycal structure (2-deoxy-2,3-didehydro-*N*-acetylneuraminic acid) (Fig. 19).

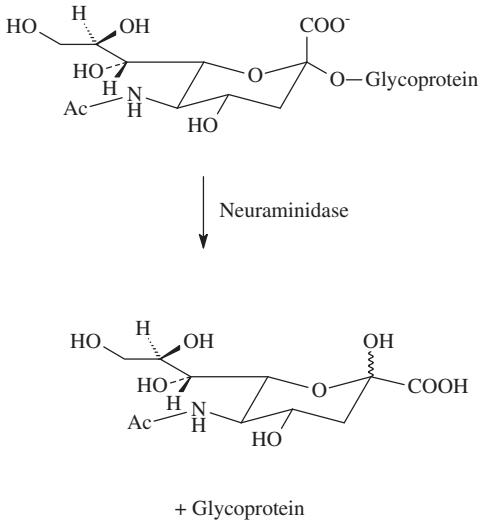


**Figure 18** The glycohydrolase activity of neuraminidase.

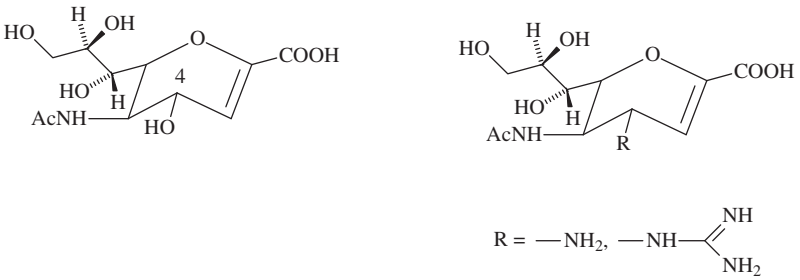


**Figure 19** Unsaturated transition-state analogs that inhibit influenza A virus neuraminidase.

One of the most important interactions between *N*-acetyl-D-neuraminic acid and its unsaturated congener with the enzyme is that between the carboxylate of both of these sialic acids and a triargininyl cluster ($Arg_{118}$, $Arg_{292}$, and $Arg_{371}$). This interaction is important both enzymatically—on binding, the sugar ring is flattened out prior to the cleavage of the glycosidic bond—and structurally, the electrostatic interaction between the negatively charged carboxylate and the positively charged amine groups of the arginines being a major contributor to the binding of substrates and substrate-based inhibitors. Based on modeling experiments, a series of 4-substituted-4-deoxy-*N*-acetyl-D-neuraminic acid derivatives were synthesized. The 4-amino analog has a $K_i$ of $4 \times 10^{-8}$ M against influenza virus neuraminidase, which represents an increase in binding affinity when compared with 2-deoxy-2,3-didehydro-*N*-

acetylneuraminic acid ($K_i$ of $4 \times 10^{-6}$ M). The replacement of the amino group with the larger more basic guanidino functionality led to an even more potent inhibitor ($K_i$ of $3 \times 10^{-11}$ M). The guanidino substituent is situated within hydrogen bond distance with two conserved acid residues. Zanamivir selectively inhibits influenza virus sialidase, while other sialidases from both bacterial and mammalian sources are not inhibited. The second approach toward new sialidase inhibitors has been concentrated on the synthesis of carbocyclic mimetics and the introduction of a lipophilic substituent in place of the glycerol side chain, which led to very potent orally active compounds (i.e., oseltamivir) for the treatment of influenza virus infections.

Computer-assisted drug design, led to the identification of zanamivir (GG167) as a specific and potent inhibitor of the enzyme and of the in vitro and in vivo replication of both influenza A and B virus [119]. Zanamivir was tailored to interact with the conserved amino acid residues within the active site of influenza A and B virus neuraminidase, and its inhibitory effect on the enzyme has proved to be predictive of the susceptibility of clinical isolates to the drug [120]. Meanwhile, zanamivir was shown to be efficacious and safe in the treatment (by inhalation) of influenza virus infections [121], and the drug has been licensed for clinical use.

Zanamivir has to be given by inhalation because of its poor oral bioavailability. In attempts to identify potentially orally bioavailable inhibitors, a series of carbocyclic transition-state-based analogs were developed in which the polar glycerol and guanidino groups (as present in zanamivir) were replaced by a lipophilic (3-pentyloxy) side chain and amino group, respectively, to give GS 4071 [122]. X-ray crystallographic studies showed that these groups of GS 4071 could be accommodated within the active site of the neuraminidase (Fig. 20). As aimed for, GS 4071, when administered as the ethyl ester prodrug (GS 4104), turned out to be orally bioavailable and was found to be effective in protecting mice and ferrets against influenza infection [123]. Subsequently, GS 4104 (oseltamivir) was found to be effective and safe in the oral treatment and prevention of influenza virus infections [124, 125] and has been licensed for clinical use.

Zanamivir and oseltamivir have paved the way for the development of similarly, structure-based, designed neuraminidase inhibitors, such as the cyclopentane derivative RWJ-270201 (with a comparable, or even better, efficacy profile in the murine influenza model) [126–128] and the pyrrolidine derivatives A-192558 and A-315675 [129–131]. The clinical potential of RWJ-270201, A-192558, and/or A-315675 in the prevention and/or treatment of influenza virus infections in humans still needs to be established.

## 7 IMP DEHYDROGENASE INHIBITORS

IMP dehydrogenase is a key enzyme in the de novo biosynthesis of purine mononucleotides: It is responsible for the NAD-dependent oxidation of IMP

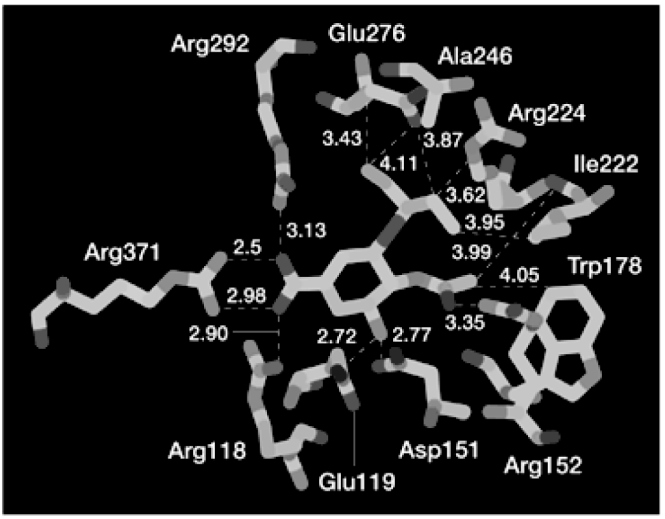


**Figure 20** Interaction of influenza neuraminidase with oseltamivir. The picture shows binding interactions of the neuraminidase inhibitor oseltamivir (Fig. 2) with influenza neuraminidase, derived from the crystal structure.

(inosine 5′-monophosphate) to XMP (xanthosine 5′-monophosphate), which is then further converted to GMP, GDP, and GTP, and from GDP, via dGDP, also to dGTP. Inhibitors of IMP dehydrogenase may be expected to affect both RNA and DNA synthesis, via reduction of the intracellular pools of GTP and dGTP, respectively. Although IMP dehydrogenase is a cellular target, inhibitors targeted at this enzyme may be expected to preferentially inhibit viral RNA and/or DNA synthesis, as there is an increased need for such syntheses in virus-infected cells.

IMP dehydrogenase can be targeted by two types of inhibitors: competitive or uncompetitive with respect to the normal substrate, IMP. To the first category belongs ribavirin, which has been officially approved for clinical use as an aerosol for the treatment of RSV infections and in combination with interferon-α for the treatment of HCV infections.

To the second category belongs mycophenolic acid (Fig. 2), an immunosuppressing agent, which has been approved as its morpholinoethyl ester prodrug, for the prevention of acute allograft rejection following kidney transplantation. The X-ray crystal structure of IMP dehydrogenase, complexed with mycophenolic acid at the active site, has been reported at high resolution (2.6 Å) [132]. New congeners of both ribavirin (Fig. 2) (i.e., EICAR) [133] and mycophenolic acid (i.e., VX-497) [134] have an activity spectrum as broad as ribavirin, but considerably greater potency. This activity spectrum encompasses both DNA and RNA viruses, and among the latter, picorna-, toga-, flavi-, bunya-, arena-, reo-, rhabdo-, and, particularly, ortho- and paramyxoviruses.

Mycophenolic acid has marked activity against yellow fever virus and, in addition, markedly potentiates the inhibitory effects of acyclic guanosine analogs (acyclovir, penciclovir, ganciclovir) against HSV, VZV, and CMV infections, which could be of great clinical utility in organ transplant recipients subject to these infections [135]. Also, mycophenolic acid potentiates the activity of guanine-derived dideoxynucleoside analogs, such as abacavir, against HIV [136], which may be further exploited as a new combination strategy in the treatment of HIV infections.

While IMP dehydrogenase inhibitors should, in their own right, be further explored for their potential in the treatment of various (+)RNA and (–)RNA virus infections, including hemorrhagic fever virus infections, current interest is mainly focused on their use in combination with (pegylated) interferon-α in the treatment of HCV infections. This stems from the successful responses that have been seen following treatment of chronic hepatitis C with ribavirin in combination with interferon-α (in patients who did not respond to interferon alone) [137].

Recently, ribavirin was shown to act as an RNA virus mutagen, forcing RNA viruses into a lethal accumulation of errors, dubbed "error catastrophe" [138, 139]. The antiviral activity of ribavirin may then result from the lethal mutagenic effect following incorporation of ribavirin into the viral genome, and, obviously, this lethal mutagenesis may be enhanced by the inhibitory effect of ribavirin (in its 5′-monophosphate form) on IMP dehydrogenase and the consequent decrease in cellular GTP pools (as mentioned above). The ability of ribavirin to force RNA viruses into error catastrophe has so far been shown only with poliovirus [138, 139], and it remains to be verified whether the theory also holds for other RNA viruses (i.e., HCV) and other ribavirin analogs (i.e., EICAR).

## 8 *S*-ADENOSYLHOMOCYSTEINE HYDROLASE INHIBITORS

*S*-adenosyl-L-homocysteine (AdoHcy, SAH) hydrolase is a key enzyme in methylation reactions depending on *S*-adenosylmethionine (SAM) as the methyl donor, including those methylation reactions that are required for the maturation of viral mRNA. In particular, (–)RNA viruses are critically dependent on these methylations for the stability and functioning of their mRNA. All capped, methylated structures consist of a N7-methyl guanosine residue linked at the 5′-hydroxyl group to the 5′-end of the mRNA strand by a triphosphate linkage. Most 5′-capped methylated structures also contain a methyl group on the 2′-hydroxy group of the penultimate nucleotide (Fig. 21). This 5-capped, methylated structure has been shown to protect mRNA form 5′-end nuclease digestion. Methylation of the capped structure increases the affinity for ribosome binding to the 5′-end of this mRNA during formation of the translational initiation complex. Uncapped or undermethylated viral mRNA is less effectively translated into viral proteins. Both

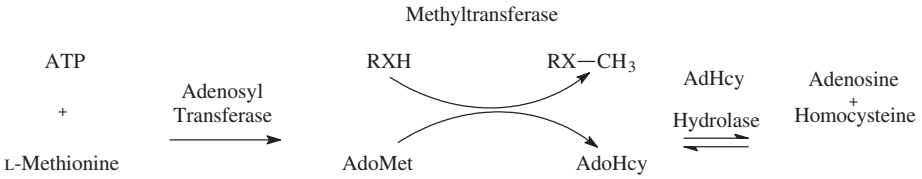


**Figure 21** Structure of capped mRNA showing the methylated terminal guanine and the methylated 2′-hydroxyl groups.

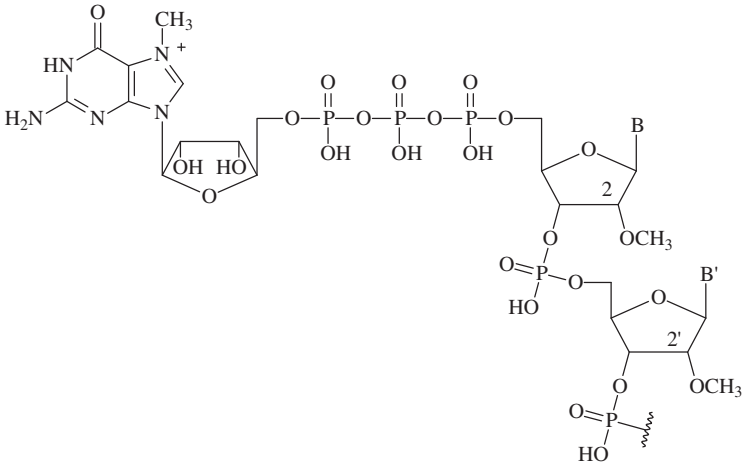


**Figure 22** Adenosylmethione (synthesized from ATP and L-methionine) functions as methyldonor in a methyltransferase reaction and is converted to adenosyl homocysteine. AdoHcy hydrolase hydrolyses AdoHcy to adenine and homocysteine.

guanine-7 methyltransferase and nucleoside-2′ methyltransferase coded by vaccinia virus require AdoMet as a methyl donor (Fig. 22). SAH is both a product and inhibitor of the methyltransferase reactions; however, SAH is rapidly hydrolyzed by SAH hydrolase into homocysteine and adenosine, and this prevents the accumulation of SAH that would otherwise lead to an inhibition of the SAM-dependent methylation reactions.

Mammalian AdoHcy hydrolase is a homotetramer containing one $NAD^+$ per subunit. The first step in the enzymatic reaction involves oxidation of the 3′-hydroxyl group of *S*-adenosyl-L-homocysteine to form 3′-keto-AdoHcy, resulting in the conversion of $NAD^+$ to NADH. The 3′-keto group increases the acidity of the C-4′ proton, allowing for abstraction of this proton by a base in the active site of the enzyme. Subsequently, β-elimination of homocysteine results in the formation of the intermediate 3′-keto-4′,5′-didehydro-5′-deoxyadenosine. Addition of water to the 5′-position of this intermediate affords 3′-keto-adenosine, which is then reduced by the enzyme-bound NADH, resulting in the formation of adenosine and regeneration of $NAD^+$.

Inhibitors of the SAH hydrolase may be expected to lead to an accumulation of SAH, and consequent inhibition of the methylation reactions. Again, as noted for IMP dehydrogenase, SAH hydrolase is a cellular target, but as virus replication increases the need for such methylations, SAH hydrolase inhibitors may confer selective antiviral activity that may well vary from one virus to another depending on their individual methylation needs.

Various adenosine analogs, that is, carbocyclic adenosine, carbocyclic 3-deazaadenosine, neplanocin A, 3-deazaneplanocin A, and their 5′-nor derivatives, have been described as potent inhibitors of SAH hydrolase [140]. Carbocyclic adenosine and replanocin A belong to the first generation of AdoHcy hydrolase inhibitors. Carbocyclic adenosine is a reversible, competitive inhibitor with a $K_i$ of 5 nM. Neplanocin A, in contrast, is an irreversible, tightly binding inhibitor with a $K_i$ of 8.4 nM. Neplanocin A inactivates the hydrolase by a "cofactor depletion" mechanism, converting the $NAD^+$ cofactor to its inactive form (NADH) with simultaneous oxidation of neplanocin A to 3′-keto-neplanocin A. This 3′-keto derivative is tightly bound to the NADH form of the enzyme (Fig. 23). Unfortunately, inhibitors of the first generation have a common problem of cellular toxicity, which precludes the clinical use of these compounds as antiviral agents. In attempts to design more specific inhibitors of AdoHcy hydrolase, two different approaches have been followed. One approach is the replacement of the adenine moiety with 3-deazaadenine, resulting in 3-deazaaristeromycin and 3-deazaneplanocin A. Another approach involves removing of the 4′-hydroxymethyl substituent, which would preclude 5′-phosphorylation by adenosine kinase and deamination by adenosine deaminase (Fig. 24).

These compounds are indeed not substrates for either adenosine deaminase and adenosine kinase, but they retain potent inhibitory activity against AdoHcy hydrolase. Inhibition of AdoHcy hydrolase by 3-deazaneplanocin and 3-deazaaristeromycin was reported to be reversible and competitive, with $K_i$ values of 0.05 and 3 nM, respectively. In contrast, the 5′-nor derivatives have been shown to inactivate AdoHcy hydrolase irreversibly. These inhibitors serve as substrates for the enzyme's oxidative activity, resulting in reduction of the enzyme-bound $NAD^+$ to NADH. The reaction stops at this point since these inhibitors are not substrate for the "hydrolytic activity" of the enzyme. These second-generation AdoHcy hydrolase inhibitors exhibit

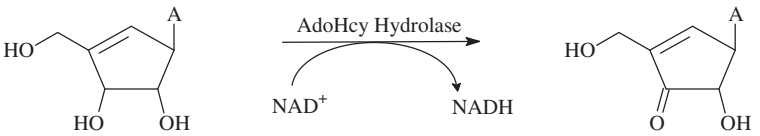


**Figure 23** Enzymatic oxidation of neplanocin to its 3′-keto derivative with concommitant reduction of $NAD^+$ to NADH. The 3′-keto derivative of neplanocin is bond to the NADH coenzyme of SAH hydrolase.

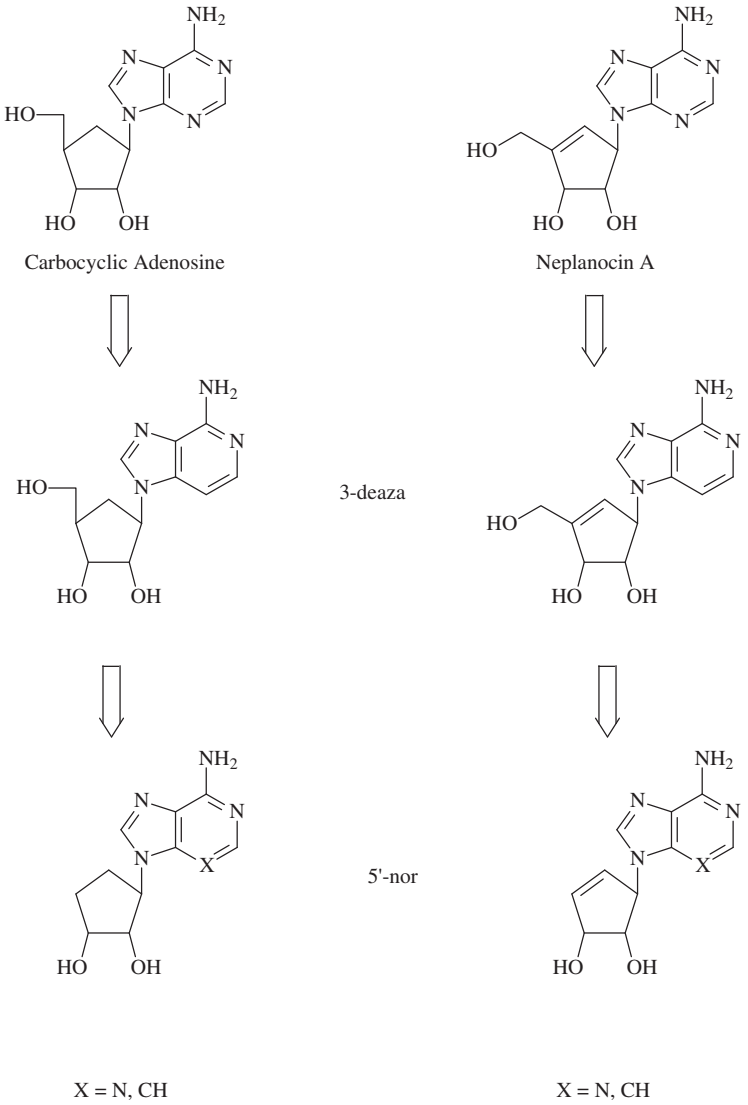


**Figure 24** Approaches followed to discover new SAH hydrolase inhibitors, i.e., replacement of the N-3 atom of the adenine base by CH and removal of the 4′-hydroxy methyl group of the sugar moiety.

broad-spectrum antiviral activity, whereas their cytotoxicity is considerably lower than that of the parent compounds. A third type of mechanism-base inhibitors of AdoHcy hydrolase are defined as compounds that use the hydrolytic activity of the enzyme to convert a prodrug to a potent drug in the active site of the enzyme, from which (*E*)-5′,6′-didehydro-6′-deoxy-6′-halo-homoadenosines are representative examples [141].

All SAH hydrolase inhibitors possess a characteristic antiviral activity spectrum, encompassing, in particular, poxviruses, (±)RNA viruses (reo) and (−)RNA viruses (bunya-, arena-, rhabdo-, filo-, ortho-, and paramyxoviruses). This includes a number of hemorrhagic fever viruses such as Ebola hemorrhagic fever. In fact, a mouse model for Ebola hemorrhagic fever has been developed [142], and the SAH hydrolase inhibitors carbocyclic 3-deazaadenosine [143] and 3-deazaneplanocin A [144] were found to protect the animals against an otherwise lethal Ebola virus infection. SAH hydrolase inhibitors thus offer real potential for the treatment of hemorrhagic fever virus infections.

## 9 CONCLUSION

The strategies reviewed here for interfering with the key events in the viral replicative cycle have the potential to target virtually all important human viral pathogens. Several such strategies, namely those aimed at viral DNA synthesis, viral polyprotein cleavage, and viral release from the cells (through the viral neuraminidase) have already provided a number of effective and useful therapeutics for the treatment of herpesvirus (HSV-1, HSV-2, VZV, CMV), retrovirus (HIV), hepadna (HBV), and influenza virus infections. Further improvements along these lines, following the same approaches, may in the future, yield more efficacious and more selective antiviral compounds. This should by no means detract from other approaches, not addressed here, that may also be envisaged to target viral compounds or virus-associated events, such as agents that specifically bind to the picornaviral capsids (i.e., pleconaril), or the HIV nucleocapsid p7 [i.e., 2,2′-dithiobisbenzamides (DIBAs)], or the influenza virus A matrix [adamantanamine derivatives (i.e., amantadine)], or glycosylation inhibitors (i.e., deoxynojirimycin derivatives), or antisense oligonucleotides or ribozymes targeted at selected viral mRNAs, or gene therapy approaches, or immunotherapy, and so forth. Whatever approach or strategy is followed to combat virus infections, the highest profit is likely to be expected if two or more of these different strategies are combined, especially in the treatment of chronic virus infections such as HIV, HBV, and HCV.

***Acknowledgments*** We thank Christiane Callebaut and Chantal Biernaux for their invaluable editorial assistance.

## REFERENCES

1. De Clercq, E. (2001). Antiviral drugs: Current state of the art. *J. Clin. Virol.*, *22*, 73–89.
2. Gallaher, W. R., Ball, J. M., Garry, R. F., Martin-Amedee, A. M., Montelaro, R. C. (1995). A general model for the surface glycoproteins of HIV and other retroviruses. *AIDS Res. Hum. Retrovir.*, *11*, 191–202.

3. Shukla, D., Liu, J., Blaiklock, P., Shworak, N. W., Bai, X., Esko, J. D., Cohen, G. H., Eisenberg, R. J., Rosenberg, R. D., Spear, P. G. (1999). A novel role for 3-*O*-sulfated heparan sulfate in herpes simplex virus 1 entry. *Cell*, *99*, 13–22.
4. Chen, Y., Maguire, T., Hileman, R. E., Fromm, J. R., Esko, J. D., Linhardt, R. J., Marks, R. M. (1997). Dengue virus infectivity depends on envelope protein binding to target cell heparan sulfate. *Nat. Med.*, *3*, 866–871.
5. Herold, B. C., Bourne, N., Marcellino, D., Kirkpatrick, R., Strauss, D. M., Zaneveld, L. J., Waller, D. P., Anderson, R. A., Chany, C. J., Barham, B. J., Stanberry, L. R., Cooper, M. D. (2000). Poly(sodium 4-styrene sulfonate): An effective candidate topical antimicrobial for the prevention of sexually transmitted diseases. *J. Infect. Dis.*, *181*, 770–773.
6. Esté, J. A., Schols, D., De Vreese, K., Van Laethem, K., Vandamme, A. M., Desmyter, J., De Clercq, E. (1997). Development of resistance of human immunodeficiency virus type 1 to dextran sulfate associated with the emergence of specific mutations in the envelope gp120 glycoprotein. *Mol. Pharmacol.*, *52*, 98–104.
7. Esté, J. A., Cabrera, C., Schols, D., Cherepanov, P., Gutierrez, A., Witvrouw, M., Pannecouque, C., Debyser, Z., Rando, R. F., Clotet, B., Desmyter, J., De Clercq, E. (1998). Human immunodeficiency virus glycoprotein gp120 as the primary target for the antiviral action of AR177 (zintevir). *Mol. Pharmacol.*, *53*, 340–345.
8. Cabrera, C., Witvrouw, M., Gutierrez, A., Clotet, B., Kuipers, M. E., Swart, P. J., Meijer, D. K., Desmyter, J., De Clercq, E., Esté, J. A. (1999). Resistance of the human immunodeficiency virus to the inhibitory action of negatively charged albumins on virus binding to CD4. *AIDS Res. Hum. Retrovir.*, *15*, 1535–1543.
9. Schols, D., Struyf, S., Van Damme, J., Esté, J. A., Henson, G., De Clercq, E. (1997). Inhibition of T-tropic HIV strains by selective antagonization of the chemokine receptor CXCR4. *J. Exp. Med.*, *186*, 1383–1388.
10. De Clercq, E. (2000). Inhibition of HIV infection by bicyclams, highly potent and specific CXCR4 antagonists. *Mol. Pharmacol.*, *57*, 833–839.
11. Baba, M., Nishimura, O., Kanzaki, N., Okamoto, M., Sawada, H., Iizawa, Y., Shiraishi, M., Aramaki, Y., Okonogi, K., Ogawa, Y., Meguro, K., Fujino, M. (1999). A small-molecule, nonpeptide CCR5 antagonist with highly potent and selective anti-HIV-1 activity. *Proc. Natl. Acad. Sci. USA*, *96*, 5698–5703.
12. Dragic, T., Trkola, A., Thompson, D. A., Cormier, E. G., Kajumo, F. A., Maxwell, E., Lin, S. W., Ying, W., Smith, S. O., Sakmar, T. P., Moore, J. P. (2000). A binding pocket for a small molecule inhibitor of HIV-1 entry within the transmembrane helices of CCR5. *Proc. Natl. Acad. Sci. USA*, *97*, 5639–5644.
13. Bridger, G. J., Skerlj, R. T., Thornton, D., Padmanabhan, S., Martellucci, S. A., Henson, G. W., Abrams, M. J., Yamamoto, N., De Vreese, K., Pauwels, R., De Clercq, E. (1995). Synthesis and structure-activity relationships of phenylenebis(methylene)-linked bis-tetraazamacrocycles that inhibit HIV replication. Effects of macrocyclic ring size and substituents on the aromatic linker. *J. Med. Chem.*, *38*, 366–378.
14. Bridger, G. J., Skerlj, R. T., Padmanabhan, S., Martellucci, S. A., Henson, G. W., Abrams, M. J., Joao, H. C., Witvrouw, M., De Vreese, K., Pauwels, R., De Clercq, E. (1996). Synthesis and structure-activity relationships of phenylenebis(methylene)-linked bis-tetraazamacrocycles that inhibit human immunodeficiency virus

replication. 2. Effect of heteroaromatic linkers on the activity of bicyclams. *J. Med. Chem.*, *39*, 109–119.

15. Bridger, G. J., Skerlj, R. T., Padmanabhan, S., Martellucci, S. A., Henson, G. W., Struyf, S., Witvrouw, M., Schols, D., De Clercq, E. (1999). Synthesis and structure-activity relationships of phenylenebis(methylene)-linked bis-azamacrocycles that inhibit HIV-1 and HIV-2 replication by antagonism of the chemokine receptor CXCR4. *J. Med. Chem.*, *42*, 3971–3981.
16. Hatse, S., Princen, K., Gerlach, L. O., Bridger, G., Henson, G., De Clercq, E., Schwartz, T. W., Schols, D. (2001). Mutation of $Asp^{171}$ and $Asp^{262}$ of the chemokine receptor CXCR4 impairs its coreceptor function for human immunodeficiency virus-1 entry and abrogates the antagonistic activity of AMD3100. *Mol. Pharmacol.*, *60*, 164–173.
17. Palczewski, K., Kumasaka, T., Hori, T., Behnke, C. A., Motoshima, H., Fox, B. A., Le Trong, I., Teller, D. C., Okada, T., Stenkamp, R. E., Yamamoto, M., Miyano, M. (2000). Crystal structure of rhodopsin: A G protein-coupled receptor. *Science*, *289*, 739–745.
18. Bridger, G. J., Skerlj, R. T., Thornton, D., Padmanabhan, S., Martellucci, S. A., Henson, G. W., Abrams, M. J., Yamamoto, N., De Vreese, K., Pauwels, R., De Clercq, E. (1995). Synthesis and structure-activity relationships of phenylenebis(methylene)-linked bis-tetraazamacrocycles that inhibit HIV replication. Effects of macrocyclic ring size and substituents on the aromatic linker. *J. Med. Chem.*, *38*, 366–378.
19. Joao, H. C., De Vreese, K., Pauwels, R., De Clercq, E., Henson, G. W., Bridger, G. J. (1995). Quantitative structural activity relationship study of bis-tetraazacyclic compounds. A novel series of HIV-1 and HIV-2 inhibitors. *J. Med. Chem.*, *38*, 3865–3873.
20. McAuley, A., Subramanian, S., Whitcombe, T. W. (1987). Synthesis and X-ray crystal-structure of a C-spirbi-[cyclam nickel (II)] complex (cyclam = 1,4,8,11-tetraazacyclotetradecane). *J. Chem. Soc.*, *8*, 539–541.
21. Strizki, J. M., Xu, S., Wagner, N. E., Wojcik, L., Liu, J., Hou, Y., Endres, M., Palani, A., Shapiro, S., Clader, J. W., Greenlee, W. J., Tagat, J. R., McCombie, S., Cox, K., Fawzi, A. B., Chou, C. C., Pugliese-Sivo, C., Davies, L., Moreno, M. E., Ho, D. D., Trkola, A., Stoddart, C. A., Moore, J. P., Reyes, G. R., Baroudy, B. M. (2001). SCH-C (SCH 351125), an orally bioavailable, small molecule antagonist of the chemokine receptor CCR5, is a potent inhibitor of HIV-1 infection *in vitro* and *in vivo*. *Proc. Natl. Acad. Sci. USA*, *98*, 12718–12723.
22. Root, M. J., Kay, M. S., Kim, P. S. (2001). Protein design of an HIV-1 entry inhibitor. *Science*, *291*, 884–888.
23. Eckert, D. M., Malashkevich, V. N., Hong, L. H., Carr, P. A., Kim, P. S. (1999). Inhibiting HIV-1 entry: Discovery of D-peptide inhibitors that target the gp41 coiled-coil pocket. *Cell*, *99*, 103–115.
24. Kilby, J. M., Hopkins, S., Venetta, T. M., DiMassimo, B., Cloud, G. A., Lee, J. Y., Alldredge, L., Hunter, E., Lambert, D., Bolognesi, D., Matthews, T., Johnson, M. R., Nowak, M. A., Shaw, G. M., Saag, M. S. (1998). Potent suppression of HIV-1 replication in humans by T-20, a peptide inhibitor of gp41-mediated virus entry. *Nat. Med.*, *4*, 1302–1307.
25. Lalezari, J. P., Henry, K., O'Hearn, M., Montaner, J. S. G., Piliero, P. J., Trottier, B., Walmsley, S., Cohen, C., Kuritzkes, D. R., Eron, J. J. Jr., Chung, J., DeMasi, R.,

Donatacci, L., Drobnes, C., Delehanty, J., Salgo, M. (2003). Enfuvirtide, an HIV-1 fusion inhibitor, for drug-resistant HIV infection in North and South America. *N. Engl. J. Med.*, *348*, 2175–2185.

26. Lazzarin, A., Clotet, B., Cooper, D., Reynes, J., Arastéh, K., Nelson, M., Katlama, C., Stellbrink, H.-J., Delfraissy, J.-F., Lange, J., Huson, L., DeMasi, R., Wat, C., Delehanty, J., Drobnes, C., Salgo, M. (2003). Efficacy of enfuvirtide in patients infected with drug-resistant HIV-1 in Europe and Australia. *N. Engl. J. Med.*, *348*, 2186–2195.
27. Lambert, D. M., Barney, S., Lambert, A. L., Guthrie, K., Medinas, R., Davis, D. E., Bucy, T., Erickson, J., Merutka, G., Petteway, S. R. Jr. (1996). Peptides from conserved regions of paramyxovirus fusion (F) proteins are potent inhibitors of viral fusion. *Proc. Natl. Acad. Sci. USA*, *93*, 2186–2191.
28. Schwartz, J. A., Lium, E. K., Silverstein, S. J. (2001). Herpes simplex virus type 1 entry is inhibited by the cobalt chelate complex CTC-96. *J. Virol.*, *75*, 4117–4128.
29. Champness, J. N., Bennett, M. S., Wien, F., Visse, R., Summers, W. C., Herdewijn, P., De Clercq, E., Ostrowski, T., Jarvest, R. L., Sanderson, M. R. (1998). Exploring the active site of herpes simplex virus type-1 thymidine kinase by X-ray crystallography of complexes with aciclovir and other ligands. *Proteins: Structure, Functions and Genetics*, *32*, 350–361.
30. De Winter, H., Herdewijn, P. (1997). Understanding the binding of 5-substituted 2′-deoxyuridine substrates to thymidine kinase of herpes simplex virus type-1. *J. Med. Chem.*, *39*, 4727–4737.
31. Wigerinck, P., Pannecouque, C., Snoeck, R., Claes, P., De Clercq, E., Herdewijn, P. (1991). 5-(5-Bromothien-2-yl)-2′-deoxyuridine and 5-(5-chlorothien-2-yl)-2′-deoxyuridine are equipotent to (*E*)-5-(2-bromovinyl)-2′-deoxyuridine in the inhibition of herpes simplex virus type I replication. *J. Med. Chem.*, *34*, 2383–2389.
32. Wigerinck, P., Snoeck, R., Claes, P., De Clercq, E., Herdewijn, P. (1991). Synthesis and antiviral activity of 5-heteroaryl-substituted 2′-deoxyuridines. *J. Med. Chem.*, *34*, 1767–1772.
33. Wigerinck, P., Kerremans, L., Claes, P., Snoeck, R., Maudgal, P. C., De Clercq, E., Herdewijn, P. (1993). Synthesis and antiviral activity of 5-thien-2-yl-2′-deoxyuridine analogues. *J. Med. Chem.*, *36*, 538–543.
34. Bohman, C., Balzarini, J., Wigerinck, P., Van Aerschot, A., Herdewijn, P., De Clercq, E. (1994). Mechanism of cytostatic action of novel 5-(thien-2-yl)- and 5-(furan-2-yl)-substituted pyrimidine nucleoside analogues against tumor cells transfected by the thymidine kinase gene of herpes simplex virus. *J. Biol. Chem.*, *269*, 8036–8043.
35. McGuigan, C., Yarnold, C. J., Jones, G., Velazquez, S., Barucki, H., Brancale, A., Andrei, G., Snoeck, R., De Clercq, E., Balzarini, J. (1999). Potent and selective inhibition of varicella-zoster virus (VZV) by nucleoside analogues with an unusual bicyclic base. *J. Med. Chem.*, *42*, 4479–4484.
36. McGuigan, C., Barucki, H., Carangio, A., Blewett, S., Andrei, G., Snoeck, R., De Clercq, E., Balzarini, J., Erichsen, J. T. (2000). Highly potent and selective inhibition of varicella-zoster virus by bicyclic furopyrimidine nucleosides bearing an aryl side chain. *J. Med. Chem.*, *43*, 4993–4997.
37. Iwayama, S., Ono, N., Ohmura, Y., Suzuki, K., Aoki, M., Nakazawa, H., Oikawa, M., Kato, T., Okunishi, M., Nishiyama, Y., Yamanishi, K. (1998). Antiherpesvirus

activities of (1′*S*,2′*R*)-9-[[1′,2′-bis(hydroxymethyl)-cycloprop-1′-yl]methyl] (A-5021) in cell culture. *Antimicrob. Agents Chemother.*, *42*, 1666–1670.

38. Ono, N., Iwayama, S., Suzuki, K., Sekiyama, T., Nakazawa, H., Tsuji, T., Okunishi, M., Daikoku, T., Nishiyama, Y. (1998). Mode of action of (1′*S*,2′*R*)-9-[[1′,2′-bis(hydroxymethyl)cycloprop-1′-yl]methyl]guanine (A-5021) against herpes simplex virus type 1 and type 2 and varicella-zoster virus. *Antimicrob. Agents Chemother.*, *42*, 2095–2102.
39. Wang, J., Froeyen, M., Hendrix, C., Andrei, G., Snoeck, R., De Clercq, E., Herdewijn, P. (2000). The cyclohexene ring system as a furanose mimic: Synthesis and antiviral activity of both enantiomers of cyclohexenylguanine. *J. Med. Chem.*, *43*, 736–745.
40. Crute, J. J., Grygon, C. A., Hargrave, K. D., Simoneau, B., Faucher, A. M., Bolger, G., Kibler, P., Liuzzi, M., Cordingley, M. G. (2002). Herpes simplex virus helicase-primase inhibitors are active in animal models of human disease. *Nature Med.*, *8*, 386–391.
41. Kleymann, G., Fischer, R., Betz, U. A., Hendrix, M., Bender, W., Schneider, U., Handke, G., Eckenberg, P., Hewlett, G., Pevzner, V., Baumeister, J., Weber, O., Henninger, K., Keldenich, J., Jensen, A., Kolb, J., Bach, U., Popp, A., Maben, J., Frappa, I., Haebich, D., Lockhoff, O., Rubsamen-Waigmann, H. (2002). New helicase-primase inhibitors as drug candidates for the treatment of herpes simplex disease. *Nature Med.*, *8*, 392–398.
42. Betz, U. A. K., Fischer, R., Kleymann, G., Hendrix, M., Rubsamen-Waigmann, H. (2002). Potent in vivo antiviral activity of the herpes simplex virus primase-helicase inhibitor BAY 57–1293. *Antimicrob. Agents Chemother.*, *46*, 1766–1772.
43. Rousseau, F. S., Kahn, J. O., Thompson, M., Mildvan, D., Shepp, D., Sommadossi, J. P., Delehanty, J., Simpson, J. N., Wang, L. H., Quinn, J. B., Wakeford, C., van der Horst, C. (2001). Prototype trial design for rapid dose selection of antiretroviral drugs: An example using emtricitabine (Coviracil). *J. Antimicrob. Chemother.*, *48*, 507–513.
44. Furman, P. A., Jeffrey, J., Kiefer, L. L., Feng, J. Y., Anderson, K. S., Borroto-Esoda, K., Hill, E., Copeland, W. C., Chu, C. K., Sommadossi, J. P., Liberman, I., Schinazi, R. F., Painter, G. R. (2001). Mechanism of action of 1-β-D-2,6-diaminopurine dioxolane, a prodrug of the human immunodeficiency virus type 1 inhibitor 1-β-D-dioxolane guanosine. *Antimicrob. Agents Chemother.*, *45*, 158–165.
45. Stoddart, C. A., Moreno, M. E., Linquist-Stepps, V. D., Bare, C., Bogan, M. R., Gobbi, A., Buckheit, R. W. Jr., Bedard, J., Rando, R. F., McCune, J. M. (2000). Antiviral activity of 2′-deoxy-3′-oxa-4′-thiocytidine (BCH-10652) against lamivudine-resistant human immunodeficiency virus type 1 in SCID-hu Thy/Liv mice. *Antimicrob. Agents Chemother.*, *44*, 783–786.
46. Otto, M. J., Arastèh, K., Schulbin, H., Beard, A., Cartee, L., Liotta, D. C., Schinazi, R. F., Murphy, R. L. (2002). Single and multiple dose pharmacokinetics and safety of the nucleoside Racivir® in male volunteers. HIV DART 2002. *Front. Drug Devel.* for Antiretroviral Therapies, Naples, Florida, USA, 15–19 December 2002. *Antiviral Res.*, *Abst.*, no. 044.
47. Schinazi, R. F., Mellors, J., Bazmi, H., Diamond, S., Garber, S., Gallagher, K., Geleziunas, R., Klabe, R., Pierce, M., Rayner, M., Wu, J. T., Zhang, H., Hammond, J., Bacheler, J., Manion, D. J., Otto, M. J., Stuyver, L., Trainor, G., Liotta, D. C.,

Erickson-Viitanen, S. (2002). DPC 817: A cytidine nucleoside analog with activity against zidovudine- and lamivudine-resistant viral variants. *Antimicrob. Agents Chemother.*, *46*, 1394–1401.

48. Dunkle, L. M., Gathe, J. C., Pedevillano, D. E., Oshana, S. C., Pottage, J. C., Rice, W. C. (2002). ACH-126,443 (β-L-Fd4C) shows potent anti-HIV activity in proof-of-princple study in treatment-experienced patients with M184V and other reverse transcriptase mutations. HIV DART 2002, Frontiers in Drug Development for Antiretroviral Therapies, Naples, Florida, USA, 15–19 December 2002. *Antiviral Res.*, *Abst.*, no. 069.
49. Katlama, C., Ghosn, J., Tubiana, R., Wirden, M., Valantin, M. A., Harmenberg, J., Mardh, G., Westling, C., Calvez, V. (2002). MIV-310 reduces HIV viral load in patients failing multiple antiretroviral therapy: Results from a 4-week phase II study. HIV DART 2002, Naples, Florida, USA, 15–19 December 2002. Abst., no. 071.
50. Herdewijn, P. (1997). Structural requirements for antiviral activity in nucleosides. *Drug Discov. Today*, *2*, 235–242.
51. Siddiqui, A. Q., McGuigan, C., Ballatore, C., Zuccotto, F., Gilbert, I. H., De Clercq, E., Balzarini, J. (1999). Design and synthesis of lipophilic phosphoramidate d4T-MP prodrugs expressing high potency against HIV in cell culture: Structural determinants for in vitro activity and QSAR. *J. Med. Chem.*, *42*, 4122–4128.
52. Saboulard, D., Naesens, L., Cahard, D., Salgado, A., Pathirana, R., Velazquez, S., McGuigan, C., De Clercq, E., Balzarini, J. (1999). Characterization of the activation pathway of phosphoramidate triester prodrugs of stavudine and zidovudine. *Mol. Pharmacol.*, *56*, 693–704.
53. Meier, C., Lorey, M., De Clercq, E., Balzarini, J. (1998). *cyclo*Sal-2′,3′-dideoxy-2′,3′-didehydrothymidine monophosphate (*cyclo*Sal-d4TMP): Synthesis and antiviral evaluation of a new d4TMP delivery system. *J. Med. Chem.*, *41*, 1417–1427.
54. Balzarini, J., Aquaro, S., Knispel, T., Rampazzo, C., Bianchi, V., Perno, C. F., De Clercq, E., Meier, C. (2000). Cyclosaligenyl-2′,3′-didehydro-2′,3′-dideoxythymidine monophosphate: Efficient intracellular delivery of d4TMP. *Mol. Pharmacol.*, *58*, 928–935.
55. De Clercq, E. (1998). The role of non-nucleoside reverse transcriptase inhibitors (NNRTIs) in the therapy of HIV-1 infection. *Antiviral Res.*, *38*, 153–179.
56. Jonckheere, H., Anné, J., De Clercq, E. (2000). The HIV-1 reverse transcription (RT) process as target for RT inhibitors. *Med. Res. Rev.*, *20*, 129–154.
57. Esnouf, R. M., Stuart, D. I., De Clercq, E., Schwartz, E., Balzarini, J. (1997). Models which explain the inhibition of reverse transcriptase by HIV-1-specific (thio)carboxanilide derivatives. *Biochem. Biophys. Res. Commun.*, *234*, 458–464.
58. Fujiwara, T., Sato, A., el-Farrash, M., Miki, S., Abe, K., Isaka, Y., Kodama, M., Wu, Y., Chen, L. B., Harada, H., Sugimoto, H., Hatanaka, M., Hinuma, Y. (1998). S-1153 inhibits replication of known drug-resistant strains of human immunodeficiency virus type 1. *Antimicrob. Agents Chemother.*, *42*, 1340–1345.
59. Andries, K., de Béthune, M.-P., Kukla, M. J., Azijn, H., Lewi, P. J., Janssen, P. A. J., Pauwels, R. (2000). R165335-TMC125, a novel non nucleoside reverse transcriptase inhibitor (NNRTI) with nanomolar activity against NNRTI resistant HIV

strains. Abstracts of the 40th Interscience Conference on Antimicrobial Agents and Chemotherapy, Toronto, Canada, 17–20 September 2000, no. 1840.

60. Corbett, J. W., Ko, S. S., Rodgers, J. D., Jeffrey, S., Bacheler, L. T., Klabe, R. M., Diamond, S., Lai, C. M., Rabel, S. R., Saye, J. A., Adams, S. P., Trainor, G. L., Anderson, P. S., Erickson-Viitanen, S. K. (1999). Expanded-spectrum nonnucleoside reverse transcriptase inhibitors inhibit clinically relevant mutant variants of human immunodeficiency virus type 1. *Antimicrob. Agents Chemother.*, *43*, 2893–2897.
61. Buckheit, R. W. Jr., Watson, K., Fliakas-Boltz, V., Russell, J., Loftus, T. L., Osterling, M. C., Turpin, J. A., Pallansch, L. A., White, E. L., Lee, J. W., Lee, S. H., Oh, J. W., Kwon, H. S., Chung, S. G., Cho, E. H. (2001). SJ-3366, a unique and highly potent nonnucleoside reverse transcriptase inhibitor of human immunodeficiency virus type 1 (HIV-1) that also inhibits HIV-2. *Antimicrob. Agents Chemother.*, *45*, 393–400.
62. Creagh, T., Ruckle, J. L., Tolbert, D. T., Giltner, J., Eiznhamer, D. A., Dutta, B., Flavin, M. T., Xu, Z. Q. (2001). Safety and pharmacokinetics of single doses of (+)-calanolide a, a novel, naturally occurring nonnucleoside reverse transcriptase inhibitor, in healthy, human immunodeficiency virus-negative human subjects. *Antimicrob. Agents Chemother.*, *45*, 1379–1386.
63. Ren, J., Nichols, C., Bird, L. E., Fujiwara, T., Sugimoto, H., Stuart, D. I., Stammers, D. K. (2000). Binding of the second generation non-nucleoside inhibitor S-1153 to HIV-1 reverse transcriptase involves extensive main chain hydrogen bonding. *J. Biol. Chem.*, *275*, 14316–14320.
64. Pelemans, H., Esnouf, R., De Clercq, E., Balzarini, J. (2000). Mutational analysis of Trp-229 of human immunodeficiency virus type 1 reverse transcriptase (RT) identifies this amino acid residue as a prime target for the rational design of new non-nucleoside RT inhibitors. *Mol. Pharmacol.*, *57*, 954–960.
65. Xiong, X., Smith, J. L., Chen, M. S. (1997). Effect of incorporation of cidofovir into DNA by human cytomegalovirus DNA polymerase on DNA elongation. *Antimicrob. Agents Chemother.*, *41*, 594–599.
66. Balzarini, J., Hao, Z., Herdewijn, P., Johns, D. G., De Clercq, E. (1991). Intracellular metabolism and mechanism of anti-retrovirus action of 9-(2-phosphonylmethoxyethyl)adenine, a potent anti-human immunodeficiency virus compound. *Proc. Natl. Acad. Sci. U.S.A.*, *88*, 1499–1503.
67. De Clercq, E., Holý, A., Rosenberg, I., Sakuma, T., Balzarini, J., Maudgal, P. C. (1986). A novel selective broad-spectrum anti-DNA virus agent. *Nature*, *323*, 464–467.
68. De Clercq, E. (1993). Therapeutic potential of HPMPC as an antiviral drug. *Rev. Med. Virol.*, *3*, 85–96.
69. De Clercq, E. (2001). Vaccinia virus inhibitors as a paradigm for the chemotherapy of poxvirus infections. *Clin. Microbiol. Rev.*, *14*, 382–397.
70. De Clercq, E. (2002). Cidofovir in the treatment of poxvirus infections. *Antiviral Res.*, *55*, 1–13.
71. De Clercq, E. (2002). Cidofovir in the therapy and short-term prophylaxis of poxvirus infections. *Trends Pharmacol. Sci.*, *23*, 456–458.
72. Pommier, Y., Pilon, A. A., Bajaj, K., Mazumder, A., Neamati, N. (1997). HIV-1 integrase as a target for antiviral drugs. *Antiviral Chem. Chemother.*, *8*, 463–483.

73. Pommier, Y., Marchand, C., Neamati, N. (2002). Retroviral integrase inhibitors year 2000: Update and perspectives. *Antiviral Res.*, *47*, 139–148.
74. Pluymers, W., Neamati, N., Pannecouque, C., Fikkert, V., Marchand, C., Burke, T. R. Jr., Pommier, Y., Schols, D., De Clercq, E., Debyser, Z., Witvrouw, M. (2000). Viral entry as the primary target for the anti-HIV activity of chicoric acid and its tetra-acetyl esters. *Mol. Pharmacol.*, *58*, 641–648.
75. Hazuda, D. J., Felock, P., Witmer, M., Wolfe, A., Stillmock, K., Grobler, J. A., Espeseth, A., Gabryelski, L., Schleif, W., Blau, C., Miller, M. D. (2000). Inhibitors of strand transfer that prevent integration and inhibit HIV-1 replication in cells. *Science*, *287*, 646–650.
76. Wai, J. S., Egbertson, M. S., Payne, L. S., Fisher, T. E., Embrey, M. W., Tran, L. O., Melamed, J. Y., Langford, H. M., Guare, J. P. Jr., Zhuang, L., Grey, V. E., Vacca, J. P., Holloway, M. K., Naylor-Olsen, A. M., Hazuda, D. J., Felock, P. J., Wolfe, A. L., Stillmock, K. A., Schleif, W. A., Gabryelski, L. J., Young, S. D. (2000). 4-Aryl-2,4-dioxobutanoic acid inhibitors of HIV-1 integrase and viral replication in cells. *J. Med. Chem.*, *43*, 4923–4926.
77. Pluymers, W., Pais, G., Van Maele, B., Pannecouque, C., Fikkert, V., Burke, T. R. Jr., De Clercq, E., Witvrouw, M., Neamati, N., Debyser, Z. (2000). Inhibition of human immunodeficiency virus type 1 integration by diketo derivatives. *Antimicrob. Agents Chemother.*, *46*, 3292–3297.
78. Pannecouque, C., Pluymers, W., Van Maele, B., Tetz, V., Cherepanov, P., De Clercq, E., Witvrouw, M., Debyzer, Z. (2002). New class of HIV integrase inhibitors that block viral replication in cell culture. *Curr. Biol.*, *12*, 1169–1177.
79. Daelemans, D., Vandamme, A.-M., De Clercq, E. (1999). Human immunodeficiency virus gene regulation as a target for antiviral chemotherapy. *Antiviral Chem. Chemother.*, *10*, 1–14.
80. Baba, M., Okamoto, M., Kawamura, M., Makino, M., Higashida, T., Takashi, T., Kimura, Y., Ikeuchi, T., Tetsuka, T., Okamoto, T. (1998). Inhibition of human immunodeficiency virus type 1 replication and cytokine production by fluoroquinoline derivatives. *Mol. Pharmacol.*, *53*, 1097–1103.
81. Turpin, J. A., Buckheit, R. W. Jr., Derse, D., Hollingshead, M., Williamson, K., Palamone, C., Osterling, M. C., Hill, S. A., Graham, L., Schaeffer, C. A., Bu, M., Huang, M., Cholody, W. M., Michejda, C. J., Rice, W. G. (1998). Inhibition of acute-, latent-, and chronic-phase human immunodeficiency virus type 1 (HIV-1) replication by a bistriazoloacridone analog that selectively inhibits HIV-1 transcription. *Antimicrob. Agents Chemother.*, *42*, 487–494.
82. Hamy, F., Felder, E. R., Heizmann, G., Lazdins, J., Aboul-ela, F., Varani, G., Karn, J., Klimkait, T. (1997). An inhibitor of the Tat/TAR RNA interaction that effectively suppresses HIV-1 replication. *Proc. Natl. Acad. Sci. U.S.A.*, *94*, 3548–3553.
83. Daelemans, D., Schols, D., Witvrouw, M., Pannecouque, C., Hatse, S., Van Dooren, S., Hamy, F., Klimkait, T., De Clercq, E., Vandamme, A. M. (2000). A second target for the peptoid Tat/transactivation response element inhibitor CGP64222: Inhibition of human immunodeficiency virus replication by blocking CXC-chemokine receptor 4-mediated virus entry. *Mol. Pharmacol.*, *57*, 116–124.
84. Chao, S. H., Fujinaga, K., Marion, J. E., Taube, R., Sausville, E. A., Senderowicz, A. M., Peterlin, B. M., Price, D. H. (2000). Flavopiridol inhibits P-TEFb and blocks HIV-1 replication. *J. Biol. Chem.*, *275*, 28345–28348.

85. Chiu, Y. L., Ho, C. K., Saha, N., Schwer, B., Shuman, S., Rana, T. M. (2002). Tat stimulates cotranscriptional capping of HIV mRNA. *Mol. Cell*, *10*, 585–597.

86. Daelemans, D., Esté, J. A., Witvrouw, M., Pannecouque, C., Jonckheere, H., Aquaro, S., Perno, C. F., De Clercq, E., Vandamme, A.-M. (1997). *S*-Adenosylhomocysteine hydrolase inhibitors interfere with the replication of human immunodeficiency virus type 1 through inhibition of the LTR transactivation. *Mol. Pharmacol.*, *52*, 1157–1163.

87. Dymock, B. W., Jones, P. S., Wilson, F. X. (2000). Novel approaches to the treatment of hepatitis C virus infection. *Antiviral Chem. Chemother.*, *11*, 79–96.

88. Kim, J. L., Morgenstern, K. A., Griffith, J. P., Dwyer, M. D., Thomson, J. A., Murcko, M. A., Lin, C., Caron, P. R. (1998). Hepatitis C virus NS3 RNA helicase domain with a bound oligonucleotide: The crystal structure provides insights into the mode of unwinding. *Structure*, *6*, 89–100.

89. Tan, S. L., Pause, A., Shi, Y., Sonenberg, N. (2002). Hepatitis C therapeutics: Current status and emerging strategies. *Nature Rev. Drug Disc.*, *1*, 867–881.

90. Paolini, C., De Francesco, R., Gallinari, P. (2000). Enzymatic properties of hepatitis C virus NS3-associated helicase. *J. Gen. Virol.*, *81*, 1335–1345.

91. Carroll, S. S., Sardana, V., Yang, Z., Jacobs, A. R., Mizenko, C., Hall, D., Hill, L., Zugay-Murphy, J., Kuo, L. C. (2000). Only a small fraction of purified hepatitis C RNA-dependent RNA polymerase is catalytically competent: Implications for viral replication and in vitro assays. *Biochemistry*, *39*, 8243–8249.

92. Patick, A. K., Potts, K. E. (1998). Protease inhibitors as antiviral agents. *Clin. Microbiol. Rev.*, *11*, 614–627.

93. Robinson, B. S., Riccardi, K. A., Gong, Y. F., Guo, Q., Stock, D. A., Blair, W. S., Terry, B. J., Deminie, C. A., Djang, F., Colonno, R. J., Lin, P. F. (2000). BMS-232632, a highly potent human immunodeficiency virus protease inhibitor that can be used in combination with other available antiretroviral agents. *Antimicrob. Agents Chemother.*, *44*, 2093–2099.

94. Erickson, J. W., Burt, S. K. (1996). Structural mechanisms of HIV drug resistance. *Ann. Rev. Pharmacol. Toxicol.*, *36*, 545–571.

95. Hodge, C. N., Aldrich, P. E., Bacheler, L. T., Chang, C. H., Eyermann, C. J., Garber, S., Grubb, M., Jackson, D. A., Jadhav, P. K., Korant, B., Lam, P. Y., Maurin, M. B., Meek, J. L., Otto, M. J., Rayner, M. M., Reid, C., Sharpe, T. R., Shum, L.L., Winslow, D. L., Erickson-Viitanen, S. (1996). Improved cyclic urea inhibitors of the HIV-1 protease: Synthesis, potency, resistance profile, human pharmacokinetics and X-ray crystal structure of DMP 450. *Chem. Biol.*, *3*, 301–314.

96. Hyland, L. J., Tomaszek, T. A. Jr., Meek, T. D. (1991). Human immunodeficiency virus-1 protease. 2. Use of pH rate studies and solvent kinetic isotope effects to elucidate details of chemical mechanism. *Biochemistry*, *30*, 8454–8463.

97. Turner, S. R., Strohbach, J. W., Tommasi, R. A., Aristoff, P. A., Johnson, P. D., Skulnick, H. I., Dolak, L. A., Seest, E. P., Tomich, P. K., Bohanon, M. J., Horng, M. M., Lynn, J. C., Chong, K. T., Hinshaw, R. R., Watenpaugh, K. D., Janakiraman, M. N., Thaisrivongs, S. (1998). Tipranavir (PNU-140690): A potent, orally bioavailable nonpeptidic HIV protease inhibitor of the 5,6-dihydro-4-hydroxy-2-pyrone sulfonamide class. *J. Med. Chem.*, *41*, 3467–3476.

98. Hagen, S. E., Domagala, J., Gajda, C., Lovdahl, M., Tait, B. D., Wise, E., Holler, T., Hupe, D., Nouhan, C., Urumov, A., Zeikus, G., Zeikus, E., Lunney, E. A., Pavlovsky, A., Gracheck, S. J., Saunders, J., VanderRoest, S., Brodfuehrer, J. (2001). 4-Hydroxy-5,6-dihydropyrones as inhibitors of HIV protease: The effect of heterocyclic substituents at C-6 on antiviral potency and pharmacokinetic parameters. *J. Med. Chem.*, *44*, 2319–2332.

99. Larder, B. A., Hertogs, K., Bloor, S., van den Eynde, C. H., DeCian, W., Wang, Y., Freimuth, W. W., Tarpley, G. (2000). Tipranavir inhibits broadly protease inhibitor-resistant HIV-1 clinical samples. *AIDS*, *14*, 1943–1948.

100. Waxman, L., Darke, P. L. (2000). The herpesvirus proteases as targets for antiviral chemotherapy. *Antiviral Chem. Chemother.*, *11*, 1–22.

101. Jarvest, R. L., Pinto, I. L., Ashman, S. M., Dabrowski, C. E., Fernandez, A. V., Jennings, L. J., Lavery, P., Tew, D. G. (1999). Inhibition of herpes proteases and antiviral activity of 2-substituted thieno-[2,3-d]oxazinones. *Bioorg. Med. Chem. Lett.*, *9*, 443–448.

102. Smith, D. G., Gribble, A. D., Haigh, D., Ife, R. J., Lavery, P., Skett, P., Slingsby, B. P., Stacey, R., Ward, R. W., West, A. (1999). The inhibition of human cytomegalovirus (hCMV) protease by hydroxylamine derivatives. *Bioorg. Med. Chem. Lett.*, *9*, 3137–3142.

103. Ogilvie, W. W., Yoakim, C., Do, F., Hache, B., Lagace, L., Naud, J., O'Meara, J. A., Deziel, R. (1999). Synthesis and antiviral activity of monobactams inhibiting the human cytomegalovirus protease. *Bioorg. Med. Chem.*, *7*, 1521–1531.

104. Borthwick, A. D., Angier, S. J., Crame, A. J., Exall, A. M., Haley, T. M., Hart, G. J., Mason, A. M., Pennell, A. M., Weingarten, G. G. . (2000). Design and synthesis of pyrrolidine-5,5-*trans*-lactams (5-oxo-hexahydro-pyrrolo[3,2-*b*]pyrroles) as novel mechanism-based inhibitors of human cytomegalovirus protease. 1. The α-methyl-*trans*-lactam template. *J. Med. Chem.*, *43*, 4452–4464.

105. Matsumoto, M., Misawa, S., Chiba, N., Takaku, H., Hayashi, H. (2001). Selective nonpeptidic inhibitors of herpes simplex virus type 1 and human cytomegalovirus proteases. *Biol. Pharm. Bull.*, *24*, 236–241.

106. Dragovich, P. S., Webber, S. E., Babine, R. E., Fuhrman, S. A., Patick, A. K., Matthews, D. A., Lee, C. A., Reich, S. H., Prins, T. J., Marakovits, J. T., Littlefield, E. S., Zhou, R., Tikhe, J., Ford, C. E., Wallace, M. B., Meador J. W. 3rd, Ferre, R. A., Brown, E. L., Binford, S. L., Harr, J. E., DeLisle, D. M., Worland, S. T. (1998). Structure-based design, synthesis, and biological evaluation of irreversible human rhinovirus 3C protease inhibitors. 1. Michael acceptor structure-activity studies. *J. Med. Chem.*, *41*, 2806–2818.

107. Dragovich, P. S., Webber, S. E., Babine, R. E., Fuhrman, S. A., Patick, A. K., Matthews, D. A., Reich, S. H., Marakovits, J. T., Prins, T. J., Zhou, R., Tikhe, J., Littlefield, E. S., Bleckman, T. M., Wallace, M. B., Little, T. L., Ford, C. E., Meador, J. W. 3rd, Ferre, R. A., Brown, E. L., Binford, S. L., DeLisle, D. M., Worland, S. T. (1998). Structure-based design, synthesis, and biological evaluation of irreversible human rhinovirus 3C protease inhibitors. 2. Peptide structure-activity studies. *J. Med. Chem.*, *41*, 2819–2834.

108. Dragovich, P. S., Prins, T. J., Zhou, R., Fuhrman, S. A., Patick, A. K., Matthews, D. A., Ford, C. E., Meador, J. W. 3rd, Ferre, R. A., Worland, S. T. (1999). Structure-based design, synthesis, and biological evaluation of irreversible human rhinovirus

3C protease inhibitors. 3. Structure-activity studies of ketomethylene-containing peptidomimetics. *J. Med. Chem.*, *42*, 1203–1212.

109. Dragovich, P. S., Prins, T. J., Zhou, R., Webber, S. E., Marakovits, J. T., Fuhrman, S. A., Patick, A. K., Matthews, D. A., Lee, C. A., Ford, C. E., Burke, B. J., Rejto, P. A., Hendickson, T. F., Tuntland, T., Brown, E. L., Meador J. W. 3rd, Ferre, R. A., Harr, J. E., Kosa, M. B., Worland, S. T. (1999). Structure-based design, synthesis, and biological evaluation of irreversible human rhinovirus 3C protease inhibitors. 4. Incorporation of P1 lactam moieties as L-glutamine replacements. *J. Med. Chem.*, *42*, 1213–1224.
110. Dragovich, P. S., Prins, T. J., Zhou, R., Brown, E. L., Maldonado, F. C., Fuhrman, S. A., Zalman, L. S., Tuntland, T., Lee, C. A., Patick, A. K., Matthews, D. A., Hendrickson, T. F., Kosa, M. B., Liu, B., Batugo, M. R., Gleeson, J. P., Sakata, S. K., Chen, L., Guzman, M. C., Meador, J. W. 3rd, Ferre, R. A., Worland, S. T. (2002). Structure-based design, synthesis, and biological evaluation of irreversible rhinovirus 3C protease inhibitors. 6. Structure-activity studies of orally 2-pyridone-containing peptidomimetics. *J. Med. Chem.*, *45*, 1607–1623.
111. Matthews, D. A., Dragovich, P. S., Webber, S. E., Fuhrman, S. A., Patick, A. K., Zalman, L. S., Hendrickson, T. F., Love, R. A., Prins, T. J., Marakovits, J. T., Zhou, R., Tikhe, J., Ford, C. E., Meador, J. W., Ferre, R. A., Brown, E. L., Binford, S. L., Brothers, M. A., DeLisle, D. M., Worland, S. T. (1999). Structure-assisted design of mechanism-based irreversible inhibitors of human rhinovirus 3C protease with potent antiviral activity against multiple rhinovirus serotypes. *Proc. Natl. Acad. Sci. U.S.A.*, *96*, 11000–11007.
112. Patick, A. K., Binford, S. L., Brothers, M. A., Jackson, R. L., Ford, C. E., Diem, M. D., Maldonado, F., Dragovich, P. S., Zhou, R., Prins, T. J., Fuhrman, S. A., Meador, J. W., Zalman, L. S., Matthews, D. A., Worland, S. T. (1999). In vitro antiviral activity of AG7088, a potent inhibitor of human rhinovirus 3C protease. *Antimicrob. Agents Chemother.*, *43*, 2444–2450.
113. Hsyu, P.-H., Pithavala, Y. K., Gersten, M., Penning, C. A., Kerr, B. M. (2002). Pharmacokinetics and safety of an antirhinoviral agent, ruprintrivir, in healthy volunteers. *Antimicrob. Agents Chemother.*, *46*, 392–397.
114. Tautz, N., Kaiser, A., Thiel, H.-J. (2000). NS3 serine protease of bovine viral diarrhea virus: Characterization of active site residues, NS4A cofactor domain, and protease-cofactor interactions. *Virology*, *273*, 351–363.
115. Llinàs-Brunet, M., Bailey, M., Fazal, G., Ghiro, E., Gorys, V., Goulet, S., Halmos, T., Maurice, R., Poirier, M., Poupart, M. A., Rancourt, J., Thibeault, D., Wernick, D., Lamarre, D. (2000). Highly potent and selective peptide-based inhibitors of the hepatitis C virus serine protease: Towards smaller inhibitors. *Bioorg. Med. Chem. Lett.*, *10*, 2267–2270.
116. Bennett, J. M., Campbell, A. D., Campbell, A. J., Carr, M. G., Dunsdon, R. M., Greening, J. R., Hurst, D. N., Jennings, N. S., Jones, P. S., Jordan, S., Kay, P. B., O'Brien, M. A., Underwood, J., Raynham, T. M., Wilkinson, C. S., Wilkinson, T. C., Wilson, F. X. (2001). The identification of α-ketoamides as potent inhibitors of hepatitis C virus NS3-4A proteinase. *Bioorg. Med. Chem. Lett.*, *11*, 355–357.
117. Sing, W. T., Lee, C. L., Yeo, S. L., Lim, S. P., Sim, M. M. (2001). Arylalkylidene rhodanine with bulky and hydrophobic functional group as selective HCV NS3 protease inhibitor. *Bioorg. Med. Chem. Lett.*, *11*, 91–94.

118. Bartenschlager, R. (2002). Hepatitis C virus replicons, potential role for drug development. *Nature Rev. Drug Disc.*, *1*, 911–916.

119. von Itzstein, M., Wu, W. Y., Kok, G. B., Pegg, M. S., Dyason, J. C., Jin, B., Van Phan, T., Smythe, M. L., White, H. F., Oliver, S. W., Colman, P. M., Varghese, J. N., Ryan, D. M., Woods, J. M., Bethell, R. C., Hotham, V. J., Cameron, J. M., Penn, C. R. (1993). Rational design of potent sialidase-based inhibitors of influenza virus replication. *Nature*, *363*, 418–423.

120. Barnett, J. M., Cadman, A., Gor, D., Dempsey, M., Walters, M., Candlin, A., Tisdale, M., Morley, P. J., Owens, I. J., Fenton, R. J., Lewis, A. P., Claas, E. C., Rimmelzwaan, G. F., De Groot, R., Osterhaus, A. D. (2000). Zanamivir susceptibility monitoring and characterization of influenza virus clinical isolates obtained during phase II clinical efficacy studies. *Antimicrob. Agents Chemother.*, *44*, 78–87.

121. Hayden, F. G., Osterhaus, A. D., Treanor, J. J., Fleming, D. M., Aoki, F. Y., Nicholson, K. G., Bohnen, A. M., Hirst, H. M., Keene, O., Wightman, K. (1997). Efficacy and safety of the neuraminidase inhibitor zanamivir in the treatment of influenza virus infections. *N. Engl. J. Med.*, *337*, 874–880.

122. Kim, C. U., Lew, W., Williams, M. A., Liu, H., Zhang, L., Swaminathan, S., Bischofberger, N., Chen, M. S., Mendel, D. B., Tai, C. Y., Laver, W. G., Stevens, R. C. (1997). Influenza neuraminidase inhibitors possessing a novel hydrophobic interaction in the enzyme active site: design, synthesis, and structural analysis of carbocyclic sialic acid analogues with potent anti-influenza activity. *J. Am. Chem. Soc.*, *119*, 681–690.

123. Mendel, D. B., Tai, C. Y., Escarpe, P. A., Li, W., Sidwell, R. W., Huffman, J. H., Sweet, C., Jakeman, K. J., Merson, J., Lacy, S. A., Lew, W., Williams, M. A., Zhang, L., Chen, M. S., Bischofberger, N., Kim, C. U. (1998). Oral administration of a prodrug of the influenza virus neuraminidase inhibitor GS 4071 protects mice and ferrets against influenza infection. *Antimicrob. Agents Chemother.*, *42*, 640–646.

124. Hayden, F. G., Treanor, J. J., Fritz, R. S., Lobo, M., Betts, R. F., Miller, M., Kinnersley, N., Mills, R. G., Ward, P., Straus, S. E. (1999). Use of the oral neuraminidase inhibitor oseltamivir in experimental human influenza: randomized controlled trials for prevention and treatment. *JAMA*, *282*, 1240–1246.

125. Nicholson, K. G., Aoki, F. Y., Osterhaus, A. D., Trottier, S., Carewicz, O., Mercier, C. H., Rode, A., Kinnersley, N., Ward, P. (2000). Efficacy and safety of oseltamivir in treatment of acute influenza: A randomised controlled trial. *Lancet*, *355*, 1845–1850.

126. Smee, D. F., Huffman, J. H., Morrison, A. C., Barnard, D. L., Sidwell, R. W. (2001). Cyclopentane neuraminidase inhibitors with potent in vitro anti-influenza virus activities. *Antimicrob. Agents Chemother.*, *45*, 743–748.

127. Sidwell, R. W., Smee, D. F., Huffman, J. H., Barnard, D. L., Bailey, K. W., Morrey, J. D., Babu, Y. S. (2001). In vivo influenza virus-inhibitory effects of the cyclopentane neuraminidase inhibitor RWJ-270201. *Antimicrob. Agents Chemother.*, *45*, 749–757.

128. Bantia, S., Parker, C. D., Ananth, S. L., Horn, L. L., Andries, K., Chand, P., Kotian, P. L., Dehghani, A., El-Kattan, Y., Lin, T., Hutchinson, T. L., Montgomery, J. A., Kellog, D. L., Babu, Y. S. (2001). Comparison of the anti-influenza virus activity of RWJ-270201 with those of oseltamivir and zanamivir. *Antimicrob. Agents Chemother.*, *45*, 1162–1167.

129. Wang, G. T., Chen, Y., Gentles, R., Sowin, T., Kati, W., Muchmore, S., Giranda, V., Stewart, K., Sham, H., Kempf, D., Laver, W. G. (2001). Design, synthesis, and structural analysis of influenza neuraminidase inhibitors containing pyrrolidine cores. *J. Med. Chem.*, *44*, 1192–1201.

130. De Goey, D. A., Chen, H. J., Flosi, W. J., Grampovnik, D. J., Yeung, C. M., Klein, L. L., Kempf, D. J. (2002). Enantioselective synthesis of antiinfluenza compound A-315675. *J. Org. Chem.*, *67*, 5445–5453.

131. Hanessian, S., Bayrakdarian, M., Luo, X. (2002). Total synthesis of A-315675: A potent inhibitor of influenza neuraminidase. *J. Am. Chem. Soc.*, *124*, 4716–4721.

132. Sintchak, M. D., Fleming, M. A., Futer, O., Raybuck, S. A., Chambers, S. P., Caron, P. R., Murcko, M. A., Wilson, K. P. (1996). Structure and mechanism of inosine monophosphate dehydrogenase in complex with the immunosuppressant mycophenolic acid. *Cell*, *85*, 921–930.

133. De Clercq, E., Cools, M., Balzarini, J., Snoeck, R., Andrei, G., Hosoya, M., Shigeta, S., Ueda, T., Minakawa, N., Matsuda, A. (1991). Antiviral activities of 5-ethynyl-1-β-D-ribofuranosylimidazole-4-carboxamide and related compounds. *Antimicrob. Agents Chemother.*, *35*, 679–684.

134. Markland, W., McQuaid, T. J., Jain, J., Kwong, A. D. (2000). Broad-spectrum antiviral activity of the IMP dehydrogenase inhibitor VX-497: A comparison with ribavirin and demonstration of antiviral additivity with alpha interferon. *Antimicrob. Agents Chemother.*, *44*, 859–866.

135. Neyts, J., Andrei, G., De Clercq, E. (1998). The novel immunosuppressive agent mycophenolate mofetil markedly potentiates the antiherpesvirus activities of acyclovir, ganciclovir, and penciclovir in vitro and in vivo. *Antimicrob. Agents Chemother.*, *42*, 216–222.

136. Margolis, D., Heredia, A., Gaywee, J., Oldach, D., Drusano, G., Redfield, R. (1999). Abacavir and mycophenolic acid, an inhibitor of inosine monophosphate dehydrogenase, have profound and synergistic anti-HIV activity. *J. Acquir. Immune Defic. Syndrom.*, *21*, 362–370.

137. Saracco, G., Ciancio, A., Olivero, A., Smedile, A., Roffi, L., Croce, G., Colletta, C., Cariti, G., Andreoni, M., Biglino, A., Calleri, G., Maggi, G., Tappero, G. F, Orsi, P. G., Terreni, N., Macor, A., Di Napoli, A., Rinaldi, E., Ciccone, G., Rizzetto, M. (2001). A randomized 4-arm multicenter study of interferon alfa-2b plus ribavirin in the treatment of patients with chronic hepatitis C not responding to interferon alone. *Hepatology*, *34*, 133–138.

138. Crotty, S., Maag, D., Arnold, J. J., Zhong, W., Lau, J. Y., Hong, Z., Andino, R., Cameron, C. E. (2000). The broad-spectrum antiviral ribonucleoside ribavirin is an RNA virus mutagen. *Nature Med.*, *6*, 1375–1379.

139. Crotty, S., Cameron, C. E., Andino, R. (2001). RNA virus error catastrophe: Direct molecular test by using ribavirin. *Proc. Natl. Acad. Sci. U.S.A.*, *98*, 6895–6900.

140. De Clercq, E., Cools, M., Balzarini, J., Marquez, V. E., Borcherding, D. R., Borchardt, R. T., Drach, J. C., Kitaoka, S., Konno, T. (1989). Broad-spectrum antiviral activities of neplanocin A, 3-deazaneplanocin A, and their 5′-nor derivatives. *Antimicrob. Agents Chemother.*, *33*, 1291–1297.

141. Yuan, C. S., Wnuk, S. F., Liu, S., Robins, M. J., Borchardt, R. T. (1994). (E)-5′,6′-didehydro-6′-deoxy-6′-fluorohomoadenosine: A substrate that measures the

hydrolytic activity of *S*-adenosylhomocysteine hydrolase. *Biochemistry*, *33*, 12305–12311.

142. Bray, M., Davis, K., Geisbert, T., Schmaljohn, C., Huggins, J. (1999). A mouse model for evaluation of prophylaxis and therapy of Ebola hemorrhagic fever. *J. Infect. Dis.*, *179*(Suppl. 1), S248–S258.
143. Huggins, J., Zhang, Z.-X., Bray, M. (1999). Antiviral drug therapy of filovirus infections: *S*-adenosylhomocysteine hydrolase inhibitors inhibit Ebola virus in vitro and in a lethal mouse model. *J. Infect. Dis.*, *179*(Suppl. 1), S240–S247.
144. Bray, M., Driscoll, J., Huggins, J. W. (2000). Treatment of lethal Ebola virus infection in mice with a single dose of an *S*-adenosyl-L-homocysteine hydrolase inhibitor. *Antiviral Res.*, *45*, 135–147.

# PROTEIN KINASE INHIBITORS IN DRUG DISCOVERY

Keykavous Parang and Gongqin Sun
*University of Rhode Island*
*Kingston Rhode Island*

1 PHOSPHORYLATION IN SIGNAL TRANSDUCTION PATHWAYS
2 CLASSIFICATION OF PROTEIN KINASES
3 GENERAL MECHANISTIC FEATURES OF PROTEIN PHOSPHORYLATION BY PROTEIN KINASES
4 GENERAL STRUCTURAL FEATURES OF PROTEIN KINASES
5 ACTIVATION OF PROTEIN KINASES
6 PROTEIN KINASES IN HUMAN DISEASES
7 IMPORTANT PROTEIN KINASES AS THERAPEUTIC TARGETS
Protein Tyrosine Kinases
Serine/Threonine Kinases
8 PROTEIN KINASES AS TARGETS FOR INHIBITOR DESIGN
General Overview of Protein Kinase Inhibitors (PKIs)
Alternative Strategies in Designing Protein Kinase Inhibitors
9 SELECTED INHIBITORS OF PROTEIN KINASES ON MARKET OR IN CLINICAL TRIALS
Receptor Tyrosine Kinase Inhibitors
Nonreceptor Tyrosine Kinase Inhibitors
Serine/Threonine Kinase Inhibitors
10 PROSPECTS AND FUTURE DIRECTIONS
References

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

## ABBREVIATIONS AND GLOSSARY

| | |
|---|---|
| ACS | acute coronary syndromes |
| AML | acute myelogenous leukemia |
| ALL | acute lymphoblastic leukemia |
| Bcr-Abl | Abelson tyrosine kinase fused to the breakpoint cluster region |
| bFGF | basic fibroblast growth factor |
| CDK | cyclin-dependent protein kinase |
| CDKI | CDK inhibitor |
| Csk | C-terminal Src kinase |
| DYRK | dual specificity, tyrosine-phosphorylated and regulated kinase |
| EGF | epidermal growth factor |
| EGFR | EGF receptor |
| Erb | estrogen receptor subtype b |
| ERK | extracellular signal-regulated protein kinase |
| FDA | U.S. Food and Drug Administration |
| FGF | fibroblast growth factor |
| FGFR | FGF receptor |
| Flk-1/KDR | fetal liver kinase-1/kinase insert domain containing receptor (VEGFR-2, VEGFR type 2) |
| Flt-1 | fms-like tyrosine kinase-1 |
| GSK3 | glycogen synthase kinase 3 |
| HER | human epidermal growth factor receptor |
| $IC_{50}$ | 50% inhibitory concentration |
| IR | insulin receptor |
| IRK | insulin receptor kinase |
| IRS | insulin receptor substrate |
| JAK | Janus kinases |
| JNK | c-jun N-terminal kinase |
| KDR | kinase insert domain containing receptor |
| Lck | lymphocyte kinase |
| MAP | mitogen-activated protein |
| MAPK | mitogen-activated protein kinase |
| MAPKK | MAP kinase kinase |
| MEK-1 | MAPK/ERK kinase 1 |
| MKK | MAP kinase kinase |
| mTOR | mammalian target of rapamycin |
| NGF | nerve growth factor |
| NRTK | nonreceptor tyrosine kinases |
| PDGF | platelet-derived growth factor |
| PDGFR | PDGF receptor |
| PKA | protein kinase A; cAMP-dependent protein kinase |
| PKC | protein kinase C |

| | |
|---|---|
| PKI | protein kinase inhibitor |
| PTK | protein tyrosine kinase |
| PTP | protein tyrosine phosphatase |
| RA | rheumatoid arthritis |
| *Ras* | rat sarcoma virus oncogene |
| *Rb* | retinoblastoma tumor suppressor gene |
| RTK | receptor tyrosine kinase |
| S | DNA synthesis stage |
| SAPK | stress-activated protein kinases |
| SH2 | Src homology 2 |
| STKs | serine/threonine kinases |
| Tek/Tie-2 | tunica internal endothelial cell kinase type 2 |
| Tie-1 | tunica internal endothelial cell kinase type 1 |
| TNF | tumor necrosis factor |
| VEGF | vascular endothelial growth factor |

## 1 PHOSPHORYLATION IN SIGNAL TRANSDUCTION PATHWAYS

Approximately 2 percent of eukaryotic genes encode protein kinases (PKs) [1, 2]. Approximately 518 PKs are predicted in the human kinome based on the information from the human genome sequence [3]. PKs have essential roles in cell-signaling pathways by interacting with extracellular ligands such as growth factors and hormones and transmitting signals across the cell membrane to the cytoplasm and the nucleus. Signal transduction is mediated by PKs through catalyzing the transfer of the γ-phosphoryl group from adenosine triphosphate (ATP) to the hydroxyl group of defined tyrosine (Tyr), serine (Ser), and/or threonine (Thr) residues in various critical protein substrates (Fig. 1). These amino acids are usually located in the consensus sequences on the protein substrate. Phosphorylated protein substrates have distinct functional properties in many different cell-signaling pathways. Abnormal phosphorylation is often a cause or consequence of diseases. In addition to the signals arising from receptor–ligand interactions as seen with growth factors,

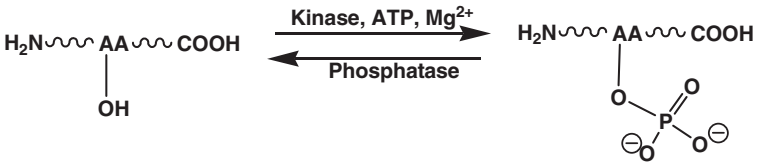


**Figure 1** General scheme for phosphorylation and dephosphorylation reactions catalyzed by PKs and PPs, respectively. PKs phosphorylate the hydroxyl group of amino acids (AA-OH = tyrosine serine, or threonine) on the protein substrate in the presence of $Mg^{2+}$ and ATP. PPs reverse the phosphorylation reaction by cleaving the O–P bond in phosphoproteins, to generate the free hydroxyl group in proteins.

environmental factors such as mechanical deformation by cell stretch or shear stress will activate PKs. On the other hand, protein phosphatases (PPs) remove the phosphate groups from distinct phosphoproteins in a reverse reaction to counterbalance the phosphorylation reaction.

## 2 CLASSIFICATION OF PROTEIN KINASES

Although 518 members of the human protein kinase superfamily can be classified into at least 7 distinct groups and many families and subfamilies [3], PKs are usually divided into two large families based on their substrate specificities: protein tyrosine kinases (PTKs) (e.g., Src tyrosine kinases, Bcr-Abl) and serine/threonine kinases (STKs) [e.g., mitogen-activated protein kinases (MAP kinases), protein kinase C (PKC), protein kinase A (PKA)] (Fig. 2). PTKs such as Src kinases and several of serine/threonine kinases such as MAP kinases play a crucial role in signal transduction pathways and regulating various cellular functions (e.g., proliferation, differentiation). Several serine/threonine kinases such as cyclin-dependent kinases (CDKs) are involved in cell cycle checkpoint control.

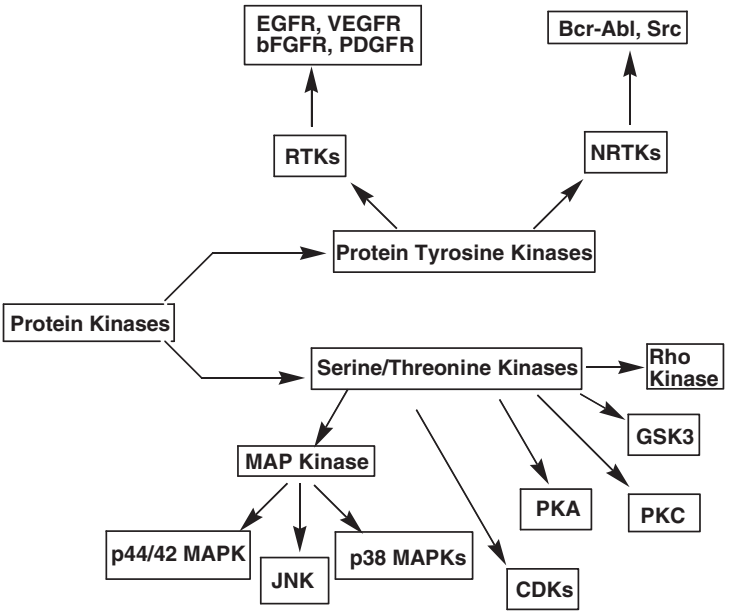


**Figure 2** Classification of PKs involved in signaling pathways into PTKs and STKs. PTKs are subdivided into the receptor tyrosine kinases (RTKs) and nonreceptor tyrosine kinases (NRTKs). Important PKs known as drug targets for each class are shown here.

It is estimated that there are approximately 100 PTKs. PTKs are subdivided into two families: receptor tyrosine kinases (RTKs; e.g., growth factor receptor kinases) and nonreceptor tyrosine kinases (NRTKs; e.g., Bcr-Abl, Src) [4]. RTKs are transmembrane cell surface proteins that are activated by the binding of ligands to their extracellular domains. The binding leads to receptor oligomerization and phosphorylation/activation of the intracellular domains by inducing autophosphorylation. Activated and phosphorylated intracellular domains then activate downstream signaling proteins such as NRTKs in the cytoplasm [4]. RTKs include the receptors for ligands, such as insulin (insulin receptor kinase; IRK), and many growth factors, such as epidermal growth factor (EGF), basic fibrobalst growth factor (FGF), platelet-derived growth factor (PDGF), vascular endothelial growth factor (VEGF), and nerve growth factor (NGF). NRTKs include members of the Src family, Fak, Jak, Abl, and Zap70. NRTKs can be regulated and activated by different mechanisms such as RTKs and other cell surface receptors such as receptors of the immune system and G-protein-coupled receptors [4].

## 3 GENERAL MECHANISTIC FEATURES OF PROTEIN PHOSPHORYLATION BY PROTEIN KINASES

To design protein kinase inhibitors (PKIs), it is important to understand the mechanistic features of the particular target kinase. Although 518 PKs share a conserved catalytic domain in sequence and structure, the catalysis and phosphorylation reaction is regulated differently by these enzymes. There is still an incomplete understanding of the mechanisms used by PKs in the phosphorylation reaction. The phosphorylation reaction occurs in the active site of PKs by the formation of a ternary complex consisting of a protein substrate, MgATP, and a PK [5]. ATP and protein substrate bind to ATP and substrate binding sites, respectively, in the catalytic domain of kinases. The transfer of the phosphoryl group takes place directly from ATP to the protein substrate through dissociative or associative mechanisms in a three-step process: (1) the initial binding of substrate to a preformed PK–ATP complex, (2) the transfer of phosphoryl group to hydroxyl group, and (3) the release of phosphorylated protein from the PK–ADP (adenosine diphosphate) complex (Fig. 3). In a fully dissociative mechanism, the bond between γ-phosphorus and ADP (leaving group) is broken in the transition state before the formation of the bond between the nucleophile (tyrosine, threonine, or serine) and phosphorus. In an associative mechanism transition state, a significant amount of bond is formed between the nucleophile and phosphorus, before the complete bond breakage between the phosphorus and the leaving group [6–8]. The distinction between the associative and dissociative mechanisms may be critical in designing transition-state and mechanism-based inhibitors of PKs. Inhibitors can be designed that could potentially inhibit the binding of ATP (ATP

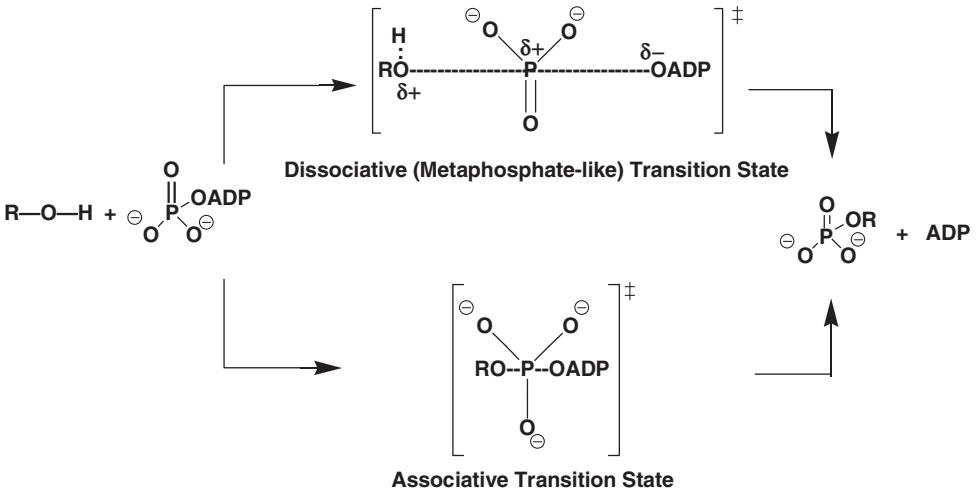


**Figure 3** Dissociative versus associative transition states for phosphoryl transfer.

binding site inhibitors), protein substrate (substrate binding site inhibitors), or both (bisubstrate analog inhibitors).

## 4 GENERAL STRUCTURAL FEATURES OF PROTEIN KINASES

The three-dimensional structures of several individual PKs have been determined by X-ray crystallography. Protein Data Bank (PDB) comprises over 160 PKs, protein kinase mutants, and protein kinase : ligand complexes. This wealth of structural information of PKs and PK–ligand complexes can be used for structure-based rational design of novel and effective PK inhibitors. For example, it has been shown from these structures that the backbone carbonyl and amide groups of the protein backbone located in the hinge region of ATP binding site potentially form three hydrogen bonds with ATP binding site inhibitors through two hydrogen bond acceptors and one hydrogen bond donor [9]. The detailed information about protein kinase structures and the interaction mode with their inhibitors have been reviewed extensively elsewhere [4, 10–13] and is beyond the scope of this chapter. Some of the crystal structures of the kinase domain of several RTKs and NRTKS that are clinically important include IRK [14], insulin-like growth factor-1 receptor [15], c-Src tyrosine kinase [16, 17], FGFR (FGF receptor) [18, 19], VEGFR [20], lymphocyte kinase (Lck) [21, 22], hematopoetic cell kinase (Hck) [23, 24], Bruton's tyrosine kinase (BtK) [25], and Abl kinase [26].

All PKs have an ~260-amino-acid conserved kinase domain folding into two structurally dissimilar lobes, N-terminal and C-terminal domains that are joined by a linker peptide coil [27]. The kinase domain is responsible for

catalytic activity and the phosphorylation reaction. The N-terminal domain of PKs or nucleotide binding site is generally ~90 amino acids, which recognizes and binds ATP utilized by all kinases. Many of the amino acid residues comprising the nucleotide binding pocket are highly conserved forming five β strands (β1 to β5) and at least one conserved helix (αC). The C-terminal lobe is predominantly α-helical and is not conserved among kinases varying in size, sequence, and topology containing the substrate binding groove, activation loop, and catalytic residues [28] (Fig. 4). The transfer of the phosphoryl group takes place directly between ATP and protein substrate in the kinase domain. In addition to kinase domain, RTKs comprise of multidomain membrane-spanning proteins.

Discussing more detailed structural features of all PKs is beyond the scope of this chapter. Cytoplasmic nonreceptor kinases contain several other domains in addition to the kinase domain. One of the well-studied examples is Src kinases that are potential drug targets and can serve as an example for the interesting structure and function that PTKs possess.

Src family kinases share common structural motifs and are made up of five components/domains: N-terminal or the fatty acid acylation domain, which targets the kinases to the plasma membrane, Src homology 3 (SH3), Src homology 2 (SH2), kinase domain (catalytic, including ATP and substrate binding sites), and CT noncatalytic regulatory domain (see Fig. 4) [17]. SH3 and SH2 are protein–protein interaction domains that have both adaptor and regulatory functions [17] by binding to their respective ligands [29]. SH3 domains are ~60 amino acids in length and bind proline-rich regions of several signaling proteins [16, 23, 30]. SH2 domains are modules of approximately 100

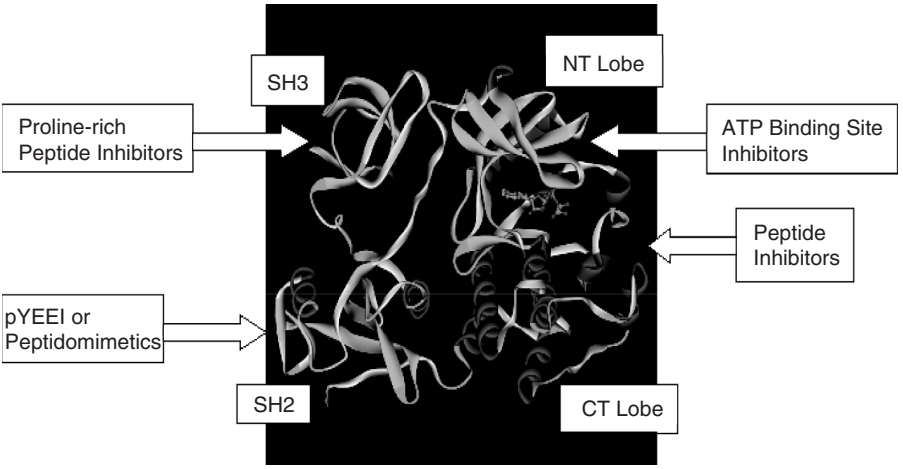


**Figure 4** General structure of Src tyrosine kinases and potential sites for inhibitors intervention. AMP-PNP, a ligand in kinase domain, is represented by ball-and-stick vectors (PDB 2SRC) [17].

amino acids that have evolved to recognize and bind to phosphotyrosine sequences located on proteins [17, 31] in response to extracellular signals [31–33]. The phosphotyrosyl group binds in a deep phosphotyrosine binding pocket of SH2 domain where it is stabilized through a large number of hydrogen bonds and electrostatic interactions with positively charged amino acids. In addition to stabilizing the inactive conformation of the kinases, the SH2 domain mediates interactions of the Src family kinases with other cellular proteins such as FAK, $p130^{cas}$, p85, PI3K, and $p68^{sam}$ [34–37] propagating downstream signaling by recruiting the SH2 domain-containing protein to its proper signaling complex and/or by altering its enzymatic activity. Inappropriate signaling through Src kinases has been linked to many pathological conditions, providing an impetus to understand the mechanism of SH2 domain recognition and to design SH2 domain binding inhibitors [29].

## 5 ACTIVATION OF PROTEIN KINASES

The activity of PKs is tightly regulated because of their central role in signal transduction [12, 14, 38]. Different kinases utilize various mechanisms to become activated such as transphosphorylation of specific residues within the activation loop in IRK, and the human Lck kinase [22, 39–43] leads to a dramatic conformational change in the structure of kinase. The conformational change exposes the substrate binding site for substrate binding by repositioning the key catalytic residues that are involved in binding to protein substrate and catalytic reaction. Another mechanism for regulation of kinase activity is phosphorylation sites N- and C-terminal to the kinase domain as seen in EphB2 receptor kinase domain [44]. Both of these mechanisms have been seen in Src family kinases [14, 16, 23, 45]. There are two major Src phosphorylation sites that have opposite effects on Src activity, Tyr416 and Tyr527. Tyr416, an autophosphorylation site, is located in a region called the activation loop of the kinase catalytic domain, which is ordered as an α-helix in inactive Src [17]. Phosphorylation of Tyr416 enhances Src activity. Tyr527 is located in the C-terminal tail of the molecule and, when phosphorylated by another PK called C-terminal Src kinase (Csk), binds to the SH2 domain of Src [17, 23]. Phosphorylation of Tyr527 by Csk leads to the kinase inactivation. This interaction locks the c-Src molecule in an inhibited closed conformation. Displacement of SH2 and/or SH3 domains, either by C-terminal tail dephosphorylation or by competitive binding of optimal SH2/SH3 domain ligands, allows the kinase domain to switch from a closed to an open conformation [17]. The kinase reactivation by different pTyr-containing peptides related to phosphotyrosine–glutamic acid–glutamic acid–isoleucine (pYEEI) known to bind the SH2 domain of Src has been demonstrated [46–49]. Src is known to be activated by SH2-mediated interaction with autophosphorylated PDGF receptor [50], and by dual SH3/SH2 interactions with sequences in the focal adhesion kinase FAK [51] and Sin [52]. In the inactive conformation of Src kinases, the

SH2 and kinase domains lie in opposite sides. A dramatic structural reorientation between the SH2 domain and the active site regions will occur upon the activation of the enzyme. In the active conformation, the SH2 domain is positioned closer to the active site region in a manner that promotes substrate binding in kinase domain [23, 53].

## 6 PROTEIN KINASES IN HUMAN DISEASES

Protein kinases are involved in most aspects of normal cellular signaling pathways and function and numerous biological events, such as apoptosis, differentiation, cell growth, and metabolism; therefore, aberrant kinase activity is believed to contribute to several diseases such as cancer [2, 54], diabetes [55, 56], cardiovascular [57], neurodegenerative [58] and inflammatory disorders [59, 60], and restenosis [61].

Protein kinase family has now become one of the most important targets for anticancer drug discovery since overexpression and enhanced PK activity have been directly associated with cell transformation, carcinogenesis, and many forms of human cancer [2, 54]. *v*-Src, a protein tyrosine kinase, was discovered more than 26 years ago as the first oncogene [62]. Since then it has been reported that over 80 percent of the oncogenes and proto-oncogenes involved in human cancers encode PTKs [2].

Activation or inhibition of a kinase in a specific disease state is not sufficient to consider it as putative therapeutic targets since dysregulation can be a consequence of the disease and may not be a contributing factor to disease pathology. Additional evidences such as genetic or physiological/cell biological data are needed to implicate a PK as a cause and not the effect of the disease and as an attractive target.

## 7 IMPORTANT PROTEIN KINASES AS THERAPEUTIC TARGETS

Herein, some of the PKs validated as targets for drug discovery are introduced. We will discuss a number of inhibitors developed against some of these kinases in Section 9 of "Combination Products: Drugs and Devices".

### Protein Tyrosine Kinases

Protein tyrosine kinases are enzymes that catalyze phosphorylation of tyrosine in many proteins by the transfer of the γ-phosphoryl group from ATP. While tyrosine phosphorylation represents a minor subset of protein phosphorylation in the cell in terms of level, it is involved in several cellular responses such as mitogenesis [63], differentiation [64], migration [65], and survival [66]. Two subfamilies of PTKs (RTKs and NRTKs) are discussed here. Typically, RTKs are linked to growth factor receptors and NRTKs are associ-

ated with the protein products of proto-oncogenes, such as *v*-src [67]. PTKs serve as molecular targets for many of the investigational new agents.

***Receptor Tyrosine Kinases*** The RTKs play instrumental roles in cancer development and have been well studied and extensively explored in anticancer drug discovery in the last decade. RTKs include EGF and its receptor (EGFR), tunica internal endothelial cell kinase type 1 (Tie-1), tunica internal endothelial cell kinase type 2 (Tek/Tie-2), VEGF and its receptor (VEGFR-2 or Flk-1/KDR), bFGF and its receptor (FGFR), and PDGF and its receptor (PDGFR). EGFR, bFGFR, VEGFR, and PDGFR kinases are discussed here.

*EGF Receptor Tyrosine Kinases* The EGF stimulates the proliferation of a large variety of cells such as epithelial cells and fibroblasts [68] by binding to and activating the EGFR and initiating intracellular signaling [68–70]. The EGF receptor kinases are also known as the EGF family of type I tyrosine kinases or ErbB tyrosine kinases. All four members in this family of RTKs, HER-1 (c-ErbB1/EGFR), HER-2 (c-ErbB2/neu), HER-3 (c-ErbB3), and HER-4 (c-ErbB4) [71, 72], are closely homologous and share common structural features such as an extracellular ligand binding domain and a cytoplasmic tyrosine kinase domain.

The EGFRs are implicated in the development and progression of certain human tumors [73], including head and neck, colon, breast, lung, prostate, and ovarian cancers. All EGFR kinases are frequently overactive in cancer cells. HER-1 is significantly overexpressed in a large variety of epithelial cancers, such as lung and breast cancers [74]. HER-2 (Neu, ErbB2) was identified in 1985 and soon became an attractive target for anticancer drug design since it is overexpressed, most commonly by gene amplification, in a number of human malignancies, including gastric, breast, colon, ovarian, and non-small-cell lung carcinoma [74]. Several pharmaceutical companies are interested in the development of agents targeting EGFR kinases.

*VEGF and bFGF Receptor Tyrosine Kinases* Two other important growth factor receptor targets are VEGFR-2 or Flk-1/KDR [75, 76] and bFGF-R, which have been associated to tumor angiogenesis. VEGFR-2 and FGFR are expressed exclusively in endothelial cells [75]; thus, they are very attractive targets for design of angiogenesis inhibitors. Some studies indicated that immunoneutralization of FGF had little or no effect on tumor angiogenesis [77, 78]. In spite of this FGFR has been explored in several diseases such as cancer, diabetic retinopathy, atherosclerosis, and rheumatoid arthritis [19]. The role of VEGF as a key regulator of angiogenesis has been confirmed [79]. There are two VEGF receptor tyrosine kinases: VEGFR1 (Flt-1) and VEGFR-2. There is now a general consensus that VEGFR2 is the major key mediator of the mitogenesis and angiogenesis in endothelial cells and is activated by binding of VEGF, a mitogen-specific protein for vascular endothelial cells [75]. It is proposed that the antiangiogenic therapy through inhibition of the

VEGFR is a safe and well-tolerated approach in treatment of cancer patients and may selectively target the tumor-associated vessels and reduce tumor-induced edema.

*PDGF Receptor Tyrosine Kinases* The PDGFRs have been reported to be overexpressed in many cancers of mesenchymal of glial origin and, thus, can also be a valuable target for these types of malignancies [80, 81]. The role of Flt-3, a member of the PDGFR tyrosine kinase family, in a variety of cancers, especially acute myelogenous leukemia (AML), has been clearly demonstrated [80].

***Nonreceptor Tyrosine Kinase Inhibitors*** The NRTKs include about one third of all PTKs and are found in cytoplasm signaling downstream of RTKs. Several important targets for anticancer drug design have been identified in the last 10 years by studying the NRTK subfamily [82] such as Bcr-Abl and Src family of kinases.

*Bcr-Abl* The Abelson PTK (c-Abl) is a nuclear protein tyrosine kinase with the unknown biological function. c-Abl is proposed to function in sensing the integrity of the genome and promoting programmed cell death. Bcr is a multifunctional cytosolic serine–threonine kinase that may play a role in regulating the activity of the Rho subfamily of small G proteins. The Philadelphia chromosome is formed by a reciprocal translocation between chromosome 9 and 22, resulting in the formation of a novel hybrid gene, which produces Bcr-Abl (the breakpoint cluster region fused to the Abelson leukemia virus), an oncoprotein responsible for Philadeliphia chromosome-positive chronic myelogenous leukemia (CML). This chromosome is found in more than 95 percent of patients with CML and about 15 to 20 percent of patients with adult acute lymphoblastic leukemia (ALL) [83] making Bcr-Abl an excellent target for the drug discovery activity against these types of cancer. Unlike c-Abl and Bcr, Bcr-Abl is both cytosolic and nuclear.

*Src Kinases* The Src family of PTKs, Src, Yes, Lck, Fyn, Lyn, Fgr, Hck, Blk, Frk, and Yrk are NRTKs [84, 85] found widely expressed in mammalian cells. Src tyrosine kinases, *v*-Src and *c*-Src, are prototypes of this well-established superfamily of PTKs. The Src family of PTKs has become key targets for both basic research and drug discovery over recent years. Src has significant structural homology to all nine other proteins in this superfamily. Enhanced Src tyrosine kinases activity has been directly linked to T-cell activation, mitogenesis, differentiation, cell transformation, and oncogenesis [54].

The Src exists as an intriguing therapeutic target for drug discovery with respect to cancer. Src has been associated with several different cancers including colon [86] and breast [87] cancers for which transformed phenotypes have been correlated with Src mutations and/or overexpression [54, 84, 88] of Src tyrosine kinase activities [89]. Src has also been implicated in the development

of osteoporosis and inflammation-mediated bone loss [90–93]. Src plays multifunctional roles in osteoblast and osteoclast activities. In osteoblasts, Src recently has been shown to be a negative regulator of osteoblast functional activity and implicated as a mediator of sex steroid-induced antiapoptotic signaling in osteoblasts [94, 95]. In osteoclasts, Src is important for osteoclast motility, activation (ruffled border formation), survival, and adhesion by mediating various signal transduction pathways [96–99]. Src is involved in bone remodeling (resorption and formation) and bone-related diseases, such as osteoporosis, rheumatoid arthritis, periodontal disease, Paget's disease, hypercalcemia of malignancy, and metastasis of certain cancers to bone [85].

Lyn activity is elevated in patients with acute myelogenous leukemia. Lck is a positive activator in T-cell signaling, and its activation is involved in autoimmune diseases such as rheumatoid arthritis [100]. Thus, Src kinases are important targets for drug discovery in several diseases such as bone-related diseases, cancer, and autoimmune diseases [54, 88, 100–104].

## Serine/Threonine Kinases

***MAP Kinases (ERK, JNK, p38 MAPKs)*** Exposure of cells to external stresses and extracellular stimuli such as ischemia/reperfusion [105, 106] will activate several members of the mitogen-activated protein kinase (MAPK) family, followed by other downstream signals [107, 108]. MAP kinases are STKs that are activated by dual phosphorylation of threonine and tyrosine residues of the Thr-X-Tyr segment in a loop located adjacent to the active site [108–110]. Prototypic members of the family present in mammalian cells include (1) extracellular signal-regulated kinases (ERK1/2; also known as p44/42 MAPK), (2) c-jun N-terminal kinases [(JNKs; also called stress-activated protein kinases (SAPKs)], (3) ERK5/big MAP kinase 1 (BMK1), and (4) p38 MAPKs. Each of these subgroups plays distinct roles in cell growth and differentiation. Each MAP kinase is phosphorylated by specific upstream kinases. Phosphorylated and activated MAP kinases move to the nucleus and regulate the phosphorylation of various substrates such as transcription factors (Fig. 5), which are involved in regulating the expression of specific sets of genes and thus mediate a specific response to the stimulus [111–116].

One important pathway causing the activation of MAPKs is the Ras-dependent MAPK cascade that is activated by growth factors (Fig. 5). The uncontrolled activation of this pathway by the overexpression of particular growth factor receptor tyrosine kinases or their mutation to constitutively active forms is now known to cause cancer. Ras activation begins its downstream signaling through association with the Raf (a serine/threonine kinase). The GTP (guanidine triphosphate)-bound dependent interaction of Ras with Raf is required for Raf kinase activation. Activated Raf phosphorylates and activates downstream kinases such as mitogen-activated ERK-activating kinase (MEK) and MAPK, respectively [117].

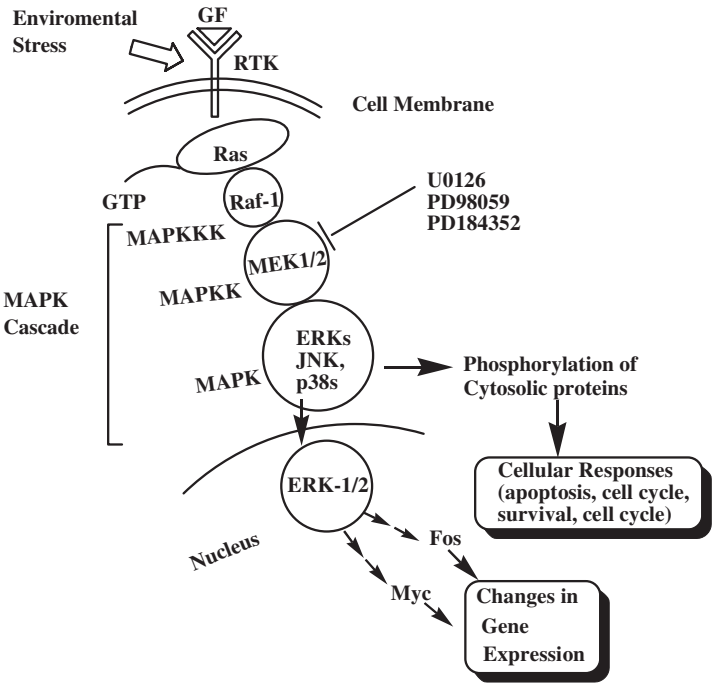


**Figure 5** Classical Ras-MAPK pathway and potential sites as drug targets. The binding of growth factors to extracellular domains of RTKs activates tyrosine kinase activities in their intracellular domains, which, in turn, activate the Ras cascade. Activation of Ras leads to sequential phosphorylation and activation of a series of enzymes involved in MAPK cascade that finally transmit the stimulatory external signals received from cytoplasmic membrane into the nucleus. Growth-factor-induced activation of pathway often leads to cell cycle progression. All MAPKs are part of a three-tiered cascade whereby a MAPKKK (in this case Raf-1) activate the downstream kinases MAPKKs (in this case MEK1/2), which, in turn, activate MAPKs (in this case, ERKs). Several MEK inhibitors discussed in text are shown here.

***Cyclin-Dependent Kinases (CDKs)*** Cell cycle is tightly regulated by members of the CDK family and cyclin regulatory subunits [118, 119]. CDKs are a family of STKs that are important in controlling entry into and progression through the cell cycle, and their activity is regulated at several levels [120] such as the binding of each CDK to a specific cyclin partner. As cells respond to the presence of mitogens, they enter the cell cycle, and the level of cyclin D1 increases leading to the activation of CDK4 and CDK6. These kinases phosphorylate the retinoblastoma tumor suppressor protein pRb, thereby abrogating its inhibition of members of the E2F family of transcription factors. The inhibition of the transcription factors triggers a program of gene expression that results in entry into the S phase of the cell cycle. The deregulation of the pRb pathway and CDK4 are important in cancer

progression [121]. Tumors may express abnormally high levels of the cyclin D1 protein, contain multiple copies of the cyclin *D1* gene [122], or contain mutations or deletions of the *pRb* gene [123–125].

***Protein Kinase C (PKC) and Protein Kinase A (PKA)*** The PKC superfamily is a member of STKs that are stimulated by $Ca^{2+}$, phospholipid, and diacylglycerol and play a critical role in intracellular signal transduction for cell proliferation and differentiation, gene expression, the control of cell differentiation and growth, and vascular permeability and proliferation. Thus, PKC enzymes are involved in cardiovascular disease, cancer, ischemia, inflammation, and central nervous system (CNS) disorders [126]. The mammalian PKC family consists of 12 different isosymes as α, βI, βII, γ, δ, ε, ζ, η, θ, ι, λ, and μ that are classified into 3 subfamilies of novel, conventional, and atypical on the basis of their cofactor requirements [127]. PKC isoforms have specific cellular localization in various tissues and cell types [128] and distinct and in some cases opposing roles in cell growth and differentiation [129, 130]. For example, the induction of β isoform in response to hyperglycemia in aortic, cardiac, retinal, and renal tissues is shown to be associated with diabetic manifestations, such as neuropathy, nephropathy, angiopathy, macular edema, retinopathy, and cardiomyopathy [131, 132]. Increased levels of PKC have been associated in several cell lines including lung [133], gastric carcinomas [134], and breast [135] with malignant transformation. However, no clear relationship has been observed in vivo. For example, with PKCβ expression being found to increase, remain the same, or decrease in colon tumors when compared with normal epithelium [136].

The PKA family represents another group of STKs and has at least two isoforms, type I and type II [137] and is known to mediate cyclic adenosine monophosphate (cAMP) effects and is involved in many endocrine and nonendocrine cell responses. The PKA holoenzyme is a tetramer consisting of a dimeric regulatory subunit (R2) and two monomeric catalytic subunits (C2). Both PKC- and PKA-mediated signaling cascades have crucial roles in glial cell proliferation and differentiation [138, 139].

***Rho Kinase*** The ROCKs, or Rho kinases, are STKs that are involved in a variety of biochemical signal transductions in the cells. In addition to cytoskeleton rearrangement, cell migration, motility, proliferation, and neurite outgrowth, this enzyme has a crucial role in cardiovascular function such as smooth muscle contraction through calcium sensitization, an event that controls vascular vessel tone [140, 141]. Thus, ROCK inhibitors have been designed as therapeutic agents for cardiovascular diseases.

***Phosphoinositide 3-kinase (PI3K)*** The PI3K has a protein kinase fold. Upon growth factor receptor stimulation, the PI3K family of enzymes is recruited and produces 30 phosphoinositide lipids that act as second messengers by binding to and activating diverse cellular target proteins [142]. For

example, the production of the phosphoinositide polyphosphate PI(3,4,5)P3, or PIP3, by PI3K has critical roles in pathways governing apoptosis, differentiation, cell proliferation, and migration; hence, abnormal regulation of PIP3 levels and the levels of their lipid products are associated with a variety of cancer types.

It is believed that resistance to radiation treatment in a number of cancers is due to activation of the PI3K pathway, suggesting that inhibition of PI3K could be an important strategy to reduce resistance in cancer radiation treatment [143].

***Glycogen Synthase Kinase 3 (GSK3)*** The GSK-3 is another cytosolic STK found in two closely related isoforms (α and β) with high homology (ca. 90 percent) at the catalytic domain [144] but distinct functional roles. Both isoforms are expressed ubiquitously in mammalian tissues. GSK3 is one of many signaling components downstream from the insulin receptor (IR). For example, GSK-3β plays a critical role in cancer (via angiogenesis, apoptosis, and tumorigenesis), CNS function (via the proteins tau and β-catenin), and glucose homeostasis [145]. Insulin-dependent glycogen synthesis is activated by the inhibition of GSK-3-dependent phosphorylation leading to lower plasma glucose. Thus, GSK-3β is a potential target for inhibitor design and treating type II diabetes [132, 146]. Additionally, due to critical roles of GSK-3 in cancer and CNS function, GSK inhibitors may have therapeutic potential for treating neurodegenerative diseases, bipolar disorder, stroke, and cancer [147]. Several potential inhibitors have been recently reviewed [148].

## 8 PROTEIN KINASES AS TARGETS FOR INHIBITOR DESIGN

Many reviews of protein kinase inhibitor discovery have appeared over the past 14 years [4, 11, 12, 38, 149, 150], and several central themes have emerged. This chapter will focus on inhibitors directed against critical molecular targets discussed above that are approved or are in clinical trials.

### General Overview of Protein Kinase Inhibitors (PKIs)

A number of agents have been approved by FDA as PKIs. These inhibitors include small molecules such as STI-571 (Gleevec) [151], ZD-1839 (Iressa) [152], and Sirolimus (Rapamycin) [153] and monoclonal antibodies such as Herceptin (Trastuzumab) [154], Erbitux (Cetuximub) [155, 156], and Bevacizumab (Avastin) [79] (Tables 1 and 2).

Several other PKIs are in clinical trials that target different stages of signal transduction such as binding to RTKs, intracellular signaling, and cell cycle regulation [150, 157, 158]. There are several issues to be considered in designing PKIs. First, the inhibitors must exhibit selectivity against a family or a member of PKs. Second, the PKs in different systems should be monitored

**TABLE 1 Selected Inhibitors of Protein Tyrosine Kinases in Market or Clinical Trials**

| Kinase Target | Agent | Trial (Disease) | Sponsor |
|---|---|---|---|
| ABL (c-Kit, PDGFR) | Gleevec (STI-571) | Approved (CML, GIST) | Novartis |
| EGFR | Iressa (ZD1839) | Approved (Lung cancer) | AstraZeneca |
| | Tarceva (OSI-774) | Phase III (Cancer) | OSI/Roche/Genentech |
| | Erbitux (Cetuximub) | Approved (Cancer) | Imclone |
| | ABX-EGF (mAb) | Phase II (Cancer) | Abgenix |
| | MDX-447 (mAb) | Phase I (Cancer) | Merck KgaA |
| | EMD 72000 (mAb) | Phase I (Cancer) | Merck KgaA |
| | Genistein | Phase II (Cancer) | NCI |
| | R3 | Phase II (Cancer) | Center of Molecular Immunology |
| | EMD-72000 (EMD-6200) | Phase I/II | Merck |
| | RH3 (mAb) | Phase II (Cancer) | York Medical Bioscience Inc. |
| EGFR, ERB2R | CI1033/PD183805 | Phase II (Cancer) | Pfizer |
| | EKB569 | Phase I (Cancer) | Wyeth-Ayerst |
| | GW2016 | Phase II (Cancer) | GlaxoSmithKline |
| | PKI166 | Phase I (Cancer) | Novartis |

| | | | |
|---|---|---|---|
| VEGFR (PDGFR, FGFR) | SU6668 | Phase I (Cancer) | Pharmacia Corporation |
| PDGFR (Flt-3) | CT53518 | Phase I (Cancer) | Millennium Pharmaceuticals |
| VEGFR | SU5416 | Phase III (Cancer) | Pharmacia Corporation |
| | PTK787/ZK222584 | Phase III (Cancer) | Novartis/Schering-Plough |
| | Anti-VEGFR2 mAb | Phase I (Cancer) | Imclone |
| | Bevacizumab (Avastin) | Approved (Cancer) | Genentech |
| | AE-941 (GW786034) | Phase II/III (Cancer) | AEterna |
| | CEP-7055 | Phase I (Cancer) | Cephalon |
| | CEP-5214 | Phase I (Cancer) | Cephalon |
| | VEGF-trap | Phase I/II (Cancer) | Regeneron |
| VEGFR (EGFR) | ZD6474 | Phase II (Cancer) | AstraZeneca |
| VEGFR (PDGFR) | SU11248 | Phase II (Cancer) | Sugen |
| VEGFR (RAF) | Bay 43-9006 | Phase III (Cancer) | Bayer/Onyx |
| VEGFR (PDGFR, c-kit, Flt-3) | AG013676 | Phase II (Cancer) | Pfizer |
| NGFR, Trk | CEP-2583 | Phase II (Cancer) | Cephalon |
| HER-2/neu | 17-AAG | Phase I (Cancer) | Kosan |
| | Trastuzumab (mAb) | Approved (Cancer) | Genentech |
| | 2C4 (mAb) | Phase I/II (Cancer) | Genentech |
| | CP-724,714 | Phase I (Cancer) | OSI Pharmaceuticals/Pfizer |
| | TAK-165 | Phase I (Cancer) | Takeda |
| | MDX-210 (mAb) | Phase I (Cancer) | Novartis |

**TABLE 2 Selected Inhibitors of Protein Serine/Threonine Kinases in Market or Clinical Trials**

| Kinase Target | Agent | Trial (Disease) | Sponsor |
|---|---|---|---|
| PKC, c-Kit, PDGFR | PKC412 | Phase II (Cancer, retinopathy) | Novartis |
| PKC | ISIS3521 | Phase III (Cancer) | ISIS Pharmaceuticals |
| | CGP41251 | Phase II (Cancer) | Novartis |
| | UCN-01 | Phase I/II (Cancer) | Kyowa Hakko Kogyo |
| | Bryostatin-1 | Phase II (Cancer) | Biotek |
| PKC-β | LY333531 | Phase I (Cancer) | Eli Lilly |
| | | Phase II/III (Diabetic neuropathy) | |
| | LY317615 | Phase I/II (Cancer) | Eli-Lilly |
| CDKs | Flavopiridol | Phase II (Cancer) | Aventis |
| | E7070 | Phase I (Cancer) | EISAI |
| | BMS-387032 | Phase I (Cancer) | Bristol-Myers Squibb |
| | UCN-01 | Phase II (Cancer) | Kyowa Hakko Kogyo |
| | CYC202 | Phase I (Cancer) | Cyclacel |
| CDK2 | CYC202 | Phase I (Cancer) | Cyclacel |
| MKK1 | PD184352 | Phase I (Cancer) | Pfizer |
| MEK | PD184352 | Phase II (Cancer) | Pfizer |
| | U-0126 | Phase I (Cancer) | Promega |
| MLK | CEP-1347 | Phase II (Neurodegeneration) | Cephalon |

| | | | |
|---|---|---|---|
| RAF | BAY43-9006 | Phase II (Cancer) | Onyx Pharmaceuticals/Bayer |
| | ISIS5132 | Phase II (Cancer) | Isis pharmaceuticals |
| | L-779,450 | Phase II (Cancer) | Merck |
| Ras | ISIS2503 | Phase II (Cancer) | Isis pharmaceuticals |
| | SCH66336 | Phase II (Cancer) | Schering-Plough |
| | BMS214662 | Phase I (Cancer) | Bristol-Myers Squibb |
| | R115777 | Phase II/III (Cancer) | Johnson and Johnson |
| mTOR | CCI779 | Phase II (Cancer) | Wyeth-Ayerst |
| | RAD001 | Phase I (Cancer) | Novartis |
| | | Phase II/II (Immunosuppressant) | |
| | Rapamycin | Approved (Immunosuppressant) | Wyeth-Ayerst |
| p38-MAPK | VX702 | Phase II (Inflammation; ACS) | Vertex Pharmaceuticals |
| | BIRB-796 | Phase III (Inflammation; RA; Crohn's; Psoriasis; Endotoxic shock) | Boehringer Ingelheim<br>Scios, Inc. |
| | SCIO-323 | Phase I (RA; Stroke; Diabetes) | Scios, Inc. |
| | SCIO-469 | Phase II (RA; Crohn's) | |
| PDK1 | UCN-01 | Phase I/II (Cancer) | Kyowa Hakko Kogyo |

since modulation of the same PK by an inhibitor may have beneficial effects and deleterious effects in two different systems. For example, inhibiting a PK required for triggering apoptosis could reduce ischemia-induced cell death in terminally differentiated cardiomyocytes but might also promote tumor formation in other organs or cell types [159]. Finally, even selective inhibition of a dysregulated PK in a particular disease state may be deleterious in a long-term use for other systems in which that same PK is not dysregulated but instead serves essential functions. For example, Herceptin (Trastuzumab, Genentech) is a monoclonal antibody that inhibits overexpressing cell surface HER2 tyrosine kinase receptor in patients with breast cancers. On the other hand, inhibition of HER2 by herceptin can cause severe cardiac dysfunction in some women receiving the therapy, suggesting a critical role for this receptor in cardiomyocyte survival [154, 159].

The comprehensive discussion of the structure–activity relationships of PKIs is beyond the scope of this chapter. For more details, readers can refer to some recent excellent reviews [13, 150, 157, 160–171]. Monovalent ligands targeting the ATP binding site, the substrate binding site, and the SH2 domain of clinically important PKs and some of novel approaches in designing PKIs such as allosteric binding site and bisubstrate analog inhibitors are discussed here.

***ATP Binding Site Inhibitors*** The PKs can be inhibited by compounds competing with ATP for binding to ATP binding site on the enzyme molecule. In general, ATP binding site inhibitors contain a rigid, largely planar structure, an aromatic ring system to complement the ATP binding pocket, and a hydrogen bond donor–acceptor pair. These groups of compounds exhibit similar hydrogen bond attachment points at the interlobe linker (hinge region) through their chemical scaffolds [172]. These hydrogen bond interactions are complemented with an additional stabilizing network of hydrophobic interactions between the planes of their aromatic ring systems and hydrophobic amino acid side chains in the binding pocket [28].

Selectivity and potency have to be considered in designing ATP binding site inhibitors. In recent years, considerable progress has been made to increase specificity and potency of inhibitors toward PKs. This strategy has led to the development of two drugs, Iressa and Gleevec [60]. Several other compounds are in advanced clinical trials.

*Selectivity* These compounds have structural similarities with ATP and bind to the enzyme instead of ATP; therefore, the kinase cannot phosphorylate proteins and signaling halts. ATP binding site inhibitors of PKs were initially thought to be unsuitable drug targets. The high degree of structural conservation in this binding site within all PKs led to the initial belief that highly selective small-molecule PKIs targeting the ATP pocket would be difficult to generate [158, 173–177].

Since the kinase protein family contains approximately 518 members, the discovery of highly specific kinase inhibitors has been slow, but substantial progress has been made. Indeed, contrary to initial concerns, there is sufficient sequence and conformational diversity in ATP binding site [177] to identify selective ATP-competitive inhibitors. The screening of unbiased chemical libraries has led to identification of several relatively selective ATP competitors [178, 179]. Several different kinds of inhibitors directed toward the ATP binding site [160, 161, 180–182] are now in clinical development as anticancer agents (Tables 1 and 2), immunosuppressant (Rapamycin) [153], neurodegeneration [CEP-1347 (KT7515)] [183], and antidiabetic neuropathy (LY333531) [184, 185]. ATP binding site inhibitors have structure diversity and include purines, quinazolines, pyrazolopyrimidines, pyrrolopyrimidines, pyridopyrimidines, maleimides, oxindoles, phthalimides, isoquinoline, naphthyridinones, quinolines, and several other classes of compounds [160, 161, 180]. Some of these compounds have shown potent inhibitory action against particular kinases ($K_i$ in nanomolar–picomolar range) [150, 160, 180–182] and can even distinguish between closely related tyrosine kinases (e.g., between Src kinases) [186, 187] and so are very promising drug leads. Although selective lead inhibitors competitive with ATP have been identified for specific PKs indicating that the initial belief was misguided, however, the process is still labor intensive mainly due to the high structural homology between protein kinase active site regions.

One approach to achieve selectivity is to change the substitution patterns in one category of the PKIs. This strategy has been used in the case for several benzopyranone-containing compounds (flavonoids) (e.g., flavopiridol, quercetin, genistein, apigenin, desmal, and PD98059) (Fig. 6). Flavonoids, in general, are competitive inhibitors with respect to ATP and lack selectivity for PTKs over STKs [188, 189]. By establishing the structure–activity relationships of flavonoids, relatively selective inhibitors have been identified [163, 190]. Flavopiridol is a CDK inhibitor (e.g., CDK1, CDK2, CDK4, and CDK7) in clinical development [191]. Genistein has been shown to inhibit tyrosine kinases such as the EGFR and $pp60^{src}$, but scarcely to inhibit the activity of serine and threonine kinases such as cAMP-dependent kinases [192, 193]. PD98059 (2′-amino-3′-methoxyflavone) is a potent, selective, and cell-permeable inhibitor of MAP kinase kinase (also known as MAPK/ERK kinase or MEK kinase) by selectively blocking the activity of MEK and/or inhibiting the phosphorylation and activation of MAP kinase [194–199].

In general, the goal is an $IC_{50}$ that is at least 100-fold lower for the target kinase over other enzymes. Sometimes it is necessary to develop not entirely selective inhibitors. For example, the antiangiogenic compound SU6668 elicits potent antiangiogenic and proapoptotic effects in vivo by inhibition of KDR, FGFR1, and PDGFRβ receptors [81]. In cancer, a combination of CDKs inhibitory and antiapoptotic (e.g., Akt) effects by a drug targeting Bcr-Abl

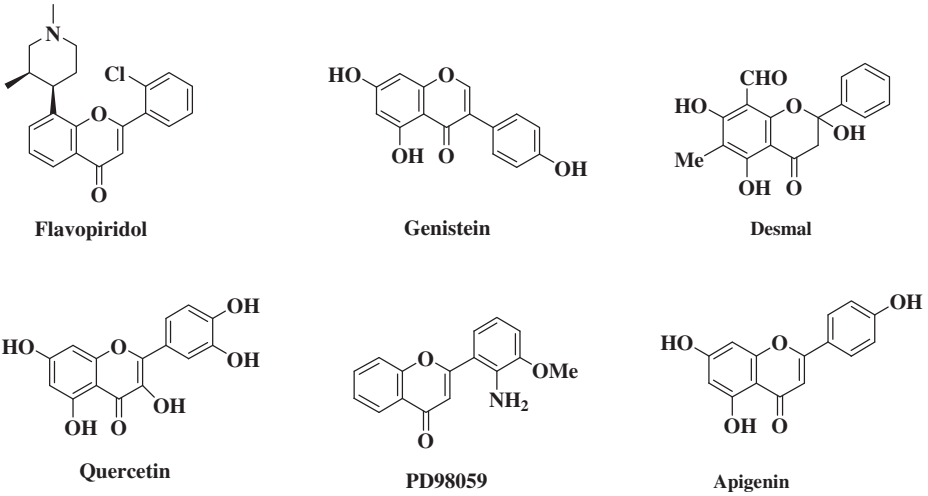


**Figure 6** Chemical structures of some flavonoids having protein kinase inhibition activity.

could be beneficial for anticancer effects. However, a p38-MAPK inhibitor for treatment of inflammatory diseases cannot be tolerated to inhibit CDKs or Akt because of enhanced toxicity.

*Potency* The PKs and a plethora of many nonkinase enzymes utilize ATP as a common substrate with different $K_m$ values. The ATP binding site inhibitors must have sufficient potency to compete with the high ATP concentration (2 to 10 mM) in vivo [158]. Enzymes with a low $K_m$ for ATP could be inhibited if ATP binding site inhibitors do not bind to the target kinase with an extremely high affinity: several orders of magnitude higher than that of ATP. In vivo, the inhibitor will be present in concentrations typically in the mid to high nanomolar range, whereas the intracellular concentration of ATP is millimolar. As a result, the ATP-binding-site-directed inhibitors usually exhibit a significantly lower potency under in vivo conditions than they do in vitro.

***Substrate Binding Site Inhibitors*** Inhibitors competing with the peptide or protein substrate for binding to substrate binding sites have also been described [180, 200–202]. These peptides are frequently in the form of uncharged phenylalanine analogs [13, 203] and comprise of peptides containing the substrate consensus motif to generate greater specificity. PKs exhibit more sequence diversity in substrate binding site and peptide and protein substrate recognition; therefore, peptide-based substrate binding site inhibitors with high specificity might be easier to discover, and they might prevent phosphorylation of some substrates but not others. These peptide-based inhibitors do not need to compete with physiologic ATP. Despite generating high selec-

tivity toward PKs, substrate peptide-based inhibitors have relatively weak inhibitory performance ($K_i$ in the high micromolar range) compared to the most potent ATP analogs [180]. There are issues such as bioavailability and susceptibility to proteolytic degradation that have to be considered when designing these inhibitors. Some closely related PKs exhibit overlapping specificities with respect to small peptides. Other than peptide-based inhibitors, no small organic molecules have been discovered to show a competition pattern at the protein–substrate binding site of PKs, possibly due to the fact that the substrate binding groove is relatively solvent exposed and shallow.

***Allosteric Inhibitors*** Regions outside the ATP binding site such as the activation loop can be used and accessed for designing inhibitors. Binding of inhibitors to this region may cause significant conformational changes in their target kinases [28]. These compounds are likely to be more selective than compounds that simply compete directly with ATP alone. For example, BIRB 796, a slow-binding inhibitor of p38 MAP kinase, was designed based on structure optimization of several compounds and their mode of binding with P38 MAP kinase in the crystal structure. BIRB 796 binding to a region that is normally occupied by the phenylalanine of the Asp-Phe-Gly ("DFG") motif converts the enzyme to a catalytically inactive conformation by changing the conformation of the active site [204]. Gleevec, a recently approved drug against CML, interacts with the aromatic side chain of the conserved phenylalanine of the DFG motif and displaces the activation loop from its normal position, resulting in trapping the unactivated form of c-Abl [28] (PDB codes 1FPU [26], 1IEP [205]).

***SH2 Domain Inhibitors*** SH2-containing targets in PTKs have been the subject of great interest by several researchers [206]. Ligands that antagonize the Src SH2 domain-dependent protein–protein interactions may provide possible therapeutic agents [29, 207–210]. SH2 domains generally recognize specific short peptide motifs containing a phosphorylated Tyr (pTyr) in the context of specific residues found within 3 to 5 amino acids C-terminal to the pTyr [31]. The Src SH2 domain has been the subject of intensive research. The SH2 inhibitors will be discussed when considering Src kinase inhibitors (Section 9 of "Combination Products: Drugs and Devices") in greater details.

***Bisubstrate Analog Inhibitors*** Bisubstrate analog inhibitors could be designed to enhance specificity and potency of PKIs. Bisubstrate analog inhibitors are designed to mimic two natural ligands that simultaneously associate with two essential binding sites of kinases. The combination of two substrates required by the target enzyme into a single molecule by an appropriate linker makes it less likely that both components will be recognized by other enzymes; therefore, enhanced enzyme specificity may be expected [211]. Additionally, if one of the substrates demonstrates specific binding to its binding pocket, a greater selectivity may arise [212, 213]. Using bisubstrate analog

inhibitors for PKs allows converting a low-affinity peptide substrate to a selective and moderately potent inhibitor for these enzymes with affinities at a low micromolar–high nanomolar range.

In general, an optimal bisubstrate analog inhibitor for PKs should have the following properties. The inhibitor should (1) inhibit the binding of both natural substrates to their binding pockets particularly by competitive inhibition, (2) be designed based on the distance, geometry, and orientation of natural substrates and their essential chemical scaffolds during the transition state, (3) enhance specificity against one particular PK, and (4) show optimal physico-chemical properties, bioavailability, and low toxicity. During the last 16 years, many different approaches have been examined for the possibility of using bisubstrate analog inhibitors against PKs [214]. These strategies fall into four categories: (1) sulfonamides and sulfonylbenzoyl derivatives [186, 215–217], (2) carboxylic acid derivatives [218–220], (3) phosphodiester derivatives [221], and (4) dipeptidyl derivatives [53, 222, 223]. The first three categories target ATP binding site and substrate binding site of PKs. The last category of compounds was designed to target SH2 and SH3 domains or the SH2 domain and substrate binding site. None of these strategies have all the characteristics for an optimal bisubstrate analog inhibitor for PKs, but substantial progress has been made in this direction.

Most bisubstrate analogs have been designed to mimic the phosphate donor (ATP) and the acceptor components (Ser-, Thr-, or Tyr-containing peptides) based in part on mechanistic and structural features of a predicted transition state for PKs. We synthesized a potent and selective peptide–ATP bisubstrate inhibitor ($K_i = 370\,\text{nM}$) that binds to ATP binding site and substrate binding site of the IRK (Fig. 7) that behaved as a competitive inhibitor versus both ATP and peptide substrates based on the kinetic analysis. A specific distance (5.7 Å) between these two components was required to achieve potent inhibition [224]. The X-ray crystallographic studies indicated that both the nucleotide and peptide portions of the bisubstrate inhibitor were binding to ATP and protein substrate binding sites, respectively (Fig. 7). The bisubstrate analog inhibitor was only a modest inhibitor against Csk since the preferred substrate peptide sequence for Csk is quite different from IRK.

## Alternative Strategies in Designing Protein Kinase Inhibitors

A number of alternative strategies have been developed for designing PKIs including substrate docking site, irreversible, and molecular chaperone inhibitors. In addition to substrate binding site, it has been shown that MAP kinases [225–227] and Csk [228] employ remote-docking-based mechanisms for substrate recognition. We anticipate that elucidation of the substrate recognition mechanisms will reveal novel approaches in developing protein substrate-competitive inhibitors.

A number of irreversible inhibitors of the erbB family of protein tyrosine kinases have been designed [229] based on the reaction of the nucleophilic Cys thiol in ATP binding site with Michael acceptor groups in the inhibitors

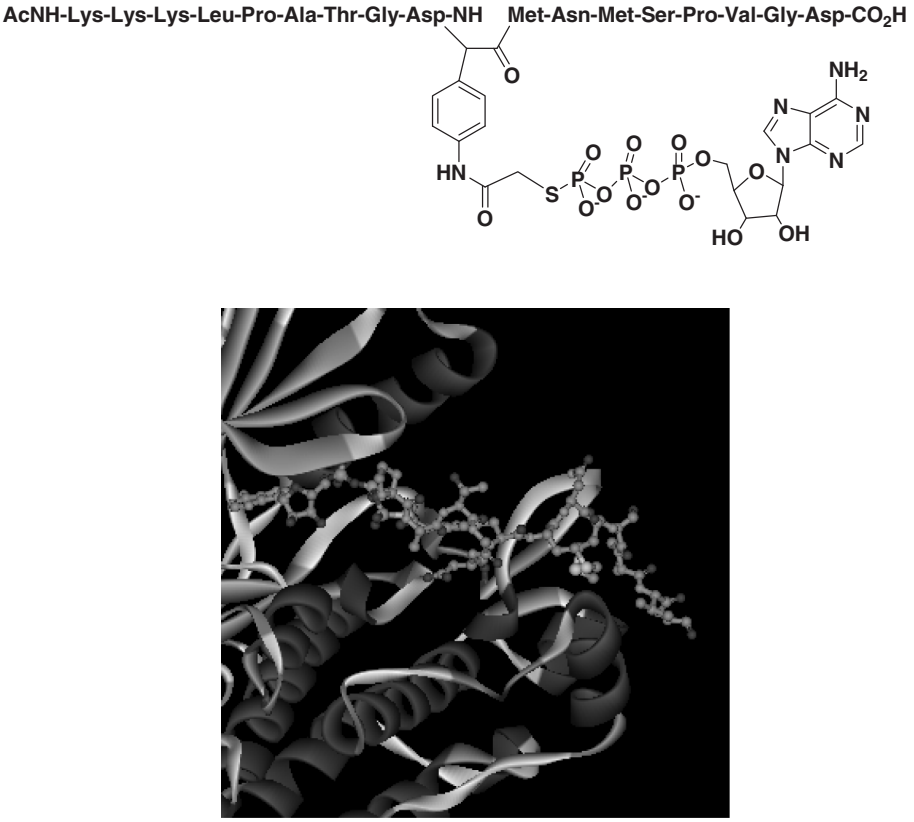


**Figure 7** Designed bisubstrate analog inhibitor for IRK and the three-dimensional structure of the bivalent tyrosine kinase inhibitor (ball-and-sticks vectors) bound to the ATP binding site and substrate binding site of IRK catalytic site (PDB code IGAG) [224].

such as CI-1033 (PD183805) [230] and EKB-569 [231] that are now in clinical trials.

Molecular chaperones such as heat shock protein 90 (Hsp90) are involved in stabilizing the conformation of PKs [232]. A number of compounds such as geldanamycin, radicicol, and 17-AAG have been studied as Hsp90 inhibitors and indirect PKIs [232–234]. It remains to be seen if any of these strategies will translate to potential therapeutic agents.

## 9 SELECTED INHIBITORS OF PROTEIN KINASES ON MARKET OR IN CLINICAL TRIALS

Tables 1 and 2 show most of the PKIs currently in clinical trials or on the market for several diseases. There are two types of inhibitors, monoclonal

antibodies (mAbs), which are directed at the extracellular domain of various RTKs, and small-molecule inhibitors. Most of PKIs are targeted against cancer, but several PKIs are designed for the treatment of a number of chronic conditions, including inflammatory and cardiovascular diseases. Both active and inactive protein kinase structures have been used in inhibitor binding studies. Detailed discussion of all these inhibitors is beyond the scope of this chapter. Herein several PKIs approved or in advanced phases of clinical studies are discussed based on validated target enzymes.

## Receptor Tyrosine Kinase Inhibitors

Recently, considerable effort in cancer drug discovery has been directed at the growth factors and growth factor receptors. There are two approaches for designing inhibitors against RTKs, targeting the extracellular ligand binding and intracellular kinase domains. Several agents such as antibodies have been designed to disrupt phosphorylation sequence by competing for receptor binding. Alternatively agents have been designed to prevent tyrosine kinase autophosphorylation. Several of the compounds developed as RTK inhibitors target more than one closely related receptor from this family.

***Targeting the Extracellular Domains of EGFR*** It has been found that compounds that interfere with EGF binding to its receptor have utility as antiprolifereative agents [235]. HER-2 receptor is a member of the EGFR family of RTKs that is involved in cell proliferation and differentiation in developing embryo and in adult tissues. It is known that cancerous breast cancer tissues overexpress the HER2 receptors on the cell surface, triggering the cell division and multiplication in tumor growth. Approximately 30 percent of all women with metastatic breast cancer overexpress the HER2 protein [236]. The HER2 extracellular domain has a constitutively open structure based on shown structural studies [237]. Genentech Inc. developed the humanized monoclonal antibody Trastuzumab directed against the extracellular domain of HER-2. The mAb received an approval of the FDA in 1998 under the name Herceptin that binds to HER2 receptor sites on the surface of breast cancer cells, blocking the receptor sites, and preventing the tumor growth by interrupting the growth signal to intracellular domains. The structure of HER2 receptor in complex with Herceptin Fab shows that Herceptin binds HER2 on the C-terminal portion of domain IV (Fig. 8). Herceptin proved to be effective in 15 percent of patients with HER2-overexpressing metatstatic breast cancer [154].

Chimeric antibody Erbitux (Cetuximub) developed by Imclone is another antibody that blocks the extracellular domain of ErbB1 and prevents signaling that triggers colorectal tumor growth. Erbitux was approved in February 2004 for the treatment of metastatic colorectal cancer [155, 156]. Other promising EGFR antibodies include Thermacin h-R3 (Cima-her) for the treatment of head and neck tumors in combination with radiotherapy.

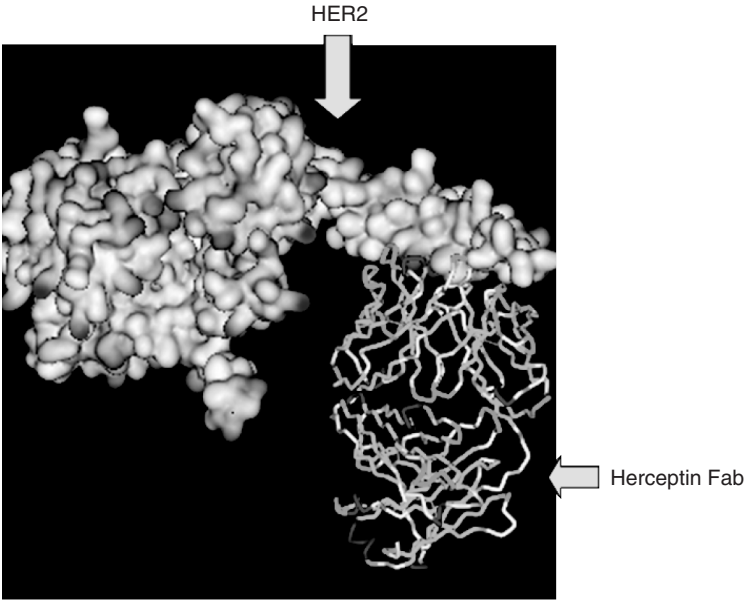


**Figure 8** Herceptin binding site. The backbone of HER2 is shown in surface format contacting the Herceptin Fab (tube format). The HER2 extracellular domain has an open structure and is a noncatalytic domain (PDB code 1N8Z) [237].

***Targeting the Intracellular Domains of EGFR*** Several potent EGFR inhibitors that bind to the intracellular kinase domain of EGFR are currently on the market or in clinical trials as anticancer agents (Table 1). Among these are ZD1839 (Iressa, gefitinib, Fig. 9), developed by AstraZeneca, and OSI-774 (Tarceva, Fig. 9), co-developed by OSI Pharmaceutical, Roche, and Genentech.

ZD1839 showed marked efficacy against several cancers, for example, prostate cancer and non-small-cell lung carcinoma [238], in human clinical trials and is approved by FDA for the treatment of metastatic non-small-cell lung cancer [239]. Clinical studies with this and other EGF receptor antagonists have not been uniformaly successful [240, 241]. For example, pulmonary toxicity was observed in several patients [242]. ZD1839 exhibited specificity for the growth inhibition of EGF-stimulated cells as compared to non-EGF-stimulated cells in cell cultures of KB human tumors tyrosine kinases with an $IC_{50}$ of 0.08 to 0.09 μM and 3.64 μM, respectively. ZD1839 did not exhibit any significant inhibitory effect on other related kinases such as MAPK, PKC, and KDR [238].

OSI-774 (Tarceva, erlotinib), an orally active and reversible potent inhibitor of EGFR tyrosine kinase with an $IC_{50}$ of 2 nM, has shown promising results in clinical trials in patients with non-small-cell lung carcinoma, head and neck, prostate, and cervical cancers (29 to 34 percent disease stabilization)

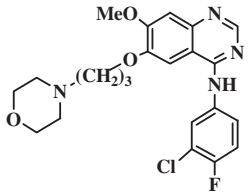

ZD1839 (Iressa™)

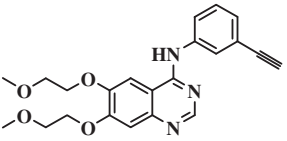

OSI-774 (Tarceva)

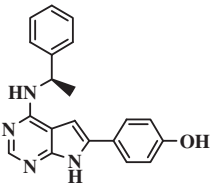

PKI116

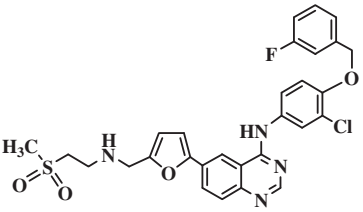

GW2016

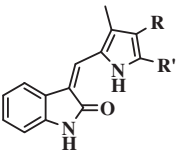

SU5402 R = $CH_2CH_2CO_2H$, R' = H
SU5416 R = H, R' = Me
SU6668 R = $CH_2CH_2CO_2H$, R' = Me

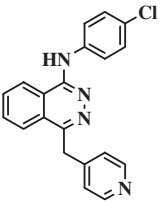

PTK787

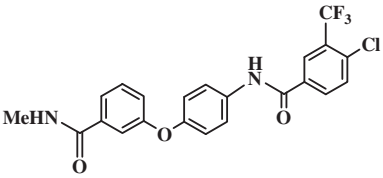

Bay 43-9006

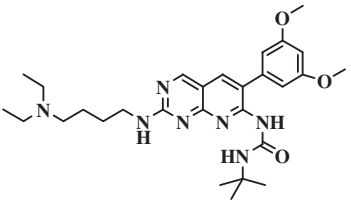

PD173074

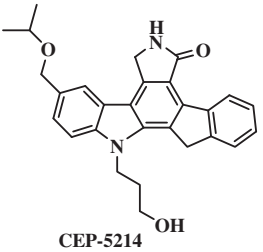

CEP-5214

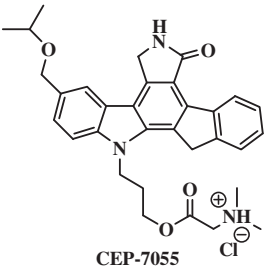

CEP-7055

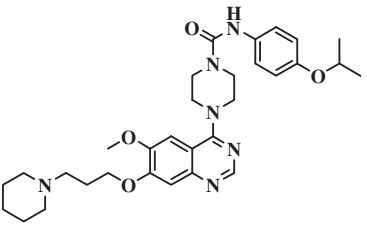

CT53518

**Figure 9** Chemical structures of RTKIs.

[243]. OSI-774 is currently in phase III clinical trials for several cancers such as pancreatic cancer and non-small-cell lung cancer. Structural studies have shown that OSI-774 binds to ATP binding site of [243] in the active form of EGFR [244].

A dual inhibitor of the HER-1 and HER-2 tyrosine kinases developed by Novartis, PKI116 is in phase I clinical trials against head and neck, breast, prostate, urinary, bladder, ovary, gliomas, and gastrointestinal tract cancers (Fig. 9). PKI116 had $IC_{50}$ values in the nanomolar range for the inhibition of the intracellular domain of HER-1 and HER-2 tyrosine kinases [245]. GW2016 (Fig. 9) is a dual tyrosine kinase inhibitor of ErbB-2 and EGFR developed by GlaxoSmithKline for the treatment of solid tumors [246].

***VEGFR and FGFR Inhibitors*** Several RTKIs have been designed to inhibit angiogenesis [247] as VEGFRIs that have application in anticancer drug development. In angiogenesis the new blood vessels are formed for solid tumors [248].

Bevacizumab (Avastin; rhuMab VEGF), an anti-VEGF humanized antibody developed by Genentech, has been approved by the FDA as an antiangiogenic anticancer drug and as a first-line therapy for metastatic colorectal cancer. Bevacizumab has a long terminal half-life of 17 to 21 days in humans [79].

The indolin-2-one family of antiangiogenic molecules, SU5402, SU5416, and SU6668 (Fig. 9), are potent inhibitors of VEGFR tyrosine kinase activity. Members of this family have a functionalized methylpyrrole ring attached to C3. The structure of the inhibitor SU5402 (Fig. 9) bound to FGFR1 tyrosine kinase domain (Fig. 10) [19] identifies the orientation of this ring in the complex. Specificity of SU5402 for FGFR1 derives in part from a hydrogen bond between the carboxyl function of SU5402 and the side chain of $Asn^{568}$ at the end of the hinge (Fig. 10).

SU5416 (Fig. 9) was developed by Sugen Inc. (Pharmacia) as a highly potent antiangiogenic antitumor agent that advanced to phase III clinical trials against glioma, Kaposi's sarcoma, and colorectal cancers by selectively targeting VEGFR-2 [248]. SU5416 enters the endothelial cells and binds to intracellular domains of VEGFR. SU5416 failed to exhibit efficacy in combination with chemotherapy [249]. Clinical trials for SU5416 have been halted by Pharmacia recently, but it remains a valuable biological tool for study of signal transduction pathways involving VEGF and its receptor (VEGFR).

SU6668, a close analog of SU5416, is another clinical angiogenesis inhibitor developed by Sugen Inc. (Fig. 9) and is undergoing various phase I clinical trials against advanced solid tumors. SU6668 is less selective than SU5416 and inhibits two other tyrosine kinases including EGFR and PDGFR in addition to VEGFR2 [81, 250, 251].

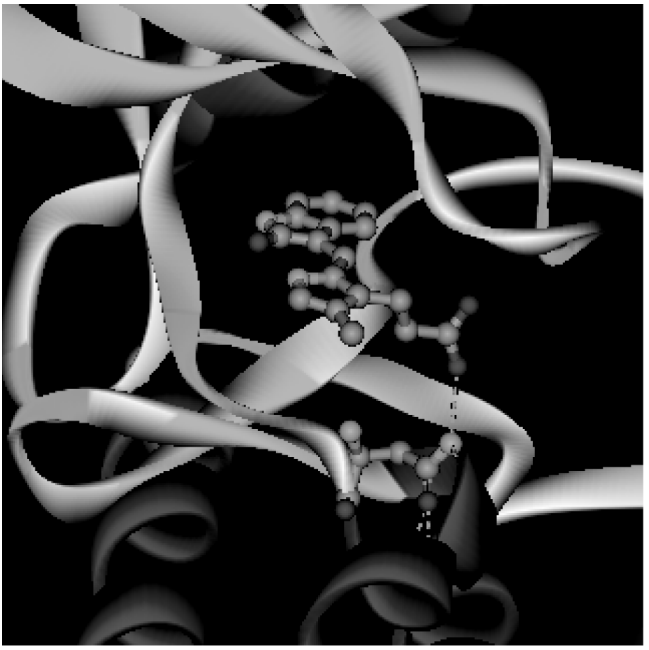

**Figure 10** Crystal structure of the tyrosine kinase domain of FGFR1 (ribbon) in complex with SU5402 inhibitor (ball-and-sticks vectors) (PDB code 1FGI) [19]. The hydrogen bonding is shown between the carboxyl function of SU5402 and the side chain of $Asn^{568}$.

PTK787/ZK222584 (Fig. 9) developed by Novartis/Schering has been introduced to target the VEGFR family [252–256] such as VEGFR1 and VEGFR2. PTK787 is effective in combination therapy against solid tumors but has limited success in monotherapy. PTK787 is in phase III clinical trials for colorectal cancer and is one of the most selective VEGFR kinase inhibitor described to date [257–259].

Other nonselective inhibitors targeting several RTKs such as VEGFRs, PDGFR, c-kit, and Flt-3 have been introduced [260]. These compounds are in several stages of clinical trials and have to be monitored for their toxicity profile because of their nonselective inhibitory effects. Bay 43–9006 (Fig. 9), an RAF kinase inhibitor and a nonselective inhibitor of RTKs including VEGFRs, is currently in phase III trials for metastatic renal-cell carcinoma [79]. PD173074 (Fig. 9) is another nonselective inhibitor of RTKs that is a potent inhibitor of the type-1 FGFR tyrosine kinase (FGFR-1) and VEGFR-2 tyrosine kinases [261].

CEP-7055, a water-soluble prodrug of CEP-5214 (Fig. 9), is currently in phase I clinical trials as an antiangiogenic antitumor agent. CEP-7055 was reported to be a potent inhibitor of all three types of VEGF receptor tyro-

sine kinases, VEGFR-1, -2, and -3 tyrosine kinases, with $IC_{50}$ values in nanomolar range and inhibited tumor formation and growth up to 90 percent in a variety of human tumor models [262]. The prodrug CEP-7055 shows a similar pharmacodynamic of the drug CEP-5214 but has a better pharmacokinetic profile.

***PDGFR Inhibitors*** SU6668 has been shown to be (Fig. 9) a nonselective PDGFR tyrosine kinase inhibitor since it also inhibits both VEGFR-2 and EGFR tyrosine kinases at similarly low concentrations [81, 251]. Recently, CT53518 (also known as MLN518) (Fig. 9) developed by Millenium Pharmaceuticals as a potent inhibitor of Flt-3, entered clinical trials against AML with impressive clinical results. CT53518 delays disease progression in a mouse model of Flt-3-associated AML and is proapoptotic in a human AML cell line [260].

## Nonreceptor Tyrosine Kinase Inhibitors

***Bcr-Abl Inhibitor*** STI-571 (Fig. 11), an ATP binding site inhibitor developed by Novartis, is very potent against tumorigenic fusion protein Bcr-Abl as described by Druker and co-workers [187, 263]. This compound was approved by the FDA in May 2001 for the treatment of CML under the name Gleevec (Glivec, imatinib mesylate) [181, 182], as the first small-molecule PTKI. Detailed pharmacological profile of STI-571 has been extensively reviewed [150, 157, 245, 264]. The drug induces 100 percent remission in the

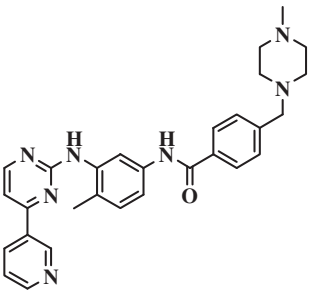


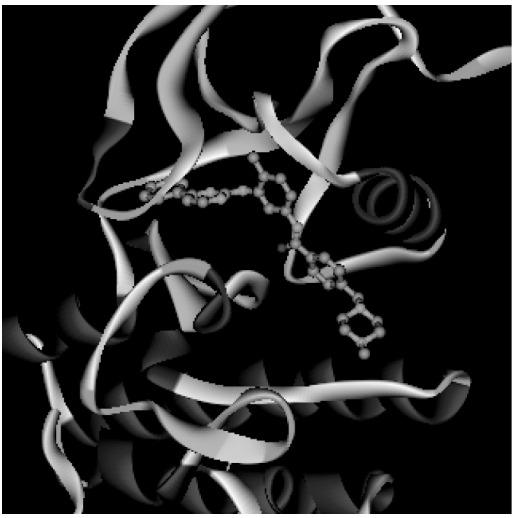

**Figure 11** Crystal structure of Gleevec (ball-and-stick vectors) with the catalytic domain of cAbl (ribbon). Gleevec targets a unique inactive conformation (PDB code 1IEP) [205].

prefully malignant chronic phase of the disease, but in advanced phases such as aggressive blast stage of the disease tends to be relatively short-lived, due to the development of de novo resistance to the drug by *BCR-ABL* gene amplification, multiple-point mutations within the kinase domain of Abl, and reactivation of the tyrosine kinase activity of the Bcr-Abl oncoprotein [242, 265, 266]. Azam et al. [242] proposed that point mutations cause steric hinderance of drug occupancy and allosteric destabilization of the autoinhibited Abl conformation. Three other kinases have been reported to be inhibited by Gleevec: the stem cell factor receptor (c-Kit), Abl-homolog Arg, and PDGF receptor [267, 268]; therefore, its use for the treatment of gastrointestinal stromal tumors in which c-Kit is overexpressed by mutation has been explored [269–271].

An inactive distinct conformation of Bcr-Abl is targeted by STI-571, making it more selective than other PTKI inhibitors. The pyridine and the pyrimidine rings of the drug are surrounded by a hydrophobic pocket in the ATP binding site (Fig. 11). The nonaromatic scaffold of the molecule is between the activation domain and the C helix, locking the kinase in an inactive conformation [26, 205]. The aspartic acid of the DFG motif in the activation domain cannot coordinate the magnesium at the catalytic site due to flipping of aspartic acid and phenyl alanine in this motif in the inactive conformation. The gatekeeper residue in Abl is $Thr^{315}$, to which Gleevec hydrogen bonds. A bulkier nonpolar residue replaces the threonine in many other kinases such as Src kinases; hence access to the pocket by STI-571 is blocked in these kinases [272]. A number of other kinases use this extra space created by $Thr^{315}$ that is not used by ATP for competing selectively with ATP [273–275].

PD173955 is a more potent inhibitor of Abl than Gleevec [205]; however, it can bind to both active (Fig. 12) and inactive conformations of Abl. Therefore, PD173955 is a less selective inhibitor than Gleevec [276].

***Src Kinase Inhibitors*** Although no drug has been approved as Src kinase inhibitor yet, research efforts in this direction have advanced significantly in recent years utilizing structure-based design, combinatorial chemistry, and high-throughput screening in designing numerous potent Src inhibitors with varying chemical structures [166]. It is hoped that Src inhibitors will soon emerge as useful therapeutics for cancer therapy. As mentioned in Section 4 of "Combination Products: Drugs and Devices", Src kinases are made up of five components/domains: N-terminal, Src homology 2 (SH2), Src homology 3 (SH3), kinase (catalytic, including NT and CT lobes), and C-terminal noncatalytic domain (Fig. 4). Currently, ATP binding site is an important target for the design of Src kinase inhibitors through an ATP-competitive inhibition mechanism. Substrate binding site and SH2, SH3 domains have also been employed as important targets for Src kinase inhibition [29, 277].

*Src SH2 Domain Inhibitors* Antagonizing the protein–protein interactions of SH2 domains with activated receptors using peptides and peptidomimiet-

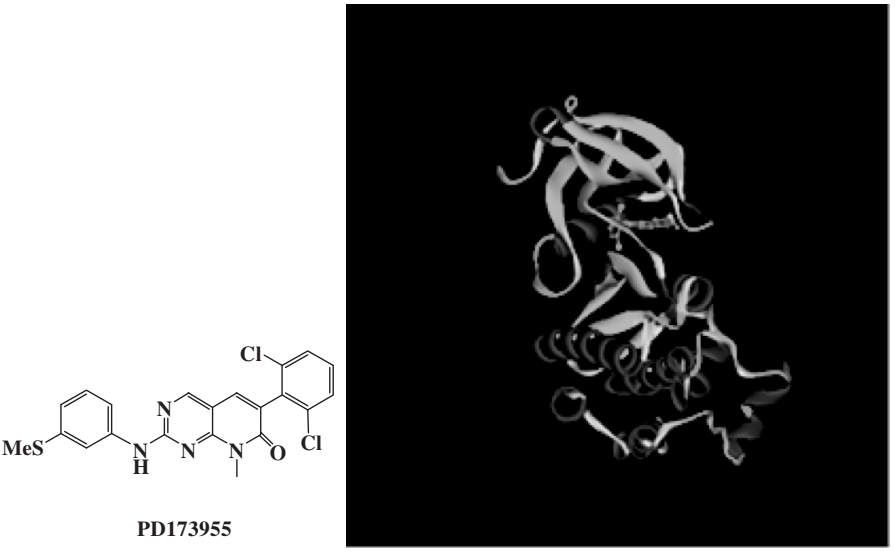


**Figure 12** Crystal structure of PD173955 (ball-and-sticks vectors) targeting the active conformation of Abl (ribbon) (PDB code 1M52) [205].

ics, has been pursued to inhibit the signaling pathways. For example, tetrapeptide pYEEI, a peptide with high affinity ($K_d$, 100 nM) for the Src SH2 domain [31], has two major interactions, that is, (i) electrostatic interactions between the pTyr phosphate moiety of the peptide and Arg residues embedded within a deep positively charged phosphotyrosine binding pocket (P site) on the SH2 domain and (ii) hydrophobic interactions between the lipophilic Ile side chain of the consensus sequence peptide and the hydrophobic pocket (P + 3 site) of the SH2 domain [278]. The crystal structure of pYEEI in complex with the Src SH2 domain reveals a very characteristic binding mode that can be described as a two-pronged plug engaging a two-hole socket (Fig. 13) [279]. The EE motif lies across the flat surface of the protein and plays the role of a linker delivering the two residues pTyr and Ile to respective binding pockets [280–282]. Tyrosyl motif-based analogs directed against pTyr binding pocket of the SH2 domain usually mimic highly charged phosphate functional groups, since recognition of ligands by SH2 domains mimics the electrostatic interactions of two negative charges and the ester oxygen for phosphate in phosphotyrosyl residues with positively charged amino acids, αA and βB arginine residues, located in pTyr binding pocket [283, 284]. Additionally pTyr residues themselves are not suitable for inhibitor design, due to the enzymatic liability of the phosphate esters to protein tyrosine phosphatases and the poor bioavailability of the doubly ionized phosphate moieties. The design of SH2 inhibitors has focused on peptidomimetic

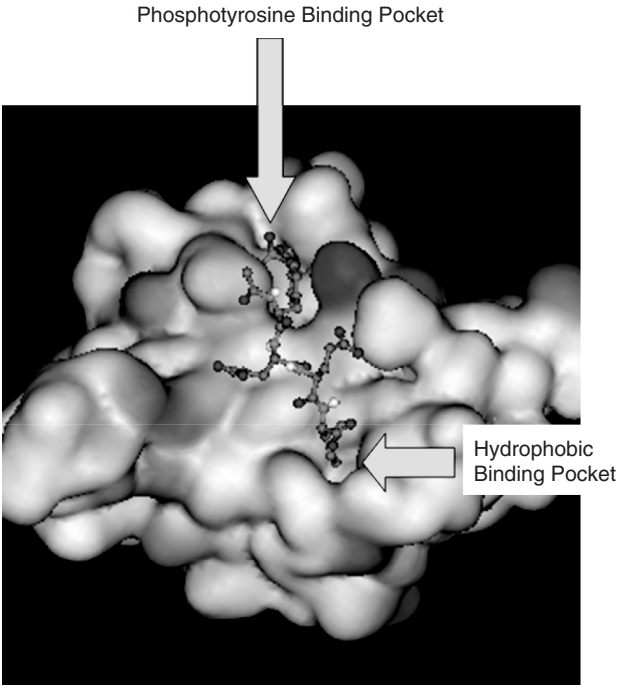


**Figure 13** X-ray crystal structure and interaction sites of pYEEI and the Src SH2 domain (PDB code 1SHD) [279].

modifications of cognate peptide sequences [285]. SH2-domain-directed pTy mimetics include phosphonate-based pTyr mimetics such as phosphonomethylphenylalanine (Pmp) and its analogs [286–294], carboxylic-acid-based pTyr mimetics such as malonyltyrosine or phenylalanine analogs and their derivatives such as carboxymethyl phenylalanine [292, 295–298], and uncharged pTyr mimetics [299], and conformationally constrained peptides [300]. Several of these inhibitors could enhance the kinase catalytic activity by switching the closed inactive to the open active conformation [46–49] by disrupting the intramolecular interactions between the $Tyr^{527}$-phosphorylated C-terminal tail and the SH2 domain. As a result, side effects might appear by the Src SH2 domain inhibitors. Much more attention has to be paid to the outcome of the Src SH2 inhibition and possible kinase activation.

*Src Protein Substrate Binding Site Inhibitors* Development of peptide inhibitors based on the peptide substrate motifs has been reported [301, 302]. The identified inhibitors are highly selective for $p60^{c\text{-}src}$ over other closely related Src family PTKs such as Lck and Lyn. However, with few exceptions, most of the inhibitors identified by this approach are only moderately potent with $IC_{50}$ values in the micromolar range [13]. The substrate site targeted

moiety, CIYKYYF, is reported to be one of the most potent peptides developed against $p60^{c\text{-}src}$ PTK ($IC_{50}$ = 0.5 μM) [303]. However, no structural, kinetics, and mechanistic studies have been proposed to confirm whether this peptide binds to the substrate binding site.

*Src ATP Binding Site Inhibitors* There exists a plethora of templates available to design ATP binding site inhibitors as Src kinase inhibitors [166, 304]. Although there are no Src inhibitor drugs on the market yet, several companies (e.g., Wyeth/Ayerst, Novartis, Ariad, Pfizer, Sugen/Pharmacia) are actively engaged in this area of drug discovery to create therapeutic agents as Src inhibitors for both osteoporosis and cancer treatment. AP22408 was developed as a bone-targeted Src kinase inhibitor that exhibited high affinities for hydroxyapatite and potent in vitro inhibition of Src-dependent, osteoclasts-mediated bone resoption [102, 304–306].

Several nonselective experimental compounds such as natural products (herbimycin A, radicolol, geldanamycin) [307–309], pyrazole[3,4-*d*]pyrimidine (PP1), PP2, SKI606, PD173955, PD180970, and SU6656 (Fig. 14) have been used for targeting the Src family of kinases. PP1 and PP2 may not be an optimal choice as Src family kinase inhibitors since they do not discriminate between different members of the Src family. For example, PP1 and PP2 inhibit Lck with $IC_{50}$ values of 50 to 60 nM. On the other hand, PP1 and PP2 also inhibited Csk, SAPK2a/p38 with $IC_{50}$ values of 520 to 1430 nM) [310]. The compound SU6656 is reported to inhibit Lck with an $IC_{50}$ of 0.04 μM as one of the most potent Lck inhibitors [179, 311]. A number of these compounds also inhibit Bcr-Abl, leading to cell cycle arrest and apopotosis of various Bcr-Abl-positive cell lines tested [312, 313]. For example, PP1 inhibits c-Abl ($IC_{50}$ = 30 nM) [310]. No molecules that demonstrate high selectivity between different Src family kinases have been identified so far. The natural product damnacanthal (Fig. 14; $IC_{50}$ for Lck, 17 nM) was 20- and 7-fold more selective for Lck over Fyn and Src, respectively [314]. The 3-(*N*-phenyl)carbamoyl-2-iminochromene (Fig. 14) [315] inhibited c-Src ($IC_{50}$ =120 nM), was < 285-fold more selective over Lyn, 570-fold more selective over Fyn, and 18-fold more selective with respect to Lck [310].

In general many of the Src kinase inhibitors published so far lack sufficient specificity desirable for clinical application or even pharmacological tools. Only few inhibitors have been tested without provoking toxicity in animal models. Src kinases inhibition seems to be still attractive since their inhibition might influence several important characteristics of tumorigenesis [84].

## Serine/Threonine Kinase Inhibitors

Serine/threonine kinases represent attractive targets for designing PKIs and therapeutic intervention. For example, the inhibition of various CDKs has been used in cancer treatment [316–318]. Fasudil (Fig. 15) is in the clinic and approved for treatment of cerebral vasospasm [319]. It has a significant

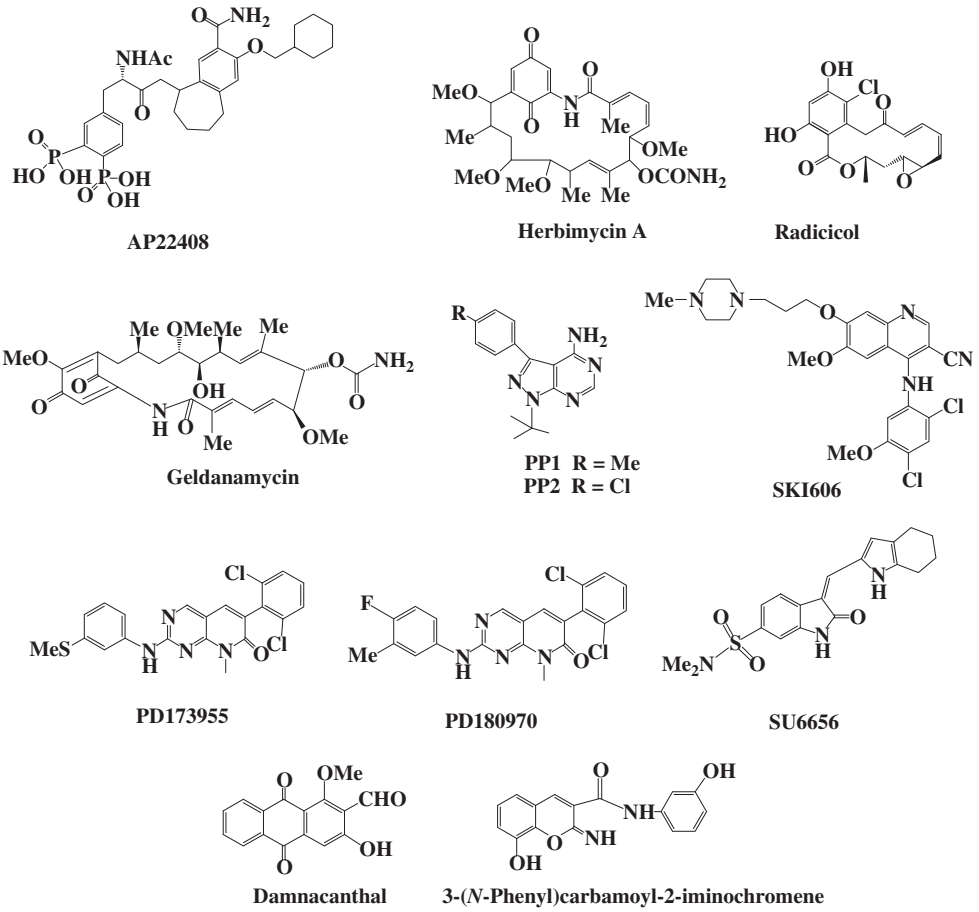


**Figure 14** ATP binding site inhibitors as Src Kinase inhibitors.

vasodilatory effect attributed to its potent inhibition of Rho-kinase signaling to myosin light-chain kinase. Some of important serine/threonine kinase inhibitors are discussed here.

***MAP Kinase Inhibitors*** Three families of stress-activated MAPKs are the p38-MAPKs, JNKs, and ERKs. We will describe important MAP kinase inhibitors in some detail.

*P38 MAPK Inhibitors* P38 MAP kinase such as p38α is activated in response to cellular stresses, growth factors, and cytokines. Activated p38α activates other kinases that phosphorylate heat shock proteins and transcription factors. Several proinflammatory and inflammatory cytokines, such as tumor necrosis factor-α (TNF-α) and interleukin 1β, are activated by increased activ-

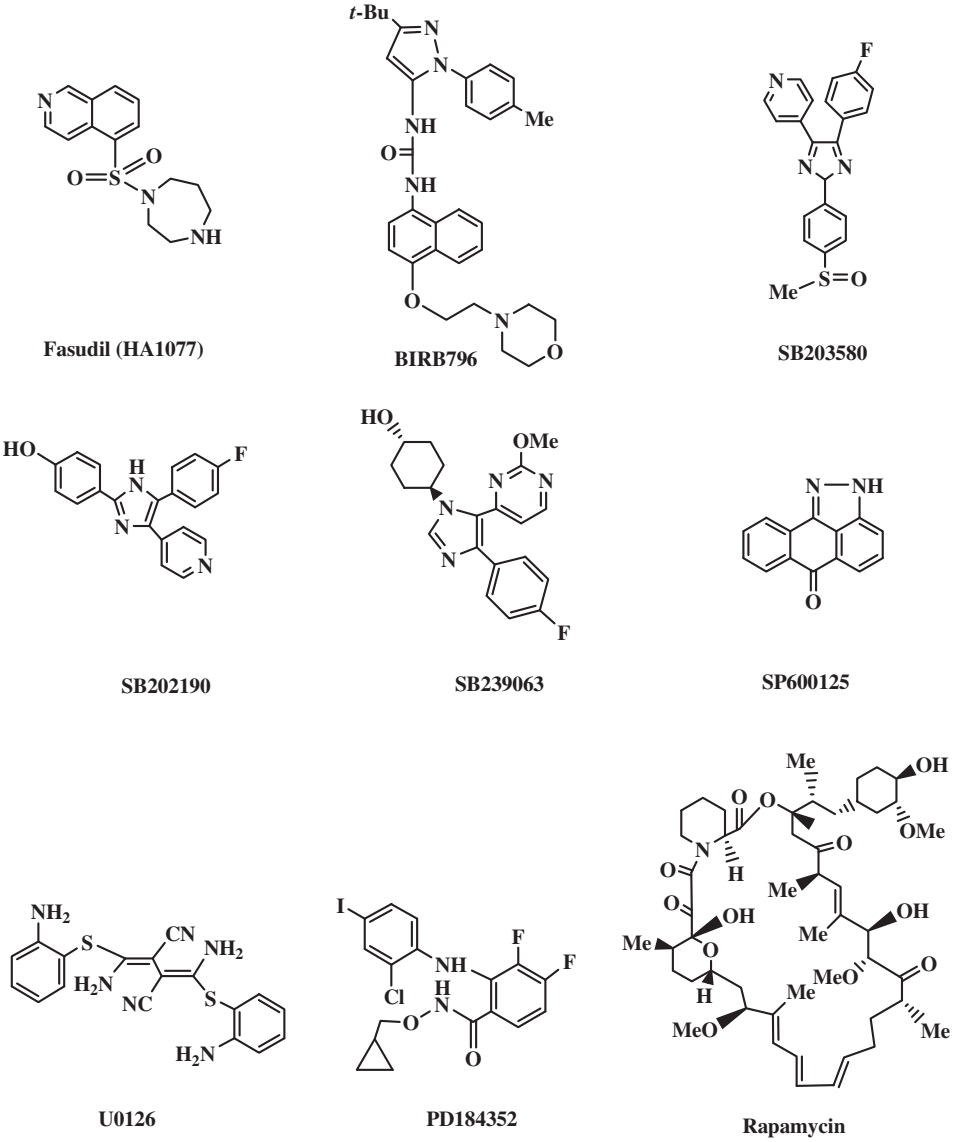


**Figure 15** Examples of MAP kinase inhibitors.

ity of p38 MAPK. These cytokines have been implicated in a number of diseases associated with inflammation, suggesting that p38 selective inhibitors could be used as therapeutic agents for the treatment of a number of inflammatory diseases such as rheumatoid arthritis [59, 167], autoimmune diseases [320–330], cancer, diabetes, and acute myocardial infarctions [331–336]. Two members of the p38 MAPK family, p38α and p38β, are activated by ischemia.

Hence, the inhibition of p38-MAPKs might have therapeutic applications for treating acute coronary syndromes (ACS) [159]. Several companies have developed p38 inhibitors [337–340].

BIRB 796, a diaryl urea, [204, 341] (Fig. 15) is a selective and potent p38α inhibitor currently in phase IIb/III clinical trials for the treatment of rheumatoid arthritis and Crohn's disease. BIRB 796 binding to the Phe residue in the conserved DFG motif of activation loop induces a large conformational change in the enzyme. The enzyme with modified conformation cannot bind to ATP [204]. In p38α this residue is much smaller, allowing access for interaction by BIRB 796, but the entry to this pocket is blocked by bulky gatekeeper amino acid side chains such as $Phe^{80}$ in CDK2 and Asp or Glu residues in most MAPK family members.

As described above, MAPKs regulate the production of cytokines and chemokines and upregulation of adhesion molecules on endothelial cells [159, 342]; hence, preventing the release of inflammatory cytokines and chemokines is a potentially promising strategy for the treatment of ACS. VX-702 (structure not disclosed) is a p38MAP kinase inhibitor in phase II clinical trials for ACS [343]. p38 MAP kinase inhibition by VX-702 leads to degradation of the cytokine messenger ribonucleic acids (mRNAs), including those coding for interleukin (IL)-1/ß, TNF, IL-10, IL-6, interferon (IFN), MIP1/ß, and IL-8 [159].

SB203580 [59, 332, 344], SB202190 [178], and SB239063 [345] (Fig. 15) are potent and specific ATP-competitive inhibitors of p38 MAP Kinase. SB239063 [345], a selective p38 MAP kinase inhibitor, was reported to have beneficial effects in the intact rat model of ischemia-reperfusion (I/R) injury, produced a dramatic reduction in the myocardial inflammatory response by reducing upregulation of P-selectin and intercellular adhesion molecule, and reduced neutrophil accumulation within the ischemic zone [159].

*JNK Inhibitors* Upstream activators of the MAPKs such as JNKs can be targeted to inhibit MAPK signaling [346]. This approach might reduce toxicity MAPKs. JNKs are activated by at least 12 different MAPK kinase kinases (MAPKKKs) and 2 MAPK kinases (MAPKKs), hence inhibiting JNKs can block signaling of all of these MAPKKKs and MAPKKs. JNK was suggested to be an attractive target for the treatment of chronic inflammatory disease, apoptotic cell death, and cancer [347]. SP600125 (Fig. 15), an inhibitor of JNK, was reported to prevent the expression of several anti-inflammatory genes in cell-based assays [347–349] but is a relatively weak and nonselective inhibitor [179]. The development of more selective JNK inhibitors will be needed to determine whether designing of specific and potent inhibitors of JNK is possible.

*ERK Inhibitors* The protein kinases MAPK ERK kinase (MEK)1/2, which is involved in activating the ERK family of MAPKs (Fig. 5), and mTOR (mammalian target of rapamycin) have critical roles in cell cycle progression. ERK

inhibitors, such as U0126, PD184352, and rapamycin/sirolimus (Fig. 15), are in clinical trials for the treatment of a variety of cancers [350]. U0126 and PD184352 are noncompetitive ATP binding site inhibitors that maintain kinases in an inactive state by preventing their phosphorylation by upstream activating kinases such as Raf. Rapamycin also inhibits the protein kinase mTOR [351] and is currently used successfully as an immunosuppressant and an inhibitor of in-stent restenosis [153].

***CDK Inhibitors*** A number of groups have identified CDK inhibitors. The inhibition of CDK2 should arrest cells in G1 phase and prevent them from entering the cell cycle. Therefore, inhibitors of CDK2 may have utility in the treatment of proliferative diseases such as cancer and psoriasis [352–364].

Flavopiridol (L86-8275, HMR1275) (Fig. 6), a synthetic flavonoid analog, is a CDK1, CDK2, and CDK4 inhibitor in clinical trials as a cancer therapy and is the first CDK inhibitor that is not competitive with ATP. Flavopiridol has shown activity in phase I and phase II trials with unknown mechanism [365]. It is not yet clear if the antiproliferative properties of flavopiridol are due to its ability to inhibit transcription through CDK9/cyclin T, its inhibition of cell cycle CDKs or through action on other targets.

Indirubin (Fig. 16) proved to be an inhibitor of CDK2 and CDK5/p25 [366]. Aminothiazole-based compounds are relatively selective CDK2 inhibitors compared with CDK1 and CDK4. The results of phase I clinical trials with BMS 387032 have recently been reported, and phase II trials are planned [367]. Olomoucine (Fig. 16) was one of the first CDK inhibitors to be developed [368]. Two of its derivatives, roscovitine [369] and purvalanol [370] (Fig. 16), are more potent ATP-competitive inhibitors of CDK1, CDK2, and CDK5 and have been used extensively to inhibit these PKs in cell-based assays. Roscovitine and purvalanol inhibit CDK2 activity with $IC_{50}$ values of 0.25 and 0.1 $\mu$M, respectively. Roscovitine inhibited DYRK1A with an $IC_{50}$ of 3.1 $\mu$M but had little effect on the other tested PKs. Purvalanol A is a 2.5-fold more potent inhibitor of CDK2 but is less selective than roscovitine and inhibits DYRK1A, ribosomal S6 kinase 1 (RSK1), ERK2, Lck, and Csk [179]. R-roscovitine (CYC202) is currently undergoing phase II clinical trials for advanced breast cancer and stage IIIB/IV non-small-cell lung cancer in combination with standard chemotherapy regimes [371].

***PKC Inhibitors*** The important role of PKC in cancer renders it a potentially suitable target for anticancer therapy. Several PKC inhibitors, such as UCN-01 and LY-317615 (Fig. 16), are now in clinical trials as anticancer agents [372]. Ruboxistaurin (LY-333531) (Fig. 16) is a selective inhibitor of PKC-β that is in phase III clinical trials for the treatment of macular edema, neuropathy, and diabetic retinopathy [373].

The microbial alkaloid staurosporine (Fig. 16) is an antiproliferative agent and potent inhibitor of PKC. Staurosporine is a competitive inhibitor of PKC's conserved ATP binding site. Staurosporine has poor specificity for PKC

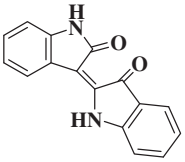

Indirubin

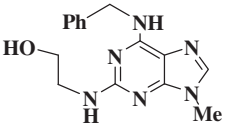

Olomoucine

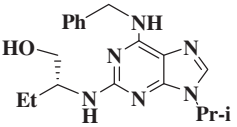

Roscovitine

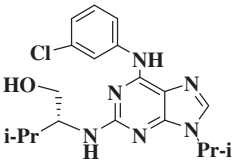

Purvalanol A

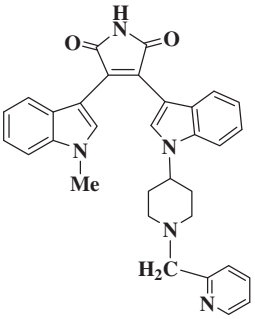

LY-317615

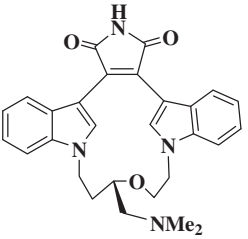

Ruboxistaurin (LY-333531)

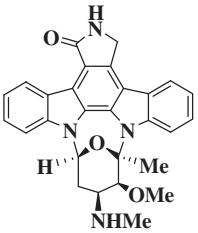

Staurosporine

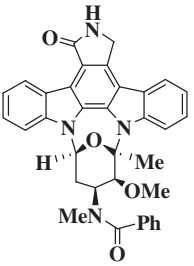

PKC412

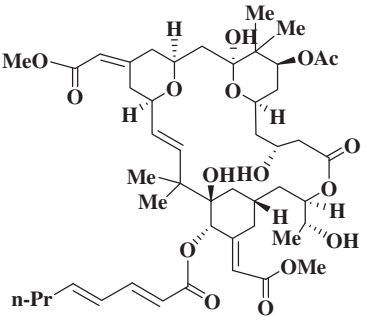

Bryostatin I

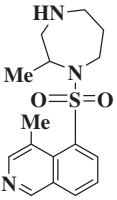

H-1152P

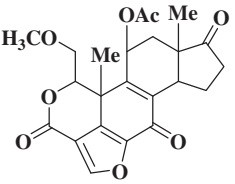

Wortmannin

**Figure 16** Examples of CDK, PKC, PI3K, and Rho kinase inhibitors.

and its isoforms and has inhibitory activity toward STKs and PTKs [374]. Staurosporine derivatives such as PKC412 (*N*-benzoyl-staurosporine) [375] and UCN-01 (7-hydroxystaurosporine) [376, 377] exhibit improved selectivity, with a potentially better therapeutic index in vivo and have been reported to enhance the effects of other cytotoxic agents.

The bryostatins have been shown to have promising antineoplastic and immunomodulatory activity in preclinical models. Bryostatin I (Fig. 16) is the prototype of this class of agents [378]. Bryostatins are potent activators of cPKC and nPKC subfamilies. However, in the presence of activating ligands, such as the tumor-promoting phorbol esters, bryostatins act as antagonists [379] possibly due to differential isoform activation [380] or nuclear translocation [381].

Antisense oligonucleotides have been developed to achieve PKC isoform-specific inhibition by inhibiting expression of target mRNA sequences. ISIS3521, an antisense phosphorothioate oligonucleotide to PKCα, has been investigated in a phase I trial [382].

***Rho-kinase Inhibitors*** Inhibition of Rho-kinase helps blood vessels to relax and increases the blood supply to cardiac tissue. Fasudil (HA1077) (Fig. 15) progressed to human clinical trials in the early 1990s. It was approved in Japan in 1995 for the treatment of cerebral vasospasm. Fasudil had no marked side effects over a 2-week period when given by intravenous injection. At micromolar concentrations, HA1077 inhibits several PKs, such as the Rho-dependent protein kinase ROCK, but it is unclear whether its clinical efficacy results from inhibition of this or other PKs, or whether it is due to a nonkinase effect. ROCK can constrict blood vessels by inhibiting smooth-muscle myosin phosphatase. A recent structure determination [383] of PKA in complex with Fasudil and with a more potent Rho-kinase inhibitor H-1152P (Fig. 16), which differs by only two methyl groups, demonstrated characteristic binding within the ATP site. It is suggested that it is the combination of residues at the ligand binding site, which generates a uniquely shaped inhibitor binding pocket that confers selectivity.

***PI3K Inhibitors*** The structure of Wortmanin (Fig. 16), a selective PI3K inhibitor, in complex with PI3K has been reported [384]. By inhibition of PI3K, Wortmanin has been shown to inhibit osteoclasts-induced bone resorption in vitro [385]. mTOR is a downstream target of the PI3K pathway that plays a role in control of cellular growth, size, and cell proliferation. Its direct inhibitor rapamycin (Fig. 15) (386) was the first agent to be identified used as an immunosuppressive agent. A related compound CCI779 is in clinical phase II for renal and glioblastoma cancer in which the level of phosphatidylinositol-3,4,5-triphosphate is elevated [387].

***GSK3 Inhibitors*** A number of small-molecule inhibitors of GSK3 for the treatment of type 2 diabetes have been proposed, including indirubins,

paullones, hymenialdisine, aminopyrimidines, and aminopyridines such as CT 99021, CT 20026, and maleimides. The discussion of these inhibitors is beyond the scope of this chapter, but they have been recently reviewed [148].

## 10 PROSPECTS AND FUTURE DIRECTIONS

Approval of several PKIs by FDA has led to the conclusion that, in the future, several diseases associated with dysregulated PKs can be treated. Herein we discussed some PKIs approved or in advanced clinical trials. PKs continue to be considered as molecular targets for cancer and other diseases, including inflammation, osteoporosis, atherosclerosis, restenosis, and stroke. PKs not only play roles in the development of diseases but also function in pathways that regulate the most basic of normal cellular processes; hence, toxicity remains a major concern. In spite of all this progress, specificity remains an area of concern for designing PKIs due to PKs structural similarities. Additionally, it is not possible to test the drugs against all kinases, and side effects are expected even after approval of drugs by FDA. Despite these obstacles, PKIs have created a lot of promises for a wide variety of diseases.

***Acknowledgment*** We acknowledge the financial support from BRIN/INBRE Program of the National Center for Research Resources, NIH, Grant number 1 P20 RR16457.

## REFERENCES

1. Hunter, T. (1987). A thousand and one protein kinases. *Cell*, *50*, 823–829.
2. Hunter, T. (2000). Signaling—2000 and beyond. *Cell*, *100*, 113–127.
3. Manning, G., Whyte, D. B., Martinez, R., Hunter, R., Sudarsanam, S. (2002). The protein kinase complement of the human genome. *Science*, *298*, 1912.
4. Hubbard, S. R., Till, J. H. (2000). Protein tyrosine kinase structure and function. *Annu. Rev. Biochem.*, *69*, 373–398.
5. Ho, M., Bramson, H. N., Hansen, D. E., Knowles, J. R. Kaiser, E. T. (1988). Stereochemical course of the phospho group transfer catalyzed by cAMP-dependent protein-kinase. *J. Am. Chem. Soc.*, *110*, 2680–2681.
6. Aqvist, J., Kolmodin, K., Florian, J., Warshel, A. (1999). Mechanistic alternatives in phosphate monoester hydrolysis: What conclusions can be drawn from available experimental data? *Chem. Biol.*, *6*, R71–R80.
7. Admiraal, S. J., Herschlag, D. (2000). The substrate-assisted general base catalysis model for phosphate monoester hydrolysis: Evaluation using reactivity comparisons. *J. Am. Chem. Soc.*, *122*, 2145–2148.
8. Mildvan, A. S. (1997). Mechanisms of signaling and related enzymes. *Proteins*, *29*, 401–416.
9. Uckun, F. M., Mao, C. (2004). Tyrosine kinases as new molecular targets in treatment of inflammatory disorders and leukemia. *Curr. Pharm. Des.*, *10*, 1083–1091.

10. Williams, D. H., Mitchell, T. (2002). Latest developments in crystallography and structure-based design of protein kinase inhibitors as drug candidates. *Curr. Opin. Pharmacol.*, *2*, 1–7.
11. Scapin, G. (2002). Structural biology in drug design: Selective protein kinase inhibitors. *Drug Discov. Today*, *7*, 601–611.
12. Hubbard, S. (2002). Protein tyrosine kinases: Autoregulation and small molecule inhibition. *Curr. Opin. Struct. Biol.*, *12*, 735–741.
13. Al-Obeidi, F. A., Wu, J. J., Lam, K. S. (1998). Protein tyrosine kinases: Structure, substrate specificity, and drug discovery. *Biopolymers*, *47*, 197–223.
14. Hubbard, S. R. (1999). Structural analysis of receptor tyrosine kinases. *Prog. Biophys. Mol. Biol.*, *71*, 343–358.
15. Munshi, S., Kornienko, M., Hall, D. L., Reid, J. C., Waxman, L., Stirdivant, S. M., Darke, P. L., Kuo, L. C. (2002). Crystal structure of the Apo, unactivated insulin-like growth factor-1 receptor kinase. Implication for inhibitor specificity. *J. Biol. Chem.*, *277*, 38797–38802.
16. Xu, W., Harrison, S. C., Eck, M. J. (1997). Three-dimensional structure of the tyrosine kinase c-Src. *Nature*, *385*, 595–602.
17. Xu, W., Doshi, A., Lei, M., Eck, M. J., Harrison, S. C. (1999). Crystal structures of c-Src reveal features of its autoinhibitory mechanism. *Mol. Cell*, *3*, 629–638.
18. Mohammadi, M., Schlessinger, J., Hubbard, S. R. (1996). Structure of the FGF receptor tyrosine kinase domain reveals a novel autoinhibitory mechanism. *Cell*, *86*, 577–587.
19. Mohammadi, M., McMahon, G., Sun, L., Tang, C., Hirth, P., Yeh, B. K., Hubbard, S. R., Schlessinger, J. (1997). Structures of the tyrosine kinase domain of fibroblast growth factor receptor in complex with inhibitors. *Science*, *276*, 955–960.
20. McTigue, M. A., Wickersham, J. A, Pinko, C., Showalter, R. E., Parast, C. V., Tempczyk-Russell, A., Gehring, M. R., Mroczkowski, B., Kan, C. C., Villafranca, J. E., Appelt, K. (1999). Crystal structure of the kinase domain of human vascular endothelial growth factor receptor 2: A key enzyme in angiogenesis. *Structure with Folding & Design*, *7*, 319–330.
21. Zhu, X., Kim, J. L., Newcomb, J. R., Rose, P. E., Stover, D. R., Toledo, L. M., Zhao, H., Morgenstem, K. A. (1999). Structural analysis of the lymphocyte-specific kinase Lck in complex with non-selective and Src family selective kinase inhibitors. *Structure with Folding & Design*, *7*, 651–661.
22. Yamaguchi, H., Hendrickson, W. A. (1996). Structural basis for activation of human lymphocyte kinase Lck upon tyrosine phosphorylation. *Nature*, *384*, 484–489.
23. Sicheri, F., Moarefi, I., Kuriyan, J. (1997). Crystal structure of the Src family tyrosine kinase Hck. *Nature*, *385*, 602–609.
24. Schindler, T., Sicheri, F., Pico, A., Gazit, A., Levitzki, A., Kuriyan, J. (1999). Crystal structure of Hck in complex with a Src family-selective tyrosine kinase inhibitor. *Mol. Cell*, *3*, 639–648.
25. Mao, C., Zhou, M., Uckun, F. M. (2001). Crystal structure of Bruton's tyrosine kinase domain suggests a novel pathway for activation and provides insights into the molecular basis of X-linked agammaglobulinemia. *J. Biol. Chem.*, *276*, 41435–41443.

26. Schindler, T., Bornmann, W., Pellicena, P., Miller, W. T., Clarkson, B., Kuriyan, J. (2000). Structural mechanism for STI-571 inhibition of abelson tyrosine kinase. *Science*, *289*, 1938–1942.
27. Taylor, S. S., Radzio-Andzelm, E. (1994). Three protein kinase structures define a common motif. *Structure*, *2*, 345–355.
28. Stout, T. J., Foster, P. G., Matthews, D. J. (2004). High-throughput structural biology in drug discovery: Protein kinases. *Curr. Pharm. Des.*, *10*, 1069–1082.
29. Sawyer, T. K. (1998). Src homology-2 domains: Structure, mechanisms, and drug discovery. *Biopolymers*, *47*, 243–261.
30. Sicheri, F., Kuriyan, J. (1997). Structures of Src-family tyrosine kinases. *Curr. Opin. Struct. Biol.*, *7*, 777–785.
31. Songyang, Z., Shoelson, S. E., Chaudhuri, M., Gish, G., Pawson, T., Haser, W. G., King, F., Roberts, T., Ratnofsky, S., Lechleider, R. J., Neel, B. G., Birge, R. B., Fajardo, J. E., Chou, M. M., Hanatusa, H., Schaffhausen, B., Cantley, L. C. (1993). SH2 domains recognize specific phosphopeptide sequences. *Cell*, *72*, 767–778.
32. Sadowski, I., Stone, J. C., Pawson, T. (1986). A noncatalytic domain conserved among cytoplasmic protein-tyrosine kinases modifies the kinase function and transforming activity of Fujinami sarcoma virus P130gag-fps. *Mol. Cell. Biol.*, *6*, 4396–4408.
33. Cantley, L. C., Auger, K. R., Carpenter, C., Duckworth, B., Graziana, A., Kapeller, R., Soltoff, S. (1991). Oncogenes and signal transduction. *Cell*, *64*, 281–302.
34. Fukui, Y., Hanafusa, H. (1991). Requirement of phosphatidylinositol-3 kinase modification for its association with p60src. *Mol. Cell. Biol.*, *11*, 1972–1979.
35. Schaller, M. D., Hilderbrand, J. D., Shannon J. D., Fox, J. W., Vines, R. R., Parsons, J. T. (1994). Autophosphorylation of the focal adhesion kinase, pp125FAK, directs SH2 dependent binding of pp60src. *Mol. Cell. Biol.*, *14*, 1680–1688.
36. Taylor, S. J., Shalloway, D. (1994). An RNA-binding protein associated with Src through its SH2 and SH3 domains in mitosis. *Nature*, *368*, 867–871.
37. Petch, L. A., Bockholt, S. M., Bouton, A., Parsons, J. T., Burridge, K. (1995). Adhesion-induced tyrosine phosphorylation of the p130 src substrate. *J. Cell. Sci.*, *108*, 1371–1379.
38. Johnson, L. N., Noble, M. E., Owen, D. J. (1996). Active and inactive protein kinases: Structural basis for regulation. *Cell*, *85*, 149–158.
39. Adams, J. A. (2003). Activation loop phosphorylation and catalysis in protein kinases: Is there functional evidence for the autoinhibitor model? *Biochemistry*, *42*, 601–607.
40. Toker, A. N., Newton, A. C. (2000). Cellular signaling: Pivoting around PDK1. *Cell*, *103*, 185–188.
41. Till, J. H., Ablooglu, A. J., Frankel, M., Bishop, S. M., Kohanski, R. A., Hubbard, S. R. (2001). Crystallographic and solution studies of an activation loop mutant of the insulin receptor tyrosine kinase: Insights into kinase mechanism. *J. Biol. Chem.*, *276*, 10049–10055.
42. Ablooglu, A. J., Frankel, M., Rusinova, E., Ross, J. B. A., Kohanski, R. A. (2001). Multiple activation loop conformations and their regulatory properties in the insulin receptor's kinase domain. *J. Biol. Chem.*, *276*, 46933–46940.

43. Orr, J. W., Newton, A. C. (1994). Requirement for negative charge on "activation loop" of protein kinase C. *J. Biol. Chem.*, *269*, 27715–27718.
44. Wybenga-Groot, L. E., Baskin, B., Ong, S. H., Tong, J., Pawson, T., Sicheri, F. (2001). Structural basis for autoinhibition of the Ephb2 receptor tyrosine kinase by the unphosphorylated juxtamembrane region. *Cell*, *106*, 745–757.
45. Moarefi, H., LaFevre-Bernt, M., Sicheri, F., Huse, M., Lee, C.-H., Kuriyan, J., Miller, W. T. (1997). Activation of the Src-family tyrosine kinase Hck by SH3 domain displacement. *Nature*, *385*, 650–653.
46. Mandine, E., Jean-Baptiste, V., Vayssiere, B., Gofflo, D., Benard, D., Sarubbi, E., Deprez, P., Baron, R., Superti-Furga, G., Lesuisse, D. (2002). High-affinity Src-SH2 ligands which do not activate Tyr(527)-phosphorylated Src in an experimental in vivo system. *Biochem. Biophys. Res. Commun.*, *298*, 185–192.
47. Liu, X., Brodeur, S. R., Gish, G., Songyang, Z., Cantley, L. C., Laudano, A. P., Pawson, T. (1993). Regulation of c-Src tyrosine kinase activity by the Src SH2 domain. *Oncogene*, *8*, 1119–1126.
48. Alonso, G., Koegl, M., Mazurenko, N., Courtneidge, S. A. (1995). Sequence requirements for binding of Src family tyrosine kinases to activated growth factors. *J. Biol. Chem.*, *270*, 9840–9848.
49. Boerner, R. J., Kassel, D. B., Barker, S. C., Ellis, B., DeLacy, P., Knight, W. B. (1996). Correlation of the phosphorylation states of $pp60^{c\text{-}src}$ with tyrosine kinase activity: The intramolecular pY530-SH2 complex retains significant activity if Y419 is phosphorylated. *Biochemistry*, *35*, 9519–9525.
50. Erpel, T., Courtneidge, S. A. (1995). Src family protein tyrosine kinases and cellular signal transduction pathways. *Curr. Opin. Cell Biol.*, *7*, 176–182.
51. Thomas, J. W., Ellis, B., Boerner, R. J., Knight, W. B., Wite, G. C., II, Shaller, M. D. (1998). SH2- and SH3-mediated interactions between focal adhesion kinase and Src. *J. Biol. Chem.*, *273*, 577–583.
52. Alexandropoulos, K., Baltimore, D. (1996). Coordinate activation of c-Src by SH3- and SH2 binding sites on a novel, p130CAS-related protein, Sin. *Genes Devel.*, *10*, 1341–1355.
53. Profit, A. A., Lee, T. R., Niu, J., Lawrence, D. S. (2001). Molecular rulers: An assessment of distance and spatial relationships of Src tyrosine kinase SH2 and active site regions. *J. Biol. Chem.*, *276*, 9446–9451.
54. Biscardi, J. S., Tice, D. A., Parsons, S. J. (1999). c-Src, receptor tyrosine kinases, and human cancer. *Adv. Cancer Res.*, *76*, 61–119.
55. Jiang, G., Zhang, B. B. (2002). PI3-kinase and its up- and down-stream modulators as potential targets for the treatment of type II diabetes. *Front. Biosci.*, *7*, 903–907.
56. Frame, S., Cohen, P. (2001). GSK3 takes centre stage more than 20 years after its discovery. *Biochem. J.*, *359*(Pt. 1), 1–16.
57. Vlahos, C. J., McDowell, S. A., Clerk, A. (2003). Kinases as therapeutic targets for heart failure. *Nat. Rev. Drug Discov.*, *2*, 99–113.
58. Mucke, H. A. (2003). CEP-1347 (Cephalon). *IDrugs*, *6*, 377–383.
59. Lee, J. C., Kumar, S., Griswold, D. E., Underwood, D. C., Votta, B. J., Adams, J. L. (2000). Inhibition of p38 MAP kinase as a therapeutic strategy. *Immunopharmacology*, *47*, 185–201.

60. Traxler, P. (2003). Tyrosine kinases as targets in cancer therapy—Successes and failures. *Expert Opin. Ther. Patents*, *7*, 215–234.
61. Gray, N., Detivaud, L., Doerig, C., Meijer, L. (1999). ATP-site directed inhibitors of cyclin-dependent kinases. *Curr. Med. Chem.*, *6*, 859–875.
62. Collett, M. S., Erikson, R. L. (1978). Protein kinase activity associated with the avian sarcoma virus src gene product. *Proc. Natl. Acad. Sci. USA*, *75*, 2021–2024.
63. Klint, P., Claesson-Welsh, L. (1999). Signal transduction by fibroblast growth factor receptors. *Front. Biosci.*, *4*, D165–D177.
64. Dominguez, M., Wasserman, J. D., Freeman, M. (1998). Multiple functions of the EGF receptor in *Drosophila* eye development. *Curr. Biol.*, *8*, 1039–1048.
65. Sieg, D. J., Hauck, C. R., Ilic, D., Klingbeil, C. K., Schaefer, E., Damsky, C. H., Schlaepfer, D. D. (2000). FAK integrates growth-factor and integrin signals to promote cell migration. *Nat. Cell Biol.*, *2*, 249–256.
66. Casaccia-Bonnefil, P., Gu, C., Chao, M. V. (1999). Neurotrophins in cell survival/death decisions. The functional roles of glial cells in health and disease. *Adv. Exper. Med. Biol.*, *468*, 275–282.
67. Van der Geer, P., Hunter, T., Lindberg, R. A. (1994). Receptor protein-tyrosine kinases and their signal transduction pathways. *Annu. Rev. Cell. Biol.*, *10*, 251–337.
68. Carpenter, G., Cohen, S. (1979). Epidermal growth factor. *Annu. Rev. Biochem.*, *48*, 193–216.
69. Carpenter, G. (1987). Receptors for epidermal growth factor and other polypeptide mitogens. *Annu. Rev. Biochem.*, *56*, 881–914.
70. Quirion, R., Araujo, D., Nair, N. P., Chabot, J. G. (1988). Visualization of growth factor receptor sites in rat forebrain. *Synapse*, *2*, 212–218.
71. Klapper, L. N., Kirschbaum, M. H., Sela, M., Yarden, Y. (2000). Biochemical and clinical implications of the ErbB/HER signaling network of growth factor receptors. *Adv. Cancer Res.*, *77*, 25–79.
72. Olayioye, M. A., Neve, R. M., Lane, H. A., Hynes, N. E. (2000). The ErbB signaling network: Receptor heterodimerization in development and cancer. *EMBO J.*, *19*, 3159–3167.
73. Sporn, M. B., Roberts, A. B. (1985). Autocrine growth factors and cancer, *Nature*, *313*, 745–747.
74. Salomon, D. S., Brandt, R., Ciardiello, F., Normanno, N. (1995). Epidermal growth factor-related peptides and their receptors in human malignancies. *Crit. Rev. Oncol. Hematol.*, *19*, 183–232.
75. Strawn, L. M., McMahon, G., App, H., Schreck, R., Kuchler, W. R., Longhi, M. P., Hui, T. H., Tang, C., Levitzki, A., Gazit, A., Chen, I., Keri, G., Orfi, L., Risau, W., Flamme, I., Ullrich, A., Hirth, K. P., Shawver, L. K. (1996). Flk-1 as a target for tumor growth inhibition. *Cancer Res.*, *56*, 3540–3545.
76. Verheul, H. M., Jorna, A. S., Hoekman, K., Broxterman, H. J., Gebbink, M. F., Pinedo, H. M. (2000). Vascular endothelial growth factor-stimulated endothelial cells promote adhesion and activation of platelets. *Blood*, *96*, 4216–4221.
77. Matsuzaki, K., Yoshitake, Y., Matuo, Y., Sasaki, H., Nishikawa, K. (1989). Monocolonal antibodies against heparin-binding growth factor II/basic fibroblast growth factor that block its biological activity: Invalidity of the antibodies for tumor angiogenesis. *Proc. Natl. Acad. Sci. USA*, *86*, 9911–9915.

78. Dennis, P. A., Rifkin, D. B. (1990). Studies on the role of basic fibroblast growth factor in vivo: Inability of neutralizing antibodies to block tumor growth. *J. Cell Physiol.*, *144*, 84–98.
79. Ferrara, N., Hillan, K. J., Gerber, H.-P., Novotony, W. (2004). Discover and development of Bevacizumab, an anti-VEGF antibody for treating cancer. *Nat. Rev. Drug Dis.*, *3*, 391–400.
80. Abu-Duhier, F. M., Goodeve, A. C., Wilson, G. A., Care, R. S., Peake, I. R., Reilly, J. T. (2001). Identification of novel FLT-3 Asp835 mutations in adult acute myeloid leukaemia. *Br. J. Haematol.*, *113*, 983–988.
81. Laird, A. D., Christensen, J. G., Li, G., Carver, J., Smith, K., Xin, X., Moss, K. G., Louie, S. G., Mendel, D. B., Cherrington, J. M. (2002). SU6668 inhibits Flk-1/KDR and PDGFRbeta in vivo, resulting in rapid apoptosis of tumor vasculature and tumor regression in mice. *FASEB J.*, *16*, 681–690.
82. Mauro, M. J., Druker, B. J. (2001). STI571: Targeting BCR-ABL as therapy for CML. *Oncologists*, *6*, 233–238.
83. Sawyers, C. L. (1999). Chronic myeloid leukemia. *N. Engl. J. Med.*, *340*, 1330–1340.
84. Warmuth, M., Damoiseaux, R., Liu, Y., Fabbro, D., Gray, N. (2003). Src family kinases: Potential targets for the treatment of human cancer and leukemia. *Curr. Pharm. Des.*, *9*, 2043–2059.
85. Susa, M., Teti, A. (2000). Tyrosine kinase Src inhibitors: Potential therapeutic application. *Drugs News Perspect.*, *13*, 169–175.
86. Weber, T. K., Steele, G., Summerhayes, I. C. (1992). Differential pp60c-src activity in well and poorly differentiated human colon carcinomas and cell lines. *J. Clin. Invest.*, *90*, 815–821.
87. Ottenholf-Kalff, A. E., Rijksen, G., van Beurden, E. A., Hennipman, A., Michels, A. A., Staal, G. E. (1992). Characterization of protein tyrosine kinase from human breast cancer. *Cancer Res.*, *49*, 4773–4778.
88. Avizienyte, E., Wyke, A. W., Jones, R. J., McLean, G. W., Westhoff, M. A., Brunton, V. G., Frame, M. C. (2002). Src-induced de-regulation of E-cadherin in colon cancer cells requires integrin signaling. *Nat. Cell Biol.*, *4*, 632–638.
89. Thomas, S. M., Brugge, J. S. (1997). Cellular functions regulated by Src family kinases. *Annu. Rev. Cell Devel. Biol.*, *13*, 513–609.
90. Shakespeare, W. C., Metcalfe, C. A. 3rd, Wang, Y., Sundaramoorthi, R., Keenan, T., Weigele, M., Bohacek, R. S., Delgarne, D. C., Sawyer, T. K. (2003). Novel bone-targeted Src tyrosine kinase inhibitor drug discovery. *Curr. Clin. Drug. Discov. Devel.*, *6*, 729–741.
91. Lowe, C., Yoneda, T., Boyce, B. F., Chen, H., Mundy, G. R., Soriano, P. (1993). Osteopetrosis in Src-deficient mice is due to an autonomous defect of osteoclasts. *Proc. Natt. Acad. Sci. USA*, *90*, 4485–4489.
92. Boyce, B. F., Yoneda, T., Lowe, C., Soriano, P., Mundy, G. R. (1992). Requirement of pp60c-Src expression for osteoclasts to form ruffled borders and resorb bone in mice. *J. Clin. Invest.*, *90*, 1622–1627.
93. Soriano, P., Montgomery, C., Geske, R., Bradley, A. (1991). Targeted disruption of the c-src proto-oncogene leads to osteopetrosis in mice. *Cell*, *64*, 693–702.
94. Marzia, M., Sims, N. A., Voit, S., Migliaccio, S., Taranta, A., Bernardini, S., Faraggiana, T., Yoneda, T., Mundy, G. R., Boyce, B. F., Baron R., Teti. A. (2000). *J. Cell Biol.*, *151*, 311–320.

95. Kousteni, S., Bellido, T., Plotkin, L. I., O'Brien, C. A., Bodenner, D. L., Han, L., Han, K., DiGregorio, G. B., Katzenellenbogen, J. A., Katzenellenbogen, B. S., Roberson, P. K., Weinstein, R. S., Jilka, R. L., Manolagas, S. C. (2001). Nongenotropic, sex-nonspecific signaling through the estrogen or androgen receptors: Dissociation from transcriptional activity. *Cell*, *104*, 719–730.

96. Abu-Amer, Y., Ross, F. P., Schlesinger, P., Tondravi, M. M., Teitelbaum, S. L. (1997). Substrate recognition by osteoclast precursors induces C-src/microtubule association. *J. Cell Biol.*, *137*, 247–258.

97. Wong, B. R., Besser, D., Kim, N., Arron, J. R., Vologodskaia, M., Hanafusa, H., Choi, Y. (1999). TRANCE, a TNF family member, activates Akt/PKB through a signaling complex involving TRAF6 and c-Src. *Mol. Cell*, *4*, 1041–1049.

98. Tanaka, S., Amling, M., Neff, L., Peyman, A., Uhlmann, E., Levy J. B., Baron, R. (1996). c-Cbl is downstream of c-Src in a signalling pathway necessary for bone resorption. *Nature*, *383*, 528–531.

99. Jeschke, M., Brandi, M.-L., Susa, M. (1998). Expression of Src family kinases and their putative substrates in the human preosteoclastic cell line FLG 29.1. *J. Bone Miner. Res.*, *13*, 1880–1889.

100. Lowell, C. A., Berton, G. (1999). Integrin signal transduction in myeloid leukocytes. *J. Leukoc. Biol.*, *65*, 313–320.

101. Missbach, M., Jeschke, M., Feyen, J., Muller, K., Glatt, M., Green, J., Susa, M. (1999). A novel inhibitor of the tyrosine kinase Src suppresses phosphorylation of its major cellular substrates and reduces bone resorption in vitro and in rodent models in vivo. *Bone*, *24*, 437–449.

102. Shakespeare, W., Yang, M., Bohacek, R., Cerasoli, F., Stebbins, K., Sundaramoorthi, R., Azimioara, M., Vu, C., Pradeepan, S., Metcalf, C., Haraldson, C., Merry, T., Dalgarno, D., Narula, S., Hatada, M., Lu, X., van Schravendijk, M. R., Adams, S., Violette, S., Smith, J., Guan, W., Bartlett, C., Herson, J., Iuliucci, J., Weigele, M., Sawyer, T. (2000). Structure-based design of an osteoclast-selective, nonpeptide src homology 2 inhibitor with in vivo antiresorptive activity. *Proc. Natl. Acad. Sci. USA*, *97*, 9373–9378.

103. Luttrell, D. K., Lee, A., Lansing, T. J., Crosby, R. M., Jung, K. D., Willard, D., Luther, M., Rodriguez, M., Berman, J., Gilmer, T. M. (1994). Involvement of pp60c-src with two major signaling pathways in human breast cancer. *Proc. Natl. Acad. Sci. USA*, *91*, 83–87.

104. Frame, M. (2002). C-Src in cancer: Deregulation and consequences for cell behavior. *Biochim. Biophys. Acta*, *1602*, 114–130.

105. Kyriakis, J. M., Avruch, J. (2001). Mammalian mitogen-activated protein kinase signal transduction pathways activated by stress and inflammation. *Physiol. Rev.*, *81*, 807–869.

106. Pombo, C. P., Bonventre, J. V., Avruch, J., Woodgett, J. R., Kyriakis, J. M., Force, T. (1994). The stress-activated protein kinases (SAPKs) are major c-Jun amino-terminal kinases activated by ischemia and reperfusion. *J. Biol. Chem.*, *269*, 26546–26551.

107. Pearson, G., Robinson, F., Beers, G. T., Xu, B. E., Karandikar, M., Berman, K., Cobb, M. H. (2001). Mitogen-activated protein (MAP) kinase pathways: Regulation and physiological functions. *Endocr. Rev.*, *22*, 153–183.

108. Cobb, M. H., Goldsmith, E. J. (1995). How MAP kinases are regulated. *J. Biol. Chem.*, *270*, 14843–14864.

109. Minden, A., Karin, M. (1997). Regulation and function of the JNK subgroup of MAP kinases. *Biochem. Biophys. Acta*, *1333*, F85–F104.

110. Hunter, T. (1995). When is a lipid kinase not a lipid kinase? When it is a protein kinase. *Cell*, *80*, 225–236.

111. Chen, Z., Gibson, T. B., Robinson, F., Silvestro, L., Pearson, G., Xu, B., Wright, A., Vanderbilt, C., Cobb, M. H. (2001). MAP kinases. *Chem. Rev.*, *101*, 2449–2476.

112. Marshall, C. J. (1995). Specificity of receptor tyrosine kinase signaling: Transient versus sustained extracellular signal-regulated kinase activation. *Cell*, *80*, 179–185.

113. Marshall, C. J. (1994). MAP kinase kinase kinase, MAP kinase kinase and MAP kinase. *Curr. Opin. Genet. Devel.*, *4*, 82–89.

114. Sugden, P. H., Clerk, A. (1997). Regulation of the ERK subgroup of MAP kinase cascades through G protein-coupled receptors. *Cell Signal*, *9*, 337–351.

115. Cobb, M. H., Boulton, T.G., Robbins, D. J. (1991). Extracellular signal-regulated kinases: ERKs in progress. *Cell Regul.*, *2*, 965–978.

116. Ono, K., Han, J. (2000). The p38 signal transduction pathway: Activation and function. *Cell Signal*, *12*, 1–13.

117. Mazzucchelli, C., Brambilla, R. (2000). Ras-related and MAPK signalling in neuronal plasticity and memory formation. *Cell Mol. Life Sci.*, *57*, 604–611.

118. Harper, J. W., Adams, P. D. (2001). Cyclin-dependent kinases. *Chem. Rev.*, *101*, 2511–2526.

119. Bramson, H. N., Corona, J., Davis, S. T., Dickerson, S. H., Edelstein, M., Frye, S. V., Gampe, R. T. Jr., Harris, P. A., Hassell, A., Holmes, W. D., Hunter, R. N., Lackey, K. E., Lovejoy, B., Luzzio, M. J., Montana, V., Rocque, W. J., Rusnak, D., Shewchuk, L., Veal, J. M., Walker, D. H., Kuyper, L. F. (2002). Oxindole-based inhibitors of cyclin-dependent kinase 2 (CDK2): Design, synthesis, enzymatic activities, and X-ray crystallographic analysis. *J. Med. Chem.*, *44*, 4339–4358.

120. Sherr, C. J. (1996). Cancer cell cycles. *Science*, *274*, 1672–1677.

121. Frizelle, S. P., Grim, J., Zhou, J., Gupta, P., Curiel, D. T., Geradts, J., Kratzke, R. A. (1998). Re-expression of p16INK4a in mesothelioma cells results in cell cycle arrest, cell death, tumor suppression and tumor regression. *Oncogene*, *16*, 3087–3095.

122. Barnes, D. M., Gillet, C. E. (1998). Cyclin D1 in breast cancer. *Breast Cancer Res. Treat.*, *52*, 1–15.

123. Roussel. M. F. (1999). The INK4 family of cell cycle inhibitors in cancer. *Oncogene*, *18*, 5311–5317.

124. Nevins, J. R. (2001). The Rb/E2F pathway and cancer. *Hum. Mol. Genet.*, *10*, 699–703.

125. Tsao, H., Benoit, E., Sober, A. J., Thiele, C., Haluska, F.G. (1998). Novel mutations in the p16/CDKN2A binding region of the cyclin-dependent kinase-4 gene. *Cancer Res.*, *58*, 109–113.

126. Goekjian, P. G., Jirousek, M. R. (1999). Protein kinase C in the treatment of disease: Signal transduction pathways, inhibitors, and agents in development. *Curr. Med. Chem.*, *6*, 877–903.

127. Dekker, L. V., Parker, P. (1994). Protein kinase C—A question of specificity. *Trends Biochem. Sci.*, *19*, 73–77.
128. Tanaka, C., Saito, N. (1992). Localization of subspecies of protein kinase C in the mammalian central nervous system. *Neurochem. Int.*, *21*, 499–512.
129. Blobe, G. C., Obeid, L. M., Hannun, Y. A. (1994). Regulation of protein kinase C and role in cancer biology. *Cancer Metastasis Rev.*, *13*, 724–730.
130. Ron, D., Kazanietz, M. G. (1999). New insights into the regulation of protein kinase C and novel phorbol esters. *FASEB J.*, *27*, 1658–1676.
131. Bullock, W. H., Magnuson, S. R., Choi, S., Gunn, D. E., Rudolph. J. (2002). Prospects for kinase activity modulators in the treatment of diabetes and diabetic complications. *Curr. Top. Med. Chem.*, *2*, 915–938.
132. Way, K. J., Katai, N., King, G. L. (2001). Protein kinase C and the development of diabetic vascular complications. *Diabetes Med.*, *18*, 945–959.
133. Takenaga, K., Takahashi, K. (1996). Effects of 12-*O*-tetradecanoylphorbol-13-acetate on adhesiveness and lung-colonising ability of Lewis lung carcinoma cells. *Cancer Res.*, *46*, 375–380.
134. Schwartz, G. K., Juang, K., Kelsen, D., Albino, A. P. (1993). Protein kinase C: A novel target for inhibiting gastric cancer cell invasion. *J. Natl. Cancer Inst.*, *85*, 402–407.
135. O'Brian, C. A., Vogel, V. G., Singletary, S. E., Ward, N. E. (1989). Elevated protein kinase C expression in human breast tumour biopsies relative to normal breast tissue. *Cancer Res.*, *49*, 3215–3217.
136. Gokmen-Polar, Y., Murray, N., Velasco, M. A., Gatlica, Z., Fields, A. P. (2001). Elevated protein kinase C βII is an early promotive event in colon carcinogenesis. *Cancer Res.*, *61*, 1375–1381.
137. Robinson-White, A., Stratakis, C. (2002). Protein kinase A signaling: "Cross-talk" with other pathways in endocrine cells. *Ann. NY Acad. Sci.*, *968*, 256–270.
138. Rajnicek, A., Britland, S., McCaig, C. (1997). Contact guidance of CNS neurites on grooved quartz: Influence of groove dimensions, neuronal age and cell type. *J. Cell. Sci.*, *110*, 2905–2913.
139. Rajnicek, A., McCaig, C. (1997). Guidance of CNS growth cones by substratum grooves and ridges: Effects of inhibitors of the cytoskeleton, calcium channels and signal transduction pathways. *J. Cell. Sci.*, *110*, 2915–2924.
140. Hu, E., Lee, D. (2003). Rho kinase inhibitors as potential therapeutic agents for cardiovascular diseases. *Curr. Opin. Investig. Drugs*, *4*, 1065–1075.
141. Riento, K., Ridley A. J. (2003). Rocks: Multifunctional kinases in cell behaviour. *Nat. Rev. Mol. Cell Biol.*, *4*, 446–456.
142. Katso, R., Okkenhaug, K., Ahmadi, K., White, S., Timms, J., Waterfield, M. D. (2001). Cellular function of phosphoinositide 3-kinases: Implications for development, homeostasis, and cancer. *Annu. Rev. Cell Devel. Biol.*, *17*, 615–675.
143. McKenna, W. G., Muschel, R. J. (2003). Targeting tumor cells by enhancing radiation sensitivity. *Genes Chromosomes Cancer*, *38*, 330–338.
144. Woodgett, J. R. (1990). Molecular cloning and expression of glycogen synthase kinase-3/factor A. *EMBO J.*, *9*, 2431–2438.
145. Kim, H.-S., Skurk, C., Thomas, S. R., Bialik, A., Suhara, T., Kureishi, Y., Birnbaum, M., Keaney, J. F. Jr., Walsh. K. (2002). Regulation of angiogenesis by glycogen synthase kinase-3beta. *J. Biol. Chem.*, *277*, 41888–41896.

146. Kaidanovich, O., Eldar-Finkelman, H. (2002). The role of glycogen synthase kinase-3 in insulin resistance and Type 2 diabetes. *Exp. Opin. Ther. Targets*, *6*, 555–561.

147. Eldar-Finkelman, H. (2002). Glycogen synthase kinase 3: An emerging therapeutic target. *Trends Mol. Med.*, *8*, 126–132.

148. Wagman, A. S., Johnson, K. W., Bussiere, D. E. (2004). Discovery and development of GSK3 inhibitors for the treatment of type 2 diabetes. *Curr. Pharm. Des.*, *10*, 1105–1137.

149. Laird, A. D., Cherrington, J. M. (2003). Small molecule tyrosine kinase inhibitors: Clinical development of anticancer agents. *Expert Opin. Investig. Drugs*, *12*, 51–64.

150. Cohen, P. (2002). Protein kinases—The major drug targets of the twenty-first century? *Nat. Rev. Drug Discov.*, *1*, 309–315.

151. Druker, B. J. (2003). Imatinib alone and in combination for chronic myeloid leukemia. *Semin. Hematol.*, *40*, 50–58.

152. Wakeling, A. E. (2002). Epidermal growth factor receptor tyrosine kinase inhibitors. *Curr. Opin. Pharmacol.*, *2*, 382–387.

153. Marx, S. O., Marks, A. R. (2001). Bench to bedside: The development of rapamycin and its application to stent restenosis. *Circulation*, *104*, 852–855.

154. Slamon, D. J., Leyland-Jones, B., Shak, S., Fuchs, H., Paton, V., Bajamonde, A., Fleming, T., Eiermann, W., Wolter, J., Pegram, M., Baselga, J., Norton, L. (2001). Use of chemotherapy plus a monoclonal antibody against HER2 for metastatic breast cancer that overexpresses HER2. *N. Engl. J. Med.*, *344*, 783–792.

155. Mitchell, P. (2004). Erbitux diagnostic latest adjunct to cancer therapy. *Nat. Biotechnol.*, *22*, 363–364.

156. Roskoski, R. Jr. (2004). The ErbB/HER receptor protein-tyrosine kinases and cancer. *Biochem. Biophys. Res. Commun.*, *319*, 1–11.

157. Fabbro, D., Ruetz, S., Buchdunger, E., Cowan-Jacob, S. W., Fendrich, G., Liebetanz, J., Mestan, J., O'Reilly, T., Traxler, P., Chaudhuri, B., Fretz, H., Zimmermann, J., Meyer, T., Giorgio, C., Furet, G., Manley, P. W. (2002). Protein kinases as targets for anticancer agents: From inhibitors to useful drugs. *Pharmacol. Ther.*, *93*, 79–98.

158. Levitzki, A. (2003). Protein kinase inhibitors as a therapeutic modality. *Acc. Chem. Res.*, *36*, 462–469.

159. Force, T., Kuida, K., Namchuk, M., Parang, K., Kyriakis, J. M. (2004). Inhibitors of protein kinase signaling pathways. *Circulation*, *109*, 1196–1205.

160. Toledo, L. M., Lydon, N. B., Elbaum, D. (1999). The structure-based design of ATP-site directed protein kinase inhibitors. *Curr. Med. Chem.*, *6*, 775–805.

161. Traxler, P. (1998). Tyrosine kinase inhibitors in cancer treatment (part II). *Expert Opin. Ther. Patents*, *8*, 1599–1625.

162. Bridges, A. J. (2001). Chemical inhibitors of protein kinases. *Chem. Rev.*, *101*, 2541–2571.

163. Myers, M. R., He, W., Hulme, C. (1997). Inhibitors of tyrosine kinases involved in inflammation and autoimmune disease. *Curr. Pharm. Des.*, *3*, 473–502.

164. Dancey, J., Sausville, E. A. (2003). Issues and progress with protein kinase inhibitors for cancer treatment. *Nat. Drug Discov.*, *2*, 296–313.

165. Nam, N. H., Parang, K. (2003). Current targets for anticancer drug discovery. *Curr. Drug Targets*, *4*, 159–179.

166. Sawyer, T., Boyce, B., Dalgarno, D., Luliucci, J. (2001). Src inhibitors: Genomics to therapeutics. *Expert Opin. Investig. Drugs*, *10*, 1327–1344.

167. Orchard, S. (2002). Kinases as targets: Prospects for chronic therapy. *Curr. Opin. Drug Discov. Devel.*, *5*, 713–717.

168. Garcia-Echeverria, C., Traxler, P., Evans, D. B. (2000). ATP site-directed competitive and irreversible inhibitors of protein kinases. *Med. Res. Rev.*, *20*, 28–57.

169. Fischer, P. M. (2004). The design of drug candidate molecules as selective inhibitors of therapeutically relevant protein kinases. *Curr. Med. Chem.*, *11*, 1563–1583.

170. Fischer, O. M., Streit, S., Hart, S., Ullrich, A. (2003). Beyond Herceptin and Gleevec. *Curr. Opin. Chem. Biol.*, *7*, 490–495.

171. Sawyer, T. K. (2004). Cancer metastatis therapeutic targets and drug discovery: Emerging small-molecule protein kinase inhibitors. *Expert Opin. Invest. Drugs*, *13*, 1–19.

172. Wu, S. Y., McNae, I., Kontopidis, G., McClue, S. J., McInnes, C., Stewart, K. J., Wang, S., Zheleva, D. I., Marriage, H., Lane, D. P., Taylor, P., Fischer, P. M., Walkinshaw, M. D. (2003). Discovery of a novel family of CDK inhibitors with the program LIDAEUS: Structural basis for ligand-induced disordering of the activation loop. *Structure* (*Camb*)., *11*, 399–410.

173. Rossé, G., Séquin, U., Mett, H., Furet, P., Traxler, P., Fretz, H. (1997). Synthesis of modified tripeptides and tetrapeptides as potential bisubstrate inhibitors of the epidermal growth factor receptor protein tyrosine kinase. *Helv. Chim. Acta*, *80*, 653–670.

174. Yuan, C. J., Jakes, S., Elliot, S., Graves, D. (1990). A rationale for the design of an inhibitor of tyrosine kinase. *J. Biol. Chem.*, *265*, 16205–16209.

175. Cushman, M., Chinnasamy, P., Chakraborti, A. K., Jurayj, J., Geahlen, R. L., Haugwitz, R. D. (1990). Synthesis of [beta-(4-pyridyl-1-oxide)-L-alanine4]-angiotensin I as a potential suicide substrate for protein-tyrosine kinases. *Int. J. Pept. Protein Res.*, *36*, 538–543.

176. Bohmer, F. D., Kaagyozov, L., Uecker, A., Serve, H., Botzki, A., Mahboobi, S., Dove, S. (2003). A single amino acid exchange inverts susceptibility of related receptor tyrosine kinases for the ATP-site inhibitor STI-571. *J. Biol. Chem.*, *14*, 5148–5155.

177. Huse, M., Kuriyan, J. (2002). The conformational plasticity of protein kinases. *Cell*, *109*, 275–282.

178. Davies, S. P., Reddy, H., Caivano, M., Cohen, P. (2000). Specificity and mechanism of action of some commonly used protein kinase inhibitors. *Biochem. J.*, *351*, 95–105.

179. Bain, J., McLauchlan, H., Elliott, M., Cohen, P. (2003). The specificities of protein kinase inhibitors: An update. *Biochem. J.*, *371*, 199–204.

180. Lawrence, D. S., Niu, J. (1998). Protein kinase inhibitors: The tyrosine-specific protein kinases. *Pharmacol. Ther.*, *77*, 81–114.

181. Topaly, J., Zeller, W. J., Fruehauf, S. (2001). Synergistic activity of the new ABL-specific tyrosine kinase inhibitor STI571 and chemotherapeutic drugs on BCR-ABL-positive chronic myelogenous leukemia cells. *Leukemia*, *15*, 342–347.

182. Thiesing, J. T., Ohno-Jones, S., Kolibaba, K. S., Druker, B. J. (2000). Efficacy of STI571, an abl tyrosine kinase inhibitor, in conjunction with other antileukemic agents against bcr-abl-positive cells. *Blood*, *96*, 3195–3199.

183. Bogoyevitch, M. A., Boehm, I., Oakley, A., Ketterman, A. J., Barr, R. K. (2004). Targeting the JNK MAPK cascade for inhibition: Basic science and therapeutic potential. *Biochim. Biophys. Acta*, *1697*, 89–101.

184. Shen, G. X. (2003). Selective protein kinase C inhibitors and their applications. *Curr. Drug Targets Cardiovasc. Haematol. Disord.*, *3*, 301–307.

185. Kim, H., Sasaki, T., Maeda, K., Koya, D., Kashiwagi, A., Yasuda, H. (2003). Protein kinase Cbeta selective inhibitor LY333531 attenuates diabetic hyperalgesia through ameliorating cGMP level of dorsal root ganglion neurons. *Diabetes*, *52*, 2102–2109.

186. Showalter, H. D., Kraker, A. J. (1997). Small molecule inhibitors of the platelet-derived growth factor receptor, the fibroblast growth factor receptor, and Src family tyrosine kinases. *Pharmacol. Ther.*, *76*, 55–71.

187. Druker, B. J., Lydon, N. B. (2000). Lessons learned from the development of an abl tyrosine kinase inhibitor for chronic myelogenous leukemia. *J. Clin. Invest.*, *105*, 3–7.

188. Graziani, Y., Erikson, E., Erikson, R. L. (1983). The effect of quercetin on the phosphorylation activity of the Rous sarcoma virus transforming gene product in vitro and in vivo. *Eur. J. Biochem.*, *135*, 583–589.

189. Hagiwara, M., Inoue, S., Tanaka, T., Nunoki, K., Ito, M., Hidaka, H. (1988). Differential effects of flavonoids as inhibitors of tyrosine protein kinases and serine/threonine protein kinases. *Biochem. Pharmacol.*, *37*, 2987–2992.

190. Fischer, P. M., Lane, D. P. (2000). Inhibitors of cyclin-dependent kinases as anti-cancer therapeutics. *Curr. Med. Chem.*, *7*, 1213–1245.

191. Sedlacek, H. H. (2001). Mechanisms of action of flavopiridol. *Crit. Rev. Oncol. Hematol.*, *38*, 139–170.

192. Akiyama, T., Ishida, J., Nakagawa, S., Ogawara, H., Watanabe, S. I., Itoh, N., Shibuy, M., Fukami, Y. (1987). Genistein, a specific inhibitor of tyrosine-specific protein kinases. *J. Biol. Chem.*, *262*, 5592–5595.

193. Nishio, K., Miura, K., Ohira, T., Heiko, Y., Saijo, N. (1994). Genistein, a tyrosine kinase inhibitor, decreased the affinity of p56lck to beta-chain of interleukin-2 receptor in human natural killer (NK)-rich cells and decreased NK-mediated cytotoxicity. *Proc. Soc. Exp. Biol. Med.*, *207*, 227–233.

194. Dudley, D. T., Pang, L., Decker, S. J., Bridges, A. J., Saltiel, A. R. (1995). A synthetic inhibitor of the mitogen-activated protein kinase cascade. *Proc. Natl. Acad. Sci. USA*, *92*, 7686–7689.

195. Pang, L., Sawada, T., Decker, S. J., Saltiel, A. R. (1995). Inhibition of MAP kinase kinase blocks the differentiation of PC-12 cells induced by nerve growth factor. *J. Biol. Chem.*, *270*, 13585–13588.

196. Waters, S. B., Holt, K. H., Ross, S. E., Syu, L. J., Guan, K. L., Saltiel, A. R., Koretzky, G. A., Pessin, J. E. (1995). Desensitization of Ras activation by a feedback disassociation of the SOS-Grb2 complex. *J. Biol. Chem.*, *270*, 20883–20886.

197. Langlois, W. J., Sasaoka, T., Saltiel, A. R., Olefsky, J. M. (1995). Negative feedback regulation and desensitization of insulin- and epidermal growth factor-stimulated p21ras activation. *J. Biol. Chem.*, *270*, 25320–25323.

198. Alessi, D. R., Cuenda, A., Cohen, P., Dudley, D. T., Saltiel, A. R. (1995). PD 098059 is a specific inhibitor of the activation of mitogen-activated protein kinase kinase in vitro and in vivo. *J. Biol. Chem.*, *270*, 27489–27494.

199. Kültz, D., Madhany, S., Burg, M. B. (1998). Hyperosmolality causes growth arrest of murine kidney cells. Induction of GADD45 and GADD153 by osmosensing via stress-activated protein kinase 2. *J. Biol. Chem.*, *273*, 13645–13651.

200. Kemp, B. E., Graves, D. J., Benjamini, E., Krebs, E. G. (1977). Role of multiple basic residues in determining the substrate specificity of cyclic AMP-dependent protein kinase. *J. Biol. Chem.*, *252*, 4888–4894.

201. Kemp, B. E., Pearson, R. B., House, C. M. (1991). Pseudosubstrate-based peptide inhibitors. *Methods Enzymol.*, *201*, 287–304.

202. Ablooglu, A. J., Till, J. H., Kim, K., Parang, K., Cole, P. A., Hubbard, S. R., Kohanski, R. A. (2000). Probing the catalytic mechanism of the insulin receptor kinase with a tetrafluorotyrosine-containing peptide substrate. *J. Biol. Chem.*, *275*, 30394–30398.

203. Niu, J., Lawrence, D. S. (1997). Nonphosphorylatable tyrosine surrogates. Implications for protein kinase inhibitor design. *J. Biol. Chem.*, *272*, 1493–1499.

204. Pargellis, C., Tong, L., Churchill, L., Cirillo, P. F., Gilmore, T., Graham, A. G., Grob, P. M., Hickey, E. R., Moss, N., Pav, S., Regan, J. (2002). Inhibition of p38 MAP kinase utilizing a novel allosteric binding site. *Nat. Struct. Biol.*, *9*, 268–272.

205. Nagar, B., Bornmann, W. G., Pellicena, P., Schindler, T., Veach, D. R., Miller, W. T., Clarkson, B., Kuriyan, J. (2002). Crystal structures of the kinase domain of c-Abl in complex with the small molecule inhibitors PD173955 and imatinib (STI-571). *Cancer Res.*, *62*, 4236–4243.

206. Vu, C. B. (2000). Recent advances in the design and synthesis of SH2 inhibitors of Src, Grb2 and ZAP-70. *Curr. Med. Chem.*, *7*, 1081–1100.

207. Vu, C. B., Corpuz, E. G., Pradeepan, S. G., Violette, S., Bartlett, C., Sawyer, T. K. (1999). Nonpeptidic SH2 inhibitors of the tyrosine kinase ZAP-70. *Bioorg. Med. Chem. Lett.*, *9*, 3009–3014.

208. Muller, G. (2000). Peptidomimetics SH2 domain antagonists for targeting signal transduction. *Topics Curr. Chem.*, *211*, 17–59.

209. Vu, C. B., Corpuz, E. G., Merry, T. J., Pradeepan, S. G., Bartlett, C., Bohacek, R. S., Botfield, M. C., Eyermann, C. J., Lynch, B. A., MacNeil, I. A., Ram, M. K., van Schravendijk, M. R., Violette, S., Sawyer, T. K. (1999). Discovery of potent and selective SH2 inhibitors of the tyrosine kinase ZAP-70. *J. Med. Chem.*, *42*, 4088–4098.

210. Cody, W. L., Lin, Z., Panek, R. L., Rose, D. W., Rubin, J. R. (2000). Progress in the development of inhibitors of SH2 domains. *Curr. Pharm. Des.*, *6*, 59–98.

211. Broom, A. D. (1989). Rational design of enzyme inhibitors: Multisubstrate analog inhibitors. *J. Med. Chem.*, *32*, 2–7.

212. Page, M. I., Jencks, W. P. (1971). Entropic contributions to rate accelerations in enzymic and intramolecular reactions and the chelate effect. *Proc. Natl. Acad. Sci. USA*, *68*, 1678–1683.

213. Bruice, T. C., Benkovic, S. J. (2000). Chemical basis for enzyme catalysis. *Biochemistry*, *30*, 6267–6274.

214. Parang, K., Cole, P. A. (2002). Bisubstrate inhibitors of protein kinases. *Pharmacol. Ther.*, *93*, 1–13.

215. Kruse, C. H., Holden, K. G., Pritchard, M. L., Field, J. A., Rieman, D. J., Greig, R. G., Poste, G. (1988). Synthesis and evaluation of multisubstrate inhibitors of an oncogene-encoded tyrosine-specific protein kinase. *J. Med. Chem.*, *31*, 1762–1767.

216. Ricouart, A., Gesquiere, J. C., Tartar, A., Sergheraert, C. (1991). Design of potent protein-kinase inhibitors using the bisubstrate approach. *J. Med. Chem.*, *34*, 73–78.

217. Traxler, P. M., Wacker, O., Bach, H. L., Geissler, J. F., Kump, W., Meyer, T., Regenass, U., Roesel, J. L., Lydon, N. (1991). Sulfonylbenzoyl-nitrostyrenes: Potential bisubstrate type inhibitors of the EGF-receptor tyrosine protein kinase. *J. Med. Chem.*, *34*, 2328–2337.

218. Uri, A., Järlebark, L., von Kügelgen, I., Schöberg, T., Undén, A., Heibronn, E. (1994). A new class of compounds, peptide derivatives of adenosine 5′-carboxylic acid, includes inhibitors of ATP receptor-mediated responses. *Biorg. Med. Chem.*, *2*, 1099–1105.

219. Uri, A., Raidaru, G., Subbi, J., Padari, K., Pooga, M. (2002). Identification of the ability of highly charges nanomolar inhibitors of protein kinases to cross plasma membranes and carry a protein into cells. *Bioorg. Med. Chem. Lett.*, *12*, 2117–2120.

220. Loog, M., Uri, A., Raidaru, G., Järv, J., Ek, P. (2000). Bi-substrate analog ligands for affinity chromatography of protein kinases. *FEBS Lett.*, *480*, 244–248.

221. Medzihradszky, D., Chen, S. L., Kenyon, G. L., Gibson, B. (1994). Solid-phase synthesis of adenosine phosphopeptides as potential bisubstrate inhibitors of protein kinases. *J. Am. Chem. Soc.*, *116*, 9413–9419.

222. Profit, A. A., Lee, T. R., Lawrence, D. S. (1999). Bivalent inhibitors of protein tyrosine kinases. *J. Am. Chem. Soc.*, *121*, 280–283.

223. Cowburn, D., Zheng, J., Xu, Q., Barany, G. (1995). Enhanced affinities and specificities of consolidated ligands for the Src homology SH3 and SH2 domains of Abelson protein-tyrosine kinase. *J. Biol. Chem.*, *270*, 26738–26741.

224. Parang, K., Till, J. H., Ablooglu, A. J., Kohanski, R. A., Hubbard, S. R., Cole, P. A. (2001). Mechanism-based design of a protein kinase inhibitor. *Nat. Struct. Biol.*, *8*, 37–41.

225. Tanoue, T., Nishida, E. (2002). Docking interactions in the mitogen-activated protein kinase cascades. *Pharmacol. Ther.*, *93*, 193–202.

226. Chang, C. I., Xu, B. E., Akella, R., Cobb, M. H., Goldsmith, E. J. (2002). Crystal structures of MAP kinase p38 complexed to the docking sites on its nuclear substrate MEF2A and activator MKK3b. *Mol. Cell.*, *9*, 1241–1249.

227. Ho, D. T., Bardwell, A. J., Abdollahi, M., Bardwell, L. (2003). A docking site in MKK4 mediates high affinity binding to JNK MAPKs and competes with similar docking sites in JNK substrates. *J. Biol. Chem.*, *278*, 32662–32672.

228. Lee, S., Lin, X., Nam, N. H., Parang, K., Sun, G. (2003). Determination of the substrate-docking site of protein tyrosine kinase C-terminal Src kinase. *Proc. Natl. Acad. Sci. USA*, *100*, 14707–14712.

229. Denny, W. A. (2002). Irreversible inhibitors of the erbB family of protein tyrosine kinases. *Pharmacol. Ther.*, *93*, 253–261.

230. Allen, L. F., Eiseman, I. A., Fry, D. W., Lenehan, P. F. (2003). CI-1033, an irreversible pan-erbB receptor inhibitor and its potential application for the treatment of breast cancer. *Semin. Oncol.*, *30*(5, Suppl. 16), 65–78.

231. Wissner, A., Overbeek, E., Reich, M. F., Floyd, M. B., Johnson, B. D., Mamuya, N., Rosfjord, E. C., Discafani, C., Davis, R., Shi, X., Rabindran, S. K., Gruber, B. C., Ye, F., Hallett, W. A., Nilakantan, R., Shen, R., Wang, Y. F., Greenberger, L. M., Tsou, H. R. (2003). Synthesis and structure-activity relationships of 6,7-disubstituted 4-anilinoquinoline-3-carbonitriles. The design of an orally active, irreversible inhibitor of the tyrosine kinase activity of the epidermal growth factor receptor (EGFR) and the human epidermal growth factor receptor-2 (HER-2). *J. Med. Chem.*, *46*, 49–63.

232. Sreedhar, A. S., Soti, C., Csermely, P. (2004). Inhibition of Hsp90: A new strategy for inhibiting protein kinases. *Biochim. Biophys. Acta*, *1697*, 233–242.

233. Neckers. L. (2002). Hsp90 inhibitors as novel cancer chemotherapeutic agents. *Trends Mol. Med.*, *8*, S55–S61.

234. Nimmanapalli, R., O'Bryan, E., Kuhn, D., Yamaguchi, H., Wang, H. G., Bhalla, K. N. (2003). Regulation of 17-AAG-induced apoptosis: Role of Bcl-2, Bcl-xL, and Bax downstream of 17-AAG-mediated downregulation of Akt, Raf-1 and Src kinases. *Blood*, *102*, 269–275.

235. Masui, H., Kawamoto, T., Sato, J. D., Wolf, B., Sato, G., Mendelsohn, J. (1984). Growth inhibition of human tumor cells in athymic mice by anti-epidermal growth factor receptor antibodies. *Cancer Res.*, *44*, 1002–1007.

236. Slamon, D. J., Clark G. M., Wong, S. G., Levin, W. J., Ullrich, A., McGuire, W. L. (1987). Human breast cancer: Correlation of relapse and survival with amplification of the HER-2/neu oncogene. *Science*, *235*, 177–182.

237. Cho, H.-S., Mason, K., Ramyar, K. X., Stanley, A. M., Gabelli, S. B., Denney, Jr., D. W., Leahy, D. J. (2003). Structure of the extracellular region of HER2 alone and in complex with the Herceptin Fab. *Nature*, *421*, 756–760.

238. Morin, M. J. (2000). From oncogene to drug: Development of small molecule tyrosine kinase inhibitors as anti-tumor and anti-angiogenic agents. *Oncogene*, *19*, 6574–6583.

239. Cohen, M., H., Williams, G. A., Sridhara, R., Chen, G., Pazdur, R. (2003). FDA drug approval summary: Gefitinib (ZD1839) (Iressa) tablets. *Oncologist*, *8*, 303–306.

240. Blagosklonny, M. V., Darzynkiewicz, Z. (2003). Why Iressa failed: Toward novel use of kinase inhibitors (outlook). *Cancer Biol. Ther.*, *2*, 137–140.

241. Dancey, J. E., Freidlin, B. (2003). Targeting epidermal growth factor receptor—Are we missing the mark? *Lancet*, *362*, 62–64.

242. Azam, M., Latek, R. R., Daley, G. O. (2003). Mechanisms of autoinhibition and STI-571/Imatinib resistance revealed by mutagenesis of BCR-ABL. *Cell*, *112*, 831–843.

243. Hidalgo, M., Siu, L. L., Nemunaitis, J., Rizzo, J., Hammond, L. A., Takimoto, C., Eckhardt, S. G., Tolcher, A., Britten, C. D., Denis, L., Ferrante, K., Von Hoff, D. D., Silberman, S., Rowinsky, E. K. (2001). Phase I and pharmacologic study of OSI-774, an epidermal growth factor receptor tyrosine kinase inhibitor, in patients with advanced solid malignancies. *J. Clin. Oncol.*, *19*, 3267–3279.

244. Stamos, J., Sliwkowski, M. X., Eigenbrot, C. (2002). Structure of the epidermal growth factor receptor kinase domain alone and in complex with a 4-anilinoquinazoline inhibitor. *J. Biol. Chem.*, *277*, 46265–46272.

245. Traxler, P., Bold, G., Buchdunger, E., Caravtti, G., Furet, P., Manley, P., O'Reilly, T., Wood, J., Zimmermann, J. (2001). Tyrosine kinase inhibitors: From rational design to clinical trials. *Med. Res. Rev.*, *21*, 499–512.

246. Tiseo, M., Loprevite, M., Ardizzoni, A. (2004). Epidermal growth factor receptor inhibitors: A new prospective in the treatment of lung cancer. *Curr. Med. Chem. Anti-Cancer Agents*, *4*, 139–148.

247. Folkman, J. (1971). Tumor angiogenesis: Therapeutic implications. *N. Engl. J. Med.*, *285*, 1182–1186.

248. Giles, F. J. (2002). The emerging role of angiogenesis inhibitors in hematologic malignancies. *Oncol.* (*Huntington*), *16*, 23–29.

249. Giles, F., Cooper, M. A., Silverman, L., Karp, J. E., Lancet, J. E., Zangari, M., Shami, P. J., Khan, K. D., Hannah, A. L., Cherrington, J. M., Thomas, D. A., Garcia-Manero, G., Albitar, M., Kantarjian, H. M., Stopeck, A. T. (2003). Phase II study of SU5416—A small-molecule, vascular endothelial growth factor tyrosine kinase receptor inhibitor—in patients with refractory myeloproliferative diseases. *Cancer*, *97*, 1920–1928.

250. Hasselbalch, H. C. (2003). SU6668 in idiopathic myelofibrosis—A rational therapeutic approach targeting several tyrosine kinases of importance for the myeloproliferation and the development of bone marrow fibrosis and angiogenesis. *Med. Hypoth.*, *61*, 244–247.

251. Laird, A. D., Vajkoczy, P., Shawver, L. K., Thurnher, A., Liang, C., Mohammadi, M., Schlessinger, J., Ullrich, A., Hubbard, S. R., Blake, R. A., Fong, T. A., Strawn, L. M., Sun, L., Tang, C., Hawtin, R., Tang, F., Shenoy, N., Hirth, K. P., McMahon, G., Cherrington. (2000). SU6668 is a potent antiangiogenic and antitumor agent that induces regression of established tumors. *Cancer Res.*, *60*, 4152–4160.

252. Breier, G., Blum, S., Peli, J., Groot, M., Wild, C., Risau, W., Reichmann, E. (2002). Transforming growth factor-beta and Ras regulate the VEGF/VEGF-receptor system during tumor angiogenesis. *Int. J. Cancer*, *97*, 142–148.

253. Ferrara, N. (2000). VEGF: An update on biological and therapeutic aspects. *Curr. Opin. Biotechnol.*, *11*, 617–624.

254. Ferrara, N. (2000). Vascular endothelial growth factor and the regulation of angiogenesis. *Recent Prog. Horm. Res.*, *55*, 15–35.

255. Ferrara, N., Gerber, H. P. (2001). The role of vascular endothelial growth factor in angiogenesis. *Acta Haematol.*, *106*, 148–156.

256. Kolch, W., Martiny-Baron, G., Kieser, A., Marme, D. (1995). Regulation of the expression of the VEGF/VPS and its receptors: Role in tumor angiogenesis. *Breast Cancer Res. Treat.*, *36*, 139–155.

257. Bold, G., Altmann, K. H., Frei, J., Lang, M., Manley, P. W., Traxler, P., Wietfeld, B., Bruggen, J., Buchdunger, E., Cozens, R., Ferrari, S., Furet, P., Hofmann, F., Martiny-Baron, G., Mestan, J., Rosel, J., Sills, M., Stover, D., Acemoglu, F., Boss, E., Emmenegger, R., Lasser, L., Masso, E., Roth, R., Schlachter, C., Vetterli, W. (2000). New anilinophthalazines as potent and orally well absorbed inhibitors of

the VEGF receptor tyrosine kinases useful as antagonists of tumor-driven angiogenesis. *J. Med. Chem.*, *43*, 2310–2323.

258. Bold, G., Altmann, K. H., Frei, J., Lang, M., Manley, P. W., Traxler, P., Wietfeld, B., Bruggen, J., Buchdunger, E., Cozens, R., Ferrari, S., Furet, P., Hofmann, F., Martiny-Baron, G., Mestan, J., Rosel, J., Sills, M., Stover, D., Acemoglu, F., Boss, E., Emmenegger, R., Lasser, L., Masso, E., Roth, R., Schlachter, C., Vetterli, W. (2000). New anilinophthalazines as potent and orally well absorbed inhibitors of the VEGF receptor tyrosine kinases useful as antagonists of tumor-driven angiogenesis. *J. Med. Chem.*, *43*, 3200–3213.
259. Wood, J. M., Bold, G., Buchdunger, E., Cozens, R., Ferrari, S., Frei, J., Hofmann, F., Mestan, J., Mett, H., O'Reilly, T., Persohn, E., Rosel, J., Schnell, C., Stover, D., Theuer, A., Towbin, H., Wenger, F., Woods-Cook, K., Menrad, A., Siemeister, G., Schirner, M., Thierauch, K. H., Schneider, M. R., Drevs, J., Martiny-Baron, G., Totzke, F. (2000). PTK787/ZK 222584, a novel and potent inhibitor of vascular endothelial growth factor receptor tyrosine kinases, impairs vascular endothelial growth factor-induced responses and tumor growth after oral administration. *Cancer Res.*, *60*, 2178–2189.
260. Smith, J. K., Mamoon, N. M., Duhe, R. J. (2004). Emerging roles of targeted small molecule protein-tyrosine kinase inhibitors in cancer therapy. *Oncol. Res.*, *14*, 175–225.
261. Mohammadi, M., Froum, S., Hamby, J. M., Schroeder, M. C., Panek, R. L., Lu, G. H., Eliseenkova, A. V., Green, D., Schlessinger, J., Hubbard, S. R. (1998). Crystal structure of an angiogenesis inhibitor bound to the FGF receptor tyrosine kinase domain. *EMBO J.*, *17*, 5896–5904.
262. Fong, T. A., Shawver, L. K., Sun, L., Tang, C., App, H., Powell, T. J., Kim, Y. H., Schreck, R., Wang, X., Risau, W., Ullrich, A., Hirth, K. P., McMahon, G. (1999). SU5416 is a potent and selective inhibitor of the vascular endothelial growth factor receptor (Flk-1/KDR) that inhibits tyrosine kinase catalysis, tumor vascularization, and growth of multiple tumor types. *Cancer Res.*, *59*, 99–106.
263. Druker, B. J., Tamura, S., Buchdunger, E., Ohno, S., Segal, G. M., Fanning, S., Zimmermann, J., Lydon, N. B. (1996). Effects of a selective inhibitor of the Abl tyrosine kinase on the growth of Bcr-Abl positive cells. *Nat. Med.*, *2*, 561–566.
264. Barnes, D. J., Melo, J. V. (2003). Management of chronic myeloid leukemia: Targets for molecular therapy. *Semin. Hematol.*, *40*, 34–49.
265. Gorre, M. E., Mohammed, M., Ellwood, K., Hsu, N., Paquette, R., Rao, P. N., Sawyers, C. L. (2001). Clinical resistance to STI-1571 cancer therapy caused by BCR-ABL gene mutation or amplification. *Science*, *293*, 876–880.
266. Roumiantsev, S., Shah, N. P., Gorre, M. E., Nicoll, J., Brasher, B. B., Sawyers, C. L., Van Etten, R. A. (2000). Clinical resistance to the kinase inhibitor STI-571 in chronic myeloid leukemia by mutation of Tyr-253 in the Abl kinase domain P-loop. *Proc. Natl. Acad. Sci. USA*, *99*, 10700–10705.
267. Buchdunger, E., Cioffi, C. L., Law, N., Stiver, D., Ohno-Jones, S., Druker, B. J., Lyndon, N. B. (2000). Abl protein-tyrosine kinase inhibitor ST1571 inhibits in vitro signal transduction mediated by c-Kit and platelet-derived growth factor receptors. *J. Pharmacol. Exp. Ther.*, *295*, 139–145.

268. Okuda, K., Weisberg, E., Gilliland, D. G., Griffin, J. D. (2001). ARG tyrosine kinase activity is inhibited by ST1571. *Blood*, *97*, 2440–2448.

269. van Oosterom, A. T., Judson, I., Verweij, J., Stroobants, S., Donato di Paola, E., Dimitrijevic, S., Martens, M., Webb, A., Sciot, R., Van Glabbeke, M., Silberman, S., Nielsen, O. S. European Organisation for Research and Treatment of Cancer Soft Tissue and Bone Sarcoma Group. (2001). Safety and efficacy of imatinib (STI571) in metastatic gastrointestinal stromal tumours: A phase I study. *Lancet*, *358*, 1421–1423.

270. Blanke, C. D., Eisenberg, B. L., Heinrich, M. C. (2001). Gastrointestinal stromal tumors. *Curr. Treat. Options Oncol.*, *2*, 485–491.

271. Heinrich, M. C., Griffith, D. J., Druker, B. J., Wait, C. L., Ott, K. A., Zigler, A. J. (2000). Inhibition of c-kit receptor tyrosine kinase activity by STI 571, a selective tyrosine kinase inhibitor. *Blood*, *96*, 925–932.

272. Nagar, B., Hantschel, O., Young, M. A., Scheffzek, K., Veach, D., Bornmann, W., Clarkson, B., Superti-Furga, G., Kuriyan, J. (2003). Structural basis for the autoinhibition of c-Abl tyrosine kinase. *Cell*, *112*, 859–871.

273. Noble, M. E., Endicott, J. A., Johnson, L. N. (2004). Protein kinase inhibitors: Insights into drug design from structure. *Science*, *303*, 1800–1805.

274. Blencke, S., Zech, B., Engkvist, O., Greff, Z., Orfi, L., Keri, G., Ullrich, A., Daub, H. (2004). Characterization of a conserved structural determinant controlling protein kinase sensitivity to selective inhibitors. *Chem. Biol.*, *11*, 691–701.

275. Blencke, S., Ullrich, A., Daub, H. (2003). Mutation of threonine 766 in the epidermal growth factor receptor reveals a hotspot for resistance formation against selective tyrosine kinase inhibitors. *J. Biol. Chem.*, *278*, 15435–15440.

276. Wisniewski, D., Lambek, C. L., Liu, C., Strife, A., Veach, D. R., Nagar, B., Young, M. A., Schindler, T., Bornmann, W. G., Bertino, J. R., Kuriyan, J., Clarkson, B. (2002). Characterization of potent inhibitors of the Bcr-Abl and the c-kit receptor tyrosine kinases. *Cancer Res.*, *62*, 4244–4255.

277. Dalgarno, D. C., Botfield, M. C., Rickles, R. J. (1998). SH3 domains and drug design: Ligands, structure, and biological function. *Biopolymers*, *43*, 383–400.

278. Kuriyan, J., Cowburn, D. (1997). Molecular peptide recognition domains in eukaryotic signaling. *Annu. Rev. Biophys. Biomol. Struct.*, *26*, 259–288.

279. Gilmer, T., Rodriguez, M., Jordan, S., Crosby, R., Alligood, K., Green, M., Kimery, M., Wagner, C., Kinder, D., Charifson, P. (1994). Peptide inhibitors of src SH3-SH2-phosphoprotein interactions. *J. Biol. Chem.*, *269*, 31711–31719.

280. Waksman, G., Kominos, D., Robertson, S. C., Pant, N., Baltimore, D., Birge, R. B., Cowburn, D., Hanfusa, H., Majer, B. J., Overduin, M., Resh, M. D., Rios, C. B., Silverman, L., Kuriyan, J. (1992). Crystal structure of the phosphotyrosine recognition domain SH2 of v-src complexed with tyrosine-phosphorylated peptides. *Nature*, *358*, 646–653.

281. Waksman, G., Shoelson, S. E., Pant, N., Cowburn, D., Kuriyan, J. (1993). Binding of a high affinity phosphotyrosyl peptide to the Src SH2 domain: Crystal structures of the complexed and peptide-free forms. *Cell*, *72*, 779–789.

282. Eck, M. J., Shoelson, S. E., Harrison, S. C. (1993). Recognition of a high-affinity phosphotyrosyl peptide by the Src homology-2 domain of p56lck. *Nature*, *362*, 87–91.

283. Grucza, R. A., Bradshaw, J. M., Futterer, K., Waksman, G. (1999). SH2 domains: From structure to energetics, a dual approach to the study of structure-function relationships. *Med. Res. Rev.*, *19*, 273–299.

284. Bradshaw, J. M., Mitaxov, V., Waksman, G. (1999). Investigation of phosphotyrosine recognition by the SH2 domain of the Src kinase. *J. Mol. Biol.*, *293*, 971–985.

285. Park, J., Fu, H., Pei, D. (2003). Peptidyl aldehydes as reversible covalent inhibitors of Src homology 2 domains. *Biochemistry*, *42*, 5159–5167.

286. Marseigne, I., Roques, B. P. (1988). Synthesis of new amino acids mimicking sulfated and phosphorylated tyrosine residues. *J. Org. Chem.*, *53*, 3621–3624.

287. Burke, T. R., Jr., Russ, P., Lim, B. (1991). Preparation of 4-[bis(tert-butoxy)phosphorylmethyl]-*N*-(fluoren-9-ylmethoxycarbonyl)-DL-phenylalanine. A hydrolytically stable analog of *O*-phosphotyrosine potentially suitable for peptide synthesis. *Synthesis*, *11*, 1019–1020.

288. Shoelson, S. E., Chatterjee, S., Chaudhuri, M., Burke, T. R., Jr. (1991). Solid-phase synthesis of nonhydrolyzable phosphotyrosyl peptide analogs with *N*α-Fmoc-(*O*,*O*-di-tert-butyl)phosphono-*p*-methylphenylalanine. *Tet. Lett.*, *32*, 6061–6064.

289. Smyth, M. S., Ford, H., Jr., Burke, T. R. Jr. (1992). A general method for the preparation of benzylic α,α-difluorophosphonic acids; non-hydrolyzable mimetics of phosphotyrosine. *Tetrahedr. Lett.*, *33*, 4137–4140.

290. Burke, T. R., Jr., Smyth, M. S., Nomizu, M., Otaka, A., Roller, P. R. (1993). Preparation of fluoro- and hydroxy-4-(phosphonomethyl)-D,L-phenylalanine suitably protected for solid-phase synthesis of peptides containing hydrolytically stable analogs of *O*-phosphotyrosine. *J. Org. Chem.*, *58*, 1336–1340.

291. Burke, T. R., Jr, Smyth, M. S., Otaka, A., Nomizu, M., Roller, P. P., Wolf, G., Case, R., Shoelson, S. E. (1994). Nonhydrolyzable phosphotyrosyl mimetics for the preparation of phosphatase-resistant SH2 domain inhibitors. *Biochemistry*, *33*, 6490–6494.

292. Yao, Z. J., King, C. R., Cao, T., Kelley, J., Milne, G. W. A., Voigt, J. H., Burke, T. R. (1999). Potent inhibition of Grb2 SH2 domain binding by non-phosphate-containing ligands. *J. Med. Chem.*, *42*, 25–35.

293. Shahripour, A., Plummer, M. S., Lunney, E. A., Para, K. S., Stankovic, C. J., Rubin, J. R., Humblet, C., Fergus, J. H., Marks, J. S., Herrera, R., Hubbell, S. E., Saltiel, A. R., Sawyer, T. K. (1996). Novel phosphotyrosine mimetics in the design of peptide ligands for pp60(src) SH2 domain. *Bioorg. Med. Chem. Lett.*, *6*, 1209–1214.

294. Sundaramoorthi, R., Kawahata, N., Yang, M. G., Shakespeare, W. C., Metcalf, C. A. III, Wang, Y., Merry, T., Eyermann, C. J., Bohacek, R. S., Narula, S., Dalgarno, D. C., Sawyer, T. K. (2003). Structure-based design of novel nonpeptide inhibitors of the Src SH2 domain: Phosphotyrosine mimetics exploiting multifunctional group replacement chemistry. *Biopolymers*, *71*, 717–729.

295. Ye, B., Akamatsu, M., Shoelson, S. E., Wolf, G., Giorgetti-Peraldi, S., Yan, X., Roller, P. P., Burke, T. R., Jr. (1995). L-*O*-(2-Malonyl)tyrosine: A new phosphotyrosyl mimetic for the preparation of Src homology 2 domain inhibitory peptides. *J. Med. Chem.*, *38*, 4270–4275.

296. Burke, T. R., Yao, Z. J., Zhao, H., Milne, G. W. A., Wu, L., Zhang, Z. Y., Voigt, J. H. (1998). Enantioselective synthesis of nonphosphorus-containing phosphotyrosyl mimetics and their use in the preparation of tyrosine phosphatase inhibitory peptides. *Tetrahedron*, *54*, 9981–9994.

297. Gao, Y., Burke, T. R. (2000). Stereoselective preparation of L-4-(2′-malonyl)phenylalanine suitably protected for Fmoc-based synthesis of potent signal transduction inhibitory ligands. *Synlett*, *1*, 134–136.

298. Tong, L., Warren, T. C., Lukas, S., Schembri-King, J., Betageri, R., Proudfoot, J. R., Jakes, S. (1998). Carboxymethyl-phenylalanine as a replacement for phosphotyrosine in SH2 domain binding. *J. Biol. Chem.*, *273*, 20238–20242.

299. Nam, N. H., Pitts, R., Sun, G., Sardari, S., Tiemo, A., Xie, M., Yan, B., Parang, K. (2004). Design of tetrapeptide ligands as inhibitors of the Src SH2 domain. *Bioorg. Med. Chem.*, *12*, 779–787.

300. Nam, N. H., Ye, G., Sun, G., Parang, K. (2004). Conformationally constrained peptide analogues of pTyr-Glu-Glu-Ile as inhibitors of the Src SH2 domain binding. *J. Med. Chem.*, *47*, 3131–3141.

301. Alfaro-Lopez, J., Yuan, W., Phan, B., Kamath, J., Lou, Q., Lam, K. S., Hruby, V. J. (1998). Discovery of a novel series of potent and selective substrate-based inhibitors of $p60^{c\text{-}src}$ protein tyrosine kinase: Conformational and topographical constraints in peptide design. *J. Med. Chem.*, *41*, 2252–2260.

302. Lou, Q., Leftwich, M. E., McKay, R. T., Salmon, S. E., Rychetsky, L., Lam, K. S. (1997). Potent pseudosubstrate-based peptide inhibitors for $p60^{c\text{-}src}$ protein tyrosine kinase. *Cancer Res.*, *57*, 1877–1881.

303. Kamath, J. R., Liu, R., Enstrom, A. M., Lou, Q., Lam, K. S. (2003). Development and characterization of potent and specific peptide inhibitors of $p60^{c\text{-}src}$ protein tyrosine kinase using pseudosubstrate-based inhibitor design approach. *J. Peptide Res.*, *62*, 260–268.

304. Metcalf, C. A. III, van Schravendijk, M. R., Dalgarno, D. C., Sawyer, T. K. (2002). Targeting protein kinases for bone disease: Discovery and development of Src inhibitors. *Curr. Pharm. Des.*, *8*, 2049–2075.

305. Sundaramoorthi, R., Shakespeare, W. C., Keenan, T. P., Metcalf, III C. A., Wang, Y., Mani, U., Taylor, M., Liu, S., Bohacek, R. S., Narula, S. S., Dalgarno, D. C., Van Schravandijk, M. R., Violette, S. M., Liou, S., Adams, S., Ram, M. K., Keats, J. A., Weigele, M., Sawyer, T. K. (2003). Bone-targeted Src kinase inhibitors: Novel pyrrolo- and pyrazolopyrimidine analogues. *Bioorg. Med. Chem. Lett.*, *13*, 3063–3066.

306. Sawyer, T. K., Bohacek, R. S., Eyermann, C. J., Kawahata, N., Metcalf, III, C. A., Shakespeare, W. C., Sundaramoorthi, R., Wang, Y., Yang, M. G. (2002). *Mini-Rev. Med. Chem.*, *2*, 475–489.

307. Uehara, Y., Hori, M., Takeuchi, T., Umezawa, H. (1986). Phenotypic change from transformed to normal induced by benzoquinonoid ansamycins accompanies inactivation of p60src in rat kidney cells infected with Rous sarcoma virus. *Mol. Cell. Biol.*, *6*, 2198–2206.

308. Murakami, Y., Mizuno, S., Hori, M., Uehara, Y. (1988). Reversal of transformed phenotypes by herbimycin A in *src* oncogene expressed rat fibroblasts. *Cancer Res.*, *48*, 1587–1590.

309. Kwon, H. J., Yoshida, M., Fukui, Y., Horinouchi, S., Beppu, T. (1992). Potent and specific inhibition of p60 v-*src* protein kinase both *in vivo* and *in vitro* by radicicol. *Cancer Res.*, *52*, 6926–6930.

310. Bishop, A. C., Shokat, K. M. (1999). Acquisition of inhibitor-sensitive protein kinases through protein design. *Pharmacol. Ther.*, *82*, 337–346.

311. Blake, R. A., Broome, M. A., Liu, X., Wu, J., Gishizky, M., Sun, L., Courtneidge, S. A. (2000). SU6656, a selective src family kinase inhibitor, used to probe growth factor signaling. *Mol. Cell Biol.*, *20*, 9018–9027.

312. Tatoon, L., Molrey, G. M., Chopra, R., Khwaja, A. (2003). The Src-selective kinase inhibitor PP1 also inhibits kit and Bcr-Abl tyrosine kinases. *J. Biol. Chem.*, *278*, 4847–4853.

313. Dorsey, J. F., Jove, R., Kraker, A. J., Wu, J. (2000). The pyrido[2,3-*d*-]pyrimidine derivative PD180970 inhibits p210Bcr-Abl tyrosine kinase and induces apoptosis of K562 leukemic cells. *Cancer Res.*, *60*, 3127–3131.

314. Faltynek, C. R., Schroeder, J., Mauvais, P., Miller, D., Wang, S., Murphy, D., Lehr, R., Kelley, M., Maycock, A., Michne, W., Miski, M., Thunberg, A. L. (1995). Damnacanthal is a highly potent, selective inhibitor of p56lck tyrosine kinase activity. *Biochemistry*, *34*, 12404–12410.

315. Huang, C. K., Wu, F. Y., Ai, Y. X. (1995). Polyhydroxylated 3-(*N*-phenyl) carbamoyl-2-iminochromene derivatives as potent inhibitors of tyrosine kinase $p60^{c\text{-}src}$. *Bioorg. Med. Chem. Lett.*, *5*, 2423–2428.

316. Vermeulen, K. D., Van Bockstaele, R., Berneman, Z. N. (2003). The cell cycle: A review of regulation, deregulation and therapeutic targets in cancer. *Cell Prolif.*, *36*, 131–149.

317. Senderowicz, A. M. (2002). Cyclin-dependent kinases as targets for cancer therapy. *Cancer Chemother. Biol. Response Modif.*, *20*, 169–196.

318. Dai, Y., Grant, S. (2003). Cyclin-dependent kinase inhibitors. *Curr. Opin. Pharmacol.*, *3*, 362–370.

319. Anchan, R. M., Reh, T. A., Angello, J., Balliet, A., Walker, M. (1991). EGF and TGF-alpha stimulate retinal neuroepithelial cell proliferation in vitro. *Neuron*, *6*, 923–936.

320. Han, J., Lee J. D., Bibbs, L., Ulevitch, R. J. (1994). A MAP kinase targeted by endotoxin and hyperosmolarity in mammalian cells. *Science*, *265*, 808–811.

321. Jiang, Y., Chen, C., Li, Z., Guo, W., Gegner, J. A., Lin, S., Han, J. (1996). Characterization of the structure and function of a new mitogen-activated protein kinase (p38beta). *J. Biol. Chem.*, *271*, 17920–17926.

322. Li, Z., Jiang, Y., Ulevitch, R. J., Han, J. (1996). The primary structure of p38 gamma: A new member of p38 group of MAP kinases. *Biochem. Biophys. Res. Commun.*, *228*, 334–340.

323. Gallagher, T. F., Seibel, G. L., Kassis, S., Laydon, J. T., Blumenthal, M. J., Lee, J. C., Lee, D., Boehm, J. C., Fier-Thompson, S. M., Abt, J. W., Soreson, M. E., Smietana, J. M., Hall, R. F., Garigipati, R. S., Bender, P. E., Erhard, K. F., Krog, A. J., Hofmann, G. A., Sheldrake, P. L., McDonnell, P. C., Kumar, S., Young, P. R., Adams, J. L. (1997). Regulation of stress-induced cytokine production by pyridinylimidazoles; inhibition of CSBP kinase. *Bioorg. Med. Chem.*, *5*, 49–64.

324. Adams, J. L., Boehm, J. C., Kassis, S., Gorycki, P. D., Webb, E. F., Hall, R., Sorenson, M., Lee, J. C., Ayrton, A., Griswold, D. E., Gallagher, T. F. (1998). Pyrimidinylimidazole inhibitors of CSBP/p38 kinase demonstrating decreased inhibition of hepatic cytochrome P450 enzymes. *Bioorg. Med. Chem. Lett.*, *8*, 3111–3116.

325. de Laszlo, S. E., Visco, D., Agarwal, L., Chang, L., Chin, J., Croft, G., Forsyth, A., Fletcher, D., Frantz, B., Hacker, C., Hanlon, W., Harper, C., Kostura, M., Li, B., Luell, S., MacCoss, M., Mantlo, N., O'Neill, E. A., Orevillo, C., Pang, M., Parsons, J., Rolando, A., Sahly, Y., Sidler, K., O'Keefe, S. J. (1998). Pyrroles and other heterocycles as inhibitors of p38 kinase. *Bioorg. Med. Chem. Lett.*, *8*, 2689–2694.

326. Revesz, L., Di Padova, F. E., Buhl, T., Feifel, R., Gram, H., Hiestand, P., Manning, U., Zimmerlin, A. G. (2000). SAR of 4-hydroxypiperidine and hydroxyalkyl substituted heterocycles as novel p38 map kinase inhibitors. *Bioorg. Med. Chem. Lett.*, *10*, 1261–1264.

327. Fijen, J. W., Zijlstra, J. G., De Boer, P., Spanjersberg, R., Tervaert, J. W., Van Der Werf, T. S., Ligtenberg, J. J., Tulleken, J. E. (2001). Suppression of the clinical and cytokine response to endotoxin by RWJ-67657, a p38 mitogen-activated protein-kinase inhibitor, in healthy human volunteers. *Clin. Exp. Immunol.*, *124*, 16–20.

328. Wadsworth, S. A., Cavender, D. E., Beers, S. A., Lalan, P., Schafer, P. H., Malloy, E. A., Wu, W., Fahmy, B., Olini, G. C., Davis, J. E., Pellegrino-Gensey, J. L., Wachter, M. P., Siekierka, J. J. (1999). RWJ 67657, a potent, orally active inhibitor of p38 mitogen-activated protein kinase. *J. Pharmacol. Exp. Ther.*, *291*, 680–687.

329. Collis, A. J., Foster, M. L., Halley, F., Maslen, C., McLay, I. M., Page, K. M., Redford, E. J., Souness, J. E., Wilsher, N. E. (2001). RPR203494 a pyrimidine analogue of the p38 inhibitor RPR200765A with an improved in vitro potency. *Bioorg. Med. Chem. Lett.*, *11*, 693–696.

330. Mclay, L. M., Halley, F., Souness, J. E., McKenna, J., Benning, V., Birrell, M., Burton, B., Belvisi, M., Collis, A., Constan, A., Foster, M., Hele, D., Jayyosi, Z., Kelley, M., Maslen, C., Miller, G., Ouldelhkim, M. C., Page, K., Phipps, S., Pollock, K., Porter, B., Ratcliffe, A. J., Redford, E. J., Webber, S., Slater, B., Thybaud, V., Wilsher, N. (2001). The discovery of RPR 200765A, a p38 MAP kinase inhibitor displaying a good oral antiarthritic efficacy. *Bioorg. Med. Chem.*, *9*, 537–554.

331. Ge, B., Gram, H., Di Padova, F., Huang, B., New, L., Ulevitch, R. J., Luo, Y., Han, J. (2002). MAPKK-independent activation of p38alpha mediated by TAB1-dependent autophosphorylation of p38alpha. *Science*, *295*, 1291–1294.

332. Lee, J. C., Laydon, J. T., McDonnell, P. C., Gallagher, T. F., Kumar, S., Green, D., McNulty, D., Blumenthal, M. J., Heys, J. R., Landvatter, S. W., Strickler, J. E., McLaughlin, M. M., Siemens, I. R., Fisher, S. M., Livi, G. P., White, J. R., Adams, J. L., Young, P. R. (1994). A protein kinase involved in the regulation of inflammatory cytokine biosynthesis. *Nature*, *372*, 739–746.

333. Purves, T., Middlemas, A., Agthong, S., Jude, E. B., Boulton, A. J., Fernyhough, P., Tomlinson, D. R. (2001). A role for mitogen-activated protein kinases in the etiology of diabetic neuropathy. *FASEB J.*, *15*, 2508–2514.

334. Igarashi, M., Wakasaki, H., Takahara, N., Ishii, H., Jiang, Z. Y., Yamauchi, T., Kuboki, K., Meier, M., Rhodes, C. J., King, G. L. (1999). Glucose or diabetes activates p38 mitogen-activated protein kinase via different pathways. *J. Clin. Invest.*, *103*, 185–195.

335. Manson, M. M., Holloway, K. A., Howells, L. M., Hudson, E. A., Plummer, S. M., Squires, M. S., Prigent, S. A. (2000). Modulation of signal-transduction pathways by chemopreventive agents. *Biochem. Soc. Trans.*, *28*, 7–12.

336. Ding, Q., Wang, Q., Evers, B. M. (2001). Alterations of MAPK activities associated with intestinal cell differentiation. *Biochem. Biophys. Soc. Trans.*, *284*, 282–288.

337. Andreakos, E. (2003). Targeting cytokines in autoimmunity: New approaches, new promise. *Expert Opin. Biol. Ther.*, *3*, 435–447.

338. Branger, J., van den Blink, B., Weijer, S., Gupta, A., van Deventer, S. J., Hack, C. E., Peppelenbosch, M. P., van der Poll, T. (2003). Inhibition of coagulation, fibrinolysis, and endothelial cell activation by a p38 mitogen-activated protein kinase inhibitor during human endotoxemia. *Blood*, *101*, 4446–4448.

339. Jackson, P. F., Bullington, J. L. (2002). Pyridinylimidazole based p38 MAP kinase inhibitors. *Curr. Topics Med. Chem.*, *2*, 1011–1020.

340. Cirillo, P. F., Pargellis, C., Regan, J. (2002). The non-diaryl heterocycle classes of p38 MAP kinase inhibitors. *Curr. Topics Med. Chem.*, *2*, 1021–1035.

341. Regan, J., Capolino, A., Cirillo, P. F., Gilmore, T., Graham, A. G., Hickey, E., Kroe, R. R., Madwed, J., Moriak, M., Nelson, R., Pargellis, C. A., Swinamer, A., Torcellini, C., Tsang, M., Moss, N. (2003). Structure-activity relationships of the p38alpha MAP kinase inhibitor 1-(5-tert-butyl-2-*p*-tolyl-2*H*-pyrazol-3-yl)-3-[4-(2-morpholin-4-yl-ethoxy)naphthalen-1-yl]urea (BIRB 796). *J. Med. Chem.*, *46*, 4676–4686.

342. Barone, F. C., Irving, E. A., Ray, A. M., Lee, J. C., Kassis, S., Kumar, S., Badger, A. M., Legos, J. J., Erhardt, J. A., Ohlstein, E. H., Hunter, A. J., Harrison, D. C., Philpott, K., Smith, B. R., Adams, J. L., Parsons, A. A. (2001). Inhibition of p38 mitogen-activated protein kinase provides neuroprotection in cerebral focal ischemia. *Med. Res. Rev.*, *21*, 129–145.

343. Hardy, L. W., Malikayil, A. (December 2003). *The Impact of Structure-Guided Drug Design on Clinical Agents.* www.currentdrugdiscovery.com, 15.

344. Cuenda, A., Rouse, J., Doza, Y. N., Meier, R., Cohen, P., Gallagher, T. F., Young, P. R., Lee, J. C. (1995). SB 203580 is a specific inhibitor of a MAP kinase homologue which is stimulated by cellular stresses and interleukin-1. *FEBS Lett.*, *364*, 229–233.

345. Gao, F., Yue, T. L., Shi, D. W., Christopher, T. A., Lopez, B. L., Ohlstein, E. H., Barone, F. C., Ma, X. L. (2002). p38 MAPK inhibition reduces myocardial reperfusion injury via inhibition of endothelial adhesion molecule expression and blockade of PMN accumulation. *Cardiovasc. Res.*, *53*, 414–422.

346. Harper, S. J., LoGrasso, P. (2001). Inhibitors of the JNK signaling pathway. *Drugs Future*, *26*, 957–973.

347. Bennett, B., Sasaki, D., Murray, B., O'leary, E., Sakata, S., Xu, W., Leisten, J. C., Motiwala, A., Pierce, S., Satoh, Y., Bhagwat, S. S., Manning, A. M., Anderson, D. W. (2001). SP600125, an anthrapyrazolone inhibitor of Jun N-terminal kinase. *Proc. Natl. Acad. Sci. USA*, *98*, 13681–13686.

348. Schnabl, B., Bradham, C. A., Bennett, B. L., Manning, A. M., Stefanovic, B., Brenner, D. A. (2001). TAK1/JNK and p38 have opposite effects on rat hepatic stellate cells. *Hepatology*, *34*, 953–963.

349. Goss, G. G., Jiang, L., Vandorpe, D. H., Kieller, D., Chernova, M. N., Robertson, M., Alper, S. L. (2001). Role of JNK in hypertonic activation of $Cl^-$-dependent $Na^+/H^+$ exchange in *Xenopus* oocytes. *Am. J. Physiol.*, *281*, 1978–1990.

350. Rubinfeld, H., Seger, R. (2004). The ERK cascade as a prototype of MAPK signaling pathways. *Methods Mol. Biol.*, *250*, 1–28.

351. Brown, E. J., Beal, P. A., Keith, C. T., Chen, J., Shin, T. B., Schreiber, S. L. (1995). Control of p70 s6 kinase by kinase activity of FRAP in vivo. *Nature*, *377*, 441–446.

352. Kimball, S. D., Webster, K. R. (2001). Cell cycle kinases and checkpoint regulation in cancer. In A. M. Doherty (Ed.), *Annual Reports in Medicinal Chemistry*. Academic, San Diego, Vol. 36, pp. 139–148.

353. Webster K. R., Kimball S. D. (2000). Novel drugs targeting the cell cycle. *Emerg. Drugs*, *5*, 45–59.

354. Ikuta, M., Kamata, K., Fukasawa, K., Honma, T., Machida, T., Hirai, H., Suzuki-Takahashi, I., Hayama, T., Nishimura, S. (2001). Crystallographic approach to identification of cyclin-dependent kinase 4 (CDK4)-specific inhibitors by using CDK4 mimic CDK2 protein. *J. Biol. Chem.*, *276*, 27548–27554.

355. Carini, D. J., Kaltenbach, R. F., Liu, J., Benfield, P. A., Boylan, J., Boisclair, M., Brizuela, L., Burton, C. R., Cox, S., Grafstrom, R., Harrison, B. A., Harrison, K., Akamike, E., Markwalder, J. A., Nakano, Y., Seitz, S. P., Sharp, D. M., Trainor, G. L., Sielecki, T. M. (2001). Identification of selective inhibitors of cyclin dependent kinase 4. *Bioorg. Med. Chem. Lett.*, *11*, 2209–2211.

356. Honma, T., Hayashi, K., Aoyama, T., Hashimoto, N., Machida, T., Fukasawa, K., Iwama, T., Ikeura, C., Ikuta, M., Suzuki-Takahashi, I., Iwasawa, Y., Hayama, T., Nishimura, S., Morishima, H. (2001). Structure-based generation of a new class of potent Cdk4 inhibitors: New de novo design strategy and library design. *J. Med. Chem.*, *44*, 4615–4627.

357. Honma, T., Yoshizumi, T., Hashimoto, N., Hayashi, K., Kawanishi, N., Fukasawa, K., Takaki, T., Ikeura, C., Ikuta, M., Suzuki-Takahashi, I., Hayama, T., Nishimura, S., Morishima, H. (2001). A novel approach for the development of selective Cdk4 inhibitors: Library design based on locations of Cdk4 specific amino acid residues. *J. Med. Chem.*, *44*, 4628–4640.

358. Fry, D. W., Bedford, D. C., Harvey, P. H., Fritsch, A., Keller, P. R., Wu, Z., Dobrusin, E., Leopold, W. R., Fattaey, A., Garrett, M. D. (2001). Cell cycle and biochemical effects of PD 0183812. A potent inhibitor of the cyclin D-dependent kinases CDK4 and CDK6. *J. Biol. Chem.*, *276*, 16617–16623.

359. Soni, R., Muller, L., Furet, P., Schoepfer, J., Stephan, C., Zumstein-Mecker, S., Fretz, H., Chaudhuri, B. (2000). Inhibition of cyclin-dependent kinase 4 (Cdk4) by fascaplysin, a marine natural product. *Biochem. Biophys. Res. Commun.*, *275*, 877–884.

360. Soni, R., O'Reilly, T., Furet, P., Muller, L., Stephan, C., Zumstein-Mecker, S., Fretz, H., Fabbro, D., Chaudhuri, B. J. (2001). Selective in vivo and in vitro effects of a small molecule inhibitor of cyclin-dependent kinase 4. *Natl. Cancer Inst.*, *93*, 436–446.

361. Knockaert, M., Greengard, P., Meijer, L. (2002). Pharmacological inhibitors of cyclin-dependent kinases. *Trends Pharmacol. Sci.*, *23*, 417–425.

362. Jeong, H. W., Kim, M. R., Son, K. H., Han, M. Y., Ha, J. H., Garnier, M., Meijer, L., Kwon, B. M. (2000). Cinnamaldehydes inhibit cyclin dependent kinase 4/cyclin D1. *Bioorg. Med. Chem. Lett.*, *10*, 1819–1822.

363. Meijer, L., Leclerc, S., Leost, M. (1999). Properties and potential-applications of chemical inhibitors of cyclin-dependent kinases. *Pharmacol. Ther.*, *82*, 279–284.

364. Fischer, P. M., Gianella-Borradori, A. (2003). CDK inhibitors in clinical development for the treatment of cancer. *Exp. Opin. Invest. Drugs*, *12*, 955–970.

365. Zhai, S., Senderowicz, A. M., Sausville, E. A., Figg, W. D. (2002). Flavopiridol, a novel cyclin-dependent kinase inhibitor, in clinical development. *Ann. Pharmacother.*, *36*, 905–911.

366. Leclerc, S., Garnier, M., Hoessel, R., Marko, D., Bibb, J. A., Snyder, G. L., Greengard, P., Biernat, J., Wu, Y. Z., Mandelkow, E. M., Eisenbrand, G., Meijer, L. (2001). Indirubins inhibit glycogen synthase kinase-3 beta and CDK5/p25, two protein kinases involved in abnormal tau phosphorylation in Alzheimer's disease. A property common to most cyclin-dependent kinase inhibitors? *J. Biol. Chem.*, *276*, 251–260.

367. Meijer, L., Raymond, E. (2003). Roscovitine and other purines as kinase inhibitors. From starfish oocytes to clinical trials. *Acc. Chem. Res.*, *36*, 417–425.

368. Vesely, J., Havlicek, L., Strnad, M., Blow, J. J., Donella-Deana, A., Pinna, L., Letham, D. S., Kato, J., Detivaud, L., Leclerc, S. (1994). Inhibition of cyclin-dependent kinases by purine analogues. *Eur. J. Biochem.*, *224*, 771–786.

369. Meijer, L., Borgne, A., Mulner, O., Chong, J. P., Blow, J. J., Inagaki, N., Inagaki, M., Delcros, J. G., Moulinoux, J. P. (1997). Biochemical and cellular effects of Roscovitine, a potent and selective inhibitor of the cyclin-dependent kinases cdc2, cdk2 and cdk5. *Eur. J. Biochem.*, *243*, 527–536.

370. Gray, N. S., Wodicka, L., Thunnissen, A. W. H., Norman, T. C., Kwon, S., Espinoza, F. H., Morgan, D. O., Barnes, G., LeClerc, S., Meijer, L. (1998). Exploiting chemical libraries, structure and genomics in the search for kinase inhibitors. *Science*, *281*, 533–538.

371. Senderowicz, A. M. (2003). Novel small molecule cyclin-dependent kinases modulators in human clinical trials. *Cancer Biol. Ther.*, *2*, S84–95.

372. Goekjian, P. G., Jirousek, M. R. (2001). Protein kinase C inhibitors as novel anticancer drugs. *Exp. Opin. Invest. Drugs*, *10*, 2117–2140.

373. Sorbera, L. A., Silvestre, J., Rabasseda, X., Castaner, J. (2000). LY-333531 mesylate hydrate: Symptomatic antidiabetic; protein kinase C inhibitor. *Drugs Future*, *25*, 1017–1026.

374. Way, K. J., Chou, E., King, G. L. (2000). Identification of PKC-isoform-specific biological actions using pharmacological approaches. *Trends Pharmacol. Sci.*, *21*, 181–187.

375. Propper, D. J., McDonald, A. C., Man, A., Thavasu, P., Balkwill, F., Braybrooke, J. P., Caponigro, F., Graf, P., Dutreix, C., Blackie, R., Kaye, S. B., Ganesan, T. S., Talbot, D. C., Harris, A. L., Twelves, C. (2001). Phase I and pharmacokinetic study of PKC412 an inhibitor of protein kinase C. *J. Clin. Oncol.*, *19*, 1485–1492.

376. Fuse, E., Tanii, K., Kurata, N., Kobayashi, H., Shimada, Y., Tamura, T., Tanigawara, Y., Lush, R. D., Headlee, D., Figg, W. D., Arbuck, S. G., Senderowicz, A. M.,

Sauseville, E. A., Akinaga, S., Kuwabara, T., Kobayashi, S. (1998). Unpredicted clinical pharmacology of UCN-01 caused by specific binding to human alpha1-acid glycoprotein. *Cancer Res.*, *58*, 3248–3255.

377. Fuse, E., Tanii, K., Takai, K., Asanome, K., Kurata, N., Kobayashi, H., Kuwabara, T., Kobayashi, S., Sugiyama, Y. (1999). Altered pharmacokinetics of a novel anti-cancer drug, UCN-01, caused by specific high affinity binding to alpha1-acid glycoprotein in humans. *Cancer Res.*, *58*, 1054–1060.
378. Schaufelberger, D. E., Koleck, M. P., Beutler, J. A., Vatakis, A. M., Atvardo, A. B., Andrews, L. V., Muschick, G. M., Forenza, S. (1991). The large scale isolation of bryostatin 1 from *Bugula nerutina* following current good manufacturing practices. *J. Nat. Prod.*, *54*, 1265–1270.
379. Smith, J. B., Smith, L., Pettit, G. R. (1985). Bryostatins: Potent, new mitogens that mimic phorbol ester tumour promoters. *Biochem. Biophys. Res. Commun.*, *132*, 939–945.
380. Szallasi, Z., Smith, C. B., Pettit, G. R., Blumberg, P. M. (1994). Differential regulation of protein kinase C isozymes by bryostatin 1 and phorbol 12-myristate 13-acetate in NIH 3t3. broblasts. *J. Biol. Chem.*, *269*, 2118–2124.
381. Hocevar, B. A., Fields, A. P. (1991). Selective translocation of βII protein kinase C to the nucleus of human promyelocytic (HL-60) leukemia cells. *J. Cell. Biol.*, *266*, 28–33.
382. Neumunaitis, J., Holmlund, J. T., Kraynak, M., Richards, D., Bruce, J., Ognoskie, N., Kwoh, T. J., Greary, R., Dorr, A., Von Hoff, D., Eckhardt, S. G. (1999). Phase I evaluation of ISIS 3521, an antisense oligodeoxynucleotide to protein kinase C-α, in patients with advanced cancer. *J. Clin. Oncol.*, *17*, 3586–3595.
383. Breitenlechner, C., Gassel, M., Hidaka, H., Kinzel, V., Huber, R., Engh, R. A., Bossemeyer, D. (2003). Protein kinase A in complex with Rho-kinase inhibitors Y-27632, Fasudil, and H-1152P: Structural basis of selectivity. *Struct. (Cambr.)*, *11*, 1595–1607.
384. Walker, E. H., Pacold, M. E., Perisic, O., Stephens, L., Hawkins, P. T., Wymann, M. P., Williams, R. L. (2000). Structural determinants of phosphoinositide 3-kinase inhibition by wortmannin, LY294002, quercetin, myricetin, and staurosporine. *Mol. Cell.*, *6*, 909–919.
385. Lakkakorpi, P. T., Wesolowski, G., Zimolo, Z., Rodan, G. A., Rodan, S. B. (1997). Phosphatidylinositol 3-kinase association with the osteoclast cytoskeleton, and its involvement in osteoclast attachment and spreading. *Exp. Cell Res.*, *237*, 296–306.
386. Hunag, S., Houghton, P. J. (2002). Inhibitors of mammalian target of rapamycin as novel antitumor agents: From bench to clinic. *Curr. Opin. Invest. Drugs*, *3*, 295–304.
387. Podsypanina, K., Lee, R. T., Politis, C., Hennessy, I., Crane, A., Puc. J., Neshat, M., Wang, H., Yang, L., Gibbons, J., Frost, P., Dreisbach, V., Blenis, J., Gaciong, Z., Fisher, P., Sawyers, C., Hedrick-Ellenson, L., Parsons, R. (2001). An inhibitor of mTOR reduces neoplasia and normalizes p70/S6 kinase activity in Pten+/– mice. *Proc. Natl. Acad. Sci. USA*, *98*, 10314–10319.

# RNA-BASED THERAPIES

STEVE PASCOLO
*CureVac, The RNA people GmbH, Paul Ehrlich Strasse 15, 72076 Tuebingen, Germany*

1 INTRODUCTION: RNA, THE SUPERLATIVE BIOMOLECULE
From the Isolation of Nuclein to the Chemical and Structural Characterization of Nuclei Acids
Natural Functions of RNA in Cells
Potentially Therapeutic Features of RNA
Advantages of RNA-Based Therapies
Natural and Modified RNA
2 DRUGS BASED ON THE CAPACITY OF RNA TO FORM DEFINED THREE-DIMENSIONAL STRUCTURES: APTAMERS
Definition of Aptamers
RNA Mimetics Suitable for Aptamer Function
Preclinical and Clinical Utilization of Aptamers
Lead Company: Eyetech Pharmaceutical, Inc.
3 SYNTHETIC RNA FOR GENE REGULATION: ANTISENSE, RIBOZYMES, AND SIRNA
Antisense and Ribozymes
Small Inhibitory RNA (siRNA)
Delivery of RNA Molecules for Suppression of Gene Expression
4 RNA FOR PROTEIN PRODUCTION: mRNA
Definition of mRNA
RNA Mimetics Suitable for mRNA Function
Preclinical and Clinical Utilization of mRNA
Lead Company: MERIX Biosciences
5 RNA FOR IMMUNOSTIMULATION: DOUBLE-STRANDED AND STABILIZED RNA
Definition of Immunostimulating RNAs
Preclinical and Clinical Utilization of Immunostimulating RNA
Lead Company: Hemispherx Biopharma, Inc.

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

6 PRODUCTION OF RNA AND LEGAL ISSUES
Manufacturing of RNA
Regulatory Aspects (Quality Management)
7 CONCLUSION AND PERSPECTIVES: RNA-BASED THERAPIES OF THE FUTURE
References

## 1 INTRODUCTION: RNA, THE SUPERLATIVE BIOMOLECULE

### From the Isolation of Nuclein to the Chemical and Structural Characterization of Nucleic Acids

Living organisms are composed of four different types of organic molecules: sugars, lipids, proteins, and nucleic acids. Nucleic acids were first isolated from the nucleus of human pus cells by Friedrich Miescher (1844–1895) in the kitchen of Tübingen's castle (Germany) in 1868. Friedrich Miescher called this substance *nuclein*. He showed that it exists in all type of cells and found that when prepared from salmon sperm, it was associated to a peptide that he named *protamine*. Further fractionation of nuclein allowed the derivation of a pure product by Richard Altmann (1852–1900) in 1889. Altman named this product *nucleic acid*. Meanwhile, Albrecht Kossel (1853–1927, winner of the Nobel prize for Medicine in 1910) used chemical hydrolysis of nuclein to discover nucleotides (1885–1901): adenine (A) and guanine (G) are purines; thymidine (T), Cytosine (C), and uracile (U) are pyrimidines. Kossel demonstrated that the nucleotides contain pentose. The existence of two different nucleic acids was seen as early as 1910. One (prepared from yeast and plants) was found to be sensitive to alkaline lysis and contain A, C, G, and U (Fig. 1*a*), the other (prepared from thymus), was alkali resistant. Phoebus Levene (1869–1940) a student of Kossel, managed in 1929 to degrade thymus-derived nuclein using enzymes and demonstrated that it consists of A, C, G, and T, a demethylated version of U. Levene and Walter Abraham Jacobs (1883–1967) showed that the pentose of the ACGT nucleic acid, as opposed to that of ACGU nucleic acid, is a 2′-deoxyribose. The ACGT nucleic acid was named deoxyribonuclei-cacid (DNA) and the ACGU nucleic acid ribonucleicacid (RNA). Research by Tornbjörn Caspersson (1911–1998) and Jean Brachet (1909–1998) proved that DNA is contained in chromosomes located in the nucleus and RNA is found principally in the cytosole. In 1953 Jim Watson (1928–) and Francis Crick (1916–2004) demonstrated that nucleotides are complementary: A pairs with T or U and G pairs with C (Fig. 1*a*). Together with Maurice Wilkins (1917–) they received the Nobel prize in 1962 for their resolution of the structure of DNA. By the end of the 1950s, the characteristics of DNA and RNA were well described. Both nucleic acids consist of four nucleotides but have different subcellular localization. Base composition is U for RNA and T for DNA (Fig. 1*a*). Chemical characteristics are deoxyribose in DNA, hydroxyribose in RNA, which is consequently sensitive to high pH

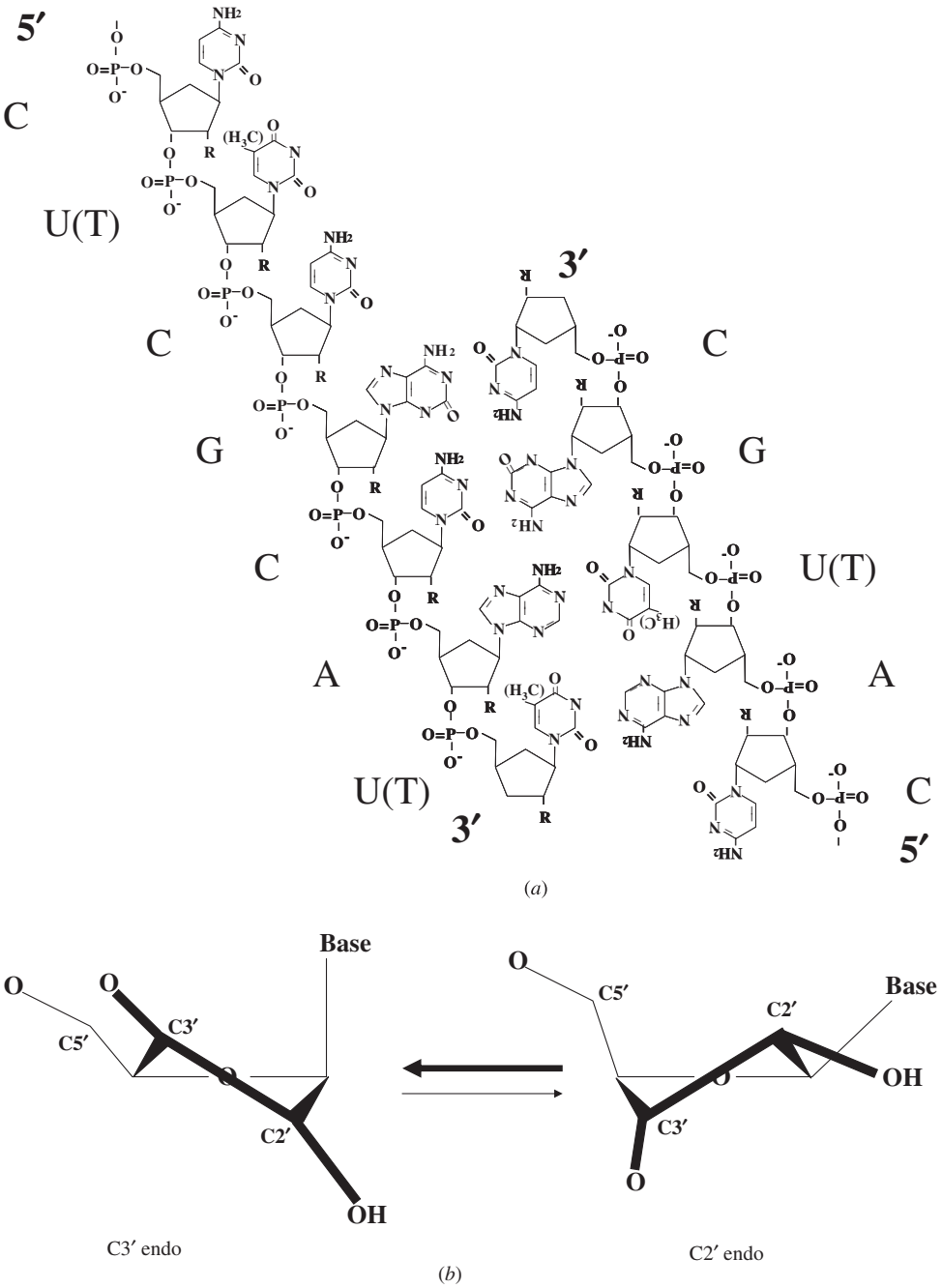


**Figure 1** Structure of nucleic acids. (*a*) Representation of two hybridized nucleic acid strands: one of the sequence $^{5'}$CUCGCAU$^{3'}$ ($^{5'}$CTCGCAT$^{3'}$ for DNA) and one of the sequence $^{5'}$CAUGC$^{3'}$ ($^{5'}$CATGC$^{3'}$ for DNA). R is H for DNA and OH for RNA. (*b*) Structure of the pentose in nucleic acids. Due to steric effects and electronic bonds, ribonucleotides contain dominantly the C3′-endo form of the ribose.

(Fig. 1*a*). Isomeric structures are furane ring in equilibrium between the 2′-endo and 3′-endo for DNA, although the 3′-endo form is favored in RNA (Fig. 1*b*). Three-dimensional structures are DNA is usually a double-stranded molecule with two antiparallel strands associated by interstrand base pairing, whereas RNA usually consists of one strand where intramolecular base pairing dictates the secondary and tertiary structure of the molecule. It was only at the beginning of the 1960s that the hypothesized roles of RNA molecules in gene expression were being experimentally documented.

## Natural Functions of RNA in Cells

At the end of the 1930s, by observing that RNA concentrations are higher in rapidly dividing cells, Casperson and Brachet presented the first experimental evidence that cellular RNA was involved in protein synthesis. The RNA-rich granules (now known as ribosomes and containing ribosomic RNA, rRNA) observed in the cytosole by Albert Claude were hypothesized by Brachet in 1946 to participate in protein synthesis. This theory was confirmed in the 1950s by experimental data coming from electron microscopy [1] and pulse-chase [2] experiments. In 1958, Crick formulated the central dogma of molecular biology [3]: DNA is transcribed into RNA, which is translated into protein. In 1961, Brenner [4] as well as Gros and co-workers [5] described unstable RNA molecules (now known as messenger RNA, mRNA), which carry genetic information from DNA in the nucleus to ribosomes in the cytosole. At the same time, another intermediate in protein synthesis was identified in the form of an aminoacyl-RNA (transfer RNA, tRNA) [6–8]. Thus, just a few years after the publication of "The Central Dogma of Molecular Biology" by Crick [3], all molecules mediating protein synthesis were identified: mRNA, rRNA, and tRNA are different forms of RNA molecules necessary in the process of gene expression.

At the end of the 1960s, another activity of RNA was discovered: double-stranded RNA structures activate the immune system [9]. More than 10 years later, a third function of RNA apart from coding–decoding genetic information and immunostimulation was reported: RNA can catalyze reactions. The discovery by Guerrier-Takada and co-workers [10] and Cech and Bass [11] of catalytic RNA (ribozymes) was acknowledged with a Nobel prize (Chemistry, 1989). Ribozymes can be used to specifically degrade mRNA, thus shutting down the expression of a defined gene. They can be seen as a catalytic version of antisense RNA [12] for inhibition of gene expression. The recent discovery of small inhibitory RNA (siRNA) is a new breakthrough in the world of molecular biology and molecular medicine [13]. Similarly to RNA antisense and ribozymes, siRNAs are molecules that naturally evolved to specifically suppress the synthesis of a defined protein, and they can be used as genetic tools to manipulate gene expression.

On reviewing RNA's natural functions, it appears that RNA is the only organic molecule that has the ability to recapitulate all biological activities nec-

essary for life: containment of genetic information, mediation and regulation of protein synthesis, scaffolding of tri-dimensional structures and enzymatic activities. Accordingly, an "RNA world" was suggested as the start of life 4.2 billion years ago. At that time, self-replicating RNA molecules or RNA complexes may have been the basis of life. They could have evolved toward organisms that contained a new type of macromolecule such as DNA (an RNA derivative that resists basic pH), lipids, and proteins that can potentate the genetic, structural, and enzymatic activities of RNA. During this evolution, RNA remained the central player in all biological processes: RNA is the genome of many viruses [human immunodeficiency virus (HIV), human T-cell lymphotropic virus (HTLV), hepatitis C virus (HCV), Semliki Forest virus, etc.]. In cells it is the obligate transient copy of genes that are sent out of the nucleus (mRNA); it is the main constituent of the whole translation machinery (tRNA and rRNA); it regulates DNA replication (RNA primers for the production of Okazaki fragments); it modulates gene expression (ribozymes, siRNA, antisenses RNA); it is used as a matrix to build new DNA sequences (RNA contained in telomerase complexes); it provides energy (mainly adenine and guanine in their triphosphate form) and signals infection (double-stranded or stabilized RNA).

In addition to these characterized activities, we are still discovering new functions of RNA molecules. For example, microRNAs (miRNA) that are structurally (short ca. 21 bases double-stranded RNA) and functionally (post-transcriptional inhibition of gene expression) similar to siRNA but transcribed from the host instead of being brought by a pathogen are now found to affect the genome [14–16]. In particular, chromatin structure and DNA methylation at the locus of a gene targeted by miRNA is modified in a way that transcription of the precise gene is inhibited. This discovery together with the demonstration that RNA serves as a matrix for DNA synthesis in chromosome's telomers challenge Cricks' central dogma: RNA is not only a product of the genome but also a feedback modulator of DNA. This brings interesting perspectives such as RNA-directed evolution, whereby outside signals would be materialized into special RNA structures (encoded by the host) capable of "educating" the genome (modify genes or gene expression). Such an RNA-mediated evolution of genomes would be a new function of RNA molecules.

## Potentially Therapeutic Features of RNA

All the natural functions of RNA have been explored as the basis for therapeutics. The capacity of RNA to form defined three-dimensional structures is exploited in aptamers (Section 2 of "Protein Kinase Inhibitors in Drug Discovery"), its functions in gene regulation are tested for the specific control of tumor growth or virus infections (Section 3 of "Protein Kinase Inhibitors in Drug Discovery"), its protein-encoding feature is used for gene therapy and vaccination (Section 4 of "Protein Kinase Inhibitors in Drug Discovery"), and its natural capacity to activate the immune system is used in adjuvant formulations (Section 5 of "Protein Kinase Inhibitors in Drug Discovery"). As shown in Figure 2,

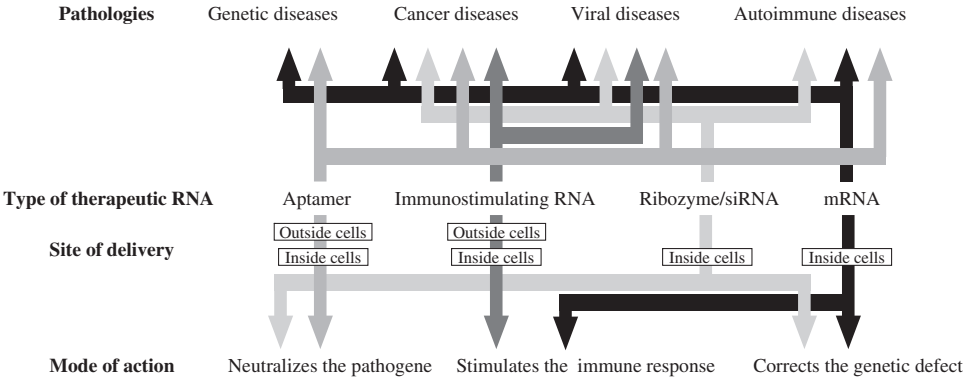


**Figure 2** Applications of RNA-based therapies. The four types of therapeutic RNA are listed; their mode and site of action is shown. They can neutralize the pathogene (inhibit virus replication, block the activity of a protein, interfere with the growth of tumor cells), stimulate the specific immune response (trigger an antitumor, antiviral, or antibacteria immune response as well as suppressing allergy or other types of autoimmunity), or correct defects due to an inherited genetic mutation (mRNA coding for the wild-type protein in the context of diseases such as cystic fibrosis or siRNA destroying an mRNA coding for an aberrant protein that has a dominant phenotype).

aptamers and mRNA approaches can virtually address all kinds of diseases, including pathologies resulting from inherited genetic defects. Gene interference mediated by siRNA and ribozymes as well as immunostimulation by RNA should prove to be especially useful in fights against infectious diseases, allergies, and cancer.

## Advantages of RNA-Based Therapies

The main advantages of RNA as an active pharmaceutical ingredient (API) are:

1. Its production and conservation are easy (see Section 6 of "Protein Kinase Inhibitors in Drug Discovery").
2. Its half-life can be exactly determined through chemical modifications (see discussion of natural and modified RNA).
3. As opposed to DNA or proteins, RNA is not immunogenic (no specific immune response against RNA was reported).
4. Through its versatile activities, RNA can theoretically be an API against any disease: for correction of pathological genetic defects, prevention or cure of infections, treatment of tumors, therapy against degenerative diseases, and control of allergies.

## Natural and Modified RNA

Chemically, RNA is a robust molecule at neutral or acidic pH (the oxygen on the 2′ position of the ribose makes it sensitive to basic pH). As opposed to most proteins, lipids, or DNA, RNA molecules can be heated to more than

95°C, frost-thawed, lyophilized, and resuspended in water solutions as many times as necessary for production, storage, and use.

The main weakness of RNA resides in the ubiquitous presence of ribonucleases (RNases). Many species of RNases are produced by prokaryotic or eukaryotic life [17]. Some RNases are secreted outside the cells; some are resident in the cytosole. The function of secreted RNases is not fully understood but may be a mechanism to prevent infection by pathogens such as virus-containing RNA genomes (influenza, Sindbis, HIV, etc.). Alternatively, it was hypothesized that neighboring cells can communicate through the secretion and uptake of RNA. In this context, ubiquitous RNases are seen as controllers of RNA-based intercell communication. In any case, RNases are abundant in blood, in and on the skin, in cells, as well as in any extracellular spaces, thus rendering RNA-based treatments inefficient compared to those based on DNA.

The principal remedy against RNase activity is the modification of the therapeutic RNA backbone (Fig. 3). Several chemical modifications made in synthetic RNA were shown to increase the molecule's half-life in RNase-contaminated milieus. Mainly, replacement of the 2′ oxygen by a fluorine atom, an ammonium ($NH_2$) or azido ($N_3$) group, replacement of the 2′-OH by O-alkyl (methyl, ethyl, propyl, methoxy-ethoxy) or O-allyl groups, replacement of an oxygen atom in the phosphorus group by a sulfur atom (phosphorothioate backbone), or replacement of the 3′OH by NH (phosphoramidate backbone) are standard chemical modifications that render RNA molecules resistant toward RNases [18]. These modified compounds are grouped under the term *RNA mimetics*. Chemical modifications can interfere with the

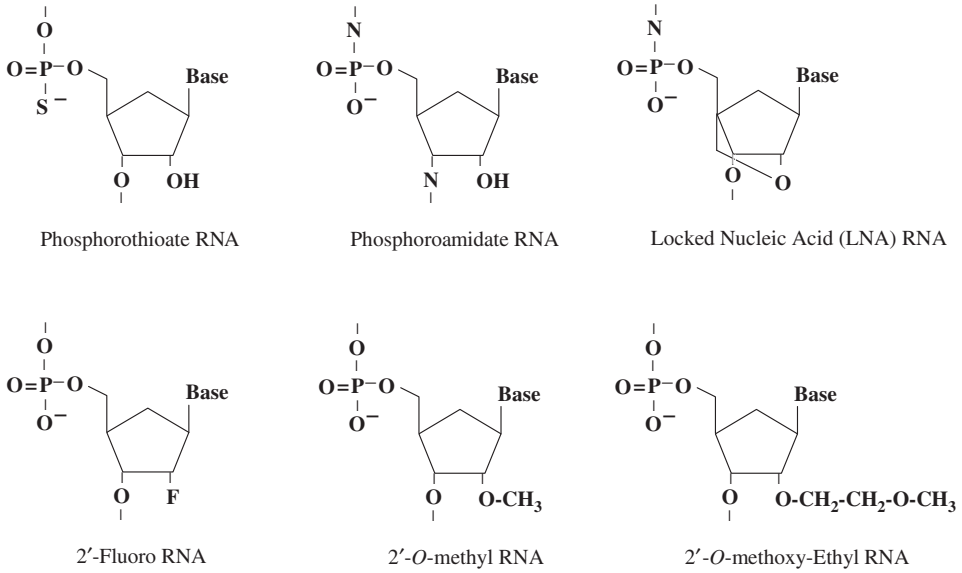


**Figure 3** Description of the most common modifications introduced in RNA (especially synthetic oligonucleotides) designed for therapeutic use.

expected structural, genetic, immunomodulating or gene expression control activity of the RNA (see discussions of RNA mimetics through out this chapter). Partial chemical modification of the therapeutic RNA where only some residues are modified usually results in a good compromise between enhanced resistance toward RNases and conservation of biological activity.

Although for some of the modifications (2′ fluoro, 2′ amino, 2′ azido), the substitution of the 2′ oxygen by other atoms or chemical groups semantically turns the fully modified RNA mimetics into DNA, the designation of all these derivative as RNA molecules is justified by three characteristics:

1. Sensitivity to RNases that usually persists in spite of being severely reduced.
2. The presence of uridine as one of the bases.
3. The dominant C3′-endo structure of the pentose. The stabilized C3′-endo structure in all RNA and RNA derivatives provides them with an increased base-pairing thermostability in comparison to DNA. Consequently, RNA–RNA or RNA–DNA hybrids are thermodynamically more stable than DNA–DNA hybrids. Recently, a method to fix the optimal C3′-endo isomeric structure of the pentose was reported: A methylene bridge connects the 2′ oxygen to the 4′ carbon of the sugar [19]. Such structurally fixed furanose rings provide the RNA oligonucleotide (termed LNA for locked nucleic acid) with not only extreme resistance toward RNase but also an increased thermostability of base pairing: the melting temperature of LNA–target RNA duplexes is several degrees higher than the melting temperature of natural RNA–RNA duplexes.

The capacity of RNA mimetics to outscore natural RNA in therapeutic uses depends on the influence of the modification(s) on the RNA's function (whether aptamers, ribozymes, siRNA, immunostimulating, or mRNA activity is required), stability, membrane permeation, pharmacokinetic properties, and biodistribution. The design of an optimal RNA mimetic requires intense research efforts but has already resulted in the release of one medicament and several promising new potential drugs that are in clinical trials. The next section will present the different technologies behind RNA-based therapeutics, the optimization of RNA mimetics for each application, the results of in vitro and in vivo preclinical or clinical studies, and a look at one of the leading company that supports the development and commercialization of each therapeutic RNA compound.

## 2 DRUGS BASED ON THE CAPACITY OF RNA TO FORM DEFINED THREE-DIMENSIONAL STRUCTURES: APTAMERS

### Definition of Aptamers

Specific interactions between RNA and proteins have been recognized. They largely depend on the secondary and tertiary structure of the partially double-

stranded nucleic acid. As exemplified by tRNAs (see review by Wittenhagen and Kelley [20]), small alterations in the RNA structure can abrogate selective association to proteins and hinder the RNA's functions. Many RNA sequences, which fold into defined structures, especially stem-loop structures (a double-stranded stem that holds a single-stranded loop of as little as four and up to several hundreds nucleotides), control important biochemical processes. For example, stability or translation of mRNA by stem-loop structures in the mRNA [21], assembling of ribosomes after precise three-dimensional folding of the ribosomal RNAs [22] or replication of HIV that is dependent on stem-loop structures in some precise parts of the HIV RNA genome [23]. The idea to use RNA molecules, that is, three-dimensional structures generated by the refolding of an RNA sequence into stem-loop domains, as highly selective ligands, functionally similar to antibodies, was turned into a process of evolution (variation, selection, and replication) called SELEX (systematic evolution of ligands by exponential enrichment) by Tuerk and Gold [24]. At the same time, Ellington and Szostak [25] coined the name *aptamer* (from the Latin *aptus*: to fit) for such in vitro selected ligands composed exclusively of nucleic acids. An aptamer is an oligonucleotide that folds into a highly specific and stable three-dimensional structure due to intrastrand hybridizations that defined several stem loops structures of different sizes. The characterization of an aptamer proceeds through two steps represented in Figure 4.

***Step 1: Selection*** From a library of random, ca. 30 to 40-mere sequences flanked by defined ca. 20-nucleotide sequences (one being also a promoter recognized by a specific RNA polymerase; Fig. 4*a*), oligonucleotides that can adapt a structure that allows them to bind to an immobilized substrate (a protein, a nucleic acid, or any type of chemical compound as well as entire cells [26]) are retained (Fig. 4*b*). The RNA oligonucleotides that cannot bind to the immobilized substrate are discarded by washing the solid support with standard salt solutions such as phosphate-buffered saline (PBS). Selectively bound oligonucleotides are eluted from the solid support by using detergents and/or heating and/or chaotropic solutions or releases of the ligand from the support.

***Step 2: Amplification*** The eluted oligonucleotides are amplified by reverse transcription polymerase chain reaction (RT-PCR) thanks to the flanking defined sequences. Subsequently, they are transcribed into RNA and reselected on the immobilized ligand.

Several (usually 10 to 20) rounds of selection/amplification result in enrichment of the random mix with RNA that specifically interact with the immobilized ligand. At the end, these aptamers are reverse transcribed, cloned into a plasmid, and sequenced. Usually, most molecules are found to have related sequences or at least a common core sequence. This core sequence can be produced by in vitro transcription of an engineered plasmid or by chemical synthesis. Its specific binding to the ligand used for selection can be verified and

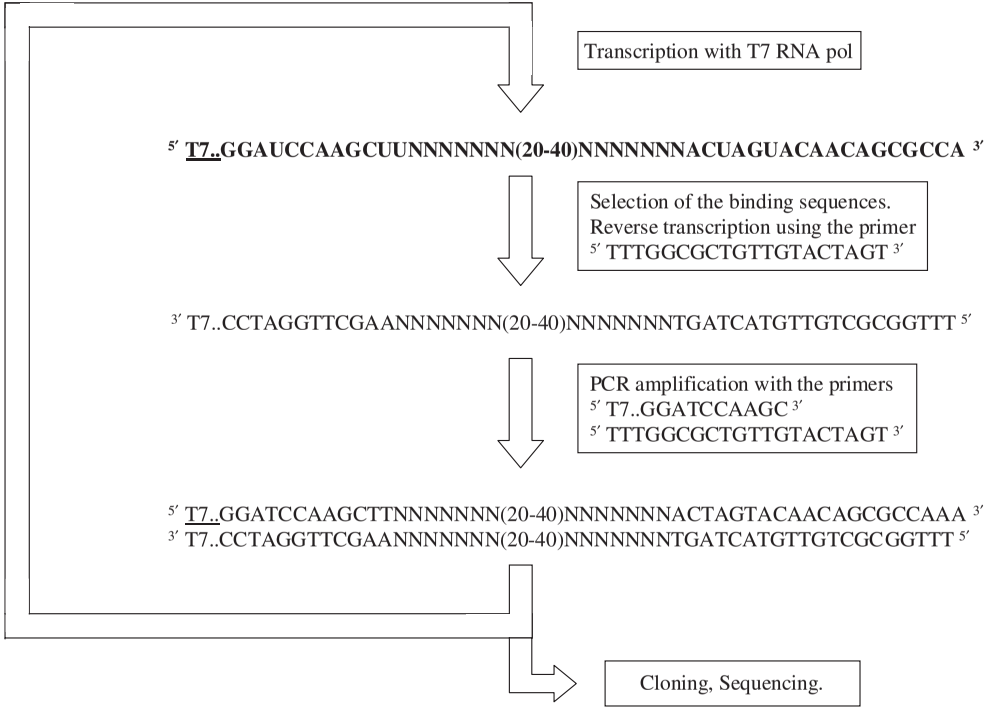


(*a*)

**Figure 4** Generation of aptamers. A synthetic oligonucleotide containing a T7 promoter sequence (underlined) followed by a random sequence (from 20 to 40 nucleotides in length usually) and a defined ca. 20 residues sequence [shown in bold in (*a*)] is coincubated with an immobilized ligand [shown in (*b*)]. The sequences that can adopt a configuration that allows them to bind to the ligand are retained. They can be eluted, reverse transcribed, and PCR amplified [as shown in (*a*)] before being transcribed. The resulting RNA molecules can be re-selected on the immobilized ligand. After several rounds (usually 10 to 20) the eluted, reverse-transcribed, and PCR-amplified molecules can be cloned and sequenced. Synthetic RNA corresponding to these sequences can be produced, stabilized by RNA modifications, and studied in vitro (affinity to their ligand, definition of the minimal active sequences) and *in vivo* (efficiency, biodistribution, half-life).

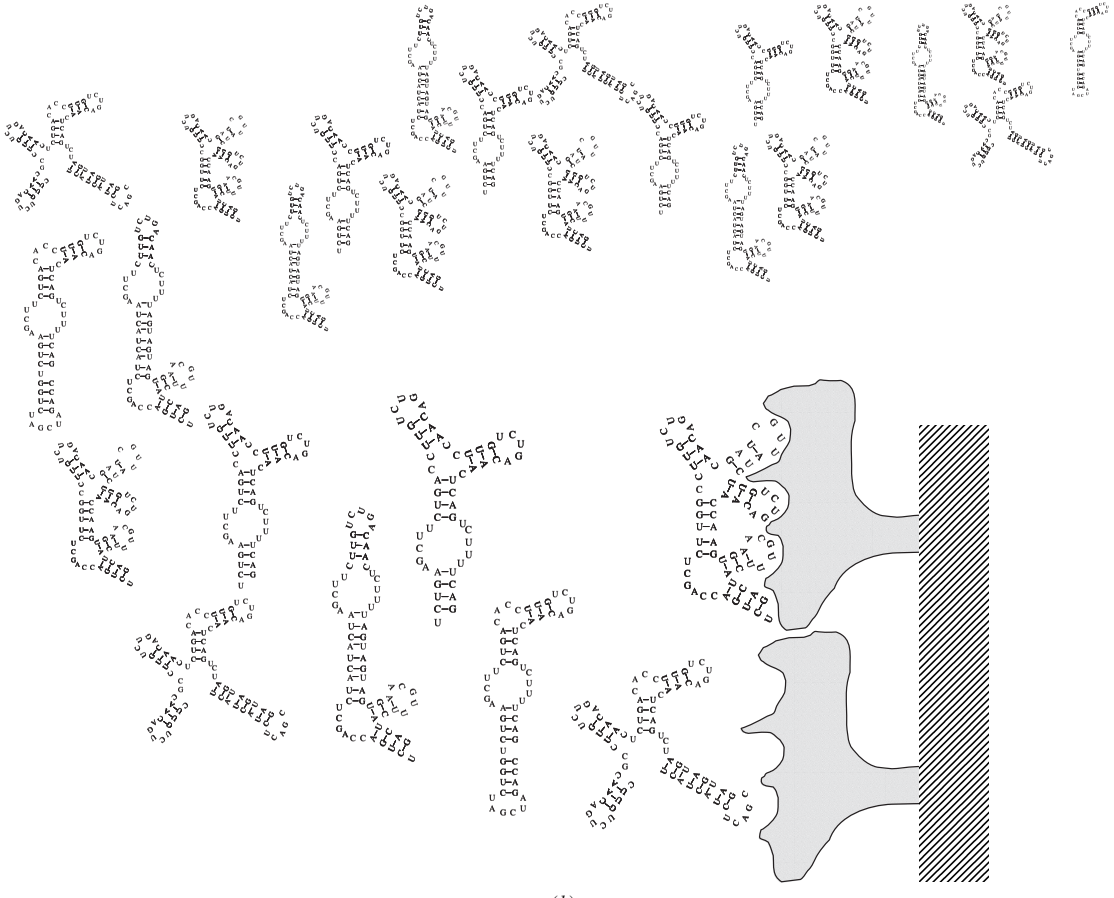

(*b*)

**Figure 4** *Continued*

further studied by standard biochemical tests [e.g., enzyme-linked immunosorbent assay (ELISA) or Biacore]. Aptamers have affinities for their cognate "antigen" equal or superior to those obtained with antibodies ($K_d$ values in the subnanomolar range). Since the whole process is done in vitro starting from a synthetic library of about $10^{14}$ different oligonucleotide sequences, aptamers can recognize any kind of ligands, whereas the isolation of high-affinity antibodies may be hindered by tolerance to self (especially for proteins conserved through evolution such as heat shock proteins). The possibility to produce large amounts of aptamers by chemical synthesis is another attractive characteristic of this technology in comparison to antibodies.

Aptamers can be DNA or RNA oligonucleotides. However, thanks to the following properties of RNA:

1. Its natural catabolism
2. The possibility to control its half-life in different physiologic milieus through the introduction of chemical modifications
3. Its lack of immunogenicity
4. The fact that it cannot modify the host genomes
5. The lack of systemic effects after in vivo application (as opposed to DNA oligonucleotides especially if they contain the so-called CpG motifs)

RNA aptamers are the safest version for therapeutic use. A large number of aptamers specific for a broad range of defined ligands have been isolated and characterized. Their inventory is presented in the aptamer database: http://aptamer.icmb.utexas.edu/ [27].

## RNA Mimetics Suitable for Aptamer Function

Usually, the SELEX process is performed with 2′fluoro pyrimidine RNA. The T7 RNA polymerase used during the amplification step is permissive to the integration of 2′fluoro UTP and 2′fluoro CTP [28]. Since most natural RNases recognize pyrimidines [29], the 2′fluoropyrimidines RNA are more resistant than natural RNA and are consequently preferred both for the in vitro selection and for direct in vivo application. The modifications used during the selection process may be required not only for the stability but also for the high-affinity binding of the aptamer to its ligand [30]. Indeed, if the 2′fluoros are involved in the determination of the precise three-dimensional structure of the RNA, their replacement with OH or other groups may modify the functional characteristics of the aptamer. Nevertheless, it was shown in several cases that on top or instead of the 2′fluoro substitutions, some modifications such as a phosphoramidate backbone [31], 2′-amino or 2′-*O*-methyl [32] substitutions as well as LNA nucleotides [33] enhance the stability of aptamers without reducing their function. Increasing the resistance of aptamers toward nucleases by any one of or a combination of the possible modifications is an empirical process and must be individually tested as it may result in a decrease or an increase of the affinity of RNA to its target molecule.

## Preclinical and Clinical Utilization of Aptamers

Potentially therapeutic aptamers have been described in diverse medical fields such as vascular biology, immunology, brain disorders, and infectious diseases [34, 35]. The selected molecular target of the aptamer may be outside the cells, for example, an extracellular receptor, a hormone, or a growth factor. In these cases, no transfection reagents are needed for aptamer applications. Of special interest for therapeutics and diagnostic use are the characterized aptamers that recognize HIV gp120 [36], thrombin [37], CTLA-4 [38], the surface protein from trypanosome [39], or prion (the pathogenic form of the prp protein: $prp^{sc}$) [40]. Each of these has direct pharmaceutical potential.

The first step toward the preclinical utilization of an aptamer is the definition of the minimal active sequence. Due to criteria of high-quality production, aptamers should not be longer than 40 bases (oligonucleotides longer than 40 bases are difficult to produce and to separate from N-1 contaminants). The second step is defining the chemical modifications that will provide the optimal stabilization with no loss of function.

With such minimal and stabilized oligonucleotides, preclinical tests can be conducted to test for efficiency in vivo. Several aptamers passed these tests and are now in human phase I and II trials. One aptamer, the anti-VEGF (vascular endothelial growth factor) pegylated aptamer (Macugen), went successfully through a phase III clinical trial and is now approved as a drug. VEGF known for inducing blood vessel permeability is an extracelluar growth factor required for neovascularisation. VEGF exists in 5 different isoforms. One of them, isoform 165, plays an important role in the abnormal blood vessel growth and leakage associated with several pathologies including the eye disease subfoveal choroidal neovascularization (CNV) secondary to age-related macular degeneration (AMD). This disease can lead to vision loss. In animal models, anti-VEGF agents have inhibited blood vessel formation and leakage. In humans, VEGF concentrations in the eyes correlates with wet AMD or diabetic macular edema (DME). Both AMD and DME are diseases of the retina that can lead to blindness. A 2′-fluoropyrimidine RNA aptamer capable of binding and of specifically inhibiting (blocks binding to its receptor) VEGF isoform 165 was characterized [41]. In a pegylated version [addition of polyethylene glycol (PEG) increases the size of the therapeutic RNA, preventing clearance of the molecule from the injected site], it showed long bioavailability (several weeks after intravitreal injection, [42]), good safety profile, and efficacy in human clinical trials [43, 44]. This anti-VEGF aptamer, called Macugen, is the first RNA oligonucleotide licensed as a drug.

## Lead Company: Eyetech Pharmaceuticals, Inc.

Eyetech Pharmaceuticals, Inc. is a biopharmaceutical company that specialized in the development and commercialization of novel therapeutics to treat diseases of the eye (see www.eyetech.com). Eyetech is traded on the NASDAQ National Market System under the symbol EYET. Its initial public

offering of 6,500,000 shares at a price of $21.00 per share was in January 2004. Today Eyetech is a company with over 150 employees with locations in New York, New Jersey and Massachusetts.

Macugen™ (Pegaptanib sodium injection), the anti-VEGF pegylated aptamer, Eyetech's lead product, is BEING developed for diseases of the eye together with Pfizer. By docking to and blocking the function of VEGF, it interferes with the development of abnormal vessel growth and leakage seen in neovascular (wet) age-related macular degeneration (AMD) and diabetic retinopathy.

In the United States, it is estimated that as many as 15 million people suffer from some form of AMD and that there are more than 1.6 million cases of wet AMD. Approximately 200,000 new cases of wet AMD arise each year in the United States. Although wet AMD represents approximately 10 per cent of all AMD cases, it is responsible for 90% of the vision loss associated with AMD. A majority of wet AMD patients experience severe vision loss in the affected eye within months to two years after diagnosis of the disease. Because AMD generally affects adults over 50 years of age, it is expected that the incidence of AMD will increase significantly as the baby boom generation ages and overall life expectancy increases.

Macugen™ has also been studied in patients with DME (diabetic macular edema). In the United States, there are approximately 500,000 people suffering from DME, with approximately 75,000 new cases each year. It is expected that the incidence of DME in the United States will increase as the number of people with diabetes increases. Because the existing treatments for both wet AMD and DME have significant limitations, there is a significant unmet medical need for a new therapy for these diseases.

In 2001, Eyetech initiated two Phase II/III pivotal clinical trials of Macugen for the treatment of wet AMD. It involved 117 medical centers and enrolled patients with subfoveal wet AMD: 578 patients in the North American trial and 612 patients in the international trial. Patients received one intravitreal injection of macugen EVERY SIX WEEKS. The primary efficacy endpoint in these trials was the proportion of patients losing less than 15 letters, or three lines, of visual acuity on the eye chart from baseline after 54 weeks. Based on the analysis of the data from the two trials, the primary efficacy endpoint was met. Macugen was approved as a drug against AMD in 2004.

Meanwhile, a phase II placebo controlled trial was done in the context of DME. Three doses were tested: 0.3 mg, 1 mg and 3 mg every six weeks for at least twelve weeks and endpoints were measured after 30 weeks. Results showed statistical significance for all doses, with the highest efficacy for the 0.3 mg dose, similarily to the AMD trial.

IN ADDITION TO AMD and DME, Macugen™ may prove to be efficient in the treatment of retinal vein occlusion (RVO), a condition that is also characterized by high VEGF levels, abnormal blood vessel growth and blood vessel leakage. RVO occurs when the circulation of a retinal vein becomes obstructed, causing blood vessel bleeding and leakage in the retina. Laser

therapy is sometimes used to treat this condition, but with limited efficacy. A phase II trial with Macugen™ started in May 2004.

## 3 SYNTHETIC RNA FOR GENE REGULATION: ANTISENSE, RIBOZYMES, AND SIRNA

### Antisense and Ribozymes

***Definition of Antisense and Ribozymes*** The first RNA molecules found to be capable of suppressing gene expression were antisense RNA [45]. They consist of a short linear antiparallel copy of the target mRNA. Through hybridization to the mRNA, antisense RNA specifically block translation [5]. This process is not reversible: One antisense RNA inhibits the translation of one mRNA molecule. For an efficient suppression of gene expression an excess of the antisense molecules compared to the targeted mRNA in the cytosole is necessary [46, 47]. Systemic delivery of synthetic antisense RNA oligonucleotides is not expected to result in a high concentration of the therapeutic molecule in the cytosole of all target cells. Thus, it would probably not be efficient in vivo for complete gene suppression. The antisense approach best suits local delivery and utilization of the more stable DNA oligonucleotides [48]. As far as RNA-based therapies are concerned, efficient in vivo suppression of gene expression can be obtained with the catalytic version of antisense RNA: the ribozyme. This type of molecule can specifically catalyze the cleavage or religation of a target RNA and, as opposed to antisense molecules, are recycled. Ribozymes were discovered in the early 1980s by Cech and Bass [11]. Their characterization arose from the studies of a self-splicing intron from *Tetrahymena thermophilia* [49]. Further studies by Altman's group also pointed to the existence of natural ribozymes capable of specifically cleaving target RNA [10]. Since these initial results, ribozymes were found in several organisms (eukaryotes, viruses, bacteria) in different forms [50]: hammerhead, hairpin, Neurospora Varkud satellite, hepatitis delta virus, self-splicing group I and group II introns, and Rnase P-ribozymes. They were shown to be capable of performing different kinds of activities such as transesterification or hydrolysis. Furthermore, in vitro selected ribozyme can catalyze the synthesis of nucleotides [51] or the creation of amide bonds [52]. The attempts toward development of therapeutic ribozymes are mainly based on two forms of naturally occurring ribozymes: hammerheads and hairpins. Hammerhead ribozymes were isolated from viroid RNA [53]. Their enzymatic core is ca. 40 nucleotides in length and consists of two substrate binding arms and a conserved catalytic domain (Fig. 6). Hammerhead ribozymes can cleave the phosphodiester bond of a target RNA after a UA, UU, or UC dinucleotide sequence (AUC and GUC triplets are optimal [54]). Hairpin ribozymes were discovered in tobacco ringspot virus satellite RNA. They are usually designed to cleave the phosphodiester bond located 5′ to a GUC sequence in the target RNA [55]. The catalytic part of hairpin ribozymes contains a domain

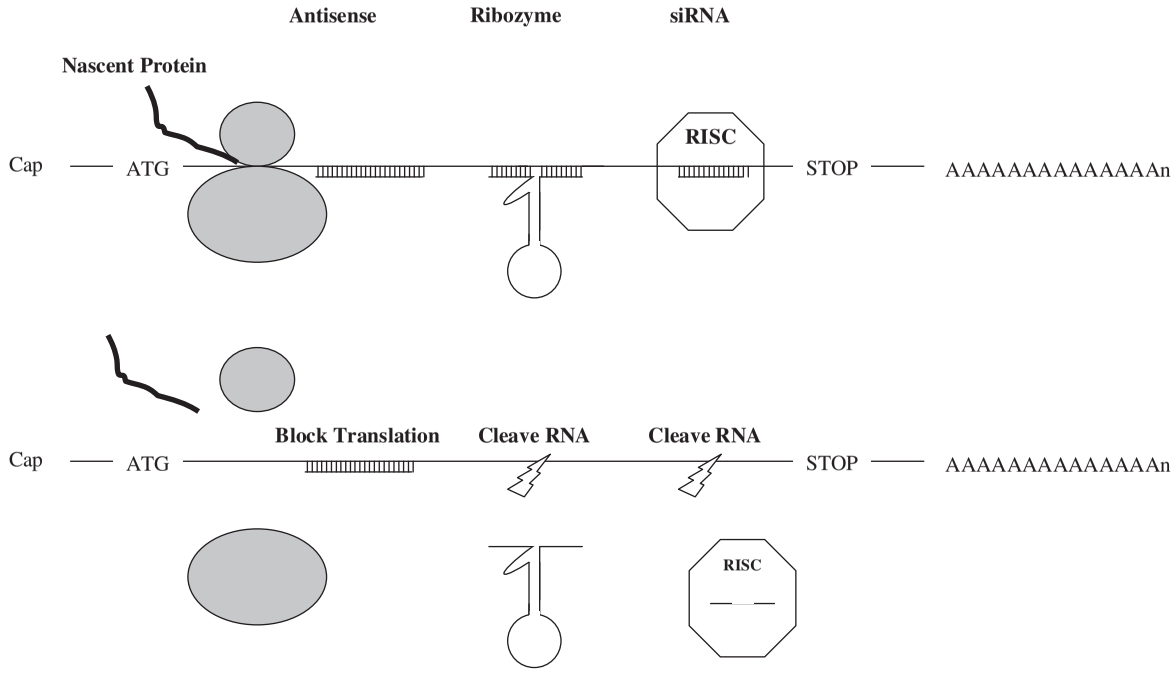


**Figure 5** Strategies of gene suppression by RNA molecules: Antisense, ribozymes and siRNA can suppress gene expression by blocking (antisense) or cutting (ribozyme and siRNA) specifically a messenger RNA. Although antisense RNA and ribozymes act autonomously, siRNA require the housekeeping RISC complex in order to perform its catalytic activity.

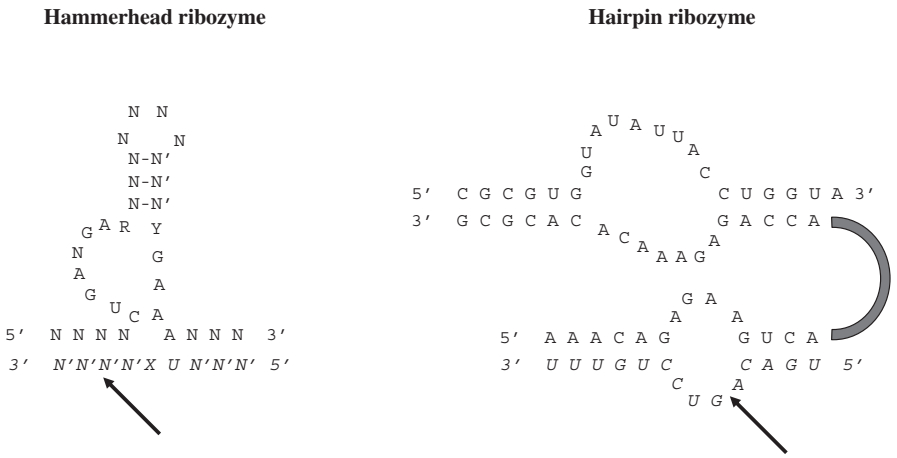


**Figure 6** Representation of hammerhead and hairpin ribozymes. N is any nucleotide, N is a nucleotide complementary to N, X is any nucleotide but G, Y is a pyrimidine, R is a purine nucleotide complementary to Y. *The arrow represents the site of cleavage on the target RNA.* Adapted from Jen and Gewirtz (2000).

composed of a loop (two single-stranded RNA sequences of several bases) flanked by double-stranded structures and a single-stranded sequence that will recognize the target RNA (Fig. 6). The complex between hairpin and target RNA should also be structured in a loop flanked by double-stranded sequences. The cleavage happens in the single-stranded sequence of the loop in the target RNA. The single-stranded parts of hairpins (contained in the loops) are very conserved and necessary for the ribozyme's activity [56]. Both hairpin and hammerhead ribozymes require metal ions for their catalytic activity. As far as cleavage activity is required, hammerheads are preferred to hairpins because they are naturally prone to be more efficient in hydrolysis than transesterification (religation), whereas hairpin ribozymes are as efficient for both reactions. This is one of the reasons why the ribozymes that are being designed, improved, produced, and actually used in therapeutic settings are hammerheads.

***RNA Mimetics Suitable for Hammerhead Ribozyme Function*** In order to be protected from quick degradation by ubiquitous RNases, synthetic ribozymes must be modified. A great variety of available modified nucleotides (listed in Natural and Modified RNA and described in Fig. 3) were tested for hammerhead ribozymes function [57–59]. Several modifications at the single-stranded catalytic site of the hammerhead inhibit its function. On the contrary, nuclease-resistant nucleotides such as 2′-fluoro, 2′-amino, 2′ alkyl, or allyl derivatives can be used at all positions involved in the recognition of the targeted sequence. There, a phosphorothioate backbone can bring further

stability. In addition, the 3′ extremity of the molecule can be occupied by a modified nucleotide that will prevent degradation by 3′–5′ exonucleases. Through these modifications, the serum half-life of modified hammerhead ribozymes could be increased to more than 10 days as compared to less than 1 min for the unmodified version.

***Preclinical and Clinical Utilizations of Hammerhead Ribozymes*** The first step toward the clinical use of hammerhead ribozymes is the characterization of the optimal accessible target sequence stretch in the mRNA to be cleaved. Messenger RNA or virus RNA genomes fold into complex, heterogeneous, and relatively unpredictable structures. Moreover, they associate with many proteins. Thus, a small part of the targeted RNA is exposed for recognition by ribozymes. To identify these sequences, a combination of computer prediction and RNAse H-sensitivity assay was developed [60]: Synthetic DNA oligonucleotides are incubated with the targeted mRNA; then, RNAse H is added. It will cleave the RNA only if the DNA probe could hybridize to it. Thus, the study of the integrity of the RNA will indicate if the sequence complementary to the DNA probe is accessible (RNA is cleaved) or not accessible (RNA is not cleaved). Only accessible sequences are considered as potential targets for ribozymes. Alternately, some random approaches were developed where optimal ribozymes are selected from libraries of active ribozymes (containing a hammerhead core catalytic sequence) having randomized substrate recognition arms [61–63]. Once the optimal target sequence is identified, the best RNA-stabilizing strategy is tested: At permissive positions (all positions involved in the recognition of the target RNA), 2′ modified nucleotides replace the natural nucleotides and a phosphorothioate backbone is included. Such RNase-resistant hammerhead ribozymes can to some extent spontaneously enter cells and show gene-suppressing activity in vitro in cell culture and in vivo after local [64] or systemic delivery [65]. In 1998, the first synthetic ribozyme ANGIOZYME entered a human clinical trial. It is a chemically stabilized RNA oligonucleotide capable of specifically cleaving the mRNA coding the high-affinity receptor for the vascular endothelial growth factor (VEGFR-1) [65]. Through blockade of angiogenesis by downregulation of VEGFR-1 expression, the ribozyme is expected to reduce tumor growth. Indeed, ANGIOZYME was shown to inhibit tumor growth in several mouse models. Daily intravenous or subcutaneous application resulted in detectable plasma levels of the synthetic RNA ribozyme [66]. Subsequently, it was tested in human phase I trials and proven to be nontoxic at doses up to 300 mg/m$^2$ administered intravenous or subcutaneous. ANGIOZYME was further evaluated in a phase II monotherapy trial involving breast cancer patients. Forty-five late-stage metastatic cancer patients with progressive disease were treated with doses of 100 mg/m$^2$ ANGIOZYME by daily subcutaneous injections. Most patients continued to progress as shown by computed tomography (CT) scans 6 weeks after the first injection. Nevertheless, the activity of the ribozyme could be recorded by a significant decrease of the amount of soluble VEGFR-1 detected in the blood at week 6 (soluble VEGFR-1 is encoded by

an alternatively spliced mRNA derived from VEGFR-1 encoding premessenger RNA; both forms of the mRNA are recognized by ANGIOZYME). A phase II trial was conducted in which Angiozyme was used in combination with a standard chemotherapy (irinotecan/5-fluorouracil/leucovorin: IFL). Eighty-three colorectal cancer patients with metastatic disease were enrolled. They received daily subcutaneous injections of 100 mg/m$^2$ of ANGIOZYME in combination with IFL. Again, reduction of soluble VEGFR-1 in the blood was detectable after several weeks of injection. But, as far as the criteria "time to progression" is concerned, there was no clear difference between the group of the ANGIOZYME + IFL-treated patients compared to historical groups of IFL-treated patients. Nevertheless, since no control group (IFL alone) was included in the phase II study, definitive conclusions cannot be drawn. Another trial in tumor patients was performed with a hammerhead ribozyme called HERZYME and designed to inhibit the expression of the human epidermal growth factor receptor type 2 (Her2). This receptor is overexpressed in many carcinomas and drives cell proliferation. Its downregulation due to the ribozyme's activity should result in diminished cell growth and in attenuation of tumor progression. HERZYME was tested in a phase I dose-escalation study in patients with Her2 overexpressing cancers (subcutaneous delivery). HERZYME was well tolerated with no significant toxicity, but no biologic activity was established at the doses tested. The specific efficiency of ANGIOZYME and HERZYME needs to be further evaluated in randomized trials (with placebo or chemotherapy-alone control groups). The third hammerhead ribozyme that entered clinical trial is called HEPTAZYME. It was designed to cleave the hepatitis C RNA and reduce viral titers. The HCV virus is associated to liver cirrhosis and hepatocellular carcinoma. In a phase II clinical trial, HEPTAZYME has demonstrated a reduction in serum HCV RNA levels in 10 percent of treated patients. The Heptazyme trial had to be stopped due to some toxic side effects reported in one animal during toxicology studies. As of this writing there is no clinical evidence of the effectiveness of synthetic ribozymes in the context of any human disease.

***One of the Leading Companies: Ribozyme Pharmaceuticals (now SIRNA Therapeutics)*** The potential of synthetic ribozymes as drugs has been developed and exploited by Ribozyme Pharmaceuticals (Boulder, Colorado). The company has expertise in oligonucleotide process development and manufacturing. Ribozyme Pharmaceuticals has established strategic corporate relationships with Archemix Corporation, atugen AG, Chiron Corporation, Fujirebio, and Geron Corporation. Ribozyme Pharmaceuticals was listed on the NASDAq under "RZYM" (now "RNAI"). It has performed studies on toxicology, biodistribution, biostability, and pharmacokinetic of modified synthetic ribozymes. Human clinical trials were sponsored by Ribozyme Pharmaceutical in the context of metastatic cancer (the hammerhead ribozymes ANGIOZYME and HERZYME, which target the mRNA coding VEGF receptor and epidermal growth factor, respectively) and hepatitis C infections (the hammerhead ribozyme HEPTAZYME).

The breakthrough discovery of siRNA [67, 68] led to a reorientation of Ribozyme Pharmaceuticals' scientists toward the utilization of this new RNA tool that is very efficient in gene suppression. Ribozyme Pharmaceuticals is now SIRNA Therapeutics (since 2003, "RNAI" on the NASDAq) and focuses primarily on the development and commercialization of siRNA-based treatments. A large part of the intellectual property of SIRNA Therapeutics is based on the patents and expertise from Ribozymes Pharmaceuticals in the areas of target validation, RNA synthesis, modifications, and delivery. SIRNA Therapeutics uses this expertise to develop a new class of siRNA. The company is in research and/or preclinical development of siRNA-based drugs capable of treating diseases such as AMD, HCV infection, and cancer (see *www.sirna.com*). In the field of oncology, SIRNA Therapeutics announced a collaboration with Eli Lilly & Company (Lilly) to jointly investigate chemically modified siRNAs against specific oncology targets provided by Lilly.

## Small Inhibitory RNA (siRNA)

***Definition of siRNA*** While engineering transgenic petunia to develop a more intense purple color than the wild variety, the Jorgensen group [68a] obtained genetically modified organisms (GMO) that developed white flowers! To explain these surprising results, they hypothesized a mechanism of posttranscriptional gene silencing. Eight years later, Fire et al. described in *Caenorhabditis elegans* the potential of double-stranded RNA molecules to specifically induce degradation of homologous single-stranded endogenous mRNA in a process that they named RNA interference (RNAi [69]). At the turn of the millennium, Elbashir and co-workers [67, 68] made a breakthrough discovery: RNAi can be obtained by small (21 to 23 mer) double-stranded synthetic RNA homologous to the target sequence. These small duplex must have 2 bases overhanging at their 3′ ends. They are called siRNA (for small inhibitory RNA) [67, 68]. The generalization of these observations in all tested organisms led to a revolution in fundamental research and molecular medicine [70, 71].

The advantages of siRNA over long double-stranded RNA molecules is that they:

1. Can be synthesized chemically.
2. Diffuse in the organism.
3. Do not induce activation of the interferon response pathway [72] (activation of the protein kinase PKR and of the 2′,5′-oligoadenylate synthetase that cause inhibition of translation by phosporylation of the translation initiation factor eIF2a and degradation of mRNA, respectively).

As opposed to ribozymes, RNAi requires housekeeping proteins to perform gene suppression: the type III RNase termed Dicer and the RISC (RNA-induced silencing complex, Fig. 6). Dicer chops long double-stranded RNA into

ca. 21-mer double-stranded RNA that possess two nucleotides overhang at both strands 3′ ends. In an adenosine triphosphate (ATP)-dependent process, RISC incorporates one strand from these ca. 21-mer double-stranded RNA. With that RNA probe, activated RISC recognizes and cleaves RNA molecules that can pair with the oligonucleotide. Using this mechanism, virtually any RNA can be specifically destroyed. Eukaryotes probably developed this enzymatic system for the recognition of double-stranded RNA (characteristic of several viruses) and destruction of homologous single-stranded RNA. Accordingly, siRNA can be regarded as the natural mediators of a cell immunity dedicated to protect the genome against pathogens (viruses or transposons).

Similarly to ribozymes and antisense technologies, the identification of an optimal target site in the mRNA to be cleaved (optimal sequence of the siRNA) is a prerequisite to further development of RNAi by siRNA. Although guidelines and websites are available (see, e.g., http://jura.wi.mit.edu [73] or EMBOSS at *http://bioweb.pasteur.fr/seqanal/interfaces/sirna.html*), the characterization of the most active siRNA for the selective degradation of a target gene requires the synthesis (enzymatically or chemically) and in vitro testing of several duplexes. Each duplex is introduced into target cells (usually tumor cells that can easily be transfected) that express the gene of interest and RNAi is measured by molecular biology tools (the amount of the targeted mRNA can be quantified by RT-PCR, Northern blot, or gene microarrays) and/or by biochemistry methods (quantification of the protein translated from the target mRNA by enzymatic tests when the activity of the target protein can be directly recorded or by ELISA, Western blots, etc.).

The sequence of a siRNA will dictate its effectiveness and specificity but will also dictate its sensitivity to RNases. The half-life can vary from minutes to days. Experimental measurements are the only way to determine the siRNA's half-life in physiologic milieus before further clinical development.

As far as therapeutic interventions are concerned, RNAi technology may translate into antivirus [HIV, human papilloma virus (HPV), hepatitis B virus (HBV), HCV] or antipathology-related proteins (e.g., oncogene, immunosuppressing agents, multidrug resistance transporters) siRNA-based drugs [74]. Although a very promising approach, the use of siRNA in modern medicine faces limits that must be solved:

1. Off target gene inhibition
2. The induction of stress by the double-stranded RNA structures
3. Inefficient intracellular penetration

The first point is crucial since the specificity of siRNA may not be as high as initially expected. Study of the global gene expression in siRNA transfected cells using gene microarrays indicated that a large number of nontargeted genes are affected by the siRNA. Actually, a complementarity between the siRNA and the target mRNA of as little as 11 continuous nucleotides is sufficient to trigger recognition and cleavage by the RISC complex [75]. Besides,

only one of the two strands of the exogenous siRNA (the antisense as compared to the mRNA sequence) should theoretically be used, but both strands may practically be incorporated into the RISC, thus increasing the possible cross-recognition of irrelevant mRNA. A better knowledge of the fundamental mechanisms of RISC activity, especially in the choice of the captured siRNA strand, may allow a reduction of "off-target" genes affected by siRNA approaches.

The induction of stress was revealed by Semizarov et al. [76]. Here again, gene microarray analysis revealed the expression of several stress-related genes after delivery of siRNA, especially at high siRNA concentrations. Concerning the third point, it was shown that some siRNA can activate antigen presenting cells similarly to long double-stranded RNA [77, 78]. This phenomenon is mediated by Toll-like receptors [78a].

Altogether, the characterization of the optimal siRNA that will specifically and efficiently destroy the target mRNA with minimal off-target activity, highest stability, good biodistribution, and no side effects (immunostimulation) requires intensive experimental in vitro and in vivo testing. Several of these critical parameters are taken into account in the freely available siRNA prediction algorhythm available at *http://jura.wi.mit.edu/siRNAext/* [73], but in vivo stability, pharmacokinetics, and biodistributiuon must be tested experimentally for each individual siRNA duplex. Concerning the intracellular delivery of exogenous siRNA, refer to Delivery of RNA Molecules for Suppression of Gene Expression below.

***RNA Mimetics Suitable for siRNA's Function*** Contrary to single-stranded RNA such as ribozymes, the short double-stranded siRNA are usually remarkably stable in physiologic milieus. Nevertheless, very large variations in siRNA half-lives (from several minutes to days) can be attributed to the sequence of the two oligonucleotides. In the case where further enhancement of the half-life of the optimal siRNA sequence deduced from in vitro experiments is required, some of the chemical modifications described in the discussion of natural and modified RNA can be introduced into the therapeutic siRNA. First, deoxy nucleotides can replace the two ribonucleotides in the 3′ overhangs of each siRNA strand [67, 68]. Modifications such as allylation or alkylation of some [79] but not all [80] 2′ OH as well as 2′fluoro substitutions and LNA nucleotides [81, 82] are tolerated with minimal or no decrease of siRNA function. Similarly, phosphorothioate backbones can be used in siRNA strands [81].

Up until now, the tested chemical modifications showed a relatively low increase in the efficiency of siRNA over time in transfected cells as well as in vivo [82, 83]. Thus, chemical modifications may not be required in the context of siRNA therapies. Nevertheless, together with the incorporation of RNase resistance data into siRNA sequence prediction programs, the usage of some chemical modifications at key positions may lead to the definition of siRNA with half-lives of several days or weeks in physiologic milieus. These improvements will result in a reduced dosage and frequency of application of siRNA for optimal activity in future therapeutic settings.

***Preclinical Evaluation of siRNA*** Several siRNA with a potential therapeutic activity were characterized. For antitumor therapies were identified siRNA that target growth factors (VEGF) [83a] and antiapoptosis molecules (Bcl-2) or siRNA that targets the mRNA coding for oncogens such as the fusion protein Bcr-Abl [83b], the constitutively activated mutated Ras proteins or mutants of P53, which dominantly inhibit the tumor suppressing activity of wild type P53 in ca. 50 percent of human malignancies [83c, 83d]. In the context of infectious diseases, were characterized siRNA specific for viral genes (HIV, HCV, HBV). Although most reports are limited to the definition of the optimal siRNA sequence using in vitro testing in cell culture systems, several investigations went further to the biodistribution and efficiency studies in preclinical setups in mice. Concerning tissue distribution and pharmacokinetics, Braasch et al. [83] used radioactive iodine to label the siRNA. After intravenous injection of naked siRNA in mice, the molecule (the radioactivity) localizes in the liver and kidney (peak at 5 min postinjection). Later, a signal is found in the bladder suggesting quick elimination though urine. Lower amounts of radioactivity are found in lung, spleen, and heart. Close to no signal can be detected in the brain. Most siRNA seem to be eliminated from the mouse body in less than 24 h but some detectable amounts of radioactivity remain up to 72 h. These data suggest that (1) when a continuous inhibition of gene expression is required, injection of siRNA should be repeated every 24 h, and (2) as far as systemic delivery is concerned, siRNA-based therapies may be mostly relevant when the target gene is expressed in liver or kidneys. For delivery to other organs, encapsulating siRNA in specific delivery vehicles and/or local delivery could be envisioned. In agreement with this biodistribution data, Zender et al. [84] showed in LacZ transgenic mice that systemic application (intravenous delivery) of siRNA designed to inhibit β-galactosidase expression, induces a three- to fourfold reduction in β-galactosidase activity in liver. When using siRNA capable of cleaving the mRNA coding caspase 8 (a necessary mediator of apoptotic cell death), the authors could induce protection of hepatocytes against Fas-mediated apoptosis. In this mouse model, acute liver failure (ALF), similar to the disease observed in patients suffering from viral hepatitis, is induced by injection of activators of Fas (either antibodies against Fas or adenovirus expressing Fas-ligand). Histopathology studies demonstrated that the anti-caspase-8 siRNA could prevent Fas-induced apoptosis in hepatocytes. Song et al. [85] obtained similar protection against induced ALF by the in vivo delivery of siRNA designed to destroy the Fas-encoding mRNA. Besides, Sorensen et al. [86] also showed that systemic delivery of siRNA resulted in selective RNAi. They studied the potential of siRNA to decrease in vivo the production of tumor necrosis factor (TNF-α) (a major player in sepsis). Their study demonstrates that intraperitoneal injection of siRNA specific for the TNF-α encoding mRNA could induce a reduction of TNF-α production after exposure to lipopolysaccharide (LPS). Moreover, mice injected with these anti-TNF-α siRNA were protected against a lethal dose of LPS.

The results of in vivo studies summarized here provide "proof of concept" that open the door for pharmaceutical use of siRNA in humans. No human clinical trial results are yet reported, the technology being too new (described in 2001) for human studies to be completed. From the results obtained in the mouse models, it can be expected that new drugs with siRNA as the active component will prove to be safe and efficient in humans.

***One of the Leading Company: Alnylam Pharmaceuticals*** Three companies were created to explore and develop siRNA: Ribopharma AG in Germany (created in 2000; located in Kulmbach, Bavaria, Germany), SIRNA (created in 2003 from Ribozyme Pharmaceuticals; located in Boulder, Colorado), and Alnylam Pharmaceuticals (created in 2002, Cambridge, Massachusetts). Ribopharma AG and Alnylam Pharmaceuticals fused in July 2003 to become Alnylam Pharmaceuticals. One of the founders of Alnylam Pharmaceuticals is Thomas Tuschl (Laboratory of RNA Molecular Biology, Rockefeller University), who discovered the structure of mammalian inhibitory RNA as a researcher at the Max-Planck-Institute for Biophysical Chemistry, in Gottingen, Germany. Alnylam Pharemaceuticals owns the first patent issued on the use of short double-stranded RNAs (siRNAs) to trigger RNAi. Alnylam Pharmaceuticals develops siRNA-based therapeutics for local (brain and eye diseases such as Parkinson's disease and wet AMD, respectively) or systemic (diabetes, obesity, autoimmune diseases, and cancer diseases) applications. No human clinical trials are disclosed on the web page from Alnylam Pharmaceuticals, www.alnylam.com, but such studies can be expected to start soon.

## Delivery of RNA Molecules for Suppression of Gene Expression

Whether antisense RNA or catalytic RNA are used, the intracellular delivery of the oligonucleotides must be achieved in vivo. The resistance of cell membranes to penetration of nucleic acids limits the efficiency of such therapies. Still, naked stabilized RNA were shown to spontaneously enter cells in the context of ribozyme studies [65, 66]. In mice, a method called hydrodynamic injection consists in a fast (several seconds) intravenous delivery of a large volume (up to 1 mL) of solution containing the therapeutic RNA resulted in efficient delivery into hepatocytes [84, 85, 87, 88]. Of course, such an application method cannot be used in humans. A feasible method is the use of a delivery vehicle such as cationic liposomes [78, 89]. In mice, liposomal formulations of siRNA were associated with the clinical manifestations of the therapeutic siRNA activity [84, 86]. The problem is that most liposomal formulations are toxic. Other vehicles such as nanoparticles or microparticles may prove to be useful for efficient and safe delivery of siRNA. All these vehicles may be targeted to a precise cell type through coating with specific ligands [90]. Cell penetrating peptides (CPP) are another tool to deliver molecules intracellularly. CPP can spontaneously cross lipidic membranes and enter

alone or with an associated cargo (protein or nucleic acid) in the cytosole. They are derived from diverse proteins such as Antennapedia or HIV TAT proteins [91]. Electrostatic interaction between an siRNA and a CPP was reported to result in efficient and specific RNAi in vitro [92]. The optimal (eventually selective) and nontoxic vehicles for siRNA is undoubtly the most needed tool for testing and developing efficient siRNA-based drugs. Many experts in this field as well as the pharmaceutical industry and biotechnology companies have advanced products that may fulfill the need for efficient delivery of siRNA. The characterization of their structure (homogeneity, stability, storage), their safety and efficiency profiles in vivo (preclinical studies, toxicology tests) as well as the possibility to produce great amounts of them at high quality good manufacturing practices (GMP) are many of the critical issues that must be addressed before new siRNA-based drugs are tested in human trials.

## 4 RNA FOR PROTEIN PRODUCTION: mRNA

### Definition of mRNA

In eukaryotes, genes are transcribed in the nucleus. This process generates pre-messenger RNA (pre-mRNA). Pre-mRNA are heavily modified: A cap structure (7-methyl guanosine triphosphate) is added to the 5′ end; some parts of the pre-mRNA (introns) are spliced out; some nucleotides of the pre-mRNA are changed by a process called editing, and a poly-A tail of usually several hundred adenine residues is added (Fig. 7). Such mature mRNA are between a minimum of ca. 200 up to several tens of thousands of bases in length. They exit the nucleus through nuclear pores. Their 5′ cap structure is recognized by eukaryotic initiation factors (eIFs) and their 3′ poly-A tail is recognized by the poly-A binding protein. Interactions of the 5′ end and 3′ end riboproteic complexes induce the beginning of translation by ribosomes. First, the small subunit of the ribosome will be recruited by the eIFs. It scans the mRNA until it reaches a AUG start codon (located in a Kozak surrounding: A/GNNNATGG) [93]. At this site, the large subunit of the ribosome associates to the small subunit and translation begins. Protein elongation continues until a stop codon (AAU, UGA, or UAG) [94]. For therapeutic use, mature messenger RNA is being produced by in vitro transcription (Fig. 7) thanks to RNA polymerase: usually T7, T3, or SP6 RNA polymerase derived from bacteriophages. See Manufacturing of RNA in GMP conditions under Section 6 of "Protein Kinase Inhibitors in Drug Discovery" for more details on mRNA production. When introduced into the cytosole of a cell, such in vitro produced mature mRNA can readily be translated into proteins. The natural catabolism of mRNA guarantees that the foreign nucleic acid will be eliminated from the cell's cytosole. The foreign protein translated from the recombinant mRNA will be expressed transiently. In a natural situation (endogenous production) intracellular mRNA has a half-life ranging from minutes (mRNA coding for

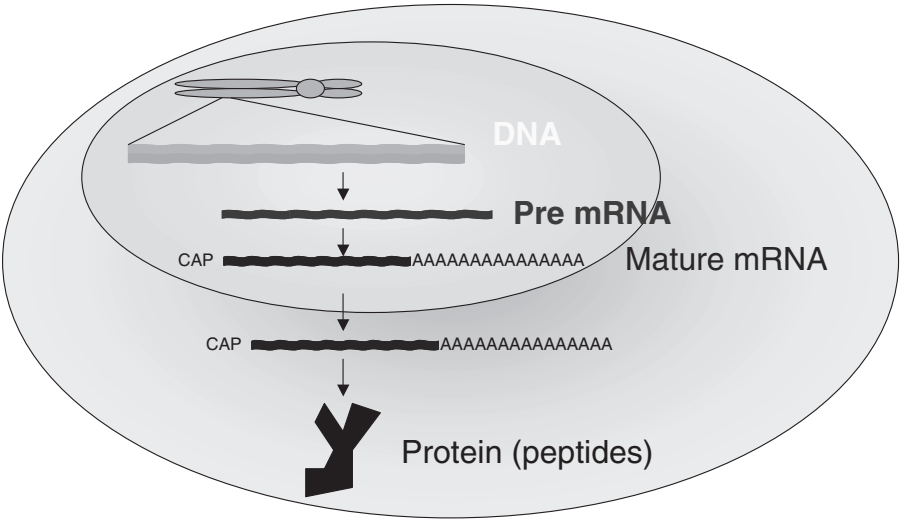


**Figure 7** Production and structure of mRNA in vivo. Genes contained in the chromosomes located in the nucleus are transcribed into pre-mRNA. This molecule is spliced (excision of introns), capped at the 5′ end and polyadenylated at the 3′ end. Its sequence may be edited. The mature mRNA leaves the nucleus through nuclear pores and its translated in proteins until its degradation by specific RNases (not shown).

proteins that control the cell cycle, e.g.) up to weeks (mRNA coding globin proteins in terminally differentiated red blood cells) [95]. Several nucleotide sequences are known to dictate the transcript's half-life [96]. Most of them are located in the 3′ untranslated region (3′ UTR) of the mRNA. For example, the A and U rich elements (AURES) signal to the cell machinery in normal conditions that the mRNA should be quickly destroyed [97]. On the other hand, pyrimidine-rich elements such as those recognized by the so-called α complex (a multiprotein structure expressed ubiquitously) allow the mRNA to be very stable [98]. The deletion of destabilization sequences such as AURES and the addition of stabilization sequences such as globin UTRs at the 3′ end of the cDNA coding for the therapeutic proteins enables in vitro production of recombinant mRNA with enhanced stability [99].

## RNA Mimetics Suitable for mRNA Function

Production of messenger RNA in vitro for clinical use depends on the activity of available purified RNA polymerases. Since only few modified NTPs are substrates for the RNA polymerases usually used for in vitro transcription (T7, SP6, and T3 RNA polymerases), the variety of chemical modifications that can be incorporated into therapeutic mRNA is very reduced in comparison to the variety of chemical modification available for synthetic RNA made by chemical synthesis. NTPs with modifications such as fluoro, azido, or amino substi-

tutions of the 2′ OH as well as NTPs with phosphorothioate modification can be used to synthesize mRNA. These, however, decrease the quality and quantity of the mRNA recovered by in vitro transcription [28] and impair translation in vivo (Probst et al., submitted). Until now, there was no NTP modifications that were shown to be permissive to RNA polymerases, to bring additional stability to the mRNA in regard of RNAse-mediated degradation and to be permissive to translation by the ribosomes in cells. Consequently, the mRNA currently used in therapies is nonmodified RNA.

## Preclinical and Clinical Utilization of mRNA

A breakthrough discovery was reported by Wolf et al. in 1990: Direct injection of naked mRNA (as well as plasmid DNA) into mouse muscle led to the production of the protein encoded by the foreign nucleic acid [100]. This indicates that somehow, and in spite of ubiquitous RNases, some cells can take up the foreign genetic information and use it for protein production in vivo. Because mRNA will be only transiently expressed, its utilization for genetic complementation (expression of a protein that is not produced from the genome because of inherited mutations) would require a frequent (daily) delivery of the drug. This aspect of mRNA-based therapies was not yet addressed since more stable vectors such as plasmid DNA or recombinant viruses seemed more appropriate. Nevertheless, safety concerns in the field of DNA-based therapies lead to the reevaluation of mRNA-based therapies for in vivo gene complementation: Theoretically, a frequent uptake of a safe "gene therapy" drug is preferable than a single uptake of a therapeutic molecule that may lead to severe side effects. Indeed, DNA or recombinant viruses may persist for some hours in some recipients and for years in others and may even cause cancer (integration of the foreign genetic information at the site of an oncogen or a tumor suppressor gene), autoimmunity (induction of anti-DNA antibodies triggering diseases such as lupus erythematosus), or modification of the human specie (integration of the foreign genetic information into the patient's germ cell genomes). Thus, a frequent application of adequately protected and delivered mRNA may become a safe alternative to DNA-based gene therapies in the field of genetic complementation.

Meanwhile, starting in the early 1990s, the potential of mRNA for transient expression of an antigen (a protein against which an immune response would be therapeutically beneficial) was investigated (for a recent review, see Pascolo [101]). First, Martinon et al. showed in mice that injection of liposome-encapsulated mRNA coding for the influenza virus nucleoprotein could trigger an anti-influenza immune response [102]. Conry et al. [102a] confirmed and extended these results by using mRNA as the active component of an antitumor immunotherapy (a vaccine whereby immune cells become competent to specifically kill tumor cells). In their experiments, mice that received intramuscular injections of naked mRNA coding for the tumor antigen carcinoembryonic antigen (CEA) were resistant to the transplantation of CEA-expressing tumor cells. In these vaccination studies, delivery of mRNA was

shown to stimulate all effector cells from the adaptive immune response: cytotoxic ($CD8^+$) T-cells capable of direct recognition and killing of target (tumor or infected) cells, B-cells that can secrete specific antibodies neutralizing the pathogenic agent, and helper ($CD4^+$) T-cells that orchestrate the immune response, enhance the activity of all immune cells, and ensure persistence of long-term memory effector cells. In 2000, two reports confirmed these results and moreover introduced tumor mRNA libraries as a tool for antitumor vaccination [102b, 102c]. In these studies, the whole tumor mRNA was injected into mice for vaccination. Such an approach could result in customized therapies where an antitumor vaccine is to be prepared from patient's tumors. Since every tumor expresses a characteristic set of immunogenic proteins specific for the tumor cells that include mutated (P53, Ras, e.g.) and recombined (Bcr-Abl, e.g.) proteins as well as the promiscuous "tumor antigens" (tissue-specific genes such as gp100, tyrosinase, or Melan-A, developmental genes, and overexpressed genes or oncogens), an autologous antitumor vaccine may be the most efficient strategy for immunotherapies. In mice, such vaccines trigger an immune response that specifically recognizes and destroys tumor cells. It prevents the dissemination of a primary tumor (blocking the formation of metastasis) and eventually results in the destruction of existing malignancies by immune cells (regression of the tumor disease). Recently, Carralot et al. showed that it is possible to tune the type of immune response triggered by mRNA-based vaccines to render it optimal for antitumor and antivirus immunotherapies [103]. In this study, granulocyte monocytes colony stimulation factor (GM-CSF), a cytokine used in the clinic for the stimulation of the growth of certain immune cells, is shown to shift the natural immunity induced by mRNA injection from a Th2 type (prone to the production of antibodies against the antigen encoded by the mRNA) to a Th1 type (prone to the proliferation of cytotoxic T-cells that can specifically kill target cells expressing the protein encoded by the mRNA vaccine). In anti-viral or antitumor immunotherapies, the mRNA plus GM-CSF form of mRNA vaccines should be used since mostly cytotoxic T-cells are required for prevention or treatment of the pathologies.

There are several other methods of mRNA-based vaccination aside from direct injection of simple mRNA. One consists in ballistic delivery where the mRNA is loaded on gold particle and delivered to the skin of patients with a gene gun [104, 105]. Cartouches containing the gold particles coated with mRNA (mRNA is precipitated on the gold microbeads by ethanol before being air-dried) are placed in the barrel of the gene gun. Pressing on the trigger while the gun is pointed to naked skin (ca. 10 cm from the skin) will result in a short release of high-pressured gas through the cartouche (the gene gun being connected to a helium bottle) that will bring the gold microbeads to the skin with high velocity. The beads will penetrate the epidermis and the dermis. Resident cells including antigen-presenting cells (APCs) such as Langerhans cells will take up the mRNA-coated beads, translate the nucleic acid, and expose the protein to the immune system to trigger an immune response. The cost of gold particles and of the gene gun as well as the difficult preclinical

setting (mice must be well shaved before being vaccinated with gene gun) have probably limited the utilization of such a promising needleless vaccination technology (no human clinical studies based on gene gun delivery of mRNA vaccines has been reported).

Another technology is the injection of autoreplicative mRNA [106–110]. Thanks to an internal ribosome entry sequence (IRES), these mRNA are bicistronic. They code for two proteins: One protein is the antigen and the other one is an RNA replicase derived from RNA viruses such as the α viruses Sindbis or Semliki Forest. After application, the bicistronic mRNA will be taken up by some cells and will be transcribed into the antigen and the replicase. Due to the presence of sequences specifically recognized by the RNA replicase in the bicistronic RNA, the replicase will multiply the foreign mRNA, thus generating a large number of proteins: antigen and replicase. This system guarantees that the antigen will be expressed for a long time and in high amounts. The uncontrolled persistence of the autoreplicative mRNA in the host and the potential of such mRNA to recombine with RNA viruses (creating new viruses) are safety concerns that may limit the use of autoreplicative mRNA in human vaccination settings.

Instead of direct intraskin delivery of mRNA for vaccination, Boczkowski and co-workers introduced a method that consists of in vitro delivery of the mRNA into autologous dendritic cells (DC) [111]. DCs are the most powerful APCs for triggering an immune response. Messenger RNA-transfected DCs produce transiently the antigen encoded by the foreign nucleic acid and present it in an optimal context to the immune system. DCs can be derived in vitro from blood using a 1-week-long cell culture protocol whereby monocytes differentiates into DCs thanks to the cytokines GM-CSF and IL-4 [112]. Afterwards, the DCs are transfected with mRNA (using either simple co-incubation or active delivery of mRNA by liposomes or electroporation) and reinfused into the organism (usually by intradermal or subcutaneous injections). Such a method triggers in mice a strong immune response against the protein encoded by the transgenic mRNA [111]. It is the most popular mRNA-based vaccine strategy with more than 40 original publications documenting its efficiency in different models (tumors and infectious disease, see review Pascolo [101]). It was transferred from the laboratory to the clinic (phase I/II studies) with the development of facilities where DCs can be cultured in GMP conditions. Tumor patients received their own in vitro generated DCs that were transfected with mRNA coding for prostate specific antigen (PSA) in patients suffering from metastatic prostate carcinoma [113], coding for CEA in patients suffering from metastatic colon cancer [114], or coding for the whole set of proteins from autologous metastasis in patients suffering from metastatic renal cell carcinoma [115]. In all three reports, the expected antitumor T-cell response could be detected in most patients and some clinical benefits were recorded. The objective efficacy of the method as an antitumor or antivirus (HIV, HCV) treatment must be evaluated in large placebo-controlled trials.

### Lead Company: MERIX Biosciences (now Argos Therapeutics)

MERIX Bioscience, Inc. (http://www.merixbio.com/index.html) was created in 1998 because of the scientific results of Prof. Gilboa. MERIX Bioscience is headquartered in Durham, North Carolina. According to its intellectual property MERIX Bioscience's therapeutic products are based on the production and utilization of mRNA-transfected dendritic cells as immunomodulators. This technology disclosed in the cornerstone article published in 1996 [111] consists of the in vitro derivation of DCs from patient's blood, transfection of such DCs by defined mRNA (coding for one tumor antigen) or mRNA libraries (derived from the patient's tumor), and reinjection of the DCs. Such a treatment was shown in mice to trigger T-cells that recognize and destroy specifically tumor cells. Three investigator-sponsored phase I/II human clinical trials indicated that the method owned by MERIX Bioscience is feasible, nontoxic and, efficient in triggering immunity against the antigens encoded by the transfected mRNA [113–115]. The impact of such a treatment on the tumor disease where clinical parameters such as "time to progression" or "time to death" are compared to control groups must be evaluated in phase IIb or III trials. MERIX Bioscience has now filed its first company-sponsored investigational new drug application with the FDA. RNA-loaded DC vaccine will be manufactured at the company's cGMP-approved vaccine manufacturing facility in Durham, North Carolina. Using mRNA-transfected DCs, MERIX Bioscience develops also treatments against infectious diseases, autoimmune disorders, and rejections of transplants.

## 5 RNA FOR IMMUNOSTIMULATION: DOUBLE-STRANDED AND STABILIZED RNA

### Definition of Immunostimulating RNAs

There exist two types of immunostimulating RNA: (1) the double-stranded RNA molecules longer than 40 base pairs and (2) the stabilized single-stranded RNA that can be as short as 20 nucleotides. In both cases, the RNA mimics structures specific from pathogens: double-stranded RNA produced during virus replication (α virus or influenza virus, e.g.) [9] or stabilized single-stranded RNA similar to those found in viruses with RNA genomes (RNA molecules condensed and stabilized by cationic compounds such as histones) [116]. Toll-like receptors (TLR) are a family of molecules conserved through evolution. They signal infection or "danger" to the immune system [117]. Mammals kept a repertoire of TLR capable of recognizing bacterial proteins (flagellin is recognized by TLR5), bacterial LPS (recognized by TLR-2 and TLR-4), bacterial and viral nonmethylated DNA (the CpG motifs recognized by TLR-9), or RNA from infectious agents: double-stranded or stabilized single-stranded RNA recognized by TLR3 or TLR-7 and TLR-8, respectively. These receptors give different activation signals that result in different effector functions. For example, double-stranded RNA (TLR-3) induces a strong

upregulation of costimulation molecules (e.g., CD86) at the surface of the responding cells (TLR-3$^+$) but low cytokine production; meanwhile stabilized single-stranded RNA (TLR-7 and -8) induces low upregulation of costimulation molecules but high cytokine secretion by responding cells. The stabilized mRNA can be mimicked by phosphorothioate RNA oligonucleotides for the stimulation of TLR-7 or -8 [118, 119]. Thus, chemically synthesized double-stranded RNA (usually a polyinosine molecule hybridized with a polycytosine nucleotide: poly I:C) or single-stranded phopshorothioate RNA as well as enzymatically produced mRNA complexed to cationic compounds are reagents available to trigger the natural mechanisms of immunostimulation.

## Preclinical and Clinical Utilization of Immunostimulating RNA

Double-stranded RNA in the form of polyinosinic-polycytidylic acid (rIn X rCn or poly I:C) were shown to be potential antitumor drugs in animal models during the 1970s [120]. When applied locally at the site of tumor transplantation in rats, they could retard or suppress tumor growth. This phenomenon is probably a consequence of local immunostimulation that results in the triggering of an antitumor immunity. Unfortunately, poly I:C showed low efficacy and high toxicity when used in humans. A derived modified compound was found to have lower toxicity and retained antitumor immunotherapeutic effect in the 1980s [121]. It consists of a mismatch-interrupted double-stranded RNA. Mismatches were introduced in the double-stranded structure using the hybridization between a polyinosinic strand and a polycitidylic strand interrupted every 12 C by uracils [rIn X r(C12,U)n]. This modified double-stranded RNA was further developed under the name Ampligen. Intravenous injections of Ampligen (10 to 80 mg per dose, twice weekly) were well tolerated in treated tumor patients. Some stabilization of the tumor disease as well as regressions were observed [122].

The immunostimulating effects and inherent anti-HIV activity [123] of Ampligen were also evaluated as therapy in acquired immunodeficiency syndrome (AIDS) patients. The AIDS patients received higher doses than the tumor patients: 200 mg intravenous twice a week. Again, no severe side effects were recorded. Virus load was decreased and immune response increased during the treatment. When used in combination with a standard anti-HIV drug (zidovudine) in a placebo-controlled trial, Ampligen (up to 700 mg intravenous twice weekly) showed a synergistic benefit: Patients receiving zidovudine with Ampligen showed a reduced tendency in CD4+ cell loss over time compared to the placebo group that received zidovidine alone [124]. Other anti-HIV drugs such as abacavir, zalcitabine, didanosine, stavudine, efavirenz, indinavir, nelfinavir, and amprenavir can also work synergistically with Ampligen in reducing HIV-related pathologies [125].

Since double-stranded RNA and stabilized RNA stimulate different receptors and consequently trigger different responses, it would be useful to compare the therapeutic activities of Ampligen to phosphorothioate oligonu-

cleotides or condensed mRNA in the context of local (injection in tumors) or systemic (adjuvant activity) applications. Meanwhile the development of DC-based therapies where DCs needs to be matured in vitro by contact to a danger signal before being reinjected offers a new therapeutic development for RNA-based adjuvant. Maturation of DCs through TLR-3, -7, or -8 may result in the triggering of the optimal maturation process that would allow the therapeutic DCs to secrete the required cytokines, express the adequate costimulation molecules, and acquire the necessary homing receptors that direct the DCs in lymph nodes. For such application, Ampligen is being tested [126].

### Lead Company: Hemispherx Biopharma, Inc.

Hemispherx Biopharma, Inc. (http://www.hemispherx.netl) is based in Philadelphia, Pennsylvania. It develops drugs for treatment of viral and immune diseases. Hemispherx Biopharma, Inc. produces Ampligen under GMP conditions and evaluates its capacity as an adjuvant for anti-HIV drugs (a phase IIb trial is ongoing in United States) or as a therapeutic against chronic fatigue syndrome (CFS).

## 6 PRODUCTION OF RNA AND LEGAL ISSUES

### Manufacturing of RNA

***Oligonucleotides (<80 Nucleotides)*** Short RNA oligonucleotides (less than ca. 80 bases) to be used as aptamers, antisense, ribozymes, siRNA, or adjuvants are produced by chemical synthesis. This process guarantees that the production is well controlled and free of any biological products. It can be upscaled to provide the amounts of oligonucleotide necessary for all phases of clinical trials and commercialization (with the existing suppliers, up to a ton of oligonucleotide per year can be produced).

Several methods of synthesis exist. Each is based on the use of building blocks that consist of nucleotides protected on reactive groups: the 5′ and 3′ groups as well as the 2′ OH (Fig. 8*a*). The most popular method of chemical synthesis is the "phosphoramidite" method [127]. The building blocks are nucleotides with an acyl protection of the amine function of the base (for A, C, and G), a dimethoxytrityl (DMT) group to protect the 5′-hydroxyl and a *t*-butyldimethylsilyl or a 2′-*O*-triisopropylsilyloxymethyl (TOM) group to protect the 2′ OH. The 3′ position has a *O*-phosphoramidite group. The chemical synthesis goes from 3′ to 5′ (reverse compared to the enzymatic synthesis). The 3′ residue of the oligonucleotide is attached on a solid support by its 3′ end. Its 5′ OH group will make a link with the 3′ phosphite group of an added building block in the presence of an appropriate activator (mostly tetrazole derivatives). The 5′ protecting group of the added nucleotide will be

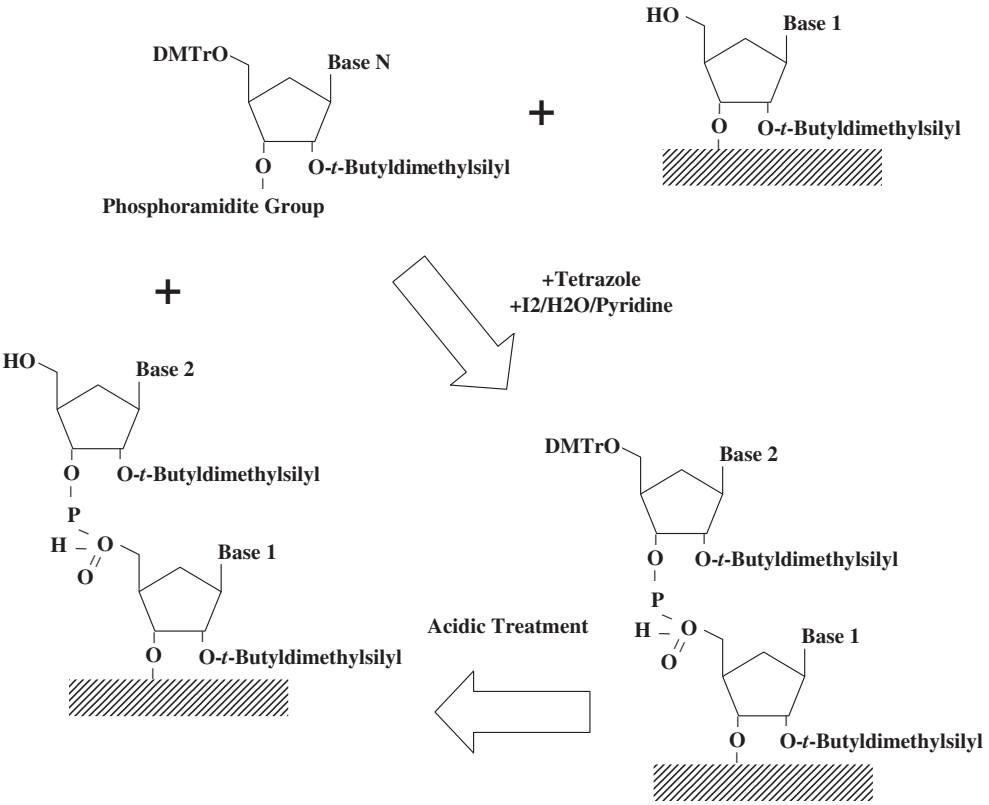


(*a*)

**Figure 8** Production of RNA. (*a*) Chemical synthesis (phosphoramidite method). The 3′ residue of the oligonucleotide is attached on a solid support by its 3′ end. Its 5′ OH group is linked with the 3′ phosphite group of an added building block in the presence of an appropriate activator (mostly tetrazole derivatives). An acidic treatment will release the DMT group and prepare the nascent oligonucleotide to receive the next base. At the end of the synthesis, the oligonucleotide is released from the solid support by basic treatment. The 2′ protection groups are then removed by a treatment with tetrabutylammonium fluoride or triethylamine trihydrofluoride. The fully deprotected oligonucleotide can be purified by HPLC. (*b*) Production of long mRNA by in vitro transcription (e.g., with the T7 RNA polymerase). The recombinant plasmid template is purified from bacteria, linearized, and in vitro transcribed to generate mRNA. The template DNA is eliminated by treatment with DNase. The final mRNA product can be purified by selective precipitation with LiCl or by anion exchange chromatography or by selection on size through the PUREmessenger technology.

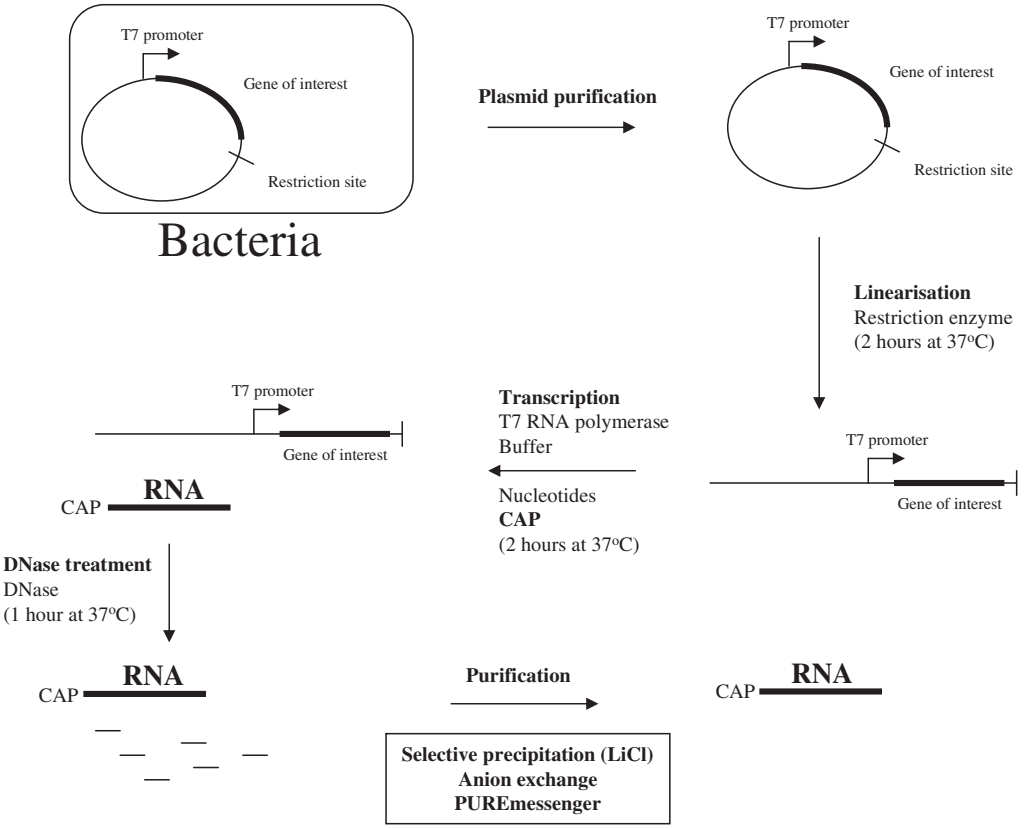


**Figure 8** *Continued*

cleaved by acidic treatment. The oligonucleotide is ready to be elongated by another nucleotide building group thanks to the same chain of reactions. The full-length oligonucleotide can be released from the solid support by basic treatment with a mixture of ammonia and methylamine (AMA), which will also deprotect the amine function of the A, C, and G residues. Finally, the soluble oligonucleotide is treated by tetrabutylammonium fluoride or triethylamine trihydrofluoride to release the 2′ protections group.

A recently described method uses a new class of silyl ethers to protect the 5′-hydroxyl in combination with an acid-labile orthoester protecting group on the 2′-hydroxyl (2′-ACE). This set of protecting groups is used with standard phosphoramidite technology [128].

The fully deprotected oligonucleotide can be purified by adequate high-performance liquid chromatography (HPLC) (anion exchange chromatography or ion pairing reversed-phase chromatography).

Many companies are offering customized synthesis of RNA oligonucleotides, but only few offer GMP-certified oligonucleotides: Avecia

(Manchester, United Kingdom; www.avecia.com) can provide hundreds of kilograms of GMP-quality, RNA oligonucleotides per year; Girindus AG (Kuensebeck, Germany; www.girindus.com), CSS (Craigavon, United Kingdom; www.css-almac.com), Lonza Ltd (Basel, Switzerland; www.lonza.com), CureVac (Tuebingen, Germany; www.curevac.com), Ambion (Austin, Texas, www.ambion.com), and Transgenomic (Omaha, Nebraska; www.transgenomic.com), among others, can also produce GMP Gligonucleotides.

***Long RNA (>80 Nucleotides)*** Messenger RNAs are always longer than 80 bases (the poly A tail being already at least 30 nucleotides). Their production requires an enzymatic synthesis. To achieve this, a genetically engineered plasmid DNA is used as a matrix. It contains a promoter recognized by a bacteriophage RNA polymerase (SP6, T7, or T3) located in front of the gene of interest. A poly-A sequence follows the gene (it should be of at least 30 A residues for the resulting mRNA to be recognized by the poly-A-binding protein necessary for translation). A unique restriction site located after the poly-A tail is used to linearize the plasmid DNA template. Through transcription this matrix generates a large amount of mRNA of a defined size (runoff transcription that stops at the site where the DNA is linearized). On the average, a DNA template is transcribed more than 100 times during a 1-h transcription reaction. The nucleotide mix in the in vitro transcription reaction contains an excess of cap nucleotide analog (N7-Me-guanosine-5′-triphosphate-5′-guanosine synthesized by chemistry) over GTP (usually a fourfold molar excess). Since the transcripts start with a G, there are statistically 80 percent of the transcripts that will start with a cap (the rest will start with a canonical G and will not be translated). DNase treatment of the transcription reaction will result in the destruction of the DNA template. The mRNA can be separated from nucleotides, enzymes, and DNA fragments by selective precipitation (e.g., with LiCl) or chromatography (e.g. anion exchange). Only two companies offer customized production of mRNA at laboratory quality or GMP quality: Ambion, (www.ambion.com) in the United States (Austin, Texas) and CureVac (www.curevac.com) in Europe (Tuebingen, Germany). These companies can generate GMP-certified batches of grams of mRNA. The homogeneity of an mRNA batch is challenged by the fact that, as is the case for oligonucleotides made by chemistry, the enzymatic production can generate shorter or longer contaminant products:

1. Short mRNA corresponds either to abortive transcriptions or to the start of transcription from a cryptic promoter (these alternative transcripts may be antisense and interfere with the biological activity of the mRNA).
2. Long mRNA that are made from traces of nonlinearized plasmids or from an upstream or antisense cryptic promoter. Here again the contaminants may affect the biological activity of the mRNA.

To solve this problem, the purification of mRNA is refined at CureVac by an innovative chromatography step called PUREmessenger. This technology not only removes all traces of contaminants (proteins, DNA fragments, and nucleotides) but also allows for the isolation of mRNA according to size. Thus shorter or longer by-products are eliminated. The so-called PUREmessenger are the highest quality in vitro transcripts available. Unexpectedly, together with their purity, their specific activity is increased. As shown in Figure 9, when transfected into eukaryotic cells, PUREmessenger mRNA is better translated than mRNA purified by classical methods such as precipitations, which do not sort mRNA on size.

## Regulatory Aspects (Quality Management)

RNA-based therapies, whether they use oligonucleotides or long messenger RNA are not classified as "gene therapies." This makes RNA-based therapies easier to implement in the clinic compared to therapeutic formulations where recombinant plasmids, viruses, or bacteria as well as DNA-transfected cells are delivered to patients.

No formal guidance from the FDA (in the United States) or the European Medicines Agency (EMEA, in Europe) is available for the production and qualification of RNA. Just one publication addresses these points and provides hints for GMP production of oligonucleotides [129]. In general, the quality of the GMP RNA must be specified by the sponsor after dialog with the producer and in agreement with the relevant authorities.

Aside from the usual final quality controls performed on any pharmaceutical ingredient (sterility, residual solvents, heavy metals, endotoxin content, pH, appearance), two particularly delicate characteristics of the final RNA batch must be measured: purity and identity. These two aspects are discussed below. Additionally, a biological activity test should be implemented when possible (e.g., testing the activity of an aptamer or of a siRNA duplex in vitro using biochemical tools or cell culture methods).

***Purity*** Measuring the purity of the RNA includes two levels of analysis that must answer two different questions. First, are there any products other than RNA in the preparation (endotoxin, microorganisms, proteins, lipids, DNA)? Second, is my RNA contaminated with some other forms of RNA: shorter, longer RNA, or RNA that does not have the expected modifications (e.g., Cap structure, 2′ substitions, phosphorothioate backbone)? The first point is easily addressed with standard tests performed on any active pharmaceutical ingredient (API). The tests are available from many companies and are compatible with the chemical nature of RNA. The second point is much more difficult to quantify. The analytical methods available will more or less detect the contaminating RNA. Because RNA molecules are complex structures, a certain amount of RNA contaminants will remain whatever purification methods are used. The apparent quality of a RNA batch depends on the tests

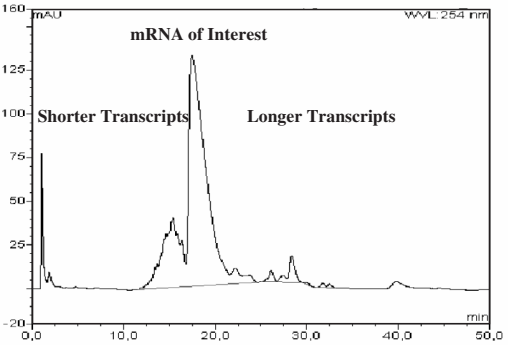


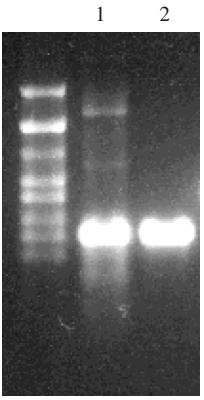


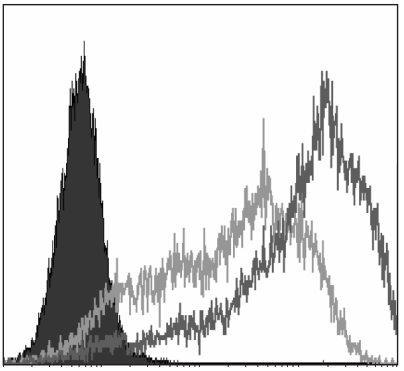


**Figure 9** PUREmessenger. In (*a*) is shown the chromatography profile of an in vitro transcript coding for EGFP (raw product). The same product is shown on a gel electrophoresis analysis in (*b*), track 1. The mRNA contained in the main peak of the chromatograph was recovered and analyzed on the gel electrophoresis shown in (*b*), track 2. This PUREmessenger mRNA is free of smaller and longer transcripts. The primary transcripts shown in track 1 and the PUREmesseger product shown in track 2 were transfected by electroporation in BHK cells. Eighteen hours later, the expression of EGFP was measured by FACS analysis. The results are shown in (*c*). The filled blue line is the negative control: Mock-transfected cells; the gray line reports the fluorescence of BHK cells transfected with the raw product; and the darker line reports the fluorescence of BHK cells transfected with the PUREmessenger product. PUREmessenger mRNA are free of any contaminants, and they are more efficient than nonpurified mRNA for protein expression.

performed, the methods used, their sensitivity, and the degree of tolerance. This qualification will follow rules mainly defined together by the customer, the provider, and the relevant legal authorities. The limits of purity should not be too stringent in order to produce large amount of therapeutic RNA for reasonable costs.

*Oligonucleotides* As far as purity of an oligonucleotide batch is concerned, mass spectrometry analysis is a gold standard. The mass of the product(s) contained in the preparation are identified and, ideally, the only detectable mass is the one of the desired RNA oligonucleotide. However, a certain amount of contaminants will always be present. HPLC preparations can remove most of the shorter or longer contaminating oligonucleotides but not all. For example, N-1 oligonucleotides are usually still detectable (N-2 and N+1 oligonucleotides can also be present). In collaboration with the authorities in charge of API in his country, the sponsor who uses the oligonucleotide in the clinic needs to set limits of impurities that can be contained in his GMP-quality oligonucleotide. Ideally, the contaminants detected by mass spectrometry should all be characterized (sequence, structure). Additional analytical methods such as polyacrylamide gel electrophoresis (PAGE) analysis, capillary electrophoresis, HPLC profile, and nuclear magnetic resonance can also be performed. The more methods, the more contaminants will be detected. For each technique, the sponsor must set limits of purity that are acceptable. Purity of up to 99 percent (HPLC profile) can be reached but requires a very stringent purification (high loss of material). Usually purities higher than 96 percent (HPLC profile) are accepted for clinical tests. This limit can be set lower for modified oligonucleotides. For example, phosphorothioate backbone modifications made during synthesis by oxidation after each coupling step are not performed at 100 percent. Consequently, some nucleotide bonds will remain phosphodiester. These contaminants are difficult to eliminate by HPLC quantitatively and will be detected by mass spectrometry. Thus, depending on the length and number of bonds that must be modified, the final oligonucleotide batch may contain more than 5 percent of oligonucleotides that miss one or several phosphorothioate bonds. Moreover, isomeric forms of the molecules coexist. For example, in a phosphorothioate oligonucleotide, a mixture of $S_p$ and $R_p$ isomers are formed (no method to chemically produce chirally pure phosphorothioate oligonucleotides is available). This isomeric heterogeneity may give complex HPLC profiles where the biologically active product is contained in several peaks. The regulatory authorities are usually well aware of these limits in the production and purification of oligonucleotides. A close collaboration between the sponsor, the manufacturer, and the authorities is necessary to fix the set of analytical methods and the degree of acceptance of impurities in the oligonucleotide preparation dedicated to clinical use.

*Messenger RNA* Since mRNA is produced with biological tools and requires DNA as a template, the first purity controls to be done on a mRNA batch

together with standard tests performed on any API (sterility, residual solvents, heavy metals, endotoxin content, pH, appearance) are the detection of contaminating DNA or proteins. The detection of DNA can be done by PCR or real-time PCR with primers specific for bacterial genomic DNA (the enzymes are produced from bacteria and the DNA matrix for transcription is usually produced from bacteria), plasmid DNA (when transcription is performed using a plasmid template), or PCR products (when the transcription is made from a PCR template, e.g., in the case of libraries). No amplicon should be detectable with 30 cycles of PCR performed on 10 ng of the mRNA batch. The detection of residual proteins can be made using a standard Bradford assay.

In addition, the purity of a defined mRNA (a single defined sequence) should be demonstrated by formaldehyde agarose gel electrophoresis analysis or ion-pair reverse-phase HPLC [130]. If a certain amount of in vitro transcribed product (more than 2 μg) is run in a gel with 0.5-cm slots, a certain amount of longer or shorter transcripts will usually be detectable [visualized by ultraviolet (UV) illumination of the ethidium bromide stained gel]. They should be characterized (purified, reverse transcribed, and sequenced) or, better, discarded during production using a purification method such as PUREmessenger from CureVac (Tuebingen, Germany). The sponsor, together with the producer and the relevant authorities, must fix the limit of detection of these longer and shorter transcript contaminants.

When mRNA libraries are needed (antitumor vaccination [115]), purity controls may be limited to the standard tests: sterility, residual solvents, heavy metals, endotoxin content, pH, appearance, together with detection of contaminating DNA and proteins.

***Identity*** The main hurdle in GMP RNA production is to prove the identity of the molecules contained in the final batch. A general identity test can be simply asking "Is there RNA in the preparation?" The answer lies in the sensitivity of the product to purified RNases. An incubation of RNA molecules with RNases should result in the complete disappearance of the nucleic acid as evidenced by the relevant analytical method (mass spectrometry, PAGE or formaldehyde/agarose gels, depending on the size of the RNA). Modified RNA that are very resistant to RNases may, however, not be suitable for such an identity test.

In addition to this general test, further precise identity tests may be required by the authorities especially in phase III clinical trials and thereafter. They are discussed bellow.

*Oligonucleotides* As far as the sequence of the oligonucleotides is concerned, mass spectrometry can be used. It gives, of course, a very good indication of product identity: The measured mass of the dominant product should be exactly the predicted mass of the oligonucleotide. It does not, however, prove the identity of the RNA: Two RNA with the same A, C, G, and U content but

different sequences will have the same mass. The sequencing of the oligonucleotide can be achieved by partial degradation of the molecule with exonucleases and mass spectrometry analysis of the resulting fragments. The mass difference between the initial oligonucleotide and the smaller degradation fragments (N-1, N-2, etc.) will indicate precisely which base is at each position. Performing this analysis separately with a 5′-exonuclease (phosphodiesterase II from bovine spleen) and a 3′-exonuclease (phosphodiesterase I from *Crotalus adamanteus*) allows confirmation of the entire oligonucleotide sequence in the batch. Again, such analysis may be difficult when the oligonucleotide contains chemical modifications that render it resistant to RNase degradation. For oligonucleotides with phosphorothioate backbones and/or DNA–RNA chimers similar enzymatic degradation methods followed by mass spectrometry analysis can be evolved to confirm the oligonucleotide's sequence and structure.

Fragmentation without enzymatic degradation can be performed by electrospray tandem mass spectrometry. The amount of fragments obtained with such a method makes the analysis difficult but may result in the confirmation of the oligonucleotide's sequence in a single experimental setup.

Manufacturers of GMP oligonucleotides have developed and qualified customized analytical methods. In conclusion, thanks to a combination of enzymatic tests and mass spectrometry measurements enabling a precise analysis of oligonucleotide's sequence and structure, it is now possible to confirm its identity. Since each oligonucleotide has a specific structure (sequence, modifications, DNA/RNA chimera, etc.), customized identity tests are required.

*Messenger RNA* Formaldehyde agarose gel electrophoresis analysis that is used to check the purity of the mRNA batch will also give information as to the identity of the mRNA. Compared to a defined RNA ladder (that must be qualified), the dominant product should have the expected size. Estimation of the size of an mRNA on a gel electrophoresis has an imprecision of ca. 10 percent of the length: a 1-kb fragment will be recognized as an RNA longer than 900 bases and shorter than 1100 bases. In this example, if the product is less than 900 bases or longer than 1100 bases, the gel electrophoresis analysis will allow the producer to recognize a problem in mRNA identity. Due to its relatively low precision, this analysis hints at the correct identity of the mRNA but provides no proof. A more precise identity test is possible after reverse transcription of the mRNA followed by PCR amplification with primer designed to recognize specifically the expected cDNA. This test will prove that at least the expected RNA sequence was present in the mRNA batch. It will, however, not demonstrate that all RNA molecules in the batch have the expected sequence. The cloning of the reverse transcribed product into a plasmid, followed by sequencing of randomly picked clones will demonstrate that virtually all RNA molecules contained in the batch have the expected sequence. This test is available but difficult to qualify at a GMP level. Never-

theless, it can be used if the regulatory authorities request an accurate identity check of the mRNA molecules contained in a GMP batch.

As far as mRNA libraries are concerned, no sequence identity test can be performed. On a formaldehyde/agarose gel electrophoresis the library should appear as a smear of fragments from 400 to 6000 bases with a maximum intensity around 1.5 kb (running 3 μg of mRNA in a 0.5-cm broad slot). Eventually, functional studies of the library can be undertaken: reverse transcription, cloning, and sequencing to confirm that most mRNA contained in the library is full-length mRNA or in vitro translation followed by Western blot analysis with antibodies specific for some housekeeping proteins to demonstrate that known proteins are translated from the mRNA batch. Again the relevance/cost ratio of these analyses must be set by the customer together with the local authorities.

A specific test necessary to qualify mRNA is the quantification of molecules that contain the 5′ cap structure. Theoretically, from the in vitro transcription setup, 80 percent of the molecules should contain the cap (usual cap:GTP ration is 4:1). Functionally, this structure is extremely important: The mRNA would not be translated into protein without the 5′ cap structure. Consequently, a test quantifying the percentage of capped molecule is required to qualify a GMP mRNA batch. To achieve this, some methods (corporate know-how, not publicly available) have been developed and validated by the two providers of GMP mRNA, Ambion (Austin, Texas) and CureVac (Tuebingen, Germany).

To conclude on the identity tests for mRNA, several methods exist to partially or accurately verify the sequence and functionality of the mRNA. Here again, the methods, the limit of sensitivity, and accuracy must be decided by the customer and the relevant authorities.

## 7 CONCLUSION AND PERSPECTIVES: RNA-BASED THERAPIES OF THE FUTURE

Aptamers, ribozymes, siRNA, immunostimulating RNA, and mRNA are as many tools that have proven in pre clinical and clinical studies to be effective against a wide variety of diseases. Production of these molecules is available at large scale and in GMP quality. The first RNA oligonucleotide that reached the market (Macugen, Eyetech) has proven the pharmaceutical potential of RNA and has paved the way for qualification and large-scale production of RNA.

Most diseases with unmet medical needs can be addressed by one or several therapies based on RNA. Cancer-related diseases, for example, may be treated by one or a combination of RNA therapies. It could consist in an antitumor vaccine (based on mRNA together with immunostimulating RNA) and/or specific inhibitors of oncogen production (ribozymes or siRNA specifically blocking the expression of growth factors) and/or aptamers that would recognize

tumor cells and induce their neutralization or destruction. Similarly, the spread of infectious diseases such as AIDS or malaria may be prevented (mRNA-based vaccines) or treated (mRNA-based vaccines plus eventually aptamers and siRNA or ribozymes) with RNA-based therapeutics.

Although the 1990s were dominated by the development of DNA-based therapies (plasmid DNA, recombinant viruses, transfected bacteria, or cells), the future clearly will rely on the safe and versatile pharmaceutical use of RNA. As the source of life 4.2 billion years ago, RNA is now regarded as a base for future therapies dedicated to improving and saving lives.

## REFERENCES


1. Palade, G. E. (1955). A small particulate component of the cytoplasm. *J. Biophys. Biochem. Cytol.*, *1*, 59–68.
2. Littlefield, J. W., Keller, E. B., Gross, J., Zamecnik, P. C. (1955). Studies on cytoplasmic ribonucleoprotein particles from the liver of the rat. *J. Biol. Chem.*, *217*, 111–123.
3. Crick, F. (1970). Central dogma of molecular biology. *Nature*, *227*, 561–563.
4. Brenner, S. J. F. M. M. (1961). An unstable intermediate carrying information from genes to ribosome for protein synthesis. *Nature*, *190*, 576–581.
5. Gros, F., Gilbert, W., Hiatt, H. H., Attardi, G., Spahr, P. F., Watson, J. D. (1961). Molecular and biological characterization of messenger RNA. *Cold Spring Harb. Symp. Quant. Biol.*, *26*, 111–132.
6. Hoagland, M. B., Stephenson, M. L., Scott, J. F., Hecht, L. I., Zamecnik, P. C. (1958). A soluble ribonucleic acid intermediate in protein synthesis. *J. Biol. Chem.*, *231*, 241–257.
7. Feldmann, H., Zachau, H. G. (1964). Chemical evidence for the 3′-linkage of amino acids to s-RNA+. *Biochem. Biophys. Res. Commun.*, *15*, 13–17.
8. Chapeville, F., Lipmann, F., Von Ehrenstein, G., Weisblum, B., Ray, W. J., Jr., Benzer, S. (1962). On the role of soluble ribonucleic acid in coding for amino acids. *Proc. Natl. Acad. Sci. USA*, *48*, 1086–1092.
9. Absher, M., Stinebring, W. R. (1969). Toxic properties of a synthetic double-stranded RNA. Endotoxin-like properties of poly I. poly C, an interferon stimulator. *Nature*, *223*, 715–717.
10. Guerrier-Takada, C., Gardiner, K., Marsh, T., Pace, N., Altman, S. (1983). The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme. *Cell*, *35*, 849–857.
11. Cech, T. R., Bass, B. L. (1986). Biological catalysis by RNA. *Annu. Rev. Biochem.*, *55*, 599–629.
12. Delihas, N. (1995). Regulation of gene expression by trans-encoded antisense RNAs. *Mol. Microbiol.*, *15*, 411–414.
13. Tuschl, T., Borkhardt, A. (2002). Small interfering RNAs: A revolutionary tool for the analysis of gene function and gene therapy. *Mol. Interv.*, *2*, 158–167.
14. Matzke, M., Matzke, A. J. (2003). RNAi extends its reach. *Science*, *301*, 1060–1061.

15. Schramke, V., Allshire, R. (2003). Hairpin RNAs and retrotransposon LTRs effect RNAi and chromatin-based gene silencing. *Science*, *301*, 1069–1074.
16. Yao, M. C., Fuller, P., Xi, X. (2003). Programmed DNA deletion as an RNA-guided system of genome defense. *Science*, *300*, 1581–1584.
17. Deutscher, M. P., (1993). Ribonuclease multiplicity, diversity, and complexity. *J. Biol. Chem.*, *268*, 13011–13014.
18. Egli, M., Gryaznov, S. M. (2000). Synthetic oligonucleotides as RNA mimetics: 2′-Modified Rnas and N3′ → > P5′ phosphoramidates. *Cell Mol. Life Sci.*, *57*, 1440–1456.
19. Koshkin, A. A., Wengel, J. (1998). Synthesis of novel 2′,3′-linked bicyclic thymine ribonucleosides. *J. Org. Chem.*, *63*, 2778–2781.
20. Wittenhagen, L. M., Kelley, S. O. (2003). Impact of disease-related mitochondrial mutations on tRNA structure and function. *Trends Biochem. Sci.*, *28*, 605–611.
21. Andrake, M., Guild, N., Hsu, T., Gold, L., Tuerk, C., Karam, J. (1988). DNA polymerase of bacteriophage T4 is an autogenous translational repressor. *Proc. Natl. Acad. Sci. USA*, *85*, 7942–7946.
22. Powers, T., Daubresse, G., Noller, H. F. (1993). Dynamics of in vitro assembly of 16 S rRNA into 30 S ribosomal subunits. *J. Mol. Biol.*, *232*, 362–374.
23. Sullenger, B. A., Gallardo, H. F., Ungers, G. E., Gilboa, E. (1990). Overexpression of TAR sequences renders cells resistant to human immunodeficiency virus replication. *Cell*, *63*, 601–608.
24. Tuerk, C., Gold, L. (1990). Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science*, *249*, 505–510.
25. Ellington, A. D., Szostak, J. W. (1990). In vitro selection of RNA molecules that bind specific ligands. *Nature*, *346*, 818–822.
26. Daniels, D. A., Chen, H., Hicke, B. J., Swiderek, K. M., Gold, L. (2003). A tenascin-C aptamer identified by tumor cell SELEX: Systematic evolution of ligands by exponential enrichment. *Proc. Natl. Acad. Sci. USA*, *100*, 15416–15421.
27. Lee, J. F., Hesselberth, J. R., Meyers, L. A., Ellington, A. D. (2004). Aptamer database. *Nucleic Acids Res.*, *32 Database issue*, D95–100.
28. Aurup, H., Williams, D. M., Eckstein, F. (1992). 2′-Fluoro- and 2′-amino-2′-deoxynucleoside 5′-triphosphates as substrates for T7 RNA polymerase. *Biochemistry*, *31*, 9636–9641.
29. Melbye, S. W., Brant, B. A., Freedberg, I. W. (1977). Epidermal nucleases. III. The ribonucleases of human epidermis. *Br. J. Dermatol.*, *97*, 355–364.
30. Lorger, M., Engstler, M., Homann, M., Goringer, H. U. (2003). Targeting the variable surface of African trypanosomes with variant surface glycoprotein-specific, serum-stable RNA aptamers. *Eukaryot. Cell*, *2*, 84–94.
31. Darfeuille, F., Arzumanov, A., Gryaznov, S., Gait, M. J., Di Primo, C., Toulme, J. J. (2002). Loop-loop interaction of HIV-1 TAR RNA with N3′ → P5′ deoxyphosphoramidate aptamers inhibits in vitro Tat-mediated transcription. *Proc. Natl. Acad. Sci. USA*, *99*, 9709–9714.
32. Darfeuille, F., Arzumanov, A., Gait, M. J., Di Primo, C., Toulme, J. J. (2002). 2′-*O*-methyl-RNA hairpins generate loop-loop complexes and selectively inhibit HIV-1 Tat-mediated transcription. *Biochemistry*, *41*, 12186–12192.

33. Darfeuille, F., Hansen, J. B., Orum, H., Di Primo, C., Toulme, J. J. (2004). LNA/DNA chimeric oligomers mimic RNA aptamers targeted to the TAR RNA element of HIV-1. *Nucleic Acids Res.*, *32*, 3101–3107.
34. Hicke, B. J., Stephens, A. W. (2000). Escort aptamers: A delivery service for diagnosis and therapy. *J. Clin. Invest.*, *106*, 923–928.
35. White, R. R., Sullenger, B. A., Rusconi, C. P. (2000). Developing aptamers into therapeutics. *J. Clin. Invest.*, *106*, 929–934.
36. Sayer, N., Ibrahim, J., Turner, K., Tahiri-Alaoui, A., James, W. (2002). Structural characterization of a 2′ F-RNA aptamer that binds a HIV-1 SU glycoprotein, gp120. *Biochem. Biophys. Res. Commun.*, *293*, 924–931.
37. Bock, L. C., Griffin, L. C., Latham, J. A., Vermaas, E. H., Toole, J. J. (1992). Selection of single-stranded DNA molecules that bind and inhibit human thrombin. *Nature*, *355*, 564–566.
38. Santulli-Marotto, S., Nair, S. K., Rusconi, C., Sullenger, B., Gilboa, E. (2003). Multivalent RNA aptamers that inhibit CTLA-4 and enhance tumor immunity. *Cancer Res.*, *63*, 7483–7489.
39. Homann, M., Goringer, H. U. (1999). Combinatorial selection of high affinity RNA ligands to live African trypanosomes. *Nucleic Acids Res.*, *27*, 2006–2014.
40. Sayer, N. M., Cubin, M., Rhie, A., Bullock, M., Tahiri-Alaoui, A., James, W. (2004). Structural determinants of conformationally selective, prion-binding aptamers. *J. Biol. Chem.*, *279*, 13102–13109.
41. Ruckman, J., Green, L. S., Beeson, J., Waugh, S., Gillette, W. L., Henninger, D. D., Claesson-Welsh, L., Janjic, N. (1998). 2′-Fluoropyrimidine RNA-based aptamers to the 165-amino acid form of vascular endothelial growth factor (VEGF165). Inhibition of receptor binding and VEGF-induced vascular permeability through interactions requiring the exon 7-encoded domain. *J. Biol. Chem.*, *273*, 20556–20567.
42. Drolet, D. W., Nelson, J., Tucker, C. E., Zack, P. M., Nixon, K., Bolin, R., Judkins, M. B., Farmer, J. A., Wolf, J. L., Gill, S. C., Bendele, R. A. (2000). Pharmacokinetics and safety of an anti-vascular endothelial growth factor aptamer (NX1838) following injection into the vitreous humor of rhesus monkeys. *Pharm. Res.*, *17*, 1503–1510.
43. (2003). Anti-vascular endothelial growth factor therapy for subfoveal choroidal neovascularization secondary to age-related macular degeneration: Phase II study results. *Ophthalmology*, *110*, 979–986.
44. Csaky, K. (2003). Anti-vascular endothelial growth factor therapy for neovascular age-related macular degeneration: Promises and pitfalls. *Ophthalmology*, *110*, 879–881.
45. Pestka, S., Daugherty, B. L., Jung, V., Hotta, K., Pestka, R. K. (1984). Anti-mRNA: specific inhibition of translation of single mRNA molecules. *Proc. Natl. Acad. Sci. USA*, *81*, 7525–7528.
46. Coleman, J., Green, P. J., Inouye, M. (1992). The use of RNAs complementary to specific mRNAs to regulate the expression of individual bacterial genes. *Biotechnology*, *24*, 253–260.

47. Izant, J. G., Weintraub, H. (1985). Constitutive and conditional suppression of exogenous and endogenous genes by anti-sense RNA. *Science*, *229*, 345–352.
48. Kurreck, J. (2003). Antisense technologies. Improvement through novel chemical modifications. *Eur. J. Biochem.*, *270*, 1628–1644.
49. Kruger, K., Grabowski, P. J., Zaug, A. J., Sands, J., Gottschling, D. E., Cech, T. R. (1982). Self-splicing RNA: Autoexcision and autocyclization of the ribosomal RNA intervening sequence of Tetrahymena. *Cell*, *31*, 147–157.
50. Doudna, J. A., Cech, T. R. (2002). The chemical repertoire of natural ribozymes. *Nature*, *418*, 222–228.
51. Unrau, P. J., Bartel, D. P. (1998). RNA-catalysed nucleotide synthesis. *Nature*, *395*, 260–263.
52. Lohse, P. A., Szostak, J. W. (1996). Ribozyme-catalysed amino-acid transfer reactions. *Nature*, *381*, 442–444.
53. Haseloff, J., Gerlach, W. L. (1988). Simple RNA enzymes with new and highly specific endoribonuclease activities. *Nature*, *334*, 585–591.
54. Eckstein, F. (1996). The hammerhead ribozyme. *Biochem. Soc. Trans.*, *24*, 601–604.
55. Anderson, P., Monforte, J., Tritz, R., Nesbitt, S., Hearst, J., Hampel, A. (1994). Mutagenesis of the hairpin ribozyme. *Nucleic Acids Res.*, *22*, 1096–1100.
56. Earnshaw, D. J., Gait, M. J. (1997). Progress toward the structure and therapeutic use of the hairpin ribozyme. *Antisense Nucleic Acid Drug Devel.*, *7*, 403–411.
57. Beigelman, L., McSwiggen, J. A., Draper, K. G., Gonzalez, C., Jensen, K., Karpeisky, A. M., Modak, A. S., Matulic-Adamic, J., DiRenzo, A. B., Haeberli, P. (1995). Chemical modification of hammerhead ribozymes. Catalytic activity and nuclease resistance. *J. Biol. Chem.*, *270*, 25702–25708.
58. Usman, N., Blatt, L. M. (2000). Nuclease-resistant synthetic ribozymes: Developing a new class of therapeutics. *J. Clin. Invest.*, *106*, 1197–1202.
59. Pieken, W. A., Olsen, D. B., Aurup, H., Williams, D. M., Heidenreich, O., Benseler, F., Eckstein, F. (1991). Structure-function relationship of hammerhead ribozymes as probed by 2′-modifications. *Nucleic Acids Symp. Ser.*, 51–53.
60. Scherr, M., Rossi, J. J., Sczakiel, G., Patzel, V. (2000). RNA accessibility prediction: A theoretical approach is consistent with experimental studies in cell extracts. *Nucleic Acids Res.*, *28*, 2455–2461.
61. Bramlage, B., Luzi, E., Eckstein, F. (2000). HIV-1 LTR as a target for synthetic ribozyme-mediated inhibition of gene expression: Site selection and inhibition in cell culture. *Nucleic Acids Res.*, *28*, 4059–4067.
62. Pierce, M. L., Ruffner, D. E. (1998). Construction of a directed hammerhead ribozyme library: Towards the identification of optimal target sites for antisense-mediated gene inhibition. *Nucleic Acids Res.*, *26*, 5093–5101.
63. Rhoades, K., Wong-Staal, F. (2003). Inverse genomics as a powerful tool to identify novel targets for the treatment of neurodegenerative diseases. *Mech. Ageing Devel.*, *124*, 125–132.
64. Flory, C. M., Pavco, P. A., Jarvis, T. C., Lesch, M. E., Wincott, F. E., Beigelman, L., Hunt, S. W., III, Schrier, D. J. (1996). Nuclease-resistant ribozymes decrease stromelysin mRNA levels in rabbit synovium following exogenous delivery to the knee joint. *Proc. Natl. Acad. Sci. USA*, *93*, 754–758.

65. Pavco, P. A., Bouhana, K. S., Gallegos, A. M., Agrawal, A., Blanchard, K. S., Grimm, S. L., Jensen, K. L., Andrews, L. E., Wincott, F. E., Pitot, P. A., Tressler, R. J., Cushman, C., Reynolds, M. A., Parry, T. J. (2000). Antitumor and antimetastatic activity of ribozymes targeting the messenger RNA of vascular endothelial growth factor receptors. *Clin. Cancer Res.*, *6*, 2094–2103.

66. Sandberg, J. A., Bouhana, K. S., Gallegos, A. M., Agrawal, A. B., Grimm, S. L., Wincott, F. E., Reynolds, M. A., Pavco, P. A., Parry, T. J. (1999). Pharmacokinetics of an antiangiogenic ribozyme (ANGIOZYME) in the mouse. *Antisense Nucleic Acid Drug Devel.*, *9*, 271–277.

67. Elbashir, S. M., Harborth, J., Lendeckel, W., Yalcin, A., Weber, K., Tuschl, T. (2001). Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. *Nature*, *411*, 494–498.

68. Elbashir, S. M., Lendeckel, W., Tuschl, T. (2001). RNA interference is mediated by 21- and 22-nucleotide RNAs. *Genes Devel.*, *15*, 188–200.

68a. Jorgensen, R. (1990). Altered gene expression in plants due to trans interactions between homologous genes. *Trends Biotech*, *8*, 340–344.

69. Fire, A., Xu, S., Montgomery, M. K., Kostas, S. A., Driver, S. E., Mello, C. C. (1998). Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature*, *391*, 806–811.

70. Caplen, N. J., Parrish, S., Imani, F., Fire, A., Morgan, R. A. (2001). Specific inhibition of gene expression by small double-stranded RNAs in invertebrate and vertebrate systems. *Proc. Natl. Acad. Sci. USA*, *98*, 9742–9747.

71. Couzin, J. (2002). Breakthrough of the year. Small RNAs make big splash. *Science*, *298*, 2296–2297.

72. Stark, G. R., Kerr, I. M., Williams, B. R., Silverman, R. H., Schreiber, R. D. (1998). How cells respond to interferons. *Annu. Rev. Biochem.*, *67*, 227–264.

73. Yuan, B., Latek, R., Hossbach, M., Tuschl, T., Lewitter, F. (2004). siRNA Selection Server: An automated siRNA oligonucleotide prediction server. *Nucleic Acids Res.*, *32*, W130–W134.

74. Sioud, M. (2004). Therapeutic siRNAs. *Trends Pharmacol. Sci.*, *25*, 22–28.

75. Jackson, A. L., Bartz, S. R., Schelter, J., Kobayashi, S. V., Burchard, J., Mao, M., Li, B., Cavet, G., Linsley, P. S. (2003). Expression profiling reveals off-target gene regulation by RNAi. *Nat. Biotechnol.*, *21*, 635–637.

76. Semizarov, D., Frost, L., Sarthy, A., Kroeger, P., Halbert, D. N., Fesik, S. W. (2003). Specificity of short interfering RNA determined through gene expression signatures. *Proc. Natl. Acad. Sci. USA*, *100*, 6347–6352.

77. Sioud, M., Sorensen, D. R. (2003). Cationic liposome-mediated delivery of siRNAs in adult mice. *Biochem. Biophys. Res. Commun.*, *312*, 1220–1225.

78. Sioud, M., Sorensen, D. R. (2004). Systemic delivery of synthetic siRNAs. *Methods Mol. Biol.*, *252*, 515–522.

78a. Hornung, V., Guenthner-Biller, M., Bourquin, C., Ablasser, A., Schlee, M., Uematsu, S., Noronha, A., Manoharan, M., Akira, S., de Fougerolles, A., Endres, S., Hartmann, G. (2005). Sequence-specific potent induction of IFN-alpha by short interfering RNA in plasmacytoid dendritic cells through TLR7. *Nat. Med.*, *11*, 263–270.

79. Amarzguioui, M., Holen, T., Babaie, E., Prydz, H. (2003). Tolerance for mutations and chemical modifications in a siRNA. *Nucleic Acids Res.*, *31*, 589–595.

80. Elbashir, S. M., Martinez, J., Patkaniowska, A., Lendeckel, W., Tuschl, T. (2001). Functional anatomy of siRNAs for mediating efficient RNAi in *Drosophila melanogaster* embryo lysate. *EMBO J.*, *20*, 6877–6888.

81. Braasch, D. A., Jensen, S., Liu, Y., Kaur, K., Arar, K., White, M. A., Corey, D. R. (2003). RNA interference in mammalian cells by chemically-modified RNA. *Biochemistry*, *42*, 7967–7975.

82. Layzer, J. M., McCaffrey, A. P., Tanner, A. K., Huang, Z., Kay, M. A., Sullenger, B. A. (2004). In vivo activity of nuclease-resistant siRNAs. *RNA*, *10*, 766–771.

83. Braasch, D. A., Paroo, Z., Constantinescu, A., Ren, G., Oz, O. K., Mason, R. P., Corey, D. R. (2004). Biodistribution of phosphodiester and phosphorothioate siRNA. *Bioorg. Med. Chem. Lett.*, *14*, 1139–1143.

83a. Zhang, et al. (2003). Vector based RNAi, a novel tool for isoform-specific knockdown of VEGF and anti-angiogenesis gene therapy of cancer. *Biochem. Biophys. Res. Commun*. *303*, 1169–1178.

84b. Scherr, et al. (2003). Specific inhibition of Bcr-abl gene expression by small interfering RNA. *Blood*, *101*, 1566–1569.

84c. Bullock and Fersht (2001). Rescuing the function of mutant p53. *Nat. Rev. Cancer*, *1*, 68–76.

84d. Martinez (2002). Synthetic small inhibiting RNAs: Efficient tools to inactivate oncogenic mutations and restore p53 pathways. *Proc. Natl. Acad. Sci. USA*, *99*, 14849–14854.

84. Zender, L., Hutker, S., Liedtke, C., Tillmann, H. L., Zender, S., Mundt, B., Waltemathe, M., Gosling, T., Flemming, P., Malek, N. P., Trautwein, C., Manns, M. P., Kuhnel, F., Kubicka, S. (2003). Caspase 8 small interfering RNA prevents acute liver failure in mice. *Proc. Natl. Acad. Sci. USA*, *100*, 7797–7802.

85. Song, E., Lee, S. K., Wang, J., Ince, N., Ouyang, N., Min, J., Chen, J., Shankar, P., Lieberman, J. (2003). RNA interference targeting Fas protects mice from fulminant hepatitis. *Nat. Med.*, *9*, 347–351.

86. Sorensen, D. R., Leirdal, M., Sioud, M. (2003). Gene silencing by systemic delivery of synthetic siRNAs in adult mice. *J. Mol. Biol.*, *327*, 761–766.

87. Lewis, D. L., Hagstrom, J. E., Loomis, A. G., Wolff, J. A., Herweijer, H. (2002). Efficient delivery of siRNA for inhibition of gene expression in postnatal mice. *Nat. Genet.*, *32*, 107–108.

88. McCaffrey, A. P., Meuse, L., Pham, T. T., Conklin, D. S., Hannon, G. J., Kay, M. A. (2002). RNA interference in adult mice. *Nature*, *418*, 38–39.

89. Templeton, N. S. (2002). Cationic liposome-mediated gene delivery in vivo. *Biosci. Rep.*, *22*, 283–295.

90. Liang, L., Liu, D. P., Liang, C. C. (2002). Optimizing the delivery systems of chimeric RNA. DNA oligonucleotides. *Eur. J. Biochem.*, *269*, 5753–5758.

91. Jarver, P., Langel, U. (2004). The use of cell-penetrating peptides as a tool for gene regulation. *Drug Discov. Today*, *9*, 395–402.

92. Simeoni, F., Morris, M. C., Heitz, F., Divita, G. (2003). Insight into the mechanism of the peptide-based gene delivery system MPG: Implications for delivery of siRNA into mammalian cells. *Nucleic Acids Res.*, *31*, 2717–2724.
93. Kozak, M. (1978). How do eucaryotic ribosomes select initiation regions in messenger RNA? *Cell*, *15*, 1109–1123.
94. Kozak, M. (1999). Initiation of translation in prokaryotes and eukaryotes. *Gene*, *234*, 187–208.
95. Ross, J. (1995). mRNA stability in mammalian cells. *Microbiol. Rev.*, *59*, 423–450.
96. Wilusz, C. J., Wormington, M., Peltz, S. W. (2001). The cap-to-tail guide to mRNA turnover. *Nat. Rev. Mol. Cell Biol.*, *2*, 237–246.
97. Chen, C. Y., Shyu, A. B. (1995). AU-rich elements: Characterization and importance in mRNA degradation. *Trends Biochem. Sci.*, *20*, 465–470.
98. Holcik, M., Liebhaber, S. A. (1997). Four highly stable eukaryotic mRNAs assemble 3′ untranslated region RNA-protein complexes sharing cis and trans components. *Proc. Natl. Acad. Sci. USA*, *94*, 2410–2414.
99. Malone, R. W., Felgner, P. L., Verma, I. M. (1989). Cationic liposome-mediated RNA transfection. *Proc. Natl. Acad. Sci. USA*, *86*, 6077–6081.
100. Wolff, J. A., Malone, R. W., Williams, P., Chong, W., Acsadi, G., Jani, A., Felgner, P. L. (1990). Direct gene transfer into mouse muscle in vivo. *Science*, *247*, 1465–1468.
101. Pascolo, S. (2004). Messenger RNA-based vaccines. *Exp. Opin. Biol. Ther.*, *4*, 1285–1294.
102. Martinon, F., Krishnan, S., Lenzen, G., Magne, R., Gomard, E., Guillet, J. G., Levy, J. P., Meulien, P. (1993). Induction of virus-specific cytotoxic T lymphocytes in vivo by liposome-entrapped mRNA. *Eur. J. Immunol.*, *23*, 1719–1722.
102a. Conry, R. M., LoBuglio, A. F., Wright, M., Sumerel, L., Pike, M. J., Johanning, F., Benjamin, R., Lu, D., Curiel, D. T. (1995). Characterization of a messenger RNA polynucleotide vaccine vector. *Cancer Res.*, *55*, 1397–400.
102b. Hoerr, I., Obst, R., Rammensee, H. G., Jung, G. (2000). In vivo application of RNA leads to induction of specific cytotoxic T lymphocytes and antibodies. *Eur. J. Immunol.*, *30*, 1–7.
102c. Granstein, R. D., Ding, W., Ozawa, H. (2000). Induction of anti-tumor immunity with epidermal cells pulsed with tumor-derived RNA or intradermal administration of RNA. *J. Invest. Dermatol.*, *114*, 632–636.
103. Carralot, J. P., Probst, J., Hoerr, I., Scheel, B., Teufel, R., Jung, G., Rammensee, H., Pascolo, S. (2004). Polarization of the immunity induced by direct injection of naked globin UTR-stabilized mRNA vaccines. *Cell Mol. Life Sci.*
104. Qiu, P., Ziegelhoffer, P., Sun, J., Yang, N. S. (1996). Gene gun delivery of mRNA in situ results in efficient transgene expression and genetic immunization. *Gene Ther.*, *3*, 262–268.
105. Vassilev, V. B., Gil, L. H., Donis, R. O. (2001). Microparticle-mediated RNA immunization against bovine viral diarrhea virus. *Vaccine*, *19*, 2012–2019.
106. Mandl, C. W., Aberle, J. H., Aberle, S. W., Holzmann, H., Allison, S. L., Heinz, F. X. (1998). In vitro-synthesized infectious RNA as an attenuated live vaccine in a flavivirus model. *Nat. Med.*, *4*, 1438–1440.

107. Racanelli, V., Behrens, S. E., Aliberti, J., Rehermann, B. (2004). Dendritic cells transfected with cytopathic self-replicating RNA induce crosspriming of CD8(+) T cells and antiviral immunity. *Immunity*, *20*, 47–58.

108. Schirmacher, V., Forg, P., Dalemans, W., Chlichlia, K., Zeng, Y., Fournier, P., von Hoegen, P. (2000). Intra-pinna anti-tumor vaccination with self-replicating infectious RNA or with DNA encoding a model tumor antigen and a cytokine. *Gene Ther.*, *7*, 1137–1147.

109. Ying, H., Zaks, T. Z., Wang, R. F., Irvine, K. R., Kammula, U. S., Marincola, F. M., Leitner, W. W., Restifo, N. P. (1999). Cancer therapy using a self-replicating RNA vaccine. *Nat. Med.*, *5*, 823–827.

110. Zhou, X., Berglund, P., Rhodes, G., Parker, S. E., Jondal, M., Liljestrom, P. (1994). Self-replicating Semliki Forest virus RNA as recombinant vaccine. *Vaccine*, *12*, 1510–1514.

111. Boczkowski, D., Nair, S. K., Snyder, D., Gilboa, E. (1996). Dendritic cells pulsed with RNA are potent antigen-presenting cells in vitro and in vivo. *J. Exp. Med.*, *184*, 465–472.

112. Sallusto, F., Lanzavecchia, A. (1994). Efficient presentation of soluble antigen by cultured human dendritic cells is maintained by granulocyte/macrophage colony-stimulating factor plus interleukin 4 and downregulated by tumor necrosis factor alpha. *J. Exp. Med.*, *179*, 1109–1118.

113. Heiser, A., Coleman, D., Dannull, J., Yancey, D., Maurice, M. A., Lallas, C. D., Dahm, P., Niedzwiecki, D., Gilboa, E., Vieweg, J. (2002). Autologous dendritic cells transfected with prostate-specific antigen RNA stimulate CTL responses against metastatic prostate tumors. *J. Clin. Invest.*, *109*, 409–417.

114. Morse, M. A., Nair, S. K., Mosca, P. J., Hobeika, A. C., Clay, T. M., Deng, Y., Boczkowski, D., Proia, A., Neidzwiecki, D., Clavien, P. A., Hurwitz, H. I., Schlom, J., Gilboa, E., Lyerly, H. K. (2003). Immunotherapy with autologous, human dendritic cells transfected with carcinoembryonic antigen mRNA. *Cancer Invest.*, *21*, 341–349.

115. Su, Z., Dannull, J., Heiser, A., Yancey, D., Pruitt, S., Madden, J., Coleman, D., Niedzwiecki, D., Gilboa, E., Vieweg, J. (2003). Immunological and clinical responses in metastatic renal cancer patients vaccinated with tumor RNA-transfected dendritic cells. *Cancer Res.*, *63*, 2127–2133.

116. Diebold, S. S., Kaisho, T., Hemmi, H., Akira, S., Reis E. S. (2004). Innate antiviral responses by means of TLR7-mediated recognition of single-stranded RNA. *Science*, *303*, 1529–1531.

117. Sieling, P. A., Modlin, R. L. (2002). Toll-like receptors: Mammalian "taste receptors" for a smorgasbord of microbial invaders. *Curr. Opin. Microbiol.*, *5*, 70–75.

118. Heil, F., Hemmi, H., Hochrein, H., Ampenberger, F., Kirschning, C., Akira, S., Lipford, G., Wagner, H., Bauer, S. (2004). Species-specific recognition of single-stranded RNA via toll-like receptor 7 and 8. *Science*, *303*, 1526–1529.

119. Scheel, B., Braedel, S., Probst, J., Carralot, J. P., Wagner, H., Schild, H., Jung, G., Rammensee, H. G., Pascolo, S. (2004). Immunostimulating capacities of stabilized RNA molecules. *Eur. J. Immunol.*, *34*, 537–547.

120. Pimm, M. V., Embleton, M. J., Baldwin, R. W. (1976). Treatment of transplanted rat tumours with double-stranded RNA (BRL 5907). I. Influence of systemic and local administration. *Br. J. Cancer*, *33*, 154–165.
121. Carter, W. A., Strayer, D. R., Hubbell, H. R., Brodsky, I. (1985). Preclinical studies with Ampligen (mismatched double-stranded RNA). *J. Biol. Response Mod.*, *4*, 495–502.
122. Brodsky, I., Strayer, D. R., Krueger, L. J., Carter, W. A. (1985). Clinical studies with ampligen (mismatched double-stranded RNA). *J. Biol. Response Mod.*, *4*, 669–675.
123. Schroder, H. C., Kelve, M., Schacke, H., Pfleiderer, W., Charubala, R., Suhadolnik, R. J., Muller, W. E. (1994). Inhibition of DNA topoisomerase I activity by 2′, 5′-oligoadenylates and mismatched double-stranded RNA in uninfected and HIV-1-infected H9 cells. *Chem. Biol. Interact.*, *90*, 169–183.
124. Thompson, K. A., Strayer, D. R., Salvato, P. D., Thompson, C. E., Klimas, N., Molavi, A., Hamill, A. K., Zheng, Z., Ventura, D., Carter, W. A. (1996). Results of a double-blind placebo-controlled study of the double-stranded RNA drug polyI:polyC12U in the treatment of HIV infection. *Eur. J. Clin. Microbiol. Infect. Dis.*, *15*, 580–587.
125. Essey, R. J., McDougall, B. R., Robinson, W. E., Jr. (2001). Mismatched double-stranded RNA (polyI-polyC(12)U) is synergistic with multiple anti-HIV drugs and is active against drug-sensitive and drug-resistant HIV-1 in vitro. *Antiviral Res.*, *51*, 189–202.
126. Adams, M., Navabi, H., Jasani, B., Man, S., Fiander, A., Evans, A. S., Donninger, C., Mason, M. (2003). Dendritic cell (DC) based therapy for cervical cancer: Use of DC pulsed with tumour lysate and matured with a novel synthetic clinically non-toxic double stranded RNA analogue poly [I]:poly [C(12)U] (Ampligen R). *Vaccine*, *21*, 787–790.
127. Scaringe, S. A., Francklyn, C., Usman, N. (1990). Chemical synthesis of biologically active oligoribonucleotides using beta-cyanoethyl protected ribonucleoside phosphoramidites. *Nucleic Acids Res.*, *18*, 5433–5441.
128. Scaringe, S. A., Wincott, F. E., Caruthers, M. H. (1998). Novel RNA synthesis method using 5′-silyl-2′-orthoester protecting groups. *J. Am. Chem. Soc.*, *120*, 11820–11821.
129. Kambhampati, R. V., Chiu, Y. Y., Chen, C. W., Blumenstein, J. J. (1993). Regulatory concerns for the chemistry, manufacturing, and controls of oligonucleotide therapeutics for use in clinical studies. *Antisense Res. Devel.*, *3*, 405–410.
130. Azarani, A., Hecker, K. H. (2001). RNA analysis by ion-pair reversed-phase high performance liquid chromatography. *Nucleic Acids Res.*, *29*, E7.

# NOVEL IMAGING AGENTS FOR MOLECULAR MR IMAGING OF CANCER

DMITRI ARTEMOV AND ZAVER M. BHUJWALLA
*The Johns Hopkins University School of Medicine*
*Baltimore, Maryland*

1 INTRODUCTION
2 CONTRAST AGENTS AND MECHANISM OF CONTRAST ENHANCEMENT IN MRI
T1 Contrast Agents
T2 Contrast Agents
Relaxation Properties of Main Classes of CA
3 MRI OF TUMOR VASCULATURE AND VASCULAR TARGETS
Low-Molecular-Weight Contrast Agents (GdDTPA)
High-Molecular-Weight Contrast Agents
Molecular Imaging of Tumor Neovasculature
4 MRI OF TUMOR CELL SURFACE RECEPTORS
MR Molecular Imaging of Isolated Cells
In Vivo MR Molecular Imaging of Receptors in Cancer
5 MRI OF INTRACELLULAR TARGETS
Contrast Agents for Long-Term Labeling of Target Cells (MRI Cell Trafficking)
Specific Contrast Agents for Intracellular Targets
6 MOLECULAR IMAGING WITH ACTIVATED MR CONTRAST AGENTS
Enzymatic Activation of CA
CEST and PARACEST Contrast Agents
7 MOLECULAR MRI AND MR SPECTROSCOPY
CSI Molecular Imaging with Reporter Molecules

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

8 DISCUSSION
References

## 1 INTRODUCTION

The response of solid tumors to therapy is traditionally measured by changes in tumor volume over the course of treatment. Because of its high spatial resolution and contrast in soft tissues, magnetic resonance imaging (MRI) provides an informative noninvasive approach to measuring tumor volumes. Additionally, nonspecific low-molecular-weight paramagnetic contrast agents (such as GdDTPA) can be used to derive perfusion characteristics of the lesion.

Novel therapeutic strategies often deviate from traditional cytotoxic chemotherapy as they do not directly kill tumor cells. As a result, tumor volume alone is no longer the ultimate response parameter. For example, one novel paradigm in tumor treatment is antiangiogenic therapy that targets tumor neovasculature. In this case treatment response can be best assessed from changes in tumor vascular parameters. Standard morphological MRI cannot detect these functional physiological parameters, and the application of dynamic MRI with appropriate contrast agents is required to determine the status of tumor vascularization. For example, dynamic MRI using a high-molecular-weight contrast agent enables quantitative measurements of tumor blood volume and vascular permeability surface area product and has been demonstrated to detect tumor response to antiangiogenic therapy [1]. Changes in the concentration of tumor metabolites detected by magnetic resonance spectroscopy (MRS) can also be an early marker of tumor response to conventional chemotherapy prior to changes in tumor volume [2, 3]. In vivo MRS might be a method of choice to detect early apoptotic events [4, 5] as well as tumor response to noncytotoxic (cytostatic) chemotherapy. In the latter case the lack of the tumor response to the therapy can be promptly addressed by switching to a more aggressive form of treatment.

Target-specific tumor therapy (based, e.g., on humanized monoclonal antibodies) is another important development in cancer treatment. The identification of a subpopulation of tumors that express therapeutic targets (such as cell surface receptors) and the monitoring of their status during treatment can help to optimize the therapeutic regimen and to improve cure. MRI with highly specific contrast agents (CA) can help to image these targets noninvasively with high spatial and temporal resolution. In the next section we present a brief overview of available contrast agents and the general mechanisms of contrast enhancement in MRI.

## 2 CONTRAST AGENTS AND MECHANISMS OF CONTRAST ENHANCEMENT IN MRI

The intensity of MR images typically depends upon the concentration of resonating nuclei (protons) in the sample and on their relaxation parameters. Relaxation processes that contribute to the measured signal can be described as T1, T2, and T2* relaxation [6]. Relaxation contrast agents that modify these relaxation rates are classified as "T1 agents" or "T2 (T2*) agents," respectively. T1 contrast agents provide a positive contrast or increased intensities in T1-weighted MR images. T2 or T2* contrast agents provide a negative contrast or reduced signals in T2 or T2* weighted MR images, respectively. To simplify our discussion we have combined T2 and T2* contrast agents as one group, discussing them separately only when necessary as a critical parameter for the method. The parameter relaxivity is used to define the effectiveness of a T1 or T2 relaxation CA. For CA based on paramagnetic metals, it is defined as a reciprocal of changes in T1 or T2 relaxation time per unit concentration of the metal $(mM \cdot s)^{-1}$. T1 and T2 relaxation processes are based on similar molecular mechanisms and most CA affect both T1 and T2 relaxation times. The classification of CA as T1 or T2 agents is therefore conditional and simply means that a particular CA under a certain set of experimental conditions will predominantly alter the T1 or T2 relaxation time.

### T1 Contrast Agents

The important advantages of T1 relaxation agents for in vivo application are (1) the long precontrast tissue T1 relaxation time of about 1.7 s and (2) the positive contrast (or signal enhancement) generated by the agent in MR images. The majority of T1 contrast agents currently used both in the clinic and in animal studies are chelate complexes of paramagnetic metal ions such as Gd(III), Mn(II), or Fe(III). Several stable complexes of Gd(III) including linear GdDTPA (diethylenetriaminepentaacetic acid) and cyclic GdDOTA (1,4,7,10-tetraazacyclododecane-1,4,7,10-tetraacetic acid) chelates are extracellular low-molecular-weight compounds with high T1 relaxivity and are the only class of paramagnetic CA approved for clinical use. Typical examples of low-molecular-weight Gd-based agents include Gadoteridol, Gd-HP-DO3A (Prohance), Gadopentate dimeglumine, Gd-DTPA (Magnevist), Gadoversetamide (Optimark), and Gadodiamide, Gd-DTPA-BMA (Omniscan).

To improve the T1 relaxivity and to prolong the plasma lifetime of the agent, multiple Gd chelates can be attached to a single polymeric backbone molecule. The resulting complexes have a longer circulation time and increased relaxivity that is advantageous for a range of clinical applications, although the potential toxicity and induction of immune response to the polymeric carrier remain to be determined. Several gadolinium-based macromolecular imaging platforms have been designed and tested in preclinical models. Protein–Gd complexes include albumin–Gd conjugates, which are classic prototypes for

intravascular imaging agents [7, 8]; poly-L-lysine–Gd [9, 10], avidin–Gd [11], as well as direct mAb–Gd conjugates [12, 13]. Polyamidoamine (PAMAM) dendrimers of different generations have also been used as a carrier for multiple Gd groups [14, 15]. For targeted intravascular imaging, several classes of very large molecular weight agents labeled with a high number of Gd atoms have been developed including cross-linked liposomes and nanoparticle emulsions [16, 17]. The typical molecular weight and the size of polymer Gd chelates varies from about 80 kD for protein-based agents to several million daltons for polymerized liposomes and emulsions with molecular diameters up to 200 nm. T1 relaxivity of macromolecular Gd complexes strongly depend on the $B_0$ magnetic field as well as on the rotational and tumbling correlation times of the complex. For targeted CA the correlation time can change significantly if the complex is immobilized by specific binding to the target receptor. The reported T1 relaxivity of the macromolecular CA ranges from 12 $(\text{mM}\cdot\text{s})^{-1}$ for 200-nm nanoparticle emulsion at $B_0$ = 4.7T [16] to 36 $(\text{mM}\cdot\text{s})^{-1}$ for generation-nine PAMAM dendrimers [14] at 20 MHz and 80 $(\text{mM}\cdot\text{s})^{-1}$ for MS-325 bound to albumin [18]. The relaxivity of a macromolecular Gd polychelate per unit concentration of the polymer increases linearly with the number of gadoliniums, as long as two Gd ions are separated by a distance that is much longer then a characteristic diffusion radius of water molecules. For a typical contact interaction between a water molecule and a metal ion of 100 ps and water diffusion coefficient of $10^{-5}$ $\text{cm}^2/\text{s}$, the minimal distance between adjacent Gd ions should be larger than about 100 Å. If this condition is fulfilled, then the T1 relaxivity of these agents (per unit concentration of the metal) should not depend on the number of attached gadolinium ions. Indeed the measured T1 relaxivity of albumin GdDTPA complexes with a labeling ratio changing in the range of 9 to 18 Gd per albumin globule remains close to 12 $(\text{mM}\cdot\text{s})^{-1}$ at 10.7 MHz [8].

## T2 Contrast Agents

Contrast agents that increase T1 relaxation and accordingly reduce T1 relaxation time also change T2 relaxation as T2 is always shorter than T1. In fact all T1 agents discussed in the previous section can also act as T2 CA. However, because in a typical in vivo situation the inherent T2 relaxation time is significantly shorter (by an order of magnitude) than the corresponding T1 relaxation time, these agents induced dramatically more significant changes in T1 relaxation of the sample.

T2 contrast agents typically possess higher magnetic moments in comparison with T1 CA, and, because of their larger molecular size, they have correspondingly longer correlation times $\tau_c$. This high magnetic moment makes them efficient at reducing T2 relaxation time and the relatively long $\tau_c$ reduces their effects on T1 relaxation time. T2 MR contrast agents usually consist of a superparamagnetic iron-oxide core that generates a macroscopic magnetic moment of the aligned electronic moments (so-called Curie spin [19]). This magnetic moment gives rise to strong local gradients of magnetic field that efficiently dephase nearby protons in an area much larger than the size of the

magnetic particle [19]. Inhomogeneous line broadening induced by these agents also results in a significant shortening of T2* relaxation time. However, because of the very short inherent T2* in biological systems at high magnetic fields $B_0$, in the majority of in vivo applications T2 contrast enhancement is measured. Smaller superparamagnetic CA can also serve as efficient T1 contrast agents at low $B_0$ of about 1.5T, where $\tau_c^{-1}$ of the CA molecule is close to the proton magnetic resonance frequency.

The magnetic core of T2 CA usually consists of monocrystalline (MION) or polycrystalline (SPIO) iron oxide with a diameter of 5 and 30 nm, respectively. To improve the stability and biocompatibility of the complex, the core is coated with a polymer such as dextran or other polysaccharide resulting in a total particle diameter of 17 to 50 nm. Dextran-coated cross-linked iron oxide nanoparticles (CLIO) can be easily modified by functionalizing surface groups in the dextran coating [20]. Alternatively, the magnetic core can be embedded into a unilamellar liposome vesicle (magnetoliposomes [21]). Magnetodendrimers are CA with a superparamagnetic iron oxide core encapsulated within a dendrimer superstructure [22, 23].

An important feature of the iron-oxide-based CA is their high T2 relaxivity, which depends upon the size of the superparamagnetic core. The T2 relaxivity of the ultrasmall monocrystalline superparamagnetic iron oxide (MION-46L) with a diameter of the magnetic core of 4 nm (2046 Fe atoms) is close to 20 $(mM{\cdot}s)^{-1}$ at 1.5T magnetic field and 25°C [24]. SPIO particles with iron core diameter of 16 nm have the typical T2 relaxivity at 1.5T of about 240 $(mM{\cdot}s)^{-1}$. For comparison the T1 relaxivity of these compounds at 1.5T magnetic field is below 10 $(mM{\cdot}s)^{-1}$ [25].

### Relaxation Properties of Main Classes of CA

Relaxation and molecular properties of different MRI agents suitable as platforms for development of targeted CA for MR molecular imaging are summarized in Table 1. MR molecular imaging presents unique challenges to researchers because of the inherently low sensitivity of MR detection (Gd concentrations in tens of micromolars) and the typically low number of imaging targets per cell. To design CA with the highest relaxivity per CA molecule, it is necessary either to use a large iron oxide core or large polymer carriers complexed with multiple Gd groups. The large molecular size prevents the efficient delivery of the agent to the imaging site because of slow extravasation and restricted diffusion of the large CA through the tumor interstitium. This problem may be less important for MRI of tumors with highly permeable vasculature [26] and is not relevant for imaging of molecular targets in the lumen of blood vessels such as the vascular endothelium [16, 17].

## 3 MRI OF TUMOR VASCULATURE AND VASCULAR TARGETS

Mechanistic approaches to characterizing the tumor vasculature include measurements of the following vascular parameters: tumor blood flow, tumor

**TABLE 1 Typical Properties of Platforms for Targeted MR Contrast Agents**

| CA Class | Representative Compound | Number of Metal Atoms | T1 Relaxivity $(mM{\cdot}s)^{-1}$ | T2 Relaxivity $(mM{\cdot}s)^{-1}$ | Magnetic Field $B_0$ $(T)$ | Molecular Diameter (nm) | Molecular Weight (Da) |
|---|---|---|---|---|---|---|---|
| Low molecular weight | GdDTPA Magnevist® | 1 [$Gd^{3+}$] | 4.5 | 5.7 | 1.5 | — | 743 |
| Protein carriers | Albumin | 19 [$Gd^{3+}$] | 11.5 | 12.4 | 0.25 | 8 | ~80,000 |
| | Poly-L-lysine | 65 [$Gd^{3+}$] | 13 | 15 | 0.47 | — | ~52,000 |
| Dendrimers | PAMAM Generation 5 | 96–128 [$Gd^{3+}$] | 22.5 | 27 | 1.5 | 6–8 | ~60,000 |
| | Gadomer-17 | 24 [$Gd^{3+}$] | 13 | | 1.5 | 5–6 | 17,500 (35,000 apparent size) |
| Iron oxide nanoparticles | MION-46 | 2064 [Fe] 4.6 nm iron core | 7 | 20 | 1.5 | 20 | ~775,000 |
| | SPIO Feridex® | ~6.5 nm $Fe_2O_3$ core | 24 | 107 | 0.47 | 70–140 | Megadalton range |
| Magnetodendrimers | | 9.5 nm $Fe_2O_3$ core | 12 | 225 | 1.5 | 40–50 | Megadalton range |
| Liposomes | ACPL (antibody-conjugated paramagnetic liposome) | ~50,000 [$Gd^{3+}$] | 12 | 11 | 1.5 | 200 | Megadalton range |
| Nanoparticle emulsions | Gd-perfluorocarbon nanoparticles | 74,000 [$Gd^{3+}$] | 34 | 50 | 1.5 | 250 | Megadalton range |

vascular volume, and vascular permeability. A traditional approach to quantifying tumor vasculature is by histological evaluation of tumor microvessel density (MVD) [27]. Histological methods, however, do not provide information about the functionality of the vasculature that is known to be partly, permanently, or intermittently collapsed in tumors [28, 29]. Dynamic contrast-enhanced MRI provides an alternative noninvasive approach to determine some of these parameters, and we will discuss several applications of dynamic MRI in the following section. Imaging of vascular endothelium receptors that are unique for the tumor neovasculature such as selectins or integrins is yet another area where MRI with targeted contrast agents was successfully used in preclinical tumor models. This molecular MRI approach may facilitate early detection of the malignant growth and/or identify molecular targets for therapeutic intervention.

## Low-Molecular-Weight Contrast Agents (GdDTPA)

Dynamic MRI studies of tumors with low-molecular-weight CA were performed in several tumor types including breast [30], brain [31], and uterine tumors [32]. GdDTPA is the most widely used Gd-based paramagnetic contrast agent. In the following sections we will use this generic name for general low-molecular-weight CA. The dynamic parameters derived from tracer kinetic analyses of these experiments, together with the assumptions made, have been discussed in several excellent studies [33, 34]. Generally, these agents are not freely diffusible; therefore, it is not possible to quantify tumor blood flow based on the two-compartment Kety model [35]. GdDTPA contrast agents rapidly extravasate from the leaky tumor vasculature, and the first-order kinetic constant defined as a transfer coefficient between the blood plasma compartment and the tumor interstitium is therefore affected by insufficient tumor perfusion (flow-limited regime) and generally depends on both PS (vascular permeability surface area product) and blood flow [36].

In bolus tracking experiments GdDTPA is injected at high doses, and the difference in magnetic susceptibility between blood vessels and extravascular space results in local magnetic field gradients or inhomogeneity broadening effects. These local gradients generate T2 or T2* contrast in spin echo or gradient echo imaging, respectively. Kinetic analysis of the dynamic susceptibility contrast in bolus tracking experiments provides a means to determine regional blood volume (rBV). While this approach is feasible for measurements of cerebral blood volume [37], its application to tumors is limited by the leakage of the low-molecular-weight CA across highly permeable tumor vasculature. Increased concentration of the CA in the tumor interstitium tends to attenuate local susceptibility gradients, and the experimental data have to be corrected for the leakage effects [38]. T2 and T2* dynamic MRI is also sensitive to the caliber of tumor vasculature, and appropriate experiments can be used to characterize micro- and macrovasculature in the tumor [37, 39–42].

## High-Molecular-Weight Contrast Agents

A generic high-molecular-weight contrast agent is one in which serum albumin is conjugated to multiple GdDTPA groups using free amines on the surface of the polymer globule [43]. This CA is characterized by a long plasma half-life (about 2 h), and after initial equilibration period (about 3 min in mice) its blood concentrations remain constant for at least 40 min after an intravenous bolus injection [1]. The CA slowly leaks across the tumor vasculature, and tissue concentrations of CA increase linearly with time. Simple linear kinetic analysis of albumin–GdDTPA concentration as a function of time is sufficient to quantify tissue blood volume (from extrapolation of kinetic curve to time $t = 0$ min) and permeability surface area product, PS, from the slope of the line [7]. Typically albumin–GdDTPA tissue concentrations are determined from changes in T1 relaxation time under assumption of fast exchange of water moleculaes between blood, interstitium, and cellular compartments. Intermediate to slow water exchange can result in underestimation of vascular volume and PS [44].

High-molecular-weight CA extravasate from the leaky tumor vasculature at a relatively slow rate as compared to small molecular weight CA such as GdDTPA. Therefore, their in vivo kinetic is (1) sensitive to a wide range of PS present in solid tumors and (2) can be modeled using a simple linear kinetic model [45]. The utility of this method was demonstrated for imaging of breast and prostate cancer models preselected for different metastatic potential. More metastatic tumors were generally characterized by increased tumor vascular volume and PS [7]. We recently also investigated the effect of the antiangiogenic agent TNP-470 on tumor vascular characteristics using MRI [1]. TNP-470 is a fumagillin derivative and its antiangiogenic effects are thought to be due to inhibition of endothelial cell proliferation [46].

Representative vascular volume and permeability data obtained for a pair of volume-matched control and treated tumors are shown in Figures 1 and 2. Treatment with TNP-470 resulted in a significant decrease in the permeable and vascular regions within the tumors. However a "compensatory" increase in levels of vascular volume and permeability was also apparent in some regions of the treated tumors and demonstrate the critical importance of imaging techniques that provide spatial information in determining the effects of antiangiogenic therapy.

## Molecular Imaging of Tumor Neovasculature

Tumor growth depends on the establishment of new vasculature recruited from the host through the process of angiogenesis [47, 48]. This neovasculature is characterized by the expression of specific molecular markers such as $\alpha_v\beta_3$ integrins. In adults these markers are mostly localized to the site of active angiogenesis and therefore can be used as targets for imaging tumor neovascularization with low background from quiescent mature vasculature [49].

One successful strategy for MRI of vascular molecular epitopes is based on using high-affinity functionalized MR contrast agents directed against a specific target. Selective retention of the CA by the target and clearance of the

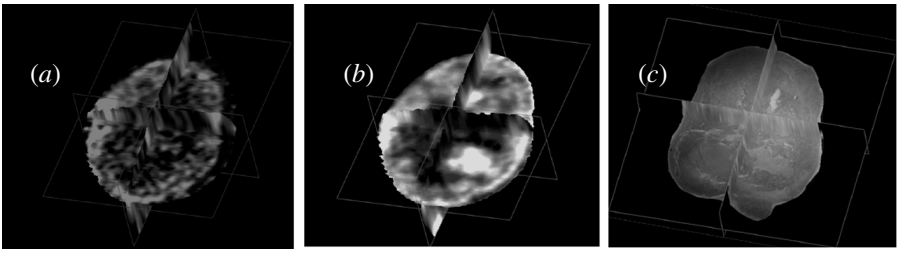


**Figure 1** Triplanar views of three-dimensional reconstructed maps of (*a*) vascular volume, (*b*) permeability, and (*c*) hematoxylin and eosin-stained histological sections for a control MatLyLu tumor (volume 405 $mm^3$).

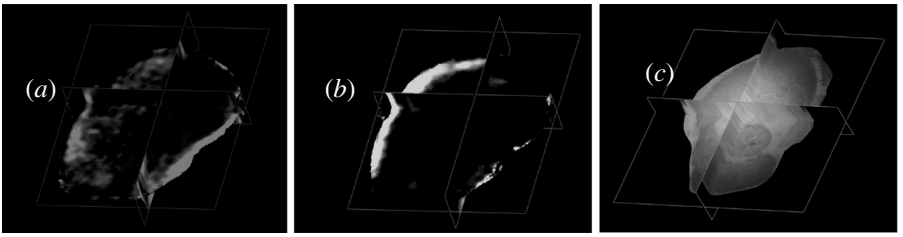


**Figure 2** Triplanar views of three-dimensional reconstructed maps of (*a*) vascular volume, (*b*) permeability, and (*c*) hematoxyline and eosin-stained histological sections for a MatLyLu tumor treated with TNP-470, 30 mg/kg every second day, total dose 90 mg/kg (volume 395 $mm^3$).

unbound "free" agent generates measurable contrast in the image. Highly specific targeting can be achieved by attaching the imaging probe to monoclonal antibodies (mAb) or mAb fragments, adapter proteins, and/or synthetic polypeptides with high specific binding to the target. The intrinsically low sensitivity of MRI and low concentrations of molecular targets require using CA with the highest available relaxivity. The accessibility of molecular targets in the tumor vasculature allows the use of large contrast agents with molecular size of up to 1 μm. Molecules of CA are delivered within the bloodstream and reach and bind the target site without extravasation and/or diffusion across tissue barriers.

E-selectin, a proinflammatory marker of endothelial cells, is expressed in proliferating endothelium and can be a marker of tumor neovasculature [50]. An MR molecular imaging agent specific for E-selectins was developed by Kang et al. [50] and is a molecular conjugate between CLIO iron oxide particle and mAb $F(ab)_2$ fragment specific for the protein. MRI of endothelial cells in culture demonstrated strong retention of the contrast agent by the cells stimulated to express E-selectins [50].

$\alpha_v\beta_3$ receptors are expressed in the vascular angiogenic endothelium [16, 17]. $\alpha_v\beta_3$ receptors were imaged in vivo in the neovasculature of a rabbit tumor model with paramagnetic polymerized liposomes [17] and in the rabbit bFGF-

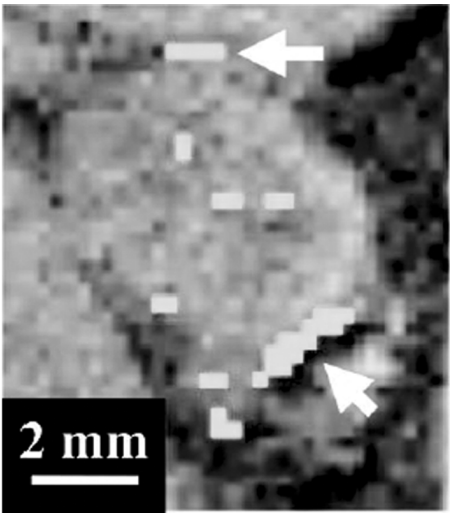


**Figure 3** Enlarged section of T1-weighted MR image showing implanted Vx-2 tumor. Yellow overlay indicates MRI signal enhancement 120 min postinjection of $\alpha_v\beta_3$-targeted nanoparticles. MRI enhancement was predominantly, although not exclusively, asymmetrically distributed along the tumor periphery proximal to blood vessels and tissue fascial interfaces (white arrows). From [51].

induced corneal micropocket model using Gd-perfluorocarbon nanoparticles [16]. These imaging nanoparticles were targeted to the receptor by mAb covalently bound or attached via biotin-avidin linkers, respectively. More recently $\alpha_v\beta_3$ receptors were imaged in Vx-2 rabbit tumor model by targeted Gd-perfluorocarbon paramagnetic nanoparticles functionalized with a $\alpha_v\beta_3$-integrin peptidomimetic antagonist selected for high-affinity binding to the receptor [51]. MRI was performed on a standard clinical 1,5T scanner and demonstrated potential feasibility of the method for clinical translation (Fig. 3).

## 4 MRI OF TUMOR CELL SURFACE RECEPTORS

### MR Molecular Imaging of Isolated Cells

Isolated cells in culture can be used as a test bed for identification of the potential tumor targets for imaging and/or therapy and for the development of surface receptor specific MRI contrast agents. Additionally, cell surface receptor CA systems can be used for long-term labeling of the cells for cell tracking MR application as an alternative to loading of cells with intracellular CA (discussed in next section). Caution should be exercised in the extrapolation of imaging results obtained in isolated cell to in vivo studies. Contrast agent pharmacokinetics, delivery properties, and stability in vivo can be determin-

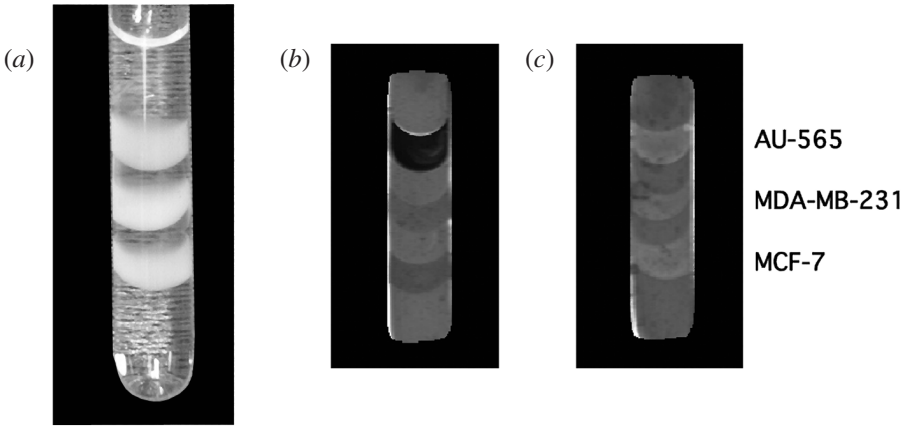


**Figure 4** MRI of breast cancer cells. (*a–c*) Demonstrate the layout and MR images of cell samples consisting of layers of AU-565, MDA-MB-231, and MCF-7 cells embedded in agarose gel in a 5-mm NMR tube. Cells were pretargeted with biotinylated Herceptin and a nonspecific biotinylated mAb (negative control), and probed with streptavidin SPIO microbeads. $T_2$ maps of the cell samples were reconstructed from eight $T_2$-weighted images acquired with RD of 8s and TE in the range 20 to 250 ms. A $T_2$ map of a cell sample probed with Herceptin is shown in (*b*) and the control cell sample treated with a nonspecific biotinylated mAb is shown in (*c*). Adapted from [55].

ing factors for imaging efficiency in animal models and eventually in clinical applications.

In cultured cancer cells the surface receptors are entirely accessible for labeling with large molecular probes as in the case of endothelial cell receptors discussed above. The most efficient probes, as seen from Table 1, are iron-oxide-based CA as they provide the highest T2 relaxivity per mole of the compound. This group of CA includes SPIO, MION, CLIO, and magnetodendrimer nanoparticles. Chemical conjugation of the superparamagnetic iron oxide particle with mAb or mAb fragments provides selective binding of the CA to the target receptor on the cell surface. Original reports on using functionalized SPIO nanoparticles as MR receptor imaging agent were published in the 1990s [52, 53]. A multistep protocol based on biotinylated primary mAb, streptavidin, and biotinylated dextran-coated SPIO nanoparticles was proposed for imaging of lymphocytes with the mAb directed against the lymphocyte common antigen [54]. A two-step labeling method was also used to image HER-2/neu receptors expressed on the surface of malignant breast cancer cells using biotinylated Herceptin mAb and streptavidin-conjugated SPIO nanoparticles [55]. The results presented in Figure 4 demonstrate a significant negative T2 contrast in the HER-2/neu expressing cell layers. The highest contrast was detected in AU-565 cells expressing $2.7 \times 10^6$ receptors/cell and the lowest in MDA-MB-231 cells with $4 \times 10^4$ receptors/cell. The

experimental data suggest a linear dependence between 1/T2 relaxation rate and concentration of the target sites for the SPIO-based CA [55].

Human engineered tranferrin receptor (ETR) was imaged in cancer cells using a conjugate of transferrin and MION nanoparticles [56, 57]. The CA was accumulated in ETR+ expressing cancer cells by internalization through receptor-mediated endocytosis mechanism.

A significant advantage of the iron-oxide-based CA is their high T2/T2* relaxivity that produces strong negative MR contrast already at nanomolar concentrations of the CA and provides sufficient sensitivity to detect cell surface receptors expressed at relatively low levels ($<10^4$ receptors/cell). However, the relatively large molecular size (>30 nm) of the superparamagnetic iron-oxide-based CA is a potential problem for in vivo applications, as discussed in the next section.

## In Vivo MR Molecular Imaging of Receptors in Cancer

In this section we present a concise review of MRI results obtained in vivo in preclinical systems. These results may be considered as preliminary steps toward translation of the methods to clinical use. When translating these applications to the bedside, in contrast to nuclear imaging, molecular MRI encounters problems. These are primarily associated with safety issues due to potential toxicity and the triggering of the immune response by high doses of targeted CA, which need to be administered to produce sufficient contrast. On the other hand results obtained in animal systems do have intrinsic importance. They can provide insights in tumor biology, the tumor microenvironment and response to various stresses that essentially require in vivo longitudinal studies. These results can also be important for the rational design of novel anticancer treatments.

***Gd-Based Contrast Agents*** We start our discussion with early attempts of in vivo MRI of various cell surface molecular targets in solid tumors using approaches similar to standard nuclear imaging protocols with radiolabeled mAb. To this end monoclonal antibodies were labeled with GdDTPA chelates, and reports on the use of the GdDTPA-mAb appear in the literature since 1985 [13, 58, 59]. Generally, these experiments produced discouraging results due to the low sensitivity of MR detection and limited contrast produced by insufficient concentration of the CA [58]. Many of these early studies also did not include appropriate control experiments. In comparison to nuclear imaging where nanomolar concentrations of immunoradioisotopes can be routinely detected in tumors, MRI detection requires concentrations of Gd above approximately 10 μM. Not more than 10 to 15 Gd complexes can be attached per antibody without significant reduction in the binding affinity [10]. Therefore, building a sufficient concentration of Gd at the target site is problematic if direct labeling of mAb with Gd is used. In a more recent study, Shahbazi-Gahrouei et al. reported on the ex vivo relaxometry of human MM-138

melanoma xenografts probed in vivo with antimelanoma GdDTPA-mAb [12]. These studies were performed with ex vivo MR at high magnetic field, with a high molar ratio of Gd to mAb. It is not clear that similar specific signal enhancement can be achieved in a typical in vivo MR setup.

The most straightforward way to increase the number of GD groups per mAb is to conjugate the antibody to a polymer carrier that is decorated by multiple Gd groups via an appropriate molecular linker. Antimucin mAb conjugated to poly-L-lysine-GdDTPA with a labeling ratio of 65 Gd ions per molecule were developed for imaging of mucinlike protein expressed in many types of gastrointestinal carcinomas [10]. The molecular weight of the conjugate was about 200 kDa, which may restrict its delivery to solid tumors. Konda et al. reported the development of a folate-conjugated dendrimer complexed with GdDTPA [60]. A fourth-generation PAMAM was used as the imaging platform for targeted MRI, and initial results were obtained for human folate receptor expressing ovarian tumor xenografts grown in nude mice [60].

***Iron-Oxide-Based Contrast Agent*** Two quite different modes of application of iron-oxide-based CA were proposed for in vivo MRI. In the first variant iron oxide nanoparticles (MION) were conjugated to mAb or other high-affinity moieties to label cell surface receptors while mostly remaining in the extracellular compartment. In a study by Weissleder et al. human polyclonal IgG was used to target MION to sites of inflammation [61]. A similar approach was used for molecular MRI of apoptosis in EL4 solid tumor models exposed to a chemotherapeutic agent [62]. SPIO particles were conjugated to the C2 domain of the protein synaptotagmin, which binds with high affinity to phosphatidylserine residues that translocate to the outer leaflet of the plasma membrane in apoptotic cells. Efficient delivery of the CA to these sites can be explained by significantly increased vascular permeability in the sites of inflammation or in treated tumors. Targeting peptides conjugated to CLIO particles also labeled with fluorophores were used to image underglycosylated MUC-1 tumor antigen [63]. Combined T2 MRI and near-infrared fluorescence imaging demonstrated a specific accumulation of the contrast in the tumor that expressed the antigen as shown in Figure 5.

The second strategy for generating specific MR contrast is to load cells with iron oxide agent using a specific plasma membrane transporter system with an appropriate substrate. Transferrin-MION contrast agent was demonstrated in vivo in engineered transferring receptor (ETR) expressing 9L glioma cells [56]. The transgenic cells internalized up to $8 \times 10^6$ of the TF-targeted CA within an hour and MION-loaded cancer cells generated strong negative T2* contrast in vivo in gradient echo MR images. Images were obtained at 1.5T 24 h after intravenous injection of 3 mg of Tf-MION to a nude mouse with two $ETR^-$ and $ETR^+$ 9L gliosarcoma tumors growing on opposite flanks of the animal. No differential contrast was detected in the T1 weighted image, and strong negative contrast was detected in the T2*-weighted image of the $ETR^+$ tumor. The efficient delivery of 17-nm MION particles to the tumor was most

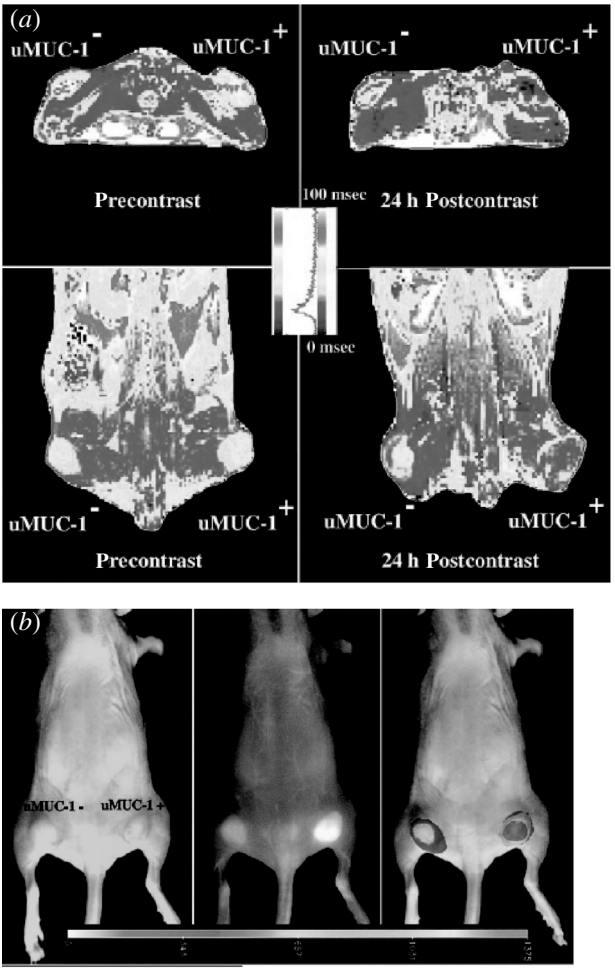


**Figure 5** (*a*) Representative T2 maps of the animals bearing underglycosylated mucin-1 antigen (uMUC-1)-negative (U87) and uMUC-1-positive (LS174T) tumors. Transverse (*top*) and coronal (*bottom*) images showed a significant (52%; $P = 0.0001$) decrease in signal intensity in uMUC-1-positive tumors 24 h after administration of the CLIO-EPPT probe. (*b*) White light (*left*), NIRF (*middle*) images, and a color-coded map (*right*) of mice bearing bilateral underglycosylated mucin-1 antigen (uMUC-1)-negative (U87) and uMUC-1-positive (LS174T) tumors. NIRF imaging was performed immediately after the MRI session. Adapted from [63].

likely a result of the favorable pharmacokinetics of the agent due to its long circulation time and hyperpermeable tumor vasculature.

***Multistep Targeting and Prelabeling Concept*** Tumor neovasculature is hyperpermeable in regions and permits extravasation of large molecules [64], which was demonstrated in several studies with macromolecular CA (above 20 nm molecular size), such as MION and SPIO [56, 62]. Solid-tumor vascularization is heterogeneous, and regions of high vascular permeability are often spatially distinct from regions of high vascular volume [7]. It was also shown that tumor perfusion is highly inhomogeneous and may have an intermittent blow flow [29]. Consequently, the delivery of high-molecular-weight imaging and therapeutic agents to tumor areas with low vascularity/flow/permeability can be significantly restricted. There is still a concern that extravasation and diffusion barriers for these nanocomplexes can significantly reduce their usefulness in viable regions of solid tumors following systemic administration [26]. Pharmacokinetics and delivery properties of targeted CA can be significantly improved by splitting the large targeted imaging/therapeutic probe to several functional components that can be administered and delivered to the tumor targets independently. This prelabeling concept is extensively used in nuclear imaging to increase the target/background ratio and to match the pharmacokinetics of radionuclides to their half-lifetimes [65].

HER-2/*neu* receptors were imaged in vivo using a two-step prelabeling strategy based on the avidin/biotin system [11]. Briefly, receptors in tumor xenografts were prelabeled with biotinylated mAb. After a 12-h interval, during which mAb were attached to the imaging targets and unbound mAb cleared from the systemic circulation, GdDTPA-labeled avidin was injected intravenously (IV). T1-weighted MRI was performed at 4.7T at different time points after the injection of avidin–GdDTPA and demonstrated positive T1 contrast in breast cancer tumor xenografts that overexpresssed the HER-2/*neu* receptor. The molecular weight of the individual components was 160 kDa for mAb and approximately 70 kDa for GdDTPA avidin, which corresponds to a molecular size smaller than 10 nm and provided efficient delivery of the CA to the interstitium of solid tumors in this model system. To further improve this approach we developed a three-component targeting system that included biotinylated primary mAb, purified avidin that served as a linker and mAb clearance agent, and biotinylated Gd-based macromolecule CA [66]. Avidin was injected 24 h after administration of the primary mAb and served two separate functions: (1) It provided selective labeling of cells with the attached biotinylated mAb and (2) avidin chase of circulating biotinylated mAb provided their rapid clearance from the blood, which significantly reduced background contrast [67]. In comparison to avidin–GdDTPA conjugates, this protocol also allows the use of optimized biotinylated paramagnetic agents with large number of Gd groups to generate contrast in MRI. An example of MRI of human HER-2/*neu* expressing BT-474 xenografts in SCID (severe combined immunodeficient) mouse is shown in Figure 6. Briefly, animals

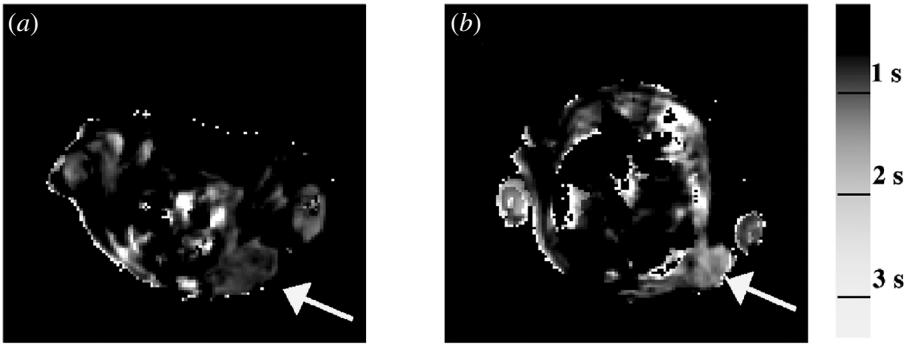


**Figure 6** T1 maps of BT-474 tumors imaged with (*a*) biotinylated and (*b*) non-biotinylated Herceptin. Tumors are indicated by arrows.

were injected IV with anti-HER-2/*neu* humanized mAb (Herceptin, Genentech) at a dose of 0.5 mg per animal. Biotinylated mAb was used for the experimental group and nonbiotinylated mAb for the control group of animals. Avidin (0.5 mg per animal) was administered IV 24 h after mAb, followed after 4 h by IV infusion of biotin-albumin-$(GdDTPA)_{20}$ contrast agent (0.5 mg/g). T1 relaxation maps were reconstructed from saturation recovery spin-echo imaging for several time points after administration of the contrast for the control and the experimental group and selective retention of the CA, and accordingly shorter T1 relaxation times were detected in BT-474 tumors treated with biotinylated Herceptin as shown in Figure 6.

***Signal Amplification Concept*** The imaging techniques discussed so far rely on imaging of cell receptors expressed in large numbers (over a million) per cell. In many biologically relevant situations, expression levels of the receptor are significantly lower. It applies both to endogenous receptors that can be markers of a disease and to molecular reporter systems linked to the expression of extracellular receptors that can be used as molecular beacons for activation of relevant genes due to changes in the signaling and/or pharmacological interventions [68]. While MRI may not compare favorably with nuclear medicine for detecting cell surface receptors, its advantage arises from the ability to combine detection of these receptors with the array of functional imaging capabilities of MR methods. To achieve sufficient contrast enhancement from a low number of available imaging targets, signal amplification is necessary. Generally, signal amplification takes place if we can generate multiple contrast generating groups per single imaging target.

An increase in the number of CA molecules attached to a single extracellular target can be achieved by using the avidin–biotin system described above [69] as each avidin/streptavidin molecule can bind four biotins, and theoretically multiple labeling steps can give rise to treelike structures formed by the

biotinylated CA. The relaxivity of individual CA units can be significantly improved by using polymer platforms labeled with multiple Gd chelates such as poly-L-lysine or dendrimer based CA [70]. Iron-oxide-based CA in principle can provide higher relaxivity in comparison with paramagnetic CA. Their large molecular size, however, can interfere with contrast delivery, and negative T2 contrast may be more difficult to detect. Recently, a novel frequency shift technique for detection of positive T2* contrast was developed that may provide an advantage for in vivo applications of SPIO base agents [71]. Another stratergy for signal amplification relies on loading target cells with CA molecules via active transport through the target receptor. Experiments with ETR receptor and MION nanoparticles conjugated to human holotransferrin is a good illustration of this approach [56]. The possible effects of the internalized CA on cell homeostasis and functions is one of the potential problems of the method.

The most effective approach to signal enhancement is to couple the targeting mechanism to a specific enzyme in such a way that the generated product acts as imaging or therapeutic CA. T1 relaxivity of a CA depends on the relation between the correlation time $\tau_c$ of the agent and $B_0$ magnetic field used for imaging. Enzymatic polymerization of small Gd complexes to a large-molecular-weight polymer product results in significant increase in $\tau_c$ and correspondingly significant increase in T1 relaxivity. Polymerization of hydroxyphenol-modified GdDOTA monomers was proposed as a sensitive probe for MRI of nanomolar concentrations of oxidoreductases (peroxidase) [72]. Low-relaxivity [3.75 $(\text{mM}\cdot\text{s})^{-1}$] monomeric substrates oxidized by the peroxidase-reduced peroxide produce activated molecules ($A^*$) that self-polymerase to form high-relaxivity [11.5 $(\text{mM}\cdot\text{s})^{-1}$] magnetic oligomers. This method was used for MRI of E-selectin expressed on the surface of endothelial cells using sandwich constructs of anti-E-selectin $F(ab')_2$ conjugated to peroxidase through DIG–anti-DIG Ab linkers [72]. It is still unclear if the method can be implemented in vivo; however, this strategy is definitely a significant step toward improving the performance of specific CA for molecular MR studies.

## 5 MRI OF INTRACELLULAR TARGETS

In previous sections we discussed various MR contrast agents suitable for molecular imaging of cellular epitopes using specific probes targeted against extracellular domain of cell surface receptors. In this chapter we will focus on intracellular MR contrast agents designed to penetrate the cell plasma membrane and accumulate in the cytoplasm and/or label specific intracellular targets. We will discuss two broad groups of intracellular CA that have recently been developed for MR imaging. The first group includes nonspecific intracellular agents primarily used for a long-term cell labeling for in vivo cell tracking. Typical examples of their application includes MR trafficking of

inoculated stem cells and MRI of the docking of armed T cells to target sites. The second group consists of targeted intracellular contrast agents that have high affinity to molecular epitopes or can be activated by the action of cell enzymes.

## Contrast Agents for Long-Term Labeling of Target Cells (MRI Cell Trafficking)

A typical experiment for MR cell trafficking includes in vitro labeling of cell population with an appropriate CA, administration of the labeled cells in vivo using systemic or local injection, and longitudinal MRI of the cell migration and accumulation to the target site(s). For a detailed discussion of MR tracking of magnetically labeled cells, we refer readers to excellent reviews by Bulte et al. [73–76].

Isolated cells labeled with iron oxide nanoparticles by phago- and/or endo/pinocytocis can be detected by MR at iron concentrations as low as 17 ng per $10^6$ cells or $8.5 \times 10^4$ particles/cell (9L glioma cells loaded with MION particles [77]). MRI of a single T cell loaded with SPIO particles was demonstrated by Dodd et al. [78]. Typically, several million SPIO or USPIO particles must be loaded into the cell to provide sufficient contrast for MR detection. Detection of a cell labeled with just a single micron-size iron oxide particle was recently demonstrated in vitro [79]. One potential benefit of this approach is that the label is not diluted by subsequent cell division, and daughter cells can be detected for a long time postimplantation. T2*-weighted gradient-echo MRI that is sensitive to local disturbances in the magnetic field produced by the concentrated SPIO particles is usually used for detection of MR contrast in vitro in isolated cells. The benefit of T2* MRI is that "long-range" effects affect T2* relaxation of distant protons within an area much larger than the size of the nanoparticle, thus significantly amplifying sensitivity of the method [54, 78].

To improve internalization of the contrast agent, magnetic nanoparticles are typically mixed with standard cell transfecting agents such as poly-L-lysine [76]. Magnetodendrimers were proposed as an efficient cell-labeling MR agent. Stem cells loaded with magnetodendrimers (9 to 14 pg iron/cell) were successfully imaged both in vitro and in vivo for as long as 6 weeks after transplantation [23]. To further improve cell uptake properties, CA nanoparticles can be linked to the human immunodeficiency virus (HIV) viral transport Tat peptide that can channel the cargo molecules across the cell plasma membrane [80]. Efficient cellular uptake of Tat–CLIO as well as Tat–macrocyclic gadolinium chelates was recently demonstrated in in vitro and in in vivo systems [81, 82]. This approach is especially useful for loading of cells with low endocytotic/pinocytotic activity such as lymphocytes. Tat–CLIO nanoparticles were used to load cytotoxic T lymphocytes (CTL) ex vivo and to track their docking to the antigen-expressing tumor xenografts in vivo by T2-weighted MRI [83]. Repetitive noninvasive detection of the adoptively transferred cell targeting

by MRI was demonstrated that may be particularly useful for evaluating novel cell-based therapies in vivo.

## Specific Contrast Agents for Intracellular Targets

Targeting intracellular epitopes is a major challenge for MR molecular imaging. Although many of potentially important biological targets are essentially intracellular such as enzymes, signaling proteins, deoxyribonucleic acid (DNA) and messenger ribonucleic acid (mRNA), most of the development in the area of MRI contrast agents was focused on extracellular CA. Indeed, in a traditional approach similar to the one used in nuclear imaging to specifically probe intracellular targets, contrast agent molecules need to freely diffuse across the cellular membrane and selectively accumulate in cells that express the marker. Large CA such as iron oxide nanoparticles primarily enter cells through endocytosis, and their retention in the cytoplasm is not target specific. To enable small gadolinium complexes to enter the cytoplasm and to provide a specific recognition of an intracellular target, a construct consisting of an amphiphilic transmembrane carrier peptide, Gd chelate, and c-Myc mRNA specific antisense peptide nuclei acid (PNA) [84] was created as shown in Figure 7. Specific uptake of the c-Myc-targeted CA was detected in rat

**Figure 7** Spatial representation of one possible configuration of the c-myc-specific Gd3 complex. The transport peptide unit (green) is given as a ribbon representation of the peptide backbone. The heavy atoms of the address peptide (PNA; light brown), the disulfide bridge between the two peptide units and the two Lys-residues (cyan) connecting the $Gd^{3+}$ complex to the PNA are displayed in a "ball-and-stick" representation. An atom color code (green, carbon; red, oxygen; blue, nitrogen) is used for the disulfide bridge and $Gd^{3+}$ complex. *Magenta*, the van der Waals spheres of the $Gd^{3+}$ ion. White, the hydrogen atoms of the two $H_2O$ molecules in complex with the $Gd^{3+}$. The representation was generated using the INSIGHT II software package. From [84].

prostate adenocarcinoma model that expresses c-Myc. The structure of the gadolinium complex is not detailed, and the stability of the compound in vivo requires further evaluation.

## 6 MOLECULAR IMAGING WITH ACTIVATED MR CONTRAST AGENTS

The major idea that drives the development of activated MR probes is to gain the ability to control relaxation properties of the CA by chemical modification of its structure or by certain external action. A broad range of potential applications of such CA can be likened to the use of the cleavable optical probe technology, where fluorescence of the probe can be controlled by changing the intermolecular distance between the fluorescent and the quenching group [85, 86].

### Enzymatic Activation of CA

One strategy to design target-specific MR contrast agents is to use a specific enzymatic reaction to activate the CA by increasing its relaxivity. The approach was originally reported by Louie et al. [87]. They designed "smart" relaxation agents in which the access of water to the first coordination sphere of a chelated paramagnetic ion is blocked with a substrate, galactopyranose, that can be removed by enzymatic cleavage by a traditional reporter gene, β-galactosidase. Following the cleavage of the blocking sugar group, the paramagnetic ion can interact directly with water protons to reduce their T1 by the inner sphere (contact) relaxation with the corresponding increase in the MR signal. Analogous approaches can be used to design CA with polypeptide blocking chains to probe protease activity. The enzyme β-galactosidase is expressed in the cell cytoplasm, therefore, the CA should be either microinjected to the cell(s) [87] or delivered to the cells by means of a certain cellular transport system such as amphiphilic Tat-type membrane translocation peptides [82]. This method can provide important information about expression patterns of endogenous enzymes or exogenous reporter genes expressed under control of a promoter region from the target gene.

A significant increase in T2 relaxivity is observed upon hybridization of the individual monodisperse iron oxide nanoparticles at the target site [88, 89]. This novel method of contrast enhancement was successfully used to study the processes of DNA cleavage or hybridization in vitro in cell-free systems [88, 90]. Cleavage of synthetic CLIO–oligonucleotide complexes with BamHI and DpnI restrictases significantly prolonged T2 relaxation time of the

samples. MRI probes for detection activity of protease enzymes was developed using a similar approach [91].

### CEST and PARACEST Contrast Agents

This new class of MR contrast agent is designed so that selective irradiation of an exchangeable proton(s) of these CA with an radio-frequency (RF) field induces a drop in the MR signal of bulk water because of transfer of the saturated magnetization from the irradiated protons to water protons by chemical exchange. Originally proposed by the group of Balaban et al. [92], CEST (chemical exchange saturation transfer) contrast was demonstrated for exchangeable amide protons of proteins [93], imino, and hydroxyl protons [94]. A significant increase in the sensitivity of detection can be gained if a polymer with a large number of equivalent exchangeable groups is used. In experiments with polyuridine (2000 uridine units) the sensitivity gain of more than 5000 per imino-proton was demonstrated, and the CA could be detected at concentration as low as 5 μM [94]. The efficiency of CEST contrast increases as the frequency offset, $\Delta\omega$, between resonance frequency of the exchangeable group and water protons increases. The frequency difference can be significantly increased from several parts per million to more than 20 ppm by attaching a paramagnetic shift reagent to the CEST agent. Experiments with a model poly-L-Arginine/Tm(HDOTP)$^{4-}$ system lowered the detection limit to 2.8 μM in a phantom at 7T magnetic field [95]. This modification of the method was named PARACEST for paramagnetic enhanced CEST. The major potential advantages of the CEST/PARACEST methods to generate MR contrast are: (1) the ability to "turn" the contrast on and off by applying saturating RF field at the resonance frequency of the exchangeable group; (2) high sensitivity that can be achieved by using polymer probes and/or PARACEST agents with large $\Delta\omega$ shifts and fast exchange rates; (3) certain types of CEST agents such as poly-amino acids can be expressed in vivo by target cells with an appropriate reporter vector; and (4) it is possible to design PARACEST probes with different $\Delta\omega$ values and image them independently using RF saturation field with the appropriate frequency offset.

The CEST method was used in vivo for endogenous amide protons of proteins and peptides to detect amide proton transfer (APT) in rat brain tumors [96]. A typical example of APT imaging is shown in Figure 8.

## 7 MOLECULAR MRI AND MR SPECTROSCOPY

Generally, MR spectroscopic imaging [chemical shift imaging (CSI), [97]] can be considered a true molecular imaging modality as it provides selective images of distribution of specific chemical compounds such as metabolites, exogenous substances and drugs.

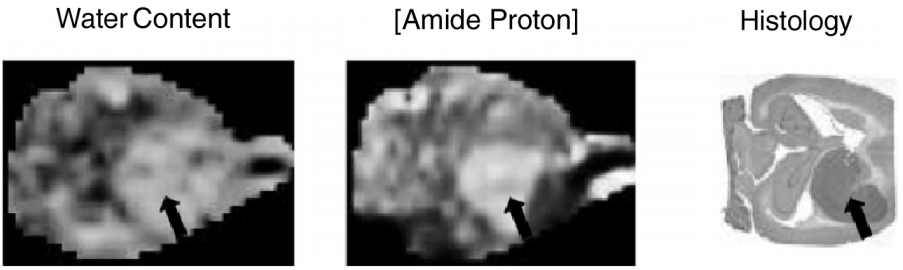


**Figure 8** Comparison of APT images with MR images and histology for a rat brain tumor. The APT-weighted image was obtained by subtracting the $S_{sat}/S_0$ images acquired at frequency-labeling offsets of ±3.5 ppm (16 scans). The tumor is visible in all of the MR images (arrow), but its contour is much clearer in the APT images. Adapted from [96].

## CSI Molecular Imaging with Reporter Molecules

In the following section we will focus on molecular aspects of imaging probes. These substances are typically designed in such a way that modification of the chemical structure of the probe induces a significant shift in the resonance frequency that can be detected by CSI. In comparison with contrast-enhanced MRI, where micromolar concentrations of CA can be detected, the low sensitivity of MR spectroscopy limits the detectable range of CSI agent concentrations at millimolar levels for protons and 19F MRS. For other nuclei such as 31P and 13C significantly higher concentrations are required.

A T1 relaxation-activated contrast agent to detect the enzyme β-galactosidase, a widely used reporter enzyme, was discussed in the previous section [87]. MR spectroscopy can also be used to assess the function of the enzyme by measuring changes in the chemical shift of a molecular probe induced by its enzymatic cleavage. 4-Fluoro-2-nitrophenyl-*b*-D-galactopyranoside was used as a prototype reporter molecule for β-galactosidase [98]. Upon the cleavage of the sugar by the enzyme, the chemical shift of 19F nucleus on the nitrophenyl ring changes by 4 to 8 ppm depending on the pH of the environment. A relatively high sensitivity of 19F MRS that approaches proton MRS, and the absence of background signals can render this method applicable to in vivo applications.

Expression of creatine kinase as a model marker gene was detected by in vivo MRS in mouse liver. Mice were injected in the tail vein with an adenoviral vector carrying the murine creatine kinase (*CK-B*) gene that induced expression of the gene in the mouse liver. Animals were given exogenous creatine, and 31P spectroscopy of the liver was performed in vivo. Production of phosphocreatine was only detected in the liver of mice injected with the *CK-B*. No spatial selection was used for MR spectroscopy apart from placing the RF surface coil onto exposed liver and shielding the surrounding tissue with a copper shield [99]. Generally the spatial resolution of 31P CSI is on the order

of about 10 $cm^3$, which can be a serious limitation for practical applications of this technique.

## 8 DISCUSSION

Magnetic resonance imaging provides a very high spatial resolution that is invaluable for morphological reference in small animals. Dynamic MRI with small molecular weight and/or nonspecific high-molecular-weight intravascular agents enables measurements of perfusion parameters that can be indicators of angiogenesis and also of tumor response to therapy. Significant effort has been concentrated on the development of targeted MR probes for truly specific molecular MRI that can differentiate between molecular targets on the cellular and subcellular levels. In spite of significant progress, several issues should be addressed to enable routine use of these imaging technologies in mainstream radiology. Some of these are outlined here.

The low sensitivity of MRI is a serious obstacle for imaging sparse molecular targets that are present at micromolar concentrations in vivo. Since the sensitivity of MR spectroscopy is in the millimolar range, efficient mechanisms of signal enhancement such as enzyme-coupled reactions have to be used to enable in vivo application of MRS for detection of molecular targets. For contrast-enhanced MRI highly efficient targeted imaging agents have to be used. For an optimized MR relaxation contrast agent the efficiency or relaxivity of the agent generally increases with the molecular size of the agent. Therefore very large CA, such as polymerized liposomes or nanoparticles, are very efficient probes for imaging vascular targets. On the other hand, to efficiently target solid tumors, the CA molecules should traverse several diffusion barriers to reach the imaging target. Tumors are characterized by highly permeable nonmature vasculature with a pore cutoff size in the range between 380 and 780 nm [26]. However, as discussed above, the delivery pattern in tumors is very nonuniform, and highly permeable areas within a tumor are often interlaced with areas of low vascular permeability and inefficient delivery [100]. Albumin–GdDTPA, blood-pool contrast agent with a molecular weight of 80 kDa leaks into the tumor interstitium in several tumor models, but not efficiently in well-vascularized viable regions [7]. Generation-4 $^{153}$Gd-folate dendrimer with a molecular weight in the range of 70 kDa demonstrated high accumulation in the high-affinity folate receptor-positive tumors [101]. On the other hand, generation-6 and higher dendrimers (MW of 240 kDa and above) are retained within the tumor vasculature [102]. Therefore, for the effective delivery of a CA to the nonvascular imaging target, its molecular weight should be restricted at the level of about 100 kDa, which corresponds to generation-4 or -5 dendrimers [14, 103] and protein carriers such as bovine serum albumin (BSA) with multiple modification sites.

Stability and in vivo pharmacokinetics of the probe molecules are important issues for the design of an efficient targeted CA. Paramagnetic CA

generally use stable Gd complexes, as their biodegradation results in a release of toxic $Gd^{3+}$ ions [104, 105]. Linear and macrocyclic chelates (DTPA and DOTA) can be used to form stable Gd complexes, and their derivatives, suitable for modification of macromolecules with Gd chelates, are commercially available (Macrocyclics, Dallas, Texas). CA delivery to cellular targets also depends on the plasma kinetics of the CA. Contrast agents with longer circulation time generally have favorable kinetics at the target site. In vivo CA clearance depends upon molecular size and surface characteristics of the agent [26], and several strategies can be used to prolong its plasma half-life time. Coating the iron core with a dextran shell results in longer circulation of iron oxide particles (LCDIO) [106]; incorporation of hydrophilic polymers such as polyethyleneglycol (PEG) to the coat reduces binding of plasma proteins and reduces macrophage uptake of the agent [21, 107]. Modification of the dendrimer surface with anionic groups reduces clearance and cytotoxicity of dendrimers in vivo [108], although both anionic and cationic surfaces showed significant liver accumulation [101].

Due to the low sensitivity of MRI to generate imaging contrast, it is necessary to build maximum local concentrations of the targeted CA at the cellular target site. Therefore, practically all available molecular targets are going to be occupied by the CA molecules. If the targets are functional cell surface receptors, then decorating them with CA molecules can interfere with normal cell signaling and may induce significant physiological response. For instance, using Herceptin as a targeting agent for HER-2/*neu* receptors can produce a therapeutic effect in HER-2/*neu* expressing tumors [109].

To provide efficient selective targeting, highly specific bispecific probes are required that can (1) recognize the molecular epitopes and (2) provide high-affinity binding sites for the CA molecule(s). Monoclonal antibodies are traditionally used for in vivo targeting. However, probes with a smaller molecular size such as mAb fragments, minibodies, diabodies, or peptides can provide better in vivo performance due to the improved delivery properties in solid tissues [69]. Biotin–avidin (streptavidin) linker system provides very high affinity for prelabeling applications but has limited applications in vivo mostly due to the induction of the host immune response against these foreign proteins [69].

For multistep labeling experiments with pretargeting of the epitopes with specific probes [110] and developing of MR contrast with CA that recognize these probes, special attention should be focused on optimized timing of imaging experiment. The available time window for a specific marker (such as mAb) on the cell surface can be limited by dissociation and internalization of the marker by endocytosis. Thus the reported half-life time of a radiolabeled mAb HER-2/*neu* receptor complex on the cell surface can be as short as 6 h [111]. On the other hand the circulating probes should be completely cleared from the system before administration of the contrast agent to provide low background in the images. Therefore, to maximize target–background contrast, a compromise between the clearance of circulating mAb and the

availability of the receptor-bound mAb on the cell surface is required. An alternative approach may include removal of circulating targeting probes by a chasing agent [67].

In terms of optimizing MRI protocols, different requirements for T1 and T2 contrast agents need to be considered. Iron-oxide-based contrast agents such as SPIO or MION nanopartricles generally provide very high T2 relaxivity; however, their delivery to the target site can be significantly restricted. Another problem with this class of agent is the negative contrast (or reduced signal intensity) produced in MR images. Recently, a novel method for MR acquisition that provides positive signal enhancement with T2 CA was proposed to improve the in vivo performance of these agents [71]. T1 contrast agents can be designed to have an optimal molecular size [112], and they produce intrinsic positive contrast in T1-weighted MRI. Modern fast acquisition MR techniques enable measuring of quantitative T1 maps [7, 113] to quantify the contrast enhancement and concentration of the molecular targets, respectively.

The rapid development of novel contrast agents will greatly improve our ability to follow modulation in vascular perfusion and molecular composition in preclinical animal models of human cancer using high-resolution MR imaging. In comparison with the competing optical and nuclear imaging modalities such as fluorescence and bioluminescence imaging, SPECT, and PET imaging, MRI provides high-resolution three-dimensional information with excellent morphological registration and does not require radioactive exposure. Clinical applications of molecular target-specific MR agents on the other side are less apparent. Major problems include high concentration of the agent required for the measurable contrast and related potential toxicity and immune response effects. The lower magnetic field used for clinical MRI also contributes to the reduced sensitivity of detection that is of paramount importance for molecular MRI. Significant progress has been made in clinical tissue-specific MR contrast agents that provide better delineation of the diseased organs [114]. However, full-scale clinical applications of molecular targeted MRI in cancer detection and therapy may not be immediately available.

***Acknowledgments*** We thank Dr. Jeff Bulte for help with preparation of the chapter. This work was supported by P50 CA103175.

## REFERENCES

1. Bhujwalla, Z. M., Artemov, D., Natarajan, K., Solaiyappan, M., Kollars, P., Kristjansen, P. E. (2003). Reduction of vascular and permeable regions in solid tumors detected by macromolecular contrast magnetic resonance imaging after treatment with antiangiogenic agent TNP-470. *Clin. Cancer Res.*, *9*, 355–362.

2. Bhujwalla, Z. M., Blackband, S. J., Wehrle, J. P., Grossman, S., Eller, S., Glickson, J. D. (1990). Metabolic heterogeneity in RIF-1 tumours detected in vivo by 31P NMR spectroscopy. *NMR Biomed.*, *3*, 233–238.
3. Aboagye, E. O., Artemov, D., Senter, P., Bhujwalla, Z. M. (1998). Intratumoral conversion of 5-fluorocytosine to 5-fluorouracil by monoclonal antibody-cytosine deaminase conjugates: Noninvasive detection of prodrug activation by magnetic resonance spectroscopy and spectroscopic imaging. *Cancer Res.*, *58*, 4075–4078.
4. Hakumaki, J. M., Brindle, K. M. (2003). Techniques: Visualizing apoptosis using nuclear magnetic resonance. *Trends Pharmacol. Sci.*, *24*, 146–149.
5. Hakumaki, J. M., Kauppinen, R. A. (2000). 1H NMR visible lipids in the life and death of cells. *Trends Biochem. Sci.*, *25*, 357–362.
6. Mansfield, P., Morris, P. G., Ordidge, R. J., Pykett, I. L., Bangert, V., Coupland, R. E. (1980). Human whole body imaging and detection of breast tumours by n.m.r. *Philos. Trans. R. Soc. Lond. B Biol. Sci.*, *289*, 503–510.
7. Bhujwalla, Z. M., Artemov, D., Natarajan, K., Ackerstaff, E., Solaiyappan, M. (2000). Vascular differences detected by MRI for metastatic versus nonmetastatic breast and prostate cancer xenografts. *Neoplasia*, *3*, 1–11.
8. Schmiedl, U., Ogan, M., Paajanen, H., Marotti, M., Crooks, L. E., Brito, A. C., Brasch, R. C. (1987). Albumin labeled with Gd-DTPA as an intravascular, blood pool-enhancing agent for MR imaging: Biodistribution and imaging studies. *Radiology*, *162*, 205–210.
9. Bogdanov, A. A., Jr., Weissleder, R., Frank, H. W., Bogdanova, A. V., Nossif, N., Schaffer, B. K., Tsai, E., Papisov, M. I., Brady, T. J. (1993). A new macromolecule as a contrast agent for MR angiography: Preparation, properties, and animal studies. *Radiology*, *187*, 701–706.
10. Gohr-Rosenthal, S., Schmitt-Willich, H., Ebert, W., Conrad, J. (1993). The demonstration of human tumors on nude mice using gadolinium-labelled monoclonal antibodies for magnetic resonance imaging. *Invest. Radiol.*, *28*, 789–795.
11. Artemov, D., Mori, N., Ravi, R., Bhujwalla, Z. M. (2003). Magnetic resonance molecular imaging of the Her-2/neu receptor. *Cancer Res.*, *63*, 2723–2727.
12. Shahbazi-Gahrouei, D., Williams, M., Rizvi, S., Allen, B. J. (2001). In vivo studies of Gd-DTPA-monoclonal antibody and gd-porphyrins: Potential magnetic resonance imaging contrast agents for melanoma. *J. Magn. Reson. Imag.*, *14*, 169–174.
13. Matsumura, A., Shibata, Y., Nakagawa, K., Nose, T. (1994). MRI contrast enhancement by Gd-DTPA-monoclonal antibody in 9L glioma rats. *Acta Neurochir. Suppl.*, *60*, 356–358.
14. Bryant, L. H., Jr., Brechbiel, M. W., Wu, C., Bulte, J. W., Herynek, V., Frank, J. A. (1999). Synthesis and relaxometry of high-generation ($G$ = 5, 7, 9, and 10) PAMAM dendrimer-DOTA-gadolinium chelates. *J. Magn. Reson. Imag.*, *9*, 348–352.
15. Kobayashi, H., Sato, N., Hiraga, A., Saga, T., Nakamoto, Y., Ueda, H., Konishi, J., Togashi, K., Brechbiel, M. W. (2001). 3D-micro-MR angiography of mice using macromolecular MR contrast agents with polyamidoamine dendrimer core with reference to their pharmacokinetic properties. *Magn. Reson. Med.*, *45*, 454–460.
16. Anderson, S. A., Rader, R. K., Westlin, W. F., Null, C., Jackson, D., Lanza, G. M., Wickline, S. A., Kotyk, J. J. (2000). Magnetic resonance contrast enhancement of

neovasculature with alpha(v)beta(3)-targeted nanoparticles. *Magn. Reson. Med.*, *44*, 433–439.

17. Sipkins, D. A., Cheresh, D. A., Kazemi, M. R., Nevin, L. M., Bednarski, M. D., Li, K. C. (1998). Detection of tumor angiogenesis in vivo by alphaVbeta3-targeted magnetic resonance imaging. *Nat. Med.*, *4*, 623–626.
18. Caravan, P., Cloutier, N. J., Greenfield, M. T., McDermid, S. A., Dunham, S. U., Bulte, J. W., Amedio, J. C., Jr., Looby, R. J., Supkowski, R. M., Horrocks, W. D., Jr., McMurry, T. J., Lauffer, R. B. (2002). The interaction of MS-325 with human serum albumin and its effect on proton relaxation rates. *J. Am. Chem. Soc.*, *124*, 3152–3162.
19. Gillis, P., Roch, A., Brooks, R. A. (1999). Corrected equations for susceptibility-induced T2-shortening. *J. Magn. Reson.*, *137*, 402–407.
20. Hogemann, D., Josephson, L., Weissleder, R., Basilion, J. P. (2000). Improvement of MRI probes to allow efficient detection of gene expression. *Bioconjug. Chem.*, *11*, 941–946.
21. Bulte, J. W., de Cuyper, M., Despres, D., Frank, J. A. (1999). Short- vs. long-circulating magnetoliposomes as bone marrow-seeking MR contrast agents. *J. Magn. Reson. Imag.*, *9*, 329–335.
22. Strable, E., Bulte, J. W., Moskowitz, B. M., Vivekanandan, K., Allen, M., Douglas, T. (2001). Synthesis and characterization of soluble iron oxide-dendrimer composites. *Chem. Mater.*, *13*, 2201–2209.
23. Bulte, J. W., Douglas, T., Witwer, B., Zhang, S. C., Strable, E., Lewis, B. K., Zywicke, H., Miller, B., van Gelderen, P., Moskowitz, B. M., Duncan, I. D., Frank, J. A. (2001). Magnetodendrimers allow endosomal magnetic labeling and in vivo tracking of stem cells. *Nat. Biotechnol.*, *19*, 1141–1147.
24. Bulte, J. W., Brooks, R. A., Moskowitz, B. M., Bryant, L. H., Jr., Frank, J. A. (1999). Relaxometry and magnetometry of the MR contrast agent MION-46L. *Magn. Reson. Med.*, *42*, 379–384.
25. Bulte, J. W., Bryant, L. H., Jr., Frank, J. A. (2003). In L. E. Feinendegen, W. W. Shreeve, W. C. Eckelman, Y.-W. Bahk, and H. N. Wagner, Jr. (Eds.), *Molecular Nuclear Medicine—The Challenge of Genomics and Proteomics to Clinical Practice*. Springer, pp. 721–740.
26. Brigger, I., Dubernet, C., Couvreur, P. (2002). Nanoparticles in cancer therapy and diagnosis. *Adv. Drug Deliv. Rev.*, *54*, 631–651.
27. Weidner, N. (1993). Tumor angiogenesis: Review of current applications in tumor prognostication. *Semin. Diagn. Pathol.*, *10*, 302–313.
28. Durand, R. E., Aquino-Parsons, C. (2001). Clinical relevance of intermittent tumour blood flow. *Acta Oncol.*, *40*, 929–936.
29. Durand, R. E. (2001). Intermittent blood flow in solid tumours—An under-appreciated source of "drug resistance". *Cancer Metastasis Rev.*, *20*, 57–61.
30. Furman-Haran, E., Margalit, R., Grobgeld, D., Degani, H. (1996). Dynamic contrast-enhanced magnetic resonance imaging reveals stress-induced angiogenesis in MCF7 human breast tumors. *Proc. Natl. Acad. Sci. USA*, *93*, 6247–6251.
31. Aronen, H. J., Cohen, M. S., Belliveau, J. W., Fordham, J. A., Rosen, B. R. (1993). Ultrafast imaging of brain tumors. *Top. Magn. Reson. Imag.*, *5*, 14–24.

32. Hawighorst, H., Knapstein, P. G., Weikel, W., Knopp, M. V., Zuna, I., Knof, A., Brix, G., Schaeffer, U., Wilkens, C., Schoenberg, S. O., Essig, M., Vaupel, P., van Kaick, G. (1997). Angiogenesis of uterine cervical carcinoma: Characterization by pharmacokinetic magnetic resonance parameters and histological microvessel density with correlation to lymphatic involvement. *Cancer Res.*, *57*, 4777–4786.
33. Tofts, P. S., Brix, G., Buckley, D. L., Evelhoch, J. L., Henderson, E., Knopp, M. V., Larsson, H. B., Lee, T. Y., Mayr, N. A., Parker, G. J., Port, R. E., Taylor, J., Weisskoff, R. M. (1999). Estimating kinetic parameters from dynamic contrast-enhanced T(1)-weighted MRI of a diffusable tracer: Standardized quantities and symbols. *J. Magn. Reson. Imag.*, *10*, 223–232.
34. Tofts, P. S. (1997). Modeling tracer kinetics in dynamic Gd-DTPA MR imaging. *J. Magn. Reson. Imaging.*, *7*, 91–101.
35. Kety, S. S. (1965). Measurement of local contribution within the brain by means of inert, diffusible tracers; examination of the theory, assumptions and possible sources of error. *Acta Neurol. Scand. Suppl.*, *14*, 20–23.
36. Larsson, H. B., Stubgaard, M., Frederiksen, J. L., Jensen, M., Henriksen, O., Paulson, O. B. (1990). Quantitation of blood-brain barrier defect by magnetic resonance imaging and gadolinium-DTPA in patients with multiple sclerosis and brain tumors. *Magn. Reson. Med.*, *16*, 117–131.
37. Weisskoff, R. M., Chesler, D., Boxerman, J. L., Rosen, B. R. (1993). Pitfalls in MR measurement of tissue blood flow with intravascular tracers: Which mean transit time? *Magn. Reson. Med.*, *29*, 553–558.
38. Donahue, K. M., Krouwer, H. G., Rand, S. D., Pathak, A. P., Marszalkowski, C. S., Censky, S. C., Prost, R. W. (2000). Utility of simultaneously acquired gradient-echo and spin-echo cerebral blood volume and morphology maps in brain tumor patients. *Magn. Reson. Med.*, *43*, 845–853.
39. Pathak, A. P., Schmainda, K. M., Ward, B. D., Linderman, J. R., Rebro, K. J., Greene, A. S. (2001). MR-derived cerebral blood volume maps: Issues regarding histological validation and assessment of tumor angiogenesis. *Magn. Reson. Med.*, *46*, 735–747.
40. Boxerman, J. L., Hamberg, L. M., Rosen, B. R., Weisskoff, R. M. (1995). MR contrast due to intravascular magnetic susceptibility perturbations. *Magn. Reson. Med.*, *34*, 555–566.
41. Deane, B. R., Lantos, P. L. (1981). The vasculature of experimental brain tumours. Part 1. A sequential light and electron microscope study of angiogenesis. *J. Neurol. Sci.*, *49*, 55–66.
42. Deane, B. R., Lantos, P. L. (1981). The vasculature of experimental brain tumours. Part 2. A quantitative assessment of morphological abnormalities. *J. Neurol. Sci.*, *49*, 67–77.
43. Hnatowich, D. J., Layne, W. W., Childs, R. L. (1982). The preparation and labeling of DTPA-coupled albumin. *Int. J. Appl. Radiat. Isot.*, *33*, 327–332.
44. Donahue, K. M., Weisskoff, R. M., Chesler, D. A., Kwong, K. K., Bogdanov, A. A., Jr., Mandeville, J. B., Rosen, B. R. (1996). Improving MR quantification of regional blood volume with intravascular T1 contrast agents: Accuracy, precision, and water exchange. *Magn. Reson. Med.*, *36*, 858–867.
45. Bhujwalla, Z. M., Artemov, D., Glockner, J. (1999). Tumor angiogenesis, vascularization, and contrast-enhanced magnetic resonance imaging. *Top. Magn. Reson. Imag.*, *10*, 92–103.

46. Teicher, B. A., Emi, Y., Kakeji, Y., Northey, D. (1996). TNP-470/minocycline/cytotoxic therapy: A systems approach to cancer therapy. *Eur. J. Cancer*, *32A*, 2461–2466.
47. Folkman, J. (1985). Tumor angiogenesis. *Adv. Cancer Res.*, *43*, 175–203.
48. Folkman, J. (1971). Tumor angiogenesis: Therapeutic implications. *N. Engl. J. Med.*, *285*, 1182–1186.
49. Brooks, P. C., Clark, R. A., Cheresh, D. A. (1994). Requirement of vascular integrin alpha v beta 3 for angiogenesis. *Science*, *264*, 569–571.
50. Kang, H. W., Josephson, L., Petrovsky, A., Weissleder, R., Bogdanov, A., Jr. (2002). Magnetic resonance imaging of inducible E-selectin expression in human endothelial cell culture. *Bioconjug. Chem.*, *13*, 122–127.
51. Winter, P. M., Caruthers, S. D., Kassner, A., Harris, T. D., Chinen, L. K., Allen, J. S., Lacy, E. K., Zhang, H., Robertson, J. D., Wickline, S. A., Lanza, G. M. (2003). Molecular imaging of angiogenesis in nascent Vx-2 rabbit tumors using a novel alpha(nu)beta3-targeted nanoparticle and 1.5 tesla magnetic resonance imaging. *Cancer Res.*, *63*, 5838–5843.
52. Josephson, L., Groman, E. V., Menz, E., Lewis, J. M., Bengele, H. (1990). A functionalized superparamagnetic iron oxide colloid as a receptor directed MR contrast agent. *Magn. Reson. Imag.*, *8*, 637–646.
53. Weissleder, R., Reimer, P., Lee, A. S., Wittenberg, J., Brady, T. J. (1990). MR receptor imaging: Ultrasmall iron oxide particles targeted to asialoglycoprotein receptors. *AJR Am. J. Roentgenol.*, *155*, 1161–1167.
54. Bulte, J. W., Hoekstra, Y., Kamman, R. L., Magin, R. L., Webb, A. G., Briggs, R. W., Go, K. G., Hulstaert, C. E., Miltenyi, S., The, T. H. (1992). Specific MR imaging of human lymphocytes by monoclonal antibody-guided dextran-magnetite particles. *Magn. Reson. Med.*, *25*, 148–157.
55. Artemov, D., Mori, N., Okollie, B., Bhujwalla, Z. M. (2003). MR molecular imaging of the Her-2/neu receptor in breast cancer cells using targeted iron oxide nanoparticles. *Magn. Reson. Med.*, *49*, 403–408.
56. Weissleder, R., Moore, A., Mahmood, U., Bhorade, R., Benveniste, H., Chiocca, E. A., Basilion, J. P. (2000). In vivo magnetic resonance imaging of transgene expression. *Nat. Med.*, *6*, 351–355.
57. Moore, A., Basilion, J. P., Chiocca, E. A., Weissleder, R. (1998). Measuring transferrin receptor gene expression by NMR imaging. *Biochim. Biophys. Acta*, *1402*, 239–249.
58. Anderson-Berg, W. T., Strand, M., Lempert, T. E., Rosenbaum, A. E., Joseph, P. M. (1986). Nuclear magnetic resonance and gamma camera tumor imaging using gadolinium-labeled monoclonal antibodies. *J. Nucl. Med.*, *27*, 829–833.
59. Unger, E. C., Totty, W. G., Neufeld, D. M., Otsuka, F. L., Murphy, W. A., Welch, M. S., Connett, J. M., Philpott, G. W. (1985). Magnetic resonance imaging using gadolinium labeled monoclonal antibody. *Invest. Radiol.*, *20*, 693–700.
60. Konda, S. D., Aref, M., Brechbiel, M., Wiener, E. C. (2000). Development of a tumor-targeting MR contrast agent using the high-affinity folate receptor: Work in progress. *Invest. Radiol.*, *35*, 50–57.
61. Weissleder, R., Lee, A. S., Fischman, A. J., Reimer, P., Shen, T., Wilkinson, R., Callahan, R. J., Brady, T. J. (1991). Polyclonal human immunoglobulin G labeled with polymeric iron oxide: Antibody MR imaging. *Radiology*, *181*, 245–249.

62. Zhao, M., Beauregard, D. A., Loizou, L., Davletov, B., Brindle, K. M. (2001). Non-invasive detection of apoptosis using magnetic resonance imaging and a targeted contrast agent. *Nat. Med.*, *7*, 1241–1244.

63. Moore, A., Medarova, Z., Potthast, A., Dai, G. (2004). In vivo targeting of under-glycosylated MUC-1 tumor antigen using a multimodal imaging probe. *Cancer Res.*, *64*, 1821–1827.

64. Monsky, W. L., Fukumura, D., Gohongi, T., Ancukiewcz, M., Weich, H. A., Torchilin, V. P., Yuan, F., Jain, R. K. (1999). Augmentation of transvascular transport of macromolecules and nanoparticles in tumors using vascular endothelial growth factor. *Cancer Res.*, *59*, 4129–4135.

65. Paganelli, G., Grana, C., Chinol, M., Cremonesi, M., De Cicco, C., De Braud, F., Robertson, C., Zurrida, S., Casadio, C., Zoboli, S., Siccardi, A. G., Veronesi, U. (1999). Antibody-guided three-step therapy for high grade glioma with yttrium-90 biotin. *Eur. J. Nucl. Med.*, *26*, 348–357.

66. Artemov, D., Foss, C., Okollie, B., Bhujwalla, Z. M. (2004). In *The Society for Molecular Imaging, the Third Annual Meeting*, St. louis. p. 386.

67. Dafni, H., Israely, T., Bhujwalla, Z. M., Benjamin, L. E., Neeman, M. (2002). Overexpression of vascular endothelial growth factor 165 drives peritumor interstitial convection and induces lymphatic drain: Magnetic resonance imaging, confocal microscopy, and histological tracking of triple-labeled albumin. *Cancer Res.*, *62*, 6731–6739.

68. Ichikawa, T., Hogemann, D., Saeki, Y., Tyminski, E., Terada, K., Weissleder, R., Chiocca, E. A., Basilion, J. P. (2002). MRI of transgene expression: Correlation to therapeutic gene expression. *Neoplasia*, *4*, 523–530.

69. Goldenberg, D. M., Chang, C. H., Sharkey, R. M., Rossi, E. A., Karacay, H., McBride, W., Hansen, H. J., Chatal, J. F., Barbet, J. (2003). Radioimmunotherapy: Is avidin-biotin pretargeting the preferred choice among pretargeting methods? *Eur. J. Nucl. Med. Mol. Imag.*, *30*, 777–780.

70. Kobayashi, H., Sato, N., Kawamoto, S., Saga, T., Hiraga, A., Haque, T. L., Ishimori, T., Konishi, J., Togashi, K., Brechbiel, M. W. (2001). Comparison of the macromolecular MR contrast agents with ethylenediamine-core versus ammonia-core generation-6 polyamidoamine dendrimer. *Bioconjug. Chem.*, *12*, 100–107.

71. Seppenwoolde, J. H., Viergever, M. A., Bakker, C. J. (2003). Passive tracking exploiting local signal conservation: The white marker phenomenon. *Magn. Reson. Med.*, *50*, 784–790.

72. Bogdanov, A., Jr., Matuszewski, L., Bremer, C., Petrovsky, A., Weissleder, R. (2002). Oligomerization of paramagnetic substrates result in signal amplification and can be used for MR imaging of molecular targets. *Mol. Imag.*, *1*, 16–23.

73. Walter, G. A., Cahill, K. S., Huard, J., Feng, H., Douglas, T., Sweeney, H. L., Bulte, J. W. (2004). Noninvasive monitoring of stem cell transfer for muscle disorders. *Magn. Reson. Med.*, *51*, 273–277.

74. Karmarkar, P. V., Kraitchman, D. L., Izbudak, I., Hofmann, L. V., Amado, L. C., Fritzges, D., Young, R., Pittenger, M., Bulte, J. W., Atalar, E. (2004). MR-trackable intramyocardial injection catheter. *Magn. Reson. Med.*, *51*, 1163–1172.

75. Lee, I. H., Bulte, J. W., Schweinhardt, P., Douglas, T., Trifunovski, A., Hofstetter, C., Olson, L., Spenger, C. (2004). In vivo magnetic resonance tracking of

olfactory ensheathing glia grafted into the rat spinal cord. *Exp. Neurol.*, *187*, 509–516.

76. Bulte, J. W., Arbab, A. S., Douglas, T., Frank, J. A. (2004). Preparation of magnetically labeled cells for cell tracking by magnetic resonance imaging. *Methods Enzymol.*, *386*, 275–299.
77. Weissleder, R., Cheng, H. C., Bogdanova, A., Bogdanov, A. (1997). Magnetically labeled cells can be detected by MR imaging. *J. Magn. Reson. Imag.*, *7*, 258–263.
78. Dodd, S. J., Williams, M., Suhan, J. P., Williams, D. S., Koretsky, A. P., Ho, C. (1999). Detection of single mammalian cells by high-resolution magnetic resonance imaging. *Biophys. J.*, *76*, 103–109.
79. Shapiro, E. M., Skrtic, S., Sharer, K., Hill, J. M., Dunbar, C. E., Koretsky, A. P. (2004). MRI detection of single particles for cellular imaging. *Proc. Natl. Acad. Sci. USA*, *101*, 10901–10906.
80. Gammon, S. T., Villalobos, V. M., Prior, J. L., Sharma, V., Piwnica-Worms, D. (2003). Quantitative analysis of permeation peptide complexes labeled with technetium-99m: Chiral and sequence-specific effects on net cell uptake. *Bioconjug. Chem.*, *14*, 368–376.
81. Prantner, A. M., Sharma, V., Garbow, J. R., Piwnica-Worms, D. (2003). Synthesis and characterization of a Gd-DOTA-D-permeation peptide for magnetic resonance relaxation enhancement of intracellular targets. *Mol. Imag.*, *2*, 333–341.
82. Bhorade, R., Weissleder, R., Nakakoshi, T., Moore, A., Tung, C. H. (2000). Macrocyclic chelators with paramagnetic cations are internalized into mammalian cells via a HIV-tat derived membrane translocation peptide. *Bioconjug. Chem.*, *11*, 301–305.
83. Kircher, M. F., Mahmood, U., King, R. S., Weissleder, R., Josephson, L. (2003). A multimodal nanoparticle for preoperative magnetic resonance imaging and intraoperative optical brain tumor delineation. *Cancer Res.*, *63*, 8122–8125.
84. Heckl, S., Pipkorn, R., Waldeck, W., Spring, H., Jenne, J., von der Lieth, C. W., Corban-Wilhelm, H., Debus, J., Braun, K. (2003). Intracellular visualization of prostate cancer using magnetic resonance imaging. *Cancer Res.*, *63*, 4766–4772.
85. Mahmood, U., Tung, C. H., Bogdanov, A., Jr., Weissleder, R. (1999). Near-infrared optical imaging of protease activity for tumor detection. *Radiology*, *213*, 866–870.
86. Weissleder, R., Tung, C. H., Mahmood, U., Bogdanov, A., Jr. (1999). In vivo imaging of tumors with protease-activated near-infrared fluorescent probes. *Nat. Biotechnol.*, *17*, 375–378.
87. Louie, A. Y., Huber, M. M., Ahrens, E. T., Rothbacher, U., Moats, R., Jacobs, R. E., Fraser, S. E., Meade, T. J. (2000). In vivo visualization of gene expression using magnetic resonance imaging. *Nat. Biotechnol.*, *18*, 321–325.
88. Perez, J. M., O'Loughin, T., Simeone, F. J., Weissleder, R., Josephson, L. (2002). DNA-based magnetic nanoparticle assembly acts as a magnetic relaxation nanoswitch allowing screening of DNA-cleaving agents. *J. Am. Chem. Soc.*, *124*, 2856–2857.
89. Bremer, C., Weissleder, R. (2001). In vivo imaging of gene expression. *Acad. Radiol.*, *8*, 15–23.

90. Perez, J. M., Josephson, L., O'Loughlin, T., Hogemann, D., Weissleder, R. (2002). Magnetic relaxation switches capable of sensing molecular interactions. *Nat. Biotechnol.*, *20*, 816–820.

91. Zhao, M., Kircher, M. F., Josephson, L., Weissleder, R. (2002). Differential conjugation of tat peptide to superparamagnetic nanoparticles and its effect on cellular uptake. *Bioconjug. Chem.*, *13*, 840–844.

92. Guivel-Scharen, V., Sinnwell, T., Wolff, S. D., Balaban, R. S. (1998). Detection of proton chemical exchange between metabolites and water in biological tissues. *J. Magn. Reson.*, *133*, 36–45.

93. Zhou, J., Payen, J. F., Wilson, D. A., Traystman, R. J., van Zijl, P. C. (2003). Using the amide proton signals of intracellular proteins and peptides to detect pH effects in MRI. *Nat. Med.*, *9*, 1085–1090.

94. Snoussi, K., Bulte, J. W., Gueron, M., Van Zijl, P. C. (2003). Sensitive CEST agents based on nucleic acid imino proton exchange: Detection of poly(rU) and of a dendrimer-poly(rU) model for nucleic acid delivery and pharmacology. *Magn. Reson. Med.*, *49*, 998–1005.

95. Aime, S., Delli Castelli, D., Terreno, E. (2003). Supramolecular adducts between poly-L-arginine and [TmIIIdotp]: A route to sensitivity-enhanced magnetic resonance imaging-chemical exchange saturation transfer agents. *Angew Chem. Int. Ed. Engl.*, *42*, 4527–4529.

96. Zhou, J., Lal, B., Wilson, D. A., Laterra, J., van Zijl, P. C. (2003). Amide proton transfer (APT) contrast for imaging of brain tumors. *Magn. Reson. Med.*, *50*, 1120–1126.

97. Brown, T. R., Kincaid, B. M., Ugurbil, K. (1982). NMR chemical shift imaging in three dimensions. *Proc. Natl. Acad. Sci. USA*, *79*, 3523–3526.

98. Cui, W., Otten, P., Li, Y., Koeneman, K. S., Yu, J., Mason, R. P. (2004). Novel NMR approach to assessing gene transfection: 4-Fluoro-2-nitrophenyl-beta-D-galactopyranoside as a prototype reporter molecule for beta-galactosidase. *Magn. Reson. Med.*, *51*, 616–620.

99. Auricchio, A., Zhou, R., Wilson, J. M., Glickson, J. D. (2001). In vivo detection of gene expression in liver by 31P nuclear magnetic resonance spectroscopy employing creatine kinase as a marker gene. *Proc. Natl. Acad. Sci. USA*, *98*, 5205–5210.

100. Alexiou, C., Arnold, W., Hulin, P., Klein, R. J., Renz, H., Parak, F. G., Bergemann, C., Lubbe, A. S. (2001). Magnetic mitoxantrone nanoparticle detection by histology, X-ray and MRI after magnetic tumor targeting. *J. Magn. Magn. Mater.*, *225*, 187–193.

101. Konda, S. D., Wang, S., Brechbiel, M., Wiener, E. C. (2002). Biodistribution of a 153 Gd-folate dendrimer, generation = 4, in mice with folate-receptor positive and negative ovarian tumor xenografts. *Invest. Radiol.*, *37*, 199–204.

102. Kobayashi, H., Kawamoto, S., Saga, T., Sato, N., Hiraga, A., Konishi, J., Togashi, K., Brechbiel, M. W. (2001). Micro-MR angiography of normal and intratumoral vessels in mice using dedicated intravascular MR contrast agents with high generation of polyamidoamine dendrimer core: Reference to pharmacokinetic properties of dendrimer-based MR contrast agents. *J. Magn. Reson. Imag.*, *14*, 705–713.

103. Kobayashi, H., Kawamoto, S., Saga, T., Sato, N., Hiraga, A., Ishimori, T., Akita, Y., Mamede, M. H., Konishi, J., Togashi, K., Brechbiel, M. W. (2001). Novel liver

macromolecular MR contrast agent with a polypropylenimine diaminobutyl dendrimer core: Comparison to the vascular MR contrast agent with the polyamidoamine dendrimer core. *Magn. Reson. Med.*, *46*, 795–802.

104. Spencer, A. W. S., Harpur, E. (1998). Gadolinium chloride toxicity in the mouse. *Hum. Exp. Toxicol.*, *17*, 633–637.
105. Runge, V. M., Dickey, K. M., Williams, N. M., Peng, X. (2002). Local tissue toxicity in response to extravascular extravasation of magnetic resonance contrast media. *Invest. Radiol.*, *37*, 393–398.
106. Moore, A., Marecos, E., Bogdanov, A., Jr., Weissleder, R. (2000). Tumoral distribution of long-circulating dextran-coated iron oxide nanoparticles in a rodent model. *Radiology*, *214*, 568–574.
107. Oussoren, C., Storm, G. (1997). Lymphatic uptake and biodistribution of liposomes after subcutaneous injection: III. Influence of surface modification with poly(ethyleneglycol). *Pharm. Res.*, *14*, 1479–1484.
108. Malik, N., Wiwattanapatapee, R., Klopsch, R., Lorenz, K., Frey, H., Weener, J. W., Meijer, E. W., Paulus, W., Duncan, R. (2000). Dendrimers: Relationship between structure and biocompatibility in vitro, and preliminary studies on the biodistribution of 125I-labelled polyamidoamine dendrimers in vivo. *J. Control Release*, *65*, 133–148.
109. Pegram, M. D., Lipton, A., Hayes, D. F., Weber, B. L., Baselga, J. M., Tripathy, D., Baly, D., Baughman, S. A., Twaddell, T., Glaspy, J. A., Slamon, D. J. (1998). Phase II study of receptor-enhanced chemosensitivity using recombinant humanized anti-p185HER2/neu monoclonal antibody plus cisplatin in patients with HER2/neu-overexpressing metastatic breast cancer refractory to chemotherapy treatment. *J. Clin. Oncol.*, *16*, 2659–2671.
110. Rosebrough, S. F. (1996). Two-step immunological approaches for imaging and therapy. *Q. J. Nucl. Med.*, *40*, 234–251.
111. Zalutsky, M. R., Xu, F. J., Yu, Y., Foulon, C. F., Zhao, X. G., Slade, S. K., Affleck, D. J., Bast, R. C., Jr. (1999). Radioiodinated antibody targeting of the HER-2/neu oncoprotein: Effects of labeling method on cellular processing and tissue distribution. *Nucl. Med. Biol.*, *26*, 781–790.
112. Kobayashi, H., Brechbiel, M. W. (2003). Dendrimer-based macromolecular MRI contrast agents: Characteristics and application. *Mol. Imag.*, *2*, 1–10.
113. Nekolla, S., Gneiting, T., Syha, J., Deichmann, R., Haase, A. (1992). Tl maps by K-space reduced snapshot-FLASH MRI. *J. Comput. Assist. Tomogr.*, *16*, 327–332.
114. Weinmann, H. J., Ebert, W., Misselwitz, B., Schmitt-Willich, H. (2003). Tissue-specific MR contrast agents. *Eur. J. Radiol.*, *46*, 33–44.

# TARGETS AND APPROACHES FOR CANCER DRUG DISCOVERY

Susan L. Mooberry
*Southwest Foundation for Biomedical Research*
*San Antonio, Texas*

1 INTRODUCTION
Advent of Targeted Therapy
Evolution of the Drug Discovery Process
Focus and Definitions
2 SIGNAL TRANSDUCTION TARGETS
HER2
The EGF Receptor Pathway
Insulin-Like Growth Factor Signaling
Bcr/Abl, c-Kit, and Platelet-Derived Growth Factor
Mammalian Target of Rapamycin (mTOR)
Methods of Drug Discovery
3 TARGETING TUMOR VASCULATURE
Inhibitors of Angiogenesis
Antivascular Agents
Methods of Drug Discovery
4 DNA TARGETS
DNA Methylation
Histone Deacetylase Inhibitors
Telomerase Inhibitors
5 ANTIMITOTICS
Microtubule-Interacting Agents
Kinesin Inhibitors
Aurora Kinase
Methods of Drug Discovery

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

6 PROMISING NEW TARGETS
Proteasome Inhibition
Heat Shock Protein 90 (Hsp90)
7 SOURCES OF SMALL-MOLECULE DRUG LEADS
Natural Products
Synthetic Chemical Libraries
8 RESOURCES FOR DRUG DISCOVERY
References

## 1 INTRODUCTION

### Advent of Targeted Therapy

Although significant advances in early diagnosis and treatment of cancer have been made in the past decade, there remains a compelling need for more effective therapies for the treatment of many different types of cancer. The identification of more specific and thereby less toxic therapies for these tumors will provide significant advantages for cancer control. With the sequencing of the human genome and identification of numerous oncogenes and tumor suppressor genes, the level of understanding of the molecular mechanisms underlying cancer has never been greater. These advances in understanding the molecular etiology of cancer led to efforts to discover and develop target-specific therapies that offer the promise of more specificity toward the key biological processes that are abnormal in cancer. Importantly, some but not all of the targeted therapies have proven efficacious in clinical trials. Single-target therapy was the initial aim of targeted therapy, but the genetic instability intrinsic to tumors causes multiple oncogenic abnormalities. Broader targeting of multiple molecular alterations that cause cancer may provide for optimum therapy for most tumors.

Along with an enhanced understanding of the molecular events that lead to tumor initiation and expansion, the discovery of the complex roles of stromal and infiltrating immune cells in tumor proliferation, neovascularization, and metastasis offers new avenues for therapeutic intervention. Angiogenesis is critical for both tumor expansion and metastasis, and the tumor vasculature is a promising new therapeutic target. The hypoxic tumor environment additionally provides therapeutic opportunities. Comprehension of many of the molecular driving forces behind cancer has given us better insights into the high degree of cross-talk between numerous signaling pathways between the tumor and its surrounding microenvironment. Elucidation of the interrelationships of these signaling networks facilitates the selection of rational drug combinations with complementary actions.

Genetic instability is a hallmark of cancer, and tumors are highly heterogeneous, providing one of the many challenges in cancer treatment. Individual cells within a tumor have different growth and survival characteristics as well as different metastatic potentials. Effective targeting of the particular processes abnormal to each tumor is critical for optimal therapy. Together with development of targeted therapies, accurate predictive diagnostic tools are necessary to identify the molecular defects of tumors to allow for individualized targeted therapy.

## Evolution of the Drug Discovery Process

Early therapies for cancer were identified in part by serendipity combined with astute observations by scientists and clinicians. Exposure of servicemen to nitrogen mustard as part of a secret government chemical warfare program during World War II led to the realization that these agents caused leucopoenia and bone marrow suppression. This knowledge was used in the first clinical tests of alkylating agents in the treatment of cancer. The observation that folate supplementation of children with leukemia caused leukemic cell proliferation led to the development of the antifolates and other antimetabolites. Extracts of the periwinkle plant were evaluated in rodents for antidiabetic properties where they caused aplasia of bone marrow and glanulocytopenia. The active constituents were isolated as the vinca alkaloids, a class of compounds with broad anticancer effects. The identification of drugs that worked and testing of these agents against a variety of different tumor types to find those that were sensitive characterized early cancer drug discovery.

Cancer drug discovery now is tied to advances in understanding the molecular causes of aberrant growth, resistance to immune surveillance and abnormal apoptotic processes, all of which occur during carcinogenesis. Today the majority of cancer drug discovery is related to identifying inhibitors of specific events that are aberrant in cancer. In this new era for cancer drug discovery, 40 percent of the compounds in clinical development are directed against novel targets, and it is estimated that 70 percent of targets for cancer drug discovery are novel, with no history of utility [1]. The prognosis for the treatment of cancer has never been better, and targeted therapy holds the key for more specific and effective therapy. Although cures may not be achieved, control of cancer as a chronic disease may be attainable.

## Focus and Definitions

The goal of this chapter is to provide an introduction to the field of cancer drug discovery. It focuses on a subset of current targets, and a brief overview of some approaches for identification of leads, and sources of materials for drug discovery efforts. The list of targets is not all inclusive but covers many validated cellular targets and several promising targets that may yield effective anticancer pharmaceuticals. The lack of inclusion of a particular target in

this introductory chapter should not in any way reduce enthusiasm for these drug targets for anticancer drug design. The chapter is focused on initial drug discovery of small-molecule inhibitors and not on the equally complex processes of lead optimization, preclinical pharmacology, in vivo testing, and clinical development. Additionally, some experimental approaches for the identification of active molecules are described. High-throughput screening (HTS) is of paramount importance in identifying new drug leads for cancer, and the details of assay design and approachs for HTS are covered in detail in other chapters of this text, so methods related to HTS are not covered in this chapter. Humanized monoclonal antibodies are discussed because they provide important proof-of-principle of the effectiveness of targeted therapy. Humanized monoclonal antibodies provide clinical utility for treating cancer, but the design of these therapies is not addressed. Gene therapy, antisense oligonucleotides, and small, interfering ribonucleic acid (siRNA) are additional viable avenues for anticancer therapies, but they are not a focus of this introductory chapter.

It is important that the cancer drug discovery community use a common language that designates the various stages of cancer drug development. These definitions were proposed years ago by the Developmental Therapeutics Branch of the National Cancer Institute [2] and are presented here to allow consistency and to acquaint those new in the field to appropriate terminology. *Cytotoxic* is used to describe activity against cells in vitro. This term does not imply any selectivity for cancerous versus normal cells. *Antitumor* is activity against a tumor in an in vivo model, and *anticancer* is used only to designate efficacy of a drug against cancer in humans following clinical testing.

## 2 SIGNAL TRANSDUCTION TARGETS

The goal of targeted therapy is to discover and develop target-specific therapies that offer the promise of more specificity toward the key biological processes that are abnormal in cancer. Abberant cellular signaling is a hallmark of cancer, with dysfunctions in receptor tyrosine kinases, intracellular tyrosine kinases, and protein phosphatases that result in increased cellular proliferation, survival, invasion, and angiogenesis. Identification of the role of specific signal transduction pathways in the carcinogenic process has led to efforts to specifically target these pathways for anticancer therapy. The most successful to date has been the targeting of Bcr/Abl that drives the oncogenic lesions in chronic myelogenous leukemia (CML). Dramatic responses are observed in the treatment of this disease with imatinib (Gleevec). Relapses occurred during therapy and whether this treatment will provide cures has yet to be seen. What has been brought to light with targeted therapy of signaling pathways is that in many cases multiple signaling pathways drive the oncogenic lesions and rational and individualized combinations of inhibitors may

be needed for optimal therapy. Targeting signaling pathways for therapeutic benefit may be more complex than originally anticipated.

## HER2

The first successful rationally developed targeted therapy for cancer was the inhibition of HER2, also known as ErbB2, a member of the epidermal growth factor (EGF) family of receptor tyrosine kinases. The oncogenic effects of *HER2/neu* were demonstrated when blockade of the receptor inhibited tumorigenesis [3]. These experiments provided valuable information about HER2 as a target, as well as the therapeutic possibilities of blockade by monoclonal antibodies. Amplification of the *HER2/neu* gene in breast cancers causes overexpression of the HER2 protein that drives breast cancer cell proliferation. Overexpression of HER2 leads to resistance to therapy and a poorer disease prognosis. It is estimated that 30 percent of all invasive breast cancers have amplified *HER2/neu*, and therapeutic targeting could have significant therapeutic benefits to those with the genetic signature of amplified *HER2/neu*. These findings led to the development of a humanized monoclonal antibody to HER2, trastuzumab (Herceptin). This antibody binds to the HER2 receptor and inhibits downstream signaling leading to growth inhibition as well as sensitization to cytotoxic chemotherapeutic drugs. The benefits of targeting HER2 are evident as a single agent and in combination with chemotherapy. With the clinical successes of the humanized monoclonal antibody trastuzumab, work is ongoing to identify small-molecule inhibitors of HER2. Importantly, the development of trastuzumab provided the first proof-of-principle that targeting the specific molecular defects of cancer is a viable approach for anticancer drug development.

## The EGF Receptor Pathway

The therapeutic value of targeting one member of the EGF pathway, HER2 has been demonstrated and EGF receptor (EGFR) (HER1), a second member of this receptor tyrosine kinase family, is being evaluated as a drug target for multiple types of cancer. Activation of the EGFR through ligand binding initiates a cascade of effects, which are depicted graphically in Figure 1. This signaling pathway activates the mitogen activated protein kinase (MAPK), phosphoinositide-3 kinase (PI3K)/Akt and the Jak/Stat pathways. These signaling cascades are implicated in cell proliferation, growth, survival, apoptosis, migration, and angiogenesis [4–6]. Dysfunction of the EGFR occurs in many different types of cancer including lung, breast, head and neck, colon, pancreatic, ovarian, and prostate cancers. Over 33 percent of solid tumors express EGFR, and it is often associated with poorer prognosis [6]. Activation of EGFR can occur due to overexpression of the receptor, expression of a constitutively active receptor, and/or amplified levels of its ligands, EGF or transforming growth factor α (TGF-α). Genetic alterations of EGFR, including

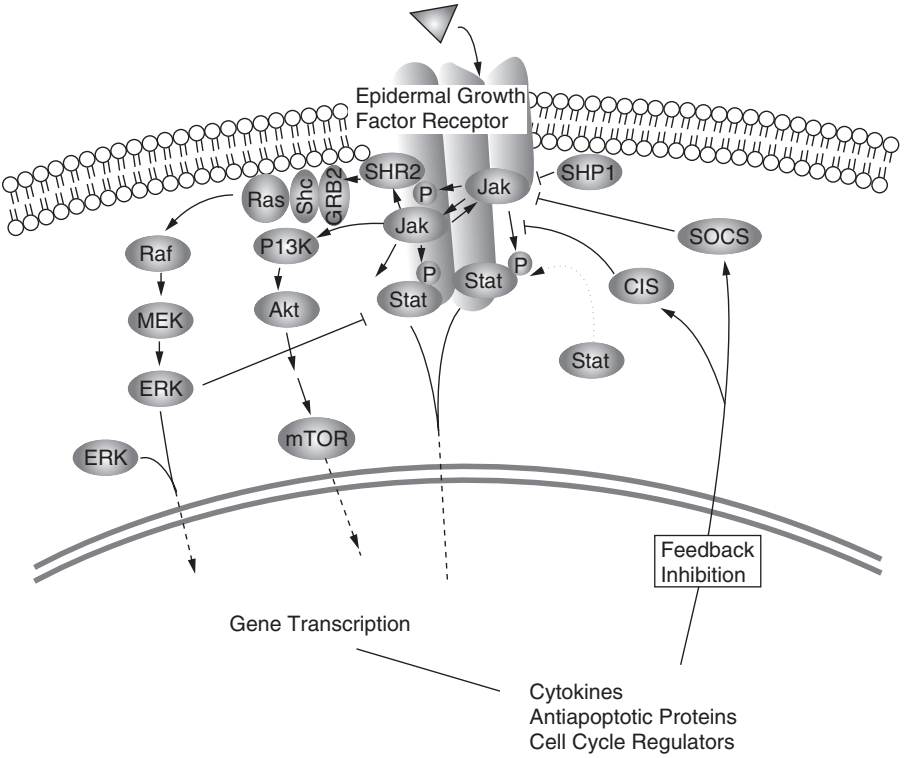


**Figure 1** The EGFR signaling pathways. After ligand activation, the EGFR phosphorylates and activates the Ras-Raf-MAP kinase PI3K/Akt, and Jak/Stat pathways. This in turn results in activation of transcription factors and modulation of the cell cycle, growth, apoptosis, and angiogenic processes. From El-Rayes and LoRusso [4].

gene amplification and mutations leading to a constitutively active kinase, contribute to the pathogenesis of many tumors [6].

The prevalence of activated EGFR signaling in cancer led to several different targeting strategies including blockade of the receptor with monoclonal antibodies, analogous to the approach used with HER2, inhibition of the tyrosine kinase domain of the receptor, and use of antisense oligonucleotides against the receptor. Cetuximab (Eributex) is a humanized anti-EGFR antibody that binds to the EGFR causing dimerization and downregulation of MAPK, PI3K, and Jak/Stat signaling, which inhibits proliferation, cell cycle progression, metastasis, and angiogenesis [4, 6]. Cetuximab is well tolerated with limiting toxicities of allergic reactions and acneform rash, and it is being used for the treatment of advanced metastatic colorectal cancer [4, 6].

A second approach to limit EGFR's effects is through small-molecule tyrosine kinase inhibitors. Inhibition of the adenosine triphosphate (ATP) binding

site has been the primary site of focus. Four different categories of inhibitors have been developed: EGFR-specific reversible inhibitors, EGFR-specific irreversible inhibitors, and reversible and irreversible pan HER inhibitors that have actions against several HER family members [4, 6]. Gefitinib (Iressa) is a reversible inhibitor of the EGFR ATP binding site. It is well tolerated, but only modest clinical responses were seen. Although the majority of non-small-cell lung cancer patients did not respond to inhibition of EGFR by gefitinib, a subgroup of 10 to 15 percent exhibited dramatic responses. Profiling of these tumors indicated that they had somatic activating mutations in the catalytic domain of the EGFR [7, 8]. The selectivity of gefitinib for these activating mutations had not been anticipated from preclinical studies. The mutations did not affect proliferation via the MAPK pathway, but they activated the Akt and Stat pathways and caused resistance to apoptosis, heightened survival, and drug resistance [9].

Antisense technologies are also being evaluated for the ability to decrease EGFR expression. What is clear though is that optimal use of these EGFR-targeted therapies may not be as apparent as first anticipated. The fact that the clinical indications for cetuximab and gefitinib do not overlap suggests that there is much more to learn about optimal use of these targeted therapies. The use of dual or pan-targeted therapies against multiple EGF family members, for example, EGFR and HER2, may also provide superior antitumor effects. Promiscuous inhibitors with broader specificity for multiple receptor kinases may be more active in tumors driven by abnormalities in several signaling pathways. Targeting several points in the signaling pathway may also yield better results.

## Insulin-Like Growth Factor Signaling

Another family of receptor tyrosine kinases, the insulin-like growth factor (IGF) family, is implicated in the proliferation and survival of cancer cells. Activation of this signaling cascade in cancer may be due to amplified levels of the insulin-like growth factors IGF-1 and IGF-2 or its receptor IGF-1R. Similar to the signaling cascade initiated by EGFR, activation of IGF-1R stimulates the PI3K and MAP kinase pathways. Individuals with higher circulating levels of IGF-1 are at greater risk for lung, breast, prostate, colon, and bladder cancer [10]. Reduction of IGF-1 levels in murine models of cancer reduced tumor burden and incidence of metastasis [11]. In cell culture systems and in animal models, inhibition of IGF-1R expression or activity inhibited cell proliferation, colony formation, and tumor growth [12]. Efforts to identify specific inhibitors of this receptor tyrosine kinase have been successful. Inhibition of IGF-1R prevents downstream activation of both the PI3K and MAP kinase pathways and, ultimately, reduction in tumor growth [12]. Whether inhibition of IGF pathways will result in therapeutic benefit remains to be seen, but targeting this signaling cascade alone, or in combination with other signaling inhibitors or cytotoxic drugs, may lead to anticancer effects.

## Bcr/Abl, c-Kit, and Platelet-Derived Growth Factor

The molecular driving force in CML is the Philadelphia chromosome ($Ph^+$), the name given to a reciprocal translocation between chromosomes 9 and 22. In this translocation the *Bcr* gene replaces the first exon of the *Abl* gene forming a chimeric oncoprotein Bcr/Abl. Bcr/Abl is deregulated and has constitutively active tyrosine kinase activities that leads to the phosphorylation of numerous cellular proteins including PI3K. Because this tumor is caused by an activating translocation in Bcr/Abl, this represented an optimal test of targeted therapies. The Bcr/Abl inhibitor imatinib (Gleevec), which binds and inhibits the ATP-binding site of the kinase, provided striking antitumor effects against CML and represents, to date, the most successful use of targeted therapy. In phase II studies using $Ph^+$ patients, a response rate of 95 percent was measured [5]. Imatinib is not specific for Bcr/Abl. It also has potent activity against c-Kit and the platelet-derived growth factor (PDGF) family of receptors. Activating mutations in the receptor tyrosine kinase c-Kit are found in gastrointestinal stromal tumors (GIST), and in some acute myeloid leukemias, small-cell lung cancer, neuroblastoma, and breast cancers [5]. Imatinib was highly effective against GIST, and it may have antitumor actions in other tumor types with activated c-Kit. The PDGF family is implicated in the pathogenesis of glioblastoma with over expression of both the receptor and the ligand, PDGF. Evidence suggests that PDGF autocrine signaling loops are involved in melanoma, neuroendocrine tumors, and tumors of the prostate, stomach, pancreas, and lung [5]. Targeting Bcr/Abl, c-Kit, and PDGF with imatinib provides utility against multiple types of cancer, and data suggest that the efficacy of imatinib may be due to its ability to target multiple tyrosine kinases [13].

## Mammalian Target of Rapamycin (mTOR)

While significant efforts are directed toward inhibiting the receptor tyrosine kinases that drive proliferation and survival, several downstream effectors of these receptors are also emerging as cancer drug targets. Focusing on central junction points for multiple signaling pathways may provide broader antitumor activity. The serine/threonine kinase mTOR (mammalian target of rapamycin) fits this criterion as it is a convergence point for several signaling cascades (Fig. 1). mTOR is phosphorylated by PI3K/Akt in response to growth factors, the estrogen receptor, and oncogenic Ras [14–16]. Through a PI3K/Akt-independent pathway, mTOR senses nutritional status and controls cell cycle progression through $G_1$ [15]. Evidence has emerged that mTOR is a central controller of cell growth and proliferation mediated through its multiple effects on protein synthesis [14–16]. mTOR, as its name implies, was identified because it is the target of the bacterial secondary metabolite rapamycin. Inhibition of mTOR causes little change in global protein synthesis, but significant changes in the cellular levels of specific proteins including c-myc and cyclin D1, which play key roles in cell cycle progression, were observed [14].

Further cell cycle changes are mediated by the ability of rapamycin to stabilize the cyclin-depenent kinase inhibitor p27 (Kip 1) leading to inhibition of cyclin-depenent kinase 2 (Cdk2). Rapamycin has effects on the cell survival pathways where it causes activation of the stress kinase ASK1. In p53 mutant cells, sustained ASK1 phosphorylation causes cell death [14].

Defects in multiple upstream proteins can activite mTOR in cancer (Fig. 1). These include activated receptor tyrosine kinases and activating *Ras* mutations that occur in approximately 30 percent of epithelial-derived tumors [15]. Amplifications of the genes for PI3K and Akt and loss-of-function mutations in the tumor suppressor phosphatase and tensin homologue (PTEN), which inhibits PI3K, occur in many tumors [14, 15]. Mutations in eukaryotic translation initiation factor 4E (eIF4E), a downstream effector of mTOR, cause malignant transformations [14, 15]. The central position of mTOR in the signaling abnormalities that define cancer makes it a logical anticancer drug target. Rapamycin is an effective inhibitor of mTOR, but its inherent instability and poor solubility prevented its development for cancer therapy. Rapamycin analogs were designed and synthesized to overcome these limitations and were evaluated clinically. The results suggest low toxicity and some clinical responses as well as disease stabilization [16]. Optimization of scheduling and patient selection may provide additional opportunities to evaluate the effectiveness of inhibiting mTOR as a therapeutic approach. Identification of a predictive biomarker would allow correlation of drug levels with clinical response [14].

In summary, targeting signaling pathways implicated in the molecular etiology of cancer holds promise for more specific and less toxic therapy for neoplastic diseases. Many hurdles exist to fully implement the hope of this approach. There is significant cross-talk and redundancy of signaling pathways that control cell growth and survival. This, together with the inherent genetic instability of cancer, leads to questions of how best to use these molecular-targeted therapies. Rational combinations of these signaling blockers represent one approach, as is the testing of agents with broader target specificity inhibiting multiple receptor tyrosine kinases. A small molecule inhibitor that targets both EGFR and HER2 provides greater activity against HER2 positive breast cancer cells than HER2 inhibition alone [17]. Additionally, there is a need to understand the characteristics of tumors that are responsive to these targeted therapies and how biomarkers can be identified for appropriate patient selection. Successful use of these targeting agents requires integrated diagnosis and therapy.

## Methods of Drug Discovery

The screening assays used to identify inhibitors of signaling cascade were initially based on biochemical kinase assays. With the availability of purified proteins and labeled peptide substrates, many different assay technologies are available to screen for inhibitors of protein kinases. New screening technologies are being developed at a rapid pace and most use a fluorescent-based

readout based on the ability of the kinase to phosphorylate an exogenous fluorescent peptide substrate. HTS utilizes techniques such as enzyme-linked immunosorbent assay (ELISA), phospho-specific protein-specific antibodies, scintillation proximity assays, biotin-labeled substrate capture assays, fluorescence polymerization (FP), and time-resolved fluorescence resonance energy transfer (TR-FRET) [18–20]. IMAP® (immobilized metal ion affinity-based fluorescence polarization) is a new technology utilizing the ability of the fluorescently labeled phosphorylated substrate to bind to immobilized metal complexes on nanoparticles [20].

Every new assay should be carefully evaluated to ensure that it is robust, provides consistent and reliable results, and that it is predictive, selective, and sensitive. The assays need to have sufficient throughput to sample a wide variety of chemistry to identify molecules with activity against the target. It is important that new screening technologies retain the sensitivities and reliability of older generations of screening technologies. HTS (96- and 384-well format) have been used, and with improvements in liquid handling capabilities, ultrahigh-throughput (uHTS) (1536+ well formats) using ultra low volumes (1 μL) are now available. Sensitivity for uHTS is provided by TR-FRET and FP detection systems [21]. Microarrayed compound screening is an ultrahigh throughout screening method (uHTS) method that eliminates the need for complex liquid handling systems. The reagents that would be added to a well of a microplate are incorporated into agarose gels that are layered over the compounds arranged on polystyrene sheets [22]. This provides capabilities to screen at a density of 8640 reactions on one sheet.

The ATP-binding site of the kinases is a primary target for identification of inhibitors and most of the protein kinase inhibitors that have been evaluated clinically are directed to this site. With advances in the modeling of the catalytic site of these kinases, virtual screening using computer modeling of the binding site and compound structures is a useful approach. Certain kinases adopt different conformations when activated by cofactors. This should be considered in the design of the binding site for optimal anticancer efficacy and specificity.

## 3 TARGETING TUMOR VASCULATURE

Tumor growth is dependent upon the generation of a functional blood supply to provide nutrients and remove metabolic waste. Without neovascularization a tumor is limited in size to 1 to 2 mm in diameter [23]. Tumor vasculature can form by a number of different processes including sprouting from existing vessels and recruitment of circulating endothelial cell precursors. Unlike normal vasculature, tumor vasculature is highly abnormal and poorly differentiated, with tortuous dilated vessels [24]. Targeting either the process of new blood vessel development, angiogenesis, or the existing tumor vasculature, with antivascular therapy, are valid approaches to control and limit cancer and

metastasis. In addition to targeting blood vessel formation, the development of a functional lymphatic system is critical to tumor metastasis. Inhibition of this process, called lymphangiogenesis, together with antiangiogenic therapy may also provide opportunities for better antitumor activities [25].

## Inhibitors of Angiogenesis

The promise of controlling tumor growth by preventing angiogenesis was proposed over 30 years ago by Judah Folkman [26]. During tumor development, there exists an imbalance between proangiogenic factors and antiangiogenic factors that leads to rapid blood vessel growth (Fig. 2). This promotion of tumor angiogenesis is called the angiogenic switch and is critical for cancer progression and metastasis [27]. Stimulation of growth factor secretion by the tumor in response to oncogenic stimulation or hypoxia can unbalance the equilibrium and cause rapid neovascularization. This angiogenic switch occurs as an early event of tumorigenesis when quiescent microscopic tumors change into rapidly expanding aggressive tumors. The search for agents that could specifically inhibit tumor angiogenesis provides an attractive target for anticancer therapy. In the adult, with a few exceptions, the normal vasculature is quiescent with balance of positive and negative regulatory factors. Antiangiogenic therapy is expected to allow more specific targeting of abnormal events. Targeting genetically stable endothelial cells might have additional advantages in that it may circumvent drug resistance mechanisms that occur in tumor cells due to genetic instability. However, recent data suggest that the genetic abnormalities of tumor tissue may also be found in tumor endothelial cells [28]. The apparent heterogeneous nature of tumor endothelial cells might need to be considered in the future. Although antiangiogenic therapy was conceived decades ago, the first antiangiogenic anticancer agent, bevacizumab (Avastin)

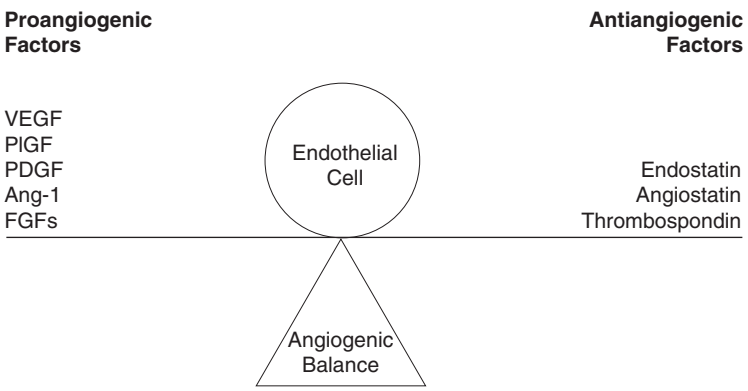


**Figure 2** Angiogenic balance. In normal endothelial cells the proangiogenic factors are balanced by antiangiogenic factors. Disruption of the balance can lead to pathogenic conditions.

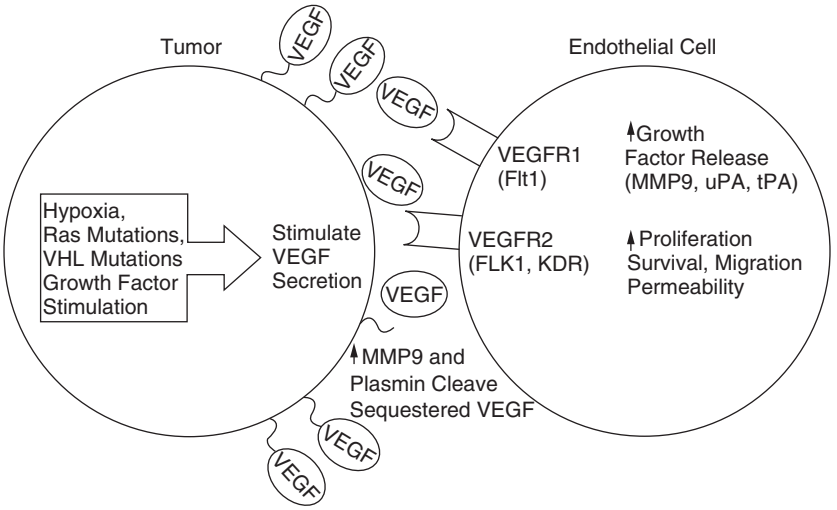


**Figure 3** The VEGF pathway. Antiangiogenic approaches that target the tumor-driven VEGF pathway include blocking VEGF with antibodies, trapping VEGF with decoy receptors, blocking VEGFR with antibodies, and inhibiting VEGFR receptor tyrosine kinase activity.

was approved by the Food and Drug Administration (FDA) in early 2004 for use in colorectal cancer.

Angiogenesis is a complex process providing many avenues for control (Fig. 3). Key signaling events leading to the angiogenic switch include growth factors and receptors of the vascular endothelial growth factor (VEGF) family, basic fibroblast growth factor (bFGF), acidic fibroblast growth factor (aFGF), the platelet-derived growth factor (PDGF) family, and protease inhibitors. Inhibition of these may provide avenues to inhibit the angiogenic switch. Endogenous antiangiogenic signals are also present. The endogenous inhibitors endostatin and angiostatin have been tested for antiangiogenic actions in tumor models where they showed antitumor and antiangiogenic effects, yet anticancer effects have not yet been achieved.

***The VEGF Axis*** The VEGF axis is critically important to normal embryonic and adult angiogenesis as well as to pathological blood vessel development. Loss of a single allele for VEGF, or the genes for VEGFR1 (Flt-1) or VEGFR2 (Flk-1, *KDR*) results in embryonic lethality. Although both normal and pathological angiogenesis involve the coordinated effort of many growth factors, the VEGF signaling cascade appears to be the rate-limiting step of new blood vessel formation [29]. The VEGF receptor ligands are a family of proteins, VEGF-A-E, and placental growth factor (PlGF). VEGF-A is the predominant VEGF with activity in angiogenesis, and its activities predominate in both normal and tumor angiogenesis [29]. The human VEGF gene contains eight exons and alternative splicing leads to the generation of several different

VEGF-A isoforms [30]. The isoforms differ in their heparin binding affinities, the smaller, acidic $VEGF_{121}$ having no affinity for heparin and the larger basic isoforms, $VEGF_{189}$ and $VEGF_{206}$, bind readily to extracellular matrix-associated heparin. The intermediate isoform, $VEGF_{165}$, has optimal biological activity and an intermediate capacity for heparin binding [30]. Tumor cells can directly secrete active VEGF or active isoforms of VEGF can be released from heparin in the extracellular matrix by the actions of proteases including plasmin and matrix metalloprotease 9 (MMP9) (Fig. 3). VEGF-C and D are implicated in lymphangiogenesis but are not thought to play a critical role in tumor angiogenesis. In addition to several different ligands, at least three different receptor types for the VEGF family have been identified (Fig. 3). The receptors of VEGF, VEGFR1–3, are receptor tyrosine kinases with varied cellular effects and localization. Currently, evidence suggests that the VEGFR2 receptor (Flt-1, KDR) plays the primary role in endothelial cell permeability, proliferation, migration, and release of endothelial stem cells from bone marrow, events critical for tumor angiogenesis [31, 32]. VEGFR1 is also present on endothelial cells, and activation can lead to induction of growth factor release and release of proteolytic enzymes including MMP9, urokinase-type plasminogen activator [33], and tissue type-plasminogen activator (tPA) [30, 31, 34]. An alternately spliced soluble isoform of VEGFR1 is implicated in blocking VEGF actions. Receptor cross-talk between VEGFR1 when activated by PlGF and VEGFR2 has been implicated in amplification of VEGF angiogenic signaling [31, 32]. VEGFR3 is involved primarily in lymphangiogenesis.

The FDA approval of the first antiangiogenic anticancer therapy, bevacizumab, a humanized monoclonal antibody to VEGF-A, highlights the importance of the VEGF pathway in tumor angiogenesis. Several different approaches to inhibit the tumor-driven VEGF pathway are active targets for cancer drug discovery. They include blockade of VEGF, with, for example, a monoclonal antibody, trapping VEGF with decoy receptors, antisense inhibition of VEGF, antibodies to VEGFR, and inhibition of VEGFR tyrosine kinase activity by small molecular protein kinase inhibitors [35].

***HIF-1α as a Therapeutic Target*** The transcription factor hypoxia-inducible factor-1 (HIF-1) is a new target for antiangiogenic therapy and antitumor therapy because it occupies a central point in the responses of tissues to hypoxia. Tumors, by virtue of aberrant blood supply and high interstitial pressure, are hypoxic, and this causes activation of adaptive survival mechanisms mediated by the transcription factor HIF-1. HIF-1 binds to hypoxia response elements (HRE) leading to the transcription of numerous proteins involved in angiogenesis, cell survival and invasion, and drug resistance. Tumors maintain high levels of HIF-1 and in many tumor types, high expression of HIF-1 is associated with a worse prognosis, validating this protein as a therapeutic target for cancer. HIF-1 is a heterodimer composed of an α and a β subunit. Both subunits are constitutively transcribed, but the cellular levels of the α

subunit are tightly controlled, providing the rate-limiting step in HIF-1 expression. Under normal oxygen conditions HIF-1α is rapidly degraded by the ubiquitin–proteasome pathway; however, under hypoxic conditions the protein is stabilized and binds to the β subunit to form the active transcription factor HIF-1. In addition to hypoxia, the expression of HIF-1α can be induced by oncogenes including *Ras*, *Src*, *HER2/neu*, and *mTOR* and by loss of tumor suppressor genes including *p53*, *PTEN*, and *VHL*. Growth factor stimulation of receptor tyrosine kinases and reactive oxygen species also stabilize HIF-1α. Over 60 genes are transactivated by HIF-1, including the genes for VEGF and VEGFR. Targeting HIF-1 may provide broad antiangiogenic and antitumor effects because of its ability to inhibit several different pathways of angiogeneic stimulation.

## Antivascular Agents

With the focus on tumor blood vessels as a target for cancer therapy, some compounds with antivascular effects have been identified. Antivascular agents differ from antiangiogenic agents in that they target existing tumor vasculature, while antiangiogenics inhibit the process of neovascularization. Vascular disrupting agents cause rapid, often within minutes, and substantial disruption of the aberrant tumor vasculature, and subsequent loss of circulation leading to tumor ischemia and necrosis [36]. Vascular disrupting agents in clinical testing include multiple-tubulin targeting drugs such as combretastatin A4 and flavonoids that appear to act by inducing the localized release of TNFα from activated macrophages in the tumor [36]. The antivascular approach, while distinct from antiangiogenic therapies, appears promising. Combinations of antiangiogenic and antivascular targeting may be used to reduce initial tumor burden and to inhibit further tumor growth.

## Methods of Drug Discovery

Diverse mechanisms are used in the identification of antiangiogenic agents. The VEGF axis represents many different avenues for drug discovery, and it has been validated as a therapeutic target for cancer. The efficacy of bevacizumab, which binds to VEGF ligand, has proven to be an effective approach to inhibiting VEGF signaling. HTS is used to identify inhibitors of the VEGF receptors tyrosine kinase activity. A high-throughput TR-FRET assay targeting VEGFR-2 measures the autophosphorylation of the receptor that occurs upon its activation [19]. The assay is homogenous in that no washing steps are required and is amenable to automation. Mechanism-based HTS to identify inhibitors of HIF-1 could address the ability of inhibitors to repress its transcriptional activation of HREs. To date most inhibitors of HIF-1 have been identified by mechanistic studies of known inhibitors [37].

The use of endothelial cells as a model has advantages in that a cell-based screen can identify new targets. Inhibition of endothelial cell proliferation,

invasion through basement membrane material [38], and the formation of capillary-like tubules on Matrigel® matrices are assays used to predict antiangiogenic effects. Reagents are available commercially to evaluate endothelial cell migration and invasion in a 96-well plate format. Endothelial cells from different vascular beds can be obtained commercially, but it is important to note that these normal cells have a finite life span. An immortalized tumor endothelial cell line has been proposed as a good model of tumor endothelial cells [39]. Assays are used to confirm that inhibitors identified in HTS or cell-based screens actually have activity against neovascularization, an in vivo process involving not only the endothelial cells but their stromal environment as well. Several different models are available to assess the ability of compounds to inhibit angiogenesis. They include embryonic zebrafish [40], chorioallantoic membrane of the chick embryo [41], and growth of vessels into a subcutaneous matrigel plug [42]. Pathological evaluations of tumor samples are used to assess antiangiogenic effects concurrent with antitumor measurements. New imaging techniques with magnetic resonance imaging (MRI) and ultrasound allow real-time evaluation of both angiogenic and antivascular effects [43, 44].

## 4 DNA TARGETS

Nitrogen mustard, a deoxyribonucleic acid (DNA) alkylating agent, led the era of modern cancer chemotherapy. DNA binding agents remain some of the most effective anticancer agents available. Cross-linking of DNA and adduct formation by alkylating agents prevents DNA replication and normal repair functions culminating in apoptosis. Significant efforts over the course of many years led to optimization of alkylating agents including platinum-based compounds. Antimetabolites prevent the synthesis of essential DNA and RNA precursors, which then causes inhibition of DNA and protein synthesis leading ultimately to apoptosis. Topoisomerases are enzymes critical for maintaining normal DNA topography. Topoisomerase inhibitors bind to both DNA and the topoisomerase enzyme leading to DNA strand breaks, which accumulate and cause cell death. Drugs that target DNA by diverse mechanisms play a major role in cancer therapy. New DNA-based targets have emerged in the past few years that have the hope of wider therapeutic windows and better specificity for cancer cells as compared to normal cells.

### DNA Methylation

Mutations in genes critical to cellular growth and survival can cause and promote cancer. Efforts are underway to target some of the specific genetic changes that occur during cancer. In addition to these genetic changes, epigenetic changes in DNA are also implicated in carcinogenesis and maintenance of the tumorigenic phenotype. Gene silencing that occurs through changes in

chromatin structure may be reversible, providing a target for therapeutic intervention. Gene expression is controlled by many factors, including accessibility of the gene for transcription. DNA methylation patterns play a key role in gene expression. Actively transcribed genes have low methylation status, while areas that are silenced are hypermethylated. DNA methytransferases (DNMT) are enzymes that catalyze the addition of a methyl group to cytosine residues. Increased methylation in the cytosine-rich CpG promoter regions causes gene silencing. Hypermethylation of CpG islands is implicated in the silencing of a number of tumor suppression genes [45, 46]. DNA methylation appears to play a central role in cancer and yet global hypomethylation results in genetic instability [45, 46]. The ability to revert cancer cells back to a normal methylation pattern may hold promise for fixing the underlying events of cancer, yet care will need to be taken to prevent global changes in DNA methylation. Three DNMTs have been identified and specificity may hold the key to solving this conundrum. Inhibition of DNMT1 by a cytosine analog or by siRNA reverted transcriptional silencing of tumor suppressor genes [46]. Antisense oligonucleotides directed against DNMT1 are effective antitumor agents and a small-molecule inhibitor of DNMT1 is being evaluated clinically. It will be critically important to affect the abnormal DNA methylation without changing global methylation [47].

## Histone Deacetylase Inhibitors

In addition to DNA methylation, histone acetylation is intricately involved in epigenetic control of gene transcription [46, 48]. Histones are the structural units that organize and condense chromatin. A nucleosome contains about 150 base pairs of DNA that is wrapped around a core of histone proteins. The amino terminus tails of histones contain lysine residues that can be acetylated by histone acetytransferases (HAT). Acetylation of these lysines leads to relaxation of the tight chromatin packaging in that region facilitating access to transcription factors and leading to transcriptional activation [48]. Histone deacetylases (HDAC) remove the acetyl group initiating a more compact chromatin structure and resultant transcriptional repression.

Histone acetylation is a dynamic process controlled by the balance between the actions of HATs and HDACs [48]. HDACs also have direct interactions with transcription factors including p53 and e2F [48]. In cancer, mutations in HAT and recruitment of HDAC, via in part by DNA methylation, are associated with transcriptional silencing of tumor suppressor genes [48, 49]. The ability to reverse the abnormal transcriptional repression is an attractive therapeutic target, and numerous HDAC inhibitors are being evaluated in preclinical and clinical models [48, 49]. Several diverse chemical classes of compounds inhibit HDAC. Trichostatin A, a microbial natural product was found to induce differentiation of leukemia cells, and additional studies showed that these effects were mediated through HDAC inhibition [48]. Other

cellular effects of HDAC inhibitors include growth suppression, cell cycle arrest, induction of the cyclin-dependent kinase inhibitor p21, and initiation of apoptosis [49]. The effects of HDAC inhibitors appear to be cell-type dependent, and concerns have been raised about the ability of some of these inhibitors to increase genetic instability and telomerase (hTERT) expression [49]. Several HDAC inhibitors have been evaluated clinically. They appear to be well tolerated, and early results show some clinical responses that are tied with inhibition the target [48, 49]. Combinations of HDAC and DNMT inhibitors are being evaluated because targeting epigenetic silencing by two different approaches may provide optimal reversal of aberrant gene repression.

## Telomerase Inhibitors

Telomeres are tandem nucleotide repeats found on the ends of chromosomes. A T-loop structure is found at the very end of the telomeres, and it differentiates the ends of the chromosomes from a chromosome break. These specialized structures prevent homologous recombination and nonhomologous end joining between chromosomes [50]. DNA polymerase cannot replicate to the very end of the chromosome because of the overhang needed for lagging strand synthesis. With progressive cell divisions telomeres are lost and aging cells contain shorter telomeres. Cellular replicative senescence occurs when the telomeres shorten to a certain threshold where they signal cell cycle arrest [51]. If this signal is lost and division continues, then chromosome ends fuse, leading to initiation of apoptosis. A normal functioning telomere prevents uncontrolled growth and cell immortality, and this is thought to help prevent cancer. Unfortunately, cancer cells escape replicative senescence and lose this control mechanism. The ability of cancer cells to escape this control is one of the hallmarks of cancer [50, 51]. Cancer cells reactivate telomerase, an enzyme that adds telomeric repeats to the ends of the chromosomes. Telomerase is composed of the enzymatic reverse transcriptase subunit, hTERT, and an RNA nucleotide template, hTR [51].

Activated telomerase is detected in almost 100 percent of cancer cell lines and in 80 percent of human tumors, and telomere-deficient mice have a low incidence of tumors [52]. Telomerase is a logical target for cancer therapeutics because cancer cells depend on active telomerase to escape replicative senescence. Telomerase inhibitors may be fairly specific because most normal cells do not express telomerase. Telomerase inhibitors are being developed. BIBR1523 is a specific and potent telomerase inhibitor that blocks cell proliferation and causes shortening of telomeres but does not initiate cancer cell death [52]. Consistent with its cytostatic actions, it did not cause tumor regression in tumor xenograft models. Concerns about the side effects of telomerase therapy have been expressed. Telomerase is expressed in hematopoetic cells and in germ cells and cells lining the intestinal crypts. Inhibition of telomerase

may cause some of the side effects associated with other DNA-interacting anticancer drugs due to lack of specificity [50–52]. Whether telomerase will prove to be a valid target for cancer therapy remains to be seen.

## 5 ANTIMITOTICS

Antimitotics are one of the oldest and most effective classes of anticancer drugs. They are used in combination with other cytotoxic chemotherapeutics in the curative therapy of childhood and adult leukemias and testicular cancer. They are also effective in the adjuvant setting, providing activity against a broad range of solid tumors and hematological malignancies. The majority of antimitotics target tubulin, but newer antimitotics that target different components of the mitotic apparatus appear to be excellent therapeutic targets for cancer.

### Microtubule-Targeting Agents

Tubulin is a validated anticancer target. Tubulin-targeting drugs are the one class of anticancer drugs that are equally effective against p53 mutant and p53 null tumors, and considering the fact that over half of human tumors contain abnormal p53, this is significant [53]. The history of the use of chemically diverse tubulin-targeting drug shows that structure cannot predict clinical utility or dose-limiting toxicities, and new tubulin-targeting agents may be expected to have different spectra of clinical activities. Microtubule-disrupting agents can be divided into two groups, microtubule stabilizers and microtubule depolymerizers. Microtubule stabilizers shift the equilibrium of cellular tubulin toward the polymerized state, and dramatic increases in the number of cellular microtubules can be observed in cells treated with these compounds. In contrast, microtubule depolymerizers cause a loss of cellular microtubules and a shift of cellular tubulin toward the cytosolic pool. At the lowest effective concentrations both classes of drugs inhibit microtubule dynamics and lead to interruption of mitosis and, ultimately, apoptosis [54]. Cancer cells are much more susceptible to tubulin-targeting drugs, and the molecular cause of this is not yet known. Identification of the nature of the susceptibility of cancer cells for microtubule-targeting agents may help in the identification of new mitotic targets for cancer therapy.

Taxol was the first microtubule stabilizer identified, and it represents one of the most effective drugs for cancer therapy to be identified in the last 20 years. Other microtubule stabilizers including taxane derivatives, the epothilones, and discodermolide are in clinical trials and the preliminary results look promising. Notably, they appear to be active in taxane-refractory tumors, even though they bind to the same binding site on tubulin. All the stabilizers tested clinically bind to tubulin on sites that overlap the Taxol binding site. A new group of stabilizers, represented by laulimalide and peloriside A,

bind to a distinct, nonoverlapping site or sites on tubulin [55]. Whether stabilizers that do not interact with the Taxol binding site have anticancer actions is not yet known.

The vinca alkaloids are the prototypical microtubule depolymerizers. Several different vincas, vincristine, vinblastine, and vinorelbine, are used clinically, and newer vincas, including vinflunine, are in clinical trials. Although structurally very similar, they have some nonoverlapping indications and to some extent differing limiting toxicities. These drugs all bind to the vinca domain of tubulin. Although colchicine is not used to treat cancer, other compounds that share a binding site with colchicine are in clinical trials. There is a third binding site on tubulin that is occupied by microtubule depolymerizers; it is the peptide binding site within the vinca domain. Agents that bind within this site noncompetitively displace the vinca alkaloids. A number of drugs that bind within this site, including dolastatin 10 and cryptophycin 52, were unsuitable for the treatment for cancer. Whether these limitations are due to the binding site specifically or to the drugs themselves is not known. New tubulin binding agents may be expected to provide advantages, such as the ability to circumvent drug resistance mechanisms.

### Kinesin Inhibitors

Until recently the only antimitotic drugs were those that bound to tubulin. The identification of monastrol, an Eg5 (KSP) inhibitor, led the way for the non-tubulin-targeting antimitotics [56, 57]. Eg5 is a mitotic kinesin that is active early in mitosis to separate the centrosomes to form a bipolar mitotic spindle. Inhibition of centrosome separation causes the formation of a monopolar mitotic spindle, which is dysfunctional and leads to mitotic arrest followed by initiation of apoptosis. Eg5 is preferentially expressed in mitotically active cells and this should provide some selectivity. It is not expressed in neural tissues, and thus Eg5-targeting agents should circumvent the peripheral neuropathies that occur with tubulin-targeting antimitotics. Eg5 is attractive as a anticancer target. Optimization of the lead monastrol led to the identification of CK0106023 [58]. CK0106023 is a potent inhibitor of Eg5 and cellular proliferation. Murine antitumor trials indicated that CK0106023 has tumor growth inhibitory activities comparable to Taxol, yet at the lowest effective concentration, it did not cause body weight loss that occurs with Taxol, indicating less toxicity [58]. Monopolar spindles in tumor sections confirmed Eg5 targeting during the in vivo trial. These data suggest that Eg5 inhibitors might have promising actions against human tumors.

### Aurora Kinases

The Aurora kinases are a family of three serine/threonine kinases and two, Aurora A and B, play critical roles in mitosis and cytokinesis. Aurora A is localized at the centrosomes and mitotic spindles, and it has been referred to as the polar Aurora [59]. Aurora A is involved in centrosome separation and spindle

assembly. Aurora B is called the equatorial Aurora because it is localized at the centromere and represents a chromosomal passenger protein [59]. Aurora B plays important roles in chromosome bi-orientation, cytokinesis, and perhaps chromosome condensation. It is also a critically important part of the spindle-assembly checkpoint that functions to prevent abnormal mitosis [59]. The diverse and central roles of Aurora A and B in cell division make these kinases encouraging targets for antimitotic effects. The Auroras are all implicated in cancer progression, and overexpression of each has been identified in cancer cell lines [59]. The gene encoding Aurora A will initiate the malignant phenotype when elevated, and gene amplification may occur in some tumors [59–61]. High expression levels of Aurora A are associated with a chromosomal instability, more aggressive disease, and poorer prognosis [60]. Elevated expression of Aurora B causes cytokinesis and chromosome segregation abnormalities leading ultimately to aneuploidy [59]. Targeting the Aurora kinases may directly target a functional abnormality that contributes to the pathogenesis of cancer.

Specific inhibitors of Aurora A have been identified. VX-680 is a potent and selective inhibitor of all three Aurora kinases, and it directed against the ATP-binding site of the kinase domain [62]. VX-680 has a 30-fold selectivity for Aurora A as compared to Aurora B. It has potent antiproliferative activities against a number of colorectal, leukemia, and lymphoma cell lines, and inhibits colony growth formation [62]. Mechanistically, VX-680 causes accumulation of cells with tetraploid DNA content, which then die by apoptosis [62]. These results suggest that it does not inhibit mitosis but cytokinesis. In xenograft models of AML, VX-680 had excellent antitumor effects and it caused tumor regression in pancreatic and colon tumor models [62]. Other Aurora kinase inhibitors have been identified [63], and it will be interesting to monitor whether Aurora kinases represent a new target for anticancer therapy. Mitosis is a complex and intricately controlled cellular event that is critical to tumor growth. Additional antimitotic targets are expected to be identified [61].

## Methods in Drug Discovery

Cell-based assays constitute the primary screening methods to identify new microtubule depolymerizers and stabilizers. Many antimitotic drugs are concentrated in the cytosol at hundreds of times the concentration in the media, and this provides exquisite sensitivity. Purified tubulin-based assays are valuable as secondary confirmation screens but are limited use for primary screening because of lack of sensitivity. Cell-based screens include microscopic, flow cytometry, and cell-morphology-based screens. A phenotypic screen that simply evaluates the effects of compounds or natural product extracts on interphase smooth muscle cells led to the identification of two new classes of microtubule stabilizers, the laulimalides [64] and taccalonolides [65], and multiple

depolymerizers, including the cryptaphycins [66] and symplostatins [67]. The assay is reliable and sensitive, but has only moderately high throughput. Use of automated microscopy combined with data algorithms to detect changes in cellular microtubule density make this assay amenable to HTS.

A flow-cytometry-based method uses a mitosis-specific antigen to quantify cancer cells in mitosis [68]. It is sensitive and has advantages over other flow-based technologies in that only a single mitosis-specific parameter is measured. This assay provides moderate throughout, but with a 96-well flow cytometry format it could be used for HTS. This assay has been used to identify a number of different new microtubule depolymerizing agents, and it has an advantage over the other two assays in that it can identify tubulin-independent antimitotics [68].

The ability of dibutyryl-cyclic AMP (db-cAMP) to cause glioma cells to round up due to tubulin polymerization is the basis for a third screening assay for tubulin-interacting agents [69]. In the presence of microtubule depolymerizers, dc-cAMP fails to initiate the morphology change. The morphology changes cause the cells to round up and lose adhesion and they can be easily removed by aspiration. Cells remaining can be quantified by any number of techniques. This assay is sensitive and amenable to HTS. Unlike the other assays it cannot detect microtubule stabilizers or antimitotics that work via a tubulin-independent mechanism [69].

High-throughput screening based on kinase activity of the mitotic kinases will be valuable approaches to identify specific inhibitors of the Aurora kinases. These techniques have already proved to be successful [63]. With the identification of other mitotic targets these types of assays will be invaluable. Modeling of the ATP binding site may also provide utility for optimization of Aurora kinase inhibitor leads.

## 6 NEW THERAPEUTIC TARGETS

### Proteosome Inhibition

Precise control of the activities of cellular signaling proteins is important for maintenance of homeostatic balance. Manipulation of protein transcription/translation as well as degradation provides opportunities to alter cellular protein levels. Protein ubiquitination followed by digestion by the proteosome is particularly important for the control of cell cycle regulatory proteins [70]. The rationale of targeting the proteosome is to revert the cellular levels of amplified proteins that contribute to the malignant phenotype to normal levels. The transcription factor nuclear factor-κB (NF-κB) is believed to play a key role in cancer [71]. NF-κB is involved in many cellular actions that define malignancy, including proliferation, invasion, angiogenesis and metastasis [72]. NF-κB is a central transcription factor in inflammation. The chemopreventive

actions of anti-inflammatory agents may be mediated through inhibition of NF-κB [72]. NF-κB is constitutively active in most cancer cell lines and in human cancer tissues. NF-κB matures from its precursor p105 form to its active p50 form by the actions of the proteosome. The activity of NF-κB is controlled by its associations with its inhibitory binding partner IκB. The degradation of IκB by the proteosome releases its inhibition of NF-κB. Inhibiting the proteosome can reduce the cellular levels of NFκB, a central prosurvival transcription factor, by two mechanisms, inhibiting NFκB maturation and preventing IκB degradation. The proteosome inhibitor bortezomib (Velcade, PS-341) is a reversible inhibitor of the 26S proteosome [71]. In vitro it has antiproliferative and apoptosis-inducing activity against drug-sensitive and drug-resistant multiple myeloma cells [73]. In clinical trials bortezomib was active against multiple myeloma, providing complete responses in 10 percent of patients and partial responses in 18 percent in patients with relapsed, refractory disease [71]. Preliminary indications suggest that inhibition of the proteosome may provide a new, effective anticancer target.

## Heat Shock Protein 90 (Hsp90)

Heat shock protein 90 is a new target for cancer therapy and early clinical trial results of a Hsp90 inhibitor, 17-allylamino-17-desmethoxygeldanamycin (17AAG), suggest that it may have promise as an anticancer target. Hsp90 is a molecular chaperone that together with its co-chaperones forms a complex that is involved in the maturation and folding of nascent polypeptide chains into the tertiary structures required for proper activities. Disruption of the maturation process leads to ubiquitination and subsequent degradation of the proteins by the proteosome. Several of the client proteins of Hsp90 are implicated in cancer cell growth and survival. Client proteins of Hsp90 implicated in carcinogenesis include the tyrosine kinases *v*-Src, Akt, Raf-1, Bcr-Abl, and HER2, and the transcription factors HIF-1α, mutant p53, and steroid receptors [74, 75]. Recent data shed additional light on the selectivity of Hsp90 inhibitors toward malignant cells. During oncogenic transformation it appears that the HS90 protein complex becomes activated in a superchaperone form complexed with it co-chaperones [76]. In this configuration Hsp90 has high ATPase activity and a higher binding affinity for 17-AAG, which binds to the amino-terminal ATP binding site. Targeting Hsp90 is an approach to inhibit the machinery that allows a cancer cell to maintain its multifaceted and ever-changing growth and survival properties.

Inhibitors of Hsp90, which target the N-terminal ATPase site, have been developed and one, 17AAG is in clinical trials. There are numerous problems with 17AAG including limited solubility and oral bioavailability, but it has allowed the test of proof-of-principle of targeting Hsp90 for anticancer therapy. New inhibitors of the N-terminal ATP binding site have been identified [77] and other mechanisms of interfering with Hsp90 may also be effective [78].

## 7 SOURCES OF SMALL-MOLECULE DRUG LEADS

With the advent of exciting new targets for cancer drug discovery and numerous sensitive and specific high-throughput screens, it is critically important to evaluate a wide range of chemistry to optimize drug discovery. There are advantages of both natural product sources and synthetic sources for chemical libraries. There are strong rationales for using not just one source but both for optimal chemical diversity.

### Natural Products

The chemical diversity found in nature in microorganisms, plants, and marine species has proved historically to be the best source for lead compounds for therapies for human diseases [79]. The majority of drugs that are in use for the treatment of cancer are natural products or are based on natural products (Fig. 4). An evaluation of all the cancer drugs approved by the FDA from the 1940s through 2002 showed that 40 percent of the anticancer drugs were natural products or natural product derivatives. Another 20 percent of the drugs represent synthetic compounds where the pharmacophore was defined by a natural product, and 30 percent were strictly synthetic [79]. Natural products have played an important role in the treatment of cancer over the

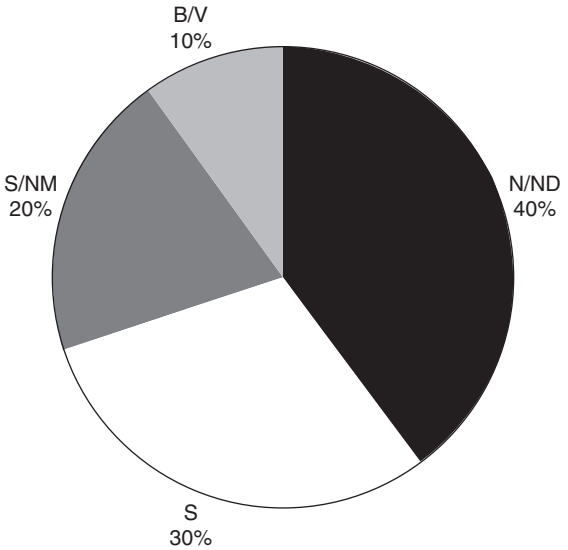


**Figure 4** Cancer drugs, 1940s–2002. Sources of drugs approved for anticancer use: natural and natural derivatives (N/ND), synthetics (S), synthetics based on natural product mimetic or pharmacophore (S/NM), and biological or vaccine (B/V). Adapted from Newman et al. [79].

past 50 years and they continue to be useful for new targets. With the disappointment of leads obtained from the first efforts of combinatorial chemistry, there is a resurgence of natural products in drug discovery. Use of natural products has been streamlined to allow for the shorter timelines of modern drug development. Many groups, including some in industry are utilizing prefractionated natural product samples, peak libraries, and are evaluating these in HTS screening assays. The use of these enriched fractions allows for accelerated lead identification.

One of the limitations of natural products can be obtaining enough material for preclinical and clinical testing. With the great strides that have been made in synthetic organic chemistry, many complex natural products can now be synthesized. This approach can often provide a source for drugs for preclinical and clinical testing, but importantly, synthetic approaches provide the opportunity for designed analogs for lead optimization. While many natural products themselves have found utility in the treatment of cancer, analogs of the natural products often have superior activities, as the source organism is not expected to have optimized the compounds for human pharmacokinetics and anticancer actions. Clearly, the complex chemicals from nature cannot yet be duplicated in synthetic chemical libraries. Natural product peak libraries are available commercially but as yet do not have the diversity that are found in academic and corporate libraries.

Supply of natural products is always a concern and has limited the clinical development of many promising lead compounds. Advances in the understanding of the genetics of complex biosynthetic pathways have been utilized in bioengineering [80]. Biosynthetic gene clusters in bacterial expression systems have been engineered to provide a renewable source of rare natural products efficiently. Manipulation of biosynthetic gene clusters provides the opportunity for metabolic engineering of a wide diversity of new chemical entities [81]. The chemical diversity of natural products as sources for new drug leads should not be overlooked and can complement large synthetic chemical libraries.

## Synthetic Chemical Libraries

Synthetic chemical libraries have many advantages over natural product libraries in that they represent structures easily amenable to chemical synthesis. Large chemical libraries, 100,000+ are available commercially for drug discovery screening. They are available in formats amenable to HTS and provide good chemical diversity at low cost. Many of the libraries are optimized for chemical diversity and for pharmaceutical characteristics. These libraries, together with HTS have been very important in the identification of new, targeted leads.

One of the most promising approaches for new drug discovery is combinatorial chemistry. Combinatorial chemistry is a relatively new technology for creating molecules en masse, compared with conventional one-molecule-at-a-

time medicinal chemistry. The technology evolved from solid-phase peptide synthesis. Originally used to create polypeptide and oligonucleotide libraries, the field has moved toward libraries of small organic molecules with potential utility as drugs. Molecular diversity can be introduced into many different molecules being prepared in parallel while tethered by a reversible linker to an insoluble polymeric support. Although initially limited to peptide molecules, the method has been expanded to include carbohydrates.

Validation of the library of the compounds produced in a combinatorial synthesis has proven crucial to success. Early practitioners failed to recognize that not all reactions worked as predicted, leading to unpredicted by-products and incomplete reactions, leading to "dirty" libraries that created irreproducible effects. The utility of the combinatorial approach for creation of fundamentally different structures is hotly contested. It is clear, however, that for lead optimization by analog synthesis, combinatorial chemistry is working and is making major contributions.

## 8 RESOURCES FOR DRUG DISCOVERY

The approaches and targets for cancer drug discovery are changing at a rapid pace, and, as mentioned in the introduction, significant efforts are being placed on cancer drug discovery of targets that have yet to be validated. It can be anticipated that some targets will lead to successful pharmaceuticals for cancer treatment while others, for a number of reasons, will fail. Keeping current on the trends in cancer drug discovery is critically important.

A wealth of information about drug discovery in general and specifics about new screening technologies has become available in the past few years. A number of excellent new journals are published with the aims of helping the scientific community keep informed about the trends of drug discovery and in specific assay techniques. The Society of Biomolecular Screening was initiated in 1994 with the goal of "Discovery through Community." The society's journal, *The Journal of Biomolecular Screening* is published on a monthly basis and provides a variety of information on specific screening approaches and technologies. National and regional meetings are sponsored by the society. The inaugural issue of *Current Drug Discovery Technologies* was published in January of 2004. The aims and scope of the journal are to provide timely information on new approaches for all aspects of modern drug discovery. *Nature Reviews Drug Discovery*, which began publication in 2002, offers an overview of the drug discovery process, review articles for new therapeutic targets, as well as articles on new tools for drug discovery. These new resources allow rapid communications of approaches and solutions for various aspects of drug discovery.

The focus of this chapter is on initial drug discovery, but turning lead identification into a clinically useful anticancer therapy requires rational drug development, which is even more complex than initial drug discovery. A new

reference, *Molecular Cancer Therapeutics: Strategies for Drug Discovery and Development* [82], was published this year and provides an excellent source for information on various aspects of cancer drug development.

***Acknowledgments*** I apologize to the many scientists who have made important contributions to cancer drug discovery and whose work and approaches were not cited due to space considerations. The goal of this chapter is to provide an introduction to the field of oncology drug discovery, and, due to the large number of exciting new targets for drug discovery, many valid and promising targets have not been discussed. Special thanks are extended to Ms. April Hopstetter and Ms. Evelyn Jackson for their excellent help with this chapter; to Drs. Pat LoRusso and Basel El-Rayes for use of their EGFR-signaling pathway figure; and to Dr. David Newman for permission to adapt his figure on the sources of anticancer drugs.

## REFERENCES

1. Booth, B., Glassman, R., Ma, P. (2003). Oncology's trials. *Nat. Rev. Drug Discov. 2*, 609–610.
2. Suffness, M., Pezzuto, J. M. (1991). Assays related to cancer drug discovery. In K. Hostettmann (Ed.), *Assays for Bioactivity*. Academic, New York, pp. 71–133.
3. Drebin, J. A., Link, V. C., Weinberg, R. A., Greene, M. I. (1986). Inhibition of tumor growth by a monoclonal antibody reactive with an oncogene-encoded tumor antigen. *Proc. Natl. Acad. Sci. USA*, *83*, 9129–9133.
4. El-Rayes, B. F., LoRusso, P. M. (2004). Targeting the epidermal growth factor receptor. *Br. J. Cancer*, *91*, 418–424.
5. Shawver, L. K., Slamon, D., Ullrich, A. (2002). Smart drugs: Tyrosine kinase inhibitors in cancer therapy. *Cancer Cell*, *1*, 117–123.
6. Laskin, J. J., Sandler, A. B. (2004). Epidermal growth factor receptor: A promising target in solid tumours. *Cancer Treat. Rev.*, *30*, 1–17.
7. Lynch, T. J., Bell, D. W., Sordella, R., Gurubhagavatula, S., Okimoto, R. A., Brannigan, B. W., Harris, P. L., Haserlat, S. M., Supko, J. G., Haluska, F. G., Louis, D. N., Christiani, D. C., Settleman, J., Haber, D. A. (2004). Activating mutations in the epidermal growth factor receptor underlying responsiveness of non-small-cell lung cancer to gefitinib. *N. Engl. J. Med.*, *350*, 2129–2139.
8. Paez, J. G., Janne, P. A., Lee, J. C., Tracy, S., Greulich, H., Gabriel, S., Herman, P., Kaye, F. J., Lindeman, N., Boggon, T. J., Naoki, K., Sasaki, H., Fujii, Y., Eck, M. J., Sellers, W. R., Johnson, B. E., Meyerson, M. (2004). EGFR mutations in lung cancer: Correlation with clinical response to gefitinib therapy. *Science*, *304*, 1497–1500.
9. Sordella, R., Bell, D. W., Haber, D. A., Settleman, J. (2004). Gefitinib-sensitizing EGFR mutations in lung cancer activate anti-apoptotic pathways. *Science*, *305*, 1163–1167.
10. Chan, J. M., Stampfer, M. J., Giovannucci, E., Gann, P. H., Ma, J., Wilkinson, P., Hennekens, C. H., Pollak, M. (1998). Plasma insulin-like growth factor-I and prostate cancer risk: A prospective study. *Science*, *279*, 563–566.

11. Wu, Y., Cui, K., Miyoshi, K., Hennighausen, L., Green, J. E., Setser, J., LeRoith, D., Yakar, S. (2003). Reduced circulating insulin-like growth factor I levels delay the onset of chemically and genetically induced mammary tumors. *Cancer Res.*, *63*, 4384–4388.
12. LeRoith, D., Helman, L. (2004). The new kid on the block(ade) of the IGF-1 receptor. *Cancer Cell*, *5*, 201–202.
13. Wong, S., McLaughlin, J., Cheng, D., Zhang, C., Shokat, K. M., Witte, O. N. (2004). Sole BCR-ABL inhibition is insufficient to eliminate all myeloproliferative disorder cell populations. *Proc. Natl. Acad. Sci. USA*. *101*, 17456–17461.
14. Bjornsti, M. A., Houghton, P. J. (2004). The TOR pathway: A target for cancer therapy. *Nat. Rev. Cancer*, *4*, 335–348.
15. Hay, N., Sonenberg, N. (2004). Upstream and downstream of mTOR. *Genes Dev.*, *18*, 1926–1945.
16. Sawyers, C. L. (2003). Will mTOR inhibitors make it as cancer drugs? *Cancer Cell*, *4*, 343–348.
17. Slamon, D. J. (2004). The FUTURE of ErbB-1 and ErbB-2 pathway inhibition in breast cancer: Targeting multiple receptors. *Oncologist*, *9*(Suppl. 3), 1–3.
18. Mallari, R., Swearingen, E., Liu, W., Ow, A., Young, S. W., Huang, S. G. (2003). A generic high-throughput screening assay for kinases: Protein kinase a as an example. *J. Biomol. Screen.*, *8*, 198–204.
19. Moshinsky, D. J., Ruslim, L., Blake, R. A., Tang, F. (2003). A widely applicable, high-throughput TR-FRET assay for the measurement of kinase autophosphorylation: VEGFR-2 as a prototype. *J. Biomol. Screen.*, *8*, 447–452.
20. Gaudet, E. A., Huang, K. S., Zhang, Y., Huang, W., Mark, D., Sportsman, J. R. (2003). A homogeneous fluorescence polarization assay adaptable for a range of protein serine/threonine and tyrosine kinases. *J. Biomol. Screen.*, *8*, 164–175.
21. Newman, M., Josiah, S. (2004). Utilization of fluorescence polarization and time resolved fluorescence resonance energy transfer assay formats for SAR studies: Src kinase as a model system. *J. Biomol. Screen.*, *9*, 525–532.
22. Anderson, S. N., Cool, B. L., Kifle, L., Chiou, W., Egan, D. A., Barrett, L. W., Richardson, P. L., Frevert, E. U., Warrior, U., Kofron, J. L., Burns, D. J. (2004). Microarrayed compound screening (microARCS) to identify activators and inhibitors of AMP-activated protein kinase. *J. Biomol. Screen.*, *9*, 112–121.
23. Folkman, J. (1990). What is the evidence that tumors are angiogenesis dependent? *J. Natl. Cancer Inst.*, *82*, 4–6.
24. Carmeliet, P., Jain, R. K. (2000). Angiogenesis in cancer and other diseases. *Nature*, *407*, 249–257.
25. Scavelli, C., Weber, E., Agliano, M., Cirulli, T., Nico, B., Vacca, A., Ribatti, D. (2004). Lymphatics at the crossroads of angiogenesis and lymphangiogenesis. *J. Anat.*, *204*, 433–449.
26. Folkman, J. (1971). Tumor angiogenesis: Therapeutic implications. *N. Engl. J. Med.*, *285*, 1182–1186.
27. Hanahan, D., Folkman, J. (1996). Patterns and emerging mechanisms of the angiogenic switch during tumorigenesis. *Cell*, *86*, 353–364.
28. Streubel, B., Chott, A., Huber, D., Exner, M., Jager, U., Wagner, O., Schwarzinger, I. (2004). Lymphoma-specific genetic aberrations in microvascular endothelial cells in B-cell lymphomas. *N. Engl. J. Med.*, *351*, 250–259.

29. Ferrara, N., Hillan, K. J., Gerber, H. P., Novotny, W. (2004). Discovery and development of bevacizumab, an anti-VEGF antibody for treating cancer. *Nat. Rev. Drug. Discov.*, *3*, 391–400.
30. Ferrara, N., Gerber, H. P., LeCouter, J. (2003). The biology of VEGF and its receptors. *Nat. Med.*, *9*, 669–676.
31. Ferrara, N. (2004). Vascular endothelial growth factor: Basic science and clinical progress. *Endocr. Rev.*, *25*, 581–611.
32. McDonald, D. M., Teicher, B. A., Stetler-Stevenson, W., Ng, S. S., Figg, W. D., Folkman, J., Hanahan, D., Auerbach, R., O'Reilly, M., Herbst, R., Cheresh, D., Gordon, M., Eggermont, A., Libutti, S. K. (2004). Report from the society for biological therapy and vascular biology faculty of the NCI workshop on angiogenesis monitoring. *J. Immunother.*, *27*, 161–175.
33. Hiratsuka, S., Nakamura, K., Iwai, S., Murakami, M., Itoh, T., Kijima, H., Shipley, J. M., Senior, R. M., Shibuya, M. (2002). MMP9 induction by vascular endothelial growth factor receptor-1 is involved in lung-specific metastasis. *Cancer Cell*, *2*, 289–300.
34. Ferrara, N. (2004). Vascular endothelial growth factor as a target for anticancer therapy. *Oncologist*, *9*(Suppl. 1), 2–10.
35. Tandle, A., Libutti, S. K., Libutti, S. K. (2003). Antiangiogenic therapy: Targeting vascular endothelial growth factors and its receptors. *Clin. Adv. Hematol. Oncol.*, *1*, 41–48.
36. Siemann, D. W., Chaplin, D. J., Horsman, M. R. (2004). Vascular-targeting therapies for treatment of malignant disease. *Cancer*, *100*, 2491–2499.
37. Powis, G., Kirkpatrick, L. (2004). Hypoxia inducible factor-1alpha as a cancer drug target. *Mol. Cancer Ther.*, *3*, 647–654.
38. Albini, A. (1998). Tumor and endothelial cell invasion of basement membranes. The matrigel chemoinvasion assay as a tool for dissecting molecular mechanisms. *Pathol. Oncol. Res.*, *4*, 230–241.
39. Walter-Yohrling, J., Morgenbesser, S., Rouleau, C., Bagley, R., Callahan, M., Weber, W., Teicher, B. A. (2004). Murine endothelial cell lines as models of tumor endothelial cells. *Clin. Cancer Res.*, *10*, 2179–2189.
40. Lawson, N. D., Weinstein, B. M. (2002). In vivo imaging of embryonic vascular development using transgenic zebrafish. *Dev. Biol.*, *248*, 307–318.
41. Ausprunk, D. H., Knighton, D. R., Folkman, J. (1975). Vascularization of normal and neoplastic tissues grafted to the chick chorioallantois. Role of host and preexisting graft blood vessels. *Am. J. Pathol.*, *79*, 597–628.
42. Passaniti, A., Taylor, R. M., Pili, R., Guo, Y., Long, P. V., Haney, J. A., Pauly, R. R., Grant, D. S., Martin, G. R. (1992). A simple, quantitative method for assessing angiogenesis and antiangiogenic agents using reconstituted basement membrane, heparin, and fibroblast growth factor. *Lab. Invest.*, *67*, 519–528.
43. Tozer, G. M. (2003). Measuring tumour vascular response to antivascular and antiangiogenic drugs. *Br. J. Radiol.*, *76*(Suppl. 1), S23–25.
44. Cosgrove, D. (2003). Angiogenesis imaging—ultrasound. *Br. J. Radiol.*, *76*(Suppl. 1), S43–49.
45. Hsieh, C. L., Jones, P. A. (2003). Meddling with methylation. *Nat. Cell. Biol.*, *5*, 502–504.

46. Szyf, M., Pakneshan, P., Rabbani, S. A. (2004). DNA demethylation and cancer: Therapeutic implications. *Cancer Lett.*, *211*, 133–143.
47. Brown, R., Plumb, J. A. (2004). Demethylation of DNA by decitabine in cancer chemotherapy. *Expert Rev. Anticancer Ther.*, *4*, 501–510.
48. Somech, R., Izraeli, S., Simon, A. J. (2004). Histone deacetylase inhibitors—a new tool to treat cancer. *Cancer Treat. Rev.*, *30*, 461–472.
49. Villar-Garea, A., Esteller, M. (2004). Histone deacetylase inhibitors: Understanding a new wave of anticancer agents. *Int. J. Cancer*, *112*, 171–178.
50. Shay, J. W. (2003). Telomerase therapeutics: Telomeres recognized as a DNA damage signal: Commentary re: K. Kraemer et al., antisense-mediated hTERT inhibition specifically reduces the growth of human bladder cancer cells. *Clin. Cancer Res.*, *9*, 3521–3525.
51. Shay, J. W., Wright, W. E. (2002). Telomerase: A target for cancer therapeutics. *Cancer Cell*, *2*, 257–265.
52. Parkinson, E. K. (2003). Telomerase as a novel and potentially selective target for cancer chemotherapy. *Ann. Med.*, *35*, 466–475.
53. O'Connor, P. M., Jackman, J., Bae, I., Myers, T. G., Fan, S., Mutoh, M., Scudiero, D. A., Monks, A., Sausville, E. A., Weinstein, J. N., Friend, S., Fornace, A. J., Jr. Kohn, K. W. (1997). Characterization of the p53 tumor suppressor pathway in cell lines of the National Cancer Institute anticancer drug screen and correlations with the growth-inhibitory potency of 123 anticancer agents. *Cancer Res.*, *57*, 4285–4300.
54. Wilson, L., Jordan, M. A. (1995). Microtubule dynamics: Taking aim at a moving target. *Chem. Biol.*, *2*, 569–573.
55. Pryor, D. E., O'Brate, A., Bilcer, G., Diaz, J. F., Wang, Y., Kabaki, M., Jung, M. K., Andreu, J. M., Ghosh, A. K., Giannakakou, P., Hamel, E. (2002). The microtubule stabilizing agent laulimalide does not bind in the taxoid site, kills cells resistant to paclitaxel and epothilones, and may not require its epoxide moiety for activity. *Biochemistry*, *41*, 9109–9115.
56. Mayer, T. U., Kapoor, T. M., Haggarty, S. J., King, R. W., Schreiber, S. L., Mitchison, T. J. (1999). Small molecule inhibitor of mitotic spindle bipolarity identified in a phenotype-based screen. *Science*, *286*, 971–974.
57. Maliga, Z., Kapoor, T. M., Mitchison, T. J. (2002). Evidence that monastrol is an allosteric inhibitor of the mitotic kinesin Eg5. *Chem. Biol.*, *9*, 989–996.
58. Sakowicz, R., Finer, J. T., Beraud, C., Crompton, A., Lewis, E., Fritsch, A., Lee, Y., Mak, J., Moody, R., Turincio, R., Chabala, J. C., Gonzales, P., Roth, S., Weitman, S., Wood, K. W. (2004). Antitumor activity of a kinesin inhibitor. *Cancer Res.*, *64*, 3276–3280.
59. Carmena, M., Earnshaw, W. C. (2003). The cellular geography of aurora kinases. *Nat. Rev. Mol. Cell. Biol.*, *4*, 842–854.
60. Meraldi, P., Honda, R., Nigg, E. A. (2004). Aurora kinases link chromosome segregation and cell division to cancer susceptibility. *Curr. Opin. Genet. Dev.*, *14*, 29–36.
61. Sausville, E. A. (2004). Aurora kinases dawn as cancer drug targets. *Nat. Med.*, *10*, 234–235.
62. Harrington, E. A., Bebbington, D., Moore, J., Rasmussen, R. K., Ajose-Adeogun, A. O., Nakayama, T., Graham, J. A., Demur, C., Hercend, T., Diu-Hercend, A., Su, M., Golec, J. M., Miller, K. M. (2004). VX-680, a potent and selective small-

molecule inhibitor of the Aurora kinases, suppresses tumor growth in vivo. *Nat. Med.*, *10*, 262–267.

63. Ditchfield, C., Johnson, V. L., Tighe, A., Ellston, R., Haworth, C., Johnson, T., Mortlock, A., Keen, N., Taylor, S. S. (2003). Aurora B couples chromosome alignment with anaphase by targeting BubR1, Mad2, and Cenp-E to kinetochores. *J. Cell. Biol.*, *161*, 267–280.
64. Mooberry, S. L., Tien, G., Hernandez, A. H., Plubrukarn, A., Davidson, B. S. (1999). Laulimalide and isolaulimalide, new paclitaxel-like microtubule-stabilizing agents. *Cancer Res.*, *59*, 653–660.
65. Tinley, T. L., Randall-Hlubek, D. A., Leal, R. M., Jackson, E. M., Cessac, J. W., Quada, J. C., Jr., Hemscheidt, T. K., Mooberry, S. L. (2003). Taccalonolides E and A: Plant-derived steroids with microtubule-stabilizing activity. *Cancer Res.*, *63*, 3211–3220.
66. Smith, C. D., Zhang, X., Mooberry, S. L., Patterson, G. M., Moore, R. E. (1994). Cryptophycin: A new antimicrotubule agent active against drug-resistant cells. *Cancer Res.*, *54*, 3779–3784.
67. Mooberry, S. L., Leal, R. M., Tinley, T. L., Luesch, H., Moore, R. E., Corbett, T. H. (2003). The molecular pharmacology of symplostatin 1: A new antimitotic dolastatin 10 analog. *Int. J. Cancer*, *104*, 512–521.
68. Anderson, H. J., de Jong, G., Vincent, I., Roberge, M. (1998). Flow cytometry of mitotic cells. *Exp. Cell. Res.*, *238*, 498–502.
69. Kokoshka, J. M., Ireland, C. M., Barrows, L. R. (1996). Cell-based screen for identification of inhibitors of tubulin polymerization. *J. Nat. Prod.*, *59*, 1179–1182.
70. King, R. W., Deshaies, R. J., Peters, J. M., Kirschner, M. W. (1996). How proteolysis drives the cell cycle. *Science*, *274*, 1652–1659.
71. Paramore, A., Frantz, S. (2003). Bortezomib. *Nat. Rev. Drug Discov.*, *2*, 611–612.
72. Aggarwal, B. B. (2004). Nuclear factor-kappaB: The enemy within. *Cancer Cell*, *6*, 203–208.
73. Hideshima, T., Richardson, P., Chauhan, D., Palombella, V. J., Elliott, P. J., Adams, J., Anderson, K. C. (2001). The proteasome inhibitor PS-341 inhibits growth, induces apoptosis, and overcomes drug resistance in human multiple myeloma cells. *Cancer Res.*, *61*, 3071–3076.
74. Neckers, L. (2002). Hsp90 inhibitors as novel cancer chemotherapeutic agents. *Trends Mol. Med.*, *8*, S55–61.
75. Isaacs, J. S., Xu, W., Neckers, L. (2003). Heat shock protein 90 as a molecular target for cancer therapeutics. *Cancer Cell*, *3*, 213–217.
76. Kamal, A., Thao, L., Sensintaffar, J., Zhang, L., Boehm, M. F., Fritz, L. C., Burrows, F. J. (2003). A high-affinity conformation of Hsp90 confers tumour selectivity on Hsp90 inhibitors. *Nature*, *425*, 407–410.
77. Vilenchik, M., Solit, D., Basso, A., Huezo, H., Lucas, B., He, H., Rosen, N., Spampinato, C., Modrich, P., Chiosis, G. (2004). Targeting wide-range oncogenic transformation via PU24FCl, a specific inhibitor of tumor Hsp90. *Chem. Biol.*, *11*, 787–797.
78. Chiosis, G., Vilenchik, M., Kim, J., Solit, D. (2004). Hsp90: The vulnerable chaperone. *Drug. Discov. Today*, *9*, 881–888.

79. Newman, D. J., Cragg, G. M., Snader, K. M. (2003). Natural products as sources of new drugs over the period 1981–2002. *J. Nat. Prod.*, *66*, 1022–1037.
80. Kerwin, S. (2002). Toward bioengineering anticancer drugs. *Chem. Biol.*, *9*, 956–958.
81. Sherman, D. H. (2002). New enzymes for "warheads." *Nat. Biotechnol.*, *20*, 984–985.
82. Prendergast, G. C. (Ed.) (2004). *Molecular Cancer Therapeutics: Strategies for Drug Discovery and Development.* Wiley, New York.

# MODELING AND INFORMATICS IN DRUG DESIGN

Prasad V. Bharatam,* Smriti Khanna, and Sandrea M. Francis
*National Institute of Pharmaceutical Education and Research (NIPER), S.A.S. Nagar, India*

**Contents**

1 Introduction
2 Computational Chemistry
  2.1 *Ab Initio* Quantum Chemical Methods
  2.2 Semiempirical Methods
  2.3 Molecular Mechanical Methods
  2.4 Energy Minimization and Geometry Optimization
  2.5 Conformational Analysis
3 Computational Biology
  3.1 *Ab Initio* Structure Prediction
  3.2 Homology Modeling
  3.3 Threading or Remote Homology Modeling
4 Computational Medicinal Chemistry
  4.1 Quantitative Structure–Activity Relationship (QSAR)
  4.2 Pharmacophore Mapping
  4.3 Molecular Docking
  4.4 *De Novo* Design
5 Pharmacoinformatics
  5.1 Chemoinformatics
  5.2 Bioinformatics
  5.3 Virtual Screening
  5.4 Neuroinformatics
  5.5 Immunoinformatics
  5.6 Drug Metabolism Informatics
  5.7 Toxicoinformatics
  5.8 Cancer Informatics
6 Future Scope
  References

**Corresponding author*.

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

## 1 INTRODUCTION

Modeling and informatics have become indispensable components of rational drug design (Fig. 1). For the last few years, chemical analysis through molecular modeling has been very prominent in computer-aided drug design (CADD). But currently modeling and informatics are contributing in tandem toward CADD. Modeling in drug design has two facets: modeling on the basis of knowledge of the drugs/leads/ligands often referred to as ligand-based design and modeling based on the structure of macromolecules often referred to as receptor-based modeling (or structure-based modeling). Computer-aided drug design is a topic of medicinal chemistry, and before venturing into this exercise one must employ computational chemistry methods to understand the properties of chemical species, on the one hand, and employ computational biology techniques to understand the properties of biomolecules on the other. Information technology is playing a major role in decision making in pharmaceutical sciences. Storage, retrieval, and analysis of data of chemicals/biochemicals of therapeutic interest are major components of pharmacoinformatics. Quite

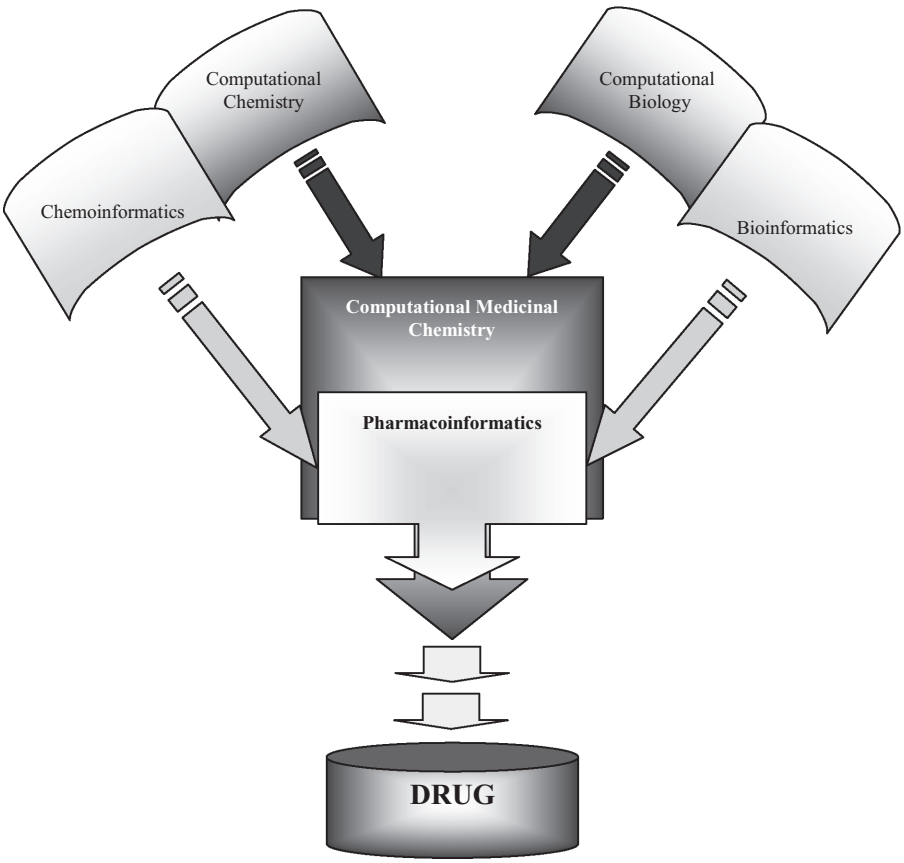


**FIGURE 1** A schematic diagram showing a flowchart of activities in computer aided drug development. The figure shows that the contributions from modeling methods and informatics methods toward the drug development are parallel and in fact not really distinguishable.

often, the efforts based on modeling and informatics get thoroughly integrated with each other, as in the case of virtual screening exercises. In this chapter, the molecular modeling methods that are in vogue in the fields of (1) computational chemistry, (2) computational biology, (3) computational medicinal chemistry, and (4) pharmacoinformatics are presented and the resources available in these fields are discussed.

## 2 COMPUTATIONAL CHEMISTRY

Two-dimensional (2D) structure drawing and three-dimensional (3D) structure building are the important primary steps in computational chemistry for which several molecular visualization packages are available. The most popular of these are ChemDraw Ultra and Chem3D Pro, which are a part of the ChemOffice suite of software packages [1]. ACD/ChemSketch [2], MolSuite [3], and many more of this kind are other programs for the same purpose. Refinement has to be carried out on all the drawings and 3D structures so as to improve the chemical accuracy of the structure on the computer screen. Structure refinement based on heuristic rules/cleanup procedures is a part of all these software packages. However, chemical accuracy of the 3D structures still remains poor even after cleanup. Further refinement can be carried out by performing energy minimization using either molecular mechanical or quantum chemical procedures. By using these methods, the energy of a molecule can be estimated in any given state. Following this, with the help of first and second derivatives of energy, it can be ascertained whether the given computational state of the molecules belongs to a chemically acceptable state or not. During this process, the molecular geometry gets modified to a more appropriate, chemically meaningful state – the entire procedure is known as geometry optimization. The geometry optimized 3D structure is suitable for property estimation, descriptor calculation, conformational analysis, and finally for drug design exercise [4–6].

### 2.1 Ab Initio Quantum Chemical Methods

Every molecule possesses internal energy ($U$), for the estimation of which quantum chemical calculations are suitable. Quantum chemical calculations involve rigorous mathematical derivations and attempt to solve the Schrödinger equation, which in its simplest form may be written as

$$H\Psi = E\Psi \tag{1}$$

$$\hat{H}_{\mathrm{el}} = \sum\left(-\frac{1}{2}\nabla_i^2\right) - \sum_i\sum_a \frac{Z_a}{|r_i - d_a|} + \frac{1}{2}\sum_i\sum_{j\neq i}\frac{1}{|r_i - r_j|} + \frac{1}{2}\sum_a\sum_{b\neq a}\frac{Z_a Z_b}{|d_a - d_b|} \tag{2}$$

where $\psi$ represents the wavefunction, $E$ represents energy, $\nabla$ represents the kinetic energy operator for electrons, $r_i$ defines the vector position of electron $i$ with vector components in Bohr radii, $Z_a$ is the charge of fixed nucleus $a$ in units of the elementary charge, and $d_a$ is the vector position of nucleus $a$ with vector components in Bohr radii.

Exact solutions to Schrödinger equation cannot be provided for systems with more than one electron. Several *ab initio* molecular orbital (MO) and *ab initio* density functional theory (DFT) methods were developed to provide expectation value for the energy. This energy can be minimized and thus the geometry of any molecule can be obtained, with high confidence level, using quantum chemical methods. During this energy estimation, the wavefunctions of every molecule can be defined, which possess all the information related to the molecule. Thus, properties like relative energies, dipole moments, electron density distribution, charge distribution, electron delocalization, molecular orbital energies, molecular orbital shapes, ionization potential, infrared (IR) frequencies, and chemical shifts can be estimated using *ab initio* computational chemistry methods. For this purpose, several quantum chemical methods like Hartree–Fock (HF), second order Moller–Plesset perturbation (MP2), coupled cluster (CCSD), configuration interaction (QCISD), many-body perturbation (MBPT), multiconfiguration self-consistent field (MCSCF), complete active space self-consistent field (CASSCF), B3LYP, and VWN were developed. At the same time to define the wavefunction, a set of mathematical functions known as *basis set* is required. Typical basis sets are 3-21G, 6-31G*, and 6-31+G*. Combination of the *ab initio* methods and basis sets leads to several thousand options for estimating energy. For reliable geometry optimization of drug molecules, the HF/6-31+G*, MP2/6-31+G*, and B3LYP/6-31+G* methods are quite suitable. When very accurate energy estimation is required, G2MP2 and CBS-Q methods can be employed. Gaussian03, Spartan, and Jaguar are software packages that can be used to estimate reliable geometry optimization and very accurate energy estimation of any chemical species. In practice, quantum chemical methods are being used to estimate the relative stabilities of molecules, to calculate properties of reaction intermediates, to investigate the mechanisms of chemical reactions, to predict the aromaticity of compounds, and to analyze spectral properties. Medicinal chemists are beginning to take benefit from these by studying drug–receptor interactions, enzyme–substrate binding, and solvation of biological molecules. Molecular electrostatic potentials, which can be derived from *ab initio* quantum chemical methods, provide the surface properties of drugs and receptors and thus they offer useful information regarding complementarities between the two [7–10].

## 2.2 Semiempirical Methods

The above defined *ab initio* methods are quite time consuming and become prohibitively expensive when the drugs possess large number of atoms and/or a series of calculations need to be performed to understand the chemical phenomena. Semiempirical quantum chemical methods were introduced precisely to address this problem. In these methods empirical parameters are employed to estimate many integrals but only a few key integrals are solved explicitly. Although these calculations do not provide energy of molecules, they are quite reliable in estimating the heats of formation. Semiempirical quantum chemical methods (e.g., AM1, PM3, SAM1) are very fast in qualitatively estimating the chemical properties that are of interest to a drug discovery scientist. MOPAC and AMPAC are the software packages of choice; however, many other software packages also incorporate these methods. Qualitative estimates of HOMO and LUMO energies, shapes of molecular orbitals, and reaction mechanisms of drug synthesis are some of the applications of

semiempirical analysis [5, 10]. When the molecules become much larger, especially in the case of macromolecules like proteins, enzymes, and nucleic acids, employing these semiempirical methods becomes impractical. In such cases, molecular mechanical methods can be used to estimate the heats of formation and to perform geometry optimization.

### 2.3 Molecular Mechanical Methods

Molecular mechanical methods estimate the energy of any drug by adding up the strain in all the bonds, angles, and torsions due to the energy of the van der Waals and Coulombic interactions across all atoms in the molecule. It reflects the internal energy of the molecule; although the estimated value is nowhere close to the actual internal energy, the relative energy obtained from these methods is indicative enough for chemical/biochemical analysis. It is made up of a number of components as given by

$$E_{\mathrm{mm}} = E_{\mathrm{bonds}} + E_{\mathrm{angles}} + E_{\mathrm{vdw}} + E_{\mathrm{torsion}} + E_{\mathrm{charge}} + E_{\mathrm{misc}} \tag{3}$$

Molecular mechanical methods are also known as force field methods because in these methods, the electronic effects are estimated implicitly in terms of force fields associated with the atoms. In Eq. 3, the energy ($E$) due to bonds, angles, and torsional angles can be estimated using the simple Hooke's law and its variations, whereas the van der Waals (vdw) interactions are estimated using the Lennard-Jones potential and the electrostatic interactions are estimated using Coulombic forces. The energy estimation, energy minimization, and geometry optimization using these methods are quite fast and hence suitable for studying the geometries and conformations of biomolecules and drug–receptor interactions. Since these methods are empirical in nature, parameterization of the force fields with the help of available spectral data or quantum chemical methods is required. AMBER, CHARMM, UFF, and Tripos are some of the force fields in wide use in computer-aided drug development [5, 10].

### 2.4 Energy Minimization and Geometry Optimization

Drug molecules prefer to adopt equilibrium geometry in nature, that is, a geometry that possesses a stable 3D arrangement of atoms in the molecule. The 3D structure of a molecule built using a 3D builder does not represent a natural state; slight modifications are required to be made on the built 3D structure so that it represents the natural state. For this purpose, the following questions need to be addressed: (1) Which minimal changes need to be made? (2) How much change needs to be made? (3) How does one know the representation at hand is the true representation of the natural state? To provide answers to these questions we can depend on energy, because molecules prefer to exist in thermodynamically stable states. This implies that if the energy of any molecule can be minimized, the molecule is not in a stable state and thus the current representation of the molecule may not be the true representation of the natural state. This also implies that we can minimize the energy and the molecular structure in that energy minimum state probably represents a true natural state. Several methods of energy minimization have been developed by

computational chemists, some of which are nonderivative methods (simplex method) but many of which are dependent on derivative methods (steepest descent, Newton–Raphson, conjugate gradient, variable metrics, etc.) and involve the estimation of the gradient of the potential energy curve [4–6]. The entire procedure of geometry modification to reach an energy minimum state with almost null gradient is known as geometry optimization in terms of the structure of the molecule and energy minimization in terms of the energy of the molecule. All computational chemistry software packages are equipped with energy minimization methods—of which a few incorporate energy minimization based on *ab initio* methods while most include the semiempirical and molecular mechanics based energy minimization methods.

## 2.5 Conformational Analysis

Molecules containing freely rotatable bonds can adopt many different conformations. Energy minimization procedures lead the molecular structure to only one of the chemically favorable conformations, called the local minimum. Out of the several local minima on the potential energy (PE) surface of a molecule, the lowest energy conformation is known as the global minimum. It is important to note all the possible conformations of any molecule and identify the global minimum before taking up a drug design exercise (Fig. 2). This is important because only one of the possible conformations of a drug, known as its bioactive conformation, is responsible for its therapeutic effect. This conformation may be a global minimum, a local minimum, or a transition state between local minima. As it is very difficult to identify the

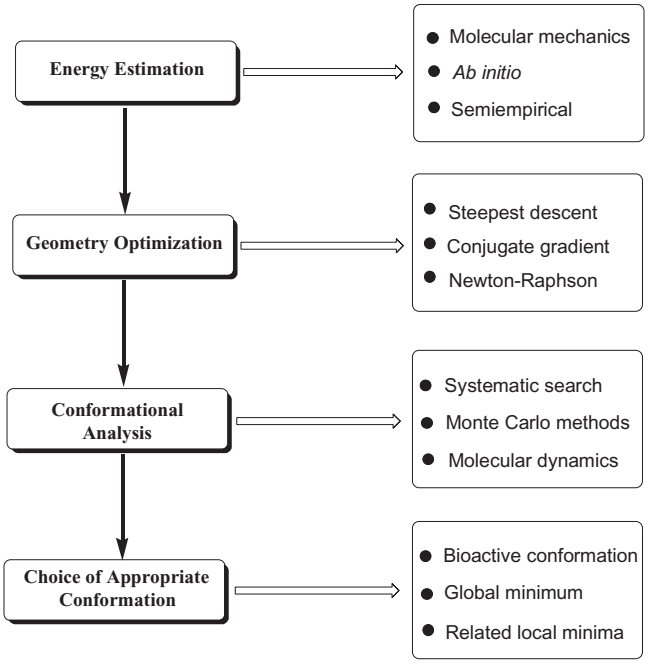


**FIGURE 2** Flowchart showing the sequence of steps during molecular modeling of drug molecules.

bioactive conformation of many drug molecules, it is common practice to assume the global minima to be bioactive. The transformation of drug molecules from one conformer to another can be achieved by changing the torsional angles. The computational process of identifying all local minima of a drug molecule, identifying the global minimum conformation, and, if possible, identifying the bioactive conformation is known as conformational analysis. This is one of the major activities in computational chemistry.

Manual conformational search is one method where the chemical intuition of the chemist plays a major role in performing the conformational analysis. Here, a chemist/modeler carefully chooses all possible conformations of a given drug molecule and estimates the energy of each conformation after performing energy minimization. This procedure is very effective and is being widely used. This approach allows the application of rigorous quantum chemical methods for the conformational analysis. The only limitation of this method arises from the ability and patience of the chemist. There is a possibility that a couple of important conformations are ignored in this approach. To avoid such problems, automated conformational analysis methods were introduced.

Various automated methods of conformational analysis include systematic search, random search, Monte Carlo simulations, molecular dynamics, genetic algorithms, and expert systems (Table 1) [4, 5]. The systematic conformational search can be performed by varying systematically each of the torsion angles of the rotatable bonds of a molecule to generate all possible conformations. The step size for torsion angle change is normally 30–60°. The number of conformations across a C—C single bond would vary between 6 and 12. With an increase in the number of rotatable bonds, the total number of conformations generated becomes quite large. The "bump check" method reduces the number of possibilities; still, the total number of conformations generated can be in the tens of thousands for drug molecules. Obviously, most of the conformations are chemically nonsignificant.

The random conformational search method employs random change in torsional angle across rotatable bonds and performs energy minimization each time; thus, a handful of chemically meaningful conformations can be generated [11–18].

Molecular dynamics is another method of carrying out conformational search of flexible molecules. The aim of this approach is to reproduce time-dependent motional behavior of a molecule, which can identify bound states out of several possible

**TABLE 1 Different Methods of Conformational Analysis**

| Methods for Conformational Analysis | Remarks |
|---|---|
| Systematic search | Systematic change of torsions |
| Random search | Conformations picked up randomly |
| Monte Carlo method | Supervised random search |
| Molecular dynamics | Newtonian forces on atoms and time dependency incorporated in conformational search |
| Genetic algorithm | Parent–child relationship along with survival of the fittest techniques employed |
| Expert system | Heuristic methods based on rules and facts employed |

states. The user needs to define step size, time of run, and the temperature supplied to the system at the beginning of the computational analysis. A simulated annealing method allows "cooling down" of the system at regular time intervals by decreasing the simulation temperature. As the temperature approaches 0 K, the molecule is trapped in the nearest local minimum. It is used as the starting point for further simulation and the cycle is repeated several times [19].

## 3 COMPUTATIONAL BIOLOGY

Computational biology is a fast growing topic and it is really not practical to distinguish this topic from bioinformatics. However, we may broadly distinguish between the two topics as far as this chapter is concerned. Molecular modeling aspects of computational biology, which lead to structure prediction, may be discussed under this heading, whereas the sequence analysis part, which leads to target identification, may be discussed under the section of pharmacoinformatics. Structure prediction of biomolecules (often referred to as "structural bioinformatics") adopts many aspects of computational chemistry. For example, energy minimization of protein receptor structure is one important step in computational biology. Molecular mechanics, molecular simulations, and molecular dynamics are employed in performing conformational analysis of macromolecules.

A rational drug design approach is very much dependent on the knowledge of receptor protein structures and is severely limited by the availability of target protein structure with experimentally determined 3D coordinates. Proteins exhibit four tiered organization: (1) primary structure defining the amino acid sequence, (2) secondary structure with α-helical and β-sheet folds, (3) tertiary structure defining the folding of secondary structure held by hydrogen bonds, and (4) quaternary structure involving noncovalent association between two or more independent proteins. Methods for identifying the primary amino acid sequence in proteins are now well developed; however, this knowledge is not sufficient enough to understand the function of the proteins, the drug–receptor mutual recognition, and designing drugs. Various experimental techniques like X-ray crystallography, nuclear magnetic resonance, and electron diffraction are available for determining the 3D coordinates of the protein structure; however, there are many limitations. It is not easy to crystallize proteins and even when we succeed, the crystal structure represents only a rigid state of the protein rather than a dynamic state. Thus, the reliability of the experimental data is not very high in biomolecules. Computational methods provide the alternative approach—although with equal uncertainty but at a greater speed. Homology modeling and *ab initio* methods are being employed to elucidate the tertiary structure of various biomolecules. The 3D structures of proteins are useful in performing molecular docking, *de novo* design, and receptor-based pharmacophore mappings. The computational methods of biomolecular structure prediction are discussed next (Fig. 3) [20, 21].

### 3.1 Ab Initio Structure Prediction

This approach seeks to predict the native conformation of a protein from the amino acid sequence alone. The predictions made are based on fundamental understanding of the protein structure and the predictions must satisfy the requirements of free-

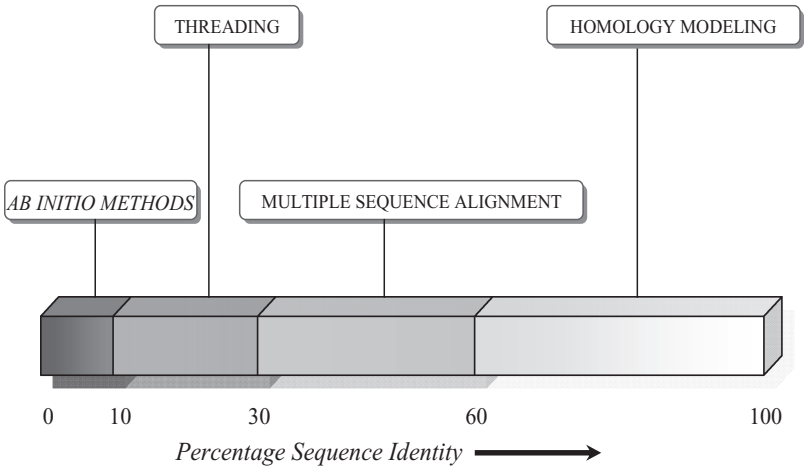


**FIGURE 3** A list of computer-aided structure prediction methods with respect to their suitability to the available sequence similarity.

energy function associated with lowest free-energy minima. The detailed representation of macromolecules should include the coordinates of all atoms of the protein and the surrounding solvent molecules. However, representing this large number of atoms and the interactions between them is computationally expensive. Thus, several simplifications have been suggested in the representations during the *ab initio* structure prediction process. These include (1) representation of side chains using a limited set of conformations that are found to be prevalent in structures from the Protein Data Bank (PDB) without any great loss in predictive ability [22] and (2) restriction of the conformations available to the polypeptides in terms of phi–psi ($\phi$–$\psi$) angle pairs [23]. Building the protein 3D structure is initiated by predicting the structures of protein fragments. Local structures of the protein fragments are generated first after considering several alternatives through energy minimization. A list of possible conformations is also extracted from experimental structures for all residues. Protein tertiary structures are assembled by searching through the combinations of these short fragments. During the assembling process, bump checking and low energy features (hydrophobic, van der Waals forces) should be incorporated. The final suggested structure is subjected to energy minimization and conformational analysis using molecular dynamics simulations. *Ab initio* structure prediction can be used to guide target selection by considering the fold of biological significance. The *ab initio* macromolecular structure prediction methods, if successful, are superior to the widely used homology modeling technique because no *a priori* bias is incorporated into the structure prediction [24].

### 3.2 Homology Modeling

Homology or comparative modeling uses experimentally determined 3D structure of a protein to predict the 3D structure of another protein that has a similar amino acid sequence. It is based on two major observations: (1) structure of a protein is uniquely determined by its amino acid sequence and (2) during evolution, the structure is more conserved than the sequence such that similar sequences adopt

practically identical structure and distantly related sequences show similarity in folds. Homology modeling is a multistep method involving the following steps: (1) obtaining the sequence of the protein with unknown 3D structure, (2) template identification for comparative analysis, (3) fold assignment based on the known chemistry and biology of the protein, (4) primary structure alignment, (5) backbone generation, (6) loop modeling, (7) side chain modeling, (8) model optimization, and (9) model validation.

The methodology adopted in homology modeling of proteins can be described as follows. The target sequence is first compared to all sequences reported in the PDB using sequence analysis. Once a template sequence is found in the data bank, an alignment is made to identify optimum correlation between template and target. If identical residues exist in both the sequences, the coordinates are copied as such. If the residues differ, then only the coordinates of the backbone elements (N, Cα, C, and O) are copied. Loop modeling involves shifting all insertions and deletions to the loops and further modifying them to build a considerably well resembling model. Modeling the side chains involves copying the conserved residues, which also includes substitution of certain rotamers that are strongly favored by the backbone. Model optimization is required because of the expected differences in the 3D structures of the target and the template. The energy minimizations can be performed using molecular mechanics force fields (either well defined and/or self-parameterizing force fields). Molecular dynamic simulations offer fast, more reliable 3D structure of the protein. Model validation is a very important step in homology modeling, because several solutions may be obtained and the scientific user should interfere and make a choice of the best generated model. Often, the user may have to repeat the process with increased caution [20, 24].

### 3.3 Threading or Remote Homology Modeling

Threading (more formally known as "fold recognition") is a method that may be used to suggest a general structure for a new protein. It is mainly adopted when pairwise sequence identity is less than 25% between the known and unknown structure. Threading technique is generally associated with the following steps: (1) identify the remote homology between the unknown and known structure; (2) align the target and template; and (3) tailor the homology model [24].

## 4 COMPUTATIONAL MEDICINAL CHEMISTRY

Representation of drug molecular structures can be handled using computational chemistry methods, whereas that of macromolecules can be handled using computational biology methods. However, finding the therapeutic potential of the chemical species and understanding the drug–receptor interactions *in silico* requires the following well developed techniques of computational medicinal chemistry.

### 4.1 Quantitative Structure–Activity Relationship (QSAR)

QSAR is a statistical approach that attempts to relate physical and chemical properties of molecules to their biological activities. This can be achieved by using easily

**TABLE 2 Different Dimensions in QSAR**

| |
|---|
| 1D QSAR: Affinity correlates with p$K_a$, log $P$, etc. |
| 2D QSAR: Affinity correlates with a structural pattern. |
| 3D QSAR: Affinity correlates with the three-dimensional structure. |
| 4D QSAR: Affinity correlates with multiple representations of ligand. |
| 5D QSAR: Affinity correlates with multiple representations of induced-fit scenarios. |
| 6D QSAR: Affinity correlates with multiple representations of solvation models. |

calculatable descriptors like molecular weight, number of rotatable bonds, and log $P$. Developments in physical organic chemistry over the years and contributions of Hammett and Taft in correlating the chemical activity to structure laid the basis for the development of the QSAR paradigm by Hansch and Fujita. Table 2 gives an overview of various QSAR approaches in practice. The 2D and 3D QSAR approaches are commonly used methods, but novel ideas are being implemented in terms of 4D–6D QSAR. The increased dimensionality does not add any additional accuracy to the QSAR approach; for example, no claim is valid which states that the correlation developed using 3D descriptors is better than that based on 2D descriptors.

***2D QSAR*** Initially, 2D QSAR or the Hansch approach was in vogue, in which different kinds of descriptors from the 2D structural representations of molecules were correlated to biological activity. The basic concept behind 2D QSAR is that structural changes that affect biological properties are electronic, steric, and hydrophobic in nature. These properties can be described in terms of Hammett substituent and reaction constants, Verloop sterimol parameters, and hydrophobic constants. These types of descriptors are simple to calculate and allow for a relatively fast analysis.

Most 2D QSAR methods are based on graph theoretical indices. The graph theoretical descriptors, also called the molecular topological descriptors, are derived from the topology of a molecule, that is, the 2D molecular structure represented as graphs. These topological connectivity indices representing the branching of a molecule were introduced by Randíc [25] and further developed by Kier and Hall [26, 27]. The graph theoretical descriptors include mainly the Kier–Hall molecular connectivity indices (chi) and the Weiner [28, 29], Hosoya [30], Zagreb [31], Balaban [32], kappa shape [33], and information content indices [32]. The electrotopological state index (E-state) [34] combines the information related to both the topological environment and the electronic character of each skeletal atom in a molecule. The constitutional descriptors are dependent on the constitution of a molecule and are numerical descriptors, which include the number of hydrogen bond donors and acceptors, rotatable bonds, chiral centers, and molecular weight (1D) [35]. Apart from that, several indicator descriptors, which define whether or not a particular indicator is associated with a given molecule, are also found to be important in QSAR. The quantum chemical descriptors include the molecular orbital energies (HOMO, LUMO), charges, superdelocalizabilities, atom–atom and molecular polarizabilities, dipole moments, total and binding energies, and heat of formation. These are 3D descriptors derived from the 3D structure of the molecule and are electronic in nature [36]. These parameters are also often clubbed with the 2D QSAR analysis.

Statistical data analysis methods for QSAR development are used to identify the correlation between molecular descriptors and biological activity. This correlation may be linear or nonlinear and accordingly the methods may be divided into linear and nonlinear approaches. The linear approaches include simple linear regression, multiple linear regression (MLR), partial least squares (PLS), and genetic algorithm–partial least squares (GA-PLS). Simple linear regression develops a single descriptor linear equation to define the biological activity of the molecule. MLR is a step ahead as it defines a multiple term linear equation. More than one term is correlated to the biological activity in a single equation. PLS, on the other hand, is a multivariate linear regression method that uses principal components instead of descriptors. Principal components are the variables found by principal component analysis (PCA), which summarize the information in the original descriptors. The aim of PLS is to find the direct correlation not between the descriptors and the biological activity but between the principal component and the activity. GA-PLS integrates genetic algorithms with the PLS approach. Genetic algorithms are an automatic descriptor selection method that incorporates the concepts of biological evolution within itself. An initial random selection of descriptors is made and correlated to the activity. This forms the first generation, which is then mutated to include new descriptors, and crossovers are performed between the equations to give the next generation. Equations with better predictability are retained and the others are discarded. This procedure is continually iterated until the desired predictability is obtained or the specified number of generations have been developed. The nonlinear approaches include an Artificial Neural Network (ANN) and machine learning techniques. Unlike the linear approaches, nonlinear approaches work on a black box principle; that is, they develop a relation between the descriptors and the activity to predict the activity, but do not give the information on how the correlation was made or which descriptors are more contributing. The ANN algorithm uses the concept of the functioning of the brain and consists of three layers. The first layer is the input layer where the structural descriptors are given as an input; second is the hidden layer, which may be comprised of more than one layer. The input is processed in this part to give the predicted values to the third output layer, which gives the result to the user. The user can control the input given and the number of neurons and hidden layers but cannot control the correlating method [37–40].

The QSAR model developed by any statistical method has to be validated to confirm that it represents the true structure–activity relationship and is not a chance correlation. This may be done by various methods such as the leave-one-out and leave-multiple-out cross-validations and the bootstrap method. The randomization test is another validation approach used to confirm the adequacy of the training set. Attaching chemical connotation to the developed statistical model is an important aspect. A successful QSAR model not only effectively predicts the activity of new species belonging to the same series but also should provide chemical clues for future improvement. This requirement, as well as the recognition that the 3D representation of the chemicals gives more detailed information, led to the development of 3D QSAR.

***3D QSAR*** 3D QSAR methods are an extension of the traditional 2D QSAR approach, wherein the physicochemical descriptors are estimated from the 3D struc-

tures of the chemicals. Typically, properties like molecular volume, molecular shape, HOMO and LUMO energies, and ionization potential are the properties that can be calculated from the knowledge of the 3D coordinates of each and every atom of the molecules. When these descriptors of series of molecules can be correlated to the observed biological activity, 3D QSAR models can be developed. This approach is different from the traditional QSAR only in terms of the descriptor definition and, in a sense, is not really 3D in nature.

Molecular fields (electrostatic and steric), which can be estimated using probe-based sampling of 3D structure of molecules within a molecular lattice, can be correlated with the reported numeric values of biological activity. Such methods proved to be much more informative as they provide differences in the fields as contour maps. The widely used CoMFA (comparative molecular field analysis) method is based on molecular field analysis and represents real 3D QSAR methods [41]. A similar approach was adopted in developing modules like CoMSIA (comparative molecular similarity index analysis) [42], SOMFA (self-organizing molecular field analysis) [43], and COMMA (comparative molecular moment analysis) [44]. Utilization and predictivity of CoMFA itself has improved sufficiently in accordance with the objectives to be achieved by it [45]. Despite the formal differences between the various methodologies, any QSAR method must include some identifiers of chemical structures, reliably measured biological activities, and molecular descriptors. In 3D QSAR, alignment (3D superimposition) of the molecules is necessary to construct good models. The main problems encountered in 3D QSAR are related to improper alignment of molecules, greater flexibility of the molecules, uncertainties about the bioactive conformation, and more than one binding mode of ligands. While considering the template, knowledge of the bioactive conformation of any lead compound would greatly help the 3D QSAR analysis. As discussed in Section 2.5, this may be obtained from the X-ray diffractions or conformation at the binding site, or from the global minimum structure. Alignment of 3D structures of molecules is carried out using RMS atoms alignment, moments alignment, or field alignment. The relationship between the biological activity and the structural parameters can be obtained by multiple linear regression or partial least squares analysis. Given next are some details of the widely used 3D QSAR approach CoMFA.

*CoMFA (Comparative Molecular Field Analysis)* DYLOMMS (dynamic lattice-oriented molecular modeling system) was one of the initial developments by Cramer and Milne to compare molecules by aligning in space and by mapping their molecular fields to a 3D grid. This approach when used with partial least squares based statistical analysis gave birth to the CoMFA approach [46]. The CoMFA methodology is a 3D QSAR technique that allows one to design and predict activities of molecules. The database of molecules with known properties is suitably aligned in 3D space according to various methodologies. After consistently aligning the molecules within a molecular lattice, a probe atom (typically carbon) samples the steric and electrostatic interactions of the molecule. Charges are then calculated for each molecule using any of the several methods proposed for partial charge estimation. These values are stored in a large spreadsheet within the module (SYBYL software) and are then accessed during the partial least squares (PLS) routine, which attempts to correlate these field energy terms with a property of interest by the use of PLS with cross-validation, giving a measure of the predictive power of the model.

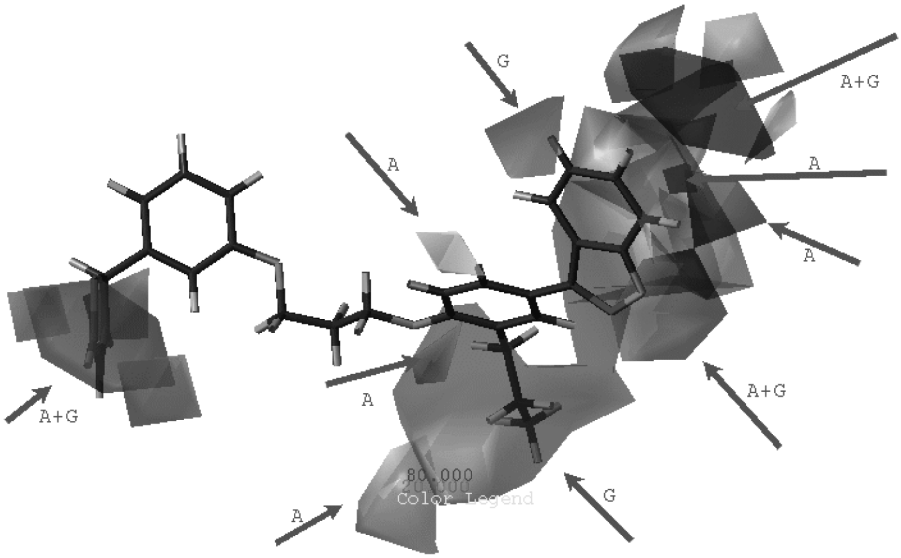


**FIGURE 4** Steric and electrostatic contour map for the dual model showing the contributions from each model. "A" depicts the contributions made by the α-model and "G" depicts the contributions made by the γ- and model. (Reproduced with permission from The American Chemical Society; S. Khanna, M. E. Sobhia, P. V. Bharatam *J Med Chem* 2005;48:3015.)

Electrostatic maps are generated, indicating red contours around regions where high electron density (negative charge) is expected to increase activity, and blue contours where low electron density (partial positive charge) is expected to increase activity. Steric maps indicate areas where steric bulk is predicted to increase (green) or decrease (yellow) activity [41, 45]. Figure 4 shows a typical contour map from CoMFA analysis. CoMSIA [42], CoMMA [44], GRID [47], molecular shape analysis (MSA) [48], comparative receptor surface analysis (CoRSA) [49], and Apex-3D [50] are other 3D QSAR methods that are being employed successfully.

***4D QSAR*** 4D QSAR analysis developed by Vedani and colleagues incorporates the conformational alignment and pharmacophore degrees of freedom in the development of 3D QSAR models. It is used to create and screen against 3D-pharmacophore QSAR models and can be used in receptor-independent or receptor-dependent modes. 4D QSAR can be used as a CoMFA preprocessor to provide conformations and alignments; or in combination with CoMFA to combine the field descriptors of CoMFA with the grid cell occupancy descriptors (GCODs) of 4D QSAR to build a "best" model; or in addition to CoMFA because it treats multiple alignments, conformations, and embedded pharmacophores, which are limitations of CoMFA [51].

***5D QSAR*** The 4D QSAR concept has been extended by an additional degree of freedom—the fifth dimension—allowing for multiple representations of the topology of the quasi-atomistic receptor surrogate. While this entity may be generated using up to six different induced-fit protocols, it has been demonstrated that the

simulated evolution converges to a single model and that 5D QSAR, due to the fact that model selection may vary throughout the entire simulation, yields less biased results than 4D QSAR, where only a single induced-fit model can be evaluated at a time (software Quasar) [52, 53].

***6D QSAR*** A recent extension of the Quasar concept to sixth dimension (6D QSAR) allows for the simultaneous consideration of different solvation models [54]. This can be achieved explicitly by mapping parts of the surface area with solvent properties (position and size are optimized by the genetic algorithms) or implicitly. In Quasar, the binding energy is calculated as

$$E_{\text{binding}} = E_{\text{ligand-receptor}} - E_{\text{desolvation,ligand}} - T\,\Delta S - E_{\text{internal strain}} - E_{\text{induced fit}} \tag{4}$$

## 4.2 Pharmacophore Mapping

A pharmacophore may be defined as the spatial arrangement of a set of key features present in a chemical species that interact favorably with the receptor leading to ligand–receptor binding and which is responsible for the observed therapeutic effect. It is the spatial arrangement of key chemical features that are recognized by a receptor and are thus responsible for biological response. Pharmacophore models are typically used when some active compounds have been identified but the 3D structure of the target protein or receptor is unknown. It is possible to derive pharmacophores in several ways, by analogy to a natural substrate, by inference from a series of dissimilar biologically active molecules (active analogue approach) or by direct analysis of the structure of known ligand and target protein.

A pharmacophoric map is a 3D description of a pharmacophore developed by specifying the nature of the key pharmacophoric features and the 3D distance map among all the key features. Figure 5 shows a pharmacophore map generated from the DISCO software module of SYBYL. A pharmacophore map may be generated from the superimposition of the active compounds to determine their common features. Given a set of active molecules, the mapping of a pharmacophore involves two steps: (1) analyzing the molecules to identify pharmacophoric features, and (2) aligning the active conformations of the molecules to find the best overlay of the corresponding features. Various pharmacophore mapping algorithms differ in the way they handle the conformational search, feature definition, tolerance definition, and feature alignment [55]. During pharmacophore mapping, generation and optimization of the molecules and the location of ligand points and site points (projections from ligand atoms to atoms in the macromolecule) are carried out. Typical ligand and site points are hydrogen bond donors, hydrogen bond acceptors, and hydrophobic regions such as centers of aromatic rings. A pharmacophore map identifies both the bioactive conformation of each active molecule and how to superimpose and compare, in three dimensions, the various active compounds. The mapping technique identifies what type of points match in what conformations of the compounds.

Besides ligand-based automated approaches, pharmacophore maps can also be generated manually. In such cases, common structural features are identified from a set of experimentally known active compounds. Conformational analysis is carried out to generate different conformations of the molecules and interfeature distances

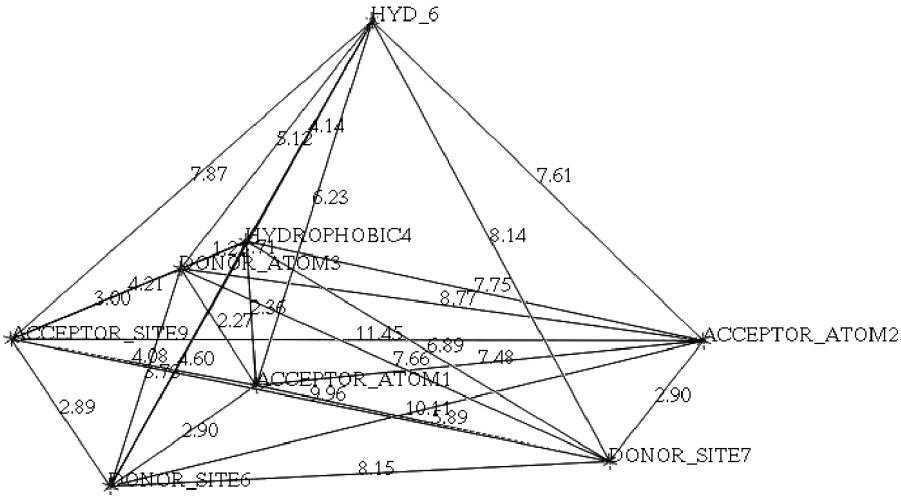


**FIGURE 5** A pharmacophore map developed from a set of GSK3 inhibitors. The pharmacophore features include hydrogen bond acceptor atoms, hydrogen bond donor atoms, hydrogen bond donor site, hydrogen bond acceptor site, and hydrophobic centers. This 3D picture also shows the distance relationship between various pharmacophoric features present in the map.

are inferred to develop the final models. The receptor mapping technique is also currently in practice to develop pharmacophore models. The important residues required for binding the pharmacophores are identified, which are employed for generating the receptor-based pharmacophores. The structure of protein can be used to generate interaction sites or grids to characterize favorable positions for ligands.

After a pharmacophore map has been derived, there are two ways to identify molecules that share its features and thus elicit the desired response. First is the *de novo* drug design, which seeks to link the disjoint parts of the pharmacophore together with fragments in order to generate hypothetical structures that are chemically novel. Second is the 3D database searching, where large databases comprising 3D structures are searched for those that match to a pharmacophoric pattern. One advantage of the second method is that it allows the ready identification of existing molecules, which are either easily available or have a known synthetic route [56, 57]. Pharmacophore mapping methods are described next.

***Distance Comparison Method (DISCO)*** The various steps involved in DISCO-based generation of a pharmacophore map are conformational analysis, calculation of the location of the ligand and site points, finding potential pharmacophore maps, and graphics analysis of the results. In the process of conformational search, 3D structures can be generated using any building program like CONCORD, from crystal structures, or from conformational searching and energy minimization with any molecular or quantum mechanical technique. Comparisons of all the duplicate conformations are excluded while comparing all the conformations. If each corresponding interatomic distance between these atoms in the two conformations is less

than a threshold (0.4 Å), then the higher energy conformation is rejected. DISCO calculates the location of site points, which can be the location of ligand atoms, or other atom-based points, like centers of rings or a halogen atom, which are points of potential hydrophobic groups. The other point is the location of the hydrogen bond acceptors or donors. The default locations of site hydrogen bond donor and acceptor points are based on literature compilations of observed intermolecular crystallographic contacts in proteins and between the small molecules. Hydrogen bond donors and acceptors such as OH and $NH_2$ groups can rotate to change the locations of the hydrogen atom.

During the process of performing pharmacophore mapping in DISCO, the user may input the tolerance for each type of interpoint distance. The user may direct the DISCO algorithm to consider all the potential points and to stop when a pharmacophore map with a certain total number of points is found. Alternatively, the user may specify the types of points, and the maximum and minimum number of each, that every superposition must include. It can also be directed to ignore specific compounds if they do not match a pharmacophore map found by DISCO. The user may also specify that only the input chirality is used for certain molecules and that only certain conformations below a certain relative energy should be considered.

The DISCO algorithm involves finding the reference molecule, which is the one with the fewest conformations. The search begins by associating the conformations of each molecule with each other. DISCO then calculates the distances between points in each 3D structure. Then it prepares the corresponding tables that relate interpoint distances in the current reference conformation and distances in every other 3D structure. Distances correspond if the point types are the same. These distances differ by no more than the tolerance limits. The clique-detection algorithm then identifies the largest clique of distances common between the reference XYZ set and every other 3D structure. It then forms union sets for the cliques of each molecule. Finally, the sets with cliques that meet the group conditions are searched [58, 59].

***CATALYST*** According to the pharmacophore mapping software CATALYST, a conformational model is an abstract representation of the accessible conformational space of a ligand. It is assumed that the biologically active conformation of a ligand (or a close approximation thereof) should be contained within this model. A pharmacophore model (in CATALYST called a hypothesis) consists of a collection of features necessary for the biological activity of the ligands arranged in 3D space, the common ones being hydrogen bond acceptor, hydrogen bond donor, and hydrophobic features. Hydrogen bond donors are defined as vectors from the donor atom of the ligand to the corresponding acceptor atom in the receptor. Hydrogen bond acceptors are analogously defined. Hydrophobic features are located at the centroids of hydrophobic atoms. CATALYST features are associated with position constraints that consist of the ideal location of a particular feature in 3D space surrounded by a spherical tolerance. In order to map the pharmacophore, it is not necessary for a ligand to possess all the appropriate functional groups capable of simultaneously residing within the respective tolerance spheres of the pharmacophoric features. However, the fewer features an inhibitor maps to, the poorer is its fit to them and the lower is its predicted affinity [60–63].

### 4.3 Molecular Docking

There are several possible conformations in which a ligand may bind to an active site, called the binding modes. Molecular docking involves a computational process of searching for a conformation of the ligand that is able to fit both geometrically and energetically into the binding site of a protein. Docking calculations are required to predict the binding mode of new hypothetical compounds. The docking procedure consists of three interrelated components—identification of the binding site, a search algorithm to effectively sample the search space (the set of possible ligand positions and conformations on the protein surface), and a scoring function. In most docking algorithms, the binding site must be predefined, so that the search space is limited to a comparatively small region of the protein. The search algorithm effectively samples the search space of the ligand–protein complex. The scoring function used by the docking algorithm gives a ranking to the set of final solutions generated by the search. The stable structures of a small molecule correspond to minima on the multidimensional energy surface, and different energy calculations are needed to identify the best candidate. Different forces that are involved in binding are electrostatic, electrodynamic, and steric forces and solvent related forces. The free energy of a particular conformation is equal to the solvated free energy at the minimum with a small entropy correction. All energy calculations are based on the assumption that the small molecule adopts a binding mode of lowest free energy within the binding site. The free energy of binding is the change in free energy that occurs upon binding and is given as

$$\Delta G_{\text{binding}} = G_{\text{complex}} - (G_{\text{protein}} + G_{\text{ligand}}) \tag{5}$$

where $G_{\text{complex}}$ is the energy of the complexed protein and ligand, $G_{\text{protein}}$ is the free energy of noninteracting separated protein, and $G_{\text{ligand}}$ is the free energy of noninteracting separated ligand.

The common search algorithms used for the conformational search, which provide a balance between the computational expense and the conformational search, include molecular dynamics, Monte Carlo methods, genetic algorithms, fragment-based methods, point complementary methods, distance geometry methods, tabu searches, and systematic searches [64].

Scoring functions are used to estimate the binding affinity of a molecule or an individual molecular fragment in a given position inside the receptor pocket. Three main classes of scoring functions are known, which include force field-based methods, empirical scoring functions, and knowledge-based scoring functions. The force field scoring functions use molecular mechanics force fields for estimating binding affinity. The AMBER and CHARMM nonbonded terms are used as scoring functions in several docking programs. In empirical scoring functions, the binding free energy of the noncovalent receptor–ligand complex is estimated using chemical interactions. These scoring functions usually contain individual terms for hydrogen bonds, ionic interactions, hydrophobic interactions, and binding entropy, as in the case of SCORE employed in DOCK4 and Böhm scoring functions (explained in detail in Section 4.4) used in FlexX. In empirical scoring functions, less frequent interactions are usually neglected. Knowledge-based scoring functions try to capture the knowledge about protein–ligand binding that is implicitly stored in the Protein Data Bank by means of statistical analysis of structural data, for example, PMF and DrugScore

functions, Wallqvist scoring function, and the Verkhivker scoring function [5, 37, 65–67]. Various molecular docking software packages are available, such as FlexX [68], Flexidock [58], DOCK [69], and AUTODOCK [70].

***FlexX*** FlexX is a fragment-based method for docking which handles the flexibility of the ligand by decomposing the ligand into fragments and performs the incremental construction procedure directly inside the protein active site. It allows conformational flexibility of the ligand while keeping the protein rigid. The base fragment or the ligand core is selected such that it has the most potential interaction groups and the fewest alternative conformations. It is placed into the active site and joined to the side chains in different conformations. Placements of the ligand are scored on the basis of protein–ligand interactions and ranked after the estimation of binding energy. The scoring function of FlexX is a modification of Böhm's function developed for the *de novo* design program LUDI. Figure 6 shows details of the interaction between a ligand and a receptor, obtained from FlexX molecular docking.

***DOCK*** DOCK is a simple minimization program that generates many possible orientations of a ligand within a user selective region of the receptor. DOCK is a program for locating feasible binding orientations, given the structures of a "ligand" molecule and a "receptor" molecule [69]. DOCK generates many orientations of one ligand and saves the best scoring orientation. The docking process is handled

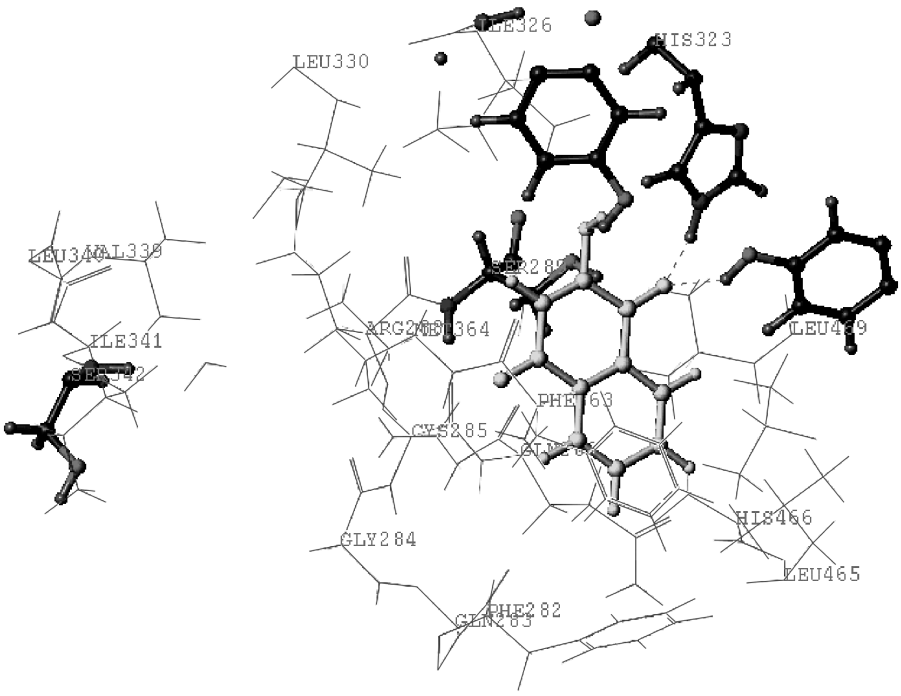


**FIGURE 6** The result of docking a ligand in the active site of PPARγ. The ligand has a hydrogen bonding interaction with histidine and tyrosine.

in four stages—ligand preparation, site characterization, scoring grid calculation, and finally docking. Site characterization is carried out by constructing site points, to map out the negative image of the active site, which are then used to construct orientations of the ligand. Scoring grid calculations are necessary to identify ligand orientations. The best scoring poses may be viewed using a molecular graphics program and the underlying chemistry may be analyzed.

There are many other widely used molecular docking software packages, like Flexidock (based on genetic algorithm), Autodock (based on Monte Carlo simulations and annealing), MCDOCK (Monte Carlo simulations), FlexE (ensemble of protein structures to account for protein flexibility), and DREAM++ (to dock combinatorial libraries).

### 4.4 De Novo Design

*De novo* design is a complementary approach to molecular docking: whereas in molecular docking already known ligands are employed, in *de novo* design, ligands are built inside the ligand binding domain. This is an iterative process in which the 3D structure of the receptor is used to build the putative ligand, fragment by fragment, within the receptor groove. Two basic types of algorithms are being widely used in *de novo* design. The first one is the "outside-in-method," in which the binding site is first analyzed to determine which specific functional groups might bind tightly. These separated fragments are then connected together with standard linker units to produce the ligands. The second approach is the "inside-out-method," where molecules are grown within the binding site so as to efficiently fit inside. *De novo* design is the only method of choice when the receptor structure is known but the lead molecules are not available. This method can also be used when lead molecules are known but new scaffolds are being sought. There are several programs developed by various researchers for constructing ligands *de novo*. GROW [71], GRID [72], CAVEAT [73], LUDI [74–77], LEAPFROG [58], GROUPBUILD [78], and SPROUT [79] are some of the *de novo* design programs that have found wide application.

***GRID*** The GRID program developed by Goodford [72] is an active site analysis method where the properties of the active site are analyzed by superimposition with a regular grid. Probe groups like water, methyl group, amine nitrogen, carboxyl oxygen, and hydroxyl are placed at the vertices of the grid and its interaction energy with the protein is calculated at each point using an empirical energy function that determines which kind of atoms and functional groups are best able to interact with the active site. The array of energy values is represented as a contour, which enables identification of regions of attractions between the probe and the protein. It is not a direct ligand generation method but positions simple fragments.

***LUDI*** LUDI, developed by Böhm [74–77], is one of the most widely used automated programs available for *de novo* design. It uses a knowledge-based approach based on rules about the energetically favorable interaction geometries of nonbonded contacts like hydrogen bonds and hydrophobic contacts between the functional groups of the protein and ligand. In LUDI the rules derived from statistical analysis of crystal packings of organic molecules are employed. LUDI is

fragment based and works in three steps. It starts by identifying the possible hydrogen bonding donors and acceptors and hydrophobic interactions, both aliphatic and aromatic, in the binding site represented as interaction site points. The site points are positions in the active site where the ligand could form a nonbonded contact. A set of interaction sites encompasses the range of preferred geometries for a ligand atom or functional group involved in the putative interaction, as observed in the crystal structure analyses. LUDI models the H-donor and H-acceptor interaction sites and the aliphatic or aromatic interaction sites. The interaction sites are defined by the distance $R$, angle $\alpha$, and dihedral angle $\omega$. The fragments from a 3D database of small molecules are then searched for positioning into suitable interaction sites such that hydrogen bonds can be formed and hydrophobic pockets filled with hydrophobic groups. The suitably oriented fragments are then connected together by spacer fragments to the respective link sites to form the entire molecule. Figure 7 shows LUDI generated fragment interaction sites inside the iNOS substrate binding domain.

An empirical but efficient scoring function is used for prioritizing the hit fragments given by LUDI. It estimates the free energy of binding ($\Delta G$) based on the

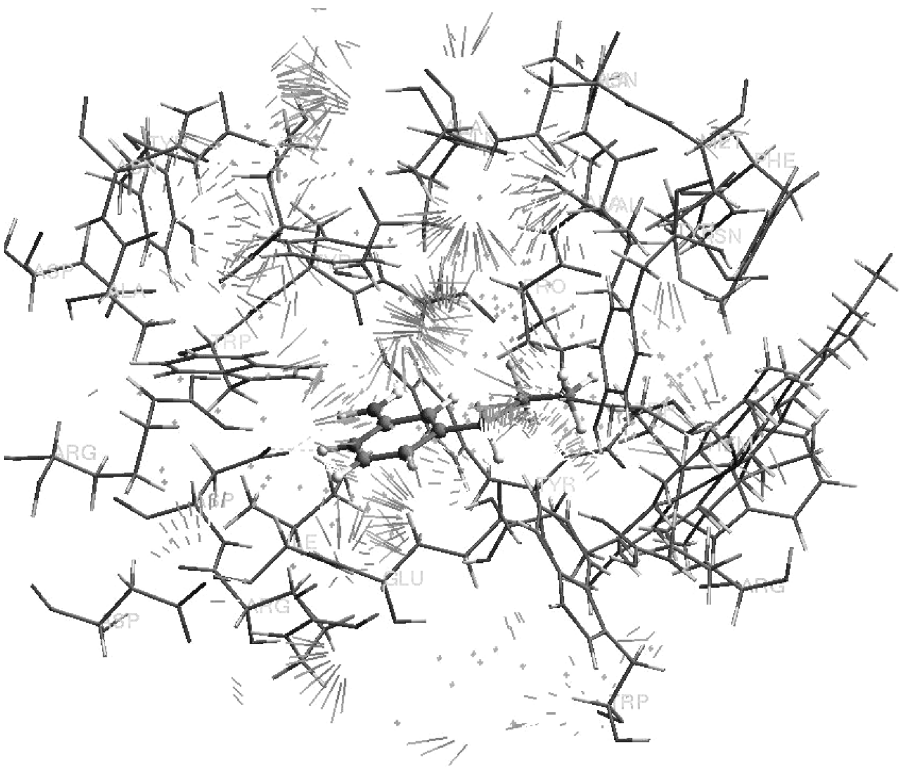

**FIGURE 7** An example of *de novo* design exercise. In the substrate binding domain of inducible nitric oxide synthase, the stick representation shows the protein structure; ball-and-stick representation belongs to the designed ligand, and the gray sticks point out the interaction sites.

hydrogen bonding, ionic interactions, hydrophobic contact areas, and number of rotatable bonds in the ligand. The LUDI scoring function is given as

$$\Delta G = \Delta G_{\mathrm{o}} + \Delta G_{\mathrm{hb}} \sum f(\Delta R) f(\Delta\alpha) + \Delta G_{\mathrm{ion}} \sum f(\Delta R) f(\Delta\alpha) + \Delta G_{\mathrm{lipo}} A_{\mathrm{lipo}} + \Delta G_{\mathrm{rot}} NR + \Delta G_{\mathrm{aro/aro}} \Delta N_{\mathrm{aro/aro}} \quad (6)$$

$\Delta G_{\mathrm{o}}$ represents the constant contribution to the binding energy due to loss of translational and rotational entropy of the fragment. $\Delta G_{\mathrm{hb}}$ and $\Delta G_{\mathrm{ion}}$ represent the contributions from an ideal neutral hydrogen bond and an ideal ionic interaction, respectively. The $\Delta G_{\mathrm{lipo}}$ term represents the contribution from lipophilic contact and the $\Delta G_{\mathrm{rot}}$ term represents the contribution due to the freezing of internal degrees of freedom in the fragment. $NR$ is the number of acyclic $sp^3$-$sp^3$ and $sp^3$-$sp^2$ bonds.

## 5 PHARMACOINFORMATICS

Information technology provides several databases, data analysis tools, and knowledge extraction techniques in almost every facet of life. In pharmaceutical sciences, several successful attempts are being made under the umbrella of pharmacoinformatics (synonymously referred to as pharmainformatics) (Fig. 8). The scope and limitations of this field are not yet understood. However, it may be broadly defined as the application of information technology in drug discovery and development. It encompasses all possible information technologies that eventually contribute to drug discovery. Chemoinformatics and bioinformatics contribute directly to drug discovery through virtual screening. Topics like neuroinformatics, immunoinformatics, vaccine informatics, and biosystem informatics contribute indirectly by providing necessary inputs for pharmaceutical design in this area. Topics like metabolomics, toxicoinformatics, and ADME informatics are contributing to this field by providing information regarding the fate of a NCE/lead *in vitro* and *in vivo* conditions. In this chapter some important aspects of these topics are presented. It is not easy to offer a comprehensive definition of this field at this stage owing to the fact that several bold attempts are being made in this field and initial signals related to a common platform are only emerging. *Drug Discovery Today* made initial efforts in this area by bringing out a supplement on this topic in which it was mainly treated as a scientific discipline with the integration of both bioinformatics and chemoinformatics [80, 81]. Recent trends in this area include several service-oriented themes including healthcare informatics [82], medicine informatics [83], and nursing informatics [84]. Here we present an overview of the current status.

### 5.1 Chemoinformatics

Chemoinformatics deals with information storage and retrieval of chemical data. This has been pioneered principally by the American Chemical Society and Cambridge Crystallographic Databank. However, the term chemoinformatics came into being only recently when methods of deriving science from the chemical databases was recognized. The integration of back-end technologies (for storing and representing chemical structure and chemical libraries) and front-end technologies (for assessing and analyzing the structures and data from the desktop) provides

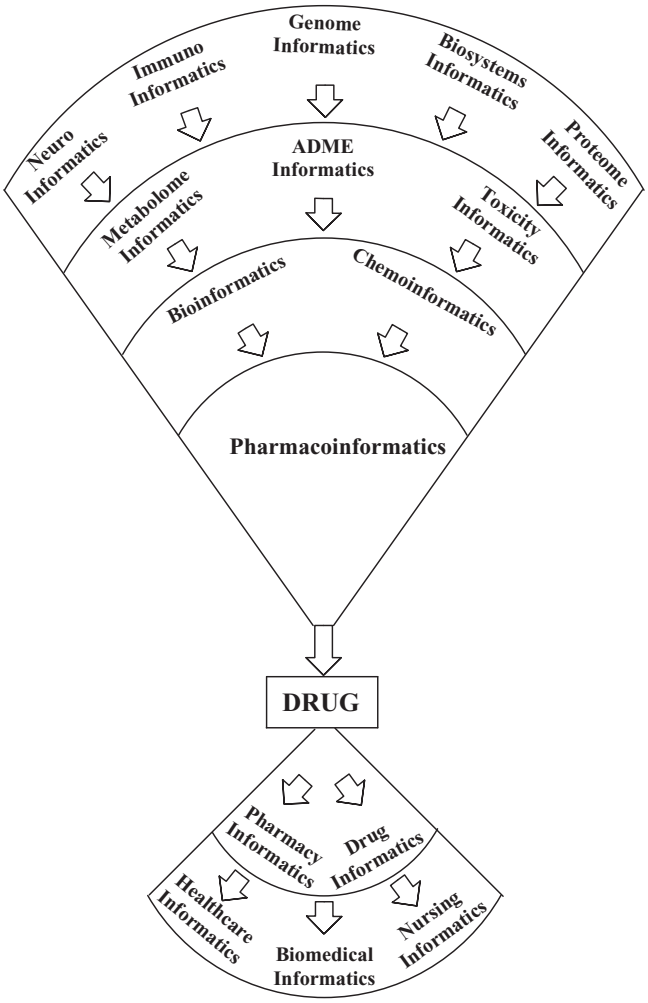


**FIGURE 8** Flowchart of informatics-based activities in pharmaceutical sciences.

opportunities in chemoinformatics. Virtual screening and high-throughput (HTS) data mining are some of the important chemoinformatics methods.

Chemical data is mostly considered as heuristic data, because it deals with names, properties, reactions, and so on. However, chemoinformatics experts devised many ways of representing chemical data (Fig. 9). (1) One-dimensional information of the chemicals can be represented in terms of IUPAC names, molecular formulas, Wiswesser line notation (WLN), SMILES notation, SYBYL line notation (SLN), and so on in addition to numerical representation of physicochemical parameters like molecular weight, molar refractivity, surface area, log $P$, and $pK_a$. (2) Two-dimensional chemical information consists of the chemical structural drawings, corresponding hashing, hash codes, connectivity tables, bond matrix, incidence matrix, adjacency matrix, bond-electron matrix, and so on. Graph theoretical procedures are extensively employed in this approach. To get a unique representation of

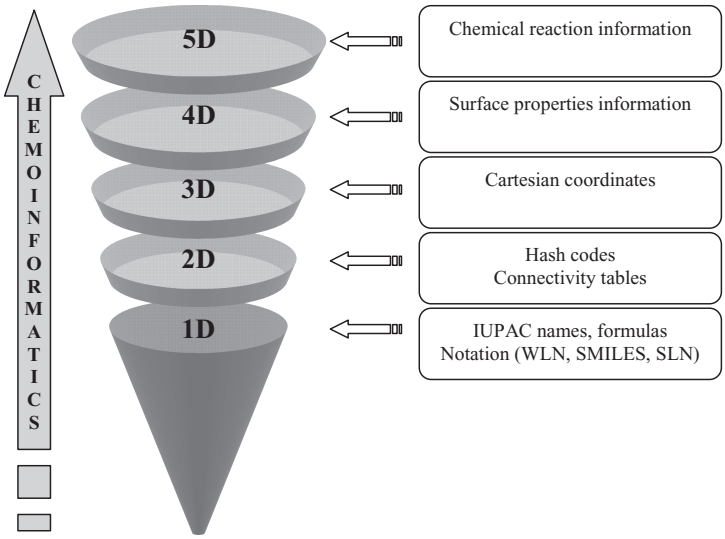


**FIGURE 9** Information flowchart in chemistry. In this context the dimensionality is not a geometrical dimensionality but an information complexity dimensionality.

chemicals, several algorithms (e.g., Morgan algorithm) were devised, which are found to be extremely important while performing chemical data mining. (3) Three-dimensional representation of structures involves the definition of the Cartesian coordinates of each and every element in the molecule. Along with connection tables, distance matrix, bond angle matrix, and torsional angle matrix, the representations of chemical 3D structures are being made in the form of formatted flat files like .mol and .sdf. (4) Fourth-dimensional information of chemicals is also available in the form of surface properties (e.g., Connolly surface, solvent accessible surface, electron density surfaces, molecular electrostatic potential (MESP), internal structure representation (molecular orbitals)). (5) Fifth-dimensional information about the chemical reactions is of course the most important information about chemical species. This involves defining chemical reaction parameters like inductive effect, resonance effect, polarizability effect, steric effect, and stereochemical effect, on the one hand, and chemical bond formation representations like Hendrickson scheme, Ugi scheme, and InfoChem scheme on the other. Some of the useful chemoinformatics databases include CAS, CSD, STN, MARPAT, UNITY-3D, MACCS-3D, CONCORD, MAYBRIDGE, NCI, and CASReact, which include chemical data in all the possible dimensions as discussed earlier [6, 10, 85–88].

Searching chemical data also requires distinctive methodologies. Search based on hash codes, graph theoretical indices (e.g., Weiner index), charge-related topological indices, Tanimoto coefficients, Carbo coefficient, Hamming distances, Euclidean distance, clique detection, and pharmacophore map searching are some of the important techniques uniquely required in chemoinformatics. Virtual generation of synthesizable chemicals is an important component of this technique. The principles of similarity and diversity need to be employed simultaneously while performing this exercise. Virtual screening is often considered under chemoinformatics, which is discussed in Section 5.3 as the most important topic of pharmacoinformatics.

Chemical reaction informatics involves the exploration of synthetic pathways and the designing of new experiments. About 15–20 million reactions are currently available in chemical reaction databases (CASReact, ChemReact, CrossFire Plus, etc.). The chemical reaction informatics would essentially assist the chemist in giving access to reaction information, deriving knowledge, predicting the course and outcome of chemical reactions, and designing syntheses. The main tasks include (1) storing information on chemical reactions, (2) retrieving the information, (3) comparing and analyzing sets of reactions, (4) defining the scope and limitations of a reaction type, (5) developing models of chemical reactivity, (6) predicting the course of the reactions, (7) analyzing reaction networks, and (8) developing methods for the design of syntheses. Chemical reaction databases consist of the following information: (1) reactants and products; (2) atom mapping, which allows you to determine which atom becomes which product atom through the reaction; (3) information regarding reacting center(s); (4) the catalyst used; (5) the atmosphere, including temperature, pressure, and composition; (6) the solvent used; (7) product yield; (8) optical purity, and (9) references to literature.

### 5.2 Bioinformatics

The area of bioinformatics encompasses various fields of molecular biology requiring data handling like genomics, proteomics, sequence analysis, and regulatory networks. Results of the Human Genome Project have triggered several activities in bioinformatics as a result of the complete sequencing of the human genome consisting of approximately 30,000 genes. The data generated in the biology laboratories is being stored in data banks like GenBank (United States), EMBL (Europe), DDBJ (Japan), and Swiss-Prot, as primary data sources. Secondary data banks like PROSITE, Profiles, and Pfam contain the fruits of analyses of the sequences in the primary databases (Table 3). These databases are available on the Web as well as on some specialized networks like EMBnet and NCBInet. Apart from the sequence data, data related to gene expression, gene products, and protein interactions are also available, whose management, analysis, and storage are the objectives of bioinformatics. Sequence analysis is the most important aspect of bioinformatics. There are many sequence analysis packages available like the Genetics Computing Group (GCG) package from Accelrys and the EMBOSS suite (European Molecular Biology Open Software Suite) from European Molecular Biology Laboratory (EMBL). Pairwise and multiple alignment algorithms were developed for determining the similarity between the sequences. Dotplot, which gives a dot matrix plot of any two sequences with respect to the amino acid comparison, is the simplest pairwise sequence analysis tool. Sequence identity, if present, becomes evident along diagonal areas in a dotplot. Dynamic programming methods like the Needleman–Wunsch algorithm (global alignment) and Smith–Waterman algorithm (local alignment) provide additional information after inserting necessary gaps. The sequence alignment program GAP from the GCG software and NEEDLE from EMBOSS are for global alignment, whereas BESTFIT from GCG and WATER from EMBOSS are for local alignment. Dynamic programming techniques provide optimal alignment between the amino acid sequences defined by the highest score. The number of matched pairs, mismatched pairs, and gaps can be used in estimating the score of a given pairwise comparison of two sequences. Percent Accepted Mutation matrix

**TABLE 3 Some Important Bioinformatics Resources**

| Category | Name | Description | Source |
|---|---|---|---|
| Sequence databases | GenBank | Genetic sequence database | National Center for Biotechnology Information http://www.ncbi.nlm.nih.gov |
| | EMBL | Nucleic acid and protein databases | European Molecular Biology Laboratory http://www.ebi.ac.uk/embl/index.html |
| | DDBJ | Nucleic acid database | http://www.ddbj.nig.ac.jp |
| | UniProt/Swiss-Prot | Protein database | http://www.uniprot.org |
| | PIR | Protein Information Resource | http://pir.georgetown.edu/pirwww/ |
| Genome databases | dbEST | Expressed Sequence Tags database | http://www.ncbi.nlm.nih.gov/dbEST/index.html |
| | GDB | Human Genome Database | http://www.gdb.org/ |
| | Ensembl | Genome database | http://www.ensembl.org/index.html |
| Secondary protein databases | Pfam | Protein family database with multiple sequence alignments and hidden Markov models | http://www.sanger.ac.uk/Software/Pfam/ |
| | Prodom | Protein family and domain database | http://protein.toulouse.inra.fr/prodom/current/html/home.php |
| | PROSITE | Protein family and domain database | http://us.expasy.org/prosite/details.html |
| Protein interaction databases | BIND | Biomolecular Interaction Network Database | http://www.bind.ca |
| | DIP | Database of Interacting Proteins | http://dip.doe-mbi.ucla.edu/ |
| | HPRD | Human Protein Reference Database | http://www.hprd.org/ |

250 (PAM250) or the BLOsum SUbstitution Matrix 62 (BLOSUM62) for protein sequences are the well accepted scoring matrices. The word *k*-tuple method is also a widely used sequence search tool, with heuristic algorithms like BLAST (Basic Local Alignment Search Tool) and FASTA. Multiple Sequence Alignment (MSA) methods like CLUSTALW and PILEUP have also been developed for alignment of three or more sequences. Highly conserved regions can be identified from such types of alignments, which is important for identifying members of the same protein family or for studying evolutionary relationships. Most of these tools are available

from the National Center for Biotechnology Information (NCBI) website (http://www.ncbi.nlm.nih.gov). An efficient search engine, ENTREZ, is also provided by NCBI for searching sequences or the related literature. The sequences are stored in various formats like GenBank, EMBL, Swiss-Prot, FASTA, and PIR. Some commercial software packages like the GCG Wisconsin package incorporate the databases and analysis tools, which can be customized for the purpose of end users [89].

Proteomics involves a detailed study of all the proteins present in a cell, their expression, post-translational modifications, interactions with drugs and proteins, and so on. Major technologies in the field of proteomics are mass spectrometry and gel electrophoresis. Proteome informatics deals with the application of informatic tools for proteome analysis. Several databases are available that contain information about the interactions between the proteins—for example, Database of Interacting Proteins (DIP) [90], the General Repository for Interaction Datasets (GRID) [91], the Biomolecular Interaction Network Database (BIND) [92], and the Human Protein Reference Database (HPRD) [93]. The general methods for extracting interaction information from the literature are natural language processing (NLP), naïve Bayes [94], decision trees [95], neural networks, nearest neighbor, and support vector machine (SVM). The prediction of protein functions can also be carried out by a comparative genome analysis. Domain fusion studies, chromosomal proximity studies, and phylogenetic profiling are some methods attempted in the postgenomic scenario to take the sequence information beyond, to the annotation of the proteome. In genome informatics, tools for genome comparison like PipMaker and Artemis Comparison Tool (ACT) and tools from NCBI like HomoloGene and LocusLink are included. Databases that integrate different computational methods to predict functional associations have also been developed, like STRING [96, 97] and POINT [98].

Target identification and target validation are the two major aspects of modern bioinformatics which are relevant to drug discovery. A relatively small number of targets (200–500) have been considered until now for the development of drugs. The traditional target identification process follows the path: assay observation → identification of the key protein → cloning of its gene. This process was bogged down by the dearth of information on several fronts, for example, structural data of proteins. With the recent advances in protein crystallography combined with the 2D- and 3D-NMR experiments, the available structural data improved drastically. This is further supplemented by the *ab initio* fold prediction from the primary sequence data. Thousands of structures can now be determined on an industrial scale. The trends in target identification made a paradigm shift in the recent past from a "deductive" approach involving assay-to-genes path to an "inductive" approach involving genes-to-assay path. The greatest challenge in bioinformatics is to facilitate this paradigm shift by providing automated tools for the identification of new genes as potential targets. Target validation involves the identification of the function of targets in disease states, such that "drugable" targets can be recognized. Experimental efforts in this context can be effectively complemented by computational methods by understanding drug–target interactions. For example, the identification of PPARγ as a target for insulin resistance could be carried out with the effective integration of the results from experimental, computational biology and bioinformatics methods. Initially, a model of *trans*-retinoid acid receptor (RXR) was

built on the basis of available crystal structures. Multiple sequence alignment was then carried out on two human retinoic acid receptors and three human PPAR subtypes. Secondary structure prediction and homology modeling were carried out to understand the structure and binding characteristics of PPARγ. Several computational techniques like pattern recognition, artificial neural networks, genetic algorithms, and alignment techniques are being employed in analyzing the data stored in databases. The advances in the area of bioinformatics should lead to an explosion in the number of available target molecules, and contribute tremendously to pharmacoinformatics.

### 5.3 Virtual Screening

Virtual screening is one of the most important technologies of pharmacoinformatics. It employs the databases and analysis tools discussed in the previous sections and pharmacophore mapping and molecular docking discussed under computational medicinal chemistry. This topic is considered under pharmacoinformatics because this technology is being heavily used for hit/lead identification through data search rather than studying molecular interactions and chemical principles. Virtual screening (also called *in silico* screening) provides a fast and cost effective tool for computationally screening compound databases in the search for novel drug leads. Various changes have occurred in the methods of drug discovery, the major ones taking place in the field of high throughput synthesis and screening techniques. The basic goal of virtual screening is the reduction of the huge chemical space of synthesized/virtual molecules and to screen them against a specific target protein virtually. Thus, the field of virtual screening has become an important component of drug discovery programs. Substantial efforts in this area have been made by large pharmaceutical companies. However, there are no well defined standards in virtual screening as yet [37].

***Virtual Libraries*** Two types of virtual libraries can be generated. One is produced by computational design, with the basic idea to design synthesizable compounds computationally. The other is a virtual library of compounds that are already synthesized. It is possible to extract millions of compounds from these sources and create databases. Virtual libraries are useful in the absence of knowledge about specific drug targets for virtual screening. Focused virtual libraries are important sometimes to save resources as the hit rate is higher in such cases [62].

***Various Approaches of Virtual Screening*** Virtual screening methods can be roughly divided into target structure-based and small molecule-based approaches, as shown in Table 4. When no structural information about the target protein is given, pharmacophore models can be used as filters for screening. These are mostly used when a set of active compounds is known that bind in a similar way to the protein. A pharmacophore capturing these features should be able to identify novel compounds with a similar pattern from the database. These pharmacophores may be generated manually or by automated software packages like CATALYST (HipHop, HypoGen) or DISCO, as described earlier. If the 3D structure of the receptor is known, a pharmacophore model can be derived based on the receptor active site. Usually all the pharmacophore searches are done in two steps: first the

**TABLE 4 Computational Methods and Tools for Virtual Screening**

| |
|---|
| Target structure-based approaches |
| Protein–ligand docking |
| Active site-directed pharmacophores |
| Molecule-based queries |
| 2D substructures |
| 3D pharmacophores |
| Complex molecular descriptors (e.g., electrotopological) |
| Volume- and surface-matching algorithms |
| Molecular fingerprints |
| Keyed 2D fingerprints (each bit position is associated with a specific chemical feature) |
| Hashed 2D fingerprints (properties are mapped to overlapping bit segments) |
| Multiple-point 3D pharmacophore fingerprints |
| Compound classification techniques |
| Cluster analysis |
| Cell-based partitioning (of compounds into subsections of *n*-dimensional descriptor space) |
| 3D/4D QSAR models |
| Statistical methods |
| Binary QSAR/QSPR |
| Recursive partitioning |

software confirms whether or not the screened compound has the required atom types or functional groups, and then it checks whether the spatial arrangement of these elements matches the query. In actual practice, receptor-based and small molecule-based approaches can be used in combination, taking into account as much information as possible.

Structure based virtual screening is carried out by docking and scoring techniques. Database of thousands of compounds can be screened against the specific target protein. FlexX [68], Flexidock [58], DOCK [52], and AUTODOCK [70] are the various docking programs employed. Molecular fingerprinting is another technique for performing virtual screening. These search tools consist of varying numbers of bits and encode different types of molecular descriptors and their values. Calculation of descriptors for a given molecule produces a characteristic bit pattern that can be quantitatively compared to others, applying a similarity metric [37].

Various filters can be used during the screening process which improves the chances of obtaining reliable hits. One such filter is the substructure filter. The database may contain certain functional groups in molecules that interact at unwanted sites and are "false positives." Many of these features are called as the substructures that are used to filter datasets. Other filters applied are the molecular weight, the number of rotatable bonds, and the calculated log $P$. All these considerations have led to the formulation of *Lipinski's rule of five* [99]. These rules suggest whether or not a molecule is going to be absorbed. The criteria for choosing better absorbed molecules are: (1) molecular weight <500, (2) log $P < 5$, (3) hydrogen bonds donors <5, and (4) hydrogen bond acceptors <10. The rule of five has been derived following a statistical analysis of known drugs. More sophisticated computational models of "drug-likeness" have been developed by employing the techniques of artificial neural networks, decision trees, and genetic algorithms [6].

It is important to realize that virtual screening is expected to produce hits, not leads. The identification of nanomolar inhibitors by database mining is an extremely rare probability. Lead optimization should be taken up to further modulate the structural features. A virtual screening exercise cannot distinguish the agonists; also, in several cases, the inactives are picked up. Hence it is better not to be very narrow toward the end of any virtual screening procedure [62, 63].

Virtual screening can be applied to target-based subset selection from the databases. Statistical approaches like binary QSAR or recursive partitioning can be applied to process HTS results and develop predictive models of biological activity. The developed models can then be employed to select candidate molecules from databases. Similarly, hits from HTS are used in fingerprint searches or compound classification analysis to identify sets of similar molecules. Based on these results, a few compounds are selected for additional testing. Many assumptions are made in virtual screening, and a positive outcome cannot be guaranteed every time. However, the overall process is extremely cost effective and fast. Virtual screening has the ability to produce leads that otherwise may not have been identified. Hence, virtual screening is emerging as a major component of pharmacoinformatics.

### 5.4 Neuroinformatics

Neuroinformatics may be defined as the organization and analysis of neuroscientific data using the tools of information technology. The information sources in neuroinformatics include behavioral sciences (psychological description) and medicinal (including drugs and diagnostic images) and biological (membranes, neurons, synapses, genes, etc.) aspects. The aim of neuroinformatics is to unravel the complex structure–function relationship of the brain in an integrative effort. Neuroscientists work at multiple levels and are producing enormous amounts of data. Distributed databases are being prepared and novel analytical tools are being generated with the help of information technology. Producing digital capabilities for web-based information management systems is one of the major objectives of neuroinformatics. Apart from data sharing, computational modeling of ion channels, neurons and neural networks, second messenger pathways, morphological features, and biochemical reaction are also often included in neuroinformatics. The initial ideas on neuroinformatics can be traced to the work of Hodgkin and Huxley, who initiated computational neuronal modeling. Current efforts in the direction include studies related to modeling the neuropsychological tests, neuroimaging, computational neuroscience, brain mapping, molecular neuroimaging, and magnetic resonance imaging.

***Important Databases in Neuroinformatics*** A probabilistic atlas and reference system for human brain is being developed as a neuroinformatics and neuroscience tool. Such a system is required because of the vast variance observed in the structure, function, and organization of human brain. It is a data source of digital images of human brain along with information on racial and ethnic conditions, education, handedness, personal traits, habits, and so on. It allows one to examine the relationship and distribution of the macro- and microscopic structure and function of human brain [100, 101]. Surface management system (SuMS) is another database that mainly deals with studies of the structure and function of cerebral cortex. All the data generated during reconstruction of cerebral cortex and subsequent flattening procedures is included in this database. SuMS (1) provides a systematic framework

for classification and storage, (2) serves as a version control system for the surface and volume datasets, (3) is an efficient data retrieval module, and (4) acts as a service request broker [102, 103]. Similarly, there are many other database systems with data analysis tools available in this field, a few of which are listed in Table 5. Several

**TABLE 5 Selected List of Important Neuroinformatics Tools and Databases**

| Databases | Brief Description | URL |
|---|---|---|
| Brain Architecture Management System (BAMS) | Repository of brain structure information; contains to date around 40,000 connections | http://brancusi.usc.edu/bkms/ |
| BrainMap | For meta-analysis of human functional brain-mapping literature | http://brainmap.org/ |
| BrainInfo | Information about the brain and its functions | http://braininfo.rprc.washington.edu/ |
| Brede Database | Neuroimaging data | http://hendrix.imm.dtu.dk/services/jerne/brede/ |
| Surface Management System (SuMS) | A surface-based database to aid cortical surface reconstruction, visualization and analysis | http://sumsdb.wustl.edu/sums/index.jsp |
| fMRIDC | Functional neuroimaging (fMRI) data (fMRI Data Center) | http://www.fmridc.org |
| LGICdb | Ligand Gated Ion Channel database | http://www.ebi.ac.uk/compneursrv/LGICdb/ |
| ModelDB | Neuronal and Network Models | http://senselab.med.yale.edu/senselab/ModelDB |
| CoCoMac | Collation of Connectivity data on the Macaque brain | http://cocomac.org/home.htm |
| L-Neuron | Computational Neuroanatomy Database | http://www.krasnow.gmu.edu/LNeuron |
| NeuroScholar | MySQL Database frontend with management of bibliography, histological and tracing data | http://www.neuroscholar.org |
| Catacomb | Components and Tools for Accessible Computer Modeling in Biology (Modeling Software for Neuroscience) | http://www.compneuro.org/catacomb/index.shtml |
| GENESIS | Neural Simulator | http://www.genesissim.org/GENESIS/ |
| Channel Lab | Single channel modeling program | http://www.synaptosoft.com/Channelab/index.html |
| NEURON | Simulation of individual neurons and networks of neurons | http://www.neuron.yale.edu |
| HHsim | Graphical Hodgkin–Huxley Simulator | http://www.cs.cmu.edu/~dst/HHsim/ |
| NEOSIM | Neural Open Simulation—for modeling of networks | http://www.neosim.org/ |
| NANS | Neuron and Network Simulator | http://vlsi.eecs.harvard.edu/research/nans.html |
| SNNAP | Simulator for Neural Networks and Action Potentials | http://snnap.uth.tmc.edu/ |

software tools have been made available over the past few years in the field of neuroinformatics. Neuroscholar [104] allows scientists to interact with information in the literature in a modular fashion. This tool permits the user to isolate fragments of data from a source and then bring them together to prepare a fact base for analysis and interpretation in a knowledge base environment. GENESIS (General Neural Simulation System) is a tool that helps in the simulation of neurosystems including subcellular components, biochemical reactions, single neurons, large neural networks, and system level models.

The Human Brain Project is one of the major initiatives under neuroinformatics. This is a broad based effort by neuroscientists and information scientists whose objective is to produce interoperable databases and analysis tools. This project included several tools for modeling simulation, information retrieval from multidisciplinary data, graphical interfaces, and integration of data analysis tools through electronic collaboration [105].

### 5.5 Immunoinformatics

Immunoinformatics is another major area in biomedical research where computational and informational technologies are playing a major role in the development of drugs and vaccines. This field is still in its infancy and it covers both modeling and informatics of the immune system and is the application of informatics technology to the study of immunological macromolecules, addressing important questions in immunobiology and vaccinology. Data sources for immunoinformatics include experimental approaches and theoretical models, both demanding validation at every stage. Major immunological developments include immunological databases, sequence analysis, structure modeling, modeling of the immune system, simulation of laboratory experiments, statistical support for immunological experimentation, and immunogenomics [106, 107]

There are many databases of relevance to immunologists, some of which are given in Table 6. IMGT, the international ImMunoGeneTics information system created in 1989, is one of the important ventures in immunoinformatics and is a knowledge resource comprising databases, tools, resources on immunoglobulins, T cell receptors, major histocompatibility complex (MHC), and related proteins of the immune system. IMGT includes sequence and genome databases with different interfaces, 3D structure database, web resources, and interactive tools for sequence and genome analysis. IMGT ONTOLOGY, which is a semantic specification of the terms to be used in immunogenetics and immunoinformatics, is available for IMGT users in the IMGT Scientific chart formalized in IMGT-ML (XML) schema [108–114]. The HIV Molecular Immunology Database is a database containing sequence and epitope maps of HIV-1 cytotoxic and helper T-cell epitopes.

Computer-aided vaccine design (CAVD) or computational vaccinology is another application of immunoinformatics involving prediction of immunogenicity. Immune interactions can be modeled using artificial neural networks (ANNs), hidden Markov models, molecular modeling, binding motifs, and quantitative matrices, of which ANN models have proved to be superior for the prediction of MHC-binding peptides [115–117]. Table 6 also gives a list of some of the tools for predicting whether or not a peptide would bind to a major histocompatibility complex. In addition to immunoinformatics, theoretical immunology is another related discipline and is the

**TABLE 6 Some Selected Immunoinformatics Databases and Tools**

| Databases and Tools | Brief Description | URL |
|---|---|---|
| IMGT, the international ImMunoGeneTics information system | A sequence, genome, and structure database for immunogenetics data | http://imgt.cines.fr |
| HIV Molecular Immunology Database | A database of HIV-specific B-cell and T-cell responses | http://www.hiv.lanl.gov/content/immunology/index.html |
| MHCBN | Comprehensive database of MHC-binding, nonbinding peptides and T-cell epitopes | http://bioinformatics.uams.edu/mirror/mhcbn/ |
| MHCPEP | Database of MHC-binding peptides | http://wehih.wehi.edu.au/mhcpep/ |
| ADABase (Mutation Registry for Adenosine Deaminase Deficiency) | Contains information on diseases and mutations associated with adenosine deaminase | http://bioinf.uta.fi/ADAbase/ |
| BCIPep | Database of B-cell epitopes | http://bioinformatics.uams.edu/mirror/bcipep/ |
| FIMM | Database of Functional Immunology | http://research.i2r.a-star.edu.sg/fimm/ |
| IPD (Immuno Polymorphism Database) | Database for the study of polymorphism in genes of the immune system | http://www.ebi.ac.uk/ipd/ |
| DHR (The Database of Hypersensitive Response) | Definition, source, and sequence of hypersensitive response (HR) proteins, etc. | http://sdbi.sdut.edu.cn/hrp/ |
| VBASE2 | Database of variable genes from the immunoglobulin loci | http://www.vbase2.org/ |
| BIMAS | Bioinformatics and Molecular Analysis Section (MHC peptide-binding prediction) | http://bimas.dcrt.nih.gov/molbio/hla_bind/ |
| SYFPEITHI | Database and prediction server of MHC ligands | http://www.syfpeithi.de/ |
| ProPred | MHC Class I and II Binding Peptide Server | http://www.imtech.res.in/raghava/propred/ http://www.imtech.res.in/raghava/propred1/ |
| nHLAPred | A neural network-based MHC Class I Binding Peptide Prediction Server | http://bioinformatics.uams.edu/mirror/nhlapred/ |
| NetMHC | Peptide-binding prediction using artificial neural networks (ANNs) and weight matrices | http://www.cbs.dtu.dk/services/NetMHC/ |
| MHCPred | Predict binding affinity for MHC I and II molecules | http://www.jenner.ac.uk/MHCPred/ |

application of mathematical modeling to diverse aspects of immunology ranging from T-cell selection in the thymus to the epidemiology of vaccination. The results from immunoinformatics are being heavily employed in defining another emerging science called Artificial Immune Systems (immunological computation or immunocomputing). The AIS computation is also of interest in modeling the immune system and solving immunological problems [118].

### 5.6 Drug Metabolism Informatics

An understanding of the pharmacokinetics of a drug can play a major role in reducing the probability of bringing a new chemical entity (NCE) with inappropriate ADME/Toxicity profile to the market. Drug metabolism and toxicity in the human body are primarily assessed during clinical trials, and preclinical assessment of the same involves study on *in vivo* and *in vitro* systems. *In silico* models for predicting pharmacokinetic properties based on the experimental results can greatly reduce the cost and time required for the experiments. These methods range from modeling approaches such as QSARs, to similarity searches as well as informatics methods like ligand–protein docking and pharmacophore modeling. Several ADME properties can be explained by simple molecular descriptors derived from the 2D chemical structure and can be used for the development of QSAR models. Such *in silico* prediction methods help chemists in judging whether or not a potential candidate may continue in the drug discovery pipeline. Metabolic biotransformation of any NCE may profoundly affect the bioavailability, activity, distribution, toxicity, and elimination of a compound; the effects of probable metabolism are now considered in the early stages of drug discovery with the help of computer-aided methods.

*In silico* prediction of metabolic biotransformation occurring at the liver cytochrome enzymes (CYP450 enzymes) are being studied [119]. Many databases and software systems are available in this field for the early prediction of substrates of CYP450 enzymes. Some of the databases and predictive systems for metabolic information of drugs are given in Table 7. The Human Drug Metabolism Database (hDMdb) project is a nonprofit, internet database of xenobiotic metabolic transformations that are observed in humans [120]. Other databases like MDL Metabolite contain xenobiotic transformation information that can also be linked to a toxicity database like MDL Toxicity database. Thus, it can give toxicity, if present, in the metabolite shown by the database. The predictive systems available for metabolism are mainly expert systems based on experimental data representing the metabolic effects (database) and/or rules derived from such data (rule-base). The rules may either be induced rules, which are quantitative, derived from a statistical analysis of the metabolic data, or knowledge-based rules derived from expert judgment. Some of the expert systems are MetabolExpert, METEOR, and META, as given in Table 7. These can also be linked to the corresponding toxicity prediction modules. METEOR covers both phase I and phase II biotransformation reactions and can analyze mass spectrometry data from metabolism studies. It can be linked to the DEREK software for toxicity prediction of the metabolites. Likewise, MetabolExpert can be linked to HazardExpert and META can be interfaced with MULTICASE, both of which are toxicity prediction modules. MetabolExpert is an open knowledge base, where the user can add his/her own rules. The META program operates from dictionaries of transformation operators, created by experts to

**TABLE 7 Databases and Tools for Metabolism Informatics**

| Databases and Tools | Brief Description | URL |
|---|---|---|
| Human Drug Metabolism Database (hDMdb) | IUPAC project for a web-based model database for human drug metabolism information | http://www.iupac.org/projects/2000/2000-010-1-700.html |
| MDL Metabolite | Comprised of a database, registration system, and browsing interface | http://www.mdl.com/products/predictive/metabolite/index.jsp |
| Accelrys' Metabolism Database | Biotransformations of organic molecules in a variety of species | http://www.accelrys.com/products/chem_databases/databases/metabolism.html |
| Biofrontier/P450 | Human cytochrome P450 information and predictive system | http://www.fqs.pl/ |
| Metabolism and Transport Drug Interaction database | Database developed by University of Washington | http://www.druginteractioninfo.org/ |
| MetabolExpert (CompuDrug, Inc.) | Predictive system for metabolic fate of a drug | http://www.compudrug.com/ |
| MEXAlert (CompuDrug, Inc.) | Rule-based prediction for first-pass metabolism | http://www.compudrug.com/ |
| META (Multicase, Inc.) | Uses dictionaries to create metabolic paths of query molecules | http://www.multicase.com/products/prod05.htm |
| METEOR (LHASA Ltd., Leeds, UK) | Predictions presented as metabolic trees | http://www.lhasalimited.org/ |

represent known metabolic paths, and is capable of predicting the sites of potential enzymatic attack and the metabolites formed [121].

### 5.7 Toxicoinformatics

Early prediction of toxicological parameters of new chemical entities (NCEs) is an important requirement in the drug discovery strategy today. This is being emphasized in the wake of many drug withdrawals in the recent past. Computational methods for predicting toxicophoric features is a cost effective approach toward saving experimental efforts and saving animal life. Current efforts in toxicoinformatics are mainly based on QSTR (quantitative structure–toxicity relationships) and rule-based mechanistic methods [122–125]. QSTR is a statistical approach, in which a correlation is developed between structural descriptors of a series of compounds and their toxicological data. In this approach, a model can be trained with the help of a set of known data, validated using many approaches, and then used for the prediction of toxicological parameters. The only limitation of this approach is that the predictive power of these models gets reduced when chemicals belonging to a class outside the series of molecules is used for the construction of the model. Toxicity prediction tools using this approach include TOPKAT and CASE/M-CASE.

TOPKAT mainly employs electrotopological descriptors based on graph theory for the development of QSTR models. TOPKAT uses linear free-energy relationships in statistical regression analysis of a series of compounds. In this software, the continuous/dichotomous toxicity end points are correlated to the structural features like electronic topological descriptors, shape descriptors, and substructure descriptors. CASE (Computer Automated Structure Evaluation) and M-CASE are toxicoinformatics software packages that have the capability to automatically generate predictive models. A hybrid QSTR artificial expert system-based methodology is adopted in CASE-based systems. A linear scale called "CASE units" is defined, which segregate the given set of molecules as active/inactive/marginally toxic species. Molecular fragments are classified in terms of biophores (fragments associated with activity) and biophobes (fragments associated with inactivity). One advantage of these packages is that they also include experimental data that is not released by the FDA, which increases the applicability of this set of packages.

Toxicity prediction tools based on "mechanistic approaches" are knowledge based-systems, where a fact base and rule base can be effectively analyzed to give qualitative information regarding the toxicity of chemical species. The expert rules included in the knowledge base are generally derived from the molecular mechanism of the drug action and hence they are known as "mechanistic approaches." Well known software packages in this category are DEREK (Deductive Estimation of Risk from Existing Knowledge), HazardExpert, and Oncologic. DEREK is a widely used toxicity prediction system. This program not only predicts the potential toxicity of a query chemical but also provides details of the logical process that leads to the predicted results. Table 8 gives a list of known resources in toxicoinformatics.

### 5.8 Cancer informatics

Application of information technology has been extended to specific subtopics of pharmaceutical sciences like cancer, diabetes, and AIDS. One important example, wherein information technology is being extensively used, is cancer informatics. The necessity of such focused subtopics of pharmacoinformatics was required because of the overflow of information and lack of integrated data formats. Major cancer informatics initiatives are being undertaken by the National Cancer Institute Center for Bioinformatics (NCICB), National Institutes of Health (United States), National Cancer Research Institute (NCRI) (United Kingdom), and the National Cancer Center (NCC) (Japan).

A web-based environment called CaCORE (Cancer Common Ontologic Reference Environment) was established, which helps in the management, redistribution, integration, and analysis of data arising from studies involving cell and molecular biology, genomics, histopathology, drug development, and clinical trials. The structural and functional components of CaCORE include (1) Enterprise Vocabulary Services (EVS), which provide terminology development, dictionary, and thesaurus services like the description-logics based NCI Thesaurus and the NCI Metathesaurus, which is a collection of biomedical vocabularies based on National Library of Medicine (NLM) and the Unified Medical Language System (UMLS); (2) The Cancer Data Standards Repository (CaDSR), which is a metadata registry for

**TABLE 8 Toxicoinformatics Tools**

| Tools | Description | URL | Predicted Endpoints |
|---|---|---|---|
| TOPKAT | QSAR | Accelrys http://www.accelrys.com/products/topkat | Mutagenicity, carcinogenicity, mammalian acute and chronic toxicities, developmental toxicities |
| M-CASE, CASE, ToxAlert, Casetox | Hybrid QSAR and expert system | Multicase Inc. http://www.multicase.com | Carcinogenicity, mutagenicity, teratogenicity, mammalian acute and chronic toxicities |
| DEREK | Knowledge based, structural rules | Lhasa Limited http://www.lhasalimited.org/ | Mutagenicity, carcinogenicity, teratogenicity/developmental toxicity, skin sensitization, acute toxicity, etc. |
| OncoLogic | Knowledge based | www.oncologic.net | Carcinogenicity |
| HazardExpert | Knowledge based | CompuDrug http://www.compudrug.com | Carcinogenicity, mutagenicity, teratogenicity, membrane irritation, neurotoxicity |
| COMPACT | QSAR; mechanistic supported by molecular modeling studies | Surrey | Cytochrome P450 metabolism |

common data elements that have been identified to simplify and standardize the data collection requirements and eligibility/exclusion criteria; and (3) the Cancer Bioinformatics Infrastructure Objects (CaBIO) module, which provides the data interface architecture [126–128].

Apart from the web-based facilities, several machine learning techniques like artificial neural networks and decision trees are also being utilized for cancer detection and diagnosis and more recently for prediction and prognosis. Several bioinformatics resources that are helpful in gene function prediction, in protein structure and function predictions, and in studies of protein–protein interactions are being employed in cancer informatics as well [129]. Several specific databases like PDQ, CGED, and FaCD (Table 9) are being heavily used by cancer informatics specialists. The NCI chemical databank is a source of all the chemicals tested for anticancer effects.

**TABLE 9 Cancer Informatics Resources**

| Databases | Brief Descvription | URL |
|---|---|---|
| PDQ (Physician Data Query) | NCI's Comprehensive Cancer Database | http://www.cancer.gov/cancertopics/pdq/cancerdatabase |
| CGED (Cancer Gene Expression Database) | Database of gene expression profile and accompanying clinical information | http://cged.hgc.jp/ |
| Mouse Retroviral Tagged Cancer Gene Database | Retroviral and transposon insertional mutagenesis in mouse tumors | http://rtcgd.ncifcrf.gov |
| The Familial Cancer Database (FaCD) | Assists in the differential diagnosis in familial cancer | http://facd.uicc.org/ |
| International Agency for Research on Cancer (IARC) p53 Mutation database | Information on TP53 gene mutations | http://www.iarc.fr/p53/ |
| The Tumor Gene Database | Information on genes that are targets for cancer-causing mutations | http://condor.bcm.tmc.edu/oncogene.html |
| SNP500Cancer Database | Central Resource for Sequence verification of SNPs | http://snp500cancer.nci.nih.gov |

## 6 FUTURE SCOPE

The field of computer-aided drug development has undergone a paradigm shift in the past five years. Earlier, this subject was dominated by chemistry and physics of drug discovery, mainly through molecular modeling, QSAR. A dramatic increase has been noted in these two topics—several practical solutions are being offered and novel concepts are being introduced. At the same time an additional component is emerging in CADD through informatics. The current status of CADD includes both modeling and informatics with equal and synergistic contributions. Although a lot has been done in this area over the past twenty years, there is still a lot of scope left for future growth. First and foremost is winning the confidence of the experimental colleague, who is still skeptical about the value of these efforts. Second is the proper integration of modeling efforts and informatics efforts. Although virtual screening has proved to be highly successful, the methodologies being adopted lack common elements. Often the methods adopted in CADD are considered as technological components, although several fundamentals exist. In fact, the fundamentals of the field of drug discovery are only emerging. CADD methods should strongly contribute in establishing these fundamentals; efforts should be concentrated in this direction.

## REFERENCES

1. http://www.cambridgesoft.com.
2. http://www.acdlabs.com/download/chemsk.html.

3. http://www.chemsw.com.
4. Goodman JM. *Chemical Applications of Molecular Modeling*. Cambridge, UK: Royal Society of Chemistry; 1998.
5. Holtje HD, Sippl W, Rognan D. *Molecular Modeling Basic Principles and Applications*, 2nd ed. Weinheim: Wiley-VCH Verlag GmbH; 2003.
6. Leach AR, Gillet VJ. *An Introduction to Chemoinformatics*. Dordrecht: Kluwer Academics; 2003.
7. Levine IN. *Quantum Chemistry*, 4th ed. Englewood Cliffs, NJ: Prentice-Hall; 1991.
8. Szabo A, Ostlund S. *Modern Quantum Chemistry*. New York: MacMillan; 1982.
9. Foresman JB, Frisch AE. *Exploring Chemistry with Electronic Structure Methods*, 2nd ed. Pittsburgh: Gaussian Inc.; 1995.
10. Gasteiger J, Engel T. *Chemoinformatics: A Textbook*. Weinheim: Wiley-VCH Verlag GmbH; 2003.
11. Howard AE, Kollman PA. An analysis of current methodologies for conformational searching of complex molecules. *J Med Chem* 1998;31:1669–1675.
12. Lipton M, Still WC. The multiple minimum problem in molecular modeling. Tree searching internal coordinate conformational space. *J Comput Chem* 1988;9:343–355.
13. Dammkoehler RA, Karasek SF, Shands EF, Marshall GR. Constrained search of conformational hyperspace. *J Comput Aided Mol Des* 1989;3:3–21.
14. Saunders M. Stochastic exploration of molecular mechanics energy surfaces. Hunting for the global minimum. *J Am Chem Soc* 1987;109:3150–3152.
15. Saunders M. Stochastic search for the conformations of bicyclic hydrocarbons. *J Comput Chem* 1989;10:203–208.
16. Ferguson DM, Raber DJ. A new approach to probing conformational space with molecular mechanics: random incremental pulse search. *J Am Chem Soc* 1989;111:4371–4378.
17. Chang G, Guida WC, Still WC. An internal-coordinate Monte Carlo method for searching conformational space. *J Am Chem Soc* 1989;111:4379–4386.
18. Saunders M, Houk KN, Wu YD, Still WC, Lipton M, Chang G, Guida, WC. Conformations of cycloheptadecane. A comparison of methods for conformational searching. *J Am Chem Soc* 1990;112:1419–1427.
19. Lybrand TP, Computer simulation of biomolecular systems using molecular dynamics and free energy perturbation methods, in *Reviews in Computational Chemistry*. Hoboken, NJ: Wiley-VCH; 1990, pp 295–320.
20. Bourne PE, Weissig H. *Structural Bioinformatics (Methods of Biochemical Analysis)*. Hoboken, NJ: Wiley-Liss; 2003.
21. Clote P, Backofen R. Computational *Molecular Biology: An Introduction*. Hoboken, NJ: Wiley; 2000.
22. Dunbrack RL Jr, Karplus M. Conformational analysis of the backbone-dependent rotamer preferences of protein sidechains. *Nat Struct Biol* 1994;1:334–340.
23. Park BH, Levitt M. The complexity and accuracy of discrete state models of protein structure. *J Mol Biol* 1995;249:493–507.
24. Sternberg MJE. *Protein Structure Prediction: A Practical Approach (The Practical Approach Series)*. New York: Oxford University Press; 1997.
25. Randíc M. Characterization of molecular branching. *J Am Chem Soc* 1975;97: 6609–6615.
26. Kier LB, Hall LH. *Molecular Connectivity in Chemistry and Drug Research*. New York: Academic Press; 1976.

27. Kier LB, Hall LH. *Molecular Connectivity in Structure–Activity Analysis*. Hoboken, NJ: Wiley; 1986.
28. Müller WR, Szymanski K, Knop JV, Trinajstic N. An algorithm for construction of the molecular distance matrix. *J Comput Chem* 1987;8:170–173.
29. Wiener H. Structural determination of paraffin boiling points. *J Am Chem Soc* 1947;69:17–20.
30. Hosoya H. Topological index. A newly proposed quantity characterizing the topological nature of structural isomers of saturated hydrocarbons. *Bull Chem Soc Jpn* 1971;44:2332–2339.
31. Bonchev D. *Information Theoretic Indices for Characterization of Chemical Structures*. Hoboken, NJ: Wiley; 1983.
32. Balaban AT. Highly discriminating distance-based topological index. *Chem Phys Lett* 1982;89:399–404.
33. Kier LB. A shape index from molecular graphs. *Quant Struct-Activity Relat* 1985;4: 109–116.
34. Kier LB, Hall LH, Frazer JW. An index of electrotopological state for atoms in molecules. *J Math Chem* 1991;7:229–241.
35. Todeschini R, Consonni V. *Handbook of Molecular Descriptors*. Weinheim: Wiley-VCH Verlag GmbH; 2000.
36. Karelson M, Lobanov VS, Katritzky AR. Quantum-chemical descriptors in QSAR/QSPR studies. *Chem Rev* 1996;96:1027–1044.
37. Abraham DJ. *Burger's Medicinal Chemistry & Drug Discovery*, 6th ed. Hoboken, NJ: Wiley-Interscience; 2003.
38. Hansch C, Leo A, Heller SR. *Exploring QSAR: Volume 1: Fundamentals and Applications in Chemistry and Biology*. Washington DC: American Chemical Society; 1995.
39. Kubinyi H, QSAR: Hansch analysis and related approaches. In *Methods and Principles in Medicinal Chemistry*. Weinheim: VCH; 1993.
40. Kubinyi H, *QSAR In Drug Design. Theory, Methods and Applications*. Leiden: ESCOM; 1993.
41. Cramer RDI, Patterson DE, Bunce JD. Comparative molecular field analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins. *J Am Chem Soc* 1988;110:5959–5967.
42. Klebe G, Abraham U, Meitzner T. Molecular similarity indices in a comparative analysis (CoMSIA) of drug molecules to correlate and predict their biological activity. *J Med Chem* 1994;37:4130–4146.
43. Robinson DD, Winn PJ, Lyne PD, Richards WG. Self-organizing molecular field analysis: a tool for structure–activity studies. *J Med Chem* 1999;42:573–583.
44. Silverman BD, Platt DE. Comparative molecular moment analysis (CoMMA): 3D-QSAR without molecular superposition. *J Med Chem* 1996;39:2129–2140.
45. Kubinyi H. Comparative molecular field analysis (CoMFA), in *The Encyclopedia of Computational Chemistry*. Chichester: Wiley; 1998, pp 448–460.
46. Wise M, Cramer RD, Smith D, Exman I. Progress in three-dimensional drug design: the use of real-time colour graphics and computer postulation of bioactive molecules in DYLOMMS. In *Quantitative Approaches to Drug Design*. Amsterdam: Elsevier; 1983, pp 145–146.
47. Pastor M, Cruciani G, Watson KAA. Strategy for the incorporation of water molecules present in a ligand binding site into a three-dimensional quantitative structure–activity relationship analysis. *J Med Chem* 1997;40:4089–4102.

48. Polanski J, Walczak B. The comparative molecular surface analysis (COMSA): a novel tool for molecular design. *Comput Chem* 2000;24:615–625.
49. Ivanciuc O, Ivanciuc T, Cabrol-Bass D. Comparative receptor surface analysis (CoRSA) model for calcium channel antagonists. *SAR QSAR Environ Res* 2002;12:93–111.
50. Hariprasad V, Kulkarni VM. A proposed common spatial pharmacophore and the corresponding active conformations of some peptide leukotriene receptor antagonists. *J Comput Aided Mol Des* 1996;10:284–292.
51. Vedani A, Briem H, Dobler M, Dollinger H, McMasters DR. Multiple-conformation and protonation-state representation in 4D-QSAR: the neurokinin-1 receptor system. *J Med Chem* 2000;43:4416–4427.
52. Vedani A, Dobler M, Dollinger H, Hasselbach KM, Birke F, Lill MA. Novel ligands for the chemokine receptor-3 (CCR3): a receptor-modeling study based on 5D-QSAR. *J Med Chem* 2005;48:1515–1527.
53. Vedani A, Dobler M. 5D-QSAR: the key for simulating induced fit? *J Med Chem* 2002;45:2139–2149.
54. Vedani A, Dobler M, Lill MA. Combining protein modeling and 6D-QSAR. Simulating the binding of structurally diverse ligands to the estrogen receptor. *J Med Chem* 2005;48:3700–3703.
55. Martin YC, Bures M, Dahaner, E. A fast approach to pharmacophore mapping and its applications to dopaminergic and benzodiazepine agonists. *J Comput Aided Mol Des* 1993;7:83–102.
56. Marriott DP, Dougall IB, Meghani P, Liu Y-J, Flower DR. Lead generation using pharmacophore mapping and three-dimensional database searching: application to muscarinic M(3) receptor antagonists. *J Med Chem* 1999;42:3210–3216.
57. Sutter J, Güner OF, Hoffmann R, Li H, Waldman M. *Pharmacophore, Perception Development and Use In Drug Design*. La Jolla, CA: International University Line, 2000, pp 504–506.
58. *SYBYL7.0*, Tripos Inc, 1699 South Hanley Rd, St Louis, MO 631444, USA. http://www.tripos.com.
59. Gardiner EJ, Artymiuk PJ, Willett P. Clique-detection algorithms for matching three-dimensional molecular structures. *J Mol Graph Model* 1997;15:245–253.
60. Sprague PW. Automated chemical hypothesis generation and database searching with CATALYST. In *Perspectives in Drug Discovery and Design*. Leiden: ESCOM Science; 1995, pp 1–20.
61. *CATALYST4.10*, Biosyn-MSI, San Diego, CA, USA, 2005. http://www.accelrys.com.
62. Bajorath J. Virtual screening in drug discovery: methods, expectations and reality. *Curr Drug Discov* 2002, pp 24–28.
63. Klebe G. *Virtual Screening: An Alternative or Complement to High Throughput Screening?* Berlin: Springer; 2000.
64. Leach AR. *Molecular Modelling: Principles and Applications*, 2nd ed. Boston: Addison Wesley Longman, Harlow; 1996.
65. Cohen N. *Guidebook on Molecular Modeling in Drug Design*. San Drego, CA: Academic Press; 1996.
66. Charifson PS. *Practical Application of Computer-Aided Drug Design*. New York: Marcel Dekker; 1997.
67. Schneider G, Bohm H-J. Virtual screening and fast automated docking methods. *Drug Discov Today* 2002;7:64–70.
68. Rarey M, Kramer B, Lengauer T, Klebe G. A fast flexible docking methods using an incremental construction algorithm. *J Mol Biol* 1996;261:470–489.

69. Kuntz ID, Blaney JM, Oatley SJ, Langridge R, Ferrin TE. A geometric approach to macromolecule–ligand interactions. *J Mol Biol* 1982;161:269–288.
70. Morris GM, Goodsell DS, Huey R, Olson AJ. Distributed automated docking of flexible ligands to proteins: parallel applications of AutoDock2.4. *J Comput Aided Mol Des* 1996;10:293–304.
71. Moon JB, Howe WJ. Computer design of bioactive molecules: a method for receptor-based *de novo* ligand design. *Proteins* 1991;11:314–328.
72. Goodford PJ. A computational procedure for determining energetically favorable binding sites on biologically important macromolecules. *J Med Chem* 1985;28:849–857.
73. Lauri G, Bartlett PA. CAVEAT: a program to facilitate the design of organic molecules. *J Comput Aided Mol Des* 1994;8:51–66.
74. Böhm HJ. LUDI: rule-based automatic design of new substituents for enzyme inhibitor leads. *J Comput Aided Mol Des* 1992;6:593–606.
75. Böhm HJ. The computer program LUDI: a new method for the *de novo* design of enzyme inhibitors. *J Comput Aided Mol Des* 1992;6:61–78.
76. Böhm HJ. A novel computational tool for automated structure-based drug design. *J Mol Recognit* 1993;6:131–137.
77. Böhm HJ. The development of a simple empirical scoring function to estimate the binding constant for a protein–ligand complex of known three-dimensional structure. *J Comput Aided Mol Des* 1994;8:243–256.
78. Rotstein SH, Murcko MA. GroupBuild: a fragment-based method for *de novo* drug design. *J Med Chem* 1993;36:1700–1710.
79. Gillet V, Johnson AP, Mata P, Sike S, Williams P. SPROUT: a program for structure generation. *J Comput Aided Mol Des* 1993;7:127–153.
80. Gund P, Sigal N. Applying information systems to high-throughput screening and analysis. In *Pharmainformatics: A Trend Guide. Drug Discor Today 1999; (Suppl)*, 25–29.
81. Bharatam PV. Pharmacoinformatics: IT solutions for drug discovery and development, *CRIPS (Current Research and Information on Pharmaceutical Sciences)*, 2003;4:2–5.
82. Hanson CW. *Healthcare Informatics*. New York: McGraw-Hill Professional; 2005.
83. Goodman KW. *Ethics, Computing, and Medicine: Informatics and the Transformation of Health Care*, 1st ed. Cambridge, UK: Cambridge University Press; 1998.
84. Saba VK, McCormick KA. *Essentials of Nursing Informatics*, 4th ed. New York: McGraw-Hill Medical; 2005.
85. Gadre SR, Shirsat RN. *Electrostatics of Atoms and Molecules*. Hyderabad: Universities Press (India) Limited; 2000.
86. Oprea TI. Chemoinformatics in drug discovery. In *Methods and Principles in Medicinal Chemistry*, Vol 23. Weinheim: Wiley-VCH Verlag GmbH; 2005.
87. Gasteiger J. Handbook on chemoinformatics: from data to knowledge, In *Representation of Molecular Structures*. Hoboken, NJ: Wiley; 2003.
88. Bajorath J. Chemoinformatics: concepts, methods, and tools for drug discovery. In *Methods in Molecular Biology*. Totowa, NJ: Humana Press; 2004.
89. Mount DW. *Bioinformatics: Sequence and Genome Analysis*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press; 2001.
90. Xenarios I, Salwinski L, Duan XJ, Higney P, Kim SM, Eisenberg D. DIP, the Database of Interacting Proteins: a research tool for studying cellular networks of protein interactions. *Nucleic Acids Res* 2002;30:303–305.
91. Breitkreutz BJ, Stark C, Tyers M. The GRID: the General Repository for Interaction Datasets. *Genome Biol* 2003;4:R23.

92. Bader GD, Betel D, Hogue CW. BIND: the Biomolecular Interaction Network Database. *Nucleic Acids Res* 2003;31:248–250.
93. Peri S, Navarro JD, Kristiansen TZ, Amanchy R, Surendranath V, Muthusamy B, Gandhi TK, Chandrika KN, Deshpande N, Suresh S, Rashmi BP, Shanker K, Padma N, Niranjan V, Harsha HC, Talreja N, Vrushabendra BM, Ramya MA, Yatish AJ, Joy M, Shivashankar HN, Kavitha MP, Menezes M, Choudhury DR, Ghosh N, Saravana R, Chandran S, Mohan S, Jonnalagadda CK, Prasad CK, Kumar-Sinha C, Deshpande KS, Pandey A. Human protein reference database as a discovery resource for proteomics. *Nucleic Acids Res* 2004;32:D497–501.
94. Marcotte EM, Xenarios I, Eisenberg D. Mining literature for protein–protein interactions. *Bioinformatics* 2001;17:359–363.
95. Wilcox A, Hripcsak G. Classification algorithms applied to narrative reports. Proc AMIA Symp 1999;455–459.
96. *STRING*, http://string.embl.de/.
97. von Mering C, Huynen M, Jaeggi D, Schmidt S, Bork P, Snel B. STRING: a database of predicted functional associations between proteins. *Nucleic Acids Res* 2003;31:258–261.
98. Tien AC, Lin MH, Su LJ, Hong YR, Cheng TS, Lee YC, Lin WJ, Still IH, Huang CY. Identification of the substrates and interaction proteins of aurora kinases from a protein–protein interaction model. *Mol Cell Proteomics* 2004;3:93–104.
99. Lipinski CA, Lombardo F, Dominy BW, Feeney PJ. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv Drug Deliv Rev* 1997;23:3–25.
100. Mazziotta J, Toga A, Evans A, Fox P, Lancaster J, Zilles K, Woods R, Paus T, Simpson G, Pike B, Holmes C, Collins L, Thompson P, MacDonald D, Iacoboni M, Schormann T, Amunts K, Palomero-Gallagher N, Geyer S, Parsons L, Narr K, Kabani N, Le Goualher G, Boomsma D, Cannon T, Kawashima R, Mazoyer B. A probabilistic atlas and reference system for the human brain: International Consortium for Brain Mapping (ICBM). *Philos Trans R Soc Lond B Biol Sci* 2001;356:1293–1322.
101. Mazziotta J, Toga A, Evans A, Fox P, Lancaster J, Zilles K, Woods R, Paus T, Simpson G, Pike B, Holmes C, Collins L, Thompson P, MacDonald D, Iacoboni M, Schormann T, Amunts K, Palomero-Gallagher N, Geyer S, Parsons L, Narr K, Kabani N, Le Goualher G, Feidler J, Smith K, Boomsma D, Hulshoff Pol H, Cannon T, Kawashima R, Mazoyer B. A four-dimensional probabilistic atlas of the human brain. *J Am Med Inform Assoc* 2001;8:401–430.
102. Dickson J, Drury H, Van Essen DC. The surface management system (SuMS) database: a surface-based database to aid cortical surface reconstruction, visualization and analysis. *Philos Trans R Soc Lond B Biol Sci* 2001;356:1277–1292.
103. Van Essen DC. Windows on the brain: the emerging role of atlases and databases in neuroscience. *Curr Opin Neurobiol* 2002;12:574–579.
104. Burns GA. Knowledge management of the neuroscientific literature: the data model and underlying strategy of the NeuroScholar system. *Philos Trans R Soc Lond B Biol Sci* 2001;356:1187–1208.
105. Shepherd GM, Mirsky JS, Healy MD, Singer MS, Skoufos E, Hines MS, Nadkarni PM, Miller PL. The Human Brain Project: neuroinformatics tools for integrating, searching and modeling multidisciplinary neuroscience data. *Trends Neurosci* 1998;21:460–468.
106. Rammensee HG. Immunoinformatics: bioinformatic strategies for better understanding of immune function. Introduction. *Novartis Found Symp* 2003;254:1–2.
107. Brusic V, Zeleznikow J, Petrovsky N. Molecular immunology databases and data repositories. *J Immunol Methods* 2000;238:17–28.

108. Robinson J, Waller MJ, Parham P, de Groot N, Bontrop R, Kennedy LJ, Stoehr P, Marsh SG. IMGT/HLA and IMGT/MHC: sequence databases for the study of the major histocompatibility complex. *Nucleic Acids Res* 2003;31:311–314.
109. Lefranc MP. IMGT, the international ImMunoGeneTics database. *Nucleic Acids Res* 2003;31:307–310.
110. Giudicelli V, Lefranc MP. Ontology for immunogenetics: the IMGT-ONTOLOGY. *Bioinformatics* 1999;15:1047–1054.
111. Petrovsky N, Schonbach C, Brusic V. Bioinformatic strategies for better understanding of immune function. *In Silico Biol* 2003;3:411–416.
112. Lefranc MP. IMGT, The international ImMunoGeneTics Information System. *Methods Mol Biol* 2004;248:27–49.
113. Lefranc MP, Giudicelli V, Ginestoux C, Bosc N, Folch G, Guiraudou D, Jabado-Michaloud J, Magris S, Scaviner D, Thouvenin V, Combres K, Girod D, Jeanjean S, Protat C, Yousfi-Monod M, Duprat E, Kaas Q, Pommie C, Chaume D, Lefranc G. IMGT-ONTOLOGY for immunogenetics and immunoinformatics. *In Silico Biol* 2004;4:17–29.
114. Lefranc MP. IMGT-ONTOLOGY and IMGT databases, tools and Web resources for immunogenetics and immunoinformatics. *Mol Immunol* 2004;40:647–660.
115. Brusic V, Zeleznikow J. Artificial neural network applications in immunology, in *Proceedings of the International Joint Conference on Neural Networks*, 1999;5:3685–3689.
116. De Groot AS, Bosma A, Chinai N, Frost J, Jesdale BM, Gonzalez MA, Martin W, Saint-Aubin C. From genome to vaccine: *in silico* predictions, *ex vivo* verification. *Vaccine* 2001;19:4385–4395.
117. De Groot AS, Sbai H, Aubin CS, McMurry J, Martin W. Immuno-informatics: mining genomes for vaccine components. *Immunol Cell Biol* 2002;80:255–269.
118. de Castro LN, Timmis JI. Artificial immune systems: a novel paradigm to pattern recognition. In *Artificial Neural Networks in Pattern Recognition*. Paisley, UK: University of Paisley; 2002, pp 67–84.
119. Korolev D, Balakin KV, Nikolsky Y, Kirillov E, Ivanenkov YA, Savchuk NP, Ivashchenko AA, Nikolskaya T. Modeling of human cytochrome P450-mediated drug metabolism using unsupervised machine learning approach. *J Med Chem* 2003;46:3631–3643.
120. Erhardt PW. A human drug metabolism database: potential roles in the quantitative predictions of drug metabolism and metabolism-related drug–drug interactions. *Curr Drug Metab* 2003;4:411–422.
121. Langowski J, Long A. Computer systems for the prediction of xenobiotic metabolism. *Adv Drug Deliv Rev* 2002;54:407–415.
122. Barratt MD, Rodford RA. The computational prediction of toxicity. *Curr Opin Chem Biol* 2001;5:383–388.
123. Pearl GM, Livingston-Carr S, Durham SK. Integration of computational analysis as a sentinel tool in toxicological assessments. *Curr Top Med Chem* 2001;1:247–255.
124. Dearden JC, Barratt MD, Benigni R, Bristol DW, Combes RD, Cronin MTD, Judson PN, Payne MP, Richard AM, Tichy M, Worth AP, Yourick JJ. The development and validation of expert systems for predicting toxicity. *ATLA* 1997;25:223–252.
125. Schultz TW, Cronin MTD, Walker JD, Aptula AO. Quantitative structure–activity relationships (QSARs) in toxicology: a historical perspective. *J Mol Structure (Theochem)* 2003;622:1–22.
126. Covitz PA, Hartel F, Schaefer C, De Coronado S, Fragoso G, Sahni H, Gustafson S, Buetow KH. caCORE: a common infrastructure for cancer informatics. *Bioinformatics* 2003;19:2404–2412.

127. Silva JS, Ball MJ, Douglas JV. The Cancer Informatics Infrastructure (CII): an architecture for translating clinical research into patient care. *Medinfo* 2001;10:114–117.
128. Hubbard SM, Setser A. The Cancer Informatics Infrastructure: a new initiative of the National Cancer Institute. *Semin Oncol Nurs* 2001;17:55–61.
129. Kihara D, Yang YD, Hawkin T. Bioinformatics resources for cancer research with an emphasis on gene function and structure prediction tools. *Cancer Informatics* 2006; pp 25–35.

# COMPUTER TECHNIQUES: IDENTIFYING SIMILARITIES BETWEEN SMALL MOLECULES

PETER MEEK, GUILLERMO MOYNA, AND RANDY ZAUHAR
*University of the Sciences in Philadelphia, Philadelphia, Pennsylvania*

**Contents**

1 Introduction
2 Computer-Aided Drug Design (CADD)
3 Harvesting Data from Small Molecules
4 Representing Molecules for Interpretation by Computers
  4.1 Importance of Continuity Within Molecular Representations
5 Defining Similarity
  5.1 Why Do We Wish to Compare Molecules?
  5.2 Utilizing External Sources of Information
6 Detecting Similarity with *in silico* Techniques
  6.1 HQSAR
  6.2 QSAR
  6.3 Superimposition
  6.4 Program Suites
  6.5 Comparative Molecular Field Analysis (CoMFA) and Related Approaches
  6.6 Shape Signatures
7 Conclusion
  Acknowledgments
  References

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

## 1 INTRODUCTION

In this chapter, current computational methodologies available for the comparison of structurally similar compounds are presented and analyzed in detail. Problems encountered throughout these types of endeavors are summarized first, giving particular emphasis to the suitability of *in silico*, or computer-based, methods for evaluating molecular similarity among members of large compound libraries. The chapter then concentrates on the description of different techniques, starting from the simplest ones, then more complex, and finishing with emerging approaches. Salient examples of a technique application found in the literature are discussed, and brief guidelines for readers interested in employing these computational tools are outlined.

## 2 COMPUTER-AIDED DRUG DESIGN (CADD)

Implementing CADD techniques depends heavily on what molecular information is provided or accessible. However, of equal or even greater importance is how the information that is searched and utilized is organized and constructed. These, of course, are the databases containing the small molecules. The format of how a molecular entry is recorded and stored is pivotal and dictates what transformations of the entries are required before any kind of search or CADD application can be performed. In fact, this chapter might be described as a series of illustrations of the different ways in which molecules can be represented in a computer, each with its potential benefits and drawbacks.

## 3 HARVESTING DATA FROM SMALL MOLECULES

Over the past few decades there has been an unprecedented explosion of information in the chemical and life sciences. Development of new biological and chemical tools has been documented and new pharmaceutical agents are discovered with ever-increasing frequency year by year [1]. At the same time, it has become more and more difficult to gain regulatory approval of new chemical entities [2] due to new and stricter legislation. There are millions of molecules that are now known to possess bioactivity against one or more targets, or to be useful as reagents in academic and industrial research. Before any computer-based technique can be implemented, these vast numbers of isolated and synthesized molecules need to be organized. As might be expected, the need to catalog this abundant information was recognized early on, and many molecules are organized into indexes (e.g., CAS system) and/or are held in commercial [3–6], industrial [7, 8], or public [9–11] databases. Given the vast volume of "chemical space" that has already been explored, it is to the advantage of the investigator to consult these knowledge bases before planning a new synthesis, and at the outset of any efforts to identify new therapeutic or bioactive compounds. In particular, synthesizing a new compound often represents an enormous effort in the laboratory, work carried out in vain if the target compound (or one similar to it) is already available from a commercial source.

Given the sheer volume of information, there is considerable need to employ computers to carry out these investigations as few human minds past, present, or future can compare one molecule to a previously memorized database of 10,000,000 molecules. Harnessing computer power is not without drawbacks; we discuss here the importance and careful considerations required to represent chemical structures. Chemical database storage and searching relies on electronic definitions of chemical structures, and the details of the methodology used are vital to defining and measuring similarity between molecules.

## 4 REPRESENTING MOLECULES FOR INTERPRETATION BY COMPUTERS

Molecules are represented based on their constituent parts and the way in which these constituent parts are connected. The components of a small molecule are atoms, which are minimally identified by element type, but which may be characterized by a number of additional descriptors, including but not limited to van der Waals radius, valence, and/or charge state. The next step beyond simply enumerating the atoms in a molecule is to provide detailed connectivity information, typically in terms of a list of bonds of recognized chemical type; in this way the molecule is defined mathematically as a graph, with the atoms as vertices and the bonds as edges of specified type. The atom and bond definitions act as the building blocks of molecules that can be represented in one of three major forms: one-dimensional (1D), two-dimensional (2D), or three-dimensional (3D). Over the years several standard computer file formats have been developed for each dimension class (Table 1), and these have steadily evolved to allow incorporation of additional data types. For instance, many applications require a detailed distribution of partial atomic charges (as opposed to simply specifying ionization state); such charges are required by some force fields, such as MMFF94 [19], and are routinely computed using rapid semiempirical methods (Gasteiger [20] and Gastegier and Huckel [21]). This information is readily included in many molecular file formats, including the structure data file sdf [22] and mol2 [14] specifications. In addition, some current formats (such as sdf) admit the possibility of "tagging" a molecule with arbitrary data, a boon in the sense of providing open-ended flexibility, but a peril in that descriptors may not be added uniformly across a database, and/or the additions may not be clearly annotated.

### 4.1 Importance of Continuity Within Molecular Representations

One of the most critical issues when applying various molecular databases is to recognize the varying levels of faithfulness between the representation of the molecule in the computer and the actual material available to the investigator. This is a function of both the database and the source of the compound. For example, suppose that a sample of a potentially bioactive compound is available in the form of a purified enantiomer. Many molecular databases store connectivity (bonding) information but do not distinguish between different stereoisomers. If the database entry corresponding to this compound is used to generate a molecular model with

**TABLE 1 Representation of the Small Molecule β-D-Glucopyranose**

*1D*[a]

SMILES string representation of β-D-glucopyranose (@ = S, @@ = R):

[H][C@@]1(O)O[C@]([H])(CO)[C@@]([H])(O)[C@]([H])(O)[C@@]1([H])O

SLN string representation of β-D-glucopyranose:

O[1]CH(CH(CH(CH(CH@1OH)OH)OH)OH)CH2OH

*2D*[b]

MarvinSketch representation of β-D-glucopyranose

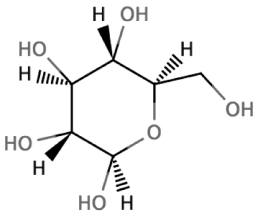


*3D*[c]

SYBYL representation of β-D-glucopyranose

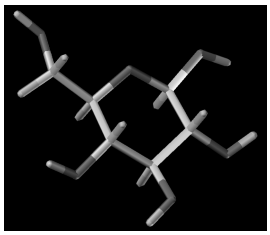

[a]Two examples: a SMILE string notation [12] and the SLN (SYBYL Line Notation [13], by Tripos [14]).
[b]Representation using the classic bond-wedge diagram (MarvinSketch [15]).
[c]The image of β-D-glucopyranose created from a mol2 file (SYBYL by Tripos). The 2D and 3D structures are collapsible to linear notation and converted into a required file format by CORINA [16]; stereoisomers are generated with STERGEN [17] and tautomers with TAUTOMER [18].

detailed atomic positions, the resulting model may not correspond to the material sample. Conversely, the investigator may have a racemic mixture, whereas the database entry assumes a specific enantiomeric form. In either case, incorrect predictions may result, since biological activity is often remarkably sensitive to the details of 3D molecular structure.

Generally, it is often the case that small changes in chemical structure or physicochemical properties can have profound implications for the specificity and magnitude of the biological effects associated with a compound. Speaking first to the effects of structural variations, we can identify classic examples of chemical isomerism, including chiral, cis/trans (*E/Z*), functional group, and positional isomerism. Small variations such as substitution of atoms and/or isomeric rearrangements can have an important impact on both molecular shape and physical properties (e.g., dipole moment, hydrophobicity) that are a function of 3D structure. Hence, it is imperative for the investigator to be aware of the details in the information stored in a database, particularly when data is absent (e.g., no representation of stereoisomerism) as well as when data is present but may not correspond to the state of the physical sample (e.g., *R* chiral form where *S* is found in the material sample, or a neutral database entry corresponding to a molecule that may be ionized at the pH of interest). It is therefore imperative to use, derive, and develop uniform molecular representation systems meticulously, without which predictions or application of *in silico* techniques would be impossible.

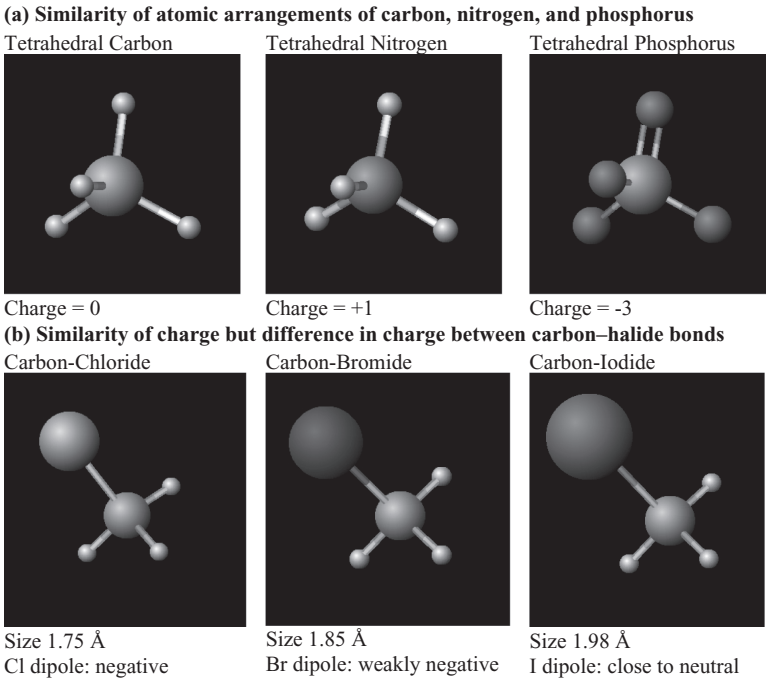


**FIGURE 1** (**a**) Depiction of three tetrahedral atom arrangements: (**left**) the central **carbon** in methane, (**center**) the central protonated **nitrogen** in ammonium, and (**right**) the **phosphorus** in phosphate all look similar but carry different charges. (**b**) Atomic arrangements of a tetrahedral carbon bonded with a single halide of increasing size: (**left**) chloromethane, (**center**) bromomethane, and (**right**) iodomethane. All molecule constructions were created with MarvinSketch [15].

To highlight this concept further, we present some examples of small molecular frameworks that are at first glance similar but that in fact differ significantly (Fig. 1). Certain atoms have similar sizes and the same bond orientation (e.g., tetrahedral carbon, phosphorus, and protonated nitrogen, Fig. 1a) yet the molecular charge distribution is considerably different. In contrast, substituting one halide for another at a specified location in a framework has a small effect on total dipole moment, and yet these atoms differ significantly in size, with the largest, iodine, occupying about the same volume as a methyl group (Fig. 1b). This raises the question: When similarity is measured, what governs how similar an atom is to another atom? In other words, how similar is O to S, or N to P or C, or an atom to a small group (e.g., I to $CH_3$)? The answer depends heavily on the context of the atom or group in question. Moreover, it is important to recognize that the physical influences of atom or small group in determining molecular properties depend on its context; such factors include hybridization state and the local chemical environment, an insight first proposed by Hammet [23] and Hansch et al. [24]. These researchers developed mathematical representations for predicting molecular properties on the basis of structure.

Yet more complexity arises if we introduce the possibility of molecular flexibility. Most molecules of biological interest admit some degree of flexibility, including bonds about which free rotation is possible and saturated ring systems that can

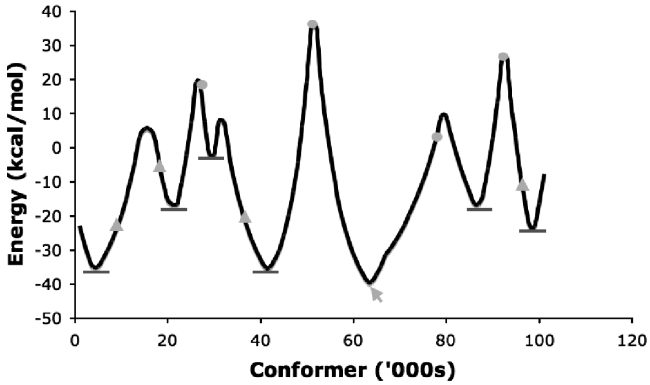


**FIGURE 2** Graph of molecular conformation versus energy. An energy profile plot is shown for each possible conformation of a hypothetical molecule in 3D space. The troughs denote local energy minima (short gray lines) and the global energy minimum (gray arrow). When exploring molecular space, finding the global energy minimum is not guaranteed. For example, if the four starting points were the gray triangles, the output would not detect the global energy minimum, but maybe five local energy minima at best. Using the start points with the gray circles would almost certainly guarantee the global energy minima and possibly all the local minima too. The problem is inherent to all experimental investigations: if the sample size is not large enough, the minima determined will not be representative of the whole. Likewise, in energy calculations if the energy barrier is too high to climb, the pockets or valleys over the hills cannot be entered into and located. This is why small molecules bound to their targets are highly sought after; a more active molecule will have a bound structure not too dissimilar to its gas phase global minima.

adopt alternate conformations (e.g., chair vs. boat forms for six-membered rings, various classes of puckering observed in pyranoses). This leads to the possibility of multiple conformers for a molecule, each representing a local minimum on the energy surface of the compound. While a molecule will often have a single global energy minimum representing the conformation most favored in terms of energy, there may be other important minima of similar energy (Fig. 2), and there is no guarantee that the global minimum corresponds to the form of the molecule that is active against a particular biological target [25]. A number of techniques are available to identify low energy conformers [14, 26]. Although many databases include a representative low energy conformer for each compound, detailed considerations of molecular flexibility are usually relegated to the final stages of analysis, after a relatively small set of interesting molecules have been identified. Nonetheless, consideration of flexibility is often critical for rationalizing differences in activity among a panel of bioactive molecules.

## 5 DEFINING SIMILARITY

Defining similarity between molecules is an inherently ambiguous undertaking. What makes two molecules similar? Should we focus on the volumes occupied by molecules, their patterns of chemical bonding, the presence of particular functional

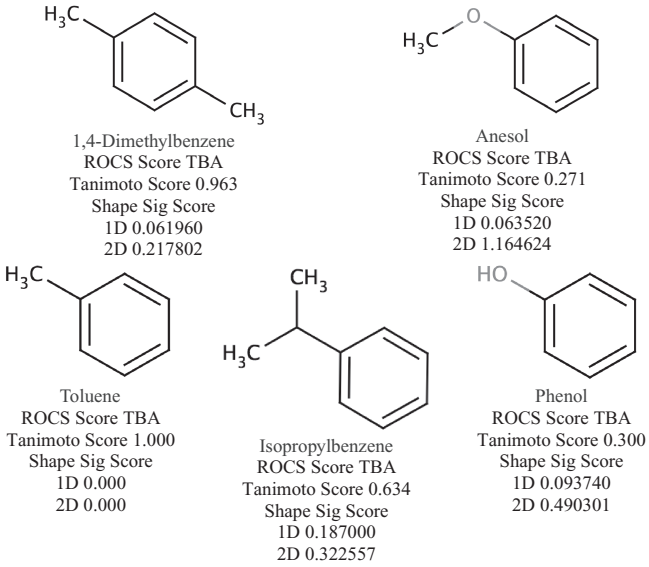


**FIGURE 3** A simple similarity example using different comparison methods. A collection of five benzene derivatives is displayed and an assessment is made of their similarity to a toluene reference using Tanimoto index [27], 1D Shape Signatures, (shape comparison), and 2D Shape Signatures (shape and mean electrostatic potential comparison) [28–30]. All three types of comparison required 3D coordinate input files (mol2). A Tanimoto score of 1.000 implies a perfect match, with 0.000 implying a complete mismatch; a Shape Signatures score of 0.000 is a perfect match and 2.000 is a complete mismatch.

groups, ionization state, hydrophobicity, dipole moment, or any or all of a host of other topological and physicochemical descriptors? In most cases the desire is to find molecules that "look like" a particular query molecule, or that are compatible with a chemical/spatial pattern (corresponding perhaps to a pharmacophore model). Note that we are not typically interested in stringent criteria; rather, the goal is to cast a fairly wide net and identify molecules that have the potential to interact with a particular target of interest. Examples of comparisons (Fig. 3) demonstrate that different tools, each employing its own valid approach to defining molecular similarity, can produce very different results. This is not in itself surprising but does highlight the need for care and caution when applying *in silico* methods to scan chemical databases for lead compounds. It is important to understand what features a particular method is "looking for," and it is just as critical to recognize what the method ignores! Similarity is itself an arbitrarily defined concept (with appreciable overlaps between techniques) and will depend entirely on the tool(s) used.

Essentially, *in silico* search techniques prove most useful when the similarity measurement (scoring index) reliably returns molecules with the same desired characteristics as the query. Such a technique will take a large dataset of unrelated molecules and reduce it to a smaller set, but with proportionally more molecules similar to the query. This process is called *enrichment*. Methods are available to quantitate enrichment [29, 31], and a measure of enrichment may provide a useful figure of merit when comparing and assessing computational search tools.

### 5.1 Why Do We Wish to Compare Molecules?

In the context of drug development (or identification of bioactivity in general), there are two motivations for comparing molecules. One is to locate molecules in a database that are similar to a known active, which might also be active against the target, and furthermore could conceivably provide better compounds as determined by any number of measures (e.g., activity, bioavailability, ease of synthesis). The second is to rationalize the efficacy of a set of known active compounds, to identify the features that are directly related to activity, and thus to determine how existing compounds can be improved. Of course, the two motivations can be combined: one might first use a library search to locate interesting lead compounds and then improve the leads by making modifications in line with established models that relate structure to activity. Clearly, at the very least, the database search component of this strategy needs to be automated, since humans do a poor job of scanning through hundreds of structures, let alone the hundreds of thousands that are available in contemporary libraries. But the central point is that drug design methods that employ measures of molecular similarity have the potential to rapidly locate lead compounds, to support the finetuning of structures to improve activity and other critical features, and to dramatically reduce costs by identifying compounds that may be readily available or easily modified.

Similarity searching is not a straightforward task: small alterations of a molecule (e.g., *cis/trans* stereoisomerism (Fig. 4a) and chirality (Fig. 4b)) can have a profound and marked effect on activity. Indeed, the protonation state of the molecule

**FIGURE 4** (**a**) Shape and *cis/trans* isomerism. This part demonstrates how a subtle change in the arrangement of a carbon–carbon double bond ($sp^2{=}sp^2$) can dramatically affect the appearance of a molecule. The bond (black dot) is either in the *cis* conformation (same side), (**i**) and (**iii**), or the *trans* conformation (opposite sides), (**ii**) and (**iv**). One should take paticular note that as small molecules are usually designed to biological targets, it is inevitable that *cis/trans* isomerism can dramatically affect the activity of a potential drug, and consequently is a major headache in chemical synthesis. (**b**) Shape and chiral isomerism. Controlling chirality poses a huge challenge for synthetic chemists. There are several chiral centers in this molecule; the only differing one is marked with an arrow, being *S* on the isomer to the left and *R* on the one to the right. Usually one optical isomer, or enantiomer, will have activity against a biological target (known as the eutomer), while the other lacks the desired activity. Due to the expense associated with the separation or enantiomers from a mixture, chirality is of paramount importance when searching molecular databases. (**c**) Shape and distribution of charge. Shown are two representations of aspartic acid—an amino acid: (**left**) the fully protonated amino acid and (**right**) the fully deprotonated amino acid. Usually the nitrogen is tetrahedral at physiological pH and the C terminus and side chain (connected to the Cα-carbon) are deprotonated. The state of the proton donors and acceptors modifies the local charge and overall shape of the molecule; hence, the small molecule being investigated needs to mimic the donor and acceptor sites to be truly similar. As a general rule, the termini are not involved in donor and acceptor sites on a protein, as the vast majority are lost in the peptide bond formation. The side chains dictate the number of donor and acceptor sites.

can affect the structure and surface charge distribution (Fig. 4c), with flexibility and rotation further complicating the matter. Hence, one could define this problem as akin to a box of left-threaded and right-threaded nuts and bolts. The different classes of bolts may have many properties in common, such that even a trained eye might sort them together, yet they have distinct and incompatible functions. This paradigm is clearly evident in the biological context and highlighted with

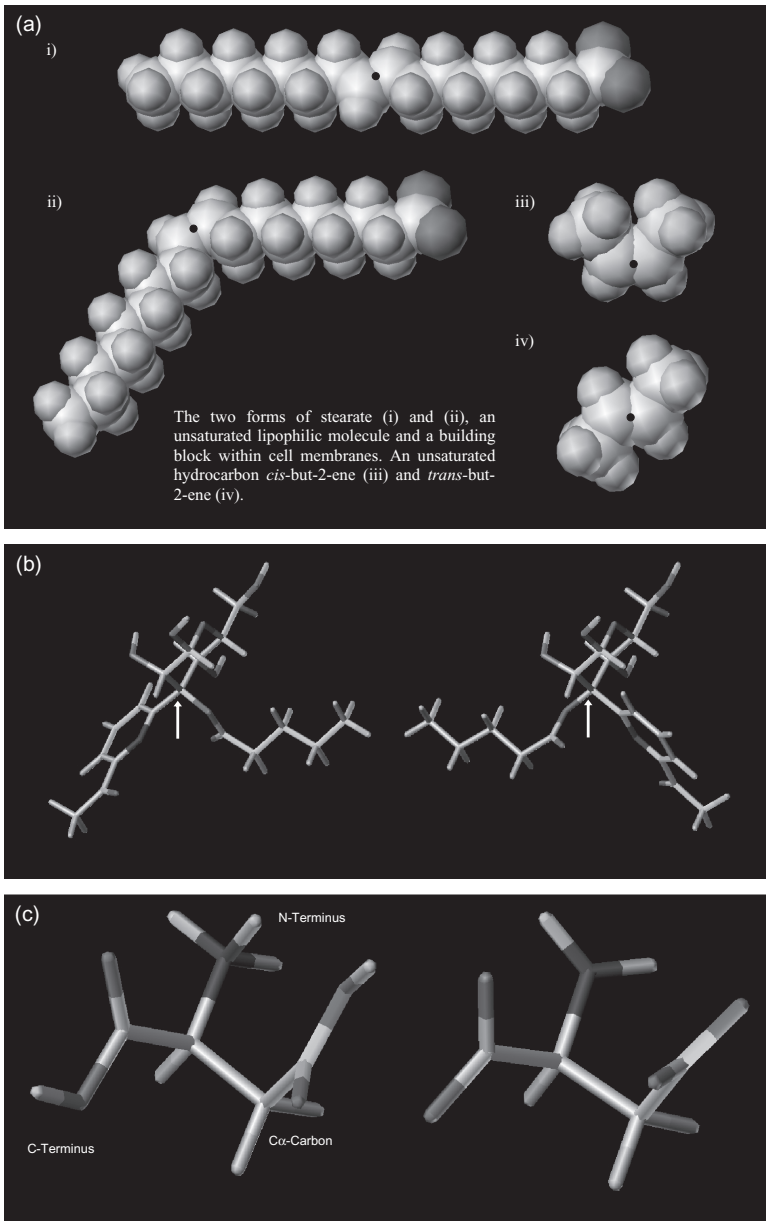

respect to enantiomer specificity (HIV serine proteases made of exclusively either D or L amino acids) [32] and the toxic effects observed with thalidomide [33].

Consequently, while attempts to measure similarity between molecules have an objective basis, the evaluations depend heavily on the algorithms employed, sometimes leading to the appearance of subjectivity (the results obtained depend on the specific tools selected). That said, it has been evident for a long while that molecular similarity provides the foundation for rational drug design. As far as the choice of tools and algorithms is concerned, these are of course ultimately evaluated at the laboratory bench, and it is empirical studies that in the end provide the validation of existing approaches, and point out new directions for improvement and development on the computational side.

### 5.2 Utilizing External Sources of Information

As discussed earlier, several scenarios admit the use of similarity searching. In some cases, a relatively small set of molecules are compared, but in others the goal is to scan a potentially enormous library of compounds. To a large extent, the form of the investigation will be determined by the molecular descriptors maintained in the target database. These vary considerably from one database to another; for example, many commonly used compound databases (National Cancer Institute [10]) include only chemical formulas in the form of SMILES strings [12], with no 3D structures. On the other hand, these same libraries may be offered by other vendors with atomic coordinates appended (Tripos version of NCI), and furthermore a number of fast tools are available to generate coordinates from SMILES strings (CORINA [16], CONCORD [34]), so this data can be added by the end user. Some databases include a range of molecular descriptors for each entry, such as $\log P$, molar refractivity, molecular weight, or custom topological descriptors (e.g., chemical fingerprints [35]). Again, these descriptors, when missing, can sometimes be added by third-party tools (MOE [36], DayLight [37]). At the same time, whether these newly added atomic coordinates and/or descriptors can readily be used with available search software is a question that needs to be addressed at the outset; if a tool requires a proprietary database format, it is an open question as to whether any modifications can be introduced by the end user. An extreme case is found when the database is accessible only via a website maintained by a company or academic laboratory, and the raw data is inaccessible except through the supplied interface. The most attractive approach, rarely realized, is to maintain a flexible database format and something akin to a relational database that can be used to construct arbitrary queries of the information held in the library.

Databases for use in similarity searching are difficult to procure or expensive to obtain [3–8]. Recently, several have become available that are utilizable and relevant to small molecule comparison [9–11]. The research conducted at the University of the Sciences in Philadelphia aims to provide a readily searchable database to the global academic community (see Section 6.6).

## 6 DETECTING SIMILARITY WITH *IN SILICO* TECHNIQUES

There is a vast expanse of *in silico* techniques that "detect" similarity between small molecules, far too many to cover all in detail here due to space constraints. However,

the aim is to give a flavor of techniques that are in common use and applied in contemporary research. We begin with established techniques that are the least complex and require the least amount of input data and move through to more complicated data intensive methods and finally emerging technologies.

## 6.1 HQSAR

HQSAR (hologram quantitative structure–activity relationships) is a powerful approach that can claim important advantages over many of the traditional QSAR (quantitative structure–activity relationship) techniques we will describe later. This is especially true when little is known about the receptor target of interest. A key advantage of HQSAR is that it requires only 2D chemical structures combined with experimental binding data; there is no need to generate atomic coordinates or alignments of structures, as is required by some 3D methods like CoMFA (comparative molecular field analysis, discussed in Section 6.5). At the same time, predictions from HQSAR can be combined with other methods to sometimes improve the overall accuracy of predicted activities. Models of high statistical quality and predictive value can be rapidly and easily generated by HQSAR and can often be used in conjunction with chemical databases to select molecules with superior biological activity, providing the model is robust. HQSAR can be used to rapidly scan chemical libraries but is also useful with small datasets.

HQSAR requires a set of training molecules in order to construct predictive models. As with any machine learning algorithm, the reliability of the model generated is largely a function of the quality of the training set; it is imperative that the training molecules be well characterized, with a common binding site and accurate $IC_{50}$ measurements.

***HQSAR Application*** HQSAR is a rapid and inexpensive means to test compounds using a computer algorithm. The computer "learns" from the input provided from real data derived from actual experiments and generates a model (Fig. 5a). This model in turn serves as the basis for predicting the activity of compounds that have been proposed as new leads, or for scanning a chemical library for compounds likely to be active (Fig. 5b). While we will describe the HQSAR methodology in more detail below, the basic idea is straightforward. It is usually true that a biologically active compound has three or four groups that are crucial to activity, and these are held at specific distances by a skeleton/backbone structure. HQSAR aims to identify these groups and to correlate the arrangement of these key components with activity.

***HQSAR: A Computational Perspective*** To understand HQSAR, an appreciation of the computational algorithm is required, and for this we need to consider the underlying representation of the data. Linear notation (more specifically SLN) is utilized to represent molecules (Table 1), [12, 13, 38], but the molecule is finally represented as a fingerprint [35]. The key importance of HQSAR fingerprint representations (Fig. 6) is that fragments are generated based on the chemical structure of each compound and need not correspond to a predefined list of functional groups and structural components. This flexibility leads to more robust database

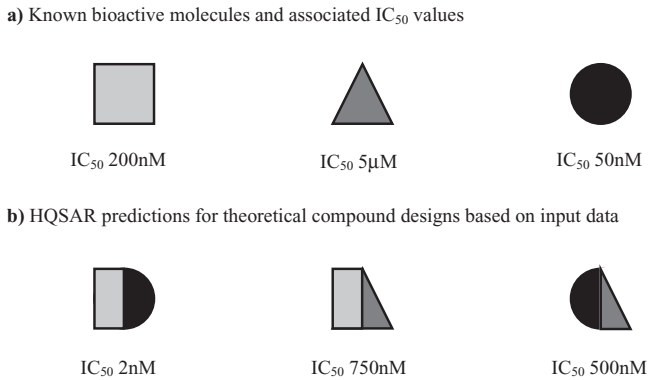


**FIGURE 5** Simplified representation of HQSAR with a small molecule training and test set. The methodology behind HQSAR consists of (**a**) the training set of known molecules and (**b**) the predictions for theoretically designed molecules. The points at the very extremities of the shapes can be considered as active groups, and the overall shape as a molecular skeleton. When an HQSAR run is conducted, the molecules in the training set (**a**) define whether particular groups contribute to or negate activity in the HQSAR model (in this case experimentally derived $IC_{50}$ values are used). Theoretical compounds incorporating variations in the component parts (**b**) are then tested with this HQSAR model. In many incidences results make sense (**lower center** and **lower right**) being between the $IC_{50}$ values of the two component parts. There are cases where the $IC_{50}$ can appear to be greater than the sum of the component parts (**lower left**), and in this case it is in a more favorable direction: 2 nM while previously the best molecule was 50 nM (**upper right**).

| | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Fingerprint | 0 | 0 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 0 |
| Hologram | 0 | 0 | 2 | 5 | 0 | 9 | 0 | 0 | 0 | 1 | 3 | 8 | 7 | 0 | 0 | 0 |

$\Sigma = 35$

**FIGURE 6** An example of a fingerprint and a hologram. The chemical structure has 35 fragments. In the case of the hashed fingerprint, the representation is purely binary, whereas in the case of the hologram, the bins contain information about the number of fragments hashed into each bin and its contribution (positive, negative, or otherwise to biological activity).

screening [37], and more effective 2D and/or similarity searches, than a structural key approach [39]. The package used [14], however, provides the capabilities of each approach.

Conventional QSARs can identify critical relationships between the properties—the geometric and the chemical characteristics of a molecular system of interest. In QSAR, the measured bioactivity of a set of compounds is correlated with structural descriptors to determine trends and predict the bioactivity of related, untested compounds. HQSAR (like combined QSAR and CoMFA) works by identifying substructural features in sets of molecules that are relevant to biological activity. A key advantage of HQSAR is the relative simplicity of the required input, which in the training phase consists of chemical (2D) structures and their experimental activity (i.e., relative binding affinity, $LD_{50}$ or $IC_{50}$).

The underlying principle of HQSAR is simply stated: since the structure of the molecule is encoded within its 2D fingerprint and that structure is a key determinant of all molecular properties including biological activity, then it should be possible

to predict the activity of a molecule from its fingerprint, which implicitly encodes the structure. Molecular fingerprints (Fig. 6) are strings of 1's and 0's (binary) that indicate the presence or absence of a particular fragment (e.g., a carbonyl, hydroxyl, or halide). The fingerprint is therefore a shorthand list of the fragments present in a structure. A *molecular hologram* extends this concept to incorporate both *branched fragments* and *counts* of the fragments that are present. These additions are important as the inclusion of branched fragments helps distinguish hybridization states, and the counting function differentiates between compounds with (for example) one, two, or more carbonyls. Thus, a hologram is a list of integers, not a bit string, and is therefore a fingerprint. In summary, a molecular hologram contains all possible molecular fragments within a molecule, including overlapping fragments, and maintains a count of the number of times each unique fragment occurs. This process of incorporating information about each fragment, and each of its constituent subfragments, implicitly encodes 3D structural information. These fragments are arranged into (or hashed) a linear array of numbers (the hologram, Fig. 6). Note that the hashing algorithm guarantees that a particular fragment will always hash to the same numerical value but does not guarantee that different fragments will always have unique hash values. If different fragments, all important for determining biological activity, should hash to the same bin positions in the hologram, it will not be possible to distinguish their effects using a partial least squares analysis (or any other approach for that matter). This phenomenon is called fragment collision and results in poor models with little predictive validity. Using holograms of different lengths ($L$) can prevent fragments that are responsible for biological activity from being hashed to the same bin.

When the various lengths of holograms are chosen at even intervals, the chances of resolving bad collisions are reduced. For example, if fragment collisions occur with the length set at 400, these collisions will not be resolved at length 200 (nor 100, 50, or 25 as these are all factors of 400). Since the 400 bin hologram is simply folded over to produce the 200 bin hologram, the colliding fragments will still end up in the same bin. For this reason, the default values for the hologram lengths are all prime numbers. This reduces the chances of seeing the same bad collisions at the various lengths. It is wise to use all available lengths so that the best model can be generated in a single HQSAR run.

Despite the fundamental simplicity of HQSAR, there are many options and much terminology connected with the application of the method, only a portion of which is described here. Each molecule in the training set is broken down into structural fragments whose size falls within a specified range; here we will denote the lower limit for the number of atoms in a fragment as $M$, and the upper limit as $N$. The parameters $M$ and $N$ are user-defined (typically with $M = 4$–$7$ and $N = 7$–$9$), as is the length ($L$) of the molecular hologram into which the fragments are hashed. Given the length $L$ of the hologram, a fragment with a particular chemical structure always hashes to the same well-defined bin. However, the hashing function can be adjusted to ignore certain molecular features; for example, the user controls whether or not fragments with the same chemical structure but opposite chirality hash into the same or different hologram bins.

The calculation of a set of molecular holograms for a training set of structures yields a data matrix of dimension $n \times L$, where $n$ is the number of compounds in the dataset and $L$ is the length of the molecular hologram. The partial least squares

technique is then employed to generate a statistical model that relates the descriptor variables (occupancy numbers of the bins in the hologram) to an observable property, for example, the biological activity expressed as $-\log IC_{50}$. The selection of the hologram length leading to the "best" HQSAR is based on the PLS analysis that gives either the highest cross-validated $q^2$ or the lowest standard error associated with the cross-validation analysis. The predictive power of the model is determined by using statistical cross-validation, the default approach being the leave-one-out (LOO) method.

An important role of a QSAR model, besides predicting the activities of untested molecules, is to provide hints about what molecular fragments may be important contributors to activity. This information is of great value to the synthetic chemist. Such information can be combined with knowledge of the maximal common structure (MCS) between compounds to get ideas for the synthesis of new molecules that might lead to suitable drug candidates. Clearly, the HQSAR method, which focuses directly on elements of chemical structure (rather than derived properties such as dipole moment or hydrophobicity), is well suited to proposing changes in chemical structure that are likely to improve activity.

The contributions of individual atoms to activity can be assessed through their representation in the hologram hash bins and can be used to color code the atoms of each molecule. Atoms that contribute positively to activity are colored toward the blue end of the spectrum; atoms that contribute negatively are colored toward the red end, and atoms that do not significantly affect activity (usually constituting the backbone or skeleton of the molecule) are colored white.

An important feature of HQSAR is the capability to recognize the maximal substructure common to the training compounds and to remove this substructure from consideration when constructing holograms. It is usually the case that a series of active molecules will share a significant common framework, and the predictive power of the method is enhanced by excluding portions of the molecule that are shared across the training set, and thus cannot be correlated with activity. Using this maximum common substructure (MCS) feature also allows the user to exclude in a consistent way molecules that lie too far outside the predictive space of the model. Moreover, identifying the MCS allows the user to better rationalize the relationship between structure and activity, and provides a template for proposing new compounds with activity that is likely to be accurately predicted by the HQSAR model. An MCS must contain at minimum seven connected atoms and is calculated purely by the molecules that constitute the training dataset; it is possible for good models to be derived without an MCS, but this is of less use when establishing and developing biological theories and of course for a centralized chemical synthesis route.

Once a HQSAR model has been calculated, the next step is to validate the model quality; the preferred method being the leave-one-out method (this is covered in detail in the next section). After model validation, proposed synthetics or other candidate theory molecules may be tested using the model. Often a search of chemical databases for molecules similar to those in the training set is conducted. Predictions of their activities, based on the model, are organized with the predicted activities in descending order. In a typical application, the database is first prescreened using a fast measure of similarity, such as the Tanimoto index, with a user-selected value for the minimum similarity cutoff (the default threshold for this index is 0.85). The results of the Tanimoto screen, which is effectively a 2D similarity

search, is carried out, and the hits found from the search are entered onto a spreadsheet. The activities of the compounds are calculated from the HQSAR model. Compounds are sorted according to their predicted measured variable (usually $IC_{50}$). These findings can be presented in simple bar charts with test compound versus predicted variable. Using the Tanimoto coefficient is not absolutely essential, but can aid in reducing the size of larger databases. Frequently a database including only those molecules proposed by medicinal chemists is used.

Successful HQSAR requires a minimum of five molecules in the training set, but preferably more than ten. These should be structurally different molecules, but with some similarity. The model should not be used to predict the activity of compounds that are structurally very different from those in the training set.

***HQSAR Example*** Vacuolar ATPase is an ATP-dependent proton pump that is important in many biological processes by acidifying specialized cell compartments. Of particular interest is the implication of V-ATPase in bone resorption by osteoclasts. Excessive bone resorption by osteoclasts and failure of osteoblasts laying down new bone are indicative of osteoporosis. The structural defects that accumulate and weaken the bones are primarily due to demineralization of the bone (removal of $Ca^{2+}$ and $PO^{3-}$). Approximately one in three postmenopausal women and one in twelve older men suffer from osteoporosis. Currently, treatment of osteoporosis is via two methods—bisphosphonates and selective estrogen receptor modulators, the latter not being the favored treatment for men. Current therapies are useful in two respects. First, the bisphosphonates rapidly increase bone mineral density by forming a dead end complex; hence, bone resorption is impaired. Second, estrogenic molecules have the beneficial effect of mimicking endogenous hormones promoting remineralization, but increase in bone density is poor compared to bisphosphonates. Estrogenics do offer a considerable advantage in that bone integrity is maintained, whereas with bisphosphonates microfractures tend to accumulate. To date, V-ATPase inhibitors are not suitable because of poor selectivity toward the osteoclast form of the enzyme and because of their more potent effect on kidney, liver, and neural forms. Structures and $IC_{50}$ data from literature sources incorporating medicinal chemistry to improve selectivity [40–45] were used to aid design of new V-ATPase inhibitors. The data was organized in a manner so HQSAR could specifically assess and augment compound activity against desired V-ATPase species (osteoclast), yet decrease activity against undesired V-ATPases (liver and kidney). HQSAR models with high predictive quality were produced and tested, and a selection of proposed compounds were synthesized and evaluated *in silico*.

All compounds from the source papers were entered into a SYBYL database and transferred to a spreadsheet manually and checked for absolute stereochemistry. While constructing HQSAR models is a straightforward process, producing a model that will be robust and have high predictive value requires care in the selection and processing of activity data. The quality of models is assessed using *cross-validation*. In this approach, subsets of the training data are selected at random, and removed. An HQSAR model is constructed using the observations that remain and is used to predict the activities of the points removed. There are different protocols for carrying out this procedure; in the "leave-one-out" technique, every element of the training set is removed, and its activity is predicted based on all the remaining observations. Whatever the protocol, the measure of the quality of the model is the

cross-validated $r^2$, denoted $q^2$, which measures the error when the points predicted are excluded when constructing the model. Cross-validation is critical not only for HQSAR but for any method that potentially uses a large number of descriptors, and where there is high probability of finding "random" variation in the attributes of the training molecules that will closely match any pattern of activity presented.

In the present case, the quality of the models (as measured by both $q^2$ and standard error) was improved by two means: normalizing $IC_{50}$ data relative to one compound of the set, or predicting differences between activity against different targets to produce a selectivity index (e.g., activity against chicken osteoclast cOc minus activity measured using bovine chromaffin granules bCG). An ideal set of compounds would contain $IC_{50}$ values that cover a wide range of activity (10–15 orders of magnitude or more) and also a diverse collection of functional groups, so that the resulting model is valid over a large chemical space. In particular, compounds that include large, complex groups (e.g., an Fmoc group, see Compound 4 in [40]) will be poorly predicted if underrepresented in the training set.

With the goal of the study to create, detect, and evaluate new compounds selective for osteoclast V-ATPase inhibition, some HQSAR results (Figs. 7–9) illustrate a few the best models produced. These were measured by achieving a cross-validated $q^2 > 0.5$ and a standard deviation < 1 between the training dataset (experimentally derived measured variable) and the predicted dataset (HQSAR model-derived measured variable) (Table 2). Within the figures the $y$-axis scale was calculated relative to bafilomycin $A_1$; in essence, this compound serves as a reference value for the HQSAR algorithm, to which all other compounds were compared (i.e., bafilomycin has an inhibitory selective index of 1 for chicken osteoclast

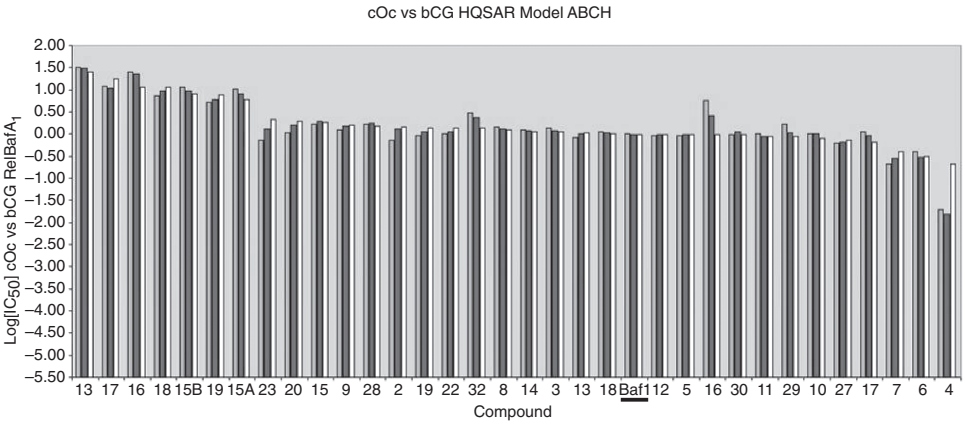


**FIGURE 7** Selection of HQSAR results I. Actual data (**light gray bars**) $\log IC_{50}$ cOc − $\log IC_{50}$ bCG of a compound measured relative to bafilomycin $A_1$ V-ATPase phosphate assay from original data [40, 44], versus predicted $\log IC_{50}$ according to HQSAR (**dark gray bars**), versus the cross-validated HQSAR prediction (**white bars**). The graph is sorted in decreasing order of the cross-validated prediction relative to bafilomycin $A_1$ specificity. The graph suggests a strong correlation between the actual measured values of cOc/bCG selectivity and the predicted values (small difference between actual and predicted data). Hence, the model is likely to be of strong predictive value. To summarize, any compound left of bafilomycin $A_1$ (**Baf1**) has an increased specificity for cOc; any compound to the right has a decreased specificity for cOc.

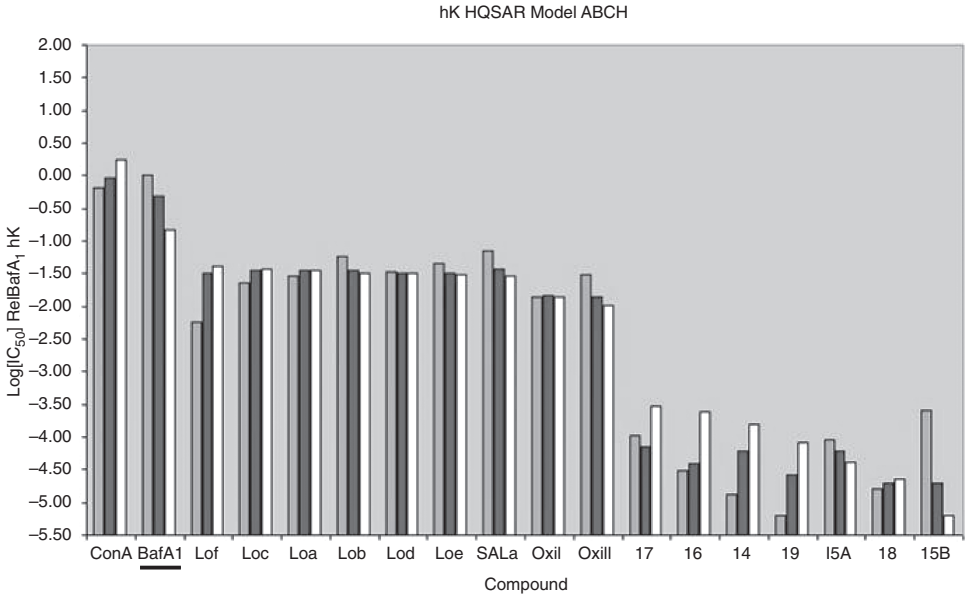


**FIGURE 8** Selection of HQSAR results II. Actual data (**light gray bars**) $\log IC_{50}$ hK of a compound measured relative to bafilomycin $A_1$ V-ATPase phosphate assay (**BafA1**) from original data [41, 44], versus predicted according to HQSAR (**dark gray bars**), versus the cross-validated HQSAR prediction (**white bars**). The graph is sorted in decreasing order of the cross-validated prediction and has a relatively strong predictive power. This data could provide a reasonable model for predicting the action of potential lead compounds against hK toxicity.

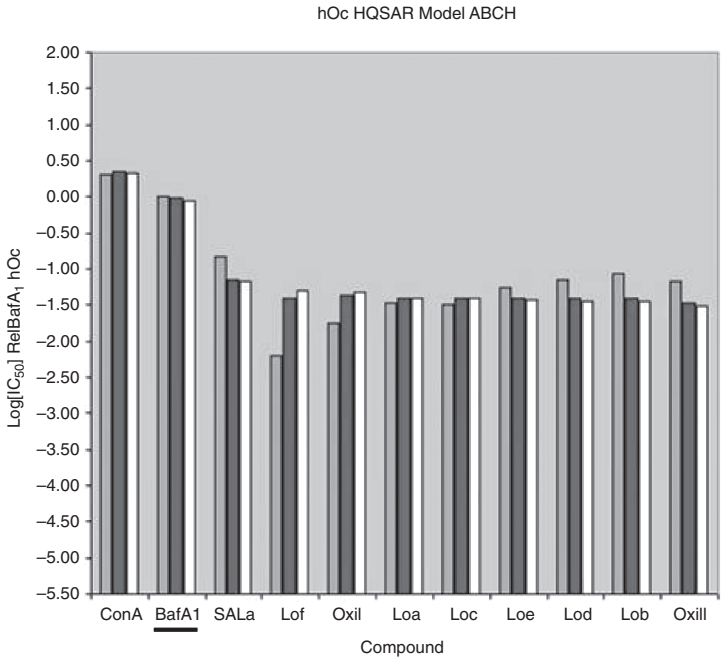


**FIGURE 9** Selection of HQSAR results III. Actual data (**light gray bars**) $\log IC_{50}$ hOc of a compound measured relative to bafilomycin $A_1$ V-ATPase phosphate assay (**BafA1**) from original data [41], versus predicted according to HQSAR (**dark gray bars**), versus the cross-validated HQSAR prediction (**white bars**). Despite the small size of this group, the HQSAR model was of high quality and may provide an indication of the activity of the potential lead compounds on hOc.

**TABLE 2 HQSAR Modelsa[a]**

| Parameter | Model | ApF | $q^2$ | STD_ERR | CVSTD_ERR | Length | Components | MCS | Pred-$r^2$ |
|---|---|---|---|---|---|---|---|---|---|
| hOc | ABC | 4 to 7 | 0.672 | 0.384 | 0.437 | 401 | 1 | N | D2S |
| hOc | ABCH | 4 to 7 | 0.708 | 0.365 | 0.412 | 401 | 1 | N | D2S |
| hOc | ACDo | 4 to 7 | 0.606 | 0.413 | 0.478 | 401 | 1 | N | D2S |
| hOc | ACHDo | 4 to 7 | 0.645 | 0.395 | 0.454 | 401 | 1 | N | D2S |
| hK | ABC | 4 to 7 | 0.857 | 0.516 | 0.681 | 199 | 2 | N | D2S |
| hK | ABCH | 4 to 7 | 0.831 | 0.472 | 0.766 | 97 | 3 | N | D2S |
| hK | ACDo | 4 to 7 | 0.890 | 0.469 | 0.598 | 61 | 2 | N | D2S |
| hK | ACHDo | 4 to 7 | 0.968 | 0.189 | 0.295 | 59 | 3 | N | D2S |
| cOc/bCG | ABC | 5 to 8 | 0.819 | 0.148 | 0.279 | 59 | 5 | N | 0.7371 |
| cOc/bCG | ABCH | 5 to 8 | 0.766 | 0.130 | 0.323 | 83 | 6 | N | 0.8109 |
| cOc/bCG | ABCCh | 5 to 8 | 0.817 | 0.124 | 0.286 | 97 | 6 | N | 0.8219 |
| cOc/bCG | ABCHCh | 5 to 8 | 0.719 | 0.115 | 0.355 | 353 | 6 | N | 0.8204 |
| cOc/bCG | ACChDo | 5 to 8 | 0.775 | 0.121 | 0.311 | 83 | 5 | N | 0.8829 |
| cOc/bCG | ACDo | 5 to 8 | 0.766 | 0.158 | 0.317 | 97 | 5 | N | 0.8335 |
| cOc | ABC | 5 to 8 | 0.413 | 0.572 | 0.904 | 53 | 5 | N | D2S |
| cOc | ABCCh | 5 to 8 | 0.449 | 0.370 | 0.890 | 199 | 6 | N | D2S |
| cOc | ACChDo | 5 to 8 | 0.490 | 0.381 | 0.857 | 353 | 6 | N | 0.3572 |
| cOc | ACDo | 5 to 8 | 0.461 | 0.452 | 0.881 | 61 | 6 | N | D2S |
| bCG | ABC | 4 to 7 | 0.577 | 0.634 | 0.895 | 59 | 5 | N | D2S |
| bCG | ABCH | 4 to 7 | 0.525 | 0.650 | 0.948 | 401 | 5 | N | D2S |
| bCG | ABCCh | 4 to 7 | 0.589 | 0.563 | 0.882 | 199 | 5 | N | 0.3111 |
| bCG | ACChDo | 4 to 7 | 0.633 | 0.360 | 0.849 | 257 | 6 | N | 0.5615 |
| bCG | ACDo | 5 to 8 | 0.568 | 0.513 | 0.921 | 307 | 6 | N | D2S |

[a]The HQSAR model accuracy was assessed by linear regression. The predictive power of the HQSAR model was calculated for each compound dataset (Pred-$r^2$), where some compounds were omitted in the HQSAR model construction. D2S = Dataset too Small for HQSAR model validation. An $r^2 > 0.35$–$0.4$ is good, with a value of 1.0 being perfect. The Parameter column is the dataset used and constitutes the dependent variable (i.e., $IC_{50}$ relative to bafilomycin $A_1$), the Model column specifies the flags used. The Atoms per Fragment (ApF) column indicates the optimum number of atoms per fragment to generate the best $q^2$ value. STD_ERR = Standard Error, and CVSTD_ERR = Cross-Validated Standard Error. Length is the optimal holographic length; in each case all available hologram lengths were used (53, 59, 61, 71, 83, 97, 151, 199, 257, 307, 353, 401). Components are the fragments that were calculated to be potentially of greatest interest to modulate activity of the ligand (the larger the number, the better and more informative the model). The maximal common structure (MCS) was only effective when comparing the plecomacrolides.

over bovine chromaffin granules). The cOc versus bCG compares the selectivity ratio of bafilomycin $A_1$ against the two biological targets (computed as the difference of $\log_{10}$ values), and then subtraction of the result from itself. The value is 0 because

$$\text{bafilomycin } A_1 \text{ selectivity (on cOc/bCG)} - \text{bafilomycin } A_1 \text{ selectivity (on cOc/bCG)} = 0$$

For compounds with greater selectivity toward cOc, the value will be >0 (*positive*) and those with less selectivity to cOc, the value will be <0 (*negative*).

The data (Fig. 7) classically demonstrates a good quality HQSAR model (relative to bafilomycin $A_1$); Indole 3 (**I3**) was the most selective experimentally derived compound for cOc versus bCG (38-fold). The predicted value from the HQSAR model also indicated **I3** as the most selective because it had the greatest magnitude. Likewise, the most nonselective compound for cOc over bCG was concanamycin derivative 4 (**4**), which had correspondingly the greatest negative magnitude. The cross-validated values indicate the predicted actvity of omitted compounds and can improve the quality of the model (by making $q^2$ closer to 1, and standard deviation closer to 0), but in these results the cross-validation impaired the quality. The quality of HQSAR models can be assessed directly by leaving out bioactive molecules that have known $IC_{50}$ data. It is advisable to select molecules over the majority of the measured data range to ensure a fair test. An HQSAR model is constructed as usual with these selected molecules omitted from the training set. The HQSAR model was then used to evaluate the predicted $IC_{50}$ of these omitted molecules. Plotting the known $IC_{50}$ versus the predicted $IC_{50}$ values should produce a graph of $y = x$, with an ideal graph having the points plotted close to the $y = x$ line. Applying linear regression evaluates the quality of the predicted activities against their known values and hence is a direct measure of the quality of an HQSAR model(s). A linear regression score ($r^2$) close to 1 is ideal with 0 being the worst possible model. If the predicted $r^2$ value is close to 1, it adds confidence to the HQSAR model (Table 2), although datasets can be too small to apply such a technique and one must rely on the HQSAR model alone (Fig. 10). When evaluating proposed new chemical inhibitors (Figs. 11–13), researchers prefer a predictive $r^2$ value of around 0.35–0.40 (or greater) when considering candidates for chemical synthesis or chemical modification.

HQSAR has been used for investigations into several other ligand classes [46–52], but while computational output is exciting and valid, links with *wet* experiments and subsequent success are pressing issues and hence such techniques often evade academic interest.

## 6.2 QSAR

In the last ten years, QSAR has become an important tool that is used by nearly every pharmaceutical, agrochemical, and biotechnology company to increase the efficiency of lead discovery. The value of the QSAR approach is that it may be used either in the absence of detailed receptor site knowledge (i.e., binding site structural information) or in conjunction with such information if it is available. A QSAR model is a multivariate mathematical relationship between a set of physicochemical

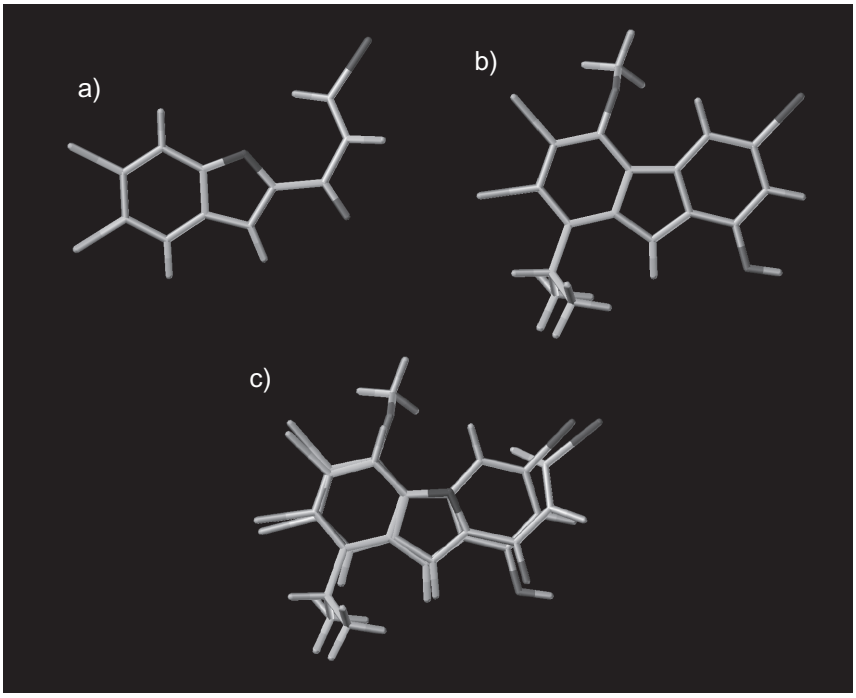


**FIGURE 10** Two small molecules that at first glance look very different (**a** and **b**). When superimposed (**c**), the fused ring structure of the second molecule (**b**) does not appear too different from the nonfused ring structure (**a**). Only by careful alignment by eye can overlaps such as these be accomplished. Slowly, computers are improving in performing these tasks.

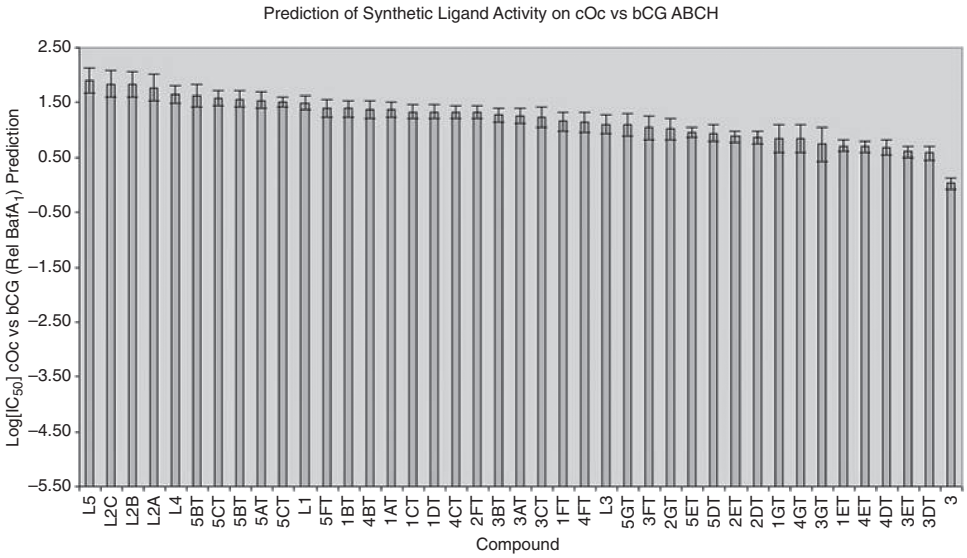


**FIGURE 11** Application of HQSAR model predictions for *in silico* ligand database assessment I. Evaluation of the database of proposed synthetic ligands for selective action on cOc over bCG V-ATPase. Database compounds of significant interest lie to the left of the graph.

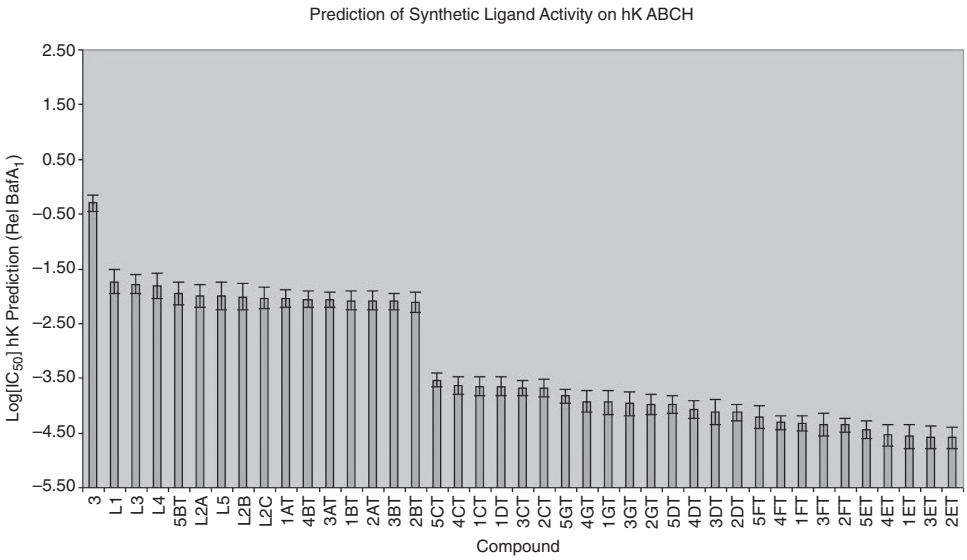


**FIGURE 12** Application of HQSAR model predictions for *in silico* ligand database assessment II. Evaluation of the database of proposed synthetic ligands sorted by $IC_{50}$ for hK V-ATPase phosphate assay. Database compounds of significant interest lie to the right of the graph.

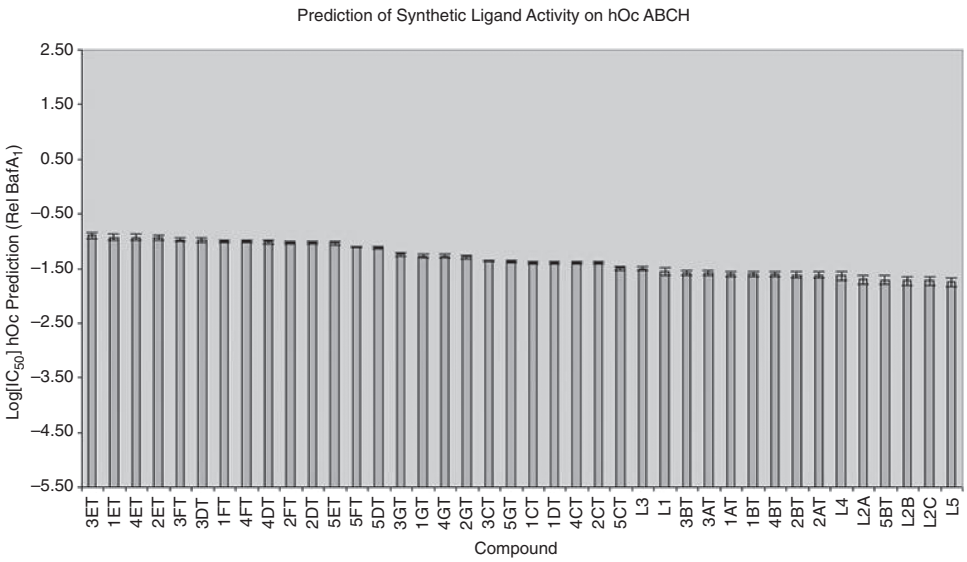


**FIGURE 13** Application of HQSAR model predictions for *in silico* ligand database assessment III. Evaluation of the database of proposed synthetic ligands sorted by $IC_{50}$ for hOc V-ATPase phosphate assay. Database compounds of significant interest lie to the left of the graph.

properties (*descriptors*) and a property of the system being studied, such as the biological activity, solubility, or mechanical behavior. QSARs correlate with cogeneric series of compounds, affinities of ligands to their binding sites, inhibition constants, rate constants, and other biological activities, either with certain structural features (e.g., Free Wilson analysis) or with atomic, group, or molecular properties, such as lipophilicity, polarizability, and electronic and steric properties (e.g., Hansch analysis). QSAR models have proved to be reliable tools for speeding up lead discovery [53–55] and have had an important place in molecular informatics throughout the world for at least two decades.

QSAR, like HQSAR, is based on measuring molecular similarity, but in the case of "classical" QSAR that similarity is based on characteristics, such as charge distribution and hydrophobicity, that derive from but are clearly secondary to molecular structure. While it is possible to rely exclusively on descriptors derived from experimental data, it is often the case that the molecular descriptors used in a QSAR study are directly computed from chemical structure; for example, a number of accurate algorithms are available to estimate log $P$ from molecular structure [56], and likewise the electrostatic properties of molecules can be calculated using *ab initio* quantum chemical methods [57–62]. So, even if molecular structure is not immediately related to function by a QSAR model, it lies close in the background of these studies. In fact, this has become increasingly the case as the chemical libraries being used have increased in size, and it has become less likely that experimental data will be uniformly available for all the compounds being considered.

The elements of a QSAR model are the training data, consisting of selected physicochemical properties, a method for generating a predictive regression model from the training data, and a set of validation data (which are usually removed from the initial training set at the beginning). A variety of mathematical techniques are available to generate predictive models, which map a set of values for the molecular descriptors to a predicted activity. The simplest is classical multivariate linear regression, but this approach becomes less satisfactory as the number of descriptors increases, and it becomes likely that some of the descriptors will be highly correlated. More satisfactory approaches include principal components analysis (PCA) [63] and partial least squares (PLS) [64], which despite some differences in mathematical approach have the same function of automatically generating "aggregate" variables that are linear combinations of descriptors and building regression models based on these. Models constructed using PCA or PLS are less complex and more robust than those based on straightforward regression. They provide the added benefit of highlighting just those descriptors that are important for explaining the activity to be predicted, and effectively dropping those that are uncorrelated with the activity to be explained. Cross-validation is critical, especially when many descriptors are in use, and is performed just as was described in the context of HQSAR (Section 6.1).

The goal of QSAR is to derive a function that relates to biological activity with some parameter(s) describing a feature of the molecule. Analyzing the correlation between biological activity and molecular parameters for a series of molecules that have already been tested forms the QSAR relationship. The concept is based on the assumption that the difference in the physical and chemical properties of molecules, whether experimentally measured or computed, accounts for the difference in their observed biological or chemical properties. Thus, in general, the QSAR method

deals with identifying and describing important structural features of molecules that are relevant to explaining variations in biological or chemical properties. The QSAR indicates the descriptors that are most statistically significant in determining the property, and studies can be focused on the molecular characteristics that those descriptors represent. QSARs thus help to make maximum use of data, whether that data is from experiment, simulation, or a database search.

QSAR is often carried out in the context of a *congeneric series*, a collection of molecules that share a common framework, but with significant variation of attached functional groups. Such a series comprises "variations on a theme," and it is this scenario that is most likely to lead to a robust, useful predictive model. While this might be seen as a drawback for classical QSAR, in that only molecules of a specified class are considered, the approach is very much compatible with most drug design methodologies where one begins with a framework that is amenable to variation using well-understood synthetic routes. The goal in this case is to develop a QSAR that rationalizes changes in activity with respect to structure, and to use this as a guide in proposing additional modifications that will further improve activity.

The "series" (a library of molecules with a shared characteristic) will consist of a common molecular framework with variance only in physicochemical descriptors such as hydrophobic constants, electronic parameters, or individual atoms. These physicochemical adaptations explain why individual molecules in a series have different biological activities. Relationships between activity and physicochemical characteristics can then be postulated, and the postulated relationship can be tested by generation of new compounds with predicted properties. The methods used to establish the equation that best describes the relationship between the property and the descriptors include regression techniques, PCA, and genetic algorithms. The expressed QSAR takes the form of either a search query or a predictive model, which can then be used to select new molecules with the specified activity from a database, or to predict the activity of individual molecules of interest. QSAR thus provides invaluable knowledge of which interactions are important to activity. This understanding provides the basis to formulate new active compounds that possess better overall therapeutic profiles, for example, compounds with increased functionality or that are more orally active. In fact, any biological or chemical activity that can be measured or derived from measurements can potentially be used as input for QSAR.

It should also be remembered that *in vitro* activity data only produce a QSAR that selects molecules that satisfy the *tested biological assay*, and do not necessarily indicate *in vivo* activity. The quality of the assay is therefore important. It has become increasingly important to conduct QSAR using more than one measured activity variable, such as adding a bioavailability measure to the activity data. The addition of *in vivo* data is difficult, as often many variables may exist that affect activity (primarily metabolism and excretion that are not tailored for), but this should be considered as an important development of QSAR.

### 6.3 Superimposition

One of the most compelling similarity techniques is not automated, can be time consuming, but is always done to get the point across clearly and effectively to a

reader or critic. Usually after several techniques have been employed to narrow down the best matches to a particular query molecule, potential candidates can be compared for similarity by 2D superimposition. The coordinates of one molecule (the query) is taken and mapped onto another molecule (the candidate) to assess how similar the two molecules are. However, the drawback is that so far only we humans can deduce where to appropriately overlay them (Fig. 10). Therefore, it is hard to implement superimposition until there is a short list of molecules: a list of thousands is not appropriate for evaluation by this method. Ultimately, this technique convinces medicinal chemists that the candidate molecule is indeed similar to another, and they will commence synthesis (providing there are not too many chiral centers, *E/Z* isomers, or complex heterocyclic ring structures).

### 6.4 Program Suites

UNITY is a program suite offered by Tripos [14], which contains a combination of techniques and methods to represent small molecules and identify potential molecules for biological trials via similarity. Another such suite available for purchase is CATALYST [65] offered by the commercial vendor Accelrys [66].

***Identifying Molecules Via Behavioral Similarities*** The principles governing the biological activity of many chemical entities used as drugs are usually due to particular groups and core structure within the molecule. These physicochemical descriptors are usually unique to a particular class of molecule. For instance, a hydrogen bond donor (or acceptor) could interact favorably with the biological target and be crucial for activity; often coordinated water molecules are critical in such cases (e.g., estrogens 1ERR [67]). Maximizing contact between the molecule and the target via van der Waals interactions is the driving force for ligand binding (i.e., a good fit), releasing water used in solvation and increasing the entropy of the system. To employ UNITY, the user needs strong biological understanding of the active ligands known and, if possible, should have an available protein structure.

Before searching large databases, it is imperative to design a robust representation of the parts of the molecule responsible for the drug class activity. This simplified representation is called a *pharmacophore*; it incorporates the features of as many drugs within the class as possible (deemed crucial for activity). If the design is of good quality, the pharmacophore will return the molecules known to be active and thus validate itself. Once the pharmacophore is designed suitably and holds well by returning other molecules within the drug class, the next step can be taken.

Large databases of molecules are available at cost and also without charge (Section 5.2) and are an excellent resource to find if there are compounds already similar to the pharmacophore. The goal is to find subtly different molecules that contain the groups and structural features essential for activity. Even if the molecule is not ideal, it may be available and amenable to medicinal chemistry.

In UNITY the procured molecular databases must be processed into a recognizable format for use by the program and often requires expertise. One must be exceptionally careful when building the database and when relying on prepared databases that the compounds are indeed represented correctly. CORINA [16] and STERGEN [17] are among the best tools for this particular job, but naming compounds uniquely is the most imperative and difficult task to accomplish. On con-

struction of UNITY databases, unique names for each molecule are responsible for most of the observed errors reported back. There are more sinister examples of errors such as the interpretation of molecules by SYBYL itself. One needs to be very vigilant with molecules having double bonds and conjugated aromatic and cyclized compounds. It is often the case that subtle changes can occur and the user is oblivious to these mistakes. A particularly good example is Raloxifene from the WDI database being interpreted incorrectly. Many PDB files contain mistakes and can also be interpreted incorrectly. When the database in constructed, it is important to note that 2D fingerprints and 2D macroscreens can be included in the database build. Often incorporating this into the UNITY database design increases the build time dramatically; but on the upside, all the searches will be considerably faster and nonmatching molecules are rapidly evaluated in a prescreen first.

These shortcomings aside, UNITY is still a powerful and reproducible means to search for similar molecules with many extra features, including the possibility of using key amino acids interacting with the ligand for structure-based drug design. What is sadly lacking is a score of how well the hits from the database fit the pharmacophore to describe the quality of the hit itself. This is rectified using an inbuilt feature to obtain Tanimoto coefficients for the returned hits: pitching these molecules against the best known drug molecules and implementing a cutoff to similarity.

***Designing New ACE Inhibitors*** Angiotensin converting enzyme (ACE) inhibitors are an excellent means to demonstrate pharmacophore design, not to mention their significant economic and medicinal importance. Heart disease affects approximately 65+ million Americans. The costs incurred via healthcare bills, lost revenue, and sick days are a serious nationwide medical and financial concern. The prime contributor to heart disease is hypertension (high blood pressure); 95% of cases are diagnosed as "essential hypertension." Blood pressure higher than 140/90 mmHg is considered seriously hypertensive and requires treatment.

ACE inhibitors work by preventing the final cleavage step of the rennin–angiotensin system; at this step angiotensin I is converted to angiotensin II by the enzyme ACE. Angiotensin II is the most potent endogenous vasoconstrictor known and increases peripheral reistance to blood flow by reducing the bore of arterioles. Inhibition of ACE would logically follow with a reduction in blood pressure and reduce the strain on the heart and circulatory system.

Recently, the crystal structure of lisinopril (Zestril by AstraZeneca) has been solved (PDB file 1O86) [68] and is considerably useful since potential molecules can be evaluated via docking into the active site of the receptor. Several ACE inhibitors combined with known $IC_{50}$ data are available [69], as are pharmacophore designs [70, 71]. To highlight the use of such information and data, we demonstrate how UNITY can be used by incorporating pharmacophore designs for database screening. Our example is somewhat simpler than that presented in the literature for clarity and ease of understanding.

The pharmacophore designed is more simplistic but serves an important purpose in orienting the mind to appreciate that pharmacophores do not need to be overly complex to be effective. As each feature is added and the pharmacophore is built up, it adds further complexity to the screening process, unlike a macroscreen or

fingerprint comparison that is rapid; a spatial or distance constraint between atoms requires a lot more computational time. Hence, a pharmacophore with one defined feature is a 1-point pharmacophore and can be identified very rapidly. Two defined features constitute a 2-point pharmacophore, and so on; typically, a pharmacophore has between 4 and 6 points, with some being very complex indeed [71]. As more points are added, the complexity of the calculations increases exponentially, so when screening, simple pharmacophores followed by a more complex one will save time in investigations. To screen a library of approximately 6 million molecules, it took approximately two weeks (in computer time) to create and screen the UNITY databases with incorporation of 2D macroscreens via fingerprints. This was six weeks in real time to manipulate and screen and modify the residual hit list data. Figure 14 shows the simple pharmacophore used incorporating the fragments conserved between all the ACE inhibitors [69], the three-carbon chain, and nitrogen in the five-membered ring. The features of this 2-point pharmacophore (Fig. 14a) were rapidly identified by the 2D Macro function and the matches were easily recognized by 1D comparisons. The 3-point pharmacophore (Fig. 14b), however, includes distance and spatial constraints, consequently requiring a lot more computational time. Thus, in screening a large library one would almost certainly conduct the search using the 2-point pharmacophore followed by the second. When testing the pharmacophores sequentially, all but one of the 23 known ACE inhibitors were returned, thus validating the pharmacophore design. The pharmacophores were used to screen an NCI database containing all possible stereoisomers of each entry and the residual dataset was organized using the Tanimoto coefficient. The top 4000 hits against

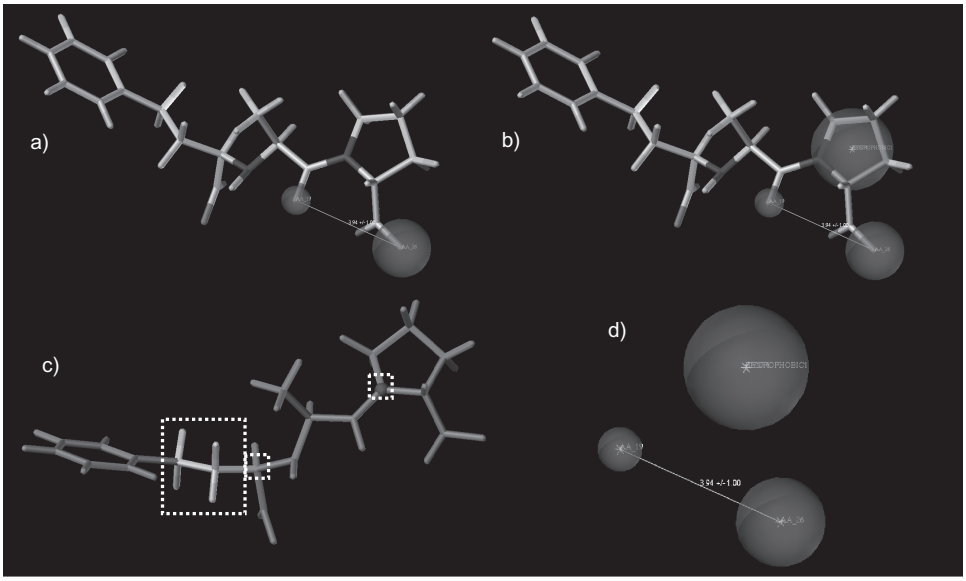


**FIGURE 14** Three pharmacophores were used in a similarity search: (**a**) 2 points + distant constraint, (**b**) 3 points + distant constraint, (**c**) fragment-based pharmacophore with four hydrogens and 3 carbons (large box) and one nitrogen (small box); (**d**) what the computer sees for (**b**).

lisinopril were docked into the PDB file 1O86 with the GOLD algorithm [72]. A comprehensive study of this investigation is currently under way.

***UNITY Summary*** UNITY demonstrates that robust design of a molecular template as a pharmacophore can prove insightful for identifying similar molecules. Although investment is required in software and expertise, it is not a particularly difficult technique to become adept with. Probably the biggest concern is maintenance and networking issues requiring specialized staff capable of system administration. This example merely scratches the surface of what UNITY can do: it is most relevant in identifying similar molecules.

### 6.5 Comparative Molecular Field Analysis (CoMFA) and Related Approaches

Comparative molecular field analysis (CoMFA) [73, 74], a proprietary method developed and marketed by Tripos, Inc. [14], is one of the most popular approaches to constructing quantitative models that link structure and activity [75–78]. Unlike conventional QSAR approaches, which use descriptors that describe the entire molecule (dipole moment, log *P*, topological indices, etc.), CoMFA is a 3D method that correlates a measured activity (e.g., log $IC_{50}$) with variations in electrostatic and steric properties for an aligned series of molecules. Since CoMFA identifies regions of the aligned structures that are important for determining activity, it allows the synthetic chemist to focus on sites of an exisiting molecular framework where variations are likely to be most productive. In favorable circumstances, a CoMFA can be thought of as an inverse model of the target receptor, since positive variations in activity require changes in the steric and electrostatic properties of the ligand that are compatible with (complementary to) those of the host receptor site. A CoMFA model can be generated to rationalize the activities of a series of ligands already characterized, and can then be used in predictive mode to estimate the activities of new molecules being proposed. CoMFA is one of a number of techniques that implicitly measure molecular shape by superimposing a regular grid on a collection of molecules.

The prerequisite for CoMFA is an aligned training set of ligands with known activities. This condition is most easily met by a congeneric series, since the shared molecular scaffold provides an immediate means to align the members of the set. In situations where a common framework is not available, CoMFA still requires that the molecules being analyzed can be superimposed in a meaningful way. One approach to this is *molecular field-fit*, where each molecule is positioned on a grid so as to maximize overlap with a preexisting set of field values (steric or electrostatic) computed on the grid vertices [79]. This approach can be used to construct a progressive alignment, where one or more seed molecules are used to generate a field on the grid, and successive molecules are aligned to the existing grid. As molecules are added to the alignment, their fields can be added to the vertices, potentially reinforcing the alignment field. Other methods that rely on a molecular field representation to carry out alignment include SEAL [80], FLUFF [81], and FIGO [82] (which combines a minimization procedure with 3D descriptors generated by the program GRID [83]).

A simple and effective approach for collections of structurally dissimilar molecules is to align them using their principal moments of inertia [84]—a procedure

that aligns molecules based on their distributions of mass. Yet another approach, applicable when a structure is available for the target protein receptor, is to dock the molecules into the common receptor site, thus aligning them [85–87]. At first this might seem an ideal approach, since the molecules will be aligned with biologically relevant conformations; however, owing to inevitable perturbations in the conformations and positions of the docked ligands, even when the molecules exhibit very similar docking modes, this procedure tends to produce CoMFA models that are inferior to those generated using a receptor-free alignment [88]; but on occasion improvements are evident [89].

Once an alignment has been constructed using an appropriate set of training molecules, the next step is to embed the aligned set in a cubical grid and to compute the interactions of a probe atom with the aligned molecules. The default probe is a carbon atom (with van der Waals parameters appropriate for a tetrahedral carbon) with a charge of +1. The probe is positioned at each vertex of the grid, and its steric (van der Waals 6–12) and electrostatic (Coulomb's law) interactions are computed separately with each of the molecules in the alignment. (It should be pointed out that the choice of probe is flexible, and a number of novel modifications have been introduced, including probes that measure hydrophobicity through the HINT potential [90–92], and use of an orientable water molecule as a probe, to measure local hydrogen-bonding potential [93].) The grid spacing is an important parameter in determining the performance of the resulting CoMFA model, and there is also some influence due to the position and orientation of the aligned molecules with respect to the grid. The importance of these factors can be evaluated by adjusting the grid spacing as well as the orientation of the aligned molecules. In addition, a grid "focusing" procedure is available, which can be used to automatically refine the grid by identifying those grid vertices on which changes in potential are highly correlated with activity [73, 74, 94]; those portions of the grid most important for creating a robust predictive model can then be refined by subdivision, effectively increasing the density of vertices in the regions of the alignment with greatest predictive importance.

The final step in the construction of a CoMFA model is the application of partial least squares (PLS) regression, a method already discussed, to generate a quantitative relationship between the field values computed on the grid and measured activity (although the technique SOMFA shows new and promising advantages [95]). In CoMFA, each vertex of the grid gives rise to at least one descriptor (a steric or electrostatic field value for each molecule in the alignment). Since these typically number in the hundreds or thousands, there is the danger that the PLS procedure will pick out vertices with potential values that correlate well with activity merely by chance. In fact, it is almost always the case that CoMFA will produce a model that predicts the activities of the training molecules with a very favorable $r^2$. To ensure that the model is robust and applicable outside the training set, it is an absolute requirement to apply cross-validation. This is usually accomplished using the leave-one-out approach discussed previously (although more sophisticated techniques are available). As a rule of thumb, CoMFA models with a cross-validated $r^2$ ($q^2$) of 0.5 or better are considered to be useful as predictive models. As with all established techniques, CoMFa is subject to evaluation [96] and improvement [95, 97].

### 6.6 Shape Signatures

Shape Signatures is a new method for detecting molecular similarity recently developed in our laboratory [98]. The fundamental motivation of the method is to generate compact descriptors that capture the features of molecular shape and polarity while avoiding details of chemical structure. By maintaining a focus on shape, our approach emphasizes those characteristics of molecules important for biological activity, making it much easier to scan chemical libraries for compounds that may be both bioactive and of novel structure.

The key feature of Shape Signatures is our approach for exploring and encoding shape and polarity information. We have adapted the method of ray tracing, widely used in computer animation and presentation graphics, as a probe of molecular geometry. To do this, we initiate a ray in the interior of a molecule, which is bounded by a triangulated representation of the solvent-accessible molecular surface (Fig. 15a), and allow the ray to propagate by the laws of optical reflection (Fig. 15b,c). Probability distributions are derived from the ray, and it is these distributions, stored as histograms, which we denote as Shape Signatures. The signatures are independent of the orientation of the molecule and can be compared very quickly using simple metrics below:

$$L_1^{1D} = \sum_i |H_i^1 - H_i^2|$$

$$L_1^{2D} = \sum_i \sum_j |H_{i,j}^1 - H_{i,j}^2|$$

The simplest signature is the probability distribution of ray-trace segment lengths, where a segment is the portion of the trace between two successive reflections. We call this a one-dimensional (1D) signature, as the domain of the distribution is one dimensional (Fig. 15e). Shape Signatures of higher dimension can easily be generated by combining ray-trace segment length with properties measured on the molecular surface; for example, by computing the joint probability distribution for observing a particular sum of segment lengths on either side of a reflection point, combined with the molecular electrostatic potential (MEP) measured at the reflection, we produce a "2D-MEP" signature with two-dimensional domain (length + electric potential), which encodes both shape and polarity information (Fig. 15f).

Although not as well developed as the ligand-based approach just described, it is also possible to apply the Shape Signatures method in *receptor-based* mode. Here, the ray-trace operation is carried out in the volume exterior to a protein receptor site, with Shape Signature histograms accumulated in the same way as in the ligand-centered approach. In this case, a match between the signatures for a potential ligand and the receptor indicates shape complementarity between the small molecule and the shape of the receptor-site volume. While harder to apply than the ligand-based method (primarily because of ambiguities in defining the binding-site volume), the receptor-based approach offers the exciting prospect of scanning chemical libraries for potential bioactive compounds on the basis of shape, without the bias of using specific chemical structure queries, and without the computational expense of database docking.

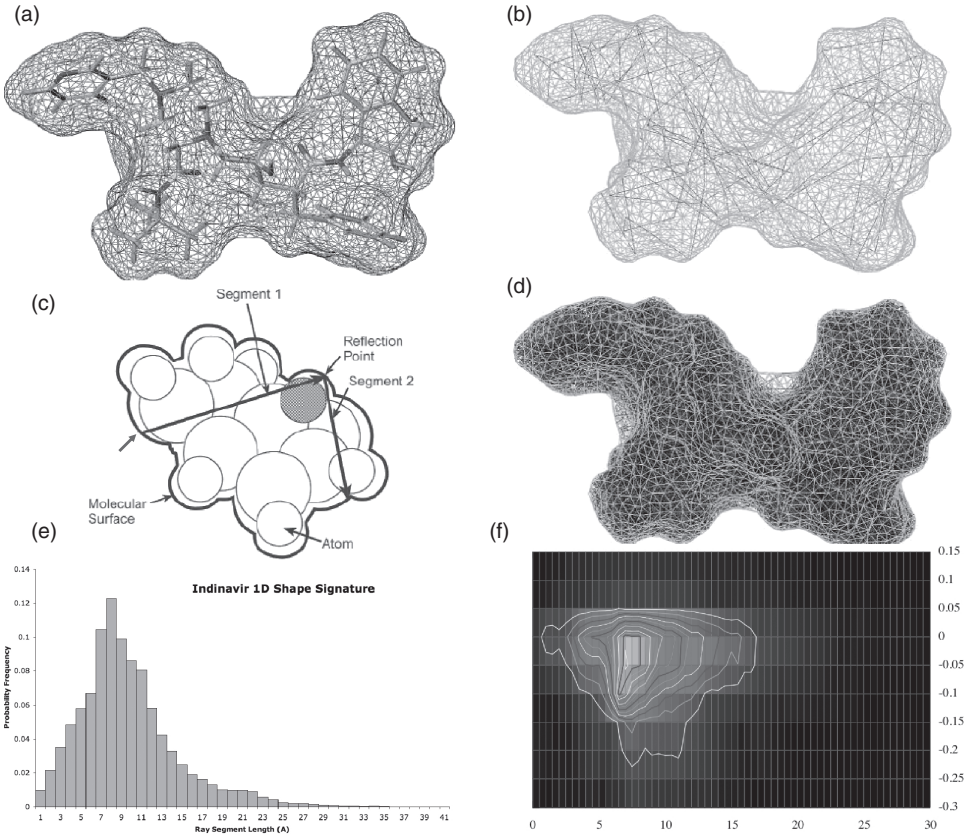


**FIGURE 15** Process of Shape Signature generation. (**a**) The structure of Indinavir enclosed in the triangulated solvent-accessible surface is generated using SMART [41]. (**b**) Propagation of a ray trace around the inside of the triangulated molecular surface. Propagation of the ray trace around the Indinavir molecule with (**c**) 100 ray-trace segments and (**d**) 10,000 ray-trace segments is shown. On completion of a Shape Signature, the generated trace (**d**) is illustrated by (**e**), a histogram denoting the probability distribution (ordinate) of ray-trace segment lengths (abscissa), a "1D Shape Signature." A Shape Signature trace defined by ray-trace segment lengths and a mean electrostatic potential (MEP) is shown by (**f**) a contour plot, a "2D Shape Signature." It is these data stored as text files that are compared when performing a database screen.

Although the method is in its infancy, Shape Signatures similarity comparisons have already been shown to be very effective for biological application [28–30, 98–100]. A current area of intensive research in our laboratory is to develop further ways of applying Shape Signatures for multiple conformers and stereoisomers, and to further develop the receptor-based strategy along with the existing ligand-based approach. It is clear that it is not unreasonable to expect to be able to carry out well over a thousand million comparisons a day on a single processor (Table 3)—numbers that are not easily attainable by other techniques. The method is trivial to

**TABLE 3 Comparison Between Speeds for Database Construction and Screening**

| Technique | Number of Molecules/Time Unit | | | | | CPU + RAM Computer Specifications |
|---|---|---|---|---|---|---|
| | Minute | Hour | Day | Month | Year | |
| | | | *Database Construction* | | | |
| Shape Signatures (Ray Tracing) | | | | | | |
| | 5.6 | $3.3 \times 10^2$ | $8.0 \times 10^3$ | $2.4 \times 10^5$ | $2.9 \times 10^6$ | 1 × 3.5 GHz Intel Pentium 4, 2 GB |
| | $1.1 \times 10^4$ | $2.1 \times 10^4$ | $5.1 \times 10^5$ | $1.5 \times 10^7$ | $1.9 \times 10^8$ | 32 × Opteron 2.6 GHz, 0.5 GB |
| Raptor Model (Generation) | | | | | | |
| | 4.5 | $2.7 \times 10^2$ | $6.5 \times 10^3$ | $1.9 \times 10^5$ | $2.4 \times 10^6$ | 1 × 1.8 GHz Pentium 4 |
| | | | *Database Screening* | | | |
| Shape Signatures (Screening) | | | | | | |
| | $1 \times 10^6$ | $6.0 \times 10^7$ | $1.4 \times 10^9$ | $4.3 \times 10^{10}$ | $5.3 \times 10^{11}$ | 1 × 3.5 GHz Intel Pentium 4, 2 GB |
| | $6.4 \times 10^7$ | $3.8 \times 10^9$ | $9.2 \times 10^{10}$ | $2.8 \times 10^{12}$ | $3.4 \times 10^{13}$ | 32 × Opteron 2.6 GHz, 0.5 GB |
| ROCS (Rapid Overlay of Chemical Structures) | | | | | | |
| | $7.6 \times 10^2$ | $4.6 \times 10^5$ | $1.1 \times 10^7$ | $3.3 \times 10^8$ | $4.0 \times 10^9$ | 1 × 3.5 GHz Intel Pentium 4, 2 GB |
| UNITY (Pharmacophore) | | | | | | |
| 3pt | 81.6 | $4.9 \times 10^3$ | $1.2 \times 10^5$ | $3.5 \times 10^6$ | $4.3 \times 10^7$ | 1 × RG14000 MIPS |
| 4pt | 68.9 | $4.1 \times 10^3$ | $9.9 \times 10^4$ | $3.0 \times 10^6$ | $3.6 \times 10^7$ | SGI 1 × 500 mHz, |
| 5pt | 56.6 | $3.4 \times 10^3$ | $8.2 \times 10^4$ | $2.4 \times 10^6$ | $3.0 \times 10^7$ | 1 GB |
| Docking GOLD 2.2 | | | | | | |
| | 0.6 | 34.7 | $8.3 \times 10^2$ | $2.5 \times 10^4$ | $3.0 \times 10^5$ | 2 × 3.0 GHz Intel Xeon, 4 GB |
| Raptor (Molecule Evaluation) | | | | | | |
| | 8.0 | $4.8 \times 10^2$ | $1.2 \times 10$ | $3.5 \times 10^5$ | $4.2 \times 10^6$ | 1 × 1.8 GHz Pentium 4 |

parallelize: both database generation and comparison—hence submission of queries via the Internet—could be implemented and accomplished in the near future. The data returned dramatically reduces the initial database size, a vital part of any successful CADD strategy [101]. Furthermore, Shape Signatures also achieves the second important quality of CADD: enrichment of molecules from the same class as the query [28–30, 98, 99]. Initial successes with Shape Signatures and estrogenic molecules (Fig. 16) were supported with experimental findings [99]. A potential novel molecule involved in analgesia has been discovered using Shape Signatures and offers considerable benefits over current compounds in this class [100]. Many groups [102–104] have previously used similar techniques for virtual-spatial recognition, but they have not been used in biological applications (with the exception of Ankerst and colleagues [105]). The Shape Signatures method is distinct from these

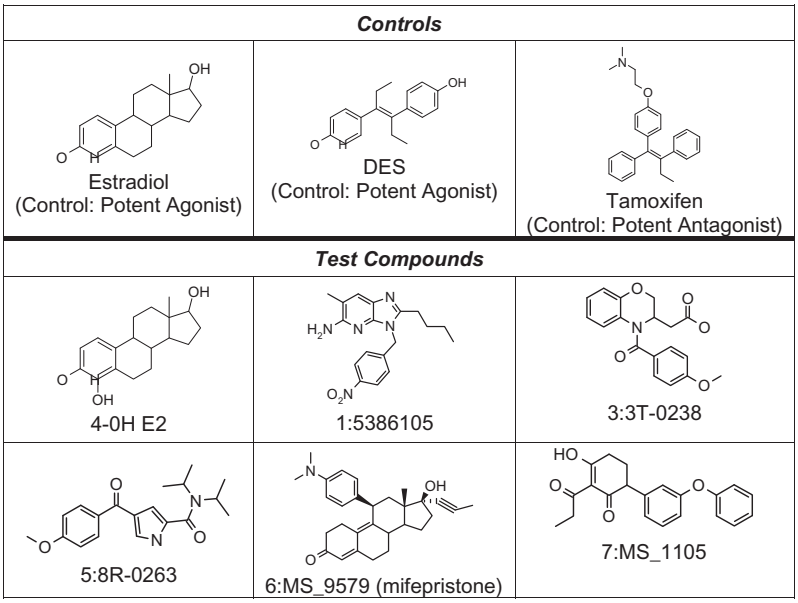


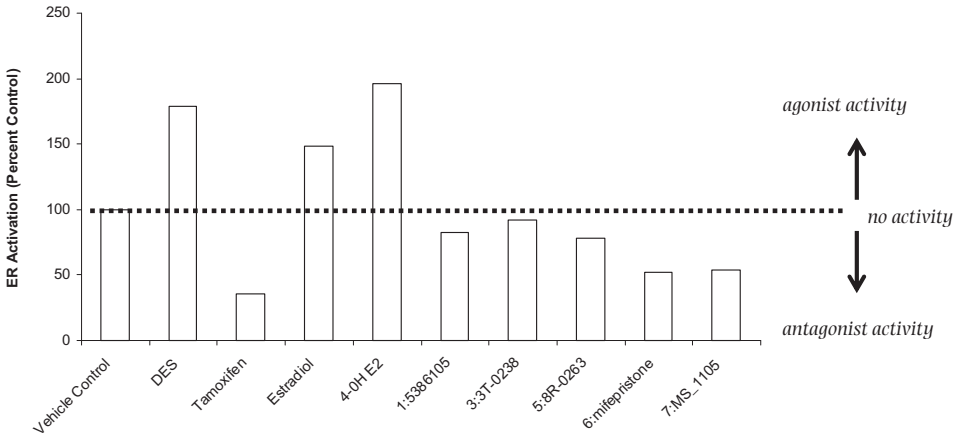


other previous attempts that use histogram comparisons for recognition of objects [102–104] or for lead discovery [105]. The Shape Signatures technique is rotationally invariant, meaning that it does not require, nor depend on, the orientation of the ligand or receptor site to obtain a reproducible histogram profile [103, 107].

As an example of the application of the Shape Signatures method, the query molecule WHI-P131 [106, 107], a tyrosine kinase inhibitor of interest as a therapeutic against several diseases (including leukemia and amyotrophic lateral sclerosis), is shown along with the top Shape Signatures hits located in the NCI database (Table 4) (were clarified by superimposition, see Section 6.3). Note the clear shape similarity between query and hits that nonetheless differ significantly in details of chemical structure. Finding these matches by other CADD methods would entail either generating a large set of structural queries based on the molecule of interest, or generating a pharmacophore model for the inhibitor. In either case, there would

**FIGURE 16** Examples of molecules selected by Shape Signatures demonstrating estrogenic activity on human estrogen receptor (ER) based on known controls. Three known estrogenic compounds (**top**)—estradiol, diethylbesterol, and tamoxifen—are all known to interact with human ER control. Using these as query molecules to screen an in-house database ($1.2 \times 10^6$ molecules), the returned molecules (**middle**) were tested via assay [48]. Taking each of the selected "hits" from the search in turn, a 25 μM sample was tested using the NR peptide ERα ELISA kit (Active Motif, Carlsbad, CA) according to manufacturer's instructions. 17β-estradiol and tamoxifen were included as part of the kit. Briefly, a precoated 96-well plate was supplied with an optimized peptide containing the consensus binding motif of ERα coactivator SRC-1. Each compound was incubated for 1 hour with MCF-7 nuclear extract and the coactivator peptide in each well. The ligand-activated ERα was first detected using primary antibody specific for ERα and further with HRP-conjugated secondary antibody. The ER competitive binding assays used a gel filtration displacement assay for estrogen receptor alpha (ERα) and was employed to assess competitive binding by selected ER antagonists among the hit compounds. Estrogen receptor binding assays were conducted in duplicate in 50 mM Tris-HCl, pH 7.5, 1 mM EDTA, 20% glycerol, and 1 mM DTT buffer. Radioligands that were used included [6,7-$^3$H]estradiol (specific activity 44 Ci/mmol, Amersham Biosciences). Binding assays were conducted on ice in a volume of 1 mL with 10 ng of purified full-length ERα and 25 nM $^3$H-estradiol in final concentration; 10 μM 17α-estradiol was used to define specific binding. Following a 1-hour incubation, assays were terminated by filtration through Whatman GF/B filters. Filters were soaked in Ecoscint liquid scintillation mixture (National Diagnostics, Somerville, NJ) and filter bound radioactivity was counted using a Beckman LS 1071 counter. (**Bottom**) One molecule was strongly agonistic (4-OH E2), while two were strongly antagonistic (6:mifepristone and 7:MS_1105). The other three molecules were only marginally antagonistic.

be the presumption of specialized knowledge to construct the queries actually used to scan the database, and a small omission in constructing that query could mean missing important hits. In contrast, the ligand-based Shape Signatures search is very easy to carry out; the investigator need only present the single compound of interest as the query.

There is one potential drawback of Shape Signatures that must be addressed up front. Because the method collapses a great deal of chemical space onto very compact descriptors, it is inevitable that a Shape Signatures search will turn up false

**TABLE 4 Shape Signatures Search of the NCI Database with WHI-P131**

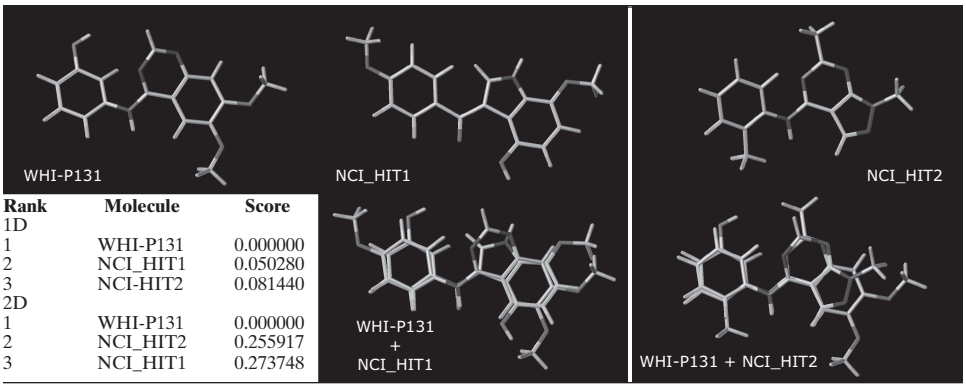


| Rank | Molecule | Score |
|---|---|---|
| 1D | | |
| 1 | WHI-P131 | 0.000000 |
| 2 | NCI_HIT1 | 0.050280 |
| 3 | NCI-HIT2 | 0.081440 |
| 2D | | |
| 1 | WHI-P131 | 0.000000 |
| 2 | NCI_HIT2 | 0.255917 |
| 3 | NCI_HIT1 | 0.273748 |

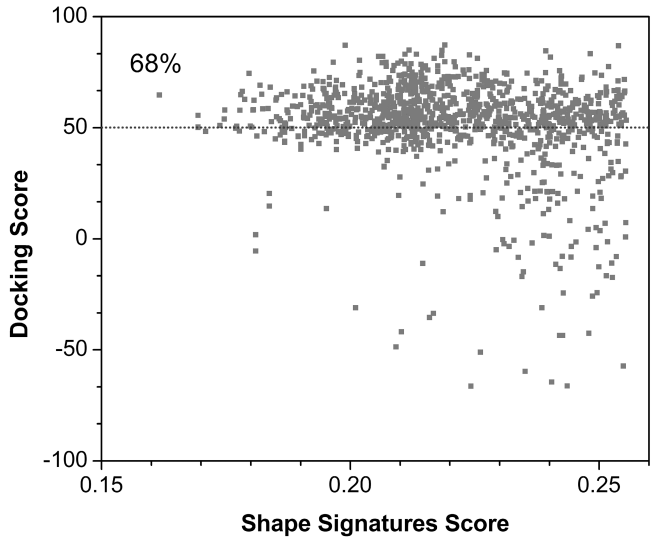


**FIGURE 17** Evaluation of Shape Signatures with the GOLD algorithm. The plot indicates Shape Signatures score (abscissa) versus GOLD (ordinate) with the dashed line at 50 indicating a good GOLD docking hit.

positives in the hit list. To test how significant this issue might be, an independent assessment of the quality of the hits produced by a Shape Signatures search with the ACE inhibitor Enalapril was established. First, docking of known ACE inhibitors that have measured $IC_{50}$ data [69] into ACE (PDB code 1O86) [68] was carried out using GOLD [72], producing a significant correlation between inhibitor $pIC_{50}$ and the GOLD score, with scores for the known actives ranging from 50 to 87. In the second phase of our study, we identified the top 250 hits produced by a Shape Signatures search for the single query, Enalapril, conducted on an in-house database of 423 drugs and the NCI database [10] of over 250,000 compounds. The hit compounds were extracted from the database and docked into the active site of ACE using GOLD. Not only did our search, conducted using a single query, immediately identify 11 of the 20 well-characterized ACE inhibitors, but the fitness scores by GOLD evaluation of all 250 hits gave >75 for 4%, and >50 for 70% of the hits (Fig. 17). Assuming that the correlation between the experimental $IC_{50}$ and the GOLD score holds, the Shape Signatures method is shown to readily identify known actives and interesting lead molecules, and with a modest proportion of false positives.

## 7 CONCLUSION

In this chapter we have merely scratched the surface of available methods to measure molecular similarity, focusing on methods with which we have experience and omitting any number of popular methods. In particular, there are a large and steadily growing body of 3D methods that are similar in spirit to CoMFA, but which nonetheless incorporate variations and refinements. In this context we must mention

approaches that use descriptors generated using GRID [83] and first-cousins to CoMFA, such as COMSIA [108]. We hope that this chapter has provided an introduction to the spirit of molecular similarity methods and will serve as a foundation for further exploration by the interested reader.

## ACKNOWLEDGMENTS

We kindly acknowledge Dr. Zhiwei Liu at the University of the Sciences in Philadelphia for her efforts in providing us with Fig. 16 for this chapter. We also extend gracious thanks to Ching Y. Wang and her supervisor, Professor William Welsh, at the University of Medicine and Dentistry of New Jersey for their efforts in constructing Fig. 15.

## REFERENCES

1. Center for Drug Evaluation and Research, 2004, Report to the Nation: Improving Public Health Through Human Drugs. (http://www.fda.gov/cder/reports/rtn/2004/rtn2004.htm).
2. Mullin R. Tufts report anticipates upturn; cites drugmakers' actions to accelerate drug development. *Chem Eng News* 2006;84:9.
3. Newman DJ, et al. Natural products as sources of new drugs over the period 1981–2002. *J Nat Prod.* 2003;66:1022–1037.
4. Laatsch H. AntiBase 2007: The Natural Product Identifier. Hoboken, NJ: John Wiley & Sons; 2007.
5. Pauli A. AmicBase 2005. Hoboken, NJ: John Wiley & Sons; 2005.
6. Thomson Scientific World Drug Index. http://www.scientific.thomson.com/products/wdi (accessed July 2006).
7. GlaxoSmithKline, Plc. http://www.gsk.com (accessed July 2006).
8. Pfizer, Inc. http://www.pfizer.com (accessed July 2006).
9. Irwin JJ, Shoichet BK. ZINC—a free database of commercially available compounds for virtual screening. *J Chem Inf Model* 2005;45:177–182. (http://zinc.docking.org).
10. NCI Database. http://cactus.nci.nih.gov/ncidb2/download.html (accessed July 2006).
11. ChemBank Database (1.1 million entries). http://chembank.broad.harvard.edu/.
12. Weininger D. SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. *J Chem Inf Comput Sci* 1988;28:31–36.
13. Ash S, et al. SYBYL Line Notation (SLN): a versatile language for chemical structure representation. *J Chem Inf Comput Sci* 1997;37:71–79.
14. Tripos, Inc. http://www.tripos.com (accessed July 2006).
15. MarvinSketch. http://www.chemaxon.com/marvin (accessed July 2006).
16. CORINA. http://www.mol-net.de/software/corina/index.html (accessed July 2006).
17. STERGEN. http://www.mol-net.de/software/stergen/index.html (accessed July 2006).
18. TAUTOMER. http://www.mol-net.de/software/tautomer/index.html (accessed July 2006).
19. Halgren TA. Merck molecular force field. III. Molecular geometries and vibrational frequencies for MMFF94. *J Comp Chem* 1995;17:553–586.

20. Gasteiger J, Marsili M. A new model for calculating atomic charges in molecules. *Tetrahedron Lett* 1978;34:3181–3184.
21. Gasteiger J, Saller H. Calculation of the charge distribution in conjugated systems by a quantification of the resonance concept. *Angew Chem Int Ed Engl* 1985;24:687–689.
22. Dalby A, et al. Description of several chemical-structure file formats used by computer-programs developed at Molecular Design Limited. *J Chem Inf Comput Sci*. 1992;32: 244–255.
23. Hammett LP. *Physical Organic Chemistry*, 1st and 2nd eds. New York: McGraw Hill; 1940, 1970. Hammond GS. A correlation of reaction rates. J Am Chem Soc 1955;77:334–338.
24. Hansch C, et al. *Substituent Constants for Correlation Analysis in Chemistry and Biology*. Hoboken, NJ: Wiley; 1979. Leo A, et al. Partition coefficients and their uses. Chem Rev 1971;71:525–554.
25. ROTATE. http://www.mol-net.de/software/rotate/index.html (accessed July 2006).
26. Example of energy minimization algorithm. http://www.chemsoc.org/exemplarchem/entries/pkirby/exemchem/mmmintro.html (accessed July 2006).
27. Weininger D, inventor; Daylight Chemical Information Systems, Inc, assignee. Method and apparatus for designing molecules with desired properties by evolving successive populations. US patent US5434796, 1995.
28. Zauhar RJ, et al. Shape Signatures: a new approach to ligand- and receptor-based molecular design. *J Med Chem*. 2003;46:5674–5690.
29. Nagarajan K, et al. Enrichment of ligands for the serotonin receptor using the Shape Signatures approach. *J Chem Inf Model* 2005;45:49–57.
30. Meek PJ, et al. Shape Signatures: speeding up computer aided drug discovery. *Drug Discov Today* 2006;11:895–904.
31. Nordling E, Homan E. Generalization of a targeted library design protocol: application to 5-HT7 receptor ligands. *J Chem Inf Comput Sci* 2004;44:2207–2215.
32. Milton RC, et al. Total chemical synthesis of a D-enzyme: the enantiomers of HIV-1 protease show reciprocal chiral substrate specificity. *Science* 1992;256:1403–1404. Errantum, *Science* 1992;257(Jul):147.
33. Thalidomide toxicity information. http://www.k-faktor.com/thalidomide (accessed July 2006).
34. CONCORD. http://www.tripos.com/index.php?family=modules,SimplePage,,,&page=sybyl_concord (accessed July 2006).
35. Barnard JM, et al. Use of Markush structure analysis techniques for descriptor generation and clustering of large combinatorial libraries. *J Mol Graph Model* 2000;18:452–463.
36. Molecular Operating Environment (MOE). http://www.chemcomp.com/software.htm (accessed July 2006).
37. James CA, Weininger D. *Daylight Theory Manual*. Daylight Chemical Information Systems, Inc; 1995.
38. Smith EG, et al. The Wiswisser line-formula chemical notation (WLN). Cheery Hill, NJ: Chemical Information Management; 1975.
39. MACCS-II. San Leandro, CA: MDL Limited; 1992.
40. Dröse S, et al. Semisynthetic derivatives of concanamycin A and C, as inhibitors of V- and P-type ATPases: structure–activity investigations and developments of photoaffinity probes. *Biochemistry* 2001;40:2816–2825.

41. Boyd MR, et al. Discovery of a novel antitumor benzolactone enamide class that selectively inhibits mammalian vacuolar-type ($H^+$)-atpases. *J Pharmacol Exp Ther* 2001;297:114–120.
42. Farina C, et al. Novel bone antiresorptive agents that selectively inhibit the osteoclast V-$H^+$-ATPase. *Farmaco* 2001;56:113–116.
43. Gagliardi S, et al. 5-(5,6-Dichloro-2-indolyl)-2-methoxy-2,4-pentadiamides: novel and selective inhibitors of the vacuolar $H^+$-atpase of osteoclasts with bone antiresorptive activity. *J Med Chem* 1998;41:1568–1573.
44. Gagliardi S, et al. Synthesis and structure–activity relationships of bafilomycin A1 derivatives as inhibitors of vacuolar $H^+$ATPase. *J Med Chem* 1998;41:1883–1893.
45. Nadler G, et al. (2*Z*,4*E*)-5-(5,6-dichloro-2-indolyl)-2-methoxy-*N*-(1,2,2,6,6- pentamethylpiperidin-4-yl)-2,4-pentadienamide, a novel, potent and selective inhibitor of the osteoclast V-ATPase. *Bioorg Med Chem Lett* 1998;8:3621–3626.
46. Chen D, et al. Holographic QSAR of selected esters. *Chemosphere* 2004;57:1739–1745.
47. Zhang H, et al. CoMFA, CoMSIA, and molecular hologram QSAR studies of novel neuronal nAChRs ligands—open ring analogues of 3-pyridyl ether. *J Chem Inf Model* 2005;45:440–448.
48. Zhu W, et al. QSAR analyses on ginkolides and their analogues using CoMFA, CoMSIA, and HQSAR. *Bioorg Med Chem* 2005;13:313–322.
49. Castilho MS, et al. Two- and three-dimensional quantitative structure–activity relationships for a series of purine nucleoside phosphorylase inhibitors. *Bioorg Med Chem* 2006;14:516–527.
50. Honorio KM, et al. Hologram quantitative structure–activity relationships for a series of farnesoid X receptor activators. *Bioorg Med Chem Lett* 2005;15:3119–3125.
51. Doddareddy MR, et al. Hologram quantitative structure–activity relationship studies on 5-HT6 antagonists. *Bioorg Med Chem* 2004;12:3815–3824.
52. Pungpo P, et al. Hologram quantitative structure–activity relationships investigations of non-nucleoside reverse transcriptase inhibitors. *Curr Med Chem* 2003;10:1661–1677.
53. Lohray BB, et al. 3D QSAR studies of *N*-4-arylacryloylpiperazin-1-yl-phenyl-oxazolidinones: a novel class of antibacterial agents. *Bioorg Med Chem Lett* 2006;16: 3817–3823.
54. Desai PV, et al. Identification of novel parasitic cysteine protease inhibitors by use of virtual screening. *J Med Chem* 2006;49:1576–1584.
55. Allan GM, et al. Modification of estrone at the 6, 16, and 17 positions: novel potent inhibitors of 17beta-hydroxysteroid dehydrogenase type 1. *J Med Chem* 2006;49: 1325–1345.
56. Eros D, et al. Reliability of log *P* predictions based on calculated molecular descriptors: a critical review. *Curr Med Chem* 2002;9:1819–1829.
57. Goller AH, et al. *In silico* prediction of buffer solubility based on quantum-mechanical and HQSAR- and topology-based descriptors. *J Chem Inf Model* 2006;46:648–658.
58. Ermondi G, et al. A combined *in silico* strategy to describe the variation of some 3D molecular properties of beta-cyclodextrin due to the formation of inclusion complexes. *J Mol Graph Model* 2006;25:296–303.
59. Clark M. Generalized fragment–substructure based property prediction method. *J Chem Inf Model* 2005;45:30–38.
60. Votano JR, et al. New predictors for several ADME/Tox properties: aqueous solubility, human oral absorption, and Ames genotoxicity using topological descriptors. *Mol Divers* 2004;8:379–391.

61. Holm R, Hoest J. Successful *in silico* predicting of intestinal lymphatic transfer. *Int J Pharm* 2004;272:189–193.
62. Lobell M, Sivarajah V. *In silico* prediction of aqueous solubility, human plasma protein binding and volume of distribution of compounds from calculated p$K_a$ and $A$ log $P$98 values. *Mol Divers* 2003;7:69–87.
63. Introduction of Principal Components Analysis (PCA). http://www.statsoft.com/textbook/stfacan.html (accessed July 2006).
64. Introducton of Partial Least Squares (PLS). http://www.statsoft.com/textbook/stpls.html (accessed July 2006).
65. CATALYST. http://www.accelrys.com/products/catalyst (accessed July 2006).
66. Accelrys, Inc. http://www.accelrys.com/products (accessed July 2006).
67. Brzozowski AM, et al. Molecular basis of agonism and antagonism in the oestrogen receptor. *Nature* 1997;389:753–758.
68. Natesh R, et al. Crystal structure of the human angiotensin-converting enzyme–lisinopril complex. *Nature* 2003;421:551–554.
69. Kamenska V, et al. The COREPA approach to lead generation: an application to ACE-inhibitors. *Eur J Med Chem* 1999;34:687–699.
70. Tzakos AG, Gerothanassis IP. Domain-selective ligand-binding modes and atomic level pharmacophore refinement in angiotensin I converting enzyme (ACE) inhibitors. *Chem Biochem* 2005;6:1089–1103.
71. Sutherland JJ, et al. Pruned receptor surface models and pharmacophores database searching. *J Med Chem* 2004;47:3777–3787.
72. Jones G, et al. Development and validation of a genetic algorithm for flexible docking. *J Mol Biol* 1997;267:727–748.
73. Cramer RD III, et al. Comparative molecular field analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins. *J Am Chem Soc* 1988;110:5959–5967.
74. CoMFa. http://www.tripos.com/index.php?family=modules,SimplePage,,,&page=sybyl_qsar_with_comfa
75. Sheng C, et al. Structure-based optimization of azole antifungal agents by CoMFA, CoMSIA, and molecular docking. *J Med Chem* 2006;49:2512–2525.
76. Menezes IR, et al. Three-dimensional models of non-steroidal ligands: a comparative molecular field analysis. *Steroids* 2006;71:417–428.
77. Demyttenaere-Kovatcheva A, et al. Identification of the structural requirements of the receptor-binding affinity of diphenolic azoles to estrogen receptors alpha and beta by three-dimensional quantitative structure–activity relationship and structure–activity relationship analysis. *J Med Chem* 2005;48:7628–7636.
78. Aguirre G, et al. New potent 5-nitrofuryl derivatives as inhibitors of *Trypanosoma cruzi* growth. 3D-QSAR (CoMFA) studies. *Eur J Med Chem* 2006;41:457–466.
79. Dixon S, et al. QMQSAR: utilization of a semiempirical probe potential in a field-based QSAR method. *J Comput Chem* 2005;26:23–34.
80. Kearsley SK, Smith GM. An alternative method for the alignment of molecular structures: maximizing electrostatic and steric overlap. *Tetrahedron Comput Methodol* 1990;3:615–633.
81. Korhonen SP, et al. Comparing the performance of FLUFF-BALL to SEAL-CoMFA with a large diverse estrogen data set: from relevant superpositions to solid predictions. *J Chem Inf Model* 2005;45:1874–1883.
82. Melani F, et al. Field interaction and geometrical overlap: a new simplex and experimental design based computational procedure for superposing small ligand molecules. *J Med Chem* 2003;46:1359–1371.

83. Goodford PJA. Computational procedure for determining energetically favorable binding sites on biologically important macromolecules. *J Med Chem* 1985;28:849–857. (GRID, version 19; Molecular Discovery Ltd, Mayfair, London, England, 2001.)
84. Collantes ER, et al. Use of moment of inertia in comparative molecular field analysis to model chromatographic retention of nonpolar solutes. *Anal Chem* 1996;68: 2038–2043.
85. Gallardo-Godoy A, et al. Sulfur-substituted alpha-alkyl phenethylamines as selective and reversible MAO-A inhibitors: biological activities, CoMFA analysis, and active site modeling. *J Med Chem* 2005;48:2407–2419.
86. Thaimattam R, et al. 3D-QSAR CoMFA, CoMSIA studies on substituted ureas as Raf-1 kinase inhibitors and its confirmation with structure-based studies. *Bioorg Med Chem* 2004;12:6415–6425.
87. Iskander MN, et al. Optimization of a pharmacophore model for 5-HT4 agonists using CoMFA and receptor based alignment. *Eur J Med Chem* 2006;41:16–26.
88. Kamath S, Buolamwini JK. Receptor-guided alignment-based comparative 3D-QSAR studies of benzylidene malonitrile tyrphostins as EGFR and HER-2 kinase inhibitors. *J Med Chem* 2003;46:4657–4668.
89. Datar PA, Coutinho EC. A CoMFA study of COX-2 inhibitors with receptor based alignment. *J Mol Graph Model* 2004;23:239–251.
90. Abraham DJ, Leo AJ. A program has recently become available that provides a functionality long missing from the molecular modeling world. HINT is a program that performs approximations of atom-based hydrophobicity based on the hydrophobic fragment work of Hansch and Leo. *Proteins* 1987;2:130–152.
91. Kellogg GE, et al. HINT—a new method of empirical hydrophobic field calculation for CoMFA. *J Comput Aided Mol Des* 1991;5:545–552.
92. Huey R, et al. Olson grid-based hydrogen bond potentials with improved directionality. *Lett Drug Des Discov* 2004;1:178–183.
93. Kim KH, et al. Use of the hydrogen bond potential function in a comparative molecular field analysis (CoMFA) on a set of benzodiazepines. *J. Comput Aided Mol Des* 1993;7:263–280.
94. Stuti G, et al. CoMFA and CoMSIA studies on a set of benzyl piperazines, piperadines, pyrazinopyridoindoles, pyrazinoisoquinolines and semi rigid analogs of diphenhydramine. *Med Chem Res* 2004;13:746–757.
95. Korhonen SP, et al. Improving the performance of SOMFA by use of standard multivariate methods. *SAR QSAR Environ Res* 2005;16:567–579.
96. Cavalli A, et al. Linking CoMFA and protein homology models of enzyme-inhibitor interactions: an application to non-steroidal aromatase inhibitors. *Bioorg Med Chem* 2000;8:2771–2780.
97. Moitessier N, et al. Combining pharmacophore search, automated docking, and molecular dynamics simulations as a novel strategy for flexible docking. Proof of concept: docking of arginine-glycine-aspartic acid-like compounds into the alphavbeta3 binding site. *J Med Chem* 2004;47:4178–4187.
98. Zauhar RJ, et al. Shape Signatures: a new approach to computer-aided ligand- and receptor-based drug design. *J Med Chem* 2003;46:5674–5690.
99. Wang CY, et al. Identification of previously unrecognized antiestrogenic chemicals using a novel virtual screening approach. *Chem Res Toxicol* 2006;19:1595–1601.
100. Zhang Q. et al. Discovery of novel triazole-based opioid receptor antagonists. *J Med Chem* 2006;49:4044–4047.
101. Zeman SP. Charting chemical space: finding new tools to explore biology. *Nature* 1–3. The 4th Horizon Symposium, Black Point Inn, Maine, USA, 20–22 May 2004.

102. Liu X, et al. *Proceedings of the 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (CVPR'03); 2003, pp 1–8.
103. Osada R, et al. Matching 3D models with shape distributions. *Shape Model Int* 2001;May:154–166.
104. Ohbuchi R, et al. Shape-similarity search of 3D models by using enhanced shape functions. *Int J Comput Appl Tech* 2005;23:70–85.
105. Ankerst M, et al. Nearest neighbor classification in 3D protein databases. In *Proceedings of the 7th Conference on Intelligent Systems for Molecular Biology* (ISMB'99); 1999, pp 34–43.
106. Amin HM, et al. Inhibition of JAK3 induces apoptosis and decreases anaplastic lymphoma kinase activity in anaplastic large cell lymphoma. *Oncogene* 2003;22: 5399–5407.
107. Jilek RJ, et al. Lead hopping method based on topomer similarity. *Chem Inf Comput Sci* 2004;44:1221–1227.
108. Klebe G. Comparative molecular similarity indices: CoMSIA. In *Kubinyi H, Folkers G, Martin YC, Eds. 3D QSAR in Drug Design*. London: Kluwer Academic Publishers; 1998, Vol 3, p 87.

# PROTEIN–PROTEIN INTERACTIONS

Kamaljit Kaur, Dipankar Das, and Mavanur R. Suresh
*University of Alberta, Edmonton, Alberta, Canada*

**Contents**

1 Introduction
2 Protein–Protein Interactions and Human Pathogenesis
  2.1 Oncogenesis
  2.2 Pathogen–Host Interaction
  2.3 Loss of Normal Protein–Protein Interaction
3 Screening of Protein–Protein Interaction Inhibitors
  3.1 Structure–Activity Relationship
  3.2 Genetic Screening Systems and Phage Display
  3.3 Yeast Two Hybrid System and Intracellular Antibodies
4 Inhibitors of Protein–Protein Interactions
  4.1 Peptide and Peptidomimetic Inhibitors
  4.2 Small-Molecule Inhibitors
  4.3 Molecules Containing Porphyrin or Peptidocalixarene Scaffolds
5 Conclusion
  References

## 1 INTRODUCTION

Living organisms are almost exclusively comprised of four classes of molecules, namely, proteins, nucleic acids, polysaccharides, and lipids. Of these, barring lipids, all other classes can be regarded as macromolecules that are built from a limited number of building blocks or monomers. In the case of proteins, such building blocks are amino acids. Proteins are formed by polymerization of essentially twenty

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

"standard" amino acids. Yet, the myriad of proteins and their diverse functions, ranging from basic metabolism to structural and reproductive functions, can be astounding and constitute the very basis of life on Earth. For instance, an *Escherichia coli* bacterium contains over 4000 different proteins participating in virtually every life sustaining function of the cell.

The Greek root of the word protein, *proteios*, meaning *of first importance*, identifies the paramount role of this class of macromolecules in eukaryotes. It is perhaps discernible that twenty amino acids can be combined in different manners to yield virtually innumerable proteins, with a variety of functions. It is, however, interesting to note that even a slight alteration of the amino acid sequence can significantly alter the structure and function of a protein molecule. A well known example is the modification of hemoglobin in sickle cell anemia. Normal hemoglobin (HbA) contains a Glu at the sixth position of each β-chain, which is replaced by Val in the sickle cell hemoglobin (HbS). This single substitution causes aggregation of HbS into stiff filaments, leading to the deformation of the red blood cells into elongated "sickled" shapes, and the consequent symptoms of sickle cell anemia.

Being macromolecules, proteins predominantly interact with other molecules, including other proteins, via weak long-range interactions. These are also referred to as nonbonded interactions and are of two types, van der Waals and electrostatic. As two interacting proteins approach very close to each other, stronger specific bonds can be formed between them. For instance, the strongest specific interaction between two proteins occurs in the form of a covalent bond in the so-called disulfide linkage. Hydrogen bonds, less strong than covalent bonds, are more common forms of specific interactions exhibited by proteins. The four common types of interactions observed are listed in Fig. 1. A second factor that modifies long-range interactions is the tertiary structure of proteins. Proteins or peptides with defined secondary structural elements like helices, strands, and coils fold into three-dimensional arrangements, which give them a tertiary structure. For example, an α-helix, a β-sheet, and a coil region in the three-dimensional solution structure of an antibacterial peptide are highlighted in Fig. 2. Sequences with less than 50 amino acids are generally considered as peptides and more than 50 amino acids are called proteins.

Protein–protein interactions are key to several biological pathways and thus are attractive targets for therapeutic intervention. Such approaches are frequently based on a sound assessment of the strong and weak interactions and the protein's secondary or tertiary structures. The modulation or disruption of protein–protein inter-

| Interactions | |
|---|---|
| Long-Range/Weak | Short-Range/Strong |
| Electrostatic (5 kcal/mol) | Covalent bonds (40 kcal/mol) |
| van der Waals (< 1 kcal/mol) | Hydrogen bonds (3-7 kcal/mol) |

**FIGURE 1** Long-range and short-range interactions between biomolecules. The energy values in parentheses are approximate.

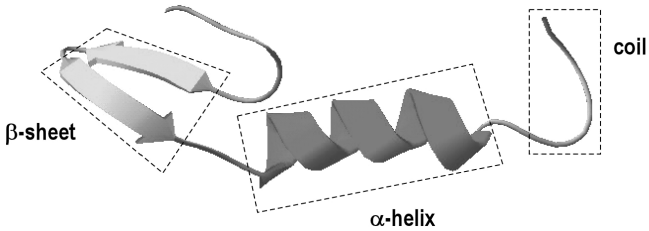


**FIGURE 2** Tertiary structure of an antibacterial peptide (leucocin A) displaying secondary structure elements, namely, α-helix, β-sheet, and coil. The figure was generated from the 1CW6 Protein Data Bank coordinates [1].

actions has been difficult owing to their large surface areas of interaction. For several extracellular protein complexes, antibodies and other proteins have been identified as successful antagonists. The main rationale for this being that large macromolecules can readily disrupt the interaction between two proteins. From a cursory glance, this appears to be an attractive therapeutic option. However, a closer inspection of the problem indicates that large proteins are generally not orally bioavailable or cell permeable. Thus, these cannot be particularly effective for targeting intracellular protein–protein interactions. In light of this, one might presume that small molecules will provide more attractive therapeutic intervention options. However, such small molecules have been less successful in this regard, since they cannot provide specific recognition needed for a large protein surface. Protein surfaces do not often present deep indentations for small molecules to bind, and affinity is achieved through summing up several weak interactions.

Recent studies report several protein complexes as targets for drug design, with some of these targets amenable to small molecule inhibition [2–7]. Here we review the interactions between some of the important protein–protein pairs, followed by the recent successes in developing peptides, peptidomimetics, or small organic molecules as inhibitors of these interactions. This field is still in its infancy as most of the compounds identified are still in preclinical stages. Recent developments made in this broad field that have pharmaceutical and clinical implications will be discussed in this chapter.

## 2 PROTEIN–PROTEIN INTERACTIONS AND HUMAN PATHOGENESIS

### 2.1 Oncogenesis

Interaction between specific regions in a protein has been found to be essential for all stages of development and homeostasis. Subsequently, many human diseases occur by either loss of essential protein–protein interaction or through the formation of a protein complex at an inappropriate time or location. Several such interactions have been found to be responsible for the onset of oncogenesis and have been well studied [4, 8]. In the following, interactions between some of the well known protein pairs that lead to oncogenesis are discussed.

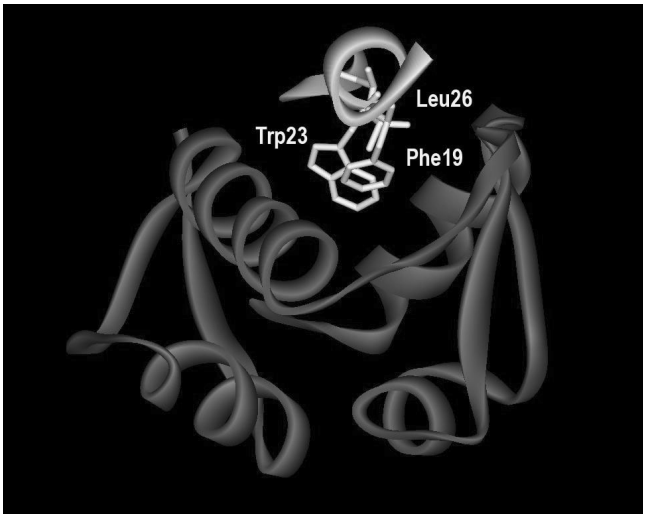


**FIGURE 3** The ribbon representation of the MDM2–p53 complex (PDB entry 1YCR [10]) displaying the hydrophobic cleft of MDM2 where the p53 peptide binds as an amphipathic α-helix. The hydrophobic side chains of Phe19, Trp23, and Leu26 of p53 inserting deep into the MDM2 cleft are shown.

***HDM2-p53*** The tumor suppressor protein p53 is involved in the maintenance of the genomic integrity of the cell. It coordinates the cellular response to DNA damage by binding to specific DNA sequences and activating genes responsible for growth arrest or apoptosis. The inactivation of p53 by the binding of the cellular oncoprotein HDM2 (human double minute 2) has been identified as an important step in tumorigenesis [9]. In normal cells, HDM2 and p53 form a negative feedback loop that helps to limit the growth-suppressing activity of p53. The identification of this key negative regulator HDM2 provided a great opportunity to manipulate the levels of the tumor suppressor p53 in cancer cells.

The mouse homolog MDM2 binds the N terminus of p53, thereby interfering with the transcriptional ability of p53. The crystal structure of the amino terminal domain of MDM2 bound to a small region on the N terminus of p53 displayed the specific interaction between a hydrophobic cleft in the MDM2 protein and an amphipathic α-helix of p53 [10]. Several van der Waals contacts augmented by two intermolecular hydrogen bonds were found at the interface. The MDM2 cleft lined with hydrophobic and aromatic amino acids interacts with the hydrophobic face of p53 amphipathic helix. As shown in Fig. 3, the side chains of Phe19, Trp23, and Leu26 from the p53 α-helical region nestle deep inside the hydrophobic pocket of MDM2. Since residue Leu22 and Trp23 of p53 are critical for its transcriptional activity, p53 is rendered transcriptionally inactive after binding to MDM2. Inhibitors of the MDM2–p53 interaction have thus been found as attractive targets to gain activity of p53 in tumor cells [11].

***Bcl-2/Bcl-$X_L$-BH3*** The proteins in the Bcl-2 family regulate apoptosis by maintaining a fine balance between the pro- and antiapoptotic proteins within the cell

[12–14]. Proapoptotic members of Bcl-2 family such as Bax, Bak, Bid, and Bad, and antiapoptotic members such as Bcl-2 and Bcl-$X_L$ exist as homodimers or mixed heterodimers. The nature of dimerization between these proteins dictates how a cell will respond to an apoptotic signal. Antiapoptotic proteins, Bcl-2, Bcl-$X_L$, or both, are overexpressed in the majority of human cancers and may play a vital role in cancer development. Therefore, inhibition of Bcl-2/Bcl-$X_L$ activity is gaining recognition for the development of potent therapeutics as anticancer drugs. Several strategies have been employed to target these proteins, including inhibition of expression levels using antisense oligonucleotides and identification of peptide ligands that affect protein–protein association [15, 16].

The antiapoptotic function of Bcl-2 and Bcl-$X_L$ is partially attributed to their ability to heterodimerize with proapoptotic members and antagonize their proapoptotic function. Three regions of the antiapoptotic proteins, namely, the Bcl-2 homology 1 (BH1), BH2, and BH3 binding sites, participate in their death-inhibiting activity and heterodimerization with the proapoptotic protein. However, only the BH3 binding site of the Bcl-2 and Bcl-$X_L$ proteins is found critical for the ability to antagonize apoptosis. Small, truncated peptides derived from the BH3 region of Bak (16 mer) and Bad (25 mer) have been found necessary and sufficient both for promoting cell death and binding to Bcl-$X_L$ [17, 18]. Furthermore, a synthetic cell permeable BH3 peptide derived from proapoptotic Bad has been shown to induce apoptosis *in vitro* and to have *in vivo* activity in human myeloid leukemia growth in severe combined immunodeficient mice [19].

The NMR solution structures of Bcl-2 and Bcl-$X_L$ alone [20, 21] and Bcl-$X_L$ in complex with Bak or Bad BH3 peptides [17, 18] have provided detailed information about the binding interactions of these proteins to the BH3 peptides (Fig. 4). The three-dimensional structures illustrate the formation of a hydrophobic cleft (the BH3 binding site) by the three BH domains of Bcl-2 and Bcl-$X_L$ in which the Bak or Bad BH3 domain binds. The overall binding motif of Bcl-$X_L$/Bak and Bcl-$X_L$/Bad

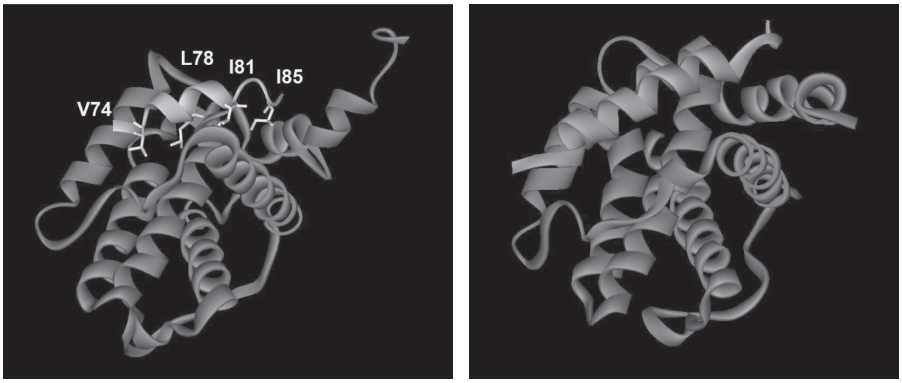


(a) (b)

**FIGURE 4** The ribbon representation of the NMR solution structures of Bcl-$X_L$ in complex with (a) Bak BH3 peptide and (b) Bad BH3 peptide derived from the PDB codes 1BXL [17] and 1G5J [18], respectively. Bak and Bad BH3 peptides bind in the hydrophobic cleft of the Bcl-$X_L$ protein. The hydrophobic side chains Val74, Leu78, Ile81, and Ile85 of Bak BH3 peptide inserting into the Bcl-$X_L$ cleft are highlighted as a stick model.

complex was found to be very similar with only a few differences at the protein–peptide interface. The Bcl-$X_L$/Bak structure shows that the hydrophobic side chains of the Bak peptide (Val74, Leu78, Ile81, and Ile85) point into the hydrophobic cleft (BH3 binding site) of Bcl-$X_L$ and stabilize complex formation. Several electrostatic interactions between the oppositely charged residues of Bcl-$X_L$ (Glu129, Arg139, and Arg100) and the Bak peptide (Arg76, Asp83, and Asp84, respectively) were also present. Similar interactions were also found between the Bcl-$X_L$ and 25 mer Bad peptide. However, the longer Bad peptide makes additional contact at the two ends of the BH3 binding site, forming a tighter complex ($K_d$ 0.6 nM) as compared to the Bcl-$X_L$/Bak 16 mer complex ($K_d$ 480 nM). These observations suggest that the BH3 binding pocket of Bcl-$X_L$ as well as Bcl-2 is essential for its antiapoptotic function, and small molecules that bind to the BH3 binding pocket of Bcl-$X_L$/Bcl-2 can block the interaction between Bcl-$X_L$/Bcl-2 and proapoptotic proteins such as Bak, Bax, and Bad.

***XIAP-Caspase*** Inhibitors of apoptosis proteins (IAPs) are important but incompletely understood negative regulators of apoptosis. Among other mechanisms, IAPs selectively bind and inhibit caspases-3, -7, and -9, but not caspase-8. Currently, there are eight members of the IAP family. Of these members, X-linked inhibitor of apoptosis protein (XIAP) is upregulated in many cancers and has thus garnered the most attention as a drug discovery target [22].

XIAP is a 57 kDa protein with three zinc-binding baculovirus IAP repeat domains (BIR1–3) and a really interesting new gene (RING)-finger that binds and inhibits caspases with nanomolar affinity (Fig. 5). The BIR2 domain inhibits caspases-3 and -7, whereas BIR3 domain inhibits caspase-9. The function of the BIR1 domain has not yet been determined. The RING finger contains an E3 ubiquitin ligase. The proapoptotic protein SMAC is an endogenous human IAP antagonist that binds and inhibits XIAP, thereby releasing caspases and reactivating apoptosis. Structural studies map the interaction between XIAP and SMAC and provide a basis for the development of small molecule XIAP inhibitors. These studies demonstrate that SMAC binds to both the BIR3 and BIR2 domains of XIAP. Crystal structure of BIR3 domain of XIAP in complex with caspase-9 [23] and SMAC [24] illustrates the key interactions between the complexes. The N terminus of the small subunit of caspase-9 binds the same shallow groove on BIR3 as the N terminus of SMAC (Fig. 6). The N-terminal 4–7 amino acids of active SMAC are necessary and sufficient for binding the BIR3 pocket of XIAP and preventing XIAP from binding and inhibiting caspase-9. A 4 mer peptide, Ala-Val-Pro-Ile, derived from SMAC binds to XIAP with ~500 nM affinity [25]. These results indicate that small molecules that

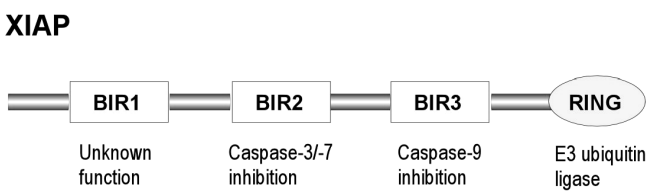


**FIGURE 5** Schematic representation of XIAP showing different domains and their functions.

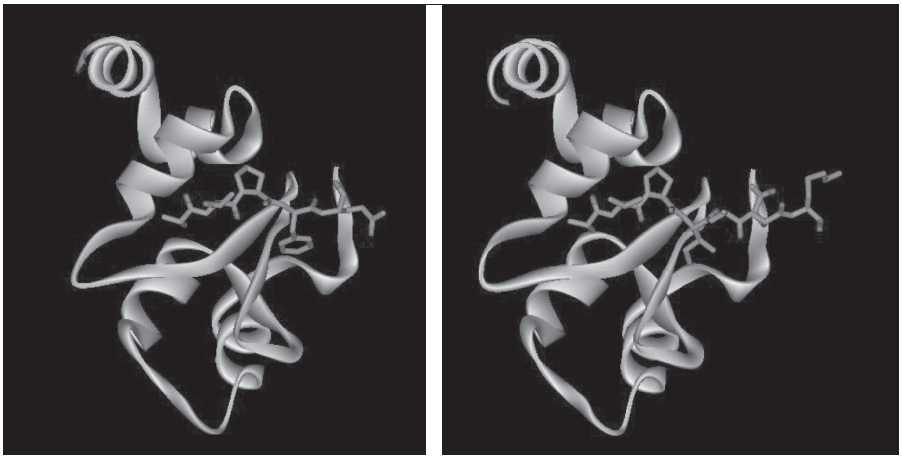

(a) (b)

**FIGURE 6** The three-dimensional structures of XIAP-BIR3 domain in complex with (a) N terminus of the small subunit of caspase-9 (Ala316, Thr317, Pro318, Phe319, and Gln320 shown as stick model) and (b) N terminus of SMAC (Ala1, Val2, Pro3, Ile4, Ala5, Gln6, and Lys7 shown as stick model). Both caspase-9 and SMAC bind in the same pocket of BIR3 as illustrated by the crystal structures of the complexes (PDB accession codes 1NW9 [23] and 1G73 [24], respectively).

mimic the actions of SMAC could be identified, providing an opportunity for a structure-based approach to the design of BIR3 inhibitors.

***Stat3 Dimerization*** Signal transducer and activator of transcription 3 (Stat3) is a cytoplasmic transcription factor that is activated in response to cytokines and growth factors. Upstream regulators of Stat3 constitute JAKs, Src, and EGF receptors and downstream targets include antiapoptotic and cell cycling genes such as Bcl-$X_L$ and cyclin D1. Stat3 is overactivated in a surprisingly large number of cancers including head and neck, breast, brain, prostate, lung, leukemia, multiple myeloma, lymphoma, and pancreas and has therefore been identified as a potential target for cancer drug development [26].

Stat3 is composed of several domains, namely, an oligomerization domain (N terminal), a coiled coil domain, a DNA-binding domain, a linker domain, a Src homology 2 (SH2) domain, a critical tyrosine at position 705 (C-terminal end), and a C-terminal transactivation domain. On activation, JAK kinase-2 phosphorylates the tyrosine residues of its coreceptor, thereby facilitating the binding of Stat3 to specific phosphotyrosine residues of JAK-2 through its SH2 domain. This leads to phosphorylation of Tyr705 on the C terminus of Stat3, followed by Stat3 dimerization by the reciprocal interaction between the SH2 domain of one monomer and the phosphorylated tyrosine of the other. The activated dimers translocate to the nucleus, where they bind to specific DNA sequences and activate gene expression. These observations have pointed out Stat3 inhibition, such as inhibition of JAK-2/Stat3 interaction and Stat3 dimerization, as a novel molecular target for the advancement of a broad anticancer therapy. The X-ray elucidation of three-dimensional

structure of the Stat3 homodimer bound to DNA reveals some details about the interaction between the two monomers and may facilitate the development of Stat3 inhibitors [27].

***Rac1-GEF*** Rho family GTPases, such as Rac1, control signaling pathways that are involved in cell adhesion, cell migration, and other cellular processes. Overexpression or upregulation of Rho GTPases has been discovered in many human tumors, including colon, breast, lung, myeloma, and head and neck squamous-cell carcinoma [28]. They can be activated through specific interaction with guanine nucleotide exchange factor (GEF) proteins that catalyze the exchange of GDP for GTP. One strategy to control tumor spreading is the selective inhibition of Rho GTPase activation by its GEF TrioN or Tiam1. Trio is a large multifunctional domain molecule with the amino-terminal module (TrioN) displaying the Rac1-specific GEF activity. Similarly, Tiam1, the T-cell invasion and metastasis gene product of the Dbl family, is shown to be an active GEF for Rac1. The three-dimensional structures of GEF–Rho protein complexes discern the specific interactions between GEFs and Rho GTPases needed for the signaling specificity mediated by Rho proteins. The cocrystal structure of Rac1/Tiam1 complex [29] shows that a domain of Tiam1, mainly dominated by α-helices, binds a shallow groove of Rac1, suggesting the presence of a small-molecule binding site (Fig. 7). A micromolar inhibitor of Rac1/TrioN interaction that selectively inhibits Rac1/Tiam1 and Rac1/TrioN versus related complexes and inhibits Rac1 activation in cells has been reported [30]. This study indicates that inhibition of the Rac1–GEF protein–protein interaction is possible, and such interactions have cellular consequences.

***Integrin $\alpha_v\beta_3$ – Fibronectin*** Another well studied example of cell adhesion proteins is the integrins. Integrins are the cell-surface receptors that act as molecular recognition sites for other proteins (for cell–cell and cell–extracellular matrix inter-

**FIGURE 7** Crystal structure of Rac1 (left) in complex with the guanine nucleotide exchange region of Tiam1 (right) determined by Worthylake et al. [29] (PDB code 1FOE). The extensive interface of the complex buries over 3000 $Å^2$ of primarily hydrophobic surface area.

actions) as well as signaling molecules transferring ligand-binding information to the cytoplasm. Integrins are heterodimeric proteins consisting of α and β subunits and typically have a high molecular mass of ~300 kDa. At least, 25 αβ integrin heterodimers have been reported and six of them are currently being evaluated in clinical trials for cancer [31]. Integrin $\alpha_v\beta_3$ has received particular attention as a potential target for anticancer drug design. The expression of $\alpha_v\beta_3$ is significantly increased on vascular cells in human tumors, but is weakly expressed on normal or quiescent endothelial cells. Since this integrin is relatively limited in its normal distribution, inhibition of its action is considered as an effective means of depriving tumors of nascent blood vessels without involving normal tissues.

Integrin recognition of the extracellular matrix ligands, such as fibronectin, collagen, and vitronectin, relies on the concerted binding of both the α and β subunits to regions of the ligand containing the Arg-Gly-Asp (RGD) sequence [32]. The RGD motif was the first integrin binding motif discovered. Several new motifs have been found since then that bind to a specific class of integrins [33]. Studies on RGD sequence have led to the discovery of cyclic pentapeptide Arg-Gly-Asp-{D-Phe}-{*N-methyl*-Val} or cyclo RGDf{NMe}V that specifically binds and inhibits integrin $\alpha_v\beta_3$ [34].

## 2.2 Pathogen–Host Interaction

Some viruses and bacteria enter eukaryotes by attachment to specific cell-surface receptors or cell-surface receptor-binding proteins. Viruses infect higher eukaryotes to reproduce themselves, whereas bacterial pathogens invade primarily to gain protection against the host immune system. Pathogens have always "enjoyed" invading human cells and have coevolved with their hosts to enable efficient entry, replication, and exit during their infectious cycles. An excellent review by Dimitrov [35] describes in depth the different virus entry mechanisms at the molecular level and opportunities for therapeutic intervention by inhibiting these processes. In this section, interaction between specific proteins of virus or bacteria and the target cell that facilitates pathogen entry into the target cell are discussed.

***Papillomavirus E2 Protein*** Infection by papillomavirus causes benign lesions that can lead to cervical cancer and other tumors [36, 37]. Papillomaviruses are small DNA viruses that infect higher eukaryotes by invading the basal layer of epithelial cells where they replicate successfully. Viral E2 protein has been found essential for replication and survival. E2 protein contains two conserved domains, the C-terminal viral DNA binding domain and the N-terminal transactivation domain that binds the viral E1 protein. Molecules that can bind these two domains of E2, thereby inhibiting the E2/DNA or E1/E2 interaction, are attractive targets for the development of therapeutics to prevent or treat papillomavirus infections. The three-dimensional structure of E1 bound to E2 reveals some important contact points between the complex [38]. The interaction surface, comprised of three helices from the N-terminal domain of E2, buries ~940 $Å^2$ surface area per protomer on E1–E2 complex formation (Fig. 8).

***HCV-Envelope Protein 2*** Hepatitis C virus (HCV) infection, another important target for antiviral drug design, causes severe medical problems, including chronic hepatitis, cirrhosis, and hepatocellular carcinoma. HCV genome is composed of a

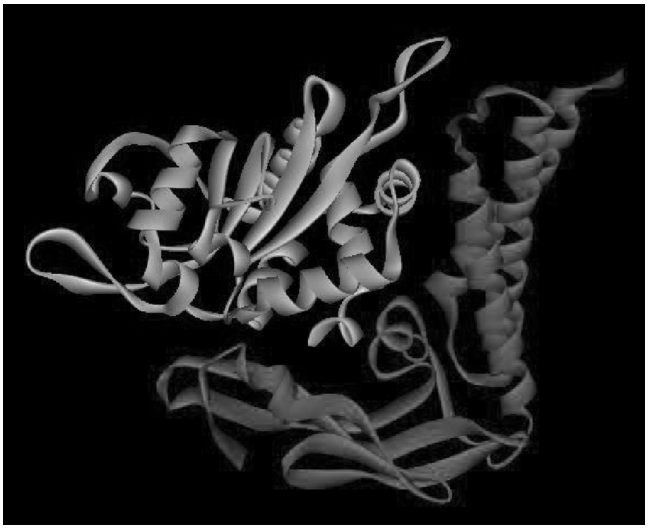

**FIGURE 8** The X-ray structure of papillomavirus E1 helicase (upper structure) in complex with its molecular partner E2 (lower structure). PDB accession code: 1TUE [38].

single-stranded positive sense RNA of approximately 9600 nucleotides that are translated into a polyprotein precursor of about 3000 amino acids. The HCV polyprotein precursor is processed by host and viral proteases to yield structural and nonstructural proteins, which are essential for replication and assembly of new viral particles. The viral envelope E2 protein initiates the infection by association with specific cell-surface receptor(s). Many groups have demonstrated that the truncated soluble versions of E2 bind specifically to hepatocytes [39, 40]. This glycoprotein is found to interact with CD81, scavenger receptor class B type 1 (SR-B1), and dendritic cell-specific intracellular adhesion molecule 3-grabbing nonintegrin (DC-SIGN). Such findings suggest that these proteins may act as receptors for HCV on the cell surface. Therefore, inhibition of interaction between E2 and the cell-surface receptors, such as CD81, has been identified as a possible target for designing anti-HCV molecules [41, 42].

***SARS-Angiotensin Receptor*** SARS-CoV is a member of the Coronaviridae, a family of positive strand RNA virus that have long been known to cause severe acute respiratory syndrome in many animals and more recently in humans. Similar to other known coronaviruses, SARS-CoV is an enveloped virus containing four structural proteins, namely, the membrane (M), envelope (E) glycoprotein, spike (S) glycoprotein, and nucleocapsid (NP) proteins [43]. The spike protein of SARS-CoV is a large type I glycoprotein and is made up of two domains, the S1 near the N terminus and the S2 near the C terminus. Unlike other coronaviruses, the spike protein of SARS-CoV is not posttranslationally cleaved in virus producing cells. The S1 and S2 domains form the globular head and the stalk of the spike protein and play an important role in specific receptor recognition and cell fusion. The S1 domain mediates receptor association whereas the S2 domain is membrane associated and likely undergoes structural rearrangements. This conformational change initiates the

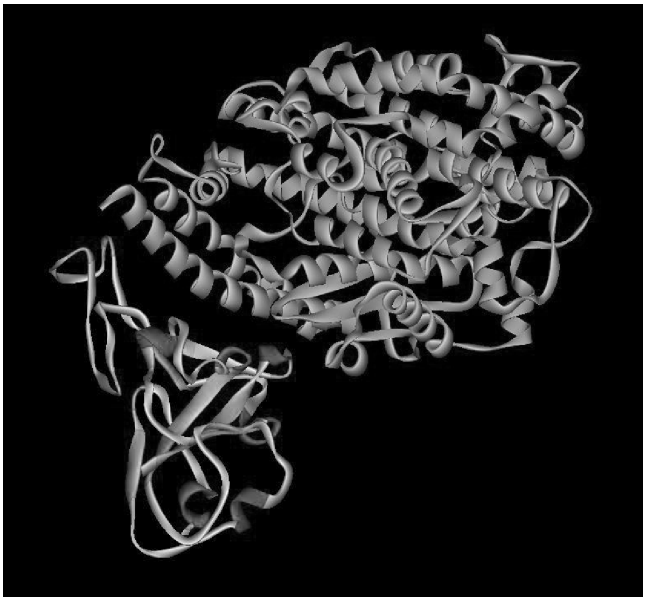

**FIGURE 9** Crystal structure of SARS-CoV spike protein RBD (lower ribbon structure) in complex with human receptor ACE2 (upper structure). PDB accession code 1AJF [47].

fusion of the virus and host cell membrane, allowing for entry of the virus. The first step in viral infection is the binding of viral proteins to certain host cell receptors. The spike protein of coronavirus is considered as the main site of viral attachment to the host cells. It has been demonstrated that a metallopeptidase, angiotensin converting enzyme 2 (ACE2) isolated from Vero E6 cells, efficiently binds the S1 domain of the SARS-CoV spike protein [44]. A discrete receptor binding domain (RBD) of the spike protein has been defined at residues 318–510 of the S1 domain [45] and this receptor binding domain is the critical determinant of virus receptor interaction and thus of viral host range and tropism. It has been demonstrated that this RBD binds ACE2 with higher affinity than does the full length S1 domain [46]. The crystal structure (Fig. 9) of SARS-CoV RBD complexed with ACE2 receptor at 2.9 Å shows that the RBD presents a gentle concave surface, which cradles the N-terminal lobe of the peptidase [47].

***Bacterial Fibronectin-Binding Proteins*** One of the mechanisms by which bacterial pathogens invade cells is by displaying fibronectin-binding proteins (FBPs) on their surface. This approach to internalize into the host cell has been adopted by some pathogenic gram-positive bacteria. FBPs contain tandem arrays of intrinsically disordered repeat sequences that bind fibronectin–integrin complexes. The NMR solution structure of a complex comprising a peptide fragment of a streptococcal fibronectin-binding protein bound to the first two domains of human fibronectin reveals the tandem β-zipper interactions between the two fragments [48]. The tandem β-zipper is created by the β-strand conformation of the repeat sequences of FBP peptide fragment that extends existing antiparallel β-sheets in both directions upon binding to fibronectin. The binding affinity of these complexes is relatively weak but

increases significantly when additional domains of both proteins are present. For example, the binding affinity ($K_A$) of two FBP repeats to pairs of fibronectin domains is ~$10^6$ $M^{-1}$. The tandem β-zipper interaction is a common phenomenon found in several pathogenic gram-positive bacteria, such as *Staphylococcus aureus, Streptococcus pyogenes*, and *Borrelia burgdorferi*, and may prove to be a widespread mechanism for bacterial foray of host cells [49–51]. Molecules that disrupt these β-zipper interactions may prove to be useful therapeutics for bacterial infections.

## 2.3 Loss of Normal Protein–Protein Interaction

Modular protein–protein interactions mediated by the tandem β-zipper have also been observed in eukaryotes [52, 53]. The LIM domains, found only in eukaryotes, are proteins with diverse functions such as transcription factors and protein kinases [54]. They are known to mediate specific protein–protein interactions through their LIM domains. Human genome encodes four LIM-only (LMO) and 12 LIM-homeodomain (LIM-HD) proteins each with a pair of tandem LIM domains at their N terminus (Fig. 10a). Three out of four LMO proteins have been implicated in oncogenesis. The LIM domains of all LMO and LIM-HD proteins bind the LIM domain-binding protein, Ldb1, through the 30 residue LIM interaction domain (LID) of this protein. Ldb1 is a ubiquitously expressed protein that contains an N-terminal dimerization domain, LID, and several other binding domains and is an

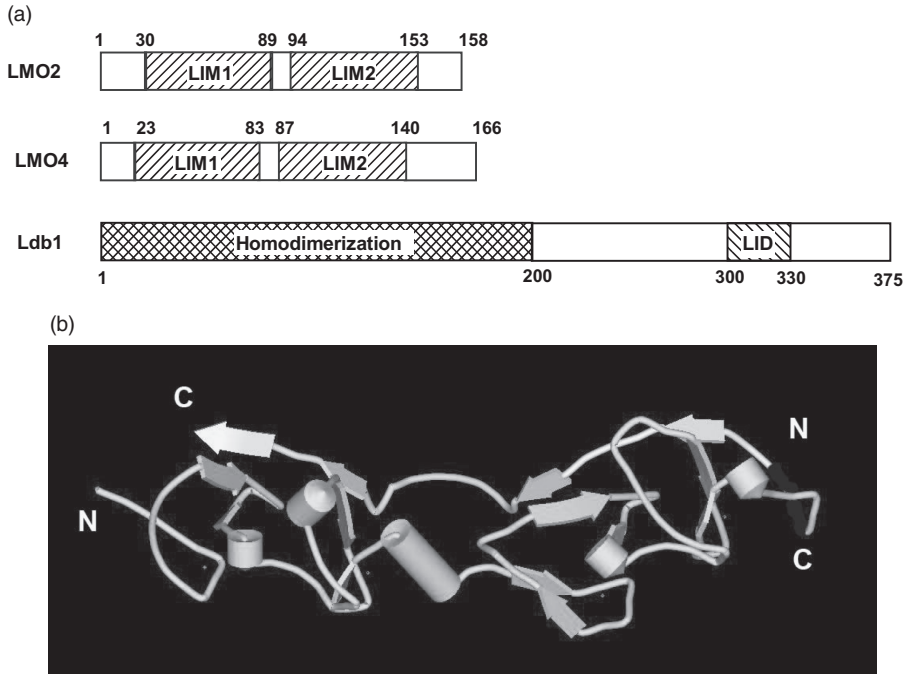


**FIGURE 10** (a) Schematic representation of different domains of LMO2, LMO4, and Ldb1 proteins. (b) The schematic of LMO4 in complex with the Ldb1-LID domain displaying the "tandem β-zipper" interaction (PDB code 1RUT) [52]. The β-strand conformation of the peptide extends the existing β-structure in the partner protein in a modular fashion.

essential cofactor that plays diverse roles in the development of complex organisms. Since LMO proteins bind the same region of Ldb1 (LID) as the LIM-HD, LMO proteins can regulate the transcriptional activity of LIM-HD by competing for binding to Ldb1. The displacement of endogenous LMO4 by ectopically expressed LMO2, as the normal binding partner for Ldb1, has been directly linked with the overexpression of LMO2 in T cells and the onset of T-cell acute lymphoblastic leukemia in children [55].

The three-dimensional structures of the complexes comprising LIM domains and Ldb1-LID present that the intrinsically disordered Ldb1-LID forms a "tandem β-zipper" upon binding to LIM domains (Fig. 10b) [52]. In the complex, the four β-strands of Ldb1-LID extend across one face of LMO4 and remain in continuous contact with the LIM domains, mainly by backbone–backbone hydrogen bonds, burying a total surface area of 3800 $Å^2$. Ldb1 is known to interact with LMO and LMO-HD proteins only and not with any other LIM domain containing proteins. Recent studies [52, 53], highlight the specific features of the interaction between Ldb1 and the LMO2 or LMO4 proteins. Structural, mutagenesis, and yeast two hybrid analysis are used to identify the key binding determinants for the complex formation. The differences in the binding interaction between the two protein complexes suggest that molecules that could bind specifically to LMO2 or LMO4 may have potential uses in the treatment of neoplastic disorders.

The protein complexes discussed above are a few examples of interaction pairs that have been identified as possible drug design targets where the interactions have been mapped at the atomic level. A wide variety of approaches are utilized in the identification of these protein–protein interaction pairs and their inhibitors. Some of the most popular approaches are discussed in the following section.

## 3 SCREENING OF PROTEIN–PROTEIN INTERACTION INHIBITORS

Several approaches have been utilized in the identification of protein–protein interaction inhibitors with the aim of developing therapies for a variety of human diseases [6, 56, 57]. An analysis of current strategies employed for the identification of lead molecules demonstrates that a search for competitors of a known binder is the basis of traditional screening as well as more modern approaches. In the following sections, some of these approaches are described in detail.

### 3.1 Structure–Activity Relationship

A common approach relies on the experimentally determined (NMR or X-ray) structure of the protein complex. In this approach, one attempts to disrupt the interfacial interactions between the two proteins by developing mimics of the interface amino acid residues (peptide fragment) for one of the binding partners. The structure of the interface peptide fragment is modified using computer docking and molecular dynamics simulations to obtain peptidomimetics or small organic molecules [3]. The resulting peptides or peptide analogues present the interacting functional groups in similar spatial orientations as the interface amino acid residues. Peptidomimetic or small molecule inhibitors have higher affinity, better selectivity, and often better pharmacokinetic properties than the parent peptide. A second

approach involves screening of a huge library of compounds using computer docking to find molecules with high affinity toward one of the binding partners of the protein complex. The lead structures identified using the above procedures are further optimized for potency and selectivity by structure–activity relationship (SAR) studies [58]. The SAR methods utilize experimental techniques, like nuclear magnetic resonance (NMR) spectroscopy, or computation methods, such as docking studies, and provide a structural perspective throughout the discovery and optimization of a lead molecule.

### 3.2 Genetic Screening Systems and Phage Display

The new library methodologies, such as phage display, allow generation of a large number of molecules with a fast screening and selection procedure to identify the most interesting lead candidates. Phage display technology has proved to be a very powerful *in vitro* technique for generating libraries containing millions of different peptides, proteins, or small molecules. Using the same technique, these libraries have been screened to identify ligands for peptide receptors, to define epitopes for monoclonal antibodies, to select substrates for enzymes, and to screen cloned antibody repertoires [59, 60].

In the phage display technique, filamentous virus is used as a platform for cloning of a DNA library (a library of genes or gene segments) encoding millions of variants of certain ligands into the phage genome and is fused to the gene encoding the phage coat or tail protein. Upon expression in the *E. coli* host in the presence of helper phage, the fusion protein (e.g., Coat protein-scFv) is incorporated into new phage particles that are assembled in the periplasmic space of the bacterium. Expression of the target gene fusion product and its subsequent incorporation into the mature phage coat results in the ligand being presented on the phage surface, while its genetic material resides within the phage genome. The proteins that are encoded by the library are expressed on the surface of phage and can be selected on the immobilized target molecule by biopanning. This interaction allows selection of high affinity binders for a variety of biomedical applications. Phages that bind the target molecule contain the gene for the protein and have the ability to replicate while nonadherent phages are washed away. This method can be used to efficiently clone genes encoding proteins with particular binding characteristics. In antibody phage display, the Fab or single chain fragment of IgG variable proteins is displayed on phage. This approach for antibody development offers advantages over immunization of animals and hybridoma technology [61]. Phage display can produce antibodies more quickly in a cost effective manner than traditional approaches. Additionally, antibody phage display techniques can potentially isolate antibodies to molecules that are not immunogenic in animals due to tolerance mechanism. Phage selection is not limited to the isolation of antibodies or short peptides. This approach has also been instrumental in studies and manipulation of a variety of other biologically active molecules and their designer variants [62].

### 3.3 Yeast Two Hybrid System and Intracellular Antibodies

Cell-based assays that monitor the intracellular behavior of target molecules, rather than binding or catalytic activity of purified proteins, are also being used in high

throughput screening of protein interaction inhibitors. These assays offer an opportunity to discover entirely new classes of compounds, molecules that act primarily by modulating protein interactions in living cells.

The yeast two hybrid system is a cell-based genetic selection assay that has been successfully used to identify protein–protein interactions *in vivo*. The model originally developed by Fields and Song [63] exploits the fact that transcription factors are comprised of two domains, a DNA binding domain and a transactivation domain. As an example, GAL4 protein of yeast (*Saccharomyces cerevisiae*) is a transcriptional activator required for the expression of genes encoding enzymes of galactose utilization. The native GAL4 protein contains two domains: an N-terminal domain that binds to specific DNA sequences but fails to activate transcription; and the C-terminal acidic domain that is necessary to activate transcription but cannot initiate function without the N-terminal domain. The basic strategy of the two hybrid system involves two proteins of interest that are expressed as two different fusion proteins. One fusion protein, known as the bait, is fused to the DNA binding domain to bind at specific sites upstream of the reporter gene. The second fusion protein, known as prey, is fused to the transactivation domain. If a physical interaction occurs between the two proteins, it brings the GAL4 domain in sufficient proximity to activate the GAL4-dependent transcription of a reporter gene. There will be no expression of the reporter gene if the two proteins do not interact in the intracellular milieu.

The two hybrid system may not be a useful tool for all protein–protein interactions. The limitation of the technique includes where the protein of interest is able to initiate GAL4-dependent transcription. Toxicity of the expressed protein or misfolding of the chimeric protein inside the cell might result in a limited activity or inaccessibility of binding site to the other protein. Furthermore, some protein–protein interactions depend on posttranslational modification (S–S bond, glycosylation, and phosphorylation) that may not appropriately occur in yeast. Two hybrid systems need the fusion protein to be targeted to the yeast nucleus and it might be a disadvantage for extracellular proteins. Weak and transient interactions are often the most interesting in signaling cascades. These are more rapidly detected in the two hybrid system in view of the significant amplification of the reporter gene in this system.

Intracellular antibodies are antibody fragments that are targeted and expressed inside the cells for interaction with cellular target antigens. This strategy can inhibit the regular function or in some cases mediate cell killing following antigen binding. Specific activity of certain intracellular proteins has been blocked by microinjection of full length antibodies [64, 65] or of hybridoma mRNA [66, 67] into the cytoplasm of various cell types. Recent advances in DNA technology and antibody engineering have allowed the development of specific, high affinity antibodies to target antigens. These probes could be targeted intracellularly as unique nontoxic therapeutics. Recombinant antibody reductants that provide many of the essential features of antibodies are suitable forms to be expressed *in vivo* or internalized efficiently inside the cells. The recombinant single chain Fv fragment (scFv) has been the most widely used for intracellular antibodies [68]. Intracellular single domain antibodies have also been isolated from yeast libraries with good antigen binding affinities [69].

The first step of intracellular antibody isolation is the derivation of the V regions of the heavy and light chains of a high affinity monoclonal antibody against a

target antigen. The VH and VL sequences could be amplified by RT-PCR of mRNA isolated from the hybridoma cells, assembled and cloned as a scFv [70]. Alternatively, one of the *in vitro* display systems, such as phage display [60], yeast display [71], or ribosome display [72] techniques, could be employed to generate the scFv libraries from immunized mouse spleen total mRNA and screen scFv libraries with the desired antigen to select specific scFv clones. Intracellular antibody capture technology [73, 74] has also been developed for *in vivo* screening of scFv libraries for a target antigen. This involves *in vitro* biopanning of diverse scFv libraries developed by phage display, followed by *in vivo* screening of antigen–antibody interaction using the yeast two hybrid system. The coding sequences of the antigen are cloned in one of the two hybrid vectors expressing the GAL4 DNA binding domain–antigen fusion protein. The coding sequence of scFv is cloned in the other two hybrid vector resulting in expression of GAL4 activation domain–scFv fusion protein. Yeast cells cotransformed with both the vectors will result in the expression of the reporter gene if the antigen and scFv interact with each other. There will be no reporter gene expression if the antibody fragment does not functionally interact with the antigen inside the yeast cells. By this technology it is possible to select and isolate intracellular antibodies, which could widely interact with the target protein inside the cells to alter or affect the protein function. Such induced intracellular protein–protein interactions could be an efficient pathogen neutralizing strategy for several viral and bacterial diseases [68, 75].

## 4 INHIBITORS OF PROTEIN–PROTEIN INTERACTIONS

As mentioned earlier, a large number of protein interaction complexes are emerging as potential targets for developing therapeutic agents. However, a big portion of these are ruled out at the onset due to the intricacies involved at the interaction site such as innate mutations and the atomic details of the binding site. The binding site may not present particular indentations, or if a pocket is present, its dimensions may be too small, or its geometry may be too shallow. Such features do not support tight binding of a drug-like molecule. Some of the above issues can be handled and a drug can be produced by generating antibodies against the target. In fact, therapeutic antibodies, including chimeric, humanized, and multivalent antibodies, and antibody fragments have been utilized in several instances and comprise over 30% of biopharmaceuticals currently undergoing clinical trials [76]. Several monoclonal antibodies against growth factors or their receptors are found effective in the treatment of solid tumors [77]. Antibodies tend to bind their targets with both high affinity and specificity and therefore block protein–protein interactions efficiently. However, antibodies are incompatible for intracellular targets, encountering problems such as poor delivery due to their relatively large size and lower stability of their disulfide-bonded structure in the reducing environment of the cell. Peptide inhibitors provide a much smaller substitute for *in vitro* inhibition of protein–protein interactions but are often not stable *in vivo* to be successful drugs. More stable variants, such as crosslinked peptides, peptide mimetics, and small molecule inhibitors, may prove to be better blocking agents for both intracellular and extracellular protein–protein interactions.

### 4.1 Peptide and Peptidomimetic Inhibitors

Specific recognition needed for a large protein surface seeks at least ~6 $nm^2$ area buried at the interface. The unique spatial distribution of the charged, polar, and hydrophobic residues at the interface are deemed important for recognition. Despite being conceptually simple, mimetics of the large interfacial area required for specific recognition remains a challenging endeavor. Nonetheless, steady progress has been made in the discovery of compounds that mimic protein surface and function.

A variety of peptide inhibitors have been reported over the last decade for blocking the MDM2–p53 association [11]. These peptides have been helpful in mapping out the interaction between the two proteins. The key interactions between MDM2 and p53 involve a relatively small area represented by three amino acids, namely, Phe19, Trp23, and Leu26 of p53. Optimization of peptides has led to the discovery of several low nanomolar inhibitors of MDM2 that have recently been reviewed by Fotouhi and Graves [11]. Peptide mimetics have also been explored in order to increase the metabolic and proteolytic stability over α-peptide inhibitors. Schepartz and colleagues [78] targeted the HDM2–p53 interaction with a 14-helical structure made of beta amino acids to display the functional groups of Phe19, Trp23, and Leu26 in the same spatial orientation as found in p53. These synthetic $\beta^3$-peptides exhibited significant helical character in aqueous buffer and one of the oligomers, **1** (Fig. 11), selectively inhibited HDM2 interaction with nanomolar affinity. Similarly, Hamilton and associates [79] have utilized terphenyl scaffold to mimic one face of α-helical peptides. Substitution of the three ortho positions of the scaffold projected one side of the molecule analogous to the i, i + 4, and i + 7 residues of an α-helix. A terphenyl derivative with three hydrophobic side chains, compound **2** (Fig. 11), was found to bind specifically at the p53 binding site of HDM2 and exhibited a $K_i$ of 182 nM. More recently, α-helical peptidomimetics with the terphenyl scaffold and more soluble terephthalamide scaffold inhibited the Bak BH3–Bcl-$X_L$ interactions in the low micromolar range [80, 81].

Verdine and co-workers [82] targeted the BID/ Bcl-$X_L$ interaction by synthesizing hydrocarbon stapled helices to mimic the amphipathic α-helix BH3 domain of BID. These molecules, for example, **3** (Fig. 11), with constraint helix became proteolytically stable, cell permeable, and bound Bcl-$X_L$ with nanomolar affinity. In a cell-based assay, these compounds induced apoptosis and in an *in vivo* experiment inhibited the growth of human leukemia xenografts. Gellman and co-workers [83] generated chimeric (α/β + α)-peptides that mimic the α-helical display of BH3 domain of Bak. These peptides are tight binders and therefore potent inhibitors ($K_i$ = 0.7 nM) of Bak/ Bcl-$X_L$ interaction.

RGD motif present in the extracellular matrix ligands, such as fibronectin, was the first integrin binding motif identified. Since then, development of RGD mimetics that bind selectively to a single integrin has been a subject of intense research [33]. Studies have primarily focused on four integrins—$\alpha_4\beta_1$, $\alpha_5\beta_1$, $\alpha_v\beta_3$, and $\alpha_{IIb}\beta_3$—that bind RGD containing ligands and are thought to have the most clinical significance. An important RGD containing molecule that emerged out of these efforts is cyclo RGDf{NMe}V pentapeptide [34]. This cyclic peptide specifically inhibits integrin $\alpha_v\beta_3$ with an $IC_{50}$ of 0.6 nM and is in Phase II clinical trials as an anticancer drug under the name cilengitide. The crystal structure (Fig. 12) of the cyclic peptide bound to the extracellular segment of $\alpha_v\beta_3$ integrin in the presence of $Mn^{2+}$ metal cation

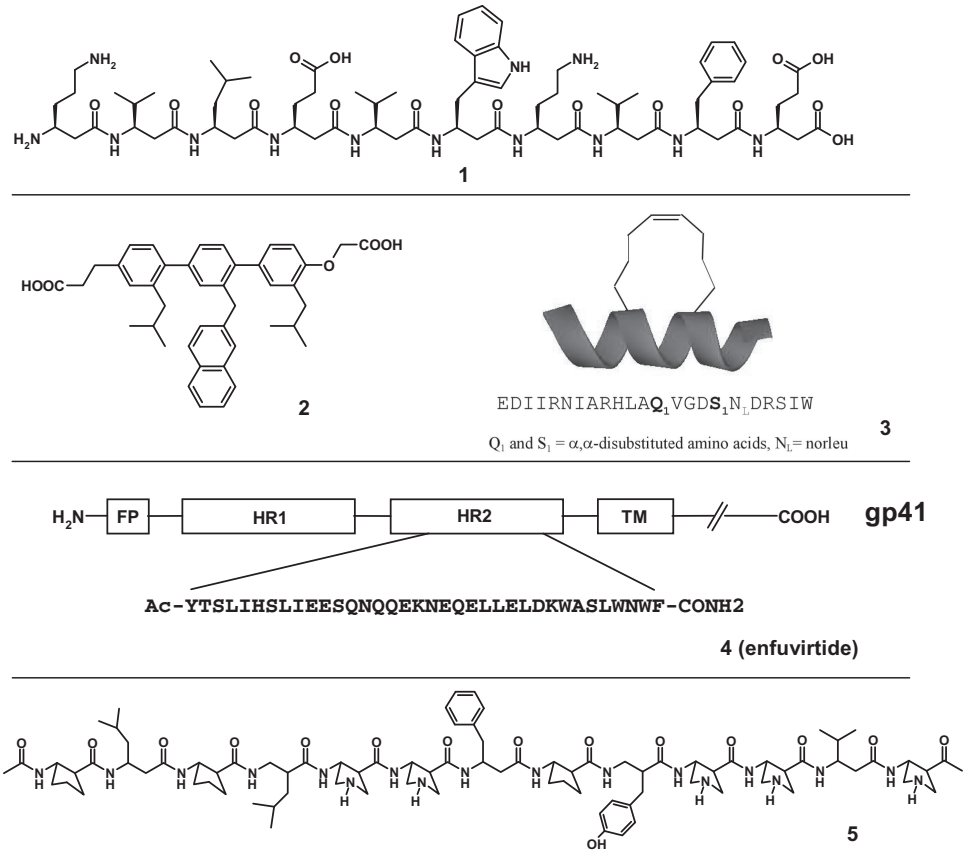


**FIGURE 11** Structure of peptide or peptidomimetic inhibitors of protein–protein interactions.

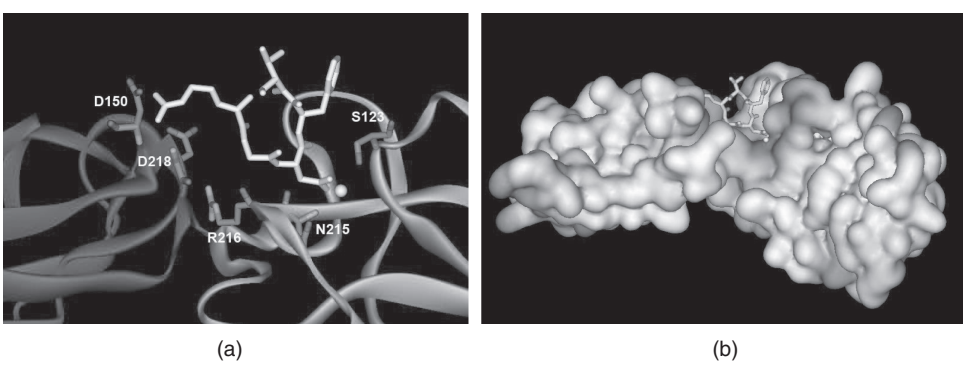


**FIGURE 12** (a) Crystal structure of cyclo RGDf{NMe}V peptide (stick model) bound in the active site of integrin $\alpha_v\beta_3$ (PDB accession code 1L5G [84]). The Asp side chain carboxyl of the cyclic peptide interacts with one of the $Mn^{2+}$. The interacting residues (stick model) from integrin $\alpha_v\beta_3$ are derived from both the α and β subunits. (b) Surface representation of the two subunits of integrin $\alpha_v\beta_3$ bound to the cyclic peptide (stick model).

reveals important binding interactions in the complex [84]. The peptide binds at the major interface between the $\alpha_v$ and $\beta_3$ subunits burying about 45% (355 Å$^2$) of its total surface area. The RGD sequence makes the main contact with the integrin subunits. The availability of the complex structure has prompted search of nonpeptide antagonists of $\alpha_v\beta_3$ integrin using detailed computer docking experiments.

Peptides have also been exploited in the design of antiviral agents [85]. A successful example in this case is the HIV antiviral drug Fuzeon (**4**), enfuvirtide, or T-20, Fig. 11) [86]. Fuzeon is a 36 amino acid α-peptide that was introduced in 2003 into the clinics as a potent HIV entry inhibitor [87]. HIV entry into the target cells takes place in several steps, beginning with the binding of viral envelope protein gp120 to CD4 receptors on the target cells, followed by the exposure of the buried transmembrane fusion protein gp41 and conformational changes for the assembly of the hexameric six-helix bundle gp41 that allows the fusion to take place. As depicted in Fig. 11, the gp41 is made of an N-terminal fusion peptide (FP), two heptad repeat regions HR1 and HR2, and a transmembrane region. Fuzeon (**4**) is a small peptide derived from the HR2 region of gp41 that competitively inhibits the last step of the viral fusion. Peptidomimetics of the HR2 region would serve as good alternate inhibitors of the HIV fusion, perhaps with better pharmacokinetic profile than the α-peptide. Hamilton and co-workers [88] have utilized the terphenyl derivatives to mimic the helical HR2 domain. The most potent molecule with hydrophobic side chains at the ortho position of the three phenyl rings efficiently inhibits HIV-1 infection in a cell fusion assay ($IC_{50}$ 15.7 μg/mL). Furthermore, *in vivo* studies of these molecules as anti-HIV agents are in progress.

Nozaki et al. [89] discovered that a small peptide fragment from the milk glycoprotein human lactoferrin is able to block the entry of hepatitis C virus (HCV) particles into the hepatocytes. Virus entry inhibitors act extracellularly by blocking the binding of the virus to the host cell and this process of shielding the virus from attachment to the target cells seems more facile compared to targeting other intracellular sites that require exacting precision. The authors demonstrated that the mechanism of action of this peptide fragment is by binding to the E2 protein of HCV, thereby blocking its entry into the host cell, rather than binding to the host cell-surface receptors. Peptide mimics of this 33 amino acid fragment that can resist proteolysis and are metabolically stable will be of great clinical interest, as there is no vaccine for HCV and current therapeutic strategies yield roughly 40% response rates. English et al. [90] have attempted to construct peptidomimetic entry inhibitors of human cytomegalovirus (HCMV). The authors have prepared and tested several β-peptides, oligomers of β-amino acid, as entry inhibitors of HCMV. The most potent β-peptide, **5** (Fig. 11), inhibited HCMV infection in a cell-based assay with an $IC_{50} \approx 30\,\mu M$.

The above examples reveal the prospect of peptides and peptidomimetics as potential therapeutics for various diseases. Currently, there are more than 40 marketed peptides worldwide and about 670 peptides are in either clinical or advanced preclinical phases. Several different classes of peptidomimetics are also entering the preclinical stages. Peptidomimetics like β-peptides, peptoids, and azapeptides are receiving particular attention as these unnatural oligomers, also called foldamers, fold into a conformationally ordered state in solution, like the natural biopolymers. These unnatural oligomers do not have disadvantageous peptide characteristics

and therefore may generate viable pharmaceuticals. They are protease resistant, resistant to metabolism, and may have reduced immunogenicity relative to peptide analogues.

### 4.2 Small-Molecule Inhibitors

A broad range of screening initiatives has helped identify several protein–protein complexes that are amenable to inhibition by small molecules. Several compounds have been identified that help characterize proteins such as MDM2, Bcl-2, and XIAP as drug targets. Additionally, small-molecule antagonists have recently been described for several new targets, including Rac1–Tiam1, β-catenin–T cell factor, and Sur-2-ESX. Several of these small molecule protein–protein inhibitors are virtually at the threshold of becoming therapeutics.

Fotouhi and Graves [11] have reviewed some interesting new scaffolds and leads as MDM2 inhibitors. Among several reported molecules, only a series of compounds termed Nutlins (**6**, Fig. 13) possessed *in vivo* activity and therefore drug-like properties [58]. The Nutlins with a core imidazoline mimic the α-helical structure of the p53 backbone. The three aryl rings on the imidazoline are presented in the same space as the side chains of Phe19, Trp23, and Leu26 of p53. Furthermore, it was demonstrated by 2D NMR spectroscopy that they bind to MDM2 at the p53-binding site. The specific interaction of Nutlins with MDM2 indeed translated to the selective growth inhibition of cells containing wild-type p53 ($IC_{50} \approx 1.5\,\mu M$) and showed 10–20-fold selectivity for cells with active versus mutated p53. Compound **6** was well tolerated, orally bioavailable, and inhibited the growth of an MDM2-overexpressing tumor in mice. It achieved a high steady-state concentration during the study, indicating good pharmacokinetic properties.

Bad and Bak proteins bind to Bcl-2 and Bcl-$X_L$ by inserting ~20 residue long α-helical Bad/Bak BH3 peptide into a hydrophobic groove [91]. Isolated BH3-like peptides also bind in this groove, suggesting the existence of a small-molecule binding site. There has been significant progress in developing compounds that bind in this groove on Bcl-2 and/or Bcl-$X_L$ and thereby augment cell death [15, 16]. Recently, several molecules, such as **7** [92], **8** [93], and **9** [94] (Fig. 13), have been developed and are moving into clinical trials. GX15-070, a small-molecule inhibitor from Gemin X, is specifically designed to inhibit all of the antiapoptotic members of the Bcl-2 protein family and is the first such small-molecule inhibitor tested in clinical trials. Phase I clinical trials of GX15-070 in patients with refractory solid tumors and lymphomas showed promising results, advancing GX15-070 into Phase II clinical trials.

Small molecules that inhibit the BIR domains of XIAP have been found to be promising candidates for the development of therapeutic XIAP inhibitors [4, 22]. Molecules have been developed to specifically target BIR2 and BIR3 regions of XIAP. This is due to the difference in the mechanism of caspase inhibition by the BIR2 and BIR3 domains and their ability to inhibit different caspases. The structural data available for the interaction between the BIR3 domain of XIAP and caspase-9 suggests that small molecules binding the BIR3 pocket of XIAP could mimic the action of SMAC and inhibit the interaction between XIAP and caspase-9. These structural studies have facilitated several research groups toward the discovery of cell-active ligands for XIAP. Tripeptide inhibitors, such as **10** and **11** (Fig. 13) with

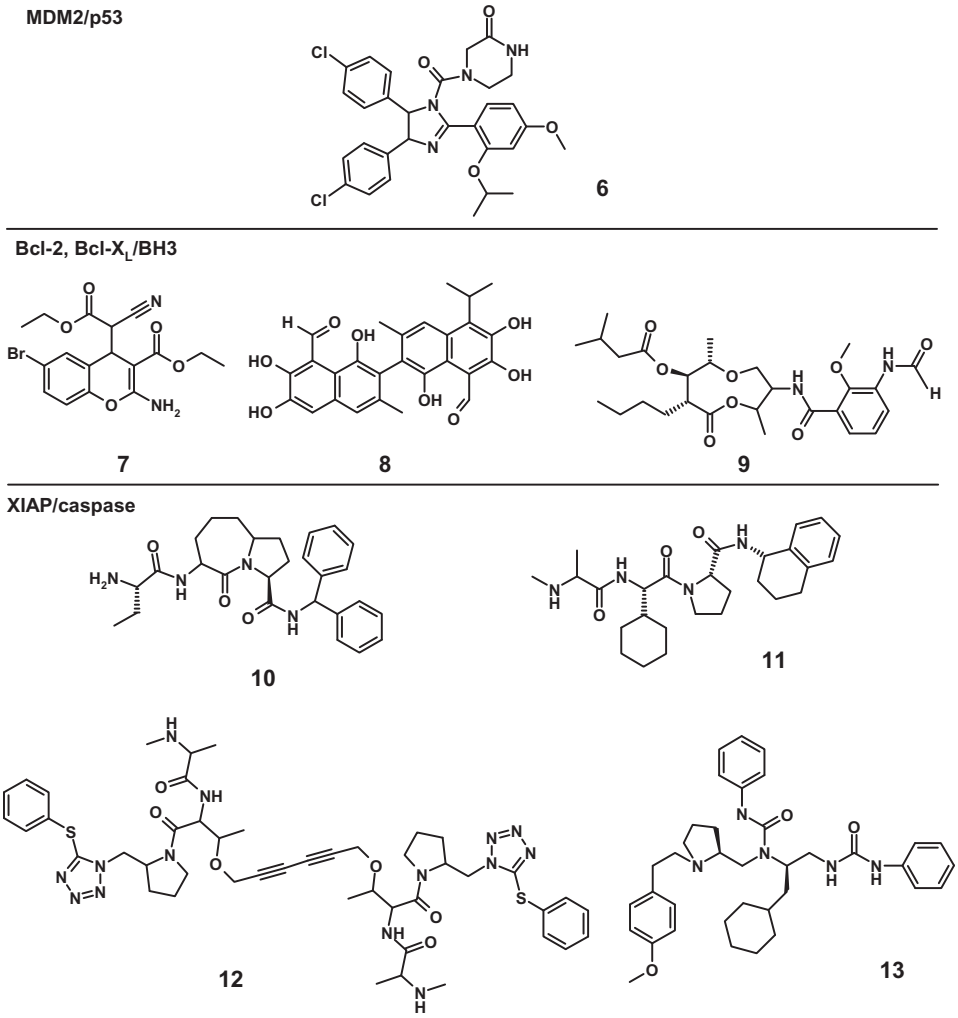


**FIGURE 13** Small-molecule inhibitors of protein–protein interactions in cancer.

unnatural amino acids were identified and tested for their binding to the BIR3 domain of XIAP by NMR spectroscopy and fluorescent polarization assay [95–97]. These molecules bound to the BIR3 domain at SMAC binding site with nanomolar affinity. Unlike **10**, the BIR3 ligand **11** could activate apoptosis in a caspase-dependent manner in the absence of additional stimuli. Li et al. [98] used computer-aided drug design to mimic SMAC peptide. Lead compound identified included a tetrazoyl thioether moiety and was modified to form a $C_2$-symmetric diyne. The compound **12** (Fig. 13) bound the BIR3 domain of XIAP with an affinity similar to SMAC peptides and also bound cIAP-1 and cIAP-2 in cells. The proposed bivalent binding mechanism of **12** to XIAP resembled the wild-type SMAC, interacting simultaneously with the BIR2 and BIR3 domains of XIAP. SMAC is a dimer and interacts with the BIR2 and BIR3 domains at the same time to inhibit XIAP.

Compounds with polyphenylurea pharmacophore were identified by screening a large combinatorial library for activation of caspase-3 in the presence of XIAP [99]. Compound **13** (Fig. 13) was found to be selective for caspases-3 and -7 over caspase-9 and did not inhibit the SMAC-XIAP interaction. The polyphenylurea inhibitors were toxic to a wide spectrum of malignant cell lines and demonstrated preferential toxicity to primary malignant cells over normal cells. In xenograft models, these compounds were found to delay the growth of tumors of the prostate, breast, and colon carcinoma cells without any unpleasant toxicity to the mice. The above examples of XIAP inhibitors clearly suggest additional studies are required to discern the feasibility of small-molecule XIAP inhibitors as potential therapeutics. However, the data already point to XIAP as an interesting target for therapy.

The strategy of inhibiting protein–protein interactions with a small molecule has also been applied to the design of antiviral and antibacterial compounds. White et al. [100] reported a small-molecule inhibitor of papilloma virus E1–E2 dimerization. The inandione inhibitor **14** (Fig. 14) binds to the hydrophobic pocket of the E2 protein as shown by the cocrystal structure of the complex (Fig. 15) [101]. A second weakly bound inandione molecule was also observed in the crystal structure, suggesting the presence of additional binding region on the E2 protein for exploiting

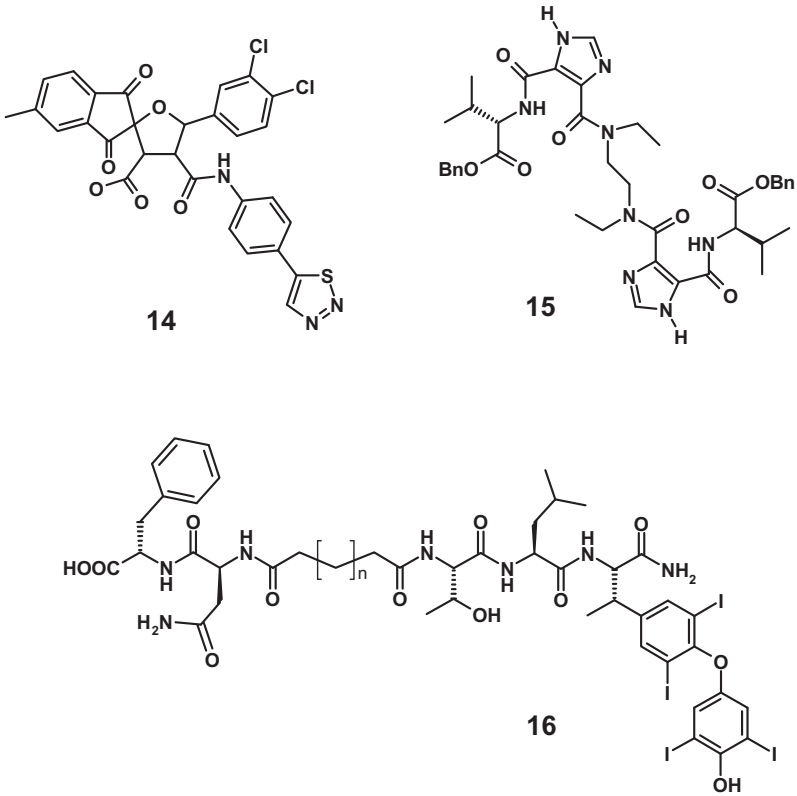


**FIGURE 14** Small-molecule inhibitors of papilloma virus E1 – E2 heterodimerization (**14**), HCV-E2 – CD81 interaction (**15**), and HIV-1 protease dimerization (**16**).

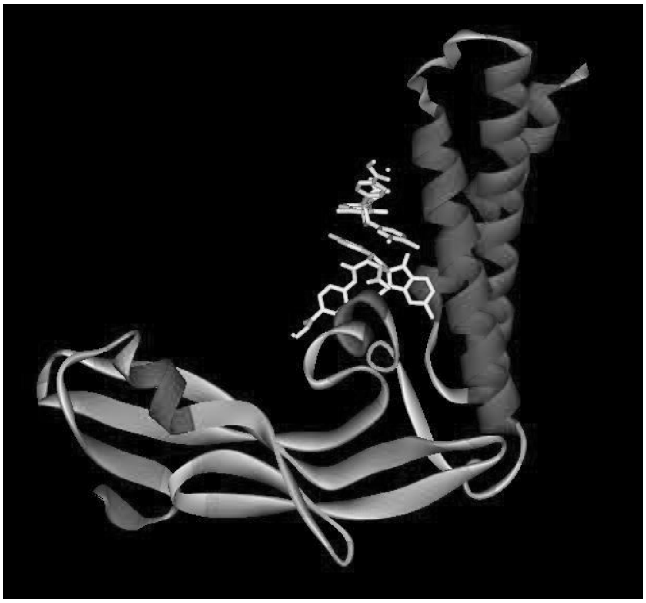

**FIGURE 15** The structure of a small-molecule inhibitor, an inandione bound to papilloma virus E2 protein (PDB accession code 1R6N [101]). The inhibitor binding to E2 prevents E1–E2 heterodimerization and disrupts viral replication. A second weakly bound inhibitor (top) suggests that additional functional groups could be added to the side chains of the inhibitor to gain binding affinity.

inhibitor design. Todd and colleagues [41] developed molecules like **15** (Fig. 14), with a novel bis-imidazole scaffold, as mimics of helix D of CD81 that reversibly inhibited binding of HCV-E2 to CD81 receptor protein.

Inhibition of HIV-1 protease dimerization is a promising strategy for anti-HIV drug design as opposed to the active-site directed inhibitors. In this regard, Chmielewski and co-workers [102] discovered a nanomolar inhibitor (**16**, $K_i$ = 71 nM, Fig. 14) of HIV-1 protease dimerization using a focused library approach. More importantly, the potent molecules of this class were equally active against wild-type and a mutant form of the enzyme. The mutant enzyme was resistant to active-site directed inhibitors, suggesting the importance of alternative drug design strategy.

### 4.3 Molecules Containing Porphyrin or Peptidocalixarene Scaffolds

Protein surface recognition by molecular architectures such as porphyrin and calixarene scaffolds has also been utilized in several instances. Hamilton and co-workers [103, 104] have used tetraphenylporphyrin derivatives to recognize the surface of cytochrome-*c* and identified subnanomolar binders. These molecules consist of peripheral anionic groups that bind positively charged Arg and Lys residues present on the cytochrome-*c* surface. The binding of phenylporphyrins induces unfolding of the protein, leading to disruption of tertiary and secondary structure. This denaturation of the protein facilitates proteolytic degradation. Another group used similar

porphyrin-based derivates for blocking potassium channels. Trauner and colleagues [105] used porphyrins to match the fourfold symmetry of the homotetrameric human $K_v1.3$ potassium channel. Using competitive binding assays, the authors showed that tetraphenylporphyrin derivatives with peripheral cationic groups strongly interact with potassium channel, thereby reducing the current through the channel.

Several synthetic receptors containing calixarene scaffolds have been designed to bind protein surfaces to block protein–protein interactions or the entry of small molecules into the active site of certain enzymes [106–109]. For example, calix[8]arene receptors decorated with basic amino acids competitively inhibit recombinant lung tryptase by binding to the acidic residues at the central junction of the tetrameric protein [107]. These molecules, most likely, bind at the entrance of the active site and block the approach of the substrate. Similarly, peptidocalix[4]arenes have been shown to bind the surface of transglutaminase, inhibiting its acitivity [108]. However, the competition assays suggest that these molecules bind to the surface of protein other than the enzyme active site, causing a conformational change in the protein or sterically blocking the approach of the substrate. Aachmann et al. [109] have designed β-cyclodextrin that binds to a specific site on the insulin surface via its solvent-exposed aromatic side chain. These studies suggest that porphyrin and calixarene scaffolds are certainly promising candidates for protein surface recognition and further work in this area may lead to novel therapeutic agents.

## 5 CONCLUSION

Over the past decade, protein complexes have become prime targets for therapeutic intervention. This has opened immense opportunities in the treatment of hitherto incurable diseases such as cancer. A large number of protein pairs have been identified as drug targets with reported successful inhibitors, suggesting the possibility of fighting disease in near future. However, with the identification of hundreds of possible drug targets in the "class" of protein–protein interaction complexes, picking targets for inhibition by peptides, peptide mimetics, or small molecules is going to be critical. Many protein pairs have interfaces where a small, linear region of one protein binds into a hydrophobic cleft of the other. Inhibitors for such interfaces have been discovered, using several approaches ranging from screening to structure-based design, that display sufficient potency and cellular activity. Unfortunately, a potent, specifically binding molecule does not necessarily make a good drug. The true potential of these molecules as therapeutics will only be realized following successful clinical trials.

## REFERENCES

1. Fregeau Gallagher NL, Sailer M, Niemczura WP, Nakashima TT, Stiles ME, Vederas JC. Three-dimensional structure of leucocin A in trifluoroethanol and dodecylphosphocholine micelles: spatial location of residues critical for biological activity in type IIa bacteriocins from lactic acid bacteria. *Biochemistry* 1997;36:15062–15072.
2. Ryan DP, Matthews JM. Protein–protein interactions in human disease. *Curr Opin Struct Biol* 2005;15:441–446.

3. Zhao L, Chmielewski J. Inhibiting protein–protein interactions using designed molecules. *Curr Opin Struct Biol* 2005;15:31–34.
4. Arkin M. Protein–protein interactions and cancer: small molecules going in for the kill. *Curr Opin Chem Biol* 2005;9:317–324.
5. Fletcher S, Hamilton AD. Protein surface recognition and proteomimetics: mimics of protein surface structure and function. *Curr Opin Chem Biol* 2005;9:632–638.
6. Sillerud LO, Larson RS. Design and structure of peptide and peptidomimetic antagonists of protein–protein interaction. *Curr Protein Pept Sci* 2005;6:151–169.
7. Pagliaro L, Felding J, Audouze K, Nielsen SJ, Terry RB, Krog-Jensen C, Butcher S. Emerging classes of protein–protein interaction inhibitors and new tools for their development. *Curr Opin Chem Biol* 2004;8:442–449.
8. Fry DC, Vassilev LT. Targeting protein–protein interactions for cancer therapy. *J Mol Med* 2005;83:955–963.
9. Bottger A, Bottger V, Garcia-Echeverria C, Chene P, Hochkeppel HK, Sampson W, Ang K, Howard SF, Picksley SM, Lane DP. Molecular characterization of the hdm2–p53 interaction. *J Mol Biol* 1997;269:744–756.
10. Kussie PH, Gorina S, Marechal V, Elenbaas B, Moreau J, Levine AJ, Pavletich NP. Structure of the MDM2 oncoprotein bound to the p53 tumor suppressor transactivation domain. *Science* 1996;274:948–953.
11. Fotouhi N, Graves B. Small molecule inhibitors of p53/MDM2 interaction. *Curr Top Med Chem* 2005;5:159–165.
12. Adams JM, Cory S. The Bcl-2 protein family: arbiters of cell survival. *Science* 1998;281:1322–1326.
13. Chao DT, Korsmeyer SJ. BCL-2 family: regulators of cell death. *Annu Rev Immunol* 1998;16:395–419.
14. Cory S, Huang DC, Adams JM. The Bcl-2 family: roles in cell survival and oncogenesis. *Oncogene* 2003;22:8590–8607.
15. O'Neill J, Manion M, Schwartz P, Hockenbery DM. Promises and challenges of targeting Bcl-2 anti-apoptotic proteins for cancer therapy. *Biochim Biophys Acta* 2004;1705:43–51.
16. Wang S, Yang D, Lippman ME. Targeting Bcl-2 and Bcl-XL with nonpeptidic small-molecule antagonists. *Semin Oncol* 2003;30:133–142.
17. Sattler M, Liang H, Nettesheim D, Meadows RP, Harlan JE, Eberstadt M, et al. Structure of Bcl-xL–Bak peptide complex: recognition between regulators of apoptosis. *Science* 1997;275:983–986.
18. Petros AM, Nettesheim DG, Wang Y, Olejniczak ET, Meadows RP, Mack J, et al. Rationale for Bcl-xL/Bad peptide complex formation from structure, mutagenesis, and biophysical studies. *Protein Sci* 2000;9:2528–2534.
19. Wang JL, Zhang ZJ, Choksi S, Shan S, Lu Z, Croce CM, Alnemri ES, Korngold R, Huang Z. Cell permeable Bcl-2 binding peptides: a chemical approach to apoptosis induction in tumor cells. *Cancer Res* 2000;60:1498–1502.
20. Petros AM, Medek A, Nettesheim DG, Kim DH, Yoon HS, Swift K, Matayoshi ED, Oltersdorf T, Fesik SW. Solution structure of the antiapoptotic protein bcl-2. *Proc Natl Acad Sci USA* 2001;98:3012–3017.
21. Muchmore SW, Sattler M, Liang H, Meadows RP, Harlan JE, Yoon HS, et al. X-ray and NMR structure of human Bcl-xL, an inhibitor of programmed cell death. *Nature* 1996;381:335–341.

22. Schimmer AD, Dalili S, Batey RA, Riedl SJ. Targeting XIAP for the treatment of malignancy. *Cell Death Differ* 2006;13:179–188.
23. Shiozaki EN, Chai J, Rigotti DJ, Riedl SJ, Li P, Srinivasula SM, Alnemri ES, Fairman R, Shi Y. Mechanism of XIAP-mediated inhibition of caspase-9. *Mol Cell* 2003;11: 519–527.
24. Web: Drugs used in the treatment of HIV infection. http://wwwfdagov/oashi/aids/viralshtml. 2000.
25. Liu Z, Sun C, Olejniczak ET, Meadows RP, Betz SF, Oost T, Herrmann J, Wu JC, Fesik SW. Structural basis for binding of Smac/DIABLO to the XIAP BIR3 domain. *Nature* 2000;408:1004–1008.
26. Jing N, Tweardy DJ. Targeting Stat3 in cancer therapy. *Anticancer Drugs* 2005;16: 601–607.
27. Becker S, Groner B, Muller CW. Three-dimensional structure of the Stat3beta homodimer bound to DNA. *Nature* 1998;394:145–151.
28. Sahai E, Marshall CJ. RHO-GTPases and cancer. *Nat Rev Cancer* 2002;2:133–142.
29. Worthylake DK, Rossman KL, Sondek J. Crystal structure of Rac1 in complex with the guanine nucleotide exchange region of Tiam1. *Nature* 2000;408:682–688.
30. Gao Y, Dickerson JB, Guo F, Zheng J, Zheng Y. Rational design and characterization of a Rac GTPase-specific small molecule inhibitor. *Proc Natl Acad Sci USA* 2004;101:7618–7623.
31. Tucker GC. Integrins: molecular targets in cancer therapy. *Curr Oncol Rep* 2006;8: 96–103.
32. Humphries MJ. The molecular basis and specificity of integrin–ligand interactions. *J Cell Sci* 1990;97:585–592.
33. D'Andrea LD, Del Gatto A, Pedone C, Benedetti E. Peptide-based molecules in angiogenesis. *Chem Biol Drug Des* 2006;67:115–126.
34. Dechantsreiter MA, Planker E, Matha B, Lohof E, Holzemann G, Jonczyk A, Goodman SL, Kessler H. N-methylated cyclic RGD peptides as highly active and selective alpha(V)beta(3) integrin antagonists. *J Med Chem* 1999;42:3033–3040.
35. Dimitrov DS. Virus entry: molecular mechanisms and biomedical applications. *Nat Rev Microbiol* 2004;2:109–122.
36. zur Hausen H. Papillomaviruses and cancer: from basic studies to clinical application. *Nat Rev Cancer* 2002;2:342–350.
37. Baseman JG, Koutsky LA. The epidemiology of human papillomavirus infections. *J Clin Virol* 2005;32:S16–S24.
38. Abbate EA, Berger JM, Botchan MR. The X-ray structure of the papillomavirus helicase in complex with its molecular matchmaker E2. *Genes Dev* 2004;18:1981–1996.
39. Bartosch B, Vitelli A, Granier C, Goujon C, Dubuisson J, Pascale S, Scarselli E, Cortese R, Nicosia A, Cosset FL. Cell entry of hepatitis C virus requires a set of co-receptors that include the CD81 tetraspanin and the SR-B1 scavenger receptor. *J Biol Chem* 2003;278:41624–41630.
40. Cormier EG, Tsamis F, Kajumo F, Durso RJ, Gardner JP, Dragic T. CD81 is an entry coreceptor for hepatitis C virus. *Proc Natl Acad Sci USA* 2004;101:7270–7274.
41. VanCompernolle SE, Wiznycia AV, Rush JR, Dhanasekaran M, Baures PW, Todd SC. Small molecule inhibition of hepatitis C virus E2 binding to CD81. *Virology* 2003;314:371–380.

42. Wagner CE, Mohler ML, Kang GS, Miller DD, Geisert EE, Chang YA, Fleischer EB, Shea KJ. Synthesis of 1-boraadamantaneamine derivatives with selective astrocyte vs C6 glioma antiproliferative activity. A novel class of anti-hepatitis C agents with potential to bind CD81. *J Med Chem* 2003;46:2823–2833.
43. Peiris JS, Guan Y, Yuen KY. Severe acute respiratory syndrome. *Nat Med* 2004; 10:S88–S97.
44. Li W, Moore MJ, Vasilieva N, Sui J, Wong SK, Berne MA, et al. Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus. *Nature* 2003;426:450–454.
45. Xiao X, Chakraborti S, Dimitrov AS, Gramatikoff K, Dimitrov DS. The SARS-CoV S glycoprotein: expression and functional characterization. *Biochem Biophys Res Commun* 2003;312:1159–1164.
46. Wong SK, Li W, Moore MJ, Choe H, Farzan M. A 193-amino acid fragment of the SARS coronavirus S protein efficiently binds angiotensin-converting enzyme 2. *J Biol Chem* 2004;279:3197–3201.
47. Li F, Li W, Farzan M, Harrison SC. Structure of SARS coronavirus spike receptor-binding domain complexed with receptor. *Science* 2005;309:1864–1868.
48. Schwarz-Linek U, Werner JM, Pickford AR, Gurusiddappa S, Kim JH, Pilka ES, et al. Pathogenic bacteria attach to human fibronectin through a tandem beta-zipper. *Nature* 2003;423:177–181.
49. Pilka ES, Werner JM, Schwarz-Linek U, Pickford AR, Meenan NA, Campbell ID, Potts JR. Structural insight into binding of *Staphylococcus aureus* to human fibronectin. *FEBS Lett* 2006;580:273–277.
50. Schwarz-Linek U, Pilka ES, Pickford AR, Kim JH, Hook M, Campbell ID, Potts JR. High affinity streptococcal binding to human fibronectin requires specific recognition of sequential F1 modules. *J Biol Chem* 2004;279:39017–39025.
51. Raibaud S, Schwarz-Linek U, Kim JH, Jenkins HT, Baines ER, Gurusiddappa S, Hook M, Potts JR. *Borrelia burgdorferi* binds fibronectin through a tandem beta-zipper, a common mechanism of fibronectin binding in staphylococci, streptococci, and spirochetes. *J Biol Chem* 2005;280:18803–18809.
52. Deane JE, Ryan DP, Sunde M, Maher MJ, Guss JM, Visvader JE, Matthews JM. Tandem LIM domains provide synergistic binding in the LMO4:Ldb1 complex. *EMBO J* 2004;23:3589–3598.
53. Ryan DP, Sunde M, Kwan AH, Marianayagam NJ, Nancarrow AL, Vanden Hoven RN, et al. Identification of the key LMO2-binding determinants on Ldb1. *J Mol Biol* 2006;359:66–75.
54. Matthews JM, Visvader JE. LIM-domain-binding protein 1: a multifunctional cofactor that interacts with diverse proteins. *EMBO Rep* 2003;4:1132–1137.
55. Hammond SM, Crable SC, Anderson KP. Negative regulatory elements are present in the human LMO2 oncogene and may contribute to its expression in leukemia. *Leuk Res* 2005;29:89–97.
56. Gadek TR. Strategies and methods in the identification of antagonists of protein–protein interactions. *Biotechniques Suppl* 2003;21–24.
57. Arkin MR, Wells JA. Small-molecule inhibitors of protein–protein interactions: progressing towards the dream. *Nat Rev Drug Discov* 2004;3:301–317.
58. Vassilev LT, Vu BT, Graves B, Carvajal D, Podlaski F, Filipovic Z, et al. *In vivo* activation of the p53 pathway by small-molecule antagonists of MDM2. *Science* 2004;303:844–848. Epub 2004 Jan 2002.
59. Benhar I. Biotechnological applications of phage and cell display. *Biotechnol Adv* 2001;19:1–33.

60. Clackson T, Hoogenboom HR, Griffiths AD, Winter G. Making antibody fragments using phage display libraries. *Nature* 1991;352:624–628.
61. Griffiths AD, Duncan AR. Strategies for selection of antibodies by phage display. *Curr Opin Biotechnol* 1998;9:102–108.
62. Hoogenboom HR, de Bruine AP, Hufton SE, Hoet RM, Arends JW, Roovers RC. Antibody phage display technology and its applications. *Immunotechnology* 1998;4:1–20.
63. Fields S, Song O. A novel genetic system to detect protein–protein interactions. *Nature* 1989;340:245–246.
64. Graessmann A, Graessmann M, Mueller C. Microinjection of early SV40 DNA fragments and T antigen. *Methods Enzymol* 1980;65:816–825.
65. Morgan DO, Roth RA. Analysis of intracellular protein function by antibody injection. *Immunol Today* 1988;9:84–88.
66. Valle G, Jones EA, Colman A. Anti-ovalbumin monoclonal antibodies interact with their antigen in internal membranes of *Xenopus* oocytes. *Nature* 1982;300:71–74.
67. Burke B, Warren G. Microinjection of mRNA coding for an anti-Golgi antibody inhibits intracellular transport of a viral membrane protein. *Cell* 1984;36:847–856.
68. Lobato MN, Rabbitts TH. Intracellular antibodies as specific reagents for functional ablation: future therapeutic molecules. *Curr Mol Med* 2004;4:519–528.
69. Tanaka T, Lobato MN, Rabbitts TH. Single domain intracellular antibodies: a minimal fragment for direct *in vivo* selection of antigen-specific intrabodies. *J Mol Biol* 2003;331:1109–1120.
70. Das D, Suresh MR. Producing bispecific and bifunctional antibodies. *Methods Mol Med* 2005;109:329–346.
71. Boder ET, Wittrup KD. Yeast surface display for screening combinatorial polypeptide libraries. *Nat Biotechnol* 1997;15:553–557.
72. Lipovsek D, Pluckthun A. *In-vitro* protein evolution by ribosome display and mRNA display. *J Immunol Methods* 2004;290:51–67.
73. Tse E, Lobato MN, Forster A, Tanaka T, Chung GT, Rabbitts TH. Intracellular antibody capture technology: application to selection of intracellular antibodies recognising the BCR-ABL oncogenic protein. *J Mol Biol* 2002;317:85–94.
74. Visintin M, Settanni G, Maritan A, Graziosi S, Marks JD, Cattaneo A. The intracellular antibody capture technology (IACT): towards a consensus sequence for intracellular antibodies. *J Mol Biol* 2002;317:73–83.
75. Lobato MN, Rabbitts TH. Intracellular antibodies and challenges facing their use as therapeutic agents. *Trends Mol Med* 2003;9:390–396.
76. Hudson PJ, Souriau C. Engineered antibodies. *Nat Med* 2003;9:129–134.
77. Hinoda Y, Sasaki S, Ishida T, Imai K. Monoclonal antibodies as effective therapeutic agents for solid tumors. *Cancer Sci* 2004;95:621–625.
78. Kritzer JA, Lear JD, Hodsdon ME, Schepartz A. Helical beta-peptide inhibitors of the p53–hDM2 interaction. *J Am Chem Soc* 2004;126:9468–9469.
79. Yin H, Lee GI, Park HS, Payne GA, Rodriguez JM, Sebti SM, Hamilton AD. Terphenyl-based helical mimetics that disrupt the p53/HDM2 interaction. *Angew Chem Int Ed Engl* 2005;44:2704–2707.
80. Yin H, Lee GI, Sedey KA, Rodriguez JM, Wang HG, Sebti SM, Hamilton AD. Terephthalamide derivatives as mimetics of helical peptides: disruption of the Bcl-x(L)/Bak interaction. *J Am Chem Soc* 2005;127:5463–5468.

81. Yin H, Lee GI, Sedey KA, Kutzki O, Park HS, Orner BP, Ernst JT, Wang HG, Sebti SM, Hamilton AD. Terphenyl-based Bak BH3 alpha-helical proteomimetics as low-molecular-weight antagonists of Bcl-xL. *J Am Chem Soc* 2005;127:10191–10196.
82. Walensky LD, Kung AL, Escher I, Malia TJ, Barbuto S, Wright RD, Wagner G, Verdine GL, Korsmeyer SJ. Activation of apoptosis *in vivo* by a hydrocarbon-stapled BH3 helix. *Science* 2004;305:1466–1470.
83. Sadowsky JD, Schmitt MA, Lee HS, Umezawa N, Wang S, Tomita Y, Gellman SH. Chimeric (alpha/beta + alpha)-peptide ligands for the BH3-recognition cleft of Bcl-XL: critical role of the molecular scaffold in protein surface recognition. *J Am Chem Soc* 2005;127:11966–11968.
84. Xiong JP, Stehle T, Zhang R, Joachimiak A, Frech M, Goodman SL, Arnaout MA. Crystal structure of the extracellular segment of integrin alpha Vbeta3 in complex with an Arg-Gly-Asp ligand. *Science* 2002;296:151–155. Epub 2002 Mar 2007.
85. Altmeyer R. Virus attachment and entry offer numerous targets for antiviral therapy. *Curr Pharm Des* 2004;10:3701–3712.
86. Moore JP, Doms RW. The entry of entry inhibitors: a fusion of science and medicine. *Proc Natl Acad Sci USA* 2003;100:10598–10602.
87. Web: Drugs used in the treatment of HIV infection. http://wwwfdagov/oashi/aids/viralshtml. Accessed 9 August 2006.
88. Ernst JT, Kutzki O, Debnath AK, Jiang S, Lu H, Hamilton AD. Design of a protein surface antagonist based on alpha-helix mimicry: inhibition of gp41 assembly and viral fusion. *Angew Chem Int Ed Engl* 2002;41:278–281.
89. Nozaki A, Ikeda M, Naganuma A, Nakamura T, Inudoh M, Tanaka K, Kato N. Identification of a lactoferrin-derived peptide possessing binding activity to hepatitis C virus E2 envelope protein. *J Biol Chem* 2003;278:10162–10173.
90. English EP, Chumanov RS, Gellman SH, Compton T. Rational development of beta-peptide inhibitors of human cytomegalovirus entry. *J Biol Chem* 2006;281:2661–2667.
91. Petros AM, Olejniczak ET, Fesik SW. Structural biology of the Bcl-2 family of proteins. *Biochim Biophys Acta* 2004;1644:83–94.
92. Wang JL, Liu D, Zhang ZJ, Shan S, Han X, Srinivasula SM, Croce CM, Alnemri ES, Huang Z. Structure-based discovery of an organic compound that binds Bcl-2 protein and induces apoptosis of tumor cells. *Proc Natl Acad Sci USA* 2000;97:7124–7129.
93. Qian J, Voorbach MJ, Huth JR, Coen ML, Zhang H, Ng SC, et al. Discovery of novel inhibitors of Bcl-xL using multiple high-throughput screening platforms. *Anal Biochem* 2004;328:131–138.
94. Tzung SP, Kim KM, Basanez G, Giedt CD, Simon J, Zimmerberg J, Zhang KY, Hockenbery DM. Antimycin A mimics a cell-death-inducing Bcl-2 homology domain 3. *Nat Cell Biol* 2001;3:183–191.
95. Sun H, Nikolovska-Coleska Z, Yang CY, Xu L, Liu M, Tomita Y, et al. Structure-based design of potent, conformationally constrained Smac mimetics. *J Am Chem Soc* 2004;126:16686–16687.
96. Sun H, Nikolovska-Coleska Z, Yang CY, Xu L, Tomita Y, Krajewski K, Roller PP, Wang S. Structure-based design, synthesis, and evaluation of conformationally constrained mimetics of the second mitochondria-derived activator of caspase that target the X-linked inhibitor of apoptosis protein/caspase-9 interaction site. *J Med Chem* 2004;47:4147–4150.
97. Sun H, Nikolovska-Coleska Z, Chen J, Yang CY, Tomita Y, Pan H, Yoshioka Y, Krajewski K, Roller PP, Wang S. Structure-based design, synthesis and biochemical testing of novel and potent Smac peptido-mimetics. *Bioorg Med Chem Lett* 2005;15:793–797.

98. Li L, Thomas RM, Suzuki H, De Brabander JK, Wang X, Harran PG. A small molecule Smac mimic potentiates TRAIL- and TNFalpha-mediated cell death. *Science* 2004;305:1471–1474.
99. Schimmer AD, Welsh K, Pinilla C, Wang Z, Krajewska M, Bonneau MJ, et al. Small-molecule antagonists of apoptosis suppressor XIAP exhibit broad antitumor activity. *Cancer Cell* 2004;5:25–35.
100. White PW, Titolo S, Brault K, Thauvette L, Pelletier A, Welchner E, et al. Inhibition of human papillomavirus DNA replication by small molecule antagonists of the E1–E2 protein interaction. *J Biol Chem* 2003;278:26765–26772.
101. Wang Y, Coulombe R, Cameron DR, Thauvette L, Massariol MJ, Amon LM, et al. Crystal structure of the E2 transactivation domain of human papillomavirus type 11 bound to a protein interaction inhibitor. *J Biol Chem* 2004;279:6976–6985.
102. Shultz MD, Ham YW, Lee SG, Davis DA, Brown C, Chmielewski J. Small-molecule dimerization inhibitors of wild-type and mutant HIV protease: a focused library approach. *J Am Chem Soc* 2004;126:9886–9887.
103. Aya T, Hamilton AD. Tetrabiphenylporphyrin-based receptors for protein surfaces show sub-nanomolar affinity and enhance unfolding. *Bioorg Med Chem Lett* 2003;13: 2651–2654.
104. Groves K, Wilson AJ, Hamilton AD. Catalytic unfolding and proteolysis of cytochrome C induced by synthetic binding agents. *J Am Chem Soc* 2004;126:12833–12842.
105. Gradl SN, Felix JP, Isacoff EY, Garcia ML, Trauner D. Protein surface recognition by rational design: nanomolar ligands for potassium channels. *J Am Chem Soc* 2003;125: 12668–12669.
106. Park HS, Lin Q, Hamilton AD. Modulation of protein–protein interactions by synthetic receptors: design of molecules that disrupt serine protease–proteinaceous inhibitor interaction. *Proc Natl Acad Sci USA* 2002;99:5105–5109.
107. Mecca T, Consoli GM, Geraci C, Cunsolo F. Designed calix[8]arene-based ligands for selective tryptase surface recognition. *Bioorg Med Chem* 2004;12:5057–5062.
108. Francese S, Cozzolino A, Caputo I, Esposito C, Martino M, Gaeta C, Troisi F, Neri P. Transglutaminase surface recognition by peptidocalix[4]arene diversomers. *Tetrahedron Lett* 2005;46:1611–1615.
109. Aachmann FL, Otzen DE, Larsen KL, Wimmer R. Structural background of cyclodextrin–protein interactions. *Protein Eng* 2003;16:905–912.

# METHOD DEVELOPMENT FOR PRECLINICAL BIOANALYTICAL SUPPORT

Masood Khan[1] and Naidong Weng[2]

[1]*Covance Laboratories Inc., Chantilly, Virginia*
[2]*Johnson & Johnson Pharmaceutical Research & Development, Raritan, New Jersey*

**Contents**

1 Preclinical Bioanalytical Support Using Liquid Chromatography with Tandem Mass Spectrometers (LC-MS/MS)
- 1.1 Introduction
- 1.2 Regulatory Requirement
- 1.3 Batch Failure Rate
- 1.4 Narrowing the Gap for Incurred Sample Analysis
- 1.5 Control Animal Samples
- 1.6 Metabolic Selectivity Coverage During Drug Development
- 1.7 Method Development Strategy
- 1.8 Method Automation Strategy: From Preclinical to Clinical
- 1.9 Matrix Effects and Recovery
- 1.10 Effect of Dosing Vehicles
- 1.11 Carry-over and Contamination
- 1.12 Misconception About Stable Isotope-Labeled Internal Standard
- 1.13 Troubleshooting Strategy
- 1.14 Conclusion

2 Preclinical Bioanalytical Support Using Ligand–Binding Assay
- 2.1 Introduction
- 2.2 Ligand-Binding Assay Technology
- 2.3 Using Ligand-Binding Assay in a GLP Environment
- 2.4 Ligand-Binding Assay Specific Challenges
- 2.5 Responding to the Challenge of Ligand-Binding Assay

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

2.6 Using Commercial Kit Based Assays
2.7 Industry and Regulatory Perspective
2.8 Conclusion
References

## 1 PRECLINICAL BIOANALYTICAL SUPPORT USING LIQUID CHROMATOGRAPHY WITH TANDEM MASS SPECTROMETERS (LC-MS/MS)

### 1.1 Introduction

Preclinical development represents a critical stage in the progression from discovery to marketed pharmaceutical drug candidates that have passed initial discovery screening and are identified to possess some drug-like properties. Serious resources and financial commitments are now being made to vigorously test the drug candidates before they enter clinical trials. From a regulatory point of view, it is also essential to ensure the welfare of the volunteers and patients participating in the clinical trials by vigorously testing the safety using appropriate animal models. A series of questions concerning the toxicity, pharmacokinetic (PK) parameters, safety assessment, formulation optimization, and so on need to be answered.

Bioanalytical support plays a pivotal role in answering these questions. Timely bioanalytical support is essential to improve the success rate of drug candidates moving along the preclinical development pipeline and allows decisions to be made early to modify/improve the drug candidates or terminate the program. By definition, the goal of preclinical development support is attrition of the drug candidates with less favorable drug-like properties. The financial benefit of eliminating candidates from the drug development pipeline has been highlighted by Lee and Kerns [1].

Bioanalytical support at the preclinical development phase presents some unique challenges. The first unique challenge is the *Good Laboratory Practice* (GLP) *requirement*. Unlike discovery support, where bioanalytical support does not typically run under GLP compliance and oftentimes generic fast gradient methods are used to rank the drug candidates, preclinical bioanalytical support is under strict GLP regulations, where the data generated undergo rigorous regulatory scrutiny. The mindset is certainly shifted from the more creative discovery bioanalytical support to more compliance-oriented development bioanalytical support. Cutting-edge technology is only used at preclinical bioanalytical support with caution and only after these technologies meet the challenge of vigorous validations for staying in compliance will they be widely accepted as one of the primary tools for preclinical bioanalytical support.

The second unique challenge is to have good *metabolic selectivity coverage*. Usually at the early stages of preclinical studies, information on metabolites may not be well characterized and a definitive human absorption, distribution, metabolism, and excretion (ADME) study is usually not run until clinical studies. New microdosing paradigms are evolving, which, over time, may garner human metabolism information earlier. Yet, from the bioanalytical point of view, the limited information on metabolites normally available presents a challenge. It is often uncertain

how many metabolites should be included in the assay. Insufficient coverage of metabolite measurement would leave inadequate coverage of the safety margin in clinical trials should a minor metabolite in an animal species become a significant one in humans. On the other hand, methods measuring too many metabolites could result in high failure rates, a highly undesirable feature that is not only costly but also leads to regulatory question regarding the validity of the entire method.

The third unique characteristic of the preclinical bioanalytical support is the very quick turnaround time for *method development, validation, and sample analysis* for multiple animal species (mouse, rat, dog, rabbit, monkey, etc.) and oftentimes for multiple matrices (plasma, urine, tissue, liver, brain, etc.). The requirements for the skill set and infrastructure of a bioanalytical laboratory supporting preclinical work could be different from those supporting clinical bioanalytical works. For clinical bioanalytical support, it is not unusual to have a group of scientists dedicated to method development and validation and then the methods are transferred to other chemists to run large sets of routine sample analysis that could last for several months to several years. For preclinical bioanalytical support, such a transition is also possible but requires a more vigorous method ruggedness test and more frequent communication between the method developer and chemists running the samples. Also, there is a need for a very quick turnaround as well as a heavy load on method development/validation instead of routine sample analysis. It is still desirable to free method developers from routine sample analysis so that they can focus on quality method development. In general, the ratio of method developers to sample analysts should be higher for supporting preclinical LC-MS/MS than for clinical LC-MS/MS.

The fourth unique characteristic is that at this stage one needs to think about a strategy of *streamlining the process at various stages of bioanalytical support*. Table 1 summarizes the major characteristics for the discovery, preclinical development, and clinical bioanalytical support. These three stages should not be viewed as three separate and consecutive ones, since each can extend well into the next stage. Streamlining the sample analysis process during the preclinical stage presents unique challenges in comparison with discovery or clinical bioanalytical supports, even though great effort has been made by many scientists in the pharmaceutical industry. In the discovery stage, a few generic methods are applied to many candidates, which make it easier to adopt a universal strategy to streamline and automate the process with the ultimate goal of generating data as quickly as possible to enable eliminating poor candidates. Data review by a quality assurance unit is not needed and reporting is usually kept quite simple. When the projects get transferred to the preclinical stage, an individual unique method needs to be developed and validated for each candidate to ensure the method meets the specific requirements specified in standard operating procedures (SOPs) as well as in regulatory guidance. The method will then be used for analyzing samples. Depending on the nature of the program, the sample numbers can vary quite significantly. The scientist needs to make sure that automation is applicable for real sample analysis [2, 3]. At the preclinical stage, emphasis is usually on method reproducibility and method specificity to ensure the highest quality data. However, one should not overlook the benefit of using automation when appropriate. Automation does reduce human errors and improve method reproducibility. Once the drug candidate moves into the clinical stage, automation and other means of improving sample processing become more important. The cost

**TABLE 1 Bioanalytical Support for Discovery, Preclinical, and Clinical Studies**

| | Discovery | Preclinical | Clinical |
|---|---|---|---|
| Regulatory requirement | Non-GLP | GLP | Although GLPs only apply to preclinical studies, clinical falls under the guidance and by practice the same standards as the GLPs |
| LC-MS/MS method | Generic | Tuned to compound; requires extensive method development | Tuned to compound; extensive method development but could leverage preclinical methods |
| Validation | Minimal | Extensive validation to multiple matrices in multiple species | Extensive validation but limited to human samples; specificity tests to coadministered compounds |
| Validation strategy | Abbreviated method validation; minimal stability test | Full validation for one species and partial validation for other species including incurred sample reproducibility; has large curve range that may be problematic due to carry-over or ionization saturation | Full validation including incurred sample analysis; very sensitive method may be needed for high potency candidate |
| Sample analysis | Small sets of samples per compound but hundreds or thousands of compounds | Moderate numbers of samples for dozens of compounds | Large numbers of samples for very few compounds |
| Sample analysis strategy | Streamline process from sample collection to data generation; use of generic methods also allows easier set up for automation | Due to relatively small sets of samples and various methods tuned for each compound, automation is feasible but may not be feasible for some methods due to limited sample volumes | Quick turnaround time for sample analysis (automated sample preparation, multiplexing, UPLC) to allow data to be released |

of automating clinical bioanalytical methods can easily be justified. The same clinical bioanalytical method can be used by multiple bioanalytical chemists within or even among different organizations to meet the demand of analyzing large sets of samples within a short period of time. The nature of clinical samples also makes automation more amenable for routine sample analysis. Sample volumes are typically several-fold higher than those in preclinical studies and are more suitable for automation. Like discovery bioanalytical support, sample analysis speed is likely to be the rate-limiting step for clinical bioanalytical support. Approaches for expediting sample analysis, such as multiplexing two high-performance liquid chromatography (HPLC) units into one mass spectrometer or ultra-pressure liquid chromatography (UPLC) using sub 2 μm columns, are frequently used.

The literature on general bioanalytical method development, validation, and sample analysis has been extensively reviewed [4–14], but a strategy focused on preclinical bioanalytical support would certainly be useful. In this chapter, we share our practical experience on preclinical bioanalytical support.

### 1.2 Regulatory Requirement

A distinguishing feature of preclinical bioanalytical support is the regulatory compliance. Standard operating procedures (SOPs) are used to address many aspects of the regulatory requirements, ranging from instrument qualification, maintenance, and calibration, to bioanalytical method validation, sample analysis, data management, and sample archiving. The goal of these SOPs is to ensure the highest data integrity that can pass regulatory scrutiny. These SOPs are usually drafted based on specific regulations and guidance from the U.S. Food and Drug Administration (FDA) and international regulatory agencies, based on various federal/state requirements and based on specific policies at each institution. The single most important regulatory guidance is the FDA guidance on bioanalytical method validation [15], which provides a general guideline on establishing important parameters, including method specificity/selectivity, sensitivity, linearity, accuracy, precision, stability, matrix effects, recovery, carry-over, and contamination. The method should also demonstrate reproducibility and accuracy of measurements for incurred samples and be suitable for its intended use. However, it is up to each institution to formulate its own SOPs based on this general guidance. It is well recognized that there is a gap between the spiked quality control (QC) samples and incurred samples [16, 17]. It is generally accepted industry-wide and by the regulatory agencies that the method should meet criteria such as selectivity, sensitivity, linearity, accuracy, precision, and stability using quality control samples that are prepared by spiking known amounts of analytes to the control matrix. However, QC samples may not entirely reflect the nature of the incurred samples and this fact is often neglected by the bioanalytical chemists, resulting in over- or underestimation of analytes due to various factors. Therefore, efforts should be made to address this gap to satisfy the regulatory compliance needs.

### 1.3 Batch Failure Rate

Batch failure rate has been particularly useful to assess the method ruggedness or reliability. High failure rates usually indicate inherent shortcomings of a given

method and could be used by the FDA to reject the submitted data. Excluding execution errors, good bioanalytical methods should achieve at least 80% passing rate. All failed batches should also be documented and reported in both validation and sample analysis reports. Under no circumstances should the failed batches be omitted from the report. If the method has failed to demonstrate the acceptable batch passing rates during sample analysis, analysis should be placed on-hold and an investigation for the root cause should be initiated. Based on the conclusions of the root-cause analysis, the method may need to be modified and revalidated. A decision on analyzing samples in the failed batches should be judicious and the reanalysis should be authorized after an investigation. Testing into compliance by repeatedly reinjecting the failed batch until it passes acceptance criteria is definitely a red flag for regulatory authorities. Excessive manual peak reintegration, especially for standards and quality control samples, is also an indication of inherent problems in method ruggedness or data integrity and this practice should be avoided. One should be aware that, with multiple analytes (drug candidates with multiple metabolites), the overall batch failure rates can be significantly worse even if each single analyte has an acceptable 80% passing rate. Therefore, including too many metabolites in a method could potentially introduce bias to otherwise useful data from the parent compound and its most important metabolites. One approach is to have a separate non-GLP method for those metabolites whose significance to humans in not yet fully understood [18]. For the drug and significant metabolites, bioanalytical methods intended for GLP studies should be vigorously tested and validated to have a thorough assessment of the method reproducibility and ruggedness. It is costly and time consuming to troubleshoot the methods in the middle of sample analysis, especially for preclinical studies where quick turnaround is required.

### 1.4 Narrowing the Gap for Incurred Sample Analysis

Due to potential differences such as protein binding and metabolite conversion, quality control samples, which are prepared by spiking analytes of interest into the control blank matrix, cannot totally mimic incurred samples. This difference may lead to a significant bias for the quantitative bioanalysis and must be carefully evaluated. The importance of assessing method accuracy and reproducibility using incurred samples has been emphasized at the 3rd AAPS/FDA Bioanalytical Method Validation Workshop (commonly referred to as the Crystal City III meeting) [19]. It was suggested that evaluation of assay reproducibility and accuracy for the incurred samples needs to be performed on each species used for GLP toxicology studies. Method reproducibility could be assessed by repeat analysis of individual or pooled incurred samples. Analyzing individual incurred samples will generate a second set of data. A clear written policy should be established for how to handle this second set of data. One approach that avoids generating a second set of data is to use pooled samples. Pooled samples provide the same type of data as individual incurred samples, but they are not reportable as study data. Nevertheless, the argument against this pooling is that bias from individual samples could be averaged out. For example, dosing vehicle that can cause significant ion suppression and therefore biased quantitation is more predominant in the early time points and pooling with later time points may diminish such an effect. One approach to solve this issue is to pool samples from similar time points. Assessment of method accu-

racy for the incurred samples presents a major challenge. It would be extremely difficult to ascertain the method accuracy for the incurred samples since the true value is unknown. The feasible approaches to ascertain some level of method accuracy for incurred samples might be the addition of analytes to the incurred samples (standard addition) or comparison of results from two orthogonal methods. Even with these assessments, situations causing bias for the quantitation of incurred samples could still exist. One such possibility is the instability of the incurred samples during the sample collection procedure and sample storage. This may be even more pronounced for the large molecule therapeutics. Neither the standard addition approach nor the orthogonal method approach would detect such a problem. Therefore, as a bioanalytical scientist, one must be judicious when developing methods and make sure any issues that can cause potential method bias or poor reproducibility are resolved prior to the method validation. One common mistake when developing separate assays for parent compound and metabolite is the failure to investigate selectivity and stability toward the potential influence of the drug and other metabolites on each other. Each individual method may only contain the analytes to be measured in that method which does not typically reflect the entire content of incurred samples. Potential conversion between known analytes or between known and unknown metabolites may occur during the sample collection, storage, preparation, or postextraction, leading to quantitation bias. The extent of the bias depends on the conversion rate and the concentration ratio among these interconverting compounds. Typically, concentrations for the calibration standards and QCs are prepared in such a way that both concentrations increase or decrease proportionally. This approach allows the use of one common stock solution to prepare standards or QCs at lower concentrations. In most cases, this approach would be fine since for most of the drug candidates such a proportioned concentration change is indeed observed (Fig. 1a). However, for drug candidates such as prodrugs, the concentration ratio of prodrug versus active drug could be lopsided, as shown in Fig. 1b. At early time points, the prodrug concentration can be significantly higher than the drug concentration. Because of the labile nature of the prodrug, the drug concentration may be overestimated because of the conversion of prodrug to drug during storage, extraction, postextraction, and in the LC-MS/MS source. One approach to investigating this problem is to prepare additional quality control samples at nonequivalent concentrations.

### 1.5 Control Animal Samples

One important aspect of bioanalytical preclinical support is to analyze samples from the control animal groups, which are not given a dose of the drug candidate. Any positive finding in these control animal samples indicates potential contamination in the dosing, collection, storage, sample preparation, or analysis process and usually warrants a thorough investigation. In order to minimize potential contamination during the sample preparation and analysis, these control animal samples may be analyzed in a separate batch from the samples from dosed animals. Multiple matrix blank samples are also analyzed along with these control animal samples to assist the investigation into the source of contamination, should it occur. Determining whether the measured concentrations in the control animals follow any pharmacokinetic (PK) profile or the parent/metabolite ratio is consistent with *in vivo* observations

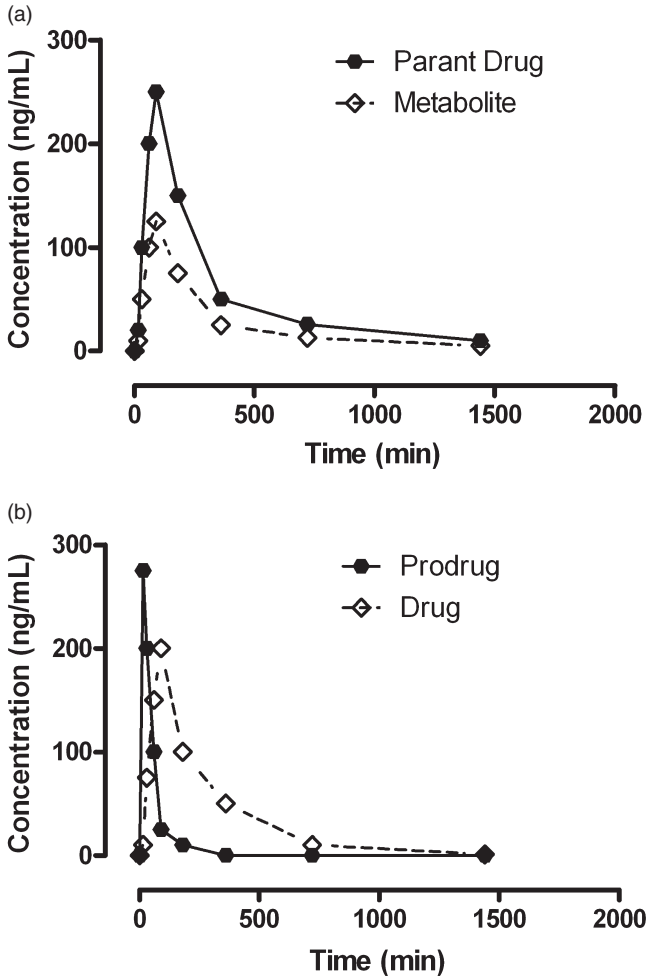


**FIGURE 1** Typical PK profiles for parent drug/metabolite and prodrug/drug. Note the parallel concentration profile between parent drug and metabolite in (a) and the lopsided concentration profile between prodrug and drug in (b).

also provides good insight into the source of contamination. It might be necessary to have a dedicated laboratory and instruments to analyze the control animal samples if a drug candidate is highly potent and is dosed in very small amount.

### 1.6 Metabolic Selectivity Coverage During Drug Development

An industry report on drug metabolites in safety testing (commonly referred to as MIST) was published in 2002 [20]. After deliberation between industry and the FDA, a draft FDA *Guidance for Industry: Safety Testing of Drug Metabolites* was issued in 2005 [21]. Extensive characterization of the pharmacokinetics of unique and/or major human metabolites would be very useful to correlate metabolites and toxicological observations. At the Crystal City III meeting, it was agreed that a

"tiered" method validation approach could be used for measuring unique and/or major metabolites in early drug development. A tiered validation approach would allow information to be gathered on human and animal metabolites with scientifically valid methods while deferring more labor-intensive GLP validations to later stages in the drug development when significant metabolites are identified. In order to identify which metabolites to measure, metabolites formed in several animal species and human liver microsomes *in vitro* are identified and their concentrations are measured. Yet, there is no guarantee that there is a direct correlation between the metabolites thus observed and the human metabolites observed *in vivo*. If unique human metabolites are found or major metabolites were observed at concentrations without safety margin from toxicology studies, it could present some unique challenges to bioanalytical scientists. It would be beneficial if human metabolites could be identified early during the drug development. With the advancement and availability of the highly sensitive accelerated mass spectrometry (AMS), one can perform exploratory human ADME by using the microdosing scheme suggested in the guidance for exploratory studies or by combining exploratory human ADME with single ascending dosing (SAD) [22].

In order to obtain meaningful results from the study, one also needs to pay attention to the metabolite stability *ex vivo*. Metabolite stability in plasma could be affected by enzymes, pH, anticoagulants, storage conditions, and freeze/thaw cycles. It is well known that acylglucuronides can degrade back to the original drug under neutral or basic pH [23]. Adjusting the pH to 3–5 would stabilize the acylglucuronides. Candidates can also be subject to hydrolysis due to esterases. An enzymatic inhibitor (sometimes more than one) is sometimes needed to stabilize both the parent compound and its metabolite(s). One should be aware that commercially purchased plasma may not have the same enzymatic activity as the plasma harvested from freshly drawn blood. It is advisable to confirm the metabolic stability using fresh matrix.

Another consideration related to metabolic specificity is the in-source fragmentation occurring at the LC-MS/MS interface. Under high temperature (300–600 °C) and high voltage (2–5 kV), phase II glucuronide metabolites may lose the glucuronide to form the original compound, which will show up in the aglycone analyte channel. If there is no chromatographic separation between them, overestimation of the aglycone compound may occur.

## 1.7 Method Development Strategy

Vigorous method development and optimization is essential in order to narrow the gap between spiked QC and incurred samples. The spiked QC samples only contain the added analytes while the incurred samples may contain additional known and unknown metabolites. In the early stages of preclinical bioanalytical support, information related to metabolite characteristics and stability may not be fully developed. As mentioned previously, metabolic specificity of the method needs to be carefully evaluated. Another unique feature of preclinical bioanalytical support is the dosing vehicle effect on the quantitation, especially when limited or minimal sample preparation and fast gradients are used.

With attaining high quality data as the ultimate goal, bioanalytical scientists strive to develop methods with good selectivity, high sensitivity, and faster throughput.

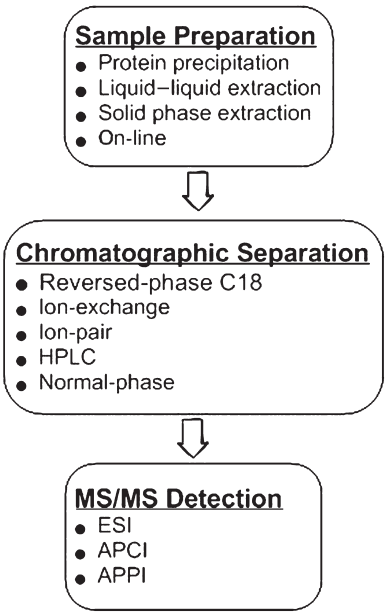


**FIGURE 2** Three distinguished but interrelated stages (sample preparation, chromatographic separation, and MS/MS detection) of analysis of biological samples using LC-MS/MS.

Bioanalytical methods (chromatographic) typically consist of three important but interrelated parts: sample preparation, chromatography, and MS detection (Fig. 2). Many techniques can be employed for each of these parts. When developing a robust method, one needs to consider all three parts as an integrated system. Sometimes trade-offs need to be balanced. For example, if a simple sample preparation procedure, such as protein precipitation, is used, one may need to compensate for the potential ion suppression or in-source conversion of glucuronide metabolites to parent compound by using a more extensive chromatographic separation. Ideally, the mechanism for sample extraction and chromatography should be orthogonal to provide better method selectivity. A combination of reversed-phase solid-phase extraction (SPE) and reversed-phase LC may not provide sufficient separation power of metabolites or interferences. Of course, one will always need to keep in mind the analyte integrity during sample extraction, postextraction, and chromatography. Although liquid chromatography with ultraviolet or fluorescence detection (LC-UV/FLU) or gas chromatography with mass spectrometer detection (GC-MS) methods are still widely used and sometimes offer better methods than LC-MS/MS [24], the use of LC-MS/MS, owing to its intrinsic advantages, has nevertheless grown exponentially in the last decade and is used extensively for the preclinical bioanalytical support for small molecule therapeutics. The principle of MS is the production of ions from analyzed compounds that are separated or filtered based on their mass to charge ratio ($m/z$). Most applications for quantitative bioanalysis use tandem mass spectrometers (MS/MS) that employ two mass analyzers: one for the precursor ion in the first quadrupole and the other for the product ion in the third quadrupole after the collision–activated dissociation of the precursor ion in a collision cell

(second quadrupole). Between the high pressure LC and the MS operated under a high-vacuum environment, interface connections that operate at atmospheric pressure, such as electrospray ionization (ESI), atmospheric-pressure chemical ionization (APCI), and atmospheric-pressure photoionization (APPI), have matured into highly reliable systems necessary for quantitative LC-MS/MS bioanalysis. On the chromatographic side, reversed-phase columns are extensively used but separation based on other retention mechanisms can also be used complementarily (e.g., normal-phase chromatography for chiral separations).

***Sample Preparation*** Protein precipitation (PPT), liquid/liquid extraction (LLE), and SPE are frequently used sample preparation techniques in the format of either individual tubes/cartridges or 96-well plates. Analytes of interest are released from protein when a protein precipitation reagent, such as organic solvents (e.g., acetonitrile, methanol), acids, or salt (e.g., ammonium sulfate), is added to the biological samples to denature the protein. These analytes stay in the supernatant after centrifugation and can be analyzed as is or can go through evaporation/reconstitution steps to make the injection compatible with the chromatographic condition. For highly protein-bound compounds, it is frequently necessary to release them from the proteins by adjusting the sample pH; otherwise, the analyte could be trapped in the precipitates. In principle, solid-phase extraction (SPE) is analogous to liquid/liquid extraction (LLE). As a liquid sample passes through the SPE, compounds are adsorbed onto the support or sorbent material. Interferences can then be selectively removed using washing solvents. Finally, the desired analytes may be selectively recovered by an elution solvent, resulting in a purified extract. The eluent can either be injected directly onto LC-MS/MS or go through the evaporation/reconstitution steps for further concentration enhancement and for compatibility with chromatographic conditions. Unlike protein precipitation, where universal extraction and recovery is usually achieved regardless of the chemical structures of the analytes, success using LLE and SPE to extract analytes largely depends on the chemistry of the molecules and the experience of the analyst. Although LLE and SPE usually give cleaner samples than protein precipitation and thus less matrix effects due to ion suppression and enhancement, their selectivity can result in differential recovery among parent compound and metabolites as well as the internal standards. This could place a limit on their use for simultaneous analysis of multiple analytes. Use of stable isotope-labeled (SIL) internal standards could compensate for this deficiency to some extent since the SIL internal standards usually have the same recovery and chromatographic characteristics as the unlabeled analytes.

***Chromatography*** With the initial sample clean-up using SPE, LLE, or PPT, unwanted compounds can still be present in higher concentrations than the analytes of interest. A second stage of cleanup, typically involving LC separation, further separates analytes of interest from the unwanted compounds. Without this further separation, those unwanted and present compounds in the MS/MS are not typically observed but present significant challenges. In the LC-MS interface, these compounds compete with analytes for ionization and cause inconsistent matrix effects that are detrimental to quantitative LC-MS/MS.

Reversed-phase LC has traditionally been used for quantitative LC-MS/MS. There are numerous types of reversed-phase columns, with the C18 column being

the most predominant. With increasing organic solvent concentrations in the mobile phase, the analyte retention decreases. However, one should be aware of the potential bimodal retention on the reversed-phase column due to the residual silanol group [25]. This bimodel retention may cause retention shifts during the run or irreproducibility of the method. Of particular interest is the use of monolithic columns (available in C18 format) operated at a high flow rate. Compared to a particulate column, the monolithic column has a reduced pressure drop but still maintains high separation efficiency at high mobile phase flow rates. This is due to its unique bimodal pore structure, which consists of macropores (2 μm) and mesopores (13 nm) [26]. The mesopores provide the surface area for achieving adequate capacity while the macropores allow high flow rates because of higher porosity, resulting in reduced flow resistance. Monolithic columns have become increasingly popular for use in ultrafast bioanalysis of drug candidates using tandem mass spectrometric detection [27–29]. Another recent development in column technology is the fused-core column, specially designed for hyperfast chromatography under the tradename Halo. Due to the small and uniform 2.7 μm particle size and fused-core technology, the analytes travel in the column with reduced diffusion, resulting in improved column efficiency and reduced column backpressure. Sub 2 μm particle columns used for UPLC has also drawn lots of attention. Special LC units, designed to operate under a backpressure as high as 15,000 psi, allow fast chromatography. Hydrophilic interaction chromatography (HILIC) became a popular choice in the last several years. HILIC is similar to normal-phase LC in that the elution is promoted by the use of polar mobile phase. However, unlike classic normal-phase LC, where the water to the mobile phase has to be kept to minimal but constant levels, water is present in a significant amount (>5%). HILIC also uses water-miscible polar organic solvents such as acetonitrile instead of water-immiscible solvents like hexane and chloroform. LC-MS/MS using HILIC on silica columns has been extensively studied and reviewed [30, 31].

***On-Line Extraction and Chromatography*** On-line extraction is another frequently used technology. With the improvement in instrumentation and column technology [32], on-line sample extraction that is coupled into LC-MS/MS has developed as an important tool in the past few years. In general, this technique employs turbulent flow chromatography, in which extraction columns are packed with retaining particles of large size (e.g., 50 μm) on a commercially available device or on a regular homemade column switch device. Biological samples are injected directly onto the extraction column, usually after dilution with an aqueous buffer. The analytes are then washed either forward or backward onto the analytical column for further chromatographic separation prior to the MS/MS detection. Common on-line extraction systems use typical reversed-phase columns for both the extraction and analysis. In our laboratory, we have successfully employed a weak-cation extraction (WCX) column and HILIC analytical column to analyze an extremely polar quaternary amine compound. This compound has no retention at all on a reversed-phase column but has excellent retention and peak shape on a HILIC column. The analyte in plasma was loaded onto a WCX column with an aqueous mobile phase and the WCX column was then washed with acetonitrile to remove endogenous compounds. The analyte was then eluted off the WCX column,

using acidified acetonitrile, onto the HILIC analytical column for a gradient elution.

***Analyte Adsorption Issue*** Loss of analytes during the collection of incurred samples, preparation of quality controls, storage, and analysis due to adsorption to the container should also be considered. Compound adsorption to storage containers often happens in sensitive LC-MS/MS assays, which causes nonlinear response and loss of sensitivity. The use of a zwitterionic detergent such as 3-[3-cholamidopropyl)-dimethylammonio]-1-propanesulfonate (CHAPS) as an additive in urine can prevent some analytes from adhering to surfaces during sample collection, storage, and preparation [33]. Protein binding of the drug candidate could also be significantly different for the quality control samples and incurred samples. If the analyte was not released from the protein prior to the extraction, quantitation bias could occur, depending on the extraction method. Xue et al. [34] observed the same issue with a liquid/liquid extraction method. For the plasma extraction method, two incubation steps were required after the addition of 5 mM ammonium acetate and the internal standard (stable isotope labeled) in acetonitrile to release the analyte bound to proteins prior to LLE with toluene.

## 1.8 Method Automation Strategy: From Preclinical to Clinical

As the candidates progress well, method development and validation for supporting human clinical trials will start well before the drugs enter the clinic. A well-established automated process, wherever possible, should be used to replace manual tasks early on in the preclinical stage. Automation results in greater performance consistency over time and in more reliable methods.

Standardization in sample collection can significantly enhance bioanalytical productivity. Preclinical samples could be collected in 96-well format [35] or transferred to 96-well format from a pierceable capped tube fitted with a nonleaking resealable polymer septum [36]. Automated 96-well plate technology for sample preparation is well established and accepted and has been shown to effectively replace manual tasks. The 96-well instruments can execute automated off-line extraction and sample cleanups. Automated SPE [37–40], LLE [41–46], and PPT [47, 48] can be performed in 96-well format. Both packed cartridges and disks in 96-well SPE formats have been successfully used. Further efficiency improvements were made when the organic extracts from SPE [49, 50] or LLE [51, 52] were injected directly onto the silica column with low aqueous–high organic mobile phases. While the commonly used SPE or LLE solvents are stronger elution solvents (not compatible) than a mobile phase on typical reversed-phase chromatography, they are weaker elution solvents (compatible) on the silica column with low aqueous–high organic mobile phases operated under HILIC. In this approach, the time-consuming and error-prone solvent evaporation and reconstitution steps are eliminated.

## 1.9 Matrix Effects and Recovery

The influence of matrix effects on quantitative bioanalytical LC-MS/MS has been well recognized [53, 54]. The U.S. Food and Drug Administration's (FDA) *Guidance*

*for Industry on Bioanalytical Method Validation* requires the assessment of matrix effects during method validation for quantitative bioanalytical LC-MS/MS methods [55]. In a broad sense, matrix effects represent the impact of the matrix on the analysis of the analytes, including ion suppression/enhancement, differentiated recovery, and altered stability, and is certainly not just limited to quantitative bioanalytical LC-MS/MS. However, it has been generally accepted and proposed in the industry guidance that matrix effects for small molecule bioanalytical analysis is defined as the analyte ionization suppression or enhancement in the presence of matrix components that could originate from endogenous compounds such as phospholipids [56], metabolites, coadministered drugs, internal standards [57], dosing vehicles [58–60], mobile phase additives [61], and plastic tubes [62]. These effects are more pronounced with ESI than with APCI [63, 64]. There are several commonly used ways of measuring matrix effects. One approach is to compare the MS response of the analyte spiked postextraction with that in a neat solution [65–67]. Since there is no extraction involved, any signal loss or enhancement in the postextraction spiked sample will be assumed due to matrix effects. Another useful approach of assessing matrix effect is postcolumn infusion of an analyte into the MS detector. The extracted blank matrix is injected by an autosampler onto the analytical column [68–70]. The purpose of postcolumn infusion with the analyte is to raise the background level so that the matrix-induced suppression will show as negative peaks. An approach for assessing the lack of inconsistent matrix effect among individual samples is to measure the consistency for the results obtained from these types of experiments with samples from individual matrix lots (typically $n > 6$) [71–76].

Constanzer et al. [77] measured both absolute matrix effects by comparing the responses from postextraction spiked samples with those from neat solutions and relative matrix effects by measuring the consistency of the response factors from spiked matrix lots. Results from LC-MS/MS methods were compared to a validated non-MS method, such as LC-UV [78, 79], LC with fluorescence detector [80], enzyme-linked immunosorbent assay (ELISA) [81, 82], and fluorescence polarization assay [83], or to a different separation method such as GC [84]. Comparable results indicate lack of matrix effects or inconsistent matrix effects since this approach does not positively identify the magnitude of the matrix effects. Huang et al. [85] used LC-UV to detect the down-field matrix peaks that caused ion suppression for the analyte of interest and modified the chromatographic condition [85]. Recovery is determined by comparing the MS response of extracted samples with those spiked (postextraction) into a blank matrix. Because both samples have the matrix constituents present, the matrix effects can be considered the same for extracted samples and postextraction spiked samples. Any differences in responses can be considered as caused by extraction recovery.

### 1.10 Effect of Dosing Vehicles

In preclinical studies, dosing vehicles typically are used at high concentrations to dissolve the test compounds. These dosing vehicles, especially polymeric vehicles such as PEG 400 and Tween 80, can cause significant signal suppression for certain analytes when minimal sample cleanup is used. Table 2 shows a real-world example of the adverse effects of dosing vehicles. The results obtained using a fast gradient LC-MS/MS method with minimal retention are significantly lower than those

**TABLE 2 Comparison of Results Obtained from the Same Set of IV Plasma Samples Containing PEG 400 When Two Different Gradient LC-MS/MS Methods Are Used**

| Time Postdose (h) | Results from the Extended Gradient Method (ng/mL) | Results from the Fast Gradient Method (ng/mL) |
|---|---|---|
| 0.067 | 753 | 261 |
| 0.133 | 712 | 219 |
| 0.2 | 761 | 239 |
| 0.333 | 523 | 176 |
| 0.5 | 394 | 104 |
| 0.75 | 279 | 72.0 |
| 1 | 116 | 28.0 |
| 1.5 | 35.0 | 7.00 |
| 2 | 6.00 | 2.00 |

obtained from an extended gradient LC-MS/MS method. When these two sets of data were fitted into a noncompartment model, values for $C_{max}$, area under the curve (*AUC*), and clearance (*Cl*) were more than 300% different [59].

Effective means of minimizing this type of effect include better chromatographic separations, better sample cleanup, and alternative ionization methods. An easy way to check dosing vehicle effects is to fortify some quality control samples with the vehicle [86]. Significant differences in the measured values between fortified quality control samples and regular quality control samples indicate potential dosing vehicle effects.

### 1.11 Carry-over and Contamination

Carry-over should be differentiated from contamination. Carry-over is usually caused by liquid handlers for the sample preparation or autosamplers for the injection. Carry-over can be estimated and to some extent managed by prearranging the samples so that the consecutive samples are not impacted by the carry-over [87]. Nevertheless, carry-over should be evaluated during method development and validation and if possible reduced. Contamination could be systematic (such as contamination of a reagent with the analyte) or could be random (such as spillover of extraction solvents of two adjacent wells in a 96-well plate). Contamination is undesirable and is usually unmanageable. Care must be taken to minimize carry-over and contamination so that they will not negatively impact the assay results.

### 1.12 Misconception About Stable Isotope-Labeled Internal Standard

Use of stable isotope-labeled analyte as the internal standard has played a significant role in advancing quantitative bioanalytical LC-MS/MS methodology. Due to the high similarity of the physical and chemical characteristics of labeled and unlabeled analytes, almost identical sample extraction recovery and matrix effects could be expected. Potential loss of analyte during the sample preparation and signal suppression/enhancement due to matrix components can be effectively compensated for. However, the stable isotope-labeled internal standard will not compensate for analyte loss due to instability or adsorption to either container or proteins. Occasionally, stable isotope-labeled internal standards have been chro-

matographically separated from the analyte and have experienced different matrix effects [88].

Yang et al. [89] presented an interesting case study for diagnosis and troubleshooting of problems associated with strong analyte–protein binding. The method was validated using a stable isotope-labeled internal standard with a LLE method using hexane as the extraction solvent. However, upon repeat analysis of the same samples, the concentration values increased fivefold from the original value. The concentration increased with each additional freeze/thaw cycle. It was found that this drug candidate has a strong protein binding and hexane is not sufficient to release the drug candidate from the protein. Freeze/thaw cycles gradually denatured the protein and weakened the binding, resulting in increases in the free and extractable drug candidate concentration. A protein precipitation (PP) method was then used to inhibit the protein binding and to release the analyte. Consistent results were then obtained.

## 1.13 Troubleshooting Strategy

Successful method troubleshooting requires full understanding of all aspects of the bioanalytical method. For example, poor dynamic range can potentially be caused by many factors. A systematic investigation of the root cause would help resolve the issues. Table 3 provides a general investigational guide for the troubleshooting of poor linearity. Similar systematic investigations could be made to other important parameters for quantitative bioanalytical methods. One important parameter to

**TABLE 3 Troubleshooting for Poor Linearity**

| Observation | Potential Root Cause |
|---|---|
| Response factors decrease at high end | Poor solubility?<br>Correct reconstitution solvent and volume: analyte soluble in matrix?<br>Saturation due to large curve range—at LC-MS/MS interface? At detector? |
| Response factors decrease at low end | Adsorption of analytes to container?<br>Adsorption of analytes to 96-well plate?<br>Loss of analytes in evaporation/reconstitution step?<br>Adsorption of analyte to injector/tubing/column/source? |
| Abnormal high response at low end of the curve | Contamination?<br>Carry-over? |
| Internal standard (ISTD) response increases with increased analyte concentration | Isotope effects or impurities from analyte?<br>Adsorption of analytes and ISTD to the container? |
| ISTD response decreases with increased analyte concentration | Suppression from analyte to ISTD? |
| ISTD not tracking analyte | LC-MS matrix effect?<br>Extraction recovery problem?<br>Instability of analyte or ISTD in extract?<br>Inadequate equilibrium for protein binding of ISTD? |

review when performing troubleshooting, especially when troubleshooting a mismatch of calibration standards and quality controls, is the response factor. The response factor is the response ratio of the analyte versus internal standard normalized against the concentrations. Response factors should be consistent throughout the entire calibration range. Table 4 shows a troubleshooting example. At first glance, the accuracy results indicate that the two low levels of quality control samples (0.25 ng/mL and 0.75 ng/mL) are biased high based on the calculated accuracy results. However, a further review of the response factors shows that these two

**TABLE 4 Troubleshooting for Poor Accuracy of Quality Control (QC) Samples**[a]

| Sample Name | Calculated Concentration (ng/mL) | Response Factor | Accuracy (%) | Sample Name | Calculated Concentration (ng/mL) | Response Factor | Accuracy (%) |
|---|---|---|---|---|---|---|---|
| CAL-200 | 205 | 0.0163 | 102.6 | QC-0.25 | 0.298 | 0.01607 | 119.2 |
| CAL-100 | 109 | 0.0173 | 109.3 | QC-0.25 | 0.298 | 0.01611 | 119.4 |
| CAL-50 | 47.7 | 0.0151 | 95.32 | QC-0.25 | 0.296 | 0.01597 | 118.5 |
| CAL-20 | 20 | 0.0158 | 100.1 | QC-0.25 | 0.299 | 0.01617 | 119.8 |
| CAL-10 | 10.3 | 0.0162 | 102.6 | QC-0.25 | 0.297 | 0.01601 | 118.8 |
| CAL-5 | 4.92 | 0.0155 | 98.35 | QC-0.25 | 0.297 | 0.01601 | 118.8 |
| CAL-2 | 2.03 | 0.0157 | 101.3 | QC-0.75 | 0.883 | 0.01774 | 117.8 |
| CAL-0.5 | 0.54 | 0.0157 | 107.9 | QC-0.75 | 0.902 | 0.01813 | 120.3 |
| CAL-0.25 | 0.241 | 0.0125 | 96.49 | QC-0.75 | 0.86 | 0.01724 | 114.6 |
| CAL-0.25 | 0.246 | 0.0128 | 98.54 | QC-0.75 | 0.865 | 0.01734 | 115.3 |
| CAL-0.5 | 0.518 | 0.0150 | 103.6 | QC-0.75 | 0.878 | 0.01762 | 117 |
| CAL-2 | 1.9 | 0.0147 | 95.15 | QC-0.75 | 0.874 | 0.01754 | 116.5 |
| CAL-5 | 4.72 | 0.0148 | 94.33 | QC-160 | 166 | 0.01641 | 103.5 |
| CAL-10 | 9.8 | 0.0155 | 98.01 | QC-160 | 174 | 0.01724 | 108.7 |
| CAL-20 | 19.8 | 0.0157 | 99.13 | QC-160 | 166 | 0.01650 | 104 |
| CAL-50 | 48.8 | 0.0155 | 97.64 | QC-160 | 168 | 0.01666 | 105.1 |
| CAL-100 | 97.9 | 0.0155 | 97.85 | QC-160 | 167 | 0.01657 | 104.5 |
| CAL-200 | 203 | 0.0161 | 101.7 | QC-160 | 170 | 0.01680 | 106 |
| ***Mean*** | | 0.01532 | 100 | QC-75 | 80.6 | 0.01704 | 107.5 |
| ***SD*** | | 0.001147 | 4.13 | QC-75 | 78.7 | 0.01664 | 104.9 |
| ***CV%*** | | 7.49 | 4.13 | QC-75 | 81.1 | 0.01715 | 108.2 |
| | | | | QC-75 | 78.8 | 0.01665 | 105 |
| | | | | QC-75 | 79.2 | 0.01674 | 105.6 |
| | | | | QC-8 | 8.91 | 0.01757 | 111.3 |
| | | | | QC-8 | 8.67 | 0.01710 | 108.3 |
| | | | | QC-8 | 8.7 | 0.01717 | 108.8 |
| | | | | QC-8 | 8.44 | 0.01665 | 105.5 |
| | | | | QC-8 | 8.73 | 0.01723 | 109.2 |
| | | | | QC-8 | 8.8 | 0.01736 | 110 |
| | | | | ***Mean*** | | 0.016878 | 111.3 |
| | | | | ***SD*** | | 0.000576 | 5.93 |
| | | | | ***CV%*** | | 3.42 | 5.33 |

[a]Pay special attention to the lower response factors but good accuracy for CAL-0.25. QC samples have consistent response factors over the entire concentration range but poor accuracy caused by the lower response factor of CAL-0.25. Because this is $1/x^2$ weighted linear regression, the lowest standard (CAL-0.25) has the most significant weighting for the calculation of QC concentrations.

levels of quality control samples have the same response factors as the other levels of quality controls. It also reveals that the response factors dropped for the calibration standards at lower concentrations even though the calculated accuracy for these standards is excellent. Further investigation revealed that the compound in the spiking solution used to prepare calibration standards was impacted by adsorption onto the container. This simple example clearly demonstrates the importance of reviewing the response factors rather than only relying on the calculated accuracy for the troubleshooting.

### 1.14 Conclusion

Bioanalytical support for preclinical studies is complicated. Hence, bioanalytical scientists need to fully understand both the underlying sciences and regulatory requirements. They should always be aware of the potential pitfalls that could occur at any stage of the study. There are a number of parameters requiring evaluation during method development and validation. The method validation strategy should be tailored for the intended use.

## 2 PRECLINICAL BIOANALYTICAL SUPPORT USING LIGAND-BINDING ASSAY

### 2.1 Introduction

Much of the information discussed in Section 1 for the bioanalysis of small molecules may also apply to the bioanalysis of large molecule therapeutics as well. However, there are some unique challenges associated with quantification of large molecule therapeutics. These are due mainly to the nature of the molecules and limitations of the technology (e.g., ligand-binding assay (LBA)), used for bioanalysis.

Recombinant proteins and humanized monoclonal antibodies comprise an ever increasing larger proportion of protein therapeutics and, as a result, have posed new challenges and special needs. These molecules may not survive the typical sample handling conditions of the chromatographic assays. Moreover, these drugs are highly potent. Consequently, their blood concentration levels tend to be very low. Hence, the method of quantification for these entities must be very gentle, specific, and highly sensitive. Ligand-binding assays such as radioimmunoassay (RIA) and enzyme-linked immunosorbent assay (ELISA) are the techniques of choice for quantification of large molecule drugs in biological samples [90, 91].

In addition, due to their high molecular weight, there is always the possibility that large molecule therapeutics may elicit an immune response directed against the drug [92]. A small change in protein structure, even though it may comprise a very small proportion of the total protein, can induce a large immunogenic response. Therefore, regulatory agencies require immunogenicity evaluation for large molecule therapeutics even in the preclinical stage [93]. Consequently, there is an increasing need to develop novel assays that can detect and characterize the anti-drug antibodies (ADA) or immunogenicity of large molecule therapeutics. To accomplish this, the laboratory has to rely on several bioanalytical techniques and technologies to measure these changes including LBA.

Like any other bioanalytical method, the intended use of the LBA helps to determine the course of development and validation of the assay. Ligand-binding assays are used for (1) the quantitation of therapeutic entities to support the toxicokinetic / pharmacokinetic studies [91, 94] and (2) monitoring the level of surrogate biomarkers to support pharmacodynamic evaluation [95]. On the other hand, qualitative or quasiquantitative assays are generally used for the determination of anti-drug antibodies [92, 96]. The LBA developed for the bioanalytical support should not only be robust, reliable, and reproducible but should also be GLP compliant from a regulatory perspective. It can then be used for production, toxicology, and Phase I though IV testing. In view of the small sample volumes available in typical preclinical studies, it is even more compelling to design and develop a robust assay to reduce sample reanalysis.

In this section we provide a brief account of the ligand-binding assay technology, specific challenges associated with the GLP-compliant method development and validation of quantitative assays, and current industry and regulatory perspectives. Method development and validation for detection of anti-drug antibodies [92, 96] or quantification of biomarkers [95] may require slightly different strategies.

### 2.2 Ligand-Binding Assay Technology

A typical ligand-binding assay utilizes an analyte-specific binder (e.g., antibody, binding protein, drug, or receptor) to capture the analyte of interest. The captured analyte is detected by the "detector molecule," which is generally an antibody labeled with a radioisotope (e.g., $^{125}$I), an enzyme (e.g., horseradish peroxidase, alkaline phosphatase), or another label (e.g., biotin, avidin). ELISA, the most commonly used ligand-binding assay, generally uses a detector molecule that is labeled with an enzyme. The extent of bound enzyme activity is measured by the changes in color intensity of the substrate solution. The color intensity is directly proportional to the concentration of analyte captured on the microtiter plate (see Fig. 3). A detailed description of the ligand-binding assay technology is beyond the scope of this chapter. Several books that provide in-depth information on ligand-binding assays, including ELISA, are included in the references [97, 98].

### 2.3 Using Ligand-Binding Assay in a GLP Environment

To support preclinical studies, an analyte-specific ligand-binding assay must be validated to conform to GLP regulations. In a GLP environment, like any other bioanalytical method, an LBA has a three-phase life cycle: (1) method development, (2) prestudy validation, and (3) in-study validation. Table 5 summarizes the assay assessment parameters that are addressed at various stages of the method life cycle. During method development, emphasis is focused on designing a robust and analyte-specific method for the intended use. This requires advanced planning and a careful consideration of (1) assay design and format, (2) critical reagent selection, (3) critical reagent stability evaluation, (4) matrix selection, (5) sample collection conditions, and (6) optimization of assay condition to achieve desired performance characteristics. Once a robust method has been developed, the prestudy validation becomes uncomplicated. The assay characteristics, including precision and accuracy, limits of quantification, and analyte stability, are formally confirmed during the

**TABLE 5 Summary of Method Validation Assessment Parameters Over the Method Life Cycle**

| Performance Parameters | Development | Prestudy Validation | *In-Study Validation* |
|---|---|---|---|
| Critical reagents | Identify and procure | Apply | Apply |
| Assay format / batch size | Establish | Apply | Apply |
| Matrix of calibrators and controls | Establish | Confirm | Apply |
| Minimal required dilution | Establish | Confirm | Apply |
| Analyte stability | Initiate | Establish | Ongoing assessment |
| Specificity | Establish | Apply | Apply |
| Selectivity | Evaluate | Confirm | Apply |
| Calibration curve fitting algorithm | Establish | Confirm | Apply |
| LLOQ and ULOQ | Evaluate | Establish | Apply |
| Precision and accuracy | Evaluate (CV and RE) | Establish (CV and RE) | Apply (TE) |
| Run acceptance QC (low, medium, high) | Evaluate | Establish | Apply |
| Dilutional linearity | Establish | Confirm | Apply—confirm if extended |
| Batch size | Evaluate | Establish | Apply—confirm if extended |
| Robustness/ ruggedness | Evaluate | Establish | Monitor |
| Parallelism | Evaluate where possible | Evaluate where possible | Establish with incurred samples |
| Run acceptance criteria | N/A | Runs accepted based on calibration curve acceptance criteria | Runs accepted based on acceptable calibration curve and QC samples following 4-6-X rule |

prestudy validation phase. These characteristics are monitored during the in-study validation phase in which study samples are analyzed [91, 94].

### 2.4 Ligand-Binding Assay Specific Challenges

The distinguishing characteristics of ligand-binding assays that separate them from chromatographic (e.g., LC-MS/MS) assays are summarized as follows.

1. The design of a sandwich ligand-binding assay takes advantage of the ability of an antibody to specifically bind to the analyte (drug). The amount of analyte

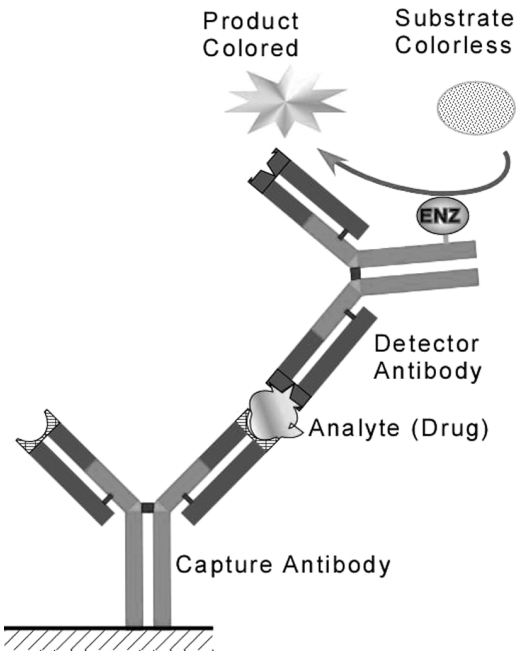


**FIGURE 3** A sandwich enzyme linked immunosorbent assay (ELISA) with colorimetric endpoint.

is measured indirectly by the binding of a detector antibody conjugated to an enzyme or any other labeling molecule (Fig. 3). In a chromatographic assay the analyte is physically separated into distinct peaks. The area under the peak is precisely measured and is directly proportional to the analyte concentration. Due to the indirect measurements in LBA, the results are somewhat less precise, as reflected by coefficient of variation ($CV \leq 20\%$), than in chromatographic assays ($CV \leq 10\%$).

2. The biological samples are generally analyzed without pretreatment in a LBA, as opposed to the process of extracting analyte from samples prior to the chromatographic assay. Consequently, numerous matrix components including binding proteins, drugs, degrading enzymes, heterophilic antibodies, and anti-animal antibodies may have negative or positive interference in the assay performance.
3. Due to the limited analyte-binding capacity of the binder molecule (e.g., capture antibody), the typical calibration curves in these assays are nonlinear, as opposed to the linear curves in chromatographic assays. Consequently, the working range of quantification in an LBA is narrower than in the linear curves of chromatographic assays (Fig. 4). In preclinical studies, relatively high doses of drugs are administered. The resulting high concentration of drug in the study samples may require sample dilutions that could be several thousandfold.

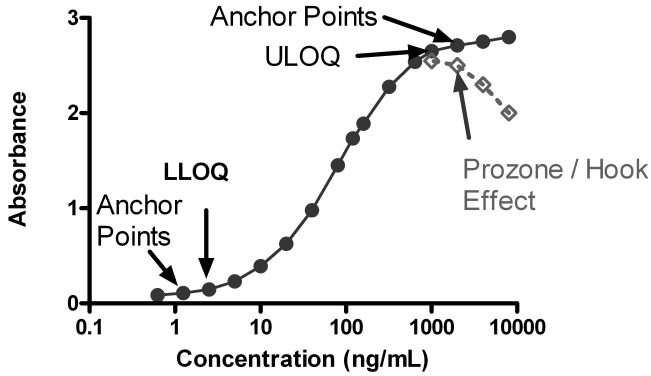


**FIGURE 4** Prozone or Hook effect Illustration. A typical sigmoid calibration curve for a sandwich ELISA is shown with "anchor points" at both ends of the curve bracketing the calibrators at the LLOQ and ULOQ levels. A high dose Prozone or Hook effect is illustrated by suppression of signal (absorbance) with analyte at concentrations far above the ULOQ level.

## 2.5 Responding to the Challenge of Ligand-Binding Assay

For all assays, the key factor is the accuracy of the reported results [15]. In view of the above characteristics, besides method reproducibility considerations, one must pay special attention to establishing the method **(1)** specificity, **(2)** selectivity, **(3)** minimal required dilution, **(4)** dilutional linearity, and **(5)** analyte (drug) stability in the biological matrix [94]. Evaluation and validation of the method repeatability, precision, and accuracy for the LBA is described in detail elsewhere [91, 94].

***Method Specificity*** Demonstration of method specificity for the LBA could be a challenging task. The LBA is generally developed during the discovery phase when specific antibodies are raised against the drug molecule and are characterized. These antibodies are used to develop the initial ELISA method. Later, the assay is used to support preclinical and clinical studies. The specificity of the method is based mainly on the ability of these antibodies to specifically interact with the drug molecule as opposed to interacting with nondrug (i.e., cross-reactant) molecules.

Antibody specificity is generally evaluated by determining the cross-reactivity of the antibody with nondrug molecules, including coadministered drugs, drug metabolites, or molecules that are structurally related to the drug. To evaluate the antibody specificity, dose–response curves are generated using drugs of interest and potential cross-reactive compounds at significantly higher concentrations than the drug itself [97, 98]. The extent of acceptable cross-reactivity may depend on the relative concentration of the potential cross-reactant present in the study sample. An antibody with low cross-reactivity (e.g., <1%) with a potential cross-reactant may not be suitable if the concentration of the cross-reactant in the sample is at a very high concentration relative to the analyte (e.g., 1000-fold). In such cases, a sample cleanup method may be required to remove the cross-reactants prior to sample analysis. Similarly, an antibody with 20% cross-reactivity with a potential contaminant may be acceptable for use if the analyte itself is expected to be in great excess.

***Method Selectivity*** In a typical LBA, the samples are analyzed without pretreatment. Numerous matrix components that are present in the sample may affect the assay performance. The ability of the method to selectively quantify the analyte of interest in the presence of a variety of matrix components defines the method selectivity [15, 91, 94]. Method selectivity evaluation and validation is one of the most crucial and demanding tasks in establishing the LBA. During method development, at least 10 individual matrix samples from 10 normal animals or subjects (from normal or patient population) are assayed unspiked and spiked with the drug at a concentration near the lower end of the calibration curve. Acceptable recovery of drug (usually within 20% of the nominal concentration) in at least 80% of the samples is considered an indication that the method is selective. At this stage of assay development (where blank matrix is not available), the calibration curve may be prepared in the assay diluent. The results of the initial selectivity screening of matrix samples may be used to identify matrix samples that can be pooled to form a blank matrix pool. This pool can then be used for preparing the calibrators and the validation or QC samples [91, 94]. Method selectivity is confirmed during the prestudy validation, where both calibrators and control samples are spiked in the pooled biological matrix. In these experiments the selectivity is evaluated by assaying at least 10 individual matrix samples unspiked and spiked with the analyte near the lower limit of quantification (LLOQ).

***Minimal Required Dilution of Sample*** In assays where a strong matrix effect is observed, it may be necessary to determine the "minimal required dilution" (MRD) of the sample, that is, the dilution at which matrix interference is minimized to an acceptable level. Since the reportable concentration of analyte is the product of the measured concentration multiplied by the dilution factor, the practical limit of quantification will be greatly affected by the MRD of sample [94]. Hence, the target should be to develop a relatively more sensitive (with lower LLOQ) method where the MRD is higher. A method with smallest MRD that has acceptable interference and sensitivity for the intended use would be ideal.

***Dilutional Linearity*** Due to the narrow range of quantification in ligand-binding assays and the possible high concentration of analyte of interest in the sample (especially in preclinical studies), a number of samples may require dilution prior to analysis. Therefore, to avoid running highly concentrated "dilution controls" with each assay during sample analysis, dilutional linearity of sample at a wide range of dilutions is demonstrated during prestudy method validation. It should be noted that the final content of biological matrix should be identical to that in the calibration curve and/or QC samples. Therefore, dilutional linearity should be established subsequent to "method selectivity" evaluation, MRD establishment, and preparation of a blank matrix pool.

In some cases, it is observed that high concentration of analyte may suppress the signal, resulting in underestimation of analyte [99]. This phenomenon is known as the "Prozone" or "Hook" effect. A typical illustration of the effect is depicted in Fig. 4. Dilutional linearity evaluation may help in identifying this effect, especially for assays used to support preclinical studies. If this effect goes undetected, it may result in a gross underestimation of drug concentration in the samples where high concentrations of analyte are expected. A typical example is illustrated in Table 6

**TABLE 6 Demonstration of Hook Effect and Dilutional Linearity**[a]

| Dilution (Fold) | Measured Concentration (pg/mL)[b] | Computed Concentration (pg/mL) | |
|---|---|---|---|
| Neat | 231 | 231 | Hook or Prozone effect |
| 2 | 365 | 729 | |
| 4 | 674 | 2,698 | |
| 80 | >ULOQ | Not computed | |
| 160 | >ULOQ | Not computed | |
| 200 | >ULOQ | Not computed | |
| 400 | 755 | 302,078 | Dilutional linearity[c] |
| 800 | 369 | 294,798 | |
| 1,600 | 178 | 285,226 | |
| 3,200 | 93 | 297,229 | |
| 6,400 | 50 | 322,447 | |
| 12,800 | 26 | 335,578 | |
| 25,600 | <LLOQ | Not computed | |

[a]Data from this table is visually depicted in Fig. 5.
[b]Method ULOQ = 25 pg/mL; ULOQ = 1000 pg/mL.
[c]Average concentration (pg/mL) = 306,226 with 6.2% CV.

and Fig. 5. A sample from an escalating dose study, where high concentration of drug is expected, may be assayed "neat" (without dilution), giving a low measured concentration. This anomalous observation may call for an investigation. In troubleshooting experiments, the sample may be re-assayed at a wide range of dilutions. In the illustration, it is observed that with increasing dilutions the measured concentration increased, reaching the ULOQ of the assay. Moreover, subsequent dilutions from 400-fold to 128,000-fold showed dilutional linearity. Computed concentrations, measured in the dilutional linearity range, provided a realistic estimate of the concentration of drug in samples. This scenario clearly illustrates the importance of Prozone or Hook effect evaluation and establishment of the dilutional linearity in an LBA.

Parallelism is another assay parameter that is conceptually similar to the dilutional linearity. It is evaluated when incurred samples become available. Plotting the responses generated by the diluted incurred samples versus the calibration curve gives a visual impression of parallelism. Various approaches to parallelism data have been described in the literature [95, 100]. A test of parallelism demonstrates that the analyte present in the sample is structurally similar to that used for generation of the calibration curve.

***Analyte Stability in Biological Matrix*** Biological matrices may vary from each other with respect to their chemical constituents including proteins and enzymes. Proteolytic enzyme composition may have a distinct effect on the stability of a drug in a biological matrix. Rat or mouse serum or plasma samples should be treated with special care. It has been our experience that an analyte that may be stable in rabbit or monkey serum may very well be highly unstable in rat or mouse serum. Therefore, in such cases, it may be necessary to collect the study samples in the

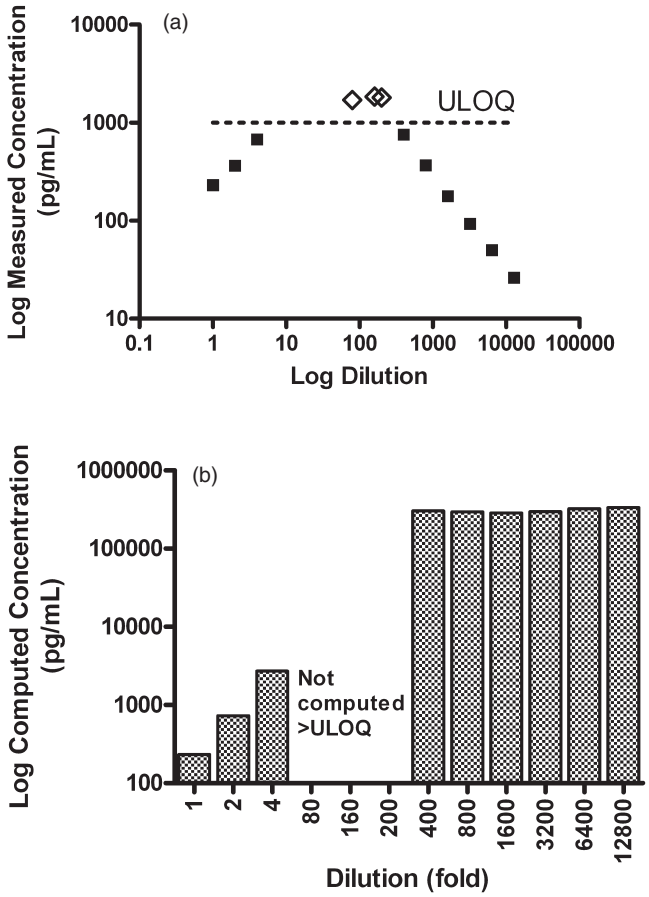


**FIGURE 5** Illustration of the significance of dilutional linearity evaluation, especially in an assay where Prozone or Hook effect is present. (a) Initially with increasing dilutions the measured concentration in a sample increased, reaching or exceeding the ULOQ of the assay, demonstrating the suppression of response with high concentration of analyte. Subsequently, six dilutions showed a linear decrease in measured concentration with increasing dilutions of the sample. (b) The computed concentrations corrected for the dilution factor from 400-fold to 128,000-fold demonstrated dilutional linearity by showing that computed concentration is independent of the dilution factor in this range.

presence of a cocktail of proteolytic enzyme inhibitors and/or keep the samples on ice during analysis.

### 2.6 Using Commercial Kit Based Assays

Often commercial kits are used for quantification of drugs that are homologs of endogenous compounds or for the measurement of biomarkers in biological samples. Kits are generally designed for use in clinical or research lab environments. In order to adopt commercial kits for GLP-compliant bioanalytical work, one must redesign the composition and format of the control samples that are used to monitor the assay performance and to make assay acceptance decisions. It is of the utmost

importance that a quality control strategy is designed in advance to evaluate and monitor the lot-to-lot variability of kits. Several multiplexed assay kits are also available commercially. Adaptation and validation of these kits for quantification of analytes of interest in biological matrices could be challenging and therefore should be thoroughly explored prior to sample analysis.

## 2.7 Industry and Regulatory Perspective

Historically, the focus of regulatory agencies has been on the chromatographic techniques for the quantification of small molecular weight drug entities that constituted the major proportion of drug development portfolios in the pharmaceutical industry. The first conference on bioanalytical method validation (commonly known as the Crystal City I conference), that was held in Crystal City, Virginia in 1990, addressed in depth the validation of chromatographic assays but paid little attention to nonchromatographic methods including immunoassays [101]. The search for LBA-specific guidance was started soon after the 1990 Crystal City conference. In 1998, one of the authors of this chapter (MK) organized a roundtable at the AAPS annual meeting in San Francisco. This was, perhaps, the first time that LBA-specific challenges were discussed on a public platform. This led to the publication of a white paper giving a comprehensive account of the pharmaceutical industry perspective on the validation of immunoassays for bioanalysis [91]. Soon after this publication in 2000, a historic workshop on bioanalytical method validation for macromolecules was sponsored by the AAPS and the FDA (1–3 March 2000). This workshop, for the first time, addressed the validation of a variety of methodologies for assays of large molecule therapeutics [102]. However, the scope of the workshop was too broad to reach a consensus. In May 2001, the FDA *Guidance for Industry: Bioanalytical Methods Validation* was issued [15]. It was thought that a focused approach was needed to address the special challenges associated with LBAs for large molecule therapeutics. Consequently, a subcommittee of the then newly formed Ligand Binding Assay Bioanalytical Focus Group (LBABFG) at the American Association of Pharmaceutical Scientists (AAPS) was set out to critically review the workshop summary report, evaluate exceptions to criteria in the guidance, and produce a consensus document. The LBABFG white paper was published in 2003 [94]. Around the same time, the AAPS sponsored a workshop (12–13 May 2003) mainly to address the "how to" of the validation of methods used for PK analysis of macromolecules [103]. Over the last four years, LBABFG white paper has provided valuable direction to LBA scientists working in regulatory compliant environments. In May 2006, the 3rd AAPS/FDA Bioanalytical Method Validation Workshop was held. With some changes, this workshop essentially endorsed the LBA method performance acceptance criteria proposed in the LBABFG white paper (Table 7). In the workshop, the recommended run acceptance rule of "4-6-30" (i.e., a minimum of 4 out of 6 QC samples are within 30% of their nominal values) has been changed to a "4-6-20" rule as stipulated in the conference report [19].

## 2.8 Conclusions

Ligand-binding assays are the technique of choice for bioanalytical support for preclinical and clinical studies for large molecule therapeutics. In view of the small sample volumes and high drug concentrations that are typically available in preclinical studies, one should design and develop a robust assay where reliable and repro-

**TABLE 7 Comparison of LBA Specific Acceptance Criteria as Recommended by the LBABFG White Paper and AAPS/FDA 3rd Conference Report**

| Performance Characteristic | LBABFG White Paper | AAPS/FDA 3rd Conference Report | Recommended Changes in Conference Report |
|---|---|---|---|
| Validation Sample %RE | ±20% ±25% at the LLOQ | ±20% ±25% at the LLOQ and ULOQ | ±25% RE extended to ULOQ level as well |
| Validation Sample %CV | ±20% ±25% at the LLOQ | ±20% ±25% at the LLOQ and ULOQ | ±25% CV extended to ULOQ level as well |
| Validation Sample %TE | ±30% ±40 at the LLOQ | ±30% ±40% at the LLOQ and ULOQ | ±40% TE extended to ULOQ level as well |
| In-study standards (excluding anchor points) | ≥75% of standard points are ±20% RE (±25% RE at the LLOQ); at least six valid standard points must remain within limits | ≥75% of standard points are ±20% RE (±25% RE at the LLOQ and ULOQ) | ±25% RE extended to standard points at ULOQ level as well; number of acceptable points in a standard curve not specified |
| In-study quality controls | 4-6-30[a] rule; at least 50% of QCs are valid at each level | 4-6-20[a] rule; at least 50% of QCs are valid at each level; three levels analyzed in duplicate (six results) | Major changes: (1) QC acceptance limit is reduced from 30% RE to 20% RE; (2) three levels of QC samples in duplicate specified |

[a]Exceptions allowed with justification.

ducible results are obtained and sample reanalysis is minimized. Moreover, the method specificity, selectivity, dilutional linearity, and reproducibility must be established and validated upfront to avoid potential problems that may be encountered during sample analysis. Method validation performance acceptance criteria for precision of measurements for LBA assays are much wider compared to that of chromatographic assays. Current white papers and workshop reports provide adequate recommendations for development, pre-study and in-study validation, and implementation of these assays to support regulatory compliant studies.

This chapter provides an overview of the current advance and trends in the bioanalytical area to support preclinical studies. It covers both conventional small molecule drugs and macromolecular BIOTEC therapeutic entities.

## REFERENCES

1. Lee MS, Kerns ED. LC/MS applications in drug development. *Mass Spectrom. Rev* 1999;18:187–279.
2. Dong H, Ouyang Z, Liu J, Jemal M. The use of a dual dye photometric calibration method to identify possible sample dilution from an automated multichannel liquid-handling system. *Clin Lab Med* 2007;27:113–122.

3. Gu H, Unger S, Deng Y. Automated Tecan programming for bioanalytical sample preparation with EZTecan. *Assay Drug Dev Technol* 2006;4:721–733.
4. Brewer E, Henion J. Atmospheric pressure ionization LC/MS/MS techniques for drug disposition studies. *J Pharm Sci* 1998;87:395–402.
5. Jemal M, Xia Y-Q. LC-MS development strategies for quantitative bioanalysis. *Curr Drug Metab* 2006;7:491–502.
6. Jemal M. High-throughput quantitative bioanalysis by LC/MS/MS. *Biomed Chromatogr* 2000;14:422–429.
7. Oliveira EJ, Watson DG. Liquid chromatography–mass spectrometry in the study of the metabolism of drugs and other xenobiotics. *Biomed Chromatogr* 2000;14:351–372.
8. Law B, Temesi D. Factors to consider in the development of generic bioanalytical high-performance liquid chromatographic–mass spectrometric methods to support drug discovery. *J Chromatogr B* 2000;748:21–30.
9. Plumb R, Dear GJ, Mallett DN, Higton DM, Pleasance S, Biddlecombe RA. Quantitative analysis of pharmaceuticals in biological fluids using high-performance liquid chromatography coupled to mass spectrometry: a review. *Xenobiotica* 2001;31:599–617.
10. Kyranos JN, Cai H, Wei D, Goetzinger WK. High-throughput high-performance liquid chromatography/mass spectrometry for modern drug discovery. *Curr Opin Biotechnol* 2001;12:105–111.
11. Triolo A, Altamura M, Cardinali F, Sisto A, Maggi CA. Mass spectrometry and combinatorial chemistry: a short outline. *J Mass Spectrom* 2001;36:1249–1259.
12. O'Connor D. Automated sample preparation and LC-MS for high-throughput ADME quantification. *Curr Opin Drug Discov Dev* 2002;5:52–58.
13. Ackermann BL, Berna MJ, Murphy AT. Recent advances in use of LC/MS/MS for quantitative high-throughput bioanalytical support of drug discovery. *Curr Top Med Chem* 2002;2:53–66.
14. Hopfgartner G, Bourgogne E. Quantitative high-throughput analysis of drugs in biological matrices by mass spectrometry. *Mass Spectrom Rev* 2003;22:195–214.
15. FDA. *Guidance for Industry: Bioanalytical Methods Validation*. US Department of Health and Human Services, Center for Drug Evaluation and Research, and Center for Veterinary Medicine, May 2001. Available at http://www.fda/cder/guidance/index.htm.
16. Jemal M, Schuster A, Whigan DB. Liquid chromatography/tandem mass spectrometry methods for quantitation of mevalonic acid in human plasma and urine: method validation, demonstration of using a surrogate analyte, and demonstration of unacceptable matrix effect in spite of use of a stable isotope analog internal standard. *Rapid Commun Mass Spectrom* 2003;17:1723–1734.
17. Jemal M, Ouyang Z, Powell ML. A strategy for a post-method-validation use of incurred biological samples for establishing the acceptability of a liquid chromatography/tandem mass-spectrometric method for quantitation of drugs in biological samples. *Rapid Commun Mass Spectrom* 2002;16:1538–1547.
18. Humphreys WG, Unger SE. Safety assessment of drug metabolites: characterization of chemically stable metabolites. *Chem Res Toxicol* 2006;19:1564–1569.
19. Viswanathan CT, Bansal S, Booth B, DeStefano AJ, Rose MJ, Sailstad J, Shah VP, Skelly JP, Swann PG, Weiner R. Workshop/conference report—Quantitative Bioanalytical methods validation and implementation: best practices for chromatographic and ligand binding assays, *AAPS* 2007;9(1): Article 4. Available at http://www.aapsj.org.
20. Baillie TA, Cayen MN, Fouda H, Gerson RJ, Green JD, Grossman SJ, Klunk LJ, LeBlanc B, Perkins DG, Shipley LA. Drug metabolites in safety testing. *Toxicol Appl Pharmacol* 2002;182:188–196.

21. US Food and Drug Administration. *Draft Guidance for Industry: Safety Testing of Drug Metabolites*. Available at www.fda.gov/cder/guidance.
22. Brown K, Tompkins EM, White INH. Applications of accelerator mass spectrometry for pharmacological and toxicological research. *Mass Spectrom Rev* 2006;25:127–145.
23. Bailey MJ, Dickinson RG. Acyl glucuronide reactivity in perspective: biological consequences. *Chem-Biol Interact* 2003;145:117–137.
24. Miles DR, Mesfin M, Mody TD, Stiles M, Lee J, Fiene J, Denis B, Boswell GW. Validation and use of three complementary analytical methods (LC-FLS, LC-MS/MS and ICP-MS) to evaluate the pharmacokinetics, biodistribution and stability of motexafin gadolinium in plasma and tissues. *Anal Bioanal Chem* 2006;385:345–356.
25. Chen Y-L, Felder L, Jiang X, Naidong W. Determination of ketoconazole in human plasma by high-performance liquid chromatography–tandem mass spectrometry. *J Chromatogr B* 2002;774:67–78.
26. Leinweber FC, Lubda D, Cabrera K, Tallarek U. Characterization of silica-based monoliths with bimodal pore size distribution. *Anal Chem* 2002;74:2470–2477.
27. Wu JT, Zeng H, Deng Y, Unger SE. High-speed liquid chromatography/tandem mass spectrometry using a monolithic column for high-throughput bioanalysis. *Rapid Commun Mass Spectrom* 2001;15:1113–1119.
28. Hsieh Y, Wang G, Wang Y, Chackalamannil S, Brisson, J-M, Ng K, Korfmacher WA. Simultaneous determination of a drug candidate and its metabolite in rat plasma samples using ultrafast monolithic column high-performance liquid chromatography/tandem mass spectrometry. *Rapid Commun Mass Spectrom* 2002;16:944–950.
29. Deng Y, Wu JT, Lloyd TL, Chi CL, Olah TV, Unger SE. High-speed gradient parallel liquid chromatography/tandem mass spectrometry with fully automated sample preparation for bioanalysis: 30 seconds per sample from plasma. *Rapid Commun Mass Spectrom* 2002:16:1116–1123.
30. Naidong W, Shou WZ, Chen, Y-L, Jiang X. Novel liquid chromatographic–tandem mass spectrometric methods using silica columns and aqueous-organic mobile phases for quantitative analysis of polar ionic analytes in biological fluids. *J Chromatogr B* 2001;754:387–399.
31. Naidong W. Bioanalytical liquid chromatography tandem mass spectrometry methods on underivatized silica columns with aqueous/organic mobile phases. *J Chromatogr B* 2003;796:209–224.
32. Herman JL. Generic method for on-line extraction of drug substances in the presence of biological matrices using turbulent flow chromatography. *Rapid Commun Mass Spectrom* 2002;16:421–426.
33. Tang YQ, Tollefson JA, Beato BD, Weng N. 2003 AAPS Annual Meeting, Salt Lake City, Utah, 26–30 October 2003.
34. Xue YJ, Pursley J, Arnold M. Liquid–liquid extraction of strongly protein bound BMS-299897 from human plasma and cerebrospinal fluid, followed by high-performance liquid chromatography/tandem mass spectrometry. *J Pharm Biomed Anal* 2007;43:1728–1736.
35. Zhang N, Rogers K, Gajda K, Kagel JR, Rossi DT. Integrated sample collection and handling for drug discovery bioanalysis. *J Pharm Biomed Anal* 2000;23:551–560.
36. Teitz DS, Khan S, Powell ML, Jemal M. An automated method of sample preparation of biofluids using pierceable caps to eliminate the uncapping of the sample tubes during sample transfer. *J Biochem Biophys Methods* 2000;45:193–204.
37. Allanson JP, Biddlecombe RA, Jones AE, Pleasance S. The use of automated solid phase extraction in the "96 well" format for high throughput bioanalysis using liquid

chromatography coupled to tandem mass spectrometry. *Rapid Commun Mass Spectrom* 1996;10:811–816.

38. Janiszewski J, Schneider R, Hoffmaster K, Swyden M, Wells D, Fouda H. Automated sample preparation using membrane microtiter extraction for bioanalytical mass spectrometry. *Rapid Commun Mass Spectrom* 1997;11:1033–1037.
39. Simpson H, Berthemy A, Burhrman D, Burton R, Newton J, Kealy M, Wells D, Wu D. High throughput liquid chromatography/mass spectrometry bioanalysis using 96-well disk solid phase extraction plate for the sample preparation. *Rapid Commun Mass Spectrom* 1998;12:75–82.
40. Eerkes A, Addison T, Naidong W. Simultaneous assay of sildenafil and desmethylsildenafil in human plasma using liquid chromatography–tandem mass spectrometry on silica column with aqueous-organic mobile phase. *J Chromatogr B* 2002;768:277–284.
41. Steinborner S, Henion J. Liquid–liquid extraction In the 96-well plate format with SRM LC/MS quantitative determination of methotrexate and its major metabolite in human plasma. *Anal Chem* 1999;71:2340–2345.
42. Jemal M, Teitz D, Ouyang Z, Khan S. Comparison of plasma sample purification by manual liquid–liquid extraction, automated 96-well liquid–liquid extraction and automated 96-well solid-phase extraction for analysis by high-performance liquid chromatography with tandem mass spectrometry. *J Chromatogr B* 1999;732:501–508.
43. Bolden RD, Hoke SH II, Eichold TH, McCauley-Myers DL, Wehmeyer KR. Semi-automated liquid–liquid back-extraction in a 96-well format to decrease sample preparation time for the determination of dextromethorphan and dextrorphan in human plasma. *J Chromatogr B* 2002;772:1–10.
44. Ramos L, Bakhtiar R, Tse FLS. Liquid–liquid extraction using 96-well plate format in conjunction with liquid chromatography/tandem mass spectrometry for quantitative determination of methylphenidate (Ritalin) in human plasma. *Rapid Commun Mass Spectrom* 2000;14:740–745.
45. Shen Z, Wang S, Bakhtiar R. Enantiomeric separation and quantification of fluoxetine (Prozac) in human plasma by liquid chromatography/tandem mass spectrometry using liquid–liquid extraction in 96-well plate format. *Rapid Commun Mass Spectrom* 2002;16:332–338.
46. Eerkes A, Shou WZ, Naidong W. Liquid/liquid extraction using 96-well plate format in conjunction with hydrophilic interaction liquid chromatography–tandem mass spectrometry method for the analysis of fluconazole in human plasma. *J Pharm Biomed Anal* 2003;31:917–928.
47. Watt AP, Morrison D, Locker KL, Evans DG. Higher throughput bioanalysis by automation of a protein precipitation assay using a 96-well format with detection by LC-MS/MS. *Anal Chem* 2000;72:979–984.
48. Shou WZ, Bu H-Z, Addison T, Jiang X, Naidong W. Development and validation of a liquid chromatography/tandem mass spectrometry (LC/MS/MS) method for the determination of ribavirin in human plasma and serum. *J Pharm Biomed Anal* 2002; 29:83–94.
49. Naidong W, Shou WZ, Addison T, Maleki S, Jiang X. Liquid chromatography/tandem mass spectrometric bioanalysis using normal-phase columns with aqueous/organic mobile phases—a novel approach of eliminating evaporation and reconstitution steps in 96-well SPE. *Rapid Commun Mass Spectrom* 2002;16:1965–1975.
50. Li AC, Junga H, Shou WZ, Bryant MS, Jiang X, Naidong W. Direct injection of solid-phase extraction eluents onto silica columns for the analysis of polar compounds isoniazid and cetirizine in plasma using hydrophilic interaction chromatography with tandem mass spectrometry. *Rapid Commun Mass Spectrom* 2004;18:2343–2350.

51. Song Q, Naidong W. Analysis of omeprazole and 5-OH omeprazole in human plasma using hydrophilic interaction chromatography with tandem mass spectrometry (HILIC-MS/MS)—eliminating evaporation and reconstitution steps in 96-well liquid/liquid extraction. *J Chromatogr B* 2006;830:135–142.

52. Naidong W, Zhou W, Song Q, Zhou S. Direct injection of 96-well organic extracts onto a hydrophilic interaction chromatography/tandem mass spectrometry system using a silica stationary phase and an aqueous/organic mobile phase. *Rapid Commun Mass Spectrom* 2004;18:2963–2968.

53. Matuszewski BK. Standard line slopes as a measure of a relative matrix effect in quantitative HPLC-MS bioanalysis. *J Chromatogr B* 2006;830:293–300.

54. Matuszewski BK, Constanzer ML, Chavez-Eng CM. Strategies for the assessment of matrix effect in quantitative bioanalytical methods based on HPLC-MS/MS. *Anal Chem* 2003;75:3019–3030.

55. FDA. *Guidance for Industry: Bioanalytical Methods Validation*. US Department of Health and Human Services, Center for Drug Evaluation and Research, and Center for Veterinary Medicine, May 2001. Available at http://www.fda/cder/guidance/index.htm.

56. Little JL, Wempe MF, Buchanan CM. Liquid chromatography–mass spectrometry/mass spectrometry method development for drug metabolism studies: examining lipid matrix ionization effects in plasma. *J Chromatogr B* 2006;833:219–230.

57. Sojo LE, Lum G, Chee P. Internal standard signal suppression by co-eluting analyte in isotope dilution LC-ESI-MS. *Analyst* 2003;128:51–55.

58. Tong XS, Wang J, Zheng S, Pivnichny JV, Griffin PR, Shen X, Donnelly M, Vakerich K, Nunes C, Fenyk-Melody J. Effect of signal interference from dosing excipients on pharmacokinetic screening of drug candidates by liquid chromatography/mass spectrometry. *Anal Chem* 2002;74:6305–6313.

59. Shou WZ, Naidong W. Post-column infusion study of the "dosing vehicle effect" in the liquid chromatography/tandem mass spectrometric analysis of discovery pharmacokinetic samples. *Rapid Commun Mass Spectrom* 2003;17:589–597.

60. Schuhmacher J, Zimmer D, Tesche F, Pickard V. Matrix effects during analysis of plasma samples by electrospray and atmospheric pressure chemical ionization mass spectrometry: practical approaches to their elimination. *Rapid Commun Mass Spectrom* 2003;17:1950–1957.

61. Mallet CR, Lu Z, Mazzeo JR. A study of ion suppression effects in electrospray ionization from mobile phase additives and solid-phase extracts. *Rapid Commun Mass Spectrom* 2004;18:49–58.

62. Mei H, Hsieh Y, Nardo C, Xu X, Wang S, Ng K, Korfmacher WA. Investigation of matrix effects in bioanalytical high-performance liquid chromatography/tandem mass spectrometric assays: application to drug discovery. *Rapid Commun Mass Spectrom* 2003; 17:97–103.

63. Pommier F, Frigola R. Quantitative determination of rivastigmine and its major metabolite in human plasma by liquid chromatography with atmospheric pressure chemical ionization tandem mass spectrometry. *J Chromatogr B* 2003;784:301–313.

64. Dams R, Huestis MA, Lambert WE, Murphy CM. Matrix effect in bio-analysis of illicit drugs with LC-MS/MS: influence of ionization type, sample preparation, and biofluid. *J Am Soc Mass Spectrom* 2003;14:1290–1294.

65. Choi BK, Hercules DM, Gusev AI. Effect of liquid chromatography separation of complex matrices on liquid chromatography–tandem mass spectrometry signal suppression. *J Chromatogr A* 2001;907:337–342.

66. Lin ZJ, Desai-Krieger D, Shum L. Simultaneous determination of glipizide and rosiglitazone unbound drug concentrations in plasma by equilibrium dialysis and liquid chromatography–tandem mass spectrometry. *J Chromatogr B* 2004;801:265–272.
67. Arnold RD, Slack JE, Straubinger RM. Quantification of doxorubicin and metabolites in rat plasma and small volume tissue samples by liquid chromatography/electrospray tandem mass spectroscopy. *J Chromatogr B* 2004;808:141–152.
68. Bonfiglio R, King RC, Olah TV, Merkle K. The effects of sample preparation methods on the variability of the electrospray ionization response for model drug compounds. *J Chromatogr B* 1999;13:1175–1185.
69. Borges V, Yang E, Dunn J, Henion J. High-throughput liquid chromatography–tandem mass spectrometry determination of bupropion and its metabolites in human, mouse and rat plasma using a monolithic column. *J Chromatogr B* 2004;804:277–287.
70. Wang Y, Hochhaus G. Simultaneous quantification of beclomethasone dipropionate and its metabolite, beclomethasone 17-monopropionate in rat and human plasma and different rat tissues by liquid chromatography–positive electrospray ionization tandem mass spectrometry. *J Chromatogr B* 2004;805:203–210.
71. Jemal M, Huang M, Mao Y, Whigan D, Powell MD. Increased throughput in quantitative bioanalysis using parallel-column liquid chromatography with mass spectrometric detection. *Rapid Commun Mass Spectrom* 2001;15:994–999.
72. Zeng W, Wang AQ, Fisher AL, Musson DG. A direct injection high-throughput liquid chromatography tandem mass spectrometry method for the determination of a new orally active alphavbeta3 antagonist in human urine and dialysate. *Rapid Commun Mass Spectrom* 2003;17:2475–2482.
73. Chavez-Eng CM, Constanzer ML, Matuszewski BK. Determination of rofecoxib (MK-0966), a cyclooxygenase-2 inhibitor, in human plasma by high-performance liquid chromatography with tandem mass spectrometric detection. *J Chromatogr B* 2000;748: 31–39.
74. Needham SR, Ye B, Smith R, Korte WD. Development and validation of a liquid chromatography–tandem mass spectrometry method for the determination of pyridostigmine bromide from guinea pig plasma. *J Chromatogr B* 2003;796:347–354.
75. Jemal M, Rao S, Gatz M, Whigan D. Liquid chromatography–tandem mass spectrometric quantitative determination of the HIV protease inhibitor atazanavir (BMS-232632) in human peripheral blood mononuclear cells (PBMC): practical approaches to PBMC preparation and PBMC assay design for high-throughput analysis. *J Chromatogr B* 2003;795:273–289.
76. Verhaeghe T, Diels L, de Vries R, de MeulDer M, de Jong J. Development and validation of a liquid chromatographic–tandem mass spectrometric method for the determination of galantamine in human heparinised plasma. *J Chromatogr B* 2003;789:337–346.
77. Constanzer ML, Chavez-Eng CM, Dru J, Kline KF, Matuszewski BK. Determination of a novel substance P inhibitor in human plasma by high-performance liquid chromatography with atmospheric pressure chemical ionization mass spectrometric detection using single and triple quadrupole detectors. *J Chromatogr B* 2004;806:243–250.
78. Zhou L, Glickman RD, Chen N, Sponsel WE, Graybill JR, Lam K-W. Determination of voriconazole in aqueous humor by liquid chromatography–electrospray ionization–mass spectrometry. *J Chromatogr B* 2002;776:213–220.
79. Lakso H, Norström A. Determination of dextropropoxyphene and nordextropropoxyphene in urine by liquid chromatography–electrospray ionization mass spectrometry. *J Chromatogr B* 2003;794:57–65.
80. Yang JZ, Bastian KC, Moore RD, Stobaugh JF, Borchardt RT. Quantitative analysis of a model opioid peptide and its cyclic prodrugs in rat plasma using high-performance

liquid chromatography with fluorescence and tandem mass spectrometric detection. *J Chromatogr B* 2002;779:269–281.

81. Inoue K, Wada M, Higuchi T, Oshio S, Umeda T, Yoshimura Y, Nakazawa H. Application of liquid chromatography–mass spectrometry to the quantification of bisphenol A in human semen. *J Chromatogr B* 2002;773:97–102.
82. Eriksson K, Östin A, Levin J-O. Quantification of melatonin in human saliva by liquid chromatography–tandem mass spectrometry using stable isotope dilution. *J Chromatogr B* 2003;794:115–123.
83. Keevil BG, Lockhart SJ, Cooper DP. Determination of tobramycin in serum using liquid chromatography–tandem mass spectrometry and comparison with a fluorescence polarisation assay. *J Chromatogr B* 2003;794:329–335.
84. Stokes P, O'Connor G. Development of a liquid chromatography–mass spectrometry method for the high-accuracy determination of creatinine in serum. *J Chromatogr B* 2003;794:125–136.
85. Huang Y, Zurlinden E, Lin E, Li X, Tokumoto J, Golden J, Murr A, Engstrom J, Conte J Jr. Liquid chromatographic–tandem mass spectrometric assay for the simultaneous determination of didanosine and stavudine in human plasma, bronchoalveolar lavage fluid, alveolar cells, peripheral blood mononuclear cells, seminal plasma, cerebrospinal fluid and tonsil tissue. *J Chromatogr B* 2004;799:51–61.
86. Leahy M, Felder L, Goytowski K, Shou W, Weng N. 52nd ASMS Conference on Mass Spectrometry and Allied Topics, Nashville, Tennessee, 23–27 May 2004.
87. Zeng W, Musson DG, Fisher AL, Wang AQ. A new approach for evaluating carry-over and its influence on quantitation in high-performance liquid chromatography and tandem mass spectrometry assay. *Rapid Commun Mass Spectrom* 2006;20:635–640.
88. Wang S, Cyronak M, Yang E. Does a stable isotopically labeled internal standard always correct analyte response? A matrix effect study on a LC/MS/MS method for the determination of carvedilol enantiomers in human plasma. *J Pharm Biomed Anal* 2007; 43:701–707.
89. Yang L, Wu N, Clement RP, Rudewicz PJ. Validation and application of a liquid chromatography–tandem mass spectrometric method for the determination of SCH 211803 in rat and monkey plasma using automated 96-well protein precipitation. *J Chromatogr B* 2004;799:271–280.
90. Khan MN. Immunoassay in drug development arena: an old player with new status. *J Clin Ligand Assay* 1999;22:242–253.
91. Findlay JWA, Smith WC, Lee JW, Nordblom GD, Das I, DeSilva BS, Khan MN, Bowsher RR. Validation of immunoassay for bioanalysis: a pharmaceutical industry perspective. *J Pharm Biomed Anal* 2000;21:1249–1273.
92. Wadhwa M, Bird C, Dilger P, Gaines-Das R, Thorpe R. Strategies for selection, measurements and characterization of unwanted antibodies induced by therapeutic biologicals. *J Immunol Methods* 2003;278:1–17.
93. Shankar G, Shores E, Wagner C, Mire-Sluis A. Scientific and regulatory considerations on the immunogenicity of biologics. *Trends Biotechnol* 2006;24:274–280.
94. DeSilva B, Smith W, Weiner R, Kelly M, Smolec J, Lee B, Khan M, Tacey R, Hill H, Celniker A. Recommendations for the bioanalytical method validation of ligand-binding assay to support pharmacokinetics assessment of macromolecules. *Pharm Res* 2003;20: 1885–1900.
95. Lee JW, Dewanarayan V, Barrett YC, Weiner R, Allison J, Fountain S, Keller S, Weinryb I, Green M, Duann L, Rogers JA, Millham R, O'Brian PJ, Sailstad J, Khan M, Ray C, Wagner JA. Fit-for-purpose method development and validation for successful biomarker measurements. *Pharm Res* 2006;32:312–328.

96. Mire-Sluis AR, Barett Y, Devanarayan V, Koren E, Liu H, Maia M, Parish T, Scott G, Shankar G, Shores E, Swanson SJ, Taniguchi G, Wierda D, Zuckerman LA. Recommendations for the design and optimization of immunoassays used in the detection of host antibodies against biotechnology products. *J Immunol Methods* 2004;289:1–16.
97. Crowther JR. *The ELISA Guidebook*, Methods in Molecular Biology, Vol 149. Totowa, NJ: Humana Press; 2001.
98. Wild D. *The Immunoassay Handbook*, 2nd ed. New York: Nature Publishing Group; 2001.
99. Rodbard D, Feldman Y, Jaffe ML, Miles Le EM. Kinetics of two-site immunoradiometeric (sandwich) assay II. *Immunochemistry* 1978;15:77–82.
100. Plikaytis BD, Holder PF, Pais LB, Maslanka SE, Gheesling LL, Carlone GM. Determination of parallism and nonparallelism in bioassay dilution curves. *J Clin Microbiol* 1994;32:2441–2447.
101. Shah VP, Midha KK, Dighe SV, McGiveray JJ, Skelly JO, Yacobi A, Laylogg T, Viswanathan CT, Cook CE, McDowall RD, Putman KA, Spector S. Analytical method validation: bioavailability, bioequivalence, and PK studies. *J Pharm Sci* 1992;81: 309–312.
102. Miller KJ, Bowsher RR, Celniker A, Gibbons J, Gupta S, Lee JW, Swanson J, Smith WC, Weiner RS. Workshop on bioanalytical method validation for macromolecules: summary report. *Pharm Res* 2001;18:1373–1383.
103. Smolec J, DeSilva B, Smith W, Weiner R, Kelly M, Lee B, Khan M, Tacey R, Hill H, Cekniker A. Bioanalytical method validation for macromolecules in support of pharmacokinetic studies. *Pharm Res* 2005;22:1425–1431.

# ANALYTICAL CHEMISTRY METHODS: DEVELOPMENTS AND VALIDATION

Izet M. Kapetanovic[1] and Alexander V. Lyubimov[2]

[1]*NIH NCI Division of Cancer Prevention, Chemoprevention Agent Development Research Group, Bethesda, Maryland*

[2]*University of Illinois at Chicago, Chicago, Illinois*

**Contents**

1 Introduction
  1.1 Background
  1.2 Purpose
2 Analytical Method Development
  2.1 Sampling and Handling
  2.2 Sample Preparation
  2.3 Analytical Techniques for Method Development
3 Analytical Method Validation
  3.1 Installation, Operation, Performance Qualification, and Maintenance of Instrumentation
  3.2 Validation Parameters and Definition
  3.3 Guidelines: ICH, FDA, AOAC, USP, ISO 9000, and ISO 17025
  3.4 Good Laboratory Practice (GLP)
  3.5 Method Validation Protocol
4 Conclusion
  References

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

# 1 INTRODUCTION

## 1.1 Background

Analytical methodology plays an integral part of drug development from the initial synthesis and manufacture through clinical trials and postmarketing monitoring. In new drug applications (NDAs), abbreviated new drug applications (ANDAs), biologics license applications (BLAs), or product license applications (PLAs), data must be submitted to establish that the analytical procedures used in testing meet proper FDA and ICH guidelines of accuracy and reliability. At the time of submission, the NDA, ANDA, BLA, or PLA should contain method validation information to support the adequacy of the analytical procedures. All analytical methods should be validated (to demonstrate that the analytical procedures are suitable for their intended use) prior to their use. The suitability of a compendial analytical procedure must be verified under the actual condition of use, but full validation is not required. As in all other preclinical pharmacology/toxicology studies, analytical assays should be conducted according to the FDA's Good Laboratory Practices (GLPs) (21 CFR part 58). Selective and sensitive validated analytical methods for the quantitative evaluation of drugs are critical for the successful conduct of preclinical safety and efficacy studies.

## 1.2 Purpose

Analytical methods are required for determination of the following:

- Active pharmaceutical ingredient (API)—manufacturing control, stability testing, shelf-life forecast.
- Synthetic contaminants, degradation products—manufacturing control, stability testing, shelf-life forecast.
- Pharmaceutical excipients (binders, disintegrants, lubricants, antioxidants and preservatives, suspending and dispersing agents, natural or artificial flavorings and colorings, coatings, etc.)—manufacturing control, stability testing. The *United States Pharmacopeia and the National Formulary* (USP-NF) contains information on the identity, strength, quality, and purity for more than 250 excipients including official test methods in analytical testing.
- Drug and metabolites in biological fluids and tissues—bioavailability, pharmacokinetics.

It is important to have validated analytical methods from the start and throughout the drug development process. The generic approach to method development involves three basic steps: sample preparation (clean up), separation, and detection. A schematic flow diagram depicting the most commonly used options is shown in Fig. 1. While general approaches to method development are similar, complexity and requirements of time and resources vary. For example, the *United States Pharmacopeia-National Formulary* (USP-NF) provides standard methods for pharmaceutical excipients, obviating a need for analytical method development, and will not be specifically addressed here. On the other hand, analysis in biological fluids and tissues is the most time and labor intensive due to complexity and variability

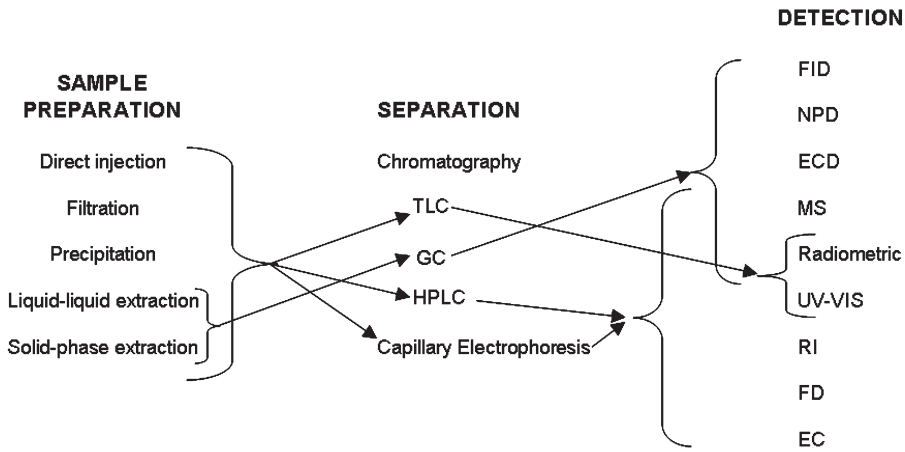


**FIGURE 1** General scheme for developing an analytical method based on separation techniques. Most commonly used approaches are depicted.

of the biological matrices and the relatively low levels of analytes. Selection of appropriate sample preparation, separation, and detection approaches depends on the nature of the sample, the nature of the sample matrix, the concentration of analyte(s) in the matrix, the laboratory levels of skill and available resources, and so on.

## 2 ANALYTICAL METHOD DEVELOPMENT

### 2.1 Sampling and Handling

The proper sampling and handling of samples is one of the most important factors in a high quality analysis. The reader is referred to an excellent resource, Chapter 7 in *Quality Assurance Principles for Analytical Laboratories* [1], for details of sampling and handling. The main principles are described below.

There are three basic activities involved in solving an analytical problem: (1) collection and handling of the appropriate sample, (2) preparation of sample for analysis, and (3) analysis using an appropriate method.

These activities are independent of each other, but one significantly influences the others. Failure to observe standard procedures for sample collection, handling, and documentation often impairs the intended purpose of the analysis itself. In many cases, the collection of samples falls outside a laboratory's control. However, the laboratory must do all it can to receive appropriate, applicable, and defensible samples.

Key points for sampling are the following:

- The laboratory shall have a sampling plan and procedures for sampling when it carries out sampling of substances, materials, or products for subsequent testing or calibration.

- Sampling plans shall, whenever reasonable, be based on appropriate statistical methods.
- Sampling procedures shall describe the selection, sampling plan, withdrawal, and preparation of a sample or samples from a substance, material, or product to yield the required information.
- The laboratory shall have procedures for recording relevant data and operations relating to sampling.

It is extremely important to perform accurate subsampling (taking the appropriate test portion for analysis) in the laboratory. The number of samples to be analyzed in a given situation is limited by the resources available for the collection of the samples or for their analysis. Analyses of multiple samples always are preferred over single samples since single samples give no information on the homogeneity of the lot that was sampled. In addition, for single samples, the sampling error is also confounded with the analytical error. It is a common practice to collect several samples for test substance and dosage formulation analyses to assess homogeneity. For example, at least three samples from the top, middle, and bottom portions of the dietary admixture or dosage formulation should be collected and usually each portion is analyzed in triplicate to assess homogeneity. However, for a bioanalytical analysis, a single sample is usually analyzed because of limited volume of sample and high cost of each analysis. Each sample should be prepared in a way that it achieves homogeneity and should be handled in a manner that prevents alteration from the original composition. Even seemingly homogeneous materials such as liquids may be subject to sedimentation or stratification.

Shipment or delivery of the sample should be done under appropriate conditions (e.g., temperature) in a proper container package and with a proper label and should be accompanied by meaningful documentation, including Material Safety Data Sheet (MSDS) and instructions for sample storage upon delivery. The documentation consists of a chain-of-custody or similarly named document that accompanies the sample as it moves to and through the laboratory. A dependable record of sample handling is important. The sample shall be accepted by a designated person (a custodian in laboratories of appreciable size) who documents the action by completing a sample accountability record. This document will contain the sample number, the name of the product, and the date received; will indicate who received it; will describe the method of shipment or delivery, the packages received, and their condition; and will provide space for recording various storage locations before and after analysis. Deliveries of the sample, or parts of the sample, to the analyst, and its return, shall also be recorded on this form, as shall the signed statement concerning the final disposition of the reserve sample. A two-part form can be used for this purpose. One copy remains with the custodian and the other moves with the sample through the laboratory and is used by the supervisor for sample management purposes. Some laboratories use a Sample Receiving Log book for sample control. The information entered in the Log book is essentially the same as described for the two-part form.

For proper handling of the test samples and calibration items, the laboratory shall have written procedures for their transportation, receipt, handling, protection, storage, retention, and/or disposal. The laboratory shall have a system for identifying

test and/or calibration items. The laboratory shall also have documented procedures and appropriate facilities to avoid deterioration, damage, loss, or cross-contamination of any test item or sample during storage and handling. All necessary environmental conditions, including special security arrangements for sample integrity as needed for some samples, shall be established, maintained, monitored, and recorded.

## 2.2 Sample Preparation

Sample preparation is a very important first and often critical step in the analytical method, especially when biological samples are involved. Typically, it is the most difficult and time-consuming step. The purpose of sample preparation is to isolate analytes of interest from interfering sample components, concentrate the analytes, and dissolve them in a suitable solvent for subsequent separation and detection. Sample preparation significantly impacts on the recovery, selectivity, sensitivity, reproducibility, and ruggedness of the analytical method. Sample cleanliness or, more often, a lack thereof may have a direct effect (e.g., full or partial chromatographic peak overlap) on detection and quantification of analytes of interest, but also an indirect effect (e.g., contamination of analytical column or mass spectrometry (MS) ion source, ionization suppression in MS). Not only can some of the interferences lead to a loss in sensitivity (e.g., ionization suppression in MS) [2], but this effect may be variable depending on individual samples [3]. However, in some instances, the coextracted materials may be helpful. For example, coextractants can neutralize active sites on glassware, injector, or column and thereby minimize adsorptive effects. Therefore, coextractants may have a positive or negative effect and should be considered during analytical method development. Consequently, it is strongly advisable that all comparisons of extracted biological samples be made to similar extracts of spiked blank biological matrix instead of neat solutions.

Commonly encountered sample problems include (1) lack of compatibility with chromatographic system (e.g., tendency to clog or degrade the analytical column), (2) being too dirty (presence of coextracted materials that interfere with separation or detection of analytes), and (3) being too dilute (sensitivity issue). More common approaches to sample preparation will be discussed later, while the less common ones (turbulent flow chromatography, monolithic chromatography, immunoaffinity, etc.) will not be included here.

Comparison of most commonly employed sample preparation techniques is presented in Table 1. In general, simpler, faster methods tend to generate dirtier samples, which tend to limit the number of injections that can be made on a column or into a chromatographic system before chromatographic or system conditions start to deteriorate.

***Direct Injection*** Direct injection approaches have been developed to facilitate and expedite analysis of biological samples and are more suitable for liquid as opposed to gas phase separation methods. They include direct injection with or without prior dilution, column switching, and restricted access medium (RAM) or internal surface reversed-phase (ISRP) column methods.

A simple dilution and injection approach is not commonly used because it requires high analyte concentration and/or high detector response to an analyte.

**TABLE 1 Relative Comparison of Commonly Used Sample Preparation Methods**

| Factor | Direct Injection | Filtration | Precipitation | LLE | SPE |
|---|---|---|---|---|---|
| Simplicity | ++++ | +++++ | +++++ | +++ | +++ |
| Speed | +++ | ++++ | ++++ | + | +++ |
| Resultant sample cleanliness | ++ | + | + | +++++ | +++++ |
| Resultant analyte concentration | + | + | + | + | +++ |
| Selectivity | + | + | + | +++ | ++++ |
| Solvent consumption | ++ | + | +++ | +++++ | ++ |
| Possible injections per column | ++ | ++ | ++ | +++++ | +++++ |

This straightforward approach may be suitable for analysis of simple matrices (e.g., API or formulated pharmaceutical product) but generally is not useful for the analysis of biological samples due to low analyte concentration and numerous complex interferences.

Column switching for separation and enrichment of analytes of interest has been in use for a long time [4]. Hagestam and Pinkerton [5] introduced an online cleaning method for HPLC analysis with a RAM/ISRP column [6–8]. RAM columns operate via both size exclusion and phase partitioning modes. Prefiltered samples are injected onto a RAM column, which retains small molecules in the inner hydrophobic pores while larger ones go to waste unretained. Subsequently, small molecules are eluted from the RAM column onto an analytical reversed-phase column, where they are concentrated prior to elution with an appropriate strength mobile phase. A number of manufacturers offer RAM columns with similar basic principles but some variations on the theme (e.g., http://www.registech.com/ram/#SPS). The advantages of this method include speed, simplicity, automation, potential increased safety for analysts due to lesser exposure, and applicability to varied biological, botanical, and environmental matrices. However, the online sample cleaning methods are, in general, subject to carry-over and injector clogging problems. Another drawback of the RAM method is a relatively low sensitivity. However, recent improvements in HPLC column technology and evolution of LC-MS and LC-MS/MS instrumentation have provided additional selectivity and sensitivity [8–10]. Hogendoorn et al. [11] demonstrated suitability of RAM columns for LC-MS/MS analysis of analytes with a wide range of polarities in human serum. Other similar techniques include size exclusion chromatography and turbulent flow chromatography (narrow bore columns packed with large particles) but will not be further discussed here.

The column switching approach utilizes switching valves to redirect the flow of a mobile phase between two columns—a cleanup column and an analytical separation column. Limitations include a requirement of mobile phase compatibility between the two columns. An example of column switching includes online solid phase extraction (SPE, to be discussed later). Good sensitivity and precision were reported in a comparison of online and offline SPE methods for measurement of six isoflavones and lignans in urine [12]. However, the online method was found to be more sensitive, precise, reproducible, and cost effective. Another variation involves a ternary-column system [13], which utilizes dual extraction columns in a

parallel configuration. This allows use of one column for extraction while the other is being equilibrated and thereby decreasing the injection cycle time and increasing the sample throughput.

***Filtration*** Various filtration methods serve similar purposes and have similar advantages and disadvantages to those of protein precipitation (see next section). There are basically two types of filters: depth and screen. Depth filters are randomly oriented fibers that retain particles throughout the matrix rather than just on the surface. They have a higher load capacity than screen filters [14]. Screen filters are polymeric membranes that have a uniform distribution of pore sizes. They are relatively thin so that there is a minimal amount of liquid retention. Screen filters clog more rapidly than depth filters.

In developing a method that requires filtration, adsorption of the analyte onto the filter must be taken into account. For dilute solutions of adriamycin, more than 95% is adsorbed to cellulose ester membranes and about 40% to polytetrafluoroethylene membranes [15]. For more concentrated solutions, as would be encountered in bulk formulation testing, filter adsorption is not as important a concern. Nevertheless, the common practice is to discard the first several milliliters of the filtrate. For protein-based products, there is significant nonspecific binding to nylon-based microporous membranes and minimal binding to hydrophilic polyvinylidene fluoride membranes [16]. Ultrafiltration removes protein and protein-bound drug, which makes it suitable for plasma-free drug concentration measurements. Dialysis, on the other hand, is primarily useful for protein binding studies.

***Precipitation*** Protein precipitation or deproteinization (commonly acetonitrile, perchloric acid, or methanol) is also a relatively simple method of sample preparation and minimally eliminates interferences from large molecules. However, interferences from small molecules remain and this approach is more suited for analytical methods that offer additional selectivity such as LC-MS and LC-MS/MS. Precipitation with acetonitrile or methanol may be combined with subsequent extraction and may be used for HPLC or LC-MS/MS analytical methods, which employ these solvents as mobile phases. In addition, matrix-based interferences can create indirect problems in MS detection due to interference with the ionization process. Furthermore, sensitivity may also be an issue as this method does not include a preconcentration step. Matrix interferences place additional strain on the analytical system. Occasionally, sample dilution prior to precipitation may be needed to avoid trapping analytes within precipitated matrix. One significant advantage of protein precipitation is that it extracts all analytes regardless of their polarity. This overcomes limitations of other extraction methods (e.g., LLE, SPE), which may not effectively extract analytes with a wide range of polarities [17]. The precipitation approach is amenable to automation.

***Liquid–Liquid Extraction (LLE)*** LLE is one of the oldest and widely used methods for sample preparation. It is based on analyte (solute) partitioning between immiscible solvents according to relative analytes' solubilities in the solvents. In order to optimize separation in terms of recovery, selectivity, and sample cleanliness, several extractions with the same or different solvent combinations may be necessary. For example, acidic analytes can be extracted from acidified aqueous solution

into organic solvent (e.g., methyl-*t*-butyl ether), followed by their extraction from methyl-*t*-butyl ether into a basic aqueous solution (e.g., ammonium hydroxide), which is neutralized, and analytes are back-extracted into methyl-*t*-butyl ether. Commonly used extraction solvents include hexane, methyl-*t*-butyl ether, ethyl acetate, chloroform, methylene chloride, and dilute acid or base. Modifiers such as a small percentage of isopropanol can also be added. Fine tuning in LLE can be carried out by judicious selection of solvents, solvent modifiers, solvent volumes, pH, and so on. Not all extractions need to involve aqueous and organic phases. There are some organic solvent combinations that are immiscible and can be used in LLE. For example, the acetonitrile : hexane combination is immiscible and can be used to clean somewhat polar analytes from lipophilic components in a biological matrix. LLE is labor intensive, slow, and not easily amenable to automation. Liquid handling stations are making automation more feasible. LLE generally requires large solvent consumption and solvent evaporation, which have environmental, safety, and health concerns. Finding a suitable solvent system and conditions may be difficult for simultaneous extraction of analytes with different polarities. There is also a tendency for emulsion formation, which introduces further complications. Adsorptive losses may result during evaporation and it may be necessary to pacify the active surface by sylation. Some solvents have a tendency for peroxide formation (e.g., diethyl ether, ethyl acetate) that can lead to analyte stability problems. Methyl-*t*-butyl ether has roughly similar polarity but a lesser propensity for peroxide formation and can often be substituted for those solvents.

***Solid-Phase Extraction (SPE)*** SPE is analogous to offline gradient liquid chromatography and has overtaken LLE in popularity as it lacks a number of the drawbacks associated with LLE. It is relatively easy to automate, requires small solvent volumes, can concentrate samples without evaporation, and yields cleaner extracts and higher recoveries for polar analytes. However, as mentioned earlier, there are cases where SPE may not be suitable for analytes with a broad range of polarities [17]. Earlier manufacturing problems in batch-to-batch reproducibility of the column materials have mostly been resolved. Cartridge columns with a wide variety of bonded phase packing materials (phases or sorbents) are commercially available (or custom prepared, if desired). Cartridge columns are obtainable in different sizes (e.g., 1 mL, 3 mL, 6 mL, 12 mL) with different amounts of packing material (e.g., 30 mg, 50 mg, 60 mg, 100 mg, 150 mg, 200 mg, 500 mg) to accommodate different sample sizes and not to mass or volume overload SPE columns. As a rough guide, nonionic silica-based sorbents can retain a mass of solute approximately equivalent to 1–5% of sorbent mass or 1–5 mg of solute per 100 mg of sorbent. Polymeric sorbents have a larger surface area and can accommodate solute up to 15% of sorbent mass. However, keep in mind that solute represents analytes plus coextracted material. As a rule of thumb, solvent volumes between 4 and 8 times the bed volume are needed to ensure appropriate conditioning, washing, and elution. Silica-based SPE columns have approximately 150 μL bed volume per 100 mg of sorbent, and polymeric sorbents about 250 μL per 100 mg of sorbent. Most commonly used sorbents comprise siloxane-bonded materials with different functional groups to provide normal-phase, reversed-phase, ion exchange, or mixed-mode chromatographic separation. General characteristics of different phases are summarized in Table 2. Some of the phases can operate in different modes. For example,

**TABLE 2 Characteristics of Common Silica-Based SPE Phases**

| | Normal | Reversed | Ion Exchange |
|---|---|---|---|
| Sorbent functional group | Silica<br>Cyano<br>Diol<br>Amino | Octadecyl (C18)<br>Octyl (C8)<br>Methyl (C1)<br>Ethyl (C2)<br>Phenyl | Strong cation exchange (SCX; p$K_a$ 2.0–3.0): benzene sulfonylpropyl<br>Strong anion exchange (SAX; p$K_a$ 11.0–12.0): quarternary amine<br>Amine (p$K_a$ 9.5–10.5) |
| Analytes | Polar | Nonpolar | Ionized |
| Elution factor(s) | Solvent strength | Solvent strength | Ionic interaction (pH, ionic strength, counterion strength) |
| Increasing elution strength | Hexane<br>Toluene<br>Chloroform<br>Methylene chloride<br>Tetrahydrofuran<br>Ethyl acetate<br>Acetone<br>Acetonitrile<br>Isopropanol<br>Methanol<br>Water | Water<br>Methanol<br>Isopropanol<br>Acetonitrile<br>Acetone<br>Ethyl acetate<br>Tetrahydrofuran<br>Methylene chloride<br>Chloroform<br>Toluene<br>Hexane | Cations:<br>$Li^+$<br>$Na^+$<br>$NH^+$<br>$K^+$, $Mg^{2+}$<br>$Ca^{2+}$<br>Anions:<br>Hydroxide, fluoride, propionate<br>Acetate, formate<br>Phosphate, carbonate<br>Sulfate<br>Nitrate<br>Citrate |

cyano phase can be used either in normal-phase or reversed-phase mode and amino phase can be used in normal phase or anion exchange phase. Undesirable secondary interactions can occur with free silanol moieties and can be minimized by end-capping.

SPE cartridges require prior conditioning to remove any impurities from them and make the sorbent compatible with the sample solvent and allow sample retention. Two conditioning washes are commonly used (each 1–2 mL per 100 mg of sorbent): the first to clean the column and the second (similar to or weaker than the sample solvent) to make the column conducive to sample retention. Sample application and elution should not be performed too rapidly as sufficient residence time must be allowed for effective analyte–sorbent interaction. In order to avoid channeling, it is also important not to allow sorbent to dry between conditionings and prior to sample application. After cartridge conditioning, liquid sample (in the weakest strength solvent possible) is applied and allowed to pass through the

column via gravity, pressure, or vacuum. Sample retained on the column should be washed with a stronger solvent than the one used in sample loading but weaker than solvent that would cause its elution. Elution of analytes should be performed with solvent strong enough to allow elution in a relatively small volume. This may avoid or minimize a need for a concentration step (e.g., evaporation) prior to sample analysis. Solvent combination may sometimes be necessary for elution (e.g., methanol : water) in order to minimize the elution volume and/or coelution of undesirable substances. Ion exchange phases require additional considerations. For effective analyte retention, sample pH should be at least 2 pH units lower or 2 pH units greater than the $pK_a$ of the corresponding cation or anion so approximately 99% of analyte remains ionized. In addition, low ionic and high solvent strength promote retention and elution, respectively. The SPE phase and the solvent system can be selected depending on specific goals: analyte elution (interference retention), interference elution (analyte retention with subsequent elution; most common), or analyte concentration. While evaporation may not be necessary when using small elution volumes, it may be advantageous in order to further increase assay sensitivity and improve chromatographic behavior (e.g., peak shapes) by dissolving the extracted sample in the mobile phase or a weaker solvent. Chromatographic artifacts can be encountered when injecting samples, especially larger sample volumes, in a solvent significantly different from the mobile phase.

More recently, miniaturized SPE formats (SPE disks and pipette tips) have been developed. These allow use of smaller sample and elution volumes, provide large flow area, and avoid channeling. SPE disks can easily be adapted to an online 96-well approach.

Comparison of plasma sample purification by LLE, automated LLE (96-well collection plate with robotic liquid handling system), and automated SPE (96-well collection plate with robotic liquid handling system) [18] found total time for automated SPE to be slightly less than that for automated LLE, and both were three times faster than manual LLE. Total time for the automated SPE could be reduced further by a factor of 2 if the evaporation step was omitted. Analyst time was lower by a factor of 25 for the two automated methods as compared to the manual LLE. Overall results in terms of accuracy and precision were comparable for the three methods.

Advantage of SPE over simpler methods like precipitation is shown in Fig. 2. More effective cleanup yielded a cleaner chromatogram, decreased ionization suppression by coeluting matrix components, and thereby improved signal to noise by more than tenfold.

Solid-phase microextraction (SPME) represents another recent variation on the SPE method [19]. This method of preparation does not use solvent and the separation takes place on a fused-silica fiber, which is coated with a suitable stationary phase. The fiber is inserted into the sample and allowed to extract analytes via absorption and/or adsorption. Analytes are then eluted either thermally or via mobile phase. This technique can be used in conjunction with GC, CE, or HPLC and is nicely suited for MS.

***Derivatization*** Derivatization may sometimes be required in order to improve volatility, stability, chromatographic behavior, sensitivity, and selectivity. Ideally, derivatization reactions should be rapid, simple, and mild and should yield a single

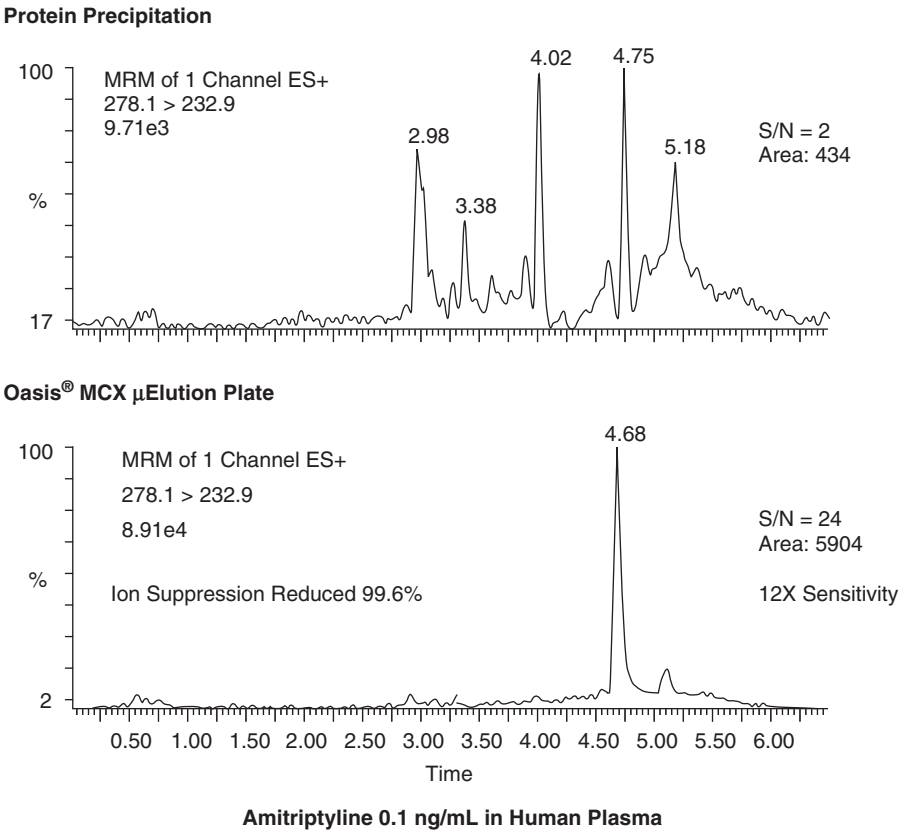


**FIGURE 2** Effect of sample cleanup on sensitivity in LC-MS. Upper chromatogram depicts sample after precipitation cleanup, while the lower tracing is for analogous sample following a SPE cleanup. (Used with permission from Waters Corp. )

product. Choice of derivatizing moiety depends on the need (e.g., stability, sensitivity, chirality), available derivatizable groups on the analyte (e.g., carboxyl, amine, hydroxyl, thiol), separation method (e.g., GC, HPLC), and detection method (e.g., electrochemical, ultraviolet, fluorescent). Tables 3 and 4 contain examples of commonly used derivatizing agents for GC and HPLC, respectively. In addition, derivatization can also improve ionization efficiency in atmospheric pressure ionization mass spectrometry (API-MS) and thereby improve sensitivity, which may be compromised by coextracted substances [20]. Derivatization can also help identify and characterize metabolites. For example, *N*-methyl-*N*-(*tert*-butyldimethylsilyl) derivatives tend to yield stable molecular (or quasimolecular) ions in mass spectrometry and thereby allow determination of molecular weight. Ions and fragments can be identified easier by using a mixture of deuterated and nondeuterated (or other stable labeled derivatizing agent) agents. A number of reviews have been published on various derivatizing techniques for the most common separation techniques including GC [21–24], HPLC [25–32], and capillary electrophoresis (CE) [26–28, 33, 34]. Other derivatization approaches have also been reported (e.g., combining sample preparation and derivatization) [35]. While derivatization offers pos-

**TABLE 3 Examples of Commonly Used GC Derivatizing Agents**

| Functional Moiety | Derivatizing Agent | Enhanced Detector Response |
|---|---|---|
| Alcohols | Heptafluorobutyric anhydride | ECD |
| | Pentafluorobenzyl chloride | ECD |
| | Bis(trimethylsilyl)trifluoroacetamide (BSTFA) | |
| | *N*-Methyl-*N*-(*t*-butyldimethylsilyl)trifluoroacetamide (MTBSTFA) | |
| Amines (primary) | Chloroacetic anhydride | ECD |
| | Pentafluorobenzoyl chloride | ECD |
| | Pentafluorobenzaldehyde | ECD |
| | Bis(trimethylsilyl)trifluoroacetamide (BSTFA) | |
| | *N*-Methyl-*N*-(*t*-butyldimethylsilyl)trifluoroacetamide (MTBSTFA) | |
| Amines | Heptafluorobutyric anhydride | ECD |
| | Pentafluorobenzyl chloride | ECD |
| | Bis(trimethylsilyl)trifluoroacetamide (BSTFA) | |
| | *N*-Methyl-*N*-(*t*-butyldimethylsilyl)trifluoroacetamide (MTBSTFA) | |
| Amino acids | Acetic acid | |
| | Heptafluorobutyric anhydride | ECD |
| | Bis(trimethylsilyl)trifluoroacetamide (BSTFA) | |
| | *N*-Methyl-*N*-(*t*-butyldimethylsilyl)trifluoroacetamide (MTBSTFA) | |
| Carboxylic acids | Pentafluorobenzyl chloride | ECD |
| | Pentafluorobenzyl alcohol | ECD |
| | Diazomethane | |
| | Hydrochloric acid + methanol | |
| | Bis(trimethylsilyl)trifluoroacetamide (BSTFA) | |
| | *N*-Methyl-*N*-(*t*-butyldimethylsilyl)trifluoroacetamide (MTBSTFA) | |
| Phenols | Heptafluorobutylimidazole | ECD |
| | Pentafluorobenzyl chloride | ECD |
| | Bis(trimethylsilyl)trifluoroacetamide (BSTFA) | |
| | *N*-Methyl-*N*-(*t*-butyldimethylsilyl)trifluoroacetamide (MTBSTFA) | |

sibilities for significant analytical improvements, appropriate caution needs to be exercised in order to avoid potential pitfalls, such as various artifacts with trimethylsilylation [36].

## 2.3 Analytical Techniques for Method Development

### *Nonseparation Methods*

#### *Absorption and Emission Spectroscopy*

ULTRAVIOLET (UV)–VISIBLE (VIS) Spectrophotometry is a technique of detecting and measuring absorption of electromagnetic radiation by molecules with π or unshared electrons. The UV-VIS spectrum is divided into 200–380 nm (UV) and 380–800 nm (VIS) regions. The Beer–Lambert law is used for measurement of ana-

**TABLE 4 Examples of Commonly Used HPLC Derivatizing Agents**

| Functional Moiety | Derivatizing Agent | Enhanced Detector Response |
|---|---|---|
| Alcohols | Benzoyl chloride | |
| | 4-Dimethylamino-1-naphthoyl nitrile | FD |
| | 3,5-Dinitrobenzyl chloride | |
| | *p*-Nitrobenzoyl chloride (4-NBCl) | |
| | *p*-Nitrophenyl chloroformate | |
| | *p*-Iodobenzensulfonyl chloride | |
| Amines (primary) | Benzoyl chloride | |
| | 4-Chloro-7-nitrobenzo-2-oxa-1,3-diazole (NBD-Cl) | FD |
| | Dansyl chloride (DnS-Cl) | FD |
| | Flourescamine | FD |
| Amines | Benzoyl chloride | |
| | 4-Chloro-7-nitrobenzo-2-oxa-1,3-diazole (NBD-Cl) | FD |
| | Dansyl chloride (DnS-Cl) | |
| | 3,5-Dinitrobenzyl chloride | |
| | *p*-Nitrobenzoyl chloride (4-NBCl) | FD |
| | *o*-Phthaldialdehyde (OPT) | FD |
| Amino acids | Dansyl chloride (DnS-Cl) | FD |
| | Flourescamine | FD |
| | Pyridoxal | FD |
| | *o*-Phthaldialdehyde (OPT) | FD |
| Carboxylic acids | Benzyl bromide | |
| | *p*-Bromophenacyl bromide | |
| | 4-bromomethyl-7-methoxycoumarin (Br-Mmc) | FD |
| | 9,10-Diaminophenanthrene | FD |
| | Phenacyl bromide | |
| Ketones/aldehydes | *p*-Nitrobenzylhydroxylamine hydrochloride | |
| | 2-Diphenylacetyl-1,3-indandione | FD |
| | 2,4-Dinitrophenyl hydrazine | |
| Phenols | 4-Chloro-7-nitrobenzo-2-oxa-1,3-diazole (NBD-Cl) | FD |
| | Dansyl chloride (DnS-Cl) | FD |
| | 3,5-Dinitrobenzyl chloride | |
| Thiols | 4-Chloro-7-nitrobenzo-2-oxa-1,3-diazole (NBD-Cl) | FD |
| | *N*-(9-Acridinyl)maleimide | FD |

lytes. While UV-VIS detectors are commonly used for HPLC analysis, UV-VIS spectrophotometry without prior chromatographic separation is not well suited for complex biological samples. It is frequently used for analysis of API or other substances in the manufacturing process and in USP methods [37]. Further improvements in UV-VIS spectrophotometry are being made, for example, UV derivative spectrophotometry [38].

INFRARED (IR) IR is a technique of detecting absorption of energy resulting from transitions between vibrational and rotational energy levels of molecules. It is in the

0.5–200 μm (20,000–50 $cm^{-1}$) region of the electromagnetic spectrum. IR is commonly used for characterization of API and in the manufacturing process [39, 40].

ATOMIC ABSORPTION SPECTROMETRY (AAS) Primary applications of AAS in the pharmaceutical area involve quantification of metal ions, metal-containing organic compounds, and inorganic molecules (e.g., arsenic trioxide, different platinum analogues, silver, mercury, copper, palladium) [41–48] in drug formulations, tissues, and biological liquids. It is also used for trace element control in natural plant products and in synthesized drug substances [49–52]. AAS is also used for indirect measurements of organic substances based on formation of ion-paired complexes with the drugs [53–56].

AAS detects absorption or emission of light in thermally vaporized and extensively atomized molecules. Atomization is achieved via either flame aspiration or electrothermal atomization (graphite furnace). The former also allows emission analysis while the latter ashes components and diminishes matrix effects. Graphite-furnace atomic absorption spectrometry (GF AAS) is well suited for analysis of biological samples as it only requires very small sample size (20 μL) and provides increased sensitivity due to a longer residence time. More details on HGAAS (hydride generation atomic absorption spectrometry) can be found in Dědina and Tsalev [57]. All aspects of the theory, instrumentation, and practical usefulness of electrothermal atomization for analytical atomic spectrometry can be found in Jackson [58]. There are also some general guides on the theory and practice of AAS [59–62]. Recently, a fully validated GF AAS method has been described for the cancer chemotherapeutic oxaliplatin [48].

*Ligand-Binding Assays* A ligand-binding assay (LBA) is a technique that uses specific antigen or antibody, capable of binding to the analyte, to identify and quantify substances. The antibody can be linked to a radioisotope (radioimmunoassay, RIA), to an enzyme that catalyzes an easily monitored reaction (enzyme-linked immunosorbent assay, ELISA), or to a highly fluorescent compound by which the location of an antigen can be visualized (immunofluorescence).

Because LBAs are indirect assays dependent on binding interactions, factors such as lipemic and hemolyzed samples, binding proteins, and anticoagulants that interfere with this process will destabilize the assay.

ELISA is a useful and powerful method in estimating ng/mL to pg/mL of test articles and metabolites in the solution, such as serum, urine, and culture supernatant. It is a relatively easy task to develop ELISA if "suitable" antibodies against materials of interest such as proteins, peptides, and drugs are readily available. ELISAs combine the specificity of antibodies with the sensitivity of simple enzyme assays by using antibodies or antigens coupled to an easily assayed enzyme. ELISAs can provide a useful measurement of antigen or antibody concentration. There are two main variations on this method: ELISA can be used to detect the presence of antigens that are recognized by an antibody or it can be used to test for antibodies that recognize an antigen. An ELISA is a five-step procedure: (1) coat the microtiter plate wells with antigen; (2) block all unbound sites to prevent false-positive results; (3) add antibody to the wells; (4) add IgG conjugated to an enzyme; and (5) react the substrate with the enzyme to produce a measurable colored product, thus indicating a positive reaction.

One of the most common types of ELISA is the "sandwich ELISA." The sandwich ELISA measures the amount of antigen between two layers of antibodies. The antigens to be measured must contain at least two antigenic sites, capable of binding to antibody, since at least two antibodies act in the sandwich. So sandwich assays are restricted to the quantitation of multivalent antigens such as proteins or polysaccharides. Sandwich ELISAs for quantitation of antigens are especially valuable when the concentration of antigens is low and/or they are contained in high concentrations of contaminating protein. When two "matched pair" antibodies are not available for the target, another option is the competitive ELISA. An advantage to the competitive ELISA is that nonpurified primary antibodies may be used. In order to utilize a competitive ELISA, one reagent must be conjugated to a detection enzyme, such as horseradish peroxidase. The enzyme may be linked to either the immunogen or the primary antibody. There are several different configurations for competitive ELISAs. The example below uses a labeled immunogen as the competitor. Briefly, an unlabeled purified primary antibody is coated onto the wells of a 96-well microtiter plate. This primary antibody is then incubated with unlabeled standards and unknowns. After this reaction is allowed to go to equilibrium, conjugated immunogen is added. This conjugate will bind to the primary antibody wherever its binding sites are not already occupied by unlabeled immunogen. Thus, the more immunogen in the sample or standard, the lower the amount of conjugated immunogen bound. The plate is then developed with substrate and color change is measured.

Ligand-binding assays, including immunoassays/antibody assays, are required to meet the established requirements for chemical assays. However, due to the nature of the ligand-binding assays, certain additional elements such as specificity must be more explicitly established and other elements such as accuracy and precision may meet a lesser standard, or be evaluated on a sample-to-sample basis. The principal elements of a method validation are discussed in ICH Q2A and Q2B and in the U.S. FDA guidance on bioanalytical method validation. The exceptions or additional elements as they apply to validation of ligand-binding assay are described [63]. The criteria for acceptance of method validation, prestudy validation, and in-study validation are described in great detail in this document. In contrast to other analytical methods, LBA are inherently nonlinear. Because four- and five-parameter logistic curves are used to create calibration curves, a large number of calibrators should be used to most accurately describe the curve. The LBA should additionally be validated for cross-reactivity to metabolites and comedications. This evaluation should be performed at high levels and in combination to thoroughly demonstrate selectivity or to determine specific cross-reactivity for each competing analyte.

Greater latitude in precision is allowed for these types of assays. The precision and accuracy should be evaluated during the method validation by analyzing 4 sets of QC samples at LLOQ, low, medium, and high levels in duplicate in 6 different batches. For assays not capable of meeting the nominal acceptance criteria, greater criteria can be set but the precision should be evaluated for each sample analyzed by preparing and analyzing multiple aliquots of each sample. For accuracy in each batch, 4 out of 6 QC samples must be within ±15% of nominal concentration but the two failed QC samples may not be at the same level (4-6-15 rule). When separation (or cleanup) is used for samples but not for calibrators, the recovery of this separation or cleanup step must be determined and used to correct reported sample

concentrations. Possible approaches to assess recovery are the use of a radiolabeled tracer analyte or an internal standard not recognized by the antibody and measured using another technique. Assessment of analyte stability should be performed in whole matrix, not treated, stripped, or prepared matrix. Matrix effects are particularly troublesome in immunoassay methods. The effect of matrix and nonspecific binding must be evaluated and documented in a number of different ways during the method validation: (1) serial dilution of reference analyte with matrix and then the evaluation of response to known concentration; (2) the calibration curve in matrix versus buffer; (3) parallelism between diluted samples and reference analyte; and (4) non specific binding. For the in-study validation, acceptance criteria should follow the 4-6-30 rule (4 out of 6 samples should be within ±30% of nominal concentration).

More recently, recommendations for the development and implementation of bioanalytical method validation for macromolecules in support of pharmacokinetic and toxicokinetic assessments were published in detail [64]. Briefly, values of ±20% (25% at the lower limit of quantification (LLOQ)) are recommended as default acceptance criteria for accuracy (% relative error (RE), mean bias) and interbatch precision (% coefficient of variation (CV)). In addition, as the secondary criterion for method acceptance, it was proposed that the sum of the interbatch precision (%CV) and the absolute value of the mean bias (%RE) be less than or equal to 30%. This added criterion is recommended to help ensure that in-study runs of test samples will meet the proposed run acceptance criteria of 4-6-30. Exceptions to the proposed process and acceptance criteria are appropriate when accompanied by a sound scientific rationale.

Usually, if possible, the LBA should be cross-validated with a chemical method such as LC/MS [65–67]. LBAs are widely used for macromolecular measurements in blood, urine, and other biological liquids [68–72], monitoring of drug–protein interactions [73], and small molecular measurements in different biological fluids [65, 74–76].

### *Isotope Assays*

RADIOISOTOPE ANALYSIS Isotopes of elements contain nuclei with the same number of protons, but different numbers of neutrons. In general, atoms where the number of protons does not equal the number of neutrons tend to be unstable. Some isotopes are stable and some are not. Both stable and unstable isotopes are used in analytical methods. An unstable isotope has an excess of neutrons in its nucleus. To eliminate the surplus, the nucleus decays to a different nucleus by emitting particles and energy. Radioactive nuclides are also known as radioisotopes.

Radioisotopes have proved to be an indispensable tool in biomedical research and have played a pivotal role in the investigation of absorption, distribution, metabolism, and excretion (ADME) properties of new chemical entities over the past several decades [77–79]. The main advantage of using radioisotopes in studying the disposition of new drug candidates is the ease of detection and the achievement of high sensitivity, especially when compounds with high specific activity are used. There are two main types of radiochemical analysis: (1) isotope dilution assay and (2) neutron activation analysis.

*Isotope Dilution Assay (IDA)* The IDA is a method of quantitative analysis based on the measurement of the isotopic abundance of a nuclide after dilution by mixing with one or more of its isotopes [80]. The activity ratio or isotopic ratio of the mixture defines the concentration of the analyte, which is a tremendous advantage for measurement since quantitative separation of the analyte from nonradioactive components is not required.

Reverse dilution analysis involves addition of nonradioactive (cold) drug or metabolite to a sample containing low levels (per weight basis) of radiolabeled drug in order to isolate a radioactive substance. Subsequent purification (usually via crystallization and recrystallizations) to constant specific activity demonstrates the original presence of the compound added.

The principal limitation to IDA is the availability of a suitable spike substance or tracer [81]. The half-life and type of radiation emitted by a radiotracer are very important, as is the purity. The half-life must be long enough so that sufficient activity is available during the analysis for good counting statistics. However, too long half-lives can be a problem because of low specific activities and storage and disposal problems. The IDA can be used as the analytical technique to quantify the absolute abundances of the isotopes excreted in the experiment.

An injection of a radioisotopically labeled drug can be used to determine blood, plasma, or urine levels and their temporal profile by liquid scintillation or gamma counting and measurement of activities in serially derived samples. Liquid scintillation counters can be used to measure total sample radioactivity. In addition, radioactivity of individual components can be determined after chromatographic separation by using an online radiometric chromatographic detector or by collecting chromatographic fractions and using a liquid scintillation detector.

More recently, accelerator mass spectrometry (AMS) has also been introduced into the field of pharmacology and toxicology [82–84]. The sensitivity is due to the fact that this technique measures radioactive atoms directly instead of merely their decay, and thus AMS is approximately one million times more sensitive than liquid scintillation. The main advantage of this technique is its exquisite sensitivity in the attamole range. High sensitivity allows for microdosing studies and human Phase 0 studies.

*Positron Isotopes* Among beta-emitters, at least 95% of all samples used in metabolism and toxicology studies are labeled with either tritium ($^{3}H$) or carbon-14 ($^{14}C$). Perhaps 1–3% of samples contain phosphorus-32 ($^{32}P$) and less than 1% include all other beta-emitters combined. Of these samples, probably 99% contain one isotope by itself with 1% or less being a mixture of two isotopes, almost always $^{3}H + {}^{14}C$. Other commonly used positron isotopes are $^{11}C$, $^{13}N$, $^{15}O$, and $^{18}F$.

When beta-emitters such as $^{3}H$ and $^{14}C$ decay, they release into their surroundings an electron (beta particle), which may have different kinetic energy.

Carbon-14 is a weak β-ray emitter (155 keV). A maximum specific activity of 62.4 mCi per milliatom of carbon or 4460 mCi/g is more than sufficient for most PK studies, where typical doses of carbon-14-labeled drugs are approximately 10–20 μCi. Long physical half-life ($5.7 \times 10^{3}$ yr) is an advantage that allows calculating specific activity without having to consider losses due to spontaneous decay. Biological half-life for a bound compound is 12 days and 40 days for an unbound isotope. Many carbon-14-labeled drugs can be prepared in the analyst's laboratory.

Tritium ($^{3}H$) is a weak β-ray emitter (i.e., maximum = 18.6 keV) with significantly lower energy than $^{14}C$. Tritium-labeled drugs of higher specific activity than those labeled with carbon-14 must be used to achieve equivalent sensitivities in detection. Tritium ($^{3}H$) has a long physical half-life of 12.3 yr and usually a biological half-life of 10–12 days, which depends on the nature of the labeled molecule.

Sulfur-35 is an artificially produced β-emitting isotope. The β particles' energy of $^{35}S$ exceeds that of $^{14}C$. This has prompted the use of $^{35}S$-labeled drugs in pharmaceutical analysis. The short physical half-life (87.9 days) and low number of drugs containing sulfur limits the overall utility of $^{35}S$ in the analysis of drugs and metabolites in biologic fluids.

Phosphorus-32 is another artificially produced β-emitting isotope. The β particles' energy of $^{32}P$ exceeds that of $^{14}C$. Because of a short physical half-life (14.3 days), it requires calculation of loss of radioactivity due to spontaneous decay. Phosphorus occurs in an extremely small number of drugs and therefore is of low importance in biopharmaceutical analysis.

*Gamma Isotopes* The most commonly used gamma isotopes are $^{125}I$, and $^{99m}Tc$. Other commonly used gamma isotopes are $^{111}In$,$^{123}I$, and $^{153}Sm$.

Gamma-emitters, upon their decay, give off discrete packets of energy (gamma rays) in the form of photons rather than particles. Gamma emission is monoenergetic as contrasted to the broad energy distribution of beta-emitters. Thus, for a radioisotope such as iodine-125 ($^{125}I$), all of the gamma emissions have the identical energy of 35 keV. Iodine-125 has been most useful in the development of radiolabeled antibodies for certain radioimmunoassays. It is also used in labeling proteins and hormones. Since a small number of therapeutic agents have iodine atoms as a part of their molecular structure, $^{125}I$ labeling is of little direct benefit in preparing radiolabeled small molecule drugs.

Technetium-99m is used primarily in imaging analysis. A short biological half-life (6 hours) and sufficient physical half-life (1 day) make it ideal for gamma imaging resolution.

Positron emitters are more appropriate for the direct labeling of drug molecules since the gamma-emitting radionuclides are rarely an integral part of the drug molecule structure. The positron (β-emitting) radionuclides are used for *in vitro* and *in vivo* experiments. The gamma-emitting radionuclides are useful for *in vivo* imaging.

The following are the most frequently used radiolabeling techniques: isotope exchange reactions, introduction of a foreign label, labeling with bifunctional chelating agents, biosynthesis, recoil labeling, and excitation labeling. Details and recommendations for radiolabeling can be found in Saha [85].

High chemical and radioactive purity of the labeled drugs is very important for successful ADME and PK studies. The loss of efficacy of a labeled compound over a period of time may result from radiolysis and depends on the physical half-life of the radionuclide, the solvent, any additive, the labeled molecule, the nature of emitted radiations, and the nature of the chemical bond between the radionuclide and the molecule. Radiolysis is decomposition by radiation of labeled compounds emitted by the radionuclides present in them. Radiolysis introduces a number of radiochemical impurities in the sample of labeled material and one should be cautious about these unwanted products. Autoradiolysis occurs when the chemical bond

breaks down by radiation from its own molecule. Indirect radiolysis is caused by the decomposition of the solvent by the radiation, producing free radicals that can break down the chemical bond of the labeled compounds. To help prevent indirect radiolysis, the pH of the solvent should be neutral because more reactions of this nature can occur at alkaline or acidic pH. Higher specific activity results in greater effect of radiolysis. The longer the half-life of the radionuclide, the more extensive the radiolysis will be. Usually a period of three physical half-lives or a maximum of 6 months is suggested for shelf life.

*Detection* Different detection means are employed for beta and gamma measurement but after the initial detection both employ essentially the same electronics to count the number of events detected as well as the same software to process the recorded results.

For beta detection, because the beta particles that are most often used have such low penetrating power, in the process known as "liquid scintillation counting," a discrete sample or a column eluate is usually mixed with special energy transfer solvents and chemicals ("scintillators"), which produce light when a radioactive decay takes place. Alternatively, a column eluate, but never a discrete sample, is allowed to flow over particulate solid ("solid scintillator") matter, which emits light when a radioactive decay occurs on the surface or in close proximity to it.

Because gamma radiation is pure energy and has no mass, it is much more penetrating than beta-particle emission of the same energy. Therefore, the preferred detector for gamma-rays is a cylindrical block of specially activated sodium iodide, one face of which is optically coupled to a photomultiplier with the sides and other face, and the photomultiplier as well, enclosed in a thin aluminum shell. Gamma radiation penetrates the shell, is stopped by the very dense sodium iodide producing light, which is then measured by the photomultiplier. Light production by this means is approximately five times that of a comparable energy beta-particle with liquid scintillator.

Whole-body autoradiography (WBA) is a routine and traditional method of detection on X-ray film. It is successfully used for tissue distribution studies [86, 87]. Whole-body radioluminography is a new method of radiation detection based on a phosphorus imaging technique, offering particular benefits of being much more sensitive and having much wider linear measurement range than traditional X-ray film techniques. This method is successfully and extensively used for tissue distribution studies [78, 88] and is especially attractive in combination with different chromatographic methods on extracts of tissues for studying the metabolic fate of the radiolabel [86].

*Neutron Activation Analysis (NAA)* NAA is one of the most sensitive techniques used for qualitative and quantitative elemental analysis of multiple major, minor, and trace elements in various samples [89, 90]. For many elements, NAA offers sensitivities that are superior to those possible by any other technique. Moreover, the accuracy and precision of the technique are such that NAA is still one of the primary analytical methods used by the National Institute of Standards and Technology (NIST) to certify the concentration of elements in Standard Reference Materials (SRMs) [91]. In typical NAA, stable nuclides ($^{A}Z$, the target nucleus) in the sample undergo neutron capture reactions in a flux of (incident) neutrons. The

radioactive nuclides ($^{A+1}Z$, the compound nucleus) produced in this activation process will, in most cases, decay through the emission of a beta particle (β–) and gamma-ray(s) with a unique half-life. A high-resolution gamma-ray spectrometer is used to detect these "delayed" gamma-rays from the artificially induced radioactivity in the sample for both qualitative and quantitative analysis. One of the principal advantages of NAA is that biological samples in the liquid or solid state can be irradiated and little or no cleanup is necessary. Samples are typically collected, irradiated, and analyzed directly in plastic holders. Care must be taken to avoid contact with metal equipment or other contaminants containing trace levels of metals. NAA is nearly free of any matrix interference effects as the vast majority of samples are completely transparent to both the probe (the neutron) and the analytical signal (the gamma-ray). Moreover, because NAA can most often be applied instrumentally (no need for sample digestion or dissolution), there is little if any opportunity for reagent or laboratory contamination. However, interferences can occur when different elements in the sample produce gamma-rays of nearly the same energy. Usually this problem can be circumvented by choosing alternate gamma-rays for these elements or by waiting for the shorter-lived nuclide to decay prior to counting. Other interferences can occur if another type of nuclear reaction concurrently produces the radionuclide of interest. The accuracy of NAA is generally high, although precision values in the range of ±5–10% (RSDs) have been quoted for this technique.

The instrumentation used to measure gamma-rays from radioactive samples generally consists of a semiconductor detector, associated electronics, and a computer-based, multichannel analyzer (MCA/computer). Most NAA labs operate one or more hyperpure or intrinsic germanium (HPGe) detectors.

Over the last several years there has been a growing interest in the use of *in situ* radiotracers to test new pharmaceuticals and dosage forms being developed for commercial distribution [92–94]. These methodologies offer significant advantages in the evaluation of encapsulations, time release, clearance, and the distribution of the pharmaceutical in animal and human models [95]. Recent work [96] evaluated the sensitivity, accuracy, and precision of NAA to measure serum iohexol (contrast agent) at concentrations necessary for estimating GFR. NAA is also used for elemental analysis of some herbal plants [97].

Readers are referred to *Fundamentals of Nuclear Pharmacy* by G. B. Saha [85] for details on theory and practice of radioisotope methods. This is a comprehensive book describing radiopharmaceuticals currently in use, covering radioactivity, radiation detectors, production of radionuclides, generators, radiopharmaceuticals, and practical recommendations on radiolabeling.

Properties of some radioactive isotopes useful in biology are presented in Table 5.

*Safety* Radioactivity users must follow the institutional Radiation Safety Office instructions, complete special training, and work in a licensed laboratory. The Radiation Safety Office conducts periodical inspections in the user's lab. As with all laboratories, no eating or drinking is allowed in labs used for radioactive work.

When radiation passes through matter, it forms ions by knocking electrons loose from their atoms. This separation of electrons from their atoms is called ionization. Ionization makes X-rays and gamma-rays different from other electromagnetic

**TABLE 5 Properties of Some Radioactive Isotopes Useful in Biology**

| Element | Half-life | Beta Energy[a] | Gamma Energy[a] |
|---|---|---|---|
| Barium ($^{137}$Ba) | 2.6 min | None | 0.662 |
| Calcium ($^{45}$Ca) | 165 days | 0.258 | 0.0125 |
| Carbon ($^{14}$C) | 5730 years | 0.156 | None |
| Cesium ($^{137}$Cs) | 30 years | 0.511, 1.176 | 0.661 |
| Chlorine ($^{36}$C1) | $3.1 \times 10^5$ years | 0.714 | None |
| Cobalt ($^{60}$Co) | 5.26 years | 0.315, 0.663, 1.49 | 1.17, 1.33, 2.16 |
| Hydrogen ($^{3}$H) | 12.26 years | 0.01861 | None |
| Iodine ($^{131}$I) | 8.07 days | Several (600 keV) | Several (364 keV) |
| Iodine ($^{125}$I) | 60 days | None | 35 keV |
| Iron ($^{59}$Fe) | 45 days | 1.573, 0.475, 0.273 | Several (1292 keV) |
| Lead ($^{210}$Pb) | 21 years | 0.015 | 0.0465 |
| Phosphorus ($^{32}$P) | 14.3 days | 1.710 | None |
| Phosphorus ($^{33}$P) | 25 days | 0.249 | None |
| Potassium ($^{42}$K) | 12.4 hours | 3.52, 1.97 | Several |
| Sulfur ($^{35}$S) | 88 days | 0.167 (similar to $^{14}$C) | None |
| Zinc ($^{65}$Zn) | 243.6 days | Positron = 0.325 | 0.34, 0.77, 1.12 |

[a]Maximum energies in million (mega) electron volts. Where the term "several" is listed, individual values can be obtained from the *Handbook of Chemistry and Physics* [188]. Specific activity is the rate of disintegration per unit quantity of the isotope and is frequently expressed as Ci/mole, mCi/mmole, or dpm/mole

radiation such as light rays or radio waves. Ionization changes the nature of biological molecules. As the amount of energy released (as the radioisotope naturally decays) increases, ionization effects increase, as well as the amount of required protection. For instance, $^{3}$H is a very low-energy radioisotope (18 keV) and paper can be used as a shield. $^{32}$P is a relatively low-energy radioisotope (1.71 MeV) and requires Plexiglas shields. $^{125}$I also is a low-energy isotope (35 keV), but because it also emits gamma-rays, lead shields are required for protection.

Radiation can cause somatic and genetic effects in biological organisms. All radioisotopes are hazardous when they get inside the body. Gamma- or high-energy beta-emitters are external hazards.

STABLE ISOTOPE ANALYSIS Stable isotopes are nonradioactive chemical tracers. Stable isotope ratios are measured by mass spectrometric analyses of either bulk phases or specific compounds. Jasper [98, 99] discussed ten major points regarding the properties and uses of stable isotopes in pharmaceuticals, including the suggestion of "isotope product authenticity" of APIs and drug products.

Stable isotopes are also very useful in pharmacokinetic studies. Stable labeled and nonlabeled drug can be simultaneously administered by different routes in order to determine absolute bioavailability. Stable labeled drug can also be administered while the subject is at steady state to determine the steady-state pharmacokinetic profile without requiring drug withdrawal [100]. Stable isotope tagging methods are also being used in proteomic research [101].

The isotope dilution assay using a stable isotope is suitable for pharmacokinetic and pharmacodynamic studies of proteins and polypeptides. This method involves the use of target protein analogues labeled with stable isotopes (such as deuterium

or $^{15}N$). These nonradioactive analogues have different molecular masses and act as internal standards. In a typical experiment, a known amount of the analogue is added to the biological sample to be analyzed prior to any sample handling. After appropriate extraction and purification processes, the final mass spectrometric analysis allows the direct identification of the target compound and its analogue by virtue of their respective molecular masses. In addition, it enables a precise quantification on the basis of the relative intensities of the observed signals (principle of isotope dilution).

The IDA completely avoids the use of radioactive material and is not susceptible to errors arising from immunological cross-reactivity with closely related compounds (e.g., precursors, analogues, maturation or degradation fragments). The measured molecular masses further allow the unambiguous identification of the target compound at a sub-pM level with a clear distinction between its endogenous and injected forms, thus allowing for simultaneous *in vivo* quantitative investigations of both native endogenous and injected compounds [102, 103].

***Separation Methods*** Relative comparison of different aspects of most commonly employed separation techniques is presented in Table 6 and characteristics of each will be discussed next.

*Chromatography* Chromatography is an analytical technique used to separate components of a mixture based on their partitioning between a flowing mobile (gaseous or liquid) phase and a stationary phase. *LC/GC* journal (http://www.lcgcmag.com/lcgc/) is a good practical source of articles about chromatographic issues and applications and periodically distributes a very helpful wallchart guide to LC (or GC) troubleshooting.

Quantification of chromatographically separated peaks is accomplished by measurements of peak areas or peak heights. For well behaved peaks, areas tend to be more precise than heights, but heights tend to be less dependent on flow rate than areas. Band broadening by neighboring peaks and noisy baselines may also affect accuracy of area measurements. Column degradation and method of integration (peak deconvolution algorithm used to identify start and stop integration markers and connection between them) may also affect accuracy.

Internal standard or external standard methods can be used for calibration and quantification. In the internal standard method, a known amount of one or more compounds with similar physicochemical properties to the analyte(s) is added

**TABLE 6 Relative Comparison of Most Commonly Used Separation Methods**

| Factor | TLC | GC | HPLC | CE |
|---|---|---|---|---|
| Resolution | + | ++++ | +++ | ++++ |
| Clean sample requirement | + | ++++ | ++ | ++ |
| Potential sensitivity | + | ++++ | ++++ | +++ |
| Need for derivatization | + | ++++ | + | + |
| Different separation mode options | + | +++ | ++++ | ++++ |
| Available sensitive/selective detectors | + | ++++ | ++++ | ++++ |
| Automation | + | ++++ | ++++ | +++ |
| System cost | + | ++++ | ++++ | +++ |

(spiked) along with known amounts of analyte(s) to calibration samples prior to sample processing. Calibration is based on peak height or area response ratio of analyte to the internal standard. When using mass spectrometric detection, stable labeled (nonradioactive) compounds make ideal standards due to their physicochemical similarity to the analytes. External standard calibration is based on samples only spiked with known amounts of analytes and therefore this approach requires reproducible recovery, dilutions, injections, and constant instrument behavior.

Carry-over is not an uncommon problem in automated chromatographic methods [104]. It can have a significant negative impact on all aspects of method development and validation. Causes for carry-over vary, for example, hardware problems (presence of dead volume in chromatographic plumbing), the nature of the solvent, or analyte solubility. The latter can often be eliminated by appropriate modification of the solvent, its pH, the ionic strength, or the addition of an organic modifier. In some cases, a wash step may be required to minimize any carry-over. While neither ICH nor FDA guidelines specifically address carry-over, chromatographers consider less than 20% of the lowest level of quantitation (LLOQ) as the general acceptance criterion.

Another potential problem during analytical procedures concerns analyte stability, including analyte interconversion [105, 106]. For example, retinoids may undergo stereoisomerization and some statins may undergo interconversion between lactone and open hydroxy carboxylic acid forms. In order to avoid analytical artifacts, appropriate analytical measures and QC samples need to be employed [105, 106].

THIN LAYER CHROMATOGRAPHY (TLC) TLC is a simple, quick, and inexpensive procedure that gives a rapid answer as to how many components are in a mixture. TLC is also used to support the identity of a compound in a mixture when the $R_f$ ($R_f$ = distance of migration of the sample zone/distance of the mobile phase front) of a compound is compared with the $R_f$ of a known compound (preferably both run on the same TLC plate). Pharmaceutical applications of TLC include analysis of starting raw materials, intermediates, pharmaceutical raw materials, formulated products, and drugs and their metabolites in biological media [107, 108].

In conventional TLC, the stationary phase is a powdered adsorbent that is fixed to an aluminum, glass, or plastic plate. The mixture to be analyzed is loaded near the bottom of the plate. The plate is placed in a reservoir of solvent so that only the bottom of the plate is submerged. This solvent is the mobile phase; it moves up the plate by capillary action, causing the components of the mixture to distribute between the adsorbent on the plate and the moving solvent, thus separating the components of the mixture so that the components are separated into separate "spots" appearing from the bottom to the top of the plate. The stationary phase (the adsorbent: silica gel or alumina) is polar, and the polarities of both the component of the mixture and the solvent used as the mobile phase are the determining factors in how fast the compound travels.

When the solvent has reached the top of the plate, the plate is removed from the developing chamber and dried, and the separated components of the mixture are visualized. Automation of locating spots is possible with commercially available spotters and plate readers.

Zone identification is confirmed by offline or online coupling of TLC with visible/ultraviolet (UV), Fourier transform infrared (FTIR), Raman, and mass spectrome-

try (MS). The most common method for detection is the use of a densitometer, which measures diffusely reflected light from the sample spot in the UV-VIS range. Video-densitometric quantitation is also used [109]. Instrumental development methods, such as overpressure layer chromatography (OPLC) or automated multiple development (AMD), can provide separations with increased resolution.

Additional information about the sample is gained by using two-dimensional TLC with two different mobile phases. After the first separation and evaporation of the mobile phase, the plate is turned by 90 degrees and a second TLC run is performed using a different type of mobile phase. A further development is the use of high performance thin layer chromatography (HPTLC). Reduction of layer thickness (down to 100 μm) and particle size (down to 5 μm) of the stationary phase leads to an improved separation within a shorter time. An HPTLC procedure takes approximately 10 min (conventional TLC up to 15–20 min). A major disadvantage is the smaller sample capacity. The very small layer thickness of approximately 10 μm and the absence of any kind of binder in combination with the framework of this stationary phase lead to new and improved properties of this ultrathin layer chromatographic (UTLC) silica-gel plate [110].

TLC elution patterns usually extrapolate to column chromatography elution patterns. Since TLC is a much faster procedure than column chromatography, TLC is used to determine the best solvent system for column chromatography. As you increase the polarity of the solvent system, all mixture components move faster (and vice versa with lowering the polarity). The ideal solvent system is simply the system that separates the components. For instance, in determining the solvent system for a flash chromatography procedure, the ideal system is the one that moves the desired component of the mixture to a TLC $R_f$ of 0.25–0.35 and that will separate this component from its nearest neighbor by the difference in TLC $R_f$ values of at least 0.20. Therefore, a mixture is analyzed by TLC to determine the ideal solvent(s) for a flash chromatography procedure.

Comparative studies have often found that HPTLC is comparative in accuracy and precision but superior to HPLC in terms of total cost and time required for pharmaceutical analyses [107, 108]. Recently, several HPTLC methods were developed and completely validated for different classes of pharmaceuticals in herbal extracts [108, 111, 112], bulk drugs and dosage forms [113–117], and plasma [118, 119].

Some advantages of the offline arrangement (HPTLC) as compared to an online process, such as column high performance liquid chromatography (HPLC), have been outlined [107, 117] and include the following:

1. There is the availability of a great range of stationary phases with unique selectivities for mixture components.
2. Ability to choose solvents for the mobile phase is not restricted by low UV transparency or the need for ultrahigh purity.
3. Repetition of densitometric evaluation can be achieved under different conditions without repeating the chromatography in order to optimize quantification since all sample fractions are stored on the plate.
4. High sample throughput can be achieved since many samples can be chromatographed simultaneously.

5. The cost of solvent purchase and disposal is minimal since the required amount of mobile phase per sample is small.
6. Accuracy and precision of quantification is high because samples and standards are chromatographed and measured under the same conditions on a single TLC plate.
7. Sensitivity limits of analysis are typically at nanogram (ng) to picogram (pg) levels.

Despite these advantages, HPLC is the preferred method of the majority of analytical laboratories for routine drug measurements because of its higher resolution, better automation, and superior quantitative detection. More detailed information on the principles, theory, practice, instrumentation, and applications of TLC and HPTLC can be found in the literature [107, 120]. A lot of practical tips are described by Elke Hahn-Deinstrop [121].

GAS CHROMATOGRAPHY (GC) Gas chromatography separates volatile mixture components by partitioning between the inert gas mobile phase and the liquid stationary phase on a solid support. Older glass or steel packed columns (2–6 m × 2–4 mm i.d.) have for the most part been superseded by thin fused-silica capillary columns (10–30 m × 0.05–0.53 mm i.d.). Capillary columns provide much greater separation efficiency (resolution) but are susceptible to sample overloading (as evidenced by long tailing peaks). Wide-bore capillary columns (0.53 mm i.d.) can be used to minimize overloading problems. Suitable choice of capillary injector (split, splitless, on-column) depends on the nature of the mixture to be analyzed (sample cleanliness, analyte concentration, analyte stability, etc.). Split injection is the most common. Only about 1% or less of the sample goes on the column. The splitless injection technique allows application of the entire sample onto a column and is useful for relatively clean samples containing very small amounts of analytes. On-column injection is made directly onto the column at a low temperature and also allows the sample to condense in a narrow band. It is useful for thermally labile analytes that may decompose in the injection port.

Resolution and run times can be optimized by appropriate column temperature programming. Some commonly observed GC problems include tailing, fronting and ghost peaks, and column bleed. Tailing peaks are usually indicative of undesirable secondary interaction of analyte(s) with the system, for example, active or contaminated surfaces (injector inlet liner, septum, transfer lines), poor column cut, or cold spots. Active or contaminated surfaces need to be pacified and/or cleaned. Column overload usually causes fronting peaks (gradual rise and abrupt peak fall). It can be avoided by decreasing the analyte amount and/or increasing the thickness of the stationary phase, internal diameter of the column, or the column temperature. Ghost peaks commonly arise from a contaminated system, most commonly the injector, and may be resolved by appropriate cleaning. Column bleed is always present to some degree but is increased by excessive column temperature, system (mainly injector and column) contamination, or system leaks allowing column exposure to oxygen. Column conditioning or correction of those problems may be needed.

**TABLE 7 Relative Comparison of Most Commonly Used GC Detectors**

| Factor | FID | NPD | ECD | MS |
|---|---|---|---|---|
| Sensitivity | 100 pg | 10 pg | 100 fg | 10 pg |
| Selectivity | + | ++ | +++ | ++++ |
| Dynamic range | ++++ | ++ | + | +++ |
| Compound sensitivity | General | Nitrogen, phosphorus | Halogens, nitrates | General |
| Structural information | + | + | + | ++++ |

GC techniques generally offer better resolution and sensitivity than other chromatographic techniques, including HPLC. However, requirements for volatile and thermally stable analytes introduce a limitation for GC analysis of some compounds. This drawback can be overcome with prior derivatization to impart volatility and/or stability.

*Detection* The most common detection methods are flame ionization, nitrogen phosphorus, electron capture, and mass spectrometry. Comparison of commonly used GC detectors is presented in Table 7.

1. *Flame Ionization Detector (FID).* FID detects most organic compounds by ionizing the column eluate in a hydrogen-air flame. It is the most commonly used GC detector and exhibits good sensitivity and wide dynamic range.
2. *Nitrogen Phosphorus Detector (NPD).* NPD, also known as a thermionic detector, responds almost exclusively and in a very sensitive manner to compounds containing nitrogen or phosphorus. It differs from FID by having a rubidium or cesium silicate bead, which when heated emits electrons. Nitrogen or phosphorus compounds affect the resultant thermionic emission, generating a signal. The bead is consumable and requires periodic replacement.
3. *Electron Capture Detector (ECD).* ECD utilizes a nickel-63 beta (electron) emitting source, which ionizes a carrier gas. Electronegative compounds (e.g., halogens, nitrates) interact with the generated ions, leading to detection of these compounds. ECD is extremely sensitive to electronegative compounds but has a somewhat limited dynamic range. It requires a general radioactive license and periodic wipe testing.
4. *Mass Spectrometer (MS).* MS will be discussed in the HPLC section below.

HIGH PERFORMANCE LIQUID CHROMATOGRAPHY (HPLC) HPLC separates the dissolved components of a mixture by their equilibration between the liquid mobile phase and a stationary phase. Separation may occur via adsorption, partitioning, ion exchange, or size exclusion. The principles described earlier for SPE sorbent phases are generally applicable to HPLC column packings. Table 8 describes the physical properties of commonly employed HPLC solvents. Commonly used mobile phases are described in Table 9.

Reversed-phase (partition) and normal-phase (adsorption) separations are most commonly employed in HPLC analyses. Silica gel and alumina adsorption chromatography has been mostly superseded by the emergence of bonded (derivatized) silicas. Bonded phases contain a constant amount of liquid phase coated on an inert solid support. This technological improvement was instrumental in resolving prior reproducibility problems. Present normal-phase technology utilizes cyano, amine, or diol moieties bonded to the stationary phase via siloxane bridges. Normal-phase

**TABLE 8 Physical Properties of Solvents Commonly Used in HPLC**

| Solvent | BP (°C) | Viscosity (mPa·s) (20 °C) | Polarity Index ($P'$) | Density (g/mL) | UV Cutoff (nm) |
|---|---|---|---|---|---|
| Acetic acid | 118 | 1.31 | 6.2 | 1.049 | 255 |
| Acetonitrile | 82 | 0.34 | 5.8 | 0.78 | 190 |
| Chloroform | 61 | 0.57 | 4.1 | 1.49 | 245 |
| Ethyl acetate | 77 | 0.46 | 4.3 | 0.90 | 255 |
| *n*-Hexane | 69 | 0.31 | 0.06 | 0.66 | 192 |
| Methanol | 65 | 0.55 | 5.1 | 0.79 | 206 |
| Methyl-*t*-butyl ether | 55 | 0.28 | 2.5 | 0.76 | 210 |
| Methylene chloride | 40 | 0.41 | 3.4 | 1.33 | 233 |
| Potassium phosphate | | | | | 215 |
| 1-Propanol | 97 | 2.26 | 4.0 | 0.80 | 210 |
| 2-Propanol | 82 | 2.86 | 4.3 | 0.79 | 205 |
| Tetrahydrofuran | 66 | 0.55 | 4.2 | 0.88 | 212 |
| Triethylamine | | 0.38 | | | |
| Trifluoroacetic acid | 72 | | | 1.54 | 210 |
| Water | 100 | 1.00 | 10.2 | 1.00 | 180 |

**TABLE 9 Commonly Used HPLC Mobile Phases**

| Compound | Reversed Phase | Normal Phase | Ion Exchange | Size Exclusion |
|---|---|---|---|---|
| Primary solvents (adjust mobile phase strength) | Methanol<br>Acetonitrile<br>Tetrahydrofuran<br>(Propanol) | Methylene chloride<br>Acetonitrile<br>Methanol<br>Methyl-*t*-butyl ether<br>Ethyl acetate | Water | Water |
| Secondary solvents (diluents) | Water | Hexane | Buffers:<br>Phosphates<br>Acetate<br>Citrate<br>Trifluoroacetate | Buffer |
| Modifiers | Buffers:<br>Phosphate<br>Acetate<br>Triethylamine<br>EDTA<br>(Ion pairing:<br>Octyl sulfonate,<br>Tertiary butyl ammonium)<br>TFA | | | |

HPLC employs a nonpolar mobile phase and a polar stationary phase. The mobile phase consists of organic solvents and more lipophilic compounds elute first. Some of the drawbacks of normal-phase HPLC include irreversible adsorption, poor reproducibility (sensitive to water content; deactivation), slow equilibration, and peak tailing.

Reversed-phase column packings employ different length alkyl chains (e.g., $C_1$, $C_2$, $C_4$, $C_8$, $C_{18}$, $C_{30}$), phenyl, or cyano moieties bonded to the stationary phase via siloxane bridges. Reversed-phase HPLC employs a polar mobile phase and a non-polar stationary phase. Elution order, at least in a first approximation, follows hydrophobicity, where polar compounds elute first. The main mode of separation is via partitioning, but there is also the possibility of secondary interactions due to residual silanol groups. These weakly acidic residual silanols can interact with analytes via ion-exchange, hydrogen bonding, or dipole–dipole mechanisms and can cause peak tailing and long retention times. Using end-capped columns (smaller groups bonded to residual silanols) or a basic modifier like triethylamine can minimize unwanted secondary interactions. Modifiers can also be used to minimize undesirable interactions with stationary phase or control analyte ionization. Ion pairing can be used to neutralize ionic or partially ionized species. Alkyl (methane, heptane) sulfonic acids or tetra alkyl (methyl, butyl) ammonium salts are commonly used as ion-pairing reagents (5–10 mM) for bases and acids, respectively. Tetra alkyl ammonium salts can also be used to block residual silanols. Addition of EDTA to the mobile phase can eliminate problems due to minor metal cation impurities.

Advantages of reversed-phase over normal-phase HPLC include the option to use optically transparent solvents (e.g., methanol, acetonitrile), greater compatibility with biological samples and electrochemical detection, relatively rapid equilibration, possibility of gradient elution, and control of secondary equilibria (equilibrium of solute in the mobile phase or stationary phase, i.e., ionization, ion pairing). The most commonly used solvents in reversed-phase HPLC are methanol and acetonitrile. Acetonitrile offers certain advantages over methanol, for example, lower UV cutoff, lower viscosity and thus lower backpressures, lesser tendency for out-gassing (bubble formation), and higher elution strength. On the other hand, buffer salts and ion-pairing agents are more soluble in methanol. The most commonly used buffers are phosphate and acetate. Phosphate buffers have a larger pH buffering range and greater transparency in UV, but lower solubility in organic solvents and lesser volatility. Gradient elution (stepwise or continuous) in reversed-phase HPLC offers a possibility of improved resolution and sensitivity, reduced run time, and better peak shapes. However, gradient chromatography can lead to baseline drifts. Reversed-phase HPLC is a workhorse for biomedical applications.

Ion-exchange chromatography is based on a reversible exchange of ions between stationary and mobile phases. Column packing uses polystyrene resins cross-linked with divinylbenzene, sulfonated to form strong cation exchanger (SCX), or quarternized amine to form strong anion exchanger (SAX). Similar principles apply to ion-exchange HPLC as already discussed for the SPE sorbent phases and will not be further discussed here.

Size exclusion (gel filtration) chromatography uses an aqueous mobile phase for separation of large organic molecules (proteins, polyptides) based on analyte molecular weight. Retention is mostly determined by size of the pores in the stationary phase. Larger molecules elute first because they tend to be excluded from the pores. Columns are selected such that pore size is consistent with the molecular weight composition of the sample.

Some of the more common HPLC problems include high, low, or cycling pressure; leaks; retention time shifting; and peak shape problems. High pressure is commonly

caused by blockage in the system (e.g., inline filter or a frit on the guard column or analytical column). This may be due to coextracted material or buffer salt precipitation. Cleaning or replacement of a filter or frits may resolve these problems. Low pressure may be caused by bubbles in the pump or a leak. Bubbles may be due to improperly degassed mobile phases or leaky check valves, which may need to be cleaned or replaced. Cycling pressure is usually caused by bubbles in a pump head or dirty check valves and may be resolved by degassing the mobile phase and purging the pump or cleaning or replacing a check valve. Retention problems may also be due to leaks, bubbles, and check valve problems. Changes in chemistry (e.g., mobile phase composition or flow) would also cause similar symptoms. Unidirectional, time-related changes in retention times are most likely due to column aging and may require column cleaning with stronger solvents or its replacement. Column aging may also cause peak shape problems and may also require column cleaning or replacement. Peak shape problems may also be due to the mobile phase (e.g., improperly prepared), lack of modifier to minimize undesirable secondary column interactions, or use of a strong injection solvent. Leaks may be due to improperly cut or tightened fittings, excessive system pressure, or worn pump or injector rotor seals.

There are several considerations to be made in selecting a suitable HPLC system, including solubility of analytes, nature of the sample matrix, nature of the functional groups on the analyte(s), required sensitivity and selectivity, and nature of the inherent detectability or ability to impart selective detectability. A good reference book on HPLC method development is by Snyder et al. [122].

*Detection* The most common HPLC detection methods are ultraviolet/visible, fluorescent, mass spectrometric, and infrared. Comparison of the most commonly used HPLC detectors is presented in Table 10. Desirable detector characteristics include high sensitivity, low drift, low noise, low dead volume, fast response, wide dynamic range, reliability, simplicity, and relative insensitivity to operational conditions (temperature, mobile phase composition, flow rate, etc.).

Sensitivity and specificity may be enhanced by appropriate derivatization. Post column derivatization allows separation to be independent of the derivatizing

**TABLE 10 Relative Comparison of Most Commonly Used HPLC Detectors**

| Factor | UV/VIS | FD | EC | RID | MS |
|---|---|---|---|---|---|
| Sensitivity | 1 ng | 1 pg | 5 pg | 100 ng | 5 pg |
| Selectivity | ++ | +++ | +++ | + | ++++ |
| Dynamic range | +++ | ++ | + | + | ++++ |
| Compound sensitivity | General | Aromatics with conjugated pi electrons | Phenols, catechols, nitrosamines | General | General |
| Structural information | +++ | ++ | ++ | + | ++++ |
| Environmental sensitivity | + | +++ | ++++ | ++++ | ++ |

moiety, which could minimize separation and also avoid on-column stability issues (e.g., *o*-phthalaldehyde).

1. *Ultraviolet/Visible (UV/VIS) and Photodiode-Array (PDA) Detector.* UV/VIS detectors are analyte property (as opposed to bulk property) detectors. They can be based on single (filter), variable (monochromator), or multiple wavelengths (multiple diodes). A fixed wavelength detector (discrete source, mostly 254 nm) is generally more sensitive than a variable one.

A variable wavelength detector (continuum source with monochromator), on the other hand, allows selection of a wavelength of maximum sensitivity and/or specificity.

A PDA (rapid scanning) provides the most information as it allows quantification at wavelength(s) of maximum sensitivity and/or selectivity, spectral resolution of chromatographically overlapping components, and collection of absorption spectra for peak identification. It collects the entire UV spectrum (all wavelengths simultaneously) several times during the elution of a chromatographic peak and provides three-dimensional data information: time, wavelength, and absorbance unit response. The PDA is commonly used for peak purity determination (Fig. 3).

2. *Fluorescence Detector (FD).* Fluorescent molecules absorb light energy and emit it at a lower frequency as electrons return to the ground state from the excited singlet state. An FD is several orders of magnitude more sensitive for compounds with appropriate functional groups (e.g., rigid aromatic compounds with conjugated pi electrons) than the UV detector. It also provides enhanced selectivity, as less than 15% of organic compounds fluoresce. Additional selectivity can be achieved with variable excitation and emission (two monochromators). Further fine-tuning can be accomplished by judicious selection of excitation and emission wavelengths. While an FD offers greater selectivity and sensitivity for fluorescent compounds, it is also more sensitive to environmental factors (e.g., temperature, pressure).

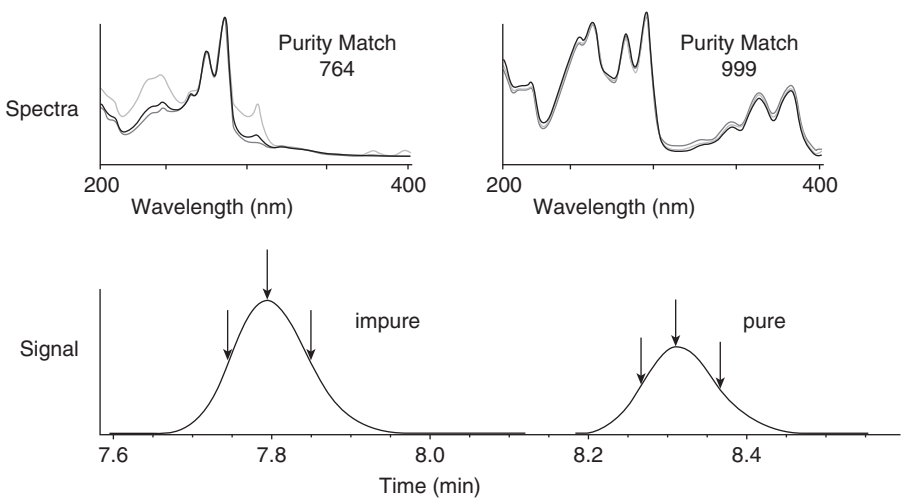


**FIGURE 3** Use of a PDA for peak purity determination. A PDA (top traces) can detect impurities even when they are not evident from the UV chromatographic signal (bottom traces). (Used with permission from C. Huber, LabCompliance.)

3. *Electrochemical (EC) Detector.* Detection with EC is based on oxidation/reduction reactions at a suitable detector. EC provides enhanced sensitivity and selectivity for analytes with appropriate functional groups (e.g., phenols, catecholamines, nitrosamines). For example, EC detection can be employed in *in vitro* drug metabolism studies where a small amount of phenolic metabolite is being formed in the presence of huge excess of substrate—the latter being transparent in this detector [123]. An EC detector can be operated in either the oxidative or reductive mode, but the former is more commonly employed and is not subject to interferences by oxygen. Electrodes require frequent cleaning, depending on the nature of the samples and mobile phase. In the most prevalent mode, amperometric, a fixed potential applied to the electrode (most commonly glassy carbon) oxidizes (or reduces) the analyte with only a small fraction of analyte (~10%) being involved in electron transfer. In coulometric mode, 100% of the analyte is involved in electron transfer and thus is independent of mobile phase flow. While it offers absolute quantitation by Faraday's law, the coulometric mode requires a large-area electrode, which is more susceptible to contamination. Resultant noise offsets 100% analyte conversion without improvement in signal-to-noise (S/N) ratio.

4. *Refractive Index Detector (RID).* While being a universal detector, this bulk property detector has several inherent limitations: it is very temperature sensitive, it requires the same mobile phase composition throughout the run, the baseline is often difficult to stabilize, and it has relatively poor sensitivity.

5. *Radiometric Detector.* A radiometric detector is primarily used for detection of beta-emitters ($^{3}H$, $^{14}C$, $^{32}P$), soft gamma-emitters, or positron-emitters. Radioactive label imparts specificity and allows detection and quantification in biological matrices without concern about nonradioactive interferences. It is especially useful for metabolite detection and pharmacokinetic studies.

6. *Mass Spectrometer Detector.* Mass spectrometric detection adds another dimension to selectivity based on mass/charge ($m/z$) ratio. In addition to quantification, it also allows identification and characterization of analytes and metabolites [124]. Increased sensitivity and selectivity for quantification can be obtained with selected ion monitoring (SIM) in which only a few selected characteristic ions are monitored. Several ionization modes are possible and are selected to best fit the analyte and sample. Electron impact (EI) is the oldest ionization mode and provides extensive fragmentation and yields important and reproducible structural information. Positive chemical ionization (PCI) provides soft ionization (little excess energy in the molecule), causing lesser fragmentation than EI, and yields information on protonated molecular ions or adducts and molecular weight. It can provide additional sensitivity and specificity over EI. Negative chemical ionization (NCI)—electron capture—provides extreme sensitivity for suitable molecules (containing halogens or nitro groups) or molecules that can be appropriately derivatized (e.g., pentafluorobenzyl) to make them amenable to NCI. Atmospheric pressure ionization (API-MS) techniques, electrospray ionization (ESI), and atmospheric pressure chemical ionization (APCI) are popular because they combine solvent elimination and ionization in a single step, but some compounds do not ionize using this approach. ESI is a softer ionization than APCI and can be useful for ionization of polar and less volatile components. APCI, on the other hand, is more robust and less susceptible to ion suppression by coeluting substances. Both

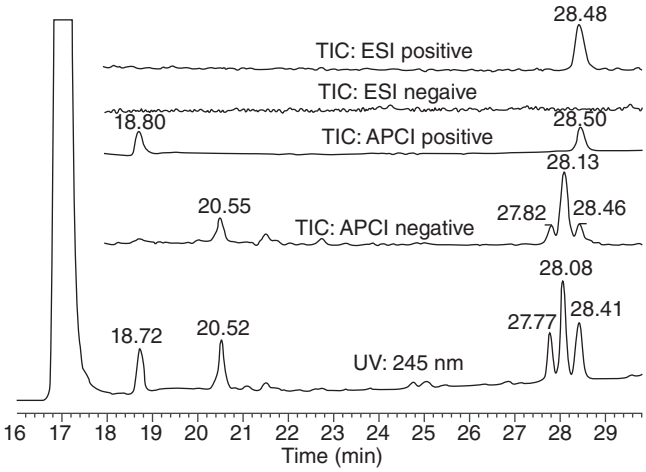


**FIGURE 4** Effect of different modes of API-MS on detection. The bottom UV trace (245 nm) indicates presence of three impurity components eluting between 27 and 29 min. The top four traces represent corresponding total ion current (TIC) response in different modes of API-MS. (Used with permission from the publisher [190].)

ESI and APCI can operate in either positive or negative mode. Sometimes significant improvements in the sensitivity of API-MS methods can be achieved by derivatization (introduction of permanently charged or readily ionized moieties in ESI or high proton or electron affinities in APCI) or mobile phase modification to facilitate adduct formation [20]. It is also important to keep in mind that response in any ionization mode is compound specific as illustrated in Fig. 4. As shown for three unknown compounds eluting between 27 and 29 minutes, response varied markedly using different modes in API-MS. Additional structural information or sensitivity/selectivity is achievable using tandem mass spectrometry or MS/MS techniques (see later discussion).

Several mass spectrometric analyzers are used for drug analyses. The most prevalent is a quadrupole. It has a mass range up to 4000 amu (atomic mass units) and represents a reasonable mixture of mass resolution, linear range, ease of use, and cost. It is also available in a triple quad configuration, TQMS (tandem mass spectrometry), where the middle quadrupole serves as a collision cell to allow for collision-induced dissociation (CID) and operation in MS/MS mode. Parent ions from the first quadrupole undergo further fragmentation in the collision cell to yield daughter ions detectable in the third quadrupole. The process is also known as selected reaction monitoring (SRM) and provides increased selectivity and sensitivity and helps overcome background interferences. MS/MS mode is also useful in peak purity assessment, as shown in Fig. 5. HPLC UV and MS monitoring showed a single chromatographic peak. Two largely abundant ions were coeluting in the MS tracings. Further fragmentation in the MS/MS mode of the larger mass ion did not generate the other coeluting smaller mass ion, suggesting that the latter was not spontaneously generated from the former and that it belongs to another species (i.e., the chromatographic peak was impure).

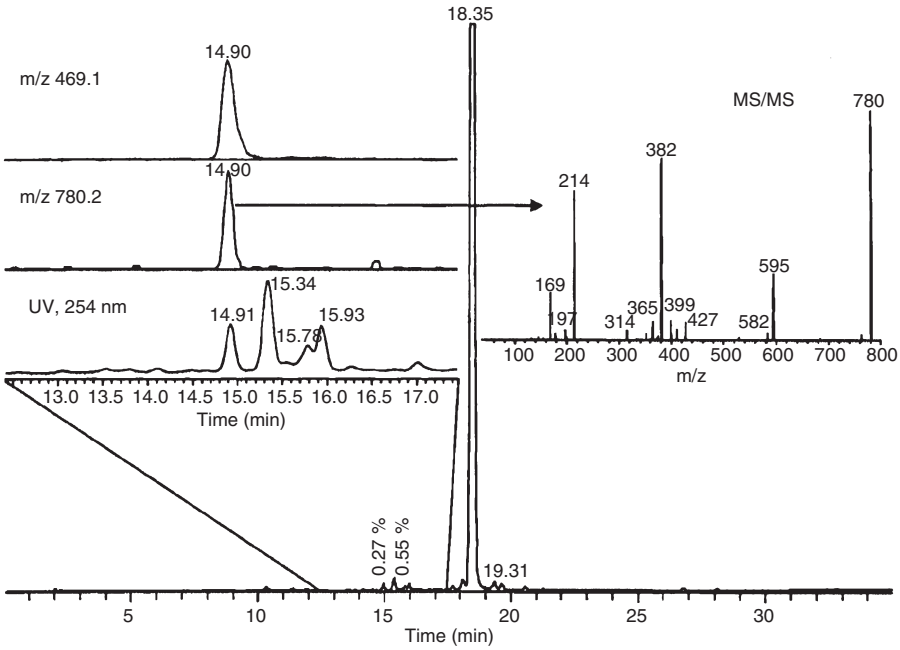


**FIGURE 5** Detection of coeluting impurity using MS/MS. Left panel shows a UV trace and expanded/enlarged UV and MS (two coeluting ions) traces. Right panel depicts MS/MS spectrum of the larger mass ion that lacks the lower mass ion. (Used with permission from the publisher [190].)

An ion trap represents another quadrupole configuration. Advantages of the ion trap include compact size and ability to trap and accumulate ions and thereby increase signal-to-noise ratio. Trapped ions can be further fragmented and ion trap instruments provide an economical alternative to tandem mass spectrometers. Serial trapping and fragmentation theoretically allow $MS^n$. Time-of-flight (TOF) analyzers offer several advantages, including high mass resolution, very high mass range, and high mass accuracy. Enhanced mass resolution with TOF instruments provides improved selectivity. A TOF analyzer [124] can also be used in MS/MS mode by combining it with another TOF, a quadrupole, or an ion trap analyzer. Use of LC/MS in the pharmaceutical arena has been reviewed in the literature [8, 125].

As already discussed, matrix effects (presence of coeluting substances) leading to ionization suppression can cause serious problems in mass spectrometric analyses, especially the EIS [2, 3].

While LC-MS offers sensitivity and selectivity and is presently a very popular technique, it is not suitable for all analytical applications [126] and is not a substitute for expertise in chemistry and chromatography.

*Capillary Electrophoresis (CE)* CE comprises a family of techniques capable of separating a variety of compounds. A recent review [127] describes this separation technique and includes several application examples. The unifying principle in CE is electrical field-driven separation, with migration rate based on analyte size and

charge in the presence of an applied voltage. An electric field is applied across a fused silica capillary with the front end of the capillary immersed in the sample. There are a number of different modes of performing CE separation, including capillary zone electrophoresis (CZE), micellar kinetic electrochromatography (MKEC), and capillary electrochromatography (CEC). CZE is most widely used due to its simplicity and versatility and is applicable to separation of any charged species, including positively and negatively charged species in the same run, but not neutral species. MKEC incorporates pseudopartitioning and thereby also allows separation of neutral species. CEC is a hybrid of CZE and HPLC. CE detectors include UV, fluorescence, electrochemical, and mass spectrometers. In general, some advantages of CE include high resolution, speed, high throughput, small sample size requirement, small reagent consumption, ability to separate most molecules (small and large, even whole organisms), feasible miniaturization, and chiral analysis. It is applicable to both small [128] and large molecules [129]. The major deficiency of CE relative to HPLC has been the higher limit of quantification resulting from small sample volumes and short path length. However, sensitivity improvements (online derivatization, preconcentration, electrokinetic injection, etc.) have been and are being made [128].

Development of various multichannel microfluid devices is under way in order to increase speed and throughput and reduce sample size requirements [130–132] in CE. These include multiplexed CE [131] and microchip electrophoresis [130].

***Chiral Methods*** The common occurrence of chirality (e.g., sugars, amino acids, enzymes) in nature has been known for several centuries, and more recently the stereoselective nature (three-dimensional drug–receptor interaction) of drugs has also become more appreciated and addressed. Recent advances in synthetic [133] and analytical methods (see later discussion) are making development of enantiomeric drugs more technically feasible. The FDA (http://www.fda.gov/cder/guidance/stereo.htm) and other regulatory agencies worldwide are recommending that enantiomers with the best drug characteristics be developed instead of racemates. Enantiomers have identical physical and chemical properties but may behave very differently in an asymmetric milieu. Stereoisomers are frequently distinguished by biological systems and may have different qualitative and quantitative pharmacokinetic and pharmacodynamic (pharmacological and/or toxicological) properties [134]. Therefore, efficacy and safety issues may exist with enantiomers [135]. For example, only the *d*-enantiomer of methylphenidate is active for treatment of the attention deficit hyperactivity disorder and is also significantly more bioavailable. Racemic terodiline has two enantiomers with similar activity for treatment of urinary incontinence. However, cardiotoxicity due to its *R*-isomer led to its withdrawal from the market. In addition, enantiomers may also have a very different metabolic profile and may affect each other's metabolism [136]. There is a growing consensus that new chiral drugs should be developed as single enantiomers (eutomers) and older racemic drugs should be reevaluated and used as pure active enantiomers if supported by clinical data [137]. During drug development, analysis of the formulated racemate should include stereochemical composition. Phase 1 or 2 pharmacokinetic data should address potential interconversion between enantio-

mers. Conventional analytical methods cannot separate enantiomers without introduction of an asymmetric environment. Direct (chiral chromatography or addition of chiral modifier to the mobile phase) and indirect (chromatography following chiral derivatization) analytical approaches can be employed for analyses of chiral compounds and have recently been reviewed [138–141]. Different separation methods (including TLC, GC, CE, and HPLC) have been employed for chiral analysis [141]. Currently, EC and HPLC are most commonly employed for chiral analysis and their application has been the subject of a number of recent reviews [140, 142–149]. Different chiral derivatization approaches have also been used [22, 24, 29, 31]. Miniaturization efforts are under way to improve throughput and speed and reduce sample volume size and include microchip EC [130] and nanotechnology [150].

## 3 ANALYTICAL METHOD VALIDATION

### 3.1 Installation, Operation, Performance Qualification, and Maintenance of Instrumentation

Current regulations do not provide clear and authoritative guidance for validation/qualification of analytical instruments. Therefore, the American Association of Pharmaceutical Scientists, the International Pharmaceutical Federation (FIP), and the International Society for Pharmaceutical Engineering (ISPE) cosponsored a workshop entitled "A Scientific Approach to Analytical Instrument Validation," held in Arlington, Virginia, on 3–5 March 2003. The report describing conclusions of this workshop is published in [151].

The various parties at this workshop agreed that processes are "validated" and instruments are "qualified." Therefore, the phrase "analytical instrument qualification (AIQ)" will be used in lieu of "analytical instrument validation" [151]. Analytical instrument qualification (AIQ) is documented evidence that an instrument performs suitably for its intended purpose and that it is properly maintained and calibrated. AIQ is often broken down into four main phases (Fig. 6):

- Design qualification (DQ) for setting functional and performance specifications (operational specifications).
- Installation qualification (IQ) for performing and documenting the installation in the selected user environment.
- Operational qualification (OQ) for testing the equipment in the selected user environment to ensure that it meets the previously defined functional and performance specifications.
- Performance qualification (PQ) for testing that the system consistently performs as intended for the selected application.

The *design qualification* **(DQ)** activity is most suitably performed by the instrument developer/manufacturer. Since the instrument design is already in place for commercial (off-the-shelf) systems, the user does not need to repeat all aspects of

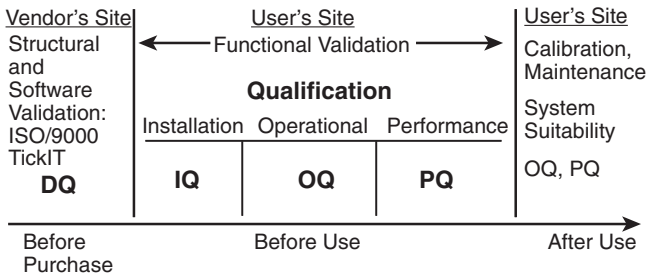


**FIGURE 6** The validation timeline. Activities under each phase are usually performed in each phase indicated in the figure. In some cases, it may be more appropriate to combine a given activity or perform it with another phase. If performed under the other phase, it is not necessary to repeat the activity under the phase where the activity is listed.

DQ. However, users should ensure that instruments have all the necessary functions and performance criteria that will enable them to be successfully implemented for the intended application and that the manufacturer has adopted a quality system for developing, manufacturing, and testing. Users should also establish that manufacturers and vendors adequately support installation, service, and training. As part of the design qualification process, the vendor should be qualified. This may be done by review of established and documented quality systems (e.g., ISO 9001) at the supplier or by a direct audit.

*Installation qualification* **(IQ)** is a documented collection of activities needed to install an instrument in the user's environment. The IQ process can be divided into two steps—*preinstallation* and *physical installation*. During preinstallation, all the information pertinent to the proper installation, operation, and maintenance of the instrument is reviewed. Site requirements and the receipt of all of the parts, pieces, and manuals necessary to perform the installation are confirmed. During physical installation, serial numbers are recorded, and all of the fluidic, electrical, and communication connections are made for components in the system. Documentation describing how the instrument was installed, who performed the installation, and other miscellaneous details should be archived. Only after a successful IQ is the instrument ready for OQ testing.

*Operational qualification* **(OQ)** is the process of demonstrating that an instrument will function according to its operational specification in the user's environment. OQ ensures that the specific modules of the system (or entire system) are operating according to the defined specifications for accuracy, linearity, and precision. This process may be as simple as verifying the module self-diagnostic routines, or may be performed in more depth by running specific tests to verify, for example, detector wavelength accuracy, flow rate, or injector precision. OQ tests can be modular or holistic. Modular testing of individual components of a system may facilitate interchange of such components without requalification and should be done whenever possible. Holistic tests, which involve the entire system, are acceptable in lieu of modular testing [152]. Having successfully completed OQ testing, the instrument is qualified for use in regulated samples testing.

The sample type of tests possible during OQ, which apply to a high performance liquid chromatography (HPLC) unit, can be applied: pump flow rate, gradient linearity, detector wavelength accuracy, detector linearity, column oven temperature, peak area precision, and peak retention time precision. A more intensive list of tests for HPLC is proposed by Grizanti and Zachowski [153]. Bansal et al. [151] consider that OQ tests may not be required to be repeated at a regular interval. Rather, when the instrument undergoes major repairs or modifications, relevant OQ tests should be repeated to verify whether the instrument continues to operate satisfactorily.

*Performance qualification* **(PQ)** is the process of demonstrating that an instrument consistently performs according to a specification appropriate for its routine use. PQ testing is conducted under actual running conditions across the anticipated working range. In practice, however, OQ and PQ frequently blend together, particularly for linearity and precision (repeatability) tests, which can be conducted more easily at the system level [153]. For HPLC, the PQ test should use a method with a well-characterized analyte mixture, column, and mobile phase. It should incorporate the essence of the System Suitability section of the general chromatography chapter ⟨621⟩ in the USP. Again, proper documentation should be archived to support the PQ process. PQ tests are performed routinely, for example, each time the instrument is used. PQ tests should be performed independent of the routine analytical testing performed on the instrument [151]. In practice, PQ can mean system suitability testing, where critical key system performance characteristics are measured and compared with documented, preset limits. For example, a well characterized standard can be injected five or six times and the standard deviation of amounts are then compared with a predefined value. The analysis of quality control (QC) samples with construction of quality control charts has been suggested as another way of performing PQ. Control samples with known amounts are interspersed among actual samples at intervals determined by the total number of samples, the stability of the system, and the specified precision. The advantage of this procedure is that the system performance is measured more or less continuously under conditions that are very close to the actual application [154].

Documented procedures should exist that instruct the operators on what to do if the system does not meet the criteria. When PQ test(s) fail to meet specifications, the instrument requires maintenance or repair. For many instruments a periodic preventive maintenance (PM) may also be recommended. Relevant PQ test(s) should be repeated after the needed maintenance or repair to ensure that the instrument remains qualified.

Suggested PM and PQ procedures and their recommended frequency for several analytical instruments are contained in Appendix A of *Quality Assurance Principles for Analytical Laboratories* [1] and in *Analytical Method Validation and Instrument Performance Verification* [155].

On completion of equipment qualification, documentation should be available that consists of:

- Design qualification document
- Vendor qualification checklist

- Installation qualification document (includes description of hardware and software)
- Procedures for testing
- Qualification test reports with signatures and dates
- Entries on instrument ID in the laboratory's instrument database
- PQ test procedures and representative results

*Software* used for analytical work can be classified into the following categories:

- Firmware
- Instrument control, data acquisition, and processing software
- Stand-alone software

1. *Firmware.* The computerized analytical instruments contain integrated chips with low-level software (firmware). Such instruments will not function without properly operating firmware, and users usually cannot alter the firmware's design or function. Firmware is thus considered a component of the instrument itself. Indeed, qualification of the hardware is not possible without operating it via its firmware. So when the hardware (i.e., analytical instrument) is qualified at the user's site, it essentially qualifies the integrated firmware. No separate on-site qualification of the firmware is needed. Any changes made to firmware versions should be tracked through change control of the instrument.

2. *Instrument Control, Data Acquisition, and Processing Software*. Software for instrument control, data acquisition, and processing for many of today's computerized instruments is loaded on a computer connected to the instrument. Operation of the instrument is then controlled via the software, leaving fewer operating controls on the instrument. Also, the software is needed for data acquisition and post-acquisition calculations. Thus, both hardware and software are critical to providing analytical results.

The manufacturer should perform the DQ, validate this software, and provide users with a summary of validation. At the user's site, holistic qualification, which involves the entire instrument and software system, is more efficient than modular validation of the software alone. Thus, the user qualifies the instrument control, data acquisition, and processing software by qualifying the instrument according to the AIQ process defined earlier.

3. *Stand-Alone Software.* An authoritative guide for validating stand-alone software, such as Laboratory Information Management System (LIMS), is available [156]. The validation process is administered by the software developer, who also specifies the development model appropriate for the software. It takes place in a series of activities planned and executed through various stages of the development cycle. The user-site testing is an essential part of the software development cycle. However, the user-site testing, though essential, is only part of the validation process for stand-alone software and does not constitute complete validation. Refer to this guide for activities needed to be performed at the user's site for testing stand-alone software used in analytical work.

***Instrument Categories*** Based on the level of qualification needed, it is convenient to categorize instruments into three groups—A, B, and C, as defined below [151]. Each group is illustrated by some example instruments. The list of instruments provided as illustration is not meant to be exhaustive, nor can it provide the exact category for an instrument at a user's site. The exact category of an instrument should be determined by the user for the specific instrument or application.

1. *Group A Instruments.* Conformance of Group A instruments to user requirements is determined by visual observation. No independent qualification process is required. Example instruments in this group include light microscopes, magnetic stirrers, mortars and pestles, nitrogen evaporators, ovens, spatulas, and vortex mixers.
2. *Group B Instruments.* Conformance of Group B instruments to user requirements is performed according to the instruments' standard operating procedures. Their conformity assessments are generally unambiguous. Installation of Group B instruments is relatively simple and causes of their failure are readily discernible by simple observations. Example instruments in this group include balances, incubators, infrared spectrometers, melting point apparatus, muffle furnaces, pH meters, pipettes, refractometers, refrigerator-freezers, thermocouples, thermometers, titrators, vacuum ovens, and viscometers.
3. *Group C Instruments.* Conformance of Group C instruments to user requirements is highly method specific, and the conformity bounds are determined by their application. Installing these instruments can be a complicated undertaking and may require the assistance of specialists. A full-qualification process, as outlined in this document, should apply to these instruments. Example instruments in this group include atomic absorption spectrometers, differential scanning calorimeters, densitometers, diode-array detectors, electron microscopes, elemental analyzers, flame absorption spectrometers, gas chromatographs, high-pressure liquid chromatographs, inductively coupled argon plasma emission spectrometers, mass spectrometers, microplate readers, near-infrared spectrometers, Raman spectrometers, thermal gravimetric analyzers, UV/VIS spectrometers, and X-ray fluorescence spectrometers.

### 3.2 Validation Parameters and Definition

Validation is a constant, evolving process starting before an instrument is placed online, and continuing long after method development and transfer. A well defined and documented validation process provides regulatory agencies with evidence that the system and method are suitable for their intended use.

The purpose of method validation [157] is to demonstrate that an analytical method is acceptable for its intended use in terms of the criteria below. Assay qualification differs from assay validation in that its intent is to demonstrate that an accepted method (e.g., compendial method from the USP) is suitable for the intended analysis under actual conditions of use. USP provides regulatory guidance for compendial method validation (USP 25-NF 20). When using USP or NF methods, it is not required to demonstrate their accuracy and reliability. Criteria to be used in

**TABLE 11 ICH Validation Characteristics**

| Procedure (ICH Q2A) Product Specification | ID Test Present/ Absent | Quantitative Limit Test ≤20% | Qualitative Limit Test ≤5% | Content Purity ≥80% | Content Range 40–60% |
|---|---|---|---|---|---|
| Accuracy | No | Yes | No | Yes | Yes |
| Repeatability | No | Yes | No | Yes | Yes |
| Specificity | Yes | Yes | Yes | Yes | Yes |
| Linearity | No | Yes | No | Yes | Yes |
| Range | No | Yes | No | Yes | Yes |
| LOD | No | No | Yes | No | No |
| LOQ | No | Yes | No | No | No |

method validation are described in ICH (International Conference on Harmonisation) guidelines Q2A (http://www.fda.gov/cder/guidance/ichq2a.pdf) [158] and Q2B (http://www.fda.gov/cder/guidance/1320fnl.pdf) [159]. Their brief descriptions are included below and details, including tolerances and acceptance criteria, can be found in those guidelines. In addition, FDA Draft Guidance 2396 (http://www.fda.gov/cder/guidance/2396dft.pdf) [160] provides additional information on analytical procedures and methods validation in CMC (chemistry, manufacturing, and controls) documentation and FDA Guidance 4252 provides information on bioanalytical method validation (http://www.fda.gov/cder/guidance/4252fnl.pdf) [161]. Systematic approaches to the acceptance/rejection decision for chromatographic methods and hypothesis testing have been reviewed [162]. ICH validation characteristics for different tests are summarized in Table 11.

***Specificity*** Specificity (some guidances use the term *selectivity*) is the ability to assess unequivocally the analyte in the presence of components that may be expected to be present (e.g., impurities, degradants, matrix).

***Accuracy and Precision*** The accuracy represents an agreement between measured and theoretical values. Precision refers to variability in the data from replicate determinations. Precision and accuracy are depicted schematically in Fig. 7. Five samples per concentration should be used to validate a bioanalytical method in terms of accuracy and precision. The mean value should be within 15% of the theoretical value with a coefficient of variation (CV) of 15%, except for LLOQ, where these should be within 20%. QC samples should represent low, middle, and high concentrations of experimental values.

***Limit of Detection (LOD) and Limit of Quantification (LOQ)*** LOD refers to the lowest amount of analyte that can be detected (distinguishable from background) but not adequately accurate or precise to be quantified. It is commonly based on 3:1 signal-to-noise ratio (SNR) or as 3 times the standard deviation of the response divided by the slope of the calibration curve.

LOQ refers to the lowest and the highest concentrations that can be quantified with adequate accuracy and precision. It is commonly based on 10:1 SNR or as 10 times the standard deviation of the response divided by the slope of the calibration curve.

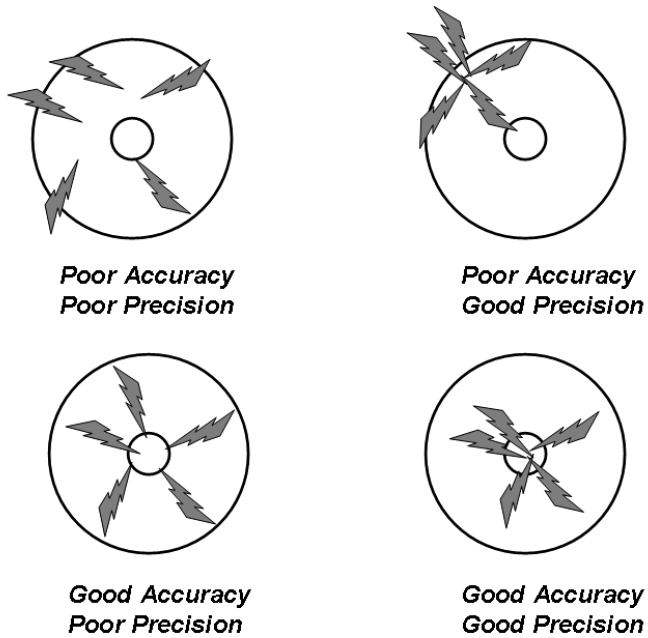


**FIGURE 7** Schematic depiction of accuracy and precision.

***Linearity and Range*** Linearity refers to direct proportionality between measurement values (response) and analyte concentrations. Commonly 5 to 8 concentrations are used to generate the standard curve. Mathematical transformations to allow linearity may be applied if shown scientifically appropriate.

Range refers to concentrations between the low and high limit of quantification and should bracket the concentration in experimental samples.

***Stability*** Potential degradation of samples, standards, controls, and critical reagents during storage (including freeze–thaw cycles when relevant) should be evaluated and attention paid to expiration dates. Stress studies can be performed by exposing analytes to acid, base, heat, UV light, oxidizers, and so on in order to assess their stability.

***Ruggedness*** Ruggedness refers to reproducibility under normal but variable conditions (e.g., different instruments, operators, laboratories, reagent lots, analytical column lots).

***Robustness*** Robustness refers to the analytical method's ability to remain unaffected by small changes in operational parameters (e.g., temperature, mobile phase composition or flow rate, injection volume) and is used to define acceptable tolerances.

***Transferability and Revalidation*** After a method has been validated, it is ready to be transferred to other laboratories that will be using the method. Documentation of the method should include a detailed written procedure, a method validation

**TABLE 12 System Suitability Parameters and Recommendations**[a]

| Parameter | Recommendation |
|---|---|
| Capacity factor ($k'$) | The peak should be well resolved from other peaks and the void volume; generally, $k' > 2.0$ |
| Repeatability | RSD ≤ 1% for $N \geq 5$ is desirable |
| Relative retention | Not essential as long as the resolution is stated. |
| Resolution ($R_s$) | $R^s > 2$ between the peak of interest and the closest eluting potential interferent (impurity, excipient, degradation product, internal standard, etc.). |
| Tailing factor ($T$) | $T \leq 2$ |
| Theoretical plates ($N$) | In general, should be >2000 |

[a]Details of determination of these parameters are presented in Ref. 189.

report, system suitability criteria, and a plan for the method's implementation. General acceptance criteria for the method's intended purpose should also be included. The new user should plan on spending time verifying method performance prior to the method being implemented.

At some time during the lifetime of a method, for one reason or another, it may become necessary to revalidate the method. Revalidation can be carried out in a reactive or proactive manner. Reactive revalidation may be performed in response to changes in incoming raw material, manufacturing batch changes, formulation changes, or any other changes (dilutions, sample preparation) in the method. A total revalidation from scratch is usually not necessary in these instances, but enough revalidation should be performed to address the issues at hand as dictated by the needs and use of the method. Proactive revalidation may be undertaken to take advantage of new technology, or perhaps to automate a previously complex, labor-intensive, and/or time-consuming manual procedure. In such cases, revalidation may be more comprehensive, depending on the undertaking. System suitability tests are based on the concept that the equipment, electronics, analytical operations, and samples constitute an integral system that can be evaluated as a whole.

***System Suitability*** System suitability is determined by checking a system to ensure system performance before or during the analysis of unknowns. For example, for HPLC, parameters such as plate count, tailing factors, resolution, and reproducibility (%RSD retention time and area for six repetitions) are determined and compared against the specifications set for the method. These parameters are measured during the analysis of a system suitability "sample" that is a mixture of main components and expected by-products. Table 12 lists the terms to be measured and their recommended limits obtained from the analysis of the system suitability sample for the HPLC analytical method.

### 3.3 Guidelines: ICH, FDA, AOAC, USP, ISO 9000, and ISO 17025

When the ICH was first established, one of the objectives was to organize an International Conference on Harmonization, hence the name given to the initiative. The name of ICH has now, perhaps, become more associated with the process of harmonization, than the actual conferences, although these have been extremely important

for ensuring that the process of harmonization was carried out in a transparent manner and that there was an open forum in which to present and discuss ICH recommendations. ICH is a joint initiative involving both regulators and industry as equal partners in the scientific and technical discussions of the testing procedures that are required to ensure and assess the safety, quality, and efficacy of medicines.

The focus of the ICH has been on the technical requirements for medicinal products containing new drugs. The vast majority of those new drugs and medicines are developed in Western Europe, Japan, and the United States of America and therefore, when the ICH was established, it was agreed that its scope would be confined to registration in those three regions.

The ICH Topics are divided into four major categories and ICH Topic Codes are assigned according to these categories: *Quality Topics*, that is, those relating to chemical and pharmaceutical Quality Assurance, which are relevant to this chapter, are divided into the following subcategories: Q1 Stability Testing, Q2 Validation, Q3 Impurity Testing, Q4 Pharmacopoeias (not implemented yet), Q5 Quality of Biotechnological Products, and Q9 Risk Management.

***Q1A (R2): Stability Testing of New Drug Substances and Products (Second Revision)*** This guideline [163] has been revised a second time in order to accommodate for the consequences of Q1F and reached Step 4 of the ICH process on 6 February 2003. This guideline provides recommendations on stability testing protocols including temperature, humidity, and trial duration. Furthermore, the revised document takes into account the requirements for stability testing in Climatic Zones III and IV in order to minimize the different storage conditions for submission of a global dossier.

***Q1B: Photostability Testing of New Drug Substances and Products*** The tripartite harmonized ICH guideline [164] was finalized (Step 4) in November 1996. This forms an annex to the main stability guideline and gives guidance on the basic testing protocol required to evaluate the light sensitivity and stability of new drugs and products.

***Q1C: Stability Testing for New Dosage Forms*** The tripartite harmonized ICH guideline [165] was finalized (Step 4) in November 1996. It extends the main stability guideline for new formulations of already approved medicines and defines the circumstances under which reduced stability data can be accepted.

***Q1D: Bracketing and Matrixing Designs for Stability Testing of Drug Substances and Drug Products*** The tripartite harmonized ICH guideline [166] was finalized (Step 4) in February 2002. This document describes general principles for reduced stability testing and provides examples of bracketing and matrixing designs.

***Q1E: Evaluation of Stability Data*** The tripartite harmonized ICH guideline [167] was finalized (Step 4) in February 2003. This document extends the main guideline by explaining possible situations where extrapolation of retest periods/shelf-lives beyond the real-time data may be appropriate. Furthermore, it provides examples of statistical approaches to stability data analysis.

***Q1F: Stability Data Package for Registration Applications in Climatic Zones III and IV*** The tripartite harmonized ICH guideline [168] was finalized (Step 4) in February 2003. This document provides guidance on specific stability testing requirements for Climatic Zones III and IV. Besides proposing acceptable storage conditions for long-term and accelerated studies, it gives guidance on data to cover situations of elevated temperature and/or extremes of humidity. The referenced literature provides information on the classification of countries according to climatic zones.

The most important for this chapter are Q2A and Q2B ICH guidelines already mentioned.

***Q2A: Text on Validation of Analytical Procedures*** The tripartite harmonized ICH text [158] was finalized (Step 4) in October 1994. This identifies the validation parameters needed for a variety of analytical methods. It also discusses the characteristics that must be considered during the validation of the analytical procedures, which are included as part of registration applications.

***Q2B: Validation of Analytical Procedures—Methodology*** The tripartite harmonized ICH text [159] was finalized (Step 4) in November 1996. It extends the guideline Q2A to include the actual experimental data required, along with the statistical interpretation, for the validation of analytical procedures.

***Q3A(R): Impurities in New Drug Substances (Revised Guideline)*** First recommended for adoption at Step 4 of the ICH process on 30 March 1995, the guideline [169] was revised under Step 2 of the ICH process on 7 October 1999 and recommended for adoption under Step 4 on 7 February 2002 by the ICH Steering Committee.

The guideline addresses the chemistry and safety aspects of impurities, including the listing of impurities in specifications, and defines the thresholds for reporting, identification, and qualification. Revision of the guideline has allowed clarification of some inconsistencies, revision of the decision tree, harmonizing with Q3B, and addressing some editorial issues.

***Q3B(R): Impurities in New Drug Products (Revised Guideline)*** This guideline [170] was revised and finalized under Step 4 in February 2003. It complements the guideline on impurities in new drug substances and provides advice in regard to impurities in products containing new, chemically synthesized drug substances. The guideline specifically deals with those impurities that might arise as degradation products of the drug substance or from interactions between drug substance and excipients or components of primary packaging materials. The guideline sets out a rationale for the reporting, identification, and qualification of such impurities based on a scientific appraisal of likely and actual impurities observed, and of the safety implications, following the principles elaborated in the parent guideline. Threshold values for reporting and control of impurities are proposed, based on the maximum daily dose of the drug substance administered in the product.

***Q3C: Impurities—Guideline for Residual Solvents*** The tripartite harmonized ICH guideline [171] was finalized (Step 4) in July 1997. This recommends the use

of less toxic solvents in the manufacture of drug substances and dosage forms, and sets pharmaceutical limits for residual solvents (organic volatile impurities) in drug products.

***Q5C: Quality of Biotechnological Products—Stability Testing of Biotechnological/Biological Products*** The tripartite harmonized ICH guideline [172] was finalized (Step 4) in November 1995. This document augments the stability guideline (Q1A above) and deals with the particular aspects of stability test procedures needed to take account of the special characteristics of products in which the active components are typically proteins and/or polypeptides.

***Q9: Quality Risk Management*** The tripartite harmonized ICH guideline [173] was finalized (Step 4) in November 2005. This guideline provides principles and examples of tools of quality risk management that can be applied to all aspects of pharmaceutical quality including development, manufacturing, distribution, and the inspection and submission/review processes throughout the lifecycle of drug substances and drug (medicinal) products, biological and biotechnological products, including the use of raw materials, solvents, excipients, packaging and labeling materials.

Two FDA guidelines—FDA Draft Guidance 2396, *Analytical Procedures and Methods Validation, Chemistry, Manufacturing and Controls Documentation* [160], and FDA Guidance 4252, *Bioanalytical Method Validation* [161]—are important documents regulating analytical methods validation in addition to ICH guidelines.

*United States Pharmacopoeia 26*, *National Formulary 21*, *Chapter 1225* describes validation of compendial methods [174]. In 2005 the USP proposed revisions to the method validation guidelines published in Chapter 1225 [175]. For the most part, the revisions were made to continue to harmonize with ICH terminology, for example, using the word "procedures" instead of "methods" [176]. The term "pharmaceutical products" is replaced by the term "pharmaceutical articles" to indicate that the guidelines apply to both drug substances and drug products. Another major change is the use of the term "intermediate precision" and the deletion of the section and use of the term "ruggedness." Use of the term ruggedness has been falling out of favor ever since implementation of the original ICH guideline on terminology [158].

Also in 2005, the USP published a proposed new chapter, Chapter 1226, entitled "Verification of Compendial Procedures" [177]. The purpose of this new general information chapter is to provide guidelines for verifying the suitability of a compendial procedure under conditions of actual use. This new chapter summarizes what is necessary to confirm that the compendial procedure works for a particular drug substance, excipients, or dosage form by *verifying* a subset of validation characteristics rather than completing a full validation. The intent is to provide guidance on how to verify that a compendial procedure that is being used for the first time will yield acceptable results utilizing the laboratory's personnel, equipment, and reagents. Verification consists of assessing selected "analytical performance characteristics," described in Chapter 1225, to generate appropriate relevant data as opposed to repeating the entire validation process.

The ISO 9000 family [178–180] was published on 15 December 2000 by the International Organization for Standardization (ISO). These revised standards are

identified by the "2000" in their designation. ISO 9001:2000 [179] specifies requirements for a quality management system for any organization that needs to demonstrate its ability to consistently provide product that meets customer and applicable regulatory requirements and aims to enhance customer satisfaction. ISO 9001:2000 has been organized in a user-friendly format with terms that are easily recognized by all business sectors. The standard is used for certification/registration and contractual purposes by organizations seeking recognition of their quality management system. ISO 9001:2000 specifies requirements for a quality management system where an organization (1) needs to demonstrate its ability to consistently provide product that meets customer and applicable regulatory requirements, and (2) aims to enhance customer satisfaction through the effective application of the system, including processes for continual improvement of the system and the assurance of conformity to customer and applicable regulatory requirements.

All requirements of this international standard are generic and are intended to be applicable to all organizations, regardless of type, size, and product provided. Recently, the ISO has published a new edition of the standard in the ISO 9000 family that defines the vocabulary and describes the fundamentals of quality management systems. ISO 9000:2005, *Quality Management Systems—Fundamentals and Vocabulary* [181], introduces no changes to the descriptions of the fundamentals of quality management systems (QMSS), as presented in the previous edition published in 2000. However, some definitions have been added and explanatory notes expanded or added.

ISO/IEC 90003:2004 [182] provides guidance for organizations in the application of ISO 9001:2000 to the acquisition, supply, development, operation, and maintenance of computer software and related support services. It replaces the old ISO 9000-3 1997 software standard. The American Society for Quality (ASQ) published ANSI/ISO/ASQ Q9000-2000 Series—*Quality Management Standards* [183], which are American National Standards on quality management and quality assurance that are internationally recognized as being identical to the ISO 9000:2000 quality standards.

The ISO/IEC Guide 25, *General Requirements for the Competence of Calibration and Testing Laboratories* [184], has been the internationally recognized basic document for accreditation of laboratories including chemical analysis. The attainment of accreditation is mandatory for some regulatory work areas and frequently is the basis of contracts for analytical work. In 2000 the draft was replaced by ISO/IEC DIS 17025, *General Requirements for the Competence of Testing and Calibration Laboratories* [185]. A new version of ISO/IEC 17025 was issued on 15 May 2005 (ISO/IEC 17025:2005) [186]. The American Association for Laboratory Accreditation (A2LA) provides accreditation and certification in accordance with ISO/IEC 17025:2005.

The ISO 17025 standard states that if testing and calibration laboratories comply with ISO 17025, they also operate in accordance with ISO 9001 or ISO 9002. However, calibration against ISO 9001 and 9002 does not itself demonstrate the competence of the laboratory to produce technically valid data and results. While ISO/IEC Guide 25 was mainly focused on technical controls, the new guide includes more management and administrative controls from the ISO 9001 and ISO 9002 quality standards. ISO/IEC 17025 is the most specific document for quality systems for laboratories. The U.S. FDA is implementing it in their labs and all regulated

laboratories may look at ISO/IEC 17025 as a basis for their operation as many elements will also be subject to FDA audits.

The impact of ISO/IEC 17025 on analytical instrumentation and test methods is similar to the impact GLP and GMP regulations have. Instrument hardware should be calibrated, tested, and verified for performance, and it should be maintained to assure a continuous performance. The performance of hardware should be periodically reverified according to a well documented schedule. Software should be validated by the user or vendor firm and verified by the end-user for proper performance. Analytical methods must be validated if nonstandard methods are used.

Specific to some extent to ISO 9000 and ISO/IEC 17025 is the requirement to trace calibration devices such as reference standards or reference test devices back to national or international standards. All tools used for calibration and performance verification of instruments should be well maintained and calibrated. Training records should be kept for all individuals who perform the calibration and verification.

### 3.4 Good Laboratory Practice (GLP)

*Analytical laboratories should follow GLPs for all preclinical drug development studies.*

The basic GLP requirements for analytical laboratories include the following:

- Established, controlled, *standard operating procedures* (SOPs) for sample handling, instrument operation, test methods, data handling, and analyst training.
- Use of *labeled, traceable reagents* properly stored and used within expiration dates.
- Use of *traceable reference standards* as testing controls.
- Maintenance of *controlled, complete laboratory notebooks*, including raw data archiving and retrieval systems.
- Having *calibrated, regularly maintained equipment* for every step in sample storage, handling, and testing.
- *Documentation* of all activities performed in operations related to GLP samples, including personnel qualifications and training records.

Attention to five steps of data acquisition for product analysis can help to focus on the goals of compliance activities: analyses should be planned, performed, monitored, recorded, and reported. Viewing these as distinct activities can greatly enhance a facility's ability to follow GLPs.

1. *Planning.* Designate a study protocol and a study director. The protocol should include descriptions of the nature and purpose of the study, sample information, dates, test methods used, justifications, and references. Any modifications to the protocol must be signed and dated. The study director is the party responsible for adhering to compliance activities. Contract laboratories should designate a study director to coordinate efforts with the client's in-house preclinical study director.

2. *Performing.* Facilities where the analytical work is done should be of suitable size, design, and construction for analyses. SOPs should describe all aspects of sample handling (such as receipt, identification, labeling, and storage) and instrument use (such as maintenance, calibration, and monitoring—including refrigerators, freezers, water baths, incubators, etc.). Equipment should be of adequate capacity and suitably located. Personnel should be experienced in the methods they perform and have documented training.

3. *Monitoring.* Analysis systems need to be examined by a quality assurance unit, who will review database information, data analysis, and the final report. A statement should be attached to the final report documenting review by this unit.

4. *Recording.* This includes documentation, data generation and analysis, and data storage and retrieval. It includes everything from the study protocol, procedures, final reports, and audits, to personnel qualifications, equipment records, and sample records.

5. *Reporting.* A final study report should be delivered to the client, describing the nature and purpose of the study, the protocols used, the samples tested, the results, references to raw data (including computer files), and any applicable literature references. All data should be signed and dated, with corrections in the protocol or report included as amendments.

### 3.5 Method Validation Protocol

A validation plan is a written plan stating how validation will be conducted, including test parameters, product characteristics, production equipment, and decision points on what constitutes acceptable results. USP Chapter 1225 [174] on validation of analytical methods specifically addresses terms and definitions, but leaves protocol and methodology open for interpretation. The ICH guidelines on method validation methodology, ICH Q2A and ICH Q2B [158, 159], and the U.S. FDA *Guidance for Industry: Bioanalytical Method Validation* [161] fill this gap. Full or partial validation may be required. Full validation (usually on three different days) is performed when developing and implementing a method for the first time, when a new drug entity is used, and when metabolites are added to an existing assay for quantification. Partial validation (may be performed in 1–2 days) is used for modification of already validated methods. Partial validation can range from as little as one intra-assay accuracy and precision determination to nearly full validation. Before outlining an experimental design or protocol, however, it is necessary to make some basic assumptions. These assumptions include:

1. Selectivity has been previously demonstrated or is measured and documented during the course of the validation protocol.
2. The method has been developed and optimized to the point where it makes sense investing time and effort in validating the method. Indeed, robustness should be the first parameter investigated.
3. Once data are generated, statistically valid approaches are used to evaluate it and make decisions, thus removing some of the subjectivity of method validation.

Given that the above three steps are addressed in one way or another, the following stepwise protocol can be proposed for method validation (each specific protocol depends on the nature of the method).

Step 1. On day 1, a linearity test (except for LBAs) over five levels for both the drug substance and dosage form is performed. Comparison of the results between the drug substance and dosage form fulfills the accuracy requirement.

Step 2. At the end of day 1, six repetitions are performed at 100% of the drug substance for repeatability.

Step 3. Steps 1 and 2 are repeated over two additional days for intermediate precision.

Step 4. LLOQ is evaluated (as needed) by analyzing the drug substance over five levels, plus six repetitions for precision.

Step 5. Baseline noise is evaluated over six repetitions of blank injections for the determination of LLOD (if required).

In this manner, a logical stepwise approach to method validation using ICH methodology can be performed. As stated previously, the use of the appropriate statistical tests (Students-*t*, Cochran, Dixon, and Fisher tests) allows for less subjective decisions to be made regarding the data, reducing method validation to a much more objective undertaking. For example, the Grubbs Procedure described in Ref. [187] may be used to exclude outliers. A rigorous use of statistics may in turn lend itself to some degree of automation in the future.

## 4 CONCLUSION

Analytical technologies continue to evolve, leading to further improvements in sensitivity and selectivity. There are many analytical tools (instrumentations, techniques, computations) presently available to analysts. Selection of tools to be used should be based on the nature of the analyte(s), analytical goals (sensitivity, selectivity, etc.), sample matrix, and sample stability. Tools should not be used blindly but with an analyst's understanding of theory, techniques, instrumentation, and their inherent limitations. While high throughput is desirable, analyses should not become only a matter of inputting samples into analytical machines and merely accepting numerical machine output. Analytical methods need to be subject to appropriate method validation (based on requirements of the most recent guidelines), including periodic instrument testing and calibration, and should incorporate appropriate quality control measures. In addition, stability of analyte(s) needs to be considered from the time of sample collection through final analytical measurement.

## REFERENCES

1. Garfield FM, Klesta E, Hirsch J, Eds. *Quality Assurance Principles for Analytical Laboratories*, 3rd ed. Gaithersburg: AOAC International; 2000.

2. Taylor PJ. Matrix effects: the Achilles heel of quantitative high-performance liquid chromatography-electrospray-tandem mass spectrometry. *Clin Biochem* 2005;38(4): 328–334.
3. Jemal M, Schuster A, Whigan DB. Liquid chromatography/tandem mass spectrometry methods for quantitation of mevalonic acid in human plasma and urine: method validation, demonstration of using a surrogate analyte, and demonstration of unacceptable matrix effect in spite of use of a stable isotope analog internal standard. *Rapid Commun Mass Spectrom* 2003;17(15):1723–1734.
4. Roth W, Beschke K, Jauch R, Zimmer A, Koss FW. Fully automated high-performance liquid chromatography. A new chromatograph for pharmacokinetic drug monitoring by direct injection of body fluids. *J Chromatogr* 1981;222(1):13–22.
5. Hagestam IH, Pinkerton TC. Internal surface reversed-phase silica supports for liquid chromatography. *Anal Chem* 1985;57:1757–1763.
6. Haque A, Stewart JT. Direct injection HPLC analysis of some non-steroidal anti-inflammatory drugs on restricted access media columns. *Biomed Chromatogr* 1999;13(1): 51–56.
7. Rbeida O, Chiap P, Lubda D, Boos KS, Crommen J, Hubert P. Development and validation of a fully automated LC method for the determination of cloxacillin in human plasma using anion exchange restricted access material for sample clean-up. *J Pharm Biomed Anal* 2005;36(5):961–968.
8. Hopfgartner G, Bourgogne E. Quantitative high-throughput analysis of drugs in biological matrices by mass spectrometry. *Mass Spectrom Rev* 2003;22(3):195–214.
9. Veuthey JL, Souverain S, Rudaz S. Column-switching procedures for the fast analysis of drugs in biologic samples. *Ther Drug Monit* 2004;26(2):161–166.
10. Heinig K, Bucheli F. Application of column-switching liquid chromatography–tandem mass spectrometry for the determination of pharmaceutical compounds in tissue samples. *J Chromatogr B Analyt Technol Biomed Life Sci* 2002;769(1):9–26.
11. Hogendoorn EA, van Zoonen P, Polettini A, Marrubini Bouland G, Montagna M. The potential of restricted access media columns as applied in coupled-column LC/LC-TSP/MS/MS for the high-speed determination of target compounds in serum. Application to the direct trace analysis of salbutamol and clenbuterol. *Anal Chem* 1998;70(7): 1362–1368.
12. Kuklenyik Z, Ye X, Reich JA, Needham LL, Calafat AM. Automated online and off-line solid-phase extraction methods for measuring isoflavones and lignans in urine. *J Chromatogr Sci* 2004;42(9):495–500.
13. Xia YQ, Whigan DB, Powell ML, Jemal M. Ternary-column system for high-throughput direct-injection bioanalysis by liquid chromatography/tandem mass spectrometry. *Rapid Commun Mass Spectrom* 2000;14(2):105–111.
14. Adamovics JA, Ed. *Chromatographic Analysis of Pharmaceuticals*, 2nd ed. New York: Marcel Dekker; 1997.
15. Bosanquet AG. Stability of solutions of antineoplastic agents during preparation and storage. *Cancer Chemother Pharmacol* 1986;17:1.
16. Pitt AM. The nonspecific protein binding of polymeric microporous membranes. *J Parenteral Sci Technol* 1987;41(2):110.
17. Catz P, Shinn W, Kapetanovic IM, Kim H, Kim M, Jacobson EL. Simultaneous determination of myristyl nicotinate, nicotinic acid, and nicotinamide in rabbit plasma by liquid chromatography–tandem mass spectrometry using methyl ethyl ketone as a deproteinization solvent. *J Chromatogr B* 2005;829(1–2):123–135.

18. Jemal M, Teitz D, Ouyang Z, Khan S. Comparison of plasma sample purification by manual liquid–liquid extraction, automated 96-well liquid–liquid extraction and automated 96-well solid-phase extraction for analysis by high-performance liquid chromatography with tandem mass spectrometry. *J Chromatogr B Biomed Sci Appl* 1999; 732(2):501–508.
19. Vas G, Vekey K. Solid-phase microextraction: a powerful sample preparation tool prior to mass spectrometric analysis. *J Mass Spectrom* 2004;39(3):233–254.
20. Gao S, Zhang ZP, Karnes HT. Sensitivity enhancement in liquid chromatography/atmospheric pressure ionization mass spectrometry using derivatization and mobile phase additives. *J Chromatogr B Analyt Technol Biomed Life Sci* 2005;825(2):98–110.
21. Ahuja S. Derivatization in gas chromatography. *J Pharm Sci* 1976;65(2):163–182.
22. Schurig V. Practice and theory of enantioselective complexation gas chromatography. *J Chromatogr A* 2002;965(1–2):315–356.
23. Halket JM, Zaikin VV. Derivatization in mass spectrometry-3. Alkylation (arylation). *Eur J Mass Spectrom (Chichester)* 2004;10(1):1–19.
24. Srinivas NR, Shyu WC, Barbhaiya RH. Gas chromatographic determination of enantiomers as diastereomers following pre-column derivatization and applications to pharmacokinetic studies: a review. *Biomed Chromatogr* 1995;9(1):1–9.
25. Danielson ND, Targove MA, Miller BE. Pre- and postcolumn derivatization chemistry in conjunction with HPLC for pharmaceutical analysis. *J Chromatogr Sci* 1988;26(8): 362–371.
26. Fukushima T, Usui N, Santa T, Imai K. Recent progress in derivatization methods for LC and CE analysis. *J Pharm Biomed Anal* 2003;30(6):1655–1687.
27. Krull IS, Deyl Z, Lingeman H. General strategies and selection of derivatization reactions for liquid chromatography and capillary electrophoresis. *J Chromatogr B Biomed Appl* 1994;659(1–2):1–17.
28. Li F, Zhang C, Guo X, Feng W. Chemiluminescence detection in HPLC and CE for pharmaceutical and biomedical analysis. *Biomed Chromatogr* 2002;17(2–3):96–105.
29. Toyo'oka T. Resolution of chiral drugs by liquid chromatography based upon diastereomer formation with chiral derivatization reagents. *J Biochem Biophys Methods* 2002; 54(1–3):25–56.
30. Ohkura Y, Kai M, Nohta H. Fluorogenic reactions for biomedical chromatography. *J Chromatogr B Biomed Appl* 1994;659(1–2):85–107.
31. Sun XX, Sun LZ, Aboul-Enein HY. Chiral derivatization reagents for drug enantioseparation by high-performance liquid chromatography based upon pre-column derivatization and formation of diastereomers: enantioselectivity and related structure. *Biomed Chromatogr* 2001;15(2):116–132.
32. Imai K. Derivatization in liquid chromatography. *Adv Chromatogr* 1987;27:215–245.
33. Bardelmeijer HA, Lingeman H, de Ruiter C, Underberg WJ. Derivatization in capillary electrophoresis. *J Chromatogr A* 1998;807(1):3–26.
34. Bardelmeijer HA, Waterval JC, Lingeman H, van't Hof R, Bult A, Underberg WJ. Pre-, on- and post-column derivatization in capillary electrophoresis. *Electrophoresis* 1997;18(12–13):2214–2227.
35. Rosenfeld JM. Solid-phase analytical derivatization: enhancement of sensitivity and selectivity of analysis. *J Chromatogr A* 1999;843(1–2):19–27.
36. Little JL. Artifacts in trimethylsilyl derivatization reactions and ways to avoid them. *J Chromatogr A* 1999;844(1–2):1–22.

37. Salgado HR, Oliveira CL. Development and validation of an UV spectrophotometric method for determination of gatifloxacin in tablets. *Pharmazie* 2005;60(4):263–264.
38. Lastra OC, Lemus IG, Sanchez HJ, Perez RF. Development and validation of an UV derivative spectrophotometric determination of losartan potassium in tablets. *J Pharm Biomed Anal* 2003;33(2):175–180.
39. Forbes RA, Persinger ML, Smith DR. Development and validation of analytical methodology for near-infrared conformance testing of pharmaceutical intermediates. *J Pharm Biomed Anal* 1996;15(3):315–327.
40. Fountain W, Dumstorf K, Lowell AE, Lodder RA, Mumper RJ. Near-infrared spectroscopy for the determination of testosterone in thin-film composites. *J Pharm Biomed Anal* 2003;33(2):181–189.
41. Kumana CR, Au WY, Lee NS, Kou M, Mak RW, Lam CW, Kwong YL. Systemic availability of arsenic from oral arsenic-trioxide used to treat patients with hematological malignancies. Eur *J Clin Pharmacol* 2002;58(8):521–526.
42. Jacobs SS, Fox E, Dennie C, Morgan LB, McCully CL, Balis FM. Plasma and cerebrospinal fluid pharmacokinetics of intravenous oxaliplatin, cisplatin, and carboplatin in nonhuman primates. *Clin Cancer Res* 2005;11(4):1669–1674.
43. Goodisman J, Hagrman D, Tacka KA, Souid AK. Analysis of cytotoxicities of platinum compounds. *Cancer Chemother Pharmacol* 2005;19:1–11.
44. Lansdown AB, Williams A, Chandler S, Benfield S. Silver absorption and antibacterial efficacy of silver dressings. *J Wound Care* 2005;14(4):155–160.
45. Al-Saleh I, Shinwari N, El-Doush I, Billedo G, Al-Amodi M, Khogali F. Comparison of mercury levels in various tissues of albino and pigmented mice treated with two different brands of mercury skin-lightening creams. *Biometals* 2004;17(2):167–175.
46. Gorter RW, Butorac M, Cobian EP. Examination of the cutaneous absorption of copper after the use of copper-containing ointments. *Am J Ther* 2004;11(6):453–458.
47. Brun PH, DeGroot JL, Dickson EF, Farahani M, Pottier RH. Determination of the *in vivo* pharmacokinetics of palladium-bacteriopheophorbide (WST09) in EMT6 tumour-bearing Balb/c mice using graphite furnace atomic absorption spectroscopy. *Photochem Photobiol Sci* 2004;3(11–12):1006–1010.
48. Brouwers EE, Tibben MM, Joerger M, van Tellingen O, Rosing H, Schellens JH. Determination of oxaliplatin in human plasma and plasma ultrafiltrate by graphite-furnace atomic-absorption spectrometry. *Anal Bioanal Chem* 2005;382(7):1484–1490.
49. Liu Y. Determination of ten trace elements in Chinese traditional medicines by atomic absorption spectrometry. *Guang Pu Xue Yu Guang Pu Fen Xi* 2000;20(3):373–375.
50. Gomez MR, Soledad C, Olsina RA, Silva MF, Martinez LD. Metal content monitoring in *Hypericum perforatum* pharmaceutical derivatives by atomic absorption and emission spectrometry. *J Pharm Biomed Anal* 2004;34(3):569–576.
51. Arce S, Cerutti S, Olsina R, Gomez MR, Martinez LD. Determination of metal content in valerian root phytopharmaceutical derivatives by atomic spectrometry. *J AOAC Int* 2005;88(1):221–225.
52. Niemela M, Kola H, Eilola K, Peramaki P. Development of analytical methods for the determination of sub-ppm concentrations of palladium and iron in methotrexate. *J Pharm Biomed Anal* 2004;35(3):433–439.
53. Abdel-Ghani NT, Youssef AF, Awady MA. Cinchocaine hydrochloride determination by atomic spectrometry and spectrophotometry. *Farmaco* 2005;60(5):419–424.
54. Khalil S, Borham N. Indirect atomic absorption spectrometric determination of pindolol, propanolol and levamisole hydrochlorides based on formation of ion-associates with

ammonium reineckate and sodium cobaltinitrite. *J Pharm Biomed Anal* 2000;22(2): 235–240.

55. Avad MM, Shalaby AA, Abdellatef HE, Hosny MM. Spectrophotometric and AAS determination of ramipril and enalapril through ternary complex formation. *J Pharm Biomed Anal* 2002;15(2):311–321.
56. Salem H, Askal H. Colourimetric and AAS determination of cephalosporins using Reineck's salt. *J Pharm Biomed Anal* 2002;29(1–2):347–354.
57. Dědina J, Tsalev DL. *Hydride Generation Atomic Absorption Spectrometry*. Hoboken, NJ: Wiley; 1995.
58. Jackson KW, Ed. *Electrothermal Atomization for Analytical Atomic Spectroscopy*. Hoboken, NJ: Wiley; 1999.
59. Butcher DJ, Sneddon J. *A Practical Guide to Graphite Furnace Atomic Absorption Spectrometry*. Hoboken, NJ: Wiley; 1998.
60. Haswell SJ, Ed. *Atomic Absorption Spectrometry: Theory Design and Applications*. St Louis, MO: Elsevier; 1991.
61. Ebdon L, Evans EH, Fisher AS, Hill SJ. *An Introduction to Analytical Atomic Spectrometry*. Hoboken, NJ: Wiley; 1998.
62. Welz B, Sperling M. *Atomic Absorption Spectrometry*, 3rd ed. Hoboken, NJ: Wiley; 1999.
63. DeSilva B, Smith W, Weiner R, Kelley M, Smolec J, Lee B. Recommendations for the bioanalytical method validation of ligand-binding assays to support pharmacokinetic assessments of macromolecules. *Pharm Res* 2003;20(11):1885–1900.
64. Smolec JM, DeSilva B, Smith W, Weiner R, Kelly M, Lee B, Khan M, Tacey R, Hill H, Celniker A, Shah V, Bowsher R, Mire-Sluis A, Findlay, JWA, Saltarelli M, Quarmby V, Lansky D, Dillard R, Ullmann M, Keller S, Karnes HT. Bioanalytical method validation for macromolecules in support of pharmacokinetic studies. *Pharm Research* 2005;22(9): 1425–1431.
65. Lyubimov AV, Garry VF, Carlson RE, Barr DB, Baker SE. Simplified urinary immunoassay for 2,4-D: validation and exposure assessment. *J Lab Clin Med* 2000;136:116–124.
66. Laloup M, Tilman G, Maes V, De Boeck G, Wallemacq P, Ramaekers J, Samyn N. Validation of an ELISA-based screening assay for the detection of amphetamine, MDMA and MDA in blood and oral fluid. *Forensic Sci Int* 2005;153(1):29–37.
67. Allen KR, Azad R, Field HP, Blake DK. Replacement of immunoassay by LC tandem mass spectrometry for the routine measurement of drugs of abuse in oral fluid. *Ann Clin Biochem* 2005;42(Pt 4):277–284.
68. Rothen-Weinhold A, Besseghir K, De Zelicourt Y, Gurny R. Development and evaluation *in vivo* of a long-term delivery system for vapreotide, a somatostatin analogue. *J Control Release* 1998;52(1–2):205–213.
69. Brown-Augsburger P, Yue XM, Lockridge JA, McSwiggen JA, Kamboj D, Hillgren KM. Development and validation of a sensitive, specific, and rapid hybridization–ELISA assay for determination of concentrations of a ribozyme in biological matrices. *J Pharm Biomed Anal* 2004;34(1):129–139.
70. Xie H, Audette C, Hoffee M, Lambert JM, Blattler WA. Pharmacokinetics and biodistribution of the antitumor immunoconjugate, cantuzumab mertansine (huC242-DM1), and its two components in mice. *Pharmacol Exp Ther* 2004;308(3):1073–1082.
71. Zhang Q, Wang GJ, Sun JG. Pharmacokinetics of recombinant human basic fibroblast growth factor in mice using a radioiodination method combined with SDS-PAGE and a sandwich enzyme-linked immunosorbent assay. *Eur J Drug Metab Pharmacokinet* 2004;29(3):163–168.

72. Ito Y, Tosh B, Togashi Y, Amagase K, Kishida T, Kishida T, Sugioka N, Shibata N, Takada K. Absorption of interferon alpha from patches in rats. *J Drug Target* 2005;13(6); 383–390.

73. Yang XX, Hu ZP, Chan SY, Zhou SF. Monitoring drug–protein interaction. *Clin Chim Acta* 2006;365(1–2):9–29.

74. Mealey KL, Peck KE, Bennett BS, Sellon RK, Swinney GR, Melzer K, Gokhale SA, Krone TM. Systemic absorption of amitriptyline and buspirone after oral and transdermal administration to healthy cats. *J Vet Intern Med* 2004;18(1):43–46.

75. Saita T, Fujito H, Mori M. A specific and sensitive assay for gefitinib using the enzyme-linked immunosorbent assay in human serum. *Biol Pharm Bull* 2005;28(10):1833–1873.

76. Higashi Y, Ikeda Y, Yamamoto R, Yamashiro M, Fujii Y. Pharmacokinetic interaction with digoxin and glucocorticoids in rats detected by radio-immunoassay using a novel specific antiserum. *Life Sci* 2005;77(9):1055–1067.

77. Dalvie D. Recent advances in the applications of radioisotopes in drug metabolism, toxicology and pharmacokinetics. *Curr Pharm Des* 2002;6(10):1009–1028.

78. McCarthy KE. Recent advances in the design and synthesis of carbon-14 labelled pharmaceuticals from small molecule precursors. *Curr Pharm Des* 2000;6:1057–1083.

79. Mao J, Xu Y, Wu D, Almassain B. Pharmacokinetics, mass balance and tissue distribution of a novel DNA alkylating agent, VNP40101M, in rats. *AAPS PharmSci* 2002;4(4) *article 24*:1–7.

80. Van Grieken R, de Bruin M. Nomenclature for Radioanalytical Chemistry: IUPAC Recommendation 1994. *Pure & Appl Chem* 1994;66:2513.

81. Fassett JD. Isotopic and nuclear analytical techniques in biological systems: a critical study. Part X. Elemental isotope dilution analysis with radioactive and stable isotopes. *Pure & Appl Chem* 1995;67(11):1943–1949.

82. White INH, Brown K. Techniques: the application of accelerator mass spectrometry to pharmacology and toxicology. *Trends Pharmacol Sci* 2004;25(8):442–447.

83. Choi MH, Skipper PL, Wishnok JS, Tannenbaum SR. Characterization of testosterone 11ß-hydroxylation by human liver microsomal cytochromes P450. *Drug Metab Dispos* 2005;33(6):714–718.

84. Sarapa N, Hsyu PH, Lappin G, Garner RC. The application of accelerator mass spectrometry to absolute bioavailability studies in humans: simultaneous administration of an intravenous microdose of $^{14}$C-nelfinavir mesylate solution and oral nelfinavir to healthy volunteers. *J Clin Pharmacol* 2005;45(10):1198–1205.

85. Saha GB. *Fundamentals of Nuclear Pharmacy*, 5th ed. New York: Springer-Verlag; 2004, p 383.

86. d'Argy R, Sundwall A. Quantitative whole-body radioluminography—future strategy for balance and tissue distribution studies. *Regul Toxicol Pharmacol* 2000;31:S57–S62.

87. Solon EG, Balani SK, Lee FW. Whole-body autoradiography in drug discovery. *Curr Drug Metab* 2002;3(5):451–462.

88. Steinke W, Archimbaud Y, Becka M, Binder R, Busch U, Dupont P, Maas J. Quantitative distribution studies in animals: cross-validation of radioluminography versus liquid-scintillation measurement. *Regul Toxicol Pharmacol* 2001;31:S33–S43.

89. Ehmann WD, Vance DE. *Radiochemistry and Nuclear Methods of Analysis*. Hoboken, NJ: Wiley; 1991.

90. Alfassi ZB. *Chemical Analysis by Nuclear Methods*. Hoboken, NJ: Wiley; 1994.

91. Robouch P, Eguskiza M, Maguregui MI, Pomme S, Pauwels J. k0-NAA, a valuable tool for reference-material producers. *Fresenius J Anal Chem* 2001;370(2–3):255–258.

92. Digenis GA, Sandefer EP, Beihn RM, Parr AF. Dual-isotope imaging of neutron-activated erbium-171 and samarium-153 and the *in vivo* evaluation of a dual-labeled bilayer tablet by gamma scintigraphy. *Pharm Res* 1991;8(10):1335–1340.

93. Awang MB, Hardy JG, Davis SS, Wilding IR, Parry SJ. Radiolabeling of pharmaceutical dosage forms by neutron activation of samarium-152. *J Labelled Compounds Radiopharmaceuticals* 1993;33(10):941–948.

94. Awang MB, Hardy JG, Davis SS, Pimm MV, Parry SJ, Wilding IR. Evaluation of $^{153}$Sm-diethylenetriaminepentaacetic acid for radiolabelling of pharmaceutical dosage forms by neutron activation. *Nucl Med Biol* 1994;21(7):905–909.

95. Ohta A, Siraishi F, Nagahara T, Tomura K. Neutron activation analysis of biological samples. *Nippon Rinsho* 1996;54(1):207–214.

96. Albert DA, Cohen AJ, Mandelbrot DA, Reinhardt CP, Dickson EW. Neutron-activation analysis: a novel method for the assay of iohexol. *J Lab Clin Med* 2003;141(2): 106–109.

97. Rajurkar NS, Pardeshi BM. Analysis of some herbal plants from India used in the control of diabetes mellitus by NAA and AAS techniques. *Appl Radiat Isot* 1997; 48(8):1059–1062.

98. Jasper JP. The increasing use of stable isotopes in the pharmaceutical industry. *Pharm Tech* 1999;23(10):106–114.

99. Jasper JP. Quantitative estimates of precision for molecular isotopic measurements. *Rapid Commun Mass Spectrom* 2001;15:1554–1557.

100. Kapetanovic IM, Kupferberg HJ. Stable isotope methodology and gas chromatography mass spectrometry in a pharmacokinetic study of phenobarbitol. *Biomed Mass Spectrom* 1980;7:47–52.

101. Schneider LV, Hall MP. Stable isotope methods for high-precision proteomics. *Drug Discov Today* 2005;10(5):353–363.

102. Stöcklin R, Arrighi J-F, Hoang-Van K, Vu L, Cerini F, Gilles N, Genet R, Markussen J, Offord RE, Rose K. Positive and negative labelling of human proinsulin, insulin and C-peptide with stable isotopes: new tools for metabolic and pharmacokinetic studies. In Chapman J, Ed. *Protein and Peptide Analysis: Advances in the Use of Mass Spectrometry,* Methods in Molecular Biology. Totowa, NJ: Humana Press; 2000, Vol 146, pp 293–315.

103. Kippen AD, Cerini F, Vadas L, Stöcklin R, Vu L, Offord RE, Rose K. Development of an isotope dilution assay for precise determination of insulin, C-peptide and proinsulin levels in non-diabetic and type II diabetic individuals with comparison to immunoassay. *J Biol Chem* 1997;272:12513–12522.

104. Vallano PT, Shugarts SB, Woolf EJ, Matuszewski BK. Elimination of autosampler carryover in a bioanalytical HPLC-MS/MS method: a case study. *J Pharm Biomed Anal* 2005;36(5):1073–1078.

105. Jemal M, Xia YQ. Bioanalytical method validation design for the simultaneous quantitation of analytes that may undergo interconversion during analysis. *J Pharm Biomed Anal* 2000;22(5):813–827.

106. Wang CJ, Pao LH, Hsiong CH, Wu CY, Whang-Peng JJ, Hu OY. Novel inhibition of cis/trans retinoic acid interconversion in biological fluids—an accurate method for determination of trans and 13-cis retinoic acid in biological fluids. *J Chromatogr B Analyt Technol Biomed Life Sci* 2003;796(2):283–291.

107. Sherma J. Chromatographic methods of analysis—thin layer chromatography. In *Encyclopedia of Pharmaceutical Technology*, 2nd ed. New York: Dekker Encyclopedias Taylor and Francis Group LLC; 2002, pp 426–439.
108. Coran SA, Giannellini V, Bambagiotti-Alberti M. High-performance thin-layer chromatographic-densitometric determination of secoisolariciresinol diglucoside in flaxseed. *J Chromatogr A* 2004;1045(1–2):217–222.
109. Petrovic M, Kastelan-Macan M, Ivankovic D, Matecic S. Video-densitometric quantitation of fluorescence quenching on totally irradiated thin-layer chromatographic plates. *J AOAC Int* 2000;83(6):1457–1462.
110. Hauck HE, Schultz M. Ultrathin-layer chromatography. *J Chromatogr Sci* 2002;40(10): 550–552.
111. Agrawal H, Kaul N, Paradkar AR, Mahadik KR. HPTLC method for guggulsterone I. Quantitative determination of E- and Z-guggulsterone in herbal extract and pharmaceutical dosage form. *J Pharm Biomed Anal* 2004;36:33–41.
112. Babu SK, Kumar KV, Subbaraju GV. Estimation of *trans*-resveratrol in herbal extracts and dosage forms by high-performance thin-layer chromatography. *Chem Pharm Bull* 2005;53(6):691–693.
113. Mahadik KR, Paradkar AR, Agrawal H, Kaul N. Stability-indicating HPTLC determination of tizanidine hydrochloride in bulk drug and pharmaceutical formulations. *J Pharm Biomed Anal* 2003;33(4):545–552.
114. Makhija SN, Vavia PR. Stability indicating HPTLC method for the simultaneous determination of pseudoephedrine and cetirizine in pharmaceutical formulations. *J Pharm Biomed Anal* 2001;25(3–4):663–667.
115. Campbell AN, Sherma J. Development and validation of a high-performance thin-layer chromatographic method with densitometric detection for determination of bisacodyl in pharmaceutical tablets. *ACTA Chromatogr* 2003;13:109–116.
116. Agrawal H, Mahadik KR, Paradkar AH, Kaul N. Stability indicating HPTLC determination of linezolid as bulk drug and in pharmaceutical dosage form. *Drug Dev Ind Pharm* 2003;29(10):1119–1126.
117. Kaul N, Agrawal H, Patil B, Kakad A, Dhaneshwar SR. Application of stability-indicating HPTLC method for quantitative determination of metadoxine in pharmaceutical dosage form. *Il Farmaco* 2005;60:351–360.
118. Pandya KK, Mody VD, Satia MC, Modi IA, Chakravarthy BK, Gandhi TP. High-performance thin-layer chromatographic method for the detection and determination of lansoprazole in human plasma and its use in pharmacokinetic studies. *J Chromatogr B Biomed Sci Appl* 1997;693(1):199–204.
119. Savale HS, Pandya KK, Gandhi TP, Modi IA, Modi RI, Satia MC. Plasma analysis of celiprolol by HPTLC: a useful technique for pharmacokinetic studies. *J AOAC Int* 2001;84(4):1252–1257.
120. Adamovics JA, ed. *Chromatographic Analysis of Pharmaceuticals*. New York: Marcel Dekker; 1997.
121. Hahn-Deinstrop E. *Applied Thin-Layer Chromatography. Best Practices and Avoidance of Mistakes*. Weinheim, Germany: Wiley-VCH; 2000.
122. Snyder LR, Kirkland JJ, Glajch JL. *Practical HPLC Method Development*, 2nd ed. Hoboken, NJ: Wiley-Interscience; 1997.
123. Clarke NJ, Rindgen D, Korfmacher WA, Cox KA. Systematic LC/MS metabolite identification in drug discovery. *Anal Chem* 2001;73(15):430A–439A.
124. Volmer DA, Sleno L. Mass analyzers: an overview of several designs and their applications, Part I. *Spectroscopy* 2005;20(11):20–26.

125. Ackermann BL, Berna MJ, Murphy AT. Recent advances in use of LC/MS/MS for quantitative high-throughput bioanalytical support of drug discovery. *Curr Top Med Chem* 2002;2(1):53–66.

126. Vallano PT, Shugarts SB, Kline WF, Woolf EJ, Matuszewski BK. Determination of risedronate in human urine by column-switching ion-pair high-performance liquid chromatography with ultraviolet detection. *J Chromatogr B Analyt Technol Biomed Life Sci* 2003;794(1):23–33.

127. Watzig H, Gunter S. Capillary electrophoresis—a high performance analytical separation technique. *Clin Chem Lab Med* 2003;41(6):724–738.

128. Altria KD, Elder D. Overview of the status and applications of capillary electrophoresis to the analysis of small molecules. *J Chromatogr A* 2004;1023(1):1–14, 720–723.

129. Hempel G. Biomedical applications of capillary electrophoresis. *Clin Chem Lab Med* 2003;41(6):720–723.

130. Belder D, Ludwig M. Microchip electrophoresis for chiral separations. *Electrophoresis* 2003;24(15):2422–2430.

131. Pang HM, Kenseth J, Coldiron S. High-throughput multiplexed capillary electrophoresis in drug discovery. *Drug Discov Today* 2004;9(24):1072–1080.

132. Koh HL, Yau WP, Ong PS, Hegde A. Current trends in modern pharmaceutical analysis for drug discovery. *Drug Discov Today* 2003;8(19):889–897.

133. Federsel HJ. Asymmetry on large scale: the roadmap to stereoselective processes. *Nat Rev Drug Discov* 2005;4(8):685–697.

134. Islam MR, Mahdi JG, Bowen ID. Pharmacological importance of stereochemical resolution of enantiomeric drugs. *Drug Safety* 1997;17(3):149–165.

135. Srinivas NR, Barbhaiya RH, Midha KK. Enantiomeric drug development: issues, considerations, and regulatory requirements. *J Pharm Sci* 2001;90(9):1205–1215.

136. Torchin CD, McNeilly PJ, Kapetanovic IM, Strong JM, Kupferberg HJ. Stereoselective metabolism of a new anticonvulsant drug candidate, losigamone, by human liver microsomes. *Drug Metab Dispos* 1996;24(9):1002–1008.

137. Waldeck B. Three-dimensional pharmacology, a subject ranging from ignorance to overstatements. *Pharmacol Toxicol* 2003;93(5):203–210.

138. Srinivas NR. Simultaneous chiral analyses of multiple analytes: case studies, implications and method development considerations. *Biomed Chromatogr* 2004; 18(10):759–784.

139. Srinivas NR. Evaluation of experimental strategies for the development of chiral chromatographic methods based on diastereomer formation. *Biomed Chromatogr* 2004; 18(4):207–233.

140. Haginaka J. Pharmaceutical and biomedical applications of enantioseparations using liquid chromatographic techniques. *J Pharm Biomed Anal* 2002;27(3–4):357–372.

141. Ward TJ. Chiral separations. *Anal Chem* 2002;74(12):2863–2872.

142. Amini A. Recent developments in chiral capillary electrophoresis and applications of this technique to pharmaceutical and biomedical analysis. *Electrophoresis* 2001;22(15): 3107–3130.

143. Andersson S, Allenmark SG. Preparative chiral chromatographic resolution of enantiomers in drug discovery. *J Biochem Biophys Methods* 2002;54(1–3):11–23.

144. Bonato PS. Recent advances in the determination of enantiomeric drugs and their metabolites in biological fluids by capillary electrophoresis-mediated microanalysis. *Electrophoresis* 2003;24(22–23):4078–4094.

145. Millot MC. Separation of drug enantiomers by liquid chromatography and capillary electrophoresis, using immobilized proteins as chiral selectors. *J Chromatogr B Analyt Technol Biomed Life Sci* 2003;797(1–2):131–159.

146. Rizzi A. Fundamental aspects of chiral separations by capillary electrophoresis. *Electrophoresis* 2001;22(15):3079–3106.

147. Scriba GK. Selected fundamental aspects of chiral electromigration techniques and their application to pharmaceutical and biomedical analysis. *J Pharm Biomed Anal* 2002; 27(3–4):373–399.

148. Scriba GK. Pharmaceutical and biomedical applications of chiral capillary electrophoresis and capillary electrochromatography: an update. *Electrophoresis* 2003;24(15): 2409–2421.

149. Misl'anova C, Hutta M. Role of biological matrices during the analysis of chiral drugs by liquid chromatography. *J Chromatogr B Analyt Technol Biomed Life Sci* 2003; 797(1–2):91–109.

150. Martin CR, Kohli P. The emerging field of nanotube biotechnology. *Nat Rev Drug Discov* 2003;2(1):29–37.

151. Bansal SK, Layloff T, Bush ED, Hamilton M, Hankinson EA, Landy JS. Qualification of analytical instruments for use in the pharmaceutical industry: a scientific approach. *AAPS PharmSciTech* 2004;5(1):E22.

152. Furman WB, Layloff TP, Tetzlaff J. Validation of computerized liquid chromatographic Systems. *J AOAC* 1994;77:1314–1318.

153. Grisanti V, Zachowski EJ. Operational and performance qualification. *LCGC North America* 2002;20 (4):356–362.

154. Huber L. *Validation and Qualification in Analytical Laboratories*. Boca Raton, FL: CRC Press; 1999.

155. Chan CC, Herman Lam H, Lee YC, Xue-Ming Zhang XM, Eds. *Analytical Method Validation and Instrument Performance Verification*. Hoboken, NJ: Wiley; 2004.

156. US Food and Drug Administration. *General Principles of Software Validation. Final Guidance for Industry and FDA Staff*. Rockville, MD: US Department of Health and Human Services, Food and Drug Administration; Jan 2002.

157. Shah VP, Midha KK, Findlay JW, Hill HM, Hulse JD, McGilveray IJ. Bioanalytical method validation—a revisit with a decade of progress. *Pharm Res* 2000;17(12): 1551–1557.

158. ICH Q2A: Text on Validation of Analytical Procedures. Fed Reg 1995;60 FR11260. http://www.fda.gov/cder/guidance/ichq2a.pdf.

159. ICH Q2B: Validation of Analytical Procedures: Methodology. Fed Reg 1997; 62 FR 27463. http://www.fda.gov/cder/guidance/1320fnl.pdf.

160. *Guidance for Industry: Analytical Procedures and Methods Validation, Chemistry, Manufacturing and Controls Documentation*. Rockville, MD: US Deptartment of Health and Human Services, Food and Drug Administration; 2000. http://www.fda.gov/cder/guidance/2396dft.pdf.

161. *Guidance for Industry: Bioanalytical Method Validation*. Rockville, MD: US Deptartment of Health and Human Services, Food and Drug Administration; 2001.

162. Hartmann C, Smeyers-Verbeke J, Massart DL, McDowall RD. Validation of bioanalytical chromatographic methods. *J Pharm Biomed Anal* 1998;17(2):193–218.

163. ICH Q1A(R2): Stability Testing of New Drug Substances and Products (Second Revision). *Fed Reg* 2003;68(225). http://www.ich.org/MediaServer.jser?@_ID=419&@_MODE=GLB.

164. ICH Q1B: Photostability Testing of New Drug Substances and Products. *Fed Reg* 1997;62(95):27115–27122. http://www.ich.org/MediaServer.jser?@_ID=412&@_MODE=GLB.
165. ICH Q1C: Stability Testing for New Dosage Forms. *Fed Reg* 1997;62(90)25634–25635. http://www.ich.org/MediaServer.jser?@_ID=413&@_MODE=GLB.
166. ICH Q1D: Bracketing and Matrixing Designs for Stability Testing of Drug Substances and Drug Products. *Fed Reg* 2003;68(11):2339–2340. http://www.ich.org/MediaServer.jser?@_ID=414&@_MODE=GLB.
167. ICH Q1E: Evaluation of Stability Data. *Fed Reg* 2004;69(110):32010–32011. http://www.ich.org/MediaServer.jser?@_ID=415&@_MODE=GLB.
168. ICH Q1F: Stability Data Package for Registration Applications in Climatic Zones III and IV. *Fed Reg* 2003;68(225):65717–65718. http://www.ich.org/MediaServer.jser?@_ID=416&@_MODE=GLB.
169. ICH Q3A(R): Impurities in New Drug Substances (Revised Guideline). *Fed Reg* 2003;68(68):6924–6925. http://www.ich.org/MediaServer.jser?@_ID=422&@_MODE=GLB.
170. ICH Q3B(R): Impurities in New Drug Products (Revised Guideline). *Fed Reg* 2003;68(220):64628–64629. http://www.ich.org/MediaServer.jser?@_ID=421&@_MODE=GLB.
171. ICH Q3C: Impurities: Guideline for Residual Solvents. *Fed Reg* 1997;62(247):67377. http://www.ich.org/MediaServer.jser?@_ID=423&@_MODE=GLB.
172. ICH Q5C: Quality of Biotechnological Products: Stability Testing of Biotechnological/Biological Products. *Fed Reg* 1996;61:36466. http://www.ich.org/MediaServer.jser?@_ID=427&@_MODE=GLB.
173. Q9: Quality Risk Management. *Fed Reg* 2005;70(151):45722–45723. http://www.ich.org/LOB/media/MEDIA1957.pdf.
174. United States Pharmacopoeial Convention. *United States Pharmacopoeia 26, National Formulary 21, ⟨1225⟩ Validation of Compendial Methods*. Rockville, MD: United States Pharmacopoeial Convention; 2003.
175. *Pharmacopeial Forum* 2005;31(2):549.
176. Swartz ME, Kull IS. Validation, qualification, or verification. *LCGC North America* 2005;23(10):47.
177. *Pharmacopeial Forum* 2005;31(2):555.
178. ISO 9000:2000 *Quality Management Systems—Fundamentals and Vocabulary*. Geneva, Switzerland: International Organization for Standardization; 2000.
179. ISO 9001:2000 *Quality Management System—Requirements*. Geneva, Switzerland: International Organization for Standardization; 2000.
180. ISO 9002:2000 *Production and Installation QMS Only*. Geneva, Switzerland: International Organization for Standardization; 2000.
181. ISO 9000:2005 *Quality Management Systems—Fundamentals and Vocabulary*. Geneva, Switzerland: International Organization for Standardization; 2005.
182. ISO/IEC 90003:2004 Software *Engineerin—Guidelines for the Application of ISO 9001:2000 to Computer Software*. Geneva, Switzerland: International Organization for Standardization; 2004.
183. ANSI/ISO/ASQ Q9000-2000 *Series*—Quality Management Standards. Milwaukee: ASQ Quality Press; 2001.
184. ISO/IEC Guide 25: *General Requirements for the Competence of Calibration and Testing Laboratories*, 3rd edition. Geneva, Switzerland: International Organization for Standardization; 1990.

185. ISO/IEC DIS 17025: *General Requirements for the Competence of Testing and Calibration Laboratories*. Geneva, Switzerland: International Organization for Standardization; 1998.
186. ISO/IEC 17025:2005 *General Requirements for the Competence of Testing and Calibration Laboratories*. Geneva, Switzerland: International Organization for Standardization; 2005.
187. Taylor JK. *Quality Assurance of Chemical Measurements*. Chelsea, MI: Lewis Publishers; 1987, p 36.
188. Lide D. *Handbook of Chemistry and Physics*, 86th ed. Boca Raton, FL: CRC Press; 2005.
189. Reviewer Guidance: Validation of Chromatographic Methods. Rockville, MD: US Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research (CDER); 1994. http://www.fdagov/cder/guidance/cmc3.pdf.
190. Ermer J, Vogel M. Applications of hyphenated LC-MS techniques in pharmaceutical analysis. *Biomed Chromatogr* 2000;14(6):373–383.

# CHEMICAL AND PHYSICAL CHARACTERIZATIONS OF POTENTIAL NEW CHEMICAL ENTITY

Adegoke Adeniji and Adeboye Adejare*
*Philadelphia College of Pharmacy, University of the Sciences in Philadelphia, Philadelphia, Pennsylvania*

**Contents**

1 Introduction
2 Determination of Mass
3 Structure Elucidation
4 Purity
5 Solubility and Lipophilicity
6 Stability
7 Proteins and Peptides
8 Miscellaneous Techniques
9 Case Studies
10 Conclusion
References

## 1 INTRODUCTION

A new chemical entity (NCE) is a compound that has not been approved as a pharmaceutical agent for a particular indication. In the United States, the approving body is the Food and Drug Administration (FDA) and most countries have a similar agency [1]. The compound may be an already approved drug being evaluated for a

**Corresponding author*.

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

new indication or a novel compound. New regulations, the ever increasing demand for drugs worldwide, and an unprecedented need to drive drug prices lower challenge the pharmaceutical companies and point to the need to put more lead compounds into the development pipeline [2]. While these chemical entities are constantly being synthesized or isolated from natural sources in various academic and industrial laboratories throughout the world, very few compounds make it to the market as drugs. It is estimated that only one out of every 5000 synthesized compounds ends up on the market as a therapeutically useful drug [3].

Elucidation of the chemical and physical properties of a potential NCE is critical to drug discovery efforts as part of "pharmaceutical profiling." Knowledge of these properties could give insight into how the compound acts or binds to the target receptor to elicit desired or undesired effects and also help with formulation studies [4, 5]. Based on these properties, some compounds could be dismissed as possible drug candidates. For instance, a compound being proposed for intravenous administration that has very poor aqueous solubility might not be favorably considered because of envisaged problems with formulation and bioavailability. Indeed, the fate of many NCEs are decided immediately after the chemical and physical properties of the compounds are determined. These properties often serve as indices by which a compound will be judged for its "drugability" and likelihood of eliciting the desired response. Thus, the journey to the market for a compound that becomes a drug begins with determination of its basic chemical and physical properties such as the mass, structure, solubility, and stability. Further physicochemical characterizations, for example, if the compound has different polymorphs, may follow should the compound exhibit a desired biological profile. However, those may not be warranted if a "cut" decision is made at an earlier stage. A stage-appropriate characterization is therefore critical for every potential NCE for maximization of resources. The most common chemical and physical characterization methods for a novel potential NCE and two examples will be described. If the potential NCE is already approved for an indication, then many pertinent chemical and physical properties would have been determined and are likely to be in literature. This brief chapter focuses on small molecules although many of the methods described have been adapted to characterize therapeutic proteins and other macromolecules.

## 2 DETERMINATION OF MASS

Most drugs are small molecules with mass in the range 100–1000 amu. There are, however, notable exceptions such as lithium salts at the lower end, while peptides, proteins, and antibodies being used in therapy could have masses much greater than 1000 amu. The mass of a compound is the sum of the atomic masses of its constituent elements. It is generally expected that a non-ionic compound with low molecular mass will be absorbed more readily than one with high mass under similar conditions, such as being in the same chemical class and dosage form. This is particularly true for orally administered drugs. Lipinski's rule of five [6, 7] indicates that a molecular mass less than or equal to 500 is essential for passive absorption of an orally administered drug. There are several methods used to determine the molecular mass of a compound. These methods include vapor density, vapor pressure osmometry, cryoscopy, elemental analysis, and undoubtedly the most popular and

important, mass spectrometry [8–11]. Mass spectrometry (MS) and elemental analysis are the two major techniques used in mass determination. MS is a very versatile tool that is applicable in several fields of science and new applications are being found. It involves a high energy bombardment of the molecule of interest to generate charged fragments, which are detected based on the mass/charge ratio. The mass of the compound is established by the mass of the molecular ion or those of its fragments. The molecular formula and the structure of the compound can also be obtained by studying the fragmentation pattern of the compound [9, 10]. MS has also been widely used in determining the structures of proteins and other macromolecules. Quite often, the molecular masses of such peptides and proteins can be inferred from fragment ions.

Elemental analysis, on the other hand, determines the composition of a sample by giving the amount (usually percentage weight) of each element that is present in the sample. It is often based on oxidation of an organic compound at elevated temperatures to produce gaseous molecules of the elements of interest, which are then separated and analyzed [11]. It is very useful as it gives values from which the empirical formula of the compound can be calculated. The molecular formula can then be determined from the empirical formula and molecular weight. By determining the proportion of the constituent elements of a compound, the purity of the compound can also be established.

## 3 STRUCTURE ELUCIDATION

Most drugs are weakly acidic or basic organic compounds [12]. The determination of structure helps to identify the functional groups on the compound, validate the synthetic or isolation pathway, and initiate structure–activity relationship studies that are very important in drug discovery [13]. It also helps to determine physicochemical parameters. The process of structure elucidation involves determining the elements present and their bonding arrangements in the compound and establishes the identity of the chemical entity. Several methods are used for structure elucidation, most of which are spectroscopic. Of these, nuclear magnetic resonance (NMR) is the most versatile and is thus a key tool in determining structures of molecules. NMR exploits the ability of certain nuclei to align themselves either with or against the field when placed in a magnetic field. Since the effective magnetic field experienced by these nuclei is influenced by the electronic environment in their proximity (e.g., the presence of a neighboring strongly electronegative atom), information about atom connectivity can be deduced from spectra analysis [14]. Both $^{1}$H- and $^{13}$C-NMR are widely used but several other nuclei such as $^{19}$F and $^{31}$P are also utilized [14–16].

Ultraviolet/Visible (UV/Vis) spectroscopy and particularly infrared (IR) spectroscopy are also widely used. UV/Vis spectroscopy gives information about the presence of chromophores (i.e., conjugated double bonds), which are responsible for selective absorption of radiation in a given wavelength range [14]. IR spectroscopy gives information about functional groups in the molecule. Both methods are based on the ability of compounds to absorb a fraction of incident radiation. In IR spectroscopy, the absorption frequency confirms the presence of particular functional groups (e.g., carbonyl, hydroxyl, and aryl) since these values are consistent

and only slightly influenced by adjacent groups in the molecule. There are charts and tables in the literature that list the absorption frequencies of common organic functional groups and they can be used as reference. The electronic collection of spectra and the availability of spectra simulating software have increased the usefulness of these and other spectroscopic techniques.

Polarimetry measures the ability of a sample to rotate plane polarized light. It is critical for the characterization of optically active or chiral compounds. The specific rotation of a chemical compound. $[\alpha]_D^T$ is defined as the observed angle of rotation $\alpha$, when light of 589 nm wavelength (the sodium D line) is passed through a sample with a pathlength of 0.1 m and a sample concentration of 1 g/ml at temperature T. It is usually measured at 20 °C. It is applicable for both qualitative and quantitative purposes since the extent of light rotation is dependent on the molecular structure and the concentration of the sample. Thus, it can be used to determine enantiomeric purity of a sample.

X-ray crystallography gives vital information about the crystal structure of a compound using diffraction techniques. It allows for measurement of atomic connectivity, bond lengths, bond angles, and torsion angles, thereby giving very detailed information about the structure of a compound [14]. The compounds, however, need to be crystallizable, which is sometimes very difficult, especially for large macromolecules such as protein and DNA complexes. For small molecules, it can be used to determine absolute configurations at chiral centers. This can become very important in discovery efforts as it can help explain compound reactivity and binding to receptors (proteins), which contain many chiral centers.

## 4 PURITY

The presence of impurities in an NCE can not only affect the physicochemical properties but could also preclude correct assessment of the bioactivity and toxicity profiles of the target compound. It is generally desirable that the compound be available in a very pure form before incorporation into any dosage form for *in vitro* or *in vivo* studies. Several methods including some that have been discussed earlier can be used to assess the purity of a compound. These methods include NMR, mass spectrometry, chromatography, UV spectroscopy, IR spectroscopy, and polarimetry. However, in many cases, the very first test to determine the purity is melting point (mp) for solids and boiling point (bp) for liquids. A broad range of values for either parameter is usually indicative of impurity [17]. The mp and bp as well as the refractive index and density are defining characteristics of a compound. Refractive index measures the ratio of the velocity of light in air to that in the sample at a particular wavelength, temperature, and pressure.

Chromatography separates mixtures based on the different rates of movement of the components along an adsorbent medium. Ever since its invention by the Russian botanist Mikhail Tswett in the early twentieth century [18], its application has increased tremendously to earn it a place of pride in analytical science. From the relatively simple thin layer chromatography (TLC) to the more technical gas chromatography (GC) and high performance liquid chromatography (HPLC), chromatographic techniques allow for separation of the components of a sample based on the ease of mobility through a stationary phase. The presence of a single band

in TLC or a single peak in HPLC or GC suggests high purity of the new compound. Various detectors can be used, from UV to fluorescence to MS. Chromatographic techniques are used not only to determine purity but also in purification [19]. The high sensitivity shown by the low limit of detection of these techniques has been employed in quantitative compound determination in plasma and other matrices, where they have been used to detect nanogram/milliliter concentrations of drugs [20, 21]. A major advantage of chromatographic techniques is that they can be adapted for quantitative determinations and/or joined with other techniques to give hyphenated methods (e.g., LC-MS). HPLC has also been used in determining purity of enantiomers [22, 23]. In such cases, a chiral column or derivatization technique to give diastereomers is employed.

## 5 SOLUBILITY AND LIPOPHILICITY

Solubility of a sample in a solvent refers to the amount of the sample that will dissolve in the given solvent at a particular temperature. It is generally desirable for a drug molecule to have some level of water solubility. Poor water solubility can present many problems from reliability and reproducibility of several *in vitro* tests to significant formulation issues [7, 24]. Solubility is typically measured using the shake flask method. The shake flask method involves adding an excess amount of a solid sample to a flask containing a specified amount of buffer; the saturated solution is then agitated until it attains equilibrium, after which the undissolved sample is separated from the solution and the concentration of the sample is determined usually by UV spectroscopy or another applicable technique [25].

Since most drugs are weak acids or bases [11], the $pK_a$ is an important parameter used to study and predict the behavior of an NCE in a physiological environment. The $pK_a$ or *dissociation constant* is a measure of the acid strength of a compound. It allows for determination of the predominant state of an ionizable compound at any given pH and therefore affects solubility of the compound. In general, aqueous solubility increases with ionization. The $pK_a$ can be determined using several methods such as capillary electrophoresis, liquid chromatography, and potentiometry.

The relevant equations are

$$pK_a = -\log K_a \tag{1}$$

Acids: $HA \rightleftharpoons H^+ + A^-$

$$K_a = [H^+][A^-]/[HA], \quad \text{where } [\,] = \text{molar concentration} \tag{2}$$

Bases: $HB^+ \rightleftharpoons H^+ + B$

$$K_a = [H^+][B]/[HB^+] \tag{3}$$

*Lipophilicity* refers to the tendency of a compound to partition into nonpolar (organic) versus aqueous environments [26]. Log *P* refers to the logarithm of the partition coefficient and it is the most common measure of lipophilicity of a compound. The partition coefficient is a constant that defines the ratio of a compound in aqueous medium to the compound in an immiscible solvent at equilibrium. The

most commonly used organic solvent for biological purposes is 2-octanol. The partition coefficient is probably one of the most important physicochemical properties that infer the crossing of a drug across biological membranes, binding to proteins, and, consequently, the activity and the ease of excretion from the body. The partition coefficient has been shown to correlate with the activity of many classes of drugs [27, 28]. A compound with log $P > 5$ is usually not considered a good candidate in drug development [6, 7] because it is considered to be too lipophilic.

$$\log P = \log_{10}(\text{Partition coefficient}) \tag{4}$$

where Partition coefficient, $P$ = [Organic]/[Aqueous].

It is generally desirable for an NCE to possess enough water solubility so that it can dissolve in body fluids while also possessing sufficient lipophilicity to cross biological membranes. Log $P$ describes the partitioning for a neutral and un-ionizable molecule, which is often not the case. For ionizable compounds, log $P$ is expected to change with pH. The relationship between log $P$ and pH is described by the log distribution coefficient (log $D$), also called the apparent partition coefficient. Log $D$, unlike log $P$, changes with pH as it depends on the acidic or basic nature of the compound. The log $D$ at pH 7.4 is often used as an index of the behavior of a sample in plasma.

$$\text{Distribution coefficient, } D = [\text{Un-ionized}]_{(o)}/[\text{Un-ionized}]_{(aq)} + [\text{Ionized}]_{(aq)} \tag{5}$$

$$\log D = \log_{10}(\text{Distribution coefficient}) \tag{6}$$

$$\log D_{(pH)} = \log P - \log(1 + 10^{(pH - pK_a)}) \quad \text{for acids} \tag{7}$$

$$\log D_{(pH)} = \log P - \log(1 + 10^{(pK_a - pH)}) \quad \text{for bases} \tag{8}$$

## 6 STABILITY

Physical and chemical stability considerations become essential to estimate how long the NCE can be stored without significant decomposition. Stability is usually determined by monitoring particular properties of the compound over a period of time, usually days or weeks and sometimes months, under specified storage conditions, which are often designed to be more unfavorable (stress testing) than the envisioned storage conditions [29, 30]. The stability of the compound is assessed in the presence of elevated temperature, light, high relative humidity (75% or higher), varying pH range, and an oxidative environment.

## 7 PROTEINS AND PEPTIDES

Peptides usually exhibit poor *in vivo* stability, pharmacokinetics, and bioavailability. These often pose a major challenge to drug discovery and development efforts. Therefore, most drugs are small molecules or peptidomimetics [31]. With advances in drug delivery and more sophisticated equipment, an increasing number of peptides and proteins are being used in therapy [32, 33].

## 8 MISCELLANEOUS TECHNIQUES

There are many other techniques used to decipher other chemical and/or physical properties of an NCE. Examples include circular dichroism (CD), Raman spectroscopy, and differential scanning calorimetry (DSC). CD is a type of absorption spectroscopy often used to determine optical properties and secondary structures of molecules. While it can be used for any optically active molecule, it is more critical in cases of biomolecules such as sugars, amino acids, peptides, and proteins when embarking on structural determination. It measures the difference in the absorption of right- and left-circularly polarized light as a function of wavelength. It is used extensively to study secondary protein structures: how the secondary structure of a molecule changes as a function of temperature or the concentration of denaturing agents, thus making it a vital tool in assessing the physical and chemical stability of peptides and proteins. Raman spectroscopy is a technique that is used to study the vibrational, rotational, and low frequency modes in a system. Its use derives from the Raman effect, which is the inelastic scatter of photons following light incidence. Its principal use is to identify a compound based on a fingerprint that is unique to the compound, since vibrational information is very specific for chemical bonds in molecules. The fingerprint region of organic molecules is in the range of 500–2000 $cm^{-1}$. More uses are being discovered for MS. High resolution MS is being accepted as a substitute for elemental analysis by several journals. Purity of amines is sometimes difficult to determine using chromatographic techniques. An MS-based technique that can help identify amine impurities using hydrogen/deuterium exchange has been developed [34].

DSC is a thermoanalytical method that measures the difference in heat absorbed by a chemical compound and a reference as a function of temperature when the two are subjected to the same regulated temperature conditions. It measures the changes in heat flow as the sample and reference undergo any physical transformation, such as phase transition. It is related to several parameters of a compound and is thus used mainly to establish identity and purity of chemical compounds among other uses such as polymorph characterization [35, 36].

## 9 CASE STUDIES

For illustration, we now examine the chemical and physical characterizations of two new potential NCEs.

***Example A.*** 4-Fluoro-*N*-(adamantan-2-yl)-benzenesulfonamide (Fig. 1) is a novel compound that was designed and synthesized as a gamma secretase inhibitor [37, 38]. Gamma secretase is an enzyme involved in amyloid precursor protein processing to give amyloid peptides, which are the building blocks of the plaques found in the brains of patients with Alzheimer's disease (AD). It has been hypothesized that gamma secretase inhibitors can halt the onset and/or progression of AD. After synthesis, the melting point for the target compound was determined to be 150–152 °C. The narrow range suggested good purity for the sample. Elucidation of the structure was carried out using IR spectroscopy and NMR, the results of which

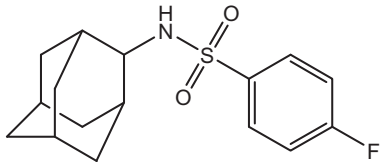


**FIGURE 1** Structure of 4-fluoro-*N*-(adamantan-2-yl)-benzenesulfonamide.

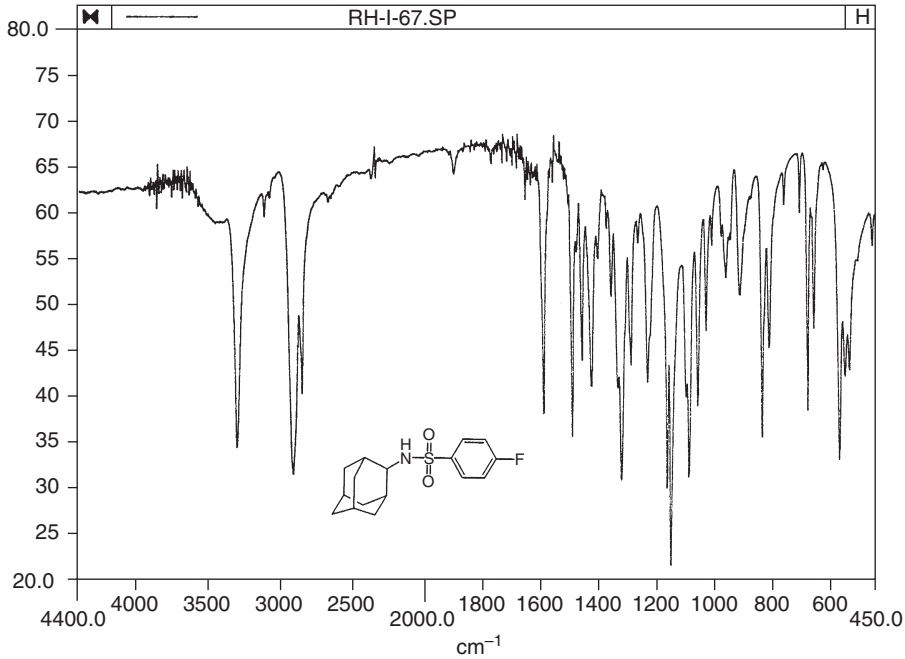


**FIGURE 2** FT-IR spectrum of 4-fluoro-*N*-(adamantan-2-yl)-benzenesulfonamide in KBr.

confirmed the presence of the expected functional groups. IR (KBr) $cm^{-1}$: 3300 (NH); 1322, 1166 (O=S=O); 1234 (Ar—F) (Fig. 2). The $^{1}H$-, $^{13}C$-, and $^{19}F$-NMR results further confirmed the presence of the expected functional groups; the results are shown in Fig. 3. The molecular mass determination using mass spectrometry and elemental analysis, the results of which are shown below, also support the proposed structure of the compound.

MS ESI (*m/z*, species, %): 308.3, $[M-H]^-$, 100% as shown in Fig. 4.

The elemental analysis of the compound showed the proportion of the constituent elements to be the following. Calculation for $C_{16}H_{20}FNO_2S$: C, 62.11; H, 6.52; F, 6.14; N, 4.53. Found: C, 62.06; H, 6.49; F, 6.43; N, 4.37. The "found" is within ±0.4 of the calculated values and thus consistent with the proposed structure.

$^{1}H$-NMR ($CDCl_3$) δ: 7.9 (s, 2H, 2 × Ar—CH), 7.4–7.1 (d, 2H, 2 × Ar—H), 4.9 (s, 1H, 1 × NH), 3.4 (s, 1H, 1 × CH), 1.8–1.5 (m, 15H, 5 × $CH_2$, 5 × CH).

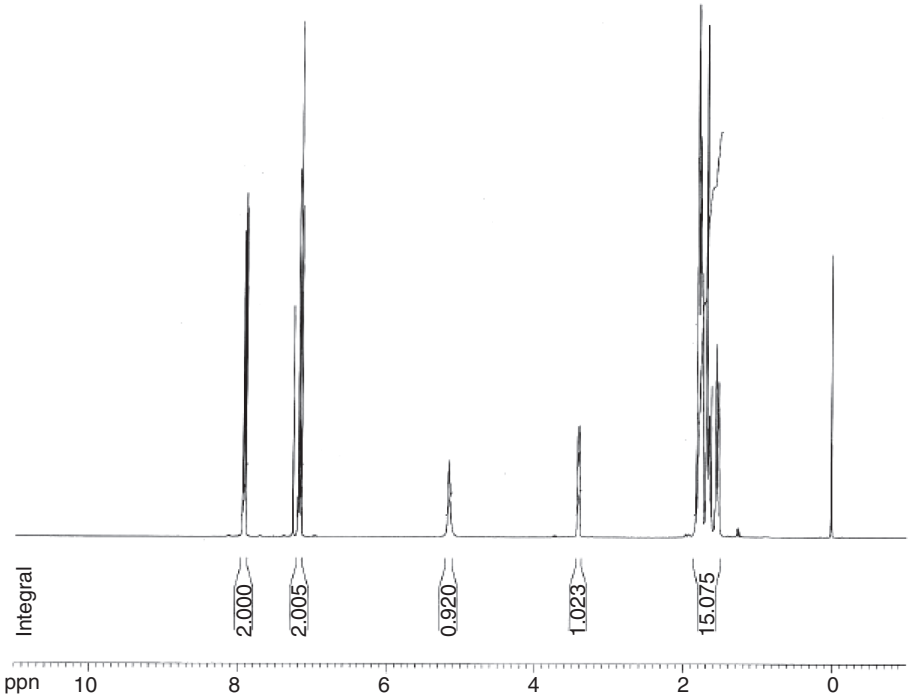


(a)

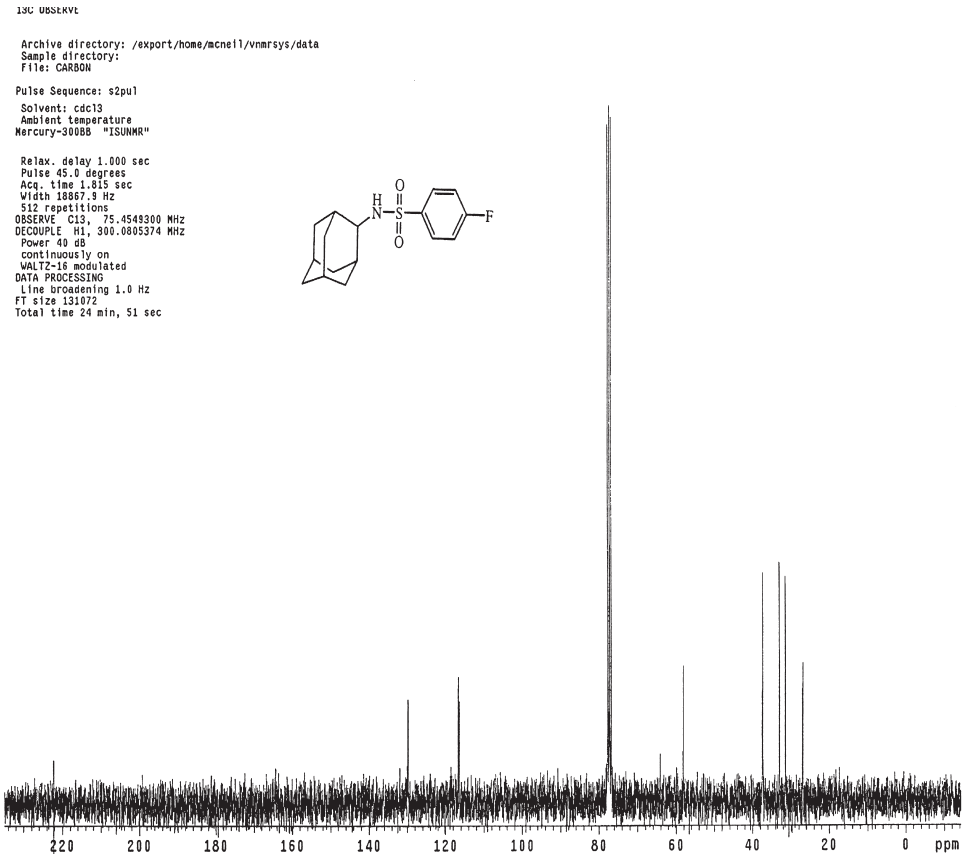


(b)

**FIGURE 3** (a) H-NMR, (b) $^{13}$C-NMR, and (c) $^{19}$F-NMR of 4-fluoro-*N*-(adamantan-2-yl)-benzenesulfonamide in $CDCl_3$.

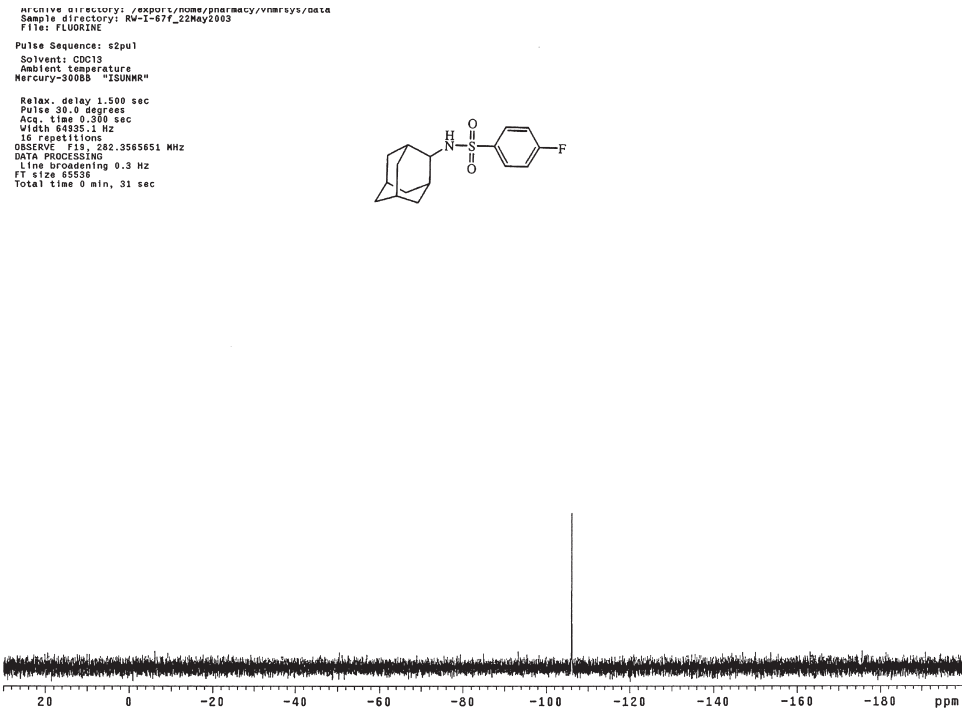


(c)

**FIGURE 3** *Continued*

$^{13}$C-NMR ($CDCl_3$) δ: 130 (s, 2C, 1 × Ar—CF, 1 × Ar—$CSO_2$), 116 (s, 4C, 4 × Ar—CH), 58 (s, 1C, 1 × CHN), 37 (s, 2C, 2 × CH), 33 (s, 2C, 2 × CH), 31 (s, 4C, 4 × $CH_2$), 27 (s, 1C, 1 × $CH_2$).

The equilibrium solubility of this compound in water and phosphate buffers (pH 7.4 and 11.50) were determined at 23.0 ± 1.0 °C by dissolving the compound in 2 mL of the appropriate solvent, shaking, and then centrifuging at 220 rpm for 48 h. The concentration of the drug was measured by UV spectrophotometry.

Determination of the p$K_a$ was conducted using the shake flask method, while the log *P*(octanol/water) was determined using both the shake flask method and HPLC with UV detection. It was also calculated using commercially available software, ChemDraw Ultra®. The agreement of the values gives some confidence in terms of getting a good handle on the lipophilicity of the compound. The results of the solubility, p$K_a$, and log *P* determination are given in Table 1.

In summary, the compound possesses adequate water solubility to dissolve in intestinal fluid and the partition coefficient of 3.4 suggests that the compound possesses acceptable lipophilicity. The compound also has molecular weight of 309.4. It has only one hydrogen bond (HB) donor group (less than the maximum of 5) and less than 10 hydrogen bond acceptor groups, with a generous count reflecting 4. Based on Lipinski's rule of five (Fig. 5), one could conclude that this compound falls

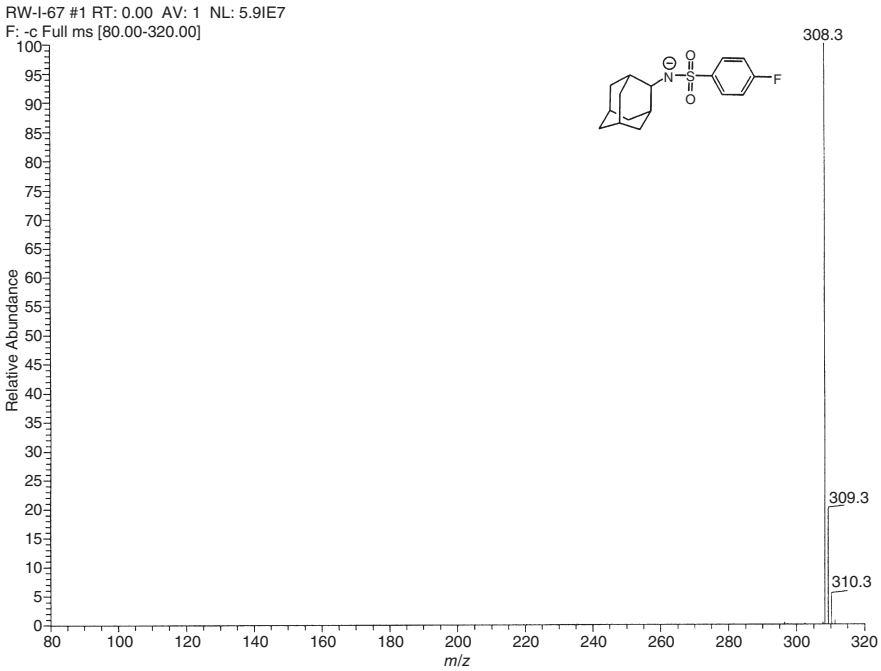


**FIGURE 4** Electrospray mass spectrum of 4-fluoro-*N*-(adamantan-2-yl)-benzenesulfonamide *m*/*z* 308.3, in MeOH.

**TABLE 1 Solubility, log *P*, and p$K_a$ of 4-Fluoro-*N*-(adamantan-2-yl)-benzenesulfonamide**

| | |
|---|---|
| Solubility (23.0 ± 1.0 °C) (μg/mL) | |
| Water | 120 ± 50 |
| Phosphate buffer (pH 7.4) | 200 ± 30 |
| Phosphate buffer (pH 11.50) | 270 ± 60 |
| Log partition coefficients | |
| Calculated | 3.44 |
| Shake flask method | 3.36 ± 0.16 |
| HPLC | 3.31 ± 0.01 |
| p$K_a$ | 10.36 ± 0.11 |

"inside the box" and thus has the appropriate physicochemical properties necessary for absorption in an oral formulation.

***Example B*** (2*R*)-3-(4-fluoro-1-naphthyloxy)-1-(*tert*-butyl-amino) 2-propanol (Fig. 6) was synthesized to determine the effect of fluorine substitution on the aromatic ring on adrenergic receptor activities [23, 39]. Similar compounds are potent beta-adrenergic receptor antagonists and are used in the control of blood pressure. Upon completion of synthesis, the melting point of the free base was found to be 56–57 °C, while that of the hydrochloride salt was 165–166 °C. The sharp melting points were indicative of good purity. The structure of the compound was determined using data

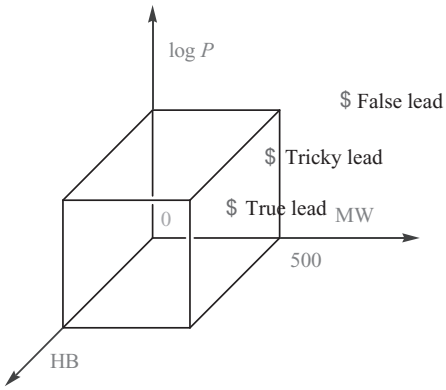


**FIGURE 5** Optimal properties for orally active compounds are found within the "log *P*-HB-MW" box.

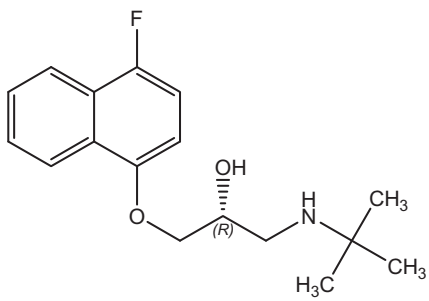


**FIGURE 6** Structure of (2*R*)-3-(4-fluoro-1-naphthyloxy)-1-(*tert*-butyl-amino) 2-propanol.

from the proton NMR analysis of the sample: $^1$H-NMR ($CDCl_3$) δ: 8.27–7.40 (m, 4H, aryl H), 7.0–7.10 (t, 1H aryl H), 6.75–6.65 (dd, 1H aryl H), 4.10 (s 2H, C**H**$_2$CHOH), 2.95–2.40 (m, 3H, C**H**OHC**H**$_2$N), 1.11 (s, 9H, C(C**H**$_3$)$_3$). This was supported by the molecular mass determination using mass spectrometry. The base peak was at 292 (M + 1) with the next major peak being 236 (M + 1 – 56), which is due to the cleavage of the *t*-butyl group. The elemental analysis of the hydrochloride salt gave $C_{17}H_{23}NFO_2Cl$, which further supported the structure proposed. The optical rotation values of the free base and hydrochloride salt were determined to be $[\alpha]_D^{25} = +3.71°$ (c = 1.96, EtOH) and $[\alpha]_D^{25} = +13.2°$ (c = 0.81, EtOH), respectively. Enantiomeric purity of the compound was evaluated by HPLC with UV detection. There was a major peak at 3.7 minutes for the *R*-enantiomer (98%) and a very minor peak at 3.15 minutes for the *S*-enantiomer.

## 10 CONCLUSION

As the cost of drug discovery and development continues to rise, the need to make important "cut" decisions as early in the process as possible is more pertinent now than ever. Along with potency and efficacy of the compound to elicit the desired

biological effects, the importance of determining appropriate chemical and physical properties cannot be overemphasized, as they could play a key role in the "cut" decisions. Advances in technology have produced several methods, many of which have been highlighted in this chapter and are at the disposal of the medicinal chemist to characterize a potential new chemical entity.

## REFERENCES


1. http://www.fda.gov/cder/about/smallbiz/exclusivity.htm.
2. Bradley D. High (throughput) mass. *Modern Drug Discov* 2003;6(4):43–45.
3. Tufts Centre for the Study of Drug Development. Backgrounder: how new drugs move through the development and approval process, 2001. http://csdd.tufts.edu/NewsEvents/RecentNews.asp?newsid=4.
4. Martin A. *Physical pharmacy*, 4th ed. Philadelphia: Lea & Febiger; 1993, Ch 4, pp 77–78.
5. Dewitte RS, Kolovanov E. *Pharmaceutical Profiling in Drug Discovery for Lead Selection*. Arlington, VA: AAPS Press; PUBX; 2004, Vol 1, Ch 2, pp 27–28.
6. Lipinski CA, Lombardo F, Dominy BW, Feeney PJ. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv Drug Deliv Rev* 1997;23:4–25.
7. Lipinski CA. Drug like properties and the cause of poor solubility and poor permeability. *J Pharmacol Toxicol Methods* 2000;44:235–249.
8. Gupta PK. Solutions and phase equilibra. In *Remington: The Science and Practice of Pharmacy*, 21st ed. Baltimore: Lippincott Williams &Wilkins; 2005, Ch 16, pp 224–228.
9. Skoog DA, Holler FJ, Nieman TA. *Principles of Instrumental Analysis*, 5th ed. Belmont, CA: Brooks/Cole and Thomson Learning; 1998, Ch 33, pp 524–528.
10. Chaudhry AK, Singh GS, Stephenson GA, Ackermann BL. Instrumental methods of analysis. In *Remington: The Science and Practice of Pharmacy*, 21st ed. Baltimore: Lippincott Williams & Wilkins; 2005, Ch 34, pp 633–636.
11. Skoog DA, Holler FJ, Nieman TA. *Principles of Instrumental Analysis*, 5th ed. Belmont, CA: Brooks/Cole and Thomson Learning; 1998, Ch 33, pp 844–845.
12. Zimmerman JJ, Feldman S. Physical–chemical properties and biological activity. In *William Foye's Principles of Medicinal Chemistry*, 2nd ed. Philadelphia: Lea & Febiger; 1981, Ch 2, pp 39–40.
13. Verkman AS. Drug discovery in academia. *Am J Physiol Cell Physiol* 2004;286: C465–C474.
14. Chaudhry AK, Singh GS, Stephenson GA, Ackermann BL. Instrumental methods of analysis. In *Remington: The Science and Practice of Pharmacy*, 21st ed. Baltimore: Lippincott Williams & Wilkins; 2005, Ch 34, pp 635–640.
15. Dalvit C, Fagerness PE, Hadden DTA, Sarver RW, Stockman BJ. Fluorine-NMR experiments for high-throughput screening: theoretical aspects, practical considerations, and range of applicability. *J Am Chem Soc* 2003;125(25):7696–7703.
16. Morgan KR, MacLachian DJ. $^{31}$P solid state NMR studies of the structure of amine-intercalated a-zirconium phosphate. 2. Titration of a-zirconium phosphate with *n*-propylamine and *n*-butylamine. *J Phys Chem* 1992;96:3458–3464.
17. Eaton DC. *Laboratory Investigations in Organic Chemistry*. New York: McGraw-Hill; 1989, Experiment 2, pp 65–67.

18. Skoog DA, Holler FJ, Nieman TA. *Principles of Instrumental Analysis*, 5th ed. Belmont, CA: Brooks/Cole and Thomson Learning; 1998, Ch 26, pp 674–677.
19. Nopper B, Kohen F, Wilchek M. A thiophilic adsorbent for the one-step high performance liquid chromatography purification of monoclonal antibodies. *Anal Biochem* 1989;180(1):66–71.
20. Permuleter JS, Loftin S, Carl JL, Karimi. M. Modified high-performance liquid chromatography with electrochemical detection method for plasma measurement of levodopa, 3-*O*-methyldopa, dopamine, carbidopa and 3,4-dihydroxyphenylacetic acid. *J Chromatogr B* 2006;836:120–123.
21. Axe BP. Trace level detection and quantification of ethyl-diazoacetate by reverse phase high performance liquid chromatography and UV detection. *J Pharm Biomed Anal* 2006;41:804–810.
22. Saito M, Morris S, Yomota C, Iwaya K, Kudo K. Determination of enantiomeric purity of hyoscyamine from scopolia extract using HPLC-CD without chiral separation. *Enantiomer* 2000;5(3–4):369–375.
23. Adejare A, Deal SA, Day MS. Syntheses and β-adrenergic binding affinities of (*R*)- and (*S*)-fluoronaphthyloxypropanolamines. *Chirality* 1999;11:144–148.
24. Lipinski CA. Avoiding investment in doomed drugs. *Curr Drug Discov* 2001;Apr:17–19.
25. Comer JEA. High throughput measurement of log *D* and p$K_a$ in drug bioavailability: estimation of solubility, permeability, absorption and bioavailability. In *Methods and Principles in Medicinal Chemistry, Volume 18*. Weinheim: Wiley-VCH; 2003, Ch 2, pp 21–45.
26. Di. L, Kerns EH. Pharmaceutical profiling in drug discovery. *Drug Discov Today* 2003;8(7):316–323.
27. Chui, W-K, Tsun-Hon Wong P, Thenmozhiyal JC. Anticonvulsant activity of phenylmethylenehydantoins: a structure–activity relationship study. *J Med Chem* 2004; 47:1527–1535.
28. Kurz H, Hofstetter A, Teschemacher H, Herz A, The importance of lipid-solubility for the central action of cholinolytic drugs. *Neuropharmacology* 1965;4(4):207–218.
29. Xu XH, Song JZ, Qaio CF, Tai J, Cheung S, Han QB. Stability and cytotoxicity of gambogic acid and its derivative gambogoic acid. *Biol Pharm Bull* 2005;28(12):2335–2337.
30. Trissel LA, Zhang Y, Xu Q. Physical and chemical stability of gemcitabine hydrochloride solutions. *J Am Pharm Assoc* 1999;39(4):509–513.
31. Lipinski C, Hopkins A. Navigating chemical space for biology and medicine. *Nature* 2004;432:855–861.
32. Kirkpatrick P, Lebbos J, LaBonte J. Fresh from the pipeline: enfuvirtide. *Nat Rev Drug Discov* 2003;2:345–346.
33. Gefter ML, Wallner BP. Peptide treatment for allergic diseases. *Clin Immunol Immunopathol* 1996;80(2):105–109.
34. Adejare A, Brown PW. Hydrogen/deuterium exchange to differentiate fragments ions from pseudomolecular ions by electrospray tandem mass spectrometry. *Anal Chem* 1997;69(8):1525–1529.
35. Skoog DA, Holler FJ, Nieman TA. *Principles of Instrumental Analysis*, 5th ed. Belmont, CA: Brooks/Cole and Thomson Learning; 1998, Ch 31, pp 805–808.
36. Martin A. *Physical Pharmacy*, 4th ed. Philadelphia: Lea & Febiger; 1993, Ch 2, pp 46–48.
37. Wells RM. *Design and synthesis of non-peptide rigid sulfonamide gamma secretase inhibitors*. MS thesis, Idaho State University, 2003.

38. Adejare A, El-Gendy AM. Membrane permeability related physicochemical properties of a novel γ-secretase inhibitor. *Int J Pharm* 2004;280:45–55.
39. Adejare A, Sciberras SS. Syntheses and β-adrenergic activities of (*R*)-fluoronaphthyloxypropanolomaine. *Pharm Res* 1997;14(4):533–536.

# PERMEABILITY ASSESSMENT

Srinivas Ganta, Puneet Sharma, and Sanjay Garg
*University of Auckland, Auckland, New Zealand*

**Contents**

1 Introduction
2 Nature and Permeability of Some Physiological Barriers
3 Drug Transfer Process Through a Physiological Membrane
4 Physicochemical Properties Determining Membrane Permeation
  4.1 Lipophilicity
  4.2 Molecular Size
  4.3 Hydrogen Bonding Property
  4.4 pH Partition (p$K_a$)
  4.5 Solubility
5 Biopharmaceutics Classification System (BCS) in Drug Product Optimization
6 Experimental Models for the Evaluation of Drug Permeation
  6.1 Predicting Drug Permeability from Physicochemical Parameters
  6.2 Cell Culture Based Models for Permeability Study
  6.3 Parallel Artificial Membrane Permeability Assay (PAMPA)
  6.4 Tissue Based Models for Permeability Study
  6.5 In Situ Model of Intestinal Drug Permeability
  6.6 Whole Animal Models
7 *In Vitro* and *In Vivo* (IVIV) Correlation
8 Conclusion
  References

## 1 INTRODUCTION

Drug discovery is a lengthy and costly process with a very low probability of success. Approximately 50% of the investigational new drugs (INDs) fail in preclinical and

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

clinical phases of the drug development process. The major reasons for such a high proportion of failures are poor biopharmaceutical properties, including low aqueous solubility, inadequate intestinal permeability, chemical instability, intestinal or hepatic metabolism, and systemic clearance. The advent of combinatorial chemistry and high throughput screening (HTS) has allowed the synthesis of thousands of compounds a day, leading to the rapid identification of therapeutically active compounds (i.e., hits and leads). Despite this dramatic increase in the speed of synthesis and screening, the number of drugs reaching the market has not increased in similar proportions.

However, the success rate is beginning to rise with the increase in availability of many *in vitro* screening techniques, which can be used for early prediction of new chemical entity (NCE) permeability through biological barriers and other physicochemical properties. It helps medicinal chemists to design compounds that exhibit the right pharmaceutical characteristics and avoid wasting valuable resources on developing molecules that are not likely to become successful drugs.

Drug permeability is considered an important parameter in drug discovery and lead optimisation; it is essential to have reliable methods of predicting the *in vivo* permeability by thoughtful use of *in vitro* permeability models.

## 2 NATURE AND PERMEABILITY OF SOME PHYSIOLOGICAL BARRIERS

The main physiological permeation barriers to be crossed by drugs are epithelia and endothelia. Epithelia cover the surface of the body and line various cavities. Endothelia line the blood capillaries so as to regulate the distribution of compounds between the blood and the interstitial fluids. Despite their extensive biochemical differences, they serve as highly selective permeability barriers, separating internal and external environments. Oral bioavailability is a highly desirable property for molecules under investigation in drug discovery, because approximately 90% of all marketed drugs are administered orally. The principal physiological barrier that drugs have to cross to enter the systemic circulation is the gastrointestinal mucosa. Alternatively, drugs can also be absorbed through the skin, cornea, and buccal, sublingual, nasal, vaginal, or rectal mucosa. Some of these routes have been explored as alternative paths for the delivery of drugs that are inactivated by first-pass metabolism. Another important physiological barrier is the blood–brain barrier (BBB), which separates the blood from the central nervous system. Due the difference in anatomical structure of the epithelia throughout the body, different drug application routes are impeded by different barriers.

## 3 DRUG TRANSFER PROCESS THROUGH A PHYSIOLOGICAL MEMBRANE

According to the *fluid mosaic model*, the structure of the cellular membrane is described as an interrupted phospholipid bilayer capable of both hydrophilic and hydrophobic interactions [1]. The membrane model for the understanding of relevant drug transport processes across the biological membranes is shown in Fig. 1.

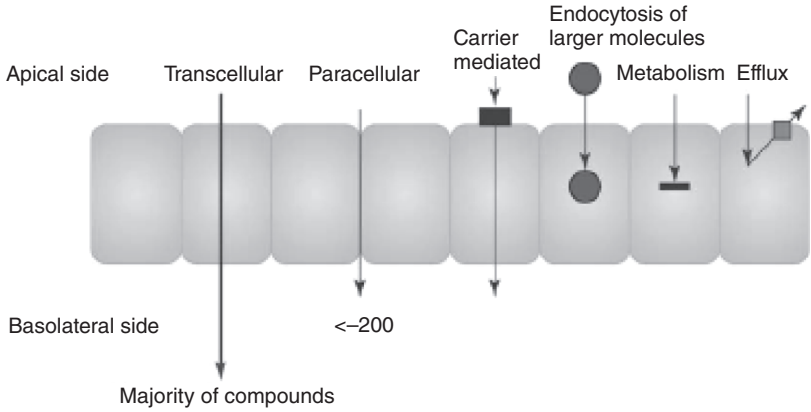


**FIGURE 1** Different routes of drug entry from the intestine into the bloodstream. (From Ref. 52, with permission from Elsevier.)

1. The majority of the lipophilic drugs are absorbed via passive diffusion across the cellular membrane—transcellular transport [2].
2. Transcellular passage is the main route for both intestinal epithelium and BBB.
3. The hydrophilic molecules (molecular weight or MW < 200) passes through the water-filled tight junctions formed by fusion of adjacent cells—paracellular transport [3].
4. The MW cutoff for paracellular transport may be 400–500 [4]. However, paracellular permeation is negligible in brain endothelium.
5. Drug molecules transported against concentration gradient by active energy consuming transporter proteins, P-glycoprotein (P-gp)—carrier mediated transport [5].
6. Some carriers can permeate in the direction of the concentration gradient without consumption of energy—facilitated transport.
7. The high MW compounds (e.g., peptides and proteins) usually take another transcellular route—endocytosis or transcytosis [6].

## 4 PHYSICOCHEMICAL PROPERTIES DETERMINING MEMBRANE PERMEATION

The tenets of molecular properties influence drug permeability through various physiological membranes. The relevant physicochemical descriptors for membrane permeability, and their interrelationships, are presented in Fig. 2 [7].

### 4.1 Lipophilicity

Because of the lipid nature of cell membranes, a molecule's lipophilicity has long been considered an important factor in drug design. The most common expression of lipophilicity is the logarithm of the partition coefficient (log $P$). Since the

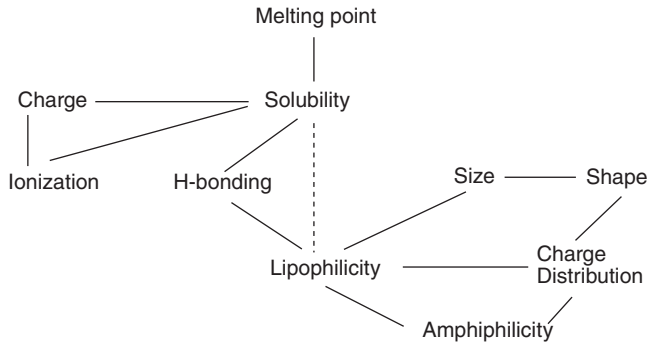


**FIGURE 2** Physicochemical properties influencing drug permeability. (From van de Waterbeemd H. *Eur J Pharm Sci* 1993; 7, with permission from Elsevier.)

membranes are primarily lipophilic in nature and with the lipid domain playing an important role in their barrier function, the relationship between log *P* and permeability is nonlinear, with a decrease in permeability at both low and high log *P* values. These nonlinearities are theorized to be due to the limited diffusion of poorly lipophilic molecules into the lipophilic cell membrane, or the preferential partitioning of highly lipophilic molecules into the phospholipid cell membrane, preventing passage through the aqueous portion of the membrane [8, 9].

### 4.2 Molecular Size

Molecular size is believed to play a distinct role in permeation processes and can be a further limiting factor in oral drug absorption. The paracellular transport highly depends on the molecular size due to the *sieving effect*, the molecular weight limit for the paracellular permeation being approximately 400–500 [4]. The bigger the molecule is, the harder it is to diffuse. Diffusion coefficients across biological membranes have been shown to be highly dependent on molecular mass. It has been concluded that the permeation of drugs with MW < 300 Da is not significantly influenced by the physicochemical properties of the drug, which will mostly permeate through aqueous channels of the membrane. By contrast, the rate of permeation is highly sensitive to molecular size for compounds with MW > 300 Da. The Lipinski rule of five proposes an upper MW limit of 500 as being the limit for orally absorbed compounds [10].

### 4.3 Hydrogen Bonding Property

The hydrogen bonding ability (an estimate of hydrophilicity) of a molecule is an important property for cellular membrane permeability [11]. Conradi et al. [3] believe that the relative contributions of both lipophilicity and hydrogen bonding must be considered, as opposed to hydrophobicity alone. Lipids within membranes contain hydrophilic parts that have hydrogen bonding acceptor groups. These groups bond to the hydrogen bond donating solutes, preventing the solutes from penetrating the membranes and slowing down the diffusion process [12].

### 4.4 pH Partition (p$K_a$)

The majority of drugs in use are either weak acids or bases and, depending on the pH, exist in an ionized or un-ionized form. Membranes are more permeable to un-ionized forms of drugs than to ionized species due to the greater lipid solubility of the un-ionized forms and the highly charged nature of the cell membranes. However, Palm et al. [13] reported that the molecules with an un-ionized fraction of <10%, a common state for many drugs in the intestinal pH range, were permeable across Caco-2 cell membranes. As discussed by these authors, a variety of examples exist where ionized molecules are permeable, deviating from classical pH partition theory.

### 4.5 Solubility

Solubility is a major factor in determining drug absorption through physiological membranes. Poor aqueous solubility is likely to result in poor absorption, since the drug diffusing through intestinal membrane is proportional to its concentration gradient between the lumen of the intestine and blood. Therefore, even drugs with high permeation rate show a low absorption. Conversely, the highly aqueous soluble drugs are well absorbed despite low or moderate permeation.

## 5 BIOPHARMACEUTICS CLASSIFICATION SYSTEM (BCS) IN DRUG PRODUCT OPTIMIZATION

The biopharmaceutics characteristics like solubility and permeability are vital in drug discovery and lead optimization due to the dependence of drug absorption on these two properties. The BCS is derived based on these two main properties [14–17]. By knowing a drug solubility and intestinal permeability, the BCS classifies drugs into one of the four categories (Fig. 3).

Class I drugs are highly soluble and highly permeable. Drug discovery programs are targeted to achieve Class I drugs, which are ideal candidates for oral delivery. Class II drugs are highly permeable across the gut by virtue of their high lipophilic-

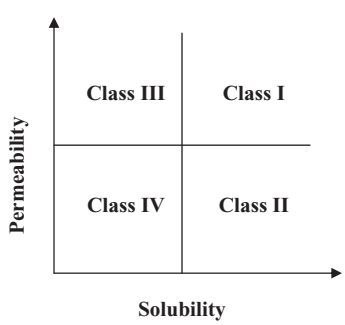


**FIGURE 3** Model depicting BCS. Drugs with high solubility and high permeability are grouped into Class I; low solubility and high permeability into Class II; high solubility and low permeability into Class III; and low solubility and low permeability into Class IV.

ity. The bioavailability of products containing these compounds is likely to be dissolution rate limited due to low solubility. However, as drug dissolution is the rate-limiting step, the *in vitro* permeability data obtained from different experimental models may not accurately predict the *in vivo* absorption. Class III drugs are highly soluble but have less intrinsic permeability due to their physicochemical properties. For this reason dissolution is very rapid but absorption is permeability rate limited. Class IV drugs exhibit low solubility and low permeability. Low and variable oral bioavailability for these drugs is anticipated because of the combined limitation of solubility and permeability. Strategies to improve both solubility and permeability should be established for these molecules.

The BCS is considerated by the U.S. Food and Drug Administration (FDA) as a fundamental principle to define the rate and extent of absorption of a drug product. It has published several guidance based on the BCS covering scale-up and postapproval changes [18] and has allowed waiver of *in vivo* bioequivalence studies [19]. However, accurate use of permeability methods to measure the drug permeability is a critical component for this classification. The scientific aspects of the BCS have been extensively discussed and have been the subject of numerous studies [20, 21].

## 6 EXPERIMENTAL MODELS FOR THE EVALUATION OF DRUG PERMEATION

The need to screen drugs for their absorption characteristics has increased significantly in recent years due to use of techniques like combinatorial chemistry. An ideal screening tool should be fast and easy to use, require small amounts of the compound, be relatively inexpensive, and give reliable predictions. In practice, such ideal testing methods are not easy to develop. While evaluating the usefulness of these different methods in drug discovery, it is important to know exactly what each technique measures and how this can help identify potentially successful or unsuccessful compounds. In the following section, we look at key techniques for evaluating and predicting the drug permeability and critically assess their advantages and limitations in drug discovery and development.

### 6.1 Predicting Drug Permeability from Physicochemical Parameters

During drug development, laboratories might use computational approaches to predict the absorption potential of new chemical entities. These computational methods are based on lipophilicity, hydrogen bonding, and molecular size. It has long been recognized that these physicochemical descriptors are related to membrane permeability [22], although this relationship is often obscured when structurally different compounds are studied [23].

***Lipophilicity: log* P *and log* D** Drug lipophilicity is commonly used as a predictor of membrane permeability because transcellular permeation is dependent on the lipophilicity of the compound. The molecule requires a certain affinity for the phospholipid structure in order to enter the cell membrane [24]. Fick's first law of diffusion describes the passive transport of the drug across the membrane as being

**TABLE 1 Computer Programs for Log *P* Calculations**

| Computer Program | Software Package/Vendor |
|---|---|
| PrologP and Prolog P | PALLAS/CompuDrug Chemistry Ltd. |
| ACD/logP and logD | Version 3.0/Advanced Chemistry Development Inc. |
| ALOGP | Tsar 3.1/Oxford Molecular Group |
| CLIP/logP | University of Lausanne |
| CLOGP | Pcmodels/Daylight CIS, CLOGP/Biobyte |
| HINT/logP | Edusoft |
| KLOGP | Multicase Inc. |
| LISP | O. Raevsky |
| LOGKOW | Syracuse Research Corp. |
| SCILOGP | SCILOGP/Scivision |
| TLOGP | TLOGP 1.0/Upstream Solutions |
| XLOGP | XLOGP V 2.0/Institute of Physical Chemistry, Peking University |
| ALOGPS | ALOGPS 2.1/Virtual Computational Chemistry Laboratory |
| MLOGP | DRAGON 3.0 |

proportional to the membrane–water partition coefficient. These coefficients are conventionally expressed as the *n*-octanol–water partition coefficient (log *P*), which is the concentration ratio of the compound between *n*-octanol and aqueous phase at equilibrium. The distribution coefficient *D* (log *D*) is pH dependent in the case of ionizable compounds.

The log *P* and log *D* are traditionally determined by the shake flask method using an *n*-octanol–water biphasic system. This has been one of the most suitable models of the lipidic biological membranes, due to the resemblance of *n*-octanol to lipids—with its long alkyl chain and the polar hydroxyl group. However, the log *P* and log *D* values for a single compound can vary as a result of different experimental conditions [25]. Both log *P* and log *D* are used for prediction of drug absorption, which can subsequently be measured by more sophisticated and standardized techniques like potentiometer titration of the compound in the biphasic system [26].

In the drug development program, an early estimate of log *P* and log *D* values can be determined by computational approaches. Many computational approaches have been developed to estimate log *P* [27, 28], including the popular *C* log *P* [29] and *A* log *P* [30] algorithms (Table 1). These approaches can calculate log *P* values using software programs and are based on data for either atomic contributions or molecular fragments, or on molecular properties such as the molecular lipophilicity potential (MLP) [31], molar volume, and hydrogen bonding.

***Liposomal Partitioning*** Liposomes are lipid bilayer vesicles that can be used as membrane models in partition studies [32] with the potentiometric titration method [33]. The phosphatidylcholine liposomes seem to provide a more accurate partition system for absorption prediction than the *n*-octanol–buffer system. Beigi et al. [34] demonstrated a chromatographic technique for partition studies in liposomal systems. Liposomes of different lipid composition were immobilized in small agarose–dextran gel beads and packed into chromatography columns. The capacity factor of 12 compounds related well to their fraction absorbed in humans.

***Chromatographic Methods*** The elution times in chromatography represent the partitioning behavior of the compound between stationary and mobile phases. Chong et al. [35] introduced immobilized artificial membrane (IAM) columns containing phosphatidylcholine or similar lipid-like ligands covalently bonded to silica particles. This lipid monolayer is similar to the lipid membrane in partition studies. This high performance liquid chromatography (HPLC) technique can be used to study the partition of a solute. In spite of the end capping, the solutes also interact with uncapped silanol groups, in addition to the lipid ligands. Once this problem is solved, IAM columns are likely to provide a good tool for absorption prediction.

***Hydrogen Bonding*** Hydrogen bonding can be measured experimentally by allowing a compound to distribute between two solvents, of which only one can form hydrogen bonds. This is expressed as $\Delta \log P$, which is the difference between the *n*-octanol–water $\log P$ and an alkane–water $\log P$ [36], where *n*-octanol is the solvent that permits hydrogen bonding. $\Delta \log P$ has shown good correlation with membrane permeability in numerous studies. El Tayar et al. [37] found correlation of $\Delta \log P$ with skin absorption. $\Delta \log P$ was also compared with brain permeability. Optimal permeability was predicted for $\log D$ between 0 and 3 and $\Delta \log P < 2$ [38]. The main shortcoming of the $\Delta \log P$ method is that it requires two experimental determinations for each compound, since *n*-octanol is miscible with alkane and it is not possible to measure $\Delta \log P$ directly.

For reasonable prediction of membrane permeability, the hydrogen bonding capacity is calculated by counting the number of hydrogen bond donor atoms (hydrogens) attached to oxygen and nitrogen and hydrogen bond acceptor atoms (oxygens and nitrogens) [10], or counting the number of lone pair electrons [39]. Recently, more accurate methods of calculating hydrogen bonding capacity have been introduced. These include MolSurf [40], which calculates hydrogen bonding properties based on quantum mechanics, and Hybot [41], which is based on experimental results for the formation of complexes between hydrogen bond donors and acceptors.

***Molecular Size*** As a molecular size descriptor, often the molecular weight is used in predicting the drug permeability. The paracellular permeability of drug compounds was successfully modeled based on MW as discussed earlier. Molecular size is not only a component of lipophilicity, but also of the diffusion coefficient ($\log D$) in physiological membranes. A rather strong dependence on molecular size has been observed for transcellular diffusion in physiological membranes. Camenisch et al. [42] reported that the physiological membrane permeability of high MW compounds is overestimated when using $\log D$ as a predictor. This effect is important for compounds with MW > 500. For compounds that have similar $\log D$ values, Caco-2 permeabilities were lower for high MW compounds as compared with compounds of lower MW. Camenisch et al. [2] reviewed suggestions for the correction of partition coefficients with MW functions that were made by several authors. In the case of BBB penetration, a correlation between logarithm of the BBB permeation and logarithm of MW also existed [43].

***Polar Surface Area (PSA)*** The polar surface area (PSA) is the molecular surface area associated with hydrogen bonding acceptor atoms (i.e., oxygen and nitrogen)

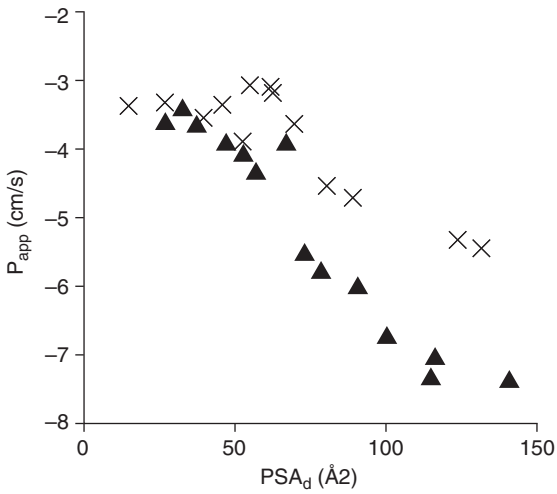


**FIGURE 4** Relationship between dynamic polar surface area and drug permeability in perfused human jejunum (×) and in Caco-2 cell monolayers (▲). (From Ref. 46, with permission from Elsevier.)

plus the area of the hydrogen atoms attached to these heteroatoms—which are known as hydrogen donors. This model has been introduced to account for the neglected intramolecular hydrogen bond formation [44]. The dynamic polar surface area (PSAd) is the statistical average of the PSA of the low energy conformations of a molecule. The areas were calculated for an aqueous environment using atomic van der Waals radii. The smaller the PSAd of a compound, the better it permeates the biological barrier. Good inverse linear correlation was observed between the permeability across a Caco-2 model and the water-accessible PSA of six β-blockers and their ester prodrugs. Figure 4 shows the PSAd correlated to permeability in the human jejunum for 13 structurally different drugs investigated by Winiwarter et al. [45]. A comparison is also seen from the Caco-2 cell monolayer permeabilities for 14 model compounds presented by Stenberg et al. [46].

The long calculation time for the low energy conformations of large and flexible molecules is a limitation with this model. Stenberg et al. [46] noted that the relationship may break if the data for drugs that undergo intestinal metabolism or active transport, or have solubility problems, is included in the calculations. Therefore, integration of all these events into a computational model is vital.

***Lipinski's Rule of Five Approach*** In a common computational approach for absorption and permeability, Lipinski et al. [10] introduced the simple "rule of five" approach for the quick evaluation of new chemical entities. This model states that absorption or permeation of a compound is more likely to be poor when the calculated MW > 500, log $P$ > 5, hydrogen bond donors > 5, and hydrogen bond acceptors > 10. The numbers of hydrogen bond donors and acceptors are simply calculated by the counting method described. If any two of these criteria are met, poor absorption or permeability is to be anticipated. This rule is based on a variety of calculated properties among several thousand drugs belonging to different therapeutic

categories; therefore, certain drugs falling outside the rule, for example, orally active drugs like antibiotics, antifungals, vitamins, and cardiac glycosides, violate the rule of five. These drugs have structural similarities that allow them to act as substrates for biological transporters; therefore, they are outside the rule of five approach.

## 6.2 Cell Culture Based Models for Permeability Study

***Caco-2 Cell Model*** The Caco-2 cell model is a well characterized and widely used *in vitro* model in the field of drug permeability studies [6, 47]. It has been considered the gold standard technique and has been used to standardize other absorption assessment methods [48, 49]. Caco-2 cells are grown on a filter support in a multiple well format and permeability is measured by the movement of molecules from one side of the cell monolayer to the other (Fig. 5).

Caco-2 cells derived from a human colon adenocarcinoma exhibit many *in vivo* intestinal cell characteristics by having tight intercellular junctions and microvilli and expressing intestinal enzymes and transporters (e.g., P-gp). Due to these similarities, the permeation characteristics of drugs across Caco-2 cell lines correlate with their human intestinal permeation characteristics; therefore, it has been recommended that the Caco-2 model can be used to predict the absorption of drugs in humans. An essential component of studies using Caco-2 cells is the culture of intact monolayer on permeable supports and the assessment of their integrity or cell damage either by measuring transepithelial electrical resistance or by using a paracellular flux marker such as mannitol [49]. Since this model is quicker to use, more convenient, and produces more reproducible data than animal studies, its use in the early stages of drug development can lead to savings in time and money and avoid the unnecessary use of experimental animals. Moreover, Caco-2 assays are relatively accessible and easier for higher throughput than the excised tissue assays.

Many drugs are absorbed by passive diffusion; therefore, physiochemical descriptors of the drugs are expected to play a vital role in this process. Caco-2 cell permeability data have frequently been plotted against log $P$ or log $D$, molecular weight, or some descriptor for H bonding [44]. The sigmoidal relationship has been identified between the hydrogen bonding properties of the solutes and permeability coefficients through Caco-2 cell monolayers [50].

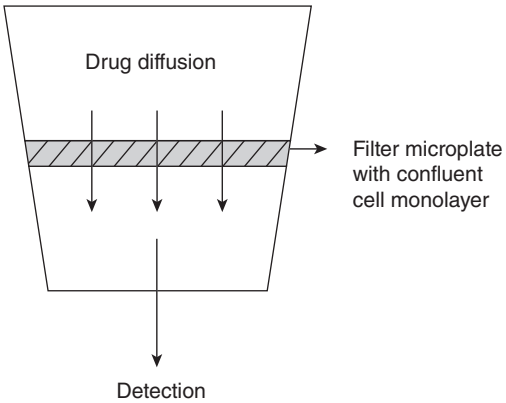


**FIGURE 5** Caco-2 cell line model.

Although, Caco-2 cells can give evidence that a compound will be well absorbed, it is difficult to predict the compounds showing low or modest permeability. Other complications with Caco-2 cells include (1) variable expression of P-gp, metabolizing enzymes, and active transport systems [35, 51]; (2) interexperimental and interlaboratory variability, such as genetic change and sensitivity to differing culture conditions and protocols [49]; and (3) Caco-2 culture is laborious and costly—around 3–21 days are required to generate a confluent layer of cells [52].

***Madin–Darby Canine Kidney (MDCK) Cell Model*** Apart from Caco-2 cells, the other models that are most frequently used for permeability studies are MDCK and LLC-PK1 cells. When the MDCK cells are cultured under standard conditions, they differentiate into polarized columnar epithelial cells and form tight cellular junctions. The main advantage of MDCK cells is shorter culture times, which can be equal to 24 hours. A good correlation was reported between permeation of passively absorbed drugs in Caco-2 and MDCK cells [53]. The permeability coefficients of hydrophilic compounds are usually lower in Caco-2 cells than in MDCK cells. Whereas Caco-2 cells originate from human colon adenocarcinoma cells, MDCK cells are from dog kidney cells, and thus the expression levels of intestinal transporters would be different in these two cell lines [54]. Hence, the MDCK model has to be extensively characterized to confirm the correlation of transport mediated drug permeability to human absorption.

***Lewis Lung Carcinoma–Porcine Kidney (LLC-PK1) Cell Model*** LLC-PK1 cells are derived from normal pig kidneys. When cultured, these cells rapidly develop into a well formed monolayer with bush boarders and microvilli on their apical cell surface [55]. Several investigators have reported the use of these cells in passive cellular diffusion [56, 57]. As with MDCK cells, these cells also possess different transporters on their surface. A study has shown that uracil was transported by sodium-dependent and -independent pathways in LLC-PK1 cells; on the other hand, in Caco-2 cells uracil was transported independently of sodium pathway [18]. These differences in the transport mechanism can be attributed to differences in the origin of the cells types and the transporters that are expressed on these cells. Thus, further studies are warranted to determine the correlation between the LLC-PK1 cells and *in vivo* human absorption.

***HT29 Cell Model*** The role of mucus on drug permeability is largely ignored, since Caco-2 cells are a widely used model in permeability studies and they are devoid of this property. Certain clones of HT29 cells produce mucus. The wild-type HT29 cells grown on media containing galactose lead to the selection of a subclone of HT29 cells that form polarized cell monolayers and secrete mucus. A study has shown that the presence of the mucus layer in HT29-H resulted in lower permeability coefficients than Caco-2 monolayers in a testosterone permeability study. This indicates that the mucus layer accounts for most of the permeability resistance to testosterone. An attempt has been made to coculture the Caco-2 cells with HT29-MTX or HT29-H cells [58] in order to provide representative characteristics of intestinal mucosa.

The major application of the cell culture based model is screening of chemical libraries for compounds that have favorable permeability characteristics. Caco-2

**TABLE 2 Cell Culture Models Used for Permeability Assessment**

| Cells | Origin | Special Characteristics |
|---|---|---|
| Caco-2 | Human colon adenocarcinoma | Well characterized and widely used Express some relevant efflux transporters; expression of influx transporters is variable |
| MDCK | Dog kidney epithelial cells | Polarized cells with low expression of ABC transporters, ideal for transfection |
| LLC-PK1 | Pig kidney epithelial cells | Polarized cells with low intrinsic transport expression, ideal for transfection |
| HT-29 | Human colon | Contains mucus-producing goblet cells |
| TC-7 | Caco-2 subclone | Similar to Caco-2 |
| IEC-18 | Rat small intestine cells | Provides a size selective barrier for paracellularly transported compounds |

cells grow very slowly in 3–21 days; this time is too long for the drug discovery industry. Other cell models (MDCK and LLC-PK1) can grow much faster and be available for studies in 5–7 days. However, recent improvements in culture media like the addition of 10% fetal bovine serum, manipulation of the seeding density, and use of BIOCOT® transwell plates can reduce the time required for cell growth. Various cell culture models used for permeability assessment are described in Table 2.

***Data Analysis from the Cell Cultured Models*** The permeability coefficients ($P_{eff}$) through the cell monolayers are calculated from the following equation [48, 59]:

$$P_{eff} = \frac{V_A}{A(C_D - C_A)} \times \frac{dC_A}{dt} \tag{1}$$

where

- $dC_A/dt$ (mg/s·mL) is the increase of drug concentration in the receiver chamber over the time period considered.
- $A$ (cm$^2$) is the membrane surface area exposed to the compound.
- $V_A$ (mL) is the solvent volume in the acceptor chamber.
- $C_A$ and $C_D$ (mg/mL) are the initial drug concentrations in the receiver and donor chambers, respectively.

Permeability rate ($dC_A/dt$) is calculated by plotting the amounts of drug measured over the linear diffusion range and determining the slope of the plot. $P_{eff}$ from the equation is only valid under the experimental conditions, where a constant concentration gradient exists, so that backdiffusion is avoided. Under that condition, $C_D$ is almost constant and $C_A$ is negligible compared to $C_D$. In several findings, experimental conditions are usually accepted as sufficiently accurate for $P_{eff}$ calculation, if the concentration difference between the donor and acceptor compartments does not diverge by more than 10% in the time interval studied [55].

### 6.3 Parallel Artificial Membrane Permeability Assay (PAMPA)

An alternative to the cell cultured models is the parallel artificial membrane permeability assay (PAMPA), which is a method for predicting passive permeability [60]. Recently, it has been suggested that artificial membrane methods such as PAMPA could soon replace high throughput screening Caco-2 assays [61], because PAMPA is an excellent biomimetic model, in terms of high throughput, reproducibility, and especially cost.

PAMPA is performed in a 96 micotiter well plate containing two parts. All the wells at the bottom part are filled with buffer solution. The top part contains a series of filter-immobilized artificial membranes composed of lipids, which match with the wells in the lower part. One-half of the filters on the top part are treated with an organic solvent, which supposedly acts as the cell membrane, and the other half are wetted with methanol/buffer. The test compound under investigation is applied to the top filters and the rate of appearance of the compound in the bottom wells should reflect the diffusion across the lipid layer. In a recent study by Lipinski et al. [62], a good correlation was found between diffusion in the PAMPA system and percentage absorption in humans for a selected series of compounds.

The filter immobilized artificial membranes are prepared by dispersing a phospholipid with the help of an organic solvent on a filter [60]. The various PAMPA models reported in the literature are presented in Table 3.

The problem of precipitation with low solubility compounds—when working at clinically relevant doses in cellular assays—is prevented in recent PAMPA models [63]. The limitations of the system include the lack of influx and efflux transporters, enzymes, and paracellular pathways.

### 6.4 Tissue Based Models for Permeability Study

Excised intestinal segments have been used to study intestinal drug absorption. The compound under investigation in a solution is applied to one side (mucosa or serosa) of the mucosa and the rate of drug absorption is determined by estimating

**TABLE 3 Membrane Compositions of PAMPA Models**

| Assay | Model | Phospholipid Constituents[a] | Organic Solvent |
|---|---|---|---|
| Egg-PAMPA | Lecithin | 10% Egg lecithin, cholesterol | *n*-Dodecane |
| BM-PAMPA | Biomimetic | 0.8% PC, 0.8% PE, 0.2% PS, 0.2% PI, 1% cholesterol | 1,7-Octadiene |
| HDM-PAMPA | Hexadecane | 100% *n*-Hexadecane | *n*-hexadecane |
| DOPC-PAMPA | Synthetic phospholipid | 2% Dioleylphosphatidyl choline | *n*-Dodecane |
| BBB-PAMPA | Blood–brain barrier | 2% Porcine brain lipid extract (PC, PE, PS, PI, PA, cerebrosides) | *n*-Dodecane |
| DS-PAMPA | Double sink | 20% Phospholipid mixture (PC, PE, PI, PA, triglycerides) | *n*-Dodecane |

[a]PC, phosphatidyl choline; PE, phosphatidyl ethanolamine; PS, phosphatidyl serine; PI, phosphatidylinositol; PA, phosphatidic acid.

either the disappearance of drug from the drug solution or appearance of drug on the serosal side. These models preserve the characteristics of the biological membranes and also help in determining absorption across different gastrointestinal segments. Limited viability and versatility are the limitations with this type of preparation.

***Everted Intestinal Segments*** Isolated intestinal segment is a simple technique used to measure the intestinal transport of a drug. In this model, the permeability assessment is carried out through a segment of the intestine, with the musculature intact. Commonly, rat intestine is used, although absorption in other animal models, such as monkey, also correlate well with human absorption [64]. To prepare the rings, a selected segment of the intestine is isolated immediately after euthanizing the animal, washed in ice-cold buffer to remove debris and digestive products, and tied at one end with a piece of suture. The closed end is carefully pushed through the intestine using a glass rod, resulting in eversion of the intestine, which is then cut into small rings, typically 2–4 mm wide [65]. Next, the rings are incubated in buffer at 37 °C containing the compound under investigation, and shaken well in a water bath. After a set time, the drug uptake is suppressed by rinsing the rings with ice-cold saline or buffer. The rings are then blotted dry, weighed, and then dissolved or processed for assay. Any segment of the intestine can be used to prepare a large number of rings. This model can be used to study both active and passive transport pathways especially that of amino acids and peptides [66]. A major disadvantage associated with this model is the lack of maintenance of viability in the isolated rings, due to absence of blood supply and nerves. Some other factors like drug metabolism and accumulation in the mucosa could lead to an overestimation of true drug absorption. For example, one such early study found that, for several beta-lactam antibiotics, absorption based on luminal disappearance was roughly twice the true amount transported across the tissue [67].

***Everted Intestinal Sacs*** In this method, a 2–4 cm section of the intestine is tied off at one end and everted similar to the intestinal ring method. The sac is filled with buffer, tied off at the other end, and placed in a flask with oxygenated (95% $O_2$/5% $CO_2$) buffer solution as a test compound. Drug absorption is measured by sampling the solution inside and outside the sac. Unlike the *in vivo* situation, everted intestinal sacs only include the mucosa plus underlying muscle layers, which could lead to biased absorption values. Although this method is relatively simple and allows several experiments to be performed using tissue from just one intestine, it differs from the *in vivo* situation and drug accumulation in the muscle layer can lead to poor recovery of the drug.

***Diffusion Cells*** Diffusion cells have been used to determine the transport of compound in living tissues, one of the first approaches being the so-called Ussing chamber. In this method, the long intestinal mucosal sheets are cut into mucosal strips of adequate size (~2 cm) and are clamped between two glass chambers filled with buffer. The compound under study is added to the donor compartment, and the accumulation of the compound at the other side of the membrane (acceptor side) is measured as a function of time. The permeability of the compound is calculated from the following equation:

$$P_{\text{eff}} = \frac{dc}{dt} \times \frac{V}{A \times C} \tag{2}$$

where $dc/dt$ is the change in concentration in the receptor compartment per unit time, $V$ is the receptor volume, and $A$ is the area available for diffusion. $P_{\text{eff}}$ is usually expressed in centimeters per second (cm/s). This method has been used to evaluate *in vitro* permeability with varying degrees of success. For example, intestinal mucosa of rat mounted in an Ussing chamber was useful in describing the regional variability in GIT absorption of a mixed series of compounds [68].

### 6.5 *In Situ* Model of Intestinal Drug Permeability

***Single-Pass Perfusion Model*** Several different models of the single-pass perfusion system have been used [69–71]. In these *in situ* models the viability of the tissue is maintained, as the whole animal is used and the blood vessels and lymphatics are intact. This technique has been widely used to study both active and passive transport pathways [69, 71, 72]. In addition, the effect of various factors such as drug concentration, intestinal region, and flow rate can be studied. In the closed loop model, a solution containing the diluted drug is added to a segment of the intestine, and the intestine is closed. After a set time the intestine is excised and the drug content is analyzed [73]. Conversely, in the open loop model, the proximal and the distal end of a segment of the intestine are canulated with a glass tube. The drug is then pumped through the intestine and the ratio of drug in and out is measured.

The steady-state effective intestinal permeability coefficient ($P_{\text{eff}}$) is calculated according to the parallel-tube model [74]:

$$P_{\text{eff}} = \frac{-Q_{\text{in}} In(C_{\text{out}}/C_{\text{in}})}{2\pi r L} \tag{3}$$

where $Q_{\text{in}}$ is the perfusion flow rate through the intestine segment (mL/s), and $C_{\text{in}}$ and $C_{\text{out}}$ correspond to the inlet and outlet concentrations (μg/mL), respectively. $r$ is the radius of the intestinal loop (cm) and $L$ is the length of the loop (cm). From the equation, it can be seen that by varying the length of the intestinal segment, the effect of transit time on drug absorption can also be studied [55, 66]. Even with these advantages, use of the single-pass perfusion model is limited because in this model the disappearance of the drug from the intestine is used to predict permeability. However, the rate of decrease in drug concentration does not always equal the rate of drug absorption; for example, drugs can interact with the lipid membrane [75]. Furthermore, the proximal and distal content of the lumen can enter the test segment of the intestine, making it difficult to control conditions such as flow rate [76].

Prediction of oral absorption in human beings is important in the early stage of the drug discovery process. *In vitro* assays in Caco-2 and MDCK models are frequently used for that purpose. However, an *in situ* technique might be more accurate in estimation of absorption. One such study [77] found that effective permeability coefficients determined in rats by single-pass intestinal perfusion for 14 compounds were correlated to the values obtained from humans.

***Single-Pass Perfusion of the Dog Intestine (Loc-I-Gut)*** The Loc-I-Gut is a very effective model for investigating regional jejunum permeability. It is a perfusion instrument consisting of six channels and two inflatable latex balloons, which are set 10 cm apart [78]. The purpose of the balloons is to isolate the segment of the intestine that is being studied and to prevent the leakage of the proximal and distal luminal content into the selected area. It has been reported that less than 2% of the luminal content leaked into the jejunum when the balloons were inflated [76]. The tube is inserted into the proximal jejunum by using fluoroscopic techniques. Air is then pumped into the balloons, inflating them and creating a 10 cm separated jejunal segment. Insertion and setting up of the tube usually takes about an hour and the infusion rate is about 2–3 mL/min. A nonabsorbable marker such as $^{14}$C-PEG 4000 is used in the perfusion solution to check the integrity of the balloons during the experiment [79]. In the dog model the drug permeability was calculated using the well-stirred model because it is assumed that the hydrodynamics are similar to the corresponding model in the human intestine. The $P_{\text{eff}}$ can be calculated as [76]

$$P_{\text{eff}} = \frac{Q_{\text{in}}(C_{\text{in}} - C_{\text{out}})/C_{\text{out}}}{2\pi r L} \tag{4}$$

where $Q_{\text{in}}$ is the perfusion flow rate, $r$ is the radius of the tube, $L$ is the length of the tube, $C_{\text{in}}$ is the infused drug concentration entering the tube, and $C_{\text{out}}$ is the diffusate concentration leaving the tube. The Loc-I-Gut model is also used for perfusion of suspended drug particles in human jejunum. In this model, the *in vivo* dissolution of the drug can be directly estimated by measuring the concentration of drug either in dissolved or in solid-state form [80]. The major limitation to this model is that it is invasive. Intubation may affect the normal physiology and function of the GI tract. Furthermore, the fluid that is used to flush out the drug from the tube may affect the absorption of the drug [81].

## 6.6 Whole Animal Models

***InteliSite Capsule*** The InteliSite® capsule (Innovative Devices LLC, Raleight, NC) is a rapid and noninvasive technique for measuring site-specific permeability of a drug [82]. The permeability is assessed through measurement of the drug in the systematic circulation. When the capsule reaches the target site it is activated. Activation is accomplished by placing the magnetic field generator next to the subject's abdomen. This magnetic field is sensed by a receiver and is converted to heat. The heat will cause the shape memory alloys to straighten. This in turn results in the rotation of the interior sleeve in relation to the outer sleeve and the drug is released. Gamma scintigraphy and pharmacokinetics analysis are used to evaluate and measure drug absorption [81]. Pithavala et al. [81] evaluated the pharmacokinetic properties of ranitidine in eight volunteer subjects using the InteliSite capsule [81]. The result of the study correlated well with data available in the literature. Thus, it can be concluded that these capsules are a valuable tool for studying regional drug absorption.

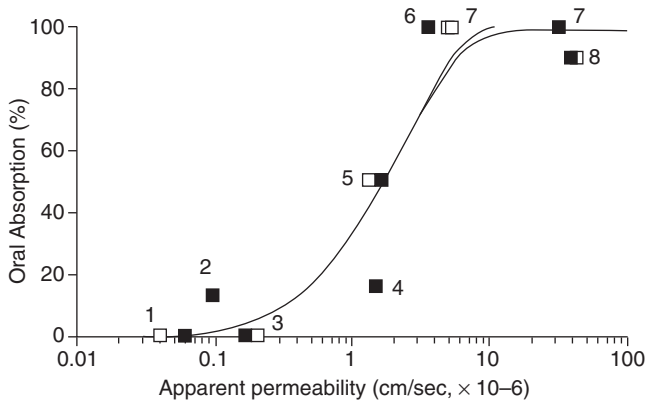


**FIGURE 6** Correlation between the permeability coefficient (in cell culture) and fraction absorbed in human (From Ref. 83, with permission from Elsevier.)

## 7 *IN VITRO* AND *IN VIVO* CORRELATION (IVIV)

In order to use any experimental model for human permeability prediction, the drug permeability from the model has to correlate well with human absorption. To validate a permeability model, compounds with known absorption in humans are used and $P_{eff}$ is calculated. Figure 6 [83] shows the general shape of the curve of fraction absorbed in human versus the *in vitro* $P_{eff}$. Any small change (such as the pH of the buffer or the intubation time) in the way that the techniques are carried out can effect the $P_{eff}$ of the drug, and the curve may shift to left or right or alter its steepness. Therefore, well designed and validated techniques are to be used within laboratories to correlate the data collected.

Furthermore, in the case of animal models, there needs to be a correlation between the animal's $P_{eff}$ and the fraction absorbed in human. It has been shown that rat and human jejunum $P_{eff}$ of passively absorbed drugs correlate well [84]. Once the correlation is demonstrated for a series of compounds, then the experimental results can be used to make intelligent decisions about those compounds that can be further studied in animals.

## 8 CONCLUSION

Identifying appropriate models for permeation study of a new molecule is a very challenging task, but there are several options available to the pharmaceutical scientist. Careful review and selection of appropriate methods is crucial to a successful drug development program. The key feature that governs the process is *in vitro* and *in vivo* correlation in permeability studies. There is a potential to develop models that can provide good correlation.

## REFERENCES

1. Singer SJ, Nicolson GL. The fluid mosaic model of the structure of cell membranes. *Science* 1972;175:720–731.

2. Camenisch G, Folkers G, van de Waterbeemd H. Review of theoretical passive drug absorption models: historical background, recent development and limitations. *Pharm Acta Helv* 1996;71:309–327.
3. Conradi RA, Burton PS, Borchardt RT. Physico-chemical and biological factors that influence a drug's cellular permeability by passive diffusion. *Methods Principles Med Chem* 1996;4:233–252.
4. Artursson P, Ungell A-L, Lofroth J-E. Selective paracellular permeability in two models of intestinal absorption: cultured monolayers of human intestinal epithelial cells and rat intestinal segments. *Pharm Res* 1993;10:1123–1129.
5. Hunter J, Hirst BH, Simmons NL. Drug absorption limited by *p*-glycoprotein-mediated secretory drug transport in human intestinal epithelial Caco-2 cell layers. *Pharm Res* 1993;10:743–749.
6. Artursson P, Palm K, Luthman K. Caco-2 monolayers in experimental and theoretical predictions of drug transport. *Adv Drug Deliv Rev* 1996;22:67–84.
7. van de Waterbeemd H. The fundamental variables of the biopharmaceutics classification system (BCS): a commentary. *Eur J Pharm Sci* 1998;7:1–3.
8. Hansch C, Clayton JM. Lipophilic character and biological activity of drugs. II. Parabolic case. *J Pharm Sci* 1973;1–21.
9. Kararli TT. Gastrointestinal absoption of drugs. *Crit Rev Ther Drug Carrier Syst* 1989;6:39–86.
10. Lipinski CA, Lombardo F, Dominyl BW, Feeney PJ. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv Drug Deliv Rev* 1997;23:3–26.
11. Diamond JM, Wright EM. Molecular forces governing non-electrolyte permeation through cell membranes. *Proc R Soc B Biol Sci* 1969;172:273–316.
12. Urtti ALS. Minimizing systemic absorption of topically administered ophthalmic drugs. *Surv Ophthalmol* 1993;37:435–456.
13. Palm K, Luthman K, Ros J, Grasjo J, Artursson P. Effect of molecular charge on intestinal epithelial drug transport: pH-dependent transport of cationic drugs. *J Pharmacol Exp Ther* 1999;291:435–443.
14. Dressman JB, Amidon GL, Fleisher D. Absorption potential: estimating the fraction absorbed for orally administered compounds. *J Pharm Sci* 1985;74:588–589.
15. Gordon LA, Hans L, Vinod PS, John RC. A theoretical basis for a biopharmaceutic drug classification: the correlation of *in vitro* drug product dissolution and *in vivo* bioavailability. *Pharm Res* 1995;12:413–420.
16. Sinko PJ, Leesman GD, Amidon GL. Predicting fraction dose absorbed in human using a macroscopic mass balance approach. *Pharm Res* 1991;8:979–988.
17. Symillides PEMMY. Toward a quantitative approach for the prediction of the fraction of dose absorbed using the absoption potential concept. *Biopharm Drug Dispos* 1989;10:43–53.
18. CDER/FDA. *Guidance for Industry*. Immediate release solid oral dosage forms: scale-up and post-approval changes. Washington DC: FDA;1995.
19. CDER/FDA. *Guidance for Industry*. Waiver of *in vivo* bioavailability and bioequivalence studies for immediate release solid oral dosage forms based on a biopharmaceutics classification system. Washington DC: FDA;2000.
20. Yu LX, Amidon GL, Polli JE, Zhao H, Mehta MU, Conner DP, Shah VP, Lesko LJ, Chen ML, Lee VHL, Hussain AS. Biopharmaceutics classification system: the scientific basis for biowaiver extensions. *Pharm Res* 2002;19.

21. Blume HH, Schug BS. The biopharmaceutics classification system (BCS): Class III drugs—better candidates for BA/BE waiver? *Eur J Pharm Sci* 1999;9:117–121.
22. Hans C, Steward AR, Iwasa J. The correlation of localization rates of benzeneboronic acids in brain and tumor tissue with substituent constants. *Mol Pharmacol* 1965;1:87–92.
23. Artursson P, Karlsson J. Correlation between oral drug absorption in humans and apparent drug permeability coefficients in human intestinal epithelial (Caco-2) cells. *Biochem Biophys Res Commun* 1991;175:880–885.
24. Kramer SD. Absorption prediction from physicochemical parameters. *Pharm Sci Technol Today* 1999;2:373–380.
25. Hansch C, Leo A. *Substituent Constants for Correlation Analysis in Chemistry and Biology*. Hoboken, NJ: Wiley-Interscience;1979.
26. Avdeef A. pH-metric log *P*. II: Refinement of partition coefficients and ionization constants of multiprotic substances. *J Pharm Sci* 1993;82:183–190.
27. Carrupt P, Testa B, Gaillard P. *Computational Approaches to Lipophilicity: Methods and Applications*. Hoboken, NJ: Wiley-VCH;1997.
28. Mannhold R, Cruciani G, Dross K, Rekker R. Multivariate analysis of experimental and computational descriptors of molecular lipophilicity. *J Comput Aided Mol Des* 1998;12:573–581.
29. Leo AJ, Hoekman D. Calculating *P*(oct) with no missing fragments; the problem of estimating new interaction parameters. *Perspect Drug Discov Des* 2000;18:19–38.
30. Ghose AK, Viswanadhan VN, Wenoloski JJ. Prediction of hydrophobic (lipophilic) properties of small organic molecules using fragmental methods: an analysis of ALOGP and CLOGP methods. *J Phys Chem A* 1998;102:3762–3772.
31. Testa B, Carrupt PA, Gaillard P, Billois F, Weber P. Lipophilicity in molecular modelling. *Pharm Res* 1996;13:335–343.
32. Balon K, Riebesehl BU, Muller BW. Drug liposome partitioning as a tool for the prediction of human passive intestinal absorption. *Pharm Res* 1999;16:882–888.
33. Avdeef A, Box KJ, Comer JEA, Hibbert C, Tam KY. pH-Metric log *P* 10. Determination of liposomal membrane–water partition coefficients of ionizable drugs. *Pharm Res* 1998;15:209–215.
34. Beigi F, Gottschalk I, Lagerquist Hagglund C, Haneskog L, Brekkan E, Zhang Y, Osterberg T, Lundahl P. Immobilized liposome and biomembrane partitioning chromatography of drugs for prediction of drug transport. *Int J Pharm* 1998;164:129–137.
35. Chong S, Dando SA, Soucek KM, Morrison RA. *In vitro* permeability through Caco-2 cells is not quantitatively predictive of *in vivo* absorption for peptide-like drugs absorbed via the dipeptide transporter system. *Pharm Res* 1996;13:120–123.
36. Seiler P. Interconversion of lipophilicities from hydrocarbon/water systems into octanol/water system. *Eur J Med Chem* 1974;9:473–479.
37. el Tayar NTR, Testa B, Carrupt PA, Hansch C, Leo A. Percutaneous penetration of drugs: a quantitative structure–permeability relationship study. *J Pharm Sci* 1991;80:744–749.
38. ter Laak AM, Tsai RS, Donne-Op den Kelder GM, Carrupt PA, Testa B, Timmerman H. Lipophilicity and hydrogen-bonding capacity of H1-antihistaminic agents in relation to their central sedative side-effects. *Eur J Pharm Sci* 1994;2:373–384.
39. Ren S, Das A, Lien EL. QSAR analysis of membrane permeability to organic compounds. *J Drug Target* 1996;4:103–107.
40. Norinder U, Svensson P. Descriptors for amino acids using MolSurf parametrization. *J Comput Chem* 1998;19:51–59.

41. Raevsky OA, Grigor'ev VY, Kireev DB, Zefirov NS. Complete thermodynamic description of H-bonding in the framework of multiplicative approach. *Quantitative Structure–Activity Relationships* 1992;11:49–63.
42. Camenisch G, Alsenz J, van de Waterbeemd H, Folkers G. Estimation of permeability by passive diffusion through Caco-2 cell monolayers using the drugs' lipophilicity and molecular weight. *Eur J Pharm Sci* 1998;6:313–319.
43. Hansch C, Bjorkroth JP, Leo A. Hydrophobicity and central nervous system agents: on the principle of minimal hydrophobicity in drug design. *J Pharm Sci* 1987;76:663–687.
44. Palm K, Luthman K, Ungell A-L, Strandlund G, Artursson P. Correlation of drug absorption with molecular surface properties. *J Pharm Sci* 1996;85:32–39.
45. Winiwarter S, Bonham NM, Ax F, Hallberg A, Lennernas H, Karlen A. Correlation of human jejunal permeability (*in vivo*) of drugs with experimentally and theoretically derived parameters. A multivariate data analysis approach. *J Med Chem* 1998;41:4939–4949.
46. Stenberg P, Luthman K, Artursson P. Virtual screening of intestinal drug permeability. *J Control Release* 2000;65:231–243.
47. Yee S. *In vitro* permeability across Caco-2 cells (colonic) can predict *in vivo* (small intestinal) absorption in man—fact or myth. *Pharm Res* 1997;14:763–766.
48. Camenisch G, Folkers G, van de Waterbeemd H. Comparison of passive drug transport through Caco-2 cells and artificial membranes. *Int J Pharm* 1997;147:61–70.
49. Hidalgo IJ. Assessing the absorption of new pharmaceuticals. *Curr Top Med Chem* 2001;1:385–401.
50. Palm K, Stenberg P, Luthman K, Artursson P. Polar molecular surface properties predict the intestinal absorption of drugs in humans. *Pharm Res* 1997;14:568–571.
51. van de Waterbeemd H. Property based design: optimization of drug absorption and pharmacokinetics. *J Med Chem* 2001;44:1313–1333.
52. Hamalainen MD, Frostell-Karlsson A. Predicting the intestinal absorption potential of hits and leads. *Drug Discov Today Technol* 2004;4:397–405.
53. Ribadeneira MD, Aungst BJ, Eyermann CJ, Huang S-M. Effects of structural modifications on the intestinal permeability of angiotensin II receptor antagonists and the correlation of *in vitro*, *in situ*, and *in vivo* absorption. *Pharm Res* 1996;13:227–233.
54. Balimane PV, Chong S. Cell culture-based models for intestinal permeability: a critique. *Drug Discov Today* 2005;10:335–343.
55. Anderberg EK, Nystrom C, Artursson P. Epithelial transport of drugs in cell culture. VII: Effects of pharmaceutical surfactant excipients and bile acids on transepithelial permeability in monolayers of human intestinal epithelial (Caco-2) cells. *J Pharm Sci* 1991;81:879–887.
56. Terada T, Saito H, Mukai M, Inui KI. Recognition of beta-lactam antibiotics by rat peptide transporters, PEPT1 and PEPT2, in LLC-PK1 cells. *Am J Physiol Renal Physiol* 1997;273:706–711.
57. Thwaites DT, Hirst BH, Simmons NL. Passive transepithelial absorption of thyrotropin-releasing hormone (TRH) via a paracellular route in cultured intestinal and renal epithelial cell lines. *Pharm Res* 1993;10:674–681.
58. Hilgendorf C, Spahn-Langguth H, Regardh CG, Lipka E, Amidon GL, Langguth P. Caco-2 versus Caco-2/HT29-MTX co-cultured cell lines: permeability via diffusion, inside- and outside-directed carrier-mediated transport. *J Pharm Sci* 2000;89:63–75.
59. Hilgers AR, Conradi RA, Burton PS. Caco-2 cell monolayers as a model for drug transport across the intestinal mucosa. *Pharm Res* 1990;7:902–910.

60. Kansy M, Senner F, Gubernator K. Physicochemical high throughput screening: parallel artificial membrane permeation assay in the description of passive absorption processes. *J Med Chem* 1998;41:1007–1010.
61. Lipinski CA. Observation on current ADMET technology: no uniformity exists. Paper presented at the *Proceedings of the Annual Meeting of the Society of Biomolecular Screening*, The Hauge, The Netherlands.
62. Lipinski CALF, Dominy BW, Feeney PJ. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development. *Adv Drug Deliv Rev* 1997;23:3–25.
63. Kansy M, Avdeef A, Fischer H. Advances in screening for membrane permeability: high-resolution PAMPA for medicinal chemists. *Drug Discov Today Technol* 2004;1:349–355.
64. Chiou WL, Buehler PW. Comparison of oral absorption and bioavailability of drugs between monkey and human. *Pharm Res* 2002;19:868–874.
65. Mummaneni V, Dressman JB. Intestine uptake of cimetidine and ranitidine in rats. *Pharm Res* 1994;11:1599–1604.
66. Stewart BH, Chan OH, Lu RH, Reyner EL, Schmid HL, Hamilton HW, Steinbaugh BA, Taylor MD. Comparison of intestinal permeabilities determined in multiple *in vitro* and *in situ* models: relationship to absorption in humans. *Pharm Res* 1995;12:693–699.
67. Sugawara M, Saitoh H, Iseki K, Miyazaki K, Arita T. Contribution of passive transport mechanisms to the intestinal absorption of beta-lactam antibiotics. *J Pharm Pharmacol* 1990;42:314–318.
68. Ungell A-L, Nylander S, Bergstrand S, Sjoberg A, Lennernas H. Membrane transport of drugs in different regions of the intestinal tract of the rat. *J Pharm Sci* 1998;87:360–366.
69. Kirn D-C, Burton PS, Borchardt RT. A correlation between the permeability characteristics of a series of peptides using an *in vitro* cell culture model (Caco-2) and those using an *in situ* perfused rat ileum model of the intestinal mucosa. *Pharm Res* 1993;10: 1710–1714.
70. Lowther N, Fox R, Faller B, Nick H, Jin Y, Sergejew T, Hirschberg Y, Oberle R, Donnelly H. *In vitro* and *in situ* permeability of a "second generation" hydroxypyridinone oral iron chelator: correlation with physico-chemical properties and oral activity. *Pharm Res* 1999;16:434–440.
71. Fagerholm U, Nilsson D, Knutson L, Lennernas H. Jejunal permeability in humans *in vivo* and rats *in situ*: investigation of molecular size selectivity and solvent drag. *Acta Physiol Scand* 1999;165:315–324.
72. Piyapolrungroj N, Zhou, YS, Li C, Liu G, Zimmermann E, Fleisher D. Cimetidine absorption and elimination in rat small intestine. *Drug Metab Dispos* 2000;28:65–72.
73. Aulton M. *Pharmaceutics: The Science of Dosage Form Design*, 2nd ed. Edinburgh; New York: Churchill Livingstone; 2002.
74. Komiya I, Park JY, Kamani A, Ho NFH, Higuchi WI. Quantitative mechanistic studies in simultaneous fluid flow and intestinal absorption using steroids as model solutes. *Int J Pharm* 1980;4:249–262
75. Balimane PV, Chong S, Morrison, RA. Current methodologies used for evaluation of intestinal permeability and absorption. *J Pharmacol Toxicol Methods* 2000;44:301–312.
76. Lennernas H, Ahrenstedt O, Hallgren R, Knutson L, Ryde M, Paalzow LK. Regional jejunal perfusion, a new *in vivo* approach to study oral drug absorption in man. *Pharm Res* 1992;9:1243–1251.
77. Salphati L, Childers K, Pan L, Tsutsui K, Takahashi L. Evaluation of a single-pass intestinal-perfusion method in rat for the prediction of absorption in man. *J Pharm Pharmacol* 2001;53:1007–1013.

78. Petri N, Tannergren C, Holst B, Mellon FA, Bao Y, Plumb GW, Bacon J, O'Leary KA, Kroon PA, Knutson L, Forsell P, Eriksson T, Lennernas H, Williamson G. Absorption/metabolism of sulforaphane and quercetin, and regulation Of phase II enzymes, In human jejunum *in vivo*. *Drug Metab Dispos* 2003;31:805–813.
79. Takamatsu N, Kim O-N, Welage LS, Idkaidek NM, Hayashi Y, Barnett J, Yamamoto R, Lipka E, Lennernãs H, Hussain A, Lesko L, Amidon GL. Human jejunal permeability of two polar drugs: cimetidine and ranitidine. *Pharm Res* 2001;18:742–744.
80. Bonlokke L, Hovgaard L, Kristensen HG, Knutson L, Lennernas H. Direct estimation of the *in vivo* dissolution of spironolactone, in two particle size ranges, using the single-pass perfusion technique (Loc-I-Gut(R)) in humans. *Eur J Pharm Sci* 2001;12:239–250.
81. Pithavala YK, Heizer WD, Parr AF, O'Connor-Semmes RL, Brouwer KLR. Use of the InteliSite® capsule to study ranitidine absorption from various sites within the human intestinal tract. *Pharm Res* 1998;15:1869–1875.
82. InteliSite data sheet, www.innovativedevices.com.
83. Bailey CA, Bryla P, Malick AW. The use of the intestinal epithelial cell culture model, Caco-2, in pharmaceutical development. *Adv Drug Deliv Rev* 1996;22:85–103.
84. Fagerholm U, Johansson M, Lennernãs H. Comparison between permeability coefficients in rat and human jejunum. *Pharm Res* 1996;13:1336–1342.

# HOW AND WHERE ARE DRUGS ABSORBED?

Marival Bermejo and Isabel Gonzalez-Alvarez
*University of Valencia, Valencia, Spain*

**Contents**

1 Introduction
2 Absorption Impact on Drug Bioavailability
3 How: Absorption Mechanisms
  3.1 Passive Diffusion
  3.2 Carrier Mediated
  3.3 Endocytosis
4 Where
  4.1 Absorption Routes: Enteral Versus Parenteral
  4.2 Routes of Drug Administration and Physiological Variables Affecting Xenobiotic Absorption
References

## 1 INTRODUCTION

A drug or a xenobiotic, in order to exert its pharmacological or toxic effects, must interact with its receptors at their action sites or therapeutic target. The essential step is therefore the absorption of the substance, that is, its entry from outside the body, unless the administration is done directly into the bloodstream. In this chapter, we review the biological barriers that the drug must cross in order to enter the body, which mechanisms are implicated in this process, and how the physiological characteristics of the administration route affect the absorption rate and extent.

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

## 2 ABSORPTION IMPACT ON DRUG BIOAVAILABILITY

The terms absorption and bioavailability are often used interchangeably when there is actually a conceptual difference between them. *Bioavailability* is a measurement of the rate and extent of therapeutically active drug that reaches the systemic circulation and is available at the site of action [1, 2]. *Absorption* is the movement of drug molecules across biological barriers from the site of administration. In some cases absorption has been defined as the movement from the administration site into the systemic circulation, but this definition could be misleading if in the pathway to the blood circulation the drug crosses a metabolic organ or tissue. The systemic availability $F_{sys}$ is the result of absorption and all possible losses before arriving at the systemic circulation:

$$F_{sys} = F_a \cdot F_m$$

where $F_a$ is the fraction absorbed and $F_m$ is the fraction escaping first-pass extraction. From the equation, it is easy to conclude that $F_a$, the fraction absorbed, represents the upper limit for systemic availability in the absence of any presystemic losses of the drug on its way to the systemic circulation. It is therefore clear, that the absorption process is relevant to the availability of the drug at the action sites.

On the other hand, it is necessary to be aware of the terminology classically applied to some administration routes: if the drug does not reach the systemic circulation, then it is considered that no absorption occurs, and therefore the administration is considered to be topical. This is the case, for instance, for ophthalmic administration or for the topical administration of drugs on the skin. The entry of drug into those tissues is called *penetration* or *permeation* but not absorption, because the drug is not intended to reach the systemic circulation. This nomenclature then leads to the conclusion that the drug is bioavailable (as it reaches its action sites) but not absorbed into the bloodstream (or does not present systemic absorption).

Throughout the chapter, the difference between absorption and bioavailability for those administration routes that involve a first pass through a metabolic organ (mainly liver) or tissue (i.e., intestinal tissue) is stressed, and the drug is considered absorbed once it is inside the body.

The next step is to analyze the factors affecting $F_a$ by the different routes of drug administration. As described in this chapter, absorption is a transport phenomenon that can be described mechanistically, in order to identify the limiting steps. The complexity of the barriers—subcellular, cellular, or multicellular—are also addressed.

## 3 HOW: ABSORPTION MECHANISMS

The complexity and number of barriers that the drug has to cross before reaching the bloodstream differs from one route of administration to other, but in all the cases the barriers are integrated by semipermeable cell membranes. They are composed of a bimolecular lipid matrix, containing mostly cholesterol and phospholipids. The lipids provide stability to the membrane and determine its permeability

characteristics. Globular proteins, of various sizes and composition, embedded in the matrix are involved in transport and function as receptors for cellular regulation. Drugs may cross a biologic barrier by passive diffusion, facilitated passive diffusion, active transport, or endocytosis.

### 3.1 Passive Diffusion

Passive diffusion involves the movement of drug molecules down a concentration or electrochemical gradient without the expenditure of energy. Passive diffusion does not involve a carrier, is not saturable, and shows a low structural specificity [3]. Passive absorption of drug can be done by two pathways: paracellular and transcellular.

A paracellular pathway consists of the diffusion of xenobiotics across the intercellular spaces or tight junctions, aqueous pores, or fenestrae [4–6]. These junctions are more or less leaky, depending on the tissue, and allow the diffusion of water and small solutes. In general, the main restriction for using this route is the molecular size followed by the electric repulsion between the ionized groups of the molecules and the ionized groups in the constituents of the aqueous pathways [7–9].

From a mathematical point of view, paracellular diffusion follows the same diffusion law that is applied to transcellular diffusion. The passive movement of molecules down a concentration gradient is described by Fick's law. The process is depicted in Fig. 1.

$$J = \frac{1}{A} \cdot \frac{dM}{dt} = P \cdot \Delta C = \frac{D \cdot (K_1 \cdot C_{\text{aq-outside}} - K_2 \cdot C_{\text{aq-inside}})}{h}$$

where $J$ represents the net flux, that is, the net rate of diffusion through the membrane surface (g/(cm$^2$·s)); $P$ is the permeability coefficient (cm/s); $D$ is the diffusion coefficient of the solute in the membrane (cm$^2$/s); $\Delta C$ is the concentration gradient across the membrane (g/cm$^3$); and $K_1$ and $K_2$ are the partition coefficients of the solute in the membrane (at both sides).

As can be seen from this equation, there are several factors that affect the diffusion of the molecules, such as lipophilicity, pH/p$K_a$, size, and area of the absorptive

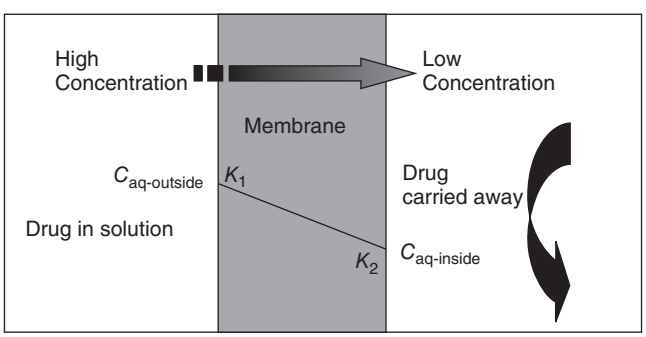


**FIGURE 1** Scheme of the process of passive diffusion across biological membranes. The driving force for diffusion is the concentration gradient. The partition coefficient of the xenobiotic in the lipid bilayer is one of the main determinants of resistance to transport.

surface. Diffusion coefficients of drugs in the membrane increase with drug lipophilicity and decrease with molecular weight. On the other hand, it is the nonionized fraction of the molecules that mostly contributes to the overall transport, as the nonionized fraction has the highest partition coefficient.

The Henderson-Hasselbalch equation can be used to calculate the relative proportions of the ionized and nonionized forms of the drug [10]:

$$\text{For acids:} \quad \%\ \text{ionized} = \frac{100}{1+10^{(\mathrm{p}K_a-\mathrm{pH})}}$$

$$\text{For bases:} \quad \%\ \text{ionized} = \frac{100}{1+10^{(\mathrm{pH}-\mathrm{p}K_a)}}$$

In other words, mass diffusion per unit area of a weakly acid drug will be higher from a more acid medium while mass diffusion per unit area of a weakly basic drug will be favored from a more alkaline medium.

A second consequence of the solute ionization in biological fluids is *ion trapping* phenomenon. When the pH of the fluids is different on both sides of the membrane, the solute will be trapped on the side favoring its ionization. A similar phenomenon is produced by the binding to macromolecules, for instance, proteins, as only the free drug is able to diffuse across the membrane.

### 3.2 Carrier Mediated

In addition to the passive diffusional processes over lipid membranes or between cells, substances can be transferred through the lipid phase of biological membranes through specialized systems, that is, active transport and facilitated diffusion [11].

***Facilitated Diffusion*** Facilitated diffusion involves the participation of a carrier molecule, so this mechanism also shows selectivity and saturation, but the transport occurs always from a higher concentration to a lower concentration area and thus it does not require any energy source [12].

One theory is that a carrier component combines reversibly with the substrate molecule at the cell membrane exterior, and the carrier–substrate complex diffuses rapidly across the membrane, releasing the substrate at the interior surface. The carrier transports only substrates with a relatively specific molecular configuration, and the process is limited by the availability of carriers.

***Active Transport*** Active transport of a xenobiotic is an energy requiring process that is saturable (i.e., is limited by the number of protein transporters present) and could proceed against a concentration gradient. Active transport can be affected by the presence of inhibitors [12]. There are two kinds of active transporters: primary active transporters use energy coming from the hydrolysis of ATP; while secondary active transporters use the energy coming from some electrochemical gradient generated by a primary transporter, for instance, a difference in $H^+$ or $Na^+$ ion concentration. Symporters transport the drug and the cotransported molecule or ion in the same direction, while antiporters transport the drug and the ion in opposite directions.

The active transport process can be described in mathematical terms using the Michaelis–Menten equation, as the binding of the solute to the carrier molecule and its translocation to the other side of the membrane have several similarities to the sustrate–enzyme interaction:

$$J_{\text{active}} = \frac{dM}{A \cdot dt} = \frac{V_{\max} \cdot C}{K_{\text{m}} + C}$$

where $V_{\max}$ is the maximal transport velocity (mass/(area·time)), $C$ is the drug concentration, and $K_{\text{m}}$ is the drug concentration at which velocity of transport is $\frac{1}{2} V_{\max}$.

When $C >> K_{\text{m}}$,

$$J \quad \frac{V_{\max} \cdot C}{C} \quad V_{\max}$$

and the rate of the transport process becomes independent of the concentration. But when the drug concentration is much lower than $K_{\text{m}}$ (i.e., $K_{\text{m}} >> C$), then

$$J \quad \frac{V_{\max}}{K_{\text{m}}} \cdot C$$

and the transport process follows an apparent first order kinetic.

When a drug is able to be absorbed by passive diffusion and active transport, the total flux is the sum of the passive and active fluxes, as can be seen in Fig. 2.

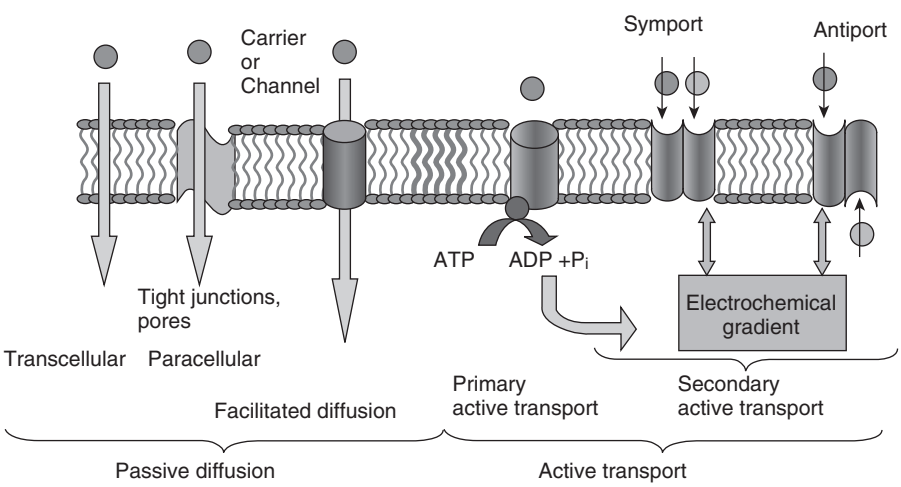


**FIGURE 2** Transport flux as a function of drug concentration for a substance transported by passive diffusion, active transport, or a combination of both processes.

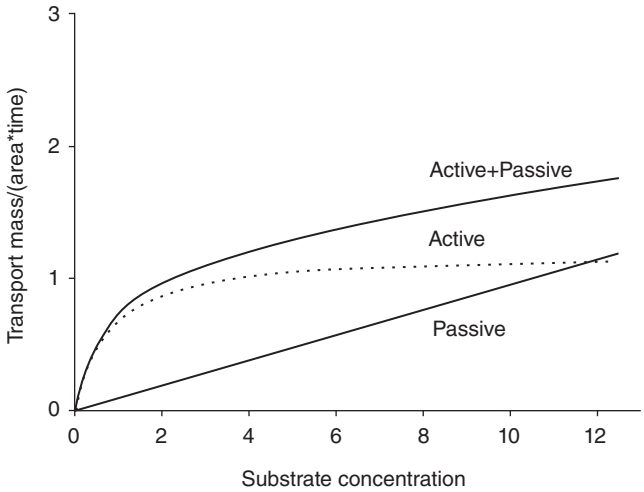


**FIGURE 3** Summary of the transport mechanisms that xenobiotics use for crossing biological membranes (with the exception of endocytosis).

### 3.3 Endocytosis

Endocytosis is also called vesicle mediated transport. The molecules or particles are engulfed by a cell. The cell membrane invaginates, encloses the fluid or particles, then fuses again, forming a vesicle that later transports them across the membrane to the inside of the cell. This mechanism also requires energy expenditure. This diffusion is dependent on the size of the pore and the molecular weight of the chemical. Phagocytosis is the type of endocytosis where an entire cell is engulfed. Pinocytosis is when the external fluid is engulfed. Receptor-mediated endocytosis occurs when the material to be transported binds to certain specific molecules in the membrane. Examples include the transport of insulin and cholesterol into animal cells. Endocytosis probably plays a minor role in drug absorption, except for protein drugs.

In summary, chemicals can be introduced into the body by various mechanisms that are represented in Fig. 3. Once the drug is in the systemic circulation it has access to its target tissue or receptors by using some of the aforementioned mechanisms.

## 4 WHERE

### 4.1 Absorption Routes: Enteral Versus Parenteral

A tablet inside our stomach is, technically, outside the body. This image helps to define the two main groups of administration routes. Anything that is administered in the intestinal tract is using the enteral route (from *enteron* = intestine). By exclusion, all the other administration routes are defined as parenteral (*par* = outside). Enteral routes include buccal and sublingual, rectal, and oral. Among the parenteral routes we can distinguish those that require a needle to introduce the drug to the

depot site (e.g., subcutaneous or intramuscular routes) and those using other mucosal surfaces or epithelial barriers (e.g., nasal, ophthalmic, or transpulmonary).

Selection of the route of administration involves the evaluation of many different factors, from the physicochemical characteristics of the drug, which determine its ability to cross the barriers, to the pharmacological action, which could require a fast absorption rate to assure a quick onset of response. Chronic therapies require ease of use and patient friendly routes. Lastly, the existence of the appropriate technology could limit the use of an administration route if a suitable dosage form is not in place.

### 4.2 Routes of Drug Administration and Physiological Variables Affecting Xenobiotic Absorption

A prerequisite to absorption by any administration route is drug dissolution. Any solid drug products (e.g., tablets) have to disintegrate and deaggregate in order to release the drug particles and to allow drug dissolution. The solubility of the drug in biological fluids at the administration site determines the gradient for diffusion, while permeability in the membrane establishes the resistance to the transport. For most administration routes these two factors—drug solubility and drug membrane permeability—are the main factors governing drug absorption.

***Enteral Routes*** Enteral absorption routes comprise all those involving any part of the gastrointestinal tract, that is, buccal and sublingual, rectal, and the so-called oral route that requires swallowing the substance, and its access to the intestine.

*Oral* The oral route of administration is the most used and convenient for patients. The gastrointestinal tract is a complex environment from the standpoint of drug absorption; therefore, a deep understanding of gastrointestinal physiology is essential to determine how and where drugs and xenobiotics are absorbed from the oral route [13, 14]. The main physiological factors that must be considered to understand nutrients and drug absorption are summarized in Table 1 and are discussed in this section.

The gastrointestinal system includes the gastrointestinal tract and all the organs that secrete substances into it. The main functions are digestion, secretion, absorption, and motility [15]. The gastrointestinal tract begins in the mouth and continues

**TABLE 1 Physiological Factors Affecting Drug Absorption from the Gastrointestinal Tract** [17]

| |
|---|
| Membrane permeability |
| Luminal pH |
| Intestinal secretions presence of food |
| Surface area |
| Transit time |
| Disease state |
| Water fluxes |
| Mucus and unstirred water layer (UWL) |

with the esophagus, which is the pathway to the stomach. The stomach is a sac-like organ with a quite acidic environment (pH ranging from 2 in fasted state to 5 or 6 in fed state) thanks to the secretion of hydrochloric acid by parietal cells. This acidic environment constitutes the first line of immune protection of the gastrointestinal tract and, on the other hand, is the first barrier against absorption, as many substances are not stable in this environment.

Digestion's final stages and most absorption occur in the next section: the small intestine. The small intestine is divided into three segments. The initial short one, the duodenum, receives the secretions of the pancreas and liver through a common duct. The pH of this segment ranges from 5 near the stomach to 7 at the final portion. The jejunum has a thicker and more vascular wall and along with the duodenum is the site where most digestion and absorption occur, so actually the small intestine has a big functional reserve. The last portion, the ileum, is the longest segment, a little narrower in diameter and with a pH rather constant at 6.5–7. Only a small portion of water, salts, and undigested material reach the large intestine. The large intestine has three parts: cecum, colon, and rectum. Between the small and large intestines there is a valve that prevents the material from going back. The large intestine has some enzymes, which are able to metabolize some drugs, along with a big population of bacteria that contribute to the fermentation of undigested residues. The large intestine is responsible for absorbing water and electrolytes and for the formation, storage, and evacuation of feces [13].

There is extensive folding on the intestinal surface. There are folds extending from the surface of the geometrical cylinder. In the surface of the folds, there are finger-like projections called villi. Villi contain absorptive cells and goblet cells. The absorptive cells are characterized by the small projections of their apical membranes that are called brush border or microvilli. The apical membrane is highly specialized and contains enzymes, binding sites, and transporters. Other important features are the tight junctions, which play a major role in regulating paracellular permeability of water and solutes. Permeability of junctional complexes is higher in the upper intestine than in the ileum.

Figure 4 shows how the combination of folds, villi, and microvilli increases the surface available for absorption about 600 times [15]. Each villus is occupied by a blind-ended lymphatic vessel (called a lacteal) and by a capillary network, whose drainage leads to the hepatic portal vein. That means that any substance absorbed in the intestinal epithelium must cross the liver before arriving in the systemic circulation. During this first passage through the liver, the nutrients, xenobiotics, and drugs can be metabolized, reducing the fraction of substance that finally arrives in the systemic circulation. This is the *first-pass effect* whose effect on bioavailability is analyzed later. On the other hand, during the passage through the liver the drug or xenobiotic could be excreted into the bile and again reach the small intestine, where it is again absorbed. This process is called enterohepatic circulation and could affect the disposition of the compound.

The large intestine has folds but no villi. The smaller surface area available for absorption and the smaller amount of fluid present in the large intestine are reflected, in general, in a lower permeability for xenobiotics; but, on the other hand, the longer residence time in this area may compensate for the lower permeability and makes the large intestine a suitable target area for prolonged release drug formulations [16, 17].

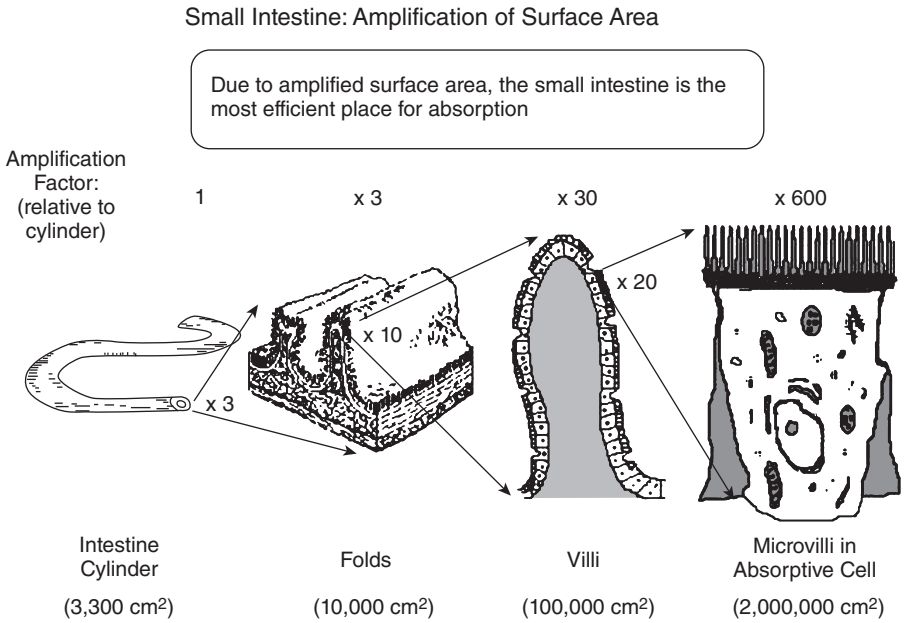


**FIGURE 4** Anatomical modifications of the intestinal system that increase the available surface for absorption. (From Ref. 15, with permission from Modern Biopharmaceutics™.)

Gastrointestinal secretions contribute to the digestion and absorption of nutrients. The main gastrointestinal secretions, their main components, and the substances they contribute to digestion are summarized in Table 2 [14].

Carbohydrates are digested into monosaccharides, proteins are hydrolized into amino acids or small peptides, and fat droplets are transformed into free fatty acids and monoglycerides. The resulting compounds from digestion as well as fat-soluble and water-soluble vitamins are absorbed by some of the absorption mechanisms explained in the previous section. All essential nutrients including proteins, carbohydrates, fats, vitamins, and minerals are substrates of some active or facilitated carrier system. There are excellent reviews about the nature and diversity of intestinal transporters [12, 18–21].

Another process of particular relevance for fat absorption that may also influence drug absorption is emulsification. Emulsification of fat droplets, thanks to mechanical disruption and the presence of emulsifying agents (bile salts and phospholipids), increase the surface available for lipase action, thus increasing the rate of lipid digestion. The digestion products are incorporated into the core of aggregates known as micelles. Micelles consist of bile salts, fatty acids, monoglycerides, and phospholipids, all clustered together with the polar ends of each molecule oriented toward the surface. The micelles release their contents near the lipid bilayer and, in consequence, the released molecules can diffuse easily across the intestinal lining. Fat-soluble vitamins follow the pathway for fat absorption, and any other highly lipophilic molecule could also be incorporated in the micellar phase. Thus, for instance, any interference with the secretion of bile decreases the absorption of fat-soluble vitamins.

**TABLE 2 Main Gastrointestinal Secretions, Their Components, and the Nutrients They Contribute to Digestion** [14]

| Secretion | Main Components | Digestion |
|---|---|---|
| Saliva | Water | |
| 1–1.5 liters/day | Electrolytes | |
| pH ~ 8 at secretion rate | Mucus | |
| | Lysozime | |
| | α-Amylase | Disaccharides |
| Gastric fluid | Mucus | |
| 1.5–2 liters/day | Hydrochloric acid | All |
| | Pepsinogen (pepsin precursor) | Proteins |
| | Intrinsic factor | Essential for vitamin $B_{12}$ absorption |
| Pancreatic secretions | Bicarbonate | Neutralizes acid from stomach |
| 1–1.5 liters/day | | |
| | Trypsin, chymotrypsin, carboxypeptidases | Protein |
| | α-Amylase | Disaccharides |
| | Pancreatic lipase | Fatty acids |
| Bile | Bile salts | Fat digestion |
| 0.5–1 liter/day | Cholesterol | |
| | Phospholipids | |
| | Bile pigments | |
| | Proteins | |
| | Bicarbonate, electrolytes | |
| Intestinal fluid | Electrolytes | |
| 2 liters/day | Water | |
| | Mucus | |

Water is the most abundant substance in chyme. Around 9 liters of ingested and secreted fluids enter the small intestine each day, but only 1.5 liters passes onto the large intestine, since 80% of the fluid is absorbed in the small intestine. Finally, only 150 grams of feces, consisting about 100 mL water and 50 g of solid material, is normally eliminated each day [14, 15]. The amount of water present in the gastrointestinal system is relevant for ensuring the complete dissolution of the administered dose. Some drugs present a dose-limited absorption when there is not enough water in the luminal fluids to allow the complete dissolution during transit time.

ABSORPTION OF DRUGS AND XENOBIOTICS The two main factors affecting rate and extent of absorption are drug permeability and solubility in intestinal fluids. Solubility of the drug determines the highest gradient for drug diffusion across the membrane. This concept is represented in Fig. 5.

Chemical stability or phenomena such as aggregation and complexation with luminal components may reduce the drug concentration in the lumen. Solubility and drug concentration in the luminal fluids are functions of the molecule $pK_a$ for weak acids and bases and the pH of the luminal fluids. The presence of different intestinal secretions and the change in secretion rate because of the presence of food changes luminal pH, as summarized in Fig. 6.

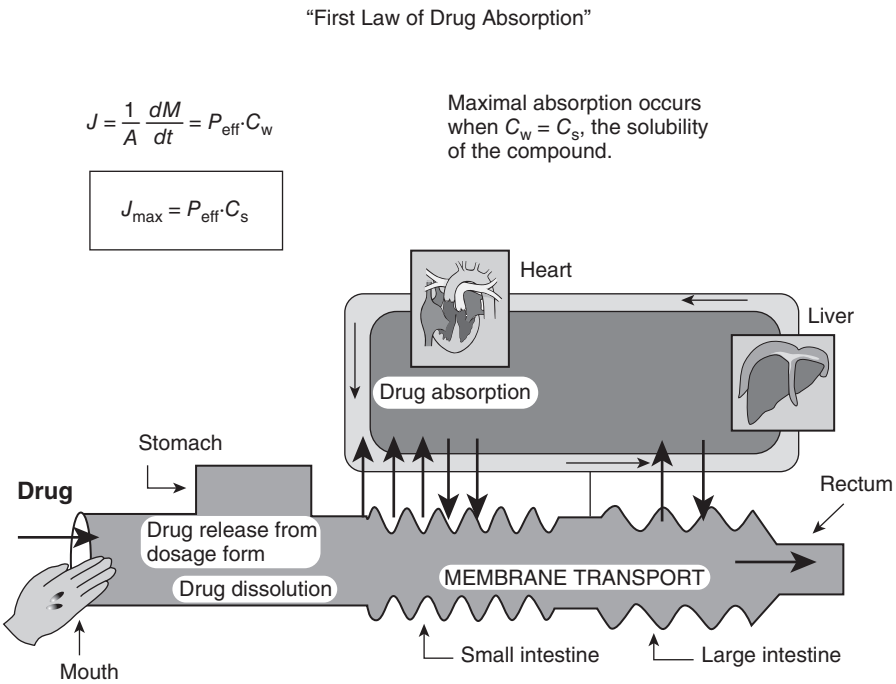


**FIGURE 5** Factors determining rate and extent of absorption in the gastrointestinal tract.

| Site | fasted | fed |
|---|---|---|
| Stomach | 1.4 - 2.1 | 3.0 - 7.0 |
| Duodenum | 4.9 - 6.4 | 5.1 - 5.2 |
| Jejunum | 4.4 - 6.5 | 5.2 - 6.2 |
| Ileum | 6.5 - 8.0 | 6.8 - 8.0 |

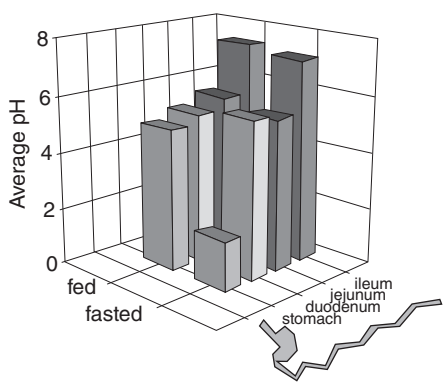


**FIGURE 6** Average intestinal pH values in the gastrointestinal tract in the fasted and fed states. Data from Dressman et al. [85].

Finally, if the drug is administered in a solid form, any factor affecting its dissolution rate will have an impact on the rate and extent of absorption, so particle size may be relevant, in particular, for those substances of low solubility.

Permeability of the drug, as well as its permeation mechanism, depends on several molecular properties. Some of them are summarized in Table 3 [22]. Drugs administered by the oral route may cross the intestinal membrane by all the

**TABLE 3 Molecular Descriptors Related to Drug Permeability in the Intestinal Epithelium**

| | |
|---|---|
| **Lipophilicity** | As drug partitioning into the cell membrane is one of the steps in membrane transport, lipophilicity is widely used as a predictor of drug permeability. Lipophilicity has two principal components, molecular size and hydrogen bonding potential [25, 87–90]. |
| **Molecular Weight (MW)** | This is a component of lipophilicity as well as diffusion coefficient in biological membranes and fluids. A rather strong dependence between transcellular diffusion and molecular size has been observed. Compounds with MW < 200 are able to pass through the intestinal membrane by the paracellular pathway along with diffusion through the transcellular route. Compounds with MW > 250 use the transcellular route but further increases of the molecular weight (MW > 500) consequently lead to a decrease in membrane diffusion [91]. |
| **Hydrogen Bonding** | The absorption ability of a molecule depends on the number and the strength of the hydrogen bonds that the molecule is able to form with water molecules, because the first step to enter into the membrane is the desolvation of the molecule. Hydrogen bonding capacity is detrimental for transport into the nonpolar environment of the cell membrane. Thus, these properties along with lipophilicity are good descriptors of drug permeation [92]. |
| **Polar Surface Area** (*PSA*) | The PSA of a molecule is defined as the area of its van der Waals surface that arises from oxygen or nitrogen atoms plus the area of the hydrogen atoms attached to these heteroatoms. As such, PSA is clearly related to the capacity to form hydrogen bonds. Drugs with PSAd < 60 $Å^2$ would be completely absorbed (fraction absorbed, FA > 90%). Drugs with PSAd > 140 $Å^2$ would be absorbed less than 10%. |
| **Nonpolar Surface Area** | Nonpolar substituents facilitate membrane transport and hydrophobic compounds generally have higher permeabilities than hydrophilic ones (with similar hydrogen bonding properties). Nonpolar surface area can also correlate with membrane permeability. In general, this parameter is included in correlations along with PSA [88, 93–95]. |

*Source*: Adapted from Ref. 22.

previously mentioned mechanisms: passive diffusion, facilitated passive diffusion, active transport, or pinocytosis.

The paracellular pathway, between the epithelial cells, is both size (MW, volume) and charge dependent [6, 23, 24]; cations seem to penetrate the negatively charged tight junctional system much more easily than anions [23]. On the other hand, the available surface area for paracellular intestinal absorption has been estimated to be about 0.01% of the total surface area of the small intestine [4, 6].

Due to the lipidic nature of the intestinal membrane, the ability of the molecules to diffuse across it is dependent on their lipophilicity [25, 26]. In particular, one of the main factors affecting drug permeability is the distribution coefficient, which depends on the pH of the luminal fluids. As has been described, the pH of the luminal contents changes along the gastrointestinal tract and it is different in the fasted versus the fed state. That affects the solubility of ionizable compounds and their distribution coefficient; so, in general, permeability also varies along the gastrointestinal tract for a given solute [27].

The unstirred water layer (UWL) adjacent to the intestinal lining was considered to be the rate-limiting step for intestinal permeability of high permeability compounds [28]. However, several *in vivo* studies have clearly reported that the thick-

ness of this UWL is significantly thinner than was previously assumed, thanks to the motility of the gastrointestinal tract [29, 30].

There is now enough evidence about the involvement of intestinal transporters in the absorption of drugs [31]. Targeting intestinal transporters by means of development of prodrugs has been a successful strategy for improving oral absorption. For example, the intestinal peptide transporter is utilized in order to increase the bioavailability of several classes of peptidomimetic drugs, especially ACE inhibitors and beta-lactam antibiotics [19]. The bioavailability of poorly absorbed drugs can be improved by utilization of the transporters responsible for the intestinal absorption of various solutes and/or by inhibiting the transporter involved in the efflux system [32–35].

The last relevant physiological factor that determines drug absorption is intestinal motility, as this function limits the intestinal transit time of any solute. Since the small intestine is the main site for drug absorption, gastric emptying is the first limiting step for the access of the drug to the intestine. The average gastric emptying time is around 30 minutes, transit time in the small intestine is around 3 hours, and transit time in the large intestine is much more variable but an average of 3 days has been reported [15, 36]. Some factors affecting the gastric emptying rates of liquids are described in Table 4.

Solid emptying is slower than liquid emptying; also, emptying depends on the caloric content of the meal (as for liquids), and particles of size bigger than 2 mm remain in the stomach until the next "housekeeping wave" that completely empties the stomach [15].

Intestinal peristalsis is stimulated by voluminous meals and hypertonic contents. Anticholinergic drugs slow intestinal movements while cholinergic drugs stimulate transit, so simultaneous administration of these drugs with food may affect absorption.

To complete the description of this route of absorption, it is necessary to remember again the difference between absorption and oral bioavailability. Win L. Chiou [1] defined absorption as the movement of drug across the outer mucosal membranes of the gastrointestinal tract; while, in general, the systemic availability ($F_{sys}$) is defined as the fraction of the amount (dose) of the xenobiotic (drug) that reaches the systemic circulation. As described previously, during first passage through the liver and also during passage through the intestinal cells, the drug could be biotransformed by the intestinal or hepatic enzymes. In mathematical terms, this process is expressed as follows:

$$F_{sys} = F_a \cdot (1 - E_g) \cdot (1 - E_h) \tag{1}$$

**TABLE 4 Factors Affecting Gastric Emptying Rates of Liquids** [15]

| Gastric Emptying | Comment |
|---|---|
| Volume | Larger volume, faster emptying rate |
| Osmotic pressure | Iso-osmotic content empties faster than hyper- or hypotonic solutions |
| pH | |
| Caloric content | Higher caloric content slows emptying |
| Viscosity | Higher viscosity delays emptying |

where $F_a$ is the fraction absorbed, $E_g$ is the gut extraction ratio, and $E_h$ is the hepatic extraction ratio. $(1 - E_g)$ is the fraction escaping intestinal metabolism and $(1 - E_h)$ represents the fraction escaping liver metabolism. In the absence of any presystemic losses, systemic bioavailability equals the fraction absorbed.

*Buccal and Sublingual* Administration of drugs using the oral cavity is an option well accepted by patients. Oral mucosa is relatively permeable with a rich blood supply [37–39]. A second advantage is the avoidance of first-pass effect and presystemic elimination in the GI tract. These factors make the oral mucosal cavity a potential alternative for systemic drug delivery [40].

Within the oral cavity, delivery of drugs can be made through the sublingual mucosa or by placing the dosage form in the mucosa lining the internal face of the cheeks, which is called the buccal route.

Oral mucosa is a stratified epithelium with a turnover time around 5–6 days [41]. It is estimated that the permeability of the buccal mucosa is 4–4000 times greater than that of the skin [42]. The buccal area has 40–50 cell layers, whereas the sublingual zone is slightly thinner. Oral mucosa is keratinized and relatively impermeable in all the areas that suffer mechanical stress, but the sublingual and buccal areas are not keratinized, rendering them more permeable [41, 43, 44]. In general, the permeabilities of the oral mucosa decrease in the order of sublingual > buccal > palatal [41].

For significant drug absorption, the drug must have a prolonged exposure to the mucosal surface so the drug taste must be acceptable to the patient [45].

The absorption potential of the buccal mucosa is influenced by the lipid solubility and therefore the permeability of the solution (osmosis), the ionization (pH), and the molecular weight of the substances. Because the pH of saliva is usually 6.5–6.9, absorption is favored for drugs with a high $pK_a$ [41, 46].

The basic drug transport mechanism for buccal epithelium is passive diffusion, as for other epithelia in the body. There are two major routes involved: transcellular (intracellular) and paracellular (intercellular) [47]. The intercellular spaces are the major barrier to permeation of lipophilic compounds and the cell membrane acts as the major transport barrier for hydrophilic compounds. The route that predominates, however, is generally the one that provides the least amount of hindrance to passage [40]. Some authors have suggested the presence of a specialized transport system for cephadroxyl in the human buccal membrane [48].

One of the major disadvantages associated with buccal drug delivery is the low flux, which results in low drug bioavailability and the lack of dosage form retention at the site of absorption. Consequently, bioadhesive polymers have extensively been employed in buccal drug delivery systems.

To sum up, the buccal mucosa offers several advantages for controlled drug delivery for extended periods of time and is a promising area for systemic delivery of orally inefficient drugs (such as potent peptide and protein drug molecules). However, the need for safe and effective buccal permeation/absorption enhancers is a crucial component for a prospective future in the area of buccal drug delivery.

Sublingual literally means "under the tongue." Sublingual mucosa is more permeable than buccal mucosa. Because of the high permeability and the rich blood supply, the sublingual route is capable of producing a rapid onset of action, but the area is continuously washed by a high amount of saliva, making the buccal area

more suitable for placement of retentive dosage forms. In clinical practice, sublingual drug administration is applied in the field of cardiovascular drugs, steroids, some barbiturates, and enzymes and it has been explored as an alternative for vitamins and minerals, which are found to be readily absorbed by this method. It could be especially useful for those who experience difficulty in swallowing tablets.

*Rectal* Absorption through the rectal mucosa is, in general, irregular and not complete so it can not be considered a first delivery option. The acceptance of the rectal route varies among countries but rectal delivery could be a convenient, alternative route when other routes are not available. The following situations represent opportunities for rectal administration:

- Patients having difficulty swallowing, nausea, vomiting, or gastric pain.
- Uncooperative or nonconscious patients.
- When access to the intravenous route is difficult.
- For drugs unstable in the gastrointestinal tract.
- For drugs with high first-pass effect.
- When a slow and prolonged input could have therapeutic interest.

For instance, rectal administration is used for anticonvulsants [49], nonnarcotic and narcotic analgesics, theophylline, antiemetics, and antibacterial agents and for inducing anesthesia in children [50, 51].

Its principal advantage is that it is independent of gastrointestinal tract motility and rate of gastric emptying [52].

However, this route should be avoided in the following situations:

- Painful anal conditions (fissures or inflamed hemorrhoids).
- Immunosuppressed patients in whom even minimal trauma could lead to formation of an abscess [53].
- Elderly patients with diarrhea.
- Patients physically unable to place the suppository.

The rectal route also has other disadvantages. There may be a great deal of variation among individuals regarding the necessary dose. Moreover, the rectal dose generally must be higher than the dose administered intravenously or orally, as absorption is slower and incomplete.

In humans, the rectum comprises the last 12–19 cm of the large intestine and the rectal epithelium is formed by a single layer of columnar or cuboidal cells and goblet cells; its surface area is about 200–400 $cm^2$. The absorbing surface area of the rectum is considerably smaller than that of the small intestine, as the former lacks villi and microvilli. However, the epithelia in the rectum and the upper intestinal tract are histologically similar, giving them comparable abilities to absorb drugs. The rectal mucosa is richly vascularized.

Drugs absorbed via the inferior and middle rectal veins can in part avoid first-pass metabolism because the blood flow from these veins bypasses the portal system and empties into the vena cava. First-pass effects are not completely avoided,

however, as there are anastomoses among inferior and middle veins with the superior one. The drugs absorbed via the superior rectal veins are transported to the liver via the portal system [54–56].

For a number of drugs the extent of rectal absorption has been reported to exceed oral values, which may reflect partial avoidance of hepatic first-pass metabolism after rectal delivery. This phenomenon has been reported for morphine, metoclopramide, ergotamine, lidocaine (lignocaine), and propranolol [54, 57].

Passive diffusion is the predominant mechanism of absorption and it is mainly dependent on the molecular weight, liposolubility, and degree of ionization of molecules. The contribution of other mechanisms, such as carrier mediated transport or pinocytosis, can be considered negligible. Paracellular diffusion is more restricted than in the upper zone of the gastrointestinal tract due to the lower permeability of the tight junctions in this area. The rate of rectal transmucosal absorption is affected by different factors: for example, the surface area available for absorption is small, particularly for drugs in solid dosage forms, which provide less extensibility than liquid forms. The volume present in the rectal cavity is also variable among patients but generally the volume is small and the viscosity is high. Diluent volume is also an important determinant of rectal drug uptake, as demonstrated with methohexital administered rectally for preprocedure sedation [58].

Lipophilic drugs are better absorbed than hydrophilic ones. Rectal pH may also influence drug uptake by altering the amount of drug that is ionised [59]. The pH of the rectal vault in children ranges from 7.2 to 12.2 [60]. This pH range favors absorption of barbiturates that will remain in a nonionized state because their $pK_a$ is near the physiologic range (~7.6). The greater lipid solubility of nonionized drugs enhances their movement across the membrane. Finally, in a similar way to the small intestine where transit time limits absorption, rectal retention of the drug determines the absorption. Retention time, in many cases, is affected by the nature of the drug excipients. Hydrophilic excipients are more irritating for the rectal mucosa and increase peristaltic movements, reducing retention time.

### *Parenteral Routes (Noninjectables)*

*Transpulmonary* The respiratory system consists of two main areas—the conducting and the respiratory regions. The first one is responsible for filtering and humidifying the air entering the respiratory region, where gas interchange takes place. The conducting region is formed by the nasal cavity, nasopharynx, bronchi, and bronchioles. Airways distal to the bronchioles and the alveoli constitute the respiratory region.

The transpulmonary route has classically been used for drugs intended to exert a local action in the lungs and more recently for systemic administration. In the first case, it allows the delivery of high drug concentration directly to the action site. In this way, the same therapeutic effect can be obtained with a fraction of the dose administered by any other systemic route, and simultaneously systemic and adverse effects are minimized. In general, a rapid onset of action can be achieved. For systemic action, the transpulmonary route offers the advantages of being a friendly needle-free option, with a huge absorptive surface area ($100\,m^2$) of a highly permeable membrane in the alveolar region (0.2–0.7 μm). Compared to the gastrointestinal tract, the possibility of enzymatic degradation is reduced, and large molecules

can be absorbed in part thanks to a prolonged residence time, as the mucociliary clearance is less marked in the lung periphery.

Nevertheless, there are also barriers against absorption through this route. The respiratory system is designed to prevent the entry of particles. The airway geometry, the humidification mechanisms, and the mucociliary clearance contribute to this process and then constitute barriers for transpulmonary absorption. Once in the lower parts of the respiratory tract, the drug has to deal with the lung surfactant, surface lining fluid, the epithelium and basement membranes, and the capillary endothelium before arriving at the blood flow.

The main factors affecting the amount of drug that is finally deposited in the deep lung are the particle size and density. The combination of both factors defines the aerodynamic diameter, which is actually the factor governing the deposition site. The relation between particle diameter and aerodynamic diameter of a spherical particle is defined by the following equation:

$$d_a = \sqrt{\frac{\rho}{\rho_a}} \cdot d$$

where $d_a$ is the aerodynamic diameter, $d$ is the particle diameter, $\rho$ is the particle density, and $\rho_a = 1\,\text{g/cm}^3$.

Particles of mean aerodynamic diameter of 1–3 μm deposit minimally in the mouth and throat and maximally in the lung's parenchymal (i.e., alveolar or "deep lung") region. Tracheobronchial deposition is maximal for particles with aerodynamic diameters between 8 and 10 μm. Particles possessing an aerodynamic diameter smaller than 1 μm (although greater than several hundred nanometers) are mostly exhaled, and particles larger than 10 μm have little chance of making it beyond the mouth.

Once deposited in the lungs, inhaled drugs are (1) cleared from the lungs, (2) absorbed into the systemic or lymphatic circulation, or (3) degraded via drug metabolism.

*Mucociliary Clearance.* Mucus is secreted by goblet cells and submucosal glands forming a double sol-gel layer covering the ciliated epithelium. Mucociliary clearance can be impaired in some pathological conditions either by impairment or ciliary movement or by the production of thicker mucus.

*Absorption.* Lipophilic molecules pass easily through the airway epithelium via passive transport. Hydrophilic molecules cross via paracellular pathways, such as tight junctions, or by active transport via endocytosis and exocytosis. The rate of protein absorption from the alveoli is size dependent. Effros and Mason [61] demonstrated an inverse relationship between alveolar permeability and molecular weight.

*Metabolism.* The lung is the only organ through which the entire cardiac output passes. The liver first-pass effect is thus avoided through this route, but metabolism during the passage through the epithelium can still happen even if the relevance of metabolic clearance in the lung for systemic availability of inhaled drug is not well understood. All metabolizing enzymes found in the liver are found to a lesser extent in the lung: phase 1 cytochrome P450 (CYP450)

enzymes, flavin-containing monooxygenases (FMOs), monoamine oxidase (MAO), aldehyde dehyrogenase, NADPH-CYP450 reductase, esterases, and proteases are all present in the lung. The monooxygenase system metabolizes fatty acids, steroids, and lipophilic xenobiotics. Esterase presents in high concentrations in alveolar macrophages, and to a lesser degree in alveolar type I and II cells. Proteins and peptides are subject to hydrolysis by proteases. However, for most proteins degradation in the alveoli is not a major clearance mechanism, with >95% of proteins, including insulin, being absorbed intact from the lung periphery.

*Nasal* The nasal mucosa is a promising site for the delivery of many drugs, including proteins and other large biomolecules that do not have good absorption through other routes. The nasal cavity is easily accessible and extensively vascularized, and compounds administered via this route avoid the hepatic first-pass effect. In this respect, it should be an ideal route for noninvasive delivery. In addition, absorption of drug at the olfactory region of the nose provides a potential pathway for a pharmaceutical compound to be available to the central nervous system. The nasal delivery of vaccines is another very attractive application in terms of efficacy and patient acceptance. However, there are a number of factors that limit its utility. The physiology of the nasal cavity presents the most significant barrier to drug absorption. Limiting factors include rapid mucociliary clearance, enzymatic degradation in the mucus layer, and low permeability of the nasal epithelium [62–66].

The nasal vestibule and atrium are the less permeable areas in the nasal cavity, whereas the respiratory region (turbinate) is the most permeable thanks to the increased surface area and the rich vasculature. This latter area has a large surface area ~150–160 $cm^2$ because of the presence of microvilli. The pH of the zone varies between 5.5 and 6.5 in adults and between 5.0 and 7.0 in children.

The mechanisms for drug absorption in this area include passive diffusion (para- and transcellular), carrier mediated transport, and transcytosis. Lipophilic drugs are well absorbed with bioavailabilities reaching 100%; but permeability is low for polar molecules (including low molecular weight drugs and large peptides and proteins). In general, absorption is restricted for drugs with molecular weights higher than 1000 daltons.

Deposition of the formulation in the anterior portion of the nose provides a longer nasal residence time but permeability is lower in this area; whereas deposition in the posterior zone, where the permeability is higher, provides shorter residence times. This shorter residence time is due to the mucociliary clearance, which must be maintained for normal physiological functions, such as the removal of dust, allergens, and bacteria. A prolonged residence time in the nasal cavity may also be achieved by using bioadhesive polymers, microspheres, and chitosan or by increasing the viscosity of the formulation. Nasal mucociliary clearance can also be stimulated or inhibited by drugs, excipients, preservatives, and/or absorption enhancers and thus affect drug delivery to the absorption site.

Several proteases and amino peptidases present in the nasal mucosa might affect the stability of proteins and peptides; nevertheless, the level of proteolytic activity is much lower than that in the gastrointestinal tract. Peptides may also form complexes with immunoglobulins (Igs) in the nasal cavity, leading to an increase in the molecular weight and a reduction of permeability.

A linear inverse correlation has been reported between the absorption of drugs and molecular weight up to 300 daltons. Absorption decreases significantly if the molecular weight is greater than 1000 daltons, except with the use of absorption enhancers.

These enhancers work by a variety of mechanisms, such as by modifying the phospholipid bilayer, leaching out protein from the membrane, or even stripping off the outer layer of the mucosa. Some of these enhancers also have an effect on the tight junctions and/or work as enzymatic inhibitors. The main problem associated with enhancers is that most of them cause significant mucosal damage at the concentrations required to enhance nasal absorption. Some of the enhancers that have been investigated include bile salts, dihydrofusidates, surfactants, and fatty acid derivatives.

Chitosan, a bioadhesive polymer, has been shown to improve the absorption by increasing the contact time between the drug and the nasal membrane and, hence, its retention on the nasal mucosa. It also had an effect on paracellular transport. Other bioadhesive materials are carbopol, cellulose agents, starch, and dextran. The use of cyclodextrins as a means of enhancing solubility and/or absorption also looks promising [67, 68].

*Ophthalmic* Systemic absorption of drugs from the ocular route is considered, in general, a nondesired effect. The main objective of the application of a drug on the front of the eye is to exert some effect in the eye and the annex structures. There are five potential targets for ophthalmic drugs: the precorneal structures (conjunctiva and eyelids), the cornea, the posterior and anterior chambers, the vitreous cavity, and the retina (Fig. 7). Preocular structures, cornea, and the anterior and

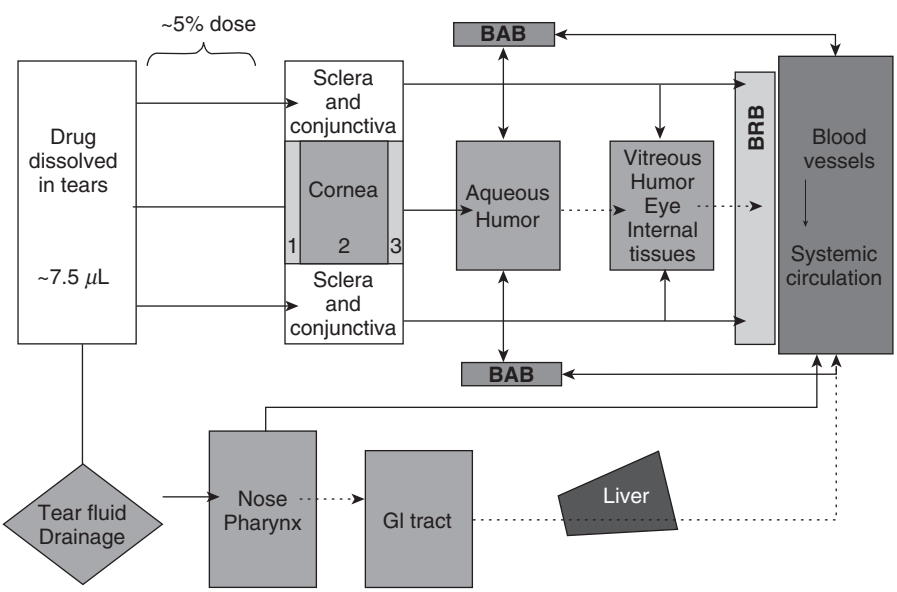


**FIGURE 7** Absorption route for drugs applied on the eye surface. BAB, blood aqueous barrier; BRB, blood retinal barrier; 1, epithelium; 2, stroma; 3, endothelium. (Adapted from Ref. 86.)

posterior chambers can be accessed by topical application but the posterior side of the eye is difficult to reach from the eye surface, and, in general, systemic administration by another route or intraocular injection is used to deliver drugs to the vitreous cavity. The main pathway for drugs to enter the anterior chamber is via the cornea. Some large and hydrophilic drugs prefer the conjunctival and scleral route, and then diffuse into the ciliary's body [69], while small lipophilic compounds permeate through the corneal pathway.

On the other hand, reaching the anterior part of the eye from the systemic circulation is difficult due to the existence of the blood–ocular (blood–aqueous and blood–retinal) barrier, similar to the hematoencephalic barrier. The tight junctions of the capillary endothelial cells restrict the entry of substances from the blood into the aqueos humor and/or into the retina [70].

Once the drug is dissolved in the tear fluid, its residence time in the conjunctival sac is short (between 2 and 5 minutes) because of the drainage of the instillation fluid with tears and the tear turnover. Drainage rates decrease with increased viscosity of the instilled fluid but increase with increased volume of the instilled fluid because the eye tends to maintain the normal tear volume around 7.5 μL. If the drug is irritating to the eye or the conjunctiva, it will induce lacrimation and decrease residence time. The drug from the lacrimal duct arrives at the nose and then pharynx and can be swallowed and reach the gastrointestinal tract. From all these routes the drug can be absorbed, and potentially it could cause adverse effects. The drug in the tears is either absorbed by the conjunctiva, sclera, or cornea. The conjunctiva is a richly irrigated mucosa covering the inside of the eyelids and sclera. Its surface is higher than the corneal surface as well as its permeability, so this absorption pathway competes with the corneal route, reducing intraocular bioavailability.

The corneal barrier is formed by a series of three layers without blood supply: epithelium, stroma, and endothelium. The outer layer is a multistratified epithelium representing 10% of the total thickness. Between the epithelium and the stroma lies Bowman's membrane, containing strong collagen fibers, which help the cornea maintain its shape. Ninety percent of the thickness corresponds to the stroma, which is constituted by a net of parallel collagen fibrils and high water content. The endothelium pumps water from the cornea, keeping it clear. Epithelium and endothelium are lipophilic, while the stroma is a barrier for hydrophobic compounds. Lipophilicity, molecular weight, and ionization degree are the main factors affecting corneal permeability as well as the degree of binding to the protein content of the lacrimal fluid [71]. Lacrimal pH oscillates around 7.0–7.4 and has a low buffering capacity; thus, the pH and buffering capacity of the instillation solution could affect ophthalmic absorption. The optimum apparent partition coefficient (octanol/pH 7.4 buffer) for corneal absorption is in the range of 100–1000. It was shown that increasing molecular size of the permeating substance decreases the rate of paracellular permeation, as the pore size does not allow permeation of big molecules. Nevertheless, transcellular diffusion is the main pathway for most drugs used in the clinic [69].

*Transdermal* The application of drug substances over the skin could have two different purposes: acting locally over the skin surface or providing drug absorption into the systemic circulation. Cutaneous administration refers, in general, to the first purpose while transdermal absorption defines the incorporation of a drug inside the body, crossing the skin in order to reach the blood vessels and exert systemic effects.

The advantage of the transdermal route over other routes is its large accessible surface area (average surface 1.8 $m^2$) and the avoidance of drug degradation in the gastrointestinal tract. Nevertheless, the skin by itself has homeostatic and protective functions, and it is a formidable barrier membrane, thanks mainly to the contribution of the stratum corneum (SC).

SKIN PHYSIOLOGY There are two important layers to human skin: the epidermis, and the dermis, which contains blood vessels. The skin contains annex structures that include sweat and sebaceous glands and hair follicles. These structures cross the skin from the dermis and open on the epidermis surface.

The stratum corneum (or horny layer) is the top layer of the skin and varies in thickness from 10 to 20 μm depending on the region of the body. It is composed of 15–20 layers of dead, flat, and keratinized epidermal cells (keratinocytes) surrounded by a lipid matrix, which renders this structure as the most significant barrier to diffusion and drug transport. The viable epidermis lies below the stratum corneum. Its cells have a greater degree of hydration, so diffusion is faster through this area. Viable epidermis contains melanocytes, which provide skin with pigmentation and Langerhans cells (antigen-presenting cells to the immune system). Finally, the dermis is the layer containing the blood vessels, sensory neurons, and a lymphatic network. The thickness of the dermis (1 mm) is approximately 100 times the thickness of the stratum corneum.

DRUG PERMEATION Generally, drug absorption into the skin occurs by passive diffusion according to Fick's law. The transport rate is directly proportional to the surface area of the skin and inversely proportional to the thickness of the stratum corneum. Regarding the physicochemical characteristics of the compound, drug hydrosolubility, lipophilicity, and molecular weight, as well as charge, are the main factors affecting permeation, as in other routes where passive diffusion is the main mechanism. There is in most cases a parabolic relationship between the drug lipophilicity (expressed as octanol–water partition coefficient) and the permeation rate [72, 73]. This kind of correlation appears because compounds with low partition coefficient (log $P$) present little partitioning into the skin and therefore low permeability, while compounds with high partition coefficient also give low permeability due to their inability to partition out of the stratum corneum. The generally accepted range of log $P$ for maximum permeation is between 1 and 3 [74]. For hydrophilic or charged molecules, resistance to transport is higher due to the lipid-rich nature of the stratum corneum and its low water content. Transport of lipophilic drug molecules is facilitated by their dissolution into intercellular lipids around the cells of the stratum corneum.

There are two main pathways by which drugs can cross the skin and reach the systemic circulation (see Fig. 8): the transfollicular and transepidermal pathways.

The *transfollicular route* is through the hair follicles and sebaceous glands. Hair follicles and sebaceous glands penetrate through the stratum corneum, allowing more direct access to the dermal microcirculation. Some hydrophilic substances can diffuse through this pathway. However, due to the low relative surface area (around 1%), very little drug actually crosses the skin via this route.

The *transepidermal pathway* consists of two paths. The *transcellular* represents the shortest path. However, the drug must cross environments of hydrophilic

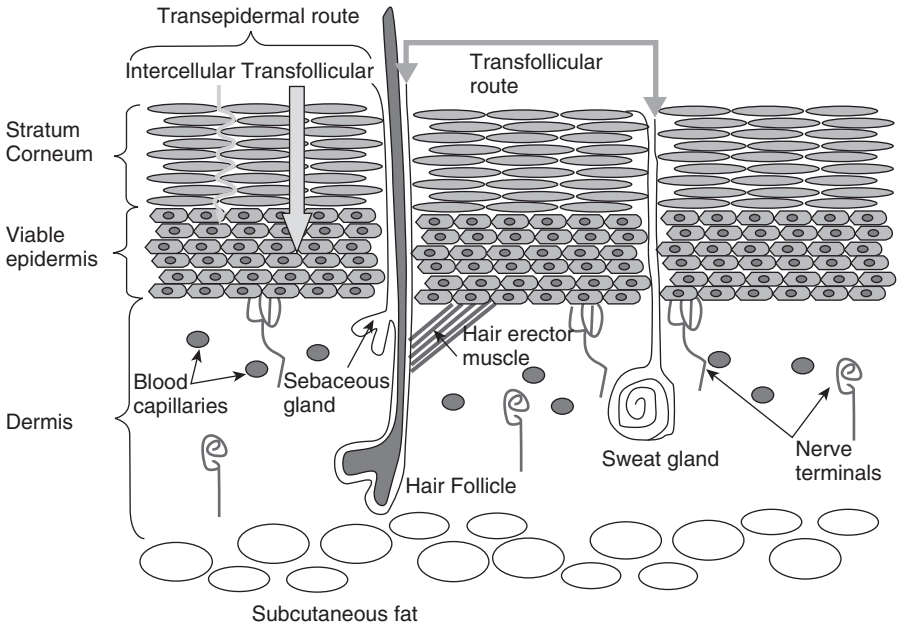


**FIGURE 8** Skin structure and main pathways for transdermal absorption.

**TABLE 5 Physiological Factors Relevant for Drug Penetration Through the Skin**

| Age | Neonates, elderly People: higher permeability [96] |
|---|---|
| Ethnicity | Controversial results in the literature: skin more permeable in Caucasians and Asians versus blacks, or no difference detected [97–99] |
| Body areas | Most permeable areas: mucous membranes, scrotal skin, eyelids<br>Intermediate permeability: face, chest/back, buttocks, abdomen, upper arms/legs<br>Less permeable zones: palms, feet soles, nails |
| Skin condition | Hydration degree: hydrated skin is more permeable [100]<br>Irritation: if stratum corneum is broken, permeability is increased [100]<br>Temperature: warmer skin is more permeable<br>Eczema: increased permeability<br>Psoriasis: thicker skin but disrupted barrier function; reported increase in permeability for low molecular drugs and for some high molecular entities [101] |

(keratinocytes) and liphophilic nature (extracellular lipids) and that means that the substance needs a balanced hydrophilic–liphophilic nature to be able to follow this route and in general most substances find high resistance to permeation. The *intercellular* pathway is the more common pathway. Nevertheless, even if the thickness of the SC is very small, the tortuosity of the diffusion pathway around the cells increases the resistance to drug penetration.

Some physiological factors affecting drug penetration across the skin are summarized in Table 5.

Examples of drugs that are administered via the transdermal route include clonidine, estradiol, fentanyl, nicotine, nitroglycerin, scopolamine, testosterone, oxybutynin, and the combination products norelgestromin/ethinyl estradiol and estradiol/norethindrone acetate. The common characteristics is that they are all molecules of low molecular weight but with enough potency to be active at low blood concentrations (few ng/mL or less) [75].

### *Vaginal*

VAGINAL PHYSIOLOGY The human vagina is a slightly S-shaped tube that communicates the cervix of the uterus with the external body surface. The tube is collapsed with the anterior and posterior walls in contact with each other. The average length of the vagina is 8–12 cm [76].

The vaginal wall is comprised of an epithelial layer, a muscular layer, and the tunica adventitia. During the menstrual cycle, the thickness of the vaginal epithelial cell layer changes by approximately 200–300 μm. The surface of the vagina is composed of numerous folds or rugae. The rugae provide distensibility, support, and an increased surface area for the vaginal wall [76].

The primary venous drainage occurs via the pudendal veins. The vaginal, uterine, vesical, and rectosigmoid veins from the middle and upper vagina lead the blood flow to the inferior vena cava, thus bypassing the hepatic portal system.

Age, hormone status, and pregnancy cause changes in vaginal physiology as well as the pH changes due to several factors including semen, menstruation, estrogen status, and bacterial colonization. Reproductive hormones control the thickness of the vaginal epithelium, with estradiol 17-β (E2) thickening the epithelium and hypoestrogenism resulting in atrophy [77]. The vaginal fluid ranges mostly a transudate from vaginal and cervical cells but also contains vulvar secretions, cervical mucus, endometrial and oviductal fluids, and microorganisms and their metabolic products.

The normal average vaginal pH in healthy women of reproductive age ranges from 3.8 to 4.2. This acidic environment is maintained by the production of lactic acid by the vaginal microflora and it constitutes part of the defensive mechanism of the vaginal mucosa. The amount and composition of the vaginal fluid also changes throughout the menstrual cycle. Women of reproductive age produce fluid at a rate of 3–4 g/4 h, while this amount decreases by 50% in postmenopausal women. The human vaginal fluid may contain enzymes, enzyme inhibitors, proteins, carbohydrates, amino acids, alcohols, hydroxylketones, and aromatic compounds.

VAGINAL ABSORPTION Vaginal delivery is a potential option for both local and systemic delivery. Vaginal mucosa possess several advantages when other routes of administration fail, such as avoidance of first-pass metabolism. On the other hand, this route of administration could be used for particular purposes, such as providing prolonged absorption from sustained release formulations or achieving local therapeutic effects while avoiding systemic administration and potential adverse effects of the drug. In general, the vaginal route provides good systemic absorption mostly for low molecular weight drugs [78].

Enzymatic activity of vaginal epithelium and vaginal fluids are two physiological factors that have shown to have influence on vaginal absorption. The external cell

layers and the basal cell layers of the vagina retain most of the enzyme activity. Among the enzymes present, proteases are likely to be the prominent barrier for the absorption of intact peptide and protein drugs into the systemic circulation [79]. The absorption of a drug that is poorly water soluble may be increased when the fluid volume is higher. However, the presence of overly viscous cervical mucus may present a barrier to drug absorption and increased fluid volume may remove the drug from the vaginal cavity and subsequently reduce absorption [80].

Since many drugs are weak electrolytes, the pH may change their degree of ionization and affect the absorption of drug. As in other membranes, the ionization decreases the permeability of the drug as the un-ionized form has the higher permeability [80].

Physicochemical properties such as molecular weight, lipophilicity, ionization, and surface charge can influence vaginal drug absorption [79]. Lipophilicity, in general, increases drug permeability but low molecular weight lipophilic drugs are likely to be absorbed more than large molecular weight lipophilic or hydrophilic drugs. On the other hand, the molecular weight cutoff above which compounds are not absorbed may be higher for the vagina than other mucosal surfaces, such as the small intestine or colon [81]. Drugs with molecular weight >300 Da have shown a higher permeability in *in vitro* models of vaginal mucosa compared to small intestine and colon [81].

In conclusion, knowledge about the relationship between physicochemical properties and human vaginal permeability is still very limited; much work needs to be done in this area.

***Parenteral Routes (Injectables)*** Parenteral drug administration comprises all the nonenteral routes. Some of the parenteral routes require a needle for placing the drug at the absorption site. The most used extravascular parenteral routes are intradermal, subcutaneous, and intramuscular, which are represented in Fig. 9. Drug has to cross different barriers to be absorbed from the depot compartment.

*Intradermal* The dermis is a layer underneath the epidermis, which contains the blood capillaries and nerve terminations. In general, the body area used to administer intradermal injections is the upper part of the arm and/or the back. It is used mainly for diagnostic agents, vaccines, or desensitization agents (a method to reduce or eliminate an organism's negative reaction to a substance or stimulus). The maximum volume to be administered by this route is 0.1 mL. Intradermal drugs diffuse slowly from the injection site into local capillaries, and the process is a little faster with drugs administered subcutaneously.

*Subcutaneous* This route of administration can be used in either short-length or chronic therapies. The drug is injected, or the delivery system is placed in the interstitial tissue beneath the dermis. The most used zones for subcutaneous injection are the upper arm, the upper thigh, the lower part of the abdomen, or the upper part of the back. Maximal volume is around 2 mL. The blood supply to the subcutaneous tissue is less than the blood supply to the underlying muscular tissue; thus, subcutaneous absorption can be slower than intramuscular absorption. Nevertheless, compared to the oral route, subcutaneous absorption can be more rapid and predictable. The following factors can alter the absorption rate through this route:

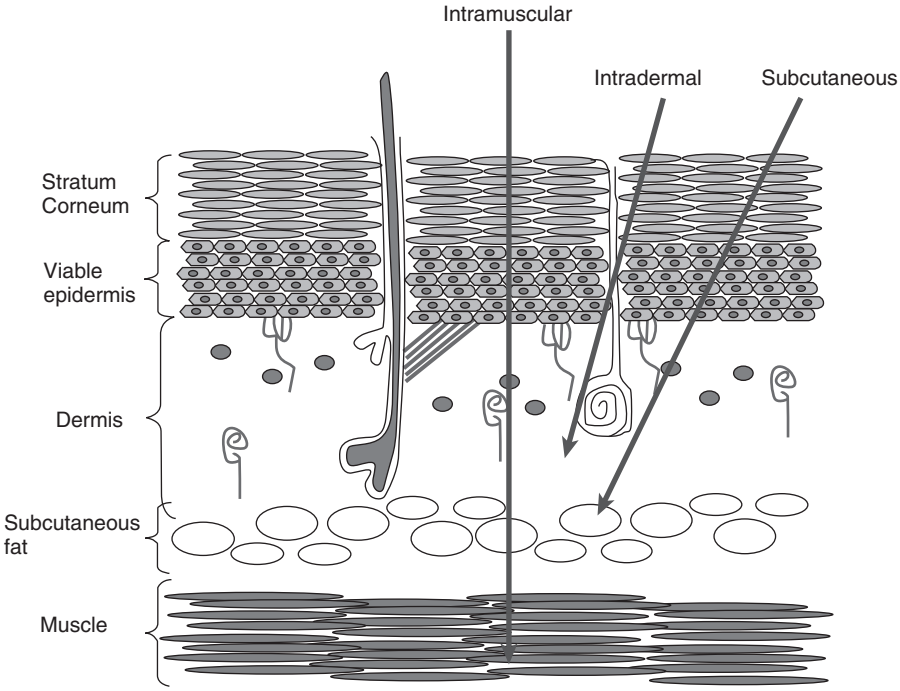


**FIGURE 9** Parenteral routes of drug administration: intradermal, subcutaneous, and intramuscular. The injection depth determines where the drug depot is formed.

heating or massaging the injection zone (to increase absorption), coadministering vasodilators or hyaluronidase (to increase diffusion rate), or administering epinephrine to cause vasoconstriction (and thus decrease absorption rate).

*Intramuscular* This route is easier and less risky to use than the intravenous one, but it is more painful and the action onset is slower as it requires an absorption step. The injection is done in the muscular layer right under the subcutaneous tissue, avoiding blood vessels and nerves. The main sites of absorption are: deltoid muscle (up to 2 mL), gluteus muscle (5 mL), vastus lateral is muscle (upper thigh 15 mL). The absorption rate through this route depends on physiological factors, such as blood flow, exercise, and injection depth. The physicochemical characteristics of the drugs can also affect absorption rate (see Table 6) and in some cases the drug precipitates in the intramuscular depot further delaying the absorption.

*Absorption Routes* Once the drug is injected, a depot is formed in contact with the surrounding tissue. From the depot the drug can reach blood capillaries or lymphatic capillaries. Absorption proceeds through the blood or lymph, depending mainly on the drug molecular weight. Drugs with molecular weight higher than 2000 daltons preferentially use lymphatic vessels, while drugs with lower molecular weight are absorbed in the blood vessels. Blood capillaries have a thin wall formed by endothelial cells constituting a lipid barrier with aqueous pores and intercellular spaces called fenestrations. Pores and fenestrations represent only 0.1–0.2% of the capillaries' surface but the flow of water and soluble substances through them can

**TABLE 6 Factors Affecting Parenteral Absorption by the Intradermal, Subcutaneous, and Intramuscular Routes**

| Physiological Factors | Physicochemical Factors |
|---|---|
| Blood flow | Lipophilicity |
| Muscular activity | Solubility |
| Vasoconstriction degree | Dissolution rate |
| Medium viscosity | $pK_a$ |

be very fast. Parenteral drug absorption occurs by passive diffusion through the lipid barrier or convective diffusion through the pores and fenestrations. Both processes can be described using Fick's diffusion law:

$$\frac{dQ_a}{dt} = \left( \frac{D \cdot P \cdot S}{L \cdot V_a} + \frac{D' \cdot S'}{\eta \cdot L' \cdot V_a} \right) \cdot Q_a$$

where $dQ_a/dt$ is absorption rate from the depot compartment; $Q_a$, the amount of drug in the depot compartment; $D$, diffusion coefficient in the lipid membrane; $P$, partition coefficient; $S$, membrane surface; $L$, membrane thickness; $V_a$, volume of parenteral deposit; $D'$, aqueous diffusion coefficient; $S'$, aqueous pores and fenestration surface; $L'$, average pore length; and $\eta$, medium viscosity.

Considering all the terms inside the parentheses as constants, the process is described as first order kinetics. In consequence, the absorption rate is directly proportional to the diffusion coefficient, available surface, and partition coefficient and inversely proportional to the membrane thickness, pore length, and medium viscosity. Absorption rate can be increased by including vasodilators or diffusion agents (hyaluronidase) in the drug formulation or decreased by using vasoconstrictors and thickening agents.

The drug's properties that exhibit more influence on drug absorption rate are those affecting dissolution rate and lipophilicity. Dissolution rate is conditioned by the particle size, crystalline structure, and polymorphs. Lipophilicity influences the drug's partitioning into the membrane and this factor depends on the drug's $pK_a$ and the pH of the fluid, as the un-ionized form is more lipophilic. All these factors are summarized in Table 6.

In the last decades new needle-free technologies have been developed for injectable delivery. Needle-free injectors are devices that do not use a needle to place the drug in its depot site. The mechanism involves high pressure to push the drug formulation through the skin to the desired site. Pressure is produced by using either a gas (carbon dioxide or nitrogen) or a spring device. The pressure forces the medication through a small opening in the device while it is held against the skin. This creates a fine stream of the medication that penetrates the skin [82, 83].

***Other Parenteral Routes*** Intraspinal (epidural and intrathecal) delivery of drugs is an alternative route used mainly for anesthesia and pain-killer medications, in particular, when patients experience intolerable side effects from systemic drugs. Intraspinal delivery can be given into the epidural space or into the intrathecal space [84].

Drugs delivered epidurally also circulate systemically. By contrast, drugs delivered intrathecally circulate only in the cerebropsinal fluid (CSF). In both cases, the drugs administered by the intraspinal route bypass the blood–brain barrier. However, epidurally administered drugs must first cross the dura (the protective outer layer of the spinal cord) before entering the CSF. Therefore, when the epidural route of delivery is used, more time and higher doses are required for the drugs to exert the same effect compared with intrathecal delivery.

## REFERENCES


1. Chiou WL. The rate and extent of oral bioavailability versus the rate and extent of oral absorption: clarification and recommendation of terminology. *J Pharmacokinet Pharmacodyn* 2001;28(1):3–6.
2. Shargel L, Yu ABC. *Applied Biopharmaceutics & Pharmacokinetics*, 4th ed. Stamford, CT: Appleton & Lange; 1999.
3. Mayerson M. *Principles of Drug Absorption. Modern Pharmaceutics edition*. New York: Marcel Dekker; 1995.
4. Artursson P. Epithelial transport of drugs in cell culture. I: A model for studying the passive diffusion of drugs over intestinal absorptive (Caco-2) cells. *J Pharm Sci* 1990;79(6):476–482.
5. Artursson P, Magnusson C. Epithelial transport of drugs in cell culture. II: Effect of extracellular calcium concentration on the paracellular transport of drugs of different lipophilicities across monolayers of intestinal epithelial (Caco-2) cells. *J Pharm Sci* 1990;79(7):595–600.
6. Artursson P, Ungell AL, Lofroth JE. Selective paracellular permeability in two models of intestinal absorption: cultured monolayers of human intestinal epithelial cells and rat intestinal segments. *Pharm Res* 1993;10(8):1123–1129.
7. Fagerholm U, Borgstrom L, Ahrenstedt O, Lennernas H. The lack of effect of induced net fluid absorption on the *in vivo* permeability of terbutaline in the human jejunum. *J Drug Target* 1995;3(3):191–200.
8. Fagerholm U, Nilsson D, Knutson L, Lennernas H. Jejunal permeability in humans *in vivo* and rats *in situ*: investigation of molecular size selectivity and solvent drag. *Acta Physiol Scand* 1999;165(3):315–324.
9. Lennernas H, Ahrenstedt O, Ungell AL. Intestinal drug absorption during induced net water absorption in man; a mechanistic study using antipyrine, atenolol and enalaprilat. *Br J Clin Pharmacol* 1994;37(6):589–596.
10. Hasselbalch K. Die Berechnung der Wasserstoffzahl des Blutes aus der freien und gebundenen Kohlensäure desselben und die Sauerstoffbindungen des Blutes als Funktion der Wasserstoffzahl. *Biochemistry* 1916;78:112.
11. Adibi SA. The oligopeptide transporter (Pept-1) in human intestine: biology and function. *Gastroenterology* 1997;113(1):332–340.
12. Sadee W, Drubbisch V, Amidon GL. Biology of membrane transport proteins. *Pharm Res* 1995;12(12):1823–1837.
13. Martinez MN, Amidon GL. A mechanistic approach to understanding the factors affecting drug absorption: a review of fundamentals. *J Clin Pharmacol* 2002;42(6):620–643.
14. DeSesso JM, Jacobson CF. Anatomical and physiological parameters affecting gastrointestinal absorption in humans and rats. *Food Chem Toxicol* 2001;39(3):209–228.

15. *Modern Biopharmaceutics*. [computer program]. Version 6.03; Amidon GL, Bermejo M. TSRL, Inc; 2003.
16. Kararli TT. Comparison of the gastrointestinal anatomy, physiology, and biochemistry of humans and commonly used laboratory animals. *Biopharm Drug Dispos* 1995;16(5): 351–380.
17. Dressman JB, Bass P, Ritschel WA, Friend DR, Rubinstein A, Ziv E. Gastrointestinal parameters that influence oral medications. *J Pharm Sci* 1993;82(9):857–872.
18. Oh DM, Amidon GL. Overview of membrane transport. *Pharm Biotechnol* 1999;12:1–27.
19. Oh DM, Han HK, Amidon GL. Drug transport and targeting. Intestinal transport. *Pharm Biotechnol* 1999;12:59–88.
20. Kunta JR, Sinko PJ. Intestinal drug transporters: *in vivo* function and clinical importance. *Curr Drug Metab* 2004;5(1):109–124.
21. Shin HC, Landowski CD, Amidon GL. Transporters in the GI tract. In Waterbeemd V-d, Lennernäs H, Pirtursson P, Eds. *Drug Bioavailability/Estimation of Solubility, Permeability and Absorption*. Weinheim, Germany: Wiley-VCH; 2003.
22. Bermejo M, Ruiz-Garcia A. Oral permeability predictions—-from *in silico* to *in vivo* models. *Business Briefing Pharma Tech* 2002;175–180.
23. Karlsson J, Ungell A, Grasjo J, Artursson P. Paracellular drug transport across intestinal epithelia: influence of charge and induced water flux. *Eur J Pharm Sci* 1999;9(1): 47–56.
24. Pade V, Stavchansky S. Estimation of the relative contribution of the transcellular and paracellular pathway to the transport of passively absorbed drugs in the Caco-2 cell culture model. *Pharm Res* 1997;14(9):1210–1215.
25. Bermejo M, Merino V, Garrigues TM, et al. Validation of a biophysical drug absorption model by the PATQSAR system. *J Pharm Sci* 1999;88(4):398–405.
26. Taylor DC, Pownall R, Burke W. The absorption of beta-adrenoceptor antagonists in rat *in-situ* small intestine; the effect of lipophilicity. *J Pharm Pharmacol* 1985; 37(4):280–283.
27. Hendriksen BA, Felix MV, Bolger MB. The composite solubility versus pH profile and its role in intestinal absorption prediction. *AAPS PharmSci* 2003;5(1):E4.
28. Lennernas H. Does fluid flow across the intestinal mucosa affect quantitative oral drug absorption? Is it time for a reevaluation? *Pharm Res* 1995;12(11):1573–1582.
29. Anderson BW, Levine AS, Levitt DG, Kneip JM, Levitt MD. Physiological measurement of luminal stirring in perfused rat jejunum. *Am J Physiol* 1988;254(6 Pt 1):G843–G848.
30. Levitt MD, Furne JK, Strocchi A, Anderson BW, Levitt DG. Physiological measurements of luminal stirring in the dog and human small bowel. *J Clin Invest* 1990;86(5): 1540–1547.
31. Tsuji A, Tamai I. Carrier-mediated intestinal transport of drugs. *Pharm Res* 1996;13(7): 963–977.
32. Benet LZ, Cummins CL, Wu CY. Unmasking the dynamic interplay between efflux transporters and metabolic enzymes. *Int J Pharm* 2004;277(1–2):3–9.
33. Fischer V, Einolf HJ, Cohen D. Efflux transporters and their clinical relevance. *Mini Rev Med Chem* 2005;5(2):183–195.
34. Wagner D, Spahn-Langguth H, Hanafy A, Koggel A, Langguth P. Intestinal drug efflux: formulation and food effects. *Adv Drug Deliv Rev* 2001;50(Suppl 1):S13–S31.
35. Suzuki H, Sugiyama Y. Role of metabolic enzymes and efflux transporters in the absorption of drugs from the small intestine. *Eur J Pharm Sci* 2000;12(1):3–12.
36. Kimura T, Higaki K. Gastrointestinal transit and drug absorption. *Biol Pharm Bull* 2002;25(2):149–164.

37. de Vries ME, Bodde HE, Verhoef JC, Junginger HE. Developments in buccal drug delivery. *Crit Rev Ther Drug Carrier Syst* 1991;8(3):271–303.
38. Squier CA. The permeability of oral mucosa. *Crit Rev Oral Biol Med* 1991;2(1):13–32.
39. de Vries ME, Bodde HE, Busscher HJ, Junginger HE. Hydrogels for buccal drug delivery: properties relevant for muco-adhesion. *J Biomed Mater Res* 1988;22(11):1023–1032.
40. Shojaei AH. Buccal mucosa as a route for systemic drug delivery: a review. *J Pharm Pharm Sci* 1998;1(1):15–30.
41. Harris D, Robinson JR. Drug delivery via the mucous membranes of the oral cavity. *J Pharm Sci* 1992;81(1):1–10.
42. Galey WR, Lonsdale HK, Nacht S. The *in vitro* permeability of skin and buccal mucosa to selected drugs and tritiated water. *J Invest Dermatol* 1976;67(6):713–717.
43. Wertz PW, Squier CA. Cellular and molecular basis of barrier function in oral epithelium. *Crit Rev Ther Drug Carrier Syst* 1991;8(3):237–269.
44. Squier CA, Cox P, Wertz PW. Lipid content and water permeability of skin and oral mucosa. *J Invest Dermatol* 1991;96(1):123–126.
45. Chan KK, Gibaldi M. Effects of first-pass metabolism on metabolite mean residence time determination after oral administration of parent drug. *Pharm Res* 1990;7(1): 59–63.
46. Streisand JB, Zhang J, Niu S, McJames S, Natte R, Pace NL. Buccal absorption of fentanyl is pH-dependent in dogs. *Anesthesiology* 1995;82(3):759–764.
47. Narawane LVHL. *Absorption Barriers. Harwood Academic*, Switzerland: AG de Boer; 1994.
48. Kurosaki YNH, Terao K, Nakayama T, Kimura T. Existence of specialized absorption mechanism for cephadroxyl, an aminocephalosporin antibiotic, in the human oral cavity. *Int J Pharm* 1992;82:165–169.
49. Graves NM, Kriel RL. Rectal administration of antiepileptic drugs in children. *Pediatr Neurol* 1987;3(6):321–326.
50. Roelofse JA, van der Bijl P, Stegmann DH, Hartshorne JE. Preanesthetic medication with rectal midazolam in children undergoing dental extractions. *J Oral Maxillofac Surg* 1990;48(8):791–797; discussion 797.
51. Malinovsky JM, Lejus C, Servin F, et al. Plasma concentrations of midazolam after i.v., nasal or rectal administration in children. *Br J Anaesth* 1993;70(6):617–620.
52. Hanning CD. The rectal absorption of opioids. In Benedetti C, Chapman CR, Giron G, Eds. *Advances in Pain Research and Therapy*. Philadelphia: Lippincott-Raven; 1990; 14:259–268.
53. Alternative routes of drug administration—advantages and disadvantages (subject review). American Academy of Pediatrics. Committee on Drugs. *Pediatrics* 1997;100(1): 143–152.
54. van Hoogdalem EJ, de Boer AG, Breimer DD. Pharmacokinetics of rectal drug administration. Part II. Clinical applications of peripherally acting drugs, and conclusions. *Clin Pharmacokinet* 1991;21(2):110–128.
55. Choonara IA. Giving drugs per rectum for systemic effect. *Arch Dis Child* 1987; 62(8):771–772.
56. Khalil SN, Florence FB, Van den Nieuwenhuyzen MC, Wu AH, Stanley TH. Rectal methohexital: concentration and length of the rectal catheters. *Anesth Analg* 1990; 70(6):645–649.
57. van Hoogdalem E, de Boer AG, Breimer DD. Pharmacokinetics of rectal drug administration. Part I. General considerations and clinical applications of centrally acting drugs. *Clin Pharmacokinet* 1991;21(1):11–26.

58. Forbes RB, Vandewalker GE. Comparison of two and ten percent rectal methohexitone for induction of anaesthesia in children. *Can J Anaesth* 1988;35(4):345–349.
59. Jantzen JP, Erdmann K, Witton PK, Klein AM. The effect of rectal pH values on the absorption of methohexital. *Anaesthesist* 1986;35(8):496–499.
60. Jantzen JP, Tzanova I, Witton PK, Klein AM. Rectal pH In children. *Can J Anaesth* 1989;36(6):665–667.
61. Effros RM, Mason GR. Measurements of pulmonary epithelial permeability *in vivo*. *Am Rev Respir Dis* 1983;127(5 Pt 2):S59–S65.
62. Arora P, Sharma S, Garg S. Permeability issues in nasal drug delivery. *Drug Discov Today* 2002;7(18):967–975.
63. Turker S, Onur E, Ozer Y. Nasal route and drug delivery systems. *Pharm World Sci* 2004;26(3):137–142.
64. Illum L. Nasal drug delivery: new developments and strategies. *Drug Discov Today* 2002;7(23):1184–1189.
65. Illum L. Nasal drug delivery—possibilities, problems and solutions. *J Control Release* 2003;87(1–3):187–198.
66. Illum L, Jabbal-Gill I, Hinchcliffe M, Fisher AN, Davis SS. Chitosan as a novel nasal delivery system for vaccines. *Adv Drug Deliv Rev* 2001;51(1–3):81–96.
67. Chavanpatil MD, Vavia PR. The influence of absorption enhancers on nasal absorption of acyclovir. *Eur J Pharm Biopharm* 2004;57(3):483–487.
68. Davis SS, Illum L. Absorption enhancers for nasal drug delivery. *Clin Pharmacokinet* 2003;42(13):1107–1128.
69. Hornof M, Toropainen E, Urtti A. Cell culture models of the ocular barriers. *Eur J Pharm Biopharm* 2005;60(2):207–225.
70. Hosoya K, Tomi M. Advances in the cell biology of transport via the inner blood–retinal barrier: establishment of cell lines and transport functions. *Biol Pharm Bull* 2005;28(1):1–8.
71. Reichl S, Dohring S, Bednarz J, Muller-Goymann CC. Human cornea construct HCC—an alternative for *in vitro* permeation studies? A comparison with human donor corneas. *Eur J Pharm Biopharm* 2005;60(2):305–308.
72. Kim MK, Lee CH, Kim DD. Skin permeation of testosterone and its ester derivatives in rats. *J Pharm Pharmacol* 2000;52(4):369–375.
73. Lopez A, Llinares F, Cortell C, Herraez M. Comparative enhancer effects of Span20 with Tween20 and Azone on the *in vitro* percutaneous penetration of compounds with different lipophilicities. *Int J Pharm* 2000;202(1–2):133–140.
74. Thomas BJ, Finnin BC. The transdermal revolution. *Drug Discov Today* 2004;9(16): 697–703.
75. Kalia YN, Merino V, Guy RH. Transdermal drug delivery. Clinical aspects. *Dermatol Clin* 1998;16(2):289–299.
76. Dezarnaulds G, Fraser IS. Vaginal ring delivery of hormone replacement therapy—a review. *Expert Opin Pharmacother* 2003;4(2):201–212.
77. Buchanan DL, Kurita T, Taylor JA, Lubahn DB, Cunha GR, Cooke PS. Role of stromal and epithelial estrogen receptors in vaginal epithelial proliferation, stratification, and cornification. *Endocrinology* 1998;139(10):4345–4352.
78. Woolfson AD, Malcolm RK, Gallagher R. Drug delivery by the intravaginal route. *Crit Rev Ther Drug Carrier Syst* 2000;17(5):509–555.
79. Alexander NJ, Baker E, Kaptein M, Karck U, Miller L, Zampaglione E. Why consider vaginal drug administration? *Fertil Steril* 2004;82(1):1–12.

80. Hussain A, Ahsan F. The vagina as a route for systemic drug delivery. *J Control Release* 2005;103(2):301–313.
81. van der Bijl P, Van Eyk AD. Comparative *in vitro* permeability of human vaginal, small intestinal and colonic mucosa. *Int J Pharm* 2003;261(1–2):147–152.
82. Furness G. Needle-Free and Auto Injectors—Management Forum Conference. An update on technology and application. 23–24 Feb 2004, London, UK. *IDrugs* 2004;7(4):329–330.
83. Ajmani D. Going from needles to needle-free injectables. *Drug Deliv Technol* 2004; http://www.drugdeliverytech.com/cgi-bin/articles.cgi?idArticle=318
84. Paice JA, Magolan JM. Intraspinal drug therapy. *Nurs Clin North Am* 1991;26(2): 477–498.
85. Dressman JB, Amidon GL, Reppas C, Shah VP. Dissolution testing as a prognostic tool for oral drug absorption: immediate release dosage forms. *Pharm Res* 1998;15(1): 11–22.
86. Loftssona T, Jarvinen T. Cyclodextrins in ophthalmic drug delivery. *Adv Drug Deliv Rev* 1999;36(1):59–79.
87. Kamm W, Hauptmann J, Behrens I, et al. Transport of peptidomimetic thrombin inhibitors with a 3-amidino-phenylalanine structure: permeability and efflux mechanism in monolayers of a human intestinal cell line (Caco-2). *Pharm Res* 2001;18(8):1110–1118.
88. Stenberg P, Luthman K, Ellens H, et al. Prediction of the intestinal absorption of endothelin receptor antagonists using three theoretical methods of increasing complexity. *Pharm Res* 1999;16(10):1520–1526.
89. Casabo VG, Nunez-Benito E, Martinez-Coscolla A, Miralles-Loyola E, Martin-Villodre A, Pla-Delfina JM. Studies on the reliability of a bihyperbolic functional absorption model. II. Phenylalkylamines. *J Pharmacokinet Biopharm* 1987;15(6):633–643.
90. Dowty ME, Dietsch CR. Improved prediction of *in vivo* peroral absorption from *in vitro* intestinal permeability using an internal standard to control for intra- and inter-rat variability. *Pharm Res* 1997;14(12):1792–1797.
91. Camenisch G, Folkers G, van de Waterbeemd H. Shapes of membrane permeability–lipophilicity curves: extension of theoretical models with an aqueous pore pathway. *Eur J Pharm Sci* 1998;6(4):325–329.
92. Norinder U, Osterberg T, Artursson P. Theoretical calculation and prediction of intestinal absorption of drugs in humans using MolSurf parametrization and PLS statistics. *Eur J Pharm Sci* 1999;8(1):49–56.
93. Palm K, Luthman K, Ungell AL, Strandlund G, Artursson P. Correlation of drug absorption with molecular surface properties. *J Pharm Sci* 1996;85(1):32–39.
94. Clark DE. Rapid calculation of polar molecular surface area and its application to the prediction of transport phenomena. 2. Prediction of blood–brain barrier penetration. *J Pharm Sci* 1999;88(8):815–821.
95. Clark DE. Rapid calculation of polar molecular surface area and its application to the prediction of transport phenomena. 1. Prediction of intestinal absorption. *J Pharm Sci* 1999;88(8):807–814.
96. Roskos KV, Maibach HI. Percutaneous absorption and age. Implications for therapy. *Drugs Aging* 1992;2(5):432–449.
97. Lotte C, Wester RC, Rougier A, Maibach HI. Racial differences in the *in vivo* percutaneous absorption of some organic compounds: a comparison between black, Caucasian and Asian subjects. *Arch Dermatol Res* 1993;284(8):456–459.
98. Gean CJ, Tur E, Maibach HI, Guy RH. Cutaneous responses to topical methyl nicotinate in black, Oriental, and Caucasian subjects. *Arch Dermatol Res* 1989;281(2):95–98.

99. Kompaore F, Marty JP, Dupont C. *In vivo* evaluation of the stratum corneum barrier function in blacks, Caucasians and Asians with two noninvasive methods. *Skin Pharmacol* 1993;6(3):200–207.
100. Tsai JC, Sheu HM, Hung PL, Cheng CL. Effect of barrier disruption by acetone treatment on the permeability of compounds with various lipophilicities: implications for the permeability of compromised skin. *J Pharm Sci* 2001;90(9):1242–1254.
101. Gould AR, Sharp PJ, Smith DR, et al. Increased permeability of psoriatic skin to the protein, plasminogen activator inhibitor 2. *Arch Dermatol Res* 2003;295(6):249–254.

# ABSORPTION OF DRUGS AFTER ORAL ADMINISTRATION

Luis Granero and Ana Polache
*University of Valencia, Valencia, Spain*

**Contents**

1 Introduction
2 Anatomy and Physiology of the Human Gastrointestinal Tract
  2.1 Basic Structure
  2.2 Stomach
  2.3 Small Intestine
  2.4 Large Intestine
  2.5 Liver, Gallbladder, and Pancreas
  2.6 Structure and Composition of Intestinal Membrane
3 Pathways of Drug Absorption
  3.1 Transcellular Pathway
  3.2 Paracellular Pathway
4 Factors Affecting Drug Intestinal Absorption
  4.1 Physicochemical Factors Affecting Gastrointestinal Drug Absorption
  4.2 Physiological Factors Affecting Gastrointestinal Drug Absorption
  4.3 Technological Factors Affecting Gastrointestinal Drug Absorption
5 Intestinal Permeability Enhancement: Possibilities and Limitations
  5.1 Structure and Function of the Intestinal Paracellular Pathway
  5.2 Intestinal Paracellular Enhancers
  5.3 Technologies in Development
  5.4 Limitations of Intestinal Promotion of Drugs
6 Methodologies for Studying Intestinal Absorption
  6.1 *In Vitro* Techniques
  6.2 *In Situ* Techniques
  6.3 *In Vivo* Techniques
  References

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

## 1 INTRODUCTION

Drugs are most commonly given orally and the oral route is the preferred one for the administration of new therapeutic agents. After oral administration, drugs must be absorbed through the gastrointestinal tract to achieve the systemic circulation and exert their pharmacological effects. The successful formulation of an optimized oral drug delivery system requires a detailed consideration and a good understanding of the intestinal absorption process, its possibilities and limitations. In fact, peroral delivery of new hydrophilic drugs, which include macromolecules frequently with null or scarce capacity of absorption through the intestinal barrier, is one of the greatest challenges in biopharmaceutical research.

Recent years have seen rapid developments in the field of intestinal absorption: more detailed insights into the structure and organization of intestinal mucosa (e.g., increased knowledge of tight junction physiology and regulation), new approaches to overcoming the problems associated with poor intestinal drug absorption (e.g., the use of paracellular enhancers), and novel methods for assessing the permeability of intestinal mucosa.

The goal of the present chapter is to give an overview and update on the concepts, possibilities, and limitations of drug absorption after oral administration. First, we introduce the anatomical and physiological aspects of the gastrointestinal tract, since they are decisive to understanding the absorption processes, and analyze the possible pathways of drug intestinal absorption. Second, the primary factors that influence oral drug absorption (e.g., physiological, physicochemical, and technological variables) are covered. Third, we analyze the use of permeation enhancers as a way to increase the rate and extent of drugs across the intestinal barrier. Fourth, we review the different methodologies (*in vitro, in situ*, and *in vivo*) that allow an investigator to establish the suitability of a drug candidate and to solve problems associated with drug intestinal absorption.

## 2 ANATOMY AND PHYSIOLOGY OF THE HUMAN GASTROINTESTINAL TRACT

The gastrointestinal tract (GIT) consists of a hollow muscular tube starting from the oral cavity, where food enters the mouth, continuing through the pharynx, esophagus, stomach, and intestines, to the rectum and anus, where food is expelled. There are various accessory organs that assist the tract by secreting enzymes to help break down food into its component nutrients. Thus, the salivary glands, liver, pancreas, and gallbladder have important functions in the digestive system. Food is propelled along the length of the GIT by peristaltic movements of the muscular walls.

The primary purpose of the GIT is to break down food into nutrients, which can be absorbed into the body to provide energy. First, food must be ingested into the mouth to be mechanically processed and moistened. Second, digestion occurs mainly in the stomach and small intestine, where proteins, fats, and carbohydrates are chemically broken down into their basic building blocks. Smaller molecules are then absorbed across the epithelium of the small intestine and subsequently enter the circulation. The large intestine plays a key role in reabsorbing excess water. Finally,

undigested material and secreted waste products are excreted from the body via defecation (passing of feces).

### 2.1 Basic Structure

The GIT is a muscular tube lined by a special layer of cells, called epithelium. The contents of the tube are considered external to the body and are in continuity with the outside world at the mouth and the anus. Although each section of the tract has specialized functions, the entire tract has a similar basic structure with regional variations. The wall is divided into four layers as follows (beginning with the luminal surface): (1) mucosa, (2) submucosa, (3) muscularis externa, and (4) serosa (mesentery). The three outer layers are similar throughout most of the tract; however, the mucosa has distinctive structural and functional characteristics.

1. *Mucosa.* The innermost layer of the digestive tract has specialized epithelial cells supported by an underlying connective tissue layer called the lamina propria. The lamina propria contains blood vessels, nerves, lymphoid tissue, and glands that support the mucosa. Depending on its function, the epithelium may be simple (a single layer) or stratified (multiple layers). Areas such as the mouth and esophagus are covered by a stratified squamous (flat) epithelium so they can survive the wear and tear of passing food. Simple columnar (tall) or glandular epithelium lines the stomach and intestines to aid secretion and absorption. The inner lining is constantly shed and replaced, making it one of the most rapidly dividing areas of the body. Beneath the lamina propria is the muscularis mucosa. This comprises layers of smooth muscle, which can contract to change the shape of the lumen.
2. *Submucosa.* The submucosa surrounds the muscularis mucosa and consists of fat, fibrous connective tissue, and larger vessels and nerves. At its outer margin there is a specialized nerve plexus called the submucosal plexus or Meissner plexus. This supplies the mucosa and submucosa.
3. *Muscularis Externa.* This smooth muscle layer has inner circular and outer longitudinal layers of muscle fibers separated by the myenteric plexus or Auerbach plexus. Neural innervations control the contraction of these muscles and hence the mechanical breakdown and peristalsis of the food within the lumen.
4. *Serosa (Mesentery).* The outer layer of the GIT is formed by fat and another layer of epithelial cells called mesothelium.

### 2.2 Stomach

The stomach is a J-shaped expanded bag, located just left of the midline between the esophagus and small intestine. It is divided into four main regions and has two borders called the greater and lesser curvatures. The first section is the *cardia*, which surrounds the cardial orifice where the esophagus enters the stomach. The *fundus* is the superior, dilated portion of the stomach that has contact with the left dome of the diaphragm. The *body* is the largest section between the fundus and the curved portion of the J. This is where most gastric glands are located and where most mixing

of the food occurs. Finally, the *pylorus* is the curved base of the stomach. Gastric contents are expelled into the proximal duodenum via the pyloric sphincter. The inner surface of the stomach is contracted into numerous longitudinal folds called *rugae*. These allow the stomach to stretch and expand when food enters. The stomach can hold up to 1.5 liters of material.

The functions of the stomach include (1) the short-term storage of ingested food, (2) mechanical breakdown of food by churning and mixing motions, (3) chemical digestion of proteins by acids and enzymes, and (4) killing bugs and germs by acidification. Most of these functions are achieved with the aid of the secretion of stomach juices by gastric glands in the body and fundus. Some cells are responsible for secreting acid and others secrete enzymes to break down proteins.

### 2.3 Small Intestine

The small intestine is composed of the duodenum, jejunum, and ileum. It averages approximately 6 m in length, extending from the pyloric sphincter of the stomach to the ileocecal valve separating the ileum from the cecum. The small intestine is compressed into numerous folds and occupies a large proportion of the abdominal cavity. The duodenum is the proximal C-shaped section that curves around the head of the pancreas. The duodenum serves a mixing function as it combines digestive secretions from the pancreas and liver with the contents expelled from the stomach. The start of the jejunum is marked by a sharp bend, the duodenojejunal flexure. It is in the jejunum where the majority of digestion and absorption occurs. The final portion, the ileum, is the longest segment and empties into the cecum at the ileocecal junction.

The small intestine performs the majority of digestion and absorption of nutrients. Partly digested food from the stomach is further broken down by enzymes from the pancreas and bile salts from the liver and gallbladder. These secretions enter the duodenum at the ampulla of Vater. After further digestion, food constituents such as proteins, fats, and carbohydrates are broken down to small building blocks and absorbed into the body.

The lining of the small intestine (Fig. 1) is made up of numerous permanent folds called plicae circulares or folds of Kerckring. Each plica has numerous villi (folds of mucosa) and each villus is covered by epithelium containing several specialized cells; some are responsible for absorption, the enterocytes, while others secrete digestive enzymes and mucus to protect the intestinal lining from digestive actions. Enterocytes in the small intestine are cells clearly specialized in absorption, having in their luminal pole the so-called microvilli (brush border). This anatomical structure is designed to increase the surface area for absorption by a factor of several hundred.

All the structure in the small intestine is conceived to increase effective surface for absorption (Fig. 2). The initial increase in surface area is due to the projection within the lumen of folds of mucosa. As indicated earlier, lining the entire epithelial surface are finger-like projections, the villi. These villi range in length from 0.5 to 1.5 m, and it has been estimated that there are about 10–40 villi/mm$^2$ of mucosal surface. Projecting from the villi surface are the microvilli, which represent the final large increase in the surface area of the small intestine. There are approximately 600 microvilli protruding from each enterocyte lining the villi. Relative to the initial

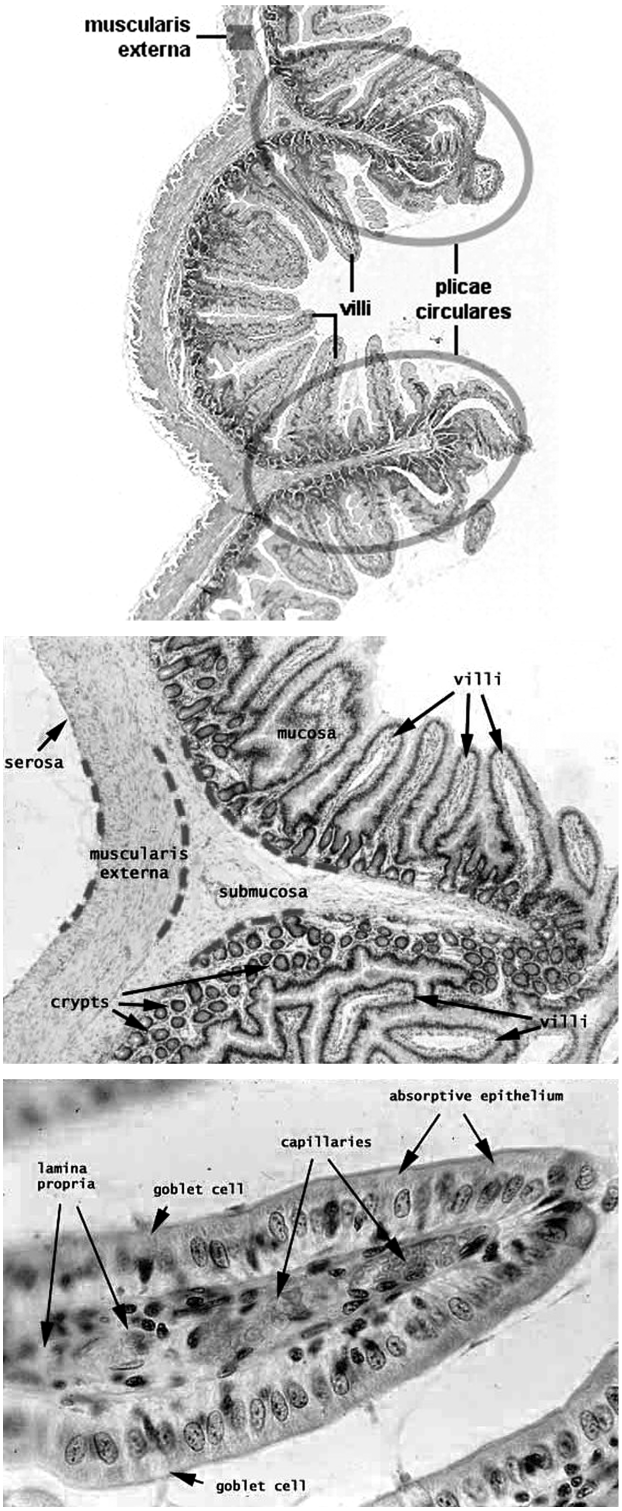


**FIGURE 1** Photomicrographs showing the structure of the small intestine (jejunum). Magnification increases from the upper to the lower panel.

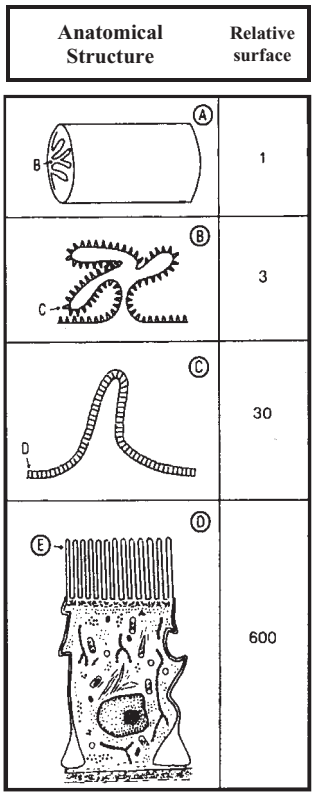


**FIGURE 2** Schematic representation of the increase in the effective surface of the small intestine mucosa, relative to a simple cylinder, as a consequence of its particular structural features: Ⓐ simple cylinder, Ⓑ plicae circulares, Ⓒ villi, Ⓓ enterocyte, and Ⓔ microvilli.

surface of a smooth cylinder, the folds, villi, and microvilli increase the effective surface area by factors of 3, 30, and 600, respectively.

## 2.4 Large Intestine

The large intestine is horseshoe shaped and extends around the small intestine like a frame. It consists of the appendix, cecum, ascending, transverse, descending, and sigmoid colon, and the rectum. It has a length of approximately 1.5 m and a width of 7.5 cm. The cecum is the expanded pouch that receives material from the ileum and starts to compress food products into fecal material. Food then travels along the colon. The wall of the colon is made up of several pouches (haustra) that are held under tension by three thick bands of muscle (taenia coli). The rectum is the final 15 cm of the large intestine. It expands to hold fecal matter before it passes through the anorectal canal to the anus. Thick bands of muscle, known as sphincters, control the passage of feces.

The mucosa of the large intestine lacks villi seen in the small intestine. Thus, the effective surface area for absorption is clearly lower than that existing in the small intestine. The mucosal surface is flat with several deep intestinal glands. Numerous

goblet cells line the glands that secrete mucus to lubricate fecal matter as it solidifies.

The functions of the large intestine can be summarized as (1) the accumulation of unabsorbed material to form feces, (2) some digestion by bacteria the bacteria responsible for the formation of intestinal gas, and (3) reabsorption of water, salts, sugar, and vitamins.

### 2.5 Liver, Gallbladder, and Pancreas

The liver, gallbladder, and pancreas, although not part of the gut, have been included, since these organs secrete materials vital to the digestive and certain absorptive functions of the gut.

The liver is a large, reddish-brown organ situated in the right upper quadrant of the abdomen. It is surrounded by a strong capsule and divided into four lobes, namely, the right, left, caudate, and quadrate lobes. The liver has several important functions. It acts as a mechanical filter by filtering blood that travels from the intestinal system. It detoxifies several drugs and endogenous metabolites including the breakdown of bilirubin and estrogen. In addition, the liver has synthetic functions, producing albumin and blood clotting factors. However, its main roles in digestion are in the production of bile and metabolism of nutrients. All drugs and nutrients absorbed by the intestines pass through the liver and are processed before traveling to the rest of the body. The bile produced by cells of the liver enters the intestines at the duodenum. Here, bile salts break down lipids into smaller particles so there is a greater surface area for digestive enzymes to act.

The gallbladder is a hollow, pear-shaped organ that sits in a depression on the posterior surface of the liver's right lobe. It consists of a fundus, body, and neck. It empties via the cystic duct into the biliary duct system. The main functions of the gallbladder are storage and concentration of bile. Bile is a thick fluid that contains enzymes to help dissolve fat in the intestines. Bile is produced by the liver but stored in the gallbladder until it is needed. Bile is released from the gallbladder by contraction of its muscular walls in response to hormone signals from the duodenum in the presence of food.

Finally, the pancreas is a lobular, pinkish-grey organ that lies behind the stomach. Its head communicates with the duodenum and its tail extends to the spleen. The organ is approximately 15 cm in length with a long, slender body connecting the head and tail segments. The pancreas has both exocrine and endocrine functions. Endocrine refers to production of hormones, which occurs in the islets of Langerhans. The islets produce insulin, glucagon, and other substances and these are the areas damaged in diabetes mellitus.

The exocrine (secretory) portion makes up 80–85% of the pancreas and is the area relevant to the gastrointestinal tract. It is made up of numerous acini (small glands) that secrete contents into ducts, which eventually lead to the duodenum. The pancreas secretes fluid rich in carbohydrates and inactive enzymes. Secretion is triggered by the hormones released by the duodenum in the presence of food. Pancreatic enzymes include carbohydrases, lipases, nucleases, and proteolytic enzymes that can break down different drugs and components of food. These are secreted in an inactive form to prevent digestion of the pancreas itself. The enzymes become active once they reach the duodenum.

## 2.6 Structure and Composition of Intestinal Membrane

Drug absorption is ultimately the penetration of the drug across the intestinal membrane and its appearance unchanged in the blood draining the GIT. The term intestinal membrane is misleading since, as we commented earlier, this membrane is not a unicellular structure, but really a number of unicellular membranes parallel to one another and separated by aqueous fluid regions bounded by these membranes. Whatever the case, drug absorption implies a movement of drug molecules across biological membranes and, therefore, it could be useful to remember their basic structure and composition.

The biological membranes are bilipid layers. In a real cell the membrane phospholipids create a spherical three-dimensional lipid bilayer shell around the cell. The hydrophobic hydrocarbon chains of the phospholipids orient toward each other, creating a hydrophobic environment within the membrane. This leaves the charged phosphate groups facing out into the hydrophilic environment. The membrane is approximately 5 nm thick. This bilipid layer is semipermeable, meaning that some molecules are allowed to pass freely (diffuse) through the membrane. Molecules can diffuse through the membrane at differing rates depending on their ability to enter the hydrophobic interior of the membrane bilayer. The most accepted biological membrane model is referred to as the *fluid mosaic model* [1]. In this model lipid bilayers are fluid, and individual phospholipids diffuse rapidly throughout the two-dimensional surface of the membrane. These fluid bilayers include proteins, cholesterol, and other types of molecules besides phospholipids. Membrane proteins diffuse throughout the membrane in the same fashion, although at a slower pace because of their massive size (a phospholipid may be 650 daltons, and a medium sized protein can be 100,000 daltons). From time to time a given phospholipid will "flip-flop" through the membrane to the opposite side, but this is uncommon. To do so required the hydrophilic head of the phospholipid to pass fully through the highly hydrophobic interior of the membrane, and for the hydrophobic tails to be exposed to the aqueous environment.

There are also molecules of cholesterol embedded in the membrane. Cholesterol is a necessary component of biological membranes. Cholesterol breaks up the van der Waals interactions and close packing of the phospholipid tails. This disruption makes the membrane more fluid. Therefore, one way for a cell to control the fluidity of its membrane is by regulating its level of cholesterol in the cell membrane. Another way is to regulate the ratio of saturated to unsaturated hydrocarbon chains of the phospholipids. A group of phospholipids with saturated hydrocarbon chains can pack close together and form numerous van der Waals bonds that hold the phospholipids to each other. Phospholipids with unsaturated hydrocarbon side chains break up those van der Waals bonds and the tight packing by preventing the phospholipids from getting close together.

The cell membrane plays host to a large amount of protein, which is responsible for its various activities. The amount of protein differs between species and according to function; however, the typical amount in a cell membrane is 50%. These proteins are undoubtedly important to a cell. Three groups of membrane proteins can be identified:

1. *Integral Proteins.* They are located spanning the membrane and thus have a hydrophilic cytosolic domain that interacts with internal molecules, a hydrophobic membrane-spanning domain that anchors it within the cell membrane, and a hydro-

philic extracellular domain that interacts with external molecules. Examples of these integral proteins, also known as transmembrane proteins, are ion channels, proton pumps, transport proteins and efflux transporters, such as P-glycoprotein, and G protein-coupled receptors.

2. *Lipid Anchored Proteins.* They are located covalently bound to single or multiple lipid molecules, which hydrophobically insert into the cell membrane and anchor the protein. The protein itself is not in contact with the membrane. Examples are G proteins.

3. *Peripheral Proteins.* They are attached to integral proteins, or associated with the head groups of membrane lipids. As such, these proteins are not in contact with the hydrophobic membrane core and therefore associate only at the cytosolic and extracellular faces. They tend to have only temporary interactions with permanent membrane proteins, and once reacted, the molecule dissociates to carry on its work in the cytoplasm. Examples are some enzymes and some hormones.

It is also important to remember that the cell membrane, being exposed to the outside environment, is an important site of cell communication. As such, a large variety of protein receptors and identification proteins, such as antigens, are present on the surface of the membrane.

## 3 PATHWAYS OF DRUG ABSORPTION

For oral drugs to be therapeutically effective, they have to possess favorable characteristics to cross the biological membrane into the systemic circulation and reach the site of action: that is, they must be absorbed. Once a drug molecule is in solution, it has the possibility to be absorbed. At the intestinal epithelial cell surface, there are two pathways that are potentially available to molecules. Transepithelial transport of drugs can be achieved either across the cell (*transcellular* or *intracellular pathway*) or through the junctions that hold the cells together (*paracellular* or *intercellular pathway*) (Fig. 3). Historically, the transcellular pathway has received more

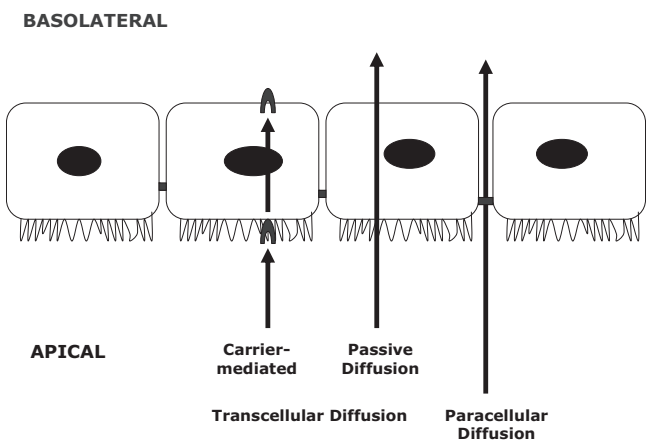


**FIGURE 3** Schematic representation of the potential transcellular and paracellular pathways of drug intestinal absorption.

attention. However, the increasing understanding of tight junction physiology and regulation has shed light on the paracellular route.

### 3.1 Transcellular Pathway

In the transcellular or intracellular route, drugs are absorbed by passive diffusion, facilitated diffusion, or an active transport system (Fig. 3). Since these absorption mechanisms are discussed in Chapter 8, we will not dwell upon it in this chapter.

*Passive diffusion* (linear or first-order kinetics) is the predominant pathway taken by most oral drugs. Passive diffusion indicates that the transfer of a compound from an aqueous phase (intestinal lumen) through a membrane may be described by physicochemical laws and by the properties of the membrane. The driving force for diffusion across the membrane is the gradient concentration of the drug across the membrane. It is essential for a molecule to have characteristics with low molecular size and relatively high lipophilicity in order to pass across the intestinal membrane by this pathway. Factors influencing transcellular passive diffusion in the GIT are well studied and are reviewed here.

*Facilitated diffusion* is a process in which a molecule specifically and reversibly binds to its carrier protein in the enterocyte membrane. It crosses the membrane associated with the carrier and is then released to the cytoplasm of the cell. As with passive diffusion, facilitated diffusion can only occur down a gradient of chemical potential, since no energy source is required. The process follows the kinetic of enzymes, being saturable and subject to inhibition by competitive inhibitors. Actually, there are not many unequivocal examples of facilitated diffusion for drug molecules.

*Active transport* of drugs requires both membrane associated carrier proteins and an energy source, in order that the drug molecules could be transported up a gradient of concentration. Active transport satisfies the following criteria: (1) it is saturable, (2) metabolic inhibitors inhibit transport, and (3) substrate analogues compete with the drug for the active site on the carrier [2]. Absorption by specialized carrier mechanism from the small intestine has been shown to exist for several drugs having chemical structures that are very similar to essential nutrients for which the intestine has a specialized transport mechanism [3].

In direct contrast with specialized transport of a drug into epithelial cells is a process refered to as *efflux transport*, which implies the facilitated movement of a drug out of the cell. Today, it is well demonstrated that this mechanism exists in epithelial cells of the gastrointestinal tract. There is a cell surface glycoprotein (P-glycoprotein, P-gp) that is responsible for this mechanism. P-gp is believed to be responsible for poor intestinal uptake of several classes of compounds. The P-gp protein could be inhibited and the process may be saturated. The former is the basis for many drug–drug and drug–nutrients interactions, and the latter limits the efficacy of the transporter. Both situations may have repercussions in the intestinal absorption of drugs.

### 3.2 Paracellular Pathway

The paracellular route can be defined as the aqueous pathway along the intercellular space between adjacent cells, which is restricted by tight junctions (TJs) at the most apical part of the cells (Fig. 3). The aqueous nature of this route makes it favorable

for the absorption (through passive diffusion) of small hydrophilic solutes (nutrients, ions, etc.). TJs are dynamic structures, which normally regulate the trafficking of nutrients, medium sized compounds (≤15 Å), and relatively large amounts of fluids between the intestinal lumen and the submucosa [4]. Tight junction permeability differs according to the type of permeant involved, its size, and its charge. The molecular size cutoff for this route is approximately 500 Da [5].

Paracellular transport was belittled for much of the last century and was not considered an effective pathway until the 1960s, when it was discovered that molecules of different sizes can cross the epithelium via the paracellular route [6]. Today, it is well known that the flux across the TJs can be considerable, and that the permeability and selectivity of the junctions can be regulated [7]. Thus, in the last decades many efforts have been made to increase paracellular transport of less absorbable drugs—the use of enhancers being the most promising and studied strategy [8]. This topic is covered in Section 5, including an overview of the structure and function of this pathway.

## 4 FACTORS AFFECTING DRUG INTESTINAL ABSORPTION

Factors affecting gastrointestinal drug absorption have been classified into three broad categories. These have been summarized in Table 1. In the following, we analyze in a brief manner their influence on intestinal drug absorption.

### 4.1 Physicochemical Factors Affecting Gastrointestinal Drug Absorption

The primary physicochemical properties of a drug influencing absorption across gastrointestinal membrane are solubility, particle size, crystal form, lipophilicity, dissociation constant, and molecular weight.

***Solubility*** Low aqueous solubility is a frequently encountered reason for poor oral absorption. Absorption requires that drug molecules be in solution at the absorption site. For that, drug molecules contained in the dosage forms must be delivered and then they must dissolve in the gastrointestinal fluids at the absorption site. Dissolution depends in part on the solubility of the drug. Together with solubility, the rate of dissolution is another variable to consider. In fact, there are cases of compounds with very low aqueous solubility, which have fully adequate oral bioavailability. This can occur because the rate of dissolution being a rapid phenomenon in spite of the overall solubility being low.

**TABLE 1 Factor Affecting Gastrointestinal Drug Absorption**

| Physicochemical Factors | Physiological Factors | Technological Factors |
|---|---|---|
| Solubility<br>Particle size<br>Crystal form<br>Lipophilicity<br>Dissociation constant | Gastric emptying and GIT motility<br>Degradation and excretion processes in the GIT | Pharmaceutical dosage form<br>Manufacturing processes |

Polar solutes are more soluble in water than in organic media, whereas the opposite is true for nonpolar solutes. When the drug is ionizable, ionized species are more soluble in water than their un-ionized counterparts. On the other hand, solubility of weak acids or bases in aqueous medium is dependent on pH. For these compounds with acidic or basic functional groups, both solubility and rate of dissolution can be improved by suitable choice of salt forms.

The presence of charged groups can aid in solubility. However, charge, particularly high charge density, can have a negative impact on drug absorption since charged species do not readily penetrate the lipoidal membrane (*vide infra*). In such cases the reduction in charge by chemical modification can be a useful strategy. In other cases, charge also can have a negative impact on drug absorption by forming water-insoluble complexes or salts with the gastrointestinal contents.

***Particle Size*** Surface area of drug particles influences the rate of drug dissolution and, therefore, drug absorption. Particle size determines the surface area of the solid. Small particles have greater surface area than larger particles and, consequently, they dissolve more rapidly. One important exception is constituted by the hydrophobic drugs. In this case, some examples have shown that the dissolution rate can increase with increasing particle size. For example, the rate of dissolution of the hydrophobic drug phenacetin increases as particle size increases from 0.11–0.15 mm to 0.50–0.71 mm [9]. It is probable that decreasing particle size of a hydrophobic drug actually decreases its effective surface area (i.e., the portion of surface without adsorbed air and actually in contact with dissolving fluid). In these cases, the addition of a surface-active agent significantly increases the effective surface area, improving the rate of dissolution and therefore the rate and extent of absorption.

Particle size appears to have little influence on the absorption of high aqueous solubility drug molecules. However, absorption of drugs such as griseofulvin, a molecule with low solubility in water, is highly dependent on the particle size. In these cases, strategies that increase particle surface area will result in improvement of drug absorption. The effective surface-area may be increased by physically reducing the particle size or by adding surface-active agents to the dosage form.

***Crystal Form*** The physicochemical properties of crystal forms are influenced by the intermolecular forces present. Crystals with weak attractive forces between molecules exhibit greater solubility than those with strong attractive forces. In this sense, solubility could be dependent on the crystal form. There are a number of drugs with different crystal forms (polymorphs) and, in these cases, they can show differences in dissolution rates and, consequently, in absorption. For example, the rate of absorption of chloramphenicol appears to be directly related to the solubility of the different polymorphs of its palmitate ester [10].

***Lipophilicity*** Gastrointestinal membranes are lipoidal in nature; therefore, they are more permeable to lipid-soluble drug substances. Lipid solubility of the diffusing species will influence, in part, passive diffusion across biological membranes. Lipid solubility of a drug is determined by the presence of nonpolar groups in the structure of the drug molecule, as well as by the presence of ionizable groups that are affected by local pH. It is logical to suppose that when an ionizable group exists in the drug molecule, it is very important, in order to improve lipophilicity, that this

group would be in the un-ionized state. High lipid solubility values must be accompanied, however, by adequate water solubility. When the water solubility is too low, a significant concentration of the drug molecule cannot be achieved at the membrane surface and absorption may be inefficient in spite of favorable lipid solubility.

The relative lipophilic to hydrophilic balance of the entire molecule can be described by the oil/water partition coefficient ($K_{o/w}$). In general, this parameter can be used as a predictor of the ability of a drug molecule to be absorbed by passive diffusion across lipoidal membranes. As $K_{o/w}$ increases, the rate of absorption increases. However, there are several exceptions to this general rule. Highly branched compounds are absorbed more slowly and small polar molecules more readily than would be expected based on their $K_{o/w}$ values.

The absorption of a hydrophilic drug may often be enhanced through appropriate structural modifications that increase $K_{o/w}$ of the compound (e.g., esterification). For example, esterification of one of the carboxylic groups of enalaprilat results in a significant increase in oral bioavailability [11].

***Dissociation Constant*** The presence of ionizable groups in the drug molecule can condition the absorption across biological membranes. The importance of ionization in drug absorption is based on the observation that the nonionized form of the drug has a greater $K_{o/w}$ than the ionized form; therefore, the nonionized form of the drug in solution penetrates lipoidal membranes of the gastrointestinal tract more rapidly than the ionized species. The rate of absorption of an ionizable drug is therefore dependent on the concentration of its nonionized species at the absorption site, which is, as predicted by the Henderson–Hasselbalch equation, a function of the $pK_a$ of the compound and the pH of the medium.

$$\text{For acidic drugs:} \quad f = \frac{1}{1+10^{(\mathrm{pH}-\mathrm{p}K_a)}} \tag{1}$$

$$\text{For basic drugs:} \quad f = \frac{1}{1+10^{(\mathrm{p}K_a-\mathrm{pH})}} \tag{2}$$

where $f$ is the nonionized fraction.

The pH of the gastrointestinal tract ranges from 1.2 to 3.5 in the stomach, 5.0 to 6.0 in the duodenum, and 6.5 to 8.0 in the jejunum and large intestine. For acidic drugs with $pK_a$ values between 2.5 and 7.5, the un-ionized fraction $f$ decreases with increases in pH. The same is true for bases with $pK_a$ values between 5 and 11.

A number of studies have related and quantified the influence of pH and $pK_a$ on drug absorption in the gastrointestinal tract. These studies resulted in the so-called pH-partition theory. Briefly, this theory states that only the nonionized form of an ionizable drug is able to penetrate biological membranes because only this form has an adequate $K_{o/w}$. As a result, acidic drugs should best be absorbed from media with $\mathrm{pH} < \mathrm{p}K_a$, whereas basic compounds would best be absorbed from media with $\mathrm{pH} > \mathrm{p}K_a$. This does not mean that acidic drugs are best absorbed in the stomach and basic drugs best absorbed in small intestine. Even though the pH-partition theory provides a useful guide in predicting general trends in passive drug absorption, some

inconsistencies have been observed. The primary limitation of this theory derives from the assumption that only nonionized drug is absorbed, when in fact the ionized form of low to medium molecular weight drugs can be absorbed by passive diffusion through the pores (actually known as the paracellular route, *vide supra*), albeit at a slower rate. In fact, some models of intestinal drug absorption [12] propose that for ionized compounds, with low to medium molecular weight (below 250), diffusion through this route is the major contributor of absorption. For extremely low molecular weight, highly hydrophilic compounds, this route may be the only one. As the molecular weight and, consequently, the $K_{o/w}$ increase, passage through the transcellular route (i.e., through the lipoidal membrane) will predominate. For intermediate molecular weight drugs, diffusion using both routes will be possible.

### 4.2 Physiological Factors Affecting Gastrointestinal Drug Absorption

Once a drug molecule is in solution, it has the potential to be absorbed. However, there are a number of variables, other than the physicochemical ones, that can condition the absorption process. In the following sections we review the main physiological factors related to drug absorption.

***Gastric Emptying and GIT Motility*** When an oral dosage form is swallowed by a patient, it travels through the GIT, starting this travel in the stomach. At this point, the dosage form must break down into small granules and particles and release the drug in order to facilitate its dissolution in the gastrointestinal fluids. Because the stomach has a reduced membrane surface area relative to small intestine, very frequently, the rate of drug absorption in this area is very low. So the rate at which the drug gets to the small intestine can condition the rate of absorption. Gastric emptying is the normal physiological process that controls the progression of the gastric contents toward the duodenum. Knowledge of the factors controlling gastric emptying is important because this process can control the rate of absorption in the small intestine.

There is a general consensus that gastric emptying (measured from the remaining volumes in the stomach after ingestion) follows first-order kinetics [13] especially when liquid or small volumes of semisolids are ingested. The half-life of the process ranges from 10 to 60 min in the case of fluids or semisolids, whereas this time increases up to 4 hours when nondisintegrating solids are ingested. Since, for the majority of the drugs, absorption occurs in the small intestine, the start of the absorption process will be conditioned by gastric emptying. When the intrinsic absorbability of the drug is elevated, gastric emptying become a rate-determining step of drug absorption. For example, paracetamol has good absorbability in the small intestine. Clements et al. [14] showed that, when this drug is orally administered, the absorption rate constant is highly correlated with the rate of gastric emptying.

Gastric emptying rate can be modified by several factors. For example, light physical activity stimulates gastric emptying, whereas strenuous exercise delays it. The volume of the ingested meal is another important factor conditioning gastric emptying. So Hunt and Macdonald [15] found that the half-life of gastric emptying increased from 7 min for a 50 mL standard meal to 50 min for a 1250 mL ingestion. The viscosity of the meal also conditions the rate of gastric emptying. Liquids abandon the stomach faster than semisolids. Nondigestible solids with size ≥ 2 mm

in diameter are handled by the stomach quite differently from liquids and will remain in the stomach for a long period (up to 12 hours). Concurrent drug therapy may also affect the rate of gastric emptying. Thus, cholinergic drugs increase the rate of emptying [16].

Delays in stomach emptying may decrease the rate of availability of a drug. The extent of availability, however, may be increased, decreased, or unaffected. For poorly water-soluble drugs, the extent of availability can be significantly reduced when the drug is ingested under fasting conditions. So Welling et al. [17] showed that erythromycin availability was reduced when 250 mg tablets were taken on a fasting stomach relative to that obtained when tablets were ingested after high fat, high protein, and high carbohydrate meals.

Concurrent drug therapy may increase or decrease intestinal motility. Such changes may increase, decrease, or have no effect on the extent of availability of the drug. So the concomitant use of metoclopramide, a drug that increases the gastrointestinal motility, can reduce the extent of availability of drugs with low values of rate of dissolution. For example, digoxin bioavailability is reduced in the presence of metoclopramide probably due to the increase in intestinal motility, so there is insufficient time for the release of digoxin from its dosage form and the subsequent dissolution in the gastrointestinal fluids before it abandons the GIT [18]. However, digoxin bioavailability increases in the presence of propantheline, a drug that slows gastrointestinal motility [18].

***Degradation and Excretion Processes in the Gastrointestinal Tract*** Several processes in the GIT can reduce the extent of availability of a drug administered by the oral route. The main processes that can reduce the extent of availability are given in Table 2.

Some drugs can interact with endogenous or exogenous substances present in the gastrointestinal fluids, forming insoluble complexes. For example, biliary salts can interact with drugs such as neomycin and kanamycin, forming insoluble and nonabsorbable complexes.

Chemical degradation, especially pH-dependent reactions, can occur in the gastrointestinal fluids. Several polypeptides, nucleotides, or fatty acids may be susceptible to enzymatic degradation by several enzymes present in the gastrointestinal fluids. The role of the gastrointestinal microflora on the metabolism of drugs has long been recognized. In humans less than 1000 organisms/mL are usually found in the gastric juice. Only if the pH is high (>4) are relatively large numbers of bacteria found. The bacterial flora in the proximal jejunum is sparse, with increasing number in the distal small intestine. The number of organisms increases markedly in the

**TABLE 2 Factors Reducing the Extent of Drug Absorption**

| Factors Acting in the Lumen of the GIT (Before Membrane Penetration) | Factors Acting After Membrane Penetration |
|---|---|
| Adsorption and nonsoluble complex formation | Gut wall metabolism |
| Chemical degradation | Gut secretion by efflux transporters |
| Enzymatic degradation | |
| Bacterial degradation | |

large intestine. Microorganisms are capable of carrying out a multitude of reactions, and some of these reactions may have toxicological significance. Most of the metabolic transformations mediated by the microflora fall into the category of hydrolytic reactions, reactions involving the removal of various groups (e.g., dehydroxylations, descarboxylations, dealkylations) and reductive reactions.

Many drugs and toxic compounds that enter the body via the GIT suffer metabolism in the mucosal cells. In fact, enterocytes in the upper segment of the small intestine express, at high concentrations, two forms of cytochrome P450 (CYP3A4 and CYP3A5) that can potentially limit the availability of drugs [19]. Oxidation reactions mediated by CYP and also conjugation reactions play an important role in the metabolism of drugs in the small intestine [20]. For example, drugs such as midazolam or felodipine are subjected to an intense first-pass effect at the intestinal membrane [21, 22]. However, although it would seem that drug metabolism within the mucosal cells would serve only to reduce its extent of availability, in some cases, this process may enhance bioavailability. For example, clindamycin palmitate is a more stable form than the clindamycin HCl solution. The palmitate ester is rapidly hydrolyzed into the parent drug and its use increases the availability of clindamycin after oral administration [23]. Molecules such as clindamycin palmitate are chemical derivatives of drugs usually made to enhance the pharmaceutical properties (e.g., lipophilicity) of the parent molecule. These derivatives are called prodrugs.

Another recently described process that reduces the extent of bioavailability is the excretion process by efflux transporters in the small intestine. In this process, the drug—once it penetrates the intestinal membrane and is in the cytoplasm of the enterocyte—is subjected to a secretion process mediated by a transporter and is translocated, again, into the intestinal lumen. This process is mediated by an ATP-binding transmembrane transporter called P-glycoprotein. P-glycoprotein is a protein that belongs to the ATP-binding cassette (ABC) family. P-glycoprotein functions as a transmembrane efflux pump that translocates its substrates from its intracellular domain to its extracellular domain. P-glycoprotein is assumed to be one of the most important ABC transporters for drug disposition in humans. It is now established that P-glycoprotein is expressed constitutively in small intestine and in other organs. As a result of its anatomical localization, P-glycoprotein limits drug entry into the body after oral drug administration as a result of its expression in the luminal (apical) membrane of enterocytes [24].

Recently, it has been shown that enterocytes simultaneously express the major drug-metabolizing enzymes CYP3A and the efflux transporter P-glycoprotein [19]. This leads to a drug efflux–metabolism alliance, which increases the access of drug to metabolism by CYP3A4 through repeated cycles of absorption and efflux, further reducing the possibility of the drug reaching the systemic circulation unaltered.

### 4.3 Technological Factors Affecting Gastrointestinal Drug Absorption

As well as the physicochemical and physiological factors affecting intestinal drug absorption, there are technological variables that could affect the rate and extent of drug absorption. We have classified these variables into two subcategories: factors depending on pharmaceutical dosage form and manufacturing factors.

***Pharmaceutical Dosage Form*** In considering, in a general manner, the availability of drugs from various classes of dosage forms, drugs administered in solution usually produce the most available drug product, assuming that the drug does not precipitate in the stomach and is not deactivated there. The second most available dosage form would be drug dispersed in a fine suspension, followed by micronized drug in a capsule, uncoated tablets, and finally by the coated tablets. In formulating and designing drug products, this ranking should be kept in mind.

The reason for this different availability is the difference in processes involved in drug release from the dosage form. Once the dosage form reaches the stomach, it must break down (if it is a solid dosage form) and release the therapeutic agent. Disintegration and dissolution are the key processes that precede the absorption process from a solid dosage form. As mentioned earlier, dissolution of the drug in the gastrointestinal fluid is the first step for drug absorption. The disintegration process increases the surface area of the dosage form and, in general, will increase the rate of dissolution. Disintegration is not a prerequisite when a pharmaceutical suspension or solution is employed.

Release of the therapeutic agent can be affected by the nonactive ingredients included in the dosage form. For example, for drugs formulated as tablets, diluents, disintegrants, binders, lubricants, surfactants, and even colorants can affect the release of the drug from the pharmaceutical dosage form as measured by the rate of dissolution in *in vitro* tests. Their proper choice becomes more critical when formulating water-insoluble drugs and when the total concentration of the drug in the dosage form is small.

***Manufacturing Processes*** In addition to the above-mentioned variables, the characteristics and processing of the dosage form could have great influence on drug availability. Variables such as process of granulation method or compressional force in production of tablets can significantly affect the bioavailability of the drug [25].

## 5 INTESTINAL PERMEABILITY ENHANCEMENT: POSSIBILITIES AND LIMITATIONS

In the last decade incredible advances have been made in the application of molecular biology and biotechnology, which have led to a revolution in the development of new therapeutic compounds. These new drugs must be properly delivered in the body to have the desired pharmacological effect. In this challenging task, it is essential to consider the biopharmaceutical and pharmacokinetic properties of the drug, particularly those concerning intestinal absorption and bioavailability (BA) after oral administration. The oral route is the preferred administration route, and it must be taken into account that the newer promising drugs, especially peptides and proteins, cannot be developed as oral products because of their null or scarce bioavailability. In general, these are hydrophilic compounds, of medium to high molecular weight, and sometimes containing strongly charged functional groups—implying that transport across the intestinal barrier occurs essentially via the paracellular pathway [26]. The contribution of this pathway to intestinal absorption is considered to be small, since this pathway occupies less than 0.1% of the total surface area of

the intestinal epithelium [27], and the presence of tight junctions (TJs) between the epithelial cells limits drug absorption. Therefore, for the above-mentioned drugs, the main cause of low bioavailability is their poor intestinal permeability. Thus, considerable attention has been directed at finding ways to increase the paracellular transport of these compounds—the use of enhancers being the most promising and studied strategy [8, 28]. In recent years, knowledge of tight junction physiology and regulation has increased [29], which has facilitated the search for compounds capable of enhancing absorption via the paracellular pathway. To obtain maximum benefit from such enhancing compounds, it is necessary to confirm that reduced membrane permeability is the cause (or at least the main cause) of poor drug bioavailability. Moreover, the activity of an absorption enhancer should be inmediate and should coincide with the presence of the drug at the absorption site [8, 30]. In these cases, an increase in intestinal absorption would be achieved, leading to an enhancement in drug bioavailability, which in turn would allow oral administration of the drug and also a reduction in inter- and intrasubject variability in plasma concentrations and therefore in therapeutic effects.

In this section, we analyze the possibilities and limitations of the most promising intestinal paracellular enhancers. The basic anatomic and physiological properties of the paracellular route are first examined in order to better understand the absorption process, which could help in the selection of the appropriate enhancer. Interest has focused on medium chain fatty acids and chitosan and its derivatives, since these are the most studied and effective enhancers. In fact, the medium chain fatty acid sodium caprate is being used as an absorption enhancer in the clinical setting in Japan, Denmark, and Sweden. There have been no reports of serious side effects [31–33]. Additionally, the results obtained with the more recent paracellular enhancers are presented. Previous reviews are recommended for in-depth insight into this topic [4, 8, 28, 34].

### 5.1 Structure and Function of the Intestinal Paracellular Pathway

The junctional complex found in the apical portion of adjacent cells is composed of three distinct regions: tight junctions or zonula occludens, zonula adherens, and macula adherens or desmosome (Fig. 4). Paracellular permeability is primarily regulated by the TJs, because this is the rate-limiting barrier of the transport pathway. TJs have been described as gates (selectively allowing the passage of small hydrophilic compounds) and as fences (forming an intermembrane diffusion barrier maintaining enterocyte polarity and excluding potentially toxic molecules). These two functions are not separate unrelated phenomena [29]. It is becoming increasingly clear that TJs in themselves constitute the product of a global polarizing process; their role therefore does not seem to correspond to that of a simple fence [35].

In the last years some light has been shed on the architecture and regulation of TJs. Because a complete review of TJ structure, function, and molecular regulation is beyond the scope of this chapter, and has previously been done [35–38], only a brief summary of the more relevant aspects of the subject and its repercussions on paracellular permeability are presented here.

Under the transmission electron microscope, tight junctions appear as a series of focal contacts between the plasma membrane of two adjacent cells. In early freeze-fracture electron microscopy sections, the tight junctional complex appeared as a

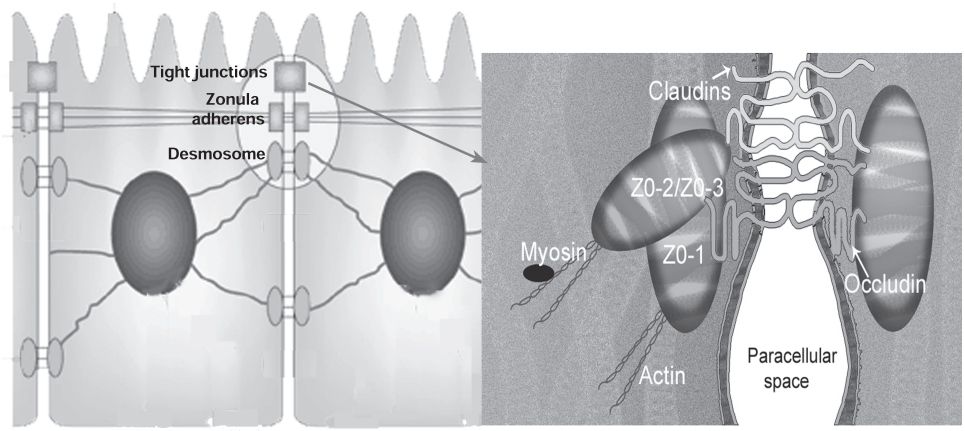


**FIGURE 4** (Left) The functional complex located in the apical part of adjacent enterocytes (tight junctions, zonula adherens, and desmosomes) is depicted. (Right) Schematic representation of the protein interactions at tight junctions. The interaction is only represented in one part of the illustration. Other proteins have been localized to the cytoplasmic surface of tight junctions, although they are not represented. For more information see the text. (From Ref. 34, with permission.)

dense network of interdigitating strands or fibrils in the plane of the plasma membrane [39]. At some points the strands showed discontinuities, which might correspond to "pores" [40], and which would imply channel-like permeability (i.e., fluctuating aqueous pores embedded in the fibrils would account for diffusion through TJs). This early concept has progressed little and it is presently also postulated that fibrils on one cell interact with fibrils on an adjacent cell to seal the paracellular space and define the permeability characteristics [36]. To date, the commonly held view is that increases in paracellular permeability under perturbed or pathological conditions result from the dilatation of existing tight junction pores [41]. In contrast, the composition of the fibrils is now well known. Fibrils are formed by at least two types of tetra-spanning transmembrane proteins: occludin and claudin (Fig. 4). Occludin is a 65 kDa phosphoprotein [42–44]. The findings reported to date are consistent with a functional role for occludin in defining the barrier [45, 46]. Claudins are a family of proteins, named from the Latin *claudere*, "to close." They appear to represent the major structural components of tight junction strands [47]. Recent studies have proposed that claudins are the pore-forming structures in TJs [48, 49]—thus strongly supporting the idea that claudins confer specific selectivity to paracellular transport. Furthermore, a dense cytoplasmic network of proteins has been described at the TJs. These are referred to as tight junction associated proteins (TJAPs), and are designated ZO-1, ZO-2, and ZO-3. As can be seen in Fig. 4, these proteins interact among each other (ZO-1 and ZO-2 are bound to each other and to 130 kDa protein, ZO-3) and also serve as a link between occludin and the actin filaments of the cytoskeleton [28]. The association of TJs with the apical perijunctional actomyosin ring seems to regulate global TJ permeability [38]. It is also known that the barrier assembly and permeability characteristics of TJs are influenced by many cellular signaling mechanisms—although these remain largely

undefined [50]. Therefore, despite rapid progress in knowledge of TJ structure and molecular physiology, their function in the context of paracellular permeability is still far from fully clear.

In summary, evidence now exists to suggest that the regulation of paracellular permeability by the TJ is a complex process, illustrated by the diversity in functional reactions of this structure. At present, apart from conventional permeability studies, several probes have become available for detecting changes in activation of intracellular proteins, such as the protein kinase C isotypes. Moreover, *in situ* imaging techniques are being used [37]. Studies in the field of diseases originated by paracellular permeability alterations (food allergies, malabsorption syndromes, and inflammatory bowel diseases such as Crohn's disease and ulcerative colitis) also contribute to our understanding of paracellular transport. In future, all such intense research can contribute to secure in-depth knowledge of the control mechanisms implicated in paracellular transport and its modulation, which may have great repercussions for drug delivery.

### 5.2 Intestinal Paracellular Enhancers

As already outlined, attempts to reduce the absorption barrier are mainly based on the use of paracellular permeation enhancers as auxiliary agents in oral drug delivery systems [8]. Before presenting the potential compounds that could be used as enhancers, several general aspects are considered.

***General Considerations*** According to Aungst [8] there are several critical issues to consider in the selection of a compound as a potential absorption enhancer for use in drug delivery: the degree of bioavailability enhancement achieved, the influence of formulation and physiological variables, the possible toxicity originated by the enhancing action, and the mechanism of permeation enhancement.

On the other hand, once the enhancer has been selected it should be highlighted that the efficacy of absorption enhancer, in terms of intensity and duration of its effect, depends on the concentration at its site of action. Therefore, it is necessary to have the means to control this concentration or at least to be able to regulate it within a certain effective concentration range. This, nevertheless, might be difficult to achieve in some cases, since many variables, as outlined before, influence the actual concentration of drug and/or enhancer at a specific site in the GIT. Alternatively, the application of controlled release dosage forms could be more effective in maintaining concentrations of enhancer in the effective range [51].

The vast majority of data published on the use of intestinal paracellular enhancers have been obtained using *in vitro* and *in situ* methodologies. The possibilities and limitations of these techniques are evaluated in the following section. The effectiveness of any compound for enhancing intestinal permeation must also be assessed by *in vivo* studies. The degree of bioavailability enhancement, as well as the effect on $C_{max}$ and the area under the curve (AUC), are the indicators most commonly used in studies of this kind. The *in vivo* setting clearly constitutes a more complex and dynamic environment, which hampers assessment of the promoting effect. Furthermore, the formulation used to administer the drug and promoter (solution, capsules, microcapsules, enteric-coated capsules, suppositories) could influence the efficacy of an absorption enhancing excipient. Similarly, the route of

administration used (peroral, intrajejunal, intracolonic, rectal) is directly related to the success of oral drug bioavailability enhancement. For more information on the factors influencing the *in vivo* performance of permeation enhancers, readers are referred to the reviews by Aungst [8] and Ward et al. [28].

***Medium Chain Fatty Acids (MCFAs)*** Based on research conducted in the last decade, it has become clear that several sodium salts of medium chain fatty acids (caprylate C8 ($CH_3-(CH_2)_6-COOH$), caprate C10 ($CH_3-(CH_2)_8-COOH$), and laurate C12 ($CH_3-(CH_2)_{10}-COOH$)) are able to enhance the paracellular permeability of hydrophilic compounds. Lindmark et al. [52] carried out a comparative study with these three fatty acids and sodium caproate (C6). They showed that C8, C10, and C12 (but not C6) exhibit dose-dependent enhancing effects on mannitol transport across cell monolayers—with C12 being the most effective enhancer. Interestingly, the lowest concentrations to enhance transport of the marker molecule were in the vicinity of their critical micellar concentration (CMC), which in turn differs considerably for each.

Among these MCFAs, sodium caprate is the most extensively studied and the only absorption-enhancing agent included in a marketed drug product. It is added in a suppository formulation intended for human use in Sweden and Japan [33]. Since this fatty acid has a low molecular weight, it could be absorbed from the intestine even more quickly than the drug itself [8]. The numerous studies conducted with C10 have addressed different aspects such as its enhancing effect on the permeability of compounds with different molecular weights (MWs), its concentration and time-dependent effects, and its toxicity and mechanism of action.

Most published data on *sodium caprate* as an absorption enhancer have been obtained using *in vitro* and *in situ* techniques. Collectively, the results obtained using these methodologies indicate that C10 can enhance intestinal permeability of low (MW 180–400) and high (MW 4000–19,000) molecular weight *marker molecules* [31, 52–56]. However, the authors point out that the effect would be significant only for substances of MW ≤ 1200 g/mol; that is, in the case of larger molecules, permeability enhancement would not result in a significant increase in the dose fraction absorbed [57]. Regarding the effects on *drug* permeability, sodium caprate enhances intestinal permeability of cefmetazole [58], ebitaride (a pentapeptide ACTH analogue) [57], peptide drugs [53, 59], and epirubicin, an anticancer drug [60]. Nevertheless, it is worth pointing out that Raiman et al. [61] have recently found that C10 (10 mM) does not affect the permeability of the bisphosphonate drug clodronate across Caco-2 cell monolayers.

In all of the studies just mentioned, the sodium caprate concentration used was in the vicinity of its CMC (i.e., 13 mM). Several research groups, using Caco-2 cells, have shown the C10 enhancement effect on membrane permeability to be concentration dependent [31, 52, 53]. We have also recently observed this *concentration-dependent effect in vitro* (Caco-2 cell) using acamprosate, an alcoholic antirelapse drug (unpublished data). However, when we performed the experiments *in situ* [62] and *in vivo* with rats, this effect disappeared. Our results agree with those obtained *in vivo* by Raoof et al. [63] in pigs. According to their data, the enhancing capacity of C10 is dose independent.

Regarding the *time-dependent effect* of C10 on intestinal permeability, studies carried out by Anderberg et al. [54], using Caco-2 cells, showed significant time-

dependent effects at 13 mM and higher concentrations (16 mM). These experiments were performed after long incubations (approximately 1 h) with C10. Since the immediate effect of the enhancer is presumably more relevant than the long-term effect *in vivo*, Lindmark et al. [32], in Caco-2 cell monolayers, determined both the long- and short-term effects. According to their results, the time-dependent effect of sodium caprate 13 mM upon permeability can be separated into two phases: an initial phase (10–20 min) in which a rapid increase in permeability was observed, and a later phase characterized by a slow but more prolonged enhancement in permeability. Kamm et al. [59] have also shown C10 absorption enhancement to be markedly dependent on incubation time in Caco-2 cells.

The local *toxicity* of sodium caprate in the small intestine is one of the main concerns with use of this fatty acid in pharmaceutical products. The toxicity of C10 has been extensively studied *in vitro*. Considering that cytotoxicity depends on the concentration and duration of exposure, comparisons among results obtained in experiments using different protocols may not always be valid. Cell damage can be assessed by several methods (morphological observations, the release of biological markers, and the recovery of transepithelial electrical resistance (TEER))—a fact that further complicates the drawing of firm conclusions. A closer examination of the reported data suggests that C10 at effective concentrations (around 13 mM) does not affect epithelial viability [64] and does not cause serious cytotoxicity—its effects moreover being reversible [53, 55, 65, 66].

The studies performed *in vivo* with sodium caprate are not as numerous as those using *in vitro* and *in situ* techniques. The drugs involved in these investigations have been peptides, antibiotics, and polar, high MW *drugs* such as antisense oligonucleotides or glycyrrhizin. When the drug and enhancer are administered as a solution, the effective C10 doses in the reported studies are comparable. Closer examination shows that all these doses are in the range of 0.1–0.5 mmol/kg. Curiously, when these amounts are transformed into concentrations, the values obtained are in some cases quite different and clearly higher (25–1000 mM) than those tested in the *in vitro* models [34]. Most investigations have been performed in rats, although dogs, pigs, rabbits, and even humans have also been used [34]. As a whole, it can be concluded that this medium chain fatty acid is capable of improving oral drug bioavailability. However, it must be stressed that only two studies have been carried out to date in humans, and no effect was reported in one of them [67].

It is worth mentioning that two studies have compared the *in vivo* enhancing effect of several medium chain fatty acids [68, 69], concluding that the strength of this effect is in the following order: caprate > laurate > caprylate. Hence, sodium caprate seems to be the most potent promoter among the MCFAs tested.

*In vivo* studies analyzing the toxicity of C10 are scarce but in concordance. This fatty acid, at the doses tested, was well tolerated by the intestinal mucosa, and no membrane damage was observed [53, 63, 68].

In conclusion, sodium caprate can be considered a promising agent for use as an enhancing excipient in drug delivery. The C10 effect is dependent on its permeability in the tissue upon which it is required to act. Further studies are needed to confirm its null or low toxicity, and to evaluate its enhancing efficacy for each particular drug substance.

***Chitosan and Its Derivatives*** High MW polymers such as chitosan and its derivatives have gained considerable attention as permeation enhancers. Because of their

high MW, these polymers are supposedly not absorbed from the gut, and systemic side effects are thus excluded. In addition, prolonged localization in the mucosa is ensured, which in turn would prolong the promoting effect. Chitosan and its derivatives are also of great interest as excipients and drug carriers in the pharmaceutical field, due to their biodegradability, biocompatibility, and recent FDA application [34].

*Chitosan* is a partially deacetylated form of chitin, which is present in crustacean shells, insects, and fungi, as well as in some microorganisms (Fig. 5). Chitosan is actually a denomination describing a series of polymers with different molecular weights (from 50 to 2000 kDa) and degrees of deacetylation. These two factors are very important for the physicochemical properties of chitosan and thus exert a major impact on its promoting effects (as discussed later).

The effect of chitosan on intestinal permeability across tight junctions was first reported by Artursson et al. [70]. These authors found chitosan glutamate to enhance $^{14}$C-mannitol transport *in vitro* (Caco-2 cell monolayers). Posteriorly, it was

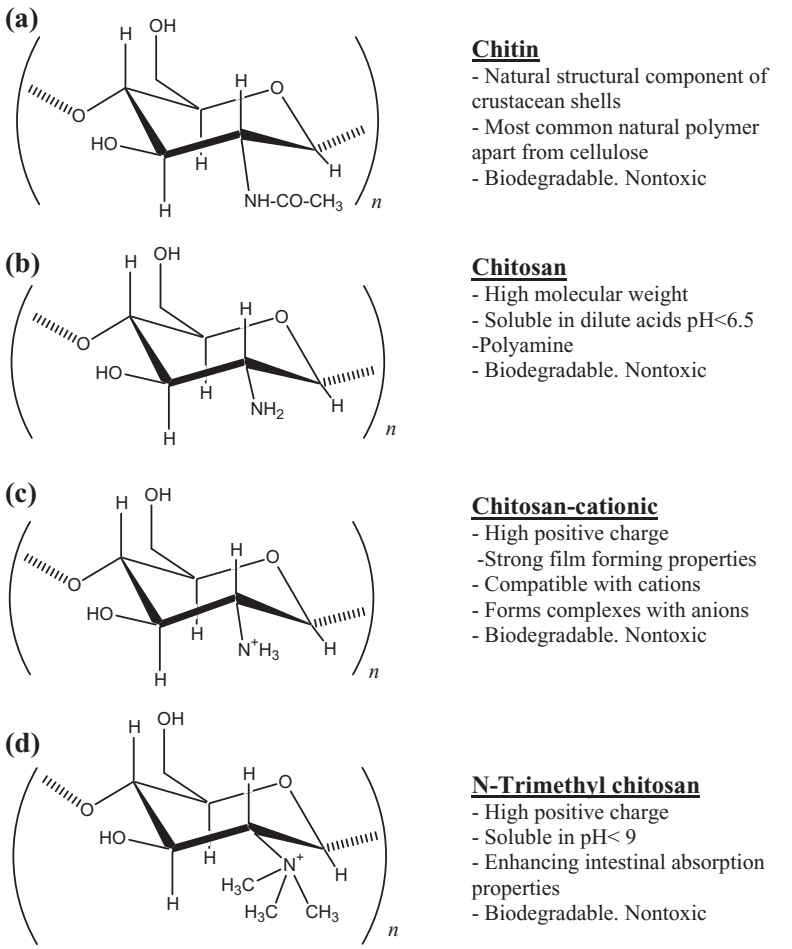


**FIGURE 5** Chemical structures and properties of chitosan and its derivatives. (From Ref. 34, with permission.)

confirmed that chitosan hydrochloride and chitosan glutamate, in a slightly acidic environment, increase the permeation of low molecular weight *markers* as well as of large hydrophilic compounds (PEG-4000 and fluorescein dextrans) in intestinal epithelial cells. It was also found that both salts are insoluble at pH 7.4 and prove to be ineffective as permeation enhancers [71, 72].

Regarding the effect of chitosan salts on *drug* intestinal permeability *in vitro*, there is general agreement that these polymers are potent absorption enhancers for poorly absorbed drugs such as atenolol and peptide drugs. Studies with buserelin, 9-deglycinamide 8-arginine vasopressin (DGAVP), and insulin have evidenced a strong increase in the transport of these drugs in the presence of chitosan glutamate and chitosan hydrochloride (acidic environment) [34].

Schipper et al. [73] reported that the molecular weight and degree of deacetylation of chitosans influence their effects on Caco-2 permeability and cytotoxicity. Accordingly, one chitosan with an intermediate degree of deacetylation and high molecular weight had good enhancing characteristics and low cytotoxicity. Chitosan toxicity has been further evaluated by other researchers based on trypan blue exclusion studies and confocal laser scanning microscopy [71]. No deleterious effects on the cells were demonstrated. Thus, the general opinion regarding chitosan salt toxicity *in vitro* appears to be that it offers a very safe toxicity profile.

*In vivo* studies of the enhancing effects of chitosan on intestinal absorption are scarce but concordant. Chitosan hydrochloride has been observed to increase drug (buserelin, octreotide acetate) bioavalibility in rats and pigs in an acidic environment [34]. Hence, the chitosan pH-dependent effect observed *in vitro* was also confirmed *in vivo*. The explanation for this phenomenon is based on the fact that chitosan, being a weak base, loses its positive charge in neutral and basic media—thus proving to be ineffective as an absorption enhancer [74].

To overcome the solubility limitations of chitosans at neutral and basic pH, *N-trimethyl chitosan chloride* (*TMC*) *derivatives* (Fig. 5d) are obtained. These derivatives display much greater aqueous solubility than chitosan at neutral and alkaline pH values [75]. Consequently, a great volume of data have been published on the intestinal enhancing properties of these polymers.

*In vitro*, TMC derivatives are especially effective in enhancing the transport of small hydrophilic compounds (e.g., mannitol), although they also improve the transport of large molecules (drugs) such as buserelin, insulin, DGAVP, and octreotide acetate [34]. Throughout the multiple studies performed *in vitro*, the researchers have abundantly shown the degree of quaternization to play an important role in the absorption-enhancing properties of these polymers, especially in neutral and basic environments. There is general agreement that the promoting effect of TMC derivatives increases with an increase in their degree of quaternization [34]. Discrepancies exist regarding the optimum degree of quaternization. Thus, maximum absorption enhancement of mannitol and buserelin was recorded with TMC 60 (degree of quaternization 60%) [31, 76]. However, it has recently been reported that the best and maximum permeation-enhancing results are achieved with TMC 49 (degree of quaternization 48.8%) [77, 78]. The few *in vivo* studies reported to date are relatively recent and have been conducted using TMC derivatives with a high degree of quaternization (TMC 40 and TMC 60), and involving the same *drugs* tested *in vitro*—that is, buserelin and octreotide acetate [31, 79]. All these data reveal increased drug bioavailability when the drug is administered intraduodenally in rats

and pigs, with TMC 40 and TMC 60 at neutral pH values. Moreover, a concentration-dependent effect of these polymers was observed [79]. No toxicity studies have been presented to date.

TMC derivatives are only suitable for improving the intestinal absorption of macromolecular therapeutic agents with neutral or basic properties due to its positive charge. Hence, when chitosan salts and TMC were evaluated for their compatibility with low molecular weight heparin (LMWH), a highly anionic polysaccharide, strong aggregation was observed with subsequent fiber formation and precipitation. To overcome this problem, another chitosan derivative has been synthesized: *mono-N-carboxymethyl chitosan* (*MCC*) [80]. This polymer is perfectly soluble in aqueous environments at neutral and alkaline pH values and is compatible with anionic and neutral compounds. MCC was evaluated as a potential absorption enhancer of LMWH *in vitro* (Caco-2 cell monolayers) and *in vivo* (rats) [80]. Both the *in vitro* and *in vivo* results indicated that MCC derivatives are capable of significantly increasing the intestinal absorption of LMWH. Moreover, these polymers were seen to present nontoxic characteristics. The authors concluded that carboxymethyl modifications of chitosan may be suitable absorption enhancers for the peroral delivery of anionic macromolecules.

In summary, it can be concluded that chitosan and its derivatives are promising excipients for use as enhancers in the peroral delivery of poorly absorbed drugs.

***Zonula Occludens Toxin (Zot)*** The great interest in the peroral administration of poorly absorbing therapeutic agents has led to the development of innovative strategies. This is the case, for instance, of *zonula occludens toxin* (Zot), a 45 kDa protein elaborated by *Vibrio cholerae* that is able to reversibly regulate tight junction permeability [81]. This toxin interacts with a specific intestinal epithelial surface receptor, with subsequent activation of a complex cascade of intracellular events that lead to a protein kinase C-dependent polymerization of actin microfilaments strategically localized to regulate the paracellular pathway, and consequently leading to opening of the TJs at a toxin concentration as low as $1.1 \times 10^{-13}$ M [82, 83]. The potential of Zot to enhance the paracellular transport of marker compounds and drugs was first investigated by Fasano et al. [84]. These authors showed that Zot reversibly enhances rabbit intestinal permeability to insulin *in vitro* in the jejunum and ileum, although no substantial changes were detected in the colon. Posteriorly, Cox et al. [85] demonstrated that Zot enhances transport across Caco-2 cell monolayers of low and *high molecular weight markers*. It was also shown that the *in vitro* permeabilities of drugs with low oral bioavailability such as paclitaxel, acyclovir, and cyclosporine and enamione anticonvulsants were increased with Zot. Furthermore, the enhancing properties of Zot were found to be reversible and nontoxic [81–85]. Recent studies have identified a smaller 12 kDa fragment of Zot, referred to as $\Delta G$ [86]—this fragment being responsible for the intrinsic tight junction modulation activity. In 2003, Salama et al. [87] reported that this biologically active fragment ($\Delta G$) is able to increase mannitol permeability across Caco-2 cell monolayers. $\Delta G$ was found to be noncytotoxic at the concentration tested. These authors also examined the *in vivo* effect of $\Delta G$ using the rat as animal model. When mannitol was administered intraduodenally with $\Delta G$ only, no significant differences were observed in terms of the pharmacokinetic parameters. However, when the active fragment was used in the presence of protease inhibitors (PIs), significant increases were

obtained for $C_{max}$ and AUC of mannitol. Thus, protease inhibitors are necessary to minimize $\Delta G$ enzymatic degradation secondary to proteases/peptidases. Furthemore, $\Delta G$ was also able to increase the AUC and $C_{max}$ for macromolecules [82]. Recently, it has been reported that $\Delta G$ significantly increased the *in vivo* oral absorption of some low bioavailable *drugs* (cyclosporin A, ritonavir, saquinavir, and acyclovir) in the presence of protease inhibitors [88]. In the opinion of the authors, these studies illustrate the potential usefulness of $\Delta G$ in enhancing oral drug delivery.

***Thiolated Polymers*** An alternative class of permeation enhancers is represented by *thiolated polymers*—also called thiomers. These are polymers in which the thiol groups are covalently bound. It has been shown that polycarbophyl polymers (PCP) display permeation-enhancing effects [89]. This property, however, could be significantly improved as a result of the covalent attachment of cysteine (Cys) to this polymer (*PCP-Cys*) (Fig 6a), as has recently been shown by Clausen and Bernkop-

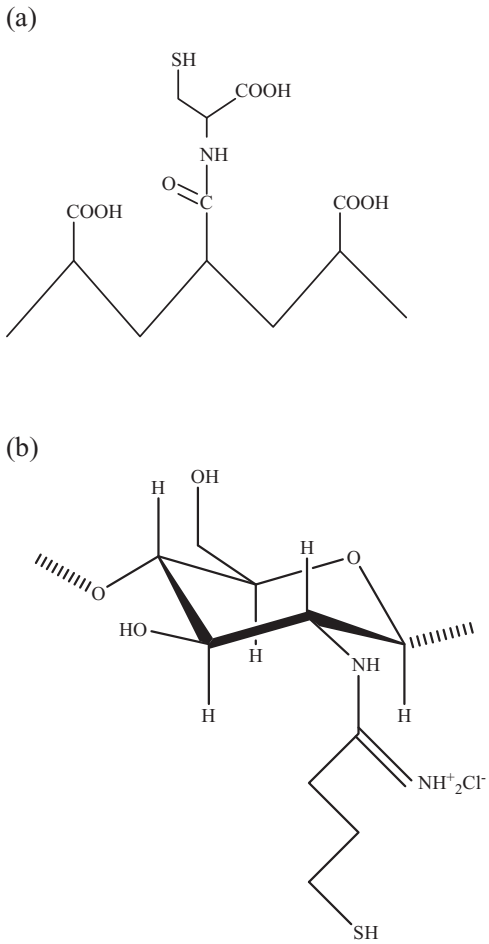


**FIGURE 6** Chemical structure of (a) polycarbophyl-cysteine conjugate and (b) chitosan-4-thio-butylamidine. (From Ref. 34, with permission.)

Schnürch [90]. Accordingly, this thiolated polymer (PCP-Cys) is able to significantly increase the transport of *marker compounds* (sodium fluorescein) and *peptide drugs* (bacitracin-fluorescein isothiocyanate and insulin-fluorescein isothiocyanate) across the intestinal mucosa of guinea pigs (*in vitro* studies). The thiol groups, covalently attached to the polymer, seem to be responsible for the improved permeation-enhancing properties of these conjugates. Furthermore, the improved mucoadhesive properties of these polymers provide a somewhat prolonged residence time, which in some cases can lead to additional improvement in drug bioavailability. They are not absorbed from the GIT due to the high molecular mass. Based on all these features, the authors concluded that the thiolated polymers could constitute a promising excipient for the development of peptide drug delivery systems. Posteriorly, this same research group explored the mechanism of action of thiolated PCP-Cys in depth [91], showing that these compounds exert their permeation-enhancing effects via glutathione. It seems that PCP-Cys can transform oxidized glutathione (GSSG) into reduced glutathione (GSH), prolonging GSH concentration at the apical membrane. GSH is reportedly capable of inhibiting protein tyrosine phosphatase (PTP) activity by almost 100%, which results in a higher extent of phosphorylated tyrosine groups on the extracellular loops of the membrane protein occludin, leading to the opennig of the TJs [92].

Other thiolated polymers such as sodium carboxymethylcellulose-cysteine, chitosan-cysteine and chitosan-4-thio-butylamidine (chitosan-TBA) (Fig. 6b) [34] have also displayed permeation-enhancing properties for hydrophilic compounds and drugs when evaluated using *in vitro* techniques. The addition of the permeation mediator GSH improves this effect. This association is referred to as a *thiomer/GSH system* [93]. Different thiomer/glutathione systems have been used depending on the hydrophilic macromolecular drugs tested. Hence, for salmon calcitonin, a peptide drug of net cationic charge, chitosan-TBA was used. In contrast, insulin and low molecular weight heparin are anionic drugs—thus being incompatible with chitosan. In these cases, an anionic thiolated polymer—poly(acrylic acid)—cysteine conjugate—was used. When *in vivo* experiments were performed, a significantly improved pharmacological efficacy/bioavailability was achieved by using these systems. These results are in good agreement with those obtained *in vitro*. The authors concluded that due to their high efficacy and minimal toxicological risks, the thiomer/GSH systems represent a promising new generation of oral permeation-enhancing delivery systems for hydrophilic macromolecules [93]. Recently, the same research group has shown that the combination of thiomer/GSH system with bromelain, a proteolytic enzyme, represents a promising strategy in order to raise the *in vivo* efficacy of orally administered hydrophilic macromolecular drugs [92].

***Nitric Oxide (NO) Donors*** In 1995, Salzman et al. [94] showed that sodium nitroprusside (SNP), an NO donor, induced a concentration-dependent increase in fluorescein sulfonic acid transport in Caco-2 cell monolayers. Similarly, Utoguchi et al. [95] reported that *S*-nitroso-*N*-acetyl-penicillamine (SNAP), another NO donor, was able to greatly enhance the rectal absorption of insulin—this effect being concentration dependent. Posteriorly, Yamamoto et al. [96] also observed this absorption-enhancing effect on 5(6)-carboxyfluorescein transport with other NO donors such as NO5 (3-(2-hydroxy-1-methylethyl)-2-nitrosohydrazino)-1-propanamine) and NO12 (*N*-ethyl-2-(1-ethyl-hydroxy-2-nitrosohydrazino)-ethanamine), using an *in vitro*

Ussing chamber method and the rat jejunum and colon. Regional differences in the promoting effect of NO12 were found (colon > jejunum). Their findings also demonstrated NO12 action to be mediated by nitric oxide, with partial inclusion of dilatation of tight junctions in the epithelium. As these agents exhibit low toxicity, the authors suggested that NO donors may be useful for enhancing the intestinal absorption of poorly absorbing drugs.

### 5.3 Technologies in Development

Gastrointestinal Permeation Enhancement Technology™ (GIPET™) is a proprietary solid-dose/microemulsion-based medium-chain fatty acid technology by Merrion Pharmaceuticals. In the first format of GIPET (GIPET I), enteric-coated tablets, comprising a pH-sensitive coating and a drug, were synthesized. The second variation of the technology (GIPET II) consisted of microemulsions of mono- and diglyceride mixtures of C8 and C10 entrapped with the drug in an enteric-coated soft gel capsule. GIPET I and II have been tested orally in rats, dogs, and humans, primarily to establish safety profiles but also to demonstrate efficay. Afterwards, these formulations have shown efficacy in human Phase I oral delivery studies of drugs comprising both single- and repeat-dosing regimes. Oral bioavailability of alendronate, desmopressin, and low molecular weight heparin in humans was increased using GIPET formulations compared with unformulated controls. Importantly, Phase I trials revealed no toxicity of concern and this was also observed in subjects receiving multiples doses of GIPET. The absorption-promoting effects of GIPET are transient and complete in less than 1 hour as shown in additional human studies [97]. According to the researchers involved in these studies, GIPET formulations have genuine potential as platform technology for safe and effective oral drug delivery of a wide range of poorly permeable drugs.

### 5.4 Limitations of Intestinal Promotion of Drugs

In 1995, Amidon et al. [98] reported the biopharmaceutical classification system (BCS) for oral delivery of inmediate release products [98], which was adopted by the FDA in 2000. The major outcome of this classification was to group major drug classes according to whether they had oral delivery issues related to solubility or permeability issues, neither of these issues, or both. So drugs that are soluble but poorly absorbed, including most peptides and newer drugs (Class III), are amenable to epithelial permeation enhancement. As outlined earlier, over the past fifteen years many attempts to promote oral absorption of poorly absorbed drugs (Class III) have been carried out but, unfortunately, some of them have failed. It must be highlighted that achieving a sucessful oral formulation with enhancers for a poorly absorbed drug implies that there is access to the appropiate intestinal region for a sufficient amount of time, release of intact soluble drug and enhancer, and an acceptable but reversible degree of epithelial cell permeability enhacement. Bearing this in mind, the limitations for paracellular permeation enhancement by oral formulation include the inability to deliver therapeutic levels over a sustained period, the requirement for massive amounts of material, and safety aspects regarding the long-term integrity of the intestinal epithelium. Additional limitations include the inability to follow through with practical and reproducible solid dose formulations in

scale-up manufacturing. Once the targeted pharmacokinetic and pharmacodynamic profile is achieved in humans, the formulation must have a safety profile to allow it to be given to patients on a repeated basis.

## 6 METHODOLOGIES FOR STUDYING INTESTINAL ABSORPTION

Assessments of rate, extent, site, and mechanism of intestinal absorption have been performed by a variety of experimental techniques in humans and animals with *in vitro*, *in situ*, and *in vivo* preparations. Details of the most common methodologies currently in use in experimental animals are given next, along with a discussion of the utility and limitations of each. It is the judicious use of these techniques that can help identify drug candidates that will be absorbed in humans. It is well recognized that human intestinal permeability cannot be accurately predicted based on a single methodology (*in vitro*, *in situ*, or *in vivo*) [99, 100] since each kind of method covers only some of the factors involved in the intestinal absorption and does not take into account other factors that may be important. Moreover, it must be highlighted that although animals provide the opportunity to gain much valuable information through screening, investigators must remain aware that the results cannot be extrapolated directly and exactly to humans.

### 6.1 *In Vitro* Techniques

Different *in vitro* methods utilizing tissues from rodents and other species have been developed to screen compounds for oral absorption and elucidate the mechanism involved in the intestinal transport, which could be decisive in obtaining maximum bioavailability. Cell cultures have also been shown to constitute a highly valuable tool in the decision making process to select candidates for *in vivo* clinical studies at early stage drug discovery and development [101].

Two *in vitro* biological methods are discussed. These are diffusion chambers equipped with intestinal tissues from animal origin and cultured cells.

***Diffusion Cells Using Tissues*** At the end of the 1960s Schultz et al. [102] developed a method for the direct measurement of unidirectional influx of amino acids from the mucosal bathing solution into the intestinal epithelium. This method was a modification of the so-called Ussing chamber widely used to evaluate drug absorption. Posteriorly, the Shultz method was slightly modified by B.G. Munck, who performed a lot of experiments in order to study the intestinal absorption of amino acids from a mechanistic point of view [103, 104]. Other researchers adopted this technique to gain insight into the absorption mechanism of some drugs (e.g., acamprosate) [105] and trace elements such as Zn [106].

In this method, a chosen segment of the intestine of the animal (rat, rabbit, or guinea pig) is excised, opened along the mesenteric attachment, and rinsed in ice-cold buffer. Each segment is mounted on a Lucite plate with the mucosal surface facing upward and a Lucite block is clamped on top of the plate. The Lucite plate is depicted in Fig. 7. In this way, four mucosal areas, each of approximately 0.35 $cm^2$, can be exposed at the bottom of the wells, where the solution is oxygenated and stirred with high rates of $O_2$ flow. The use of two blocks allows eight measurements

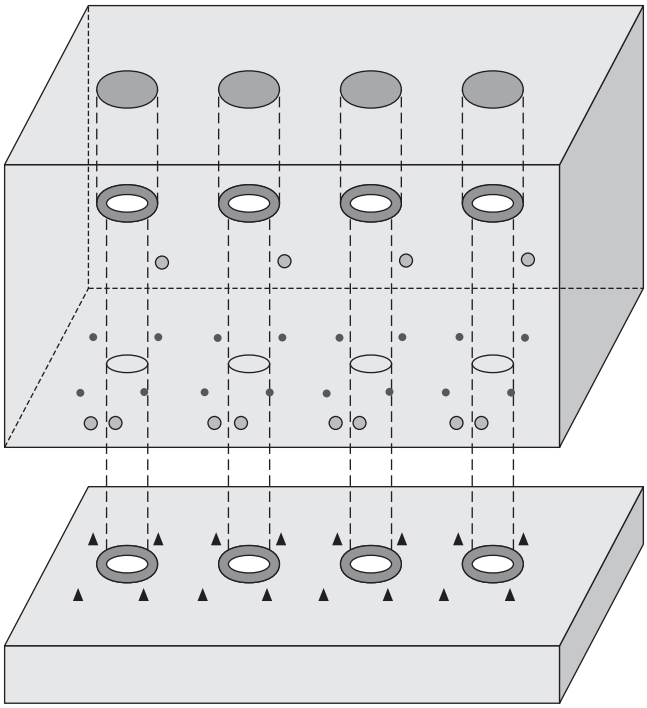

**FIGURE 7** Schematic representation of a Lucite plate, where the segment of the intestine is mounted, and the Lucite block that is clamped on top of the plate.

from each rat. With guinea pigs and rabbits, 4 blocks can be used, allowing 16 measurements from each animal. The serosal surface of the tissue rests on moistened filter paper and is not exposed to the mucosal solution. After mounting, the tissues are preincubated at 37 °C, in a properly designed chamber (Fig. 8). The tissues are incubated for 15 min with a drug-free solution containing 5 mM glucose, which was then withdrawn. The well and mucosal surface are gently wiped with soft paper to remove the adhered incubation fluid before the test solution (containing the radiolabeled drug to test and radiolabeled PEG-4000) is injected. The 0.5 min incubation period is terminated by aspiration of the incubation fluid and flushing of the well with an ice-cold 300 mM mannitol solution. The mannitol wash serves to terminate the exposure both by diluting the remaining test solution and by suddenly cooling the tissues. The exposed tissues are then punched out, briefly rinsed in ice-cold mannitol solution, blotted on hard filter paper, and extracted for 18 h in 0.1 M $HNO_3$. Aliquots of tissue extract and test solution are assayed for drug content. Usually radiolabeled drugs are used since, as outlined before, the exposition time is very short so the amount of transported compound is really low. So the extract and tissue solution are analyzed by liquid scintillation spectrometry. The amount of radiolabeled PEG-4000 was used to correct for extracellular contamination; thus corrected, the radiolabeled drug activity is used to calculate the rate of influx across the brush border membrane. The unidirectional drug influx, $J$, is expressed in concentration/$cm^2 \cdot h$.

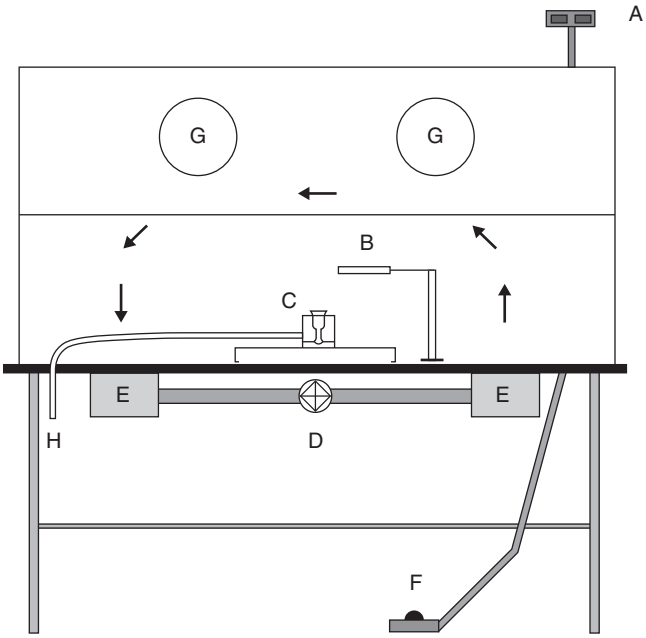


**FIGURE 8** Schematic representation of the chamber used to incubate and oxygenate the tissue mounted in the Lucite plate: A, digital time and temperature controller; B, thermic sound; C, Lucite plate; D, air turbine; E, air inlet and outlet; F, chronometer pulsator; G, access to chamber inside; H, oxygen inlet.

This technique is simple and rapid, and a reduced number of animals, are used in comparison with the *in situ* and *in vivo* methods. Conditions of temperature, oxygenation, and availability of nutrients and energy sources can be controlled more closely for mechanistic studies than is possible *in situ* or *in vivo*. Moreover, researchers can study differences in regional absorption of drugs, which is not possible in cell cultures. The limitations include the problems associated with the use of radiolabeled compounds and the lack of mesenteric circulation, which is not a physiological condition. Another drawback of any *in vitro* system is the intrinsic variability that can be seen in the permeation data.

***Intestinal Cell Cultures*** Within the past decades the use of monolayers of human intestinal epithelial cells for evaluating drug absorption has become possible. Actually, it was impossible to culture normal, mature enterocytes so colon adenocarcinoma cell lines (Caco-2 cells, HT29, T84) are used. Although isolated from a colon tumor, these cells all show an enterocyte-like differentiation; however, the degree of differentiation is variable (Table 3). These cells are relatively well defined while there is also no interference by feedback and/or other mechanism, which are present in intact organs or animals.

The most used cell line from human origin is the Caco-2 cell line. Even though these cells display colon-like properties regarding transepithelial electrical resistance (TEER) and passive permeability to drugs, their morphology, expression of brush border enzymes, and carriers for nutrients resemble more closely small

**TABLE 3 Degree of Differentation of Epithelial Cell Lines Originating from Human Colon Carcinoma**

| | Differentiation | | |
|---|---|---|---|
| Cell Line | Tight Junctions | Apical Brush Border (Microvilli + Enzyme) | Mucus Secretion |
| Caco-2 cells | + | + | 0 |
| HT29 | + | +/– | +/– |
| T84 | + | 0 | 0 |

*Source*: Adapted from Ref. 120.

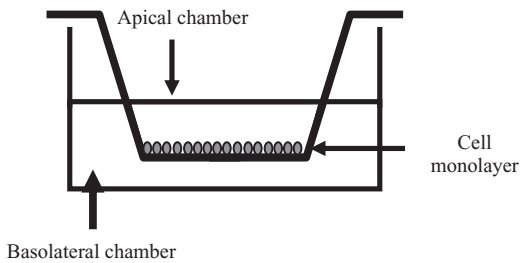


**FIGURE 9** Schematic setup for transport studies in culture cell lines. Sampling can be carried out from the apical and/or basolateral chamber.

intestinal enterocytes [107]. Several studies have demonstrated the usefulness of these cells as a model for intestinal drug transport. According to published results, the properties of this cell line are the same as normal intestinal epithelium in *in vivo* model systems and are useful for studying transport of hydrophilic as well as lipophilic compounds. A vast amount of work has been done in characterizing the Caco-2 cell line. Several carrier and receptor systems have been identified. Moreover, in some studies, a correlation between human intestinal drug absorption after oral administration and permeability in Caco-2 cells was established [108].

Figure 9 illustrates the culturing system whereby the Caco-2 cells are seeded, usually onto polycarbonate filter support on which a polarized cell monolayer is formed within a few days and differentiates into absorptive intestinal cells. The microporous support allows access to drug to both the apical and basolateral sides of the epithelium, which means that drug transport could be studied in both the apical to basolateral and the basolateral to apical. The result of transport experiments can be expressed as the permeability, $P_{app}$, in cm/s.

The tight junctions in the Caco-2 cell monolayers resemble the junctions in the human colon, resulting in far higher TEER than typical for small intestine tissue. The higher TEER makes these cells a useful and widespread method for studying both pharmacological pretreatments and direct additives on cell integrity because the paracellular permeability of a typical nonpermeant such as PEG-4000 or mannitol can easily be monitored. The direct measurement of TEER is also used as a method to evaluate cell monolayer integrity. To measure TEER values, a four-electrode system, which measures the resistance of a cell monolayer on permeable support, could be used. The results of TEER measurements can be expressed as

absolute values or as percentages of the initial values, which is preferable when TEER values are measured after various treatments since variability in the starting values between monolayers exists and different measurement setups might result in different absolute values.

One of the major advantages of the use of intestinal cell cultures lies in its possible application to high throughput screening (HTS) strategies. This model has a high potential for HTS because the cells are rather simple to culture in high quantities and in a reproducible manner. Keep in mind that to time and cost-effectively develop a new drug application, the selection of candidates for clinical studies has to be made during an early-decision process made with *in vitro* data. So *in vitro* permeability and solubility screening has become a routine measurement in drug discovery and development. In some cases, the fraction of drug absorbed in humans could be predicted by *in vitro* Caco-2 cell permeability. In addition, *in vitro* permeability and solubility data can be used to classify compounds according to the BCS system and, subsequently, to direct formulation optimization strategies [101]. Another important advantage is that, as analyzed in the previous section, Caco-2 cells are also a valuable tool to investigate absorption-enhancement of poorly absorbable drugs. Furthermore, the mechanism of action of absorption-enhancing compounds may be elucidated and if combined with microscopical techniques, the paracellular enhancement could be visualized [108].

Apart from these advantages, Caco-2 cells also have striking disadvantages. First, these are cells with tighter tight junctions than those present in the small intestine. This might give rise to underpredictions of the permeabilities of hydrophilic compounds with a low *in vivo* uptake. Second, transporter substrates usually demonstrate lower permeability than their actual permeability *in vivo*, due to the low transporter expression levels in Caco-2 cells [101]. Third, between laboratories, or even batch-to-batch, variability can occur. These limitations should be taken into consideration when assessing the "developability" of a drug and during the decision making process for selecting candidates, before processing a drug to clinical studies.

### 6.2 *In Situ* Techniques

The isolation of segments of the gastrointestinal tract *in situ* in anesthetized animals with intact mesenteric blood flow is widely used to gain valuable information as to the site and mechanism of drug absorption and permeability assessment [2]. Different *in situ* preparations (single-pass perfusion, recirculating perfusion, oscillating perfusion, and static perfusion) have been used. One of the most employed techniques is the *in situ* rat gut preparation described by Doluisio et al. [109]. In short, after anesthetizing the rats, the abdomen is opened through a middle incision and a selected segment of intestine is cannulated at either end. After rinsing with physiological saline to eliminate fecal residue and debris, an isotonic drug solution properly buffered is perfused at 37 °C. At selected times (usually every 5 min for a total of 30 min), the solution is forced out of alternate ends of the segment and sampled. Drug concentration in the samples is quantified and used for calculations. The correspoding kinetic parameters (first-order absorption rate constant, $k_a$, or $V_{max}$ and $K_m$) can be estimated. In some cases, it is necessary to correct for water reabsorption in order to obtain the actual concentration. A nonabsorbable water marker, such as

phenol red, or radiolabeled PEG-4000 could be used. Since our previous experience with these markers was not entirely satisfactory, an alternative method based on the direct measurement, at fixed times, of the remaining volumes of the test solutions perfused independently in selected animals was developed [110]. This technique has proved to be very useful in studying the correlation between gastrointestinal absorption and partition constants for homologous series of xenobiotics [110, 111], the influence of synthetic and natural bile acid surfactants on passive diffusion of xenobiotics [112, 113], and the mechanism of drug absorption. For example, intestinal transport of baclofen, cefadroxil, and cefuroxime axetil through carrier-mediated transport has been evidenced [114–116]. More recently, we have reported, by using this technique, that acamprosate, an anti-craving drug, is absorbed in the small intestine of the rat by passive diffusion [100]. All these investigations have helped to solve problems related with the low bioavailability of the drugs studied. The influence of the potential inhibitors and/or enhancers as well as the participation of the P-gp in the absorption of drugs could also be evaluated [100, 117, 118].

With the use of any of the *in situ* techniques, intact circulation is maintained in such a way that absorbed drugs are taken away from the basement membrane in a more normal physiological manner than in the *in vitro* techniques discussed above. Particularly, with the *in situ* rat gut preparation, absorption rates that are near to *in vivo* values are obtained [12], which allow one to fit absorption models more easily to experimental data. Recently, the *in situ* single-pass perfusion model has proved to be a robust means of assessing permeability for BCS classification [119]. The drawbacks inherent in these techniques are the effects of anesthesia and the short time that the preparations are viable.

### 6.3 *In Vivo* Techniques

The *in vivo* studies in laboratory animals are needed to validate the results obtained *in vitro* or *in situ* in order to avoid the risk of inaccurate predictions. Furthermore, it should be borne in mind that *in vivo* studies are also required during formulation development.

Usually, animals were placed under anesthesia and subjected to surgical cannulation of a vein (e.g., jugular vein) to facilitate blood sampling. Oral administration of the drug is performed usually by gastric intubation. If it is requiered, the animals can be housed in metabolic cages in order to collect urine and feces. Once the sampling period is over, the samples are properly processed to obtain plasma and urine concentrations, which are used for kinetic calculations ($C_{max}$, AUC, bioavailability). Whole animal studies are also useful for investigating specific parts of the absorption process. For example, chronic portal vein cannulae may be inserted surgically and maintained patent for sampling portal venous blood after oral administration. A portal-vein cannula also allows drug infusions to simulate presentation of drug to the liver by controlled-release devices.

One of the main disadvantages of *in vivo* animal studies is that they are impractical for the screening of a large number of compounds due to the intensive work, slow speed, high cost, and lack of automation. Additionally, *in vivo* oral bioavailability studies in animals are not always predictive of bioavailability in humans. Nevertheless, in spite of these limitations, *in vivo* studies provide some initial estimates of how well a compound is absorbed, the variables that affect absorption

(including enhancers), and the probability that the compound undergoes first-pass metabolism.

## REFERENCES


1. Singer SJ, Nicolson GL. The fluid mosaic model of the structure of cell membranes. *Science*. 1972;175:720–731.
2. Loper AE, Gardner CR. Gastrointestinal absorption of drugs. In Swarbrick J, Boylan JC, Eds. *Encyclopedia of Pharmaceutical Technology*, Volume 6. New York: Marcel Dekker; 1988, pp 385–413.
3. Tsuji A, Tamai I. Carrier-mediated intestinal transport of drugs. *Pharm Res* 1996;13(7): 963–977.
4. Salama NN, Eddington ND, Fasano A. Tight junction modulation and its relationship to drug delivery. *Adv Drug Deliv Rev* 2006;58:15–28.
5. He Y-L, Murby S, Warhurst G, Gifford L, Walker D, Ayrton J, Eastmond R, Rowland M. Species differences in size discrimination in the paracellular pathway reflected by oral bioavailability of poly(ethylene glycol) and D-peptides. *J Pharm Sci* 1998;87:626–633.
6. Lindeman P, Solomon AK. Permeability of luminal surface of intestinal mucosal cells. *J Gen Physiol* 1962;45:801–810.
7. Cerejido M. *Tight Junctions*. Boca Raton, FL: CRC Press; 1992.
8. Aungst BJ. Intestinal permeation enhancers. *J Pharm Sci* 2000;89(4):429–442.
9. Finholt P. Influence of formulation on dissolution rate. *Acad Pharm Sci* 1974;106–146.
10. Aguiar AJ, Krc J Jr, Kinkel AW, Samyn JC. Effect of polymorphism on the absorption of chloramphenicol from chloramphenicol palmitate. *J Pharm Sci* 1967;56:847–853.
11. Ulm EH, Hichens M, Gomez HJ, Till AE, Hand E, Vassil TC, Biollaz J, Brunner HR, Schelling JL. Enalapril maleate and a lysine analogue (MK-521): disposition in man. *Br J Clin Pharm* 1982;14:357–362.
12. Pla-Delfina JM, Moreno J. Intestinal absorption–partition relationships: a tentative functional nonlinear model. *J Pharmacokinet Biopharm* 1981;9:191–215.
13. Hunt JN. Gastric emptying and secretion in man. *Physiol Rev* 1959;39:491–533.
14. Clements JA, Heading RC, Nimmo WS, Prescott LF. Kinetics of acetaminophen absorption and gastric emptying in man. *Clin Pharmacol Ther* 1978;24:420–431.
15. Hunt JN, Macdonald I. The influence of volume on gastric emptying. *J Physiol* 1954;126:459–474.
16. Nimmo J, Heading RC, Tothill P, Prescott LF. Pharmacological modification of gastric emptying: effects of propantheline and metoclopromide on paracetamol absorption. *Br Med J* 1973;1(5853):587–589.
17. Welling PG, Huang H, Hewitt PF, Lyons LL. Bioavailability of erythromycin stearate: influence of food and fluid volume. *J Pharma Sci* 1978;67:764–766.
18. Manninen V, Apajalahti A, Melin J, Karesoja M. Altered absorption of digoxin in patients given propantheline and metoclopramide. *Lancet* 1973;24(7800):398–400.
19. Suzuki H, Sugiyama Y. Role of metabolic enzymes and efflux transporters in the absorption of drugs from the small intestine. *Eur J Pharm Sci* 2000;12:3–12.
20. Lin JH, Chiba M, Baillie TA. Is the role of the small intestine in first-pass metabolism overemphasized? *Pharm Rev* 1999;51:135–158.
21. Regardh CG, Edgar B, Olsson R, Kendall M, Collste P, Shansky C. Pharmacokinetics of felodipine in patients with liver disease. *Eur J Clin Pharmacol* 1989;36:473–479.

22. Paine MF, Shen DD, Kunze KL, Perkins JD, Marsh CL, McVicar JP, Barr DM, Gillies BS, Thummel KE. First-pass metabolism of midazolam by the human intestine. *Clin Pharmacol Ther* 1996;60:14–24.
23. Barr WH, Riegelman S. Intestinal drug absorption and metabolism. I. Comparison of methods and models to study physiological factors of *in vitro* and *in vivo* intestinal absorption. *J Pharm Sci* 1970;59:154–163.
24. Fromm MF. Importance of P-glycoprotein at blood–tissue barriers. *Trends Pharmacol Sci* 2004;25:423–429.
25. McGinity JW, Stavchansky SA, Martin A. *Bioavailability in Tablet Technology*. New York: Marcel Dekker; 1981;269–449.
26. Hayashi M, Sakai T, Hasegawa Y, Nishikawahara T, Tomioka H, Iida A, Shimizu N, Tomita M, Awazu S. Physiological mechanism for enhancement of paracellular drug transport. *J Control Release* 1999;62:141–148.
27. Pappenheimer JR. Physiological regulation of transepithelial impedance in the intestinal mucosa of rats and hamsters. *J Membr Biol* 1987;100:137–148.
28. Ward PD, Tippin TK, Thakker DR. Enhancing paracellular permeability by modulating epithelial tight junctions. *Pharm Sci Technol* 2000;3(10):346–358.
29. Madara JL. Regulation of the movement of solutes across tight junctions. *Annu Rev Physiol* 1998;60:143–159.
30. Thanou MM, Florea BI, Langemeÿer MW, Verhoef JC, Junginger HE. N-trimethylated chitosan chloride (TMC) improves the intestinal permeation of the peptide drug buserelin *in vitro* (Caco-2 cells) and *in vivo* (rats). *Pharm Res* 2000;17:27–31.
31. Sakai M, Imai T, Ohtake H, Azuma H, Otagiri M. Effects of absorption enhancers on the transport of model compounds in Caco-2 cell monolayers: assessment by confocal laser scanning microscopy. *J Pharm Sci* 1997;86(7):779–785.
32. Lindmark T, Kimura Y, Artursson P. Absorption enhancement through intracellular regulation of tight junction permeability by medium chain fatty acids in Caco-2 cells. *J Pharmacol Exp Ther* 1998;284:362–369.
33. Takahashi K, Murakami T, Yumoto R, Hattori T, Higashi Y, Yata N. Decanoic acid induced enhancement of rectal absorption of hydrophilic compounds in rats. *Pharm Res* 1994;11(10):1401–1404.
34. Cano-Cebrián MJ, Zornoza T, Granero L, Polache A. Intestinal absorption enhancement via the paracellular route by fatty acids, chitosans and others: a target of drug delivery. *Curr Drug Deliv* 2005;2:9–22.
35. Cereijido L, Shoshani L, Contreras RG. Molecular physiology and pathophysiology of tight junctions. I. Biogenesis of tight junctions and epithelial polarity. *Am J Physiol Gastrointestinal Liver Physiol* 2000;279:G477–G482.
36. Mitic LL, Van Itallie CM, Anderson JM. Molecular physiology and pathophysiology of tight junctions I. Tight junction structure and function: lessons from mutant animals and proteins. *Am J Physiol Gastrointestinal Liver Physiol* 2000;279:G250–G254.
37. Karczewski J, Groot J. Molecular physiology and pathophysiology of tight junctions III. Tight junction regulation by intracellular messengers: differences in response within and between epithelia. *Am J Physiol Gastrointestinal Liver Physiol* 2000;279:G660–G665.
38. Nusrat A, Turner JR, Madara JL. Molecular physiology and pathophysiology of tight junctions. IV. Regulation of tight junctions by extracellular stimuli: nutrients, cytokines, and immune cells. *Am J Physiol Gastrointestinal Liver Physiol* 2000;279:G851–G857.
39. Gumbiner B. Structure, biochemistry, and assembly of epithelial tight junctions. *Am J Physiol* 1987;253:C749–C758.

40. Claude P. Morphological factors influencing transepithelial permeability: a model for the resistance of the zonula occludens. *J Membr Biol* 1978;39:219–332.
41. Watson CJ, Rowland M, Warhurst G. Functional modeling of tight junctions in intestinal cell monolayers using polyethylene glycol oligomers. *Am J Physiol Cell Physiol* 2001;281:C388–C397.
42. Furuse M, Hirase T, Itoh M, Nagafuchi A, Yonemura S, Tsukita S. Occludin: a novel integral membrane protein localizing at tight junctions. *J Cell Biol* 1993;123:1777–1788.
43. Ando-Akatsuka Y, Saitou M, Hirase T, Kishi M, Sakakibara A, Itoh M, Yonemura S, Furuse M, Tsukita S. Interspecies diversity of the occludin sequence: cDNA cloning of human, mouse, dog, and rat-kangaroo homologues. *J Cell Biol* 1996;133:43–47.
44. Saitou M, Ando-Akatsuka Y, Itoh M, Furuse M, Inazawa J, Fujimoto K, Tsukita S. Mammalian occludin in epithelial cells: its expression and subcellular distribution. *Eur J Cell Biol* 1997;73:222–231.
45. Tsukita S, Furuse M. Occludin and claudins in tight-junction strands: leading or supporting players? *Trends Cell Biol* 1999;9:268–273.
46. Van Itallie CM, Anderson JM. Occludin confers adhesiveness when expressed in fibroblasts. *J Cell Sci* 1997;110:1113–1121.
47. Furuse M, Sasaki H, Fujimoto K, Tsukita S. Single gene product, claudin-1 or -2, reconstitutes tight junction strands and recruits occludin in fibroblasts. *J Cell Biol* 1998;143; 391–401.
48. Simon DB, Lu Y, Choate KA, Velazquez H, Al-Sabban E, Praga M, Casari G, Bettinelli A, Colussi G, Rodríguez SJ, McCredie D, Milford D, Sanjad S, Lifton RP. Paracellin-1, a renal tight junction protein required for paracellular $Mg^{2+}$ resorption. *Science* 1999;285; 103–106.
49. Tsukita S, Furuse M. Pores in the wall: claudins constitute tight junction strands containing aqueous pores. *J Cell Biol* 2000;149:13–16.
50. Anderson JM, Van Itallie CM. Tight junctions and the molecular basis for regulation of paracellular permeability. *Am J Physiol* 1995;269:G467–G475.
51. De Boer A, Breimer D. GI-tract absorption enhancement: general introduction and simulation of time–concentration-effect profiles of absorption enhancers. In de Boer AG, Ed. *Drug Absorption Enhancement*. Chur, Switzerland: Harwood Academic Publishers; 1994, Vol 3, pp 155–176.
52. Lindmark T, Nikkila T, Artursson P. Mechanisms of absorption enhancement by medium chain fatty acids in intestinal epithelial Caco-2 cell monolayers. *J Pharmacol Exp Ther* 1995;275(2):958–964.
53. Chao AC, Nguyen JV, Broughall M, Griffin A, Fix JA, Daddona PE. *In vitro* and *in vivo* evaluation of effects of sodium caprate on enteral peptide absorption and on mucosal morphology. *Int J Pharm* 1999;191(1):15–24.
54. Anderberg EK, Lindmark T, Artursson P. Sodium caprate elicits dilatations in human intestinal tight junctions and enhances drug absorption by the paracellular route. *Pharm Res* 1993;10(6):857–864.
55. Quan YS, Hattori K, Lundborg E, Fujita T, Murakami M, Muranishi S, Yamamoto A. Effectiveness and toxicity screening of various absorption enhancers using Caco-2 cell monolayers. *Biol Pharm Bull* 1998;21(6):615–620.
56. Lindmark T, Schipper N, Lazarova L, de Boer AG, Artursson P. Absorption enhancement in intestinal epithelial Caco-2 monolayers by sodium caprate: assessment of molecular weight dependence and demonstration of transport routes. *J Drug Target* 1998;5(3):215–223.

57. Artursson P, Karlsson J. Correlation between oral drug absorption in humans and apparent drug permeability coefficients in human intestinal epithelial (Caco-2) cells. *Biochem Biophys Res Commun* 1991;175:880–885.
58. Tomita M, Sawada T, Ogawa T, Ouchi H, Hayashi M, Awazu S. Differences in the enhancing effects of sodium caprate on colonic and jejunal drug absorption. *Pharm Res* 1992;9(5):648–653.
59. Kamm W, Jonczyk A, Jung T, Luckenbach G, Raddatz P, Kissel T. Evaluation of absorption enhancement for a potent cyclopeptidic alpha(nu)beta(3)-antagonist in a human intestinal cell line (Caco-2). *Eur J Pharm Sci* 2000;10:205–214.
60. Lo YL, Huang JD. Effects of sodium deoxycholate and sodium caprate on the transport of epirubicin in human intestinal epithelial Caco-2 cell layers and everted gut sacs of rats. *Biochem Pharmacol* 2000;59(6):665–672.
61. Raiman J, Törmälehto S, Yritys K, Junginger H, Mönkkönen J. Effects of various absorption enhancers on transport of clodronate through Caco-2 cells. *Int J Pharm* 2003;261: 129–136.
62. Zornoza T, Cano MJ, Polache A, Granero L. Pharmacology of acamprosate: an overview. *CNS Drug Rev* 2003;9(4):359–374.
63. Raoof AA, Ramtoola Z, McKenna B, Yu RZ, Hardee G, Geary RS. Effect of sodium caprate on the intestinal absorption of two modified antisense oligonucleotides in pigs. *Eur J Pharm Sci* 2002;17:131–138.
64. Söderholm JD, Oman H, Blomquist L, Veen J, Lindmark T, Olaison G. Reversible increase in tight junction permeability to macromolecules in rat ileal mucosa *in vitro* by sodium caprate, a constituent of milk fat. *Dig Dis Sci* 1998;43(7):1547–1552.
65. Uchiyama T, Sugiyama T, Quan YS, Kotani A, Okada N, Fujita T, Muranishi S, Yamamoto A. Enhanced permeability of insulin across the rat intestinal membrane by various absorption enhancers: their intestinal mucosal toxicity and absorption-enhancing mechanism of *N*-lauryl-beta-D-maltopyranoside. *J Pharm Pharmacol* 1999;51(11):1241–1250.
66. Yamamoto A, Okagawa T, Kotani A, Uchiyama T, Shimura T, Tabata S, Kondo S, Muranishi S. Effects of different absortion enhancers on the permeation of ebirbatide, an ACTH analogue, across intestinal membranes. *J Pharm Pharmacol* 1997;49(11): 1057–1061.
67. Lennernäs H, Gjellan R, Hällgren R, Graffner J. The influence of caprate on rectal absorption of phenoxymethylpenicillin: experience from an *in-vivo* perfusion in humans. *J Pharm Pharmacol* 2002;54:499–508.
68. Ishizawa T, Hayashi M, Awazu S. Enhancement of jejunal and colonic absorption of fosfomycin by promoters in the rat. *J Pharm Pharmacol* 1987;39:892–895.
69. Sasaki K, Yonebayashi S, Yoshida M, Shimizu K, Aotsuka T, Takayama K. Improvement in the bioavailability of poorly absorbed glycyrrhizin via various non-vascular administration routes in rats. *Int J Pharm* 2003;265:95–102.
70. Artursson P, Lindmark T, Davis SS, Illum L. Effect of chitosan on the permeability of monolayers of intestinal epithelial cells (Caco-2). *Pharm Res* 1994;11:1358–1361.
71. Kotzé AF, Lueßen HL, de Leeuw BJ, de Boer AG, Verhoef JC, Junginger HE. Comparison of the effect of different chitosan salts and *N*-trimethyl chitosan chloride on the permeability of intestinal epithelial cells (Caco-2). *J Control Release* 1998;51:35–46.
72. Kotzé AF, Lueßen HL, de Boer AG, Verhoef JC, Junginger HE. Chitosan for enhanced intestinal permeability: prospects for derivatives soluble in neutral and basic environments. *J Pharm Sci* 1998;7:145–151.
73. Schipper NGM, Vårum KM, Artursson P. Chitosans as absorption enhancers for poorly absorbable drugs. 1: Influence of molecular weight and degree of acetylation on drug

transport across human intestinal epithelial (Caco-2) cells. *Pharm Res* 1996;13(11):1686–1692.

74. Van der Lubben IM, Verhoef C, Borchard G, Junginger HE. Chitosan and its derivatives in mucosal drug and vaccine delivery. *Eur J Pharm Sci* 2001;14:201–207.
75. Kotzé AF, Lueßen HL, de Leeuw BJ, Verhoef JE, Brussee J, Junginger HE. *N*-trimethyl chitosan chloride as a potential absorption enhancer across mucosal surfaces: *in vitro* evaluation in intestinal epithelial cells (Caco-2). *Pharm Res* 1997;14:1197–1202.
76. Thanou MM, Kotzé AF, Scharringhausen T, Lueßen HL, de Boer AG, Verhoef JC, Junginger HE. Effect of degree of quaternization of *N*-trimethyl chitosan chloride for enhanced transport of hydrophilic compounds across intestinal caco-2 cell monolayers. *J Control Release* 2000;64:15–25.
77. Jonker C, Hamman JH, Kotzé AF. Intestinal paracellular permeation enhancement with quaternised chitosan: *in situ* and *in vitro* evaluation. *Int J Pharm* 2002;238:205–213.
78. Hamman JH, Schultz CM, Kotzé AF. *N*-trimethyl chitosan chloride: optimum degree of quaternization for drug absorption enhancement across epithelial cells. *Drug Dev Ind Pharm* 2003;29:161–172.
79. Thanou MM, Verhoef JC, Marbach P, Junginger HE. Intestinal absorption of octreotide: *N*-trimethyl chitosan chloride (TMC) ameliorates the permeability and absorption properties of the somatostatin analogue *in vitro* and *in vivo*. *J Pharm Sci* 2000;89:951–957.
80. Thanou MM, Nihot MT, Jansen M, Verhoef JC, Junginger HEJ. Mono-*N*-carboxymethyl chitosan (MCC), a polyampholytic chitosan derivative, enhances the intestinal absorption of low molecular weight heparin across intestinal epithelia *in vitro* and *in vivo*. *J Pharm Sci* 2001;90:38–46.
81. Fasano A, Bernadette B, Pumplin D, Wasserman SS, Tall BD, Ketley JM, Kaper JB. *Vibrio cholerae* produces a second enterotoxin, which affects intestinal tight junctions. *Proc Natl Acad Sci USA* 1991;88:5242–5246.
82. Salama NN, Fasano A, Thakar M, et al. The effect of $\Delta G$ on the transport and oral absorption of macromolecules. *J Pharm Sci* 2004;93(5):1310–1319.
83. Wang W, Uzzau S, Goldblum E, Fasano A. Human zonulin, a potential modulator of intestinal tight junctions. *J Cell Sci* 2000;113:4435–4440.
84. Fasano A, Uzzau S. Modulation of intestinal tight junctions by zonula occludens toxin permits enteral administration of insulin and other macromolecules in an animal model. *J Clin Invest* 1997;99(6):1158–1164.
85. Cox D, Raje S, Gao H, Salama NN, Eddington ND. Enhanced permeability of molecular weight markers and poorly bioavailable compounds across Caco-2 cell monolayers using the absorption enhancer, zonula occludens toxin. *Pharm Res* 2002;19:1680–1688.
86. Di Pierro M, Lu R, Uzzau S, Wang W, Margaretten K, Pazzani C, Maimone F, Fasano A. Zonula occludens toxin structure–function analysis. Identification of the fragment biologically active on tight junctions and of the zonulin receptor binding domain. *J Biol Chem* 2001;276:19160–19165.
87. Salama NN, Fasano A, Ruliang L, Eddington ND. Effect of the biologically active fragment of zonula occludens toxin, delta G, on the intestinal paracellular transport and oral absorption of mannitol. *Int J Pharm* 2003;251:113–121.
88. Salama NN, Fasano A, Thakar M, Eddington ND. The impact of $\Delta G$ on the oral bioavailability of low bioavailable therapeutic agents. *J Pharmacol Exp Ther* 2005;312:199–205.
89. Lueßen HL, Rentel CO, Kotzé AF, Lehr CM, de Boer AG, Verhoef JC, Junginger HE. Mucoadhesive polymers in peroral peptide drug delivery. IV. Polycarbophil and chitosan

are potent enhancers of peptide transport across intestinal mucosae in vitro. *J Control Release* 1997;45(1):15–23.

90. Clausen AE, Bernkop-Schnürch A. *In vitro* evaluation of the permeation-enhancing effect of thiolated polycarbophil. *J Pharm Sci* 2000;89:1253–1261.
91. Clausen AE, Kast CE, Bernkop-Schnürch A. The role of glutathione in the permeation enhancing effect of thiolated polymers. *Pharm Res* 2002;19(5):602–608.
92. Guggi D, Bernkop-Schnürch A. Improved paracellular uptake by the combination of different types of permeation enhancers. *Int J Pharm* 2005;288:141–150.
93. Bernkop-Schnürch A, Kast CE, Guggi D. Permeation enhancing polymers in oral delivery of hydrophilic macromolecules: thiomer/GSH systems. *J Control Release* 2003;93: 95–103.
94. Salzman AL, Menconi MJ, Unno N, Ezzel RM, Casey DM, Gonzalez PK, Fink MP. Nitric oxide dilates tight junctions and depletes ATP in cultured Caco-2BBe intestinal epithelial monolayers. *Am J Physiol* 1995;268:6361–6373.
95. Utoguchi N, Watanabe Y, Shida T, Matsumoto M. Nitric oxide donors enhance rectal absorption of macromolecules in rabbits. *Pharm Res* 1998;15:870–876.
96. Yamamoto A, Tatsumi H, Maruyama M, Uchiyama T, Okada N, Fujita T. Modulation of intestinal permeability by nitric oxide donors: implications in intestinal delivery of poorly absorbable drugs. *J Pharm Exp Ther* 2001;296:84–90.
97. Leonard TW, Lynch J, McKenna MJ, Brayden DJ. Promoting absorption of drugs in humans using medium-chain fatty acid-based solid dosage forms: GIPET. *Expert Opin Drug Deliv* 2006;3(5):685–692.
98. Amidon GL, Lennernas H, Shah VP, Crison JR. A theoretical basis for a biopharmaceutic drug classification: the correlation of *in vitro* drug product dissolution and *in vivo* bioavailability. *Pharm Res* 1995;12(3):413–420.
99. Balimane PV, Chong S, Morrison RA. Current methodologies used for evaluation of intestinal permeability and absorption. *J Pharmacol Toxicol Methods* 2000;44(1):301–312.
100. Zornoza T, Cano-Cebrian MJ, Nalda-Molina R, Guerri C, Granero L, Polache A. Assessment and modulation of acamprosate intestinal absorption: comparative studies using *in situ*, *in vitro* (CACO-2 cell monolayers) and *in vivo* models. *Eur J Pharm Sci* 2004;22(5):347–356.
101. Sun D, Yu L, Hussain MA, Wall DA, Smith RL, Amidon GL. *In vitro* testing of drug absorption for drug "developability" assessment: forming an interface between *in vitro* preclinical data and clinical outcome. *Curr Opin Drug Discov Dev* 2004;7(1):75–85.
102. Schultz SG, Curran PF, Chez RA, Fuisz RE. Alanine and sodium fluxes across mucosal border of rabbit ileum. *J Gen Physiol* 1967;50:1241–1260.
103. Munck BG. Transport of imino acids and non-α–amino acids across the brush-border membrane of the rabbit ileum. *J Membr Biol* 1985;83:15–24.
104. Munck LK, Munck BG. Chloride-dependence of amino acid transport in rabbit ileum. *Biochim Biophys Acta* 1990;1027:17–20.
105. Más-Serrano P, Granero L, Martín-Algarra RV, Guerri C, Polache A. Kinetic study of acamprosate absorption in rat small intestine. *Alcohol Alcohol* 2000;35(4):324–330.
106. Condomina J, Zornoza-Sabina T, Granero L, Polache A. Kinetics of zinc transport *in vitro* in rat small intestine and colon: interaction with copper. *Eur J Pharm Sci* 2002;16:289–295.
107. Wilson G, Hassan IF, Dix CJ, Wlliamsson I, Shah R, Mackay R, Artursson P. Transport and permeability properties of human Caco-2 cells: an *in vitro* model of the intestinal epithelial cell barrier. *J Control Release* 1990;11:25–40.

108. Noach A, Hurni M, De Boer A, Breimer D. The paracellular approach: drug transport and its enhancement via the paracellular pathway. In de Boer A, Ed. *Drug Absorption Enhancement*. Chur, Switzerland: Harwood Academic Publishers; 1994, Vol 3, pp 291–324.
109. Doluisio JT, Billups NF, Dittert LW, Sugita ET, Swintosky JV. Drug absorption. I. An *in situ* rat gut technique yielding realistic absorption rates. *J Pharm Sci* 1969;58:1196–1200.
110. Martín-Villodre A, Plá-Delfina JM, Moreno J, Pérez-Buendía D, Miralles J, Collado E, Sánchez-Moyano E, Del Pozo A. Studies on the reliability of a bihyperbolic functional absorption model. I. Ring-substituted anilinas. *J Pharmacokinet Biopharm* 1986;14(6): 615–633.
111. Casabó VG, Núñez-Benito E, Martínez-Coscollá A, Miralles-Loyola E, Martín-Villodre A, Plá-Delfina JM. Studies on the reliability of a bihyperbolic functional absorption model. II. Phenylalkylamines. *J Pharmacokinet Biopharm* 1987;15(6):633–643.
112. Bermejo MV, Pérez-Varona AT, Segura-Bono MJ, Martín-Villodre A, Plá-Delfina JM, Garrigues TM. Compared effects of synthetic and natural bile acid surfactants on xenobiotic absorption I. Studies with polysorbate and taurocholate in rat colon. *Int J Pharma* 1991;69:221–231.
113. Garrigues TM, Collado EF, Fabra-Campos S, Pérz-Buendía MD, Martín-Villodre A, Plá-Delfina JM. Absorption–partition relationships for true homologous series of xenobiotics as a possible approach to study mechanisms of surfactants in absorption. III. Aromatic amines and cationic surfactants. *Int J Pharm* 1989;57:189–196.
114. Merino M, Peris-Ribera JE, Torres-Molina F, Sánchez-Picó A, García-Carbonell MC, Casabó VG, Martín-Villodre A, Plá-Delfina JM. Evidence of a specialized transport mechanism for the intestinal absorption of baclofen. *Biopharm Drug Dispos* 1989;10: 279–297.
115. Sánchez-Pico A, Peris-Ribera JE, Toledano C, Torres-Molina F, Casabo VG, Martin-Villodre A, Pla-Delfina JM. Nonlinear intestinal absorption kinetics of cefradoxil in the rat. *J Pharm Pharmacol* 1989;41:179–185.
116. Ruiz-Carretero P, Merino-Sanjuan M, Nacher A, Casabo VG. Pharmacokinetic models for the saturable absorption of cefuroxime axetil and saturable elimination of cefuroxime. *Eur J Pharm Sci* 2004;21(2–3):217–223.
117. Polache A, Plá-Delfina JM, Merino M. Partially competitive inhibition of intestinal baclofen absorption by beta-alanine, a nonessential dietary amino acid. *Biopharm Drug Dispos* 1991;12:647–660.
118. Rodriguez-Ibanez M, Nalda-Molina R, Montalar-Montero M, Bermejo MV, Merino V, Garrigues TM. Transintestinal secretion of ciprofloxacin, grepafloxacin and sparfloxacin: *in vitro* and *in situ* inhibition studies. *Eur J Pharmacokinet Biopharm* 2003;55(2):241–246.
119. Kim JS, Mitchell S, Kijek P, Tsume Y, Hilfinger J, Amidon GL. The suitability of an *in situ* perfusion model for permeability determinations: utility for BCS class I biowaiver requests. *Mol Pharm* 2006;3(6):686–694.
120. Anderberg EK, Artursson P. Cell cultures to access drug absorption enhancement. In de Boer A, Ed. *Drug Absorption Enhancement*. Chur, Switzerland: Harwood Academic Publishers; 1994, pp 101–118.

# DISTRIBUTION: MOVEMENT OF DRUGS THROUGH THE BODY

Jayanth Panyam[1] and Yogesh Patil[2]

[1]*University of Minnesota, Minneapolis, Minnesota*
[2]*Wayne State University, Detroit, Michigan*

**Contents**

1 Introduction
  1.1 Drug Distribution
  1.2 Effect of Distribution on Drug Action
2 Drug Distribution in Different Compartments
  2.1 Distribution Among Plasma/Blood, Tissues, Organs, and Body Fluids
  2.2 Volume of Distribution ($V_D$)
3 Factors Affecting Drug Distribution
  3.1 Rate of Distribution
  3.2 Extent of Distribution
  3.3 Role of Transporters
  3.4 Pharmacogenetic Factors
4 Modifying Drug Distribution
  4.1 Alterations in Chemical Structure: Prodrug Design
  4.2 Pegylation
  4.3 Use of Another Drug to Modify Distribution
  4.4 Encapsulation in Delivery Systems
5 Conclusion
  References

## 1 INTRODUCTION

### 1.1 Drug Distribution

Drug distribution refers to the reversible movement of drug from blood to various tissues of the body [1]. Following entry into the systemic circulation, a drug

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

distributes into different tissues, the rate and extent of which is dependent on the drug's physicochemical properties and the blood flow to the tissues. The rate and extent of distribution to different tissues, in turn, affects the duration and magnitude of therapeutic effect and toxicity [2]. This chapter explores the various factors that affect drug distribution and the different approaches available to modify drug distribution.

### 1.2 Effect of Distribution on Drug Action

The intensity and duration of therapeutic effects derived from drugs other than those exerting an immediate, irreversible action depend, theoretically, on maintaining adequate concentrations of the active form of the drug at receptor sites [3]. For the purpose of this chapter, receptor sites are considered to be target sites located anywhere in the body that combine with a drug or its active metabolite to produce a pharmacological effect. To elicit a therapeutic effect, a drug has to distribute from the systemic circulation to its site of action [4]. Therefore, it is important to consider the effect of distribution in selecting an appropriate dose, dosage form, dosage interval, and route of administration; appropriate modification of these parameters permits attainment of effective concentrations of the active form of a drug at receptor sites [5].

For blood levels of a drug to correlate with pharmacological effect, the concentration of the drug in blood must be in equilibrium with the concentration of drug at the receptor site through which it produces its effect [6]. Furthermore, the pharmacological response being investigated must bear a direct relationship to the drug's concentration at the receptor site [5, 6]. Certain drugs, such as monoamine oxidase and cholinesterase inhibitors, must accumulate at receptor sites to a certain concentration before pharmacological effects will ensue [7]. During this initial period of drug accumulation at receptor sites, no direct relationship between the drug level in blood and pharmacological effects may be seen. Thus, the distribution process could make a graded response (response proportional to drug concentration), giving the impression of a quantal response (response does not have a linear relationship with drug concentration) [8].

## 2 DRUG DISTRIBUTION INTO DIFFERENT COMPARTMENTS

### 2.1 Distribution Among Plasma/Blood, Tissues, Organs, and Body Fluids

After absorption into the bloodstream, drugs tend to distribute into all body water. The total body water can be divided into intravascular, extracellular, and intracellular water [9]. For a 70 kg person, the total body water is about 42 L. The intravascular water accounts for about 3 L of the total body water while the extracellular and intracellular waters account for 16 L and 23 L, respectively [2]. Drugs that permeate freely through cell membranes become distributed, in time, throughout the body water. Drugs that pass readily through or between capillary endothelial cells, but do not penetrate other cell membranes, are distributed into the extracellular fluid space [10]. Occasionally, the drug molecule may be so polar, large, or highly

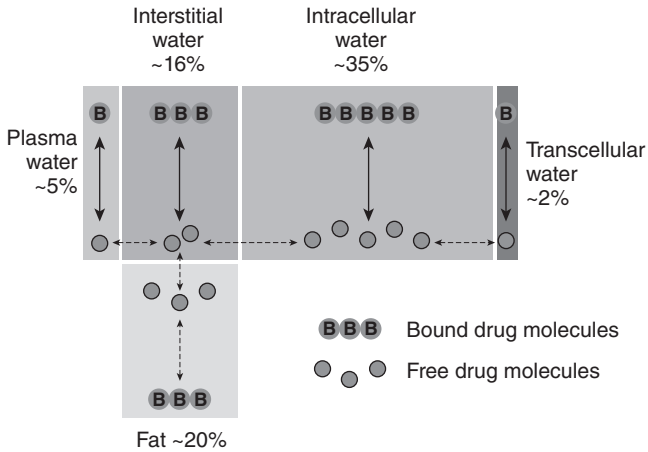


**FIGURE 1** The main body fluid compartments, expressed as a percentage of body weight. Drug molecules exist in bound or free form in each compartment, but only the free drug is able to move between the compartments. (From Ref. 11, with permission.)

bound to plasma proteins that it remains in the intravascular space after IV administration [2].

In all these body fluid compartments, drug molecules are present in either protein-bound or unbound form (Fig. 1), with the equilibrium existing between the unbound form of the drug in different tissues [11]. Furthermore, drugs that are weak acids or bases will exist as an equilibrium mixture of ionized or un-ionized forms, the position of the equilibrium depending on the pH of the compartment. Un-ionized and lipophilic drug molecules can generally diffuse from interstitial space across the cell membrane into the cytoplasm of a cell, while ionized and highly polar drugs cannot readily diffuse across the cell membrane [12].

The effect of pH on diffusion of drugs across a cell membrane could be understood from the gastric pouch experiments [13]. It was shown that drugs cross the gastric epithelium in their nonionized form, and that the ionized form penetrates very slowly, if at all. In the experiments, various acidic and basic drugs were administered intravenously to dogs with gastric pouches, and the concentrations of drug in gastric juice and plasma were measured. At steady state, basic drugs appeared in gastric juice in concentrations ranging from 1 to 40 times that of plasma. In contrast, acidic drugs appeared in gastric juice in low concentrations, which ranged from 0 to 0.6 that of plasma. The results were explained in terms of a model system in which gastric juice is separated from plasma by a barrier permeable only to the nonionized form of a weak electrolyte (Fig. 2) [7]. At steady state, the concentrations of nonionized drug in plasma and gastric juice are the same (correcting for the degree of plasma binding), but the concentrations of the ionized form are unequal because of the difference in the pH of the two fluids. Accordingly, the total concentration of drug (ionized plus nonionized) on both sides of the gastric mucosa is a function of the pH of the two fluids and the dissociation constant of the drug. From this relationship, it can readily be calculated that a basic drug will be more concentrated in tissues with acidic pH than in plasma, and an acidic drug will be more concentrated in plasma than in tissues with acidic pH [7].

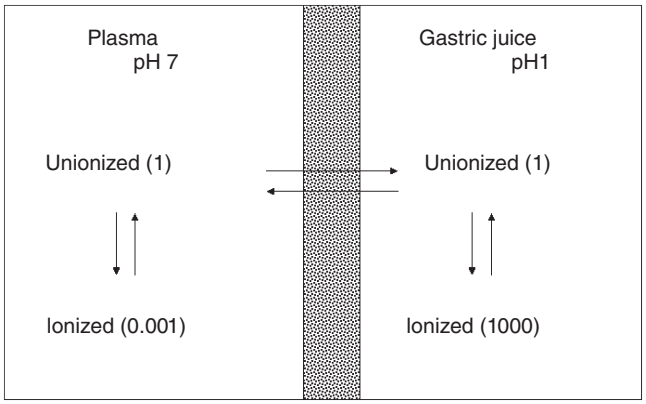


**FIGURE 2** The theoretical distribution of an organic base, p$K_a$ 4, between plasma and gastric juice, assuming that the fluids are separated by a boundary permeable only to the unionized drug molecule. (From Ref. 7, with permission.)

Drugs may also undergo redistribution in the body after initial high levels are achieved in tissues that have a rich vascular supply. As the plasma concentration falls, drug readily diffuses back from a tissue with initial high concentration into the circulation to be quickly redistributed to other tissues with high blood-flow rates, such as the muscles. Then, over time, the drug also becomes deposited in tissues with poor blood supply, such as the fat depots [14]. Some drug molecules are distributed to eliminating organs like liver and kidney. Many drugs are not distributed equally throughout the body but tend to accumulate in certain specific tissues or fluids [15, 16]. For example, chloroquine mainly accumulates in liver, tetracycline in bone and teeth, and iodine in thyroid glands.

## 2.2 Volume of Distribution ($V_D$)

Anatomically, the total volume of body water available for a drug to distribute into is about 42 L [2]. If all drugs were to distribute into all of body water equally, then the initial plasma concentration of all drugs for a given intravenous bolus dose of the drug would be similar. However, this is rarely the case. The plasma concentration of the drug appears to be as if the total body water was less than or greater than 42 L. The reason for the discrepancy is twofold. First, most drugs do not distribute evenly in all body water. Second, only the drug concentration in the intravascular compartment is measurable. Two extreme cases of distribution should be considered for a better understanding. Some molecules like Evan's blue are strictly localized within the intravascular compartment. When the concentration of Evan's blue in the plasma is measured, it appears as if the total volume in which the molecule is distributed is small. For example, following a 100 mg IV bolus dose, the initial plasma concentration measured would be about 100 mg/3 L = 33 mg/L. Now, if the dose (100 mg) and initial concentration (33 mg/L) were used to estimate the volume, this would be about 3 L. For the other extreme, consider a drug molecule like quinine, which accumulates extensively in intracellular water. When the concentration of such a drug molecule in the plasma is measured, it appears as if the total volume in

which the drug is distributed is large. For example, consider a drug molecule that shows 90% partitioning out of the intravascular water. Thus, following a 100 mg dose, 10 mg will stay in the intravascular water while 90 mg will be distributed into extravascular water. The initial concentration measured for this drug after an IV bolus dose would theoretically be about 10 mg/3 L = 3.3 mg/L. If the dose and initial concentration were used to estimate the volume, this would be about 33 L.

In order to account for these differences, the term *apparent volume of distribution* is used. Apparent volume of distribution ($V_D$) is the volume that must be considered in estimating the amount of drug in the body from the concentration of drug found in the sampling compartment [1]. Volume of distribution relates the concentration of the drug in plasma ($C_P$) and the amount of drug in the body ($A$), as in the following equation:

$$V_D = \frac{A}{C_P}$$

Some drugs have $V_D$ as small as 5 L while others have a $V_D$ of 50,000 L. Since $V_D$ is not a true anatomic/physiologic space, it is called *apparent* volume of distribution. Volume of distribution is a measure of the tendency of a drug to move out of blood/plasma to other tissue sites. Thus, the larger the volume of distribution, the more extensively is the drug distributed outside the vasculature. Volume of distribution is useful in the determination of plasma drug concentration when a known amount of drug is in the body or conversely in the determination of dose required to achieve a particular plasma concentration. Examples of drugs with small and large volumes of distribution are shown in Table 1 [11]. It is important to note that volume of distribution is dependent on a patient's body weight and is often quoted with the units of L/kg [11].

## 3 FACTORS AFFECTING DRUG DISTRIBUTION

There are several factors that affect the rate and extent of drug distribution. Blood perfusion to a tissue and permeability of the drug across cell membrane play an important role in determining the rate of distribution. Extent of distribution is mainly affected by molecular weight, lipid solubility, $pK_a$, and protein binding of the drug. Presence of drug efflux transporters and genetic differences in transporter activity in different patient populations also significantly affect drug distribution.

### 3.1 Rate of Distribution

Distribution of drugs from blood to different tissues is a two-step process [11]. The first step involves the transport of drug by blood to different tissues. The second step is the transport of the drug from blood into the tissue across the cell membrane. For lipophilic, small molecular weight drugs that can cross cell membranes efficiently, transport to the tissue is the rate-limiting step, while for hydrophilic and large molecular weight drugs, transport across the cell membrane is the rate-limiting step. Thus, depending on the drug's characteristics, drug distribution could be either permeability limited or perfusion limited.

**TABLE 1 Distribution Volumes for Some Drugs Compared with Volume of Body Fluid Compartments**

| Volume (L/kg body weight) | Compartment | Volume of Distribution ($V_D$; L/kg body weight) | |
|---|---|---|---|
| 0.05 | Plasma | 0.05–0.1 | Heparin |
| | | | Insulin |
| | | 0.1–0.2 | Warfarin |
| | | | Sulfamethoxazole |
| | | | Glibenclamide |
| | | | Atenolol |
| 0.2 | Extracellular fluid | 0.2–0.4 | Tubocurarine |
| | | 0.4–0.7 | Theophylline |
| 0.55 | Total body water | | Ethanol |
| | | | Neostigmine |
| | | | Phenytoin |
| | | 1–2 | Methotrexate |
| | | | Indomethacin |
| | | | Paracetamol |
| | | | Diazepam |
| | | | Lidocaine |
| | | 2–5 | Glyceryl trinitrate |
| | | | Morphine |
| | | | Propranolol |
| | | | Digoxin |
| | | | Chlorpromazine |
| | | >10 | Nortriptyline |

***Permeability-Limited Distribution*** The cell membrane separates extra- and intracellular compartments. The cell membrane allows a bidirectional, dynamic, and selective exchange of organic molecules, ions, and gas molecules between the two compartments. Small, noncharged lipid molecules pass through the membrane freely. Small polar molecules (carbon dioxide, water) can also pass cell membranes easily following their concentration gradient. Most hydrophilic molecules and macromolecules cannot cross the cell membrane effectively [17].

Thus, cell permeability is the limiting factor in the tissue distribution of polar and high molecular weight drugs. For such drugs, equilibrium is reached faster in tissues without permeability constraints, even though these tissues may have a lower perfusion rate. For example, brain has a perfusion rate of 0.5 mL/min/mL while muscle has a perfusion rate of 0.025 mL/min/mL. However, the cell junctions in the brain capillary endothelium are tighter than those found in the muscle capillary endothelium. Lipophilic drugs that can cross the cell membrane efficiently reach equilibrium faster in brain than in muscle while polar molecules reach equilibrium faster in muscles [18].

Cell permeability varies in characteristics depending on the tissue [19]. The endothelial cells of the brain capillaries, which are more tightly joined to one another than those found in other capillaries, contribute to the slower diffusion of water-soluble drugs into brain. Another barrier to water-soluble drugs is the glial

connective tissue cells (astrocytes), which form an astrocytic sheath close to the basement membrane of the brain capillary endothelium [14]. The tight junctions between brain vascular endothelial cells lead to high endothelial electrical resistance, in the range of 1500–2000 $\Omega \cdot cm^2$ (pial vessels), as compared to 3–33 $\Omega \cdot cm^2$ in other tissues [20, 21]. Thus, polar compounds cannot enter the brain effectively but can enter the interstitial fluids of most other tissues [22].

Even within a tissue, the distribution of a drug may vary between the different regions of the tissue. For example, drugs penetrate into the brain cortex more rapidly than into white matter, probably because of the greater delivery rate of drug via the bloodstream to the tissue [23]. The consequences of the diverse rates of entry of different drugs into the CNS include the following: (1) water-soluble or ionized drugs will not enter the CNS very well; (2) low ionization, low plasma–protein binding, and a fairly high lipid–water partition coefficient confer ready penetration; and (3) direct injections into the CSF often produce unexpected effects [24, 25].

***Perfusion-Limited Distribution*** For most lipophilic drugs with small molecular weight, cell membranes do not create any barrier to their distribution [26]. For such drugs, distribution to different tissues is mainly governed by the blood perfusion rate to that tissue. Perfusion rate is often expressed as milliliter of blood per minute per unit volume of tissue. Perfusion rate could be as much as 10 mL/min/mL for lungs to only 0.025 mL/min/mL for resting muscle (Table 2) [1].

**TABLE 2 Blood Flow, Perfusion Rate, and Relative Size of Different Organs and Tissues Under Basal Conditions in a Standard 70 kg Human[a]**

| Organ[b] | Percent of Body Volume | Blood Flow (mL/min) | Percent of Cardiac Output | Perfusion Rate (mL/min/mL of Tissue) |
|---|---|---|---|---|
| Adrenal glands | 0.03 | 25 | 0.2 | 1.2 |
| Blood | 7 | (5000)[b] | (100) | — |
| Bone | 16 | 250 | 5 | 0.02 |
| Brain | 2 | 700 | 14 | 0.5 |
| Fat | 20[c] | 200 | 4 | 0.03 |
| Heart | 0.4 | 200 | 4 | 0.6 |
| Kidneys | 0.5 | 1100 | 22 | 4 |
| Liver | 2.3 | 1350 | 27 | 0.8 |
| Portal | 1.7 (gut) | (1050) | (21) | — |
| Arterial | — | (300) | (6) | — |
| Lungs | 1.6 | (5000) | (100) | 10 |
| Muscle (inactive) | 43 | 750 | 15 | 0.025 |
| Skin (cool weather) | 11 | 300 | 6 | 0.04 |
| Spleen | 0.3 | 77 | 1.5 | 0.4 |
| Thyroid gland | 0.03 | 50 | 1 | 2.4 |
| *Total Body* | *100* | *5000* | *100* | *0.071* |

[a]Compiled and adapted from data in Guyton AC. *Texbook of Medical Physiology*, 7th ed. Philadelphia: Saunders; 1986, p 230; Lentner C, Ed. *Geigy Scientific Tables, Volume 1*. Edison, NJ: Ciba-Geigy; 1981; and Davies B, Morris T. Physiological parameters in laboratory animals and humans. *Pharm Res* 1993;10:1093–1095.

[b]Some organs (e.g., stomach, intestines, and pancreas) are not included.

[c]Includes fat within organs.

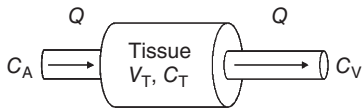


**FIGURE 3** Perfusion-limited tissue distribution: $Q$ = rate of blood flow; $C_A$ = arterial concentration; $C_V$ = venous concentration; $V_T$ = tissue volume; and $C_T$ = tissue concentration.

All other factors (membrane permeability, protein binding, pH, etc.) remaining equal, well-perfused tissues take up more drug than poorly perfused tissues. There is a direct correlation between tissue perfusion rate and time to distribute to a tissue. This can be explained as follows.

Consider a tissue T whose volume is $V_T$ and the concentration of the drug in the tissue is $C_T$ (Fig. 3). The rate of blood flow to the tissue is $Q$. If $C_A$ is concentration of the drug in arterial blood and $C_V$ is the concentration of the drug in venous blood, then the rate of presentation of the drug to the tissue is given by $Q \cdot C_A$, while the rate of leaving of drug from the tissue is given by $Q \cdot C_V$. The net rate of drug uptake by the tissue is $Q \cdot (C_A - C_V)$.

The amount of the drug in the tissue is obtained using the volume of the tissue and the concentration of the drug in the tissue: amount of drug in tissue = $V_T \cdot C_T$. The relative affinity of the drug for the tissue can be defined in terms of a partition coefficient $K_P$, where $K_P = C_T/C_V$. Rearranging this relationship, and substituting for $C_T$ in the above equation, we get

$$\text{Amount of drug in tissue} = V_T \cdot K_P \cdot C_V$$

The exit of the drug from the tissue can be defined by the fractional rate of exit, which is given by

$$\text{Fractional rate of exit} = k_T = \text{Rate of exit/Amount in tissue}$$

Substituting for rate of exit and the amount of the drug in the tissue, we get

$$k_T = (Q \cdot C_V)/(V_T \cdot K_P \cdot C_V) = (Q/V_T)/K_P$$

where $Q/V_T$ is perfusion rate of tissue and $K_P$ is the partition coefficient of the drug for the tissue. The time to distribute into a tissue is given by tissue distribution half-life (time for 50% distribution):

$$\text{Tissue distribution half-life} = 0.693/k_T = 0.693K_P/(Q/V_T)$$

If each organ has the same ability to store the drug ($K_P$ is equal), then the distribution half-life is governed by the blood flow $Q$ and the volume (size) $V_T$ of the organ. A large blood flow to the organ decreases the distribution time; whereas a large organ size increases the distribution time because a longer time is needed to fill a large organ volume with drug. Also, it is important to note that if a tissue has a long distribution half-life, a long time is needed for the drug to leave the tissue

when blood level decreases. If arterial concentration is maintained constant, tissue concentration goes up but rate of uptake decreases with time. Tissue concentration ($C_T$) at any time $t$ is given by the following relationship:

$$C_T = K_P C_A (1 - e^{-k_T t})$$

Equilibrium in tissue concentration and the loss of drug from tissue take longer the poorer the perfusion and the greater the partitioning of drug into the tissue.

### 3.2 Extent of Distribution

Factors like molecular weight, lipid solubility, pH–p$K_a$, and protein binding play an important role in determining the extent of drug distribution. It is important to note that all the above factors also affect the rate of drug distribution into the tissue.

***Molecular Weight*** The diffusion of a molecule across a cell membrane depends mainly on its molecular size, the diffusion coefficient for small molecules being inversely proportional to the square root of molecular weight. Consequently, while large molecules diffuse at a slower rate than small ones, the effect of molecular weight for most drug molecules is modest. Most drugs fall within the molecular weight range 200–1000, and variations in aqueous diffusion rate have only a small effect on their overall pharmacokinetic behavior. However, beyond this molecular weight range, diffusion of molecules across cell membranes is considerably limited. For compounds with molecular diameter above 100 Å, transfer across membranes is significantly slow [11].

***Lipid Solubility*** Distribution of drugs is significantly influenced by lipid solubility. Lipid solubility affects the ability of the drug to bind to plasma proteins and to cross lipid membrane barriers [27]. A drug needs to be lipid soluble to penetrate membranes, unless there is an active transport system for the drug or the drug is so small that it can pass through the aqueous channels in the membrane. For weakly acidic and weakly basic drugs, ionization, and therefore lipid solubility, is pH dependent (see pH–p$K_a$ discussion). As discussed earlier, very high lipid solubility can result in a drug initially partitioning preferentially into highly vascularized lipid-rich areas. Subsequently, these drugs slowly redistribute into body fat, where they may remain for long periods of time.

***pH–p$K_a$*** Most drugs are weak acids or bases and therefore exist in both un-ionized and ionized forms, the ratio of the two forms varying with pH. For a weak base, the ionization reaction is

$$BH^+ \xleftrightarrow{K_a} B + H^+$$

The negative logarithm of the acid dissociation constant is designated by the symbol p$K_a$ and is given by the Henderson–Hasselbalch equation:

$$pK_a = pH + \log_{10}([BH^+]/[B])$$

For a weak acid,

$$AH \xleftrightarrow{K_a} A^- + H^+$$

$$pK_a = pH + \log_{10}([AH]/[A^-])$$

Drug accumulates on the side of a membrane where pH favors greater ionization of drug. This is known as the pH-partition hypothesis [28]. Only un-ionized nonpolar drug can permeate through the lipid membrane, and at equilibrium, the concentration of the un-ionized species is the same on both sides, but there will be more total drug on the side on which the degree of ionization is greater. Basic drugs tend to accumulate in tissues with pH values lower than the $pK_a$ of the drug; conversely, acidic drugs concentrate in regions of higher pH, provided that the free drug is sufficiently lipid soluble to be able to penetrate the membranes that separate the compartments. Even small differences in pH across membranes, such as those that exist between CSF (pH 7.3) and plasma (pH 7.4), milk (pH 6.5–6.8) and plasma, renal tubular fluid (pH 5.0–8.0) and plasma, and inflamed tissue (pH 6.0–7.0) and healthy tissue (pH 7.0–7.4), can lead to unequal distribution of drugs [11].

Figure 4 shows how a weak acid (e.g., aspirin, $pK_a$ 3.5) and a weak base (e.g., pethidine, $pK_a$ 8.6) would be distributed at equilibrium between three body compartments, namely, plasma (pH 7.4), alkaline urine (pH 8), and gastric juice (pH 3). Within each compartment, the ratio of ionized to un-ionized drug is governed by the $pK_a$ of the drug and the pH of that compartment [11]. It is assumed that the un-ionized species can cross the membrane and therefore reaches an equal concentration in each compartment. The ionized species is assumed not to cross at all. The result is that, at equilibrium, the total (ionized + un-ionized) concentration of the drug will be different in the different compartments, with an acidic drug being concentrated in the compartment with high pH ("ion trapping"), and vice versa [29]. The concentration gradients produced by ion trapping can theoretically be very large, if there is a large pH difference between compartments. Thus, aspirin would be concentrated more than fourfold with respect to plasma in alkaline renal tubule, and about 6000-fold in plasma with respect to the acidic gastric contents. Such large gradients, however, are unlikely to be achieved in reality for two main reasons. First, the attribution of total impermeability to the charged species is not realistic, and even a small permeability will considerably attenuate the concentration difference that can be reached. Second, body compartments rarely approach equilibrium. Neither the gastric contents nor the renal tubular fluid stands still, and the resulting flux of drug molecules reduces the concentration gradients well below the theoretical equilibrium conditions. The pH-partition mechanism nonetheless correctly explains some of the qualitative effects of pH changes in different body compartments on the pharmacokinetics of weakly acidic or basic drugs, particularly in relation to penetration of the blood–organ barriers. Values of $pK_a$ for some common drugs are shown in Fig. 5 [1].

***Protein Binding*** Binding to proteins in tissue, dissolution in adipose tissue, formation of nondiffusible complexes in tissues such as bone, incorporation into specific storage granules, or binding to selective sites in tissues all impede movement of drugs in the body and account for differences in the cellular and organ distribution

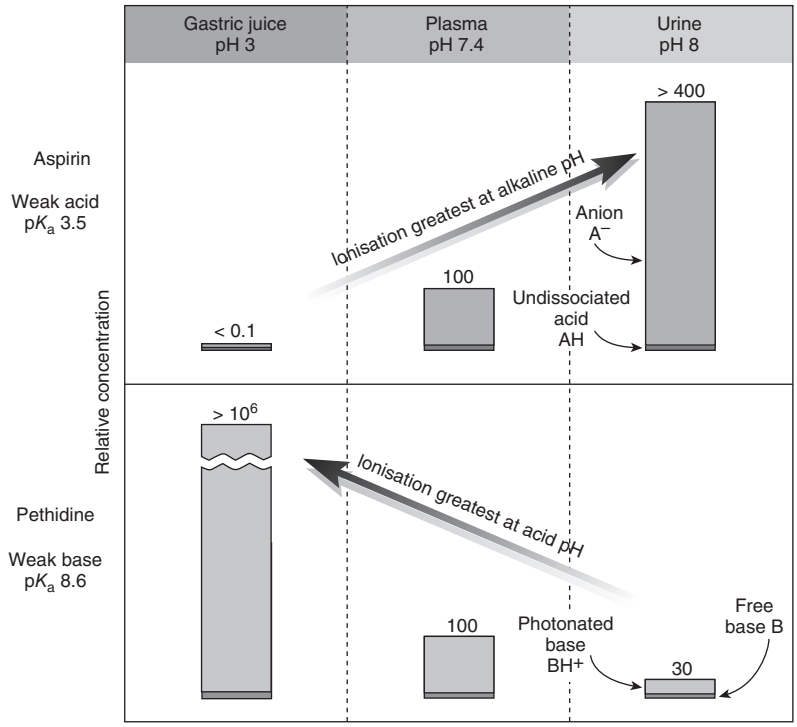


**FIGURE 4** Theoretical partition of a weak acid (aspirin) and a weak base (pethidine) between aqueous compartments (urine, plasma, and gastric juice) according to the pH difference between them. Numbers represent relative concentrations (total plasma concentration = 100). It is assumed that the uncharged species in each case can permeate the cellular barrier separating the compartments and therefore reaches the same concentration in all three. Variations in the fractional ionization as a function of pH give rise to the large total concentration differences with respect to plasma. (From Ref. 11, with permission.)

of particular drugs [30, 31]. Extensive plasma–protein binding will cause more of the drug to stay in the central blood compartment, because only the unbound or free fraction of a drug can diffuse out of capillaries into tissues. Therefore, drugs that bind strongly to plasma proteins tend to have lower volumes of distribution. Similarly, drugs that bind extensively to tissue proteins (proteins outside the vascular compartment) tend to have large volumes of distribution [32].

The most important binding of drugs in circulation is to plasma albumin, which comprises about 50% of the total plasma proteins and binds the widest range of drugs [33]. Acidic drugs commonly bind to albumin, while basic drugs often bind to $\alpha_1$-acid glycoproteins and lipoproteins (Table 3) [34]. Many endogenous substances, steroids, vitamins, and metal ions are bound to globulins. A drug may become bound to plasma proteins to a greater or lesser degree, depending on a number of factors, for example, plasma pH, concentration of plasma proteins, concentration of the drug, the presence of another agent with a greater affinity for the limited number of binding sites, and the presence of acute-phase proteins during active inflammatory conditions [9]. The degree of plasma–protein binding and the affinity of a drug

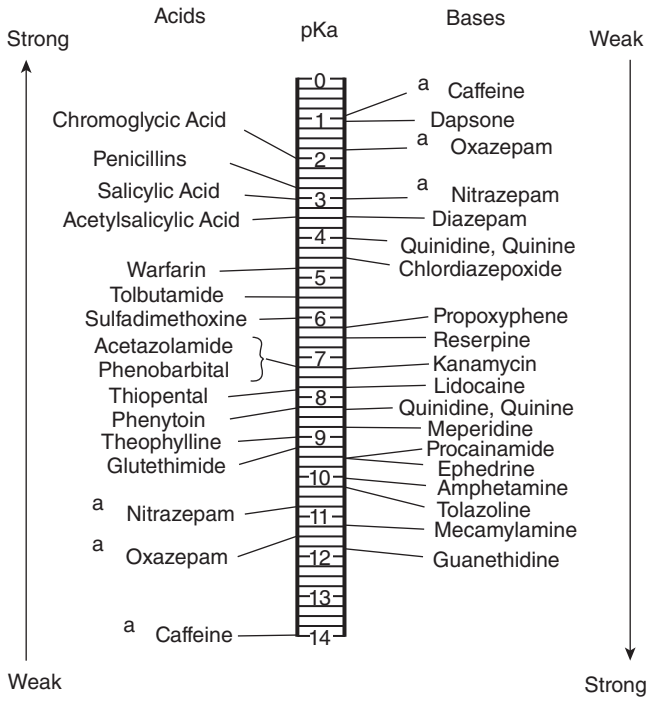


**FIGURE 5** The p$K_a$ values of acidic and basic drugs vary widely. Some of drugs are amphoteric (a): that is, they have both acidic and basic functional groups. (From Ref. 1, with permission.)

**TABLE 3 Potential Binding Sites of Proteins for Various Drugs**

| Binding Sites | Drugs |
|---|---|
| For Acidic Agents: | |
| Albumins | Bilirubin, bile acids, fatty acids, vitamin C, salicylates, sulfonamides, barbiturates, phenylbutazone, penicillins, tetracyclines, probenecid |
| For Basic Agents: | |
| Globulins, $\alpha_1$, $\alpha_2$, $\beta_1$, $\beta_2$, $\gamma$ | Adenosine, quinacrine, quinine, streptomycin, chloramphenicol, digoxin, ouabain, coumarin |

for the nonspecific protein-binding sites are of great clinical significance in some instances and much less so in others [35]. For example, a potentially toxic compound (such as dicumarol) may be 98% bound, but if for any reason it becomes only 96% bound, then the concentration of the free active drug that becomes available in the plasma is doubled, with potentially harmful consequences. The concentration of a drug administered in overdose may exceed the binding capacity of the plasma protein and lead to an excess of free drug, which can distribute into various target tissues and produce exaggerated effects. Of equal importance is the readiness with which drugs dissociate from plasma proteins. Those that are more tightly bound tend to have much longer elimination half-lives, because they are released gradually from

the plasma protein reservoir. The long-acting sulfonamides are good examples of this phenomenon.

### 3.3 Role of Transporters

Drug transporters, proteins that transport drugs across the cell membrane, either from or to the intravascular compartment, play an important role in determining drug disposition. Transporters have been shown to be important in the disposition of endogenous compounds, drugs, and other xenobiotics in many organs such as the intestine, liver, kidney, and brain [36–39]. This section focuses on the role of primary transport proteins like ATP-binding cassette transporters and secondary and tertiary transporters like organic cation and anion transporters in the distribution of drugs.

***Primary Transporters (ATP-Binding Cassette Transporters)*** It is now well recognized that membrane efflux transporters, especially P-glycoprotein (P-gp), play an important role in determining the absorption, distribution, metabolism, excretion, and toxicology (ADMET) behaviors of many drugs and molecules in development. Many of the important mammalian efflux transporters are members of the adenosine triphosphate (ATP)-binding cassette (ABC) superfamily, where transport is driven by hydrolysis of ATP. The first member of this family to be discovered was P-gp. P-gp is a 170 kDa, integral membrane protein found in most organisms from bacteria to humans. It was originally identified as a key reason for the development of multidrug resistance in certain cancers [40]; however, constitutive expression of P-gp in many normal tissues such as intestinal epithelia, hepatocytes, kidney proximal tubules, blood–brain barrier endothelia, and placental trophoblast demonstrates its protective roles in limiting drug absorption and distribution, contributing to pharmacokinetics, and potentially impacting pharmacodynamics and toxicity [41, 42]. A list of P-gp substrates is shown in Table 4. The tissue localization and probable role of transport proteins in drug absorption, distribution, and excretion are depicted in Fig. 6.

P-gp is only one of 48 known transporters belonging to seven subfamilies (A–G) [43]. So the potential for other transporters to affect drug disposition is high. As the importance of transport proteins in drug disposition emerges, it is also clear that these transport proteins are saturable, inducible, can be inhibited, and display some degree of polymorphism—factors that need to be considered with respect to variability in drug disposition and response [44–47].

As discussed earlier, the extent of membrane permeability is largely governed by the physicochemical properties of drugs, and lipophilicity is generally considered a key determinant in extent of drug absorption, hepatic transport, and brain penetration. However, increased lipophilicity alone is not predictive of increased permeability when the transport proteins contribute to drug transport across the membrane. While determining the kinetics of drug movement across a barrier, contributions of both passive diffusion and active transport have to be considered. The flux of a drug across a membrane is linearly related to drug concentration, if only passive diffusion is present. Active transport (forward transport or efflux) is usually saturable at high concentrations. If both passive diffusion and active transport are present, net flux will depend on the relative contribution of the two processes and the concentration

**TABLE 4 List of P-Glycoprotein Substrates**

| |
|---|
| Anthracyclines (doxorubicin, daunomycin, epirubicin) |
| Acridines (m-AMSA) |
| Azatoxins (azatoxin) |
| Benxzoheptalene compounds (colchicine) |
| Benzothiazepines (diltiazem) |
| Dihydropyridines (azidopine, nicardipine) |
| Epipodophyllotoxins (etoposide) |
| Isoquinolines (cepharantine) |
| Macrolides (FK506) |
| Organometallic cations (99mTc-sestamibi) |
| Peptides (actinomycin D, bleomycin, cyclosporin A, valinomycin) |
| Phenothiazines (chlorpromazine) |
| Phenylalkylamines (verapamil, tiapamil) |
| Pyrroloindoles (mitomycin C) |
| Quinolines (chloroquine, quinidine) |
| Rhodamines (rhodamine 123) |
| Steroids (aldosterone, corticosterone, cortisol, dexamethasone, testosterone) |
| Taxanes (paclitaxel, docetaxel) |
| Vinca alkaloids (vincristine, vinblastine) |
| Other (digoxin, Hoechst 33342) |

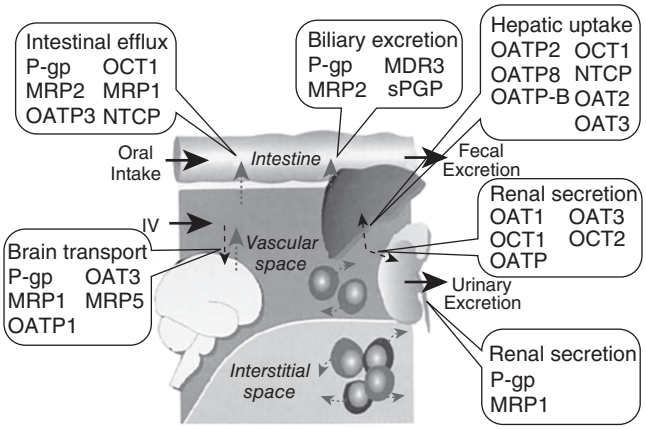


**FIGURE 6** Tissue localization and role of transport proteins in drug disposition. (From Ref. 19, with permission.)

of drug. At very high concentrations, the contribution of active transport to the kinetics of drug flux across the membrane is minimized.

***Secondary and Tertiary Transporters (Driven by an Exchange or Cotransport of Intracellular and/or Extracellular Ions)*** Besides the ATP family of transporters, other transporters, such as the organic anion transporters (OATs) and the organic cation transporters (OCTs) are present in the human body and significantly influence drug distribution [48].

Organic anion transporters (OATs) play an essential role in the disposition of clinically important anionic drugs, including antiviral drugs, antitumor drugs, antibiotics, antihypertensives, and anti-inflammatory agents. The activities of OATs are directly linked to drug toxicity and drug–drug interactions. So far, four members of the OAT family have been identified: OAT1, OAT2, OAT3, and OAT4 [49]. These transporters share several common structural features including 12 transmembrane domains, multiple glycosylation sites localized in the first extracellular loop between transmembrane domains 1 and 2, and multiple phosphorylation sites present in the intracellular loop between transmembrane domains 6 and 7 and in the carboxyl terminus. The impact of these structural features on the function of these transporters has just begun to be explored. OAT1 and OAT3 are predominantly expressed in the kidney and brain. OAT2 is predominantly expressed in liver. OAT4 is present mainly in placenta and kidney. These transporters are multispecific with a wide range of substrate recognition [49].

OAT1 has been shown to interact with a wide range of organic anion drugs such as nonsteroidal anti-inflammatory drugs (NSAIDs), β-lactam antibiotics, antiviral drugs, diuretics, antitumor drugs, and angiotensin-converting enzyme inhibitors. The prototype substrate for OAT1 is para-aminohippuric acid (PAH) [50]. OAT2 interacts with organic anion drugs, such as NSAIDs, and antibiotics. Unlike for OAT1, PAH is a low affinity substrate for OAT2. For OAT3, the prototype substrates are sulfate- and glucuronide-conjugated steroids. OAT3 also interacts with various drugs and endogenous substances such as NSAIDs, antitumor drugs, H2-receptor antagonists, prostaglandins, diuretics, angiotensin-converting enzyme inhibitors, β-lactam antibiotics, and various neurotransmitter metabolites. PAH is a low affinity substrate for OAT3. OAT4 interacts with sulfate-conjugated steroids, antibiotics, and ochratoxin A and shows very little transport of PAH. Compared with OAT1 and OAT3, OAT4 has much narrower substrate specificity [51].

Organic cation transporters (OCTs) are critical in drug absorption, targeting, and disposition. It has become increasingly clear that multiple mechanisms are involved in organic cation transport in the key tissues responsible for drug absorption and disposition: kidney, liver, and intestine. Drugs from a wide array of clinical classes—including antihistamines, skeletal muscle relaxants, antiarrhythmics, and adrenoceptor blocking agents—are organic cations. In addition, several endogenous bioactive amines—such as dopamine, choline, and *N*1-methylnicotinamide (NMN)—are organic cations. Since many of these molecules ($pK_a$ 8–12) are polar and positively charged at physiologic pH, OCTs generally are involved in the absorption, distribution, and elimination of these compounds [52].

### 3.4 Pharmacogenetic Factors

Pharmacogenetics deals with inherited differences in response to drugs. It is well recognized that most medications exhibit wide interpatient variability in their efficacy and toxicity. For many drugs, these interindividual differences are due, in part, to polymorphisms in genes encoding drug-metabolizing enzymes, drug transporters, and/or drug targets (e.g., receptors, enzymes). Pharmacogenomics is aimed at elucidating the genetic basis for differences in drug efficacy and toxicity, and it uses genome-wide approaches to identify the network of genes that govern an individual's response to drug therapy.

***Polymorphism*** The best recognized examples are genetic polymorphisms of drug-metabolizing enzymes, which affect about 30% of all drugs [53]. Loss of function of thiopurine *S*-methyltransferase (TPMT) results in severe and life-threatening hematopoietic toxicity if patients receive standard doses of mercaptopurine and azathioprine. Gene duplication of cytochrome P4502D6 (CYP2D6), which metabolizes many antidepressants, has been identified as a mechanism of poor response in the treatment of depression. There is also a growing list of genetic polymorphisms in drug targets that have been shown to influence drug response. In this section, we discuss genetic polymorphisms in drug transporters that are involved in drug distribution.

As previously described, many drugs are substrates of active transporters, membrane proteins that maintain cellular homeostasis by importing and exporting endogenous compounds. Because of their localization in intestinal, hepatic, and renal epithelial cells, these transporters are important in the absorption, distribution, bioavailability, and elimination of many drugs. Moreover, they can be important in targeting drugs to organs because they are localized in blood–organ barriers.

Genetic polymorphisms affecting a transporter's expression or affinity for substrates can alter drug concentrations at the site of action despite similar blood concentrations. MDR1, MRPs, OATPs, OCTs, OATs, and nucleoside transporters are of particular interest because they transport exogenous substrates, including drugs, as well as endogenous compounds. P-gp, the product of *ABCB1* gene, has received much attention because its substrates include many important drugs. Systemic screening initially revealed 15 genetic variants of human *ABCB1*. A total of 28 SNPs have now been identified; those in exons 21 (G2677T) and 26 (C3435T) are of particular interest because they affect expression or function [47, 55]. Although several studies have addressed the association of these variants with disposition and effects of P-gp substrates, controversy remains about the influence of different variants on pharmacokinetics and pharmacodynamics. Tissues studied so far have shown an average eight- to tenfold difference in P-gp expression. The 3435CC and 3435TT genotypes show a two- to threefold difference in P-gp expression in duodenum, kidney, peripheral leukocytes, and placenta, with substantial overlap between genotypes. This modest difference suggests a moderate impact of the *ABCB1* genotypes on the disposition and effects of P-gp substrates; nongenetic factors probably play an important role in modifying P-gp expression. Differences of ~25–35% in the bioavailability and renal clearance of digoxin in relation to the exon 21 or exon 26 SNP have been reported [56–59]. Several studies [60, 61] have addressed the relevance of *ABCB1* polymorphism to dose requirements, blood concentrations, chronic rejection, and chronic nephrotoxicity in renal transplant patients receiving the calcineurin inhibitors cyclosporin A and tacrolimus.

P-gp expressed on the luminal side of endothelial cells of the brain capillaries significantly limits the transfer of many drugs from blood to brain, as evidenced by a huge increase in brain-to-blood concentration in the P-gp knockout mouse [54]. Because many CNS-active drugs are P-gp substrates, differences in *ABCB1* expression at the blood–brain barrier could help explain why patients with identical plasma drug concentrations respond differently and have different side effects. The consequences of genetic polymorphism have been assessed *in vivo* for another transporter of relevance for drug therapy, OATP-C (SCP1B1), which facilitates the uptake of drug substrates from the blood into the hepatocyte. The relatively common

**TABLE 5 Genetic Polymorphisms of Human Transporters**

| | | |
|---|---|---|
| BSEP | Conjugates | Not yet elucidated |
| MDR-1 | Natural product anticancer drugs, CYP3A4 substrates, digoxin | Not yet elucidated |
| MRPs | Glutathione, glucuronide, and sulfate conjugates, nucleoside antivirals | Not yet elucidated |

variant OATP-C*5 is associated with markedly reduced transporter function [62]. Carriers of the *5 allele have high plasma concentrations of the OATP-C substrate pravastatin [63–65], suggesting impaired uptake of pravastatin by hepatocytes. Indeed, pravastatin concentrations in hepatocytes are low, which results in less inhibition of cholesterol synthesis as assessed by decreased lathosterol concentration and lathosterol/cholesterol ratio [65]. Whether cholesterol-lowering efficacy is impaired in carriers of these variants during long-term treatment is yet unknown. A profound impact of OATP-C polymorphism was recently demonstrated for the antidiabetic drug repaglinide, for which AUC values were approximately three times higher in carriers of the variant *5 allele than in wild-type subjects. This effect was associated with a more pronounced reduction of blood glucose levels [65].

Table 5 provides a list of human drug transporters that exhibit functional genetic polymorphisms and their substrates [66].

## 4 MODIFYING DRUG DISTRIBUTION

A number of drugs often exhibit undesirable pharmacokinetic properties. Some of these include unfavorable distribution profiles such as lack of penetration into target tissues or binding to specific tissues that could result in toxicity. A number of approaches are available to modify the distribution profile of a drug. These include alterations in the chemical structure of the drug, use of another drug that alters the pharmacokinetics of the drug under investigation, and encapsulation of the drug in a delivery system.

### 4.1 Alterations in Chemical Structure: Prodrug Design

Prodrug design strategies have been employed to improve the efficacy of drugs with undesirable pharmacokinetic properties such as chemical instability and lack of specificity. Targeted prodrug design represents a strategy for site-directed and efficient drug delivery. Targeting of drugs to transporters and receptors to aid in site-specific carrier-mediated absorption is emerging as a novel and clinically significant approach. Various prodrugs have been successful in achieving the goals of enhanced availability and are therefore considered to be an important tool in biopharmaceutics.

Strategies in targeted prodrug design include antibody-directed enzyme prodrug therapy, gene-directed enzyme prodrug therapy, and peptide transporter-associated prodrug therapy. The term *prodrug* or *proagent* was first introduced to signify pharmacologically inactive chemical derivatives that could be used to alter the

physicochemical properties of drugs, in a temporary manner, to increase their usefulness and/or to decrease associated toxicity [67]. Often, use of the term prodrug implies a covalent link between a drug and a chemical moiety, although some authors also use it to characterize some forms of salts of the active drug molecule. Although there is no universal definition for a prodrug itself, and the definition may vary, generally prodrugs can be defined as pharmacologically inert chemical derivatives that can be converted *in vivo* to active drug molecules, enzymatically or nonenzymatically, to exert a therapeutic effect. Ideally, the prodrug should be converted to the original drug at the target site of action, followed by subsequent rapid elimination of the released derivatizing group [68, 69].

Prodrugs can be designed to target specific enzymes or carriers by considering enzyme–substrate specificity or carrier–substrate specificity in order to overcome various undesirable drug properties. This type of "targeted-prodrug" design requires considerable knowledge of particular enzymes or carrier systems, including their molecular and functional characteristics. Targeted prodrug designs are classified into two categories: (1) targeting specific enzymes and (2) targeting specific membrane transporters.

***Prodrug Design Targeting Enzymes*** In prodrug design, enzymes can be recognized as presystemic metabolic sites or prodrug–drug *in vivo* reconversion sites. Usually, targeting enzymes to reduce the presystemic metabolism is more successfully achieved by irreversible chemical modification rather than by a prodrug approach. Therefore, our discussion focuses on the enzymes as *in vivo* reconversion targets for prodrugs. The enzyme-targeted prodrug approach can be used to improve oral drug absorption, as well as site-specific drug delivery. In the case of improving oral drug absorption, gastrointestinal enzymes may be the main targets for prodrug design, and the use of a nutrient moiety as a derivatizing group permits more specific targeting for gastrointestinal enzymes to improve oral drug absorption [70]. These prodrugs have the additional advantage of producing nontoxic nutrient by-products when they regenerate the active drugs *in vivo*.

Site-specific drug delivery can be obtained from tissue-specific activation of a prodrug, which is the result of metabolism by an enzyme that is either unique for the tissue or is present at a higher concentration (compared with other tissues); thus, it activates the prodrug more efficiently. This type of site-specific drug delivery has been of particular interest in cancer chemotherapy. Appropriately designed prodrugs have been found to be effective in the treatment of animal tumors possessing high levels of an activating enzyme [71, 72]. However, clinical results were disappointing when it was found that human tumors containing appropriately high levels of the activating enzymes were rare and that the high levels of activating enzymes were not associated with any particular type of tumor [73]. Recently, new therapies have been proposed to overcome this limitation of prodrug therapy. These new approaches are referred to as ADEPT (antibody-directed enzyme prodrug therapy) and GDEPT (gene-directed enzyme prodrug therapy), which attempt the localization of prodrug activation enzymes into specific cancer cells prior to prodrug administration.

***Prodrug Design Targeting the Membrane Transporters*** Although the classical approach to improve membrane permeability of polar drugs uses lipophilic derivatives to increase passive membrane penetration, the targeted prodrug approach uses

transporters designed for facilitating membrane transport of polar nutrients such as amino acids and peptides. There is direct and indirect evidence for the participation of carrier-mediated membrane transport mechanisms, where several hydrophilic compounds seem to be absorbed efficiently via specific transporters [74]. Therefore, targeting specific membrane transporters is particularly important when prodrugs are polar or charged. Prodrugs can be designed to structurally resemble endogenous compounds and to be transported by specific carrier proteins. In this case, prodrugs may have the additional advantage of producing nontoxic by-products when prodrugs are converted to the parent drug molecules. The brain uptake of the potent glycine-NMDA receptor antagonists, such as 7-chlorokynurenic acid and 5,7-dichlorokynurenic acid, was significantly improved by their respective prodrugs, L-4-chlorokynurenine and L-4,6-dichlorokynurenine, which are amino acid derivatives [70]. L-4-chlorokynurenine was shown to be rapidly delivered into the brain by the large neutral amino acid transporter of the blood–brain barrier and to be converted intracellularly to its parent drug, 7-chlorokynurenic acid [70]. Furthermore, there have been some reports on prodrug design targeting peptide transporters, including peptidyl derivatives of methyldopa and alafosfalin and tripeptidyl prodrugs of foscarnet [75–77]. Developing prodrugs targeting specific membrane carriers requires considerable knowledge of the carrier proteins, including their distribution and substrate specificity.

## 4.2 Pegylation

Pegylation refers to the modification of a therapeutic agent by the attachment of poly(ethylene glycol) (PEG) molecules through covalent conjugation to the therapeutic agent. Pegylation may be an effective method of delivering therapeutic proteins and modifying their pharmacokinetic properties, in turn modifying pharmacodynamics, via a mechanism dependent on altered binding properties of the native protein [78]. PEG moieties are inert, long-chain amphiphilic molecules produced by linking repeating units of ethylene oxide [79]. A large number of potential PEG molecules are available, and they can be produced in different configurations, including linear or branched structures, and in different molecular weights (Fig. 7). Using pegylation to increase the size and molecular weight of a therapeutic protein

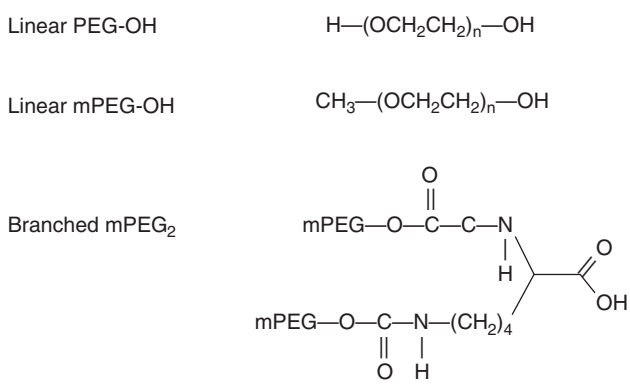


**FIGURE 7** Structural formulas of poly(ethylene glycol) (PEG) molecules. mPEG = monomethoxypoly(ethylene glycol). (From Ref. 80, with permission.)

alters the immunological, pharmacokinetic, and pharmacodynamic properties of the protein in ways that can extend its potential uses [80, 81]. Goals for chemically coupling PEG to peptide and protein drugs include decreased renal clearance and, for some products, more sustained absorption after subcutaneous administration as well as restricted distribution [82]. These pharmacokinetic changes may result in more constant and sustained plasma concentrations, which can lead to increases in clinical effectiveness when the desired effects are concentration dependent. Maintaining drug concentrations at or near a target concentration for an extended period of time is often clinically advantageous and is particularly useful in antiviral therapy, since constant antiviral pressure should prevent replication and may thereby suppress the emergence of resistant variants [83]. Additionally, PEG modification may decrease adverse effects caused by the large variations in peak-to-trough plasma drug concentrations associated with frequent administration and by the immunogenicity of unmodified proteins. Pegylated proteins may have reduced immunogenicity because PEG-induced steric hindrance can prevent immune recognition [84]. Furthermore, pegylation can enhance targeting of peptides and proteins to tumor tissue through enhanced permeation and retention (EPR) effect [78]. Tumor tissues have a very porous vasculature, and large molecular weight compounds (like pegylated drugs) have enhanced permeation into tumor tissue. Once permeated into the tumor, they are retained in the tumor for a prolonged period of time due to poor drainage from tumor. This EPR effect associated with pegylation has also been used to improve the therapeutic effectiveness of a number of drug molecules like doxorubicin [85, 86]. Pegylation has also been used to improve the circulation times of nanometer-size delivery systems such as liposomes and nanoparticles [87]. Table 6 lists the potential types of PEG conjugates.

**TABLE 6 Types of PEG Conjugates**

| Conjugate Type | Properties and Applications[a] |
|---|---|
| Small molecule drugs | Improved solubility, controlled permeability through biological barriers, longevity in bloodstream, controlled release |
| Affinity ligands and cofactors | Used in aqueous two-phase partitioning systems for purification and analysis of biological macromolecules and cells; enzymatic reactors |
| Peptides | Improved solubility, conformational analysis, biologically active conjugates |
| Proteins | Resistance to proteolysis, reduced immunogenicity and antigenicity, longevity in bloodstream, tolerance induction; uses include therapeutics, organic-soluble reagents, bioreactors |
| Saccharides | New biomaterials, drug carriers |
| Oligonucleotides | Improved solubility, resistance to nucleases, cell membrane permeability |
| Lipids | Used for preparation of PEG-grafted liposomes |
| Liposomes and particulates | Longevity in bloodstream, RES evasion |
| Biomaterials | Reduced thrombogenicity, reduced protein and cell adherence |

[a]PEG, poly(ethylene glycol); RES, reticuloendothelial system.

### 4.3 Use of Another Drug to Modify Distribution

A drug molecule that affects the pharmacokinetics of a second drug can be used clinically to modify the disposition, and therefore the pharmacodynamics, of the second drug. This phenomenon has been used previously to modify the absorption and elimination of a number of drugs. Classic examples include the use of probenecid to decrease the urinary excretion of penicillin [88] and the use of vasoconstrictors to decrease the rate of absorption of local anesthetics from subcutaneous injection site [89]. Use of hyperosmotic mannitol to transiently open up the tight junctions of the blood–brain barrier represents one of the earliest attempts to use one agent to modify the distribution of another drug [90]. With increasing knowledge of the effects of transporters on drug distribution and pharmacokinetics, a number of attempts have been made to use the inhibitors of these transporters to improve the distribution profile of drugs. Because the use of inhibitors of P-gp have been the most studied, this is discussed further.

As discussed earlier, P-gp is overexpressed in many normal tissues such as the capillary endothelium of the blood–brain barrier (BBB), the intestinal epithelium, and also in tumor tissues. Overexpression of P-gp in the blood–brain barrier results in the reduced transport of P-gp substrates into the brain. The significance of this problem is highlighted by the estimations that up to 98% of the newly developed small molecules will not cross the BBB [91]. This limits the number of drugs available for treating diseases like epilepsy [92], stroke and brain injury [93], brain cancer [94], HIV infection of the brain [95], and amyotrophic lateral sclerosis [96]. Numerous preclinical studies and a few clinical studies have indicated the potential of P-gp inhibitors to enhance brain delivery of P-gp substrates [91, 94, 97–99]. P-gp inhibition was evaluated as a means of increasing the brain delivery of the peripherally acting opioid loperamide [100]. Healthy volunteers received a single oral dose of loperamide with or without the P-gp inhibitor quinidine. The central effect of loperamide (change in ventilatory response in response to carbon dioxide) was only observed in volunteers receiving both loperamide and quinidine, indicating the ability of the P-gp inhibitor to enhance the brain delivery of P-gp substrates. Other case studies have demonstrated the use of the P-gp inhibitor verapamil to enhance the brain delivery of anticonvulsant drugs in refractory epilepsy [101, 102].

Many important anticancer agents like doxorubicin and paclitaxel are substrates of P-gp. Overexpression of P-gp in tumor cells results in the development of drug resistance. Initial clinical trials with P-gp inhibitors to treat resistant tumors were performed with "first generation" P-gp inhibitors such as cyclosporin, which were already in use for other indications [103]. Absence of confirmation of P-gp expression in the tumors and P-gp inhibitor toxicity at doses administered to achieve serum concentrations comparable to those that were effective in animal models resulted in the failure of these drugs in clinical trials [104]. Second generation inhibitors (e.g., PSC 833) were developed solely for the purpose of overcoming drug resistance [105]. These agents were tested in clinical trials in various malignancies for which there was evidence that P-gp is expressed or associated with a poorer therapeutic outcome [106]. One major limitation of these trials, however, was the reduction in anticancer drug doses that was required with concurrent administration of inhibitor [107]. P-gp inhibitors increased the serum levels of the coadministered

chemotherapeutic agent. A number of studies found that reduction in the dose led to a number of patients being undertreated, which could have contributed to the failure of these combination treatments [107]. Pharmacokinetic interactions between the P-gp inhibitor and the drug could also result from the inhibitor's ability to inhibit other proteins involved in drug metabolism such as cytochrome P450 [108] P-gp inhibitors with fewer pharmacokinetic interactions are being developed [109], and functional assays to verify the role of P-gp in drug resistance, such as sestamibi imaging, are proving helpful in assessing the development of the newer inhibitors [110]. A number of other approaches to overcome P-gp that are currently in development include the use of monoclonal antibodies against P-gp [111, 112], antisense oligonucleotides [113], pH-sensitive polymeric micelles [114], Pluronic(r) copolymers [115], peptide-based transmembrane inhibitors [116], and inhibitors of signal transduction [117].

### 4.4 Encapsulation in Delivery Systems

A delivery system is often used to encapsulate a drug, because of the following advantages: (1) prolonged drug availability in the body because of sustained release of the drug from the delivery system; (2) enhanced availability of the drug at the target site—by choosing appropriate formulation parameters, a favorable tissue distribution profile could be obtained; and (3) enhanced stability of the drug due to protection from drug-metabolizing enzymes. Sustained (or continuous) release of a drug from a delivery system involves slow diffusion of the drug out of a polymeric matrix and/or slow degradation of the polymer over time. Pulsatile release is sometimes the preferred method of drug delivery, as it closely mimics the way in which the body naturally produces hormones such as insulin. It is achieved by using drug-carrying polymers that respond to specific stimuli (e.g., exposure to light, changes in pH or temperature) [118].

Choice of the delivery system often depends on the route of administration, nature of the drug, and nature of the disease. Applications involving intravenous administration require the use of nanometer-size delivery systems. This is necessary to reach distant target sites perfused by fine capillaries (diameter $\approx 1\,\mu m$) and to prevent embolism. Colloidal drug carrier systems such as micelles, vesicle and liquid crystal dispersions, as well as nanoparticle dispersions consisting of small particles of 10–400 nm diameter have been found useful as intravenous drug delivery systems. Figure 8 demonstrates different types of drug carriers. When developing these formulations, the goal is to obtain systems with optimized drug loading and release properties, long shelf-life, and low toxicity [119].

Micelles formed by self-assembly of amphiphilic block copolymers (5–50 nm) in aqueous solutions are of great interest for drug delivery applications [120]. Drugs can be physically entrapped in the core of block copolymer micelles and transported at concentrations that can exceed their intrinsic water solubility. Moreover, the hydrophilic blocks can form hydrogen bonds with the aqueous surroundings and form a tight shell around the micellar core. As a result, the contents of the hydrophobic core are effectively protected against hydrolysis and enzymatic degradation. In addition, the corona may prevent recognition by the reticuloendothelial system and therefore preliminary elimination of the micelles from the bloodstream.

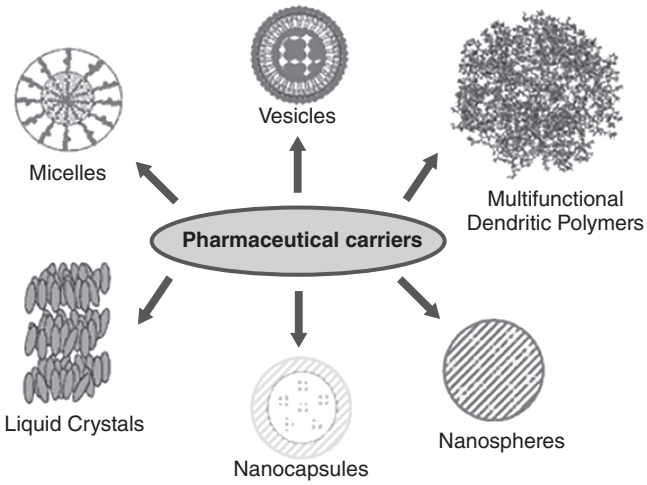


**FIGURE 8** Schematic of various pharmaceutical carriers. (From Ref. 120, with permission.)

Functionalization of block copolymers with crosslinkable groups can increase the stability of the corresponding micelles and improve their temporal control. Substitution of block copolymer micelles with specific ligands is a very promising strategy to achieve targeted drug delivery.

Liposomes are a form of vesicles that consist either of many, a few, or just one phospholipid bilayer [121]. The polar character of the liposomal core enables polar drug molecules to be encapsulated. Amphiphilic and lipophilic molecules are solubilized within the phospholipid bilayer according to their affinity toward the phospholipids. Channel proteins can be incorporated without loss of their activity within the hydrophobic domain of vesicle membranes, acting as a size-selective filter, only allowing passive diffusion of small solutes such as ions, nutrients, and antibiotics. Thus, drugs that are encapsulated in a nanocage functionalized with channel proteins are effectively protected from premature degradation by proteolytic enzymes. The drug molecule, however, is able to diffuse through the channel, driven by the concentration difference between the interior and the exterior of the nanocage.

Dendrimers are nanometer-sized, highly branched, and monodisperse macromolecules with symmetrical architecture [122]. They consist of a central core, branching units, and terminal functional groups. The core, together with the internal units, determines the environment of the nanocavities and consequently their solubilizing properties, whereas the external groups determine the solubility and chemical behavior of these polymers. Targeting effectiveness is affected by attaching targeting ligands at the external surface of dendrimers, while their stability and protection from the phagocytosis is achieved by functionalization of the dendrimers with PEG chains.

Nanocapsules are vesicular systems in which the drug is confined to a cavity surrounded by a polymer membrane, while nanospheres are matrix systems in which the drug is physically and uniformly dispersed [123]. Biodegradable polymeric nanoparticles have many applications in the controlled release of drugs, in targeting

particular organs/tissues, as carriers of DNA in gene therapy, and in their ability to deliver proteins, peptides, and genes to the target tissue. Like for other nanocarriers, pharmacokinetics and the biodistribution profile of nanoparticles may be altered by varying the surface properties (e.g., addition of PEG chains) and attaching specific tissue-targeting ligands on the surface.

## 5 CONCLUSION

Distribution plays an important role in determining the magnitude and duration of a drug's therapeutic effect. Drug distribution is influenced by physical properties of the drug as well as physiologic factors such as blood perfusion, protein binding, and transporter activity. A number of physical and chemical methods are available to alter the distribution of the drug. These allow the therapeutic use of drug molecules that would otherwise have unfavorable distribution in the body.

## REFERENCES

1. Tozer TN. *Clinical Pharmacokinetics Concepts and Applications*, 3rd ed. Baltimore: Lippincott Williams & Wilkins; 1995, pp 137–140.
2. Shargel L. *Applied Biopharmaceutics and Pharmacokinetics*, 4th ed. New York: McGraw-Hill; 1999, pp 281–291.
3. Dvorchik BH, Vesell ES. Pharmacokinetic interpretation of data gathered during therapeutic drug monitoring. *Clin Chem* 1976;22:868–878.
4. MacKichan JJ. Pharmacokinetic consequences of drug displacement from blood and tissue proteins. *Clin Pharmacokinet* 1984;9(Suppl 1):32–41.
5. Lin JH, Lu AY. Role of pharmacokinetics and metabolism in drug discovery and development. *Pharmacol Rev* 1997;49:403–449.
6. Aarons LJ, Rowland M. Kinetics of drug displacement interactions. *J Pharmacokinet Biopharm* 1981;9:181–190.
7. Vesell ES. Clinical pharmacology: a personal perspective. *Clin Pharmacol Ther* 1985;38: 603–612.
8. Aarons L, Salisbury R, Alam-Siddiqi M, Taylor L, Grennan DM. Plasma and synovial fluid kinetics of flurbiprofen in rheumatoid arthritis. *Br J Clin Pharmacol* 1986;21: 155–163.
9. Schuhmann G, Fichtl B, Kurz H. Prediction of drug distribution *in vivo* on the basis of *in vitro* binding data. *Biopharm Drug Dispos* 1987;8:73–86.
10. Lloyd JB. Lysosome membrane permeability: implications for drug delivery. *Adv Drug Deliv Rev* 2000;41:189–200.
11. Rang HP, Ritter JM, Moore PK, Lamb P. *Pharmacology*, 4th ed. Edinburgh: Churchill Livingstone; 2001, pp 91–105.
12. Hogben CA, Tocco DJ, Brodie BB, Schanker LS. On the mechanism of intestinal absorption of drugs. *J Pharmacol Exp Ther* 1959;125:275–282.
13. Shore ML. Biological applications of kinetic analysis of a two-compartment open system. *J Appl Physiol* 1961;16:771–782.
14. Schanker LS. Passage of drugs across body membranes. *Pharmacol Rev* 1962;14: 501–530.

15. Olanoff LS, Anderson JM. Controlled release of tetracycline–III: A physiological pharmacokinetic model of the pregnant rat. *J Pharmacokinet Biopharm* 1980;8:599–620.
16. Ingbar SH, Freinkel N. The influence of ACTH, cortisone, and hydrocortisone on the distribution and peripheral metabolism of thyroxine. *J Clin Invest* 1955;34:1375–1379.
17. Artursson P, Karlsson J. Correlation between oral drug absorption in humans and apparent drug permeability coefficients in human intestinal epithelial (Caco-2) cells. *Biochem Biophys Res Commun* 1991;175:880–885.
18. Taylor DC, Pownall R, Burke W. The absorption of beta-adrenoceptor antagonists in rat *in-situ* small intestine; the effect of lipophilicity. *J Pharm Pharmacol* 1985;37:280–283.
19. Simons K, Vaz WL. Model systems, lipid rafts, and cell membranes. *Annu Rev Biophys Biomol Struct* 2004;33:269–295.
20. Crone C, Christensen O. Electrical resistance of a capillary endothelium. *J Gen Physiol* 1981;77:349–371.
21. Butt AM, Jones HC, Abbott NJ. Electrical resistance across the blood–brain barrier in anaesthetized rats: a developmental study. *J Physiol* 1990;429:47–62.
22. Tamai I, Tsuji A. Transporter-mediated permeation of drugs across the blood–brain barrier. *J Pharm Sci* 2000;89:1371–1388.
23. Tsuji A. Small molecular drug transfer across the blood–brain barrier via carrier-mediated transport systems. *NeuroRx* 2005;2:54–62.
24. Ecker GF, Noe CR. *In silico* prediction models for blood–brain barrier permeation. *Curr Med Chem* 2004;11:1617–1628.
25. Suzuki H, Sugiyama Y. [Kinetic analysis of the disposition of hydrophilic drugs in the central nervous system (CNS): prediction of the CNS disposition from the transport properties in the blood–brain and blood–cerebrospinal fluid barriers]. *Yakugaku Zasshi* 1994;114:950–971.
26. Upton RN, Doolette DJ. Kinetic aspects of drug disposition in the lungs. *Clin Exp Pharmacol Physiol* 1999;26:381–391.
27. Avdeef A. Physicochemical profiling (solubility, permeability and charge state). *Curr Top Med Chem* 2001;1:277–351.
28. Obata K, Sugano K, Saitoh R, Higashida A, Nabuchi Y, Machida M, Aso Y. Prediction of oral drug absorption in humans by theoretical passive absorption model. *Int J Pharm* 2005;293:183–192.
29. Sue YJ, Shannon M. Pharmacokinetics of drugs in overdose. *Clin Pharmacokinet* 1992;23:93–105.
30. Levy G. Effects of plasma protein binding of drugs on duration and intensity of pharmacological activity. *J Pharm Sci* 1976;65:1264–1265.
31. Tocco DJ, deLuna FA, Duncan AE, Hsieh JH, Lin JH. Interspecies differences in stereoselective protein binding and clearance of MK-571. *Drug Metab Dispos* 1990;18:388–392.
32. Lin JH, deLuna FA, Ulm EH, Tocco DJ. Species-dependent enantioselective plasma protein binding of MK-571, a potent leukotriene D4 antagonist. *Drug Metab Dispos* 1990;18:484–487.
33. Eap CB, Cuendet C, Baumann P. Binding of *d*-methadone, *l*-methadone, and *dl*-methadone to proteins in plasma of healthy volunteers: role of the variants of alpha 1-acid glycoprotein. *Clin Pharmacol Ther* 1990;47:338–346.
34. Ruiz-Cabello F, Erill S. Abnormal serum protein binding of acidic drugs in diabetes mellitus. *Clin Pharmacol Ther* 1984;36:691–695.

35. Ryan DP, Matthews JM. Protein-protein interactions in human disease. *Curr Opin Struct Biol* 2005;15:441–446.
36. Kim RB. Transporters and drug discovery: why, when, and how. *Mol Pharm* 2006;3:26–32.
37. Zhang L, Strong JM, Qiu W, Lesko LJ, Huang SM. Scientific perspectives on drug transporters and their role in drug interactionst. *Mol Pharm* 2006;3:62–69.
38. Loo TW, Clarke DM. Recent progress in understanding the mechanism of P-glycoprotein-mediated drug efflux. *J Membr Biol* 2005;206:173–185.
39. Choo EF, Leake B, Wandel C, Imamura H, Wood AJ, Wilkinson GR, Kim RB. Pharmacological inhibition of P-glycoprotein transport enhances the distribution of HIV-1 protease inhibitors into brain and testes. *Drug Metab Dispos* 2000;28:655–660.
40. Gottesman MM, Ling V. The molecular basis of multidrug resistance in cancer: the early years of P-glycoprotein research. *FEBS Lett* 2006;580:998–1009.
41. Calcagno AM, Ludwig JA, Fostel JM, Gottesman MM, Ambudkar SV. Comparison of drug transporter levels in normal colon, colon cancer, and Caco-2 cells: impact on drug disposition and discovery. *Mol Pharm* 2006;3:87–93.
42. Schinkel AH, Jonker JW. Mammalian drug efflux transporters of the ATP binding cassette (ABC) family: an overview. *Adv Drug Deliv Rev* 2003;55:3–29.
43. Dean M, Hamon Y, Chimini G. The human ATP-binding cassette (ABC) transporter superfamily. *J Lipid Res* 2001;42:1007–1017.
44. Sugiyama Y, Ito K. Future prospects for toxicokinetics: prediction of drug disposition and adverse effects in humans from *in vitro* measurements of drug metabolism, transport and binding. *J Toxicol Sci* 1998;23(Suppl 4):647–652.
45. Greiner B, Eichelbaum M, Fritz P, Kreichgauer HP, von Richter O, Zundler J, Kroemer HK. The role of intestinal P-glycoprotein in the interaction of digoxin and rifampin. *J Clin Invest* 1999;104:147–153.
46. Edwards DJ, Fitzsimmons ME, Schuetz EG, Yasuda K, Ducharme MP, Warbasse LH, Woster PM, Schuetz JD, Watkins P. 6 ,7 -Dihydroxybergamottin in grapefruit juice and Seville orange juice: effects on cyclosporine disposition, enterocyte CYP3A4, and P-glycoprotein. *Clin Pharmacol Ther* 1999;65:237–244.
47. Kerb R, Hoffmeyer S, Brinkmann U. ABC drug transporters: hereditary polymorphisms and pharmacological impact in MDR1, MRP1 and MRP2. *Pharmacogenomics* 2001;2:51–64.
48. Inui KI, Masuda S, Saito H, Cellular and molecular aspects of drug transport in the kidney. *Kidney Int* 2000;58:944–958.
49. You G. Towards an understanding of organic anion transporters: structure–function relationships. *Med Res Rev* 2004;24:762–774.
50. Masuda S, Inui K. [Molecular mechanisms on drug transporters in the drug absorption and disposition]. *Nippon Rinsho* 2002;60:65–73.
51. Sekine T, Cha SH, Endou H. The multispecific organic anion transporter (OAT) family. *Pflugers Arch* 2000;440:337–350.
52. Thomas MC, Tikellis C, Kantharidis P, Burns WC, Cooper ME, Forbes JM. The role of advanced glycation in reduced organic cation transport associated with experimental diabetes. *J Pharmacol Exp Ther* 2004;311:456–466.
53. Meyer UA. The molecular basis of genetic polymorphisms of drug metabolism. *J Pharm Pharmacol* 1994;46(Suppl 1):409–415.
54. Schinkel AH. Pharmacological insights from P-glycoprotein knockout mice. *Int J Clin Pharmacol Ther* 1998;36:9–13.

55. Schwab M, Eichelbaum M, Fromm MF. Genetic polymorphisms of the human MDR1 drug transporter. *Annu Rev Pharmacol Toxicol* 2003;43:285–307.
56. Sakaeda T, Nakamura T, Okumura K. Pharmacogenetics of MDR1 and its impact on the pharmacokinetics and pharmacodynamics of drugs. *Pharmacogenomics* 2003;4:397–410.
57. Kurata Y, Ieiri I, Kimura M, Morita T, Irie S, Urae A, Ohdo S, Ohtani H, Sawada Y, Higuchi S, Otsubo K. Role of human MDR1 gene polymorphism in bioavailability and interaction of digoxin, a substrate of P-glycoprotein. *Clin Pharmacol Ther* 2002;72:209–219.
58. Gerloff T, Schaefer M, Johne A, Oselin K, Meisel C, Cascorbi I, Roots I. MDR1 genotypes do not influence the absorption of a single oral dose of 1 mg digoxin in healthy white males. *Br J Clin Pharmacol* 2002;54:610–616.
59. Verstuyft C, Schwab M, Schaeffeler E, Kerb R, Brinkmann U, Jaillon P, Funck-Brentano C, Becquemont L. Digoxin pharmacokinetics and MDR1 genetic polymorphisms. *Eur J Clin Pharmacol* 2003;58:809–812.
60. Haufroid V, Mourad M, Van Kerckhove V, Wawrzyniak J, De Meyer M, Eddour DC, Malaise J, Lison D, Squifflet JP, Wallemacq P. The effect of CYP3A5 and MDR1 (ABCB1) polymorphisms on cyclosporine and tacrolimus dose requirements and trough blood levels in stable renal transplant patients. *Pharmacogenetics* 2004;14:147–154.
61. Krijt J, Vackova M, Kozich V. Measurement of homocysteine and other aminothiols in plasma: advantages of using tris(2-carboxyethyl)phosphine as reductant compared with tri-*N*-butylphosphine. *Clin Chem* 2001;47:1821–1828.
62. Tirona RG, Kim RB. Pharmacogenomics of organic anion-transporting polypeptides (OATP). *Adv Drug Deliv Rev* 2002;54:1343–1352.
63. Ieiri I, Suzuki H, Kimura M, Takane H, Nishizato Y, Irie S, Urae A, Kawabata K, Higuchi S, Otsubo K, Sugiyama Y. Influence of common variants in the pharmacokinetic genes (OATP-C, UGT1A1, and MRP2) on serum bilirubin levels in healthy subjects. *Hepatol Res* 2004;30:91–95.
64. Mwinyi J, Johne A, Bauer S, Roots I, Gerloff T. Evidence for inverse effects of OATP-C (SLC21A6) 5 and 1b haplotypes on pravastatin kinetics. *Clin Pharmacol Ther* 2004;75:415–421.
65. Niemi M, Schaeffeler E, Lang T, Fromm MF, Neuvonen M, Kyrklund C, Backman JT, Kerb R, Schwab M, Neuvonen PJ, Eichelbaum M, Kivisto KT. High plasma pravastatin concentrations are associated with single nucleotide polymorphisms and haplotypes of organic anion transporting polypeptide-C (OATP-C, SLCO1B1). *Pharmacogenetics* 2004;14:429–440.
66. Evans WE, Johnson JA. Pharmacogenomics: the inherited basis for interindividual differences in drug response. *Annu Rev Genomics Hum Genet* 2001;2:9–39.
67. Albert A. Chemical aspects of selective toxicity. *Nature* 1958;182:421–422.
68. Stella VJ, Charman WN, Naringrekar VH. Prodrugs. Do they have advantages in clinical practice? *Drugs* 1985;29:455–473.
69. Banerjee PK, Amidon GL. Physicochemical property modification strategies based on enzyme substrate specificities I: rationale, synthesis, and pharmaceutical properties of aspirin derivatives. *J Pharm Sci* 1981;70:1299–1303.
70. Han HK, Amidon GL. Targeted prodrug design to optimize drug delivery. *AAPS Pharm Sci* 2000;2:E6.
71. Cobb LM, Hacker T, Nolan J. NAD(P)H nitroblue tetrazolium reductase levels in apparently normoxic tissues: a histochemical study correlating enzyme activity with binding of radiolabelled misonidazole. *Br J Cancer* 1990;61:524–529.

72. Knox RJ, Connors TA. Prodrugs in cancer chemotherapy. *Pathol Oncol Res* 1997;3: 309–324.
73. Bagshawe KD. Antibody-directed enzyme prodrug therapy for cancer: its theoretical basis and application. *Mol Med Today* 1995;1:424–431.
74. Mizuma T, Ohta K, Hayashi M, Awazu S. Comparative study of active absorption by the intestine and disposition of anomers of sugar-conjugated compounds. *Biochem Pharmacol* 1993;45:1520–1523.
75. Swaan PW, Tukker JJ. Carrier-mediated transport mechanism of foscarnet (trisodium phosphonoformate hexahydrate) in rat intestinal tissue. *J Pharmacol Exp Ther* 1995;272:242–247.
76. Grappel SF, Giovenella AJ, Nisbet LJ. Activity of a peptidyl prodrug, alafosfalin, against anaerobic bacteria. *Antimicrob Agents Chemother* 1985;27:961–963.
77. Oh DM, Han HK, Amidon GL. Drug transport and targeting. Intestinal transport. *Pharm Biotechnol* 1999;12:59–88.
78. Caliceti P, Veronese FM. Pharmacokinetic and biodistribution properties of poly(ethylene glycol)–protein conjugates. *Adv Drug Deliv Rev* 2003;55:1261–1277.
79. Veronese FM, Pasut G. PEGylation, successful approach to drug delivery. *Drug Discov Today* 2005;10:1451–1458.
80. Harris JM, Martin NE, Modi M. Pegylation: a novel process for modifying pharmacokinetics. *Clin Pharmacokinet* 2001;40:539–551.
81. Francis GE, Delgado C, Fisher D, Malik F, Agrawal AK. Polyethylene glycol modification: relevance of improved methodology to tumour targeting. *J Drug Target* 1996;3: 321–340.
82. Mukai Y, Yoshioka Y, Tsutsumi Y, Phage display and PEGylation of therapeutic proteins. *Comb Chem High Throughput Screen* 2005;8:145–152.
83. Florence AT, Jani PU. Novel oral drug formulations. Their potential in modulating adverse effects. *Drug Saf* 1994;10:233–266.
84. Delgado C, Francis GE, Fisher D. The uses and properties of PEG-linked proteins. *Crit Rev Ther Drug Carrier Syst* 1992;9:249–304.
85. Crawford J. Clinical uses of pegylated pharmaceuticals in oncology. *Cancer Treat Rev* 2002;28(Suppl A):7–11.
86. Crawford J. Clinical benefits of pegylated proteins in oncology. *Cancer Treat Rev* 2002;28(Suppl A):1–2.
87. Bhadra D, Bhadra S, Jain P, Jain NK. Pegnology: a review of PEG-ylated systems. *Pharmazie* 2002;57:5–29.
88. Bergholz H, Erttmann RR, Damm KH. Effects of probenecid on plasma/tissue distribution of $^{14}$C-benzylpenicillin in rats. *Experientia* 1980;36:333–334.
89. Adriani J, Naraghi M. The pharmacologic principles of regional pain relief. *Annu Rev Pharmacol Toxicol* 1977;17:223–242.
90. Black KL, Ningaraj NS. Modulation of brain tumor capillaries for enhanced drug delivery selectively to brain tumor. *Cancer Control* 2004;11:165–173.
91. Pardridge WM. Blood–brain barrier drug targeting: the future of brain drug development. *Mol Interv* 2003;3:90–105, 51.
92. Volk HA, Loscher W. Multidrug resistance in epilepsy: rats with drug-resistant seizures exhibit enhanced brain expression of P-glycoprotein compared with rats with drug-responsive seizures. *Brain* 2005;128:1358–1368.
93. Liu XD, Pan GY, Xie L, Hou YY, Lan W, Su Q, Liu GQ. Cyclosporin A enhanced protection of nimodipine against brain damage induced by hypoxia-ischemia in mice and rats. *Acta Pharmacol Sin* 2002;23:225–229.

94. Fellner S, Bauer B, Miller DS, Schaffrik M, Fankhanel M, Spruss T, Bernhardt G, Graeff C, Farber L, Gschaidmeier H, Buschauer A, Fricker G. Transport of paclitaxel (Taxol) across the blood–brain barrier *in vitro* and *in vivo*. *J Clin Invest* 2002;110:1309–1318.
95. Kim RB, Fromm MF, Wandel C, Leake B, Wood AJ, Roden DM, Wilkinson GR. The drug transporter P-glycoprotein limits oral absorption and brain entry of HIV-1 protease inhibitors. *J Clin Invest* 1998;101:289–294.
96. Kirkinezos IG, Hernandez D, Bradley WG, Moraes CT. An ALS mouse model with a permeable blood–brain barrier benefits from systemic cyclosporine A treatment. *J Neurochem* 2004;88:821–826.
97. Bauer B, Hartz AM, Fricker G, Miller DS. Modulation of P-glycoprotein transport function at the blood–brain barrier. *Exp Biol Med (Maywood)* 2005;230:118–127.
98. Boyle FM, Eller SL, Grossman SA. Penetration of intra-arterially administered vincristine in experimental brain tumor. *Neuro-oncol* 2004;6:300–305.
99. Kemper EM, Boogerd W, Thuis I, Beijnen JH, van Tellingen O. Modulation of the blood–brain barrier in oncology: therapeutic opportunities for the treatment of brain tumours? *Cancer Treat Rev* 2004;30:415–423.
100. Sadeque AJ, Wandel C, He H, Shah S, Wood AJ. Increased drug delivery to the brain by P-glycoprotein inhibition. *Clin Pharmacol Ther* 2000;68:231–237.
101. Iannetti P, Spalice A, Parisi P. Calcium-channel blocker verapamil administration in prolonged and refractory status epilepticus. *Epilepsia* 2005;46:967–969.
102. Summers MA, Moore JL, McAuley JW. Use of verapamil as a potential P-glycoprotein inhibitor in a patient with refractory epilepsy. *Ann Pharmacother* 2004;38:1631–1634.
103. Ferry DR, Traunecker H, Kerr DJ. Clinical trials of P-glycoprotein reversal in solid tumours. *Eur J Cancer* 1996;32A:1070–1081.
104. Shabbits JA, Krishna R, Mayer LD. Molecular and pharmacological strategies to overcome multidrug resistance. *Expert Rev Anticancer Ther* 2001;1:585–594.
105. Twentyman PR, Bleehen NM. Resistance modification by PSC-833, a novel non-immunosuppressive cyclosporin [corrected]. *Eur J Cancer* 1991;27:1639–1642.
106. Oza AM. Clinical development of P glycoprotein modulators in oncology. *Novartis Found Symp* 2002;243:103–115; discussion 115–118, 180–185.
107. Fracasso PM, Brady MF, Moore DH, Walker JL, Rose PG, Letvak L, Grogan TM, McGuire WP. Phase II study of paclitaxel and valspodar (PSC 833) in refractory ovarian carcinoma: a gynecologic oncology group study. *J Clin Oncol* 2001;19:2975–2982.
108. Ma MK, McLeod HL, Westervelt P, Fracasso PM. Pharmacokinetic study of infusional valspodar. *J Clin Pharmacol* 2002;42:412–418.
109. Agrawal M, Abraham J, Balis FM, Edgerly M, Stein WD, Bates S, Fojo T, Chen CC. Increased 99mTc-sestamibi accumulation in normal liver and drug-resistant tumors after the administration of the glycoprotein inhibitor, XR9576. *Clin Cancer Res* 2003;9:650–656.
110. Bigott HM, Prior JL, Piwnica-Worms DR, Welch MJ. Imaging multidrug resistance P-glycoprotein transport function using microPET with technetium-94m-sestamibi. *Mol Imaging* 2005;4:30–39.
111. Haus-Cohen M, Assaraf YG, Binyamin L, Benhar I, Reiter Y. Disruption of P-glycoprotein anticancer drug efflux activity by a small recombinant single-chain Fv antibody fragment targeted to an extracellular epitope. *Int J Cancer* 2004;109:750–758.
112. Mechetner EB, Roninson IB. Efficient inhibition of P-glycoprotein-mediated multidrug resistance with a monoclonal antibody. *Proc Natl Acad Sci USA* 1992;89:5824–5828.

113. Kang H, Fisher MH, Xu D, Miyamoto YJ, Marchand A, Van Aerschot A, Herdewijn P, Juliano RL. Inhibition of MDR1 gene expression by chimeric HNA antisense oligonucleotides. *Nucleic Acids Res* 2004;32:4411–4419.
114. Lee ES, Na K, Bae YH. Doxorubicin loaded pH-sensitive polymeric micelles for reversal of resistant MCF-7 tumor. *J Control Release* 2005;103:405–418.
115. Batrakova EV, Li S, Elmquist WF, Miller DW, Alakhov VY, Kabanov AV. Mechanism of sensitization of MDR cancer cells by Pluronic block copolymers: selective energy depletion. *Br J Cancer* 2001;85:1987–1997.
116. Tarasova NI, Seth R, Tarasov SG, Kosakowska-Cholody T, Hrycyna CA, Gottesman MM, Michejda CJ. Transmembrane inhibitors of P-glycoprotein, an ABC transporter. *J Med Chem* 2005;48:3768–3775.
117. Xu D, Ye D, Fisher M, Juliano RL. Selective inhibition of P-glycoprotein expression in multidrug-resistant tumor cells by a designed transcriptional regulator. *J Pharmacol Exp Ther* 2002;302:963–971.
118. Langer R. Drug delivery and targeting. *Nature* 1998;392:5–10.
119. Kayser O, Lemke A, Hernandez-Trejo N. The impact of nanobiotechnology on the development of new drug delivery systems. *Curr Pharm Biotechnol* 2005;6:3–5.
120. Muller Goymann CC. Physicochemical characterization of colloidal drug delivery systems such as reverse micelles, vesicles, liquid crystals and nanoparticles for topical administration. *Eur J Pharm Biopharm* 2004;58:343–356.
121. Hart SL. Lipid carriers for gene therapy. *Curr Drug Deliv* 2005;2:423–428.
122. Iwamura M. [Dendritic systems for drug delivery applications]. *Nippon Rinsho* 2006;64:231–237.
123. Mayer C. Nanocapsules as drug delivery systems. *Int J Artif Organs* 2005;28:1163–1171.

# THE BLOOD–BRAIN BARRIER AND ITS EFFECT ON ABSORPTION AND DISTRIBUTION

A. G. de Boer[1] and P. J. Gaillard[2]
[1]*University of Leiden, Leiden, The Netherlands*
[2]*to-BBB Technologies BV, Leiden, The Netherlands*

**Contents**

1 Introduction
2 Barriers in the Brain
  2.1 Blood–Brain Barrier (BBB)
  2.2 Blood–Cerebrospinal Fluid Barrier (BCSFB) and Ependyma
3 Methods to Measure Drug Transport to the Brain
  3.1 *In Vitro* Methods
  3.2 *In Vivo* Methods
4 Strategies for Drug Delivery to the Brain
  4.1 Local Brain Delivery by Direct Injection/Infusion of Drugs into the Brain
  4.2 Global Brain Delivery by Vascular Nonviral Drug Administration
5 Conclusion
  References

## 1 INTRODUCTION

The central nervous system (CNS) is a sanctuary site and is protected by various barriers. These regulate brain homeostasis and the transport of endogenous and exogenous compounds by controlling their selective and specific uptake, efflux, and

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

metabolism in the brain. Unfortunately, potential drugs for the treatment of most brain diseases are therefore often not able to cross these barriers. As a result, various drug delivery and targeting strategies are currently being developed to enhance the absorption and distribution of drugs into the brain. This chapter discusses the biology and physiology of the blood–brain barrier (BBB) and the blood–cerebrospinal fluid barrier with respect to drug transport (absorption, distribution), the *in vitro* and *in vivo* methods to measure BBB transport, and the possibilities to deliver large molecular drugs, by viral and receptor-mediated nonviral drug delivery, to the (human) brain.

## 2 BARRIERS IN THE BRAIN

There are three barriers that limit drug transport to the brain parenchyma. These are the blood–brain barrier (BBB) localized in the capillaries in the brain, the blood–cerebrospinal fluid barrier (BCSFB) that is presented by the choroid plexus epithelium in the ventricles, and the ependyma that is an epithelial layer of cells covering the brain tissue in the ventricles and limits the transport of compounds from the CSF to the brain tissue.

### 2.1 Blood–Brain Barrier (BBB)

In 1885 Ehrlich [1] was the first to show evidence for the existence of a barrier between blood and brain. He injected vital dyes intravenously and found that, in contrast to other tissues, it did not stain the brain [1]. His successor, Goldman [2], injected these dyes into the cerebral spinal fluid, after which staining of the brain was observed, but not of the peripheral organs. Since this seminal discovery, much research has been performed on the (patho) physiology and pharmacology of the BBB (see also other reviews [3–9]).

The current knowledge is that the BBB is situated at the interface of blood and brain and its primary function is to maintain the homeostasis of the brain. Furthermore, the BBB is not uniform throughout the brain, since the capillaries in the circumventricular organs (CVOs) are fenestrated [10, 11]. Figure 1 gives a schematic representation of the barriers present in the CNS.

The human BBB has a total blood vessel length of approximately 600 km. In fact, every cubic centimeter of the cortex comprises the amazing sum of 1 km of blood vessels. It has an estimated surface area of approximately 20 $m^2$, which is claimed to be similar to the BCSFB [12, 13]. However, the surface area of the BCSFB is facing the CSF and not the blood, which makes the BBB, based on total blood flow and its wide vascular bed, functionally the most important global influx barrier for solutes to reach the brain [10, 11].

The BBB is mainly formed by brain capillary endothelial cells (BCECs) [14], although other cell types such as pericytes, astrocytes, and neuronal cells also play an important role in the function of the BBB [15–19]. BCECs are different from peripheral endothelial cells, as can be seen schematically in Fig. 2 (in which the specific surroundings of the brain capillaries are shown). BCECs have specific characteristics, such as tight junctions, which prevent paracellular transport of small and large (water-soluble) compounds from blood to the brain [14, 20, 21]. Furthermore,

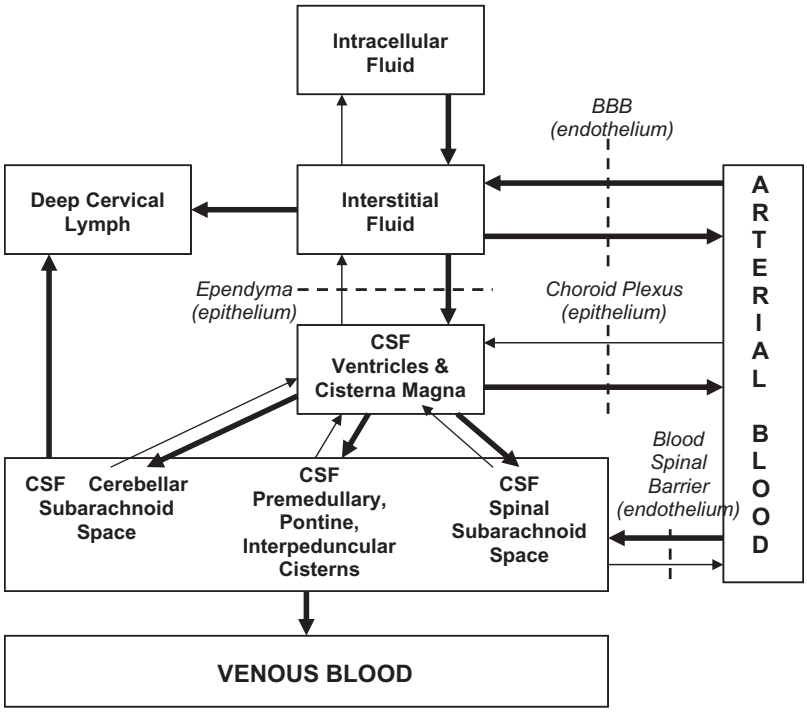


**FIGURE 1** The various barriers that one can find in the brain (broken lines) representing the blood–brain barrier, the blood–cerebrospinal fluid (CSF) barrier, the brain–CSF barrier, and the blood–spinal barrier. The blood–brain barrier has the largest surface area and is therefore considered to be the most important influx barrier for solutes to reach the brain. In addition, paths are shown for fluid movement (solid arrows) between cerebral intracellular fluid (ICF), intrastitial fluid (ISF), cerebrospinal fluid (CSF), blood, and lymphatics. Thick arrows represent major paths of fluid movement under normal conditions. Thin arrows represent minor paths of fluid movement under normal conditions. (Adapted from Ref. 45.)

transcellular transport from blood to brain is limited as a result of low vesicular transport, high metabolic activity, and a lack of fenestrae [15]. It functions as a physical, a metabolic [22–24], and an immunological barrier [25]. In addition, receptors at the BBB may play a role in brain signaling (e.g., insulin receptor [26], LRP1,2 receptor [27]). These specific characteristics of the BBB are induced and maintained by the (endfeet of) astrocytes, surrounding the BCECs [15, 28], as well as by neuronal endings, which can directly innervate the BCECs [14, 29]. Pericytes also play a role at the BBB, as they share the continuous capillary basement membrane with the BCECs [30]. Their phagocytotic activity forms an additional BBB property [10]. Furthermore, pericytes regulate endothelial homeostasis by inducing the endothelial release of plasminogen-activator inhibitor-1 (PAI-1), thereby negatively regulating brain endothelial fibrinolysis [31].

Until now, many transport systems have been discovered that play an important role in maintaining BBB integrity and brain homeostasis but also influence drug transport to the brain. These comprise carrier-mediated transport (CMT) and receptor-mediated transport (RMT) systems including cationic and anionic influx

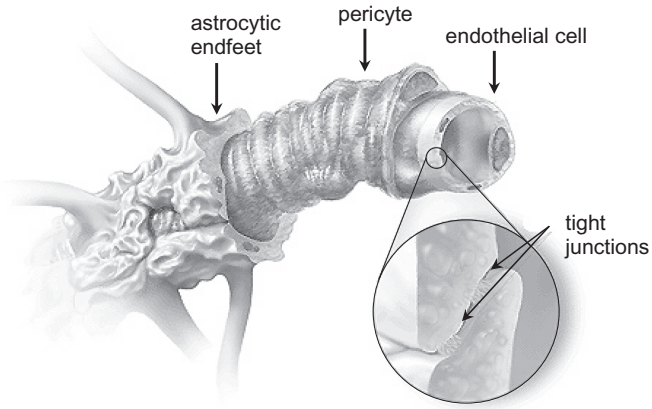


**FIGURE 2** Schematic representation of surrounding pericytes (covering about 20–30% of the capillary surface) and astrocytic endfeet projecting on the endothelial cells of the cerebral capillaries that induce and maintain the blood–brain barrier. In contrast, endothelial cells of peripheral capillaries do not form a tight barrier because they lack the specific input of these brain cells. See text for details. (Reprinted with permission from G. Miller, *Science* 2002;297:1116–1118. Illustration: C Slayden. © 2002 AAAS.)

and efflux systems like P-glycoprotein [32, 33], multidrug resistance proteins [34, 35], nucleoside transporters, organic anion transporters (OATs), organic cation transporters (OCTs), large amino acid transporter, and the RMT systems like the transferrin-1 and -2 receptors, the melanotransferrin receptor [27], and the scavenger receptors SB-AI and SB-BI [36–38].

Because of the complex interactions between cell types and the transport of compounds to and from the brain, as well as the dynamic regulation of the BBB properties (e.g., receptor expression, formation of tight junctions), the "multitasking" BBB [39], or the "neurovascular unit" [5, 40], is considered to be an organ protecting the brain [36]. In addition, these properties are influenced by disease and drug effects, which will change the functionality of the BBB under such conditions and will influence drug delivery to the brain [6, 7, 41].

### 2.2 Blood–Cerebrospinal Fluid Barrier (BCSFB) and Ependyma

The BCSFB is a rather complicated system (see other reviews also [42–44]). It comprises mainly the choroidal and arachnoidal epithelium giving access to the ventricular and subarachnoidal CSF, respectively. Considering its location and the direction of the CSF flow, the choroidal epithelium at the choroid plexuses (CPs) can be considered to be the most important part of the BCSFB, which are located in the lateral ventricles and in the third and fourth ventricles. The CP functions as a physical, an enzymatic, and an immunological barrier and plays a role in drug metabolism, drug transport, repair, and signaling. It presents phase I–III enzymes [43]. Phase I enzymes are used for functionalization of drugs and include cytochrome P450 isoform (CYP2B1,2) and monoamine oxidase. Phase II enzymes are used for conjugation of drugs and include glutathione and glucuronosyl transferases. Phase III enzymes include transporters like Na-dicarboxylate cotransporter, the ascorbic acid trans-

porter, the organic anion transporter (OAT), the organic anion transporter polypeptide 1 and 2 (OATP1,2), the organic cation transporter (OCT), equilibrating and concentrative nucleotide transporters (ENTs and CNTs), and multidrug transporters like P-gp and MRP1 [4, 33, 34, 36]. Furthermore, the CP seems to be involved in repair by secreting neuroprotective compounds and acting as a site of neurogenesis [43]. In addition, the CPs express many receptors and several of them are involved in signaling between the immune system and the brain ([46]) and pathology of the CP has been found in many CNS disease conditions [43].

In humans the total volume of the CSF is about 160 mL and the formation rate is about 0.35–0.40 mL/min. It is produced mainly by the CPs (60%) and the other part (40%) via ultrafiltration by brain capillaries [47]. It flows from the lateral to the third and fourth ventricles and subsequently into the cisterna magna and other large basal cisterns (see Fig. 1). From there the CSF flows posterior and downward into the subarachnoid space around the spinal cord and upward around the cerebral hemispheres. Via the arachnoidal villi in the subarachnoidal space the CSF flows outward into venous blood.

Since the endothelium in the blood capillaries in the CP is fenestrated, resistance to drug transport seems to be presented by (gap) junctions of the CP epithelium, which are more permeable than the tight junctions of the BBB endothelium [42]. Moreover, blood flow in the CP blood capillaries seems to be 5–10 times higher than the mean cerebral blood flow [43, 46, 48]. In addition, it has been calculated that the total surface area of all CPs is of the same order of magnitude as the entire BBB [13]. All these facts together allow for the possibility of substantial drug transport via the CPs into the CSF. However, as mentioned before, the surface area of the BCSFB is facing the CSF and not the blood, which makes drug transport to the CSF less effective. Moreover, from the point of view of drug transport to the brain parenchyma, the BCSFB has some disadvantages. Once a drug is in the CSF, an additional barrier for the transport of molecules is presented by the ependyma covering the ventricles. The ependyma is a single layer of epithelial cells connected by (gap) junctions and seems to be rather permeable, particularly to small molecular lipophilic compounds, while permeability for macromolecular drugs has been claimed (see also in Section 4.1, Direct Injection/Infusion of Macromolecular Drugs) [42].

## 3 METHODS TO MEASURE DRUG TRANSPORT TO THE BRAIN

### 3.1 *In Vitro* Methods

Research on drug transport across the BBB and BBB functionality has been very much enhanced by the availability of *in vitro* BBB (co)culture systems and many attempts have been made to develop such systems [49–60]. In addition, the use of such systems allows a detailed investigation of BBB related phenomena at the (sub)cellular level in the absence of feedback systems from the rest of the body. This makes it much easier to study *in vitro* BBB transport and BBB functionality by (pharmacological) intervention techniques like the application of receptor agonists and antagonists, blockers of transporters and enzymes, interference RNA and (anti)gene approaches, and the influence of disease.

BBB (co)culture systems comprise primary isolated cells (including their isolation procedures) or cell lines, the use of endothelial cells alone or cocultured with astrocytes (rat fetal or $C_6$-glioma), the application of epithelial cells as a BBB model, and the use and availability of human cell lines [61]. One of the most important issues is the variability and transference of data between laboratories [62] and the pitfalls in performing experiments and evaluating permeability data [63].

### *Primary Isolated Cells*

*Comparison of Isolation Procedures* Preparation of the *in vitro* BBB from primary isolated cells involves the isolation of capillaries and culture of BCECs alone or in combination with astrocytes or astrocyte conditioned medium. The first step in the culture of BCECs is the isolation of brain capillaries or brain capillary fragments. These can be isolated from human, bovine, porcine, and rat brain by various procedures, all starting with the grey matter of the brain [64]. Although human material would be most ideal from a scientific point of view, it is less easily available, and therefore most routinely applied procedures start from animal brains. Preferentially, bovine or porcine brains are used because of their larger mass and therefore the large amount of brain capillaries that can be isolated. The isolation procedures that have been used most frequently can be classified into nonenzymatic mechanical, combined mechanical–enzymatic, or enzymatic procedures [65–69]. The typical main steps in the various procedures are shown in Table 1.

The *enzymatic procedure* [65, 66] is rather time consuming (about 8 h). Following mincing of the brains into 1–2 mm cubes, two enzymatic and two separation steps are performed to isolate brain capillaries. First, an incubation step is performed with dispase followed by a dextran centrifugation/separation; and second, an incubation step with dispase/collagenase is followed by a Percoll gradient separation.

The *mechanical procedure* [55, 67] is fast and lasts 1–2 h for the isolation of capillaries from bovine brain. The brains are homogenized using two different potters. Subsequently, two filtration steps are applied: first, one isolating the material that stays behind on a 150 μm mesh, and second, one that collects the material, including capillaries that go through a 200 μm mesh.

The *combined mechanical–enzymatic procedure* [68, 69] lasts for about 4 h. It comprises the homogenization of brains by a Wheaton homogenizer followed by a filtration step through a 150 μm nylon mesh. The trapped material is subsequently treated with collagenase/DNAse and filtered through a 200 μm nylon mesh. The capillary fraction is collected by centrifugation and may be stored at –80 °C by freezing in a mix of fetal calf serum and 10% (v/v) DMSO.

*Culture of BCECs* Brain capillaries may be plated and cultured in culture flasks. The BCECs, which have grown out of the capillaries, may be isolated by (micro)trypsinization and plated in culture flasks or on (polycarbonate) porous filters. Cecchelli's group [55, 67] claims that the mechanical procedure results in pure BCEC cultures: first, because of the isolation of brain capillaries (not arterioles and venules) that is provided by a selective adhesion at an extracellular matrix of corneal endothelial cells, and second, because of a microtrypsinization that avoids contamination by other cell types. In the other two methods, collagen (human placenta type IV) and human plasma fibronectin [69] or rat-tail collagen with fibronectin [65, 66]

**TABLE 1 Overview of the Main Steps in the Various Brain Capillary and Brain Capillary Endothelial Cell (BCEC) Isolation Procedures**

| Mechanical [55, 67] | Mechanical–Enzymatic [68, 69] | Enzymatic [65, 66] |
|---|---|---|
| Bovine brain grey matter | Bovine brain grey matter | Bovine brain grey matter |
| Homogenization (with two different potters) | Homogenization (Wheaton homogenizer) | Minced to 1–2 mm cubes |
| Filtration step 1 (<150 μm mesh) | Filtration step 1 (<150 μm mesh) | Dispase incubation (2.5 h) |
| Filtration step 2 (>200 μm mesh) | Collagenase/DNAse incubation (1 h) | Dextran centrifugation |
| Selective adhesion of capillaries at CEC[a] matrix) | Filtration step 2 (>200 μm mesh) | Collagenase/dispase incubation (4 h) |
| Growing BCECs | Selective adhesion of capillaries at collagen type IV)[b] and fibronectin[c] matrix | Percoll gradient |
| Collection of BCECs by microtrypsinization | Growing BCECs | Selective adhesion of capillaries at rat-tail collagen and fibronectin matrix |
| Plating BCECs and astrocytes at bottom of well | Collection of BCECs by trypsinization | Growing BCECs |
| | Plating BCECs and astrocytes at the other side of the filter | Collection of BCECs by trypsinization<br>Plating BCECs |

[a]CEC, corneal endothelial cell.
[b]Human placenta type IV collagen.
[c]Human plasma fibronectin.
*Source*: From Ref. 362, with permission.

replace the selective isolation of capillaries by the corneal endothelial cell matrix. This avoids the periodical isolation of corneal endothelial cells. The selectivity of the latter two methods is further guaranteed/enhanced by the application of either a Percoll gradient [65] or the time for attachment (2–4 h) of the capillary fragments at the collagen/fibronectin coating [69]. In addition, the application of 50% ACM in the procedure of Gaillard et al. [69] enhances the selective outgrowth of BCECs. Subsequently, in one step BCECs can be selectively collected following a short trypsinization time (<1.5 min). By this procedure one can obtain enough BCECs to get about 1000 (24 wells) filters from one isolation. The method of Audus and Borchardt [65, 66] also results in the isolation of capillaries and subsequently in a large amount of endothelial cells. In addition, the procedure of Dehouck et al. [55, 67] results in a substantial amount of BCECs to cover filters (about 300), but BCECs have to be passed at least 4–7 times. Unfortunately, passing primary cultured cells will eventually lead to a loss of BBB properties. In this respect, the preservation of properties is in our opinion an important issue and it has been shown that several can be reinduced by astrocytes [70–72].

The mechanical isolation procedure reduces the risk of losing surface molecules, which can occur during enzymatic treatment, particularly when aspecific proteases are applied and when these are not reinduced. This risk is less following collagenase treatment since this is a more selective enzyme. Therefore, induction of surface molecules is particularly important for enzymatically isolated cells, since it helps to reinduce BBB properties/molecules. In addition, it may be worthwhile to know that endothelial cells have the special property to transdifferentiate from one phenotype to another by the influence of environmental factors [73].

*Importance of Astrocytes in BBB Culture Systems* BCECs may be grown to confluency alone or in the presence of ACM or astrocytes. The use of astrocytes or ACM enhances the expression of various surface molecules and transporters and increases paracellular tightness [49–51, 68, 70–72].

Most frequently, astrocytes isolated from newborn rats and rat $C_6$-glioma cells have been used. Together with bovine or porcine BCECs, one obtains xenogenic coculture systems, which are very useful in studying drug transport and BBB functionality. In addition, preliminary data from our lab show that it is possible to construct an allogenic system of bovine BCECs with bovine astrocytes. However, these astrocytes were obtained from brains of 6–8 month old calves, which grow significantly slower than astrocytes of newborn rats and prolong culture time. A rat allogenic BBB coculture system has been established comprising rat BCECs and fetal rat astrocytes [74], and a mouse syngenic BBB coculture model has also been developed [75]. However, the disadvantage of these systems is that the amount of endothelial cells that can be obtained from mouse and rat brain is rather small in comparison to bovine and porcine brain and therefore these procedures are not yet suitable for large scale production of cocultures.

Astrocytes can be seeded at the bottom of (mostly polycarbonate) porous filters and at the bottom of the well. This may give rise to differences in endothelial functionality since factors released by the astrocytes influence BBB endothelial cells [67, 70–72]. Therefore, to study BBB functionality, particularly under acute stress conditions, endothelial cells and astrocytes should be in contact or in close approximation to each other for an optimal feedback. This is less necessary in studying only endothelial permeability of drugs. However, it may be argued that due to pharmacodynamic effects of drugs, increased or decreased endothelial permeability may be found caused by a less adequate feedback by astrocytes and astrocytic factors.

*Other Cells in BBB Cultures* Other cells that may occur following isolation of BCECs are pericytes and fibroblasts. Pericytes are normally present at the BBB and are surrounded by a basement membrane and have been shown to induce separately specific enzymes in BCECs [76]. In addition, it has been argued that pericytes are necessary to establish an *in vivo*-like BBB cell culture model [31]. Fibroblasts are present in brain tissue, but not very much is known about their influence at the BBB. Other cells may also have an important action in BBB cell culture systems, such as neurons, microglia, and monocytes, and it may be argued that combinations of these cells should be present in a BBB cell culture system for optimal functionality. This has been described for an *in vitro* BBB model comprising human BCECs, human fetal astrocytes, human macrophages, and neuronal (human teratocarcinoma) cells [77]. Other systems have been developed comprising cultured monolayers of brain

endothelial cell lines or selected epithelial cell lines, combined with astrocyte and neuron cultures that present a novel three-dimensional technique for the screening of neurotoxic compounds [78].

As mentioned before, human material is rather difficult to obtain regularly in large quantities. A good alternative is the use of human umbilical vein endothelial cells (HUVECs). These can be rather easily obtained and when cultured under proper conditions, such as with astrocytes, a human-like BBB system may be obtained [79]. Alternatively, human endothelial cells from skin capillaries can be used in combination with postnatal rat astrocytes [80]. Such systems are convenient for several purposes (such as transcellular lipophilic/hydrophilic transport and carrier/receptor-mediated transport) and are easy to obtain in large quantities. Still, it may be expected that the functionality of such cells is different from primary cultured BCECs (with respect to tightness, up/downregulation of transporters, etc.) although this has not been demonstrated until now.

***Application of Barrier Enhancers*** It has been shown that several compounds like cAMP, phosphodiesterase inhibitors, glucocorticoids, and interferon-α,β may be used to enhance the tightness of BBB endothelial monolayers [68, 69, 81, 82]. In addition, the application of basic fibroblast growth factor (bFGF; secreted by astrocytes [83, 84]) is important in growing/establishing functional and tight BBB cell culture systems [55] and in the induction of BBB properties [84].

Application of these enhancers has increased the ease of culture and the properties of the cell culture system. However, apart from the lack of functional influence of astrocytes, the question that still has to be answered is how these compounds influence other aspects of BBB functionality in addition to endothelial tightness.

***Endothelial Cell Lines*** Various immortalized brain capillary endothelial cell lines have been derived. These are made by transfection of these cells with an immortalizing gene. An overview of the most frequently used cell lines is given in Table 2. These cell lines have various properties that resemble *in vivo* BBB capillary endothelial cells. The advantage of cell lines is their ease in culture, their purity, and the fact that there is no need for a periodic isolation of capillaries from brains. Since these cell lines are obtained by immortalization, it may be expected that their functionality will be changed compared to primary cultured BCECs. This is illustrated by the fact that most of these cell lines are rather permeable (paracellular leakiness), which limits their use in studying BBB transport. The application of these cells is particularly useful for (immuno)histochemical identification of extra- and intracellular molecules, for studying morphological aspects of endothelial cells, and so on. In addition, barrier enhancers (see also previous paragraph) may be used to improve the tightness of these monolayers. Recently, a human brain immortalized endothelial cell line was established by transfection of the human telomerase or SV40 T antigen. This cell line (hCMEC/D3) represents a stable, well characterized, and well differentiated human brain endothelial cell line and offers opportunities for BBB research [61].

***Tumor Cell Systems*** The application of rat $C_6$-glioma cells makes the culture of astrocytes much easier. This provides the same advantages as with cell lines. Table 3 gives an overview of the various sources of astrocytes and their application in

**TABLE 2 Overview of the Used Immortalized Brain Capillary Endothelial Cell Lines**

| Cell Line | Source | References |
|---|---|---|
| SV-BEC | Bovine | 100 |
| t-BBEC-117 | Bovine | 101 |
| BBEC-SV | Bovine | 102 |
| RBEC1 | Rat | 103 |
| TR-BBB | Rat | 104 |
| RBE4 | Rat | 105 |
| CR3 | Rat | 106 |
| RCE-T1 | Rat | 107 |
| GPNT | Rat | 108, 109 |
| GP8.3 | Rat | 110 |
| TM-BBB | Mouse | 111 |
| MBEC | Mouse | 112 |
| S5C | Mouse | 113 |
| PBMEC/C1-2 | Porcine | 114 |
| SV-HEC | Human | 115 |
| BB19 | Human | 116 |
| HBEC-51 | Human | 117 |

*Source*: From Ref. 363, with permission.

**TABLE 3 Astrocytes Used in Coculture with Endothelial or Epithelial Cells of Various Sources**

| Astrocyte | Coculture with Endothelium[a] |
|---|---|
| Newborn rat astrocyte | BCECs (bovine) [55, 67, 69] |
| Newborn rat astrocyte cell line ($A_7$) | BCECs (bovine) [118] |
| $C_6$-rat glioma | BCECs (bovine) [119] |
| $C_6$-rat glioma | HUVECs (human) [79] |
| $C_6$-rat glioma | RBE4 (rat) [120] |
| $C_6$-rat glioma | Aortic (bovine) [121] |
| TR-AST (rat)[b] [122] | |
| Newborn mouse astrocytes | BCECs (mouse) [75] |
| Human fetal astrocytes | BCECs (human) [77] |
| **Astrocyte** | **Coculture with Epithelium[a]** |
| $C_6$-rat glioma | ECV304 (human) [81] |
| $C_6$-rat glioma | MDCK (dog) [80] |
| Human 1321N1 astrocyte cell line | ECV304 (human) [81] |

[a]HUVECs, human umbilical vein endothelial cells; RBE4, rat brain immortalized capillary endothelial cell line; ECV, T24 human bladder epithelial carcinoma cell line; MDCK, dog kidney epithelial cell line; BCECs, brain capillary endothelial cells.
[b]Conditionally immortalized astrocyte type II.

*Source*: From Ref. 362, with permission.

BBB coculture systems. The application of tumor astrocytic cell lines introduces the consequence of a tumor phenotype in BBB cocultures. This may be very helpful if one is interested in the blood–tumor barrier in the brain because the homeostasis of the tumor–BBB endothelium may be different. Surface molecules and transport-

ers may be up- or downregulated and may give leads for tumor targeting. These arguments make these BBB systems more suitable to study only passive hydrophilic or lipophilic drug transport. Since active transport processes may be differentially regulated under tumor conditions, the results of such transport studies are less relevant for those under "healthy" BBB conditions. Therefore, these tumor BBB coculture systems are not suitable to study active transport processes and functionality under other than tumor conditions. Apart from the tumor phenotype, these BBB coculture systems differ with respect to the source of the endothelial cells, while the ECV304 (recently characterized as a T24 bladder epithelial carcinoma cell, see [58]) and MDCK (dog kidney epithelial cell line) cell lines represent epithelium. It has recently been shown that the basal transendothelial electrical resistance (TEER) of ECV304 monolayers could be enhanced by human 1321N1 astrocytes and primary rat astrocytes and that relatively tight monolayers could be obtained [81]. However, the replacement of endothelial cells by epithelial cells may have some consequences. In the first place, endothelial cells are longitudinal cells with a thickness of 2–4 μm in contrast to MDCK cells that are columnar cells with a thickness of more than 20 μm. Second, (drug) metabolism by epithelial cells may be quite different from endothelial cells; and third, transporters may be up- or downregulated and absent or mutated.

***Variability and Transference of Results Between BBB Culture Systems*** BBB culture systems have been the subject of a Concerted Action within the EEC entitled *Drug Transport Across the BBB: New Experimental Strategies*, involving 21 research groups from 9 European countries [85]. The focus of this Concerted Action was on the optimization, harmonization, and validation of brain-microvessel-endothelial cell culture and microdialysis to develop and study new experimental strategies for drug transport to the brain. The various BBB culture systems have been described in the report of this Concerted Action including the sources of variability [62]. Furthermore, pitfalls in evaluating BBB data should be noted in order to avoid bias and misinterpretation of results [63]. Here we want to indicate and compare the various properties of the most frequently applied BBB (co)culture systems. In addition, others have indicated the need for validated BBB culture systems [86, 87]. It has been discussed elsewhere that the transference of data from various BBB culture systems and the evaluation of the results would be enhanced if these were obtained in a well validated context [62]. Presently, there is no laboratory-to-laboratory validation of BBB systems and only individual researchers characterize and validate their BBB systems by the expression of various surface molecules and enzymes, paracellular marker molecules ($^{14}$C-mannitol, $^{14}$C-sucrose, fluorescein, and FITC-dextran MW 4000), and the measurement of the transendothelial electrical resistance (TEER) [82, 88]. In addition, since TEER values measured before the experiment do not guarantee BBB tightness during the experiment, this method is very helpful to estimate BBB performance during and after the experiment also. Efforts have been focused on the use and validation of *in vitro* BBB systems to estimate BBB transport of drugs and on the *in vitro*/*in vivo* correlation of BBB drug transport by the ECVAM (European Center for Validation of Alternative Methods) [89, 90].

Furthermore, the misinterpretation of results due to biased *in vitro* conditions may be a serious point [63]. Physiologically relevant *in vitro* conditions (pH, serum

proteins, unstirred water layer, cellular retention, and use of artificial membranes) should be established as much as possible and recommendations have been made to optimize *in vitro* systems accordingly [63].

***Application of BCEC Cocultures in Research*** BBB (co)cultures are useful to study a whole variety of phenomena. These include transport processes like passive hydrophilic (paracellular) transport, passive lipophilic transport, carrier- or receptor-mediated (transcellular) transport, and fluid-phase and adsorptive-mediated transport. The discovery of efflux transporters like P-glycoprotein and MRPs (reviewed in [34–38, 91]) and influx transporters like the amino acid, the monocarboxylic acid, the cationic drug, the hexose, the nucleotide, and the peptide transport systems (reviewed by [92]) have contributed much to our understanding of BBB transport. In addition, drug metabolism [22–24], the effect of specific agents and conditions on the functionality of the BBB including disease state, the visualization of drug transport routes by confocal laser scanning microscopy (CFLM) [93], and the measurement of the transendothelial electrical resistance [80] can be studied.

Particularly, disease state is an important BBB condition to be studied, since it has been shown to change BBB functionality and therefore drug transport to the brain [80]. In addition, inflammatory conditions, which may occur in multiple sclerosis, Alzheimer disease, AIDS-related dementia, meningitis, and encephalitis, can be induced at the *in vitro* BBB by applying lipopolysaccharide (LPS), TNF-α, IL-1β, and NO [6, 7, 9, 75]. Under such disease conditions, BBB permeability, functionality, and drug transport can be studied together with (pharmacological) intervention. The application of functional genomics and the phage-display technique to BBB coculture systems is in this respect very interesting in order to identify transcribed genes and expressed surface molecules under various conditions including disease state.

***Conclusions*** The presently available *in vitro* BBB (co)culture systems have their specific properties and can be used for specific goals (see Table 4). These goals can be passive hydrophilic and lipophilic drug transport, the study of active transport processes, and BBB functionality. It may be clear that for the study of passive drug transport through the BBB, sophisticated BBB systems are not required. However, since monolayers may be leaky, a wrong interpretation may be obtained on the contribution of para- and transcellular transport to the total transport and corrections should be made [94]. BBB coculture systems that resemble as much as possible the *in vivo* BBB must be used when active transport processes and the study of BBB functionality are involved. Tumor (co)culture systems are more useful to study the tumor BBB, while other systems [67, 69] are more useful to study active BBB processes and BBB functionality under healthy and disease conditions.

The influence of BBB pathology and the influence of disease mediators like LPS, NO, and radical oxygen species/radical nitrogen species (ROS/RNS) on BBB permeability is well known, particularly with respect to paracellular permeability [6, 7, 9, 75]. In addition, disease may up- or downregulate active transcellular transport systems (transporters and transcytosis mechanisms). Less is known about the influences of the pharmacodynamic effects of drugs at the BBB, which may increase (e.g., glucocorticoids) [82, 95] or even decrease (doxorubicin, vinblastine given together

**TABLE 4 Overview of the Various *In Vitro* BBB Coculture Systems Involving the Use of $C_6$-Rat Glioma Cells**

| E/A Coculture | Useful for Study of | Not Ideally Suited for Study of | References |
|---|---|---|---|
| Endothelium/astrocyte coculture | | | |
| HUVEC (human)/ $C_6$-glioma (rat) | L; H; CLSM | BBB-F; BBB-D; DT; DM; DE | 79 |
| BCEC (human)/ human fetal A | L; H; CLSM; DT; DM; DE "human"-BBB; BBB-F; BBB-D | | 77 |
| BCEC (bovine)/ newborn (rat) A-cell line | L; H; CLSM | BBB-F; BBB-D; DT; DM; DE | 118 |
| BCEC (bovine)/ newborn rat A | L; H; CLSM; DT; DM; DE "normal"-BBB; BBB-F; BBB-D | | 67, 69 |
| BCEC (mouse)/ newborn mouse A | L; H; CLSM; DT; DM; DE "normal"-BBB; BBB-F; BBB-D | | 75 |
| BCEC (bovine)/ $C_6$-glioma (rat) | L; H; tumor-BBB; CLSM | BBB-F; BBB-D; DT; DM; DE | 119 |
| RBE4 (rat)/ $C_6$-glioma (rat) | L; H; tumor-BBB; CLSM | BBB-F; BBB-D; DT; DM; DE | 120 |
| Epithelium/astrocyte coculture | | | |
| MDCK (dog)/ $C_6$-glioma (rat) | L; H; CLSM | BBB-F; BBB-D; DT; DM; DE | 123 |
| ECV304 (human)/ $C_6$-glioma (rat) | L; H; CLSM | BBB-F; BBB-D DT; DM; DE | 81 |
| ECV304 (human)/ 1321N1 (human) | L; H; CLSM | BBB-F; BBB-D DT; DM; DE | 81 |

Abbreviations: E/A, endothelial or epithelial coculture with astrocytes; A, astrocyte; L, passive lipophilic transport; H, passive hydrophilic (paracellular transport); BBB-F, BBB functionality; BBB-D, BBB diseases; DM, drug metabolism; DE, drug effects; DT, BBB drug influx and efflux transport systems; CLSM, confocal laser scanning microscopy.

*Source*: From Ref. 362, with permission.

with a P-gp inhibitor) paracellular tightness [72, 96, 97]. In addition, several diseases in the brain (multiple sclerosis, Alzheimer disease, AIDS-related dementia, meningitis, epilepsy, tumors, etc.) are the primary or secondary cause of BBB permeability. Therefore, it is important to consider the pharmacodynamic (therapeutical or toxicological) effects of drugs at the BBB also.

Variability [62] and physiologically relevant *in vitro* conditions [63] are probably the main obstacles to the transference of results between laboratories. An interesting example is the expression of P-glycoprotein in BCECs depending on cell culture conditions. Barrand et al. [98] have shown that mdr1a–P-gp in cultured rat BCECs was downregulated and changed into the expression of mdr1b–P-gp. In addition, it has been shown that P-gp expression in bovine BCECs could be upregulated again by rat astrocytes [72]. Similar data have been published by Zhang et al. [99], where the expression of MRPs (at mRNA level) in brain capillaries was qualitatively and

quantitatively different from those in BCEC monolayers. Therefore, it may be stressed again, that the systems to be used are well characterized morphologically, biochemically, and functionally. Together with the use of validation markers [86] and physiologically relevant *in vitro* conditions, this will lead to an optimal transference of BBB data between laboratories [62, 63, 89, 90].

### 3.2 *In Vivo* Methods

Although the *in vitro* systems have evolved into sophisticated and functional models of the BBB, they may result in quantitative and qualitative differences in BBB transport due up- or downregulation of transporters and species differences. Therefore, it is necessary to verify and validate the obtained *in vitro* data *in vivo*. This requires the application of animal-based methods. There are several methods; however, their applicability depends on the sensitivity and selectivity to measure drug concentrations in the brain of fast and poor BBB/brain penetrating compounds, the estimation of local concentrations in the brain or whole brain distribution, and the measurement of single-time concentrations versus concentration time profiles. Therefore, it is essential to understand the possibilities and limitations of these methods for the estimation of BBB transport including the role of drug influx and efflux transporters at the BBB [36].

Here we briefly discuss the following methods in the context of their possibilities and limitations to estimate BBB transport: brain uptake index (BUI), multiple pass techniques, brain efflux index (BEI), *in situ* brain perfusion, CSF (cerebrospinal fluid) sampling (unit impulse response), positron emission tomography (PET), magnetic resonance techniques (imaging (MRI), spectroscopy (MRS)), quantitative autoradiography (QAR), and intracerebral microdialysis. Most of these methods are invasive in contrast to the noninvasive ones like PET and NMR.

***Brain Uptake Index (BUI)*** This is one of the oldest techniques (1970) to estimate the uptake (via the BBB and/or blood–CSF barrier) of drugs/compounds into the brain [124, 125]. The method is widely used, rapid, and relatively cheap and the surgery is easy. It is a single passage technique and comprises the intra-arterial administration (common carotid artery) of a solution containing a reference (mostly a $^{14}$C-labeled compound) and a test compound (preferentially a $^{3}$H-labeled compound) under anesthesia. The basic idea is that the reference compound can easily penetrate the brain and the test compound less. Following a short time (15 s) the animal is decapitated and brain concentrations of test and reference compounds are measured and related to the plasma concentration. This ratio is a measure for the brain uptake of the test compound that is expressed as a percentage of the penetration of the reference compound. This procedure is very suitable for compounds that are labile or fast metabolized. Its drawback is that the exposure to the brain is very short and results from BUI studies cannot easily be related to cerebrovascular permeability. These are dependent on blood flow, brain region, and time between injection and decapitation. Therefore, this procedure is not suitable for poorly penetrating compounds including peptides and proteins. Another limitation is that the total injected dose does not reach the whole brain since about 20% of the drug enters the internal carotid artery while 80% reaches the external carotid artery.

***Brain Efflux Index (BEI)*** The BEI has recently (1996) been developed [126, 127]. It comprises the estimation of the efflux of compounds from the brain following microinjection (maximal 1 μL) at a specific site in the brain, compared to a reference compound ($^{14}$C-carboxy-inulin) that has a limited BBB permeability. Following decapitation at variable times, the brain and plasma concentration of compound and reference can be estimated. The BEI is expressed as the percentage of the ratio of drug effluxed from the brain and the drug injected into the brain. The apparent elimination rate constant can be calculated by plotting the log(BEI) versus time. Subsequently, the efflux clearance of compounds from the brain can be calculated. The technique is cheap and the surgery is moderately easy.

***Multipassage Techniques*** These techniques have been developed to circumvent several of the drawbacks of the BUI method and were originally designed by Brodie et al. [128]. The exposure is much longer and thus measurable brain concentrations of poorly penetrating compounds can be obtained. Later on, mathematical models were developed to calculate the permeability surface area (PS) product after an intravenous bolus administration [129]. Generally, the concentration of the drug in the brain is divided by the integral of the plasma drug concentration and time (AUC). This results in the estimation of the clearance of drug from blood or plasma to brain and is expressed as the PS product for BBB and/or BCSFB transport. For an accurate estimation of the PS of slowly penetrating compounds, it is essential to correct for the amount of drug in the vascular bed of the brain. Furthermore, it has been speculated that the PS product was time dependent [130].

The multiple-passage/multiple-time-point techniques provide information about the time course of brain uptake [128, 131, 132]. Originally, these experiments were performed by maintaining plasma steady-state concentrations and following the time course of brain uptake by sacrificing groups of animals at different time points. BBB transport can be estimated when the concentration difference between brain or CSF and plasma relative to the plasma concentration is plotted versus time. Later on, this procedure was further developed to intravenous bolus administration [133]. Ultimately, the PS products can be calculated by the Renkin–Crone equation [134]. The interesting advantage of this equation is that the amount of drug penetrating the brain can also be estimated by positron emission tomography (PET) and computer tomography (CT) scans, which allow a noninvasive measurement of BBB permeability in humans [135]. An additional advantage is that the regional blood volume does not have to be estimated. However, since brain sampling can only be done after sacrificing the animal, it is not possible to study the time profile of BBB transport in a single animal. The technique is cheap and the surgery easy to perform.

***In Situ Brain Perfusion Techniques*** These were developed (1984) to avoid (non-brain) metabolism of the test compound and to fully control the passage of the BBB and/or the BCSFB. These techniques do not require estimation of the regional blood volumes to correct brain concentrations. In this technique the common carotid artery is ligated and the ($^{14}$C-labeled) reference and ($^{3}$H-labeled) test compounds are perfused retrograde via the external into the internal carotid artery under anesthesia [136]. Following perfusion the animal is decapitated and the brain is collected and analyzed for reference and test compounds. The use of two test compounds (a

highly diffusible one and the other an intravascular marker compound) allows an accurate estimation of the PS product [129]. The problem with this technique is that in rats the common carotid artery has branches between the neck region and the brain. Therefore, to ensure the delivery of compounds to the brain, these branches have to be ligated. Guinea pigs do not have these branches and are therefore more convenient animals to use in this respect [137]. The technique is cheap and surgery is easy.

***Unit Impulse Response Procedure*** This procedure together with numerical deconvolution [138–140] has been applied (1989) to estimate BBB and/or BCSFB transport based on serial CSF sampling in the *cisterna magna* [141]. This procedure allows the construction of a complete time–transport profile in a single anesthetized animal. It requires the estimation of a weighing function comprising the characterization of the disposition of a drug in the brain. This is achieved by injecting a compound in the lateral ventricle of the brain, followed by sampling of the CSF. Subsequently, the compound is given intravenously and plasma and CSF samples have to be collected. The BBB input profile can be obtained by a mathematical procedure called deconvolution [139, 140]. Principally, one can obtain information on passive and active transport processes in the brain. Essentially, the cumulative cleared plasma volume is plotted against time. When a straight line is obtained, its slope represents the uptake clearance. In the case of nonlinearity, the $V_{max}$ and $K_m$ can in principle be estimated. The method has been applied to lipophilic drugs like atenolol, acetaminophen, and antipyrine [141]. The advantage of this technique is that it avoids the occurrence of backdiffusion from brain to blood. The costs of the technique and the complexity of the surgery are moderate.

***Positron Emission Tomography (PET)*** Regional BBB and BCSFB permeability and drug transport can be investigated by PET [142]. This technique employs the intravenous administration (bolus or infusion) of positron emitting isotopes like $^{18}F$, $^{13}N$, $^{11}C$, or $^{15}O$. By this technique brain and plasma concentrations can be followed over a long time. Subsequently, one analyzes the data by pharmacokinetic models that describe the transport of tracers (uptake, distribution, and elimination). In 1955, this technique was developed by Davson [143]. The advantage of the technique is that the transport of tracers can be visualized and studied in whole brain over time. However, the technique is expensive and the preparation and stability of the tracers are matters of concern. Essentially, tracer concentrations are measured and no distinction can be made between parent compound, metabolite, and protein bound or unbound tracer. Nevertheless, this technique is sensitive and therefore very helpful in diagnostic purposes such as the localization of tumors in the brain and also in studying local differences in BBB permeability and BBB transport. P-gp functionality has been investigated with $^{11}C$-verapamil in mdr1a (+/+) and (–/–) mice [142, 144]. The PET procedure underestimated the uptake of $^{11}C$-verapamil in the brain. However, in patients, pretreatment with cyclosporin A increased $^{11}C$-verapamil levels in GLC4/P-gp tumors (184%) and in brains (1280%). This pharmacokinetic effect was clearly visualized with PET [145].

***Quantitative Autoradiography (QAR)*** This method was initially applied in 1951 by Campbell [146] and later on used to visualize the distribution of radioactive

tracers across the BBB and/or the BCSFB [131]. Typically, these are $^{3}$H- of $^{14}$C-labeled compounds that are given intravenously. Following administration, the animals are sacrificed after various times, brains collected, and slices prepared. Subsequently, autoradiographs are prepared that can be scanned by a computer driven densitometer. The technique is cheap and easy to perform. The spatial resolution of QAR is high; however, the time resolution is rather low since one animal has to be used for one time point. Other drawbacks are that no distinction can be made between parent compound and metabolites, and protein bound and unbound compounds.

***Intracerebral Microdialysis*** This method was developed in the late 1950s and measures particularly local concentrations of compounds in the extracellular fluid (ECF) of the brain or the CSF. It is an invasive method and longitudinal, I-shaped or semicircular probes can be used [147]. The time to perform experiments has to be evaluated carefully with respect to periprobe tissue reactions that may occur [148]. The method has been successfully applied in the estimation of BBB transport of various compounds [148–151]. The advantage of this method is that concentration–time profiles can be obtained of unbound compounds in the extracellular fluid in the brain. The spatial resolution of this technique is relatively high, and a highly sensitive and selective analytical technique is essential for estimation of concentrations of compounds in the ECF. Presently, HPLC combined with mass spectroscopy methods result in high selectivity and sensitivity. The surgery is moderately difficult and the costs are moderate.

An essential factor is the estimation of the *in vivo* recovery of the microdialysis probe. This describes the relation between the concentration of the compound in the fluid around the probe and that in the dialysate. *In vitro* recovery is mainly dependent on the perfusate composition and flow rate, the physicochemical characteristics of the compound, temperature, and probe characteristics. In addition, *in vivo* recovery is influenced by the para- and transcellular transport of compounds through brain tissue and the blood supply in terms of vascularization and blood flow [147]. *In vivo* recovery can be estimated by the "nonet flux" (measurement of recovery under serial steady-state conditions in one animal) [152] and the "dynamic nonet flux" (measurement of recovery under one but different steady-state conditions in different groups of animals) [153] methods and by the internal standard, the retrodialysis, and reverse dialysis methods (reviewed in Ref. 147).

Intracerebral microdialysis allows the estimation of the local BBB time transport profile of unbound compounds in the ECF of the brain and in the CSF. It can distinguish between parent compound and metabolite(s). The analytical technique to estimate the substances in the ECF or CSF must be very sensitive and selective. In addition, the conversion of dialysate concentrations to real extracellular concentrations in the brain is very much determined by the procedure to estimate *in vivo* recovery.

***Magnetic Resonance Techniques*** These comprise magnetic resonance imaging (MRI, initially applied in the late 1970s) and magnetic resonance spectroscopy (MRS, initially applied in the 1950s) and are noninvasive techniques. They do not require the use of ionizing radiation or the use of iodinated contrast compared to X-ray computed tomography (CT). They comprise the direct or indirect monitoring

and recording of the spatiotemporal distribution of molecular or cellular processes for biochemical, biologic, diagnostic, or therapeutic applications [154]. The techniques employ $^{1}$H-, $^{13}$C-, and $^{31}$P-labeled compounds that are injected intravenously. This allows evaluation of a variety of parameters in the brain related to anatomy, physiology, and metabolism, like macrophage infiltration, cytotoxic edema, cerebral blood flow and volume, and metabolic processes [155]. Depending on the technique used, slice-based as well as true 3D images can be obtained [156]. MRI techniques have been developed into functional MRI (fMRI) techniques, which measure regional blood volume and flow, regional diffusion of water (diffusion tensor imaging—DTI) [157, 158], and regional blood oxygenation level dependence (BOLD) [159, 160]. In addition, when combining a contrast enhancing principle (superparamagnetic iron oxide—SPIO; $Mn^{2+}$, $Gd^{3+}$) [160] with a target-specific carrier moiety (receptor ligand, antibody, cell), highly specific information on targeted molecular interactions can be obtained by molecular (brain) imaging [161]. This is very important from the diagnostic and therapeutic (e.g., disease progression) point of view. However, the problem is to deliver these agents through the BBB, and successful strategies have to be developed to accomplish this. Nevertheless, MRI has been used to investigate BBB permeability and leakage involved in brain diseases including neurodegenerative diseases [157, 162]; various brain tumors [163, 164], multiple sclerosis [165, 166], early infarcts, and inflammatory lesions can be studied [157] in addition to age related changes in brain tissue [158]. Moreover, since there are many disease models in small animals, MRI can be used to study these (brain) disease processes dynamically [156].

MRS has been used successfully to quantify phenylalanine (Phe) concentrations in the brain of patients with phenylketonuria and to estimate its BBB transport (PKU) [167]. In addition, it was shown that BBB transport parameters determined the clinical outcome in phenylketonuria [168]. In studies with multiple sclerosis (MS) patients, it was demonstrated that the earliest detectable event in the development of a new lesion was the increase in permeability of the BBB that was associated with inflammation [169]. It was shown that when the inflammation declined, edema resolved. With $^{1}$H-MRS a correlation was shown between a decrease in *N*-acetylaspartate (NAA) concentrations (a marker of neuronal or axonal damage) and clinical disability [170].

With these techniques one can differentiate between parent compound and its metabolites as well as between free and unbound drug concentrations. The sensitivity and its spatial resolution are limited but resolutions ranging from 300 to 500 μm throughout the human brain and 50 to 100 μm through the rodent brain should be possible in reasonable scan times [159]. The latter are presently up to 2 hours while another disadvantage is that the technique is expensive [162]. Nevertheless, magnetic resonance techniques are very convenient to estimate BBB time transport profiles of poorly and fast brain-penetrating compounds.

***Conclusions*** Various methods have been developed to estimate drug transport to and from the brain. Table 5 shows an overview of the typical properties of the reviewed methods. Presently the brain perfusion method is one the most frequently used methods in animals. It allows the estimation of brain uptake of fast and poorly penetrating compounds at single times. In addition, when more animals are used a concentration–time profile can be obtained. The BEI method is applicable in the

**TABLE 5 Applicability[a] of Various Methods to Estimate Drug Transport and Distribution in the Brain**

| Method | Fast Penetrating Compounds | Poor Penetrating Compounds | Single Concentration at Single Time | Concentration–Time Course in Single Animal | Brain Distribution[b] | Costs | Surgery |
|---|---|---|---|---|---|---|---|
| BUI | + | – | + | – | – | Cheap | Easy |
| Multipassage techniques | + | + | + | – | – | Cheap | Easy |
| BEI | + | + | + | – | – | Cheap | Moderately easy |
| Brain perfusion | + | + | ++ | – | – | Cheap | Easy |
| Impulse response method | + | + | ++ | + | – (CSF) | Cheap | Moderately difficult |
| PET | + | + | + | ++ (human) | + (WB) | Expensive | Easy |
| QAR | + | + | + | – | ++ (WB) | Cheap | Easy |
| MRI/MRS | + | + | | ++ (human) | + (WB) | Expensive | Easy |
| Intracerebral microdialysis | + | + | | + | + (local) | Moderate | Moderately difficult |

[a]+, applicable; ++, very good applicability; –, not appplicable.
[b]WB, whole brain; CSF, only in cerebrospinal fluid.

same way and allows the measurement of brain efflux of compounds. The BUI, BEI, multipassage techniques, brain perfusion, and the impulse response method (only in CSF) do not give information on brain distribution. In contrast, PET, QAR, MRI/MRS, and intracerebral microdialysis provide information on the brain concentration–time course and brain distribution of compounds in one single animal. However, the resolution varies, with that of QAR and intracerebral microdialysis > PET and MRI/MRS. In addition, QAR gives information on whole brain distribution of compounds and intracerebral microdialysis only gives information about the unbound concentration in the brain area where the probe has been implanted. Both methods are suitable for routine use in animals while intracerebral microdialysis has been used to monitor patients with brain trauma [171]. In particular, PET and MRI/MRS can be applied in humans and can allow the estimation of the concentration–time course of compounds and their distribution in the whole brain. Moreover, the advantage of MRI/MRS is that a distinction can be made between metabolites and unbound and bound drug.

## 4 STRATEGIES FOR DRUG DELIVERY TO THE BRAIN

For many diseases of the brain, such as Alzheimer disease, Parkinson disease, stroke, depression, schizophrenia, epilepsy, and migraine headache, the drugs on the market are far from being ideal, let alone curative. A significant part of the problem is the poor BBB penetration of most of the drugs in development against neuronal targets for treatment of these disorders. This includes approximately 98% of the small molecules and nearly 100% of large molecules, such as recombinant proteins or gene-based medicines [12]. Therefore, much effort is put toward delivery and targeting of drugs to the brain [10, 12, 172]. Drug delivery to the brain can be achieved via several methods, including local invasive (direct injection/infusion) delivery, induction of enhanced permeability, and the application of global physiological targeting strategies. Apart from the delivery of small molecules (e.g., <500 Da) to the brain, the global delivery of large hydrophilic molecules, including enzymes, interference RNA (RNAi), and genes, present a serious problem.

### 4.1 Local Brain Delivery by Direct Injection/Infusion of Drugs into the Brain

Because of the lack of ability to target via the vascular route to the brain, invasive methods, like direct injection or infusion and convection enhanced delivery, have been applied to macromolecular drugs as well as to the delivery of viral vectors. These are reviewed in the following paragraphs.

***Direct Injection/Infusion of Macromolecular Drugs*** Intracerebral ventricular (ICV) or intrathecal drug infusion comprises direct injection/infusion of drug into the CSF. However, as mentioned before, CSF is completely drained into the venous circulation, and drugs still have to cross the ependymal brain–CSF barrier. This has shown to be feasible particularly for many small (mostly lipophilic [45, 172]) drugs following intraventricular administration. In addition, it was shown that compounds with MW < 5000 rapidly penetrate the ependyma but that penetration into the brain parenchyma was limited due to diffusion, tortuosity, transcapillary loss, cell uptake,

and binding [173, 174]. As a result, the infused drug has minimal access to the parenchyma by diffusion (Fig. 3C) [175]. In addition, it was shown that compounds with MW > 5000 very poorly penetrated the brain parenchyma even when they were administered intraventricularly. It was shown that following a 3–5 h infusion of $^{14}$C-inulin in the lateral ventricle of dogs, only an appreciable uptake in tissues lining the ventricle was found [176] and intraventricular injection of $^{125}$I-brain-derived neurotrophic growth factor [177] in rats resulted in a very poor uptake by the brain parenchyma. Very recently, it was elegantly demonstrated that the uptake of $^{125}$I-insulin-like growth factor-1 (MW 7.7 kDa) following intraventricular infusion was very poor by brain parenchyma and was rapidly cleared from the cerebrospinal fluid compartment mainly into blood [178]. In addition, these authors stated that under healthy conditions the CSF flow direction will probably be different from that under disease conditions and that further investigations are required in this respect. Furthermore, experimental data in human parkinsonian patients indicate that intraventricular GDNF infusion by a Medtronic device was biologically active but did not improve parkinsonism symptoms, possibly because GDNF (dimer with MW 33–45 kDa) did not sufficiently reach the target tissues [179].

Direct infusion into the brain has also been applied. The delivery of GDNF by direct infusion into the putamen of parkinsonian patients in one study showed

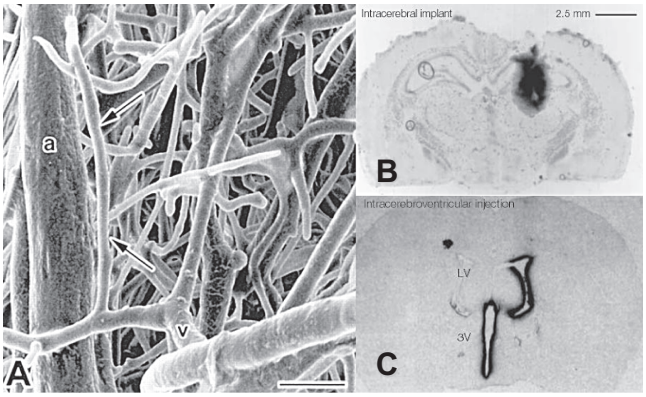


**FIGURE 3** Drug delivery via the vascular route will enable widespread distribution of the drug to each single neuron within the brain (note that the bar in panel A indicates a length of 25 μm, which is about the size of a single neuron). (A) Scanning electron micrograph of a vascular cast of a mouse brain (a, artery; v, vein). (Reprinted from Satomi et al., Cerebral vascular abnormalities in a murine model of hereditary hemorrhagic telangiectasia. *Stroke* 2003;34(33):783–789. © 2003, with permission from Lippincott Williams & Wilkins.) (B) Minimal diffusion of [$^{125}$I]-nerve growth factor (NGF) after intracerebral implantation of a biodegradable polymer (note that the bar indicates a length of 2.5 mm, which was also the size of the implant). (Reprinted from Krewson et al., Distribution of nerve growth factor following direct delivery to brain interstitium. *Brain Res* 1995;680(1–2):196–206. © 1995, with permission from Elsevier.) (C) ICV injection of [$^{125}$I]-brain-derived neurotrophic factor (BDNF). (Reprinted from Yan et al., Distribution of intracerebral ventricularly administered neurotrophins in rat brain and its correlation with trk receptor expression. *Exp Neurol* 1994;127(1):23–36. © 1994, with permission from Elsevier.) Note that the neurotrophin does not distribute into the brain beyond the ipsilateral ependymal surface.

general improvement over time [180] while a Phase II clinical trial could only demonstrate "potential efficacy" [181]. Indeed, Amgen decided to confirm its previous decision to stop a Phase II clinical trial with GDNF infusions into the putamen of 48 patients because "scientific results indicated that allowing patients to continue treatment could potentially cause permanent harm, complicating an already devastating disease" [182].

A variant of this is *convection enhanced drug delivery*, where a positive pressure infusion in brain parenchyma has been applied to increase drug uptake [183–185]. In addition, uptake was significantly improved when the viscosity of the infusate was increased [183], which seems to be caused by a similar mechanism (osmotic opening) as has been applied by Kroll and Neuwelt [186]. Good results have been shown by large volume infusion of naked DNA in the tail vein or legs where temporarily blood flow was inhibited; however, this is not applicable to the brain [187, 188].

Furthermore, brain implants have been used for the local delivery of drugs to the brain (Fig. 3B). Indeed, it has been shown that polymeric implants (poly(ethylene-*co*-vinyl acetate)) containing $^{125}$I-nerve-growth factor into rat brain were effective for local delivery, but unfortunately not for global delivery, of the drug [189].

In conclusion, because of the fenestrated capillaries in the CPs and their total surface area that is comparable to the BBB, drug transport into the CSF could be substantial. However, apart from the need of direct injection into brain parenchyma or infusion into the ventricles, for drugs that have to reach their target by passive diffusion, the direction of the CSF flow in the ventricles, the presence of the ependymal barrier (particularly for compounds with MW > 5000), and the direction of the flow of the interstitial fluid in the brain parenchyma are major obstacles for an effective global penetration into the brain.

***Direct Injection/Infusion of Viral Vectors*** Gene therapy [190–193] is meant to deliver genetic material with a therapeutic function encoding proteins (e.g., enzymes) or siRNA/shRNA to somatic cells [194]. In addition, gene therapy could provide a long-term effect following one single administration. However, genes are hydrophilic and charged and are large molecules that cannot pass cell membranes and tight cellular layers like the BBB. Therefore, means have to be found and developed to transport genes to the desired site of action.

Viruses have evolved over millions of years to obtain optimal mechanisms for gene delivery to host cells, which makes them applicable as a biological vector system to deliver genetic material to brain cells. There is a broad range of viral vectors available but the most commonly used are adeno-associated virus (AAV) vectors and lentivirus (LV) vectors. Important issues in viral gene delivery are stable transgene expression, limited immunogenicity, the induction of an inflammatory response, the (lack of) cell specific targeting efficiency, safety, toxicity, and the need for packaging cell lines [195].

Brain tropism (e.g., brain targeting) varies from virus to virus and has been shown not to be sufficient to pass the BBB following intravenous administration; therefore, the virus has to be injected directly into the brain. Moreover, various viral vectors have affinities for specific brain cells [190–196]. This could be improved if viral vectors could be produced with a homing/targeting device at their surface. Furthermore, the selectivity of gene transcription in the brain can be strongly enhanced by

application of brain selective promoters like the myelin basic protein (MBP) promoter [197], the neuron-specific enolase (NSE) promoter [198–200], the platelet-derived growth factor-β (PDGB-β) promoter [201, 202], the glial fibrillary acid protein (GFAP) promoter [203], and synthetic promoters like the CMV plus human growth hormone first intron enhancer and the CMV-chicken β-actin promoter [204, 205]. Moreover, the addition of a posttranscriptional regulatory element of the woodchuck hepatitis virus (WPRE) has enhanced the expression of GFP [206, 207]. Similarly, application of the replication origin (oriP) and Epstein–Barr nuclear antigen (EBNA-1) elements [208] can enhance episomal transcription of plasmids [209]. However, it has been indicated that application of such elements could induce tumorgeneity, which makes them not readily applicable in humans [210].

The Tet and Cre-Lox systems are other regulatory systems that have been applied, particularly for viruses that integrate their genetic material into the genome [211]. The Tet system can upregulate the expression of GFP in rat brain in the absence of doxycycline [212] while the Cre-Lox system allows a conditional but irreversible gene modification [213].

The *adeno-associated virus* (AAV) is a human parvovirus and contains a ssDNA. There are approximately 35 known serotypes. Several of them have been engineered into recombinant viral vectors from which the rAAV1, rAAV2, and rAAV5 are most widely used to deliver genes to the brain [193, 214–217]. In particular, these serotypes are most effective in the transduction of neurons [215, 218]. The attractive feature of the virus is that its replication is dependent on coinfection with a lytic helper virus, which precludes the reversion into a replication competent virus [219].

rAAV vectors are capsid virions composed of a nucleic acid genome surrounded by a proteinaceous shell. The composition of the capsid determines which cells they bind to and may enter. They enter the cell following binding to heparin sulfate proteoglycan (HSPG) receptors. Two other coreceptors seem to be helpful in this and finally the dsDNA enters the nucleus through the pore [220, 221].

Once in the cell, the rAAV genome is maintained as an extrachromosomal element or episomal element [190, 222–224], which eliminates the risk of proto-oncogene activation or endogenous gene inactivation, which may occur following insertion into the host genome [223]. Nevertheless, random integration of rAAV material does irregularly occur [190, 223, 225]. The episome might replicate along with the endogenous cellular DNA but can also be maintained as a nonreplicating extrachromosomal element [190]. In nondividing cells, the episome can be maintained as a stable element, resulting in long-term transgene expression in cells [190, 222–224, 226]. In addition, a safe and long-term correction of genetic diseases in animal models following a single administration has been demonstrated [191]. These viruses can be produced under current Good Manufacturing Practices (cGMPs) as outlined in the Code of Federal Regulations (21CFR) in the United States [191]. This all guarantees that pure and high-quality rAAV vectors can be produced for human use.

Expression of rAAV2 delivered vectors in animals has been shown for brain-derived neurotrophic factor (BDNF) in atrophy of spinal neurons [227]; and glial cell line mediated neurotrophic factor (GDNF), tyrosine hydroxylase (TH), aromatic L-amino acid decarboxylase (AADC), and guanosine triphosphate cyclohydrolase (GCH) in Parkinson disease [228–234]. In addition, rAAV vectors are

effective in the treatment of lysosomal storage diseases like mucopolysaccharidosis type I, II, and VII [235] and Niemann–Pick A [236]. In addition, it was demonstrated in mice that by targeting the deep cerebellar nuclei with AAV1 vectors capable of axonal transport, it was possible to get a widespread distribution throughout the cerebellum [236].

The major problem in delivering rAAV vectors to the brain is to get global brain transduction. This holds particularly for disorders caused by single-gene mutations such as lysosomal storage disorders and leukodystrophies, where a global brain transduction is desirable. Some vectors have been shown to have an enhanced neurotropism when expressing peptides that mimic binding domains for cytoplasmic dynein or NMDA receptors [237]. In addition, other viral vectors have been constructed that express capsid proteins that function as homing devices for targeted delivery of genes [238].

Another disadvantage is the occurrence of immune responses when animals have been treated with rAAV vectors, since this has resulted in killing of the transduced cells by cytotoxic T cells [190, 223, 225, 239, 240]. It was argued that this was caused by a slow uncoating of the rAAV in contrast to rapid uncoating [241, 242]. In addition, a humoral response can be elicited, which results in the production of antibodies. Moreover, memory cells are produced that will boost a strong humoral response when the same antigen is detected again. In this way, memory cells prevent the possibility of vector readministration [223].

Another feature is the packaging capacity of viruses. For rAAVs this is rather small (~4.5 kb) [190, 225, 243]. However, this may be increased up to 36 kb [244] by using (gutless) vectors, where as many viral genes as possible are deleted, only using those for the packaging sequence and sequences that define the beginning and end of the viral genome [190, 224, 240].

*Lentivirus* (LV) *vectors* are very promising vehicles to deliver therapeutic genes into the brain [190, 193, 194]. LV vectors are capsid virions surrounded by a lipid bilayer envelope [190]. LV vectors enter the cell following binding to HSPG; following membrane fusion, the virus moves along the microtubule by dynein-mediated transport. Finally, the dsDNA enters the nucleus through the pore [190]. LV vectors have a moderate packaging capacity with a maximum of 8 kb. LV vectors can be used to transduce a broad range of dividing as well as nondividing cells. In contrast to the rAAV vectors, they integrate into the host genome [196] and should therefore provide a longer and more stable transgene expression. However, they could also activate a proto-oncogene [190, 223, 224, 245] but since most CNS cells are terminally differentiated, the risk of tumorigenesis seems to be diminished [190, 225]. Nevertheless, the risk that vital genes are inactivated by random integration exists [190, 223, 245].

Furthermore, LV vectors seem to have a low immunogenicity. This seems to be caused by the fact that the genomes of these vectors do not encode any viral proteins. Indeed, low immunogenicity and prolonged transgene expression in the presence of preexisting lentiviral immunity was found, which should be encouraging for future use of these vectors in brain gene therapy [246].

Although viral delivery systems enter the cell by receptor-mediated endocytosis, this cell entry is not selective/specific since they can enter various cells, including brain cells, because of the lack of the expression of a selective/specific homing device. Therefore, rAAV and LV vectors can only effectively be delivered to the

brain by direct local injection/infusion. However, the diffusion from the site of administration in the brain is minimal; therefore, the region of cells that is transduced is restricted to a few millimeters from this site [190, 224]. To compensate for poor brain diffusion, usually the vector is administered by injections at multiple sites to increase the area of vector delivery. However, injection is associated with injury at the injection site, which could evoke a damaging response.

LV vectors have already been applied for gene therapy of various (brain) diseases like lysosomal storage diseases, Sandhoff disease and mucopolysaccharidosis I [247, 248], Parkinson disease [249, 250], ALS [251], epilepsy [252], prion disease [253], brain tumors [254], and the delivery of growth factors [250, 255].

In summary, one can say that today viral gene delivery to the brain can be efficient. Many applications have been demonstrated for brain gene therapy by rAAV and LV vectors, showing an increasing interest in the use of LV vectors. Critical points in this respect are global versus local delivery including the ability to reach the desired site of action (targeting efficiency). This may be improved considerably when homing/targeting devices are expressed at the outer surface of the virus particle [238], which will allow the vascular and global delivery of these systems to the CNS, thereby avoiding damage caused by direct injection/infusion. The selectivity of delivery can be further enhanced when tissue-selective/specific promoters are included in the vectors. rAAV vectors stay mainly episomal, while LV vectors integrate into the genome with the possibility to activate proto-oncogenes. Furthermore, immunological reactions may occur, but these seem to be reduced following application of LV vectors. In addition, rAAV and LV vectors have a rather small to moderate packaging capacity and can safely be produced under well controlled conditions.

With respect to invasive brain drug delivery strategies, one may conclude that these can be effective for local delivery (e.g., tumors) but not for the administration of drugs against more widespread diseases, including diffuse tumors and metastases. Therefore, since every neuron has its own capillary, the only route for an effective global delivery of biopharmaceutical drugs to the brain is the vascular route by application of nonviral systems. This is discussed further in the next section.

### 4.2 Global Brain Delivery by Vascular Nonviral Drug Administration

The advantage of the vascular route is the widespread transport by blood of the infused drug across the whole brain [256]. As has been mentioned, approximately each neuron has its own brain capillary for oxygen supply as well as the supply of other nutrients (see also Fig. 3A). This means that the vascular route is a very promising one for drug targeting and delivery to the whole brain. However, the main problem for drug transport to the brain is BBB passage.

***Drug Transport Across the BBB*** There are various possibilities for drug transport across the BBB as shown in Fig. 4. Passive diffusion depends on lipophilicity and molecular weight [257]. Furthermore, the ability of a compound to form hydrogen bonds will limit its diffusion through the BBB [258]. In general, Lipinski's rule of five, as well as Abraham's equation, can be used to predict the passive transport of a drug molecule across the BBB [259–261]. Transport of hydrophilic compounds via the paracellular route is limited, while lipophilic drugs smaller than 400–600 Da may freely enter the brain via the transcellular route.

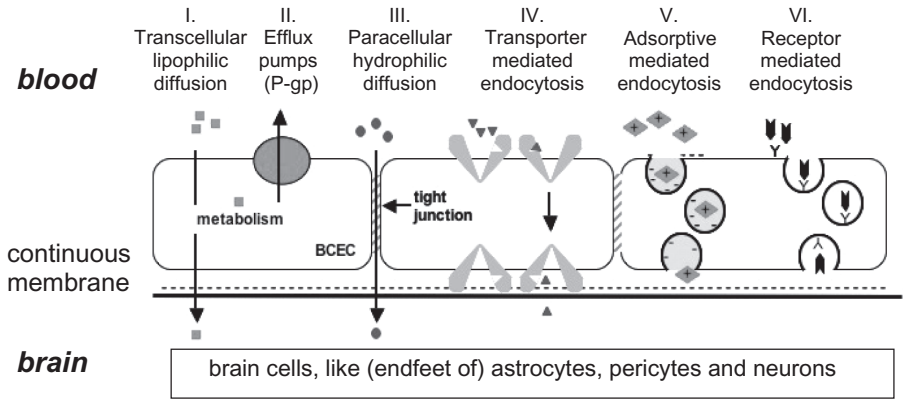


**FIGURE 4** The various transport processes that may occur at the blood–brain barrier. (Adapted from Ref. 297.)

Active transport systems can be divided into absorptive-, carrier-, or receptor-mediated transcytosis. Absorptive-mediated transcytosis is initiated by the binding of polycationic substances (such as most cell-penetrating peptides) to negative charges on the plasma membrane [262, 263]. This process does not involve specific plasma membrane receptors. Upon binding of the cationic compound to the plasma membrane, endocytosis occurs, followed by the formation of endosomes. However, vesicular transport is actively downregulated in the BBB to protect the brain from nonspecific exposure to polycationic compounds. Therefore, forcing drugs to enter the brain by absorptive-mediated transcytosis goes against the neuroprotective barrier function, as demonstrated for anionic and cationic nanoparticles that disrupt the BBB [264].

Carrier-mediated transcytosis is used for the delivery of nutrients, such as glucose, amino acids, and purine bases, to the brain [265–267]. At least eight different nutrient transport systems have been identified, each transporting a group of nutrients of the same structure. Carrier-mediated transcytosis is substrate selective and the transport rate is dependent on the degree of occupation of the carrier [267, 268]. Therefore, only drugs that closely mimic the endogenous carrier substrates will be taken up and transported into the brain.

In contrast, receptor-mediated transcytosis enables larger molecules, such as peptides, proteins, and genes, to specifically enter the brain. Classical examples of receptors involved in receptor-mediated transcytosis are the insulin receptor [269], the transferrin receptor [270, 271], and the transporters for low density lipoprotein [272], leptin [273], and insulin-like growth factors [274].

Besides many influx mechanisms, several efflux mechanisms exist at the BBB as well. The best known is P-glycoprotein (P-gp) [33–37, 275, 276]. P-gp is a transmembrane protein, located at the apical membrane of the BCEC. It has a high affinity for a wide range of cationic and lipophilic compounds and therefore limits transport to the brain of many drugs including cytotoxic anticancer drugs, antibiotics, hormones, and HIV protease inhibitors [277]. Other multidrug resistance (MDR) efflux mechanisms at the BBB include the MDR related protein (MRP), such as MRP 1, 2, 5, and 6 [278]. In addition, as mentioned before (see Section 2.1), many other

transporters are present at the BBB, like the organic anion transporter (influx and efflux), the organic cation transport system (influx), and the nucleoside transporter system (influx) [36, 37].

In conclusion, research over the years has shown that the BBB is a dynamic organ, which combines restricted diffusion to the brain for endogenous and exogenous compounds with specialized transport mechanisms for essential nutrients.

***Global Brain Delivery by Enhanced Passive Vascular Drug Delivery*** Large molecules (e.g., drug conjugates) or drug delivery systems (nanosized carrier systems including polymers, emulsions, micelles, liposomes, and nanoparticles) may reach the brain by passive targeting similar to small molecules, meaning that their distribution is mainly determined by physicochemical and physiological conditions [279–281]. Mostly this results in a very poor distribution into the brain. However, enhanced distribution may be obtained when the permeability of blood vessels in the target tissue has been increased due to disease conditions such as in tumors. Such "enhanced permeability and retention" has been observed in many cases [282, 283], showing increased accumulation of drug in those (mostly peripheral) areas. Unfortunately, drug delivery to the brain by enhanced permeability and retention is not effective in treating human brain diseases.

Enhanced drug delivery to the brain has been achieved in various other ways also. One is the osmotic disruption/shrinking of the BBB by intracarotid administration of a hypertonic mannitol solution. By subsequent administration of drug, substantial increased concentrations can be achieved in brain or tumor tissue. However, because the BBB is temporarily opened under such conditions, which has to be done under anesthesia, it has the disadvantage also that neurons may be damaged permanently due to unwanted blood components entering the brain [284]. Therefore, this procedure is only applicable in life-threatening brain diseases (e.g., tumors) [186].

A less "aggressive" approach seems to be the intracarotid administration of alkylglycerol [285], which enhances mainly drug transport by the paracellular route. In rats with brain implanted $C_6$-glioma, increased tissue fluorescence was found in both tumor tissue and brain surrounding the tumor following coadministration of FITC-dextran (40 kDa). Similar data on brain tissue were observed following coadministration of methotrexate to nude mice. However, anesthesia is needed to perform this type of administration, which limits its applicability to life-threatening brain diseases like brain tumors. In addition, no human data are available yet.

Another procedure is the application of a bradykinin-analogue (RMP-7) that has been shown to increase BBB permeability by opening of the tight junctions via a receptor-mediated mechanism (Ceroport®), which resulted in an enhanced uptake of carboplatin in $C_6$-glioma implanted in rat brain [286]. However, a Phase II study could not show improved carboplatin efficacy in combination with RMP-7 [287]. In addition, a Phase III study with RMP-7 and carboplatin with radiotherapy was stopped because of no increased efficacy.

At last, a general aspect about these agents is the possibility that besides opening of the BBB and the associated brain disturbances, application in the treatment of brain tumors could also cause an enhanced incidence of tumor dissemination via vascular spread of tumor cells [288].

Enhanced delivery can be achieved also by application of so-called protein transduction domains (PTDs) (for reviews see Refs. 289 and 290). These PTDs are typical

amino acid sequences that are capable of enhancing delivery of large molecules into cells. Typical examples are the Tat-PTD [291], the homeodomain of Antennapedia [292], SynB-vectors [293], and others [294, 295]. These peptides are basic molecules (and therefore cations) that are able to enhance protein uptake by cells mainly by increased adsorptive-mediated endocytosis. Toxicity profiles of the various PTDs are an important issue [296] but because this approach lacks cell or tissue selectivity it seems to be particularly suitable for local drug delivery. Clinical proof of principle has recently been demonstrated with SynB-vectors to enhance the transport of morphine-6-glucuronide to the brain (www.syntem.com).

Recently, therapeutic gene silencing of apolipoprotein B has been demonstrated in mice following administration of a high dose (50 mg/kg) of modified siRNAs [297]. This was achieved by coupling cholesterol to siRNA. Increased uptake by liver, heart, kidney, adipose, and lung tissue was seen but not by brain. The uptake into these tissues was mainly established by lipophilization of the siRNA by cholesterol and the resulting increase in plasma half-life from 6 min (naked siRNA) to 95 min (chol-siRNA).

A similar approach has been applied by coupling lipophilic and amphiphilic groups (stearoyl groups or Pluronic block copolymers) to polypeptides to enhance the uptake of these compounds by the brain. However, no selectivity in brain uptake has been demonstrated by applying such modifications [298].

In conclusion, all these approaches are able to enhance more or less the uptake into various tissues but lack selective/specific homing devices that are needed to target drugs to the brain and to selectively/specifically increase drug delivery to the brain.

***Global Brain Delivery by Active Vascular Receptor-Mediated (Nonviral) Drug Targeting*** Active physiological or disease-induced drug targeting strategies involve the application of a homing device or technology to apply endogenous transport mechanisms for site selective/specific delivery in the body [273, 275, 299]. With respect to the BBB, this particularly involves receptor-mediated transcytosis systems at the BBB to reach extracellular or intracellular targets in the brain. The advantage of active targeting is the increase of the amount of drug in the target tissue, thereby increasing the pharmacological response and reducing (peripheral) side effects. However, when the homing device is not specific, side effects can still occur. Interestingly, targeting efficiency can be increased when the target is disease-induced such as occurs for the diphtheria toxin receptor under inflammatory disease conditions (see following paragraphs). The focus of the following paragraphs is on active targeting by receptor-mediated transcytosis, and several examples of receptor-mediated transcytosis that have successfully been employed to target drugs to the brain are reviewed.

In general, receptor-mediated transcytosis occurs in three steps: receptor-mediated endocytosis of the compound at the luminal (blood) side, movement through the endothelial cytoplasm, and exocytosis of the drug at the abluminal (brain) side of the brain capillary endothelium [12]. Upon receptor–ligand internalization, chlathrin-coated vesicles are formed, which are approximately 120 nm in diameter [300, 301]. These vesicles may transport their content to the other side of the cell or go into a route leading to (protein) degradation. Indeed, at least two important routes for degrading proteins have been identified including the lysosomal and the

ubiquitin-proteasome route [302–304]. Nevertheless, protein delivery to the brain has been shown to be effective. Therefore, receptor-mediated transcytosis allows the specific delivery/targeting of larger drug molecules or drug-carrying particles (like liposomes or nanoparticles) to the brain.

*Transferrin Receptor* The most widely characterized receptor-mediated transcytosis system for the targeting of drugs to the brain is the transferrin receptor (TfR). TfR is a transmembrane glycoprotein consisting of two 90 kDa subunits. A disulfide bridge links these subunits and each subunit can bind one transferrin molecule [305]. The TfR is expressed mainly on hepatocytes, erythrocytes, intestinal cells, and monocytes, as well as on endothelial cells of the BBB [306, 307]. Furthermore, in the brain the TfR is expressed on choroid plexus epithelial cells and neurons [305]. The TfR mediates cellular uptake of iron bound to transferrin.

Drug targeting to the TfR can be achieved by either using the endogenous ligand, transferring, or using an antibody directed against the TfR (OX-26 anti-rat TfR). Each of these targeting vectors has its advantages and disadvantages. For transferrin, the *in vivo* application is limited due to high endogenous concentrations of transferrin in plasma and the likely overdosing with iron when one tries to displace the endogenous transferrin with exogenously applied transferrin-containing systems. However, recent studies in our group [309] have shown that liposomes tagged with transferrin are suitable for drug delivery to BBB endothelial cells *in vitro*, even in the presence of serum. OX-26 does not bind to the transferrin binding site and is therefore not displaced by endogenous transferrin.

The TfR is responsible for iron transport to the brain. So far, the intracellular trafficking of transferrin and OX-26 upon internalization via the TfR has not yet been elucidated. Some literature reports suggest transcytosis of transferrin across the BCECs, while others claim endocytosis of transferrin, followed by an intracellular release of iron and a subsequent return of apo-transferrin to the apical side of the BCEC [268, 309, 310]. Moos and Morgan [305] have shown that the transcytosis of iron exceeds the transcytosis of transferrin across the BBB, supporting the second theory. Furthermore, these authors have proposed a new theory in which the TfR–transferrin complex is transcytosed to the basolateral side of the BCEC, where transferrin remains bound to the TfR, but iron is released into the brain extracellular fluid [311]. Subsequently apo-transferrin bound to the TfR will recycle back to the apical side of the BBB. This theory is supported by data from Zhang and Pardridge [309], who found a 3.5-fold faster efflux from brain to blood of apo-transferrin than holo-transferrin. In addition, in a recent publication Deane et al. [312] illustrated that free iron is rapidly taken up by brain capillaries and subsequently released into the brain extracellular fluid and CSF at controlled moderate to slow rates.

The mechanism of transcytosis of OX-26 is not yet fully elucidated either. Pardridge and colleagues have shown efficient drug targeting and delivery to the brain *in vivo* by applying OX-26 (examples can be found in [265, 313–315]. In contrast, Broadwell et al. [316] have shown that both transferrin and OX-26 are able to cross the BBB, but that the transcytosis of transferrin is more efficient. Furthermore, Moos and Morgan [317] have shown that OX-26 mainly accumulates in the BCECs and not in the postcapillary compartment. In addition, iron deficiency did not increase OX-26 uptake in rats. Our data as well as literature reports show that iron deficiency causes an increase in TfR expression [306, 318, 319]. Therefore, it is

expected that the uptake of OX-26 would increase as well. The data by Moos and Morgan [317] suggest that OX-26 transcytosis might result from a high-affinity accumulation by the BCECs, followed by a nonspecific exocytosis at the basolateral side of the BCECs. In addition, these authors found a periventricular localization of OX-26, which suggests that OX-26 probably also is transported across the BCSFB.

Although the mechanism of transcytosis of transferrin and OX-26 may not yet be fully elucidated, it is important to realize that drug delivery to the brain via the TfR is efficient. By these means, vasoactive intestinal peptide (VIP), brain-derived neurotrophic factor (BDNF), basic fibroblast growth factor (bFGF), epidermal growth factor (EGF), and peptide nucleic acids, as well as pegylated immunoliposomes containing plasmid DNA encoding for β-galactosidase, tyrosine hydroxylase, and short hairpin RNAs, have all been made available to the brain [320]. However, OX-26 is an antibody against the rat TfR and does not bind to the human TfR, making it impossible to translate this technology to the clinic. Moreover, rat antibodies will cause immunogenic reactions in humans, unless they are humanized. The preparation of humanized or chimeric antibodies is difficult and in some cases this may lead to a loss of affinity for the target receptor [321]. In addition, one can argue that the administration of antibodies directed against such an important uptake mechanism involved in iron homeostasis poses a risk for human application.

Still, preferably, a targeting vector directed to the TfR would be small and nonimmunogenic and should initialize internalization of the TfR upon binding. Xu et al. [322] have used a single-chain antibody Fv fragment against the human TfR, which was tagged with a lipid anchor for insertion into a liposomal bilayer [323]. The molecular weight of this antibody fragment, including the lipid anchor, was approximately 30 kDa. In addition, Lee et al. [323] have used a phage-display technique to find small peptide ligands for the human TfR. They obtained 7 mer and 12 mer peptides that bind to a different binding site than transferrin and are internalized by the TfR. Although these small peptides can also exert immunogenic reactions in humans, they are promising ligands for drug targeting to the human TfR on the BBB.

*Insulin Receptor* Another widely characterized, classical, receptor-mediated transcytosis system for the targeting of drugs to the brain is the insulin receptor. Again, just as for the TfR system, Pardridge and colleagues have documented use of the insulin receptor for the targeted delivery of drugs to the brain.

The insulin receptor is a large 300 kDa protein and is a heterotetramer of two extracellular α and two transmembrane β subunits. Each β chain contains a tyrosine kinase activity in its cytosolic extension. The α and β subunits are coded by a single gene and are joined by disulfide bonds to form a cylinder structure. Primarily, insulin binds and changes shape of the receptor to form a tunnel to allow entry of molecules such as glucose into the cells. The insulin receptor is a tyrosine kinase receptor and induces a complex cellular response by phosphorylating proteins on their tyrosine residues. The binding of a single insulin molecule into a pocket created by the two α chains effects a conformational change in the insulin receptor, so that the β chains approximate one another, and carries out transphosphorylation on tyrosine residues. This autophosphorylation is necessary for the receptor to internalize into

endosomes. The endosomal system has been shown to be a site where insulin signaling is regulated, but also the degradation of endosomal insulin occurs there. Most of the insulin is degraded, but less so in endothelial cells [324], whereas the receptors are largely recycled to the cell surface. Endocytosis is not necessary for insulin action but probably is important for removing the insulin from the cell, so the target cell for insulin responds in a time-limited fashion to the hormone. This endocytosis mechanism of the insulin receptor has been exploited for the targeting of drugs to the brain.

As for transferrin, the *in vivo* application of insulin as the carrier protein is limited, mainly due to the high concentrations of insulin needed and the resulting lethal overdosing with insulin. Therefore, drug or gene delivery, for instance, to rhesus monkeys is performed with the murine 83–14 MAb that binds to the exofacial epitope on the α subunit of the human insulin receptor. The MAb has a BBB permeability surface area (PS) product in the primate that is ninefold greater than murine MAbs to the human TfR [325]. By using this MAb, Pardridge and co-workers have successfully made available to the brain of primates radiolabeled amyloid-β peptide$_{1\text{-}40}$ (A$\beta_{1\text{-}40}$), serving as a diagnostic probe for Alzheimer disease, and pegylated immunoliposomes containing plasmid DNA encoding for β-galactosidase [325].

Unfortunately, the 83-14 MAb cannot be used in humans owing to immunogenic reactions to this mouse protein. However, genetically engineered, effective, forms of the MAb have now been produced, which may allow for drug and gene delivery to the human brain [326]. Still, one can argue that the administration of antibodies directed against such an important mechanism involved in glucose homeostasis poses a risk for human application.

*LRP1 and LRP2 Receptors* During the past few years, the LRP1 and LRP2 (also known as megalin or glycoprotein 330) receptors have been exploited to target drugs to the brain in a similar fashion as the transferrin and insulin receptors. Both LRP1 and LRP2 receptors belong to the structurally closely related cell surface LDL receptor gene family. Both receptors are multifunctional, mutiligand scavenger and signaling receptors. A large number of substrates are shared between the two receptors, like lipoprotein lipase (LPL), $\alpha_2$-macroglobulin ($\alpha_2$M), receptor-associated protein (RAP), lactoferrin, tissue- and urokinase-type plasminogen activator (tPA/uPA), plasminogen activator inhibitor (PAI-1), and tPA/Upa:PAI-1 complexes. More specific ligands for the LRP1 receptor are melanotransferrin (or P97), thrombospondin 1 and 2, hepatic lipase, factor VIIa / tissue-factor pathway inhibitor (TFPI), factor VIIIa, factor IXa, A$\beta_{1\text{-}40}$, amyloid-β precursor protein (APP), C1 inhibitor, complement C3, apolipoprotein E (apo E), *Pseudomonas* exotoxin A, HIV-1 Tat protein, rhinovirus, matrix metalloproteinase 9 (MMP-9), MMP-13 (collagenase-3), sphingolipid activator protein (SAP), pregnancy zone protein, antithrombin III, heparin cofactor II, $\alpha_1$-antitrypsin, heat shock protein 96 (HSP-96), and platelet-derived growth factor (PDGF, mainly involved in signaling) [327–331], where apolipoprotein J (apo J, or clusterin), Aβ bound to apo J and apo E, aprotinin, and very low density lipoprotein (VLDL) are more specific for the LRP2 receptor [332, 333].

The group of Richard Béliveau first reported that melanotransferrin/P97 was actively transcytosed across the BBB and suggested that this was mediated by the

LRP1 receptor [331]. Melanotransferrin is a membrane-bound transferrin homolog that can also exist in a soluble form and is highly expressed on melanoma cells compared to normal melanocytes. Intravenously applied melanotransferrin delivers the majority of its bound iron to the liver and kidney, where only a small part was taken up by the brain [334]. After conjugation to melanotransferrin, the group of Béliveau was able to successfully deliver doxorubicin to brain tumors in animal studies [335, 336]. This melanotransferrin-mediated drug targeting technology (now designated NeuroTrans™) is currently further developed by BioMarin Pharmaceuticals Inc. (Novato, CA, USA) for the delivery of enzyme replacement therapies to the brain. Interestingly, together with researchers from BioMarin, Pan et al. [337] recently reported on the efficient transfer of RAP across the BBB by means of the LRP1/LRP2 receptors, suggesting a novel means of protein-based drug delivery to the brain. RAP is a 39 kDa protein that functions as a specialized endoplasmic reticulum chaperone assisting in the folding and trafficking of members of the LDL receptor family. In yet unpublished results, the group of Béliveau have now filed a patent application on the use of the LRP2-specific ligand aprotinin, and more specifically on functional derivatives thereof (e.g., angio-pep1), thereby providing a noninvasive and flexible method and a carrier for transporting a compound or drug across the BBB [338]. Aprotinin (Trasylol®) is known as a potent inhibitor of serine proteases such as trypsin, plasmin, and tissue and plasma kallikrein and is the only pharmacologic treatment approved by the U.S. Food and Drug Administration to reduce blood transfusion in coronary artery bypass grafting [339].

In addition to being a tumor marker protein, melanotransferrin is also associated with brain lesions in Alzheimer disease and is a potential marker of the disorder [340]. In addition, the proposed receptor for melanotransferrin, LRP1, has genetically been linked to Alzheimer disease and may influence APP processing and metabolism and Aβ uptake by neurons through $\alpha_2$M ($\alpha_2$M is one of the Aβ carrier proteins next to, e.g., apo E, apo J, transthyretin, and albumin) [332, 341]. Furthermore, a close relationship with RAGE (receptor for advanced glycation end products) in shuttling Aβ across the BBB has been described [332, 341]. In addition, the LRP2 receptor has also been described to mediate the uptake of Aβ complexed to apo J and apo E across the BBB [333, 342, 343]. This complex interaction with Alzheimer disease makes the safety of the use of the LRP1/LRP2 receptors for the targeting of drugs to the brain difficult to predict in human application; especially when one considers that the complex signaling function of these receptors is included in the assessment (e.g., the control of permeability of the BBB, vascular tone, and the expression of MMPs [327]), as well as the fact that both receptors are critically involved in the coagulation-fibrinolysis system. On top of that, melanotransferrin was also reported to be directly involved in the activation of plasminogen [344], and high plasma concentrations of melanotransferrin are needed to deliver drugs to the brain, resulting perhaps in dose limitations because of the high iron load in the body.

The same line of reasoning for the interactions at the level of the uptake receptors may apply to the use of RAP and aprotinin (derivatives). On the other hand, the latter has already been successfully applied to humans usually without severe side effects, indeed making the peptide derivatives potentially safe drug carriers. As for RAP, though, no results on the efficacy or capacity of the aprotinin peptides as carrier for drugs have been made available yet.

*Diphtheria Toxin Receptor* Recently, our group has identified a novel human applicable carrier protein (known as CRM197) for the targeted delivery of conjugated proteins across the BBB [345, 346]. Uniquely, CRM197 has already been used as a safe and effective carrier protein in human vaccines for a long time [347] and recently also as a systemically active therapeutic protein in anticancer trials [348]. This has resulted in a large body of prior knowledge on the carrier protein, including its transport receptor and mechanism of action, receptor binding domain, conjugation and manufacturing process, and kinetic and safety profile both in animals and humans. CRM197 delivers drugs across the BBB by the well characterized, safe, and effective mechanism called receptor-mediated transcytosis. From the literature, it was already known that CRM197 uses the membrane-bound precursor of heparin-binding epidermal growth factor-like growth factor (HB-EGF) as its transport receptor [349]. This precursor is also known as the diphtheria toxin receptor (DTR). In fact, CRM197 is a nontoxic mutant of diphtheria toxin. Membrane-bound HB-EGF is constitutively expressed on the BBB, neurons, and glial cells [350]. Moreover, HB-EGF expression is strongly upregulated on cerebral blood vessels by ischemic stroke and in gliomas [351, 352], which may lead to a site-selective improvement of the therapeutic efficacy of the targeted drugs in the brain.

By means of a dynamic cell culture model of the BBB, our group was able to demonstrate the functional expression of the DTR, safety of the CRM197 carrier protein, and specific transport efficacy of CRM197 carrier protein conjugated to a 40 kDa enzyme (horseradish peroxidase, HRP, serving as a "model" protein drug) and DTR-targeted pegylated liposomes containing HRP. In addition, the *in vivo* proof-of-principle with this novel brain drug targeting technology was demonstrated by the specific brain uptake of DTR-targeted HRP in guinea pigs [345, 346].

Although HB-EGF is expressed in species including human, monkey, rat, and mouse with a similar tissue distribution, only rats and mice are resistant to diphtheria toxin because of an amino acid substitution in the receptor-binding domain on HB-EGF that reduces binding of diphtheria toxin to rodent HB-EGF [346, 353]. Fortunately, a transgenic mouse (conditionally) expressing the human DTR was recently generated by Cha and co-workers [354], allowing researchers to specifically study the brain drug delivery technology in mice as well.

Another known complication of the bacterial CRM197 protein is that neutralizing antibodies against diphtheria toxin may develop or may already be present in serum of the recipient because of earlier vaccinations, thereby reducing the efficacy of the drug delivery system. There are, however, several lines of evidence that such an immune response to CRM197 can occur, but, most importantly, that it is not really a problem in the clinic, at least not for the treatment of acute indications. In fact, the clinical studies performed by Buzzi et al. [348] indicate that preexisting levels of neutralizing antibodies were rather decreased 30 days after repeated treatment with CRM197.

Another interesting aspect of the DTR is the fact that this receptor is strongly upregulated under inflammatory disease conditions [346] such as occur in many brain diseases like Alzheimer disease, Parkinson disease, multiple sclerosis, ischemia, encephalitis, epilepsy, and tumors [351]. On the one hand, such a situation may enhance the therapeutic effect by disease-induced targeting. On the other hand, this may be used to image (inflammatory) disease areas in the brain. For instance, it has

already been shown that the DTR is upregulated in the brain following seizures [355]. By applying CRM197-coated liposomes loaded with an MRI-enhancing agent (e.g., gadolinium), it would be possible to image such disease areas, as has been done for the glucose receptor [356]. In particular, this would be helpful to investigate the extent of damage/recovery and disease progression/therapy in time of many brain diseases, particularly epilepsy, ischemia, and brain tumors.

## 5 CONCLUSION

There is still an unmet need for the treatment of brain diseases. In addition to the presently applied small molecular drugs, there are many good biopharmaceutical drugs that, unfortunately, cannot enter the brain in sufficient quantities to be effective. Therefore, (new) technologies have to be applied and developed to accomplish this. In this chapter we have tried to give insight into the present *in vitro* [357] and *in vivo* [358] techniques to study and improve drug absorption and distribution of biopharmaceutical drugs (e.g., proteins, RNAi, and genes) to the brain.

*In vitro* techniques to study BBB drug transport and BBB functionality should use an endothelial/astrocyte coculture system for establishing a tight and functional BBB system. Some systems with BBB barrier enhancers (e.g., cAMP, phosphodiesterase inhibitors, glucocorticoids) could be used to estimate BBB drug transport but may lack BBB functionality. From the *in vivo* methods, the noninvasive MRI techniques seem to be very promising and impressive advancements have been made to image (diseased) brain by application of MRI contrast-enhancing agents. The development of technologies to actively target such agents to the (diseased) brain will open a wide area for diagnostic investigations and the possibility to monitor disease progression in the brain.

With respect to the therapeutic treatment of brain diseases, the only human applied technology to deliver (biopharmaceutical) drugs to the brain is by viral delivery. Up to now viral delivery to the brain has to be performed by direct injection/infusion into the brain because of insufficient brain targeting due to the lack of a homing device, which makes it only sufficient for local delivery. Therefore, the most promising technology in our opinion is active vascular nonviral receptor-mediated targeting. Vascular nonviral active targeting has been shown to be effective and selective for global delivery of drugs to the brain in several animals. By using antibodies against receptors (e.g., transferrin and insulin receptor), it has been demonstrated that proteins (enzymes), RNAi, and genes could be targeted to the brain. However, these applications cannot be transferred directly to the human situation: first, because the antibodies used in animals cannot (without humanization) directly be applied in humans; and second, antibodies are meant to stick to their counterpart, which means that administration of these will reduce the availability of receptors at the BBB and in the brain. Moreover, several of these receptors (insulin receptor and the multiligand LRP1 and LRP2 receptors) are involved in signaling too [27, 359]. Therefore, brain targeting by such antibodies may interfere with both the transport of endogenous ligands (e.g., insulin, transferrin) and brain signaling and cause downregulation of receptors; and therefore these antibodies should not be used for chronic administration. Furthermore, the melanotransferrin receptor seems to be involved in the release of plasmin [344] and therefore with the

blood clotting cascade, while the LRP1 receptor seems to be involved with apolipprotein J transport in a complex with Alzheimer's amyloid-β [343]. Interference with such transport systems poses a potential hazard. In this respect, the DTR seems to be a human applicable, safe, and effective uptake receptor for the targeting of drugs to the brain. Moreover, the DTR does not have an endogenous ligand, so there is no competition for transport. Furthermore, CRM197 is already safely applied to humans (illustrating that binding to the DTR *per se* does not result in serious side effects), where other carrier systems involve potential safety hazards in human application. However, even though specific brain uptake of DTR-targeted enzyme was already established in guinea pigs, the technology now awaits further *in vivo* validation in terms of kinetics of brain distribution and efficacy of targeted drugs in relevant disease models of the CNS.

## REFERENCES


1. Ehrlich P. *Das Sauerstoff-Bedurfnis des Organismus: eine farbenanalytische Studie*. Berlin: Hirschwald; 1885.
2. Goldman EE. Vitalfarbung am Zentralnervensystem. *Abh Preuss Akad Wiss Phys Math K1* 1913;1:1–60.
3. Abbott NJ. Dynamics of CNS barriers: evolution, differentiation, and modulation. *Cell Mol Neurobiol* 2005;25(1):5–23.
4. Begley DJ, Brightman MW. Structural and functional aspects of the blood–brain barrier. *Prog Drug Res* 2003;61:39–78.
5. Hawkins BT, Davis TP. The blood–brain barrier/neurovascular unit in health and disease. *Pharmacol Rev* 2005;57(2):173–185.
6. Abbott NJ. Inflammatory mediators and modulation of blood–brain barrier permeability. *Cell Mol Neurobiol* 2000;20(2):131–147.
7. Gaillard PJ, de Boer AG, Breimer DD. Pharmacological investigations on lipopolysaccharide-induced permeability changes in the blood–brain barrier *in vitro*. *Microvasc Res* 2003;65(1):24–31.
8. Pardridge WM. Molecular biology of the blood–brain barrier. *Mol Biotechnol* 2005;30(1): 57–70.
9. de Vries HE, Kuiper J, de Boer AG, van Berkel ThJC, Breimer DD. The role of the blood–brain barrier in neuroinflammatory diseases. *Pharmacol Rev* 1997;49(2):143–155.
10. Begley DJ. Delivery of therapeutic agents to the central nervous system: the problems and the possibilities. *Pharmacol Ther* 2004;104(1):29–45.
11. Pardridge WM. Blood–brain barrier biology and methodology. *J Neurovirol* 1999;5(6): 556–569.
12. Pardridge WM. Drug targeting, drug discovery, and brain drug development. In Pardridge WM, Ed. *Brain Drug Targeting—The Future of Brain Drug Development*. New York: Cambridge University Press; 2001, pp 1–12.
13. Keep RG, Jones HC. A morphometric study on the development of the lateral ventricle choroids plexus, choroids plexus capillaries and ventricular ependyma in the rat. *Brain Res Dev Brain Res* 1990;56:47–53.
14. Rubin LL, Staddon JM. The cell biology of the blood–brain barrier. *Annu Rev Neurosci* 1999;22:11–28.

15. Pardridge WM. Overview of blood–brain barrier transport biology and experimental methodologies. In Pardridge WM, Ed. *Peptide Drug Delivery to the Brain*. New York: Raven Press; 1991, pp 52–98.
16. Janzer RC, Raff MC. Astrocytes induce blood–brain barrier properties in endothelial cells. *Nature* 1987;325(6101):253–257.
17. Gaillard PJ, van der Sandt IC, Voorwinden LH, Vu D, Nielsen JL, de Boer AG, Breimer DD. Astrocytes increase the functional expression of P-glycoprotein in an *in vitro* model of the blood–brain barrier. *Pharm Res* 2000;17(10):1198–1205.
18. Lai CH, Kuo KH. The critical component to establish *in vitro* BBB model: pericyte. *Brain Res Brain Res Rev* 2005 [Epub ahead of print].
19. Abbott NJ. Astrocyte-endothelial interactions and blood–brain barrier permeability. *J Anat* 2002;200(6):629–638.
20. Brightman MW, Reese TS. Junctions between intimately apposed cell membranes in the vertebrate brain. *J Cell Biol* 1969;40(3):648–677.
21. Reese TS, Karnovsky MJ. Fine structural localization of a blood–brain barrier to exogenous peroxidase. *J Cell Biol* 1967;34(1):207–217.
22. Minn A, Ghersi-Egea J-F, Perrin R, Leinigner B, Siest G. Drug metabolizing enzyme in the brain and cerebral microvessels. *Brain Res Rev* 1991;16:65–82.
23. el-Bacha RS, Minn A. Drug metabolizing enzymes in cerebrovascular endothelial cells afford a metabolic protection to the brain. *Cell Mol Biol (Noisy-le-grand)* 1999;45(1): 15–23.
24. Ghersi-Egea JF, Leininger-Muller B, Cecchelli R, Fenstermacher JD. Blood–brain interfaces: relevance to cerebral drug metabolism. *Toxicol Lett* 1995;82–83:645–653.
25. Wekerle H. Immune protection of the brain—efficient and delicate. *J Infect Dis* 2002; 186(Suppl 2):S140–S144.
26. Biessels GJ, Bravenboer B, Gispen WH. Glucose, insulin and the brain: modulation of cognition and synaptic plasticityin health and disease: a preface. *Eur J Pharmacol* 2004; 490(1–3):1–4.
27. Herz J, Strickland DK. LRP: a multifunctional scavenger and signaling receptor. *J Clin Invest* 2001;108(6):779–784.
28. Debault LE, Cancilla PA. gamma-Glutamyl transpeptidase in isolated brain endothelial cells: induction by glial cells *in vitro*. *Science* 1980;207(4431):653–655.
29. Hendry SH, Jones EG, Beinfeld MC. Cholecystokinin-immunoreactive neurons in rat and monkey cerebral cortex make symmetric synapses and have intimate associations with blood vessels. *Proc Natl Acad Sci USA* 1983;80(8):2400–2404.
30. Balabanov R, Dore-Duffy P. Role of the CNS microvascular pericyte in the blood–brain barrier. *J Neurosci Res* 1998;53(6):637–644.
31. Kim JA, Tran ND, Li Z, Yang F, Zhou W, Fisher MJ. Brain endothelial hemostasis regulation by pericytes. *J Cereb Blood Flow Metab* 2005; Jul 13 [Epub ahead of print].
32. Cordon-Cardo CJ, O'Brien JP, Casals D, Rittman-Grauer L, Biedler JL, Melamed MR, Bertino JR. Multidrug-resistance gene P-glycoprotein is expressed by endothelial cells at blood–brain barrier sites. *Proc Natl Acad Sci USA* 1989;86:695–698.
33. Schinkel AH, Wagenaar E, van Deemter L, Mol CAAM, Borst P. Absence of the mdr1a P-glycoprotein in mice affects tissue distribution and pharmacokinetics of dexamethasone, digoxin, and cyclosporin A. *J Clin Invest* 1995;96(4):1698–1705.
34. Borst P, Evers R, Kool M, Wijnholds J. The multidrug resistance protein family. *Biochem Biophys Acta* 1999;1461:347–557.
35. Borst P, Evers R, Kool M, Wijnholds J. A family of drug transporters: the multidrug resistance-associated proteins. *J Natl Cancer Inst* 2000;92:1295–1302.

36. de Boer AG, van der Sandt IC, Gaillard PJ. The role of drug transporters at the blood–brain barrier. *Annu Rev Pharmacol Toxicol* 2003;43:629–656.
37. Lee G, Dallas S, Hong M, Bendayan R. Drug transporters in the central nervous system: brain barriers and brain parenchyma considerations. *Pharmacol Rev* 2001;53(4):569–596.
38. Loscher W, Potschka H. Role of drug efflux transporters in the brain for drug disposition and treatment of brain diseases. *Prog Neurobiol* 2005;76(1):22–76.
39. Abbott NJ. Physiology of the blood–brain barrier and its consequences for drug transport to the brain. International Congress Series 1277, *Proceedings of the Esteve Foundation Symposium*, Volume *11*. Amsterdam: Elsevier; 2005, pp 3–18.
40. Begley DJ. ABC transporters and the blood–brain barrier. *Curr Pharm Des* 2004;10(12): 1295–1312.
41. Lo EH, Singhal AB, Torchilin VP, Abbott NJ. Drug delivery to damaged brain. *Brain Res Brain Res Rev* 2001;38(1–2):140–148.
42. Johanson CE, Duncan JA, Stopa EG, Baird A. Enhanced prospects for drug delivery and brain targeting by the choroid-plexus-CSF route. *Pharm Res* 2005;22(7):1011–1037.
43. Emerich DF, Skinner SJM, Borlongan CV, Vasconcellos AF, Thanos CG. The choroids plexus in the rise, fall and repair of the brain. *BioEssays* 2005;27:262–274.
44. Strazielle N, Khuth ST, Ghersi-Egea J-F. Detoxification systems, passive and specific transport for drugs at the blood–CSF barrier in normal and pathological situations. *Adv Drug Deliv Res* 2004;56:1717–1740.
45. Shen DD, Artru AA, Adkinson KK. Principles and applicability of CSF sampling for the assessment of CNS drug delivery and pharmacodynamics. *Adv Drug Deliv Rev* 2004;56:1825–1857.
46. Chodobski A, Szmydynger-Chodobska J. Choroid plexus, target for polypeptides and site of their synthesis. *Microsc Res Tech* 2001;52:65–82.
47. Abbott NJ. Evidence for bulk flow of brain interstitial fluid: significance for physiology and pathology. *Neurochem Int* 2004;45(4):545–552.
48. Faraci FM, Kinzenbaw D, Heistad DD. Effect of endogenous vasopressin on blood flow to choroids plexus during hypoxia and intracranial hypertension. *Am J Physiol* 1994;266: H393–H398.
49. Joó F. The blood–brain barrier *in vitro*: ten years of research on microvessels isolated from the brain. *Neurochem Int* 1985;7:1–25.
50. Joó F. The cerebral microvessels in culture, an update. *J Neurochem* 1992;58:1–17.
51. Joó F. The blood–brain barrier *in vitro*: the second decade. *Neurochem Int* 1933;23(6):499–521.
52. Takakura Y, Audus KL, Borchardt RT. Blood–brain barrier: transport studies in isolated brain capillaries and in cultured brain endothelial cells. *Adv Pharmacol* 1991;22:137–165.
53. Eddy EP, Maleef BE, Hart TK, Smith PL. *In vitro* models to predict blood–brain barrier permeability. *Adv Drug Deliv Rev* 1997;23(1–3):185–198.
54. de Boer AG, Breimer DD. Reconstitution of the blood–brain barrier in cell culture for studies of drug transport and metabolism. *Adv Drug Deliv Rev* 1996;22:251–264.
55. Cecchelli R, Dehouck B, Descamps L, Fenart L, Buée-Scherrer V, Duhem C, Lundquist S, Rentfel M, Torpier G, Dehouck MP. *In vitro* model for evaluating drug transport across the blood–brain barrier. *Adv Drug Deliv Rev* 1999;36:165–178.
56. Pardridge WM. Blood–brain barrier biology and methodology. *J Neurovirol* 1999;5(6): 556–569.

57. Bradbury MWB. The blood–brain barrier. Transport across the cerebral endothelium. *Circ Res* 1985;57:213–222.
58. Gumbleton M, Audus KL. Progress and limitations in the use of *in vitro* cell cultures to serve as a permeability screen for the blood–brain barrier. *J Pharm Sci* 2001;90(11):1681–1698.
59. Cucullo L, Aumayr B, Rapp E, Janigro D. Drug delivery and *in vitro* models of the blood–brain barrier. *Curr Opin Drug Discov Dev* 2005;8(1):89–99.
60. Reichel A, Begley DJ, Abbott NJ. An overview of *in vitro* techniques for blood–brain barrier studies. *Methods Mol Med* 2003;89:307–324.
61. Weksler BB, Subileau EA, Perriere N, Charneau P, Holloway K, Leveque M, Tricoire-Leignel H, Nicotra A, Bourdoulous S, Turowski P, Male DK, Roux F, Greenwood J, Romero IA, Couraud PO. Blood–brain barrier-specific properties of a human adult Brain endothelial cell line. *FASEB J* 2005; [Epub ahead of print].
62. de Boer AG, Gaillard PJ, Breimer DD. The transference of results between blood–brain barrier cell culture systems. *Eur J Pharm Sci* 1999;8(1):1–4.
63. Youdim KA, Avdeef A, Abbott NJ. *In vitro* trans-monolayer permeability calculations: often forgotten assumptions. *Drug Discov Today* 2003;8(21):997–1003.
64. de Boer AG, Sutanto W, Eds. *Drug Transport Across the Blood–Brain Barrier: New Experimental Strategies*. Amsterdam: Harwood Scientific Publisher; 1997.
65. Audus KL, Borchardt RT. Characterization of an *in vitro* blood–brain barrier model system for studying drug transport and metabolism. *Pharm Res* 1986;3:81–87.
66. Audus KL, Rose JM, Wang W, Borchardt RT. In Pardridge W, Ed. *An Introduction to the Blood–Brain Barrier: Methodology and Biology*. New York: Cambridge University Press; 1998, p 86.
67. Dehouck M-P, Méresse S, Delorme P, Fruchart J-C, Cecchelli R. An easier, reproducible, and mass-production method to study the blood–brain barrier *in vitro*. *J Neurochem* 1990;54:1798–1801.
68. Rubin LL, Hall DE, Parter S, Barbu K, Cannon C, Horner HC, Janatpour M, Liaw CW, Manning K, Morales J, Tanner LI, Tomaselli KJ, Bard F. A cell culture model of the blood–brain barrier. *J Cell Biol* 1991;115(6):1725–1735.
69. Gaillard PJ, Voorwinden LH, Nielsen JL, Ivanov A, Atsumi R, Engman H, Ringbom C, de Boer AG, Breimer DD. Establishment and functional characterization of an *in vitro* model of the blood–brain barrier, comprising a co-culture of brain capillary endothelial cells and astrocytes. *Eur J Pharm Sci* 2001;12:215–222.
70. Janzer RC, Raff MC. Astrocytes induce blood–brain barrier properties in endothelial cells. *Nature* 1987;325:253–257.
71. Meresse S, Dehouck MP, Delorme P, Bensaid M, Tauber JP, Delbart C, Fruchart JC, Cecchelli R. Bovine brain endothelial cells express tight junctions and monoamine oxidase activity in long-term culture. *J Neurochem* 1989;53(5):1363–1371.
72. Gaillard PJ, van der Sandt ICJ, Voorwinden LH, Vu D, Nielsen JL, de Boer AG, Breimer DD. Astrocytes increase the functional expression of P-glycoprotein in an *in vitro* model of the blood–brain barrier. *Pharm Res* 2000;17(10):1198–1205.
73. Augustin HG, Kozian DH, Johnson RC. Differentiation of endothelial cells: analysis of the constitutive and activated endothelial cell phenotypes. *BioEssays* 1994;16(12):901–906.
74. Parkinson FE, Friesen J, Krizanac-Bengez L, Janigro D. Use of a three-dimensional *in vitro* model of the rat blood–brain barrier to assay nucleoside efflux from brain. *Brain Res* 2003;980(2):233–241.

75. Coisne C, Dehouck L, Faveeuw C, Delplace Y, Miller F, Landry C, Morissette C, Fenart L, Cecchelli R, Tremblay P, Dehouck B. Mouse syngenic *in vitro* blood–brain barrier model: a new tool to examine inflammatory events in cerebral endothelium. *Lab Invest* 2005;85(6):734–746.
76. Ramsauer M, Kunz J, Krause D, Dermietzel R. Regulation of a blood–brain barrier-specific enzyme expressed by cerebral pericytes (pericytic aminopeptidase N/pAPN) under cell culture conditions. *J Cerebral Blood Flow Metab* 1998;18:1270–1281.
77. Mukhtar M, Pomerantz RJ. Development of an *in vitro* blood–brain barrier model to study molecular neuropathogenesis and neurovirologic disorders induced by human immunodeficiency virus type 1 infection. *J Human Virol* 2000;3:324–334.
78. Tahti H, Nevala H, Toimela T. Refining *in vitro* neurotoxicity testing—the development of blood–brain barrier models. *Altern Lab Anim* 2003;31(3):273–276.
79. Hurst RD, Fritz IB. Properties of an immortalised vascular endothelial/glioma cell co-culture model of the blood–brain barrier. *J Cell. Physiol* 1996;167:81–88.
80. de Boer AG, Gaillard PJ. Unpublished results, 2005.
81. Tan KH, Dobbie MS, Felix RA, Barrand MA, Hurst RD. A comparison of the induction of immortalized endothelial cell impermeability by astrocytes. *Neuroreport* 2001;12(7):1329–1334.
82. Franke H, Galla H-J, Beuckmann CT. An improved low-permeability *in vitro*-model of the blood–brain barrier: transport studies on retinoids, sucrose, haloperidol, caffeine and mannitol. *Brain Res* 1999;818:65–71.
83. Sobue K, Yamamoto N, Yoneda K, Hodgson ME, Yamashiro K, Tsuruoka N, Tsuda T, Katsuya H, Miura Y, Asai K, Kato T. Induction of blood–brain barrier properties in immortalized bovine brain endothelial cells by astrocytic factors. *Neurosci Res* 1999;35(2):155–164.
84. Reuss B, Dono R, Unsicker K. Functions of fibroblast growth factor (FGF)-2 and FGF-5 in astroglial differentiation and blood–brain barrier permeability: evidence from mouse mutants. *J Neurosci* 2003;23(16):6404–6412.
85. de Boer AG, de Vries HE, Gaillard PJ, Breimer DD. In de Boer AG, Sutanto W, Eds. *Drug Transport Across the Blood–Brain Barrier: New Experimental Strategies*. Amsterdam: Harwood Scientific Publisher; 1997.
86. Garberg P. Alternatives to laboratory animals. *Altern Lab Anim* 1998;26(6):821–825.
87. Takakura Y, Audus KL, Borchardt RT. Blood–brain barrier: transport studies in isolated brain capillaries and in cultured brain endothelial cells. *Adv Pharmacol* 1991;22:137–165.
88. Gaillard PJ, de Boer AG. Relationship between permeability status of the blood–brain barrier and *in vitro* permeability coefficient of a drug. *Eur J Pharm Sci* 2000;12:95–102.
89. Prieto P, Blaauboer BJ, de Boer AG, Boveri M, Cecchelli R, Clemedson C, Coecke S, Forsby A, Galla HJ, Garberg P, Greenwood J, Price A, Tahti H, European Centre for the Validation of Alternative Methods. Blood–brain barrier *in vitro* models and their application in toxicology. The report and recommendations of ECVAM Workshop 49. *Altern Lab Anim* 2004;32(1):37–50.
90. Garberg P, Ball M, Borg N, Cecchelli R, Fenart L, Hurst RD, Lindmark T, Mabondzo A, Nilsson JE, Raub TJ, Stanimirovic D, Terasaki T, Oberg JO, Osterberg T. *In vitro* models for the blood–brain barrier. *Toxicol In Vitro* 2005;19(3):299–334.
91. Schinkel AH. P-glycoprotein, a gatekeeper in the blood–brain barrier. *Adv Drug Deliv Rev* 1999;36(2–3):179–194.

92. Tamai I, Tsuji A. Transporter-mediated permeation of drugs across the blood–brain barrier. *J Pharm Sci* 2000;89(11):1371–1388.
93. Jaehde U, Masereeuw R, de Boer AG, Fricker G, Nagelkerke JF, Vonderscher J, Breimer DD. Quantification and visualization of the transport of octreotide, a somatostatin analogue, across monolayers of cerebrovascular endothelial cells. *Pharm Res* 1994;11(3): 442–448.
94. Sorensen M, Steenberg B, Knipp GT, Wang W, Steffansen B, Frokjaer S, Borchardt RT. The effect of beta-turn structure on the permeation of peptides across monolayers of bovine brain microvessel endothelial cells. *Pharm Res* 1997;14(10):1341–1348.
95. Gaillard PJ, van der Meide PH, de Boer AG, Breimer DD. Glucocorticoid and type I interferon interactions at the blood–brain barrier: relevance for drug therapies for multiple sclerosis. *NeuroReport* 2001;12(10):2198–2193.
96. Fenart L, Buee-Scherrer V, Descamps L, Duhem C, Poullain MG, Cecchelli R, Dehouck MP. Inhibition of P-glycoprotein: rapid assessment of its implication in blood–brain barrier integrity and drug transport to the brain by an *in vitro* model of the blood–brain barrier. *Pharm Res* 1998;15(7):993–1000.
97. van der Sandt ICJ, Gaillard PJ, Voorwinden LH, de Boer AG, Breimer DD. P-glycoprotein inhibition leads to enhanced disruptive effects by anti-microtubule cytostatics at the *in vitro* blood–brain barrier. *Pharm Res* 2001;18(5):587–592.
98. Barrand MA, Robertson KJ, von Weikersthal SF. Comparisons of P-glycoprotein expression in isolated rat brain microvessels and in primary cultures of endothelial cells derived from microvasculature of rat brain, epididymal fat pad and from aorta. *FEBS Lett* 1995;374(2):179–183.
99. Zhang Y, Han H, Elmquist WF, Millen DW. Expression of various multidrug resistance-associated protein MRP homologues in brain microvessel endothelial cells. *Brain Res* 2000;876:148–153.
100. Durieu-Trautmann O, Fooignant-Chaverot N, Perdomo J, Gounon P, Strosberg AD, Couraud PO. Immortalization of brain capillary endothelial cells with maintenance of structural characteristics of the blood–brain barrier endothelium. *In Vitro Cell Dev Biol* 1991;27(A):771–778.
101. Sobue K, Yamamoto N, Yoneda K, Hodgson ME, Yamashiro K, Tsuroika N, Tsuda T, Katsuya H, Miuara Y, Asai K, Kato T. Induction of blood–brain barrier properties in immortalized bovine brain endothelial cells by astrocytic factors. *Neurosci Res* 1999;35: 155–164.
102. Stins MF, Prasadarao NV, Zhou J, Arditi M, Kim KS. Bovine brain microvascular endothelial cells transfected with SV40-large T antigen: development of an immortalized cell line to study pathophysiology of CNS disease. *In Vitro Cell Dev Biol Anim* 1997;33:243–247.
103. Kido Y, Tamai I, Okamoto M, Suzuki F, Tsuji A. Functional clarification of MCT1-mediated transport of monocarboxylic acids at the blood–brain barrier using *in vitro* cultured cells and *in vivo* BUI studies. *Pharm Res* 2000;17(1):55–62.
104. Terasaki T, Hosoya K. Conditionally immortalized cell lines as a new *in vitro* model for the study of barrier functions. *Biol Pharm Bull* 2001;24(2):111–118.
105. Roux F, Durieue-Trautmann O, Chaverot N, Claire M, Mailly P, Bourre JM, Strosberg AD, Couraud PO. Regulation of gamma-glutamyl transpeptidase and alkaline phosphatase activities in immortalized rat brain microvessel endothelial cells. *J Cell Physiol* 1994;159:101–113.
106. Lechardeur D, Schwartz B, Paulin D, Scherman D. Induction of blood–brain barrier differentiation in a rat brain-derived endothelial cell line. *Exp Cell Res* 1995;220(1): 161–170.

107. Mooradian DL, Diglio CA. Production of a transforming growth factor-beta-like growth factor by RSV-transformed rat cerebral microvascular endothelial cells. *Tumour Biol* 1991;12:171–181.

108. Regina A, Romero IA, Greenwood J, Adamson P, Bourre JM, Couraud PO, Roux F. Dexamethasone regulation of P-glycoprotein activity in an immortalized rat brain endothelial cell line, GPNT. *J Neurochem* 1999;73:1954–1963.

109. Pham YT, Regina A, Farinotti R, Couraud PO, Wainer IW, Roux R, Gimeniz F. Interactions of racemic mefloquine and its enantiomers with P-glycoprotein in an immortalised rat brain capillary endothelial cell line, GPNT. *Biochem Biophys Acta* 2000;1524:212–219.

110. Greenwood J, Pryce G, Devine L, Male DK, Dos Santos WLC, Calder VL, Adamson P. SV40 large T immortalised cell lines of the rat blood–brain and blood–retinal barriers retain their phenotypic and immunological characteristics. *J Neuroimmunol* 1996;71: 51–63.

111. Asaba H, Hosoya K, Takanaga H, Ohtsuki S, Tamura E, Takizawa T, Terasaki T. Blood–brain barrier is involved In the efflux transport of a neuroactive steroid, dehydroepiandrosterone sulphate, via organic anion transporting polypeptide 2. *J Neurochem* 2000;75:1907–1916.

112. Tatsuta T, Naito M, Oh-hara T, Sugawara I, Tsuruo T. Functional involvement of P-glycoprotein in blood–brain barrier. *J Biol Chem* 1992;267:20383–20393.

113. Wijsman JA, Shivers RR. Immortalized mouse brain endothelial cells are ultrastructurally similar to endothelial cells and respond to astrocyte-conditioned medium. *In Vitro Cell Dev Biol Anim* 1998;34:777–784.

114. Teifel M, Friedl P. Establishment of the permanent microvascular endothelial cell line PBMEC/C1–2 from porcine brains. *Exp Cell Res* 1996;228:50–57.

115. Muruganandam A, Herx LM, Monette R, Durkin JP, Stanimirovic DB. Development of immortalized human cerebromicrovascular endothelial cell line as an *in vitro* model of the human blood–brain barrier. *FASEB J* 1997;11:1187–1197.

116. Prudhomme JG, Sherman IW, Land KM, Moses AV, Stenglein S, Nelson JA. Studies of *Plasmodium falciparum* cytoadherence using immortalized human brain capillary endothelial cells. *Int J Parasitol* 1996;26:647–655.

117. Xiao L, Yang C, Dorovinin-Zis K, Tandon NN, Ades E, Lal AA, Udhayakumar V. *Plasmodium falciparum*: involvement of additional receptors in the cytoadherence of infected erythrocytes to microvascular endothelial cells. *Exp Parasitol* 1996;84:42–55.

118. Beuckmann CT, Dernbach K, Hakvoort A, Galla HJ. A new astrocytic cell line which is able to induce a blood–brain barrier property in cultured brain capillary endothelial cells. *Cytotechnology* 1997;24(1):11–17.

119. Raub TJ. Signal transduction and glial cell modulation of cultured brain microvessel endothelial cell tight junctions. *Am J Physiol* 1996;271:C495–C503.

120. Lagrange P, Romero IA, Minn A, Revest PA. Transendothelial permeability changes induced by free radicals in an *in vitro* model of the blood–brain barrier. *Free Radic Biol Med* 1999;27(5–7):667–672.

121. Stanness KA, Guateo E, Janigro D. A dynamic model of the blood–brain barrier "*in vitro*." *Neurotoxicology* 1996;17(2):481–496.

122. Tetsuka K, Hosoya K, Ohtsuki S, Takanaga H, Yanai N, Ueda M, Obinata M, Terasaki T. Acidic amino acid transport characteristics of a newly developed conditionally immortalized rat type 2 astrocyte cell line (TR-AST). *Cell Structure and Function* 2001;26(4): 197–203.

123. Veronesi B. Characterization of the MDCK cell line for screening neurotoxicants. *NeuroToxicology* 1996;17(2):433–443.
124. Oldendorf WH. Measurement of brain uptake of radiolabeled substances using a tritiated water internal standard. *Brain Res* 1970;24:372–376.
125. Oldendorf WH. Blood–brain barrier permeability to drugs. *Annu Rev Pharmacol* 1974;14:239–248.
126. Kakee A, Terasaki T, Sygiyama Y. Brain efflux index as a novel method of analyzing efflux transport at the blood–brain barrier. *J Pharmacol Exp Ther* 1996;277:1550–1559.
127. Kakee A, Terasaki T, Sugiyama Y. Selective brain to blood efflux transport of para-aminohippuric acid across the blood–brain barrier: *in vivo* evidence by use of the brain efflux index method. *J Pharmacol Exp Ther* 1997;283(3):1018–1025.
128. Brodie BB, Kurz H, Shanker LS. The importance of dissociaton constant and lipid-solubility in influencing the passage of drugs into the cerebrospinal fluid. *J Pharmacol Exp Ther* 1960;130:20–25.
129. Ohno K, Pettigrew KD, Rapoport SI. Lower limits of cerebrovascular permeability to nonelectrolytes in the conscious rat. *Am J Physiol* 1978;235:H299–H307.
130. Preston E, Haas N. Defining the lower limits of blood–brain barrier permeability: factors affecting the magnitude and interpretation of permeability-area products. *J Neursci Res* 1986;16:709–719.
131. Smith QR. In Neuwelt EA, Ed. *Implications of the Blood–Brain Barrier and Its Manipulation, Volume 1: Basic Science Aspects*. New York: Plenum Publishing Corp; 1989, pp 85–118.
132. Smith QR. A review of blood–brain barrier transport techniques. *Methods Mol Med* 2003;89:193–208.
133. Patlak CS, Blasberg RG, Fenstermacher JD. Graphical evaluation of blood-to-brain transfer constants from multiple-time uptake data. *J Cerebral Blood Flow Metab* 1983;3:1–7.
134. Crone C, Levitt DG. In Renkin EM, Michel CC, Eds. *Handbook of Physiology*. Bethesda, MD: American Physiological Society; 1984, Vol 4, Part I, pp 411–466.
135. Patlak CS, Blasberg RG. Graphical evaluation of blood-to-brain transfer constants from multiple-time uptake data. Generalizations. *J Cerebral Blood Blow Metab* 1985;5: 584–590.
136. Takasato Y, Rapoport SI, Smith QR. An *in situ* brain perfusion technique to study cerebrovascular transport in the rat. *Am J Physiol* 1984;247:H484–H493.
137. Zlokovic BV, Begley DJ, Djuricic BM, Mitrovic DM. Measurement of solute transport across the blood–brain barrier in the perfused guinea pig brain: method and application to *N*-methyl-alpha-aminoisobutyric acid. *J Neurochem* 1986;46:1444–1451.
138. Vaughan DP, Dennis M. Mathematical basis of point-area deconvolution method for determining *in vivo* input functions. *J Pharm Sci* 1978;67:663–665.
139. Langenbucher F. Numerical convolution/deconvolution as a tool for correlating *in vitro* with *in vivo* drug availability. *Pharm Ind* 1982;44:1166–1172.
140. Langenbucher F. Improved understanding of convolution algorithms correlating body response with drug input. *Pharm Ind* 1982;44:1275–1278.
141. van Bree JBMM, Baljet AV, van Geyt A, de Boer AG, Danhof M, Breimer DD. The unit impulse response procedure for the pharmacokinetic evaluation of drug entry into the central nervous system. *J Pharmacokinet Biopharm* 1989;17:441–462.
142. Hendrikse NH, Schinkel AH, De Vries EGE, Fluks E, Van der Graaf WTA, et al. Complete *in vivo* reversal of P-glycoprotein pump function in the blood–brain barrier visualized with positron emission tomography. *Br J Pharmacol* 1998;124(7):1413–1418.

143. Davson H. A comparative study of the aqueous humour and cerebrospinal fluid in the rabbit. *J Physiol Lond* 1955;129:111–133.
144. Hendrikse NH, Franssen EJ, van der Graaf WT, Vaalburg W, De Vries EG. Visualization of multidrug resistance *in vivo*. *Eur J Nucl Med* 1999;26(3):283–293.
145. Hendrikse NH, de Vries EG, Eriks-Fluks L, van der Graaf WT, Hospers GA, Willemsen AT, Vaalburg W, Franssen EJ. A new *in vivo* method to study P-glycoproteIn transport In tumors and the blood–brain barrier. *Cancer Res* 1999;59(10):2411–2416.
146. Campbell D. Track autoradiography with iron-59 and sulphur-35 with quantitative evaluation. *Nature* 1951;167(4242):274–275.
147. de Lange ECM, de Boer AG, Breimer DD. Methodological issues in microdialysis sampling for pharmacokinetic studies. *Adv Drug Deliv Rev* 2000;45:125–148.
148. de Lange ECM, Zurcher C, Danhof M, de Boer AG, Breimer DD. Repeated microdialysis perfusions: periprobe tissue reactions and BBB permeability. *Brain Res* 1995;702:261–265.
149. de Lange ECM, de Boer AG, Breimer DD. Microdialysis for pharmacokinetic analysis of drug transport to the brain. *Adv Drug Deliv Rev* 1999;36:211–227.
150. Sawchuk RJ, Elmquist WF. Microdialysis in the study of drug transporters in the CNS. *Adv Drug Deliv Rev* 2000;45(2–3):295–307.
151. Xie R, Hammerlund-Udenaes M, de Boer AG, de Lange ECM. The role of P-glycoprotein in blood–brain barrier transport of morphine: transcortical microdialysis studies in mdr1a (–/–) and mdr1a (+/+) mice. *Br J Pharmacol* 1999;128:563–568.
152. Lonnroth P, Jansson PA, Smith U. A microdialysis method allowing characterization of intercellular water space in humans. *Am J Physiol* 1987;253:E228–E231.
153. Olson RJ, Justice JB. Quantitative microdialysis under transient conditions. *Anal Chem* 1993;65:1017–1022.
154. Jenkins BG, Chen YI, Kuestermann E, Makris NM, Nguyen V, et al. An integrated strategy for evaluation of metabolic and oxidative defects in neurodegenerative illness using magnetic resonance techniques. *Ann NY Acad Sci* 1999;893:214–242.
155. Del Sole A, Falini A, Ravasi L, Ottobrini L, De Marchis D, Bombardieri E, Lucignani G. Anatomical and biochemical investigation of primary brain tumours. *Eur J Nucl Med* 2001;28:1851–1872.
156. Barkhof F, van Walderveen M. Characterization of tissue damage in multiple sclerosis by nuclear magnetic resonance. *Philos Trans R Soc Lond B Biol Sci* 1999;354(1390):1675–1686.
157. Bradley WG Jr. Magnetic resonance imaging of the central nervous system. *Neurol Res* 1984;6(3):91–106.
158. Moller HE, Ullrich K, Weglage J. *In vivo* proton magnetic resonance spectroscopy in phenylketonuria. *Eur J Pediatr* 2000;159(Suppl 2):S121–S125.
159. Weglage J, Wiedermann D, Denecke J, Feldmann R, Koch HG, et al. Individual blood–brain barrier phenylalanine transport determines clinical outcome in phenylketonuria. *Ann Neurol* 2001;50(4):463–467.
160. McDonald WI. Rachelle Fishman–Matthew Moore Lecture. The pathological and clinical dynamics of multiple sclerosis. *Neuropathol Exp Neurol* 1994;53(4):338–343.
161. Simone IL, Tortorella C, Federico F. The contribution of (1)H-magnetic resonance spectroscopy in defining the pathophysiology of multiple sclerosis. *Ital J Neurol Sci* 1999;20(5 Suppl):S241–S245.
162. Ederoth P, Tunblad K, Bouw R, Lundberg CJ, Ungerstedt U, Nordstrom CH, Hammarlund-Udenaes M. Blood–brain barrier transport of MorphIne In patients with severe brain trauma. *Br J Clin Pharmacol* 2004;57(4):427–435.

163. Thakur ML, Lentle BC, SNM. Radiological Society of North America (RSNA) Joint SNM/RSNA Molecular Imaging Summit Statement. *J Nucl Med* 2005;46(9):11N–13N, 42N.
164. Pirko I, Fricke ST, Johnson AJ, Rodriguez M, Macura SI. Magnetic resonance imaging, microscopy, and spectroscopy of the central nervous system in experimental animals. *NeuroRx* 2005;2(2):250–264.
165. Jasanoff A. Functional MRI using molecular imaging agents. *Trends Neurosci* 2005;28(3): 120–126.
166. Koretsky AP. New developments in magnetic resonance imaging of the brain. *NeuroRx* 2004;1(1):155–164.
167. Rudin M, Rausch M, Stoeckli M. Molecular imaging in drug discovery and development: potential and limitations of nonnuclear methods. *Mol Imaging Biol* 2005;7(1):5–13.
168. Inglese M, Ge Y. Quantitative MRI: hidden age-related changes in brain tissue. *Top Magn Reson Imaging* 2004;15(6):355–363.
169. Gothelf D, Furfaro JA, Penniman LC, Glover GH, Reiss AL. The contribution of novel brain imaging techniques to understanding the neurobiology of mental retardation and developmental disabilities. *Ment Retard Dev Disabil Res Rev* 2005;11(4):331–339.
170. Jacobs AH, Kracht LW, Gossmann A, Ruger MA, Thomas AV, Thiel A, Herholz K. Imaging in neurooncology. *NeuroRx* 2005;2(2):333–347.
171. Sipkins DA, Gijbels K, Tropper FD, Bednarski M, Li KC, Steinman L. ICAM-1 expression in autoimmune encephalitis visualized using magnetic resonance imaging. *J Neuroimmunol* 2000;104(1):1–9.
172. Liu X, Chen C. Strategies to optimize brain penetration in drug discovery. *Curr Opin Drug Discov Dev* 2005;8(4):505–512.
173. Nicholson C, Sykova, E. Extracellular space structure revealed by diffusion analysis. *Trends Neurosci* 1998;21(5):207–215.
174. Fenstermacher J, Kaye T. Drug "diffusion" within the brain. *Ann N Y Acad Sci* 1988;531: 29–39.
175. Aird RB. A study of intrathecal, cerebrospinal fluid-to-brain exchange. *Exp Neurol* 1984;86(2):342–358.
176. Rall DP. Transport through the ependymal linings. *Prog Brain Res* 1968;29:159–172.
177. Yan Q, Matheson C, Sun J, Radeke MJ, Feinstein SC, Miller JA. Distribution of intracerebral ventricularly administered neurotrophins in rat brain and its correlation with trk receptor expression. *Exp Neurol* 1994;127(1):23–36.
178. Nagaraja TN, Patel P, Gorski M, Gorevic PD, Patlak CS, Fenstermacher JD. In normal rat, intraventricularly administered insulin-like growth factor-1 is rapidly cleared from CSF with limited distribution into brain. *Cerebrospinal Fluid Res* 2005;26:2–5.
179. Nutt JG, Burchiel KJ, Comella CL, Jankovic J, Lang AE, Laws ER Jr, Lozano AM, Penn RD, Simpson RK Jr, Stacy M, Wooten GF, ICV GDNF Study Group. Implanted intracerebroventricular, glial cell line-derived neurotrophic factor. Randomized, double-blind trial of glial cell line-derived neurotrophic factor (GDNF) in PD. *Neurology* 2003;60(1):69–73.
180. Patel NK, Bunnage M, Plaha P, Svendsen CN, Heywood P, Gill SS. Intraputamenal infusion of glial cell line-derived neurotrophic factor in PD: a two-year outcome study. *Ann Neurol* 2005;57(2):298–302.
181. Slevin JT, Gerhardt GA, Smith CD, Gash DM, Kryscio R, Young B. Improvement of bilateral motor functions in patients with Parkinson disease through the unilateral intraputaminal infusion of glial cell line-derived neurotrophic factor. *J Neurosurg* 2005;102(2):216–222.

182. Amgen Press Release. 11 Feb 2005.
183. Mardor Y, Rahav O, Zauberman Y, Lidar Z, Ocherashvilli A, Daniels D, Roth Y, Maier SE, Orenstein A, Ram Z. Convention-enhanced dud delivery increased efficacy and magnetic resonance image monitoring. *Cancer Res* 2005;65(15):6858–6863.
184. Ohata K, Marmarou A. Clearance of brain edema and macromolecules through the cortical extracellular space. *J Neurosurg* 1992;77:387–396.
185. Chen MY, Lonser RR, Morrison PF, Governale LS, Oldfield EH. Variables affecting CED to the striatum: a systemic examination of rate of infusion, cannula size, infusate concentration, and tissue–cannula sealing time. *J Neurosurg* 1999;90:315–320.
186. Kroll RA, Neuwelt EA. Outwitting the blood–brain barrier for therapeutic purposes: osmotic opening and other means. *Neurosurgery* 1998;42:1083–1100.
187. Wells DJ. Opening the floodgates: clinically applicable hydrodynamic delivery of plasmid DNA to skeletal muscle. *Mol Ther* 2004;10(2):207–208.
188. Wolff JA, Budker V. The mechanism of naked DNA uptake and expression. *Adv Genet* 2005;54:3–20.
189. Krewson CE, Klarman ML, Saltzman WM. Distribution of nerve growth factor following direct delivery to brain interstitium. *Brain Res* 1995;680(1–2):196–206.
190. Davidson BL, Breakefield XO. Viral vectors for gene delivery to the nervous system. *Nat Rev Neurosci* 2003;4(5):353–364.
191. Snyder RO, Francis J. Adeno-associated viral vectors for clinical gene transfer studies. *Curr Gene Ther* 2005;5(3):311–321.
192. Deglon N, Hantraye P. Viral vectors as tools to model and treat neurodegenerative disorders. *J Gene Med* 2005;7(5):530–539.
193. de Lima MC, Da Cruz MT, Cardoso AL, Simoes S, de Almeida LP. Liposomal and viral vectors for gene therapy of the central nervous system. *Curr Drug Targets CNS Neurol Disord* 2005;4(4):453–465.
194. Fountaine TM, Wood MJ, Wade-Martins R. Delivering RNA interference to the mammalian brain. *Curr Gene Ther* 2005;5(4):399–410.
195. Kootstra NA, Verma IM. Gene therapy with viral vectors. *Annu Rev Pharmacol Toxicol* 2003;43:413–439.
196. Kay MA, Glorioso JC, Naldini L. Viral vectors for gene therapy: the art of turning infectious agents into vehicles of therapeutics. *Nat Med* 2001;7(1):33–40.
197. Chen H, McCarty DM, Bruce AT, Suzuki K. Oligodendrocyte-specific gene expression in mouse brain: use of a myelin-forming cell type-specific promoter in an adeno-associated virus. *J Neurosci Res* 1999;55(4):504–513.
198. Niwa H, Yamamura K, Miyazaki J. Efficient selection for high-expression transfectants with a novel eukaryotic vector. *Gene* 1991;108(2):193–199.
199. Klein RL, Meyer EM, Peel AL, Zolotukhin S, Meyers C, Muzyczka N, King MA. Neuron-specific transduction in the rat septohippocampal or nigrostriatal pathway by recombinant adeno-associated virus vectors. *Exp Neurol* 1998;150(2):183–194.
200. Fu H, Samulski RJ, McCown TJ, Picornell YJ, Fletcher D, Muenzer J. Neurological correction of lysosomal storage in a mucopolysaccharidosis IIIB mouse model by adeno-associated virus-mediated gene delivery. *Mol Ther* 2002;5(1):42–49.
201. Peel AL, Zolotukhin S, Schrimsher GW, Muzyczka N, Reier PJ. Efficient transduction of green fluorescent protein in spinal cord neurons using adeno-associated virus vectors containing cell type-specific promoters. *Gene Ther* 1997;4(1):16–24.
202. Paterna JC, Moccetti T, Mura A, Feldon J, Bueler H. Influence of promoter and WHV post-transcriptional regulatory element on AAV-mediated transgene expression in the rat brain. *Gene Ther* 2000;7(15):1304–1311.

203. Zhang Y, Schlachetzki F, Zhang YF, Boado RJ, Pardridge WM. Normalization of striatal tyrosine hydroxylase and reversal of motor impairment in experimental parkinsonism with intravenous nonviral gene therapy and a brain-specific promoter. *Hum Gene Ther* 2004;15(4):339–350.
204. Mandel RJ, Rendahl KG, Spratt SK, Snyder RO, Cohen LK, Leff SE. Characterization of intrastriatal recombinant adeno-associated virus-mediated gene transfer of human tyrosine hydroxylase and human GTP-cyclohydrolase I in a rat model of Parkinson's disease. *J Neurosci* 1998;18(11):4271–4284.
205. Owen R, Mandel RJ, Ammini CV, Conlon TJ, Kerr DS, Stacpoole PW, Flotte TR. Gene therapy for pyruvate dehydrogenase E1alpha deficiency using recombinant adeno-associated virus 2 (rAAV2) vectors. *Mol Ther* 2002;6(3):394–399.
206. Donello JE, Loeb JE, Hope TJ. Woodchuck hepatitis virus contains a tripartite post-transcriptional regulatory element. *J Virol* 1998;72(6):5085–5092.
207. Loeb JE, Cordier WS, Harris ME, Weitzman MD, Hope TJ. Enhanced expression of transgenes from adeno-associated virus vectors with the woodchuck hepatitis virus post-transcriptional regulatory element: implications for gene therapy. *Hum Gene Ther* 1999;10(14):2295–2305.
208. Trojan J, Blossey BK, Johnson TR, Rudin SD, Tykocinski M, Ilan J, Ilan J. Loss of tumorigenicity of rat glioblastoma directed by episome-based antisense cDNA transcription of insulin-like growth factor I. *Proc Natl Acad Sci USA* 1992;89(11):4874–4878.
209. Zhang Y, Zhang YF, Bryant J, Charles A, Boado RJ, Pardridge WM. Intravenous RNA interference gene therapy targeting the human epidermal growth factor receptor prolongs survival in intracranial brain cancer. *Clin Cancer Res* 2004;10(11):3667–3677.
210. Snudden DK, Smith PR, Lai D, Ng MH, Griffin BE. Alterations in the structure of the EBV nuclear antigen, EBNA1, in epithelial cell tumours. *Oncogene* 1995;10(8):45–52.
211. Blau HM, Rossi FM. Tet B or not tet B: advances in tetracycline-inducible gene expression. *Proc Natl Acad Sci USA* 1999;96(3):7–9.
212. Kafri T, van Praag H, Gage FH, Verma IM. Lentiviral vectors: regulated gene expression. *Mol Ther* 2000;1:516–521.
213. Pfeifer A, Brandon EP, Kootstra N, Gage FH, Verma IM. Delivery of the Cre recombinase by a self-deleting lentiviral vector: efficient gene targeting *in vivo*. *Proc Natl Acad Sci USA* 2001;98(20):450–455.
214. Daly TM. Overview of adeno-associated viral vectors. *Methods Mol Biol* 2004;246: 157–165.
215. Burger C, Nash K, Mandel RJ. Recombinant adeno-associated viral vectors in the nervous system. *Hum Gene Ther* 2005;16(7):1–91.
216. Gao G, Alvira MR, Somanathan S, Lu Y, Vandenberghe LH, Rux JJ, Calcedo R, Sanmiguel J, Abbas Z, Wilson JM. Adeno-associated viruses undergo substantial evolution in primates during natural infections. *Proc Natl Acad Sci USA* 2003;100(10):81–86.
217. Grimm D, Kay MA. From virus evolution to vector revolution: use of naturally occurring serotypes of adeno-associated virus (AAV) as novel vectors for human gene therapy. *Curr Gene Ther* 2003;3(4):1–304.
218. McCown TJ, Xiao X, Li J, Breese GR, Samulski RJ. Differential and persistent expression patterns of CNS gene transfer by an adeno-associated virus (AAV) vector. *Brain Res* 1996;713(1–2):99–107.
219. Muzyczka N, Berns KI. Parvoviridae: the viruses and their replication. In Knipe DM, Howley PM, Eds. *Field's Virology*. New York: Lippincott, Williams; 2001, pp 2327–2360.
220. Lu Y. Recombinant adeno-associated virus as delivery vector for gene therapy—a review. *Stem Cells Dev* 2004;13(1):3–45.

221. Fitzsimons HL, Bland RJ, During MJ. Promoters and regulatory elements that improve adeno-associated virus transgene expression in the brain. *Methods* 2002;28(2):7–36.
222. Davidson BL, Paulson HL. Molecular medicine for the brain: silencing of disease genes with RNA interference. *Lancet Neurol* 2004;3(3):5–9.
223. Somia N, Verma IM. Gene therapy: trials and tribulations. *Nat Rev Genet* 2000;1(2): 1–9.
224. Hsich G, Sena-Esteves M, Breakefield XO. Critical issues in gene therapy for neurologic disease. *Hum Gene Ther* 2002;13(5):9–604.
225. Tang G, Chiocca A. Gene transfer and delivery in central nervous system disease. *Neurosurg Focus* 1997;3(3):e2.
226. Latchman DS, Coffin RS. Viral vectors in the treatment of Parkinson's disease. *Mov Disord* 2000;15(1):17.
227. Ruitenberg MJ, Blits B, Dijkhuizen PA, te Beek ET, Bakker A, van Heerikhuize JJ, Pool CW, Hermens WT, Boer GJ, Verhaagen J. Adeno-associated viral vector-mediated gene transfer of brain-derived neurotrophic factor reverses atrophy of rubrospinal neurons following both acute and chronic spinal cord injury. *Neurobiol Dis* 2004; 15(2):4–406.
228. Kirik D, Georgievska B, Burger C, Winkler C, Muzyczka N, Mandel RJ, Bjorklund A. Reversal of motor impairments in parkinsonian rats by continuous intrastriatal delivery of L-dopa using rAAV-mediated gene transfer. *Proc Natl Acad Sci USA* 2002;99(7):8–13.
229. Shen Y, Muramatsu SI, Ikeguchi K, Fujimoto KI, Fan DS, Ogawa M, Mizukami H, Urabe M, Kume A, Nagatsu I, Urano F, Suzuki T, Ichinose H, Nagatsu T, Monahan J, Nakano I, Ozawa K. Triple transduction with adeno-associated virus vectors expressing tyrosine hydroxylase, aromatic-L-amino-acid decarboxylase, and GTP cyclohydrolase I for gene therapy of Parkinson's disease. *Hum Gene Ther* 2000;11(11):9–19.
230. Passini MA, Watson DJ, Vite CH, Landsburg DJ, Feigenbaum AL, Wolfe JH. Intraventricular brain injection of adeno-associated virus type 1 (AAV1) in neonatal mice results in complementary patterns of neuronal transduction to AAV2 and total long-term correction of storage lesions in the brains of beta-glucuronidase-deficient mice. *J Virol* 2003;77(12):34–40.
231. Desmaris N, Verot L, Puech JP, Caillaud C, Vanier MT, Heard JM. Prevention of neuropathology in the mouse model of Hurler syndrome. *Ann Neurol* 2004;56(1):8–76.
232. Kordower JH, Emborg ME, Bloch J, Ma SY, Chu Y, Leventhal L, McBride J, Chen EY, Palfi S, Roitberg BZ, Brown WD, Holden JE, Pyzalski R, Taylor MD, Carvey P, Ling Z, Trono D, Hantraye P, Deglon N, Aebischer P. Neurodegeneration prevented by lentiviral vector delivery of GDNF in primate models of Parkinson's disease. *Science* 2000; 290(5492):767–773.
233. Consiglio A, Quattrini A, Martino S, Bensadoun JC, Dolcetta D, Trojani A, Benaglia G, Marchesini S, Cestari V, Oliverio A, Bordignon C, Naldini L. *In vivo* gene therapy of metachromatic leukodystrophy by lentiviral vectors: correction of neuropathology and protection against learning impairments in affected mice. *Nat Med* 2001;7(3):310–316.
234. Mandel RJ, Snyder RO, Leff SE. Recombinant adeno-associated viral vector-mediated glial cell line-derived neurotrophic factor gene transfer protects nigral dopamine neurons after onset of progressive degeneration in a rat model of Parkinson's disease. *Exp Neurol* 1999;160(1):205–214.
235. Bosch A, Perret E, Desmaris N, Heard JM. Long-term and significant correction of brain lesions in adult mucopolysaccharidosis type VII mice using recombinant AAV vectors. *Mol Ther* 2000;1(1):63–70.

236. Dodge JC, Clarke J, Song A, Bu J, Yang W, Taksir TV, Griffiths D, Zhao MA, Schuchman EH, Cheng SH, O'riordan CR, Shihabuddin LS, Passini MA, Stewart GR. Gene transfer of human acid sphingomyelinase corrects neuropathology and motor deficits in a mouse model of Niemann–Pick type A disease. *Proc Natl Acad Sci USA*, 2005;Nov 21 [Epub ahead of print].
237. Xu J, Ma C, Bass C, Terwilliger EF. A combination of mutations enhances the neurotropism of AAV-2. *Virology* 2005;341(2):203–214.
238. Wang X, Kong L, Zhang GR, Sun M, Geller AI. Targeted gene transfer to nigrostriatal neurons in the rat brain by helper virus-free HSV-1 vector particles that contain either a chimeric HSV-1 glycoprotein C-GDNF or a gC-BDNF protein. *Brain Res Mol Brain Res* 2005;139(1):88–102.
239. Jooss K, Chirmule N. Immunity to adenovirus and adeno-associated viral vectors: implications for gene therapy. *Gene Ther* 2003;10(11):955–963.
240. Barkats M, Bilang-Bleuel A, Buc-Caron MH, Castel-Barthe MN, Corti O, Finiels F, Horellou P, Revah F, Sabate O, Mallet J. Adenovirus in the brain: recent advances of gene therapy for neurodegenerative diseases. *Prog Neurobiol* 1998;55(4):333–341.
241. Lowenstein PR, Castro MG. Inflammation and adaptive immune responses to adenoviral vectors injected into the brain: peculiarities, mechanisms, and consequences. *Gene Ther* 2003;10(11):946–954.
242. Thomas CE, Storm TA, Huang Z, Kay MA. Rapid uncoating of vector genomes is the key to efficient liver transduction with pseudotyped adeno-associated virus vectors. *J Virol* 2004;78(6):3110–3122.
243. Landles C, Bates GP. Huntington and the molecular pathogenesis of Huntington's disease. Fourth in molecular medicine review series. *EMBO Rep* 2004;5(10):958–963.
244. Thomas CE, Schiedner G, Kachaned S, Castro MG, Loewenstein PR. Peripheral infection with adenovirus causes unexpected long-term brain inflammation in animals injected intracranially with first-generation, but not with high capacity adenovirus vectors: toward realistic long-term neurological gene therapy for chronic diseases. *Proc Natl Acad Sci USA* 2000;97:7482–7487.
245. Schlachetzki F, Zhang Y, Boado RJ, Pardridge WM. Gene therapy of the brain: the transvascular approach. *Neurology* 2004;62(8):1275–1281.
246. Abordo-Adesida E, Follenzi A, Barcia C, Sciascia S, Castro MG, Naldini L, Lowenstein PR. Stability of lentiviral vector-mediated transgene expression in the brain in the presence of systemic antivector immune responses. *Hum Gene Ther* 2005;16(6):741–751.
247. Tsuji D, Kuroki A, Ishibashi Y, Itakura T, Itoh K. Metabolic correction in microglia derived from Sandhoff disease model mice. *J Neurochem* 2005;94(6):1631–1638.
248. Kobayashi H, Carbonaro D, Pepper K, Petersen D, Ge S, Jackson H, Shimada H, Moats R, Kohn DB. Neonatal gene therapy of MPS I mice by intravenous injection of a lentiviral vector. *Mol Ther* 2005;11(5):776–789.
249. Fjord-Larsen L, Johansen JL, Kusk P, Tornoe J, Gronborg M, Rosenblad C, Wahlberg LU. Efficient *in vivo* protection of nigral dopaminergic neurons by lentiviral gene transfer of a modified Neurturin construct. *Exp Neurol* 2005;195(1):49–60.
250. Georgievska B, Jakobsson J, Persson E, Ericson C, Kirik D, Lundberg C. Regulated delivery of glial cell line-derived neurotrophic factor into rat striatum, using a tetracycline-dependent lentiviral vector. *Hum Gene Ther* 2004;15(10):934–944.
251. Ralph GS, Radcliffe PA, Day DM, Carthy JM, Leroux MA, Lee DC, Wong LF, Bilsland LG, Greensmith L, Kingsman SM, Mitrophanous KA, Mazarakis ND, Azzouz M. Silencing mutant SOD1 using RNAi protects against neurodegeneration and extends survival in an ALS model. *Nat Med* 2005;11(4):429–433.

252. Vezzani A. Gene therapy in epilepsy. *Epilepsy Curr* 2004;4(3):87–90.
253. Crozet C, Lin YL, Mettling C, Mourton-Gilles C, Corbeau P, Lehmann S, Perrier V. Inhibition of PrPSc formation by lentiviral gene transfer of PrP containing dominant negative mutations. *J Cell Sci* 2004;117(Pt 23):5591–5597.
254. Steffens S, Tebbets J, Kramm CM, Lindemann D, Flake A, Sena-Esteves M. Transduction of human glial and neuronal tumor cells with different lentivirus vector pseudotypes. *J Neurooncol* 2004;70(3):281–288.
255. Wong LF, Ralph GS, Walmsley LE, Bienemann AS, Parham S, Kingsman SM, Uney JB, Mazarakis ND. Lentiviral-mediated delivery of Bcl-2 or GDNF protects against excitotoxicity in the rat hippocampus. *Mol Ther* 2005;11(1):89–95.
256. Pardridge WM. Drug and gene delivery to the brain: the vascular route. *Neuron* 2002;36(4):555–558.
257. Bradbury MW. *The Concept of a Blood–Brain Barrier*. Chichester: Wiley; 1979.
258. Pardridge WM. Transport of small molecules through the blood–brain barrier: biology and methodology. *Adv Drug Deliv Rev* 1995;15:5–36.
259. Stein WD. *The Movement of Molecules Across Cell Membranes*. New York: Academic Press; 1967.
260. Abraham MH, Chadha HS, Mitchell RC. Hydrogen bonding. 33. Factors that influence the distribution of solutes between blood and brain. *J Pharm Sci* 1994;83(9):1257–1268.
261. Lipinski CA, Lombardo F, Dominy BW, Feeney PJ. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv Drug Deliv Rev* 2001;46(1–3):3–26.
262. Bickel U, Yoshikawa T, Pardridge WM. Delivery of peptides and proteins through the blood–brain barrier. *Adv Drug Deliv Rev* 2001;46(1–3):247–279.
263. Vorbrodt AW. Ultracytochemical characterization of anionic sites in the wall of brain capillaries. *J Neurocytol* 1989;18(3):359–368.
264. Lockman PR, Koziara JM, Mumper RJ, Allen DD. Nanoparticle surface charges alter blood–brain barrier integrity and permeability. *J Drug Target* 2004;12(9–10):635–641.
265. Abbott NJ, Romero IA. Transporting therapeutics across the blood–brain barrier. *Mol Med Today* 1996;2(3):106–113.
266. Tamai I, Tsuji A. Transporter-mediated permeation of drugs across the blood–brain barrier. *J Pharm Sci* 2000;89(11):1371–1388.
267. Tsuji A, Tamai II. Carrier-mediated or specialized transport of drugs across the blood–brain barrier. *Adv Drug Deliv Rev* 1999;36(2–3):277–290.
268. Smith QR. Drug delivery to brain and the role of carrier mediated transport. *Adv Exp Med Biol* 1993;331:83–93.
269. Duffy KR, Pardridge WM. Blood–brain barrier transcytosis of insulin in developing rabbits. *Brain Res* 1987;420(1):32–38.
270. Moos T, Morgan EH. Transferrin and transferrin receptor function in brain barrier systems. *Cell Mol Neurobiol* 2000;20(1):77–95.
271. Pardridge WM, Eisenberg J, Yang J. Human blood–brain barrier transferrin receptor. *Metabolism* 1987;36(9):892–895.
272. Dehouck B, Fenart L, Dehouck MP, et al. A new function for the LDL receptor: transcytosis of LDL across the blood–brain barrier. *J Cell Biol* 1997;138(4):877–889.
273. Bjorbaek C, Elmquist JK, Michl P, et al. Expression of leptin receptor isoforms in rat brain microvessels. *Endocrinology* 1998;139(8):3485–3491.

274. Duffy KR, Pardridge WM, Rosenfeld RG. Human blood–brain barrier insulin-like growth factor receptor. *Metabolism* 1988;37(2):136–140.
275. Loscher W, Potschka H. Drug resistance in brain diseases and the role of drug efflux transporters. *Nat Rev Neurosci* 2005;6(8):591–602.
276. Loscher W, Potschka H. Role of drug efflux transporters in the brain for drug disposition and treatment of brain diseases. *Prog Neurobiol* 2005;76(1):22–76.
277. van der Sandt IC, De Boer AG, Breimer DD. Implications of Pgp for the transport and distribution of drugs into the brain. In Sharma HS, Wesman J, Eds. *Blood–Spinal Cord and Brain Barriers in Health and Disease*. San Diego, CA: Elsevier Academic Press; 2004, pp 63–72.
278. Zhang Y, Han H, Elmquist WF, Miller DW. Expression of various multidrug resistance-associated protein (MRP) homologues in brain microvessel endothelial cells. *Brain Res* 2000;876(1–2):148–153.
279. Yokoyama M. Drug targeting with nano-sized carrier systems. *J Artif Organs* 2005;8(2):77–84.
280. Kreuter J. Influence of the surface properties on nanoparticle-mediated transport of drugs to the brain. *J Nanosci Nanotechnol* 2004;4(5):484–488.
281. Marcucci F, Lefoulon F. Active targeting with particulate drug carriers in tumor therapy: fundamentals. *Drug Discov Today* 2004;9(5):219–228.
282. Matsumura Y, Maeda H. A new concept for macromolecular therapeutics in cancer chemotherapy: mechanism of tumoritropic accumulation of proteins and the antitumor agent smancs. *Cancer Res* 1986;46(12 Pt 1):6387–6392.
283. Gabizon A, Papahadjopoulos D. Liposome formulations with prolonged circulation time in blood and enhanced uptake by tumors. *Proc Natl Acad Sci USA* 1988;85(18):6949–6953.
284. Miller G. Drug targeting. Breaking down barriers. *Science* 2002;297(5584):1116–1118.
285. Erdlenbruch B, Alipour M, Fricker G, Miller DS, Kugler W, Eibl H, Lakomek M. Alkylglycerol opening of the blood–brain barrier to small and large fluorescence markers in normal and C6 glioma-bearing rats and isolated rat brain capillaries. *Br J Pharmacol* 2003;140:1201–1210.
286. Matsukado K, Inamura T, Nakano S, Fukui M, Bartus RT, Black KL. Enhanced tumor uptake of carboplatin and survival in glioma-bearing rats by intra-carotid infusion of bradykinin analog, RMP-7. *Neurosurgery* 1996;39:125–133.
287. Prados MD, Schold SC Jr, Fine HA, Jaeckle K, Hochberg F, Mechtler L, Fetell MR, Phuphanich S, Feun L, Janus TJ, Ford K, Graney W. A randomized, double-blind, placebo-controlled, phase 2 study of RMP-7 in combination with carboplatin administered intravenously for the treatment of recurrent malignant glioma. *Neuro-oncol* 2003;5(2):96–103.
288. Packer RJ, Krailo M, Mehta M, Warren K, Allen J, Jakacki R, Villablanca JG, Chiba A, Reaman G. Phase 1 study of concurrent RMP-7 and carboplatin with radiotherapy for children with newly diagnosed brainstem gliomas. *Cancer* 2005;104(6):1281–1287.
289. Dietz GP, Bahr M. Delivery of bioactive molecules into the cell: the Trojan horse approach. *Mol Cell Neurosci* 2004;27(2):85–131.
290. Lindgren M, Hallbrink M, Prochiant A, Langel U. Cell-penetrating peptides. *Trends Pharmacol Sci* 2000;21(3):99–103.
291. Jeang KT, Xiao H, Rich EA. Multifaceted activities of the HIV-1 transactivator of transcription, *Tat. J Biol Chem* 1999;274(41):28837–28840.

292. Perez F, Joliot A, Bloch-Gallego E, Zahraoui A, Triller A, Prochiantz A. Antennapedia homeobox as a signal for the cellular internalization and nuclear addressing of a small exogenous peptide. *J Cell Sci* 1992;102(Pt 4):717–722.
293. Rousselle C, Clair P, Lefauconnier JM, Kaczorek M, Scherrmann JM, Temsamani J. New advances in the transport of doxorubicin through the blood–brain barrier by a peptide vector-mediated strategy. *Mol Pharmacol* 2000;57(4):679–686.
294. Langedijk JP, Olijhoek T, Schut D, Autar R, Meloen RH. New transport peptides broaden the horizon of applications for peptidic pharmaceuticals. *Mol Divers* 2004;8(2):101–111.
295. Wender PA, Mitchell DJ, Pattabiraman K, Pelkey ET, Steinman L, Rothbard JB. The design, synthesis, and evaluation of molecules that enable or enhance cellular uptake: peptoid molecular transporters. *Proc Natl Acad Sci USA* 2000;97(24):13003–13008.
296. Saar K, Lindgren M, Hansen M, Eiriksdottir E, Jiang Y, Rosenthal-Aizman K, Sassian M, Langel U. Cell-penetrating peptides: a comparative membrane toxicity study. *Anal Biochem* 2005;345(1):55–65.
297. Soutschek J, Akinc A, Bramlage B, Charisse K, Constien R, Donoghue M, Elbashir S, Geick A, Hadwiger P, Harborth J, John M, Kesavan V, Lavine G, Pandey RK, Racie T, Rajeev KG, Rohl I, Toudjarska I, Wang G, Wuschko S, Bumcrot D, Koteliansky V, Limmer S, Manoharan M, Vornlocher HP. Therapeutic silencing of an endogenous gene by systemic administration of modified siRNAs. *Nature* 2004;432(7014):173–178.
298. Batrakova EV, Vinogradov SV, Robinson SM, Niehoff ML, Banks WA, Kabanov AV. Polypeptide point modifications with fatty acid and amphiphilic block co-polymers for enhanced brain delivery. *Bioconjug Chem* 2005;16(5):793–802.
299. Laakso T, Andersson J, Artursson P, Edman P, Sjoholm I. Acrylic microspheres *in vivo*. X. Elimination of circulating cells by active targeting using specific monoclonal antibodies bound to microparticles. *Life Sci* 1986;38(2):183–190.
300. Conner SD, Schmid SL. Regulated portals of entry into the cell. *Nature* 2003; 422(6927):37–44.
301. Gumbleton M, Abulrob AG, Campbell L. Caveolae: an alternative membrane transport compartment. *Pharm Res* 2000;17(9):1035–1048.
302. Sachse M, Urbe S, Oorschot V, Strous GJ, Klumperman J. Bilayered clathrin coats on endosomal vacuoles are involved in protein sorting toward lysosomes. *Mol Biol Cell* 2002;13(4):1313–1328.
303. Cuervo AM. Autophagy: many paths to the same end. *Mol Cell Biochem* 2004; 263(1–2):55–72.
304. Ciechanover A. Proteolysis: from the lysosome to ubiquitin and the proteasome. *Nat Rev Mol Cell Biol* 2005;6(1):79–87.
305. Moos T, Morgan EH. Transferrin and transferrin receptor function in brain barrier systems. *Cell Mol Neurobiol* 2000;20(1):77–95.
306. Morgan EH. Iron metabolism and transport. In Zakin D, Boyer TD, Eds. *Hepatology. A Textbook of Liver Disease*. Philadelpia: Saunders; 1996, pp 526–554.
307. Ponka P, Lok CN. The transferrin receptor: role in health and disease. *Int J Biochem Cell Biol* 1999;31(10):1111–1137.
308. Visser CC, Stevanovic S, Voorwinden LH, van Bloois L, Gaillard PJ, Danhof M, Crommelin DJ, de Boer AG. Targeting liposomes with protein drugs to the blood–brain barrier *in vitro*. *Eur J Pharm Sci* 2005;25(2–3):299–305.
309. Zhang Y, Pardridge WM. Rapid transferrin efflux from brain to blood across the blood–brain barrier. *J Neurochem* 2001;76(5):1597–1600.

310. Pardridge WM, Buciak JL, Friden PM. Selective transport of an anti-transferrin receptor antibody through the blood–brain barrier *in vivo*. *J Pharmacol Exp Ther* 1991;259(1):66–70.
311. Moos T, Morgan EH. The significance of the mutated divalent metal transporter (DMT1) on iron transport into the Belgrade rat brain. *J Neurochem* 2004;88(1):233–245.
312. Deane R, Zheng W, Zlokovic BV. Brain capillary endothelium and choroid plexus epithelium regulate transport of transferrin-bound and free iron into the rat brain. *J Neurochem* 2004;88(4):813–820.
313. Pardridge WM, Boado RJ, Kang YS. Vector-mediated delivery of a polyamide ("peptide") nucleic acid analogue through the blood–brain barrier *in vivo*. *Proc Natl Acad Sci USA* 1995;92(12):5592–5596.
314. Shi N, Boado RJ, Pardridge WM. Receptor-mediated gene targeting to tissues *in vivo* following intravenous administration of pegylated immunoliposomes. *Pharm Res* 2001;18(8):1091–1095.
315. Zhang YF, Boado RJ, Pardridge WM. Absence of toxicity of chronic weekly intravenous gene therapy with pegylated immunoliposomes. *Pharm Res* 2003;20(11):1779–1785.
316. Broadwell RD, Baker-Cairns BJ, Friden PM, Oliver C, Villegas JC. Transcytosis of protein through the mammalian cerebral epithelium and endothelium. III. Receptor-mediated transcytosis through the blood–brain barrier of blood-borne transferrin and antibody against the transferrin receptor. *Exp Neurol* 1996;142(1):47–65.
317. Moos T, Morgan EH. Restricted transport of anti-transferrin receptor antibody (OX26) through the blood–brain barrier in the rat. *J Neurochem* 2001;79(1):119–129.
318. Van Gelder W, Huijskes-Heins MI, Cleton-Soeteman MI, Van Dijk JP, Van Eijk HG. Iron uptake in blood–brain barrier endothelial cells cultured in iron-depleted and iron-enriched media. *J Neurochem* 1998;71(3):1134–1140.
319. Visser CC, Voorwinden LH, Crommelin DJ, Danhof M, De Boer AG. Characterization and modulation of the transferrin receptor on brain capillary endothelial cells. *Pharm Res* 2004;21(5):761–769.
320. Pardridge WM. Drug and gene targeting to the brain via blood–brain barrier receptor-mediated transport systems. *Int Congress Series* 2005;1277:49–59.
321. Pardridge WM. Vector discovery: genetically engineered Trojan horses for drug targeting. In Pardridge WM, Ed. *Brain Drug Targeting—The Future of Brain Drug Development*. New York: Cambridge University Press; 2001, pp 126–154.
322. Xu L, Tang WH, Huang CC, et al. Systemic p53 gene therapy of cancer with immunolipoplexes targeted by anti-transferrin receptor scFv. *Mol Med* 2001;7(10):723–734.
323. Lee JH, Engler JA, Collawn JF, Moore BA. Receptor mediated uptake of peptides that bind the human transferrin receptor. *Eur J Biochem* 2001;268(7):2004–2012.
324. Bottaro DP, Bonner-Weir S, King GL. Insulin receptor recycling in vascular endothelial cells. Regulation by insulin and phorbol ester. *J Biol Chem* 1989;264(10):5916–5923.
325. Wu D, Yang J, Pardridge WM. Drug targeting of a peptide radiopharmaceutical through the primate blood–brain barrier *in vivo* with a monoclonal antibody to the human insulin receptor. *J Clin Invest* 1997;100(7):1804–1812.
326. Coloma MJ, Lee HJ, Kurihara A, et al. Transport across the primate blood–brain barrier of a genetically engineered chimeric monoclonal antibody to the human insulin receptor. *Pharm Res* 2000;17(3):266–274.
327. Herz J, Strickland DK. LRP: a multifunctional scavenger and signaling receptor. *J Clin Invest* 2001;108(6):779–784.
328. Herz J. LRP: a bright beacon at the blood–brain barrier. *J Clin Invest* 2003;112(10): 1483–1485.

329. Boucher P, Gotthardt M, Li WP, Anderson RG, Herz J. LRP: role in vascular wall integrity and protection from atherosclerosis. *Science* 2003;300(5617):329–332.
330. Yepes M, Sandkvist M, Moore EG, et al. Tissue-type plasminogen activator induces opening of the blood–brain barrier via the LDL receptor-related protein. *J Clin Invest* 2003;112(10):1533–1540.
331. Demeule M, Poirier J, Jodoin J, et al. High transcytosis of melanotransferrin (P97) across the blood–brain barrier. *J Neurochem* 2002;83(4):924–933.
332. Deane R, Wu Z, Zlokovic BV. RAGE (yin) versus LRP (yang) balance regulates Alzheimer amyloid beta-peptide clearance through transport across the blood–brain barrier. *Stroke* 2004;35(11 Suppl 1):2628–2631.
333. Chun JT, Wang L, Pasinetti GM, Finch CE, Zlokovic BV. Glycoprotein 330/megalin (LRP-2) has low prevalence as mRNA and protein in brain microvessels and choroid plexus. *Exp Neurol* 1999;157(1):194–201.
334. Richardson DR, Morgan EH. The transferrin homologue, melanotransferrin (p97), is rapidly catabolized by the liver of the rat and does not effectively donate iron to the brain. *Biochim Biophys Acta* 2004;1690(2):124–133.
335. Gabathuler R, Kolaitis G, Brooks RC, et al. WO0213843 (2002).
336. Gabathuler R, Arthur G, Kennard M, et al. Development of a potential vector (NeuroTrans) to deliver drugs across the blood–brain barrier. *Int Congress Series* 2005; 1277:171–184.
337. Pan W, Kastin AJ, Zankel TC, et al. Efficient transfer of receptor-associated protein (RAP) across the blood–brain barrier. *J Cell Sci* 2004;117(Pt 21):5071–5078.
338. Béliveau R, Demeule M. WO2004060403 (2004).
339. Sedrakyan A, Treasure T, Elefteriades JA. Effect of aprotinin on clinical outcomes in coronary artery bypass graft surgery: a systematic review and meta-analysis of randomized clinical trials. *J Thorac Cardiovasc Surg* 2004;128(3):442–448.
340. Jefferies WA, Food MR, Gabathuler R, et al. Reactive microglia specifically associated with amyloid plaques in Alzheimer's disease brain tissue express melanotransferrin. *Brain Res* 1996;712(1):122–126.
341. Deane R, Du Yan S, Submamaryan RK, et al. RAGE mediates amyloid-beta peptide transport across the blood–brain barrier and accumulation in brain. *Nat Med* 2003;9(7):907–913.
342. Zlokovic BB. Cerebrovascular transport of Alzheimer's amyloid beta and apolipoproteins J and E: possible anti-amyloidogenic role of the blood–brain barrier. *Life Sci* 1996;59(18):1483–1497.
343. Zlokovic BV, Martel CL, Matsubara E, et al. Glycoprotein 330/megalin: probable role in receptor-mediated transport of apolipoprotein J alone and in a complex with Alzheimer disease amyloid beta at the blood–brain and blood–cerebrospinal fluid barriers. *Proc Natl Acad Sci USA* 1996;93(9):4229–4234.
344. Demeule M, Bertrand Y, Michaud-Levesque J, et al. Regulation of plasminogen activation: a role for melanotransferrin (p97) in cell migration. *Blood* 2003;102(5):1723–1731.
345. Gaillard PJ, Brink A, De Boer AG. Diphtheria toxin receptor-targeted brain drug delivery. *Int Congress Series* 2005;1277:185–195.
346. Gaillard PJ, Brink A, De Boer AG.WO2004069870 (2004).
347. Anderson P. Antibody responses to *Haemophilus influenzae* type b and diphtheria toxin induced by conjugates of oligosaccharides of the type b capsule with the nontoxic protein CRM197. *Infect Immun* 1983;39(1):233–238.

348. Buzzi S, Rubboli D, Buzzi G, et al. CRM197 (nontoxic diphtheria toxin): effects on advanced cancer patients. *Cancer Immunol Immunother* 2004;53(11):1041–1048.
349. Raab G, Klagsbrun M. Heparin-binding EGF-like growth factor. *Biochim Biophys Acta* 1997;1333(3):F179–F199.
350. Mishima K, Higashiyama S, Nagashima Y, et al. Regional distribution of heparin-binding epidermal growth factor-like growth factor mRNA and protein in adult rat forebrain. *Neurosci Lett* 1996;213(3):153–156.
351. Mishima K, Higashiyama S, Asai A, et al. Heparin-binding epidermal growth factor-like growth factor stimulates mitogenic signaling and is highly expressed in human malignant gliomas. *Acta Neuropathol (Berl)* 1998;96(4):322–328.
352. Tanaka N, Sasahara M, Ohno M, et al. Heparin-binding epidermal growth factor-like growth factor mRNA expression in neonatal rat brain with hypoxic/ischemic injury. *Brain Res* 1999;827(1–2):130–138.
353. Mitamura T, Higashiyama S, Taniguchi N, Klagsbrun M, Mekada E. Diphtheria toxin binds to the epidermal growth factor (EGF)-like domain of human heparin-binding EGF-like growth factor/diphtheria toxin receptor and inhibits specifically its mitogenic activity. *J Biol Chem* 1995;270(3):1015–1019.
354. Cha JH, Chang MY, Richardson JA, Eidels L. Transgenic mice expressing the diphtheria toxin receptor are sensitive to the toxin. *Mol Microbiol* 2003;49(1):235–240.
355. Opanashuk LA, Mark RJ, Porter J, Damm D, Mattson MP, Seroogy KB. Heparin-binding epidermal growth factor-like growth factor in hippocampus: modulation of expression by seizures and anti-excitotoxic action. *J Neurosci* 1999;19(1):133–146.
356. Luciani A, Olivier JC, Clement O, Siauve N, Brillet PY, Bessoud B, Gazeau F, Uchegbu IF, Kahn E, Frija G, Cuenod CA. Glucose-receptor MR imaging of tumors: study in mice with PEGylated paramagnetic niosomes. *Radiology* 2004;231(1):135–142.
357. de Boer AG, Gaillard PJ. *In vitro* models of the blood–brain barrier: when to use which? *Curr Med Chem Central Nervous Syst Agents* 2002;2(3):203–209.
358. de Boer AG, Gaillard PJ. *In vivo* methods to estimate drug transport to the brain across the blood–brain barrier. *Med Chem Rev Online* 2005;2(2):127–131.
359. Plum L, Schubert M, Brüning JC. The role of insulin receptor signaling in the brain. *Trends Endocrinol Metab* 2005;16(2):59–65.

# TRANSPORTER INTERACTIONS IN THE ADME PATHWAY OF DRUGS

Yan Zhang[1] and Donald W. Miller[2]

[1]*Drug Metabolism and Biopharmaceutics, Incyte Corporation, Wilmington, Delaware*
[2]*University of Manitoba, Winnipeg, Manitoba, Canada*

**Contents**

1 Introduction
2 Transport Systems Impacting Drug ADME
3 Computational Modeling of Transporter Interactions
  3.1 Transporter-Based Modeling
  3.2 Substrate-Based Modeling
  3.3 Applied Computer Modeling Systems in ADME
4 *In Vitro* Systems for Identification of Transporter Interactions
  4.1 *In Vitro* Accumulation Assays for Identification of Transporter Processes
  4.2 *In Vitro* Permeability Assays for Identification of Transporter Processes
5 *In Vivo* Methods for Assessing Transport Interactions
6 Conclusion
  References

## 1 INTRODUCTION

Approximately 30% of the human genome encodes membrane proteins [1]. It has been estimated that as many as 1200 of these genes code for potential drug transporters [2]. Given the number of transport proteins expressed on cell membranes,

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

it stands to reason that interactions of therapeutic agents with endogenous transporter systems could have a major impact on the absorption, distribution, metabolism, and excretion (ADME) properties of drugs. Transporter interactions can impact on the ADME pathway in two distinct and important ways. First, if a drug is a substrate for a specific transporter involved in absorption, distribution, or elimination, the transporter can influence the ADME properties of the drug itself. Examples of this include increased intestinal absorption of cephalosporins through transport by a peptide transporter in the intestine [3, 4], and decreased intestinal absorption of taxol by the efflux transporter, P-glycoprotein [5, 6]. Second, a drug may act as either an inhibitor or inducer of a selective transporter involved in the ADME properties, thereby altering the ADME properties of another drug. In this case, the result would be a drug–drug interaction that changes the pharmacokinetics of the drug. Athough such interactions were in the past considered to be entirely drug metabolism (cytochrome P450-mediated) related, it is now recognized that transporters may also be involved [7, 8].

Identification of the transport properties of compounds during the drug discovery process is essential for optimization of the ADME properties of therapeutic agents. In this regard, the pharmaceutical scientist has a variety of tools with which to probe transporter interactions of compounds. These include cell-based uptake, absorption and transport assays, *in situ* models that can isolate the pertinent absorption, distribution, or elimination tissues, and computational models that can simulate and/or predict transporter interactions. In addition, transgenic mouse models have made it possible to examine the impact of selected transport systems on the ADME process at the whole animal level. This chapter identifies some of the more important transporters that can influence the ADME properties of drugs. In addition, a strategy for evaluating compounds for transporter interactions is discussed that incorporates *in silico*-based modeling and *in vitro* and *in vivo* transporter studies.

## 2 TRANSPORT SYSTEMS IMPACTING DRUG ADME

There are several types of transport systems involved in moving solutes and macromolecules across cellular barriers. These include primary active transport systems that rely directly on ATP hydrolysis to move solutes across the membrane. Examples include the drug efflux transporters such as P-glycoprotein (P-gp), multidrug resistance associated protein (MRP), and breast cancer resistant protein (BCRP). Transporters in this group belong to the ATP binding cassette (ABC) superfamily of proteins [9–11]. In general, these transporters consist of 6–17 membrane spanning domains, one or more nucleotide binding domains, and one or more drug binding domains [10–12]. As these transporters rely directly on ATP hydrolysis, they have the capability of transporting the solutes and macromolecules against a concentration gradient. In mammalian cells, the direction of transport is usually directed toward removing compounds from the cell (i.e., outwardly directed).

Drug efflux transporters have been closely linked with the development of drug resistance in cancer [9, 10], bacterial [13], and parasitic [14] cells. These same drug efflux transporters are also present in normal tissue such as epithelial cells of the intestine, liver, and kidney, and endothelial cells forming the blood–brain barrier [15–18]. While the drug efflux transporters serve an important protective function,

preventing the absorption and distribution and speeding the elimination of potentially toxic xenobiotics, due to their rather promiscuous nature (i.e., ability to transport a wide variety of chemical compositions), they also have a substantial impact on the ADME properties of drugs. P-glycoprotein, the most widely studied drug efflux transporter, contributes to the low oral bioavailability of several drugs [19]. Furthermore, even those drugs that can overcome P-gp efflux in the intestine are still subject to restricted tissue distribution and increased clearance [20]. For these reasons, much effort has been placed on the identification of potential drug efflux transporter interactions early in the drug development process.

Another type of transport system present in the cell is secondary active transporters. In contrast to primary active transporters that rely on ATP hydrolysis, the secondary active transporters utilize electrochemical gradients to supply the energy required to move the solutes into or out of the cell. Thus, the "energy" for these transporters comes from the movement of ions, typically either sodium or hydrogen, down their concentration gradient. Examples of secondary active transporters include the sodium-dependent amino acid transporters found in the intestine and blood–brain barrier [21, 22], and the hydrogen ion-dependent peptide transporters (PepT1 and PepT2) found in the intestine, kidney, lung, and blood–cerebrospinal fluid barrier [23–25].

From a drug transport perspective, PepT1 and PepT2 deserve special mention. Both of these transporters were the first mammalian transporters identified that use the electrochemical proton gradient as a driving force [26]. The "preferred" substrates for PepT1 and PepT2 are di- and tripeptides [27]. This property of PepT transporters has been exploited to aid in the absorption of peptide and peptidomimetic agents in the intestine and lung [28, 29]. An example of this is the angiotensin converting enzyme (ACE) inhibitors [28]. However, unlike many of the solute and nutrient transporters that have relatively rigid structural requirements for transport substrates, the PepT transporters accommodate a wide variety of compounds. Consequently, not only are PepT1 and PepT2 able to aid in the intestinal absorption (PepT1) and kidney reabsorption (PepT2) of peptidomimetic drugs [28], but they can also influence the ADME properties of antibiotics such as cephalosporin [3, 4], and nucleoside-based antiviral prodrugs such as valacyclovir [30]. Although PepT1 and PepT2 share the same set of substrates, they differ in their affinity and capacity. Whereas PepT1 is a low affinity, high capacity transporter, PepT2 has high affinity and low capacity for the same substrates [23]. In addition, the presence of PepT2 on pulmonary epithelial cells [24, 29, 31] and epithelial cells of the choroid plexus, which form the blood–cerebrospinal fluid barrier [25], suggests this transporter could also influence drug distribution of selected therapeutic agents in these tissues.

A third type of transporter class is the solute carriers that allow the passage of compounds across the cellular barrier through a process involving facilitated diffusion [32]. In facilitated diffusion, the transported molecule interacts with the protein carrier, and through conformational changes in the protein, the molecule is allowed to pass through the lipid membrane. Cellular uptake of many nutrients such as glucose [33], various nucleotides [34, 35], and creatine [36] are known to occur through facilitated diffusion processes. These carriers display the same characteristics of saturability and solute selectivity as observed with the active transporters. However, unlike primary or secondary active transport systems, the carriers involved

in facilitated diffusion processes are not sensitive to metabolic inhibitors. In addition, the carriers involved in facilitated diffusion are bidirectional in that the solute will follow a concentration gradient whether that is directed into or out of the cell. While such carriers are important for nutrient absorption and distribution, the impact of facilitated diffusion on ADME properties of drugs is minimal, due in large part to the stringent structural requirements for these transporters.

## 3 COMPUTATIONAL MODELING OF TRANSPORTER INTERACTIONS

A number of criteria have to be met to successfully develop a drug-like compound. It has to be absorbed into the bloodstream and be distributed and bound to targeted tissues. Furthermore, in order to avoid toxic effects, many drugs require biotransformation for efficient elimination from the body. All of these aspects are significantly mediated or influenced, at some level, by transporter proteins. This has made the identification of transporter interactions a very important part of the ADME process. The following section summarizes the various options available for examining transport-mediated influences in the ADME pathway.

Prediction of the passive permeability of a compound using computational modeling is relatively straightforward and is aided by a very large training set of compounds with established physicochemical properties (i.e., lipophilicity, H-bonding potential, solubility, and stability) with which to generate the computer simulations. However, prediction of permeability when transporter interactions are involved is much more difficult. Indeed, when correlating actual permeability as accessed in either cell based or *in vivo* systems, to that predicted by computer modeling, it is often the outliers that have either metabolism or transport issues influencing permeability. Despite the added complexity, computational modeling of transporter interactions is an area of intense interest. Numerous efforts have been made to characterize various transporter/carrier proteins, including the PepT1, and the drug efflux pumps such as P-gp and MRP. However, little is known about these proteins at the molecular level due to the limited availability of high resolution crystal structures. Determination of crystal structure for transport proteins is important not only for understanding the function of these transporters but also for defining the structure–activity properties required for compounds transported by specific systems. The two general approaches used to model drug transporter systems include transporter-based modeling and substrate-based modeling [37]. Each of these approaches to computer modeling of transporter interactions is described.

### 3.1 Transporter-Based Modeling

The transporter-based approach focuses on the three-dimensional structure of the transport protein to assist in understanding the drug transport process. This approach uses the known or predicted three-dimensional structure of the protein to map out the topography of the transporter in the cell membrane. As would be predicted, the better the information is concerning the three-dimensional structure of the transporter being examined, the better the resulting computational model will be. However, complete knowledge of the three-dimensional structure is not essential

as a transporter-based approach can be applied when the target transporter shares the same number of transmembrane domains (TMDs) and an adequate level (generally >30%) of amino acid sequence identity with a second transporter, that has a known crystal structure available. This approach can be used even when the two transporters are not functionally related. For example, modeling of the apical sodium-dependent bile acid transporter (ASBT) was based on the crystal structure of bacteriorhodopsin [38]. In addition, the three-dimensional structure of the human facilitative glucose transporter (GLUT1) was predicted by using the crystal structure of a bacterial transporter, glycerol-3-phosphate transporter (GlpT), as a template [39]. Furthermore, important insights regarding the functionality of a target transporter can still be generated by modeling partial domains of the transporter when there is an absence of appropriate templates for a full-length transporter [37]. In studying the sulfate transporter, Rouached et al. [40] avoided modeling the transmembrane domains in their entirety because of the lack of appropriate templates. However, they identified the bacterial SpoIIAA protein as a template for the functionally unknown C-terminal sulfate transporter and anti-sigma antagonist (STAS) domain and directly generated a three-dimensional model for this domain [40].

In the transporter-based approach, the structure of a target transporter can be calculated using homology modeling techniques, including several commercially available programs, such as Modeller [41], InsightII homology [42], and ICM (MolSoft, San Diego, CA) [37]. Among these, Modeller is the most widely applied program. It follows four general steps, including template selection, sequence alignment, model generation, and model validation. These four steps are repeated as a cycle until no further improvements to the model are observed. Among the four steps, validation is an essential process to ensure accuracy of the generated model. Modeller generates several models based on the sequence alignment, which are, in turn, evaluated by internal self-consistency checks and external programs that verify structural protein quality. Examples of external programs used to verify protein structures are Procheck [43] and WHAT IF [44]. Generally, these programs develop rules and parameters extracted from available crystal structures and apply these to the model structure. The fitness that is represented by a "fit score" indicates the compatibility of the model to the available crystal structures [37]. The transporter-based modeling approach now enables the elucidation of three-dimensional structures for a large number of transporters that would otherwise not be available to scientists. Examples include the glucose transporter (GLUT1) [39], P-gp [45], and MRP1 [46].

Both P-gp and MRP1 belong to the ABC superfamily of transporters that have been extensively studied. The initial models of both P-gp and MRP are based on the crystal structures of bacterial transporters. P-gp was initially modeled using the crystal structure of the *Escherichia coli* lipid A transporter (MsbA), which shares ~30% homology with P-gp [45]. Subsequently, Stenham and colleagues [47] constructed an atomic detail model for P-gp based on assessment of the crystal structure of MsbA in relation to their disulfide cross-linking data. The two diverging models of P-gp were later harmonized in accordance with a theory proposed by Lee et al. [48], who postulated that P-gp could exist in two dynamic conformations—an open conformation [45] and a closed conformation [47]. Similarly, MRP1 was modeled based on the crystal structure of *Vibrio cholerae* MsbA [46]. The model implied that Phe594 forms a hydrophobic pocket with four residues, therefore successfully

predicting the functional importance of the Phe594 in transmembrane helix 11 of the MRP1 [46].

The transporter-based approach provides detailed protein structural information. Based on the topographical characterization of the protein, one can identify potential areas/pockets where substrate and/or inhibitor interact with the transporter proteins. It should be noted that this approach is not feasible for proteins without a suitable TMD template. To date, this approach has been applied mostly to transporters possessing 7 or 12 TMDs with the exception of MRP1, which has 17 TMDs [37]. For those transporters for which no suitable TMD template is available, techniques such as substrate-based methods that do not require knowledge of transporter structure can be applied.

### 3.2 Substrate-Based Modeling

Substrate-based modeling correlates biological activity of a series of substrates or inhibitors for a transporter with their molecular descriptors or chemical features. Based on the compiled properties of the substrates and inhibitors, the models can then predict potential transporter interactions. This approach has been applied successfully to generate pharmacophore (the spatial arrangement of the hydrophobicity, hydrogen bonding status, charged state, and other features of a compound) and quantitative structure–activity relationship (QSAR) models for many membrane transporter proteins, such as P-gp [49], MRP2 [50], human peptide transporter 1 (hPEPT1) [51, 52], organic cation transporters (OCTs) [53], organic anion transporter polypeptides (OATPs) [54], and nucleoside transporter [55, 56].

Several substrate-based computational programs are commercially available to generate these modelings. The most widely used programs include CATALYST (Accelrys Inc., San Diego, CA) and distance comparison (DISCO) [57] for pharmacophore generation. For more quantitative structure–activity relationship (QSAR) predictions, comparative molecular field analysis (CoMFA) [58] and comparative molecular similarity index analysis (CoMSIA) [59] are commercially available modeling programs. Both CATALYST and DISCO programs attempt to determine common features based on the superposition of active compounds (i.e., transporter substrates or inhibitors) using different algorithms [37]. For example, HIPHOP is one of the algorithms of CATALYST which generates a pharmacophore based on common chemical features among active molecules. In an effort to identify novel hPEPT1 inhibitors, Ekins and co-sworkers [52] used three relatively high affinity ligands—Gly-Sar, bestatin, and enalapril—to generate a common features (HIPHOP) pharmacophore. This consisted of two hydrophobic features, a hydrogen bond donor, acceptor, and a negative ionizable feature. The pharmacophore was then used to search the Comprehensive Medicinal Chemistry (CMC) database of more than 8000 drug-like molecules and retrieved 145 virtual hits mapping to the pharmacophore features. Furthermore, the pharmacophore was also able to identify known hPEPT1 substrates and inhibitors in further database mining of more than 500 commonly prescribed drugs [52]. In contrast to pharmacophore modeling, a 3D-QSAR model consists of a mathematical equation describing potency as a function of 3D interaction fields around aligned training set compounds [60]. Both CoMFA and CoMSIA are commonly used algorithms for 3D-QSAR generation. They are based on the assumption that changes in binding affinities of ligands are related to changes in

molecular properties, represented by interaction fields [58, 59]. Recently, Biegel et al. [61] determined the structural requirements for the substrates of the renal type peptide transporter (PEPT2) by 3D-QSAR using the CoMSIA program. The model was built based on 83 compounds (including dipeptides, tripeptides, and β-lactam antibiotics) covering a wide range of affinity constants. An analysis of the contour maps provided by the CoMSIA method gave insights into the requirements for high affinity substrates of PEPT2. Furthermore, an additional 3D-QSAR model based on the same compounds was generated and correlated with affinity data of the intestinal peptide transporter (PEPT1). By comparing the CoMSIA contour plots, differences in substrate selectivity between PEPT1 and PEPT2 were determined [61].

Substrate-based modeling can often provide insight in the ligand–protein binding or inhibition process and assist in the identification of other molecules that contain the same pharmacophore features. Furthermore, the correlation between the variations in ligand or inhibitor binding affinities and the variations in their respective structural features can assist in the design of more potent inhibitors or substrates with higher affinity. However, disadvantages are also present in the substrate-based modeling approach. First, the training datasets that the models are based on are generated from different sources, that is, different cell type, species, and laboratories. Therefore, the data are often not directly comparable, and the quality and consistency of the data might also be in question. Second, although a large body of literature involving the QSAR of various transporters can be found, it is often based on datasets comprised of transport inhibitors, not necessarily transport substrates. This is especially true for the QSAR modeling of P-gp and MRP, as much of the early work was aimed at finding compounds that could inhibit these transporters and thus reverse multidrug resistance [62]. The major reason inhibition data is frequently used to establish the computational models is because generation of this data is relatively higher throughput and simpler in experimental design compared to determination of the kinetic data ($K_m$, $V_{max}$) for transported substrates [37].

### 3.3 Applied Computer Modeling Systems in ADME

Within the ADME process, absorption and elimination (especially through metabolic pathways) have been the most extensively modeled. With respect to absorption, most of the current computational approaches used to predict absorption are based on the assumption that absorption is a passive process. In contrast, computer models for predicting elimination pathways place more emphasis on active processes such as enzyme–drug interactions. Thus, examination of both processes provides a very diverse exposure to the computational models currently employed in early ADME studies.

One of the simplest and most widely used computational approaches to estimate passive absorption or permeation was developed by Lipinski and co-workers in 1997 [63]. The authors selected a compound library consisting of 2287 drug candidates with good absorption properties according to the data of intestinal epithelial permeability and solubility. Based on the physicochemical properties of these compounds, the authors proposed the "rule of five" (ROF), taking into account hydrogen bonding, molecular weight, and lipophilicity.

Although the ROF allows quick screening of many NCEs and provides an initial indication of human intestinal absorption (HIA) of those compounds, there are some weaknesses associated with this model. First, the ROF does not apply to compounds that are subject to active transport processes. For example, therapeutic classes, such as antibiotics, antifungals, vitamins, and cardiac glycosides, consistently fall outside the ROF parameter cutoffs [63] and yet are effectively absorbed in the GI system via various transport systems. Second, the model was created using those compounds with high absorption as its database; only a limited number of compounds with a low absorption were included.

While the ROF is a model more based on the physicochemical properties of a compound, there are several physiological based models that also have great potential for use within both the early and later stages of drug discovery. Two examples of these models are iDEA™ (*in vitro* determination for the estimation of ADME) and GastroPlus™. GastroPlus was developed by Simulations Plus, Inc. (Lancaster, CA; http://www.simulations-plus.com/). It simulates GI absorption and pharmacokinetics for drugs dosed orally or intravenously in humans and animals. The simulation is based on the advanced compartmental absorption and transit (ACAT) model [64], which is an advanced version of the one originally reported by Yu and Amidon [65]. The ACAT model is a flexible physiologically based simulation with nine compartments corresponding to different segments of the digestive tract including stomach, seven small intestinal compartments, and colon. Each compartment is modeled with accurate physiological information regarding volumes, transit times, length, and radius [64]. This model takes into account pH dependency of basic parameters, solubility and permeability, transport of drug material through the GI tract, and absorption of drug through the intestinal wall into the portal vein. In addition to these, GastroPlus provides the ability for customization by allowing absorption constants to be set individually for each intestinal compartment [66]. With this physiologically based simulation software program, the rate, extent, and gastrointestinal location of oral drug absorption can be predicted [67]. For predictions based on measured permeability, the GastroPlus model has been trained to accept values for human jejunal permeability as input. The determination for human permeability ($P_{eff}$) is based on a combination of experimental data inputs from both *in vivo* human values and *in situ* rat or dog intestinal permeability values. These values then further convert to human values using a built-in correlation [67, 68]. However, because there is no relation between the Caco-2 data and the human $P_{eff}$ with this model, human $P_{eff}$ cannot be predicted using Caco-2 data [67]. Therefore, a preliminary step is transformation based on a correlation when Caco-2 data is used. Normally, a correlation of human $P_{eff}$ against Caco-2 permeability can be built using reference compounds by users before the simulation. By doing this, the Caco-2 permeability is converted to effective human jejunal permeability, $P_{eff}$, based on the established correlation [68].

Another computational model for the simulation of absorption is iDEA, a predictive ADME simulation system developed by Lion Bioscience Inc. (San Diego, CA; http://www.lionbioscience.com/). This particular model was developed based on a proprietary database that coupled oral bioavailability data from human clinical trials with preclinical *in vitro* data characterizing various parameters of the drug including solubility, Caco-2 permeability, protein binding, and metabolic stability in human cryopreserved hepatocytes. This model can simulate human physiology and

accounts for regional variations in intestinal permeability, solubility, surface area, and fluid movement [66, 68]. iDEA is based on the STELLA (Structural Thinking Experimental Learning Laboratory with Animation) simulation software and comprises structure-based Caco-2 and absorption modules (SBM) and a physiologically based (PBM) absorption module [69, 70]. Therefore, it allows the prediction of Caco-2 permeability or the fraction of dose absorbed ($F_a$) of drugs from their structures. It also predicts the rate and extent of human oral drug absorption, by taking into account physiological parameters such as intestinal pH, transit time, and blood flow. Since there are large interlaboratory differences in the absolute values of Caco-2 permeability, it may be necessary to "customize" the simulation package to the individual laboratory. This is done by determining the Caco-2 permeability of several marker compounds for comparison the Caco-2 values used in the simulation software package. In doing this, one can determine how Caco-2 permeability values from Lion Bioscience Inc. correlate with the in-house Caco-2 permeability data and modify the simulation model accordingly.

Overall, the computational approaches for predicting oral drug absorption are still in their infancy. Neither the GastroPlus nor the iDEA system has been evaluated completely. Part of the issues with evaluating the extent to which either model predicts drug absorption in the GI tract is the limited access to a large data bank for fraction of dose absorbed. While such information is often available in-house, public access is limited by most pharmaceutical companies due to issues of confidentiality. An inherent limitation with the iDEA system is that compounds suspected to be metabolized or substrates of active transporters or carriers in the intestine have been excluded from model building [67]. To this extent, GastroPlus shows strengths in its ability to integrate additional data, including transport/efflux processes and intestinal metabolism [68]. Recently, Tubic et al. [71] demonstrated that P-gp was involved in the talinolol nonlinear dose-dependent absorption using GastroPlus simulation modeling. Furthermore, it was observed that the results predicted from the model were comparable to the finding of a Phase I dose escalation study [71].

Compared to GastroPlus, iDEA has a simpler operating structure. Once the server is installed, it runs as a web application on the corporate Intranet. Therefore, the deployment is simple and training is not a big issue. In this respect, iDEA could be aimed at use by non-DMPK specialists (e.g., medicinal chemists), who could access the absorption potential of compounds with a quick and easy tool. In contrast to the iDEA system, GastroPlus is well designed and suitable for those scientist dealing on a day-to-day basis with intestinal absorption issues. Because of the more complex modeling parameters, GastroPlus requires more thorough training compared to the iDEA system [68].

## 4 *IN VITRO* SYSTEMS FOR IDENTIFICATION OF TRANSPORTER INTERACTIONS

Identification of transporter interactions with drugs and determination of their impact on the ADME profile at an early stage in drug development has resulted in increased emphasis being placed on relatively high throughput *in vitro* screening methods. Commonly used cell-based approaches for examining transporter

**TABLE 1 Cell-Based Approaches for Examining Transporter Interactions in the ADME Process**

| Cell System | ADME Property | Transporter Applications | Experimental Design |
|---|---|---|---|
| Intestinal brush border membrane vesicles | Intestinal absorption | • Identification of substrates/inhibitors<br>• Determination of transporter kinetics | • Uptake studies |
| Caco-2 cell line | Intestinal absorption | • Identification of substrates/inhibitors<br>• Determination of transporter kinetics<br>• Determination of transporter polarity | • Uptake studies<br>• Permeability studies<br>• Bidirectional permeability studies |
| Isolated brain microvessel endothelial cells | Blood–brain barrier (brain distribution) | • Identification of substrate/inhibitors<br>• Determination of transporter kinetics | • Uptake studies |
| Cultured brain microvessel endothelial cells | Blood–brain barrier (brain distribution) | • Identification of substrate/inhibitors<br>• Determination of transporter kinetics<br>• Determination of transporter polarity | • Uptake studies<br>• Permeability studies<br>• Bidirectional permeability studies |
| Liver canalicular membrane vesicles | Hepatic/biliary excretion | • Identification of substrates/inhibitors<br>• Determination of transporter kinetics | • Uptake studies |
| Cultured hepatocytes | Hepatic/biliary excretion | • Identification of substrate/inhibitors<br>• Determination of transporter kinetics<br>• Determination of transporter polarity | • Uptake studies<br>• Permeability studies<br>• Bidirectional permeability studies |
| Madin–Darby canine kidney (MDCK) cell line | General absorption, distribution, and excretion | • Identification of substrate/inhibitors<br>• Determination of transporter kinetics<br>• Determination of transporter polarity | • Uptake studies<br>• Permeability studies<br>• Bidirectional permeability studies |

interactions and their applications in ADME processes are listed in Table 1. These include isolated cell or membrane preparations from various tissues. Examples are the use of isolated brain microvessel endothelial cells to study transporters at the blood–brain barrier [72, 73] and the use of membrane vesicles prepared from the brush-border membrane of intestinal epithelial cells [74] or the canalicular membrane of liver hepatocytes [75] to study transporter processes in the intestine and liver, respectively. While isolated cell and cell membranes typically have the highest levels of transporter expression, use of these preparations requires a great deal of starting tissue and determination of actual permeability

and polarized transport is difficult. In addition, assay validation with the isolated cell and cell membrane preparations is difficult. Thus, efforts have been made to develop cell culture systems to model the different ADME transport and permeability barriers (see Table 1). The advantages of cell culture systems are that one can examine polarized transporter processes and actually evaluate both cellular accumulation and permeability across confluent monolayers. Thus, the influence of transporters on the actual permeability of a compound can be determined. For those cell culture systems that utilize established, immortalized cells lines (i.e., Caco-2, MDCK) there are the added advantages of reproducibility and low interassay variability. In addition, the cell lines can also be transfected to express specific transporters of interest. Examples are the MDCK cell lines that have been transfected with MDR1, MRP1, or MRP2 drug efflux transporters [76–78]. The availability of the transfected cell lines provides important research tools for honing in on transporter dynamics and function.

### 4.1 *In Vitro* Accumulation Assays for Identification of Transporter Processes

Experiments examining the transport properties of compounds in cell-based systems fall into three major categories. These include studies based on cell or vesicle accumulation, unidirectional transport, or bidirectional transport. Studies examining the cellular or membrane vesicle accumulation of a compound can provide a fast and reliable method for identifying both transporter substrates and inhibitors. An example of this is the use of various membrane vesicle preparations to identify the transporter-mediated uptake and efflux of the lipid lowering drug pitavastatin [79]. These studies examined both the acid and lactone forms of pitavastatin. In the case of the acid form, uptake into the cell was dependent on multiple solute transporters, including members of the organic anion transporter (OAT) and the organic anion transporter polypeptide (OATP) family. In contrast, the lactone form was not a substrate for transport into the cell [79]. With regard to transporter-mediated efflux from the cell, the lactone form of pitavastatin was a substrate for P-gp, while the acid form displayed BCRP efflux activity [79]. These studies are a prime example of how vesicular and cellular accumulation can be used to delineate the structure–activity requirements for transport processes.

Cell accumulation assays are also helpful for identification of compounds that can inhibit drug transport. An example is the use of isolated brain capillaries and cultured brain microvessel endothelial cells in drug screening assays for drug efflux transporters [73, 80]. In these studies, changes in the cellular accumulation of fluorescent probes for P-gp and MRP were used to quantitatively assess the inhibitor properties of a series of compounds. Identification of potential inhibitors of drug efflux transport proteins, especially P-gp, are increasingly part of the early testing in the drug development process. This is due to the emphasis placed on knowing whether a compound is a P-gp inhibitor and/or inducer at the time a new drug application (NDA) is filed with the Food and Drug Administration (FDA). In addition, since many inhibitors of drug transport can also act as substrates, cellular accumulation assays are often used as a first screen for potential transporter interactions of compounds. Those agents showing activity can then be examined for substrate properties.

A final application for the cellular accumulation assays is quantitative determination of transporter kinetics. Using the cellular accumulation of fluorescent probes

for P-gp and MRP transport, Bachmeier et al. [81] determined and compared the Michaelis–Menten kinetics of these drug efflux transporters in primary cultured brain microvessel endothelial cells, MDR1 transfected MDCK cells, and freshly isolated brain microvessels. In these studies, P-gp related transport displayed similar $K_m$ values across the different BBB models; however, $V_{max}$ was significantly higher in the freshly isolated brain microvessels [81]. For MRP related transport, differences in both $K_m$ and $V_{max}$ were noted in the various BBB models [81].

### 4.2 *In Vitro* Permeability Assays for Identification of Transporter Processes

Drug interactions with transport systems can also be evaluated using actual cell monolayer permeability. These studies can be performed unidirectionally with and without various transport inhibitors to determine the impact of the transport system on actual permeability of the compound. This type of experiment typically examines the absorptive, distributional, or excretory transport. As with the cellular accumulation assays, transported compounds will display concentration-dependent monolayer permeability. An example is the permeability studies for the amino acid phenylalanine in Caco-2 monolayers [82]. These studies showed that the transport of L-phenylalanine from the apical to basolateral side of Caco-2 cells was concentration dependent. The transport process was saturable with a very high affinity ($K_m$ = 0.019 mM) to the amino acid transporter in Caco-2 cells [82]. Studies performed in the presence of specific transport inhibitors will also influence permeability. For absorptive transporters such as PepT, inhibition of the transporter results in reduced permeability in cell expression of PepT [82]. For drug efflux transporters, such as P-gp, inhibitors will increase the absorptive or distributive permeability of substrates [72, 78]. Alternatively, bidirectional permeability studies can be performed to identify potential transport systems. In these studies permeability is assessed in the apical-to-basolateral (A-B) direction and also in the basolateral-to-apical (B-A). By examining the resulting B-A/A-B ratio of the permeability coefficients, one can determine whether absorptive/uptake transporters (ratios significantly less than 1) or efflux transporters (ratios significantly greater than 1) are involved. Bidirectional permeability studies in either Caco-2 or MDCK MDR1 cells have been very useful for identifying those compounds that are transported by drug efflux transport proteins [83, 84]. While most commonly used to identify secretory transport systems, bidirectional permeability studies are also useful for the identification of absorptive transport properties as well. Studies by Li and colleagues [85] used bidirectional permeability in Caco-2 cell monolayers to aid in the identification of amino acid prodrugs of the antiviral levovirin, which utilize the PepT system. These studies were able to identify ester-based amino acid prodrugs with PepT substrate and inhibitor properties. Such compounds are anticipated to result in improved oral bioavailability for levovirin.

## 5 *IN VIVO* METHODS FOR ASSESSING TRANSPORT INTERACTIONS

Examination of drug transporter interactions at the whole animal level is much more complex than cell-based assays. However complex, ultimate determination of

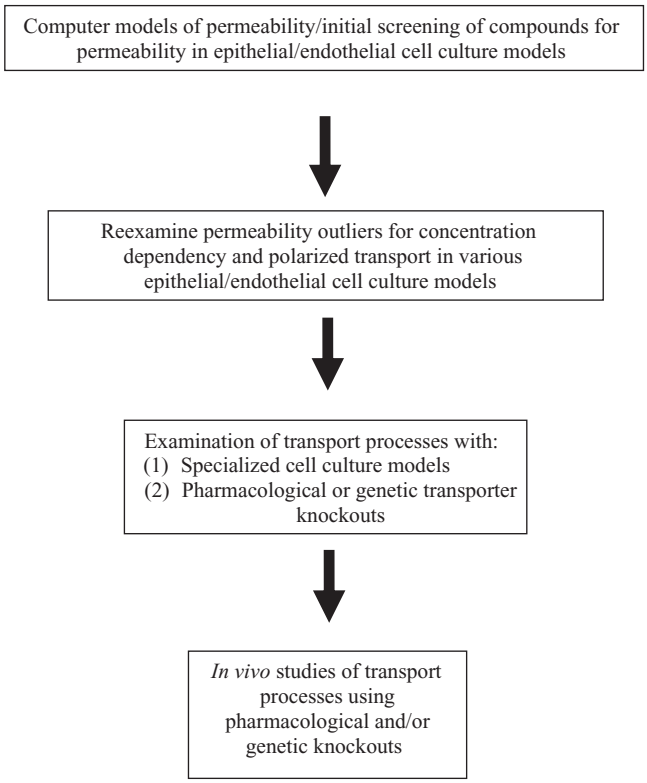


**FIGURE 1** Experimental process for determining transporter-related processes in ADME.

the impact that transporters have on the ADME properties of a drug require whole animal studies. Thus, *in vivo* studies examining transporter influences on the ADME properties of a drug are typically performed after extensive *in vitro* evaluation of the transport properties of the lead compound(s) (see Fig. 1). The exception to this is the cassette dosing of drugs to rodents during the early drug development stage. In these studies, a series of compounds are administered together, to obtain an early glimpse of the ADME properties in the *in vivo* setting [86–88]. However, it should be noted that performing the cassette dosing at both low and high doses of the compounds can provide preliminary indications of drug transporter interactions, as drugs that are substrates for transporters will demonstrate concentration dependency [88].

The advantage to performing *in vivo* studies during the final stages of the preclinical development of a drug is that a great deal of information concerning drug transporter interactions has already been identified. This allows the *in vivo* studies to hone in on the appropriate transporter using either pharmacological or genetic inhibition of transport properties. In the past, *in vivo* demonstration of transport relied on dose dependency and the use of pharmacological agents that inhibited transport processes. However, through advances in molecular biology, researchers now have the ability to selectively "knock out" drug transporters either through

creation of transgenic animals [89] or, more recently, through the use of small inhibitory RNA (siRNA) [90], to transiently knock down transporter expression. An example of the power that transgenic knockouts bring to *in vivo* studies of transporter interactions in the ADME process can be found in the MDR1 transgenic knockout mice [89, 91]. Studies with these mice have established quite conclusively the influence that P-glycoprotein has in the oral absorption, brain penetration, and elimination of selected drugs [91]. Similar examples can be found with the PEPT2 knockout mice, where levels of model PEPT2 substrates have been reduced in blood and cerebral spinal fluid owing to the lack of transporter at the kidney and blood–cerebrospinal fluid barriers [92].

The use of transgenic animals for studying ADME processes is not without its limitations. This is especially true when dealing with transporters. The main problems with transgenic animals in the study of ADME properties of a drug revolve around the related issues of transporter promiscuity and transporter compensation. For many transporters, there are various subtypes or homologs. Examples include the MRP drug efflux transporter family with over nine different homologs [93–95], and the organic anion transporter (OAT) with at least five homologs in humans [96]. Given the overlapping substrate and inhibitor properties of these transporters [17, 95] it is no wonder that transgenic knockout mice for single homologs of *mrp* have not yielded the dramatic results observed with *mdr1* knockout mice [97, 98]. The other issue to keep in mind is the upregulation of compensatory transport systems in the various transgenic knockout models. For example, *mdr1* knockout mice deficient in P-gp appear to have enhanced BCRP transport. This is suggested by the studies of Salama and colleagues [99], which evaluated the tissue accumulation of the HIV protease inhibitor nelfinavir, in wild-type and *mdr1* knockout mice. While significant increases in the brain accumulation of nelfinavir were observed in the *mdr1* knockout mice compared to wt controls, an even greater increase was observed when the P-gp/BCRP inhibitor GF120918 was administered to the knockout mice. As these mice had no P-gp, the increased brain accumulation of nelfinavir in the *mdr1* knockout mice exposed to GF120918 is likely due to inhibition of BCRP. These findings are consistent with the studies of Cisternino and colleagues [100] that show an increased expression of BCRP in brain microvessel endothelial cells from *mdr1* knockout mice compared to wt mice, and suggest compensatory increases in BCRP can occur in selected tissues as a result of P-gp deficiency.

*In vivo* application of siRNA to selectively knockdown transporter expression is relatively new. The technique involves the introduction of single chain oligonucleotides into the cell that bind to specific single chain RNA sequences preventing transcription into protein [90]. The use of siRNA to knock down the expression of a variety of protein targets has been demonstrated in cell culture systems [90]. While there are reports in the literature of successful use of siRNA technology *in vivo*, applications to transporter relevant topics are lacking. The main obstacle to application of siRNA technology for the study of ADME-related transport processes is the effective delivery of siRNA to the relevant tissue/organs [101]. However, as methods for the efficient introduction of siRNA into targeted tissue improve, it is highly likely that this technique will have future applications for ADME-related studies in the whole animal.

## 6 CONCLUSION

The increasing cost of drug development is partially due to the failure to identify undesirable compounds at an early enough stage of the drug discovery process. Therefore, development of tools capable of predicting the ADME (absorption, distribution, metabolism, and excretion) properties of the new chemical entities (NCEs) becomes a significant way to reduce the cost for drug failure and to improve the chances of designing drugs with optimal ADME characteristics. In this regard, *in vitro* models capable of screening large numbers of compounds and computational (*in silico*) based models have greatly streamlined the early ADME process. This is especially true for identification of transporter interactions at the early stages of drug development. Using a rationale approach, such as that outlined in Fig. 1, that incorporates early *in silico* and cell-based assays to identify potential compounds that interact with transport systems coupled with *in vivo* studies of the ADME properties of identified compounds, one can optimize the identification of transporter-related issues and evaluate their potential impact on the ADME properties of drugs in the developmental pipeline.

## REFERENCES


1. Yeh AP, McMillan A, Stowell MH. Rapid and simple protein-stability screens: application to membrane proteins. *Acta Crystallogr D Biol Crystallogr* 2006;62:451–457.
2. Sakaeda T, Nakamura T, Okumura K. Pharmacogenetics of drug transporters and its impact on the pharmacotherapy. *Curr Top Med Chem* 2004;4:1385–1398.
3. Boll M, Markovich D, Weber WM, Korte H, Daniel H, Murer H. Expression cloning of a cDNA from rabbit small intestine related to proton-coupled transport of peptides, beta-lactam antibiotics and ACE-inhibitors. *Pflugers Arch* 1994;429:146–149.
4. Ganapathy ME, Prasad PD, Mackenzie B, Ganapathy V, Leibach FH. Interaction of anionic cephalosporins with the intestinal and renal peptide transporters PEPT 1 and PEPT 2. *Biochim Biophys Acta* 1997;1324:296–308.
5. Malingre MM, Richel DJ, Beijnen JH, Rosing H, Koopman FJ, Ten Bokkel Huinink WW, Schot ME, Schellens JH. Coadministration of cyclosporine strongly enhances the oral bioavailability of docetaxel. *J Clin Oncol* 2001;19:1160–1166.
6. Chiou WL, Wu TC, Ma C, Jeong HY. Enhanced oral bioavailability of docetaxel by coadministration of cyclosporine: quantitation and role of P-glycoprotein. *J Clin Oncol* 2002;20:1951–1952; author reply 1952.
7. Christians U, Schmitz V, Haschke M. Functional interactions between P-glycoprotein and CYP3A in drug metabolism. *Expert Opin Drug Metab Toxicol* 2005;1:641–654.
8. Holtzman CW, Wiggins BS, Spinler SA. Role of P-glycoprotein in statin drug interactions. *Pharmacotherapy* 2006;26:1601–1607.
9. Gottesman MM, Pastan I. Biochemistry of multidrug resistance mediated by the multidrug transporter. *Annu Rev Biochem* 1993;62:385–427.
10. Barrand MA, Bagrij T, Neo SY. Multidrug resistance-associated protein: a protein distinct from P-glycoprotein involved in cytotoxic drug expulsion. *Gen Pharmacol* 1997;28:639–645.
11. Bates SE, Robey R, Miyake K, Rao K, Ross DD, Litman T. The role of half-transporters in multidrug resistance. *J Bioenerg Biomembr* 2001;33:503–511.

12. Borst P, Evers R, Kool M, Wijnholds J. A family of drug transporters: the multidrug resistance-associated proteins. *J Natl Cancer Inst* 2000;92:1295–1302.
13. van Veen HW, Konings WN. Multidrug transporters from bacteria to man: similarities in structure and function. *Semin Cancer Biol* 1997;8:183–191.
14. Klokouzas A, Shahi S, Hladky SB, Barrand MA, van Veen HW. ABC transporters and drug resistance in parasitic protozoa. *Int J Antimicrob Agents* 2003;22:301–317.
15. Thiebaut F, Tsuruo T, Hamada H, Gottesman MM, Pastan I, Willingham MC. Cellular localization of the multidrug-resistance gene product P-glycoprotein in normal human tissues. *Proc Natl Acad Sci USA* 1987;84:7735–7738.
16. Cordon-Cardo C, O'Brien JP, Casals D, Rittman-Grauer L, Biedler JL, Melamed MR, Bertino JR. Multidrug-resistance gene (P-glycoprotein) is expressed by endothelial cells at blood–brain barrier sites. *Proc Natl Acad Sci USA* 1989;86:695–698.
17. Borst P, Evers R, Kool M, Wijnholds J. The multidrug resistance protein family. *Biochim Biophys Acta* 1999;1461:347–357.
18. Seetharaman S, Barrand MA, Maskell L, Scheper RJ. Multidrug resistance-related transport proteins in isolated human brain microvessels and in cells cultured from these isolates. *J Neurochem* 1998;70:1151–1159.
19. Chan LM, Lowes S, Hirst BH. The ABCs of drug transport in intestine and liver: efflux proteins limiting drug absorption and bioavailability. *Eur J Pharm Sci* 2004;21:25–51.
20. Lin JH. Drug–drug interaction mediated by inhibition and induction of P-glycoprotein. *Adv Drug Deliv Rev* 2003;55:53–81.
21. Smith QR, Stoll J. Blood–brain barrier amino acid transport. In Pardridge WM, Ed. *Introduction to the Blood–Brain Barrier*. Cambridge: Cambridge University Press; 1988, pp 188–197.
22. Castagna M, Shayakul C, Trotti D, Sacchi VF, Harvey WR, Hediger MA. Molecular characteristics of mammalian and insect amino acid transporters: implications for amino acid homeostasis. *J Exp Biol* 1997;200:269–286.
23. Daniel H, Herget M. Cellular and molecular mechanisms of renal peptide transport. *Am J Physiol* 1997;273:F1–F8.
24. Groneberg DA, Nickolaus M, Springer J, Doring F, Daniel H, Fischer A. Localization of the peptide transporter PEPT2 in the lung: implications for pulmonary oligopeptide uptake. *Am J Pathol* 2001;158:707–714.
25. Teuscher NS, Novotny A, Keep RF, Smith DE. Functional evidence for presence of PEPT2 in rat choroid plexus: studies with glycylsarcosine. *J Pharmacol Exp Ther* 2000;294:494–499.
26. Paulsen IT, Skurray RA. The POT family of transport proteins. *Trends Biochem Sci* 1994;19:404.
27. Leibach FH, Ganapathy V. Peptide transporters in the intestine and the kidney. *Annu Rev Nutr* 1996;16:99–119.
28. Shu C, Shen H, Hopfer U, Smith DE. Mechanism of intestinal absorption and renal reabsorption of an orally active ace inhibitor: uptake and transport of fosinopril in cell cultures. *Drug Metab Dispos* 2001;29:1307–1315.
29. Groneberg DA, Fischer A, Chung KF, Daniel H. Molecular mechanisms of pulmonary peptidomimetic drug and peptide transport. *Am J Respir Cell Mol Biol* 2004;30: 251–260.
30. Ganapathy ME, Huang W, Wang H, Ganapathy V, Leibach FH. Valacyclovir: a substrate for the intestinal and renal peptide transporters PEPT1 and PEPT2. *Biochem Biophys Res Commun* 1998;246:470–475.

31. Groneberg DA, Eynott PR, Doring F, Dinh QT, Oates T, Barnes PJ, Chung KF, Daniel H, Fischer A. Distribution and function of the peptide transporter PEPT2 in normal and cystic fibrosis human lung. *Thorax* 2002;57:55–60.
32. Stein WD. Facilitated diffusion: the simple carrier. In Stein WD, Ed. *Transport and Diffusion Across Cell Membranes*. New York: Academic Press; 1986, pp 231–362.
33. Wright EM, van Os CH, Mircheff AK. Sugar uptake by intestinal basolateral membrane vesicles. *Biochim Biophys Acta* 1980;597:112–124.
34. Wohlhueter RM, Marz R, Plagemann PG. Thymidine transport in cultured mammalian cells. Kinetic analysis, temperature dependence and specificity of the transport system. *Biochim Biophys Acta* 1979;553:262–283.
35. Harley ER, Paterson AR, Cass CE. Initial rate kinetics of the transport of adenosine and 4-amino-7-(beta-D-ribofuranosyl)pyrrolo[2,3-d]pyrimidine (tubercidin) in cultured cells. *Cancer Res* 1982;42:1289–1295.
36. Ku CP, Passow H. Creatine and creatinine transport in old and young human red blood cells. *Biochim Biophys Acta* 1980;600:212–227.
37. Chang C, Swaan PW. Computational approaches to modeling drug transporters. *Eur J Pharm Sci* 2006;27:411–424.
38. Zhang EY, Phelps MA, Banerjee A, Khantwal CM, Chang C, Helsper F, Swaan PW. Topology scanning and putative three-dimensional structure of the extracellular binding domains of the apical sodium-dependent bile acid transporter (SLC10A2). *Biochemistry* 2004;43:11380–11392.
39. Salas-Burgos A, Iserovich P, Zuniga F, Vera JC, Fischbarg J. Predicting the three-dimensional structure of the human facilitative glucose transporter glut1 by a novel evolutionary homology strategy: insights on the molecular mechanism of substrate migration, and binding sites for glucose and inhibitory molecules. *Biophys J* 2004;87:2990–2999.
40. Rouached H, Berthomieu P, El Kassis E, Cathala N, Catherinot V, Labesse G, Davidian JC, Fourcroy P. Structural and functional analysis of the C-terminal STAS (sulfate transporter and anti-sigma antagonist) domain of the Arabidopsis thaliana sulfate transporter SULTR1.2. *J Biol Chem* 2005;280:15976–15983.
41. Sali A, Blundell TL. Comparative protein modelling by satisfaction of spatial restraints. *J Mol Biol* 1993;234:779–815.
42. Greer J. Comparative modeling methods: application to the family of the mammalian serine proteases. *Proteins* 1990;7:317–334.
43. Laskowski RA, MacArthur MW, Moss DS, Thornton JM. PROCHECK: a program to check the stereochemical quality of protein structures. *J Appl Crystallogr* 1993;26: 283–291.
44. Vriend G. WHAT IF: a molecular modeling and drug design program. *J Mol Graph* 1990;8:52–56, 29.
45. Seigneuret M, Garnier-Suillerot A. A structural model for the open conformation of the mdr1 P-glycoprotein based on the MsbA crystal structure. *J Biol Chem* 2003;278: 30115–30124.
46. Campbell JD, Koike K, Moreau C, Sansom MS, Deeley RG, Cole SP. Molecular modeling correctly predicts the functional importance of Phe594 in transmembrane helix 11 of the multidrug resistance protein, MRP1 (ABCC1). *J Biol Chem* 2004;279: 463–468.
47. Stenham DR, Campbell JD, Sansom MS, Higgins CF, Kerr ID, Linton KJ. An atomic detail model for the human ATP binding cassette transporter P-glycoprotein derived from disulfide cross-linking and homology modeling. *FASEB J* 2003;17:2287–2289.

48. Lee JY, Urbatsch IL, Senior AE, Wilkens S. Projection structure of P-glycoprotein by electron microscopy. Evidence for a closed conformation of the nucleotide binding domains. *J Biol Chem* 2002;277:40125–40131.
49. Yates CR, Chang C, Kearbey JD, Yasuda K, Schuetz EG, Miller DD, Dalton JT, Swaan PW. Structural determinants of P-glycoprotein-mediated transport of glucocorticoids. *Pharm Res* 2003;20:1794–1803.
50. Ng C, Xiao YD, Lum BL, Han YH. Quantitative structure–activity relationships of methotrexate and methotrexate analogues transported by the rat multispecific resistance-associated protein 2 (rMrp2). *Eur J Pharm Sci* 2005;26:405–413.
51. Gebauer S, Knutter I, Hartrodt B, Brandsch M, Neubert K, Thondorf I. Three-dimensional quantitative structure–activity relationship analyses of peptide substrates of the mammalian $H^+$/peptide cotransporter PEPT1. *J Med Chem* 2003;46:5725–5734.
52. Ekins S, Johnston JS, Bahadduri P, D'Souza VM, Ray A, Chang C, Swaan PW. *In vitro* and pharmacophore-based discovery of novel hPEPT1 inhibitors. *Pharm Res* 2005;22:512–517.
53. Suhre WM, Ekins S, Chang C, Swaan PW, Wright SH. Molecular determinants of substrate/inhibitor binding to the human and rabbit renal organic cation transporters hOCT2 and rbOCT2. *Mol Pharmacol* 2005;67:1067–1077.
54. Chang C, Pang KS, Swaan PW, Ekins S. Comparative pharmacophore modeling of organic anion transporting polypeptides: a meta-analysis of rat Oatp1a1 and human OATP1B1. *J Pharmacol Exp Ther* 2005;314:533–541.
55. Chang C, Swaan PW, Ngo LY, Lum PY, Patil SD, Unadkat JD. Molecular requirements of the human nucleoside transporters hCNT1, hCNT2, and hENT1. *Mol Pharmacol* 2004;65:558–570.
56. Hu H, Endres CJ, Chang C, Umapathy NS, Lee EW, Fei YJ, Itagaki S, Swaan PW, Ganapathy V, Unadkat JD. Electrophysiological characterization and modeling of the structure–activity relationship of the human concentrative nucleoside transporter 3 (hCNT3). *Mol Pharmacol* 2006;69:1542–1553.
57. Martin YC, Bures MG, Danaher EA, DeLazzer J, Lico I, Pavlik PA. A fast new approach to pharmacophore mapping and its application to dopaminergic and benzodiazepine agonists. *J Comput Aided Mol Des* 1993;7:83–102.
58. Cramer RD 3rd, Patterson DE, Bunce JD. Recent advances in comparative molecular field analysis (CoMFA). *Prog Clin Biol Res* 1989;291:161–165.
59. Klebe G, Abraham U, Mietzner T. Molecular similarity indices in a comparative analysis (CoMSIA) of drug molecules to correlate and predict their biological activity. *J Med Chem* 1994;37:4130–4146.
60. Chang C, Ray A, Swaan P. *In silico* strategies for modeling membrane transporter function. *Drug Discov Today* 2005;10:663–671.
61. Biegel A, Gebauer S, Brandsch M, Neubert K, Thondorf I. Structural requirements for the substrates of the $H^+$/peptide cotransporter PEPT2 determined by three-dimensional quantitative structure–activity relationship analysis. *J Med Chem* 2006;49:4286–4296.
62. Stouch TR, Gudmundsson O. Progress in understanding the structure–activity relationships of P-glycoprotein. *Adv Drug Deliv Rev* 2002;54:315–328.
63. Lipinski CA, Lombardo F, Dominy BW, Feeney PJ. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv Drug Deliv Rev* 1997;23:3–25.
64. Agoram B, Woltosz WS, Bolger MB. Predicting the impact of physiological and biochemical processes on oral drug bioavailability. *Adv Drug Deliv Rev* 2001;50(Suppl 1): S41–S67.

65. Yu LX, Amidon GL. A compartmental absorption and transit model for estimating oral drug absorption. *Int J Pharm* 1999;186:119–125.
66. Boobis A, Gundert-Remy U, Kremers P, Macheras P, Pelkonen O. *In silico* prediction of ADME and pharmacokinetics. Report of an expert meeting organised by COST B15. *Eur J Pharm Sci* 2002;17:183–193.
67. Bohets H, Annaert P, Mannens G, Van Beijsterveldt L, Anciaux K, Verboven P, Meuldermans W, Lavrijsen K. Strategies for absorption screening in drug discovery and development. *Curr Top Med Chem* 2001;1:367–383.
68. Parrott N, Lave T. Prediction of intestinal absorption: comparative assessment of GASTROPLUS and IDEA. *Eur J Pharm Sci* 2002;17:51–61.
69. Grass GM. Simulation models to predict oral drug absorption from *in vitro* data. *Adv Drug Deliv Rev* 1997;23:199–219.
70. Norris DA, Leesman GD, Sinko PJ, Grass GM. Development of predictive pharmacokinetic simulation models for drug discovery. *J Control Release* 2000;65:55–62.
71. Tubic M, Wagner D, Spahn-Langguth H, Bolger MB, Langguth P. *In silico* modeling of non-linear drug absorption for the P-gp substrate talinolol and of consequences for the resulting pharmacodynamic effect. *Pharm Res* 2006;23:1712–1720.
72. Batrakova EV, Li S, Miller DW, Kabanov AV. Pluronic P85 increases permeability of a broad spectrum of drugs in polarized BBMEC and Caco-2 cell monolayers. *Pharm Res* 1999;16:1366–1372.
73. Bachmeier CJ, Miller DW. A fluorometric screening assay for drug efflux transporter activity in the blood–brain barrier. *Pharm Res* 2005;22:113–121.
74. Hashimoto T, Nomoto M, Komatsu K, Haga M, Hayashi M. Improvement of intestinal absorption of peptides: adsorption of B1-Phe monoglucosylated insulin to rat intestinal brush-border membrane vesicles. *Eur J Pharm Biopharm* 2000;50:197–204.
75. Shilling AD, Azam F, Kao J, Leung L. Use of canalicular membrane vesicles (CMVs) from rats, dogs, monkeys and humans to assess drug transport across the canalicular membrane. *J Pharmacol Toxicol Methods* 2006;53:186–197.
76. Zhang Y, Benet LZ. Characterization of P-glycoprotein mediated transport of K02, a novel vinylsulfone peptidomimetic cysteine protease inhibitor, across MDR1-MDCK and Caco-2 cell monolayers. *Pharm Res* 1998;15:1520–1524.
77. Bakos E, Evers R, Szakacs G, Tusnady GE, Welker E, Szabo K, de Haas M, van Deemter L, Borst P, Varadi A, Sarkadi B. Functional multidrug resistance protein (MRP1) lacking the N-terminal transmembrane domain. *J Biol Chem* 1998;273:32167–32175.
78. Evers R, Kool M, van Deemter L, Janssen H, Calafat J, Oomen LC, Paulusma CC, Oude Elferink RP, Baas F, Schinkel AH, Borst P. Drug export activity of the human canalicular multispecific organic anion transporter in polarized kidney MDCK cells expressing cMOAT (MRP2) cDNA. *J Clin Invest* 1998;101:1310–1319.
79. Fujino H, Saito T, Ogawa S, Kojima J. Transporter-mediated influx and efflux mechanisms of pitavastatin, a new inhibitor of HMG-CoA reductase. *J Pharm Pharmacol* 2005;57:1305–1311.
80. Bubik M, Ott M, Mahringer A, Fricker G. Rapid assessment of P-glycoprotein-drug interactions at the blood–brain barrier. *Anal Biochem* 2006;358:51–58.
81. Bachmeier CJ, Trickler WJ, Miller DW. Comparison of drug efflux transport kinetics in various blood–brain barrier models. *Drug Metab Dispos* 2006;34:998–1003.
82. Yamashita S, Hattori E, Shimada A, Endoh Y, Yamazaki Y, Kataoka M, Sakane T, Sezaki H. New methods to evaluate intestinal drug absorption mediated by oligopeptide transporter from *in vitro* study using Caco-2 cells. *Drug Metab Pharmacokinet* 2002;17: 408–415.

83. Troutman MD, Thakker DR. Novel experimental parameters to quantify the modulation of absorptive and secretory transport of compounds by P-glycoprotein in cell culture models of intestinal epithelium. *Pharm Res* 2003;20:1210–1224.
84. Wang Q, Rager JD, Weinstein K, Kardos PS, Dobson GL, Li J, Hidalgo IJ. Evaluation of the MDR-MDCK cell line as a permeability screen for the blood–brain barrier. *Int J Pharm* 2005;288:349–359.
85. Li F, Hong L, Mau CI, Chan R, Hendricks T, Dvorak C, Yee C, Harris J, Alfredson T. Transport of levovirin prodrugs in the human intestinal Caco-2 cell line. *J Pharm Sci* 2006;95:1318–1325.
86. Korfmacher WA, Cox KA, Ng KJ, Veals J, Hsieh Y, Wainhaus S, Broske L, Prelusky D, Nomeir A, White RE. Cassette-accelerated rapid rat screen: a systematic procedure for the dosing and liquid chromatography/atmospheric pressure ionization tandem mass spectrometric analysis of new chemical entities as part of new drug discovery. *Rapid Commun Mass Spectrom* 2001;15:335–340.
87. Raynaud FI, Fischer PM, Nutley BP, Goddard PM, Lane DP, Workman P. Cassette dosing pharmacokinetics of a library of 2,6,9-trisubstituted purine cyclin-dependent kinase 2 inhibitors prepared by parallel synthesis. *Mol Cancer Ther* 2004;3:353–362.
88. Smith NF, Hayes A, Nutley BP, Raynaud FI, Workman P. Evaluation of the cassette dosing approach for assessing the pharmacokinetics of geldanamycin analogues in mice. *Cancer Chemother Pharmacol* 2004;54:475–486.
89. Schinkel AH, Smit JJ, van Tellingen O, Beijnen JH, Wagenaar E, Van Deemter L, Mol CA, Van der Valk MA, Robanus-Maandag EC, te Riele HP, et al. Disruption of the mouse *mdr1a* P-glycoprotein gene leads to a deficiency in the blood–brain barrier and to increased sensitivity to drugs. *Cell* 1994;77:491–502.
90. Toub N, Malvy C, Fattal E, Couvreur P. Innovative nanotechnologies for the delivery of oligonucleotides and siRNA. *Biomed Pharmacother* 2006;60:607–620.
91. Schinkel AH, Mol CA, Wagenaar E, van Deemter L, Smit JJ, Borst P. Multidrug resistance and the role of P-glycoprotein knockout mice. *Eur J Cancer* 1995;31A:1295–1298.
92. Ocheltree SM, Shen H, Hu Y, Keep RF, Smith DE. Role and relevance of peptide transporter 2 (PEPT2) in the kidney and choroid plexus: *in vivo* studies with glycylsarcosine in wild-type and PEPT2 knockout mice. *J Pharmacol Exp Ther* 2005;315:240–247.
93. Kool M, de Haas M, Scheffer GL, Scheper RJ, van Eijk MJ, Juijn JA, Baas F, Borst P. Analysis of expression of cMOAT (MRP2), MRP3, MRP4, and MRP5, homologues of the multidrug resistance-associated protein gene (MRP1), in human cancer cell lines. *Cancer Res* 1997;57:3537–3547.
94. Kool M, van der Linden M, de Haas M, Baas F, Borst P. Expression of human MRP6, a homologue of the multidrug resistance protein gene MRP1, in tissues and cancer cells. *Cancer Res* 1999;59:175–182.
95. Kruh GD, Guo Y, Hopper-Borge E, Belinsky MG, Chen ZS. ABCC10, ABCC11, and ABCC12. *Pflugers Arch* 2007;453:675–684.
96. Anzai N, Kanai Y, Endou H. Organic anion transporter family: current knowledge. *J Pharmacol Sci* 2006;100:411–426.
97. Sun H, Johnson DR, Finch RA, Sartorelli AC, Miller DW, Elmquist WF. Transport of fluorescein in MDCKII-MRP1 transfected cells and mrp1-knockout mice. *Biochem Biophys Res Commun* 2001;284:863–869.
98. Sugiyama D, Kusuhara H, Lee YJ, Sugiyama Y. Involvement of multidrug resistance associated protein 1 (Mrp1) in the efflux transport of 17beta estradiol-D-17beta-glucuronide ($E_2$17betaG) across the blood–brain barrier. *Pharm Res* 2003;20:1394–1400.

99. Salama NN, Kelly EJ, Bui T, Ho RJ. The impact of pharmacologic and genetic knockout of P-glycoprotein on nelfinavir levels in the brain and other tissues in mice. *J Pharm Sci* 2005;94:1216–1225.
100. Cisternino S, Mercier C, Bourasset F, Roux F, Scherrmann JM. Expression, up-regulation, and transport activity of the multidrug-resistance protein Abcg2 at the mouse blood–brain barrier. *Cancer Res* 2004;64:3296–3301.
101. Aigner A. Delivery systems for the direct application of siRNAs to induce RNA interference (RNAi) in vivo. *J Biomed Biotechnol* 2006:1–15.

# ACCUMULATION OF DRUGS IN TISSUES

Krishnamurthy Venkatesan,[1] Deepa Bisht,[1] and Mohammad Owais[2]

[1]*National JALMA Institute for Leprosy and Other Mycobacterial Diseases, Agra, India*
[2]*Aligarh Muslim University, Aligarh, India*

**Contents**

1 Introduction
2 Movement of Drug Molecules Across Cell Barriers
3 Absorption of Drugs
  3.1 Oral, Sublingual, and Rectal Administration
  3.2 Application to Epithelial Surfaces
  3.3 Administration by Inhalation
  3.4 Administration by Injection
4 Drug Distribution
  4.1 Rate of Distribution
  4.2 Blood Perfusion
  4.3 Extent of Blood Distribution
  4.4 Physiological Barriers and Reservoirs Influencing Drug Distribution
5 Important Proteins that Contribute to Drug Binding
6 Volume of Distribution
7 Influence of Lipophilicity and Lysosomal Accumulation on Tissue Drug Distribution
8 Partition of the Drug into Fat and Other Tissues
9 Special Drug Delivery Systems
  Acknowledgments
  References

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

## 1 INTRODUCTION

In order to exert desirable pharmacological action, drugs need to achieve an adequate concentration in their target tissues. The term pharmacokinetics refers to absorption, distribution, metabolism, excretion, and toxicity. This chapter deals, in detail, with the aspects concerned with drug distribution and disposition.

The two fundamental processes that determine the concentration of a drug at any moment and in any region of the body are (1) translocation of the drug molecules and (2) the chemical transformations they undergo due to metabolic action. Drug molecules move around the body in two ways: bulk flow transfer (i.e., in the bloodstream) and diffusional transfer. The chemical nature of the drug makes no difference to its transfer by bulk flow as the cardiovascular system provides a very fast long-distance distribution system. In contrast, diffusional characteristics differ markedly between different drugs. In particular, ability to cross hydrophobic diffusion barriers is strongly influenced by lipid solubility of the drug molecule. Aqueous diffusion is also part of the overall mechanism of drug transport since it is this process that delivers drug molecules to and from the nonaqueous barriers.

Once the drug reaches the bloodstream, it is distributed to all organs including tissues that are not relevant to its pharmacological or therapeutic effect. Thus, after absorption, the drug may not only get reversibly associated with its site of action (receptor) but may get bound to plasma proteins or may accumulate in various storage sites or may enter into tissues that are not involved in its primary action (although they may be involved in metabolism, and excretion). This part of pharmacokinetics, which deals with distribution, metabolism, and excretion, is termed drug disposition because these three phases precisely decide the fate of a drug after absorption. The drug may be evenly, or unevenly, or selectively distributed in different body compartments, which are physiological in character rather than anatomical entities.

## 2 MOVEMENT OF DRUG MOLECULES ACROSS CELL BARRIERS

Plasma membrane limits the boundary of a cell from the rest of the body parts including aqueous compartments. In fact, a single layer of membrane separates intracellular and extracellular compartments. An epithelial barrier such as the gastrointestinal mucosa or renal tubule consists of a layer of cells tightly connected with each other so that molecules traverse at least two cell membranes—inner and outer. Vascular epithelium is still more complicated, with its variation in anatomical disposition and vascular flow between tissues. In some organs, especially the central nervous system and placenta, there are tight junctions between the cells and the endothelium is covered with an impermeable layer of periendothelial cells. In organs such as the liver and spleen, the endothelium is discontinuous, allowing free passage between cells.

There are four main ways by which small molecules cross cell membranes—(1) by diffusing directly through the lipids of the plasma membrane; (2) by diffusing through aqueous pores formed by special proteins that traverse the lipid bilayer; (3) by combining with a transmembrane carrier protein that binds a drug molecule on one side of the membrane, then changes conformation and releases it on the

other; or (4) by pinocytosis. Of these four ways, diffusion through lipid and carrier-mediated transport are very important in relation to pharmacokinetic mechanisms. There is yet another mechanism of transport, by which some drugs are actively transported into the cell by energy-dependent pumps [1]. In some cases, two or more drugs may share the same pump, giving rise to drug–drug interactions A different energy-dependent efflux pump may also pump them out of the cell. It is thought that multidrug resistance (mdr) transporter may play a role in lowering intracellular drug concentrations via an efflux mechanism. Increased expression of P-glycoprotein is often associated with lower intracellular accumulation of certain drugs due to efflux of the drug from the cells [2–5].

## 3 ABSORPTION OF DRUGS

Absorption is defined as the passage of a drug from the site of administration into the plasma. Absorption is therefore important for all routes of administration, except intravenous injection. The main routes of administration are oral, sublingual, rectal, topical application to other epithelial surfaces (e.g., skin, cornea, vagina, and nasal mucosa), inhalation, and injection—subcutaneous, intramuscular, intravenous, and intrathecal [6].

### 3.1 Oral, Sublingual, and Rectal Administration

***Oral Administration*** Although most drugs are swallowed by the oral route, little absorption occurs until the drug reaches the small intestine. The mechanism of drug absorption is the same as for other epithelial barriers, namely, the passive diffusion at a rate determined by the ionization and lipid solubility of the drug molecules. In a few instances intestinal absorption depends on carrier-mediated transport rather than simple lipid diffusion. Examples of the drugs under this category are levodopa and fluorouracil. In general, iron is absorbed via specific carriers in the surface membranes of the jejunal mucosa, and calcium is absorbed by means of vitamin D-dependent carrier systems. The biochemical properties of the drug such as molecular weight, particle size, and pH, along with local factors such as food, intestinal motility, and blood flow, all affect gastrointestinal (GI) absorption. The presence of agents that may affect gastric pH or motility or interact with the drug to produce binding will prevent or slow the absorption. Examples of these agents and the absorption factors affected are antacids (pH), vasodilators (GI perfusion), cathartics (GI motility), and calcium (binding). Some drugs are given to patients as prodrugs since the absorption of the prodrug may be superior to the parent drug (hence increased bioavailability). Examples of prodrugs are levodopa and azarabine, prodrugs of dopamine and azauridine, respectively. The term *bioavailability* indicates the proportion of the drug that passes into the systemic circulation after oral administration, taking into consideration both absorption and metabolic degradation in the gut wall or liver before reaching the systemic circulation.

***Sublingual Administration*** A few drugs are placed under the tongue (taken sublingually) for their direct absorption into the small blood vessels that lie underneath the tongue. The absorption is rapid and the drug enters the systemic circulation

without entering the portal system and so escapes first-pass metabolism. The sublingual route is especially good for nitroglycerin, which is used for treatment of angina.

***Rectal Administration*** Rectal administration is useful for drugs that are required both to produce a local effect (e.g., anti-inflammatory drugs for treating ulcerative colitis) and to produce systemic effects. The spatial positioning of the rectal suppositories avoiding the upper hemorrhoidal vein can avoid the *first-pass effect*, as the encapsulated drug molecules will be directly released into the inferior vena cava. Nevertheless, the popularity of this mode of route has abated because of erratic absorption behavior and also a great deal of advancement in parenteral preparations.

### 3.2 Application to Epithelial Surfaces

***Cutaneous Administration*** Cutaneous administration is used when a local effect on the skin is required (e.g., topically applied steroids). Appreciable absorption may nonetheless occur and lead to systemic effects. Most drugs are absorbed very poorly through unbroken skin. Transdermal dosage forms in which the drug is incorporated as a stick-on patch applied in the area of thin skin are increasingly used. Several drugs are available in this form, for example, estrogen for hormone replacement.

***Nasal Sprays*** Some drugs like peptide hormone analogues (e.g., antidiuretic hormone, gonadotropin-releasing hormone, and calcitonin), if given orally, will be inactive as they will quickly be destroyed in the gastrointestinal tract, and so they are administered in nasal sprays so that adequate absorption takes place through nasal mucosa. Intranasal administration has advantages in several cases. For example, clinical experience suggests that intranasal sumatriptan (a potent serotonin $5HT_{1B/1D}$ agonist widely used in the treatment of migraine) has more advantages over oral and subcutaneous administration, such as more rapid onset of the effect and use in patients with gastrointestinal complaints [7].

***Eye Drops*** Many drugs are applied as eye drops and they get absorbed through the epithelium of the conjunctival sac. Systemic side effects are not caused while desirable local effects are best achieved. One example is dorzolamide (a carbonic anhydrase inhibitor), which is given as eye drops to lower ocular pressure in glaucoma patients. In some cases administration of eye drops may result in unwanted side effects on account of delivery of absorbed material from the eye to the systemic circulation.

### 3.3 Administration by Inhalation

This route is best used for volatile and gaseous anesthetics and the lung serves as the route of administration and elimination. Rapid absorption on account of large surface area and high vascularity of the lung tissue makes it possible to achieve rapid adjustments of plasma concentration of the drugs. Drugs used for their effects on the lung are also given by inhalation, usually as an aerosol. Inhalation therapy

is commonly used in the treatment of asthma as it results in a high concentration of active drug at the site of action. The degree of lung deposition of an inhaled drug is influenced by several factors—the one most important is the inhalation maneuvers. Other possible factors are related to the properties of the formulation, such as particle size and hygroscopic nature. Glucocorticoids (e.g., beclomethasone) and bronchodilators (e.g., salbutamol) are given by this route, which results in high local concentration in the lung while minimizing systemic side effects. Minimizing the systemic absorption by chemical modification of a drug will reduce the side effects. For example, ipatropium, a muscarinic receptor antagonist, a quaternary ammonium ion analogue of atropine, is poorly absorbed into systemic circulation.

### 3.4 Administration by Injection

***Intravenous Injection*** This is the fastest and the most certain route of drug administration. Bolus injection produces a very high concentration of drug first in the right heart and lungs, and then in the systemic circulation. The rate of injection of the drug decides the peak concentration reaching the tissues. Steady intravenous infusion avoids the uncertainties of absorption from other sites while avoiding high peak plasma concentrations resulting from bolus injection. Several antibiotics and anesthetics such as propofol and diazepam, used for treatment of patients with status epilepticus, are administered intravenously for better bioavailability of the drug.

***Subcutaneous or Intramuscular Injection*** Subcutaneous or intramuscular injection of drugs usually produces a faster effect than oral administration, but the rate of absorption depends on the site of injection and on local blood flow. Diffusion through the tissue and removal of the injected drug by local blood flow are the rate-limiting factors.

***Intrathecal (Intraspinal) Injection*** Injection of a drug into the subarachnoid space via a lumbar puncture needle is often used for some specialized purposes. Methotrexate is administered intrathecally in the treatment of certain childhood leukemias. Regional anesthesia can be produced by intrathecal administration of a local anesthetic, such as bupivacaine. Some antibiotics (e.g., aminoglycosides) cross the blood–brain barrier very slowly and in rare cases of nervous system infections they are given intrathecally.

## 4 DRUG DISTRIBUTION

Following absorption from the gastrointestinal tract or direct administration through another route, the drug is distributed to the tissues and organs throughout the body. Most drugs do not localize to one particular site, but are distributed to a number of organs and tissues. Distribution refers to both rate of distribution and extent of distribution. The rate of distribution depends on the membrane permeability and the blood perfusion, while the extent of distribution is influenced by factors such as lipid solubility of the drug, pH–p$K_a$, plasma protein binding, and intracellular binding [8, 9].

### 4.1 Rate of Distribution

***Membrane Permeability*** Membrane permeability refers to the ability of a drug to cross the cell membranes. This may occur by passive diffusion, where the drug simply moves from an area of higher concentration to one of lower concentration. Non-polar drug molecules dissolve freely in the lipid bilayer of the membrane and are translocated across it by diffusion. The permeability coefficient, $P$, and the concentration gradient across the membrane will determine the number of drug molecules that will cross per unit area in unit time. In general, two physicochemical factors that significantly contribute to the permeability coefficient are solubility of the drug and its diffusivity across the membrane. Solubility is expressed as a partition coefficient for the substance to be distributed between the membrane phase (lipid environment) and the aqueous environment and the diffusivity is a measure of the mobility of the molecules within membrane lipid and is expressed as a diffusion coefficient. Since the diffusion coefficient varies only slightly among different drug molecules, the partition coefficient remains as the most important variable influencing drug diffusion. Lipid solubility is one of the important determinants of the pharmacokinetic characteristics of a drug. Many properties, such as rate of absorption from the gut, penetration into brain and the other tissues, and the extent of renal elimination, can be predicted from the drug's solubility profile.

Lipid-soluble compounds pass trough the cell membranes very rapidly, whereas water-soluble compounds penetrate more slowly at a rate dependent on their size. Low molecular weight drugs pass through by simple diffusion. The transfer is slow for drugs with molecular diameter above 100 Å. There are two deviations to the typical capillary structure, which result in variation from normal drug permeability. First, the permeability is greatly increased in the renal capillaries by pores in the membrane of the endothelial cells and also in specialized hepatic capillaries (sinusoids), where a complete lining is generally lacking. This results in more extensive transfer of drugs out of the capillary bed. Second, the brain capillaries seem to have impermeable walls restricting the transfer of molecules from blood to brain tissue (blood–brain barrier). Lipid-soluble compounds can readily be transferred but the transfer of polar compounds is severely restricted by the barrier.

The transfer rate of ionizable drugs across the capillary membrane depends on the $pK_a$ of the drug molecule as well as the pH of the blood. Many drugs are weak acids or bases and therefore exist in the un-ionized as well as the ionized form. The ratio of un-ionized to ionized or vice versa will depend on the pH of the environment. The ionized species has very low lipid solubility and is virtually unable to permeate membrane except where a specific transport mechanism exists. The lipid solubility of the un-ionized species will depend on the chemical nature of the drug. Most of the un-ionized drug molecules are adequately lipid soluble. There are, however, exceptions to this rule while dealing with many drug molecules. For example, aminoglycoside antibiotics are sufficiently hydrophilic in the uncharged form, here the presence of hydroxyl groups (electronegative oxygen) in the sugar molecules of the drug that plays an important role in imparting solubility to polar solvents like water.

Ionization affects not only the rate of drug permeation across membranes but also the steady-state distribution of the drug molecules between two or more aqueous compartments like plasma, urine, and gastric juice, if a pH difference exists between

them. Within each compartment the ratio of ionized to un-ionized species of the drug, as determined by the p$K_a$ of the drug and the pH of the environment, will be different and acidic drug will be concentrated in the compartment with high pH and vice versa. This is called ion trapping. The larger the pH difference between two aqueous compartments, the greater will be the concentration gradient produced by ion trapping.

### 4.2 Blood Perfusion

Membrane permeability tends to restrict the transfer and distribution of drugs once they are delivered to the tissue. The other major factor that determines the rate of drug distribution is blood perfusion. The rate at which blood perfuses to different organs varies widely. Total blood flow is greatest to brain, kidneys, liver, and muscle with highest perfusion rates to brain, kidney, liver, and heart. Hence total drug concentration would rise most rapidly in these organs. Highly perfused organs, namely, heart, liver, kidney, and brain, usually receive drug within minutes after its absorption. Distribution to tissues like muscle, fat, and skin may require longer periods of time (minutes to hours). Certain organs such as the adrenals also have large perfusion rates. The relative restriction/contribution of the twin factors of membrane permeability and blood perfusion to drug distribution can better be understood with examples of transfer of thiopental and penicillin to brain and muscle. Thiopental gets into the brain faster than muscle, while penicillin gets into muscle more quickly than it gets into brain. Thiopental is only partly ionized and passes into the brain or muscle easily. But since brain has a perfusion rate approximately 20 times higher than muscle, thiopental transfers in and out of brain more quickly. On the other hand, penicillin, which is quite polar, transfers faster in muscle than brain on account of lesser restriction of muscle capillaries to the drug.

There is growing evidence that the reduction of blood flow in soft tissues is closely associated with the significant impairment of the rate of tissue penetration by antimicrobial agents. For example, in patients with peripheral arterial occlusive disease, it has been demonstrated that the peak concentrations are significantly lower in the interstitium of ischemic soft tissue in comparison with that in healthy and well perfused tissue.

### 4.3 Extent of Drug Distribution

Several physiological barriers influence the extent of drug distribution. Chemicophysical properties, molecular weight, degree of plasma protein binding, and intracellular binding are among the most important intrinsic characteristics that condition the extent of drug distribution in the body [10, 11]. Besides its ability to bind with plasma protein, the size and degree of ionization as well as lipid solubility of a drug molecule specifically regulate its extravascular transfer. Host-related factors such as site of infection, presence of biological barriers (blood–brain barrier, placental barrier, blood–CSF barrier, prostate, ocular membranes), and local pH may contribute to the pharmacological action of drug molecules. Sometimes these factors can inhibit a particular drug molecule as well. Inactivation of the antibacterial agent in the tissue can occur at infection sites by binding to interstitial fluid proteins, soluble intracellular proteins, or subcellular structures (nucleic acids, leukocytic chromatin cell membranes, and mucopolysaccharides).

### 4.4 Physiological Barriers and Reservoirs Influencing Drug Distribution

***Blood–Brain Barrier*** A major problem in drug delivery to brain is the presence of the blood–brain barrier (BBB), which is now well established as a unique membranous barrier that tightly segregates the brain from the blood influx. Endothelial cells of brain capillaries differ from most other capillaries of the body, as they are very tightly joined and lack intracellular pores. In addition, these brain capillaries are also enveloped by less permeable glial cells. Morphologically, this constitutes the *blood–brain barrier* [12, 13]. Anatomically, there exists a dual barrier in the CNS: the blood–brain barrier and blood–CSF barrier (located in the choroid plexus). The BBB places certain constraints on the passage of drugs from the blood to the brain and the CSF. As a rule only lipid-soluble nonionized forms of drugs penetrate more easily into brain as compared to water-soluble ionized forms of drugs [14, 15]. The volatile anesthetics like ether and chloroform, ultrashort acting barbiturates like thiopental, narcotic analgesics like morphine and heroin, dopamine precursor like L-dopa, sympathomimetics like amphetamine and ephedrine, and drugs like diazepam and propranolol can pass through the BBB, while polar compounds like dopamine, serotonin, and streptomycin and quaternary substances like *d*-tubocurarine, hexamethonium, neostigmine, and acetylcholine do not penetrate the BBB. The following five regions of the brain, on account of their relative permeability, do not form part of the BBB—pituitary gland, pineal body, area prostrema near the floor of the fourth ventricle, median eminence, and choroid plexus capillaries. The BBB is further reinforced by a high concentration of P-glycoprotein (P-gp) active drug efflux transporter protein in the luminal membranes of the cerebral capillary endothelium. This efflux transporter actively removes a broad range of drug molecules from the endothelial cell cytoplasm before they cross into the brain parenchyma [13].

***Blood–CSF Barrier*** The second barrier that a systemically administered drug encounters before entering the central nervous system (CNS) is known as the blood–CSF barrier (BCB). Since CSF can exchange molecules with the interstitial fluid of the brain parenchyma, the passage of drug molecules from the blood into the CSF is also carefully regulated by the BCB. Physiologically, the BCB is found in the epithelium of choroid plexus, which are arranged in a manner that limits the passage of molecules and cells into the CSF. The choroid plexus and the arachnoid membrane act together at the barriers between the blood and the CSF. [13]. Only nonionizable lipid-soluble drugs can pass from blood to the CSF as in the case of the BBB where the epithelial cells of choroid plexuses are lined by occluding zonulae. But the brain–CSF barrier is composed of epithelial cells lining the ventricles and occluding zonulae do not connect these. Hence the CSF–brain barrier is extremely permeable to drug molecules from the CSF to brain cells [16]. The significance of this fact is that drugs like penicillin that have poor penetration through the BBB can cross the brain–CSF barrier if given by the intrathecal route and reach the brain tissue in sufficient concentration. In the clinical setting, drug concentrations in CSF are sometimes used as a surrogate for drug concentrations at the target site within the brain. But multiplicity of compartments of the brain and involvement of several factors in the transport of dugs from plasma into the brain and distribution within the brain limit the applicative value of CSF drug levels in prediction of brain target concentrations [17].

***Placental Barrier*** The placental membrane readily allows the transfer of nonpolar lipid-soluble drugs mainly by passive diffusion. However, the other transport mechanisms like active transport and pinocytosis are also operative. Thus, lipid-soluble nonionizable drugs like hypnotics, narcotics, general anesthetics, cardiac glycosides, alcohol, neuroleptics, and certain antibiotics can readily cross the placental barrier, while nonionizable and very high molecular weight substances cannot cross the barrier. Antihypertensive drugs that are lipid soluble bind with proteins to lesser extent, and when in un-ionized form generally pass through the placental barrier with ease [18].

***Special Compartments for Drug Distribution*** Besides plasma proteins, the tissues of the body can also act as reservoirs for drugs, but usually the sites where drugs accumulate are not those where they exert their physiological action. These stored drugs are in equilibrium with the drug in plasma and are only slowly released from these storage sites to provide enough circulating drug to produce biological effects. Consequently, this type of storage, unlike plasma protein binding, represents "a site of loss" rather than "a depot" for continued drug action.

*Cellular Reservoir* A drug may have a great affinity for plasma proteins, yet primarily be distributed in tissue if the tissue has an even higher affinity for the drug. This affinity could be due to several reasons, for example, binding to tissue proteins (albumin) or top nucleoproteins. Examples are digoxin and emetine to skeletal muscle, heart, liver, kidney (bound to muscle proteins), and retina (nucleoproteins); cadmium, lead, and mercury in kidney (muscle protein, metallothionein); and chlorpromazine in eye (affinity to retinal pigment melanin).

*Fat Reservoir* Highly lipid-soluble drugs like thiopentone, DDT, and phenoxybenzamine get selectively accumulated in fat and adipose tissue. The fat, however, is a sluggish reservoir due to lesser blood flow; but if body fat starts depleting, as occurs during starvation, the stored drug may be mobilized and toxicity may occur.

*Transcellular Reservoir* Aqueous humor (e.g., chloramphenicol and prednisolone), CSF (e.g., amino sugars and sucrose), endolymph, joint fluid (e.g., ampicillin), and pleural (e.g., imipramine and methadone), pericardial, and peritoneal sacs can also serve as drug reservoirs.

*Bone and Connective Tissue Reservoirs* Many drugs and substances like tetracycline, cisplatin, lead, arsenic, and fluorides form a complex with bone salts and get deposited in nails, bones, and teeth. The antifungal drug griseofulvin has an affinity for keratin precursor cells and is selectively accumulated in the skin and fingernails. Bone may become a reservoir for the slow release of toxic substances like lead and arsenic and for the anticancer drug cisplatin into the blood.

*Biliary Reservoir* The lipophilic drugs with relatively high molecular weight (~325 Da in rats and greater than 500 Da in humans) are generally metabolized in liver and eventually accumulate in bile. Subsequent biliary excretion of such drugs can significantly impact their systemic exposure, which in turn regulates their pharmacological effect on the one hand and their toxicity on the other. Drugs (such as

mycophenolic acid and digoxin) excreted into bile often undergo some degree of reabsorption along the gastrointestinal tract [19, 20]. Enterohepatic recycling is also a very important physiological process for bile salt homeostasis. In humans, more than 95% of the bile salt is reabsorbed in the distal ileum [21]. The biliary clearance of a drug would be extremely valuable in elucidating the potential mechanism operative in hepatobiliary toxicity, predicting drug–drug interaction, and identifying enterohepatic versus enteroenteric recirculation. Several clinical studies have shown that drug–drug interaction can result in changes in drug absorption due to inhibition of transport proteins. For example, inhibition of P-glycoprotein results in increased absorption of talinolol [22]. The liver plays a central role in the biotransformation and removal of xenobiotics and endogenous compounds from the blood; while the majority of small, lipophilic compounds flow through the space of Disse and enter the hepatocyte via the basolateral membrane by simple diffusion. On the other hand, more polar and bulky molecules require transport systems to cross the sinusoidal membrane. After reaching the hepatocyte, drugs can be transported into bile either as an uncharged species or as more hydrophilic metabolites produced after phase I and phase II biotransformation, or they can be excreted into the bloodstream by basolateral transport proteins. In addition to bile salt transport, hepatic transport proteins facilitate excretion of end products of metabolism, such as bilirubin or xenobiotics, into bile for subsequent removal via the intestine.

*Plasma Protein Binding as Reservoir* Although some drugs are simply dissolved in plasma water, most of them get bound to plasma components, such as albumin, globulin, transferrin, and ceruloplasmin (and also through the BBB), get metabolized, and are excreted in urine, saliva, CSF, and milk. As the free or unbound drug is eliminated from the body, more drug dissociates from the drug–plasma protein complex, glycoproteins, and α- and β-lipoproteins to replenish the free drug levels in plasma. Drugs usually are bound to plasma and cellular proteins in a reversible manner and in dynamic equilibrium, according to the law of mass action [23]. Thus, extensive plasma protein binding may serve as a circulating drug reservoir. But it is only the free fraction of the drug that is pharmacologically active, while the protein bound component is inert. Similarly, it is the free fraction of the drug that diffuses through the capillary wall (and also through the BBB), gets metabolized, and is excreted in urine, saliva, CSF, and milk.

## 5 IMPORTANT PROTEINS THAT CONTRIBUTE TO DRUG BINDING

1. *Albumin.* Of the plasma proteins, albumin is the major drug binding and transport protein of the circulatory system. Drug binding to plasma albumin occurs to a variable extent. It consists of three homologous domains. Each domain consists of two subdomains, A and B. The principal sites of ligand binding to HSA are located in hydrophobic cavities in subdomains IIA and IIIB. HSA is known to undergo different pH-dependent conformational transitions, the N-F transition between pH 5.0 and 3.5, the F-E transition between pH 3.5 and 1.2, and the N-B transition

between pH 7.0 and 9.0. The N-F isomerization involves unfolding and separation of domain III from the rest of the molecule without significantly affecting the rest of the molecule. The I state is characterized by unfolding of domain III together with partial but significant loss of native conformation of domain I. The domain II of HSA remains unaffected in the intermediate state. That the N-B transition has physiological significance is suggested by the fact that under increased $Ca^{2+}$ concentration in blood plasma, the B isomer predominates. Moreover, it is believed that the transport function of albumin is controlled through this transition or akin to it.

Most of the acidic drugs (such as warfarin, penicillin, sulfonamides, phenylbutazone, and salicylic acid), in principle, bind to albumin. Two binding sites for acidic drugs have been identified on human serum albumin. Although less specific, domain I can bind with a variety of structurally unrelated drugs like acenocoumarin, warfarin, dicoumarol, phenylbutazone, phenytoin, valproic acid, sulfonamides, and endogenous substances like bilirubin. On the other hand, site II is more specific. With the exception of benzodiazepines, acidic drugs such as ibuprofen, naproxen, and analogues of probenecid, clofibric acid, cloxacillin, and endogenous substances like tryptophan bind to this site. Some drugs bind to both the sites with different affinities, and examples are salicylic acid, tolubutamide, and indomethacin [24–26].

2. $\alpha_2$-*Acid Glycoprotein.* While acidic drugs primarily bind to albumin, many lipophilic basic drugs, such as quinidine, imipramine, lidocaine, chlorpromazine, propranolol and its analogues, and spinperidol bind avidly to $\alpha_2$-acid glycoprotein (AAG). The drug binding to AAG in health and disease has been extensively reviewed by Kremer et al. [27].

3. *Lipoproteins.* The role of lipoproteins as a transport system for drugs has been largely overlooked compared to the extensive reports on drug binding with albumin and AAG. However, there are reports that lipoproteins could be implicated in the binding process; for example, propofol binds extensively to all lipoprotein fractions [28, 29].

4. *Tissue Proteins and Nucleoproteins.* Some drugs, particularly those having high apparent volume of distribution, specifically bind to tissue proteins (e.g., digoxin, emetine, and chloroquine). The steroidal drugs can bind to specific receptor and nuclear proteins present on nucleus.

5. *Miscellaneous Binding Proteins.* Only a few drugs are significantly bound to other special proteins; for example, corticosteroids are bound to steroid binding globulin, transcortin; thyroxine to α-globulin; and antigens to γ-globulin.

## 6 VOLUME OF DISTRIBUTION

The distribution of drugs in the body depends on their lipophilicity and protein binding. Low plasma binding or high tissue binding or high lipophilicity means an extensive tissue distribution. In pharmacokinetics, the distribution is described by the parameter $V_d$, the apparent volume of distribution [10]. The apparent volume of distribution, $V_d$, is defined as the volume of fluid required to contain the total

amount, $Q$, of drug in the body at the same concentration as that present in the plasma, $C_p$.

$$V_d = Q/C_p$$

Values of $V_d$ for many drugs have been measured. The drugs with high plasma protein binding remain largely restricted to the vascular compartment and tend to have lower volumes of distribution. Some drugs are confined to the plasma compartment. The plasma volume is about 0.05 L/kg body weight. A few drugs such as heparin ($V_d$ of 0.05–0.1 L/kg) are confined to plasma volume because the molecule is so large that it cannot cross the capillary wall easily. More often, the retention of a drug in the plasma following a single dose reflects the strong binding to plasma protein. However, the free drug in the interstitial fluid only exerts a pharmacological effect.

Following multiple dosing, equilibration occurs with an ensuing increase in $V_d$ values. Drugs such as warfarin, sulfamethoxazole, glibenclamide, and atenolol are confined to the plasma compartment. Few drugs are distributed in the extracellular compartment, which has a volume of about 0.2 L/kg and many polar compounds such as vecuronium, gentamicin, and carbenicillin show an approximate $V_d$ of 0.2 L/kg. These drugs cannot easily enter cells because of their low lipid solubility and they do not traverse the blood–brain or placental barriers easily. Drugs like tubocurarine ($V_d$ 0.2–0.4 L/kg) and theophylline ($V_d$ 0.4–0.7 L/kg) are also distributed in the extracellular fluid. Some other drugs are distributed throughout the body water. Total body water represents about 0.55 L/kg, which approximates the distribution volume of relatively lipid-soluble drugs that cross cell membranes, such as phenytoin and ethanol. There are many drugs with the volume of distribution greater than the total body volume on account of binding of drug outside the plasma compartment, or partitioning into body fat. Examples are morphine, tricyclic antidepressants, and haloperidol.

## 7 INFLUENCE OF LIPOPHILICITY AND LYSOSOMAL ACCUMULATION ON TISSUE DRUG DISTRIBUTION

Distribution of a drug in the body is dependent on the permeation properties, the blood flow rates in various tissues, and plasma and tissue uptake. The distribution of drugs *in vivo* is largely determined by uptake competition among individual tissues [30, 31]. The lipophilic drugs are characterized by extensive accumulation in tissues, which leads to a high volume of distribution. Nonspecific binding to cellular membrane (phospholipids) and uptake by acidic compartments (mainly lysosomes) are responsible for such a distribution pattern [32]. Lysosomal trapping is an important mechanism of distribution of basic psychotropic drugs [33]. Lysosomal trapping has been reported to contribute to the uptake of despiramine and chloroquine in various tissues [34]. Basic drugs are distributed widely in various tissues in the following order: lung, fat, heart, kidney, brain, gut, muscle, and bone. The fat volume in the body influences the deposition kinetics. There is a good correlation in various tissues between the tissue–plasma concentration ratio and the octanol–water coefficient among various drugs.

## 8 PARTITION OF THE DRUG INTO FAT AND OTHER TISSUES

Drug accumulation or concentration in a tissue or biological fluid occurs when the influx of a drug is greater than the efflux. However, concentration gradients do not govern the rate of active transport. Thus, active transport can play a part in accumulation. Fat represents a large nonpolar compartment. In practice, this is important only for a few drugs mainly because the effective fat–water partition coefficient is relatively low for most drugs. For example, sequestration of morphine by body fat is of little importance as the drug's lipid–water partition coefficient is only 0.4, although the drug is sufficiently lipophilic to cross the blood–brain barrier. On the other hand, a drug such as thiopental, with a fat–water partition coefficient of about 10, accumulates substantially in body fat. The second factor that limits the sequestration of a drug in body fat is the amount of blood supply. Since fat receives less than 2% of cardiac output, drugs are delivered to body fat rather slowly. But some drugs like clofazimine (a riminophenazine with anti-inflammatory properties), used in the treatment of leprosy and other mycobacterial infections, extensively accumulates in fatty tissues and organs with in the reticuloendothelial system and the drug is retained for a long time even after stopping administration of the drug [35]. Sequestration also involves either drug ionization or drug binding to proteins and other macromoleules, preventing efflux by diffusion or active transport.

## 9 SPECIAL DRUG DELIVERY SYSTEMS

Several approaches are being explored to improve drug delivery. Some of these are polymeric carriers [36], biologically erodable microspheres [37], nanoparticles [38, 39], prodrugs [40, 41], antibody–drug conjugates, liposomes, and coated implantable devices [42, 43].

Many drugs can be formulated using simple vehicles; however, a large number of chemical entities pose problems because of the dose and poor solubility in simple solvents and cosolvent systems. As a result, there is a need for advanced formulations that will permit the administration of challenging drugs efficiently. For that purpose, drugs can be attached to the polymeric carriers or dispersed in suitable (colloidal) vehicles. In the past, surfactant systems that solubilized the drug and thereby provided a transparent product for injection were in use; for example, cremophor was a material of choice—unfortunately, it has been associated with anaphylactic reactions. Workers in the field hence sought alternative lipid carriers such as liposomes and lipid emulsions. Although fat emulsions have been in trend for decades for parenteral nutrition and these systems have been used with success as drug carriers, yet they are avoided, as such preparations allowed the drugs to rapidly exit the system and reach the bloodstream, hence limiting the possibilities of drug targeting [44]. Successful homing of drugs to the desired biological compartment of the host depends on the intrinsic properties of the drug, or it can also be achieved by manipulation of the carrier/delivery system, as little can be done to influence the target and its surroundings. Liposomes have been developed as carriers for the delivery of therapeutic and diagnostic agents against various infectious diseases. Incidentally, the particulate nature of the vesicles may facilitate passive homing of the entrapped drug molecules to the macrophages, which harbor many of the important pathogens in

their intracellular compartments such as *Mycobacterium, Leishmania*, and *Candida* species, belonging to three different major classes of microbes, namely, bacteria, protozoans, and fungi, respectively. Moreover, macrophages upon interaction with particulate drug delivery vehicles may act as the secondary depot for them, thereby helping in targeted delivery of entrapped drug molecules.

Particulate novel drug delivery vehicles such as liposomes, microspheres, and niosomes deliver entrapped material to macrophages and dendritic cells mainly. Liposomes have been documented as microscopic vesicles composed of phospholipid bilayers surrounding aqueous compartments. Liposomes have been used extensively as drug carriers. Potential applications of this delivery system include cancer chemotherapy, enzyme therapy, immunomodulation, antimicrobial therapy, metal detoxification, diagnostics, and topical therapy [45–49].

Injecting drugs encapsulated in liposomes has an advantage: since they are colloidal in nature, they are recognized as foreign particles and can readily be taken up by phagocytic cells of the immune system (cf. macrophages); thus, the drugs rapidly accumulate in the mononuclear phagocyte system (MPS). The inherent tendency of the liposomes to concentrate in MPS can be exploited in enhancing the nonspecific host defense against various intracellular pathogens by entrapping chemotactic or immunomodulatory chemical entities in them besides using them as carriers of antibiotics against various intracellular infections. Liposomal systems have been optimistically considered as "magic bullets" for more than three decades. Drug delivery with liposomes as the carrier system provides options and opportunities for designing biostable and/or site-specific drug therapy. Engineered or tailored versions of liposomes that offer the potentials of exquisite levels of specificity for drug targets are receiving much attention these days. Depending on the site of targeting, liposomes may be modified by either grafting chemotactic ligands such as peptides, polysaccharides, or affinity ligands like antibodies to the liposomal surface directly or may be coated via poly(ethylene glycol) anchors [50–52].

The delivery of diagnostic, therapeutic, and immunomodulatory agents to lymph nodes is mandatory for treatment of diseases with lymphatic involvement [53–55]. Saccharide modified liposomes showed enhanced absorption from the injection site and enhanced lymph node uptake compared to control liposomes [56]. Liposomes coated with nonspecific human antibodies injected subcutaneously showed a modestly increased lymphatic absorption and lymph node uptake compared to liposomes without the coating [57]. For example, administration of anti-HLA-DR bearing immunoliposomes resulted in a threefold higher accumulation in regional lymph nodes [58]. Poly(ethylene glycol) (PEG) coated liposomes were found to avoid uptake by the lymphatic system as compared to the plain liposomes [59, 60]. A novel method to enhance lymph node uptake of biotin-coated liposomes has been explored to facilitate the uptake of the drug-bearing liposomes [61].

The amino-phospholipids are generally confined to the inner leaflet of the plasma membrane. Translocation of phosphatidylserine (PS) to the outer leaflet serves as a signal for triggering recognition by macrophages [62]. Studies on the interaction between liposomes and macrophages have shown that larger particles may be phagocytosed more efficiently by macrophages than smaller liposomes [63]. Similarly, grafting of tetrapeptide tuftsin on the liposome surface would therefore enable us not only to "home" the liposomized drug to macrophages but also to stimulate these cells nonspecifically against infections [64–69].

Analysis of the intracellular trafficking patterns of the liposomal antigens reveals that after being phagocytosed by macrophages, liposomal antigen readily escapes from the endosomes into the cytoplasm of the macrophages [70]. Moreover, liposomes made up of lipids with fusogenic properties have been shown to deliver their entrapped molecules in the cytosol of the target cells more efficiently than conventional forms of liposomes. The membrane lipid composition of microorganisms exhibits a large amount of anionic phospholipids that play a pivotal role in membrane–membrane fusion [71]. Liposomes made up of lipids from various microorganisms have been used for targeting to macrophages as well as dendritic cells [72–75]. Recently, we have demonstrated that antigen entrapped in liposomes made up of lipids from *Escherichia coli* and edible yeast generate strong humoral as well as cell-mediated responses far better then conventional egg PC liposomes [76–78]. Moreover, sterically stabilized liposomes have been found to possess increased stability and to play a major role in $CD8^+$ T-cell response by targeting mature as well as immature dendritic cells. In a more recent study, PorA, a major antigen of *Neisseria meningitidis*, was purified and reconstituted in different types of targeted liposomes, by using mannose or phosphatidylserine as targeting moieties, or with positively charged liposomes for targeting dendritic cells (DCs). The use of targeted PorA liposomes resulted in an improved uptake by and activation of DCs and an increased localization in draining lymph nodes [79].

The interaction of liposomes with phagocytic cells is followed by either stable adsorption to the cell surface or cellular uptake of intact vesicles by an energy-dependent mechanism leading to lysosomal degradation of the liposomes and their contents. Liposome adsorption to the cell surface seems to be the rate-limiting step, since it can be assumed that stably adsorbed vesicles are more susceptible to subsequent uptake than vesicles that are loosely attached to the cell surface. Several prerequisites have to be successfully met that eventually enable liposomes to deliver biologically active agents to macrophages. Essentially, liposomes must readily bind to and be phagocytosed by free and fixed phagocytes; they must prevent degradation of entrapped drug; they must retain the encapsulated agent for delivery to the intracellular compartment of RES cells; and they must localize to macrophages in organs where metastasis or macrophage-associated disorders occur. Moreover, it has been shown that liposome with specific lipid composition or those harboring fusogenic peptide on their surface can specifically fuse with the target cells.

In general, liposomes are the most widely studied carrier in drug targeting to macrophages. However, the extent of liposome binding and subsequent ingestion by macrophages depends on a number of features of the lipid vesicles. These include composition, type, size, and surface properties of liposomes. For example, PS containing negatively charged liposomes can associate more effectively and deliver the liposomal content more efficiently to the macrophages than PS containing neutral liposomes [80–83]. Similarly, smaller liposomes deliver drugs more effectively than larger ones; presumably they are internalized more efficiently [84]. Some studies implied that positively charged and large sized liposomes could, however, improve liposome uptake in comparison with their counterparts [85, 86].

It was also observed that liposome uptake increased linearly with the incubation time and concentration. Higher uptake was observed with smaller and negatively charged liposomes. Inclusion of increasing amounts of cholesterol and sphingomyelin resulted in a decrement in the uptake by macrophages. The same study also

revealed that the uptake of liposomes differed considerably in different mouse strains [80–83].

The fatty acyl chains of the phospholipids used in the preparation of liposomes can regulate their *in vivo* stability. The efficacy of a liposomal formulation of a drug can be affected by the composition of liposomes [87]. Encapsulation of ampicillin in vesicles of solid liposomes—prepared from distearoylphosphatidylcholine (DSPC) and dipalmitoylphosphatidylglycerol (DPPG)—as compared with fluid type liposomes (composed of cholesterol, PC, and PS) resulted in delay of intracellular killing of *Listeria monocytogenes.* Besides charge, size and lamellarity of the liposomes are important characteristics that regulate their uptake by macrophages. Systematic evaluation of multilamellar vesicles (MLVs) with different phospholipid composition reveals that certain classes of phospholipids are recognized preferentially by macrophages. Inclusion of negatively charged phospholipids, such as phosphatidylserine (PS) or phosphatidylglycerol, in MLVs consisting of phosphatidylcholine (PC) greatly enhances their binding to and phagocytosis by macrophages. In contrast, neutral MLVs composed exclusively of PC are not efficiently taken up by macrophages. Such enhancement of phagocytosis has been shown to occur in mouse peritoneal macrophages, mouse Kupffer cells, rodent alveolar macrophages, human peripheral blood monocytes, and human alveolar macrophages [88]. Normal rat alveolar macrophages were incubated with two different types of liposomes—MLV (<5 lamellae) and REV (reverse evaporation vesides) (<5 lamellae)—and adjusted to contain equal amounts of lipid and entrapped muramyldipeptide. Similarly, it was found that the extent of tumoricidal activity was higher in the case of MLV than that of REV. Recently developed poly(ethylene glycol)-coated liposomes, known as Stealth liposomes, are not readily taken up by the macrophages in the reticuloendothelial system; however, they do stay in the circulation for a relatively long period of time [89–91]. It is simply because of the fact that long PEG chains impart steric hindrance in the interaction of liposomes with lipoproteins. In addition, it has been shown that incorporation of different polymers into liposomes can enhance liposome circulation time in a concentration-dependent manner. On the other hand, peptide-grafted liposomes were found to be more leishmanicidal than its ungrafted counterpart. This is because the peptide-grafted liposomes are phagocytosed more efficiently by macrophages [92]. Recently, it has been demonstrated that incorporation of a peptide from HIV-1 can also facilitate intracellular delivery of protein drugs [93].

Macrophages possess various receptors such as Fc receptors, complement, fibronectin lipoprotein, mannosyl, galactosyl, and many other receptors. These macrophage surface receptors determine the control of activities such as activation, recognition, endocytosis, and secretion. A useful approach for promoting the uptake of liposomal content by macrophages is to incorporate ligands capable of interacting with macrophage surface receptors [94]. Various research groups showed that uptake of ligand-incorporatcd liposomes was significantly higher than liposomes without ligands [95–101]. For example, mannose receptors on the macrophage surface have been exploited by developing neoglycoprotein and mannose residue on the surface of interacting drug carriers [101–103]. Studies on the interaction of mannosylated liposomes with macrophages indicate that targeted liposomes present much higher affinity for those cells than the unconjugated vesicles.

Liposomes composed of mannosylated *myo*-inositol (extracted from the cell wall of microorganisms) have been shown to be taken up by peritoneal macrophages.

Liposomes were composed of egg phosphatidylcholine, cholesterol, and dicetylphosphate with or without monoarachidic acid esters (MAEs). The liposomes with MAEs were rapidly eliminated from blood. More than 60% of the eliminated liposomes were recovered in the liver and spleen [98]. The modification did not, however, affect the distribution to the other tissues. The modified liposomes are taken up by Kupffer cells of the liver and resident macrophages of the spleen via mannose receptor-mediated binding endocytosis. Moreover, both *in vitro* and *in vivo* uptake of liposomes have been increased extraordinarily by targeting these carriers through Fc surface receptors of macrophages. Coupling of rabbit immunoglobulin with liposomes increases the uptake by rat liver macrophages more than five times compared with control liposomes [103–105]. *In vivo* study demonstrated that 80–85% of the coupled liposomes were accumulated in the liver within 1 h of the injection. This was because of the presence of the Fc moiety in immunoglobulin-coupled liposomes. Mouse monoclonal antibody modified with W-succinimidyl-3-(2-pyridyldithio) propionate in order to conjugate with liposomes containing dideoxycytidine triphosphate by covalent bonding was found to home entrapped material to the macrophages. The uptake of antibody-conjugated liposomes was four to six times higher than that of plain liposomes. This increase in uptake was thought to be due to the Fc receptor-mediated binding [105].

In a manner similar to liposomes, size, surface property, composition, concentration, and hydrophilicity or lipophilicity of microspheres and nanoparticles play a significant role in the uptake by macrophages. The events of phagocytosis include contact with pseudopods of macrophages followed by their engulfment into the cytoplasm by lamellipods [106]. Hydrophobic and relatively large microspheres are more susceptible to phagocytosis than their hydrophilic counterparts. Likewise, nanoparticles with lipophilic coating are taken up by macophages more readily as compared to their hydrophilic counterparts. Coating the particle surface with opsonic materials and activating macrophages with various activating factors can improve the extent of phagocytosis. Incubation time and dose of the vehicles can also control the process of phagocytosis.

The influence of surface charge and size of microspheres on their phagocytosis by mouse peritoneal macrophages was studied by using polystyrene and phenylated polyacrolein microspheres of different diameter as well as modified cellulose microspheres with different surface charge [107]. It was found that efficiency of uptake was maximum when size range of microsphere was on the order of 1–2 μm. For both negatively and positively charged particles, the extent of phagocytosis was increased with increasing zeta potentials, and was the lowest when zeta potential was zero [108]. However, there was no significant difference in the phagocytosis between the cationic and anionic surfaces when compared at a zeta potential of the same absolute value. Modified cellulose microspheres were allowed to incubate with macrophages in order to study the influence of the hydrophobicity on macrophage phagocytosis. Hydrophobic microspheres prepared from benzoyl cellulose were the most susceptible to phagocytosis and the nonionic hydrophilic ones were the least susceptible. It was observed that an optimal surface hydrophobicity is necessary for the microspheres to be phagocytosed [109].

Similarly, nanoparticles made from polyalkylcyanoacrylate and polymethylmethacrylate and human serum albumin microspheres have been used to study the influence of various parameters that can regulate uptake by human macrophages. The incorporation of lipophilic polymethylmethacrylate in nanoparticles was found to

result in better phagocytosis, when compared with polyalkylcyanoacrylate nanoparticles of similar size range [109]. Polybutylcyanoacrylate nanoparticles coated with lipophilic Pluronic F68—a biocompatible poloxamer—increased phagocytosis by nearly 50%, while Pluronic F108 had no influence. Nanoparticles of the same material were phagocytosed to a larger extent if they were of larger diameter. For example, phagocytosis of nanoparticles made from human serum albumin of 1.5 μm in diameter was higher than that of 200 nm diameter nanoparticles. Recently, it was found that poly(ethylene glycol)-distearate-incorporated microspheres can modify the extent of the phagocytosis depending on the concentration of the excipient used in their preparation [110]. It seems that microsphere association or uptake by macrophages might be a saturable process. For a given cell density, uptake of particles into macrophages was also dose dependent. In one study, the most avid uptake was observed with a microsphere dose of 1 mg/L; however, there was a gradual reduction in the uptake as microsphere dose was increased further [111].

In a manner that is typical of a particulate delivery system, coating of microspheres with opsonic materials and amphiphiles can modify significantly the extent of phagocytosis by macrophages, depending on the state of macrophage activation. Several proteins such as γ-globulin, human fibronectin, bovine tuftsin, and gelatin enhance the phagocytosis, while bovine serum albumin reduces the phagocytosis of cellulose microspheres [112]. It was also observed that the presence of fetal calf serum increases the phagocytosis of gelatin-grafted cellulose microspheres, while showing no effect on other protein-grafted microspheres. The influence of cross-linking and concentration of gelatin microspheres on the phagocytosis by mouse peritoneal macrophages has also been studied. Glutaraldehyde-mediated cross-linking of gelatin microspheres was found to facilitate interferon targeting to macrophages [113]. The study showed that phagocytosis of microspheres decreased with decreasing concentration of gelatin and glutaraldehyde and was proportional to the amount of microspheres added, until a saturation of phagocytosis was observed at higher doses of microspheres. It was also demonstrated that precoating or surface immobilization with gelatin was the most effective method to enhance the phagocytosis among all other opsonic proteins. Phagocytic uptake of poly styrene microspheres, coated with a series of perfluoroalkylated amphiphiles derived from phosphocholine and poly(ethylene glycol), was studied in the presence or absence of serum and peritoneal macrophages were used as phagocytic cells [114]. Phagocytic uptake tests carried out at 37 °C showed that microspheres coated with any surfactants cause a decrement in the phagocytosis under both conditions—either in the presence or absence of serum. However, the extent of decrease varied among the surfactants, and in most instances, the presence of serum had no influence on the phagocytosis when microspheres were coated with the same surfactant.' Microspheres from some biodegradable substances such as copolymers of polylactic acid and polyglycolic acid, cross-linked potato starch, hydroxyethyl starch, dextran, lichenan, and mannan are found to be successfully phagocytosed by macrophages [115–119].

When encapsulation materials produced from nanoparticles in the 1–100 nm size range have a larger surface area for the same volume, smaller pore size improved both solubility and structural properties. Hence, the drug nanoparticles, irrespective of the method of production, represent a technology that may overcome solubility problems and bioavailability problems of all poorly soluble drugs. The transforma-

tion of any drug into drug nanoparticles, leading to an increase in saturation solubility and dissolution velocity and providing the general feature of an increased adhesiveness to surfaces, is one of the most important achievements. Surface modification of the drug nanocrystals can further increase the benefits, by producing mucoadhesive nanosuspensions for oral application (e.g., targeting to the stomach or colon) or surface-modified site-specific nanoparticles for intravenous injection (e.g., targeting to the brain or bone marrow). A fusion of the novel nanosuspension technology with the traditional dosage forms (e.g., incorporating drug nanoparticles into pellets or tablets for oral delivery) is also a noteworthy advantage.

## ACKNOWLEDGMENTS

The authors thank Prashant Sharma, research scholar, for his help in the preparation of the manuscript for this chapter. We are also grateful to the coordinator of the unit, Prof. M. Saleemuddin, for the facilities provided to complete this task.

## REFERENCES

1. Jones K, Hoggard PG, Sales SD, Khoo S, Davey R, Back DJ. Differences in the intracellular accumulation of HIV-protease inhibitors *in vitro* and the effect of active transport. *AIDS* 2001;15:675–681.
2. Leveque D, Jehl F. P-glycoprotein and pharmacokinetics. *Anticancer Res* 1995;15: 331–336.
3. Sadeque AJ, Wandel C, He H, Shah S, Wood AJ. Increased drug delivery to the brain by P-glycoprotein inhibition. *Clin Pharmacol Ther* 2000;68:231–237.
4. Taylor EM. The impact of efflux transporters in the brain on the development of drugs for CNS disorders. *Clin Pharmacokinet* 2002;41:81–92.
5. Meaden ER, Hoggard OG, Newton P, Tjia JF, Aldam D, Cornforth D, Lloyd J, Williams J, Back DJ, Khoon SH. P-glycoprotein and MRP1 expression and reduced ritonavir and saquinavir accumulation in HIV-infected individuals. *J Antimicrob Chemother* 2002;50:583–588.
6. http://www.merck.com. Introduction: administration and kinetics of drugs. Merck Manual Home Edition: Updated 1 Feb 2003.
7. Fuseau E, Petricoul O, Moore KHP, Barrow A, Ibbotson T. Clinical pharmacokinetics of intranasal sumatriptan. *Clin Pharmacokinet* 2002;41:801–811.
8. Peterson L, Gerding D. Influence of protein binding of antibiotics in serum pharmacokinetics and extra-vascular penetration: clinically useful concepts. *Rev Infect Dis* 1980;2:340–348.
9. http://www.boomer.org/c/p/1/Ch18/Ch1803.html. Factors affecting drug distribution.
10. Oie S. Drug distribution and binding. *J Clin Pharmacol* 1986;26:583–586.
11. Mendell GL. Uptake, transport delivery and intracellular activity of antimicrobial agents. *Pharmacotherapy* 2005;25:130S–133S.
12. Abbon NJ, Romero JA. Transporting therapeutics across the blood–brain barrier. *Mol Med Today* 1996;2:106–113.
13. Misra A, Ganesh S, Shahiwala A, Shah SP. Delivery to the central nervous system: a review. *J Pharm Pharmaceut Sci* 2003;6:252–273.

14. Audus KL, Chikhale PJ, Miller DW, Thompson SE, Borchadt RT. Brain uptake of drugs: chemical and biological factors. *Adv Drug Res* 1992;23:1–64.
15. Oldendorf WH. Lipid solubility and drug penetration of the blood–brain barrier. *Proc Exp Biol Med* 1974;14:813–816.
16. Barling RWA, Selkont JB. The penetration of antibiotics into cerebrospinal fluid and brain tissue. *J Antimicrob Chemother* 1992;4:203–227.
17. de Lange ECM, Danhof M. Considerations in the use of cerebrospinal fluid pharmacokinetics to predict brain target concentrations in the clinical setting: implications of the barriers between blood and brain. *Clin Pharmacokinet* 2002;10:691–703.
18. Khedun SM, Maharaj B, Moodley J. Effects of antihypertensive drugs on the unborn child: what is known, and how should this influence prescribing? *Pediatr Drugs* 2000;2:419–436.
19. Bullingham RE, Nicolls AJ, Kamm BR. Clinical pharmacokinetics of mycophenolate mofetil. *Clin Pharmacokinet* 1998;34:429–455.
20. Roberts MS, Magnusson BM, Burczynski FJ, Weiss M. Enterohepatic circulation: physiological, pharmacokinetics and clinical implications. *Clin Pharmacokinet* 2002;41: 751–790.
21. Pauli-Magnus C, Stiegar B, Meier Y, Kullak-Ublick GA, Meier PJ. *J Hepatol* 2005;43: 342–357.
22. Schwarz UI, Gramatte T, Krappweis J, Oertel R, Kirch W. P-glycoprotein inhibitor erythromycin increases oral bioavailability of talinolol in humans. *Int J Clin Pharmacol Ther* 2000;38:161–167.
23. Curry SH. Binding to plasma protein. In Curry SH, Ed. *Drug Disposition and Pharmacokinetics with a Combination of Pharmacological and Clinical Relationships*, 3rd ed. Oxford, UK: Blackwell Scintific Publications; 1980.
24. Tillement JP, Zini R, d'Athis P, Vassent G. Binding of certain acidic drugs to human albumin: theoretical and practical estimation of fundamental parameters. *Eur J Clin Pharmacol* 1974;7:307–313.
25. Koch-Weser J, Sellers EM. Binding of drugs to serum albumin (2 parts) *N Engl J Med* 1976;294:311–316, 526–531.
26. Urien S, Nguyen P, Berlioz S, Bree F, Vacherot F, Tillement JP. Characterization of discrete classes of binding sites of human serum albumin by application of thermodynamic principles. *Biochem J* 1994;302:69–72.
27. Kremer JMH, Wilting J, Janssen LMH. Drug binding to human $\alpha_2$-acid glycoprotein in health and disease. *Pharmacol Rev* 1988;40:1–47.
28. Lemaire M, Urien S, Albengres E, et al. Lipoprotein binding of drugs. In Reidenberg MM, Erill S, Eds. *Drug Protein Binding*. New York: Praeger, 1986, pp 93–108.
29. Zamacona MK, Suarez E, Garcia E, Aguirre C, Calvo R. The significance of lipoproteins in serum binding variations of propofol. *Anasth Analg* 1998;87:1147–1151.
30. Barza M. Principles of tissue penetration antibiotics. *J Antimicrob Chemother* 1981;3:45–65.
31. Craig WA. Tissue binding of antimicrobials: a brief review. *Infection* 1976;4(Suppl 2):S160–S163.
32. Yogokawa K, Ishizaki J, Ohkuma S, Miyamoto K. Influence of lipophilicity and lysosomal accumulation on tissue distribution kinetics of basic drugs: a physiologically based pharmacokinetic model. *Methods Find Exp Clin Pharmacol* 2002;24:81–93.
33. Daniel WA. Mechanisms of cellular distribution of psychotropic drugs. Significance for drug action and interactions. *Prog Neuro-psychopharmacol Biol Psychiatr* 2003;27: 66–73.

34. Daniel WA, Bickel MH, Honegger UE. The contribution of lysosomal trapping in the uptake of desipramine and chloroquine by different tissues. *Pharmacol Toxicol* 1995;77: 402–406.
35. Venkatesan K. Clinical pharmacokinetic considerations in the treatment of patients with leprosy. *Clin Pharmacokinet* 1989;16:365–386.
36. Langer R. 1994 Whittacker lecture: polymers for drug delivery and tissue engineering. *Annu Biomed Eng* 1995;23:101–111.
37. Mathiovitz E, Jacob JS, Jong YS, et al. Biologically erodable microspheres as potential oral drug delivery systems. *Nature* 1997;386:410–414.
38. Lobenberg R, Araujo L, Kreuter J. Body distribution of azidothymidine bound to nanoparticles after oral administration. *Eur J Pharm Biopharmt* 1997;44:127–132.
39. Lobenbeg R, Kreuter J. Macrophage targeting of azidothymidine: a promising strategy for AIDS therapy. *AIDS Res Hum Retroviruses* 1996;12:1709–1715.
40. Bodor N, Kaminski JJ. Prodrugs and site-specific chemical delivery systems. *Annu Rep Med Chem* 1987;22:303–313.
41. Lambert DM. Rationale and applications of lipids as prodrug carriers. *Eur J Pharm Sci* 2000;11:S15–S27.
42. Chonn A, Cullis PR. Recent advances in liposomal drug delivery systems. *Curr Opin Biotechnol* 1995;6:698–708.
43. Lian T, Rodney J, Ho Y. Trends in developing liposome drug delivery systems. *J Pharm Sci* 2001;90:667–680.
44. Gregoriadis G, Ed., *Liposomes as Drug Carriers, Recent Trends and Progress*. Chichester: Wiley; 1988.
45. Martin IJ. Specialized drug delivery systems, manufacturing and production technology. In Tyle P, Ed., *Drugs and Pharmaceutical Series*. Volume 41. New York: Marcel Dekker; 1990, pp 267–316.
46. Alving CR. Liposomes as carriers of antigens and adjuvants. *J Immunol Methods* 1991;140:1–13.
47. Fidler IJ, Kleinerman ES. Clinical application of liposomes containing macrophage activators for therapy of cancer metastasis. *Adv Drug Deliv Rev* 1994;13:325–340.
48. Gregoriadis G, Gursel M, McCormack B. Liposomes as immunological adjuvants and vaccine carriers. *J Control Release* 1996;41:49–56.
49. Bendas G. Immunoliposomes: a promising approach to targeting cancer therapy. *Biodrugs* 2001;15:215–224.
50. Muggia PM. Liposomal encapsulated anthracyclines: new therapeutic horizons. *Curr Oncol Rep* 2001;3:156–162.
51. Vyas SP, Sihorkar V. Endogenous ligands and carriers in non-immunogenic site-specific drug delivery. *Adv Drug Deliv Rev* 2000;43:101–164.
52. Sunamoto J, Sakai K, Sato T, Kondo H. Molecular recognition of polysaccharide-anchored liposomes. Importance of sialic acid moiety on liposomal surface. *Chem Lett* 1988;43:1781–1784.
53. Puri N, Weyland EH, Abdel-Rahman SM, Sinko PJ. An investigation of the intradermal route as an effective means of immunization for microparticulate vaccine delivery systems. *Vaccine* 2000;18:2600–2612.
54. Muranishi S, Takuya F, Murakami M, Yamamoto A. Potential for lymphatic targeting of peptides. *J Control Release* 1997;46:157–164.
55. Harvie P, Desormeaux A, Gagne N, Tremblay M, Poulin L, Beauchamp D, Bergeron MG. Lymphoid tissue targeting of liposome-encapsulated 29,39-dideoxyinosine. *AIDS* 1995;9: 701–707.

56. Wu MS, Robbins JC, Bugianesi RL, Ponpipom MM, Shen TY. Modified *in vivo* behaviour of liposomes containing synthetic glycolipids. *Biochim Biophys Acta* 1981;674:19–29.
57. Mangat S, Patel H. Lymph node localization of nonspecific antibody coated liposomes. *Life Sci* 1985;36:1917–1925.
58. Dufresne I, Desormeaux A, Bestman-Smith I, Gourde P, Tremblay MJ, Bergeron MG. Targeting lymph nodes with liposomes bearing anti-HLA-DR Fab2 fragments. *Biochim Biophys Acta* 1999;1421:284–294.
59. Oussoren C, Storm G. Lymphatic uptake and biodistribution of liposomes after subcutaneous injection. III Influence of surface modification with poly(ethyleneglycol). *Pharm Res* 1997;14:1479–1484.
60. Allen TM, Hansen CB, Guo LSS. Subcutaneous administration of liposomes: a comparison with the intravenous and intraperitoneal routes of injection. *Biochim Biophys Acta* 1993;1150:9–16.
61. Phillips WT, Klipper R, Goins B. Novel method of greatly enhanced delivery of liposomes to lymph nodes. *Pharm Exp Ther* 2000;295:309–313.
62. Schwartz RS, Tanaka Y, Fidler IJ, Chiu DT, Lubin B, Schroit AJ. Increased adherence of sickled and phosphatidylserine-enriched human erythrocytes to cultured human peripheral blood monocytes. *J Clin Invest* 1985;75:1965–1972.
63. Senior J. Fate and behaviour of liposomes *in vivo*: a review of controlling factors. *Ther Drug Carrier Syst* 1987;3:123–193.
64. Owais M, Ahmed I, Krishnakumar K, Jain RK, Bachhawat BK, Gupta CM. Tuftsin-bearing liposomes as drug vehicles in treatment of experimental aspergillosis. *FEBS Lett* 1993;326:56–58.
65. Masood AK, Faisal SM, Haque W, Owais M. Immunomodulator tuftsin augments antifungal activity of amphotericin B against experimental murine candidiasis. *J Drug Target* 2002;10:185–192.
66. Masood A, Khan TH, Nasti Khanam S, Mallick AI, Ahmad FirozHaq W, Ahmed N, Owais M. Co-administration of tuftsin and liposomized nystatin can combat less susceptible *C albicans* in neutropenic mice. *FEMS Immunol Med Micobiol* 2004;42:249–258.
67. Singhal A, Bali A, Jain RK, Gupta CM. Specific interactions of liposomes with PMN leukocytes upon incorporating tuftsin in their bilayers. *FEBS Lett* 1984;178:109–113.
68. Gupta CM, Puri A, Jain RK, Bali A, Anand N. Protection of mice against *Plasmodium berghei* infection by a tuftsin derivative. *FEBS Lett* 1986;205:351–354.
69. Guru PY, Agrawal AK, Singha UK, Singhal A, Gupta CM. Drug targeting in *Leishmania donovani* infections using tuftsin-bearing liposomes as drug vehicles. *FEBS Lett* 1989;245:204–208.
70. Alving CR. Liposomal vaccines: clinical status and immunological presentation for humoral and cellular immunity. *Ann NY Acad Sci* 1995;754:143–152.
71. Pluschke G, Overath P. Influence of changes in polar head group composition on the lipid phase transition and characterization of a mutant containing only saturated phospholipid acyl-chains. *J Biol Chem* 1981;268:3207–3212.
72. Owais M, Gupta CM. Liposome mediated cytosolic delivery of macromolecules and its possible use in vaccine development. *Eur J Biochem* 2000;267:3946–3956.
73. Ahmad N, Masood AK, Owais M. Fusogenic potential of prokaryotic membrane lipids: implication in vaccine development. *Eur J Biochem* 2001;268:5667–5675.
74. Sproitt GD, Dicaire CJ, Gurnani K, Deschatelets LA, Krishnan L. Liposome adjuvants prepare from the total polar lipids of *Haloferax volcanii, Planococcus* spp. *and Bacillus*

*firmus* differ in ability to elicit and sustain immune responses. *Vaccine* 2004;22:2154–2162.

75. Krishnan L, Dicaire CJ, Patel GB, Sprott GD. Archeosome vaccine adjuvants induce strong humoral, cell mediated and memory responses: comparison to conventional liposome and alum. *Infect Immun* 2000;68:54–63.
76. Faisal MS, Masood AK, Tahseen HN, Nadeem A, Owais M. Antigen entrapped in the escheriosomes leads to the generation of $CD4^+$ helper and $CD8^+$ cytotoxic T-cell response. *Vaccine* 2003;21:2383–2393.
77. Owais M, Khan MA, Agrewala JN, Bist D, Gupta CM. Use of liposomes as an immunopotentiating delivery system: in perspective of vaccine development. *Scand J Immunol* 2001;54:125–132.
78. Agrewala JN, Owais M, Gupta CM, Mishra GC. Antigen incorporation into liposomes results in the enhancement of IL-4 and IgG1 secretion: evidence for preferential expansion of Th-2 cells. *Cytokines Mol Ther* 1996;2:59–65.
79. Arigita C, Bevaart L, Everse LA, Koning GA. Liposomal meningococcal B vaccination: role of dendritic cell targeting in the development of a protective immune response. *Infect Immun* 2003;71:5210–5218.
80. Duzgunes N, Pretzer E, Simoes S, Slepushkin V, Konopka K, Flasher D, de Lima MC. Liposome-mediated delivery of antiviral agents to human immunodeficiency virus-infected cells. *Mol Membr Biol* 1999;16:111–118.
81. Raz A, Bucana C, Fogler WE, Poste G, Fidler IJ. Biochemical, morphological and ullrastructural studies on the uptake of liposomes by murine macrophages. *Cancer Res* 1981;41:487–494.
82. Hsu MJ, Juliano RL. Interactions of liposome with reticulo-endothelial system. II. Nonspecific and receptor mediated uptake of liposomes by mouse peritoneal macrophages. *Biochim Biophys Acta* 1982;720:411–419.
83. Heath TD, Lopez NG, Papahadjospoulous D. The effect of liposome size and surface charge on liposome mediated delivery of methotrexate–aspartate to cells *in vitro*. *Biochim Biophys Acta* 1985;820:74–84.
84. Allen TM, Austin GA, Chonn A, Lin L, Lee KC. Uptake of liposomes by cultured mouse bone marrow macrophages: influence of liposome composition and size. *Biochim Biophys Acta* 1991;106:56–64.
85. Schwendener RA, Lagocki PA, Rahman YH. The effects of charge and size on the interaction of unilamnellar liposomes with macrophages. *Biochim Biophys Acta* 1984;772: 93–101.
86. Rahman YE, Cerny EA, Patel KR, Lau EH, Wright BJ. Differential uptake of liposomes varying in size and lipid composition by parenchymal and Kupffer cells of mouse liver. *Life Sci* 1982;31:2061–2071.
87. Bakker-Woudenbcrg IAJM, Lokerse AF, Roerdink FH. Effect of lipid composition on the activity of liposome-entrapped ampicillin against intracellular Listeria monocytogenes. *Antimicrob Agents Chemother* 1988;32:1560–1564.
88. Maruyama K. Liposomal technology in cancer therapy. *Biother Jpn* 1998;12:1051–1058.
89. Torchilin VP, Trabetskoy VS, Whiteman KR, Caliceti P, Femiti P, Veronese PM. New synthetic amphiphilic polymers for steric protection of liposomes *in vivo*. *J Pharm Sci* 1995;84:1049–1053.
90. Torchilin VP. Polymer coated long circulating microparticulate pharmaceuticals. *J Microencaps* 1998;15:1–19.
91. Torchilin VP. Affinity liposomes *in vivo*: factors influencing target accumulation. *J Mol Recogn* 1996;9:335–346.

92. Banerjee G, Medda S, Basu MK. A novel peptide-grafted liposomal delivery system targeted to macrophages. *Antimicrob Agents Chemother* 1998;42:348–351.
93. Torchilin VP, Rammohan R, Weissig V, Levchenko TS. TAT peptide on the surface of liposomes affords their efficient intracellular delivery even at low temperature and in the presence of metabolic inhibitors. *Proc Natl Acad Sci USA* 2001;15:8786–8791.
94. Kaneda Y. Virosomes: evolution of the liposomes as a targeted drug delivery system. *Adv Drug Deliv Rev* 2000;43:197–205.
95. Gregoriadis G. Liposome as immunological adjuvants: approaches to immunopotentiation including ligand-mediated targeting to macrophages. *Res Immunol* 1992;143:178–185.
96. Aramaki Y, Murai M, Tsuchiya S. Activation of Fc receptor mediated phagocytosis by mouse peritoneal macrophages following the intraperitoneal administration of liposomes. *Pharm Res* 1994;11:518–521.
97. Dutta M, Bandyopadhyay R, Basu MK. Neoglycosylated liposomes as efficient ligand for the evaluation of specific sugar receptors on macrophages in health and in experimental leishmaniasis. *Parasitology* 1994;109:139–147.
98. Yachi K, Kikuchi H, Yamauchi H, Hirota S, Tomikawa M. Distribution of liposomes containing mannobiose esters of fatty acid in rats. *J Microencaps* 1995;12:377–388.
99. Shao J, Ma JKH. Characterization of mannosylated liposome system for drug targeting to alveolar macrophages. *Drug Deliv* 1997;4:43–48.
100. Muller CD, Schuber F. Neo-mannosylated liposomes: synthesis and interaction with mouse Kupffer cells and resident peritoneal macrophages. *Biochim Biophys Acta* 1989;986:97–105.
101. Kole L, Sarkar K, Mahato SB, Das PK. Neoglycoprotein conjugated liposomes as macrophage specific drug carrier in the therapy of leishmaniasis. *Biochem Biophys Res Commun* 1994;200:351–358.
102. Barrat G, Term JP, Yapo A, Petit JF. Preparation and characterisation of liposome containing mannosylated phospholipids capable of targeting to macrophages. *Biochim Biophys Acta* 1986;862:153–164.
103. Derksen JTP, Morselt HWM, Kalicharan D, Hul-staert CE, Scherphof GL. Interaction of immunoglobul in-coupled liposomes with rat liver macrophages *in vitro*. *Exp Cell Res* 1987;167:105–115.
104. Derksen JTP, Morselt HWM, Scherptiof GL. Uptake and processing of immunogliobulin-coated liposomes by sub-populations of rat liver macrophages. *Biochim Biophys Acta* 1988;971:127–136.
105. Bettageri GV, Black CDV, Szebeni J, Wahl LM, Weinstein JN. Fc-receptor-mediated targeting of antibody-bearing liposomes containing dideoxycytidine triphosphate to human monocyte/macrophage. *J Pharm Pharmacol* 1993;45:48–53.
106. Schafer V, Briesen HV, Riibsamen-Waigmann H, Steffan AM, Royer C, Kreuter J. Phagocytosis and degradation of human serum albumin microspheres and nanoparticles in human macrophages. *J Microencaps* 1994;11:261–269.
107. Tabata Y, Ikada Y. Effect of the size and surface charge of polymer microspheres on their phagocytosis by macrophage. *Biomaterials* 1988;9:356–362.
108. Roser M, Fischer D, Kissel T. Surface-modified biodegradable albumin nano- and microspheres. II: Effect of surface charges on *in vitro* phagocytosis and biodistribution in rats. *Eur J Pharm Biopharm* 1998;4:255–263.
109. Schafer V, von Briesen H, Andreesen R, Steffan AM, Royer C, Troster S, Kreuter J, Waigmann HR. Phagocytosis of nanoparticles by human immunodeficiency virus

(HIV)-infected macrophages: a possibility for antiviral drug targeting. *Pharm Res* 1992;9:541–546.

110. Lacasse FX, Filion MC, Phillips NC, Escher E, McMullen JN, Hildgen P. Influence of surface properties at biodegradable microsphere surfaces: effects of plasma protein adsorption and phagocytosis. *Pharm Res* 1998;15:312–317.
111. Akhtar S, Lewis KJ. Antisense oligonucleotide delivery to cultured macrophages is improved by incorporation into sustained-release biodegradable polymer microspheres. *Int J Pharm* 1997;151:57–67.
112. Ikada Y, Tabata Y. Phagocytosis of bioactive microspheres. *J Bioact Compat Polym* 1986;1:32–46.
113. Tabata Y, Ikada Y. Synthesis of gelatin microspheres containing interferon. *Pharm Res* 1989;6:422–427.
114. Privitera N, Naon R, Riess VIG. Phagocytic uptake by peritoneal macrophages of microspheres coated with phosphocholine or polyethylene glycol phosphate-derived perfluoro alkylated surfactants. *Int J Pharm* 1995;120:73–82.
115. D'Souza M, DeSouza P. Site specific microencapsulated drug targeting strategies—liver and gastro-intestinal tract targeting. *Adv Drug Deliv Rev* 1995;17:247–254.
116. Mullerad J, Apte RN, Cohen S. Delivery of IL-1 microspheres to tumor macrophages. *Proc Int Symp Control Release Bioact Mater* 1995;22:512–513.
117. Okawa T, Ichimal H, Ishida T, Kawata M, Kaguba S, Mamba K, Makita T. Implication of activation of intraperitoneal macrophages with biodegradable microspheres. *Cell Biol Int Rep* 1989;13:547–553.
118. Kanke M, Geissler RG, Deborah P, Kaplan A, DeLuca PP. Interaction of microspheres with blood constituents. *J Parent Sci Technol* 1988;42:157–165.
119. Artursson P, Arro Edman P, Ericsson JLE, Sjoholm I. Biodegradable microspheres V: stimulation of macrophages with microparticles made of various polysaccharides. *J Pharm Sci* 1987;76:127–133.

# SALT AND COCRYSTAL FORM SELECTION

Ann W. Newman, Scott L. Childs, and Brett A. Cowans
*SSCI, Inc., West Lafayette, Indiana*

**Contents**

1 Introduction
2 Salts and Cocrystals
3 Form Selection Process
  3.1 Counterion/Guest Selection
  3.2 Salt/Cocrystal Formation
  3.3 Characterization
  3.4 Evaluation
  3.5 Polymorph Screen
  3.6 Crystallization and Quantitation
4 Conclusion
  References

## 1 INTRODUCTION

Form selection is critical to the successful development of an active pharmaceutical ingredient (API) [1–5]. A number of physical properties vary with the solid form (a partial list is given in Table 1) and several properties such as crystallinity, solubility, and stability need to be measured and evaluated to determine whether a candidate will be acceptable for development [6, 7]. If other options are needed, a different crystalline form, a different salt, or a cocrystal may be explored. It has been estimated that approximately half of all administered drugs are formulated as salts [8]. With more poorly soluble drugs being found and developed, form selection can

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

**TABLE 1 Properties that Can Vary with Solid Form**

| |
|---|
| Bioavailability |
| Hygroscopicity |
| Dissolution rate |
| Solubility |
| Physical stability |
| Chemical stability |
| Compressibility |
| Chargeability |
| Filterability |
| Flow |
| Millability |
| Melting point |
| Color |
| Excipient compatibility |
| Morphology |
| Density |
| Taste |
| Toxicology |
| Levels of gastric irritation |
| Clinical results |

improve properties, such as solubility and ultimately bioavailability, and increase the potential viability of a candidate. Form selection is therefore an integral part of the drug development process.

Various approaches to form selection exist during development. One general approach is to perform sufficient work initially in order to choose a good candidate early (before the Investigational New Drug (IND) application is filed and clinical trials are started). This approach is believed to reduce the time to market. Another approach is to initially limit the time and money spent choosing a form early in development since many compounds will be eliminated. For this latter scenario, a form is chosen that is believed to be adequate for early trials, knowing that changing forms later will have the added cost and time commitment for bridging studies. In practice, the actual approach will be some combination of these two extremes.

Clearly, the risk assessment for each compound must be evaluated based on the information available for the drug being developed. The dosage form to be developed also needs to be considered when choosing the form, since the desired properties for an oral dosage form will be significantly different from those needed for a suspension or an IV formulation. For most compounds, a complete picture for form selection and an understanding of the solid-state properties should be developed and presented at the end of Phase II clinical trials.

It should be noted that form selection also influences intellectual property position. Crystalline forms including salts and cocrystals are patentable. Additional form selection studies may be carried out late in development for intellectual property protection on improved products. Understanding the solid-state properties of the API is critical to these endeavors.

This chapter discusses solid form screening and the selection process for pharmaceutical compounds with the main focus on salts and cocrystals. Information on

counterion/guest selection, preparation, characterization, and property evaluation are also presented. The need for polymorph screening as part of the selection process is included, and the role of quantitative methods to determine the solid forms present in API and drug product is also discussed.

## 2 SALTS AND COCRYSTALS

When ionizable groups are present in an API molecule, a salt can be formed which often provides better physical properties, such as increased water solubility or melting point. In addition, nonionizable compounds and salts may form cocrystals with the desired properties. Working definitions for these materials are given below and the concepts are shown schematically in Fig. 1.

*Salt.* Acid/base chemistry creates a new species; a free acid or free base is needed for salt formation. An acid–base complex is formed that contains the API along with additional nontoxic molecular species (counterion) in the same crystal structure.

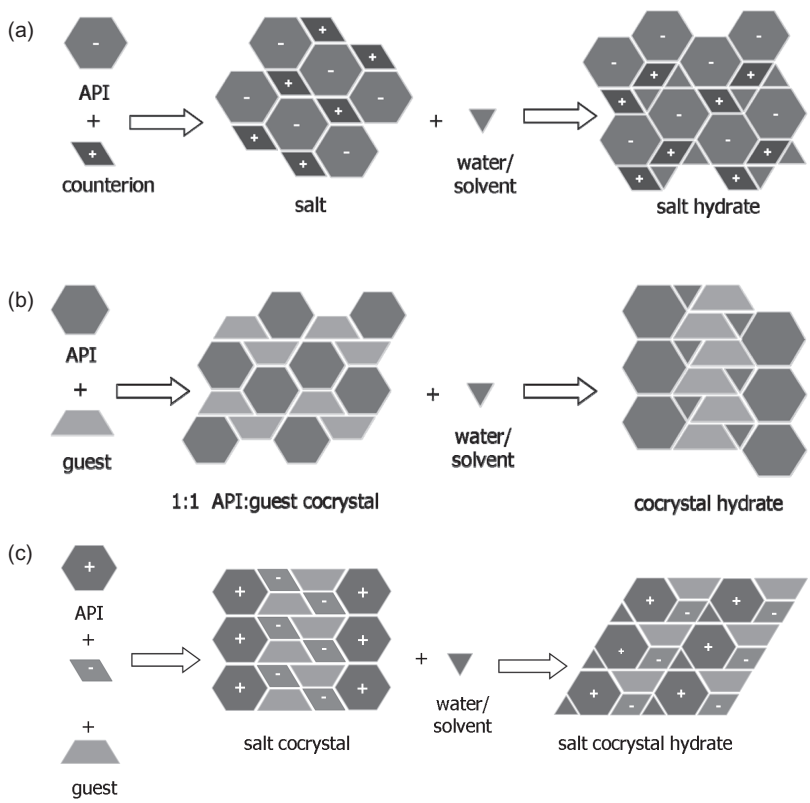


**FIGURE 1** Schematic of salt and cocrystal formation: (a) salt and salt hydrate/solvate, (b) cocrystal and cocrystal hydrate/solvate, and (c) salt cocrystal and salt cocrystal hydrate/solvate.

*Cocrystal.* A crystalline molecular complex of an API or a salt of an API; it consists of two or more components, with each component being an atom, ionic compound, or molecule.

Salts are commonly produced and used as drug substances in the pharmaceutical industry, and numerous studies of salts and salt selections have been reported [9–13]. Cocrystals are comparatively new to the pharmaceutical industry but have been studied and used in other industries for years [14]. As a solid form, cocrystals perform essentially the same function as salts according to the regulatory definition of an active moiety [15]: "the molecule or ion, excluding those appended portions of the molecule that cause the drug to be an ester, salt (including a salt with hydrogen or coordination bonds), or other noncovalent derivative (such as a complex, chelate, or clathrate) of the molecule, responsible for the physiological or pharmacological action of the drug substance." For all compounds, including cocrystals, the regulatory requirement is not the categorization of the particular form; rather, it is to adequately and accurately characterize the system, to understand scientifically the form that is being developed, and to transfer this information to the regulatory agency. From this perspective, salts, hydrates, and cocrystals are identical in that they all contain the essential active moiety plus additional ions or molecules that are present primarily to impart specific physical properties on the solid form containing the active moiety.

A number of variables need to be considered when selecting a form, such as target solubility, dosage form to be developed, loading in the dosage form, physical and chemical stability, and amount of material available. It should be emphasized that there is more than one way to select a salt or cocrystal since numerous factors should be considered in the selection process. A tiered approach to salt/cocrystal selection is helpful when screening a number of possible candidates. Common properties used to evaluate materials include crystallinity, hygroscopicity, solubility, melting point, density, and stability. An example of a decision tree designed to select a material for an oral dosage form is given in Fig. 2a. The variables included in this selection include crystallinity, hygroscopicity, solubility, and polymorphism. These decision trees can be modified for different situations. Properties needed for an injectable formulation are usually different from those required for an oral dosage form, as shown in Fig. 2b. Because the properties needed for the selection process cover a broad range, it is imperative that a multidisciplinary approach be used to choose the solid form. The three main departments that deal directly with the solid form issues are chemical development (to produce the API with the desired properties), analytical (for testing of materials), and pharmaceutics (for preformulation and formulation development). Other groups, such as discovery, toxicology, and pharmacokinetics may also provide useful information during the selection process.

Salt/cocrystal selection involves experimentally identifying and characterizing a variety of different multicomponent solids by incorporating a number of secondary molecular species into the lattice with the active moiety. By taking a broader view concerning the types of secondary molecules that are experimentally investigated and not limiting the counterions tested based on $pK_a$, a wider variety of solid forms with potentially useful properties can be considered in the process of form selection. The cocrystal discussion in this chapter is intended to act as a guide in the expansion of the targeted multicomponent solids that are investigated during form selection.

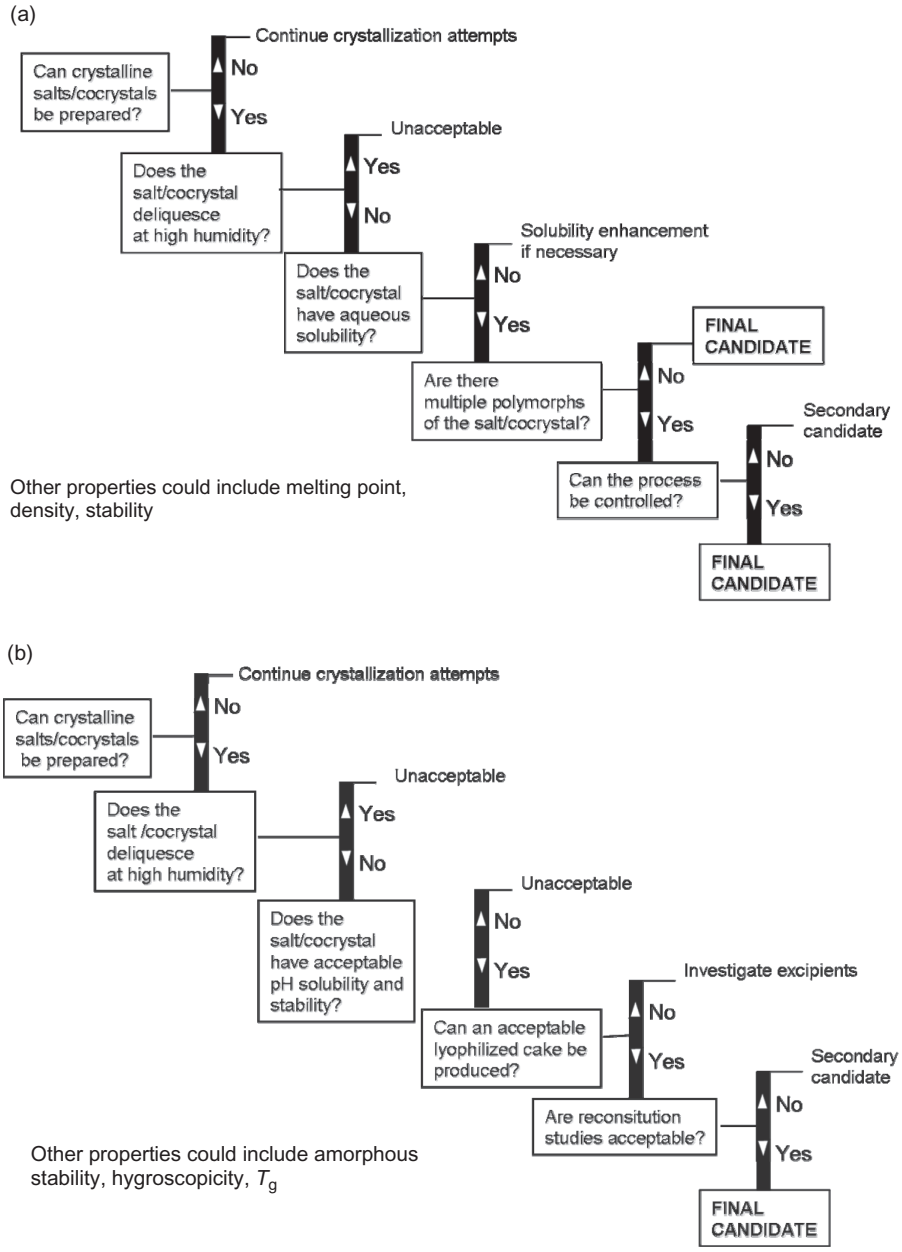


**FIGURE 2** Decision trees for salt/cocrystal selection: (a) example for an oral dosage form and (b) example for an injectable formulation.

## 3 FORM SELECTION PROCESS

The form selection process is divided into five sections: counterion/guest selection, formation, characterization, evaluation, and polymorph screening. Each section provides an overview of common methodologies used.

### 3.1 Counterion/Guest Selection

Pharmaceutically acceptable counterions and cocrystal guests are often used to simplify the regulatory concerns surrounding the product, since information about the materials is already known and has been filed previously with the regulatory agency. The list of pharmaceutically acceptable compounds typically includes molecules that are readily found in marketed products. Figure 3 shows the most common counterions found in marketed products based on databases compiled in our laboratory; the results are similar to those reported elsewhere [8]. If acceptable toxicology can be demonstrated for a new entity and it provides significant property improvement, a strong case can be made for using a new substance as a guest or counterion.

There are other considerations when determining a list of possible counterions and guests. For instance, the loading in the drug product may make a high molecular weight counterion or guest unacceptable. Commonly, sodium or potassium are used as counterions for acidic APIs when high loading is a factor, but small organic species may still be preferable over inorganic ions. The type of dosage form also needs to be considered since there will be limitations for injectable, ocular, and inhalation formulations that may be different from those encountered with oral

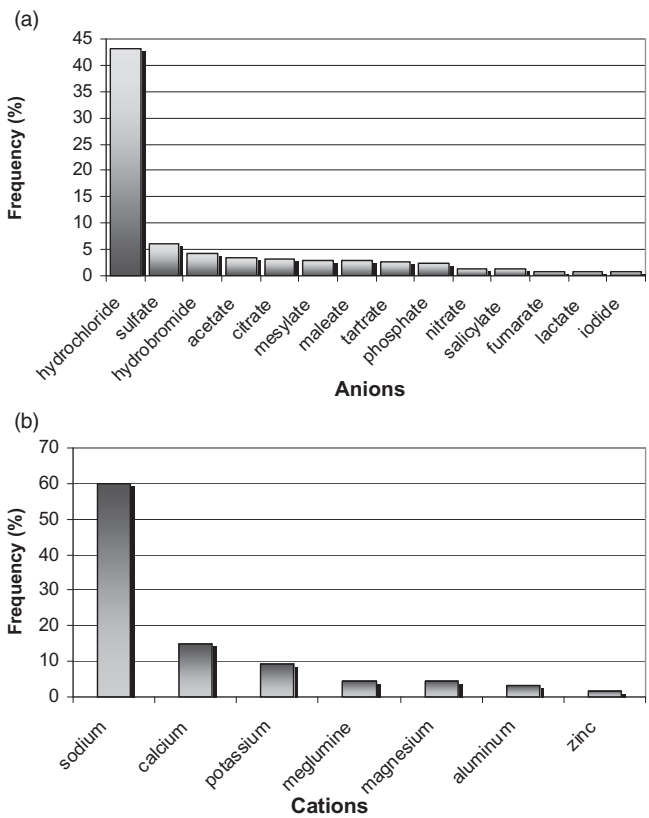


**FIGURE 3** Common counterions found in marketed pharmaceutical products: (a) anions and (b) cations.

formulations [8]. The indication expected for the API may also have limitations regarding other components that can be present.

For salts, the p$K_a$ of the relevant functional groups on the drug molecule is considered initially. A two to three unit separation between the p$K_a$ values of the functional groups on the API and the counterion is commonly invoked when making salts [16]. This is fine as a guideline, but salts and cocrystals can be formed outside this range based on the decreased solubility of the salt formed. Hydrochloride (HCl) salts are commonly produced for pharmaceutical bases, but there are issues that should be considered when using HCl [8, 17]. Some issues to consider for an HCl salt include the common ion effect *in vivo*, corrosive properties during production, and loss of HCl during processing. Many companies have a list of acceptable counterions that are used for every salt selection. Although this is an acceptable place to start, many of the counterions on the list may not be ideal for the molecule being developed. It is important to look at the chemistry and the solid-state aspects of salts and cocrystals while compiling a list of possible counterions and guests that are to be empirically evaluated.

A cocrystal screen is an empirical exercise but can be guided by an analysis of the API in order to determine the potential for cocrystallization and to gather information to assist in the selection of the most appropriate guest molecules. The number and arrangement of the hydrogen bond donors and acceptors, salt forming ability, size and conformational flexibility, stability of the homomeric structure, physical properties, and solubility requirements are all considered during the initial evaluation. Crystal engineering also relies heavily on information contained in the Cambridge Structural Database [18]. The crystal engineering approach used to design an effective cocrystal screen is based on academic studies [19–21] and the application of this methodology to pharmaceuticals has generated a number of recent successes [22].

The ability of the selected guests to form hydrogen bonded synthons with the API that are energetically more favorable than the homomeric intermolecular interactions can be an important consideration in guest selection [23]. This is consistent with the observation that after the intramolecular hydrogen bonds form, the strongest remaining hydrogen bond donors typically interact with the best acceptor sites [24]. If these groups are on the same molecule, then a homomeric form is favored; however, by choosing API:guest combinations that place the strongest donor and acceptor on different molecular species, the potential for cocrystallization is increased.

It should be noted that many of the compounds used as counterions in salt selections can also be used as guests for cocrystals. We find that pharmaceutically acceptable, weak carboxylic acids typically used in salt screens are particularly useful guest molecules. One of the most effective ways for a company to initiate cocrystal screening is by using these carboxylic acids to extend the range of a traditional salt screen of a basic API. By removing the traditional limitations on counterion selection based on p$K_a$ values, the potential for finding stable solid forms that meet the physical property requirements is increased. The concern over whether a solid is a cocrystal or a salt should be secondary to the consideration of the properties of the solid form.

Pharmaceutically "unacceptable" compounds may be employed if cocrystal formation is used initially as a purification or isolation step. The cocrystal can then be

broken to isolate the API [25]. The API can be used upon isolation/crystallization or a cocrystal or salt with a pharmaceutically acceptable guest could be produced for development.

Although finding cocrystals can be challenging, the unique benefits of cocrystals outweigh the possible difficulties because they provide a screening option for APIs that cannot form salts. Until now, nonionizable or poor salt-forming APIs typically have had no recourse for crystalline solid form variation other than the adventitious formation of hydrates or solvates. Cocrystallization provides a systematic approach to form selection for these APIs.

### 3.2 Salt/Cocrystal Formation

After selecting a list of possible counterions and guests, the crystallization experiments to be performed need to be planned. There are a number of issues to consider when producing salts and cocrystals. This section outlines some of these issues, such as solvent choice, solution crystallization experiments, nonsolvent options, and possible stoichiometries.

One of the first considerations in salt or cocrystal formation from solvents is the solubility of the API as well as the counterions or guests. A common solvent or miscible solvent system with acceptable solubility must be found for the experiments. This appears to be straightforward, but it can become a major issue for poorly soluble drugs. This is especially problematic for cocrystal formation. When a strong salt forms, charge balance requirements necessitate the crystallization of the anion and cation together and the effects of solubility of either of the individual ions (and conjugate acid/base) are minimized compared to the analogous situation of cocrystallizing two independent nonionized components. It has been shown that the phase diagram for a cocrystal is related to the solubility of the individual components [26]. For example, experiments that involve solvent systems in which the guest is very soluble and the API has very low solubility are less likely to produce a solid phase containing both components.

Other considerations are methods of preparation and the type of screening methods used. Automated methods using 96 well plates, for example, screen a large number of counterions or guests with less material [27, 28]. The amount of each sample produced is determined by the solubilities of the components and is typically small (1–3 mg). Due to the small sample size, characterization is usually limited to optical microscopy, X-ray powder diffraction (XRPD), and/or Raman spectroscopy. Crystallization conditions are limited to solvent based experiments, such as evaporations or cooling when solubility permits, and a limited temperature range is available for most plates. If solids are not formed initially, recrystallization attempts may be more difficult because of experimental limitations with the plate. Crystalline materials that are produced in the plates need to be produced on a larger scale for further characterization to confirm stoichiometry and volatile content. Traditional (manual) preparation methods typically screen a smaller number of counterions or guests and will use more material per experiment. These methods produce adequate material for preliminary characterization or additional recrystallization attempts if necessary. The use of automated versus manual techniques is a decision that is usually based on material availability and known properties.

Numerous solvent methods have been used to produce salts and cocrystals. Commonly, both the API and counterion/guest are dissolved and the solutions added together. Parameters such as the order of acid/base addition, temperature, or rate of addition can all play an important role in salt or cocrystal formation. For cocrystal experiments, the solubility of the API is often lower than that of the guest compounds being screened. This can lead to the preferential nucleation of a homomeric API solid phase instead of cocrystal formation. To reduce the effects of solubility differences, yet still retain the advantages of high throughput methods, slurry experiments are often employed in the screening process.

Salt exchanges in solution or with the aid of a resin are also possible choices depending on the properties of the API [29]. Gases, such as HCl or HBr (hydrogen bromide), have also been bubbled into API solutions to produce salts [30]. Preparation techniques such as precipitation and evaporation are commonly used to produce the solid. If precipitation does not readily occur once the components are mixed, other methods may be needed to produce solids, such as concentration of the solution, antisolvent addition, lower temperatures, or a combination of these procedures.

Nonsolvent methods can also be used. Fusion methods based on Kofler techniques have been reported for a variety of systems [31]. More recently, solid-state grinding, with or without small amounts of solvent, has been used successfully for cocrystal and salt formation [32]. Even though these, or even some solvent based methods, may not be suitable for a large scale process, the seeds obtained from these experiments can be used to produce larger amounts of the desired material.

Stoichiometry is another consideration in salt and cocrystal formation. More than one acidic or basic site may be present for the API. If more than one site is available on the API, control of stoichiometry may be an issue if the $pK_a$ values are similar. Diacid counterions, such as fumarate and malate, can also be used to investigate different stoichiometries in salts or cocrystals. An example of a system with two basic sites with similar $pK_a$ values is given in Table 2. Two different stoichiometries of a hydrobromide (HBr) salt were targeted. Based on the yield and elemental data, it was obvious that the disalt was formed in both cases, regardless of the targeted stoichiometry. Additional XRPD data showed that both experiments produced the same crystal form of the disalt.

In general, there are common solvents and procedures, such as evaporation, that are used for salt and cocrystal formation, but they will not work in every case. It is

**TABLE 2 Characterization Data on a Hydrobromide Salt**

| Assay | 1:1 HBr:API | 2:1 HBr:API | Expected Values for Disalt |
|---|---|---|---|
| C | 49.77 | 49.92 | 49.48 |
| H | 4.05 | 3.97 | 3.97 |
| N | 14.83 | 14.91 | 15.05 |
| Br | 28.86 | 29.20 | 28.63 |
| Yield | Poor yield | Good yield | — |
| XRPD | Form A | Form A | — |

important to consider the properties of the API and design experiments that are appropriate and have the greatest chance of success. A variety of methods will also produce a wider range of crystalline solids to choose from during form selection and development.

### 3.3 Characterization

Characterization is essential for identification of the solid produced during the screen and provides the information necessary to determine if the new material can easily be developed. Information from several techniques is often required to confirm that a salt or cocrystal was prepared and to determine the chemical and physical properties. Numerous references are available that describe solid-state characterization techniques for new salts and cocrystals [1, 12, 13, 33, 34]. The theory and instrumentation of commonly used techniques are not reviewed in this section, but these techniques are discussed in terms of the information provided on new salts/cocrystals.

X-ray powder diffraction (XRPD) is often used to determine if the material is crystalline and to identify the solid form. A unique XRPD pattern does not always indicate that a new salt or cocrystal was formed and the data should be compared to the XRPD patterns of known forms of the API and the counterion/guest. As shown in Fig. 4, an HCl salt preparation experiment (diethyl ether fast evaporation) using Form A of the free amine resulted in the formation of a new solid form of the free amine (Form B). The HCl salt was produced from acetone using a combination of slow cool and slow evaporation techniques. The presence, or absence, of the HCl was confirmed using elemental analysis.

Raman spectroscopy is another fast and efficient tool used to determine salt or cocrystal formation [35]. Unique spectra obtained on any new material should be compared to the spectra obtained on the initial components, as shown in Fig. 5. In this case, the Raman spectrum indicates that a cocrystal of theophylline and salicylic acid was formed and was confirmed by other methods of analysis. Infrared (IR) spectroscopy can also be used to determine salt formation. Shifts in the carbonyl bands will indicate salt formation but can also indicate a different crystal form, so comparison with known forms is critical. The presence of $NH_2$ peaks in an IR spectrum can confirm that the ionized species is present in the new material. Amorphous materials usually exhibit broad IR or Raman peaks, but they can also exhibit a unique spectrum that can make interpretation more difficult.

Solid-state nuclear magnetic resonance (ssNMR) spectroscopy is a powerful tool for studying solid forms of salts or cocrystals [36, 37]. Solid-state NMR can confirm that a salt or cocrystal was formed and provide structural or conformational information [38, 39]. Solvated forms can be identified as well. In some cases, ssNMR can distinguish solid forms that are indistinguishable by other techniques [40].

Solution nuclear magnetic resonance (NMR) spectroscopy can provide a variety of information on new solids. In many cases, a solution proton NMR can reveal if a salt is formed, even if the counterion or guest is not directly observable. The chemical shift of protons that are at or near the site of charge transfer or complex formation will change compared to the chemical shift of those protons in the neutral or uncom-

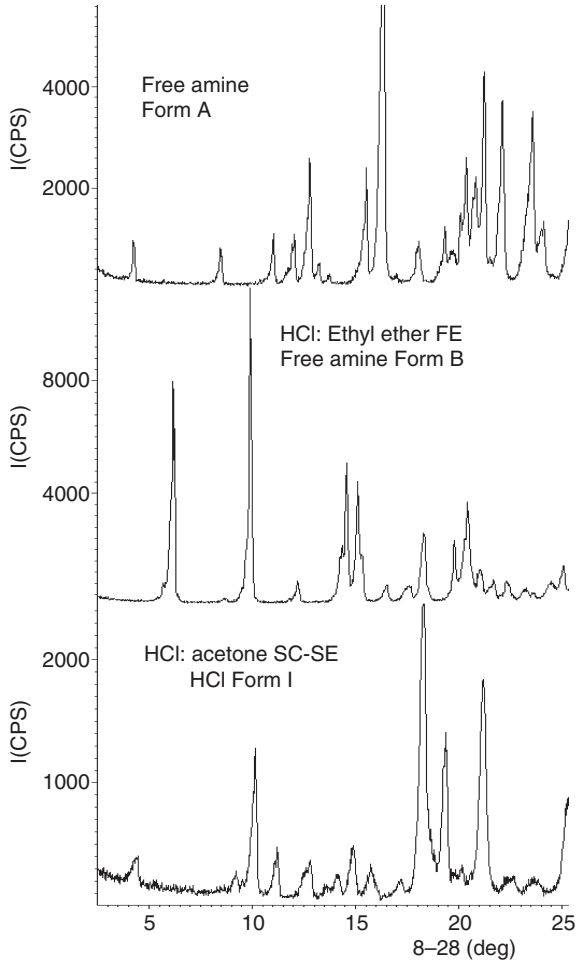


**FIGURE 4** XRPD data for HCl salt attempts: (top) free amine Form A, (middle) HCl salt attempt from ethyl ether fast evaporation that resulted in free amine Form B, and (bottom) HCl salt attempt from acetone slow cool/slow evaporation that resulted in HCl salt Form I. The presence or absence of HCl was confirmed by other methods.

plexed species. An example is given in Fig. 6, which compares the spectral region of interest for an acetate salt and the free base. The NH resonance of the secondary amine at approximately 1.9 ppm in the free base spectrum is shifted to approximately 9.5 ppm (inset) as the protons of the ammonium salt. In addition, methylene proton resonances alpha to the ammonium ion have shifted approximately 0.2 ppm downfield compared to the position of those resonances in the free base.

The stoichiometry of an organic counterion or guest can be determined as well. A stoichiometry of approximately 1:1 acetate:API was determined from the peak integration for the acetate salt (Fig. 6). If an organic solvent is present, the amount can be determined and indicates whether the material is a solvate or not. This type of information is crucial in determining the developability of the new solid and can easily be obtained on a few milligrams of material.

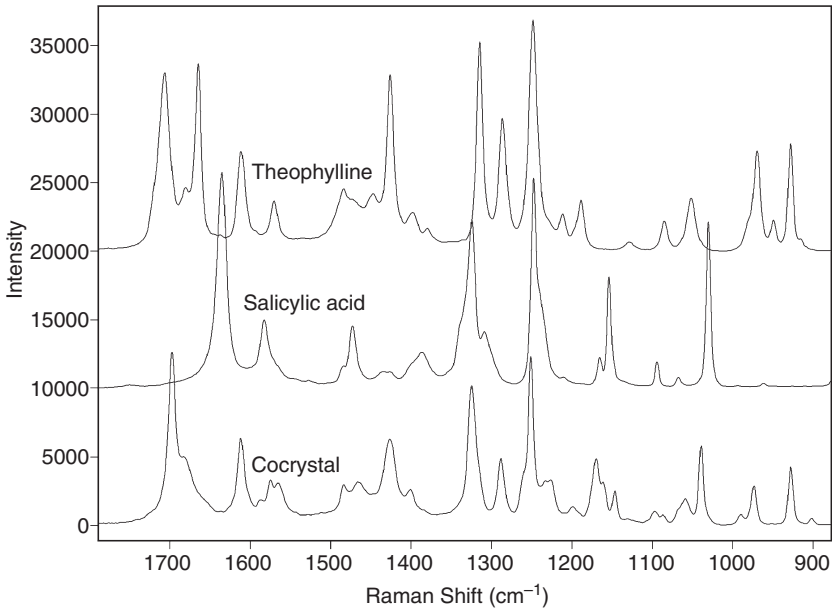


**FIGURE 5** Raman spectra of (top) theophylline, (middle) salicylic acid, and (bottom) theophylline:salicylic cocrystal.

Differential scanning calorimetry (DSC) can also be used for the characterization of new solids if the free base or acid has a distinct thermal curve [11, 12]. DSC can be very sensitive to small amounts of the initial material depending on the strength of the melting endotherm. An example is given in Fig. 7. The free base exhibits a melting endotherm at approximately 228 °C while the salt exhibits a melting endotherm at approximately 250 °C. There is an additional broader endotherm observed for the salt at approximately 149 °C, which later proved to be due to the volatization of solvent from the crystalline material.

Other characterization can also be performed depending on the amount of material available. Confirming the stoichiometry with elemental analysis is a common practice. Measuring the chemical purity with HPLC is important for many API materials that may be prone to degradation under acidic or basic conditions. Determining the solvent content using NMR, gas chromatography (GC), thermogravimetry (TG), TG-IR, KF, or other methods is critical in assessing the developability of a new salt or cocrystal.

### 3.4 Evaluation

Once acceptable salts or cocrystals are found based on the characterization data, other properties need to be evaluated to choose the best compound for development. This assessment is performed as the latter part of the tiered flowchart, as shown in the examples in Fig. 2. Any number of factors can be investigated during this period and they should be tailored to the known issues of the API as well as the anticipated dosage form.

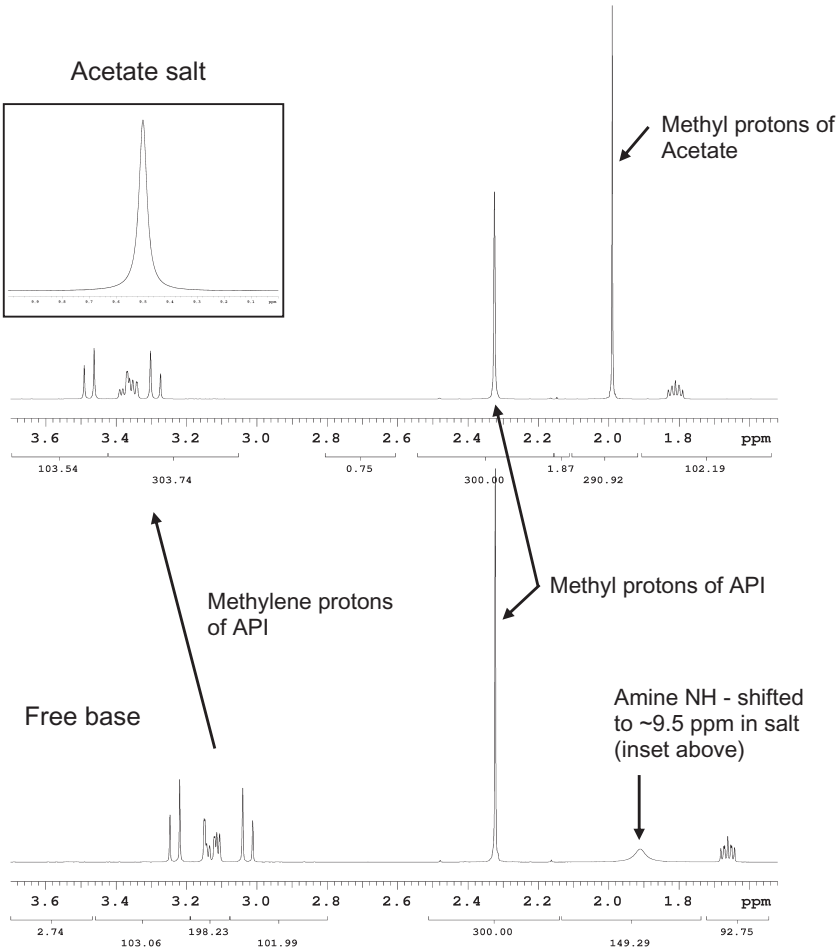


**FIGURE 6** Solution proton NMR spectra of (top) acetate salt and (bottom) free base. Shifting observed for methylene and amine protons in acetate salt spectrum is due to salt formation. Stoichiometry of the acetate salt can also be estimated by integration of the methyl proton resonances for the acetate and the API.

A wide range of acceptable values are possible for the properties identified as important, and they depend on the development plan and dosage form [9]. In general, a solubility of 0.1–10 mg/mL is targeted for a solid oral dosage form and greater than 10 mg/mL for parenteral formulations. A pH range of 3–10 for an aqueous solution is targeted for intravenous delivery. A melting point of greater than 100 °C is usually desired for acceptable processing, especially if drying or grinding steps are anticipated. Hygroscopicity can be acceptable if the material does not deliquesce at 60–75% RH or if minimal water uptake occurs during normal processing relative humidity conditions (typically 30–75% RH). Hygroscopicity is commonly evaluated by the amount of water sorbed by the material, but other changes, such as hydrate formation, should also be investigated when assessing

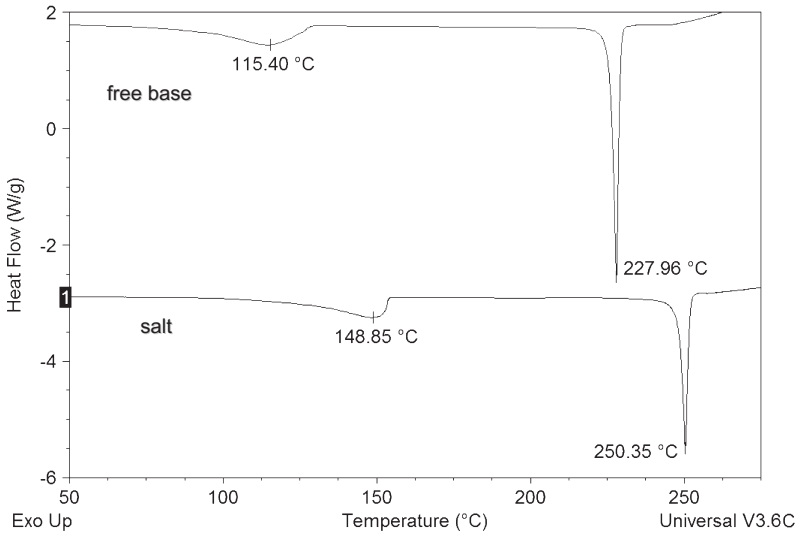


**FIGURE 7** DSC curves of a pharmaceutical (top) free base and (bottom) salt. High temperature peaks are melting endotherms and low temperature broad endotherms are due to volatilization.

hygroscopicity [41]. Chemical and physical stability can be determined using accelerated stability conditions, such as 40 °C/75% RH over 1–2 weeks. Chemical stability can be evaluated by HPLC determination of impurities and physical stability can be evaluated by XRPD to determine a change in crystal structure. Crystal size and shape can be important to handling properties in processing and manufacturing, and an early assessment of these properties can save significant development time. The propensity of a salt or cocrystal to exist as multiple crystal forms can be evaluated at this stage with an abbreviated polymorph screen. Information on the formation of crystalline solvates from certain solvents can be used as an early guide for the crystallization step and will narrow the possible solvent choices. It should be noted that even salts or cocrystals with numerous forms can easily be developed once a controlled process is in place to produce the desired form.

This type of tiered approach was reported for BMS-180431 [11]. Seven crystalline salts were initially prepared: sodium, calcium, zinc, magnesium, potassium, lysine, and arginine. Tier 1 tested the hygroscopicity and three salts moved on to the next tier (magnesium, lysine, and arginine). Tier 2 encompassed solubility and crystal form changes and two salts, arginine and lysine, exhibited acceptable properties. The last tier tested stability (4 week evaluation) and excipient compatibility. The arginine salt was found to be the best candidate with the lysine salt as the alternate choice. The entire process was completed in approximately 6 weeks.

The evaluation process can be as simple or as complex as needed for the development of a particular compound. Numerous factors need to be considered when choosing the best salt or cocrystal for development, and it needs to be tailored to the compound. In some cases, there may not be a salt or cocrystal that exhibits all the desired attributes. When this occurs, a risk assessment must be performed to determine what extra considerations may be needed to develop the compound. If

a compound is hygroscopic above 30% RH but the other attributes are acceptable, a decision needs to be made whether controlling manufacturing conditions below this level is feasible and cost effective. These types of decisions require a multidisciplinary approach among numerous groups.

### 3.5 Polymorph Screen

Once a salt or cocrystal is chosen, a more extensive polymorph screen should be performed to find the best crystal form for development [42]. Based on screens performed in our lab, it was found that 87% of the compounds investigated resulted in more than one form including amorphous material (Fig. 8a) [43]. Multiple polymorphs, defined as forms having the same chemical composition, were found in 51% of the cases, and hydrates were found in 37% of the screens. As shown in Fig. 8b, two to four forms were found for the largest number of compounds, but many cases exist where significantly more forms were obtained.

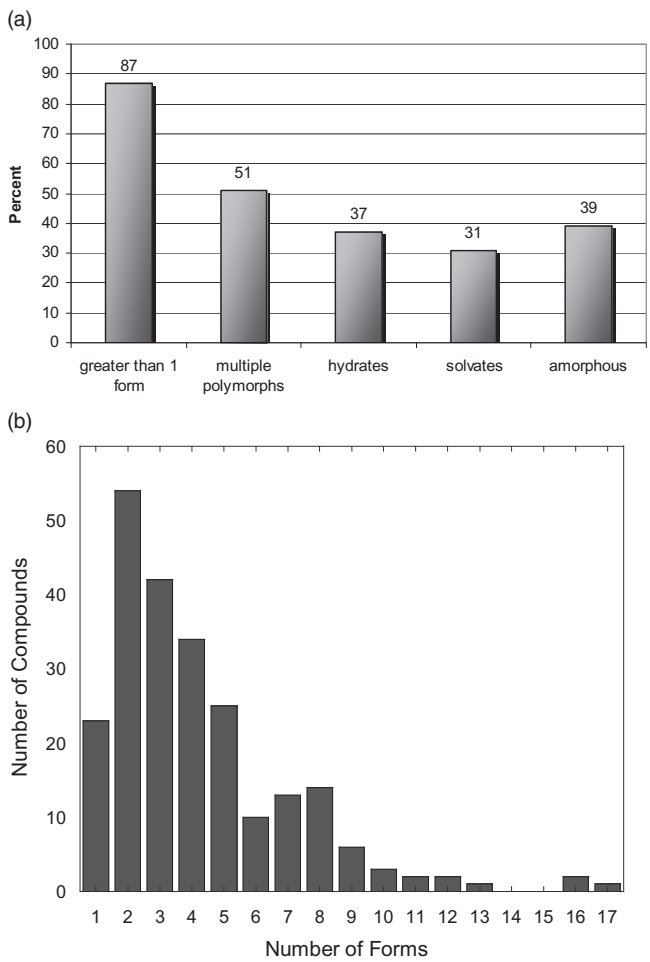


**FIGURE 8** Forms observed in polymorph screens.

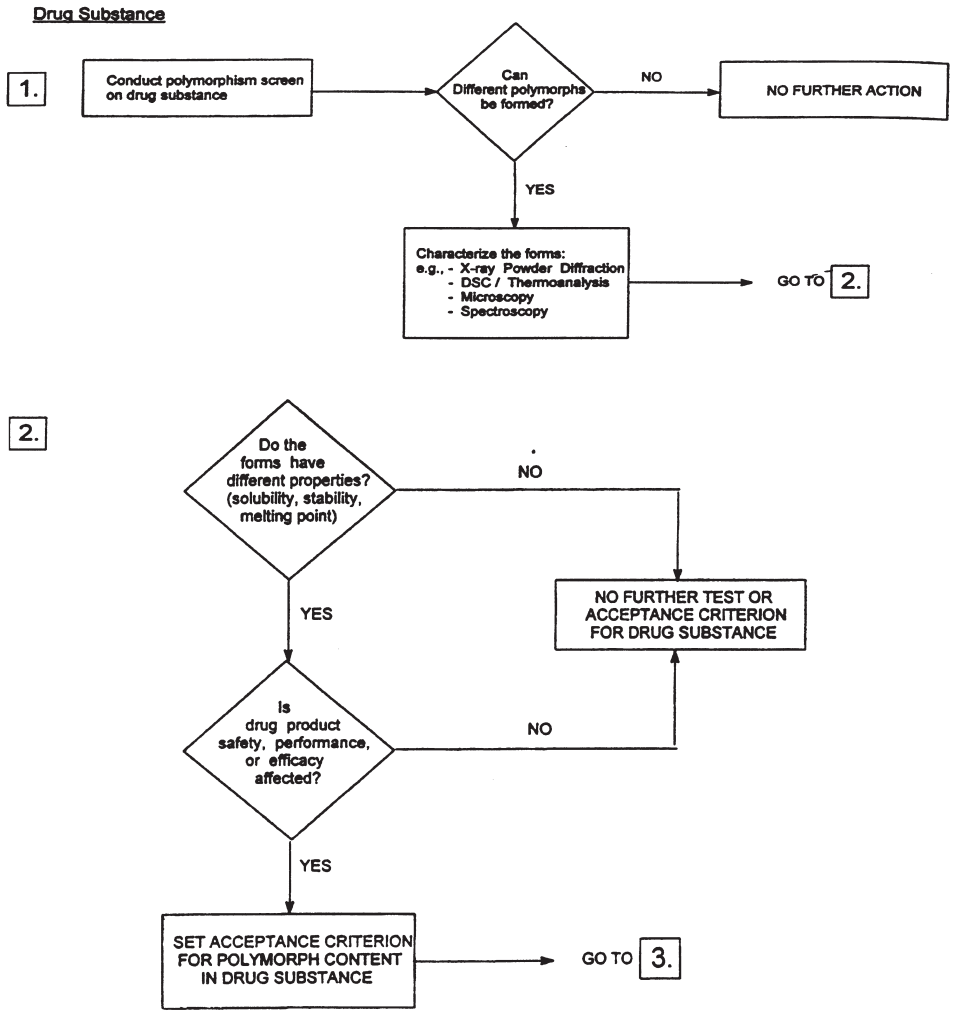


**FIGURE 9** Decision trees for polymorph screening and property determination [44, 45].

A polymorph screen is recommended by regulatory agencies. The U.S. Food and Drug Administration (FDA) has recommended guidelines for polymorph screens [44, 45] and these decision trees are presented in Fig. 9. The purpose of a polymorph screen is to find multiple forms of a compound and it can be considered as a search for seeds. The decision trees are available to help guide the development of an API that can exist as more than one form. Issues such as stability, solubility, and performance in the dosage form need to be addressed for different forms of an API.

Polymorph screens can and should be performed at various stages of the development process, depending on the type of information that is required [46]. One automated screen early in development is not always sufficient to find the relevant forms for development. Possible screens and when they could be performed are summarized in Table 3 and Fig. 10. The screen should be tailored to the type of

**TABLE 3 Possible Polymorph Screens During Development**

| Type of Screen | Goal |
|---|---|
| Early | • Propensity for polymorphism<br>• Indication of most stable forms<br>• Indication of stable hydrates |
| Full | • Most stable form (enantiotropy)<br>• Most stable hydrate<br>• Other forms |
| Focused | • Selected form studied under GMP manufacturing conditions |
| Comprehensive | • Increase certainty of knowing most stable form<br>• Find form giving improved drug product<br>• Patent protection of all forms<br>• Find noninfringing forms<br>• Reduce patent costs by finding new forms at one time |

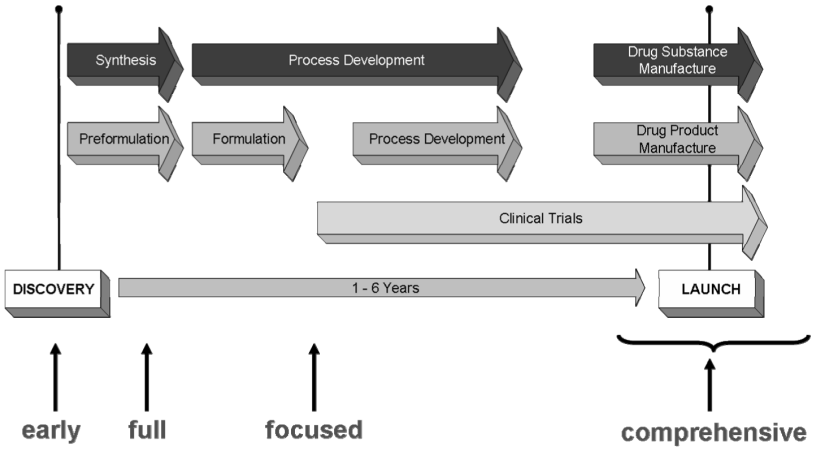


**FIGURE 10** Types of polymorph screens at various stages of development.

information required at that stage of development and, as the API progresses, additional information may be needed. The number and type of screens performed for an API will be determined based on the properties of the drug substance, the dosage form, and special issues that may come up during development.

Common steps in a polymorph screen are:

1. Characterize starting material.
2. Generate solid samples.
3. Collect XRPD data on solids produced.
4. Classify the XRPD patterns.
5. Scale up if needed.
6. Characterize each form.
7. Compare properties.
8. Determine relative thermodynamic stability.
9. Select form.

A wide range of crystallization techniques used in polymorph screens have been reported and reviewed [47, 48]. The majority of them are solvent based. A variety of factors are considered when choosing solvent systems for a screen, including solubility of the API, solvents to be used in manufacturing, a range of polarities and functionalities, and a range of water activities. Choices should not be limited to Class 3 solvents at this stage of the selection process. Crystallization experiments can include evaporations, slurries, cools, diffusion (vapor or liquid), solvent drop grinding, and antisolvent addition. The solubility of the API in organic solvents can help guide the type of crystallization experiments to perform. Material that is poorly soluble in most solvents would lend itself to slurry experiments or elevated temperature experiments with subsequent cooling. Material that is highly soluble may oil out at concentrations that are too high. Crystallizations covering both thermodynamic and kinetic conditions should be included since manufacturing processes commonly employ kinetic crystallization conditions (such as antisolvent addition). Experiments intended to influence the nucleation event should also be considered [49]. Nonsolvent based experiments, such as thermal treatment, sublimation, grinding, and compression [50], are also valuable and should be included whenever possible. Crystallization conditions used in a screen do not have to mimic manufacturing conditions. Seeds produced from small scale experiments can commonly be used to make larger amounts of desired forms.

The number and types of experiments should vary from compound to compound since crystallization conditions will change with the properties of the API. The choice of manual versus automated experiments also needs to be determined. A combination of solvent based experiments (automated and manual) along with nonsolvent based crystallizations will provide a more comprehensive screen and cover more of the crystallization space for the compound of interest. It is not necessary, and is sometimes counterproductive, to perform thousands of the same type of experiment, which will cover a limited portion of the crystallization space.

Analysis of the solids produced during the screen is the next step. A variety of techniques are used to characterize new materials, such as XRPD, thermal methods, and spectroscopic methods. A multidisciplinary approach is needed to understand the polymorphic system of an API. One technique will not provide all the necessary information on new crystal forms of the material and how they relate to one another. Numerous references are available that describe characterization of new forms [13, 32, 42, 46]. Analytical techniques in this section are discussed in terms of the information provided on new solids obtained during a polymorph screen.

XRPD is commonly used to determine if the samples are crystalline and it is a direct measurement of the crystal form. Once data are collected, the patterns are classified into different groups that may represent different forms of the material. At this stage, a different pattern may represent a pure material, preferred orientation, or a mixture of forms. Additional characterization information is needed on a representative sample to determine if a pure phase has been produced.

Once crystallinity has been established, determining the presence of solvents is important. Crystal forms can be solvated or unsolvated. In many cases, solvated forms are dismissed in a screen because they cannot readily be used as an API. However, solvates can play an important role in processing and crystallization development and can help explain the formation of one form over another from a certain solvent or process. Solvates would usually not be chosen for development,

but upon drying a new unsolvated form may be produced that has all the properties required for the targeted product. Hydrates can be considered a subset of solvates, with water being the solvent incorporated into the lattice. Various types of hydrates have been characterized and discussed [51]. Two common types are stoichiometric and variable hydrates. Stoichiometric hydrates contain water in molar ratios. Those that are stable over a wide range of relative humidity conditions are possible candidates for development. Others may be unstable and unsuitable for further consideration. Variable hydrates typically contain water in nonstoichiometric amounts and the amount of water readily changes with relative humidity conditions. These types of hydrates are difficult to characterize, control, and develop.

Melting point can be a factor during development and should be determined for the various forms. DSC, hot-stage microscopy, or a melting point apparatus can be used to measure this property. A higher melting solid, generally greater than 100 °C, is desirable. A low melting material is more difficult to dry, mill, or compress. DSC data can also provide information on high temperature crystal form changes that may impact development [42], as well as information on physical purity of the sample. Other data can be invaluable in determining if a sample is a pure form or a physical mixture of two or more forms. Solid-state NMR has been used to characterize many pharmaceutical materials and can commonly provide information on mixtures that is difficult to obtain by other methods [35, 36]. IR spectroscopy and Raman spectroscopy are also commonly used to differentiate crystalline forms [32].

Once the forms have been characterized, it is important to understand the relationship between the forms. A schematic, such as the one given in Fig. 11 [52], can be constructed to visualize changes in the forms during development. An early

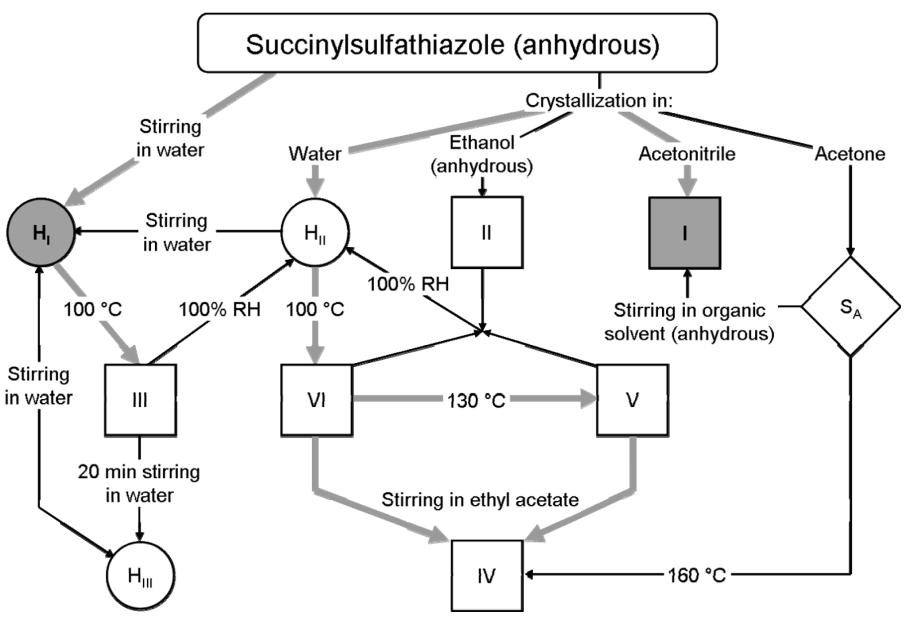


**FIGURE 11** Schematic of succinylsulfathiazole forms produced under various conditions [52].

understanding of possible form changes during processing, such as upon drying or exposure to water, will help to design acceptable processes that will minimize form changes [53, 54].

The thermodynamically most stable form should be identified if more than one acceptable unsolvated form is found in the screen [46, 55, 56]. The most stable crystalline form is defined as that with the lowest free energy and it will also exhibit the lowest solubility. In a monotropic system, the relative free energy for the two forms remains the same along the entire temperature range, as shown in the energy–temperature diagram in Fig. 12a. The heat of fusion curves show that the more stable form has the higher melting point and the greater heat of fusion. In an enantiotropic

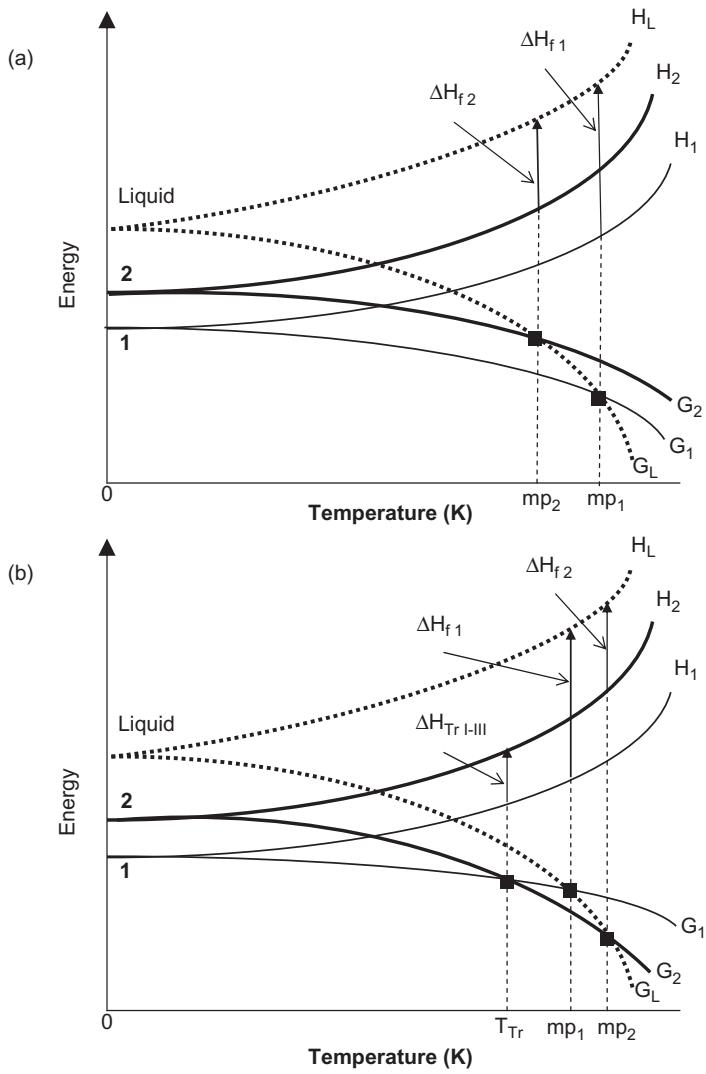


**FIGURE 12** Energy–temperature diagrams for (a) monotropic and (b) enantiotropic systems.

**TABLE 4 Burger and Ramburger Rules to Determine the Thermodynamically Stable Form** [57]

| Rule | Description |
|---|---|
| Heat of transition rule | If an endothermic transition is observed at some temperature below the melting point, it may be assumed that there are two forms related enantiotropically.<br>If an exothermic transition is observed below the melting point, it may be assumed that there are two forms related monotropically or the transition temperature is higher. |
| Heat of fusion rule | If the higher melting form has the lower heat of fusion, the two forms are usually enantiotropic; otherwise they are monotropic. |
| IR rule | If the first absorption band in the infrared spectrum of a hydrogen-bonded molecular crystal is higher for one modification than for the other, that form may be assumed to have the larger entropy. |
| Density rule | If one modification of a molecular crystal has a lower density than the other, it may be assumed to be less stable at absolute zero. |

system, the relative stabilities change at a temperature defined as the transition temperature, as shown in Fig. 12b. In energy–temperature diagrams, the free energy curves can cross, but the enthalpy curves do not. Based on this, the more stable form below the transition temperature in an enantiotropic system will exhibit the lower melting point and the higher heat of fusion. There are four rules proposed by Burger and Ramburger to determine the type of relationship between forms [57] and these are summarized in Table 4. Other methods to determine the stability include plotting solubility data at different temperatures in a van't Hoff plot [58] as well as eutectic mixtures [59]. This information is important for crystallization processes since crystallization above the transition temperature will favor one form and below the transition temperature will favor the other. It is also critical for long term stability since the metastable form may transform to the more stable form over time. A metastable form with sufficient stability to provide an acceptable shelf life is another option to consider.

At this stage, other factors need to be considered in the form selection process. Information on solubility, intrinsic solubility, modified absorption potential [60], and chemical and physical stability will help in selection of the best candidate for development.

### 3.6 Crystallization and Quantitation

Once a crystal form is found, crystallization method development will be needed to determine the best conditions to produce material with optimal characteristics, with crystal form being only one consideration. Crystallization conditions to produce the salt or cocrystal directly are the best option. However, a recrystallization step may be needed to produce the desired crystal form with acceptable purity and yield. Numerous factors, such as crystallization equipment, solvent, temperature, cooling rate, antisolvent addition, and other parameters, will need to be investigated [61].

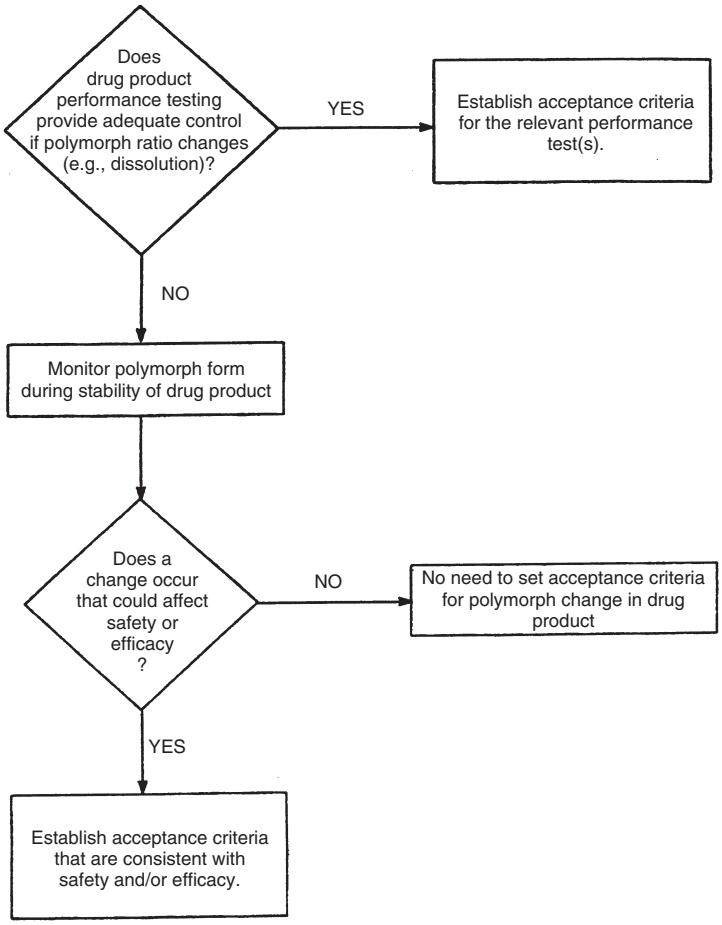


**FIGURE 13** Decision tree for polymorphic form in drug product [44, 45].

Seeding is also a good option for many systems. Knowing the transition temperature for an enantiotropic pair will be necessary to determine the best crystallization conditions.

When multiple forms are known, the process needs to be controlled to make the desired form of the API consistently. An assay, qualitative or quantitative, may be needed for confirmation, as outlined in Fig. 9. Quantitation of forms may also need to be addressed for the drug product as shown in the flowchart in Fig. 13. Common analytical techniques used to quantitate the forms in solid samples include XRPD, DSC, IR and Raman spectroscopy, and NMR. Moisture balance data can also be used to quantitate amorphous materials in crystalline solids. Specificity needs to be determined for the technique being used for method development. Guidelines for quantitative method development have been reported [62] and the parameters to be investigated include specificity, linearity, range, accuracy, precision, detection limit, and quantitation limit. Different types of quantitative methods can be developed, as summarized in Table 5. Different assays can be used at various stages of development and to obtain suitable data for specific situations. An assay may be

**TABLE 5 Summary of Solid-State Quantitative Assays**

| Quantitative Method | Scope of Assay | Result |
|---|---|---|
| Limit test | A limit of detection for a technique is determined (such as 2%) | Result is specified as "less than 2% Form B present" |
| Specification assay | An assay limit is determined for a technique (such as 95%) | Result is specified as "greater than 95% Form A present" |
| Full quantitative method | A minimum quantifiable limit (MQL) (such as 2%) and linear range (such as 2–25%) are determined | "X% of Form B" if within the quantitation range of 2–25%, or "less than 2% Form B" if below MQL, or "greater than 25% Form B" if above linear range |

needed to show that a new process is producing the desired form and may eventually be sunsetted after control is established over a period of time.

## 4 CONCLUSION

Significant information is needed to select a form for development. Making salts and cocrystals of the API for evaluation is a small part of the form selection process. Characterization of physical properties and understanding the interaction of various forms in a solid system are critical for the success of a formulated product and can significantly reduce time to market. A qualitative or quantitative assay may be needed for the API and/or drug product if multiple forms are found. The information obtained during form selection can be used for regulatory and intellectual property documentation and can help extend the life of an important franchise.

## REFERENCES

1. Byrn SR, Pfeiffer RR, Stowell JG. *Solid-State Chemistry of Drugs*, 2nd ed. West Lafayette, IN: SSCI Inc; 1999.
2. Clas SD. The importance of characterizing the crystal form of the drug substance during drug development. *Curr Opin Drug Discov Dev* 2003;6:550–560.
3. Curatolo W. Physical–chemical properties of oral drug candidates in the discovery and exploratory development settings. *Pharm Sci Technol Today* 1998;1:387–393.
4. Gardner CR, Walsh CT, Almarsson O. Drugs as materials: valuing physical form in drug discovery. *Nat Rev Drug Discov* 2004;3:926–934.
5. Venkatesh S, Lipper RA. Role of the development scientist in compound lead selection and optimization. *J Pharm Sci* 2000;89:145–154.
6. Fiese EF. General pharmaceutics—the new physical pharmacy. *J Pharm Sci* 2003;92(7): 1331–1342.
7. Huang LF, Tong WQ. Impact of solid state properties on developability assessment of drug candidates. *Adv Drug Deliv Rev* 2004;56:321–334. Kibbey CE, Poole SK, Robinson

B, Jackson JD, Durham D. An integrated process for measuring the physicochemical properties of drug candidates in a preclinical discovery environment. *J Pharm Sci* 2001;90:1164–1175. Wenlock MC, Austin RP, Barton P, Davis AM, Leeson PD. A comparison of physiochemical property profiles of development and marketed oral drugs. *J Med Chem* 2003;46:1250–1256.

8. Stahl PH, Wermuth CG, Eds. *Handbook of Pharmaceutical Salts*. Hoboken, NJ:Wiley-VCH; 2002.
9. Balbach S, Korn C.. Pharmaceutical evaluation of early development candidates "the 100 mg-approach." *Int J Pharm* 2004;275:1–12.
10. Bastin RJ, Bowker MJ, Slater BJ. Salt selection and optimisation procedures for pharmaceutical new chemical entities. *Org Proc Res Dev* 2000;4:427–435.
11. Morris KR, Fakes MG, Thakur AB, Newman AW, Singh AK, Venit JJ, Spagnuolo CJ, Serajuddin ATM. An integrated approach to the selection of optimal salt form for a new drug candidate. *Int J Pharm* 1994;105:209–217.
12. Giron D. Characterisation of salts of drug substances. *J Thermal Anal Calorimetry* 2003;73:441–457.
13. Newman AW, Stahly GP. Form selection of pharmaceutical compounds. In: Ohannessian L, Ed. *Handbook of Pharmaceutical Analysis*, *Volume 117*. New York: Marcel Dekker; 2001, pp 1–57.
14. Braga D, Brammer L, Champness NR. New trends in crystal engineering. *Crystengcomm* 2005;7:1–19. Erk P, Hengelsberg H, Haddow MF, van Gelder R. The innovative momentum of crystal engineering. *Crystengcomm* 2004;6:474–483. Hulliger J, Roth SW, Quintel A, Bebie H. Polarity of organic supramolecular materials: a tunable crystal property. *J Solid State Chem* 2000;152:49–56. Iimura N, Ohashi Y, Hirata H. Complex formation of perfumes with cationic surfactants and the enhanced thermal stability. *Bull Chem Soc Jpn* 2000;73:1097–1103. Wong MS, Pan F, Bosshard C, Guenter P. Supramolecular synthesis of molecular co-crystals with highly optimized chromophoric orientation for second-order nonlinear optics. *Polym Mater Sci Eng* 1996;75:132–133. Wong MS, Pan F, Gramlich V, Bosshard C, Gunter P. Self-assembly of an acentric co-crystal of a highly hyperpolarizable merocyanine dye with optimized alignment for nonlinear optics. *Adv Materials* 1997;9:554–557.
15. 21 CFR 314.108(a) and Draft Guidance *Applications Covered by Section 505(b)(2)*, http://www.fda.gov/cder/guidance/2853dft.htm.
16. Gould PL. Salt selection of basic drugs. *Int J Pharm* 1986;33:201–217.
17. Miyazaki S, Oshiba M, Nadai T. Precaution on use of hydrochloride salts in pharmaceutical formulation. *J Pharm Sci* 1981;70(6):594–596.
18. Bruno IJ, Cole JC, Lommerse JPM, Rowland RS, Taylor R, Verdonk ML. IsoStar: a library of information about nonbonded interactions. *J Comput Aided Mol Des* 1997; 11(6):525–537. Chisholm J, Pidcock E, Van De Streek J, Infantes L, Motherwell S, Allen FH. Knowledge-based approaches to crystal design. *Crystengcomm* 2006;8:11–28. Gillon AL, Feeder N, Davey RJ, Storey R. Hydration in molecular crystals—a Cambridge Structural Database analysis. *Crystal Growth Des* 2003;3:663–673. Steiner T. The hydrogen bond in the solid state. *Angew Chem Int Ed* 2002;41:48–76.
19. Aakeroy CB, Salmon DJ. Building co-crystals with molecular sense and supramolecular sensibility. *Crystengcomm* 2005;7:439–448.
20. Anthony A, Desiraju GR, Jetti RKR, Kuduva SS, Madhavi NNL, Nangia A, Thaimattam R, Thalladi VR. Crystal engineering: some further strategies. *Mater Res Bull* 1998;1–18.
21. Desiraju GR. Supramolecular synthons in crystal engineering—a new organic-synthesis. *Angew Chem Int Ed Engl* 1995;34:2311–2327.

22. Childs SL, Chyall LJ, Dunlap JT, Smolenskaya VN, Stahly BC, Stahly GP. Crystal engineering approach to forming cocrystals of amine hydrochlorides with organic acids. Molecular complexes of fluoxetine hydrochloride with benzoic, succinic, and fumaric acids. *J Am Chem Soc* 2004;126:13335–13342. Fleischman SG, Kuduva SS, McMahon JA, Moulton B, Walsh RDB, Rodriguez-Hornedo N, Zaworotko MJ. Crystal engineering of the composition of pharmaceutical phases: multiple-component crystalline solids involving carbamazepine. *Crystal Growth Des* 2003;3:909–919.McMahon JA, Bis JA, Vishweshwar P, Shattock TR, McLaughlin OL, Oswald MJ, Allan DR, McGregor PA, Motherwell WDS, Parsons S, Pulham CR. The formation of paracetamol (acetaminophen) adducts with hydrogen-bond acceptors. *Acta Crystallogr B Struct Sci* 2002;58:1057–1066.Remenar JF, Morissette SL, Peterson ML, Moulton B, MacPhee JM, Guzman HR, Almarsson O. Crystal engineering of novel cocrystals of a triazole drug with 1,4-dicarboxylic acids. *J Am Chem Soc* 2003;125:8456–8457.Rodriguez-Spong B, Price CP, Jayasankar A, Matzger AJ, Rodriguez-Hornedo N. General principles of pharmaceutical solid polymorphism: a supramolecular perspective. *Adv Drug Deliv Rev* 2004;56:241–274.Trask AV, Motherwell WDS, Jones W. Pharmaceutical cocrystallization: engineering a remedy for caffeine hydration. *Crystal Growth Des* 2005;5:1013–1021.Walsh RDB, Bradner MW, Fleischman S, Morales LA, Moulton B, Rodriguez-Hornedo N, Zaworotko MJ. Crystal engineering of the composition of pharmaceutical phases. *Chem Commun* 2003;186–187.
23. Aakeroy CB, Desper J, Helfrich BA. Heteromeric intermolecular interactions as synthetic tools for the formation of binary co-crystals. *Crystengcomm* 2004;6:19–24.
24. Etter MC. Hydrogen-bonds as design elements in organic-chemistry. *J Phys Chem* 1991;95:4601–4610.
25. Amos JG, Indelicato JM, Pasini CE, Reutzel SM. US patent 6,001,996. 1999.
26. Nehm SJ, Rodriguez-Spong B, Rodriguez-Hornedo N. Phase solubility diagrams of cocrystals are explained by solubility product and solution complexation. *Crystal Growth Des* 2006;6:592–600.
27. Monissette SL, Almarsson O, Peterson ML, Remenar JF, Read J, Lemmo AV, Ellis S, Cima MJ, Gardner CR. High-throughput crystallization: polymorphs, salts, co-crystals and solvates of pharmaceutical solids. *Adv Drug Deliv Rev* 2004;56:275–300. Remenar JF, MacPhee JM, Larson BK, Tyagi VA, Ho JH, McIlroy DA, Hickey MB, Shaw PB, Almarsson O. Salt selection and simultaneous polymorphism assessment via high-throughput crystallization: the case of sertraline. *Org Proc Res Dev* 2003;7:990–996. Ware EC, Lu DR. An automated approach to salt selection for new unique trazodone salts. *Pharm Res* 2004;21:177–184.
28. Stephenson GA. Applications of X-ray powder diffraction in the pharmaceutical industry. *Rigaku J* 2005;22(1):2–15.
29. Kondritzer AA, Ellin RI, Edberg LJ. Investigation of methyl pyridinium-2-aldoxime salts. *J Pharm Sci* 1961;50(2):109.
30. Koehler HM, Hefferren JJ. Mineral salts of lidocaine. *J Pharm Sci* 1964;53(9): 1126.
31. Childs SL, Chyall LJ, Dunlap JT, Coates DA, Stahly BC, Stahly GP. A metastable polymorph of metformin hydrochloride: isolation and characterization using capillary crystallization and thermal microscopy techniques. *Crystal Growth Des* 2004;4:441–449. Rai RN, Varma KBR. Phase diagram and dielectric studies of binary organic materials. *Mater Lett* 2000;44:284–293. Zalac S, Khan MZI, Gabelica V, Tudja M, Mestrovic E, Romih M. Paracetamol-propyl phenazone interaction and formulation difficulties associated with eutectic formation in combination solid dosage forms. *Chem Pharm Bull* 1999;47:302–307. Kuhnert-Brandstatter M, Geiler M, Wurian I. *Sci Pharm* 1983;51:34–41. Kofler A, Kolsek J. *Mikrochimi Acta* 1969;2:408–435.

32. Pedireddi VR, Jones W, Chorlton AP, Docherty R. Creation of crystalline supramolecular arrays: a comparison of co-crystal formation from solution and by solid state grinding. *Chem Commun* 1996;987–988. Shan N, Toda F, Jones W. Mechanochemistry and co-crystal formation: effect of solvent on reaction kinetics. *Chem Commun* 2002;2372–2373. Trask AV, Haynes DA, Motherwell WDS, Jones W. Screening for crystalline salts via mechanochemistry. *Chem Commun* 2006;51–53. Trask AV, van de Streek J, Motherwell WDS, Jones W. Achieving polymorphic and stoichiometric diversity in cocrystal formation: importance of solid-state grinding, powder X-ray structure determination, and seeding. *Crystal Growth Des* 2005;5:2233–2241.
33. Brittain HG, Ed. *Physical Characterization of Pharmaceutical Solids*. New York: Marcel Dekkar; 1995.
34. Brittain HG, Bogdanowich SJ, Bugay DE, DeVincentis J, Lewen G, Newman AW. Physical characterization of pharmaceutical solids. *Pharm Res* 1991;8(8):963–973.
35. Mehrens SM, Kale UJ, Qu XG. Statistical analysis of differences in the Raman spectra of polymorphs. *J Pharm Sci* 2005;94:1354–1367.
36. Tishmack PA, Bugay DE, Byrn SR. Solid-state nuclear magnetic resonance spectroscopy—pharmaceutical applications. *J Pharm Sci* 2003;92:441–474.
37. Stephenson GA, Forbes RA, Reutzel-Edens SM. Characterization of the solid state: quantitative issues. *Adv Drug Deliv Rev* 2001;48:67–90.
38. Padden BE, Zell MT, Dong ZD, Schroeder SA, Grant DJW, Munson EJ. Comparison of solid-state $^{13}C$ NMR spectroscopy and powder X-ray diffraction for analyzing mixtures of polymorphs of neotame. *Anal Chem* 1999;71(16):3325–3331.
39. Reutzel-Edens SM, Bush JK. Solid-state NMR spectroscopy of small molecules: from NMR crystallography to the characterization of solid oral dosage forms. *Am Pharm Rev* 2002;5/2:112–115.
40. Stahly GP, Bates S, Andres MC, Cowans BA. Discovery of a new polymorph of dehydroepiandrosterone (prasterone) and solution of its crystal structure from X-ray powder diffraction data. *Crystal Growth Des* 2006;6:925–932.
41. Reutzel-Edens SM, Newman AW. The physical characterization of hygroscopicity in pharmaceutical solids. In Hilfiker R, Ed. *Polymorphism*. Weinheim:Wiley-VCH; 2006, pp 235–258.
42. Giron D, Mutz M, Garnier S. Solid-state of pharmaceutical compounds—impact of the ICH Q6 guideline on industrial development. *J Thermal Anal Calorimetry* 2004;77:709–747. Singhal D, Curatolo W.. Drug polymorphism and dosage form design: a practical perspective. *Adv Drug Deliv Rev* 2004;56:335–347. Raw AS, Furness MS, Gill DS, Adams RC, Holcombe FO, Yu LX.. Regulatory considerations of pharmaceutical solid polymorphism in abbreviated new drug applications (ANDAs). *Adv Drug Deliv Rev* 2004;56:397–414. Yu LX, Furness MS, Raw AS, Outlaw KPW, Nashed NE, Ramos E, Miller SPF, Adams RC, Fang F, Patel RM, Holcombe FO, Chiu YY, Hussain AS.. Scientific considerations of pharmaceutical solid polymorphism in abbreviated new drug applications. *Pharm Res* 2003;20:531–536.
43. Stahly GP. Polymorphism in crystals: fundamentals, predications and industrial practice. Presented at the American Chemical Society ProSpectives Meeting, Tampa, FL, 23–26 Feb 2003.
44. Q6A Specifications, Test procedures and acceptance criteria for new drug substances and new drug products: chemical substances. *Fed Reg* 1997;62:62889–62910.
45. Byrn SA, Pfeiffer R, Ganey M, Hoiberg C, Poochikian G. *Pharm Res* 1995;12:945–954.
46. Blagden N, Davey RJ. Polymorph selection: Challenges for the future? *Crystal Growth Des.* 2003;3:873–885.

47. Guillory JK. Generation of polymoprhs, hydrates, solvates, and amorphous solids. In Brittain HG, Ed. *Polymorphism in Pharmaceutical Solids*. New York: Marcel Dekker; 1999, Ch 5, pp 183–226.
48. Bernstein J. *Polymorphism in Molecular Crystals*. London:Oxford Science Publications; 2002.
49. Gracin S, Uusi-Penttila M, Rasmuson AC. Influence of ultrasound on the nucleation of polymorphs of *p*-aminobenzoic acid. *Crystal Growth Des* 2005;5:1787–1794. Davey RJ, Allen K, Blagden N, Cross WI, Lieberman HF, Quayle MJ, Righini S, Seton L, Tiddy GJT. Crystal engineering—nucleation, the key step. *Crystengcomm* 2002;257–264.
50. Fabbiani FPA, Allan DR, David WIF, Moggach SA, Parsons S, Pulham CR. High-pressure recrystallisation—a route to new polymorphs and solvates. *Crystengcomm* 2004;6:505–512.
51. Morris KM, Rodriquez-Hornedo N. Hydrates. In Swarbrick J, Boylan JC, Eds. *Encyclopedia of Pharmaceutical Technology*, *Volume 7*. New York: Marcel Dekker; 1993, pp 393–440.
52. Burger A, Grießer UJ. The polymorphic drug substances of the European Pharmacopoeia. Part 7: Physical stability, hygroscopicity and solubility of succinylsulfathiazole crystal forms. *Eur J Pharm Biopharm* 1991;37:118–124.
53. Herbstein FH. Diversity amidst similarity: a multidisciplinary approach to phase relationships, solvates, and polymorphs. *Crystal Growth Des* 2004;4:1419–1429.
54. Morris KR, Griesser UJ, Eckhardt CJ, Stowell JG. Theoretical approaches to physical transformations of active pharmaceutical ingredients during manufacturing processes. *Adv Drug Deliv Rev* 2001;91–114.
55. Miller JM, Collman BM, Greene LR, Grant DJW, Blackburn AC. Identifying the stable polymorph early in the drug discovery–development process. *Pharm Dev Technol* 2005;10:291–297.
56. Grunenberg A, Henck J-O, Siesler HW. Theoretical derivation and practical application of energy/temperature diagrams as an instrument in preformulation studies of polymorphic drug substances. *Int J Pharm* 1996;129:147–158.
57. Burger A, Ramburger R. On the polymorphism of pharmaceuticals and other molecular crystals. I. *Mikrochim Acta (Wien) II* 1979;259–272. Burger A, Ramburger R. On the polymorphism of pharmaceuticals and other molecular crystals. *II. Mikrochim Acta (Wien) II* 1979;273–316.
58. Behme RJ, Brooke D. Heat of fusion measurement of a low melting polymorph of carbamazepine that undergoes multiple-phase changes during differential scanning calorimetry analysis. *J Pharm Sci* 1991;80(10):986–990. Behme RJ, Brooke D, Farney RF, Kensler TT. Characterization of polymorphism of gepirone hydrochloride. *J Pharm Sci* 1985;74(10):1041–1046.
59. Yu L, Stephenson GA, Mitchell CA, Bunnell CA, Snorek SV, Bowyer JJ, Borchardt TB, Stowell JG, Byrn SR. Thermochemistry and conformational polymorphism of a hexamorphic crystal system. *J Am Chem Soc* 2000;122(4):585–591.
60. Sanghvi T, Ni N, Yalkowsky SH. A simple modified absorption potential. *Pharm Res* 2001;18(12):1794–1796. Anderson NG. *Practical Process Research and Development*. San Diego, CA: Academic Press; 2000.
61. International Conference on Harmonisation; Q2B validation of analytical procedures: methodology. *Fed Reg* 1997;62:27464.
62. Newman AW, Byrn SR. Solid-state analysis of the active pharmaceutical ingredient in drug products. *Drug Discov Today* 2003;8(9):898–905.

# DISSOLUTION

A.K. Tiwary, Bharti Sapra, and Subheet Jain
*Punjabi University, Patiala, Punjab, India*

**Contents**

1 Introduction
2 Factors Influencing Drug Dissolution from Dosage Forms
  2.1 Formulation Variables
  2.2 Process Variables
3 Dissolution Test Designing
  3.1 Media Selection
  3.2 Apparatus Selection
  3.3 Dissolution Apparatus Operating Parameters
4 Analysis of Dissolution Data
5 *In Vitro–In Vivo* Correlation
  5.1 Physiological Concerns
  5.2 Correlation
  5.3 Biowaivers for Changes in the Manufacture of a Drug Product
  References

## 1 INTRODUCTION

Dissolution of the active ingredients is of paramount importance for the successful performance of a dosage form. Dissolution testing is a convenient tool for ensuring intra- and interbatch performance compliance of dosage forms. Furthermore, the release profiles obtained from *in vitro* dissolution tests can be used for predicting *in vitro–in vivo* correlation models. Thus, carefully conducted dissolution tests are useful for ensuring dosage form functionality.

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

The physicochemical characteristics of a drug and its site of action/absorption together dictate the type of dosage form to be developed. The plethora of additives and technologies available in the hands of the product development scientist make it possible to design a dosage form possessing the predetermined drug release attributes. Great strides in the excipient and technology arena, however, make it imperative to critically evaluate their influence on *in vitro* drug dissolution during the preformulation phase. This involves proper selection of dissolution media, dissolution apparatus, operation variables, and interpretation of generated data. Although a preliminary estimate of the dosage form's *in vivo* performance can be gauged by applying *in vitro–in vivo* correlation models, it is essential to test the dosage form in animals to demonstrate its efficacy and safety. Therefore, the functionality of dosage form is first demonstrated in appropriate animal model(s) during the preclinical phase before testing it in human beings. Figure 1 gives a schematic representation of the steps to be followed during the course of dosage form development with respect to dissolution testing of a dosage form.

## 2 FACTORS INFLUENCING DRUG DISSOLUTION FROM DOSAGE FORMS

Dissolution of drug from a dosage form is unequivocally the most important indicator of its anticipated *in vivo* performance. Both nature of formulation ingredients and methods of incorporating them influence drug release from dosage forms. Therefore, formulation and process variables demand careful evaluation during the initial product development phase in order to achieve the predetermined drug dissolution profile.

### 2.1 Formulation Variables

A variety of excipients are available for use in solid dosage forms. Functional differences in excipients are known to exist depending on the material and usage and it is difficult to generalize findings for all the drugs. Therefore, the potential of these excipients to modify drug release from dosage forms should be critically evaluated during formulation. Although a detailed discussion on all excipients would not be possible, a few categories needing special attention are discussed next.

***Diluents*** Diluents are used in solid dosage forms as main bulk forming agents. Being present in large quantities, they can exert profound influence on the stability of the active ingredient, ease of manufacturing, drug release, and bioavailability profile. Table 1 lists important diluents used for direct compression and wet granulation. It is evident from this table that an excipient influences compression kinetics, tablet properties, and eventually functionality of the dosage form.

Furthermore, the effect of precompression, which is often used for reworking tablet/capsule granules, on drug release should be evaluated especially while attempting dry granulation. The quality of tablets containing a combination of microcrystalline cellulose and lactose was found to increase after reworking when the proportion of spray dried lactose was increased. However, the tablet weight uniformity decreased [5]. Generally, recompression reduces tablet strength and this

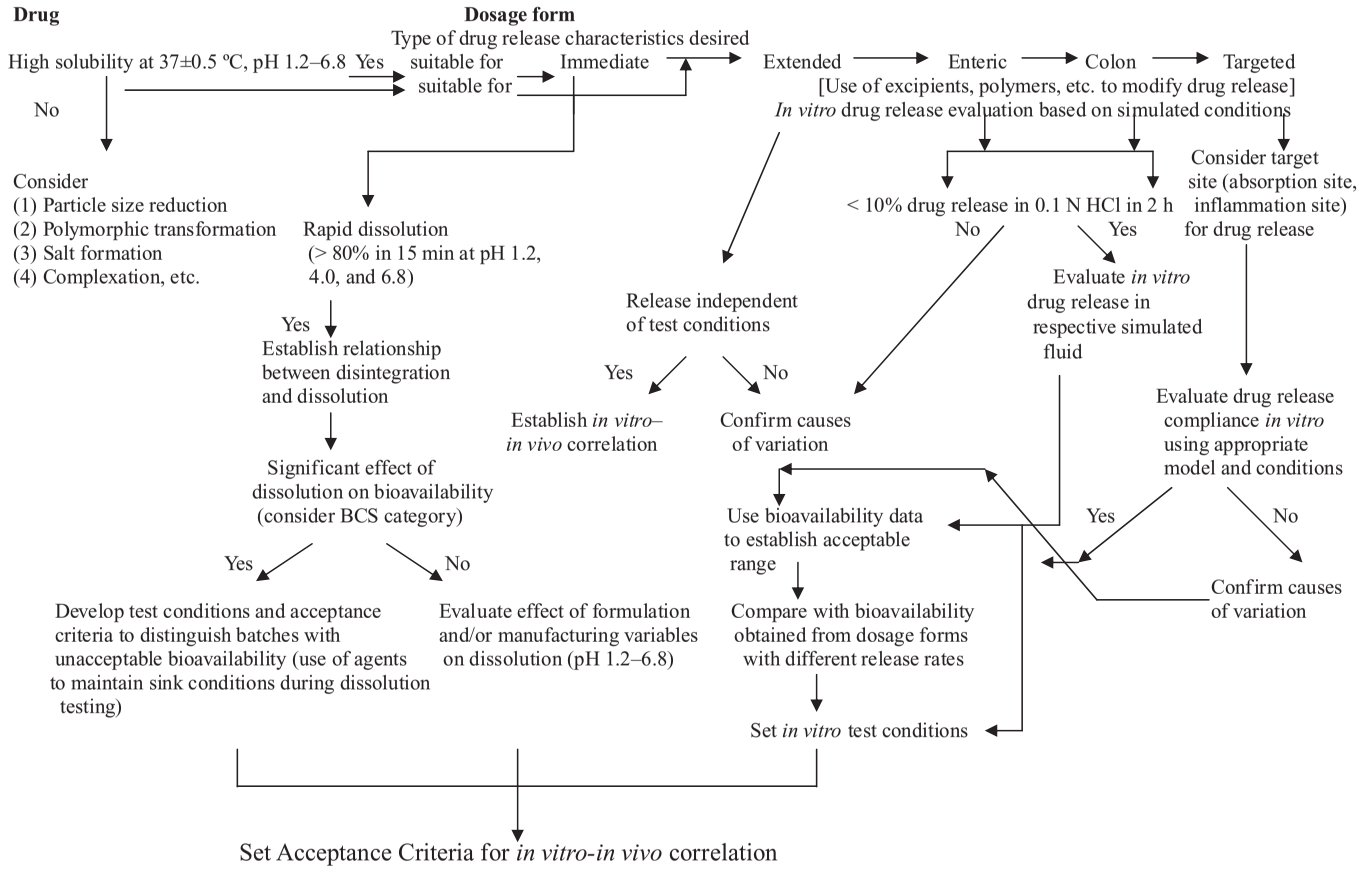


**FIGURE 1** Schematic representation of steps to be followed during dosage form development.

**TABLE 1 A Few Important Diluents and Their Properties**

| Excipient (Trade Name) | Properties | | Suitable for | Effect on Drug Dissolution | References |
|---|---|---|---|---|---|
| | Advantages | Disadvantages | | | |
| Microcrystalline cellulose (Avicel, PH-101) | Good compressibility, forms hard tablets at low pressure, fast tablet disintegration, available in different particle sizes for use in different types or tablets | Fair flowability, costly | • Direct compression<br>• Wet granulation<br>• With nonaqueous binding system<br>• Higher carrying capacity<br>• Hard tablet at low applied pressure<br>• Controlled release when combined with lactose, starch, etc. | Capillarity explains the penetration of water into tablet, thereby destroying cohesive bonds between particles | 1–3 |
| Powdered cellulose | Little inherent lubrication property, inexpensive | Poor flowability, poor compressibility, poor binding, high dilution potential | • Wet granulation | Tablets often disintegrate readily; but chances of physical entrapment are also there, which may lead to delayed dissolution | 1, 2 |
| Dicalcium phosphate | Fairly good compressibility, excellent flow property | Soluble in acidic pH of stomach | • Direct compression<br>• Wet granulation with nonaqueous binding system | Negative effect on dissolution by using increased proportions with microcrystalline cellulose; because of hydrophobic nature, may limit the dispersion of drug particles, hence may lead to slow dissolution | 1, 2 |

| | | | | | |
|---|---|---|---|---|---|
| Spray dried lactose | Good compressibility, excellent flow property, low hygroscopicity, low reactivity with active ingredient; excellent physical, chemical, and microbiological stability | Browning due to the presence of 5-hydroxy methyl 2-furaldehyde when combined with amines, moisture, lactates, phosphates; has relatively poor dilution potential | • Direct compression<br>• Wet granulation with nonaqueous binding system | Aids in dissolution | 1, 2, 4 |
| Starch directly compressible (Sta-Rx-1500) | Acts as local desiccant to help stabilize moisture sensitive drugs, free flowing, self lubricating | Tablets with high content are soft and may be difficult to dry, susceptible to softening when combined with 0.5% magnesium stearate | • Direct compression | Starch granules swell in dissolution media forming a viscous matrix that retards dissolution medium penetration and hence decrease the dissolution | 1, 2 |
| Compressible sugar (invert sugar, malt dextrins) | Highly compressible, inexpensive nonhygroscopic (invert sugar is hygroscopic at elevated temperature and high humidity) | Readily soluble, calorigenic, not acceptable for diabetes | • Tablet at high moisture content requires high ejection force<br>• Low lubricant requirement | | 1, 2 |

**TABLE 1** ***Continued***

| Excipient (Trade Name) | Properties | | Suitable for | Effect on Drug Dissolution | References |
|---|---|---|---|---|---|
| | Advantages | Disadvantages | | | |
| Mannitol | Nonhygroscopic, needs combination with other directly compressible excipients for good compression, relatively inert with high heat stability | Poor flowability and requires high lubricant addition, expensive | • Mouth dissolving<br>• Buccal tablet<br>• Chewable tablet<br>• Generally wet granulated using auxillary binder to produce compressible granules | Not relevant | 1, 2 |
| Ethyl cellulose (TOEC-Aqualon) | High ethoxyl content and low viscosity, excellent flowability, drug release triggered by circumferential failure of the coat, greater lag time (3–16 h) by increasing coating compression force (5-L5KN) than other grades | Relatively expensive | • Compression coating | May have a serious retadant effect on disintegration and drug dissolution release | 1, 2 |
| Pregelatinized starch (PCS PC-10) | Free flowing, good binding ability, good disintegrating capacity, high dilution potential, some lubricant action, good oil absorption capacity | Requires high pressure to produce hard table; binding gets reduced by lubricants; magnesium stearate softens the tablet | • Direct compression | Effect is concentration dependent, generally improves dissolution | 1, 2 |

reduction becomes more significant when initial compression is carried out at higher pressure. Hence, these tablets can be expected to exhibit faster drug release.

*Directly Compressible Vehicles* Direct compression is preferred over wet granulation as it requires fewer unit operations and less space, saves time, and labor costs, and hence is more economical. Tablets prepared by direct compression exhibit less variation in dissolution profile after storage [6]. In addition, as directly compressed tablets disintegrate, discrete drug particles are released—unlike granules from wet granulation—which ensures faster as well as complete drug dissolution. It is well known that if fragmentation predominates during compression, an increase in specific surface area leads to enhanced dissolution. However, if consolidation is the predominating phenomenon, an increase in compression pressure decreases dissolution. Therefore, it is essential to evaluate the compression behavior of compressible vehicles before using them as diluents in solid dosage forms.

LACTOSE Lactose exhibits polymorphism depending on crystallization conditions. Crystalline lactose undergoes fragmentation while amorphous lactose undergoes plastic deformation during compression. Tablets prepared from amorphous lactose show greater hardness with increasing water content [7].

α-Lactose monohydrate (100#) is generally used in direct compression due to its good flow property. It undergoes fragmentation during compression and exhibits poor binding [8] and the tablet strength decreases with increase in particle size [9]. Tablets prepared by combining microcrystalline cellulose and α-lactose exhibit greater hardness with an increase in proportion of the former [9]. α-Lactose monohydrate exhibits good flowability, compressibility, and hardness when mixed with polyethylene glycol 6000 [10]. A multifunctional coprocessed, directly compressible diluent containing lactose, polyvinyl pyrrolidone, and croscarmellose sodium was found to exhibit better flow properties, compressibility, and disintegration than lactose monohydrate [11]. The binding capacity of α-lactose monohydrate increases considerably after dehydration, wherein it changes from single crystals to aggregates of anhydrous α-lactose particles [12]. Tablets containing anhydrous α-lactose show slow disintegration and dissolve during disintegration [5]. Anhydrous β-lactose monohydrate is 10-fold more soluble than α-lactose monohydrate. The slowing effect is suggested to be due to the presence of small pore diameters and precipitation of anhydrous α-lactose as α-lactose monohydrate in these pores. The smaller initial solubility of α-lactose monohydrate produces this effect and the tablets exhibit higher dissolution [13].

Spray dried lactose contains a mixture of crystals of lactose monohydrate and spherical agglomerates of small crystals bound together by glassy or amorphous material. The hardness of tablets was found to increase with decrease in particle size [14]. Furthermore, disintegration time is reported to increase with an increase in compression force [15]. Addition of lubricant to spray dried lactose prolongs the disintegration time and drug dissolution [16].

The granulated form of α-lactose monohydrate (agglomerated lactose) possesses good flow and binding property but not as good as spray dried lactose. The presence of increasing amounts of β-lactose in agglomerated lactose is suggested to impart strong intergranular cohesion [17].

CELLULOSE DERIVATIVES Microcrystalline cellulose (MCC) finds wide application in the formulation of tablets by both direct compression and wet granulation. MCC possesses good flow property, hardness, and friability. It is available in different grades depending on physical state, powder (PH-101) and granular (PH-102) as well as particle size. It can be used for controlling the release of both water-soluble and water-insoluble drugs when combined with other diluents like lactose, starch, and dibasic calcium phosphate [18]. The granular form (PH-102) is reported to exhibit better flow [15]. The denser versions of PH-101 or PH-102—PH-301 and PH-302, respectively—have improved flowability but reduced compressibility [19]. Tablets containing higher proportion of PH-101 were found to exhibit higher crushing strength and lower disintegration time, while those containing PH-102 and PH-200 exhibited lower crushing strength and shorter disintegration time [20]. An increase in the ratio of length to diameter of particles (Celous KR 801) yields tablets with higher tensile strength [21]. Similarly, presence of cellulose II polymorph in Avicel PH-102 gives more elasticity to particles than Avicel PH-102 that contains cellulose I polymorph [22].

SUGARS Sucrose is widely used in the manufacture of chewable tablets [23]. However, its use is restricted due to concerns regarding microbial growth, high calorie contribution, and moisture sensitivity. Di-Pac consists of 97% sucrose and 3% modified dextrin [24]. It is directly compressible but tends to cake and lose its fluidity at high moisture level. Due to cocrystallization of sucrose and dextrin, Di-Pac undergoes plastic deformation and produces harder tablets that dissolve slower than those obtained from sucrose alone [25]. Mannitol is water soluble and nonhygroscopic and is an important ingredient in chewable tablets due to its negative heat of solution that produces a cooling effect in the buccal cavity. Emdex is produced by hydrolysis of starch and consists of aggregates of dextrose microcrystals intermixed with higher molecular weight sugars. Theophylline tablets prepared with emdex were found to exhibit higher mechanical strength, rapid disintegration, and faster drug release than the tablets prepared with maltodextrin M 450 [26]. The rate of release of water-insoluble drug was found to be significantly affected by addition of lubricants in tablets prepared from maltodextrin (M 700, M 500, and M 150) while the release of water-soluble drug remained unaffected [27].

Starch may be obtained from corn, wheat, potatoes, tapioca, or rice. Tablets containing a high concentration of starch are usually soft. An increase in starch content and precompression force and a decrease in granule size was found to enhance the dissolution rate of salicylic acid tablets [28]. Presence of high initial moisture content in starch often leads to poor tablet properties. Therefore, directly compressible varieties are used for improving processing characteristics.

Sta-Rx-1500, a partially gelatinized cornstarch, is free flowing and directly compressible. If compressed alone, it is self-lubricating. However, on addition of 5–10% ingredient that does not possess lubricating property, it requires additional lubricant, usually a glidant, such as colloidal silicone. It softens on mixing with magnesium stearate [29]. Compressible starch gives faster dissolution as compared to other normal starches [20, 30]. Rice starch (Era Tab) produced tablets of terfenadine with higher crushing strength and lower friability than those produced from Sta-Rx-1500, dicalcium phosphate, or Avicel PH-101 [31]. Uni-pure, a fully gelatinized maize starch, was found to possess strong binding properties with faster dissolution and disintegration [32].

DICALCIUM PHOSPHATE Dicalcium phosphate dihydrate is the most common inorganic material used as a diluent in direct compression. The slightly alkaline nature (pH 7.0–7.4) makes it less suitable for use with drugs that are prone to alkali degradation [33]. Emcompress, containing aggregates of dicalcium phosphate, is directly compressible. However, it requires addition of a large quantity of lubricant for easy ejection from dies. Because Emcompress undergoes fragmentation during consolidation, the added lubricant significantly reduces dissolution of drugs [34].

*Coprocessed Diluents for Direct Compression* Coprocessing of well established diluents is a method that is used for obtaining an integrated product, which possesses superior functionality. A few well known coprocessed, directly compressible diluents are summarized in Table 2. Although improved functionality depends on the specific purpose related to drug delivery, all directly compressible diluents, in general, need careful evaluation with respect to flowability, tablet characteristics, and reworkability [35].

COPROCESSED LACTOSE Ludipress consists of α-lactose monohydrate (93.4%), polyvinyl pyrrolidone (3.2%), and crospovidone (3.4%). In spite of the presence of disintegrant, tablets containing Ludipress take more time to disintegrate than those containing α- lactose monohydrate, Tablettose, or anhydrous β-lactose [36]. Although addition of lubricant is essential, the mixing time was not found to influence the crushing strength of tablets [37]. The disintegration time of tablets compressed using Ludipress at 100 MPa pressure did not change, whereas that for Cellactose tablets got prolonged [38, 39]. Cellactose contains α-lactose monohydrate (75%) and cellulose (25%). Good compressibility is attributed to synergistic effect by fragmentation of lactose and plastic deformation of cellulose [40, 41]. Investigators have reported better compressibility of Cellactose as compared to Ludipress, Fast Flo Lactose, Tablettose, Di-Pac, anhydrous lactose [37, 42], and a mixture of Avicel PH-101 and Tablettose [43]. Reduction in compression pressure, which does not eliminate macropores, was reported to decrease the disintegration time of Cellactose tablets [44]. Tablets prepared from granules containing lactose coprocessed with microcrystalline cellulose and dextrin [45], dicalcium phosphate, and microcrystalline cellulose [46] were found to exhibit satisfactory properties.

Pharmatose DCL 40 contains β-lactose (95%) and anhydrous lacitol (5%). It is spherical in shape, exhibits good flowability, and is comparatively stable on exposure to high humidity [23].

StarLac is produced by spray drying of α-lactose monohydrate and maize starch [47]. Its tablets exhibited very good flow behavior, acceptable crushing force due to presence of lactose, rapid disintegration due to starch, and rapid release of theophylline [48].

COPROCESSED MICROCRYSTALLINE CELLULOSE A codried mixture of microcrystalline cellulose and β-cyclodextrin was found to exhibit improved flowability and compressibility [49]. However, coprocessing microcrystalline cellulose with sodium lauryl sulfate by ultrasonic homogenization followed by spray drying yielded a product that was inferior to microcrystalline cellulose for tableting paracetamol [50]. The PH-M series of microcrystalline cellulose is reported to be suitable for rapidly disintegrating tablets of acetaminophen and ascorbic acid [51]. The tensile strength of tablets prepared from coprocessed granules containing microcrystalline cellulose

**TABLE 2 A Few Important Commercially Available Coprocessed Diluents and Their Functionality**

| Brand Name | Adjuvants | Properties |
|---|---|---|
| Cellactose | Coprocessed lactose and microcrystalline cellulose (MCC) | Among directly compressible fillers, MCC is the most compressible and has the highest dilution potential because of its high binding and good disintegrating properties. MCC is of interest when combined with other less compressible excipients as an added component in coprocessed direct tableting excipients. |
| XyliTab | Directly compressible xylitol and sodium CMC | Xylitab granulates provide the convenience of direct compression. Use of Xylitab avoids the need for lengthy traditional granulation processes and allows dry mixing of tablet components. |
| Ludipress | β-Lactose and crospovidone, PVP | Direct compression auxiliary with medium porosity and good flow properties. |
| Starlac | Lactose and maize starch | Direct compression excipient, improved flowability, rapid disintegration of tablets, increases crushing strength. |
| Pharmatose DCL 11 | Anhydrous lactose, maize starch | Excellent flow properties and porous surface structure ensure excellent mixing properties as a consequence; DCL 11 permits very low tablet weight variation and high drug content uniformity; DCL 11 is only slightly affected by the level of lubrication. |
| Pharmatose DCL 14 | Mixture of fine crystals of lactose monohydrate and amorphous lactose | Directly compressible product, better compaction properties, uniform particle size, high degree of flowability. |
| Avicel CE 15 | MCC and guar gum | Incorporates particle-engineered products and process technology; provides excellent control of sensory factors that affect palatability; imparts such pleasing sensory features to chewable tablets that the perceived taste of drugs is remarkably altered; produces comparably softer tablets that are less friable and disintegrate rapidly. |
| Carbofarma GS 12 | Calcium carbonate (90%) and pregelatinized starch (10%) | High fluidity (easy workability at high speeds), high density (allows high calcium dosage in small tablets), excellent palatability (no awful feeling especially in chewable formulations). |
| Carbofarma GM 11 | Calcium carbonate (90%) and maltodextrin (10%) | High fluidity, high density, and excellent palatability. |
| Plasdone S-630 | Vinyl acetate and vinyl pyrrolidone | Versatile wet granulation tablet binder, viscosity modifier, crystal growth inhibitor, and solubilizer in liquid dosage forms; available in pyrogen-free "C" grade suitable for parenteral and ophthalmic applications; soluble in water and all pharmaceutically acceptable solvents; protective colloid/stabilizer, demulcent, and lubricant in ophthalmic solutions, solubilizer by complexation. |
| BarcroftTM CS90 | Spray dried calcium carbonate and starch | Highly compressible and free flowing. Simply adjusting the compression force used can allow one to make both chewable and conventional tablets. |
| Di-Pac | Sucrose and dextrin | Excellent flowability, high compressibility, low hygroscopicity, sweet and nonreactive. |
| Advantose FS 95 | Codried fructose | Turns fructose into an excellent excipient for pharmaceutical, nutraceuticals, and chewable vitamin tablets, improved flow properties, excellent compressibility. |

and calcium carbonate decreases with an increase in the amount of calcium carbonate [52].

Prosolv contains microcrystalline cellulose (98%) coprocessed with colloidal silicone dioxide (2%). It has better flowability and compressibility [19, 53] and the tablets maintain tensile strength even in the presence of magnesium stearate, unlike microcrystalline cellulose [54]. A marked enhancement in disintegration time was observed in tablets coated with Prosolv as compared to Avicel [55].

Low crystallinity cellulose is prepared by reacting cellulose with 85% w/w phosphoric acid fumes at room temperature for 1 h, then at 50 °C until a viscous solution is formed. This is poured into water to obtain a fine powder [56–58]. Kumar et al. [59] found that tablets obtained by compressing low crystallinity cellulose (LCPC) were intact even after 6 h, whereas those obtained by compressing LCPC-2000 disintegrated in 4 min. LCPC-4000 was found to be the most ductile material and exhibited the highest compression and compaction characteristics [59].

***Lubricants*** Lubricants are used for reducing the friction between granules and the die wall during compression. Antiadherents prevent sticking to the punch, while glidants improve the flowability of granules. They act by either fluid (hydrodynamics) lubrication by providing a continuous layer of fluid lubricant or by boundary lubrication that results from adherence of polar portions of molecules with carbon chains to metallic surfaces. Additional mechanisms by which antiadherents and glidants improve granule properties relate to electrostatic charge, gas adsorption, and van der waals forces [60, 61].

Lubricants are hydrophobic in nature. Their presence on the granule/tablet surface may decrease the area of contact between the particles and interfere in particle–particle bonding, thereby providing a less cohesive tablet [62]. Furthermore, the coating of lubricant on particles renders the surface hydrophobic and may increase the disintegration time while decreasing the dissolution rate.

An increase in concentration of calcium stearate, glyceryl monostearate, magnesium stearate, or stearic acid from 0.1% to 5% was observed to decrease the dissolution rate of salicylic acid and aspirin, whereas no effect was observed when talc was used [63]. It is important to ascertain whether the lubricant exhibits interaction with excipients or not. Although magnesium stearate possesses higher specific surface area than sodium stearyl fumarate, its adhesion to drug–crospovidone agglomerates results in a decrease in dissolution rate [64]. The negative effect of lubricants on dissolution can be counteracted by choosing diluents that possess both high solubility and fragmentation [65]. The deleterious effect of magnesium stearate (0.5%) on dissolution of ketorolac tromethamine was alleviated by hydrophobic anionic surfactants (sodium *N*-lauroyl sarcosinate, sodium stearoyl 2-lactylate, and sodium stearate), nonionic surfactants (polaxmer 188), and cationic surfactant (cetyl pyridinium chloride). However, lipophilic surfactant (glyceryl monostearate) was ineffective in counteracting the effect of magnesium stearate [66]. The deleterious effect of lubricant on drug dissolution can also be overcome by including superdisintegrants in the formulation [67].

The effect of lubricant on dissolution depends on the aqueous solubility of the drug. Magnesium stearate did not influence the dissolution rate of hydrochlorothiazide, sorivudine, and aztreonam capsules stored at 50 °C for six months. Although caking was observed in capsules of all the drugs on storage at 40 °C/75% RH, the

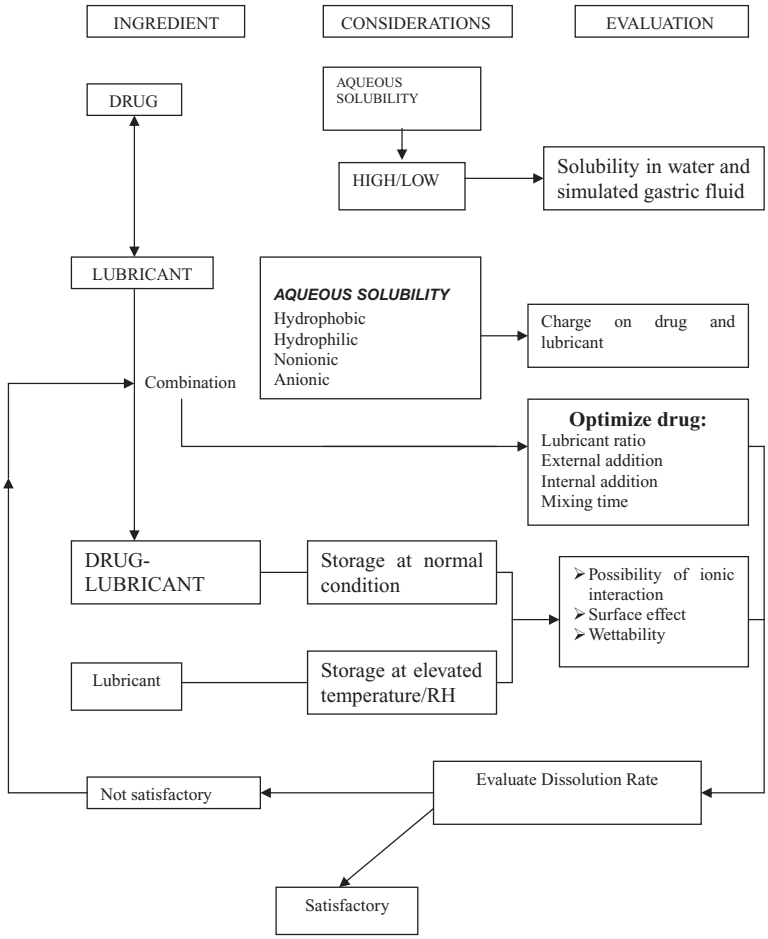


**FIGURE 2** Evaluating the influence of lubricant on drug dissolution.

decrease in dissolution rate of only sorivudine could be attributed to this effect. On the other hand, aztreonam, due to its inherent high aqueous solubility, did not exhibit any discernible effect on dissolution [68].

The addition of lubricants is known to decrease tablet hardness and results in incomplete drug dissolution [64, 69–72]. External addition of magnesium stearate was found to produce tablets that disintegrated immediately and rapidly released trypsin as compared to those containing internally added lubricant. This was ascribed to higher water penetration rate in tablets prepared by external addition of magnesium stearate [73]. This aspect becomes important when a high dose of poorly wettable drug has to be incorporated in tablet formulation [74]. A decision tree for evaluating the influence of lubricants on drug dissolution during preformulation phase is depicted in Fig. 2.

***Disintegrants*** The dissolution of a solid dosage form depends to a large extent on its mode of disintegration. The properties of disintegrant and diluent together

dictate the disintegration as well dissolution of a solid dosage form. Disintegrants acting by swelling mechanism generally increase the dissolution efficiency with an increase in compression pressure [74]. The fragmentation at high compression pressure was suggested to be responsible for increasing dissolution rate of tablets containing starch [75]. Slightly swelling disintegrant (potato starch) containing diazepam tablets were found to exhibit more disintegration time with an increase in mixing time with magnesium stearate. On the other hand, tablets containing strongly swelling disintegrant (sodium starch glycolate) were not significantly affected. Furthermore, *in vitro* dissolution of potato starch containing tablets could be correlated with *in vivo* absorption by employing a high stirring rate [76]. This suggests an overwhelming role of swelling capacity of disintegrant in influencing drug dissolution from dosage forms.

Systems containing high quantity of hydrophobic drug possess an area with pronounced hydrophobicity where wetting acts as the rate-limiting step in dissolution. In such systems, the incorporation of disintegrant does not affect drug dissolution [77]. An increase in hydrophobicity of tablet formulation decreases the efficiency of superdisintegrants [78]. Storage of wet granulated tablets containing superdisintegrants revealed a decrease in dissolution efficiency, with the effect being more pronounced for croscarmellose than for crospovidone or sodium starch glycolate [79]. This is due to slow loss of swelling capacity of disintegrants upon storage of wet granulated tablets.

Incorporation of a large concentration of superdisintegrant reduces the release of poorly soluble drugs due to formation of a viscous layer around the granules during the dissolution process [80]. Lundqvist et al. [81] reported a reduction in disintegration and an increase in friability time when the amount of disintegrant pellets compressed into tablets was increased. However, disintegrants possessing a higher true particle density enhance the disintegration time of pellets. Furthermore, although barium sulfate or calcium carbonate were more effective than magnesium oxide or ferric oxide, use of higher concentration produced an adverse effect due to the large increase in pellet density [81]. Therefore, both the type and concentration of superdisintegrant seem to be critical for obtaining optimum drug release characteristics.

Furthermore, the performance of a disintegrant varies with the pH of the medium. Significant differences in dissolution of acetaminophen from tablets containing sucrose and Primogel or Ac-Di-Sol are attributed to the depressed functioning of Primogel coupled with stronger binding of sucrose in acidic medium than that in neutral medium [82]. Similarly, tablets of terfenadine showed improved disintegration/dissolution with superdisintegrants in the order Crospovidone > Ac-Di-Sol > Primogel > low substituted hydroxypropyl cellulose. Starch decreases dissolution rate by forming a barrier in the form of swollen grains around the particles [83].

It is important to note that the solubility of the major component (diluent) in a tablet formulation can significantly affect disintegration and dissolution. Water-soluble materials tend to dissolve rather than disintegrate, while insoluble materials produce a rapidly disintegrating tablet if an appropriate quantity of disintegrant is included in the formulation [78]. For soluble formulations, the development of disintegration force could be hindered [84]. In addition, the role of soluble excipients like lactose that possibly act passively by hydrogen bond annihilation cannot be ruled out when present in tablet formulations [5]. As the quantity of water

penetrating into a tablet is limited, the soluble excipients can be expected to partially consume the available water, leaving only a part of it for development of disintegration force. Lopez-Solis and Villafuerte-Robles [85] found greater dissolution improvement of norfloxacin when tablets contained a dissolving diluent, Pharmatose DCL 11 and less hygroscopic disintegrant PVP XL 10, as compared to tablets containing more hygroscopic disintegrant like croscarmellose or starch 1500. Both sodium starch glycolate and croscarmellose sodium showed significantly reduced water uptake than crospovidone in 0.1 N HCl. Tablets containing the former disintegrant revealed significant increase in disintegration time for slowly disintegrating (lactose based) tablets as compared to rapidly disintegrating (dicalcium phosphate based) tablets. The dependence of dissolution rate of hydrochlorothiazide on both water uptake and solubility of ingredients in the testing medium suggests the importance of swelling capacity of disintegrants in determining drug release from tablets [86].

Furthermore, the surface area of disintegrated agglomerates during dissolution needs to be critically considered while judging drug dissolution from tablets. Difference in dissolution rate of aspirin from tablets containing different superdisintegrants was attributed to their behavior in promoting disintegration. Video images of the dynamic process indicated Ac-Di-Sol, Primogel, or Polyplasdone XL 10 promote, respectively, rapid disintegration, slow disintegration, and rapid disintegration but into large agglomerates [87].

Therefore, drug release from tablets is intimately connected with the performance of disintegrant present in the formulation. Solubility of diluent, water uptake, swelling behavior, development of disintegration force, and disintegration behavior of disintegrant are critical parameters to be considered while optimizing a tablet dosage form with respect to drug dissolution.

***Release Modifiers*** Polymers are used for modifying the release of drugs from solid dosage forms. They may be employed in matrix form or as a coating for this purpose. Furthermore, the physical form (tablets, granules, beads, pellets, etc.) of the dosage form may vary. Figure 3 gives an overview of the decision making process while selecting a release modifier for solid dosage forms. The choice of polymer(s) for modifying drug release depends on many factors, as summarized in Table 3. The drug release characteristics from dosage forms become more complex when combinations of release modifiers are used. It would not be possible to discuss all the release modifiers here with respect to their mechanism of drug release. Hence, a few important factors that may influence drug dissolution in the presence of release modifiers are discussed next.

*Polymer Characteristics* The desired drug release profile governs the choice of release modifier(s). Polymers possessing pH-independent solubility (Carbopol and HPC) provide continuous drug release in acidic as well as alkaline buffer media. Polymers that have pH-dependent solubility (Eudragits) are suitable for achieving drug release at a particular pH or site in the gastrointestinal tract. Polymers that are insoluble at all pH values (ethyl cellulose) need to be combined with other polymers to obtain the desirable drug release profile. Such polymers may be combined with pH-sensitive polymers to obtain drug release at specific pH values [88] or with polysaccharides to obtain drug release in the colon [89].

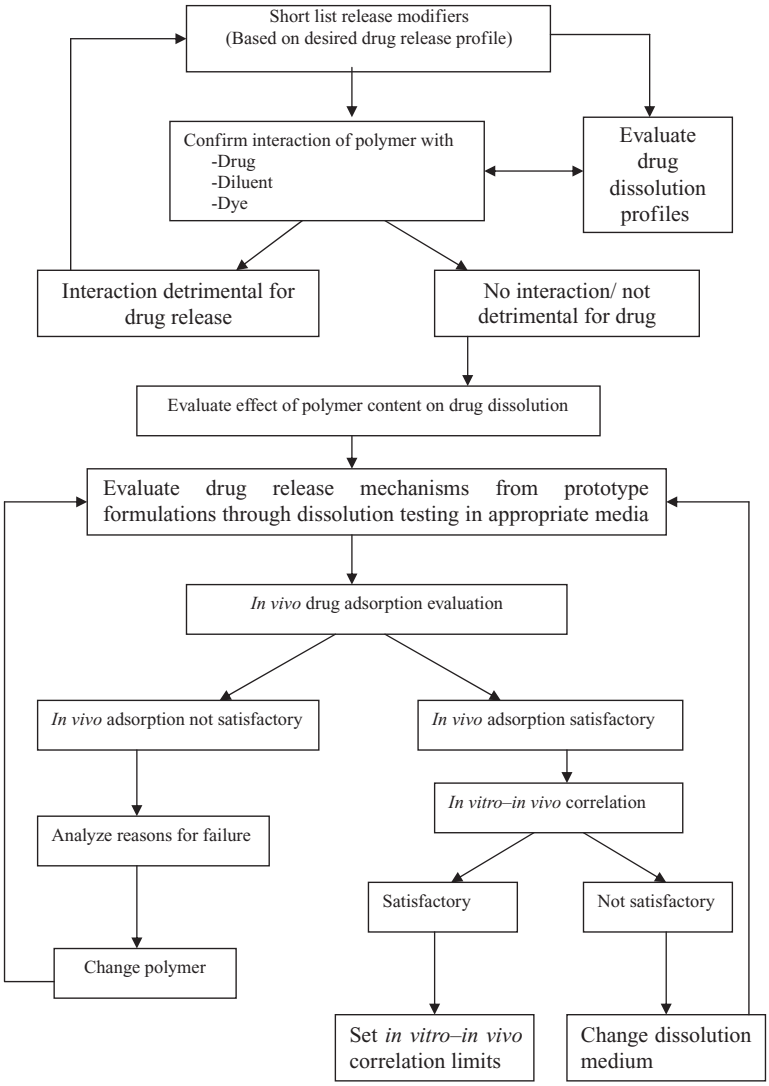


**FIGURE 3** Decision tree for evaluating the influence of release modifiers on drug dissolution.

*Polymer Content* The selected polymer or polymer combination is developed into the prototype formulation and evaluated for drug release profile through dissolution testing. Usually, preparation of a series of formulations with different polymer contents exhibiting different dissolution profiles [90] helps in elucidating the drug release mechanism(s). The drug release rate from a tablet formulated using polymer(s) depends on the molar ratio of polymer(s) [91, 92]. Chiao and Price [93] observed that release rate of propranolol hydrochloride from cellulose acetate butyrate microspheres increased with higher drug to polymer ratio and decreased with increasing diameter.

**TABLE 3 Dissolution as a Tool for Accessing Functionality of Release Modifiers**

| Purpose (Dosage Form) | Polymer Characteristic Desired | Dissolution Behavior Desired | Correlation with Absorption[a] |
|---|---|---|---|
| Release in stomach | Good solubility in 0.1 N HCl | Drug level detectable immediately. Peak level corresponds to disintegration. | Drug detectable in plasma soon after ingestion. Concentration falls depending on $t_{50}$ |
| Continuous release in stomach (floating) | Restricted solubility in 0.1 N HCl; swells in acidic medium, thereby delaying drug release. | Drug released with little delay; concentration rises continuously. | Drug detectable with little delay and with continuous rise for 3–5 h (fed state trial essential) |
| Sustained release | pH-independent solubility. Continuous release over extended time period irrespective of pH. | Drug released slowly with delay. Release continues at constant rate in pH-progression media. | Drug detectable in blood with little delay. Constant concentration maintained over prolonged period. |
| Colon release | Not soluble in both acidic and alkaline pH; degradation due to bacteroids present in cecal contents. | No/little (<10%) drug released on sequential exposure to 0.1 N HCl (2 h) and pH 7.5 (3 h). Total release in pH 6.8 containing cecal contents. | No drug detectable until 2 h after oral ingestion; abrupt rise in concentration after 5 h. (Assurance of no effect of pH on drug release essential.) |
| Site-specific release | Soluble at pH of the site. | Drug released only after exposure to particular pH. | No drug detected until dosage form reaches the desired site. |
| Time release | Slow erosion over the desired time interval. | Very little drug released over a particular time interval; abrupt release after that. | Negligible drug in blood until a few hours after ingestion; abrupt increase thereafter. (Assurance of no effect of pH, gastric emptying, and dietary factors essential.) |

[a]Assuming high solubility, high gastrointestinal permeability, and passive absorption.

The absolute amount of polymers also influences drug release. Extended release tablets of cephalexin showed reduced release in the presence of higher amounts of HPMC [94]. Release of ascorbic acid from tablets revealed that $t_{50}$ and $t_{90}$ was delayed from 1 h to 1.4 h and from 4.55 h to 7.2 h, respectively, when the total polymer content was increased to 60 g/500 g in comparison to 45 g/500 g. Similar

results were obtained when Kollicoat SR 30D was used in formulating ibuprofen granules [95]. However, slight modification in HPMC concentration does not significantly affect drug release rate [96].

*Polymer–Drug Interaction* Various interactions taking place between the drugs and polymers may lead to change in drug release pattern, where drug solubility in external medium as well as its diffusion capacity within the polymer network play an important role. The ability of Eudragit polymers to adsorb acidic drugs from a solution at pH 7.4 is well known. The ionized acid molecules carry a negative charge to ammonium groups linked to polymer backbone. This allows electrostatic binding to take place [97, 98]. This phenomenon occurs when the pH of the dissolution medium is higher than the $pK_a$ of the drug [97]. In the case of tris-HCl buffer (pH 7.4), chloride ions compete with drug molecule anions for active sites in the polymer backbone, thus resulting in a lower adsorption rate of drugs [99].

Many acids have been used to prepare chitosan based controlled release drug delivery systems. Such conditions may not be desirable for some drugs, especially for an acid labile drug [100]. It was further confirmed by Nunthanid et al. [101] that the drug release behavior was sustained from high molecular weight chitosan acetate films. In addition, they observed sustained release of theophylline from tablets prepared with spray dried chitosan and as its amount increased, the drug release occurred at a much slower rate [101].

*Polymer–Diluent Interaction* Like other factors, polymer–diluent interaction may also affect drug release. Microcrystalline cellulose allowed rapid penetration of aqueous medium, resulting in rapid erosion of polymer matrices. Lactose played an important role in imparting a physical barrier affecting diclofenac release kinetics by reducing tortuosity of diffusion pathway [102]. Anionic dyes like erythrosine, panceau 4 R, sunset yellow, and tartrazine are immiscible with chitosan citrate. But brilliant blue and green FS at concentrations of 0.02–1.0% w/w are miscible with chitosan. This could be due to reduced interaction because of the presence of positive charge on these dye molecules. Drug dissolution with these coated tablets was pH dependent and also corresponded to the ability of chitosan to protonate in the medium. Faster drug dissolution occurs from coated tablets colored with brilliant blue than from those colored with green FS. This is due to the fact that brilliant blue exhibits more water solubility due to the presence of more $SO_3H$ groups [103]. In another study, the miscibility of dye and chitosan was found to depend on the molecular configuration and ionic group of dye molecule. Drug dissolution from tablets coated with green FS is slightly slower than that from tablets coated with plain films and core tablet, respectively [104].

*Polymer–Excipient Interaction* Drug–excipient, excipient–excipient, and polymer–excipient interactions are of major preformulation consideration in the pharmaceutical industry and usually they can be predicted at early stages of product development. Interactions in the liquid state are more complicated than in the solid state, as sometimes predictions can be complicated by the dispersion medium, change in pH of the medium, ionic strength, a difference in the solubility parameter of the drug, or temperature fluctuations [98]. It is reported that PEG 6000 and Eudragit L-30 D have similar solubility parameters. This increases their affinity

toward each other [105]. Chances of interaction of Eudragit L-30D with PEG 6000 increase in an acidic medium due to hydrogen bonding between the ether oxygen of the oxyethylene groups of PEG 6000, resulting in faster dissolution [98]. PEG usually interacts with molecules that have hydrogen bonding functions as polymethacrylic acid polymers [105, 106].

***Physicochemical Characteristics of Drug*** The physicochemical properties of pharmaceuticals, including solubility and dissolution rate, can be influenced by the degree of crystallinity, solvation rate, crystal form, particle size/surface area, and so on. The solubility of a drug is influenced by many factors, which in turn affect the dissolution behavior. The rate of dissolution assumes great importance if it is lower than the rate of diffusion to the site of absorption or the rate of absorption of the drug. In such cases, there is a need for manipulating the drug characteristics in order to achieve the required dissolution profile.

Decreasing the particle size is the most common approach for increasing the solubility of a drug. However, the role of particle size should be considered for drugs exhibiting solubility of 0.1 mg/mL or less. This is because the chances of toxic effects associated with faster dissolution of a drug often increase with an increase in the drug solubility [107]. It is important to note that decreasing the particle size of a hydrophobic drug, phenytoin, actually decreases the proportion of particles in contact with dissolution medium. Furthermore, the small particles have air adsorbed on their surface due to which they float on the surface of the dissolution medium, thereby exhibiting lower dissolution [108]. The stability of different salts/coprecipitates in the gastro intestinal fluids needs to be critically evaluated. The most soluble sodium salt of penicillin is the least bioavailable form as it is most unstable in the digestive fluids [109].

Deaggregation of the drug particles is necessary to aid faster dissolution as evidenced by significantly different plasma concentrations obtained after administration of different brands of chloramphenicol capsules [108].

The availability of the weakly acidic drug para amino-salicylic acid is only 77% of that of its salt forms. A correlation in the rank order of availability of its potassium, calcium, or sodium salt forms indicated the pivotal role of their respective drug solubility in influencing the absorption of para amino-salicylic acid [108].

The beneficial effect of modifying the nonionized drugs into salt forms was reported for tetracycline and tolbutamide. The dissolution rate of weakly basic drug (tetracycline) decreases, whereas for the weakly acidic drug (tolbutamide) it increases as the pH of the dissolution medium is increased. However, the dissolution of salt forms of aluminum acetyl salicylate, sodium warfarin, or pamoate salt of benzamphetamine was found to be lower than that of their nonionized forms. This was suggested to be due to precipitation of an insoluble particle or film on the surface of the tablet rather than in bulk solution, which decreased the effective surface area by preventing deaggregation of the particles. Furthermore, the properties of the inactive ingredients present in the dosage form may also be influenced by the pH of the dissolution medium and may exert profound influence on dissolution characteristics of drugs (phenytoin, nitrofurantoin) [108]. Table 4 lists a few drugs whose dissolution is known to be influenced by physicochemical properties. In view of the above, it seems essential to compare the solubility and dissolution profile of the active ingredient/salt form with that of the dosage form in an

**TABLE 4 Influence of Physicochemical Factors on Dissolution Rate of Drugs**

| Drug | Characteristic Feature Influencing Dissolution | Remarks | References |
|---|---|---|---|
| Folic acid | Particle size | Formulations made from 9.0 μm particles release higher amount at 60 min as compared to formulations containing 24.4 μm sizes. Difference not statistically significant. | 110 |
| Griseofulvin | Particle size | Unmodified griseofulvin is required to dissolve NLT 70% and micronized griseofulvin is required to dissolve NLT 85% of stated amount in presence of 4% or 0.54% w/v SLS. | 111 |
| Nitrofurantoin | Particle size | Particles smaller than 10 μm are more rapidly and completely absorbed from tablet dosage form than macro-crystalline form (74–177 μm) from capsule. | 108 |
| Phenacetin | Particle size | Dissolution rate increased with increasing size. Addition of Tween-80 to dissolution medium increases the dissolution rate as particle size is decreased. | 108 |
| Para amino-salicylic acid (PAS) | Salt form | Solubility of nonionized PAS = 1 g/600 mL<br>Potassium PAS = 1 g/10 mL<br>Calcium PAS = 1 g/7 mL<br>Sodium PAS = 1 g/2 mL<br>Salt forms are more rapidly available than nonionized drug. | 108 |
| Tolbutamide | pH effect | As pH increases, dissolution rate increases. | 108 |
| Tetracycline | pH effect | As pH increases, dissolution rate decreases. | 108 |
| Ampicillin, glutethemide, theophylline | Anhydrous forms (solvate form) | At room temperature anhydrous forms exhibit higher dissolution rate than their corresponding hydrates due to difference in free energy of hydration. | 112 |
| Azetazolamide, carbamazepine, indomethacin | Metastable forms (phase transformations) | Because of the higher free energy they dissolve more rapidly than their corresponding stable forms. | 113 |
| Diclofenac sodium | Solid-solid transition | Anhydrous form dissolves faster than hydrated form. Storage under uncontrolled environment conditions (RH ≤ 60%, 25 °C) transforms the anhydrous form to hydrous form; adverse effect is anticipated during dosage form manufacturing too. | 114 |
| Chloramphenicol palmitate (polymorphic forms: A and B) | Polymorphism | As percentage of polymorph B increases, extent and rate of availability also increase due to increase in rate of dissolution. But since polymorph B is less stable, the less soluble polymorphic form A is used in formulations. | 108 |
| Diclofenac | Complexation (with cyclodextrin) | Time for dissolving 50% of drug is decreased. Dissolution efficiency of complex is 40.31 as compared to 1.21 of drug alone. | 115 |
| Triamterene | Complexation (with cyclodextrin) | 95% of drug dissolves in 2 h, whereas complexed drug dissolves within 10 min. | 116 |

appropriate medium in order to discern the effect of factors other than the inherent characteristics of the drug molecule.

## 2.2 Process Variables

### *Granulation Procedure*

*Wet/Dry Granulation* The type of granulation process can influence the dissolution rate of tablets [117]. A wet granulation process imparts hydrophilic properties to granules and hence is known to improve the dissolution rate of poorly soluble drugs [118]. Tablets of cephalexin formulated by a dry granulation process exhibited slower dissolution than those formulated by using a wet granulation technique. This was attributed to slow penetration of dissolution medium into tablets prepared by dry granulation [94]. On the other hand, dissolution of naproxen sodium tablets was fastest when they were manufactured by a direct compression method. The exothermic reaction during wet granulation produced the hydrate form of naproxen, which dissolved at a slower rate [119].

HPMC swells in contact with dissolution medium [120, 121] and forms a thick gel, which in turn may decrease the pore size and hence reduce drug release. However, in the presence of PVP (highly water-soluble additive) tablets prepared by wet granulation may undergo faster dissolution due to contact of the dissolution medium inside the swollen HPMC [94]. The dissolution rate of ticlopidine hydrochloride was improved by incorporating cornstarch in wet granulated tablets as compared to dry granulation. This could be attributed to better bonding and better homogeneity of starch disintegrant within the tablets [122]. Phenylpropanolamine tablets prepared with different granulation techniques resulted in different dissolution profiles. Partial melt and melt granulated tablets resulted in slower drug release than that obtained after dry blending or wet grinding. This was perhaps due to formation of different matrix structures from different methods. Furthermore, insufficient heat generated during compression does not seem enough to melt the wax and does not form internal structure similar to that produced in melt granulated and partial melt granulated tablets [123]. Faster dissolution of phenobarbital sodium was obtained from tablets prepared by compression with Avicel than that from tablets prepared by wet granulation with gelatin [124]. Tablets prepared by wet granulation using ethyl cellulose and lactose, acacia mucilage and lactose, or starch paste and lactose released the drug in the order: acacia mucilage and lactose = starch paste and lactose > acacia mucilage and lactose [117]. Therefore, the type of additive seems to play an important role in influencing drug dissolution besides the method of granulation.

*Granulating Solvent* Ghorab and Adeyeye [125] granulated ibuprofen with β-cyclodextrin using water, ethanol (95%), and isopropanol as the granulating solvents. Isopropanol and water, as granulating solvents, enhanced the dissolution of the oven-dried batches more than ethanol [125].

*Drying Process* A significant effect on particle size as well as on release rate of ibuprofen from microspheres was observed when the cooling time from 65 °C to room temperature was varied between 2 min and 100 min. The drug release was

maximum when drying was rapid. Granules prepared by a faster cooling process seemed to be more porous and had a less smooth surface [126].

*Granule Size* Granule size is not of considerable importance while studying dissolution rate of a drug, especially when the granules are softer [127, 128]. The effects of granule size and concentration of magnesium stearate as lubricant on the dissolution rate of paracetamol tablets were studied. The results obtained show that dissolution rate was increased as the granule size of the tablet was increased for tablets prepared with 1.5% magnesium stearate as lubricant. However, tablets prepared with different granule size exhibited no pronounced effect on the dissolution characteristics when the concentration of magnesium stearate used was 0.75% [129].

*Mixing* Some physical interactions between drug and excipient can take place during the dry blending phase of powders/granules. Blending with lubricants/glidants/antiadherents results in their absorption on particle surfaces, thereby increasing the hydrophobicity of drug particles. This decreases drug dissolution [2]. The dissolution rate of prednisolone was found to depend on the length of time for which ingredients were mixed with magnesium stearate [130].

***Compression Force*** The relationships between compression force and dissolution can be of four types:

1. Dissolution rate is increased as compression force is increased (e.g., use of starch paste as granulating agent).
2. Dissolution is slowed as the compressional force is increased (e.g., use of methylcellulose as granulating agent).
3. Dissolution rate is faster as compression force is increased to a maximum value. After this, the dissolution rate decreases with increase in compression force (e.g., use of gelatin solution as granulating agent).
4. Dissolution rate is slowed down as compression force is increased. However, further increase in compression pressure leads to an increase in dissolution rate [131].

Generally, an increase in compression force should decrease the bonding of particles because of crushing or cleavage of particles that in turn may increase the dissolution rate. Investigations on spironolactone tablets employed a granulated mixture of calcium carbonate and starch. The effect of manufacturing variables like compression force, crushing strength, or friability was studied by using tablets having the same composition but prepared under different conditions. The tablets compressed from each mixture exhibited a significant relationship between compression force and crushing strength or friability. Tablet strength increased linearly with increase in applied pressure. However, tablets prepared from large granulates showed an increase in disintegration time with increase in compression force, whereas those prepared from smaller granulates exhibited no dependence on compression force. In addition, disintegration times were shorter for the latter type of tablets. Furthermore, tablets prepared from smaller granulates required less time to dissolve [132].

Compression alters the size of drug particles, which in turn can affect the dissolution rate [133]. In a study conducted by Ganderton et al. [134], the dissolution rate of prednisolone tablets was examined at different compressional pressures. The dissolution rate of prednisolone from tablets prepared at low pressure was less compared to tablets manufactured at high pressures. With further increase in pressure, depressed dissolution rates were observed. This could be attributed to reduced breakage of tablets compressed at low pressure and release of stress along with loss of small air bubbles leading to disruption in tablets compressed at very high pressure. However, rebonding of material in tablets compressed at very high pressure produced stronger and denser tablets, which were not readily penetrated by dissolution medium [134]. A difference in pellet density of 19% was observed after compression at different pressures. The dissolution rate for the lot with lowest density made at lowest compression was significantly greater than other lots. This could be due to "channeling" that occurs within the pellet, thereby increasing the effective dissolving surface [135]. A study conducted by Chiao and Price [93] revealed faster release of propranolol hydrochloride from compressed tablets than from uncompressed microspheres. This may be attributed to rupture of microspheres at higher compression pressure. A compression force $<200\,kg/cm^2$ to the inner core tablet influenced the release behavior of drug but a force $>200\,kg/cm^2$ delayed the lag time [136]. Hence, it could be concluded that the effect of compression pressure is dependent on the pressure range evaluated and on the properties of drug, filler, binder, and so on. Therefore, it is rather difficult to predict the effect of compression pressure on the dissolution rate. Nevertheless, compression does influence drug release and this could be evaluated in relation to the properties of other formulation components.

## 3 DISSOLUTION TEST DESIGNING

Drug dissolution testing is a fundamental part of drug product development and manufacturing and is also employed as a quality control tool to monitor batch-to-batch consistency of drug release from a product. The dissolution test has extensively been used as a test to reflect the bioavailability of a drug product in humans. The advancements made for obtaining better correlation between dissolution and systemic availability include use of different media, evaluation of entire drug release profile, use of dialysis membrane, and flow through cell.

Drug classification based on permeability as well as solubility is gaining recognition and may be extremely useful during the early stages of dissolution methodology development. Changing the dissolution medium allows for testing the dosage form under conditions similar to the gastrointestinal tract. For a commercial product, dissolution testing is primarily used (1) to confirm manufacturing and product consistency, (2) to evaluate product quality during its shelf life, (3) to assess post-approval changes, and (4) to minimize the need for bioequivalency studies [137].

While developing a dissolution test for poorly soluble compounds during the early drug development phase, the process should focus on assessing relevant physical and chemical properties and dosage form design, as these play a major role in selecting the dissolution medium and apparatus. However, the strategy for designing dissolution tests depends on availability of additional data as well.

### 3.1 Media Selection

The choice of medium depends on the purpose of dissolution test. For batch-to-batch quality testing, selection of dissolution medium is based partially on the solubility data and the dose range of drug in the product to ensure that sink conditions may be justifiable, if it is shown to be more discriminating or if it provides reliable data, which otherwise can only be obtained with the addition of surfactants. On the other hand, when a dissolution test is used for evaluating the biopharmaceutical properties of a dosage form, it is more important that the proposed biorelevant test closely simulates the gastrointestinal environment rather than necessarily producing sink conditions [138].

The dissolution characteristics of oral conventional formulations should first be evaluated using test media within the physiological pH range (pH 1.2–6.8) and between pH 1.2 and 7.5 for modified release formulations. During method development, it may be useful to measure the pH of a test medium before and after the run to observe whether the pH changes during the test. An aqueous medium without surfactants is preferred, but an aqueous medium with surfactants may be used either to increase the probability of establishing an *in vivo* relationship or to increase the drug solubility in case of low solubility compounds [139, 140]. Table 5 highlights this aspect because different pharmacopoeias utilize different surfactants for testing drug release. The need for surfactants as well as their type and concentration need to be justified [143, 144]. The amount of surfactant needed depends on the critical micelle concentration of surfactant and the degree to which the compound partitions into the surfactant micelles. Furthermore, the critical micelle concentration of

**TABLE 5 Use of Different Surfactants at Various Concentrations During Dissolution Testing of Tablets** [111, 141, 142]

| Surfactant | Concentration | Drug Product (Tablet) | Official in |
|---|---|---|---|
| Sodium lauryl sulfate | 0.5% w/v | Medroxypyrogesterone acetate | USP |
| | 1% w/v | Megestrol | USP |
| | 0.09% w/v | Norethindone and ethinyl estradiol | USP |
| | 0.09% w/v | Norethindone and mestranol | USP |
| | 0.02% w/v | Norethindone acetate | USP |
| | 0.05% w/v | Penicillamine | USP |
| | 0.2% w/v | Praziquantel | USP |
| | 0.1% w/v | Spironolactone | USP |
| | 0.1% w/v | Spironolactone and hydrochlorothiazide | USP |
| | 4% w/v | Griseofulvin | USP, IP |
| | 1% w/v | Carbamazepine | USP |
| | 0.3% w/v | Estradiol | USP |
| | 0.54% w/v | Ultramicrosize griseofulvin | USP |
| Polysorbate 80 | 5 ppm in water | Levonorgesterol and ethinyl estradiol | USP |
| Polysorbate 20 | 1% v/v | Lacidipine | BP |
| Sodium dodecyl sulfate | 0.07% w/v | Cyproterone | BP |
| | 1.5% w/v | Griseofulvin | BP |
| | 0.1% w/v | Spironolactone | BP |
| | 0.3% w/v | Nimodipine | BP |

surfactant depends on the characteristics of the surfactants and the ionic strength of the dissolution medium.

Development of medium for colon-specific drug delivery systems is a great challenge. This is due to the diversity in rationale behind developing a delivery system. In addition, inadequate understanding of the colon's hydrodynamics and motility complicates the development of *in vitro* evaluation. USP dissolution methods using multiple-pH media are employed for such products [145].

Simple aqueous media such as simulated gastric fluid without pepsin (SGFsp) can be satisfactorily used for BCS-I (high solubility, high permeability) drugs. Milk as a biorelevant medium may be useful for detecting special food-formulation interactions. However, for BCS-II drugs (low solubility, high permeability), the choice of dissolution medium shall depend on the ionization behavior of the drug. For neutral BCS-II drugs, the presence of solubility enhancers seems to be necessary. However, for weakly basic BCS-II compounds, whose dissolution in stomach is of prime importance, the use of SGFsp medium seems to be important while assessing the initial dissolution. Furthermore, comparison of dissolution behavior in fluids that simulate conditions in the proximal small intestine in the fed state and also in fluids that simulates condition in the proximal small intestine in the fasted state is needed to establish the effect of meals on absorption of these drugs [146].

### 3.2 Apparatus Selection

Physical and chemical properties of a drug as well as formulation play an important role in selection of dissolution test apparatus, especially in the case of poorly soluble compounds. Sometimes, the existing dissolution apparatus may need modifications to accommodate new release mechanisms; for example, in the case of a nondisintegrating dosage form, there is a requirement of delivery orifice. On the other hand, a disintegrating delivery system faces problems as there is the requirement of transferring the dosage form (without losing any portion) to a different medium. In the case of modified release delivery systems, the challenging task is to change the medium to obtain a pH gradient or simulation of fed and fasted conditions. Both USP apparatus III (reciprocating cylinder) and apparatus IV (flow through cell) are useful when the drug release from the dosage form is to be tested in different media [111, 147, 148]. However, it is important to note that both type and purpose of using a particular dissolution apparatus are often different in USP, BP, and IP (Table 6). Despite these differences, each pharmacopoeia aims to closely mimic the *in vivo* conditions. Still, endeavors are needed to achieve similarity in dissolution testing apparatus for various dosage forms among different pharmacopoeias.

In addition, modifications in existing apparatus can be done depending on its applicability to the type of dosage forms. Burns et al. [149] found the modified paddle (BP type II paddle) method best for studying the dissolution release profile of floating propanolol capsules. In this apparatus paddle blades were set at the surface of the dissolution medium [149].

The surface area of a pure drug may be held constant in order to evaluate the dissolution rate. Agitation intensity is another alternative, where alteration of stationary film thickness surrounding the particles is reflected in a change in dissolution rate constant and should be carefully controlled.

**TABLE 6 Critical Differences in Dissolution Rate Test Apparatus Used in Various Pharmacopoeias** [111, 141, 142]

| Dissolution Apparatus | Used for Dosage Form | | | Critical Differences | | |
|---|---|---|---|---|---|---|
| | USP | BP | IP | USP | BP | IP |
| 1 | Coated tablet<br>Uncoated tablet<br>Delayed release<br>Extended release | Conventional | Conventional<br>Coated<br>Uncoated | Basket | Basket | Paddle |
| 2 | Conventional<br>Enteric coated<br>Extended release | Conventional<br>Extended release<br>Delayed release | Conventional<br>Coated<br>Uncoated | Paddle | Paddle | Basket |
| 3 | Extended release<br>Modified release dosage form<br>Bead type | Modified release dosage form | (Not official) | Reciprocating cylinder | Flow through cell apparatus | (Not official) |
| 4 | Modified release (especially when active ingredient has limited solubility/soft gelatin, bead products, suppositories/ poorly soluble drugs) | (Not official) | (Not official) | Flow through cell apparatus | (Not official) | (Not official) |
| 5 | Transdermal | (Not official) | (Not official) | Paddle over disk | (Not official) | (Not official) |
| 6 | Transdermal | (Not official) | (Not official) | Cylinder | (Not official) | (Not official) |
| 7 | Non disintegrating modified release dosage form<br>Transdermal | (Not official) | (Not official) | Reciprocating holder | (Not official) | (Not official) |

### 3.3 Dissolution Apparatus Operating Parameters

The operating parameters of dissolution testing apparatus also need to be evaluated critically before finalizing the dissolution conditions for a drug product. These parameters include medium volume, temperature, aeration/deaeration, sinker evaluation, sampling time point specification (single point or multiple point), and agitation speed.

***Volume of Dissolution Medium*** Table 7 lists pharmacopoeial dissolution testing requirements of a few drugs representing the biopharmaceutical classification scheme (Classes I–IV).

It is evident from this table that the recommended volume of dissolution medium varies from 500 to 1000 mL. A volume of 900 mL is the most common volume, while using apparatus 1 and apparatus 2 according to USP, BP, and IP specifications. The volume can be increased to 2–4 L in special cases (e.g., in the case of low solubility compounds) and then the standard vessels are rejected and large vessels are used for conducting dissolution testing. The volume can be decreased to 100–250 mL in the case of highly potent and low dose dosage forms [150, 151]. Hence, in both cases, proper justification is required for increasing/decreasing the volume of the testing medium.

***Temperature of Dissolution Medium*** The standard temperature conditions for dissolution testing of conventional oral (to be swallowed) dosage forms is 37 ± 0.5 °C [111, 141, 142]. Drug solubility is temperature dependent. Hence, control of temperature during the dissolution testing is very crucial. However, for products that are intended to be dissolved/dispersed in water and then administered orally, a temperature of 17.5 ± 2.5 °C needs to be employed [111]. A comparison of dissolution profiles of HALO™ at 36.5 °C and 37 °C indicated a considerable decrease in release (54% after 60 min at 36.5 °C, 63% after 60 min at 37 °C). Similarly, release of propranolol was observed to be 73% at 37 ± 0.2 °C as compared to 84% at 37 ± 0.5 °C after 300 min. Therefore, proper control of temperature during dissolution testing seems to be important [149].

***Aeration/Deaeration*** The dissolved gas/air in the dissolution medium can influence dissolution of drug particles. Dissolved air can lead to a change in the pH of the dissolution medium. Furthermore, air bubbles may act as a barrier for dissolution if they are present on the dosage form or in the mesh of the basket. In addition, aeration can retard dissolution as it can push the particles away from the center and toward the walls of the dissolution vessel [152, 153].

***Sinker Selection*** USP recommends some nonreactive material wire when the dosage forms tend to float [111]. The sinker should be evaluated for its ability to maintain the dosage form at the bottom of the vessel without inhibiting drug release. Paddle apparatus tablet sinkers were used by Cappola to prevent tablets from sticking to the bottom of the dissolution vessel. Overall, dissolution for all tablets with sinkers showed more rapid and complete drug release than tablets without sinkers [154].

**TABLE 7 Pharmacopoeial Dissolution Requirements of a Few Drug Products Belonging to the Biopharmaceutical Classification Scheme (Classes I–IV)**

| Drug product | Pharmacopoeial Requirements | | |
|---|---|---|---|
| | USP | BP | IP |
| | | *BCS Class I* | |
| Aspirin (acetyl salicylic acid) | 0.05 M acetate buffer (pH 4.5), 500 mL, apparatus 1 (50 rpm), NLT 80% of Q[a] dissolved in 30 min. | Acetate buffer (pH 4.5), apparatus 1 (50 rpm), NLT 70% of Q dissolved in 45 min. | Not recommended. |
| Aspirin tablet (delayed release) | Apparatus 1 (100 rpm), 90 min (buffer stage). | 0.1 N HCl, 1000 mL, apparatus 1 (100 rpm), NLT 5% of Q in 2 h; mixed phosphate buffer (pH 6.8), 900 mL, NLT 70% of Q in 45 min. | Not recommended. |
| Digoxin tablet | 0.1 N HCl, 500 mL, apparatus 1 (120 rpm), NLT 805 of Q dissolved in 60 min (average of 12 tablets has LT 75% of Q in 60 min). If amount dissolved in 60 min is MT 95% for any individual tablet, amount dissolved in 15 min is NMT 90% for each tablet. | Water, 600 mL, 120 rpm, basket apparatus, NLT 75% of Q is dissolved in 60 min. | Water, 600 mL, 120 rpm, basket apparatus, NLT 75% of Q is dissolved in 60 min. |
| Prednisolone tablet | Water, 900 mL, apparatus 2 (50 rpm), NLT 70% of Q is dissolved in 30 min. | Water, 900 mL, apparatus 2 (50 rpm), NLT 70% of Q is dissolved in 45 min. | Water, 900 mL, paddle (50 rpm), NLT 70% of Q is dissolved in 30 min. |
| Quinine $SO_4$ tablet | 0.1 N HCl, 900 mL, apparatus 1 (100 rpm), NLT 85% of Q is dissolved in 30 min. | 0.1 N HCl, 900 mL, basket apparatus (100 rpm), NLT 70% of Q is dissolved in 45 min. | 0.1 N HCl, 900 mL, apparatus 2 (100 rpm), NLT 70% of Q is dissolved in 30 min. |
| Quinine $SO_4$ ER tablet | 0.1 N HCl (900 mL), apparatus 1 (100 rpm) 20–50% dissolved in 0.125 D h; 43–73% dissolved in 0.5 D hNLT 70% dissolved in 1.5 D h | Not recommended. | Not recommended. |
| Quinine $BiSO_4$ ER tablet | Not recommended. | 0.1 N HCl (900 mL), basket (100 rpm), NLT 70% of Q is dissolved in 45 min. | Not recommended. |
| Warfarin (sodium) tablet | Water, 900 mL, apparatus 2 (50 rpm), NLT 80% of Q is dissolved in 30 min. | 0.68% w/v potassium dihydrogen orthophosphate (pH 6.8), 900 mL, basket (100 rpm), NLT 70% of Q is dissolved in 45 min. | 0.68% w/v potassium dihydrogen orthophosphate (pH 6.8), 900 mL, paddle (100 rpm), NLT 70% of Q is dissolved in specific time. |

**TABLE 7** ***Continued***

| Drug product | Pharmacopoeial Requirements | | |
|---|---|---|---|
| | USP | BP | IP |
| Levodopa tablet | 0.1 N HCl, 900 mL, apparatus 1 (100 rpm), NLT 75% of Q is dissolved in 30 min. | 0.1 N HCl, 900 mL, apparatus 1 (100 rpm), NLT 70% of Q is dissolved in 45 min. | 0.1 N HCl, 900 mL, apparatus 2 (75 rpm), NLT 75% of Q is dissolved in 30 min. |
| Lithium carbonate tablet | Water, 900 mL, apparatus 1 (100 rpm), NLT 60% of Q is dissolved in 30 min. | Not recommended. | Water, 900 mL, paddle (100 rpm), NLT 60% of Q is dissolved in 30 min. |
| Paracetamol tablet | Not recommended. | Phosphate buffer (pH 5.8), 900 mL, apparatus 2 (50 rpm), NLT 70% of Q is dissolved in 45 min. | Phosphate buffer (pH 7.6), 900 mL, paddle (50 rpm), NLT 80% of Q is dissolved in 30 min. |
| Diazepam tablet | 0.1 N HCl, 900 mL, apparatus 1 (100 rpm), NLT 85% of Q is dissolved in 30 min. | 0.1 N HCl (900 ml), apparatus 1 (100 rpm), NLT 70% of Q is dissolved in 45 min. | Water (900 mL), paddle apparatus (100 rpm), NLT 85% of Q is dissolved in 45 min. |
| Theophylline tablet | Water, 900 mL, apparatus 2 (50 rpm), NLT 80% of Q is dissolved in 45 min. | Not recommended. | Not recommended. |
| | | *BCS Class II* | |
| Griseofulvin tablet | Water, 4% sodium lauryl sulfate (SLS), 1000 mL, apparatus 2 (100 rpm), NLT 70% of Q is dissolved in 60 min. | Water, 1.5% sodium dodecyl sulfate, 1000 mL, apparatus 2 (100 rpm), NLT 70% of Q is dissolved in 45 min. | Water, 4% SLS, 900 mL, paddle apparatus (100 rpm), NLT 70% of Q is dissolved in 60 min. |
| Dapsone tablet | Dilute HCl (2 in 100), 1000 mL, apparatus 1 (100 rpm), NLT 70% of Q is dissolved in 60 min. | 0.1 M HCl, 900 mL, apparatus 1 (100 rpm), NLT 70% of Q is dissolved in 45 min. | 2% v/v HCl, 900 mL, paddle apparatus (100 rpm), NLT 70% of Q is dissolved in 60 min. |
| Ibuprofen | Phosphate buffer (pH 7.2), 900 mL, apparatus 1 (150 rpm), NLT 75% of Q is dissolved in 30 min. | Not recommended. | Phosphate buffer (pH 7.2), 900 mL, paddle apparatus (100 rpm), NLT 50% of Q is dissolved in 30 min. |
| Naproxen tablet | Phosphate buffer (pH 7.4), 900 mL, apparatus 2 (50 rpm), NLT 80% of Q is dissolved in 45 min. | Phosphate buffer (pH 7.4), 900 mL, apparatus 2 (50 rpm), NLT 70% of Q is dissolved in 45 min. | Not recommended. |
| Naproxen enteric coated | Not recommended. | 0.1 N HCl, 900 mL, apparatus 2 (50 rpm), 2 h in acid stage, NLT 5% of Q in acid stage. Buffer stage: 900 mL, 45 min, NLT 70% of Q in buffer stage. | Not recommended. |

| | | | |
|---|---|---|---|
| Trimethoprim | 0.01 N HCl, 900 mL, apparatus 2 (50 rpm), NLT 75% of Q is dissolved in 45 min. | Not recommended. | Not recommended. |
| | | *BCS Class III* | |
| Acyclovir tablet | Not recommended | 0.1 M HCl, 900 ml, apparatus 2 (50 rpm), NLT 70% of Q is dissolved in 45 min. | Not recommended. |
| Chlorpromazine HCl tablet | 0.1 N HCl, 900 mL, apparatus 1 (50 rpm), NLT 80% of Q is dissolved in 30 min. | 0.1 M HCl, 900 mL, apparatus 2 (50 rpm), NLT 70% of Q is dissolved in 45 min. | Not recommended. |
| Rythromycin stearate tablet | Phosphate buffer 0.05 M (pH 6.8), 900 mL, apparatus 2 (100 rpm), NLT 75% of Q is dissolved in 120 min. | 2.722% Sodium acetate (pH 5.0), 900 mL, apparatus 2 (50 rpm), NLT 70% of Q is dissolved in 45 min. | Not recommended. |
| Hydrochlorothiazide tablet | 0.1 N HCl, 900 mL, apparatus 1 (100 rpm), NLT 60% of Q is dissolved in 60 min. | Not recommended. | 0.1 M HCl, 900 mL, paddle apparatus (100 rpm), NLT 60% of Q is dissolved in 45 min. |
| | | *BCS Class IV* | |
| Acetazolamide tablet | 0.1 N HCl, 900 mL, apparatus 1 (100 rpm), NLT 75% of Q is dissolved in 60 min. | Not recommended. | Not recommended. |
| Cefixime tablet | Potassium phosphate buffer 0.05 M (pH 7.2) 900 mL, apparatus 1 (100 rpm), NLT 75% of Q is dissolved in 45 min. | Not recommended. | Not recommended. |
| Furosemide | Phosphate buffer (pH 5.8), 900 mL, apparatus 2 (50 rpm), NLT 80% of Q is dissolved in 60 min. | Phosphate buffer (pH 5.8), 900 mL, apparatus 2 (50 rpm), NLT 80% of Q is dissolved in 45 min. | Not recommended. |
| Hydralazine tablet | 0.1 N HCl, 900 ml, apparatus 1 (100 rpm), NLT 60 % of Q is dissolved in 30 min. | Not recommended. | Not recommended. |
| Sulfasalazine tablet | Phosphate buffer (pH 7.5), 900 mL, apparatus 1 (100 rpm), NLT 85% of Q is dissolved in 60 min. | Not recommended. | Not recommended. |

[a]Q, stated/labeled amount.

***Sampling Time Point Specification*** Sampling at early time points is used to show that there is little/no probability of dose dumping/accumulation. Drug release during the initial period should not be more than the specified amount. An intermediate time point is chosen to evaluate the *in vitro* drug release profile. Lastly, the final time point is used to explain almost complete release of drug from the respective dosage form. In the case of products with more than one active ingredient, drug release is determined at different time intervals for different active ingredients of the drug product [152]. Single point sampling during dissolution does not yield an estimate of the dissolution pattern of drug from a dosage form. Table 8 summarizes a few important drug products that are required to undergo single/multiple time point sampling. It is evident that multiple sampling assumes more importance while testing a modified release dosage form because it becomes essential to evaluate the entire dissolution profile rather than only the extent of drug dissolved.

***Rotational Speed*** Variation in the amount of drug dissolved can be due to variation of the hydrodynamics in the vicinity of a tablet. Fluctuations in flow can affect shearing of the tablet surface, deagglomeration of particles, mass transfer from solid to liquid, and mixing of tablet fragments and thus eventually influencing the hydrodynamics. The literature reveals that a change in rotational speed can alter the dissolution rate [28, 155]. Kukura et al. [156] showed that increasing the stirring from 50 to 100 rpm increased the intensity of shear force exerted by the fluid but did not improve the homogeneity of the spatial distribution of shear. Dissolution testing of carbamazepine tablets using the paddle method in hydroalcoholic medium was conducted at 100 rpm and 75 rpm. The results showed a decrease in the amount of carbamazepine dissolved at lower rotation speed [157]. An increase in rate and extent of ranitidine dissolved was observed along with less individual tablet variability when the basket method at 30 rpm was used instead of the paddle method at 500 rpm [154].

***Altered Dissolution Environment*** To study the impact of slight variations in vessel dimensions, a metal strip was placed to simulate altered dissolution environment. During testing, the metal strip is expected to force the product/product-deaggregates to remain off center, resulting in a large surface area for dissolution. In addition, the contact of dissolution fluid with product should also increase as the distance from the center increases. However, results with or without a metal strip were observed to be dependent, which could be due to product sensitivity to the density of disintegrated particles or gradual change in available surface of the dissolution vessel. Nondisintegrating tablets showed higher dissolution than disintegrating tablets with a metal strip in place due to production and settling of denser particles at the bottom [158]. Hence, slight modifications in the dimensions/shape of dissolution vessels do not seem to exert a remarkable influence on drug release rate provided there is no appreciable change in the hydrodynamics inside the vessel.

***Viscosity of Medium*** Dissolution rate decreases with an increase in the viscosity of the dissolution medium especially when the dissolution process is diffusion controlled [159].

**TABLE 8 A Few Examples of Single Point and Multiple Point Correlation**

| Drug Product | Dissolution Condition | Recommended Units Sampling | | Official in |
|---|---|---|---|---|
| Aspirin tablet | 0.05 M acetate buffer, 500 mL, pH 4.5, apparatus 1, 50 rpm, 30 min. | NLT 80% of Q is dissolved in 30 min. | | USP |
| Buffered aspirin tablet | 0.05 M acetate buffer, 500 mL, pH 4.5, apparatus 2, 75 rpm, 30 min. | NLT 80% of Q is dissolved in 30 min. | | USP |
| Aspirin ER tablet | Test 1: 0.1 N HCl, 900 mL, apparatus 2, 60 rpm, time 1 h; 4 h. | Tolerance time | Amount dissolved | USP |
| | | 1 h | 20–55% | |
| | | 4 h | NLT 80% | |
| Diazepam capsule | 0.1 N HCl, 900 mL, apparatus 1, 100 rpm, time 45 min. | NLT 85% of Q is dissolved in 45 min. | | USP |
| Diazepam ER capsule | Simulated gastric fluid TS (without enzymes), 900 mL, apparatus 1, 100 rpm, 0.042 D h, 0.167 D h, 0.333 D h, 0.5 D h. | Time (h) | Amount dissolved | USP |
| | | 0.042 D | 15–27% | |
| | | 0.167 D | 49–66% | |
| | | 0.333 D | 76–96% | |
| | | 0.5 D | 85–115% | |
| Methylphenidate HCl tablet | Water, 900 mL, apparatus 1, 100 rpm, 45 min. | NLT 755 of Q is dissolved in 45 min. | | USP |
| Methylphenidate HCl ER tablet | Water, 500 mL, apparatus 2, 50 rpm 0.125 D h, 0.25 D h, 0.438 D h, 0.625 D h, 0.875 D h. | Time (h) | Amount dissolved | |
| | | 0.125 | 20% and 50% | |
| | | 0.25 | 35% and 70% | |
| | | 0.438 | 53% and 83% | |
| | | 0.625 | 70% and 95% | |
| | | 0.875 | NLT 80% amount dissolved | |

**TABLE 8** ***Continued***

| Drug Product | Dissolution Condition | Recommended Units Sampling | | Official in |
|---|---|---|---|---|
| Oxprenolol HCl tablet | 0.1 N HCl, 900 mL, apparatus 1, 100 rpm, 30 min. | NLT 80% of Q is dissolved in 30 min. | | USP |
| Oxprenolol HCl ER tablet | Acid medium (0.1 N HCl), 900 mL. Dissolution medium (simulated intestinal fluid Ts: without enzyme), 900 mL, apparatus 1, 100 rpm, 1 h in acid. | Time | Amount dissolved | USP |
| | | 1 (acid medium) | Between 15% and 45% | |
| | | 1 (dissolution medium) | Between 30% and 60% | |
| | | 3 (dissolution medium) | Between 50% and 80% | |
| | | 7 (dissolution medium) | NLT 75% | |
| Procainamide HCl tablet | 0.1 N HCl, 900 mL, apparatus 1, 100 rpm, 75 min. | NLT 80% of Q in 75 min. | | USP |
| Procainamide HCl ER tablet | Test 1: 0.1 N HCl, 900 mL, apparatus 2, 50 rpm, time 1, 4, 6 hr | Time | Amount dissolved | USP |
| | | 1 | Between 30% and 60% | |
| | | 4 | Between 60% and 90% | |
| | | 6 | NLT 75% | |
| | Test 2: acid stage (1 h), buffer stage (0.05 M, pH 7.5, $PO_4$ buffer), NLT 8 h | Time | Amount dissolved | |
| | | 1 | Between 30% and 60% | |
| | | 4 | Between 60% and 90% | |
| | | 8 | NLT 80% | |
| | Test 3: medium same as in test 2; apparatus 2, 50 rpm; time 1, 3, 6, 8 h | Time | Amount dissolved | |
| | | 1 | Between 25% and 50% | |
| | | 3 | Between 40% and 75% | |
| | | 6 | Between 65% and 90% | |
| | | 8 | NLT 80% | |

***Ionic Strength*** Ionic strength of the dissolution medium plays an important role in the dissolution process [160]. Dissolution behavior of extended release dosage forms that use hydrophilic gel forming polymers seems to be significantly affected by small changes in the ionic strength of the dissolution medium. Certain ionic salts and drugs are reported to cause failure of HPMC based extended release products [161]. HPMC based extended release diclofenac sodium tablets had significantly different release profiles in dissolution media of the same pH but different ionic strengths. The dissolution rate of theophylline from extended release granules prepared using HPMC as a retarding agent was observed to increase with increasing ionic strength of the dissolution medium whereas an initial decrease followed by an increase in the dissolution rate of theophylline was observed from commercially available extended release matrix tablet formulation with an increase in the ionic strength of the dissolution medium [162, 163]. Alterations in dissolution profiles may be due to changes in thermal gelation point (critical temperature that determines the sol to gel transition of polymers in aqueous media) of the polymers [161], increased surface erosion leading to dissolution of polymer chains, increased osmotic pressure [161, 162], and/or lower polymer solubility [164].

***pH of Medium*** pH of the dissolution medium influences solubility of the drug. In addition, it can affect the solubility/stability of the excipients. Both factors may produce an entirely different dissolution profile if the pH of the dissolution medium is not selected carefully. Furosemide tablets were observed to dissolve faster as the pH of the medium was increased. The tablet brand that dissolved poorly at pH 4.6 also exhibited poor bioavailability [165]. The intrinsic dissolution rate of piroxicam increased with an increase in pH of the medium [144]. Thus, careful selection of pH and buffers composing the dissolution medium seems to be essential for *in vitro* dissolution studies as well as for prediction of bioavailability.

## 4 ANALYSIS OF DISSOLUTION DATA

Whenever a new solid dosage form is developed or produced, it is necessary to ensure that drug dissolution occurs in an appropriate manner. In addition, for a pharmaceutical company to obtain approval for a generic product, there are a series of criteria that must be addressed to demonstrate that the product is bioequivalent to a proprietary counterpart. Under certain conditions, one of these criteria is to demonstrate that the dissolution curves of the two products are similar. The *in vivo* bioavailability study investigates the rate and extent of drug absorption in humans. The fact that drug absorption depends on the dissolved state of the drug has encouraged pharmaceutical companies to explore the relationship between *in vivo* drug bioavailability and *in vitro* dissolution. It has been suggested that *in vitro* dissolution testing can be used as a surrogate for *in vivo* bioequivalence studies to assess equivalence between the test and reference formulations, and for postapproval changes. If dissolution profile similarity is demonstrated for the formulations before and after the changes, then expensive *in vivo* bioequivalence testing can be waived off. Various procedures have been proposed to compare dissolution profile data (Fig. 4). The utility of these methods is briefly discussed next.

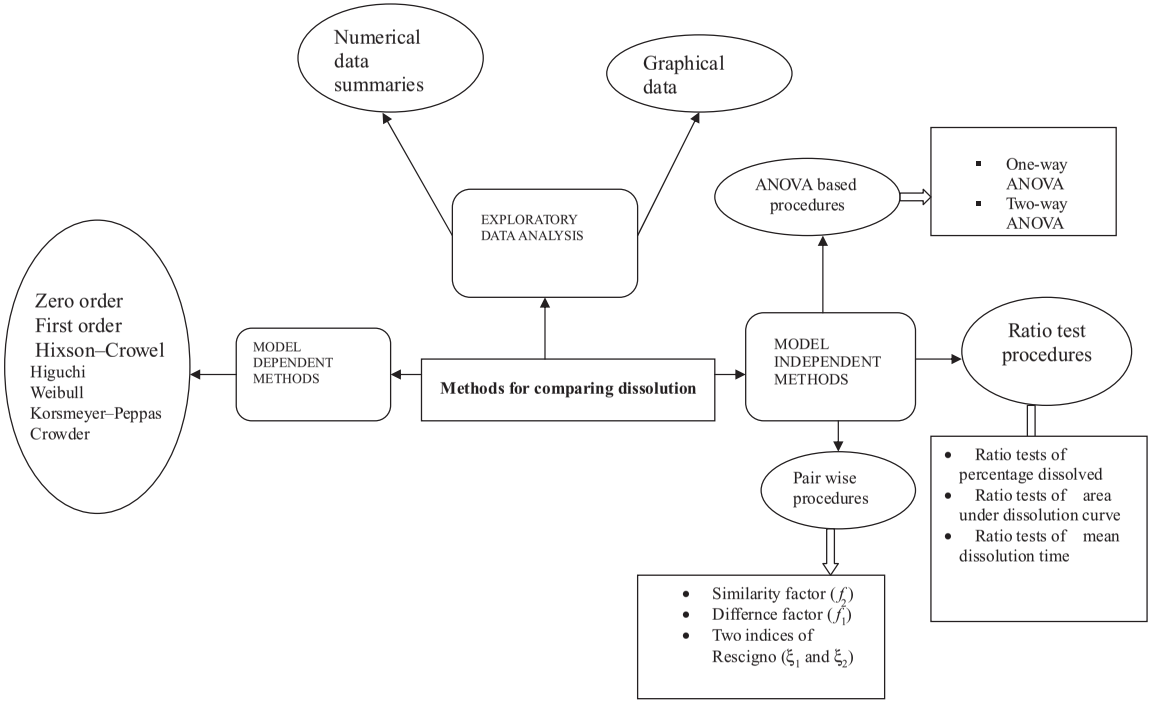


**FIGURE 4** Various methods used for evaluating dissolution data.

***Exploratory Data Analysis*** As a first step, it is useful to perform exploratory data analysis to compare dissolution profile data. The U.S. FDA does not currently endorse these methods. In this method the dissolution profile data are illustrated graphically by plotting mean dissolution profile data for each formulation with error bars extending to two standard errors at each dissolution time point. Then the data are summarized numerically and 95% confidence intervals are evaluated in mean dissolution profiles at each dissolution time point. This method is useful in obtaining an improved understanding of dissolution data. It can also be used as a first step to compare the dissolution profile data (in a graphical as well as numerical manner). However, the analysis of dissolution profiles becomes difficult if the error bars for two formulations overlap (graphically as well as numerically) [166, 167].

***Statistical Methods (Model-Independent Methods)*** Using one-way ANOVA, statistical comparisons of the mean dissolution data at each dissolution time point are performed. One-way ANOVA is equivalent to the *t*-test in the case where the dissolution profile data of two formulations is compared. This method takes the variability in the dissolution profile data into account while comparing each time point. Two-way ANOVA is used in data analysis when both formulation and time are class variables (in the case of immediate release data when one-way ANOVA is not applicable). But the major limitation is that this method ignores the correlation between the dissolution time points [168, 169].

***Multivariate Approach (MANOVA) Methods*** These are based on repeated measures, where time is the repeated factor and percent drug dissolved is the dependent variable. The method is more precise as it takes into account the variability as well as the correlation structure in data. However, this method may not be too informative, because it doesn't contain information about the nature of differences between mean dissolution profiles. In addition, it is difficult to implement this method tactically [170].

***Chow–Ki Method*** This is another statistical method, which is comparable to the approach used for assessing the average bioequivalence of two formulations. Equivalence limits for similarity at each dissolution time point are derived from Q (as defined in USP). It uses the concept of local and global similarity to assess closeness of test to reference dissolution profiles. This method is relatively easy to implement. Correlation between successive time points is accounted for by using an autoregressive time series model. Power and Type I error of this method are unknown at this stage. Type I error is the probability of concluding that the test and reference mean profiles are different when in fact they are similar. Type II error is the probability of concluding that the test and reference mean profiles are the same when they are actually different [168].

***Model-Independent Method*** This uses the dissolution data in their native form. Two mathematical methods are described in this method. The first is the difference factor equation, which is also called the $f_1$ equation, and the second is the similarity factor, also called the $f_2$ equation. The FDA approves both equations for dissolution profile comparison but the similarity equation is preferred.

$$f_1 = \left( \frac{\sum_{t=1}^{n} |R_t - T_t|}{\sum_{t=1}^{n} R_t} \right) \times 100$$

$$f_2 = 50 \log_{10} \left\{ \left( 1 + \frac{1}{n} \sum_{t=1}^{n} w_t (R_t - T_t)^2 \right)^{-0.5} \times 100 \right\}$$

where $n$ is the number of dissolution sample time points; $w_t$ is the optional weighting factor, and $R_t$ and $T_t$ are the mean percent drug dissolved at each time point, respectively, for each reference and test product. Evaluation requires the use of a minimum of three time points (except zero), not less than 12 individual values for every time point for each formulation, and the standard deviation should not be more than 10% from the second to the last time point. When the calculated values of $f_1$ lie between 0 and 15 and those of $f_2$ lie above 50, the dissolution profiles are interpreted to be similar. In $f_2$ estimation, bias could take place due to the contribution of variances in the percentage of drug dissolved in test and reference preparations measured at a particular time point. Unbiased similarity factor ($f_2^*$) is calculated to determine the effect of variances of the percentage of drug dissolved from test and reference products at particular time points on $f_2$:

$$f_2^* = 50 \left( \log_{10} \left\{ 1 + \left( \frac{1}{n} \right) \left[ \sum_{t=1}^{n} (R_t - T_t)^2 - \sum_{t=1}^{n} (S_r^2 - S_t^2)/N \right] \right\}^{-0.5} \times 100 \right)$$

where $S_r^2 - S_t^2$ is the variance in the percentage of drug dissolved measured at any time point of reference and test preparations and $N$ is the number of units on which dissolution is conducted. By this method it is easy to compare the data. In addition, it also provides a single number to describe the comparison of dissolution profile data. However, it is not possible to know the Type I and Type II errors [171–173].

***Rescigno's Indices*** Rescigno's indices have been developed for comparing blood-plasma concentration functions. $\xi_i$ and $\xi_2$ are considered analogous to $f_1$ and $f_2$:

$$\xi_i = \left[ \frac{\int_0^{t_n} |R_t - T_t|^i dt}{\int_0^{t_n} |R_t + T_t|^i dt} \right]^{1/i}$$

where $R_t$ and $T_t$ are the mean percentages of drug dissolved, respectively, for reference and test formulations at each time point and $t_n$ is the final dissolution time point. $\xi_i$ can be thought of as a function of the weighted average of the vertical distances between test and reference mean profiles at each time point. More precisely, $\xi_i$ is the absolute value of vertical distance and $\xi_2$ is the square of vertical distance.

The denominator of $\xi_i$ is a scaling factor. The main advantage of this method is that, unlike $f_1$, the value of $\xi_i$ remains unchanged if reference and test are interchanged. But this method does not take into account variability and correlation structure data like $f_1$ and $f_2$ equations, and also it is difficult to calculate Rescigno's indices. Moreover, there is no definite criterion for concluding difference or similarity data profile [174].

***Korsmeyer–Peppas Model*** This is a simple, semiempirical model relating exponentially the drug release to the elapsed time. This model uses the following equation for evaluation of dissolution data:

$$f_t = at^n$$

where $a$ is a constant incorporating structural and geometric characteristics of the dosage form and $n$ is a release exponent indicative of the drug release mechanism. This equation can be written

$$f_t = \frac{M_t}{M_\infty} \quad \text{(fractional release of drug)}$$

where $M_t/M_\infty$ is a function of $t$ (fractional release of drug). Modifying the above equation to accommodate lag time generates the following equation:

$$\frac{M(t-l)}{M_\infty} = a(t-l)^n$$

where $l$ is the lag time.

In the case of a burst effect, the equation becomes

$$\frac{M_t}{M_\infty} = at^n + b$$

In the absence of lag time or burst effect, $at^n$ is used and the resulting model, also known as the power law, has been used, very frequently, to describe the drug release from several modified release dosage forms [175, 176].

It is a simple and empirical model and describes drug release from modified release dosage forms [175, 177, 178]. This model assumes that the rate-limiting step to drug release is erosion of matrix. The time-dependent diffusional resistance to eroding matrix is hypothesized not to influence drug release [179].

***Crowder Model*** This is a mechanistic model. It uses repeated measures regression analysis techniques for comparing data from different formulations. It is considered superior to other modeling based approaches and takes covariance structure into account while statistically comparing the data. However, its practical utility is limited because it may be difficult to analyze the data using conventional statistical packages [180].

***Baker–Lonsdale Model*** This has been developed from the Higuchi model:

$$f_t = \frac{3}{2}\left[1-\left(1-\frac{M_t}{M_\infty}\right)^{2/3}\right]-\frac{M_t}{M_\infty} = Kt$$

where $K$ is the release rate constant (slope), and $M_t$; and $M_\infty$ are the amounts of drug released, respectively, at time $t$ and infinite period. This model has been utilized for linearizing release data from several formulations of microcapsules or microspheres [181–183].

The Hopfenberg model utilizes the following equation to describe the dissolution data:

$$\frac{M_t}{M_\infty} = 1-\left(1-\frac{K_0 t}{c_0 a_0}\right)^n$$

where $M_t$ is the amount of drug dissolved in time $t$; $M_\infty$ is the total amount of drug dissolved when dosage form is exhausted; $M_t/M_\infty$ is the fraction of dose dissolved; $K_0$ is the erosion rate constant; $C_0$ is the initial drug concentration in matrix; and $A_0$ is the initial radius of spheres/cylinder/half-thickness of the slab, where on $A$ value of 1, 2, or 3 is used for a slab, a cylinder, or a sphere, respectively. This model is generally useful while studying release from surface eroding devices (slab, cylinder, sphere) [184].

***Higuchi Model*** Higuchi developed several theoretical models to study the release of water-soluble and poorly soluble drugs incorporated in semisolid and/or solid matrixes. Generally, a simplified equation is used for analyzing dissolution data:

$$Q_t = K_{\mathrm{H}}\mathrm{t}$$

where $K_{\mathrm{H}}$ is the Higuchi dissolution constant. This model is widely used to describe drug dissolution from modified release dosage forms [185–187].

***Hixson–Crowell Model*** This has been used to describe the release profile by considering the diminishing surface of the drug particles during dissolution:

$$Q_0^{1/3} - Q_t^{1/3} = K_s t$$

where $K_s$ is the constant incorporating the surface–volume relationship; $Q_0$ and $Q_t$ are, respectively, the initial amount of drug and the remaining amount of drug in the dosage form at time $t$. Upon simplification, the equation can be written

$$(1-f_t)^{1/3} = 1-K_\beta t$$

where $f_t = 1-(Q_t/Q_0)$, and $f_t$ represents the rug-dissolved fraction at time $t$ and $K_\beta$ is the release constant. This model is applied to dosage forms (e.g., tablets) where the dissolution occurs in planes that are parallel to the drug surface—that is, if tablet

dimensions diminish proportionally in such a way that the geometrical form remains constant all the time. The major assumption made is that the release rate is limited by the drug particles' dissolution rate and not by diffusion through the polymeric matrix [188].

***Weibull Function*** This is a modeling based method used for evaluating the dissolution data profiles:

$$m = 1 - \exp\left(\frac{-(t - Ti)^b}{a}\right)$$

where $a$ defines the time scale of the process, $T_i$ is the lag time before the onset of the dissolution or release process (in most cases zero), $b$ is the shape parameter, $m$ is the accumulated fraction of drug, and $t$ is the solution time of the drug.

Upon rearranging, this equation becomes

$$\log[-\ln(1-m)] = b\log(t - T_i) - \log a$$

On the basis of this equation, the plot between log of dissolved amount of drug versus log of time will be linear for dosage forms. The main limitation of this model is the lack of a kinetic basis for its use due to the nonphysical nature of the parameters. In addition, it cannot characterize the dissolution kinetic properties of a drug. Furthermore, there is no single parameter related to the intrinsic dissolution rate of the drug. Lastly, it is of limited use for establishing *in vitro–in vivo* (IVIV) correlation [189–191].

***Order of Drug Release Analysis*** Zero order and first order kinetics and ratio test procedures are modeling based methods. Drug dissolution from dosage forms that do not disaggregate and release the drug slowly is characterized by zero order kinetics (assuming that area doesn't change):

$$Q_t = Q_0 + K_0 t$$

where $Q_t$ is the amount of drug dissolved in time $t$, $Q_0$ is the initial amount of drug in solution, and $K_0$ is the zero order release rate constant. This method can be used to describe drug dissolution of modified release dosage forms [192, 193].

First order drug release can be represented by

$$\ln Q_t = \ln Q_0 + K_1 t$$

where $K_1$ is the first order release constant. This equation is modified to yield the following equation:

$$\ln Q_t = \ln Q_0 + K_1 t S$$

where $S$ is the solid area accessible to dissolution [192].

***Ratio Test Procedures*** These compare dissolution profiles of two formulations at a particular time point. Ratio tests of percentage dissolved can be performed to give a 90% confidence interval for the mean ratio of percentage dissolved [194].

## 5 *IN VITRO–IN VIVO* CORRELATION

Ethical constraints make it imperative to generate pharmacological and toxicological data in laboratory animals and *in vitro* systems before human testing can begin. Two fundamental challenges faced by pharmaceutical scientists pertain to scaling up of pharmacokinetic data from animals to humans and extrapolation of the *in vitro* data to the *in vivo* situation.

### 5.1 Physiological Concerns

Drug absorption is influenced by physicochemical characteristics of the drug as well as physiological factors. While physicochemical factors are species independent, physiological factors are species dependent. Therefore, it is essential to examine the bioavailability of a drug in animals with respect to the physiological differences with human beings. Often, the knowledge of different absorption or first-pass metabolism in animals helps in reasonably extrapolating the absorption of drugs to that in human beings (Table 9). Furthermore, differences in the gastrointestinal physiology of monkeys and dogs with respect to human beings play a significant role in addressing the differences observed in dosage form performance (Table 10). It is evident from Table 11 that solution is emptied from the stomach within 1 h and it is only slightly influenced by feeding. $t_{lag}$, $t_{max}$, and MAT of tablet were much longer than those of solution. The gastric emptying of tablets in the fed state is delayed more than the solution dosage forms. It is noteworthy that agitation intensity in the gastrointestinal tract of monkeys is closer to that in humans than that in dogs. This leads to approximately similar gastric emptying time of solid dosage forms in monkeys and humans [207].

The rate of distribution of a drug to the organs or tissues is determined by the rate of blood flow and ease with which the drug molecules cross the capillary wall and penetrate the cells. Like most physiological parameters, blood flow and circulation time can be extrapolated across species by use of an allometric equation [210, 211]:

$$y = aw^b$$

where $y$ is the blood flow, $a$ is the allometric coefficient, $b$ is the allometric exponent, and $w$ is the body weight. The allometric relationship between blood circulation time (seconds) and total body weight is $21\,w^{0.21}$ [212].

The smaller animal species deliver drugs faster and more frequently to organs of elimination. This is expected to result in rapid elimination of drugs in smaller animals than in humans [210, 211]. Furthermore, differences in amino acid sequences and the presence of different isoforms and enzymes in different species may lead to significantly different patterns of drug metabolism (Table 12). Drugs and their metabolites are usually eliminated from the body via urine and bile. The relative contribution of biliary and urinary excretion to overall elimination of drugs depends

**TABLE 9 A Few Examples of Interspecies Differences in Drug Bioavailability**

| Drug | Bioavailability | | | | Remarks | References |
|---|---|---|---|---|---|---|
| | Dog | Rat | Monkey | Human | | |
| Indinavir (solution) | 72% | 24% | 19% | 40–60% | Decreased bioavailability in rat and monkey due to extensive hepatic first-pass metabolism. | 195 |
| L-365,260: cholecystokinin receptor antagonist (suspension) | 9% | 14% | — | — | Limited bioavailability was attributed to poor absorption as a result of low aqueous solubility. | 196 |
| L-365,260: cholecystokinin receptor antagonist (solution in PEG 600) | 70% | 50% | — | — | Improved bioavailability due to greater amount of drug soluble in PEG 6000. | 196 |
| Melatonin (oral dose 10 mg/kg)<br>Intraperitoneal dose | 100%<br>74% | 53.5% | 100% | — | Suggested lack of substantial hepatic first-pass of melatonin in rats. However, the oral bioavailability of melatonin in dogs decreased to 16.9% following a 1 mg/kg oral dose, indicating dose-dependent bioavailability in dogs. | 197 |

on the nature of the drugs and is significantly affected for molecules possessing molecular weight greater than 7500 kDa [220]. Many drugs are excreted unchanged by the kidneys. The rate of renal excretion depends on renal blood flow, glomerular filteration rate (GFR), tubular secretion, and reabsorption [201]. Both GFR and number of nephrons show a good allometric relationship. Like GFR, renal excretion of drugs also shows a good allometric relationship across species. Thus, the renal clearance of drugs in humans can be extrapolated from animal data by use of an allometric approach. But this approach requires at least four or five animal species in order to obtain a proper allometric relationship. An alternative to predict human renal clearance is to use the ratio of GFR between rats and humans. The ratio of renal clearance of various drugs between rats and humans is roughly equal to the ratio of GFR between these two species [221].

**TABLE 10 Summary of Physiological Parameters that May Influence Drug Bioavailability in Different Species**

| Species | Biliary Excretion Rate (of Organic Chemicals) | Plasma Volume (L/kg) | Body Weight (kg) | Maximum Life Period (years) | Relative Length of Small Intestine (%) | Average Absolute Length of Small Intestine (m) | pH of Small Intestine | Transit Time | Glomerular Filteration Rate (mL/min/kg) | References |
|---|---|---|---|---|---|---|---|---|---|---|
| Rabbit | Relatively poor | 0.0314 | 3.0 | 8.0 | 60 | 1.51 | 6–8 | *a* | 4.8 | 198–200 |
| Pig | *a* | *a* | *a* | *a* | 78 | 14.16–18.29 | 6–7.5 (fasted) | *a* | *a* | 198, 199 |
| Mouse | Good | *a* | 0.02 | 2.7 | *a* | *a* | *a* | *a* | 10 | 201 |
| Monkey | Relatively poor | 0.0448 | 3.5 | 20 | *a* | *a* | 5.6–9 (fasted) | *a* | *a* | 199, 201, 202 |
| Dog | Good | 0.0515 | 12.5 | 20 | 85 | 2.48–4.14 | 6.2–7.5, 4.5–7.5 (fasted) | 0.5–2 h | 4.0 | 198, 199, 203, 204 |
| Rat | Good | 0.0313 | 0.25 | 4.7 | 64 | 0.82 | 6.5–7.1 | 1.5 h | 8.7 | 198, 203, 205 |
| Human | Relatively poor | 0.0436 | 70 | 93 | 79 | 6.25 | 6.8–8.6, 5–7 (fasted) | 2.7–8.5 h | 1.8 | 198, 201, 202, 205, 206 |

[a]Not found.

**TABLE 11 Gastric Emptying Time of Dosage Forms Under Fasted and Fed State in Animals and Humans**

| Type of Dosage Form | Parameter | Human | | Monkeys | | Dog | | References |
|---|---|---|---|---|---|---|---|---|
| | | Fasting | Fed | Fasting | Fed | Fasting | Fed | |
| Solution | MAT (h) | 1.3 ± 0.3 | 2.6 ± 0.2 | 0.8 ± 0.4 | 1.8 ± 1.4 | 0.7 ± 0.4 | 1.3 ± 0.6 | 207, 208 |
| | *t*lag (h) | 0.1 ± 0.1 | 0.4 ± 0.1 | 0.1 ± 0.0 | 0.0 ± 0.0 | 0.2 ± 0.2 | 0.3 ± 0.1 | |
| | *t*max (h) | 1.0 | 3.0 | 1.0 ± 0.2 | 1.8 ± 1.3 | 1.0 | 1.0 | |
| Controlled release tablet | MAT (h) | 3.4 ± 0.5 | 6.0 ± 1.6 | 3.6 ± 2.0 | Plasma concentration rose rapidly, *t*lag was delayed | 2.1 ± 0.6 | Not calculated, delayed compared to human | 207, 208 |
| | *t*lag (h) | 1.7 ± 0.6 | 3.3 ± 0.8 | 2.3 ± 1.5 | | 1.2 ± 0.5 | | |
| | *t*max (h) | 2.0 | 6.0 | 3.7 ± 2.1 | | 3.0 | | |
| Coumarin (solution) | *t*max (h) | 0.2 | — | 0.21 ± 0.03 | — | — | — | 209 |
| Rolipram (suspension) | *t*max (h) | 0.38 ± 0.22 | — | 1.7 ± 0.9 | — | — | — | 209 |
| Zomepira (solution) | *t*max (h) | 0.6 ± 0.3 | — | 1.2 ± 0.8 | — | — | — | 209 |

**TABLE 12 Summary of Drug Metabolizing Enzymes Present in Liver of Different Species**

| Species | Gender/Strain | Liver Weight | Enzyme | References |
|---|---|---|---|---|
| Dog | Mixed/beagle | 360 g (sd = 20) | Cytochrome P450 (CYP) CYP1A1/1A2; ethoxyresorufin O-deethylation<br>CYP1A2; phenacetin O-deethylation<br>CYP2A6; coumarin 7-hydroxylation<br>CYP2B6; pentoxyresorufin O-dealkylation<br>CYP2C9; phenytoin p-hydroxylation<br>CYP3A4; human 3A4 antibodies<br>CYP2E1; human 2E1 antibodies<br>CYP2C9; human 2C9 antibodies<br>CYP3A/2B; ethylmorphine N-demethylation | 213–216 |
| Rat | Female and male/Sprague–Dawley | 10.6 g (sd = 3) | Cytochrome P450 (CYP)<br>CYP3A/2B; ethylmorphine N-demethylation<br>CYP3A4/5; nifedipine oxidation<br>CYP3A4/5; erythromycin N-demethylation<br>CYP3A4; benzphetamine N-demethylation<br>CYP2E1; aniline p-hydroxylation<br>CYP2D6; bufuralol 1-hydroxylation<br>CYP2C19; S-mephenytoin 4′-hydroxylation<br>CYP2C9; phenytoin p-hydroxylation<br>CYP2B6; pentoxyresorufin O-dealkylation<br>CYP2A6; coumarin 7-hydroxylation<br>CYP1A2; phenacetin O-deethylation | 216 |

| | | | | |
|---|---|---|---|---|
| Mouse | Mixed/ICR | Not found | GST; chlorodinitrobenzene glutathione transferase | 217 |
| Human | Mixed | 1700 g | Cytochrome P450 (CYP) t | 213–216, 218 |
| | | | CYP1A1/1A2; ethoxyresorufin O-deethylation | |
| | | | CYP1A2; phenacetin O-deethylation | |
| | | | CYP2A6; coumarin 7-hydroxylation | |
| | | | CYP2B6; pentoxyresorufin O-dealkylation | |
| | | | CYP2C8/2C9; tolbutamide hydroxylation | |
| | | | CYP2C9; phenytoin p-hydroxylation | |
| | | | CYP2C19; S-mephenytoin 4′-hydroxylation | |
| | | | CYP2D6; bufuralol 1′-hydroxylation | |
| | | | CYP2E1; aniline p-hydroxylation | |
| | | | CYP3A; chlorpromazine S-oxygenation | |
| | | | CYP3A4; benzphetamine N-demethylation | |
| | | | CYP3A4/5; nifedipine oxidation | |
| | | | CYP3A/2B; ethylmorphine N-demethylation | |
| Rabbit | Mixed/New Zealand white | Not found | GST; chlorodinitrobenzene glutathione transferase | 217, 219 |
| | | | GST; dichloronitrobenzene glutathione transferase | |
| | | | GST; ethacrynic acid glutathione transferase | |
| Monkey | Male/ cynomolgus | 62.5 g | Cytochrome P450 (CYP) content | 213 |
| | | | CYP1A2; phenacetin O-deethylation | |
| | | | CYP2A6; coumarin 7-hydroxylation | |
| | | | CYP2D6; dextromethorphan O-demethylation | |
| | | | CYP2E1; chlozoxazone 6-hydroxylation | |
| | | | CYP2E1; ethoxycoumarin O-deethylation | |

## 5.2 Correlation

The development and subsequent validation of an *in vitro–in vivo* correlation (IVIVC) is an increasingly important component of dosage form optimization. An IVIVC is a relationship (preferably linear) between a biological parameter ($C_{max}$, $t_{max}$, or AUC) produced by a dosage form and an *in vitro* characteristic (e.g., *in vitro* dissolution data). The IVIVC guidance developed by the U.S. FDA states that the main objective of developing and evaluating an IVIVC is to enable the dissolution test to serve as a surrogate for *in vivo* bioavailability studies. An IVIVC has been defined as "a predictive mathematical model describing the relationship between an *in vitro* property of a dosage form and an *in vivo* response." Generally, rate or extent of drug dissolution is the *in vitro* property while plasma drug concentration/amount of drug absorbed is the *in vivo* property.

The development of correlation usually involves three steps: (1) formulation development with different release rates, (2) obtaining *in vitro* dissolution profiles and *in vivo* plasma concentration profiles for these formulations, and (3) estimation of *in vivo* absorption or *in vitro* dissolution time course using an appropriate technique for each formulation.

*Level A* correlations are the most common type of correlation submitted to the FDA. It is usually estimated by a two-stage procedure comprising deconvolution followed by comparison of the fraction of drug absorbed to the fraction of drug dissolved. A correlation of this type is generally linear and represents a point-to-point relationship between the *in vitro* and *in vivo* input rate. In a linear correlation, the *in vitro* and *in vivo* data curves are superimposable or may be made superimposable with the help of a scaling factor. One of the alternative approaches to develop Level A correlation is a convolution procedure that models the relationship between *in vitro* dissolution and plasma concentration in a single step.

*Level B* correlation uses the principles of statistical moment analysis. The mean *in vitro* dissolution time is compared to either the mean residence time or the mean *in vivo* dissolution time. It also uses all of the *in vitro* and *in vivo* data, but it is not considered to be a point-to-point correlation. Level B correlation does not reflect the actual *in vivo* plasma level curve.

*Level C* correlation establishes a single point relationship between a dissolution parameter like $t_{50}$ and a pharmacokinetic parameter like AUC. It does not reflect the complete shape of the plasma concentration time curve, which is an essential factor.

*Multiple Level C* correlates one or several pharmacokinetic parameters ofinterest to the amount of drug dissolved at several time points of the dissolution profile [222, 223].

The USP recognizes these levels of IVIVC. However, the applicability of IVIVC depends on the inherent permeability of the molecules across the gastrointestinal membrane. Table 13 summarizes the possibility of IVIVC for immediate as well as extended release oral products on the basis of the biopharmaceutical classification scheme (BCS).

A successful IVIVC utilizes (1) *in vitro* values to evaluate different batches of a pharmaceutical product as a quality control check to ensure the desired physiologic performance, and (2) *in vitro* values as a tool for developing a series of dosage forms to obtain a desired *in vivo* performance. These objectives can only be fulfilled if a

**TABLE 13 Possibility of IVIVC of IR and ER Dosage Forms According to BCS** [223]

| Type of Formulation | Class | Solubility | Permeability | IVIVC (Possible/ Not Possible) |
|---|---|---|---|---|
| Immediate release (IR) | I | High | High | May not be possible |
| | II | Low | High | May not be possible |
| | III | High | Low | May not be possible |
| | IV | Low | Low | May be possible |
| Extended release (ER) | Ia | High + site dependent | High + site dependent | Level A |
| | Ib | High + site dependent | Site dependent with narrow absorption window | Level C |
| | IIa | Low + site dependent | High + site dependent | Level A |
| | IIb | Low + site dependent | Site dependent with narrow absorption window | May not be possible |
| | Va | Variable | Variable | May not be possible |
| | Vb | Variable | Variable | Level A |

relationship between *in vitro* and *in vivo* parameters is confirmed. The approaches for IVIVC are summarized in Fig. 5.

IVIVCs can be divided into nonquantitative and quantitative correlations. In the case of nonquantitative correlations, the two variables are not related to each other mathematically as observed in rank order correlation (not popular now) [224–226]. In the case of quantitative correlations, the variables correlate with each other through a linear or nonlinear equation. A quantitative correlation can be established by using well known estimated values of important *in vitro* dissolution processes and *in vivo* parameters; for instance, amount dissolved at a specific time point can be correlated with fraction absorbed or absorption rate constant. These correlations can be categorized as single point correlations/or multiple point correlations.

Single point correlations are not very informative and sometimes they are not easy to apply practically [227, 228]. On the other hand, point-to-point or multiple point correlations are informative and are established in two ways. In the first case, a relationship is established between the actual time course of the *in vitro* dissolution and the time course of dissolution in the lumen or arrival in the general circulation, which is estimated by using a deconvolution method (observed concentration in the bloodstream versus time profile). In the second case, the relationship between actual plasma concentrations versus time profile is evaluated. In both cases, there is a requirement of intravenous or oral solution data or in the case of extended release product of high solubility, oral data from immediate release solid dosage form is required. In the case of multiple point correlation, *in vitro* dissolution

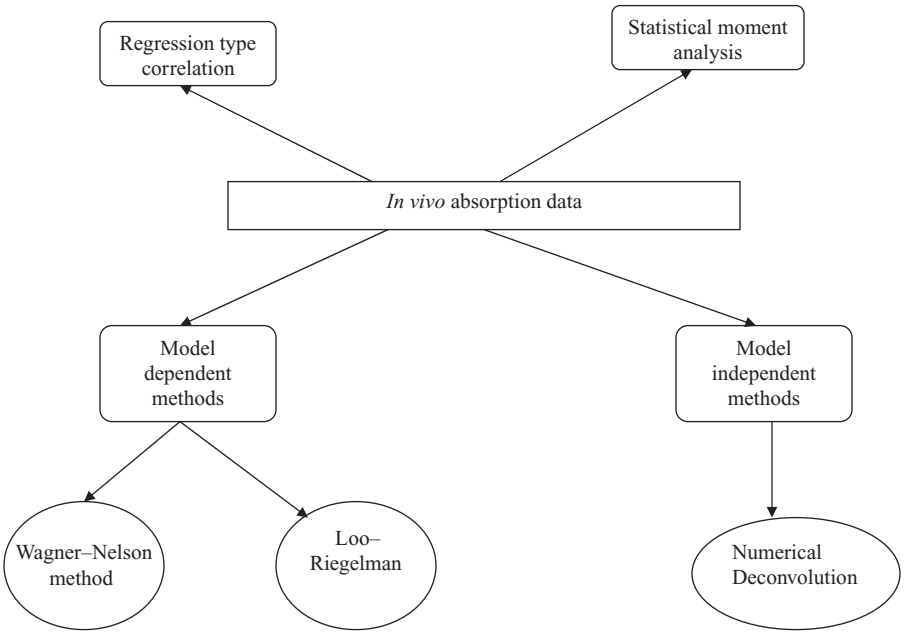


**FIGURE 5** Approaches for *in vivo* data analysis.

datasets must be compared with each other at the time of evaluation as well as application [222, 229]. In addition to the well recognized methods of IVIVC, a few useful methods are summarized in Table 14.

### 5.3 Biowaivers for Changes in the Manufacture of a Drug Product

1. *Biowaivers Without an IVIVC.* For formulations consisting of beads in capsules, with the only difference between strengths being the number of beads, approval of lower strengths without an IVIVC is possible, provided bioavailability data are available for the highest strength.

2. *Biowaivers Using an IVIVC.* Nonnarrow therapeutic Index Drugs: If an IVIVC is developed with the highest strength, any lower strengths may be granted approval if these strengths are proportional or qualitatively the same. Furthermore, the *in vitro* dissolution profiles of all strengths should be similar, and all strengths should have the same release mechanisms. In addition, changes in release controlling excipients in the formulations should be within the range of release controlling excipients of established correlations.

3. *Biowaivers Using an IVIVC.* Narrow therapeutic Index Drugs: If external predictability of IVIVC is established, following waivers are likely to be granted, if the release rate of at least two formulations have been studied for IVIVC development. If an IVIVC is developed with the highest strength, any lower strength may be granted approval if the strengths are proportional or qualitatively the same. Furthermore, the *in vitro* dissolution profiles of all strengths should be similar and

**TABLE 14 A Few Methods of IVIVC with Respect to Time Profile**

| Correlation Method | Remarks | References |
|---|---|---|
| **Characterization of time profile** | | |
| Distribution functions (e.g., Weibull function) | Represented by *in vitro* and *in vivo* time profiles, for example, (a) *in vitro* release profile corresponds to distribution of drug released at time $t$ and the corresponding probability distribution function profile characterizes the release rate; (b) plasma concentration profile represents drug distribution in plasma at any time and its cumulative distribution function represents the drug absorbed and already eliminated. | 230–232 |
| Semi-invariants (moments) | Area ($K_0$), mean ($K_1$), variance ($K_2$), skewness ($K_3$), and kurtosis ($K_4$) are semi-invariants. All semi-invariants are defined in terms of integrals of time profile between $t = 0$ and $t = \infty$. Variance, skewness, and kurtosis are called higher rate semi-invariants. | 230–232 |
| **Comparison of time profiles** | | |
| General aspects | Model-independent techniques compare data pairs observed at corresponding time values, | 230–232 |
| Model-dependent/-independent comparison | where time is only a class effect (e.g., paired $t$-test, ANOVA). Model-dependent techniques are better as they assume that observed data pairs belong to a general distribution function. | |
| Horizontal/vertical comparison | Vertical comparison represents the value obtained at a given time (i.e., extent characteristics of the process) and horizontal comparison represents the time required to reach a certain ordinate value (i.e., rate aspect of process). | 230–232 |
| Comparison by semi-invariants | For the comparison of two differential plasma concentration profiles the index of bioequivalence is used (discussed in data analysis). | 230–232 |
| Rescigno's indices ($\xi_1$ and $\xi_2$) | | |
| Model-independent indices | | 230–232 |
| Moore–flanner index ($f_1$) | $f_1 = \frac{\sum \lvert R - T \rvert}{\sum R}$ | |
| Moore–flanner index ($f_2$) | $f_2 = 50 \log \left\{ \frac{100}{\left(\sqrt{1 + \sum (R^{\wedge} - T^{\wedge})2}\right)/N} \right\}$ | |
| Statistical consideration | Statistical comparisons are performed in terms of dissimilarity rather than similarity. If computed statistics or indices exceed a predefined decision limit, both samples are considered as different. | 230–232 |
| Decision intervals and limits | The Chow–Ki model, multivariate aspects, and bootstrap techniques are used. | |

all strengths should have the same release mechanism. Biowaiver for the approval of new strengths is applicable to strengths lower than the highest strength, within the dosing range that has been established to be safe and effective, provided that the new strengths are proportional or qualitatively the same, have the same release mechanisms, have similar *in vitro* dissolution profiles, and are manufactured using the same type of equipment and the same process at the same site as other strengths that have bioavailability data available. The generic products can also qualify for this biowaiver (a) if bioequivalence has been established for all strengths of reference listed products, (b) if dose proportionality has been established for reference listed products, and (c) if bioequivalence is established between the generic product and reference listed products and all strengths are compositionally proportional or qualitatively the same and have a similar release mechanism as well as *in vitro* dissolution profiles.

4. *Biowaivers When in Vitro Dissolution Is Independent of Dissolution Test Conditions.* Narrow and nonnarrow therapeutic Index Drugs: Biowaivers may be granted if dissolution data are submitted in compendial medium and in three other media (water, 0.1 N HCl, USP buffer pH 6.8). In addition, if *in vitro* dissolution should be shown to be independent of dissolution test conditions after change is made in drug product manufacturing, biowaivers may be granted.

5. *Situations for Which IVIVC Is Not Recommended.* IVIVC is not recommended (a) for approval of a new formulation (extended release) of an approved drug product when the new formulation has a different release mechanism, (b) for approval of a dosage strength higher/lower than doses that have been shown to be safe and effective in clinical trials, (c) for approval of another sponsor's product, even with the same release controlling mechanism, and (d) for approval of a formulation change involving a nonrelease controlling excipient in the drug product that may significantly affect drug absorption [222, 233, 234].

## REFERENCES

1. Czeisler JL, Perlman KP. *Encyclopedia of Pharmaceutical Technology*, Volume 4. New York: Marcel Dekker; 1991, pp 40–52.
2. Scolk KGV, Zogiio M, Carstensen JT. *Pharmaceutical Dosage Forms: Tablets*, Volume I. New York: Marcel Dekker; 1980, pp 73–105.
3. Shah MA, Wilson RG. Some effects of humidity and heat on the tableting properties of microcrystalline cellulose formulations. I. *J Pharm Sci* 1968;57:181–182.
4. Brownley CA, Lachman L. Browning of spray-processed lactose. *J Pharm Sci* 1964;53:452–454.
5. Van Kamp HV, Bolhuis GK, Kussendrager KD, Lerk CF. Studies on tableting properties of lactose. IV. Dissolution and disintegration properties of different types of crystalline lactose. *Int J Pharm* 1986;28:229–238.
6. Shangraw RF. Direct compression tableting. In *Encyclopedia of Pharmaceutical Technology*, Volume 4 2nd ed. New York: Marcel Dekker; 1980, pp 85–160.
7. Lerk CF. Consolidation and compaction of lactose. *Drug Dev Ind Pharm* 1993;19:2359–2398.

8. Roberts RJ, Rowe RC. Effect of punch velocity on the compaction of a variety of materials. *J Pharm Pharmacol* 1985;37:377–384.
9. Fell JT, Newton JM. The tensile strength of tablets. *J Pharm Sci* 1968;20:657–658.
10. Gohel MC, Patel LD, Amin AF, Jogani PD, Bajaj SB, Patel GJ. Studies in improvement of flow and compressional characteristics of lactose. *Int J Pharm Excipients* 1999;1: 86–92.
11. Gohel MC, Jogani PD. Functionality testing of a multifunctional directly compressible adjuvant containing lactose, polyvinyl pyrrolidone, and croscarmellose sodium. *Pharm Technol* 2002;25:64–82.
12. Wong DY, Wright P, Aulton ME. Deformation of alpha-lactose monohydrate and anhydrous alpha-lactose monocrystals. *Drug Dev Ind Pharm* 1988;14:2109–2126.
13. Shukla AJ, Price JC. Effect of moisture content on compression properties of directly compressible high beta-content anhydrous lactose. *Drug Dev Ind Pharm* 1991; 17:2067–2081.
14. Alpar O, Hersey JA, Shotton E. The compression properties of lactose. *J Pharm Pharmacol* 1970;22:1S–7S.
15. Bolhuis GK, Lerk CF. Comparative evaluation of excipients for direct compression. Part 1. *Pharma Weekb* 1973;108:469–481.
16. Vromans H, Bolhuis GK, Lerk CF, Kussendrager KD. Studies on tableting properties of lactose. VIII. The effect of variations in primary particle size, percentage of amorphous lactose and addition of a disintegrant on the disintegration of spray-dried lactose tablets. *Int J Pharm* 1987;39:201–206.
17. Bolhuis GK, Zuurman K. Tableting properties of experimental and commercially available lactose granulations for direct compression. *Drug Dev Ind Pharm* 1995;21: 2057–2071.
18. Bavitz JF, Schwartz JB. Direct compression vehicles, part II. *Drug Cosmetic Ind* 1976;118(4):60–77.
19. Hwang R, Peck GR. A systematic evaluation of the compression and tablets characteristics of various types of microcrystalline cellulose. *Pharm Technol* 2001;24:112–132.
20. Obae K, Iijima H, Imada K. Morphological effect of microcrystalline cellulose particles on tablet tensile strength. *Int J Pharm* 1999;182:155–164.
21. Lahdenpaa E, Niskanen M, Yliruusi J. Crushing strength, disintegration time and weight variation of tablets compressed from three Avicel pH grades and their mixtures. *Eur J Pharm Biopharm* 1997;43:315–322.
22. Reus-Medina M, Lanz M, Kumar V, Leuenberger H. Comparative evaluation of the powder properties and compression behaviour of a new cellulose based direct compression excipient and Avicel PH 102. *J Pharm Pharmacol* 2004;56:951–956.
23. Bolhuis GK, Chowhan ZT. *Materials for direct Compression, Pharmaceutical Powder Compaction Technology*, Volume 7. New York: Marcel Dekker; 1996, pp 419–499.
24. Shangraw RF. Direct compression tableting. In *Encyclopedia of Pharmaceutical Technology*, Volume 4, 2nd ed. New York: Marcel Dekker; 1988, pp 85–160.
25. Rizzuto AB, Chen AC, Veiga ME. Modification of the sucrose crystal structure to enhance pharmaceutical properties of excipient and drug substances. *Pharm Technol* 1984;8:32–39.
26. Olmo IG, Ghaly ES. Evaluation of two dextrose-based directly compressible excipients. *Drug Dev Ind Pharm* 1998;24:771–778.
27. Papadimitriou E, Efentakis M, Choulis NH. Evaluation of maltodextrins as excipients for direct compression and tablets and their influence on the rate of dissolution. *Int J Pharm* 1992;86:131–136.

28. Levy G. Effect of certain tablet formulation factors on dissolution rate of the active ingredient I. *J Pharm Sci* 1963;52:1039–1046.
29. Manudhane KS, Contractor AM, Kim HY, Shangraw RF. Tableting properties of a directly compressible starch. *J Pharm Sci* 1969;61:616–620.
30. Underwood TW, Cadwallader DE. Influence of various starches on dissolution rate of salicylic acid from tablets. *Pharm Sci* 1972;58:239–243.
31. Hsu SH, Tsai TR, Chuo WH, Cham TM. Evaluation of era-tab as direct compression excipients. *Drug Dev Ind Pharm* 1997;23:711–716.
32. Heinze F. Starch—the natural excipient choice. *Manufacturing Chemist* 2002;73:40–42.
33. Rees JE, Rue PJ. Time dependent deformation of some direct compression excipients. *J Pharm* 1978;30:601–607.
34. Khan KA, Rhodes, CT. Effect of variation in compaction force on properties of six direct compression tablet formulation. *J Pharm Sci* 1976;65:1835–1837.
35. Gohel MC. A review of co-processed directly compressible excipients. *J Pharm Pharm Sci* 2005;8:76–93.
36. Whiteman M, Yarwood RJ. Evaluation of six lactose-based materials as direct compression tablet excipients. *Drug Dev Ind Pharm* 1988;14:1023–1040.
37. Plaizier-Vercammen JA, Van Den Bossche H. Evaluation of the tableting properties of a new excipient for direct compression. *Drugs Made in Germany* 1993;36:133–137.
38. Schmidt PC, Rubensdorfer CJ. Evaluation of ludipress as a multipurpose excipient for direct compression. Part 1. Powder characteristics and tableting properties. *Drug Dev Ind Pharm* 1994;20:2899–2925.
39. Schmidt PC, Rubensdorfer CJ. Evaluation of ludipress as a multipurpose excipient for direct compression. Part II: Inactive blending and tableting with micronized glibenclamide. *Drug Dev Ind Pharm* 1994;20:2927–2952.
40. Garr JS, Rubinstein MH. Compaction properties of a cellulose-lactose direct compression excipient. *Pharm Tech Int* 1991;3:24–27.
41. Armstrong NA, Roscheisen G, Al-Aghbar MR. Cellactose as a tablet diluent. *Manufacturing Chemist* 1996;67:25–26.
42. Reimerdes D. The near future of tablet excipients. *Manufacturing Chemist* 1993; 64:14–15.
43. Belda PM, Mielck JB. The tabletting behaviour of cellactose compared with mixtures of celluloses with lactoses. *Eur J Pharm Biopharm* 1996;42:325–330.
44. Casalderrey M, Souto C, Concheiro A, Gomea-Amoza JL, Martinez-Pacheco RA. Comparison of cellactose with two ad hoc processed lactose–cellulose blends as direct compression excipients. *Chem Pharm Bull* 2000;48: 458–463.
45. Gohel MC, Modi CJ, Jogani PD. Functionality testing of a coprocessed diluent containing lactose and microcrystalline cellulose. *Pharm Technol* 1999;22:40–46.
46. Gohel MC, Bariya S, Jogani PD. Investigation in direct compression characteristics of coprocessed adjuvant containing lactose, microcrystalline cellulose and dicalcium phosphate. *Pharm Dev Technol* 2003;8:143–152.
47. Meggle AG. Technical information at www.megglepharma.de/en/product/uebersich/starlac.
48. Hauschild K, Picker KM. Evaluation of new coprocessed compound based on lactose and maize starch for tablet formulation. *AAPS PharmSci* 2004;6:1–12.
49. Tsai T, Wu J, Ho H, Sheu M. Modification of physical characteristics of microcrystalline cellulose by codrying with β-cyclodextrin. *J Pharm Sci* 1998;87:117–122.

50. Levis SR, Deasy PB. Pharmaceutical applications of size reduced grades of surfactant co-processed microcrystalline cellulose. *Int J Pharm* 2001;230:25–33.
51. Ishikawa T, Mukai B, Shiraishi S, Utoguchi N, Fujii M, Matsumoto M, Watanaba Y. Preparation of rapidly disintegrating tablet using new types of microcrystalline cellulose (PH-M Series) and low substituted-hydroxypropylcellulose or spherical sugar granules by direct compression method. *Chem Pharm Bull* 2001;49:134–139.
52. Lourdes Garzo'n M, Villafuerte L. Compactibility of mixtures of calcium carbonate and microcrystalline cellulose. *Int J Pharm* 2002;231:33–41.
53. Luukkonen P, Schaefer T, Hellen L, Juppo AM, Yliruusi J. Rheological characterization of microcrystalline cellulose and silicified microcrystalline cellulose wet masses using a mixer torque rheometer. *Int J Pharm* 1999;188:181–192.
54. Allen JD. Improving DC with SMCC. *Manufacturing Chemist* 1996:67:19–23.
55. Kachimanis K, Nikolakakis I, Malamataris S. Tensile strength and disintegration of tableted silicified microcrystalline cellulose: influences of interparticle bonding. *J Pharm Sci* 2003;92:1489–1501.
56. Wei S. Preparation and physical/mechanical evaluation of new crystallinity forms of cellulose as pharmaceutical excipients. Minneapolis, MN: University of Minnesota; 1991.
57. Banker GS, Wei S. Low crystallinity cellulose. U.S. patent 5417984. 23 May 1995.
58. Wei S, Kumar V, Banker GS. Phosphoric acid mediated depolymerization and decrystallinization of cellulose. Preparation of low crystallinity cellulose—a new pharmaceutical excipient. *Int J Pharm* 1996;142:175–191.
59. Kumar V, Kothari SH, Banker GS. Effect of agitation rate on the generation of low crystallinity cellulose from phosphoric acid. *J Appl Polym Sci* 2001;82:2624–2628.
60. Jones TM. *Symposium On Powders*. Dublin: Society of Cosmetic Chemists of Great Britain; 1969.
61. Peleg M, Mannheim CH. Effect of conditioners on the flow properties of powdered sucrose. *Powder Technol* 1973;7:45–50.
62. Matsuda Y, Minamida Y, Hayashi S. Comparative evaluation of tablet lubricants: effect of application method on tablet hardness and ejectability after compression. *J Pharm Sci* 1976;65:1155–1160.
63. Iranlove TA, Parrot EL. Effects of compression force, particle size, and lubricants on dissolution rate. *J Pharm Sci* 1978;67:535–539.
64. Chowhan ZT, Chi LH. Drug–excipient interactions resulting from powder mixing. IV: Role of lubricants and their effect on *in vitro* dissolution. *J Pharm Sci* 1986; 75:542–545.
65. Westerberg M, Nystrom C. Physicochemical aspects of drug release. XII. The effect of some carrier particle properties and lubricant admixture on drug dissolution from tableted ordered mixtures. *Int J Pharm* 1991;69:129–141.
66. Ong JTH, Chowhan ZT, Samuels JG. Drug–excipient interactions resulting from powder mixing. VI. Role of various surfactants. *Int J Pharm* 1993;96:231–242.
67. Desai DS, Rubitski BA, Varia SA, Newman AW. Physical interactions of magnesium stearate with starch derived disintegrants and their effects on capsule and tablet dissolution. *Int J Pharm* 1993;91:217–226.
68. Desai DS, Rubitski BA, Bergum JS, Varia SA. Effect of various fomulation factors on dissolution stability of aztreonam, hydrochlorothiazide, and sorivudine capsules. *Drug Dev Ind Pharm* 1994;110:249–255.

69. Shah AC, Mlodozeniec AR. Mechanism of surface lubrication: influence of duration of lubricant excipient mixing on processing characteristics of powders and properties of compressed tablets. *J Pharm Sci* 1997;66:1377–1382.
70. Chowhan ZT, Chi LH. Drug–excipient interactions resulting from powder mixing. V: Role of sodium lauryl sulphate. *Int J Pharm* 1986;60:61–78.
71. Otsuka M, Gao J, Matsuda Y. Effects of mixer and mixing time on the pharmaceutical properties of theophylline tablets containing various kinds of lactose as diluents. *Drug Dev Ind Pharm* 1993;19:333–348.
72. Kikuta J, Kitamori N. Effect of mixing time on the lubricating properties of magnesium stearate and the final characteristics of the compressed tablet. *Drug Dev Ind Pharm* 1994;20:343–355.
73. Otsuka M, Sato M, Matsuda Y. Comparative evaluation of tableting compression behaviours by methods of internal and external lubricant addition: inhibition of enzymatic activity of trypsin preparation by using external lubricant addition during the tableting compression process. *AAPS PharmSci* 2001;3:1–11.
74. Khan KA, Rooke DJ. Effect of disintegrant type upon the relationship between compressional pressure and dissolution efficiency. *J Pharm Pharmacol* 1976;28:633–636.
75. Zhang Y, Law Y, Chakrabarti S. Physical properties and compact analysis of commonly used direct compression binders. *AAPS PharmSci* 2003;4(article 62):1–11.
76. Proost JH, Bolhuis GK, Lerk CF. The effect of the selling capacity of disintegrants on the *in vitro* and *in vivo* availability of diazepam tablets, containing magnesium stearate as a lubricant. *Int J Pharm* 1983;13:287–296.
77. Sjokvist E, Nystrom C, Alden M. Physicochemical aspects of drug release. IX: Investigation of some factors that impair dissolution of drugs from solid particulate dispersion systems. *Int J Pharm* 1989;54:161–170.
78. Johnson JR, Wang LH, Gordon MS, Chowhan ZT. Effect of formulation solubility and hygroscopicity on disintegrant efficiency in tablets prepared by wet granulation, in terms of dissolution. *J Pharm Sci* 1991;80:469–471.
79. Gordan MS, Rudrarju VS, Rhie JK, Chowhan ZT. The effect of aging on the dissolution of wet granulated tablets containing super disintegrants. *Int J Pharm* 1993;97:119–131.
80. Bolhuis GK, Zuurman K, Wierik GHP. Improvement of dissolution of poorly soluble drugs by solid deposition on a super disintegrant. II. The choice of super disintegrants and effect of granulation. *Eur J Pharm BioPharm* 1997;5:63–69.
81. Lundqvist AEK, Podczeck F, Newton JM. Influence of disintegrant type and proportion on the properties of tablets produced from mixtures of pellets. *Int J Pharm* 1997;147:95–107.
82. Chen CR, Cho SL, Lin CK, Lin YH, Chiang ST, Wu HL. Dissolution difference between acidic and neutral media of acetaminophen tablets containing a super disintegrant and a soluble excipient. II. *Chem Pharm Bull* 1998;46:478–481.
83. Sallam E, Ibrahim H, Dahab RA, Shubair M, Khalil E. Evaluation of fast disintegrants in terfenadine tablets containing a gas evolving disintegrant. *Drug Dev Ind Pharm* 1998;24:501–507.
84. Carmella C, Colombo P, Conte U, La Manna A. Tablet disintegrate update; the dynamic approach. *Drug Dev Ind Pharm* 1987;13:2111–2145.
85. Lopez-Solis J, Villafuerte-Robles L. Effect of disintegrants with different hygroscopicity on dissolution of Norfloxacin/Pharmatose DCL 11 tablets. *Int J Pharm* 2001;216: 127–135.
86. Zao N, Augsburger LL. The influence of swelling capacity of superdisintegrants in different pH media on the dissolution of hydrochlorothiazide from directly compressed tablets. *AAPS PharmSci* 2005;6:E120–E126.

87. Zao N, Augsburger LL. Functionality comparison of 3 classes of superdisintegrants in promoting aspirin tablet disintegration and dissolution. *AAPS PharmSci* 2005;6:E634–E640.
88. Cheng G, An F, Zou MJ, Sun J, Hao XH, He YX. Time- and pH-dependent colon-specific drug delivery for orally administered diclofenac sodium and 5-aminosalicylic acid. *World J Gastroenterol* 2004;10:1769–1774.
89. Wakerly Z, Fell JT, Attwood D, Parkins D. Pectin/ethylcellulose film coating formulations for colonic drug delivery. *Pharm Res* 1996;13:1210–1212.
90. Prabakaran D, Singh P, Kanaujia P, Mishra V, Jaganathan KS, Vyas SP. Controlled porosity osmotic pumps of highly aqueous soluble drug containing hydrophilic polymers as release retardants. *Pharm Dev Technol* 2004;9:435–442.
91. Rao VM, Haslam JL, Stella VJ. Controlled and complete release of a model poorly water-soluble drug, prednisolone, from hydroxypropyl methylcellulose matrix tablets using (SBE) (7m)-beta-cyclodextrin as a solubilizing agent. *J Pharm Sci* 2001;90: 807–817.
92. Tiwari D, Sause R, Madan PL. Evaluation of polyxyethylene homopolymers for buccal bioadhesive drug delivery device formulations. *AAPS PharmSci* 1999;1(article 13):1–8.
93. Chiao CS, Price JC. Formulation, preparation and dissolution characteristics of propranolol hydrochloride microspheres. *J Microencapsulation* 1994;11:153–159.
94. Saravanan M, Natraj KS, Ganesh KS. Hydroxypropyl methylcellulose based cephalexin extended release tablets: influence of tablet formulation, hardness and storage on *in vitro* release kinetics. *Chem Pharm Bull* 2003;51:978–983.
95. Bordaweka MS, Zia H. Evaluation of polyvinyl acetate dispersion as a sustained release polymer for tablets. *Drug Deliv* 2006;13:121–131.
96. Bravo SA, Lamas MC, Salomon CJ. *In-vitro* studies of diclofenac sodium controlled-release from biopolymeric hydrophilic matrices. *J Pharm Pharm Sci* 2002;5:213–219.
97. Pignatello R, Ferro M, Puglisi G. Preparation of solid dispersions of nonsteroidal anti-inflammatory drugs with acrylic polymers and studies on mechanisms of drug-polymer interactions. *AAPS PharmSci* 2002;3(article 10):1–11.
98. Adeyeye MC, Mwangi E, Katondo B, Jain A, Ichikawa H, Fukumori Y. Dissolution stability studies of suspensions of prolonged-release diclofenac microcapsules prepared by the Wurster process: I. Eudragit-based formulation and possible drug-excipient interaction. *J Microencapsulation* 2005;22:333–342.
99. Jenquin MR, McGinity JW. Characterization of acrylic resin matrix films and mechanisms of drug–polymer interactions. *Int J Pharm* 10:23–34.
100. Karlsen J. Excipient properties of chitosan. *Manufacturing Chemist* 1991;62:18–19.
101. Nunthanid J, Laungtana-anan M, Sriamornsak P, Limmatvapirat S, Puttipipatkhachorn S, Lim, LY, Khor E. Characterization of chitosan acetate as a binder for sustained release tablets. *J Control Release* 2004;99:15–26.
102. Gao P, Nixon P, Skoug J. Diffusion in HPMC gels. II. Prediction of drug release rates from hydrophilic matrix extended-release dosage forms. *Pharm Res* 1995;12:965–971.
103. Phaechamud T, Koizumi T, Ritthidej GC. Chitosan citrate as film former: compatibility with water-soluble anionic dyes and drug dissolution from coated tablets. *Int J Pharm* 2000;198:97–111.
104. Ritthidej GC, Phaechamud TW. Effect of anionic water-soluble dyes on film coating properties of chitosan acetate. *Drug Dev Ind Pharm* 2003;29:585–594.
105. Breikkreutz J. Leakage of enteric (Eudragit L)-coated dosage forms in simulated gastric juice in the presence of poly(ethylene glycol). *J Control Release* 2000;67:79–88.

106. Mathur AV, Hammonds KF, Klier J, Scranton AB. Equilibrium swelling of poly(methacrylic acid-ethylene glycol) hydrogels: effect of swelling medium and synthesis conditions. *J Control Release* 1998;54:177–184.
107. Chasseaud LF, Taylor T. Bioavailability of drugs from formulations after oral administration. *Annu Rev Pharmacol* 1974;14:35–46.
108. Bourne DWA, Dittert LW. *Modern Pharmaceutics* 4th ed. New York: Marcel Dekker; 1996, pp 122–153.
109. Banakar UV, Makoid MC. *Drug Development Process: Increasing Efficiency and Cost-Effectiveness*. New York: Marcel Dekker Taylor & Francis CRC; 1996, pp 117–168.
110. Du J, Hoag SW. Characterization of excipient and tableting factors that influence folic acid dissolution, friability and breaking strength of oil- and water-soluble multi vitamin with minerals tablets. *Drug Dev Ind Pharm* 2003;29:1137–1147.
111. *United States Pharmacopoia* 24 and *National Formulary 19*. USP publishers. Rockville, MD: US Pharmacopoeia; 2001.
112. Urakami K, Shono Y, Higashi A, Umemoto K, Godo M. A novel method for estimation of transition temperature for polymorphic pairs in pharmaceuticals using heat of solution and solubility data. *Chem Pharm Bull* 2002;50:263–267.
113. Hancock BC, Parks M. What is the true solubility advantage of amorphous pharmaceuticals? *Pharm Res* 2001;17:397–404.
114. Bartolomei M, Bertocchi P, Antoniella E, Rodomonte A. Physicochemical charaterization and intrinsic dissolution study of new hydrate form of diclofenac sodium: comparison with anhydrous form. *J Pharm Biomed Anal* 2006;40:1105–1113.
115. Manca ML, Zaru M, Ennas G, Valenti D, Sininco C, Loy G, Fadda AM. Diclofenac-β-cyclodextrin binary system: physicochemical characterization and *in vitro* dissolution and diffusion study. *AAPS PharmSci Tech* 2005;6(article 58):E464–E472.
116. Mukne AP, Nagarsenker MS. Triamterene-β-cyclodextrin systems: preparation, characterization and *in vivo* evaluation. *AAPS PharmSci Tech* 2004;5(article 19):1–9.
117. Marlowe E, Shangraw RF. Dissolution of sodium salicylate from tablet matrices prepared by wet granulation and direct compression. *J Pharm Sci* 1967;56:498–504.
118. Abdou HM. *Remington's Pharmaceutical Sciences*, 17th ed. Easton, PA: Mack Publishing; 1986, pp 653–666.
119. Bansal P, Haribhakti K, Subramanian V, Plakogiannis F. Effect of formulation and process variables on the dissolution profile of naproxen sodium from tablets. *Drug Dev Ind Pharm* 1994;20:2151–2156.
120. Mitchell K, Ford JL, Armstrong DJ, Elliott PNC, Hogan JE, Rostron C. The influence of the particle size of hydroxypropylmethylcellulose K15M on its hydration and performance in matrix tablets. *Int J Pharm* 1999;100:175–179.
121. Gao P, Meury RH. Swelling of hydroxypropyl methylcellulose matrix tablets. 1. Characterization of swelling using a novel optical imaging method. *J Pharm Sci* 1996;85: 725–731.
122. Chowhan ZT, Yang IC. Effect of intergranular versus intragranular cornstarch on tablet friability and *in vitro* dissolution. *J Pharm Sci* 1983;72:983–988.
123. Zhang YE, Tchao R, Schwartz JB. Effect of processing methods and heat treatment on the formation of wax matrix tablets for sustained drug release. *Pharm Dev Technology* 2001;6:131–144.
124. Finholt P. In Leeson LJ, Carstenson JT, Eds. *Dissolution Technology*. Industrial Pharmaceuticals Technology Section. Washington DC: Apha; 1974, pp 136–137.
125. Ghorab MK, Adeyeye MC. Enhancement of ibuprofen dissolution via wet granulation with beta-cyclodextrin. *Pharm Dev Technol* 2001;6:305–304.

126. Al-Kassas RS, Gilligan CA, Po ALW. Processing factors affecting particle size and *in vitro* drug release of sustained release ibuprofen microspheres. *Int J Pharm* 1993;94: 59–67.
127. Kanig JL, Rudnic EM. The mechanism of disintegrant action. *Pharm Technol* 1984; 8:50–62.
128. Levy G, Gumtow RH. Effect of certain formulation factors on dissolution rate of the active ingredient III: tablet lubricants. *J Pharm Sci* 1963;52:1139–1144.
129. Basak SC, Sivakamasundari T, Shivagamasundari S, Manavalan R. Influence of granule size and lubricant concentration on the dissolution of paracetamol tablets. *Indian J Pharm Sci* 2003;65:299–301.
130. Lerk CF, Bolhuis GK, Smallenbroek AJ, Zuurman K. Interaction of tablet disintegrants and magnesium stearate during mixing II. Effect on dissolution rate. *Pharm Acta Helv* 1982;57:282–286.
131. Van Oudtshoorn MC, Potgieter FJ, de Blaey CJ, Polderman J. The influence of compression and formulation on the hardness, disintegration, dissolution, absorption and excretion of sulphadimidine tablets. *J Pharm Pharmacol* 1971;23:583–586.
132. Massimo G, Santi P, Colombo G, Nicoli S, Zani F, Colombo P, Bettini R. The suitability of disintegrating force kinetics for studying the effect of manufacturing parameters on spironolactone tablet properties. *AAPS PharmSci* 2003;4(article17):1–7.
133. Tuladhar MD, Carless JE, Summers MP. The effects of polymorphism, particle size and compression pressure on the dissolution rate of phenylbutazone tablets. *J Pharm Pharmacol* 1983;35:269–274.
134. Ganderton D, Hadgraft JW, Rispin WT, Thompson AG. The breakup and dissolution of phenindione tablets. *Pharm Acta Helv* 1967;42:152–162.
135. Kent JS. Implant pellets I: effects of compression pressure on *in vivo* dissolution of delmadinone acetate pellets. *J Pharm Sci* 1976;65:89–92.
136. Lin KH, Lin SY, Li MJ. Compression forces and amount of outer coating layer affecting the time-controlled disintegration of the compression-coated tablets prepared by direct compression with micronized ethylcellulose. *J Pharm Sci* 2001;90:2005–2009.
137. Food and Drug Administration. Guidance for Industry: Immediate release solid oral dosage forms; scale-up and postapproval changes: chemistry, manufacturing, and controls, *in vitro* dissolution testing, and *in vivo* bioequivalance documentation. Rockville, MD: FDA Nov 1995.
138. Prabhu S, Jacknowitz AI, Stout PJ. A study of factors controlling dissolution kinetics of zinc complexed protein suspensions in various ionic species. *Int J Pharm* 2001;217: 71–78.
139. Tang L, Khan SU, Muhammad NA. Evaluation and selection of biorelevant dissolution media for a poorly water soluble new chemical entity. *Pharm Dev Technol* 2001; 6:531–540.
140. Noory C, Tran N, Ouderkirk L, Shah V. Steps for development of a dissolution test for sparingly water soluble drug products. *Dissolution Technol* 2000;Article 3:1–5.
141. *British Pharmacopoeia*. London: British Pharmacopoeia Commission Bernan Press; 2003.
142. Government of India Ministry of Health and Family Welfare. *The Pharmacopoeia of India*. Delhi, India: Controller of Publication; 1996.
143. Crison JR, Weiner ND, Amidon GL. Dissolution media for *in vitro* testing of water insoluble drugs: effect of surfactant purity and electrolyte on *in vitro* dissolution of carbamazepine in aqueous solutions of sodium lauryl sulphate. *J Pharm Sci* 1997;86: 384–388.

144. Jinno J, Oh D, Crison JR, Amidon GL. Dissolution of ionizable water insoluble drugs: the combined effect of pH and surfactant. *J Pharm Sci* 2000;89:268–274.

145. Li J, Yang L, Ferguson SM, Hudson TJ, Watanabe S, Katsuma M, Fix JA. *In vitro* evaluation of dissolution behaviour for a colon specific drug delivery system (CODES™) in multi-pH media using United States Pharmacopeia apparatus II and III. *AAPS PharmSciTech* 2002;3(article 33):1–9.

146. Galia E, Nicolaides E, Hörter D, Löbenberg R, Reppas C, Dressman JB. Evaluation of various dissolution media for predicting *in vivo* performance of class I and II drugs. *Pharm Res* 1998;15:698–705.

147. Crison JR. Developing dissolution test for modified release dosage form: general considerations. http://wwwdissolutiontechcom/Dtresour/299articlesCrison,htm.

148. Moller H, Wirbitzki E. Regulatory aspects of modified release dosage form: special cases of dissolution testing using the flow-through system. *Bollettino Chim Farma* 1993;132: 105–115.

149. Burns SJ, Corness D, Hay G, Higginbottom S, Whelan I, Attwood D, Barnwell SG. Development and validation of an *in vitro* dissolution method for a floating dosage form with biphasic release characteristics. *Int J Pharm* 1995;121:37–44.

150. Crail DJ, Tunis A, Dansereau R. Is the use of a 200 ml vessel suitable for dissolution of low dose drug products? *Int J Pharm* 2004;269:203–209.

151. Yu LX, Carlin AS, Amidon GL, Hussain AS. Feasibility studies of utilizing disk intrinsic dissolution rate to classify drugs. *Int J Pharm* 2004;270:221–227.

152. Banakar UV. *Pharm Dissolution Testing*. New York: Marcel Dekker; 1992, pp 107–187.

153. Sarapu A, Clark J. Elimination of tablet air entrapment using USP 1 rotating-basket dissolution apparatus. *J Pharm Sci* 1980;69:129.

154. Cappola ML. A better dissolution method for ranitidine tablets USP. *Pharm Dev Technol* 2001;6:11–17.

155. Hamlin WE, Nelson E, Ballard BE, Wagner JG. Loss of sensitivity in distinguishing real differences in dissolution rates due to increasing intensity of agitation. *J Pharm Sci* 1962;51:432–435.

156. Kukura J, Baxter JL, Muzzio FJ. Shear distribution and variability in the USP apparatus 2 under turbulent conditions. *Int J Pharm* 2004;279:9–17.

157. Shah VP, Konecny JJ, Everett RL, McCullough B, Noorizadeh AC, Skelly JP. *In vitro* dissolution profile of water insoluble drug dosage forms in the presence of surfactants. *Pharm Res* 1989;6:612–618.

158. Qureshi SA, Shabnam J. Cause of high variability in drug dissolution testing and its impact on setting tolerances. *Eur J Pharm Sci* 2001;12:271–276.

159. Wurster DE, Polli GP. Investigation of drug release from solids. V. Simultaneous influence of adsorption and viscosity on the dissolution rate. *J Pharm Sci* 1964;53:311–314.

160. Bodmeier R, Guo X, Sarabia RE, Skultety PF. The influence of buffer species and strength on diltiazem HCl release from beads coated with aqueous cationic polymer dispersions. Eudragit RS, RL 30D. *Pharm Res* 1996;13:52–56.

161. Fagan PG, Harrison PJ, Shankland N. A correlation between cloud point and disintegration of hydroxyalkylcellulose controlled release matrices. *J Pharm Pharmacol* 1989;41: 25.

162. Jalil R, Ferdous AJ. Effect of viscosity increasing agent and electrolyte concentration on the release rate of theophylline from HPMC based sustained release capsules. *Drug Dev Ind Pharm* 1993;19:2637–2643.

163. Li WPA, Wong LP, Gilligan CA. Characterization of commercially available theophylline sustained or controlled-release systems: *in vitro* drug release profiles. *Int J Pharm* 1990;66:111–130.

164. Bonferoni MC, Rossi S, Ferrari F, Stavik F, Pena-Romero A, Caramella C. Factorial analysis of the influence of dissolution medium on drug release from carrageenan–diltiazem complexes. *AAPS PharmSciTech* 2000;1(article 15):1–8.

165. Prasad VK, Rapaka RS, Knight PW, Cabana BE. Dissolution medium—a critical parameter to identify bioavailability problems of furosemide tablets. *Int J Pharm* 1982;11: 81–90.

166. O'Hara TO, Dunne A, Butler J, Devane J. A review of methods used to compare dissolution profile data. *AAPS PharmSciTech* 2000;1(article 15):214–223.

167. Demirtürk E, Öner L. Evaluation of *in vitro* dissolution profile comparison methods of immediate release gliclazide tablet formulations. *Hacettepe Univ J Faculty Pharm* 2005;25:1–10.

168. Chow SC, Ki FYC. Statistical comparison between dissolution profiles of drug products. *J BioPharm Stat* 1997;7:241–258.

169. Qazi S, Samuel NKP, Venkatachalam TK, Uckun FM. Evaluating dissolution profiles of an anti-HIV agent using ANOVA and non-linear regression models in JMP software. *Int J Pharm* 2003;252:27–39.

170. Tsong Y, Hammerstorm T, Sathe P, Shah VP. Statistical assessment of mean difference between two dissolution data sets. *Drug Info J* 1996;30:1105–1112.

171. Sathe P, Tsong Y, Shah VP. *In vitro* dissolution profilee comparison: statistics and analysis, model dependent approach. *Pharm Res* 1996;13:1799–1803.

172. Shah VP, Tsong Y, Sathe P, Liu JP. *In vitro* dissolution profile comparison and analysis of the similarity factor, $f_2$. *Pharm Res* 1998;15:889–896.

173. Peh KK, Wong CF. Application of similarity factor in development of controlled release diltiazem tablet. *Drug Dev Ind Pharm* 2000;26:723–730.

174. Rescigno R. Bioequivalence. *Pharm Res* 1992;9:925–928.

175. Lin SY, Yang JC. *In vitro* dissolution behaviour of some sustained release theophylline dosage forms. *Pharm Acta Helv* 1989;4:236–240.

176. Sangali ME, Giunchedi P, Maggi L, Conte U, Gazzaniga A. Inert monolithic device with a central hole for constant drug release. *Eur J Pharm Sci* 1994;40:370–373.

177. Korsmeyer RW, Gurny R, Doelker EM, Buri P, Peppas NA. Mechanism of solute release from porous hydrophilic polymers. *Int J Pharm* 1983;15:25–35.

178. Kim H, Fassihi R. Application of binary polymer system in drug release rate modulation. 2. Influence of formulation variables and hydrodynamic conditions on release kinetics. *J Pharm Sci* 1997;86:323–328.

179. El-Arini SK, Leuenberger H. Dissolution properties of praziquantel–PVP systems. *Pharm Acta Helv* 1998;73:89–94.

180. Crowder MJ. Keep timing the tablets: statistical analysis of pill dissolution rates. *Appl Stat* 1996;45:323–334.

181. Baker RW, Losdale HS. *Controlled Release: Mechanisms and Rates*. New York: Plenum Press, 1974, pp 15–71.

182. Jun HW, Lai JW. Preparation and *in vitro* dissolution tests of egg albumin microcapsules of nitrofurantoin. *Int J Pharm* 1983;16:65–77.

183. Shukla AJ, Price JC. Effect of drug (core) particle size on the dissolution of theophylline from microspheres made from low molecular weight cellulose acetate propionate. *Pharm Res* 1989;6:418–421.

184. Hopfenberg HB. In Paul DR, Harris FW. Eds. *Controlled Release Polymeric Formulations*. ACS Symposium Series 33. Washington DC: American Chemical Society; 1976, pp 26–31.
185. Higuchi T. Mechanism of sustained action medication. Theoretical analysis of rate of release of solid drugs dispersed in solid matrices. *J Pharm Sci* 1963;52:1145–1149.
186. Desai SJ, Singh P, Simonelli AP, Higuschi WI. Investigation of factors influencing release of solid drug dispersed in inert matrices. III. Quantitative studies involving the polyethylene plastic matrix. *J Pharm Sci* 1966;55:1230–1234.
187. Desai SJ, Singh P, Simonelli AP, Higuschi WI. Investigation of factors influencing release of solid drug dispersed in inert matrices. IV. Some studies involving the polyvinyl chloride matrix. *J Pharm Sci* 1966;55:1235–1239.
188. Niebergall PJ, Milosovich G, Goyan JE. Dissolution rate studies. II. Dissolution of particles under conditions of rapid agitation. *J Pharm Sci* 1963;52:236–241.
189. Langenbucher F. Linearization of dissolution rate curves by the Weibull distribution. *J Pharmacol* 1972;24:979–981.
190. Pedersen PV, Myrick JW. Versatile kinetic approach to analysis of dissolution data. *J Pharm Sci* 1978;67:1450–1455.
191. Christensen FN, Hansen FY, Bechgaard H. Physical interpretation of parameters in the Rosin–Rammler–Sperling–Weibull distribution for drug release from controlled release dosage forms. *J Pharm Pharmacol* 1980;32:580–582.
192. Costa P, Lobo JMS. Modeling and comparison of dissolution profiles. *Eur J Pharm Sci* 2001;13:123–133.
193. Varelas CG, Dixon DG, Steiner C. Zero order release from biphasic polymer hydrogels. *J Control Release* 1995;34:185–192.
194. Polli JE, Rekhi GS, Shah VP. Methods to compare dissolution profiles. *Drug Info J* 1996;30:1113–1120.
195. Lin JH, Chiba M, Balani SK, Chen I-W, Kwei GY-S, Vastag KJ, Nishine JA. Species differences in the pharmacokinetics and metabolism of indinavir, a potent HIV protease inhibitor. *Drug Metab Dispos* 1996;24:1111–1120.
196. Lin JH, Storey DE, Chen I-W, Xu X. Improved absorption of L-365, 260, a poorly soluble drug. *Biopharm Drug Dispos* 1996;17:1–16.
197. Yeleswara K, McLaughlin LG, Knipe JO, Schabdach D. Pharmacokinetics and oral bioavailability of exogenous melatonin in preclinical animal models and clinical implications. *J Pineal Res* 1997;22:45–51.
198. Clemens ET, Stevens CE. A comparison of gastrointestinal transit time in ten species of mammal. *J Agric Sci* 1980;94:735–737.
199. Kararli TT. Comparison of the gastrointestinal anatomy, physiology, and biochemistry of humans and commonly used laboratory animals. *Biopharm Drug Dispos* 1995;16: 351–380.
200. Obach RS, Baxter JG, Liston TE, Silber M, Jones BC, MACintyre F, Rance DJ, Wastall P. The prediction of human pharmacokinetic parameters from preclinical and *in vitro* metabolism data. *J Pharmacol Exp Ther* 1997;283:46–58.
201. Renkin EM, Gilmore JP. *Handbook of Physiology—Renal Physiology*. Baltimore, MD: Williams and Wilkins; 1973, pp 185–248.
202. Dressman JB, Yamada K. *Pharmaceutical Bioequivalence*. New York: Marcel Dekker; 1991, pp 235–266.
203. Davies B, Morris T. Physiological parameters in laboratory animals and humans. *Pharmaceutical Research*, 10, 1093–1095.

204. Miyabayashi T, Morgan JP, Atitola MAO, Muhumuza L. Small intestinal emptying time in normal beagle dogs. A contrast radiographic study. *Vet Radiol* 1986;27:197–209.
205. Ilet KF, Tee LBG, Reeves PhT, Minchin RF. Metabolism of drugs and other xenobiotics in the gut lumen and wall. *Pharmacol Ther* 1990;46:67–93.
206. Degen LP, Phillips SF. Variability of gastrointestinal transit in healthy women and men. *Gut* 1996;39:299–305.
207. Ikegami K, Tagawa K, Narisawa S, Osawa T. Suitability of the cynomolgus monkey as an animal model for drug absorption studies of oral dosage forms from the viewpoint of gastrointestinal physiology. *Biol Pharm Bull* 2003;1442–1447.
208. Kaniwa N, Aoyagi N, Ogata H, Ejima A. Gastric emptying rates of drug preparations. I. Effects of size of dosage form, food and species on gastric emptying rates. *J Pharmcobio Dynam* 1988;11:563–570.
209. Chiou1 WL, Buehler1 PW. Comparison of oral absorption and bioavailability of drugs between monkey and human. *Pharm Res* 2002;19:868–873.
210. Edwards NA. Scaling of renal function in mammals. *Comp Biochem Physiol* 1975;52A:63–66.
211. Boxenbaum H. Interspecies variation in liver weight, hepatic blood flow, and antipyrin intrinsic clearance in extrapolation of benzodiazepines and phenytoin. *J Phamacokinet Biopharm* 1980;8:165–176.
212. Stahl WR. Scaling of respiratory variable in mammals. *J Appl Physiol* 1969;22: 453–460.
213. Knaak JB, Al-Bayati MA, Raabe OG, Blancatos JN. Development of *in vitro* $V_{max}$ and $K_m$ values for the metabolism of isofenphos by P-450 liver enzymes in animals and human. *Toxicol Appl Pharmacol* 1993;120:106–113.
214. Sharer JE, Shipley LA, Vandenbranden MR, Binkley SN, Wrighton SA. Comparisons of phase I and phase II *in vitro* hepatic enzyme activities of human, dog, rhesus monkey, and cynomolgus monkey. *Drug Metab Dispos* 1995;23:1231–1241.
215. Chauret N, Gauthier A, Martin J, Nicoll-Griffith D. *In vitro* comparison of cytochrome P450-mediated metabolic activities in human, dog, cat, and horse. *Drug Metab Rev* 1997;25:1130–1136.
216. Shimida T, Mimura M, Inoue K, Nakamura S, Oda H, Ohmori S, Yamazaki H. Cytochrome P450-dependent drug oxidation activities in liver microsomes of various animal species including rats, guinea pigs, dogs, monkeys, and humans. *Arch Toxicol* 1997;71: 401–408.
217. Igarashi T, Tomihari N, Ohmori S, Ueno K, Kitagawa H, Satoh T. Comparison of glutathione *S*-transferase in mouse, guinea pig, rabbit and hamster liver cytosol to those in rat liver. *Biochem Int* 1986;13:641–648.
218. Stevens JC, Shipley LA, Cashman JR, Vandenbranden M, Wrighton SA. Comparison of human and rhesus monkey *in vitro* phase I and phase II hepatic drug metabolism activities. *Drug Metab Rev* 1993;21:753–759.
219. Puri S, Kohli KK. Differences in hepatic drug metabolizing enzymes and their response to lindane in rat, rabbit and monkey. *Pharmacol Toxicol* 1995;77:136–141.
220. Kobayashi S, Murray S, Watson D, Sesardic D, Davies DS, Boobies AB. The specificity of inhibition of debrisoquine 4-hydroxylase activity in quinidine and quinine in rat is the reverse of that in man. *Biochem Pharmacol* 1989;38:2795–2799.
221. Lin JH. Species similarities and differences in pharmacokinetics. *Drug Metab Dispos* 1995;17:221–223.
222. Guidance for Industry. Extended release oral dosage forms: development, evaluation and application of *in vitro/in vivo* correlations. Washington DC: Food and Drug Administration CDER; Sept 1997.

223. Young D, Davane JG, Butler J. *In Vitro–In Vivo Correlations*. New York: Plenum Press; 1997.
224. Nelson E. Comparative dissolution rates of weak acid and their sodium salts. *J Am Pharm Assoc* 1958;47:297–300.
225. Schirmer RE, Kleber JW, Black HR. Correlation of dissolution, disintegration and bioavailability of aminosalicylic acid tablets. *J Pharm Sci* 1973;62:1270–1274.
226. Needham TE, Shah K, Kotzan J, Zia H. Correlation of aspirin excretion with parameter from different dissolution methods. *J Pharm Sci* 1978;67:1070–1073.
227. Fairweather WR. Investigating relationships between *in vivo* and *in vitro* pharmacological variables for the purpose of prediction. *J Pharmacokinet BioPharm* 1977;5:405–418.
228. Concherio A, Vila-Jato JL, Martinez-Pacheco R, Seijo B, Ramos T. *In vitro-in vivo* correlations of eight nitrofurantoin tablet formulations: effect of various technological factors. *Drug Dev Ind Pharm* 1987;13:501–516.
229. Wagner JG. *Pharmacokinetics for the Pharmceutical Scientist*. New York: Technomic Publishing; 1993, pp 259–269.
230. Bennett CA, Franklin NL. *Statistical Analysis in Chemistry and the Chemical Industry*. Hoboken, NJ: Wiley; 1963.
231. Johnson NL, Leone FC. *Statistics and Experimental Design in Engineering and the Physical Sciences*. Hoboken, NJ: Wiley; 1977.
232. Dressman J, Krämer J. *Pharmaceutical Dissolution Testing*. New York: Taylor & Francis Group; 2005.
233. *Guidance for Industry*. Waiver of *in vivo* bioavailability and bioequivalance studies for immediate-release solid oral dosage forms based on a biopharmaceutics classification system. Washington DC: Food and Drug Administration CDER, Aug 2000.
234. Burgess DJ, Hussain AS, Ingallinera TS, Chen ML. Assuring quality and performance of sustained and controlled release parenterals: workshop report. *AAPS PharmSci* 2002;4(article 7):1–11.

# STABILITY: PHYSICAL AND CHEMICAL

Eric M. Gorman,[1] Brian E. Padden,[2] and Eric J. Munson[1]
[1]*The University of Kansas, Lawrence, Kansas*
[2]*Schering-Plough Research Institute, Summit, New Jersey*

**Contents**

1 Introduction
2 Solution Stability
  2.1 Chemical Stability
  2.2 Physical Stability
3 Solid-State Stability
  3.1 Physical Stability
  3.2 Chemical Stability
  3.3 Identifying True Solid-State Chemical Reactions
4 Kinetics
  4.1 Kinetic Models
  4.2 Temperature
  4.3 pH
  4.4 Buffers
  4.5 Relative Humidity
5 Accelerated Stability Testing
  5.1 Isothermal Kinetics
  5.2 Nonisothermal Kinetics
6 Improving Stability
7 Methods of Analysis

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

8 Regulatory Guidelines
9 Further Reading
Acknowledgments
References

## 1 INTRODUCTION

The pharmaceutical properties of a drug, such as solubility, dissolution, and toxicology, can change dramatically when a drug undergoes chemical or physical degradation. This can severely limit the drug's efficacy and safety. After a compound is synthesized it must undergo several steps before being used by the patient. Some of these steps include formulation into the final product, storage during manufacturing and distribution, and dispensing to the patient. Understanding the stability of a potential drug is a vital component to the pharmaceutical development process because if the drug changes during any one of these steps it may not have the desired pharmaceutical effects. Therefore, the evaluation of drug stability becomes part of the development process as early as candidate selection and continues for the lifetime of the product. For the purposes of this chapter, however, we only consider drug stability up to and through the formulation process.

There are four main reasons to study the stability of drugs [1]:

1. *Quantitation of Drug Loss*. It is essential to know how fast the drug degrades so that an accurate shelf life can be determined.
2. *Identification and Quantitation of Degradation Products*. It is important to know what the degradation products are, as they may be toxic, and to know at what levels they are produced.
3. *Change in Aesthetic Properties Upon Degradation*. If a drug product changes notably in its aesthetic appearance it is presumed adulterated, even if by all other criteria the drug product is perfectly safe for use. This typically occurs if one of the degradation products either has an offensive odor or is highly colored. An example is tetracycline degrading to form epianhydrotetracycline, which is red [2, 3].
4. *Degradation of Drug After Administration*. Even if the drug is stable in its formulation, it is important to know how it will degrade under physiologic conditions; for example, degradation at low pH for an orally administered drug.

There have been numerous articles and books published on the subject of stability. The best way to begin a stability study is to perform a thorough search of the literature to find similar compounds that have been studied previously, which may provide some insight into potential reaction mechanisms and products. In doing this, a foundation is laid for the studies that are to be performed on the new compound.

## 2 SOLUTION STABILITY

Stability studies on new drug candidates are typically performed in the solution state. These studies are used as an aid in candidate selection and are intended to identify compounds that would likely pose a significant challenge during the drug development process. Solution stability is also important as a guide for the formulation process. For example, if a drug is very sensitive to low pH then an enteric-coated tablet may be required for oral drug delivery.

### 2.1 Chemical Stability

Chemical stability is a measure of the susceptibility of a compound to undergo a chemical reaction in which atoms are rearranged via bond breakage and bond formation to produce a different chemical entity. Drugs have a myriad of different chemical structures and are thus subject to a variety of different possible mechanisms of degradation. Most drugs degrade by one or more of the following mechanisms: hydrolysis, dehydration, isomerization, racemization, decarboxylation, elimination, oxidation, or photodegradation [1, 4, 5]. It may also be possible to have reactions between two drug molecules [6] or between a drug molecule and an excipient within a formulation [7].

Hydrolysis is one of the most frequent degradation mechanisms observed in pharmaceuticals for two main reasons. First, drugs are typically dissolved in media primarily composed of water, which results in a relative abundance of water that is available to react. Second, ester and amide functional groups degrade primarily via hydrolysis and these are two very common functional groups in pharmaceuticals.

Proteins and peptides are steadily becoming more popular as pharmaceuticals. These biopharmaceuticals, as they are called, are subject to the same degradation reactions as "traditional" small molecule pharmaceuticals, and in some cases these reactions are given specific names to describe both the reaction type and the amino acid that is involved [8, 9]. For example, the hydrolysis of the amino acid asparagine to form aspartate and isoaspartate is referred to as deamidation. Biopharmaceuticals are inherently more complex than small molecules because of their increased size and higher levels of organization (primary, secondary, tertiary, and quaternary structure), which leads to significantly increased complexity when the stability of these compounds is considered [9–13].

### 2.2 Physical Stability

Physical stability is a measure of the propensity of a compound to undergo changes that do not involve the breaking or formation of covalent bonds. In this case the compound has the exact same chemical structure but the physical state has changed in some way. For small molecules, two potential physical stability problems in solution are precipitation and aggregation [14, 15]; however, these are both relatively rare when compared with chemical degradation mechanisms of small molecules. On the other hand, physical stability is a major concern in the case of biopharmaceuticals. If the protein/peptide denatures (typically through the loss of quaternary and tertiary structure) then it can adsorb to surfaces (such as glass or plastic), aggregate, and precipitate [16]. In the case of biopharmaceuticals, in particular, these physical

changes typically are not easily reversed [10, 17]. One additional mechanism of physical degradation is vaporization. If the drug has a sufficiently high vapor pressure it can evaporate from the formulation whether it is in the solution or solid state. This phenomenon has been observed in nitroglycerin and can result in large variability in the drug content of tablets after being stored in the same container [18].

## 3 SOLID-STATE STABILITY

Pharmaceuticals are typically formulated in the solid state for several reasons. One major advantage is increased patient compliance, because it is easy to dose the pharmaceutical. It is also discreet, meaning that a patient can use it in a social setting and not draw the attention of other people while doing so. Another advantage is that it is typically cheaper than developing and producing a sterile product for parenteral administration. One of the main reasons for formulating a pharmaceutical in the solid state is that typically the compound is more stable in the solid state compared to the solution state. However, the solid state presents its own unique challenges.

### 3.1 Physical Stability

In the solid state a compound can be either amorphous or crystalline. A solid is amorphous if it lacks long-range order in the arrangement of the molecules. Amorphous materials can exist in one of two states: glassy or rubbery. The glassy state (low temperature state) is a more rigid state while the rubbery state (high temperature state) is a much more mobile state and hence is less stable. An amorphous material has a characteristic glass transition temperature ($T_g$), which is the temperature at which the material changes from the glassy to the rubbery state, or vice versa [19, 20].

If the molecules in the solid state are packed in a regular pattern with long-range order, then the solid is said to be crystalline. The different conformations or orientations in which molecules are able to pack are referred to as crystal forms. If the molecules pack in one crystalline form the material may be called Form I and if the molecules pack in a different crystalline form the material may be called Form II. The naming of these crystal forms is arbitrary and usually reflects the order in which they were discovered. They are usually represented by either Roman numerals (e.g., I, II, III, IV), letters of the Latin alphabet (e.g., A, B, C, D), letters of the Greek alphabet (e.g., α, β, γ, δ), or Arabic numerals (e.g., 1, 2, 3, 4). These different crystalline forms must be chemically identical in order to be called polymorphs. If the solid also contains solvent molecules in the crystal structure then it is referred to as a solvate, and if that solvent is water then it is specifically called a hydrate. The stoichiometry of the water molecules present in the crystal structure is represented by the name (e.g., a monohydrate has one molecule of water per molecule of the compound of interest, a dihydrate has two molecules of water, etc.). On the other hand, if the crystal structure contains no water then it is referred to as an anhydrate. If multiple anhydrous forms of the same compound exist they are said to be polymorphs. Additionally, if multiple monohydrate forms of the same compound exist they are also said to be polymorphs. Other terms (such as solvatomorphs and pseu-

dopolymorphs) have been used to describe solvates, but these are less favorable because their definitions are not as clear. Crystal habits are another phenomenon of crystalline solids, where crystals have the same internal structure (molecular packing) but have different external shapes or morphologies (e.g., needles, plates), which leads to differences in their pharmaceutical properties, including mixing, flowability, and dissolution [21].

The stability of a polymorph is represented by its free energy. The free energy of a polymorph is proportional to its solubility. Since polymorphs have different free energies, they will also have different solubilities [19]. Therefore, the solubility of polymorphs can be used to determine their relative stability. Only the form with lowest free energy at a specified temperature and pressure is said to be stable. All other forms are said to be metastable relative to that form. If at any temperature the relative solubility difference between two polymorphs is maintained, as shown in Fig. 1a, then the polymorphs are said to be related monotropically. However, if at some temperature the relative solubility of the polymorphs inverts, as shown in Fig. 1b, then they are said to be enantiotropically related. Unfortunately, solubility studies have their drawbacks when it comes to identifying monotropic and enantiotropic systems. One of the major problems is a limited temperature range over which the measurements can be made. While many times the data is linear on a van't Hoff plot (as in Fig. 1), the extrapolation of this data outside the experimental range is often prone to large errors [21]. Differential scanning calorimetry (DSC) can be a very useful tool in identifying monotropic and enantiotropic systems, because it provides heats of transition and fusion for polymorphs and polymorphic conversions. These values can then be interpreted using the rules developed by Burger and Ramberger [22, 23]:

1. *Heat* (*or Enthalpy*) *of Transition Rule*. If a solid–solid transition is endothermic then the polymorphs are related enantiotropically; if the transition is exothermic then they are related monotropically.
2. *Heat* (*or Enthalpy*) *of Fusion Rule*. If a polymorph has a lower heat of fusion but a higher melting point than another polymorph, then the two forms are related enantiotropically; if not, the relationship is monotropic.

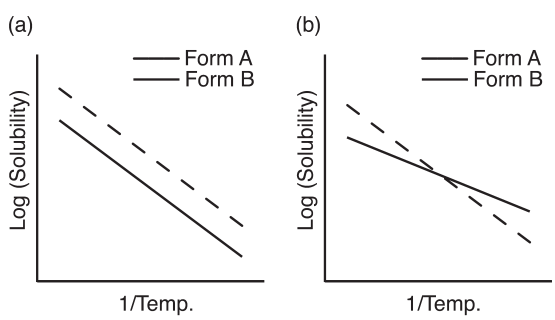


**FIGURE 1** Theoretical van't Hoff plots showing relative water solubility of two polymorphs related (a) monotropically and (b) enantiotropically.

Figure 2a shows the heat of transition rule. In the monotropic system, the metastable Form B converts, at some temperature, to the stable Form A; since Form A has a lower enthalpy ($H_A$) the conversion is an exothermic process. This also holds true for the enantiotropic system if Form B converts to Form A before the crossover temperature, where Form B becomes the stable form and Form A becomes metastable. On the other hand, if Form A converts to Form B in the enantiotropic system, at a temperature above the crossover temperature, then since Form B has a higher enthalpy ($H_B$) the conversion will be endothermic. Figure 2b shows the heat of fusion rule. The point at which the free-energy curve of each polymorph crosses the liquid line in the energy–temperature diagrams is the temperature at which it melts. In the monotropic system the metastable Form B melts at a lower temperature and has a smaller heat of fusion ($\Delta_F H$) than the stable Form A, which has a higher

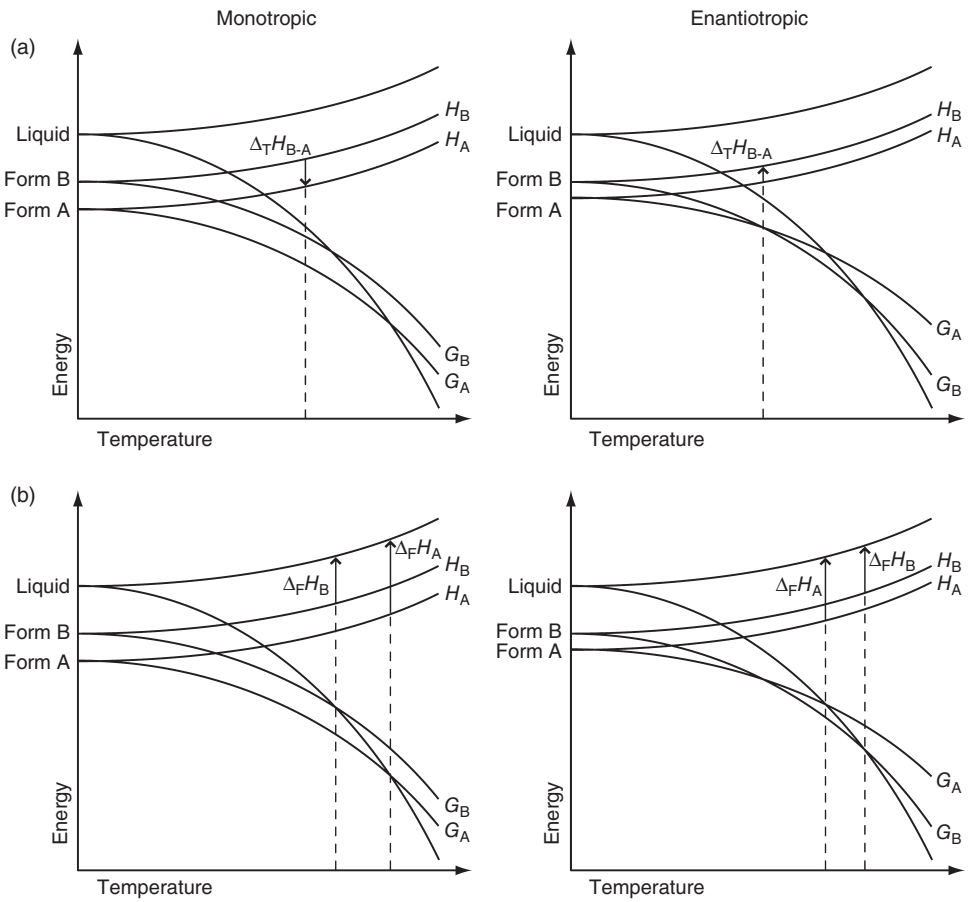


**FIGURE 2** Theoretical energy–temperature diagrams for two polymorphs—Form A and Form B—showing the enthalpy ($H$) and free energy ($G$) of each polymorph as a function of temperature. (a) Energy–temperature diagrams illustrating the heat of transistion ($\Delta_T H$) rule for both a monotropic and an enantiotropic system. (b) Energy–temperature diagrams illustrating the heat of fusion ($\Delta_F H$) rule for both a monotropic and enantiotropic system.

melting temperature and a larger $\Delta_F H$. Conversely, in the enantiotropic system, Form A melts at a lower temperature with a higher $\Delta_F H$ and Form B melts at a higher temperature with a lower $\Delta_F H$. This occurs because the free-energy lines cross but the enthalpy lines do not cross. Additionally, different polymorphs may have different rates of dissolution.

Occasionally, the solubility of the stable form is too low for the drug to reach therapeutic levels when administered to patients. This may require the use of a metastable form with a higher solubility to be used in a formulation. Typically, an amorphous material will have a significantly higher solubility and dissolution rate than any crystalline form of a material and thus is also less stable. Caution must be used when developing a formulation because many of the processing techniques that are used in the pharmaceutical industry can cause polymorphic conversions or other changes to the physical state of a solid, making it vital to identify and control these properties throughout the development process and especially in the final formulated product. One rule of thumb is that if an amorphous material is stored 50 °C below its $T_g$, then its mobility will be sufficiently low to prevent crystallization or other degradation [19, 20, 24, 25].

One example of the consequences of polymorphic change in a formulation is ritonavir [26], a protease inhibitor, that was discovered in 1992 at Abbott Laboratories and was given FDA approval in 1996. It was formulated and marketed as a semisolid capsule (containing a nearly saturated solution) and a liquid formulation. Despite attempts to produce and identify other crystal forms, only one (Form I) was known to exist. Then in 1998 the semisolid formulation began failing dissolution tests because the drug was precipitating in the formulated product. When the precipitate was analyzed it was found to be in a previously unknown crystal form (Form II) and this new form was found to be more thermodynamically stable and thus less soluble. Form II spread throughout their production facilities and eventually they were only able to produce the new less soluble form.

While small molecules are almost always more stable in a crystalline form, the same is not necessarily true for biopharmaceuticals. Pikal and Rigsbee [27] showed that insulin is actually more stable when it is amorphous.

### 3.2 Chemical Stability

All of the chemical reactions that are possible in the solution state are also possible in the solid state. Reactions in crystalline solids are governed by the topochemical postulate [28], which states that reactions will occur with a minimum of atomic and molecular movement. In the solid state, functional groups on a molecule or neighboring molecules may be held closer or farther away from one another than they are in the solution state, which may change the rates of the reactions or even the degradation products that are formed [29]. Additionally, polymorphs and solvates may have different reactivities because of the differences in their crystal structures.

### 3.3 Identifying True Solid-State Chemical Reactions

Water is often present in the solid state, whether it is atmospheric, in the excipients of the formulation, or in the drug itself (e.g., a hydrate). This raises the question of

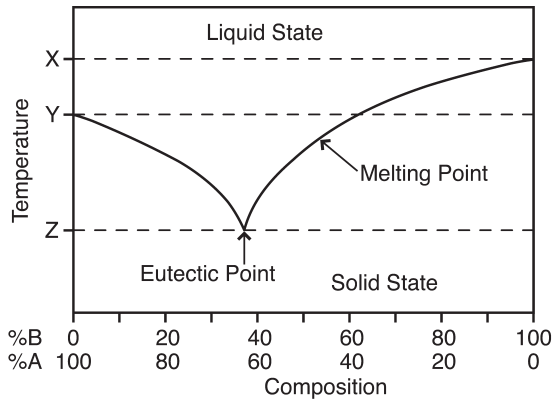


**FIGURE 3** Phase diagram depicting how the melting point of the material changes as the hypothetical degradation of A proceeds to Form B. (Adapted from Ref. 21.)

whether a reaction actually occurs in the solid state or a solution state, where the molecules actually dissolve in the water, react, and then precipitate back out of solution. Five criteria have been developed to identify true solid-state reactions [21].

A reaction occurs in the solid state if:

1. In solution the reaction is slower or does not occur.
2. Pronounced differences are found in the reactivity of compounds with very similar chemical structures when they are in the solid state.
3. Reaction products are different from those observed in solution.
4. Different crystalline forms of the same compound have different reactivities or yield different reaction products.
5. The reaction occurs at a temperature below the eutectic point of a mixture of the starting material and products.

The fifth criterion is illustrated in Fig. 3, where A reacts to form B. A has a melting point of Y and B has a melting point of X. As B is formed it depresses the observed melting point because it acts as an impurity in A. At some critical composition (eutectic point) the melting point begins to rise again because A begins to be the impurity in B. The fifth criterion is fulfilled if the reaction occurs at a temperature below Z.

## 4 KINETICS

In the most general sense drug stability is a very simple concept. A drug reacts via some mechanism, at some rate, and produces a product. However, reactions can become very complex when variables such as temperature, pH, physical state, and so on are taken into account.

### 4.1 Kinetic Models

A chemical reaction can be represented by a rate expression, such as

$$\frac{-d[\mathrm{A}]}{dt} = k[\mathrm{A}]^x[\mathrm{B}]^y[\mathrm{C}]^z \cdots \tag{1}$$

The order of the reaction is given by the sum of the powers ($x$, $y$, $z$, . . .) in the rate expression. In the simplest system the rate of the reaction does not depend on the concentration of any of the reactants. This is called a zero-order reaction and is given by

$$\frac{-d[\mathrm{A}]}{dt} = k \tag{2}$$

The integrated form of this equation is

$$[\mathrm{A}]_t = [\mathrm{A}]_0 - kt \tag{3}$$

Equation 4 represents a first-order reaction and Eq. 5 gives the integrated form:

$$\frac{-d[\mathrm{A}]}{dt} = k[\mathrm{A}] \tag{4}$$

$$[\mathrm{A}]_t = [\mathrm{A}]_0 e^{-kt} \tag{5}$$

Equation 6 represents one example of a second-order reaction and Eq. 7 gives the integrated form:

$$\frac{-d[\mathrm{A}]}{dt} = k[\mathrm{A}]^2 \tag{6}$$

$$\frac{1}{[\mathrm{A}]_t} = kt + \frac{1}{[\mathrm{A}]_0} \tag{7}$$

The order of a reaction is determined by observing the reaction over time and then fitting it to the models to determine which one fits the observed data. Most reactions will be zero, first, or second order. However, if several reactions are occurring simultaneously the model may become more complex. Once the order of the reaction is determined it is important to verify that the order matches the mechanism. For example, hydrolysis reactions should depend on the concentration of both the drug and the water. However, these reactions often appear to follow first-order kinetics because the media in which they occur are composed primarily of water. The concentration of water is so large that it does not change significantly over the course of the reaction, so the rate of the reaction only depends on the concentration of the drug because there is essentially an infinite supply of water for the reaction.

Since this reaction is second order, but appears to be first order, it is referred to as a pseudo-first-order reaction.

Solid-state reactions often display more complicated kinetics, which leads to difficulties in understanding the kinetics of the system. There are four main steps in a solid-state reaction [21].

1. *Loosening of Molecules*. This step occurs at nucleation/defect sites, which are imperfections in the crystal that typically form during the crystallization process or during processing (e.g., grinding) and allow sufficient mobility for the reaction to occur.
2. *Molecular Change*. This step is the actual reaction of interest and is typically slower than the same reaction in solution due to the restricted motion.
3. *Solid Solution Formation*. Early in the reaction only a small amount of the product has been formed and is randomly mixed with the reactant (this mixture is the solid solution), eventually enough product forms and the reactants and products separate.
4. *Separation of Product*. When the product separates from the reactant it forms a new phase that can be amorphous or crystalline.

It is often difficult to fully understand the kinetics of a reaction in the solid state. Several models are commonly used (shown in Table 1) and they can be grouped

**TABLE 1 Solid-State Kinetic Models**

| Name | Equation[a] |
|---|---|
| Prout–Tompkins | $\ln\left(\frac{\alpha}{1-\alpha}\right) = kt + c$ |
| Avrami–Erofe'ev | $[-\ln(1-\alpha)]^n = kt \quad n = \frac{1}{4}, \frac{1}{3}, \frac{1}{2}, \frac{2}{3}$, or 1 |
| One-dimensional phase boundary (zero order) | $1 - \alpha = kt$ |
| Two-dimensional phase boundary (contracting area) | $1 - (1-\alpha)^{1/2} = kt$ |
| Three-dimensional phase boundary (contracting volume) | $1 - (1-\alpha)^{1/3} = kt$ |
| Diffusion-controlled one-dimensional | $\alpha^2 = kt$ |
| Diffusion-controlled two-dimensional | $(1-\alpha)\ln(1-\alpha) + \alpha = kt$ |
| Diffusion-controlled three-dimensional | $1 - \frac{2}{3}\alpha - (1-\alpha)^{2/3} = kt$ |
| Jander equation | $[1 - (1-\alpha)^{1/3}]^2 = kt$ |
| Power law | $a^n = kt \quad n = \frac{1}{4}, \frac{1}{3}, \frac{1}{2}$, 1, or 2 |
| First order | $\ln(\alpha) = kt$ |
| Second order | $\frac{1}{1-\alpha} = kt$ |

[a] $\alpha$ = fraction decomposed, $k$ = rate constant, $t$ = time, and $c$ = constant.

into four types: reactions involving nucleation, reactions controlled by phase boundaries, diffusion-controlled reactions, and other reactions [21].

The two models that fall into the nucleation category are the Prout–Tompkins and Avrami–Erofe'ev equations. Prout and Tompkins [30, 31] developed an equation that assumes a solid-state reaction begins at a nucleation site, branches out in chains, and is terminated more rapidly as the number of nuclei increase. This can be likened to a polymer chain reaction in which there is an initiation, propagation, and termination step. The Avrami–Erofe'ev equation [32–34] also assumes that the reaction starts at nuclei; however, it assumes that the nuclei grow and engulf other nuclei and the directionality of this growth is interpreted from the value of $n$.

If a reaction begins on the surface of a crystal and advances toward the interior, then it falls into the phase boundary category. There are three equations in this group and each represents a different dimensionality of the growth. The one-dimensional or zero-order rate equation (identical to the solution-state zero-order rate equation) assumes that the reaction only moves in one direction across the crystal (e.g., in a rod-shaped crystal the reaction begins at one or both ends and the reaction front travels down the crystal to the opposite end). If the reaction progresses in two dimensions (e.g., in a rod-shaped crystal the reaction initiates on the circumference and then travels radially toward the long axis of the rod) then it is represented by the two-dimensional or contracting area equation [35]. Finally, if the reaction travels in three dimensions (e.g., in a rod-shaped crystal the reaction initiates on both the ends and the circumference and then travels toward the center of the rod) then it can be modeled by the three-dimensional or contracting volume equation [35].

The third group contains the diffusion-controlled reactions and, as the name implies, these equations describe reactions in which two reactants must come together to form the product. As with the phase boundary equations there are three equations that describe the dimensionality of the diffusion: one-, two-, and three-dimensional diffusion-controlled rate equation [35–39]. The Jander equation also falls into this category and it is a simplified form of the three-dimensional diffusion-controlled rate equation.

The final group is composed of all the other equations that are used but do not have any physical interpretation. The power-law equation has been used in solid-state kinetic studies; however, it has no theoretical basis for interpretation except that when $n = 2$ it is identical to the one-dimensional diffusion equation. The first- and second-order rate equations are sometimes used to describe solid-state reactions; however, the order of a reaction in the solid state is not understood or defined as well as it is in the solution state and therefore it has little interpretive value. On the other hand, the zero-order (one-dimensional phase boundary) equation does have a physical meaning in the solid state.

Typically in solid-state kinetic studies the data is fit to each of the equations described earlier, and in the case of the Avrami–Erofe'ev and power-law equations, where $n$ can have multiple values, all of the possible values of $n$ are used to fit the data. This results in approximately 19 different fits of the same set of data and for most fits the correlation will be very good ($R^2 > 0.9$). If information is known about the reaction mechanism, then it can be used to narrow the possible fits to use. For example, if it is known that the reaction does involve two reactants that are not held closely in the crystal structure, then it is likely that one of the diffusion models would

be more representative of the reaction rate. However, typically these distinctions cannot be made and the models with the best fits ($R^2 > 0.98$) are analyzed to see if they can yield some information about the mechanism of the reaction. For example, do the nucleation equations seem to fit better or do the phase boundary equations seem to fit better? One model is then chosen and used to measure the degradation rates in the subsequent stability studies.

### 4.2 Temperature

Drugs must be stable enough to be stored for long periods of time (~2 years); however, when trying to develop a stable formulation it is not desirable to wait for a long period of time to find out if the formulation will be stable enough to survive storage. Therefore, a method for speeding up stability measurements is needed. It is well known that the Arrhenius equation (Eq. 8) describes the relationship between the activation energy ($E_a$) of a reaction and the rate of the reaction ($k$) at a given temperature ($T$), where $A$ is a frequency term and $R$ is the gas constant. The Arrhenius equation can also be linearized (Eq. 9) to simplify data fitting and analysis.

$$k = A\exp\left(\frac{-E_a}{RT}\right) \tag{8}$$

$$\ln(k) = \ln(A) + \left(\frac{-E_a}{RT}\right) \tag{9}$$

Therefore, if the rate of degradation is measured at multiple temperatures it is possible to measure $E_a$ for the reaction by plotting $\ln(k)$ versus $1/T$ to produce an Arrhenius plot, and determine $E_a$ from the slope ($-E_a/R$). The stability of the formulation can then be predicted at room temperature or any other storage conditions such as refrigeration (assuming that the reaction mechanism does not change [40]). While $E_a$ and $A$ are typically assumed to be constants, in theory they are not. The Eyring equation (Eq. 10) can be used to account for the temperature dependency of both $E_a$ and $A$, and as with the Arrhenius equation it can be linearized (Eq. 11) to simplify data analysis.

$$k = \left[\frac{k_B T}{h}\exp\left(\frac{\Delta S^{\ddagger}}{R}\right)\right]\exp\left(\frac{-\Delta H^{\ddagger}}{RT}\right) \tag{10}$$

$$\ln\left(\frac{k}{T}\right) = \left[\ln\left(\frac{k_B}{h}\right) + \left(\frac{\Delta S^{\ddagger}}{R}\right)\right] + \left(\frac{-\Delta H^{\ddagger}}{RT}\right) \tag{11}$$

The Eyring plot is produced by plotting $\ln(k/T)$ versus $1/T$, where $\Delta H^{\ddagger}$ can be obtained from the slope and used to predict the stability at lower temperatures (although the Arrhenius plot is much more commonly used for this purpose). The Arrhenius and Eyring equations form the basis for the accelerated stability tests that are used to evaluate and compare potential formulations in the development process (for more information see Section 5).

**TABLE 2 Sensitivity of Rate Constants to Changes in Temperature**

| | $k_2/k_1$ | |
|---|---|---|
| $E_a$ (kcal/mol) | $\Delta T = 1\,°C$ | $\Delta T = 0.1\,°C$ |
| 10 | 1.058 | 1.006 |
| 20 | 1.119 | 1.011 |
| 30 | 1.184 | 1.017 |

When studying stability it is important to maintain a constant temperature because, as shown by the Arrhenius and Eyring equations, if the temperature changes the rate constant will change. The relevant degradation reactions that most pharmaceuticals undergo in the solution state have an $E_a$ between 10 and 30 kcal/mol and Table 2 shows how sensitive the rate constant of a reaction with an $E_a$ in this range is to changes in the temperature.

Table 2 demonstrates the need for accurate temperature control during stability studies. If the reaction being studied has an $E_a$ of 30 kcal/mol, the measured degradation rate can be off by as much as 18.4% if the temperature is off by only 1 °C. Therefore, ovens that are used for stability studies should have the ability to maintain temperatures as accurately as possible.

In the solid state it is usually assumed that reactions will exhibit Arrhenius kinetics (linear Arrhenius plot); however, this is not always the case. One potential reason for non-Arrhenius kinetics is a change in the mechanism of degradation. Additionally, the physical state of the material may change. For example, an amorphous material may not exhibit Arrhenius kinetics if the temperature range being studied spans the $T_g$ of the material [41], or the material may even crystallize at higher temperatures. Also, a crystalline material may undergo a polymorphic conversion to a different crystalline form at some temperature in the range that is being studied [40].

### 4.3 pH

Hydrogen ($H^+$) and hydroxide ions ($OH^-$) can act as catalysts in reactions. Therefore, the observed rate of a reaction ($k_{obs}$) is actually the sum of the multiple of the rate constants of each reaction mechanism and the concentration (or more accurately the activity) of the corresponding catalyst:

$$k_{obs} = k_{H^+}[H^+] + k_{H_2O} + k_{OH^-}[OH^-] \tag{12}$$

The observed rate of degradation can be measured at numerous pH values and then plotted, as in Fig. 4, to yield a pH–rate profile. Buffers are typically employed to maintain a constant pH during the stability studies; however, they can dramatically affect the results if not properly accounted for (see Section 4.4 on buffers). These profiles can be very simple as they are in Fig. 4a–e, or very complex as in Fig. 4f. If there is no catalysis ($k_{H^+} << k_{H_2O} >> k_{OH^-}$; Fig. 4a) the pH–rate profile should just be a flat line parallel to the x-axis (no change in rate). However, if hydrogen ions catalyze the reaction ($k_{H^+} \geq k_{H_2O} >> k_{OH^-}$; Fig. 4b) then the left side

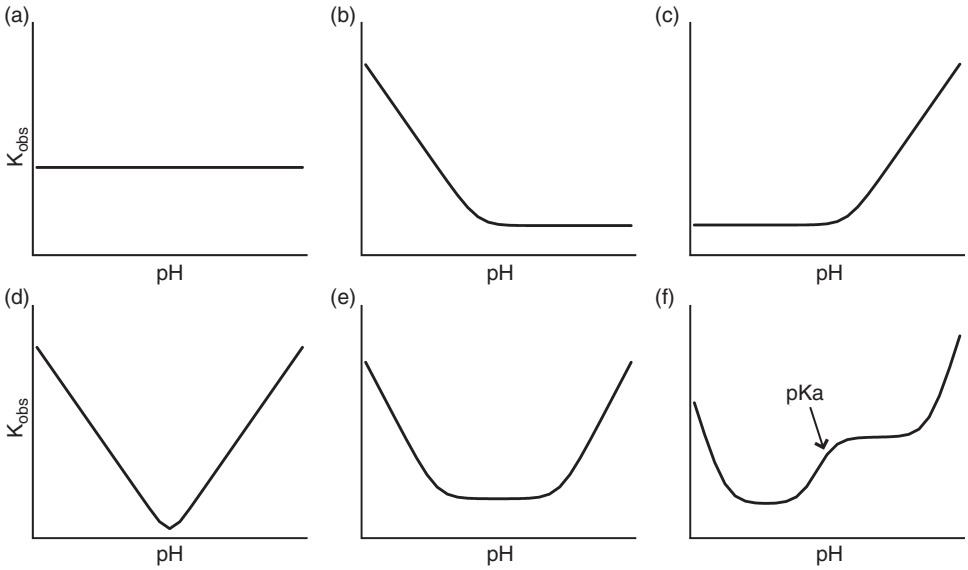


**FIGURE 4** Theoretical pH–rate profiles showing (a) no acid or base catalysis, (b) only acid catalysis, (c) only base catalysis, (d) acid and base catalysis in V-shaped profile, (e) acid and base catalysis in U-shaped profile, and (f) acid and base catalysis of a weak base.

of the profile should curve up, and if hydroxide ions catalyze the reaction ($k_{H^+} << k_{H_2O} \leq k_{OH^-}$; Fig. 4c) then the right side of the profile should curve up. Subsequently, if both ions catalyze the reaction then both ends should curve up, and in this case there are two types of profiles that are possible: the V shaped and the U shaped. In the V shape (Fig. 4d) the $k_{H^+}$ and $k_{OH^-}$ are significantly larger than the $k_{H_2O}$ such that the two sloped regions meet at a point and the profile resembles the letter V. However, in the U shape (Fig. 4e) all three rate constants are much closer to the same value so that the center of the profile is flattened and the entire profile resembles the letter U. If the molecule that is being studied is a weak acid or a weak base (Fig. 4f), the neutral and the ionized forms can have different reactivities and the $H^+$ and $OH^-$ may have different abilities to catalyze the reactions of the two forms. This results in a much more complex equation to describe $k_{obs}$ (Eq. 13) where $k$ is the rate of the reaction of the ionized form and $k'$ is the rate of the neural form and $f_{DH^+}$ and $f_D$ are the fractions of the ionized and neutral forms of a weak base, respectively.

$$k_{obs} = (k_{H^+}[H^+] + k_{H_2O} + k_{OH^-}[OH^-])f_{DH^+} + (k'_{H^+}[H^+] + k'_{H_2O} + k'_{OH^-}[OH^-])f_D \tag{13}$$

$$f_{DH^+} = \frac{[H^+]}{K_a + [H^+]} \tag{14}$$

$$f_D = \frac{K_a}{K_a + [H^+]} \tag{15}$$

It is possible to show that $k_{H_2O}f_{DH^+}$ and $k'_{H^+}[H^+]f_D$ are mathematically equivalent and that $k_{OH^-}[OH^-]f_{DH^+}$ and $k'_{H_2O}f_D$ are also mathematically equivalent; therefore, Eq. 13 can be simplified to produce Eq. 16:

$$k_{obs} = k_{H^+}[H^+]f_{DH^+} + k_{H_2O}f_{DH^+} + k'_{H_2O}f_D + k'_{OH^-}[OH^-]f_D \tag{16}$$

When examining a pH–rate profile the regions of acid and base catalysis should have a slope of –1 and +1, respectively. If the slopes are closer to zero then the reaction may be pH dependent but not acid or base catalyzed. The individual rate constants can be determined manually or the pH–rate profile can be fit to an equation, such as Eq. 16, that appropriately describes $k_{obs}$ using a software program that is capable of performing nonlinear regression analysis. If a software program is used to fit the profile the data should be weighted because the $[H^+]$ can span as many as 12 orders of magnitude.

When it comes to the solid state, pH has no definition. However, differences have been observed in the solid state when proteins are prepared from solutions of different pH values. Therefore, effective pH or "pH" has been defined as the pH of a solution before it is converted to a solid (typically by lyophilization) or after the solid is reconstituted. While the physical basis of this "pH" is not yet fully understood it can be a very helpful parameter when trying to develop a solid formulation for peptides and proteins [41].

The pH– and "pH"–rate profiles play a very important role in formulation development. It provides an easy mechanism by which an appropriate formulation pH or "pH" can be identified visually and it provides an idea of how vital pH control is. For example, if a drug has a V-shaped pH–rate profile it may be vital to buffer the formulation because a small change in the pH may result in a large change in the degradation rate; however, if the drug has a U-shaped profile the formulation could be prepared near the center of the flat region and it may not require buffering since a small change in the pH will result in little to no change in the stability.

### 4.4 Buffers

As stated previously, buffers are commonly used in stability studies to maintain the pH of the solution over the course of the experiment, since the $k_{obs}$ is dependent on the concentration of the $H^+$ and $OH^-$. However, the buffer can also catalyze ($k_{cat}$) the reaction through general acid or general base catalysis; although different ionized or un-ionized states of the buffer may have more or less catalytic activity. If this occurs then more terms must be added to the $k_{obs}$ equation (Eq. 12) to account for this catalysis:

$$k'_{obs} = k_{obs} + k_{cat}[\text{Buffer}]_{total} \tag{17}$$

$$k_{cat} = k_{BH^+}f_{BH^+} + k_B f_B \tag{18}$$

In this example the buffer is a weak base with a single ionized state. When the pH is changed to study the stability at another pH, the $k_{cat}$ will also change because the fractions of the ionized ($f_{BH^+}$) and neutral ($f_B$) buffer components will change.

Therefore, it is necessary to determine the $k_{obs}$ so that the buffer catalysis does not impact the pH–rate profile. By increasing the concentration of the buffer while maintaining the pH, it is possible to extrapolate the $k_{obs}$ from a plot of $k'_{obs}$ versus the total concentration of the buffer (Fig. 5), where the $y$-intercept is $k_{obs}$ and the slope is $k_{cat}$. The $k_{obs}$ of studies at different pH values can then be combined to create a pH–rate profile. The $k_{cat}$ can also be plotted versus the fraction of one of the ionized forms of the buffer (e.g., $f_B$) to determine the $k_{BH^+}$ and $k_B$ (Fig. 6); this may help in identifying which buffer to use in a formulation by helping to determine

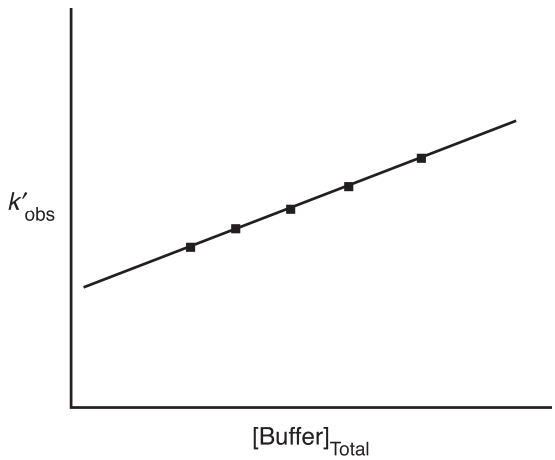


**FIGURE 5** Theoretical plot of $k'_{obs}$ versus total buffer concentration showing the effects of general acid/base catalysis on $k'_{obs}$ when buffer concentration is changed.

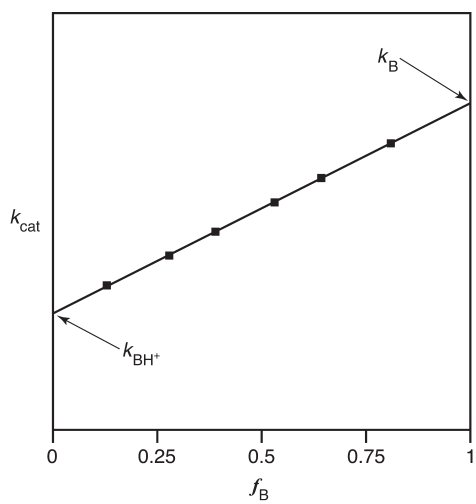


**FIGURE 6** Theoretical plot of $k_{cat}$ versus the fraction of the buffer that is in the neutral form, showing that the ionized form of the buffer is less catalytic than the neutral form.

which buffers do not catalyze the degradation or which do so only slightly. Note that ionic strength can also affect the rate of degradation; therefore, it is also important that the ionic strength be kept constant as the pH and the buffer concentration are changed.

### 4.5 Relative Humidity

In the solution state a drug is already dissolved in a solvent and thus water vapor has little to no interaction with the drug. However, solids can interact with water vapor in multiple ways. The interaction can be a physical one in which the solid takes up the water. If the solid is amorphous the water could increase the mobility of the matrix since water acts as a plasticizer to decrease the $T_g$ of amorphous materials, or if the mobility increases sufficiently the drug may be able to crystallize [20]. On the other hand, a crystalline solid, either anhydrous or hydrated, could take up water stoichiometrically to form a hydrate or a higher order hydrate, respectively. Conversely, a hydrated form could lose water to the atmosphere if the humidity drops below a particular level. Water can also be involved in chemical reactions in the solid state. In the case of a hydrolysis reaction, the water from the atmosphere could diffuse into the solid and then react with the drug. The water may also produce a physical change, such as an increase in mobility, which may subsequently lead to the increased rate of a chemical reaction due to the increased mobility of the drug molecule. Additionally, there is a possibility that the water may form a thin layer on the surface of the particles in which the drug could dissolve and then degrade [42–44].

Since water can play such a vital role in the degradation of drugs in the solid state it is vital to characterize how water vapor interacts with solid materials. This is typically performed by water vapor sorption analysis in which the material is stored at multiple relative humidities while the temperature is held constant until the material reaches equilibrium. The water content of the material stored at each relative humidity is then measured and plotted against the corresponding relative humidity to form a water vapor sorption isotherm. The isotherm provides important information concerning the stability of different polymorphs or amorphous material to relative humidity. For example, if a drug has two anhydrate forms (A and B) by comparing their isotherms at 25 °C it can be seen that at 20% relative humidity Form A takes up 1 mole of water for every mole of drug and thus forms a monohydrate, and subsequently at 75% relative humidity this monohydrate produced from Form A takes up another mole of water for every mole of drug to produce a dihydrate. Form B does not take up any significant amount of water and remains anhydrous at all relative humidities. This information can be very important when preparing to develop a formulation. If Form B could not be used in the formulation, the consequences of the formation of the mono- and dihydrates must be considered (i.e., differences in the stability, solubility, or dissolution rate of Form A and the hydrates).

Deliquescence occurs when a solid, typically amorphous, adsorbs enough water vapor to liquefy. Conversely, if water evaporates to leave behind a solid this process is called efflorescence. When these processes occur the material is not simply labeled deliquescent or efflorescent; instead the temperature and relative humidity at which the water gain or loss occurs should be reported [21].

## 5 ACCELERATED STABILITY TESTING

The degradation of drugs means that eventually they will no longer be effective because too much of the drug will have degraded and the delivered dose will be too small. Therefore, drugs are given a shelf life that indicates how long they can be stored and still be guaranteed effective. For some drugs, this lifetime is defined as the amount of time that it takes for 10% of the drug to degrade or the $t_{90}$. However, occasionally the degradation of the drug produces a toxic compound, and the amount of this toxic substance that is produced will determine the shelf life of the formulation [40, 45]. The $t_{90}$ for a first-order degradation reaction can be determined from the rate constant using

$$t_{90} = \frac{-\ln(0.9)}{k} \tag{19}$$

However, when determining if a potential formulation will be stable enough it is undesirable to wait for a year or longer to see if more than 10% of the drug has degraded. Thus, accelerated stability studies are used to stress the formulation (increased temperature and relative humidity) in an effort to predict the shelf life of the formulation under typical storage conditions. By stressing the formulation the study can be completed in several months and then, once an appropriate formulation is found, an extended stability study (~2 years) under typical storage conditions can be performed to verify the predictions.

While the principles are the same for accelerated stability testing in the solution and solid states, it is important to note that additional complications can arise in solids. For example, the drug may have polymorphs that are related enantiotropically and either some or all of the temperatures used may be above the crossover temperature; therefore, the accelerated stability testing may be done on one polymorph while the study is meant to predict the stability of the other polymorph. Also, the polymorphs could be related monotropically and some or all of the temperatures may be above the temperature at which the metastable form converts to the more stable form. However, it is not common to develop a formulation that contains a metastable polymorph, but it could be formed unintentionally in the processing steps used to produce the formulation. Additionally, it is possible that, when performing an accelerated stability study on an amorphous material, the material may be in its rubbery state and under normal storage conditions it would be in the glassy state, depending on the $T_g$ of the material. In all of these cases, the accelerated stability study would give erroneous results and would not correctly predict the stability under normal storage conditions. It would not be possible to detect these errors unless the stability is studied under normal storage conditions, which may take a year or longer. Therefore, it is vital to thoroughly characterize solid-state systems before performing accelerated stability studies [46].

### 5.1 Isothermal Kinetics

Isothermal kinetics studies rely on the relationships between temperature and rate constant (as shown previously in Eq. 8 and 10) to predict the shelf life of a formulation. These studies are termed isothermal because the temperature is held constant

to measure the rate constant of the degradation. The rate constant is measured at multiple temperatures and subsequently fit to the Arrhenius equation (Eq. 9) to obtain the $E_a$ of the reaction. The $E_a$ of the reaction can subsequently be used to predict the rate of degradation and thus the $t_{90}$ at lower temperatures using Eq. 8. This can be used to determine if the formulation is sufficiently stable or if a change must be made to increase the stability.

### 5.2 Nonisothermal Kinetics

Nonisothermal kinetics studies were first proposed by Rogers [47] as a way to simplify stability studies. In a nonisothermal kinetics study the temperature is increased throughout the course of the study. This allows the $E_a$ to be determined from only one experiment and eliminates the need for measuring the degradation rate at multiple temperatures. The degradation of a drug that follows first-order kinetics can be described by Eq. 20 [47], which is achieved by substituting Eq. 8 into Eq. 5:

$$[\mathrm{D}]_t = [\mathrm{D}]_0 \exp[-tA\exp(-E_a/RT)] \tag{20}$$

$[\mathrm{D}]_t$ and $[\mathrm{D}]_0$ are the concentrations of the drug at the time $t$ and zero, respectively; $t$ is the time; $A$ is the frequency term from the Arrhenius equation; $R$ is the gas constant; and $T$ is the temperature. However, Eq. 20 describes the isothermal degradation kinetics; King et al. [48] derived Eq. 21, which can be fit via nonlinear regression to directly estimate both $E_a$ and $t_{90}$.

$$[\mathrm{D}]_t = [\mathrm{D}]_0 \exp\left\{-t\left(\frac{0.1054}{t_{90}}\right)\exp\left[\left(\frac{E_a}{R}\right)\left(\frac{1}{298}-\frac{1}{T(t)}\right)\right]\right\} \tag{21}$$

In this equation $T(t)$ is the temperature at time $t$. One major drawback to this method is that the quality of the prediction depends on the experimental design; thus, significant care must be taken to choose the correct experiment length, sampling interval, and temperature program (rate of temperature change) in order to get accurate predictions [49]. Unfortunately, isothermal studies are needed to obtain the correct variables for accurate nonisothermal studies. In practice, there is very little, if any, gain from performing nonisothermal studies. For this reason, these studies are not routinely performed. They do occasionally experience a resurgence in popularity due to their potential to reduce the number of experiments that are required. Additionally, nonisothermal kinetics also require very accurate temperature control but have the added requirement that the oven also be able to precisely increase the temperature according to a predefined program.

## 6 IMPROVING STABILITY

There are several ways to produce a stable formulation when the drug that it contains is inherently unstable. In general, this is accomplished by either adding excipients or making chemical modifications to the drug. If the drug is moderately stable and only a slight increase in stability is needed, then excipients will probably be

sufficient, although they will only slow the degradation. However, if the molecule is extremely unstable, it will probably require changes to its chemical structure to obtain sufficient stability.

Aside from the previously mentioned excipients (e.g., buffers to control pH, desiccants to control humidity), there are many excipients that can be used to increase stability. The drug may undergo photodegradation. If so, the formulation can be protected from light by being placed in an opaque container, a photoprotective coat can be applied to the outside of a tablet, or a colorant can be added to the formulation to absorb the light. Exposure to light can also cause the formation of compounds that can oxidize drug molecules. If the drug is sensitive to oxidation, an antioxidant (e.g., α-tocopherol, butylated hydroxyanisole, ascorbic acid, thioglycolic acid) could be added to the formulation to reduce the oxidation rate of the drug. Additionally, oxygen adsorbent can be used to reduce the oxygen concentration and thereby reduce the concentration of the reactive oxygen species [1].

Some drugs can be stabilized either by forming a complex or partitioning into another environment. There are several ways that a ligand can protect a drug molecule when forming a complex: the ligand can act as a barrier that blocks a reactant from reaching the drug molecule or it can restrict the freedom of the drug molecule and hold it in a more stable conformation. Cyclodextrins, in particular, have proved to be very useful in their ability to complex with and stabilize drugs; however, cyclodextrins can be toxic and therefore cyclodextrin derivatives are typically used instead [50]. In some cases liposomes or micelles can be used in a formulation to increase the stability of a drug, either through interactions at the "oil–water" interface or by the partitioning of the drug into the lipid phase [1].

In some cases excipients are not able to adequately stabilize a drug. In these cases chemical modification of the drug will likely be required, thereby creating a new chemical entity and requiring that the entire stability testing process be repeated. One potential solution is to develop a prodrug that will be converted to the active pharmaceutical after it is administered to the patient. Additionally, it may be necessary to remove the unstable functional group in the drug. This may create problems if that functional group is also needed for the pharmacological activity; if this is the case, it may be necessary to identify a similar functional group that will still result in the necessary pharmacological activity but is less reactive. Even if the unstable functional group is removed, the drug will almost certainly still degrade, but the degradation would likely be much slower, and occur through another mechanism [1].

## 7 METHODS OF ANALYSIS

When performing stability studies it is necessary to use an analytical method that is capable of determining the extent of the reaction by following either the disappearance of the drug or appearance of its degradation products, or ideally both. This requires a method that can identify and quantify either the drug or its products in a mixture containing the drug, all of its degradation products, excipients, and any impurities [51].

Chemical stability is commonly measured using a chromatography technique, such as HPLC, to separate the reactants and products and is coupled with a detec-

tion method for identification and quantification, including electromagnetic spectroscopy (e.g., UV or visible light absorption), mass spectrometry, and electrochemical detection [51]. The only difference between studying the solution and solid states is that the solid must be dissolved before analysis. Since the studies typically can go for long periods of time (days to months), it is not uncommon to store the samples until the end of the experiment and then perform the analysis of all of the samples at once to help reduce the day-to-day variability of the instruments. The samples are typically stored by freezing; however, it is imperative that a control be performed to verify that no changes occur in the samples when they are frozen [52].

Physical stability requires a much broader set of techniques. For peptides and proteins there are several spectroscopic techniques that can be used to follow changes in the secondary and tertiary structure. Circular dichroism and infrared spectroscopy can be used to follow changes in the secondary structure, while intrinsic fluorescence can be used to follow changes in the tertiary structure. UV-Vis spectroscopy and DSC can be used to monitor denaturation [10]. Light scattering techniques and size exclusion chromatography are available to analyze aggregation.

When it comes to solids there are several techniques available to identify and track physical changes. X-ray diffractometry can be used to determine the crystal structure of a crystalline material, and to identify and quantitate polymorphs, solvates, and amorphous materials [21, 53]. Hot-stage microscopy can be used to follow changes in solids during heating by changes in birefringence [21, 54]. Thermogravimetric analysis (TGA) can be used to determine water or other volatile content (identify solvates and determine their stochiometry) and to follow stability to thermal decomposition to form gaseous products [21]. Additionally, Karl Fischer titration can be used to determine water content. DSC can provide a variety of information including melting points of polymorphs, $T_g$ of amorphous material [20], and identification of monotropic versus enantiotropic systems [21]. Infrared and Raman spectroscopy can be utilized to identify and quantitate polymorphs and solvates [21]. Water vapor sorption can be used to characterize how different solid forms (polymorphs, solvates, and amorphous materials) interact with atmospheric water. Solid-state NMR spectroscopy can provide mobility information (such as $T_g$) and identify polymorphs, solvates, and amorphous materials [55]. Solid-state NMR spectroscopy is gaining popularity for several reasons, including that it is nondestructive, that it can often differentiate the drug from the excipients in a formulation when other techniques cannot, and that it can be used to study many dosage forms *in situ* [21]. Solid-state NMR spectroscopy can also provide crystal structure information, such as the number of crystallographically inequivalent sites [56].

## 8 REGULATORY GUIDELINES

In the pharmaceutical industry, stability testing is typically performed according to guidelines established by the International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH). The ICH groups "stability" among its "Quality Topics" and thus identifies each individual "Quality" guideline with a Q-code. Some of the relevant guidances are shown in Table 3 and are available online at http://www.ich.org.

**TABLE 3 ICH Stability Guidelines**

| Q-Code | Topic |
|---|---|
| Q1A(R2) | Stability Testing of New Drug Substances and Products |
| Q1B | Stability Testing: Photostability Testing of New Drug Substances and Products |
| Q1C | Stability Testing for New Dosage Forms |
| Q1D | Bracketing and Matrixing Designs for Stability Testing of New Drug Substances and Products |
| Q1E | Evaluation of Stability Data |
| Q1F | Stability Data Package for Registration Applications in Climatic Zones III and IV |

There are also national and regional guidelines on stability testing. In the United States, the U.S. Food and Drug Administration (FDA) is the regulatory body and the FDA's Center for Drug Evaluation and Research (CDER) issues stability guidances. These documents can be found online at http://www.fda.gov/cder/guidance/index.htm and http://www.fda.gov/cber/guidelines.htm.

## 9 FURTHER READING

Stability is a very broad topic that has been widely studied for many years due to the need to understand the degradation of every potential pharmaceutical before being approved by the FDA, and every new compound offers its own new and unique challenges. There have been countless articles and numerous books written on the subject; thus, it is impossible to condense all of this information into one chapter. The following articles and books are good resources for further information about specific topics that have been discussed in this chapter.

Yoshioka and Stella [1] have written a book that provides extensive information regarding the stability of drugs (both small molecules and biopharmaceuticals) primarily in the solution state. They provide many examples for each of the possible degradation mechanisms and present a thorough discussion of kinetics and the factors that affect the chemical stability of drugs. Additionally, they include the ICH guidelines on stability and photostability testing of new drug substances and products (Q1A and Q1B) and the major concerns that have been raised by the European Union, the United States, and Japan regarding these guidelines. However, note that the ICH guidelines are revised occasionally. Therefore, it is best to refer to the most up-to-date copies, which are available at the ICH website (www.ich.org).

Byrn, Pfeiffer, and Stowell [21] wrote a useful book concerning drugs in the solid state. In the book they provide a good introduction to solid-state chemistry along with some of the more common methods of analysis that are used to study solids: X-ray diffractometry, microscopy, DSC, TGA, solubility, dissolution, particle-size analysis, infrared spectroscopy, and solid-state NMR spectroscopy. They include extensive discussions and examples of both crystalline (polymorphs, hydrates, and solvates) and amorphous materials and the physical and chemical changes that can occur within them, complete with a discussion about solid-state kinetics.

In addition to the two books discussed above there are also many books and review articles that have been written about stability but are more focused on specific issues. The photodegradation of drugs and formulations is described in a book

edited by Tønnesen [4], including information on how photostability testing is performed, how to treat and interpret data from photostability tests, and many other issues concerning the exposure of drugs to light. The book series entitled Pharmaceutical Biotechnology [11] contains several volumes concerning the stability and characterization of proteins and peptides. Additionally, Chi et al. [17] wrote a review paper discussing the physical stability of proteins in solution and the mechanism and driving forces for aggregation. There is another book series, Drugs and the Pharmaceutical Sciences, which also contains several volumes with very helpful information regarding the stability of drugs [12, 14, 18]. Connors, Amidon, and Kennon [57] wrote a book in which they discuss the kinetics of degradation and provide stability monographs of 30 drugs, including acetaminophen, aspirin, and vitamin A. There are also several review articles that discuss amorphous materials [19, 20], crystalline materials [53], polymorphism [58], the effects of water on drugs in the solid state [42, 44], and accelerated aging studies [40]. Finally, there are countless articles that have been published about the stability of specific compounds that may help to understand the degradation of a new drug. Therefore, a good stability study always begins with a thorough literature search.

## ACKNOWLEDGMENTS

The authors thank Dr. Valentino J. Stella for his discussions regarding the stability of drugs in the solution state and for informing the authors of several of the examples provided in this chapter. The authors also thank Dr. Dewey H. Barich for his help in reviewing the manuscript for this chapter. EMG was supported by a Biotechnology Training Grant (NIH 29488) from the National Institute of General Medical Sciences. This material is based on work supported by the National Science Foundation under Grant No. 046214. Any opinions, findings, and conclusions or recommendations expressed in this chapter are those of the authors and do not necessarily reflect the views of the National Science Foundation.

## REFERENCES

1. Yoshioka S, Stella VJ. *Stability of Drugs and Dosage Forms*. New York: Kluwer Academic/Plenum Publishers; 2000, p 268.
2. Hoener BA, Sokoloski TD, Mitscher LA, Malspeis L. Kinetics of dehydration on epitetracycline in solution. *J Pharm Sci* 1974;63:1901–1904.
3. Sokoloski TD, Mitscher LA, Juvarkar JV, Hoener B. Rate and proposed mechanism of anhydrotetracycline epimerization in acid solution. *J Pharm Sci* 1977;66:1159–1165.
4. Moore DE. Standardization of photodegradation studies and kinetic treatment of photochemical reactions. In Tønnesen HH, Ed. *Photostability of Drugs and Drug Formulations*. Bristol, PA: Taylor & Francis; 1996, pp 63–82.
5. Greenhill JV, McLelland MA. Photodecomposition of drugs. In Ellis GP, West GB, Eds. *Progress in Medicinal Chemistry. Progress in Medicinal Chemistry Series*, Vol 27. Amsterdam: Elsevier Science; 1990, pp 51–121.
6. Bundgaard H. Polymerization of penicillins: kinetics and mechanism of di- and polymerization of ampicillin in aqueous solution. *Acta Pharma Suecica* 1976;13(1):9–26.

7. Bundgaard H, Larsen C. Piperazinedione formation from reaction of ampicillin with carbohydrates and alcohols in aqueous solutions. *Int J Pharm* 1979;3:1–11.
8. Pogocki D, Schoneich C. Chemical stability of nucleic acid-derived drugs. *J Pharm Sci* 2000;89(4):443–456.
9. Oliyai C, Schoneich C, Wilson GS, Borchardt RT. Chemical and physical stability of protein pharmaceuticals. *Top Pharm Sci 1991: Proc 51st Int Congr Pharm Sci FIP* 1992;23–46.
10. Krishnamurthy R, Manning MC. The stability factor: importance in formulation development. *Curr Pharm Biotechnol* 2002;3(4):361–371.
11. Meyer JD, Ho B, Manning MC. Effects of conformation on the chemical stability of pharmaceutically relevant polypeptides. In Carpenter JF, Manning MC, Eds. *Rational Design of Stable Protein Formulations. Pharmaceutical Biotechnology Series*, Vol 13. New York: Kluwer Academic/Plenum Publishers; 2002, pp 85–107.
12. McNally EJ, Lockwood CE. The importance of a thorough preformulation study. In McNally EJ, Ed. *Protein Formulation and Delivery. Drugs and the Pharmaceutical Sciences Series*, Vol 99. New York: Marcel Dekker; 2000, pp 111–138.
13. Wang W. Lyophilization and development of solid protein pharmaceuticals. *Int J Pharm* 2000;203(1–2):1–60.
14. Lucks J-S, Müller BW, Klütsch K. Parenteral fat emulsions: structure, stability, and applications. In Nielloud F, Marti-Mestres G, Eds. *Pharmaceutical Emulsions and Suspensions. Drugs and the Pharmaceutical Sciences Series*, Vol 105. New York: Marcel Dekker; 2000, pp 229–257.
15. Heurtault B, Saulnier P, Pech B, Proust J-E, Benoit J-P. Physico-chemical stability of colloidal lipid particles. *Biomaterials* 2003;24(23):4283–4300.
16. Yoshioka S, Aso Y, Izutsu K, Terao T. Aggregates formed during storage of β-galactosidase in solution and in the freeze-dried state. *Pharm Res* 1993;10(5):687–691.
17. Chi EY, Krishnan S, Randolph TW, Carpenter JF. Physical stability of proteins in aqueous solution: mechanism and driving forces in nonnative protein aggregation. *Pharm Res* 2003;20(9):1325–1336.
18. Guillory JK, Poust RI. Chemical kinetics and drug stability. In Banker GS, Rhodes CT, Eds. *Modern Pharmaceutics. Drugs and the Pharmaceutical Sciences Series*, Vol 121. New York: Marcel Dekker; 2002, pp 139–166.
19. Kaushal AM, Gupta P, Bansal AK. Amorphous drug delivery systems: molecular aspects, design, and performance. *Crit Rev Ther Drug Carrier Syst* 2004; 21(3):133–193.
20. Craig DQM, Royall PG, Kett VL, Hopton ML. The relevance of the amorphous state to pharmaceutical dosage forms: glassy drugs and freeze dried systems. *Int J Pharm* 1999;179(2):179–207.
21. Byrn SR, Pfeiffer RR, Stowell JG. *Solid-State Chemistry of Drugs*, 2nd ed. West Lafayette, IN: SSCI, Inc; 1999, p 576.
22. Burger A, Ramberger R. On the polymorphism of pharmaceuticals and other molecular crystals. I. Theory of thermodynamics rules. *Mikrochim Acta* 1979;72:259–271.
23. Burger A, Ramberger R. On the polymorphism of pharmaceuticals and other molecular crystals. II. Applicability of thermodynamic rules. *Mikrochim Acta* 1979;72:273–316.
24. Watanabe T, Wakiyama N, Kusai A, Senna M. Drug–carrier interaction in solid dispersions prepared by co-grinding and melt-quenching. *Ann Chim (Cachan, France)* 2004;29(1):53–66.
25. Hancock BC. Disordered drug delivery: destiny, dynamics and the Deborah number. *J Pharm Pharmacol* 2002;54(6):737–746.

26. Chemburkar SR, Bauer J, Deming K, Spiwek H, Patel K, et al. Dealing with the impact of ritonavir polymorphs on the late stages of bulk drug process development. *Org Process Res Dev* 2000;4(5):413–417.
27. Pikal MJ, Rigsbee DR. The stability of insulin in crystalline and amorphous solids: observation of greater stability for the amorphous form. *Pharm Res* 1997;14(10):1379–1387.
28. Cohen MD, Schmidt GMJ. Topochemistry. I. A survey. *J Chem Soc* 1964;June: 1996–2000.
29. Cohen MD, Schmidt GMJ, Sonntang FI. Topochemistry. II. The photochemistry of trans-cinnamic acids. *J Chem Soc* 1964;June:2000–2013.
30. Prout EG, Tompkins FC. Thermal decomposition of $KMnO_4$. *Trans Faraday Soc* 1944;40:488–498.
31. Brown ME. The Prout–Tompkins rate equation in solid-state kinetics. *Thermochim Acta* 1997;300:93–106.
32. Avrami M. Kinetics of phase change. II. Transformation–time relations for random distribution of nuclei. *J Chem Phys* 1940;8:212–224.
33. Avrami M. Kinetics of phase change. III. Granulation, phase change and microstructure. *J Chem Phys* 1941;9:177–184.
34. Erofe'ev BV. Generalized equation of chemical kinetics and its application in reactions involving solids. *C R Acad Sci. USSR* 1946;52:511–514.
35. Sharp JH, Brindley GW, Narahari Achar BN. Numerical data for some commonly used solid state reaction equations. *J Am Ceram Soc* 1966;49:379–382.
36. Holt JB, Cutler IB, Wadsworth ME. Rate of thermal dehydration of kaolinite in vacuum. *J Am Ceram Soc* 1962;45:133–136.
37. Ginstling AM, Brounshteĭn BI. The diffusion kinetics of reactions in spherical particles. *J Appl Chem USSR Engl Trans* 1950;23:1327–1338.
38. Valensi G. Kinetics of oxidation of metallic spherules and powders. *C R Acad Sci USSR* 1936;202:309–312.
39. Carter RE. Kinetic model for solid-state reactions. *J Chem Phys* 1961;34:2010–2015.
40. Waterman KC, Adami RC. Accelerated aging: prediction of chemical stability of pharmaceuticals. *Int J Pharm* 2005;293(1–2):101–125.
41. Lai MC, Topp EM. Solid-state chemical stability of proteins and peptides. *J Pharm Sci* 1999;88(5):489–500.
42. Stewart PJ, Tucker IG. Prediction of drug stability. Part 5: Physical stability. *Aust J Hosp Pharm* 1985;15(4):236–246.
43. Ahlneck C, Zografi G. The molecular basis of moisture effects on the physical and chemical stability of drugs in the solid state. *Int J Pharm* 1990;62(2–3):87–95.
44. Airaksinen S, Karjalainen M, Shevchenko A, Westermarck S, Leppänen E, et al. Role of water in the physical stability of solid dosage formulations. *J Pharm Sci* 2005;94: 2147–2165.
45. Stella VJ. Chemical and physical bases determining the instability and incompatiblity of formulated injectable drugs. *J Parent Sci Technol* 1986;40(4):142–163.
46. Lachman L. Physical and chemical stability testing of tablet dosage forms. *J Pharm Sci* 1965;54(10):1519–1526.
47. Rogers AR. An accelerated storage test with programmed temperature rise. *J Pharm Pharmacol* 1963;15(Suppl):101T–105T.
48. King S-YP, Kung M-S, Fung H-L. Statistical prediction of drug stability based on nonlinear parameter estimation. *J Pharm Sci* 1984;73(5):657–662.

49. Yoshioka S, Aso Y, Uchiyama M. Statistical evaluation of nonisothermal prediction of drug stability. *J Pharm Sci* 1987;76(10):794–798.
50. Albers E, Müller BW. Cyclodextrin derivatives in pharmaceutics. *Crit Rev Ther Drug Carrier Syst* 1995;12(4):311–337.
51. Mehta AC. Analytical issues in the chemical stability testing of drugs in solution. *Anal Proc (London)* 1995;32(2):67–70.
52. Chilamkurti RN. Formulation development of frozen parenteral dosage forms. *J Parent Sci Technol* 1992;46(4):124–129.
53. Datta S, Grant DJ. Crystal structures of drugs: Advances in determination, prediction and engineering. *Nat Rev Drug Discov* 2004;3(1):42–57.
54. Windram VA, Threlfall TL. Chemical microscopy in the pharmaceutical industry. *Anal Proc (London)* 1992;29(3):108–110.
55. Bugay DE. Solid-state nuclear magnetic resonance spectroscopy: theory and pharmaceutical applications. *Pharm Res* 1993;10(3):317–327.
56. Ripmeester JA. Application of solid state $^{13}C$ NMR to the study of polymorphs, clathrates and complexes. *Chem Phys Lett* 1980;74(3):536–538.
57. Connors KA, Amidon GL, Kennon L. *Chemical Stability of Pharmaceuticals. Hoboken*. NJ: Wiley; 1979, p 367.
58. Haleblain J, McCrone W. Pharmaceutical applications of polymorphism. *J Pharm Sci* 1969;58(8):911–929.

# DOSAGE FORMULATION

ALEXANDER V. LYUBIMOV
*University of Illinois at Chicago, Chicago, Illinois*

**Contents**

1 Introduction
  1.1 Background
  1.2 Objectives/Definitions
2 General Principles of Dosage Formulation in Preclinical Studies
  2.1 Dosage Formulation Preparation
  2.2 Role of the Route of Administration
  2.3 Role of Absorption/Enhancement of Oral Absorption
  2.4 Role of Animal Species
  2.5 Protein/Peptide, DNA, Antisense Oligonucleotides (AS-ONs), siRNA, and Other Biological Formulations
  2.6 Enhanced Delivery and Semisolid or Sustained Release Preparations (Micronization and Nanoparticles)
  2.7 Vehicles and Selection of Excipients
3 Types and Characteristics of Dosage Formulations by Route of Administration
  3.1 Oral Formulations
  3.2 Intravenous and Other Parenteral Formulations
  3.3 Alternatives to Parenteral Administration (Nasal, Buccal, Transdermal, Rectal, and Vaginal)
  3.4 Topical Formulations for Local Disorders (Dermal, Ocular, Ear)
  3.5 Inhalation Formulations (Vapors and Aerosols)
4 Conclusion
  References

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

# 1 INTRODUCTION

## 1.1 Background

Advances in biotechnology have allowed the economical and large scale production of therapeutically important complex polymers of amino acids (peptides and proteins), nucleosides (antisense molecules), and other biological drugs to be used to combat poorly controlled diseases [1]. This rapid progress in molecular biology required substantial progress in the formulation and development of delivery systems for the next generation of drugs. Rational drug design does not necessarily mean rational drug delivery, which strives to incorporate into a molecule the properties necessary for optimal transfer between the site of administration and the pharmacological target site in the body [2]. In addition, the advent of high throughput screening has affected contemporary drug pipelines based in large part on the nature of the assay, which includes the evaluation of organic solutions (e.g., DMSO) of drug candidates in immobilized receptor matrices. As a result, new drug leads tend to be more lipophilic. On top of this, the hit to lead process reduces still further aqueous solubility, meaning that the formulator is often faced with significant challenges with regard to generating appropriately orally bioavailable dosage forms. Various new vehicles and delivery systems, such as liposomes, biodegradable microspheres, biopolymers, hydrogels, and bioavailability enhancers, have been introduced during the last decades to improve delivery and bioavailability of new chemical and biological drugs. Safety evaluation of new vehicles and excipients as well as their role on efficacy and toxicity of new drugs is an important part of new formulations assessment.

Drug formulation is a rapidly evolving field. The new formulations, their function and characterization, absorption mechanisms, and novel routes and modes of delivery are described in many journals including (but obviously not limited to) the following journals: *International Journal of Pharmaceutics* (http://www.sciencedirect.com/science/journal/03785173); *AAPS PharmSciTech; Journal of Control Release* (http://www.sciencedirect.com/science/journal/01683659); *AAPS PharmSciTech* (http://www.aapspharmscitech.org), an online-only journal of the American Association of Pharmaceutical scientists; *European Journal of Pharmaceutical Sciences* (http://sciencedirect.com/science/journal/092809870); *European Journal of Pharmaceutics and Biopharmaceutics* (http://www.sciencedirect.com/science/journal/09396411); and *Advanced Drug Delivery Reviews* (http://www.sciencedirect.com/science/journal/0169409X).

## 1.2 Objectives/Definitions

Formulation is a preparation of pharmaceutical product in various dosage forms containing a new active ingredient(s) (AI) [3]. The purpose of drug formulation is to determine experimentally all the variables necessary to develop an optimal formula and working directions for making the pharmaceutical product. Pharmaceutical technology applied to developing a drug formulation comprises selection of materials and procedures that are adaptable to various processes that lend themselves to inclusion in specific dosage forms (tablets, injections, capsules, ointments, etc.). Dosage formulations can include sterile products, tablets, capsules, topical

preparations, and nonconventional products such as semisolid or sustained release preparations. Formulations can be developed to enhance drug solubility, improve bioavailability, enhance drug stability, target tissue delivery, and treat specific patient populations. Stability and release properties of formulations can be evaluated *in vitro* prior to their use in animals.

Clinical drug formulation is a very complex multistep process that is significantly simplified at the stage of preclinical dosage formulations. A dosage form for preclinical study should optimize chances for a compound to exert its pharmacological action, determine the ways of solubilizing the compound, and determine photosensitivity and other aspects of stability. In a preclinical toxicity study, it is important that the formulation does not change the toxic effect of the active ingredient (either mask or enhance it). It is very important because in many cases the final clinical formulations may be quite different in composition from the dosage forms used for safety assessment in animals.

*Stability* is one of the most important factors that determine whether a compound or mixture of compounds can be developed in a therapeutically useful product. The stability of a dosage formulation can be defined as its degree of resistance to chemical and physical changes. The efficacy of the preparation must remain constant (or changed only within the specified limits) until the date of expiration. The formulation must be chemically, physically, microbiologically, therapeutically, and toxicologically stable.

Drug stability may be influenced by internal and external factors.

*Internal Factors.* Potential drug–drug (when mixtures are used) or drug–excipient interactions.

*External Factors.* The test articles (drugs) and dosage formulations come into contact with containers or packaging materials. For example, alkali from the glass containers, heavy metals from the metal containers, components of cork or rubber stopper, and plastics can penetrate and react with preparations. Various storage conditions may indicate three types of changes: (1) physical and colloidal–physical changes, (2) chemical changes, and (3) microbiological changes. Proper storage conditions should be determined and followed to minimize these changes.

There are numerous types of potential undesired interactions between the drug and excipients, which can occur during the process of formulation preparation, storage, and administration. For example, solubility, ionization pH effects, formation of complexes, and degradation reactions are possible. These changes may be *intentional* (e.g., prolonged release or enhanced delivery) or *unintentional* (e.g., discoloration, precipitation, liquefaction). The second group is usually referred to as incompatibilities. Incompatibilities are undesired physical, chemical, colloidal, or biopharmaceutical processes that take place during preparation, storage, or administration of the drug formulation. Incompatibilities may take place in visible or hidden manners. Visible incompatibilities include the following processes: (1) changes in solubility or degree of dispersion, turbidity, precipitation, coagulation, aggregation, crystallization, or crystal growth; (2) changes involving viscosity, physical state, crystal form, liquefaction, solidification, separation of phases, and uptake or loss of moisture; and (3) discoloration or change in odor or taste

due to chemical transformation such as oxidation, hydrolysis, or enzyme degradation [3].

Hidden incompatibilities may be the result of interactions between active ingredients and excipients or packing materials and environmental conditions and include chemical, physical, physicochemical, colloidal, technological, and biopharmaceutical interactions. Some typical chemical interactions include spatial rearrangements, hydrolysis, the formation of esters, oxidation and reduction, and polymerization and depolymerization. Incompatibilities may also be related to the electric charge of molecules (e.g., formation of new salts, double salts, and intermolecular associations and complexes as a result of the interaction of AI and excipients, which have opposite charge). Physical interactions include adhesion, absorption, molecular adhesion, and ionic absorption.

To assess stability of the dosage formulation in a preclinical safety study, usually purity of active ingredient in formulation is determined over time. A formulation is considered unstable when the purity of the active ingredient (AI) decreases more than 10%. This may depend on accuracy of the analytical method and type of formulation (e.g., diet formulations may be considered stable up to a 15–20% decrease of AI). When chromatographic methods are used for purity determination, the appearance of additional peaks may be another sign of AI decomposition over time. Stability of formulation under the storage and application conditions should be determined prior to any preclinical safety assessment. It is not acceptable to simply assume that the drug will be stable in the diet or ointment even if it is known to be stable for a long period of time at room temperature in a powder form. Mixing with the diet (for oral studies) or with "inert" excipient (e.g., Vaseline for dermal administration) or even with water can significantly shorten stability of very stable drugs. Stability at storage conditions of formulations determined prior to toxicological experiment will dictate how often formulations need to be prepared in a repeated dosing study. Stability at room temperature is important to know even in a single dose experiment to assure formulation stability throughout the time of administration, which may be long when a significant number of small animals per dose group are used (especially important for dermal applications and lengthy infusions).

## 2 GENERAL PRINCIPLES OF DOSAGE FORMULATION IN PRECLINICAL STUDIES

### 2.1 Dosage Formulation Preparation

The test article is rarely administered as a neat compound and in most cases a dosage formulation is prepared for dosing the animals. It may be as simple as a gelatin capsule filled with the drug for oral administration in dogs or a solution in water for gavage dosing of rodents. It can also be a complicated mixture of the drug, vehicle(s), surfactants, and so on. The desirable characteristics of the "ideal" formulation and its preparation are presented in Table 1. It is important to test solubility of dosage solutions or homogeneity/suspendability of suspensions prior to the dosing of animals. We suggest that the solubility/suspendability test be run with the proposed formulations several days prior to dosing. This will allow for adjustments of a formulation, if necessary. An example of the form to be completed to record the result of these tests is presented in Table 2.

**TABLE 1 Desirable Characteristics of a Dosing Formulation and Its Preparation**

1. Preparation of the formulation should not involve heating of the test material anywhere near to the point where its chemical or physical characteristics are altered.
2. If the material is a solid and it is to be assessed for dermal effects, its shape and particle size should be preserved.
3. Multicomponent test materials (mixtures) should be formulated so that the administered form accurately represents the original mixture (i.e., components should not be selectively suspended or taken into solution).
4. Formulation should preserve the chemical stability and identity of the test material.
5. Formulation should be such as to minimize total test volume. Use just enough solvent or vehicle.
6. Formulation should be easy to administer accurately.
7. The pH of dosing formulations should be between 5 and 9, if possible.
8. Acids or bases should not be used to divide the test material (for both humane reasons and to avoid pH partitioning in either the gut or the renal tubule).
9. If a parenteral route is to be employed, final solutions should be as nearly isotonic as possible.

*Source*: From Ref. 4, with permission.

The test substance can be administered to animals at a constant concentration across all dose levels (i.e., varying the dose volume) or at a constant dose volume (i.e., varying the dose concentration). The toxicity observed by administration of a constant concentration may be different from that observed for a constant dose volume. For instance, when a large volume of oil is given orally, gastrointestinal motility is increased, causing diarrhea and decreasing the time available for absorption of the test substance in the GI tract.

The volume of water or solvent–water mixture used to dissolve the chemical should be kept low, since excess quantities may distend the stomach and cause rapid gastric emptying. In addition, large volumes of water may carry the chemical through membrane pores and increase the rate of absorption. Thus, if dose-dependent absorption is suspected, it is important that the different doses are given in the same volume of solution [4].

Local irritation by a test substance generally decreases when the material is diluted. If the objective of the study is to establish systemic toxicity, the test substance should be administered in a constant volume to minimize gastrointestinal irritation that may, in turn, affect its absorption. If, however, the objective is to assess the irritation potential of the test substance, then it should be administered undiluted.

## 2.2 Role of the Route of Administration

The intended route of administration in humans dictates a route of administration in animals during preclinical studies. However, the drug formulations in animals are often different from the clinical formulations. The specific route of administration is a significant determinant of the commonly used formulations in preclinical testing. Specific types of formulations by route of administration are described in Section 3.

Generally, pharmaceutical companies prefer to develop an oral route of administration for a new chemical entity. Dosage forms associated with this route are

**TABLE 2 Form for Preclinical Formulation Testing for Solubility/Suspensibility**

SOLUBILITY/SUSPENSION TESTING

Chemical: _____

Performed By: _____

Date Performed: _____

The test article was mixed with the following vehicle(s):

| Vehicle | Volume of Vehicle | Quantity of Test Article | Results |
|---|---|---|---|
| Water | | | |
| Corn Oil | | | |
| Peanut Oil | | | |
| | | | |
| | | | |
| | | | |

Since the test article was not soluble in the above vehicles, it was suspended in the following vehicle(s):

| Vehicle | Volume of Vehicle | Quantity of Test Article | Results |
|---|---|---|---|
| | | | |
| | | | |
| | | | |

A = Test article was completely soluble.

B = Test article appeared partially soluble.

C = Test article appeared insoluble.

D = Test article was adequately suspended.

E = Test article was not adequately suspended.

— = Vehicle not utilized.

Comments: _____

_____

_____

Read and Reviewed by: _____ Date: _____

relatively cheap to manufacture, are simpler for patients to handle, and are widely accepted by the public. Because this is the preferred route of administration, it becomes crucial to evaluate the oral bioavailability of lead compounds in the earlier stage of the development process. Oral bioavailabilty of a compound can be a major selection criterion for the development candidate. Drawbacks of the low bioavailable compound from a pharmaceutical development standpoint are inherent variability in the oral bioavailablity, a requirement of large quantities of drug substance to establish a safety margin in preclinical safety studies, and, depending on the pharmacology of the molecule, a high dose to elicit the desired effect.

In contrast, with only a few exceptions, most new anticancer compounds are initially developed for intravenous (IV) use, despite some drawbacks such as the morbidity associated with gaining IV access; risk of IV catheter-related infection, thrombosis, and extravasation; and patients' preference for oral therapy when equally effective [5]. Important reasons for choosing IV use for initial drug development are the fact that usually less gastrointestinal toxicity occurs, there is immediate 100% bioavailability and instantaneous pharmacodynamic effects, and there is a possibility to modify the dosing rate or even halt the infusion if necessary. Solubility of the compound is a specific demand for IV administration, even for the newer chemotherapeutic agents, which are known to be poorly water soluble. Classical solubility approaches, which are discussed subsequently, include the use of colloidal systems, prodrug development, or solubilization techniques.

Colloidal systems such as liposomes, microcapsules, microspheres, nanoparticles, or macromolecule complexes may protect the anticancer drug from premature degradation or (chemical) inactivation within the systemic circulation. Prodrugs are inactive derivatives that release the active drug following spontaneous degradation or enzymatic reactions. Solubilization is the process of uptake of drugs through complex formation into, for example, oligomers of dextrose and fatty acids, through cosolvent systems (such as ethanol, polyethylene glycol, and glycerol), or through surfactant systems. The surfactant systems consist of amphoteric compounds (e.g., lecithin or gelatin), ionic surfactants (e.g., sodium palmitate), or nonionic surfactants (e.g., Tween 80 and Cremophor EL (CrEL)). Some of these solubilization approaches are discussed later in this chapter.

### 2.3 Role of Absorption/Enhancement of Oral Absorption

In the early stages of development, it is important to determine whether the dosage form targeted for clinical use has an acceptable level of bioavailability. To evaluate the dosage form in a clinical study is time consuming and costly. Large animals (e.g., dogs and monkeys) are used for assessment of bioavailability of different dosage formulations. Examples of the usefulness of such studies are presented in this volume in Chapter 32, Bioavailability and Bioequivalence Studies.

Oral administration is the most convenient mode of drug delivery and is associated with superior patient compliance as compared to other modes of drug intake. However, oral administration has only limited use for important drugs, from various pharmacological categories, that have poor oral bioavailability due to incomplete absorption and/or degradation in the gastrointestinal (GI) tract [6]. Some of these drugs are characterized by a narrow absorption window (NAW) at the upper part of the GI tract. This is because the proximal part of the small intestine exhibits

extended absorption properties (including larger gaps between the tight junctions, and dense active transporters). Despite the extensive absorption properties of the duodenum and jejunum, the extent of absorption at these sites is limited because the passage through this region is rapid. Enhancing the gastric residence time (GRT) of a NAW drug may significantly improve the net extent of its absorption. To further increase the GRT of drugs, a gastroretentive dosage form (GRDF) can be developed. It is quite complex to achieve extensive retention of the GRDF since the natural activity of the stomach is to evacuate its contents into the intestine. The development of the GRDF has generated enormous interest and the pharmaceutical aspects of these developments are reviewed elsewhere [7, 8]. Controlled release (CR) dosage forms have been extensively used to improve therapy of many important medications. However, in the case of NAW drugs this pharmaceutical approach cannot be utilized since it requires sufficient colonic absorption of the drug (which is, by definition, not the case for NAW agents). On the other hand, incorporation of the drug in a controlled release gastroretentive dosage form (CR-GRDF) can yield significant therapeutic advantages due to a variety of pharmacokinetic (PK) and pharmacodynamic (PD) factors. Hoffman et al. [9] reviewed these aspects in rat and dog animal models in order to suggest rational selection of drugs for which CR-GRDF would be a beneficial strategy. These authors found that a CR-GRDF formulation was superior compared to other modes of administration for levodopa and riboflavin, but not for metformin.

Incomplete oral bioavailability has various causes. These include poor dissolution or low aqueous solubility, degradation of the drug in gastric or intestinal fluids, poor intestinal membrane permeation, and presystemic intestinal or hepatic metabolism. Absorption of peptides, peptide analogues, or other polar, high molecular weight drugs may be significantly enhanced by using excipients that increase intestinal permeability. For some compounds, permeation through the intestinal epithelium is hindered by their active transport from the enterocyte back into the intestinal lumen. The secretory transporters involved may include P-glycoprotein (P-gp), the family of multidrug resistance-associated proteins (MRPs), and possibly others. For substrates of these secretory transporters, inhibiting secretory transport can increase permeation in the absorptive direction. This can be accomplished with pharmacologically inactive excipients.

Before permeation-enhancing excipients are evaluated for a particular drug of interest, it is important to understand the cause of its low oral bioavailability. Ideally, first-pass metabolism and solubility are not limiting oral bioavailability. Furthermore, the role of secretory transport should be assessed, because excipients that increase absorption by inhibiting secretory transport are different from those that increase absorption by other mechanisms.

Several reviews address intestinal permeation-enhancing excipients and evaluate how far some of the more advanced technologies have come toward fulfilling these criteria [6, 9–13].

### 2.4 Role of Animal Species

The key issue of an assessment of new dosage formulations in animal studies is to extrapolate the results to humans. The bioavailability of drug formulation in animals and the ability to predict its bioavailability in humans is one of the most important

factors. An important issue in determining the choice of animal species for bioavailability testing is its size. In most circumstances, it is not possible to administer intact, solid pharmaceutical formulations—tablets and capsules, for example—in small rodents such as mice, rats, or hamsters. If such dosage forms are to be tested, larger species, such as rabbits, dogs, and monkeys, are more appropriate.

Another problem associated with the use of smaller species is repetitive blood sampling. Urine measurements of drug excretion can be used for all animal species to determine bioavailability, overcoming many of the blood sampling difficulties. A reasonable proportion of the drug, however, must be excreted unchanged via kidney to make the results meaningful. In addition, extensive first-pass metabolism would preclude this approach if the total metabolite excretion is being used.

None of the animal models are ideal, and each has its advantages and disadvantages. Table 3 compares different animal species as a model for human bioavailability testing [14]. Nevertheless, many investigators have used animals for bioavailability testing and have obtained valuable information that has helped in their understanding of their product. The rabbit appears to be the least suitable animal model for bioavailability testing, despite its relatively large size (large enough to test solid formulations) and low cost. This is because it has a long gastric emptying time, and both its gastrointestinal pH and microflora bear little resemblance to those in humans. Table 4 presents gastric and intestinal pH values of common laboratory animals. The rat has the major disadvantage of being too small to test most solid

**TABLE 3 Advatages (+) and Disadvantages (–) of Different Animal Species as a Model for Humans in Bioavailability Testing**

| Parameters | Rat | Rabbit | Dog | Monkey |
|---|---|---|---|---|
| Cost | + | + | – | – |
| Ease of urine collection | + | + | + | – |
| Ease of blood sampling | – | + | + | + |
| Ease of handling | + | + | + | – |
| Intact solid formulation | – | + | + | + |
| Intestinal pH | + | – | + | + |
| Gastrointestinal transit time | – | – | – | – |
| Gastrointestinal microflora | – | – | + | + |
| Drug metabolism | – | (– +) | (– +) | + |
| Biliary excretion | – | + | – | + |

*Source*: From Ref. 14, with permission.

**TABLE 4 Gastric and Intestinal pH Values for Different Animal Species**

| | pH Values | | | | | |
|---|---|---|---|---|---|---|
| Gastrointestinal section | Rat | Guinea Pig | Rabbit | Dog | Monkey | Human |
| Stomach | 3.3–5.5 | 4.1–4.5 | 1.9 | 3.5–5.5 | 2.8–4.8 | 1.5–3.5 |
| Duodenum | 6.5–6.7 | 7.6–7.7 | 6.0–6.8 | 6.2 | 5.8 | 5.0–7.0 |
| Jejunum | 6.7–6.8 | 7.7–8.1 | 6.8–7.5 | 6.2–6.6 | 5.8–6.0 | 6.0–7.0 |
| Ilium | 6.8–7.1 | 8.1 | 7.5–8.0 | 6.6–7.5 | 6.1 | 7.0 |
| Colon | 6.6 | 6.7 | 7.2 | 6.5 | 5.1 | 5.5–7.0 |

*Source*: From Ref. 14, with permission.

formulations, as well as being difficult to obtain multiple blood samples from repetitive sampling. In addition, the gastrointestinal microflora is different from that of humans. This species does have the advantage of having a gastrointestinal pH similar to that in humans and, if urinary data are used for bioavailability testing, it is easy to obtain total urine collections using metabolism cages.

In most respects, the monkey offers the closest physiological and metabolic model for humans and would be the first choice if not for its extremely high cost and handling requirements. Additionally, the monkey is capable of transmitting several human pathogens, and care must be taken when handling analytical samples. Obviously, if primate facilities are available, the use of this species becomes more attractive. Dogs offer a reasonable compromise since (1) they are large enough to test solid dosage forms, (2) their gastrointestinal pH is similar to that in humans, (3) sampling is convenient, (4) they are easy to handle, and (5) special precautions are not needed when handling the biological samples after collection.

Throughout drug development, toxicology/safety studies are conducted in two species. As most drugs are intended for oral administration, the rat and dog are the primary species used for reasons of availability, costs, and long history of use. However, not all drugs are targeted for oral delivery and other routes (intravenous, dermal, etc.) and species (rabbit for reproductive toxicology, monkey) may be used.

The rat is the species most commonly used with oral administration—the route of choice—and is the primary rodent utilized in toxicology studies. Oral administration by gavage as a solution or suspension is the most frequently employed dosage form in rodents and is used in studies of up to 6–9 months duration. The primary consideration is the choice of vehicle. Solutions in water are chosen if the solubility of the drug is sufficient to deliver the highest dose proposed. For drugs of low or limited solubility, a dispersing agent for use in an aqueous milieu is selected. A vehicle of 0.5% aqueous Tween 80 (i.e., polysorbate 80) has a long history of use for a wide range of drugs. Cellulose and its derivatives (e.g., methyl-, carboxy-), starch, and different natural oils (e.g., corn, sesame) are often used as a vehicle for oral gavage formulations. For studies of greater duration than 9 months (i.e., lifetime carcinogenicity studies), oral drugs are usually administered as admixture in the diet of the rats. Dogs are often dosed with capsules.

Literature review of species differences in pharmacokinetics and bioavailabilty of different dosage formulations are presented in Section 2.5 of Chapter, Bioavailability and Bioequivalence Studies.

## 2.5 Protein/Peptide, DNA, Antisense Oligonucleotides (AS-ONs), siRNA, and Other Biological Formulations

***Protein/Peptide Formulations*** Protein pharmaceuticals have high specificity and activity at relatively low concentrations, in comparison with small chemical drugs. These features have made protein pharmaceuticals indispensable in combating human diseases. Currently, the U.S. Food and Drug Administration (FDA) has approved over 30 different recombinant DNA-derived proteins (e.g., erythropoietin, interferon α-2a/b, somatropin and follitropin β), and many more are in an advanced stage of development [15].

The most challenging task in the development of protein pharmaceuticals is to deal with physical and chemical instabilities of proteins. Protein instability is one of the major reasons why protein pharmaceuticals are administered traditionally

through injection rather than taken orally like most small chemical drugs [16]. Problems such as acid catalyzed degradation in the stomach, proteolytic breakdown in the GI tract, poor permeability across the gastrointestinal mucosa, and first-pass metabolism during transfer across the absorption barrier and in the liver must be overcome for the efficient delivery of drugs into the bloodstream [17]. In order to achieve successful oral delivery of protein drugs, they need to be protected from the harsh environment in the stomach. For designing oral dosage forms, the formulator must consider that the natural pH environment of the GI tract varies from acidic (pH ~ 1.2) in the stomach to slightly alkaline in the intestine (pH ~ 7.4) [18]. The review by George and Abraham [19] describes the sources and physical and chemical properties of alginate and chitosan (naturally occurring biopolymers) that enable them to become suitable for protein delivery, the mechanisms of hydrogel formation, modifications that increase their protein encapsulation efficiency, and the recent trends in their application.

There are considerable hurdles to be overcome before practical use can be made of therapeutic peptides and proteins due to chemical and enzymatic instability, poor absorption through biological membranes, rapid plasma clearance, peculiar dose–response curves, and immunogenicity [1]. Many attempts have addressed these problems by chemical modifications or by coadministration of adjuvants to eliminate undesirable properties of peptides and proteins [20]. Cell permeant peptides and their delivery enhancement are reviewed by Ülo Langel [21]. Different types of protein/peptide formulations need to be developed based on clinical needs, patient compliance; delivery methods; drug stability, storage, and distribution; and market competitiveness. Having a clear vision of the type of formulation that is desired will allow one to design a better formulation. Liquid formulations have generally been preferred due to the convenience of manufacturing and use. However, protein drugs may not be stable enough to be handled as a liquid formulation. Dried formulations (e.g., lyophilized formulations) or suspension formulations have been successfully used to overcome stability problems. In addition, specific applications and/or delivery routes may demand an appropriate type of formulation such as a spray-dried formulation for pulmonary delivery.

Maximizing the stability of protein/peptide is one of the most important objectives of formulation optimization. Once the key degradation products, responsible stresses, and analytical methods are identified, the protein or peptide can be stabilized by optimizing various formulation variables. Often, well established scientific principles can be used to solve some stability problems. In other instances, solving stability issues can best be achieved through utilizing empirical approaches. When an optimized formulation does not provide sufficient stability, then alternative approaches like lyophilization are indicated. Table 5 provides some typical solutions for routine stability problems.

***Gene Therapies*** Gene delivery has been regarded as a powerful tool for curing disease by replacing defective genes, substituting missing genes, or silencing unwanted gene expression. Gene therapy holds great promise for treating various forms of diseases with a genetic origin including cystic fibrosis, different forms of cancer, and cardiovascular disorders. Basically, there are two types of gene carriers that deliver foreign DNA into the diseased target cell population. DNA delivery systems are viral and nonviral vectors [22]. Viral vectors are the most effective because of their evolutionary optimization for this purpose. However, recently

**TABLE 5 Problems and Solutions of Protein and Peptide Drug Development**

| Problems | Solutions |
|---|---|
| Deamidation, cyclic imide formation | pH optimization |
| Aggregation, precipitation | pH optimization; addition of sugars, salts, amino acids, and/or surfactants |
| Truncation | pH optimization, protease removal |
| Oxidation | Excipient purity analysis, addition of free-radical scavengers (mannitol, sorbitol), use of a competitive inhibitor (methionine) |
| Surface denaturation, adsorption | Addition of surfactants or excipient proteins |
| Dehydration | Addition of sugars or amino acids |

reported safety issues such as random recombination, oncogenic potential, and immunogenicity [23, 24] have set back the rapid development of viral vectors. As compared to viral vectors, nonviral systems have a low or absent immunogenicity, can relatively easily be scaled-up, and have great flexibility with regard to vector modification and DNA incorporation [25, 26]. Nonviral vectors are safe to use but less efficient. In light of safety concerns, nonviral delivery systems have been developed for gene therapy experiments. Among those, cationic liposomes are widely used for almost all animal cells because they have nonspecific ionic interaction and low toxicity properties [27–29]. However, there are some limitations for cationic liposomes. When they are used for *in vivo* transfection, they are unstable. Therefore, many polymeric cationic systems such as gelatin, polyethylenimine (PEI), poly(L-lysines), tetraaminofullerene, poly(L-histidine)-graft-poly(L-lysines), DEAE-dextrans, cationic dendrimers, and chitosan have been studied for *in vitro* as well as *in vivo* application [30, 31].

The clinical use of gene therapy treatments, however, is restricted, mainly because of the absence of safe and efficient gene delivery technologies [32–34]. In the past decade, nonviral approaches, particularly cationic lipid or polymer-based systems, have received rapidly growing interest, because they offer several advantages over the viral counterparts. Nevertheless, currently investigated gene delivery polymers like polyethylenimine (PEI), poly(L-lysine) (PLL), polyamidoamine dendrimers (PAMAM), and poly(2-(dimethylamino)ethyl methacrylate) (pDMAEMA) have not yet advanced to clinical evaluation, mainly due to their acute cytotoxicities and low *in vivo* transfection activity [24, 35–37].

Recently, biodegradable cationic polymers such as poly(4-hydroxy-L-proline ester) [38, 39], poly[α-(4-aminobutyl)-L-glycolic acid][40], poly(L-lysine) conjugates [41], linear poly(ester amine)s [42–45], poly(ester amine) networks [46], cationic polyphosphazenes [47], and degradable polyethylenimine [48, 49] have been investigated as nonviral carriers for gene delivery. Generally, polyplexes based on these polymers showed lower cytotoxicity and comparable *in vitro* transfection activities as compared to the nondegradable counterparts. Furthermore, the degradation properties of these polymeric vectors can be employed as a valuable tool to regulate the unpacking and release of the DNA inside the cells. Linear poly(ester amine)s as recently reported by Langer and co-workers are of particular interest because the applied synthesis is straightforward and polymers with a great variety of different structures and properties can readily be prepared [42, 43, 45]. Chitosan oligosac-

charides (COSs) chemically modified with deoxycholic acid (DOCA) have low toxicity and possess a good transfection activity [50].

Zhong et al. [51] report a versatile family of degradable hyperbranched poly(ester amine)s that contain primary, secondary, and tertiary amino groups in their structures for nonviral gene delivery. These hyperbranched poly(ester amine)s could easily be synthesized in high yields by a Michael-type conjugate addition of trifunctional amines, containing both a primary and a secondary amino group to diacrylates [52]. As vectors for gene delivery, these hyperbranched polymers may have several advantages: (1) in general they exhibit much better solubility in water than their linear counterparts [53]; (2) primary amines present in the periphery of these hyperbranched polymers can interact more efficiently with the phosphate groups of DNA, resulting in the formation of nanosized polymer/DNA complexes, analogous to the behavior of hyperbranched PEI that was shown to condense DNA into a smaller size particle than linear PEI [54]; (3) the high density of secondary and tertiary amines will give rise to high buffering capacity over a wide pH range, facilitating endosomal escape of polymer/DNA complexes ("proton sponge hypothesis") [55]; and (4) due to their high functionality, the structure of hyperbranched polymers can easily be modified to further improve transfection efficacy and/or achieve targeted gene delivery *in vivo*. For example, hyperbranched PEI, consisting of 25%, 50%, and 25% of primary, secondary, and tertiary amines, is one of the most efficient polymeric gene carriers studied *in vitro* and *in vivo* [55–58].

A significant limitation in using nonviral gene therapy approaches for the treatment of disease has been poor delivery systems leading to poor cell transfection and protein expression [59]. Interest remains high in developing a robust nonviral system due to toxicity concerns associated with the alternative viral based approaches. The use of mechanical/physical delivery systems in combination with plasmid based vectors has produced encouraging results with regard to the high protein expression levels achieved [60–62]. These methodologies such as electroporation or hydrodynamic pressure are fairly complex in design and require special equipment and/or training; however, they are beginning to be clinically evaluated for various diseases and may ultimately prove to have therapeutic applicability [63]. Alternative nonviral approaches are directed at combining plasmid DNA with lipid or polymer based molecules that act to facilitate DNA uptake and expression by promoting interaction with cellular membranes, protecting DNA against endogenous nucleases and trafficking plasmid to the nucleus [64–66].

Polymeric gene carrier systems have been widely used for gene therapy applications and are emerging as a viable alternative to the current systems [35]. These systems are attractive because of ease of use and low cost. Furthermore, excellent molecular flexibility allows for sophisticated modifications and incorporation of novel chemistries with little difficulty [67, 68]. Use of polymers for gene therapy applications also offers the possibility of repeated injections in order to maximize and/or optimize the therapeutic applicability.

Polyethylenimine (PEI) as a gene delivery polymer has a distinct advantage over other nonviral gene delivery systems due to its inherent endosomolytic activity [69]. However, despite its use in animal studies, the clinical development of polyethylenimines has been sluggish due to molecular heterogeneity and acute toxicity [70, 71]. The transfection activity and cytotoxicity of PEI is directly related to its molecular size. Specifically, the larger the molecular size the higher the transfection

activity and cytotoxicity. Different strategies have been described to improve polymer activity without compromising safety including incorporating degradable linkages into high molecular weight polymers to improve clearance [49] or incorporating biocompatible membrane-permeating moieties into low molecular weight polymers to improve activity [36, 72].

Fewell et al. [59] describe the chemical synthesis and the physical and chemical properties of a novel condensing lipopolymer based on a low molecular weight polyethylenimine covalently linked to cholesterol (CHOL) and methoxypolyethylene glycol (PEG) molecules forming PEG–PEI–CHOL (PPC). The polymer offers a high degree of protection against DNases and high *in vitro* and *in vivo* transfection activity. Lv et al. [73] discuss the toxicity of main cationic gene vectors (cationic lipids and cationic polymers); emphasis is placed on the relationship between toxicity and structure of these compounds. The authors evaluate the structural features of cationic compounds and discuss which groups may increase the toxicity, what kind of linkages have relatively short half-life, and how proper modifications will decrease the toxicity of cationic lipids and cationic polymers for gene delivery.

*Antisense oligonucleotides (AS-ONs)* are specific inhibitors of gene expression. As single strands of DNA, these molecules hybridize to complementary RNA and induce RNase H-mediated degradation of the target RNA [74, 75]. Initially introduced by Zamecnik and Stephenson [76] and Stephenson and Zamecnik [77], AS-ONs represent a promising tool in fighting viral, malignant, and inflammatory diseases. Since unmodified phosphodiester oligonucleotides are susceptible to degradation by exo- and endonucleases, the more stable phosphorothioate oligonucleotides (S-ONs) have become the most extensively used antisense agents. In these molecules, one of the nonbridging oxygens of the phosphate backbone is substituted by a sulfur atom [78]. Several clinical trials with S-ONs are in progress for treatment of chronic lymphocytic leukemia, multiple myeloma, malignant melanoma, pancreatic carcinoma, and non-small-cell lung carcinoma [79, 80]. However, a major problem for the development of antisense therapy is the ineffective cellular uptake of antisense molecules. Oligonucleotides are relatively large hydrophilic molecules with a molecular weight ranging from 5000 to 10,000 Da and do not passively diffuse across cell membranes. Conflicting theories for oligonucleotide transport into the cell have been reported: (1) receptor-dependent [81], (2) independent [82], (3) via pinocytosis [83], or (4) via unknown mechanism (e.g., caveolar or potocytotic) [84]. However, in order to be active, AS-ONs must preferentially be present either in the nucleus or in the cytoplasm. The intracellular localization of ONs is confined to the endocytotic compartments. Only a small portion of ONs are somehow released into the cytoplasm [85].

Improvement of cellular uptake is a major topic of ongoing antisense research. Numerous potential delivery systems such as viral capsid structures, nanoparticles, and liposomes have been explored for efficient antisense delivery [86–88]. Currently, most research focuses on liposomes. Anionic liposomes can be loaded with ONs and then be linked to specific antibodies for tissue targeting [89]. Unfortunately, for these liposomes encapsulation of ONs is quite inefficient. In contrast, incubation of ONs with cationic lipids results in the high efficient formation of liposome–ON complexes by electrostatic interaction. The cellular uptake of ONs with cationic liposomes is strongly enhanced by a factor of 15–25 over ON only, resulting in a 1000-fold overall increase in ON activity [90–92]. This tremendous increase in

activity may be explained by the specific nuclear localization of ONs. In contrast to the cytoplasmic localization of noncomplexed ONs, ONs delivered by cationic liposomes accumulate predominantly in the cell nucleus [93].

Antisense oligonucleotides are being widely investigated for the downregulation of genes [94, 95]. To inhibit protein production, the ONs have to reach the cytoplasm or nucleus of the cells, where they act by specific binding to the target mRNA or DNA. In spite of this simple action mechanism, different barriers still limit the antisense activity. Actually, before the ONs can reach their target site, they first have to cross the cellular membrane, escape from the endosomal compartment, leave their pharmaceutical carriers (i.e., the delivery system), and hybridize to the target sequence [96].

Rapid degradation of phosphodiester ONs is known to be one of the factors limiting their success as therapeutics. One approach to improve antisense activity is to develop ONs with increased nuclease stability so that intracellular degradation should not be an issue [97]. However, modification of the ON backbone often induces undesirable features. The first generation of chemically modified ONs, namely, phosphothioate ONs, are known to frequently exhibit nonspecific, nonantisense effects, generally attributed to their increased nonspecific protein binding when compared to their phosphodiester analogues [98]. The second and third generation antisense compounds all work to some extent but have one or more problems such as solubility, delivery issues, lack of RNase H activation, or simply the cost of synthesis [99].

Another approach to improve the antisense activity of ONs is the development of suitable delivery systems [96, 100]. The delivery system, often composed of cationic lipids or cationic polymers, is designed to help the ONs enter the cell and escape from the endosomes. Clearly, in cases where (degradable) phosphodiester ONs (PO-ONs) are used, the protection of the ONs against nucleases, given by their carrier, becomes an important aspect. Ideally, the delivery system should protect the complexed ONs during the different steps of the trafficking pathway. In general, cationic carriers are believed to protect the complexed ONs against degradation. However, some carriers can establish an antisense effect with degradable phosphodiester ONs, while other carriers fail [101, 102]. The lack of antisense activity has been attributed to insufficient ON protection against nucleases with the carriers used, although no direct proof of this hypothesis was given.

*RNAi* is a recently developed technique to silence proteins in a sequence-specific manner by inhibiting mRNA and consequently reducing protein expression [103]. The functional mediator of RNA interference is a short dsRNA oligonucleotide called small interfering RNA (siRNA). A growing number of investigations are being reported on efforts to use siRNA as a candidate therapeutic agent [104]. To achieve therapeutic success, however, several hurdles must be overcome including rapid clearance, nuclease degradation, and inefficient intracellular localization. Most efforts reported to date are based on studies in rodents and rodent models of disease. With the exception of local administration into eye tissues, the results show low activity due to a strong dependence on methods for intracellular delivery, as found for antisense and gene therapy approaches. One surprising observation is that siRNA composed of unmodified RNA is rapidly cleared by the kidneys [105], indicating low binding to proteins or cells in the blood and a surprising lack of nuclease degradation as the major barrier. A delivery strategy for use of siRNA as a thera-

peutic modality should therefore firstly reduce glomerular filtration, and secondly optimize intracellular delivery to target cells, while also minimizing exposure to nuclease and distribution to nontarget tissues. Nanoparticle delivery systems developed with a ligand–polymer–polymer conjugate appear attractive to meet these needs and indeed neovasculature targeted forms, using a "cyclic" RGD peptide as the ligand, have produced successful results [106, 107].

*Peptide nucleic acids (PNAs)* are rapidly emerging as powerful tools in DNA hybridization and antisense techniques. PNAs are synthetic oligonucleotide analogues in which the natural nucleic acid backbone is replaced by an uncharged pseudopeptide backbone consisting of *N*-(2-aminoethyl)-glycine units linked to the natural purine and pyrimidine bases [108]. They form very stable complexes with complementary DNA or RNA strands with very high affinity, partially as a result of the neutrality and flexibility of the artificial backbone.

PNAs possess a strong resistance to enzymatic degradation in cells, fluids, and tissues, showing an exceptional stability in biological environments [109]. In addition, the unique chemical, physical, and biological properties of PNAs have been exploited to produce powerful biomolecular tools, antisense and antigene agents, molecular probes, and biosensors. Important new applications that could not be performed using natural oligonucleotides have emerged. Duplexes between PNA and DNA or RNA are in general thermally more stable than the corresponding DNA–DNA or DNA–RNA duplexes [109, 110]. The sequence dependence of the stability, however, is more complex than that found for DNA–DNA complexes because of the inherent asymmetry of the duplex. In fact, PNA–DNA duplexes show significantly increased stability when the purines are in the PNA strand. Thus, on top of the dependence on G–C content, the stability of PNA–DNA duplexes also depends on the purine fraction of the PNA strand [111]. In general, it is also observed that the thermal stability of PNA–PNA duplexes exceeds that of PNA–RNA duplexes, which again are more stable than PNA–DNA duplexes.

Very importantly, but not surprisingly, the stability of PNA–DNA duplexes is almost unaffected by the ionic strength of the medium (actually the stability decreases slightly with increasing $Na^+$ concentration due to counterion release upon duplex formation). This is in sharp contrast, of course, to the behavior of DNA–DNA (or RNA) duplexes, whose stability decreases dramatically at low ionic strength because of the required counterion shielding of the phosphate backbone [112].

PNA hybridization kinetics have also been studied by surface plasmon resonance (SPR) technology and the results, although they cannot be considered conclusive, indicate no major difference in PNA and DNA hybridization duplex formation [110]. Although it is not the intention to give an elaborate account of the ongoing development of gene therapeutic drugs based on PNA oligomers, it is fair to say that PNA does possess many of the properties desired for an antisense reagent. It binds strongly and with excellent sequence specificity to complementary mRNA, it has very high biological stability, and targeting of specific mRNAs with PNA has been shown to strongly inhibit its translation. This has been demonstrated in several biological systems *in vitro*, and the introduction of various novel methods for improved delivery of PNA to eukaryotic cells has recently also allowed good efficacy in cell culture *ex vivo* and, sometimes, *in vivo* [113–116]. In the last two years, PNA antisense and antigene technology has entered the realm of biological and preclinical studies. This is primarily due to the development of a number of methods able to improve PNA permeability across cell membranes [117–119] and to allow

efficient delivery of PNA oligomers to eukaryotic cells. Two novel delivery systems based on polymeric core–shell microspheres and autologous loaded opsonized RBC systems were able to specifically inhibit the expression of COX-2 or iNOS proteins in murine macrophages at nanomolar concentration, that is, with 3 log lower concentration with respect to those required by PNA linked to hydrophobic peptide [120]. Authors concluded that core–shell polymeric microspheres and loaded RBC represent promising delivery strategies for *in vivo* targeting of antisense PNAs to macrophages during inflammatory processes.

### 2.6 Enhanced Delivery and Semisolid or Sustained Release Preparations (Micronization and Nanoparticles)

In order to improve the specific delivery and bioavailability of drugs with low therapeutic index, several drug carriers such as liposomes, microparticles, nanoassociates, nanoparticles, drug polymer conjugates, and polymeric micelles have been developed. In recent years, polymeric micelles have been the object of growing scientific attention [121].

***Micelles*** One of the widely used techniques for augmenting the water solubility of drug candidates is through modification of the solubilization medium by using cosolvents, complexing agents (i.e., cyclodextrins), or surfactants [122, 123]. In an aqueous medium, above a given concentration termed the critical micellar concentration (CMC), surfactants self-assemble into a colloidal dispersion of molecular aggregates called micelles (<50–100 nm) [124] and have been shown to solubilize hydrophobic drugs inside the core formed by the lipophilic part of the surfactant. The hydrophilic moieties form the shell or corona of these nanocarriers [124–128]. In addition to the increase in drug solubility, micelles can protect the incorporated drugs by insulating them from the aqueous environment.

Besides classic surfactants, micelles can also be formed from amphiphilic polymers. These polymers often provide better kinetic and thermodynamic stability than conventional surfactants [124–128]. However, they exhibit a number of disadvantages including a complex manufacturing process, which requires organic solvents followed by a dialysis step or an evaporation process depending on polymer polarity [125–128]. These limitations were addressed by the introduction of self-assembling polymeric micelles such as mmePEG$_{750}$P(CL-*co*-TMC) (monomethyletherpoly(oxyethylene glycol$_{750}$)-poly(caprolactone-*co*-trimethylene carbonate)). This amphiphilic copolymer aggregates spontaneously into 20 nm micelles upon contact with water [129] and is intended for oral delivery of poorly soluble drugs.

Micelles as drug carriers are able to provide a set of unbeatable advantages—they can solubilize poorly soluble drugs and thus increase their bioavailability, they can stay in the body (in the blood) long enough to provide gradual accumulation in the required area, their size permits them to accumulate in body regions with leaky vasculature, they can be targeted by attachment of a specific ligand to the outer surface, and they can be prepared in large quantities easily and reproducibly [125]. Being in a micellar form, the drug (poorly soluble drug, first of all) is well protected from possible inactivation under the effect of biological surroundings, it does not provoke undesirable side effects, and its bioavailability is usually increased.

The micelle is structured in such a way that the outer surface of the micelle exposed to the aqueous surrounding consists of components that are hardly reactive

toward blood or tissue components. This structural peculiarity allows micelles to stay in the blood (tissues) rather long without being recognized by certain proteins and/or phagocytic cells. This longevity is an extremely important feature of micelles as drug carriers. Long-circulating pharmaceuticals and pharmaceutical carriers currently represent a fast growing area of biomedical research; for example, see Refs. 130–134.

***Micronization*** Developing intravenous formulations of hydrophobic drugs is particularly challenging. Possessing intravenous formulations of drugs can open new markets for drugs like antibiotics that could often be initially prescribed in a hospital setting. In addition, intravenous formulations can expand the market for the oral dosage formulation because physicians typically prefer to discharge patients on the oral formulation of the same drug administered intravenously in the hospital.

Many hydrophobic drugs are comprised of particles that are relatively large and therefore have a limited surface area available for interaction with water. These hydrophobic drugs are often formulated in less than ideal ways in order to make them dissolve. It is possible to increase the dissolution rate of hydrophobic drugs by increasing their aggregate surface area. To accomplish this, many pharmaceutical companies use a process, called micronization, which entails grinding hydrophobic drugs into smaller microparticles. However, the drug particles produced by micronization are often still not small enough to adequately improve dissolution, or to be administered intravenously. New technologies have been developed that convert drugs that do not dissolve well in water—hydrophobic drugs—into microparticles or nanoparticles of the drugs embedded in small microparticles. Alternatively, oils like Cremophor are used to dissolve the drugs. However, these oils are often not well tolerated and can require prolonged infusion rather than rapid injections. In addition, some hydrophobic drugs can be formulated into soft gelatin capsules, but these are only suitable for oral administration and encapsulate only a small volume of drug, requiring the administration of many capsules. Sometimes development of these drugs must be terminated because no suitable formulation can be found.

A continuous basal insulin supply may help to improve the condition of diabetics. A body weight gain in diabetic rats was observed as fast as in normal rats after a single injection of insulin-loaded microspheres in 28 days [135]. *In vitro* release of insulin from poly(lactide-*co*-glycolide) (PLGA) microspheres was improved by physically mixing microspheres prepared by water-in-oil-in-water double emulsion and solid-in-oil-in-water emulsion methods [136]. In this study, insulin-loaded microspheres maintained basal insulin level in diabetic rabbits for 40 days. Histological evaluations suggested that the insulin-loaded microspheres were biocompatible.

Macromolecular drugs such as peptides and proteins are unable to overcome the mucosal barriers and/or are degraded before reaching the bloodstream. Among the approaches explored so far in order to optimize the transport of these macromolecules across mucosal barriers, the use of nanoparticulate carriers represents a challenging but promising strategy. Nanostructures based on the mucoadhesive polysaccharide chitosan-coated systems showed an important capacity to enhance the intestinal absorption of the model peptide, salmon calcitonin, as shown by the important and long-lasting decrease in the calcemia levels observed in rats [137].

Alonso [138] reviews recent advances in the design of polymeric nanosystems intended to be used as carriers for nasal vaccine delivery. The information accumu-

lated regarding the *in vivo* behavior of these nanocarriers indicates that they are able to facilitate the transport of the associated antigen across the nasal epithelium, thus leading to efficient antigen presentation to the immune system. Furthermore, the results suggest that not only the size and surface properties but also the polymer composition and the structural architecture of the nanosystems are critical for the optimization of these antigen carriers.

***Dendrimers*** Dendrimers are one of the emerging delivery systems with the capability to present such hydrophobic agents in a formulation with better prospective [139]. Dendrimers are synthetic, highly branched, monodisperse macromolecules of nanometer dimensions. These dendritic macromolecules with a large number of surface terminal groups and interior cavities offer a better opportunity for delivery by becoming charged and acting as static covalent micelles [140]. Started in the mid-1980s, the research investigations into the synthetic methodology and physical and chemical properties of these macromolecules are increasing exponentially with growing interest in this field. Although dendrimer drug delivery is in its infancy and toxicity is still a problem, it offers several attractive features. It provides a uniform platform for drug attachment that has the ability to bind and release drugs through several mechanisms. Properties associated with these dendrimers such as uniform size, water solubility, modifiable surface functionality, and available internal cavities make them attractive for biological and drug delivery applications. These are biocompatible, nonimmunogenic, and water soluble and possess terminal functional groups for binding various targeted or guest molecules [141]. The host–guest properties of dendrimers based on hydrophobic and ionic interactions apart from physical entrapment have been thoroughly studied [142]. The involvement of dendrimer in enhancing solubility of acidic anti-inflammatory drug through ionic interaction has been explored [143]. Dendrimers have also been explored for intracellular delivery of the anti-inflammatory drug ibuprofen [144]. As the dendrimer has many end functional groups, these would determine their solubility and physical and chemical interaction in the immediate surrounding environment [145].

***Liposomes*** Liposomes are widely used as nanosized drug delivery vehicles for active and passive targeting [146]. Liposomes may also "plug" and "seal" the damaged myocyte membranes and protect cells against ischemic and reperfusion injury *in vitro* [147] and the delivery of the ATP into the ischemic myocytes at the same time [148, 149]. As a drug delivery system, liposomes can significantly alter the pharmacokinetics and pharmacodynamics of entrapped drugs, for example, by enhancing drug uptake, delaying rapid drug clearance, and reducing drug toxicity [150–152].

For example, pegylated liposomal formulation of 2′-deoxyinosine (a 5-fluorouracil modulator), through an optimized pharmacokinetic profile (near seven times clearance reduction as compared with the free form), displays a potent tumor reduction effect in xenografted mice [153].

## 2.7 Vehicles and Selection of Excipients

FDA guidance [154] means by *new excipients* any inactive ingredients that are intentionally added to therapeutic and diagnostic products, but that (1) we believe

are not intended to exert therapeutic effects at the intended dosage, although they may act to improve product delivery (e.g., enhance absorption or control release of the drug substance); and (2) are not fully qualified by existing safety data with respect to the currently proposed level of exposure, duration of exposure, or route of administration. Examples of excipients include fillers, extenders, diluents, wetting agents, solvents, emulsifiers, preservatives, flavors, absorption enhancers, sustained release matrices, and coloring agents.

In addition to their intended use, excipient candidates should also be qualified as appropriate pharmaceutical ingredients. Generally, it is preferable to select excipients that have been used in marketed products with a relevant route of delivery. Furthermore, excipients that have been used with a similar dosing frequency, history of chronic use, and patient population are also preferable. Otherwise, the approval and safety of the excipients need to be carefully examined. If the excipient is known to be safe on the basis of solid science or has a proven clinical safety record, it can be considered equivalent to approved excipients. The list of excipients used for parenteral pharmaceuticals is available in the literature [155–157]. The current information from the Inactive Ingredient Database can be found at the FDA website at http://www.accessdata.fda.gov/scripts/cder/iig/index.cfm.

When it is necessary to introduce other excipients with minimal safety records, a significant risk associated with the excipients will be added to the product development process and additional preclinical and clinical studies may be needed. Not all excipients are inert substances; some have been shown to be potential toxicants. The Federal Food, Drug, and Cosmetic Act of 1938 (the Act) was enacted after the tragedy of the elixir of sulfanilamide in 1937 in which an untested excipient was responsible for the death of many children who consumed the pharmaceutical. The Act required manufacturers to perform safety testing of pharmaceuticals and submit new drug applications (NDAs) demonstrating safety before marketing. Since that time, the FDA has become aware that certain other excipients used in commerce can cause serious toxicities in consumers of prescription and over-the-counter (OTC) drug products in the United States and other countries.

The FDA's *Guidance for Industry: Nonclinical Studies for the Safety Evaluation of Pharmaceutical Excipients* [154] describes the types of toxicity data that the Agency uses in determining whether a potential new excipient is safe for use in human pharmaceuticals. It discusses recommended safety evaluations for excipients proposed for use in OTC and generic drug products, and describes testing strategies for pharmaceuticals proposed for short-term, intermediate, and long-term use. It also describes recommended excipient toxicity testing for pulmonary, injectable, and topical pharmaceuticals.

Another important requirement in qualifying an excipient is the purity of the raw material. Depending on its historical use as a pharmaceutical ingredient, excipients are available in several different pharmaceutical grades, for example, USP (U.S. Pharmacopoeia), Ph. Eur. (European Pharmacopoeia), and JP (Japan Pharmacopoeia). These pharmaceutical grade materials should be considered as a primary resource, but the quality provided may not be good enough for specific product development. For example, significant stability problems can be found with some impurities even at concentrations below their specification (e.g., metal ions, peroxides, protease, and reducing sugars). These problems are more prominent in low protein concentration products due to a high impurity/protein ratio, although problems like visible precipitation may be independent of protein concentration. If an

adjustment to the existing specification is necessary, it is critical to look into the availability of GMP quality raw materials with modified specifications as early as possible once the potential problem is identified. The use of excipients derived from animals or humans should be avoided if possible due to the risks associated with transmissible diseases like bovine spongiform encephalopathy, Creutzfeldt–Jakob disease, hepatitis virus, and HIV. Numerous regulatory guidelines have been issued to discourage the use of animal/human derived excipients (Note for Guidance on Minimizing the Risk of Transmitting Animal Spongiform Encephalopathy Agents Via Medicinal Products—EMEA (4/99); Note for Guidance on Plasma-Derived Medicinal Products—EMEA (7/98); Development Pharmaceutics for Biotechnological and Biological Products. Annex to Note for Guidance on Development). When animal derived excipients have to be included in a product, the manufacturer will need to fulfill its responsibilities to justify their selection and implement adequate safety measures.

A review of commercially available oral and injectable solution formulations presented by Robert Strickley [158] reveals that the solubilizing excipients include *water-soluble organic solvents* (polyethylene glycol 300, polyethylene glycol 400, ethanol, propylene glycol, glycerin, *N*-methyl-2-pyrrolidone, dimethylacetamide, and dimethylsulfoxide), *nonionic surfactants* (Cremophor EL, Cremophor RH 40, Cremophor RH 60, *d*-α-tocopherol polyethylene glycol 1000 succinate, polysorbate 20, polysorbate 80, Solutol HS 15, sorbitan monooleate, poloxamer 407, Labrafil M-1944CS, Labrafil M-2125CS, Labrasol, Gellucire 44/14, Softigen 767, and mono- and di-fatty acid esters of PEG 300, 400, or 1750), *water-insoluble lipids* (castor oil, corn oil, cottonseed oil, olive oil, peanut oil, peppermint oil, safflower oil, sesame oil, soybean oil, hydrogenated vegetable oils, hydrogenated soybean oil, and medium-chain triglycerides of coconut oil and palm seed oil), *organic liquids/semisolids* (beeswax, *d*-α-tocopherol, oleic acid, medium-chain mono- and diglycerides), various *cyclodextrins* (α-cyclodextrin, β-cyclodextrin, hydroxypropyl-β-cyclodextrin, and sulfobutylether-β-cyclodextrin), and *phospholipids* (hydrogenated soy phosphatidylcholine, distearoylphosphatidylglycerol, l-α-dimyristoylphosphatidylcholine, l-α-dimyristoylphosphatidylglycerol). The chemical techniques to solubilize water-insoluble drugs for oral and injection administration include pH adjustment, cosolvents, complexation, microemulsions, self-emulsifying drug delivery systems, micelles, liposomes, and emulsions.

An excellent reference for excipient information is the APA's *Handbook of Pharmaceutical Excipients* [159]. A very extensive and detailed list of vehicles and excipients commonly used in preclinical animal studies is presented by S. Gad in his book *Drug Safety Evaluation* [4, pp. 494–500]. In addition to this, we present below in more detail three other commonly or more recently used vehicles—Cremophor, chitosan, and cyclodextrin.

***Cremophor EL (Polyoxyethyleneglycerol Triricinoleate)*** The heterogeneous nonionic surfactant CrEL is a white to off-white viscous liquid with an approximate molecular weight of 3 kDa and a specific gravity (25 °C/25 °C) of 1.05–1.06, and it is produced by the reaction of castor oil with ethylene oxide at a molar ratio of 1:35 [160]. Castor oil is a colorless or pale yellow fixed oil obtained from the seeds of *Ricinus communis*, with an extremely high viscosity, and consists mainly of the glycerides of ricinoleic, isoricinoleic, stearic, and dihydroxystearic acids. CrEL is usually of highly variable composition, with the major component identified as

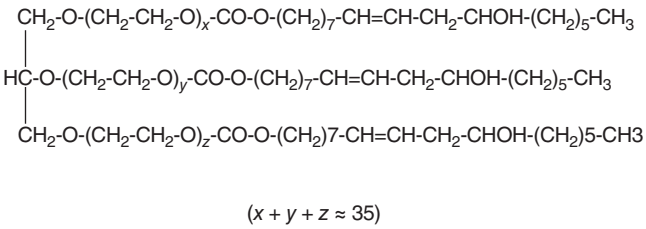


**FIGURE 1** Chemical structures of the main component of CrEL (polyoxyethyleneglycerol triricinoleate 35).

oxylated triglycerides of ricinoleic acid (i.e., polyoxyethyleneglycerol triricinoleate 35) (Fig. 1). Polyvinyl chloride (PVC)-free equipment for CrEL administration is obligatory, since CrEL is known to leach plasticizers from PVC infusion bags and polyethylene-lined tubing sets, which can cause severe hepatic toxicity [161]. CrEL is being used as a vehicle for the solubilization of a wide variety of hydrophobic drugs, including anesthetics, photosensitizers, sedatives, immunosuppressive agents, and (experimental) anticancer drugs. The amount of CrEL administered with these drugs averages 5 mL (range 1.5–10.3 mL), although paclitaxel is an exception as the amount of CrEL is much higher per administration, approximately 26 mL.

Recent investigations have revealed that CrEL, a widely used formulation vehicle, is a biologically and pharmacologically active ingredient of various commercially available drugs. For example, when used in paclitaxel administrations, an exceptionally large amount of CrEL is inevitably coadministered with the IV infusions, causing important biological events that can lead to serious acute hypersensitivity reactions and neurological toxicity, depending on the dose and duration of infusion [5, 162–164 for a review]. In addition, the substantial effects of CrEL on the disposition of other coadministered drugs (e.g., anthracyclines) with a narrow therapeutic window in polychemotherapeutic regimens also can be potentially hazardous to the patient [165]. In view of the inherent problems associated with the use of CrEL, it can be anticipated that there may be a therapeutic advantage from using paclitaxel formulations in which CrEL is absent [5]. Currently, a large variety of new (CrEL-free) formulation vehicles for paclitaxel are in (pre)clinical development, including cosolvent systems (Tween 80/ethanol/Pluronic L64), water-soluble polymers (e.g., polyethylene glycols), emulsions (e.g., triacetin), liposomes, albumin, cyclodextrins, nanocapsules, and microspheres (reviewed in Ref. 166).

In contrast, the *in vitro* and *in vivo* observations of myeloprotective effects related to the use of CrEL in combination with some agents, such as cisplatin, might be used to reformulate such agents with CrEL in order to achieve an optimization of their therapeutic window [167–169]. Iwase et al. [170, 171] have shown that Cremophor EL augments the cytotoxicity of hydrogen peroxide in rat thymocytes.

Based on our laboratory experience (data not published), we have seen some hypersensitivity reactions to CrEL in dogs and would recommend replacement of this vehicle with less toxic vehicles in preclinical dosage formulation, especially in dog studies.

***Chitosan*** Chitin is a copolymer of *N*-acetyl-glucosamine and *N*-glucosamine units randomly or block distributed throughout the biopolymer chain depending on the

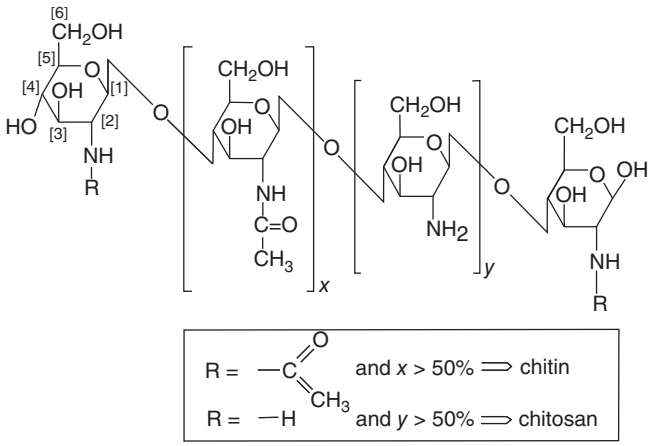


**FIGURE 2** Schematic representation of the chitin and chitosan depicting the copolymer character of the biopolymers.

processing method used to derive the biopolymer (Fig. 2). When the number of *N*-acetyl-glucosamine units is higher than 50%, the biopolymer is termed chitin. Conversely, when the number of *N*-glucosamine units is higher, the term chitosan is used. Chitosan has been the better researched version of the biopolymer because of its ready solubility in dilute acids, rendering chitosan more accessible for utilization and chemical reactions.

Today, the production, chemistry, and applications of chitin and chitosan are well known. Commercially, chitin and chitosan are obtained from shellfish sources such as crabs and shrimps. It is likely that future sources of chitin and chitosan will come from biotechnology innovations, especially when medical applications are the focus.

Chitosan, a cationic polysaccharide obtained by alkaline *N*-deacetylation of chitin, is one of the most widely utilized polysaccharides. The sugar backbone of chitosan consists of β-1,4-linked *d*-glucosamine. In its structure, chitosan is very similar to cellulose, except for the amino group replacing the hydroxyl group on the C-2 position [172]. Recently, chitosan has been widely employed in pharmaceutical and biomedical fields owing to its unique properties such as nontoxicity, biocompatibility, and biodegradability [173, 174]. These characteristics make chitosan an excellent candidate for various biomedical applications such as drug delivery, tissue engineering, and gene delivery [173–178].

Chitosan has been investigated as an excipient in the pharmaceutical industry, to be used in direct tablet compression, as a tablet disintegrant, for the production of controlled release solid dosage forms, or for the improvement of drug dissolution [174]. Compared to traditional excipients, chitosan has been shown to have superior characteristics and especially flexibility in its use. Furthermore, chitosan has been used for production of controlled release implant systems for delivery of hormones over extended periods of time. Lately, the transmucosal absorption promoting characteristics of chitosan have been exploited, especially for nasal and oral delivery of polar drugs to include peptides and proteins and for vaccine delivery. These

properties, together with the very safe toxicity profile, make chitosan an exciting and promising excipient for the pharmaceutical industry for present and future applications.

The transfection efficiency of low molecular weight chitosan (LMWC, molecular weight of 22 kDa) was significantly higher than naked DNA and higher than poly-L-lysine (PLL); it was also less cytotoxic than PLL [176]. For the further development of efficient gene carriers, highly purified chitosan oligosaccharides (COSs) were chemically modified with deoxycholic acid (DOCA). Owing to the amphiphilic characters, the DOCA-conjugated COSs (COSDs) formed self-aggregated nanoparticles in aqueous milieu. COSDs showed great potential as gene carriers with a high level of gene transfection efficiencies, even in the presence of serum. Considered with the negligible cytotoxic effects, DOCA-modified chitosan oligosaccharides can be considered as potential candidates for efficient nonviral gene carriers [50]. A vehicle containing 1.5% chitosan at pH 6.7 increased the bioavailability of intraduodenally administered buserelin in rats from 0.1% to 5.1% [179].

***Cyclodextrins*** Cyclodextrins (CDs) are cyclic oligosaccharides consisting of six or more glucose units linked by α-(1,4)-glycosidic bonds. α-, β-, and γ-Cyclodextrins (α-CD, β-CD, and γ-CD) are composed of six, seven, and eight unsubstituted glucose units, respectively (Fig. 3). In nature, cyclodextrins are produced as a storage form of carbohydrate (reserve substrate) by different microorganisms [180]. On an industrial scale, cyclodextrins are obtained by enzymatic hydrolysis of starch with cyclodextrin-glucosyltransferases. These amylolytic enzymes are produced by different strains of bacilli (e.g., *Bacillus macerans, Bacillus circulans*, and *Bacillus firmus*) and other organisms of other species (e.g., *Klebsiella oxytoca, Micrococcus* sp., and *Thermoanaerobacter* sp.). Cyclodextrin-glucosyltransferase (CGTase) cleaves the helical amylose molecule at regular intervals of six, seven, or eight glucose units, forming at the same time a ring by an intramolecular glucosyltransferase reaction [181].

Cyclodextrins have attracted the interest of the pharmaceutical, cosmetic, and—more recently—the food industry because they can entrap lipophilic compounds in the cavity of their torus-like structure [182–184]. Because of their larger ring, γ-CD and β-CD can host a wider range of molecules than α-CD. The usefulness of α-CD as a carrier and protectant of lipophilic molecules is therefore limited. However,

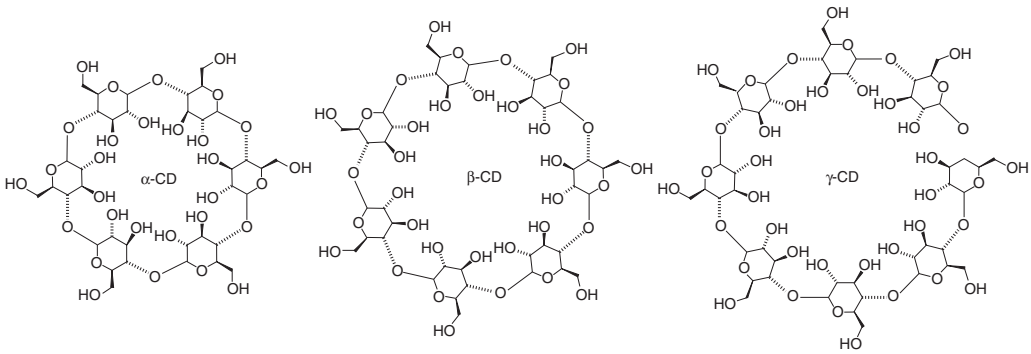


**FIGURE 3** Chemical structure of the three main types of cyclodextrins.

α-CD merits interest as a potential food ingredient for a completely different reason. It has been found that α-CD is not digested by pancreatic amylase and therefore has the physiological properties of a water-soluble, dietary fiber. Consisting of α-(1,4)-linked glucose molecules, it structurally and metabolically resembles retrograded or crystalline nongranular starch ("resistant starch" of the RS3 type according to Englyst's classification) [185]. However, unlike resistant starch, α-CD is freely soluble in water (about 13 g/100 mL at 25 °C), yielding clear solutions with a low viscosity. Because of these properties, α-CD is very well suited for use as a dietary fiber not only in solid but also in semiliquid foods (e.g., yogurt) and all types of beverages.

The structural features of CDs are evaluated for their effect on complexation performance in pharmaceuticals [186]. Optimal specifications, quality production, and safety of each CD are presented. The current and future regulatory process facing excipients is summarized, and the current regulatory status of CDs in Japan, the United States, and Europe is presented.

Cyclodextrin complexation seems to be an attractive delivery system for proteins and peptides helping to overcome chemical and enzymatic instability, poor absorption through biological membranes, and rapid plasma clearance of these macromolecules [187]. The results from the study by Oda et al. [188] suggest that β-CD is considered as one of the most suitable solubilizing agents for evaluating poorly water-soluble drugs using *in situ* loop and perfusion techniques. The capacity to dissolve phenytoin was great in β-CD and hydroxypropyl β-cyclodextrin, followed by Tween 80. Those of methanol, dimethyl sulfoxide, dimethyl acetoamide, and polyethylene glycol 400 were much lower than expected.

However, CDs have been reported to interact with cell membrane constituents such as cholesterol, phospholipids, and phosphatidylinositols, resulting in the induction of hemolysis of RBCs [189–191]. Additionally, randomly methylated β-cyclodextrin (M-β-CD) disrupted the structures of lipid rafts and caveolae [192, 193], which are lipid microdomains formed by lateral assemblies of cholesterol and sphingolipids in the cell membrane, through extraction of cholesterol from the microdomains [194]. The magnitude of the hemolytic activity of the parent CDs is reported to increase in the order of γ-CD < α-CD < β-CD [189, 195]. Recently, various hydrophilic CD derivatives have been developed to improve aqueous solubility and complexation ability of parent CDs. Of the various hydrophilic CD derivatives, hydroxypropylated CDs, sulfobutyl ether CDs, and branched CDs are able to lower the hemolytic activity of each parent CD [196]. Conversely, methylation of the hydroxyl group at the 2 and 6 positions of a glucopyranose unit of β-CD is known to severely augment the hemolytic activity in high concentrations [197]. Arima et al. [198] and Motoyama et al. [199] reported that 2,6-di-*O*-methyl-α-CyD (DM-α-CyD) is predominantly likely to release CD14, a glycosylphosphatidylinositol (GPI)-anchored protein, from lipid rafts through the extractions of phospholipids from murine macrophages without cytotoxicity compared to α-CyD and 2-hydroxypropyl-α-cyclodextrin (HP-α-CyD). Recently, Motoyama et al. [200] showed that DM-α-CyD has higher hemolytic and morphological change activity than α-CyD and HP-α-CyD through more extraction of phospholipids including sphingomyelin and proteins, not cholesterol, from RBC membranes than α-CyD and HP-α-CyD. These toxicity reports for different types of cyclodextrins should be taken into account for their selection in dosage formulations.

## 3 TYPES AND CHARACTERISTICS OF DOSAGE FORMULATIONS BY ROUTE OF ADMINISTRATION

### 3.1 Oral Formulations

Pharmaceutical preparations for peroral administration form the majority of drug products most frequently administered. Within this group, availability for absorption usually decreases as solutions > suspensions > capsules > compressed tablets > coated tablets. Changes in this order may occur but are exceptions [201].

***Solutions*** This dosage form presents the drug component in a form most suitable for absorption. Hence, if there is no complexation or micellization with components of the formula, and if there is no degradation in the stomach, this dosage form will give maximal bioavailability and a rapid onset of action. The other advantage is that since solutions are homogeneous mixtures, the medication is uniformly distributed throughout the preparation. A number of drugs irritate the gastric mucosa when given in a concentrated form, such as a capsule or a tablet. This irritancy may be reduced by an adequate dilution factor if the drug is administered as a solution.

The disadvantage is that the solution form is bulky and hence requires increased costs for containers, handling, and storage. A drug in solution is usually the form most susceptible to degradation. Because solutions are, in many instances, fair to good microbiological growth media, adequate care must be taken to ensure a low initial bacterial count and prevention of growth during storage. The formulations are usually kept under refrigerated conditions to increase stability and decrease microbiological growth.

***Emulsions*** An emulsion is a thermodynamically unstable heterogeneous system consisting of at least one immiscible liquid intimately dispersed in another in the form of droplets. Oil emulsions (o/w) are usually manufactured to make the active oil phase easier to handle and sometimes more bioavailable. For example, griseofulvin, a poorly water-soluble drug, shows an increase in the rate and extent of absorption when dispersed in corn oil and then emulsified (o/w) [202]. The smaller amounts of drug present in the upper region of the small intestine at any particular time after oral administration of the corn oil emulsion, coupled with the inhibitory effect of lipid on proximal small intestine motility, allows the drug more time to dissolve in, and be absorbed from, the region of the intestinal tract where absorption is optimal. Also, the presence of emulsified corn oil in the intestinal tract stimulates bile excretion.

Emulsions, however, present problems with stability, not only of the emulsion itself, but also of the active ingredient. Changes in the solubility of emulsifiers and their hydrophilic–lipophilic balance, the viscosity of the emulsion, and partitioning of ingredients should always be considered.

***Suspensions*** A suspension is a two-phase system composed of a solid material dispersed in a liquid. A suspension may be considered the most desirable dosage form for a particular drug for a number of reasons. Certain drugs are chemically unstable when in solution but stable when a suitable salt or derivative is suspended. The suspension presents most of the advantages of the solution dose form: ease in

swallowing, greater ease in administration of unusually large doses, and infinite variability of dosage. The disadvantages of suspensions are that they are heterogeneous systems: homogeneity of dose throughout the entire use period is difficult to maintain, or components may settle before the dose is measured. To overcome this problem, a suspending agent is usually added to increase viscosity and hence slow settling in the system.

***Powders and Granules*** Powders and granules, in general, provide a better environment for maintaining stability of the active drug than solutions, emulsions, and suspensions. They may be used to advantage in a number of different ways as follows:

1. The powder volume is usually less than an equivalent liquid volume.
2. The powder or granule could be administered by being mixed with the animal's food. Less common, the powder may be formulated as a soluble powder, to be added to drinking water.
3. There is good stability (may be difficult with some premix components) and easy tailoring of dose.
4. Specialized dosing equipment or unnecessary handling of the animal is not required.

A disadvantage would be that the dose actually consumed may vary widely, depending on the adequacy of the mix and the animal's eating/drinking habits. A medication that changes the odor and taste characteristics of the usual diet (or water) may cause problems with palatability.

***Capsules*** Capsules are the most common formulations used in dogs. Unlike other dosage forms, the capsule is a tasteless, easily administered and digested container for different materials such as powders, granules, pellets, suspensions, emulsions, or oils. There are two basic types of capsules. The hard gelatin capsule [203] is usually used for solid fill formulations and the soft gelatin capsule [204] for liquid or semisolid fill formulations. Soft gelatin capsules should be used in preference to hard gelatin capsules probably only if the fill is fluid or if bioavailability of the hard gelatin capsule formulation does not meet with requirements. The capsule offers an advantage in that the particle size and distribution of the original starting compound are rarely altered by the final filling process. Capsules are efficient in hiding the taste and odor of an unpleasant drug. Filling of the capsules takes time and labor costs are therefore high. The gelatin capsule does not offer protection from oxygen or moisture, although if it has an opaque shell, it will protect the contents from light. Hence, information on the stability and moisture sensitivity of the formulation must be known before the gelatin shell dosage form is selected.

***Paste*** Use of an oral paste formulation should be considered to overcome some of the dosing problems associated with other conventional dose delivery systems [205, 206]. A well formulated paste can provide accurate dose delivery, the dose being easily individually tailored to the animal's body weight. A paste affords a high degree of safety to both the administrator and animal.

**TABLE 6 Advantages and Disadvantages of a Prolonged Release Delivery System**

| Advantages | Disadvantages |
|---|---|
| 1. The maintenance of a relatively constant drug concentration in blood can reduce fluctuations of drug concentrations in tissue and at biological target sites for a more uniform pharmacological response. | 1. Controlled release formulations are comparatively more costly to manufacture. |
| 2. The incidence and intensity of side effects that might be caused by excessively high peak plasma concentrations resulting from the administration of conventional dosage forms may be reduced. | 2. They do not permit termination. |
| 3. The number and frequency of doses is decreased; hence, there is a reduction in labor costs and trauma to the animal. | 3. Drugs with a low margin of safety may cause concern. |
| 4. Slow accumulation (so-called enhanced permeability and retention effect, EPR) in pathological sites with affected and leaky vasculature (such as tumors, inflammations, and infarcted areas) and improved or enhanced drug delivery in those areas can be achieved. | 4. A type of first-pass effect could be experienced. |
| 5. Gradual release of a drug from a dosage form may reduce or prevent irritation to the gastrointestinal mucosa by drugs that are irritating to the tissue in high concentration. | |

***Tablets or Boluses*** A tablet or bolus offers many advantages over other dose forms. The tablet offers relative ease of administration of an "accurate" dose and is readily adaptable to various dose sizes of medicinal substances. Tablets are portable and compact. They usually present the fewest problems with stability. However, the tablet form is usually not available in the early stages of preclinical safety drug assessment. Tablets also pose a problem because an animal can easily eject the tablet when out of sight of the doser, if the tablet has not been properly administered. To achieve optimal drug delivery to the intended site of absorption or action, the active ingredient should be protected from stomach function. Encapsulation or ptotective coating to achieve this should be considered. Gastric irritants, gastric degradable drugs, glandular products, and gastric nauseants, as well as drugs intended for action in the intestine, should be enteric coated to attain maximum effectiveness with least drug loss or side effects [205].

***Controlled Release Pharmaceuticals*** Long-circulating pharmaceuticals and new pharmaceutical carriers are rapidly developing. These dosage forms allow maintenance of a required level of a pharmaceutical agent in the blood for extended time intervals for better drug availability. The main advantages and disadvantages of prolonged release delivery are presented in Table 6.

### 3.2 Intravenous and Other Parenteral Formulations

The intravenous (IV) dosage form must be a solution and thus the physicochemical properties of the drug are taken into consideration. Once the formulation is established, its compatibility with blood must be determined to establish venous and perivenous tolerance. In addition, the duration of injection (or infusion) is

**TABLE 7 Advantages and Disadvantages of Parenteral Therapy and Drug Formulations**

| Advantages | Disadvantages |
|---|---|
| 1. The rate of onset of action can easily be controlled by the type of formulation and by the site of injection. | 1. IV infusion warrants special apparatus (i.e., pumps) and sometimes animal surgery prior to dose (e.g., surgically implanted jugular catheter or VAP). |
| 2. Nausea and vomiting are overcome. | 2. The dosage form has to be administered by trained personnel (especially for IV) and requires strict adherence to aseptic technique. |
| 3. Degradation of some drugs in the GI tract is avoided. | 3. Once a drug is injected into the veins, tissue, or other body fluids, its removal is impossible. |
| 4. Onset of action is faster and blood levels are more predictable. | 4. Subcutaneous and intramuscular injections can produce severe local irritation and tissue necrosis. |
| 5. Recommended for drugs that are poorly absorbed from the gastrointestinal tract. | 5. Parenteral administrations (especially IV infusion) are time consuming. |
| 6. Rapid onset of pharmacological effect. | 6. Special care has to be taken in packaging parenterals. |
| 7. When an animal is uncooperative or unconscious, parenteral administration is warranted. | |
| 8. Parenteral administration can be used for local effects when desired (e.g., in anesthesiology). | |

evaluated to establish the safety margin; however, it should be stated that this factor may not be applicable to human use because of anatomical and physiological differences.

Parenteral dose forms include aqueous, aqueous organic, and oily solutions, emulsions, suspensions, and solid forms for implantation. The parenterals should be sterile and pyrogen free; they are, if possible, buffered close to normal physiological pH and are preferably isotonic with body fluids [201]. The advantages and disadvantages of parenteral therapy should be considered and are presented in Table 7.

Severe local irritation and even tissue necrosis at the site of injection may be due to the solvent used [207, 208], or in many cases, the active drug results in a reaction. Rasmussen [209] has studied the tissue-damaging effect at the site of intramuscular injection of various preparations of antibiotics, other chemotherapeutic agents (sulfonamides, trimethoprim), certain drugs (lidocaine, diazepam, digoxin), and some vehicles. After intramuscular injection of physiological saline or sterile water, little or no tissue reaction was observed, while vehicles containing glycerol or propylene glycol caused severe damage at the injection site in swine, hens, and rabbits [210–212]. Pain is also likely to occur after intramuscular injection. Formulation techniques can often be used to overcome irritation and pain at the injection site.

### 3.3 Alternatives to Parenteral Administration (Nasal, Buccal, Transdermal, Rectal, and Vaginal)

***Nasal*** Recent progress in biotechnology led to an enormous increase in the number of peptide and protein drugs that usually require parenteral administration to become therapeutically useful [213, 214]. Nonparenteral administration across mucosal surfaces (e.g., rectal, buccal, vaginal, nasal, and pulmonary epithelia) has been considered as an alternative route of application to replace parenteral injections [215, 216]. Among them, nasal delivery of peptides attracted strong interest because of its favorable biopharmaceutical properties. An overview is given by Gizurarson [217] on the relevance of the intranasal administration of drugs, peptides, and vaccines, and some physiological factors that present a barrier to the use of this route. The structure of the nasal epithelium, its high blood perfusion, and lower detrimental enzyme concentrations may result in higher bioavailabilities as compared to oral administration [218]. Nasal drug administration facilitates self-medication, thereby improving patient compliance. However, protein and peptide drugs show frequently low and variable bioavailabilities after nasal administration, for example, for insulin <1% [219]. Therefore, nasal insulin therapy necessitates coadministration of absorption enhancers such as surfactants, bile salts, fusidic acid derivatives, medium-chain fatty acids, chelators, or enzyme inhibitors to increase absorption and hence bioavailability [220, 221]. Although these absorption enhancers significantly increase transmucosal peptide transport, local tolerability, and safety concerns related to membrane damage, ciliary toxicity, mucus discharge, and epithelial disruption need to be overcome [222].

Consequently, the design of appropriate peptide carrier systems avoiding harmful additives has become a topic of intensive research. Nanoscale colloidal carriers from hydrophilic and biocompatible polymers, such as chitosan [223], were found to be useful for nasal vaccine [224] and insulin delivery [225] without employing additional penetration enhancers. Not only size and surface properties but also the polymer composition and architecture affected the functional properties of nanocarriers, such as transport of macromolecules across biological surfaces. In general, nanoscale dimensions favor transport of particles across mucosal epithelia [226, 227].

Moreover, nanocarriers with bioadhesive properties (e.g., nanoparticles (NP) coated with Carbopol) prevented rapid nasal clearance [228]. Increasing residence time in the nasal cavity was thought to improve drug absorption [229, 230]. Through the design of polymer properties, nanoparticle characteristics such as hydrophilicity and surface charge can be manipulated. In particular, polycations such as poly(L-lysine), protamine, polyethylene imine, chitosan, and dextran derivatives seem to increase mucosal permeability [231–233]. The combination of enhancing properties, mediated through cationic surface charges, with the formation of nanoparticulate carriers was achieved by encapsulating insulin into chitosan NP by ionotropic gelation, which yielded a significant blood glucose reduction in rabbits after nasal administration [225]. Recent study demonstrates that the nanocomplexes significantly enhanced insulin absorption, suggesting that amphiphilic biodegradable comb-polymers offer a promising approach for nasal peptide delivery [234].

In the development of a nasal drug delivery system (NDDS), formulation characteristics and device capabilities must be harmonized in order for consistent delivery into the nasal cavity. The approach to improve nasal bioavailability is the use of

polymeric gel vehicles to increase nasal residence times and to control the rate of drug absorption. The aerosol droplet size distribution (DSD) is an important variable in defining the efficiency of aerosolized drugs. Low viscosity or shear-thinning vehicle systems were effectively atomized into small droplets using different nasal pump sprays, as previously reported [235]. There have been many reports that solutions of mixtures of certain polymers, surfactants, and excipients can exhibit molecular interactions that affect the rheological and physicochemical properties of the solutions [236]. The nature of these interactions can affect the ability of the solution to be aerosolized into small droplets and may alter the stability and liberation of the active components. The characteristic of nasal aerosol generation is dependent on a combination of actuation force, viscosity, rheological properties (e.g., changes in viscosity and appearance of the thixotropic system), surface tension, and pump design [235].

In the pharmaceutical industry, polymers are routinely used in the formulation of gels and in the stabilization of emulsions. In a recent publication, the insulin gel was formulated for intranasal administration using the combination of carbopol and hydroxypropyl methylcellulose as gelling agent [237]. The use of bioadhesive nasal gel containing insulin not only promoted the prolonged contact between the drug and the absorptive sites in the nasal cavity but also facilitated direct absorption of medicament through the nasal mucosa in rat and human studies. This study further demonstrates that administration of insulin intranasally in gel form is a pleasant and painless alternative to injectable insulin.

***Buccal*** Among the various transmucosal routes, buccal mucosa has excellent accessibility, an expanse of smooth muscle, and relatively immobile mucosa, hence suitable for administration of retentive dosage forms. Direct access to the systemic circulation through the internal jugular vein bypasses drugs from the hepatic first-pass metabolism, leading to high bioavailability. Other advantages, such as low enzymatic activity, suitability for drugs or excipients that mildly and reversibly damage or irritate the mucosa, painless administration, easy drug withdrawal, facility to include permeation enhancer/enzyme inhibitor or pH modifier in the formulation, and versatility in designing multidirectional or unidirectional release systems for local or systemic actions, project buccal adhesive drug delivery systems as promising options for continued research [238, 239]. Morishita et al. [240] demonstrate in a rat model that 20% Pluronic PF-127 gels containing unsaturated fatty acids are potential formulations for the buccal delivery of insulin. Sudhakar et al. [239] discuss the implication of various approaches for buccal adhesive delivery strategies applied for the systemic delivery of orally less inefficient drugs, in addition to the widely used local drug delivery. The growth rate for transmucosal drug delivery systems is expected to increase 11% annually through 2007. Worldwide market revenues are at $3B with the United States at 55%, Europe at 30%, and Japan at 10%. However, the effect of salivary scavenging and accidental swallowing of the delivery system and the barrier property of buccal mucosa stand as major limitations in the development of buccal adhesive drug delivery systems.

An ideal buccal adhesive system must have the following properties:

1. Should adhere to the site of attachment for several hours.
2. Should release the drug in a controlled fashion.

3. Should provide drug release in a unidirectional way toward the mucosa.
4. Should facilitate the rate and extent of drug absorption.
5. Should not cause any irritation or inconvenience to the patient.
7. Should not interfere with normal functions, such as talking or drinking.

Buccal adhesive drug delivery systems using matrix tablets, films, layered systems, disks, microspheres, ointments, and hydrogels have been studied and reported by several research groups. However, limited studies exist on novel devices that are superior to those of conventional buccal adhesive systems for the delivery of therapeutic agents through buccal mucosa. A number of formulation and processing factors can influence properties of the buccal adhesive system. There are numerous important factors, including biocompatibility (both the drug–device and device–environment interfaces), reliability, durability, environmental stability, accuracy, delivery scalability, and permeability, to be considered while developing such formulations. Traditionally, pharmaceutically acceptable polymers were used to enhance the viscosity of products to aid their retention in the oral cavity. Dry mouth is treated with artificial saliva solutions that are retained on mucosal surfaces to provide lubrication. These solutions contain sodium CMC as bioadhesive polymer.

Several buccal adhesive delivery devices were developed at the laboratory scale by many researchers either for local or systemic actions. They are broadly classified into three groups presented in Table 8.

*Selection of Animal Species for Buccal Studies* Apart from the specific methodology used to study buccal drug permeation characteristics, special attention is warranted for the selection of experimental animal species for such experiments. Many researchers have used small animals including rats and hamsters for permeability studies [241, 242]. But unlike humans, most laboratory animals have a totally keratinized oral lining and hence are not suitable. The rat has a buccal mucosa with a very thick, keratinized surface layer. The rabbit is the only laboratory rodent that has nonkeratinized mucosal lining similar to human tissue, but the sudden transition to keratinized tissue at the mucosal margins makes it hard to isolate the desired nonkeratinized region [243].

Among the larger experimental animals, monkeys are not practical models because of the difficulties associated with their maintenance. Dogs [244, 245] are easy to maintain and less expensive than monkeys [246] and their buccal mucosa is nonkeratinized and has a close similarity to that of the human buccal mucosa. However, we found in our laboratory that it is difficult to precisely deliver tablets intrabuccally to dogs as they tend to swallow tablets. Pigs also have nonkeratinized buccal mucosa similar to that of humans and their inexpensive handling and main-

**TABLE 8 Classification of Buccal Adhesive Delivery Devices**

| Solid Buccal Adhesive Dosage Forms | Semisolid Buccal Adhesive Dosage Forms | Liquid Buccal Adhesive Dosage Forms |
|---|---|---|
| Tablets | Gels | Solutions |
| Microparticles | Patches/films | Suspensions |
| Wafers | | |
| Lozenges | | |

tenance costs make them a highly suitable animal model for buccal drug delivery studies. In fact, the oral mucosa of pigs resembles that of humans more closely than any other animal in terms of structure and composition [247].

Adhesion of buccal adhesive drug delivery devices to mucosal membranes leads to an increased drug concentration gradient at the absorption site and therefore improved bioavailability of systemically delivered drugs. In addition, buccal adhesive dosage forms have been used to target local disorders at the mucosal surface (e.g., mouth ulcers) to reduce the overall dosage required and minimize side effects that may be caused by systemic administration of drugs. Researchers are now looking beyond traditional polymer networks to find other innovative drug transport systems. Much of the development of novel materials in controlled release buccal adhesive drug delivery is focusing on the preparation and use of a responsive polymeric system using a copolymer with desirable hydrophilic/hydrophobic interaction, block or graft copolymers, complexation networks responding via hydrogen or ionic bonding, and new biodegradable polymers especially from natural edible sources. At the current global scenario, scientists are finding ways to develop buccal adhesive systems through various approaches to improve the bioavailability of orally less inefficient drugs by manipulating the formulation strategies like inclusion of pH modifiers, enzyme inhibitors, and permeation enhancers. Novel buccal adhesive delivery systems, where drug delivery is directed toward buccal mucosa by protecting the local environment, is also gaining interest. Currently, solid dosage forms and liquids and gels applied to the oral cavity are commercially successful. The future direction of buccal adhesive drug delivery lies in vaccine formulations and delivery of small proteins/peptides. Microparticulate bioadhesive systems are particularly interesting as they offer protection to therapeutic entities as well as the enhanced absorption that results from increased contact time provided by the bioadhesive component. Exciting challenges remain to influence the bioavailability of drugs across the buccal mucosa. Many issues are yet to be resolved before safe and effective delivery through buccal mucosa. Successfully developing these novel formulations requires assimilation of a great deal of emerging information about the chemical nature and physical structure of these new materials.

***Transdermal*** In recent years, the development of transdermal dosage forms has been attracting increasing attention, owing to the several advantages that this administration route offers [248]. Transdermal delivery systems, when compared with conventional formulations, generally show a better control of blood levels, a reduced incidence of systemic toxicity, no hepatic first-pass metabolism, and higher patient compliance [249, 250]. In addition, the transdermal route provides sustained and controlled delivery [250]. It also allows continuous input of drugs with short biological half-lives and can eliminate pulsed entry into systemic circulation, which often causes undesirable side effects [251–254]. Release of drug from transdermal patches is controlled by the chemical properties of drug and delivery form as well as the physiological and physicochemical properties of the biological membrane [255]. Technological discoveries, over the last decade, have proved the feasibility of using several methodologies for enhancing transdermal drug delivery [256]. With a diverse set of tools to enhance skin permeability, the future of transdermal drug delivery looks brighter. The challenge now lies in converting these discoveries into useful products using newer excipients and technologies [257].

However, drug delivery via the skin is not an easy task because of the formidable barrier properties of the stratum corneum (SC). The majority of drugs do not appear to penetrate the skin at a sufficiently high rate to have therapeutic effectiveness. The feasibility of the transdermal route is thus limited to powerful actives presenting the appropriate features such as appropriate low molecular weight and high lipophilicity. Different approaches have been studied to overcome these limitations in order to develop efficient transdermal delivery systems [258–264].

The implication of skin permeation of drug on release rate profiles of the experimental formulations should not be ignored because the skin is known to have a substantial role in variation of release kinetics [265]. At an early stage and in a steady-state of skin permeation, diffusion of drug through appendages is considered to be significant. Selection of receptor fluid is also important for *in vitro* studies. A biphasic characteristic of the study fluid is desirable as the diffusion for the drug molecules through skin is routed through both aqueous and nonaqueous heterogeneous media. PEG 400 and water or normal saline are commonly selected to provide biphasic characteristics of the liquid [266].

The enhancement of skin flux with increase of drug concentrations may be due to accumulation of a greater amount of drug on the skin surface. The improvement in skin flux with the increase of PVP content may be owing to its antinucleating effect that converts the crystalline drug into an amorphous state, which generally possesses a high energy state with high solubility. The enhancement in drug solubility provides increased thermodynamic activity, which facilitates the skin permeation of drug [267].

One longstanding approach for improving transdermal drug delivery uses penetration enhancers (also called sorption promoters or accelerants), which penetrate into skin to reversibly decrease the barrier resistance [256]. Numerous compounds have been evaluated for penetration-enhancing activity, including sulfoxides (such as dimethylsulfoxide, DMSO), Azones (e.g., laurocapram), pyrrolidones (e.g., 2-pyrrolidone, 2P), alcohols and alkanols (ethanol or decanol), glycols (e.g., propylene glycol, PG, a common excipient in topically applied dosage forms), surfactants (also common in dosage forms), and terpenes. Different aspects of enhanced transdermal delivery are reviewed in Refs. 256 and 268.

Another way of enhancing transdermal delivery is iontophoresis. Iontophoresis involves application of small amounts of current to push charged drug molecules through skin, resulting in higher fluxes of the drug molecules for which permeation through skin is otherwise negligible. It provides the advantages of improved patient compliance, avoidance of first-pass metabolism, control rate of drug release from the patch, and the possibility of programmed delivery [269, 270]. The self-contained iontophoretic patches successfully delivered granisetron hydrochloride by iontophoresis and depot formation was observed in the dermal and subcutaneous structures in the skin [271]. In addition, the prodrug approach and, most recently, the use of colloidal carriers represent very promising strategies to improve transdermal delivery [260, 261].

***Rectal*** This route is used for local action on the rectum and lower colon. Glucose, digested proteins, and anesthetics are occasionally administered by high colonic irrigation to obtain systemic effect.

***Vaginal*** Vaginal tablets are ovoid or pear-shaped and prepared by granulation and compression. They can be formulated to exhibit two types of release mechanisms: (1) a slow release dissolution that retains the tablet's original shape; this tablet is ideal for drugs requiring low concentrations in the cavity for long periods; and (2) effervescent and disintegrating tablets that release quickly and ensure rapid distribution of the active drug for total local effect throughout the cavity. Both forms may often contain a buffer to maintain or change the vaginal pH to that required for normal physiologic vaginal flora.

Vaginal pessaries require specialized manufacturing equipment. The mass is usually a glycogelatin, theobroma oil, or synthetic base solid at room temperature but that dissolves or is fluid at body temperature. Pessaries pose special stability problems and should only be considered if the vaginal tablet does not give satifactory drug efficacy. Intravaginal and intrauterine drug delivery devices are often used.

### 3.4 Topical Formulations for Local Disorders (Dermal, Ocular, Ear)

Topical preparations may be used for local protective or therapeutic reasons (dusting powders, solutions, suspensions, lotions, liniments, creams, ointments, aerosols).

***Dermal Formulations*** Delivery of drugs to the skin is an effective and targeted therapy for local dermatological disorders. This route of drug delivery has gained popularity because it avoids first-pass effects, gastrointestinal irritation, and metabolic degradation associated with oral administration. Topical gel formulations provide a suitable delivery system for drugs because they are less greasy and can easily be removed from the skin. Percutaneous absorption of drugs from topical formulations involves the release of the drug from the formulation and permeation through skin to reach the target tissue. The release of the drug from topical preparations depends on the physicochemical properties of the vehicle and the drug employed. In order to enhance drug release and skin permeation, methods such as the selection of a suitable vehicle [271], coadministration of a chemical enhancer [268], and iontophoresis [272] have been studied.

The important considerations in the formulation of a dermatological preparation are, however, whether it is to be applied to a broken wound or an abrasion, whether it is to be rubbed into the affected area, whether it has to exhibit adhesiveness upon application to the skin, and whether it will deliver the active ingredient to the site required. Dusting powders, lotions, and aerosols are recommended formulations for application to abraded sites: lotions, liniments, creams, and ointments are best for unabraded sites.

For local therapeutic and systemic effects, the following should be considered: (1) the rate of dissolution of the drug in the vehicle, (2) the rate of diffusion of solubilized drug through the vehicle to the skin, and (3) the rate of permeation of the drug through the stratum corneum.

Formulation overcomes the problem of drug dissolution in the vehicle. Either the vehicle can be changed or the chemical structure of the active drug can be suitably altered (e.g., salt formation). The vehicle also governs the rate of diffusion and release of the drug to the stratum corneum. Thus, the vehicle is exceptionally

important in determining topical bioavailability [273–278]. Nevertheless, vehicle design is often ignored in the development of a suitable delivery system.

In the design of a suitable delivery system, the following should be considered [201]:

1. Permeability of the stratum corneum may be increased if it is hydrated by a suitable vehicle (dimethyl sulfoxide can result in superhydration [279, 280]) or by occlusive dressings or vehicles. Transport in some species, however, may be via skin appendages; hence, agents promoting increased stratum corneum permeability may have no or even a negative effect [281].
2. The thermodynamic activity of the drug in the vehicle should be higher than in the stratum corneum [282].
3. The formulation should be buffered to a pH such that the active ingredient is in the un-ionized state.
4. Keratolytics, lipid and polar solvents (acetone and alcohol), surfactants after protracted use, and some vehicles may cause damage to the stratum corneum, thus increasing penetrability.
5. Viscosity of the medium is inversely related to flux.

There are various topical gel formulations on the market for both local and systemic delivery of drugs and several others are in clinical trials [283]. Various delivery systems, such as liposomes [284], biodegradable microspheres [285, 286], and hydrogels [287], have been investigated in order to circumvent the susceptibility of peptides to degradation during formulation, storage, and administration. Proteins and peptides are compatible with hydrogels, such as nonionic cellulose polymers and poloxamer gels, which are nontoxic and exhibit reversible thermal characteristics. Poloxamers have been reported as suitable gel systems for various proteins such as insulin [240], α-chymotrypsin, and lactate dehydrogenase [288] and peptides such as deslorelin and gonadotropin-releasing hormone (GnRH) [289]. Pluronic F127 (PF127), a type of poloxamer, has been shown to enhance the stability of proteins, such as urease, and interleukin-2 [290]. These formulations are targeted to deliver both small molecules and macromolecules. Spantide II, a neurokinin-1 receptor (NK-1R) antagonist, for the treatment of inflammatory skin disorders (e.g., dermatitis and psoriasis) showed significant retention in epidermis and dermis from lotion and gel formulations when the lotion and gel formulations contained *N*-methyl-2-pyrrolidone (NMP) as a penetration enhancer [291].

Recently, patch devices were introduced for local dermal (as well as systemic) drug delivery [292, 293]. The optimized bioadhesive patch device offered a more patient-compliant and convenient alternative to tetracaine percutaneous anesthetic gel, particularly where large areas of skin are to be treated [292].

***Ocular Formulations*** Ophthalmic preparations are sterile aqueous or oily solutions, suspensions, emulsions, or ointments for topical administration by instillation. Ophthalmic solutions are usually isotonic and buffered to minimize irritation to the eye. All multiple-dose eye preparations must contain a bacteriostatic agent. These, however, cannot be used in the injured eye or during surgical intervention in the anterior eye chamber, because of possible irritation.

Solutions exhibit a fast drug pulse delivery: they produce an initial high dosage that rapidly lapses to very low concentrations. Ophthalmic suspensions should preferably have an aqueous vehicle containing the drug of low solubility. The suspension duration of action is more prolonged than in aqueous solution. Their disadvantage is the possibility of irritation due to suspended crystals or particles. Ointments remain in the eye longer than do solutions, both in the precorneal tear film and the conjunctival fornices, thereby increasing absorption of active ingredients [294]. The commonly used ointment bases may be toxic to the interior of the eye and should not be used when the cornea has been penetrated. Invasion of the ointment base into the internal chambers of the eye causes toxic endothelial damage, corneal edema, vascularization, and scarring [294].

Topical therapy with corticosteroids is quite common in the treatment of ocular inflammatory disorders. Many nonsteroidal anti-inflammatory drugs (NSAIDs) have been tested as ocular anti-inflammatory agents [295–297] so as to diminish the well documented ocular side effects caused by corticosteroids. Ketorolac tromethamine (KT), an aryl-acetic acid NSAID, is nonirritating to the eye at 0.5% wt/vol concentration [298]. Aqueous ocular drop of KT is an effective and safe anti-inflammatory agent for topical use following cataract surgery and intra ocular lens implantation [299–301].

Compared with aqueous drop, sesame and soybean oil drops of ketorolac provided higher ocular availability followed by ophthalmic ointment in rabbit ocular studies. The ointment formulation provided maximum sustained effect. Ketorolac aqueous drop with BAC and EDTA improved the rate of ocular absorption though not the extent of absorption [302]. Sesame and soybean oil drops containing 0.2% (wt/vol) ketorolac free acid and benzyl alcohol (0.5% vol/vol) and ophthalmic ointment containing 0.5% (wt/wt) KT (in dissolved state) showed higher *in vitro* transcorneal permeation with minimum corneal damage [303].

Iontophoresis as a noninvasive technique for ocular drug delivery has been investigated for many years. The technology of ocular iontophoresis has reached maturity in the aspect of device development. Therapeutic levels of drugs can be reached, in the anterior and posterior segments of the eye, after applying a convenient and noninvasive iontophoretic treatment [304]. It is clearly seen that ocular iontophoresis has clinical potential and importance as a local delivery system for many drugs. The iontophoretic treatment is already a promising tool for delivering anti-inflammatory and antibiotic drugs to the eye. Iontophoresis will soon be used in delivering chemotherapeutic agents for retinoblastoma and ocular melanomas or as home-use device for treating glaucoma. It is a matter of time until iontophoresis will routinely be used in the ophthalmic field [304].

Transscleral iontophoresis is a good potential alternative for multiple intravitreal injections or systemic therapy used for posterior ocular disorders, such as endophthalmitis, uveitis, retinitis, optic nerve atrophy, pediatric retinoblastoma, and age-related macular degeneration (AMD). The search for an alternative drug delivery system derives from the serious intraocular complications occurring after recurrent intravitreal injections including retinal detachment, vitreous hemorrhage, endophthalmitis, and cataract. One recent study [305] showed that compared to IV administration, transscleral Coulomb Control iontophoresis (CCI) achieved higher and more sustained tissue concentrations of methylprednisolone with negligible systemic absorption. These data demonstrate that high levels of methylprednisolone

can safely be achieved in intraocular tissues and fluids of the rabbit eye, using CCI. Results from another study [306] suggest that focal delivery of carboplatin using subconjunctival injection or iontophoretic delivery transmits drug more effectively than intravenous delivery into the target tissues of the vitreous, choroid, retina, and optic nerve in the rabbit eye. Both focal applications resulted in significantly higher peak concentrations of carboplatin in the choroid, retina, optic nerve, and vitreous than those obtained after intravenous delivery. Focal chemotherapeutic delivery resulted in dramatically decreased carboplatin levels in the blood plasma compared with intravenous delivery.

***Ear Formulations*** Otic dosage forms are intended for administration either on the outer ear or into the auditory canal. They include a number of dosage forms: solutions, suspensions, ointments, otic cones, and powders. Their primary use is either to remove wax or supply local drug treatment. Direct delivery of new therapies for treatment of sensorineural hearing loss to the fluids of the inner ear is necessary because of the presence of a blood–labyrinth drug barrier, which is anatomically and functionally similar to the blood–brain barrier [307, 308]. Direct delivery also has significant potential advantages for therapeutic application. Drugs are largely unaltered by metabolic changes that inevitably occur with other routes of administration, and have ready access to the sensory cells of the inner ear (the hair cells) and the synaptic regions of hair cells [309]. Delivery directly to the inner ear can avoid undesirable systemic side effects that some drugs may produce. The performance of an extracorporeal reciprocating perfusion system in guinea pigs was recently described by Chen et al. [310].

### 3.5 Inhalation Formulations (Vapors and Aerosols)

Many biologically active peptides have been discovered recently and have attracted attention as new drugs. Because of transport and enzymatic barriers, clinical dosage forms of these peptides have been primarily parenteral forms [311]. Development of sustained release forms of these peptide drugs is also being actively researched [312–314] and the pulmonary route would seem to be a promising alternative for delivering them, because many drugs that are poorly absorbed from other mucosal sites are well absorbed from the lungs [315]. This route of administration offers several advantages over the conventional gastrointestinal pathway, including large surface area, extensive vasculature, easily permeable membrane, and low intracellular and extracellular enzymatic activity [316–319]. Recent clinical and preclinical reports reveal that delivery of peptide drugs such as leuprolide acetate and insulin is feasible through the pulmonary route [320–323]. However, the bioavailability of the drugs having relatively high molecular weight is still poor through the pulmonary route compared with the parenteral route.

Recent research is focused on the delivery of systemically acting drugs via the pulmonary route [324]. Dry powder inhalations are a promising application form for peptides and proteins for systemic delivery because they overcome the drawbacks of oral and invasive delivery forms [315].

The administration of liposome-encapsulated drugs by aerosols seems to be a feasible way of targeting these delivery systems to the lung. The tolerability and safety of liposome aerosols have previously been tested in animals as well as in

human volunteers; no untoward effects have been recognized [325, 326]. Liposomes are also known to sustain the release of the entrapped drug(s) and to decrease the mucociliary clearance of the drug(s) because of their surface viscosity. Therefore, more effective and sustained systemic absorption of a drug would be attained by administering the drug containing liposomes in the respiratory tract. The developed liposomal dry powder inhaler of Leucinostatin demonstrated approximately 50% bioavailability compared with the SC route in rats [311]. This study justifies the role of the pulmonary route as a promising alternative to the presently available SC route. Green et al. [327] review the formulation process of inhalable drugs. They showed that high pressure media milling (HPMM) yields stable, fine-particle dispersion of medicament in hydrofluorocarbone propellant (HCP), which enhances a metered dose inhaler. HCP is currently mostly used for inhalation dose formulation as a more environmentally friendly product than previously used chlorofluorocarbons. However, eliminating chlorine from modern propellants introduced a solubility challenge for which the HPMM process helps to compensate.

## 4 CONCLUSION

Rapid developments in the field of molecular biology and gene technology resulted in generation of many macromolecular drugs including peptides, proteins, polysaccharides, and nucleic acids in great numbers and possessing superior pharmacological efficacy with site specificity and devoid of untoward and toxic effects. However, the main impediment for the oral delivery of these drugs as potential therapeutic agents is their extensive presystemic metabolism and instability in an acidic environment, resulting in inadequate and erratic oral absorption. This requires development of new drug delivery systems and increased usage of other routes of administration compared to the traditional oral and parenteral routes. These new therapies and new delivery systems require new formulations, which should be evaluated for toxicity and efficacy in preclinical studies using the proposed routes of administration.

## REFERENCES

1. Irie T, Uekama K. Cyclodextrins in peptide and protein delivery. *Adv Drug Deliv Rev* 1999;36:101–123.
2. Pearlman R, Wang Y J, Eds. *Formulation, Characterization, and Stability of Protein Drugs*. New York: Plenum Press; 1996.
3. Rászcz I. In *Drug Formulation*. Hoboken, NJ: Wiley; 1989.
4. Gad SC, Ed. *Drug Safety Evaluation*. Hoboken, NJ: Wiley-Interscience; 2002.
5. Gelderblom H, Verweij J, Nooter K, Sparreboom A. Cremophor EL, the drawbacks and advantages of vehicle selection for drug formulation. *Eur J Cancer* 2001;37(13): 1590–1598.
6. Aungst BJ. Intestinal permeation enhancers. *J Pharm Sci* 2000;89(4):429–442.
7. Deshpande AA, Rhodes CT, Shah NH, Malick AW. Control-release drug delivery systems for prolonged gastric residence: an overview. *Drug Dev Ind Pharm* 1996;22: 531–539.

8. Hwang S-J, Park H, Park K. Gastric retentive drug-delivery systems. *Crit Rev Ther Drug Carrier Syst* 1998;15:243–284.
9. Hoffman A, Stepensky D, Lavy E, Eyal S, Klausner E, Friedman M. Pharmacokinetic and pharmacodynamic aspects of gastroretentive dosage forms. *Int J Pharm* 2004; 277(1–2):141–153.
10. Aungst BJ, Saitoh H, Burcham DL, Huang S-M, Mousa SA, Hussain MA. Enhancement of the intestinal absorption of peptides and non-peptides. *J Control Release* 1996;41: 19–31.
11. Fix JA. Strategies for delivery of peptides utilizing absorption-enhancing agents. *J Pharm Sci* 1996;85:1282–1285.
12. Hochman J, Artursson P. Mechanisms of absorption enhancement and tight junction regulation. *J Control Release* 1994;29:253–267.
13. Swenson ES, Curatolo WJ. Intestinal permeability enhancement for proteins, peptides and other polar drugs: mechanisms and potential toxicity. *Adv Drug Deliv Rev* 1992;8: 39–92.
14. Ings RMJ. In Smolen VF, Ball L, Eds. *Animal Studies and the Bioavailability Testing of Drug Products in Control Drug Bioavailability*. Hobohen, NJ: Wiley;1984, p 43.
15. Eriksson HJC, Hinrichs WLJ, van Veen B, Somsen GW, de Jong GJ, Frijlink HW. Investigations into the stabilisation of drugs by sugar glasses: I. Tablets prepared from stabilised alkaline phosphatase. *Int J Pharm* 2002;249: 59–70.
16. Wang W. Instability, stabilization, and formulation of liquid protein Pharmaceuticals. *Int J Pharm* 1999;185:129–188.
17. Xing L, Dawei C, Liping X, Rongquing Z. Oral colon-specific drug delivery for bee venom peptide: development of a coated calcium alginate gel beads-entrapped liposome. *J Control Release* 2003;93:293–300.
18. Shargel L, Yu A. In *Applied Biopharmaceutics and Pharmacokinetics*, 4th ed. New York: McGraw-Hill;1999, Ch 5.
19. George M, Abraham TE. Polyionic hydrocolloids for the intestinal delivery of protein drugs: alginate and chitosan—a review. *J Control Release* 2006;14(1):1–14.
20. Sayani AP, Chien YW. Systemic delivery of peptides and proteins across absorptive mucosae. *CRC Crit Rev Ther Drug Carrier Syst* 1996;13:85–184.
21. Langel Ü. In *Cell-Penetrating Peptides: Processes and Applications*. New York: CRC Press; 2002.
22. Weecharangsan W, Opanasopit P, Ngawhirunpat T, Rojanarata T, Apirakaramwong A. Chitosan lactate as a nonviral gene delivery vector in COS-1 cells. *AAPS PharmSciTech* 2006;7(3): Article 66.
23. De Smedt SC, Demeester J, Hennink WE. Cationic polymer based gene delivery systems. *Pharm Res* 2000;17:113–126.
24. Lundstrom K. Latest development in viral vectors for gene therapy. *Trends Biotechnol* 2003;21:117–122.
25. Li S, Huang L. Nonviral gene therapy: promises and challenges. *Gene Ther* 2000;7:31–34.
26. Luo D, Saltzman WM. Synthetic DNA delivery systems. *Nat Biotechnol* 2000;18:33–37.
27. Felgner JH, Kumar R, Sridhar CN, Wheeler CJ, Tsai YJ, Border R, Ramsey P, Martin M, Felgner PL. Enhanced gene delivery and mechanism studies with a novel series of cationic lipid formulations. *J Biol Chem* 1994;269:2550–2561.
28. Farhood H, Serbina N, Huang L. The role of dioleoyl phosphatidylethanolamine in cationic liposome mediated gene transfer. *Biochim Biophys Acta* 1995;1235:289–295.

29. Zabner J, Fasbender AJ, Moninger T, Poellinger KA, Welsh MJ. Cellular and molecular barriers to gene transfer by a cationic lipid. *J Biol Chem* 1995;270:18997–19007.
30. Lee KY, Kwon IC, Kim YH, Jo WH, Jeong SY. Preparation of chitosan self-aggregates as a gene delivery system. *J Control Release* 1998;51:213–220.
31. Mansouri S, Lavigne P, Corsi K, Benderdour M, Beaumont E, Fernandes JC. Chitosan–DNA nanoparticles as non-viral vectors in gene therapy: strategies to improve transfection efficacy. *Eur J Pharm Biopharm* 2004;57:1–8.
32. Thomas CE, Ehrhardt A, Kay MA. Progress and problems with the use of viral vectors for gene therapy. *Nat Rev Genet* 2003;4:346–358.
33. Pfeifer A, Verma IM. Gene therapy: promises and problems. *Annu Rev Genomics Hum Genet* 2001;2:177–211.
34. Anderson WF. Human gene therapy. *Nature* 1998;392:25–30.
35. Anwer K, Rhee BG, Mendiratta SK. Recent progress in polymeric gene delivery systems. *Crit Rev Ther Drug Carrier Syst* 2003;20:249–293.
36. Han S, Mahato RI, Sung YK, Kim SW. Water-soluble lipopolymer for gene delivery. *Bioconjug Chem* 2001;12:337–345.
37. Merdan T, Kopecek J, Kissel T. Prospects for cationic polymers in gene and oligonucleotide therapy against cancer. *Adv Drug Deliv Rev* 2002;54:715–758.
38. Lim YB, Choi YH, Park JS. A self-destroying polycationic polymer: biodegradable poly(4-hydroxy-l-proline ester). *J Am Chem Soc* 1999;121:5633–5639.
39. Putnam D, Langer R. Poly(4-hydroxy-l-proline ester): low-temperature polycondensation and plasmid DNA complexation. *Macromolecules* 1999;32:3658–3662.
40. Lim YB, Han SO, Kong HU, Lee Y, Park JS, Jeong B, Kim SW. Biodegradable polyester poly[alpha-(4 aminobutyl)-l-glycolic acid], as a non-toxic gene carrier. *Pharm Res* 2000; 17:811–816.
41. Bikram M, Ahn CH, Chae SY, Lee MY, Yockman JW, Kim SW. Biodegradable poly(ethylene glycol)-*co*-poly(L-lysine)-*g*-histidine multiblock copolymers for nonviral gene delivery. *Macromolecules* 2004;37:1903–1916.
42. Lynn DM, Langer R. Degradable poly(beta-amino esters):synthesis characterization and self-assembly with plasmid DNA. *J Am Chem Soc* 2000;122:10761–10768.
43. Anderson DG, Lynn DM, Langer R. Semi-automated synthesis and screening of a large library of degradable cationic polymers for gene delivery. *Angew Chem Int Ed Engl* 2003;42:3153–3158.
44. Liu Y, Wu DC, Ma YX, Tang GP, Wang S, He CB, Chung TS, Goh S. Novel poly(amino ester)s obtained from Michael addition polymerizations of trifunctional amine monomers with diacrylates: safe and efficient DNA carriers. *Chem Commun* 2003; 2630–2631.
45. Anderson DG, Akinc A, Hossain N, Langer R. Structure/property studies of polymeric gene delivery using a library of poly(beta-amino esters). *Mol Ther* 2005;11:426–434.
46. Lim YB, Kim SM, Suh H, Park JS. Biodegradable, endosome disruptive, and cationic network-type polymer as a highly efficient and nontoxic gene delivery carrier. *Bioconjug Chem* 2002;13:952–957.
47. Luten J, van Steenis JH, van Someren R, Kemmink J, Schuurmans-Nieuwenbroek NME, Koning GA, Crommelin DJA, van Nostrum CF, Hennink WE. Water-soluble biodegradable cationic polyphosphazenes for gene delivery. *J Control Release* 2003;89:483–497.
48. Ahn CH, Chae SY, Bae YH, Kim SW. Biodegradable poly(ethylenimine) for plasmid DNA delivery. *J Control Release* 2002;80:273–282.

49. Petersen H, Merdan T, Kunath F, Fischer D, Kissel T. Poly(ethylenimine-*co*-l-lactamide-*co*-succinamide): a biodegradable polyethylenimine derivative with an advantageous pH-dependent hydrolytic degradation for gene delivery. *Bioconjug Chem* 2002;13: 812–821.
50. Chae SY, Son S, Lee M, Jang M-K, Nah J-W. Deoxycholic acid-conjugated chitosan oligosaccharide nanoparticles for efficient gene carrier. *J Control Release* 2005;109(1–3): 330–344.
51. Zhong Z, Song Y, Engbersen JFJ, Lok MC, Hennink WE, Feijen J. A versatile family of degradable non-viral gene carriers based on hyperbranched poly(esteramine)s. *J Control Release* 2005;109(1–3):317–329.
52. Gao C, Tang W, Yan DY. Synthesis and characterization of water-soluble hyperbranched poly(ester amine)s from diacrylates and diamines. *J Polym Sci Polym Chem* 2002;40: 2340–2349.
53. Jikei M, Kakimoto M. Hyperbranched polymers: a promising new class of materials. *Prog Polym Sci* 2001;26:1233–1285.
54. Wightman L, Kircheis R, Rossler V, Carotta S, Ruzicka R, Kursa M, Wagner E. Different behavior of branched and linear polyethylenimine for gene delivery *in vitro* and *in vivo*. *J Gene Med* 2001;3:362–372.
55. Boussif O, Lezoualch F, Zanta MA, Mergny MD, Scherman D, Demeneix B, Behr JP. A versatile vector for gene and oligonucleotide transfer into cells in culture and in-vivo-polyethylenimine. *Proc Natl Acad Sci USA* 1995;92:7297–7301.
56. Abdallah B, Hassan A, Benoist C, Goula D, Behr JP, Demeneix BA. A powerful nonviral vector for *in vivo* gene transfer into the adult mammalian brain: polyethylenimine. *Hum Gene Ther* 1996;7:1947–1954.
57. Kircheis R, Wightman L, Wagner E. Design and gene delivery activity of modified polyethylenimines. *Adv Drug Deliv Rev* 2001;53:341–358.
58. Fischer D, Bieber T, Li YX, Elsasser HP, Kissel T. A novel non-viral vector for DNA delivery based on low molecular weight branched polyethylenimine: effect of molecular weight on transfection efficiency and cytotoxicity. *Pharm Res* 1999;16:1273–1279.
59. Fewell JG, Matar M, Slobodkin G, Han S-O, Rice J, Hovanes B, Lewis DH, Anwer K. Synthesis and application of a non-viral gene delivery system for immunogene therapy of cancer. *J Control Release* 2005;109:288–298.
60. Zhang G, Budker V, Williams P, Subbotin V, Wolff JA. Efficient expression of naked DNA delivered intraarterially to limb muscles of nonhuman primates. *Hum Gene Ther* 2001;12:427–438.
61. Fewell JG, MacLaughlin F, Mehta V, Gondo M, Nicol F, Wilson E, Smith LC. Gene therapy for the treatment of hemophilia B using PINC-formulated plasmid delivered to muscle with electroporation. *Mol Ther* 2001;3:574–583.
62. Herweijer H, Wolff JA. Progress and prospects: naked DNA gene transfer and therapy. *Gene Ther* 2003;10:453–458.
63. Database on Human Gene Transfer Trials (GeMCRIS). http://www4.od.nih.gov/oba/RAC/GeMCRIS_public.htm.
64. Ma H, Diamond SL. Non viral gene therapy and its delivery systems. *Curr Pharm Biotechnol* 2001;2:1–17.
65. Davis ME. Non-viral gene delivery systems. *Curr Opin Biotechnol* 2002;13:128–131.
66. Anwer K, Kao G, Rolland A, Driessen WH, Sullivan SM. Peptide-mediated gene transfer of cationic lipid/plasmid DNA complexes to endothelial cells. *J Drug Target* 2004;12: 215–221.

67. Wang DA, Narang AS, Kotb M, Gaber AO, Miller DD, Kim SW, Mahato RI. Novel branched poly(ethylenimine)–cholesterol water-soluble lipopolymers for gene delivery. *Biomacromolecules* 2002;3:1197–1207.
68. Furgeson DY, Cohen RN, Mahato RI, Kim SW. Novel water insoluble lipoparticulates for gene delivery. *Pharm Res* 2002;19:382–390.
69. Sonawane ND, Szoka FC Jr, Verkman AS. Chloride accumulation and swelling in endosomes enhances DNA transfer by polyamine–DNA polyplexes. *J Biol Chem* 2003;278: 44826–44831.
70. Chollet P, Favrot MC, Hurbin A, Coll JL. Side-effects of a systemic injection of linear polyethylenimine–DNA complexes. *J Gene Med* 2002;4:84–91.
71. Godbey WT, Wu KK, Mikos AG. Poly(ethylenimine) and its role in gene delivery. *J Control Release* 1999;60:149–160.
72. Furgeson DY, Chan WS, Yockman JW, Kim SW. Modified linear polyethylenimine–cholesterol conjugates for DNA complexation. *Bioconjug Chem* 2003;14:840–847.
73. Lv H, Zhang S, Wang B, Cui S, Yan J. Toxicity of cationic lipids and cationic polymers in gene delivery. *J Control Release* 2006;114(1):100–109.
74. Crooke ST. Molecular mechanisms of antisense drugs: RNase H. *Antisense Nucleic Acid Drug Dev* 1998;8:133–134.
75. Lindner LH, Brock R, Arndt-Jovin D, Eibl H. Structural variation of cationic lipids: minimum requirement for improved oligonucleotide delivery into cells. *J Control Release* 2006;110(210):444–456.
76. Zamecnik PC, Stephenson ML. Inhibition of Rous sarcoma virus replication and cell transformation by a specific oligodeoxynucleotide. *Proc Natl Acad Sci USA* 1978;75: 280–284.
77. Stephenson ML, Zamecnik PC. Inhibition of Rous sarcoma viral RNA translation by a specific oligodeoxyribonucleotide. *Proc Natl Acad Sci USA* 1978;75:285–288.
78. De Clercq E, Merigan TC. Requirement of a Stable Secondary Structure for the Antiviral Activity of Polynucleotides. *Nature* 1969;222:1148–1152.
79. Alberts SR, Schroeder M, Erlichman C, Steen PD, Foster NR, Moore DF Jr, Rowland KM Jr, Nair S, Tschetter LK, Fitch TR. Gemcitabine and ISIS-2503 for patients with locally advanced or metastatic adenocarcinoma: a North Central Cancer Treatment Group phase II trial. *J Clin Oncol* 2004;22:4944–4950.
80. Kim R, Emi M, Tanabe K, Toge T. Therapeutic potential of antisense Bcl-2 as a chemosensitizer for cancer therapy. *Cancer* 2004;101:2491–2502.
81. Loke SL, Stein CA, Zhang XH, Mori K, Nakanishi M, Subasinghe C, Cohen JS, Neckers LM. Characterization of oligonucleotide transport into living cells. *Proc Natl Acad Sci USA* 1989;86:3474–3478.
82. Wu-Pong S, Weiss TL, Hunt CA. Antisense c-myc oligodeoxyribonucleotide cellular uptake. *Pharm Res* 1992;9:1010–1017.
83. Yakubov LA, Deeva EA, Zarytova VF, Ivanova EM, Ryte AS, Yurchenko LV, Vlassov VV. Mechanism of oligonucleotide uptake by cells: involvement of specific receptors? *Proc Natl Acad Sci USA* 1989;86:6454–6458.
84. Zamecnik P, Aghajanian J, Zamecnik M, Goodchild J, Witman G. Electron micrographic studies of transport of oligrodeoxynucleotides across eukaryotic cell membranes. *Proc Natl Acad Sci USA* 1994;91:3156–3160.
85. Wagner RW, Matteucci MD, Lewis JG, Gutierrez AJ, Moulds C, Froehler BC. Antisense gene inhibition by oligonucleotides containing C-5 propyne pyrimidines. *Science* 1993;260:1510–1513.

86. Bertling WM, Gareis M, Paspaleeva V, Zimmer A, Kreuter J, Nurnberg E, Harrer P. Use of liposomes, viral capsids, and nanoparticles as DNA carriers. *Biotechnol Appl Biochem* 1991;13:390–405.
87. Gewirtz AM, Sokol DL, Ratajczak MZ. Nucleic acid therapeutics: state of the art and future prospects. *Blood* 1998;92:712–736.
88. Zobel HP, Junghans M, Maienschein V, Werner D, Gilbert M, Zimmermann H, Noe C, Kreuter J, Zimmer A. Enhanced antisense efficacy of oligonucleotides adsorbed to monomethylaminoethylmethacrylate methylmethacrylate copolymer nanoparticles. *Eur J Pharm Biopharm* 2000;49:203–310.
89. Zelphati O, Degols G, Loughrey H, Leserman L, Pompon A, Puech F, Maggio AF, Imbach JL, Gosselin G. Inhibition of HIV-1 replication in cultured cells with phosphorylated dideoxyuridine derivatives encapsulated in immunoliposomes. *Antiviral Res* 1993;21:181–195.
90. Capaccioli S, DiPasquale G, Mini E, Mazzei T, Quattrone A. Cationic lipids improve antisense oligonucleotide uptake and prevent degradation in cultured cells and in human serum. *Biochem Biophys Res Commun* 1993;197:818–825.
91. Bennett CF, Chiang MY, Chan H, Shoemaker JE, Mirabelli CK. Cationic lipids enhance cellular uptake and activity of phosphorothioate antisense oligonucleotides. *Mol Pharmacol* 1992;41:1023–1033.
92. Weyermann J, Lochmann D, Zimmer A. Comparison of antisense oligonucleotide drug delivery systems. *J Control Release* 2004;100:411–423.
93. Lappalainen K, Urtti A, Soderling E, Jaaskelainen I, Syrjanen K, Syrjanen S. Cationic liposomes improve stability and intracellular delivery of antisense oligonucleotides into CaSki cells. *Biochim Biophys Acta* 1994;1196:201–208.
94. Sazani P, Kole R. Therapeutic potential of antisense oligonucleotides as modulators of alternative splicing. *J Clin Invest* 2003;112:481–486.
95. Remaut K, Lucas B, Braeckmans K, Sanders NN, Demeester J, De Smedt SC. Protection of oligonucleotides against nucleases by pegylated and non-pegylated liposomes as studied by fluorescence correlation spectroscopy. *J Control Release* 2005;110(1): 212–226.
96. Shi F, Hoekstra D. Effective intracellular delivery of oligonucleotides in order to make sense of antisense. *J Control Release* 2004;97:189–209.
97. Agrawal S. Importance of nucleotide sequence and chemical modifications of antisense oligonucleotides. *Biochim Biophys Acta* 1999;1489:53–68.
98. Brown DA, Kang SH, Gryaznov SM, Dedionisio L, Heidenreich O, Sullivan S, Xu X, Nerenberg MI. Effect of phosphorothioate modification of oligodeoxynucleotides on specific protein-binding. *J Biol Chem* 1994;269:26801–26805.
99. Hogrefe RI. An antisense oligonucleotide primer. *Antisense Nucleic Acid Drug Dev* 1999;9:351–357.
100. Akhtar S, Hughes MD, Khan A, Bibby M, Hussain M, Nawaz Q, Double J, Sayyed P. The delivery of antisense therapeutics. *Adv Drug Deliv Rev* 2000;44:3–21.
101. Lucas B, Van Rompaey E, Remaut K, Sanders N, De Smedt S, Demeester J. On the biological activity of anti-ICAM-1 oligonucleotides complexed to non-viral carriers. *J Control Release* 2004;96:207–219.
102. Dheur S, Dias N, van-Aerschot A, Herdewijn P, Bettinger T, Remy JS, Helene C, Saison-Behmoaras ET. Polyethylenimine but not cationic lipid improves antisense activity of 3′-capped phosphodiester oligonucleotides. *Antisense Nucleic Acid Drug Dev* 1999;9: 515–525.

103. Huppi K, Martin SE, Caplen NJ. Defining and assaying RNAi in mammalian cells. *Mol Cell* 2005;17:1–10.
104. Schiffelers RM, Mixson AJ, Ansari AM, Fens MHAM, Tang Q, Zhou Q, Xu J, Molema G, Lu PY, Scaria PV. Transporting silence: design of carriers for siRNA to angiogenic endothelium. *J Control Release* 2005;109(1–3):5–14.
105. Braasch DA, Paroo Z, Constantinescu A, Ren G, Oz OK, Mason RP, Corey DR. Biodistribution of phosphodiester and phosphorothioate siRNA. *Bioorg Med Chem Lett* 2004;14:1139–1143.
106. Schiffelers RM, Ansari A, Xu J, Zhou Q, Tang Q, Storm G, Molema G, Lu PY, Scaria PV, Woodle MC. Cancer siRNA therapy by tumor selective delivery with ligand-targeted sterically stabilized nanoparticle. *Nucleic Acids Res* 2004;32:149.
107. Kim B, Tang Q, Biswas PS, Xu J, Schiffelers RM, Xie FY, Ansari AM, Scaria PV, Woodle MC, Lu PY, Rouse BT. Inhibition of ocular angiogenesis by siRNA targeting vascular endothelial growth factor pathway genes: therapeutic strategy for herpetic stromal keratitis. *Am J Pathol* 2004;165:2177–2185.
108. Nielsen PE, Egholm M, Berg RH, Buchardt O. Sequence-selective recognition of DNA by strand displacement with a thymine-substituted polyamide. *Science* 1991;254:1497–1500.
109. Egholm M, Buchardt O, Christensen L, Behrens C, Freier SM, Driver DA, Berg RH, Kim SK, Norden B. PNA hybridizes to complementary oligonucleotides obeying the Watson–Crick hydrogen-bonding rules. *Nature* 1993;365:566–568.
110. Jensen KK, Orum H, Nielsen PE, Norden B. Kinetics for hybridization of peptide nucleic acids (PNA) with DNA and RNA studied with the BIAcore technique. *Biochemistry* 1997;36:5072–5077.
111. Giesen U, Kleider W, Berding C, Geiger A, Orum H, Nielsen PE. A formula for thermal stability prediction of PNA/DNA duplexes. *Nucleic Acids Res* 1998;26:5004–5006.
112. Tomac S, Sarkar M, Ratilainen T, Wittung P, Nielsen PE, Norden B, Graslund A. Ionic effects on the stability and conformation of peptide nucleic acid complexes. *J Am Chem Soc* 1996;118:5544–5552.
113. Eriksson M, Nielsen PE. PNA nucleic acid complexes. Structure, stability and dynamic. *Q Rev Biophys* 1996;29:369–394.
114. Pooga M, Soomets U, Hallbrink M, Valkna A, Saar K, Rezaei K, Kahl U, Hao JX, Xu XJ, Wiesenfeld-Hallin Z, Hokfelt T, Bartfai A, Langel U. Cell penetrating PNA constructs regulate galanin receptor levels and modify pain transmission *in vivo*. *Nat Biotechnol* 1998;16:857–861.
115. Hamilton SE, Simmons CG, Kathiriya IS, Corey DR. Cellular delivery of peptide nucleic acids and inhibition of human telomerase. *Chem Biol* 1999;6:343–351.
116. Fraser GL, Holmgren J, Clarke PBS, Wahlestedt C. Antisense inhibition of delta-opioid receptor gene function *in vivo* by peptide nucleic acids. *Mol Pharmacol* 2000;57:725–731.
117. Wittung P, Kajanus J, Edwards K, Nielsen P, Norden B, Malmstrom BG. Phospholipid membrane-permeability of peptide nucleic-acid. *FEBS Lett* 1995;365:27–29.
118. Buchardt O, Egholm M, Berg RH, Nielsen PE. Peptide nucleic-acids and their potential applications in biotechnology. *Trends Biotechnol* 1993;11:384–386.
119. Meier C, Engels JW. Peptide nucleic-acids (PNAs)—unusual properties of non-ionic oligonucleotide analogs. *Angew Chem Int Ed* 1992;31:1008–1010.
120. Chiarantini L, Cerasi A, Fraternale A, Millo E, Benatti U, Sparnacci K, Laus M, Ballestri M, Tondelli L. Comparison of novel delivery systems for antisense peptide nucleic acids. *J Control Release* 2005;109(1–3):24–36.

121. Sezgin Z, Yüksel N, Baykara T. Preparation and characterization of polymeric micelles for solubilization of poorly soluble anticancer drugs. *Eur J Pharm Biopharm* 2006;64(3): 261–268.

122. Strickley R. Solubilizing excipients in oral and injectable formulations. *Pharm Res* 2004;21:201–230.

123. Mathot F, van Beijsterveldt L, Préat V, Brewster M, Ariën A. Intestinal uptake and biodistribution of novel polymeric micells after oral administration. *J Control Release* 2006;111(1–2):47–55.

124. Lavasanifar A, Samuel J, Kwon G. Poly(ethylene oxide)-*block*-poly(L-amino acid) micelles for drug delivery. *Adv Drug Deliv Rev* 2002;54:169–190.

125. Torchilin V. Structure and design of polymeric surfactant-based drug delivery systems. *J Control Release* 2001;73:137–172.

126. Jones M-C, Leroux J-C. Polymeric micelles—a new generation of colloidal drug carriers. *Eur J Pharm Biopharm* 1999;48:101–111.

127. Kataoka K, Harada A, Nagasaki Y. Block copolymer micelles for drug delivery: design characterization and biological significance. *Adv Drug Deliv Rev* 2001;47:113–131.

128. Kwon G, Okano T. Soluble self-assembled block copolymers for drug delivery. *Pharm Res* 1999;5:597–600.

129. Ould-Ouali L, Ariën A, Rosenblatt J, Nathan A, Twaddle P, Matalenas T, Borgia M, Arnold S, Leroy D, Dinguizli M, Rouxhet L, Brewster M, Préat V. Biodegradable self-assembling PEG-copolymer as vehicle for poorly water-soluble drugs. *Pharm Res* 2004;21:1581–1590.

130. Müller RH. In *Colloidal Carriers for Control Drug Delivery and Targeting*. Stuttgart Germany: Wissenschaftliche Verlagsgesellschaft Boca Raton, FL: CRC Press; 1991.

131. Lasic DD, Martin F, Eds. *Stealth Liposomes*. Boca Raton, FL: CRC Press; 1995.

132. Cohen S, Bernstein H, Eds. *Microparticulate Systems for the Delivery of Proteins and Vaccines*. New York: Marcel Dekker; 1996.

133. Torchilin VP, Trubetskoy VS. Which polymers can make nanoparticulate drug carriers long-circulating? *Adv Drug Deliv Rev* 1995;16:141–155.

134. Torchilin VP. How do polymers prolong circulation time of liposomes? *J Liposome Res* 1996;6:99–116.

135. Takenaga M, Yamaguchi Y, Kitagawa A, Ogawa Y, Mizushima Y, Igarashi R. A novel insulin formulation can keep providing steady levels of insulin for much longer periods *in-vivo*. *J Pharm Pharmacol* 2002;54:1189–1194.

136. Kang F, Singh J. Preparation, *in vitro* release *in vivo* absorption and biocompatibility studies of insulin-loaded microspheres in rabbits. *AAPS PharmSciTech* 2005;6(3): E487–E494.

137. Prego C, García M, Torres D, Alonso MJ. Transmucosal macromolecular drug delivery. *J Control Release* 2005;101(1–3):151–162.

138. Alonso MJ. Nanoparticles as carriers for nasal vaccine delivery. *Expert Rev Vaccines* 2005;4(2):185–196.

139. Asthana A, Chauhan AS, Diwan PV, Jain NK. Poly(amidoamine) (PAMAM) dendritic nanostructures for controled site-specific delivery of acidic anti-inflammatory active ingredient. *AAPS PharmSciTech* 2005;6(3):E536–E542.

140. Beezer AE, King ASH, Martin IK, Mitchel JC, Twyman LJ, Wain CF. Dendrimers as potential drug carriers; encapsulation of acidic hydrophobes within water soluble PAMAM derivatives. *Tetrahedron* 2003;59:3873–3880.

141. Patri AK Jr,Majoros IJ Jr, Baker JR Jr. Dendritic polymer macromolecular carriers for drug delivery. *Curr Opin Chem Biol* 2002;6:466–471.
142. Bosman AW, Janssen HM, Meijer EW. About dendrimers: structure physical properties, and applications. *Chem Rev* 1999;99:1665–1688.
143. Milhem OM, Myles C, McKeown NB, Attwood D, Emanuele AD. Poly-amidoamine starburst dendrimers as solubility enhancers. *Int J Pharm* 2000;197:239–241.
144. Kolhe P, Misra E, Kannan RM, Kannan S, Lieh-Lai M. Drug complexation *in vitro* release and cellular entry of dendrimers and hyperbranched polymers. *Int J Pharm* 2003;259:143–160.
145. Zeng F, Zimmerman SC. Dendrimers in supramolecular chemistry: from molecular recognition to self-assembly. *Chem Rev* 1997;97:1681–1712.
146. Lasic DD, Papahajopoulos D, Eds. *Medical Application of Liposomes*. Amsterdam, The Netherlands: Elsevier; 1998.
147. Khaw BA, Torchilin VP, Vural I, Narula J. Plug and seal: prevention of hypoxic cardiocyte death by sealing membrane lesions with antimyosin-liposomes. *Nat Med* 1995;1:1195–1198.
148. Liang W, Levchenko TS, Khaw BA, Torchilin VP. ATP-containing immunoliposomes specific for cardiac myosin. *Curr Drug Deliv* 2004;1:1–7.
149. Verma DD, Hartner WC, Levchenko TS, Bernstein EA, Torchilin VP. ATP-loaded liposomes effectively protect the myocardium in rabbits with an acute experimental myocardial infarction. *Pharm Res* 2005;22(12):2115–2120.
150. Kimelberg HK, Mayhew EG. Properties and biological effects of liposomes and their uses in pharmacology and toxicology. *CRC Crit Rev Toxicol* 1978;6:25–79.
151. Szoka F, Papahadjopoulos D. Liposomes: preparation and characterization. In Knight CG, Ed. *Liposomes: From Physical Structure to Therapeutic Applications*. New York: Elsevier/North-Holland Biomedical Press; 1981, pp 51–82.
152. Poznansky MJ, Juliano RL. Biological approaches to the Control delivery of drugs: a critical review. *Pharm Rev* 1984;36:277–336.
153. Fanciullino R, Giacometti S, Aubert C, Fina F, Martin P-M, Piccerelle P, Ciccolini J. Development of stealth liposome formulation of 2′-deoxyinosine as 5-fluorouracil modulator: *in vitro* and *in vivo* study. *Pharm Res* 2005;22(12):2051–2057.
154. FDA. *Guidance for Industry: Nonclinical Studies for the Safety Evaluation of Pharmaceutical Excipients*. FDA, Center for Drug Evaluation and Research (CDER) Center for Biologics Evaluation and Research (CBER); May 2005.
155. Powell MF, Nguyen T, Baloian L. Compendium of excipients for parenteral formulations. *PDA J Pharm Sci Technol* 1998;52:238–311.
156. Nema S, Washkuhn RJ, Brendel RJ. Excipients and their use in injectable products. *PDA J Pharm Sci Technol* 1997;51:166–171.
157. *Inactive Ingredient Guide: Inactive Ingredients for Currently Marketed Drug Products*. Rockville, MD: FOI Services Inc; 1996.
158. Strickley RG. Parenteral formulations of small molecules therapeutics marketed in the United States (1999). Part III. *PDA J Pharm Sci Technol* 2000;54(2):152–169.
159. Rowe RC, Sheskey PJ, Owen SC, Eds. *Handbook of Pharmaceutical Excipients*, 5th ed. Washington DC: American Pharmaceutical Association; 2005.
160. Hoffman H. Polyoxyethyleneglycerol triricinoleate 35 DAC 1979. *Pharm Zeit* 1984;129: 730–733.
161. Goldspiel BR. Guidelines for administration. In McGuire WP, Rowinsky EK, Eds. *Paclitaxel in Cancer Treatment*. New York: Marcel Dekker; 1995, pp 175–186.

162. Windebank AJ, Blexrud MD, Groen PC. Potential neurotoxicity of the solvent vehicle for cyclosporine. *J Pharmacol Exp Ther* 1994;268:1051–1056.
163. Watkins J, Ward AM, Appleyard TN. Adverse reactions to intravenous anaesthetic induction agents. *Br Med J* 1997;2:1084–1085.
164. Szebeni J, Muggia FM, Alving CR. Complement activation by Cremophor EL as a possible contributor to hypersensitivity to paclitaxel: an *in vitro* study. *J Natl Cancer Inst* 1998;90:300–306.
165. Wandel C, Kim RB, Stein CM. "Inactive" excipients such as Cremophor can affect *in vivo* drug disposition. *Clin Pharmacol Ther* 2003;73:394–396.
166. van Zuylen L, Verweij J, Sparreboom A. Role of formulation vehicles in taxane pharmacology. *Invest New Drugs* 2001;19:125–141.
167. de Vos AI, Nooter K, Verweij J, Loos WJ, Brouwer E, De Bruijn P, Ruijgrok EJ, van der Burg ME, Stoter G, Sparreboom A. Differential modulation of cisplatin accumulation in leucocytes and tumor cell lines by the paclitaxel vehicle Cremophor EL. *Ann Oncol* 1997;8:1145–1150.
168. Ma J, Verweij J, Planting AST, Kolker HJ, Loos WJ, de Boer-Dennert M, van der Burg ME, Stoter G, Schellens JH. Docetaxel and paclitaxel inhibit DNA adduct formation and intracellular accumulation of cisplatin in human leukocytes. *Cancer Chemother Pharmacol* 1996;36:382.
169. Badary OA, Abdel-Naim AB, Khalifa AE, Hamada FMA. Differential alteration of cisplatin myelotoxicity by the paclitaxel vehicle Cremophor EL. *Naunyn-Schmiedeberg Arch Pharmacol* 2000;261:339–344.
170. Iwase K, Oyama Y, Tatsuishi T, Yamaguchi J, Nishimura Y, Kanada A, Kobayashi M, Maemura Y, Ishida S, Okano Y. Cremophor EL augments the cytotoxicity of hydrogen peroxide in lymphocytes dissociated from rat thymus glands. *Toxicol Lett* 2004;154: 143–148.
171. Yamaguchi J-Y, Nishimura Y, Kanada A, Kobayashi M, Mishima K, Tatsuishi T, Iwase K, Oyama Y. Cremophor EL, a non-ionic surfactant promotes $Ca^{2+}$-dependent process of cell death in rat thymocytes. *Toxicology* 2005;211(3):179–186.
172. Nah JW, Jang MK. Spectroscopic characterization and preparation of low molecular, water-soluble chitosan with free-amine group by novel method. *J Polym Sci Part A Polym Chem* 2002;40:3796–3803.
173. Thanou M, Verhoef JC, Junginge HE. Chitosan and its derivatives as intestinal absorption enhancers. *Adv Drug Deliv Rev* 2001;50(1):S91–S101.
174. Illum L. Chitosan and its use as a pharmaceutical excipient. *Pharm Res* 1998;15: 1326–1331.
175. Khor E, Lim LY. Implantable applications of chitin and chitosan. *Biomaterials* 2003;24:2339–2349.
176. Lee M, Nah JW, Kwon Y, Koh JJ, Ko KS, Kim SW. Water-soluble and low molecular weight chitosan-based plasmid DNA delivery. *Pharm Res* 2001;18:427–431.
177. Koping-Hoggard M, Tubulekas I, Guan H, Edwards K, Nilsson M, Varum KM. Chitosan as a nonviral gene delivery system. Structure–property relationships and characteristics compared with polyethylenimine *in vitro* and after lung administration *in vivo*. *Gene Ther* 2001;8:1108–1121.
178. Roy K, Mao HQ, Huang SK, Leong KW. Oral gene delivery with chitosan–DNA nanoparticles generates immunologic protection in a murine model of peanut allergy. *Nat Med* 1999;5:387–391.
179. Lueßen HL, de Leeuw BJ, Langemeÿer WE, de Boer AG, Verhoef JC, Junginger HE. Mucoadhesive polymers in peroral peptide drug delivery. VI. Carbomer and chitosan

improve the intestinal absorption of the peptide drug buserelin *in vivo*. *Pharm Res* 1996;13:1668–1672.

180. Lina BAR, Bär A. Subchronic oral toxicity studies with α-cyclodextrin in rats. *Regul Toxicol Pharmacol* 2004;39(1):14–26.
181. Schmid G. Cyclodextrin glycosyltransferase production: yield enhancement by overexpression of cloned genes. *TIBTECH* 1989;7:244–248.
182. Allegre M, Deratani A. Cyclodextrin uses: from concept to industrial reality. *Agro Food Industry Hi-Tech* 1994;5:9–17.
183. Le Bas G, Rysanek N. Structural aspects of cyclodextrins. In Duchêne D, Ed. *Cyclodextrins and Their Industrial Uses*. Paris: Editions de Santé; 1987, pp 105–130.
184. Schmid G. Preparation and application of γ-cyclodextrin. In Duchêne D, Ed. *New Trends in Cyclodextrins and Derivatives*. Paris: Editions Santé; 1991, pp 27–54.
185. Englyst HN, Kingman SM, Cummings JH. Classification and measurement of nutritionally important starch fractions. *Eur J Clin Nutr* 1992;46 Suppl (2):S33.
186. Thompson DO. Cyclodextrins—enabling excipients: their present and future use in pharmaceuticals. *Crit Rev Ther Drug Carrier Syst* 1997;14 (1):1–104.
187. Uekama K, Hirayama F, Irie T. Cyclodextrin drug carrier systems. *Chem Rev* 1998;98: 2045–2076.
188. Oda M, Saitoh H, Kobayashi M, Aungst BJ. β-Cyclodextrin as a suitable solubilizing agent for *in situ* absorption study of poorly water-soluble drugs. *Int J Pharm* 2004; 280(1–2):95–102.
189. Ohtani Y, Irie T, Uekama K, Fukunaga K, Pitha J. Differential effects of α-, β- and γ-cyclodextrins on human erythrocytes. *Eur J Biochem* 1989;186:17–22.
190. Debouzy JC, Fauvelle F, Crouzy S, Girault L, Chapron Y, Goschl M, Gadelle A. Mechanism of α-cyclodextrin induced hemolysis. 2. A study of the factors controlling the association with serine-, ethanolamine-, and choline-phospholipids. *J Pharm Sci* 1998;87: 59–66.
191. Fauvelle F, Debouzy JC, Crouzy S, Goschl M, Chapron Y. Mechanism of α-cyclodextrin-induced hemolysis. 1. The two-step extraction of phosphatidylinositol from the membrane. *J Pharm Sci* 1997;86:935–943.
192. Galbiati F, Razani B, Lisanti MP. Emerging themes in lipid rafts and caveolae. *Cell* 2001;106:403–411.
193. Simons K, Ehehalt R. Cholesterol, lipid rafts, and disease. *J Clin Invest* 2002;110: 597–603.
194. Anderson RG, Jacobson K. A role for lipid shells in targeting proteins to caveolae, rafts, and other lipid domains. *Science* 2002;296:1821–1825.
195. Irie T, Otagiri M, Sunada M, Uekama K, Ohtani Y, Yamada Y, Sugiyama Y. Cyclodextrin-induced hemolysis and shape changes of human erythrocytes *in vitro*. *J Pharmacobiodyn* 1982;5:741–744.
196. Irie T, Uekama K. Pharmaceutical applications of cyclodextrins. III. Toxicological issues and safety evaluation. *J Pharm Sci* 1997;86:147–162.
197. Uekama K, Otagiri M. Cyclodextrins in drug carrier systems. *Crit Rev Ther Drug Carrier Syst* 1987;3:1–40.
198. Arima H, Nishimoto Y, Motoyama K, Hirayama F, Uekama K. Inhibitory effects of novel hydrophilic cyclodextrin derivatives on nitric oxide production in macrophages stimulated with lipopolysaccharide. *Pharm Res* 2001;18:1167–1173.
199. Motoyama K, Arima H, Nishimoto Y, Miyake K, Hirayama F, Uekama K. Involvement of CD14 in the inhibitory effects of dimethyl-α-cyclodextrin on lipopolysaccharide signaling in macrophages. *FEBS Lett* 2005;579:1707–1714.

200. Motoyama K, Arima H, Toyodome H, Irie T, Hirayama F, Uekama K. Effect of 2,6-di-*O*-methyl-α-cyclodextrin on hemolysis and morphological change in rabbit's red blood cells. *Eur J Pharm Sci* 2006;29(2):111–119.
201. Pope DC, Baggot JD. The basis of selection of the dosage form. In Blodinger I, Ed. *Formulation of Veterinary Dosage Forms*. New York: Marcel Dekker; 1983, pp 1–70.
202. Bates TR, Sequeira JA. Bioavailability of micronized griseofulvin from corn oil-in-water emulsion, aqueous suspension and commercial tablet dosage forms in humans. *J Pharm Sci* 1975;64(5):793–797.
203. Jones BE. *Manuf Chem Aerosol News* 1969;40(2):25.
204. Lazor T. *Drug Cosmet Ind* 1974;114(3):42.
205. Pope DG. In Monkhouse, DC, Ed. *Animal Health Products, Design and Evaluation*. Washington DC: American Pharmaceutical Association; p 78.
206. Pope DG. *Aust Vet Pract* 1980;10:57.
207. Hem SL, Bright DR, Banker GS, Page JP. *Drug Dev Commun* 1974–75;1:471.
208. Spiegel AJ, Noseworthy MM. Use of nonaqueous solvents in parenteral products. *J Pharm Sci* 1963;52:917–927.
209. Rasmussen F. In Van Miert ASJPAM, Frens J, Van der Kreek FW, Eds. *Trends in Veterinary Pharmacology and Toxicology*. Amsterdam: Elsevier; 1980, p 27.
210. Rasmussen F, Svendsen O. Tissue damage and concentration at the injection site after intramuscular injection of chemotherapeutics and vehicles in pigs. *Res Vet Sci* 1976;20: 55–60.
211. Blom L, Rasmussen F. Tissue damage at the infection site after intramuscular injection of drugs in hens. *Br Poult Sci* 1976;17:1–4.
212. Svendsen O, Rasmussen F, Nielson P, Steiness E. The loss of creatine phosphokinase (CK) from intramuscular injection sites in rabbits. A predictive tool for local toxicity. *Acta Pharmacol Toxicol (Copenhagen)* 1979;44:324–328.
213. Andersson L, Blomberg L, Flegel M, Lepsa L, Nilsson B, Verlander M. Large-scale synthesis of peptides. *Biopolymers* 2000;55:227–250.
214. Pontiroli AE. Peptide hormones: review of current and emerging uses by nasal delivery. *Adv Drug Deliv Rev* 1998;29:81–87.
215. Chetty DJ, Chien YW. Novel methods of insulin delivery: an update. *Crit Rev Ther Drug Carrier Syst* 1998;15:629–670.
216. Simon M, Kissel T. Away with the needle. Noninvasive administration routes for insulin: improved quality of life for diabetics. *Pharm Unserer Zeit* 2001;30:136–141.
217. Gizurarson S. The relevance of nasal physiology to the design of drug absorption studies. *Adv Drug Deliv Rev* 1993;11(3):329–347.
218. Gizurarson S, Bechgaard E. Study of nasal enzyme activity towards insulin *in vitro*. *Chem Pharm Bull* 1991;39:2155–2157.
219. Deurloo MJ, Hermens WA, Romeyn SG, Verhoef JC, Merkus FW. Absorption enhancement of intranasally administered insulin by sodium taurodihydrofusidate (STDHF) in rabbits and rats. *Pharm Res* 1989;6:853–856.
220. Hinchcliffe M, Illum L. Intranasal insulin delivery and therapy. *Adv Drug Deliv Rev* 1999;35:199–234.
221. Sanchez A, Ygartua P, Fos D. Role of surfactants in the bioavailability of intranasal insulin. *Eur J Drug Metab Pharmacokinet* 1991;3:120–124.
222. Chandler SG, Illum L, Thomas NW. Nasal absorption in the rats: II. Effect of enhancers on insulin absorption and nasal histology. *Int J Pharm* 1991;76:61–70.

223. Prego C, Garcia M, Torres D, Alonso MJ. Transmucosal macromolecular drug delivery. *J Control Release* 2005;101:151–162.

224. Koping-Hoggard M, Sanchez A, Alonso MJ. Nanoparticles as carriers for nasal vaccine delivery. *Expert Rev Vaccines* 2005;4:185–196.

225. Fernandez-Urrusuno R, Calvo P, Remunan-Lopez C, Vila-Jato JL, Alonso MJ. Enhancement of nasal absorption of insulin using chitosan nanoparticles. *Pharm Res* 1999;16: 1576–1581.

226. Brooking J, Davis SS, Iluum L. Transport of nanoparticles across the rat nasal mucosa. *J Drug Target* 2001;9:267–279.

227. Vila A, Sanchez A, Evora C, Soriano I, McCallion O, Alonso MJ. PLA–PEG particles as nasal protein carriers: the influence of the particle size. *Int J Pharm* 2005;292:43–52.

228. Takeuchi H, Yamamoto H, Kawashima Y. Mucoadhesive nanoparticulate systems for peptide drug delivery. *Adv Drug Deliv Rev* 2001;47:39–54.

229. Dondeti P, Zia H, Needham T. Bioadhesive and formulation parameters affecting nasal absorption. *Int J Pharm* 1996;127:115–133.

230. Jimenez-Castellanos MR, Zia H, Rhodes CT. Mucoadhesive drug delivery systems. *Drug Dev Ind Pharm* 1993;19:143–194.

231. Granger DN, Kvielys PR, Perry MA, Taylor AE. Charge selectivity of rat intestinal capillaries—influence of polycations. *Gastroenterology* 1986;91:1443–1446.

232. Maitani Y, Machida Y, Nagai T. Influence of molecular weight and charge on nasal absorption of dextran and DEAE-dextran in rabbits. *Int J Pharm* 1992;76:43–49.

233. Illum L, Farraj NF, Davis SS. Chitosan as novel nasal delivery system for peptide drugs. *Pharm Res* 1994;11:1186–1189.

234. Simon M, Wittmar M, Kissel T, Linn T. Insulin containing nanocomplexes formed by self-assembly from biodegradable amine-modified poly(vinyl alcohol)-graft-poly(L-lactide): bioavailability and nasal tolerability in rats. *Pharm Res* 2005;22(11): 1879–1886.

235. Dayal P, Shaik MS, Singh M. Evaluation of different parameters that affect droplet-size distribution from nasal sprays using the Malvern Spraytec. *J Pharm Sci* 2004;93: 1725–1742.

236. Malmsten M, Ed. *Surfactants and Polymers in Drug Delivery*. Lancaster, PA: Marcel Dekker; 2002.

237. D'Souza R, Mutalik S, Venkatesh M, Vidyasagar S, Udupa N. Insulin gel as an alternate to parenteral insulin: formulation preclinical and clinical studies. *AAPS Pharm Sci Tech* 2005;6(2):E184–189.

238. Alur HH, Johnston TP, Mitra AK. Peptides and proteins: buccal absorption. In Superbrick J, Boylan JC, Eds. *Encyclopedia of Pharmaceutical Technology*. New York: Marcel Dekker; 2001, pp 193–218.

239. Sudhakar Y, Kuotsu K, Bandyopadhyay AK. Buccal bioadhesive drug delivery—a promising option for orally less efficient drugs. *J Control Release* 2006;114(1):15–40.

240. Morishita M, Barichello JM, Takayama K, Chiba Y, Tokiwa S, Nagai T. Pluronic F-127 gels incorporating highly purified unsaturated fatty acids for buccal delivery of insulin. *Int J Pharm* 2001;212(2):289–293.

241. Aungst BJ, Rogers NJ. Comparison of the effects of various transmucosal absorption promoters on buccal insulin delivery. *Int J Pharm* 1989;53:227–235.

242. Siegel IA, Gordon HP. Surfactant—induced increase of permeability Of rat oral mucosa to non-electrolytes *in vivo*. *Arch Oral Biol* 1985;30:43–47.

243. Squier CA, Wertz PW. In Rathbone MJ. Ed. *Structure and Function of the Oral Mucosa and Implications for Drug Delivery Oral Mucosal Drug Delivery*. New York: Marcel Dekker 1996, pp 1–26.
244. Ishida M, Machida Y, Nambu N, Nagai T. New mucosal dosage forms of insulin. *Chem Pharm Bull* 1981;84:810–816.
245. Barsuchn CL, Olanoff LS, Gleason DD, Olanoff EL, Gleason DD, Adkins EL, Ho NFH. Human buccal absorption of flubiprofen. *Clin Pharmacol Ther* 1988;44:225–231.
246. Mehta M, Kemppainen BW, Stafford RG. *In vitro* penetration of tritium labeled water (THO) and [$^3$H] Pb Tx-3 (a red tide toxin) through monkey buccal mucosa and skin. *Toxicol Lett* 1991;55:185–194.
247. Squier CA, Cox P, Wertz PW. Lipid content and water permeability of skin and oral mucosa. *J Invest Dermatol* 1991;96:123–126.
248. Puglia C, Filosa R, Peduto A, de Caprariis P, Rizza L, Bonina F, Blasi P. Evaluation of alternative strategies to optimize ketorolac transdermal delivery. *AAPS Pharm Sci Tech* 2006;7(3):Article 64.
249. Parikh DK, Ghosh TK. Feasibility of transdermal delivery of fluoxetine. *AAPS PharmSciTech* 2005;6:E144–E149.
250. Satturwar PM, Fulzele SV, Dorle AK. Evaluation of polymerized rosin for the formulation and development of transdermal drug delivery system: a technical note. *AAPS PharmSciTech* 2005;6(4):E649–E654.
251. Chien YW. Comparative Control skin permeation of nitroglycerin from marketed transdermal delivery systems. *J Pharm Sci* 1983;72:968–970.
252. Loftsson T, Gildersleeve N, Border N. The effect of vehicle additives on the transdermal delivery of nitroglycerin. *Pharm Res* 1987;4:436–444.
253. Corbo M, Liu JC, Chien YW. Bioavailability of propranolol following oral & transdermal administration in rabbits. *J Pharm Sci* 1990;79:584–587.
254. Aqil M, Sultana Y, Ali A. Matrix type transdermal drug delivery systems of metoprolol tartrate: *in vitro* characterization. *Acta Pharm* 2003;53:119–125.
255. Rao PR, Ramakrishna S, Diwan PV. Drug release kinetics from polymeric films containing propranolol hydrochloride for transdermal use. *Pharm Dev Technol* 2000;5: 465–472.
256. Williams AC, Barry BW. Penetration enhancers. *Adv Drug Deliv Rev* 2004;56:603–618.
257. Bohme K. Buprenorphine in a transdermal therapeutic system—a new option. *Clin Rheumatol* 2002;21:S13–S16.
258. Foldvari M. Non-invasive administration of drugs through the skin: challenges in delivery system design. *Pharm Sci Technol Today* 2000;3:417–425.
259. Barry BW. Novel mechanisms and devices to enable successful transdermal drug delivery. *Eur J Pharm Sci* 2001;14:101–114.
260. Doh HJ, Cho WJ, Yong CS, Choi HG, Kim JS, Lee CH, Kim DD. Synthesis and evaluation of ketorolac ester prodrugs for transdermal delivery. *J Pharm Sci* 2003;92:1008–1017.
261. Alsarra IA, Bosela AA, Ahmed SM, Mahrous GM. Proniosomes as a drug carrier for transdermal delivery of ketorolac. *Eur J Pharm Biopharm* 2005;59:485–490.
262. Cho YA, Gwak HS. Transdermal delivery of ketorolac tromethamine: effects of vehicles and penetration enhancers. *Drug Dev Ind Pharm* 2004;30:557–564.
263. Tiwari SB, Udupa N. Investigation into the potential of iontophoresis facilitated delivery of ketorolac. *Int J Pharm* 2003;260:93–103.
264. Müller RH, Mäder K, Gohla S. Solid lipid nanoparticles for Controled drug delivery: a review of the state of the art. *Eur J Pharm Biopharm* 2000;50:161–177.

265. Johnson ME, Blankschtein D, Linger R. Evaluation of solute permeation through the stratum corneum: lateral bilayer diffusion as the primary transport mechanism. *J Pharm Sci* 1997;86:1162–1172.
266. Burger GT, Cheryl ML. Animal care and facilities. In Hayes AW, Ed. *Principles and Methods of Toxicology*, 2nd ed, New York: Raven Press Ltd; 1989, pp 521–552.
267. Rao PR, Diwan PV. Formulation and *in vitro* evaluation of polymeric films of diltiazem hydrochloride and indomethacin for transdermal administration. *Drug Dev Ind Pharm* 1998;24:327–336.
268. Marjukka ST, Bouwstra JA, Urtti A. Chemical enhancement of percutaneous absorption in relation to stratum corneum structural alterations. *J Control Release* 1999;59:149–161.
269. Guy RH, Kalia YN, Delgado-Charro MB, Merino V, Lopez A, Marro D. Iontophoresis: electrorepulsion and electroosmosis. *J Control Release* 2000;64:129–132.
270. Conjeevaram R, Banga AK, Zhang L. Electrically modulated transdermal delivery of fentanyl. *Pharm Res* 2002;19:440–444.
271. Behl C, Char HS, Patel SB, Mehta DB, Piemontese D, Malick AW. *In vivo* and *in vitro* skin uptake and permeation studies; critical considerations and factors, which affect them. In Shah BP, Maibach HI, Eds. *Topical Drug Bioavailability, Bioequivalence and Penetration*. New York: Plenum Press; 1993, pp 225–259.
272. Green PG, Flanagan M, Shroot B, Guy RH. Iontophoretic drug delivery. In Walters KA, Hadgraft J, Eds. *Pharmaceutical Skin Penetration Enhancement*. New York: Marcel Dekker; 1993, pp 311–333.
273. Malone T, Haleblian JK, Poulsen BJ, Burdick KH. Development and evaluation of ointment and cream vehicles for a new topical steroid fluclorolone acetonide. *Br J Dermatol* 1974;90:187–195.
274. Poulsen BJ, Young E, Coquilla V, Katz M. Effect of topical vehicle composition on the *in vitro* release of fluocinolone acetonide and its acetate ester. *J Pharm Sci* 1968;57: 928–933.
275. Oishi J, Ushio Y, Narahara M, Takehara M, Nakagawa T. Effect of vehicles on percutaneous absorption. I. Characterization of oily vehicles by percutaneous absorption and trans-epidermal water loss test. *Chem Pharm Bull (Tokyo)* 1976;24:1765–1773.
276. Ostrenga J, Steinmetz C, Poulsen B. Significance of vehicle composition. I. Relationship between topical vehicle composition, skin penetrability and clinical efficacy. *J Pharm Sci* 1971;60:1175–1179.
277. Ostrenga J, Steinmetz C, Poulsen B, Yett S. Significance of vehicle composition. II Prediction of optimal vehicle composition. *J Pharm Sci* 1971;60:1180–1183.
278. Coldman MF, Poulsen BJ, Higuchi T. Enhancement of percutaneous absorption by the use of volatile: nonvolatile systems as vehicles. *J Pharm Sci* 1969;58:1098–1102.
279. Stoughton RB, Fritsch W. Influence of dimethylsulfoxide (DMSO) on human percutaneous absorption. *Arch Dermatol* 1964;90:512–517.
280. Maibach HT, Feldman RJ. The effect of DMSO on percutaneous penetration of hydrocortisone and testosterone in man. *Ann NY Acad Sci* 1967;141:423–427.
281. Pitman IH, Rostas SJ. Topical drug delivery to cattle and sheep. *J Pharm Sci* 1981;70: 1181–1194.
282. Higuchi T. *J Soc Cosmet Chem* 1960;11:85.
283. Gordon RD, Peterson TA. Myths about transdermal drug delivery. *Drug Deliv Technol* 2003;3:1–4.
284. Tian J, Yin CH. *In vitro* and *in vivo* assessment of liposomes containing recombinant human interleukin-2. *Proc Int Symp Control Release Bioact Mate* 1998;25:439–440.

285. Hora MS, Rana RK, Nunberg JH, Tice TR, Gilley RM, Hudson ME. Control release of interleukin-2 from biodegradable microspheres. *Biotechnology (NY)* 1990;8:755–758.

286. Johnston TP, Punjabi MA, Froelich CJ. Sustained delivery of interleukin from poloxamer 407 gel matrix following intraperitoneal injection in mice. *Pharm Res* 1992;9:425–434.

287. Hennink W, Franssen EO, van Dijk-Wolthuis WNE, Talsma H. Dextran hydrogels for the Control release of proteins. *J Control Release* 1997;48:107–114.

288. Stratton LP, Dong A, Manning M, Carpenter JF. Drug delivery matrix containing native protein precipitates suspended in a poloxamer gel. *J Pharm Sci* 1997;86:1006–1010.

289. Wenzel JG, Balaji KS, Koushik K, Navarre C, Duran SH, Rahe CH, Kompella UB. Pluronic F127 gel formulations of deslorelin and GnRH reduce drug degradation and sustain drug release and effect in cattle. *J Control Release* 2002;85:51–59.

290. Wang P, Johnston TP. Enhanced stability of two model proteins in an agitated solution environment using poloxamer 407. *J Parenter Sci Technol* 1993;47:183–189.

291. Kikwai L, Babu RJ, Prado R, Kolot A, Armstrong CA, Ansel JC, Singh M. *In vitro* and *in vivo* evaluation of topical formulations of spantide II. *AAPS PharmSciTech* 2005;6(4):E565–E572.

292. Wolfson AD, McCafferty DF, Moss GP. Development and characterization of a moisture-activated bioadhesive drug delivery system for percutaneous local anaesthesia. *Int J Pharm* 1998;169(1):83–94.

293. Campbell CB. Ophthalmic agents: ointments or drops? *Vet Med Small Animal Clin* 1979;74:971–974.

294. McCarron PA, Donnelly RF, Zawislak A, Woolfson AD. Design and evaluation of a water-soluble bioadhesive patch formulation for cutaneous delivery of 5-aminolevulinic acid to superficial neoplastic lesions. *Eur J Pharm Sci* 2006;27(2–3):268–279.

295. Cooper CA, Bergamini MVW, Leopold IH. Use of flurbiprofen to inhibit corneal neovascularization. *Arch Ophthalmol* 1980;98:1102–1105.

296. Kraff MC, Sanders DR, Mcguigan L, Rannan MG. Inhibition of blood–aqueous humor barrier breakdown with diclofenac: a fluorometric study. *Arch Ophthalmol* 1990;108: 380–383.

297. Searle AE, Pearce JL, Shaw DE. Topical use of indomethacin on the day of cataract surgery. *Br J Ophthalmol* 1990;74:19.

298. Mahoney JM, Waterbury LD. (±) 5 Benzoyl-1,2-dihydro-3H pyrrolo [1.2a] pyrrole-1-carboxylic acid (RS 37619): a non irritating ophthalmic anti-inflammatory agent. *Invest Ophthalmol Vis Sci* 1983;24:151–159.

299. Flach AJ, Kraff MC, Sanders DR, Tanenbaum L. The quantitative effect of 0.5% ketorolac tromethamine solution and 0.1% dexamethasone sodium phosphate solution on post surgical blood aqueous barrier. *Arch Ophthalmol* 1988;106:480–483.

300. Heier J, Cheetham JK, Degryse R. Ketorolac tromethamine 0.5% ophthalmic solution in the treatment of moderate to severe ocular inflammation after cataract surgery: a randomized vehicle–control clinical trial. *Am J Ophthalmol* 1999;127:253–259.

301. Solomen KD, Cheetham JK, Degryse R, Brint SF, Rosenthal A. Topical ketorolac tromethamine 0.5% ophthalmic solution in ocular inflammation after cataract surgery. *Ophthalmology* 2001;108:331–337.

302. Malhotra M, Majumdar DK. *In vivo* ocular availability of ketorolac following ocular instillations of aqueous, oil, and ointment formulations to normal corneas of rabbits: a technical note. *AAPS PharmSciTech* 2005;6(3):E523–E526.

303. Malhotra M, Majumdar DK. *In vitro* transcorneal permeation of ketorolac from oil based ocular drops and ophthalmic ointment. *Indian J Exp Biol* 1997;35:1324–1330.

304. Eljarrat-Binstock E, Domb AJ. Iontophoresis: a non-invasive ocular drug delivery. *J Control Release* 2006;110(3):479–489.

305. Behar-Cohen FF, El Aouni A, Gautier S, Davis GJ, Chapon P, Parel JM. Transscleral Coulomb-control iontophoresis of methylprednisolone into the rabbit eye: influence of duration of treatment current intensity and drug concentration on ocular tissue and fluid levels. *Exp Eye Res* 2002;74(1):51–59.

306. Hayden BC, Jockovich M-E, Murray TG, Voigt M, Milne P, Kralinger M, Feuer WJ, Hernandez E, Parel JM. Pharmacokinetics of systemic versus focal carboplatin chemotherapy in the rabbit eye: possible implication in the treatment of retinoblastoma. *Invest Ophthalmol Vis Sci* 2004;45:3644–3649.

307. Inamura N, Salt AN. Permeability changes of the blood–labyrinth barrier measured *in vivo* during experimental treatments. *Hear Res* 1992;61(1–2):12–18.

308. Juhn SK, Rybak LP. Labyrinthine barriers and cochlear homeostasis. *Acta Oto-Laryngol* 1981;91(5–6):529–534.

309. Tonndorf J, Duvall AJ, Reneau JP. Permeability of intracochlear membranes to various vital stains. *Ann Otol Rhinol Laryngol* 1962;71:801–841.

310. Chen Z, Kujawa SG, McKenna MJ, Fiering JO, Mescher MJ, Borenstein JT, Leary Swan EE, Sewell WF. Inner ear drug delivery via a reciprocating perfusion system in the guinea pig. *J Control Release* 2005;110(1):1–19.

311. Shahiwala A, Misra A. A preliminary pharmacokinetic study of liposomal leuprolide dry powder inhaler: a technical note. *AAPS PharmSciTech* 2005;6(3):E482–E486.

312. Jiang G, Qiu W, Deluca PP. Preparation and *in vitro/in vivo* evaluation of insulin-loaded poly(acryloyl-hydroxyethyl starch)-PLGA composite microspheres. *Pharm Res* 2003;20: 452–459.

313. Hildebrand GE, Tack JW. Microencapsulation of peptides and proteins. *Int J Pharm* 2000;196:173–176.

314. Haeckel S, Sachse S, Albayrak C, Müller RH. Preparation and characterization of insulin-loaded microcapsules produced by the induced phase separation method. In *Proceedings of the World Meeting, Association of Industrial Galenique Pharmacy/Association of Primate Veterinarians*; 25–28, May 1998, Paris France, pp 507–508.

315. Wall DA. Pulmonary absorption of peptides and proteins. *Drug Deliv* 1995;2:1–20.

316. Stanley SD. Delivery of peptide and non-peptide drugs through the respiratory tract. *Pharm Sci Tech Today* 1999;12:450–459.

317. Niven RW. Delivery of biotherapeutics by inhalation aerosol. *Crit Rev Ther Drug Carrier Syst* 1995;12:151–231.

318. Patton JS, Trinchero P, Platz RM. Bioavailability of pulmonary delivered peptide and protein: interferon alpha calcitonin and parathyroid hormones. *J Control Release* 1994;28:79–85.

319. Edwards DA, Hanes J, Caponetti G. Large porous particles for pulmonary drug delivery. *Science* 1997;276:1868–1871.

320. Berelowitz M, Becker G. Inhaled insulin: clinical pharmacology and clinical study results. *Respir Drug Deliv* 2000;7:151–154.

321. Clauson PG, Balent B, Okikawa J. PK-PD of four different doses of pulmonary insulin delivered with AERx Diabetes management system. *Respir Drug Deliv* 2000;7:155–161.

322. Adjel A, Hui J, Finley R, Lin T, Fort F. Pulmonary bioavailability of leuprolide acetate following multiple dosing to beagle dogs: some pharmacokinetic and pre-clinical issues. *Int J Pharm* 1994;107:57–66.

323. Garber RA, Cappelleri JC, Kourides IA, Gelgend RA, Chandler LP, Gorkin L. Improved patient satisfaction with inhaled insulin in subjects with Type I diabetes mellitus: results from a multi-center randomized Control trial. *Diabetes* 1999;48:A13.
324. Ganderton D. Targeted delivery of inhaled drugs: current challenges and future trends. *J Aerosol Med* 1999;12:3–8.
325. Saari M, Vidgren MT, Koskinen MO, Turjanmaa VMH, Nieminen MM. Pulmonary distribution and clearance of two beclomethasone liposome formulations in healthy volunteers. *Int J Pharm* 1999;181:1–9.
326. Waldrep JC, Gilbert BE, Knight CM. Pulmonary delivery of beclomethasone liposome aerosol in volunteers Tolerance and safety. *Chest* 1997;111:316–323.
327. Green J, Gommeren E, Creazzo J, Pharmaceutical K. Pharmaceutical aerosols: enhancing the metered dose inhaler. *Drug Delivery Technol* 2004;4(6):48–53.

# CYTOCHROME P450 ENZYMES

Eugene G. Hrycay and Stelvio M. Bandiera
*University of British Columbia, Vancouver, British Columbia, Canada*

**Contents**

1 Introduction
2 Identification of CYP Enzymes
  2.1 Characterization of CYP Enzymes
  2.2 Detection and Isolation
  2.3 Occurrence and Distribution
  2.4 Nomenclature of CYP Enzymes
  2.5 Genetic Polymorphism
3 CYP Enzyme Reactions
  3.1 CYP Substrates
  3.2 Types of Reactions Catalyzed by CYP Enzymes
  3.3 CYP Enzyme Marker Activities
  3.4 CYP Reaction Cycle
  3.5 CYP Inhibitors
4 Structure and Regulation
  4.1 Overview of Individual CYP Enzymes
  4.2 CYP1 Family
  4.3 CYP2 Family
  4.4 CYP3 Family
  4.5 CYP4 Family
  4.6 Regulation of CYP Enzyme Expression
5 Conclusion
  Acknowledgments
  References

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

## 1 INTRODUCTION

The cytochrome P450 (CYP) enzymes, also known as the microsomal mixed function oxidase system, are the predominant biotransformation pathway in the body for lipid-soluble endogenous and xenobiotic compounds. CYP enzymes play a key role in the biotransformation of pharmaceutical agents and thus are major determinants of duration of action and clearance of drugs. CYP enzymes are also involved in complications arising from drug therapy, such as drug interactions, which may be caused by induction or inhibition of CYP enzymes. Moreover, CYP enzymes may contribute to disease states. For example, CYP enzymes are implicated in the etiology of some cancers because chemical procarcinogens require enzymatic conversion to carcinogens, a process that is mediated by CYP enzymes. This chapter presents aspects of the characterization, function, and regulation of CYP enzymes, in mammals generally and in humans specifically, that are relevant to our understanding of how CYP enzymes can metabolize a large variety of drugs.

CYP enzymes are membrane-bound proteins localized predominantly in the smooth endoplasmic reticulum of liver and other tissues and are quantitatively and qualitatively among the most important enzymes of the phase I pathway for xenobiotic biotransformation [1, 2]. CYP enzymes are responsible for oxidizing hydrophobic compounds to more polar metabolites for further biotransformation or subsequent excretion. CYP enzymes function as terminal oxidases of an electron transport chain in which one atom of oxygen from molecular oxygen is incorporated into the substrate and the other atom is reduced to water. The multicomponent electron transport chain also includes a flavoprotein, NADPH-dependent cytochrome P450 reductase, which serves to transfer reducing equivalents (electrons) supplied by NADPH (and sometimes NADH) to the CYP enzymes. CYP, as the terminal oxidase of the chain, is the enzyme active site, the oxygen- and substrate-binding site, and the determinant of substrate specificity (Fig. 1).

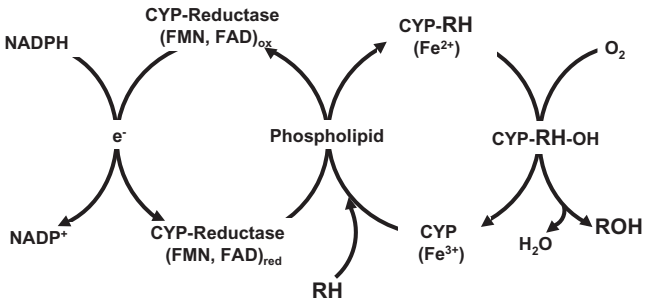


**FIGURE 1** Electron transfer and metabolism of substrates by the microsomal CYP system. Electrons, transferred through a series of oxidation–reduction reactions from NADPH to CYP by NADPH-dependent CYP reductase, are used to reduce molecular oxygen with the concomitant oxidation of the substrate. FMN, flavin mononucleotide; FAD, flavin adenine dinucleotide; RH, substrate; ROH, oxidized product.

## 2 IDENTIFICATION OF CYP ENZYMES

### 2.1 Characterization of CYP Enzymes

CYP enzymes contain a heme prosthetic group, which gives the enzymes a reddish brown color. The heme group is iron protoporphyrin IX and is the same heme moiety found in hemoglobin. As is the case for hemoglobin, the iron protoporphyrin IX group in CYP enzymes has a strong affinity for molecular oxygen but binds carbon monoxide with higher affinity. The iron atom of the heme prosthetic group can exist in two oxidation states, $Fe^{2+}$ (reduced) and $Fe^{3+}$ (oxidized). The heme moiety bound to CYP enzymes is essential for activity, and loss of the heme group or "poisoning" by carbon monoxide is accompanied by loss of enzymatic activity.

In all CYP enzymes, the heme prosthetic group (iron protoporphyrin IX) is bound to a single polypeptide chain composed of approximately 500 amino acids. Considerable evidence indicates that the amino terminus of the apoprotein anchors CYP in the lipid membrane bilayer, with most of the enzyme including the active site exposed on the cytoplasmic side of the endoplasmic reticulum [3, 4]. The central iron atom of the heme group is bound noncovalently to the sulfur atom of a cysteine residue situated close to the C terminus of the apoprotein chain.

Interactions between the heme group and apoprotein produce unique spectral and magnetic properties. Normally, the iron in the heme prosthetic group of CYP is in the oxidized state, which exhibits low affinity for carbon monoxide and absorbs visible light with a spectral maximum at a wavelength of approximately 418 nm. However, if the iron is in the reduced state, it can then bind carbon monoxide with high affinity, and the resulting reduced carbon monoxide-bound CYP complex exhibits a large spectral maximum at 450 nm. This characteristic spectral property is a signature of CYP enzymes from which the name, cytochrome P450, was derived. Other heme proteins, such as hemoglobin, do not produce an absorbance at 450 nm when complexed with carbon monoxide.

Another essential component for CYP activity is NADPH-dependent CYP reductase, which is also located primarily on the surface of the phospholipid bilayer of the smooth endoplasmic reticulum in close proximity to the CYP enzymes. NADPH-dependent CYP reductase forms a temporary complex with CYP, permitting the flow of electrons to the heme prosthetic group and the resulting oxidation of the substrate. NADPH-dependent CYP reductase has two domains, one containing flavin adenine dinucleotide and the second containing flavin mononucleotide. Two electrons are acquired from NADPH and migrate from FAD to FMN, then to the heme group of CYP. In some cases, CYP enzymes can also accept electrons from cytochrome $b_5$, a small membrane-bound heme protein that receives electrons from NADH.

### 2.2 Detection and Isolation

Studies by Axelrod [5] and by Brodie and co-workers [6] in the mid-1950s revealed the existence of an enzyme system in liver that converted drugs and aromatic compounds into more polar metabolites and that required NADPH for the reaction. This enzyme system was subsequently identified as a carbon monoxide-binding pigment in liver microsomes by Klingenberg [7] and Garfinkel [8].

Microsomes is the name given to the endoplasmic reticulum membrane fraction, which can be isolated by breaking open and grinding cells to prepare a whole cell homogenate. If the homogenate is subjected to centrifugation at approximately 9000 *g*, cell debris, nuclei, and mitochondria spin down in the pellet. The resulting supernatant suspension, called the S9 fraction, contains both cytosolic and microsomal proteins. When the supernatant is spun at higher speed (e.g., 100,000 *g*), the resulting supernatant fraction contains the cytosolic fraction and the pellet contains the microsomal fraction. The pellet can be resuspended with 0.25 M sucrose or with 0.1 M potassium phosphate buffer containing 10% glycerol to give the microsomes, which can be stored at –80 °C for substantial periods of time. As long as the microsomal fraction is sufficiently concentrated (i.e., has a protein concentration >10 mg/mL), the CYP activity of the frozen microsomal fraction can be maintained for up to five years. The total CYP concentration of tissue preparations such as the S9 or microsomal fractions can be determined in a scanning spectrophotometer by measuring the height of the reduced carbon monoxide-bound CYP complex at 450 nm as reported by Omura and Sato [9].

The identity of CYP enzymes involved in the metabolism of a specific drug can be determined using a combination of approaches. Well characterized, specific inhibitory antibodies (either polyclonal or monoclonal) against various CYP enzymes are a valuable tool for identification of CYP enzymes in S9 and microsomal fractions. Chemical inhibitors (discussed later), especially highly selective inhibitors, can also be used. Finally, purified or recombinant enzymes can be used to determine the intrinsic ability of an individual or panel of CYP enzymes to catalyze a particular reaction or metabolize a specific drug. Once the CYP enzymes involved in the metabolism of a specific drug have been identified, it may be necessary to measure levels of the CYP enzymes in the liver or other relevant tissue. Immunological assays with specific antibodies, such as immunoblot analysis, are the preferred method for quantifying tissue levels of individual CYP enzymes.

### 2.3 Occurrence and Distribution

CYP enzymes are present in most organisms including mammals, birds, fish, reptiles, amphibians, insects, plants, fungi, and bacteria [10, 11]. The finding that CYP enzymes are found in every class of biota suggests that the diverse CYP enzymes identified to date have evolved from a single ancestral gene over a period of 1.36 billion years [12].

In humans and other mammals, CYP enzymes are found in almost all tissues including liver, intestine, lung, kidney, stomach, brain, adrenal gland, gonads, heart, nasal and tracheal mucosa, and skin [13, 14]. Concentrations and expression of individual CYP enzymes differ between tissues. Liver contains the largest number and the highest levels of CYP enzymes involved in drug biotransformation [2].

The endoplasmic reticular membrane is a structure with an extensive surface area in most cells. In rat liver, for example, the endoplasmic reticular membrane comprises 7–11 $m^2$/g of liver weight [15], and total microsomal protein represents 1.8–2% of liver wet weight. CYP enzymes comprise from 4% to 6% of total microsomal protein. This translates to a typical specific content of 0.8–1.1 nmol of total CYP per milligram of microsomal protein or approximately 20 nmol of total CYP per gram rat liver. In human liver, the total CYP content is lower, with CYP enzymes typically

**TABLE 1 Total CYP Content in Various Human and Rat Organs**

| Organ | Human (nmol/mg microsomal protein) | Rat (nmol/mg microsomal protein) |
|---|---|---|
| Liver | 0.30–0.60 | 0.8–1.1 |
| Adrenal | 0.23–0.54 | 0.5 |
| Small intestine | 0.03–0.21 | 0.02–0.13 |
| Kidney | 0.03 | 0.05–0.2 |
| Lung | 0.01 | 0.035–0.05 |
| Brain | 0.10 | 0.025–0.05 |
| Testis | 0.005 | 0.07–0.12 |
| Skin | Not determined | 0.05 |
| Mammary | <0.001 | 0.001–0.003 |

*Source*: Compiled from Refs. 15–20.

comprising 1.5–3% of total microsomal protein, giving a specific content of 0.3–0.6 nmol of total CYP per milligram of microsomal protein (Table 1), or approximately 5 nmol of total CYP per gram human liver. For tissues other than liver, the CYP content, expressed in terms of microsomal protein or tissue weight, is lower (Table 1).

### 2.4 Nomenclature of CYP Enzymes

Multiple CYP enzymes have been identified in diverse species. Each CYP enzyme is encoded by a separate gene. To date, nearly 4000 CYP genes have been identified [21]. CYP enzymes are named and categorized according to their amino acid sequences. A systematic nomenclature has been developed to classify CYP enzymes into families and subfamilies based on the degree of similarity of their primary protein structure [22]. According to this system, CYP enzymes within the same family should exhibit at least 40% amino acid sequence identity, and CYP enzymes that share less than 40% identity in amino acid sequences are assigned to different families, which are designated using CYP as an abbreviation for cytochrome P450, followed by an Arabic numeral (e.g., CYP1, CYP2, CYP3, CYP4) [12]. Within a family, members that share more than 55% identity are assigned to the same subfamily, which is designated by a capital letter following the Arabic numeral (e.g., CYP2A, CYP2B, CYP2C) [12]. Individual enzymes are identified using a final number to specify the individual protein (e.g., CYP2A1, CYP2A2, CYP2B1).

To date, over 300 different CYP enzymes representing 50 families, 82 subfamilies, and 67 different species have been identified [12, 15]. In humans, 57 CYP genes belonging to 18 families and 42 subfamilies have been identified thus far [11, 23] (Table 2). In addition, more than 50 pseudogenes have been found. Pseudogenes are defective genes that do not produce functional proteins. They are thought to be relics of gene duplications in which one of the copies has degenerated and lost its function.

Although all members of the CYP superfamily possess highly conserved regions of amino acid sequences, there is variation in the primary sequences among individual CYP enzymes and among related enzymes across species. Small changes in amino acid sequences are reflected in profound differences in substrate specificity.

**TABLE 2 Human CYP Enzymes**

| Family | Subfamily | Individual CYP Enzymes |
|---|---|---|
| CYP1 | CYP1A | CYP1A1, CYP1A2 |
| | CYP1B | CYP1B1 |
| CYP2 | CYP2A | CYP2A6, CYP2A7, CYP2A13 |
| | CYP2B | CYP2B6 |
| | CYP2C | CYP2C8, CYP2C9, CYP18, CYP19 |
| | CYP2D | CYP2D6 |
| | CYP2E | CYP2E1 |
| | CYP2F | CYP2F1 |
| | CYP2J | CYP2J2 |
| | CYP2R | CYP2R1 |
| | CYP2S | CYP2S1 |
| | CYP2U | CYP2U1 |
| | CYP2W | CYP2W1 |
| CYP3 | CYP3A | CYP3A4, CYP3A5, CYP3A7, CYP3A43 |
| CYP4 | CYP4A | CYP4A9, CYP4A11 |
| | CYP4B | CYP4B1 |
| | CYP4F | CYP4F2, CYP4F3, CYP4F8, CYP4F11, CYP4F12 |
| | CYP4V | CYP4V2 |
| | CYP4X | CYP4X1 |
| | CYP4Z | CYP4Z1 |
| CYP5 | CYP5A | CYP5A1 |
| CYP7 | CYP7A | CYP7A1 |
| | CYP7B | CYP7B1 |
| CYP8 | CYP8A | CYP8A1 |
| | CYP8B | CYP8B1 |
| CYP11 | CYP11A | CYP11A1 |
| | CYP11B | CYP11B1, CYP11B2 |
| CYP17 | CYP17 | CYP17 |
| CYP19 | CYP19 | CYP19 |
| CYP20 | CYP20 | CYP20 |
| CYP21 | CYP21A | CYP21A1 |
| CYP24 | CYP24 | CYP24 |
| CYP26 | CYP26A | CYP26A1 |
| | CYP26B | CYP26B1 |
| | CYP26C | CYP26C1 |
| CYP27 | CYP27A | CYP27A1 |
| | CYP27B | CYP27B1 |
| | CYP27C | CYP27C1 |
| CYP39 | CYP39 | CYP39 |
| CYP46 | CYP46 | CYP46 |
| CYP51 | CYP51 | CYP51 |

*Source:* Adapted from D. Nelson website [396].

In addition, CYP enzymes differ in spectral, electrophoretic, and immunologic properties, modes of regulation, and inducibility.

Of the 18 CYP gene families identified in mammals thus far [12, 15], the major hepatic drug-metabolizing enzymes belong to families CYP1, CYP2, and CYP3. CYP enzymes belonging to families such as CYP5, CYP7, CYP8, CYP24, CYP27,

and CYP51 are involved in the biosynthesis and catabolism of endogenous compounds.

The human CYP1 family, for example, consists of three enzymes, CYP1A1 and CYP1A2 in the CYP1A subfamily, and CYP1B1 in the CYP1B subfamily, although all three enzymes are not expressed in all tissues. Enzymes belonging to the CYP1 family are normally present in liver at very low levels but can be highly induced by treatment with planar halogenated and nonhalogenated aromatic compounds such as 3-methylcholanthrene, β-naphthoflavone, and 2,3,7,8-tetrachlorodibenzo-*p*-dioxin (TCDD) [15, 24–26].

The CYP2 family is represented by five major subfamilies, CYP2A, CYP2B, CYP2C, CYP2D, and CYP2E, in liver. CYP2E1, which is the only member of the CYP2E subfamily in various species and is involved in the metabolism of ethanol and other low molecular weight organic solvents, is induced by ethanol, acetone, and the antitubercular drug, isoniazid [24, 27]. The CYP2C enzymes, a quantitatively important group of enzymes in human and rodent liver, appear to be relatively refractory to induction by most drugs and other xenobiotics [27].

The CYP3 family is represented in liver by only one subfamily, CYP3A. CYP3A enzymes are inducible by glucocorticoids, phenobarbital, and macrolide antibiotics [15, 24, 28]. CYP3A enzymes are the most predominant CYP enzymes in human liver and intestine and are involved in the metabolism of a large percentage of therapeutic agents [1].

The CYP4 family, which is represented in liver by CYP4A and CYP4F subfamilies, is less important in xenobiotic metabolism. CYP4A and CYP4F enzymes catalyze the metabolism of fatty acids, leukotrienes, eicosanoic acids, and prostaglandins and are inducible by hypolipidemic agents, such as clofibrate, and by phthalates and other peroxisomal proliferators [15, 29].

In contrast to the CYP enzymes, NADPH-dependent CYP reductase is encoded by one gene. Thus the single form of NADPH-dependent CYP reductase, which exists in human liver, must be capable of interacting with the multiple CYP enzymes found in each cell. Moreover, NADPH-dependent CYP reductase is expressed at a lower level than CYP, with a reductase to CYP ratio of approximately 1:10 or less. NADPH-dependent CYP reductase is slightly inducible (a maximal two- to three-fold increase) by some of the same compounds (e.g., 3-methylcholanthrene, phenobarbital, TCDD) that induce CYP enzymes.

A list of CYP enzymes that have been detected in human tissues is presented in Table 3.

### 2.5 Genetic Polymorphism

Genetic polymorphism is a common phenomenon in the CYP superfamily. A polymorphism arises from a single base change or multiple nucleotide base changes in a gene and leads to a variant form of the coded protein. The mutation can occur in both coding and noncoding regions of the gene. The variant form of the polymorphic CYP enzyme may have enzymatic activity that is identical to that of the native enzyme, or it may exhibit greater or lesser activity, or it may not be expressed. For example, more than 60 polymorphisms are known for CYP2D6 [44]. The different polymorphic variants are designated by the addition of an asterisk followed by a number to the CYP enzyme name. For example, CYP2D6*1 is a commonly

**TABLE 3 CYP Enzymes Detected in Various Human Organs**

| Organ | CYP Genes Expressed (mRNA or Protein) |
|---|---|
| Liver[a] | CYP1A2, CYP2A6, CYP2B6, CYP2C8, CYP2C9, CYP2C18, CYP2C19, CYP2D6, CYP2E1, CYP3A4, CYP3A5, CYP3A7, CYP4A11 |
| Small intestine[a,b] | CYP1A1 (5.6), CYP2C9 (8.4), CYP2C19 (1), CYP2D6 (0.5), CYP2J2 (0.9), CYP3A4 (43), CYP3A5 (16) |
| Nasal mucosa | CYP2A6, CYP2A13, CYP2B6, CYP2C, CYP2J2, CYP3A |
| Trachea | CYP2A6, CYP2A13, CYP2B6, CYP2C, CYP2J2, CYP2S1, CYP3A, CYP4X1 |
| Lung | CYP1A1, CYP1A2, CYP1B1, CYP2A6, CYP2A13, CYP2B6, CYP2C8, CYP2C18, CYP2D6, CYP2E1, CYP2F1, CYP2J2, CYP3A4, CYP3A5, CYP4B1 |
| Stomach | CYP1A1, CYP1A2, CYP2C, CYP2J2, CYP2S1, CYP3A4 |
| Colon | CYP1A1, CYP1A2, CYP1B1, CYP2J2, CYP2S1, CYP3A4, CYP3A5, CYP4F12 |
| Kidney | CYP1B1, CYP2A6, CYP2B6, CYP2E1, CYP2R1, CYP2S1, CYP3A5, CYP4A11, CYP4F2, CYP4F12 |
| Skin | CYP1A1, CYP1A2, CYP1B1, CYP2A6, CYP2B6, CYP2C9, CYP2C18, CYP2C19, CYP2D6, CYP2E1, CYP2S1, CYP3A4, CYP3A5, CYP4A11 |
| Brain | CYP1A1, CYP2B6, CYP2D6, CYP2E1, CYP2U1 |
| Mammary | CYP1A1, CYP1B1, CYP2C, CYP2D6, CYP3A4, CYP3A5, CYP4Z1 |
| Placenta | CYP1A1, CYP2E1, CYP2F1, CYP2S1, CYP3A4, CYP3A5, CYP4B1 |

[a]CYP enzyme levels in liver and small intestine were detected by protein expression, whereas CYP enzymes in other tissues were detected mainly by mRNA expression.
[b]CYP enzyme levels in the small intestine were determined by immunoblot analysis using intestinal microsomes from 31 individuals. The numbers shown in parentheses in the table are mean CYP enzyme levels, expressed as picomoles per milligram microsomal protein, in the small intestine [14]. Note: CYP1A1 was detected in 3, and CYP3A5 was detected in 11, of the 31 samples, and CYP1A2, CYP2A6, CYP2B6, CYP2C8, and CYP2E1 were below the limit of detection [14].

*Source*: Compiled from Refs. 2, 13, 14, and 30–43.

occurring variant of CYP2D6 that codes for an enzyme with high activity toward a number of substrates including debrisoquine and sparteine, whereas CYP2D6*9 codes for an enzyme with low activity. CYP2D6*4 and CYP2D6*5 are nonfunctional alleles, meaning that the CYP2D6 protein is not expressed. Although allelic variation can occur with any CYP gene, polymorphisms of CYP2A6, CYP2C9, CYP2C19, and CYP2D6 are the best characterized (Table 4) [45].

Enzyme polymorphism is inherited and is a major cause of interindividual and ethnic differences in the rates at which individuals metabolize drugs. The activities of the different polymorphic enzymes may vary by more than 100-fold when tested with the same drug substrate. Thus individuals bearing a nonfunctional allele such as CYP2D6*5 or an allele that codes for a CYP enzyme with low activity, such as CYP2D6*9, will metabolize debrisoquine, a drug that is hydroxylated at the 4-position by CYP2D6, more slowly than individuals bearing an allele that codes for a CYP enzyme with high activity, such as CYP2D6*1. People in the former group are classified as poor/slow metabolizers and people in the latter group are classified as extensive/fast metabolizers. Moreover, other people may have multiple copies of a variant allele, such as CYP2D6*2, and are classified as ultrarapid metabolizers [46].

**TABLE 4 Human CYP Polymorphisms with Relatively High Population Frequency**

| CYP Enzyme | Allele[a] | Function |
|---|---|---|
| CYP2A6 | *2 | Inactive |
| | *4 | Inactive |
| CYP2D6 | *2 | Decreased activity |
| | *3 | Inactive |
| | *4 | Inactive |
| | *5 | Inactive |
| | *6 | Inactive |
| | *7 | Inactive |
| | *9 | Decreased activity |
| | *10 | Decreased activity |
| | *17 | Decreased activity |
| | *2xN | Ultrarapid activity |
| CYP2C9 | *2 | Decreased activity |
| | *3 | Decreased activity |
| CYP2C19 | *2 | Inactive |
| | *3 | Inactive |
| | *4 | Inactive |
| | *5 | Inactive |

[a]The *1 allele, which is not presented in the table, codes for CYP enzymes with high activity (i.e., represents the rapid/extensive metabolizer phenotype). The allelic variants listed in the table represent polymorphisms that occur with relatively high frequency and have functional relevance. Many more CYP allelic variants have been detected but are not listed in the table.

*Source*: Adapted from Ref. 45.

## 3 CYP ENZYME REACTIONS

### 3.1 CYP Substrates

The CYP enzyme system is capable of catalyzing the oxidative biotransformation of an almost limitless number of xenobiotics including drugs, solvents, anesthetics, dyes, petroleum products, pesticides, halogenated aromatic hydrocarbons such as polychlorinated biphenyls (PCBs), and plant products. An exhaustive list of substrates for the mammalian CYP enzymes is presented in Table 5. These substrates share a common feature in that most are lipid soluble and are converted by CYP into more water-soluble metabolites. This facilitates their eventual excretion from the body. Xenobiotics are metabolized by a relatively small number of CYP enzymes belonging to families CYP1, CYP2, and CYP3. Many lipid-soluble physiological compounds including fatty acids, steroids, eicosanoids, retinoids, and amino acids also serve as substrates of CYP1, CYP2, and CYP3 enzymes. A characteristic of the CYP1, CYP2, and CYP3 enzymes that contributes to their ability to metabolize such a diverse array of different substrates is that they possess broad and overlapping substrate selectivity.

By contrast, a large subset of CYP enzymes, especially those found in steroidogenic organs such as the adrenal gland and gonads, but also including enzymes present in the liver, are not involved in xenobiotic metabolism. These CYP enzymes

**TABLE 5 Substrates for Mammalian CYP Enzymes**[a]

| Category/Subcategory of Substrates | Examples of Substrates[b] |
|---|---|
| | *Animal/Fungal/Plant Natural Products* |
| Alkaloids | Californine, protopine |
| Components of garlic | Diallyl disulfide, diallyl sulfide |
| Coumarins, furanocoumarins, psoralens | Bergamottin, BFC, coumarin, 7-ethoxycoumarin, methoxsalen |
| Drugs | Caffeine, cocaine, ipomeanol, morphine, nicotine, quinidine, quinine |
| Flavonoids | Aminoflavone, apigenin, 7,8-benzoflavone, biochanin A, daidzein, eriodictyol, flavanone, galangin, genistein, hesperetin, kaempferide, methoxyflavanone, naringenin, prunetin, tangeretin, tamarixetin |
| Flavorants, herbal constituents, spices | Aristolochic acids, capsaicin, glabridin |
| Indoles | Ibogaine, indole, 3-methylindole |
| Mycotoxins | Territrems A, B, C |
| Phytoalexins | Resveratrol |
| Retinoids | All-*trans*-retinoic acid, retinal, 9-*cis*-retinoic acid, 13-*cis*-retinoic acid |
| Safroles | Elemicin, myristicin |
| Terpenes and derivatives | Betulinic acid, eucalyptol, germander, limonenes, (R)-(+)-pulegone, verbenone |
| Vitamins | 1α,25-Dihydroxyvitamin $D_3$, 1α-hydroxyvitamin $D_2$, 1α-hydroxyvitamin $D_3$, vitamin A, vitamin $B_2$, vitamin $D_2$, vitamin $D_3$, vitamin E |
| Various natural products | Benzylisothiocyanate, odorants, oleuropein, pheromones, pyrrolizidine, tannins, tentoxin, yohimbine |
| | *Endogenous Compounds* |
| Amino acids, amines | Arginine, *N*-hydroxyarginine, octopamine, synephrine, tryptamine, tyramine |
| β-Carbolines | Methoxydimethyltryptamine, pinoline |
| Heme compounds | Bilirubin |
| Pineal hormones | Melatonin |
| Hydrocarbons | Acetone, acetol, $H_2O_2$ |
| Lipids | Eicosanoic acids, fatty acids, leukotrienes, lipid peroxides, lipoxins, prostaglandins, thromboxanes |
| Steroid hormones and various steroids | Androgens, bile acids, cholesterol, estrogens, glucocorticoids, lanosterol, mineralocorticoids, progestins, steroid hydroperoxides |
| | *Xenobiotics* |
| Alcohols, aldehydes, alkanes, and related compounds | Acetal, acetoacetate, butanol, butanone, 3,4-dimethylhexane, ethanal, ethane, ethanol, glycerol, hexane, 2-hexanone, methanol, methyl-*t*-butyl ketone, *p*-nitrobenzaldehyde, pentane, pentanol, propanol, nonane, *p*-tolualdehyde |

| | |
|---|---|
| Alkenes, alkynes | Biphenyl acetylene, divinyl ether, ethene, fluroxene, propene |
| Amines and related compounds | Acetanilide, acetylhydrazine, aniline |
| Azo compounds, azoles | Azoxymethane, methylazoxymethanol, pyrazole |
| Benzene and derivatives | Benzene, bromobenzene, 2,6-di-*t*-butyl-4-methylphenol, nitrobenzene, phenol |
| Carcinogens[c] | |
| Alkaloids, alkenes, allylarenes, spices | β-Asarone, butadiene, estragole, isosafrole, methyleugenol, safrole, senecionine, styrene |
| Amines and derivatives | Acetylaminofluorene, aminoazotoluene, aminobiphenyl, benzidine, dimethylaminoazobenzene, Glu-P-1, Glu-P-2, HMP, IQ, MeIQ, MeIQx, 2-naphthylamine, PhIP, Trp-P-1, Trp-P-2 |
| Carbamates, ethers | Diethylether, ethyl carbamate, methyl-*t*-butyl ether, vinyl carbamate |
| Estrogens | Diethylstilbestrol, 17β-estradiol |
| Halogenated hydrocarbons | Benzyl chloride, $CCl_4$, chloroform, bis(chloromethyl)ether, dibromochloropropane, 1,4-dichlorobenzene, dichloromethane, dichloropropane, dimethylcarbamoyl chloride, epichlorohydrin, ethylene dibromide, ethylene dichloride, methyl chloride, methylene chloride, MOCA, trichloroethane, trichloroethylene, Tris-BP, vinyl bromide, vinyl chloride, vinylidene chloride |
| Hydrazines, nitriles, triazenes | Acrylonitrile, agaritine, dimethylhydrazine, dimethylphenyltriazene, procarbazine |
| Mycotoxins | Aflatoxin $B_1$, aflatoxin $G_1$, aflatoxin $M_1$, ochratoxin A, sterigmatocystin |
| Nitro compounds | DEN, DMN, *p*-nitroanisole, 4-nitrobiphenyl, 2-nitrofluorene, nitrosodimethylmorpholine, nitrosomethylbenzylamine, NNAL, NNK, NNN |
| Polycyclic aromatics | Aminoanthracene, 6-aminochrysene, benzo[*a*]pyrene, chrysene, benzophenanthrene, dibenzanthracene, DMBA, MC, naphthalene, nitrobenzo[*a*]pyrene, 6-nitrochrysene, nitropyrene, phenanthrene |
| Catechols, hydroquinones | α-Methyldopa, *m*-AMSA |
| Fungicides | Furametpyr, *o*-phenylphenol |
| Halogenated hydrocarbons | Bromobenzene, chlorobenzene, 1,2-dichlorobenzene, enflurane, halothane, methoxyflurane, tetrachloroethane, tetrachloroethylene, trichlorobenzene |
| Heteroaromatics, nitroaromatics | Isaxonine, lasciocarpaine, *p*-nitrophenol, *N*-nitrosopyrrolidine, suprofen, thiophene, ticrynafen |
| Hydroperoxides | BOOH, *t*-butylhydroperoxide, cumene hydroperoxide, DBOOH, EBOOH, TMOOH |
| Insect repellents | *N*,*N*-Diethyl-*m*-toluamide |
| Oral contraceptives | Desogestrel, dienogest, ethinylestradiol, gestodene, medroxyprogesterone, mestranol, norethisterone |
| Organic solvents, petroleum products | $CCl_4$, chloroform, methylene chloride, pyridine, toluene, vinyl chloride, xylene |
| Pesticides | Aldrin, azobenzene, carbaryl, carbofuran, chlordanes, chlorpyrifos, chlortoluron, DCBN, DDT, diazinon, dieldrin, endosulfan, endrin, heptachlor, kepone, methiocarb, methoxychlor, parathion, phorate, pyrethrins, rotenone, strychnine |
| Plasticizers | Di-(2-ethylhexyl)phthalate |

**TABLE 5** ***Continued***

| Category/Subcategory of Substrates | Examples of Substrates[b] |
|---|---|
| Polyhalogenated aromatics | PBBs, PBDEs, PCBs, TCDD, TCDF |
| Sulfur compounds | Carbon disulfide, thioacetamide, thiobenzamide |
| | *Therapeutic/Pharmacologic Drugs*[c,d] |
| Abortifacients | Mifepristone |
| Alcohol deterrents | Diethyldithiocarbamate, disulfiram |
| Analgesics | |
| Antimigraine drugs | Almotriptan, butalbital, divalproex, eletriptan, dihydroergotamine, ergotamine, flunarizine, lisuride, naratriptan, propranolol, valproic acid, zolmitriptan |
| Antineuralgia drugs | Capsaicin, carbamazepine, duloxetine |
| COX-3 inhibitors | Acetaminophen, aminopyrine, antipyrine |
| Opioid analgesics | Alfentanil, buprenorphine, codeine, dihydrocodeine, ethylmorphine, fentanyl, hydrocodone, hydromorphone, ketobemidone, meperidine, methadone, morphine, noralfentanil, oxycodone, pentazocine, phenazocine, propoxyphene, sufentanil, tramadol |
| Various analgesics | Acetaminophen, acetanilide, carisoprodol, phenacetin |
| Anesthetics, general | |
| Opioid anesthetics | Alfentanil, fentanyl, sufentanil |
| Fluorinated hydrocarbons | Desflurane, enflurane, fluroxene, halothane, isoflurane, methoxyflurane, sevoflurane |
| Various general anesthetics | Ketamine, phencyclidine, propofol, thiamylal, thiopental |
| Anesthetics, local | Bupivacaine, levobupivacaine, lidocaine, ropivacaine |
| Anorectic/antiobesity drugs | Dexfenfluramine, fenfluramine, phenmetrazine, phentermine, sibutramine |
| Antiacne drugs | Retinoic acid, erythromycin, metronidazole, salicylic acid, tazarotenic acid |
| Antiandrogenics | Bicalutamide, flutamide, finasteride, valproate |
| Anti-alopecia drugs | Finasteride, minoxidil |
| Anti-Alzheimer drugs | DMP543, donepezil, galantamine, milameline, tacrine |
| Anti-amyotrophic lateral sclerosics | Riluzole |
| Antiasthmatics/bronchodilators/ nasal decongestants | Flunisolide, fluticasone, methoxyphenamine, montelukast, paraxanthine, pseudoephedrine, salmeterol, theobromine, theophylline, tiaramide, zafirlukast, zileuton |
| Anti-ADHD drugs | d-Amphetamine, imipramine, methamphetamine, methylphenidate, nortriptyline, tamoxetine |
| Antibacterials | |
| Antileprotics | Dapsone, rifabutin |
| Anti-MAC drugs | Rifabutin |
| Antirickettsials | Chloramphenicol, rifampicin |

| | |
|---|---|
| Antituberculars | Isoniazid, rifabutin, rifalazil, rifampicin |
| Fluoroquinolones | Ciprofloxacin, grepafloxacin, pefloxacin |
| Macrolides | Azithromycin, clarithromycin, erythromycin, miocamycin, roxithromycin, troleandomycin |
| Various antibacterials | Chloramphenicol, clindamycin, rifabutin, metronidazole, sulfamethoxazole, sulfamidine, trimethoprim |
| Antibacterials, topical | Sulfadiazine silver |
| Anti-benign prostatic hyperplasia drugs | Alfuzosin, finasteride, prazosin, tamsulosin |
| Anticancer/antineoplastic drugs | All-*trans*-retinoic acid, aminoflavone, anastrozole, AQ4N, artemisinin, $O^6$-benzylguanine, berberine, bexarotene, bicalutamide, bropirimine, busulfan, *t*-butylhydroxyanisole, carmustine, coumarin, cyclophosphamide, dacarbazine, daidzein, DF203, docetaxel, doxorubicin, dutasteride, ellipticine, etoposide, exemestane, fadrozole, finasteride, flutamide, ftorafur, gefitinib, genistein, hexamethylmelamine, ifosfamide, imatinib, imiquimod, irinotecan, karenitecin, ketoconazole, letrozole, lonafarnib, mitotane, mitoxantrone, MMDX, oracin, paclitaxel, procarbazine, sirolimus, tamoxifen, targretin, tauromustine, tegafur, teniposide, thioguanine, 6-thiopurine, thioTEPA, toremifene, trofosfamide, urethane, valproic acid, vinblastine, vincristine, vindesine, vinorelbine |
| Antidepressants | |
| Monamine oxidase inhibitors | Amiflamine, brofaromine, iproniazid, moclobemide, phenelzine, tranylcypromine |
| NMRIs, tricyclic antidepressants | Amitriptyline, clomipramine, desipramine, doxepin, imipramine, medifoxamine, norclomipramine, nortriptyline, opipramol, trimipramine |
| SNRIs, S&NRIs, SSRIs | Citalopram, duloxetine, escitalopram, fluoxetine, fluvoxamine, maprotiline, milnacipran, norfluoxetine, paroxetine, reboxetine, sertraline, triazoledione, tomoxetine, venlafaxine |
| Various antidepressants | Adinazolam, bupropion, clonazepam, cotinine, gepirone, iprindole, iproclozide, mianserin, minaprine, mirtazapine, nefazodone, trazodone, triazoledione |
| Anti-dermatitis herpetiformis drugs | Dapsone |
| Antidiabetics | Acetohexamide, chlorpropamide, CPO, glibenclamide, gliclazide, glimepiride, glipizide, nateglinide, phenformin, pioglitazone, repaglinide, rosiglitazone, tolbutamide, troglitazone |
| Antiepileptics | Carbamazepine, clobazam, clomethiazole, clonazepam, diazepam, ethosuximide, felbamate, ganaxolone, losigamone, mephenytoin, mephobarbital, nimetazepam, phenobarbital, phenytoin, primidone, tiagabine, topiramate, trimethadione, valproic acid, zonisamide |
| Antierectile dysfunction drugs | Sildenafil, vardenafil, yohimbine |
| Antiestrogenics | Tamoxifen, toremifene |
| Antifungals | Clotrimazole, fluconazole, itraconazole, ketoconazole, miconazole, sirolimus, terbinafine, voriconazole |
| Antigout, uricosuric drugs | Benzbromarone, colchicine, dexamethasone, hydrocortisone, indomethacin, methylprednisolone, phenylbutazone, sulfinpyrazone, tienilic acid, zoxazolamine |

**TABLE 5** ***Continued***

| Category/Subcategory of Substrates | Examples of Substrates[b] |
|---|---|
| Antihistamines | Azelastine, brompheniramine, chlorpheniramine, cinnarizine, clemastine, cyproheptadine, diphenhydramine, doxylamine, ebastine, flunarizine, hydroxyzine, loratadine, medrylamine, mequitazine, orphenadrine, pheniramine, promethazine, rupatadine, tripelennamine |
| Antihyperlipoproteinemics | |
| Fibrates | Bezafibrate, ciprofibrate, clofibrate, fenofibrate, gemfibrozil |
| HMGRIs | Atorvastatin, cerivastatin, fluvastatin, lovastatin, pravastatin, simvastatin |
| Antihyperprolactinemics | Lisuride, terguride |
| Anti-inflammatory drugs | |
| COSAIDs | Budesonide, dexamethasone, fluticasone, mometasone, prednisolone |
| NSAIDs | |
| Anti-FMF drugs | Colchicine |
| COX-2 selective inhibitors | Celecoxib, etoricoxib, lumiracoxib, rofecoxib, valdecoxib |
| Various NSAIDs | Aceclofenac, acetylsalicylic acid, alclofenac, berberine, diclofenac, etodolac, flurbiprofen, ibufenac, ibuprofen, indomethacin, ketorolac, lisofylline, lornoxicam, mefenamic acid, meloxicam, naproxen, oxyphenbutazone, phenylbutazone, piroxicam, salicylate, suprofen, tazofelone, tenoxicam, tiaramide, tolmetin, zomepirac |
| Anti-malignant hyperthermia drugs | Dantrolene |
| Antimanics | Carbamazepine |
| Antinarcolepsy drugs | d-Amphetamine, methylphenidate, modafinil |
| Anti-OCD drugs | Clomipramine, fluoxetine, fluvoxamine, paroxetine, sertraline |
| Anti-opioid addiction drugs | Buprenorphine, methoxycoronaridine, vanoxerine |
| Anti-osteoporosis drugs | Raloxifene |
| Anti-parkinsonian drugs | Amantadine, benztropine, bromocriptine, lisuride, orphenadrine, procyclidine, rasagiline, ropinirole, terguride, tolcapone |
| Antiprogestins | Valproate |
| Antipruritics | Camphor, cyproheptadine, doxepin |
| Antipsoriatics | Bergapten, retinoic acid, methoxsalen, phenol, salicylic acid, tazarotenic acid |
| Antipyretics | Aceclofenac, acetaminophen, acetanilide, acetylsalicylic acid, alclofenac, aminopyrine, ibuprofen, indomethacin, naproxen, phenacetin, tiaramide |
| Antirheumatics | |
| Corticosteroids | Cortisone, dexamethasone, methylprednisolone, prednisolone |
| Various antirheumatics | Azathioprine, capsaicin, chloroquine, hydroxychloroquine sulfate, leflunomide |
| Antiseptics | Naphthalene, phenol, terpineol |
| Antithyroid drugs | Methimazole, methylthiouracil, propylthiouracil |
| Antitussives | Codeine, dextromethorphan, dihydrocodeine, dimemorfan, dropropizine, ethylmorphine, hydrocodone |
| Anti-urinary incontinence drugs | Oxybutynin, scopolamine, tolterodine |

| | |
|---|---|
| Antivirals | |
| Anti-influenza drugs | Amantadine, chloroquine |
| Anti-hepatitis B drugs | Pradefovir |
| Anti-herpetics | Valacyclovir |
| Anti-HIV drugs | Amprenavir, atazanavir, capravirine, delavirdine, efavirenz, indinavir, lopinavir, methoxycoronaridine, nelfinavir, nevirapine, ritonavir, saquinavir, tipranavir, zidovudine |
| Anti-papillomavirus drugs | Imiquimod, resiquimod |
| Anxiolytics/hypnotics/sedatives | Adinazolam, alpidem, alprazolam, brofaromine, brotizolam, buspirone, chlordiazepoxide, citalopram, clomethiazole, clomipramine, clonazepam, diazepam, escitalopram, fenobam, flunitrazepam, fluvoxamine, gespirone, imipramine, meprobamate, mexazolam, midazolam, nimetazepam, nordiazepam, opipramol, oxazepam, paroxetine, pimozide, sertraline, temazepam, triazolam, zaleplon, zolpidem, zopiclone |
| Anti-insomnials | Brotizolam, eszopiclone, zaleplon, zolpidem, zopiclone |
| Cardiovascular drugs | |
| ADR receptor agonists | Octopamine, synephrine |
| Antianginals | Amiodarone, amlodipine, bunitrolol, diltiazem, isosorbide dinitrate, metoprolol, nicardipine, nifedipine, oxprenolol, perhexiline, pronethalol, propranolol, ranolazine, timolol |
| Antiarrhythmics | Ajmaline, amiodarone, aprindine, azimilide, bepridil, bunitrolol, cibenzoline, digoxin, diltiazem, diprafenone, disopyramide, lidocaine, mexiletine, oxprenolol, phenytoin, procainamide, pronethalol, propafenone, propranolol, quinidine, sparteine, timolol, verapamil |
| Anticoagulants | Acenocoumarol, dicoumarol, phenprocoumon, warfarin |
| Antiglaucomics | Betaxolol, brinzolamide, dorzolamide, pilocarpine, pindolol, timolol |
| Antihypotensives | Desglymidodrine |
| Antihypertensives | |
| α-ADR receptor antagonists | Alfuzosin, indoramin, labetolol, prazosin, tamsulosin, yohimbine |
| β-ADR receptor antagonists | Betaxolol, bisoprolol, bufuralol, bunitrolol, bupranolol, carteolol, carvedilol, labetolol, metoprolol, nebivolol, oxprenolol, pindolol, pronethalol, propranolol, timolol |
| ACE inhibitors | Captopril, enalapril |
| AII receptor antagonists | Candesartan, irbesartan, losartan, tasosartan, valsartan |

**TABLE 5** ***Continued***

| Category/Subcategory of Substrates | Examples of Substrates[b] |
|---|---|
| Calcium channel blockers | Amlodipine, barnidipine, bepridil, diltiazem, felodipine, gallopamil, isradipine, lomerizine, mibefradil, nicardipine, nifedipine, niludipine, nilvaldipine, nimodipine, nisoldipine, nitrendipine, norverapamil, oxodipine, pranidipine, verapamil |
| Various antihypertensives | Ajmaline, debrisoquine, eplerenone, guanabenz, guanoxan, pargyline, pheniprazine |
| Antiplatelets | Acetylsalicylic acid, cilostazol, clopidogrel, seratrodast, sulfinpyrazone, ticlopidine |
| Cardiac inotropics | Digitoxin, digoxin, pimobendan, vesnarinone |
| Lipid peroxidase inhibitors | Tirilazad |
| Vasodilators | Bosentan, buflomedil, cinnarizine, dihydralazine, flosequinan, flunarizine, glyceryltrinitrate, hydralazine, isosorbide dinitrate, minoxidil, nimodipine, perhexiline |
| Diuretics | Caffeine, furosemide, perhexiline, tienilic acid, torasemide, triamterene |
| Estrogenics | 17α-Ethinylestradiol |
| Gastrointestinal drugs | |
| Antidiarrheals | Berberine, loperamide |
| Antiemetics | Aprepitant, diphenidol, dolasetron, dronabinol, ezlopitant, granisetron, meclozine, metoclopramide, ondansetron, scopolamine, tropisetron, zatosetron |
| Antigastroenteritis drugs | Berberine |
| Antiulceratives | Cimetidine, esomeprazole, lansoprazole, omeprazole, pantoprazole, rabeprazole, ranitidine |
| Prokinetic drugs | Ecabapide, metoclopramide, mosapride |
| Immunosuppressants/ immunomodulators | Azathioprine, busulfan, cyclosporin A, imiquimod, laquinimod, lisofylline, mycophenolic acid, resiquimod, roquinimex, sirolimus, tacrolimus |
| Mucolytics | Ambroxol |
| Multidrug resistance modulators | Valspodar |
| Muscle relaxants | Carisoprodol, chlordiazepoxide, chlorzoxazone, cyclobenzaprine, dantrolene, diazepam, meprobamate, nimetazepam, orphenadrine, oxybutynin, tizanidine, tolterodine, zoxazolamine |
| Mydriatics | Atropine, scopolamine, yohimbine |
| Neuroleptics | Aripiprazole, bromoperidol, chlorpromazine, clozapine, etoperidone, fluperlapine, fluphenazine, haloperidol, harmaline, harmine, iloperidone, mazapertine, methoxyamphetamine, methoxyphenamine, olanzapine, perazine, perospirone, perphenazine, pimozide, protriptyline, quetiapine, remoxipride, risperidone, sertindole, thioridazine, tiospirone, trazodone, trifluperidol, ziprasidone, zotepine, zuclopenthixol |
| Nootropics | Nefiracetam, tacrine |
| Oxytocics | Sparteine |

| | |
|---|---|
| Parasiticides | |
| Anthelmintics | Albendazole, diethylcarbamazine, ivermectin, naphthalene, praziquantel, pyrantel, thiabendazole |
| Antimalarials | Amodiaquine, arteether, artelinic acid, artemether, artemisinin, artesunate, berberine, chloroquine, chlorproguanil, halofantrine, lumefantrine, mefloquine, primaquine, proguanil, pyrimethamine, quinidine, quinine, tafenoquine |
| Ectoparasiticides | Carbaryl, deltamethrin, lindane, malathion, methoxychlor, permethrin |
| Various parasiticides | Metronidazole, pentamidine, sulfamethoxazole, tinidazole |
| Photosensitizers | Methoxsalen, trioxsalen |
| Signal transduction inhibitors | Imatinib |
| Smoking cessation aids | Bupropion |
| Stimulants | |
| Hallucinogenics | Ibogaine, LSD, MBDB, MDA, MDE, MDMA, mescaline, methoxyamphetamine, phencyclidine, THC |
| Various stimulants | d-Amphetamine, benzphetamine, caffeine, methamphetamine, methylphenidate, modafinil |
| Topical keratolytics | Salicylic acid |
| Ultraviolet screens | *p*-Aminobenzoic acid |

[a]Substrates presented in this Table are metabolized by CYP enzymes derived primarily from human and rodent species.

[b]Abbreviations: AII, angiotensin II; ACE, angiotensin converting enzyme; ADHD, attention-deficit hyperactivity disorder; ADR, adrenergic; AQ4N, 1,4-bis{[2-(dimethylamino-*N*-oxide)ethyl]amino}-5,8-dihydroxyanthracene-9,10-dione; BFC, 7-benzyloxy-4-trifluoromethylcoumarin; BOOH, 2-*t*-butyl-4-hydroperoxy-4-methyl-2,5-cyclohexadienone; CNS, central nervous system; COSAIDs, corticosteroidal anti-inflammatory drugs; COX, cyclooxygenase; CPO, 5-{4-[3-(4-cyclohexyl-2-propylphenoxy)propoxy]phenyl}-1,3-oxazolidine-2,4-dione; DBOOH, 2,6-di-*t*-butyl-4-methyl-4-hydroperoxy-2,5-cyclohexadienone; DCBN, 2,6-dichlorobenzonitrile; DDT, dichlorodiphenyltrichloroethane; DEN, diethylnitrosamine; DF203, 2-(4-amino-3-methylphenyl)benzothiazole; DMBA, 7,12-dimethylbenz[*a*]anthracene; DMN, dimethylnitrosamine; DMP543, 10,10-bis(2-fluoro-4-pyridinylmethyl)-9(10*H*)anthracenone; EBOOH, 2,6-di-*t*-butyl-4-ethyl-4-hydroperoxy-2,5-cyclohexadienone; FMF, familial Mediterranean fever; Glu-P-1, 2-amino-6-ethyldipyrido[1,2-*a*:3′,2′-*d*]imidazole; Glu-P-2, 2-aminodipyrido[1,2-*a*:3′,2′-*d*]imidazole; HIV, human immunodeficiency virus; HMGRIs, 3-hydroxy-3-methylglutarylcoenzyme A reductase inhibitors; HMP, hexamethylphosphoramide; IQ, 2-amino-3-methylimidazo[4,5-*f*]quinoline; LSD, lysergic acid diethylamide; MAC, *Mycobacterium avium* complex; MBDB, *N*-methylbenzodioxazolylbutanamine; MC, 3-methylcholanthrene; MDA, 3,4-methylenedioxyamphetamine; MDE, 3,4-methylenedioxyethylamphetamine; MDMA, 3,4-methylenedioxymethamphetamine; MeIQ, 2-amino-3,5-dimethylimidazo[4,5-*f*]quinoline; MeIQx, 2-amino-3,8-dimethylimidazo[4,5-*f*]quinoline; MMDX, methoxymorpholinodoxorubicin; MOCA, 4,4′-methylene-bis(2-chloroaniline); NMRIs, nonselective monamine reuptake inhibitors; NNAL, 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanol; NNK, 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone; NNN, nornitrosonicotine; NSAIDs, nonsteroidal anti-inflammatory drugs; OCD, obsessive–compulsive disorder; PBBs, polybrominated biphenyls; PBDEs, polybrominated diphenyl ethers; PCBs, polychlorinated biphenyls; PhIP, 2-amino-1-methyl-6-phenylimidazo[4,5-*b*]pyridine; PPAR, peroxisome proliferator-activated receptor; SNRIs, selective noradrenaline reuptake inhibitors; S&NRIs, serotonin and noradrenaline reuptake inhibitors; SSRIs, selective serotonin reuptake inhibitors; TCDD, 2,3,7,8-tetrachlorodibenzo-*p*-dioxin; TCDF, 2,3,7,8-tetrachlorodibenzofuran; THC, $\Delta^9$-tetrahydrocannabinol; thioTEPA, triethylenethiophosphoramide; TOOH, 4-hydroperoxy-2,4,6-trimethyl-2,5-cyclohexadienone; Tris-BP, tris(2,3-dibromopropyl)phosphate; Trp-P-1, 3-amino-1,4-dimethyl-5H-pyrido[4,3-*b*]indole; Trp-P-2, 3-amino-1-methyl-5H-pyrido[4,3-*b*]indole.

[c]Some of the substrates listed as carcinogens or drugs are considered to be procarcinogens or prodrugs, respectively, and need to be bioactivated by CYP enzymes [15, 47–59].

[d]Therapeutic properties of the drugs were determined using human subjects. Drugs are listed under several alphabetically arranged categories/subcategories and are further classified under specific pharmacologic/chemical subheadings within a category. Certain drugs that are used for more than one indication are classified under more than one section. Drugs have been selected on the basis of past and present importance. Certain drugs are no longer marketed as therapeutic agents.

*Source*: Compiled from Refs. 10 and 60–108. A comprehensive listing of substrates for CYP enzymes can be found in book chapters [2, 15, 50, 53, 55], a review article [89], and websites [397, 398].

have much narrower substrate specificity and function exclusively in the biotransformation of physiological compounds. For example, CYP5 and CYP8A are involved in the production of thromboxane and prostacyclin. CYP11, CYP17, CYP19, and CYP21 play important roles in steroid biosynthesis. CYP7, CYP8B, and CYP27A are involved in the synthesis of bile acids, and CYP51 is involved in cholesterol biosynthesis. CYP24 and CYP27B metabolize vitamin D, while enzymes in the CYP26 family metabolize retinoic acid [23].

### 3.2 Types of Reactions Catalyzed by CYP Enzymes

CYP reactions are typically oxidative reactions that expose or add a functional group on to the substrate molecule. In addition to catalyzing the biotransformation of xenobiotics, CYP enzymes are also involved in the metabolism of endogenous compounds and the biosynthesis of signaling molecules used to control homeostasis and development [21].

Depending on the structure of the substrate, CYP enzymes catalyze a wide range of monooxygenase reactions (i.e., insertion of one atom of oxygen) including aromatic and aliphatic hydroxylations, *N*-, *O*-, and *S*-dealkylations, deaminations, *N*-, *P*-, and *S*-oxidations, and dehalogenations. Less commonly, CYP enzymes can catalyze reduction reactions under conditions of low oxygen tension (Table 6). CYP-catalyzed biotransformation often results in inactivation or a decrease in the biological activity of the substrate, although it can also lead to formation of chemically reactive metabolites that can bind to macromolecules such as proteins and DNA and initiate toxic, teratogenic, and carcinogenic events [47, 48]. Other products may be stable molecules possessing intrinsic biological activity that differs from that of the substrate.

It must be emphasized that one CYP enzyme may catalyze several pathways of drug metabolism, and one pathway may be catalyzed by several CYP enzymes. For example, human CYP3A4 can catalyze the hydroxylation of testosterone, *N*-demethylation of alfentanil, and *N*-oxidation of nifedipine, to name just a few reactions, and a drug such as acetaminophen can be metabolized by CYP1A2, CYP2E1, and CYP3A4. Furthermore, CYP enzymes often exhibit regio- and stereoselectivity toward a particular substrate. A good example is testosterone, which undergoes selective ring hydroxylation by various CYP enzymes at either the alpha or beta positions at carbons 2, 6, 7, and 16. Unfortunately, there is no relationship between various CYP enzymes and their abilities to catalyze specific types of reactions so that it is not possible, for example, to characterize an individual CYP enzyme as strictly an *N*-demethylase or an aromatic hydroxylase.

### 3.3 CYP Enzyme Marker Activities

An ideal marker enzyme activity should be catalyzed by a single CYP enzyme. However, in view of the versatile and nonspecific nature of CYP enzymes, it is difficult to find completely specific marker activities for individual CYP enzymes under all conditions.

Human and rodent liver contains more than 25 or 30 different CYP enzymes that can, under *in vitro* conditions, participate to varying degrees in the biotransformation of a given xenobiotic. However, at the concentration of substrates (e.g., drugs)

**TABLE 6 Reactions Catalyzed by CYP Enzymes**

| Types of Reactions[a] | Examples of Substrates[a,b] | Product(s) Formed |
|---|---|---|
| **Oxidative reactions** | | |
| Alicyclic carbon hydroxylation | Acetohexamide | *trans*-4-Hydroxyacetohexamide |
| | Cyclophosphamide | 4-Hydroxycyclophosphamide |
| | Digitoxin | Digoxin |
| | Minoxidil | 4′-Hydroxyminoxidil |
| Aliphatic amine hydroxylation | Amantadine, benzylamphetamine, chlorphentermine, nicotine, phenmetrazine, phentermine | *N*-Hydroxy derivatives |
| Aliphatic carbon hydroxylation | *n*-Alkanes, barbiturates, bile acids, cyclohexane, fatty acids, prostaglandins, steroids, drugs (e.g., capsaicin, chlorpropamide, flutamide, glutethimide, ibuprofen, meprobamate, phenylbutazone, valproic acid) | ω- and (ω-1)-Hydroxylation; hydroxylation at other sites |
| Alkene epoxidation | Aflatoxin $B_1$ | Aflatoxin $B_1$ 2,3-oxide |
| | Carbamazepine | Carbamazepine 10,11-oxide |
| | Diethylstilbestrol | Diethylstilbestrol oxide |
| | Styrene | Styrene oxide |
| Aromatic amide hydroxylation | Acetaminophen, 2-acetylaminofluorene | *N*-Hydroxy derivatives |
| Aromatic carbon epoxidation | Benzene, benzo[*a*]pyrene, bromobenzene, DMBA, 4-ipomeanol, naphthalene, phenytoin | Epoxide derivatives |
| Aromatic carbon hydroxylation | Amphetamine, chlorpromazine, coumarins, diazepam, imipramine, lidocaine, lovastatin, mephenytoin, methadone, phenobarbital, polycyclic aromatics, promethazine, propranolol, salicylic acid, warfarin | Hydroxylation at different sites (certain reactions may proceed via an epoxide intermediate) |
| Aromatization | Felodipine, nicardipine, nifedipine | Aromatization derivatives |
| Arylamine hydroxylation | Aniline | Phenylhydroxylamine |
| | Dapsone | *N*-Hydroxydapsone |
| Cyclization | Proguanil | Cycloguanil |

**TABLE 6** ***Continued***

| Types of Reactions[a] | Examples of Substrates[a,b] | Product(s) Formed |
|---|---|---|
| *N*-Dealkylation | Aminopyrine | Monomethyl-4-aminoantipyrine |
| | Caffeine | 1,7-Dimethylxanthine |
| | Diazepam | Nordiazepam |
| | Halofantrine | Desbutylhalofantrine |
| | Iproniazid | Isoniazid |
| | Lidocaine | Monoethylglycylxylidine |
| | (*S*)-Mephenytoin | Nirvanol |
| | Mephobarbital | Phenobarbital |
| | Methamphetamine | Amphetamine |
| | Promethazine | 10-(2-Aminopropyl)phenothiazine |
| | Propranolol | Desisopropylpropranolol |
| | Sibutramine | Norsibutramine, dinorsibutramine |
| | Sildenafil | Norsildenafil |
| *O*-Dealkylation | Artelinic acid | Dihydroquinghaosu |
| | Astemizole | Desmethylastemizole |
| | Codeine | Morphine |
| | 7-Ethoxycoumarin | 7-Hydroxycoumarin |
| | Ethylmorphine | Morphine |
| | Methoxyamphetamine | Hydroxyamphetamine |
| | Phenacetin | Acetaminophen |
| | Reboxetine | Desethylreboxetine |
| *S*-Dealkylation | 2-Benzylthio-5-trifluoromethylbenzoic acid | 2-Thio-5-trifluoromethylbenzoic acid |
| | Methitural | Desmethylmethitural |
| | Methylmercaptan | Mercaptan |
| | 6-Methylthiopurine | 6-Thiopurine |
| Deamination | Amphetamine | Phenylacetone |
| | Bisdesmethylbrompheniramine | 3-(*p*-Bromophenyl)-3-pyridyl-propionate |
| | (*S*)-(+)-α-Methyldopamine | 3,4-Dihydroxyphenylacetone |
| | Propranolol | Desisopropylpropranolol |
| Dehalogenation | Chloramphenicol | Oxamic acid derivative |
| | Chloroform | Phosgene |
| | Dibromomethane | Hydrogen bromide |
| | Halothane | Trifluoroacetic acid |

| | | |
|---|---|---|
| Dehydrogenation | *N*-Methylformamide, valproic acid, zafirlukast | Dehydrogenated derivatives |
| Denitrification | 2-Nitropropane | Nitrite, acetone |
| Denitrosation | *N*-Dimethylnitrosamine | Nitrite, methylamine |
| Descarbethoxyation | Loratadine | Descarboethoxyloratadine |
| Desulfuration | Chlorpyrifos, diazinon, parathion | Oxon derivatives |
| | Thiopental | Pentobarbital |
| Heterocyclic ring oxidation | Phenmetrazine | 3-Oxophenmetrazine |
| *N*-Hydroxylation | Acetaminophen, acetylhydrazine, 6-aminochrysene, aminoglutethimide, amphetamine, isoniazid, phenacetin, phenmetrazine, phenytoin, riluzole | *N*-Hydroxy derivatives |
| Hydroxylation at allylic carbon atoms | Hexobarbital | 3′-Hydroxyhexobarbital |
| | Pentazocine | *trans*- and *cis*-Hydroxypentazocine |
| | Quinidine | 3-Hydroxyquinidine |
| | THC | 7-Hydroxy-THC |
| Hydroxylation at benzylic carbon atoms | Amphetamine | α-Hydroxyamphetamine |
| | Methaqualon | 2′-Hydroxymethylmethaqualon |
| | 3-Methylcholanthrene | 3-Hydroxymethylcholanthrene |
| | Metoprolol | α-Hydroxymetoprolol |
| | Tolbutamide | 4-Hydroxytolbutamide |
| Hydroxylation of carbon atoms adjacent to a carbonyl or imine group | Diazepam | 3-Hydroxydiazepam |
| | Flurazepam | 3-Hydroxyflurazepam |
| | Glutethimide | 4-Hydroxyglutethimide |
| | Nimetazepam | 3-Hydroxynimetazepam |
| Hydroxylation of indole ring | Dolasetron, indole, ondansetron, tropisetron | Hydroxy derivatives |
| Isoxazole ring cleavage | Leflunomide | α-Cyanoenol derivative |
| Methyl hydroxylation | Dacarbazine, etoricoxib, ibuprofen, 3-methylindole, midazolam, montelukast, naproxen, omeprazole, terfenadine, tolterodine, triazolam, zolpidem | Hydroxymethyl derivatives |
| *N*-Oxidation | Dapsone, dolasetron, mephentermine, nifedipine, quinidine, senecionine, sparteine, trimethylamine | *N*-Oxidation derivatives |
| *P*-Oxidation | Diethylphenylphosphine | *P*-Oxidation derivative |
| *S*-Oxidation | Chlorpromazine, cimetidine, lansoprazole, metiamide, montelukast, promethazine, thioacetamide, thiobenzamide, thioguanine, thioridazine, zotepine | Sulfoxide derivatives |

**TABLE 6** ***Continued***

| Types of Reactions[a] | Examples of Substrates[a,b] | Product(s) Formed |
|---|---|---|
| Pyridine ester deesterification | Dehydrogenated pranidipine | OPC-13463 |
| Ring formation | Pulegone | Menthofuran |
| Transannular dioxygenation | 9,10-Dimethyl-1,2-benz[*a*]anthracene | 9,10-Dimethyl-1,2-benz[*a*]anthracene -9,10-epidioxide |
| **Reductive reactions** | | |
| *t*-Amine *N*-oxide reduction | AQ4N | AQ4 |
| | *N*,*N*-Dimethylaniline *N*-oxide | *N*,*N*-Dimethylaniline |
| | Imipramine *N*-oxide | Imipramine |
| Arene oxide reduction | Benzo[*a*]pyrene-4,5-oxide | Benzo[*a*]pyrene |
| Azido reduction | Zidovudine | 3′-Amino-3′-deoxythymidine |
| Azo reduction | Azobenzene | Aniline |
| | Azsulfidine | Sulfapyridine, *p*-aminosalicylic acid |
| | Prontosil | 2,4-Diaminoaniline, sulfanilamide |
| Chromate reduction | Chromate (VI) | Chromate (III) |
| Hydroperoxide reduction | *t*-Butyl hydroperoxide, cumene hydroperoxide, $H_2O_2$, linoleic acid hydroperoxide, steroid hydroperoxides | Hydroxy derivatives, other products |
| Nitro reduction | Chloramphenicol | Dichloro-*N*-[2-hydroxy-1-(hydroxymethyl)-2-(4-aminophenyl) ethyl]acetamide |
| | Nitrobenzene | Aniline |
| | *p*-Nitrobenzoic acid | *p*-Aminobenzoic acid |
| Oxygen reduction | Oxygen | $H_2O$, $H_2O_2$, superoxide |
| Reductive dehalogenation | Carbon tetrachloride | Chloroform |
| | Halothane | Chlorodifluoroethylene |
| Reductive dehydrogenation | Hexachlorocyclohexane | Hexachlorocyclohexene |
| Reductive ring cleavage | Zonisamide | 2-Sulfamoylacetylphenol |

[a]The examples listed are not meant to be comprehensive but are meant to illustrate the diversity of types of reactions and substrates.

[b]Abbreviations: AQ4, 1,4-bis-{[2-dimethylamino)ethyl]amino}-5,8-dihydroxyanthracene-9,10-dione; AQ4N, 1,4-bis-{[2-(dimethylamino-*N*-oxide)ethyl]amino}-5,8-dihydroxyanthracene-9,10-dione; DMBA, 7,12-dimethylbenz[*a*]anthracene; OPC-13463, methyl-2,6-dimethyl-4-(3-nitrophenyl)-3-carboxy-5-pyridinecarboxylate; THC, $\Delta^9$-tetrahydrocannabinol.

*Source*: Compiled from Refs. 15, 50–55, 60, 61, 89, and 109–122.

typically found in tissue after exposure, enzyme kinetics favors a single form of CYP as being the primary catalyst of metabolism for that substrate [27, 49]. Representative examples of CYP enzymes in human liver and their characteristic activities are presented in Table 7. The number of competing CYP enzymes, the amounts of individual enzyme present, and the catalytic efficiencies of each enzyme determine the rate and predominant pathway of metabolism of a compound.

**TABLE 7 Marker Activities and Inhibitors of the Major Human Hepatic CYP Enzymes**

| CYP | Characteristic Marker Activities | Inhibitors[a] |
|---|---|---|
| CYP1A1 | 7-Ethoxyresorufin *O*-deethylation | 9-Hydroxyellipticine |
| CYP1A2 | 7-Methoxyresorufin *O*-demethylation | 7,8-Benzoflavone, fluvoxamine, furafylline |
| | Phenacetin *O*-deethylation | 7,8-Benzoflavone, fluvoxamine, furafylline |
| CYP1B1 | 17β-Estradiol 4-hydroxylation | Propofol |
| CYP2A6 | Coumarin 7-hydroxylation | Diethyldithiocarbamate, methoxsalen |
| CYP2B6 | Bupropion hydroxylation | ThioTEPA, ticlopidine |
| | Efavirenz 8-hydroxylation | ThioTEPA |
| | 7-Ethoxytrifluoromethylcoumarin *O*-deethylation | 9-Ethynylphenanthrene |
| | (*S*)-Mephenytoin *N*-demethylation | ThioTEPA |
| CYP2C8 | Amadiaquine *N*-deethylation | Quercetin |
| | Paclitaxel 6α-hydroxylation | Quercetin |
| CYP2C9 | Celecoxib methyl hydroxylation | Sulfaphenazole |
| | Diclofenac 4′-hydroxylation | Sulfaphenazole, tienilic acid |
| | (*S*,*R*)-Flurbiprofen 4′-hydroxylation | Sulfaphenazole |
| | Naproxen *O*-demethylation | Sulfaphenazole, tienilic acid |
| | Tolbutamide *p*-methyl hydroxylation | Sulfaphenazole, tienilic acid |
| | Torasemide tolylmethyl hydroxylation | Sulfaphenazole |
| CYP2C19 | (*S*)-Mephenytoin 4′-hydroxylation | Fluconazole, tranylcypromine |
| CYP2D6 | Bufuralol 1′-hydroxylation | Quinidine |
| | Debrisoquine 4-hydroxylation | Ajmalicine, quinidine |
| CYP2E1 | Chlorzoxazone 6-hydroxylation | Chlormethiazole, 4-methylpyrazole |
| | Lauric acid ω1-hydroxylation | Diethyldithiocarbamate |
| CYP3A4 | Alfentanil *N*-dealkylation | Ketoconazole, troleandomycin |
| | Alprazolam 4-hydroxylation | Ketoconazole |
| | Midazolam 1′-hydroxylation | Ketoconazole |
| | Nifedipine *N*-oxidation | Gestodene, troleandomycin |
| | (*S*)-Quinidine 3-hydroxylation | Ketoconazole |
| | Testosterone 6β-hydroxylation | Troleandomycin |
| CYP4A11 | Lauric acid ω-hydroxylation | 10-(Imidazolyl)decanoic acid, TS-011 |

[a]Abbreviations: thioTEPA, triethylenethiophosphoramide; TS-011, [*N*-(3-chloro-4-morpholin-4-yl) phenyl-*N*-hydroxyimidoformamide].

*Source*: Compiled from Refs. 2, 15, 50, 51, 53, 55, 89, 123–149, and 398.

### 3.4 CYP Reaction Cycle

The fact that CYP enzymes have been conserved throughout evolution suggests that they serve a vital role in the preservation of life. Many believe the endogenous substrates of CYP enzymes are steroids or some as yet unidentified endogenous compounds [49, 150]. Alternatively, CYP enzymes may serve a critical protective role against the possible harmful effects of exogenous compounds ingested in the diet to which all organisms are exposed continually during their lifetimes. It has also been suggested that, from an evolutionary standpoint, the primary biological role of CYP can be considered to be the detoxification of molecular oxygen ($O_2$) in tissue to water [15]. This is accomplished by first activating oxygen with reducing equivalents supplied by NADPH, but the activation of oxygen is dangerous if not tightly regulated. The regulation is achieved by a mechanism that prevents CYP from activating oxygen unless a substrate molecule is bound to the enzyme to act as a recipient for the reactive oxygen insertion [15].

The basic aspects of the catalytic mechanism and the interactions of the components of the CYP system are generally agreed upon. The overall reaction for a typical CYP-mediated monooxygenation can be represented as follows:

$$RH + NAD(P)H + O_2 \rightarrow ROH + NAD(P) + H_2O$$

where RH is the substrate and ROH is the oxidized product (Fig. 1).

The catalytic cycle can be divided into several discrete steps but basically involves oxygen activation and oxygen insertion (Fig. 2). Binding of substrate to the CYP enzyme occurs in the first step. The substrate binds to a site on the enzyme in close proximity to the heme moiety, which results in displacement of the water molecule that normally occupies the sixth coordination site of the heme iron. Substrate binding also facilitates a change in spin state and a reduction of the heme iron, from the low spin ferric ($Fe^{3+}$) to the high spin ferrous ($Fe^{2+}$) state, by the addition of an electron from NADPH. Molecular oxygen is then able to bind to the high spin ferrous heme iron to form a dioxygen ($Fe^{2+}—O_2$) complex. The addition of a second electron results in conversion to a peroxoiron ($Fe^{2+}OOH$) complex. Protonation and cleavage of the O—O bond produces water and a reactive iron–oxo ($FeO^{3+}$) complex, which transfers its oxygen to the substrate [15, 21, 151–153]. The oxidized substrate is then released, permitting CYP to return to its initial state and repeat the cycle. Oxidation (oxygen insertion) of the substrate typically occurs at the most reactive carbon–hydrogen bond.

### 3.5 CYP Inhibitors

Many clinically important drug interactions involve the effect of one drug on the metabolism of another. Enzyme inhibition, specifically inhibition of CYP enzymes, is a very common mechanism behind drug interactions. Some CYP inhibitors and the CYP enzymes affected are listed in Table 7.

The effect of enzyme inhibition is to reduce the rate of metabolism of drugs, taken concurrently with the inhibitor, that are also substrates of the inhibited CYP enzyme. The affected drugs begin to accumulate in the body and their therapeutic activity and side effects are increased. For example, if erythromycin, an inhibitor of

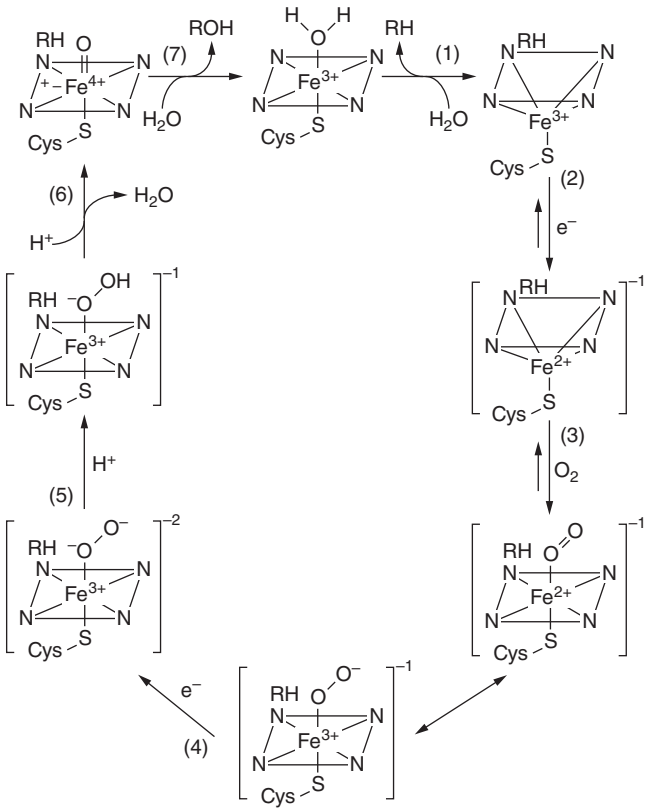


**FIGURE 2** CYP catalytic cycle. Enzymatic steps by which CYP enzymes catalyze the hydroxylation of a substrate (RH) to a product (ROH). Adapted from Ref. 153.

CYP3A4, is taken by a patient who is also being treated with carbamazepine, which is extensively metabolized by CYP3A4, this may lead to toxicity due to higher concentrations of carbamazepine.

Inhibition is dose related, so that inhibition will not appear until a sufficient concentration of inhibitor is present in the liver. The effects are often maximal when steady-state plasma concentrations of the inhibitor are reached. Moreover, at low concentrations, the inhibitor may be relatively selective for a single CYP enzyme, whereas at high concentrations, the inhibitor may be relatively nonselective and several CYP enzymes will be inhibited.

There are several mechanisms by which drugs and other compounds inhibit CYP enzymes. The commonest mechanism is simple competition by two drugs for metabolism by the same CYP enzyme. The two drugs compete at the substrate binding site to produce reversible and rapid inhibition of the metabolism of one of the drugs, usually the drug with the lower affinity for the enzyme or the drug that is present at substantially lower concentration. Simple competitive inhibition is usually short-acting and is possible with any two drugs. A second more selective type of reversible inhibition is produced by drugs containing an azole group, such as the antifungal agents, ketoconazole, miconazole, and fluconazole, or by drugs containing an

imidazole group, such as cimetidine. In this type of inhibition, a basic nitrogen atom of the azole or imidazole group interacts, via its free electron pair, with the iron in the heme group of CYP and forms a relatively stable but reversible substrate–enzyme complex. A more slowly reversible form of inhibition is produced when a drug is metabolized to a reactive metabolite that binds in a noncovalent fashion with the iron in the CYP heme group. Examples of drugs that produce this type of inhibitory metabolite (i.e., metabolism-dependent inhibitors) include nortriptyline, amiodarone, and the macrolide antibiotics such as erythromycin and clarithromycin. This type of inhibition often increases with repeated dosing and the effects may continue for several weeks after stopping the drug if the drug has a long half-life. Irreversible inhibition is relatively rare but the effects are long-lasting. Irreversible inhibitors such as gestodene, ethinyl estradiol, levonorgestrel, and spironolactone are metabolized to reactive metabolites that form covalent bonds with either the heme group or the apoprotein of CYP, a process that inactivates the enzyme and is only overcome when the inhibited CYP enzyme turns over (i.e., is destroyed and new enzyme is synthesized).

## 4 STRUCTURE AND REGULATION

### 4.1 Overview of Individual CYP Enzymes

The total CYP level in liver may vary two- to threefold, at most, among humans. However, there is considerably greater variation in the enzyme activities and protein levels of individual CYP enzymes within a population. Table 8 illustrates the variability and range of CYP enzyme levels that have been measured for 14 different enzymes in human liver microsomes. The interindividual variation in human hepatic

**TABLE 8 CYP Enzyme Levels[a] in Human Liver Microsomes**

| CYP Enzyme | Content (pmol CYP/mg microsomal protein) | Percentage of Total CYP |
|---|---|---|
| CYP1A1 | Not detected | 0 |
| CYP1A2 | <1–45 | <0.5–15 |
| CYP1B1 | Not detected | 0 |
| CYP2A6 | <1–68 | 4–17 |
| CYP2B6 | 1–39 | <1–7 |
| CYP2C8 | 1–60 | 12–15 |
| CYP2C9 | 9–100 | 14–18 |
| CYP2C18 | <1 | <0.5 |
| CYP2C19 | 0.5–19 | 1–17 |
| CYP2D6 | 1.5–10 | 1–4 |
| CYP2E1 | 7–49 | 3–9 |
| CYP3A4 | 40–140 | 18–29 |
| CYP3A5 | <1 | <0.5 |
| CYP4A11 | Not determined | 10 |

[a]CYP enzyme levels were compiled from Refs. 55, and 155–158. CYP enzyme levels were measured in individual human liver samples by immunoquantification. Values represent the approximate range of CYP levels reported.

CYP levels and activities is related to several factors including genetic polymorphism, enzyme inhibition, and enzyme induction.

### 4.2 CYP1 Family

Two CYP enzymes are present in the 1A subfamily, CYP1A1 and CYP1A2. These enzymes are highly conserved among species [15]. Human CYP1A1 and CYP1A2 share approximately 70% amino acid sequence identity.

CYP1A1 and CYP1A2 preferentially metabolize planar aromatic molecules but exhibit distinct substrate selectivity (Table 9). CYP1A1 and CYP1A2 both catalyze the *O*-dealkylation of 7-methoxyresorufin and 7-ethoxyresorufin and are involved in the bioactivation of chemical carcinogens [2, 159]. CYP1A1 preferentially catalyzes the hydroxylation of benzo[*a*]pyrene leading to the 7,8-epoxide intermediate, whereas CYP1A2 preferentially catalyzes the *N*-hydroxylation of aromatic amines including 4-aminobiphenyl, 2-naphthylamine, and 2-acetylaminofluorene [26, 51]. With respect to drug metabolism, CYP1A2 catalyzes the *N*-hydroxylation of acetaminophen, leading to formation of the reactive benzoquinoneimine metabolite, *N*3-demethylation of caffeine to paraxanthine, and *O*-dealkylation of phenacetin. Ellipticine is an inhibitor of human CYP1A1 and furafylline is a potent metabolism-dependent inhibitor of human CYP1A2 [2, 13].

CYP1A1 is not normally expressed in the liver of humans and most mammals but is detected in extrahepatic tissues such as lung and placenta in humans. CYP1A2 is expressed constitutively and almost exclusively in the liver [26, 158, 160]. Hepatic levels of CYP1A2 vary widely, exhibiting as much as a 40- to 100-fold difference, among individuals [155, 158].

CYP1B1 is the only member of the CYP1B subfamily. Recombinant CYP1B1 catalyzes the oxidation of 7,12-dimethylbenz[*a*]anthracene and the 4-hydroxylation of 17β-estradiol (Table 9). CYP1B1 protein has not been detected in liver [158] and is found primarily in extrahepatic tissues including thymus and the adrenal cortex [55]. CYP1B1 is linked to eye development as mutations in the human CYP1B1 gene are associated with congenital glaucoma [180].

CYP1A1, CYP1A2, and CYP1B1 are inducible enzymes, which are upregulated by the activation of the aryl hydrocarbon receptor (AhR) [26, 181]. In humans, CYP1A enzymes can be induced by polycyclic aromatic hydrocarbons found in charbroiled meat, cigarette smoke, and cruciferous vegetables. Omeprazole, a proton pump inhibitor, has also been shown to induce CYP1A enzymes [2]. In rodents, CYP1A1 levels increase from undetectable to approximately 40% of the total CYP content in the liver, following treatment with xenobiotic inducers such as 3-methylcholanthrene [160]. CYP1A1 levels in the lung, small intestine, and kidneys are also induced. CYP1A2 levels in the liver are increased but CYP1A2 remains undetectable in extrahepatic tissues following administration of inducers such as 3-methylcholanthrene [160].

### 4.3 CYP2 Family

The CYP2 family is the largest CYP family present in humans in terms of number of different CYP enzymes identified. CYP2 enzymes are responsible for steroid

**TABLE 9 Substrates for Human CYP1 Enzymes**[a]

| CYP | Category of Substrates | Examples of Substrates[b] |
|---|---|---|
| CYP1A1 | Benzene derivatives | *o*-Aminoazotoluene, toluene |
| | Carcinogens | |
| | Amino compounds | 2-Acetylaminofluorene, 2-aminoanthracene, 4-aminobiphenyl, 6-aminochrysene, 2-aminofluorene, IQ, MeIQ, MeIQx, methoxyaminoazobenzene, NNK, PhIP, Trp-P-1, Trp-P-2 |
| | Polycyclic aromatics | Benz[*a*]anthracene-3,4-diol, benzo[*g*]chrysene-11,12-diol, benzo[*b*]fluoranthene-9,10-diol, benzo[*a*]perylene, benzo[*a*]pyrene-7,8-diol, chrysene-1,2-diol, dibenz[*a*,*h*]anthracene, dibenzo[*a*,*l*]pyrene-11,12-diol, DMBA, DMBA-3,4-diol, 5,6-dimethylchrysene-1,2-diol, 1-ethylpyrene, 5-methylchrysene-1,2-diol, 6-nitrochrysene, 2-naphthylamine, 2-nitronaphthalene, nitropyrene, phenanthrene |
| | Various carcinogens | Aflatoxin $B_1$, diethylstilbestrol, 17β-estradiol, ochratoxin A |
| | Chloroaromatics | 1,2,4-Trichlorobenzene, 1,2,3,5-tetrachlorobenzene |
| | Coumarins, resorufins | Coumarin, 7-ethoxycoumarin, ethoxyresorufin |
| | Endogenous compounds | Tryptamine, uroporphyrinogen |
| | Flavonoids | 7,8-Benzoflavone, eriodictyol, galangin, genistein, hesperetin, kaempferide, methoxyflavanone |
| | Fungicides | Furametpyr |
| | Retinoids | All-*trans*-retinoic acid, retinal, retinol |
| | Steroids | Androstenedione, estrone, pregnenolone, progesterone, testosterone |
| | Drugs | Abecarnil, albendazole, aminopyrine, amiodarone, amodiaquine, benzydamine, bufuralol, caffeine, capsaicin, carvedilol, chloroquine, chlorzoxazone, cinnarizine, clomethiazole, dacarbazine, diclofenac, ellipticine, flunarizine, flutamide, fluvastatin, lisofylline, lisuride, MDE, nicardipine, nicotine, nifedipine, pentamidine, phenacetin, phencyclidine, pranidipine, propranolol, pyrantel, quinine, riluzole, ropivacaine, tegafur, terguride, theophylline, thiabendazole, tolbutamide, toremifene, troglitazone, warfarin, zotepine |
| CYP1A2 | Benzene derivatives | *o*-Aminoazotoluene, toluene |
| | β-Carbolines | Harmaline, harmine |
| | Carcinogens | |
| | Amines and derivatives | 2-Acetylaminofluorene, 2-aminoanthracene, aminoazotoluene, 4-aminobiphenyl, 6-aminochrysene, 2-aminofluorene, benzidine, dimethylaminoazobenzene, furylfuramide, Glu-P-1, Glu-P-2, IQ, MeIQ, MeIQx, methoxyaminoazobenzene, NNK, PhIP, Trp-P-1, Trp-P-2, 2-naphthylamine |
| | Polycyclic aromatics | Benz[*a*]anthracene-3,4-diol, benzo[*a*]pyrene-7,8-diol, 1,3-dinitropyrene, DMBA, DMBA-3,4-diol, 5-methylchrysene-1,2-diol, 1-methylpyrene, nitropyrene, naphthalene, 6-nitrochrysene, phenanthrene |
| | Various carcinogens | Aflatoxin $B_1$, butadiene, 1,4-dichlorobenzene, 17β-estradiol, methyleugenol, ochratoxin A |

**TABLE 9** ***Continued***

| CYP | Category of Substrates | Examples of Substrates[b] |
|---|---|---|
| | Chloroaromatics | Bromodichloromethane, 1,2-dichlorobenzene, 1,2,4-trichlorobenzene |
| | Coumarins, resorufins | BFC, coumarin, 7-ethoxycoumarin, methoxyresorufin |
| | Endogenous compounds | Arachidonic acid, linoleic acid, $PGH_2$, phosphatidylcholine, uroporphyrinogen |
| | Flavonoids | Aminoflavone, apigenin, biochanin A, daidzein, formononetin, galangin, genistein, hesperetin, kaempferide, kaempferol, methoxyflavanone, naringenin, prunetin, tangeretin, tamarixetin |
| | Fungicides | Furametpyr, *o*-phenylphenol |
| | Hormones | Estrone, melatonin, progesterone, testosterone |
| | Insect repellents | *N*,*N*-Diethyl-*m*-toluamide |
| | Pesticides | Ametryne, atrazine, chlorpyrifos, diazinon, methoxychlor, parathion |
| | Phytoalexins | Resveratrol |
| | Pneumotoxins, other toxins | 3-Methylindole, myristicin |
| | Retinoids | Retinal, 13-*cis*-retinoic acid, retinol |
| | Xanthines | 1,7-Dimethylxanthine, 8-ethylfurafylline, furafylline, paraxanthine |
| | Drugs | Acenocoumarol, acetaminophen, acetanilide, albendazole, aminopyrine, amiodarone, amitriptyline, amodiaquine, antipyrine, benzydamine, $O^6$-benzylguanine, bropirimine, bufuralol, bupivacaine, bupropion, caffeine, capsaicin, carbamazepine, carvedilol, chlordiazepoxide, chlorpromazine, chlorzoxazone, cilostazol, cinnarizine, ciprofloxacin, cisapride, clomipramine, clozapine, cyclobenzaprine, dacarbazine, dapsone, desipramine, diazepam, diclofenac, diethylcarbamazine, dihydralazine, doxepin, ellipticine, enoxacin, etoposide, fenfluramine, flunarizine, flunitrazepam, fluoxetine, flutamide, fluvoxamine, ftorafur, grepafloxacin, guanabenz, haloperidol, hydromorphone, iloperidone, imipramine, imiquimod, ipomeanol, leflunomide, lidocaine, lisofylline, maprotiline, MBDB, MDE, MDMA, (*S*)-mephenytoin, methadone, metoclopramide, mexiletine, mianserin, mirtazapine, naproxen, nicardipine, nicotine, nifedipine, nordiazepam, nortriptyline, olanzapine, ondansetron, pefloxacin, perazine, perphenazine, phenacetin, pimobendan, pimozide, pranidipine, praziquantel, primaquine, proguanil, propafenone, propofol, propranolol, ranitidine, rasagiline, resiquimod, riluzole, ritonavir, rofecoxib, ropinirole, ropivacaine, tacrine, tamoxifen, tauromustine, tegafur, terbinafine, theobromine, theophylline, thiabendazole, tizanidine, tolbutamide, tolcapone, toremifene, triamterene, verapamil, warfarin, zileuton, zolmitriptan, zolpidem, zotepine, zoxazolamine |

**TABLE 9** ***Continued***

| CYP | Category of Substrates | Examples of Substrates[b] |
|---|---|---|
| CYP1B1 | Carcinogens | |
| | Amino compounds | 2-Aminofluorene, IQ, MeIQ, MeIQx, methoxyaminoazobenzene, NNK, PhIP, Trp-P-1, Trp-P-2 |
| | Polycyclic aromatics | 2-Aminoanthracene, 6-aminochrysene, benz[*a*]anthracene-1,2-diol, benzo[*g*]chrysene-11,12-diol, benzo[*b*]fluoranthene-9,10-diol, benzo[*c*]phenanthrene-3,4-diol, benzo[*a*]pyrene, benzo[*a*]pyrene-7,8-diol, chrysene-1,2-diol, dibenzo[*a*,*l*]pyrene-11,12-diol, DMBA, DMBA-3,4-diol, 5,6-dimethylchrysene-11,12-diol, dinitropyrene, 9-hydroxybenzo[*a*]pyrene, 1-ethylpyrene, 5-methylchrysene, 5-methylchrysene-1,2-diol, 1-methylpyrene, 6-nitrochrysene, nitrofluoranthene, nitropyrene |
| | Various carcinogens | Aflatoxin $B_1$, 17β-estradiol |
| | Coumarins, resorufins | Coumarin, 7-ethoxycoumarin, ethoxyresorufin |
| | Flavonoids | Eriodictyol, genistein, hesperetin |
| | Steroid hormones | Estrone, progesterone, testosterone |
| | Fatty acids | Arachidonic acid |
| | Retinoids | Retinal, retinol |
| | Drugs | Albendazole, amodiaquine, artelinic acid, artemisinin, artesunate, bufuralol, caffeine, DF203, docetaxel, ellipticine, flutamide, nifedipine, pyrimethamine, theophylline, thiabendazole, tolbutamide |

[a]Involvement of specific human CYP enzymes in the metabolism of substrates listed in Tables 9–12 was determined using *in vivo* and *in vitro* procedures. *In vivo* techniques included administering drugs and other substrates together with selective CYP inhibitors to human subjects and measuring metabolites in plasma and urine. *In vitro* studies included incubating human hepatic microsomes with substrate along with selective CYP inhibitors or antibodies and measuring inhibition of enzyme activity and product formation. Other *in vitro* studies included incubating purified human hepatic CYP enzymes, or a panel of cDNA expressed human CYP enzymes, with substrate and measuring product formation. Liver is the primary site of expression for the majority of CYP enzymes listed in Tables 9–12.

[b]Abbreviations: BFC, 7-benzyloxy-4-trifluoromethylcoumarin; DF203, 2-(4-amino-3-methylphenyl)benzothiazole; DMBA, 7,12-dimethylbenz[*a*]anthracene; Glu-P-1, 2-amino-6-methyldipyrido[1,2-*a*:3′,2′-*d*]imidazole; Glu-P-2, 2-aminodipyrido[1,2-*a*:3′,2′-*d*]imidazole; IQ, 2-amino-3-methylimidazo[4,5-*f*]quinoline; MBDB, *N*-methylbenzodioxazolylbutanamine; MDE, 3,4-methylenedioxyethylamphetamine; MDMA, 3,4-methylenedioxymethamphetamine; MeIQ, 2-amino-3,5-dimethylimidazo[4,5-*f*]quinoline; MeIQx, 2-amino-3,8-dimethylimidazo[4,5-*f*]quinoline; NNK, 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone; $PGH_2$, prostaglandin $H_2$; PhIP, 2-amino-1-methyl-6-phenylimidazo[4,5-*b*]pyridine; Trp-P-1, 3-amino-1,4-dimethyl-5H-pyrido[4,3-*b*]indole; Trp-P-2, 3-amino-1-methyl-5H-pyrido[4,3-*b*]indole.

*Source*: Compiled from Refs. 10, 51–56, 130, and 161–179. A comprehensive listing of substrates for CYP1 enzymes can be found in book chapters [2, 15, 50, 53, 55] and a review article [89].

hydroxylation and the biotransformation of a wide variety of drugs and xenobiotics, as well as many endogenous compounds (Table 10).

The CYP2A subfamily consists of three genes in humans, namely, CYP2A6, CYP2A7, and CYP2A13. CYP2A6, CYP2A7, and CYP2A13 exhibit >90% identity

**TABLE 10 Substrates for Human CYP2 Enzymes**

| CYP | Category of Substrates | Examples of Substrates[a] |
|---|---|---|
| CYP2A6 | Carcinogens | |
| | Amino compounds | 6-Aminochrysene, PhIP |
| | Nitro compounds | *p*-Nitroanisole, DEN, DMN, NNAL, NNK, NNN |
| | Various carcinogens | Aflatoxin $B_1$, butadiene, chloroform, 17β-estradiol, MOCA, phenanthrene, safrole |
| | Components of garlic | Diallyl sulfide |
| | Coumarins | Coumarin, 7-ethoxycoumarin, psoralen |
| | Ethers, phenols | Ethyl *t*-butyl ether, methyl *t*-butyl ether, *p*-nitrophenol |
| | Fluorinated hydrocarbons | Halothane, methoxyflurane, sevoflurane |
| | Indoles, xanthines | 1,7-Dimethylxanthine, indole, 3-methylindole, paraxanthine |
| | Insect repellants | *N*,*N*-Diethyl-*m*-toluamide |
| | Retinoids | All-*trans*-retinoic acid |
| | Steroid hormones | Progesterone, testosterone |
| | Terpenes | Verbenone |
| | Drugs | Acetaminophen, aminopyrine, amodiaquine, artelinic acid, artesunate, bupropion, chlorzoxazone, cinnarizine, cisapride, clomethiazole, clozapine, cotinine, cyclobenzaprine, cyclophosphamide, fadrozole, flunarizine, flunitrazepam, ftorafur, ifosfamide, letrozole, lidocaine, lisofylline, losigamone, medifoxamine, methoxsalen, mexiletine, naproxen, nicotine, nifedipine, omeprazole, phenacetin, primaquine, propofol, quinoline, ritonavir, tacrine, tamoxifen, tegafur, valproic acid, zidovudine |
| CYP2B6 | Alkanes | Hexane |
| | Aromatic hydrocarbons | *o*-Phenylphenol, toluene |
| | Nitrosoamines | *N*-Nitrosomethylbutylamine, *N*-nitrosomethylethylamine, nitrosomorpholine |
| | Carcinogens | |
| | Polycyclic aromatics | 6-Aminochrysene, benzo[*a*]pyrene-7,8-diol, dibenzo[*a*]pyrene, DMBA, naphthalene, phenanthrene |
| | Various carcinogens | Aflatoxin $B_1$, butadiene, chloroform, 17β-estradiol, isoprene, DMN, NNK, styrene |
| | Coumarins, resorufins | Coumarin, 7-ethoxycoumarin, ethoxyresorufin, benzyloxyresorufin, pentoxyresorufin |
| | Fatty acids | Arachidonic acid, lauric acid |
| | Flavonoids | Flavanone, methoxyflavanone |
| | Fluorinated hydrocarbons | Chlorodifluoroethane, halothane, methoxyflurane, sevoflurane |
| | Insect repellents | *N*,*N*-Diethyl-*m*-toluamide |

**TABLE 10** ***Continued***

| CYP | Category of Substrates | Examples of Substrates[a] |
|---|---|---|
| | Pesticides | Acetochlor, ametryne, butachlor, chlorpyrifos, diazinon, endosulfan, methoxychlor, parathion, terbutryne |
| | Polychlorinated aromatics | 2,4,5,2′,4′,5′-Hexachlorobiphenyl |
| | Retinoids | All-*trans*-retinoic acid |
| | Steroid hormones, sterols | 5β-Cholestane-3α,7α,12α-triol, estrone, testosterone |
| | Terpenes | Verbenone |
| | Drugs | Alfentanil, aminopyrine, amitriptyline, antipyrine, arteether, artelinic acid, artemisinin, artesunate, benzphetamine, $O^6$-benzylguanine, bergamottin, bupropion, capsaicin, carteolol, cinnarizine, clobazam, clomethiazole, cyclophosphamide, diazepam, diclofenac, diethylcarbamazine, efavirenz, erythromycin, ethylmorphine, flunitrazepam, ifosfamide, ketamine, lidocaine, MDA, MDE, MDMA, (*S*)-mephenytoin, mephobarbital, mianserin, midazolam, nevirapine, nicotine, pentobarbital, primaquine, procainamide, promethazine, propofol, quinoline, ropivacaine, selegiline, seratrodast, sertraline, tamoxifen, tazofelone, temazepam, thioTEPA, tinidazole, trofosfamide, valproic acid, verapamil, zotepine |
| CYP2C8 | Carcinogens | Benzo[*a*]pyrene, benzo[*a*]pyrene-7,8-diol, naphthalene, styrene |
| | Fatty acids | Arachidonic acid, 9,10-epoxystearic acid, leukotoxin |
| | Fluorescent probes | Dibenzylfluorescein |
| | Pesticides | Chlorpyrifos, diazinon, methoxychlor, parathion |
| | Retinoids | All-*trans*-retinoic acid, 13-*cis*-retinoic acid, retinol, tazarotenic acid |
| | Steroids | Androstenedione, testosterone |
| | Drugs | Albendazole, aminopyrine, amiodarone, amodiaquine, antipyrine, artelinic acid, benzphetamine, bufuralol, capsaicin, carbamazepine, cerivastatin, chloroquine, cisapride, clozapine, CPO, dapsone, diazepam, diclofenac, fluvastatin, ftorafur, gallopamil, halofantrine, karenitecin, lidocaine, lonafarnib, (*R*)-mephenytoin, mephobarbital, mirtazapine, mycophenolic acid, naproxen, omeprazole, paclitaxel, perospirone, phenprocoumon, phenytoin, pioglitazone, pyrantel, repaglinide, rosiglitazone, sulfadiazine, tegafur, temazepam, terbinafine, tiaramide, tienilic acid, tolbutamide, torasemide, trimethoprim, troglitazone, verapamil, warfarin, zidovudine, zopiclone |

| | | |
|---|---|---|
| CYP2C9 | Carcinogens | |
| | Polycyclic aromatics | Benzo[*a*]pyrene, benzo[*a*]pyrene-7,8-dihydrodiol, dibenzo[*a*]pyrene, naphthalene, phenanthrene |
| | Various carcinogens | Butadiene, 17β-estradiol, MeIQ, methyleugenol, safrole |
| | Components of garlic | Diallyl disulfide |
| | Coumarins | 7-Ethoxycoumarin |
| | Fatty acids | Arachidonic acid, 9,10-epoxystearic acid, isoleukotoxin, leukotoxin, linoleic acid |
| | Flavonoids | Biochanin A, formononetin, tamarixetin |
| | Fluorescent probes | Dibenzylfluorescein |
| | Oral contraceptives | Desogestrel |
| | Retinoids | Retinoic acid, retinol |
| | Steroids | 5α-Androstane-3α,17β-diol, testosterone |
| | Terpenes | Limonene |
| | Drugs | Aceclofenac, acenocoumarol, acetylsalicylic acid, albendazole, amitriptyline, antipyrine, artelinic acid, benzbromarone, bosentan, bufuralol, bupropion, candesartan, capsaicin, carvedilol, celecoxib, chloramphenicol, chlorpheniramine, cinnarizine, clozapine, cyclophosphamide, dapsone, dextromethorphan, diazepam, diclofenac, dicoumarol, dolasetron, dorzolamide, dronabinol, eletriptan, ellipticine, etodolac, etoricoxib, flunarizine, flunitrazepam, fluoxetine, flurbiprofen, fluvastatin, glimepiride, glipizide, hexobarbital, hydromorphone, ibuprofen, ifosfamide, imipramine, indomethacin, irbesartan, ketamine, ketobemidone, lornoxicam, losartan, lovastatin, lumiracoxib, mefenamic acid, meloxicam, (*R*)-mephenytoin, mirtazapine, moclobemide, montelukast, naproxen, nateglinide, nelfinavir, nevirapine, nicotine, ondansetron, perazine, perphenazine, phenobarbital, phenprocoumon, phenylbutazone, phenytoin, piroxicam, pravastatin, proguanil, propofol, pyrantel, ritonavir, rosiglitazone, selegiline, seratrodast, sertraline, sildenafil, simvastatin, sulfadiazine, sulfamethoxazole, sulfinpyrazone, suprofen, tafenoquine, tamoxifen, targretin, tauromustine, tegafur, temazepam, tenoxicam, terbinafine, THC, thiamylal, tiaramide, tienilic acid, tolbutamide, torasemide, trimethadione, trimethoprim, tropisetron, valproic acid, valsartan, venlafaxine, verapamil, voriconazole, warfarin, zafirlukast, zidovudine, zileuton, zopiclone |
| CYP2C18 | Pesticides | Methoxychlor |
| | Retinoids | All-*trans*-retinoic acid |
| | Steroids | Progesterone |
| | Drugs | Aminopyrine, antipyrine, clobazam, diazepam, imipramine, lansoprazole, lidocaine, mephenytoin, naproxen, perphenazine, piroxicam, propofol, tienilic acid, tolbutamide, verapamil, warfarin |

**TABLE 10** ***Continued***

| CYP | Category of Substrates | Examples of Substrates[a] |
|---|---|---|
| CYP2C19 | Carcinogens | Benzo[*a*]pyrene-7,8-dihydrodiol, benzo[*a*]pyrene-7,8-diol, chloroform, 17β-estradiol, methyleugenol |
| | Components of garlic | Diallyl disulfide |
| | Fatty acids | Arachidonic acid |
| | Hormones | Melatonin |
| | Indoles | Indole |
| | Insect repellents | *N*,*N*-Diethyl-*m*-toluamide |
| | Pesticides | Chlorpyrifos, diazinon, furametpyr, methoxychlor, parathion |
| | Prostaglandins | 9,11-Epoxymethano-$PGH_2$, 11,9-epoxymethano-$PGH_2$, 9,11-diazo-15-deoxy-$PGH_2$ |
| | Retinoids | Retinal, retinoic acid |
| | Steroids | Desogestrel, progesterone, testosterone |
| | Terpenes | Limonene |
| | Drugs | Acenocoumarol, adinazolam, albendazole, aminopyrine, amitriptyline, azelastine, benzbromarone, bufuralol, capsaicin, carisoprodol, chlorproguanil, citalopram, clobazam, clomethiazole, clomipramine, clozapine, CPO, dapsone, dextromethorphan, diazepam, diphenylhydantoin, divalproex, eletriptan, esomeprazole, etoricoxib, flunitrazepam, glibenclamide, hexobarbital, imipramine, lansoprazole, loratadine, (*R*)- and (*S*)-mephenytoin, mephobarbital, methadone, methoxycoronaridine, metoprolol, mianserin, moclobemide, naproxen, nelfinavir, nirvanol, nordiazepam, nortriptyline, omeprazole, pantoprazole, pentamidine, perazine, perphenazine, phenobarbital, phenytoin, piroxicam, praziquantel, proguanil, propranolol, pyrimethamine, quinine, rabeprazole, ranitidine, ritonavir, rofecoxib, selegiline, sertraline, tauromustine, temazepam, terbinafine, thalidomide, THC, tolbutamide, topiramate, valproic acid, venlafaxine, verapamil, voriconazole, warfarin, zidovudine, zotepine |
| CYP2D6 | Amines | Octopamine, synephrine, tryptamine, tyramine |
| | β-Carbolines | Harmaline, harmine, methoxydimethyltryptamine, pinoline |
| | Carcinogens | Butadiene, 17β-estradiol, NNK, safrole |
| | Components of garlic | Diallyl disulfide |
| | Fungicides | *o*-Phenylphenol |
| | Insect repellents | *N*,*N*-Diethyl-*m*-toluamide |
| | Pesticides | Chlorpyrifos, diazinon, methoxychlor, parathion |
| | Steroids | Progesterone |

| | | |
|---|---|---|
| | Drugs | Ajmaline, albendazole, alprenolol, amiflamine, aminopyrine, amitriptyline, amodiaquine, aprindine, aripiprazole, artemisinin, azelastine, barnidipine, benztropine, betaxolol, bisoprolol, brofaromine, bromoperidol, buflomedil, bufuralol, bunitrolol, bupranolol, capsaicin, captopril, carteolol, carvedilol, chloroquine, chlorpheniramine, chlorpromazine, cibenzoline, cinnarizine, citalopram, clemastine, clomipramine, clozapine, codeine, cyclobenzaprine, dapsone, debrisoquine, delavirdine, desglymidodrine, desipramine, desmethylcitalopram, dexfenfluramine, dextromethorphan, diazepam, diclofenac, diltiazem, diphenhydramine, diprafenone, dolasetron, donepezil, doxepin, doxylamine, duloxetine, efavirenz, eletriptan, encainide, ethylmorphine, etoricoxib, ezlopitant, fenfluramine, fentanyl, flecainide, flunarizine, fluoxetine, fluperlapine, fluphenazine, fluvastatin, fluvoxamine, galantamine, gallopamil, gepirone, guanoxan, haloperidol, hydrocodone, hydromorphone, hydroxyzine, ibogaine, imipramine, indinavir, indoramin, karenitecin, labetalol, lidocaine, lisuride, lopinavir, loratadine, lovastatin, maprotiline, MBDB, MDA, MDE, MDMA, meperidine, mequitazine, methadone, methamphetamine, methoxyamphetamine, methoxyphenamine, metoclopramide, metoprolol, mexiletine, mianserin, milameline, minaprine, mirtazapine, moclobemide, morphine, nebivolol, nefazodone, nelfinavir, nevirapine, norcodeine, norfluoxetine, nortriptyline, olanzapine, ondansetron, opipramol, oxycodone, oxyprenolol, paroxetine, pentazocine, perhexiline, perospirone, perphenazine, phenformin, pilocarpine, pimozide, pindolol, praziquantel, primaquine, procainamide, promethazine, propafenone, propoxyphene, propranolol, *N*-propylajmaline, pyrantel, quetiapine, quinidine, ranitidine, remoxipride, risperidone, ritonavir, rofecoxib, ropivacaine, selegiline, sertindole, sertraline, simvastatin, sparteine, sumatriptan, tamoxifen, tamsulosin, tauromustine, terguride, thebaine, thioridazine, timolol, tolterodine, tomoxetine, tramadol, trazodone, trifluperidol, trimipramine, tripelennamine, tropisetron, venlafaxine, yohimbine, zolpidem, zotepine, zuclopenthixol |
| CYP2E1 | Alkanes and related compounds | Acetaldehyde, acetonitrile, diethylether, ethanol, glycerol |
| | Amides | Dimethylformamide, methylformamide |
| | Aromatic hydrocarbons | Aniline, naphthylhydrazine, *p*-nitrophenol, phenol, pyridine, toluene, xylene |
| | Carcinogens | |
| | Amine derivatives | DEN, DMN, NNAL, NNK, NNN |
| | Halogenated hydrocarbons | $CCl_4$, chloroform, 1,4-dichlorobenzene, dichloromethane, 1,2-dichloropropane, ethylene dibromide, ethylene dichloride, methyl chloride, methylene chloride, tetrachloroethane, trichloroethane, trichloroethylene, vinyl bromide, vinyl chloride |
| | Polycyclic aromatics | Benzo[*a*]pyrene-7,8-diol, 1-methylpyrene |
| | Various carcinogens | Acrylonitrile, benzene, butadiene, ethyl carbamate, isoprene, methyleugenol, *p*-nitroanisole, safrole, styrene, vinyl carbamate |

**TABLE 10** ***Continued***

| CYP | Category of Substrates | Examples of Substrates[a] |
|---|---|---|
| | Chloroaromatics | 1,2-Dichlorobenzene, 1,2,4-trichlorobenzene, 1,2,3,5-tetrachlorobenzene, pentachlorobenzene |
| | Components of garlic | Diallyl disulfide, diallyl sulfide |
| | Coumarins | 7-Ethoxycoumarin |
| | Endogenous compounds | Acetone, $PGH_2$, phosphatidylcholine |
| | Fatty acids | Lauric acid, linoleic acid, oleic acid |
| | Flavonoids | Genistein |
| | Fluorinated hydrocarbons | Desflurane, enflurane, halothane, isoflurane, methoxyflurane, sevoflurane |
| | Indoles | Indole, 3-methylindole |
| | Insect repellents | *N*,*N*-Diethyl-*m*-toluamide |
| | Pesticides | Methoxychlor, parathion |
| | Drugs | Acetaminophen, albendazole, artesunate, bupropion, caffeine, capsaicin, chloroquine, chlorzoxazone, dacarbazine, dapsone, disulfiram, dorzolamide, eszopiclone, felbamate, fluoxetine, isoniazid, ondansetron, phenobarbital, primaquine, quinoline, ritonavir, salicylate, sulfadiazine, tamoxifen, tegafur, theobromine, theophylline, thiamylal, tolcapone, trimethadione, vesnarinone, zidovudine |
| CYP2J2 | Fatty acids | Arachidonic acid, linoleic acid |
| | Steroids | Testosterone |
| | Vitamins | 1α-Hydroxyvitamin $D_3$, vitamin $D_2$, vitamin $D_3$ |
| | Drugs | Astemizole, bufuralol, diclofenac, ebastine |
| CYP2R1 | Vitamins | 1α-Hydroxyvitamin $D_2$, 1α-hydroxyvitamin $D_3$, vitamin $D_2$, vitamin $D_3$ |

[a]Abbreviations: CPO, 5-{4-[3-(4-cyclohexyl-2-propylphenoxy)propoxy]phenyl}-1,3-oxazolidine-2,4-dione; DEN, diethylnitrosamine; DMN, dimethylnitrosamine; MBDB, *N*-methylbenzodioxazolylbutanamine; MDA, 3,4-methylenedioxyamphetamine; MDE, 3,4-methylenedioxyethylamphetamine; MDMA, 3,4-methylenedioxymethamphetamine; MeIQ, 2-amino-3,5-dimethylimidazo[4,5-*f*]quinoline; MOCA, 4,4 -methylene-bis(2-chloroaniline); NNAL, 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanol; NNK, 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone; NNN, nornitrosonicotine; $PGH_2$, prostaglandin $H_2$; PhIP, 2-amino-1-methyl-6-phenylimidazo[4,5-*b*]pyridine; THC, $\Delta^9$-tetrahydrocannabinol; thioTEPA, triethylenethiophosphoramide.

*Source*: Compiled from Refs. 40, 112, and 182–252. A detailed listing of substrates for CYP2 enzymes can be found in book chapters [2, 15, 50, 53, 55] and a review article [89].

at the amino acid level [253]. CYP2A6 is expressed primarily in liver. CYP2A7, which was originally named CYP2A4, is considered nonfunctional. CYP2A13 is expressed mainly in the nasal mucosa, followed by the trachea and lung [13]. Expression of CYP2A6 is subject to genetic polymorphism (see Table 4), with four allelic variants having been described. CYP2A6 is involved in the metabolism of several compounds found in tobacco. For example, CYP2A6 catalyzes the *C*-oxidation of nicotine, hydroxylation of cotinine at the 3′-position, hydroxylation of *N*-nitrosonornicotine at the 5′-position, and oxidation of 4-(methylnitrosoamino)-1-(3-pyridyl)-1-butanone and *N*-nitrosodiethylamine [253]. In terms of drug substrates, CYP2A6 preferentially catalyzes the hydroxylation of coumarin at the 7-position, which serves as a marker activity for this enzyme. Methoxypsoralen (methoxsalen), a structural analogue of coumarin, is a potent inhibitor of CYP2A6 [2].

There is only one CYP2B enzyme in humans, CYP2B6. A second CYP2B gene, CYP2B7, was identified as a splice variant of CYP2B6 but is not transcribed into protein [254]. CYP2B6 is expressed mostly in the liver [15, 254]. In a panel of 19 human liver microsome samples, CYP2B6 protein expression showed a 100-fold interindividual variability [255]. CYP2B6 shows sex and ethnic differences in expression and activity [256]. CYP2B6 is involved in the metabolism of relatively few drugs. It has been shown to catalyze the 4-hydroxylation of cyclophosphamide, and the *N*-demethylation of benzphetamine, (*S*)-mephenytoin, and diazepam [2]. Moreover, CYP2B6 catalyzes the *O*-dealkylation of 7-benzyloxyresorufin, methoxychlor, and 7-ethoxycoumarin [253]. However, the activities are not specific to CYP2B6 as other CYP enzymes such as CYP2C9 and CYP3A4 also contribute to these reactions. In rats, members of the CYP2B subfamily include CYP2B1, CYP2B2, and CYP2B3 [396]. Hepatic CYP2B1 and CYP2B2 are highly inducible in the rat following exposure to phenobarbital [15, 257]. Unlike the rat, little is known about induction of CYP2B6 in humans. In cultured human hepatocytes, CYP2B6 expression was increased following treatment with rifampicin, dexamethasone, and phenobarbital [255].

The CYP2C subfamily consists of four genes in humans, namely, CYP2C8, CYP2C9, CYP2C18, and CYP2C19. The four CYP2C enzymes are >80% identical in terms of amino acid sequence [15]. All four CYP2C enzymes are expressed primarily in liver (see Table 8), with CYP2C9 typically present at a higher level than CYP2C8, CYP2C18, or CYP2C19. CYP2C9 and CYP2C19 exhibit genetic polymorphism (see Table 4), with two allelic variants described for CYP2C9 and five allelic variants described for CYP2C19. CYP2C enzymes are involved in the metabolism of several important drugs. For example, CYP2C8 catalyzes the 6α-hydroxylation of taxol and the 4-hydroxylation of retinol and all-*trans* retinoic acid. CYP2C9 catalyzes tolbutamide hydroxylation, tienilic acid hydroxylation, phenytoin 4′-hydroxylation, diclofenac 4′-hydroxylation, and (*S*)-warfarin 7-hydroxylation [2]. CYP2C19 preferentially catalyzes the hydroxylation of (*S*)-mephenytoin at the 4′-position, which serves as a marker activity for this enzyme. No specific reactions have been identified for CYP2C18, which is expressed in trace amounts in liver. Sulfaphenazole is a potent selective inhibitor of CYP2C9 [2]. Inhibitors of CYP2C8 and CYP2C19 are known but they are not very specific.

There is a single functional member of the CYP2D subfamily in humans. CYP2D6 is expressed primarily in liver. Low levels have also been detected in extrahepatic tissues. CYP2D6 is a highly polymorphic enzyme. More than 60 allelic variants,

which affect the amount of expression of the enzyme and its catalytic activity, have been described (see Table 4). The polymorphisms affecting CYP2D6 have clinical relevance because this enzyme is involved in the metabolism of at least 100 different drugs (see Table 10). For example, CYP2D6 catalyzes the 4-hydroxylation of propranolol and debrisoquine, *O*-demethylation of dextromethorphan, 1′-hydroxylation of bufuralol, 5′-hydroxylation of propafenone, and hydroxylation of sparteine, all of which can be used as marker activities for this enzyme. Quinidine is a relatively selective inhibitor of CYP2D6 [2].

CYP2E1 is the only member of the CYP2E subfamily in humans. It is expressed constitutively in liver, lung, and kidney [2]. Relatively few drugs are metabolized by CYP2E1, but this enzyme catalyzes ethanol oxidation, *N*1- and *N*7-demethylation of caffeine, *N*-oxidation of acetaminophen, and chlorzoxazone 6-hydroxylation, which serves as a marker activity for CYP2E1. CYP2E1 can be induced by ethanol and isoniazid. 4-Methylpyrazole is a relatively selective inhibitor for CYP2E1 [2].

CYP2F and CYP2J are two minor subfamilies with only one functional gene in each. CYP2F1 is expressed primarily in human lung. It has been shown to catalyze the oxidation of ethoxycoumarin and propoxycoumarin, and of 4-ipomeanol and 3-methylindole, two lung toxicants [13]. CYP2J2 is expressed in liver and intestine. It has been shown to catalyze the hydroxylation of ebastine, a second-generation antihistamine [357].

Additional enzymes in the CYP2 family include CYP2S1, CYP2U1, and CYP2W1, but their function remains unknown, as substrates for these enzymes have not been identified [11].

### 4.4 CYP3 Family

CYP3A enzymes are the predominant CYP enzymes expressed in human liver and intestine [14, 154]. The human CYP3A subfamily consists of CYP3A4, CYP3A5, CYP3A7, and CYP3A43. CYP3A enzymes catalyze the biotransformation of a wide variety of xenobiotic compounds, as well as bile acids and steroids (Table 11). Indeed, CYP3A substrates represent diverse chemical structures including cyclosporin A, which is the largest CYP substrate known. The prominent role played by CYP3A enzymes in drug metabolism is highlighted by an estimate that CYP3A enzymes are responsible for the biotransformation of greater than 50% of all prescription medications [154].

CYP3A enzymes catalyze the *N*-dealkylation of alfentanil, benzphetamine, codeine, dextromethorphan, erythromycin, imipramine, and tamoxifen, hydroxylation of midazolam at the 1′-position, hydroxylation of alprazolam at the 4-position, and *N*-oxidation of dihydropyridines such as nifedipine [50]. A catalytic activity that is common to CYP3A enzymes in many mammalian species, and is often used as a marker activity, is the stereoselective hydroxylation of testosterone and cortisol at the 6β-position [15] (see Table 7). Inhibitors of CYP3A include the macrolide antibiotics, antifungal agents such as clotrimazole and ketoconazole, HIV protease inhibitors, particularly ritonavir, and gestodene [2].

CYP3A4 is the single most abundant CYP enzyme in human liver and small intestine [14, 154]. All human liver and small intestine samples appear to contain CYP3A4, but levels vary by 10-fold or more among individuals [14, 155]. In contrast, CYP3A5 is expressed at a lower level (approximately one-third or less) than

**TABLE 11 Substrates for Human CYP3A4**

| | Category of Substrates | Examples of Substrates[a] |
|---|---|---|
| CYP3A4 | Azole antifungals | Clotrimazole, fluconazole, itraconazole, ketoconazole, miconazole, voriconazole |
| | Bile acids | Chenodeoxycholic acid, cholic acid, deoxycholic acid, lithocholic acid, taurochenodeoxycholic acid |
| | Carcinogens | |
| | Mycotoxins | Aflatoxin $B_1$, aflatoxin $G_1$, ochratoxin A, sterigmatocystin |
| | Polycyclic aromatics | 6-Aminochrysene, benzo[*a*]pyrene, benzo[*a*]pyrene-7,8-diol, DDB[*a*]P, DDB[*b*]F, DDMBA, 1,6-dinitropyrene, 1-ethylpyrene, 1-methylpyrene, naphthalene, nitropyrene, phenanthrene |
| | Various carcinogens | 17β-Estradiol, MeIQ, methoxyaminoazobenzene, MOCA, NNK, NNN, senecionine, Tris-BP |
| | Chloroaromatics | 1,2-Dichlorobenzene, 1,2,4-trichlorobenzene, 1,2,3,5-tetrachlorobenzene, pentachlorobenzene |
| | Coumarins | BFC |
| | Endogenous compounds | 9,10-Epoxystearic acid, *N*-hydroxyarginine, leukotoxin, $PGH_2$ |
| | Ergot alkaloids | Bromocriptine, dihydroergotamine |
| | Flavonoids | Apigenin, 7,8-benzoflavone, genistein, kaempferol, naringenin, tamarixetin |
| | Fluorescent probes | Dibenzylfluorescein |
| | Fungicides | Furametpyr, *o*-phenylphenol |
| | Hormones | 1α,25-Dihydroxyvitamin $D_3$ |
| | Insect repellents | *N*,*N*-Diethyl-*m*-toluamide |
| | Macrolides | Clarithromycin, erythromycin, miocamycin, roxithromycin, sirolimus, tacrolimus, troleandomycin |
| | Mycotoxins, other toxins | Myristicin, territrems B and C |
| | Pesticides | Acetochlor, alachlor, aldrin, ametryne, atrazine, butachlor, carbofuran, chlorpyrifos, chlortoluron, diazinon, endosulfan, methoxychlor, parathion, terbutryne |
| | Retinoids | Retinal, all-*trans*-retinoic acid, 9-*cis*-retinoic acid |
| | Rifamycins and derivatives | Benzoxazinorifamycins, rifalazil, rifampicin |
| | Steroid hormones, steroids | Androstenedione, budesonide, corticosterone, cortisol, dehydroepiandrosterone, dexamethasone, DHEA, dutasteride, estrone, 17α-ethinylestradiol, finasteride, fluticasone, gestodene, levonorgestrel, lilopristone, medroxyprogesterone, mestranol, methandienone, methylprednisolone, mifepristone, norethisterone, onapristone, prednisolone, prednisone, pregnenolone, progesterone, testosterone |
| | Terpenes | Eucalyptol, limonene |

**TABLE 11** ***Continued***

| Category of Substrates | Examples of Substrates[a] |
|---|---|
| Drugs | Acetaminophen, adinazolam, albendazole, alfentanil, alfuzosin, alpidem, alprazolam, ambroxol, aminopyrine, amiodarone, amitriptyline, amlodipine, amprenavir, anastrozole, antipyrine, aprepitant, AQ4N, aripiprazole, arteether, artelinic acid, artemether, artemisinin, astemizole, atazanavir, atorvastatin, azelastine, azimilide, barnidipine, benzphetamine, bepridil, bezafibrate, bicalutamide, bisoprolol, bosentan, brinzolamide, bromoperidol, brotizolam, bupivacaine, buprenorphine, bupropion, buspirone, busulfan, capravirine, capsaicin, carbamazepine, carvedilol, cerivastatin, chloroquine, chlorpromazine, cibenzoline, cimetidine, ciprofibrate, cisapride, citalopram, clindamycin, clobazam, clofibrate, clomethiazole, clomipramine, clonazepam, clopidogrel, clozapine, cocaine, codeine, colchicine, cyclobenzaprine, cyclophosphamide, cyclosporine, dantrolene, dapsone, delavirdine, desipramine, desmethyladinazolam, dextromethorphan, diazepam, diclofenac, digitoxin, dihydralazine, dihydrocodeine, diltiazem, disopyramide, docetaxel, dolasetron, donepezil, dorzolamide, doxepin, doxorubicin, ebastine, ecabapide, efavirenz, eletriptan, ellipticine, enalapril, eplerenone, esomeprazole, eszopiclone, ethosuximide, ethylmorphine, etizolam, etoperidone, etoposide, etoricoxib, exemestane, ezlopitant, felbamate, felodipine, fenofibrate, fentanyl, flosequinan, flunitrazepam, flutamide, fluvastatin, galantamine, gallopamil, ganaxolone, gefitinib, gemfibrozil, gepirone, germander, glibenclamide, glyceryltrinitrate, granisetron, halofantrine, haloperidol, hydrocodone, hydromorphone, ifosfamide, imatinib, imipramine, imiquimod, indinavir, ipomeanol, irinotecan, isoniazid, isosorbide dinitrate, isradipine, ivermectin, karenitecin, ketamine, ketobemidone, lansoprazole, laquinimod, letrozole, lidocaine, lisofylline, lisuride, lonafarnib, loperamide, lopinavir, loratadine, losartan, lovastatin, lumefantrine, MBDB, MDA, MDE, MDMA, medifoxamine, mefloquine, meloxicam, methadone, mexazolam, mianserin, mibefradil, midazolam, mifepristone, mirtazapine, MMDX, modafinil, montelukast, mosapride, mycophenolic acid, nateglinide, navelbine, nefazodone, nefiracetam, nelfinavir, nevirapine, nicardipine, nifedipine, niludipine, nilvaldipine, nimodipine, nisoldipine, nitrendipine, noralfentanil, |

nordiazepam, nortriptyline, norverapamil, omeprazole, ondansetron, orphenadrine, oxodipine, oxybutynin, paclitaxel, pantoprazole, perazine, perospirone, perphenazine, phencyclidine, phenprocoumon, pimozide, pioglitazone, piroxicam, pradefovir, pranidipine, pravastatin, praziquantel, primaquine, propafenone, proguanil, pyrilamine, quetiapine, quinidine, quinine, rabeprazole, raloxifene, ranolazine, reboxetine, repaglinide, resiquimod, risperidone, ritonavir, rofecoxib, ropinirole, ropivacaine, roquinimex, rupatadine, salmeterol, saquinavir, selegiline, seratrodast, sertindole, sertraline, sibutramine, sildenafil, simvastatin, sterinodole, sufentanil, sulfadiazine, sulfamethoxazole, sulfamidine, tafenoquine, tamoxifen, tamsulosin, targretin, tasosartan, tauromustine, tazofelone, temazepam, teniposide, tentoxin, terbinafine, terfenadine, terguride, THC, theophylline, thiamylal, thioTEPA, tiagabine, ticlopidine, tinidazole, tipranavir, tirilazad, tolmetin, tolterodine, toremifene, trazodone, triazolam, triazoledione, trimethadione, trimethoprim, trofosfamide, troglitazone, tropisetron, valdecoxib, valspodar, vanoxerine, vardenafil, venlafaxine, verapamil, vesnarinone, vinblastine, vincristine, vindesine, vinorelbine, warfarin, yohimbine, zafirlukast, zaleplon, zatosetron, zidovudine, zileuton, ziprasidone, zopiclone, zolpidem, zomepirac, zonisamide, zotepine

---

[a]Abbreviations: AQ4N, 1,4-bis{[2-(dimethylamino-*N*-oxide)ethyl]amino}-5,8-dihydroxyanthracene-9,10-dione; BFC, 7-benzyloxy-4-trifluoromethylcoumarin; DDB[*a*]P, 7,8-dihydroxy-7,9-dihydrobenzo[*a*]pyrene; DDB[*b*]F, 9,10-dihydroxy-9,10-dihydrobenzo[*b*]fluoranthene; DDMBA, 3,4-dihydroxy-3,4-dihydro-7,12-dimethylbenz[*a*]anthracene; DHEA, dehydroepiandrosterone; MBDB, *N*-methylbenzodioxazolylbutanamine; MDA, 3,4-methylenedioxyamphetamine; MDE, 3,4-methylenedioxyethylamphetamine; MDMA, 3,4-methylenedioxymethamphetamine; MeIQ, 2-amino-3,5-dimethylimidazo[4,5-*f*]quinoline; MMDX, methoxymorpholinodoxorubicin; MOCA, 4,4′-methylene-bis(2-chloroaniline); NNK, 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone; NNN, nornitrosonicotine; $PGH_2$, prostaglandin $H_2$; THC, $\Delta^9$-tetrahydrocannabinol; thioTEPA, triethylenethiophosphoramide; Tris-BP, tris-(2,3-dibromopropyl)phosphate.

*Source*: Compiled from Refs. 59, 118, 199, 212, 221, 238, 244, 247, and 258–344. A comprehensive listing of substrates for CYP3A enzymes can be found in book chapters [2, 15, 50, 53, 55], a review article [89], and websites [397, 398].

CYP3A4 and in only 10–30% of human liver and intestine samples [14, 155]. CYP3A5 has also been detected in human kidney [2]. The lack of CYP3A5 expression in most humans is thought to be due to a splice site mutation [345]. CYP3A7 is considered to be a fetal enzyme and accounts for 30–50% of total CYP content in fetal liver [345]. CYP3A7 is now recognized to be expressed in some adult livers [345]. CYP3A43 mRNA is expressed at low levels in the liver and at relatively high levels in the testis and prostate. In earlier studies, a fifth CYP3A enzyme, CYP3A3, was identified. This enzyme showed strong homology to CYP3A4. It is now believed that CYP3A3 is in fact an allelic variant of CYP3A4 [345]. Both CYP3A5 and CYP3A7 have similar substrate specificities as CYP3A4. The substrate specificity of CYP3A43 is unknown since only low levels of expression have been achieved in recombinant systems [345]. The human CYP3A enzymes can be induced by treatment with drugs such as phenobarbital, rifampicin, phenytoin, and troglitazone [50], and this induction can result in clinically relevant drug interactions.

In rats, the CYP3A subfamily includes CYP3A1, CYP3A2, CYP3A9, CYP3A18, and CYP3A62 [346, 396]. A sixth rat CYP3A enzyme, CYP3A23, has also been reported in the literature, but this is now believed to be an allelic variant of CYP3A1 [346, 347]. CYP3A2, CYP3A9, CYP3A18, and CYP3A62 display age- and sex-specific regulation and induction [348, 349]. For example, CYP3A2 is constitutively expressed in hepatic microsomes from immature rats of both sexes and mature male rats but not in adult female rats [348]. CYP3A18 is expressed predominantly in adult male rats, whereas CYP3A9 is expressed at higher levels in adult female rats than in male rats [350]. Dexamethasone treatment increases hepatic microsomal levels of CYP3A2 in immature rats and adult male but not adult female rats [348]. CYP3A1 is undetectable in hepatic microsomes from untreated rats of either sex but following administration of dexamethasone, hepatic levels of CYP3A1 are increased dramatically [348]. CYP3A1 can account for 30–37% of total CYP content in hepatic microsomes prepared from dexamethasone-treated immature and mature rats of both sexes. Expression of hepatic CYP3A9, CYP3A18, and CYP3A62 enzymes is also increased by administration of dexamethasone [346, 348]. Other inducers of rat CYP3A enzymes include phenobarbital, rifampicin, isosafrole, and pregnenolone 16α-carbonitrile [348].

### 4.5 CYP4 Family

Members of the CYP4 family generally metabolize endogenous fatty acid substrates such as lauric, palmitic, and arachidonic acids and their derivatives including prostaglandins, thromboxane, prostacyclin, and leukotrienes (Table 12). CYP4 enzymes play a minor role in the metabolism of drugs and other xenobiotics.

The CYP4A subfamily consists of CYP4A9 and CYP4A11. CYP4A11 is expressed in human liver and kidney. The expression and function of CYP4A9 is not known at present. CYP4A11 preferentially catalyzes the ω (terminal carbon) hydroxylation of fatty acids, although it can also catalyze hydroxylation at the ω-1 (penultimate carbon) position [2]. Hence the characteristic activity of CYP4A11 is hydroxylation of lauric acid at the 12-(ω) position (see Table 7). CYP4A enzymes in rats are inducible by peroxisomal proliferators such as clofibrate. Induction of CYP4A11 by clofibrate has also been reported in primary cultures of human hepatocytes [360].

**TABLE 12 Substrates for Human CYP4 Enzymes**

| CYP | Category of Substrates | Examples of Substrates[a] |
|---|---|---|
| CYP4A11 | Fatty acids, leukotrienes | Arachidonic acid, lauric acid, $LTB_4$, linoleic acid, palmitic acid, 9,10-epoxystearic acid, isoleukotoxin, leukotoxin |
| | Eicosanoic acids | 15-HETE |
| | Prostaglandins | 9,11-Epoxymethano-$PGH_2$, 11,9-epoxymethano-$PGH_2$, 9,11-diazo-15-deoxy-$PGH_2$ |
| | Drugs | Artesunate, pentamidine |
| CYP4F2 | Aromatic hydrocarbons | 7-Ethoxycoumarin, *p*-nitroanisole |
| | Fatty acids, leukotrienes, lipoxins | Arachidonic acid, hydroxystearic acids, $LTB_4$, 6-*trans*-$LTB_4$, oleic acid, lipoxin $A_4$, 9,10-epoxystearic acid, isoleukotoxin, leukotoxin |
| | Eicosanoic acids | 5,6-EET, 8,9-EET, 11,12-EET, 14,15-EET, 5-HETE, 8-HETE, 9-HETE, 11-HETE, 12-HETE, 15-HETE |
| CYP4F3B | Fatty acids, leukotrienes | Arachidonic acid, 9,10-dihydroxyoctadec-12-enoic acid, 12,13-dihydroxyoctadec-9-enoic acid, 9,10-dihydroxystearic acid, $LTB_4$, 9,10-epoxystearic acid, isoleukotoxin, leukotoxin |
| | Eicosanoic acids | 5,6-DHET, 8,9-DHET, 11,12- DHET, 14,15-DHET, 5,6-EET, 8,9-EET, 11,12-EET, 14,15-EET |
| CYP4F11 | Fatty acids, leukotrienes | Arachidonic acid, $LTB_4$ |
| | Eicosanoic acids | 8-HETE |
| | Drugs | Benzphetamine, chlorpromazine, erythromycin, ethylmorphine, imipramine, theophylline, verapamil |
| CYP4F12 | Fatty acids, leukotrienes | Arachidonic acid, $LTB_4$ |
| | Prostaglandins | 11,9-Epoxymethano-$PGH_2$, 9,11-diazo-15-deoxy-$PGH_2$ |
| | Drugs | Ebastine |

[a]Abbreviations: DHET, dihydroxyeicosatrienoic acid; EET, epoxyeicosatrienoic acid; HETE, hydroxyeicosatetraenoic acid; $LTB_4$, leukotriene $B_4$; $PGH_2$, prostaglandin $H_2$.

*Source*: Compiled from Refs. 15, 29, 55, 89, and 351–361.

The human CYP4F subfamily consists of five genes, CYP4F2, CYP4F3, CYP4F8, CYP4F11, and CYP4F12 [358, 359]. CYP4F enzymes have been detected at the level of mRNA in numerous tissues in humans including leukocytes, liver, lung, kidney, intestine, heart, brain, and skin [358, 359]. CYP4F enzymes are involved in the oxidation of fatty acids, leukotrienes, and prostaglandins (Table 12).

### 4.6 Regulation of CYP Enzyme Expression

The expression of CYP genes in many species is highly regulated by age, hormonal factors including sex steroids, growth hormone, and thyroid hormones, diet and nutritional status, and disease states. Indeed, regulation of hepatic CYP enzymes is complex. Although the mechanisms and consequences of induction of CYP enzymes have been studied intensively, relatively little is known about the mechanisms by which CYP enzymes are regulated by physiological factors. In humans, CYP enzyme

expression appears to be regulated mainly via induction by xenobiotics (i.e., drugs, environmental compounds including those present in cigarette smoke, dietary components, and ethanol) and possibly by hormones. CYP enzyme induction has consequences in terms of drug pharmacokinetics, accumulation, and toxicity.

Induction of CYP enzymes usually occurs at the level of transcription and results in *de novo* protein synthesis. There are several mechanisms that involve interactions with regulatory elements in the relevant CYP gene, by which CYP genes can be induced. Most commonly, inducers bind to a receptor protein, which then binds to regulatory elements in the DNA, and activates gene transcription and translation. Alternatively, the inducer can interact, either directly or indirectly, with a repressor protein that would normally be bound to the operator region of the CYP gene and repress its expression. Binding of the inducer would release the repressor and result in increased transcription of the structural region of the gene [15]. Specific xenobiotic receptor proteins have now been identified as being involved in gene induction by four distinct mechanisms [369, 370]. Inducing agents can bind to four different xenobiotic receptors, namely, the aryl hydrocarbon receptor (AhR), the constitutive androstane receptor (CAR), the pregnane X receptor (PXR), and the peroxisomal proliferator-activated receptor (PPAR). Typical inducers of human CYP enzymes are presented in Table 13.

The Ah receptor is a ligand-activated nuclear transcription factor that controls the expression of a battery of genes including CYP1A1, CYP1A2, and CYP1B1 [371, 372]. The AhR is ubiquitously expressed in most organs and is conserved across many species including mammals and invertebrates [181]. Once activated, the AhR translocates to the nucleus, where it forms a heterodimer with the aryl hydrocarbon nuclear translocator (ARNT) [373]. This heterodimer complex then binds to a

**TABLE 13 Inducers of Human CYP Enzymes**

| CYP | Inducers[a] |
|---|---|
| CYP1A1 | Chargrilled meat, omeprazole, PCBs, polycyclic aromatics (e.g., DMBA, MC), TCDD |
| CYP1A2 | Chargrilled meat, cruciferous vegetables (e.g., broccoli, Brussels sprouts), hyperforin, omeprazole, PCBs, polycyclic aromatics, TCDD |
| CYP1B1 | PCBs, polycyclic aromatics, TCDD |
| CYP2A6 | Phenobarbital |
| CYP2B6 | Nevirapine, phenobarbital, dexamethasone, troglitazone |
| CYP2C8 | Rifampicin |
| CYP2C9 | Hyperforin, phenobarbital, rifampicin |
| CYP2C19 | Rifampicin, ritonavir |
| CYP2D6 | Unknown |
| CYP2E1 | Ethanol, isoniazid |
| CYP3A4 | Bosentan, carbamazepine, cortisol, dexamethasone, efavirenz, felbamate, hyperforin, nevirapine, phenobarbital, phenytoin, primidone, rifampicin, rifapentin, sulfadimidine, topiramate, troglitazone |
| CYP4A11 | Clofibrate |

[a]Abbreviations: DMBA, 7,12-dimethylbenz[*a*]anthracene; MC, 3-methylcholanthrene; PCBs, polychlorinated biphenyls; TCDD, 2,3,7,8-tetrachlorodibenzo-*p*-dioxin.

*Source*: Compiled from Refs. 19, 26, 158, and 360–370. A comprehensive listing of CYP inducers can be found in book chapters [2, 15, 50, 55], a review article [89], and a website [398].

xenobiotic response element in the promoter region upstream of the CYP1A1 and CYP1A2 genes. Binding of the heterodimer complex to the xenobiotic response element results in recruitment of coregulatory proteins and relaxation of chromatin structure, which promotes transcription [181]. One of the most potent inducers of CYP1A expression is the highly toxic compound, TCDD. Other xenobiotic AhR ligands include 3-methylcholanthrene, indolo[2,3-*b*]carbazole, and coplanar PCBs [181] (Table 14). Endogenous AhR ligands include tryptamine, bilirubin, and indirubin [181].

Phenobarbital is a prototype agent for a variety of structurally unrelated inducers that elicit induction of CYP2B enzymes in a number of mammalian species. Induction of the CYP2B enzymes occurs mainly through the activation of CAR, a novel nuclear receptor. Following exposure to phenobarbital-type inducers, CAR translocates from the cytoplasm to the nucleus, where it binds and activates a phenobarbital-responsive element upstream of the CYP2B genes [374]. CAR, which is highly expressed in liver, is known to dimerize with the retinoid X receptor (RXR) and can constitutively activate its target genes even in the absence of inducers. Normally, endogenous androstane-like steroids bind to CAR and maintain it in an inactive state. The binding of phenobarbital-type inducers interferes with the binding of the inhibitory steroids so that receptor activity is derepressed and gene transcrip-

**TABLE 14 Receptors and Xenobiotic Activators/Ligands of Human CYP Genes**

| Receptors[a] | Xenobiotic Activators/Ligands[a] | Regulated CYP Genes |
|---|---|---|
| AhR | Guanabenz, omeprazole, PCBs, polycyclic aromatics (e.g., benzo[*a*]pyrene, MC), TCDD, TCDF, thiabendazole | CYP1A1, CYP1A2, CYP1B1 |
| CAR | CITCO, hyperforin, methoxychlor, phenobarbital, phenytoin, rifampicin | CYP2A6, CYP2B6, CYP2C8, CYP2C9, CYP2C19, CYP3A4 |
| PPARα | Bezafibrate, clofibrate, fenofibrate, fenoprofen, flufenamic acid, gemfibrozil, herbicides, ibuprofen, indomethacin, phthalates, Wy-14,643 | CYP4A11 |
| PXR | Amprenavir, avasimibe, bosentan, carbamazepine, chlorpyrifos, CITCO, clotrimazole, cypermethrin, dexamethasone, DDT, dieldrin, efavirenz, endosulfan, endrin, etoposide, glibenclamide, hyperforin, lovastatin, mifepristone, methoxychlor, nelfinavir, nifedipine, nonylphenol, paclitaxel, phenobarbital, phenytoin, rifabutin, rifampicin, ritonavir, simvastatin, tamoxifen, TCPOBOP, topiramate, triclosan, troglitazone | CYP2A6, CYP2B6, CYP2C8, CYP2C9, CYP2C19, CYP3A4 |

[a]Abbreviations: AhR, aryl hydrocarbon receptor; CAR, constitutive androstane receptor; CITCO, 6-(4-chlorophenyl)imidazo[2,1-*b*][1,3]thiazole-5-carbaldehyde *O*-(3,4-dichlorobenzyl)oxime; DDT, dichlorodiphenyltrichloroethane; MC, 3-methylcholanthrene; PCBs, polychlorinated biphenyls; PPARα, peroxisome proliferator-activated receptor α; PXR, pregnane X receptor; TCDD, 2,3,7,8-tetrachlorodibenzo-*p*-dioxin; TCDF, 2,3,7,8-tetrachlorodibenzofuran; TCPOBOP, 1,4-bis[2-(3,5-dichloropyridyloxy)]benzene; Wy-14,643, [4-chloro-6-(2,3-xylidino)pyrimidynylthio]acetic acid.

*Source*: Compiled from Refs. 23, 26, 94, 154, 181, 366, and 369–395.

tion can proceed [369, 370]. CYP2B enzymes can be induced by a variety of antimicrobials, barbiturates, and pesticides. Some examples include clotrimazole, rifampicin, phenytoin, methoxychlor, and dichlorodiphenyltrichloroethane (DDT) [374, 375].

Another novel nuclear receptor, PXR (also known as PAR and SXR) [376], mediates induction of CYP3A genes in response to glucocorticoids such as dexamethasone and drugs such as rifampicin [28] (Table 14). CYP3A4 and CYP3A5 enzymes are inducible following administration of rifampicin although CYP3A5 is induced to a lesser extent than CYP3A4 [345]. Induction of the CYP3A subfamily occurs mainly through activation of PXR. PXR binds as a heterodimer with RXR to xenobiotic response elements in the proximal promoter regions of the CYP3A genes [154]. Induction of the CYP3A subfamily shows distinct species-specific differences. For example, rifampicin is a potent inducer of human CYP3A4 but is a weak inducer of rat CYP3A2 [154]. Studies have suggested that CAR may also be involved in transcriptional activation of the CYP3A genes since CAR can bind to xenobiotic response elements in the promoter region of the CYP3A genes [154]. Experimental evidence indicates that the glucocorticoid receptor may participate in the induction of CYP3A4, but the molecular mechanism is not clear [377].

The peroxisomal proliferator-activated receptor (PPAR) is highly expressed in liver and mediates peroxisome proliferation and induction of CYP4A enzymes [369] (Table 14).

## 5 CONCLUSION

Insight into the biotransformation of drugs by CYP and other drug-metabolizing enzymes is a prerequisite for the evaluation of drug safety and risk assessment. Moreover, comprehensive knowledge of the regulation and activity of CYP enzymes and an awareness of possible drug interactions is needed to avoid possible adverse effects associated with the common practice of prescribing several drugs simultaneously.

There are still many gaps in our knowledge of CYP enzymes. There is a practical need for better and more specific reagents to measure CYP enzyme activity and CYP enzyme levels in various tissues. For example, the ready availability of highly specific inhibitory antibodies and potent specific chemical inhibitors for each CYP enzyme would be highly desirable. This would allow investigators to identify the individual CYP enzymes involved in the metabolism of a particular drug and to quantify the enzymes in both hepatic and extrahepatic tissues. Also needed are highly selective and safe *in vivo* probes of CYP activity to allow investigators to estimate CYP levels in patients without the need to take liver biopsy samples. The large interindividual variability in CYP enzyme levels complicates drug therapy and efficacy. Expression profiling studies with larger banks of human tissues and human subjects are needed to discover the extent of variation of individual CYP protein levels, to delineate the causes of the variation, and to map distribution patterns of CYP enzymes within a population.

Recent advances are expanding our knowledge of the CYP enzymes and providing therapeutic benefits. For example, the development of CYP knockout mouse models is enabling investigators to explore the role and importance of CYP enzymes

in normal physiology. Pharmacogenomic testing is also currently receiving much attention because of the increasing awareness of the benefits of incorporating pharmacogenomics into the selection of drug regimens. The selection of proper drug regimen and dose based on the patient's individual genetic makeup could eliminate the unpredictable response of drug treatment, which may arise because of genetic polymorphisms that affect drug metabolism, clearance, and tolerance. Lastly, computational models can be used to predict the metabolism of specific drugs by CYP enzymes, to "design out" unfavorable drug interactions, and to help guide the drug design process.

## ACKNOWLEDGMENTS

Research in the authors' laboratory has been supported by grants from the Natural Sciences and Engineering Research Council of Canada, the Canadian Institutes of Health Research, and the Toxic Substances Research Initiative (Health Canada). The authors thank Dr. Ted Lakowski and Mr. Patrick Edwards for their highly valued input in the preparation of this chapter.

## REFERENCES

1. Wrighton SA, Stevens JC. The human hepatic cytochromes P450 involved in drug metabolism. *Crit Rev Toxicol* 1992;22:1–21.
2. Parkinson A. Biotransformation of xenobiotics. In Klaasen CD, Ed. *Casarett and Doull's Toxicology: The Basic Science of Poisons*, 6th ed. New York: McGraw Hill; 2001, pp 133–224.
3. Peterson JA, Graham-Lorence SE. Bacterial P450s: structural similarities and functional differences. In *Ortizde Montellano PR, Ed. Cytochrome P450: Structure, Mechanism, and Biochemistry*, 2nd ed. New York: Plenum Press; 1995, pp 151–180.
4. von Wachenfeldt C, Johnson EF. Structures of eukaryotic cytochrome P450 enzymes. In *Ortiz de Montellano PR, Ed. Cytochrome P450: Structure, Mechanism, and Biochemstry*, 2nd ed. New York: Plenum Press; 1995, pp 183–223.
5. Axelrod J. The enzymic demethylation of ephedrine. *J Pharmacol Exp Ther* 1955;114: 430–438.
6. Brodie B, Axelrod J, Cooper JR, Gaudette L, LaDu BN, Mitoma C, Udenfriend S. Detoxification of drugs and other foreign compounds by liver microsomes. *Science* 1955;121:603–604.
7. Klingenberg M. Pigments of rat liver microsomes. *Arch Biochem Biophys* 1958;75: 376–386.
8. Garfinkel D. Studies on pig liver microsomes. I. Enzymic and pigment composition of different microsomal fractions. *Arch Biochem Biophys* 1958;77:493–509.
9. Omura T, Sato R. The carbon monoxide-binding pigment of liver microsomes. *J Biol Chem* 1964;239:2370–2378.
10. Bandiera SM. Cytochrome P450 enzymes as biomarkers of PCB exposure and modulators of toxicity. In Robertson LW, Hansen LG, Eds. *PCBs: Recent Advances in Environmental Toxicology and Health Effects*. Lexington: University Press of Kentucky, 2001, pp 185–192.

11. Guengerich FP, Wu ZL, Bartleson CJ. Function of human cytochrome P450s: characterization of the orphans. *Biochem Biophys Res Commun* 2005;338:465–469.
12. Nelson DR, Kamataki T, Waxman DJ, Guengerich FP, Estabrook RW, Feyereisen R, Gonzalez FJ, Coon MJ, Gunsalus IC, Gotoh O, Okuda K, Nebert DW. The cytochrome P450 superfamily: update on new sequences, gene mapping, accession numbers, and nomenclature. *Pharmacogenetics* 1996;6:1–42.
13. Ding X, Kaminsky LS. Human extrahepatic cytochromes P450: function in xenobiotic metabolism and tissue-selective chemical toxicity in the respiratory and gastrointestinal tracts. *Annu Rev Pharmacol Toxicol* 2003;43:149–173.
14. Paine MF, Hart HL, Ludington SS, Haining RL, Rettie AE, Zeldin DC. The human intestinal cytochrome P450 "pie." *Drug Metab Dispos* 2006;34:880–886.
15. Lewis DFV. *Cytochromes P450: Structure, Function and Mechanism*. London: Taylor and Francis; 1996.
16. Bhamre S, Anandatheerathavarada HK, Shankar SK, Boyd MR, Ravindranath V. Purification of multiple forms of cytochrome P450 from a human brain and reconstitution of catalytic activities. *Arch Biochem Biophys* 1993;301:251–255.
17. Hellmold H, Rylander T, Magnusson M, Reihnér E, Warner M, Gustafsson J-Å. Characterization of cytochrome P450 enzymes in human breast tissue from reduction mammaplasties. *J Clin Endocrinol Metab* 1998;83:886–895.
18. Mason JI, Estabrook RW, Purvis JL. Testicular cytochrome P450 and iron–sulfur protein as related to steroid metabolism. *Ann NY Acad Sci* 1973;212:406–419.
19. Lee IP, Suzuki K, Mukhtar H, Bend JR. Hormonal regulation of cytochrome P-450-dependent monooxygenase activity and epoxide-metabolizing enzyme activities in testis of hypophysectomized rats. *Cancer Res* 1980;40:2486–2492.
20. Watanabe M, Abe T. Aryl hydrocarbon hydroxylase and its components in human lung microsomes. *Gann* 1981;72:806–810.
21. Denisov IG, Markis TM, Sligar SG, Schlichting I. Structure and chemistry of cytochrome P450. *Chem Rev* 2005;105:2254–2277.
22. Nebert DW, Adesnik M, Coon MJ, Estabrook RW, Gonzalez FJ, Guengerich FP, Gunsalus IC, Johnson EF, Kemper B, Levin W, Phillips IR, Satoh R, Waterman MR. The P450 gene superfamily: recommended nomenclature. *DNA* 1987;6:1–11.
23. Honkakoski P, Negishi M. Regulation of cytochrome P450 (CYP) genes by nuclear receptors. *Biochem J* 2000;347:321–337.
24. Okey AB. Enzyme induction in the cytochrome P-450 system. *Pharmacol Ther* 1990;45:241–298.
25. Miners JO, McKinnon RA. CYP1A. In Levy RH, Thummel KE, Trager WF, Hansten PD, Eichelbaum M, Eds. *Metabolic Drug Interactions*. Philadelphia: Lippincott Williams & Wilkins; 2000, pp 61–73.
26. Nebert DW, Dalton TP, Okey AB, Gonzalez FJ. Role of aryl hydrocarbon receptor-mediated induction of the CYP1 enzymes in environmental toxicity and cancer. *J Biol Chem* 2004;279:23847–23850.
27. Ryan DE, Levin W. Purification and characterization of hepatic microsomal cytochrome P450. *Pharmacol Ther* 1990;45:153–239.
28. Schuetz EG, Brimer C, Schuetz JD. Environmental xenobiotics and the antihormones cyproterone acetate and spironolactone use the nuclear hormone pregnenolone X receptor to activate the *CYP3A23* hormone response element. *Mol Pharmacol* 1998;54:1113–1117.

29. Hoch U, Zhang Z, Kroetz DL, Ortiz de Montellano PR. Structural determination of the substrate specificities and regioselectivities of the rat and human fatty acid ω-hydroxylases. *Arch Biochem Biophys* 2000;373:63–71.
30. Imaoka S, Ogawa H, Kimura S, Gonzalez FJ. Complete cDNA sequence and cDNA-directed expression of CYP4A11, a fatty acid omega-hydroxylase expressed in human kidney. *DNA Cell Biol* 1993;12:893–899.
31. Kharasch ED, Hankins DC, Thummel KE. Human kidney methoxyflurane and sevoflurane metabolism: intrarenal fluoride production as a possible mechanism of methoxyflurane nephrotoxicity. *Anesthesiology* 1995;82:689–699.
32. Hakkola J, Pasanen M, Hukkanen J, Pelkonen O, Mäenpää J, Edwards RJ, Boobis AR, Raunio H. Expression of xenobiotic-metabolizing cytochrome P450 forms in human full-term placenta. *Biochem Pharmacol* 1996;51:403–411.
33. Huang Z, Fasco MJ, Figge HL, Keyomarsi K, Kaminsky LS. Expression of cytochromes P450 in human breast tissue and tumors. *Drug Metab Dispos* 1996;24:899–905.
34. Zheng Y-M, Fisher MB, Yokotani N, Fujii-Kuriyama Y, Rettie AE. Identification of a meander region praline critical for heme binding to cytochrome P450: implications for the catalytic function of human CYP4B1. *Biochemistry* 1998;37:12847–12851.
35. Gervot L, Rochat B, Gautier JC, Bohnenstengel F, Kroemer H, de Berardinis V, Martin H, Beaune P, de Waziers I. Human CYP2B6: expression, inducibility and catalytic activities. *Pharmacogenetics* 1999;9:295–306.
36. Rylander T, Neve EPA, Ingelman-Sundberg M, Oscarson M. Identification and tissue distribution of the novel human cytochrome P450 2S1 (CYP2S1). *Biochem Biophys Res Commun* 2001;281:529–535.
37. Smith G, Wolf CR, Deeni YY, Dawe RS, Evans AT, Comrie MM, Ferguson J, Ibbotson SH. Cutaneous expression of cytochrome P450 CYP2S1: individuality in regulation by therapeutic agents for psoriasis and other skin diseases. *Lancet* 2003;361:1336–1343.
38. Miksys S, Tyndale RF. The unique regulation of brain cytochrome P450 2 (CYP2) family enzymes by drugs and genetics. *Drug Metab Rev* 2004;36:313–333.
39. Swanson HI. Cytochrome P450 expression in human keratinocytes: an aryl hydrocarbon receptor perspective. *Chem-Biol Interact* 2004;149:69–79.
40. Kinobe RT, Parkinson OT, Mitchell DJ, Gillam EMJ. P450 2C18 catalyzes the metabolic bioactivation of phenytoin. *Chem Res Toxicol* 2005;18:1868–1875.
41. Savas U, Hsu M-H, Griffin KJ, Bell DR, Johnson EF. Conditional regulation of the human CYP4X1 and CYP4Z1 genes. *Arch Biochem Biophys* 2005;436:377–385.
42. Chinta SJ, Kommaddi RP, Turman CM, Strobel HW, Ravindranath V. Constitutive expression and localization of cytochrome P-450 1A1 in rat and human brain: presence of a splice variant form in human brain. *J Neurochem* 2005;93:724–736.
43. Karlgren M, Miura S-I, Ingelman-Sundberg M. Novel extrahepatic cytochrome P450s. *Toxicol Appl Pharmacol* 2005;207:S57–S61.
44. Raimundo S, Fischer JU, Eichelbaum M, Griese, E-U, Schwab M, Zanger UM. Elucidation of the genetic basis of the common metaboliser phenotype for drug oxidation by CYP2D6. *Pharmacogenetics* 2000;20:577–581.
45. Brockmöller J, Kirchheiner J, Meisel C, Roots I. Pharmacogenetic diagnostics of cytochrome P450 polymorphisms in clinical drug development and in drug treatment. *Pharmacogenomics* 2000;1:125–152.
46. Zanger U, Eichelbaum M, CYP2D6. In Levy RH, Thummel KE, Tager WF, Hansten PD, Eichebaum M, Eds. *Metabolic Drug Interactions*. Philadelphia: Lippincott Williams & Wilkins; 2000, pp 87–94.

47. Guengerich FP, Liebler DC. Enzymatic activation of chemicals to toxic metabolites. *CRC Crit Rev Toxicol* 1985;14:259–262.
48. Conney AH. Induction of microsomal enzymes by foreign chemicals and carcinogenesis by polycyclic aromatic hydrocarbons: G.H.A. Clowes Memorial Lecture. *Cancer Res* 1982;42:4875–4917.
49. Guengerich FP. Reactions and significance of cytochrome P450 enzymes. *J Biol Chem* 1991;266:10019–10022.
50. Williams DA. Drug metabolism. In Williams DA, Lemke TL, Eds. *Foye's Principles of Medicinal Chemistry*, 5th ed. Baltimore: Lippincott Williams & Wilkins; 2002, pp 174–233.
51. Guengerich FP, Shimada T. Oxidation of toxic and carcinogenic chemicals by human cytochrome P-450 enzymes. *Chem Res Toxicol* 1991;4:391–407.
52. Gonzalez FJ, Gelboin HV. Role of human cytochromes P450 in the metabolic activation of chemical carcinogens and toxins. *Drug Metab Rev* 1994;26:165–183.
53. Guengerich FP, Human cytochrome P450 enzymes. In Ortiz de Montellano PR, Ed. *Cytochrome P450: Structure, Mechanism, and Biochemistry*, 2nd ed. New York: Plenum Press; 1995, pp 473–535.
54. Low LK, Metabolic changes of drugs and related organic compounds. In Delgado JN, Remers WA, Eds. *Textbook of Organic Medicinal and Pharmaceutical Chemistry*. Philadelphia: Lippincott Williams & Wilkins; 1998, pp 43–122.
55. Lewis DFV, *Guide to Cytochromes P450: Structure and Function*. London: Taylor and Francis; 2001.
56. Shimada T, Oda Y, Gillam EMJ, Guengerich FP, Inoue K. Metabolic activation of polycyclic aromatic hydrocarbons and other procarcinogens by cytochromes P450 1A1 and P450 1B1 allelic variants and other human cytochromes P450 in *Salmonella typhimurium* NM2009. *Drug Metab Dispos* 2001;29:1176–1182.
57. Patterson LH, Murray GI. Tumour cytochrome P450 and drug activation. *Curr Pharm Design* 2002;8:1335–1347.
58. Rooseboom M, Commandeur JNM, Vermeulen NPE. Enzyme-catalyzed activation of anticancer prodrugs. *Pharmacol Rev* 2004;56:53–102.
59. Vignati L, Turlizzi E, Monaci S, Grossi P, de Kanter R, Monshouwer M. An *in vitro* approach to detect metabolite toxicity due to CYP3A4-dependent bioactivation of xenobiotics. *Toxicology* 2005;216:154–167.
60. Hrycay EG, O'Brien PJ. Cytochrome P-450 as a microsomal peroxidase in steroid hydroperoxide reduction. *Arch Biochem Biophys*, 1972;153:480–494.
61. Hrycay EG, O'Brien PJ, van Lier JE, Kan G. Pregnene 17α-hydroperoxides as possible precursors of the adrenosteroid hormones. *Arch Biochem Biophys*, 1972;153:495–501.
62. Hrycay EG, Gustafsson J-Å, Ingelman-Sundberg M, Ernster L. The involvement of cytochrome P-450 in hepatic microsomal steroid hydroxylation reactions supported by sodium periodate, sodium chlorite, and organic hydroperoxides. *Eur J Biochem* 1976;61:43–52.
63. Hrycay EG, Bandiera SM. Spectral interactions of tetrachlorobiphenyls with hepatic microsomal cytochrome P450 enzymes. *Chem-Biol Interact* 2003;146:285–296.
64. Aoyama T, Yamano S, Guzelian PS, Gelboin HV, Gonzalez FJ. Five of 12 forms of Vaccinia virus-expressed human hepatic cytochrome P450 metabolically activate aflatoxin $B_1$. *Proc Natl Acad Sci USA* 1990;87:4790–4793.
65. Larrey D, Tinel M, Letteron P, Maurel P, Loeper J, Belghiti J, Pessayre D. Metabolic activation of the new tricyclic antidepressant tianeptine by human liver cytochrome P450. *Biochem Pharmacol* 1990;40:545–550.

66. Czerwinski M, McLemore TL, Philpot RM, Nhamburo PT, Korzekwa K, Gelboin HV, Gonzalez FJ. Metabolic activation of 4-ipomeanol by complementary DNA-expressed human cytochromes P-450: evidence for species-specific metabolism. *Cancer Res* 1991;51:4636–4638.
67. Tanimoto Y, Kaneko H, Ohkuma T, Oguri K, Yoshimura H. Site-selective oxidation of strychnine by phenobarbital inducible cytochrome P-450. *J Pharmacobio-Dynamics* 1991;14:161–169.
68. Shou M, Korzekwa KR, Krausz KW, Crespi CL, Gonzalez FJ, Gelboin HV. Regio- and stereo-selective metabolism of phenanthrene by twelve cDNA-expressed human, rodent, and rabbit cytochromes P-450. *Cancer Lett* 1994;83:305–313.
69. Rauschenbach R, Gieschen H, Husemann M, Salomon B, Hildebrand M. Stable expression of human cytochrome P450 3A4 in V79 cells and its application for metabolic profiling of ergot derivatives. *Eur J Pharmacol* 1995;293:183–190.
70. Caccia S, Metabolism of the newer antidepressants. *Clin Pharmacokinet* 1998;34: 281–302.
71. Daly JW. Thirty years of discovering arthropod alkaloids in amphibian skin. *J Nat Products* 1998;61:162–172.
72. Strolin-Benedetti M, Brogin G, Bani M, Oesch F, Hengstler JG. Association of cytochrome P450 induction with oxidative stress *in vivo* as evidenced by 3-hydroxylation of salicylate. *Xenobiotica* 1999;29:1171–1180.
73. Prior TI, Chue PS, Tibbo P, Baker GB. Drug metabolism and atypical antipsychotics. *Eur Neuropsychopharmacol* 1999;9:301–309.
74. Baker GB, Urichuk LJ, McKenna KF, Kennedy SH. Metabolism of monoamine oxidase inhibitors. *Cell Mol Neurobiol* 1999;19:411–426.
75. Rotzinger S, Bourin M, Akimoto Y, Coutts RT, Baker GB. Metabolism of some "second-" and "fourth-" generation antidepressants: iprindole, viloxazine, bupropion, mianserin, maprotiline, trazodone, nefazodone, and venlafaxine. *Cell Mol Neurobiol* 1999;19:427–442.
76. Chouinard G, Lefko-Singh K, Teboul E. Metabolism of anxiolytics and hypnotics: benzodiazepines, buspirone, zoplicone, and zolpidem. *Cell Mol Neurobiol* 1999;19: 533–552.
77. Lanza DL, Yost GS. Selective dehydrogenation/oxygenation of 3-methylindole by cytochrome P450 enzymes. *Drug Metab Dispos* 2001;29:950–953.
78. O'Neil MJ, Smith A, Heckelman PE, Budavari S. *The Merck Index: An Encyclopedia of Chemicals, Drugs, and Biologicals*, 13th ed. Whitehouse Station: Merck & Co; 2001.
79. Rogers JF, Nafziger AN, Bertino JS. Pharmacogenetics affects dosing, efficacy, and toxicity of cytochrome P450-metabolized drugs. *Am J Med* 2002;113:746–750.
80. Kelly P, Kahan BD. Metabolism of immunosuppressant drugs. *Curr Drug Metab* 2002;3:275–287.
81. Miyazawa M, Shindo M, Shimada T. Metabolism of (+)- and (−)-limonenes to respective carveols and perillyl alcohols by CYP2C9 and CYP2C19 in human liver microsomes. *Drug Metab Dispos* 2002;30:602–607.
82. Zhang W, Ramamoorthy Y, Tyndale RF, Glick SD, Maisonneuve IM, Kuehne ME, Sellers EM. Metabolism of 18-methoxycoronaridine, an ibogaine analog, to 18-hydroxycoronaridine by genetically variable CYP2C19. *Drug Metab Dispos* 2002;30:663–669.
83. Otake Y, Walle T. Oxidation of the flavonoids galangin and kaempferide by human liver microsomes and CYP1A1, CYP1A2, and CYP2C9. *Drug Metab Dispos* 2002;30: 103–105.

84. Tolleson WH, Doerge DR, Churchwell MI, Marques MM, Roberts DW. Metabolism of biochanin A and formononetin by human liver microsomes in vitro. *J Agric Food Chem* 2002;50:4783–4790.

85. Breinholt VM, Offord EA, Brouwer C, Nielsen SE, Brøsen K, Friedberg T. *In vitro* investigation of cytochrome P450-mediated metabolism of dietary flavonoids. *Food Chem Toxicol* 2002;40:609–616.

86. Hijazi Y, Boulieu R. Contribution of CYP3A4, CYP2B6, and CYP2C9 isoforms to N-demethylation of ketamine in human liver microsomes. *Drug Metab Dispos* 2002;30:853–858.

87. Canadian Pharmacists Association. *Compendium of Pharmaceuticals and Specialties*. Ottawa, Canada: Canadian Pharmacists Association, 2002.

88. Williams DA, Lemke TL, *Foye's Principles of Medicinal Chemistry*, 5th ed. Baltimore: Lippincott Williams & Wilkins, 2002.

89. Rendic S. Summary of information on human CYP enzymes: human P450 metabolism data. *Drug Metab Rev* 2002;34:83–448.

90. Jann MW, Shirley KL, Small GW. Clinical pharmacokinetics and pharmacodynamics of cholinesterase inhibitors. *Clin Pharmacokinet* 2002;41:719–739.

91. Smith JA, Newman RA, Hausheer FH, Madden T. Evaluation of *in vitro* drug interactions with karenitecin, a novel, highly lipophilic camptothecin derivative in phase II clinical development. *J Clin Pharmacol* 2003;43:1008–1014.

92. Reilly CA, Ehlhardt WJ, Jackson DA, Kulanthaivel P, Mutlib AE, Espina RJ, Moody DE, Crouch DJ, Yost GS. Metabolism of capsaicin by cytochrome P450 produces novel dehydrogenated metabolites and decreases cytotoxicity to lung and liver cells. *Chem Res Toxicol* 2003;16:336–349.

93. Li X-Q, Björkman A, Andersson TB, Gustafsson LL, Masimirembwa CM. Identification of human cytochrome P450s that metabolise anti-parasitic drugs and predictions of *in vivo* drug hepatic clearance from *in vitro* data. *Eur J Clin Pharmacol* 2003;59:429–442.

94. Chen G, Bunce NJ. Polybrominated diphenyl ethers as Ah receptor agonists and antagonists. *Toxicol Sci* 2003;76:310–320.

95. Breinholt VM, Rasmussen SE, Brøsen K, Friedberg TH. *In vitro* metabolism of genistein and tangeretin by human and murine cytochrome P450s. *Pharmacol Toxicol* 2003;93: 14–22.

96. Yun C-H, Lee HS, Lee H-Y, Yim S-K, Kim K-H, Kim E, Yea S-S, Guengerich FP. Roles of human liver cytochrome P450 3A4 and 1A2 enzymes in the oxidation of myristicin. *Toxicol Lett* 2003;137:143–150.

97. Paul LD, Springer D, Staack RF, Kraemer T, Maurer HH. Cytochrome P450 isoenzymes involved in rat liver microsomal metabolism of californine and protopine. *Eur J Pharmacol* 2004;485:69–79.

98. Neels HM, Sierens AC, Naelaerts K, Scharpé SL, Hatfield GM, Lambert WE. Therapeutic drug monitoring of old and newer anti-epileptic drugs. *Clin Chem Lab Med* 2004;42:1228–1255.

99. Slaughter D, Takenaga N, Lu P, Assang C, Walsh DJ, Arison BH, Cui D, Halpin RA, Geer LA, Vyas KP, Baillie TA. Metabolism of rofecoxib *in vitro* using human liver subcellular fractions. *Drug Metab Dispos* 2003;31:1398–1408.

100. Zhou S, Koh H-L, Gao Y, Gong Z-Y, Lee EJD. Herbal bioactivation: the good, the bad and the ugly. *Life Sci* 2004;74:935–968.

101. Naritomi Y, Terashita S, Kagayama A. Identification and relative contributions of human cytochrome P450 isoforms involved in the metabolism of glibenclamide and lansopra-

zole: evaluation of an approach based on the *in vitro* substrate disappearance rate. *Xenobiotica* 2004;34:415–427.

102. Ma X, Idle JR, Krausz KW, Gonzalez FJ. Metabolism of melatonin by human cytochromes P450. *Drug Metab Dispos* 2005;33:489–494.
103. Obach RS, Cox LM, Tremaine LM. Sertraline is metabolized by multiple cytochrome P450 enzymes, monamine oxidases, and glucuronyl transferases in human: an *in vitro* study. *Drug Metab Dispos* 2005;33:262–270.
104. Death AK, McGrath KCY, Handelsman DJ. Valproate is an anti-androgen and anti-progestin. *Steroids* 2005;70:946–953.
105. Hodgson E, Rose RL. Human metabolism and metabolic interactions of deployment-related chemicals. *Drug Metab Rev* 2005;1:1–39.
106. van Schaik RHN. Cancer treatment and pharmacogenetics of cytochrome P450 enzymes. *Invest New Drugs* 2005;23:513–522.
107. Mutch E, Williams FM. Diazinon, chlorpyrifos and parathion are metabolized by multiple cytochromes P450 in human liver. *Toxicology* 2006;224:22–32.
108. Jeurissen SMF, Bogaards JJP, Boersma MG, ter Horst JPF, Awad HM, Fiamegos YC, van Beek TA, Alink GM, Sudhölter EJR, Cnubben NHP, Rietjens IMCM. Human cytochrome P450 enzymes of importance for the bioactivation of methyleugenol to the proximate carcinogen 1′-hydroxymethyleugenol. *Chem Res Toxicol* 2006;19:111–116.
109. Chen C, Tu M-H. Transannular dioxygenation of 9,10-dimethyl-1,2-benzanthracene by cytochrome P-450 oxygenase of rat liver. *Biochem J* 1976;160:805–808.
110. Vaz ADN, Coon MJ. Hydrocarbon formation in the reductive cleavage of hydroperoxides by cytochrome P450. *Proc Natl Acad Sci USA* 1987;84:1172–1176.
111. Yang CS, Lu AYH, The diversity of substrates for cytochrome P450. In Guengerich FP, Ed. *Mammalian Cytochromes P450*. Boca Raton, FL; CRC Press, 1987, pp 1–17.
112. Koop DR. Oxidative and reductive metabolism by cytochrome P450 2E1. *Fed Am Soc Exp Biol J* 1992;6:724–730.
113. Nakasa H, Komiya M, Ohmori S, Rikihisa T, Kiuchi M, Kitada M. Characterization of human liver microsomal cytochrome P450 involved in the reductive metabolism of zonisamide. *Mol Pharmacol* 1993;44:216–221.
114. Chang TKH, Gonzalez FJ, Waxman DJ. Evaluation of triacetyloleandomycin, α-naphthoflavone and diethyldithiocarbamate as selective chemical probes for inhibition of human cytochromes P450. *Arch Biochem Biophys* 1994;311:437–442.
115. Veal GJ, Back DJ. Metabolism of zidovudine. *Gen Pharmacol* 1995;26:1469–1475.
116. Coon MJ, Vaz ADN, Bestervelt LL. Peroxidative reactions of diversozymes. *Fed Am Soc Exp Biol J* 1996;10:428–434.
117. Pan-Zhou X-R, Cretton-Scott E, Zhou X-J, Yang M-X, Lasker JM, Sommadossi J-P. Role of human liver P450s and cytochrome $b_5$ in the reductive metabolism of 3′-azido-3′-deoxythymidine (AZT) to 3′-amino-3′-deoxythymidine. *Biochem Pharmacol* 1998;55: 757–766.
118. Grace JM, Skanchy DJ, Aguilar AJ. Metabolism of artelinic acid to dihydroqinqhaosu by human liver cytochrome P4503A. *Xenobiotica* 1999;29:703–717.
119. Kudo S, Okumura H, Miyamoto G, Ishizaki T. Cytochrome P450 isoforms involved in carboxylic acid ester cleavage of Hantzsch pyridine ester of pranidipine. *Drug Metab Dispos* 1999;27:303–308.
120. Patterson LH, McKeown SR, Robson T, Gallagher R, Raleigh SM, Orr S. Antitumour prodrug development using cytochrome P450 (CYP) mediated activation. *Anti-Cancer Drug Design* 1999;14:473–486.

121. Guengerich FP. Common and uncommon cytochrome P450 reactions related to metabolism and chemical toxicity. *Chem Res Toxicol* 2001;14:611–650.

122. Kalgutkar AS, Nguyen HT, Vaz ADN, Doan A, Dalvie DK, McLeod DG, Murray JC. *In vitro* metabolism studies on the isoxazole ring scission in the anti-inflammatory agent leflunomide to its active α-cyanoenol metabolite A771726: mechanistic similarities with the cytochrome P450-catalyzed dehydration of aldoximes. *Drug Metab Dispos* 2003;31: 1240–1250.

123. Lesca P, Beaune P, Monsarrat B. Ellipticines and human liver microsomes: spectral interaction with cytochrome P-450 and hydroxylation. Inhibition of aryl hydrocarbon metabolism and mutagenicity. *Chem-Biol Interact* 1981;36:299–309.

124. Inaba T, Tyndale RE, Mahon WA. Quinidine: potent inhibition of sparteine and debrisoquine oxidation in vivo. *Br J Clin Pharmacol* 1986;22:199–200.

125. Tassaneeyakul W, Birkett DJ, Veronese ME, McManus ME, Tukey RH, Quattrochi LC, Gelboin HV, Miners JO. Specificity of substrate and inhibitor probes for human cytochromes P450 1A1 and 1A2. *J Pharmacol Exp Ther* 1993;265:401–407.

126. Brosen K, Skjelbo E, Rasmussen BB, Poulsen HE, Loft S. Fluvoxamine is a potent inhibitor of cytochrome P4501A2. *Biochem Pharmacol* 1993;45:1211–1214.

127. Kunze KL, Trager WF. Isoform-selective mechanism-based inhibition of human cytochrome P450 1A2 by furafylline. *Chem Res Toxicol* 1993;6:649–656.

128. Strobl GR, von Kruedener S, Stockigt J, Guengerich FP, Wolff T. Development of a pharmacophore for inhibition of human liver cytochrome P-450 2D6: molecular modeling and inhibition studies. *J Med Chem* 1993;36:1136–1145.

129. Mäenpää J, Juvonen R, Raunio H, Rautio A, Pelkonen O. Metabolic interactions of methoxsalen and coumarin in humans and mice. *Biochem Pharmacol* 1994;48:1363–1369.

130. Burke MD, Thompson S, Weaver RJ, Wolf CR, Mayer RT. Cytochrome P450 specificities of alkoxyresorufin *O*-dealkylation in human and rat liver. *Biochem Pharmacol* 1994;48:923–936.

131. López-Garcia MP, Dansette PM, Mansuy D. Thiophene derivatives as new mechanism-based inhibitors of cytochromes P-450: inactivation of yeast-expressed human liver cytochrome P-450 2C9 by tienilic acid. *Biochemistry* 1994;33:166–175.

132. Halpert JR, Guengerich FP, Bend JR, Correia MA. Contemporary issues in toxicology: selective inhibitors of cytochromes P450. *Toxicol Appl Pharmacol* 1994;125:163–175.

133. Baldwin SJ, Bloomer JC, Airton AD, Clarke SE, Chenery RJ. Ketoconazole and sulphaphenazole as the respective selective inhibitors of P450 3A and 2C9. *Xenobiotica* 1995;25:261–270.

134. Wienkers LC, Wurden CJ, Storch E, Kunze KL, Rettie AE, Trager WF. Formation of (*R*)-8-hydroxywarfarin in human liver microsomes. A new metabolic marker for the (*S*)-mephenytoin hydroxylase, P4502C19. *Drug Metab Dispos* 1996;24:610–614.

135. Roberts ES, Hopkins NE, Foroozesh M, Alworth WL, Halpert JR, Hollenberg PF. Inactivation of cytochrome P450s 2B1, 2B4, 2B6, and 2B11 by arylalkynes. *Drug Metab Dispos* 1997;25:1242–1248.

136. Moody GC, Griffin SJ, Mather AN, McGinnity DF, Riley RJ. Fully automated analysis of activities catalysed by the major human liver cytochrome P450 (CYP) enzymes: assessment of human CYP inhibition potential. *Xenobiotica* 1999;29:53–75.

137. Hasler JA, Estabrook R, Murray M, Pikuleva I, Waterman M, Capdevila J, Holla V, Helvig C, Falck JR, Farrell G, Kaminsky LS, Spivack SD, Boitier E, Beaune P. Human cytochromes P450. *Mol Aspects Med* 1999;20:1–137.

138. Chan WK, Delucchi AB. Resveratrol, a red wine constituent, is a mechanism-based inactivator of cytochrome P450 3A4. *Life Sci* 2000;67:3103–3112.

139. Rochat B, Morsman JM, Murray GI, Figg WD, McLeod HL. Human CYP1B1 and anticancer agent metabolism: mechanism for tumor-specific drug inactivation? *J Pharmacol Exp Ther* 2001;296:537–541.

140. Rae JM, Soukhova NV, Flockhart DA, Desta Z. Triethylenethiophosphoramide is a specific inhibitor of cytochrome P450 2B6: implications for cyclophosphamide metabolism. *Drug Metab Dispos* 2002;30:525–530.

141. Li X-Q, Björkman A, Andersson TB, Ridderström M, Masimirembwa CM. Amodiaquine clearance and its metabolism to *N*-desethylamodiaquine is mediated by CYP2C8: a new high affinity and turnover enzyme-specific probe substrate. *J Pharmacol Exp Ther* 2002;300:399–407.

142. Václavíková R, Horský S, Šimek P. Paclitaxel metabolism in rat and human liver microsomes is inhibited by phenolic antioxidants. *Naunyn Schmiedebergs Arch Pharmacol* 2003;368:200–209.

143. Turpeinen M, Nieminen R, Juntunen T, Taavitsainen P, Raunio H, Pelkonen O. Selective inhibition of CYP2B6-catalyzed bupropion hydroxylation in human liver microsomes in vitro. *Drug Metab Dispos* 2004;32:626–631.

144. Walsky RL, Obach RS. Validated assays for human cytochrome P450 activities. *Drug Metab Dispos* 2004;32:647–660.

145. Locuson, CW II, Suzuki H, Rettie AE, Jones JP. Charge and substituent effects on affinity and metabolism of benzbromarone-based CYP2C19 inhibitors. *J Med Chem* 2004;47:6768–6776.

146. Dutreix C, Peng B, Mehring G, Hayes M, Capdeville R, Pokorny R, Seiberling M. Pharmacokinetic interaction between ketoconazole and imatinib mesylate (Glivec) in healthy subjects. *Cancer Chemother Pharmacol* 2004;54:290–294.

147. Miyata N, Seki T, Tanaka Y, Omura T, Taniguchi K, Doi M, Bandou K, Kametani S, Sato M, Okuyama S, Cambj-Sapunar L, Harder DR, Roman RJ. Beneficial effects of a new 20-hydroxyeicosatetraenoic acid synthesis inhibitor, TS-011 [*N*-(3-chloro-4-morpholin-4-yl)phenyl-*N*-hydroxyimidoformamide], on hemorrhagic and ischemic stroke. *J Pharmacol Exp Ther* 2005;314:77–85.

148. Kim M-J, Kim H, Cha I-J, Park J-S, Shon J-H, Liu K-H, Shin J-G. High-throughput screening of inhibitory potential of nine cytochrome P450 enzymes *in vitro* using liquid chromatography/tandem mass spectrometry. *Rapid Commun Mass Spectrom* 2005;19: 2651–2658.

149. Moon YJ, Wang X, Morris ME. Dietary flavonoids: effects on xenobiotic and carcinogen metabolism. *Toxicol In Vitro* 2006;20:187–210.

150. Guengerich FP. Enzymatic oxidations of xenobiotic chemicals. *CRC Crit Rev Biochem Mol Biol* 1990;25:97–153.

151. De Voss JJ, Sibbessen O, Zhang Z, Ortiz De Montellano PR. Substrate docking algorithms and prediction of the substrate specificity of cytochrome $P450_{cam}$ and its L244A mutant. *J Am Chem Soc* 1997;119:5480–5498.

152. Josephy PD, Mannervik B, Ortiz de Montellano PR. Cytochrome P450. In *Molecular Toxicology*. Oxford: Oxford University Press; 1997, pp 209–252.

153. Yasui H, Hayashi S, Sakurai H. Possible involvement of singlet oxygen species as multiple oxidants in P450 catalytic reactions. *Drug Metab Pharmacokinet* 2005;20:1–13.

154. Kliewer SA, Goodwin B, Willson TM. The nuclear pregnane X receptor: a key regulator of xenobiotic metabolism. *Endocr Rev* 2002;23:687–702.

155. Shimada T, Yamazaki H, Mimura M, Inui Y, Guengerich FP. Interindividual variations in human liver cytochrome P450 enzymes involved in the oxidation of drugs, carcinogens and toxic chemicals: studies with liver microsomes of 30 Japanese and 30 Caucasians. *J Pharmacol Exp Ther* 1994;270:414–423.
156. Edwards RJ, Adams DA, Watts PS, Davies DS, Boobis AR. Development of a comprehensive panel of antibodies against the major xenobiotic metabolizing forms of cytochrome P450 in humans. *Biochem Pharmacol* 1998;56:377–387.
157. Rodrigues AD. Integrated cytochrome P450 reaction phenotyping. *Biochem Pharmacol* 1999;57:465–480.
158. Chang TKH, Chen J, Pillay V, Ho J-Y, Bandiera SM. Real-time polymerase chain reaction analysis of CYP1B1 gene expression in human liver. *Toxicol Sci* 2003;71:11–19.
159. Nerurkar PV, Park SS, Thomas PE, Nims RW, Lubet RA. Methoxyresorufin and benzyloxyresorufin: substrates preferentially metabolized by cytochromes P4501A2 and 2B, respectively, in the rat and mouse. *Biochem Pharmacol* 1993;46:933–943.
160. Sesardic D, Cole KJ, Edwards RJ, Davies DS, Thomas PE, Levin W, Boobis AR. The inducibility and catalytic activity of cytochromes P450c (P450IA1) and P450d (P450IA2) in rat tissues. *Biochem Pharmacol* 1990;39:499–506.
161. Gu L, Gonzalez FJ, Kalow W, Tang BK. Biotransformation of caffeine, paraxanthine, theobromine and theophylline by cDNA-expressed human CYP1A2 and CYP2E1. *Pharmacogenetics* 1992;2:73–77.
162. Jayyosi Z, Villoutreix J, Ziegler JM, Batt AM, de Maack F, Siest G, Thomas PE. Identification of cytochrome P-450 isozymes involved in the hydroxylation of dantrolene by rat liver microsomes. *Drug Metab Dispos* 1993;21:939–945.
163. Shou M, Korzekwa KR, Crespi CL, Gonzalez FJ, Gelboin HV. The role of 12 cDNA-expressed human, rodent, and rabbit cytochromes P450 in the metabolism of benzo[*a*]pyrene and benzo[*a*]pyrene *trans*-7,8-dihydrodiol. *Mol Carcinog* 1994;10: 159–168.
164. Rodrigues AD, Kukulka MJ, Roberts EM, Ouellet D, Rodgers TR. [*O*-methyl $^{14}$C]naproxen *O*-demethylase activity in human liver microsomes: evidence for the involvement of cytochrome P4501A2 and P4502C9/10. *Drug Metab Dispos* 1996;24:126–136.
165. Bloomer JC, Clarke SE, Chenery RJ. *In vitro* identification of the P450 enzymes responsible for the metabolism of ropinirole. *Drug Metab Dispos* 1997;25:840–844.
166. Shimada T, Gillam EMJ, Sutter TR, Strickland PT, Guengerich FP, Yamazaki H. Oxidation of xenobiotics by recombinant human cytochrome P450 1B1. *Drug Metab Dispos* 1997;29:617–622.
167. Lee SH, Slattery JT. Cytochrome P450 isozymes involved in lisofylline metabolism to pentoxifylline in human liver microsomes. *Drug Metab Dispos* 1997;25:1354–1358.
168. Sanderink G-J, Bournique B, Stevens J, Petry M, Martinet M. Involvement of human CYP1A isoenzymes in the metabolism and drug interactions of riluzole in vitro. *J Pharmacol Exp Ther* 1997;282:1465–1472.
169. Kinzig-Schippers M, Fuhr U, Zaigler M, Dammeyer J, Rüsing G, Labedzki A, Bulitta J, Sörgel F. Interaction of pefloxacin and enoxacin with the human cytochrome P450 enzyme CYP1A2. *Clin Pharmacol Ther* 1999;65:262–274.
170. Szotakova B, Wsol V, Skalova L, Kvasnickova E. Role of cytochrome P4501A in biotransformation of the potential anticancer drug oracin. *Exp Toxicol Pathol* 1999;51: 428–431.
171. Chen H, Howald WN, Juchau MR. Biosynthesis of all-*trans*-retinoic acid from all-*trans*-retinol: catalysis of all-*trans*-retinol oxidation by human P-450 cytochromes. *Drug Metab Dispos* 2000;28:315–322.

172. Hu M, Krausz K, Chen J, Ge X, Li J, Gelboin HL, Gonzalez FJ. Identification of CYP1A2 as the main isoform for the phase I hydroxylated metabolism of genistein and a prodrug converting enzyme of methylated isoflavones. *Drug Metab Dispos* 2003;31:924–931.

173. Granfors MT, Backman JT, Laitila J, Neuvonen PJ. Tizanidine is mainly metabolized by cytochrome P450 1A2 in vitro. *Br J Clin Pharmacol* 2003;57:349–353.

174. Peng W-X, Wang L-S, Li H-D, El-Aty AMA, Chen G-L, Zhou H-H. Evidence for the involvement of human liver microsomes CYP1A2 in the mono-hydroxylation of daidzein. *Clin Chim Acta* 2003;334:77–85.

175. Smith KS, Smith PL, Heady TN, Trugman JM, Harman WD, Macdonald TL. *In vitro* metabolism of tolcapone to reactive intermediates: relevance to tolcapone liver toxicity. *Chem Res Toxicol* 2003;16:123–128.

176. Piver B, Fer M, Vitrac X, Merillon J-M, Dreano Y, Berthou F, Lucas D. Involvement of cytochrome P450 1A2 in the biotransformation of *trans*-resveratrol in human liver microsomes. *Biochem Pharmacol* 2004;68:773–782.

177. Fuhr U, Kober S, Zaigler M, Mutschler E, Spahn-Langguth H. Rate-limiting biotransformation of triamterene is mediated by CYP1A2. *Int J Clin Pharmacol Ther* 2005;43:327–334.

178. Goda R, Nagai D, Akiyama Y, Nishikawa K, Ikemoto I, Aizawa Y, Nagata K, Yamazoe Y. Detection of a new *N*-oxidized metabolite of flutamide, *N*-[4-nitro-3-(trifluoromethy l)phenyl]hydroxylamine, in human liver microsomes and urine of prostate cancer patients. *Drug Metab Dispos* 2006;34:828–835.

179. Shimada T, Guengerich FP. Inhibition of human cytochrome P450 1A1-, 1A2-, and 1B1-mediated activation of procarcinogens to genotoxic metabolites by polycyclic aromatic hydrocarbons. *Chem Res Toxicol* 2006;19:288–294.

180. Stoilov I, Akarsu AN, Sarfarazi M. Identification of three different truncating mutations in the cytochrome P450 1B1 (CYP1B1) gene as the principal cause of primary congenital glaucoma (buphthalmos) in families linked to the GLC3A locus on chromosome 2p21. *Hum Mol Genet* 1997;6:641–647.

181. Fujii-Kuriyama Y, Mimura J. Molecular mechanisms of AhR functions in the regulation of *cytochrome P450* genes. *Biochem Biophys Res Commun* 2005;338:311–317.

182. Guengerich FP, Kim D-H, Iwasaki M. Role of human cytochrome P-450 IIE1 in the oxidation of many low molecular weight cancer suspects. *Chem Res Toxicol* 1991;4:168–179.

183. Narimatsu S, Kariya S, Isozaki S, Ohmori S, Kitada M, Hosokawa S, Masubuchi Y, Suzuki T. Involvement of CYP2D6 in oxidative metabolism of cinnarizine and flunarizine in human liver microsomes. *Biochem Biophys Res Commun* 1993;193:1262–1268.

184. Clement B, Jung F. *N*-hydroxylation of the antiprotozoal drug pentamidine catalyzed by rabbit liver cytochrome P-450 2C3 or human liver microsomes, microsomal retroreduction, and further oxidative transformation of the formed amidoximes: possible relationship to the biological oxidation of arginine to $N^{G}$-hydroxyarginine, citrulline, and nitric oxide. *Drug Metab Dispos* 1994;22:486–497.

185. Fischer V, Vickers AEM, Heitz F, Mahadevan S, Baldeck J-P, Minery P, Tynes R. The polymorphic cytochrome P-4502D6 is involved in the metabolism of both 5-hydroxytryptamine antagonists, tropisetron and ondansetron. *Drug Metab Dispos* 1994;22:269–274.

186. Miners JO, Rees DLP, Valente L, Veronese ME, Birkett DJ. Human hepatic cytochrome P450 2C9 catalyzes the rate-limiting pathway of torsemide metabolism. *J Pharmacol Exp Ther* 1995;272:1076–1081.

187. Yumibe N, Huie K, Chen K-J, Snow M, Clement RP, Cayen MN. Identification of human liver cytochrome P450 enzymes that metabolize the nonsedating antihistamine loratadine: formation of descarboethoxyloratadine by CYP3A4 and CYP2D6. *Biochem Pharmacol* 1996;51:165–172.
188. Tracy TS, Marra C, Wrighton SA, Gonzalez FJ, Korzekwa KR. Studies of flurbiprofen 4′-hydroxylation: additional evidence suggesting the sole involvement of cytochrome P450 2C9. *Biochem Pharmacol* 1996;52:1305–1309.
189. Ring BJ, Catlow J, Lindsay TJ, Gillespie T, Roskos LK, Cerimele BJ, Swanson SP, Hamman MA, Wrighton SA. Identification of the human cytochromes P450 responsible for the *in vitro* formation of the major oxidative metabolites of the antipsychotic agent olanzapine. *J Pharmacol Exp Ther* 1996;276:658–666.
190. Liu C, Zhuo X, Gonzalez FJ, Ding X. Baculovirus-mediated expression and characterization of rat CYP2A3 and human CYP2D6: role in metabolic activation of nasal toxicants. *Mol Pharmacol* 1996;50:781–788.
191. Kariya S, Isozaki S, Uchino K, Suzuki T, Narimatsu S. Oxidative metabolism of flunarizine and cinnarizine by microsomes from β-lymphoblastoid cell lines expressing human cytochrome P450 enzymes. *Biol Pharm Bull* 1996;19:1511–1514.
192. Bonnabry P, Leemann T, Dayer P. Role of human liver microsomal CYP2C9 in the biotransformation of lornoxicam. *Eur J Clin Pharmacol* 1996;49:305–308.
193. Sanwald P, David M, Dow J. Characterization of the cytochrome P450 enzymes involved in the *in vitro* metabolism of dolasetron. Comparison with other indole-containing 5-HT3 antagonists. *Drug Metab Dispos* 1996;24:602–609.
194. Ren S, Yang JS, Kalhorn TF, Slattery JT. Oxidation of cyclophosphamide to 4-hydroxycyclophosphamide and deschloroethylcyclophosphamide in human liver microsomes. *Cancer Res* 1997;57:4229–4235.
195. Kudo S, Uchida M, Odomi M. Metabolism of carteolol by cDNA-expressed human cytochrome P450. *Eur J Clin Pharmacol* 1997;52:479–485.
196. Chiba M, Xu X, Nishime JA, Balani SK, Lin JH. Hepatic microsomal metabolism of montelucast, a potent leukotriene $D_4$ receptor antagonist, in humans. *Drug Metab Dispos* 1997;25:1022–1031.
197. Kamimura H, Oishi S, Matsushima H, Watanabe T, Higuchi S, Hall M, Wood SG, Chasseaud LF. Identification of cytochrome P450 isozymes involved in metabolism of the $\alpha_1$-adrenoceptor blocker tamsulosin in human liver microsomes. *Xenobiotica* 1998;28:909–922.
198. Hiroi T, Imaoka S, Funae Y. Dopamine formation from tyramine by CYP2D6. *Biochem Biophys Res Commun* 1998;249:838–843.
199. Chesne C, Guyomard C, Guillouzo A, Schmid J, Ludwig E, Sauter T. Metabolism of meloxicam in human liver involves cytochromes P4502C9 and 3A4. *Xenobiotica* 1998;28:1–13.
200. Ko J-W, Desta Z, Flockhart DA. Human *N*-demethylation of (*S*)-mephenytoin by cytochrome P450s 2C9 and 2B6. *Drug Metab Dispos* 1998;26:775–778.
201. Lasker JM, Wester MR, Aramsombatdee E, Raucy JL. Characterization of CYP2C19 and CYP2C9 from human liver: respective roles in microsomal tolbutamide, *S*-mephenytoin, and omeprazole hydroxylations. *Arch Biochem Biophys* 1998;353:16–28.
202. Gentile DM, Verhoeven CHJ, Shimada T, Back DJ. The role of CYP2C in the *in vitro* bioactivation of the contraceptive steroid desogestrel. *J Pharmacol Exp Ther* 1998;287:975–982.
203. Obach RS, Pablo J, Mash DC. Cytochrome P4502D6 catalyzes the *O*-demethylation of the psychoactive alkaloid ibogaine to 12-hydroxyibogamine. *Drug Metab Dispos* 1998;26:764–768.

204. Venkatakrishnan K, von Moltke LL, Greenblatt DJ. Relative quantities of catalytically active CYP 2C9 and 2C19 in human liver microsomes: application of the relative activity factor approach. *J Pharm Sci* 1998;87:845–853.
205. Coller JK, Somogyi AA, Bochner F. Comparison of (*S*)-mephenytoin and proguanil oxidation *in vitro*: contribution of several CYP isoforms. *Br J Clin Pharmacol* 1999;48: 158–167.
206. de Groot MJ, Ackland MJ, Horne VA, Alex AA, Jones BC. A novel approach to predicting P450 mediated drug metabolism. CYP2D6 catalyzed *N*-dealkylation reactions and qualitative metabolite predictions using a combined protein and pharmacophore model for CYP2D6. *J Med Chem* 1999;42:4062–4070.
207. Ekins S, Wrighton SA. The role of CYP2B6 in human xenobiotic metabolism. *Drug Metab Rev* 1999;31:719–754.
208. Kobayashi K, Abe S, Nakajima M, Shimada N, Tani M, Chiba K, Yamamoto T. Role of human CYP2B6 in *S*-mephobarbital *N*-demethylation. *Drug Metab Dispos* 1999;27: 1429–1433.
209. Bourrié M, Meunier V, Berger Y, Fabre G. Role of cytochrome P-4502C9 in irbesartan oxidation by human liver microsomes. *Drug Metab Dispos* 1999;27:288–296.
210. Bachus R, Bickel U, Thomsen T, Roots I, Kewitz H. The *O*-demethylation of the antidementia drug galanthamine is catalyzed by cytochrome P450 2D6. *Pharmacogenetics* 1999;9:661–668.
211. Chow T, Hiroi T, Imaoka S, Chiba K, Funae Y. Isoform-selective metabolism of mianserin by cytochrome P-450 2D. *Drug Metab Dispos* 1999;27:1200–1204.
212. Niwa T, Shiraga T, Mitani Y, Terakawa M, Tokuma Y, Kagayama A. Stereoselective metabolism of cibenzoline, an antiarrhythmic drug, by human and rat liver microsomes: possible involvement of CYP2D and CYP3A. *Drug Metab Dispos* 2000;28:1128–1134.
213. Gillam EMJ, Notley LM, Cai H, de Voss JJ, Guengerich FP. Oxidation of indole by cytochrome P450 enzymes. *Biochemistry* 2000;39:13817–13824.
214. Pelkonen O, Rautio A, Raunio H, Pasanen M. CYP2A6: a human coumarin 7-hydroxylase. *Toxicology* 2000;144:139–147.
215. Tang C, Shou M, Mei Q, Rushmore TH, Rodrigues AD. Major role of human liver microsomal cytochrome P450 2C9 (CYP2C9) in the oxidative metabolism of celecoxib, a novel cyclooxygenase-II inhibitor. *J Pharmacol Exp Ther* 2000;293:453–459.
216. Thijssen HH, Flinois J-P, Beaune PH. Cytochrome P4502C9 is the principal catalyst of racemic acenocoumarol hydroxylation reactions in human liver microsomes. *Drug Metab Dispos* 2000;28:1284–1290.
217. Scheen AJ. Pharma-clinics medication of the month. Nebivolol (Nobiten). *Rev Méd Liège* 2000;56:788–791.
218. Oda Y, Hamaoka N, Hiroi T, Imaoka S, Hase I, Tanaka K, Funae Y, Ishizaki T, Asada A. Involvement of human liver cytochrome P4502B6 in the metabolism of propofol. *Br J Clin Pharmacol* 2001;51:281–285.
219. Brachtendorf L, Jetter A, Beckurts KT, Hölscher AH, Fuhr U. Cytochrome P-450 enzymes contributing to demethylation of maprotiline in man. *Pharmacol Toxicol* 2002;90:144–149.
220. Desta Z, Wu GM, Morocho AM, Flockhart DA. The gastroprokinetic and antiemetic drug metoclopramide is a substrate and inhibitor of cytochrome P450 2D6. *Drug Metab Dispos* 2002;30:336–343.
221. Jacobson PA, Green K, Birnbaum A, Remmel RP. Cytochrome P450 isozymes 3A4 and 2B6 are involved in the *in vitro* metabolism of thiotepa to TEPA. *Cancer Chemother Pharmacol* 2002;49:461–467.

222. Matsumoto S, Hirama T, Matsubara T, Nagata K, Yamazoe Y. Involvement of CYP2J2 on the intestinal first-pass metabolism of antihistamine drug, astemizole. *Drug Metab Dispos* 2002;30:1240–1245.

223. Usmani KA, Rose RL, Goldstein JA, Taylor WG, Brimfield AA, Hodgson E. *In vitro* human metabolism and interactions of repellent *N, N*-diethyl-*m*-toluamide. *Drug Metab Dispos* 2002;30:289–294.

224. Marill J, Capron CC, Idres N, Chabot GG. Human cytochrome P450s involved in the metabolism of 9-*cis*- and 13-*cis*-retinoic acids. *Biochem Pharmacol* 2002;63:933–943.

225. Evans DC, O'Connor D, Lake BG, Evers R, Allen C, Hargreaves R. Eletriptan metabolism by human CYP450 enzymes and transport by human P-glycoprotein. *Drug Metab Dispos* 2003;31:861–869.

226. Skinner MH, Kuan H-Y, Pan A, Sathirakul K, Knadler MP, Gonzales CR, Yeo KP, Reddy S, Lim M, Ayan-Oshodi M, Wise SD. Duloxetine is both an inhibitor and a substrate of cytochrome P4502D6 in healthy volunteers. *Clin Pharmacol Ther* 2003;73:170–177.

227. Miyazawa M, Sugie A, Shimada T. Roles of human CYP2A6 and 2B6 and rat CYP2C11 and 2B1 in the 10-hydroxylation of (–)-verbenone by liver microsomes. *Drug Metab Dispos* 2003;31:1049–1053.

228. Kajita J, Fuse E, Kuwabara T, Kobayashi H. The contribution of cytochrome P450 to the metabolism of tegafur in human liver. *Drug Metab Pharmacokinet* 2003;18:303–309.

229. Ward BA, Gorski JC, Jones DR, Hall SD, Flockhart DA, Desta Z. The cytochrome P450 2B6 (CYP2B6) is the main catalyst of efavirenz primary and secondary metabolism: implication for HIV/AIDS therapy and utility of efavirenz as a substrate marker of CYP2B6 catalytic activity. *J Pharmacol Exp Ther* 2003;306:287–300.

230. Le Gal A, Dréano Y, Lucas D, Berthou F. Diversity of selective environmental substrates for human cytochrome P450 2A6: alkoxyethers, nicotine, coumarin, *N*-nitrosodiethylamine, and *N*-nitrosobenzylmethylamine. *Toxicol Lett* 2003;144:77–91.

231. Cheng JB, Motola DL, Mangelsdorf DJ, Russell DW. De-orphanization of cytochrome P450 2R1. *J Biol Chem* 2003;278:38084–38093.

232. Yu A-M, Idle JR, Krausz KW, Küpfer A, Gonzalez FJ. Contribution of individual cytochrome P450 isozymes to the *O*-demethylation of the psychotropic β-carboline alkaloids harmaline and harmine. *J Pharmacol Exp Ther* 2003;305:315–322.

233. Attar M, Dong D, Ling K-HJ, Tang-Liu DD-S. Cytochrome P450 2C8 and flavin-containing monooxygenases are involved in the metabolism of tazarotenic acid in humans. *Drug Metab Dispos* 2003;31:476–481.

234. Benetton SA, Borges VM, Chang TKH, McErlane KM. Role of individual human cytochrome P450 enzymes in the *in vitro* metabolism of hydromorphone. *Xenobiotica* 2004;34:335–344.

235. Wójcikowski J, Pichard-Garcia L, Maurel P, Daniel WA. The metabolism of the piperazine-type phenothiazine neuroleptic perazine by the human cytochrome P-450 isoenzymes. *Eur Neuropsychopharmacol* 2004;14:199–208.

236. Tougou K, Gotou H, Ohno Y, Nakamura A. Stereoselective glucuronidation and hydroxylation of etodolac by UGT1A9 and CYP2C9 in man. *Xenobiotica* 2004;34:449–461.

237. Giraud C, Tran A, Rey E, Vincent J, Tréluyer J-M, Pons G. *In vitro* characterization of clobazam metabolism by recombinant cytochrome P450 enzymes: importance of CYP2C19. *Drug Metab Dispos* 2004;32:1279–1286.

238. Ufer M, Svensson JO, Krausz KW, Gelboin HV, Rane A, Tybring G. Identification of cytochromes $P_{450}$ 2C9 and 3A4 as the major catalysts of phenprocoumon hydroxylation *in vitro*. *Eur J Clin Pharmacol* 2004;60:173–182.

239. Hirani VN, Raucy JL, Lasker JM. Conversion of the HIV protease inhibitor nelfinavir to a bioactive metabolite by human liver CYP2C19. *Drug Metab Dispos* 2004;32: 1462–1467.

240. Akimoto M, Iida I, Itoga H, Miyata A, Kawahara S, Kohno Y. The *in vitro* metabolism of desglymidodrine, an active metabolite of prodrug midodrine, by human liver microsomes. *Eur J Drug Metab Pharmacokinet* 2004;29:179–186.

241. Maurer HH, Kraemer T, Springer D, Staack RF. Chemistry, pharmacology, toxicology, and hepatic metabolism of designer drugs of the amphetamine (ecstasy), piperazine, and pyrrolidinophenone types: a synopsis. *Ther Drug Monit* 2004;26:127–131.

242. Kubo M, Koue T, Inaba A, Takeda H, Maune H, Fukuda T, Azuma J. Influence of itraconazole co-administration and CYP2D6 genotype on the pharmacokinetics of the new antipsychotic Aripiprazole. *Drug Metab Pharmacokinet* 2005;20:55–64.

243. Winter HR, Unadkat JD. Identification of cytochrome P450 and arylamine *N*-acetyltransferase isoforms involved in sulfadiazine metabolism. *Drug Metab Dispos* 2004; 33:969–976.

244. Yasar Ü, Annas A, Svensson J-O, Lazorova L, Artursson P, Al-Shurbaji A. Ketobemidone is a substrate for cytochrome P4502C9 and 3A4, but not for P-glycoprotein. *Xenobiotica* 2005;35:785–796.

245. Bland TM, Haining RL, Tracy TS, Callery PS. CYP2C-catalyzed delta(9)-tetrahydrocannabinol metabolism: kinetics, pharmacogenetics and interaction with phenytoin. *Biochem Pharmacol* 2005;70:1096–1103.

246. Nakashima A, Kawashita H, Masuda N, Saxer C, Niina M, Nagae Y, Iwasaki K. Identification of cytochrome P450 forms involved in the 4-hydroxylation of valsartan, a potent and specific angiotensin II receptor antagonist, in human liver microsomes. *Xenobiotica* 2005;35:589–602.

247. Kajosaari LI, Laitila J, Neuvonen PJ, Backman JT. Metabolism of repaglinide by CYP2C8 and CYP3A4 *in vitro*: effect of fibrates and rifampicin. *Basic Clin Pharmacol Toxicol* 2005;97:249–256.

248. Obach RS, Cox LM, Tremaine LM. Sertraline is metabolized by multiple cytochrome P450 enzymes, monoamine oxidases, and glucuronyl transferases in human: an *in vitro* study. *Drug Metab Dispos* 2005;33:262–270.

249. Kimura M, Yamazaki H, Fujieda M, Kiyotani K, Honda G, Saruwatari J, Nakagawa K, Ishizaki T, Kamataki T. CYP2A6 is a principal enzyme involved in hydroxylation of 1,7-dimethylxanthine, a main caffeine metabolite, in humans. *Drug Metab Dispos* 2005;33:1361–1366.

250. Cederbaum AI. CYP2E1—biochemical and toxicological aspects and role in alcohol-induced liver injury. *Mt Sinai J Med* 2006;73:657–672.

251. Aiba I, Yamasaki T, Shinki T, Izumi S, Yamamoto K, Yamada S, Terato H, Ide H, Ohyama Y. Characterization of rat and human CYP2J enzymes as vitamin D 25-hydroxylases. *Steroids* 2006;71:849–856.

252. Casabar RCT, Wallace AD, Hodgson E, Rose RL. Metabolism of endosulfan-alpha by human liver microsomes as a simultaneous *in vitro* probe for CYP2B6 and CYP3A4. *Drug Metab Dispos* 2006;34:1779–1785.

253. Pritchard MP, Wolf CR. Other CYP: 2A6, 2B6, 4A. In Levy RH, Thummel KE, Trager WF, Hansten PD, Eichelbaum M, Eds. *Metabolic Drug Interactions*. Philadelphia: Lippincott Williams & Wilkins; 2000, pp 145–160.

254. Hanna IH, Reed JR, Guengerich FP, Hollenberg PF. Expression of human cytochrome P450 2B6 in *Escherichia coli*: characterization of catalytic activity and expression levels in human liver. *Arch Biochem Biophys* 2000;376:206–216.

255. Ekins S, Vandenbranden M, Ring BJ, Gillespie JS, Yang TJ, Gelboin HV, Wrighton SA. Further characterization of the expression in liver and catalytic activity of CYP2B6. *J Pharmacol Exp Ther* 1998;286:1253–1259.

256. Lamba V, Lamba J, Yasuda K, Strom S, Davila J, Hancock ML, Fackenthal JD, Rogan PK, Ring B, Wrighton SA, Schuetz EG. Hepatic CYP2B6 expression: gender and ethnic differences and relationship to CYP2B6 genotype and CAR (constitutive androstane receptor) expression. *J Pharmacol Exp Ther* 2003;307:906–922.

257. Souček P, Gut I. Cytochrome P-450 in rats: structures, functions, properties and relevant human forms. *Xenobiotica* 1992;22:83–103.

258. Sattler M, Guengerich FP, Yun C-H, Christians U, Sewing K-F. Cytochrome P-450 3A enzymes are responsible for biotransformation of FK506 and rapamycin in man and rat. *Drug Metab Dispos* 1992;20:753–761.

259. Ward S, Back DJ. Metabolism of gestodene in human liver cytosol and microsomes *in vitro*. *J Steroid Biochem Mol Biol* 1993;46:235–243.

260. Zhou-Pan X-R, Sérée E, Zhou X-J, Placidi M, Maurel P, Barra Y, Rahmani R. Involvement of human liver cytochrome P450 3A in vinblastine metabolism: drug interactions. *Cancer Res* 1993;53:5121–5126.

261. Zhou X-J, Zhou-Pan X-R, Gauthier T, Placidi M, Maurel P, Rahmani R. Human liver microsomal cytochrome P450 3A isozymes mediated vindesine biotransformation. Metabolic drug interactions. *Biochem Pharmacol* 1993;45:853–861.

262. Huskey S-EW, Dean DC, Miller RR, Rasmusson GH, Chiu S-HL. Identification of human cytochrome P450 isozymes responsible for the *in vitro* oxidative metabolism of finasteride. *Drug Metab Dispos* 1995;23:1126–1135.

263. Pichard L, Gillet G, Bonfils C, Domergue J, Thénot J-P, Maurel P. Oxidative metabolism of zolpidem by human liver cytochrome P450s. *Drug Metab Dispos* 1995;23:1253–1262.

264. Labroo RB, Thummel KE, Kunze KL, Podoll T, Trager WF, Kharasch ED. Catalytic role of cytochrome P4503A4 in multiple pathways of alfentanil metabolism. *Drug Metab Dispos* 1995;23:490–496.

265. Ono S, Hatanaka T, Miyazawa S, Tsutsui M, Aoyama T, Gonzalez FJ, Satoh T. Human liver microsomal diazepam metabolism using cDNA-expressed cytochrome P450s: role of CYP2B6, 2C19 and the 3A subfamily. *Xenobiotica* 1996;26:1155–1166.

266. Wienkers LC, Steenwyk RC, Sanders PE, Pearson PG. Biotransformation of tirilazad in human: 1. Cytochrome P450 3A-mediated hydroxylation of tirilazad mesylate in human liver microsomes. *J Pharmacol Exp Ther* 1996;277:982–990.

267. Kumar GN, Rodrigues AD, Buko AM, Denissen JF. Cytochrome P450-mediated metabolism of the HIV-protease inhibitor ritonavir (ABT-538) in human liver microsomes. *J Pharmacol Exp Ther* 1996;277:423–431.

268. Tateishi T, Krivoruk Y, Ueng YF, Wood AJ, Guengerich FP, Wood M. Identification of human liver cytochrome P-450 3A4 as the enzyme responsible for fentanyl and sufentanil *N*-dealkylation. *Anesth Analg* 1996;82:167–172.

269. Sanwald P, David M, Dow J. Characterization of the cytochrome P450 enzymes involved in the *in vitro* metabolism of dolasetron. Comparison with other indole-containing 5-HT3 antagonists. *Drug Metab Dispos* 1996;24:602–609.

270. Delaforge M, Andre F, Jaouen M, Dolgos H, Benech H, Gomis JM, Noel JP, Cavelier F, Verducci J, Aubagnac JL, Liebermann B. Metabolism of tentoxin by hepatic cytochrome P-450 3A isozymes. *Eur J Biochem* 1997;250:150–157.

271. Iatsimirskaia E, Tulebaev S, Storozhuk E, Utkin I, Smith D, Gerber N, Koudriakova T. Metabolism of rifabutin in human enterocyte and liver microsomes: kinetic parameters,

identification of enzyme systems, and drug interactions with macrolides and antifungal agents. *Clin Pharmacol Ther* 1997;61:554–562.

272. Surapaneni SS, Clay MP, Spangle LA, Paschal JW, Lindstrom TD. *In vitro* biotransformation and identification of human cytochrome P450 isozyme-dependent metabolism of tazofelone. *Drug Metab Dispos* 1997;25:1383–1388.
273. Prueksaritanont T, Gorham LM, Ma B, Liu L, Yu X, Zhao JJ, Slaughter DE, Arison BH, Vyas KP. *In vitro* metabolism of simvastatin in humans: identification of metabolizing enzymes and effect of the drug on hepatic P450s. *Drug Metab Dispos* 1997;25: 1191–1199.
274. Teramura T, Fukunaga Y, van Hoogdalem EJ, Watanabe T, Higuchi S. Examination of metabolic pathways and identification of human liver cytochrome P450 isozymes responsible for the metabolism of barnidipine, a calcium channel blocker. *Xenobiotica* 1997;27:885–900.
275. Butler AM, Murray M. Biotransformation of parathion in human liver: participation of CYP3A4 and its inactivation during microsomal parathion oxidation. *J Pharmacol Exp Ther* 1997;280:966–973.
276. von Moltke LL, Greenblatt DJ, Grassi JM, Granda BW, Fogelman SM, Harmatz JS, Kramer SJ, Fabre LF, Shader RI. Gepirone and 1-(2-pyrimidinyl)-piperazine *in vitro*: human cytochromes mediating transformation and cytochrome inhibitory effects. *Psychopharmacology* 1998;140:293–299.
277. Renwick AB, Mistry H, Ball SE, Walters DG, Kao J, Lake BG. Metabolism of zaleplon by human hepatic microsomal cytochrome P450 isoforms. *Xenobiotica* 1998;28:337–348.
278. Hashizume T, Mise M, Terauchi Y, O L, Fujii T, Miyazaki H, Inaba T. *N*-dealkylation and hydroxylation of ebastine by human liver cytochrome P450. *Drug Metab Dispos* 1998;26:566–571.
279. Yaich M, Popon M, Medard Y, Aigrain EJ. *In-vitro* cytochrome P450 dependent metabolism of oxybutynin to *N*-deethyloxybutynin in humans. *Pharmacogenetics* 1998;8:449–451.
280. Haaz MC, Rivory L, Riche C, Vernillet L, Robert J. Metabolism of irinotecan (CPT-11) by human hepatic microsomes: participation of cytochrome P-450 3A and drug interactions. *Cancer Res* 1998;58:468–472.
281. Zeng Z, Andrew NW, Arison BH, Luffer-Atlas D, Wang RW. Identification of cytochrome P4503A4 as the major enzyme responsible for the metabolism of ivermectin by human liver microsomes. *Xenobiotica* 1998;28:313–321.
282. Postlind H, Danielson A, Lindgren A, Andersson SHG. Tolterodine, a new muscarinic receptor antagonist, is metabolized by cytochromes P450 2D6 and 3A in human liver microsomes. *Drug Metab Dispos* 1998;26:289–293.
283. Baune B, Flinois JP, Furlan V, Gimenez F, Taburet AM, Becquemont L, Farinotti R. Halofantrine metabolism in microsomes in man: major role of CYP 3A4 and CYP 3A5. *J Pharm Pharmacol* 1999;51:419–426.
284. Becquemont L, Mouajjah S, Escaffre O, Beaune P, Funck-Brentano C, Jaillon P. Cytochrome P-450 3A4 and 2C8 are involved in zopiclone metabolism. *Drug Metab Dispos* 1999;27:1068–1073.
285. Salphati L, Benet LZ. Metabolism of digoxin and digoxigenin digitoxosides in rat liver microsomes: involvement of cytochrome P4503A. *Xenobiotica* 1999;29:171–185.
286. Nebbia C, Ceppa L, Dacasto M, Carletti M, Nachtmann C. Oxidative metabolism of monensin in rat liver microsomes and interactions with tiamulin and other

chemotherapeutic agents: evidence for the involvement of cytochrome P-450 3A subfamily. *Drug Metab Dispos* 1999;27:1039–1044.

287. Guengerich FP. Cytochrome P-450 3A4: regulation and role in drug metabolism. *Annu Rev Pharmacol Toxicol* 1999;39:1–17.
288. Suzuki A, Iida I, Tanaka F, Akimoto M, Fukushima K, Tani M, Ishizaki T, Chiba K. Identification of human cytochrome P-450 isoforms involved in metabolism of *R*(+)- and *S*(−)-gallopamil: utility of *in vitro* disappearance. *Drug Metab Dispos* 1999;27:1254–1259.
289. Shiraga T, Kaneko H, Iwasaki K, Tozuka Z, Suzuki A, Hata T. Identification of cytochrome P450 enzymes involved in the metabolism of zotepine, an antipsychotic drug, in human liver microsomes. *Xenobiotica* 1999;29:217–229.
290. Wienkers LC, Allievi C, Hauer MJ, Wynalda MA. Cytochrome P-450-mediated metabolism of the individual enantiomers of the antidepressant agent reboxetine in human liver microsomes. *Drug Metab Dispos* 1999;27:1334–1340.
291. May-Manke A, Kroemer H, Hempel G, Bohnenstengel F, Hohenlöchter B, Blaschke G, Boos J. Investigation of the major human hepatic cytochrome P450 involved in 4-hydroxylation and *N*-dechlorethylation of trofosfamide. *Cancer Chemother Pharmacol* 1999;44:327–334.
292. Zalma A, von Moltke LL, Granda BW, Harmatz JS, Shader RI, Greenblatt DJ. *In vitro* metabolism of trazodone by CYP3A: inhibition by ketoconazole and human immunodeficiency viral protease inhibitors. *Biol Psychiatry* 2000;47:655–661.
293. Gantenbein M, Attolini L, Bruguerolle B, Villard P-H, Puyoou F, Durand A, Lacarelle B, Hardwigsen J, Le-Treut Y-P. Oxidative metabolism of bupivacaine into pipecolylxylidine in humans is mainly catalyzed by CYP3A. *Drug Metab Dispos* 2000;28:383–385.
294. Kasahara M, Suzuki H, Komiya I. Studies on the cytochrome P450 (CYP)-mediated metabolic properties of miocamycin: evaluation of the possibility of a metabolic intermediate complex formation with CYP, and identification of the human CYP isoforms. *Drug Metab Dispos* 2000;28:409–417.
295. Prakash C, Kamel A, Cui D, Whalen RD, Miceli JJ, Tweedie D. Identification of the major human liver cytochrome P450 isoform(s) responsible for the formation of the primary metabolites of ziprasidone and prediction of possible drug interactions. *Br J Clin Pharmacol* 2000;49(Suppl 1):35S–42S.
296. Lefèvre G, Bindschedler M, Ezzet F, Schaeffer N, Meyer I, Thomsen MS. Pharmacokinetic interaction trial between co-artemether and mefloquine. *Eur J Pharm Sci* 2000;10:141–151.
297. Rawden HC, Kokwaro GO, Ward SA, Edwards G. Relative contribution of cytochromes P-450 and flavin-containing monoxygenases to the metabolism of albendazole by human liver microsomes. *Br J Clin Pharmacol* 2000;49:313–322.
298. Kobayashi K, Mimura N, Fujii H, Minami H, Sasaki Y, Shimada N, Chiba K. Role of human cytochrome P450 3A4 in metabolism of medroxyprogesterone acetate. *Clin Cancer Res* 2000;6:3297–3303.
299. Fitzpatrick JL, Ripp SL, Smith NB, Pierce WM Jr, Prough RA. Metabolism of DHEA by cytochromes P450 in rat and human liver microsomal fractions. *Arch Biochem Biophys* 2001;389:278–287.
300. Pearce RE, Gotschall RR, Kearns GL, Leeder JS. Cytochrome P450 involvement in the biotransformation of cisapride and racemic norcisapride *in vitro*: differential activity of individual human CYP3A isoforms. *Drug Metab Dispos* 2001;29:1548–1554.
301. Hyland R, Roe EGH, Jones BC, Smith DA. Identification of the cytochrome P450 enzymes involved in the *N*-demethylation of sildenafil. *Br J Clin Pharmacol* 2001; 51:239–248.

302. Igel M, Sudhop T, von Bergmann K. Metabolism and drug interactions of 3-hydroxy-3-methylglutaryl coenzyme A-reductase inhibitors (statins). *Eur J Clin Pharmacol* 2001;57:357–364.

303. Ishigami M, Honda T, Takasaki W, Ikeda T, Komai T, Ito K, Sugiyama Y. A comparison of the effects of 3-hydroxy-3-methylglutaryl-coenzyme A (HMG-CoA) reductase inhibitors on the CYP3A4-dependent oxidation of mexazolam *in vitro*. *Drug Metab Dispos* 2001;29:282–288.

304. Kassahun K, McIntosh IS, Shou M, Walsh DJ, Rodeheffer C, Slaughter DE, Geer LA, Halpin RA, Agrawal N, Rodrigues AD. Role of human liver cytochrome P4503A in the metabolism of etoricoxib, a novel cyclooxygenase-2 selective inhibitor. *Drug Metab Dispos* 2001;29:813–820.

305. Cherstniakova SA, Bi D, Fuller DR, Mojsiak JZ, Collins JM, Cantilena LR. Metabolism of vanoxerine, 1-[2-[bis(4-fluorophenyl)methoxy]ethyl]-4-(3-phenylpropyl) piperazine, by human cytochrome P450 enzymes. *Drug Metab Dispos* 2001;29:1216–1220.

306. Ibrahim A, Karim A, Feldman J, Kharasch E. The influence of parecoxib, a parenteral cyclooxygenase-2 specific inhibitor, on the pharmacokinetics and clinical effects of midazolam. *Anesth Analg* 2002;95:667–673.

307. Eagling VA, Wiltshire H, Whitcombe IW, Back DJ. CYP3A4-mediated hepatic metabolism of the HIV-1 protease inhibitor saquinavir *in vitro*. *Xenobiotica* 2002;32:1–17.

308. Yan Z, Caldwell GW, Wu WN, McKown LA, Rafferty B, Jones W, Masucci JA. *In vitro* identification of metabolic pathways and cytochrome P450 enzymes involved in the metabolism of etoperidone. *Xenobiotica* 2002;32:949–962.

309. Chen Q, Ngui JS, Doss GA, Wang RW, Cai X, DiNinno FP, Blizzard TA, Hammond ML, Stearns RA, Evans DC, Baillie TA, Tang W. Cytochrome P450 3A4-mediated bioactivation of raloxifene: irreversible enzyme inhibition and thiol adduct formation. *Chem Res Toxicol* 2002;15:907–914.

310. Cook CS, Berry LM, Kim DH, Burton EG, Hribar JD, Zhang L. Involvement of CYP3A in the metabolism of eplerenone in humans and dogs: differential metabolism by CYP3A4 and CYP3A5. *Drug Metab Dispos* 2002;30:1344–1351.

311. Guay DR. Clinical pharmacokinetics of drugs used to treat urge incontinence. *Clin Pharmacokinet* 2003;42:1243–1285.

312. Kim SY, Suzuki N, Laxmi YRS, Rieger R, Shibutani S. α-Hydroxylation of tamoxifen and toremifene by human and rat cytochrome P450 3A subfamily enzymes. *Chem Res Toxicol* 2003;16:1138–1144.

313. Tréluyer JM, Bowers G, Cazali N, Sonnier M, Rey E, Pons G, Cresteil T. Oxidative metabolism of amprenavir in the human liver. Effect of the CYP3A maturation. *Drug Metab Dispos* 2003;31:275–281.

314. Galetin A, Clarke SE, Houston JB. Multisite kinetic analysis of interactions between prototypical CYP3A4 subgroup substrates: midazolam, testosterone, and nifedipine. *Drug Metab Dispos* 2003;31:1108–1116.

315. Greenblatt DJ, von Molke LL, Giancarlo GM, Garteiz DA. Human cytochromes mediating gepirone biotransformation at low substrate concentrations. *Biopharm Drug Dispos* 2003;24:87–94.

316. Wynalda MA, Hutzler JM, Koets MD, Podoll T, Wienkers LC. *In vitro* metabolism of clindamycin in human liver and intestinal microsomes. *Drug Metab Dispos* 2003; 31:878–887.

317. Cockshott ID. Bicalutamide: clinical pharmacokinetics and metabolism. *Clin Pharmacokinet* 2004;43:855–878.

318. Dalvie DK, O'Connell TN. Characterization of novel dihydrothienopyridinium and thienopyridinium metabolites of ticlopidine *in vitro*: role of peroxidases, cytochromes P450, and monoamine oxidases. *Drug Metab Dispos* 2004;32:49–57.
319. Bu H-Z, Pool WF, Wu EY, Raber SR, Amantea MA, Shetty BV. Metabolism and excretion of capravirine, a new non-nucleoside reverse transcriptase inhibitor, alone and in combination with ritonavir in healthy volunteers. *Drug Metab Dispos* 2004;32:689–698.
320. Sanchez RI, Wang RW, Newton DJ, Bakhtiar R, Lu P, Chiu S-HL, Evans DC, Huskey S-EW. Cytochrome P450 3A4 is the major enzyme involved in the metabolism of the substance P receptor antagonist aprepitant. *Drug Metab Dispos* 2004;32:1287–1292.
321. Galetin A, Brown C, Hallifax D, Ito K, Houston JB. Utility of recombinant enzyme kinetics in prediction of human clearance: impact of variability, CYP3A5, and CYP2C19 on CYP3A4 probe substrates. *Drug Metab Dispos* 2004;32:1411–1420.
322. Berecz R, Dorado P, De La Rubia A, Caceres MC, Degrell I, Llerena A. The role of cytochrome P450 enzymes in the metabolism of risperidone and its clinical relevance for drug interactions. *Curr Drug Targets* 2004;5:573–579.
323. Ferrari A, Coccia CPR, Bertolini A, Sternieri E. Methadone—metabolism, pharmacokinetics and interactions. *Pharmacol Res* 2004;50:551–559.
324. Naritomi Y, Terashita S, Kagayama A. Identification and relative contributions of human cytochrome P450 isoforms involved in the metabolism of glibenclamide and lansoprazole: evaluation of an approach based on the *in vitro* substrate disappearance rate. *Xenobiotica* 2004;34:415–427.
325. Picard N, Cresteil T, Prémaud A, Marquet P. Characterization of a phase 1 metabolite of mycophenolic acid produced by CYP3A4/5. *Ther Drug Monit* 2004;26:600–608.
326. Usmani KA, Hodgson E, Rose RL. *In vitro* metabolism of carbofuran by human, mouse, and rat cytochrome P450 and interactions with chlorpyrifos, testosterone, and estradiol. *Chem-Biol Interact* 2004;150:221–232.
327. Kalgutkar AS, Vaz ADN, Lame ME, Henne KR, Soglia J, Zhao SX, Abramov YA, Lombardo F, Collin C, Hendsch ZS, Hop CECA. Bioactivation of the nontricyclic antidepressant nefazodone to a reactive quinine-imine species in human liver microsomes and recombinant cytochrome P450 3A4. *Drug Metab Dispos* 2005;33:243–253.
328. Zhu M, Zhao W, Jimenez H, Zhang D, Yeola S, Dai R, Vachharajani N, Mitroka J. Cytochrome P450 3A-mediated metabolism of buspirone in human liver microsomes. *Drug Metab Dispos* 2005;33:500–507.
329. Ghosal A, Chowdhury SK, Tong W, Hapangama N, Yuan Y, Su A-D, Zbaida S. Identification of human liver cytochrome P450 enzymes responsible for the metabolism of lonafarnib (Sarasar). *Drug Metab Dispos* 2005;34:628–635.
330. Picard N, Cresteil T, Djebli N, Marquet P. *In vitro* metabolism study of buprenorphine: evidence for new metabolic pathways. *Drug Metab Dispos* 2005;33:689–695.
331. Tuvesson H, Hallin I, Persson R, Sparre B, Gunnarsson PO, Seidegård J. Cytochrome P450 3A4 is the major enzyme responsible for the metabolism of laquinimod, a novel immunomodulator. *Drug Metab Dispos* 2005;33:866–872.
332. Duisken M, Sandner F, Blömeke B, Hollender J. Metabolism of 1,8-cineole by human cytochrome P450 enzymes: identification of a new hydroxylated metabolite. *Biochim Biophys Acta* 2005;1722:304–311.
333. Kitamura A, Mizuno Y, Natsui K, Yabuki M, Komuro S. Characterization of human cytochrome P450 enzymes involved in the *in vitro* metabolism of perospirone. *Biopharm Drug Dispos* 2005;26:59–65.
334. Le Tiec C, Barrail A, Goujard C, Taburet AM. Clinical pharmacokinetics and summary of efficacy and tolerability of atazanavir. *Clin Pharmacokinet* 2005;44:1035–1050.

335. Beulz-Riché D, Grudé P, Puozzo C, Sautel F, Filaquier C, Riché C, Ratanasavanh D. Characterization of human cytochrome P450 isoenzymes involved in the metabolism of vinorelbine. *Fundam Clin Pharmacol* 2005;19:545–553.

336. Tokairin T, Fukasawa T, Yasui-Furukori N, Aoshima T, Suzuki A, Inoue Y, Tateishi T, Otani K. Inhibition of the metabolism of brotizolam by erythromycin in humans: *in vivo* evidence for the involvement of CYP3A4 in brotizolam metabolism. *Br J Clin Pharmacol* 2005;60:172–175.

337. Kassahun K, Skordos K, McIntosh I, Slaughter D, Doss GA, Baillie TA, Yost GS. Zafirlukast metabolism by cytochrome P450 3A4 produces an electrophilic α,β-unsaturated iminium species that results in the selective mechanism-based inactivation of the enzyme. *Chem Res Toxicol* 2005;18:1427–1437.

338. Peng F-C, Chang C-C, Yang C-Y, Edwards RJ, Doehmer J. Territrems B and C metabolism in human liver microsomes: major role of CYP3A4 and CYP3A5. *Toxicology* 2006;218:172–185.

339. Xu Y, Hashizume T, Shuhart MC, Davis CL, Nelson WL, Sakaki T, Kalhorn TF, Watkins PB, Schuetz EG, Thummel KE. Intestinal and hepatic CYP3A4 catalyze hydroxylation of 1α,25-dihydroxyvitamin $D_3$: implications for drug-induced osteomalacia. *Mol Pharmacol* 2006;69:56–65.

340. Tanaka E, Nakamura T, Inomata S, Honda K. Effects of premedication medicines on the formation of the CYP3A4-dependent metabolite of ropivacaine, 2′, 6′-pipecoloxylidide, on human liver microsomes *in vitro*. *Basic Clin Pharmacol Toxicol* 2006;98: 181–183.

341. Chen Q, Doss GA, Tung EC, Liu W, Tang YS, Braun MP, Didolkar V, Strauss JR, Wang RW, Stearns RA, Evans DC, Baillie TA, Tang W. Evidence for the bioactivation of zomepirac and tolmetin by an oxidative pathway: identification of glutathione adducts *in vitro* in human liver microsomes and *in vivo* in rats. *Drug Metab Dispos* 2006;34:145–151.

342. Cho TM, Rose RL, Hodgson E. *In vitro* metabolism of naphthalene by human liver microsomal cytochrome P450 enzymes. *Drug Metab Dispos* 2006;34:176–183.

343. Lin C-C, Fang C, Benetton S, Xu G-F, Yeh L-T. Metabolic activation of pradefovir by CYP3A4 and its potential as an inhibitor or inducer. *Antimicrob Agents Chemother* 2006;50:2926–2931.

344. Najib J. Eszopiclone, a nonbenzodiazepine sedative-hypnotic agent for the treatment of transient and chronic insomnia. *Clin Ther* 2006;28:491–516.

345. Daly AK. Significance of the minor cytochrome P450 3A isoforms. *Clin Pharmacokinet* 2006;45:13–31.

346. Matsubara T, Kim HJ, Miyata M, Shimada M, Nagata K, Yamazoe Y. Isolation and characterization of a new major intestinal CYP3A form, CYP3A62, in the rat. *J Pharmacol Exp Ther* 2004;309:1282–1290.

347. Rekka E, Evdokimova E, Eeckhoudt S, Labar G, Calderon PB. Role of temperature on protein and mRNA cytochrome P450 3A (CYP3A) isozymes expression and midazolam oxidation by cultured rat precision-cut liver slices. *Biochem Pharmacol* 2002; 64:633–643.

348. Cooper KO, Reik LM, Jayyosi Z, Bandiera S, Kelley M, Ryan DE, Daniel R, McCluskey SA, Levin W, Thomas PE. Regulation of two members of the steroid-inducible cytochrome P450 subfamily (3A) in rats. *Arch Biochem Biophys* 1993;301:345–354.

349. Mahnke A, Strotkamp D, Roos PH, Hanstein WG, Chabot GG, Nef P. Expression and inducibility of cytochrome P450 3A9 (CYP3A9) and other members of the CYP3A subfamily in rat liver. *Arch Biochem Biophys* 1997;337:62–68.

350. Kawai M, Bandiera SM, Chang TKH, Bellward GD. Growth hormone regulation and developmental expression of rat hepatic CYP3A18, CYP3A9, and CYP3A2. *Biochem Pharmacol* 2000;59:1277–1287.
351. Castle PJ, Merdink JL, Okita JR, Wrighton SA, Okita RT. Human liver lauric acid hydroxylase activities. *Drug Metab Dispos* 1995;23:1037–1043.
352. Powell PK, Wolf I, Lasker JM. Identification of CYP4A11 as the major lauric acid omega-hydroxylase in human liver microsomes. *Arch Biochem Biophys* 1996;335: 219–226.
353. Jin R, Koop DR, Raucy JL, Lasker JM. Role of human CYP4F2 in hepatic catabolism of the proinflammatory agent leukotriene B4. *Arch Biochem Biophys* 1998;359:89–98.
354. Kikuta Y, Kusunose E, Kusunose M. Characterization of human liver leukotriene $B_4$ ω-hydroxylase P450 (CYP4F2). *J Biochem* 2000;127:1047–1052.
355. Bylund J, Bylund M, Oliw EH. cDNA cloning and expression of CYP4F12, a novel human cytochrome P450. *Biochem Biophys Res Commun* 2001;280:892–897.
356. Kikuta Y, Kusunose E, Kusunose M. Prostaglandin and leukotriene ω-hydroxylases. *Prostaglandins Other Lipid Mediators* 2002;68–69:345–362.
357. Hashizume T, Imaoka S, Mise M, Terauchi Y, Fujii T, Miyazaki H, Kamataki T, Funae Y. Involvement of CYP2J2 and CYP4F12 in the metabolism of ebastine in human intestinal microsomes. *J Pharmacol Exp Ther* 2002;300:298–304.
358. Kalsotra A, Turman CM, Kikuta Y, Strobel HW. Expression and characterization of human cytochrome P450 4F11: putative role in the metabolism of therapeutic drugs and eicosanoids. *Toxicol Appl Pharmacol* 2004;199:295–304.
359. Le Quéré V, Plée-Gauthier E, Potin P, Madec S. Human CYP4F3s are the main catalysts in the oxidation of fatty acid epoxides. *J Lipid Res* 2004;45:1446–1458.
360. Raucy JL, Lasker J, Ozaki K, Zoleta V. Regulation of CYP2E1 by ethanol and palmitic acid and CYP4A11 by clofibrate in primary cultures of human hepatocytes. *Toxicol Sci* 2004;79:233–241.
361. Kroetz DL, Xu F. Regulation and inhibition of arachidonic acid ω-hydroxylases and 20-HETE formation. *Annu Rev Pharmacol Toxicol* 2005;45:413–438.
362. Petersen KU. Omeprazole and the cytochrome P450 system. *Aliment Pharmacol Ther* 1995;9:1–9.
363. Probst-Hensch NM, Tannenbaum SR, Chan KK, Coetzee GA, Ross RK, Yu MC. Absence of the glutathione S-transferase M1 gene increases cytochrome P4501A2 activity among frequent consumers of cruciferous vegetables in a Caucasian population. *Cancer Epidemiol Biomarkers Prev* 1998;7:635–638.
364. Rae JM, Johnson MD, Lippman ME, Flockhart DA. Rifampin is a selective, pleiotropic inducer of drug metabolism genes in human hepatocytes: studies with cDNA and oligonucleotide expression arrays. *J Pharmacol Exp Ther* 2001;299:849–857.
365. Raucy JL. Regulation of CYP3A4 expression in human hepatocytes by pharmaceuticals and natural products. *Drug Metab Dispos* 2003;31:533–539.
366. Dingemanse J, van Giersbergen PLM. Clinical pharmacology of bosentan, a dual endothelin receptor antagonist. *Clin Pharmacokinet* 2004;43:1089–1115.
367. Komoroski BJ, Zhang S, Cai H, Hutzler JM, Frye R, Tracy TS, Strom SC, Lehmann T, Ang CYW, Cui YY, Venkataramanan R. Induction and inhibition of cytochromes P450 by the St. John's wort constituent hyperforin in human hepatocyte cultures. *Drug Metab Dispos* 2004;32:512–518.
368. Larsen JT, Brøsen K. Consumption of charcoal-broiled meat as an experimental tool for discerning CYP1A2-mediated drug metabolism *in vivo*. *Basic Clin Pharmacol Toxicol* 2005;97:141–148.

369. Waxman DJ. P450 gene induction by structurally diverse xenochemicals: central role of nuclear receptors CAR, PXR, and PPAR. *Arch Biochem Biophys* 1999;369:11–23.

370. Savas U, Griffin KJ, Johnson EF. Molecular mechanisms of cytochrome P-450 induction by xenobiotics: an expanded role for nuclear hormone receptors. *Mol Pharmacol* 1999;56:851–857.

371. Rowlands JC, Gustafsson, J-Å. Aryl hydrocarbon receptor-mediated signal transduction. *Crit Rev Toxicol* 1997;27:109–134.

372. Whitlock JP Jr. Induction of cytochrome P4501A1. *Annu Rev Pharmacol Toxicol* 1999;39:103–125.

373. Hankinson O. The aryl hydrocarbon receptor complex. *Annu Rev Pharmacol Toxicol* 1995;35:307–340.

374. Wang H, Negishi M. Transcriptional regulation of cytochrome P450 2B genes by nuclear receptors. *Curr Drug Metab* 2003;4:515–525.

375. Li HC, Dehal SS, Kupfer D. Induction of the hepatic CYP2B and CYP3A enzymes by the proestrogenic pesticide methoxychlor and by DDT in the rat. Effects on methoxychlor metabolism. *J Biochem Toxicol* 1995;10:51–61.

376. Blumberg B, Sabbagh W Jr, Bolado J Jr, van Meeter CM, Ong ES, Evans RM. SXR, a novel steroid and xenobiotics sensing nuclear receptor. *Genes Dev* 1998;12:3195–3205.

377. Burk O, Wojnowski L. Cytochrome P450 3A and their regulation. *Naunyn Schmiedebergs Arch Pharmacol* 2004;369:105–124.

378. Lehmann JM, Lenhard JM, Oliver BB, Ringgold GM, Kliewer SA. Peroxisome proliferator-activated receptors α and γ are activated by indomethacin and other nonsteroidal anti-inflammatory drugs. *J Biol Chem* 1997;272:3406–3410.

379. Gonzalez FJ, Peters JM, Cattley RC. Mechanism of action of the nongenotoxic peroxisome proliferators: role of the peroxisome proliferator-activated receptor α. *J Natl Cancer Inst* 1998;90:1702–1709.

380. Moore LB, Parks DJ, Jones SA, Bledsoe RK, Consler TG, Stimmel JB, Goodwin B, Liddle C, Blanchard SG, Willson TM, Collins JL, Kliewer SA. Orphan nuclear receptors constitutive androstane receptor and pregnane X receptor share xenobiotic and steroid ligands. *J Biol Chem* 2000;275:15122–15127.

381. Staudinger JL, Goodwin B, Jones SA, Hawkins-Brown D, MacKenzie IK, LaTour A, Liu Y, Klaassen CD, Brown KK, Reinhard J, Willson TM, Koller BH, Kliewer SA. The nuclear receptor PXR is a lithocholic acid sensor that protects against liver toxicity. *Proc Natl Acad Sci USA* 2001;98:3369–3374.

382. Goodwin B, Moore LB, Stoltz CM, McKee DD, Kliewer SA. Regulation of the human *CYP2B6* gene by the nuclear pregnane X receptor. *Mol Pharmacol* 2001;60:427–431.

383. Ripp SL, Fitzpatrick JL, Peters JM, Prough RA. Induction of *CYP3A* expression by dehydroepiandrosterone: involvement of the pregnane X receptor. *Drug Metab Dispos* 2002;30:570–575.

384. Maglich JM, Stoltz CM, Goodwin B, Hawkins-Brown D, Moore JT, Kliewer SA. Nuclear pregnane X receptor and constitutive androstane receptor regulate overlapping but distinct sets of genes involved in xenobiotic detoxification. *Mol Pharmacol* 2002;62: 638–646.

385. Denison MS, Nagy SR. Activation of the aryl hydrocarbon receptor by structurally diverse exogenous and endogenous chemicals. *Annu Rev Pharmacol Toxicol* 2003;43: 309–334.

386. Handschin C, Meyer UA. Induction of drug metabolism: the role of nuclear receptors. *Pharmacol Rev* 2003;55:649–673.

387. Chen Y, Ferguson SS, Negishi M, Goldstein JA. Induction of human *CYP2C9* by rifampicin, hyperforin, and phenobarbital is mediated by the pregnane X receptor. *J Pharmacol Exp Ther* 2004;308:495–501.
388. Pascussi J-M, Gerbal-Chaloin S, Drocourst L, Assénat E, Larrey D, Pichard-Garcia L, Vilarem M-J, Maurel P. Cross-talk between xenobiotic detoxication and other signalling pathways: clinical and toxicological consequences. *Xenobiotica* 2004;34:633–664.
389. van Raalte DH, Li M, Pritchard PH, Wasan KM. Peroxisome proliferator-activated receptor (PPAR)-α: a pharmacological target with a promising future. *Pharm Res* 2004;21:1531–1538.
390. Chen Y, Kissling G, Negishi M, Goldstein JA. The nuclear receptors constitutive androstane receptor and pregnane X receptor cross-talk with hepatic nuclear factor 4α to synergistically activate the human *CYP2C9* promoter. *J Pharmacol Exp Ther* 2005;314:1125–1133.
391. Kretschmer XC, Baldwin WS. CAR and PXR: xenosensors of endocrine disrupters? *Chem-Biol Interact* 2005;155:111–128.
392. Tirona RG, Kim RB. Nuclear receptors and drug disposition gene regulation. *J Pharm Sci* 2005;94:1169–1186.
393. Ferguson SS, Chen Y, LeCluyse EL, Negishi M, Goldstein JA. Human *CYP2C8* is transcriptionally regulated by the nuclear receptors constitutive androstane receptor, pregnane X receptor, glucocorticoid receptor, and hepatic nuclear factor 4α. *Mol Pharmacol* 2005;68:747–757.
394. Chang TK, Waxman DJ. Synthetic drugs and natural products as modulators of constitutive androstane receptor (CAR) and pregnane X receptor (PXR). *Drug Metab Rev* 2006;38:51–73.
395. Itoh M, Nakajima M, Higashi E, Yoshida R, Nagata K, Yamazoe Y, Yokoi T. Induction of human CYP2A6 is mediated by the pregnane X receptor with peroxisome proliferator-activated receptor-γ coactivator 1α. *J Pharmacol Exp Ther* 2006;319: 693–702.
396. Nelson D. Cytochrome P450 homepage, 18 July 2006. http://drnelson.utmem.edu/CytochromeP450.html.
397. Ragueneau-Majlessi I, Hachad H, Levy RH. M&T drug interaction database. 2004. http://didbase@u.washington.edu.
398. Flockhart D. Drug interactions. 29 August, 2006. http://www.drug-interactions.com.

# METABOLISM KINETICS

Charles W. Locuson and Timothy S. Tracy
*University of Minnesota, Minneapolis, Minnesota*

**Contents**

1 Introduction
2 Typical Kinetics
  2.1 Michaelis–Menten Equation and Its Use in Drug Metabolism
  2.2 The Case of Multiple Substrates
  2.3 Use of the Michaelis–Menten Equation and Its Parameters
3 Atypical Kinetics Displayed by Drug Metabolizing Enzymes
  3.1 Sigmoidal Kinetics with a Single Substrate
  3.2 Biphasic Kinetics (Nonasymptotic)
  3.3 Hyperbolic (Nonessential) Activation
  3.4 Substrate Inhibition
  3.5 Explanations for Atypical Kinetics in P450 Enzymes and Their Detection
  3.6 Artifactual Sources of Atypical Kinetics
4 Conclusion
  Acknowledgment
  References

## 1 INTRODUCTION

It is common to apply kinetic models to *in vitro* data in an attempt to determine parameters, such as intrinsic clearance, that can be used to make correlations to the *in vivo* situation. These estimations are generally predicated on the data following a hyperbolic profile such that the Michaelis–Menten equation can be applied. This

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

allows determination of $V_m$ and $K_m$ and thus the determination of intrinsic clearance. However, it is increasingly being recognized that, for many substrates, these "typical kinetic" conditions may not exist and "atypical kinetics" may be observed. Unfortunately, application of the Michaelis–Menten equation to these "atypical" situations results in misestimation of $V_m$ and $K_m$ and, subsequently, intrinsic clearance. Thus, it is important to understand the assumptions behind typical kinetics in drug metabolism reactions, the different types of atypical kinetic profiles that may also be observed, and what models and equations can be used to estimate kinetic parameters from each of these types of data.

## 2 TYPICAL KINETICS

### 2.1 Michaelis–Menten Equation and Its Use in Drug Metabolism

One of the most widely taught and used biochemical principles is embodied in an equation, which has remained the very foundation of enzymology for nearly a century. This equation allows quantitative descriptions of many enzyme-catalyzed reactions in terms of rate and concentration and therefore helps us describe the molecular workings of the cell. Enzymology has aided in the diagnoses of diseases as well as the development of therapies used to treat them, and at the heart of enzymology is the Michaelis–Menten equation:

$$v = \frac{V_m \cdot [S]}{K_m + [S]} \tag{1}$$

where reaction velocity, ν, is described as a function of substrate concentration, [S], multiplied by a limiting rate, $V_m$ (also known as $V$ or $V_{max}$ whose subscript is not to be confused with that of $K_m$), and divided by the sum of [S] and constant $K_m$, whose subscript refers to its name, the Michaelis constant.

Following up on the theory of A. J. Brown and V. Henri that enzymatic catalysis results from the formation of an intermediate enzyme–substrate complex, Michaelis and Menten reaffirmed Henri's basic assumption.[1] Mainly, the effect of substrate concentration on velocity could be described by an equilibrium between enzyme and substrate and that their complexation can lead to product.

$$\mathrm{E} + \mathrm{S} \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} \mathrm{ES} \xrightarrow{k_2} \mathrm{E} + \mathrm{P} \tag{2}$$

Derivation of the rate equation for Eq. 2 based on the initial concentration of enzyme and substrate, and the definition of the substrate dissociation constant, $K_S$, gives

$$v = \frac{k_2 \cdot \mathrm{E}_0 \cdot [S]}{K_S + [S]} \tag{3}$$

[1]The following details were compiled largely from Cornish-Bowden [6], and for further details, consultation of this book is highly recommended.

where $E_0$ is the initial enzyme concentration. By avoiding any assumption about equilibrium in the first step of Eq. 2 (i.e., substrate binding to enzyme), Briggs and Haldane actually made the final contribution to the Michaelis–Menten equation with their steady-state treatment. Their steady-state assumption was that the concentration of the enzyme–substrate intermediate would become constant soon after initiation of the reaction. Using the same scheme (Eq. 2), $K_S$ in Eq. 3 is substituted with the Michaelis constant:

$$K_m = \frac{k_{-1} + k_2}{k_1} \tag{4}$$

using the steady-state assumption as shown in the derivation available in most biochemistry textbooks. This is probably the preferred treatment unless specific details regarding substrate binding are known that enable the use of $K_S$. Whether $K_m$ or $K_S$ is used, the form of the equation is the same:

$$v = \frac{k_{cat} \cdot E_0 \cdot [S]}{K_m + [S]} \tag{5}$$

with the substitution of $k_2$ for $k_{cat}$ (the observed rate[2] of reaction) so that no assumption is made about the number of steps that occur between substrate binding and the chemistry described by $k_{cat} \times E_0$. In other words, $k_2$ in Eq. 2 does not reflect the rate of catalysis except for the simplest two-step mechanism such that $k_{cat}$ is used to represent the observed rate of reaction if there are more than two steps, even if the true number of steps is not known. For the same reason $K_m$ is often more complex than shown in Eq. 4. $E_0$ and $k_{cat}$ are then multiplied to produce the parameter $V_m$ to yield the familiar Eq. 1.

Equation 2 does not consider the reversibility of the $k_2$ step, which would result in the subtraction of the reverse rate in the numerator ($k_P \times E_0 \times [P]$) and a term that includes a $K_m$ for product P in the denominator. Even so, the Michaelis–Menten equation in its commonest form (Eq. 1) adequately describes the kinetics of metabolism for many drugs by enzymes like the cytochromes P450. Ideally, the rate of reaction at each substrate concentration would follow the method of *initial rates*, meaning several measurements over a limited amount of time (i.e., less than a few minutes) are subjected to linear regression to determine a rate. This prevents excessive amounts of product from building up and reversibly (or irreversibly) inhibiting the enzyme, or undergoing significant reversal back to substrate. However, this is often not feasible if time-expensive methods like liquid-chromatography separation or even extraction into organic solvent are needed to measure metabolized drug. In addition, even an incubation carried out with recombinant P450 is

[2] $k_{cat}$, the *catalytic constant* or *turnover number*, describes the number of turnovers an enzyme can accomplish in a given unit of time. Determination of $k_{cat}$ requires accurately estimating $V_m$, therefore requiring the use of a sufficient range of substrate concentration to ensure saturation, and knowledge of $E_0$ ($k_{cat} = V_m/E_0$). $V_m$ is determined preferably by nonlinear regression (although certain linear plots have been used for estimates). The initial enzyme concentration, $E_0$ or $E_{total}$, is in the same concentration units as $V_m$ (e.g., μM/time) so that the unit of $k_{cat}$ is time$^{-1}$.

much more complex than most may realize. Using the steady-state treatment, a P450 incubation may be thought of as possessing a $K_m$ not only for drug, but for oxygen, P450 oxidoreductase (which has a $K_m$ for NADPH), and possibly any other partially rate-limiting steps required for catalysis. Fortunately, most of these variables are kept constant and/or saturating. Considering the number of variables in P450–drug incubations and some of the difficulties involved in analyzing the products of such reactions, the Michaelis–Menten equation nevertheless remains remarkably useful.

Therefore, although proper controls must be carried out (see below), triplicate end-point measurements and their analysis with the Michaelis–Menten equation are probably the most fundamental and useful experiments in drug metabolism.

## 2.2 The Case of Multiple Substrates

Cofactors are heavily relied upon by drug metabolizing enzymes. The UDP-glucuronosyltransferases (UGTs), glutathione *S*-transferases (GSTs), flavin monooxygenases (FMOs), sulfotransferases (more correctly, sulfonyltransferases), *N*-acetyltransferases (NATs), and thiopurine *S*-methyltransferase (TPMT) all utilize nucleotide-containing cofactors for oxygen activation (e.g., FMO) or for conjugation (e.g., all others). FMOs can remain poised for oxidation without drug substrate ever binding as long as saturating levels of NADPH are present. UGTs, on the other hand, require a ternary complex of enzyme and two substrates (UDP-glucuronic acid and the aglycone substrate to be conjugated) to form before catalysis can occur.

Unfortunately, the number of parameters in rate equations expands substantially when describing models involving additional reaction steps. For instance, take the example of the kinetic mechanism of UGT that generally occurs through the ordered binding of two cosubstrates [1]. If the reverse reaction is ignored and any amount of product formed during the course of the reaction is considered negligible, the rate equation requires concentration information and estimation of $K_m$ values for both substrates:

$$v = \frac{V_m[A][B]}{K_{iA}K_{mB} + K_{mB}[A] + K_{mA}[B] + [A][B]} \tag{6}$$

where A and B represent the two substrates each with their own $K_m$ ($K_{mA}$ and $K_{mB}$) and $K_{iA}$ is the dissociation constant for the EA complex. Even with the aforementioned assumption, a substantial number of data points would be needed to fit this rate equation due to the number of parameters involved. However, if substrate B is always present in saturating concentrations and care is taken to avoid the buildup of products, Eq. 6 *reduces eventually to the Michaelis–Menten equation* (Eq. 1). In fact, this experimental and analytical simplification is probably the standard and preferred method for determining $V_m$ as well as the $K_m$ for each substrate. With further analysis this method also has the benefit of revealing features of the kinetic mechanism such as whether substrate binding is random or ordered, or if the first substrate is turned over and leaves before the second substrate can be turned over (i.e., Ping-Pong mechanism) [2].

### 2.3 Use of the Michaelis–Menten Equation and Its Parameters

Most are familiar with the rectangular hyperbola that results from plotting reaction velocity ($y$-axis) versus substrate concentration ($x$-axis) and that these plotted data can be fit to the Michaelis–Menten equation using nonlinear regression (although the curve itself was never used by Michaelis and Menten and therefore is *not* called a Michaelis–Menten curve). Therefore, the parameters $V_m$ and $K_m$, as best fit by the nonlinear version of the equation (Eq. 1), are reviewed.

At low [S], Eq. 1 demonstrates that velocity becomes proportional to [S]:

$$v \approx \frac{k_{cat} \cdot E_0 \cdot [S]}{K_m} = \frac{V \cdot [S]}{K_m} \tag{7}$$

The middle expression of Eq. 7 is important because a very useful parameter is obtained from the ratio of $k_{cat}/K_m$ known as the *specificity constant*. $k_{cat}/K_m$ (units of time$^{-1}$ × concentration$^{-1}$) is an excellent evaluation of substrate specificity for an enzyme. Imagine two substrates for the same enzyme that can each be turned over without the other present. The substrate with the highest specificity constant will correctly predict that it will be turned over faster than the other substrate if both were present at any equimolar concentration. Even if one substrate has a larger $k_{cat}$, a substrate with a higher $k_{cat}/K_m$ will be turned over faster if the two are compared at the same concentration.

Low substrate concentrations (i.e., much lower than $K_m$) are also relevant to *in vivo–in vitro* correlations. If the therapeutic dosage of a drug results in sub-$K_m$ plasma concentrations (the most common situation), where kinetics are largely first order for the enzyme(s) metabolizing it, Eq. 7 can be defined as the intrinsic clearance ($CL_{int} = V_m/K_m$, units time$^{-1}$), which is the clearance of drug from a tissue via metabolism independent of blood flow, protein binding, and other restrictions. Although the units are different from that for $k_{cat}/K_m$, $V_m/K_m$ is still a specificity constant and the comparison of $CL_{int}$ values for different drugs with the same enzyme may prove useful in predicting the reduced $CL_{int}$ for *each* drug involved in a drug–drug interaction or polypharmacy.

Next, when the substrate concentration is equal to $K_m$, $K_m$ can be defined operationally as the *concentration of substrate where velocity equals* $\frac{1}{2}V_m$. Often, $K_m$ in this sense is used as a measure of the affinity of substrate for enzyme as if they were in equilibrium ($K_S$); however, without additional information regarding rate constants, $K_m$ should never be assumed to be anything more than an estimate of $K_S$ with no guarantees attached. Finally, when substrate is said to be saturating so that every enzyme molecule is bound to substrate, $K_m$ becomes negligible and $v \approx V_m$. In other words, the rate is limited by the amount of enzyme present, and hence, $V_m$ is perhaps better described as a limiting rate rather than a maximum.

## 3 ATYPICAL KINETICS DISPLAYED BY DRUG METABOLIZING ENZYMES

Atypical kinetics is a term that has come to be used to describe any *in vitro* drug metabolism kinetics that do not fit the hyperbolic function seen when velocity is

plotted versus substrate concentration. *Allosterism* is another term used frequently in the P450 literature that refers to the ability of a ligand (a.k.a. effector) other than the substrate being turned over to bind in a distinct, noncatalytic site on an enzyme and affect the rate of metabolism of the substrate. The effector can be a second (or third or fourth, etc.) substrate molecule binding to the same enzyme or it can be a molecule that is structurally dissimilar from the substrate. Several lines of evidence suggest that some of the allosteric sites in P450s 3A4 and 2C9 are really distinct sites within the same binding pocket that substrate occupies; however, we think allosterism is still a suitable term to define this phenomenon since the P450s most likely do not bind every substrate in the same orientation or subregion of their active sites anyway.

Much attention has been given to the atypical kinetics of P450s in drug metabolism, but it is becoming apparent that phase II enzymes and drug transporters may occasionally display atypical kinetics likely stemming from the simultaneous occupancy of a single protein by multiple ligand (in addition to the normal substrates, if there are multiple substrates required for catalysis). In fact, almost all of the theoretical mechanisms that could give rise to atypical kinetics are plausible with drug metabolizing enzymes whether it occurs via one substrate (*homotropic*) or combinations of substrate and effector (*heterotropic*). Using P450s as an example, it is completely reasonable to hypothesize atypical kinetics arise from effector binding that modifies P450–P450, P450–P450 reductase, or P450–cytochrome-$b_5$ interactions, where the site can be in the same binding pocket as substrate or remote and be either homotropic or heterotropic.

### 3.1 Sigmoidal Kinetics with a Single Substrate

Sigmoidicity in substrate–velocity plots usually brings to mind cooperativity, which is generally thought of as arising from multimeric enzymes, like the textbook example of hemoglobin. Here, the sites where substrate binds are equivalent on each enzyme so cooperativity, not allosterism (e.g., effect of BPG on $O_2$ binding by hemoglobin), is used to describe the kinetic phenomenon. Evidence, however, now suggests P450s may display homotropic activation (or homotropic cooperativity), where a monomeric enzyme displays sigmoidal kinetics (Fig. 1) presumably through substrate binding to both a catalytic site and an effector site [3–5]. In any case, a useful expression called the Hill equation (Eq. 8) is adequate for describing many cases of this type of kinetics with just a few parameters:

$$v = \frac{V_m[S]^n}{K^n + [S]^n} \tag{8}$$

where $V_m$ and $K$ are analogous to the parameters in the Michaelis–Menten equation (but the parameter[3] $K$ is *not* called the Michaelis constant) and $n$ is the Hill coefficient. The Hill equation is empirical with respect to the parameter $n$ because while

[3]It is not essential that $K$ be to the power $n$ except that since $n$ is often not a whole integer, $K$ as the substrate concentration that gives the half-maximal velocity will have the same units as [S], but not be directly relatable to it.

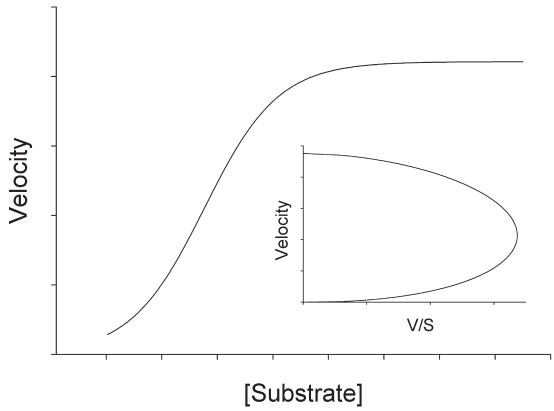


**FIGURE 1** Graphical representation of a kinetic profile exhibiting sigmoidal (autoactivation) kinetics. Inset demonstrates the Eadie–Hofstee plot of a sigmoidal kinetic profile.

it measures relative cooperativity, it does not necessarily equal the number of binding sites for substrate because it is often not an integer. A more biochemically logical description of multiple binding sites is obtained with the use of the Adair equation (described in Cornish-Bowden [6]); however, the parameters of the Hill equation remain practical for most applications, including the study of drug metabolizing enzymes. For example, substrates (e.g., testosterone, carbamazepine, pyrene) that induce sigmoidal kinetics by CYP3A4 can all be meaningfully compared to each other using the Hill equation to determine which ones have the highest efficiency ($V_m/K$) and cooperativity ($n$).

When cooperativity is caused by an effector other than the target substrate in a monomeric enzyme (a.k.a. *heterotropic activation*) [7, 8], for instance, during inhibition screening, the Hill equation can still be used. Or, to better define the $V_m$ and $K_m$ parameters for substrate when it is known that the effector turnover rate is insignificant or much less than it is for substrate, Eq. 9 may prove useful by giving $K_m$ and $V_m$ in the absence or presence of effector (subscripts 1 and 2, respectively):

$$v = \frac{\dfrac{V_{m1} \cdot [S]}{K_{m1}} + \dfrac{V_{m2} \cdot [S]^2}{K_{m1} \cdot K_{m2}}}{1 + \dfrac{[S]}{K_{m1}} + \dfrac{[S]^2}{K_{m1} \cdot K_{m2}}} \tag{9}$$

Equation 9, although working under the assumption of equilibrium, will also give the $K_m$ and $V_m$ of a substrate metabolized in two distinct subsites within the substrate binding pocket, each with different binding affinities and velocities. More complex rate equations have been developed for sigmoidal kinetics in drug metabolism and more rare mechanisms behind cooperativity can be found in Cornish-Bowden [6], and these sources should be consulted for a full description.

### 3.2 Biphasic Kinetics (Nonasymptotic)

As defined here, a biphasic kinetic profile does not follow saturation kinetics and has two distinct phases. (*Note that sigmoidal kinetics may also be biphasic but exhibits saturation.*) At low substrate concentrations, the kinetic profile exhibits curvature similar to that observed with hyperbolic kinetics; however, at high substrate concentrations, the velocity of the reaction continues to increase in a linear fashion (Fig. 2), as opposed to becoming asymptotic as would be expected with hyperbolic kinetics. This type of profile most commonly occurs when the metabolism of a substrate is carried out by more than one enzyme in a multienzyme system. However, when using a single, purified enzyme source, it is most likely the result of substrate binding in a productive orientation to more than one region within the active site, but these binding regions exhibit different affinities for the substrate and also differential turnover rates. In this case, a single enzyme behaves as if it were a multienzyme system. The enzyme thus exhibits a low $K_m$/low $V_m$ component (responsible for the semihyperbolic nature of the profile) and another high $K_m$/high $V_m$ component (producing the linear portion of the profile). In order to model this type of kinetic profile, one must use equations that describe the kinetics of one substrate interacting with two binding sites within the same enzyme species, such as Eq. 10.

$$v = \frac{(V_{m1} \cdot [S]) + (CL_{int} \cdot [S]^2)}{K_{m1} + [S]} \tag{10}$$

The parameters $V_{m1}$ and $K_{m1}$ are estimated from data representing the curved portion of the plot at lower substrate concentrations and are the estimates of $V_m$ and $K_m$ for the low $V_m$, low $K_m$ site, respectively. $CL_{int}$ describes the linear portion of the plot exhibited at higher substrate concentrations and is the ratio of $V_{max2}/K_{m2}$ (i.e., the high $V_{max}$, high $K_m$ component of the profile). Because this upper portion of the plot is linear, one cannot estimate the actual $V_m$ and $K_m$ parameters for this portion of the profile since saturation is not achieved. Thus, $CL_{int}$ is the slope (rate) for this linear portion of the kinetic profile.

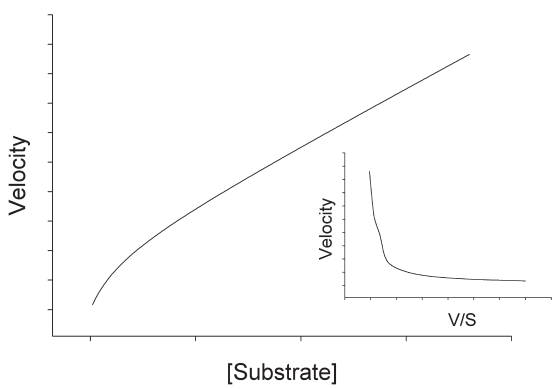


**FIGURE 2** Graphical representation of a kinetic profile exhibiting biphasic kinetics. Inset demonstrates the Eadie–Hofstee plot of a biphasic kinetic profile.

### 3.3 Hyperbolic (Nonessential) Activation

The most straightforward kind of activation in terms of fitting $K_m$ and $V_m$ values to velocity versus [S] plots is probably hyperbolic activation. The name refers to the replot of $K_m/V_m$ against reciprocal activator concentration, which is hyperbolic, not linear. (These plots for sigmoidal and biphasic kinetics are also not expected to be linear.) It is nonessential, because the effector is not necessary for catalysis to occur. While $V_m$ and often $K_m$ change for a substrate in the presence of an activating effector, the shape of the velocity vesus [S] curve remains a hyperbola. Therefore, the term hyperbolic activation has a more important meaning in drug metabolism kinetics because it can also refer to homo- or heterotropic activation that results in hyperbolic curves rather than sigmoidal or biphasic curves. As one might expect, changes in $K_m$ and $V_m$ can be fit at each activator concentration using the Michaelis–Menten equation. One of the best documented cases of this heterotropic activation is the activation of flurbiprofen (S) hydroxylation by the effector dapsone (B) in P450 2C9 and a mechanistic scheme has been proposed (Fig. 3). A graphical representation of the kinetic profile for this hyperbolic activation is shown in Fig. 4.

The equation (Eq. 11) for fitting multiple curves representing different activator concentrations has a manageable number of parameters including the usual $K_m$ (which is assumed to be an estimate of $K_S$ in this equilibrium model) and $V_m$ for substrate in the absence of activator:

$$v = \frac{V_m \cdot [S]}{K_m \dfrac{1+[B]/K_B}{1+\beta[B]/\alpha K_B} + [S]\dfrac{1+[B]/\alpha K_B}{1+\beta[B]/\alpha K_B}} \tag{11}$$

where [B] is the concentration of activator, $\alpha$ is the factor by which $K_S$ for substrate changes, and $\beta$ can be thought of as the factor by which $V_m$ changes (see Fig. 3 for further explanation of parameters).

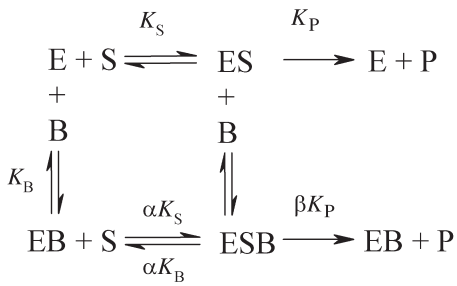


**FIGURE 3** Kinetic mechanism for nonessential activation [2] demonstrates that the effector, B, is not essential for product formation. Furthermore, it can be seen that this scheme is flexible as $\alpha$ and $\beta$ can be greater or less than unity. In the case of flurbiprofen, metabolism in the presence of effector dapsone results in a decrease in $\alpha$, indicating an *increase* in affinity of P450 2C9 for flurbiprofen. This was further supported by equilibrium binding experiments. Parameter $\beta$ was increased, indicating the velocity for oxidation of flurbiprofen has increased. Activation could also be achieved with $\alpha > 1$ but this would require higher substrate concentrations to be observed. Alternatively, if $\alpha < 1$, but $\beta < 1$, activation could result at low substrate concentrations and then result in inhibition at higher substrate concentrations.

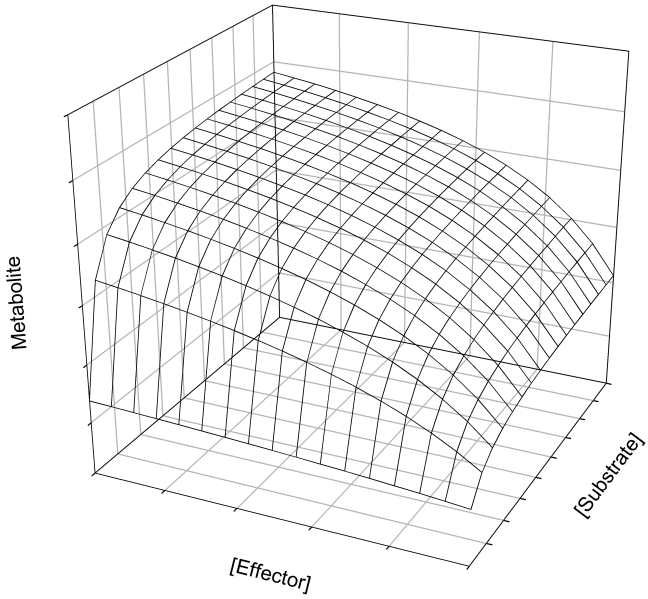


**FIGURE 4** Graphical representation of a kinetic profile exhibiting heterotropic cooperativity (activation).

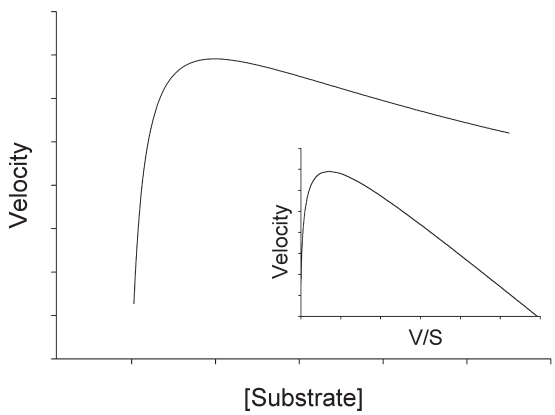


**FIGURE 5** Graphical representation of a kinetic profile exhibiting substrate inhibition kinetics. Inset demonstrates the Eadie–Hofstee plot of a substrate inhibition kinetic profile.

### 3.4 Substrate Inhibition

In the case of substrate inhibition kinetics, as the concentration of substrate is increased, to determine $V_m$ an inflection point in the velocity versus [S] curve occurs so that rather than plateau, the velocity starts to decrease and continues to do so as [S] is raised further (Fig. 5). Assuming the reactions were carried out appropriately (see Section 3.6), two likely scenarios arise when substrate inhibition is noted:

1. If the experiment is conducted with P450, the most likely cause is analogous to uncompetitive inhibition where two or more ligands (in this case, two substrate molecules) binding to the enzyme decrease its ability to catalyze substrate oxidation.
2. If the experiment is conducted with UGT, inhibition is likely the result of aglycone substrate binding right after a glucuronide conjugate is released from a previous round of catalysis, thereby trapping UDP and forming an inactive ternary complex. For many UGTs there appears to be a preference to bind UDP-glucuronic acid first, followed by aglycone, suggesting an ordered ternary complex mechanism.

In either case, $K_m$ and $V_m$ can be estimated via use of the appropriate equation. It should be noted, however, that the derived parameters (e.g., $V_m$ and $K_m$) are not truly equivalent to the usual situation. Because $V_m$ is never truly reached (or to simulate it unless the inhibition is very weak), the true $K_m$ is also not discernible. In the past, researchers have incorrectly fit these types of data by either eliminating the points showing decreasing velocity or simply applying the Michaelis–Menten equation to the entire dataset. Lin and colleagues [9] have demonstrated the necessity to fit the entire dataset to the proper equation and the hazards of not doing so. For substrate inhibition occurring through promiscuous binding of a second substrate (case 1), the following rate equation (Eq. 12) is useful. It is derived from the uncompetitive inhibition model using an extra substrate molecule instead of inhibitor, therefore leading to a squared term in the denominator when multiplied out:

$$v = \frac{V_m[S]}{K_m + [S] + \frac{[S]^2}{K_{si}}} \tag{12}$$

However, in the course of kinetic analyses one may wish to gain additional information from the available kinetic data, allowing for estimation of additional kinetic parameters that may provide more insight into the processes taking place. Thus, more complicated equations for kinetic modeling of substrate inhibition kinetics have been derived (e.g., Eq. 13) that allow for the estimation of interaction factors and additional kinetic parameters.

$$v = \frac{V_m\left(\frac{1}{K_S} + \frac{\beta \cdot [S]}{\alpha K_i K_S}\right)}{\frac{1}{[S]} + \frac{1}{K_S} + \frac{1}{K_i} + \frac{[S]}{\alpha K_S K_i}} \tag{13}$$

In the case of Eq. 13, $K_S \approx K_m$, $K_i$ is the dissociation constant of substrate binding to the inhibitory site, $\alpha$ is the factor by which the dissociation ($K_S$ and $K_i$) of substrate at both sites changes when a second substrate is bound, and $\beta$ is the factor by which $V_m$ changes when a second substrate is bound. Because of the number of parameters being estimated in Eq. 13, modeling with this

equation will generally require substantially more data points than required for Eq. 12.

For substrate inhibition mechanisms involving ordered ternary complex, such as the UGTs (case 2), Eq. 14 is applicable:

$$v = \frac{V_m[A][B]}{K_{iA}K_{mB} + K_{mB}[A] + K_{mA}[B] + [A][B]\left(1 + \frac{[B]}{K_{SiB}}\right)} \tag{14}$$

where $K_{SiB}$ is *not* a dissocation constant.

### 3.5 Explanations for Atypical Kinetics in P450 Enzymes and Their Detection

Reports on the actual mechanism(s) behind the activation kinetics of drug metabolizing enzymes (mainly P450) are few compared to the kinetic models that have been devised to fit nonhyperbolic velocity data [10–12]. The approaches that have been taken to understand the interplay of substrate, effector, and enzyme are, however, diverse and include isotope effects, fluorescence spectroscopy [13], and nuclear magnetic resonance [14, 15] in addition to assays for distinct steps in the P450 catalytic cycle [16]. One of the most influential hypotheses behind the action of effectors that prompted the use of these varied techniques is that the effector binding site could be located in the active site with the substrate. Over the past few years evidence for simultaneous ligand binding to the same active site has now been provided for at least two microsomal enzymes, human CYP3A4 and CYP2C9 , and three bacterial enzymes, CYP102A1 (BM3) , CYP107A1 (eryF) [17], and CYP158A2 [18] as discussed below. Of course, allosterism can result from binding at remote sites or through protein–protein interactions.

***Presence of Multiple Ligands Temporally Bound in the Same Active Site*** In the case of 3A4, the substrate pyrene could be used to demonstrate multiple ligand binding to the same enzyme active site during catalysis. Pyrene induces a sigmoidal kinetic response with increasing concentrations, and its planarity, coupled with its multiring fused aromatic structure, gives it a large surface for intermolecular pyrene–pyrene pi-stacking (a.k.a. excimer formation). After showing binding of pyrene to 3A4 by visible difference spectrophotometry and changes in fluorescence of pyrene and the enzyme, differences in the excitation spectra (absorbance of monomer and excimer in the ground state) gave multiple indications that there are excimers in the presence of enzyme distinct from those that form in solution. Crystal structures of 3A4 do demonstrate active volumes sufficiently large enough (>1000 Å) for multiple ligands in their closed conformations [19]. Some investigators have even suggested triply occupied active sites [20], but defining the nature of atypical kinetics in terms of different subsites still remains a challenge due to the dependency on the structure of both substrate and effector.

Shortly after the 3A4/pyrene report, a series of isotope effect experiments with P450 BM3 suggested that multiple fatty acids could bind to the same enzyme active site simultaneously. This explained why the F87A mutant produced a different metabolic profile of hydroxylated products in the presence of laurate. Mutation of

F87 to alanine results in an increase in hydroxylation of the ω-1 position as laurate concentrations are raised. Since the alanine substitution gives a smaller side chain, and therefore more space, it is unclear why the ratio of metabolites would change unless the presence of a second substrate or the collapse of the enzyme constricts the available volume. This is where isotope effects appear to be capable of detecting multiple ligand binding without the need for fluorescent molecules. The theory is that given two enzyme-bound ligands in rapid equilibrium, mixing protio substrate (all hydrogens) and substrate substituted with deuteriums (the deuterium–carbon bond possesses a lower vibrational energy and requires more energy for abstraction) at the site of metabolism will favor metabolism of the nonlabeled substrate. Hence, a bias in the rate of labeled and un labeled substrate metabolism is produced and it favors the hydrogen-containing substrate (i.e., $k_H/k_D$ ratio is greater than unity). In this case, the bias for ω-1 hydroxylation of palmitate was increased even more when deuterated laurate was added. Further validation of this technique is awaited for its use as a powerful tool with possibly untapped potential.

There are now crystal structures for two of the three major families of human microsomal P450s, 2 and 3 [21], but none have captured multiple ligands in the same pocket. However, crystallography has provided convincing electron density for multiple ligand binding in the same pocket for two bacterial enzymes. While not exactly a high throughput technology, even with the use of robotics, crystallography has generated structures that are invaluable for the predictive modeling of drug–drug interactions. P450 eryF in complex with two molecules of 9-aminophenanthrene and, more recently, P450 158A2 with two molecules of flaviolin have been reported. Both structures show one ligand above the heme with the second ligand lying above the first. Interestingly, both ligands are aromatic and planar, possibly owing some of their binding to each other. As of now, there are no examples with two different ligands bound, nor does a single structure indicate the order of their binding or the extent at which they move about inside the enzyme.

With the help of crystallography, nuclear magnetic resonance spectroscopy (NMR) has the capability to provide more details about the sequence and location of binding without the need for assigning every resonance. Yoon et al. [15] have reported on the use of protein NMR spectroscopy to study the binding of 9-aminophenanthrene to P450 eryF using enzyme expressed with $^{15}$N-phenylalanine so that all phenylalanine residues would have detectable amide nitrogen nuclei. Titration of 9-aminophenanthrene to levels that are substoichiometric to enzyme altered the signal intensity for Phe residues that are different from those whose intensity changed in the presence of higher levels of 9-aminophenanthrene ([substrate] > [enzyme]). Further arguments regarding the time scale of on and off rates and their interpretation in terms of the mechanism of cooperative binding were also drawn from this study. Surely, the ability to distinguish two binding events at different locations inside the enzyme will greatly complement other analytical methods, and studies with human P450s will no doubt bring a more detailed understanding to drug–drug interactions induced by atypical kinetics.

An additional use of NMR involves the enhanced relaxation rate of substrate nuclei near the iron of the heme, which is paramagnetic in the ferric state [22]. Using P450 2C9, Hummel et al. [14] studied the proton relaxation times of the substrate flurbiprofen in the absence and presence of heteroactivator dapsone. Time-averaged distances of each substrate proton can be estimated by carrying out a $T_1$ relaxation

experiment both with the ferric enzyme and the ferrous carbon monoxide complex, which is not paramagnetic. $T_1$ refers to the longitudinal relaxation and refers to the time it takes for the irradiated substrate protons to recess back to alignment with the magnetic field. The site of hydroxylation (4 -proton) was found to be closer to the iron of the heme group in the presence of heteroactivator, suggesting that its hyperbolic activation may result from increased collision frequency of substrate with the reactive iron-oxene or enhanced spin state conversion via heme dehydration.

Another dimension to the action of dapsone on flurbiprofen metabolism by P450 2C9 was provided by a study that compared the coupling of NADPH consumption to metabolite formation versus shunting to the side products superoxide, $H_2O_2$, and water. Catalytic coupling can be defined as the fraction of metabolite formed per NADPH consumed and is thus a measure of enzyme efficiency distinct from $k_{cat}/K_m$. In addition to increasing the rate of flurbiprofen metabolite formation, the effector dapsone was found to stimulate the consumption of NADPH, but decrease the formation of $H_2O_2$. Overall, the coupling of flurbiprofen metabolism was increased. Therefore, the distance of a substrate relative to the heme iron as altered by effectors appears to have multiple consequences as shown for P450 2C9. Possible effector-mediated changes in the position of substrate could displace the water that occupies the sixth coordination position of the iron to increase its reduction potential, keep the substrate closer to the active oxidant, and therefore decrease uncoupling of the catalytic cycle.

### 3.6 Artifactual Sources of Atypical Kinetics

During the conduct of *in vitro* drug metabolism experiments, it is prudent to be alert for the possible occurrence of atypical kinetics. However, the researcher must also be aware that there are several potential artifactual causes of atypical kinetic profiles that may lead to a misintepretation of the data. Thus, extreme care must be taken to assure the validity of the results. The comments below address several of these potential sources of error and how to minimize their impact on the kinetic profiles observed.

One of the fundamental assumptions when conducting enzyme kinetic experiments is that the substrate (and effector concentration if drug–drug interactions are being studied) remains constant throughout the experiment. For substrates and effectors that exhibit substantial turnover, one must worry about substrate/effector depletion. This will result in constantly changing substrate/effector concentrations throughout the experiment and invalidate the assumption of steady state. Related to constantly changing (and decreasing) drug concentrations are potential sources of error when the actual substrate/effector concentration is lower than believed. This can occur if either compound undergoes substantial nonspecific binding to the incubation matrix (i.e., protein) or when the drug has very low solubility. In these cases, the free drug concentration (i.e., that available to interact with the enzyme) is substantially lower than believed, resulting in overestimates of kinetic parameters such as $K_m$, and $K_i$.

Although less of an issue in today's environment of widespread LC/MS/MS availability, insufficient analytical sensitivity to accurately and reproducibly quantitate metabolite production at the extreme low end of the substrate concentration–

velocity curve may result in erroneous assignment of the kinetic profile. This is most problematic in the case of sigmoidal kinetics, where inadequate sensitivity hinders determining the less steep slope during the initial portion of the curve. One must also be cautious when using multienzyme systems (e.g., human liver microsomes) for kinetic characterization since multiple enzymes are present. This is particularly problematic with respect to observations of biphasic kinetics, since more than one enzyme may be producing the metabolite. In this case, the kinetic phenomena observed are actually a hybrid of the concerted effects of multiple enzymes and result in *apparent* kinetic constants.

Finally, the kinetic profile observed may be affected by incubation conditions and preparation. Choice of buffer salt to be used in the incubation can result in differing kinetic profiles [23] and the presence of certain organic solvents (particularly at concentrations above 2%) have also been shown to activate enzyme kinetic processes [24]. Finally, the inclusion (or exclusion) of cytochrome-$b_5$ can result in different kinetic profiles observed [25]. Thus, one must pay particular attention to controls and conditions when making the assignment of atypical kinetic phenomena.

## 4 CONCLUSION

Proper estimation of kinetic parameters is critical to successful extrapolation of *in vitro* data to the *in vivo* situation. In addition to standard hyperbolic kinetics, sigmoidal (autoactivation), biphasic, substrate inhibition, and heterotropic activation have all been observed in the kinetic profiles of drug metabolism reactions. The most common hypothesis for the occurrence of these types of "atypical" kinetic profiles is based on data suggesting that two (or more) substrate molecules (or a substrate and effector molecule) bind within the enzyme active site simultaneously and alter the enzyme kinetics. The occurrence of these atypical kinetic profiles complicates the determination of *in vitro–in vivo* correlations and thus requires the use of more complex equations appropriate to the type of profile observed in order to properly estimate the kinetic parameters.

## ACKNOWLEDGMENT

This work was supported in part by a grant from the National Institutes of Health (#GM 063215).

## REFERENCES

1. Luukkanen L, Taskinen J, Kurkela M, Kostiainen R, Hirvonen J, Finel M. Kinetic characterization of the 1A subfamily of recombinant human UDP-glucuronosyltransferases. *Drug Metab Dispos* 2005;33:1017–1026.
2. Segel IH. *Rapid Equilibrium Bireactant and Terreactant Systems*. Hoboken, NJ: Wiley; 1975.

3. Shou M, Mei Q, Ettore MWJr , Dai R, Baillie TA, Rushmore TH. Sigmoidal kinetic model for two co-operative substrate-binding sites in a cytochrome P450 3A4 active site: an example of the metabolism of diazepam and its derivatives. *Biochem J* 1999;340(Pt 3):845–853.
4. Korzekwa KR, Krishnamachary N, Shou M, Ogai A, Parise RA, Rettie AE, Gonzalez FJ, Tracy TS. Evaluation of atypical cytochrome P450 kinetics with two-substrate models: evidence that multiple substrates can simultaneously bind to cytochrome P450 active sites. *Biochemistry* 1998;37:4137–4147.
5. Ueng YF, Kuwabara T, Chun YJ, Guengerich FP. Cooperativity in oxidations catalyzed by cytochrome P450 3A4. *Biochemistry* 1997;36:370–381.
6. Cornish-Bowden A. *Fundamentals of Enzyme Kinetics*. London: Portland Press; 2004.
7. Rock DA, Perkins BN, Wahlstrom J, Jones JP. A method for determining two substrates binding in the same active site of cytochrome P450BM3: an explanation of high energy omega product formation. *Arch Biochem Biophys* 2003;416:9–16.
8. Hutzler JM, Hauer MJ, Tracy TS. Dapsone activation of CYP2C9-mediated metabolism: evidence for activation of multiple substrates and a two-site model. *Drug Metab Dispos* 2001;29:1029–1034.
9. Lin Y, Lu P, Tang C, Mei Q, Sandig G, Rodrigues AD, Rushmore TH, Shou M. Substrate inhibition kinetics for cytochrome P450-catalyzed reactions. *Drug Metab Dispos* 2001;29:368–374.
10. Galetin A, Clarke SE, Houston JB. Quinidine and haloperidol as modifiers of CYP3A4 activity: multisite kinetic model approach. *Drug Metab Dispos* 2002;30:1512–1522.
11. Galetin A, Clarke SE, Houston JB. Multisite kinetic analysis of interactions between prototypical CYP3A4 subgroup substrates: midazolam, testosterone, and nifedipine. *Drug Metab Dispos* 2003;31:1108–1116.
12. Kenworthy KE, Clarke SE, Andrews J, Houston JB. Multisite kinetic models for CYP3A4: simultaneous activation and inhibition of diazepam and testosterone metabolism. *Drug Metab Dispos* 2001;29:1644–1651.
13. Dabrowski MJ, Schrag ML, Wienkers LC, Atkins WM. Pyrene–pyrene complexes at the active site of cytochrome P450 3A4: evidence for a multiple substrate binding site. *J Am Chem Soc* 2002;124:11866–11867.
14. Hummel MA, Gannett PM, Aguilar JS, Tracy TS. Effector-mediated alteration of substrate orientation in cytochrome P450 2C9. *Biochemistry* 2004;43:7207–7214.
15. Yoon MY, Campbell AP, Atkins WM. "Allosterism" in the elementary steps of the cytochrome P450 reaction cycle. *Drug Metab Rev* 2004;36:219–230.
16. Hutzler JM, Wienkers LC, Wahlstrom JL, Carlson TJ, Tracy TS. Activation of cytochrome P450 2C9-mediated metabolism: mechanistic evidence in support of kinetic observations. *Arch Biochem Biophys* 2003;410:16–24.
17. Cupp-Vickery J, Anderson R, Hatziris Z. Crystal structures of ligand complexes of P450eryF exhibiting homotropic cooperativity. *Proc Natl Acad Sci USA* 2002;97:3050–3055.
18. Zhao B, Guengerich FP, Bellamine A, Lamb DC, Izumikawa M, Lei L, Podust LM, Sundaramoorthy M, Kalaitzis JA, Reddy LM, Kelly SL, Moore BS, Stec D, Voehler M, Falck JR, Shimada T, Waterman MR. Binding of two flaviolin substrate molecules, oxidative coupling, and crystal structure of *Streptomyces coelicolor* A3(2) cytochrome P450 158A2. *J Biol Chem* 2005;280:11599–11607.
19. Scott EE, Halpert JR. Structures of cytochrome P450 3A4. *Trends Biochem Sci* 2005;30:5–7.

20. He YA, Roussel F, Halpert JR. Analysis of homotropic and heterotropic cooperativity of diazepam oxidation by CYP3A4 using site-directed mutagenesis and kinetic modeling. *Arch Biochem Biophys* 2003;409:92–101.
21. Johnson EF, Stout CD. Structural diversity of human xenobiotic-metabolizing cytochrome P450 monooxygenases. *Biochem Biophys Res Commun* 2005;338:331–336.
22. Yao H, Costache AD, Sem DS. Chemical proteomic tool for ligand mapping of CYP antitargets: an NMR-compatible 3D QSAR descriptor in the Heme-Based Coordinate System. *J Chem Inf Comput Sci* 2004;44:1456–1465.
23. Hutzler JM, Powers FJ, Wynalda MA, Wienkers LC. Effect of carbonate anion on cytochrome P450 2D6-mediated metabolism *in vitro*: the potential role of multiple oxygenating species. *Arch Biochem Biophys* 2003;417:165–175.
24. Hickman D, Wang JP, Wang Y, Unadkat JD. Evaluation of the selectivity of *in vitro* probes and suitability of organic solvents for the measurement of human cytochrome P450 monooxygenase activities. *Drug Metab Dispos* 1998;26:207–215.
25. Jushchyshyn MI, Hutzler JM, Schrag ML, Wienkers LC. Catalytic turnover of pyrene by CYP3A4: evidence that cytochrome b5 directly induces positive cooperativity. *Arch Biochem Biophys* 2005;438:21–28.

# DRUG CLEARANCE

Sree D. Panuganti[1] and Craig K. Svensson[2]
[1]*Purdue University, West Lafayette, Indiana*
[2]*Osmetech Molecular Diagnostics, Pasadena, California*

**Contents**

1 Introduction
2 Routes of Drug Elimination
  2.1 Potential Routes of Elimination
  2.2 Characterization of Elimination
3 Renal Clearance
  3.1 Basic Kidney Structure and Function
  3.2 Mechanisms of Renal Clearance
  3.3 Determination of Renal Clearance and Mechanisms of Renal Clearance
  3.4 Factors Influencing Renal Clearance
4 Hepatic Clearance
  4.1 Basic Liver Structure and Function
  4.2 Mechanism of Drug Transport Across the Hepatocytes
  4.3 Role of Liver in First-Pass Metabolism
  4.4 Biliary Clearance
  4.5 Determination of Hepatic Clearance
  4.6 Models of Hepatic Clearance
5 Effect of Disease on Drug Clearance
  References

## 1 INTRODUCTION

Drug discovery and development is a complex process involving various sequential stages broadly grouped as preclinical and clinical evaluation. Both phases require the characterization of the pharmacokinetics of the molecule being evaluated as a

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

potential drug candidate. In drug discovery and development, it is of particular interest to understand the pharmacokinetic behavior of potential drug candidates as early as possible in order to select the most promising compounds for further development while not wasting unnecessary resources on those compounds that will not prove to be viable for clinical use. While many pharmacokinetic processes contribute to the viability of drug candidates, clearance processes arguably have the greatest impact on determining whether or not a compound possesses the pharmacokinetic characteristics amenable to further development of the candidate. The clearance of a drug candidate is frequently the primary factor in determining the bioavailability, dosing frequency, duration of action, and viable routes of administration for the clinical use of the compound. Thus, the rate and routes of drug clearance can be critical in determining whether or not the development of a candidate compound should move forward. It is therefore imperative that development scientists possess an understanding of the determinants and characterization of drug clearance. Our objective in this chapter is to provide an overview of clearance concepts, ranging from a description of the routes by which drugs are cleared from the body to pathophysiological conditions that may alter these processes qualitatively and quantitatively. In subsequent chapters, authors will describe the importance of drug clearance in the overall context of pharmacokinetic analysis.

## 2 ROUTES OF DRUG ELIMINATION

Avoidance of accumulation of drugs subsequent to repeated exposure necessitates varied pathways for elimination of such compounds from the body. The rate of elimination will be a key determinant of the time for which critical body sites are exposed to effective concentrations of a drug. Hence, the duration of pharmacologic effect *in vivo* is generally determined by the rate of elimination. The routes of elimination are important in determining the likelihood of altered drug disposition in various disease states, as well as the potential for drug–drug interactions. Based on the anticipated patient population for a new chemical entity, it may be preferable to have a drug that is primarily eliminated by the liver, as opposed to the renal route, as the primary pathway of elimination. The importance of such factors in assessment of compounds under development makes it imperative that development scientists possess a good understanding of the potential pathways and means for quantifying drug elimination.

### 2.1 Potential Routes of Elimination

Mechanisms of drug elimination are largely determined by the physiochemical characteristics of the ingested compounds. Small water-soluble compounds are readily eliminated unchanged by the kidney, while large molecules and lipophilic agents must undergo biotransformation in the liver prior to being excreted by the kidney. While generally quantitatively minor, elimination by alternative routes may be of clinical/practical importance.

***Renal Elimination*** It may be argued that essentially all drugs are eliminated to some degree via the renal route. While large lipophilic molecules are not directly

eliminated by the kidney, their biotransformation products (metabolites) are generally more water soluble and subject to renal elimination. This reality is important to recognize when considering the potential impact of renal disease on drug elimination. While the elimination of a lipophilic or large molecular weight parent compound may not be altered in the presence of renal disease, the elimination of its metabolites may be altered significantly. If those metabolites exhibit pharmacologic or toxicologic effects, renal disease may alter patient response, despite the lack of alterations in the disposition of the parent compound. As it represents one of the two most important pathways of drug elimination, renal elimination will be considered in further detail in a subsequent section.

***Hepatic Elimination*** The liver is an organ of critical importance for drug elimination. This is in part due to the fact that disease- and drug-induced alterations in metabolic capacity can have such a profound impact on the ability of the liver to metabolize drugs. It is, however, important to recognize that the liver also eliminates numerous drugs unchanged through its capacity for biliary excretion. As one of the two major organs for drug elimination (the other being the kidney), concepts of hepatic clearance are covered in further depth in a subsequent section.

***Pulmonary Elimination*** As the general population is aware of the use of ethanol measurements in exhaled breath to approximate blood ethanol concentrations, the ability to put pulmonary excretion of xenobiotics to practical use is obvious. While pulmonary elimination is a minor route for ethanol elimination, the constant ratio of alveolar to blood ethanol concentration permits extrapolation of the latter from the former. The number of compounds for which the pulmonary route is quantitatively significant in terms of the overall elimination of the drug is small and mostly limited to gaseous anesthetics. The ability to measure $^{13}CO_2$ or $^{14}CO_2$ in excreted breath has been utilized as an indirect means of measuring drug elimination for compounds that undergo oxidative metabolism in the liver. Several such tests are commercially available as a means of phenotyping the metabolic capacity of patients via specific cytochrome P450s (e.g., CYP3A4). These tests are dependent on the one-carbon oxidation that gives rise to $CO_2$, which is then excreted in exhaled breath.

***Other Routes of Elimination*** While also quantitatively minor, additional routes of elimination (e.g., tears, sweat, hair) may be of practical or clinical significance. For example, while the fraction of a dose of drug that is excreted in tears is quite small, it may have very important implications for patients. Rifampin is a compound that is eliminated in tears and is capable of staining contacts a yellow-orange color when worn by patients during therapy with this drug. Another pathway of elimination that is of minor quantitative importance is hair. However, measurement of drug concentration in hair permits assessment of drug exposure over a longer time period than possible with blood or urine. This fact is applied in the numerous commercially available kits to measure drugs of abuse in hair samples. Its noninvasive nature makes this means of exposure assessment attractive.

### 2.2 Characterization of Elimination

There are two primary ways in which drug elimination may be quantified. The first is through estimation of the *half-life* of a drug in some reference fluid (e.g., plasma).

The half-life represents the time required for 50% of the drug to be eliminated from the body. For drugs that are eliminated in a first-order fashion, this time is independent of dose or concentration. However, since many drugs are eliminated by processes that involve the direct interaction of drug with a macromolecule (e.g., an enzyme or transporter protein), large doses will result in saturation of enzymatic or transport process and give rise to altered half-lives (compared to that observed with lower doses). This is important to keep in mind in preclinical studies where large doses may be utilized. The importance of dose-dependent pharmacokinetics and the application in toxicokinetics is discussed in other chapters (see Chapters 31 and 36 in this volume).

The most useful parameter for quantifying drug elimination *in vivo* is *clearance*. Assessment of elimination using clearance is preferred over that of half-life, since the latter is a function of both clearance and distribution (see Chapters 31 and 36 in this volume). For this reason, half-life fails to provide insight into the mechanism of drug elimination and cannot be used to predict the likely impact of disease or drug interactions on drug elimination. Clearance is defined as *the volume of blood from which all of the drug would appear to be removed per unit time*. As such, it is expressed in terms of volume per unit time and may be related to real physiological volumes, such as hepatic blood flow or renal plasma flow.

An understanding of the concept of clearance can be gained by consideration of the perfusion of an isolated organ with blood containing drug (Fig. 1). The rate of drug entry into and out of the organ can be described as

$$\text{Rate in} = QC_a$$
$$\text{Rate out} = QC_v$$

where $Q$ is the organ perfusion rate, $C_a$ is the concentration of drug in arterial blood, and $C_v$ is the concentration of drug in venous blood. Whenever $C_v < C_a$, elimination of drug occurs in the organ and it is referred to as a clearing organ. The rate of elimination can be determined as

$$\text{Rate of elimination} = QC_a - QC_v = Q(C_a - C_v)$$

A common means for describing the efficiency by which a single organ removes a drug from blood is the *extraction ratio (E)*, where

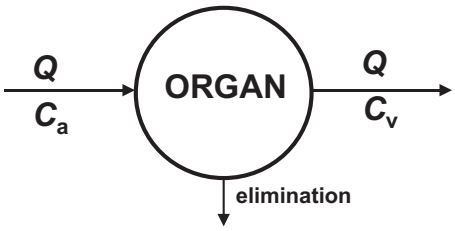


**FIGURE 1** Schematic representation of a clearing organ. $C_a$ designates arterial drug concentration, $C_v$ is the venous drug concentration, and $Q$ represents the blood flow to the organ. This schematic assumes that blood flow across the organ does not change significantly.

$$E = \frac{\text{Rate of elimination}}{\text{Rate of entry}}$$

$$E = \frac{Q(C_a - C_v)}{QC_a} = \frac{C_a - C_v}{C_a}$$

The clearance ($CL$) of the organ for a specific compound can be given as

$$CL = QE = Q\left(\frac{C_a - C_v}{C_a}\right)$$

The above equation reveals two potential extremes for $CL$. The first is when $E \rightarrow 1$, under which conditions $CL \sim Q$. Hence, for a drug that displays a high extraction ratio, the clearance is said to be *perfusion rate limited*; meaning that the rate-limiting step in the clearance is the delivery of drug to the clearing organ. Under such conditions, changes in organ perfusion rate will result in proportional changes in drug clearance. In contrast, when $E$ is very low, changes in perfusion have little impact on drug clearance.

From the above equations, one can readily develop the means to calculate the total body clearance for a drug, $CL_T$, where

$$CL_T = \frac{\text{Elimination rate}}{\text{Concentration in blood}} = \frac{dX/dt}{C}$$

$$CL_T = \frac{\int_0^\infty (dX/dt)dt}{\int_0^\infty C\,dt}$$

where

$$\int_0^\infty \frac{dX}{dt}dt = \text{Total amount of drug eliminated}$$

and

$$\int_0^\infty C\,dt = AUC_0^\infty$$

where $AUC$ is the area under the drug concentration versus time curve, $X$ is the amount of drug in the body, and $t$ is time. When the drug is administered intravenously, 100% of the drug enters the systemic circulation. Therefore, after intravenous administration, the total amount of drug ultimately eliminated is equal to the intravenous dose ($D_{IV}$). Hence,

$$CL_T = \frac{D_{IV}}{AUC_0^\infty}$$

From the above equation, it can be seen that the total clearance (sometimes also referred to as the *systemic* clearance) can readily be determined after intravenous

(IV) administration of a drug. The total clearance can be determined after non-intravenous routes of administration by accounting for the fraction of drug that reaches the systemic circulation (described in more detail in Chapter 36 in this volume).

An important principle related to the use and determination of organ-specific clearance is the additivity of clearance. Specifically, the total clearance is the sum of the individual organ clearance such that, for example,

$$\text{Total clearance} = \text{Renal clearance} + \text{Hepatic clearance}$$

This additivity arises from the fact that organs of elimination are fractionally perfused in parallel. An exception to this principle is pulmonary clearance, which arises as a result of the fact that the entire cardiac output traverses the pulmonary bed prior to being fractionally distributed to other organs of elimination (e.g., liver and kidney).

## 3 RENAL CLEARANCE

### 3.1 Basic Kidney Structure and Function

The most important functions of the kidney include regulation of body water content, mineral composition, and acidity. The kidney also plays an important role in the removal of endogenous metabolic waste products and xenobiotics and their metabolites from the body. Indeed, the major organ in mammals for the clearance of drugs and their metabolites is the kidney. Humans have two kidneys located in the upper rear region of the abdominal cavity. The functional units of kidney are called nephrons and each kidney is composed of over one million nephrons. Figure 2 illustrates different components of the nephron.

Each nephron consists of vascular and tubular components. The glomerulus is made up of a network of glomerular capillaries and Bowman's capsule, which is a

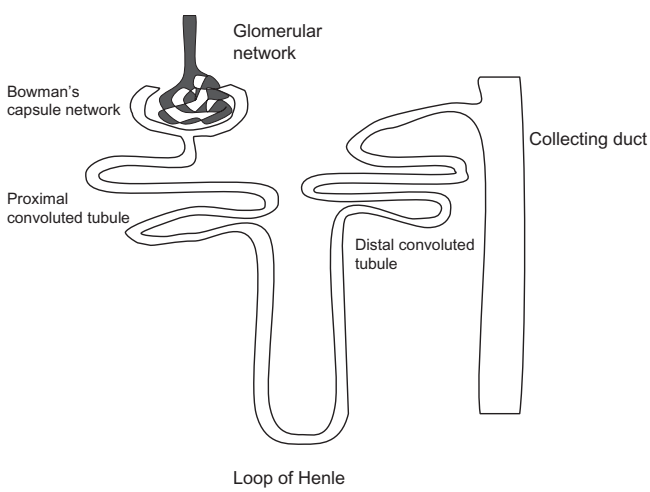


**FIGURE 2** Components of the renal nephron.

thin double membrane that surrounds the glomerulus and functions as a filter. Blood is filtered at the glomerulus so that only protein-free filtrate is allowed to pass into the nephron. The resulting glomerular filtrate travels through the proximal convoluted tubule, where approximately 80% of the water is reabsorbed. The loop of Henle connects the proximal convoluted tubule to the distal convoluted tubule. A large collecting duct collects kidney filtrate from the distal convoluted tubules of each nephron and carries it to the bladder.

The kidney can be divided into two portions histologically: the outer renal cortex and the inner renal medulla. The renal cortex contains the glomeruli, proximal and distal convoluted tubules, the outer portion of the loop of Henle, and collecting ducts, whereas the renal medulla contains the lower ends of the loop of Henle and collecting ducts.

### 3.2 Mechanisms of Renal Clearance

The kidney excretes endogenous and xenobiotic compounds and their metabolites by three distinct mechanisms: glomerular filtration, tubular secretion, and tubular reabsorption. Glomerular filtration and tubular secretion remove drugs from the circulation, whereas tubular reabsorption can be considered as a redistribution mechanism by which drug moves from the inner tubule back into the circulation. While drugs may also be metabolized in the kidney, the quantitative contribution of this route is insignificant in terms of the overall clearance of drugs.

***Glomerular Filtration*** Glomerular filtration is a passive process. In this process drugs and other endogenous compounds of molecular weight less than 6 kDa are filtered effectively. Drug molecules that are bound to plasma proteins, such as albumin, will not be filtered. In other words, only unbound drug in plasma water is available for glomerular filtration. The glomerular filtration rate (GFR) can be determined using marker compounds that undergo filtration only (i.e., do not undergo secretion or reabsorption) and are not bound to plasma proteins. The renal clearance of such compounds is equal to the GFR. The most common marker compounds used to measure GFR are inulin (MW 5200) and creatinine (MW 131). The rate of filtration of a drug can be given as

$$\text{Rate of filtration} = GFR \times C_{\text{fup}}$$

where $C_{\text{fup}}$ is the concentration of free (unbound) drug in plasma. Furthermore, since $C_{\text{fup}} = f_{\text{up}} \times C$,

$$\text{Rate of filtration} = f_{\text{up}} \times GFR \times C$$

If drug is only filtered and all of the drug filtered is excreted in the urine, the rate of drug excretion will be equal to the rate of filtration. Hence

$$CL_{\text{R}} = \frac{\text{Rate of excretion}}{\text{Plasma concentration}} = \frac{f_{\text{up}} \times GFR \times C}{C} = f_{\text{up}} \times GFR$$

Thus the renal clearance of a compound that is only filtered can be determined based on a knowledge of the protein binding and the GFR (as approximated by inulin or creatinine).

The glomerular filtration rate varies from individual to individual, but in healthy subjects ranges from 110 to 130 mL/min (~180 L/day). While single-point determinations of GFR are often used to estimate drug clearance, it should be recognized that GFR is subject to circadian variation. GFR is highest in the active phase (morning) and lowest in the inactive phase (evening), with an amplitude of variation of about 20–30% [1, 2].

***Tubular Secretion*** The kidney receives about 10–25% of the cardiac output, or about 700–1500 mL/min, assuming a cardiac output of 6000 mL/min [3], of which 60% will be the plasma flow. Approximately 18% of the effective plasma flow is filtered at the glomeruli. The remaining renal blood bathing the tubules permits access to the secretory processes in these tubules.

When the renal clearance of a drug exceeds $f_{up} \times GFR$, the drug must undergo active tubular secretion. There are two distinct systems involved in tubular secretion: one for organic anions (acids) and one for organic cation (bases). These systems are made up of multiple transporters. Both the anion and cation transport systems are located in the proximal convoluted tubule of the nephron. The most important transporters from a drug interaction perspective are the organic anion transporters. This process is thought to be localized primarily at the basolateral membrane of the proximal convoluted tubule. The transport of the organic anions across this membrane occurs through a three-step process (Fig. 3a).

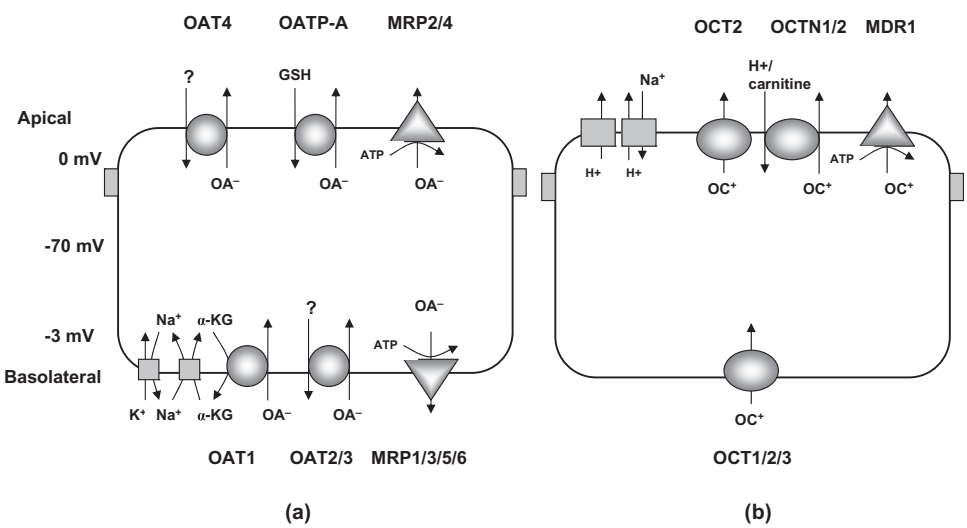


**FIGURE 3** Functional and proposed models of organic anion (a) and cation (b) transporters in renal tubules. $OA^-$, organic anion; OAT, organic anion transporter; OATP, organic anion transporting polypeptide; $OC^+$, organic cation; OCT, OCTN, organic cation transporter; MRP, multidrug resistance associated protein; MDR, multidrug resistance P-glycoprotein. (Adapted from Ref. 19.)

The organic anion transport system (OAT1) is dependent on a cell membrane Na/K-ATPase that maintains the intracellular electrochemical gradient. Initially, ATP is hydrolyzed to drive the sodium pump. In the second step, the resulting $Na^+$ gradient drives α-ketogluterate (α-KG) cotransport across the basolateral membrane. The third step in this process is the exchange of α-KG with the organic anion ($OA^-$), as α-KG markedly stimulates the transport of organic anions. Once inside the proximal tubule cell, the organic acid is eliminated on the luminal side of the cell by a carrier-mediated transport system [4–6].

At present, it is difficult to predict whether a new drug candidate will interact with renal tubular secretion mechanisms based on structure–activity relationships alone. However, certain generalizations can be made based on transport studies using analogue compounds with different characteristics. In general, substrates with two negative charges will interact with anion transport systems, which are influenced by the molecular distance between the charges. Substances that interact with the organic anion transport system must have a hydrophobic area and two partial negative charges, a hydrophobic area and one negative ionic charge, or a large hydrophobic area and a partial negative charge [3].

Comparatively less is known about the organic cation transport system. This is partly due to the toxic side effects of the marker compounds on the cardiovascular system that could be used to study the transport process. As the organic anion transporters are responsible for the tubular secretion of organic acids, the organic cation transport systems are responsible for weak bases; often nitrogen-containing compounds that are ionized at physiological pH. Similar to anion transport, the organic cation transport also proceeds in three steps (Fig. 3b). In contrast to filtration, protein-bound toxicants are available for active transport. Binding to plasma proteins usually does not hinder tubular secretion of drugs because of the dynamic equilibrium that exists between free and bound drug. As free drug is removed and transported across the tubular epithelium, immediate dissociation of the drug–protein complex usually occurs [7, 8].

Since tubular secretion is an active process, there may be competitive inhibition of the secretion of one compound by coadministration of a second compound secreted via the same transporter. This can actually be taken advantage of clinically to prolong the systemic exposure to a drug readily secreted by the kidney. For example, probenecid inhibits the secretion of penicillin and coadministration of these two agents can be used to decrease the dose and frequency of penicillin that needs to be administered.

***Tubular Reabsorption*** When the $CL_R < f_{up} \times GFR$, it is apparent that a portion of the filtered drug must undergo reabsorption. Most of the filtrate that enters the renal tubules is reabsorbed, with urine output accounting for less than 1% of the filtrate fraction that enters the tubules. This creates a large concentration gradient across the tubule and plasma, representing a driving force for passive reabsorption of solutes back into the circulation. Small molecular weight substances that are sufficiently lipophilic and un-ionized will be reabsorbed. In contrast, compounds that have large molecular weight are polar or ionized will not be reabsorbed. As many drugs are either weak bases or acids, the pH of the glomerular filtrate can greatly influence the extent of tubular reabsorption. Even small changes in pH of filtrate, either because of diet or drugs, can result in significant variations in the percentage

of drug reabsorbed or excreted. For example, the p$K_a$ of phenobarbital (a weak acid) is 7.2 with an ionic strength equal to plasma. At a plasma pH of 7.4, the drug exists in ionized form and at pH values less than 7.2 the un-ionized form of the drug dominates and will be excreted. In the case of a drug overdose, it is possible to increase the excretion of some drugs by suitable adjustment of urine pH.

### 3.3 Determination of Renal Clearance and Mechanisms of Renal Clearance

***General Model for Renal Clearance*** The total renal clearance of a drug is the sum of all the renal elimination mechanisms involved in the elimination of that specific drug:

$$CL_R = CL_{Fil} + CL_{ATS} - CL_{TR}$$

where $CL_{Fil}$ is the filtration clearance, $CL_{ATS}$ is the active tubular secretion clearance, and $CL_{TR}$ is the reabsorption clearance [3, 9]. As indicated previously, the filtration clearance of a drug is given as $f_{up} \times GFR$, while the active tubular secretion clearance can be described as

$$CL_{ATS} = \frac{Q_{RPF} f_{up} CL_{usint}}{Q_{RPF} + f_{up} CL_{usint}}$$

where $Q_{RPF}$ is renal plasma flow and $CL_{usint}$ is secretory clearance of unbound drug. Since the amount of drug reabsorbed represents a fraction of that which is filtered and secreted ($F_{TR}$), the renal clearance can be expressed as

$$CL_R = f_{up} GFR + \frac{Q_{RPF} f_{up} CL_{usint}}{Q_{RPF} + f_{up} CL_{usint}} - F_{TR}\left( f_{up} GFR + \frac{Q_{RPF} f_{up} CL_{usint}}{Q_{RPF} + f_{up} CL_{usint}} \right)$$

The above equation represents the complex relationship between plasma protein binding, renal plasma flow, glomerular filtration rate, and intrinsic secretory activity. More complex models that take into account urine flow, pH gradient, and transport saturation are needed to fully account for the many physiological variables governing renal clearance. While this complexity may leave the reader with the impression that determination and assessment of renal clearance is hopelessly complex, such is not the case. Simple measurement of renal clearance can provide valuable insight into mechanisms when compared with the renal clearance of marker compounds, as described in subsequent sections.

***Determination of Renal Clearance from Urine and Plasma Concentrations*** There are numerous methods available for calculation of renal clearance from measurements of plasma and urine drug concentrations. One common method is based on knowledge of the additivity of clearance. After a single intravenous dose of a drug, the fraction eliminated renally ($f_R$) is given as

$$f_R = \frac{Ae^{\infty}}{D_{IV}}$$

where $Ae^{\infty}$ is the total amount of drug excreted unchanged in urine. Determination of this value necessitates the collection of urine until all the drug has been eliminated from the body. If blood samples have also been obtained so that total clearance can be calculated as described previously, the renal clearance is determined as

$$CL_R = f_R \times CL_T$$

Alternatively, using the same data, one can calculate renal clearance from the relationship

$$CL_R = \frac{Ae^{\infty}}{AUC_0^{\infty}}$$

Renal clearance can also be calculated using excretion rate determinations rather than total drug excretion, where

$$CL_R = \frac{\Delta Ae/\Delta t}{C_{mid}}$$

$C_{mid}$ represents the plasma concentration at the midpoint of the urine collection, $\Delta Ae$ is the amount of drug excreted in urine unchanged over the collection interval, and $\Delta t$ is the collection interval. As the renal clearance may vary somewhat between collection intervals, the best way to utilize this method is to determine the excretion rate over several different intervals. This can then be plotted against the midpoint concentration for each excretion rate. The result will provide a straight line with a slope equal to the renal clearance. Alternatively, if the drug is administered as a constant infusion until steady state is achieved,

$$CL_R = \frac{(\Delta Ae/\Delta t)_{ss}}{C_{ss}}$$

***Assessment of Mechanism(s) of Renal Elimination from Renal Clearance Determinations*** Renal clearance estimates are commonly used to determine the mechanism of renal elimination of compounds. If the renal clearance is approximately equal to $f_{up} \times$ GFR, filtration is the predominant mechanism. If renal clearance is less than $f_{up} \times GFR$, tubular reabsorption must occur, while a renal clearance greater than $f_{up} \times GFR$ indicates that tubular secretion is occurring in addition to filtration. Since renal clearance is the sum of all renal elimination mechanisms, a renal clearance equal to $f_{up} \times GFR$ does not rule out other mechanisms of renal elimination, as it is possible that secretion and reabsorption occur to an equal extent and thus negate one another.

As indicated previously, the most commonly used markers for estimating GFR are inulin and endogenous creatinine [1]. Inulin is one of the most accurate methods to estimate GFR. However, determination of inulin clearance is not practical for routine clinical use as it requires the infusion of an exogenous substance. In contrast, estimation of GFR via creatinine does not involve the administration of an additional substance.

Experimentally, concurrent determination of the renal clearance of a drug of interest and the clearance of inulin can be used to probe the mechanisms of renal elimination for the compound of interest. Once the renal clearance values are determined, the ratio of drug clearance to inulin clearance is determined, as shown in Table 1. Adjustments for plasma protein binding may need to be taken into account when using this method for elucidating the mechanisms of drug elimination.

As a means of more routine estimation of GFR, clearance of the endogenous marker creatinine can be determined by

$$CL_{\mathrm{cr}} = \frac{U_{\mathrm{cr}} \times V}{S_{\mathrm{cr}}}$$

where $U_{\mathrm{cr}}$ is the urine creatinine concentration, $V$ is the urine flow rate, and $S_{\mathrm{cr}}$ is the serum creatinine concentration sampled at the midpoint of the urinary collection period.

Creatinine clearance is often considered as an ideal method to assess GFR because (1) daily creatinine production is constant, (2) serum and urinary creatinine concentrations are easily measured in the laboratory, and (3) by making the urinary collection period sufficiently long, problems associated with short collection intervals can be avoided [10, 11]. Although the results obtained by creatinine clearance estimates correlate well with estimates of GFR, some deviation from GFR is to be expected, especially at very high and very low GFR values [12]. For most applications of GFR estimated through creatinine clearance, this deviation is insignificant. Creatinine clearance can also be estimated through use of $S_{\mathrm{cr}}$, body weight, height, and sex. Such estimates are particularly useful for determining initial dosage regimens for drugs in a clinical setting.

A drug that is highly extracted by the kidney, meaning it undergoes extensive active tubular secretion, will exhibit a renal clearance that approximates renal plasma flow. A widely used model compound that exhibits these characteristics is *p*-aminohippuric acid (PAH). PAH has been widely used to determine the impact of interventions of renal plasma flow and to compare the renal clearance of a compound to the renal plasma flow in the same subjects.

**TABLE 1 Use of Clearance Ratios to Determine Mechanisms of Renal Clearance**

| Clearance Ratio | Probable Mechanisms |
|---|---|
| $\frac{CL_{\mathrm{drug}}}{CL_{\mathrm{inulin}}} < 1$ | Filtration and reabsorption |
| $\frac{CL_{\mathrm{drug}}}{CL_{\mathrm{inulin}}} = 1$ | Filtration |
| $\frac{CL_{\mathrm{drug}}}{CL_{\mathrm{inulin}}} > 1$ | Filtration and active tubular secretion |

While a renal clearance less than $f_{up} \times GFR$ is indicative of tubular reabsorption, confirmation of this mechanism can be obtained by perturbations in urine flow and pH. As the concentration gradient is the driving force for reabsorption, increasing urine flow in the presence of reabsorptive mechanisms will result in an increase in the renal clearance. While pH manipulations may alter the renal tubular reabsorption of some drugs, drugs whose ionization does not change within the range of physiologically achievable urine pH (~5–8) will not exhibit an altered renal clearance despite changes in urine pH. Hence, the observation of pH-dependent renal clearance for ionizable compounds confirms the presence of tubular reabsorption, but such changes will not be expected for neutral compounds.

### 3.4 Factors Influencing Renal Clearance

Factors that influence the ability of the kidney to eliminate drugs may cause marked changes in the pharmacokinetics of some compounds. Such factors include age, disease, plasma protein binding, drug concentration or dose, and drug interactions at the site of elimination.

***Age*** Many functions of the kidney are not fully developed at birth. As a result, some xenobiotics are eliminated more slowly in newborns than in adults. Similarly, the aging kidney exhibits reductions both in mass and in GFR. These changes may be associated with other age-related conditions, such as atherosclerosis and hypertension, which reduce renal blood flow. An age-dependent decline in the number of nephrons also occurs. The decline in GFR is in turn responsible for the reduced renal clearance of drugs that are normally removed by the kidney. With age the capacity to excrete acid load is diminished independent of decreased GFR. The diminished capacity of older subjects to excrete acid load could be due to an insufficient decrease in the tubular reabsorption.

***Drug–Protein Binding*** The degree of plasma protein binding affects the filtration of drugs, as protein–xenobiotic complexes are too large to pass through the pores of the glomeruli. Whether or not changes in plasma protein binding alter renal clearance is dependent on the renal extraction ratio for the drug in question. For example, furosemide exhibits a low extraction ratio and its renal clearance changes proportionally to the free fraction of drug in plasma [13]. In contrast, the renal clearance of drugs that exhibit a high renal extraction ratio, and are therefore extensively secreted, is not significantly altered by changes in protein binding. Binding to plasma proteins usually does not hinder tubular secretion of drugs because of the rapid equilibrium that exists between free and bound drug. As free drug is removed and transported across the tubular epithelium, immediate dissociation of the drug–protein complex usually occurs—freeing more drug for secretion.

***Concentration or Dose*** As tubular secretion involves the interaction of drug with transport carriers, which are limited in number, there is the potential for saturation of these membrane transport proteins as the dose or concentration of drug is increased. In the absence of saturation, the renal clearance of a drug will be independent of dose or concentration. However, when doses are administered that give rise to concentrations that saturate the transporters, the renal clearance will decrease.

Similarly, if the tubular reabsorption is by a carrier-mediated transport process (as is the case for ascorbic acid), increasing concentrations will saturate the carriers, resulting in an *increase* in the renal clearance (since the fraction reabsorbed will be reduced and more drug will be excreted in the urine). Such saturation has practical implications for the dosing of compounds that display saturation within the range of therapeutic concentrations. It is also important for the development scientist to keep the potential for such dose- or concentration-dependent changes in mind when utilizing high doses in preclinical studies.

A limited number of drugs exhibit concentration-dependent plasma protein binding at concentrations in the range achieved after therapeutic doses. An increase in the free fraction of a drug as the dose/concentration is increased will result in an increase in the filtration clearance (since this value is equal to $f_{up} \times GFR$). In this scenario, increasing the dose of a drug may fail to give rise to the expected increases in steady-state concentration since clearance is increased with dose. Several nonsteroidal anti-inflammatory drugs, such as naproxen, exhibit this type of dose/concentration-dependent changes in renal clearance.

***Drug Interactions*** The most well known drug–drug interactions that occur at the level of renal elimination are those in which one drug acts as a competitive inhibitor for another drug at a renal transport site. Competitive inhibitors will decrease the renal secretion of other substrates, but will not impact the filtration or extent of reabsorption that occurs passively. Such inhibition will result in a decrease in the clearance of a drug, which in turn increases the steady-state plasma concentration and results in a prolonged half-life of the drug.

Other sites of drug–drug interaction, though less common, can also affect the renal elimination of drugs. For example, antacids can alter urine pH and change the fraction of an ionizable drug that undergoes tubular reabsorption. Drugs that inhibit the renal metabolism of drugs represent another mechanism of potential interaction.

## 4 HEPATIC CLEARANCE

### 4.1 Basic Liver Structure and Function

The liver is the second largest organ in the body, weighing about 1.5 kg. The liver is located in the right upper quadrant of the abdomen, just below the diaphragm. Main functions of the liver include bile production, detoxification of waste products, storage of iron, vitamins, and trace elements, and the biotransformation of toxic compounds for excretion by the kidneys. Because of its location and ability to efficiently extract a wide variety of compounds from the portal circulation, the liver plays a central role in removing toxic materials before their entry into the systemic circulation. The extracting ability of the liver may result in a substantial decrease in the systemic availability of drugs after oral administration. In addition to its ability to biotransform a wide variety of toxicants, the liver may also directly excrete drugs and/or their metabolites into bile. Thus, the liver possesses two mechanisms for xenobiotic elimination—biotransformation and biliary excretion.

The liver is anatomically divided into two lobes, right and left. Each lobe is further divided into approximately one million lobules. These hepatic lobules are the functional units of the liver. The liver is composed of two types of epithelial cells: hepatocytes, which account for approximately 80% of the nuclear population, and cholangiocytes (epithelial cells that line intrahepatic bile ducts), which account for 3–5% of the liver cell population. A lobule is a hexagonal arrangement of plates of hepatocytes radiating outward from a central vein in the center. Roughly 75% of blood entering the liver arises from venous blood draining into the portal vein. Importantly, all of the venous blood returning from the small intestine, stomach, pancreas, and spleen converges into the portal vein, as does a portion of that which drains from the large intestine. The remaining 25% of the blood supply to the liver is arterial blood delivered through the hepatic artery (Fig. 4). The hepatic vascular system has several unique characteristics relative to other organs.

Terminal branches of the hepatic portal vein and hepatic artery empty together and mix as they enter sinusoids in the liver. Hence, the sinusoids are exposed to drug that originates from the gastrointestinal system as well as from the systemic circulation; and such exposure occurs simultaneously. This fact may have important implications in modeling hepatic clearance. The sinusoids are distensible vascular channels bounded circumferentially by hepatocytes. Blood flows through the sinusoids and empties into the central vein of each lobule. As blood flows through the sinusoids, nutrients and other endogenous and exogenous compounds are distributed into the hepatocytes. Hepatocytes are arranged in plates with their apical surfaces facing and surrounding the sinusoids. The basal faces of adjoining hepatocytes are welded together by junctional complexes to form canaliculi, the first channel in the biliary system. Hepatocytes secrete bile into the canaliculi, and those secretions flow antiparallel to the blood flow in sinusoids. At the ends of the

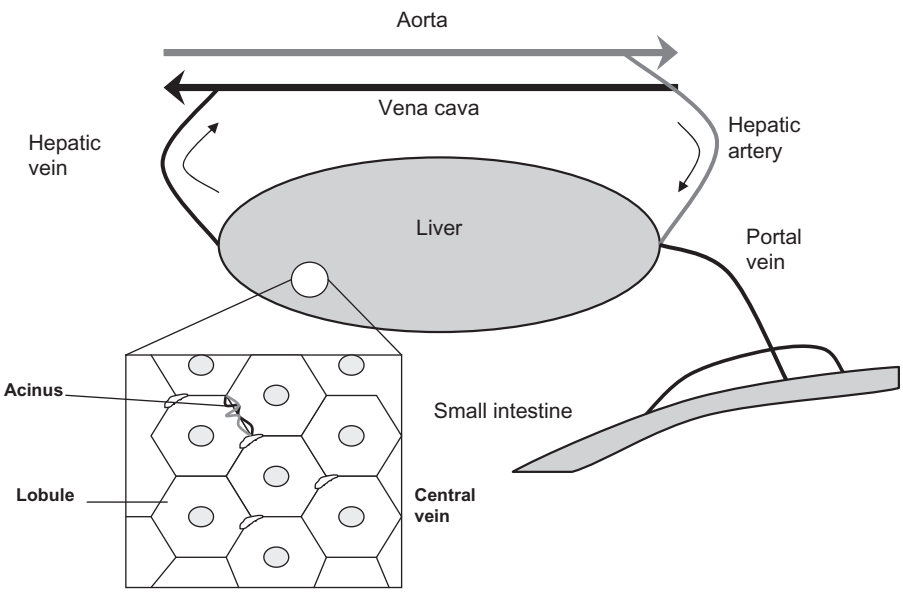


**FIGURE 4** Hepatic vascular system and lobular structure.

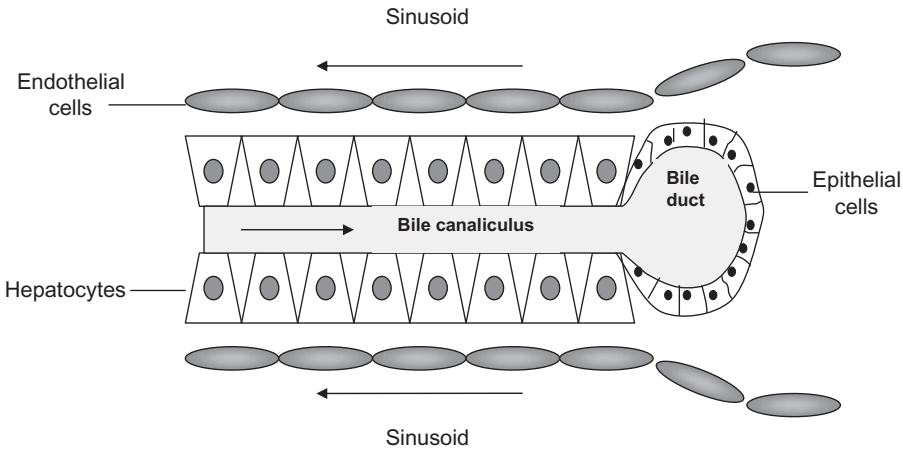


**FIGURE 5** Representation of bile flow from bile canaliculus into the bile duct.

canaliculi, bile flows into bile ducts, which are true ducts lined with epithelial cells. Bile ducts thus begin in very close proximity to the terminal branches of the portal vein and hepatic artery.

Hepatocytes are the chief functional cells of the liver and perform a large number of metabolic, endocrine, and secretory functions. As mentioned previously, roughly 80% of the mass of the liver is hepatocytes. In three dimensions, hepatocytes are arranged in plates that are in membrane-to-membrane contact with one another. The cells are polygonal in shape and their sides are in contact either with sinusoids (sinusoidal face or apical face) or neighboring hepatocytes (basal faces). A portion of the basal faces of hepatocytes are modified, giving rise to bile canaliculi (Fig. 5). Bile originates as secretions from the basal surface of hepatocytes, which collect in canalicular channels. These secretions flow toward the periphery of lobules and into bile ductules and interlobular bile ducts, ultimately collecting in the hepatic duct outside the liver.

### 4.2 Mechanism of Drug Transport Across the Hepatocytes

The portal venous blood contains all of the products absorbed from the gastrointestinal (GI) tract, such as major and minor nutrients originating from ingested food, some endogenous substrates secreted into and reabsorbed from the GI tract, and drugs or other ingested xenobiotics. Drugs delivered to the liver via the portal vein or hepatic artery can be taken up by the hepatocytes and metabolized to more polar compounds. Drugs and their metabolites may then redistribute back into the sinusoids, from where they reach the systemic circulation. Alternatively, drugs or their metabolites may be secreted into the intestinal tract via the biliary system. These secreted compounds and their metabolites can be reabsorbed from the intestine or excreted through feces. Uptake of solutes from sinusoidal blood involves stepwise processes: transport, across the sinusoidal membrane, intrahepatocellular transport, and transport across the canalicular membrane.

***Transport Across the Sinusoidal Membrane*** Once in the sinusoidal blood, drugs may be absorbed from the sinusoids by an active transport process or diffuse passively across the hepatocyte plasma membranes at the apical face. For drugs undergoing an active transport process, traversing the membrane involves several steps. The first is reversible binding of the drug to the membrane transport protein on the sinusoidal side of the cell membrane. Traversing the lipid bilayer itself could occur by a flip-flop process involving the hydrophilic region of the molecule. To aid this process, the carboxyl group of fatty acids becomes protonated by hydrogen bonding either with constituents in the membrane/water interface or constituents within the membrane itself. The uptake process is completed by dissociation of the drug at the cytosolic side of the cell membrane.

Specific plasma membrane transporters exist in the sinusoidal and canalicular membranes of hepatocytes (Fig. 6). These include $Na^+$-dependent and $Na^+$-independent anionic transporters, as well as cationic transporters. These plasma membrane transporters in humans include $Na^+$-taurocholate cotransporting polypeptide (NTCP), human organic anion transporting polypeptides (OATP1 and

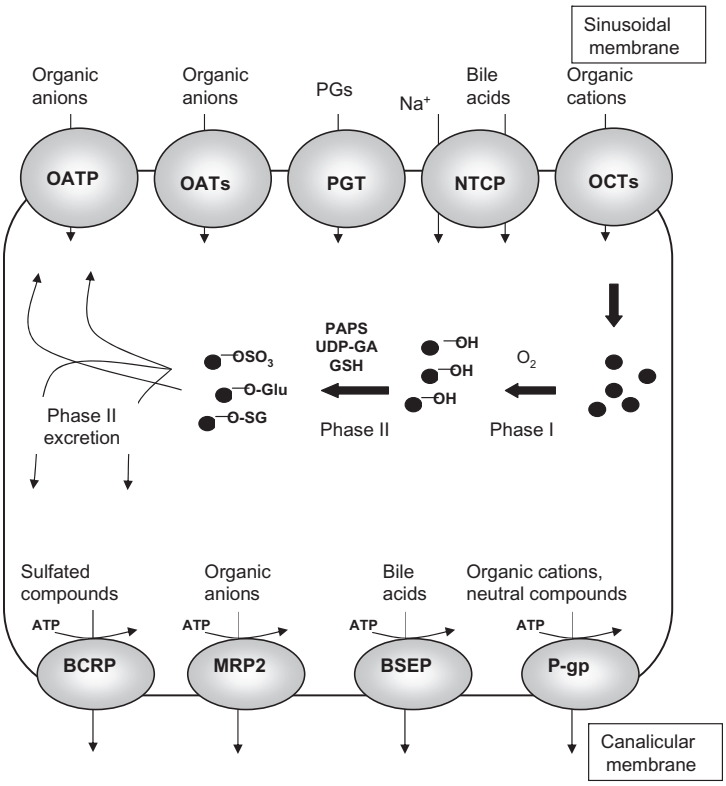


**FIGURE 6** Uptake and efflux transporters on sinusoidal and canalicular membranes. OAT, organic anion transporter; OATP, organic anion transporting polypeptide; PGT, prostaglandin transporter; NTCP, $Na^+$-taurocholate cotransporting polypeptide; OCT, organic cation transporter; BCRP, breast cancer resistance protein; MRP, multidrug resistance associated protein; BSEP, bile salt export pump; P-gp, P-glycoprotein.

OATP2), prostaglandin transporter (PGT), and various organic cation transporters; including the organic cation transporter associated with uptake of hydrophilic aliphatic and certain aromatic cations (OCT1). These membrane transporters are involved in the transport of bile acids, organic anions, and organic cations.

***Intrahepatocellular Transport*** Intrahepatocellular drug transport may occur through several processes. These include cytoplasmic diffusion, protein mediated diffusion, cytoplasmic flow, vesicular transport, and drug transfer from intracellular membranes to intracellular proteins. Cytoplasmic diffusion and protein mediated diffusion of drugs are important in intracellular drug disposition. The transfer of solutes from the sinusoidal membrane to the canalicular membrane involves several different steps. These include lateral diffusion of the unbound ligand through the cytoplasm, lateral diffusion of the unbound ligand through membranes, vesicular transport, and transport by cytosolic binding proteins. It is also possible that transporters facilitate the uptake of some solutes by intrahepatic organelles such as lysosomes and mitochondria. Lipophilic basic drugs accumulate in the relatively acidic hepatic lysosomes as a result of the pH difference between lysosomes and the cytoplasm, but may also be due to binding to lipophilic substances and/or aggregation within lysosomes. The sequestration of cationic drugs in the liver also occurs by both slow and fast binding to proteins and membrane sites, as well as ion trapping into mitochondria and lysosomes. The accumulation of cationic lipophilic drugs in the liver, probably primarily due to trapping in mitochondria and lysosomes as well as binding, results in a slow elution and a low availability of cationic lipophilic drugs for excretion in the bile and enterohepatic transport.

***Transport Across the Canalicular Membrane*** Transport of solutes across the canalicular membrane into the bile occurs via active transport by several membrane proteins. Mechanistically, biliary elimination of anionic compounds, including glutathione *S*-conjugates, is mediated by multidrug resistance associated protein (MRP) transporter MRP2, whereas bile salts are excreted by a bile salt export pump (BSEP). MRP2 can recognize many kinds of organic anions and is responsible for the biliary excretion of many endogenous compounds and drugs. Also expressed in the canalicular membrane is P-glycoprotein, which mediates the biliary excretion of hydrophobic, mostly organic cationic and neutral metabolites. All of the canalicular membrane transport proteins are driven by ATP hydrolysis to transport a number of drug conjugates that undergo enterohepatic recycling, including glucuronide, sulfate, and glutathione conjugates. Figure 6 illustrates various sinusoidal and canalicular membrane transporters involved in the uptake and efflux of solutes.

Although hepatocytes lining the sinusoids are polarized (i.e., transport compounds from sinusoidal surface to canalicular surface), there is a possibility for back-transport into the sinusoids. For instance, certain multidrug resistance associated protein (MRP) transporters (MRP1, MRP3, and MRP6) exist at the sinusoidal surface of the membrane [14]. These active transporters can transport certain drugs and metabolites from hepatocytes back into sinusoids. In some cases there is preferential excretion of the solutes back into the sinusoids rather than into the bile, thus minimizing potential enterohepatic recycling. Other drug anion conjugates are also preferentially excreted into the sinusoids by a saturable transport mechanism.

In general, transporters such as MRP1 are present in the sinusoidal membrane only at very low levels in quiescent cells.

### 4.3 Role of Liver in First-Pass Metabolism

Although every tissue has some ability to metabolize drugs, the liver is the principal organ of drug metabolism. Other tissues that display considerable activity include the gastrointestinal tract, the lungs, the skin, and the kidneys. Drugs that are administered orally are absorbed from the gastrointestinal tract, carried via the hepatic portal vein to the liver. Many drugs are absorbed intact from the small intestine and transported to the liver, where they undergo extensive metabolism by the liver before the systemic organs are exposed to the drug. This removal of a drug by the liver, before the drug has become available in the systemic circulation, is called the *first-pass effect*. First-pass metabolism can occur in the gut and/or the liver, since the entire dose of the drug absorbed after oral administration must pass through these organs prior to reaching the systemic circulation. This process may greatly limit the bioavailability of orally administered drugs, such that alternative routes of administration must be employed to achieve therapeutically effective blood levels.

### 4.4 Biliary Clearance

Xenobiotic compounds excreted into bile are often divided into three classes on the basis of the ratio of their concentration in bile versus that in plasma. Class A substances have a ratio of nearly 1 and include sodium, potassium, glucose, mercury, thallium, cesium, and cobalt. Class B substances have a ratio of bile to plasma greater than 1. Class B substances include bile acids, bilirubin, sulfobromophthalein, lead, arsenic, manganese, and many other xenobiotics. Class C substances have a ratio below 1 (e.g., inulin, albumin, zinc, iron, gold, and chromium). Compounds rapidly excreted into bile are most likely to be found among class B substances.

The mechanism by which the body directs some compounds to the biliary excretion is as yet unclear. However, it is apparent that molecular weight of the compound is a key factor in determining biliary excretion. The molecular weight threshold for humans is estimated to be 500–600. Compounds with molecular weights less than this threshold are excreted primarily in the urine. This knowledge may be useful in drug design when one objective is to achieve a lead compound exhibiting primarily biliary excretion—which obviates the need for dosage adjustment in renal disease.

Drugs that undergo biliary secretion have the potential to be reabsorbed from the GI tract and exhibit enterohepatic recycling. Enterohepatic recycling prolongs the half-life of drugs primarily excreted in bile. Interruption of this cycling through the oral administration of nonabsorbable adsorbants (e.g., activated charcoal) may enhance the clearance of affected compounds. In addition, since the gallbladder empties episodically in humans, drugs that undergo significant enterohepatic recycling may exhibit discontinuous plasma concentrations (as a portion of the excreted dose is periodically reabsorbed, resulting in increases in the plasma concentration of drug).

### 4.5 Determination of Hepatic Clearance

Determination of the hepatic clearance ($CL_H$) of a compound is generally achieved through applying the principle of the additivity of clearance. In particular, the total and renal clearance are determined as described previously. If there is no evidence of nonrenal/nonhepatic elimination of the drug, the $CL_H$ is given as

$$CL_H = CL_T - CL_R$$

This means of determination for hepatic clearance obviously requires numerous assumptions that must be validated experimentally. Hepatic clearance may also be estimated from *in vitro* data, taking into account protein binding and liver blood flow.

### 4.6 Models of Hepatic Clearance

Predicting the impact of pharmacological, physiological, and pathophysiological interventions on hepatic clearance necessitates the elucidation of the determinants of hepatic clearance such that a quantitative model can be developed. Among the various mathematical models proposed to describe hepatic clearance, the venous equilibrium model (also known as the well-stirred model) represents the simplest model that appears to provide reasonably robust predictive capacity [15, 16].

***Venous Equilibrium Model of Hepatic Clearance*** The venous equilibrium model assumes the liver to be represented as a single, homogeneous, well-stirred compartment, such that all metabolic enzymes in the liver are exposed to the same concentration of drug (Fig. 7). This model simplistically assumes that drug bathing the sinusoids is in equilibrium with that in hepatic venous blood (hence the name venous equilibrium). Such a model obviously does not account for various anatomical complexities present in the liver, such as its network of branching tubes each with a zonal distribution of enzymes. It is also physiologically naive to envision drug concentration as unchanging as blood traverses the sinusoidal bed. Nevertheless, despite the fact that experimental evidence suggests that the hepatic sinusoid is not

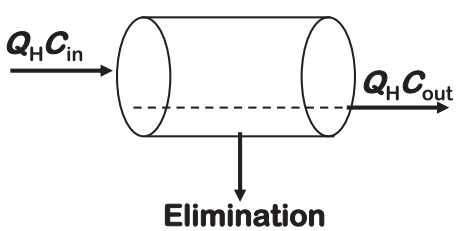


**FIGURE 7** Schematic of the venous equilibrium or well-stirred model of hepatic clearance. This model in essence envisions drug concentration dropping immediately upon entry into the hepatic sinusoids and not declining further as blood traverses across the sinusoidal bed. Hence the drug concentration bathing the sinusoids is viewed as being in equilibrium with the concentration in venous blood. $Q_H$, blood flow; $C_{in}$, concentration of drug in blood entering the liver; $C_{out}$, concentration of drug in venous blood exiting the liver. Dashed line represents the drug concentration in the sinusoids. (Adapted from Ref. 20.)

well stirred [17], the model has been found to provide good estimates of the impact of various changes on hepatic clearance and drug concentration. This model identifies three independent, noninteracting determinants of hepatic clearance: hepatic blood flow ($Q_H$), fraction of unbound drug in blood ($f_{ub}$), and the unbound intrinsic hepatic clearance ($CL_{uint}$) The $CL_{uint}$ represents the ability of the liver to remove drug from blood in the absence of confounding factors (e.g., protein binding and blood flow) and is determined by the ability of the liver to metabolize or secrete (via biliary secretion) the drug. Accordingly, the hepatic clearance ($CL_H$) and hepatic extraction ratio ($E_H$) for the venous equilibrium model can be expressed as [15, 16, 18, 19]

$$CL_H = \frac{Q_H f_{ub} CL_{uint}}{Q_H + f_{ub} CL_{uint}}$$

$$E_H = \frac{f_{ub} CL_{uint}}{Q_H + f_{ub} CL_{uint}}$$

This model provides important insight into the impact of changes in blood flow, protein binding, and drug metabolism/transport on hepatic clearance—particularly at the extreme values of the unbound intrinsic hepatic clearance. For example, when $Q_H >> f_{ub}CL_{uint}$, the value of $CL_H$ will approach $f_{ub}CL_{uint}$ (i.e., $CL_H \approx f_{ub}CL_{uint}$). For such drugs, changes in hepatic clearance will be proportional to changes in protein binding or the unbound intrinsic clearance. Hence, coadministration of an enzyme inducer that increases the metabolism of such a drug will result in an increase in the hepatic clearance of the agent. Moreover, increases in the free fraction of drug will also result in proportional increases in the clearance of the drug. This indicates that the clearance of the drug is "restricted" to the free drug. Such drugs are therefore sometimes referred to as exhibiting restrictive clearance. Drugs in this category are also often classified as possessing a low intrinsic clearance. Agents that exhibit a hepatic extraction ratio < 0.3 can be viewed as possessing a low intrinsic clearance. Importantly, however, changes in hepatic blood flow will not alter the clearance of such drugs, provided the decrease is not sufficient to cause hypoxia and result in hepatocellular damage. Examples of drugs that fall into this category are diazepam, phenytoin, tolbutamide, and warfarin.

At the opposite extreme are those drugs for which $Q_H << f_{ub}CL_{uint}$, sometimes denoted as high intrinsic clearance drugs. Under these conditions, the value of $CL_H$ will approach $Q_H$ (i.e., $CL_H \approx Q_H$). Such drugs are referred to as exhibiting blood flow-dependent clearance (or perfusion rate limited), meaning that changes in blood flow result in proportional changes in clearance. Under these conditions, increases in enzyme or transport activity will not result in further increases in hepatic clearance, as drug is already being cleared as rapidly as it is delivered to the liver. Modest reductions in enzyme activity will also not result in significant alterations in drug clearance. However, if an inhibitor of metabolism or transport causes a magnitude of reduction such that the assumption that $Q_H << f_{ub}CL_{uint}$ is no longer true, coadministration of an inhibitor with a drug exhibiting a high intrinsic clearance may result in significant changes in drug clearance. Drugs with an extraction ratio > 0.7 are generally classified as exhibiting a high intrinsic clearance. Moreover, drug is removed so rapidly under these conditions that it is not restricted to free drug.

This model also provides important insight into factors that influence the systemic availability of drugs after oral administration. If a drug is completely absorbed after oral administration and is not metabolized by the gut, the fraction of drug that reaches the systemic circulation can be described as

$$F = 1 - E_{\mathrm{H}}$$

Since

$$E_{\mathrm{H}} = \frac{f_{\mathrm{ub}} CL_{\mathrm{uint}}}{Q_{\mathrm{H}} + f_{\mathrm{ub}} CL_{\mathrm{uint}}}$$

$F$ can be determined as

$$F = 1 - E_{\mathrm{H}}$$
$$F = 1 - \frac{f_{\mathrm{ub}} CL_{\mathrm{uint}}}{Q_{\mathrm{H}} + f_{\mathrm{ub}} CL_{\mathrm{uint}}}$$
$$F = \frac{Q_{\mathrm{H}}}{Q_{\mathrm{H}} + f_{\mathrm{ub}} CL_{\mathrm{uint}}}$$

The relationship given above indicates that when $Q_{\mathrm{H}} >> f_{\mathrm{ub}} CL_{\mathrm{uint}}$, $F \rightarrow 1$. Such drugs will not exhibit any significant first-pass effect during passage through the liver. Hence, coadministration of an enzyme inducer or inhibitor will not significantly alter the systemic availability of drugs with these characteristics. In contrast, when $Q_{\mathrm{H}} << f_{\mathrm{ub}} CL_{\mathrm{uint}}$, $F \rightarrow 0$ and little of the drug will reach the systemic circulation after oral administration. Under these circumstances, changes in enzyme activity, as produced by inhibitors or inducers, may have profound effects on the systemic availability of drugs. Among the drugs that exhibit a high first-pass effect and low systemic availability are propranolol, verapamil, and nitroglycerin. Because of the high first-pass effect, oral doses of such drugs are substantially (sometimes an order of magnitude) greater than intravenous doses in order achieve the desired pharmacologic effect. In addition, there is generally much more inter- and intrapatient variability in the pharmacokinetics of drugs subject to extensive metabolism. As a consequence, one objective in lead optimization is often to utilize rationale drug design to eliminate first-pass metabolism.

This model of hepatic drug clearance also demonstrates the importance of understanding the role of first-pass metabolism in the interpretation of values of drug clearance obtained from the ratio of dose and area under the drug concentration versus time curve. This is most readily seen for a drug that is eliminated solely by hepatic metabolism and completely absorbed after oral administration. For such compounds, the hepatic clearance after intravenous administration can be determined as

$$CL_{\mathrm{H}} = \frac{D_{\mathrm{IV}}}{AUC_{\mathrm{IV}}}$$

where $D_{IV}$ is the intravenous dose and $AUC_{IV}$ is the area under the drug concentration versus time curve for that dose. Similarly, the hepatic clearance after oral administration is given as

$$CL_H = \frac{F \times D_o}{AUC_o}$$

where $F$ is the systemic availability, $D_o$ is the oral dose, and $AUC_o$ is the area under the drug concentration versus time curve after that dose. Remembering that

$$CL_H = \frac{Q_H f_{ub} CL_{uint}}{Q_H + f_{ub} CL_{uint}}$$

and

$$F = \frac{Q_H}{Q_H + f_{ub} CL_{uint}}$$

one can substitute these equalities for $CL_H$ and $F$, yielding

$$\frac{Q_H f_{ub} CL_{uint}}{Q_H + f_{ub} CL_{uint}} = \frac{Q_H}{Q_H + f_{ub} CL_{uint}} \times \frac{D_o}{AUC_o}$$

This relationship simplifies to

$$f_{ub} CL_{uint} = \frac{D_o}{AUC_o}$$

Importantly, note that the ratio of oral dose to $AUC$ does not provide the value for hepatic clearance, but rather $f_{ub} CL_{uint}$. When calculated in this fashion, this is referred to as the oral clearance ($CL_o$). For a low intrinsic clearance drug, this value will be essentially equal to the hepatic clearance. However, for a high intrinsic clearance drug, the value will be much higher than the hepatic clearance and will exceed the value for hepatic blood flow.

This relationship also provides a means by which the unbound intrinsic hepatic clearance and hepatic blood flow can be determined experimentally. If an intravenous dose of radiolabeled drug and an oral dose of "cold" drug are administered simultaneously, the hepatic and intrinsic clearances can be calculated from the above relationships. Once these values are determined, the blood flow can be calculated from the knowledge of these two parameters. While similar determinations can be made by administering nonradiolabeled drug by both routes on different days, simultaneous administration by both routes obviates the need to account for day to day variability in hepatic blood flow or enzyme activity.

***Parallel Tube Model of Hepatic Clearance*** The parallel tube model represents the liver sinusoids as identical parallel tubes. This model also assumes that metabolic

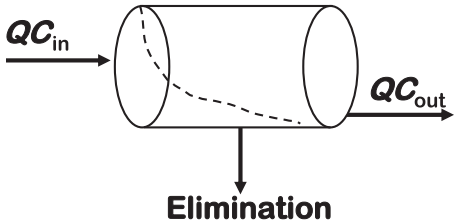


**FIGURE 8** Schematic representation of the parallel tube model of hepatic drug clearance. This model envisions the liver as being comprised of a series of parallel tubes (one of which is shown) with drug concentration declining as blood traverses through the sinusoidal space. See Fig. 7 for abbreviations. (Adapted from Ref. 20.)

enzymes are uniformly distributed along the sinusoid [20]. In this model, drug concentration is envisioned as declining exponentially as drug traverses the sinusoidal tube (Fig. 8). This leads to different equalities for extraction ratio, hepatic clearance, and systemic availability:

$$E = 1 - e^{-f_{ub}CL_{uint}/Q_H}$$
$$CL_H = Q_H(1 - e^{-f_{ub}CL_{uint}/Q_H})$$
$$F = e^{-f_{ub}CL_{uint}/Q_H}$$

The differences between the parallel tube and venous equilibrium models in prediction of the effect of changes in blood flow, enzyme activity, and protein binding are insignificant for drugs that exhibit a low intrinsic clearance (i.e., $E < 0.3$). For intermediate drugs (i.e., $0.3 < E < 0.7$) the differences are also generally insignificant except with oral administration. The difference in model prediction is most clearly seen for drugs with high intrinsic clearance values. Using a variety of experimental models, most commonly the isolated perfused rat liver, numerous investigators have demonstrated substantial differences in the predictive capacity of these two models for such drugs. However, it should be noted that these differences are probably not of clinical significance for the overwhelming number of therapeutic agents. Therefore, the use of the most simplistic model provides important insight for the practical application of clearance concepts.

***Other Models of Hepatic Clearance*** In addition to the two models described earlier, additional models of hepatic clearance have been proposed. The distributed model is a modification of the parallel tube model, where each tube has a distinct blood flow rate and metabolic capacity [19]. The dispersion model adds complexity via consideration of the transit time of the drug in the model. A series compartment model has also been proposed, which envisions the liver to be comprised of a series of well-stirred compartments. The review of Morgan and Smallwood [21] is recommend for readers interested in further consideration of various models of hepatic clearance.

## 5 EFFECT OF DISEASE ON DRUG CLEARANCE

From the foregoing discussion, it should be obvious that disease states that alter the functional capacity of organs of elimination may result in substantial changes in drug clearance. It is also important to recognize that the contribution of a given organ of elimination to the overall elimination of a drug may change dramatically in the face of significant pathology of another organ of elimination. Consider, for example, a drug that is eliminated 85% by the kidneys unchanged and 15% by hepatic metabolism. For such a drug, alteration of hepatic metabolism by administration of an enzyme inhibitor is unlikely to have a clinically significant impact on the disposition of the compound. In contrast, in the presence of renal disease such that hepatic metabolism becomes the predominant route of elimination, interaction with an enzyme inhibitor may be of substantial significance.

The impact of renal disease on the renal clearance of drugs has been the most widely studied disease-induced alteration of drug disposition. This is, in part, due to the ready ability to quantify kidney function via creatinine clearance. Hence, it has been possible to evaluate the impact of various degrees of functional deficit on the renal elimination of drugs. Renal diseases, mainly acute and chronic renal failure, are associated with nephron loss to varying degrees depending on the stage of the renal disease. Bricker's "intact nephron" hypothesis provides an explanation for the kidney's ability to compensate and preserve homeostasis despite a significant loss of nephron function in renal disease [22]. During renal failure, regardless of etiology, injury occurs to the nephrons in a progressive manner. Significant damage to groups of nephrons will eliminate them from contributing to the maintenance of normal renal function. The remaining intact nephrons will compensate by experiencing cellular hypertrophy. This growth process will enable them to accept larger blood volumes for clearance, thus contributing to maintenance of glomerulotubular balance and the excretion of greater solute levels, resulting in compensation. Thus, varieties of adaptations compensate for the decreased GFR and allow a new steady state of external balance to exist. However, in the process of adapting, a second component becomes disordered. The classic examples are the secondary hyperparathyroidism and uremic syndrome of chronic renal failure. Thus, renal diseases not only influence the overall elimination of drugs directly but also result in secondary abnormalities.

Progressive renal disease evolves to a multiorgan syndrome such that the disposition of drugs via nonrenal routes may be altered. For example, patients with renal disease may accumulate endogenous compounds that displace drugs from plasma protein binding sites. This has the potential to increase the hepatic clearance of low intrinsic clearance drugs. Furthermore, the accumulation of the endogenous products normally excreted in the kidney may impair the functional capacity of the liver, such that drug metabolism (and thus hepatic clearance) is impaired.

The kidney makes the major contribution to excretion of unchanged drug and also the excretion of metabolites. A range of physiological and pathophysiological states may influence the efficiency of renal clearance as described earlier. It is evident that renal drug clearance is altered to a clinically significant extent in a number of disease states. Even though the intact nephrons compensate for the decreased clearance by adaptation, they cannot completely reestablish the normal kidney function. For a drug, if the renal clearance contributes only 25–30% to

overall clearance, the renal impairment may not influence the total body clearance significantly; while for the drugs that are predominantly cleared by the kidneys, total clearance is significantly affected by renal impairment. Therefore, drug removal by artificial means (renal replacement therapy) such as hemodialysis becomes essential in subjects with substantial loss of renal function. Similar to renal clearance, the extent to which the drug is affected by dialysis is determined primarily by physicochemical characteristics such as molecular size, protein binding, volume of distribution, hydrophilicity, and plasma clearance of the drug. In addition, technical aspects of the dialysis procedure such as characteristics of dialysis membrane and dialysate flow rates may also determine the extent of drug removal [23, 24]. Drugs that are significantly cleared by the kidney often undergo substantial removal during dialysis and dosing adjustments are often required for such drugs. Only the unbound drug is available for filtration and drugs with high protein binding are poorly cleared by dialysis. The influence of renal diseases on protein binding of drugs is not well understood. It has been reported that low albumin levels are found in critically ill patients, which may increase the unbound fraction of many drugs with possible deleterious effects. These patients also often have increased levels of acid α-glycoprotein, which may increase protein binding of some drugs. Thus, the unbound fraction in healthy volunteers and in patients with renal impairment may differ substantially from the unbound fraction of drugs in critically ill patients receiving dialysis. Similarly, in critically ill patients the actual volume of distribution may differ significantly from that of healthy subjects, and it also exhibits great inter- and intraindividual variation [24, 25].

The ability of the liver to metabolize drugs depends on hepatic blood flow and enzyme activity, both of which can be affected by liver damage. Hepatic diseases can selectively modify the kinetics of drug metabolism in the liver [26]. Previous studies on the effects of hepatic disease have shown that cytochrome P450 (CYP) enzymes are more susceptible to hepatocellular injury than are NADPH-cytochrome P450 reductase or phase II enzymes, such as UDP-glucuronosyltransferases involved in drug conjugation reactions. Patients with severe liver diseases such as cirrhosis or severe hepatitis with liver failure have significant impairment of CYP enzymes. Moreover, in patients with liver damage, expression of some CYP isoforms remains unaffected while other isoforms are significantly reduced. For instance, in liver cirrhosis expression of CYP1A2 and CYP3A4 is decreased, while in cirrhosis with cholestasis expression of CYP2C and CYP2E isoforms is reduced [27, 28]. Therefore, a thorough knowledge of the particular enzyme involved in the metabolism of a drug and the impact of hepatic damage on that enzyme is essential to provide a reasonable basis for dosage adjustment in patients with hepatic impairment. In addition, liver failure can influence the binding of a drug to plasma proteins. These changes can occur alone or in combination; when they coexist their effect on drug kinetics can be synergistic. The kinetics of drugs with a low hepatic extraction are sensitive to hepatic failure rather than to liver blood flow changes. While the kinetics of many drugs are altered by liver disease, quantifying the required dosing changes remains a challenge. At present, there is no satisfactory test that provides a quantitative measure of liver function, with which drug clearance is highly correlated.

Liver diseases may also impair the biliary excretion of drugs. Cholestasis is a hepatic abnormality in which there is stagnation of bile flow, which may arise from

a physical obstruction of the biliary tree (extrahepatic cholestasis) or a decrease in the secretion of the bile by the hepatocytes (intrahepatic cholestasis). The impairment of biliary secretions may result in the accumulation of metabolites, such as glucuronide conjugates, that may be transported into blood. Under these conditions, the overall time that a compound remains in the body increases considerably. There are other clinical conditions in which the metabolism of drugs is altered, including heart failure and other liver diseases.

It is also important to recognize that chronic renal failure can significantly affect the disposition of both low and high hepatic extraction drugs, which are cleared predominantly by the liver. Renal diseases have also been suggested to affect biliary clearance through the accumulation of some P-glycoprotein substrates in the plasma, which may in turn inhibit the secretion of some drugs [9, 29].

## REFERENCES


1. van Acker BA, Koomen GC, Koopman MG. Discrepancy between circadian rhythms of inulin and creatinine clearance. *J Lab Clin Med* 1992;120:400–410.
2. Koopman MG, Koomen GC, Krediet RT. Circadian rhythm of glomerular filtration rate in normal individuals. *Clin Sci* 1989;77:105–111.
3. Bonate PL, Reith K, Weir S. Drug interactions at the renal level: implications for drug development. *Clin Pharmacokinet* 1998;34:375–404.
4. Ullrich KJ, Rumrich G. Contraluminal transport systems in the proximal renal tubules involved in secretion of organic ions. *Am J Physiol* 1989;254:F453–F462.
5. Burkhardt G, Ullrich KJ. Organic anion transport across the contraluminal membrane KG dependence on sodium. *Kidney Int* 1989;36:370–377.
6. Lee W, Kim RB. Transporters and renal drug elimination. *Annu Rev Pharmacol Toxicol* 2004;44:137–166.
7. Rennick BR. Renal tubule transport of organic cations. *Am J Physiol* 1981;240: F83–F89.
8. Dorian C, Gattene VH II, Klaassen CD. Renal cadmium deposition and injury as a result of accumulation of cadmium-metallothionein (CdMT) by proximal concoluted tubules—a light microscope autoradiographic study with $^{109}$CdMT. *Toxicol Appl Pharmacol* 1992;114:173–181.
9. Dreisbach AW, Lartora JJL. The effect of chronic renal failure on hepatic drug metabolism and drug disposition. *Semin Dialysis* 2003;16:45–50.
10. Rule AD, Larson TS, Bergstralh EJ, Slezak JM, Jacobsen SJ, Cosio FG. Using serum creatinine to estimate glomerular filtration rate: accuracy in good health and in chronic kidney disease. *Ann Intern Med* 2004;141:929–937.
11. Rule AD, Gussak HM, Pond GR, Bergstralh EJ, Stegall MD, Cosio FG, Larson, TS. Measured and estimated GFR in healthy potential kidney donors. *Am J Kidney Dis* 2004;43:112–119.
12. Branten AJ, Vervoort G, Wetzels JF. Serum creatinine is a poor marker of GFR in nephrotic syndrome. *Nephrol Dial Transplant* 2005;20:707–711.
13. Hall SD, Rowland M. Influence of fraction unbound upon the renal clearance of furosemide in the isolated perfused rat kidney. *J Pharmacol Exp Ther* 1985;232:263–268.
14. Chandra P, Brouwer KLR. The complexities of hepatic drug transport: current knowledge and emerging concepts. *Pharm Res* 2004;21:719–735.

15. Perrier D, Gibaldi M. Clearance and biological half-life as indices of intrinsic hepatic metabolism. *J Pharmacol Exp Ther* 1974;191:17–24.
16. Wilkinson GR, Shand DG. A physiological approach to hepatic drug clearance. *Clin Pharmacol Ther* 1975;18:377–390.
17. Weisiger RA, Mendel CM, Cavalieri RR. The hepatic sinusoid is not well-stirred: estimation of the degree of axial mixing by analysis of lobular concentration gradients formed during uptake of thyroxine by the perfused liver. *J Pharm Sci* 1986;75:263–268.
18. Ridgeway D, Tuszynski JA, Tam YK. Reassessing models of hepatic extraction. *J Biol Phys* 2003;29:1–21.
19. Ahmad AB, Bennette PN, Routland M. Models of hepatic drug clearance: discrimination between the well-stirred and parallel tube models. *J Pharm Pharmac* 1983;35:219–224.
20. Pang SK, Rowland M. Hepatic clearance of drugs I. Theoretical considerations of a "well-strirred" model and a "parallel tube" model. Influence of hepatic blood flow, plasma and blood cell binding and the hepatocellular enzymatic activity on hepatic drug clearance. *J Pharmacokinet Biopharm* 1977;5:625–653.
21. Morgan DJ, Smallwood RA. Clinical significance of pharmacokinetic models of hepatic elimination. *Clin Pharmacokinet* 1990;18:61–76.
22. Bricker NS. On the meaning of the intact nephron hypothesis. *Am J Med* 1969;46:1–11.
23. Hudson JQ, Comstock TJ, Feldman GM. Evaluation of an *in vitro* dialysis system to predict drug removal. *Nephrol Dial Transplant* 2004;19:400–405.
24. Driscoll DF, McMahon M, Blackburn GL, Bistrain BR. Phenytoin toxicity in a critically ill, hypoalbuminemic patient with normal serum drug concentrations. *Crit Care Med* 1988;16:1248–1249.
25. Brugge JF. Pharmacokinetics and drug dosing adjustments during continuous venovenous hemofiltration or hemodiafiltration in critically ill patients. *Acta Anesthesiol Scand* 2001;45:929–934.
26. Rodighiero V. Effects of liver disease on pharmacokinetics. An update. *Clin Pharmacokinet* 1999;37:399–431.
27. Blouin RA, Farrell GC, Ionnides C, Renton KW, Watlington CO. Impact of diseases on detoxication. *J Biochem Mol Toxicol* 1999;13:215–218.
28. George J, Murray M, Byth K, Farrell GC. Differential alterations of cytochrome P450 proteins in livers from patients with severe chronic liver disease. *Hepatology* 1995;21:120–128.
29. Nolin TD, Frye RF, Matzke GR. Hepatic drug metabolism and transport in patients with kidney disease. *Am J Kidney Dis* 2003;42:906–925.

# *IN VITRO* METABOLISM IN PRECLINICAL DRUG DEVELOPMENT

Olavi Pelkonen,[1] Ari Tolonen,[1,2] Miia Turpeinen,[1] and Jouko Uusitalo[2]

[1]*University of Oulu, Oulu, Finland*
[2]*Novamass Analytical Ltd., Oulu, Finland*

**Contents**

1 Introduction
2 *In Vitro* Biological Systems to Catalyze Metabolism
  2.1 Metabolic Competence of the *In Vitro* System
  2.2 From Recombinants to Perfusions
  2.3 Current Types and Components of *In Vitro* Systems
  2.4 Measuring Induction Potential *In Vitro*
  2.5 Species Considerations
3 Assignment of Metabolizing Enzymes
  3.1 Affinity (Inhibition) Studies
  3.2 Enzyme-Selective Substrates, Inhibitors, and Antibodies
  3.3 Recombinant Enzymes
  3.4 Enzyme Kinetic Characterization of Principal Metabolic Reactions
4 Identification and Quantification of Metabolites
  4.1 First Incubations—Tentative Identification of Metabolites
  4.2 Absolute Structures—Reference Metabolites and NMR Studies
  4.3 Quantitative Studies
  4.4 Requirements for Analytical Instrumentation

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

5 Active/Reactive/Toxic Metabolites
6 Testing Configurations
6.1 Sample Preparation
6.2 Single Enzyme/Activity Systems
6.3 Medium to High Throughput Systems
6.4 Analytical Considerations
7 Priorities of *In Vitro* Metabolism Studies and *In Vitro–In Vivo* Extrapolation
7.1 Priorities of *In Vitro* Metabolism Studies
7.2 *In Vitro–In Vivo* Extrapolation
8 Validation of *In Vitro* Systems
9 Conclusion
References

## 1 INTRODUCTION

In drug therapy, some important questions concerning the efficient and safe use of a drug are the following: (1) What is the size of the dose? (2) How often should it be given? (3) Is there interference from other simultaneously used drugs or other chemicals? (4) Are there important idiosyncratic reactions due to a drug? (5) Are there important patient characteristics to be considered? [1]. The first two questions relate to basic pharmacokinetic characteristics of a drug, principally to its bioavailability and clearance. The third question relates to potential interactions. The fourth concerns the possibility of metabolism-related toxicities, and the fifth relates to endogenous, host-related and genetic features of an individual patient to whom a drug is administered. At the basic level, we need to know about the drug substance (1) its clearance and factors contributing to it, (2) major and minor, including potentially toxic, metabolites (i.e., metabolite profile), and (3) drug-metabolizing enzymes catalyzing (at least) rate-limiting routes, and their inhibition and induction and physiological and pathological behavior.

In this chapter, metabolism in preclinical drug development is the principal topic. However, before going into details, some more general concerns during drug discovery and development should be expressed. Drug development is the optimization of not a single feature but multiple features, even regarding metabolism and kinetics [2, 3]. Consequently, optimization is always a trade-off between several independent or interdependent features. Thus, there are no preset answers to the optimization problem; instead, it is a continuous cross-talk between various optimization tasks. It is therefore difficult to say that one *in vitro* test is more important than another, because the relative importance depends on the individual substance under study and its unique characteristics. However, this is not to say that we do not need a general strategy to study drug metabolism and pharmacokinetics by *in vitro* methods. We need a general strategy, but it has to be modified depending on the substance, its stage of development, and many other considerations.

In Table 1, an attempt has been made to put metabolism into a wider perspective in pharmacokinetic (PK) processes. Optimization of metabolic features in the context of pharmacokinetics is divided into characteristics, which should be pre-

**TABLE 1 Metabolic Processes, Characteristics, and Factors that Should Be Predicted on the Basis of *In Vitro* Screening Assays or Assumed During the *In Vitro–In Vivo* Extrapolationa**

| PK Process | Characteristic | Factors | Assays |
|---|---|---|---|
| Systemic exposure | | | |
| Hepatic clearance | Hepatic blood flow | Protein binding permeability | Animals *in vivo* |
| | Hepatic metabolic clearance | CYP enzymes, phase II enzymes | *In vitro* and cell based (see text) |
| | Biliary clearance | Transporters | Liver perfusion |
| Bioavailability | | | |
| First-pass clearance | Intestinal clearance | Efflux transporters and intestinal metabolism | Caco-2 |
| | Hepatic clearance | See above | |
| Interactions | | | |
| Metabolism | Inhibition | CYP enzymes | *In vitro* microsomes, recombinant CYPs or cell based |
| | | Phase II enzymes | as above |
| | Induction | | Hepatocytes, receptor-binding |
| Metabolic activation | Reactive metabolites, oxygen radicals, etc. | Producing enzymes | *In vitro* systems |

[a]For some illustrative examples concerning various PK aspects, see chapters in a book by Pelkonen et al. [6].

dicted on the basis of *in vitro* information, factors (i.e., biological phenomena or constituents mediating them), and assays or screens (i.e., systems available to measure a factor or a characteristic). Furthermore, an additional entry depicts metabolic activation, which is thought to be of importance for "idiosyncratic" reactions (see Section 5).

One more introductory note: the literature on subjects in this chapter is very large and making proper acknowledgments to all important contributions is impossible. Selection of references thus reflects our somewhat personal view, but hopefully contains the most useful ones, from which the reader can go further.

## 2 *IN VITRO* BIOLOGICAL SYSTEMS TO CATALYZE METABOLISM

The questions posed at the beginning of this chapter will be definitively and unequivocally answered during the clinical phase, including postmarketing pharmacovigilance, of drug development. However, this is too late from the early drug development point of view and thus, to avoid failures and withdrawals later on, the basic information for answering these questions should preferably be produced in the nonclinical phase of development. In the current paradigm of nonclinical drug development,

*in vitro* screening systems are used for producing the appropriate answers through correlation and extrapolation, which actually are absolute requirements for the judicious use of *in vitro* studies to support drug development. Because of interspecies differences and extrapolation uncertainties, human-derived or humanized *in vitro* systems are preferably used. Important goals for using various human liver derived *in vitro* systems are the following: (1) to elucidate and determine principal metabolic routes of a new chemical entity (NCE) and to tentatively identify principal metabolites; (2) to identify phase I enzymes, especially cytochrome P450 (CYP), catalyzing the principal oxidation routes (primary metabolites) and phase II enzymes catalyzing the conjugated metabolites, and to gain some quantitative data on their significance for the overall metabolic fate of an NCE; and (3) to provide useful background information for characterizing potential interactions and physiological, genetic, and pathological factors affecting the kinetics and variability of an NCE in the *in vivo* situation.

### 2.1 Metabolic Competence of the *In Vitro* System

When screening potential drug candidates, one is faced with a question. How large or restricted should the set of enzymes be to cover a reasonable number of potential drug-metabolizing enzymes; that is, what is the optimal range of enzymes to be screened? In the end, each drug is an individual, with its characteristic physicochemical properties, metabolic reactions, metabolizing enzymes, kinetics, and so on. Regarding metabolism, thus far the main focus has been on CYP enzymes. This is natural, because the CYP enzyme superfamily is responsible for approximately 70–80% of the rate-limiting phase I metabolism of drugs [4]. However, there are a large number of other important enzymes participating in the metabolism of drugs. For example, should we routinely screen for various UGT or SULT enzymes and at which stage of the development? Table 2 gives an overview of drug-metabolizing enzymes, which should be considered when studying the metabolic competence of a given *in vitro* screening system [5, 6]. Although it is impossible to give unequivocal recommendations for how comprehensive metabolic competence should be employed in the *in vitro* test system, it is always useful to consider this matter and explicitly state objectives of the specific test and reasons for selecting the specific biological preparation, with a specified set of enzymes.

### 2.2 From Recombinants to Perfusions

Enzyme sources (i.e., biological preparations), are another concern. Most important possibilities are briefly described in Table 3, which presents an overview of current biological preparations for use as enzyme sources and some of their advantages and disadvantages. Liver microsomes have been the preferred source of enzymes for *in vitro* metabolism screening. Recombinant expressed enzymes have become more and more usable, because of easy availability and lack of variability. Isolated and/or cultured cells have a larger complement of enzymes in more natural cellular surroundings. These aspects have been very thoroughly covered recently [13, 14]. There are strong arguments in favor of using human primary hepatocytes [15], but obviously their availability for early and extensive screening is rather difficult. There are only a few studies in which several enzyme sources have been

**TABLE 2 Drug–Metabolizing Enzymes (Preferably Human) that Should/Could/Might Be Included in the *In Vitro Assays***

| Enzyme Classes/Enzymes | (Types of) Drugs Metabolized (Examples) | Sources Available | Need of Incorporation into *In Vitro* Systems |
|---|---|---|---|
| Cytochrome P450 enzymes (about 10–13 drug-metabolizing enzyme forms) | Practically 90% of all drug substances | Hepatocytes most versatile and comprehensive; other tissues and cell types selectively | Incorporation of P450-competent enzyme source is necessary for most purposes |
| An example: CYP1A1 | Few pharmaceutics, principally polycyclic aromatic hydrocarbons and other carcinogens | Recombinant enzyme; mainly extrahepatic expression (e.g., placenta from smokers) | Induction usually required for expression; incorporation not necessary routinely |
| An example: CYP3A4 | A majority, >50%, of all clinically used drugs | Hepatocytes, recombinant | Incorporation necessary for comprehensive screening |
| Flavin-monooxygenases (FMOs) (five forms) | Compounds with secondary and tertiary amines or sulfydryl groups (chlorpromazine, desipramine, methimazole) | Hepatocytes most versatile and comprehensive | Incorporation comes automatically with hepatocytes; specific incorporation only in special circumstances |
| Prostaglandin H synthase | PAH-diols, aflatoxin B1, aromatic amines | Liver, kidney, bladder (microsomes) | Incorporation advisable for toxicity screening? |
| Alcohol/aldehyde dehydrogenases and oxidases | Various compounds with alcohol and aldehyde functions (ethanol) | Many tissues; cytosol | Some incorporation comes automatically with hepatocytes or hepatic homogenate; specific incorporation only in special circumstances |
| Monoamine oxidase | Selegiline, moclobemide | Many tissues; mitochondria | Incorporation advisable in specific situations |

**TABLE 2** ***Continued***

| Enzyme Classes/Enzymes | (Types of) Drugs Metabolized (Examples) | Sources Available | Need of Incorporation into *In Vitro* Systems |
|---|---|---|---|
| Esterases/hydrolases/peptidases | Compounds with cleavable ester/amide bond (procaine, succinylcholine, lidocaine) | Many tissues including blood and blood cells | At least some activity present when liver preparation is incorporated into the test system |
| Reductases | Many substances with azo, nitro, and carbonyl functions (chloramphenicol, naloxone); importance not well characterized | Present in hepatocytes (cytosol, some in microsomes) | Activity needs special circumstances; need of incorporation not adequately defined |
| UDP-glucuronosyltransferases (UGTs); glucuronide conjugation | Most drugs with suitable O-, S-, and N-functional groups (morphine, diazepam, paracetamol) | Hepatocytes most versatile and comprehensive (microsomes); subcellular systems need UDPGA | Incorporation of UGT-competent enzyme source is advisable for most purposes |
| Sulfotransferases (SULTs); sulfate conjugation | Phenols, alcohols, aromatic amines (paracetamol, methyldopa) | Hepatocytes most versatile and comprehensive (cytosol); subcellular systems need PAPS | Incorporation not adequately defined |
| GSH transferases (GSTs); glutathione conjugation | Epoxides, arene oxides, nitro groups, hydroxylamines (ethacrynic acid) | Hepatocytes most versatile and comprehensive (cytosol, microsomes) | Incorporation not adequately defined |
| Acyl-CoA glycinetransferase; amino acid conjugation | Acyl-CoA derivatives of carboxylic acids (salicylic acid) | Hepatocytes (mitochondria) | Incorporation advisable only in special cases |
| *N*-acetyltransferases (NATs); acylation | Amines (sulfonamides, isoniazid, clonazepam, dapsone) | Hepatocytes (cytosol) | Incorporation advisable only in special cases |
| Methyl transferases; methylation | Catecholamines, phenols, amines (L-dopa, thiouracil) | Various tissues (cytosol) | Incorporation advisable only in special cases |

*Sources*: Adapted from Refs. 5 and 41.

**TABLE 3 Comparison of *In Vitro* Enzyme Sources Used in Preclinical Research**

| Enzyme Sources | Availability | Advantages | Disadvantages |
|---|---|---|---|
| Liver homogenates [7] | Relatively good; commercially available | Contains basically all hepatic enzymes | Liver architecture lost; cofactor addition necessary |
| Microsomes [7] | Relatively good; transplantations or commercial sources | Contains most important rate-limiting enzymes; relatively inexpensive; easy storage | Contains only phase I enzymes and UGTs; requires strictly specific substrates or antibodies for individual DMEs; cofactor addition necessary |
| Recombinant CYP enzymes [8] | Commercially available | Can be utilized with HTS substrates; role of individual CYPs in the metabolism can be easily studied | The effect of only one enzyme at a time can be studied |
| Primary hepatocytes [9, 10, 15] | Difficult to obtain; relatively healthy tissue needed; commercially available | Contains the whole complement of DMEs cellularly integrated; induction effect of an NCE can be studied; cryopreservation possible | Requires specific techniques and well established procedures; levels of many DMEs decrease rapidly during cultivation |
| Liver slices [11] | Difficult to obtain; fresh tissue needed. | Contains the whole complement of DMEs and cell–cell connections; induction effect of an NCE can be studied; cryopreservation possible | Requires specific techniques and well established procedures; limited viability |
| Immortalized Cell lines [12] | Available at request; only a few adequately characterized cell lines exist | Nonlimited source of enzymes | Expression of most DMEs is poor; genotype/phenotype instability |

*Sources*: Adapted from Refs. 5 and 41.

rigorously compared [16–18], but it seems that the currently used enzyme systems—recombinant enzymes, human liver microsomes and/or homogenates, liver slices, and hepatocytes—all give fairly reliable results, if their inherent restrictions are taken into consideration.

### 2.3 Current Types and Components of *In Vitro* Systems

During the development of an NCE, according to the current paradigm, several types of investigations are usually carried out, the earlier the better. The enzyme sources in these studies are usually human-derived systems. A summary of the major *in vitro* methods is provided in Table 4. The most important objectives of these studies are the elucidation of metabolic stability of an NCE, identification of metabolites and metabolic routes, and identification of CYP forms metabolizing an NCE. The following paragraphs summarize briefly these approaches. More details are given in subsequent sections.

1. The metabolic stability of an NCE determines its future as a drug candidate. By determining the time and concentration dependence of disappearance and/or metabolite formation *in vitro* in an appropriate system, its hepatic clearance *in vivo* can be predicted [23, 24].
2. Metabolite identification, at least at the tentative level, can be developed from incubations with human liver cells and other preparations, for example, homogenates or microsomes (see Section 4).

**TABLE 4 *In Vitro* Studies for the Characterization of Metabolism and Metabolic Interactions of Potential Drugs**[a]

| *In Vitro* Test | Preparations | Parameters | Extrapolations |
|---|---|---|---|
| Metabolic stability | Microsomes, homogenates, cells, slices | Disappearance of the parent molecule or appearance of (main) metabolites | Intrinsic clearance, interindividual variability |
| Metabolite identification | Same as above | Tentative identification of metabolites, for example, by LC-TOF-MS and LC-MS-MS | Metabolic routes, qualitative (semiquantitative, if possible) metabolic chart |
| Identification of metabolizing enzymes | Microsomes with inhibitors or inhibitory antibodies; recombinant individual enzymes | Assignment and relative ability of enzymes to metabolize a compound | Prediction of effects of various genetic, environmental, and pathological factors; interindividual variability |
| Enzyme inhibition | Microsomes, recombinant enzymes | Inhibition of model activities by a substance | Potential drug–drug interactions |
| Enzyme induction | Cells, slices, permanent cell lines (if available), constructs | Induction of model activities (or mRNA); receptor binding (e.g., PXR or CAR) | Induction potential of a substance |

[a]For some salient background data and examples, see Refs. 15 and 19–22.

3. After characterizing the metabolic stability and metabolic routes of an NCE, the *in vivo* prediction requires clarification of the enzymes that participate in the *in vitro* biotransformation of the NCE. After determining the initial velocity conditions and enzyme kinetic parameters for identified rate-limiting pathways, some of the main tools and approaches used in enzyme assignment are CYP selective chemical inhibitors and antibodies, cDNA expressed CYPs, correlation analysis, and measures of affinities of an NCE for CYPs (by inhibition).

Numerous compounds have been characterized for their inhibitory potency against different CYPs. Many of them are selective for the desired enzyme only at relatively low concentrations. Today, there are several commercial sources for CYP-specific inhibitory antibodies. Inhibitory antibodies raised specifically against a certain CYP form are a good tool in distinguishing between CYP forms.

cDNA expressed enzymes are convenient tools when a specific activity or a selective chemical inhibitor cannot be used in metabolic studies. Recombinant enzymes are used to ascertain the role of a certain CYP in the metabolism of an NCE. Still, the biotransformation of an NCE by a single CYP does not necessarily mean its participation in the reaction *in vivo*.

For correlation analysis, a well characterized bank of human liver samples is needed. In correlation analysis, the measured CYP-specific activities are correlated against the rate of the metabolic pathway of an NCE in every individual liver sample. Correlation analysis gives information about the possible extent of the contribution of certain CYPs to the reaction under study.

4. The effect of an NCE on characteristic CYP-selective activities is studied by coincubating series of dilutions of an NCE with a specific substrate. By comparing the effects of an NCE on the CYP-specific activities to the respective effects of diagnostic inhibitors, a tentative prediction of the *in vivo* situation can be made.

### 2.4 Measuring Induction Potential *In Vitro*

Induction of xenobiotic-metabolizing enzymes is an adaptive cellular response that usually leads to enhanced metabolism and termination of the pharmacological action of drugs. A majority of the drug-metabolizing CYP enzymes are inducible, including the most important CYP3A4. Induction of human CYP enzymes is difficult to study because there are no human liver cell lines that express the full complement of CYP enzymes or reproduce the induction observed *in vivo* [25]. Thus, there is increasing interest in the development of mechanism-based test systems for CYP induction, and proof of principle has been demonstrated for the nuclear receptors CAR, PXR, and PPARα. The main *in vitro* methods used are direct and indirect binding assays as well as cell-based reporter gene protocols [26, 27].

Because these systems are based ultimately on the inducer/receptor interaction, they cannot detect inducers that require *in vivo* transformation to active species or inducers that act via an alternative mechanism. For instance, dexamethasone may increase the expression of the PXR and CAR receptors and thus induce CYP enzymes or synergize with other inducers. Even though the mechanism-based induc-

tion screens may not detect every CYP inducer, they are still expected to be valuable as preliminary screens.

### 2.5 Species Considerations

Species differences in drug-metabolizing enzymes are large and more often than not unpredictable. Ortologous enzymes can be rather closely related in terms of sequence homology and still display large quantitative and qualitative differences toward potential substrates and inhibitors. A much cited example is concerned with mouse hepatic Cyp2a4 and Cyp2a5, which differ in only 11 amino acids and still there is an almost complete reversal of substrate specificity concerning testosterone and coumarin [28]. On the other hand, metabolic reaction of a given compound may be catalyzed by enzymes belonging to different subfamilies in different species (see examples in Ref. 29). Currently, it is practically impossible to predict species differences in metabolism. Thus, it is advisable to perform some comparative studies in the preclinical phase. For example, metabolic stability studies give at least a tentative view about differences in clearance rates between different species. Metabolite identification in hepatic preparations from different species may point to important considerations when trying to extrapolate results from *in vivo* animal studies to human risk assessment. The earlier these types of comparative studies are being performed, the better their results can be taken into consideration, for example, in the selection of appropriate species for certain types of toxicity studies.

## 3 ASSIGNMENT OF METABOLIZING ENZYMES

In terms of potential clinical consequences, the assignment of metabolizing enzymes is perhaps the most important single task in *in vitro* metabolic studies. For the identification of the CYP (or basically any other) enzyme responsible for the metabolite formation, a number of different and complementary approaches are used. The indirect method utilizes specific CYP enzyme inhibitors, which are each added in turn into the incubation with the study compound in the presence of liver microsomal preparation, and the formation of the metabolite is monitored quantitatively. The decreasing metabolite formation indicates that the CYP responsible for the metabolite formation is being inhibited. Different inhibitor concentrations can be used to calculate the $IC_{50}$ value of the inhibitor toward the metabolite formation, but often only single inhibitor concentration (high enough for good inhibition of certain CYP but low enough to keep the inhibition CYP-selective) is used. The use of CYP-specific antibodies is carried out similarly to the inhibitors. Another often used and complementary method is the use of recombinant CYP enzymes, where the involvement of a certain CYP in the metabolite formation can be directly detected as a metabolite peak in LC-MS. In addition, CYP activity phenotyped liver preparations are used to reflect the formation of the metabolites. The increasing metabolite formation paralleling the increasing activity of a certain CYP enzyme in the different liver preparations suggests also a probable involvement of that selective CYP isoform in the biotransformation.

### 3.1 Affinity (Inhibition) Studies

Inhibition of CYP enzymes is the most common cause of drug–drug interactions and has led to the removal of several drugs from the market during the past few years [30, 31]. Thus, studying the inhibitory effect of NCEs is extremely important; an inhibitory profile would also give an idea about the affinity spectrum of the compound.

The most common method is to compare metabolite formation of the known CYP-selective biotransformation in a liver microsomal incubation with and without the study compound. The probe substrate in a biotransformation reaction is chosen so that its monitored metabolite is known to be formed specifically via one certain CYP enzyme. These studies are often carried out as a cocktail-type approach using a number of different CYP-specific probe substrates (3–10) simultaneously in the one incubation, and the formed metabolites are usually analyzed with a single LC/MS-MS method developed for the simultaneous analysis of all the different metabolites (see Ref. 32). The cocktail approach is especially good for the screening of possible CYP interactions of an NCE, as the one incubation and one LC/MS run give information of all studied CYP enzymes in a time-efficient manner. If high inhibition potential toward certain CYPs in a cocktail-type screening for an NCE is detected, then the more accurate $IC_{50}$ values can be evaluated using single-probe metabolite reactions and also using metabolite spiked standard samples for more accurate quantitation in LC/MS.

### 3.2 Enzyme-Selective Substrates, Inhibitors, and Antibodies

Selection of model drugs and reactions is an important consideration in many *in vitro* testing systems. For example, recent surveys [33, 34] give lists of CYP-selective substrates and inhibitors and some potential alternatives for situations in which preferable probes cannot be used. A validation study of several CYP substrates has also been published [35]. Without going into extensive details here, one important consideration is clinical relevance: probe substances should preferably be clinically used drugs for which relatively good and comprehensive *in vivo* data is available [36]. This would make *in vitro–in vivo* extrapolations more reliable, because these probes could then be used as comparators for molecules under study. Table 5 compiles currently used substrates and biotransformations as well as inhibitors, which display at least some selectivity toward specific CYP enzymes.

In the metabolizing enzyme identification studies employing diagnostic inhibitors, the approach is essentially the same as in the CYP interaction studies, but the detected analytes are the metabolites of the studied compound, and the concentration of the diagnostic inhibitor in the incubation is varied. The use of diagnostic inhibitors has become rather routine in the early screening of promising candidate drugs.

### 3.3 Recombinant Enzymes

Isolated heterogeneous human CYP enzymes, expressed as single enzymes at a time from cDNA in bacterial, yeast, and mammalian cells, have been commercially avail-

**TABLE 5 Summary of Human Hepatic Drug-Metabolizing CYP Enzymes and Their Selected Probe Substrates and Inhibitors Used in *In Vitro* and *In Vivo* Studies**

| CYP | Percentage (%) in Liver[a] | Substrate | Inhibitor | Other Characteristics |
|---|---|---|---|---|
| 1A2 | ~10 | Ethoxyresorufin<br>Melatonin<br>Caffeine<br>Phenacetin | Furafylline<br>Fluvoxamine | Inducible<br>Polymorphic |
| 2A6 | ~8 | Coumarin<br>Nicotine | Tranylcypromine | Inducible<br>Polymorphic |
| 2B6 | ~2 | Bupropion<br>Efavirenz<br>Cyclophosphamide | Thio-Tepa<br>Ticlopidine | Inducible<br>Polymorphic |
| 2C8 | ~5 | Paclitaxel<br>Amodiaquine<br>Rosiglitazone | Montelukast<br>Quercetin | Polymorphic |
| 2C9 | ~20 | *S*-warfarin<br>Diclophenac<br>Tolbutamide<br>Losartan | Sulfaphenazole | Polymorphic<br>Inducible |
| 2C19 | ~2 | Omeprazole<br>*S*-mephenytoin<br>Proguanil | Fluconazole | Polymorphic<br>Inducible |
| 2D6 | ~2 | Dextromethorphan<br>Debrisoquine<br>Bufuralol<br>Propranololol | Quinidine<br>Paroxetine | Polymorphic |
| 2E1 | ~15 | Chlorzoxazone<br>Ethanol | Pyridine<br>Disulfiram | Inducible |
| 3A4 | ~40 | Midazolam<br>Testosterone<br>Simvastatin<br>Nifedipine<br>Erythromycin | Ketoconazole<br>Itraconazole | Inducible |

[a]Relative and absolute amounts of hepatic P450 proteins vary highly among people. Rounded values are based on a meta-analysis by Rowland Yeo et al. [45].

*Sources*: Data adapted from Refs. 37–44.

able for several years. Recombinant CYPs have been adopted as frontline tools in early drug development. These systems can be utilized to ascertain whether an NCE is a substrate for a particular CYP form and what metabolite is generated by that specific enzyme. Moreover, recombinant enzymes can be used as small-scale bioreactors to generate usable amounts of metabolic product [46–48]. They have been used also for clearance predictions and drug–drug interaction studies [49]. It should be kept in mind, however, that recombinant expressed enzymes are not in their natural microsomal environment and consequently there is a concern about the extent of applicability of the findings.

### 3.4 Enzyme Kinetic Characterization of Principal Metabolic Reactions

It is of importance to determine enzyme kinetic characteristics of at least the principal and rate-limiting reactions of a substance for scaling up and predicting *in vivo* kinetics of an NCE. If there is only one rate-limiting reaction catalyzed by a single enzyme, this investigation is easier, but if there are several reactions and multiple enzymes, the task is obviously more difficult and the prediction may become more imprecise. The importance of enzyme kinetic characterization becomes apparent in the extrapolation process: the first step in the scheme is determining "enzyme efficiency or enzyme intrinsic clearance," essentially reflecting $V_{max}/K_m$, which is then scaled to liver unit weight.

## 4 IDENTIFICATION AND QUANTIFICATION OF METABOLITES

Analytical tools are at the heart of metabolic studies during early drug development. Metabolism studies start with analyses of the parent compound in simple metabolic stability studies, in which the substrate loss is being determined, and continue to more complex structural analyses on the identification of major and minor metabolites produced by the catalysis of oxidative and conjugative enzymes, and ending with development of routine analytical methods for the identification of the respective enzymes and also addressing other drug interaction-related metabolism studies.

Although the measurement of the disappearance of an NCE in human liver preparations (i.e., "metabolic stability") sounds rather simple, it is actually a rather complicated undertaking, including the development of an assay for the parent compound, and contains a lot of caveats, starting with chemical stability and binding problems. However, if this assay is expanded with the identification of metabolites produced, it gives very useful information for drug development and for planning of subsequent experiments. With current mass spectrometry (MS) techniques and the use of liver preparations, microsomes, the S9 supernatant, or homogenate, fortified with all appropriate cofactors, a tentative understanding of metabolites and metabolic routes can be obtained. On this basis, educated guesses about the involvement of potential metabolizing enzymes can also be made. Current analytical repertoire makes it relatively easy to devise appropriate routine assays for measuring metabolism of a compound. Naturally, it depends on the results of the identification of principal metabolites for a compound, but if a compound is metabolized, a routine method has to be developed for the identification of metabolizing enzymes.

### 4.1 First Incubations—Tentative Identification of Metabolites

The approach used for elucidating the *in vitro* biotransformation of an NCE is dependent on the needed specificity of the data to be obtained [50]. For an NCE with no known metabolites or any information about metabolic behavior, the easiest and time-efficient way to identify the *in vitro* metabolites is to produce all oxidative and conjugative metabolites with a single incubation with cultured hepatocytes or liver homogenates that include all CYP enzymes, as well as the principal conjugative enzymes, at least UGT, SULT, GST, and NAT (Table 2). After incubating the drug

substance with cultured hepatocytes or in liver homogenate together with the cofactors NADPH, UDPGA, GSH, and PAPS, needed for function of the enzymes, the metabolites formed are identified usually by liquid chromatography–mass spectrometry (LC/MS) techniques [51–53]. If some more specific information concerning the metabolism is needed, other types of liver preparations can be utilized, and the number of added cofactors can be varied. For example, if only phase I oxidative reactions need to be studied, liver microsomal fraction is used instead of the homogenate, as it contains all the oxidative CYP enzymes but lacks the GST and SULT enzymes. However, glucuronide conjugation may also occur with microsomal fraction, if cofactor UDPGA is used in incubation (for more detailed description, see Sections 1 and 2).

After the metabolite peak has been found from the chromatographic (LC/MS) data, the biotransformation can be identified according to the mass spectrum obtained for the metabolite. With modern LC/MS instruments equipped with atmospheric pressure ion sources, most commonly electrospray (ESI) [52, 53], usually only molecular ions are detected in the spectrum. This enables the elucidation of the shift in the molecular weight during the biotransformation from the substrate to metabolite. Some common shifts in molecular weights caused by metabolic reactions are given in the Table 6. Depending on the MS instrument type used, fragment ion MS data can also be obtained from the metabolite, enabling the elucidation of the biotransformation site in the substrate. With mass spectrometers capable of high resolution and good mass accuracy, accurate mass data from the metabolites can be

**TABLE 6 Changes in Molecular Mass Due to the Most Common Metabolic Reactions**

| Biotransformation | Mass Change (u) | Charasteristic Fragmentation[a] (NL = Neutral Loss) |
|---|---|---|
| Dehydrogenation (oxidation) | –2 | |
| Demethylation | –14 | |
| Desethylation | –28 | |
| Oxidative desulfuration | –32 | |
| Hydrogenation (reduction) | +2 | |
| Methylation | +14 | |
| Hydroxylation | +16 | |
| N/S-oxidation | +16 | |
| Epoxidation | +16 | |
| Acetylation | +42 | |
| Sulfation | +80 | –80 u (NL of $SO_3$) |
| Glucuronidation | +176 | –176 u (NL of $C_6H_8O_6$) |
| Glutathionation | +305[b] | –129 u (NL of $C_5H_7NO_3$) |
| | | –275 u (NL of $C_{10}H_{17}N_3O_6$, aryl-GSH) |
| Amino acid conjugation | | |
| Glycine | +57 | *m/z* 76 (Gly + $H^+$), *m/z* 74 (Gly – $H^+$) |
| Taurine | +107 | *m/z* 126 (Tau + $H^+$), *m/z* 124 (Tau – $H^+$) |

[a]The fragments of phase I metabolites are typically the same as for the substrate, or differ equally from the molecular weight.
[b]The glutathionation may lead to a number of different molecular mass changes; the addition of 305 u is the most simple case where glutathione replaces hydrogen atom in the substrate.

obtained, enabling differentiation in the biotransformation with a different molecular formula but the same nominal mass change, for example, distinquishing simultaneous demethylation and hydroxylation from hydrogenation, even though both reactions lead to an increase in molecular weight of two mass units. Additional H/D exchange studies can be conducted to distinguish different types of oxidations, for example, hydroxylation versus N-oxidation, giving the same mass shift but completely different biotransformation [54].

### 4.2 Absolute Structures—Reference Metabolites and NMR Studies

In some cases the data obtained by LC/MS methods as above does not give a clear enough picture about the structure of the formed metabolite, or the obtained tentative structure of the metabolite has to be elucidated in a more detailed manner. In these cases, the most simple and unambiguous approach is to synthesize the possible structures obtained for the metabolite (after LC/MS), and the LC/MS behavior of the synthesized compounds is compared with the unknown metabolite to confirm/exclude the elucidated structure. This of course may take more time for the chemists to synthesize the desired structure(s). To reliably conclude that the synthesized compound and the metabolite have the exact same structure, more than one single chromatographic method (different columns, different eluent pH) should be used to verify the similar LC/MS behavior.

An alternative possibility is to isolate and purify the metabolite from the incubation matrix, and to elucidate the exact structure using nuclear magnetic resonance (NMR) methods [55, 56]. However, the sensitivity of the NMR is not even close to the level of mass spectrometric methods, and about 100 nanograms of sample has to be obtained from purification to be able to acquire definitive NMR data in sensible time (i.e., overnight), and orders of magnitude more for more insensitive two-dimensional NMR experiments. Fortunately, in many cases, the most sensitive simple NMR measurement (i.e., basic one-dimensional $^{1}$H NMR) may already help in confirmation of the structure. Also online coupled LC/NMR instruments [57] are now available to avoid the purification step, but the chromatographic separation in these systems always suffers in comparison to more easily coupled LC/MS techniques.

### 4.3 Quantitative Studies

Besides the qualitative LC/MS studies in metabolite identification, quantitative data is also needed from a number of different studies with *in vitro* liver preparations. Among these, the most typical are the P450 enzyme interaction studies, direct drug–drug interaction studies, metabolizing CYP enzyme identification studies, and the estimation of metabolic stability (already mentioned). To save time, these analyses are often carried out semiquantitatively, that is, without spiked standard samples, and the results are reported as relative LC/MS peak areas of each detected compound in different samples, and these areas are expressed in percentages by comparing them to reference incubation samples (e.g., without an inhibitor or zero time incubation). In this stage of drug development, this sort of approach usually gives the accuracy needed, as the main purpose is to see if there are interactions or not,

or to evaluate the general level of metabolic stability. Also, in this phase the metabolites themselves are usually not available to be used as external standards for quantitation (for creating a calibration curve), and so if any spiked standard samples are used, the metabolites are quantified as "substrate equivalents." This in turn may sometimes lead to large differences between the results obtained and the real metabolite concentrations, as the mass spectrometric response between the drug and its metabolite may vary a lot due to different ionization and fragmentation properties, especially if the biotransformation occurs in the same functional group of the compound where the ionization occurs in the ionization process of the LC/MS analysis.

The main issue in the analyses of early drug development is to have good specificity, to be sure to monitor the correct compound (metabolite), and to have good linearity of detection response, so that the relative peak areas, obtained without calibration curves created by spiked standard samples, are representative of the concentration differences in different samples. When the need for high sensitivity is added to the list of requirements, the usual analytical system of choice is LC/MS-MS with triple quadrupole mass spectrometers.

### 4.4 Requirements for Analytical Instrumentation

Although the range of modern analytical instrumentation commercially available for metabolism studies is very large, today's metabolism studies are carried out using high performance liquid chromatography coupled online with mass spectrometry (LC/MS), and additionally some other detector types, such as UV-diode array (DAD, PDA), fluorescence, or radioactive detectors, are used. The additional detectors are solely used for "detecting purposes"—not for identification of the compound but for its quantitation or detection if the mass spectrometric response of the analyte is poor. The fluorescence detector can be used as a very sensitive detector for certain types of fluorescing compounds if suitable excitation and emission wavelengths are chosen, and the radioactive detector can be used for screening and quantification of the metabolites of the radiolabeled substrate. However, these are rarely needed in *in vitro* studies; the LC/MS instruments can give many types of data from a single run, including detection and identification of the metabolites, and at least semiquantitative concentration estimates.

Generally, any kind of mass spectrometer with HPLC compatible ion source can be used in metabolite idenfication and quantitation, but unfortunately, none of the mass spectrometer types is the optimal instrument for all kinds of studies (qualitative vs. quantitative), and thus at least two different types of instrument are required if all the studies are to be conducted at a state-of-the-art level. The strengths and drawbacks of the most common types of mass spectrometer studies are given in Table 7. The table includes only the time-of-flight/QTOF [58, 59], ion trap [60, 61], and triple quadrupole [62, 63] type instruments, but recently a number of hybrid instruments containing features from many instrument types have been introduced by manufacturers. Of these new instrument types, most useful in the analysis of small drug-like molecules are the triple quadrupole/linear ion trap instrument, where the last quadrupole of a traditional triple quadrupole instrument is replaced by a linear ion trap to increase the full scan sensitivity [64], and the ion-trap-TOF instrument, where the collision cell of QTOF is replaced by an ion trap to enable $MS^n$

**TABLE 7 Applicability of the Most Common Type Mass Spectrometers for *In Vitro* Drug Metabolism Studies (in LC/MS)**

| Parameters | Time-of-Flight (TOF-MS) | Ion Trap (IT-MS) | Triple Quadrupole (QQQ-MS) | Quadrupole-TOF (QTOF-MS) |
|---|---|---|---|---|
| Strengths | • High full scan sensitivity<br>• High resolution for exact mass measurement<br>• Very fast data acquisition (for HTS)<br>• Easy to operate | • High full scan sensitivity<br>• CID MS/MS and even $MS^3$ possibility | • Very high sensitivity for known analytes (MRM)<br>• High linear range<br>• High quality CID MS/MS, also with precursor ion and neutral loss scanning<br>• Very fast data acquisition (in MRM mode) | • High full scan sensitivity<br>• High resolution for exact mass measurement<br>• Very fast data acquisition (for HTS)<br>• CID MS/MS |
| Drawbacks | • Poor linear range<br>• No real MS/MS possibility (only "in-source" MS/MS) | • Poor linear range<br>• No exact mass possibility<br>• No precursor ion or neutral loss scanning MS/MS<br>• Slow scanning speed | • No exact mass possibility<br>• Poor full scan sensitivity (for screening) | • Poor linear range<br>• No precursor ion or neutral loss scanning MS/MS |
| Optimal use | Excellent for metabolite screening and identification of biotransformations | For metabolite screening and their tentative identifications; identification of biotransformation site | Excellent for quantitative analysis of known analytes; identification of biotransformation site; HTS applications | For metabolite screening; identification of biotransformations and their sites; HTS applications |
| Price | 200,000–300,000 € | 100,000–160,000 € | 160,000–300,000 € | 350,000–600,000 € |

experiments with high resolution for fragment ions [65]. For higher mass resolution studies, mostly needed when working with biomolecules, the Orbitrap [66] and ion cyclotron resonance [67] mass spectrometers offer superior performance.

The basic rule of thumb is that the time-of-flight mass spectrometers are the instruments of choice for a screening type of analysis (biotransformation screening and tentative metabolite identification), whereas the triple quadrupole instruments are superior in quantitative work. For metabolite screening the instrument needed should have good sensitivity with wide scan range to detect all unexpected metabolites simultaneously; also, it should be able to give qualitative data for at least tentative identification of the metabolites. The time-of-flight instruments have all this, the qualitative data coming from easy operation, accurate mass measurements, and in-source fragment ion data.

## 5 ACTIVE/REACTIVE/TOXIC METABOLITES

It would be very useful to know whether metabolism is needed for the biological action (pharmacologic or toxicologic) of drugs. Metabolic activation, which may be pharmacologically useful (prodrug metabolism) or toxicologically adverse, is an established primary mechanism of action for many drugs and toxicants [68–71]. Especially during early drug development, a screening system able to detect the formation of reactive drug metabolites would provide crucial information for the development program. In toxicity risk assessment, a robust and validated screening system is urgently needed, because metabolism is often needed for tissue toxicity, immunotoxicity, genotoxicity, and/or carcinogenicity to be initiated. In Table 8,

**TABLE 8 Potential *In Vitro* Screening Systems for Detecting the Formation of Reactive Metabolites**

| Type of Assay | Rationale | Example |
|---|---|---|
| Inhibition of drug metabolism | "Suicide" binding of a reactive metabolite with the enzyme | Tienilic acid |
| Formation of GSH conjugate | Binding of a reactive metabolite with GSH | Paracetamol |
| Covalent binding | Detection of bound reactive metabolite by radioactivity or by MS | Many carcinogens and mutagens |
| Lipid peroxidation | Detection of reactive oxygen species and/or peroxides by thiobarbituric acid reagent | Many halogenated substances |
| Target cell toxicity | Toxicity, for example, cell viability of metabolites produced either in the target cell itself or exogenously | Practically all substances are ultimately toxic to cells at high enough concentrations; mediation of toxicity by metabolism has to be incorporated into the test |

several potential screening systems are summarized. Generally, the generation of short-lived reactive metabolites is relatively straightforward to observe in various *in vitro* systems, although, it has to be stressed, none of them are adequately validated [72]. Because reactive metabolites are usually short-lived and unstable, they cannot usually be detected as such, but only after binding to trapping agents, cellular macromolecules, or other cell components. However, it is difficult to predict whether the formation of metabolites capable of binding with trapping agents or macromolecules will ultimately result in serious toxicity *in vivo* [73].

During early drug development, assays for mechanism-based inhibition are perhaps more widely used than other assays for detecting reactive metabolites. Mechanism-based inhibition can occur via the formation of metabolite intermediate complexes or via the strong covalent binding of reactive intermediates to the protein or heme of the CYP. Mechanism-based inhibition is terminated by enzyme resynthesis and is therefore usually long-lasting [74, 75]. In some cases, the metabolic product inactivates the enzyme completely. This is referred to as suicide inhibition. The most important phenomenon of mechanism-based inhibition is the time-, concentration-, and NADPH-dependent enzyme inactivation [2, 74]. Classical mechanism-based inhibitors include furafylline (CYP1A2) [76, 77] and gestodene (CYP3A4) [78]. It is worth noting that many compounds display both mechanism-based and competitive modes of inhibition; one such example is ticlopidine [79].

Although the need to detect and identify reactive metabolites is widely recognized for drug development, there are large gaps in our knowledge and also uncertainties about which kinds of strategies to employ. Detecting reactive metabolites does not necessarily mean that tissue toxicity or immunotoxicity will ensue. Metabolic activation seems to be behind many immunotoxic manifestations, but unequivocal evidence has been difficult to produce [80, 81]. One of the more important problems has been the selectivity of binding of reactive intermediates. Previously, it was thought that binding of reactive metabolites with cellular macromolecules is essentially a nonselective process. Now it has become more and more apparent that the process is primarily a selective process and even slightly different reactive species display large differences in their binding targets [82]. Development of screening methods that would differentiate toxicologically significant reactive metabolites from inactive ones would be a significant advance. Interesting analyses for future developments in research on metabolic activation from the industrial perspective are provided by Evans et al. [83] and Baillie [73].

## 6 TESTING CONFIGURATIONS

With the advent of combinatorial chemistry and large libraries of chemicals, a need for high throughput screening (HTS) has become ever more pressing in the pharmaceutical industry. Most major drug companies make use of various HTS configurations for efficient screening. Regarding the metabolic properties of NCEs, screenings of metabolic stability and drug–drug interactions have advanced to a considerable extent [22]. Often, rather sophisticated detection systems based on fluorescence are used in HTS systems. In the following, MS detection techniques are described.

### 6.1 Sample Preparation

Because of relatively defined incubation conditions with liver preparations or hepatocytes, sample preparation for LC/MS can usually be kept very simple. Usually the simple centrifugation of the incubation sample and careful pipetting of the supernatant into an HPLC autosampler vial is enough for an LC/MS compatible sample. However, when performing quantitative analyses, even for relative peak areas without external standards, the use of an internal standard is preferable. If long sample lists are acquired (e.g., overnight from an autosampler), the high buffer and salt concentrations together with other biomolecules or high concentration incubation media components may contaminate the mass spectrometer ion source, leading to general decrease in the sensitivity level toward the end of the sample list. By using an internal standard in the samples and using the relative peak area between the analyte and the internal standard as a result, instead of a plain analyte peak area, the effect of ion source contamination on the results can be avoided (as long as the sensitivity level stays acceptable).

### 6.2 Single Enzyme/Activity Systems

Many *in vitro* metabolism studies—such as interaction studies with CYP-selective inhibitors and antibodies, inhibition screens using studied NCEs with CYP-selective substrates, and studies with recombinant enzymes—are relatively routine and robust from the biochemical and analytical point of view, and there is a wide literature background for using these tools in various *in vitro* settings. Detailed descriptions of the methodological aspects of the more widely used CYP-selective substrates and inhibitors can be found in recent articles [35] (see also Table 5).

### 6.3 Medium to High Throughput Systems

Thorough surveys of medium to high throughput systems for *in vitro* absorption, distribution, metabolism, and excretion (ADME) have recently been published [84, 85]. When identifying the metabolizing or interacting CYP enzymes for larger numbers of compounds, the use of recombinant CYP enzymes and diagnostic CYP inhibitors are both suitable for HTS mode. With diagnostic CYP inhibitors, only one inhibitor concentration is used, low enough to be specific for inhibition of the target CYP but high enough for adequate inhibition if the study compound is metabolized via the same CYP. In the HTS approach the LC/MS analysis time is a crucial factor, and therefore very specific analytical methods for detecting only the monitored metabolites in a very short time are needed. The study compounds may also be incubated as mixtures of compounds, and the LC/MS method is adjusted to detect all the desired metabolites from each compound in a single analysis, again increasing the analytical challenge. However, the possible interaction with the simultaneously incubated compounds may affect the results obtained; especially with recombinant enzymes, the formation of monitored metabolite is a sign of the involvement of the certain CYP enzyme in the metabolite formation.

For fast screening of P450 inhibition, it is possible to employ a cocktail-type analysis of known CYP-selective biotransformations in liver microsomal incubations. A number of assays for different CYP enzymes have been developed and

published [22]. The most simple cocktails contain only a few important CYP isoforms, whereas the most complete assays contain all nine of the most important CYPs and also more than one probe reaction for some isoforms [32]. If a HTS for a large number of study compounds is needed, usually only one concentration per compound is needed to give a rough estimate of the level of interaction with each CYP. However, the development and use of the cocktail approach is much more challenging than the use of single-substrate incubations, as the substrate concentrations have to be optimized so as not to cause any interactions between each other; also the reliable LC/MS analysis of many chemically different metabolites with a single run is quite challenging. If very high throughput is needed, the studied compounds may be delivered into the incubations as cocktails of several compounds, and the go or no-go decision is then given to all of the compounds tested together. However, this approach contains a risk of losing good lead candidates, as only one potent inhibitor in the mixture of simultaneously tested compounds may lead to wasting the other, noninhibiting, compounds.

### 6.4 Analytical Considerations

From the analytical point of view, the issues covered in Section 4 are all valid here as well. The quantitative analyses in these studies are usually carried out as relative LC/MS peak areas, without any spiked calibration standard samples. The excellent linear range and specificity of triple quadrupole mass spectrometers makes them the best choice for LC/MS instrumentation, enabling very sensitive detection of the low metabolite concentrations. Also, when detecting metabolites from several compounds incubated simultaneously in the same well/vessel, the triple quadrupole instruments provide the most specific detection for each analyte. Even structural analogues with the same molecular weight and the same retention times can still be separated, as long as the compounds have different fragmentation pathways in the collision-induced dissociation. However, with HTS applications and constantly changing analytes, the time-of-flight mass spectrometers (or ion trap instruments) may offer a good alternative for instrumentation, as the excellent sensitivity for acquiring data over a high mass range enables the detection of all different analytes, without the need to adjust the detection parameters before the analysis, as is the case with triple quadrupole instruments. Therefore, with LC/TOF-MS and LC/QTOF-MS, the same generic method may be used for a large set of completely different compounds in HTS applications [86]. With QTOF, the MS/MS option in data acquisition is available. However, if the application is for P450 inhibition screening, where detected compounds remain the same all the time, the triple quadrupole mass spectrometer is the best possible instrument for the analysis.

When the aim in analyses is to produce as much data as possible from a single LC/MS run, many software packages can control the data acquisition with so-called data-dependent operations [87], which may be particularly useful with MS/MS instruments, as the detection of high abundance ions with the certain $m/z$ in the spectrum of a high abundance ion may trigger some other type of data acquisition using this ion as a precursor mass—such as product ion scanning and selected reaction monitoring. This way, both qualitative and quantitative data can be acquired simultaneously. If the molecular weights of the studied metabolites are known

before analysis, the same acquisition functions can of course be set manually, which usually provides better quality data.

## 7 PRIORITIES OF *IN VITRO* METABOLISM STUDIES AND *IN VITRO–IN VIVO* EXTRAPOLATION

One of the important questions concerning the starter *in vitro* test of metabolism is: Which one is best as a first test? Because each molecule is rather unique in terms of metabolism and kinetics, we would need a rather comprehensive test, which should cover as many salient features of a compound as possible. But this requirement is contradictory to simplicity, rapidity, and economy, which are needed for early testing.

There is no fixed single strategy for performing metabolic studies during early drug development. Rather, major drug companies employ their own slighly variable schemes. Table 9 presents one possible strategy of metabolism studies in early drug development, which is described below. Whatever strategy is adopted, assessment of metabolic stability is very important to perform early on in the drug development program, because it determines, despite many caveats inherent in current testing systems, to a large extent the usefulness of the molecule and gives an important value for *in vitro–in vivo* extrapolation.

### 7.1 Priorities of *In Vitro* Metabolism Studies

Determination of basic physicochemical properties is the prerequisite for every subsequent property determination. When developing adequate analytical methods for NCEs, it is important to know solubility, p$K_a$, lipophilicity, and also chemical

**TABLE 9 Example of a Rough Priority Scheme for *In Vitro* Studies Related to Metabolic Properties of an NCE**

| Properties of an NCE | Remarks |
|---|---|
| Solubility, lipophilicity, p$K_a$, chemical stability, possible inpurities, plasma protein binding (free fraction) | These data are absolutely required before metabolism studies to prevent unpleasant surprises; free fraction (plasma protein binding) needed for proper extrapolation |
| Metabolic stability in microsomes/ homogenates or hepatocytes | A crucial piece of information for the extrapolation of hepatic clearance |
| Metabolite identification in microsomes, homogenates, or hepatocytes | Human hepatocytes are an enzyme source of choice if available |
| Inhibitory interactions in human liver microsomes | Also help pinpoint enzymes with affinity toward the studied compound |
| Identification of metabolizing CYPs | Primarily employed: diagnostic inhibitors, recombinant enzymes, correlation analysis (if a liver bank is available) |
| Induction of drug metabolism | Preferably in human hepatocytes if available |
| Metabolic stability and identification of metabolites in animal liver preparations | To search for appropriate species for certain toxicity studies |

stability. A cost-effective approach is to integrate method development, simple degradation study, and physicochemical property characterization together with plasma protein binding and nonspecific binding. It is feasible to start *in vitro* metabolism and permeation studies after the physicochemical, degradation, and binding properties of an NCE are already known. It is easier to incubate, to analyze, and most importantly to predict the *in vivo* situation, if possible interfering problems are dealt with before starting the metabolism experiments.

Metabolism data using human liver preparations are irreplaceable if the drug candidates are intended for human use. Since a rat is not human, it is not feasible to start metabolism studies with rat or any other animal liver preparations; their time is later, when animal models for toxicological studies are selected. A good practice is to start metabolism studies with human and several obvious animal models, but resources might be wasted if the compound is rejected. The first human *in vitro* metabolism studies should be conducted with a system involving all the possible drug-metabolizing enzymes and cofactors for the enzymes. High initial concentrations of parent compound should be used and all the possible (predicted and unpredicted, major and minor) metabolites should be identified. Special attention needs to be addressed to those metabolites potentially associated with toxicity, such as acyl glucuronides, epoxides, glutathione pathway metabolites, and hydroxyl amines, because their presence indicates a need for lead optimization. This primary metabolism screen for tentative metabolic stability and metabolite identification could best be conducted with a LC-TOF-MS or LC-QTOF-MS instrument. The preferred liver preparation would be human liver homogenate (pool of several individuals) due to its relatively easy access compared to hepatocytes. A permanent hepatocyte-like cell line would be highly desirable, but no such cell line has yet been found or developed.

For the compounds that survive the primary metabolism screen, several second tier studies are needed. First, kinetic metabolic stability and metabolite formation studies are needed for proper *in vivo* prediction. This is usually done in human liver microsomes, unless the metabolites found in the primary screen indicate homogenate, S9, or cytosol has to be used, with several time points, concentrations, and replicates. Second, rate-limiting CYP enzymes, or other enzymes if indicated by primary screen, should be identified to assess potential drug–drug interactions and also kinetic aspects. Human liver microsomal incubations of an NCE together with CYP-specific inhibitors or antibodies give a clear indication of which CYP enzymes are involved. If the picture is completed by incubations with recombinant CYP enzymes and both tests point to the same CYP enzyme, an involvement of a certain CYP enzyme in a particular biotransformation can be relatively reliably confirmed. Third, a microsomal study to be conducted at this stage is CYP enzyme inhibition screen. A simple cocktail approach with one concentration of NCE aiming at "% activity inhibited" will usually do the trick. If something risky is observed, more thorough investigations including mechanistic studies should be done.

In the third round and for the molecules that remain, one should study CYP induction in human hepatocytes. Once the hepatocytes are employed, it will be cost efficient to study the other metabolic properties on hepatocytes also: confirmation of the previous investigations in microsomes/homogenates/S9. Thus, one should again study metabolic stability, metabolite identification, CYP inhibition and induction, and metabolizing CYP enzymes. A good add-on would be a study of accumula-

tion of the parent and metabolites in hepatocyte cells, because this information may be crucial for the interpretation of *in vitro–in vivo* extrapolation.

Finally, the rat (or mouse or dog or monkey) becomes relevant. If the candidate drug passes the *in vitro* tests with human preparations, a multispecies metabolism study would be appropriate to find out the most relevant species for toxicological investigations. Special attention should be given to the metabolites formed—not metabolic stability—as every relevant human metabolite should be found from the toxicological species.

### 7.2 *In Vitro–In Vivo* Extrapolation

Metabolic clearance (intrinsic clearance, $CL_{int}$) is an important determinant for extrapolation purposes. $CL_{int}$ is a direct measure of the efficacy of an enzyme to metabolize a substrate. Knowing that the drug concentrations in *in vivo* situations are usually far below their $K_m$ values (i.e., under linear velocity conditions), $CL_{int}$ is equal to the ratio of $V_{max}/K_m$ [88, 89]. When the contribution of multiple enzymes to the metabolism is assumed, the net *in vitro* $CL_{int}$ in the whole study system can be expressed as the sum of each metabolic pathway [89, 90].

Several elements limit the reliability of predicting $CL_{int}$ on the basis of *in vitro* studies. The most important factor to be taken into account is the nonspecific binding of the substrate to the microsomal protein fraction. Generally, for drugs with high protein binding affinity, underestimations with $CL_{int}$ can occur if protein binding is taken into account. However, inclusion of both microsomal and plasma protein binding usually results in good agreement between extrapolated and actual clearance values [91, 92].

Altered hepatic blood flow may have an impact on the accuracy of the *in vitro–in vivo* extrapolation. On the basis of their hepatic clearances, drugs can be classified either as low clearance or high clearance compounds. In the first case, the hepatic blood flow has a minor effect on the total clearance of the drug (enzyme-limited clearance), whereas in the latter case, changes in the hepatic blood flow will have a drastic effect on the total clearance of the drug (flow-limited clearance). However, the clinical outcome is hard to predict, as conditions affecting portal blood flow usually involve concurrent complications [2, 93].

In general, *in vitro–in vivo* extrapolations are performed with two objectives in mind: prediction of the intrinsic clearance ($CL_{int}$) of an NCE and prediction of potential drug–drug interactions. Other pharmacokinetic parameters, such as plasma half-life, volume of distribution, and oral bioavailability, have also to be considered for the estimation of *in vivo* kinetics on the basis of *in vitro* studies. Many *in vivo* factors affect the results of extrapolation. One of these is the binding to plasma and tissue proteins and, ultimately, the distribution volume of the drug. Details of various models and equations can be found in recent reviews [94, 95].

## 8 VALIDATION OF *IN VITRO* SYSTEMS

The main purpose of *in vitro* tests used in preclinical departments of drug companies is to aid drug development, not registration. Naturally, companies are interested in

the precision and accuracy of *in vitro* tests, but there is no requirement of formal validation. Consequently, validation is done primarily in-house. If the results of the tests are being used in regulatory applications, then a certain amount of validation—even if not formal—is required. FDA, EMEA, and authorities in Japan have all published some guidelines under which circumstances of *in vitro* drug interaction results can be used to support marketing authorization. The current view is that prediction of drug–drug interactions on the basis of *in vitro* studies is most advanced from the validation and reliability points of view [1, 96–98].

*In vitro* systems are generally useful in identifying compounds with severe potential liabilities, such as metabolic instability, very high affinity to a principal metabolizing enzyme, or metabolism by a polymorphic enzyme (e.g., CYP2D6). Whether false positives in these respects are rare or common, is not known. Also, the frequency of false negatives is not known. Discussions with company scientists about their unpublished data have led to a view that the predictive power of various *in vitro* tests tends to weaken when the number and structural variability of tested substances increase. There is no consensus about what is the acceptable frequency of false positives or negatives, or more precisely, what are the acceptable tolerances for findings. The only remedy for these uncertainties is validation according to the established principles [99].

## 9 CONCLUSION

Metabolism is a major determinant governing both pharmacokinetics and clinical response of the majority of drugs and a great deal of effort is now directed at assessing key metabolic parameters in the early stages of drug development. Several *in vitro* methods are now available for determination of metabolic features, often yielding data that reasonably well predict *in vivo* behavior of the studied drug molecules [1]. Further development and refining of these methods will provide us with methods having increased precision and robustness, allowing for highly reliable analysis and prediction of metabolic features. In addition, analogous methodology will be employed on predicting absorption, organ uptake, and efflux mediated by various transporter systems, plasma protein binding, and cellular determinants of intrinsic clearance. As an outcome of this development, we are already witnessing a diminishing number of drug candidates being withdrawn from clinical studies (or the market) due to major kinetic problems, such as strong metabolism induction or interaction potential.

An important consideration concerns the integration of results from various *in vitro* tests. Pharmacokinetics is actually an integrated whole, and different pieces are valuable only in the context of the whole. Development of modeling and simulation tools, for example, physiologically based pharmacokinetic models, will aid in incorporation of the results from *in vitro* studies into realistic models and in understanding genetic, physiological, and pathological factors affecting the behavior of an NCE when a substance is administered for the first time to a human being. Although much has already been done, there is still a need for reliable and robust pharmacokinetic models that would incorporate various factors and processes affecting the fate of a molecule. With the help of these *in silico* models, information from *in vitro* tests becomes continuously integrated into the database.

The use of human-derived biological systems seems rather natural, because we are interested primarily in human metabolism and kinetics. Human hepatocyte assay is regarded as the gold standard for metabolic screening, but there are several problems in this approach. There is an urgent need to develop a renewable, immortal hepatocyte-like permanent cellular model, which would express all the necessary enzymes and transporters at the level resembling *in vivo* hepatocyte. One promising candidate is a recently developed hepatic cell line, HepaRG, which expressed most enzymes and transporters at a relatively high level [100, 101].

An important drawback, at least in some academic laboratories that work to develop *in vitro* techniques to measure drug metabolism, is the absolute necessity of expensive mass spectrometric techniques for analytical work. However, we feel that rational and productive research on *in vitro* drug metabolism tests is not possible without up-to-date analytical techniques. Analytical chemistry is at the heart of early drug development.

## REFERENCES


1. Pelkonen O, Raunio H. *In vitro* screening of drug metabolism during drug development: Can we trust the predictions? *Expert Opin Drug Metab Toxicol* 2005;1:49–60.
2. Lin JH, Lu AYH. Role of pharmacokinetics and metabolism in drug discovery and development. *Pharmacol Rev* 1997;49:403–449.
3. Van de Waterbeemd H, Gifford E. ADMET *in silico* modelling: towards prediction paradise? *Nat Rev Drug Dev* 2003;2:192–203.
4. Guengerich FP, Rendic S. Human cytochrome P450 enzymes (human CYPs): human cytochrome P450 enzymes, a status report summarizing their reactions, substrates, inducers, and inhibitors—1st update. *Drug Metab Rev* 2002;34:1–450.
5. Coecke S, Ahr H, Blaauboer BJ, Bremer S, Casati S, Castell J, Combes R, Corvi R, Crespi CL, Cunningham ML, Elaut G, Eletti B, Freidig A, Gennari A, Ghersi-Egea JF, Guillouzo A, Hartung T, Hoet P, Ingelman-Sundberg M, Munn S, Janssens W, Ladstetter B, Leahy D, Long A, Meneguz A, Monshouwer M, Morath S, Nagelkerke F, Pelkonen O, Ponti J, Prieto P, Richert L, Sabbioni E, Schaack B, Steiling W, Testai E, Vericat JA, Worth A. Metabolism: a bottleneck in *in vitro* toxicological test development. The Report and Recommendations of ECVAM Workshop 54. *Altern Lab Anim* 2006;34:49–84.
6. Pelkonen O, Baumann A, Reichel A. *Pharmacokinetic Challenges in Drug Discovery*. Ernst Schering Research Foundation Workshop 37. Berlin: Springer; 2002.
7. Kremers P. Liver microsomes: a convenient tool for metabolism studies but. . . . In Boobis AR, Kremers P, Pelkonen O, Pithan K, Eds. *European Symposium on the Prediction of Drug Metabolism in Man: Progress and Problems*. Office for Official Publications of the European Communities; 1999, pp 38–52.
8. Rodrigues AD. Integrated cytochrome P450 reaction phenotyping: attempting to bridge the gap between cDNA-expressed cytochromes P450 and native human liver microsomes. *Biochem Pharmacol* 1999;57:465–480.
9. Guillouzo A. Acquisition and use of human *in vitro* liver preparations. *Cell Biol Toxicol* 1995;11:141–145.
10. Guillouzo A, Morel F, Fardel O, Meunier B. Use of human hepatocyte cultures for drug metabolism studies. *Toxicology* 1993;82:209–219.

11. Ferrero JL, Brendel K. Liver slices as a model in drug metabolism. *Adv Pharmacol* 1997;43:131–169.
12. Allen DD, Caviedes R, Cardenas AM, Shimahara T, Segura-Aguilar J, Caviedes PA. Cell lines as *in vitro* models for drug screening and toxicity studies. *Drug Dev Ind Pharm* 2005;31:757–768.
13. Brandon EFA, Raap CD, Meijerman I, Beijnen JH, Schellens JHM. An update on *in vitro* test methods in human hepatic drug biotransformation research: pros and cons. *Toxicol Appl Pharmacol* 2003;189:233–246.
14. Plant N. Strategies for using *in vitro* screens in drug metabolism. *DDT* 2004;9:328–336.
15. Gomez-Lechon MJ, Donato MT, Castell JV, Jover R. Human hepatocytes in primary culture: the choice to investigate drug metabolism in man. *Curr Drug Metab* 2004;5:443–462.
16. Andersson TB, Sjöberg H, Hoffmann K-J, Boobis AR, Watts P, Edwards RJ, Lake BG, Price RJ, Renwick AB, Gómez-Lechón MJ, Castell JV, Ingelman-Sundberg M, Hidestrand M, Goldfarb PS, Lewis DFV, Corcos L, Guillouzo A, Taavitsainen P, Pelkonen O. An assessment of human liver-derived *in vitro* systems to predict the *in vivo* metabolism and clearance of almokalant. *Drug Metab Dispos* 2001;29:712–720.
17. Salonen JS, Nyman L, Boobis AR, Edwards RJ, Watts P, Lake BG, Price RJ, Renwick AB, Gómez-Lechón MJ, Castell JV, Ingelman-Sundberg M, Hidestrand M, Guillouzo A, Corcos L, Goldfarb PS, Lewis DFV, Taavitsainen P, Pelkonen O. Comparative studies on the CYP-associated metabolism and interaction potential of selegiline between human liver-derived *in vitro* systems. *Drug Metab Dispos* 2003;31:1093–1102.
18. Pelkonen O, Myllynen P, Taavitsainen P, Boobis AR, Watts P, Lake BG, Price RJ, Renwick AB, Gómez-Lechón MJ, Castell JV, Ingelman-Sundberg M, Hidestrand M, Guillouzo A, Corcos L, Goldfarb PS, Lewis DFV. Carbamazepine: a "blind" assessment of CYP-associated metabolism and interactions in human liver-derived *in vitro* systems. *Xenobiotica* 2001;31:321–343.
19. Thompson TN. Early ADME in support of drug discovery: the role of metabolic stability studies. *Curr Drug Metab* 2000;1:215–241.
20. Gebhardt R, Hengstler JG, Muller D, Glockner R, Buenning P, Laube B, Schmelzer E, Ullrich M, Utesch D, Hewitt N, Ringel M, Hilz BR, Bader A, Langsch A, Koose T, Burger HJ, Maas J, Oesch F. New hepatocyte *in vitro* systems for drug metabolism: metabolic capacity and recommendations for application in basic research and drug development, standard operation procedures. *Drug Metab Rev* 2003;35:145–213.
21. Masimirembwa CM, Bredberg U, Andersson TB. Metabolic stability for drug discovery and development: pharmacokinetic and biochemical challenges. *Clin Pharmacokinet* 2003;42:515–528.
22. Ansede JH, Thakker DR. High-throughput screening for stability and inhibitory activity of compounds toward cytochrome P450-mediated metabolism. *J Pharm Sci* 2004;93: 239–255.
23. Obach RS, Baxter JG, Liston TE, Silber BM, Jones BC, MacIntyre F, Rance DJ, Wastall P. The prediction of human pharmacokinetic parameters from preclinical and *in vitro* metabolism data. *J Pharmacol Exp Ther* 1999;283:46–58.
24. Clarke SE, Jeffrey P. Utility of metabolic stability screening: comparison of *in vitro* and *in vivo* clearance. *Xenobiotica* 2001;31:591–598.
25. Pelkonen O, Hukkanen J, Honkakoski P, Hakkola J, Viitala P, Raunio H. *In vitro* screening of cytochrome P450 induction potential. In Pelkonen O, Baumann A, Reichel A, Eds. *Pharmacokinetics Challenges in Drug Discovery*. Ernst Schering Research Foundation Workshop 37. Berlin: Springer Verlag; 2002, pp 105–137.

26. Honkakoski P, Negishi M. Regulation of cytochrome P450 (CYP) genes by nuclear receptors. *Biochem J* 2000;347:321–337.
27. Honkakoski P. Nuclear receptors CAR and PXR in metabolism and elimination of drugs. *Curr Pharmacogenomics* 2003;1:75–85.
28. Lindberg RL, Negishi M. Alteration of mouse cytochrome P450coh substrate specificity by mutation of a single amino-acid residue. *Nature* 1989;339:632–634.
29. Pelkonen O, Breimer DD. Role of environmental factors in the pharmacokinetics of drugs: considerations with respect to animal models, P-450 enzymes, and probe drugs. In Welling PG, Balant LP, Eds. *Handbook of Experimental Pharmacology*, Volume 110. Berlin: Springer-Verlag; 1994, pp 289–332.
30. Friedman MA, Woodcock J, Lumpkin MM, Shuren JE, Hass AE, Thompson LJ. The safety of newly approved medicines: do recent market removals mean there is a problem? *JAMA* 1999;281:1728–1734.
31. Lasser KE, Allen PD, Woolhandler SJ, Himmelstein DU, Wolfe SM, Bor DH. Timing of new black box warnings and withdrawals for prescription medications. *JAMA* 2002; 287:2215–2220.
32. Turpeinen M, Uusitalo J, Jalonen J, Pelkonen O. Multiple P450 substrates in a single run: rapid and comprehensive *in vitro* interaction assay. *Eur J Pharm Sci* 2005;24:123–132.
33. Tucker GT, Houston JB, Huang S-M. EUFEPS Conference Report. Optimising drug development: strategies to assess drug metabolism/transporter interaction potential—towards a consensus. *Eur J Pharm Sci* 2001;13:417–428.
34. Yuan R, Madani S, Wei X-X, Reynolds K, Huang S-M. Evaluation of cytochrome P450 probe substrates commonly used by the pharmaceutical industry to study *in vitro* drug interactions. *Drug Metab Dispos* 2002;30:1311–1319.
35. Walsky RL, Obach RS. Validated assays for human cytochrome P450 activities. *Drug Metab Dispos* 2004;32:647–660.
36. Streetman DS, Bertino JS Jr, Nafziger AN. Phenotyping of drug-metabolizing enzymes in adults: a review of *in-vivo* cytochrome P450 phenotyping probes. *Pharmacogenetics* 2000;10:187–216.
37. Bertz RJ, Granneman GR. Use of *in vitro* and *in vivo* data to estimate the likelihood of metabolic pharmacokinetic interactions. *Clin Pharmacokinet* 1997;32:210–258.
38. Kremers P. *In vitro* tests for predicting drug–drug interactions: the need for validated procedures. *Pharmacol Toxicol* 2002;91:209–217.
39. Pelkonen O, Maenpaa J, Taavitsainen P, Rautio A, Raunio H. Inhibition and induction of human cytochrome P450 (CYP) enzymes. *Xenobiotica* 1998;28:1203–1253.
40. Pelkonen O. Human CYPs: *in vivo* and clinical aspects. *Drug Metab Rev* 2000; 34:37–46.
41. Pelkonen O, Turpeinen M, Uusitalo J, Rautio A, Raunio H. Prediction of drug metabolism and interactions on the basis of *in vitro* investigations. *Basic Clin Pharmacol Toxicol* 2005;96:167–175.
42. Lewis DF. Human cytochromes P450 associated with the phase 1 metabolism of drugs and other xenobiotics: a compilation of substrates and inhibitors of the CYP1, CYP2 and CYP3 families. *Curr Med Chem* 2003;10:1955–1972.
43. Ingelman-Sundberg M. Pharmacogenetics of cytochrome *P*450 and its applications in drug therapy: the past, present and future. *Trends Pharmacol Sci* 2004;25:193–200.
44. Totah RA, Rettie AE. Cytochrome P450 2C8: substrates, inhibitors, pharmacogenetics, and clinical relevance. *Clin Pharm Ther* 2005;77:341–352.

45. Rowland Yeo K, Rostami-Hodjegan A, Tucker GT. Abundance of cytochromes P450 in human liver: a meta analysis. *Br J Clin Pharmacol* 2004;57:687–688.
46. Friedberg T, Henderson CJ, Pritchard MP, Wolf CR. *In vivo* and *in vitro* recombinant DNA technology as a powerful tool in drug development. In Woolf TF, Ed. *Handbook of Drug Metabolism*. New York: Marcel Dekker; 1999, pp 322–362.
47. Moody GC, Griffin SJ, Mather AN, McGinnity DF, Riley RJ. Fully automated analysis of activities catalysed by the major human liver cytochrome P450 (CYP) enzymes: assessment of human CYP inhibition potential. *Xenobiotica* 1999;29:53–75.
48. McGinnity DF, Riley RJ. Predicting drug pharmacokinetics in humans from *in vitro* metabolism studies. *Biochem Soc Trans* 2001;29(Pt 2):135–139.
49. Galetin A, Brown C, Hallifax D, Ito K, Houston JB. Utility of recombinant enzyme kinetics in prediction of human clearance: impact of variability, CYP3A5, and CYP2C19 on CYP3A4 probe substrates. *Drug Metab Dispos* 2004;32:1411–1420.
50. Nassar A-EF, Talaat RE. Strategies for dealing with metabolite elucidation in drug discovery and development. *DDT* 2004;9:317–327.
51. Kostiainen R, Kotiaho T, Kuuranne T, Auriola S. Liquid chromatography/atmospheric pressure ionization–mass spectrometry in drug metabolism studies. *J Mass Spectrom* 2003;38:357–372.
52. Cole MJ, Janiszewski JS, Fouda HG. Electrospray mass spectrometry in contemporary drug metabolism and pharmacokinetics. In Pramanik BN, Ganguly AK, Gross ML, Eds. *Applled Electrospray Mass Spectrometry*. New York: Marcel Dekker; 2002, pp 211–249.
53. Lee MS, Kerns EH. LC/MS applications in drug development. *Mass Spectrom Rev* 1999;18:187–279.
54. Tolonen A, Turpeinen M, Uusitalo J, Pelkonen O. A simple method for differentiation of monoisotropic drug metabolites with hydrogen–deuterium exchange LC/MS. *Eur J Pharm Sci* 2005;25:155–162.
55. Pellecchia M, Sem DS, Wutrich K. NMR in drug discovery. *Nat Rev Drug Discov* 2002;1:211–219.
56. Keifer PA. NMR spectroscopy in drug discovery: tools for combinatorial chemistry, natural products, and metabolism research. *Prog Drug Res* 2000;55:137–211.
57. Corcoral O, Spraul M. LC-NMR-MS in drug discovery. *Drug Discov Today* 2003;8:624–631.
58. Zhang N, Fountain ST, Bi H, Rossi DT. Quantification and rapid metabolite identification in drug discovery using API time-of-flight LC/MS. *Anal Chem* 2000;72:800–806.
59. Plumb RS, Stumpf CL, Granger JH, Castro-Perez J, Haselden JN, Dear GJ. Use of liquid chromatography/time-of-flight mass spectrometry and multivariate statistical analysis shows promise for the detection of drug metabolites in biological fluids. *Rapid Commun Mass Spectrom* 2003;17:2632–2638.
60. March RE, Todd JFJ, Eds. *Practical Aspects of Ion Trap Mass Spectrometry*. Boca Raton, FL: CRC Press; 1995.
61. Kantharaj E, Tuytelaars A, Proost PEA, Ongel Z, van Assouw HP, Gilissen RAHJ. Simultaneous measurement of drug metabolic stability and identification of metabolites using ion-trap mass spectrometry. *Rapid Commun Mass Spectrom* 2003;17:2661–2668.
62. Brewer E, Henion J. Atmospheric pressure ionization LC/MS/MS techniques for drug disposition studies. *J Pharm Sci* 1998;87:395–402.

63. Xu X, Veals J, Korfmacher WA. Comparison of conventional and enhanced mass resolution triple-quadrupole mass spectrometers for discovery bioanalytical applications. *Rapid Commun Mass Spectrom* 2003;17:832–837.
64. Hopfgartner G, Varesio E, Tschappat V, Grivet C, Bourgogne E, Leuthold LA. Triple quadrupole linear ion trap mass spectrometer for the analysis of small molecules and macromolecules. *J Mass Spectrom* 2004;39:845–855.
65. Collings BA, Campbell JM, Mao D, Douglas DJ. A combined linear ion trap time-of-flight system with improved performance and $MS^n$ capabilities. *Rapid Commun Mass Spectrom* 2001;15:1777–1795.
66. Hu Q, Noll RJ, Li H, Makarov A, Hardman M, Cooks RG. The Orbitrap: a new mass spectrometer. *J Mass Spectrom* 2005;40:430–443.
67. Marshall AG, Hendrickson CL, Jackson GS. Fourier transform ion cyclotron resonance mass spectrometry: a primer. *Mass Spectrom Rev* 1998;17:1–35.
68. Park BK, Kitteringham NR, Maggs JL, Pirmohamed M, Williams DP. The role of metabolic activation in drug-induced hepatotoxicity. *Annu Rev Pharmacol Toxicol* 2005; 45:177–202.
69. Walgren JL, Mitchell MD, Thompson DC. Role of metabolism in drug-induced idiosyncratic hepatotoxicity. *Crit Rev Toxicol* 2005;35:325–361.
70. Zhou S, Chan E, Duan W, Huang M, Chen Y-Z. Drug bioactivation, covalent binding to target proteins and toxicity relevance. *Drug Metab Rev* 2005;1:41–213.
71. Kaplowitz N. Idiosyncratic drug hepatotoxicity. *Nat Rev Drug Discov* 2005;4:489–499.
72. Baillie TA, Cayen MN, Fouda H, Gerson RJ, Green JD, Grossman SJ, Klunk LJ, LeBlanc B, Perkins DG, Shipley LA. Contemporary issues in toxicology. Drug metabolites in safety testing. *Toxicol Appl Pharmacol* 2002;182:188–196.
73. Baillie TA. Future of toxicology—metabolic activation and drug design: challenges and opportunities in chemical toxicology. *Chem Res Toxicol* 2006;19:889–893.
74. Halpert JR. Structural basis of selective cytochrome P450 inhibition. *Annu Rev Pharmacol Toxicol* 1995;35:29–53.
75. Kent UM, Juschyshyn MI, Hollenberg PF. Mechanism-based inactivators as probes of cytochrome P450 structure and function. *Curr Drug Metab* 2001;2:215–243.
76. Sesardic D, Boobis AR, Murray BP, Murray S, Segura J, de la Torre R, Davies DS. Furafylline is a potent and selective inhibitor of cytochrome P450IA2 in man. *Br J Clin Pharmacol* 1990;29:651–663.
77. Kunze KL, Trager WF. Isoform-selective mechanism-based inhibition of human cytochrome P450 1A2 by furafylline. *Chem Res Toxicol* 1993;6:649–656.
78. Back DJ, Houlgrave R, Tjia JF, Ward S, Orme ML. Effect of the progestogens, gestodene, 3-keto desogestrel, levonorgestrel, norethisterone and norgestimate on the oxidation of ethinyloestradiol and other substrates by human liver microsomes. *J Steroid Biochem Mol Biol* 1991;38:219–225.
79. Turpeinen M, Nieminen R, Juntunen T, Taavitsainen P, Raunio H, Pelkonen O. Selective inhibition of CYP2B6-catalyzed bupropion hydroxylation in human liver microsomes *in vitro*. *Drug Metab Dispos* 2004;36:626–631.
80. Uetrecht J. Screening for the potential of a drug candidate to cause idiosyncratic drug reactions. *DDT* 2003;8:832–837.
81. Williams DP, Park BK. Idiosyncratic toxicity: the role of toxicophores and bioactivation. *DDT* 2003;8:1044–1050.
82. Dennehy MK, Richards KAM, Wernke GR, Shyr GR, Liebler DC. Cytosolic and nuclear protein targets of thiol-reactive electrophiles. *Chem Res Toxicol* 2006;19:20–29.

83. Evans DC, Watt AP, Nicoll-Griffith DA, Baillie TA. Drug–protein covalent adducts: an industry perspective on minimizing the potential for drug bioactivation in drug discovery and development. *Chem Res Toxicol* 2004;17:3–16.
84. van de Waterbeemd H. High-throughput and *in silico* techniques in drug metabolism and pharmacokinetics. *Curr Opin Drug Discov Dev* 2002;5:33–43.
85. Wunberg T, Hendrix M, Hillisch A, Lobell M, Meier H, Schmeck C, Wild H, Hinzen B. Improving the hit-to-lead process: data-driven assessment of drug-like and lead-like screening hits. *Drug Discov Today* 2006;11:175–180.
86. Castro-Perez J, Plumb R, Granger JH, Beattie I, Joncour K, Wright A. Increasing throughput and information content for *in vitro* drug metabolism experiments using ultra-performance liquid chromatography coupled to a quadrupole time-of-flight mass spectrometer. *Rapid Commun Mass Spectrom* 2005;19:843–848.
87. Lopez LL, Yu X, Cui D, Davis MR. Identification of drug metabolites in biological matrices by intelligent automated liquid chromatography/tandem mass spectrometry. *Rapid Commun Mass Spectrom* 1999;12:1756–1760.
88. Houston JB. Utility of *in vitro* drug metabolism data in predicting *in vivo* metabolic clearance. *Biochem Pharmacol* 1994;47:1469–1479.
89. Iwatsubo T, Hirota N, Ooie T, Suzuki H, Shimada N, Chiba K, Ishizaki T, Green CE, Tyson CA, Sugiyama Y. Prediction of *in vivo* drug metabolism in the human liver from *in vitro* metabolism data. *Pharmacol Ther* 1997;73:147–171.
90. Ito K, Iwatsubo T, Kanamitsu S, Nakajima Y, Sugiyama Y. Quantitative prediction of *in vivo* drug clearance and drug interactions from *in vitro* data on metabolism, together with binding and transport. *Annu Rev Pharmacol Toxicol* 1998;38:461–499.
91. Obach RS. Prediction of human clearance of twenty-nine drugs from hepatic microsomal intrinsic clearance data: an examination of *in vitro* half-life approach and nonspecific binding to microsomes. *Drug Metab Dispos* 1999;27:1350–1359.
92. Obach RS. The prediction of human clearance from hepatic microsomal metabolism data. *Curr Opin Drug Discov Dev* 2001;4:36–44.
93. Rowland M, Tozer TN. *Clinical Pharmacokinetics: Concepts and Applications*, 3rd ed. Philadelphia: Lippincott Williams & Wilkins; 1994.
94. Houston JB, Galetin A. Progress towards prediction of human pharmacokinetic parameters fron *in vitro* technologies. *Drug Metab Rev* 2003;35:393–415.
95. Shiran MR, Proctor NJ, Howgate EM, Rowland-Yeo K, Tucker GT, Rostami-Hodjegan A. Prediction of metabolic drug clearance in humans: *in vitro–in vivo* extrapolation vs allometric scaling. *Xenobiotica* 2006;36:567–580.
96. Bjornsson TD, Callaghan JT, Einolf HJ, Fischer V, Gan L, Grimm S, Kao J, King SP, Miwa G, Ni L, Kumar G, McLeod J, Obach RS, Roberts S, Roe A, Shah A, Snikeris F, Sullivan JT, Tweedie D, Vega JM, Walsh J, Wrighton SA. The conduct of *in vitro* and *in vivo* drug–drug interaction studies: a Pharmaceutical Research and Manufacturers of America (PhRMA) perspective. *Drug Metab Dispos* 2003;31:815–832.
97. Ito K, Brown HS, Houston JB. Database analyses for the prediction of *in vivo* drug–drug interactions from *in vitro* data. *Br J Clin Pharmacol* 2004;57:473–486.
98. Obach RS, Walsky RL, Venkatakrishnan K, Gaman EA, Houston JB, Tremaine LM. The utility of *in vitro* cytochrome P450 inhibition data in the prediction of drug–drug interactions. *J Pharmacol Exp Ther* 2006;316:336–348.
99. Hartung T, Bremer S, Casati S, Coecke S, Corvi R, Fortaner S, Gribaldo L, Halder M, Hoffmann S, Roi AJ, Prieto P, Sabbioni E, Scott L, Worth A, Zuang V. A modular approach to the ECVAM principles on test validity. *Altern Lab Anim* 2004;32: 467–472.

100. Aninat C, Piton A, Glaise D, Le Charpentier T, Langouet S, Morel F, Guguen-Guillouzo C, Guillouzo A. Expression of cytochromes P450, conjugating enzymes and nuclear receptors in human hepatoma HepaRG cells. *Drug Metab Dispos* 2006;34:75–83.
101. Le Vee M, Jigorel E, Glaise D, Gripon P, Guguen-Guillouzo C, Fardel O. Functional expression of sinusoidal and canalicular hepatic drug transporters in the differentiated human hepatoma HepaRG cell line. *Eur J Pharm Sci* 2006;28:109–117.

# UTILIZATION OF *IN VITRO* CYTOCHROME P450 INHIBITION DATA FOR PROJECTING CLINICAL DRUG–DRUG INTERACTIONS

JANE R. KENNY, DERMOT F. MCGINNITY, KEN GRIME, AND ROBERT J. RILEY

*AstraZeneca R&D Charnwood, Loughborough, United Kingdom*

**Contents**

1 Introduction
2 Cytochrome P450 and Its Role in Drug–Drug Interactions
  2.1 Clinical Consequences of Cytochrome P450 Mediated Drug–Drug Interactions
  2.2 Mechanisms of Cytochrome P450 Inhibition
3 Assessment of Cytochrome P450 Inhibition *In Vitro*
  3.1 Enzyme Systems
  3.2 Probe Substrate Selection
  3.3 Analysis Methods
4 Generation of *In Vitro* Cytochrome P450 Inhibition Data
  4.1 Reversible Inhibition ($IC_{50}$ and $K_i$)
  4.2 Mechanism-Based Inhibition ($K_I$, $k_{inact}$, and $IC_{50}$)
5 *In Vitro–In Vivo* Extrapolation of Cytochrome P450 Based Drug–Drug Interactions
  5.1 Approach for Reversible Inhibitors
  5.2 Approach for Mechanism-Based Inhibitors
  5.3 Key Concepts
  5.4 FDA Guidelines
6 Conclusion
References
Appendix A
Appendix B

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

## 1 INTRODUCTION

Adverse drug reactions (ADRs) are recognized as a significant cause of hospitalizations and up to 2.8% of hospital admissions may be a consequence of drug–drug interactions (DDIs) [1, 2]. The possibility of DDIs presents whenever two or more drugs are administered simultaneously; an ever more likely occurrence in an aging population often exposed to polypharmacy [3]. Moreover, it has been estimated that >90% of ADRs are also associated with dose-dependent pharmacokinetic/pharmacodynamic (PK/PD) events, which are one of the top five causes of death in the United States, resulting in over 100, 000 deaths each year [1, 2, 4]. While this is alarming, it is noteworthy that of the drugs associated with preventable drug-related hospital admissions, only four drug groups account for more than 50% of ADRs [5]. Indeed, drug concentration-dependent pharmacological reactions (classified as type A adverse reactions) account for approximately 75% of all ADRs and in principle should be preventable [6]. Such statistics highlight the need for improved mechanistic understanding and clear risk avoidance strategies, both during the drug discovery/development process and in the clinic.

DDIs may occur at the metabolic level with impact on drug clearance at the site of gastrointestinal absorption, on plasma or tissue binding, by inhibition or induction of hepatic drug metabolizing enzymes, and on carrier mediated transport (including renal and hepatic uptake and biliary efflux) [7]. All these interactions may induce changes in the PK profiles of the drugs involved, resulting in altered PD response or promotion of a toxicological effect not associated with the primary pharmacology [8, 9]. Several clinically relevant and severe DDI cases, such as those resulting in the withdrawal of terfenadine and cerivastatin [10], have increased the focus on metabolism-based interactions, particularly those involving cytochrome P450 (CYP) enzymes. Indeed, CYP inhibition is reported to account for as much as half of all reported DDIs [11]. These metabolic interactions can occur whenever the metabolism of one "victim" drug is inhibited by the presence of another "perpetrator" drug. The resultant decrease in clearance of the victim drug leads to increased systemic concentrations that are pronounced if the inhibited pathway is the major route of elimination. Several CYP mediated DDIs have resulted in severe clinical reactions and fatalities in extreme cases [10, 12], so real emphasis is placed on the prediction and hence avoidance of DDIs in drug discovery [7, 13–15]. Indeed, there has been substantial academic and industrial effort directed toward the accurate prediction of DDIs with the aim of reducing their impact in the clinic by designing out such interactions.

DDIs as a potential source or contributor to drug attrition, either directly or indirectly, has contributed to the industry focus on this area. In the late 1980s poor drug metabolism and pharmacokinetics (DMPK) properties and toxicity were predominantly responsible for the discontinued development of new chemical entities (NCEs) during late stage clinical trials [16, 17]. In response to this, and to increase competitive advantage (e.g., by avoiding product interaction labels), the pharmaceutical industry has increasingly aligned DMPK to the early drug discovery phase. The failure of NCEs at a late stage is a cost the pharmaceutical industry can ill afford; it is estimated that the total preapproval costs for a new drug are in excess of 800 million U.S. dollars [18]. The industry has responded to both the clinical and

economic factors by front-loading DDI screens early in the discovery phase to predict the risk of DDI for NCEs. High throughput automated assays to evaluate the impact of CYP inhibition are now common as early as the hit identification stage [19]. Increased importance is placed on evaluation of the risk for NCEs participating in DDIs, as a victim and/or perpetrator of such interactions. However, despite the focus that DDIs and potentially associated ADRs have received over the last two decades, it was recently reported that there has been little reduction in their clinical impact [6, 17], although with a drug development time in excess of ten years it is perhaps too early to identify the impact of such DDI screens on these clinical statistics.

In the following sections several examples of DDIs are introduced to demonstrate the clinical consequences of CYP inhibition. The kinetics of reversible and irreversible inhibition are presented along with discussion of *in vitro* methodology. Accurate *in vitro* kinetic data is an absolute prerequisite for *in vitro–in vivo* extrapolation (IVIVE), and the current literature models for predicting clinical inhibition are presented. Overall, this chapter aims to summarize the current knowledge base and the complexities of CYP inhibition and to provide a guide for embarking on prediction of clinical DDIs with particular emphasis on the drug discovery setting.

## 2 CYTOCHROME P450 AND ITS ROLE IN DRUG–DRUG INTERACTIONS

The CYP family of enzymes are the primary facilitators of oxidative biotransformation in the liver and are responsible for the metabolism, at least in part, of the majority of human drugs [20]. These heme-containing enzymes are primarily situated in the lipid bilayer of the endoplasmic reticulum of hepatocytes, but are also located in extrahepatic tissues, including the gastrointestinal tract (GIT), kidney, and lung [21]. Fifty-seven different human CYP enzymes have been identified [22] and of these, 23 human isoforms are from the CYP1, CYP2, and CYP3 families that are recognized as important in xenobiotic metabolism [23]. However, it has been estimated that approximately 90% of human drug oxidations can be attributed to just five CYPs, namely, CYP1A2, CYP2C9, CYP2C19, CYP2D6, and CYP3A4/5 [24, 25]. CYP3A4 is the most abundant of these and can constitute up to 50% of total hepatic CYP content [26]. This enzyme is also the most prolific, with approximately 60% of all oxidative biotransformations being attributed to CYP3A4 [27]. The CYP family metabolize a very broad range of substrates, from endogenous steroids and vitamins to small drug molecules of all charge types [28]. For example, CYP3A4 metabolizes large lipophilic molecules including testosterone, erythromycin, and cyclosporine as well as smaller molecules like midazolam and nifedipine. CYP2C9 favors medium sized acidic or neutral molecules such as diclofenac and tienilic acid; CYP2D6 favors basic, lipophilic molecules such as dextromethorphan and metoprolol, while CYP1A2 has affinity for neutral planar molecules such as theophylline and phenacetin [28]. Since CYPs are the major enzyme family in xenobiotic/drug metabolism, they retain a central role in the mediation of clinically relevant DDIs.

### 2.1 Clinical Consequences of Cytochrome P450 Mediated Drug–Drug Interactions

There is a wealth of literature around CYP mediated DDIs. The most documented DDIs resulting in severe clinical adverse reactions are listed in Table 1. In the following section examples of drugs as either victims or perpetrators of DDIs will be presented with specific cases exemplifying (1) reversible and irreversible inhibition at the major CYPs and (2) the role of intestinal metabolism and enzyme polymorphisms within the patient population.

The antihistamine terfenadine, a victim of DDI, was withdrawn because of patient fatality following coadministration with ketoconazole [10]. Competitive inhibition of CYP3A4-dependent terfenadine methyl-hydroxylation by ketoconazole resulted in up to a 40-fold increase in the area under the plasma concentration–time curve (*AUC*) of terfenadine. This increased concentration manifested itself clinically as QT prolongation and torsades de pointes [37]. Other victims of DDI are simvastatin and lovastatin. Both are 3-hydroxy-3-methylglutaryl coenzyme A (HMG-CoA) reductase inhibitors, which although exceedingly well tolerated, incur a small risk of myopathy or potentially fatal rhabdomyolysis, particularly when coadministered with medications that increase their systemic exposure [44]. These statins both have a substantial CYP3A4 component to their metabolism, and large (hepatic and intestinal) first-pass metabolism makes them particularly susceptible to interactions via CYP3A4 inhibition. Intestinal metabolism may be effectively abolished in the presence of a potent oral CYP3A4 inhibitor, resulting in an increase in plasma concentration of up to threefold for drugs such as cyclosporine, midazolam, lovastatin, and simvastatin [44–46]. This was elegantly confirmed in clinical studies where midazolam was administered either via the oral or intravenous route alongside the CYP3A4 inhibitors itraconazole and fluconazole [46]. Following an intravenous dose the *AUC* of midazolam was increased 3.5-fold in the presence of itraconazole; however, the *AUC* increased sevenfold when midazolam was administered orally [46]. Clearly, for drugs with a high hepatic extraction, the impact of metabolic inhibition will give rise to large changes in oral exposure, compounded by any effect of inhibition on intestinal metabolism.

Another statin, cerivastatin, was anticipated to have less potential for CYP3A4 mediated DDIs than the other statins due to the additional role of CYP2C8 in its metabolism [47]. Prior to launch it was thought that cerivastatin would provide a safer alternative to simvastatin and lovastatin due to its alternative metabolic pathway. Indeed, interaction studies with erythromycin and itraconazole, two potent inhibitors of CYP3A4, have shown little impact on cerivastatin plasma levels [48–50]. However, coadministration with gemfibrozil increases the *AUC* of cerivastatin about sixfold [33]. This is assumed to be due to inhibition of CYP2C8 by gemfibrozil and its glucuronide metabolite. Interestingly, the glucuronide metabolite demonstrates a 20-fold increase in inhibitory potency at CYP2C8 over its parent ($IC_{50}$ of 80 μM and 4 μM for gemfibrozil and its glucuronide metabolite, respectively [51]) and is actively concentrated in the liver [51, 52]. It is also worth noting that gemfibrozil also inhibits the OATP1B1-mediated hepatic uptake of statins and that this mechanism could well have contributed to this and indeed other gemfibrozil–statin interactions [52, 53]. During cerivastatin monotherapy, the incidence rate of rhabdomyolysis was 10–100 times higher than with the other statins and gemfibrozil

**TABLE 1 Examples of Marketed Drugs Involved in Clinically Significant Drug–Drug Interactions**

| Predominant CYP Isoform | Victim Drug | Perpetrator Drug | Maximum Clinical AUC Increase of Victim Drug | Clinical Consequences | References |
|---|---|---|---|---|---|
| CYP1A2 | Theophylline | Ciprofloxacin | Variable | Theophylline toxicity | 29 |
| | Tizanidine | Ciprofloxacin | 10-fold | Excess hypotensive and sedative effects | 30 |
| | Tacrine | Fluvoxamine | ~sixfold | Hepatotoxicity | 31, 32 |
| CYP2C8 | Cerivastatin[a] | Gemfibrozil | sixfold | Rhabdomylosis | 33 |
| CYP2C9 | Tolbutamide | Sulfaphenazole | 5.3-fold | Hypoglycemia | 34 |
| | Warfarin | Fluconazole | 50% decrease in clearance | Enhanced hypoprothrombinemic effect | 35 |
| CYP2D6 | Desipramine | Fluoxetine | ~sixfold | | 31 |
| | Metoprolol | Mibefradil[a] | 4–5-fold | Effect on cardioselectivity | 36 |
| CYP3A4 | Terfenadine[a] | Ketoconazole | ~40-fold | Torsades des pointes | 2, 10, 37, 38 |
| | Lovastatin | Itraconazole | ~20-fold | Rhabdomylosis | 39, 40 |
| | Simvastatin | Mibefradil[a] | 5–10-fold | Rhabdomylosis | 41 |
| | Midazolam | Erythromycin | threefold | Excess sedation | 42 |
| | Cisapride[a] | Erythromycin | Variable | QT prolongation | 12, 43 |

[a]Withdrawn from the market between 1990 and 2001.

greatly increased this risk; the number of rhabdomyolysis cases reported to the U.S. Food and Drug Administration with the gemfibrozil–cerivastatin combination was 533 [53]. As a result, cerivastatin was voluntarily withdrawn from the market in 2003.

Adverse events can be manifest without particularly large changes in the *AUC* of the victim drug. In epileptic patients stabilized on carbamazepine therapy, addition of the antidepressant agent viloxazine has resulted in side effects of carbamazepine intoxication (fatigue, dizziness, and ataxia). This is associated with an *AUC* elevation of carbamazepine (approximately twofold) and carbamazepine 10,11-epoxide (approximately 1.3-fold) [54]. The most likely mechanism for this interaction is CYP3A4 inhibition by viloxazine [55]. This example shows how a modest elevation can be detrimental if the therapeutic window of the victim drug is narrow; in this case resulting in an exaggerated pharmacological effect.

Mibefradil, a novel calcium antagonist for the treatment of hypertension and angina and a perpetrator of DDI, was voluntarily withdrawn from the market in 1997, less than a year after its launch [36]. Several severe adverse reactions were observed during mibefradil therapy in combination with various drugs including simvastatin and lovastatin [36, 41]. During its development, mibefradil had been identified as having inhibitory effects on CYP3A4 and CYP2D6, consequently coadministration was contraindicated with terfenadine and metoprolol (a CYP2D6 substrate) and the labeling carried a list of drugs that might require dose adjustment. When mibefradil was withdrawn, an additional complication arose; life-threatening reactions occurred if therapy was replaced with β-blockers or calcium channel antagonists within 5 days of discontinuing mibefradil therapy [36]. This was attributed to the long elimination half-life of mibefradil, but subsequently, it has been observed that mibefradil is a time-dependent inhibitor of CYP3A4 [56, 57] and also inhibits P-gp *in vitro* ($K_i$ of 1.6 μM [58]). The time-dependent nature of its CYP3A4 inhibition could explain the long "wash-out" period of mibefradil mediated inhibition. There are suggestions in the literature that some of the severe interactions observed with mibefradil could have been predicted from the *in vitro* and *in vivo* data available at the time [36, 45, 59].

The extent of DDIs can vary substantially within the population due to variability in enzyme expression. This variation is observed in the ciprofloxacin–theophylline interaction and has implications for the severity of the clinical outcome. The inhibitory effect of ciprofloxacin at CYP1A2 can result in varying changes in theophylline levels across a patient population. In sensitive patients, theophylline clearance can decrease significantly, increasing the risk of theophylline related toxicity [29, 60]. In another case, ciprofloxacin inhibits the CYP1A2 mediated metabolism of tizanidine, resulting in AUC increases of up to 10-fold, dangerously escalating its hypotensive and sedative effects [30]. CYP1A2 is the primary clearance mechanism for both theophylline and tizanidine [61, 62], and there is considerable interindividual variability in the *in vivo* metabolic activity of CYP1A2 (demonstrated by caffeine clearance [63]); therefore, inhibition results in significant changes in plasma levels in sensitive patients only.

As is the case for all polymorphic CYPs, the CYP2D6 related DDIs are complicated by the genetic polymorphisms observed for this enzyme (between 5% and 10% of the Caucasian population are CYP2D6 poor metabolizers). Investigation of potential DDIs observed via CYP2D6 inhibition may be confounded by the pres-

ence of individuals with the poor metabolizer phenotype within the study group [64]. It is evident that poor metabolizers are unlikely to suffer DDIs through CYP2D6 inhibition and therefore will not display the same magnitude of interaction predicted from *in vitro* studies [6]. CYP2D6 is inhibited by antipsychotic drugs and the propensity for interactions is increased for this patient population as psychopharmacologic medications are often taken for long periods of time and commonly coprescribed with other medications [65]. For example, the CYP2D6 inhibitor fluoxetine has been demonstrated to substantially increase the plasma levels of desipramine (between four- and sixfold) [31].

The clinical interactions presented here highlight the need to assess an NCE's involvement in potential DDIs from *both* perspectives of inhibition, either as a victim drug where metabolism is altered by a coadministered inhibitor or as the perpetrator of DDI. The cases also highlight the importance of fully defining the elimination and metabolic profile of an NCE: reaction phenotyping can lend greater understanding to potential interactions as can an understanding of the target patient population and likely coadministered drugs.

## 2.2 Mechanisms of Cytochrome P450 Inhibition

There are different mechanisms through which CYP inhibition may occur, all of which are a function of the metabolic capacity of the enzyme. In nearly all cases of clinically relevant *reversible* inhibition, the perpetrator is a substrate competing for the same CYP mediated metabolism. Simple Michaelis–Menten kinetics often satisfactorily describes CYP metabolism *in vitro*, provided that the enzyme and substrate are in thermodynamic equilibrium.

***Enzyme Kinetics*** In the reaction scheme shown in Fig. 1, a steady-state equilibrium is assumed to be reached (the ES complex rapidly reaches a constant value such that $d[\mathrm{ES}]/dt = 0$). This assumption is key to the derivation of the Michaelis–Menten equation, which underpins not only prediction of metabolic clearance from *in vitro* data but also prediction of DDIs (since the prediction of DDI magnitude from *in vitro* data always starts with a modification of the Michaelis–Menten equation). The reason that the investigator needs to be aware of this is that experimental conditions are important in making this assumption. For $d[\mathrm{ES}]/dt = 0$, the reverse reaction rate should be negligible and the formation of the [ES] complex should not significantly alter the substrate concentration [S]. In other words, the substrate concentration should greatly exceed the enzyme concentration and the substrate concentration should not be significantly depleted during the incubation (less than 20% metabolism). To this end, it is noteworthy that in a typical human liver microsome (HLM) preparation, CYP3A4 concentration may be as much as 0.25 μM when the incubation contains 1 mg/mL HLM.

Equations 1 and 2 are the Michaelis–Menten equation (1) and its rearrangement (2), where $v$ is the observed rate of metabolite formation, $V_{\max}$ is the maximum rate of this metabolite formation, $K_m$ is the Michaelis–Menten constant representing the enzyme:substrate affinity and is the concentration that supports half the maximum rate of metabolite formation, and [S] is the substrate concentration at the enzyme active site. The derivation of these equations is shown in Appendix A.

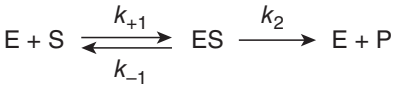


**FIGURE 1** Schematic representing Michaelis–Menten kinetics, where E is enzyme, S is substrate, $k_{+1}$ is the association rate constant describing ES formation from E + S, $k_{-1}$ is the dissociation rate constant from ES to E + S, ES is substrate-bound enzyme, P is the product of enzyme, and $k_2$ is the dissociation rate constant from ES to E + P.

$$v = \frac{V_{\max} \cdot [S]}{K_m + [S]} \tag{1}$$

$$= \frac{V_{\max}}{1 + K_m/[S]} \tag{2}$$

Only simple one-site binding can be accurately described by hyperbolic Michaelis–Menten kinetics but there is substantial literature evidence that atypical kinetics and multisite binding exist for the CYP enzymes. This is most clear for CYP3A4 [66] but is also observed for CYP2C9 [67, 68]. Positive or negative cooperative effects, where a bound substrate molecule influences the binding of a second molecule, have implications in terms of probe substrate selection and make full kinetic characterizations complex. It must be noted, however, that while cooperativity is clearly demonstrable *in vitro*, there is little evidence of its influence *in vivo* [69, 70]. Only simple one-site models of inhibition will be discussed here and detailed explanation of atypical (or sigmoidal) kinetics in inhibition can be found elsewhere in the scientific literature [14, 66, 71–73].

***Reversible Inhibition*** Having explained the basis of simple CYP kinetics, the details of inhibition can be described. Reversible CYP inhibition can be categorized into three main forms: competitive, noncompetitive, and uncompetitive. Figure 2 describes each type of inhibition and the rate equations associated with them are presented below. In competitive inhibition the substrate and inhibitor compete for the same binding site within the enzyme; inhibition therefore decreases as the enzyme becomes saturated with substrate. Competitive inhibitors interfere with substrate binding so as to raise $K_m$ by a factor of $[I]/K_i$ without affecting $V_{\max}$ (i.e., the affinity of the substrate for the enzyme, $K_m$, is decreased by a factor dependent in magnitude on the ratio of inhibitor concentration and the affinity of the inhibitor for the enzyme, $K_i$ (Eq. 3)):

$$v = \frac{V_{\max} \cdot [S]}{K_m(1 + [I]/K_i) + [S]} \tag{3}$$

It is common to relate the potency of competitive inhibition to $IC_{50}$ (the concentration of inhibitor for which the reaction rate is suppressed by 50%). Equation 4 is derived from Eq. 3 when $v$ becomes $v/2$ and I becomes $IC_{50}$.

$$IC_{50} = K_i(1 + [S]/K_m) \tag{4}$$

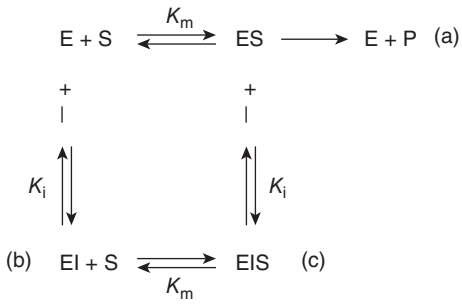


**FIGURE 2** Schematic of competitive inhibition, representing Michaelis–Menten kinetics. Pathway (a), competitive inhibition; pathway (b), uncompetitive inhibition; pathway (c), noncompetitive inhibition. E is enzyme, S is substrate, ES is substrate-bound enzyme, P is the product of enzyme and substrate, I is inhibitor, EI is inhibitor-bound enzyme, EIS is inhibitor- and substrate-bound enzyme, $K_m$ is the Michaelis–Menten constant, and $K_i$ is the inhibition constant. Note that $K_m$ can be described as $(k_2 + k_{-1}/k_{+1})$—see Appendix A. However, if it is assumed that rapid formation of the ES complex occurs when the enzyme and substrate are mixed; $k_{+1} >> k_2$ and $K_m$ can be described as $k_{-1}/k_{+1}$.

In noncompetitive inhibition the inhibitor binds at a site distinct from the substrate binding site and can interact with either the substrate-free or the substrate-bound form of the enzyme. It is assumed that the inhibitor binds to both states of the enzyme with the same affinity and so the degree of inhibition is independent of substrate concentration. The binding affinity ($K_m$) is unaffected but the inhibitor acts to remove active enzyme and consequently the maximal reaction velocity ($V_{max}$) is decreased. The metabolic rate of noncompetitive inhibition is described by

$$v = \frac{V_{max}/(1+[I]/K_i)\cdot[S]}{K_m + [S]} \tag{5}$$

The substrate-independent nature of inhibition is further demonstrated by the $IC_{50}$, which is equivalent to the inhibitor constant $K_i(IC_{50} = K_i)$, derived from Eq. 5 when $v$ becomes $v/2$ and [I] becomes $IC_{50}$.

In uncompetitive inhibition (which is extremely rare for CYP metabolism [71]), only the substrate-bound form of the enzyme is open to inhibitor binding; the inhibitor cannot bind to free enzyme and so the degree of inhibition will increase with substrate concentration. The metabolic rate in the presence of an uncompetitive inhibitor can be described by Eq. 6 and the corresponding $IC_{50}$ by Eq. 7:

$$v = \frac{V_{max}/(1+[I]/K_i)\cdot[S]}{K_m(1+[I]/K_i)+[S]} \tag{6}$$

$$IC_{50} = K_i(1 + K_m/[S]) \tag{7}$$

***Mechanism-Based Inhibition*** Irreversible inhibition or mechanism-based inactivation is an unusual occurrence with most enzymes. It is, however, observed in reactions catalyzed by CYP in somewhat higher frequency [74]. In contrast with

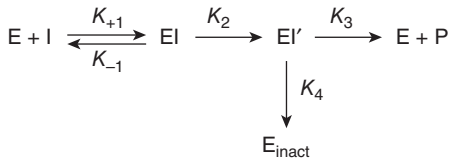


**FIGURE 3** Schematic of mechanism-based inhibition, where E is enzyme, I is inhibitor, EI is inhibitor-bound enzyme, EI′ is the initial complex formed between metabolic intermediate and enzyme, P is the product of enzyme and inhibitor complex that escapes reactive binding in the active site, $E_{inactive}$ is the inactive enzyme with irreversibly bound metabolic intermediate, $k_{+1}$ is the association rate constant for E + I forming EI, $k_2$ is the association rate constant for EI forming EI′, $k_3$ is the dissociation rate constant for EI′ to E + P, and $k_4$ is the rate constant from EI′ to $E_{inact}$.

reversible inhibition, the effects of mechanism-based inhibition (MBI) are more profound after multiple dosing and the recovery period, typically several days, is independent of continued exposure to the drug [75, 76]. This MBI transpires when a CYP is quasi-irreversibly or irreversibly inhibited by reactive intermediates formed during the metabolic process. The inhibitory intermediates may form covalently bound adducts to the CYP heme or apoprotein; alternatively, they may form a metabolic inhibitory complex (MIC) in the active site. This MBI is characterized by NADPH-, concentration-, and time-dependent inactivation. Additionally, the inactivation rate is diminished in the presence of a competing substrate, enzyme activity is not restored by dialysis or filtration (as inhibitor is covalently bound), and *de novo* synthesis of protein is required to recover metabolic activity [77]. Figure 3 shows a schematic representation of this inhibition [78] and the inhibition for the initial rate of enzyme deactivation can be described as follows.

From Fig. 3 it can be seen that the change in concentration over time of inhibitor-bound enzyme, [EI], the initial complex formed between metabolic intermediate and enzyme, [EI′], and the inactive enzyme with irreversibly bound metabolic intermediate, [$E_{inact}$], respectively can be expressed as

$$d[\mathrm{EI}]/dt = k_{+1}[\mathrm{I}]\cdot[\mathrm{E}] - (k_{-1} + k_2)[\mathrm{EI}] \tag{8}$$

$$d[\mathrm{EI}']/dt = k_2[\mathrm{EI}] - (k_3 + k_4)[\mathrm{EI}'] \tag{9}$$

$$d[\mathrm{E}_{\mathrm{inact}}]/dt = k_4[\mathrm{EI}'] \tag{10}$$

where [E] is the concentration of the unbound active enzyme and [I] is the concentration of the mechanism-based inhibitor.

The parameters $k_{inact}$ (the maximal inactivation rate constant) and $K_I$ (the inhibitor concentration that supports half the maximal rate of inactivation) are described by Eqs. 11 and 12, respectively, using the association constants shown in Fig. 3.

$$k_{\mathrm{inact}} = k_2 k_4/(k_2 + k_3 + k_4) \tag{11}$$

$$K_{\mathrm{I}} = \frac{(k_3 + k_4)(k_{-1} + k_2)}{(k_2 + k_3 + k_4)k_1} \tag{12}$$

The apparent inactivation rate constant of the enzyme, $k_{obs}$, is described by

$$k_{obs} = \frac{k_{inact}[I]}{[I] + K_I} \tag{13}$$

The derivation of these equations is included in Appendix B to show the mathematical assumptions that must be made to allow such parameters to be defined experimentally.

The partition ratio ($r$) is a useful additional parameter to $k_{inact}$ and $K_I$. It is defined as the number of moles of reactive metabolite escaping the enzyme without inactivating it (product) per number of moles of reactive metabolite inactivating the enzyme (trapped product) ($r = k_3/k_4$; Fig. 3). As the partition ratio approaches 0, inhibition is more efficient; for every cycle of the enzyme and substrate that results in product, a cycle also results in inactivated enzyme. For a potent inhibitor, the reactive species is less likely to escape the active site without adduct formation. While the partition ratio may be helpful in ranking compounds in order of inhibitory efficiency (as is the $k_{inact}/K_I$ ratio), any use beyond this is limited.

## 3 ASSESSMENT OF CYTOCHROME P450 INHIBITION *IN VITRO*

Over the last decade there have been tremendous advances in the understanding of both the theoretical and technical aspects of generating *in vitro* kinetic data to engender predictions of DDIs [7, 11, 13, 21, 71, 79, 80]. The advantages of using *in vitro* assays to study inhibition include allowing inexpensive and rapid assessment of inhibition potential and consequent evaluation of DDI risk, facilitating detailed kinetic analysis of drugs known to cause DDIs in the clinic, and helping to further develop our understanding of these interactions and ultimately striving toward the avoidance of DDIs for NCEs. However, despite the undoubted usefulness of *in vitro* models, there are limitations, primarily due to the fact that, to date, there is no consensus toward a model that quantitatively predicts interactions in humans from *in vitro* data. Indeed, there is no single method for generating *in vitro* data. Despite the lack of consensus, there are some fundamentally common elements to the existing methodologies and these form the basis of the remainder of this chapter. Prior to modeling data, accurate *in vitro* assessments must be made. There are many parameters to consider when embarking on inhibition studies to generate good *in vitro* data, the key elements of which are discussed here.

### 3.1 Enzyme Systems

Historically, the majority of inhibition studies have been conducted using subcellular fractions of liver. Clear species differences exist in hepatic enzyme content and function [81]. In most cases a homolog of the human enzyme is present in the preclinical species. For example, CYP3A1 and CYP3A2 are the major 3A isoforms in the rat and share substrate crossover to human CYP3A4 and CYP3A5. Likewise CYP2C11 in rodent liver is homologous to CYP2C9 in humans. However, the species-specific isoforms of CYP1A, CYP2C, CYP2D, and CYP3A do show appreciable interspecies differences in terms of catalytic activity [81]. While laboratory

species can be a useful tool for investigating pharmacokinetic species differences and for providing confidence in IVIVE, clinical DDI potential must be investigated *in vitro* using human derived tissues and/or enzymes because of such species differences.

Human liver microsomes (HLMs) are prepared from liver homogenates by differential ultracentrifugation, first removing the cytosolic (S9) fraction, then with further ultracentrifugation collecting the subcellular fraction, which includes membrane-bound enzymes [82]. The utility of HLMs is clear; they can be isolated from a relatively small amount of human liver, can be stored almost indefinitely at –80 °C, contain functional CYPs (and other enzymes such as UDP-glucuronosyltransferases and flavin monooxygenases), and can be used as a single donor or pooled to provide a representation of enzyme expression in the population. However, due to the presence of multiple CYP isoforms with overlapping substrate selectivity, kinetic investigations must be thorough and carefully planned as inhibitors may be substrates for more than one CYP isoform. Consequently, for CYP inhibition studies using HLMs, the use of selective probe reactions is paramount and often relies on careful adjustment of substrate concentration to ensure selectivity.

In the last decade, the use of recombinant human CYP (rCYP) proteins, expressed in yeast, baculovirus-infected insect cells, or *Escherichia coli*, has become widespread and indeed now forms the backbone of screening efforts in the pharmaceutical industry. There are considerable advantages to using a single rCYP enzyme in an assay as it allows manipulation of substrates and inhibitors without the complications of enzyme selectivity. A notable example of this is CYP3A5, which has broad substrate crossover with CYP3A4, making it difficult to characterize reactions that are specific to or of equal affinity at CYP3A5 with respect to CYP3A4 [83]. A sometimes-perceived disadvantage of rCYP is the increased enzyme activity per unit of membrane protein compared to that of HLM. However, the high CYP expression levels give rise to low incubational protein content and low non-specific substrate and inhibitor binding, resulting in more accurate estimations of unbound $K_{+1}$ values. Finally, the utility of rCYP in reaction phenotyping and inhibition studies has been well validated [84–86] and the versatility of rCYPs lends them to enhanced throughput screening.

Isolated human hepatocytes are an alternative *in vitro* tool and can be used to add refinement to data generated using subcellular fractions, as the impact of drug transporter activity on $K_i$, accounting for intracellular free drug concentrations, can be investigated. As whole cells, hepatocytes express not only all membrane-bound enzymes but also hepatic cytosolic enzymes; moreover, they contain the necessary cofactors. Under the correct conditions, human hepatocytes retain some functionality for many days, express the transcriptional machinery to regulate enzymes, and contain functional proteins for active drug transport across membranes [87]. The utility of hepatocytes for investigation of metabolic stability [88–90] and DDIs mediated by CYP induction [87, 91, 92] is well established. More recently, hepatocytes have been used to describe the impact of reversible [68, 93, 94] and time-dependent CYP inhibition *in vitro* [95, 96]. Primary hepatocytes in culture provide the closest *in vitro* model to human liver and may offer advantages when predicting clinical DDI. Metabolites of one pathway may lead to inhibition of another; this phenomenon is indiscernible using single rCYPs and exemplifies the usefulness of hepatocytes [97].

There is little published comparative data of inhibition constants generated in rCYPs and human hepatocytes. Data generated in this laboratory demonstrates $IC_{50,apparent}$ values generated in human hepatocytes, using probe substrates specific to the CYP isoform, were systematically higher than those determined using the respective rCYP [68]. This was predominantly the result of greater nonspecific binding in hepatocytes compared to rCYPs, as for the majority of compounds tested there was a good concordance between the respective $IC_{50,unbound}$ values (Fig. 4a).

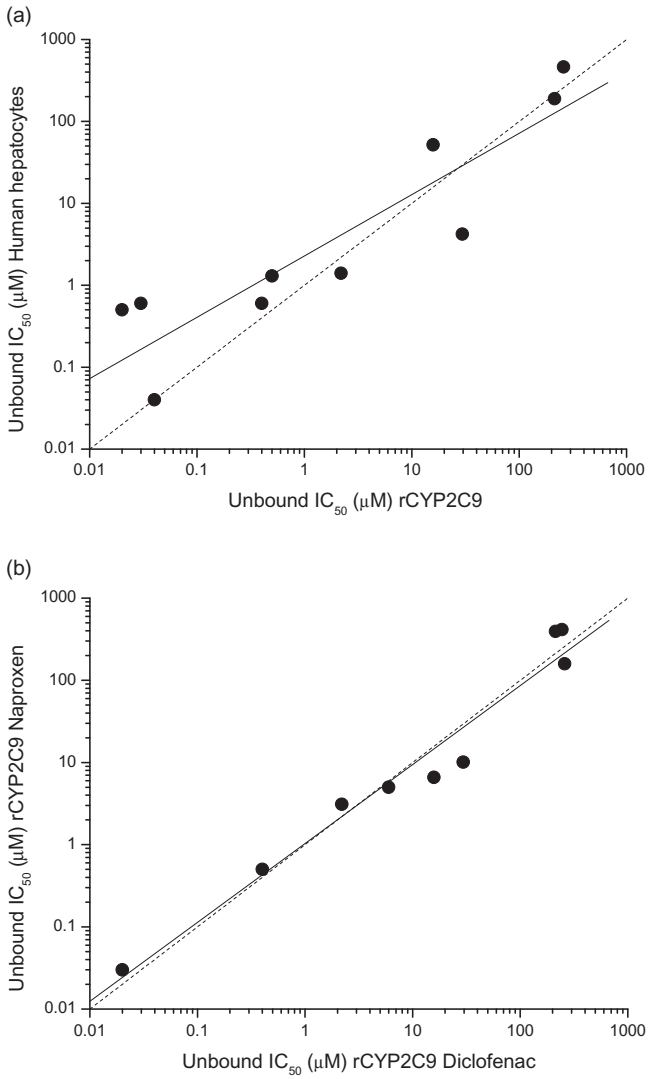


**FIGURE 4** (a) Relationship between $IC_{50,unbound}$ for CYP2C9 in human hepatocytes and rCYPs. The dotted line is unity. The solid line indicates linear regression of the data ($r^2 = 0.88$, $p < 0.0001$) (b) $IC_{50,unbound}$ comparisons using naproxen and diclofenac as substrates for rCYP2C9. The dotted line is unity. The solid line indicates linear regression of the data ($r^2 = 0.77$, $p < 0.002$). (Part (b) from Ref. 68.)

It can be speculated that the impact of hepatic transporters may explain some discrepancies in $IC_{50,unbound}$ between rCYP and hepatocytes as compound can be actively effluxed from or taken up into the hepatocyte, altering the concentration of drug exposed to the enzyme [97]. Additionally, the impact of time-dependent CYP inhibition in primary cultures of human hepatocytes has been evaluated, again using a cocktail of CYP substrates, which allows the maximum amount of data to be extracted from this limited resource [95]. $K_I$ and $k_{inact}$ values were estimated from nonlinear regression analysis and compared with values obtained from rCYPs and HLMs (Table 2). The values generated are also comparable to several literature reports of $K_I$ and $k_{inact}$ estimates using rCYP and HLMs and the results from our laboratory using rCYP, HLMs, and human hepatocytes were consistent with these

**TABLE 2 Kinetic Parameters of Time-Dependent Inhibitors in Different *In Vitro* Matrices**

| Compound | CYP Isoform | Enzyme Source | $K_I$ (μM) | $k_{inact}$ ($min^{-1}$) | Reference |
|---|---|---|---|---|---|
| Tienilic acid | CYP2C9 | rCYP2C9 | 2 | 0.19 | 95 |
| | | rCYP2C9 (2C10) | 4 | 0.20 | 99 |
| | | Human hepatocytes | 2 | 0.05 | 95 |
| AZ1 | CYP2C9 | rCYP2C9 | 30 | 0.02 | 95 |
| | | Human hepatocytes | 19 | 0.02 | 95 |
| Fluoxetine | CYP2C19 | rCYP2C19 | 0.4 | 0.5 | 95 |
| | | HLM | 8[a] (0.8) | 0.03 | 95 |
| | | Human hepatocytes | 0.2 | 0.04 | 95 |
| | CYP3A4 | rCYP3A4 | 2 | 0.03 | 95 |
| | | HLM | 5[a] (0.5) | 0.01 | 95 |
| | | HLM | 5[a] | 0.02 | 100 |
| | | Human hepatocytes | 1 | 0.01 | 95 |
| Erythromycin | CYP3A4 | rCYP3A4 | 9 | 0.12 | 98 |
| | | rCYP3A4 | 5 | 0.12 | 101 |
| | | HLM | 16[a] | 0.07 | 102 |
| | | HLM | 82[a] | 0.07 | 103 |
| | | HLM | 13[a] | 0.02 | 104 |
| | | HLM | 10[a] | 0.08 | 105 |
| | | HLM | 11[a] | 0.05 | 106 |
| | | HLM | 15[a] | 0.07 | 96 |
| | | Human hepatocytes | 11 | 0.07 | 95 |
| Troleandomycin | CYP3A4 | rCYP3A4 | 0.3 | 0.12 | 98 |
| | | rCYP3A4 | 0.2 | 0.15 | 101 |
| | | HLMs | 2[a] | 0.03 | 96 |
| | | Human hepatocytes | 0.4 | 0.05 | 95 |

[a]Apparent $K_I$ values. All other values are *unbound* $K_I$ estimates.

*Source*: From Ref. 95.

reports. Indeed, the CYP3A4-dependent $K_I$ and $k_{inact}$ values in human hepatocytes are comparable to values determined in rCYP3A4 and HLMs (Table 2) when corrected for nonspecific binding in the incubation.

Finally, an element of interlaboratory variability can be expected for different tissue preparations. However, an interlaboratory comparison found the rank order of five HLM preparations was conserved across five laboratories, all using different methodologies, despite differences in absolute values (based on protein content, CYP content, and activity) [107]. Irrespective of the enzyme system used, there are some common factors that must be considered for all *in vitro* inhibition incubations. The assay must be designed such that linear initial rate conditions for the reaction under investigation are assured. To make certain of this, preliminary experiments will be required to ensure product formation is linear with respect to both time and enzyme (protein) concentration [72]. Additionally, protein concentration should be kept as low as reasonably possible to reduce the impact of nonspecific binding [108]. As discussed with respect to hepatocytes and rCYPs, the fraction unbound in the incubation ($f_{uinc}$) can have a pronounced effect on the apparent $K_i$, $K_I$, and $IC_{50}$ determined (Fig. 5). This is especially the case for lipophilic inhibitors such as ketoconazole, for which the apparent $IC_{50}$ can vary between 0.005 and 0.3 μM [109]. It is important that such drug binding terms be incorporated by default when making IVIVE to improve the accuracy of *in vitro* data [88], and the importance of considering the fraction unbound in an *in vitro* incubation ($f_{uinc}$) when relating *in vitro* and *in vivo* data was proposed many years ago by Gillette (1963). The term itself can readily be determined and indeed predicted with accuracy from physiochemical properties [88, 110, 111]. Another important factor is the amount of solvent vehicle, which must be minimized ideally to 1% (or less) of the final incubation volume, as activation or inhibition of CYP activity can be manifest through solvent effects alone (Fig. 6). Acetonitrile appears the most favored solvent, having minimal inhibitory effects, despite it being able to activate CYP1A2; although DMSO is commonly used to ensure adequate NCE solubility, it can result in inhibition of metabolism, especially for basic CYP3A4 substrates [88, 112]. Solvent effects

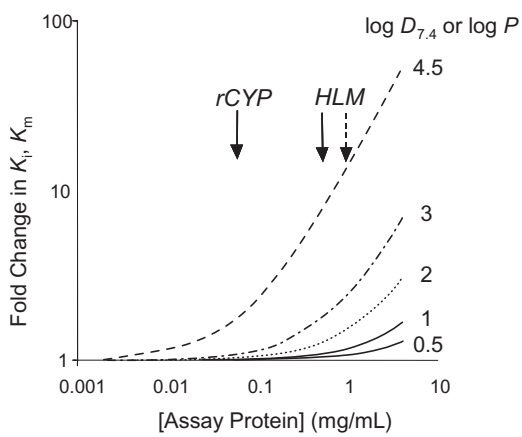


**FIGURE 5** Simulated effects of microsomal or hepatocyte concentration on $K_i$ or $K_m$ as a function of log $P$ or log $D_{7.4}$-dependent nonspecific binding. (From Ref. 88.)

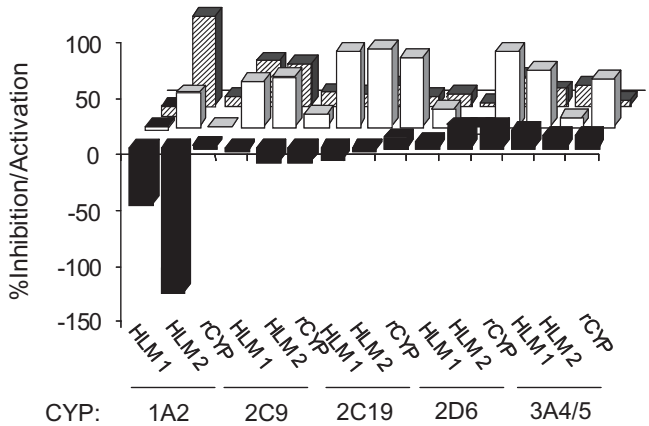


**FIGURE 6** Overview of potential effects of commonly used organic solvents (1%, v/v) on CYP activities with recombinant proteins and HLMs (solid bars = acetonitrile, open bars = DMSO, hatched bars = methanol). (From Ref. 112.)

can also be selective for probe and enzyme source [113], so it is advisable to check the effects of an alternative solvent for key studies.

### 3.2 Probe Substrate Selection

An important component of *in vitro* inhibition study design is the selection of appropriate probe substrates. Nonselective probes in HLM or human hepatocyte assays can result in hybrid inhibition measurements where multiple enzymes (with varying sensitivity to the inhibitor) can contribute to the reaction under investigation [14]. To avoid such complications, a selective marker substrate is required for each CYP isoform being assessed. These probes should meet the following criteria: possess high affinity for the specific enzyme; be metabolized to a specific product metabolite; provide the ability to detect the product separately from the substrate (using the analytical system of choice); and, finally, the turnover of the probe substrate should be optimized in each enzyme system and be linear with respect to time resulting in less than 20% loss of parent compound to maintain constant reaction velocity throughout the experiment [106, 114]. Industry representatives and regulatory bodies have made recommendations for preferred probe substrates for each CYP isoform (Table 3), many of which have the additional benefit of being clinically relevant [117, 120]. These substrates have been rigorously validated for human CYPs [115] and the usefulness and caveats associated with them are well documented [14].

An advantage to the availability of highly selective probes is that, coupled with the improved analytical sensitivity afforded by advances in detection systems such as LC/MS-MS, a cassette of substrates can be used to evaluate the inhibitory potential of a compound upon multiple CYP isoforms in a single incubation. The usefulness of cassette assays derives from increasing the throughput of inhibition screening, although they do require careful optimization and validation to ensure selectivity is maintained [116, 118, 121].

**TABLE 3 Specific Probe Substrate Reactions for Individual CYP Enzymes**

| CYP | Substrate | Reaction | Analytical Method | Comments[a] | References |
|---|---|---|---|---|---|
| CYP1A2 | Phenacetin[b] | O-deethylation | LC/MS-MS | High specificity at <100 μM ($K_m \approx 50$ μM) | 115, 116 |
| | Ethoxyresorufin (EROD) | O-deethylation | Fluorescence | $K_m \approx 0.2$ μM; not clinically relevant | 117 |
| | Caffeine | N3-demethylation | UV-HPL | Low affinity $K_m \approx 600$ μM | 117 |
| CYP2A6 | Coumarin | 7-Hydroxylation | LC/MS-MS Fluorescence | High specificity $K_m \approx 0.5$ μM | 115, 117 |
| CYP2B6 | Bupropion[b] | Hydroxylation | LC/MS-MS | High specificity $K_m \approx 90$ μM | 115, 118 |
| | *S*-Mephenytoin | N-demethylation | LC/MS-MS | Not CYP2B6 specific $K_m \approx 2000$ μM (CYP2C9 high affinity $K_m \approx 150$ μM) | 14 |
| CYP2C8 | Amodiaquine[b] | N-deethylation | LC/MS-MS | High specificity $K_m \approx 2$ μM | 115, 118 |
| CYP2C9 | Diclofenac[b] | 4′-Hydroxylation | LC/MS-MS | Good specificity $K_m \approx 2$ μM (some CYP2C19 $K_m \approx 80$ μM) | 115, 116 |
| | Tolbutamide | Hydroxylation | LC/MS-MS Radiometric | $K_m \approx 200$ μM (some CYP2C19 $K_m \approx 500$ μM) | 115, 119 |
| CYP2C19 | *S*-Mephenytoin[b] | 4′-Hydroxylation | LC/MS-MS | High specificity $K_m \approx 40$ μM | 115, 116 |
| CYP2D6 | Bufuralol[b] | 1-Hydroxylation | LC/MS-MS | High affinity component for CYP2D6 $K_m \approx 5$ μM (CYP2C19 and CYP2C9 at high concentrations) | 116 |
| | Dextromethorphan | O-demethylation | LC/MS-MS | | 115 |
| CYP2E1 | Chlorzoxazone | 7-Hydroxylation | LC/MS-MS | Questionable specificity $K_m \approx 700$ μM | 115, 117 |
| CYP3A4 | Midazolam[b] | 1′-Hydroxylation | LC/MS-MS | High specificity benzodiazepine $K_m \approx 3$ μM | 56, 115, 116 |
| | Testosterone | 6β-Hydroxylation | UV-HPLC Radiometric | High specificity steroid $K_m \approx 50$ μM | 56, 115, 117 |
| | Nifedipine | Oxidations | UV-HPLC | High specificity dihydropyridine $K_m \approx 50$ μM | 87 |

[a]Comments from Ref. 14.

[b]Primary Substrate used in this laboratory.

An additional aspect to be considered for *in vitro* assays is that CYP inhibition can be substrate dependent, particularly for CYP3A4 [122]. CYP3A4 is thought to bind substrates and inhibitors in multiple modes and/or binding sites [66, 123]. As a result, inhibitory interactions observed with one probe substrate may not be representative of those observed with other substrates [19]. This is due to the cooperative nature of substrate–inhibitor interactions within the active site [124]. Kenworthy et al. [122] classified 10 CYP3A4 substrates into three distinct groups: a benzodiazepine group (including diazepam, midazolam, triazolam, and dextromethorphan), a large molecular weight group (including testosterone, erythromycin, and cyclosporine), and finally nifedipine and benzyloxyresorufin (BROD), which fit in neither of the previous groups and are distinct from each other. These substrate groups were classified according to analysis of their behavior as probe reactions for the effect of 34 compounds on CYP3A4 mediated metabolism, with different inhibitor profiles being observed with different substrate groups. For example, the $IC_{50}$ for inhibition of CYP3A4 mediated nifedipine oxidation by haloperidol was 0.1 μM, while for dextromethorphan N-demethylation it was >100 μM [122]. It is postulated that haloperidol is able to inhibit CYP3A4 by binding at more than one site, with the effect being dependent on the substrate, reflecting the cooperative nature of CYP3A4.

There is also increasing evidence that similar multisite kinetics and substrate-dependent interactions are observed for CYP2C9 *in vitro* [67–69]. Therefore, while there is little evidence of auto-/heteroactivation *in vivo* in humans [69, 70, 125], it is good practice to evaluate the inhibitory potential of NCEs against CYP3A4 and CYP2C9 using at least two structurally dissimilar probe substrates, to afford a better predictive quality to the *in vitro* data as *in vivo* effects can be substrate dependent. In our laboratory, routine screens are conducted with drug-like probes for both CYP3A4 (midazolam) and CYP2C9 (diclofenac); for key compounds, a second and potentially a third probe is used to fully evaluate inhibition of these isoforms, for example, erythromycin and naproxen for CYP3A4 and CYP2C9, respectively. Figure 4b shows the effect of using either diclofenac or naproxen as probe substrate to measure the $IC_{50}$ for a range of compounds demonstrating CYP2C9 inhibition. On average, it was found that the $IC_{50,apparent}$ generated in an assay using naproxen as a substrate could be approximately 1.5-fold times that using diclofenac. This could be corrected from a consideration of $IC_{50,unbound}$ [68] (see Fig. 4b).

### 3.3 Analysis Methods

Several analytical methods are available for determining the impact of inhibition on a probe substrate. The selection of an analytical endpoint depends on the probe under investigation, the availability of the detection system, and the throughput required. LC/MS-MS is probably the system of choice for the pharmaceutical industry due to the selectivity and sensitivity it affords [115, 126]. This, combined with the fact that LC/MS-MS is the backbone of analysis within DMPK laboratories and therefore readily available, has driven its increased use in CYP inhibition assays. There are several LC/MS-MS methods that enable routine analysis of metabolites generated in microsomal and rCYP incubations, and importantly LC/MS-MS facilitates the use of clinically relevant probes. Additionally, the sensitivity afforded by LC/MS-MS has allowed multiple reaction monitoring of substrate cassettes, facilitating the investigation of test inhibitors for multiple CYP isoforms [20, 116, 121]. In

our laboratory we routinely screen for CYP inhibition using a substrate cassette and a rCYP cocktail; the details of this and other assays used are shown in Table 4.

An alternative to LC/MS-MS is HPLC-radioflow. This allows separation of metabolites of $^{14}C$ or $^{3}H$ labeled probe substrates. It is, however, a time-consuming technique requiring meticulous optimization of conditions and long run times, and LC/MS-MS is a more rapid alternative [115, 116, 118, 126]. A higher throughput radiometric endpoint assay is available and is flexible, sensitive, robust, and free from analytical interference [119, 127]. This laboratory uses a range of automated assays with [*O*- or *N*-methyl-$^{14}C$]-substrates, which liberate [$^{14}C$]-formaldehyde as the product of enzyme-specific oxidative demethylation [119]. However, the extra considerations involved with the use of radioactivity, such as the highly regulated disposal of isotopes and additional safety requirements, does constrain their use perhaps to second substrate selection.

Fluorometric assays are commonly used for CYP inhibition studies, especially in early drug discovery, as their speed and cost advantages lend them to enhanced throughput (compared to detailed manual assays) or true HTS (miniaturized assays employing robotics to utilize 96-, 384-, or 1536-well plate formats with rapid endpoints). Indeed, they are widely used throughout the pharmaceutical industry. Fluorescent-based assays are very amenable to multiwell plate assays; but despite their utility in rapidly generating large sets of inhibition data there are drawbacks. There can be fluorescent interference from the test compound or its product and issues with fluorescent quenching as well as lack of probe molecule specificity for an individual CYP [20]. The nonselective nature of fluorescent probes can, of course, be overcome by the use of single rCYPs. A potential concern for these assays is the non-drug-like qualities of most fluorescent probes, particularly in terms of clinically relevant CYP interactions. A detailed study evaluating inhibition data generated using fluorescent probes compared to that of conventional drug probes suggested that in general only a weak correlation between the two endpoints existed, with both endpoints not detecting inhibition by a significant number of compounds [19, 88]. Cohen et al. [19] suggest that concerns over fluorogenic probes not resembling drug molecules are unfounded and that it is the architecture of the CYP active site, permitting binding of structurally diverse molecules, that underlies interprobe differences in $IC_{50}$ values, be they fluorescent or not. In our laboratory, fluorometric assays have been widely replaced by LC/MS-MS-based assays but are still used as an alternative substrate assay to cross-check inhibition data for CYP3A4 and CYP2C9, as it is prudent to assess inhibition in detailed studies using multiple probes and alternative endpoints as least for these enzymes.

Finally, luminescence-based CYP selective probes have recently become commercially available, such as those used in the Promega P450-Glo™ assays. These assays provide a luminescent method for measuring CYP activity from recombinant and native sources. A conventional CYP reaction is performed by incubating the CYP with a luminogenic CYP substrate and NADPH regeneration system. The substrates in the P450-Glo assays are derivatives of beetle luciferin ((4*S*)-4,5-dihydro-2-(6-hydroxybenzothiazolyl)-4-thiazolecarboxylic acid) and are substrates for CYP but not luciferase. The derivatives are converted by CYP to a luciferin product that is detected in a second reaction with a luciferin detection reagent. The amount of light produced in the second reaction is directly proportional to the activity of the CYP (http://www.promega.com). Such assays lend themselves to true

**TABLE 4 Assay Conditions Used for $IC_{50}$ Generation in the Authors' Laboratory Utilizing Fluorescent or Radiometric Endpoints and a Single rCYP Isoform and LC/MS-MS Endpoint with a Cocktail of rCYPs and Cassette of Substrates[a]**

| Isoform | Substrate | Final [Substrate] (μM) | [CYP] (pmol/mL) | Incubation Time (min) | Endpoint | Vehicle (1% v/v) | Inhibitor ($IC_{50}$; μM) |
|---|---|---|---|---|---|---|---|
| CYP1A2 | Ethoxyresorufin | 0.7 | 17 | 15 | $\lambda_{ex}$ 530 nm; $\lambda_{em}$ 590 nm | DMSO | Fluvoxamine (0.17) |
| CYP2C9 | 7-Methoxy-4-(trifluoromethyl) coumarin | 50 | 50 | 20 | $\lambda_{ex}$ 410 nm; $\lambda_{em}$ 535 nm | DMSO | Sulfaphenazole (0.3) |
| CYP2C19 | 7-Methoxy-4-(trifluoromethyl) coumarin | 50 | 57 | 20 | $\lambda_{ex}$ 405 nm; $\lambda_{em}$ 535 nm | MeCN | Omeprazole (3) |
| CYP2D6 | 3-[2-(*N*,*N*-diethyl-*N*-methylammonium)ethyl]-7-methoxy-4-methlycoumarin | 10 | 56 | 15 | $\lambda_{ex}$ 530 nm; $\lambda_{em}$ 590 nm | DMSO | Quinidine (0.060) |
| CYP3A4 | 7-Benzyloxy-4-(trifluoromethyl) coumarin | 10 | 9 | 15 | $\lambda_{ex}$ 405 nm; $\lambda_{em}$ 535 nm | MeOH /MeCN | Ketoconazole (0.007) |
| CYP2C9 | Naproxen | 109 | 70 | 15 | Liquid scintillation counting | DMSO | Sufaphenazole (0.45) |
| CYP2C19 | Diazepam | 21 | 40 | 15 | Liquid scintillation counting | DMSO | Omeprazole (3) |
| CYP2D6 | Dextromethorphan | 5 | 20 | 15 | Liquid scintillation counting | DMSO | Quinidine (0.025) |
| CYP3A4 | Erythromycin | 19 | 25 | 10 | Liquid scintillation counting | DMSO | Ketoconazole (0.035) |
| 5 CYP cocktail | 1A2–Phenacetin | 26 | 15 | | | | α-Naphthoflavone(0.014) |
| | 2C9–Diclofenac | 2 | 5 | | | | Sulfaphenazole (0.26) |
| | 2C19–S-mephenytoin | 31 | 3 | 10 | LC/MS-MS | DMSO | Tranylcypromine (3.60) |
| | 2D6–Bufurolol | 9 | 5 | | | | Quinidine (0.018) |
| | 3A4–Midazolam | 3 | 5 | | | | Ketoconazole (0.0036) |
| | | | *Total 33* | | | | |
| 3 CYP cocktail | 2B6–Bupropion | 20 | 18 | | | | Ticlopidine (0.55) |
| | 2C8–Amodiaquine | 2 | 1 | 10 | LC/MS-MS | DMSO | Quercetin (5.12) |
| | 3A5–Midazolam | 2 | 5 | | | | Ketoconazole (0.17) |
| | | | *Total 24* | | | | |

[a]See Refs. 116, 118, and 119. Data is mean $n > 50$.

**TABLE 5 Some Considerations in the Selection of CYP Assay Endpoints**

| | Assay Type | | | |
|---|---|---|---|---|
| Consideration | Radiolabeled | Fluorescent | LC/MS-MS | Luminescent |
| Cost | ++ | +++ | + | ++ |
| Throughput | + | +++ | ++ | HTS |
| Accessibility | ++ | ++ | + | ++ |
| Interference | +++ | + | ++ | + |
| Health and safety | + | +++ | +++ | +++ |
| Published information | ++ | +++ | +++ | + |
| Marketed drug substrates | ++ | — | +++ | — |

HTS capabilities, providing quick readouts on multiple well plates. Table 5 shows a comparison of LC/MS-MS, fluorometric, radiometric, and luminescent endpoints to aid consideration of assay choice.

Using fluorescent or luminescent assays for enhanced/HTS and rapid generation of $IC_{50}$ data may be invaluable for the development of structure–activity relationships (SARs) for the different CYPs [69, 128]. The utility of high throughput CYP inhibition screens has allowed generation of large datasets, facilitating development of statistical *in silico* approaches. To this end, quantitative *in silico* models now exist for prominent CYPs, often based on physical-chemical properties alone. These models can classify compounds as "noninhibitors" with reasonable accuracy (~80% of compounds were classified correctly for CYP3A4 and CYP2D6 [128]) and can guide medicinal chemistry away from CYP interactions at early stages of drug discovery programs [128, 129]. In theory, these *in silico* models allow predictions of inhibitory potency to be made even before a compound has been synthesized [129].

## 4 GENERATION OF *IN VITRO* CYTOCHROME P450 INHIBITION DATA

Fundamentally, *in vitro* kinetic data must be accurate, meaningful, and relevant. As discussed previously, all assays should be preoptimized for time and protein linearity to ensure the reaction proceeds under linear conditions. The objective is to determine $K_i$ for predicting the likelihood and extent of DDIs. Experimentally, this is achieved by either defining the $IC_{50}$ or by directly determining a $K_i$. From Eqs. 4 and 7, it is clear that if $[S] = K_m$, the $K_i$ can be estimated to within twofold from an accurate $IC_{50}$ determination. This may well be less than the error in determining $K_i$ by an extensive (and laborious) fully inhibited Michaelis–Menten study. Furthermore, the mechanism of inhibition must be defined and, if irreversible, generation of $K_I$ and $k_{inact}$ is required for accurate assessment of DDIs.

When generating *in vitro* inhibition data, particularly when evaluating the propensity of NCEs to behave as CYP inhibitors, it is important to include the appropriate positive controls. To this end, known inhibitors should be included in any incubations as markers, and it has been suggested that $K_i$ or $IC_{50}$ values generated

for these standard inhibitors should fall within threefold of the median literature value for an assay to be deemed acceptable [130].

### 4.1 Reversible Inhibition ($IC_{50}$ and $K_i$)

An $IC_{50}$ is generated using a single concentration of substrate. Inhibitor concentrations should cover at least two log units (e.g., 0, 0.1, 0.3, 1, 3, 10 μM) to ensure a range of measurements from negligible to virtually complete inhibition. The percentage of total enzyme activity remaining (from control incubations) is plotted against log inhibitor concentration [I] to generate an $IC_{50}$ curve. Plotting log [I] results in more accurate confidence intervals when using nonlinear regression, as the data is equally spaced on a log axis, allowing symmetrical confidence intervals to be calculated. In the authors' laboratory, an automated assay to determine $IC_{50}$ is adopted; $K_i$ is estimated to be half the $IC_{50}$ (as [S] = $K_m$). The automated assay utilizes a pool of recombinant CYPs (CYP1A2, CYP2C9, CYP2C19, CYP2D6, and CYP3A4) to minimize nonspecific binding in the incubation, at protein concentrations and incubation times designed to ensure the selective metabolism of a substrate cassette at concentrations equivalent to the $K_m$ (phenacetin, diclofenac, *S*-mephenytoin, bufuralol, and midazolam, respectively). LC/MS-MS analysis of specific metabolite formation allows comparison of peak area in the presence of increasing concentration of test inhibitor against a DMSO control. Recently, an additional assay was developed for CYP2B6, CYP2C8, and CYP3A5 due to the emerging importance of these enzymes [118]. This second automated assay again uses a pool of rCYPs and a substrate cassette of bupropion, amodiaquine, and midazolam, respectively. Table 4 shows the reaction conditions for these assays.

An alternative approach is to experimentally define the inhibition constant $K_i$. To do this accurately, the inhibitor concentration should span a range around the suspected $K_i$ (based on a prior $IC_{50}$ estimation) using at least five concentrations, for example, $0K_i$, $0.3K_i$, $K_i$, $3K_i$, and $10K_i$. Typically, 8–10 substrate concentrations should be used to investigate each concentration of inhibitor and these should span from $K_m/3$ to $3K_m$. The range of substrate concentration needs to be expanded for increased inhibitor concentrations, since the $K_m$ shifts up when the substrate's metabolism is competitively inhibited. $K_i$ determinations therefore require large numbers of incubations, as all parts of the velocity–substrate concentration profile need to be defined for each inhibitor concentration. Accurate determinations of any one of the parameters ($K_m$, $V_{max}$, $K_i$) depends on accurate determination of all three.

Once generated, the data can be visually inspected for the mechanism of inhibition; this is commonly achieved using two types of graphical representation—the Eadie–Hofstee plot ($v$ against $v$/[S]) and the Dixon plot (1/$v$ against [I]). The type of inhibition can be determined from the shape of the linearized data and examples of these plots for each type of reversible inhibition are shown in Fig. 7. Kinetic parameters should never be determined from linear transforms such as these as doing so generally leads to inaccurate estimation of $K_i$ (the two-dimensional nature of the data fit does not fully describe the effect of changing inhibitor concentrations on different concentrations of substrate). The plots are solely for the purpose of aiding interpretation of the inhibition type. Consequently, nonlinear regression software is an absolute requirement to facilitate iterative modeling of changing sub-

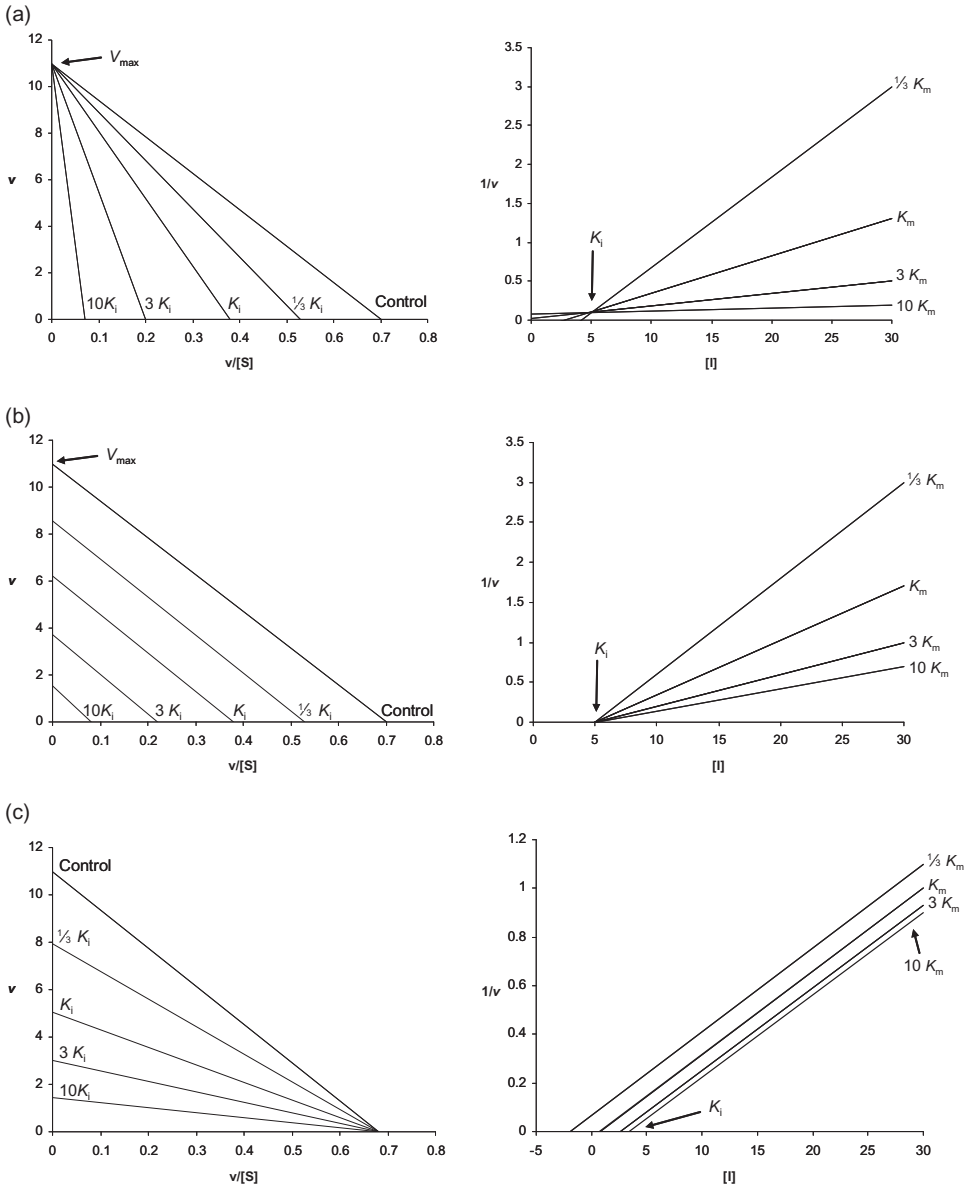


**FIGURE 7** Graphical representation of inhibition plots: Eadie–Hofstee plots ($v$ against $v$/[S] on the left) and Dixon plots (1/$v$ against [I] on the right) for (a) competitive inhibition, (b) noncompetitive inhibition, (c) and uncompetitive inhibition.

strate and inhibitor concentrations, fitting the untransformed data to the appropriate rate equation (for competitive, noncompetitive, or uncompetitive inhibition, Eqs. 3, 5, and 6, respectively). In practice, this means nonlinear regression modeling of all the substrate concentration, velocity, and inhibitor concentration data simultaneously using the equation for competitive inhibition (Eq. 3). The data should be

weighted by $1/y$, to compensate for unequal experimental variance in the different experimental points. (Velocity–substrate concentration plots tend to have a constant relative error associated such that the residual—the difference between predicted and observed velocity—will be larger at higher reaction velocities. This is true because the error is the same at low and high velocities, giving rise to larger standard deviations from the mean observed velocity at higher velocities.) The same dataset should then be reanalyzed using the equation for noncompetitive inhibition (Eq. 5). The two mathematical models can then be discriminated by evaluation of "goodness-of-fit" criteria. For a full explanation, see Mannervik [131]. Generally, goodness-of-fit criteria (e.g., Akaike values) and accuracy of parameter estimates (standard deviation and coefficient of variance) should be compared for both models and the most appropriate model (type of inhibition) chosen.

### 4.2 Mechanism-Based Inhibition ($K_I$, $k_{inact}$, and $IC_{50}$)

True irreversible inhibition or mechanism-based inhibition (MBI) of CYPs can only be defined according to the criteria detailed previously and may be established by mechanistic studies such as dialysis of the inhibited CYP or spectral shifts in absorbance [100]. MBI is not readily detected in standard *in vitro* assay formats. Time-dependent inhibition (TDI) is a collective term for a change in potency of CYP inhibitors during an *in vitro* incubation or dosing period *in vivo*. Such changes usually result in an increase in inhibitory potency; potential mechanisms include the formation of more inhibitory metabolites and MBI. In the following assays, it is the time-, concentration-, and/or NADPH-dependent nature of the inhibition under investigation.

In essence, an assay to detect TDI involves a preincubation step, where enzyme is incubated with and without NADPH and in the presence of varying inhibitor concentrations. This reaction is allowed to proceed over a predetermined time course, before dilution of the reaction mixture into a second incubation that includes a probe substrate to assess the degree of enzyme inactivation [77]. To accurately define the inhibitory constants of TDI, the conditions governing the two separate incubations must be carefully chosen. A useful critical evaluation of experimental design, methodology, and data analysis for TDI protocols in published literature highlighted not only the variation between laboratories but the subsequent impact on the predicted effects of TDI *in vivo* if the *in vitro* kinetic parameters are incorrectly defined [79].

The experimental design of a TDI assay must adhere to the same conditions as those for reversible inhibition (linear enzyme reactions, etc.) and protocols have to be adjusted to correctly define the kinetics of different inhibitors. These adjustments are dependent on the potency of the inhibitor under investigation, as potent inhibitors require considerably shorter preincubation times than weaker inhibitors [75, 79]. To this end, the time course of the pre-incubation must be such that the inhibition at each concentration of inhibitor is linear with respect to time—at least over the initial preincubation period. Allowing the preincubation reaction to proceed over a large time course with multiple dilution points into the second assay (0, 0.5, 1, 2, 3, 5, 10, 20, and 30 min) should gauge this; however, this is not necessarily practical due to the large reaction volume required in the preincubation to sustain multiple dilution points. In practice, two separate assays may have to be conducted: an

initial assessment where preincubation occurs over an intermediate time course (such as 5, 10, and 20 min), and a second assay where times are adjusted for more accurate definition of the inhibition profile (based on inspection of the data generated in the initial assessment). In addition to this, it is vital to ensure all the CYP inactivation occurs in this preincubation, to allow accurate measurement of inhibited enzyme with respect to incubation time, without concurrent inhibition and probe substrate metabolism in the second incubation. To this end, the dilution of the preincubation reaction mixture into the second incubation should be as large as possible (in general, 10–50-fold) while still guaranteeing analytical sensitivity for the probe substrate metabolite. Additionally, the concentration of CYP selective substrate in the final incubation should be high relative to its $K_m$ (in general, 5–10-fold), again to minimize further CYP inactivation after the preincubation step, and the final incubation time should be minimized. Finally, since the enzyme can undergo some inactivation in the absence of inhibitor, the enzyme activity remaining following each preincubation time point should be determined by comparison to control (enzyme incubated without inhibitor). In the authors' laboratory, two automated assays with an LC/MS-MS endpoint are used to assess TDI potential [98] and have been modified to allow detection of weaker time-dependent inhibitors [75]. An initial screen generates an $IC_{50}$ value using HLMs and a cassette of CYP specific probe substrates. A single inhibitor concentration is tested using a single preincubation time of 30 min and a 20-fold dilution. If inhibition is detected in this $IC_{50}$ assay, a second assay is then used to generate values for $K_I$ and $k_{inact}$ from rCYPs. This assay adheres to the strict criteria detailed earlier in order to accurately define the kinetic parameters. Five preincubation times (up to 23 min) are used with six inhibitor concentrations and a dilution factor of 20-fold. There is a good correlation between $k_{inact}/K_I$ ratio (a measure of inhibitory efficiency) generated in rCYP and the $IC_{50}$ from the HLM assay, thus allowing NCE evaluation from a reasonable throughput assay [98] (see Fig. 8).

The analysis of data generated in TDI assays must be carefully conducted. First, the natural log of the percent control activity remaining is plotted against time. The slope of the linear portion for each inhibitor concentration is equivalent to the inactivation rate constant, $k$ (or $k_{obs}$). Graphical representations of these plots are shown in Fig. 9. Using nonlinear regression, this data can be fitted to Eq. 13 to estimate $k_{inact}$ and $K_I$. A common fault is to use linear Kitz–Wilson plots rather than nonlinear regression, as two-dimensional fitting of the data cannot yield accurate estimations for $k_{inact}$ and $K_I$. [79]. It is also worth noting that if all the CYP inactivation occurs in the preincubation step, the point where the log-linear fits of each inhibitor concentration cross the $y$ axis (time = 0) represents the reversible inhibition component. However, if the dilution of the preincubation is not sufficient, considerable reversible inhibition may occur during the second incubation, making data interpretation more problematic [98].

Finally, an additional utility of a TDI screen in drug discovery may be as a "flag" for reactive metabolite formation. Electrophilic species are often generated during CYP MBI and there is a high chance that these may escape from the CYP active site and react elsewhere in the body, either via detoxification processes (glutathione deactivation) or more worryingly via reaction with proteins or nucleic acids. It is not at all surprising that the same chemical groups implicated in reactive metabolite formation are also known to inhibit CYPs irreversibly [75]. Perhaps linked to this,

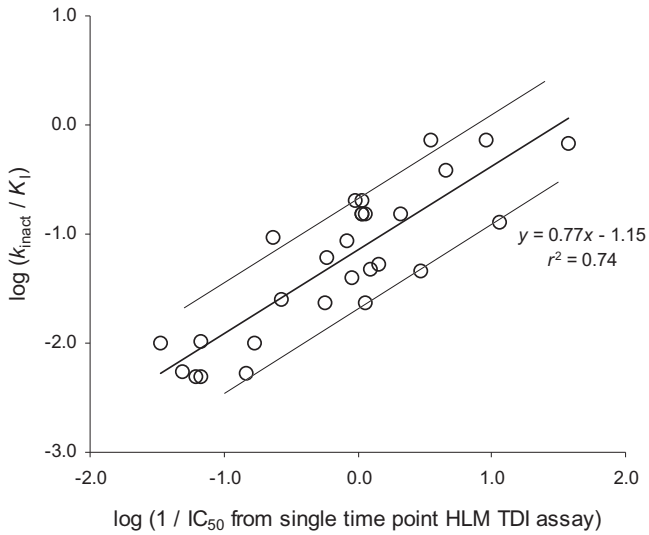


**FIGURE 8** Plot of log unbound $IC_{50}$ against log ($k_{inact}/K_I$) for 28 time-dependent inhibitors of CYP3A4. $IC_{50}$ values were generated using the automated single time point, single inhibitor concentration human liver microsomal time-dependent inhibition screen [98]. These values were adjusted for incubational binding to give unbound $IC_{50}$ values (according to the mathematical model of Austin et al. [111]). $K_I$ and $k_{inact}$ were estimated using the 3 time point, 6 inhibitor concentration, recombinant CYP automated time-dependent inhibition assay. The solid line is the line of best fit, with dashed lines representing threefold deviation from the line. (From Ref. 98.)

it is found that many drugs exhibiting TDI are also associated with adverse drug reactions (e.g., erythromycin, diclofenac, tacrine, valproic acid, halothane, suprofen, ticlopidine, tienilic acid, isoniazid, zileuton) [132, 133]. Using TDI screening, alongside cold trapping and/or radiolabeled binding studies, may be advantageous in minimizing the potential for reactive metabolite formation as early as possible in a drug discovery screening cascade.

## 5 *IN VITRO–IN VIVO* EXTRAPOLATION OF CYTOCHROME P450 BASED DRUG–DRUG INTERACTIONS

The clinical implications of DDIs and the importance of generating accurate *in vitro* data have been discussed; from this it follows that IVIVE requires careful data handling using the approach outlined below to generate predictions of clinical impact. Over the last 10–15 years, predicting the extent of DDI in a semiquantitative way has been the subject of many publications and the most effective models are presented here.

### 5.1 Approach for Reversible Inhibitors

If the metabolism of a drug is reversibly inhibited, the magnitude of the effect will be $1 + [I]/K_i$ (sometimes termed the inhibition index, $I_i$), provided that the concen-

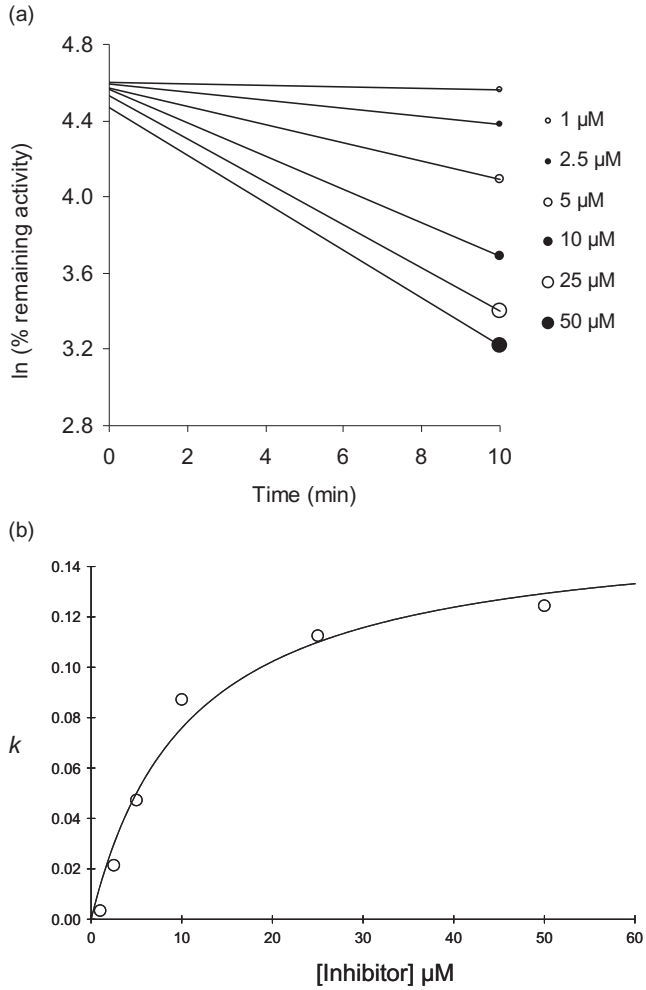


**FIGURE 9** Graphical representation of irreversible (time-dependent) inhibition. (a) Inhibition of control activity against time for increasing inhibitor concentration; and (b) nonlinear regression of $k$ against inhibitor concentration giving $k_{inact}$ of 0.16 min$^{-1}$ and $K_I$ of 10.6 μM (circles are observed data points and line is predicted data fit).

tration of substrate undergoing inhibition is well below its $K_m$ value and that it is eliminated by the single pathway being inhibited [134]. This is because the magnitude of effect is dependent not only on how much affinity the inhibitor has for the enzyme ($K_i$) but also on the concentration of inhibitor available. For competitive inhibition, the $K_m$ of the victim drug will be increased to $K_m + K_m$ [I]/$K_i$ (or $K_m(1 +$ [I]/$K_i$)) and therefore the intrinsic clearance ($CL_{int}$) (as $CL_{int} = V_{max}/K_m$) will be decreased to $V_{max}/(K_m(1 +$ [I]/$K_i)$ or $CL_{int}/(1 +$ [I]/$K_i)$.

For an IV administered victim drug, the following is true:

$$CL = \text{Dose}/AUC \tag{14}$$

$$CL/CL' = AUC'/AUC \tag{15}$$

where $CL$ is the *in vivo* clearance in the absence and $CL'$ in the presence of inhibitor, respectively, and $AUC$ is area under the plasma–concentration time curve in the absence and $AUC'$ in the presence of inhibitor′, respectively.

$$\frac{AUC'}{AUC} = \frac{Q_h \cdot CL_{int} \cdot f_{ub}}{Q_h + CL_{int} \cdot f_{ub}} \cdot \frac{Q_h + CL'_{int} \cdot f_{ub} \cdot 1/I_i}{Q_h \cdot CL'_{int} \cdot f_{ub} \cdot I_i} \tag{16}$$

$$= \frac{Q_h^2 \cdot CL_{int} \cdot f_{ub} + Q_h^2 \cdot CL_{int} \cdot f_{ub} \cdot 1/I_i \cdot CL_{int} \cdot f_{ub}}{Q_h^2 \cdot f_{ub} \cdot CL_{int} \cdot 1/I_i + Q_h^2 \cdot f_{ub} \cdot CL_{int} \cdot f_{ub} \cdot CL_{int} \cdot 1/I_i} \tag{17}$$

$$= \frac{Q_h + (f_{ub} \cdot CL_{int})/I_i}{Q_h/I_i + (f_{ub} \cdot CL_{int})/I_i} \tag{18}$$

$$\frac{AUC'}{AUC} = \frac{Q_h \cdot I_i + f_{ub} \cdot CL_{int}}{Q_h + f_{ub} \cdot CL_{int}} \tag{19}$$

where $Q_h$ is hepatic blood flow and $I_i$ is the inhibition index $1 + [I]/K_i$. If $f_{ub} \cdot CL_{int} << Q_h$, there is a high inhibitory effect and the $AUC$ ratio approximates to $I_i$ $(1 + [I]/K_i)$. If $f_{ub} \cdot CL_{int} >> Q_h$, there is little effect since the $AUC$ ratio approximates to 1.

For an orally administered victim drug, the fraction escaping hepatic first-pass extraction in the presence ($F'$) or absence ($F$) of inhibitor is

$$\frac{F'}{F} = \frac{1 - CL'_h/Q_h}{1 - CL_h/Q_h} \tag{20}$$

Rearranging the well stirred model [135],

$$CL_{int} = \frac{CL_h}{f_{ub}(1 - CL_h/Q_h)} \tag{21}$$

$$= \frac{CL_h}{F \cdot f_{ub}} \tag{22}$$

It follows that

$$F = \frac{CL_h}{CL_{int} \cdot f_{ub}} \tag{23}$$

Therefore,

$$F = \frac{CL_{int} \cdot f_{ub} \cdot Q_h}{CL_{int} \cdot f_{ub} + Q_h} \cdot \frac{1}{CL_{int} \cdot f_{ub}} \tag{24}$$

$$= \frac{Q_h}{CL_{int} \cdot f_{ub} + Q_h} \tag{25}$$

Putting this back into Eq. 20, we find

$$\frac{F'}{F} = \frac{Q_h}{CL_{int} \cdot f_{ub}/I_i + Q_h} \cdot \frac{CL_{int} \cdot f_{ub} + Q_h}{Q_h} \tag{26}$$

$$\frac{F'}{F} = \frac{CL_{int} \cdot f_{ub} + Q_h}{(CL_{int} \cdot f_{ub}/I_i) + Q_h} \tag{27}$$

So, when $f_{ub} \cdot CL_{int} << Q_h$, there is little inhibitor effect since the ratio of the fraction escaping first pass is approximately 1. When $f_{ub} \cdot CL_{int} >> Q_h$, the ratio of the fraction escaping first pass in the presence and absence of inhibitor approximates to $1 + [I]/K_i$.

It is recognized as a "rule of thumb" that inhibitors with $[I]/K_i < 0.1$ represent a low DDI risk, those with $[I]/K_i$ between 0.1 and 1 are moderate risk, and those with $[I]/K_i > 1$ are high risk. The relationship between the *AUC* ratio and $[I]/K_i$ ratio is depicted in Fig. 10 and is the most simple model of CYP inhibition [80, 130, 136, 137]. A more detailed approach, that is, the fundamental basis of most IVIVE for CYP inhibition, expands on the inhibition index $(1 + [I]/K_i)$ by including the relative metabolic contribution (*fm*) of each inhibited CYP to the *in vivo* metabolism of the substrate under investigation:

$$\frac{AUC'}{AUC} = \frac{1}{\frac{fm_{CYP}}{1+[I]/K_i} + (1 - fm_{CYP})} \tag{28}$$

The impact of *fm* on the magnitude of inhibition can be visualized in Fig. 11 [56, 68, 138]. From this plot it is clear that $fm > 0.5$ is required for inhibition to generate significant increases in *AUC* and that predictions are very sensitive to the *fm* used when it approaches 1. Within a population, *fm* will vary and individuals most at risk

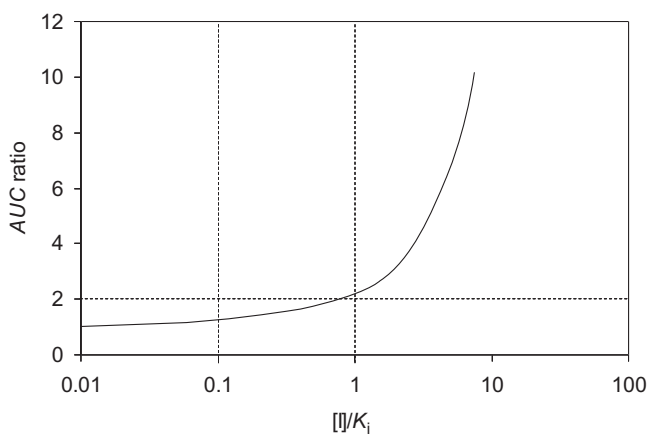


**FIGURE 10** Qualitative zoning of the prediction of the magnitude and risk of drug–drug interactions mediated by reversible inhibition (*AUC* ratio) using $[I]/K_i$ ratio. Low risk, <0.1; medium risk, 0.1–1; high risk, >1.

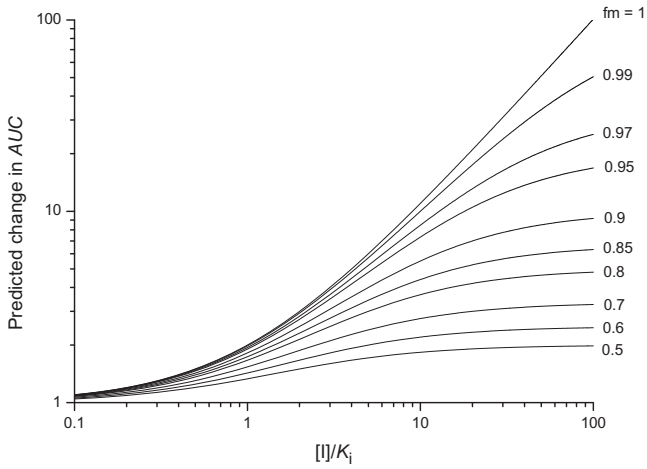


**FIGURE 11** Simulation of predicted change in *AUC* as a function of $[I]/K_i$ with different *fm* values. The simulation is based on Eq. 28, where *fm* is the fraction of substrate clearance mediated by the inhibited metabolic pathway. (Adapted from Ref. 98.)

of DDI will be those with the highest *fm*. The model can be expanded further to include inhibition of multiple CYP isoforms and the effect of each contribution to overall clearance of the victim drug and other routes of elimination need to be considered [136, 138].

The choice of which [I] to use in these predictions has been a source of much debate in the literature [7, 13, 80, 88, 139, 140]. Classical pharmacology states that only the unbound fraction of a drug is able to cross membranes and interact with a target protein; from this it follows that a drug associated with proteins in the plasma or tissue is unavailable for interaction. It follows that the use of unbound drug concentrations and inhibition constants should be the starting point in IVIVE and doing so can give good predictions when other relevant factors are included [88]. Changes in the unbound fraction of a drug in the blood ($f_{ub}$) is in itself a potential mechanism of DDI; $f_{ub}$ is increased when one drug is displaced by other drugs at the site of plasma protein binding. These interactions rarely cause serious clinical problems [7, 141]. However, despite the rationale for using corrections of unbound drug concentrations to predict clearance [108, 142], the use of drug binding terms in predictive IVIVE models remains controversial [13, 88, 97, 139].

Across the spectrum of literature on the subject, inconsistencies in the use of drug binding terms and the choice of *in vivo* inhibitor concentration may have clouded the general understanding and inadvertently reduced confidence in IVIVE for anything other than *qualitative* DDI predictions (average steady plasma concentration, $[I]_{av}$, peak steady-state plasma concentration, $[I]_{max}$, maximum hepatic input concentration $[I]_{in}$; see Table 6) [7, 13, 80, 144]. The suggestion that total blood concentrations can be used with unbound *in vitro* $K_i$ values to achieve good predictivity is less than helpful, since it leaves the reader potentially unsure as to the impact of blood binding *in vivo*. Binding to blood components may not reach equilibrium during the short time period in the portal vein, leading to a greater free-drug concentration than may be predicted and underestimation of DDI [136]. However, the

**TABLE 6 Choice of *In Vivo* Inhibitor Concentration Advocated in the Literature in Recent Years**

| Which [I] | Equation | $K_i$ Adjusted for $f_{uinc}$? | Author Comments | References |
|---|---|---|---|---|
| Maximum *unbound* hepatic input concentration, $[I]_{in}$, where [I] is $[I]\cdot f_{ub}$ | $[I]_{av} = \frac{D/\tau}{CL/F}$ <br> $[I]_{in} = [I]_{av} + \frac{k_a \cdot F_a \cdot D}{Q_h}$ | No | The incidence of false-negative predictions was largest using $[I_{av}]$ for unbound drug concentration and smallest using $[I_{in}]$ for total drug concentration for each of the CYP enzymes. The use of $[I_{in}]$ is recommended because true negatives can be identified and, in contrast to the use of other [I] values, false negatives eliminated. | 137 |
| Maximum *total* hepatic input concentration, $[I]_{in}$ | $[I]_{in} = [I]_{av} + \frac{k_a \cdot F_a \cdot D}{Q_h}$ | No | True negatives can be identified and, in contrast to the use of other values for [I], false negatives are eliminated. True positives are also predicted well and, although the incidence of false-positive predictions is quite high, the use of hepatic input concentration is recommended. | 80 |
| Maximum *total* hepatic input concentration $[I]_{in}$ | $[I]_{in} = [I]_{av} + \frac{k_a \cdot F_a \cdot D}{Q_h}$ | No | The use of $[I]_{in}$ incorporating both $fm_{CYP}$ and refined $k_a$ values resulted in the most successful prediction overall. | 143 |
| Average *total* steady plasma concentration, $[I]_{av}$, or maximum *total* hepatic input concentration, $[I]_{in}$ | $[I]_{in} = [I]_{av} + \frac{k_a \cdot F_a \cdot D}{Q_h}$ <br> $[I]_{av} = \frac{D/\tau}{CL/F}$ | Yes | Either the average systemic plasma concentration after repeated oral administration ($[I]_{av}$) or the maximum hepatic input concentration ($[I]_{in}$) | 56 |
| Average *total* steady plasma concentration, $[I]_{av}$ | $[I]_{av} = \frac{D/\tau}{CL/F}$ | Yes | The use of $[I]_{av}$ as the [I] surrogate generated the most successful predictions as judged by several criteria. Using $[I]_{av}$ incorporation of either plasma protein binding of inhibitor or gut wall CYP3A4 inhibition did not result in a general improvement of DDI predictions. | 13 |
| Maximum *unbound* hepatic input concentration, $[I]_{in,u}$, where [I] is $[I]\cdot f_{ub}$ | $[I]_{in} = [I]_{av} + \frac{k_a \cdot F_a \cdot D}{Q_h}$ | Yes ([protein] low so $f_{uinc}$ assumed negligible) | The use of estimated unbound hepatic inlet $C_{max}$ during the absorptive phase yielded the most accurate predictions of the magnitudes of DDI. | 139 |

"free-drug" theory should be regarded as the default situation and if successful predictions cannot be made, the reason(s) should be sought.

Equation 29 shows the maximum hepatic input of inhibitor (the sum of the contribution from the systemic circulation and the absorption phase; $[I]_{in}$) calculated using the absorption rate constant ($k_a$), the fraction absorbed ($F_a$), and dose [7].

$$[I]_{in} = [I]_{av} + (k_a \cdot F_a \cdot \text{Dose})/Q_h) \tag{29}$$

Often it is assumed that the rate of absorption is instantaneous (approaching $0.1\,\text{min}^{-1}$); however, changes in $k_a$ can account for marked differences in $[I]_{in}$ and the value chosen should be carefully considered [143].

In the authors' laboratory, good predictions have been made using unbound maximum hepatic concentration $[I]_{in,u}$ along with correction of $K_i$ for binding in the incubation ($f_{uinc}$) [88]. Using a published dataset of clinical DDIs [80], an 84% success rate was achieved for prediction of DDI at CYP3A4 and a 91% success rate predicting CYP2D6 interactions (Fig. 12) [88]. There were some noticeable outliers in the dataset, however; these highlighted that subtle changes in some parameters (e.g., plasma protein values for a highly bound drug) can result in large discrepancies in calculated unbound concentration and that measured *AUC* changes for drugs with low oral bioavailability are subject to large errors. Additionally, it was apparent that the impact of the $[I]/K_i$ estimate on *AUC* change becomes small once $fm_{CYP}$ falls below 0.85 (Fig. 11) [88]. The clinical dataset is substantially smaller for CYP2C9 than for CYP3A4 and CYP2D6, but it was found that using $[I]_{in,u}$ and $K_{i,u}$ to predict changes in *AUC* resulted in 14 correct assignations and 1 false negative [68]. However, the desire to avoid false negatives in a drug discovery screening

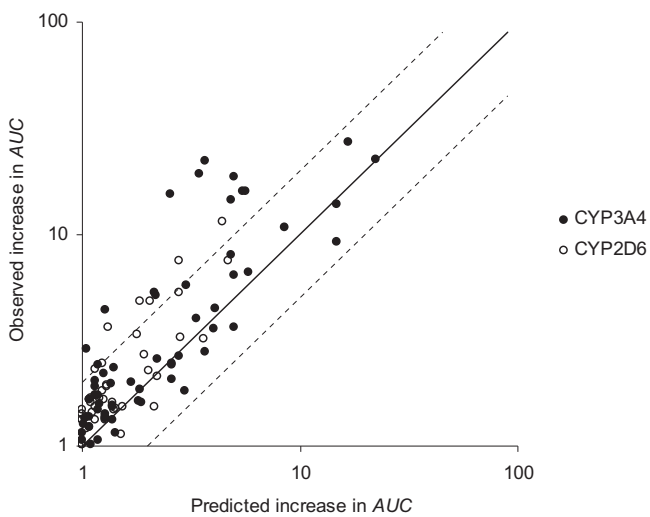


**FIGURE 12** Observed and predicated *AUC* values for CYP3A4 and CYP2D6 DDIs. The square box represents twofold increase in *AUC*, the solid line is unity and the dashed lines are twofold errors in either direction. (Adapted from Ref. 88.)

cascade perhaps explains the popularity for using *total* concentration terms for [I], despite the decreased accuracy in absolute prediction afforded by doing so.

The multifaceted nature of CYP inhibition *in vivo* requires a great deal of understanding for accurate prediction, and the reasons behind the lack of a unifying model can be appreciated. As the IVIVE field progresses, it is anticipated that prediction accuracy will increase as there is profound interest and utility in such models within the pharmaceutical industry and considerable academic effort being made toward this goal. The ultimate aim is an integrated approach toward CYP inhibition prediction, which encompasses the role of all CYP isoforms and is underpinned by sound pharmacokinetic principles.

### 5.2 Approach for Mechanism-Based Inhibitors

Two models are currently presented in the literature and they rely on some key parameters for accurate predictions of DDI [7, 100]. As for reversible inhibition, the relevant concentration of inhibitor at the target CYP and the fraction of substrate clearance affected by inhibiting the enzyme are fundamental. In addition to this, predictions of MBI also employ the *in vivo* rate of enzyme synthesis ($k_{synth}$) and of degradation ($k_{degrad}$), and the *in vitro* inactivation kinetic parameters ($K_I$ and $k_{inact}$). Where the currently available models differ is in the complexity of predicting the effect of changing inhibitor concentration over a dosing interval and the impact of this on the overall amount of inactivated CYP.

A model first presented by Hall and co-workers and subsequently expanded by the same group [100, 145, 146] is used most widely. This model assumes that the rates of enzyme synthesis and degradation are unaffected by enzyme inactivation and rely on the steady-state concentration of active enzyme being proportional to the ratio of the rate of synthesis and the rate of degradation. In the presence of inhibitor, the rate of degradation is then the sum of the *in vivo* degradation rate constant ($k_{degrad}$) and the rate constant describing the drug-induced inactivation ($\lambda$), where

$$\lambda = ([\mathrm{I}] \cdot k_{inact} / [\mathrm{I}] + K_I) \tag{30}$$

The model including inactivation of intestinal CYP and the fraction of drug metabolized by the inhibited pathway in the absence of the inhibitor is given in by

$$\frac{AUC}{AUC'} = \frac{F'}{F} \cdot \left(1 + \frac{fm}{([\mathrm{I}] \cdot k_{inact})/(k_{degrad} \cdot (K_I + 1))} + (1 - fm)\right)^{-1} \tag{31}$$

This model has been shown to predict well for verapamil [145] and other mechanism-based inhibitors [147]. The accuracy of prediction is dependent on the $k_{degrad}$ used, especially for substrates undergoing substantial ($fm > 0.9$) metabolism by the inhibited enzyme [56].

As with reversible CYP inhibition, a nonlinear regression model encompassing the various parameters predicting changing inhibitor and substrate concentrations affords the most sophisticated approach. To that end, an elegant model was proposed by Ito et al. [7], which predicts the effect of macrolide antibiotics on CYP3A4 [104]. A real advantage of this model is that it allows visualization of the plasma concentration–time profiles for the affected substrate and the inhibitor, and more importantly the change in active enzyme concentration over time. This model encompasses not only $k_{inact}$ and $K_I$ and the rate of resynthesis of active enzyme but also the changing concentration of the inhibitor as a function of its own disposition. Three compartments are represented: portal vein, liver, and systemic blood. The rate of change of inhibitor over time is modeled in each compartment with and without coadministered drug and a final compartment describes inactivation of hepatic CYP (Fig. 13). The use of this model requires nonlinear regression in a seven-compartment model which, while possible, is not necessarily practicable. Consequently, in the authors' laboratory, an analogous model has been written and implemented using a Microsoft Excel spreadsheet format [75], allowing simple manipulation and interrogation of this complex model.

Alongside the validation of this model presented by Ito et al. [104], where the interaction between erythromycin and clarithromycin with midazolam in humans was accurately predicted, the isolated perfused rat liver model (IPRL) has been used to demonstrate the effect of known mechanism-based inhibitors and NCEs [148]. Using the IPRL allowed ready validation of the predicted impact of TDI *in vivo*. It provides a bridge between the *in vitro* effect and *in vivo* impact for compounds with existing clinical data and NCEs, using pre-clinical species to bring confidence to predicted effect in humans. An adapted five-compartment model was used to fit the data and the *ex vivo* effects, although modest, were well predicted. The interplay between inhibitory potency and pharmacokinetic parameters was demonstrated; a less potent inhibitor with longer PK duration can impact hepatic CYPs to the same extent as a much more potent inhibitor with short duration. This was confirmed for troleandomycin and erythromycin in the IPRL; troleandomycin was found to be over 100-fold more efficient at inhibiting CYP3A2 than erythromycin *in vitro*, yet in the IPRL only a two fold increase over the effect of erythromycin was observed, reflecting the clinical situation for these drugs [149, 150]. This point was further confirmed by a study in cultured human hepatocytes performed in the authors' laboratory [95], where a compound that is a relatively weak time-dependent inhibitor of CYP2C9 was observed to have a dramatic impact on CYP2C9 activity over a prolonged culture period due to the large unbound concentration in the incubation combined with its metabolic stability. However, when the *in vivo* effect was modeled using predicted human PK parameters and dose, only a 5% decrease in *in vivo* CYP2C9 activity was predicted (Fig. 14).

### 5.3 Key Concepts

The literature regarding IVIVE of pharmacokinetics and drug interactions is vast, not only in terms of predicting human pharmacokinetic parameters but also the impact of potential DDIs and the correct use of *in vitro* data to engender the most accurate prediction of clinical DDIs. A summary of the underlying key concepts not already covered for CYP mediated DDIs is presented here.

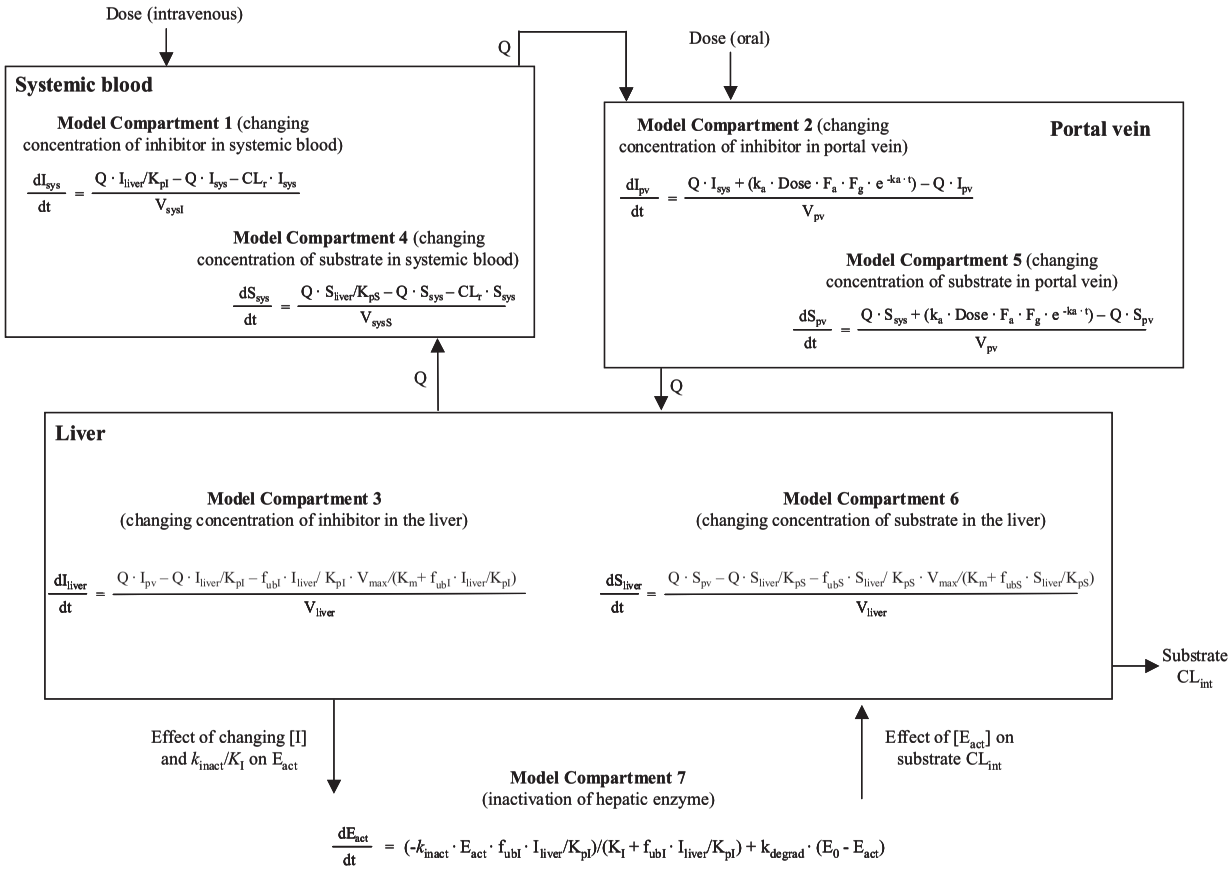


**FIGURE 13** Model present by Ito et al. [7] for predicting effect of TDI. Seven compartments are described representing the changing inhibitor concentration in systemic blood (compartment 1), the changing inhibitor concentration in the portal vein following an oral dose (compartment 2), the changing inhibitor concentration in the liver (compartment 3), the changing substrate concentration in systemic blood (compartment 4), the changing substrate concentration in the portal vein following an oral dose (compartment 5), the changing substrate concentration in the liver (compartment 6), and the inactivation of hepatic enzyme (compartment 7), where I and S are the concentrations of inhibitor and substrate, respectively; $I_{sys}$, $S_{sys}$ and $I_{liver}$, $S_{liver}$ are the concentration in the systemic circulation and liver, respectively; $V_{sysI}$, $V_{sysS}$ represent the volumes of distribution in the central compartment; $K_{pI}$, $K_{pS}$ are the liver-to-blood concentration ratios; $f_{ubS}$, $f_{ubI}$ are the fraction unbound in blood; $I_{pv}$, $S_{pv}$ represent the concentrations in the portal vein; $V_{pv}$ represents the volume of the portal vein; Q is blood flow, $CL_r$ represents renal clearance; $k_a$ is the absorption rate constant; $F_a$ is fraction absorbed from an oral dose; and $F_g$ represents the intestinal availability.

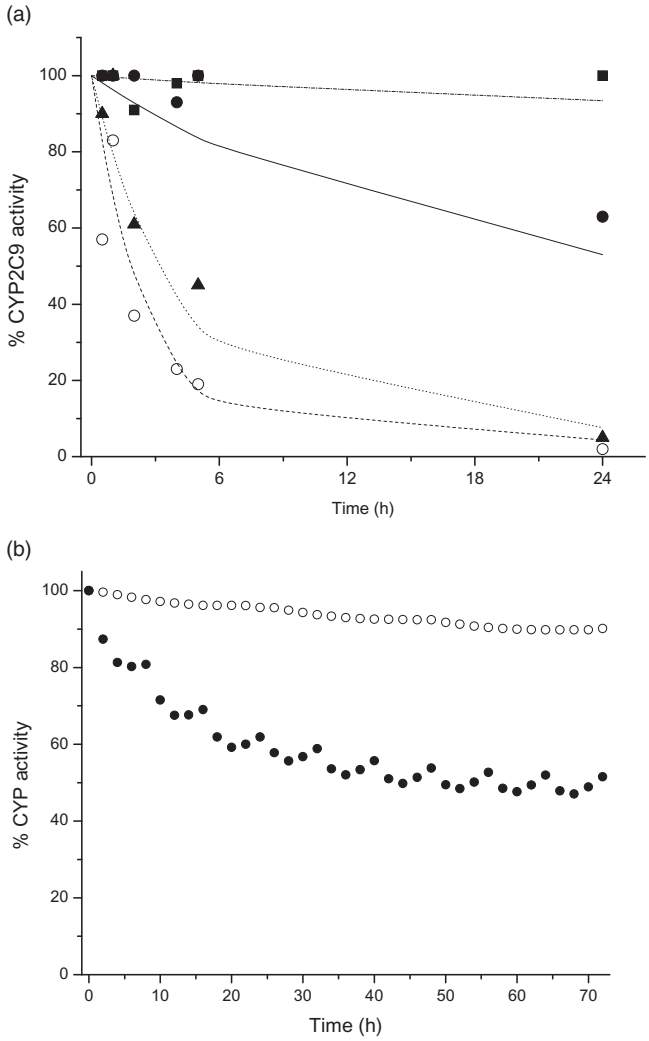


**FIGURE 14** Effect of CYP2C9 inhibition observed *in vitro* and predicted *in vivo*. (a) CYP2C9 activity in cultured human hepatocytes after incubation with AZ1: CYP2C9-dependent diclofenac 4′-hydroxylation after incubation with AZ1 was determined after incubation with 0.1 μM (solid squares), 1 μM (solid circles), 5 μM (solid triangles), and 10 μM (open circles) AZ1. The dash-dot, solid, dotted, and dashed lines indicate nonlinear regression of the 0.1, 1, 5, and 10 μM AZ1 data, respectively. (b) Simulated profiles of CYP activity in human liver after oral dosing with AZ1 and erythromycin, respectively, using the methodology of Ito et al. [104]: CYP2C9 activity in human liver was simulated after oral dosing with AZ1 (70 mg every 24 h for 72 h open circles) and CYP3A4 activity in human liver after oral dosing with erythromycin (500 mg every 8 h for 72 h closed circles) (From Ref. 95.).

Making quantitative predictions of *in vivo* DDI using *in vitro* data requires consideration not only of inhibitor $K_i$ (or $K_I$ and $k_{inact}$ for mechanism-based inhibition), but also $f_{uinc}$, $f_{u,b}$, $CL$, volume, $fm_{CYP}$, dose, and $k_a$. Use of such factors has resulted in relatively accurate predictions of DDI for reasonably large datasets [75, 88, 139, 147], with some exceptions (possibly compounds with an active component to their

hepatic uptake/efflux—clearly there is more to learn in this area). For detailed predictions of individual DDIs, computer-modeling software should be used to simulate not only the impact of the scenarios detailed earlier, but also the changing drug concentrations with time as a result of the individual PK parameters; this allows a detailed interrogation of the interaction. If a worst-case scenario is to be provided, the maximum unbound inhibitor concentration at the enzyme site should be used. For this, knowledge of the extent to which the inhibitor concentrates in the hepatocyte is important (if no information is available, 10× the maximum unbound blood concentration has been proposed [7]). Additionally, assuming that elimination of substrate is by a single hepatic CYP predicts a worse-case scenario for the inhibited pathway *in vivo*. These approaches are of utility in a discovery setting when selecting compounds for further progression, but it can be appreciated that it generally leads to an overestimation of the inhibitory effect, especially when combined with the assumption of complete and instantaneous absorption of the inhibitor from an oral dose.

Refinement and improved predictivity is afforded by including the contribution of the inhibited CYP to the total clearance of the substrate ($fm_{CYP}$) and more realistic estimations of the fraction and rate of absorption for the inhibitor, essentially by estimating not only the concentration of inhibitor in the liver more accurately but also its impact on the metabolic pathway of the substrate. Information on the *in vivo* $fm_{CYP}$ for specific substrates can be found in the literature [68, 138, 151]. As demonstrated by the clinical cases of DDI presented previously, CYP inhibition has greatest impact when a substrate is predominantly cleared by the inhibited pathway; consequently, multiple pathways for drug elimination reduce the chance of clinically significant DDI. However, consideration must be given to the target patient population as, for example, renal impairment can dramatically affect the elimination pathway of a drug.

Recently, an *in vitro* approach aimed at overcoming the difficultly of assessing inhibitor concentration at the enzyme was presented [152]. It involves using human hepatocytes suspended in human plasma to generate $K_i$ without the need for correction for free concentration at the active site (with the caveat that active transport in hepatocytes *in vitro* may not fully represent that *in vivo*). Using this method, the authors claim accurate prediction of several clinically relevant CYP DDIs by including the CYP phenotypic metabolism for each substrate. This method requires a large range of inhibitor concentrations in the *in vitro* incubations and relies heavily on acute analytical sensitivity. A limitation of this and other methods utilizing $fm_{CYP}$ is the level of understanding required for each substrate under investigation (optimal concentration range for determining $K_i$ and an accurate estimation of the contribution of different CYPs to the substrates' *in vivo* metabolism). This method is one approach to minimizing the difficulty of predicting [I] at the active site.

Other points of consideration for IVIVE of DDI include using appropriate substrate–inhibitor combinations *in vitro* so the *in vivo* interaction is correctly assessed; this further emphasizes the importance of using clinically relevant probe substrates in the *in vitro* determinations. Consideration must also be given to the metabolites of an inhibitor, which can often have a significant inhibitory effect in their own right. If they persist in the systemic circulation or are actively concentrated in the liver, their influence can lead to underprediction of the magnitude of the interaction if they are discounted.

There are several pharmacokinetic parameters to be considered for both substrate and inhibitor when making predictions of DDIs. These considerations essentially require estimating levels of drug within the body, be it in the circulating plasma, liver, portal vein, or gastrointestinal tract. Dose and clearance are the primary determinants for these parameters. For drugs in preclinical development, accurate prediction of the relative impact of DDI relies on reasonable assessment of dose level, frequency, and route. It follows that by minimizing dose the potential of DDI will also be minimized. The route of administration of the victim drug must also be considered. A compound with high (intestinal and hepatic) first-pass extraction will behave very differently in the presence of an inhibitor when administered orally as opposed to intravenously. For a drug with a high extraction ratio, the $C_{max}$ will be significantly increased in the presence of an inhibitor but the elimination half-life will be quite similar in the presence and absence of inhibitor. Conversely, for a compound with a low extraction ratio, the same inhibitor will have limited effect on $C_{max}$ following an oral dose but the elimination half-life may be considerably larger; this will also be the case following an IV dose [130].

Predictions of CYP inhibition made in isolation of other contributing factors, albeit sometimes unknown, may only result in an approximation of *in vivo* effect. The ultimate goal is to be able to predict *in vivo* disposition by encompassing not only multiple mechanisms of CYP (and indeed other enzyme) inhibition but also the potential inductive capacity and transporter interactions for the compound under evaluation. Many drugs exhibit numerous interactive attributes, one such example is the antiviral protease inhibitor ritonavir. *In vitro* ritonavir has been demonstrated to induce CYP3A4 expression through activation of the Pregnane X receptor [153]; it is also a potent mechanism-based inhibitor of CYP3A4 [154] and a substrate and inducer of the P-glycoprotein transporter [155]. Despite accurate *in vitro* assessment of its inhibitory capacity, predictions of the *in vivo* effect of ritonavir are often erroneous due to varying contribution of all these factors and also are dependent on the time course of drug exposure. Initially, ritonavir exerts an inhibitory effect on CYP3A4 mediated metabolism *in vivo* but with prolonged exposure the inductive capacity of the drug predominates. For example, short-term low-dose administration of ritonavir produces a large and significant impairment of triazolam clearance and enhancement of clinical effects [156] with the *AUC* of triazolam increasing from 13.6 to 553 ng/mL · h following 1 day of ritonavir treatment but following 10 days of treatment falling to 287 ng/mL·h [157]. Accurate prediction of such situations becomes difficult. A unifying model allowing all factors to be encompassed, although complex, would be beneficial.

As discussed, a comprehensive, integrated, and systematic approach is required for IVIVE of CYP mediated DDIs. To this end there exists Simcyp® (http://www.simcyp.com), a population-based prediction software tool that offers such an evaluation for CYP mediated DDIs. This software incorporates physiological, genetic, and epidemiological information, which, together with *in vitro* data, facilitates the modeling and simulation of the time course and fate of drugs in representative virtual patient populations. Simcyp uses information from routine *in vitro* studies generated in most drug discovery departments and with this information predicts individual pharmacokinetic parameters (such as $F_{abs}$, $CL$, and $V_{ss}$) and simulates CYP mediated drug interactions and the likely associated population variability. This allows prediction of outcomes in relevant patient populations, identifying those

individuals at most risk from a DDI, not just a single value in an average individual. The software has the potential to simulate the effect of combined CYP induction and inhibition and can also be applied to DDIs involving time-dependent inhibition. Simcyp may assist in the evaluation and optimization of candidate drugs, in the estimation of early human doses and exposures in the clinic, and the prioritization and planning of suitable *in vitro* and clinical interaction studies in the development phase. Tools such as Simcyp are attracting a lot of interest in the pharmaceutical industry and may facilitate a more aligned strategy and as such are worthy of further investment and validation.

### 5.4 FDA Guidelines

The FDA defines a significant DDI as a two fold (or more) increase in the *AUC* of the substrate drug in the presence of an inhibitor. The two fold threshold is given as inter-individual variation and polymorphisms in CYP expression within the population could account for changes less than this [158]. If a risk of DDI is predicted from *in vitro* data, a clinical study evaluating the risk and a thorough investigation of the mechanism of CYP inhibition (along with understanding of the potential coadministered drugs in the target patient population) may make it possible to launch a candidate drug, providing it carries a product label stating possible risk of DDI. For example, if a drug has been determined to be a "strong" inhibitor of CYP3A, a warning about an interaction with "sensitive CYP3A substrates" and CYP3A substrates with narrow therapeutic range may be required [158].

The FDA has released guidance documents, which reflect current thinking in the field specifically aimed at industry evaluating metabolism-based drug interactions; *Guidance for Industry: Drug Metabolism/Drug Interaction Studies in the Drug Development Process—Studies in Vitro*; and *Guidance for Industry: In Vivo Drug Metabolism/Drug Interaction Studies—Study Design, Data Analysis and Recommendations for Dosing and Labelling* [159]. The *in vitro* studies guide, as well as providing a summary of drug metabolism and relevant concepts, describes *in vitro* techniques for evaluating the potential of metabolism-based interactions, the correlation of *in vitro* and *in vivo* findings, and the timing of *in vitro* studies and subsequent product labeling. It defines the goals of evaluating *in vitro* drug metabolism as: (1) to identify all of the major metabolic pathways that affect the test drug and its metabolites, including the specific enzymes responsible for elimination and the intermediates formed; and (2) to explore and anticipate the effects of the test drug on the metabolism of other drugs and the effects of other drugs on its metabolism.

## 6 CONCLUSION

Undoubtedly, using *in vitro* data to make accurate predictions of clinical DDIs involving inhibition of CYP is possible providing that there is understanding of the various processes involved. The availability of clinical data is central to building predictive models and is required for feedback on their accuracy and further iteration of these algorithms. Misinterpretation of the risk of CYP inhibition can be

costly both in terms of patient safety and pharmaceutical resources; in a drug discovery setting predicting the worst-case scenario is a prudent approach.

In brief, CYP inhibition may/can result in serious clinical adverse reactions, the prediction of which is a maturing science. Accurate predictions are possible and the pitfalls and progress in using *in vitro* inhibition data to project clinical drug–drug interactions have been presented. Clearly, accurate predictions start with the production of high quality *in vitro* data, require rational examination of the kinetics of inhibition, and, finally, require the application of sound pharmacokinetic theory. The approaches discussed here should guide the researcher in implementing this best practice.

## REFERENCES


1. Pirmohamed M, James S, Meakin S, Green C, Scott AK, Walley TJ, Farrar K, Park BK, Breckenridge AM. Adverse drug reactions as cause of admission to hospital: prospective analysis of 18,820 patients. *Br Med J* 2004;329(7456):15–19.
2. Lazarou J, Pomeranz BH, Corey PN. Incidence of adverse drug reactions in hospitalised patients. *JAMA* 1998;279(15):1200–1205.
3. Frankfort SV, Tulner LR, van Campen JPCM, Koks CHW, Beijnen JH. Evaluation of pharmacotherapy in geriatric patients after performing complete geriatric assessment at a diagnostic day clinic. *Clin Drug Invest* 2006;26(3):169–174.
4. Jankel C, Fitterman LK. Epidemiology of drug–drug interactions as a cause of hospital admissions. *Drug Safety* 1993;9:51–59.
5. Howard RL, Avery AJ, Slavenburg S, Royal S, Pipe G, Lucassen P, Pirmohamed M. Which drugs cause preventable admissions to hospital? A systematic review. *Br J Clin Pharmacol* 2007;63(2):136–147.
6. Shah RR. Mechanistic basis of adverse drug reactions: the perils of inappropriate dose schedules. *Expert Opin Drug Safety* 2005;4(1):103–128.
7. Ito K, Iwatsubo T, Kanamutsu S, Ueda K, Suzuki H, Sugiyama Y. Prediction of pharmacokinetic alterations caused by drug–drug interactions: metabolic interaction in the liver. *Pharmacol Rev* 1998;50(3):387–411.
8. Kumar V, Wahlstrom JL, Rock DA, Warren CJ, Gorman LA, Tracy TS. CYP2C9 inhibition: impact of probe selection and pharmacogenetics on *in vitro* inhibition profiles. *Drug Metab Dispos* 2006;34(12):1966–1075.
9. Shou M. Prediction of pharmacokinetics and drug–drug interactions from *in vitro* metabolic data. *Curr Opin Drug Discov Dev* 2005;8(1):66–77.
10. Zimmermann M, Duruz H, Guinand O, Broccard O, Levy P, Lacatis D, Bloch A. Torsades de pointes after treatment with terfenadine and ketoconazole. *Eur Heart J* 1992; 13(7):1002–1003.
11. Backmann KA, Lewis JD. Predicting inhibitory drug–drug interactions and evaluating interaction reports using inhibition constants. *Ann Pharmacother* 2005;39:1064–1072.
12. Wysowsk DK, Bacsanyi J. Cisapride and fatal arrhythmia. *N Engl J Med* 1996; 335(4):290–291.
13. Brown HS, Galetin A, Halifax D, Houston JB. Prediction of *in vivo* drug–drug interactions from *in vitro* data: factors affecting prototypic drug–drug interactions involving CYP2C9, CYP2D6 and CYP3A4. *Clin Pharmacokinet* 2006;45(10):1035–1050.

14. Venkatakrishnan K, von Moltke LL, Obach RS, Greenblatt DJ. Drug metabolism and drug interactions: application and clinical value of *in vitro* models. *Curr Drug Metab* 2003;4:423–459.
15. Lin JH, Lu AYH. Interindividual variability in inhibition and induction of cytochrome P450 enzymes. *Annu Rev Pharmacol Toxicol* 2001;41:535–567.
16. Prentis RA, Lis Y, Walker SR. Pharmaceutical innovation by the seven UK-owned pharmaceutical companies (1964–1985). *Br J Clin Pharmacol* 1988;25:387–396.
17. Schuster D, Laggner C, Langer T. Why drugs fail—a study on side effects in new chemical entities. *Curr Pharm Des* 2005;11:3545–3559.
18. DiMasi JA, Hansen RW, Grabowski HG. The price of innovation: new estimates of drug development costs. *J Health Econ* 2003;22:151–185.
19. Cohen LH, Remley MJ, Raunig D, Vaz ADN. *In vitro* drug interactions of cytochrome P450: an evaluation of fluorogenic to conventional substrates. *Drug Metab Dispos* 2003;31:1005–1015.
20. Wienkers LC, Heath TG. Predicting *in vivo* drug interactions from *in vitro* drug discovery data. *Nat Rev (Drug Disco)* 2005;4:825–833.
21. Donato MT, Gomez-Lechón MJ. Inhibition of P450 enzymes: an *in vitro* approach. *Curr Enzyme Inhib* 2006;2:281–304.
22. Lewis DFV. 57 varieties: the human cytochromes P450. *Pharmacogenomics* 2004; 5(3):305–318.
23. Girault I, Rougier N, Chesné C, Lideraue R, Beaune P, Biéche I, De Waziers I. Simultaneous measurement of 23 isoforms from the human cytochrome P450 families 1 to 3 by quantitative reverse transcriptase-polymerase chain reaction. *Drug Metab Dispos* 2005;33:1803–1810.
24. Tanaka E. Clinically important pharmacokinetic drug–drug interactions: role of cytochrome P450 enzymes. *J Clin Pharm Ther* 1998;23:403–416.
25. Bertz RJ, Granneman GR. Use of *in vitro* and *in vivo* data to estimate the likelihood of metabolic pharmacokinetic interactions. *Clin Pharmacokin* 1997;32(3):210–258.
26. Shimada T, Yamazaki H, Mimura M, Yukiharu I, Guengerich FP. Interindividual variations in human liver cytochrome P450 enzymes involved in oxidation of drugs, carcinogens and toxic chemicals; studies with liver microsomes of 30 Japanese and 30 Caucasians. *J Pharmacol Exp Ther* 1994;270:414–423.
27. Dresser GK, Spence JD, Bailey DG. Pharmacokinetic–pharmacodynamic consequences and clinical relevance of cytochrome P450 inhibition. *Clin Pharmacokinet* 2000; 38(1):41–57.
28. Lewis JD, Bachmann KA. Cytochrome P450 enzymes and drug–drug interactions: an update on the superfamily. *J Pharm Technol* 2006;22:22–31.
29. Batty KT, Davis TME, Ilett KF, Dusci LJ, Langton SR. The effect of ciprofloxacin on theophylline pharmacokinetics in healthy subjects. *Br J Clin Pharmacol* 1995; 39(3):305–311.
30. Granfors MT, Backman JT, Neuvonen M, Neuvonen PJ. Ciprofloxacin greatly increases concentrations and hypotensive effect of tizanidine by inhibiting its cytochrome P450 1A2-mediated presystemic metabolism. *Clin Pharmacol Ther* 2004;76(6):598–606.
31. Hemeryck A, Belpaire FM. Selective serotonin reuptake inhibitors and cytochrome P-450 mediated drug–drug interactions: an update. *Curr Drug Metab* 2002;3:13–37.
32. Becquemont L, Bot MA, Le Riche C, Beaune P. Influence of fluvoxamine on tacrine metabolism *in vitro*: potential implication for hepatotoxicity *in vivo*. *Fundam Clin Pharmacol* 1996;10(2):156–157.

33. Backman JT, Kyrklund C, Neuvonen M, Neuvonen PJ. Gemfibrozil greatly increases plasma concentrations of cerivastatin. *Clin Pharmacol Ther* 2002;72:685–691.
34. Veronese ME, Miners JO, Randles D, Gregov D, Birkett DJ. Validation of the tolbutamide metabolic ratio for population screening with use of sulfaphenazole to produce model phenotypic poor metabolizers. *Clin Pharmacol Ther* 1990;47:403–411.
35. Crussell-Porter LL, Rindone JP, Ford MA, Jaskar DW. Low-dose fluconazole therapy potentiates the hypoprothrombinemic response of warfarin sodium. *Arch Intern Med* 1993;153(1):102–104.
36. Krayenbühl JC, Vozeh S, Kondo-Oestreicher M, Dayer P. Drug–drug interactions of new active substances: mibefradil example. *Eur J Clin Pharmacol* 1999;55:559–565.
37. Boxenbaum H. Cytochrome P450 3A4 *in vivo* ketoconazole competitive inhibition: determination of $K_i$ and dangers associated with high clearance drugs in general. *J Pharm Pharm Sci* 1999;2(2):47–52.
38. Honig PK, Woosley RL, Zamani K, Conner DP, Cantilena LR, Jr. Changes in the pharmacokinetics and electrocardiograph pharmacodynamics of terfenadine with concomitant administration of erythromycin. *Clin Pharmacol Ther* 1992;52(3):231–238.
39. Ishigam M, Uchiyama M, Kondo T, Iwabuchi H, Inoue SI, Takasaki W, Ikeda T, Komai T, Ito K, Sugiyama Y. Inhibition of *in vitro* metabolism of simvastatin by itraconazole in humans and prediction of *in vivo* drug–drug interactions. *Pharm Res* 2001;18(5): 622–631.
40. Corpier CL, Jones PH, Suki WN, Lederer ED, Quinones MA, Schmidt SW, Young JB. Rhabdomyolysis and renal injury with lovastatin use. Report of two cases in cardiac transplant recipients. *JAMA* 1998;260(2):239–241.
41. Schmassmann-Suhijar D, Bullingham R, Gasser R, Schmutz J, Haefeli WE. Rhabdomyolysis due to interaction of simvastatin with mibefradil [letter]. *Lancet* 1998;351: 1929–1930.
42. Kantola T, Kivisto KT, Neuvonen PJ. Erythromycin and verapamil considerably increase serum simvastatin and simvastatin acid concentrations. *Clin Pharmacol Ther* 1998;64(2): 177–182.
43. Michalets EL, Williams CR. Drug interactions with cisapride: clinical implications. *Clin Pharmacokinet* 2000;39(1):49–75.
44. Jacobsen W, Kirchner G, Hallensleben K, Mancinelli L, Deters M, Hackbarth I, Baner K, Benet LZ, Sewing K, Christians U. Small intestinal metabolism of the 3-hydroxy-3-methylglutaryl-coenzyme A reductase inhibitor lovastatin and comparison with pravastatin. *J Pharmacol Exp Ther* 1999;291(1):131–139.
45. Spoendlin M, Peters J, Welker H, Bock A, Thiel G. Pharmacokinetic interaction between oral cyclosporin and mibefradil in stabilized post-renal-transplant patients. *Nephrol Dial Transplant* 1998;13:1787–1791.
46. Olkkola KT, Ahonen J, Neuvonen PJ. The effect of the systemic antimycotics, itraconazole and fluconazole, on the pharmacokinetics and pharmacodynamics of intravenous and oral midazolam. *Anesth Analg* 1996;82(3):511–516.
47. Muck W. Rational assessment of the interaction profile of cerivastatin supports its low propensity for drug interactions. *Drugs* 1998;56(1s):15–23.
48. Muck W, Ochmann K, Rohde G, Unger S, Kuhlmann J. Influence of erythromycin pre- and co-treatment on single-dose pharmacokinetics of the HMG-CoA reductase inhibitor cerivastatin. *Eur J Clin Pharmacol* 1998;53:469–473.
49. Kantola T, Kivisto KT, Neuvonen PJ. Effect of itraconazole on cerivastatin pharmacokinetics. *Eur J Clin Pharmacol* 1999;54:851–855.

50. Mazzu AL, Lasseter KC, Shamblen EC, Agarwal V, Lettieri J, Sundaresen P. Itraconazole alters the pharmacokinetics of atorvastatin to a greater extent than either cerivastatin or pravastatin. *Clin Pharmacol Ther* 2000;68:391–400.
51. Shitara Y, Sugiyama Y. Pharmacokinetic and pharmacodynamic alterations of 3-hydroxy-3-methylglutaryl coenzyme A (HMG-CoA) reductase inhibitors: drug–drug interactions and interindividual differences in transporter and metabolic enzyme functions. *Pharmacol Ther* 2006;112(1):71–105.
52. Shitara Y, Hirano M, Sato H, Sugiyama Y. Gemfibrozil and its glucuronide inhibit the organic anion transporting polypeptide 2 (OATP2/OATP1B1:SLC21A6)-mediated hepatic uptake and CYP2C8-mediated metabolism of cerivastatin: analysis of the mechanism of the clinically relevant drug–drug interaction between cerivastatin and gemfibrozil. *J Pharmacol Exp Ther* 2004;311:228–236.
53. Neuvonen PJ, Mikko Niemi M, Backman JT. Drug interactions with lipid-lowering drugs: mechanisms and clinical relevance. *Clin Pharmacol Ther* 2006;80(6):565–581.
54. Pisani F, Narbone MC, Fazio A, Crisafulli P, Primerano G, D'Agostino AA, Oteri G, Di Perri R. Increased serum carbamazepine levels by viloxazine in epileptic patients. *Epilepsia* 1986;25:482–285.
55. Rotzinger S, Bourin M, Akimoto Y, Coutts RT, Bake GB. Metabolism of some "second"– and "fourth"–generation antidepressants: iprindole, viloxazine, bupropion, mianserin, maprotiline, trazodone, nefazodone, and venlafaxine. *Cell Mol Neurobiol* 1999;19(4): 427–442.
56. Galetin A, Ito K, Halifax D, Houston JB. CYP3A4 substrate selection and substitution in the prediction of potential drug–drug interactions. *J Pharmacol Exp Ther* 2005; 314(1):180–190.
57. Prueksaritanont T, Ma B, Tang C, Meng Y, Assang C, Lu P, Reider PJ, Lin JH, Baillie TA. Metabolic interactions between mibefradil and HMG-CoA reductase inhibitors: an *in vitro* investigation with human liver preparations. *Br J Clin Pharmacol* 1999;47(3):291–298.
58. Wandel C, Kim RB, Guengerich PF, Wood AJJ. Mibefradil is a P-glycoprotein substrate and a potent inhibitor of P-glycoprotein and CYP3A4 *in vitro*. *Drug Metab Dispos* 2000;28:895–898.
59. Reynolds KS. Drug interactions: regulatory perspective. In Piscitelle S, Rodvold K, Eds. *Drug Interactions in Infectious Diseases*, 2nd eds. Totowa, NJ: Humana Press; 2005.
60. Venkatakrishnan K, von Moltke LL, Greenblatt DJ. Effects of the antifungal agents on oxidative drug metabolism: clinical relevance. *Clin Pharmacokinet* 2000;8(2): 111–180.
61. Granfors MT, Backman JT, Jouko L, Neuvonen PJ. Tizanidine is mainly metabolized by cytochrome P450 1A2 *in vitro*. *Br J Clin Pharmacol* 2004;57(3):349–353.
62. Ha HR, Chen J, Freiburghaus AU, Follath F. Metabolism of theophylline by cDNA-expressed human cytochromes P-450. *Br J Clin Pharmacol* 1995;39(3):321–326.
63. Ilett KF, Castleden WM, Vandongen YK, Stacey MC, Butler MA, Kadlubar FF. Acetylation phenotype and cytochrome P450IA2 phenotype are unlikely to be associated with peripheral arterial disease. *Clin Pharmacol Ther* 1993;54(3):317–322.
64. Dorado P, Berecz R, Penas-Lledo EM, Caceres MC, Llerena A. Clinical implications of CYP2D6 genetic polymorphism during treatment with antipsychotic drugs. *Curr Drug Targets* 2006;7(12):1671–1680.
65. Preskorn SH, Harvey AT. Cytochrome P450 enzymes and psychopharmacology. In Bloom FE, Kupfer DJ, Eds. *Psychopharmacology—The Fourth Generation of Progress*, 4threv ed. Philadelphia: Raven Press; 1995.

66. Kenworthy KE, Clarke SE, Andrews J, Houston JB. Multisite kinetic models for CYP3A4: simultaneous activation and inhibition of diazepam and testosterone metabolism. *Drug Metab Dispos* 2001;29(12):1644–1651.
67. Hutzler JM, Hauer MJ, Tracy TS. Dapsone activation of CYP2C9-mediated metabolism: evidence for activation of multiple substrates and a two-site model. *Drug Metab Dispos* 2001;29(7):1029–1034.
68. McGinnity DF, Tucker J, Trigg S, Riley RJ. Prediction of CYP2C9-mediated drug–drug interactions: a comparison using data from recombinant enzymes and human hepatocytes. *Drug Metab Dispos* 2005;33(11):1700–1707.
69. Egnell A-C, Eriksson C, Albertson N, Houston B, Boyer S. Generation and evaluation of a CYP2C9 heteroactivation pharmacophore. *J Pharmacol Exp Ther* 2003;307:878–887.
70. Egnell A-C, Houston B, Boyer S. *In vivo* CYP3A4 heteroactivation is a possible mechanism for the drug interaction between felbamate and carbamazepine. *J Pharmacol Exp Ther* 2003;305(3):1251–1262.
71. Zhang Z, Wong YN. Enzyme kinetics for clinically relevant CYP inhibition. *Curr Drug Metab* 2005;6:241–257.
72. Tracy TS, Hummell MA. Modelling kinetic data from *in vitro* drug metabolising enzyme experiments. *Drug Metab Rev* 2004;26(2):231–242.
73. Houston JB, Kenworthy KE, Galetin A. Typical and atypical enzyme kinetics. In Lee JS, Obach RS, Fisher MB, Eds. *Drug Metabolising Enzymes, Cytochrome P450 and Other Enzymes in Drug Discovery and Development*. New York: Marcel Dekker; 2003, pp 211–254.
74. Hollenberg PF. Characteristics and common properties of inhibitors, inducers, and activators of CYP enzymes. *Drug Metab Rev* 2002;34(1–2):17–35.
75. Riley RJ, Grime K, Weaver R. Time-dependent CYP inhibition. *Expert Opin Drug Metab Toxicol* 2007;3(1):51–66.
76. Zhou S, Yung Chan S, Cher Goh B, Chan E, Duan W, Huang M, McLeod HL. Mechanism-based inhibition of cytochrome P450 3A4 by therapeutic drugs. *Clin Pharmacokinet* 2005;44(3):279–304.
77. Silverman RB. Mechanism-based enzyme inactivation. In *Chemistry and Enzymology*, Vol 1. Boca Raton, FL: CRC Press; 1988, pp 3–30.
78. Waley SG. Kinetics of suicide substrates: practical procedures for determining parameters. *Biochem J* 1985;227:843–849.
79. Ghanbari F, Rowland-Yeo K, Bloomer JC, Clarke SE, Lennard MS, Tucker GT, Rostami-Hodjegan A. A critical evaluation of the experimental design of studies of mechanism-based enzyme inhibition, with implications for *in vitro–in vivo* extrapolation. *Curr Drug Metab* 2006;7:315–334.
80. Ito K, Brown HS, Houston JB. Database analyses for the prediction of *in vivo* drug–drug interactions from *in vitro* data. *Br J Clin Pharmacol* 2004;57(4):473–486.
81. Martignoni M, Groothuis GMM, de Kanter R. Species differences between mouse, rat, dog, monkey and human CYP-mediated drug metabolism, inhibition and induction. *Expert Opin Drug Metab Toxicol* 2006;2(6):875–894.
82. Gill HJ, Tingle MD, Park BK. N-Hydroxylation of dapsone by multiple enzymes of cytochrome P450: implications for inhibition of hemotoxicity. *Br J Clin Pharmacol* 1995;40(6):531–538.
83. Soars MG, Grime K, Riely RJ. Comparative analysis of substrate and inhibitor interactions with CYP3A4 and CYP3A5. *Xenobiotica* 2006;36(4):287–299.
84. Tang W, Wang WW, Lu AYH. Utility of recombinant cytochrome P450 enzymes: a drug metabolism perspective. *Curr Drug Metab* 2005;6:503–517.

85. Proctor NJ, Tucker GT, Rostami-Hodjegan A. Predicting drug clearance from recombinantly expressed CYPs: intersystem extrapolation factors. *Xenobiotica* 2004;34(2):151–178.
86. McGinnity DF, Riley RJ. Predicting drug pharmacokinetics in humans from *in vitro* metabolism studies. *Biochem Soc Transact* 2001;29(2):135–139.
87. Gomez-Lechón MJ, Donato MT, Castell JV, Jover R. Human hepatocytes in primary culture: the choice to investigated drug metabolism in man. *Curr Drug Metab* 2004;5:443–462.
88. Grime K, Riley RJ. The impact of *in vitro* binding on *in vitro–in vivo* extrapolations, projections of metabolic clearance and clinical drug–drug interactions. *Curr Drug Metab* 2006;7(3):251–264.
89. Griffin SJ, Houston JB. Comparison of fresh and cryopreserved rat hepatocyte suspensions for the prediction of *in vitro* intrinsic clearance. *Drug Metab Dispos* 2004;32:552–558.
90. Houston JB, Carlile DJ. Prediction of hepatic clearance from microsomes, hepatocytes and liver slices. *Drug Metab Rev* 1997;29:891–922.
91. Dickins M. Induction of cytochromes P450. *Curr Top Med Chem* 2004;4(16): 1745–1766.
92. LeCluyse EL. Human hepatocyte culture systems for the *in vitro* evaluation of cytochrome P450 expression and regulation. *Eur J Pharm Sci* 2001;13(4):343–368.
93. Oleson FB, Berman CL, Li AP. An evaluation of the P450 inhibition and induction potential of daptomycin in primary human hepatocytes. *Chem Biol Interact* 2004; 150:137–147.
94. Li AP, Lu C, Brent JA, Pham C, Fackett A, Ruegg CE, Silber PM. Cryopreserved human hepatocytes: characterization of drug-metabolizing enzyme activities and applications in higher throughput screening assays for hepatotoxicity, metabolic stability, and drug–drug interaction potential. *Chem Biol Interact* 1999;121:17–35.
95. McGinnity DF, Berry AJ, Kenny JR, Grime K, Riley RJ. Evaluation of time-dependent cytochrome P450 inhibition using cultured human hepatocytes. *Drug Metab Dispos* 2006;34:1291–1300.
96. Zhao P, Kunze KL, Lee CA. Evaluation of time-dependent inactivation of CYP3A in cryopreserved human hepatocytes. *Drug Metab Dispos* 2005;33:853–861.
97. Soars MG, McGinnity DF, Grime K, Riely RJ. The pivotal role of hepatocytes in drug discovery. *Chem-Biol Interact* 2007;168(1):2–15.
98. Atkinson A, Kenny JR, Grime K. Automated assessment of time-dependent inhibition of human cytochrome P450 enzymes using liquid chromatography-tandem mass spectrometry analysis. *Drug Metab Dispos* 2005;33:1637–1647.
99. Lopez-Garcia MP, Dansette PM, Mansuy D. Thiophene derivatives as new mechanism-based inhibitors of cytochromes P-450: inactivation of yeast-expressed human liver cytochrome P-450 2C9 by tienilic acid. *Biochemistry* 1994;33(1):166–175.
100. Mayhew BS, Jones DR, Hall SD. An *in vitro* model for predicting *in vivo* inhibition of cytochrome P450 3A4 by metabolic intermediate complex formation. *Drug Metab Dispos* 2000;28:1031–1037.
101. Chan WK, Delucchi AB. Resveratrol, a red wine constituent, is a mechanism-based inactivator of cytochrome P450 3A4. *Life Sci* 2000;67(25):3103–3112.
102. Kanamitsu S, Ito K, Green CE, Tyson CA, Shimada N, Sugiyama Y. Prediction of *in vivo* interaction between triazolam and erythromycin based on *in vitro* studies using human liver microsomes and recombinant human CYP3A4. *Pharm Res* 2000;17(4):419–426.

103. Yamano K, Yamamoto K, Katashima M, Kotaki H, Takedomi S, Matsuo H, Ohtani H, Sawada Y, Iga T. Prediction of midazolam-CYP3A inhibitors interaction in the human liver from *in vivo*/*in vitro* absorption, distribution, and metabolism data. *Drug Metab Dispos* 2001;29(4):443–452.
104. Ito K, Ogihara K, Kanamitsu S, Itoh T. Prediction of the *in vivo* interaction between midazolam and macrolides based on *in vitro* studies using human liver microsomes. *Drug Metab Dispos* 2003;31(7):945–954.
105. Dai R, Wei X, Luo G, Sinz M, Marathe P. Metabolism-dependent P450 3A4 inactivation with multiple substrates. Abstract from 12th North American ISSX Meeting, Providence, RI. *Drug Metab Rev* 2003;35:341.
106. McConn DJ, Zhao Z. Integrating *in vitro* kinetic data from compounds exhibiting induction, reversible inhibition and mechanism-based inactivation: *in vitro* study design. *Curr Drug Metab* 2004;5:141–146.
107. Boobis AR, McKillop D, Robinson DT, Adams DA, McCormick DJ. Interlaboratory comparison of the assessment of P450 activities in human hepatic microsomal samples. *Xenobiotica* 1998;28(5):493–506.
108. Obach RS. Prediction of human clearance of twenty-nine drugs from hepatic microsomal intrinsic clearance data: an examination of *in vitro* half-life approach and nonspecific binding to microsomes. *Drug Metab Dispos* 1999;27:1350–1359.
109. Riley RJ. The potential pharmacological and toxicological impact of P450 screening. *Curr Opin Drug Discov Dev* 2001;4:45–54.
110. Austin RP, Barton P, Mohmed S, Riley RJ. The binding of drugs to hepatocytes and its relastionship to physicochemical properties. *Drug Metab Dispos* 2005;33(3):419–425.
111. Austin RP, Barton P, Cockroft SL, Wenlock MC, Riley RJ. The influence of nonspecific microsomal binding on apparent intrinsic clearance, and its prediction from physicochemical properties. *Drug Metab Dispos* 2002;30(12):1497–1503.
112. Riley RJ, Grime K. Metabolic screening *in vitro*: metabolic stability, CYP inhibition and induction. *Drug Discov Today* 2004;1(4):365–372.
113. Tang C, Shou M, Rodrigues AD. Substrate-dependent effect of acetronitrile on human liver microsomal cytochrome P450 2C9 (CYP2C9) activity. *Drug Metab Dispos* 2000;28:567–572.
114. Baranczewski P, Stanczak A, Sundberg K, Svensson R, Wallin A, Jansson J, Garberg P, Postlind H. Introduction to *in vitro* estimation of metabolic stability and drug interactions of new chemical entities in drug discovery and development. *Pharmacol Rep* 2006;58:453–472.
115. Walsky RL, Obach RS. Validated assays for human cytochrome P450 activities. *Drug Metab Dispos* 2004;32:647–660.
116. Weaver R, Grime K, Beatie IG, Riley RJ. Cytochrome P450 inhibition using recombinant proteins and mass spectrometry/multiple reaction monitoring technology in a cassette incubation. *Drug Metab Dispos* 2003;31:955–966.
117. Tucker GT, Houston JB, Huang S. Optimising drug development: strategies to assess drug metabolism/transport interaction potential—towards a consensus. *Pharm Res* 2001;18(8):1071–1080.
118. O'Donnell CJ, Grime K, Courtney P, Slee D, Riley RJ. The development of a cocktail CYP2B6, CYP2C8 and CYP3A5 inhibition assay and a preliminary assessment of utiliy in a drug discovery setting. *Drug Metab Dispos* 2007;35(3):381–385.
119. Moody GC, Griffin SJ, Mather AN, McGinnity DF, Riley RJ. Fully automated analysis of activities catalysed by the major human liver cytochrome P450 (CYP) enzymes: assessment of human CYP inhibition potential. *Xenobiotica* 1999;29(1):53–75.

120. US Food and Drug Administration, Center for Drug Evaluation and Research website. Drug Development And Drug Interactions information page. http://www.fda.gov/cder/drug/drugInteractions/.
121. Bu H-Z, Magis L, Knuth K, Teitelbaum P. High-throughput cytochrome P450 (CYP) inhibition screening via a cassette probe-dosing strategy. VI. Simultaneous evaluation of inhibition potential of drugs on human hepatic isozymes CYP2A6, 3A4, 2C9, 2D6 and 2E1. *Rapid Commun Mass Spectrom* 2001;15(10):741–748.
122. Kenworthy KE, Bloomer JC, Clarke SE, Houston JB. CYP3A4 drug interactions: correlation of 10 *in vitro* probe substrates. *Br J Clin Pharmacol* 1999;48(5):716–727.
123. Shou M, Lin Y, Lu P, Tang C, Mei Q, Cui D, Tang W, Ngui JS, Lin CC, Singh R, Wong BK, Yergey JA, Lin JH, Pearson PG, Baillie TA, Rodrigues AD, Rushmore TH. Enzyme kinetics of cytochrome P450-mediated reactions. *Curr Drug Metab* 2001;2(1):17–36.
124. Ueng YF, Kuwabara T, Chun YJ, Guengerich FP. Cooperativity in oxidations catalyzed by cytochrome P450 3A4. *Biochemistry* 1997;36(2):370–381.
125. Zhang Z, Li Y, Shou M, Zhang Y, Ngui JS, Stearns RA, Evans DC, Baillie TA, Tang W. Influence of different recombinant systems on the cooperativity exhibited by cytochrome P4503A4. *Xenobiotica* 2004;34(5):473–486.
126. Janiszewski JS, Rogers KJ, Whalen KM, Cole MJ, Liston TE, Duchosiav E, Fouda HG. A high-capacity LC-MS system for the bioanalysis of samples generated from plate based metabolic screening. *Anal Chem* 2001;73(7):1495–1501.
127. DiMarco A, Marcucci I, Verdirame M, Perez J, Sanchez M, Pelaez F, Chaudhary A, Laufer R. Development and validation of a high throughput radiometric CYP3A4 inhibition assay using tritiated testosterone. *Drug Metab Dispos* 2005;33:349–358.
128. Jensen BF, Vind C, Padkjr SB, Brockhoff PB, Refsgaard HHF. *In silico* prediction of cytochrome P450 2D6 and 3A4 inhibition using Gaussian kernel weighted k-nearest neighbor and extended connectivity fingerprints, including structural fragment analysis of inhibitors versus noninhibitors. *J Med Chem* 2007;50(3):501–511.
129. Refsgaard HHF, Jensen BF, Christensen IT, Hagen N, Brockhoff PB. *In silico* prediction of cytochrome P450 inhibitors. *Drug Dev Res* 2006;67:417–429.
130. Bachmann KA. Inhibition constants, inhibitor concentrations and the prediction of inhibitory drug drug interactions: pitfalls, progress and promise. *Curr Drug Metab* 2006;7:1–14.
131. Mannervik B. Regression analysis in design and analysis. In Purich DL, Ed. *Contemporary Enzyme Kinetics and Mechanism*, 2nd ed. San Diego, CA: Academic Press; 1996.
132. Kalgutkar AS, Gardner I, Obach RS, Shaffer CL, Callegari E, Henne KR, Mutlib AE, Dalvie DK, Lee JS, Nakai Y, O'Donnell JP, Boer J, Harriman SP. A comprehensive listing of bioactivation pathways of organic functional groups. *Curr Drug Metab* 2005; 6(3):161–225.
133. Li AP. A review of the common properties of drugs with idiosyncratic hepatotoxicity and the "multiple determinant hypothesis" for the manifestation of idiosyncratic drug toxicity. *Chem Biol Interact* 2002;142(1–2):7–23.
134. Segel IH. Simple inhibition systems. In *Enzyme Kinetics: Behaviour and Analysis of Rapid Equilibrium and Steady-State Enzyme Systems*. Hoboken, NJ: Wiley; 1975.
135. Wilkinson GR, Shand DG. A physiological approach to hepatic drug clearance. *Clin Pharmacol Ther* 1975;18:377–390.
136. Rostami-Hodjegan A, Tucker G. *In silico* simulations to assess the *in vivo* consequences of *in vitro* metabolic drug–drug interactions. *Drug Discov Today* 2004;1(4):441–448.
137. Houston JB, Galetin A. Progress towards prediction of human pharmacokinetic parameters from *in vitro* technologies. *Drug Metab Rev* 2003;35(4):393–415.

138. Ito K, Halifax D, Obach RS, Houston JB. Impact of parallel pathways of drug elimination and multiple cytochrome P450 involvement on drug–drug interactions: CYP2D6 paradigm. *Drug Metab Dispos* 2005;33(6):837–844.
139. Obach RS, Walsky RL, Venkatakrishnan K, Gaman EA, Houston JB, Termaine LM. The utility of *in vitro* cytochrome P450 inhibition data in the prediction of drug–drug interactions. *J Pharmacol Exp Ther* 2006;316:336–348.
140. Kato M, Tachibana T, Ito K, Sugiyama Y. Evaluation of methods for predicting drug–drug interactions by Monte Carlo simulation. *Drug Metab Pharmacokinet* 2003;18(2): 121–127.
141. Rowland M, Tozer TN. Interacting drugs. In Rowland M, Tozer TN, Eds. *Clinical Pharmacokinetics: Concepts and Applications*. Philadelphia: Williams & Wilkins; 1995, pp 267–289.
142. Riley RJ, McGinnity DF, Austin RP. A unified model for predicting human hepatic, metabolic clearance from *in vitro* intrinsic clearance data in hepatocytes and microsomes. *Drug Metab Dispos* 2005;33(9):1304–1311.
143. Brown HS, Ito K, Galetin A, Houston JB. Prediction of *in vivo* drug–drug interactions from *in vitro* data: impact of incorporating parallel pathways of drug elimination and inhibitor absorption rate constant. *Br J Clin Pharmacol* 2005;60(5):508–518.
144. Blanchard N, Richert L, Coassolo P, Lave T. Qualitative and quantitative assessment of drug–drug interaction potential in man, based on $K_i$, $IC_{50}$ and inhibitor concentration. *Curr Drug Metab* 2006;5(2):147–156.
145. Wang YH, Jones DR, Hall SD. Prediction of cytochrome P450 3A inhibition by verapamil enantiomers and their metabolites. *Drug Metab Dispos* 2004;32(2):259–266.
146. Ernest CS, Hall SD, Jones DR. Mechanism-based inactivation of CYP3A by HIV protease inhibitors. *J Pharmacol Exp Ther* 2005;312(2):583–591.
147. Obach RS, Walsky RL, Venkatakrishnan K. Mechanism based inactivation of human cytochrome P450 enzymes and the prediction of drug–drug interactions. *Drug Metab Dispos* 2007;35(2):246–255.
148. Kenny JR, Grime K. Pharmacokinetic consequences of time-dependent inhibition using the isolated perfused rat liver model. *Xenobiotica* 2006;36(5):351–365.
149. Warot D, Bergougnan L, Lamiable D, Berlin I, Bensimon G, Danjou P, Puech AJ. TroleandomycIn–triazolam interaction in healthy volunteers: pharmacokinetic and psychometric evaluation. *Eur J Clin Pharmacol* 1987;32:389–393.
150. Phillips JP, Antal EJ, Smith RB. A pharmacokinetic drug interaction between erythromycin and triazolam. *J Clin Psychopharmacol* 1986;6:297–299.
151. Galetin A, Burt H, Gibbons L, Houston JB. Prediction of time-dependent CYP3A4 drug–drug interactions: impact of enzyme degradation, parallel elimination pathways, and intestinal inhibition. *Drug Metab Dispos* 2006;34(1):166–175.
152. Lu C, Miwa GT, Prakash SR, Gan L, Balani SK. A novel method for the prediction of drug–drug interactions in humans based on *in vitro* CYP phenotypic data. *Drug Metab Dispos* 2007;35(1):79–85.
153. Luo G, Cunningham M, Kim S, Burn T, Lin J, Sinz M, Hamilton G, Rizzo C, Jolley S, Gilbert D, Downey A, Mudra D, Graham R, Carroll K, Xie J, Madan A, Parkinson A, Christ D, Selling B, Lecluyse E, Gan L. CYP3A4 induction by drugs: correlation between a pregnane X receptor reporter gene assay and CYP3A4 expression in human hepatocytes. *Drug Metab Dispos* 2002;30(7):795–804.
154. von Moltke LL, Durol ALB, Duan SX, Greenblatt DJ. Potent mechanism-based inhibition of human CYP3A *in vitro* by amprenavir and ritonavir: comparison with ketoconazole. *Eur J Clin Pharmacol* 2000;56(3):259–261.

155. Perloff MD, Stoermer E, von Moltke LL, Greenblatt DJ. Rapid assessment of P-glycoprotein inhibition and induction *in vitro*. *Pharm Res* 2003;20(8):1177–1183.
156. Greenblatt DJ, von Moltke LL, Harmatz JS, Durol ALB, Daily JP, Graf JA, Mertzanis P, Hoffman JL, Shader RI. Differential impairment of triazolam and zolpidem clearance by ritonavir. *J Acquir Immune Defic Syndr Hum Retrovirol* 2000;24(2):129–136.
157. Culm-Merdek KE, von Moltke LL, Gan L, Horan KA, Reynolds R, Harmatz JS, Court MH, Greenblatt DJ. Effect of extended exposure to grapefruit juice on cytochrome P450 3A activity in humans: comparison with ritonavir. *Clin Pharmacol Ther* 2006;79(3):243–254.
158. US Food and Drug Administration, Center for Drug Evaluation and Research website. http://www.fda.gov/CDER.
159. US Food and Drug Administration, Center for Drug Evaluation and Research website. Guidance web page. http://www.fda.gov/CDER/guidance.

## APPENDIX A

From Fig. 1 it can be seen that

$$d[\mathrm{ES}]/dt = k_1 \cdot [\mathrm{E}] \cdot [\mathrm{S}] - k_{-1} \cdot [\mathrm{ES}] - k_2 \cdot [\mathrm{ES}] \tag{A.1}$$

But at steady state, $d[\mathrm{ES}]/dt = 0$; therefore,

$$k_1 \cdot [\mathrm{E}] \cdot [\mathrm{S}] = k_{-1} \cdot [\mathrm{ES}] + k_2 \cdot [\mathrm{ES}] \tag{A.2}$$

The total enzyme concentration is

$$[\mathrm{E}]_0 = [\mathrm{E}] + [\mathrm{ES}] \tag{A.3}$$

and the total substrate concentration is

$$[\mathrm{S}]_0 = [\mathrm{S}] + [\mathrm{ES}] \tag{A.4}$$

and therefore

$$[\mathrm{E}] = [\mathrm{E}]_0 - [\mathrm{ES}] \tag{A.5}$$

and

$$[\mathrm{S}] = [\mathrm{S}]_0 - [\mathrm{ES}] \tag{A.6}$$

Therefore, substituting Eqs. A.5 and A.6 into Eq. A.2 gives

$$k_1 \cdot [\mathrm{S}]([\mathrm{E}]_0 - [\mathrm{ES}]) = (k_{-1} + k_2) \cdot [\mathrm{ES}] \tag{A.7}$$

Dividing through by [ES] gives

$$\frac{k_1 \cdot [\mathrm{E}]_0 \cdot [\mathrm{S}]}{[\mathrm{ES}]} - k_1 \cdot [\mathrm{S}] = k_{-1} + k_2 \tag{A.8}$$

and therefore

$$k_{-1} + k_2 + k_1 \cdot [\mathrm{S}] = k_1 \cdot \frac{[\mathrm{E}]_0 \cdot [\mathrm{S}]}{[\mathrm{ES}]} \tag{A.9}$$

and

$$[\mathrm{ES}] = \frac{k_1 \cdot [\mathrm{E}]_0 \cdot [\mathrm{S}]}{k_1 \cdot [\mathrm{S}] + k_{-1} + k_2} \tag{A.10}$$

Dividing through by $k_1$ gives

$$[\mathrm{ES}] = \frac{[\mathrm{E}]_0 \cdot [\mathrm{S}]}{[\mathrm{S}] + (k_{-1} + k_2/k_1)} \tag{A.11}$$

since from the reaction scheme in Fig. 1 the overall dissociation rate constant for enzyme–substrate complex, $[\mathrm{ES}] = (k_{-1} + k_2)/k_1$. This is referred to as the Michaelis–Menten constant $K_\mathrm{m}$. Therefore, Eq. A.11 becomes

$$[\mathrm{ES}] = \frac{[\mathrm{E}]_0 \cdot [\mathrm{S}]}{[\mathrm{S}] + K_\mathrm{m}} \tag{A.12}$$

Referring again to the reaction scheme in Fig. 1, the rate of reaction, $v$, or product formation, $d[\mathrm{P}]/dt = k_2 \cdot [\mathrm{ES}]$. Substituting this into Eq. A.12 gives

$$\frac{d[\mathrm{P}]}{dt} = v = \frac{k_2 \cdot [\mathrm{E}]_0 \cdot [\mathrm{S}]}{[\mathrm{S}] + K_\mathrm{m}} \tag{A.13}$$

where $[\mathrm{S}] >> K_\mathrm{m}$ and $v$ tends to $k_2 \cdot [\mathrm{E}]_0$ such that the maximal reaction rate can be described as $V_{\max} = k_2 \cdot [\mathrm{E}]_0$. Equation A.13 therefore becomes

$$v = \frac{V_{\max} \cdot [\mathrm{S}]}{K_\mathrm{m} + [\mathrm{S}]} \tag{A.14}$$

$$= \frac{V_{\max}}{1 + K_\mathrm{m}/[\mathrm{S}]} \tag{A.15}$$

## APPENDIX B

$$d[\mathrm{EI}]/dt = k_{+1} \cdot [\mathrm{I}] \cdot [\mathrm{E}] - (k_{-1} + k_2) \cdot [\mathrm{EI}] \tag{B.1}$$

$$d[\mathrm{EI}']/dt = k_2 \cdot [\mathrm{EI}] - (k_3 + k_4) \cdot [\mathrm{EI}'] \tag{B.2}$$

$$d[\mathrm{E}_{\mathrm{inact}}]/dt = k_4 \cdot [\mathrm{EI}'] \tag{B.3}$$

The *initial* rate of enzyme degradation, $v_{\text{inact}}$, is described by Eq. B.1, where $k_{\text{obs}}$ is the apparent inactivation rate constant of the enzyme

$$v_{\text{inact}} = d[\text{E}_{\text{inact}}]/dt = k_4 \cdot [\text{EI}'] = k_{\text{obs}} \cdot [\text{E}]_0 \tag{B.4}$$

It is assumed that $[\text{E}_{\text{inact}}]$ is negligible before $[\text{EI}']$ reaches steady state; therefore, the total enzyme concentration, $[\text{E}]_0$, remains the same whether it is free, in complex with inhibitor, [EI], or inactivated $[\text{EI}']$:

$$[\text{E}]_0 = [\text{E}] + [\text{EI}] + [\text{EI}'] \tag{B.5}$$

Referring to Fig. 3, when the levels of the inhibitor-bound enzyme are at steady state, $d[\text{EI}]/dt = 0$,

$$k_1 \cdot [\text{E}] \cdot [\text{I}] = (k_{-1} + k_2) \cdot [\text{EI}] \tag{B.6}$$

therefore,

$$[\text{EI}] = \frac{k_1 \cdot [\text{E}] \cdot [\text{I}]}{k_{-1} + k_2} \tag{B.7}$$

and when the levels of the initial complex formed between metabolic intermediate and enzyme are at steady state, $d[\text{EI}']/dt = 0$,

$$k_2 \cdot [\text{EI}] = (k_3 + k_4) \cdot [\text{EI}'] \tag{B.8}$$

therefore,

$$[\text{EI}'] = \frac{k_2 \cdot [\text{EI}]}{k_3 + k_4} \tag{B.9}$$

Substituting Eq. B.4 into Eq. B.6 gives

$$[\text{EI}'] = \frac{k_2 \cdot k_1 \cdot [\text{I}] \cdot [\text{E}]}{(k_3 + k_4) \cdot (k_{-1} + k_2)} \tag{B.10}$$

Substituting Eqs. B.8 and B.9 into Eq. B.5 gives

$$[\text{E}]_0 = [\text{E}] + \frac{k_1 \cdot [\text{I}] \cdot [\text{E}]}{(k_{-1} + k_2)} + \frac{k_2 \cdot k_1 \cdot [\text{I}] \cdot [\text{E}]}{(k_3 + k_4) \cdot (k_{-1} + k_2)} \tag{B.11}$$

To simplify Eq. B.11, the common denominator $(k_3 + k_4) \cdot (k_{-1} + k_2)$ is used:

$$[\text{E}]_0 = \frac{(k_3 + k_4) \cdot (k_{-1} + k_2)}{(k_3 + k_4) \cdot (k_{-1} + k_2)} \cdot [\text{E}] + \frac{(k_3 + k_4) \cdot k_1 \cdot [\text{I}] \cdot [\text{E}]}{(k_3 + k_4) \cdot (k_{-1} + k_2)} + \frac{k_2 \cdot k_1 \cdot [\text{I}] \cdot [\text{E}]}{(k_3 + k_4) \cdot (k_{-1} + k_2)} \tag{B.12}$$

$$= [\mathrm{E}] \cdot \frac{(k_3 + k_4) \cdot (k_{-1} + k_2) + (k_3 + k_4) \cdot k_1 \cdot [\mathrm{I}] + k_2 \cdot k_1 \cdot [\mathrm{I}]}{(k_3 + k_4) \cdot (k_{-1} + k_2)} \tag{B.13}$$

Therefore,

$$[\mathrm{E}] = \frac{(k_3 + k_4) \cdot (k_{-1} + k_2) \cdot [\mathrm{E}]_0}{k_2 \cdot k_1 \cdot [\mathrm{I}] + (k_3 + k_4) \cdot k_1 \cdot [\mathrm{I}] + (k_3 + k_4) \cdot (k_{-1} + k_2)} \tag{B.14}$$

Substituting Eq. B.13 into Eq. B.10 gives

$$[\mathrm{EI}'] = \frac{k_2 \cdot k_1 \cdot [\mathrm{I}] \cdot [\mathrm{E}]}{(k_3 + k_4) \cdot (k_{-1} + k_2)} \cdot \frac{(k_3 + k_4) \cdot (k_{-1} + k_2) \cdot [\mathrm{E}]_0}{k_2 \cdot k_1 \cdot [\mathrm{I}] + (k_3 + k_4) \cdot k_1 \cdot [\mathrm{I}] + (k_3 + k_4) \cdot (k_{-1} + k_2)} \tag{B.15}$$

$$= \frac{k_2 \cdot k_1 \cdot [\mathrm{I}] \cdot [\mathrm{E}]_0}{k_2 \cdot k_1 \cdot [\mathrm{I}] + (k_3 + k_4) \cdot k_1 \cdot [\mathrm{I}] + (k_3 + k_4) \cdot (k_{-1} + k_2)} \tag{B.16}$$

Dividing through by $k_1$ gives

$$[\mathrm{EI}'] = \frac{k_2 \cdot [\mathrm{I}] \cdot [\mathrm{E}]_0}{k_2 \cdot [\mathrm{I}] + (k_3 + k_4) \cdot [\mathrm{I}] + ((k_3 + k_4) \cdot (k_{-1} + k_2)/k_1)} \tag{B.17}$$

Then dividing through by $(k_2 + k_3 + k_4)$ gives

$$[\mathrm{EI}'] = \frac{(k_2/(k_2 + k_3 + k_4)) \cdot [\mathrm{I}] \cdot [\mathrm{E}]_0}{[\mathrm{I}] + \dfrac{(k_3 + k_4) \cdot (k_{-1} + k_2)}{(k_2 + k_3 + k_4) \cdot k_1}} \tag{B.18}$$

From Eq. B.18, we have

$$k_4 \cdot [\mathrm{EI}'] = \frac{(k_2 \cdot k_4/(k_2 + k_3 + k_4)) \cdot [\mathrm{I}] \cdot [\mathrm{E}]_0}{[\mathrm{I}] + \dfrac{(k_3 + k_4) \cdot (k_{-1} + k_2)}{(k_2 + k_3 + k_4) \cdot k_1}} \tag{B.19}$$

Substituting Eq. B.17 into Eq. B.3 gives

$$k_{\mathrm{obs}} = \frac{(k_2 \cdot k_4/(k_2 + k_3 + k_4)) \cdot [\mathrm{I}]}{[\mathrm{I}] + \dfrac{(k_3 + k_4) \cdot (k_{-1} + k_2)}{(k_2 + k_3 + k_4) \cdot k_1}} \tag{B.20}$$

$$k_{\mathrm{obs}} = \frac{k_{\mathrm{inact}} \cdot [\mathrm{I}]}{[\mathrm{I}] + K_{\mathrm{I}}} \tag{B.21}$$

where $k_{\text{inact}}$ is the maximal inactivation rate constant and $K_{\text{I}}$ is the inhibitor concentration that supports half the maximal rate of inactivation and, referring to Fig. 3, can be described by

$$k_{\text{inact}} = k_2 \cdot k_4 / (k_2 + k_3 + k_4) \tag{B.22}$$

$$K_{\text{I}} = \frac{(k_3 + k_4) \cdot (k_{-1} + k_2)}{(k_2 + k_3 + k_4) \cdot k_1} \tag{B.23}$$

# *IN VIVO* METABOLISM IN PRECLINICAL DRUG DEVELOPMENT

Sevim Rollas
*Marmara University, Istanbul, Turkey*

**Contents**

1 Introduction
2 *In Vivo* Metabolism of Drug Candidates
  2.1 Isolation and Identification of Metabolites
  2.2 Synthesis of Metabolites
  2.3 Protein Binding of Metabolites
  2.4 Preclinical Experimental Studies
3 *In Vivo* Metabolism of Prodrugs
  3.1 Ester and Amide Prodrugs
  3.2 Other Prodrugs
  3.3 Metabolic Stability of Prodrugs
4 Acyl Glucuronides
5 Role of Active Metabolites in Preclinical Drug Development
6 Current Approaches for Evaluating Preclinical *In Vivo* Metabolism
7 Conclusion
  References

## 1 INTRODUCTION

Drug biotransformation plays an important role in the absorption, distribution, metabolism, excretion, and toxicology (ADMET) properties. Undesirable ADMET properties are the cause of many drug development failures [1–3]. Many toxic side effects (carcinogenicity, tissue necrosis, apoptosis, hypersensitivity, teratogenicity) of

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

drugs are directly attributable to the formation of chemically reactive metabolites [4–6]. The liver is a major site for the metabolism of xenobiotics (drugs and other exogenous compounds) and endogenous compounds. Other tissues such as kidney, lungs, adrenals, placenta, brain, intestinal mucosa, and skin have some degree of drug-metabolizing capability. Drug metabolism reactions have been divided into two main classes, phase I (functionalization) includes oxidation, reduction, and enzymatic hydrolysis reactions and phase II (conjugation) includes glucuronidation, sulfation, acetylation, methylation, amino acid and glutation conjugation, and mercapturic acid formation reactions. In phase I reactions, the cytochrome P450 (CYP) enzymes are responsible for oxidative biotransformation of many organic compounds [7]. Other drug-metabolizing enzymes, flavin-containing monooxygenase, monoamine oxidase, epoxide hydrolase, UDP-glucuronosyltransferase, glutation *S*-transferase, sulfotransferase, methyltransferase, and *N*-acetyltransferase may also be important when the drug is primarily metabolized by a non-P450 enzyme [8]. Glucuronidation is a well known major drug-metabolizing reaction (pathway) in humans and changes the structure of the parent drug, and thus its chemical and biological reactivity.

Individual differences in drug biotransformation exist. Some people metabolize a drug rapidly so that therapeutically effective blood and tissue levels are not achieved; others metabolize the same drug slowly so that toxicity results. Most CYP enzymes that belong to the CYP 1, 2, or 3 families are polymorphic [9]; so it may be diffucult to predict a clinical response to a particular dose of a drug. Genetic polymorphisms have played an important role in these differences [10].

Metabolites can be obtained from plasma, urine, and bile after administration of compounds to laboratory animals and are used in preclinical *in vivo* studies as a reference standard [11–13]. The synthesis of metabolites is mostly done with parent compounds or is performed by total synthesis [14]. In addition, microbial methods can be used to obtain metabolites. The last two methods are very important for active metabolite synthesis.

Prodrug approches are very valuable in drug development. The goal of prodrug development is to solve specific pharmaceutical or pharmacological problems. The prodrug itself is inactive and is converted to the active drug *in vivo*. Early screening of their *in vitro* and/or *in vivo* plasma stability provides vital information for a stability profile of prodrug candidates [15, 16].

The application of well established preclinical *in vitro* and *in vivo* methodologies, such as gas chromatography–mass spectrometry (GC-MS) and liquid chromatography–mass spectrometry (LC-MS), is essential for drug discovery. Sensitive and specific analytical methods—for example liquid chromatography–high resolution nuclear magnetic resonance spectrometry [17, 18], LC-MS/MS (tandem mass spectrometry) [19–22], liquid chromatography–nuclear magnetic resonance/mass spectrometry (LC-NMR/MS) [23], and quadrupole time-of-flight mass spectrometry [24]—provide the identification and quantitation of metabolites. In general, this chapter is devoted to the preclinical *in vivo* metabolism of drug candidates. In the first part of this chapter, isolation, identification, synthesis, protein binding, and preclinical experimental studies are discussed, while in the following parts, *in vivo* metabolism of prodrugs, acyl glucuronides, and the role of active metabolites are covered.

## 2 *IN VIVO* METABOLISM OF DRUG CANDIDATES

In preclinical drug discovery, investigation of the *in vitro, in vivo* metabolism and toxicity screening of a drug candidate together play a very important role in the drug's success before entering into clinical use [25]. Rapid metabolism is one of the most important problems in achieving therapeutic drug levels [26]. Liver microsomal incubation has routinely been carried out by pharmaceutical companies to survey the metabolisim of potential drugs. Phase I and phase II drug-metabolizing enzymes affect the overall therapeutic and toxicity profiles of a drug. Several *in vitro* models, such as isolated perfused livers, liver tissue slices, freshly isolated hepatocytes in suspension, primary hepatocyte cultures, cell lines, microsomes, mitochondria, and expression systems, in particular, have been developed to profile human metabolism [1]. *In silico* (computational) techniques are tools that are used to predict the metabolic profiles of drugs that are in the design phase, but there is still insufficient data on the efficacy of *in silico* techniques. That is why *in vivo* metabolism studies should be performed on pharmacophore groups for an *in silico* database. *In vivo* metabolism is still being studied for conventional drugs in humans or animals for the detection of unknown metabolites [27–31]. During drug development, *in vivo* drug absorption, metabolic fate, and hepatic first-pass metabolism studies generally are performed in rats, dogs, and/or monkeys [32–34]. Indeed, the results of *in vitro* metabolism may not be the same as results of *in vivo* metabolism in animals and humans. Reliable human tests should be carried out. Human hepatocytes are used to model biotransformation in the human liver to predict human clearance [35–37]. Dog hepatocytes have also been used. However, the most successful results have been provided using human hepatocytes [38, 39]. Investigating the stereoselective metabolism of optical isomers of drugs is crucial in distinguishing their pharmacokinetic behaviors. The formation of sulfone, hydroxy, and 5-*O*-demethyl metabolites from *S*- and *R*-omeprazole using liver microsomes has been reported [40]. The metabolism of the optical isomers of omeprazole exhibits a significant stereoselectivity. Current strategies in high throughput chemistry are focused on high quality lead compounds [41]. Therefore animal testing is the most important step for lead identification and *in vitro–in vivo* correlation. The preclinical evaluation of selected lead compounds is represented in Fig. 1.

### 2.1 Isolation and Identification of Metabolites

Metabolite isolation and identification are important processes in discovery and development of new drugs. In many cases, the isolation and identification of metabolites are diffucult because of the low concentration or instability of these compounds in biological matrix. Therefore metabolites are synthesized and characterized by spectroscopic methods. The drug metabolism methodologies may be divided into isolation technologies (e.g., extraction, fractioning, chromatography) and chemical characterization (by means of spectroscopic techniques). For *in vivo* characterization of metabolites, biological samples are collected and it is determined what extraction methods (e.g., preparative solid phase extraction, solid phase extraction, liquid–liquid extraction, or protein precipitation (deproteination)) are necessary for analysis of the samples. Several studies reported use of extraction processes [13, 14,

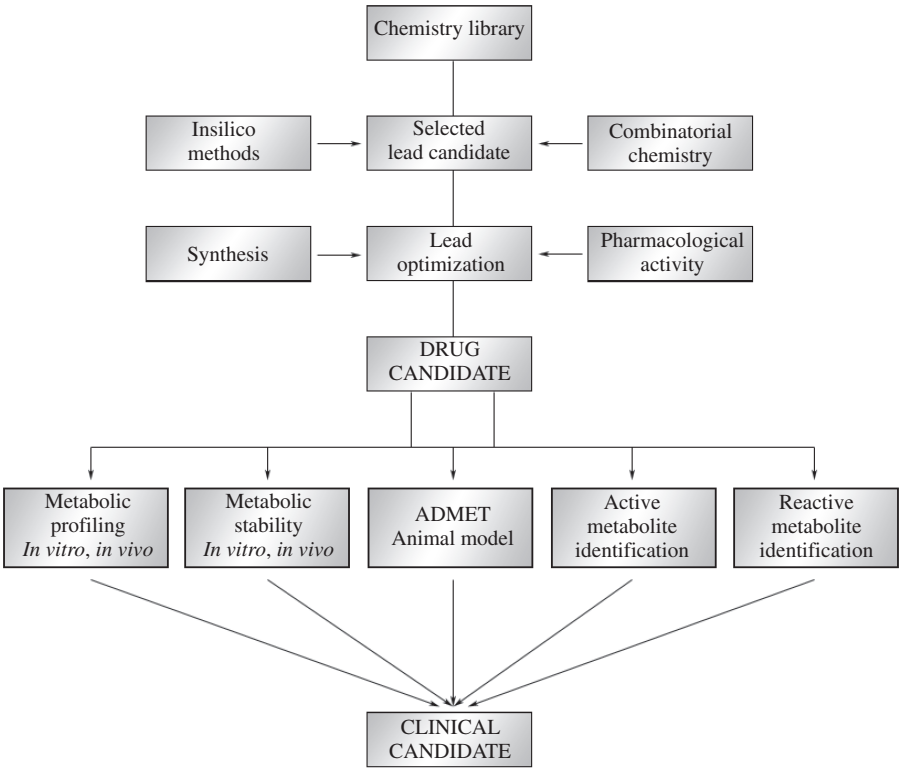


**FIGURE 1** Preclinical evaluation of selected lead candidates.

19, 34, 42–46]. Recently, the use of turbulent flow chromatography with online solid phase extraction and column switching for sample preparation has been reported by Ong et al. [47]. Development of an analytical methodology suitable for accurate determination of both drug and metabolite in biological fluids is an advantange for the examination of the metabolic profile of a drug. The well known technique of liquid chromatography–tandem mass spectrometry (LC-MS/MS), can provide information for the identification for drug metabolites [25, 48]. High resolution nuclear magnetic resonance (NMR) spectroscopy is widely used for structure elucidation of organic compounds and their metabolites in biological samples [49–51]. The most high throughput analytical method for substrate and metabolite analysis is liquid chromatography with parallel NMR and mass spectrometry (LC-NMR/MS) [23, 49]. Timm and colleagues [42] reported another method for the analysis of oral platelet aggregation inhibitor, a double protected prodrug, and its metabolites by using HPLC—column switching combined with turbo ion spray single quadrupole mass spectrometry after precipitation of plasma protein by 0.5 M perchloric acid. Deproteination is a suitable method for sample preparation. Acetonitrile, methanol, perchloric acid, and trichloroacetic acid are used as deproteination agents [52–56]. Hydrophilic glucuronide conjugates may be hydrolyzed by β-glucuronidase in acetate buffer [28, 29, 57] or by acid or alkali [46, 58, 59] before being analyzed. In recent years, ultra performance liquid chromatography (UPLC) [60] has been used

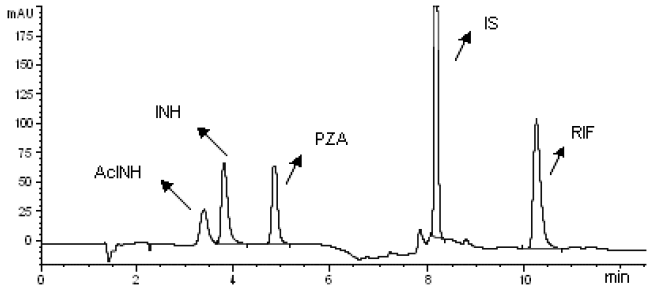


**FIGURE 2** HPLC chromatogram of AcINH (acetylisoniazide), INH (isoniazide), PZA (pyrazinamide), IS (internal standard), and RIF (rifampin). Pore size is 4 μm Pore 3.9 × 150 mm long Nova-Pak $C_{18}$ column. Signals were monitored by diode array detector. Gradient elution was used.

for drug and metabolite identification in biological fluids [61, 62]. Johnson and Plump [62] reported human metabolism of acetaminophen using UPLC with quadrupole time-of-flight mass analysis. Dear et al. [52] also reported a new approach by using UPLC for *in vivo* metabolite identification. These novel fast and successful methods may be applied to preclinical *in vivo* metabolite identification of drug candidates.

Metabolites alter the quantity of drug during pharmacokinetic or drug monitoring studies if an efficient separation is not provided. In our study on the monitoring of isoniazide, pyrazinamide, and rifampicin, for the separation of isoniazid and its major metabolite acetylisoniazid, we developed an HPLC analysis method represented in Fig. 2 for the drug monitoring study [55].

## 2.2 Synthesis of Metabolites

Metabolite identification and synthesis are important processes in the development of new lead candidates for drug metabolite profiling, pharmacokinetic studies (interference of metabolite), pharmacological activity testing (for active metabolite), metabolite quantification, CYP identification, and toxicity testing. Therefore a synthesis of methods are needed to obtain authentic metabolites. The synthesis of metabolites that are not easily obtained by chemistry methods may be produced by microbial biotransformation of the drugs. Chemical synthesis [14, 63] and microbial methods [64–66] are reported in the literature. Sulfadimethoxine *N*-glucuronide was prepared by Bridges et al. [67] starting from sodium sulfadimethoxine and methyl 2,3,4-tri-*O*-acetyl-1-bromoglucuronate. Ketotifen [68] and soraprazam [69] glucuronides have also been prepared synthetically. The N-oxide metabolite of the model compound *N*-benzyl-*N*-methylaniline was prepared by oxidizing with hydrogen peroxide [70]. Sulfinpyrazone sulfide metabolite was prepared by He et al. [71] by the reduction of sulfinpyrazone with iron powder. The hydroxamic acid metabolites are usually obtained from molecules containing an ester functional group [72].

The use of microbial models is a suitable method to produce a sufficient quantity of metabolites [73, 74]. Most bacterial species have P450 enzymes [75]. Enantiomeri-

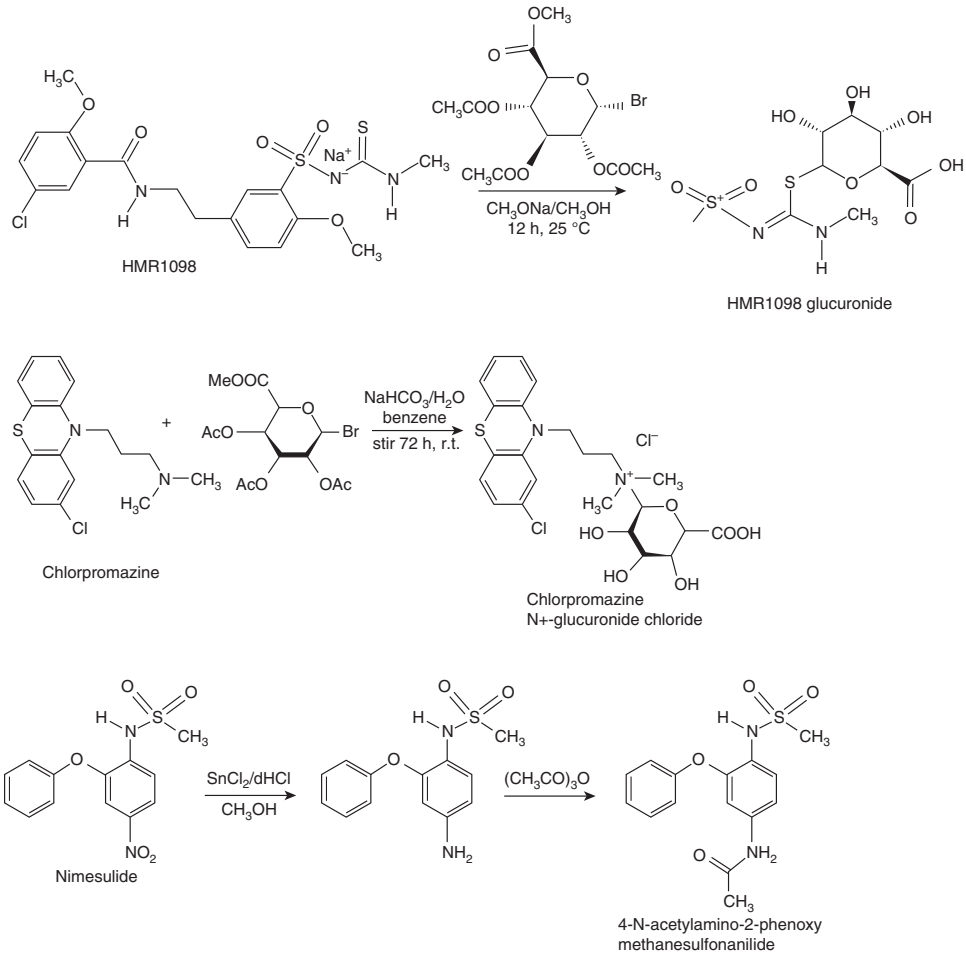


**FIGURE 3** Examples for synthesis of some metabolites.

cally pure sulfoxides may be synthesized by biocatalytic methods using microorganisms or isolated enzymes. Ricci et al. [76] prepared aromatic sulfoxides from sulfides employing Basidiomycetes. Zhang et al. [77] reported biotransformation of amitriptyline by the filamentous fungus *Cunninghamella elegans*. The amitriptyline N-oxide metabolite that is produced from fungus can also be obtained by the oxidation of amitriptyline with 3-chloroperoxybenzoic acid to compare with the microbial product. The biosynthesis of drug glucuronides may be performed using suitable animal microsomes to obtain glucuronide metabolites [78]. The synthesis route of some metabolites is shown in Fig. 3 [20, 79, 80].

## 2.3 Protein Binding of Metabolites

Most drugs and/or their metabolites bind to plasma proteins and other blood components such as red blood cells [81–84]. Generally, free drug molecules can arrive

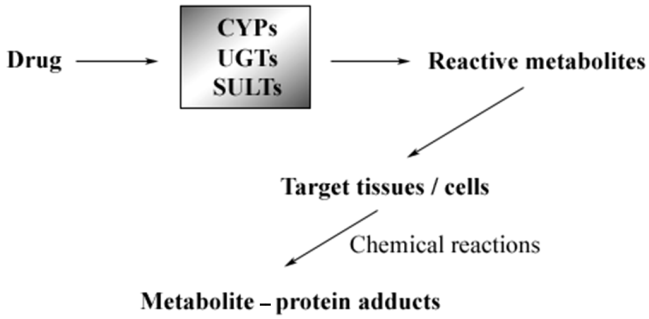


**FIGURE 4** Formation of metabolite–protein adducts of drugs.

at the site of action. They are distributed into tissues for direct binding to tissue proteins [85]. Binding of a drug to plasma protein influences the drug's ADME processes. Yamazaki and Kanaoka [86] have developed a computational method that can provide precise and useful prediction of plasma protein binding for new drug candidates. However, preclinically, the *in vivo* studies must also be performed on drug candidates. The binding of drugs and/or their metabolites to plasma and tissue proteins is reversible. In some cases, they may covalently bind to plasma and tissue proteins, resulting of drug–protein or metabolite–protein adducts [87–89]. In most cases, the covalent binding by drugs is via their reactive metabolites [90–93]. Covalent binding to tissue proteins has been demonstrated for several metabolites such as acetaminophen–quinoneimine metabolite [94], carbamazepine–10,11-epoxide metabolite [95], furosemide–epoxide metabolite [88], ibuprofen–glucuronide metabolite [96], clozapine–nitrenium ion derived from clozapine N-oxide metabolite [97], and valproic acid–4-ene metabolite [98]. Formation of metabolite–protein adducts of drugs is represented in Fig. 4.

### 2.4 Preclinical Experimental Studies

The application of human-based preclinical *in vitro* hepatic metabolism (hepatocyte or liver microsome) is a must for the evaluation of drug metabolism and the selection of drug candidates with a probability of clinical success. However, *in vitro* metabolic stability and metabolite profiling studies should be correlated with *in vivo* experiments with relevant laboratory animal species for prediction of *in vivo* human drug properties [99]. Sample preparation is the most important element in *in vivo* metabolism studies. Conventional solid phase extraction is performed with a $C_{18}$ cartridge [82, 100, 101]. Recently, Oasis HLB (hydrophilic–lipophilic balance), MCX (mixed mode, cation-exchange), and MAX (mixed mode, anion-exchange) cartridges have been developed to prepare samples [102–104]. Cleanup of the samples by means of solid phase extraction is shown Fig. 5.

Examples of *in vivo* metabolism studies are given next. The metabolite profile of 5-(2-ethyl-2*H*-tetrazol-5-yl)-1-methyl-1,2,3,6-tetrahydropyridine (Lu 25-109), a muscarine agonist, has been reported in mice, rats, dogs, and humans in plasma and

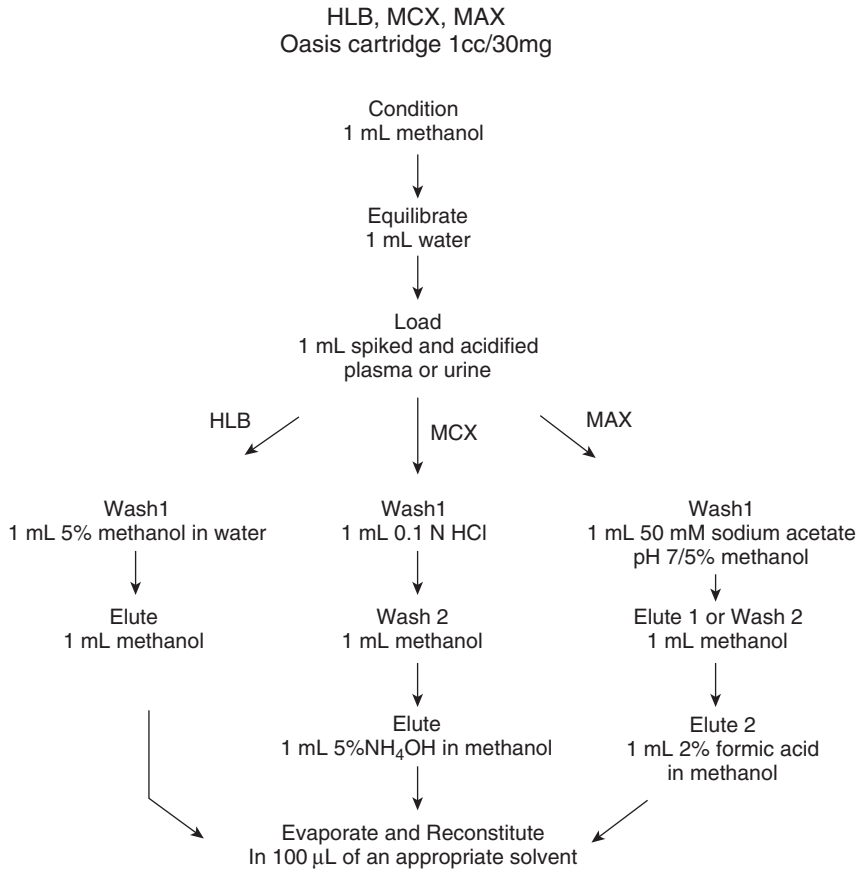


**FIGURE 5** Cleanup of the samples by HLB, MCX, and MAX cartridges.

urine after oral administration of [$^{14}$C]Lu 25-109. The *N*-deethyl metabolite is a major metabolite in human plasma. The formation of a pyridine derivative has only been shown in rats [105].

The myocardial binding of the high potency antipsychotic benzisoxazol derivative risperidone and its active metabolite, 9-hydroxyrisperidone, was studied by using equilibrium dialysis *in vitro* and using intraperitoneal administration *in vivo* on plasma and tissue samples of guinea pig by Titier et al [106]. Risperidone and its metabolite concentrations have been determined by the HPLC method with UV detection. The plasma protein and tissue binding of 9-hydroxyrisperidone was lower than that of the more lipophilic risperidone. This method can be applied for other drugs and for therapeutic drug monitoring [106].

In our laboratory, *in vivo* metabolic pathway selected model compounds—1,2,4-triazole-3-thiones [82, 107], hydrazide [108], thiourea derivative [109], azo compounds [110, 111], and quinazolinone [112] and pyridazine [113] rings compounds—were investigated in rats (Fig. 6). The potential metabolites of substrate, which cannot be provided, were synthesized. Unknown metabolites were

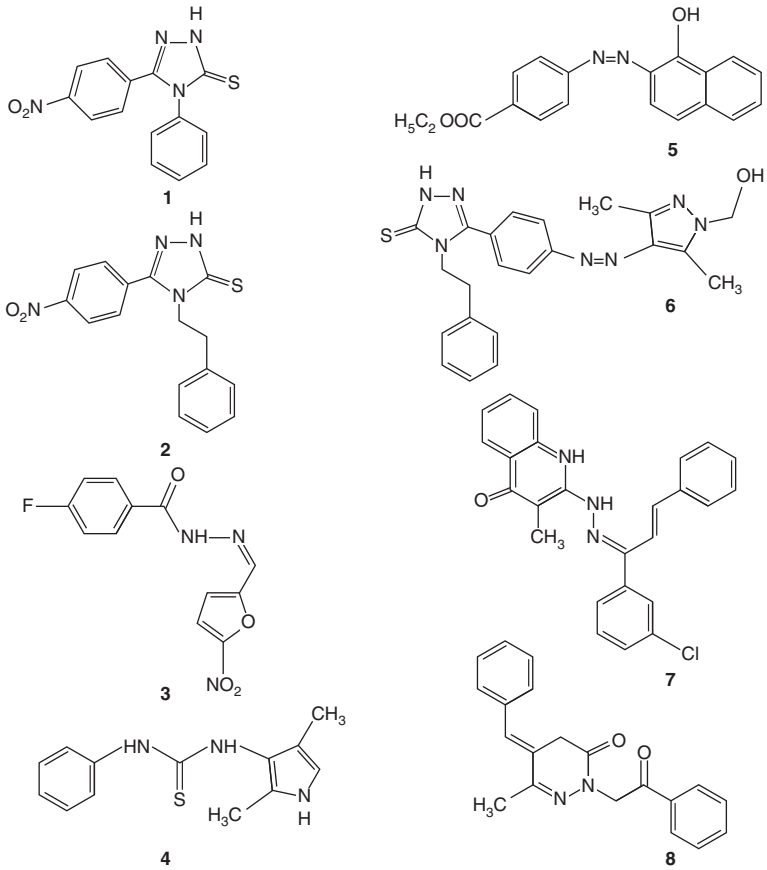


**FIGURE 6** Examples of model compounds investigated for their *in vivo* meta-bolism: (**1**) 5-(4-nitrophenyl)-4-phenyl-2,4-dihydro-3*H*-1,2,4-triazole-3-thione [82]; (**2**) 5-(4-nitrophenyl)-4-(2-phenylethyl)-2,4-dihydro-3*H*-1,2,4-triazole-3-thione [107]; (**3**) 4-fluorobenzoic acid [(5-nitro-2-furanyl)methylene]hydrazide [108]; (**4**) *N*-phenyl-*N*′-(3,5-dimethyl-pyrazole-4-yl)thiourea [109]; (**5**) ethyl 4-[(2-hydroxy-1-naphthyl)azo]benzoate [110]; (**6**) 4-phenylethyl-5-[4-(1-(2-hydroxy-ethyl)-3,5-dimethyl-4-pyrazolylazo)-phenyl]-2,4-dihydro-3*H*-1,2,4-triazole-3-thione [111]; (**7**) 2-[1′-phenyl-3′-(3-chlorophenyl)-2′-propenylyden]hydrazino-3-methyl-4(3*H*)-quinazolinone [112]; (**8**) 3-oxo-5-benzylidene-6-methyl-(4*H*)-2-(benzoylmethyl)pyridazine [113].

collected during HPLC studies for mass analysis. Desulfurization reaction did not occur in the 1,2,4-triazole-3-thiones model compounds. Their oxygen analogues were synthesized and used as a reference standard in these studies. Examples of model compounds investigated for their *in vivo* metabolism are shown in Fig. 6.

4-Fluorobenzoic acid [(5-nitro-2-furanyl)methylene]hydrazide (substrate) is active against *Staphylococcus aureus* ATCC 29213 (MIC value 3.25 μg/mL) [114]. It was administered intraperitoneally in doses of 50 or 100 mg/kg. Blood samples were collected after administration. The substrate and its potential metabolites were synthesized and separated using HPLC [108]. The substrate and its potential metabolites were not detected in plasma. After plasma proteins were denaturated, 4-

fluorobenzoic acid (1.5 h) was detected. However, it rapidly metabolizes via a hydrolytic reaction in rats and shows poor metabolic stability *in vivo*.

## 3 *IN VIVO* METABOLISM OF PRODRUGS

Drug activation has been known since the identification of prontosil as a bioreductive prodrug of sulfanylamide. The aim of prodrug development is to solve specific pharmaceutical or pharmacological problems. Various drugs with groups that can be modified by chemical reactions can be converted into prodrugs. They can be synthesized to improve drug stability and absorption, to reduce diverse effects, to extend the duration of action, to increase water solubility or other desirable properties, and to allow site-specific delivery [115]. Generally, synthetic or natural compounds with therapeutic potential as bioactive agents contain alcohol, phenol, ester, thioether, amino, or nitro groups in their structure. In the synthesis of prodrugs, mainly esters, amides, ethers, phosphamides, hydroxamic acids, imines, N-oxides, Mannich bases, azo groups, glycosides, peptides, salts, polymers, and complexes from the parent drug are obtained. The anti-inflammatory drug sulindac is a prodrug and its two active metabolites—sulindac sulfide and sulindac sulfone—result from reduction and oxidation, respectively [116]. This type of prodrug is called a *bioprecursor*. Examples of other bioprecursors are diazepam, primidone, and clofibrate. Quantitative prediction of the effect of a prodrug's enzymatic catalysis *in vivo* is diffucult due to the variety and complexity of enzyme systems [117]. The conversion of a prodrug into a parent drug is represented schematically in Fig. 7.

### 3.1 Ester and Amide Prodrugs

Esterification and amidation are two important reactions for the synthesis of prodrugs. Many drugs contain free carboxylic acid and/or an amine group. Chandrasekaran et al. [118] reported that the nonsteroidal anti-inflammatory drug indomethacin conjugated with triethylene glycol at the carboxylic acid to obtain its ester and amide prodrugs. Sriram et al. [119] synthesized the amino acid ester prodrug of stavudine for the effective treatment of HIV/AIDS. Wu et al. [120] synthesized lipophilic metronidazol esters using enzymes as catalyst. The best enzymatic transesterfication is provided by *Candida antarctica* lipase acrylic resin.

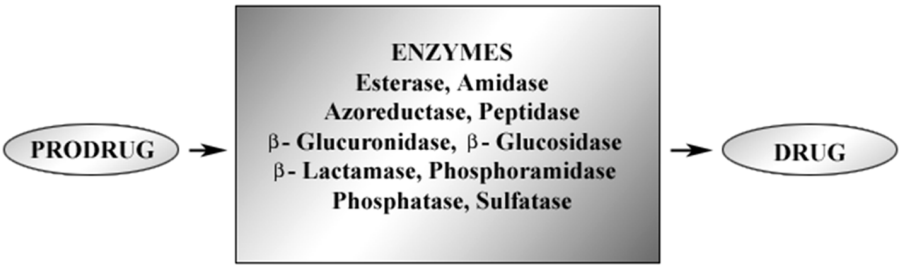


**FIGURE 7** Conversion of prodrug to parent drug.

### 3.2 Other Prodrugs

Generally, prodrugs are designed for dealing with problems like rapid metabolism or poor absorption or reducing diverse effects seen with conventional drugs. In recent years, tumor-activated prodrug (TAP) strategies have become a very important approach in cancer therapy [121]. These prodrugs can be selectively activated in tumor tissue. The aromatic and aliphatic N-oxides, quinones, aromatic nitro groups, and cobalt complexes are the most important structures in developing hypoxia-selective TAPs [122]. Monoclonal antibody–enzyme conjugates (antibody-directed enzyme prodrugs) have recently been developed and rituximab, cetuximab, and trastuzumab are being used in the clinical setting [123,124]. β-Lactamase-dependent prodrugs have also been developed for antibody-directed enzyme therapy [125].

Another class of prodrugs is the dendritic prodrugs, which can improve drug delivery and solubility: thus decreasing dosages and prolonging drug release [126–128]. Najlah et al. [129] reported that the design, synthesis, and characterization of polyamidoamine (PAMAM) dendrimer-based prodrugs are described by selecting as the model compound naproxen, a poorly water-soluble drug. Propranolol-$G_3$ PAMAM dendrimer conjugates were also synthesized to increase propranolol solubility [130]. Salamonczyk [131] prepared phosphorus-based dendrimers of acyclovir.

### 3.3 Metabolic Stability of Prodrugs

The metabolic stability of a drug candidate in liver microsomes of different animal species is determined in order to assess the potential of this drug to form undesired, potentially toxic, or pharmacologically inactive/active metabolites. In the determination of metabolic stability, *in vitro* models were widely used in order to investigate the metabolic fate of drug candidates. Liver microsomes and S9 (cytosolic and mitochondrial) fractions as drug-metabolizing enzyme sources are used for metabolic stability studies [132]. Di et al. [133] have suggested a single time-point microsomal stability assay and incubation time of 15 min for metabolic stability profiling at the early stage of drug discovery. The duration and/or the intensity of action of most drugs are determined by their rate of metabolism by hepatic enzymes [134]. The metabolic stability procedure involves the following [15, 16, 135]:

- Preparation of buffers
- Preparation of stock solution of test compound (in DMSO or water)
- Incubation (at 37 °C)
- Sample preparation
- Determination of loss of parent compound (using LC-MS, LC-MS/MS)

Screening of plasma stability of the prodrugs is very important for rapid conversion in plasma [136]. However, *in vitro* hydrolysis rates of prodrugs of ester or amide must be correlated with *in vivo* hydrolysis rates [137]. Lai and Khojasteh-Bakht [138] reported a new method for prodrug stability—an automated online liquid chromatographic–mass spectrometric method.

## 4 ACYL GLUCURONIDES

Glucuronidation reaction represents the major route for elimination and detoxification of drugs. The biosynthesis of glucuronides is catalyzed by the uridine diphosphate glucuronosyltransferases, which are localized in hepatic endoplasmic reticulum. However, the glucuronidation of some drugs has been observed in human kidney microsomes [135]. Seventeen human UDP-glucuronosyltransferases have been identified to date [139]. The mechanism of glucuronidation is a nucleophilic substitution reaction and the resulting glucuronide has the 1β-glycosidic configuration.

The rate of glucuronidation is dependent on the nucleophilic character of the substrate and enzyme concentration. The most systematic investigations were initiated in the 1930s by R. T. Williams and collaborators [140, 141]. Acyl glucuronide metabolites of carboxylic acid may be hydrolyzed under physiological conditions. Therefore the glucuronides produced can potentially interfere with the pharmacokinetics of the parent drug. In this case, early pharmacokinetic study is important in animal models to regulate dosing and duration of action.

On the other hand, glucuronide conjugates are capable of cellular injury (e.g., hepatotoxicity, carcinogenesis) by facilitating the formation of reactive electrophilic intermediates and their transport into target tissues [87]. Acyl glucuronides are chemically unstable in aqueous solution and undergo the intramolecular acyl migration. The acyl group may migrate to the C-2, C-3, or C-4 position of the glucuronic acid from the 1β position [142]. In addition, unstable acyl glucuronides have also been shown to form covalent adducts with proteins under *in vivo* conditions. The existence of genetic polymorphism in the rate of drug acylation has important consequences in drug therapy. Chen et al. [143] developed a rapid *in vitro* method for the assessment of the stability of acyl glucuronides. Many nonsteroidal anti-inflammatory drugs, such as diclofenac, diflunisal, zomepirac, and naproxen, and other drugs, such as valproic acid, citalopram, and furosemide, which bear the carboxylic functional group, have been metabolized to form acy lglucuronides [49, 144, 145].

## 5 ROLE OF ACTIVE METABOLITES IN PRECLINICAL DRUG DEVELOPMENT

Drugs can metabolize inactive, reactive, or pharmacologically active molecules [146, 147]. Fura et al. [148] and Guengerich et al. [149] define an active metabolite as the active metabolic product possessing the same pharmacological activity as the parent drug. Pharmacologically active metabolites should be recognized at the drug discovery stage, because the potential contribution of an active metabolite to total activity is an important factor for pharmacokinetic properties of a drug candidate. Therefore active metabolite kinetics should be studied in animal models. Klein et al. [150] reported a population pharmacokinetic model for irinotecan and two of its metabolites, SN-38 and SN-38 glucuronide. The active metabolite 7-hydroxycamptothecin (SN-38) is 300–1000 times more potent than its parent molecule irinotecan. Therefore it is important to develop dosing strategies for irinotecan. Mak and Weglicki [151] reported on the antioxidant properties of 4-hydroxypropranolol, a major metabolite of propranolol. It is four- to eightfold more potent than vitamin E. As a result, 4-hydroxypropranolol may contribute to the cardiovascular therapeutic

benefits of propranolol because of its endothelial cytoprotective efficacy against cell injury.

Active metabolites play a role as lead candidates during lead optimization [148, 152]. For example, ezetimibe was found during the lead optimization phase [152]. During drug development, it is sometimes impossible to characterize the pharmacological activity of active metabolites because of the difficulties in their synthesis by common methods. They may be more active than their respective parent molecules. However, a number of active metabolites have been marketed as drugs: for example, acetaminophen, cetirizine, imipramine, desloratadine, fexofenadine, mesoridazone, nortriptyline, and oxazepam [148].

Metabolism of the antifungal drug itraconazole (ITZ) involves both phase I (oxidative) and phase II (conjugative) pathways. One of these metabolites, hydroxyitraconazole (OH-ITZ), possesses antifungal activity similar to that of ITZ [153]. Therefore, the pharmacokinetic behavior of ITZ may be affected by its active metabolites. Gasparro et al. [154] reported that the active metabolite *R*-desmethyldeprenyl of *R*-deprenyl has been shown to be more active than R-deprenyl. Ciclesonide, an inhaled corticosteroid with prolonged anti-inflammatory activity, is being developed for the treatment of asthma. Nave et al. [155] reported ciclesonide is converted to an active metabolite, desisobutyryl-ciclesonide, which undergoes reversible fatty acid conjugation in *in vitro* studies. Cannell et al. [144] demostrated that nonsteroidal anti-inflammatory drugs (diflunisal, zomepirac, and diclofenac), acyl glucuronides, and their rearrangement isomers had antiproliferative activity on human adenocarcinoma HT-29 cells in culture.

In conclusion, active metabolites may:

- Alter pharmacokinetic properties of parent compounds.
- Cause enzyme induction or inhibition.
- Alter pharmacokinetics of drugs when used together.
- Show different pharmacological activity.
- Be toxic.

## 6 CURRENT APPROACHES FOR EVALUATING PRECLINICAL *IN VIVO* METABOLISM

Metabolism is one of the most important steps in ADMET studies. Metabolites affect pharmacokinetic results [156]. *In vivo* pharmacokinetic behavior involves permeability, metabolic stability, inhibition, and induction of CYP isoenzymes and plasma protein binding. A new approach may be the use of ultra performance liquid chromatography [52], online hyphenated LC-NMR/MS [48], and an integrated high capacity solid phase extraction–MS/MS system [157] for *in vivo* metabolite identification in the future. Metabolism is also one of the most complicated of the pharmacokinetic properties. Genetic, environmental, and physiological factors may affect drug metabolism. Computational approaches for predicting CYP-related metabolism is very important for the evaulation of a successful new drug [158]. Therefore, CYP inhibitors, CYP inducers, CYP regioselectivity, and CYP substrates have been studied for the development of a reliable predictive model.

Computational or virtual ADME [2, 26, 159–162] and metabolism models derived from past *in vitro* and *in vivo* data have been developed to contribute to new drug discovery in the pharmaceutical industry [163, 164]. Klopman et al. [165] evaluated a computer model for prediction of intestinal absorption in humans. This model may be applied to particularly active metabolites. Recent studies have shown that one of the current trends will be drug-induced autoimmunity. Idiosyncratic drug reactions that do not occur in most patients are an important clinical problem [166]. Most of these reactions appear to be immune-mediated reactions. It is nessesary to develop new animal models, in addition to the present models, for preclinical studies. However, the newest approaches of drug discovery may be the mechanism-based, the function-based, and the physiology-based approaches [167].

## 7 CONCLUSION

In the last few years, much effort in drug discovery and development has focused on defining the metabolic profile, toxicity, stereoselective metabolism, and pharmacokinetics of new drug candidates. Understanding the role of drug–metabolizing enzymes in drug metabolism is an important area of research for human toxicology testing. *In vitro* and *in vivo* models have been used as part of high throughput screening programs to characterize lead identification, drug metabolic profiling, metabolic stability, drug permeability, drug solubility, and drug safety. *In vitro* studies generally provide a valuable insight into a drug's metabolic profile in humans; however, sometimes there are discrepancies between *in vitro* results and *in vivo* findings. In the future, the use of *in silico* predictive models and ultra developed analysis techniques for preclinical *in vivo* studies will be an essential requirement for the discovery and development of successful new drugs.

## REFERENCES

1. Sivaraman A, Leach JK, Townsend S, Iida T, Hogan BJ, Stolz DB, Fry R, Samson LD, Tannenbaum SR, Griffith LG. A microscale *in vitro* physiological model of the liver: predictive screens for drug metabolism and enzyme induction. *Curr Drug Metabolism* 2005;6:569–591.
2. Bugrim A, Nikolskaya T, Nikolsky Y. Early prediction of drug metabolism and toxicity: systems biology approach and modeling. *Drug Discov Today* 2004;9:127–135.
3. Ekins S, Ring BJ, Grace J, McRobie-Belle DJ, Wrighton SA. Present and future *in vitro* approaches for drug metabolism. *Pharmacol Toxicol Methods* 2000;44:313–324.
4. Wells PG, Kim PM, Laposa RR, Nicol CJ, Parman T, Winn LM. Oxidative damage in chemical teratogenesis. *Mutat Res* 1997;396:65–78.
5. Williams DP. Toxicophores: investigations in drug safety. *Toxicology* 2006;226:1–11.
6. Park KB, Dalton-Brown E, Hirst C, Williams DP. Selection of new chemical entities with decreased potential for adverse drug reactions. *Eur J Pharmacol* 2006;549:1–8.
7. Rendic S. Summary of information on human CYP enzymes: human P450 metabolism data. *Drug Metab Rev* 2002;34:438–448.
8. Lu AYH, Wang RW, Lin JH. Cytochrome P450 *in vitro* reaction phenotyping: a re-evaluation of approaches used for P450 isoform identification. *Drug Metab Dispos* 2003;31:345–350.

9. Agundez JAG. Cytochrome P450 gene polymorphism and cancer. *Curr Drug Metab* 2004;5:211–224.
10. Dorne JLCM, Walton K, Renwick AG. Human variability in xenobiotic metabolism and pathway-related uncertainty factors for chemical risk assessment: a review. *Food Chem Toxicol* 2005;43:203–216.
11. Williams RT. Studies in detoxification 13. The biosynthesis of aminophenyl- and sulphonamidoaminophenyl- glucuronides in the rabbit and their action on haemoglobin *in vitro*. *Biochem J* 1943;37:329–333.
12. Turgeon J, Pare JRJ, Lalande M, Grech-Belanger O, Belanger PM. Isolation and structural characterization by spectroscopic methods of two glucuronide metabolites of mexiletine after N-oxidation and deamination. *Drug Metab Dispos* 1992;20:762–769.
13. Sidelman UG, Christiansen E, Krogh L, Cornett C, Tjornelund J, Hansen SH. Prufication and $^{1}$H NMR spectroscopic characterization of phase II metabolites of tolfenamic acid. *Drug Metab Dispos* 1997;25:725–731.
14. Lantz RJ, Gillespie TA, Rash TJ, Kuo F, Skinner M, Kuan HY, Knadler MP. Metabolism, excretion, and pharmacokinetics of duloxetine in healthy human subjects. *Drug Metab Dispos* 2003;31:1142–1150.
15. Di L, Kerns EH, Hong Y, Chen H. Develoment and application of high throughput plasma stability assay for drug discovery. *Int J Pharm* 2005;297:110–119.
16. Kerns EH, Di L. Pharmaceutical profiling in drug discovery. *Drug Discov Today* 2003; 8:316–323.
17. Mutlib AE, Chen H, Nemeth GA, Markwalder JA, Seitz SP, Gan LS, Christ DD. Identification and characterization of efavirenz metabolites by liquid chromatography/mass spectrometry and high field NMR: species differences in the metabolism of efavirenz. *Drug Metab Dispos* 1999;27:1319–1333.
18. Zhang KE, Hee B, Lee CA, Liang B, Potts BCM. Liquid chromatography–mass spectrometry and liquid chromatography–NMR characterization of *in vitro* metabolites of a potent and irreversible peptidomimetic inhibitor or rhinovirus 3C protease. *Drug Metab Dispos* 2001;29:729–734.
19. Bramer SL, Tata PNV, Vengurlekar SS, Brisson JH. Method for the quantitative analysis of cilostazol using LC/MS/MS. *J Pharm Biomed Anal* 2001;26:637–650.
20. Kucukguzel SG, Kucukguzel I, Oral B, Sezen S, Rollas S. Detection of nimesulide metabolites in rat plasma and hepatic subcellular fractions by HPLC-UV/DAD and LC-MS/MS studies. *Eur J Drug Metab Pharmacokinet* 2005;30:127–134.
21. Lampinen-Salomonsson M, Beckman E, Bondesson U, Hedeland M. Detection of altrenogest and its metabolites in post administration horse urine using liquid chromatography tandem mass spectrometry—increased sensitivity by chemical derivatization of glucuronic acid conjugate. *J Chromatogr B* 2006;833:245–256.
22. Guo L, Qi M, Jin X, Wang P, Zhao H. Determination of active metabolite of prulifloxacin in human plasma by liquid chromatography–tandem mass spectrometry. *J Chromatogr B* 2006;832:280–285.
23. Corcoran O, Spraul M. LC-NMR-MS in drug discovery. *Drug Discov Today* 2003; 8:624–631.
24. Ishigai M, Langridge JI, Bordoli RS. A new approach for dynamics of enzyme-catalyzed glutation conjugation by electrospray quadrupole/time–of-flight mass spectrometry. *Anal Biochem* 2001;298:83–92.
25. Lee M-Y, Dordick JS. High-throughput human metabolism and toxicity analysis. *Curr Opin Biotechnol* 2006;17:1–9.

26. Tarbit MH, Berman J. High-throughput approaches for evaluating absorption, distribution, metabolism and excretion properties of lead compounds. *Curr Opin Chem Biol* 1998;2:411–416.
27. Schaber G, Wiatr G, Wachsmuth H, Dachtler M, Albert K, Gaertner I, Breyer-Pfaff U. Isolation and identification of clozapine metabolites in patient urine. *Drug Metab Dispos* 2001;29:923–931.
28. Strickmann DB, Blaschke G. Isolation of an unknown metabolite of non-steroidal anti-inflammatory drug etodolac and its identification as 5-hydroxy etodolac. *J Pharm Biomed Anal* 2001;25:977–984.
29. Desmoulin F, Gilard V, Martino R, Malet-Martino M. Isolation of an unknown metabolite of capecitabine, an oral 5-fluorouracil prodrug, and its identification by nuclear magnetic resonance and liquid chromatography–tandem mass spectrometry as a glucuroconjugate of 5′-deoxy-5-fluorocytidine, namely 2′-(β-D-glucuronic acid)-5′-deoxy-5-fluorocytidine. *J Chromatogr B* 2003;792:323–332.
30. Walles M, Thum T, Levsen K, Borlak J. Metabolism of verapamil: 24 new phase I and phase II metabolites identified in cell cultures of rat hepatocytes by liquid chromatography–tandem mass spectrometry. *J Chromatogr B* 2003;798:265–274.
31. Aresta A, Carbonara T, Palmisano F, Zambonin CG. Profiling urinary metabolites of naproxen by liquid chromatography–electrospray mass spectrometry. *J Pharm Biomed Anal* 2006;41:1312–1316.
32. Bertrand M, Jackson P, Walther B. Rapid assessment of drug metabolism in the drug discovery process. *Eur J Pharm Sci* 2000;11(Suppl 2):S61–S62.
33. Ward KW, Smith BR. A comprehensive quantitative and qualitative evaluation of extrapolation of intravenous pharmacokinetic parameters from rat, dog and monkey to humans. I. Clearance. *Drug Metab Dispos* 2004;32:603–611.
34. Mattiuz EL, Ponsler GD, Barbuch RJ, Wood PG, Mullen JH, Shugert RL, Li Q, Wheeler WJ, Kuo F, Conrad PC, Sauer J-M. Disposition and metabolic fate of atomoxetine hydrocloride: pharmacokinetics, metabolism, and excretion in the Fischer 344 rats and beagle dog. *Drug Metab Dispos* 2003;31:88–97.
35. Wilkening S, Stahl F, Bader A. Comparison of primary human hepatocytes and hepatoma cell line HEPG2 with regard to their biotransformation properties. *Drug Metab Dispos* 2003;31:1035–1042.
36. Naritomi Y, Terashita S, Kagayama A, Sugiyama Y. Utility of hepatocytes in predicting drug metabolism: comparison of hepatic intrinsic clearance in rats and humans *in vivo* and *in vitro*. *Drug Metab Dispos* 2003;31:580–588.
37. Soars MG, Burchell B, Riley RJ. *In vitro* analysis of human drug glucoronidation and prediction of *in vivo* metabolic clearance. *J Pharmacol Exp Ther* 2001;301:382–390.
38. McGinnity DF, Soars MG, Urbanowicz RA, Riley RJ. Evaluation of fresh and cryopreserved hepatocytes as *in vitro* drug metabolism tools for the prediction of metabolic clearance. *Drug Metab Dispos* 2004;32:1247–1253.
39. Iwatsubo T, Hirota N, Ooide T, Suzuki H, Shimada N, Chiba K, Ishizaki T, Green CE, Tyson CA, Sugiyama Y. Prediction of *in vivo* drug metabolism in the human liver from *in vitro* metabolism data. *Pharmacol Ther* 1997;73:147–171.
40. Abelo A, Anderson TB, Antonsson M, Naudot AK, Skanberg I, Weidolf L. Stereoselective metabolism of omeprazole by human cytochrome P450 enzymes. *Drug Metab Dispos* 2000;28:966–972.
41. Steinmeyer A. The hit-to-lead process at Schering AG: strategic aspects. *ChemMedChem* 2006;1:31–36.
42. Timm U, Birnböck H, Erdin R, Hopfgartner G., Zumbrunnen R. Determination of oral platelet aggregation inhibitor Sibrafiban® in rat, dog and human plasma utilising

HPLC–column switching with turbo ion spray single quadrupole mass spectrometry. *J Pharm Biomed Anal* 1999;21:151–163.

43. Wang XJ, Jin YX, Ying JY, Zeng S, Yao TW. Determination of rutin deca(H-) sulfate sodium in rat plasma using ion-pairing liquid chromatography after ion-pairing solid-phase extraction. *J Chromatogr B* 2006;833:231–235.
44. Wang PG, Wei JS, Kim G, Chang M, El-Shourbagy T. Validation and application of a high-performance liquid chromatography–tandem mass spectrometric method for simultaneous quantification of lopinavir and ritonavir in human plasma using semi-automated 96-well liquid–liquid extraction. *J Chromatogr A* 2006;1130:302–307.
45. Dumasia MC. *In vivo* biotransformation of metoprolol in the horse and on-column esterification of the aminocarboxylic acid metabolite by alcohols during solid phase extraction using mixed mode columns. *J Pharm Biomed Anal* 2006;40:75–81.
46. Oruc EE, Kocyigit-Kaymakcioglu B, Yilmaz-Demircan F, Gurbuz Y, Kalaca S, Kucukguzel SG, Ulgen M, Rollas S. An high performance liquid chromatographic method for the quantification of cotinine in the urine of preschool children. *Pharmazie* 2006; 61:823–827.
47. Ong VS, Cook KL, Kosora CM, Brubaker WF. Quantitative bioanalysis: an integrated approach for drug discovery and development. *Int J Mass Spectrom* 2004;238:139–152.
48. Constanzer ML, Chavez-Eng CM, Fu I, Woolf EJ, Matuszewski BK. Determination of dextromethorphan and its metabolite dextrorphan in human urine using high performance liquid chromatography with atmospheric pressure chemical ionization tandem mass spectrometry: a study of selectivity of a tandem mass spectrometric assay. *J Chromatogr B* 2005;816:297–308.
49. Yang Z. Online hyphenated liquid chromatography–nuclear magnetic resonance spectroscopy–mass-spectrometry for drug metabolite and nature product analysis. *J Pharm Biomed Anal* 2006;40:516–527.
50. Sidelmann UG, Bjornsdottir I, Shockcor JC, Hansen SH, Lindon JC, Nicholson JK. Directly coupled HPLC-NMR and HPLC-MS approaches for the rapid characterisation of drug metabolites in urine: application to the human metabolism of naproxen. *J Pharm Biomed Anal* 2001;24:569–579.
51. Vanderhoeven SJ, Lindon JC, Troke J, Nicholson JK, Wilson ID. NMR spectroscopic studies of the transacylation reactivity 1-β-*O*-acyl glucuronide. *J Pharm Biomed Anal* 2006;41:1002–1006.
52. Dear GJ, Patel N, Kelly PJ, Webber L, Yung M. TopCount coupled to ultra-performance liquid chromatography for the profiling of radiolabeled drug metabolites in complex biological samples. *J Chromatogr B* 2006;844:96–103.
53. Gemesi LI, Kapas M, Szeberenyi SZ. Application of LC-MS analysis to the characterisation of the *in vitro* and *in vivo* metabolite profiles of RGH-1756 in the rat. *J Pharm Biomed Anal* 2001;24:877–885.
54. Mohri K, Okoda K, Benet LZ. Stereoselective metabolism of benoxaprofen in rats. Biliary excretion of benoxaprofen taurine conjugate and glucuronide. *Drug Metab Dispos* 1998;26:332–337.
55. Unsalan S, Sancar M, Bekce B, Clark PM, Karagoz T, Izzettin FV, Rollas S. Therapeutic monitoring of isoniazid, pyrazinamide and rifampicin in tuberculosis patients using LC. *Chromatographia* 2005;61:595–598.
56. Nobilis M, Vybiralova Z, Sladkova K, Lisa M, Holcapek M, Kvetina J. High-performance liquid-chromatographic determination of 5-aminosalicylic acid and its metabolites in bood plasma. *J Chromatogr A* 2006;1119:299–308.
57. Zhang JY, Yuan JJ, Wang YF, Roy H, Bible JR, Breau AP. Pharmacokinetics and metabolism of a COX-2 inhibitor, valdecoxib in mice. *Drug Metab Dispos* 2003;31:491–501.

58. Kamata T, Nishikawa M, Katagi M, Tsuchihashi H. Optimized glucuronide hydrolysis for the detection of psilocin in human urine samples. *J Chromatogr B* 2003; 796:421–427.
59. Groff M, Riffel K, Song H, Lo M-W. Stabilization and determination of a PPAR agonist in human urine using automated 96-well liquid–liquid extraction and liquid chromatography/tandem mass spectrometry. *J Chromatogr B* 2006;842:122–130.
60. Wren SAC. Peak capacity in gradient ultra performance liquid chromatography (UPLC). *J Pharm Biomed Anal* 2005;38:337–343.
61. Li R, Dong L, Huang J. Ultra performance liquid chromatography–tandem mass spectrometry for the determination of epirubicin in human plasma. *Anal Chim Acta* 2005;546:167–173.
62. Johnson KA, Plumb R. Investigation of the human metabolism of acetaminophen using UPLC and exact mass oa-TOF MS. *J Pharm Biomed Anal* 2005;39:805–810.
63. Yu W, Dener JM, Dickman DA, Grothaus P, Ling Y, Liu L, Havel C, Malesky K, Mahajan T, O'Brain C, Shelton EJ, Sperandio D, Tong Z, Yee R, Mordenti JJ. Identification of metabolites of the tryptase inhibitor CRA-9249: observation of a metabolite derived from an unexpected hydroxylation pathway. *Bioorg Med Chem Lett* 2006;16:4053–4058.
64. Funhoff EG, Salzmann J, Bauer U, Witholt B, Blein JB. Hydroxylation and epoxidation reactions catalyzed by CYP153 enzymes. *Enzyme Microbial Technol* 2007;40:806–812.
65. Medici R, Lewkowicz ES, Iribarren AM. Microbial synthesis of 2,6-diaminopurine nucleosides. *J Mol Catal B: Enzymatic* 2006;39:40–44.
66. Keppler AF, Porto ALM, Schoenlein-Crusius IH, Comasseto JV, Andrade LH. Enzymatic evaluation of different *Aspergillus* strains by biotransformation of cyclic ketones. *Enzyme Microbial Technol* 2005;36:967–975.
67. Bridges JW, Kibby MR, Williams RT. The structure of the glucuronide of sulphadimethoxine formed in man. *Biochem J* 1965;96:829–836.
68. Mey U, Wachsmuth H, Breyer-Pfaff U. Conjugation of the enantiomers of ketotifen to four isomeric quarternary ammonium glucuronides in humans *in vivo* and in liver microsomes. *Drug Metab Dispos* 1999;27:1281–1292.
69. Senn-Bilfinger J, Ferguson JR, Holmes MA, Lumbard KW, Huber R, Zech K, Hummel RP, Zimmermann PJ. Glucuronide conjugates of Soraprazan (BY359), a new potassium-competitive acid blocker (P-CAB) for the treatment of acid-related diseases. *Tetrahedron Lett* 2006;47:3321–3323.
70. Kucukguzel I, Ulgen M, Gorrod JW. *In vitro* hepatic metabolism of *N*-benzyl-*N*-methylaniline. *Il Farmaco* 1999;54:331–337.
71. He M, Rettie AE, Neal J, Trager WF. Metabolism of sulfinpyrazone sulfide and sulfinpyrazone by human liver microsomes and cDNA-expressed cytochrome p450s. *Drug Metab Dispos* 2001;29:701–711.
72. Mshvidobadze EV, Vasilevsky SF, Elguero J. A new route to pyrazolo[3,4-c] and [4,3-c]pyridinones via heterocyclization of *vic*-substituted hydroxamic acids of acetylenylpyrazoles. *Tetrahedron* 2004;60:11875–11878.
73. Duhart BT, Zhang JD, Deck J, Freeman JP, Cerniglia CE. Biotransformation of protriptyline by filamentous fungi and yeasts. *Xenobiotica* 1999;29:733–746.
74. Li L, Liu R, Ye M, Hu X, Wang Q, Bi K, Guo D. Microbial, metabolism of evodiamine by *Penicillium janthinellum* and its application for metabolite identification in rat urine. *Enzyme Microbial Technol* 2006;39:561–567.
75. Lewis DFV, Wiseman A. A selective review of bacterial forms of cytochrome P450 enzymes. *Enzyme Microbial Technol* 2005;36:377–384.

76. Ricci LC, Commasseto JV, Andrabe LH, Capelari M, Cass QB, Porto ALM. Biotransformations of aryl alkyl sulfides by whole cells of white-rot Basidiomycetes. *Enzyme Microbial Technol* 2005;36:937–946.
77. Zhang D, Evans FE, Freeman JP, Duhart B Jr, Cerniglia CE. Biotransformation of amitriptyline by Cunninghamella elegans. *Drug Metab Dispos* 1995;23:1417–1425.
78. Soars MG, Mattiuz ED, Jackson DA, Kulanthaivel P, Ehlhardt WJ, Wrighton SA. Biosynthesis of drug glucuronides for use as authentic standards. *J Pharmacol Toxicol Methods* 2002;47:161–168.
79. Ethell BT, Riedel J, Englert H, Jantz H, Oekonomopulos R, Burchell B. Glucuronidation of HMR-1098 in human microsomes: evidence for the involvement of UGT1A1 in the formation of *S*-glucuronides. *Drug Metab Dispos* 2003;31:1027–1034.
80. Luo H, Hawes EM., McKay G, Midha KK. Synthesis and charecterization of quaternary ammonium-linked glucuronide metabolites of drugs with an aliphatic tertiary amine group. *J Pharm Sci* 1992;81:1079–1083.
81. Tonn GR, Kerr CR, Axelson JE. *In vitro* protein binding of propafenone and 5-hydroxypropafenone in serum, in solutions of isolated serum proteins, and to red blood cells. *J Pharm Sci* 1992;81:1098–1103.
82. Oruc EE, Rollas S, Kabasakal L, Uysal MK. The *in vivo* metabolism of 5-(4-nitrophenyl)-4-phenyl-2,4-dihydro-3*H*-1,2,4-triazole-3-thione in rats. *Drug Metab Drug Interact* 1999;15:127–140.
83. Degelman CJ, Ebner T, Ludwig E, Happich S, Schildberg FW, Koebe HG. Protein binding capacity *in vitro* changes metabolism of substrates and influences the predictability of metabolic patways *in vivo*. *Toxicol In Vitro* 2004;18:835–840.
84. Acharya MR, Sparreboom A, Sausville EA, Conley BA, Doroshow JH, Venitz J, Figg WD. Interspecies differences in plasma protein binding of MS-275, a novel histon deacetylase inhibitor. *Cancer Chemother Pharmacol* 2006;57:275–281.
85. Yang XX, Hu ZP, Chan SY, Zhou SF. Monitoring drug–protein interaction. *Clin Chim Acta* 2006;365:9–29.
86. Yamazaki K, Kanaoka M. Computational prediction of the plasma protein-binding present in diverse pharmaceutical compounds. *J Pharm Sci* 2004;93:1480–1494.
87. Bailey MJ, Dickinson RG. Acyl glucuronide reactivity in perspective: biological consequences. *Chem-Biol Interact* 2003;145:117–137.
88. Park K, Williams DP, Naisbitt DJ, Kitteringham NR, Pirmohamed M. Investigation of toxic metabolites during drug development. *Toxicol Appl Pharmacol* 2005;207:425–434.
89. Day SH, Mao A, White R, Schulz-Utermoehl T, Miller R, Beconi MG. A semi-automated method measuring the potential for protein covalent binding in drug discovery. *J Pharmacol Toxicol Methods* 2005;52:278–285.
90. Dahms M, Spahn-Langguth H. Covalent binding of acidic drugs via reactive intermediates: detection of benoxaprofen and flunoxaprofen protein adducts in biological material. *Pharmazie* 1996;51:874–881.
91. Dong JQ, Liu J, Smith PC. Role of benoxaprofen acyl glucuronides in covalent binding to rat plasma and liver proteins in vivo. *Biochem Pharmacol* 2005;70:937–948.
92. Grillo MP, Benet LZ. Studies on the reactivity of clofibryl-*S*-acyl CoA thioester with glutathione *in vitro*. *Drug Metab Dispos* 2002;30:55–62.
93. Fan PW, Bolton JL. Bioactivation of tamoxifen to metabolite E quinone methide: reaction with glutathione and DNA. *Drug Metab Dispos* 2001;29:891–896.
94. Muldrew KL, James LP, Coop L, McCullough SS, Hendrickson HP, Hinson JA, Mayeux PR. Determination of acetaminophen–protein in mouse liver and serum and human

serum after hepatotoxic doses of acetaminophen using high-performance liquid chromatography with electrochemical detection. *Drug Metab Dispos* 2002;30:446–451.

95. Bu HZ, Kang P, Deese AJ, Zhao P, Pool WF. Human *in vitro* glutathione and protein adducts of carbamazepine-10,11-epoxide, a stable and pharmacologically active metabolite of carbamazepine. *Drug Metab Dispos* 2005;33:1920–1924.
96. Castillo M, Lam YW, Dooley MA, Stahl E, Smith PC. Disposition and covalent binding of ibuprofen and its acyl glucuronide in the elderly. *Clin Pharmacol Ther* 1995; 57:636–644.
97. Williams DP, Pirmohamed M, Naisbitt DJ, Maggs JL, Park BK. Neutrophil cytotoxicity of the chemically reactive metabolite(s) of clozapine: possible role in agranulocytosis. *J Pharmacol Exp Ther* 1997;283:1375–1382.
98. Sadeque AJM, Fisher MB, Korzekwa KR, Gonzales FJ, Rettie AE. Human CYP2C9 and CYP2A6 mediate formation of the hepatotoxin 4-ene-valproic acid. *J Pharmacol Exp Ther* 1997;283:698–703.
99. Li AP. Preclinical *in vitro* screening assays for drug-like properties. *Drug Discov Today* 2003;2:179–185.
100. Flores JR, Nevado JJB, Salcedo AMC, Diaz MPC. Non-aqueous capillary zone electrophoresis method for the analysis of paroxetine, tamoxifen and their main metabolites in urine. *Anal Chim Acta* 2004;512:287–295.
101. Meng QH, Gauthier D. Simultaneous analysis of citalopram and desmethylcitalopram by liquid chromatography with fluorescence detection after solid-phase extraction. *Clin Biochem* 2005;38:282–285.
102. Francois-Bouchard M, Simonin G, Bossant M-J, Boursier-Neyret C. Simultaneous determination of ivabradine and its metabolites in human plasma by liquid chromatography–tandem mass spectrometry. *J Chromatogr B* 2000;745:261–269.
103. Djabarouti S, Boselli E, Allaouchiche B, Ba B, Nguyen AT, Gordien JB, Bernadou JM, Saux MC, Breilh D. Determination of levofloxacin in plasma, bronchoalveolar lavage and bone tissues by high-performance liquid chromatography with ultraviolet detection using a fully automated extraction method. *J Chromatogr B* 2004;799:165–172.
104. Rizzo M, Ventrice D, Sarro VD, Gitto R, Caruso R, Chimirri A. SPE-HPLC determination of new tetrahydroisoquinoline derivatives in rat plasma. *J Chromatogr B* 2005;821: 15–21.
105. Christensen EB, Andersen JB, Pedersen H, Jensen KG, Dalgaard L. Metabolites of [$^{14}$C]-5(2-ethyl-2*H*-tetrazol-5-yl)-1-methyl-1,2,3,6-tetrahydropyridine in mice, rats, dogs, and humans. *Drug Metab Dispos* 1999;27:1341–1349.
106. Titier K, Deridet E, Moore N. *In vivo* and *in vitro* myocardial binding of risperidone and 9-hydroxyrisperidone. *Toxicol Appl Pharmacol* 2002;180:145–149.
107. Oruc EE, Kabasakal L, Rollas S. The *in vivo* metabolism of 5-(4-nitrophenyl)-4-(2-phenylethyl)-2,4-dihydro-3*H*-1,2,4-triazole-3-thione in rats. *Eur J Drug Metab Pharmacokinet* 2003;28:113–118.
108. Gulerman NN, Oruc EE, Kartal F, Rollas S. *In vivo* metabolism of 4-fluorobenzoic acid [(5-nitro-2-furanyl)methylene]hydrazide in rats. *Eur J Drug Metab Pharmacokinet* 2000;25:103–108.
109. Kocyigit-Kaymakcioglu B, Rollas S, Kartal-Aricioglu F. *In vivo* metabolism of *N*-phenyl-*N′*-(3,5-dimethylpyrazole-4-yl)thiourea in rats. *Eur J Drug Metab Pharmacokinet* 2003;28(4):273–278.
110. Bekce B, Sener G, Oktav M, Ulgen M, Rollas S. *In vitro* and *in vivo* metabolism of ethyl 4-[(2-hydroxy-1-naphthyl)azo]benzoate. *Eur J Drug Metab Pharmacokinet* 2005; 30:91–97.

111. Kaymakcioglu BK, Oruc EE, Unsalan S, Kabasakal LEE, Rollas S. High-pressure liquid chromatographic analysis for identification of *in vitro* and *in vivo* metabolites of 4-phenethyl-5-[4-(1-(2-hydroxyethyl)3,5-dimethyl-4-pyrazolylazo)phenyl]-2,4-dihydro-3*H*-1,2,4-triazole-3-thione in rats. *J Chromatogr B* 2006;831:184–189.
112. Kaymakcioglu BK, Aktan Y, Suzen S, Gokhan N, Koyuncuoglu S, Erol K, Yesilada A, Rollas S. *In vivo* metabolism of 2-[1′-phenyl-3′-(3-chlorophenyl)-2′-propenylyden] hydrazino-3-methyl-4(3*H*)-quinazolinone in rats. *Eur J Drug Metab Pharmacokinet* 2005;3:255–260.
113. Oruc EE, Unsal O, Balkan A, Ozkanli F, Goren MZ, Terzioglu MZ, Rollas S. The *in vivo* metabolism of 3-oxo-5-benzylidene-6-methyl-(4*H*)-2-(benzoylmethyl)pyridazine in rats. *Eur J Drug Metabol Pharmacokinet* 2006;31:21–25.
114. Rollas S, Gulerman N, Erdeniz H. Synthesis and antimicrobial activity of some new hydrazones of 4-fluorobenzoic acid hydrazide and 3-acethyl-2,5-disubstitute-1,3,4-oxadiazolines. *Il Farmaco* 2002;57:171–174.
115. Sincula AA, Yalkowsky SH. Rationale for design of biologically reversible drug derivatives. Prodrugs. *J Pharm Sci* 1975;64:181–210.
116. Chen Y-L, Jong Y-J, Wu S-M. Capillary electrophoresis field-amplified sample stacking and electroosmotic flow suppressant for analysis of sulindac and its two metabolites in plasma. *J Chromatogr A* 2006;1119:176–182.
117. Testa B. Prodrug research: futile or fertile? *Biochem Pharmacol* 2004;68:2097–2106.
118. Chandrasekaran S, Al-Ghananeem AM, Riggs RM, Crooks PA. Synthesis and stability of two indomethacin prodrugs. *Bioorg Med Chem Lett* 2006;16:1874–1879.
119. Sriram D, Yogeeswari P, Srichakravarthy N, Bal TR. Synthesis of stavudine amino acid ester prodrugs with broad-spectrum chemotherapeutic properties for the effective treatment of HIV/AIDS. *Bioorg Med Chem Lett* 2004;14:1085–1087.
120. Wu Q, Wang M, Chen ZC, Lu DS, Lin XF. Enzymatic synthesis of metronidazole esters and their monosaccharide ester derivatives. *Enzyme Microbial Technol* 2006;39: 1258–1263.
121. Chari RVJ. Targeted delivery of chemotherapeutics: tumor-activated prodrug therapy. *Adv Drug Delivery Rev* 1998;31:89–104.
122. Denny WA. Prodrug strategies in cancer therapy. *Eur J Med Chem* 2001;36:577–595.
123. Senter PD, Springer CJ. Selective activation of anticancer prodrugs by monoclonal antibody-enzyme conjugates. *Adv Drug Delivery Rev* 2001;53:247–264.
124. Baselga J. Epidermal growth factor receptor pathway inhibitors. *Update Cancer Ther* 2006;I:299–310.
125. Smyth TP, O'Donnell ME, O'Connor MJ, Ledger JO. β-Lactamase-dependent prodrugs—recent developments. *Tetrahedron* 2000;56:5699–5707.
126. Tang S, June SM, Howell BA, Chai M. Synthesis of salicylate dendritic prodrugs. *Tetrahedron Lett* 2006;47:7671–7675.
127. Li X, Wu Q, Lv D-S, Lin X-F. Controllable synthesis of polymerizable ester and amide prodrugs of acyclovir by enzyme in organic solvent. *Bioorg Med Chem* 2006; 14:3377–3382.
128. Chandrasekar D, Sistla R, Ahmad FJ, Khar RK, Dwan PV. The development of folate–PAMAM dendrimer conjugates for targeted delivery of anti-arthritics and their pharmacokinetics and biodistribution in arthritic rats. *Biomaterials* 2006;28:504–512.
129. Najlah M, Freeman S, Attwood D, Emanuele DA. Synthesis, characterization and stability of dendrimer prodrugs. *Int J Pharm* 2006;308:175–182.
130. Emanuele AD, Jevprasesphant R, Penny J, Attwood D. The use of a dendrimer–propranolol prodrug to bypass efflux transporters and enhance oral bioavailability. *J Controlled Release* 2004;95:447–453.

131. Salamonczyk GM. Acyclovir terminated thiophosphate dendrimers. *Tetrahedron Lett* 2003;44:7449–7453.
132. Ansede JH, Thakker DR. High-throughput screening for stability and inhibitory activity of compounds toward cytochrome P450-mediated metabolism. *J Pharm Sci* 2004; 93:239–255.
133. Di L, Kerns EH, Gao N, Li SQ, Huang Y, Bourassa JL, Huryn DM. Experimental design on single-time point high-throughput microsomal stability assay. *J Pharm Sci* 2004;93:1537–1544.
134. Mandagere AK, Thompson TN, Hwang KK. Graphical model for estimating oral bioavailability of drugs in humans and other species from their Caco-2 permeability and *in vitro* liver enzyme metabolic stability rates. *J Med Chem* 2002;45:304–311.
135. Soars MG, Riley RJ, Findlay KAB, Coffey MJ, Riley RJ, Burchell B. Evidence for significant differences in microsomal drug glucuronidation by canine and human liver and kidney. *Drug Metab Dispos* 2001;29:121–126.
136. Di L, Kerns EH, Li SQ, Petusky SL. High throughput microsomal stability assay for insoluble compounds. *Int J Pharm* 2006;317:54–60.
137. Lorenzi PL, Landowski CP, Song X, Borysko KZ, Breitenbach JM, Kim JS, Hilfinger JM, Townsend LB, Drach JC, Amidon GL. Amino acid ester prodrugs of 2-bromo-5,6-dichloro-1-(beta-dribofuranosyl)benzimidazole enhance metabolic stability *in vitro* and *in vivo*. *J Pharmacol Exp Ther* 2005;314:883–890.
138. Lai F, Khojasteh-Bakht SC. Automated online liquid chromatographic/mass spectrometric metabolic study for prodrug stability. *J Chromatogr B* 2005;814:225–232.
139. Miners JO, Knights KM, Houston JB, Mackenzie PI. *In vitro–in vivo* correlation for drugs and others eliminated by glucuronidation in humans: pitfalls and promises. *Biochem Pharmacol* 2006;71:1531–1539.
140. Pryde JP, Williams RT. The biochemistry and physiology of glucuronic acid I. The structure of glucuronic acid of animal origin. *Biochem J* 1933;27:1197–1204.
141. Williams RT. Studies in detoxication II. (a) The conjugation of isomeric 3-menthanols with glucuronic acid and the asymmetric conjugation of *dl*-menthol and *dl*-isomenthol in the rabbit. (b) *d*-Isomenthylglucuronide, a new conjugated glucuronic acid. *Biochem J* 1938;32:1849–1855.
142. Ebner T, Heinzel G, Prox A, Beschke K, Wachsmuth H. Disposition and chemical stability of telmisartan 1-*O*-acylglucuronide. *Drug Metab Dispos* 1999;27:1143–1149.
143. Chen Z, Holt TG, Pivnichny JV, Leung K. A simple *in vitro* model to study the stability of acylglucuronides. *J Pharmacol Toxicol Methods* 2006;55:91–95.
144. Cannell GR, Vesey DA, Dickinson RG. Inhibition of proliferation of HT-29 colon adenocarcinoma cell by carboxylate NSAIDs and their acyl glucuronides. *Life Sci* 2001;70:37–48.
145. Shipkova M, Wieland E. Glucuronidation in therapeutic drug monitoring. *Clin Chim Acta* 2005;358:2–23.
146. Gad SC. Active drug metabolites in drug development. *Curr Opin Pharmacol* 2003; 3:98–100.
147. Fura A. Role of pharmacologically active metabolites in drug discovery and development. *Drug Discov Today* 2006;11:133–142.
148. Fura A, Shu YZ, Zhu M, Hanson RL, Roongta V. Discovering drugs through biological transformation: role of pharmacologically active metabolites in drug discovery. *J Med Chem* 2004;47:4339–4351.
149. Guengerich FP. Cytochrome P450s and other enzymes in drug metabolism and toxicity. *AAPS J* 2006;8:101–111.

150. Klein CE, Gupta E, Reid JM, Atherton PJ, Sloan JA, Pitot HC, Ratain MJ, Kastrissios H. Population pharmacokinetic model for irinotecan and two of its metabolites, SN-38 and SN-38 glucuronide. *Clin Pharmacol Ther* 2002;72:638–647.
151. Mak IT, Weglicki WB. Potent antioxidant properties of 4-hydroxyl-propranolol. *J Pharmacol Exp Ther* 2004;303:85–90.
152. Davis HR. Ezetimibe: first in a new class of cholesterol absorption inhibitors. *Int Cong Ser* 2004;1262:243–246.
153. Wong JW, Nisar U-R, Yuen KH. Liquid chromatographic method for the determination of plasma itraconazole and its hydroxy metabolite in pharmacokinetic/bioavailability studies. *J Chromatogr B* 2003;798:355–360.
154. Gasparro DM, Almeida DRP, Fülöp F. *Ab initio* multi-dimensional conformational analysis of *R*-(−)-desmethyldeprenyl: a potent neuroprotective metabolite of *R*-(−)-deprenyl. *J Mol Structure THEOCHEM* 2005;725:75–83.
155. Nave R, Meyer W, Fuhst R, Zech K. Formation of fatty acid conjugates of ciclesonide active metabolite in the rat lung after 4-week inhalation of ciclesonide. *Pulm Pharmacol Ther* 2005;18:390–396.
156. Eddershaw PJ, Beresford AP, Bayliss MK. ADME/PK as part of a rational approach to drug discovery. *Drug Discov Today* 2000;5:409–414.
157. Kerns EH, Kleintop T, Little D, Tobien T, Mallis L, Di L, Hu M, Hong Y, McConnell OJ. Integrated high capacity solid phase extraction–MS/MS system for pharmaceutical profiling in drug discovery. *J Pharm Biomed Anal* 2004;34:1–9.
158. Crivori P, Poggesi I. Computational approaches for predicting CYP-related metabolism properties in the screening of new drugs. *Eur J Med Chem* 2006;41:795–808.
159. Lipinski CA. Drug-like properties and the causes of poor solubility and poor permeability. *J Pharmacol Toxicol Methods* 2000;44:235–249.
160. Dickins M, Waterbeemd H. Simulation models for drug disposition and drug interactions. *Drug Discov Today Biosilico* 2004;2:38–45.
161. Davis AM, Riley RJ. Predictive ADMET studies, the challenges and The opportunities. *Curr Opin Chem Biol* 2004;8:378–386.
162. Huisinga W, Telgmann R, Wulkow M. The virtual laboratory to pharmacokinetics: design principles and concepts. *Drug Discov Today* 2006;11:800–805.
163. Ekins S, Waller CL, Swaan PW, Cruciani G, Wrighton SA, Wikel JH. Progress in predicting human ADME parameters *in silico*. *J Pharmacol Toxicol Methods* 2000;44:251–272.
164. Ekins S, Nikolsky Y, Nikolskaya T. Techniques: application of systems biology to absorption, distribution, metabolism, excretion and toxicity. *Trends Pharmacol Sci* 2005; 26(4):202–209.
165. Klopman G, Stefan L, Saiakhov RD. A computer model for the prediction of intestinal absorption in humans. *Eur J Pharm Sci* 2002;17:253–263.
166. Uetrecht J. Curreent trends in drug-induced autoimmunity. *Autoimmun Rev* 2005;4: 309–314.
167. Sams-Dodd F. Drug discovery. Selecting the optimal approach. *Drug Discov Today* 2006;11:465–472.

# *IN VITRO* EVALUATION OF METABOLIC DRUG–DRUG INTERACTIONS: SCIENTIFIC CONCEPTS AND PRACTICAL CONSIDERATIONS

ALBERT P. LI
*In Vitro ADMET Laboratories Inc., Columbia, Maryland*

**Contents**

1 Introduction
2 Mechanisms of Adverse Drug–Drug Interactions
3 Drug Metabolism
4 CYP Isoforms
5 Human *In Vitro* Experimental Systems for Drug Metabolism
  5.1 Hepatocytes
  5.2 Liver Postmitochondrial Supernatant (PMS)
  5.3 Human Liver Microsomes
  5.4 Recombinant P450 Isoforms (rCYP)
  5.5 Cytosol
6 Mechanisms of Metabolic Drug–Drug Interactions
7 Mechanism-Based Approach for the Evaluation of Drug–Drug Interaction Potential
8 Experimental Approaches for the *In Vitro* Evaluation of Drug–Drug Interaction Potential
  8.1 Study 1: Metabolic Phenotyping 1—Metabolite Identification
  8.2 Study 2: Metabolic Phenotyping 2—Identification of Major Metabolic Pathways
  8.3 Study 3: Metabolic Phenotyping 3—Identification of P450 Isoform Pathways (P450 Phenotyping)
  8.4 Study 4: CYP Inhibitory Potential

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

8.5 Study 5: Enzyme Induction Potential
8.6 Study 6: *In Vitro* Empirical Drug–Drug Interactions
9 Data Interpretation
9.1 Pathway Evaluation
9.2 P450 Inhibition
9.3 P450 Induction
10 Conclusion
References

## 1 INTRODUCTION

Simultaneous coadministration of multiple drugs to a patient is highly probable. A patient may be coadministered multiple drugs to allow effective treatment of a single disease (e.g., cancer, HIV infection) or multiple disease or disease symptoms. It is now known that drug–drug interactions may have serious, sometimes fatal, consequences. Serious drug–drug interactions have led to the necessity of a drug manufacturer to withdraw or limit the use of marketed drugs. Examples of fatal drug–drug interactions are shown in Table 1. As illustrated by the examples in Table 1, a major mechanism of adverse drug–drug interactions is the inhibition of the metabolism of a drug by a coadministered drug, thereby elevating the systemic burden of the affected drug to a toxic level.

Besides toxicity, loss of efficacy can also result from drug–drug interactions. In this case, the metabolic clearance of a drug is accelerated due to the inducing effects of a coadministered drug on drug metabolism. A well known example is the occurrence of breakthrough bleeding and contraceptive failures in women taking oral contraceptives who were coadministered the enzyme inducer rifampin [7]. Examples of drug–drug interactions leading to the loss of efficacy are shown in Table 2.

Estimation of drug–drug interaction potential is therefore an essential element of drug development. Screening for drug–drug interaction in early phases of drug development allows the avoidance of the development of drug candidates with high potential for adverse drug interactions. Estimation of drug–drug interaction potential is a regulatory requirement—it is required for New Drug Applications (NDAs) to the U.S. FDA [11]. In this chapter, the scientific principles, technologies, and experimental approaches for the preclinical evaluation of drug–drug interactions are reviewed.

## 2 MECHANISMS OF ADVERSE DRUG–DRUG INTERACTIONS

Adverse effects in a patient due to coadministration of multiple drugs can be due to pharmacological or pharmacokinetic drug–drug interactions defined as follows:

*Pharmacological Interactions.* Adverse effects that occur due to combined pharmacological activities, leading to exaggerated pharmacological effects. An example of pharmacological interaction is the serious, sometimes fatal, drop in blood pressure due to coadministration of nitroglycerin and Sildenafil [12].

**TABLE 1 Drugs that Have Been Withdrawn from the Market Due to Fatal Interactions with Coadministered Drugs**

| Drug–Drug Interaction | Mechanism of Interactions | References |
|---|---|---|
| Terfenadine–ketoconazole interaction, leading to fatal arrhythmia (torsades de pointes). Terfenadine was withdrawn from the market in January 1997 and replaced by a safer alternative drug (fexofenadine), which is the active metabolite of terfenadine. | Terfenadine is metabolized mainly by CYP3A4 and has been found to interact with CYP3A4 inhibitors (e.g., ketoconazole), leading to elevation of plasma terfenadine level that reached cardiotoxic levels. | 1–3; www.fda.gov/bbs/topics/answers/ans00853.html |
| Mibefradil interaction with multiple drugs, leading to serious adverse effects. Mibefradil interactions with statins has led to rhabdomyolysis. Mibefradil was withdrawn from the market in June 1998, less than a year after it was introduced to the market in August 1997. | Mibefradil is a potent CYP3A4 inhibitor known to elevate the plasma levels of over 25 coadministered drugs to toxic levels. Statins, especially simvastatin and cerivastatin, are known to cause rhabdomyolysis. | 4; www.fda.gov/bbs/topics/answers/ans00876.html |
| Sorivudine–5-fluorouracil (5-FU) interaction, leading to severe or fatal gastrointestinal and bone marrow toxicities. Soruvidine was withdrawn from the market in 1993. | Sorivudine inhibits dihydropyrimidine dehydrogenase, an enzyme pathway responsible for fluoropyrimidine metabolism. | 5 |
| Gemfibrozil–cerivastatin interaction, leading to rhabdomyolysis. Cerivastatin was withdrawn from the market in August 2001. | Inhibition of cerivastatin metabolism by gemfibrozil, apparently due to CYP2C8 inhibitory effects of gemfibrozil. | 6; www.fda.gov/medwatch/safety/2001/Baycol2.html |

*Pharmacokinetic Interactions.* Adverse effects that occur due to altered body burden of a drug as a result of a coadministered drug that can occur because of the ability of one drug to alter the absorption, distribution, metabolism, and excretion (ADME properties) of the coadministered drug. Of the ADME properties, drug metabolism represents the most important and prevalent mechanism for pharmacokinetic interactions.

## 3 DRUG METABOLISM

All drugs administered to a patient are subject to biotransformation. Orally administered drugs are first subjected to metabolism by the intestinal epithelium and, upon

**TABLE 2 Drug–Drug Interactions Leading to Loss of Efficacy**

| Drug–Drug Interaction | Mechanism | References |
|---|---|---|
| Oral contraceptive–rifampin interactions, leading to breakthrough bleeding and contraceptive failure | Rifampin accelerates the metabolism of the estrogenic component (e.g., 17 α-ethinylestradiol) of oral contraceptives via induction of the metabolizing enzymes (CYP3A4 and estrogen sulfotransferases) | 7, 8 |
| Cyclosporin–rifampin interaction, leading to rejection of transplanted organs | Rifampin induces CYP3A, leading to accelerated metabolic clearance of cyclosporine to nonimmunosuppressive level | 9 |
| St. John's wort (SJW) interactions with prescribed drugs, leading to loss of efficacy | SJW (*Hypericum perforatum*) is an herbal medicine found to contain ingredients that can induce CYP3A4, CYP2C9, CYP1A2, and various transporters, leading to clinically observed accelerated metabolic clearance and/or loss of efficacy of a large number of drugs including warfarin, phenprocoumon, cyclosporine, HIV protease inhibitors, theophylline, digoxin, and oral contraceptives. The incidents with SJW illustrate the importance of the evaluation of potential drug–drug interaction of herbal medicines. | 10 |

absorption into the portal circulation, are metabolized by the liver before entering the systemic circulation. While multiple tissues have certain degrees of biotransformation capacity, it is generally accepted that hepatic metabolism represent the most important aspect of drug metabolism.

Drug metabolism can be classified into the following major categories:

*Phase I Oxidation.* This generally is described as the addition of an oxygen atom (e.g., as a hydroxyl moiety) to the parent molecule. Phase I oxidation is carried out by multiple enzyme pathways, including the various isoforms of the cytochrome P450 (CYP) family and the non-P450 biotransformation enzymes such as flavin-containing monooxygenase (FMO) and monamine oxidase (MAO).

*Phase II Conjugation.* Phase II conjugation represents enzyme reactions that lead to the addition of a highly water-soluble molecule to the chemical that is being metabolized, leading to highly water-soluble "conjugates" that allow efficient excretion. Examples of phase II enzymes are UDP-glucuronosyl transferase (UGT), sulfotransferase (ST), and glutathione-*S*-transferase (GST). Conjugation reactions often occur with the hydroxyl moiety of the parent structure or with the oxidative metabolites.

The major drug-metabolizing enzymes and subcellular locations are summarized in Table 3.

**TABLE 3 Major Pathways for Drug Metabolism, Enzymes, Subcellular Locations, and *In Vitro* Experimental System Containing the Enzymes**[a]

| Major Classification | Enzyme | Subcellular Location | Representative *In Vitro* Experimental System |
|---|---|---|---|
| Phase I oxidation | Cytochrome P450 mixed function monooxygenases | Endoplasmic reticulum | Microsomes, S9, hepatocytes |
| | Monoamine oxidase | Mitochondria | Hepatocytes |
| | Flavin-containing monooxygenase | Endoplasmic reticulum | Microsomes, S9, hepatocytes |
| | Alchohol/aldehyde dehydrogenase | Cytosol | S9, hepatocytes |
| | Esterases | Cytosol and endoplasmic reticulum | Microsomes, S9, hepatocytes |
| Phase II conjugation | UDP-dependent glucuronyl transferase | Endoplasmic reticulum | Microsomes, S9, hepatocytes |
| | Phenol sulfotransferases; estrogen sulfotransferase | Cytosol | S9, hepatocytes |
| | N-acetyl transferase | Endoplasmic reticulum | Microsomes, S9, hepatocytes |
| | Soluble glutathione-S-transferases (GST) | Cytosol | S9, hepatocytes |
| | Membrane-bound GST | Endoplasmic retuculum | Microsomes, S9, hepatocytes |

[a]These enzymes are grouped into phase I oxidation and phase II conjugation enzymes, although it is now believed that such classification may not be possible for all drug-metabolizing enzymes. Representative *in vitro* experimental systems containing these enzymes are shown to guide the selection of the most relevant approach for specific enzyme pathways. It is apparent that intact hepatocytes represent the most complete *in vitro* system for drug metabolism studies as they contain all the key hepatic drug-metabolizing enzyme pathways.

## 4 CYP ISOFORMS

Cytochrome P450-dependent monooxygenases (CYP) are the drug-metabolizing enzymes often involved in metabolic drug–drug interactions. The CYP family is represented by a large number of isoforms, each having selectivity for certain chemical structures. The major hepatic human CYP isoforms are CYP1A2, CYP2A6, CYP2B6, CYP2C8, CYP2C9, CYP2C19, CYP2D6, CYP2E1, and CYP3A4. Of these isoforms, the CYP3A isoforms are the most important in drug metabolism. CYP3A isoforms (CYP3A4 and CYP3A5) collectively represent the most abundant hepatic CYP isoforms (approximately 26%), followed by CYP2C isoforms (approximately 17%). In terms of the isoforms involved in drug metabolism, CYP3 isoforms are known to be involved in the metabolism of the most number of drugs (approximately 33%), followed by CYP2C isoforms (approximately 25%) [13].

P450 isoforms are known to have specific substrates, inhibitors, and inducers (Table 4).

**TABLE 4 Major Human P450 Isoforms[a] Involved in Drug Metabolism**

| CYP Isoform | Substrate | Inhibitor | Inducer |
|---|---|---|---|
| CYP1A2 | Phenytoin | Furafylline | Omeprazole |
| CYP2A6 | Coumarin | Tranylcypromine | Rifampin |
| CYP2B6 | Bupropion | Ticlopidine | Rifampin |
| CYP2C8 | Taxol | Quercetin | Rifampin |
| CYP2C9 | Tolbutamide | Sulfaphenazole | Rifampin |
| CYP2C19 | S-mephenytoin | Omeprazole | Rifampin |
| CYP2D6 | Dextromethorphan | Quinidine | None |
| CYP2E1 | Chlorzoxazone | Diethyldithiocarbamate | Ethanol |
| CYP3A4 | Testosterone | Ketoconazole | Rifampin |

[a]The individual isoforms and examples of isoform-specific substrates, inhibitors, and inducers are shown.

## 5 HUMAN *IN VITRO* EXPERIMENTAL SYSTEMS FOR DRUG METABOLISM

Substantial species–species differences occur in drug metabolism pathways, especially for CYP isoforms. Because of the species–species differences, human *in vitro* hepatic experimental systems rather than nonhuman animals are viewed as the most relevant to the evaluation of xenobiotic properties, including human drug metabolism and metabolism-based drug–drug interactions [14–17]. The following are the commonly used *in vitro* experimental systems for the evaluation of metabolism-based drug–drug interactions.

### 5.1 Hepatocytes

Hepatocytes are the parenchymal cells of the liver, which are responsible for hepatic biotransformation of xenobiotics. Isolated hepatocytes represent the most physiologically relevant experimental system for drug metabolism studies as they contain all the major hepatic drug-metabolizing enzyme pathways that are not physically disrupted such as cell free fractions. Furthermore, the drug-metabolizing enzymes and cofactors in the hepatocytes are present at physiological concentrations. Freshly isolated hepatocytes and cryopreserved hepatocytes are generally believed to represent the most complete *in vitro* system for the evaluation of hepatic drug metabolism [18].

In the past, the use of human hepatocytes has been severely limited by their availability, as studies would be performed only if human livers were available for hepatocyte isolation. Furthermore, hepatocyte isolation from human livers is not a technology available to most drug metabolism laboratories. This limitation has been overcome in the past decade due to the advancements in the procurement of human livers for research, and the commercial availability of isolated human hepatocytes. The application of human hepatocytes in drug metabolism studies is also greatly aided by the successful cryopreservation of human hepatocytes to retain drug metabolism activities [8, 19, 20]. Recently, the usefulness of cryopreserved human hepatocytes was further extended through the development of technologies to cryopreserve human hepatocytes to retain their ability to be cultured as attached cultures (plateable cryopreserved hepatocytes), which can be used for longer term

**TABLE 5 Viability and Plateability (Ability of Hepatocytes to Be Cultured as Monolayer Cultures) of the Various Lots of Cryopreserved Human Hepatocytes[a]**

| Lot | Yield (cells/vial) | Viability (trypan blue) | Plating | Confluency |
|---|---|---|---|---|
| HU4003 | 4.5 × 106 | 86% | Yes | 100% |
| HU4001 | 6.0 × 106 | 80% | No | 20% |
| HU4004 | 6.0 × 106 | 80% | No | 30% |
| HU4000 | 7.2 × 106 | 93% | Yes | 100% |
| HU4013 | 7.3 × 106 | 92% | Yes | 75% |
| HU4016 | 6.2 × 106 | 81% | Yes | 100% |
| HU4021 | 5.4 × 106 | 89% | Yes | 70% |
| HU4022 | 5.5 × 106 | 91% | Yes | 80% |
| HU4026 | 5.85 × 106 | 91% | No | 10% |
| HU4027 | 5.9 × 106 | 92% | No | 30% |
| HU4028 | 3.2 × 106 | 83% | Yes | 50% |
| HU4023 | 2.1 × 106 | 89% | No | 20% |
| HU4029 | 6.0 × 106 | 90% | Yes | 80% |

[a]Hepatocytes manufactured by APSciences Inc. in partnership with CellzDirect Inc.

studies such as enzyme induction studies [21]. Examples of the viability and plateability of cryopreserved human hepatocytes prepared in our laboratory are shown in Table 5.

### 5.2 Liver Postmitochondrial Supernatant (PMS)

Liver PMS is prepared by first homogenizing the liver, and then centrifuging the homogenate at a speed of either 9000 *g* or 10,000 *g* to generate the supernatants S9 or S10, respectively. Liver PMS contains both cytosolic and microsomal drug-metabolizing enzymes but lacks mitochondrial enzymes.

### 5.3 Human Liver Microsomes

Liver microsomes are the 100,000 *g* pellet for the PMS. Microsome preparation procedures in general involve the homogenization of the liver, dilution of the homogenate with approximately 4 volumes of sample weight with a buffer (e.g., 0.1 M Tris-HCl, pH 7.4, 0.1 M KCl, 1.0 mM EDTA, 1.0 mM PMSF [22]) followed by centrifugation at 9000–14,000 *g* to remove nonmicrosomal membranes, and then at 100,000–138,000 *g* to pellet the microsomes [23]. Microsomes contain the smooth endoplasmic reticulum, which is the site of the major phase I oxidation pathway, the P450 isoforms, and esterases, as well as a major conjugating pathway, UGT.

### 5.4 Recombinant P450 Isoforms (rCYP)

These are microsomes derived from organisms transfected with genes for individual human P450 isoforms (e.g., bacteria, yeast, mammalian cells) [24–26] and therefore contain only one specific human isoform. The major human P450 isoforms involved in drug metabolism are available commercially as rCYP. This experimental system is widely used to evaluate the drug-metabolizing activities of individual P450 isoforms [15, 16].

**TABLE 6 Comparison of the Key *In Vitro* Drug-Metabolizing Experimental Systems: Liver Microsomes (Microsomes), Liver Postmitochondrial Supernatant (S9), Liver Cytosol (Cytosol), and Hepatocytes and Their Major Drug-Metabolizing Enzymes[a]**

| *In Vitro* System | P450 | MAO | UGT | ST | GST |
|---|---|---|---|---|---|
| Microsomes | + | − | +[b] | − | +[c] |
| S9 | + | − | +[b] | +[b] | + |
| Cytosol | − | − | −[b] | +[b] | +[d] |
| Hepatocytes | + | + | + | + | + |

[a]P450, cytochrome P450; MAO, monoamine oxidase; UGT, UDP-glucuronosyl transferase; ST, sulfotransferase; GST, glutathione-*S*-transferase.
[b]Activity of this drug-metabolizing enzyme requires the addition of specific cofactors, for instance, UDP-glucuronic acid (UDPGA) for UGT activity, and 3′-phosphoadenosine 5′-phosphosulfate (PAPS) for ST activity.
[c]Membrane-bound GST but not the soluble GST is found in the microsomes.
[d]Soluble GST but not membrane-bound GST are found in the cytosol.

### 5.5 Cytosol

The supernatant after the 100,000 *g* centrifugation for microsome preparation is the cytosol, which is practically devoid of all membrane associated enzymes. *N*-acetyl transferases, sulfotransferases, and dehydrogenases are examples of cytosolic enzymes. While drug–drug interaction studies are mainly studied using liver microsomes, there are cases of drug–drug interactions involving phase II pathways that can be studied using liver cytosol [27].

A comparison of the different *in vitro* experimental systems and their drug-metabolizing enzymes are shown in Table 6.

## 6 MECHANISMS OF METABOLIC DRUG–DRUG INTERACTIONS

Metabolic drug–drug interaction results from the alteration of the metabolic clearance of one drug by a coadministered drug. There are two major pathways of metabolic drug–drug interactions:

*Inhibitory Drug–Drug Interaction.* When one drug inhibits the drug metabolism enzyme responsible for the metabolism of a coadministered drug, the result is a decreased metabolic clearance of the affected drug, resulting in a higher than desired systemic burden. For drugs with a narrow therapeutic index, this may lead to serious toxicological concerns. Most fatal drug–drug interactions are due to inhibitory drug–drug interactions.

*Inductive Drug–Drug Interactions.* Drug–drug interactions can also be a result of the acceleration of the metabolism of a drug by a coadministered drug. Acceleration of metabolism is usually due to the induction of the gene expression, leading to higher rates of protein synthesis and therefore higher cellular content of the induced drug–metabolizing enzyme and a higher rate of metabolism of the substrates of the induced enzyme. Inductive drug–drug interactions can lead to a higher metabolic clearance of the affected drug, leading to a decrease in plasma concentration and loss of efficacy. Inductive drug–drug interactions can also lead to a higher systemic burden of metabolites, which, if toxic, may lead to safety concerns.

## 7 MECHANISM-BASED APPROACH FOR THE EVALUATION OF DRUG–DRUG INTERACTION POTENTIAL

Due to the realization that it is physically impossible to evaluate empirically the possible interaction between one drug and all marketed drugs, and that most drug-metabolizing enzyme pathways are well defined, a mechanism-based approach is used for the evaluation of drug–drug interaction potential of a new drug or drug candidate [15, 16, 28]. This mechanism-based approach is now also recommended by the U.S. FDA (www.fda.gov/cber/gdlns/interactstud.htm). The approach consists of the following major studies.

*Metabolic Phenotyping.* Metabolic phenotyping is defined as the identification of the major pathways involved in the metabolism of the drug in question. The reasoning is that if the pathways are known, one can estimate potential interaction of the drug in question with known inhibitors or inducers of the pathway.

*Evaluation of Inhibitory Potential for Drug-Metabolizing Enzymes.* The ability of the drug in question to inhibit the activities of known pathways for drug metabolism is evaluated. If a drug is an inhibitor of a drug-metabolizing enzyme pathway, it will have the potential to cause inhibitory drug interactions with coadministered drugs that are substrates of the inhibited pathway.

*Induction Potential for Drug-Metabolizing Enzymes.* The ability of the drug in question to induce drug-metabolizing enzyme activities is evaluated. If the drug in question is an inducer of a specific pathway, it will have the potential to cause inductive drug interactions with coadministered drugs that are substrates of the induced pathway.

## 8 EXPERIMENTAL APPROACHES FOR THE *IN VITRO* EVALUATION OF DRUG–DRUG INTERACTION POTENTIAL

Because of the known species–species differences in drug metabolism, it is now believed that *in vitro*, human-based, experimental systems are more appropriate than nonhuman animal models for the evaluation of drug–drug interactions. *In vitro* positive findings are usually confirmed with *in vivo* clinical studies. The typical preclinical studies for drug–drug interactions [15, 16, 28] (www.fda.gov/cber/gdlns/interactstud.htm) are described next.

### 8.1 Study 1: Metabolic Phenotyping 1—Metabolite Identification

The objective of this study is to identify the major metabolites of the drug in question. For this study, the drug in question is incubated with an appropriate *in vitro* metabolic system to allow the formation of metabolites [15, 16]. Metabolites are then identified using analytical chemical approaches. The *in vitro* experimental system of choice is human hepatocytes, with high performance liquid chromatography/mass spectrometry (HPLC/MS) or tandem mass spectrometry (HPLC/MS/MS) as the most convenient analytical tool to identify the metabolites.

The metabolites are generally identified as metabolites of phase I oxidation or phase II conjugation. If phase I oxidation is concluded as the major pathway for the

oxidative metabolism of the drug, Experiment 2 will be performed to evaluate which of the several oxidative pathways are involved. Phase II conjugation pathways can generally be identified by the identity of the metabolite and subsequent experiments to further identify the pathways may not be necessary. For instance, if the metabolite is a glucuronide, UGT can be identified as the enzyme involved.

A typical experimental design is as follows:

- *In vitro System.* Cryopreserved human hepatocytes pooled from two donors (male, female).
  - Three drug concentrations: 1, 10, and 100 μM.
  - Hepatocyte concentration: 0.5–1.0 million hepatocytes per mL.
  - Three incubation times: 1, 2, and 4 hours (suspension culture); up to 24 hours (attached culture).
  - Incubation in 24 well plates at 37 °C.
  - Organic solvent (e.g., acetonitrile) to terminate reaction and to extract medium and intracellular metabolites.
  - Stored frozen until analysis.
- *Analytical Chemistry.* HPLC-MS/MS.
  - Quantification of disappearance of parent chemical in all samples.
  - Identification of metabolites from 100 μM samples.
  - Detection of metabolites in 1 and 10 μM samples.

### 8.2 Study 2: Metabolic Phenotyping 2—Identification of Major Metabolic Pathways

If oxidative metabolites are found to be the major metabolites, it is necessary to evaluate which major oxidative pathways are involved in the metabolism. This is performed via the use of liver microsomes and experimental conditions that would inhibit a specific pathway. The major pathways and experimental conditions are shown in Table 7.

As P450 pathways are considered the most important for metabolic drug–drug interactions, the study with the general P450 inhibitor, 1-aminobenzotriazole (ABT)

**TABLE 7 Experimental Conditions to Reduce the Activity of the Major Drug-Metabolizing Enzyme Pathways (P450 Isoforms (CYP), Flavin-Containing Monooxygenases (FMOs), and Monoamine Oxidase (MAO)) Using the *In Vitro* Experimental Systems for Drug Metabolism (Liver Microsomes (Microsomes), Postmitochondrial Supernatant (S9), and Hepatocytes)**

| *In Vitro* System | Condition | Inactivated Pathway(s) |
|---|---|---|
| Microsomes | NADH omission | CYP, FMO |
| Microsomes or hepatocytes | 1-Aminobenzotriazole treatment | CYP |
| Microsomes | Heat (45 °C) inactivation | FMO |
| S9 | Pargyline treatment | MAO |

*Source*: Adapted from http://www.fda.gov/cder/guidance/6695dft.pdf.

is the one that should be performed. ABT is known to inhibit all eight human P450 isoforms involved in drug metabolism [29]. Inhibition of metabolism of a test article by ABT would indicate that the test article is metabolized by the P450 pathway. A typical study with ABT is as follows:

- Human liver microsomes (0.5 mg protein/mL).
- Experiment 1: Evaluation of experimental conditions for the accurate quantification of metabolic clearance.
  - Incubation with three concentrations of test article (e.g., 0.1, 1, 10 μM) and three incubation times (e.g., 15, 30, and 60 minutes).
  - Quantification of test article disappearance.
- Experiment 2: Reaction phenotyping.
  - Incubation with one concentration of the test article at one incubation time (chosen from Experiment 1) in the presence and absence of three concentrations of ABT (100, 200, and 500 μM).
  - Quantification of test article disappearance and evaluation of the effects of ABT treatment.

### 8.3 Study 3: Metabolic Phenotyping 3—Identification of P450 Isoform Pathways (P450 Phenotyping)

If ABT is found to inhibit the metabolism of the drug or drug candidate in Study 2, P450 metabolism is ascertained. The next step is to identify which P450 isoforms are involved in the metabolism, a process termed P450 phenotyping [30, 31]. There are several major approaches for this study.

***Liver Microsome and Isoform-Selective Inhibitors*** In this experiment, the test article is incubated with human liver microsomes in the presence and absence of individual selective inhibitors for the eight major CYP isoforms. The ability of an inhibitor to inhibit metabolism of the test article would indicate that the pathway inhibited by the inhibitor is involved in metabolism. For instance, if ketoconazole, a potent CYP3A4 inhibitor, is found to inhibit the metabolism of the test article, then CYP3A4 is concluded to be involved in the metabolism of the test article. It is also a common practice to assign the degree of involvement by the maximum percent inhibition. For instance, if the maximum inhibition, expressed as percentages of the total metabolism in the absence of inhibitor, by sulfaphenazole (CYP2C9 inhibitor) and ketoconazole (CYP3A4 inhibitor) are 20% and 80%, respectively, it can be concluded that the CYP2C9 is involved in 20% and CYP3A4 in 80% of the metabolism of the test article. It is important to realize that the inhibitors are isoform selective rather than isoform specific, so data interpretation must be performed carefully to avoid an inaccurate assignment of enzyme pathways [31]. It is always prudent to confirm the results of this study with results using a different approach (e.g., using rCYP).

***Incubation with Individual rCYPs*** In this experiment, individual rCYP isoforms are used to evaluate which P450 isoforms are involved in the metabolism [30]. The

test article is incubated with each rCYP and its disappearance quantified. A rCYP that leads to the disappearance of the test article would indicate that the isoform is involved in the metabolism of the test article. For instance, if rCYP2C19 incubation leads to the disappearance of the test article, then CYP2C19 is concluded to be involved in the metabolic clearance of the test article. It is important to realize that these studies are performed with a single P450 isoform and therefore lack competing enzyme pathways. Metabolism by a rCYP isoform may not be relevant *in vivo* because of higher affinity pathways.

***Correlation Study with Human Liver Microsomes*** In this experiment, the test article is incubated with multiple lots of human liver microsomes that have been previously characterized for the activities of the individual CYP isoforms [32]. The rate of metabolic clearance of the test article is then plotted against the CYP activities of the different lots of microsomes. A linear correlation between activity and rate of disappearance for a specific CYP would indicate that this pathway is involved in the metabolism of the test article. This study requires the evaluation of at least 10 liver microsome lots with well-distributed gradations of activities.

***Liver Microsome/Inhibitor Study Design*** In general, studies with liver microsomes are believed to be more relevant than studies with rCYP, as studies with individual rCYP does not allow competition in metabolism for isoforms with different affinities for the substrate and therefore may overemphasize the participation of low affinity pathways. It is important to use substrate concentrations similar to expected plasma concentrations. An artifactually high concentration would cause the substrate to be metabolized similarly by high and low affinity enzyme pathways [33]. Using liver microsomes with physiologically relevant substrate concentrations should provide the best results. A typical liver microsome experiment with inhibitors is as follows.

- Human liver microsomes (0.5 mg/mL).
- Experiment 1: Metabolic stability study.
  - Incubation with three concentrations of test article (e.g., 0.1, 1, 10 μM) and three incubation times (e.g., 15, 30, and 60 minutes).
  - Quantification of test article disappearance.
- Experiment 2: Reaction phenotyping.
  - Incubation with one concentration of the test article at one incubation time (chosen from Experiment 1) in the presence and absence of isoform-specific inhibitors.
  - Quantification of test article disappearance.

The isoform-specific inhibitors suggested by the U.S. FDA are shown in Table 8.

***Evaluation of CYP Isoform Contributions Using Both Liver Microsomes and rCYP Isoforms*** It is also possible to calculate the relative contribution of individual isoforms using data from both liver microsomes and rCYP isoforms using the following approach [34, 35]:

**TABLE 8 Preferred and Acceptable P450 Isoform-Specific Inhibitors Suggested by U.S. FDA in the September 2006 Draft Guidance Document for Drug–Drug Interaction Evaluation and Preferred Inhibitors Used in the *In Vitro* ADMET Laboratories (IVAL)**

| CYP | FDA Preferred Inhibitor | FDA Acceptable Inhibitor | IVAL Preferred Inhibitor |
|---|---|---|---|
| 1A2 | Furafylline | α-Napthoflavone | Furafylline |
| 2A6 | Tranylcypromine, methoxsalen | Pilocarpine, tryptamine | Tranylcypromine |
| 2B6 | | Ticlopidine, sertraline | Ticlopidine |
| 2C8 | Quercetin | Trimethoprim, gemfibrozil, rosiglitazone | Quercetin |
| 2C9 | Sulfaphenazole | Fluconazole | Sulfaphenazole |
| 2C19 | | Ticlopidine | Omeprazole |
| 2D6 | Quinidine | | Quinidine |
| 2E1 | | Diethyldithiocarbamate | Diethyldithiocarbamate |
| 3A4/5 | Ketoconazole, itraconazole | Troleandomycin, verapamil | Ketoconazole |

- The relative activity factor for individual isoforms (using isoform-specific substrates) is calculated first. This is necessary as each lot of liver microsomes would have different relative amounts of each P450 isoform. $V_{max}$ and $K_m$ values are determined for each isoform using isoform-specific substrates for both liver microsomes and rCYP isoforms. The relative activity factor (RAF) is calculated using the following equation:

$$RAF = \frac{(V_{max}/K_m \text{ of CYP in microsomes})}{V_{max}/K_m \text{ of rCYP}}$$

- Contribution of a specific CYP isoform to the metabolism of a test article is then calculated using the following equation:

$$\text{Contribution of CYP (\%)} = RAF \times \frac{V(\text{rCYP})}{V(\text{microsomes})}$$

### 8.4 Study 4: CYP Inhibitory Potential

The objective of this study is to evaluate if the drug or drug candidate in question is an inhibitor of a specific P450 isoform. This study can be performed with rCYP, human liver microsomes, and human hepatocytes.

***rCYP Studies*** rCYP studies represent the most convenient and rapid study for the evaluation of CYP inhibitory potential. As the study involves substrates that form metabolites that can be quantified by fluorescence, the laborious and time-consuming HPLC or LC/MS sample analysis is not required. For this reason, most drug development laboratories perform rCYP inhibition assays as a screen for P450

inhibitory potential of their drug candidates. The study involves the incubation of individual rCYP isoforms with the test article at various concentrations (e.g., seven concentrations plus solvent control) in triplicate, and a substrate that can be metabolized by the specific isoform. As the reaction contains only one isoform, isoform-specific substrates are not required to be used. The requirement is that the substrate generates metabolites that can be measured by a plate reader with the capability to quantify florescence.

***Liver Microsome Studies*** Liver microsomes represent the most appropriate experimental system for the evaluation of the interaction of a drug with P450 isoforms. For the evaluation of CYP inhibitory potential, the test article is incubated with liver microsomes in the presence of individual isoform-specific substrates. The isoform-specific substrates and the metabolites quantified are shown in Table 9.

***Human Hepatocyte Studies*** rCYP and human liver microsomes are cell-free systems, allowing direct interaction of the test article with the P450 isoforms. *In vivo*, the inhibitor is initially absorbed into the systemic circulation and then interacts with the enzymes after penetration through the hepatocyte plasma membrane. Once inside the cytoplasm, the inhibitor may be metabolized by phase I and/or phase II metabolism and/or actively transported out of the hepatocytes, for instance, via bile excretion. Furthermore, there may be transporters present to actively uptake the inhibitor. The result is that the intracellular concentration of the inhibitor may be substantially different from the plasma concentration. Results with rCYP and human liver microsomes may not be useful to estimate *in vivo* inhibitory effects based on plasma concentrations if the intracellular concentration of the inhibitor is not known.

The use of intact human hepatocytes may allow a more accurate extrapolation of *in vitro* results to *in vivo*. The study is performed using intact human hepatocytes incubated with isoform-specific substrate and the test article. The intact plasma membrane and the presence of all hepatic metabolic pathways and cofactors allow distribution and metabolism of the test article. The resulting inhibitory effect there-

**TABLE 9 P450 Isoform-Specific Substrates and Their Respective Metabolites**[a]

| CYP | Substrate | Metabolite |
|---|---|---|
| 1A2 | Phenacetin | Acetaminophen |
| | Ethoxyresorufin | Resorufin |
| 2A6 | Coumarin | 7-OH-coumarin |
| 2B6 | Bupropion | Hydroxypropion |
| 2C8 | Taxol | 6-α-Hydroxypaclitaxel |
| 2C9 | Tolbutamide | 4′-Hydroxytolbutamide |
| 2C19 | *S*-Mephenytoin | 4-OH-mephenytoin |
| 2D6 | Dextromethorphan | Dextrophan |
| 2E1 | Chlorzoxazone | 6-Hydroxychlorzoxazone |
| 3A4/5 | Testosterone | 6-β-Hydroxytestosterone |

[a]These substrates are used in *in vitro* experimental systems such as liver microsomes, liver S9, or hepatocytes in which multiple isoforms are expressed.

fore should be physiologically more relevant to the *in vivo* situation than results with cell-free systems.

It is recommended that inhibition studies with intact hepatocytes be performed if inhibitory effects of a drug or drug candidate have been observed with rCYP or liver microsomes to allow a more accurate prediction of the extent of *in vivo* inhibitory effects. Time-dependent inhibition of P450 can also be studied using intact human hepatocytes [36]. One precaution with the use of intact hepatocytes is to concurrently measure cytotoxicity. As dead hepatocytes are not active in drug metabolism, without cytotoxicity information, cytotoxic drug concentrations could be interpreted as inhibitory concentrations.

A recent advancement is to use intact hepatocytes suspended in whole human plasma for inhibition studies to allow correction for plasma protein binding [37]. As drugs *in vivo* are always in contact with 100% human blood, this is conceptually sound and therefore deserves further investigation on its general applicability. One disturbing finding in our laboratory is that testosterone, a compound that is readily metabolized *in vivo*, is not metabolized by intact human hepatocytes in whole plasma (A. P. Li, unpublished).

***$IC_{50}$, $K_i$, $K_{inact}$ and [I]/$K_i$ Determinations*** Enzyme inhibition data are often presented as $IC_{50}$, the concentration of the inhibitor to cause 50% inhibition at one chosen substrate concentration; $K_i$, the inhibition constant (dissociation constant from the inhibitor–enzyme complex) determined by enzyme kinetic analysis (e.g., Dixon plot); and $K_{inact}$, the time-dependent inhibition constant for mechanism-based inhibitors. $IC_{50}$ values can be estimated from the study described earlier. A positive inhibition, defined as dose-dependent inhibition, with the inhibited activity lower than 50% of that of the negative control, will require further experimentation to define $K_i$ for a better evaluation of *in vivo* inhibitory potential. Furthermore, study to determine $K_{inact}$ may be performed to evaluate if the inhibitor acts via covalent binding to the active site of the enzyme, leading to time-dependent irreversible inhibition.

$IC_{50}$ is generally determined by plotting the log of the relative activity (activity in the presence of the inhibitor as a percentage of the activity of the negative solvent control), and then estimating the concentration yielding 50% relative activity using linear regression analysis. $IC_{50}$ can also be calculated from the relationship between inhibitor concentrations and percent control activity with the aid of a nonlinear regression program such as SCIENTIST (Micromath, Salt Lake City, UT) [38].

$K_i$ can be determined using a Dixon plot with the reciprocal of the activity as the *y*-axis and inhibitor concentration as the *x*-axis. Results with at least two substrate concentrations below $V_{max}$ are plotted,with $K_i$ calculated as the negative of the *x*-intercept [39]. $K_i$ can also be estimated with the aid of nonlinear regression analysis software such as SYSTAT (SPPS, Inc., Chicago, IL) [40].

Most P450 inhibitors act via reversible (competitive or noncompetitive) mechanisms and their inhibitory potential can be estimated from their $IC_{50}$ or $K_i$ values. Some inhibitors are "mechanism-based" or "time-dependent" inhibitors, which can cause irreversible inhibition due to the formation of reactive metabolites by the CYP isoform, leading to covalent binding to the active site and thereby causing

irreversible inhibition of the affected enzyme molecule [41]. Irreversible inhibitors therefore will have prolonged inhibition of the enzyme even after clearance of the drug in question. $K_{inact}$ is a measure of the potency of such "mechanism-based" inhibitors.

$K_{inact}$ can be determined using the following approach [42]:

1. Plot the relative activity (activity in the presence of the inhibitor as a percentage of the activity of the solvent or negative control) versus time and determine the slope at each inhibitor concentration.
2. Plot (1/slope) versus (1/inhibitor concentration) (Kitz–Wilson plot). $K_{inact}$ is calculated as the reciprocal of the *y*-intercept, and $K_i$ as the negative of the reciprocal of the *x*-intercept.

$[I]/K_i$, the ratio of the anticipated or known steady-state plasma drug concentration to $K_i$, is generally used to determine the likelihood of clinical drug–drug interactions [43, 44]. A general rule of thumb suggested by the U.S. FDA (http://www.fda.gov/cder/guidance/6695dft.pdf) is as follows:

- $[I]/K_i < 0.1$: Unlikely to cause *in vivo* drug–drug interactions.
- $[I]K_i = 1$: Possible to cause *in vivo* drug–drug interactions.
- $[I]/K_i > 1$: Likely to cause *in vivo* drug–drug interactions.

$K_i$ is estimated by an experiment with varying inhibitor and substrate concentrations. A typical $K_i$ study is as follows:

- *In vitro* experimental system: rCYP, human liver microsomes or hepatocytes.
- Inhibitor concentrations: Five (ideally yielding 10–90% inhibition of activity).
- Substrate concentrations: Minimum of two for the Dixon plot; three is recommended.
- Timepoint: one (within the linear time course) if time course is known; multiple (e.g., 5, 10, and 15 minutes) if time course under the experimental conditions has not been established.
- $K_i$ is determined by a Dixon plot, plotting the reciprocal of activity versus inhibitor concentration. The negative of the *x*-coordinate value corresponding to the intercept of the plots for the low and high substrate concentrations is the $K_i$.

For mechanism-based inhibitors, $K_{inact}$ is estimated by an experiment with varying inhibitor concentration and preincubation time. A typical $K_{inact}$ study is as follows:

- *In vitro* experimental system: rCYP, human liver microsomes or hepatocytes.
- Preincubation time (preincubation of enzyme with inhibitor): Five values (e.g., 5, 10, 15, 20, 30 minutes).
- Inhibitor concentration: Five (ideally yielding 10–90% inhibition of activity).
- Substrate concentration: one.

- Substrate incubation time: one (within the linear time course) if time course is known; multiple (e.g., 5, 10, and 15 minutes) if time course under the experimental conditions has not been established.
- $K_{inact}$ is determined by the following approach:
  - Plot activity as a percentage of the solvent control versus time.
  - Estimate the first-order inactivation constants at each inhibitor concentration by multiplying the slope of the linear regression analysis by 2.303.
  - Determine $t_{1/2}$ of the inactivation reaction as $0.693/k$.
  - Plot the Kitz–Wilson plot of $t_{1/2}$ versus the reciprocal of the inhibitor concentration and estimate $K_{inact}$ as the $y$-intercept, and $K_i$ as the reciprocal of the $x$-intercept.

### 8.5 Study 5: Enzyme Induction Potential

Enzyme induction is a major mechanism for drug–drug interactions. Induction of a drug-metabolizing enzyme by one drug would lead to the enhanced metabolism of coadministered drugs that are substrates of the induced enzyme.

Experimental evaluation of enzyme induction involves the treatment of human hepatocytes for several days with the test article followed by evaluation of enzyme activities using P450 isoform-specific substrates [45, 46]. As freshly isolated hepatocytes possess endogenous activities, which may be the result of inducers present in the donor's systemic circulation, the isolated hepatocytes are cultured for 2–3 days to allow the P450 enzyme activities to return to a basal level. Testing for induction potential is initiated by treatment of the cultured hepatocytes for 2–3 days to allow full expression of the induced enzyme. Induction is generally evaluated by measuring enzyme activity, as activity represents the most relevant endpoint for drug–drug interaction. Both freshly isolated and plateable cryopreserved human hepatocytes can be used for the induction study [21, 47, 48].

As of this writing, all known inducers of P450 isoforms *in vivo* are inducers *in vitro* [21]. The known human P450 inducers are shown in Table 10.

**TABLE 10 Clinically Demonstrated Human Enzyme Inducers and Their Respective *In Vitro* Induction Results as Well as Their Association with Severe Hepatotoxicity**

| *In Vivo* Enzyme Inducer | *In Vitro* Human Hepatocyte Induction Finding | Severe Clinical Hepatotoxicity |
|---|---|---|
| Carbamazepine | + | + |
| Dexamethasone | + | – |
| Isoniazid | + | + |
| Omeprazole | + | + |
| Phenobarbital | + | + |
| Phenytoin | + | + |
| Rifampin | + | + |
| Rifapentine | + | – |
| Rifabutin | + | – |
| Troglitazone | + | + |
| St. John's wort | + | + |

The typical experimental procedures for an enzyme induction study are as follows:

- Day 0: Plate human hepatocytes (freshly isolated or plateable cryopreserved human hepatocytes).
- Day 1: Refresh medium.
- Day 2: Refresh medium.
- Day 3: Change medium to that containing test article, solvent control, or positive controls.

  Minimum of three test article concentrations, with the high concentration at least one order of magnitude greater than expected plasma concentration.

  If plasma concentration not known, evaluate concentrations ranging over at least two orders of magnitude (e.g., 1, 10, 100 μM).
- Day 4: Refresh treatment medium.
- Day 5: Refresh treatment medium.
- Day 6: Measure activity (*in situ* incubation with isoform-specific substrates).

The isoform-specific substrates described earlier for CYP inhibition studies are generally used for enzyme induction studies.

The known CYP inducers are now known to induce either CYP1A and/or CYP3A, with inducers of other inducible isoforms such as CYP2A6, CYP2C9, and CYP2C19 found also to be CYP3A inducers. For general enzyme induction evaluation for drug–drug interaction, it may be adequate to simply screen for CYP1A and CYP3A induction. If CYP3A induction is observed, then investigations into CYP2A6, CYP2C9, and CYP2C19 induction are warranted.

The two most common confounding factors for P450 induction studies are as follows:

1. *Inducers that Are Also Inhibitors.* The co-occurrence of P450 inhibition and induction (i. e., the compound is both an inhibitor and inducer) can confound induction results. Ritonavir is an example of a CYP3A4 inducer [49] that is also a potent CYP3A4 inhibitor [50]. The inhibitory effects can overcome any induction effects using activity as an endpoint. For the evaluation of enzyme induction potential of inhibitors, Western blotting for the amount of enzyme proteins would be most appropriate. Studies with mRNA expression would provide data to distinguish between induction of gene expression or protein stabilization as mechanisms. As in the case of ritonavir, induction effects persist after the clearance of the drug from the systemic circulation, leading to enhanced clearance of drugs that are substrates of the induced pathways. It is important to define the induction potential of a drug even if it is found to be an enzyme inhibitor.
2. *Cytotoxic Compounds.* Induction effects can be masked by the decrease of cell viability, as most induction assays quantify substrate metabolism *in situ* (in the same cell culture plate in which the cells are cultured) and assume that there is no change in cell number. Cytotoxicity evaluation therefore should always be performed concurrently with induction studies. In the presence of cytotoxicity, activity should be corrected by the viability for comparison with negative control activity to assess induction potential.

A compound is concluded to be an inducer if reproducible, statistically significant, and dose-dependent induction effects are observed. U.S. FDA recommends the use of the criterion of "40% or higher of the activity of positive controls" as a positive response (www.fda.gov/cber/gdlns/interactstud.htm).

### 8.6 Study 6: *In Vitro* Empirical Drug–Drug Interactions

The physiological significance of the findings based on the mechanistic approach may be substantiated by *in vitro* drug–drug interactions between frequently coadministered drugs that are likely to have interaction with the drug in question [28]. This is particularly important if the drug in question is either a CYP3A4 substrate or a CYP3A4 inhibitor. As CYP3A4 is now known to have different affinities for different substrates and inhibitors [51], the interaction potential for a drug and a particular coadministered drug may be substantially different from that estimated by using a surrogate substrate of CYP3A4.

This study can be performed with liver microsomes or hepatocytes. The use of hepatocytes probably would allow the development of data more relevant to humans *in vivo*.

## 9 DATA INTERPRETATION

The studies described previously allow one to develop data for the estimation of drug–drug interaction potential of the drug or drug candidate in question. Accurate prediction of *in vivo* effects is possible only through thorough and scientifically sound interpretation of the data. While every novel chemical structure will provide a unique set of data and therefore requires individualized data interpretation and/or further experimentation, the following guidelines can be used to aid the evaluation of the data generated.

### 9.1 Pathway Evaluation

Possible outcomes of a study are as follows:

1. The test article is not metabolized by liver microsomes or hepatocytes. This is indicated by the lack of metabolite formation and lack of parent disappearance in Studies 1 and 2. Hepatic metabolism is not involved in the metabolic clearance of the compound. There should be no concern with coadministered drugs that can alter drug-metabolizing enzyme activities.
2. The test article is metabolized but not metabolized by P450 isoforms. As P450-related drug–drug interactions are the most prevalent, non-P450 drug–drug interactions should be considered on a case-by-case basis. For instance, MAO interaction may be important if the drug in question may be coadministered with known MAO substrates or inhibitors. UGT substrates, for instance, may have drug interactions with UGT inhibitory drugs such as probenecid.
3. The test article is metabolized by a single P450 isoform. This represents the easiest data to interpret, albeit not a good scenario for a drug candidate. A drug that is metabolized predominantly by a single P450 isoform will very likely have drug–drug interactions with inhibitors of the isoform. The known cases of serious drug–drug interactions often involve a single P450 pathway,

with CYP3A4 being the most prominent. Drugs that have been withdrawn due to fatal drug–drug interactions are often CYP3A4 substrates or potent CYP3A4 inhibitors. Because of the role of CYP2C8 in the metabolism of statins, which are widely prescribed to combat hypercholesterolemia, CYP2C8 has become the second-most important isoform for drug–drug interactions. Cerivastatin, a CYP2C8 substrate, was withdrawn from the market in August 2001 after reports of fatal interactions with the CYP2C8 inhibitor gemfibrozil [52].

4. The test article is metabolized by multiple P450 isoforms. This is generally interpreted that the test article may not have serious interactions with a specific inhibitor of one of the P450 isoforms, as the metabolic clearance can be carried out by the unaffected pathways. However, there are examples of drugs that have been found to be metabolized by multiple pathways but later were found in clinical or postmarketing studies to have interactions with potent inhibitors of a specific pathway. An example is the antifungal terbinafine, which has been characterized using human liver microsomes and rCYP isoforms to be metabolized by multiple P450 isoforms: CYP1A2, CYP2C8, CYP2C9. CYP2C19, CYP2D6, and CYP3A4, leading to the conclusion that "the potential for terbinafine interaction with other drugs is predicted to be insignificant" [53]. In the same study, as terbinafine was a competitive inhibitor of CYP2D6, it was concluded that it would have interactions with CYP2D6 substrates. *In vivo* studies confirmed the CYP2D6 inhibitory effects as predicted by *in vitro* studies; however, it was also observed clinically that rifampin, a CYP3A4 inducer, caused a 100% increase in terbinafine clearance (www.fda.gov/medwatch/safety/2004/jan_PI/Lamasil_PI.pdf). One possible explanation of this is that, upon CYP3A4 induction, the total metabolism of terbinafine is greatly enhanced due to the high capacity of CYP3A4 for this substrate. It is therefore important to realize that if a drug is metabolized by multiple isoforms, it may still have significant drug interactions with inducers of isoforms with high capacity for the metabolism of the drug.

### 9.2 P450 Inhibition

The outcomes of P450 inhibition studies may include the following:

1. No inhibition is observed. If no inhibitory effects are observed with rCYP, microsomes, and hepatocytes, the substance in question is considered not to have the potential to cause inhibitory metabolic drug–drug interactions *in vivo*. As of now, there are no examples of *in vivo* enzyme inhibitors that are not inhibitors *in vitro*.
2. Significant inhibition is observed. A practical definition of significant inhibition is that the test article is found to cause dose-dependent and >50% inhibition of one or more P450 isoforms at the concentrations evaluated. The conclusion is that the test article is a potent inhibitor. As described earlier, the physiological significance is determined by the $[I]/K_i$ value, with any $[I]/K_i$ value of 0.1 or higher as possible or likely to cause *in vivo* drug–drug interactions. It is recommended that $[I]/K_i$ values obtained from cell-free systems (microsomes and rCYP) are confirmed with intact hepatocytes to aid an accurate prediction

of *in vivo* effects. If the results with hepatocytes are also determined to be significant, *in vivo* studies will need to be performed to estimate human *in vivo* drug–drug interaction potential.

3. No time-dependent inhibition is observed. The inhibitor is not a mechanism-based inhibitor.
4. Time-dependent inhibition is observed. The inhibitor is a time-dependent inhibitor. *In vivo* studies will need to be performed to further define its drug–drug interaction potential.
5. Additional safety concern. A time-dependent inhibitor may need to be further studied to define its hepatotoxic potential, as a number of time-dependent P450 inhibitors are found to cause idiosyncratic hepatotoxicity.

### 9.3 P450 Induction

The following outcomes may be observed:

1. No induction is observed. The substance evaluated is not an enzyme inducer if P450 inhibitory and cytotoxic potential are eliminated as confounding factors.
2. Induction is observed. The substance evaluated is observed to cause dose-dependent and physiologically significant induction (e.g., induced activity greater than twofold the negative control activity). If the doses found to be positive are within clinical plasma concentrations (e.g., within 10× of plasma $C_{max}$), *in vivo* studies may be needed to further define the test article's *in vivo* enzyme induction and the subsequent drug–drug interaction potential.
3. Additional safety concern. Enzyme inducers may need to be further evaluated for their hepatotoxic potential, as a large number of enzyme-inducing drugs are found to cause severe hepatotoxicity.

## 10 CONCLUSION

Drug–drug interactions can have serious adverse consequences and therefore should be evaluated accurately before a new drug is introduced to the human population. Due to the scientific advances in the understanding of key human drug-metabolizing pathways, and the availability of human *in vitro* systems for drug metabolism studies, human drug–drug interaction evaluations, especially drug metabolism-related interactions, can be performed rapidly and efficiently. A scientific, mechanism-based approach to evaluate drug–drug interactions remains the most appropriate approach:

1. Understanding of the major drug-metabolizing pathways in the metabolism of the drug or drug candidate in question will allow assessment of its potential interactions with existing drugs that are inhibitors or inducers of the pathways involved.
2. A careful and exhaustive evaluation of the inhibitory potential of the drug or drug candidate in question toward the major human drug metabolism enzymes

will allow assessment of its potential to cause interactions with existing drugs that are substrates of the inhibited enzymes.

3. Evaluation of induction potential of the drug or drug candidate in question for the inducible human drug-metabolizing enzymes will allow assessment of potential interactions with drugs that are substrates of the induced enzymes.

This approach is mainly applied toward P450 isoforms but can also be applied to non-P450 drug-metabolizing enzyme pathways. The next wave of major advances in drug–drug interactions is anticipated to be approaches for the evaluation of the interactions between drugs and drug transporters.

The success achieved with the scientific-based approaches in the evaluation of drug–drug interactions is a result of the extensive scientific research in the identification and characterization of drug-metabolizing enzymes, the definition of the mechanisms of metabolism-based drug–drug interactions, and the development, characterization, and intelligent application of the human-based *in vitro* experimental models for drug metabolism. Similar approaches should be adopted for the evaluation of other major adverse drug effects (e.g., idiosyncratic drug toxicity), which so far have eluded the routine drug safety evaluation approaches. It is through an open mind—a willingness to venture toward the development of a hypothesis, the testing of a hypothesis, and the development and adoption of approaches to investigate a problem based on the best science—that the field of drug safety evaluation can move forward.

## REFERENCES

1. Vazquez E, Whitfield L. Seldane warnings. *Posit Aware* 1997;8:12.
2. Carlson AM, Morris LS. Coprescription of terfenadine and erythromycin or ketoconazole: an assessment of potential harm. *J Am Pharm Assoc (Wash)* 1996;NS36:263–269.
3. Von Moltke LL, Greenblatt DJ, Duan SX, Harmatz JS, Wright CE, Shader RI. Inhibition of terfenadine metabolism *in vitro* by azole antifungal agents and by selective serotonin reuptake inhibitor antidepressants: relations to pharmacokinetic interactions *in vivo*. *J Clin Psychopharmacol* 1996;16:104–112.
4. Omar MA, Wilson JP. FDA adverse event reports on statin associated rhabdomyolysis. *Ann Pharmacother* 2002;36:288–295.
5. Diasio RB. Sorivudine and 5-fluorouracil; a clinically significant drug–drug interaction due to inhibition of dihydropyrimidine dehydrogenase. *Br J Clin Pharmacol* 1998;46: 1–4.
6. Ozdemir O, Boran M, Gokce V, Uzun Y, Kocak B, Korkmaz S. A case with severe rhabdomyolysis and renal failure associated with cerivastatin–gemfibrozil combination therapy—a case report. *Angiology* 2000;51:695–697.
7. Zhang H, Cui D, Wang B, Han YH, Balimane P, Yang Z, Sinz M, Rodriqus AD. Pharmacokinetic drug interactions involving 17alpha-ethinylestradiol: a new look at an old drug. *Clin Pharmacokinet* 2007;46:133–157.
8. Li AP, Hartman NR, Lu C, Collins JM, Strong JM. Effects of cytochrome P450 inducers on 17 alpha-ethinyloestradiol (EE2) conjugation by primary human hepatocytes. *Br J Clin Pharmacol* 1999;48:733–742.
9. Capone D, Aiello C, Santoro GA, Gentile A, Stanziale P, D'Alessandro R, Imperatore P, Basile V. Drug interaction between cyclosporine and two antimicrobial agents, josamycin

and rifampicin, in organ-transplanted patients. *Int J Clin Pharmacol Res* 1996;16: 73–76.

10. Henderson L, Yue QY, Berqquist C, Gerden B, Arlett P. St. John's wort (*Hypericum perforatum*): drug interactions and clinical outcomes. *Br J Clin Pharmacol* 2002;54: 349–356.
11. Huang SM, Lesko LJ, Williams RL. Assessment of the quality and quantity of drug–drug interaction studies in recent NDA submissions: study design and data analysis issues. *J Clin Pharmacol* 1999;39:1006–1014.
12. Schalcher C, Schad K, Brunner-La Rocca HP, Schindler R, Oechslin E, Scharf C, Suetsch G, Bertel O, Kiowski W. Interaction of sildenafil with cAMP-mediated vasodilation *in vivo*. *Hypertension* 2003;40:763–767.
13. Guengerich FP. Cytochrome P450s and other enzymes in drug metabolism and toxicity. *AAPS J* 2006;8:E101–E111.
14. Li AP, Maurel P, Gomez-Lechon MJ, Cheng LC, Jurima-Romet M. Applications of primary human hepatocytes in the evaluation of P450 induction. *Chem Biol Interact* 1997;107:5–16.
15. Li AP. Screening for human ADME/Tox drug properties in drug discovery. *Drug Discov Today* 2001;6:357–366.
16. Li AP. *In vitro* approaches to evaluate ADMET drug properties. *Curr Top Med Chem* 2004;4:701–706.
17. MacGregor JT, Collins JM, Sugiyama Y, Tyson CA, Dean J, Smith L, Andersen M, Curren RD, Houston JB, Kadlubar FF, Kedderis GL, Krishnan K, Li AP, Parchment PE, Thummel K, Tomaszewski JE, Ulrich R, Vickers AE, Wrighton SA. *In vitro* human tissue models in risk assessment: report of a consensus-building workshop. *Toxicol Sci* 2001;59:17–36.
18. Hewitt NJ, Lechon MJ, Houston JB, Hallifax D, Brown HS, Maurel P, Kenna JG, Gustavsson L, Lohmann C, Skonberg C, Huillouzo A, Tuschi G, Li AP, Elcluyse E, Groothuis GM, Hengstler JG. Primary hepatocytes: current understanding of the regulation of metabolic enzymes and transporter proteins, and pharmaceutical practice for the use of hepatocytes in metabolism, enzyme induction, transporter, clearance, and hepatoxicity studies. *Drug Metab Rev* 2007;39:159–234.
19. Li AP, Gorycki PD, Hengstler JG, Kedderis GL, Koebe HG, Rahmani R, de Sousas G, Silva JM, Skett P. Present status of the application of cryopreserved hepatocytes in the evaluation of xenobiotics: consensus of an international expert panel. *Chem Biol Interact* 1999;121:117–123.
20. Li AP, Lu C, Brent JA, Pham C, Fackett A, Ruegg CE, Silber PM. Cryopreserved human hepatocytes: characterization of drug-metabolizing enzyme activities and applications in higher throughput screening assays for hepatotoxicity, metabolic stability, and drug–drug interaction potential. *Chem Biol Interact* 1999;121:17–35.
21. Li AP. Human hepatocytes: isolation, cryopreservation and applications in drug development. *Chem Biol Interact* 2007;168:16–29.
22. Raucy J, Lasker JM. Isolation of P450 enzymes from human livers. *Methods Enzymol* 1991;206:577–594.
23. Nelson AC, Huang W, Moody DE. Human liver microsome preparation: impact on the kinetics of L-α-acetylmethadol (LAAM) *N*-demethylation and dextromethorphan *O*-demethylation. *Drug Metab Dispos* 2001;29:319–325.
24. Barnes HJ, Arlotto MP, Waterman MR. Expression and enzymatic activity of recombinant cytochrome P450 17-alpha-hydroxylase in *Escherichia coli*. *Proc Natl Acad Sci USA* 1991;88:5597–5601.
25. Friedberg T, Pritchard MP, Bandera M, Hanlon SP, Yao D, McLaughlin LA, Ding S, Burhell B, Wolf CR. Merits and limitations of recombinant models for the study of human

P450-mediated drug metabolism and toxicity—an intralaboratory comparison. *Drug Metab Rev* 1999;31:523–544.

26. Donato MT, Jimenez N, Castell JV, Gomez-Lechon J. Fluorescence-based assays for screening nine cytochrome P450 (P450) activities in intact cells expressing individual human P450 enzymes. *Drug Metab Dispos* 2004;32:699–706.
27. Vtric F, Haefeli WE, Drewe, J, Krahenbuhl S, Wenk M. Interaction of ibuprofen and probenecid with metabolizing enzyme phenotyping procedures using caffeine as the probe drug. *Br J Clin Pharmacol* 2003;55:191–198.
28. Li AP. Scientific basis of drug-drug interactions: mechanism and preclinical evaluation. *Drug Inf J* 1998;32:657–664.
29. Emoto C, Murase S, Sawada Y, Jones BC, Iwasaki K. *In vitro* inhibitory effects of 1-aminobenzotriazole on drug oxidations catalyzed by human cytochrome P450 enzymes: a comparison with SKF-525A and ketoconazole. *Drug Metab Pharmacokinet* 2003;18: 287–295.
30. Rodriques AD. Integrated cytochrome P450 reaction phenotyping: attempting to bridge the gap between cDNA-expressed cytochromes P450 and native human liver microsomes. *Biochem Pharmacol* 1999;57:465–480.
31. Lu AYH, Wang RW, Lin JH. Commentary: Cytochrome P450 *in vitro* reaction phenotyping: a re-evaluation of approaches for P450 isoform identification. *Drug Metab Dispos* 2003;31:345–350.
32. Ring BJ, Gillespie JS, Eckstein JA, Wrighton SA. Identification of human cytochromes P450 responsible for atomozetine metabolism. *Drug Metab Dispos* 2002;30:319–323.
33. Renwick AB, Surry D, Price RJ, Lake BG, Evans DC. Metabolism of 7-benzyloxy-4-trifluoromethylcoumarin by human hepatic cytochrome P450 isoforms. *Xenobiotica* 2004;30:955–969.
34. Crespi CL. Xenobiotic-metabolizing human cells as tools for pharmacological and toxicological research. *Adv Drug Res* 1995;26:179–235.
35. Uttamsingh V, Lu C, Miwa G, Gan LS. Relative contributions of the five major human cytochromes P450, 1A2, 2C9, 2C19, 2D6, and 3A4, to the hepatic metabolism of the proteasome inhibitor bortezomib. *Drug Metab Dispos* 2005;33:1723–1728.
36. McGinnity DF, Berry AJ, Kenny JR, Grime K, Riley RJ. Evaluation of time-dependent cytochrome P450 inhibition using cultured human hepatocytes. *Drug Metab Dispos* 2006;34:1291–1300.
37. Lu C, Miwa GT, Prakash SR, Gan LS, Balani SK. A novel model for the prediction of drug–drug interactions in humans based on *in vitro* cytochrome p450 phenotypic data. *Drug Metab Dispos* 2007;35:79–85.
38. Chiba M, Jin L, Neway W, Vacca JP, Tata JR, Chapman K, Lin JH. *Drug Metab Dispos* 2001;29:1–3.
39. Kim JY, Baek M, Lee S, Kim SO, Dong MS, Kim BR, Kim DH. Characterization of the selectivity and mechanism of cytochrome P450 inhibition by dimethyl-4,4′-dimethoxy-5,6,5′,6′-dimethylenedioxybiphenyl-2,2′-dicarboxylate. *Drug Metab Dispos* 2001;29: 1555–1560.
40. Wen X, Wang JS, Backman JT, Kivisto KT, Neuvonen PJ. Gemfibrozil as an inhibitor of human cytochrome P450 2C9. *Drug Metab Dispos* 2001;29:1359–1361.
41. Walsh CT. Suicide substrates, mechanism-based enzyme inactivators: recent developments. *Annu Rev Biochem* 1984;53:493–535.
42. Madeira M, Levine M, Chang TKH, Mirfazaelian A, Bellward G. The effect of cimetidine on dextromethorphan *O*-demethylase activity of human liver microsomes and recombinant CYP2D6. *Drug Metab Dispos* 2004;32:460–467.

43. Brown HS, Galetin A, Hallifax D, Houston B. Prediction of *in vivo* drug–drug interactions from *in vivo* data: factors affecting prototypic drug–drug interactions involving CYP2C9, CYP2D6 and CYP3A4. *Clin Pharmacokinet* 2006;45:1035–1050.
44. Kato M, Tachibana T, Ito K, Sugiyama Y. Evaluation of methods for predicting drug–drug interactions by Monte Carlo simulation. *Drug Metab Pharmacokinet* 2003;18:121–127.
45. Li AP, Rasmussen A, Xu L, Kaminski DL. Rifampicin induction of lidocaine metabolism in cultured human hepatocytes. *J Pharmacol Exp Ther* 1995;274:673–677.
46. Li AP. Primary hepatocyte cultures as an *in vitro* experimental model for the evaluation of pharmacokinetic drug–drug interactions. *Adv Pharmacol* 1997;43:103–130.
47. Roymans D, Van Looveren C, Leone A, Parker JB, McMillan M, Johnson MD, Koganti A, Gilissen R, Silber P, Mannens G, Meuldermans W. Determination of cytochrome P450 1A2 and P450 3A4 induction in cryopreserved human hepatocytes. *Biochem Pharmacol* 2004;67:427–437.
48. Roymans D, Annaert P, Van Houdt J, Weygers A, Noukens J, Sensenhauser C, Silva J, van Looveren C, Hendrickx J, Mannens G, Meuldermans W. Expression and induction potential of cytohromes P450 in human cryopreserved hepatocytes. *Drug Metab Dispos* 2005;33:1004–1016.
49. Hariparsad N, Nallani S, Sane RS, Buckley DJ, Buckley AR, Desai PB. Induction of CYP3A4 by efavirenz in primary human hepatocytes: comparison with rifampin and phenobarbital. *J Clin Pharmaol* 2004;44:1273–1281.
50. Lillibridge JH, Liang BH, Kerr BM, Webber S, Quart B, Shetty BV, Lee CA. Characterization of the selectivity and mechanism of human cytochrome P450 inhibition by the human immunodeficiency virus–protease inhibitor nelfinavir mesylate. *Drug Metab Dispos* 1998;26:609–616.
51. Wang RW, Newton DJ, Liu N, Atkins WM, Lu AYH. Human cytochrome P-450 3A4: *in vitro* drug–drug interaction patterns are substrate-dependent. *Drug Metab Dispos* 2000;28:360–366.
52. Backman JT, Kyrklund C, Neuvonen M, Neuvonen PJ. Gemfibrozil greatly increases plasma concentrations of cerivastatin. *Clin Pharmaol Ther* 2002;72:685–691.
53. Vickers AE, Sinclair JR, Zollinger M, Heitz F, Glanzel U, Johanson L, Fischer V. Multiple cytochrome P450s involved in the metabolism of terbinafine suggest a limited potential for drug–drug interactions. *Drug Metab Dispos* 1999;27:1029–1038.

# MECHANISMS AND CONSEQUENCES OF DRUG–DRUG INTERACTIONS

Dora Farkas, Richard I. Shader, Lisa L. von Moltke, and David J. Greenblatt
*Tufts University School of Medicine, Boston, Massachusetts*

**Contents**

1 Principles of Drug–Drug Interactions
- 1.1 Introduction
- 1.2 Mechanisms of Drug–Drug Interactions
- 1.3 Methods and Systems for Analyzing Drug–Drug Interactions
- 1.4 Pharmacokinetic Principles of Drug–Drug Interactions
- 1.5 Conclusions

2 Case Study A: Drug–Drug Interactions with Triazolam and Other Benzodiazepines
- 2.1 Introduction
- 2.2 Metabolic Interactions
- 2.3 *In Vitro* Studies
- 2.4 Pharmacokinetic Clinical Studies
- 2.5 Pharmacokinetic/Pharmacodynamic Integration
- 2.6 Conclusions

3 Case Study B: Drug Interactions with Herbs and Natural Food Products
- 3.1 Introduction
- 3.2 Regulation of Herbal Supplements
- 3.3 Influence of Herbal Constituents on Drug Metabolism *In Vitro*
- 3.4 Clinically Significant Drug Interactions
- 3.5 Conclusions

References

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

# 1 PRINCIPLES OF DRUG–DRUG INTERACTIONS

## 1.1 Introduction

Understanding the mechanisms and consequences of drug–drug interactions is essential for the development of new pharmaceuticals and for the design of multidrug regimens. Drug interactions occur when one drug (the "perpetrator") changes the pharmacokinetic and/or the pharmacodynamic actions of another drug (the "victim"). This interaction can lead to changes in the clearance of the victim drug, which can influence its efficacy and safety. For drugs with narrow therapeutic indices, such as warfarin, digoxin, and theophylline, small changes in plasma concentrations can lead to toxicity.

Even after drugs undergo rigorous safety and efficacy testing during clinical trials, they can be withdrawn from the market due to the unanticipated possibility of toxic interactions with other drugs. For example, in the late 1990s the U.S. FDA recommended the withdrawal of the antihistamine terfenadine after it was discovered that its coadministration with the antifungal ketoconazole was associated with cardiac toxicity, which in a few instances was fatal. Terfenadine has been replaced by its metabolite, fexofenadine, which has similar efficacy but no known cardiac toxicity [1].

Drug–drug interactions are especially a concern for patients taking several drugs concurrently, such as the elderly and patients with serious medical diseases such as cancer or HIV. It is estimated that almost 40% of the elderly take five or more drugs simultaneously [2] and that 90% of them also use over-the-counter (OTC) medications, some of which may also contribute to drug interactions [3]. For example, the $H_2$ receptor antagonist cimetidine inhibits several cytochrome P450s and interacts with the CYP1A2 substrate theophylline [4]. Elderly patients are also thought to be 3–10 times more likely than younger patients to be susceptible to adverse drug–drug interactions due to changes in hepatic and renal clearance [5, 6]. In addition, with the increasing use of complementary alternative medicines such as herbal supplements, the probability of drug–drug interactions is increased. For example, it is estimated that over 50% of cancer patients take herbal remedies, even though there is evidence for adverse drug interactions with several herbs such as kava-kava, *Ginkgo biloba*, ginseng, and echinacea [6, 7].

Drug–drug interactions are not limited to prescription drugs, OTC drugs, and herbal remedies. Other substances such as foods, beverages, and excipients in drug tablets can enhance or inhibit the disposition of drugs. Grapefruit juice, for example, is an inhibitor of CYP3A, which is responsible for metabolizing an estimated 50% of prescription drugs. Further clinical trials demonstrated that grapefruit juice only inhibits intestinal CYP3A, and therefore, it is only expected to influence the pharmacokinetics of drugs that undergo significant enteric CYP3A metabolism [8], such as short half-life benzodiazepines [9]. More recent evidence suggests that certain other fruit juices also have the potential to inhibit CYP3A. For example, Seville orange juice was shown to inhibit enteric metabolism of felodipine [10] to a similar extent as grapefruit juice. In addition, *in vitro* and animal studies suggest that pomegranate juice is also a significant inhibitor of CYP3A [11, 12].

Interestingly, certain tablet excipients also influence absorption and bioavailability. For example, the cremophores EL and RH 40, which enhance lipid solubility,

have been shown to increase the oral bioavailability of saquinavir and digoxin, respectively [13–15]. However, other excipients such as Tween 80 and HS15 can adversely affect the disposition of drugs by inhibiting the drug transporter P-glycoprotein (also known as ABCB1 and MDR1) and increasing intracellular levels of P-glycoprotein substrates such as the chemotherapy agent daunorubicin [15].

Drug–drug interactions can occur through four mechanisms of drug disposition, collectively known as ADME—*a*bsorption, *d*istribution, *m*etabolism, and *e*limination. Absorption of a drug across a lipid bilayer is influenced by several factors, including physicochemical properties of the drug, such as solubility, lipophilicity, the pH of the environment, gastric emptying rate, gastrointestinal motility, and bile secretion. Antacids such as $H_2$ receptor antagonists can inhibit the absorption of certain drugs, such as ketoconazole, which need an acidic environment, and the anxiolytic drug clorazepate, which needs an acidic environment to be converted to its metabolite, desmethyldiazepam [16]. Furthermore, multivalent cations, such as calcium, magnesium, zinc, iron, and aluminum, which form chemical complexes, can inhibit the absorption of the antibiotics tetracycline and ciprofloxacin [17, 18]. Thus, the consumption of milk or nutritional supplements containing these minerals is contraindicated with tetracycline. Gastrointestinal motility is also impaired by the anticholinergic effects of marijuana or enhanced by the gastrokinetic agent metoclopramide. A decrease (or increase) in gastrointestinal motility can increase (or decrease) the exposure of the intestine to drugs and may result in toxicity [17].

It was previously believed that when a drug is displaced from its plasma protein binding site by another drug, the victim drug's unbound concentration will increase, leading to increased efficacy or possibly toxicity [19]. There are actually very few cases where displacement from plasma protein causes clinically significant interactions [20, 21]. Exceptions include drugs with high extraction and narrow therapeutic indices. For example, extensively albumin-bound compounds may theoretically displace a drug such as warfarin; however, this effect is only transient and rarely of clinical importance [22, 23]. Some interactions were previously attributed primarily to displacement but were found to be a result of inhibition of clearance. Phenylbutazone, for example, inhibits the metabolism of warfarin [24, 25], while sulfonamides reduce the clearance of tolbutamide [26, 27].

Altered metabolism (including phase I and phase II) is the most common reason for drug interactions since many metabolic enzymes can be inhibited or induced by drugs. Inhibition and induction of transporter proteins, particularly P-glycoprotein, is also a clinically significant pathway for drug–drug interactions and will be discussed in more detail in the following sections.

Clearly, it is impossible for the prescribing physician to remember all potential drug interactions. However, an understanding of the principles of drug–drug interactions can aid in the design of a safe multidrug regimen. New analytical techniques and modeling tools combined with knowledge databases are enabling researchers to make better predictions of possible drug interactions. The goals of this chapter are to discuss (1) the common mechanisms of drug–drug interactions, (2) different experimental systems and modeling tools used to investigate possible drug interactions, and (3) pharmacokinetic principles of drug interactions. In Sections 2 and 3, we illustrate clinically significant drug–drug and drug–herb interactions in two different case studies.

## 1.2 Mechanisms of Drug–Drug Interactions

***Metabolic Interactions*** While there are several possible mechanisms for drug interactions, most of the interactions occur through alterations of hepatic or enteric drug metabolism. Metabolic interactions occur when a perpetrator drug inhibits or induces the activity of a drug-metabolizing enzyme, resulting in either decreased or increased metabolism of a victim drug. Most of the known drug interactions involve the cytochrome P450s (phase I metabolism); however, interactions may also involve phase II enzymes such as the UDP glucuronidases [28, 29]. Table 1 provides selected examples of some common prescription drugs that are inhibitors or inducers of the cytochrome P450s and the transporter P-glycoprotein. According to most studies, CYP2D6 is not an easily inducible enzyme. However, there are some studies that suggest that CYP2D6 might be induced by phenobarbital [30] and carbamazepine [31].

The cytochrome P450s implicated in most drug–drug interactions are CYP1A2, CYP2B6, CYP2C8, CYP2C9, CYP2C19, CYP2D6, and CYP3A4. The CYP3A family, particularly CYP3A4, is considered to be exceptionally important, since it mediates the clearance of many common prescription medications [29, 32]. Many prescription drugs, such as the antifungals ketoconazole and fluconazole, the macrolide antibiotic clarithromycin, and the antidepressant fluoxetine, are potent inhibitors of CYP3A and can lead to significant drug interactions. As mentioned previously, the antihistamine terfenadine was withdrawn from the market after it was discovered that coadministration with ketoconazole could lead to cardiac toxicity. The toxicity resulted from the accumulation of excess terfenadine in the bloodstream, since the CYP3A-mediated metabolism of terfenadine was inhibited by ketoconazole [33]. Antifungal agents also interact with other classes of metabolic enzymes. It is thought that fluconazole and miconazole can interact with the metabolism of warfarin, possibly by inhibiting CYP2C9 [6, 34]. However, not all CYP450 inhibitors lead to clinically significant interactions. For example, the antibiotics azithromycin and dirithromycin are weak CYP3A inhibitors and have not been observed to cause toxic drug–drug interactions [35–37].

Selective serotonin reuptake inhibitors (SSRIs) are also known to interact with most of the major cytochrome P450s. CYP1A2 is inhibited by fluvoxamine and to a lesser extent by citalopram, and CYP2D6 is known to be inhibited significantly by fluoxetine and paroxetine, and to a lesser extent by sertraline and citalopram. Many antineoplastic and antiretroviral drugs also interact with the CYP450s, which can be potentially dangerous when several of these drugs are coadministered. The anticancer drug paclitaxel, frequently used to treat breast and ovarian cancers, has been reported to cause toxicity when coadministered with CYP3A inhibitors such as the antiretrovirals delavirdine and saquinavir. Paclitaxel was also found to interact with the chemotherapeutic agent valspodar, possibly through the inhibition of P-glycoprotein by valspodar [6, 38].

Complicating clinical practice is the interindividual variability due to genetic differences in the drug metabolism enzymes [6]. Genetic differences in drug metabolism were first suggested in the 1950s. Hughes et al. [39] reported significant interindividual variability in the acetylation of isoniazid and observed that "slow acetylators" were more likely to experience toxicity from the drug [40]. In the years following this discovery, many epidemiological studies were carried out to confirm

**TABLE 1 Prescription Drugs as Inducers and Inhibitors of the Major Cytochrome P450s and P-Glycoprotein**

| CYP1A2 | CYP2B6 | CYP2C9 | CYP2C19 | CYP2D6 | CYP2E1 | CYP3A4 | P-Glycoprotein |
|---|---|---|---|---|---|---|---|
| | | | *Inducer(s)* | | | | |
| Omeprazole | Rifampin | Rifampin | Rifampin | None known | Isoniazid | Carbamazepine<br>Dexamethasone<br>Rifampin<br>Phenytoin | Rifampin |
| | | | *Inhibitors* | | | | |
| Fluvoxamine<br>Cimetidine | Ketoconazole<br>Clopidogrel | Sulfaphenzole<br>Fluvoxamine<br>Fluconazole<br>Miconazole | Fluoxetine<br>Fluvoxamine | Fluoxetine<br>Quinidine<br>Methadone<br>Paroxetine | Disulfiram | Clotrimazole<br>Itraconazole<br>Ketoconazole<br>Miconazole<br>Ritonazir<br>Nefazodone | Verapamil<br>Doxorubicin<br>Clarithromycin<br>Ritonavir<br>Omeprazole |

*Source*: Adapted from Refs. 15, 32, 43, 46, and 144.

interindividual differences in drug metabolism. For example, the cytochrome P450s CYP2D6 and CYP2C9 have significant genetic polymorphisms and they also metabolize an estimated 25% [41] and 16% [42] of prescription drugs, respectively. CYP2D6, in particular, has at least 80 allelic variants, resulting in a range from poor to ultrarapid metabolizers, and it is also inhibited by drugs such as fluoxetine and methadone [43]. CYP2C9 has five alleles with varying prevalences among different ethnic groups. Genetic variations in CYP2C9 are also a concern, since CYP2C9 is responsible for metabolizing substrates with narrow therapeutic indices such as warfarin and phenytoin [44].

Some of the most significant drug–drug interactions are due to mechanism-based, or irreversible, inhibition. In this case, the drug binds to the enzyme and can modify the heme, protein, or the complex through covalent binding and inactivate it irreversibly [45]. Mechanism-based inhibitors include paroxetine (CYP2D6) [28], tamoxifen, erythromycin, and fluoxetine (CYP3A4). The CYP3A inhibitor in grapefruit juice, 6′, 7′-dihydroxybergamottin, is also a mechanism-based inhibitor. In clinical trials where midazolam was used as a CYP3A probe, a 10 oz glass of white grapefruit juice led to approximately a 65% increase in midazolam's area under the curve (AUC). Subsequent administration of midazolam demonstrated that complete recovery of CYP3A takes about 3 days [8].

Although less investigated, drug–drug interactions also occur through enzyme induction. Induction is a much slower process than inhibition since it involves transcriptional activation of genes. CYP1A2 clearance is increased by cruciferous vegetables, charcoal-broiled beef, cigarette smoke, and omeprazole. The antibacterial rifampin is an inducer of CYP2C9, CYP2C19 and CYP3A4. CYP3A4 can be induced by a variety of compounds including barbiturates, glucocorticoids, phenytoin, carbamazepine, and St. John's wort [46]. Induction of cytochrome P450s typically occurs through the nuclear receptors primarily AhR, CAR, and PXR. In addition to interindividual differences among the cytochrome P450s, there are also differences in nuclear receptors [47, 48] and in their regulatory proteins [49, 50]. Chang et al. [51] measured the mRNAs of CYP2B6, CAR, and PXR in 12 human liver samples. They found a 240-fold interindividual variability among samples in CAR mRNA levels, and a 278-fold variability in the mRNA levels of CYP2B6. The variability in PXR was about 27-fold and it also correlated well with the variabilities in CYP2B6 and CAR [50]. Individual differences in drug disposition could also be due to genetic variations in drug transporters. In fact, there is recent evidence to suggest that interindividual variability in the metabolism of phenytoin and fexofenadine could be attributed to genetic differences in the transporter P-glycoprotein [52, 53].

***Interactions Through Nonneuronal Efflux Transporters*** Transporters are ubiquitously located throughout the body and play an important role in the disposition of drugs, particularly in the liver, the intestine, and the kidneys. While some transporters can interfere with the delivery of drugs inside the cells, they are essential for maintaining a physiologic barrier at several locations, such as the blood–brain barrier [54], testes, and placenta [15]. Transporters include both uptake and efflux proteins and have been classified into several families, including the ATP-binding cassette transporters (ABCs), the organic anion transporting polypeptides (OATPs), organic anion transporters (OATs), and the organic cation transporters (OCTs). Inhibition or induction of transporters can lead to increased or decreased drug

concentration inside the cells, and therefore a better understanding of these transporters is essential for rational drug design and pharmacokinetic modeling.

The ABC family is further subdivided into several subfamilies such as ABCA, ABCB, and ABCC. The most widely recognized transporter is the multidrug resistance protein, MDR1, also known as ABCB1 or P-glycoprotein [55], and in this chapter we use P-glycoprotein as the model for demonstrating the importance of drug–drug interactions through transporters.

Some drugs such as the immunosuppressant cyclosporin A are interesting in that they are both substrates and inhibitors of P-glycoprotein [15]. For example, it has been shown *in vitro* that cyclosporin A inhibited transport of the cardiac drug digoxin, yet digoxin did not inhibit transport of cyclosporin A [56]. P-glycoprotein is especially critical in maintaining the intactness of the blood–brain barrier and cyclosporin A has led to the accumulation of vinblastine in the brain of mice [57]. In P-glycoprotein knockout mice, a normally harmless dose of the opiate loperamide (used as an antidiarrheal in humans) was lethal as the drug gained entry into the central nervous system [58, 59]. In rats, cyclosporine inhibited P-glycoprotein in other organs such as the testes, where it led to the accumulation of doxorubicin [60]. Cardiac toxicity in rats was also caused by another inhibitor of P-glycoprotein, verapamil, which increased the cardiac levels of the chemotherapeutic agent idarubicin [61].

While the presence of P-glycoprotein is associated with drug efflux from cells, it is important to point out that not all substrates of P-glycoprotein have poor bioavailability. Digoxin, in particular, has an oral bioavailability of 50–85% [62, 63]. Ritonavir, another P-glycoprotein substrate, has a bioavailability of 60% [63, 64]. The reason for the high bioavailability of these drugs is that bioavailability is also influenced by influx processes, which might outweigh the effects of the efflux transporters for some drugs. In addition, similar to enzymes, P-glycoprotein efflux is a saturable process. The $K_m$ for many P-glycoprotein substrates, such as digoxin, verapamil, and indinavir, have been determined and the values are in the micromolar range. Thus, in theory, when the plasma concentration of the drug is significantly higher than its $K_m$ value, the drug might be expected to have high bioavailability [62, 63].

Inhibition of P-glycoprotein can occur through multiple mechanisms and can be classified as competitive inhibition, noncompetitive inhibition, or cooperative stimulation. Inhibition can occur if the inhibitor occupies an active binding site or if it inhibits the ATP hydrolysis process that is necessary for efflux. Verapamil inhibits P-glycoprotein by occupying an active binding site, vanadate interacts with ATP hydrolysis, while cyclosporine inhibits P-glycoprotein through both mechanisms [63, 65]. Cooperative inhibition occurs when two inhibitors act together to inhibit P-glycoprotein. P-glycoprotein is thought to have at least two binding sites, and sometimes inhibitors bind to both sites, leading to synergism between the two binding sites to inhibit P-glycoprotein [66]. Inhibitors that were found to act at both sites included vinblastine, tamoxifen, and quinidine [67]. Modeling inhibition can be quite complex, because in addition to the multiple binding sites on P-glycoprotein, there are also allosteric interactions when multiple binding sites are occupied.

Induction of P-glycoprotein has also been observed in response to common P450 inducers such as 3-methylcholanthrene [68] and dexamethasone [69]. Certain cytotoxic agents such as adriamycin, daunomycin, and mitoxantrone have also been found to induce P-glycoprotein in rodent cells [70]. There is now data to suggest

that induction of P-glycoprotein may occur through multiple mechanisms including PXR. Thus, species differences in P-glycoprotein induction are probably due in part to differences in the activation of species-specific PXR [63]. However, even within the same organism, P-glycoprotein is not necessarily induced to the same extent in every tissue, and the time until maximum induction can also vary significantly. Jette et al. [71] compared induction of P-glycoprotein by 10 mg/kg/day cyclosporine A in various organs in the rat. Whereas P-glycoprotein increased by more than 250% in the stomach, it only increased 69% in the lungs [63]. Furthermore, maximal response was observed after only 5 days in the heart, while it took 15 days for maximal response in the spleen and testes.

***Pharmacoenhancement or Augmentatics*** In the previous sections we have discussed examples of adverse drug–drug interactions. However, the pharmacokinetics of HIV combination protease inhibitors demonstrate the principles of pharmacoenhancement, where the "perpetrator" drug actually improves the effectiveness of the "victim" drug. In this section we illustrate the principles of pharmacoenhancement by discussing the pharmacokinetics of HIV protease inhibitors.

Protease inhibitors (PIs), introduced in the mid-1990s, are now standard therapy for HIV. However, in spite of their effectiveness, there is poor patient compliance, leading to treatment failure within 1 year in 40–60% of cases [72, 73]. Reasons for poor patient compliance include unpleasant side effects, the need for frequent dosing, and food- or fluid-dependent administration regimens. Poor patient compliance allows the viral load to get very high and can lead to resistance to the treatment. Combination PIs, where multiple drugs are combined within one formulation, are more effective in suppressing viral replication by increasing the overall exposure to the boosted drug. In addition, clinical trials have suggested better patient compliance because of fewer food restrictions, simpler dosing regimen, as well as lower interindividual variabilities.

Many PIs are combined with ritonavir as the "perpetrator" drug. Ritonavir, which is usually prescribed as 600 mg twice a day, is generally not well tolerated since it can lead to nausea, diarrhea, and other unpleasant side effects. However, a 100 mg dose of ritonavir has been shown to boost the effectiveness of several other PIs [72]. All approved HIV PIs are predominantly metabolized by CYP3A4, and they are all substrates for P-glycoprotein. Of all PIs, ritonavir is the most potent inhibitor of both CYP3A4 and P-glycoprotein [74]. The combination PI consisting of 400 mg of lopinavir and 100 mg of ritonavir has been shown to be well tolerated and effective [75] and is now a marketed drug as a twice-a-day prescription. The FDA has now approved other once-a-day regimens such as amprenavir and atazanavir for patients who are treatment-experienced. The effectiveness of administering lopinavir/saquinavir once-a-day is being currently evaluated [76]. Other drugs boosted by ritonavir include indinavir, amprenavir, saquinavir, lopinavir, atanazavir, and nelfinavir. Other PIs, such as efivarenz and nevirapine, are inducers of CYP3A4; thus, their inclusion in a treatment regimen might require an increase in the dose of other drugs [74].

The specific type of pharmacoenhancement with ritonavir depends on the individual PIs. The effects will be either (1) to increase $C_{max}$, $C_{min}$, and AUC and modestly increase $t_{1/2}$ (primarily $C_{max}$ boosting) or (2) to increase $t_{1/2}$ and $C_{min}$ and modestly increase AUC (primarily $t_{1/2}$ boosting). An increase in $C_{min}$ is associated with higher

efficacy and an increase in $C_{max}$ also leads to higher drug exposure but could also result in toxicity unless the drug is well tolerated. A boosting of $C_{max}$, $C_{min}$, and AUC (the first type of effect) is observed with lopinavir and saquinavir, which have high first-pass metabolism. Indinavir and amprenavir have short half-lives and ritonavir primarily increases the $t_{1/2}$ and $C_{min}$ and modestly increases AUC for these drugs (the second effect). It is also fortunate that the AUC and the $C_{max}$ are not significantly boosted since an increase in the $C_{max}$ for these drugs has been associated with adverse side effects [77].

Double-boosted PIs are also gaining popularity, especially in treatment-experienced patients. The advantages of this type of treatment are that (1) the levels of several different PIs are increased simulataneously and they are effective against viruses that might have resistance against one particular drug, and (2) addition of a second booster might further enhance the pharmacokinetic properties of the treatment. For example, addition of saquinavir to an atazanavir/ritonavir combination was reported to increase the $C_{max}$ of atazanavir more than just adding ritonavir. The AUCs of atazanavir and saquinavir also increased compared to single-boosted regimens [78, 79].

While ritonavir seems to be effective in combination PIs, not all patients can tolerate it due to its adverse side effects. Another booster drug is the nonnucleoside reverse transcriptase inhibitor delavirdine, which has been shown to increase the half-life, AUC, and $C_{max}$ of nelfinavir. Nelfinavir is metabolized partially by CYP2D6, which is missing in 7% of Caucasians and 1% of Asians [74]. Other individuals are ultrarapid CYP2D6 metabolizers and need higher doses to reach therapeutic drug concentrations. Thus, interindividual variability in drug-metabolizing phenotype also needs to be taken into consideration when prescribing the specific combination PI therapy.

### 1.3 Methods and Systems for Analyzing Drug–Drug Interactions

***In Vitro Systems*** Several systems are available for studying drug–drug interactions and the choice of system depends on the predicted type of interaction. Clearly, the most relevant system is the human body; however, clinical trials are expensive and only a few compounds can be tested per trial. Furthermore, species differences may limit the usefulness of animal experiments. *In vitro* and *in silico* (computational) studies can be useful in eliminating unlikely drug–drug interactions and determining the probable interactions that could be further investigated in clinical or animal studies. Microsomes are frequently used for inhibition studies because they are inexpensive and allow for simulataneous monitoring of several enzymes including the cytochrome P450s, the flavin monooxygenases, and the UDP-glucuronosyltransferases.

Induction studies are carried out in hepatocytes, since the cellular machinery is necessary for upregulating the transcription of genes. However, primary hepatocytes are known to rapidly lose metabolic functions after isolation; thus, special media and culture conditions are necessary for maintaining them. The main challenge regarding hepatocyte culture is that most factors stimulate either regeneration or differentiation but not both [40]. Long-term cultures of hepatocytes usually require a mitogenic compound, such as epidermal growth factor (EGF), which stimulates DNA synthesis. However, it has been shown that EGF leads to a downregulation

of the metabolic enzymes, specifically CYP1A, CYP2B [80, 81], CYP2C11 [82], and CP3A [83]. Cell density also influences cell proliferation and metabolism. A low cell density increases cell proliferation but reduces the expression and inducibility of metabolic enzymes [84]. Culturing hepatocytes on structural matrix proteins, such as collagen, fibronectin, and laminin, has been found to maintain cell morphology and metabolic functions better than plating them on plastic. Culturing hepatocytes in a sandwich configuration is even more suitable for maintaining metabolically active cells. The most common matrices for sandwich cultures are collagen type I and Matrigel, an extract derived from the basement membrane of the Engelbreth–Holm–Sworm mouse sarcoma [85]. Both collagen–collagen and collagen–Matrigel sandwiches have been shown to maintain morphology and albumin secretion for several weeks [86], and they are also inducible for both CYP1A2 and CYP3A4 in systems with human hepatocytes [87].

In addition to optimizing cell culture parameters, special cell lines have been engineered that constitutively express cytochrome P450s and are also inducible for specific enzymes such as CYP3A4. For example, the DPX-3A4 cell line, which contains a luciferase-linked PXR promoter, has been shown to be inducible by a number of compounds such as rifampin, ginseng, and kava-kava [88]. Other cell lines such as the Fa2N-4 are inducible for multiple enzymes, including CYP1A2, CYP2C9, CYP3A4, and UGT1A and the MDR1 transporter [89]. The WIF-B9 cell line, which is a fusion of rat hepatoma and human fibroblasts, has been shown to constitutively express both human and rat cytochrome P450s [90].

For transporter studies involving P-glycoprotein, the human intestinally derived Caco-2 cell line is the most frequently used. The Madin–Darby canine kidney (MDCK) cells are also frequently tested since they require shorter culture period (3–5 days) and have high levels of P-glycoprotein expression [91, 92]. The two main issues with using cells lines for transporter studies are (1) the influence of passage number on the expression of P-glycoprotein and (2) expression of other transporter proteins. In particular, Caco-2 cells typically express the breast cancer resistance protein (BCRP) and the multidrug resistance-associated protein 2 (MRP2) [93–95].

Pharmacokinetic studies can also be carried out in more physiologic *in vitro* systems such as a bioartificial liver. The main challenge with such systems is designing conditions that can maintain hepatocytes long enough for the studies. Cell lines are easier to maintain than primary hepatocytes, but they usually do not express drug-metabolizing enzymes. Several new cell lines have been developed for this purpose such as a pig hepatocyte cell line and a couple of human-derived lines. The pig hepatocytes have been shown to express most of the cytochrome P450s as well as glutathione *S*-transferase and UDP-glucuronosyltransferase [96]. A HepG2-derived cell line was also found to express CYP3A4 at even higher levels than primary human hepatocytes [97]. Furthermore, a human hepatocellular carcinoma cell line has been engineered to be inducible for CYP3A4. This cell line, FLC-5, was cultured in a radial-flow bioreactor and shown to be inducible for CYP3A4 via PXR activation [98]. Thus, the development of these cell lines is a promising step in designing physiologically based perfused bioreactors for drug metabolism and drug interaction studies.

***Computational Modeling and Data Processing Tools*** Computational (or *in silico*) tools are also used widely in the pharmaceutical industry to evaluate possible

metabolism and toxicity pathways for new drugs. While it is not possible to completely characterize most compounds *in silico*, a lot of the technology is already in place and incorporation of new data could lead to more reliable tools in the near future [99]. Quantitative structure–activity relationship (QSAR) models have been compiled for over 2000 compounds to determine possible substrates for the CYP450s [100]. Although many other pathways and endogenous factors influence the pharmacokinetics of drugs, these models can aid in modeling the binding potential of drugs to enzymes. Structure–activity models have been developed for the major cytochrome P450s, including CYP3A4, CYP2D6, and CYP2C9, and results from these models have shown over 90% correlation with experimental data [101]. More in-depth QSAR models have been built for individual P450s, such as CYP2C9, to examine interactions of the enzyme with a variety of potential inhibitors and also to determine the effect of mutations on enzyme activity [102]. Special computer-based models have also been developed for the prediction of drug–drug interactions, which incorporate all known data on drug metabolism enzymes. One example is Q-DIPS (quantitative drug interactions prediction system), which correctly predicted a potent inhibition of CYP3A4 by ketoconazole, and a weaker inhibition by fluconazole [103]. QSAR models have also been developed for drug transporters, especially P-glycoprotein. These models have identified several potential modulators of transporters in order to overcome drug resistance to many chemotherapeutic compounds [104, 105]. Thus, these models can be useful for discovering new compounds that could enter the drug pipelines; however, the development process of new drugs still relies heavily on *in vitro* and *in vivo* experimentation.

Another challenge of drug development is integration of all known data about a certain enzyme or the affected molecular pathway affected. Integration of data has become especially time consuming with the high throughput methods of genomics, proteomics, and metabonomics, which generate vast amounts of data. The goal of systems biology is to combine experimentally generated data with modeling tools in order to discover potential drug interactions and new therapeutic targets. Although many models have been developed, they are far from being complete, since only 15% of the human genome has known function. Nevertheless, these knowledge databases combined with high throughput screening have become useful for modeling ADME/Tox, discovering potential drug interactions and for visualizing the biological processes influenced by drugs [106].

### 1.4 Pharmacokinetic Principles of Drug–Drug Interactions

***Kinetics*** Pharmacokinetic modeling is an essential tool for understanding drug disposition and for designing safe and effective treatment regimens. The simplest approach assumes linear, first-order Michaelis–Menten kinetics. This type of kinetics is described by

$$v = \frac{V_{\max}[\mathrm{S}]}{K_{\mathrm{M}} + [\mathrm{S}]}$$

where $v$ is the reaction velocity, [S] is the substrate concentration, and $K_{\mathrm{M}}$ is the substrate concentration at which the reaction velocity reaches half of the maximal velocity, $V_{\max}$. When $[\mathrm{S}] << K_{\mathrm{M}}$ the reaction displays linear kinetics; thus, the rate is directly proportional to the substrate concentration. When [S] get close to $K_{\mathrm{M}}$

the enzyme starts becoming saturated and eventually the velocity, $v$, reaches its maximum, $V_{max}$.

If a drug is metabolized by more than one enzyme, the velocity of the reaction is the sum of the individual rates:

$$v = \frac{V_{max(1)}[S]}{K_{M(1)} + [S]} + \frac{V_{max(2)}[S]}{K_{M(2)} + [S]}$$

In this case each enzyme reaction has its own $V_{max}$ and $K_M$, denoted by the appropriate subscripts. In order to recognize such a phenomenon, the data needs to be viewed on an Eadie–Hofstee plot ($v$ vs. $v/[S]$), which will illustrate the biphasic characteristic of the reaction [28].

Inhibition and induction occur through different mechanisms and there are different mathematical models and *in vitro* tools for studying them. Inhibition is primarily a chemical phenomenon involving a reversible or irreversible binding of the inhibitor to the enzyme. For this reason inhibition is usually concentration dependent and is usually quantified by the inhibition constant $K_i$ or the 50% inhibitory concentration, the $IC_{50}$. Inhibition can be competitive, noncompetitive, or uncompetitive. In competitive inhibition, the inhibitor binds to the active pocket of the enzyme and blocks the binding of the substrate. In noncompetitive inhibition, the inhibitor and the substrate bind independently to different sites, and in uncompetitive inhibition, the inhibitor binds to the substrate–enzyme complex and renders the complex inactive. In all three cases, the inhibitor interferes with the reaction velocity and the Michaelis–Menten kinetics are then modified accordingly [46].

For competitive inhibition,

$$v = \frac{V_{max}[S]}{[S] + K_M(1 + [I]/K_i)}$$

For noncompetitive inhibition,

$$v = \frac{V_{max}[S]}{(1 + [I]/K_i)(K_M + [S])}$$

For uncompetitive inhibition,

$$v = \frac{V_{max}[S]}{[S] + K_M + [I][S]/K_i}$$

The mechanism of inhibition can be determined from plotting experimental data. Figure 1 shows the plots of $1/v$, the reciprocal of the velocity, versus [I], the inhibitor concentration, for the three types of inhibition (also known as Dixon plots) [107]. Note that [I] and $1/v$ will be positive for all experimental data and the diagrammed intersection points can only be determined by extrapolating the lines. When the mechanism is competitive, the lines will intersect at the point where $[I] = -K_i$ and $1/v = 1/V_{max}$, and when it is noncompetitive the intersection point is at $[I] = -K_i$ and $1/v = 0$. In uncompetitive inhibition, the slopes are independent of [S] and hence

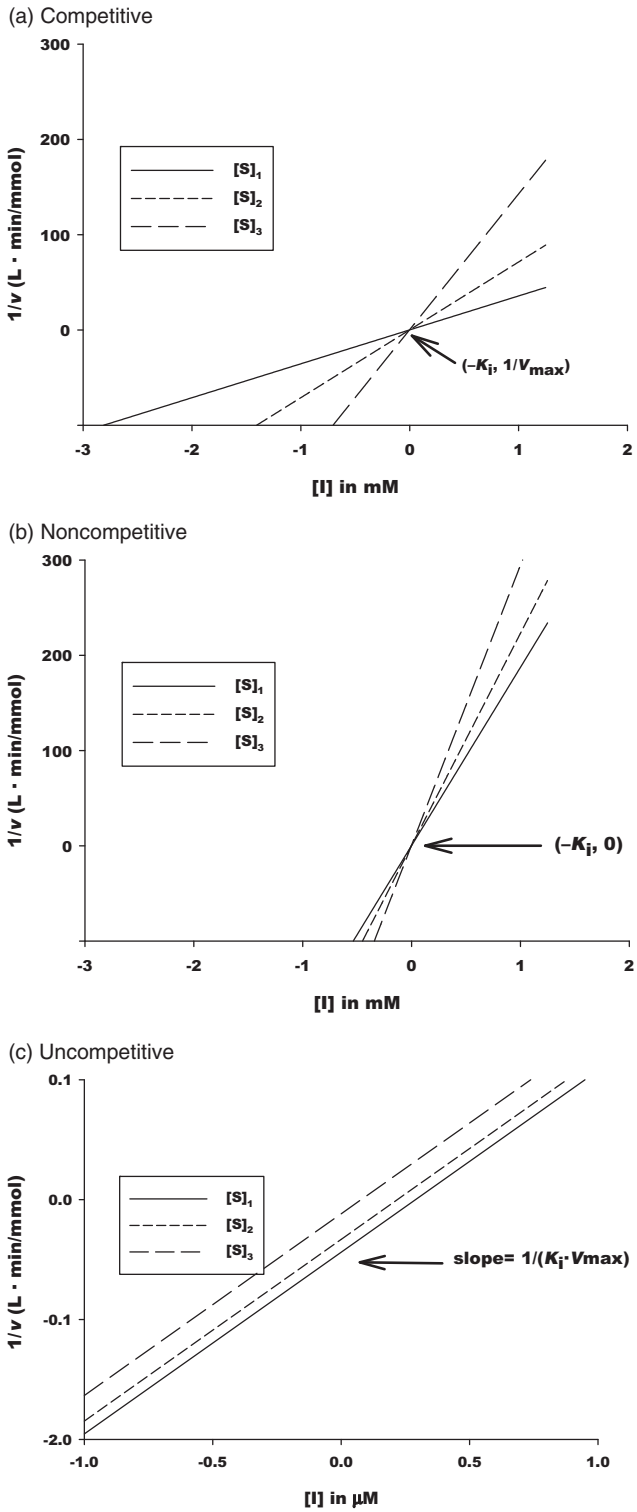


**FIGURE 1** Relationship between concentration of inhibitor, [I], and the reciprocal of velocity ($1/v$) at different substrate concentrations for (a) competitive, (b) uncompetitive, and (c) noncompetitive inhibition with the following parameters: $K_m = 4.7 \times 10^{-5}$ M; $V_{max} = 22\,\mu$Mol/(L·min); $K_i = 3 \times 10^{-4}$ M. The three substrate concentrations were $[S]_1 = 2 \times 10^{-4}$ M, $[S]_2 = 1 \times 10^{-4}$ M, and $[S]_3 = 5 \times 10^{-5}$ M.

**TABLE 2 Pharmacokinetic Changes Due to Drug–Drug Interactions**

| Mechanism of Interaction | Change in Systemic Exposure | Theoretical Change in Exposure[a] |
|---|---|---|
| Competitive Inhibition | Increased | $\frac{AUC_i}{AUC} = 1 + \frac{[I]}{K_i}$ |
| Noncompetitive inhibition | Increased | $\frac{AUC_i}{AUC} = 1 + \frac{[I]}{K_i}$ |
| Uncompetitive inhibition | Increased | $\frac{AUC_i}{AUC} = 1 + \left(\frac{[I]}{K_i}\right)\left(\frac{[S]}{[S]+K_M}\right)$ |
| Mechanism-based inhibition (irreversible inactivation) | Increased | $\frac{AUC_i}{AUC} = \frac{V_{max}}{V_{max(inhibited)}}$ |
| Induction | Decreased | $\frac{AUC_i}{AUC} = \frac{V_{max}}{V_{max(inhibited)}}$ |

[a] $AUC_i$ denotes the area under the curve after administration of the inhibitor. These equations assume that $[S] << K_M$.

*Source*: Adapted from Ref. 28.

the lines will be parallel [107]. Table 2 summarizes the changes in systemic exposure during different drug interaction scenarios [28].

Determining the type of inhibition mechanism and calculation of the inhibition constant, $K_i$, is very labor intensive, since the reaction velocity needs to be determined at several different inhibitor and substrate concentrations. Before determining the type of inhibition, it is helpful to investigate whether the inhibition is mechanism based, in which the case the inhibitor would irreversibly inactivate the enzyme. In this case, the inhibitor is preincubated with the enzyme for some time period before addition of the substrate. This preincubation allows the inhibitor to bind to the enzyme, and if the mechanism is irreversible, the reaction should proceed significantly slower than without preincubation.

Calculation of the $IC_{50}$ is useful for determining whether the inhibitor is likely to have an *in vivo* significance. The $IC_{50}$ can be determined by plotting the percentage of control (uninhibited) reaction velocity ($R$) versus the inhibitor concentration [I], at a constant substrate concentration. The resulting sigmoidal curve can be modeled with nonlinear regression and described with the following equation:

$$R = 100\left(1 - \frac{E_{max}C^A}{C^A + IC^A}\right)$$

where $E_{max}$ is the maximum fractional inhibition, $C$ is the concentration of the inhibitor, $IC$ is the concentration that results in a reaction velocity of $50(2 - E_{max})$, which is the velocity at half the maximal inhibition, and $A$ is an exponent relating the sigmoidal shape of the curve. The true $IC_{50}$, producing a velocity that is 50% of the control value, is calculated as [108]

$$\mathrm{IC}_{50} = \frac{IC}{(2E_{\max} - 1)^{1/A}}$$

Using this approach, the calculation of $IC_{50}$ is model independent, since its value does not depend on the biochemical mechanism of inhibition (competitive vs. noncompetitive vs. uncompetitive). Lower $IC_{50}$ values indicate greater inhibitory potency and $IC_{50}$ will always be greater than or equal to $K_i$, the inhibition constant. If the mechanism of inhibition is "noncompetitive," they are equal [109]. If the mechanism of inhibition is reversible (competitive or noncompetitive), the preincubation process will not affect inhibitory potency, or may actually decrease inhibitory potency (higher $IC_{50}$) due to metabolic consumption of inhibitor in the microsomal system. If the inhibition is irreversible, the preincubation will increase inhibitory potency (lower $IC_{50}$).

Many metabolic enzymes obey atypical or nonlinear enzyme kinetics. CYP3A4 exhibits a sigmoidal saturation curve due to the existence of multiple binding sites on which substrates can bind cooperatively [110]. The $v$ versus [S] plot is described by the following equation:

$$v = \frac{V_{\max}[\mathrm{S}]^n}{S_{50}^n + [\mathrm{S}]^n}$$

in which $n$ refers to a Hill coefficient (which has to be greater than 1.0), and $S_{50}$ is the substrate concentration at which the velocity is half of $V_{\max}$ [28].

Another possible complication in modeling drug–drug interactions is the possibility of concurrent inhibition and induction. Ritonavir is an example, since it is an inhibitor of CYP3A, but long-term administration induces the enzyme through the PXR receptor [111]. St. John's wort was found to produce the same effect; certain components of it are potent inhibitors of CYP3A, yet St. John's wort also activates CYP3A through PXR [112]. Inhibition and induction can be distinguished from each other since they occur on different time scales. Inhibition is an immediate phenomenon and can be studied in microsomes, which do not have the cellular machinery for inducing enzymes. Induction is a time-consuming process, where transcription alone can take 4–6 hours; however, in clinical trials it could take days to see any inductive effects, especially if inhibition is competing with induction. For example, in a trial where ritonavir and alprazolam were coadministered, there was a small decrease in the plasma concentration of alprazolam after 12 days, suggesting that induction was overcoming inhibition [28].

Besides metabolism, efflux transport is also a form of clearance and modeling of transporter systems is useful for determining the most important parameters in this process. Several computational models have been developed to study transporters. Jang et al. [113] modeled the P-glycoprotein-mediated efflux of the anticancer drug paclitaxel. The P-glycoprotein-mediated efflux rate can described as

$$\text{P-gp} - \text{efflux rate} = \frac{J_{\max} \cdot C_{\max}}{K_{\mathrm{M,P\text{-}gp}} + C_{\mathrm{free,c}}}$$

In this equation, which is analogous to the Michaelis–Menten model, $J_{\max}$ is the maximum efflux rate per cell, $K_{\mathrm{M,P\text{-}gp}}$ is the dissociation constant of the efflux process,

and $C_{free,c}$ is the free drug concentration inside the cells [114]. The concentration of the drug is then modeled using mass balance equations, which describe the amount of paclitaxel inside and outside the cells as a function of time. Using this approach, it was determined that the most important factor influencing the transport process is the extracellular drug concentration, followed by intracellular binding capacity and binding affinity, and finally P-glycoprotein expression. Furthermore, it was also shown that, similar to enzyme kinetics, P-glycoprotein efflux is a saturable process [115]. Pharmacokinetic modeling can also be used to determine which transport process is most relevant for a certain drug and what the rate constants are for a particular process [116].

***In Vitro/In Vivo Scaling*** Clearance is an important parameter for *in vitro/in vivo* scaling, particularly in drug–drug interaction studies. *In vitro* studies are usually conducted in liver microsomes and several assumptions need to be made for scaling. First, one must assume that liver metabolism is the major route of clearance and that oxidative metabolism is much more significant than other forms of metabolism such as conjugation, hydrolysis, and reduction. Second, one must also assume that the rates of metabolism *in vivo* will be similar to those *in vitro*. The scaling factor is usually a function of the microsomal protein per gram liver and the weight of the liver in comparison to body weight. Livers are usually assumed to have 45–50 mg microsomes/gram liver. If one is to scale up from hepatocyte experiments, 120,000,000 hepatocytes/gram liver could be used as a scaling factor [28, 117]. If one is scaling up from hepatocyte experiments, it is not necessary to assume that oxidative metabolism is dominant since hepatocytes are capable of carrying out the other forms of biotransformation.

One of the challenges of scaling *in vitro* to *in vivo* is estimating the inhibitor concentration, [I]. In a microsomal incubation, [I] can be calculated directly from the experimental conditions. However, *in vivo* it is not clear what the proper [I] is, since the inhibitor concentration at the site of metabolism cannot be calculated from the administered dose. Some estimates of [I] include total inhibitor concentration in the plasma, unbound concentration in the plasma, or concentration at the target organ such as the intrahepatic concentration. So far, no general estimate has been found to correlate *in vitro* [I] to an *in vivo* [I] in every specific case. Complicating factors include intestinal metabolism, transporters, and flow-dependent clearance. In general, the larger the $IC_{50}$ values, the less potent the inhibitor and the less likely that a clinically significant interaction will occur. Conversely, small $IC_{50}$ values usually indicate a high likelihood of a clinical interaction. Large $IC_{50}$ values are considered to be those higher than 100 μM, while small ones are usually less than 1 μM. The challenge is to interpret "intermediate" values of $IC_{50}$. Calculating $[I]/K_i$ can also be helpful in determining the likelihood of an *in vivo* interaction. The larger the ratio the more likely the possibility of an interaction. If the ratio is less than 0.5, the interaction is unlikely; but if $[I]/K_i$ is higher than 5, a clinically significant interaction is probable [109, 118].

## 1.5 Conclusions

Drug–drug interactions can occur through multiple pathways, where the perpetrator drug can influence the absorption, distribution, metabolism, and elimination of the

victim drug. Based on empirical data, the most common interactions occur through the inhibition or induction of the cytochrome P450s and the efflux transporter P-glycoprotein. Clinically significant drug interactions can lead to an increase in systemic blood concentrations (possibly causing toxicity), or a decrease in the plasma concentrations (corresponding to lack of efficacy). The success of combination HIV protease inhibitors illustrates the principle of pharmacoenhancement, where the combined formulation of two drugs is designed such that one drug improves the efficacy of the second drug. This improvement usually occurs by increasing the AUC, $C_{max}$, or the half-life of the boosted drug, thus increasing the overall systemic exposure to the compound. In this section we have also discussed the pharmacokinetic principles of drug–drug interactions as well as the challenges associated with predicting *in vivo* interactions from *in vitro* data. It is clear that it is currently very difficult to make clinically significant predictions based on *in vitro* data, even with the use of mathematical modeling. In Case Study A, we discuss the pharmacokinetics of benzodiazepines, particularly triazolam. We show several aspects of drug interactions with benzodiazepines, including *in vitro* and *in vivo* studies, prediction of clinical interactions, differences between males and females and different age groups, and also pharmacokinetic/pharmacodynamic integration. As is clear from the discussion so far, over-the-counter preparations (including herbal supplements) and foods can significantly contribute to drug–drug interactions. In Case Study B, we review the most commonly used herbs, mechanisms of drug interactions, and reports of clinically significant adverse events.

## 2 CASE STUDY A: DRUG–DRUG INTERACTIONS WITH TRIAZOLAM AND OTHER BENZODIAZEPINES

### 2.1 Introduction

Among the most commonly prescribed benzodiazepines are triazolam, midazolam, lorazepam, diazepam, and alprazolam. Triazolam, a short half-life triazolobenzodiazepine, is now one of the accepted clinical probes for monitoring CYP3A activity in human studies as it is metabolized primarily by CYP3A [119]. Since triazolam is a hypnotic, it has also been useful for pharmacodynamic experiments. The purpose of this case study is to review the series of *in vitro* and *in vivo* experiments that explored the pharmacokinetics and pharmacodynamics of triazolam and related benzodiazepines in the presence of other drugs, particularly CYP3A inhibitors. This case study also discusses (1) contributions from different CYP3A isoforms, (2) hepatic versus intestinal CYP3A, (3) effects of benzodiazepines in young versus elderly individuals, and (4) pharmacokinetic/pharmacodynamic integration.

### 2.2 Metabolic Interactions

Triazolam is metabolized primarily by CYP3A4 and it forms the α-hydroxy and 4-hydroxy metabolites [120]. These compounds have less activity than triazolam and are further conjugated to the glucuronide metabolites [121]. Triazolam metabolism is known to be inhibited by several classes of CYP3A inhibitors such as macrolide antibiotics [122], azole antifungal agents [123], and some SSRI antidepressants [124].

On the other hand, drugs that induce CYP3A would be expected to impair the effectiveness of triazolam. Coadministration of rifampin, an inducer of CYP3A4, greatly reduces the pharmacodynamic effects of triazolam [125]. Coadministration of triazolam with ritonavir poses an interesting clinical dilemma because ritonavir has been shown to concurrently induce and inhibit CYP3A [126]. During short-term administration, ritonavir inhibits the clearance the triazolam, but during long-term dosage the net effect of the coadministration is not easily predicted [111].

### 2.3 *In Vitro* Studies

The metabolism of triazolam was examined *in vitro* in human liver microsomes in the presence of several CYP3A inhibitors including ketoconazole, erythromycin, and several types of SSRIs in order to assess the likelihood of a clinical interaction [33]. Without inhibition, the $V_{max}$ and $K_M$ values for the formation of the 4-OH metabolite were 10.3 nM/min/mg of protein and 304 μM. For the α-OH pathways, these values were 2.4 nM/min/mg and 74 μM, respectively [124]. However, the $V_{max}/K_M$ ratios were the same for both metabolites, suggesting that each pathway contributes equally to clearance.

Ketoconazole was found to be the most potent inhibitor with a $K_i$ of 0.006 μM for the α-OH metabolite and 0.023 for the 4-OH metabolite. The weakest inhibitor was erythromycin with $K_i$ values of 36.6 μM and 111 μM for the α-OH and 4-OH pathways, respectively. Several SSRIs were also inhibitors of CYP3A, including fluoxetine, norfluoxetine, and sertraline.

The metabolism of midazolam was also examined *in vitro* in the presence of SSRIs and azole antifungal agents [127]. Midazolam also forms the α-OH and 4-OH metabolites, and without inhibition the formation of the α-OH metabolite was a higher affinity process ($K_M$ = 3.3 μM) than the formation of the 4-OH metabolite ($K_M$ = 57 μM). Based on the $V_{max}/K_M$ ratios, it was estimated that α-OH accounted for 95% of intrinsic clearance. Competitive inhibition constants ($K_i$) versus the formation of the α-OH metabolite for the antifungal agents ketoconazole, itraconazole, and fluconazole were 0.0037 μM, 0.27 μM, and 1.27 μM, respectively. The $K_i$ values for the α-OH pathway for fluoxetine and its metabolite norfluoxetine were 11.5 μM and 1.44 μM, respectively, which was consistent with previous clinical findings, which suggested that fluoxetine impairs the clearance of CYP3A substrates. Furthermore, these results suggest that most of the inhibition is due to the metabolite norfluoxetine rather than the parent compound fluoxetine.

The CYP3A family is thought to consist of at least three isoforms—CYP3A4, CYP3A5, and CYP3A7. While CYP3A4 is considered to be the dominant isoform, CYP3A5 has been detected in 10–25% of adult livers [128] and CYP3A7 is thought to be present in up to 50% of adult livers, particularly from African-Americans [129–131]. The metabolism of both triazolam and midazolam, as well as testosterone and nifedipine, was examined in human liver microsomes and recombinant CYP3A4 and CYP3A5 to investigate the contribution of each isoform in the metabolism of these drugs [132]. Overall, CYP3A4 contributed more than CYP3A5 in the metabolism of all the compounds. Ketoconazole significantly inhibited metabolism in microsomes and recombinant CYP3A4. The inhibition for CYP3A5 was 5–19-fold less than for CYP3A4 for all substrates, suggesting that CYP3A5 might not be as significant as CYP3A4 in drug interactions with ketoconazole. Although there seems

to be a significant interindividual variability in CYP3A5 content, with some samples having up to 25% as much CYP3A5 as CYP3A4, the overall net contribution is probably dominated by CYP3A4.

### 2.4 Pharmacokinetic Clinical Studies

It is important to note that clinically significant drug–drug interactions are not very common and, most frequently, coadministration of two drugs results in no change in the pharmacokinetic or pharmacodynamic profile of either drug. Occasionally, there is a change in the kinetic profile but it is only rarely large enough to be clinically significant [133].

Inhibition of triazolam metabolism by ketoconazole was confirmed in clinical trials where a 200 mg dose of ketoconazole reduced the oral clearance of 0.125 mg triazolam ninefold and prolonged half-life fourfold [124]. This degree of inhibition was correctly predicted by an *in vitro–in vivo* scaling model, which assumed competitive inhibition:

$$\text{Fractional decrement} = \frac{I}{I + K_{\mathrm{i}}(1 + S/K_{\mathrm{M}})}$$

If $S$ is much less than $K_{\mathrm{M}}$, this equation can be approximated as

$$\text{Fractional decrement} = \frac{I}{I + K_{\mathrm{i}}}$$

This equation holds true in the case of noncompetitive inhibition regardless of the substrate concentration [124].

Triazolam is a relatively high extraction drug, metabolized primarily by CYP3A4, with oral bioavailability around 50%. The plasma levels of high extraction drugs are expected to be influenced significantly when combined with inhibitors of the key metabolic enzymes. Figure 2 shows the pharmacokinetic and pharmacodynamic profiles of triazolam when the drug was coadministered with ketoconazole. After coadministration, there was a 10-fold decrease in the clearance of triazolam, which was accompanied by a drop in the pharmacodynamic marker, the digit-symbol-substitution test (DSST). On the other hand, alprazolam is a low extraction drug (also metabolized by CYP3A4) with a bioavailability of 90%. After coadministration with ketoconazole, its clearance was only reduced threefold, in contrast to the 10-fold decrease observed with triazolam [134].

The expression of hepatic and intestinal CYP3A are influenced by different environmental factors and thus their contributions to the metabolism of drugs need to be distinguished. For example, 6,7-dihydroxybergamottin, a component in grapefruit juice, primarily inhibits intestinal CYP3A and is thus only expected to interfere with drugs whose metabolism is primarily carried out by enteric CYP3A. In order to distinguish between the contributions of hepatic and intestinal metabolism, a drug is usually administered in two separate trials: once orally, when the drug undergoes both hepatic and intestinal metabolism, and then intravenously when the contribution from intestinal metabolism is considered negligible. Midazolam is also one

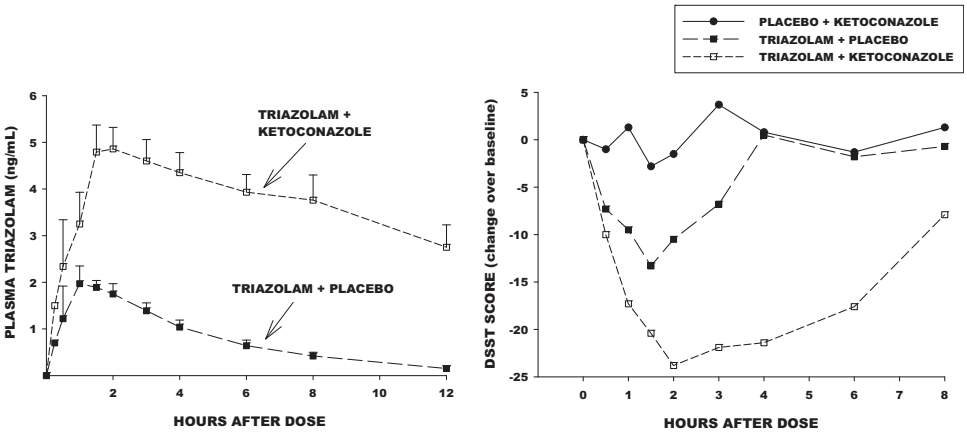


**FIGURE 2** Pharmacokinetic and pharmacodynamic interactions between triazolam and ketoconazole. Healthy male volunteers received 0.25 mg triazolam and 200 mg ketoconazole or placebos orally. The diagram on the left shows the mean (+SE) plasma concentration of triazolam (ng/mL) with and without ketoconazole. The diagram on the right illustrates the results of the digit-symbol-substitution test (DSST) under the three conditions listed [134].

of the accepted CYP3A clinical probes and its metabolism was studied in the presence of ketoconazole to determine contributions from intestinal and hepatic CYP3A [129]. After ketoconazole administration the AUC of intravenous midazolam increased fivefold, whereas after oral administration it increased 16-fold. The bioavailability of the intestinal component also increased more significantly than the hepatic component after ketoconazole therapy. These results suggest that intestinal CYP3A contributes significantly to midazolam clearance after oral dosage. Interestingly, females had a higher clearance of midazolam than males. Most other studies with benzodiazepines found no statistically significant difference between males and females, although females in general have higher clearances than males, possibly because females have increased enteric CYP3A [135]. However, the sample size in this study was small (three females and six males), thus larger studies would be needed to confirm a difference.

On the other hand, the clearance of CYP3A substrates is found to decrease with age, particularly in men, possibly due to decreases in hepatic and renal clearance [136]. A clinical trial with triazolam showed that elderly women (over 60 years of age) had a small increase in AUC and a decrease in clearance compared with younger women (20–36 years of age) but these differences were not statistically significant. However, in elderly men the AUC values were 75% higher and clearances were 28% lower than in younger men (Fig. 3) [137]. These findings concur with previous research, which suggested that age-related decrease in clearance of CYP3A substrates might be more significant in men.

## 2.5 Pharmacokinetic/Pharmacodynamic Integration

The goal of pharmacokinetic/pharmacodynamic (PK/PD) studies is to correlate changes in the plasma levels of drugs with observed pharmacodynamic effects. In

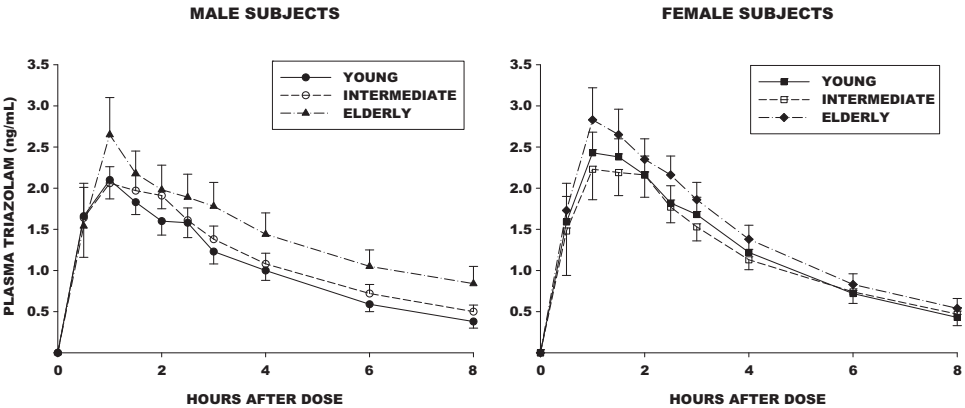


**FIGURE 3** Mean (±SE) plasma triazolam levels (ng/mL) for male and female volunteers in different age groups after administration of 0.25 mg triazolam. The age groups were young (20–36 years old), intermediate (40–56 years old), and elderly (60–75 years old) [137].

the previous section we discussed that the pharmacokinetics of triazolam were more affected than the pharmacokinetics of alprazolam. Pharmacodynamic studies consisting of electroencephalographic (EEG) beta activity and the digit-symbol-substitution test (DDST) also showed a similar pattern. For alprazolam, ketoconazole increased EEG and DSST by factors of 1.35 and 2.29, respectively, whereas these numbers increased by factors of 2.51 and 4.33, respectively, for triazolam [134]. Thus, the consequences of coadministering a high extraction compound such as triazolam with a CYP3A inhibitor could be clinically more important than administering the CYP3A inhibitor with a low extraction drug such as alprazolam.

The pharmacodynamics of triazolam in elderly volunteers followed a pattern similar to the pharmacokinetics. Probably due to decreased clearance, elderly subjects experienced more pronounced pharmacodynamic effects such as sedation and impairment of motor coordination [137–139]. For example, the values for percent decrement in the DSST after triazolam administration were –5.9 ± 1.9 and –8.6 ± 1.9 in young and elderly women, respectively (Fig. 4) [137]. These values in young and elderly men were –5.4 ± 2 and –11.4 ± 2, respectively. It was of note that self-rated sedation actually decreased in elderly subjects after triazolam administration despite an increase in the observer-rated sedation and a decrease in the percent beta amplitude. It is possible that elderly subjects were not aware of their sedation or their reporting did not match with their true sedation levels. Thus, prescribing benzodiazepines for the elderly could be a concern if they are not fully aware of their sedation or cannot report the effects of these drugs.

Another concern with an elderly population is the increase in the likelihood of drug interactions within multidrug regimens. Steroid hormones, such as progesterone, have been known to exert effects on the central nervous system [140]. Thus, the pharmacodynamic interaction between triazolam and progesterone has been investigated in postmenopausal women, since this population previously was likely to be concurrently taking both drugs [141]. In a randomized trial, one group received intravenous triazolam plus oral progesterone (300 mg) or intravenous triazolam plus oral placebo. The pharmacokinetic parameters such as AUC were similar between

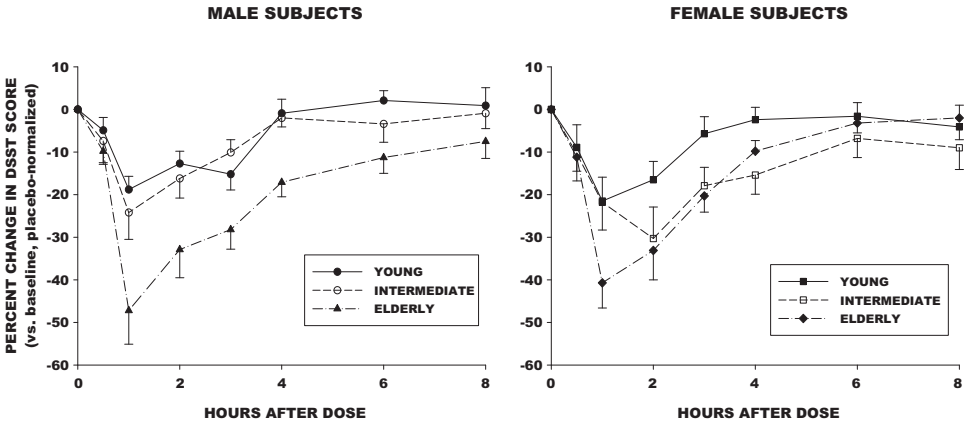


**FIGURE 4** Percent change (±SE) in digit-symbol-substitution test (DSST) scores for male and female volunteers in different age groups after administration of 0.25 mg triazolam. The age groups were young (20–36 years old), intermediate (40–56 years old), and elderly (60–75 years old) [137].

both groups but pharmacodynamic testing showed that progesterone potentiated the sedative effects of triazolam, suggesting that women taking progesterone might have an increased sensitivity to triazolam or other benzodiazepines. The pharmacodynamic effects of progesterone are attributed to a metabolite 3α-hydroxy-5α-dihydroprogesterone, which modulates the GABA receptor complex and regulates brain excitability [141, 142].

Pharmacodynamic effects are influenced not only by plasma levels but also by the concentration of drug in the brain and the binding of the drug to its receptor. Interestingly, the pharmacodynamic effects of triazolam in the presence of ketoconazole are less potent than expected based on plasma levels. Thus, the effects of ketoconazole on triazolam receptor binding were investigated in mice who received 50 mg/kg ketoconazole and 0.1–0.3 mg/kg of triazolam [143]. Ketoconazole was found in the brains at 31% of the plasma levels and its coadministration also increased triazolam levels in the brain compared to controls. *In vitro* binding studies also demonstrated that ketoconazole inhibits triazolam displacement of [$^3$H]flunitrazepam binding by 36–89% depending on ketoconazole concentration. Thus, these data suggest that, in addition to inhibiting CYP3A metabolism, ketoconazole also interferes with triazolam binding in the brain and could impair its pharmacodynamic effects.

The efflux transporter P-glycoprotein, or MDR1, has been shown to be involved in drug disposition in many organs including the liver, kidney, intestines, and brain. As an efflux transporter, MDR1 can limit the transport of drugs into epithelial cells and enhance the excretion of drugs from hepatocytes and renal tubules [144]. The role of P-glycoprotein is especially critical in the blood–brain barrier, where the endothelial cells lining the blood vessels are joined so closely together that only lipophilic drugs can enter via passive diffusion [63]. Since triazolam needs to cross the blood–brain barrier to reach its receptor, one of the questions is whether efflux transport by P-glycoprotein influences the disposition of triazolam in the brain. The effect of P-glycoprotein on triazolam distribution was investigated by comparing

brain levels of ketoconazole and triazolam in P-glycoprotein-deficient mice and matched controls [145]. Interestingly, triazolam levels were the same in both mouse types, but P-glycoprotein-deficient mice had higher levels of ketoconazole. Coadministration of the two drugs did not alter their levels in the brain. Thus, while triazolam is not a P-glycoprotein substrate in the blood–brain barrier, ketoconazole may be transported by P-glycoprotein.

### 2.6 Conclusions

We have discussed the methods for investigating different aspects of drug–drug interactions. An initial estimate of clinical significance comes from *in vitro* studies with microsomes, which can be later confirmed with clinical trials. In order to differentiate between enteric and hepatic metabolism it is necessary to administer a drug in clinical trials orally and intravenously and compare the clearances from each administration. *In vitro* studies have also shown that, among the CYP3A isoforms, CYP3A4 is probably an important contributor to the metabolism of benzodiazepines. It is also important to specify the study population in clinical trials. As we have shown, there are significant differences between age groups, both in terms of clearance of benzodiazepines and in their pharmacodynamic responses to the drugs. The elderly seem particularly susceptible to the sedative effects of benzodiazepines. This effect may be partially due to lower hepatic clearance. However, the concentration–effect relationship was also more pronounced in the elderly, suggesting that the same concentration produces a more significant effect than in younger volunteers. While a substantial difference between males and females could not be consistently demonstrated in benzodiazepine trials, CYP3A is known to be influenced by the hormonal environment. Thus, sex differences should also be considered as variables in human studies.

## 3 CASE STUDY B: DRUG INTERACTIONS WITH HERBS AND NATURAL FOOD PRODUCTS

### 3.1 Introduction

The consumption of herbal supplements is increasing among both adults and children. Although estimates of the percentage of adults using herbals varies depending on survey methods (ranging from 12% to over 50%), the sales of herbals have been increasing about 25% a year with the approximate yearly sales now being over 4 billion dollars [146]. Herbal usage has become especially common among chronically ill patients, including cancer and HIV patients, who are on multidrug regimens and have an increased risk of drug–herbal interactions. There have also been reports of as many as 45% of adults giving their children herbal supplements as well as 45% of pregnant women consuming herbals [147, 148].

Increased consumption of herbals is a concern for drug manufacturers since the FDA requires proper labeling of pharmaceuticals regarding possible adverse interactions. For example, the consumption of grapefruit juice is contraindicated on the labeling of several drugs, such as lovastatin, atorvastatin, and cyclosporine. Grapefruit juice is also reported to interact with other pharmaceuticals including calcium channel antagonists, sedatives, and HIV protease inhibitors [149]. The FDA is very

aware of clinically significant interactions with herbs and natural food products. In February of 2000 the FDA issued a warning after clinical trials showed a substantial decrease in the plasma levels of indinavir when combined with St. John's wort. The FDA also warned against coadministration of St. John's wort with all other nonnucleoside reverse transcriptase protease, inhibitors, such as nevirapine [147, 150].

While herbs consist of chemical compounds just like pharmaceuticals, it is challenging to conduct a systematic investigation of potential herb–drug interactions because (1) herbs are regulated differently than pharmaceuticals, (2) they consist of several (sometimes over 100) active ingredients, and (3) there are many different preparations and batch-to-batch variations. Many reports of herb–drug interactions originate primarily from case studies rather than controlled trial studies, therefore cause–effect relationships are difficult to prove [151]. In this section we discuss the regulation of herbal products, mechanisms of drug–herb interactions, and clinically significant adverse events.

### 3.2 Regulation of Herbal Supplements

Herbs are considered dietary supplements and are regulated in the United States by the Dietary Supplement Health and Education Act of 1994. This legislation is significantly more lenient than European and Japanese regulations, which treat herbal products the same way as pharmaceuticals [152]. Herbal product manufacturers are allowed to make claims on the function of herbs without proof regarding safety and efficacy; however, they must also state that these statements have not been evaluated by the FDA. Under the Dietary Supplement Health and Education Act, the herb manufacturers are responsible for monitoring the safety of herbal products [153] but the safety guidelines are not well established. Surprisingly, according to surveys almost 60% of adults believe that herbs need approval before being sold [152]. However, manufacturers of herbal and dietary supplements do not need to report adverse events to the FDA, and the FDA needs to prove that supplements are unsafe before they can be removed from the market [154].

Commercially available herbal preparations vary substantially in their compositions and frequently their labeling does not reflect their true compositions. Thin-layer chromatographic analysis of 59 commercial echinacea preparations showed that the labeling was consistent with the content in only 31 samples and 6 samples contained no echinacea at all [155]. Another study with ginseng products showed that there was a 15–200-fold variation in the concentration of ginsenosides and eleutherosides, which are considered the biologically active ingredients [146, 156].

The most commonly used herbs in the United States are echninacea, garlic, *Ginkgo biloba*, saw palmetto, ginseng, grape seed extract, green tea, St. John's wort, bilberry, and aloe [146]. In order to illustrate the major concepts regarding herbal and food–drug interactions, this case study discusses primarily St. John's wort, *Ginkgo biloba*, garlic, and echinacea.

### 3.3 Influence of Herbal Constituents on Drug Metabolism *In Vitro*

Herbs are complex mixtures of chemicals such as flavonoids, polyphenols, alkaloids, triterpenoids, and anthraquinones—all of which have been shown to modulate CYP450 activity *in vitro* [157]. As will be discussed, multiple components may

induce or inhibit an enzyme, but frequently they do not lead to clinically significant interactions.

Hyperforin and hypericin are the two components in St. John's wort believed to be responsible for the antidepressant effects. However, the inhibition and induction capacities of St. John's wort are attributed primarily to hyperforin [158]. *In vitro* studies with cDNA-expressed cytochrome P450 enzyme systems showed that hyperforin is a significant inhibitor of CYP1A2, CYP2C9, CYP2C19, and CYP3A4 with $IC_{50}$ values of 3.87, 0.01, 0.02, and 4.20 μM, respectively. In general, potent inhibitors are considered those with $IC_{50}$ values less than 10 μM [159], although clinically significant interactions are expected when the $IC_{50}$ values are less than 1 μM. Hyperforin also induces intestinal P-glycoprotein in isolated cells [160] and volunteers [161]. This effect is also thought to be mediated by the binding of hyperforin to the pregnane X receptor (PXR) inducing CYP2B6, CYP3A4, and P-glycoprotein [147, 162, 163]. Other herbal inhibitors of P-glycoprotein include ginsenosides, piperine, sylmarin, and catechins from green tea [164].

Consumption of garlic is thought to be beneficial for many health conditions including atherosclerosis and peptic ulcer disease and as a chemopreventative agent for gastric tumors [165]. However, extracts from garlic, in particular, the volatile components, diallyl sulfide, dipropyl sulfide, and dipropyl disulfide, have been shown to interact with a number of CYP450s *in vitro* [166, 167]. CYP2C9, CYP2C19, CYP3A4, CYP3A5, CYP3A7, and CYP2E1 were shown to be inhibited, whereas CYP1A2, CYP2B, and CYP3A were induced [168–170]. Interestingly, garlic seemed to have little effect on CYP2D6 activity (less than 10% inhibition). In the study by Foster et al. [169], modulation of CYP450 was compared across several garlic preparations, specifically garlic oil, odorless garlic preparations, freeze-dried garlic tablets, and three types of garlic bulbs (common, elephant, and Chinese). CYP3A4 was significantly inhibited by all of the products and in the case of the tablets inhibition was over 95%. Inhibition of three CYP3A isoforms (CYP3A4, CYP3A5, CYP3A7) was also compared by the three types of garlic bulbs. All of the isoforms were inhibited, but CYP3A4 was affected the most with inhibition ranging from almost 40% to over 58%. Interestingly, these garlic preparations had little effect on P-glycoprotein compared with verapamil, the positive control [169].

An investigation of 29 constituents of *Ginkgo biloba* revealed inhibitory potential for all of the five major cytochrome P450s (CYP1A2, CYP2C9, CYP2C19, CYP2D6, and CYP3A) in human liver microsomes [171]. Quercetin and amentoflavone inhibited three or more isoforms and nine constituents inhibited CYP3A. The mean $IC_{50}$ values ranged from 0.035 to over 100 μM, although most of the values were between 1 and 30 μM. In cDNA-expressed cytochrome P450 enzyme systems (CYP1A2, CYP2C9, CYP2C19, CYP2D6, and CYP3A4), ginkgolic acids I and II inhibited CYP1A2, CYP2C9, CYP2C19, CYP2D6, and CYP3A4 with the $IC_{50}$ values being 4–5 μM for CYP1A2 and CYP2C19, around 2 μM for CYP2C9, 7–11 μM for CYP2D6, and 6–7 for CYP3A4 [159]. Interestingly, *Ginkgo biloba* increased clearance of the calcium channel blocker nicardipine in rats, suggesting induction of CYP3A2 [172]. Overall, *Ginkgo biloba* increased the expression of CYP2B1/2 and CYP3A1/2 in rats but did not affect CYP1A1/2, CYP2C11, or CYP4A1 [172, 173].

While *Ginkgo biloba* modulated CYP3A *in vitro* and in animal studies, there are few controlled clinical case studies suggesting herb–drug interactions through

enzyme modulation by *Ginkgo biloba* [166, 174]. It must be noted that the absence of evidence for clinically significant CYP3A4 inhibition in humans does not necessarily mean that such an interaction could not occur. *In vitro* studies have demonstrated potent inhibition of this enzyme and conflicting findings on *Ginkgo biloba* in clinical trials could be due to variations in *Ginkgo biloba* preparations [175].

### 3.4 Clinically Significant Drug Interactions

While many of these investigations have been successful at identifying drug-inducing or drug-inhibiting components, many *in vitro* studies did not correctly predict clinically significant interactions [166]. As mentioned in Case Study A, clinically significant drug interactions are relatively rare and occur primarily when the therapeutic index is very narrow or if the perpetrator is a very significant modulator of a metabolic enzyme. Table 3 shows representative examples of clinically significant drug–herb interactions.

Warfarin, an example of a drug with a narrow therapeutic index, is metabolized almost exclusively by CYP2C9 [176]. In a small clinical study, American ginseng has been shown to cause a slight but statistically significant decrease in the AUC of warfarin as well as the INR (International Normalized Ratio, a number used to characterize blood clotting) of patients, suggesting that ginseng could decrease the efficacy of warfarin. *Ginkgo biloba* contains several inhibitors of CYP2C9, the most potent being amentoflavone with an $IC_{50}$ around 0.035 μM [171]. Several anectodal case reports suggested potentiation of warfarin coagulation by *Ginkgo biloba*; this phenomenon could not be demonstrated in controlled clinical trials [22, 177].

Herb–drug interactions are a particular concern for the elderly since they (1) are usually on multidrug regimens, increasing the likelihood of interactions, and (2) have decreased hepatic and renal clearance compared to a younger population [178]. Clinical studies with hyperforin in healthy elderly volunteers (60–76 years) showed a 140% induction of CYP3A4 compared to a 98% induction in young volunteers, despite the fact that the young volunteers were administered 2.5 times the dose of the elderly. Interestingly, the serum concentrations were also similar (42.6 ng/mL in

**TABLE 3 Clinically Significant Interactions Between Herbs and Prescription Drugs**

| Herb/Food Product | Clinical Interaction | Potential Mechanism |
|---|---|---|
| St. John's wort | Digoxin [181] | Induction of P-gp |
| | Simvastatin [182], alprazolam, midazolam [180, 185] | Induction of CYP3A4 |
| | Amitriptyline [183], indinavir [186], nevirapine [187] | Induction of CYP3A4 and P-gp |
| | Buspirone [184] | Inhibition of serotonin reuptake |
| Echinacea | Midazolam [188] | Induction of CYP3A4 |
| | Caffeine [188] | Induction CYP 1A2 |
| Garlic | Ritonavir [191], saquinavir [192] | Induction of transporter, not necessarily P-gp [169] |
| | Warfarin [195] | Enhancement of anticoagulation |
| *Ginko biloba* | Aspirin, warfarin [196], ibupofen [197], rofecoxib [147, 198] | Enhancement of anticoagulation |

young vs. 51.2 ng/mL in old) [179]. This data suggests that the elderly are particularly susceptible to the inductive effects of hyperforin on CYP3A4 activity.

The antidepressant effect of St. John's wort has been confirmed in clinical trials [164] and is attributed partially to hyperforin, which inhibits the uptake of neurotransmitters in synapses. However, St. John's wort has been linked to many clinically significant and even potentially fatal drug interactions. Long-term (2 weeks) St. John's wort administration has been shown to induce both intestinal and hepatic CYP3A4 [147, 180]. St. John's wort has been reported to interact with protease inhibitors, oral contraceptives, and anticoagulants, as well as cyclosporine, digoxin, amitryptiline, and theophylline. Effects of these interactions included organ rejection (with cyclosporine), unplanned pregnancies and intermenstrual bleeding (with contraceptives), and decreases in the efficacies of antiretrovirals and anticoagulants [158].

In a clinical study, St. John's wort was also shown to decrease plasma levels of digoxin, probably through the induction of P-glycoprotein [181]. Clinical studies also reported decreased plasma concentration of simvastatin but not pravastatin when combined with St. John's wort [182]. Due to hyperforin's ability to inhibit the reuptake of brain neurotransmitters, such as serotonin, St. John's wort can also interact with antidepressants and anxiolytic drugs. In clinical studies, St. John's wort decreased the plasma and urine concentrations of the tricyclic antidepressant amitriptyline, presumably through the induction of P-glycoprotein and CYP3A4 [183]. In a reported case study, a patient experienced serotonin syndrome after combining St. John's wort with the anxiolytic drug buspirone, which is a 5-HT1 receptor agonist [184]. In multiple clinical studies, the plasma levels of the CYP3A4 probes alprazolam and midazolam were decreased after treatment with St. John's wort [180, 185]. The plasma concentrations of the antiretroviral drugs indinavir [186] and nevirapine [187] were also reduced when the drugs were combined with St. John's wort. Such an interaction could have serious implications, including reduced treatment efficacy and drug resistance [147].

Echinacea represents 10% of the U.S. herb market and is commonly used for immune system stimulation [155]. Echinacea is also known to modulate CYP3A4 and in human studies it was observed to induce hepatic CYP3A4 but inhibit intestinal CYP3A4 [188]. Specifically, echinacea increased the systemic but not the oral clearance of midazolam by 34%. Increase in systemic clearance is probably due to induction of CYP3A4 (since in this case the contribution of intestinal CYP3A4 is considered negligible) whereas with oral clearance, hepatic induction might be counteracted with intestinal inhibition of CYP3A4. Echinacea also reduced the oral clearance of caffeine (CYP1A2 substrate) and tolbutamide (CYP2C9 substrate), although the latter increase was not statistically significant. The clearance of the CYP2D6 substrate dextromethorphan was not affected. Thus, echinacea would be expected to interact with drugs that are substrates for CYP1A2 and CYP3A4 [188].

Garlic has been cultivated for over 5000 years as a culinary and medicinal herb and it was one of the best-selling herbs in 2002 [175, 189]. In clinical studies garlic did not influence the clearance of caffeine, dextromethorphan, and the CYP3A4 substrate alprazolam [190]. While these clinical studies suggest that garlic does not affect CYP3A4, garlic has been shown to decrease the AUC of ritonavir [191] and saquinavir [192]. Thus, the interaction may occur through a transporter, but not necessarily P-glycoprotein [169]. Garlic did not influence the metabolism of

acetaminophen [193], but it did enhance the anticoagulant effects of warfarin, leading to an increase in clotting time [194, 195].

*In vitro* studies reported that *Ginkgo biloba* is an inhibitor of CYP1A2, CYP2C9, CYP2C19, CYP2D6, and CYP3A4 [171]. Furthermore, several case studies have suggested that this interaction could be of clinical importance; however, this has not been confirmed by clinical studies. In case studies *Ginkgo biloba* has been reported to lead to increased bleeding in patients taking nonsteroidal anti-inflammatory drugs such as aspirin[196], ibuprofen [197], and the cyclo-oxygenase-2 inhibitor rofecoxib [147, 198], and in one case the combination of *Ginkgo biloba* and ibuprofen led to intracerebral bleeding and death [197]. While in a case study *Ginkgo biloba* has been reported to increase the anticoagulant effect of warfarin [196], a randomized clinical trial showed that recommended doses of *Ginkgo biloba* did not affect the pharmacokinetics or pharmacodynamics of warfarin [199].

### 3.5 Conclusions

*In vitro* and clinical studies have clearly demonstrated the potential for serious and possibly fatal interactions between herbal supplements and prescription medications. In general, the public views these supplements as natural and safe, and more than half of adults believe that herbals need to be approved before reaching the market. With the FDA becoming increasingly aware of herb–drug interactions and requiring proper labeling (as in the case of grapefruit juice), drug developers need to consider the possibility of coadministration with herbal supplements. While it would be impossible to examine all the possible herb–drug interactions, the *in vitro* effects of most commonly used herbs have been examined. Thus, based on metabolic studies, one could predict which herbs a drug might interact with.

St. John's wort is of particular concern, due to its potent induction of both CYP3A4 and P-glycoprotein. Clinical studies have confirmed that this herb can lead to potent interactions with several prescription drugs such as digoxin, antidepressants, benzodiazepines, and antiretrovirals. In the case of cyclosporine, an interaction could lead to organ rejection and be fatal. Other herbs such as echinacea, garlic, and *Ginkgo biloba* have also been reported to interact with prescription drugs. While these are some of the most commonly used herbs, they are not the only ones suspected of drug interactions. Other herbs such as kava-kava, ginger, ginseng, and green tea have also been reported to lead to adverse effects when coadministered with certain prescription drugs, and further clinical studies are necessary to the assess the effects of these herbs on medications [147].

## ACKNOWLEDGMENTS

Supported by Grants AT-003540, AG-017880, and AI-058784 from the United States Department of Health and Human Services.

## REFERENCES

1. Pratt C, Brown AM, Rampe D, Mason J, Russell T, Reynolds R, Ahlbrandt R. Cardiovascular safety of fexofenadine HCl. *Clin and Exp Allergy* 1999;29(Suppl 3): 212–216.

2. Jorgensen T, Johansson S, Kennerfalk A, Wallander MA, Svardsudd K. Prescription drug use, diagnoses, and healthcare utilization among the elderly. *Ann Pharmacother* 2001;35:1004–1009.
3. Hanlon JT, Fillenbaum GG, Ruby CM, Gray S, Bohannon A. Epidemiology of over-the-counter drug use in community dwelling elderly: United States perspective. *Drugs Aging* 2001;18:123–131.
4. Nix DE, Di Cicco RA, Miller AK, Boyle DA, Boike SC, Zariffa N, Jorkasky DK, Schentag JJ. The effect of low-dose cimetidine (200 mg twice daily) on the pharmacokinetics of theophylline. *J Clin Pharmacol* 1999;39:855–865.
5. Anantharaju A, Feller A, Chedid A. Aging liver. A review. *Gerontology* 2002;48: 343–353.
6. Blower P, de Wit R, Goodin S, Aapro M. Drug–drug interactions in oncology: Why are they important and can they be minimized? *Crit Rev Oncol/Hematol* 2005;55:117–142.
7. Werneke U, Earl J, Seydel C, Horn O, Crichton P, Fannon D. Potential health risks of complementary alternative medicines in cancer patients. *Br J Cancer* 2004;90:408–413.
8. Greenblatt DJ, von Moltke LL, Harmatz JS, Chen G, Weemhoff JL, Jen C, Kelley CJ, LeDuc BW, Zinny MA. Time course of recovery of cytochrome P450 3A function after single doses of grapefruit juice. *Clin Pharmacol Ther* 2003;74:121–129.
9. Greenblatt DJ, Patki KC, von Moltke LL, Shader RI. Drug interactions with grapefruit juice: an update. *J Clin Psychopharmacol* 2001;21:357–359.
10. Malhotra S, Bailey DG, Paine MF, Watkins PB. Seville orange juice–felodipine interaction: comparison with dilute grapefruit juice and involvement of furocoumarins. *Clin Pharmacol Ther* 2001;69:14–23.
11. Hidaka M, Okumura M, Fujita K, Ogikubo T, Yamasaki K, Iwakiri T, Setoguchi N, Arimori K. Effects of pomegranate juice on human cytochrome P450 3A (CYP3A) and carbamazepine pharmacokinetics in rats. *Drug Metab Dispos* 2005;33:644–648.
12. Hidaka M, Fujita K, Ogikubo T, Yamasaki K, Iwakiri T, Okumura M, Kodama H, Arimori K. Potent inhibition by star fruit of human cytochrome P450 3A (CYP3A) activity. *Drug Metab Dispos* 2004;32:581–583.
13. Martin-Facklam M, Burhenne J, Ding R, Fricker R, Mikus G, Walter-Sack I, Haefeli WE. Dose-dependent increase of saquinavir bioavailability by the pharmaceutic aid Cremophor EL. *Br J Clin Pharmacol* 2002;53:576–581.
14. Tayrouz Y, Ding R, Burhenne J, Riedel KD, Weiss J, Hoppe-Tichy T, Haefeli WE, Mikus G. Pharmacokinetic and pharmaceutic interaction between digoxin and Cremophor RH40. *Clin Pharmacol Ther* 2003;73:397–405.
15. Balayssac D, Authier N, Cayre A, Coudore F. Does inhibition of P-glycoprotein lead to drug–drug interactions? *Toxicol Lett* 2005;156:319–329.
16. Ochs HR, Steinhaus E, Locniskar A, Knüchel M, Greenblatt DJ. Desmethyldiazepam kinetics after intravenous, intramuscular, and oral administration of clorazepate dipotassium. *Klin Wochenschrift* 1982;60:411–415.
17. Venkatakrishnan K, Shader RI, Greenblatt DJ. Concepts and mechanisms of drug disposition and drug interactions. In Ciraulo DA, Shader RI, Greenblatt DJ, Creelman W, Eds. *Drug Interactions in Psychiatry*. Philadelphia: Lippincott Williams &Wilkins; 2005, pp 1–46.
18. Kato R, Ueno K, Imano H, Kawai M, Kuwahara S, Tsuchishita Y, Yonezawa E, Tanaka K. Impairment of ciprofloxacin absorption by calcium polycarbophil. *J Clin Pharmacol* 2002;42:806–811.
19. Greenblatt DJ, Sellers EM, Koch-Weser J. Importance of protein binding for the interpretation of serum or plasma drug concentrations. *J Clin Pharmacol* 1982;22:259–263.

20. MacKichan JJ. Protein binding drug displacement interactions fact or fiction? *Clin Pharmacokinet* 1989;16:65–73.
21. Sansom LN, Evans AM. What is the true clinical significance of plasma protein binding displacement interactions? *Drug Safety* 1995;12:227–233.
22. Greenblatt DJ, von Moltke LL. Interaction of warfarin with drugs, natural substances, and foods. *J Clin Pharmacol* 2005;45:127–132.
23. Benet LZ, Hoener BA. Changes in plasma protein binding have little clinical relevance. *Clin Pharmacol Ther* 2002;71:115–121.
24. O'Reilly RA, Trager WF, Motley CH, Howald W. Stereoselective interaction of phenylbutazone with [$^{12}$C/$^{13}$C]warfarin pseudoracemates in man. *J Clin Invest* 1980;65: 746–753.
25. Banfield C, O'Reilly R, Chan E, Rowland M. Phenylbutazone–warfarin interaction in man: further stereochemical and metabolic considerations. *Br J Clin Pharmacol* 1983;16:669–675.
26. Rolan PE. Plasma protein binding displacement interactions—why are they still regarded as clinically important? *Br J Clin Pharmacol* 1994;37:125–128.
27. Hansen JM, Christensen LK. Drug interactions with oral sulphonylurea hypoglycaemic drugs. *Drugs* 1977;13:24–34.
28. Venkatakrishnan K, von Moltke LL, Obach RS, Greenblatt DJ. Drug metabolism and drug interactions: application and clinical value of *in vitro* models. *Curr Drug Metab* 2003;4:423–459.
29. Williams JA, Hyland R, Jones BC, Smith DA, Hurst S, Goosen TC, Peterkin V, Koup JR, Ball SE. Drug–drug interactions for UDP-glucuronosyltransferase substrates: a pharmacokinetic explanation for typically observed low exposure ($AUC_i$/AUC) ratios. *Drug Metab Dispos Biol Fate Chem* 2004;32:1201–1208.
30. Spina E, Avenoso A, Campo GM, Caputi AP, Perucca E. Phenobarbital induces the 2-hydroxylation of desipramine. *Ther Drug Monitor* 1996;18:60–64.
31. Ono S, Mihara K, Suzuki A, Kondo T, Yasui-Furukori N, Furukori H, de Vries R, Kaneko S. Significant pharmacokinetic interaction between risperidone and carbamazepine: its relationship with CYP2D6 genotypes. *Psychopharmacology (Berl)* 2002;162:50–54.
32. Parkinson A. Biotransformation of xenobiotics. In Klaassen C, Ed. *Casarett & Doull's Toxicology*. New York: McGraw-Hill; 1996, pp 113–186.
33. von Moltke LL, Greenblatt DJ, Duan SX, Harmatz JS, Shader RI. *In vitro* prediction of the terfenadine–ketoconazole pharmacokinetic interaction. *J Clin Pharmacol* 1994;34: 1222–1227.
34. Holbrook AM, Pereira JA, Labiris R, McDonald H, Douketis JD, Crowther M, Wells PS. Systematic overview of warfarin and its drug and food interactions. *Arch Intern Med* 2005;165:1095–1106.
35. Thummel KE, Wilkinson GR. *In vitro* and *in vivo* drug interactions involving human CYP3A. *Annu Rev Pharmacol Toxicol* 1998;38:389–430.
36. Baciewicz AM, and Baciewicz FA Jr. Ketoconazole and fluconazole drug interactions. *Arch Intern Med* 1993;153:1970–1976.
37. Gillum JG, Israel DS, Polk RE. Pharmacokinetic drug interactions with antimicrobial agents. *Clin Pharmacokinet* 1993;25:450–482.
38. Advani R, Fisher GA, Lum BL, Hausdorff J, Halsey J, Litchman M, Sikic BI. A phase I trial of doxorubicin, paclitaxel, and valspodar (PSC 833), a modulator of multidrug resistance. *Clin Cancer Res* 2001;7:1221–1229.
39. Hughes H, Biehl J, Jones A, Schmidt L. Metabolism of isoniazid in man as related to the occurence of peripheral neuritis. *Am Rev Tuberculosis* 1954;70:266–273.

40. Farkas D, Tannenbaum SR. *In vitro* methods to study chemically-induced hepatotoxicity: a literature review. *Curr Drug Metab* 2005;6:111–125.
41. Davis MP, Homsi J. The importance of cytochrome P450 monooxygenase CYP2D6 in palliative medicine. *Support Care Cancer* 2001;9:442–451.
42. Schwarz UI. Clinical relevance of genetic polymorphisms in the human CYP2C9 gene. *Eur J Clin Invest* 2003;33 (Suppl 2):23–30.
43. Ingelman-Sundberg M. Genetic polymorphisms of cytochrome P450 2D6 (CYP2D6): clinical consequences, evolutionary aspects and functional diversity. *Pharmacogenomics J* 2005;5:6–13.
44. Rettie AE, Jones JP. Clinical and toxicological relevance of CYP2C9: drug–drug interactions and pharmacogenetics. *Annu Rev Pharmacol Toxicol* 2005;45:477–494.
45. Zhou S, Chan E, Lim LY, Boelsterli UA, Li SC, Wang J, Zhang Q, Huang M, Xu A. Therapeutic drugs that behave as mechanism-based inhibitors of cytochrome P450 3A4. *Curr Drug Metab* 2004;5:415–442.
46. Alfaro CA, Piscitelli SC. Drug interactions. In Atkinson AJ, Daniels CE, Dedrick RL, Grudzinskas CV, Markey SP. eds. *Principles of Clinical Pharmacoglogy*. Boston: Academic Press; 2001, pp 167–180.
47. Ma Q, Lu AY. Origins of individual variability in P4501A induction. *Chem Res Toxicol* 2003;16:249–260.
48. Hustert E, Zibat A, Presecan-Siedel E, Eiselt R, Mueller R, Fuss C, Brehm I, Brinkmann U, Eichelbaum M, Wojnowski L, Burk O. Natural protein variants of pregnane X receptor with altered transactivation activity toward CYP3A4. *Drug Metab Dispos Biol Fate Chem* 2001;29:1454–1459.
49. Cao H, Hegele RA. Human aryl hydrocarbon receptor nuclear translocator gene (ARNT) D/N511 polymorphism. *J Hum Genet* 2000;45:92–93.
50. Tang C, Lin JH, Lu AY. Metabolism-based drug–drug interactions: what determines individual variability in cytochrome P450 induction? *Drug Metab Dispos Biol Fate Chem* 2005;33:603–613.
51. Chang TK, Bandiera SM, Chen J. Constitutive androstane receptor and pregnane X receptor gene expression in human liver: interindividual variability and correlation with CYP2B6 mRNA levels. *Drug Metab Dispos Biol Fate Chem* 2003;31:7–10.
52. Kerb R, Aynacioglu AS, Brockmöller J, Schlagenhaufer R, Bauer S, Szekeres T, Hamwi A, Fritzer-Szekeres M, Baumgartner C, Ongen HZ, Guzelbey P, Roots I, Brinkmann U. The predictive value of MDR1, CYP2C9, and CYP2C19 polymorphisms for phenytoin plasma levels. *Pharmacogenomics J* 2001;1:204–210.
53. Kim RB, Leake BF, Choo EF, Dresser GK, Kubba SV, Schwarz UI, Taylor A, Xie HG, McKinsey J, Zhou S, Lan LB, Schuetz JD, Schuetz EG, Wilkinson GR. Identification of functionally variant MDR1 alleles among European Americans and African Americans. *Clin Pharmacol Ther* 2001;70:189–199.
54. Begley DJ. ABC transporters and the blood–brain barrier. *Curr Pharm Design* 2004;10:1295–1312.
55. Chandra P, Brouwer KL. The complexities of hepatic drug transport: current knowledge and emerging concepts. *Pharm Res* 2004;21:719–735.
56. Okamura N, Hirai M, Tanigawara Y, Tanaka K, Yasuhara M, Ueda K, Komano T, Hori R. Digoxin–cyclosporin A interaction: modulation of the multidrug transporter P-glycoprotein in the kidney. *J Pharmacol Exp Ther* 1993;266:1614–1619.
57. Saito T, Zhang ZJ, Tokuriki M, Ohtsubo T, Shibamori Y, Yamamoto T, Saito H. Cyclosporin A inhibits the extrusion pump function of P-glycoprotein in the inner ear of mice treated with vinblastine and doxorubicin. *Brain Res* 2001;901:265–270.

58. Sadeque AJ, Wandel C, He H, Shah S, Wood AJ. Increased drug delivery to the brain by P-glycoprotein inhibition. *Clin Pharmacol Ther* 2000;68:231–237.
59. Schinkel AH, Wagenaar E, Mol CA, van Deemter L. P-glycoprotein in the blood–brain barrier of mice influences the brain penetration and pharmacological activity of many drugs. *J Clin Invest* 1996;97:2517–2524.
60. Hughes CS, Vaden SL, Manaugh CA, Price GS, Hudson LC. Modulation of doxorubicin concentration by cyclosporin A in brain and testicular barrier tissues expressing P-glycoprotein in rats. *J Neuro-Oncol* 1998;37:45–54.
61. Kang W, Weiss M. Influence of P-glycoprotein modulators on cardiac uptake, metabolism, and effects of idarubicin. *Pharm Res* 2001;18:1535–1541.
62. Kramer WG, Reuning RH. Use of area under the curve to estimate absolute bioavailability of digoxin. *J Pharm Sci* 1978;67:141–142.
63. Lin JH. Drug–drug interaction mediated by inhibition and induction of P-glycoprotein. *Adv Drug Deliv Rev* 2003;55:53–81.
64. Lin JH. Human immunodeficiency virus protease inhibitors. From drug design to clinical studies. *Adv Drug Deliv Rev* 1997;27:215–233.
65. Szabo K, Welker E, Bakos E, Müller M, Roninson I, Varadi A, Sarkadi B. Drug-stimulated nucleotide trapping in the human multidrug transporter MDR1. Cooperation of the nucleotide binding domains. *J Biol Chem* 1998;273:10132–10138.
66. Shapiro AB, Ling V. Positively cooperative sites for drug transport by P-glycoprotein with distinct drug specificities. *Eur J Biochem* 1997;250:130–137.
67. Ayesh S, Shao YM, Stein WD. Co-operative, competitive and non-competitive interactions between modulators of P-glycoprotein. *Biochim Biophys Acta* 1996;1316:8–18.
68. Fardel O, Lecureur V, Corlu A, Guillouzo A. P-glycoprotein induction in rat liver epithelial cells in response to acute 3-methylcholanthrene treatment. *Biochem Pharmacol* 1996;51:1427–1436.
69. Fardel O, Morel F, Guillouzo A. P-glycoprotein expression in human, mouse, hamster and rat hepatocytes in primary culture. *Carcinogenesis* 1993;14:781–783.
70. Chin KV, Chauhan SS, Pastan I, Gottesman MM. Regulation of mdr RNA levels in response to cytotoxic drugs in rodent cells. *Cell Growth Differ* 1990;1:361–365.
71. Jette L, Beaulieu E, Leclerc JM, Beliveau R. Cyclosporin A treatment induces overexpression of P-glycoprotein in the kidney and other tissues. *Am J Physiol* 1996; 270:F756–F765.
72. Boffito M, Maitland D, Samarasinghe Y, Pozniak A. The pharmacokinetics of HIV protease inhibitor combinations. *Curr Opin Infect Dis* 2005;18:1–7.
73. van Heeswijk RP, Veldkamp A, Mulder JW, Meenhorst PL, Lange JM, Beijnen JH, Hoetelmans RM. Combination of protease inhibitors for the treatment of HIV-1-infected patients: a review of pharmacokinetics and clinical experience. *Antiviral Ther* 2001;6:201–229.
74. Moyle G. Use of HIV protease inhibitors as pharmacoenhancers. *AIDS Reader* 2001;11:87–98.
75. Cvetkovic RS, Goa KL. Lopinavir/ritonavir: a review of its use in the management of HIV infection. *Drugs* 2003;63:769–802.
76. Scott JD. Simplifying the treatment of HIV infection with ritonavir-boosted protease inhibitors in antiretroviral-experienced patients. *Am J Health-System Pharm* 2005; 62:809–815.
77. Moyle GJ, Back D. Principles and practice of HIV-protease inhibitor pharmacoenhancement. *HIV Med* 2001;2:105–113.

78. Kashuba AD. Drug-drug interactions and the pharmacotherapy of HIV infection. *Topics HIV Med* 2005;13:64–69.
79. von Hentig NH, Mueller A, Haberl A. The ATSAQ-1 cohort study; pharmakinetic interactions of atazanavir (ATV) and saquinavir (SAQ) in a ritonavir (RTV) boosted protease inhibitor regimen. *XV International AIDS Conference*; 2004.
80. de Smet K, Loyer P, Gilot D, Vercruysse A, Rogiers V, Guguen-Guillouzo C. Effects of epidermal growth factor on CYP inducibility by xenobiotics, DNA replication, and caspase activations in collagen I gel sandwich cultures of rat hepatocytes. *Biochem Pharmacol* 2001;61:1293–1303.
81. de Smet K, Beken S, Depreter M, Roels F, Vercruysse A, Rogiers V. Effect of epidermal growth factor in collagen gel cultures of rat hepatocytes. *Toxicol in Vitro* 1999;13: 579–585.
82. Ching KZ, Tenney KA, Chen J, Morgan ET. Suppression of constitutive cytochrome P450 gene expression by epidermal growth factor receptor ligands in cultured rat hepatocytes. *Drug Metab Dispos Biol Fate Chem* 1996;24:542–546.
83. Greuet J, Pichard L, Ourlin JC, Bonfils C, Domergue J, Le Treut P, Maurel P. Effect of cell density and epidermal growth factor on the inducible expression of CYP3A and CYP1A genes in human hepatocytes in primary culture. *Hepatology* 1997;25:1166–1175.
84. LeCluyse EL. Human hepatocyte culture systems for the *in vitro* evaluation of cytochrome P450 expression and regulation. *Eur J Pharm Sci* 2001;13:343–368.
85. Moghe PV, Coger RN, Toner M, Yarmush ML. Cell–cell interactiOns are essential for MaIntenance of hepatocyte FunctiOn In collagen gel but not On Matrigel. *Biotechnol Bioeng* 1997;56:706–711.
86. Moghe PV, Berthiaume F, Ezzell RM, Toner M, Tompkins RG, Yarmush ML. Culture matrix configuration and composition in the maintenance of hepatocyte polarity and function. *Biomaterials* 1996;17:373–385.
87. LeCluyse E, Madan A, Hamilton G, Carroll K, DeHaan R, Parkinson A. Expression and regulation of cytochrome P450 enzymes in primary cultures of human hepatocytes. *J Biochem Mol Toxicol* 2000;14:177–188.
88. Raucy JL. Regulation of CYP3A4 expression in human hepatocytes by pharmaceuticals and natural products. *Drug Metab Dispos* 2003;31:533–539.
89. Mills JB, Rose KA, Sadagopan N, Sahi J, de Morais SM. Induction of drug metabolism enzymes and MDR1 using a novel human hepatocyte cell line. *J Pharmacol Exp Ther* 2004;309:303–309.
90. Biagini CP, Bender V, Borde F, Boissel E, Bonnet MC, Masson MT, Cassio D, Chevalier S. Cytochrome P450 expression-induction profile and chemically mediated alterations of the WIF-B9 cell line. *Biol Cell* 2005;98:23–32.
91. Polli JW, Wring SA, Humphreys JE, Huang L, Morgan JB, Webster LO, Serabjit-Singh CS. Rational use of *in vitro* P-glycoprotein assays in drug discovery. *J Pharmacol Exp Ther* 2001;299:620–628.
92. Tang F, Horie K, Borchardt RT. Are MDCK cells transfected with the human MRP2 gene a good model of the human intestinal mucosa? *Pharm Res* 2002;19:773–779.
93. Taub ME, Podila L, Ely D, Almeida I. Functional assessment of multiple P-glycoprotein (P-gp) probe substrates: influence of cell line and modulator concentration on P-gp activity. *Drug Metab Dispos Biol Fate Chem* 2005;33:1679–1687.
94. Taipalensuu J, Tornblom H, Lindberg G, Einarsson C, Sjoqvist F, Melhus H, Garberg P, Sjostrom B, Lundgren B, Artursson P. Correlation of gene expression of ten drug efflux proteins of the ATP-binding cassette transporter family in normal human jejunum and

in human intestinal epithelial Caco-2 cell monolayers. *J Pharmacol Exp Ther* 2001;299: 164–170.

95. Prime-Chapman HM, Fearn RA, Cooper AE, Moore V, Hirst BH. Differential multidrug resistance-associated protein 1 through 6 isoform expression and function in human intestinal epithelial Caco-2 cells. *J Pharmacol Exp Ther* 2004;311:476–484.
96. Donato MT, Castell JV, Gomez-Lechon MJ. Characterization of drug metabolizing activities in pig hepatocytes for use in bioartificial liver devices: comparison with other hepatic cellular models. *J Hepatol* 1999;31:542–549.
97. Omasa T, Kim K, Hiramatsu S, Katakura Y, Kishimoto M, Enosawa S, Ohtake H. Construction and evaluation of drug-metabolizing cell line for bioartificial liver support system. *Biotechnol Prog* 2005;21:161–167.
98. Iwahori T, Matsuura T, Maehashi H, Sugo K, Saito M, Hosokawa M, Chiba K, Masaki T, Aizaki H, Ohkawa K, Suzuki T. CYP3A4 inducible model for *in vitro* analysis of human drug metabolism using a bioartificial liver. *Hepatology* 2003;37:665–673.
99. Bugrim A, Nikolskaya T, Nikolsky Y. Early prediction of drug metabolism and toxicity: systems biology approach and modeling. *Drug Discov Today* 2004;9:127–135.
100. Korolev D, Balakin KV, Nikolsky Y, Kirillov E, Ivanenkov YA, Savchuk NP, Ivashchenko AA, Nikolskaya T. Modeling of human cytochrome P450-mediated drug metabolism using unsupervised machine learning approach. *J Med Chem* 2003;46:3631–3643.
101. Manga N, Duffy JC, Rowe PH, Cronin MT. Structure-based methods for the prediction of the dominant P450 enzyme in human drug biotransformation: consideration of CYP3A4, CYP2C9, CYP2D6. *SAR QSAR Environ Res* 2005;16:43–61.
102. Jones JP, He M, Trager WF, Rettie AE. Three-dimensional quantitative structure–activity relationship for inhibitors of cytochrome P4502C9. *Drug Metab Dispos Biol Fate Chem* 1996;24:1–6.
103. Bonnabry P, Sievering J, Leemann T, Dayer P. Quantitative drug interactions prediction system (Q-DIPS): a computer-based prediction and management support system for drug metabolism interactions. *Eur J Clin Pharmacol* 1999;55:341–347.
104. Langer T, Eder M, Hoffmann RD, Chiba P, Ecker GF. Lead identification for modulators of multidrug resistance based on *in silico* screening with a pharmacophoric feature model. *Arch Pharm (Weinheim)* 2004;337:317–327.
105. Rebitzer S, Annibali D, Kopp S, Eder M, Langer T, Chiba P, Ecker GF, Noe CR. *In silico* screening with benzofurane- and benzopyrane-type MDR-modulators. *Farmaco* 2003; 58:185–191.
106. Ekins S, Nikolsky Y, Nikolskaya T. Techniques: application of systems biology to absorption, distribution, metabolism, excretion and toxicity. *Trends Pharmacol Sci* 2005; 26:202–209.
107. Segel I. Enzymes. *Biochemical Calculations*. Hoboken, NJ: John Wiley & Sons, 1976; pp 208–323.
108. von Moltke LL, Greenblatt DJ, Grassi JM, Granda BW, Duan SX, Fogelman SM, Daily JP, Harmatz JS, Shader RI. Protease inhibitors as inhibitors of human cytochromes P450: high risk associated with ritonavir. *J Clin Pharmacol* 1998;38:106–111.
109. von Moltke LL, Greenblatt DJ, Schmider J, Wright CE, Harmatz JS, Shader RI. *In vitro* approaches to predicting drug interactions *in vivo*. *Biochem Pharmacol* 1998; 55:113–122.
110. Shou M, Mei Q, Ettore MW Jr, Dai R, Baillie TA, Rushmore TH. Sigmoidal kinetic model for two co-operative substrate-binding sites in a cytochrome P450 3A4 active site: an example of the metabolism of diazepam and its derivatives. *Biochem J* 1999;340 (Pt 3):845–853.

111. Greenblatt DJ, von Moltke LL, Daily JP, Harmatz JS, Shader RI. Extensive impairment of triazolam and alprazolam clearance by short-term low-dose ritonavir: the clinical dilemma of concurrent inhibition and induction. *J Clin Psychopharmacol* 1999;19: 293–296.
112. Obach RS. Inhibition of human cytochrome P450 enzymes by constituents of St. John's wort, an herbal preparation used in the treatment of depression. *J Pharmacol Exp Ther* 2000;294:88–95.
113. Jang SH, Wientjes MG, Au JL. Kinetics of P-glycoprotein-mediated efflux of paclitaxel. *J Pharmacol Exp Ther* 2001;298:1236–1242.
114. Hunter J, Jepson MA, Tsuruo T, Simmons NL, Hirst BH. Functional expression of P-glycoprotein in apical membranes of human intestinal Caco-2 cells. Kinetics of vinblastine secretion and interaction with modulators. *J Biol Chem* 1993;268:14991–14997.
115. Jang SH, Wientjes MG, Au JL. Interdependent effect of P-glycoprotein-mediated drug efflux and intracellular drug binding on intracellular paclitaxel pharmacokinetics: application of computational modeling. *J Pharmacol Exp Ther* 2003;304:773–780.
116. Hoffmaster KA, Zamek-Gliszczynski MJ, Pollack GM, Brouwer KL. Multiple transport systems mediate the hepatic uptake and biliary excretion of the metabolically stable opioid peptide [D-penicillamine2,5]enkephalin. *Drug Metab and Dispos Biol Fate Chem* 2005;33:287–293.
117. Zahlten RN, Stratman FW. The isolation of hormone-sensitive rat hepatocytes by a modified enzymatic technique. *Arch Biochem Biophys* 1974;163:600–608.
118. Bertz RJ, Granneman GR. Use of *in vitro* and *in vivo* data to estimate the likelihood of metabolic pharmacokinetic interactions. *Clin Pharmacokinet* 1997;32:210–258.
119. Greenblatt DJ, von Moltke LL, Harmatz JS, Durol AL, Daily JP, Graf JA, Mertzanis P, Hoffman JL, Shader RI. Differential impairment of triazolam and zolpidem clearance by ritonavir. *J Acquir Immune Defic Syndr* 2000;24:129–136.
120. Kronbach T, Mathys D, Umeno M, Gonzalez FJ, Meyer UA. Oxidation of midazolam and triazolam by human liver cytochrome P450IIIA4. *Mol Pharmacol* 1989;36:89–96.
121. Eberts FS Jr, Philopoulos Y, Reineke LM, Vliek RW. Triazolam disposition. *Clin Pharmacol Ther* 1981;29:81–93.
122. Phillips JP, Antal EJ, Smith RB. A pharmacokinetic drug interaction between erythromycin and triazolam. *J Clin Psychopharmacol* 1986;6:297–299.
123. Varhe A, Olkkola KT, Neuvonen PJ. Oral triazolam is potentially hazardous to patients receiving systemic antimycotics ketoconazole or itraconazole. *Clin Pharmacol Ther* 1994;56:601–607.
124. von Moltke LL, Greenblatt DJ, Harmatz JS, Duan SX, Harrel LM, Cotreau-Bibbo MM, Pritchard GA, Wright CE, Shader RI. Triazolam biotransformation by human liver microsomes *in vitro*: effects of metabolic inhibitors and clinical confirmation of a predicted interaction with ketoconazole. *J Pharmacol Exp Ther* 1996;276:370–379.
125. Villikka K, Kivistö KT, Backman JT, Olkkola KT, Neuvonen PJ. Triazolam is ineffective in patients taking rifampin. *Clin Pharmacol Ther* 1997;61:8–14.
126. Hsu A, Granneman GR, Bertz RJ. Ritonavir. Clinical pharmacokinetics and interactions with other anti-HIV agents. *Clin Pharmacokinet* 1998;35:275–291.
127. von Moltke LL, Greenblatt DJ, Schmider J, Duan SX, Wright CE, Harmatz JS, Shader RI. Midazolam hydroxylation by human liver microsomes *in vitro*: inhibition by fluoxetine, norfluoxetine, and by azole antifungal agents. *J Clin Pharmacol* 1996;36:783–791.
128. Wrighton SA, Ring BJ, Watkins PB, Vandenbranden M. Identification of a polymorphically expressed member of the human cytochrome P-450III family. *Mol Pharmacol* 1989;36:97–105.

129. Tsunoda SM, Velez RL, von Moltke LL, Greenblatt DJ. Differentiation of intestinal and hepatic cytochrome P450 3A activity with use of midazolam as an *in vivo* probe: effect of ketoconazole. *Clin Pharmacol Ther* 1999;66:461–471.
130. Schuetz JD, Beach DL, Guzelian PS. Selective expression of cytochrome P450 CYP3A mRNAs in embryonic and adult human liver. *Pharmacogenetics* 1994;4:11–20.
131. Kuehl P, Zhang J, Lin Y, Lamba J, Assem M, Schuetz J, Watkins PB, Daly A, Wrighton SA, Hall SD, Maurel P, Relling M, Brimer C, Yasuda K, Venkataramanan R, Strom S, Thummel K, Boguski MS, Schuetz E. Sequence diversity in CYP3A promoters and characterization of the genetic basis of polymorphic CYP3A5 expression. *Nat Genet* 2001;27:383–391.
132. Patki KC, von Moltke LL, Greenblatt DJ. *In vitro* metabolism of midazolam, triazolam, nifedipine, and testosterone by human liver microsomes and recombinant cytochromes p450: role of CYP3A4 and CYP3A5. *Drug Metab Dispos Biol Fate Chem* 2003; 31:938–944.
133. Greenblatt DJ, von Moltke LL. Drug–drug interactions: clinical perspective. In Rodrigues AD, ed. *Drug–Drug Interactions*. New York: Marcel Dekker; 2002, pp 565–584.
134. Greenblatt DJ, Wright CE, Von Moltke LL, Harmatz JS, Ehrenberg BL, Harrel LM, Corbett K, Counihan M, Tobias S, Shader RI. Ketoconazole inhibition of triazolam and alprazolam clearance: differential kinetic and dynamic consequences. *Clin Pharmacol Ther* 1998;64:237–247.
135. Gorski JC, Jones DR, Haehner-Daniels BD, Hamman MA, O'Mara EM Jr, Hall SD. The contribution of intestinal and hepatic CYP3A to the interaction between midazolam and clarithromycin. *Clin Pharmacol Ther* 1998;64:133–143.
136. Cotreau MM, von Moltke LL, Greenblatt DJ. The influence of age and sex on the clearance of cytochrome P450 3A substrates. *Clin Pharmacokinet* 2005;44:33–60.
137. Greenblatt DJ, Harmatz JS, von Moltke LL, Wright CE, Shader RI. Age and gender effects on the pharmacokinetics and pharmacodynamics of triazolam, a cytochrome P450 3A substrate. *Clin Pharmacol Ther* 2004;76:467–479.
138. Greenblatt DJ, Divoll M, Abernethy DR, Moschitto LJ, Smith RB, Shader RI. Reduced clearance of triazolam in old age: relation to antipyrine oxidizing capacity. *Br J Clin Pharmacol* 1983;15:303–309.
139. Greenblatt DJ, Harmatz JS, Shapiro L, Engelhardt N, Gouthro TA, Shader RI. Sensitivity to triazolam in the elderly. *N Engl J Med* 1991;324:1691–1698.
140. Callachan H, Cottrell GA, Hather NY, Lambert JJ, Nooney JM, Peters JA. Modulation of the GABAA receptor by progesterone metabolites. *Proc R Soc London Ser B Biol Sci* 1987;231:359–369.
141. McAuley JW, Reynolds IJ, Kroboth FJ, Smith RB, Kroboth PD. Orally administered progesterone enhances sensitivity to triazolam in postmenopausal women. *J Clin Psychopharmacol* 1995;15:3–11.
142. McAuley JW, Kroboth PD, Stiff DD, Reynolds IJ. Modulation of [$^3$H]flunitrazepam binding by natural and synthetic progestational agents. *Pharmacol Biochem Behav* 1993;45:77–83.
143. Fahey JM, Pritchard GA, von Moltke LL, Pratt JS, Grassi JM, Shader RI, Greenblatt DJ. Effects of ketoconazole on triazolam pharmacokinetics, pharmacodynamics and benzodiazepine receptor binding in mice. *J Pharmacol Exp Ther* 1998;285:271–276.
144. Lin JH, Yamazaki M. Role of P-glycoprotein in pharmacokinetics: clinical implications. *Clin Pharmacokinet* 2003;42:59–98.
145. von Moltke LL, Granda BW, Grassi JM, Perloff MD, Vishnuvardhan D, Greenblatt DJ. Interaction of triazolam and ketoconazole in P-glycoprotein-deficient mice. *Drug Metab Dispos Biol Fate Chem* 2004;32:800–804.

146. Bent S, Ko R. Commonly used herbal medicines in the United States: a review. *Am J Med* 2004;116:478–485.
147. Izzo AA. Herb–drug interactions: an overview of the clinical evidence. *Fundam Clin Pharmacol* 2005;19:1–16.
148. Ernst E. Are herbal medicines effective? *Int J Clin Pharmacol Ther* 2004;42:157–159.
149. Greenblatt DJ, Patki KC, von Moltke LL, Shader RI. Drug interactions with grapefruit juice: an update. *J Clin Psychopharmacol* 2001;21:357–359.
150. James, JS. St. John's Warning: Do Not Combine with Protease Inhibitors, NNRTI's. http://www.aids.org/atn/a-337-02.html. 2005.
151. Fugh-Berman A, Ernst E. Herb–drug interactions: review and assessment of report reliability. *Br J Clin Pharmacol* 2001;52:587–595.
152. Lewis JD, Strom BL. Balancing safety of dietary supplements with the free market. *Ann Intern Med* 2002;136:616–618.
153. Dietary Supplement Health Education Act of 1994. http://www.fda.gov/opacom/laws/dshea.html.
154. Dietary Supplements: Questions Answers. US Food Drug Aministration; Center for Food Safety Applied Nutrition; 2001. http://www.cfsan.fda.gov/~dms/ds-faq.html.
155. Gilroy CM, Steiner JF, Byers T, Shapiro H, Georgian W. Echinacea and truth in labeling. *Arch Intern Med* 2003;163:699–704.
156. Harkey MR, Henderson GL, Gershwin ME, Stern JS, Hackman RM. Variability in commercial ginseng products: an analysis of 25 preparations. *Am J Clin Nutr* 2001;73:1101–1106.
157. Zhou S, Gao Y, Jiang W, Huang M, Xu A, Paxton JW. Interactions of herbs with cytochrome P450. *Drug Metab Rev* 2003;35:35–98.
158. Ioannides C. Pharmacokinetic interactions between herbal remedies and medicinal drugs. *Xenobiotica* 2002;32:451–478.
159. Zou L, Harkey MR, Henderson GL. Effects of herbal components on cDNA-expressed cytochrome P450 enzyme catalytic activity. *Life Sci* 2002;71:1579–1589.
160. Perloff MD, von Moltke LL, Stormer E, Shader RI, Greenblatt DJ. Saint John's wort: an *in vitro* analysis of P-glycoprotein induction due to extended exposure. *Br J Pharmacol* 2001;134:1601–1608.
161. Dresser GK, Schwarz UI, Wilkinson GR, Kim RB. Coordinate induction of both cytochrome P4503A and MDR1 by St John's wort in healthy subjects. *Clin Pharmacol Ther* 2003;73:41–50.
162. Moore LB, Goodwin B, Jones SA, Wisely GB, Serabjit-Singh CJ, Willson TM, Collins JL, Kliewer SA. St. John's wort induces hepatic drug metabolism through activation of the pregnane X receptor. *Proc Nat Acad Sci USA* 2000;97:7500–7502.
163. Wentworth JM, Agostini M, Love J, Schwabe JW, Chatterjee VK. St John's wort, a herbal antidepressant, activates the steroid X receptor. *J Endocrinol* 2000;166:R11–R16.
164. Zhou S, Lim LY, Chowbay B. Herbal modulation of P-glycoprotein. *Drug Metab Rev* 2004;36:57–104.
165. Levi F, Pasche C, La Vecchia C, Lucchini F, Franceschi S. Food groups and colorectal cancer risk. *Br J Cancer* 1999;79:1283–1287.
166. Delgoda R, Westlake AC. Herbal interactions involving cytochrome P450 enzymes: a mini review. *Toxicol Rev* 2004;23:239–249.
167. Guyonnet D, Belloir C, Suschetet M, Siess MH, Le Bon AM. Liver subcellular fractions from rats treated by organosulfur compounds from Allium modulate mutagen activation. *Mutat Res* 2000;466:17–26.

168. Wu CC, Sheen LY, Chen HW, Kuo WW, Tsai SJ, Lii CK. Differential effects of garlic oil and its three major organosulfur components on the hepatic detoxification system in rats. *J Agric Food Chem* 2002;50:378–383.

169. Foster BC, Foster MS, Vandenhoek S, Krantis A, Budzinski JW, Arnason JT, Gallicano KD, Choudri S. An *in vitro* evaluation of human cytochrome P450 3A4 and P-glycoprotein inhibition by garlic. *J Pharm Pharm Sci* 2001;4:176–184.

170. Chen HW, Tsai CW, Yang JJ, Liu CT, Kuo WW, Lii CK. The combined effects of garlic oil and fish oil on the hepatic antioxidant and drug-metabolizing enzymes of rats. *Br J Nutr* 2003;89:189–200.

171. von Moltke LL, Weemhoff JL, Bedir E, Khan IA, Harmatz JS, Goldman P, Greenblatt DJ. Inhibition of human cytochromes P450 by components of *Ginkgo biloba*. *J Pharm Pharmacol* 2004;56:1039–1044.

172. Shinozuka K, Umegaki K, Kubota Y, Tanaka N, Mizuno H, Yamauchi J, Nakamura K, Kunitomo M. Feeding of *Ginkgo biloba* extract (GBE) enhances gene expression of hepatic cytochrome P-450 and attenuates the hypotensive effect of nicardipine in rats. *Life Sci* 2002;70:2783–2792.

173. Ginkgo. In LaGow B, Ed. *PDR for Herbal Medicine*. Montvale, NJ: Thomson PDR; 2004, pp. 368–378.

174. Gurley BJ, Gardner SF, Hubbard MA, Williams DK, Gentry WB, Cui Y, Ang CY. Cytochrome P450 phenotypic ratios for predicting herb-drug interactions in humans. *Clin Pharmacol Ther* 2002;72:276–287.

175. Sparreboom A, Cox MC, Acharya MR, Figg WD. Herbal remedies in the United States: potential adverse interactions with anticancer agents. *J Clin Oncol* 2004;22:2489–2503.

176. Yamazaki H, Shimada T. Human liver cytochrome P450 enzymes involved in the 7-hydroxylation of *R*- and *S*-warfarin enantiomers. *Biochem Pharmacol* 1997;54:1195–1203.

177. Engelsen J, Nielsen JD, Winther K. Effect of coenzyme Q10 and *Ginkgo biloba* on warfarin dosage in stable, long-term warfarin treated outpatients. A randomised, double blind, placebo-crossover trial. *Thromb Haemost* 2002;87:1075–1076.

178. Cotreau MM, von Moltke LL, Greenblatt DJ. The influence of age and sex on the clearance of cytochrome P450 3A substrates. *Clin Pharmacokinet* 2005;44:33–60.

179. Gurley BJ, Gardner SF, Hubbard MA, Williams DK, Gentry WB, Cui Y, Ang CY. Clinical assessment of effects of botanical supplementation on cytochrome P450 phenotypes in the elderly: St John's wort, garlic oil, Panax ginseng and *Ginkgo biloba*. *Drugs Agin* 2005;22:525–539.

180. Markowitz JS, Donovan JL, DeVane CL, Taylor RM, Ruan Y, Wang JS, Chavin KD. Effect of St John's wort on drug metabolism by induction of cytochrome P450 3A4 enzyme. *JAMA* 2003;290:1500–1504.

181. Jöhne A, Brockmöller J, Bauer S, Maurer A, Langheinrich M, Roots I. Pharmacokinetic interaction of digoxin with an herbal extract from St. John's wort (*Hypericum perforatum*). *Clin Pharmacol Ther* 1999;66:338–345.

182. Sugimoto K, Ohmori M, Tsuruoka S, Nishiki K, Kawaguchi A, Harada K, Arakawa M, Sakamoto K, Masada M, Miyamori I, Fujimura A. Different effects of St. John's wort on the pharmacokinetics of simvastatin and pravastatin. *Clin Pharmacol Ther* 2001;70:518–524.

183. Jöhne A, Schmider J, Brockmöller J, Stadelmann AM, Störmer E, Bauer S, Scholler G, Langheinrich M, Roots I. Decreased plasma levels of amitriptyline and its metabolites on comedication with an extract from St. John's wort (*Hypericum perforatum*). *J Clin Psychopharmacol* 2002;22:46–54.

184. Dannawi M. Possible serotonin syndrome after combination of buspirone and St John's wort. *J Psychopharmacol* 2002;16:401.
185. Wang Z, Gorski JC, Hamman MA, Huang SM, Lesko LJ, Hall SD. The effects of St John's wort (*Hypericum perforatum*) on human cytochrome P450 activity. *Clin Pharmacol Ther* 2001;70:317–326.
186. Piscitelli SC, Burstein AH, Chaitt D, Alfaro RM, Falloon J. Indinavir concentrations and St John's wort. *Lancet* 2000;355:547–548.
187. de Maat MM, Hoetelmans RM, Matht RA, van Gorp EC, Meenhorst PL, Mulder JW, Beijnen JH. Drug interaction between St John's wort and nevirapine. *AIDS* 2001;15:420–421.
188. Gorski JC, Huang SM, Pinto A, Hamman MA, Hilligoss JK, Zaheer NA, Desai M, Miller M, Hall SD. The effect of Echinacea (*Echinacea purpurea* root) on cytochrome P450 activity *in vivo*. *Clin Pharmacol Ther* 2004;75:89–100.
189. Blumenthal M. Herbs continue slide in mainstream market: sales down 14 percent. *HerbalGram* 2003;58:71.
190. Markowitz JS, DeVane CL, Chavin KD, Taylor RM, Ruan Y, Donovan JL. Effects of garlic (*Allium sativum* L.) supplementation on cytochrome P450 2D6 and 3A4 activity in healthy volunteers. *Clin Pharmacol Ther* 2003;74:170–177.
191. Gallicano K, Foster B, Choudhri S. Effect of short-term administration of garlic supplements on single-dose ritonavir pharmacokinetics in healthy volunteers. *Br J Clin Pharmacol* 2003;55:199–202.
192. Piscitelli SC, Burstein AH, Welden N, Gallicano KD, Falloon J. The effect of garlic supplements on the pharmacokinetics of saquinavir. *Clin Infect Dis* 2002;34:234–238.
193. Gwilt PR, Lear CL, Tempero MA, Birt DD, Grandjean AC, Ruddon RW, Nagel DL. The effect of garlic extract on human metabolism of acetaminophen. *Cancer Epidemiol Biomarkers Prev* 1994;3:155–160.
194. Hu Z, Yang X, Ho PC, Chan SY, Heng PW, Chan E, Duan W, Koh HL, Zhou S. Herb–drug interactions: a literature review. *Drug* 2005;65:1239–1282.
195. Evans V. Herbs and the brain: friend or foe? The effects of ginkgo and garlic on warfarin use. *J Neurosci Nurs* 2000;32:229–232.
196. Matthews MK Jr. Association of *Ginkgo biloba* with intracerebral hemorrhage. *Neurology* 1998;50:1933–1934.
197. Meisel C, Jöhne A, Roots I. Fatal intracerebral mass bleeding associated with *Ginkgo biloba* and ibuprofen. *Atherosclerosis* 2003;167:367.
198. Hoffman T. Ginkgo, Vioxx and excessive bleeding—possible drug–herb interactions: case report. *Hawaii Med J* 2001;60:290.
199. Jiang X, Williams KM, Liauw WS, Ammit AJ, Roufogalis BD, Duke CC, Day RO, McLachlan AJ. Effect of ginkgo and ginger on the pharmacokinetics and pharmacodynamics of warfarin in healthy subjects. *Br J Clin Pharmacol* 2005;59:425–432.

# SPECIES COMPARISON OF METABOLISM IN MICROSOMES AND HEPATOCYTES

NIELS KREBSFAENGER
*Schwarz Biosciences, Monheim, Germany*

**Contents**

1 Introduction and General Aspects
  1.1 Relevance of Liver Preparations for Metabolism Studies *in Vitro*
  1.2 Liver Microsomes Versus Hepatocytes
  1.3 Toxicity Testing in Primary Hepatocyte Metabolism Studies
  1.4 Species Characteristics and Differences in Metabolism
  1.5 Regulatory and Strategic Aspects
2 Materials, Methods, and Technical Aspects
  2.1 Preparation and Characterization of Liver Microsomes and Hepatocytes
  2.2 Experimental Study Design
  2.3 Commercial Suppliers and CROs
3 Conclusion
  References

## 1 INTRODUCTION AND GENERAL ASPECTS

Species differences in metabolism may have significant impact on pharmacokinetics and toxicity of drugs. Therefore detailed knowledge of comparative metabolism across species is key for species selection in preclinical safety testing and interpretation of any animal data in respect to relevance for humans.

From the safety perspective, the primary concern is to identify any unique or major human metabolites. Early identification will allow for timely assessment of

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

potential safety issues in the development process and thereby reduce the risk for subjects included in clinical trials as well as for possible delays in the development process.

Since in early drug development animal and human *in vivo* metabolism data are usually lacking, *in vitro* species comparisons provide useful information to support a preliminary species selection for preclinical safety testing and interpretation of animal data.

### 1.1 Relevance of Liver Preparations for Metabolism Studies *In Vitro*

The liver is the major organ for biotransformation of xenobiotics including drugs. The majority of xenobiotic metabolizing enzymes (XMEs) are expressed in the liver and usually their abundance is higher than in other organs or tissues. Several standard methods have been established to investigate hepatic metabolism *in vitro* and there is comprehensive knowledge of their potential and limitations. Therefore, liver preparations such as tissue slices, hepatocytes, or microsomes are valuable and accepted tools for comparative studies on drug metabolism across species *in vitro* (see Section 1.5). As orally administered drugs enter first the liver via the portal vein before being distributed in the systemic circulation, and metabolism may have substantial impact on the toxicity of xenobiotics (see Chapter 28 in this volume), hepatotoxicity is a general concern in drug development (see Section 1.3). Therefore, toxicity testing in hepatocytes along with metabolism studies may provide additional useful information in early drug development.

However, it should be kept in mind that the liver is not the only metabolically competent organ. For example, cytochrome P450 (CYP) 3A4 is also expressed at high levels in the gut and may substantially contribute to metabolism of drugs even before they enter the liver. CYP1A1 expression is significantly induced in the lung by smoking but there is no such enzymatic activity in the liver, and several prodrugs are cleaved by plasma esterases. Accordingly, *in vitro* metabolism data based on liver preparations alone are not always fully predictive for the situation *in vivo*. Therefore, additional testing using, for example, pulmonary, intestinal, or kidney (microsomal) preparations may be appropriate in specific cases.

### 1.2 Liver Microsomes Versus Hepatocytes

Liver microsomes and primary hepatocytes are the most common *in vitro* systems to assess species differences in metabolism *in vitro*. However, they differ in many respects (Table 1).

Microsomes are subcellular membrane fractions obtained by differential centrifugation (see Section 2.1), which contain primarily the membrane-bound (smooth endoplasmic reticulum) phase I XMEs (CYPs). Liver microsomes of various species are commercially available (see Section 2.3) and are standard for screening of metabolic stability, CYP profiling, and inhibition (see Chapter 21 in this volume). Besides CYP-dependent metabolism, flavin-containing monooxygenase (FMO) and uridine diphosphoglucuronosyltransferase (UGT) metabolism can also be investigated in microsomes if the appropriate cofactors are added to the incubation (i.e., $FADH_2$ or UDP-glucuronic acid, respectively). Epoxide hydrolase

**TABLE 1 Comparative Characteristics and Properties of Liver Microsomes and Primary Hepatocytes**

| | | Primary Hepatocytes | |
|---|---|---|---|
| Characteristic | Liver Microsomes | Cryopreserved | Freshly Isolated |
| **Metabolic competence** | Phase I (CYPs, EH)<br>No transporters<br>High enzyme activities<br>FMOs and UGTs (if supplemented with specific cofactors)<br>Carboxylesterases (depending on preparation method) | Phases I and II<br>Transporters<br>Low enzyme activities | Phases I and II<br>Transporters<br>Low to intermediate enzyme activities |
| **Cofactors** | To be supplemented | Intrinsic | Intrinsic |
| **Batch characterization** | Usually precharacterized (enzyme activities, protein content, genotyping) | Usually precharacterized (enzyme activities, inducibility, viability, genotyping) | Limited precharacterization (characterization usually in parallel to experiment) |
| **Incubation time (type)** | Up to 1 hour (subcellular suspension) | Up to 2–4 hours (suspension)<br>Days—weeks (plated[a]) | Up to 2–4 hours (suspension)<br>Days—weeks (plated) |
| **Toxicity testing** | Not possible | Limited | Various parameters possible |
| **Species available**[b] | Many | Many | Some |
| **Availability** | Excellent | Good | Unpredictable |
| **Costs** | Less expensive | Expensive | Most expensive |

[a]Plating of cryopreserved hepatocytes is not possible with all species.
[b]See Table 6.

(EH) and—depending on the preparation method—carboxylesterase activity may also be maintained. Enzyme activities are comparably high and compensate for the limited incubation time due to enzyme degradation over time.

The major limitation of liver microsomes is the lack of phase II metabolic competence. Predictions for compounds that are readily conjugated may therefore be misleading. For example, compounds with free amino, carboxy, or hydroxy functions may be acetylated, glucuronidated, or sulfated without prior functionalization. Microsomal incubations may therefore overestimate the metabolic stability of such compounds. Similarly, the comparison of metabolic profiles over species is evidently limited if phase II metabolism is involved in biotransformation.

Primary hepatocytes overcome the obstacle of lacking phase II metabolism and provide the advantages of a "living" cellular test system such as intrinsic formation of cofactors, regeneration of XMEs (though not constant over time), and the

possibility to implement the assessment of liver function and toxicological parameters (see Section 1.3). Hepatocytes allow also for investigation of XME induction. In addition, test substance concentrations at the enzyme are expected to be closer to the physiological situation compared to noncellular microsomal incubations due to the preserved natural orientation for linked enzymes and intact cellular and subcellular compartmentation as well as presence of transporters (see Chapter 12 in this volume). Overall, primary hepatocytes provide a much more physiological test situation. Freshly isolated primary hepatocytes and some batches of plateable cryopreserved hepatocytes can be cultured for up to weeks ahead, maintaining hepatocyte properties and forming organ-like structures (e.g., polarity, junctional complexes, bile canaliculi, and glycogen particles) [1, 2], although nonparenchymal cells are usually underrepresented.

The major limitations of primary hepatocytes are the high interbatch variability and lower enzyme activities, which may further decrease significantly during prolonged incubation. Also, primary hepatocytes are considerably more expensive than liver microsomes. Another drawback with freshly isolated primary hepatocytes as compared to cryopreserved hepatocytes and microsomes is the limited availability and lack of precharacterization; that is, the batch characteristics usually are only determined in parallel with the actual experiment.

### 1.3 Toxicity Testing in Primary Hepatocyte Metabolism Studies

During the last decade liver toxicity has been one of the most frequent reasons for pharmacovigilance safety reports and the withdrawal from the market of an approved medicinal product [3].

Accumulation of the parent drug itself or its metabolites in the liver (see Chapter 13 in this volume) and formation of reactive metabolites[1] are potential cause for concern. Even at therapeutic dose levels, drugs may affect expression of XMEs (induction or inhibition) and increase production of reactive metabolites. Therefore, the use of metabolically competent and potentially inducible hepatocytes is recommended as an *in vitro* tool for the early detection and mechanistic investigation of potential hepatotoxicity [3]. The concomitant assessment of liver function and toxicological parameters within comparative metabolism studies in hepatocytes offers the chance to link potential hepatotoxicity with critical metabolite(s) and identify possible species differences early on. This may help to interpret *in vivo* findings and their relevance for humans (Fig. 1). However, it should be stressed that cultured hepatocytes clearly do not represent the complex *in vivo* situation and important mechanisms of liver toxicity such as those mediated via an immunological cascade or idiosyncratic reactions are not represented.

Many of the standard clinical chemistry parameters for liver injury measured in plasma samples in *in vivo* studies can also be assessed in the supernatant of hepatocyte cultures *in vitro* (Table 2). *In vivo*, increases in the levels of alanine aminotransferase (ALT) and aspartate aminotransferase (AST) in serum, in combination with increased bilirubin levels, are actually considered to be the most

[1]Potential structural alerts for hepatotoxicity are furans, quinones, epoxides, thiophenes, carboxylic acids, phenoxyl radicals, phenols, acyl halides, acyl glucuronides, aniline radicals, and aromatic and hydroxy amines.

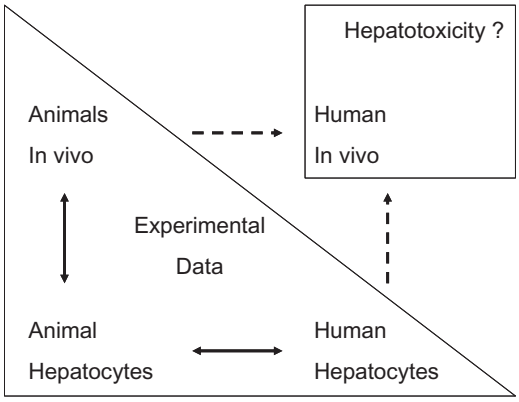


**FIGURE 1** Prediction of human hepatotoxicity. Combining human *in vitro*, animal *in vitro*, and animal *in vivo* data improves the prediction of human hepatotoxicity *in vivo*. Human hepatocytes can provide the "missing link" for species extrapolation while animal *in vivo* data "test" the relevance of the respective *in vitro* data.

**TABLE 2 Examples for Liver Function and Toxicological Parameters in Hepatocytes**

| Parameter | Correlation/Function | Assessment |
|---|---|---|
| **Cell morphology and general appearance** | General cell viability, necrotic or apoptotic cells, extra- or intracellular precipitates, etc. | Microscopical "in-life" examination of cells in culture well |
| **Albumin secretion** | Synthesis competence | Clinical chemistry of supernatant |
| **ALT** | Hepatocellular damage | Clinical chemistry of supernatant |
| **AST** | Hepatocellular damage | Clinical chemistry of supernatant |
| **LDH** | Hepatocellular damage | Clinical chemistry of supernatant |
| **γ-GT** | Hepatobiliary damage | Clinical chemistry of supernatant |
| **aP** | Hepatobiliary damage | Clinical chemistry of supernatant |
| **GLDH** | Mitochondrial damage | Clinical chemistry of supernatant |
| **Lactate** | Mitochondrial damage | Clinical chemistry of supernatant |
| **ATP** | Mitochondrial damage | Terminal intracellular detection |
| **GSH** | Redox status | Terminal intracellular detection |
| **Trypan blue exclusion** | Membrane integrity | Terminal staining of cells in well |
| **Neutral red uptake** | Membrane integrity | Terminal staining of cells in well |
| **MTT reduction** | Mitochondrial activity | Terminal staining of cells in well |
| **XTT reduction** | Mitochondrial activity | Terminal staining of cells in well |
| **Sulforhodamine B staining** | Protein content | Terminal staining of cells in well |
| **Hoechst 33258 staining** | DNA content | Terminal sampling of cells/DNA |
| **Expression profiling** | mRNA (or protein) levels | Terminal sampling of cells/mRNA |

relevant signal of liver toxicity, and an increase of ALT activity in the range of two- to fourfold and higher compared to concurrent control average or individual pre-treatment values should raise concern as an indication of potential hepatic injury [3]. ALT is considered a more specific and sensitive indicator of hepatocellular injury than AST.

**TABLE 3 General Rules on Interspecies Differences in DMPK**

| Species | DMPK Characteristics |
|---|---|
| **Human** | Polymorphisms (e.g., CYP2C9, CYP2C19, CYP2D6, NAT1, NAT2) |
| **Dog** | Low acetylation, high capacity for deacetylation; different absorption due to higher pH in gastrointestinal tract than in humans (consider use of synthetic gastric fluid to mimic human situation) |
| **Rat** | Often gender differences that are not observed in other species; abundant tetrahydrofolate (protects, e.g., against methanol ocular damage) |
| **Rabbit** | Low sulfation |
| **(Mini)Pig** | Low sulfation; gastrointestinal conditions similar to humans |
| **Cat** | Low glucuronidation; high sulfation |

In addition, in cultured hepatocytes conventional cytotoxicity staining or expression profiling can be terminally performed (Table 2). Toxicogenomic data may support the elucidation of hepatotoxic mechanisms or interspecies differences and the development of appropriate biomarkers for liver toxicity.

### 1.4 Species Characteristics and Differences in Metabolism

As pointed out earlier, the knowledge of qualitative and quantitative interspecies differences in metabolism is key in species selection for preclinical safety studies and interpretation of any animal data.

Although there are some "general rules" on species differences in drug metabolism and pharmacokinetics (DMPK) (Table 3) like a decreasing rate of metabolism with an increasing size of organism, that is, mouse > rat > dog > human (see Chapter 29 in this volume for allometric scaling), which is also reflected *in vitro*, experimental testing of individual compounds is *conditio sine qua non* as there are also numerous more subtle but often decisive distinctions and the metabolic fate of a new chemical entity (NCE) cannot be reliably predicted today. The complexity of interspecies differences in the XME pattern is probably best exemplified by the many species differences in CYP isoforms [4]. Similarly, several species differences in isoforms have been identified for the major phase II XMEs: UGTs [5], sulfotransferases (SULTs) [6], *N*-acetyltransferases (NATs) [7], and glutathione *S*-transferases (GSTs) [8]. Information on typical phase I and II XME activities in liver preparations of different species is available on the websites of commercial vendors (Table 4) or, for example, in Refs. 9 and 10 (see Section 2.1).

For example, species comparison of 7-ethoxycoumarin (7-EC) phase I and II metabolism *in vitro*, which is typically used for (pre)characterization of human and animal hepatocytes (Table 5), suggests that human metabolism is best represented in cynomolgus monkeys with respect to the metabolite profile (glucuronidated metabolite (7-HCG) > sulfated metabolite (7-HCS) > O-deethylated metabolite (7-HC)); however, rats would be the more appropriate model with respect to the rate of metabolism (monkey ≈ rabbit ≈ dog > rat > human) [10].

### 1.5 Regulatory and Strategic Aspects

The importance of species comparisons in metabolism for safety assessment of NCEs and their metabolite(s) during drug development is reflected in several

**TABLE 4 Examples of Commercial Sources for Liver Preparations and/or Specialized CROs for *In Vitro* Metabolism Studies (Status August 2006)**

| CRO/Vendor | Website @ www. | Comment[a] |
|---|---|---|
| BD Biosciences | bdbiosciences.com (gentest.com) | M, cH, fH, Services |
| GenPharmTox BioTech AG | genpharmtox.com | Services |
| In Vitro Technologies, Inc. | invitrotech.com | M, cH, fH, Services |
| tebu-bio | tebu-bio.com | M, cH, Services |
| XenoTech, LLC | xenotechllc.com | M, cH, fH, Services |

[a]M, microsomes; cH, cryopreserved hepatocytes; fH, freshly isolated hepatocytes.

**TABLE 5 Typical Marker Reactions for (Pre)Characterization of Human and Animal Liver Microsomes and Hepatocytes**

| XME | Marker Reaction |
|---|---|
| | *Phase I* |
| CYP(s) | 7-Ethoxycoumarin O-deethylation (formation of 7-hydroxycoumarin) |
| Human isoenzyme | |
| CYP1A2 | Phenacetin O-deethylation |
| CYP2A6 | Coumarin 7-hydroxylation |
| CYP2B6 | *S*-mephenytoin N-demethylation |
| CYP2C8 | Taxol 6-hydroxylation |
| CYP2C9 | Diclofenac 4′-hydroxylation |
| CYP2C19 | *S*-mephenytoin 4′-hydroxylation |
| CYP2D6 | Bufuralol 1′-hydroxylation |
| CYP2E1 | Chlorzoxazone 6-hydroxylation |
| CYP3A4/5 | Testosterone 6β-hydroxylation |
| | *Phase II* |
| UGT(s) | 7-Hydroxycoumarin glucuronidation |
| SULT(s) | 7-Hydroxycoumarin sulfation |
| GST(s) | 1-Chloro-2,4-dinitro-benzene glutathione conjugation |
| NAT1 (human) | *p*-Aminobenzoic acid N-acetylation |
| NAT2 (human) | Sulfamethazine N-acetylation |

regulatory guidance documents (status August 2006; see Chapter 37 in this volume):

- ICH M3, 1997: *Guidance for Industry on Nonclinical Safety Studies for the Conduct of Human Clinical Trials for Pharmaceuticals* [11].
- EMEA, 2000: *Note for Guidance on Repeated Dose Toxicity* [12].
- FDA/CDER, 2005: *Draft Guidance for Industry on Safety Testing of Drug Metabolites* [13].
- FDA/CDER, 1997: *Guidance for Industry: Drug Metabolism/Drug Interaction Studies in the Drug Development Process: Studies In Vitro* [14].

According to ICH M3, "exposure data in animals should be evaluated prior to human clinical trials. Further information on ADME[2] in animals should be made available to compare human and animal metabolic pathways. Appropriate information should usually be available by the time the Phase I (Human Pharmacology) studies have been completed."

The EMEA *Note for Guidance on Repeated Dose Toxicity* emphasizes the importance of comparative metabolism data for species selection in preclinical safety studies: "Within the usual spectrum of laboratory animals used for toxicity testing, the species should be chosen based on their similarity to humans with regard to pharmacokinetic profile including biotransformation. Exposure to the main human metabolite(s)[3] should be ensured. If this can not be achieved in toxicity studies with the parent compound, specific studies with the metabolite(s) should be considered."

And the FDA/CDER *Draft Guidance for Industry on Safety Testing of Drug Metabolites* encourages that "attempts be made to identify as early as possible during the drug development process differences in drug metabolism in animals used in nonclinical safety assessments compared to humans. It is especially important to identify metabolites that may be unique to humans."

Although the ultimate confirmation and justification on species selection can only be made after animal and human *in vivo* metabolite profiles are available (see Chapters 23 and 27 in this volume), this is, not before Clinical Phase I, *in vitro* species comparisons including human material provide an important—and the only—possibility to rationalize the species selection for pre-IND safety studies; usually on genotoxicity and repeated dose toxicity in one rodent and one nonrodent animal species [11] (see Chapter 37 in this volume). They further offer the possibility to compare human and several standard laboratory and potentially also nonstandard animal species in a relatively efficient approach. In addition, though obviously preferable, in contrast to *in vivo* studies, radiolabeled test substance is often not necessarily required to obtain at least qualitatively comparable metabolite profiles and "identify" the major metabolite(s) if state-of-the-art analytical equipment, usually HPLC/MS,[4] is available (see Chapter 27 in this volume).

While the early identification of unique or major metabolites will allow for timely assessment of potential safety issues, the discovery of unique or major human metabolites late in drug development can cause development delays and could have possible implications for marketing approval.

Accordingly, sponsors are encouraged by regulators "to conduct *in vitro* studies to identify and characterize unique human or major metabolites early in drug development" [13]. Liver preparations such as microsomes, hepatocytes, or tissue slices are considered appropriate for this purpose. This approach is also regarded as the "best practice" in the U.S. pharmaceutical industry [15, 16].

*In vitro* metabolism studies are not classical safety studies and regulatory agencies in the Unites States, Europe, and Japan therefore do not require compliance

[2]ADME: absorption, distribution, metabolism, excretion.

[3]The FDA/CDER *Draft Guidance for Industry on Safety Testing of Drug Metabolites* defines major metabolites primarily as those identified in human plasma that account for greater than 10% of drug-related material (administered dose or systemic exposure, whichever is less).

[4]HPLC, high performance liquid chromatography; MS, mass spectrometry.

with "Good Laboratory Practice" (GLP) regulations. However, testing under GLP may be considered since this may add to quality and improved processing and documentation, especially if studies are outsourced to contract research organizations (CROs), and there appears to be increasing interest in GLP compliance by regulators also for safety-related studies.

## 2 MATERIALS, METHODS, AND TECHNICAL ASPECTS

### 2.1 Preparation and Characterization of Liver Microsomes and Hepatocytes

***Preparation of Liver Microsomes*** The generalized procedure for preparation of (liver) microsomes by differential centrifugation is shown in Fig. 2. Tissue samples are homogenized and centrifuged at a lower force (e.g., 9000 *g*) to form a premicrosomal pellet containing cell debris, nuclei, peroxisomes, lysosomes, and mitochondria. The resulting postmitochondrial supernatant is subsequently centrifuged at a higher force (e.g., 138,000 *g*) to precipitate the microsomes. The cytosolic supernatant may directly be used in metabolism studies; the microsomal pellet is resuspended in buffer and is ready for direct use or storage below –70 °C until use.

A more detailed description of the procedure is given in Ref. 17. The assessment of different protocols [18–24] shows that there are a number of methodological variables such as the number of strokes used to homogenize the liver samples (four to eight), the centrifugation parameters (9000 *g* for 20 min to 18,000 *g* for 10 min in the first step and 100,000–143,000 *g* for 60–90 min in the second step), compositions of the homogenization and final suspension buffers (EDTA, potassium chloride,

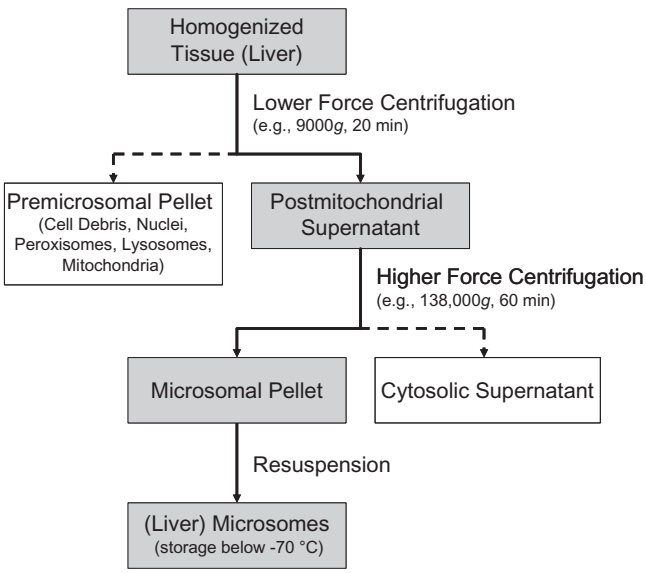


**FIGURE 2** Preparation of (liver) microsomes. The generalized differential centrifugation procedure for preparation of microsomes is shown. For more detailed information (e.g., on tissue homogenization, buffer composition, repetition of steps), please refer to the references as given in the text.

glycerol, sucrose), and repetition of steps for more thorough extraction. Relative volumes, concentrations, and dilutions of the samples also differ. These variables may significantly affect the recovery and specific activities of microsomal components but do not appear to affect enzyme kinetics [17].

***Preparation of Freshly Isolated Hepatocytes*** There are two principal methods for isolation of primary hepatocytes: (1) *in situ* liver perfusion in anesthetized animals and (2) perfusion of human or animal liver samples after liver resection.

For the first method, the animal is anesthetized and the liver is perfused *in situ* via the vena portae with EGTA[5]/washing buffer to remove blood and prevent clotting of the capillaries. Subsequently, perfusion is continued with collagenase buffer to dissociate the cells from their matrix. After perfusion the liver is removed from the animal, the liver capsula is removed, and the cells are dissociated carefully in suspension buffer. The liver cell suspension is filtered and the filtrate is gently centrifuged for sedimentation of the hepatocytes. The cell pellet is washed in suspension buffer, centrifuged again, and resuspended in suspension buffer.

For the second method, the resected liver tissue is immediately transferred into ice-cold suspension buffer for storage and transport up to some hours until further processing. Liver samples are perfused with EGTA/washing buffer to remove blood and prevent clotting of the capillaries. Perfusion is performed by (several) blunt-end cannulae inserted into vessels of the cut surface. Thereafter, perfusion is continued with collagenase buffer to dissociate the cells from their matrix. The tissue is transferred into a large Petri dish with suspension buffer and the liver cells are gently scraped out with a spatula. The liver cell suspension is filtered and the filtrate is gently centrifuged for sedimentation of the hepatocytes. The cell pellet is washed in suspension buffer, centrifuged again, and resuspended in suspension buffer.

Detailed procedures for the isolation of primary rat and human hepatocytes by these methods are described in Ref. 9. Like with the preparation of microsomes, multiple variations of the general procedures have been described [25–31]. In principle, both methods are suitable for different species such as mouse, rat, rabbit, dog, minipig, or monkey. However, applicability may be limited by size of the organ or tissue sample or due to practical limitations; for example, human and sometimes animal livers may not be available for perfusion *in situ* but only after resection.

The freshly isolated primary hepatocytes may be used directly in suspension [9, 32], culture plated on collagen-coated wells (Fig. 3) [1, 26, 27, 29, 33], or cryopreserved and precharacterized for later use [9, 32, 34, 35].

***(Pre)characterization of Liver Preparations*** Microsomes and cryopreserved hepatocytes are usually precharacterized for XME activities and total protein and CYP content or cell viability and inducibility, respectively. Liver preparations of individual human donors are usually supplemented with information on the respective donor, such as age, gender, race, smoking habits, and premedication, and sometimes are also genotyped for polymorphic XMEs such as CYP2C9, CYP2C19, CYP2D6, NAT1, or NAT2. This enables one to select batches of preferred characteristics for the intended type of study and compound under investigation (Section 2.3).

[5]EGTA, ethylene glycol bis(β-aminoethyl-ether)-*N*,*N*,*N*′,*N*′-tetraacetic acid.

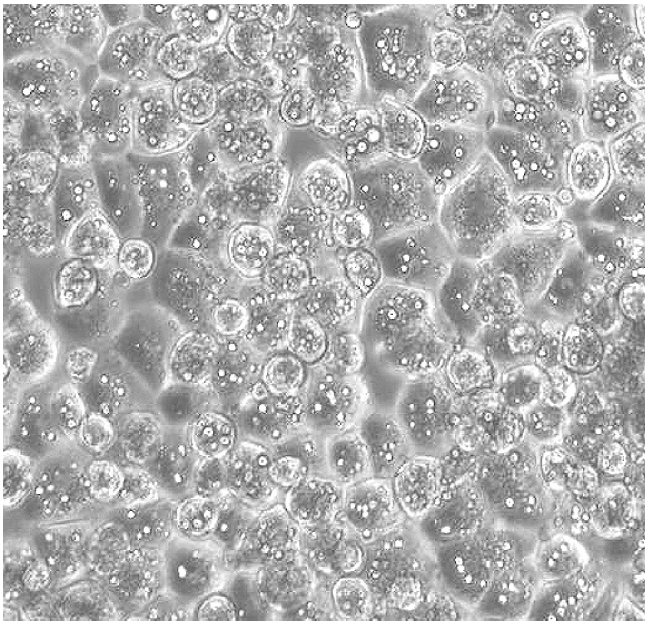

**FIGURE 3** Light microscopical picture of plated primary human hepatocytes in culture. (Courtesy of GenPharmTox BioTech AG, Martinsried, Germany.)

If freshly isolated primary hepatocytes are directly employed after isolation, precharacterization is usually limited to cell viability and donor information but further indepth characterization (e.g., for enzyme activities) may only be performed concomitant to the actual experiment.

Typical marker reactions for (pre)characterization of human and animal liver microsomes and hepatocytes are listed in Table 5. A more comprehensive list of preferred and acceptable substrates, inhibitors, and inducers for *in vitro* experiments and the respective concentration ranges is given on the FDA website *Drug Development and Drug Interactions: Table of Substrates, Inhibitors and Inducers* at http://www.fda.gov/cder/drug/drugInteractions/tableSubstrates.htm#major. Information on typical marker substrate activities for phase I and II XMEs in liver preparations of different species are available on the websites of commercial vendors (Table 4) or, for example, in Refs. 9 and 10.

## 2.2 Experimental Study Design

In the following, the principal experimental study designs for investigating metabolic stabilities and metabolite profiles of NCEs in liver microsomes, hepatocyte suspensions, and cultured hepatocytes are described. Depending on the test item properties and intended purpose of the study (e.g., for screening or a development project), manifold variations may be advisable, which cannot be discussed in detail here. However, some general points to consider will be highlighted later in the subsection "Points to Consider."

***Metabolic Stability and Metabolite Profile in Liver Microsomes*** An aliquot of test item stock solution(s) is added to the prewarmed reaction mix(es) including liver microsomes of the desired species, incubation buffer, and NADPH or a NADPH regenerating system and is incubated at 37 °C for different time intervals (usually 10–60 min). The reaction is terminated by addition of organic solvent (e.g., ice-chilled acetonitrile or methanol) and kept cool during sample preparation. In many cases, after centrifugation of the precipitated microsomal components, the supernatant can be directly submitted to HPLC-UV/VIS,[6] -FLD, -RD, and/or -MS analysis.

The loss of parent compound and/or formation of metabolite(s) is calculated compared to a zero time point control, a control without test item, and/or a control without NADPH to correct for chemical degradation, non-CYP-dependent metabolism, and artifact peaks. From this the rate of metabolism (i.e., metabolic half-life, intrinsic clearance, or simply percentage of test compound metabolized/remaining after a certain time) and/or qualitative information on the metabolite profile (i.e., minor, intermediate, major abundance of phase I metabolites) is derived.

***Metabolic Stability and Metabolite Profile in Hepatocyte Suspensions*** An aliquot of test item stock solution(s) is added to the prewarmed reaction mix(es) including hepatocytes of the desired species and incubation buffer and is incubated at 37 °C under slight agitation (e.g., at 40 rpm in a shaking water bath) for different time intervals (usually 15–120 min, maximum 4 hours). The reaction is terminated by addition of organic solvent (e.g., ice-chilled acetonitrile or methanol) and kept cool during sample preparation. In many cases, after centrifugation of the precipitated hepatocytes and other components, the supernatant can be directly submitted to analysis. The loss of parent compound and/or formation of metabolite(s) is calculated compared to a zero time point control and/or a control with heat-inactivated hepatocytes to correct for chemical degradation and artifact peaks. From this the rate of metabolism and/or qualitative information on the phase I and II metabolite profile is derived.

***Metabolic Stability and Metabolite Profile in Cultured Hepatocytes*** An aliquot of test item stock solution(s) is added to the preincubated hepatocyte culture (attachment phase of at least 4 hours, for long-term cultures often overnight) including hepatocytes of the desired species and incubation buffer and is incubated at 37 °C and 5% $CO_2$ for different time intervals (usually 2–24 hours, maximum several days). The reaction is terminated by addition of organic solvent (e.g., ice-chilled acetonitrile or methanol) and/or removal of the supernatant and kept cool during sample preparation. In many cases, after centrifugation of the precipitated hepatocytes and other components, the supernatant can be directly submitted to analysis.

The loss of parent compound and/or formation of metabolite(s) is calculated compared to a zero time point control and/or a control with heat-inactivated hepatocytes to correct for chemical degradation and artifact peaks. From this the rate of metabolism and/or qualitative information on the metabolite profile is derived.

As discussed in Section 1.3, it should be noted that cultured hepatocytes in addition allow for the concomitant assessment of several liver function and

[6]HPLC, high performance liquid chromatography; UV/VIS, ultraviolet/visible light detector; FLD, fluorescence detector; RD, radiometric detection; MS, mass spectrometry.

toxicological parameters, thereby offering the chance to link potential hepatotoxicity with critical metabolite(s) and identify possible species differences early in drug development.

***Points to Consider*** Low solubility of the test item in aqueous buffer at physiological pH is often a challenge in *in vitro* drug metabolism studies. Precipitation in the incubation mix is difficult to note due to the turbidity of the suspension; light microscopy with cultured hepatocytes may sometimes reveal precipitates, but in general solubility may already be limited before any precipitate is visible. The situation is even complicated by potentially solubilizing effects of lipids and proteins within the incubation.

Usually, stock solutions of the test item in organic solvent such as dimethyl sulfoxide, acetonitrile, ethanol, or methanol are prepared to enhance solubility for the serial dilution as well as within the actual incubation. If the high concentration stock solution was prepared in solvent, serial dilutions should also be prepared with solvent—not medium—to assure correct concentrations of the lower concentration stock solutions (i.e., to prevent precipitation during the serial dilution). Usually, final concentrations of up to 1% organic solvent are employed and tolerated by (sub)cellular test systems. However, it should be noted that different solvents interact with different XMEs even at considerably lower concentrations [36]. Another aspect linked with test item solubility is the recovery rate, which, if possible, may be worthwhile examining for proper interpretation of *in vitro* metabolism data.

Test item concentrations generally vary between 0.1 and 100 μM. In most cases, in early development the actual concentration in plasma or liver *in vivo* is not known but will usually be within this range. For determination of the rate of metabolism, lower concentrations of 0.1 or 1 μM are recommended in order not to overload the test system, while higher test item concentrations of 10 or 100 μM will facilitate detection of (minor) metabolites due to a higher absolute amount of metabolism. It should, however, be noted that metabolite profiles (*in vitro* and *in vivo*!) may differ for very high versus very low test item concentrations due to other sets of XMEs being preferred; for example, UGTs are low affinity but high capacity enzymes while SULTs show high affinity but low capacity. Appropriate test system concentrations, buffer composition, and detailed incubation/cultivation procedures may be derived from the commercial vendor's information (Section 2.3) or the referenced literature.

Incubation times are generally limited by loss of enzyme activity to about 1 or 2 hours for microsomes and hepatocyte suspensions, respectively. If minor metabolism is observed, incubation times may be extended (for up to 4 hours in suspension cultures) or cultured hepatocytes should be used, enabling incubation for up to several days. Chemical stability of the test item under the incubation conditions should also be considered.

For examination of metabolite profiles (see Chapter 27 in this volume), the use of a radiolabeled test item is obviously preferable as drug-related peaks are easier to differentiate from background and artifacts; metabolites can—at least approximately—easily be quantified and recovery can be assessed. However, metabolite profiles can also be derived from cold test substance if appropriate detection is possible, for example, by UV/VIS or FLD for absorbing compounds or by full-scan or targeted mass detection. In each case, it should be carefully considered which kind

of metabolites would be detected—or missed—with the respective analytical method given its limitations and the structure and properties of the test compound (see Chapter 27 in this volume).

### 2.3 Commercial Suppliers and CROs

There are a number of commercial sources for liver preparations and/or CROs for outsourcing *in vitro* metabolism studies (Table 4). Standard species for which microsomes and hepatocytes are available are listed in Table 6. Preparations from other species are sometimes also available as "custom preparations," however, usually at substantially higher costs and sometimes unacceptable timelines.

When purchasing liver preparations, besides precharacterization data, quality, price, and a well established shipment procedure should be considered. Especially when liver preparations of "critical" species such as minipig (agricultural species) or monkey (strict animal welfare legislation) are purchased from vendors abroad, customs may delay delivery and biomaterials may degrade if not maintained appropriately. Therefore, especially in the case of freshly isolated hepatocytes, although there are elaborate shipment procedures established nowadays, it may be preferable to choose a local supplier in order to ensure the best possible quality and prompt delivery.

While some milligrams of test substance per species should usually be sufficient, it is difficult to give even a general rule on costs and timelines for outsourcing studies on species comparisons/metabolite profiles due to the many different possible study designs (number of replicates, time points, controls, etc.) and analytical methods (cold versus radiolabeled test substance, metabolite profile "scan" versus metabolite identification, etc.).

**TABLE 6 Standard Species[a] for Which Microsomes and Hepatocytes Are Commercially Available at the Vendors Listed in Table 4 (Status August 2006)**

| Species | Strain | Liver Microsomes | Cryopreserved Hepatocytes | Freshly Isolated Hepatocytes |
|---|---|---|---|---|
| Human | Individual donor | + | + | + |
| | Pooled | + | + | |
| Monkey | Cynomolgus | + | + | + |
| | Rhesus | + | + | |
| | Marmoset | + | + | |
| Minipig | Goettingen | + | + | |
| | Yucatan | + | | |
| | Sinclair | + | | |
| Dog | Beagle | + | + | + |
| Rabbit | New Zealand White | + | + | |
| Guinea pig | Dunkin–Hartley | + | + | |
| Hamster | Golden Syrian | + | | |
| Rat | Sprague–Dawley | + | + | + |
| | Wistar Han | + | + | |
| | Fischer 344 | + | | |
| Mouse | CD-1 | + | + | + |
| | B6C3F1 | + | | |

[a]Additional species may be available as "custom preparations."

## 3 CONCLUSION

Species differences in metabolism may have significant impact on pharmacokinetics and toxicity of drugs. Therefore detailed knowledge of comparative metabolism across species is key for species selection in preclinical safety testing and interpretation of any animal data in respect to relevance for humans. Although there are some "general rules" on species differences in drug metabolism and pharmacokinetics, experimental testing of individual compounds is *conditio sine qua non*, as there are also numerous more subtle but often decisive distinctions and the metabolic fate of an NCE cannot be reliably predicted today.

Liver preparations such as tissue slices, hepatocytes, or microsomes are valuable and accepted tools for comparative studies on drug metabolism across species *in vitro*. While liver microsomes of various species are easily commercially available and are standard for screening of metabolic stability, CYP profiling, and inhibition, freshly isolated or cryopreserved hepatocytes in suspension represent a well established, commercially readily available, and easy-to-handle *in vitro* system that generally correctly predicts interspecies differences in phase I and II metabolism of NCEs. If longer incubation periods are required (i.e., if the rate of metabolism is low or minor metabolites are to be assessed), hepatocyte culture systems may be used. Such systems allow in addition for the concomitant assessment of liver function and toxicological parameters within comparative metabolism studies and thereby offer the chance to link potential hepatotoxicity with critical metabolite(s) and identify possible species differences early on. This may help to interpret *in vivo* findings and their relevance for humans.

Although all *in vitro* systems have their inherent limitations and the ultimate confirmation and justification on species selection can only be made after animal and human *in vivo* metabolite profiles are available, *in vitro* species comparisons including human material provide an important—and the only—possibility to rationalize the species selection for early *in vivo* safety studies, and they further offer the possibility to compare human and several standard laboratory and potentially also nonstandard animal species in a relatively efficient approach.

## REFERENCES

1. Koebe HG, Pahernik S, Eyer P, Schildberg FW. Collagen gel immobilization: a useful cell culture technique for long-term metabolic studies on human hepatocytes. *Xenobiotica* 1994;24(2):95–107.
2. Yamamoto N, Wu J, Zhang Y, Catana AM, Cai H, Strom S, Novikoff PM, Zern MA. An optimal culture condition maintains human hepatocyte phenotype after long-term culture. *Hepatol Res* 2006;35:169–177.
3. EMEA. *Guideline on Detection of Early Signal of Drug-Induced Hepatotoxicity in Non-Clinical Studies*. EMEA/CHMP/SWP/150115/2006.
4. Nelson DR, Koymans L, Kamataki T, Stegeman JJ, Feyereisen R, Waxman DJ, Waterman MR, Gotoh O, Coon MJ, Estabrook RW, Gunsalus IC, Nebert DW. P450 superfamily: update on new sequences, gene mapping, accession numbers and nomenclature. *Pharmacogenetics* 1996;6(1):1–42.

5. Mackenzie PI, Bock WK, Burchell B, Guillemette C, Ikushiro S, Iyanagi T, Miners JO, Owens IS, Nebert DW. Nomenclature update for the mammalian UDP glycosyltransferase (UGT) gene superfamily. *Pharmacogenetic Genomics* 2005;15(10):677–685.
6. Blanchard RL, Freimuth RR, Buck J, Weinshilboum RM, Coughtrie MW. A proposed nomenclature system for the cytosolic sulfotransferase (SULT) superfamily. *Pharmacogenetics* 2004;14(3):199–211.
7. Vatsis KP, Weber WW, Bell DA, Dupret J-M, Evans DA, Grant DM, Hein DW, Lin HJ, Meyer UA, Relling MV, Sim E, Suzuki T, Yamazoe Y. Nomenclature for *N*-acetyltransferases. *Pharmacogenetics* 1995;5(1):1–17.
8. Mannervik B, Board PG, Hayes JD, Listowsky I, Pearson WR. Nomenclature for mammalian soluble glutathione transferases. *Methods Enzymol* 2005;401:1–8.
9. Gebhardt R, Hengstler JG, Müller D, Glöckner R, Buenning P, Laube B, Schmelzer E, Ullrich M, Utesch D, Hewitt N, Ringel M, Reder-Hilz B, Bader A, Langsch A, Koose T, Burger H-J, Maas J, Oesch F. New hepatocyte *in vitro* systems for drug metabolism: metabolic capacity and recommendations for application in basic research and drug development, standard operation procedures. *Drug Metab Rev* 2003;35(2&3):145–213.
10. Li AP, Lu C, Brent JA, Pham C, Fackett A, Ruegg CE, Silber PM. Cryopreserved human hepatocytes: characterization of drug-metabolizing enzyme activities and applications in higher throughput screening assays for hepatotoxicity, metabolic stability, and drug-drug interaction potential. *Chem Biol Interact* 1999;121(1):17–35.
11. ICH M3. *Guidance for Industry on Nonclinical Safety Studies for the Conduct of Human Clinical Trials for Pharmaceuticals*; 1997.
12. EMEA. *Note for Guidance on Repeated Dose Toxicity*; 2000. EMEA/CPMP/SWP/1042/99 corr.
13. FDA/CDER. *Draft Guidance for Industry on Safety Testing of Drug Metabolites*; 2005.
14. FDA/CDER. *Guidance for Industry: Drug Metabolism/Drug Interaction Studies in the Drug Development Process: Studies In Vitro*; 1997.
15. Baillie TA, Cayen MN, Fouda H, Gerson RJ, Green JD, Grossman SJ, Klunk LJ, LeBlanc B, Perkins DG, Shipley LA. Drug metabolites in safety testing. *Toxicol Appl Pharmacol* 2002;182:188–196.
16. Hastings KL, El-Hage J, Jacobs A, Leighton J, Morse D, Osterberg R. Drug metabolites in safety testing. *Toxicol Appl Pharmacol* 2003;190(1):91–92.
17. Nelson AC, Huang W, Moody DE. Variables in human liver microsome preparation: impact on the kinetics of L-α-acetylmethadol (LAAM) N-demethylation and dextromethorphan O-demethylation. *Drug Metab Dispos* 2001;29(3):319–325.
18. Boobis AR, Brodie MJ, Kahn GC, Fletcher DR, Saunders JH, Davies DS. Monooxygenase activity of human liver in microsomal fractions of needle biopsy specimens. *Br J Clin Pharmacol* 1980;9:11–19.
19. Raucy J, Lasker JM. Isolation of P450 enzymes from human livers. *Methods Enzymol* 1991;206:577–594.
20. Kharasch ED, Thummel KE. Identification of cytochrome P450 2E1 as the predominant enzyme catalyzing human liver microsomal defluorination of sevoflurane, isoflurane, and methoxyflurane. *Anesthesiology* 1993;79:795–807.
21. Guengerich FP. In Hayes A, Ed. *Analysis and Characterization of Enzymes in Principles and Methods of Toxicology*. New York: Raven Press; 1994, pp 1259–1313.
22. Rodrigues AD, Kukulka MJ, Surber BW, Thomas SB, Uchic JT, Rotert GA, Michel G, Thome-Kromer B, Machinist JM. Measurement of liver microsomal cytochrome P450 (CYP2D6) activity using [*O*-methyl-$^{14}$C]dextromethorphan. *Anal Biochem* 1994;219: 309–320.

23. de Duve C. Tissue fractionation past and present. *J Cell Biol* 1971;50:20d–55d.
24. Papac DI, Franklin MR. N-benzylimidazole, a high magnitude inducer of rat hepatic cytochrome P-450 exhibiting both polycyclic aromatic hydrocarbon- and phenobarbital-type induction of phase I and phase II drug-metabolizing enzymes. *Drug Metab Dispos* 1988;16:259–264.
25. DeLeve LD. Cellular target of cyclophosphamide toxicity in the murine liver: role of glutathione and site of metabolic activation. *Hepatology* 1996;24(4):830–837.
26. Hengstler JG, Utesch D, Steinberg P, Ringel M, Swales N, Biefang K, Platt KL, Diener B, Boettger T, Fischer T, Oesch F. Cryopreserved primary hepatocytes as an *in vitro* model for the evaluation of drug metabolism and enzyme induction. *Drug Metab Rev* 2000;32:81–118.
27. Li AP, Roque AP, Beck DJ, Kaminski DL. Isolation and culturing of hepatocytes from human livers. *Methods Cell Sci* 1992;14(3):139–145.
28. Moldeus P, Hogberg J, Orrenius S. Isolation and use of liver cells. *Methods Enzymol* 1978;51:60–70.
29. Ryan CM, Carter EA, Jenkins RL, Sterling LM, Yarmush ML, Malt RA, Tompkins RG. Isolation and long-term culture of human hepatocytes. *Surgery* 1993;113:48–54.
30. Seglen PO. Preparation of isolated rat liver cells. *Methods Cell Biol* 1976;13:29–83.
31. Steffan AM, Gendrault JL, McCuskey RS, McCuskey PA, Kirn A. Phagocytosis, an unrecognized property of murine endothelial liver cells. *Hepatology* 1986;6(5):830–836.
32. Elaut G, Papeleu P, Vinken M, Henkens T, Snykers S, Vanhaecke T, Rogiers V. Hepatocytes in suspension. *Methods Mol Biol* 2006;320:255–263.
33. Dunn JC, Yarmush ML, Koebe HG, Tompkins RG. Hepatocyte function and extracellular matrix geometry: long-term culture in a sandwich configuration. *FASEB J* 1989;3(2):174–177. Erratum in: *FASEB J* 1989;3(7):1873.
34. Loretz LJ, Li AP, Flye MW, Wilson AG. Optimization of cryopreservation procedures for rat and human hepatocytes. *Xenobiotica* 1989;19(5):489–498.
35. Gomez-Lechon MJ, Lahoz A, Jimenez N, Castell VJ, Donato MT. Cryopreservation of rat, dog and human hepatocytes: influence of preculture and cryoprotectants on recovery, cytochrome P450 activities and induction upon thawing. *Xenobiotica* 2006;36(6): 457–472.
36. Busby WF Jr, Ackermann JM, Crespi CL. Effect of methanol, ethanol, dimethyl sulfoxide, and acetonitrile on *in vitro* activities of cDNA-expressed human cytochromes P-450. *Drug Metab Dispos* 1999;27(2):246–249.

# METABOLITE PROFILING AND STRUCTURAL IDENTIFICATION

Mehran F. Moghaddam
*Celgene, San Diego, California*

**Contents**

1 Introduction
2 Why Is Metabolite Profiling Important?
  2.1 Biotransformation-Assisted Drug Design and Discovery
  2.2 Discovery of Pharmacologically Active Metabolites
  2.3 Detection of Toxicologically Active Metabolites
3 How Are Metabolites Generated?
4 What Are Biotransformation Reactions and How Do They Generate Reactive Metabolites?
5 How Are Metabolites Isolated and Identified?
  5.1 Mass Spectrometry
  5.2 LC/NMR
  5.3 Other Analytical Techniques
  5.4 Useful Chemical Modification techniques
  5.5 A Case Study
6 Conclusion
Acknowledgments
References

## 1 INTRODUCTION

In the pharmaceutical and agrichemical arena, transformation of a compound into its metabolite(s) via *in vivo* or *in vitro* biochemical reactions is referred to as biotransformation reactions. Metabolite profiling refers to the process of identifying

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

all drug related metabolites generated through biotransformation reactions in animals or in *in vitro* systems. In animals, this process may result in the presence of metabolites in the circulation and in any and all tissues as well as the excreta, collectively referred to as animal matrices. In general, biotransformation reactions deactivate/detoxify xenobiotics and render them more water soluble and hence more amendable to excretion (Fig. 1a). This leads to clearance of drugs from the body. However, there are clear examples of bioactivation of compounds after biotransformation, which may lead to generation of pharmacologically or toxicologically active metabolites. As depicted in Fig. 1b, biotransformation can be beneficial in timely degradation of drugs so they do not accumulate in the body and exert their effects continuously. Also, generation of pharmacologically active metabolites can be viewed as a beneficial aspect of biotransformation. At the same time, the process of biotransformation can have deleterious effects on drug action when it results in rapid clearance of a drug or transforms it into toxic metabolites.

The goal of the ensuing discussion is to provide readers unfamiliar with the field of drug metabolite identification and profiling with an overall knowledge of the concepts involved and to enable them to understand the literature in this area of science.

## 2 WHY IS METABOLITE PROFILING IMPORTANT?

Metabolite profiling should be conducted as early in the process of drug discovery and development as possible for several reasons.

### 2.1 Biotransformation-Assisted Drug Design and Discovery

By identifying metabolite(s) of a compound in *in vitro* systems early in the discovery process, one may uncover the metabolic "weak spots" on a molecule. Medicinal chemists should be informed as soon as metabolites have been identified in order to stabilize "weak spots" on the experimental molecules via chemical modifications. This allows medicinal chemists to engender metabolic stability into and lower metabolic clearance of the chemical series of interest in a timely fashion [1]. Additionally, metabolites with chemical features capable of generating adverse effects should be pointed out to guide chemists to eliminate progenitors to such moieties.

### 2.2 Discovery of Pharmacologically Active Metabolites

In some instances, biotransformation of a drug can lead to a pharmacologically active metabolite (Fig. 1a). Although one may envision a more extensive definition, the literature defines a pharmacologically active metabolite as a metabolite with activity against the same pharmacological target as the parent drug [2, 3]. An example of a published pharmacologically active metabolite is Clarinex™ (desloratadine), which is a metabolite of Claritin™ (loratadine) with a 10-fold

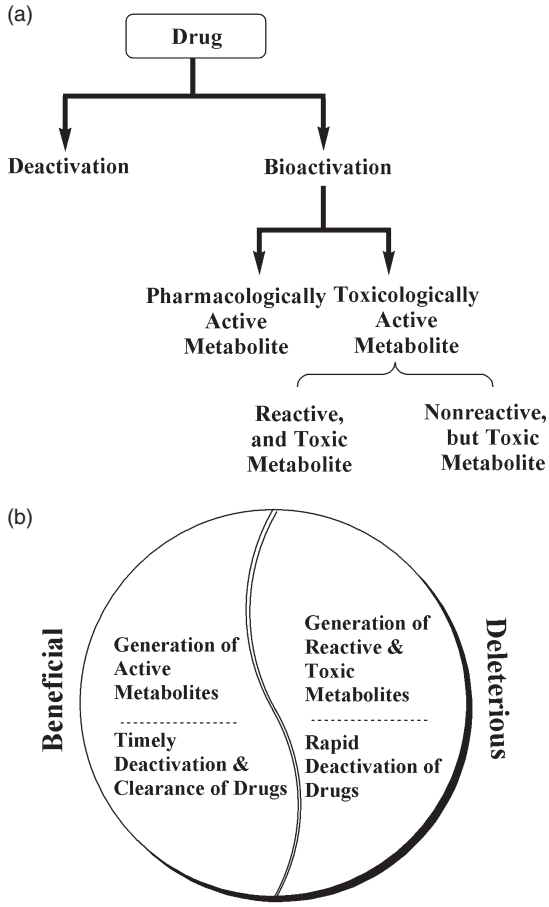


**FIGURE 1** (a) Drug metabolism results in deactivation or bioactivation of a drug. Bioactivation of a drug leads to formation of either pharmacologically or toxicologically active metabolites. Toxic metabolites may or may not be chemically reactive in nature. (b) Drug metabolism is a process that can be viewed as beneficial or deleterious depending on the outcome. It may be beneficial if it generates pharmacologically active metabolites and degrades the drug slowly. However, it will be viewed as a deleterious process if it generates reactive metabolites and/or results in rapid clearance of drugs.

increase in its potency, longer half-life, and greater exposure [2]. Other examples include 6-*O*-glucuronide of morphine with higher levels of exposure and lower incidences of side effects, Allegra™ (fexofenadine) with no cardiac side effects as opposed to terfenadine, its parent drug, and Zyrtec™ (cetirizine) with a higher affinity for H1 receptor and a lack of distribution to brain (therefore nonsedative) as opposed to its parent drug, Atarax™ (hydroxyzine).

Early knowledge of the metabolic profile of a compound allows one to monitor it in plasma during pharmacokinetic (PK) and efficacy studies. In the discovery setting, this establishes the relevance of metabolites generated *in vitro* to *in vivo*

findings and potentially uncovers pharmacologically active metabolites. Knowledge of active metabolites can assist in establishment of better PK/PD (pharmacokinetic/pharmacodynamic) correlations. Determination of the extent of contribution of an active metabolite to the efficacy of a drug is needed for more accurate dose selection in the clinic. Furthermore, pharmacologically active metabolites can be synthesized and tested as potential drugs themselves [3].

### 2.3 Detection of Toxicologically Active Metabolites

Early identification of metabolites may assist in safety assessment of molecules in the discovery process. From a regulatory and safety perspective, it is critical to demonstrate that all human metabolites are present in animal species tested in toxicology studies. This could be partially assessed by conducting *in vitro* studies with hepatocytes, S9 fractions, or microsomes from human and preclinical species, as discussed later.

In 2002, a group of scientists from the pharmaceutical industry proposed a guideline, "metabolites in safety testing" or "MIST," for assessing the safety of metabolites [4]. It was primarily proposed that human metabolites present in circulation at 25% or more of the total drug related material should be considered for testing. In 2005, in order to address this issue, the U.S. Food and Drug Administration (FDA) put forth a draft guidance (http://www.fda.gov/cder/guidance/6366dft.pdf). This document makes recommendations on when and how to identify, characterize, and evaluate major metabolites or metabolites unique to humans for safety. Unique metabolites are defined as metabolites that occur in humans only, and therefore have not been adequately tested for safety in preclinical species in toxicology testing. Major metabolites are defined as those identified in human plasma that account for greater than 10% of the drug related material (administered dose or systemic exposure, whichever is less), which again were not present in the preclinical species at significant amounts. The FDA recommends that attempts be made to identify any major metabolites or metabolites unique to humans, as early as possible so these molecules can be synthesized and properly tested for safety in a timely manner. It is warned that discovery of major metabolites or metabolites unique to humans late in the development process can cause delays and have possible implications for marketing approval of new drugs.

Although a cutoff point of 10% of circulating drug related material is used by the FDA, the Environmental Protection Agency (EPA) regards the threshold for metabolite identification and toxicological testing at 5% [5]. There are clear examples of metabolites circulating at levels below 10%, which have been reported to cause toxicity. Therefore it has been stated by the FDA that the issue of safety concerns should be handled on a case-by-case basis regardless of how a "major metabolite" is defined [5].

Some examples cited by the FDA include felbamate, cyclophosphamide, and acetaminophen. Felbamate, used for the treatment of epilepsy, has been associated with aplastic anemia and hepatotoxicity, both attributed to a reactive metabolite, atropaldehyde [6]. Atropaldehyde is found as a mercapturic acid urinary metabolite (~2% of felbamate concentration in urine) and mercapturic alcohol (~13% of felbamate concentration in urine). Cyclophosphamide is not directly cytotoxic. However, it is metabolized to several toxic metabolites, one of which, 4-

hydroxycyclophosphamide, represented about 8% of total plasma exposure [7]. Acetaminophen liver toxicity is attributed to *N*-acetyl-*p*-benzoquinonimine (NAPQI), detected in urine as thioether metabolites. The thioethers constitute approximately 9% of a therapeutic dose of acetaminophen [8]. Figure 2 shows the decision tree presented by the FDA regarding metabolite testing for safety. It is important to note that reactive metabolites often are too reactive to remain in the matrix. This makes their detection in circulation difficult and they may only be observed as conjugates in excreta or not detected at all (bound residues in tissues). Reactive metabolites have been implicated in idiosyncratic reactions as well. This is a subject worthy of a lengthy discussion; however, it is not the focus of this chapter so the reader is referred to a recent review article on this subject [9]. In brief, more than 10% of acute liver failures from consumption of drugs are due to idiosyncratic reactions. Liver failure is now a leading cause of drug failure in the clinic. Remarkably, most routine animal toxicology testing fails to predict this problem in humans. In fact, idiosyncratic reactions do not occur in most patients at any dose, but, when they take place they can be fatal. These reactions are characterized by immune reactions and are generally referred to as hypersensitivity, allergic, type B, or type II reactions.

Smith and Obach [10] have proposed that decisions regarding safety testing of metabolites should be based on absolute abundance of the molecules rather than their relative abundance (compared to the parent) as was proposed by MIST or FDA guidelines. This suggestion was based on the consideration that the presence of a metabolite as some percentage of the parent drug does not convey the concentrations to which an animal or target cells are exposed. For example, given equal potencies, exposing an animal to a metabolite present at 1% of the dose of drug A dosed at 100 mg/kg would constitute a greater risk than exposing the same animal to a metabolite present at 50% of drug B dosed at 1 mg/kg. These authors have

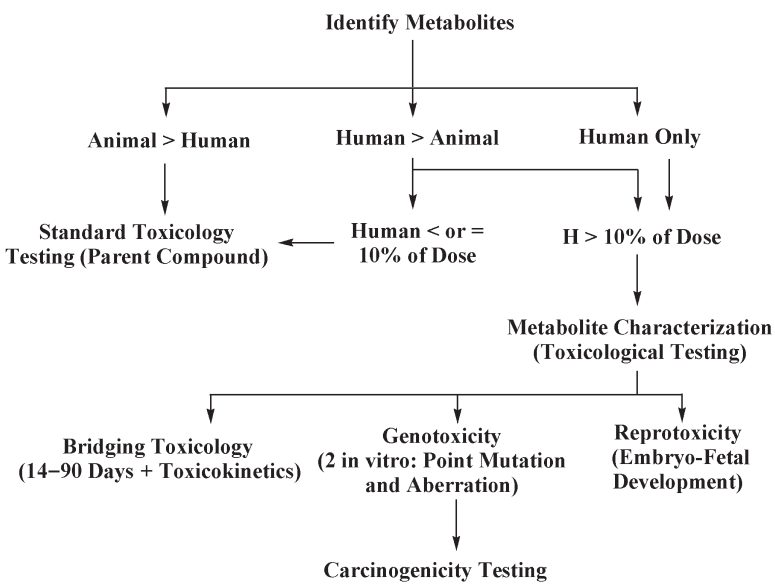


**FIGURE 2** FDA decision tree for safety testing of metabolites.

proposed a new flowchart for metabolite testing. The reader is encouraged to follow the development of this matter carefully by looking for future FDA guidelines.

It is noteworthy that all of the guidelines or recommendations just discussed require a human study using radiolabeled drugs to conduct thorough metabolite identification and profiling studies before accurate measurements of each circulating metabolite can be assessed.

## 3 HOW ARE METABOLITES GENERATED?

Although many organs are capable of metabolizing xenobiotics in animals, the liver is universally recognized as the major site of drug metabolism. Therefore, it is common to use hepatic subcellular fractions such as microsomes and S9 fractions for the generation of metabolites. Some of the advantages offered by the use of subcellular fractions include relative ease of preparation (Fig. 3), ease of storage, commercial availability, relative low cost, flexibility in the manner of use, amenability to high throughput screening, and utility in mechanistic studies with inhibitors. The disadvantages of the use of microsomes and S9 fractions include potential

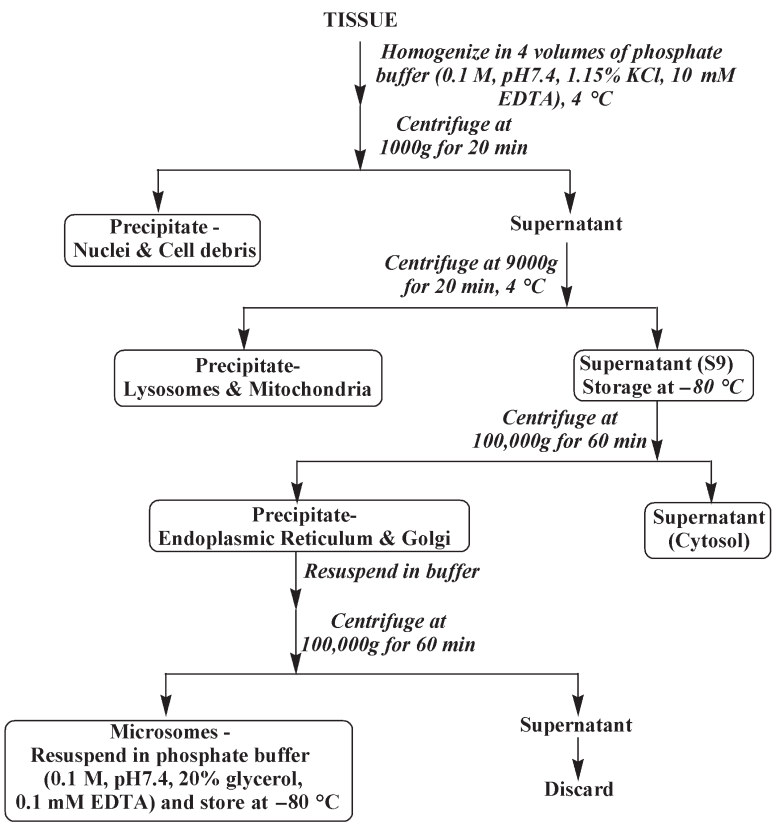


**FIGURE 3** A scheme for preparation of subcellular fractions. Buffers other than phosphate buffer may be used provided the pH values indicated are observed.

inactivation of some enzymes during preparations (FMOs), loss of cellular compartmentalization, lack of all cellular enzymes resulting in limited metabolism, and absence of cofactors, which will need to be added or regenerated during the course of an incubation. Consequently, the utility of intact hepatocytes offers a viable and advantagous alternative for generation of hepatic metabolites. When prepared correctly [11], these cells have all their subcellular fractions and compartments intact and functioning, do not require cofactors, will provide sequential metabolism of compounds, and are therefore more representative of *in vivo* metabolism. Additionally, hepatocytes are correctly predictive of interspecies differences in drug metabolism [12]. The most important factors prohibiting the use of hepatocytes are that it is more complicated to isolate intact fresh hepatocytes than subcellular fractions, it is more costly to procure commercial cryopreserved hepatocytes, and there are limitations associated with cell viability during incubation and freezer storage. It has been recommended that the incubation of hepatocyte suspensions should not exceed 4 hours [12]. It was demonstrated that while this period is sufficiently long to determine metabolic stability and to allow generation of the main metabolites of a test compound, it may be too short to allow generation of some minor, particularly phase II, metabolites [12].

Regardless of which *in vitro* system is used, it is recommended that human biological reagents be utilized in parallel to those from animal species used for pharmacokinetic and efficacy studies. This allows the investigator to quickly determine species differences in metabolite generation and have a chance for an early look into the possibility of formation of metabolites unique to humans. Furthermore, such studies may unravel species differences in rates of clearance and assist in selection of animal species most relevant to the human. Additionally, they assist in determination of whether the metabolic clearance is the cause of rapid clearance in animal species used in PK studies.

In addition to the *in vitro* mixtures, animal matrices are valuable sources of metabolites. These matrices include plasma, urine, bile, feces, and specific tissues. In order to better understand when and how to best utilize these matrices, one must understand the anatomy and physiology of the test subject. Figure 4 is a diagram of the anatomical features of an animal as it relates to drug metabolism. As can be observed, after oral administration (left side of the diagram) a drug enters the stomach and is exposed to gastric hydrochloric acid. In humans, the pH of the stomach can be as low as 1 during the fasted state and as high as 5 in the fed state [13]. In animals, varying levels of acidity are observed in different species [14]. The low pH may result in chemical degradation of drugs in the stomach. The basicity of the gastrointestinal tract is elevated going from the stomach to the colon and rectum, where the pH is around 8 [15]. In the small intestine, hydrolytic enzymes secreted by the pancreas may metabolize a drug. The large intestine has a large reservoir of microbial populations, which can perform hydrolytic and reductive biotransformations of the drugs and/or their metabolites. These metabolites, along with any unchanged drug molecules, are excreted in the feces.

The major site of absorption is the small intestine. A drug molecule and/or its metabolites or degradation products can be absorbed through the wall of the small intestine and enter the portal vein. The intestinal epithelial cells, also known as enterocytes, are enzymatically active and can metabolize xenobiotics. Once these molecules enter the portal vein, they are carried to the liver, where further

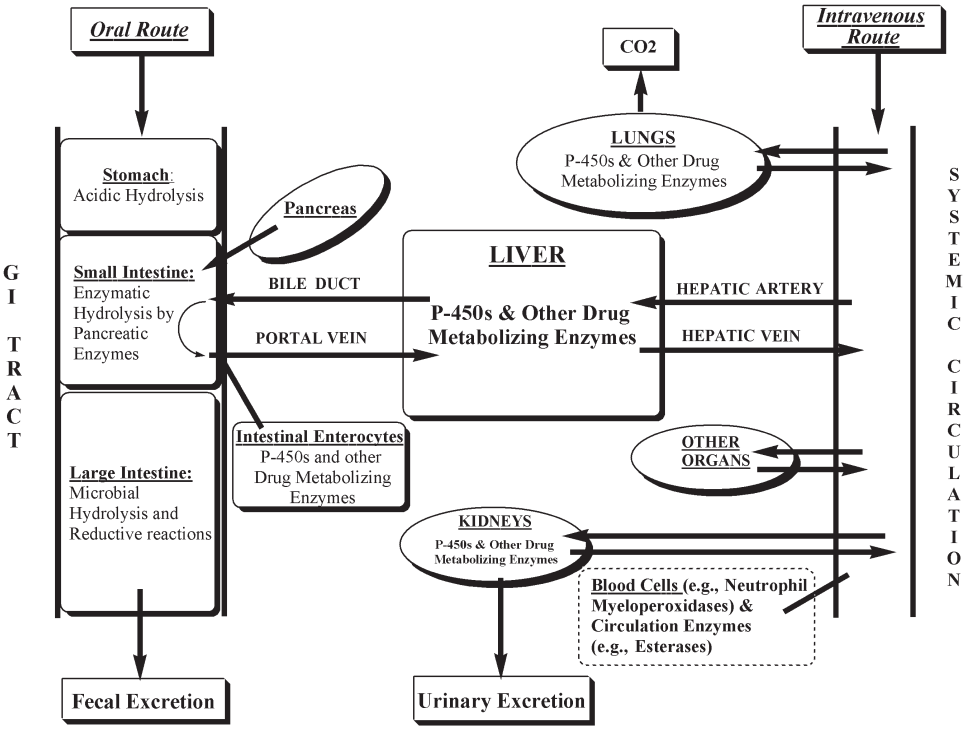


**FIGURE 4** A schematic of mammalian anatomy as it relates to drug metabolism.

metabolism may take place; a process referred to as "first-pass effect." Some of the unmetabolized drug and/or its metabolites may be secreted into the bile and return to the small intestine via the bile duct. A fraction of these molecules can be reabsorbed through the small intestine wall and enter the portal circulation again, a process referred to as "enterohepatic recirculation." When molecules leave the liver they enter the hepatic vein and move toward the systemic circulation, which in turn carries them to the peripheral tissues and organs. These molecules can go back into the systemic circulation and travel back to the liver and bile duct. Other molecules present in circulation can enter the lung and kidney, both of which are metabolically active organs. In the lung, drug molecules can be converted into metabolites or be extensively metabolized to $CO_2$ and exhaled. In the kidney, both parent drugs and their metabolites can be filtered out of circulation at the glomerulus or be secreted out by the epithelial cells of the loop of Henle into the urine [13]. These molecules are then excreted in urine (or may be reabsorbed within the loop of Henle). Blood itself is a metabolizing tissue, which has circulating enzymes such as esterases. Additionally, cells such as neutrophils in the blood contain myeloperoxidases capable of metabolizing drugs [16].

In the following sections, examples describe how some of the matrices described earlier may be used in metabolite profiling. In PK and efficacy studies, plasma samples are prepared routinely from serially collected systemic blood. Identification of circulating metabolites in PK studies can prove very valuable in validating the importance of the metabolites identified in *in vitro* studies. Also, by simultaneous

monitoring of a parent compound and its circulating metabolites, one may be able to identify pharmacologically active metabolites. This is helpful when there are inconsistencies between maximum plasma concentration and the time to reach that concentration and the maximum efficacy and the time to reach that effect for a drug candidate. Of course, it is important to note that delayed effects are not always due to late forming metabolites and can simply be due to delayed pharmacological effects of a parent molecule. When animals are dually cannulated in the bile duct and portal vein (Fig. 4), metabolite profiling of the blood from the portal vein offers a way to differentiate between hepatic and prehepatic (intestinal) metabolism. In these studies, it is imperative to have animals dually cannulated to prevent hepatic metabolites from enterohepatic recirculation via the bile duct (Fig. 4).

## 4 WHAT ARE BIOTRANSFORMATION REACTIONS AND HOW DO THEY GENERATE REACTIVE METABOLITES?

Enzymatic reactions responsible for the transformation of xenobiotics to their metabolites are referred to as biotransformation reactions. As mentioned previously, these reactions are generally designed to render xenobiotics more water soluble and therefore more amenable to excretion. The liver is the main site of biotransformation. However, other organs, although usually not to the same extent, are capable of participating in this process. These reactions reduce the half-life of a compound and diminish its oral bioavailability due to the first-pass effect in the liver. Traditionally, biotransformation reactions are grouped into phase I and phase II reactions (Table 1).

**TABLE 1 List of Phase I and II Drug Metabolizing Enzymes**

| Enzyme | Reaction |
|---|---|
| Phase I | |
| Cytochrome P450s (CYP450) | Oxidation and reduction |
| Monoamine oxidases (MAOs) | Oxidation |
| Flavin-containing monooxygenases (FMOs) | Oxidation |
| Alcohol dehydrogenases (ADHs) | Oxidation and reduction |
| Aldehyde dehydrogenases (ALDHs) | Oxidation and reduction |
| Xanthine oxidases (XOs) | Oxidation |
| Epoxide hydrolases (EHs) | Hydrolysis |
| Carboxylesterases and peptidases | Hydrolysis |
| Carbonyl (keto) reductases | Reduction |
| Phase II | |
| UDP-glucuronosyltransferase | Glucuronidation |
| UDP-glycosyltransferase | Glycosidation |
| Sulfotransferase | Sulfation |
| Methyltransferase | Methylation |
| Acetyltransferase | Acetylation |
| Amino acid transferases | Amino acid conjugation |
| Glutathione-*S*-transferase | Glutathione conjugation |
| Fatty acid transferases | Fatty acid conjugation |

Phase I reactions include hydrolytic, reductive, and oxidative reactions. Esterases, amidases, and epoxide hydrolases are among the hydrolytic enzymes. There are many reductive enzymes such as keto-reductases that catalyze phase I reactions. Finally, oxidative reactions of phase I enzymes are catalyzed by cytochrome P450s, dehydrogenases (i.e., alcohol or aldehyde dehydrogenases), and flavin-containing monooxygenases (FMOs). These highly prevalent oxidative reactions serve to introduce hydroxyl groups on xenobiotics as well as to modify molecules by oxidative N-, S-, and O-dealkylations, which in turn serve to generate free amines, sulfhydryls, and alcohols. A survey of 300 commonly prescribed drugs from multiple therapeutic areas has revealed that cytochrome P450s are responsible for the biotransformation of about half of the marketed drugs [17]. Cytochrome P450s are a superfamily of heme-containing isozymes. They are embedded primarily in the lipid bilayer of the endoplasmic reticulum of liver cells, which are converted into microsomes during homogenization of liver samples. These enzymes have broad substrate specificities and require reduced nicotinamide adenine dinucleotide phosphate (NADPH) as a cofactor. The dominant cytochrome (CYP) isozymes involved in human metabolism of drugs are CYP3A4, CYP2D6, CYP2C9, CYP2C19, CYP1A2, and CYP2E1; ranked according to their order of importance [18]. CYP3A4 were found to be responsible for ~40% of CYP-catalyzed drug metabolism.

Phase II reactions catalyzed by transferases are characterized as conjugation reactions (Table 1). These include glucuronidation, sulfation, glutathione and amino acid conjugation, acetylation, and methylation reactions of drugs. Glucuronidation reactions, catalyzed by uridine-diphosphoglucuronosyltransferase or UDPGT (microsomal) using UDP-glucuronic acid (UDPGA) as the endogenous reagent, are the most common phase II reactions. UDPGTs are ubiquitous to most tissues [18]. The net result of this reaction is addition of β-glucuronic acids to alcohols, carboxylic acids, amines, and sulfhydryls. These glucuronides can be deconjugated by β-glucuronidases, which are often used as biochemical tools to free the parent compound during the metabolite identification process. Drug glucuronide conjugates formed in the liver or intestinal enterocytes are excreted into the small intestine after secretion into the bile. Once in the small intestine, these conjugates can be hydrolyzed back to the parent drug by β-glucuronidases secreted into the lumen of the small intestine by the pancreas. The drug molecule can then be reabsorbed (enterohepatic recirculation). The kidney is also a significant site of glucuronidation and is capable of excretion of glucuronides into the urine.

It is not the objective of this chapter to discuss all the biotransformation reactions extensively, as this subject has been reviewed at length in comprehensive review articles [19, 20]. However, a brief overview of this subject is presented to provide the reader with the basic understanding of the subject. Selected examples of phase I biotransformation reaction mechanisms are presented in Fig. 5. These general reactions show oxidation of double bonds, heteroatoms, and carbons. Mechanisms of oxidation alpha to heteroatoms like nitrogen, oxygen, and sulfur, which ultimately can lead to cleavage of the carbon–heteroatom bonds, are also depicted. It is worth discussing the reaction mechanism for glucuronidation (Fig. 6) because of its prevalence and implication in the development of adverse effects, which will be discussed later. As can be observed, the lone pairs of electrons from an alcohol conduct an SN2 attack on the anomeric carbon of glucuronic acid, which has a phosphoester linkage to the uridinyldiphosphate molecule in UDPGA. This results

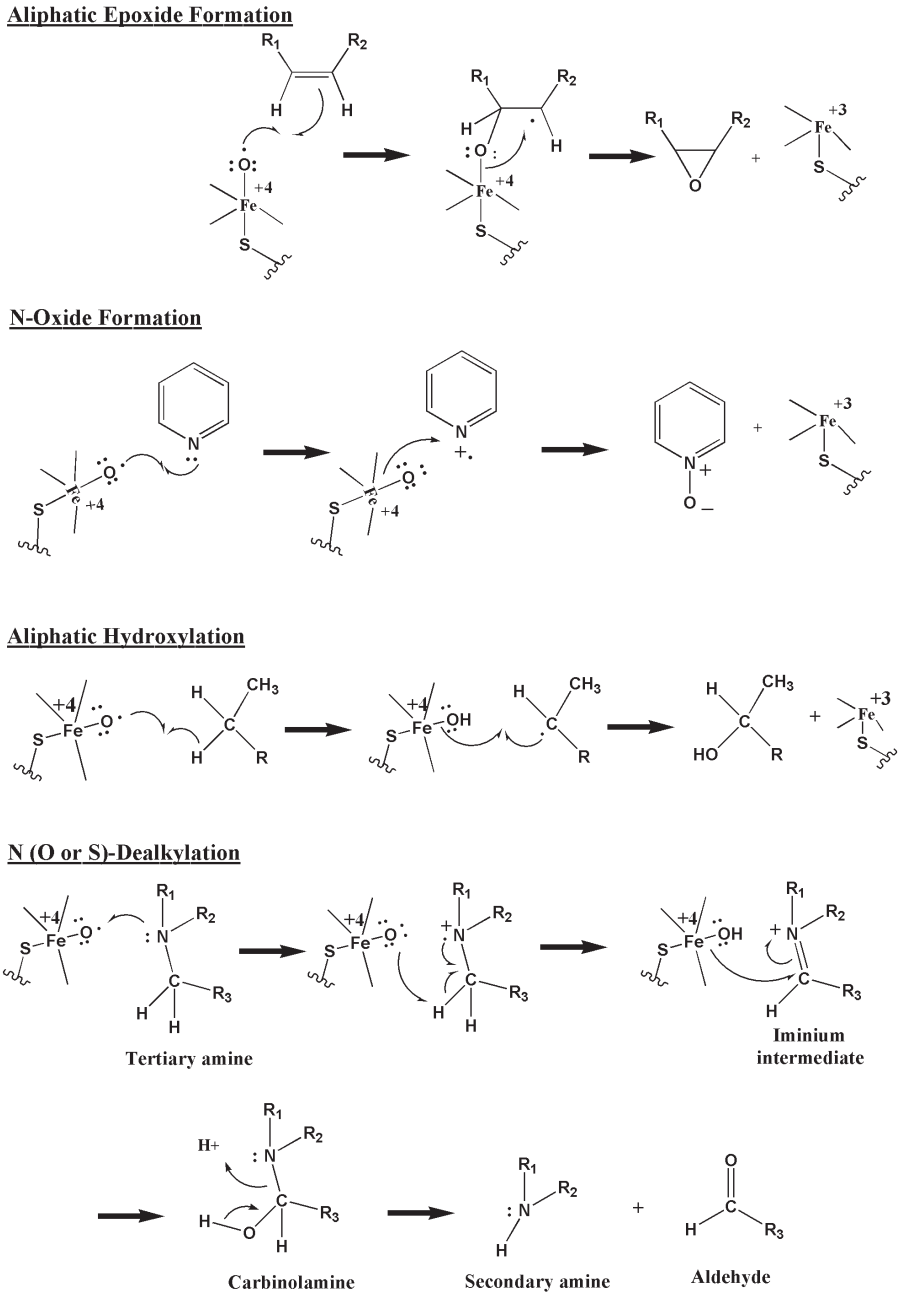


**FIGURE 5** Selected examples of P450 oxidative reaction mechanisms.

in a substitution of the UDP, an inversion of the bond into a β-bond and hence, formation of a β-glucuronide conjugate.

Reactive metabolites have become an increasingly important topic of discussion in recent years due to the adverse effects associated with them. They are the unintended products of what was traditionally referred to as "detoxification pathways"

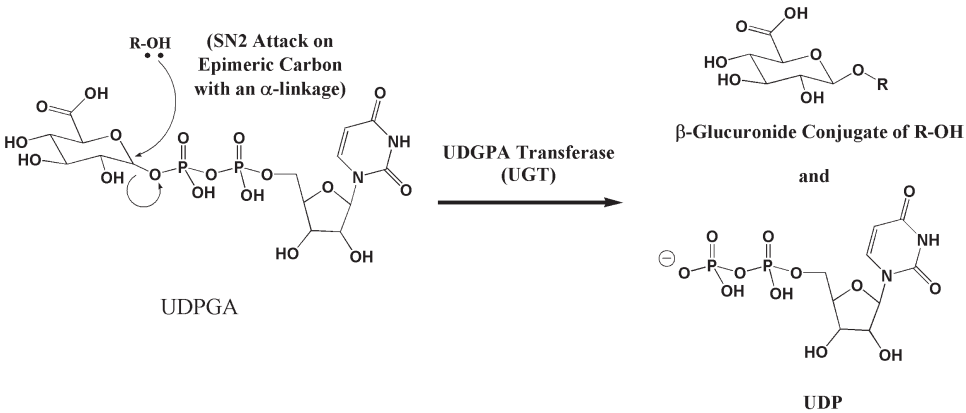


**FIGURE 6** UDPGA transferase mechanism of glucuronidation. A conjugating nucleophile displaces the uridinyldiphospho moiety of UDPGA by an SN2 attack on the anomeric carbon of glucuronic acid.

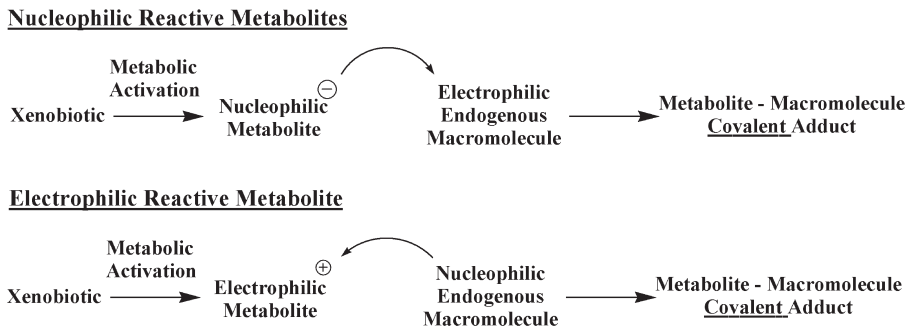


**FIGURE 7** Covalent binding to cellular constituents can be conducted by nucleophilic or electrophilic reactions.

[21]. In general, reactive metabolites can be categorized as electrophilic or nucleophilic compounds (Fig. 7). After metabolic activation into a nucleophilic compound, a xenobiotic can react with an electrophilic endogenous molecule and form a stable covalent bond. Xenobiotics with a potential to form electrophilic metabolites draw reactions from nucleophilic endogenous molecules and form adducts. In such a way, reactive metabolites can cause enzyme deactivation, lead to off target toxicity, and cause idiosyncratic reactions. As shown in Fig. 8, the methylenedioxyphenyl moiety present in many xenobiotics such as Ecstasy (methylenedioxyamphetamine, MDA) can be metabolized to a nucleophilic carbene metabolite, which can covalently bind the heme group within the active site of P450s like CYP2D6, hence resulting in severe toxicity [22]. Demethylenation of the methylenedioxy moiety may play a role in the toxicity of such groups due to formation of quinone type intermediates [20, 23]. Aniline is an example of a compound that can be metabolized to form an electrophilic reactive metabolite, namely, a reactive nitrene (Fig. 8).

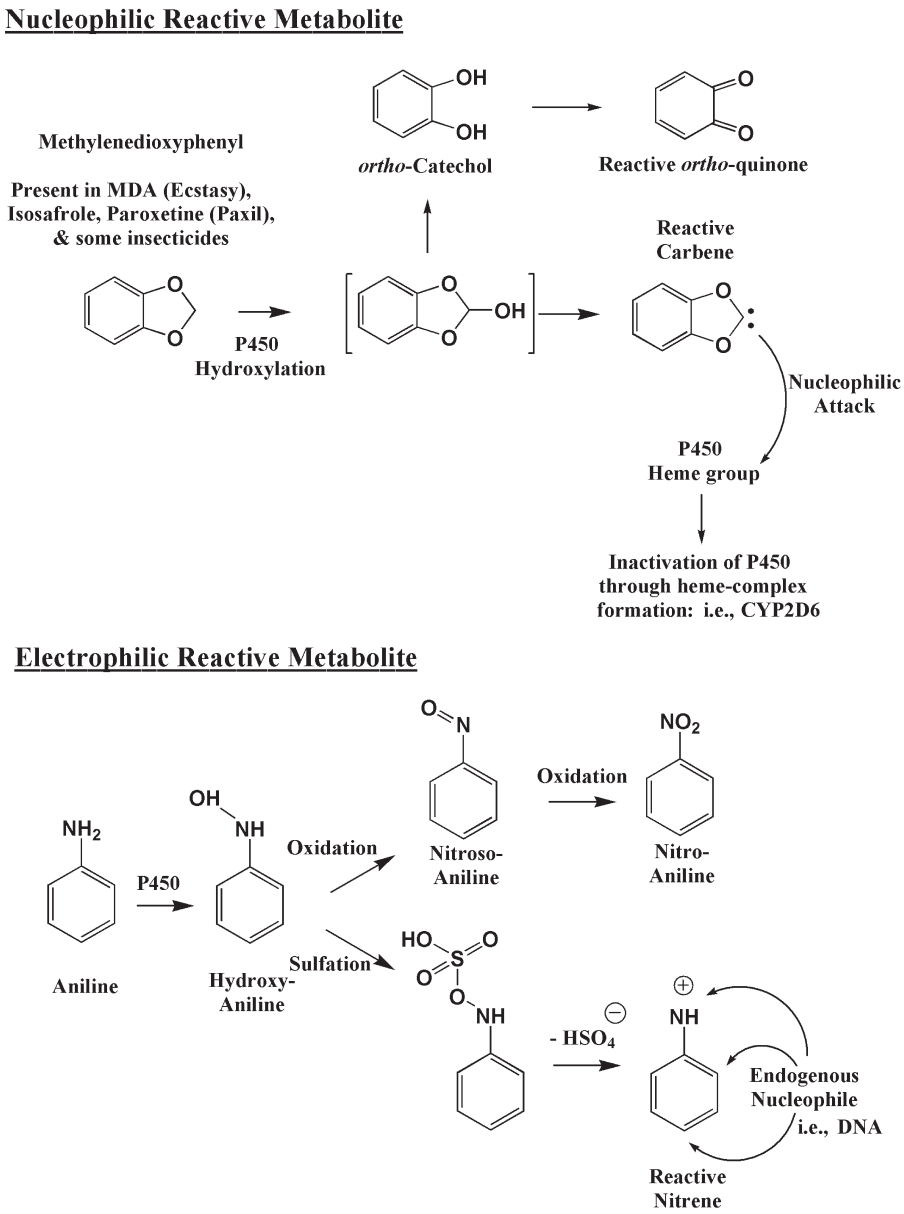


**FIGURE 8** Selected examples of nucleophilic and electrophilic reactions of reactive metabolites.

Diclofenac, a widely used nonsteroidal anti-inflammatory drug (NSAID), can cause a rare but serious hepatotoxicity [24]. Around four in 100,000 users (180 confirmed cases in the first three years of marketing in the U.S.) of diclofenac users developed severe liver damage with an 8% rate of fatality [24]. Diclofenac contains two masked anilines and a carboxylic acid moiety, all of which provide opportunities for reactive metabolite formation. Figure 9 describes biotransformation of the

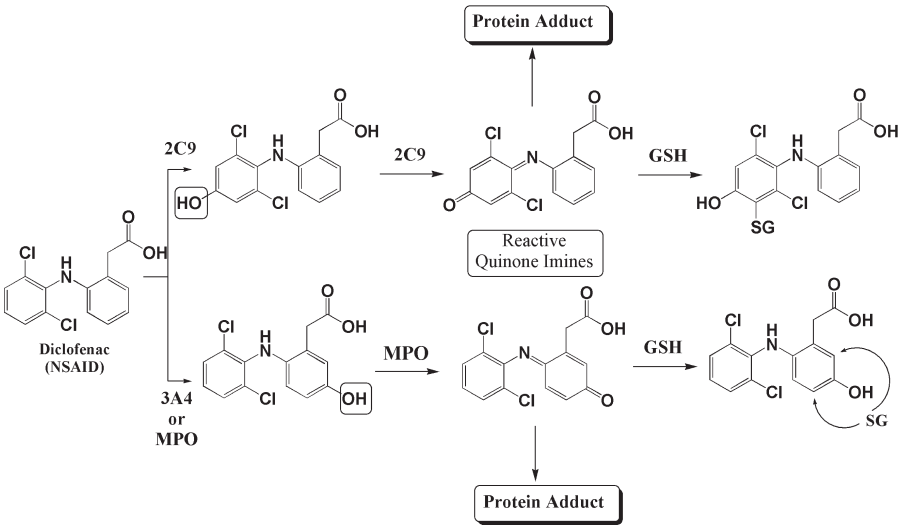


**FIGURE 9** Potential bioactivation of diclofenac to reactive species.

anilinic groups of diclofenac into electrophilic reactive metabolites capable of binding nucleophilic molecules. As shown, the two aromatic rings can be hydroxylated *para* to the nitrogen by CYP2C9 and CYP3A4 or myeloperoxidases (MPOs). Further metabolism results in the two respective *para*-aminoquinones, which can undergo nucleophilic attacks at *meta* positions to the nitrogen by nucleophiles like glutathione or macromolecules to form covalent adducts. The carboxylic acid moiety on this, as well as many other drugs, is a suitable site for acyl glucuronidation.

A glucuronide conjugate of a carboxylic acid is referred to as an acyl glucuronide. Acyl glucuronides, but not ether glucuronides (of alcohols), have been implicated in irreversible binding of proteins resulting in haptenization and immune reactions. The first documentation of such adduct formation was by Smith et al. [25] using zomepirac, an NSAID with adverse effects, in human plasma and *in vitro* using albumin. The mechanism of toxin formation was described as shown in Fig. 10, based on mechanisms previously described for reaction of bilirubin glucuronides [26], glycosylation of hemoglobin [27–29], and albumin [30, 31]. As illustrated in Fig. 10, a 1-*O*-acyl glucuronide, which is formed via the mechanism described in Fig. 6, can undergo several routes of metabolism. The arrow labeled as 1 shows intramolecular migration of the constituent. During this process the hydroxyl group on C2 of the glucuronic acid substitutes the hydroxyl group of the anomeric carbon (C1) via an SN2 reaction. This process is referred to as "acyl migration" and the acyl group can continue to migrate to the other hydroxyl groups on C3 and C4. The transacylated metabolite, 2-*O*-acyl glucuronide, can exist in two isomeric forms, one of which can undergo a ring opening to yield a reactive aldehyde. Reaction of this aldehyde with an amino group on a protein, referred to as "glycation" of the protein, results in a Schiff base formation. This constitutes a stable covalent linkage of the acyl group to the protein mediated by the glucuronic acid. At this point, the acyl group–protein complex is haptenized and recognizable by the immune system as an antigen. Formation of antibodies against this hapten will be problematic and a

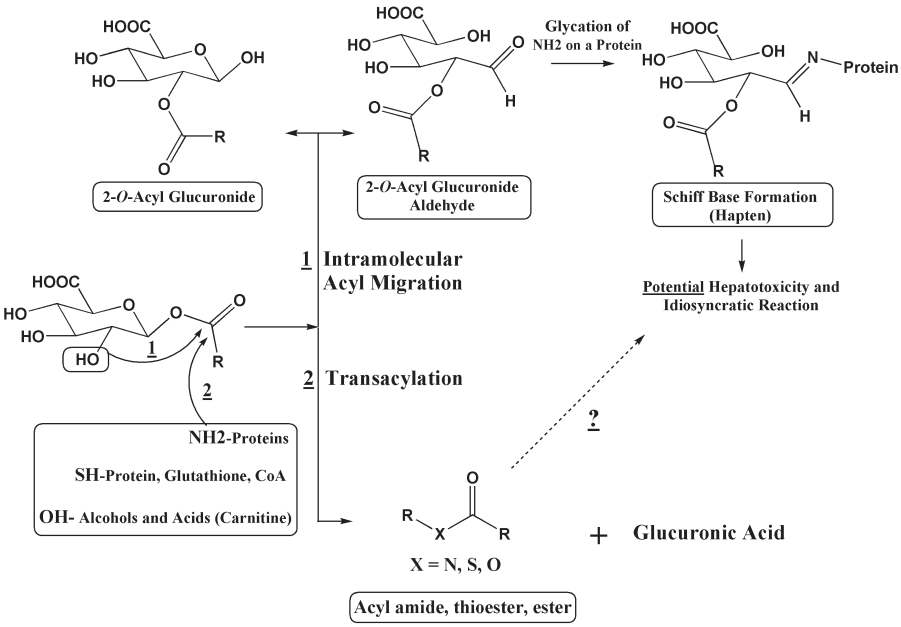


**FIGURE 10** Formation of an acyl glucuronide and its reactivity.

potential reason for idiosyncratic reactions. Indeed, circulating antibodies against diclofenac–liver protein adducts have been detected in the sera of seven out of seven patients with diclofenac-induced hepatotoxicity [24].

Additionally, the same observation was made in 12 of 20 subjects on diclofenac without hepatotoxicity, and none of the control subjects had this circulating antibody [24]. The other path for an acyl glucuronide metabolism is path 2 in Fig. 10. In this case, a nucleophilic moiety such as an amine, sulfhydryl, or hydroxyl group of another molecule will displace the glucuronide moiety to form an acyl amide, thioester, or ester. If these transacylating groups are positioned on macromolecules, haptenization may take place depending on the stability of the covalent bond, this time through a direct linkage of the xenobiotic to a macromolecule.

In order to understand whether reactive metabolites are formed by a drug or drug candidate and to estimate the extent of their formation, elaborate methods have been developed using trapping agents [32]. From a practical standpoint, during the course of drug discovery and development, most compounds with reactive metabolites, if capable of generating significant adducts and adverse effects, will be excluded due to their poor pharmacokinetics (i.e., exposure) or *in vitro* and *in vivo* toxicity. Furthermore, it is difficult to justify eliminating a compound from the discovery process simply based on its ability to form reactive metabolites, unless these data are coupled with toxicity data. This is because there are *in vivo* mechanisms to degrade most covalently bound adducts and prevent adverse effects. Until there are elaborate correlations established to link particular types and levels of adduct formation to incidences of adverse effects and idiosyncratic reactions, the utility of reactive metabolite trapping experiments in the discovery setting is unclear. This is particularly true in most discovery groups because to quantitate the extent of adduct

formation, radiolabeled drugs are needed and these generally are not available in the early discovery process. However, trapping experiments may be useful in confirming or removing doubts about specific chemical moieties' ability to generate reactive metabolites if such moieties are associated with toxicity.

## 5 HOW ARE METABOLITES ISOLATED AND IDENTIFIED?

Many techniques are employed for isolation and identification of metabolites. Due to the complexity of matrices that contain metabolites, often one has to use a chromatographic method to separate metabolites from each other and from unrelated chemical entities such as endogenous compounds. In the past, gas or liquid chromatography (GC or LC) was used to accomplish such a task. Analytes were first extracted from the matrices using volatile organic solvents and injected onto a GC column. The compounds of interest eluting off the GC column could be condensed and isolated into cooled vessels and reconstituted in appropriate solvents for the purpose of structural analysis. In the case of liquid chromatography, analytes eluting from the LC column were collected manually or using an automated fraction collector. After evaporation of the LC solvent, analytes of interest were reconstituted into solvents appropriate for structural analysis.

With development of tandem gas chromatography/mass spectrometry (GC/MS), the need for isolation of analytes was eliminated. GC/MS methods for the study of metabolites had a limitation in that the metabolites of interest had to be volatile or had to be derivatized to become volatile enough to elute off a GC column and enter into the mass spectrometer [33]. This added a degree of complexity to metabolism studies, because not only a derivatization step was required, but also the heat used to volatilize the analytes could cause chemical degradation of the analytes. With the advent of tandem liquid chromatography/mass spectrometry (LC/MS), the need for sample collection was alleviated. In LC/MS studies, each analyte eluting from the LC system enters the mass spectrometer for structural analysis. In addition to LC/MS techniques, there are other methods that couple the LC separation to a variety of techniques such as UV, IR, and NMR. The main focus of this discussion is modern LC/MS techniques due to the dominant role of mass spectrometry in metabolite identification.

### 5.1 Mass Spectrometry

The goal of this discussion is to provide the reader with a basic understanding of mass spectrometry as it relates to the identification of metabolites. Therefore, detailed information regarding the architecture of mass spectrometers will not be provided. A mass spectrometer is an instrument designed to distinguish gas-phase ions according to their mass-to-charge ($m/z$) ratios. Figure 11 depicts an oversimplification of an LC/MS instrument. Generally, analytes separated by the LC column enter a chamber referred to as an ion source before entering the high vacuum chamber of a mass spectrometer via a vacuum interface component. The mass analyzer components, as well as the mass detector unit, are placed in a high vacuum chamber to minimize the possibility of intramolecular collisions and to assist in movement of the ions toward the detector. The mass analyzers may be

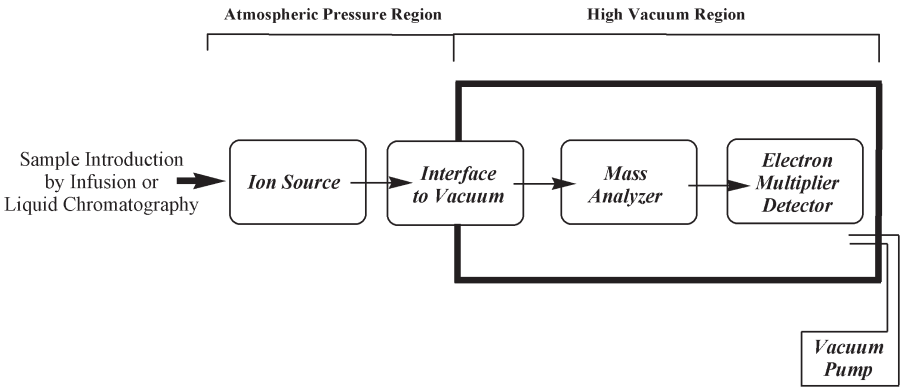


**FIGURE 11** An oversimplified depiction of the components in a triple-quadrupole mass spectrometer.

operated with electric, magnetic, or both fields. These fields are utilized to move ions to the detector after they enter the high vacuum environment in order to produce a signal. The mass-to-charge ratios (and not just the mass) of ions are responsible for separation of the ions in the electrical and/or magnetic fields.

***Ion Generation*** Analytes are introduced into a mass spectrometer via an interface called a probe. Details of how an ionization probe operates are not discussed here. It is only pointed out that the analytes suspended in the HPLC solvent pass through a capillary tube within the probe and are introduced into the ion source chamber (discussed later) as a fine mist of droplets. There are several ways in which analytes may be ionized in LC/MS. Atmospheric pressure ionization (API) is a term used commonly to refer to the technique used to generate ions in the ion source and prior to entry into the chambers operated under vacuum. API techniques include electrospray ionization (ESI), atmospheric pressure chemical ionization (APCI), and atmospheric pressure photoionization (APPI). All of these are referred to as soft ionization techniques.

- In the ESI technique, the ions are generated in the mobile phase of the HPLC by adding a proton donor such as acetic or formic acid or a proton acceptor such as ammonium hydroxide. ESI is a solution phase, soft ionization technique used for any type of analyte, regardless of its volatility. With this technique, special attention should be paid to the $pK_a$ of the analyte and the pH of the mobile phase to ensure ionization and to determine whether positive or negative ions will be generated. This in turn determines whether the mass spectrometer should be operated under positive or negative modes. While ESI can be used for most analytes, it seems to be very effective in analysis of polar compounds with large mass ranges. In ESI, the ions are generated in $[M + H]^+$ or $[M - H]^-$ depending on the mode used.
- In the APCI technique, the analytes reach the ion source in a neutral state. They become vaporized within the HPLC mobile phase using temperature and application of a high voltage needle positioned into the corona discharge. This leads

to their gas phase ionization through a complex process. APCI is typically used for analytes of medium to low polarity that have some volatility. This technique is often used with small molecules with molecular weights of up to ~1500 daltons and is extremely robust and not subject to minor changes in buffers and/or buffer strength. As was the case for ESI, APCI can be operated in both positive and negative ion modes.

- APPI is the newest of the three techniques. It is another gas-phase ionization technique, where the ionization is accomplished using a UV light source. The ions are generated in positive mode only and in the form of $[M]^+$ and $[M + H]^+$.

Once the ions are formed they leave the ion source and enter the first vacuum region of the mass spectrometer, commonly referred to as "Interface to Vacuum," through a narrow orifice leading to an ion transfer capillary tube. These ions then move toward the ion guides (quadrupoles 00 and 0 or Q00 and Q0). Ion guides are multipole rod assemblies that operate with radiofrequency (RF) voltage. In this environment, all ions have a stable trajectory and pass through the system while being focused into ion beams on their way to the mass analyzer. All the neutral and oppositely charged ions are pumped away or crash into the surface before the orifice of the capillary tube. Regardless of the type of ionization, the differences in the design and utility of the mass analyzer differentiates different types of mass spectrometers.

### *Mass Analyzers*

*Triple Quadrupole* In a triple quadrupole or "triplequad" mass spectrometer, the mass analyzer is divided into three segments, each referred to as a quadrupole or a quad. The reason for the name is that each quadrupole consists of four poles or rods assembled parallel to each other (Fig. 12). Rods opposite each other in the quadrupole assembly are considered a pair. Voltages of the same amplitude and polarity are applied to an opposing pair of rods. The voltages applied to the other pair are the same amplitude, but opposite polarity. As the polarity of voltages in each pair of rods oscillates between positive and negative, the ions of interest within each quadrupole are focused and directed toward the detector end of the instrument. In

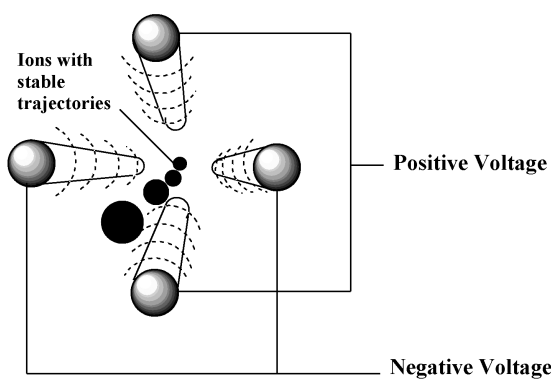


**FIGURE 12** Movement of ions through a quadrupole.

addition to focusing ions, quadrupoles act as selection devices. The voltages used are of AC and DC nature and because the frequency of the AC voltage is in the radiofrequency range, it is referred to as RF voltage. When selecting for a specific ion, the amplitudes of RF and DC voltages are kept constant, only ions with an $m/z$ that resonate with that particular condition will have stable trajectories to pass through the quadrupole assembly to be detected. All other ions will be destabilized and crash onto the rods and become eliminated. In cases where the RF and DC voltages are ramped up, a process referred to as scanning upward, ions of successively higher $m/z$ ratios will be mobilized and reach the detector through stable trajectories. Therefore the ratio of the RF to DC voltages determines the ability of the mass spectrometer to separate ions of different $m/z$ ratios.

The first and third quads (Q1 and Q3) are mass analyzers, which, depending on the type of experiment, can be configured in a specific manner to detect an ion of interest. The second quad (Q2) is a collision cell filled with an inert gas like argon when conducting collision-induced dissociation (CID) or MS/MS ($MS^2$) experiments. CID is a transfer of energy to an ion through collision with a neutral molecule (inert gas) resulting in vibration, cleavage, and/or rearrangement of one or more bonds on the ion. The weakest bonds in an analyte cleave first and to a greater extent resulting in fragments. The resulting fragments are used to assemble the structure of the parent ion. One limitation of a triple quad mass spectrometer is that it will not conduct MS experiments beyond $MS^2$. While for most metabolite identification projects $MS^2$ data is sufficient in determining general sites of metabolism on a molecule, for a more detailed analysis one may resort to an ion-trap instrument (as described later). Triple quads can be used to conduct several types of experiments. These include full scan, selected ion monitoring, product or daughter ion scan, precursor or parent ion scan, selected or multiple reaction monitoring (SRM or MRM), and constant neutral loss scan.

FULL SCAN MASS SPECTROMETRY Full scan is an $MS^1$ experiment, which is used to detect any molecule with an $m/z$ within a certain range in the analyzed sample. A full scan experiment is inclusive of all ionizable molecules in the matrix that enter the mass analyzer and fit the mass range criteria designated by the operator. Figure 13 depicts the steps involved in conducting a full scan experiment. As shown in Fig. 13, in an $MS^1$ experiment, Q1 is set to scan the mass range of interest, while Q2 and Q3 are turned off and Q2 does not contain any collision gas. Alternatively, Q1 and Q2 can be turned off while only Q3 is in the scanning mode. The mass range of interest is scanned by ramping RF/DC voltages to select ions from the lowest to highest $m/z$ repeatedly and once every few milliseconds (scan time). Every ion exiting Q1 then gains a clear path through Q2 and Q3 to an electron multiplier detector and its $m/z$ is recorded. In an LC/MS experiment, as analytes are separated by the chromatography column, they enter the mass spectrometer one by one and this, coupled with the fast scanning capability across a mass range, generates a total ion chromatogram (TIC), which includes all ions eluting from the HPLC. The data output obtained from a full scan experiment is a TIC and a mass spectrum associated with each TIC signal (Fig. 13). It should be noted that the selected mass range is shown on the $x$-axis of each mass spectrum.

As was mentioned before, full scan experiments can be used to gather data on all the ionizable compounds in a mixture. Once this process is finished, one can

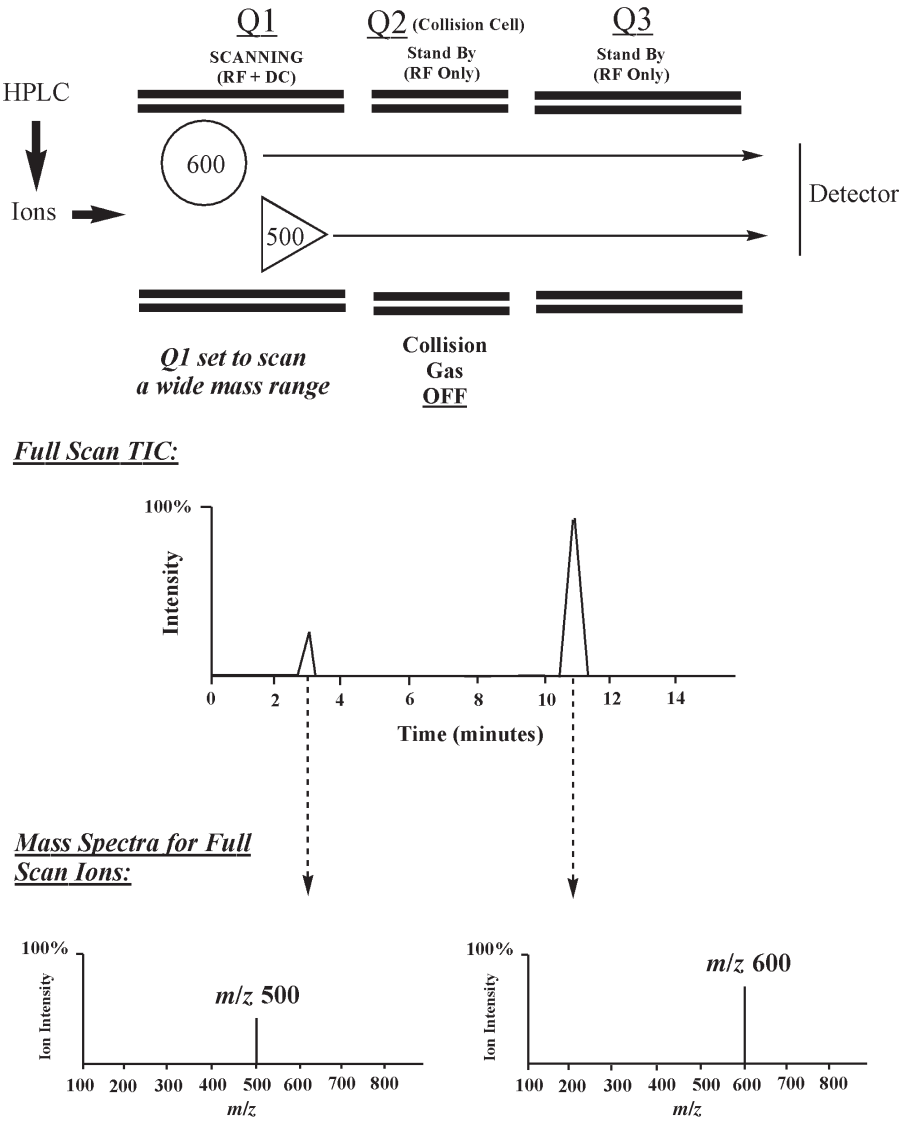


**FIGURE 13** A schematic of a full scan experiment. All ions pass through the triple quadrupole mass spectrometer to be detected. A full scan TIC and mass spectra are generated.

search for a specific ion (a process referred to as "mass extraction"), to detect the presence of compounds with a specific $m/z$. The limitation of this experiment is that one cannot distinguish the compound of interest from isobaric compounds because a full scan dataset only provides information on the overall mass of a compound, and not its structural features. In cases when the operator is limited to the use of a single quad mass spectrometer and is unable to perform $MS^2$ experiments, any TIC peak that shows the $m/z$ of interest may be collected as it elutes from the HPLC and further analyzed by other techniques such as NMR.

SELECTED ION MONITORING This experiment is useful to detect minute quantities of a known analyte in a complex mixture. Therefore, while it is useful in detection of known metabolites, it is not particularly useful in identification of unknown metabolites. Selected ion monitoring (SIM) or selected ion recording (SIR) is another $MS^1$ experiment. In contrast to a full scan experiment, where a mass range is scanned repeatedly, in an SIM experiment the mass spectrometer is configured to acquire and record ions with one or a few selected mass-to-charge ratios. SIM is therefore a more sensitive experiment than full scan because all the scan time is used to focus on a specific mass rather than a broad mass range, as was the case for full scan.

PRODUCT OR DAUGHTER ION SCANNING This $MS^2$ experiment is useful in understanding the fragmentation pattern of a molecule in the mass spectrometer. In a product ion scanning experiment, a particular ion (parent or precursor) entering Q1 is fragmented through collision with an inert gas in Q2 (collision-induced dissociation, CID) to give rise to its product ions (this utility will be discussed later). In the example presented in Fig. 14, the Q1 is set to select only ions with $m/z$ 600 into Q2, while Q2 is filled with the collision gas and Q3 is scanning for a wide mass range ($m/z$ 100–800 as seen on the $x$-axis of the product ion mass spectrum) selected by the operator to capture all fragments. Therefore, in this example, the ion with $m/z$ 600 passes from Q1 to Q2 (or the collision cell) and fragments. These fragments, $m/z$ 150 and 450, as well as any unfragmented parent ion are detected and captured in the product ion mass spectrum. The product ion TIC will represent the parent ion of $m/z$ 600 detected by Q1 and is linked to the mass spectrum containing the fragment ion masses detected by Q3; in this case 150, 450, and 600.

The intensity of the TIC peak depends on the number of ions with $m/z$ 600 detected in Q1. The intensity of ion signals in the resulting mass spectrum depends on the abundance of each ion after CID. Therefore, in this particular example, the intensity of the TIC depends on how many ions of $m/z$ 600 enter Q1, which is a function of the concentration of the analytes in the sample and its ionization potential. On the other hand, the intensity of the fragment ions in Q3 depends on how well the parent ion fragmented in Q2. By elevating collision energy in Q2, one can completely shatter the ion at $m/z$ 600 and even fragment its product ions at $m/z$ 150 and 450 further into their respective product ions. Respectively, by lowering the collision energy, one will observe more of the parent and less of the product ions in the product ion spectrum. It is recommended to optimize the experimental condition such that around 5–10% of the parent ion is detected in the product ion spectrum to prevent excessive fragmentation.

PRECURSOR OR PARENT ION SCANNING This experiment is useful in understanding the ions from which a particular fragment originates. In a precursor ion scanning experiment, all the ions produced in the ion source enter Q1 where they are scanned and sequentially transmitted to Q2. After CID in Q2, the product ions enter into Q3 where only the product ion of interest will be selected and transmitted to the detector. In the example given in Fig. 15, Q1 is set to scan a wide range of ions ($m/z$ 100–800), and therefore all ions within that range are scanned and transmitted to Q2. CID of all of these ions generates their product ions, which all enter into Q3. However, Q3 is set to scan for a specific product ion of $m/z$ 150. In this case, only one of the two molecules gives rise to that product ion. Generation of product ion

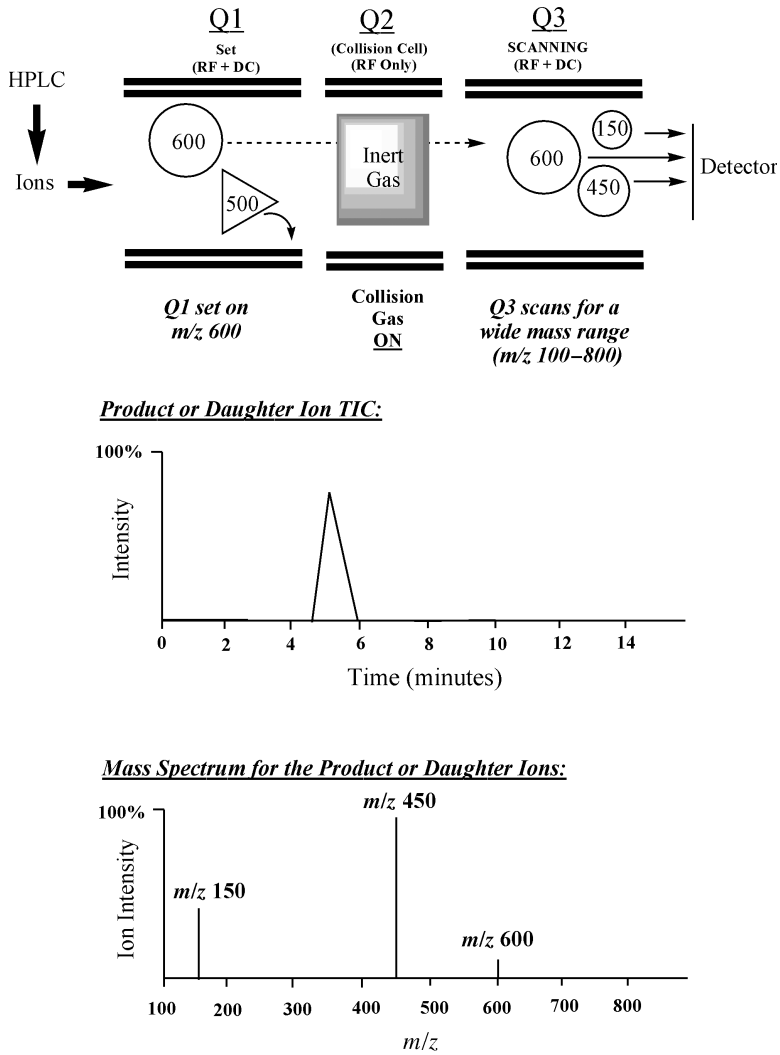


**FIGURE 14** A schematic of a product or daughter ion scan experiment. Selected ions enter the collision cell and their product ions are detected. A product ion TIC and mass spectra are generated.

*m*/*z* 150 is linked to the parent ion with *m*/*z* 600. The data output of this experiment is a TIC of the precursor ion linked to its mass spectrum. Generally, the choice of which product ion to scan in Q3 depends on a previously conducted $MS^2$ experiment (i.e., infusion of the parent compound). Ideally, one optimizes the experimental conditions (i.e., collision energy) for the fragment ion of interest and uses the same experimental conditions for conducting the precursor ion scanning experiment. This maximizes the sensitivity of precursor ion detection.

CONSTANT NEUTRAL LOSS SCANNING This $MS^2$ experiment is useful in detecting conjugates (i.e., glucuronide or sulfate) of known compounds. In a constant neutral

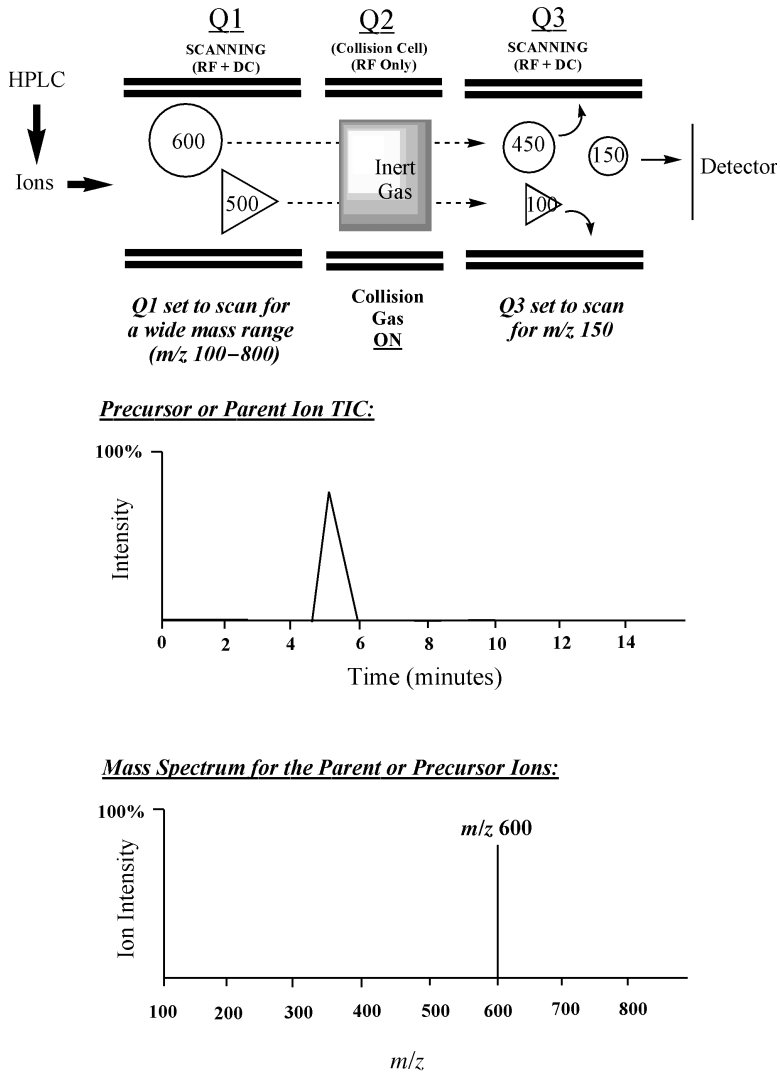


**FIGURE 15** A schematic of a precursor or parent ion scan experiment. Precursors to selected product ions are determined and precursor TIC and mass spectra are generated.

loss experiment, Q1 and Q3 are linked together and they scan at the same rate, over the same mass range. The respective mass ranges, however, are offset by a mass of a neutral moiety. For example, for detection of a glucuronide conjugate, Q3 will scan 176 atomic mass units below Q1 in order to identify a precursor ion, which after CID in Q2 will give rise to an ion 176 mass units lower than itself. This is indicative of the loss of a glucuronide moiety (from the precursor molecule). In the example depicted in Fig. 16, Q1 and Q3 are scanned and only one product ion ($m/z$ 676) is found that is 176 atomic mass units higher than an ion ($m/z$ 500) scanned and transmitted from Q3. The data output of this experiment includes a TIC that shows signals from all parent ions capable of losing $m/z$ 176 and this TIC is linked to the mass spectra of such ions.

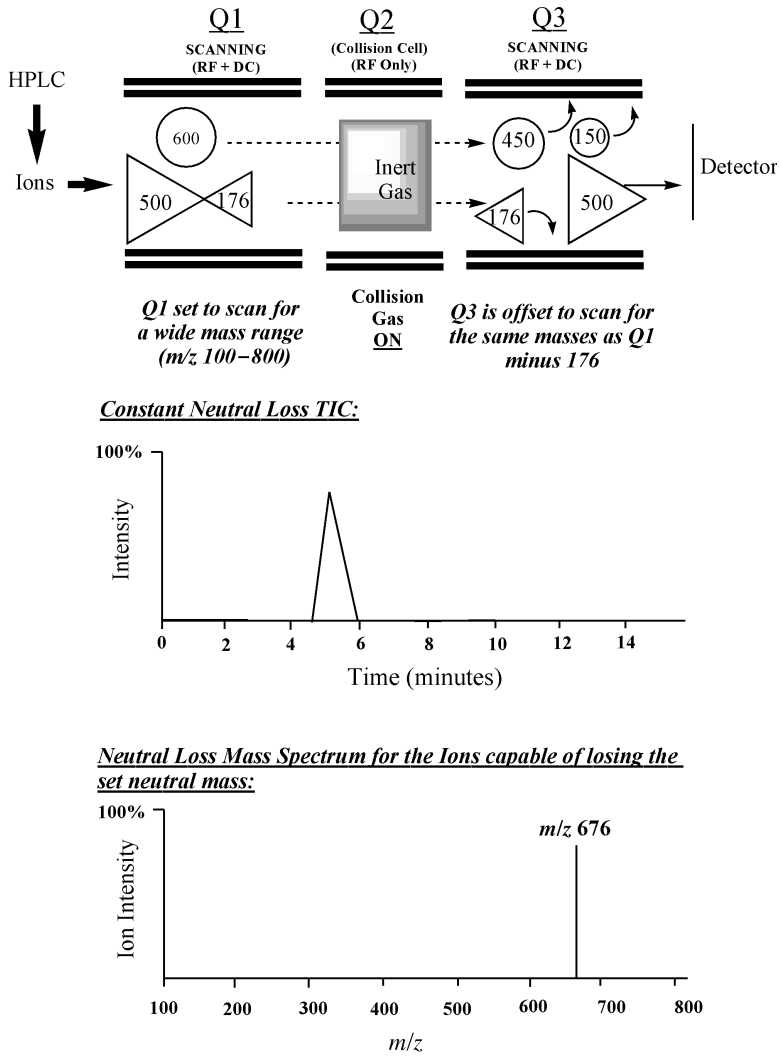


**FIGURE 16** A schematic of a constant neutral loss scan experiment. Specific precursors to specific product ions are determined. A product ion TIC and mass spectra are generated.

A constant neutral loss is difficult to optimize for detection of unknown conjugated metabolites due to unavailability of synthetic metabolites. In the absence of an optimized method, one risks over- or underfragmenting ions of interest. By using too high of a collision energy, one may not only fragment all of the conjugated metabolite but may also go too far and fragment the product ion (deconjugated ion). By using inadequate levels of collision energy, one may not fragment all of the conjugated metabolites into the product ion. In both cases the sensitivity of constant neutral loss experiments is compromised. Therefore, the operator is left with limited choices. In cases when it is the parent drug itself that has been conjugated, the recommendation is to use a level of collision energy such that minimal fragmentation (~5%) of the parent drug is obtained. In most cases, this is enough energy to

tease the conjugated drug apart into the drug, without fragmenting the drug significantly.

SELECTED REACTION MONITORING OR MULTIPLE REACTION MONITORING (MRM) This $MS^2$ experiment is very sensitive and useful in detection of trace levels of analytes in a complex mixture. Generally, this experiment is used for quantitation purposes. However, one can potentially use this method to monitor for existence of a particular metabolite if its fragmentation pattern is known. In setting up for an MRM experiment, one must have prior knowledge of the *m*/*z* of the molecule of interest and its major fragment resulting from CID of the molecule of interest. This information can be hypothetical or obtained from an infusion experiment prior to the LC/MS experiment utilizing MRM. In an MRM experiment, the intensity of the signal obtained is based on transition of parent *m*/*z* to the fragment *m*/*z*. In these studies, all the ions produced in the ion source enter Q1, where they are scanned and only the selected parent ion of interest is transmitted to Q2. After CID in Q2, the product or daughter ions enter into Q3, where only the ion of interest will be transmitted to the detector. In the example depicted in Fig. 17, the operator has previously determined that the parent molecule (*m*/*z* 600) fragments into a major product ion with *m*/*z* 150 under the employed experimental conditions. Therefore, the targeted transition will be 600 → 150. During the conduct of the study, all product ions enter Q1 and become scanned to select *m*/*z* 600. Only the parent ion with *m*/*z* 600 is transmitted to Q2 and fragmented. If and when a product ion of *m*/*z* 150 is detected

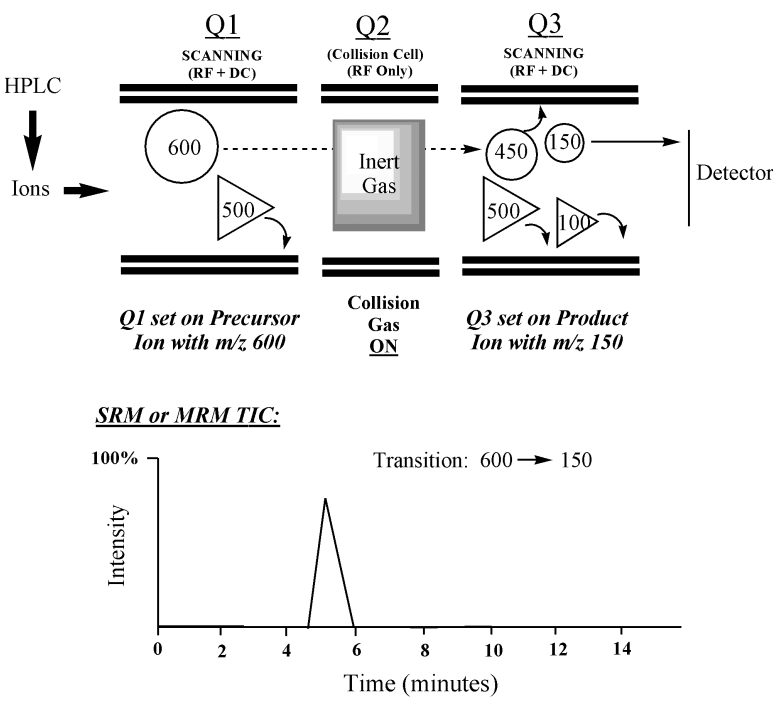


**FIGURE 17** A schematic of a selected or multiple reaction monitoring (MRM) experiment. A specific ion transition is detected and only a TIC is generated.

by Q3, then a TIC signal representing the transition of 600 → 150 is recorded. All other precursor ions with *m*/*z* 600, which do not yield a product ion of 150, are dismissed. As depicted in Fig. 17, the output of an MRM experiment is a TIC, which represents a particular ion transition. Mass spectral data are not obtained in an MRM experiment.

*Ion Trap* As mentioned before, the differences between mass spectrometers are based mainly on the mass analyzers. Generally, all the steps before the mass analyzer in an ion trap mass spectrometer are similar to those described for triplequads. But there is a fundamental difference between an ion trap and a triple quadruple mass spectrometer. Unlike the experiments performed on triplequads that are tandem in space, experiments conducted on ion traps are tandem in time.

Figure 18a simplifies components of a conventional ion trap. After ions enter the trap through the entrance endcap, RF is applied to the ring electrodes to attract or repel the ions. At any one moment the entrance and exit endcaps have an RF of the opposite charge to that of the ring electrodes. This oscillation between opposite

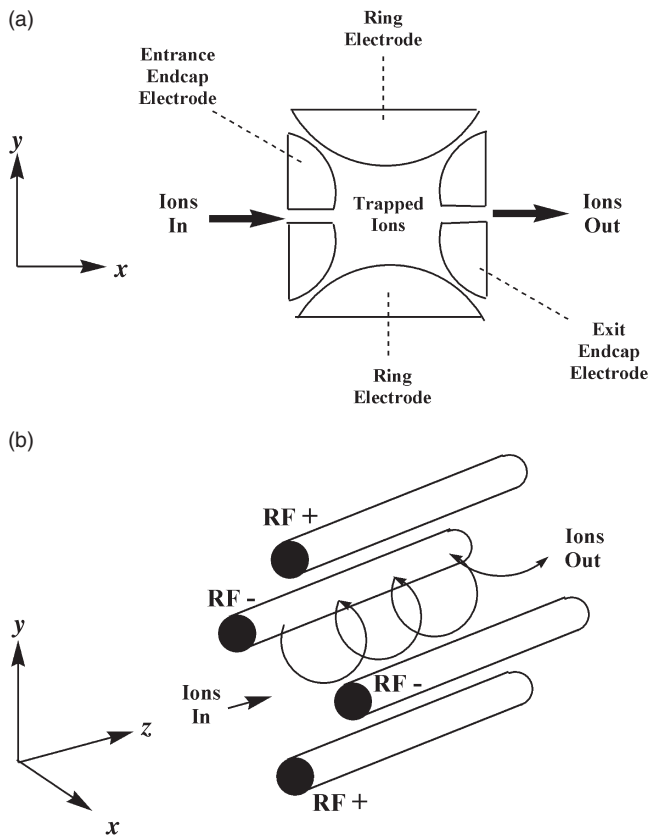


**FIGURE 18** A schematic of different types of ion trap mass analyzers. (a) A traditional ion trap with a cubic configuration and *xy* dimensions for ion movement. (b) A linear ion trap with an elongated configuration and *xyz* dimensions for ion movements.

polarities keeps ions moving around within the trap by being attracted to or rejected by the electrically charged components. Helium gas is used to stabilize the ion movement within the trap and prevent chaotic large orbital motions that result in poor spectral resolution. There are four basic functions in the operation of ion traps and these include ion collection, isolation, excitation, and ejection.

During ion collection, ions enter into the ion trap and are retained there. In the simplest possible type of experiment—a full scan—all ions are trapped, scanned, and ejected to give rise to a spectrum that displays the $m/z$ of all ions that were collected. In ion isolation, all the ions are first collected in the trap as described earlier and then specific RF/DC voltages are applied to eject ions that are not of interest from the trap. This allows only the ion(s) of interest to be retained and transmitted to the detector. This feature can be used for SIM experiments. Also, for the purposes of $MS^2$ after isolation of the ion of interest, voltages are applied to excite the precursor ion. This occurs by increasing the vibrational energy and its collision with the helium gas in the trap to cause its fragmentation. For an $MS^3$ experiment of a precursor ion, all the ions are collected and the precursor ion is isolated by ejecting all other ions. The precursor ion is excited to fragment ($MS^2$), the fragment of interest is isolated by ejecting all other fragments, and the fragment of interest is excited to generate $MS^3$ fragments. The $MS^3$ fragments are then scanned and detected.

For all practical purposes, the mechanisms for trapping ions in linear ion traps are similar to the traditional ion traps. The basic difference between the two systems is depicted in Fig. 18b. The linear ion trap is made up of a quadrupole and the trapped ions move through the triplequad toward the detector in a corkscrew fashion ($z$-axis). Therefore, in the linear ion trap, the ions are trapped in the $xy$ plane as well as the $z$-axis, providing a longer trajectory path to the detector, during which more of the undesirable ions are ejected. This makes linear ion traps more sensitive instruments than their predecessor ion traps.

It is not possible to conduct the $MS^2$ experiments such as precursor ion, product ion, constant neutral loss scanning, or MRM experiments using ion traps. Ion traps can only conduct full scan and SRM experiments. However, one significant advantage of the ion traps over the triplequad mass spectrometer is their ability to conduct data-dependent acquisition experiments. In this type of experiment, as analytes pass through the HPLC column and into the ion trap, the sequence of events taking place in the ion trap are a full scan followed by an $MS^n$ of any detected ion. This is an extremely powerful technique in that it eliminates the tedious need to first examine the full scan data and then to select ions to study further using $MS^n$ in a different experiment. At the end of each data-dependent acquisition experiment, one can examine the full scan ($MS^1$) as well as $MS^n$ data from an ion of interest. The inherent difficulty remains to know what ions are drug related and therefore of interest. In order to facilitate identification of drug related material, one may analyze blank samples at the same time and identify TIC signals in the samples that are absent in the blank samples using baseline subtractions. Alternatively, one may conduct a manual search for $MS^2$ fragments, which are expected to be present in the metabolites of a drug, in the $MS^2$ spectra obtained during the course of a data-dependent acquisition experiment. In fact, the net result obtained from this manual exercise is equivalent to that of a precursor ion scan experiment conducted on a triplequad.

Another manner by which this type of dataset is useful is to devise a list of hypothetical metabolites based on knowledge of the structure of the parent molecule and potential enzyme reactions responsible for degrading it. This subject requires a considerable level of knowledge of drug metabolizing enzymes and is discussed briefly at a later point. One may then manually search the full scan data for $m/z$ of these potential metabolites and, if found, study the $MS^n$ data that is already obtained in the data-dependent acquisition in detail to decipher its structure. Another potential manner by which one may identify drug related compounds is to use inline UV detection. This method requires that the drug and its metabolites exhibit intense UV absorptivities. Otherwise, utility of this technique in metabolite profiling requires major quantities of the drug and its metabolites in the matrix. Finally, the best method for linking a metabolite to a drug at this time remains the use of radiolabeled drugs and online radiochemical detectors.

*Quadrupole Linear Ion Trap* As mentioned previously the triplequad systems have the advantage of conducting experiments such as precursor ion and constant neutral loss scanning as well as MRM experiments for quantitation. However, triplequads cannot perform fragmentations beyond $MS^2$ for structural analysis. Ion traps can perform $MS^n$, however, they are not capable of conducting precursor ion, constant neutral loss scanning, and MRM experiments. The Q-Trap™ system is a hybrid of a triple quadrupole and a linear ion trap system. For example, a Q-Trap instrument may be set up to conduct a large set of MRMs and product ion ($MS^3$) and/or neutral loss scan experiments at the same time. This makes a Q-Trap a very elegant system capable of both quantitation and metabolite identification [34].

*Accurate Mass Techniques* Accurate mass refers to good mass accuracy. This is often confused with the concept of "high resolution." The two concepts are not related. It is much easier to obtain good mass accuracy with a high resolution instrument; however, it is possible to have one without the other. Simply put, resolution refers to how well the separation is between two adjacent masses, while accurate mass refers to how precise the mass assignment is for an ion. Therefore it is possible to accurately assign masses to different ions without high resolution; however, to eliminate interference from an isobaric compound one needs to first resolve it from the ion of interest using a high resolution instrument. In order to better understand the power of high resolution mass spectrometers and the utility of accurate mass measurements, one needs to understand the concepts of "mass resolution" and "mass defect."

The term "resolution" in the context of mass spectrometry refers to the ability of the instrument to distinguish between two adjacent peaks (masses) in a mass spectrum. Mass resolution is calculated by dividing the mass of one ion by the difference between that mass and the next higher mass ($R = M/\Delta M$). The higher the resolution of an instrument, the more effectively it distinguishes between two adjacent masses. By rearranging the above formula, one can estimate what adjacent ions may be separated on mass spectrometers of different resolutions ($\Delta M = M/R$). For example, a mass spectrometer with a resolution of 1000 can separate an ion with $m/z$ 400.2000 from another ion only if it is more than 0.4002 amu ($\Delta M$) apart. This means the next highest mass that can be resolved from $m/z$ 400.2000 is $m/z$ 400.6002.

However, a second mass spectrometer with a resolution of 10,000 can separate an ion with $m/z$ 400.2000 from another ion even if it is 0.0400 amu ($\Delta M$) apart. This means the next highest mass that can be resolved from $m/z$ 400.2000 is $m/z$ 400.2400.

Mass defect is the difference between the monoisotopic accurate (exact) mass of an atom and its nominal mass. For example, the monoisotopic accurate mass of oxygen is 15.9949 and its nominal mass is 16. Therefore, the calculated mass defect for oxygen is (15.9949 – 16) or –0.0051. Mass defect of a molecule is the sum total of its atomic mass defects. For example, the mass defect of a molecule with molecular formula $C_8H_{14}N_3O_3$ will be 0.1036 and its accurate and nominal masses will be 200.1036 and 200, respectively. Using a low resolution mass spectrometer, it would not be possible to distinguish $C_8H_{14}N_3O_3$ from $C_8H_{16}N_4O_2$ because they both have the same nominal mass of 200 and exhibit $m/z$ 201.1 ($MH^+$ in positive ion mode). However, because their atomic composition is different, they each have different mass defects. As mentioned previously, the mass defect of $C_8H_{14}N_3O_3$ is 0.1036 while that of $C_8H_{16}N_4O_2$ is 0.1275. This means that their accurate masses in positive ion mode will be 201.1036 and 201.1275, respectively, and therefore will be separated with the right mass resolution.

Using the formula $R = M/\Delta M$, one can calculate that a high resolution mass spectrometer with a resolution of ~8000 ($M$ = 201.1036, $\Delta M$ = 0.0239) will be able to separate the two isobaric compounds $C_8H_{14}N_3O_3$ and $C_8H_{16}N_4O_2$ from each other. Assuming that these were unknowns, once distinguished from each other, different molecular formulas could be assigned to these compounds based on their accurate masses. This added feature of high resolution mass spectrometer software provides an investigator with elemental composition of unknown ions and their fragments. This is an extremely powerful tool for metabolite structure assignments of unknowns with a high degree of confidence.

Currently, the more frequently used high resolution mass spectrometers for small molecule structural work include the quadrupole time-of-flight (Q-TOF) [35], the TSQ Quantum AM [36], and the hybrid LTQ/Orbitrap mass spectrometers [37].

### 5.2 LC/NMR

Nuclear magnetic resonance (NMR) spectroscopy has been used widely for structural elucidation of natural products and metabolites [38]. In brief, an NMR instrument includes a strong magnet (up to 800 megahertz) that generates a homogeneous field, a radiofrequency transmitter, a receiver, a recorder, a calibrator, and an integrator. The sample is dissolved in the appropriate deuterated solvent, placed in an NMR tube, suspended in the magnetic field, and spun. The appropriate combinations of radiofrequency and magnetic fields are used to study the environment around $^1H$, $^{13}C$, $^{15}N$, $^{19}F$, and other nuclei as well as their relationship to other nuclei in a molecule. This technique is based on absorption of electromagnetic radiation by the sample as a function of the magnetic environment the molecule experiences inside the magnet. A plot of the frequencies of the absorption peaks versus peak intensities is the data output of an NMR experiment [39]. From an NMR plot, information regarding chemical shifts (specific magnetic environment), multiplicity of signals (the interaction between neighboring nuclei), integration of signals, and intramolecular relationships can be extracted. In the case of $^1H$-NMR, once protons

on an analyte have been characterized using one-dimensional $^{1}$H-NMR, two-dimensional NMR experiments such as $^{1}$H–$^{1}$H correlation spectroscopy (COSY) may assist in establishing the relative positions of these protons on a molecule. Positional relationships of protons and carbons can be established by several different types of $^{1}$H–$^{13}$C two-dimensional NMR experiments.

Integration of LC with modern NMR has created a powerful technique for identification of metabolites and without a need for prior isolation. A limitation of this technique is the necessity of using expensive deuterated HPLC solvents and its relative lack of sensitivity. The stop-flow LC system allows for prolonged acquisition of data and in part overcomes the sensitivity issue. Additionally, with the development of solvent suppression softwares to hide signals from protonated solvents and CryoFlowProbes™ to improve sensitivity, the use of NMR has become even more feasible. In fact, with current technology the use of protonated organic modifiers mixed with $D_2O$ and sample sizes as small as 5 μL are possible [38]. Additionally, interfacing LC NMR and MS will provide an even more powerful and complementary technique [40]. For example, chemical moieties such as carboxylic acids that are $^{1}$H-NMR silent due to their proton–deuterium exchange are easily detectable by MS. Conversely, closely eluting or even coeluting isomeric or isobaric compounds are likely to be missed by MS but not by NMR [38]. In the arena of drug metabolism, urine after solid-phase extraction and cleanup appears to be the most analyzed matrix using NMR [40]. Bile has presented a challenge due to its content of bile salts, micelles, and detergents [38]. Finally, NMR unlike MS is not a destructive method and the analyzed samples can be retrieved after analysis.

### 5.3 Other Analytical Techniques

***Radioactivity*** Compounds radiolabeled with $^{3}$H or $^{14}$C provide excellent and facile means for metabolite detection. This is the most effective method for establishing drug relevance. The radionuclide of choice is $^{14}$C because $^{3}$H may exchange with water *in vivo* or *in vitro* and create confusion during metabolite profiling. However, because of the lower cost and ease of radiosynthesis, $^{3}$H has been employed in metabolism studies, particularly in late stage discovery programs. The decision on the location of the radionuclide on a molecule should be a joint decision among the medicinal chemists, radiochemists, and biotransformation scientists. It is crucial to choose a site on a molecule that is stable and is subject to minimal or no biotransformation reactions. In cases where there is an expectation of cleavage of a molecule due to metabolism, sites on either side of the cleavage may be labeled to ensure the presence of radionuclides in all or most of the metabolites. Radiolabeled drug candidates are generally used during preclinical or clinical development due to the cost of synthesis and the timelines associated with radiosynthesis. The matrices containing these compounds and their metabolites can be analyzed by HPLC while collecting fractions. These fractions can be analyzed on liquid scintillation counters after adding scintillation fluids to reconstruct a radiochromatogram. It is also possible to collect fractions into microplates, dry the solvents under nitrogen, and count the radioactivity on a TopCount microplate counter [41].

Alternatively, an online detection of radioactivity may be employed using a variety of radioactivity detectors to establish a radiochromatogram of the eluting

metabolites. HPLC–radiochemical detection requires the use of scintillation fluids for sensitivity. This eliminates the possibility of isolation of radiolabeled metabolites for MS or NMR analysis. It is possible to use online radiochemical detectors that do not utilize scintillation fluid and allow for fraction collection and metabolite isolation. However, this mode of radiodetection is considerably less sensitive than the previous method. A more recent development has been the invention of LC-accurate radioisotope counting (LC-ARC™). This system uses liquid scintillation fluid and, due to its stop-flow design, it is the most sensitive online mode of radiodetection available. The limit of detection for this method is 5–20 DPM for $^{14}C$ and 10–40 DPM for $^{3}H$, which enables this system to generate thorough metabolite profiles for radiolabeled compounds. It is possible to collect parallel fractions from this instrument that do not contain any scintillation fluids and allow for metabolite isolation. Additionally, it is possible to interface this technology to LC/MS systems [41].

***Infrared (IR) Detection*** The electromagnetic region between the visible and microwave regions is called infrared radiation. For determination of unknown structures, the region between 4000 $cm^{-1}$ and 666 $cm^{-1}$ is of the greatest utility [39]. An IR spectrum is characteristic of an entire molecule. However, certain chemical moieties generate the same signals regardless of what molecule they reside on. The presence of specific signals hints at the presence of certain chemical moieties in a molecule but does not provide any information about the structural makeup or architecture of a molecule. This technique may be useful in providing complementary or confirmatory information for structure elucidation of unknowns, as demonstrated in the case study in Section 5.5.

***Ultraviolet (UV) Detection*** Wavelengths in the UV region of the electromagnetic spectrum are expressed in nanometers (nm) and are between 200 and 380 nm. Absorption of UV by a molecule results in the elevation of electrons from ground state orbitals to higher energy orbitals in an excited state [39]. For practical purposes, UV absorption is limited to conjugated unsaturated bonds. Two important factors in UV absorption are the wavelength at which maximal absorption takes place ($\lambda_{max}$) and the intensity of absorption ($\varepsilon$, molar absorptivity or extinction coefficient). Both these factors are dependent on the nature of the conjugated unsaturated bonds in a molecule. UV detection of a metabolite in a biological matrix depends on its concentration, intensity of absorption, and presence of interference from other UV active compounds. There are clear examples of studies where UV monitoring of metabolites were helpful in identification of drug related metabolites [42]. However, in other cases, UV detectors are not very informative due to the reasons mentioned previously.

## 5.4 Useful Chemical Modification Techniques

Derivatization reactions may be used to change physicochemical properties of molecules so they behave differently on an HPLC column and/or ionize differently in mass spectrometers. Also, the derivatized analytes may generate informative signals in their mass or NMR spectra. For example, during metabolism of nitrogen- or sulfur-containing compounds, N- or S-oxides may form. By using mass spectrometry alone, one may not be able to decipher whether the site of oxidation was on a carbon

**TABLE 2 List of Derivatizing Agents**

| Group | Reaction | Reagent |
|---|---|---|
| Hydroxyl | Acetylation | Acetic anhydride/pyridine |
| | Methylation | Diazomethane |
| | Dansylation | Dansyl chloride |
| Amine | Acetylation | Acetic anhydride/pyridine |
| | Dansylation | Dansyl chloride |
| Carboxyl | Methylation | Diazomethane, (trimethysilyl)diazomethane, methanol/HCl |
| | Reduction | Lithium aluminum hydride |
| N-oxides and S-oxides | Reduction | Titanium trichloride |

or a heteroatom within a segment of a molecule. Titanium trichloride ($TiCl_3$) has been used successfully to reduce N-oxides back to amines and establish the position of oxidation on the nitrogen rather than a carbon or even sulfur [43]. The reduction reaction using $TiCl_3$ was found to be efficient even in the presence of biological matrices, which makes this reagent valuable for biotransformation studies. Table 2 provides selected reagents that are common for chemical modifications of metabolites. A more comprehensive list of derivatizing reagents has been complied by Knapp [44].

### 5.5 A Case Study

In order to demonstrate the utility of some of the methods discussed earlier, metabolite profiling and identification of the compound in Fig. 19 is discussed [45]. In the actual study, this test compound was radiolabeled, and therefore metabolite profiling was performed using a radiochemical detector. Once metabolites of the parent compound were detected using a radiochemical detector, they were studied using LC/MS, NMR, and IR. In the absence of radioactivity, precursor ion scanning of selected fragments of the parent molecule would have determined which analytes were related to the parent compound. The following scenario can address metabolite identification of this compound without the use of radioactivity. First, the fragmentation pattern of the parent compound ($m/z$ 414) is determined by infusing it into the mass spectrometer and conducting an $MS^2$ experiment. In this case, the product ions were $m/z$ 68, 88, 116, 132, 185, 236, 283, 299, 300, and 386. Figure 19a shows a partial interpretation of these fragments. Ions at $m/z$ 299 and 116 represent valuable fragments because they are generated as a result of a fragmentation that divides the molecule into two complementary segments, A and B (Fig. 19b). Next, precursor ions to each of the diagnostic product ions ($m/z$ 299 and 116) are found in the biological matrices using precursor ion scans. In this case, a precursor ion scan of the product ion at $m/z$ 116 would yield two TIC signals, both with $m/z$ 414, consistent with that of the parent molecule and an isomer. This indicates that the only molecules in the mixture capable of generating fragments at $m/z$ 116 (segment A) after CID are the parent molecule and its isomer. In contrast, a precursor ion scan of the product ion at $m/z$ 299 would yield seven signals in the precursor ion TIC. This indicates that there are seven chromatographically distinct compounds in the mixture that yield a fragment of $m/z$ 299. This experiment would also provide the

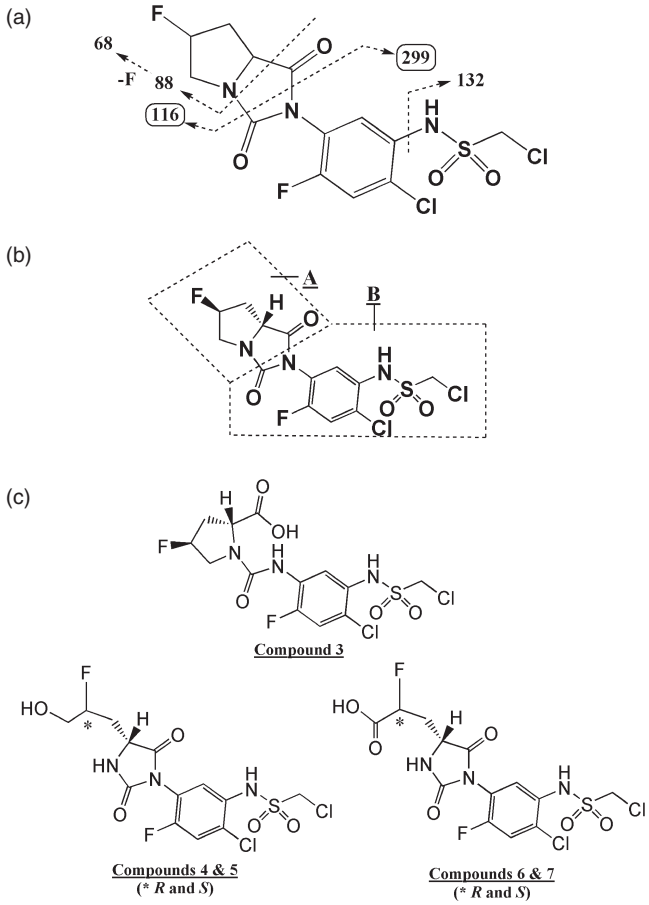


**FIGURE 19** A case study. (a) The fragmentation pattern of the molecule of interest in a triple quadrupole mass spectrometer. (b) The 116/299 fragmentation results in two diagnostic segments in the molecule. (c) Structures of metabolites 3–7.

*m*/*z* associated with each TIC signal. These would be *m*/*z* 414 (unmetabolized parent and its isomer), 432 (three of the TIC signals), and 446 (two of the TIC signals). From this we know that there are two compounds with *m*/*z* consistent with that of the parent ($MH^+$, *m*/*z* 414), three metabolites with *m*/*z* 432 ($MH^+$ + 18), and two metabolites with *m*/*z* 446 ($MH^+$ + 32).

The next step would be to conduct product ion scans of each of these ions (432 and 446) to obtain fragmentational information on each one. The product ion scans showed that the ions with $MH^+$ at *m*/*z* 414 have fragmentation patterns identical to that of the parent molecule. The retention time of the first compound confirmed its identity as the unmetabolized parent compound. Stereochemical considerations (Fig. 19) indicated the second compound to be a metabolite resulting from an epimerization of the parent to its diastereomer (later confirmed by use of a synthetic standard). The third compound had an $MH^+$ (*m*/*z* 432 or 414 + 18) consistent with a hydrolysis product of the parent. In this metabolite a molecule of water was determined to be added to the hydantoin ring because the product ions of this metabolite

included *m*/*z* 88 and 134. This indicates that while the fluoropyrrolidine moiety (*m*/*z* 88) was still intact, as was the case in the parent molecule, the hydantoin ring system had undergone hydrolysis leading to a ring opening (*m*/*z* 134 or 116 + 18). The structure was confirmed with a synthetic standard at a later time. Compounds 4 and 5 also exhibited *m*/*z* 434. The product ion scans of both these metabolites were qualitatively identical to each other and different from that of compound 3. The product ion spectra for compounds 4 and 5 showed that segment B was still intact. Therefore a net 18 atomic mass units (amu) had been added to segment A. Also, a CID ion of *m*/*z* 414 consistent with the loss of $H_2O$ ($MH^+$ – 18) was observed with both these metabolites, which was indicative of the presence of an alcohol group on these metabolites. However, due to inadequacy of the LC/MS data, the precise position of oxidation on segment A could not be determined using an $MS^2$ experiment on a quadrupole mass spectrometer. The final structural assignments were based on NMR and IR, as discussed later. Although this was not attempted at the time, this problem could potentially be resolved using an ion trap mass spectrometer by isolating the ion at *m*/*z* 134 and conducting further CID experiments on that fragment to obtain secondary and maybe even tertiary diagnostic fragments to enable structural assignments.

Compounds 6 and 7 (*m*/*z* 446) also exhibited qualitatively identical mass spectra to each other. As was the case for compounds 4 and 5, both these metabolites lost water after CID. Addition of 32 amu to the parent molecule (or 14 amu to compounds 4 and 5) could have resulted from a net addition of water and a carbonyl group to the parent molecule. However, as for compounds 4 and 5, the precise position of oxidation on segment A could not be determined using an $MS^2$ experiment and the final structural assignments were based on NMR and IR.

Compounds 4–7 were isolated (20–50 μg) using HPLC for detailed IR and NMR analysis. The IR analysis started by obtaining IR signals on the parent molecule. This work established the diagnostic signals. The IR absorption bands at 1726 and 1785 $cm^{-1}$ were attributed to the hydantoin ring, because five-membered cyclic imide rings generally have two absorption bands in the carbonyl region. Additionally, the ratio of intensities for these absorption bands (A1726/A1785) was 6.6, which was consistent with that of five-membered rings (six-membered rings have a ratio of ~2). These observations were not made for compound 3, where the hydantoin ring was no longer intact. In compounds 4–7, the same hydantoin signals as in the parent compound were observed, indicating that the hydantoin rings were intact and that the site of biotransformation on all of these molecules was on the fluorine-containing pyrrolidine ring.

In compounds 4 and 5, $^1$H-NMR data supported that segment B of these molecules were intact as previously suggested. Two-dimensional $^1$H–$^1$H COSY experiments supported a contiguous coupling network of protons. This network extends from the hydantoin proton (ring junction) to the methylene protons on carbon alpha to the fluorine-bearing carbon and the proton geminal to the fluorine. This established that the bonds between carbons in this area were not cleaved. However, the diastereotopic relationship of the methylene protons adjacent to the pyrrolidine nitrogen had been relieved, indicating cleavage of the C—N bond in the pyrrolidine ring and the presence of a hydroxyl group on that carbon. This established compounds 4 and 5 as diastereomeric hydroxyl metabolites, shown in Fig. 19c. For compounds 6 and 7, addition of 32 amu could have been due to (1) hydrolytic

cleavage of the hydantoin ring followed by oxidation of a carbon on the fluoropyrrolidine ring to a carbonyl; (2) addition of two oxygen atoms to the fluoropyrrolidine ring; or (3) oxidation of the carbon alpha to the fluorine-bearing carbon of a carboxylic acid. The IR data proved that the hydantoin ring was intact and ruled out the first possibility. As was the case for compounds 4 and 5, data from the two-dimensional $^{1}H$–$^{1}H$ COSY experiments supported a contiguous coupling network of protons from the hydantoin proton to the proton geminal to the fluorine atom. Further examination of the data revealed that the signals due to the methylene protons alpha to the nitrogen in the pyrrolidine ring were lacking. These, along with biosynthetic considerations, led to the conclusion that oxidation of the alcohol groups on compounds 4 and 5 to carboxylic acid had yielded compounds 6 and 7. IR data confirmed the presence of carboxylate moieties on compounds 6 and 7 as shown in Fig. 19c.

This study serves as one example of the logical steps involved in structural profiling and identification of a compound. There are numerous such examples in the literature with variations in approach. Readers are encouraged to study these publications to develop a better understanding of this field.

## 6 CONCLUSION

Metabolite profiling and identification is a crucial step in discovery and development of drugs. This activity assists medicinal chemists in designing more metabolically stable and safer drugs, pharmacologists in uncovering active metabolites, and toxicologists in describing potential causative agents for adverse effects. Scientists working in the area of metabolite profiling and identification should keep up with new advances in analytical technologies such as mass spectrometry and NMR. However, it is also imperative that they are well informed on the basic principles of anatomy, physiology, pharmacokinetics, enzymology, and chemistry. This allows for a more comprehensive interpretation of data for their colleagues in medicinal chemistry, pharmacology, and toxicology groups.

## ACKNOWLEDGMENT

I would like to extend my sincere gratitude to Ms. Yang Tang, Ms. Samantha Richardson, and Dr. Michael Shirley for their timely assistance in reviewing this chapter.

## DEDICATION

With the hope for more extensive discovery of pharmaceutical drugs to alleviate human suffering, I would like to dedicate this chapter to the everlasting memory of my beloved grandmother, Mrs. Kobra Malayeri, whose rich and happy life came to an end in a tragic, painful, and lengthy struggle with cancer on October 13, 2004. She will not be forgotten.

## REFERENCES


1. Palani A, Shapiro S, Josien H, Bara T, Clader JW, Greenlee WJ, Cox K, Strizki JM, Baroudy BM. Synthesis, SAR, and biological evaluation of oximino-piperidino-piperidine amides. 1. Orally bioavailable CCR5 receptor antagonists with potent anti-HIV activity. *J Med Chem* 2002;45:3143–3160.
2. Fura A, Su Y-Z, Shu M, Hanson RL, Roongta V, Humphreys WG. Discovering drugs through biological transformation: role of pharmacologically active metabolites in drug discovery. *J Med Chem* 2004;47:1–13.
3. Gad SC. Active drug metabolites in drug development. *Curr Opin Pharmacol* 2003;3:98–100.
4. Baillie TA, Cayen MN, Fouda H, Gerson RJ, Green JD, Grossman SJ, Grossman SJ, Klunk LJ, LeBlanc B, Perkins DG, Shipley LA. Metabolites in safety testing. *Toxicol Appl Pharmacol* 2002;182:188–196.
5. Hastings KL, El-Hage J, Jacobs A, Leighton J, Morse D, Osterberg R. Drug metabolites in safety testing, Letter to the Editor. *Toxicol Appl Pharmocol* 2003;190:91–92.
6. Thompson CD, Barthen MT, Hopper DW, Miller TA, Quigg M. Quantification in patient urine samples of felbamate and three metabolites: acid carbamate and two mercapturic acids. *Epilepsia* 1999;40:769–776.
7. Sladek NE, Doeden D, Powers JF, Krivit W. Plasma concentrations of 4-hydroxycyclophosphamide and phosphoramide mustard in patients repeatedly given high doses of cyclophosphamide in preparation for bone marrow transplantation. *Cancer Treat Rep* 1984;68:1247–1254.
8. Manyike PT, Kharasch ED, Kalhorm TF, Slattery JT. Contribution of CYP2E1 and CYP3A to acetaminophen reactive metabolite formation. *Clin Pharmacol Ther* 2000;67:275–282.
9. Kaplowitz N. Idiosyncratic drug hepatotoxicity. *Nat Rev* 2005;4:489–499.
10. Smith DA, Obach RS. Seeing through the MIST: abundance versus percentage. Commentary on metabolites in safety testing. *Drug Metab Dispos* 2005;33:1409–1417.
11. Lau YY, Sapidov E, Cui X, White RE, Cheng K-C. Development of a novel *in vitro* model to predict hepatic clearance using fresh, cryopreserved, and sandwich-cultured hepatocytes. *Drug Metab Dispos* 2002;30:1446–1454.
12. Gebhardt R, Hengstler JG, Muller D, Glockner R, Buenning P, Laube B, Schmelzer E, Ullrich M, Utesch D, Hewitt N, Ringel M, Hilz BR, Bader A, Langsch A, Koose T, Burger H-J, Maas J, Oesch F. New hepatocyte *in vitro* systems for drug metabolism: metabolic capacity and recommendations for application in basic research and drug development, standard operation procedures. *Drug Metabol Rev* 2003;35:145–213.
13. Guyton AC, Hall JE. *Textbook of Medical Physiology*, 10th ed. Philadelphia: Saunders; 2000.
14. Davies B, Morris T. Physiological parameters in laboratory animals and humans. *Pharm Res* 1993;10:1093–1095.
15. Ritschel WA, Kearns GL. *Handbook of Basic Pharmacokinetics*, 5th ed. Washinton DC: American Pharmaceutical Association; 1999.
16. Zhao CL, Uetrecht JP. Metabolism of ticlopidine by activated neutrophils: implications for ticlopidine-induced agranulocytosis. *Drug Metab Dispos* 2000;28:726–730.
17. Bertz RJ, Granneman GR. Use of *in vitro* and *in vivo* data to estimate the likelihood of metabolic pharmacokinetic interactions. *Clin Pharmacokinet* 1997;32:210–258.

18. Daly AK, Cholerton S, Armstrong M, Idle JR. Genotyping for polymorphisms in xenobiotic metabolism as a predictor of disease susceptibility. *Environ Health Prespect* 1994;102:55–61.
19. Dalvie DK, Kalgutkar AS, Khojasteh-Bakht SC, Obach RS, O'Donnell JP. Biotransformation reaction of five-membered aromatic heterocylic rings. *Chem Res Toxicol* 2002;15:269–299.
20. Kalgutkar AS, Gardner I, Obach RS, Schaffer CL, Callegari E, Henne KR, Mutlib AE, Dalvie DK, Lee JS, Nakai Y, O'Donnell JP, Boer J, Harriman SP. A comprehensive listing of bioactivation pathways of organic functional groups. *Curr Drug Metab* 2005;6: 161–225.
21. Moghaddam MF, Grant DF, Cheek JM, Greene JF, Williamson KC, Hammock BD. Bioactivation of leukotoxins to their toxic diols by epoxide hydrolase. *Nat Med* 1997;3:562–566.
22. Keseru GM, Kolossvary I, Szekely I. Inhibitors of cytochrome P450 catalyzed insecticide metabolism: a rational approach. *Int J Quantum Chem* 1999;73:123–135.
23. Bolton JL, Trush MA, Penning TM, Dryhurst G, Monks TJ. Roles of quinones in toxicology. *Chem Res Toxicol* 2000;13:135–160.
24. Aithal GP, Ramsay L, Daly AK, Sonchit N, Leathart JBS, Alexander G, Kenna JG, Caldwell J, Day CP. Hepatic adducts, circulating antibodies, and cytokine polymorphisms in patients with diclofenac hepatotoxicity. *Hepatology* 2004;39:1430–1440.
25. Smith PC, McDonagh AF, Benet LZ. Irreversible binding of zomepirac to plasma protein *in vitro* and *in vivo*. *J Clin Invest* 1986;77:934–939.
26. McDonagh AF, Palma LA, Lauff JJ, Wu TW. Origin of mammalian biliprotein and rearrangement of bilirubin glucuronides *in vivo* in the rat. *J Clin Invest* 1984;74:763–770.
27. Bunn HF, Gabbay KH, Gallop PM. The glycosylation of hemoglobin: relevance to diabetes mellitus. *Science (Wash DC)* 1987;200:21–27.
28. Koenig RJ, Blobstein SH, Cerami A. Structure of carbohydrate of hemoglobin $A_{ic}$. *J Biol Chem* 1977;252:2992–2997.
29. Higgins PJ, Bunn HF. Kinetic analysis of the nonenzymatic glycosylation of hemoglobin. *J Biol Chem* 1981;256:5204–5208.
30. Garlick RL, Mazar JS. The principal site of non-enzymatic glycosylation of human serum albumin *in vivo*. *J Biol Chem* 1983;258:6142–6146.
31. Shaklai NR, Garlick RL, Bunn HR. Nonenzymatic glycosylation of albumin alters its conformation and function. *J Biol Chem* 1984;259:3812–3817.
32. Evans DC, Watt AP, Nicoll-Griffith DA, Baillie TA. Drug–protein adducts: an industry perspective on minimizing the potential for drug bioactivation in drug discovery and development. *Chem Res Toxicol* 2004;17:3–16.
33. Bauer E, McDougall J, Cameron BD. The *trans–cis* isomerization of *trans*-4′-(2-hydroxy-3,5-dibromo-benzylamino)cyclohexanol *in vivo* and *in vitro* in different species. *Xenobiotica* 1986;7:625–633.
34. Xia Y-Q, Miller JD, Bakhtiar R, Franklin RB, Liu DQ. Use of a quadrupole linear ion trap mass spectrometer in metabolite identification and bioanalysis. *Rapid Commun Mass Spectrom* 2003;17:1137–1145.
35. Wrona M, Mauriala T, Bateman KP, Mortishire-Smith RJ, O'Connor D. "All-in-one" analysis for metabolite identification using liquid chromatography/hybrid quadrupole time-of-flight mass spectrometry with collision energy switching. *Rapid Commun Mass Spectrom* 2005;19:2597–2602.
36. Jemal M, Ouyang Z, Zhao W, Zhu M, Wells WW. A strategy for metabolite identification using triple-quadrupole mass spectrometry with enhanced resolution and accurate mass capability. *Rapid Commun Mass Spectrom* 2003;17:2732–2740.

37. Peterman MS, Duczak N, Kalgutkar AS, Lame ME, Soglia JR. Application of a linear ion trap/orbitrap mass spectrometer in metabolite characterization studies: examination of the human liver microsomal metabolism of the non-tricyclic anti-depressant nefazodone using data-dependent accurate mass measurements. *J Am Soc Mass Spectrom* 2006;17: 363–375.
38. Corcoran O, Spraul M. LC-NMR-MS in drug discovery. *Drug Discov Today* 2003;8:624–631.
39. Silverstein RM, Bassler GC, Morrill TC. *Spectrometric Identification of Organic Compounds*, 4th ed. Hoboken, NJ: Wiley; 1981.
40. Borlak J, Walles M, Elend M, Thum T, Preiss A, Levsen K. Verapamil: identification of novel metabolites in cultures of primary human hepatocytes and human urine by LC-$MS^n$ and LC-NMR. *Xenobiotica* 2003;33:655–676.
41. Nassar A-EF, Parmentier Y, Martinet M, Lee DY. Liquid chromatography–accurate radioisotope counting and microplate scintillation counter technologies in drug metabolism studies. *J Chromatogr Sci* 2004;42:348–353.
42. Shirley MA, Bennani YL, Boehm MF, Breau AP, Pathirana C, Ulm EH. Oxidative and reductive metabolism of 9-*cis*-retinoic acid in the rat. Identification of 13,14-dihydro-9-*cis*-retinoic acid and its taurine conjugate. *Drug Metab Dispos* 1996;24:293–302.
43. Kulanthaivel P, Barbuch RJ, Davidson RS, Yi P, Rener GA, Mattiuz EL, Hadden CE, Goodwin LA, Ehlhardt WJ. Selective reduction of N-oxides to amines: application of drug metabolism. *Drug Metab Dispos* 2004;32:966–972.
44. Knapp DR. *Handbook of Analytical Derivatization Reactions*. Hoboken, NJ: Wiley; 1979.
45. Moghaddam MF, Brown A, Budevska BO, Lam Z, Payne WG. Biotransformation, excretion kinetics, and tissue distribution of an *N*-pyrrolo[1,2-*c*]imidazolylphenyl sulfonamide in rats. *Drug Metab Dispos* 2001;29:1162–1170.

# LINKAGE BETWEEN TOXICOLOGY OF DRUGS AND METABOLISM

Ruiwen Zhang and Elizabeth R. Rayburn
*University of Alabama at Birmingham, Birmingham, Alabama*

**Contents**

1 Introduction
2 Drug Metabolism and Its Role in Toxicology
  2.1 General Principles of Drug Metabolism
  2.2 Metabolism and Metabolites in Toxicology
  2.3 Examples for Pharmacogenetics and Pharmacogenomics of Therapeutic Drugs
  2.4 Examples of the Link Between Toxicology of Drugs and Metabolism
3 Methods Used in Linking Drug Metabolism and Toxicology
  3.1 General Principles
  3.2 *In Vivo* Models
  3.3 *In Vitro* Models
  3.4 Human Cells *In Vitro*
  3.5 *In Silico* Models
  3.6 Predictive Models
4 Toxicity/Metabolism in Drug Development
  4.1 Toxicity, Toxicogenomics, and Metabolism
  4.2 "Omics," Toxicology, and Risk Assessment
  4.3 Future Directions
Acknowledgments
References

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

## 1 INTRODUCTION

Modern drug therapy relies on better understanding of molecular biology and genetics, the systems biology of diseases, and gene–environment and gene–drug interactions (Fig. 1a). In the last two decades, progress has been made in identifying, cloning, sequencing, and characterizing pathogenic genes (foreign and endogenous) important to disease onset and development, leading to novel approaches to drug target validation and to drug discovery and development (Fig. 1a). The fast-growing knowledge and technologies of genomics, proteomics, and systems biology are revolutionizing the process of drug discovery and development. For example, increasing numbers of molecular targets (DNA, RNA, and proteins), high-throughput screening technologies, and high-yield chemical syntheses have resulted in ever-increasing numbers of candidate compounds that are available for pharmacology and toxicology testing. In addition, genetic-based therapeutics, such as gene therapy and RNA/DNA-based therapeutics (aptamer, antisense, ribozyme, RNAi, miRNA, etc.), have been challenging the traditional paradigm of pharmacology and toxicology studies of new drugs [1–3].

Although our ability to identify candidate lead compounds with high potency and specificity has largely been improved, many candidate therapeutic agents are prevented from reaching the market due to inappropriate pharmacokinetic/ADME (absorption, distribution, metabolism, and excretion) properties and drug-induced toxicity; both are highly affected by many factors including pathogens, host genes, and drug–environment interactions (Fig. 1b). It is desirable that poorly behaved

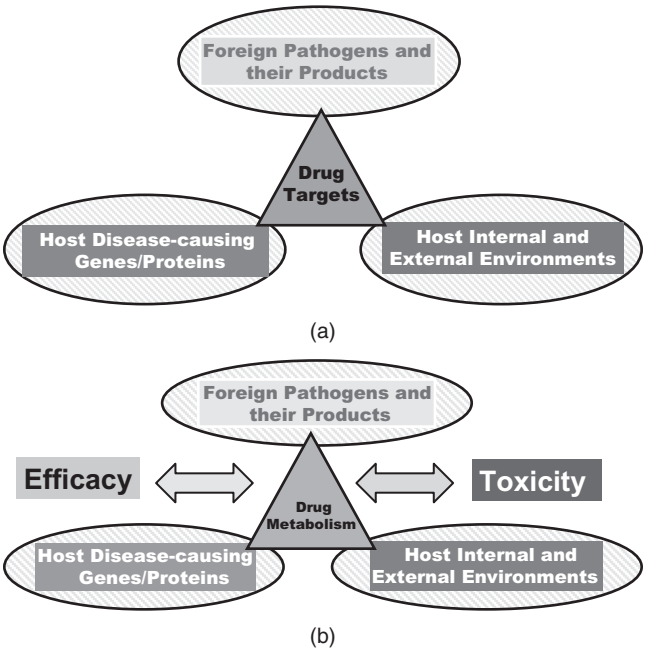


**FIGURE 1** (a) Drug may target various subjects. (b) Drug metabolism may be affected by various factors.

compounds are removed early in the discovery and early development phases rather than during the more costly late development phases, especially during clinical development. This strategy has been termed "failing early and fast." As a consequence, over the past decades, ADME and toxicity (ADMET) screening studies have been incorporated earlier into the drug discovery phase. The intent of this chapter is to introduce the fundamental ADMET concepts, key tools, reagents, and experimental approaches for predicting human pharmacokinetics and assessing "drug-like" molecules, with an emphasis on the link between toxicology and metabolism (Figs. 1b and 2a).

Pharmacokinetic (PK) studies, which evaluate the way in which a drug interacts with various barriers within a system, are an integral part of preclinical and clinical drug testing. A typical PK study may be divided into two sections, with respect to the fate of a given drug in the body: *absorption*, the movement of a drug into the bloodstream, and *disposition*, which involves the distribution, metabolism, and excretion of a drug. Preclinical PK studies often provide qualitative and/or quantitative measurements of these events, using various parameters and models to describe the speed, extent, and underlying mechanisms for a chemical entity to enter and remain in various systems/tissues of interest, generally including the plasma and blood cells, body fluids, various host normal tissues such as brain, heart, lungs, liver, kidneys, spleen, bone marrow, and skin, and drug-targeted tissues such as the tumor and infected tissues [4–6]. The related field of pharmacodynamics, which evaluates the effects of a drug and its metabolites on the test system under various physiological and pathological conditions, is often evaluated in tandem with pharmacokinetics. Pharmacodynamic (PD) studies are often carried out with a disease model that mimics the intended human diseases for the test drug and employ various biomarkers to link the biological effects of a drug with its use by various routes of administration under various conditions.

Also of equal importance for drug design are toxicokinetic studies. Toxicokinetics, as the name suggests, provides information on the behaviors and fate of a toxicant/drug in the test systems, with relation to the occurrence and time course of toxic events. While PK and PD studies provide information relative to drug distribution, metabolism, removal, and pharmacological activity, toxicokinetic studies are used to predict the behavior and safety of a compound/drug [7]. Knowing the distribution and especially the speed and extent of metabolism of a drug is helpful in determining its future efficacy and toxicity. Moreover, PK/PD and toxicity/safety studies are a routine part of human clinical trials, and the objectives of preclinical PK/PD and toxicology studies are to predict potential problems earlier in the drug development process. The results of these studies can be extrapolated to give at least a general idea of the likely effects of a particular compound in humans [8]. While still imperfect, numerous *in vitro*, *in vivo*, and *in silico* systems have been developed to evaluate the pharmacokinetics, pharmacodynamics, and toxicity of compounds prior to their use in humans. Of note, there is a trend to move toxicity testing to earlier phases of drug development.

Although typical pharmacology and toxicology studies can give an idea of the distribution and effects of a drug or compound with high predictive value for most cases, unexpected differences in the response to a given drug or compound among various patients have given rise to yet other fields of study—namely, pharmacogenetics, pharmacogenomics, and toxicogenomics—that investigate the effects of genes

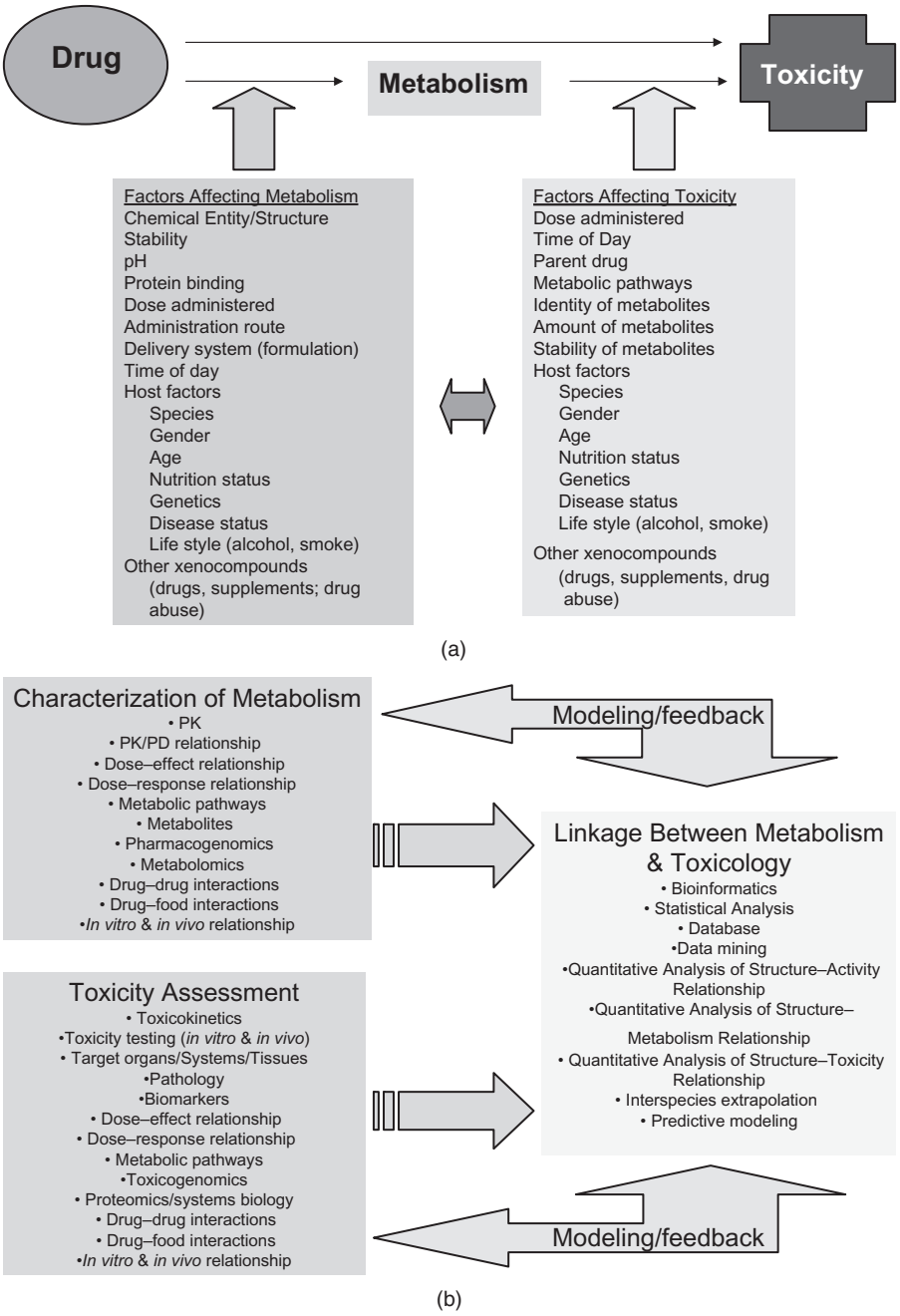


**FIGURE 2** (a) The linkage between drug metabolism and toxicity. Listed are factors affecting the metabolic and toxic responses to a drug. (b) The framework for studies aiming at the linkage between drug metabolism and toxicity.

of interest on the efficacy and toxicity of a drug or compound, utilizing various technologies in the fields of genetics, genomics, proteomics, and molecular biology [9, 10]. The differential genetic makeup of patients, including differences in the level of gene expression, polymorphisms, and gene deficiencies, can greatly alter the efficacy and toxicity of various therapies [9, 10]. Although many genes and their expression may be involved in various PK/PD and toxicology processes, the majority of pharamcogenomics/pharmacogenetics studies have been focused on genes related to drug metabolism. While differences in expression of the cytochrome P450 enzymes may be the most important factors that affect drug metabolism, there are ever-increasing numbers of various other genes that are shown to determine drug efficacy, metabolism, excretion, and toxicity (discussed in detail in Section 2). One of the well-documented cases is dihydropyrimidine dehydrogenase (DPD), a key metabolizing enzyme of 5-fluorouracil, one of the most clinically used anticancer agents. DPD deficiency has clearly been linked to severe 5-fluorouracil toxicity, including death. Pharmacogenetic and pharmacogenomic characteristics of DPD have been well defined by various experimental models, in the general population, and in cancer patients [11].

Another related area of study, metabolomics/metabonomics, has also been gaining recognition in recent years. This area investigates the time-dependent metabolic responses of organisms or physiological systems and the resulting metabolites of chemical entities, foods, and other compounds [12, 13]. The field is of particular interest for preclinical studies because it can be used to determine endpoints for further toxicity studies (clinical chemistry, histopathology, etc.) and to determine potential biomarkers [14–18]. It is also useful for screening purposes in order to eliminate agents with unfavorable profiles while still in early phases of development [19–21]. Because the metabolism of compounds differs based on a number of drug, patient/host, and environmental factors, the relatively new "-dynamics" and "-omics" areas are increasingly important for determining the efficacy and potential risks of new therapeutic agents. Figure 2a demonstrates the complexity of the interaction between these factors and their impact on drug metabolism, and thus toxicity.

Figures 1 and 2 provide a framework for concepts, scientific reasoning, and experimental design of studies aimed at linking drug metabolism and toxicity, the focal point of this chapter. As illustrated in Fig. 2b, three major components should be included in these studies: (1) the profile and mechanisms of drug metabolism; (2) the profile and mechanisms of drug toxicity; and (3) the methods and models for linking drug toxicity and metabolism. For drugs that are intended for human use, the major purpose of these preclinical studies should be to predict the potential risks to humans under both pathological and physiological conditions. While there have been significant improvements, there is still a need to develop more predictive models and better computational programs that can facilitate data analysis and extrapolation of the results from experimental models to human patients [22]. Recent progress made in bioinformatics and computational biology will certainly facilitate the fine-tuning of the modeling process and the development of models with better predictive values. The ability to predict the metabolism and toxicity of a compound and their relationship will both improve the results of human clinical trials and the likelihood of success of the new agents.

## 2 DRUG METABOLISM AND ITS ROLE IN TOXICOLOGY

### 2.1 General Principles of Drug Metabolism

As shown in Fig. 2a, the drug metabolism process relies on a number of factors. The human body encounters numerous foreign compounds (xenobiotics) during the course of an average day. While some compounds may be consciously taken in, such as food, vitamins, and pharmaceuticals, others, such as pollutants and bacteria with their toxins, are absorbed or consumed because of their ubiquitous presence in the environment.

All of these compounds need to be utilized or removed from the system, and this is accomplished by either metabolism or excretion. Here, we focus on metabolism, more specifically drug metabolism, which is the biotransformation of various drugs to (1) inactive compounds that can readily be excreted from the body or (2) more active compounds that can exert an effect (either positive or negative). Drug metabolism can result in four major outcomes:

1. A decrease in the biological activity of drugs and xenobiotics.
2. An increase in excretion due to the formation of a more water-soluble metabolite.
3. Conversion of a prodrug to an active drug, or conversion of an active drug to a different active compound.
4. Generation of toxic or carcinogenic metabolites.

Which of these processes occurs depends on a variety of factors. While the identity of the agent is the most important, its pharmacokinetics/pharmacodynamics within the human body is a close second. There are several factors that can affect the way that the body responds to xenobiotics.

- *Age.* Infants and children may not have a fully developed system to deal with xenobiotics and may then have different rates of drug metabolism and responses to the compound and toxicities. Older persons also may have differences in their ability to metabolize compounds due to changes that occur to the key metabolic systems, primarily the liver and kidneys.
- *Drug Interactions.* When several drugs are present in the body at the same time, they can cause adverse or unexpected side effects and can cause changes in metabolism. For example, combining two chemotherapeutic agents or a chemotherapeutic agent with an anti-inflammatory agent changes the pharmacokinetics of the chemotherapeutic agents [23, 24]. Common food items can also influence the metabolism of different compounds. It is important to note that there is an increased risk for drug–drug interactions in elderly persons and those who are hospitalized due to their higher frequency of polypharmacy. There is also growing concern over the increase in people who are now more actively self-medicating with over-the-counter agents.
- *Disease States.* Diseases that affect major excretory and metabolic organs (specifically those affecting the liver and kidneys) will change the rate of metabolism of a drug. Also, disease states that involve parenteral feeds can affect the

action and activity of a drug. Moreover, diseases that affect blood flow to various organs will also affect drug metabolism.

- *Occupation.* The occupation of a person in some cases can dictate metabolism. For example, pesticide applicators may have increased drug metabolism because liver enzymes can be induced by chronic exposure to pesticides. For example, exposure of male mice to piperonyl butoxide (PBO), a synergistic active ingredient in pesticides, leads to overexpression of cytochrome P450 genes, such as CYP1A1, CYP2A5, CYP2B9, and CYP2B10 [25].
- *Time of Day.* The expression and types of enzymes and cofactors expressed as well as the speed with which metabolism takes place varies with the host's circadian rhythm. For example, the expression of DPD and dThdPase varies with time in both human patients and experimental animals. This has implications for the metabolism of 5-fluorouracil and its toxicity [26, 27].
- *Genetics.* As mentioned in the Introduction, pharmacogenomics and toxicogenomics are now being studied to determine the metabolism and potential toxic reactions to drugs. Genetics and ethnicity are becoming increasingly recognized for their importance in drug metabolism. For example, one of the major drug metabolic pathways utilizes the cytochrome P450 system, which has several isoforms. One such isoform, CYP2D6, is responsible for metabolizing almost 20–30% [28] of all drugs, and there are about 51 [29] known allelic variants of this isoform.

  In Caucasian individuals with CYP2D6*4 or *5 alleles, 12–21% [29] have problems metabolizing drugs that are substrates for CYP2D6.

  Fifty-one percent [29] of the alleles in the Asian population are the CYP2D6*10 allele, which has a proline substitution for serine at position 34. This leads to the improper folding of the protein, which in turn affects its stability. Patients with this mutation have a reduced ability to metabolize some drugs, such as neuroleptics and antidepressants.

  The CYP2D6*17 allele is seen in about 20–35% [29] of black Africans.

  About 10–16% [29] of the Semitic population (people from Ethiopia, Saudi Arabia, and the Middle East) are ultrarapid metabolizers because they have 2–13 copies of the functional gene, resulting in a very high amount of the enzyme. Individuals with these mutations often do not respond well to neuroleptics and antidepressants because they metabolize the drugs too fast.

### 2.2 Metabolism and Metabolites in Toxicology

Drug metabolism involves various enzymatic pathways and transport systems that can convert xenobiotics to a form that can be more readily excreted. Although the enzymes needed for drug metabolism are present in almost every tissue, the rate at which compounds can be metabolized in a given tissue depends on the amounts and types of metabolic enzymes present in that tissue. Compared to other organs, the liver has very high concentrations of drug-metabolizing enzymes, which are also present in the GI tract, kidneys, lungs, and skin.

While the major function of drug metabolism is to inactivate and eliminate foreign compounds from the body, it can also generate reactive compounds that

**TABLE 1 Classification of Phase I and Phase II Reactions**

| Phase I | Phase II |
|---|---|
| Oxidation | Glucuronidation |
| Reduction | Sulfation |
| Hydrolysis | Acetylation |
| Hydration | Methylation |
| Dethioacetylation | Glutathione conjugation |
| Isomerization | Amino acid conjugation |
| | Fatty acid conjugation |
| | Condensation |

may be harmful. The generation of electrophiles is of particular concern as they can react with cellular components such as DNA and RNA and induce genetic mutations. It is essential to carefully evaluate the advantages and disadvantages of administration prior to the use of any drug. In addition to conventional testing methods ($LD_{50}$ and MTD in animals, and safety evaluations in preliminary human clinical trials), the newer fields of pharmacogenomics, toxicogenomics, and metabolomics are being used to assess the spectrum of metabolites generated and their potential toxicity prior to administration of an agent to human patients.

Drug metabolism generally occurs through two categories of enzymatic reactions: classified as either phase I or phase II reactions. Phase I reactions structurally alter the original compound and are often regarded as ones that prepare a xenobiotic for conjugation in phase II. Phase I reactions include oxidation, reduction, hydrolysis, hydroxylation, dethioacetylation, and isomerization reactions. Phase II reactions, which are catalyzed by enzymes found in hepatocytes, are usually conjugation reactions and involve the addition of a large bulky side group to a compound. In many cases, phase II reactions produce water-soluble and charged metabolites that have reduced biological activity and can readily be excreted in urine or bile. It is worthwhile to note that phase II reactions can precede phase I reactions. For example, the drug isoniazid undergoes phase II metabolism first and results in the formation of an *N*-acetyl conjugate, which can then undergo a phase I hydrolysis reaction [30]. Table 1 gives a summary of the phase I and phase II reactions. Figure 3 is a scheme that displays the fate of administered drugs.

### *Drug Metabolism: Effects on Efficacy and Toxicity of the Test Drug*

*Phase I Reactions* Phase I reactions are carried out by the mixed function oxidase (MFO) system. The most important MFO enzymes include the cytochrome P450s (CYP450). To date, the CYP450 MFO is regarded as the most powerful *in vivo* oxidizing system. This family of monooxygenases can be broken down into the liver microsomal MFO system and adrenal mitochondrial MFO system.

- The liver microsomal MFO system is made up of two components—NADPH cytochrome P450 reductase and cytochrome P450 (CYP450)—which are membrane-bound hemoproteins that use NADPH as a reducing agent and require molecular oxygen as well as FAD and FMN.

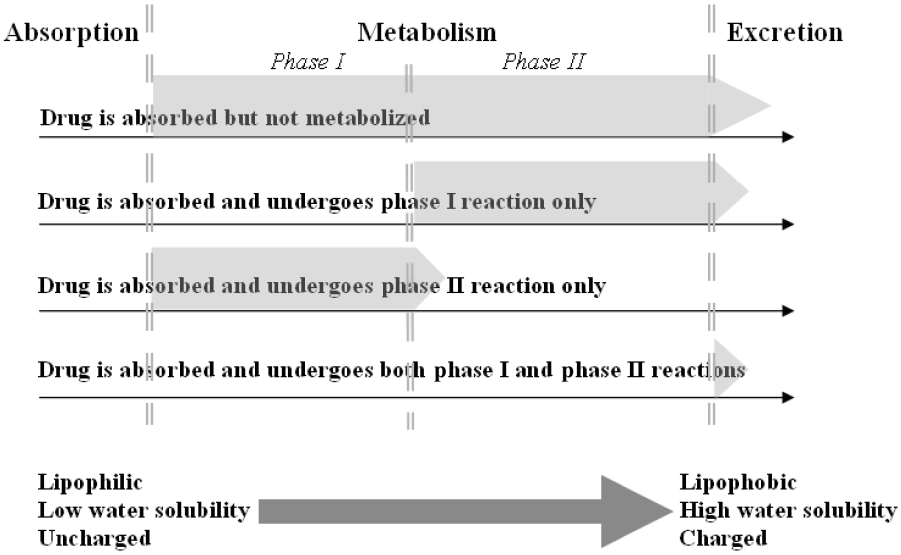


**FIGURE 3** The fate of administered drugs. Drugs can be absorbed and undergo different fates but ultimately they are all excreted in urine, bile, or feces. As drugs are metabolized, they have decreased lipophilicity, increased water solubility, and increased charge.

- Instead of CYP450 reductase, the adrenal mitochondrial MFO system is made up of an iron–sulfur protein, adrenodoxin, and an adrenodoxin reductase. These two proteins are involved in transferring electrons from NADPH to CYP450 in steroid biosynthesis and bile metabolism. As such, the adrenal MFO system is not a major player in drug metabolism. However, it is important to note that, in the future, it may be necessary to make use of this system for specific agents, particularly to evaluate and optimize drugs that mimic endogenous steroids.

The liver microsomal MFO system catalyzes numerous reactions by transferring electrons from NADPH to CYP450. It does so by activating the iron component of CYP450, enabling the insertion of an oxygen atom into a substrate (drug). A scheme depicting this reaction appears in Fig. 4.

NADPH–CYP450 reductase catalyzes nonspecific reactions, and its main role is the transfer of electrons to CYP450 through the use of FAD and FMN. Thus, specificity depends on the different isoforms of the CYP450s. In humans, there are 57 CYP genes and 33 CYP pseudogenes, which can be divided into families and subfamilies. Table 2 [35] shows the relative abundance of some of the isoforms and their importance in drug metabolism. Substrate specificity is often hard to assess because one enzyme may be involved in more than one metabolic pathway and one pathway of drug metabolism can involve several different enzymes.

Although the CYP450 catalytic cycle is very important in drug metabolism, it can also create some problems, such as autooxidation. In this process, if the substrate cannot acquire the electrons properly, the system can generate superoxide anions, hydrogen peroxide, or water. This process is of major concern as the CYP450 system has a success rate ranging from 1% to 100%. The production of free radicals can

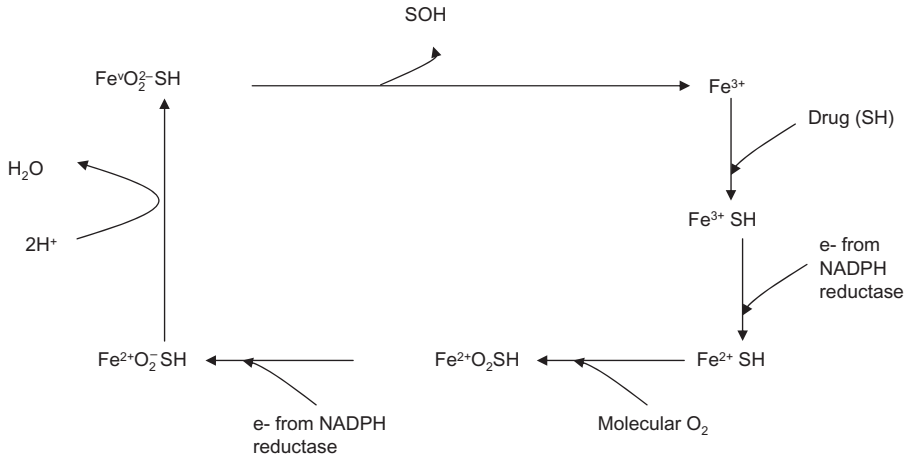


**FIGURE 4** Drug (here represented by SH) binds to the oxidized form of CYP450, which accepts one electron from NADPH reductase. This reduced form can bind molecular oxygen. Another electron is added and the molecule rearranges to generate the oxidized form of the drug (here SOH), water, and the oxidized form of CYP450.

**TABLE 2 Different Expression Levels of CYP450 and Their Contributions to Metabolism**

| Isoform | Percentage of Hepatic CYP450 Content | Isoforms | Percentage Contribution to Metabolism |
|---|---|---|---|
| 3A | 30 | 3A 4/5 | 36 |
| 2C | 18 | 2D6 | 19 |
| 1A2 | 13 | 2C 8/9 | 16 |
| 2E1 | 7 | 1A2 | 11 |
| 2A6 | 4 | 2C19 | 8 |
| 2D6 | 1.5 | 2E1 | 4 |
| 2B6 | 0.2 | 2A6 | 4 |
| Others | 26.3 | 2B6 | 3 |

result in toxicity, and determination of possible reactive oxygen and nitrogen species that can be generated from the drug is important for predicting its toxicity.

Some phase I reactions do not involve the CYP450 MFO system, although these mostly represent reactions of endogenous compounds. However, they can be participants in drug metabolism. Following is a list of these types of reactions.

1. *Oxidations.*

(a) Alcohol oxidation is performed by alcohol dehydrogenase.

(b) Aldehyde oxidation is performed by aldehyde dehydrogenase and aldehyde oxidase.

(c) Xanthine oxidase is concerned with the metabolism of xanthine-containing drugs such as caffeine and allopurinol.

(d) Monoamine oxidase (A and B forms) will oxidatively deaminate primary, secondary, and tertiary amines.

(e) Flavin-containing monooxygenases (FMOs) can oxidize nitrogen-, sulfur-, and phosphorus-containing drugs, by using NADPH and molecular oxygen, in a similar manner as MFOs. However, FMOs will not catalyze the oxidation of negatively charged compounds such as organic acids.

(f) Peroxidase-dependent cooxidation is important in extrahepatic tissues, especially those with low CYP450 levels and activity. In this reaction, peroxidases couple the reduction of hydrogen peroxide and tissue hydroperoxide to the oxidation of other compounds.

2. *Reductions.* Other processes catalyze the reduction of azo compounds, nitro compounds, epoxides, heterocyclic ring compounds, and halogenated hydrocarbons. Examples of such reactions include the reduction of protonsil to triaminobenzene and sulfanilamide, and the successive reductions of nitrobenzene to nitrosobenzene to phenylhydroxylamine to aniline.

3. *Hydrolysis.* The addition of $H_2O$ to compounds such as esters, amides, and carbamates by peptidases, amidases, and esterases results in the generation of new metabolites.

4. *Epoxide Hydrolase.* Epoxide hydrolases catalyze the hydration of epoxides. Since epoxides are known to be reactive species that can bind to DNA and RNA and thus induce mutations, epoxide hydrolase activity serves as a protective measure to decrease toxicity.

*Phase II Reactions* As explained previously, phase II reactions usually consist of the addition of a bulky charged group to phase I reaction products to facilitate excretion of drugs. Phase II reactions also involve several enzymes that are capable of catalyzing a wide variety of reactions. Table 3 lists phase II reactions, their enzymes, and the functional groups involved.

1. *Glucuronidation.* Glucuronidation involves the addition of UDP glucuronic acid to the substrate and is catalyzed by UDP-glucuronyltransferases (UGTs), which are microsomal membrane-dependent enzymes. It is the main phase II reaction, as UDP glucuronic acid can bind to O, N, C, or S atoms. While UGTs seem to have low specificity in drug metabolism, their specificity increases when dealing with

**TABLE 3 Summary of Phase II Reactions**

| Reaction | Enzyme | Functional Group |
|---|---|---|
| Glucuronidation | UDP-glucuronyltransferase | $-OH$, $-COOH$, $-NH_2$, $-SH$ |
| Glycosidation | UDP-glycosyltransferase | $-OH$, $-COOH$, $-SH$ |
| Sulfation | Sulfotransferase | $-NH_2$, $-SO_2NH_2$, $-OH$ |
| Methylation | Methyltransferase | $-OH$, $-NH_2$ |
| Acetylation | Acetyltransferase | $-NH_2$, $SO_2NH_2$, $-OH$, $-COOH$ |
| Amino acid, glutathione, and fatty acid conjugation and condensation | Glutathione-*S*-transferase | Epoxide, organic halide, $-OH$, various |

endogenous substrates. There are three subfamilies of UGT: UGT1A, UGT2A, and UGT2B, which contain the 15 known isoforms of UGT. While the liver is the main site of glucuronidation, other organs such as the stomach and kidneys have high levels of some UGT isoforms [31]. Drugs that undergo glucuronidation include morphine and codeine.

2. *Sulfation.* Sulfation plays an important role in drug metabolism, especially when phenols and phenol-containing compounds, alcohols, amines, thiols, and some steroids are present. Sulfotransferases (SULTs) catalyze the transfer of a sulfate moiety from 3′-phosphoadenosine 5′-phosphosulfate (PAPS, a cofactor) to the substrate. Since this sulfate moiety has a very low $pK_a$ (~2), sulfate conjugates will be charged at physiologic pH values, making them more water soluble and rapidly excreted. Similar to UGTs, SULTs are mostly found in the liver, as well as extrahepatic organs. It is important to note that the availability of PAPS is a rate-limiting condition of sulfation reactions, and that while there are both membrane-bound and membrane-free forms of SULTs, only the membrane-free forms are involved in drug metabolism. To date, 10 isoforms of cytosolic SULTs have been identified, and all of them have been reported to participate in the activation of procarcinogens. Unstable esters are frequently formed, which can spontaneously rearrange, producing reactive electrophiles capable of binding to DNA and RNA. Drugs that undergo sulfation include ethanol, acetaminophen, dopamine, and minoxidil (Rogaine™).

3. *Acetylation.* Aromatic amines and sulfonamides are the most common substrates for acetylation by *N*-acetyltransferase (NAT). During this process, the acetyl group from acetyl coenzyme A is transferred to the drug substrate, which contains a primary amine, a hydroxyl, or a sulfhydryl group. Unlike glucuronidation and sulfation, acetylation does not produce conjugates that have improved water solubility and excretion. Drugs that undergo acetylation include isoniazid, *p*-aminobenzoic acid, and hydralazine. NATs can be classified as NAT1 or NAT2 and each are regulated independently of the other. Point mutations in NAT2 have been associated with polymorphisms seen in the metabolism of isoniazid, which is often used to identify rapid and slow acetylators. For example, a point mutation at Ile114Thr is often seen in Caucasians but very rarely in Orientals. The latter population has a mutation at nucleotide 857 of NAT2, which results in lower level of active enzyme. Ninety percent of North Africans and 50% of Caucasians are slow acetylators, while 10–30% of the Korean and Japanese populations are slow acetylators [32]. Such genetic polymorphisms will affect the dose response to drugs and increase drug toxicities in slow acetylators.

4. *Methylation.* Drugs that contain O, N, and S atoms can be methylated by methyltransferases, which can be classified into four groups—catechol-*O*-methyltransferase (COMT), phenol-*O*-methyltransferase (POMT), thiopurine *S*-methyltransferase (TPMT), and thiol methyltransferase (TMT). Although the reactions have high substrate specificity, the methyl donor is *S*-adenosyl-methionine (SAM), irrespective of which enzyme is used. Methylation, like acetylation, does not necessarily result in increased water solubility or excretion. Substrates that undergo methylation include histamine, *L*-dopa, and 6-mercaptopurine.

5. *Amino Acid Conjugation.* Drugs that contain a carboxyl group can be conjugated with amino acids in a two-step reaction that requires acyl CoA ligase and amino acid acyltransferase. One of the most important amino acid conjugations is

in bile metabolism. Amino acids such as glycine, glutamine, or taurine (nonprotein amino acid) can be conjugated with their substrates to form products that can be readily excreted in urine [33]. One example of an amino acid conjugation substrate is salicylic acid.

6. *Glutathione Conjugation.* Glutathione *S*-transferase (GST) catalyzes the conjugation reaction of some drugs with glutathione, which is a tripeptide in which glutamate is bonded to cysteine through the γ-COOH group of glutamate. The other amino acid present in this molecule is glycine. As glutathione (GSH) is a nucleophile, it can bind reactive electrophiles present in the body and thus plays an important protective role. While GSH can spontaneously react with reactive electrophiles, GST can increase the rate of this reaction, as it converts GSH into its active anionic form. Once GSH conjugates have been formed, they undergo other reactions through peptidase and acetylase activities, to mainly produce cysteine and mercapturate conjugates, which can then be excreted in urine. More than 20 GSTs have been discovered and they can be classified as either cytosolic or microsomal [34].

The cytosolic forms are more important for drug metabolism and can participate in conjugation, reduction, and isomerization reactions, while the microsomal forms interact with endogenous substrates. Although concentrations of GSH and GST are relatively high in cells, depletion of GSH can occur, leading to toxicity. For example, high doses of acetaminophen are toxic because the drug covalently binds to GSH and effectively depletes GSH from the system. Acetaminophen bioprocessing usually involves glucuronidation and sulfation reactions, but in high doses, these systems are overloaded and CYP450 starts to process the drug. This produces the toxic metabolite *N*-acetyl-*p*-benzoquinone imine (NAPQI) [35], which can attack cellular components. In most cases, NAPQI is eliminated by GSH, but in the case of overdose, there can be a decreased amount of GSH and thus hepatocellular necrosis and death can occur.

Similar to SULTs, NATs, and P450 enzymes, there are also GST polymorphisms. GST has five members (mu, alpha, pi, theta, and sigma). The mu (GSTM1*0) and theta (GSTT1*0) genotypes express a null phenotype, and in individuals with the GSTM1*0 allele (50% of the Caucasian population, ~60% in Chinese and Korean populations), it may indicate a genetic predisposition to lung, colon, and bladder cancers [36–38].

While GSH is usually protective, it has been shown that GST can lead to increased drug resistance to chemotherapeutic agents, possibly through the inhibition of MAP kinase activity and apoptosis [34], which are both mechanisms that are important for the action of many anticancer therapies. TLK199 is a GSH analogue that can produce a GSH inhibitor, TLK117, and thus decrease the effects of GSH on anticancer agents [34].

***Drug Metabolism: Effects on Efficacy and Toxicity of Other Drugs*** When more than one medication is taken by a patient, there is always a risk that the patient may suffer side effects that are due to the interaction between the different drugs. Drugs may compete for binding sites on plasma proteins or transporters, or can alter the activity of the metabolic enzymes (most notably the CYPs). For example, ketoconazole, used as an antifungal drug, is a potent inhibitor of several CYPs including CYP3A4. When drugs that use CYP3A4, such as anti-HIV viral protease inhibitors, are coadministered with ketoconazole, they cannot be metabolized and their plasma

concentration increases, leading to toxicity [39, 40]. Some agents can also induce CYPs in a way to increase their own metabolism or to affect the metabolism of other drugs present in the body. For example, the herbal medicine St. John's wort will induce CYP3A4, leading to increased metabolism of other drugs such as oral contraceptives. The toxicity and efficacy of a drug can also be influenced by diet. For example, when the antihistamine terfenadine was taken with grapefruit juice, it led to arrhythmias. This was due to the fact that grapefruit juice contains furanocoumarins, which are inhibitors of CYP3A4, and terfenadine is a prodrug that requires this specific CYP for bioactivation. Terfenadine levels in the plasma would rise and cause arrhythmias [41].

***Toxicokinetics, Toxicodynamics, and Toxicogenomics*** In order to predict these types of toxicities, the past few decades have witnessed the exponential expansion of the boundaries of scientific knowledge. Core discoveries, such as DNA and its structure, as well as technical milestones including genetic engineering and initial quantification of cellular processes and their pathways, have been made. Such findings have triggered the evolution of drug evaluation and toxicology from its traditional aspects to one that now includes genetics. These discoveries have led to the development of toxicogenomics, pharmacogenomics, and better ways to study pharmacodynamics and toxicodynamics [42, 43].

In light of the discovery of the multiple enzyme isoforms that are present in metabolic pathways, such as the CYP450s, it is important to consider how genetic diversity plays a role in the metabolism and side effects of xenobiotics. The development of new techniques in the field of toxicogenomics is proving very useful in the elucidation of the genetic components of diseases. For example, by using nucleic acid microarrays, protein chips, and denaturing HPLC (DHPLC), one can identify pathways and mechanisms that are affected by administration of different agents. These methods allow the comparison between multiple nucleic acid sequences and their variants concurrently [42, 43]. This is particularly useful as toxicity is now being recognized as the result of multiple gene interactions, rather than a largely uncharacterized response within the entire body or within certain organ systems [42].

This new trend necessitates the investigation of genetic variability and its relation to phenotypic observations, as well as its importance in drug metabolism and subsequent toxicity. It is clear that toxicogenetics and pharmacogenetics will now play an integral part in drug development and are likely to be increasingly important in the future. Moreover, the use of these strategies by various methods can predict the patient response to treatment, thus helping physicians to select the most appropriate treatment options for individual patients that will improve the efficacy and minimize toxicity of the therapy [42–44].

### 2.3 Examples for Pharmacogenetics and Pharmacogenomics of Therapeutic Drugs

***Example 1: 5-Fluorouracil*** 5-Fluorouracil (5-FU), an antimetabolite of pyrimidine, is commonly used in the treatment of head, neck, breast, ovarian, and gastrointestinal tract cancers [45–47]. However, it is associated with toxicities ranging from stomatitis, diarrhea, skin changes, myelosuppression, and neurological com-

plications [45]. While these symptoms may be observed at standard doses of 5-FU, they have also been shown to be dependent on the dose and schedule of drug administration. A genetic basis to these toxic effects has also been discovered and is thought to be the main cause of the difference in toxicities between patients. The enzyme dihydropyrimidine dehydrogenase is responsible for the catabolism of 5-FU, converting it (about 85% of administered drug) [48] to its inactive and less toxic metabolite, dihydrofluorouracil, and increasing its excretion. 5-FU can also undergo an anabolic pathway, which accounts for its host and tumor toxicity. In patients, the aim is to achieve a balance between the two pathways to limit the amount of toxicity to normal tissues through the catabolic pathway, while not reducing the effects of its anabolic metabolites in tumor cells. However, this balance can be affected in patients who have a partial or complete deficiency of DPD [47]. The dihydropyrimidine dehydrogenase gene (*DPYD*) is known to have 39 alleles [49], and a point mutation at the DPYD*2A allele has been linked to deficiency in DPD [47, 50–52]. It is clear that the alteration in 5-FU metabolism can lead to toxicity to the host. The main factor involved in its toxicity is genetics, emphasizing the need for more accurate risk and safety assessments based on individual patient profiles.

***Example 2: Thiopurine Methyltransferase*** Azathioprine and 6-mercaptopurine (6-MP) are immunosuppressant drugs that are used to treat and prevent reoccurrence of diseases such as acute lymphoblastic leukemia (ALL), inflammatory bowel disease (IBD), ulcerative colitis, rheumatoid arthritis, and organ rejection in transplant patients [53, 54]. The metabolism of these two drugs requires thiopurine methyltransferase (TPMT), which is a cytosolic enzyme that can catalytically methylate these drugs. TPMT has been found to have several alleles that can be used to classify the general population as normal/high, intermediate, and deficient methylators. In Caucasians, this accounts for 90%, 10%, and 0.3–0.6% of the population, respectively [47, 54]. Three main alleles have been linked to deficient methylators, namely, TPMP*20, TPMP*21, and TPMP*22 [54], and these result in a reduction in enzyme activity. Other alleles are also known to be involved in the deficiency (TPMT*2, *3A, *3B, *3C) [47]. This can lead to toxic accumulation of metabolites such as azathioprine and 6-MP along the thiopurine metabolic pathway. Hence, similar to Example 1, patient genetics greatly influence the toxicity of administered drugs.

### 2.4 Examples of the Link Between Toxicology of Drugs and Metabolism

Drug metabolism is vital to the bioactivation, inactivation, and eventual clearance of drugs, but it also represents a major source of toxicity. There are many naturally toxic parent compounds; however, a large portion of therapeutic drugs exert acute or chronic toxic effects only following metabolism. During the drug discovery process, it is important to carefully perform risk and safety assessments of drugs, and to keep in mind issues such as dose-dependent toxicity, species differences with respect to drug action, metabolism and toxicity, and long- and short-term toxicity, as well as reproductive toxicity. In addition to these factors, there are numerous others that affect the potential toxicity of a compound following metabolism. Some of these are listed in Table 4.

**TABLE 4 Examples of Factors that Affect Metabolism and the Subsequent Toxicity of Drugs**

| Factor | Subclassification | Example |
|---|---|---|
| Identity of the compound | Innate toxicity of the parent compound | Vinca alkaloid compounds bind to and disrupt microtubules, resulting in cell cycle arrest [55]. |
| | Presence of toxic constituent atoms or structures | Methoxyflurane releases free inorganic fluorine in the blood, leading to nephrotoxicity [56]. |
| | Generation of free radicals | Many compounds lead to the generation of free radicals following metabolism (e.g., hydroquinones in cigarette tar lead to the production of free radicals and subsequent DNA damage) [57]. |
| | Metabolism/ activation | Prodrugs (e.g., capecitabine) result in different toxicity compared to their parent compound (5-FU) [56]. |
| Polypharmacy | Drug–drug interactions | Warfarin binds to albumin. When phenylbutazone is administered to the same patient, phenylbutazone competes for warfarin's binding site, displacing warfarin and causing the level of free warfarin in the plasma to increase [26]. |
| | Activation or repression of enzyme activity | Drugs can influence the activity of metabolic enzymes, including cytochrome P450s. For example, ketoconazole inhibits CYP3A4. Anti-HIV protease inhibitors are also metabolized by CYP3A4. When the enzyme's activity is repressed, they cannot be metabolized, and their plasma concentration increases [31]. |
| Host status | Genetics and ethnicity | There are 51 known allelic variants of CYP2D6<br>• 12–21% of Caucasians with CYP2D6*4 or CYP2D6*5 alleles have problems metabolizing 2D6 substrates.<br>• Within the Asian population, the CYP2D6*10 allele leads to improper 2D6 protein folding, reducing the ability to metabolize neuroleptics and antidepressants.<br>• About 10–16% of the Semitic population are ultra-rapid metabolizers because they have 2–13 copies of the functional gene [34]. |
| | Age | Elderly patients have different expression and activity of metabolic enzymes compared to adults, and typically metabolize drugs more slowly [58]. |
| | Concommittant disease | Various disease states can influence metabolism (e.g., cirrhosis, gastrointestinal disorders, hypertension) [59]. |
| Epidemiological factors | Diet | Consumption of grapefruit juice by patients taking the antihistamine terfenadine led to arrhythmias because of high levels of terfenadine in the plasma due to CYP3A4 inhibition by furanocoumarins in the juice [31]. |
| | Occupation | Many occupations can lead to exposure to activators/ repressors of metabolic enzymes (e.g., farmers are likely to experience increased exposure to pesticides) [27]. |

## 3 METHODS USED IN LINKING DRUG METABOLISM AND TOXICOLOGY

### 3.1 General Principles

As part of the regulatory process, toxicologists face the need for both qualitative and quantitative studies to assess hazardous risks versus benefits of substances that are potentially toxic. Qualitative studies provide the toxicologist with information about the possible hazardous effects of the toxicant via various exposure routes and at various doses. They also provide information about the mode of action, progression of toxic effects, and susceptibility of subjects to the agent [60–63]. To understand the biological factors that affect the dose–response curve, a quantitative relationship among the varying factors must be understood.

Modern toxicology has three key assumptions: (1) toxicity in humans can be predictively modeled in other organisms; (2) selection of an appropriate model is key to accurate prediction in humans; and (3) to understand the relevance to humans, you must understand the strengths and weaknesses of the model selected [60–64]. *In vivo*, *in vitro*, mathematical, and computer models (*in silico*) are all used to study the effects of metabolism on the toxicity of drugs and xenobiotic compounds. Table 5 shows the different methods used to assess the metabolism and toxicity of target drugs and their advantages and disadvantages.

**TABLE 5 A Summary of the Systems and Methods Used to Evaluate Metabolism and Toxicity, and Their Advantages and Disadvantages**

| System | Examples | Acquisition | Maintenance | Expense | Human Correlation |
|---|---|---|---|---|---|
| *In Vivo* | Nonhuman primates | Difficult | Difficult | High | High |
| | Rodents | Easy | Easy | Low | Moderate–high |
| | Other mammals | Moderate | Moderate | Moderate | Moderate–high |
| | Nonmammalian models | Easy–moderate | Easy–moderate | Low–moderate | Low–moderate |
| | Lower animals | Easy | Easy | Low | Low–moderate |
| *In Vitro* | Primary cells | Moderate | Moderate–difficult | Low | Moderate (tissue dependent) |
| | Cancer cells | Easy | Easy–moderate | Low | Moderate (cell and study dependent) |
| | Isolated organs | Moderate–difficult | Difficult | Moderate | Moderate (cell and study dependent) |
| | Microsomes | Easy–moderate | Easy–moderate | Low–moderate | Low–moderate |
| | P450 kits | Easy | Easy | Low–moderate | Low–moderate |
| *In Silico* | Scaling | Easy–moderate | Easy–moderate | Low | Moderate (depends on model) |
| | QSAR | Moderate–difficult | Moderate–difficult | Moderate–high | Moderate, but improving |
| | PBBK | Moderate–difficult | Moderate–difficult | Moderate–high | Low–moderate, but improving |
| | BMD | Easy–moderate | Easy–moderate | Moderate–high | Moderate |

### 3.2 *In Vivo* Models

*In vivo* modeling has long been the gold standard in toxicity testing, and extensive *in vivo* studies are still required prior to use of new agents in human clinical trials. *In vivo* models provide the ability to assess the actions or effects of particular drugs and their metabolites in an intact animal before human trials [60–64]. While other research models may provide useful information, only *in vivo* models provide the full picture of the pharmacokinetics, metabolism, and effects of the agent and its metabolites. In order to most effectively use animals for predicting human efficacy and toxicity, it is vitally important that the proper model is chosen.

***Nonhuman Primates*** Although they are the most closely related (genetically) to humans, nonhuman primates (most often chimpanzees, gibbons, macaques, and baboons) make up only a small percentage of the animals used for biomedical research. However, nonhuman primates are susceptible to many of the same diseases as humans, so primates are used for studies of surgery and diagnostics, as well as for studies of drug efficacy and toxicity of agents that cannot be accurately evaluated in other animals. The most common use for primates is to study new agents for the prevention and treatment of HIV/AIDS.

***Nonrodent Mammalian Models*** Many models are available for use in biomedical research, each serving to closely model different human systems. While not the most popular models, cats and dogs will continue to play a vital role in biomedical research and for toxicity and PK studies. Because of their similarity with human anatomy and physiology, dogs and cats make excellent models for some diseases and for evaluating toxicity to certain systems. Dogs are especially well suited to studies of compounds with potential cardiotoxicity. They are also frequently used to study renal toxicity of compounds due to their high incidence of kidney disease, and dogs have been a valuable source of preclinical data used to develop novel therapeutic agents [65]. Due to the similarity in anatomy and biochemistry of the cat brain to that of humans, cats have served as important models in the field of neurophysiology and synaptic response. Research using this model has also focused on leukemia, a disease shared by cats and humans, providing information important in understanding immunodeficiency and cancer [66]. Numerous other animals, including mini-pigs and ferrets, are used to study specific types of agents.

***Rodent Models*** Rodents are the most frequently used model for toxicology testing. Rodents have many advantages that make them valuable as an *in vivo* model system, including their small size, short gestation period, and ease of housing and care. Popular examples include hamsters, rats, and mice. Hamsters provide a model that is relatively free of spontaneous disease but is susceptible to several introduced pathogens. They also have a characteristic cheek pouch that provides an excellent site for the evaluation of proposed carcinogenic agents. Short life spans, rapid reproduction, and a virtual absence of spontaneous disease make the hamster a good model for reproductive toxicology. With gestation time of 15–18 days and a very regular estrus cycle of 4 days, the hamster has the shortest gestation time of all laboratory animals, giving it an advantage over other rodent models [67].

Rats are in the same subfamily as mice, Muridae. The majority of laboratory rats are domesticated varieties of the Norway rat. With their well-defined genetic makeup,

this extensively used model has been purpose-bred to be susceptible to specific human diseases, such as hypertension. A wide variety of transgenic and chemically induced rat disease models are also available for laboratory use. Additionally, rats have a ridge that covers the opening of the esophagus, making them unable to vomit. This enables rats to be used for extensive oral/ingestion toxicity studies. Rats are the most commonly used animal to test median lethal doses ($LD_{50}$).

Disadvantages to the use of rats compared to mice include their size (requiring larger amounts of test compounds compared to mice) and longer life span. Additionally, while the rat genome has been studied extensively, the mouse genome is more well defined. In addition, the rat's blood supply to the heart makes it a poor model for heart-related studies.

The mouse model is the most frequently used model in *in vivo* research. Mouse subspecies origins can be traced back to the beginning of human civilization when mice became commensal with human settlements [68]. Coat color mutations of mice have been noted and recorded for millennia, with widespread domestication of the mouse by the 1700s [68]. Haldane's report in 1915 began the genetic mapping of the mouse. Haldane described the linkage between pink-eye color and albino loci that were later defined as part of mouse chromosome 7. Knowledge of the mouse genome grew slowly over the next 50 years until the discovery of recombinant DNA technology. By 2002, 96% of the mouse genome had been sequenced, leading researchers to even further specialize the mouse model.

There are now a wide variety of mouse models available for evaluating the efficacy and toxicity of compounds for various diseases including cancer, diabetes, lupus, and infectious diseases. Even for just the study of cancer, there are thousands of possible models, including xenograft models (where human tumor cells are injected subcutaneously into immunocompromised animals), syngeneic models (where mouse cancer cells are implanted subcutaneously into normal or immunocompromised mice), transgenic models (with overexpression or targeted knockout of specific gene expression), and chemically induced models (where agents with proven carcinogenic activity are administered to animals in order to determine whether the formation of tumors can be slowed or prevented).

While rabbits are not technically rodents, they are often grouped within the same category. Rabbits have been used at length in biomedical testing as well as in consumer product safety testing. Rabbits are easy to handle, and their large ear veins make blood collection easily accessible. The rabbit, like the rat, does not have the ability to vomit. Extensive evaluations of atherosclerosis and potential therapies for atherosclerosis have been done in rabbits because studies have demonstrated that they develop lesions similar to human fatty streaks when fed a high fat diet [69]. The rabbit model is also often used in irritation testing that is required by law for certain consumer products. Rabbits are also used to evaluate pyrogenicity and muscle irritation by medical devices that are administered parenterally [64].

Guinea pigs were once a very popular subject for scientific research; however, they are used with much less frequency than other rodent models today. While in the past they were used for a variety of studies, including the development of a vaccines and antiviral agents, guinea pigs are now most commonly used to study sensitivity and hypersensitivity to agents for toxicity testing and for studying potential preventive and therapeutic approaches for juvenile diabetes and preeclampsia

[70, 71]. Mice and rats are now used for most of the studies that were previously done in guinea pigs.

***Nonmammalian Models*** While they are not currently used as a primary model for evaluating new therapeutic agents, fish (especially zebrafish, *Danio rerio*) represent an excellent model for studying embryonic development and exposure to environmental toxicants and are gaining in favor for evaluating pharmaceutical agents [72]. The zebrafish genome is well characterized, and many transgenic and genetically modified fish have been generated [73]. Moreover, it is possible to inject antisense oligonucleotides or siRNA into embryonic fish to generate fish with knocked down genes. Like rodents, the small fish are inexpensive, easy to breed, and relatively easy to maintain. However, given that they are not mammals, they do not represent an accurate translational model for all types of toxicity testing.

Other animals, including avian models (most frequently chickens and embryonic chicks) and frogs (*Xenopus laevis* and other *Xenopus* species) are also frequently used to study embryonic development and gene function [74, 75]. The genetics of these other animals are still being determined (although the *Xenopus* genome has been fairly well characterized), although animals with gene mutations or polymorphisms have been obtained for many species. However, because their genome is not as well characterized as that of other research animals, and because they do not possess some of the anatomical features of mammals, these models are less frequently used for drug development and evaluation.

Finally, lower organisms, including fruit flies (*Drosophila melanogaster*) and roundworms (*Caenorhabitis elegans*) are also being used to study gene function and could provide novel models for toxicity testing. The advantage of these models is that they have been extensively characterized, with the *C. elegans* genome being the first to be completely sequenced, and with all of its cells having been studied and mapped [76, 77]. Drosophila are also frequently subjected to gene manipulation to study gene function and development. Both of these lower organisms are easily bred and maintained in labs and, given their relative simplicity, can facilitate studies of reproductive and developmental toxicity. However, while they represent excellent models for some studies, the use of these models for drug design and evaluation studies is limited.

Despite their many advantages, the use of animals is generally costly, can be controversial, and is not necessarily correlative with human results. Other *in vitro* and computer modeling systems are being used to evaluate the toxicity of compounds and are especially useful for preliminary studies so that fewer animals will be required.

### 3.3 *In Vitro* Models

In addition to the above disadvantages, there are increasing numbers of compounds that need to be screened. Therefore, faster and less expensive models of toxicity evaluation have been established. The use of *in vitro* models for the screening of cytotoxic drugs began in the 1940s with the development of nitrogen mustard derivatives. With the development of chemically defined media in the 1950s, qualitative evaluations in cells grown as monolayers in Petri dishes became easier, and by using protein dyes, dose–response relationships could be determined [78].

Compounds affect cells directly in *in vitro* systems until medium containing compound is removed, allowing observation of direct toxic effects [79]. The advances in *in vitro* methods have allowed for elucidation of many of the mechanisms underlying clinical toxicity. Moreover, in many cases there is a high degree of correlation between the results obtained from animal studies and cell culture studies, providing an alternative to animal models. This allows for both high throughput screening and decreased use of laboratory animals.

The use of *in vitro* models is now generally well accepted as a means of prescreening for toxicity. In early stages, *in vitro* models can assist drug design via permeability (barrier penetration) assays and structure–activity relationships. The preliminary *in vitro* models often are used to decide the fate of compounds and whether they will be selected to continue in the development pipeline.

*In vitro* toxicity testing can involve everything from bacteria or yeast, to isolated organ preparations or embryos, to cultured human or animal cells. Isolated organ preparations were among the first systems used for *in vitro* toxicity studies. While the data from isolated organ studies may be the most similar to the *in vivo* situation, the nature of these types of studies (require animals, often laborious surgical procedures, and there is often large variation between samples) largely negates their usefulness. Similarly, the use of embryos is labor intensive and not practical for studying anything other than developmental toxicity. Rather, cultured cells (immortalized or isolated and cultured from primary tissue sources) now are the most commonly used model for *in vitro* studies. For example, hepatocytes can be used to evaluate hepatotoxicity, and other cell types (such as endothelial cells or colon cells) can be used to determine whether a drug is likely to enter the bloodstream, to be absorbed by oral administration, or to cross the blood–brain barrier.

### 3.4 Human Cells *In Vitro*

***Human Tissues In Vitro*** Special consideration must be given to the use of human tissue or primary human cells in *in vitro* modeling. *In vitro* modeling using human tissue could offer valuable preclinical research data and possibly increased safety, but there are still numerous legal and ethical issues involved in the use of human tissues [80, 81]. Improvements in cell culture techniques, along with increasing availability of human tissues, have led to increased use of human primary and immortalized cells in toxicology and pharmacodynamic studies [82]. Human *in vitro* models are being used with physiological models to predict quantitative metabolism, transport, clearance, and pharmacodynamic outcome [82]. The use of human cells may be able to tease apart differences between human responses and animal responses, including differential expression and activity of genes involved in detoxification or in converting drugs to toxic metabolites.

Various methods exist for evaluating the effects of compounds on cells, including methods to evaluate viability (trypan blue staining), proliferation (MTT or MTS assays, BrdUrd incorporation), apoptosis (Annexin-V staining/flow cytometry), and cell cycle progression (propidium iodide staining/flow cytometry) [5, 83]. Due to the ease of isolation of proteins, DNA, and RNA from cells, cell culture also lends itself to evaluation of effects on individual molecular targets. Of particular importance are gene arrays, which allow changes in numerous genes to be observed and quantitated together from a small sample. While the resulting array data are incredibly

complex, better statistical methods and data mining software are allowing for interpretation of the effects of compounds and metabolites on gene expression, and vice versa.

***Primary Cell Cultures*** Primary cell cultures are prepared directly from tissue removed from an organism during biopsy or necropsy. As a result, primary cultures have high variability in the types of cells that can grow from a given sample. The well-differentiated heterogeneous nature of primary cell culture allows for it to retain many complex biochemical properties of the tissue of origin [64]. However, cell viability is variable and depends on factors such as damage to the cell membrane, cell isolation (anoikis), and the nutrients, growth factors, and hormones available in the media. While heterogeneous cultures are sometimes desirable, more often, single cell types (e.g., B lymphocytes) are required. Many cells can be extracted or purified from heterogeneous samples using gating procedures that eliminate or allow for the passage of cells with expression of certain surface proteins [64]. Other cells can be selected for by the type of container and medium in which they are grown. As a result of individual genetics, the growth kinetics differ between cultures, making results obtained from primary cultures difficult to reproduce. Other complications encountered with primary cell lines are their finite life span in culture and the loss of specific cell functions and metabolic capabilities [64].

***Permanent and Immortalized Cell Lines*** Most *in vitro* systems are established with permanent cell lines, which will grow consistently in culture. Often derived from primary cell line subcultures, permanent cell lines will form continuous diploid or clonal lines that will grow consistently for extended periods of time. Of the population of diploid cells, at least 75% of the culture will contain the same complement genetic material as the host organism [64]. These cells can typically undergo approximately 40–50 divisions in culture.

Continuous cell lines often originate from tumor cells. Spontaneous or induced transformation (e.g., via infection with Epstein–Barr virus) of a primary cell line results in a cell line that is immortalized. Immortalized cells retain many, but not all, of the characteristics of the primary cells from which they are derived. Transformation can often result in genetic differences from the tissue of origin, including a heteroploid or aneuploid karotype [64]. However, most of the immortalized and tumor-derived cell lines are actually clonal. Clonal cell lines are formed from a genetically homogeneous subline through single cell isolation. Through reisolation and cloning, genetic drift is minimized and the homogeneity of the cell line can be maintained. Via selection media, desirable phenotypes or markers can be selected for. Transformed cells and established tumor cell lines can grow indefinitely and often can grow in suspension since the cells have lost their need for attachment [64].

***Cell-Free Systems*** Still other *in vitro* systems exist (in addition to cell cultures, and isolated organ or embryo cultures) for evaluating drug metabolism and the presence of potentially toxic metabolites. These typically involve incubation of the target drug with metabolic enzymes or tissues expressing the enzymes. Incubation of a drug with microsomes (isolated vesicles containing metabolic enzymes) can indicate the likely metabolites of the drug [84]. Microsome studies can also be used to assess

differences in metabolism of compounds by mutant enzymes or different enzyme isoforms or to assess whether certain compounds activate or repress the activity of the metabolic enzymes [85]. While whole hepatocytes can also be used, isolated microsomes allow for easier examination of the metabolism of a compound by specific enzymes and complexes. However, microsomes do not contain phase II enzymes [86], so results obtained from microsome studies should be considered carefully. In some cases, it may be preferable to use whole hepatocyte preparations, which contain both phase I and II enzymes [86].

To evaluate the effects on particular cytochrome P450 enzymes, or to determine which enzymes metabolize a drug, there are also kits available containing preparations of specific P450 enzymes (available from Human Biologics International, BD Biosciences, Sigma-Aldrich, and others). While not necessarily reflective of what occurs *in vivo*, these kits and microsome preparations are an excellent way to collect preliminary data on likely metabolites of new drugs.

### 3.5 *In Silico* Models

With the advances in technology in the last few decades, the development of *in silico* (computer-based) testing offers a vast array of toxicity testing requiring small sample sizes [87]. While traditional predictive scaling models utilize equations to determine starting and likely maximum doses for more advanced studies, newer *in silico* studies use computer simulations to model the interaction of various compounds with important receptors and to predict efficacy and toxicity. Computer-aided models focus on replacing animal models in toxicity testing through structure-based techniques, quantitative structure–activity relationship (QSAR) methodologies, and PBBK (physiologically based biokinetic) modeling [88]. Preliminary testing has been very promising. QSAR can be used for a wide range of toxicity endpoints including lethality [89–91]. The results of QSAR analysis continue to improve as more compounds are evaluated and can be compared. PBBK can be used to predict toxicity by simulating the kinetic behavior of the compound of interest using available data. Systems biology attempts to use the interaction of the complete system, that is, metabolic, regulatory, signaling, and transport processes. By understanding the interactions between small molecules and their targets, models should be able to better predict toxic effects of test compounds. Besides reducing the number of animals needed in toxicity testing, reliable *in silico* testing could also predict and prevent failure of a compound in later phases of development [90, 91].

Mathematical modeling of toxicity often involves the use of a dose–response model to define the benchmark dose (BMD) that will give a standard response. The user-friendly BMD software is readily available and is a good tool for establishing a reference point/point of departure within an observable dose range [92]. BMD can be applied with as few as two dose groups and one control group, and it makes use of all data points on a dose–response curve.

The major limitation of these alternative models is data reliability. The toxicity predicted through the use of computer simulations is only as accurate as the data used. However, as the state of the science improves through the resolution of dynamic protein structures, more defined metabolic pathways, and characterized genetic variations within the human population, it is likely that *in silico* technology will also improve.

Traditional toxicity testing involved large numbers of animals and it was often impossible to predict the type of toxicity that would result. Current toxicity testing in animals has been greatly streamlined, partly by the use of *in silico* and *in vitro* models to predict possible target organs and useful endpoints. It is likely that as our knowledge of genetics and pharmacogenetics improves, our use of animal models will be further streamlined, and results from the few animal studies that are done will allow more accurate prediction of human toxicity. While it is likely that the number of animals used can be further decreased, it is unlikely that it will ever be possible to completely avoid their use. No other research model can provide investigators with so much information or can reproduce the complex environment found within a living animal.

Fortunately, however, as more data is generated using these various *in vitro*, *in vivo,* and *in silico* models, it will become easier to develop an understanding of the SARs that exist among different drugs and classes of molecules. This will then facilitate the *in silico* studies, further eliminating the need for using a large number of animals. Combined, all of these technologies will help to ensure that unsafe and ineffective agents do not enter into human clinical trials, thus saving human lives, preserving human quality of life, and guarding the bottom line for the companies developing novel agents.

### 3.6 Predictive Models

Recently, significant progress has been made in predictive toxicology that, from a chemistry perspective, uses a structure–activity relationship (SAR) built on available test data (chemical characteristics, pharmacokinetics, pharmacodynamics, metabolism, toxicity, toxicokinetics, pharmacogenetics, pharmacogenomics, gene profiling, metabolomics, and bioassays) to predict the potential biological activity (e.g., toxicity ) of a chemical /drug based solely on its molecular structure and computed properties. Of note, the predictive value of any predictive toxicology approaches that are based on SAR modeling are entirely dependent on and limited by the availability, quality, and quantity of the reference data. In the field of linking metabolism and toxicology, the basic data for SAR modeling include, but are not limited to, the following:

- *In vitro* and *in vivo* toxicology studies
- Characteristics of the chemicals/drugs and their analogues
- *In vitro* and *in vivo* metabolic studies
- *In vivo* and *in vitro* pharmacology screening
- *In vitro* and *in vivo* pharmacology profiling
- Biomarkers/experimental endpoints
- Human accidental exposure studies

Among these data, the following characteristics are particularly important:

- Dose–effect relationship
- Dose–response relationship
- *In vitro*/*in vivo* relationship

- Therapeutic index/window
- Interspecies scaling

To predict biological activities from chemical structure, there are many computational approaches and programs available; interested readers are directed to several excellent recent reviews [93–96]. Such approaches have proven capabilities when applied to well-defined toxicity endpoints or well-documented metabolic pathways. However, these approaches are less suited to the challenges of global toxicity prediction, that is, for predicting the potential toxicity of structurally diverse chemicals across a wide range of endpoints of regulatory and pharmaceutical concern. In this field, there are several major trends improving the predictive values of the models:

- Large-scale data generation involving more fundamental, interdisciplinary technologies such as systems biology, genomics, proteomics, and metabolomics.
- Increasing public information resources to support these efforts.
- In-depth, well-focused mechanistic studies on chemical biology and toxicology using advanced molecular and genetic approaches (e.g., gene transfer, gene knockout, transgenic animal models) and analytical methods (e.g., real-time quantitative PCR, gene arrays, LC/MS, MS/MS).

Figure 5 illustrates a systems approach for SAR modeling to predict the linkage between metabolism and toxicology. The role of bioinformatics (Fig. 5) in toxicology

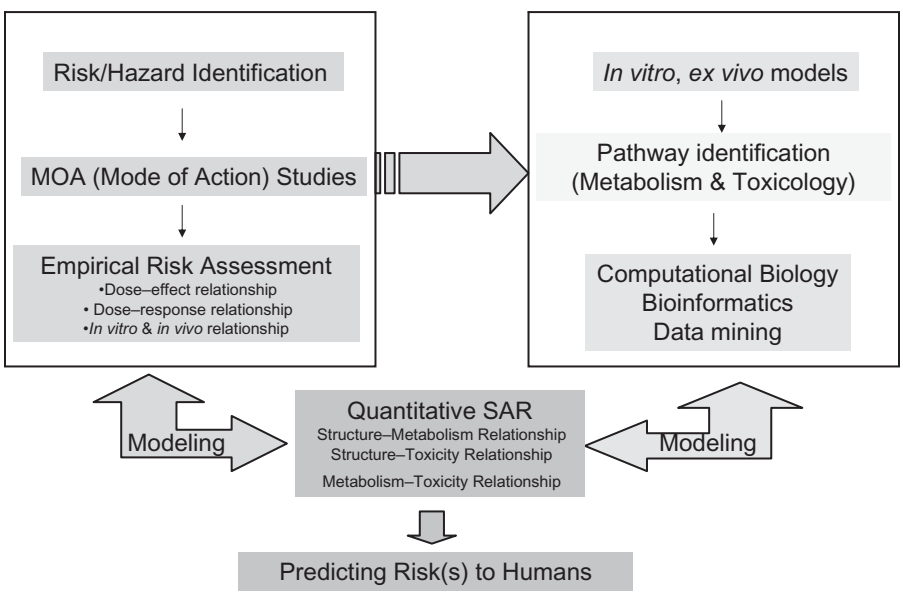


**FIGURE 5** Systems approaches to SAR modeling of the linkage between drug metabolism and toxicity.

data analysis and data mining is critical to SAR modeling, including the following: (1) increasing standardization and digitization of available toxicity data and migration of these data into the public domain; (2) establishing database standards and models that facilitate data integration and mining for both historical and new findings; and (3) linking these toxicology data with chemical structures, leading to hypothesis-driven SAR modeling. Of note, the predictive value of a SAR model should ultimately be confirmed by experimental or clinical data.

## 4 TOXICITY/METABOLISM IN DRUG DEVELOPMENT

### 4.1 Toxicity, Toxicogenomics, and Metabolism

One of the most important steps in preclinical drug development is the evaluation of the metabolism and toxicity of the compound. Many effective drugs fail during toxicity testing. As recent events in the news have shown, the toxicity of compounds is often difficult to predict using traditional toxicity testing. Even extensive animal studies do not give the full picture of toxicity. Moreover, while clinical trials give the most accurate prediction of the response of patients, there are still frequently problems when the new agent is used in a larger population. Depending on the risk–benefit ratio, some toxic agents are still useable, but major toxicity not only costs the pharmaceutical industry time and money that would have been better used for other drugs in the pipeline, but can also result in life-threatening toxicity for patients.

Even when the majority of a population does not experience any major toxicity from a compound, there are genetic variations to important metabolic enzymes (i.e., the cytochrome P450s) that can result in increased or decreased metabolism of a drug, leading to decreased efficacy or increased toxicity. Preclinical testing, including computer modeling, and especially cell culture will be important as more becomes known about the human genome. Pharmacogenomics and toxicogenomics (which study the effect of genes on the efficacy, metabolism, and toxicity of compounds) will likely be used to dictate which agents will continue in development, and whether there are subpopulations of patients who should be cautioned against the use of certain agents.

Although the P450 enzymes may be the most important and well-known metabolic enzymes, there are many others that can affect drug pharmacokinetics and dynamics and can result in otherwise unforeseen toxicity, including examples mentioned in Section 2 [50, 54].

While genetic polymorphisms may preclude the use of drugs for certain patients, prior knowledge of potential toxicity caused by these differences can save human lives and ensure that a therapeutic agent can remain in the clinic for other patients. As mentioned, certain mutations within the DPYD gene (which codes for dihydropyrimidine dehydrogenase) result in life-threatening toxicity for patients administered a routine dose of 5-FU. Demonstrating the advances in genetics, molecular biology, and instrumentation, a simple breath test using $^{13}$C-labeled uracil was developed that can detect patients who are likely to have DPYD mutations. These patients can be administered alternative therapies while their genotype is confirmed using other methods, such as DHPLC (denaturing HPLC) [97].

### 4.2 "Omics," Toxicology, and Risk Assessment

Being able to predict major toxicity, either for all patients or for certain subsets of patients, could result in saved lives and can lead to a more efficient drug development program. Pharmacogenomics and toxicogenomics can help predict the overall toxicity of compounds and can be used for screening certain subpopulations who may have metabolic differences that could lead to toxicity or inefficacy. In this way, clinical trials and subsequent large-scale use of new agents could be tailored to decrease toxicity and increase efficacy for patients.

Evaluating toxicogenomics and pharmacogenomics is becoming a trend among pharmaceutical companies. For example, Bristol-Myers Squibb started a Department of Applied Genomics in 2001 to evaluate pharmacogenomics and toxicogenomics. Most other major pharmaceutical companies are following suit. Although still in its infancy, pharmacogenomics and toxicogenomics programs have the potential to improve drug development programs by eliminating compounds earlier in the program, allowing more money to be invested in the development of other compounds.

The toxicogenomics "revolution" was sparked by the advent of microarray technology. These arrays utilize small samples of cDNA from cells or tissues to determine the effect of compounds on gene expression [98, 99]. By examining the types of genes that are up- or downregulated by a given compound, it is possible to predict whether a compound might be toxic, and to gain insight into the type of toxicity it causes (e.g., oxidative stress) [99].

Nevertheless, changes at the gene expression level are not necessarily sufficient for assessing the potential risks of new compounds. Often, while there are no changes in gene transcription, there are alterations in protein expression, stability, or activity level. Moreover, different pathways may be activated even while expression levels remain the same. Hence, other "omics" technologies are necessary to more effectively evaluate the effects of a given compound. Proteomics, typically evaluated using two-dimensional (2D) electrophoresis gels (where proteins are separated by charge and then by molecular weight) and mass spectroscopy, can be used to evaluate the expression level and, to a degree, the posttranslational modifications made to numerous proteins. Enzyme assays are usually required to determine the activity of proteins. The newer field of metabolomics (or metabonomics) evaluates the relationship between the genetic profile of an organism/cell and the complement of small molecules present in the plasma/serum or tissues/cells following administration of the agent. Combining all of these "omics" strategies can lead to a more global picture of the response of cells, tissues, or animals to new compounds, can define which genes are important for processing and reacting to the compound, and can be used to predict potential problems with efficacy, resistance, and toxicity.

### 4.3 Future Directions

One of the most exciting outcomes of toxicogenomics is the potential to predict the type of toxicity that may result from administration of a particular agent. Along with this is the ability to determine potential biomarkers that can be used to assess the risk for individual patients. In addition to the already known polymorphisms

(i.e., to the cytochrome P450 enzymes), gene and protein expression levels at baseline and following preliminary dosing (using a lower dose of the agent) can be used to determine whether a particular patient is likely to respond to the agent or is especially sensitive to the compound, potentially leading to toxicity. Biomarkers can also be used to detect whether a patient has been exposed to an environmental toxicant [100].

The "omics" studies also present several ethical issues. In addition to perhaps leading to a mandate for more individualized therapy, or at least a screening for polymorphisms in the most common and important metabolic enzymes, there are many other potential applications for "omics" data. For example, toxicogenomics data could be used to predict whether an individual would be more/less sensitive to toxic chemicals. These data could be used to determine which individuals would be allowed to engage in work where exposure could occur [101]. While this type of practice could prevent toxicity due to exposure to common chemicals, it would also constitute a new (and still legal in some states) type of employment discrimination. It may become necessary to pass federal legislation banning discrimination on the basis of genetic information.

On a more basic level, "omics" studies can provide insight into the complex reactions within living organisms. While none of the *in silico*, *in vitro*, or *in vivo* models are currently sufficient for determining the potential metabolites and toxicity of a compound, the addition of pharmacogenomics, toxicogenomics, and metabolomics may result in a more accurate picture of the toxicity of the compound in humans.

In conclusion, while advances have been made in the fields of toxicology, pharmacology, and molecular biology, there is still a lot of work yet to be done in order to make the drug development process faster, easier, and more effective. Study of the toxicity of both the parent compound and its metabolites can be greatly facilitated by examining the interplay between the genome of the experimental model or patient and the agent being administered. Moreover, when the full spectrum of metabolites is known, a more accurate picture of the toxicity and pharmacokinetics of the agent can be derived. With advances in our knowledge about disease processes, new technology, data processing, and statistical analysis, the evaluation of the toxicity of new agents will become faster and more accurate.

## ACKNOWLEDGMENTS

We would like to thank Dr. Donald Hill, Ms. Kajal Buckoreelall, Ms. Scharri Ezell, and Ms. Angela Venn for their excellent assistance in preparation of this chapter. The work is partially supported by NIH/NCI grants (R01CA80698, R01CA112029, and R01CA121211) and a DoD Prostate Cancer Research Program grant (W81XWH-06-1-0063).

## REFERENCES

1. Rayburn E, Wang H, He J, Zhang R. RNA silencing technologies in drug discovery and target validation. *Lett Drug Des Discov* 2005;2:1–18.
2. Que-Gewirth NS, Sullenger BA. Gene therapy progress and prospects: RNA aptamers. *Gene Ther* 2007;14:283–291.

3. Zhang Z, Wang H, Li M, Rayburn E, Agrawal S, Zhang R. Novel MDM2 p53-independent functions identified through RNA silencing technologies. *Ann NY Acad Sci* 2005;1058:205–214.
4. Walker DK. The use of pharmacokinetic and pharmacodynamic data in the assessment of drug safety in early drug development. *Br J Clin Pharmacol* 2004;58(6):601–608.
5. Wang H, Cai Q, Zeng X, Yu D, Agrawal S, Zhang R. Anti-tumor activity and pharmacokinetics of a mixed-backbone antisense oligonucleotide targeted to RIα subunit of protein kinase A after oral administration. *Proc Natl Acad Sci USA* 1999;96: 13989–13994.
6. Wang H, Rayburn E, Wang W, Kandimalla ER, Agrawal S, Zhang R. Immunomodulatory oligonucleotides as novel therapy for breast cancer: pharmacokinetics, *in vitro* and *in vivo* anti-cancer activity and potentiation of antibody therapy. *Mol Cancer Ther* 2006;5(8):2106–2114.
7. Welling PG. Differences between pharmacokinetics and toxicokinetics. *Toxicol Pathol* 1995;23(2):143–147.
8. Watanabe PG. Toxicokinetics in the evaluation of toxicity data. *Regul Toxicol Pharmacol* 1988;8(4):408–413.
9. Pander J, Gelderblom H, Guchelaar HJ. Insights into the role of heritable genetic variation in the pharmacokinetics and pharmacodynamics of anticancer drugs. *Expert Opin Pharmacother* 2007;8:1197–1210.
10. Goldstein DB, Need AC, Singh R, Sisodiya SM. Potential genetic causes of heterogeneity of treatment effects. Am *J Med* 2007;120(4 Suppl 1):S21–S25.
11. Mattison LK, Johnson MR, Diasio RB. A comparative analysis of translated dihydropyrimidine dehydrogenase cDNA; conservation of functional domains and relevance to genetic polymorphisms. *Pharmacogenetics* 2002;12:133–144.
12. Nicholson JK, Lindon JC, Holmes E. "Metabonomics": understanding the metabolic responses of living systems to pathophysiological stimuli via multivariate statistical analysis of biological NMR spectroscopic data. *Xenobiotica* 1999;29:1181–1189.
13. Thomas CE, Ganji G. Integration of genomic and metabonomic data in systems biology—are we "there" yet? *Curr Opin Drug Discov Dev* 2006;9:92–100.
14. Robertson DG, Reily MD, Baker JD. Metabonomics in preclinical drug development. *Expert Opin Drug Metab Toxicol* 2005;1:363–376.
15. Robertson DG, Reily MD, Sigler RE, Wells DF, Paterson DA, Braden TK. Metabonomics: evaluation of nuclear magnetic resonance (NMR) and pattern recognition technology for rapid *in vivo* screening of liver and kidney toxicants. *Toxicol Sci* 2000;57: 326–337.
16. Nicholls AW, Holmes E, Lindon JC, Shockcor JP, Farrant RD, Haselden JN, Damment SJ, Waterfield CJ, Nicholson JK. Metabonomic investigations into hydrazine toxicity in the rat. *Chem Res Toxicol* 2001;14:975–987.
17. Connor SC, Wu W, Sweatman BC, Manini J, Haselden JN, Crowther DJ, Waterfield CJ. Effects of feeding and body weight loss on the $^{1}$H-NMR-based urine metabolic profiles of male Wistar Han rats: implications for biomarker discovery. *Biomarkers* 2004;9: 156–179.
18. Lindon JC, Holmes E, Nicholson JK. Metabonomics in pharmaceutical R&D. *FEBS J* 2007;274:1140–1151.
19. Bailey NJ, Oven M, Holmes E, Nicholson JK, Zenk MH. Metabolomic analysis of the consequences of cadmium exposure in Silene cucubalus cell cultures via $^{1}$H NMR spectroscopy and chemometrics. *Phytochemistry* 2003;62:851–858.

20. Griffin JL. Metabonomics: NMR spectroscopy and pattern recognition analysis of body fluids and tissues for characterisation of xenobiotic toxicity and disease diagnosis. *Curr Opin Chem Biol* 2003;7:648–654.
21. Makinen VP, Soininen P, Forsblom C, Parkkonen M, Ingman P, Kaski K, Groop PH, Ala-Korpela M, FinnDiane Study Group. Diagnosing diabetic nephropathy by $^{1}$H NMR metabonomics of serum. *MAGMA* 2006;19:281–296.
22. Pelkonen O, Raunio H. *In vitro* screening of drug metabolism during drug development: Can we trust the predictions? *Expert Opin Drug Metab Toxicol* 2005;1:49–59.
23. Posey JA, Wang H, Hamilton J, Delgrosso A, Zhang R, Freda T, Zamboni WC. Phase-I dose escalation and sequencing study of docetaxel and continuous infusion topotecan in patients with advanced malignancies. *Cancer Chemother Pharmacol* 2005;56:182–188.
24. Wang H, Wang Y, Rayburn ER, Hill DL, Rinehart JJ, Zhang R. Dexamethasone as a chemosensitizer for breast cancer chemotherapy: potentiation of the antitumor activity of adriamycin, modulation of cytokine expression, and pharmacokinetics. *Int J Oncol* 2007;30:947–953.
25. Muguruma M, Nishimura J, Jin M, Kashida Y, Moto M, Takahashi M, Yokouchi Y, Mitsumori K. Molecular pathological analysis for determining the possible mechanism of piperonyl butoxide-induced hepatocarcinogenesis in mice. *Toxicology* 2006;228: 178–187.
26. Daher GC, Zhang R, Soong S-J, Diasio RB. Circadian variation of pyrimidine catabolizing enzymes in rat liver: possible relevance to 5-fluorodeoxyuridine chemotherapy. *Drug Metab Dispos* 1991;19:285–287.
27. Zhang R, Lu Z, Diasio R, Liu T, Soong S-J. The time of administration of 3′-azido-3′-deoxythymidine (AZT) determines its host toxicity with possible relevance to AZT chemotherapy. *Antimicrob Agents Chemother* 1993;37:1771–1776.
28. Ingelman-Sundberg M. Pharmacogenetics of cytochrome P450 and its applications in drug therapy: the past, present and future. *Trends Pharmacol Sci* 2004;25:193–200.
29. Eichelbaum M, Ingelman-Sundberg M, Evans WE. Pharmacogenomics and individualized drug therapy. *Annu Rev Med* 2006;57:119–137.
30. Sharer JE, Shipley LA, Vandenbranden MR, Binkley SN, Wrighton SA. Comparisons of phase I and phase II *in vitro* hepatic enzyme activities of human, dog, rhesus monkey, and cynomolgus monkey. *Drug Metab Dispos* 1995;23:1231–1241.
31. Fisher MB, Paine MF, Strelevitz TJ, et al. The role of hepatic and extrahepatic UDP-glucuronosyltransferases in human drug metabolism. *Drug Metab Rev* 2001;33:273–297.
32. Zang Y, Zhao S, Doll MA, States JC, Hein DW. The T341C (Ile114Thr) polymorphism of *N*-acetyltransferase 2 yields slow acetylator phenotype by enhanced protein degradation. *Pharmacogenetics* 2004;14:717–723.
33. Rendic S. Summary of information on human CYP enzymes: human P450 metabolism data. *Drug Metab Rev* 2002;34:83–448.
34. http://www.merck.com/mrkshared/mmanual/section22/chapter298/298e.jsp.
35. Hinson JA, Reid AB, McCullough SS, James LP. Acetaminophen-induced hepatotoxicity: role of metabolic activation, reactive oxygen/nitrogen species, and mitochondrial permeability transition. *Drug Metab Rev* 2004;36:805–822.
36. Townsend DM, Tew KD. The role of glutathione-*S*-transferase in anti-cancer drug resistance. *Oncogene* 2003;22:7369–7375.
37. Tukey RH, Strassburg CP. Human UDP-glucuronosyltransferases: metabolism, expression, and disease. *Annu Rev Pharmacol Toxicol* 2000;40:581–616.

38. Landi S. Mammalian class theta GST and differential susceptibility to carcinogens: a review. *Mutat Res* 2000;463:247–283.
39. Giacomini KM, Krauss RM, Roden DM, Eichelbaum M, Hayden MR, Nakamura Y. When good drugs go bad. *Nature* 2007;446:975–977.
40. Dresser GK, Spence JD, Bailey DG. Pharmacokinetic–pharmacodynamic consequences and clinical relevance of cytochrome P450 3A4 inhibition. *Clin Pharmacokinet* 2000;38:41–57.
41. Benton RE, Honig PK, Zamani K, Cantilena LR, Woosley RL. Grapefruit juice alters terfenadine pharmacokinetics, resulting in prolongation of repolarization on the electrocardiogram. *Clin Pharmacol Ther* 1996;59:383–388.
42. Aardema M, MacGregor J. Toxicology and genetic toxicology in the new era of "toxicogenomics": impacts of "-omics" technologies. *Mutat Res* 2002;499:13–25.
43. Waters M, Olden K, Tennant R. Toxicogenomic approach for assessing toxicant-related disease. *Mutat Res* 2003;544:415–424.
44. Thybaud V, Le Fevre AC, Boitier E. Application of toxicogenomics to genetic toxicology risk assessment. *Environ Mol Mutagen* 2007;48:369–379.
45. Diasio RB, Beavers TL, Carpenter JT. Familial deficiency of dihydropyrimidine dehydrogenase: biochemical basis for familial pyrimidinemia and severe 5-fluorouracil-induced toxicity *J Clin Invest* 1998;81:47–51.
46. Lu Z, Zhang R, Diasio RB. Purification and characterization of dihydropyrimidine dehydrogenase from human liver. *J Biol Chem* 1992;267:17102–17109.
47. Gardiner SJ, Begg EJ. Pharmacogenetics, drug-metabolizing enzymes, and clinical practice. *Pharmacol Rev* 2006;58:521–590.
48. Diasio RB, Johnson MR. Dihydropyrimidine dehydrogenase: its role in 5-fluorouracil, clinical toxicity and tumor resistance. *Clin Cancer Res Editorial* 1999;5:2672–2673.
49. Seck K, Riemer S, Kates R, Ullrich T, Lutz V, Harbeck N, Schmitt M, Kiechle M, Diasio R, Gross E. Analysis of the DPYD gene implicated in 5-fluorouracil catabolism in a cohort of Caucasian individuals. *Clin Cancer Res* 2005;11:5886–5892.
50. Lu Z, Zhang R, Diasio RB. Dihydropyrimidine dehydrogenase activity in human peripheral blood mononuclear cells and liver: population characteristics, newly identified deficient patients, and clinical implication in 5-fluorouracil chemotherapy. *Cancer Res* 1993;53:5433–5438.
51. Lu Z, Zhang R, Carpenter JT, Diasio RB. Decreased dihydropyrimidine dehydrogenase activity in population of patients with breast cancer: implication for 5-fluorouracil-based chemotherapy. *Clin Cancer Res* 1998;4:325–329.
52. Gardiner SJ, Begg EJ, Robinson RA. The effect of dihydropyrimidine dehydrogenase deficiency on outcomes with fluorouracil. *Adverse Drug React Toxicol Rev* 2002;21:1–16.
53. www.nlm.nih.gov/medlineplus/druginfo/uspdi/202077.html.
54. Schaeffeler E, Eichelbaum M, Reinisch W, Zanger UM, Schwab M. Three novel thiopurine *S*-methyltransferase allelic variants (*TPMT*20, *21, *22*)—association with decreased enzyme function. *Hum Mutat* 2006;27:976.
55. Jordan MA. Mechanism of action of antitumor drugs that interact with microtubules and tubulin. *Curr Med Chem Anti-Cancer Agents* 2002;2:1–17.
56. Pelusi J. Capecitabine versus 5-FU in metastatic colorectal cancer: considerations for treatment decision-making. *Community Oncol* 2006;3:19–27.
57. Pryor WA. Cigarette smoke radicals and the role of free radicals in chemical carcinogenicity. *Environ Health Perspect* 1997;105:875–882.

58. Hurria A, Lichtman SM. Pharmacokinetics of chemotherapy in the older patient. *Cancer Control* 2007;14:32–43.
59. Rivory LP, Slaviero KA, Clarke SJ. Hepatic cytochrome P450 3A drug metabolism is reduced in cancer patients who have an acute-phase response. *Br J Cancer* 2002;87:277–280.
60. Guengerich FP, MacDonald JS. Applying mechanisms of chemical toxicity to predict drug safety. *Chem Res Toxicol* 2007;20:344–369.
61. Mattes WB. Cross-species comparative toxicogenomics as an aid to safety assessment. Expert *Opin Drug Metab Toxicol* 2006;2:859–874.
62. Seed J, Carney EW, Corley RA, Crofton KM, DeSesso JM, Foster PM, Kavlock R, Kimmel G, Klaunig J, Meek ME, Preston RJ, Slikker W Jr, Tabacova S, Williams GM, Wiltse J, Zoeller RT, Fenner-Crisp P, Patton DE. Overview: using mode of action and life stage information to evaluate the human relevance of animal toxicity data. *Crit Rev Toxicol* 2005;35:664–672.
63. Lu Z, Zhang R, Diasio RB. Comparison of dihydropyrimidine dehydrogenase from human, rat, pig and cow liver: biochemical and immunological properties. *Biochem Pharm* 1993;46:945–952.
64. Gad SC. The mouse ear swelling test (MEST) in the 1990s. *Toxicology* 1994;93:33–46.
65. Niemeyer GP, Welch JA, Tillson M, Brawner W, Rynders P, Goodman S, Dufresne M, Dennis J, Lothrop CD Jr. Renal allograft tolerance in DLA-identical and haploidentical dogs after nonmyeloablative conditioning and transient immunosuppression with cyclosporine and mycophenolate mofetil. *Transplant Proc* 2005;37:4579–4586.
66. Rohn JL, Gwynn SR, Lauring AS, Linenberger ML, Overbaugh J. Viral genetic variation, AIDS, and the multistep nature of carcinogenesis: the feline leukemia virus model. *Leukemia* 1996;10:867–869.
67. Hendry WJ 3rd, Sheehan DM, Khan SA, May JV. Developing a laboratory animal model for perinatal endocrine disruption: the hamster chronicles. *Exp Biol Med* 2002;227:709–723.
68. Mouse Genome Sequencing Consortium, Waterston RH, Lindblad-Toh K, Birney E, et al. Initial sequencing and comparative analysis of the mouse genome. *Nature* 2002;420:520–562.
69. Yanni AE. The laboratory rabbit: an animal model of atherosclerosis research. *Lab Anim* 2004;38:246–256.
70. Gad SC. *Animal Models in Toxicology*, 2nd ed. Philadelphia: Taylor & Francis; 2007, pp 334–402.
71. Reid ME. *The Guinea Pig in Research*. Washington, DC: Human Factors Research Bureau; 1958, pp 62–70.
72. Rubinstein AL. Zebrafish assays for drug toxicity screening. *Expert Opin Drug Metab Toxicol* 2006;2:231–240.
73. http://zfin.org/cgi-bin/webdriver?MIval=aa-ZDB_home.apg.
74. Claudio L, Bearer CF, Wallinga D. Assessment of the US Environmental Protection Agency methods for identification of hazards to developing organisms, Part II: The developmental toxicity testing guideline. *Am J Ind Med* 1999;35:554–563.
75. http://www.xenbase.org/common/.
76. Watts DJ, Strogatz SH. Collective dynamics of "small-world" networks. *Nature* 1998;393: 440–442.
77. The *C. elegans* Sequencing Consortium. Genome sequence of the nematode *C. elegans*: a platform for investigating biology. *Science* 1998;282:2012–2018.

78. Dendy PP. The use of *in vitro* methods to predict tumour response to chemotherapy. *Br J Cancer Suppl* 1980;4:195–198.
79. Kikkawa R, Fujikawa M, Yamamoto T, Hamada Y, Yamada H, Horii I. *In vivo* hepatotoxicity study of rats in comparison with *in vitro* hepatotoxicity screening system. *J Toxicol Sci* 2006;31:23–34.
80. Thasler WE, Schlott T, Kalkuhl A, Plan T, Irrgang B, Jauch KW, Weiss TS. Human tissue for *in vitro* research as an alternative to animal experiments: a charitable "honest broker" model to fulfill ethical and legal regulations and to protect research participants. *Altern Lab Anim* 2006;34:387–392.
81. Meslin EM, Quaid KA. Ethical issues in the collection, storage, and research use of human biological materials. *J Lab Clin Med* 2004;144:229–234.
82. MacGregor JT, Collins JM, Sugiyama Y, Tyson CA, Dean J, Smith L, Andersen M, Curren RD, Houston JB, Kadlubar FF, Kedderis GL, Krishnan K, Li AP, Parchment RE, Thummel K, Tomaszewski JE, Ulrich R, Vickers AE, Wrighton SA. *In vitro* human tissue models in risk assessment: report of a consensus-building workshop. *Toxicol Sci* 2001;59:17–36.
83. Wang W, Zhao Y, Rayburn E, Hill D, Wang H, Zhang R. Anti-cancer activity and structure–activity relationships of natural products isolated from fruits of *Panax ginseng*. *Cancer Chemother Pharmacol* 2007;59:589–601.
84. Miners JO, Knights KM, Houston JB, Mackenzie PI. *In vitro–in vivo* correlation for drugs and other compounds eliminated by glucuronidation in humans: pitfalls and promises. *Biochem Pharmacol* 2006;71:1531–1539.
85. Suresh D, Srinivasan K. Influence of curcumin, capsaicin, and piperine on the rat liver drug-metabolizing enzyme system *in vivo* and in vitro. *Can J Physiol Pharmacol* 2006;84:1259–1265.
86. Gomez-Lechon MJ, Castell JV, Donato MT. HepaTocytes—the choice To investigate drug metabolism and toxicity in man: *in vitro* variability as a reflection of *in vivo*. *Chem Biol Interact* 2007;168:30–50.
87. Mayne JT, Ku WW, Kennedy SP. Informed toxicity assessment in drug discovery: systems-based toxicology. *Curr Opin Drug Discov Dev* 2006;9:75–83.
88. Gubbels-van Hal WM, Blaauboer BJ, Barentsen HM, Hoitink MA, Meerts IA, Van der Hoeven JC. An alternative approach for the safety evaluation of new and existing chemicals, an exercise in integrated testing. *Regul Toxicol Pharmacol* 2005;42:284–295.
89. Ekins S, Andreyev S, Ryabov A, Kirillov E, Rakhmatulin EA, Sorokin S, Bugrim A, Nikolskaya T. A combined approach to drug metabolism and toxicity assessment. *Drug Metab Dispos* 2005;34:495–503.
90. Ekins S, Nikolksy Y, Nikolskaya T. Techniques: application of systems biology to absorption, distribution, metabolism, excretion and toxicity. *Trends Pharmacol Sci* 2005;26:202–209.
91. Kruhlak NL, Contrera JF, Benz RD, Matthews EJ. Progress in QSAR toxicity screening of pharmaceutical impurities and other FDA regulated products. *Adv Drug Deliv Rev* 2007;59:43–55.
92. Barlow S, Renwick AG, Kleiner J, Bridges JW, Busk L, Dybing E, Edler L, Eisenbrand G, Fink-Gremmels J, Knaap A, Kroes R, Liem D, Muller DJ, Page S, Rolland V, Schlatter J, Tritscher A, Tueting W, Wurtzen G. Risk assessment of substances that are both genotoxic and carcinogenic report of an international conference organized by EFSA and WHO with support of ILSI Europe. *Food Chem Toxicol* 2006;44:1636–1650.
93. Duch W, Swaminathan K, Meller J. Artificial intelligence approaches for rational drug design and discovery. *Curr Pharm Des* 2007;13:1497–1508.

94. Norinder U, Bergstrom CA. Prediction of ADMET properties. *Chem Med Chem* 2006;1:920–937.
95. Sardari S, Dezfulian M. Cheminformatics in anti-infective agents discovery. *Mini Rev Med Chem* 2007;7:181–189.
96. Patzel V. *In silico* selection of active siRNA. *Drug Discov Today* 2007;3–4:139–148.
97. Mattison LK, Ezzeldin H, Carpenter M, Modak A, Johnson MR, Diasio RB. Rapid identification of dihydropyrimidine dehydrogenase deficiency by using a novel 2–13C-uracil breath test. *Clin Cancer Res* 2004;10:2652–2658.
98. Nuwaysir EF, Bittner M, Trent J, Barrett JC, Afshari CA. Microarrays and toxicology: the advent of toxicogenomics. *Mol Carcinog* 1999;24:153–159.
99. Pennie WD, Woodyatt NJ, Aldridge TC, Orphanides G. Application of genomics to the definition of the molecular basis for toxicity. *Toxicol Lett* 2001;31:353–358.
100. Sabbioni G, Sepai O, Norppa H, Yan H, Hirvonen A, Zheng Y, Jarventaus H, Back B, Brooks LR, Warren SH, Demarini DM, Liu YY. Comparison of biomarkers in workers exposed to 2,4,6-trinitrotoluene. *Biomarkers* 2007;12:21–37.
101. Weinstein M, Widenor M, Hecker S. Health and employment practices: ethical, legal, and social implications of advances in toxicogenomics. *AAOHN J* 2005;53:529–533.

# ALLOMETRIC SCALING

William L. Hayton[1] and Teh-Min Hu[2]
[1]*The Ohio State University, Columbus, Ohio*
[2]*National Defense Medical Center, Taipei, Taiwan*

**Contents**

1 Introduction
  1.1 Role of Allometric Scaling in Preclinical Development
  1.2 Historical Perspectives
2 Allometric Equation
  2.1 Mathematical Description, $Y = aM^b$
  2.2 Interpretation of $a$ and $b$
  2.3 Using the Allometric Equation
3 Scaling Clearance
  3.1 Physiological Basis and Theoretical Considerations
  3.2 Empirical Observations
  3.3 Recommended Approach
4 Scaling Volume of Distribution
  4.1 Physiological Basis and Theoretical Considerations
  4.2 Empirical Observations
  4.3 Limitations and Challenges
5 Scaling Half-Life
  5.1 Rationale and Theoretical Considerations
  5.2 Empirical Observations
  5.3 Limitations and Challenges
6 Conclusion
  References

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

# 1 INTRODUCTION

## 1.1 Role of Allometric Scaling in Preclinical Development

Allometric scaling is useful in the preclinical drug development phase to project pharmacokinetic parameter values determined experimentally in a few animal species to other species and ultimately to humans. Such projection permits early decisions about the developmental prospects for candidate drugs and forms a basis for selection of drug dosage to provide a desired level of systemic exposure. Projection of human pharmacokinetic parameter values from preclinical animal studies permits calculation of a first dose in humans that achieves a desired systemic exposure. Allometric scaling also plays a quality control function in that failure to observe an expected scaling relationship raises a concern about the validity of pharmacokinetic parameter values on hand or about possibly unusual interspecies variability. The purpose of this chapter is to present allometric scaling techniques in the context of the preclinical drug development process.

## 1.2 Historical Perspectives

***Allometry in Biology*** The term *allometry* was first coined by Huxley and Teissier [1] in 1936. The *allometric growth* can be described by an algebraic formula, $Y = aM^b$, in which the sizes ($Y$) of the parts of an organism follow a power law relationship with the overall body mass ($M$) [1]. While the original concept of allometry was limited to the size-dependent morphological or structural changes within or between species, it has been applied to other size-related attributes, such as biochemical and physiological processes. To date, allometry (or allometric scaling) is often used to denote any size-dependent power law relationship found in biology.

A topic pursued for more than a century is the allometric scaling of metabolic rate across species. Over 160 years ago, Sarrus and Rameaux first recognized the importance of size as a determinant of metabolic rate ($MR$) (cited by White and Seymour [2]). In 1883 the German scientist Max Rubner [3] described a proportional relationship between $MR$ and body surface area ($BSA$) in dogs. He then suggested that metabolic rate is proportional to BSA and therefore to body mass ($M$) to the power of two-thirds, that is, $MR \propto M^{2/3}$, which has been considered as the *surface law*. In 1932 Kleiber [4] concluded that basal metabolic rate was not proportional to surface area, but scaled with an exponent close to three-fourths. Kleiber's finding was later supported by Brody's [5] famous mouse-to-elephant curve. *Kleiber's law* has been generalized as the quarter-power law, which has been considered as ubiquitous and "universal" in biology [6].

***Allometry in Pharmacokinetics*** While some ponder theories of allometry, biomedical and pharmaceutical scientists face the practical question of how to extrapolate physiological, pharmacological, and toxicological data from animals to humans. In 1949 Adolph [7] listed 34 allometric equations that correlated morphological, physiological, and biochemical parameters with interspecies body mass. For example, creatinine clearance followed the allometric equation with a $b$ value of 0.69 [7]. His quantitative power law relationships suggested overwhelming similarities across

species, which formed a biological basis for extrapolation of experimental data from one animal species to another [8]. Adolph's work also had a significant impact on modern development of physiologically based pharmacokinetics (*PBPK*).

Animal scale-up of pharmacokinetics was actively pursued as early as the 1950s. Pinkel [9] proposed the use of body surface area (*BSA*) as a criterion of drug dosage in cancer chemotherapy, based on the observation of species invariant dose when expressed on a $mg/m^2$ basis for five antineoplastic agents. Freireich et al. [10] confirmed Pinkel's observation by evaluating the predictive relationship between toxic doses in five adult mammals with adult human data for 18 anticancer drugs. Furthermore, Mellett's [11] analysis of kinetic data for amethopterin and cyclophosphamide in six species, including humans, suggested that an allometric relationship applied for *AUC* and half-life ($t_{1/2}$).

The concept of scaling pharmacokinetics gained momentum in the early 1970s. Dedrick et al. [12, 13] assumed that pharmacokinetic processes correlated with internal physiological processes and demonstrated that drug concentration profiles in several lab animals were superimposable when a normalization procedure based on the concept of *equivalent time* was adopted. The *equivalent time* was chosen to correlate with body mass to the one-quarter power, based on the similarity principle [14] and the allometric relationship between half-life of methotrexate and body mass of several mammals [13]. This was the first attempt where the similarity principle in biology was applied in animal scale-up of pharmacokinetics [12]. Klotz et al. [15] showed that diazepam plasma clearance was proportional to species body mass to the two-thirds power. Soon thereafter, Weiss et al. [16] detailed the dependence of pharmacokinetic parameters on body mass. They drew the link between pharmacokinetic parameters and body mass by first introducing equations that related pharmacokinetic parameters (hepatic clearance, renal clearance, etc.) to physiological parameters (liver blood flow, *GFR*, etc.), followed by application of the allometric equations for the physiological parameters. Thus, for a drug eliminated by glomerular filtration, the renal clearance would be $CL_{\text{renal}} = GFR = aM^{b}$, where $b$ was cited as 0.82 [16].

In 1980 Boxenbaum [17] characterized interspecies scaling of liver weight, hepatic blood flow, and antipyrine clearance. A log–log plot of antipyrine intrinsic clearance versus body mass yielded a straight line that well described the data from 10 out of 11 mammalian species, but the human value was a pronounced outlier [17]. The seven fold deviation between estimated and observed human antipyrine intrinsic clearance suggested that extrapolation of human pharmacokinetic parameters from animal data using simple allometry may not generally apply. Later, Boxenbaum [18] introduced maximum life span potential (*MLP*) into the allometric equation, producing an allometric relationship between intrinsic clearance of antipyrine per *MLP* and body mass. Boxenbaum's papers drew significant attention and the concept of allometric scaling in pharmacokinetics was well discussed in the 1980s [8, 19–25]. Meanwhile, the suitability of the allometric equation for predicting human pharmacokinetic parameters was further examined [26–29]. Sawada et al. [28] predicted total body clearance, renal clearance, hepatic clearance, volume of distribution, and half-life for six β-lactam antibiotics in humans. Extrapolations were performed using data from mice, rats, rabbits, dogs, and monkeys. For the first time, differences in the plasma protein binding among species was considered an important determinant for the prediction of volume of distribution [28]. Sawada et al. [29] further showed

that similarities among animals and humans in tissue distribution of 10 basic drugs failed to produce a successful prediction of the human value. To predict the volume of distribution it was suggested that unbound fraction in plasma ($f_{up}$) should be included for analysis. In the same period, prediction of human pharmacokinetic parameters using the allometric equation was applied to other antibiotics, such as cephalosporins, monobactams, and macrolides [26, 30].

In the early 1990s an insightful review by Ings [31] heralded a new era for allometry in pharmacokinetics. As this new era began the concept and technique of allometric scaling had become familiar to drug development scientists. But it had also become apparent that the allometric scaling technique was not universally reliable. While for humans the technique sometimes gave acceptable predictions for the pharmacokinetics of those drugs whose elimination was either renal or liver-blood-flow dependent, poor predictions often occurred for low clearance drugs whose elimination was primarily via metabolism. Moreover, the application of the technique was limited to only a few drug classes. These shortcomings contributed to increased research activity in the 1990s. The features of interspecies pharmacokinetic studies in this period include (1) a publication boom focused on interspecies pharmacokinetic scaling for old and new agents and small versus large molecules (proteins); (2) the use of single animal species to extrapolate human pharmacokinetics; (3) the integration of *in vitro* metabolic data into the allometric prediction; (4) the use of the *rule of exponents* along with additional studies of correction factors such as *MLP* and brain weight; and (5) more successful examples of human dose estimation for Phase I studies.

A commentary by Bonate and Howard [32] highlighted an inherent problem faced by those interested in prediction—the uncertainty associated with the prediction. The authors argued that a large bias was associated with publication of predominantly successful allometric scaling attempts, which could lead to a false sense of security with regard to use of allometric scaling to prospectively select dose for a first-time-in-human study. One major problem pointed out by the authors was that prediction intervals are often too wide to be of practical use [32]. By using linear regression and log–log coordinates for prediction, the authors cautioned that an excellent point estimate might not prescribe a good prediction. Moreover, they questioned whether the use of "correction factors" in allometric scaling was appropriate [32, 33].

Hu and Hayton [34] tackled the uncertainty issue in a different way. They carefully analyzed clearance values for 115 xenobiotics from published studies in which at least three species were used for the purpose of interspecies comparison of pharmacokinetics. The allometric exponent for each xenobiotic was calculated along with its confidence interval (*CI*). While it was of interest to note that for each individual drug either 0.67 or 0.75 appeared to be the central tendency of the allometric exponent, the confidence interval of the exponent was often large due to considerable random variability in the clearance values from each species [34].

After so many years of endeavor, it is somewhat discouraging that no matter which allometric method is used for predicting human pharmacokinetic parameters, some drugs will be very poorly predicted. The incorporation of *in vitro* metabolic data improves the prediction performance. Perhaps the most urgent task for the future is to understand what makes certain drugs so unique and how to improve the predictability of these drugs.

## 2 ALLOMETRIC EQUATION

### 2.1 Mathematical Description, $Y = aM^b$

As used here, *scaling* refers to relationships between objects and their representation (e.g., a scale model where 1 cm represents 1 m) or to relationships between elements of a series of similar objects that vary in size only. For example, when geometric bodies such as squares, rectangles, and cubes vary in size, their corresponding linear dimensions differ by a constant ratio. Two geometrically similar rectangles that differ in width by a factor of 2 will differ similarly in length. Length is then said to scale in direct proportion to width, while area scales in direct proportion to the square of width. Geometrically similar bodies are also referred to as *isometric* bodies (*iso* means *equal*).

When bodies are not geometrically similar they are referred to as *allometric* (*allo* means *different*). Characterization of scaling relationships for a series of geometrically dissimilar bodies, for example, mammalian species, is therefore termed *allometric scaling*. For organisms, many aspects of their morphology and physiology change systematically as body mass ($M$) changes, according to the power function:

$$Y = aM^b \tag{1}$$

where $a$ and $b$ are constants, and $Y$ is the particular variable of interest. Because Eq. 1 relates so many biological variables to body mass, it has become known as the *allometric equation*. In its logarithmic form, it is

$$\log Y = \log a + b \log M \tag{2}$$

Use of the allometric equation for scaling in biology was popularized by Huxley [35]. During the ensuing decades, use of the equation became widespread even though no satisfactory underlying theoretical basis supported its use [7, 36]. The allometric equation was often invoked whenever a linear relationship between log $Y$ and log $M$ was observed. Critics of the nondiscriminatory, routine use of the allometric equation have pleaded for more thoughtful application of the equation, noting that use of the "cookbook approach to study the comparative biology of size has falsely simplified a complex problem" and that "use of the method has been considered equivalent to study of the problem" [37].

The foundations for a theoretical underpinning of the allometric equation were laid by the landmark paper of West, Brown, and Enquist [6], which were strengthened in following years [38, 39]. The paper proposed that organisms utilize a similarly structured material supply network, which provides the common mechanism that underlies the allometric equation and in addition explains the often observed quarter-power (and its multiples) value for the body mass exponent, $b$. The theory is based on three assumptions: (1) the supply network has "a space-filling fractal-like branching pattern"; (2) "the final branch of the network . . . is a size-invariant unit"; and (3) "the energy required to distribute resources is minimized." An extensive variety of quantifiable biological characteristics conform to the theory, which removes much of the empirical nature of the allometric equation [40].

### 2.2 Interpretation of *a* and *b*

Empirically the values of the body mass coefficient $a$ and exponent $b$ are simply those that best fit the allometric equation to the particular dataset. The value of $a$ is the value of $Y$ when the value of $M$ equals unity. The $b$ value shows the sensitivity of the $Y$ value to a change in the $M$ value. As is often the case, $b < 1$ and a change in $M$ results in a less-than-proportional change in $Y$. In the special case of $b = 1$, $Y$ varies in direct proportion to $M$. While Heusner [41] argued that the $a$ value is not a constant but rather increases with the size of $M$, applications in scaling pharmacokinetics have generally ignored this possibility.

The $b$ value is often interpreted in comparison with the values obtained from simple geometrical relationships; for example, a value of two-thirds that occurs when the variable $Y$ is proportional to body surface area rather than body mass [42]. In this regard, the allometric scaling of basal metabolic rate in mammals is of interest as a potential surrogate for the body mass sensitivity of drug metabolism rate. Kleiber [43] proposed for basal metabolism a $b$ value of three-fourths, a value that was disputed for several decades, with an alternative value of two-thirds proposed [41, 44–46]. The two-thirds value garnered support from a heat-dissipation mechanism, where it was argued that heat loss is proportional to body surface area and that heat loss is an important constraint that forced its production per unit body mass to decline as body mass increased, in proportion to body surface area. The three-quarters value lacked a compelling mechanistic explanation until the model of West, Brown, and Enquist [6], which convincingly supports the three-fourths value [40].

### 2.3 Using the Allometric Equation

Allometry has been broadly defined as "the study of size and its consequences" [47], and more narrowly described as the specific use of the allometric equation to analyze the body mass influence on quantifiable biological phenomena [37]. In preclinical drug development, the narrow description more aptly applies as the usual problem involves characterization of the association between body mass and quantitative pharmacokinetic and biopharmaceutic parameters—for example, apparent volume of distribution, clearance, half-life, and bioavailability—and using the quantitative association to make predictions. Both inter- and intraspecies applications are possible.

***Choice of Species*** Preclinical drug development uses laboratory animals to gain insights about the pharmacology and toxicology of candidate drugs. Because animals and human beings show marked congruence in their genome, physiology, and biochemistry, it is hoped that insights obtained from animal studies will be applicable to humans. Dose–response relationships, metabolite profiles, bioavailability properties, pharmacokinetic behaviors, and so forth observed in animal studies are hoped to be predictive for humans. Decades of drug development research have validated the utility of animal studies in preclinical development, and informed the uncertainty involved in extrapolation of animal results to humans, but have also borne witness to marked divergences, which sometimes have been unpredictable and difficult to explain.

Thus, an important and hard-to-answer question is: Which species should be used in preclinical studies? While it is tempting to believe that those species with the closest phylogenetic relationship to humans will be the most reliable for prediction of drug behavior in humans, it has become apparent that substantial differences in metabolic clearance occur between closely related species. Environmental and dietary adaptations are also important determinants that shape quantitative differences (and similarities) in drug behavior among species. In addition, nonscientific considerations such as cost and public acceptance of experimentation with particular species cannot be divorced from the selection of species to be studied. It appears uncontroversial to suggest that test species should be from the Class *Mammalia*, and from the 4000+ members of this class the species most commonly used are rat, dog, rabbit, mouse, and monkey [31] (Table 1). While popularity of usage does not necessarily point to the species that provide the best prediction of behavior in humans, there is scant evidence to suggest that alternative species would be better in this regard. A recent review concluded that no animal species consistently shows absorption, distribution, metabolism, and excretion behaviors similar to those of humans [48]. Animals "humanized" by transgenesis may improve this situation [48, 49].

Ideally, the number of test species to use is at least four, particularly if the test animal body mass range does not encompass 70 kg. This recommendation is based on a review of clearance ($CL$) values for 115 xenobiotics subjected to allometric analysis, where failure to find a statistically significant correlation between log $CL$ and log $M$ occurred for 24 compounds [34]. The correlation failure rate declined as the number of species studied increased: a total of 20 studies used only three species and 16 of those failed to find correlation. The correlation failure rate was dramatically lower for $n \geq 4$: for $n = 4$, five failures to find correlation out of 43 studies; for $n = 5$, 2/31; for $n = 6$, 1/15; and for $n = 7$, 0/6.

**TABLE 1 Incidence of Use of Animal Species in Scaling Clearance Values for Xenobiotics**

| Species | Usage (% of 115 Xenobiotics) |
|---|---|
| Rat | 97 |
| Dog | 73 |
| Rabbit | 62 |
| Mouse | 56 |
| Monkey | 53 |
| Pig | 7.8 |
| Guinea pig | 6.1 |
| Sheep | 5.2 |
| Hamster | 5.2 |
| Baboon | 3.5 |
| Chimpanzee | 2.6 |
| Cat | 2.6 |
| Cow | 2.6 |
| Goat | 0.9 |

*Source*: From Ref. 34, with permission.

While at least four species is optimum, data from fewer species is not unusual and effort has been expended toward development of scaling techniques for this case [50–55]. Allometry using three and even two species was investigated using a 103 nonpeptide xenobiotic training set, along with nonallometric single species approaches to *CL* estimation based on the product of hepatic extraction ratio estimated in animal and human hepatic blood flow [50, 51]. For the 103 drug candidates, volume of distribution (*V*), total body clearance, and mean residence time (*MRT*) were determined from the plasma (serum) concentration–time profile after intravenous (IV) administration in rat, dog, monkey, and human, and it was possible to compare the various prediction techniques using animal data with measured human values. Allometry based on all three animal species, and pairs of species (rat–monkey, rat–dog, and dog–monkey), predicted human *V* and *CL* values within an error of –50% to +100% (exception: dog–monkey pair for *CL*), for about two-thirds of the test compounds, and MRT was predicted within –50% to +100% for about two-thirds of compounds using allometry based on three species and on the rat–dog pair only.

***Ideal Body Mass Range*** Clearly, the selected species should span a "reasonable range" of body mass values, and if the objective is to scale the animal values to estimate the corresponding human value, it is desirable to include species with *M* values below and above 70 kg. However, resource constraints often make it difficult to study animals larger than 70 kg and very few published studies of pharmacokinetic parameter scaling include species larger than humans (Table 1). While pharmacokinetic parameter values frequently conform to the allometric equation, use of a small number of values from experimental animals to prospectively predict human values by extrapolation beyond the range of *M* values is "outright perilous" [36].

It is challenging to state precisely what an appropriate *M* range would be; two orders of magnitude seems reasonable, although not based on theoretical principles. For example, if rat were the smallest species (0.2 kg), then the largest species should be on the order of 20 kg. Also, while it is desirable to avoid extrapolating beyond the *M* value range, it is often necessary to do so, in which case one should minimize the extrapolation. A given error in the extrapolation line would be expected to produce a larger error in the extrapolated value the further the extrapolation. For example, if mouse (0.02 kg) were the smallest species and the largest species were 2 kg, the two-orders-of-magnitude suggestion would be met, but the extrapolated human value would be less certain due to the extrapolation from 2 kg to 70 kg. In this case, the *M* range should be increased to minimize extrapolation.

***Optimal Species for Scale-up to Human*** Drug development studies mostly utilize species from the mouse–rat–rabbit–monkey–dog group (the standard five) and data are insufficient to fully address this question beyond these five species. Furthermore, only these species have well established background pathology and other species would not be used for toxicology testing even though they might be advantageous for extrapolation of pharmacokinetic behavior to human [31]. As preclinical pharmacokinetic data are usually obtained as part of the toxicology studies, it is pertinent to consider whether some of the five most commonly used species would serve better than the others for prediction of human pharmacokinetic parameter values.

Rat is nearly universally used in preclinical pharmacokinetic studies and its pharmacokinetic parameter values generally relate well to corresponding human values in allometric scaling analyses [51]. A review of the 61 drugs for which *CL* values were available in human and at least three species showed that qualitatively most of the standard five species were close to the fitted allometric equation line and that the human *CL* was reasonably predicted most of the time (data used in Ref. 34). Significant deviations were relatively few although more prevalent for rabbit and monkey than for mouse, rat, and dog. On the other hand, in their analysis of a 103 compound dataset, Ward and Smith [50, 51] concluded that monkey was somewhat superior to other species for prediction of human pharmacokinetic parameter values when single species extrapolation was conducted. It therefore appears that none of the standard five species stands out as particularly superior or inferior to the others.

***Fitting the Allometric Equation to Data*** Approaches that can be used to fit the allometric equation to *Y,M* data include linear regression of the logarithms of *Y* and *M*, and nonlinear least squares or maximum likelihood fitting of the allometric equation to the *Y,M* data. Both approaches assume that the *M* values are independent variables and that the error associated with their determination is zero. It is therefore essential to accurately determine the mass (weight) of each animal used for determination of the corresponding *Y* value, with the associated experimental error being negligible compared with that of the *Y* value.

While individual *Y,M* values for each animal could be fitted, it is customary practice to average the values from each species and to fit the allometric equation to the means. On log–log coordinates with a two order of magnitude range for *M*, individual values for each species cluster close together, and as a practical matter, there is little difference between analysis of individual or lumped data, with regard to the appearance of the plot and the values of *a* and *b*.

Avoid application of allometry to *M*-normalized *Y* values as this can create spurious correlation [56]. Weighting the *Y* values is not usually done, probably because the linearized form of the equation is usually fit to log *Y*, log *M* points and the magnitude of the range is thereby much reduced. If the equation is fitted directly to *Y*,*M* data, a weighting function should be used to prevent the excessive influence of the larger *Y* values on the *a* and *b* values; for example, $Y^{-1}$ or $Y^{-2}$, or by the reciprocal of the variance of the means. Pertinent statistical characterizations of the goodness of fit should also be calculated and reported; namely, correlation coefficient and its statistical significance, and variance for *a* and *b*. Prothero [56] developed by simulation an empirical equation to calculate the standard error (*se*) for *b* from the mean percent deviation (*PD*) of the points about the line, the number (*n*) of points, and the log of the body mass range (*pWR*), where *WR* is $M_{max}/M_{min}$:

$$se = \frac{0.019(PD)}{pWR \times n^{1/2}} \tag{3}$$

This alternative to the value supplied by standard statistical packages alerts the user to suspect unreliability in the latter when it is smaller than predicted by Eq. 3.

## 3 SCALING CLEARANCE

### 3.1 Physiological Basis and Theoretical Considerations

Clearance is generally defined as the proportionality constant that relates the steady-state rate of drug elimination to its concentration in a reference fluid, which is commonly plasma or serum and sometimes whole blood [57]. For expediency plasma concentration ($C_p$) is used in the following sections with the understanding that it represents the other reference fluids as well. The total body (systemic, plasma) clearance ($CL$) is a primary pharmacokinetic parameter, often determined experimentally as the ratio of dose to area under the plasma concentration–time profile after IV administration of the dose:

$$CL = \text{Dose} \Big/ \int C_p \, dt$$

The total body clearance is the sum of the clearance values contributed by the various organs involved in the elimination of the particular substance. Hepatic clearance ($CL_H$) is the major contributor to $CL$ of drugs; it contributes more than three-fourths the $CL$ for about 60% of the drugs used in therapeutics [58]. For most hepatically cleared drugs, metabolism is the major mechanism, with biliary excretion of importance for a relatively small number of drugs, usually those that are amphipathic and have molecular weight threshold that differs from species to species in the range 400 ± 100. Renal clearance ($CL_R$) is a distant second to $CL_H$ in overall importance, being the principal contributor to $CL$ for only about 15% of drugs in current use. For interspecies scaling of $CL$ values it is helpful to know the relative contributions of the liver, kidneys, and other organs to the total body clearance.

Physiologically based models of clearance provide a useful basis for understanding interspecies scaling relationships [16]. These models incorporate the plasma flow ($Q$) to the clearing organ, the free fraction of drug in the plasma ($f_{up}$), and the intrinsic activity of the clearance process within the organ ($CL_{int}$). Each of these components varies among species, and for each there is the possibility that its allometric scaling relationship differs from the others. In the following three sections, expectations for allometric scaling relationships are developed based on relevant physiologically based clearance models.

***Hepatic Clearance*** Of the several available models of $CL_H$ [57], the well stirred or venous equilibrium model will be used, as other models lead to similar expectations with regard to allometric scaling relationships. The standard representation of the well stirred liver model of $CL_H$ is [59]

$$CL_H = \frac{Q_H f_{up} CL_{int}}{Q_H + f_{up} CL_{int}} \tag{4}$$

$CL_{int}$ here represents the activity of metabolism enzymes or the activity of biliary transport mechanisms.[1] There are two limiting cases, one where $Q_H >> f_{up} CL_{int}$ (low

[1]It may be drug concentration dependent if the metabolism enzyme and/or the transport mechanism are significantly occupied by the drug. This greatly complicates characterization of the pharmacokinetics and thereby the scaling of pharmcokinetic parameter values across species. This situation is not addressed in the present chapter.

extraction case) and the other where $Q_H << f_{up}CL_{int}$ (high extraction case). Most drugs fall into the low extraction category and for these $CL_H \cong f_{up}CL_{int}$. The extraction class is readily identified by comparison of the values of $CL_H$ and $Q_H$; $Q_H$ values are available in the literature [60]. For low extraction drugs the ratio $CL_H/Q_H$ is below 0.25 and for high extraction drugs it is above 0.75. In the low extraction case, interspecies differences in both $f_{up}$ and $CL_{int}$ contribute to interspecies differences in $CL_H$. To remove the influence of $f_{up}$, it is recommended that $CL_{int}$ be the hepatic clearance parameter to scale; it is calculated as the ratio $CL_H/f_{up}$. The human $CL_H$ value is then calculated as the product of the measured human $f_{up}$ and the human $CL_{int}$ predicted by allometric scaling. Interspecies scaling of $CL_{int}$ is problematic; there are examples of high interspecies variability and of a tendency for the predicted human $CL_{int}$ value to substantially exceed the measured value. Predicted human $CL_{int}$ values by allometry applied to animal data should therefore be regarded with appropriate skepticism. Approaches to dealing with this uncertainty are described in Section 3.2.

For the high extraction class of compounds, $CL_H \cong Q_H$, and the expectation in allometric scaling is that $CL_H$ would scale as does $Q_H$; $f_{up}$ is not an issue in this case. Liver blood flow for the standard laboratory animals and humans is well fitted by the allometric equation with the $b$ value being 0.894 [17]. For a high extraction drug therefore, the expectation is that the $b$ value for $CL_H$ would be about 0.9, similar to that for $Q_H$.

For intermediate extraction drugs ($CL_H/Q_H$ between 0.25 and 0.75), $CL_H$ shows sensitivity to $f_{up}$, $Q_H$, and $CL_{int}$. The influence of $f_{up}$ cannot be removed by scaling $CL_H/f_{up}$ as an $f_{up}$ term remains on the right-hand side of Eq. 4. This case is perhaps the most difficult to handle as all three independent determinants can vary in different ways across species, causing $CL_H$ values to vary erratically, the result being no systematic way to scale them.

***Renal Clearance*** Drug elimination by the kidneys is quantitatively less important than that by the liver; it contributes more than 75% to total clearance for ~15% of drugs used in therapeutics [58]. Renal clearance ($CL_R$) may be more complex than hepatic clearance as it may involve a combination of glomerular filtration, tubular secretion, tubular reabsorption, and metabolism. Physiological models are available for each process, but their combination is difficult since the processes act in series and are therefore not simply additive. For the filtration component, $f_{up} \times GFR$ is its contribution to $CL_R$, which may then be reduced due to subsequent tubular reabsorption. The tubular secretion contribution may be modeled along the lines of the well stirred liver model, where $Q_R$ replaces $Q_H$ and represents the postglomerular plasma flow. Tubular reabsorption may also reduce the secretion clearance. Finally, tubular reabsorption of drug from the preurine of the tubule may be passive or involve transporter, and it may be dependent on preurine flow. Clearly, several physiological processes may contribute to the $CL_R$ value; those operative for a particular test compound and their relative contributions are generally only poorly understood at the preclinical stage of development. Nevertheless, $CL_R$ values typically scale relatively well across test animal species and the human $CL_R$ value is typically well predicted by allometric scaling of animal $CL_R$ values. Appropriate adjustments for interspecies differences in $f_{up}$ should be considered; see Role of Protein Binding in Section 3.2. For example, both filtered and secreted test sub-

**TABLE 2 Body Mass Exponent (*b*) Values for Renal Clearance of Renal Function Test Substances [7]**

| Test Substance | Renal Clearance by | *b* |
|---|---|---|
| Urea | Filtration | 0.72 |
| Inulin | Filtration | 0.77 |
| Creatinine | Filtration | 0.69 |
| Diodrast | Filtration + secretion | 0.89 |
| Hippurate | Filtration + secretion | 0.80 |

stances followed the allometric equation reasonably well, and *b* values based on measured renal clearance values in rat, rabbit, dog, and human were in the range 0.69–0.89 (Table 2) [7]. These substances are little bound (urea, inulin, creatinine) to plasma proteins or highly extracted (diodrast, hippurate) so $f_{up}$ was not a confounding variable.

***Total Body Clearance*** Total body clearance may be dominated by a single organ and in that case scaling of *CL* would conform to the expectation for the major pathway. In the event that multiple organs were involved (e.g., liver and kidneys contribute significantly), *CL* in principle could be a sum of allometric expressions:

$$CL = a_L M^{b_L} + a_R M^{b_R} \tag{5}$$

If the *b* values differed significantly, the relative contribution of each pathway would vary across species and a simple log *CL*–log *M* graph would be curved. In practice, *b* values for the hepatic and renal pathways are not sufficiently different that this behavior is observed, and other pathways generally do not contribute significantly.

### 3.2 Empirical Observations

The following subsections present approaches to the allometric scaling of drug clearance and the performance of each approach as experienced in preclinical drug development. The unembellished application of the allometric equation is considered first and this is followed by a series of modifications aimed at improvement of its performance.

***Simple Allometry*** A study of the application of Eq. 1 to interspecies scaling of drug clearance for 115 xenobiotics in three or more species showed that *CL* generally followed the allometric equation (Fig. 1) (Fig. 3 in Ref. 34). While the equation failed to fit *CL* values for 24 of the compounds, as indicated by the failure to show a statistically significant correlation, most of the failures could be attributed to use of only three species (see Section 2.3, Choice of Species). For 91 compounds, the *CL* values followed the allometric equation; in most of these cases four or more species were used. From the set of 91 compounds, 68 included humans as

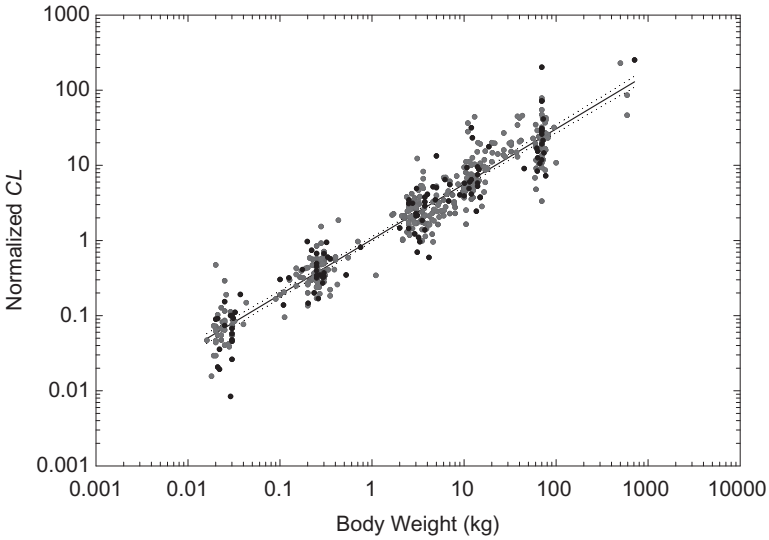


**FIGURE 1** *CL* values for 91 substances determined in three or more species. Values were normalized by adjustment of all values of *a* to 1 [34].

one of the species. For all but 7 of the 68 compounds, the observed human *CL* value deviated from that predicted by the fitted allometric equation by less than 50%; that is, the measured value was within a range of one-half to twice the predicted value, which is a reasonable marker of precision [52, 55].

This relatively rosy record of success (~90%) has to be tempered by two realities. First, because the human *CL* was included in the fitted datasets, the human *CL* would be expected to lie closer to the predicted value than if it had been omitted from the values used to fit the equation. Second, the data for the various compounds were from the literature, mostly from papers that reported on the allometric scaling of the compound. As negative results are rarely reported, it is not known how often the allometric equation fails to scale *CL* values across species.

Ward and Smith [51] compiled from the literature *CL* values for 103 nonpeptide xenobiotics in rat, dog, monkey, and human after IV administration. Allometric scaling of the three animal species' *CL* values to predict the human *CL* value had a 66% success rate, based on whether the human *CL* was correctly predicted to be in the low, moderate, or high *CL* category, defined as a percentage of hepatic blood flow (<30%, 30–70%, and >70%). The success rate differed across categories, being 80–90% for low *CL* substances and only 30–50% for high *CL* substances.

It might be concluded that allometric scaling of human drug *CL* using three or more animal species is a viable approach. For renally eliminated drugs, the success rate is quite high, particularly when renal elimination is primarily by glomerular filtration, plasma protein binding is low, and tubular reabsorption is minimal. For more complex renal clearance mechanisms, the success rate would tend to decline. For drugs eliminated hepatically, the success rate is not as high as when the kidney is the dominant pathway. There are dozens of hepatic drug metabolism enzymes that across species vary in their activity and substrate selectivity. That along with

interspecies variability in the extent of plasma protein binding make the success rates reported above surprisingly high.

To improve the predictive capability of allometry, several embellishments have been proposed. The first to consider is the use of brain weight or maximum life span potential along with $M$.

***Maximum Life Span Potential and Brain Weight*** In scaling the $CL$ of extensively metabolized drugs, investigations found early on that the human $CL$ value often fell considerably below the value predicted in animals. Examples include the intrinsic hepatic clearances of antipyrine and phenytoin, which were below the values predicted by application of the allometric equation to animal data by factors of 7 and 4.4, respectively [17]. Compared with other mammals, humans were also noted to have relatively long life spans, and life span was observed to relate to brain weight. As enhanced longevity may necessitate slowed energy consumption, it was thought that drug metabolism rate may also correlate inversely with life span. A quite successful test of this proposition with antipyrine, phenytoin, and clonazepam intrinsic clearances involved allometric scaling of the product of $CL_{\text{int}}$ and $MLP$, the maximum life span potential [18]. This adjustment brought the human $CL_{\text{int}}$ values into alignment with the animal values (e.g., antipyrine) (Fig. 2).

A variation of this approach is to use brain weight ($BW$, literature values) as a second independent variable in the simple allometric equation. For example, antipyrine $CL_{\text{int}}$ (mL/min) in 15 mammalian species including humans [21] is given by

$$CL_{\text{int}} = 0.3865 M^{1.316} BW^{-0.6188} \tag{6}$$

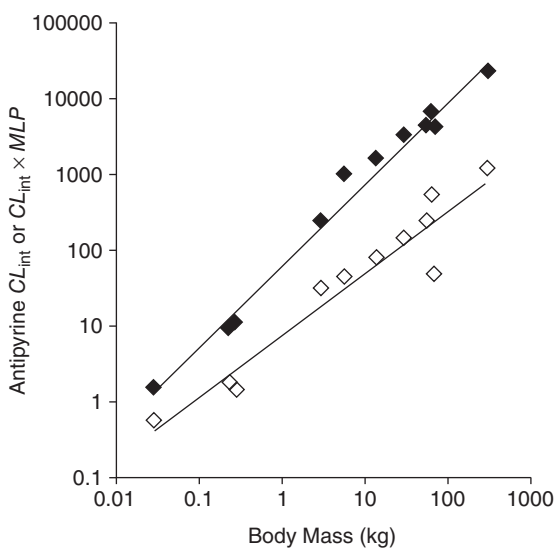


**FIGURE 2** Allometric scaling of antipyrine intrinsic clearance ($CL_{\text{int}}$) alone (◇) and after multiplication by $MLP$ (◆). $CL_{\text{int}}$ from Ref. 21 and MLP from Ref. 18. $CL_{\text{int}}$ for human and pig moved from below and above the allometric line (lower plot) to be near the allometric line for $MLP \times CL_{\text{int}}$ (upper plot).

where $M$ and $BW$ are in kilograms. The coefficient and exponents were determined by fitting the equation to the data using a computer-based algorithm. All 15 calculated $CL_{int}$ values were within –67.4% to +57.6% of the observed values.

However, not all drugs eliminated by metabolism show overprediction of human clearance upon allometric scaling of animal values; for example, Boxenbaum [18] noted that while some do (antipyrine, benzodiazepines, propranolol), others such as tolbutamide and cyclophosphamide do not. More recent work indicates that human $CL$ values may fall above the allometric line as often as they are below it [34]. Use of $MLP$ or $BW$ generally moves the human $CL_{int}$ value more that it does the values from animals. Thus, the $MLP$ or $BW$ technique would tend to move a human value that was on the simple allometric line to well above the line. As a *prospective* technique to predict human $CL_{int}$, use of $MLP$ or $BW$ is problematic as the result would be to increase the number of overpredictions as the number of underpredictions was decreased. From the standpoint of first-dose-in-human estimation, this may be undesirable as overprediction of $CL$ would lead to a less conservative first human dose.

***The "Rule of Exponents"*** For prediction of human $CL$ from animal values, Mahmood [61] and Mahmood and Balian [62, 63] examined the predictive capacity of simple allometry, the $MLP$ and $BW$ techniques described in the previous section, and a variant that involved allometric scaling of the $CL \times BW$ product. These studies pointed to the empirical use of the $b$ value obtained from simple allometry to select one of the following methods to predict human $CL$:

1. For $0.55 < b < 0.75$, use the simple allometric equation.[2]
2. For $0.75 < b < 1.0$, use the $MLP$ technique:

$$CL_{\text{human}} = \frac{a70^b}{8.18 \times 10^5} \tag{7}$$

   where $a$ and $b$ are obtained from fitting the allometric equation to the product $CL_{\text{animal}} \times MLP$ (h), and $8.18 \times 10^5$ h is the human $MLP$.
3. For $b > 1.0$, use

$$CL_{\text{human}} = \frac{a70^b}{1.53} \tag{8}$$

   where $a$ and $b$ are obtained from fitting the allometric equation to the product $CL_{\text{animal}} \times BW$ (kg), and 1.53 kg is the human $BW$.

This empirical approach based on the $b$ value after simple allometry is known as the *rule of exponents*. For the set of 46 test substances, $CL$ values were available in three or more of mouse, rat, rabbit, guinea pig, goat, monkey, and dog. The predicted human $CL$ was within the range 0.5–2.0 times the observed $CL$ for 33

[2]The upper limit (0.75) of the $b$ range was given as 0.70 in Refs. 62 and 63 and as 0.75 in Ref. 61. The 0.75 value was used in this analysis.

substances (72%), and an additional 6 (13%) substances were included in the range 0.2–5.0.

A comprehensive evaluation of the approaches described above for prediction of human *CL* utilized a 103 compound dataset with *CL* values determined in rat, dog and monkey [52, 55]. Using the limits 0.5–2.0 for the ratio of predicted to observed *CL* in humans to define acceptable predictive performance, none of the allometric approaches was particularly successful. Simple allometry with no correction factor performed better (53% of acceptable predicted human *CL*) than did incorporation of *MLP* (38%) and *BW* (45%), and use of the rule of exponents to select among the three approaches (49%). In addition, subsets of test compounds grouped according to physicochemical property ($c \log P$, polar surface area, number of H-bond acceptors, number of rotatable bonds) and route of elimination failed to improve predictive performance. This discouraging analysis suggests that allometry applied to *CL* values determined in animals will not reliably predict the human *CL* value. As noted earlier (Section 2.3, Choice of Species) use of only three animal species does not provide a robust dataset for projection to other species and the relatively poor performance reported by Nagilla and Ward [52] may reflect in part this constraint.

The search for more reliable techniques to improve the allometric-based approach to prediction of human *CL* has more recently turned to incorporation of *in vitro* measurements such as activity of hepatic enzymes and plasma protein binding. These efforts are summarized in the following three subsections.

**In Vitro *Data for* In Vivo *Prediction*** Liver weight follows the allometric equation in mammals with a $b$ value of 0.849 [17], which is similar to that for hepatic blood flow. Thus, in the common mammalian species hepatic blood flow per unit mass of liver is invariant, at about $1.5\,\mathrm{L\,min^{-1}}$ (kg liver)$^{-1}$ and failure to scale *CL* among species is generally not attributable to liver mass or blood flow issues. It has been known for some time that *in vitro* enzyme kinetic parameters can be quantitatively related to *in vivo* hepatic clearance in a single species. This *in vitro*/*in vivo* correspondence was first demonstrated in rat by Rane et al. [64]. For seven compounds having a wide range of hepatic extraction ratio, they successfully predicted *in vivo* hepatic extraction ratio from $K_M$ and $V_{max}$ determined in $9000\,g$ supernatant or microsomal preparations along with liver mass and blood flow, and blood free fraction. Others reported similar success [57]. The feasibility of *in vivo* hepatic $CL_{int}$ prediction ($CL_{int} = V_{max}/K_M$) pointed to the potential to use $K_M$ and $V_{max}$ values in animals to predict the human $CL_{int}$.

A significant problem remained, however. Among the dozens of hepatic enzymes available for xenobiotic metabolism, relatively few are primarily involved in the metabolism of any particular compound. As there are interspecies differences in the relative proportions of the various enzymes and structural differences that can lead to interspecies differences in intrinsic activity and metabolic pathways, it is not so simple a matter to use *in vitro* metabolism parameters determined in laboratory animals to predict $CL_{int}$ in humans [48, 65]. Possible reasons for failure to predict human $CL_{int}$ include nonhepatic metabolism, slow equilibrium between blood and hepatocytes, and involvement of transporters [66]. Recent effort has therefore focused on integration of *in vitro* measures of metabolic activity with *in vivo* $CL_{int}$ values in animals to improve the accuracy of human $CL_{int}$ prediction.

***Incorporation of In Vitro Data in Allometry*** The first successful combination of *in vitro* and *in vivo* data to predict the human $CL_{int}$ took the approach of adjusting each *in vivo* animal clearance ($CL_{an}$) value by multiplying by the human/animal ratio of *in vitro* $V_{max}/K_M$ [67, 68]. This effectively adjusted each animal $CL$ to the value it would have had if the overall intrinsic activity of drug metabolism enzymes were the same in the test animal and humans on a unit mass basis; that is, $CL_{an}$ could be "humanized" ($CL_{an}^{h}$) by the following adjustment:

$$CL_{an}^{h} = CL_{an} \frac{(V_{max}/K_M)_{human}}{(V_{max}/K_M)_{an}} \tag{9}$$

This adjustment to $CL_{an}$ reduces the problem associated with interspecies differences in metabolism pathways and enzyme activities so that $CL_{an}^{h}$ values will more likely follow the allometric equation and will also more likely predict the *in vivo* human $CL$. It is only appropriate for low or intermediate hepatic extraction drugs since for high extraction drugs hepatic blood flow rather than $CL_{int}$ is the principal determinant of $CL_H$. Because liver size and blood flow scale according to the allometric equation, adjustments for their interspecies differences are unnecessary. At least three animal species should be used.

The *in vitro* $V_{max}$ and $K_M$ values were determined using a hepatic microsome preparation. The impact of inefficiency in the microsomal isolation procedure is minimized if it is similar for all determinations, as assured by simultaneous determination of *in vitro* parameters using the same protocol. Lavé et al. [67] also used an *in vitro* hepatocyte preparation to adjust the *in vivo* $CL_{an}$ with equally good results. For hepatocytes, *in vitro* $CL_{int}$ (volume/$10^6$ cells) = Dose/$AUC(0 - t)$, where $AUC(0 - t)$ is the area under the drug concentration–time profile in the hepatocyte suspension over the duration of the incubation. The hepatocyte method for assessment of *in vitro* $CL_{int}$ is the method of choice as it provides the widest array of enzyme activities and generally shows the same metabolite pattern as is seen *in vivo* [69].

***Role of Protein Binding*** Both renal and hepatic clearance may be affected by the degree of plasma binding; namely, when clearance is predominantly composed of low extraction hepatic or active renal tubular secretion clearance, or glomerular filtration. Xenobiotic binding affinity can vary considerably among species, due to differences in the concentration of the particular protein and to differences in the affinity of the xenobiotic for the binding protein [70, 71]. The influence of interspecies variability on the allometric scaling of $CL$ can be reduced by calculation of the "unbound" clearance [72] before the allometric analysis. This adjustment can be added to that shown in Eq. 9:

$$CL_{an}^{h} = CL_{an} \frac{(V_{max}/K_M)_{human}}{(V_{max}/K_M)_{an}} \frac{f_{up,human}}{f_{up,an}} \tag{10}$$

***Oral Clearance*** Clearance determined as Dose/$\int C_p\, dt$ after oral administration of test substance is uncertain due to lack of knowledge about the bioavailability, characterized as the fraction of dose that reaches the systemic circulation, $F$:

$$CL_{\text{oral}} = \frac{CL}{F} = \frac{Dose}{\int C_{\text{p}} dt} \tag{11}$$

Interspecies variability in $F$ due to species-to-species differences in bioavailability (incomplete absorption, presystemic degradation) will be reflected in the apparent oral clearance ($CL_{\text{oral}} = CL/F$) values and will contribute to their scatter around the fitted allometric line. A special case is when $F < 1$ due solely to presystemic hepatic first-pass elimination of the test compound. Here, $CL/F$ is in fact $f_{\text{up}}CL_{\text{int}}$, assuming that $C_{\text{p}} << K_{\text{M}}$. For this special case, the adjustments described in Eqs. 9 and 10 may be carried out and the human $CL$ estimated using the allometric equation. When $F < 1$ for reasons other than hepatic first-pass extraction, there is no adjustment procedure available.

***Projections Based on One or Two Species*** Lin [48] noted that the per kg rat/human $GFR$ ratio was 4.8, which is generally found for the rat/human $CL_{\text{R}}$ ratio of renally cleared drugs. He suggested that the human $CL_{\text{R}}$ ($\text{mL min}^{-1}\text{kg}^{-1}$) could therefore be estimated as the rat $CL_{\text{R}}$ ($\text{mL min}^{-1}\text{kg}^{-1}$) divided by 4.8. This seems appropriate since renal clearance is scaled relatively well by the allometric equation.

Ward and Smith [51] examined two-species allometry for estimation of human $CL$, using 103 test compounds and their $CL$ values determined in rat, dog, monkey, and human. Rat–dog and rat–monkey, but not dog–monkey, pairs provided correct qualitative estimates of human $CL$ category (low, intermediate, or high fraction of hepatic blood flow) with about 60% success. They also investigated calculation of the human $CL$ according to the animal $CL$ expressed as a fraction of liver blood flow multiplied by human liver blood flow:

$$CL_{\text{man}} = \frac{CL_{\text{an}}}{Q_{\text{H,an}}} \times Q_{\text{H,human}} \tag{12}$$

While $CL$ was assumed to be primarily hepatic, a check of the 103 test compound list against a database of fraction of dose eliminated in urine (*Goodman and Gilman's the Pharmacological Basis of Therapeutics*, 9th and 10th editions [58]) showed that about one-fourth of the 55 test compounds in the database were predominantly (>50%) eliminated in the urine of humans. Use of Eq. 12 makes the tacit assumption that the ratio $CL/Q_{\text{H}}$ is a constant across species, which seems unfounded even when hepatic clearance is the dominant contributor.

### 3.3 Recommended Approach

1. Determine the organ(s) primarily involved in clearance of the test compound. Generally, this will be liver or kidneys with the likelihood of liver being greater.
2. Determine $CL$ after IV administration if possible in a minimum of four lab animal species. Fewer than four will raise the uncertainty considerably in the extrapolated human $CL$ value. $CL$ value for each species should be an average from several animals with interfacility differences minimized. Body mass

should span a range of 100, and the mass of the largest animal should be ~10 kg.

3. If renal clearance is the primary elimination pathway, simple allometry will generally predict the human *CL*. If $f_{up}$ varies among species, use the unbound clearance in the allometric analysis.
4. If hepatic clearance is the primary elimination pathway, assess whether the clearance is low or high extraction. If the latter, simple allometry will generally predict the human *CL*.
5. If the primary elimination pathway is hepatic and low extraction, prediction of the human *CL* by simple allometry has a relatively low probability of success. Use $f_{up}$ and *in vitro* intrinsic clearance measurements in the animals and humans to adjust the *in vivo* animal clearance values prior to allometric analysis.
6. Recognize that the extrapolated human *CL* value may be far from the true value. Extrapolation is considered successful if it is within a range of one-half to twice the true value.

## 4 SCALING VOLUME OF DISTRIBUTION

### 4.1 Physiological Basis and Theoretical Considerations

Drug distribution is often expressed in terms of the volume of distribution ($V_d$), which is usually defined as the ratio of drug amount in the body to the plasma (or blood) concentration. It is generally believed that only the unbound drug can freely diffuse across membranes that separate vascular and tissue compartments. Changes in drug protein binding in plasma and tissue can therefore affect drug distribution in the body, which can be described by the following relationship [59]:

$$V_d = V_p + \left(\frac{f_{up}}{f_{ut}}\right) V_t \tag{13}$$

$V_p$ and $V_t$ are the volume of plasma and tissue, respectively, and $f_{up}$ and $f_{ut}$ are the unbound fraction of drug in plasma and tissue, respectively. Considering allometric scaling of $V_d$, three scenarios are apparent: (1) $V_d \approx V_p$, (2) $V_d >> V_p$, and (3) $V_d = V_p + V_t$.

When a drug is extensively bound to plasma protein (i.e., $f_{up} \to 0$), the drug is confined in the vascular space; therefore, $V_d \to V_p$, which may satisfy condition 1. For a drug that fulfills the first condition, the interspecies scaling of $V_d$ would be expected to follow that of blood volume. Generally, the total blood volume tends to scale across species with the first power of body mass ($M^{0.99}$ [7] and $M^{1.02}$ [73]). Therefore, for drugs with extremely high protein binding in plasma,

$$V_d \approx aM^{1.0} \tag{14}$$

On the other hand, for a drug that is extensively bound to tissue protein (i.e., $f_{ut}$ very small), $V_d$ may be significantly larger than $V_p$ and the second condition

may be fulfilled. Usually if $V_d \geq 30\,L$, $V_p$ has only a minor effect on $V_d$ [59]; in this case,

$$V_d \approx \left(\frac{f_{up}}{f_{ut}}\right) V_t \tag{15}$$

As indicated in Eq. 15, both $f_{up}$ and $f_{ut}$ would affect the magnitude of $V_d$; therefore, if interspecies variation of protein binding were substantial, it would be difficult to find an allometric relationship for $V_d$.

When a drug has very low plasma and tissue binding (i.e., $f_{up}$ and $f_{ut} \rightarrow 1$), the $V_d$ of the drug will reflect the volume of total body water (condition 3). Accordingly, the scaling of $V_d$ would correlate with the scaling of the total body water (*TBW*). The reported *b* value is 0.89 for the *TBW* [74]; therefore, for a drug that satisfies condition 3, the allometric relationship could be

$$V_d = aM^{0.89} \tag{16}$$

The scaling relationship of the mass (or volume) of the major drug distribution organs could also be indicative for the scaling of $V_d$. The scaling exponent of major organs ranged from 0.70 to 0.99 [7].

The $V_d$ defined in the compartmental models usually does not represent a real physiological volume; however, it is an important pharmacokinetic parameter for the estimation of clinical loading dose. The rationale behind the scaling of $V_d$ in drug development is at least twofold: (1) the volume of distribution of the central compartment ($V_c$) can be used to estimate initial plasma concentration, which may serve as an additional indicator for safety/toxicity for first-time dosing in humans; (2) combined with a known *CL* value, the value of $V_c$ can be used to estimate the half-life, since $t_{1/2} = 0.693(V_d/CL)$. For compounds exhibiting multicompartmental kinetics, other volume terms ($V_\beta$, $V_{ss}$) have also been considered. $V_{ss}$ would be expected to scale according to the principles embodied in Eqs. 14–16. $V_\beta$ includes influences of distribution and elimination kinetic parameters and its scaling according to the allometric equation is less certain.

### 4.2 Empirical Observations

Generally, good correlation between body mass and $V_c$ across species is expected. Most reported allometric exponents for $V_d$ scattered about 1.0. Some drugs have *b* values that deviate significantly from 1.0 [75]. Mahmood [76] showed that $V_c$ can be more accurately predicted than $V_\beta$ and $V_{ss}$. It has been suggested that extrapolating $V_\beta$ or $V_{ss}$ from animal data is of limited significance for the first-time dosing in humans, because both parameters can be estimated later from human data [75]. Obach et al. [77] compared four different methods to predict human $V_{ss}$ and demonstrated that when the allometric method was used, a better prediction was obtained when protein binding was taken into account. The main purpose of predicting $V_d$ was to use the predicted $V_d$ value in combination with a predicted *CL* value for predicting human $t_{1/2}$, "which is a more meaningful parameter with decision-making impact on the drug discovery and development processes" [77].

Interspecies scaling of $V_d$ has also been applied to protein drugs and the reported allometric exponents were around 1.0. Mordenti [78] analyzed the volume of distribution of five human proteins in humans and laboratory animals as a function of body mass using allometric scaling techniques. The analyses showed that the volume data for each protein were satisfactorily described by an allometric equation. The allometric exponent for $V_c$ ranged from 0.83 to 1.05, and the allometric exponent for $V_{ss}$ ranged from 0.84 to 1.02. In another study, Mordenti [79] reported that the allometric exponents for recombinant human factor VIII were 1.04 and 0.84 for $V_c$ and $V_{ss}$, respectively.

### 4.3 Limitations and Challenges

Prediction involves uncertainty. To establish a meaningful allometric relationship, it is always desirable to include as many animal species as possible. However, in the real world of drug development, pharmacokinetic data are often obtained from a limited number of preclinical species. It becomes a challenging task to predict any pharmacokinetic parameters when data are only available from two or three species. Mahmood and Balian [80] showed that allometric scaling of $V_d$ from only two species is as predictive as that observed from three or more species. Ward and Smith [50], however, showed a contradictory result, indicating that three-species allometry provided improved quantitative prediction compared with two-species allometry. Moreover, single-species extrapolation has been proposed for making early go/no-go human PK decisions for drug discovery candidates [50, 81]. When the interspecies variation in pharmacokinetics, for a given drug candidate, cannot be understood prospectively, the claim of superiority of any projection based on limited data (i.e., from one to three species) may be futile. In other words, the overall prediction may be equally good or equally bad, depending on how one determines the prediction performance. For estimating volume of distribution, it may imply knowing in advance whether the protein binding (plasma and tissue) of a given drug is likely to be so variable across species. In this regard, the use of *in vitro* binding data may be helpful. Further studies may aim at characterizing the molecular and physicochemical properties of drugs that show high interspecies variability.

## 5 SCALING HALF-LIFE

### 5.1 Rationale and Theoretical Considerations

Half-life is an important pharmacokinetic parameter in that (1) it provides the basis for the selection of dosage regimen and (2) it offers prediction of the time to reach steady-state concentration and of the degree of accumulation. In drug development, the utility of scaling half-life across species lies mainly in predicting the dosing frequency of a lead compound. This is true because a frequently asked question in the compound selection process is: Will compound X be a once daily drug? [77].

Theoretically, half-life can be written

$$t_{1/2} = 0.693\left(\frac{V_d}{CL}\right) \tag{17}$$

If one assumes that the $V_d$ scales with $M^{1.0}$ and $CL$ scales as $M^{0.75}$, then the expected scaling relationship would be

$$t_{1/2} = aM^{0.25} \tag{18}$$

Practically, there are three ways to predict the half-life in humans: (1) direct extrapolation from animal data using the allometric relationship for half-life, if there is any; (2) indirect prediction using Eq. 17, where $V_d$ and $CL$ in humans are predicted *a priori* [63, 77, 82]; and (3) indirect estimation from predicted mean residence time ($MRT$), using the following relationship [76]:

$$t_{1/2} = 0.693MRT \tag{19}$$

### 5.2 Empirical Observations

Predicting the half-life in humans from animal data is challenging. One study reported that the exponents of the half-life for drugs varied from –0.066 to 0.547, with the average of 0.19 [76]. Riviere et al. [83] conducted an interspecies allometric analysis of the $t_{1/2}$ of 44 drugs across veterinary and laboratory animal species. A total of 11 drugs (25%) showed statistically significant correlations between log $t_{1/2}$ and log $M$. The exponent $b$ value for this group of drugs ranged from 0.1 to 0.415, with an average of 0.24 ± 0.09. The remaining 33 drugs showed lack of correlation; many of these drugs were low hepatic extraction drugs. Lack of quality data (e.g., narrow body mass range, multiple data sources) may account for the lack of correlation observed with some drugs [83].

### 5.3 Limitations and Challenges

Difficulty in obtaining consistent correlation between $t_{1/2}$ and body mass across species might be due to the fact that $t_{1/2}$ is a hybrid parameter of $V_d$ and $CL$; that is, both drug distribution and elimination processes determine the value of $t_{1/2}$. Successful prediction of $t_{1/2}$ tended to occur for drugs that are mainly renally excreted and have low protein binding [26]. It would be more difficult to obtain an allometric relationship for drugs that follow complex disposition mechanisms, such as extensive metabolism involving multiple enzymatic systems and high plasma and/or tissue binding. Use of the above-mentioned indirect methods may improve the predictability of $t_{1/2}$. Moreover, the reliability of prediction may be increased by the incorporation of *in vitro* metabolic data [77, 84] and of calculated molecular properties [53]. The main focus of predicting $t_{1/2}$ should be placed on assigning compounds to appropriate dosing regimen categories (e.g., once daily, twice daily), rather than on a precise estimate of $t_{1/2}$ value.

## 6 CONCLUSION

Allometric scaling can be useful in the projection of preclinical pharmacokinetic parameter values to humans. Apparent volume of distribution will generally scale successfully, more so when interspecies differences in plasma protein binding are

accounted for. Also, renal clearance generally will scale successfully to humans and accounting for interspecies differences in plasma protein binding will improve the success rate. Scaling metabolic clearance is problematic but progress on incorporation of *in vitro* metabolism parameters and plasma protein binding to adjust animal clearance values for interspecies differences in intrinsic metabolism activity has improved the prospects for successful prediction of human *CL* for compounds cleared primarily by metabolism. In future, transgenic animals with human intrinsic enzyme activities may further improve predictive success.

## REFERENCES


1. Huxley JS, Teissier G. Terminology of relative growth. *Nature* 1936;137:780–781.
2. White CR, Seymour RS. Allometric scaling of mammalian metabolism. *J Exp Biol* 2005;208:1611–1619.
3. Rubner N. Ueber den Einfluss der Koerpergroesse auf Stoff und Kraftwechsel. *Z Biol* 1883;19:535–562.
4. Kleiber M. Body size and metabolism. *Hilgardia* 1932;6:315–332.
5. Brody S. *Bioenergetics and Growth*. New York: Reinhold Publishing Corporation; 1945.
6. West GB, Brown JH, Enquist BJ. A general model for the origin of allometric scaling laws in biology. *Science* 1997;276:122–126.
7. Adolph EF. Quantitative relations in the physiological constitutions of mammals. *Science* 1949;109:579–585.
8. Davidson IW, Parker JC, Beliles RP. Biological basis for extrapolation across mammalian species. *Regul Toxicol Pharmacol* 1986;6:211–237.
9. Pinkel D. The use of body surface area as a criterion of drug dosage in cancer chemotherapy. *Cancer Res* 1958;18:853–856.
10. Freireich EJ, Gehan EA, Rall DP, Schmidt LH, Skipper HE. Quantitative comparison of toxicity of anticancer agents in mouse, rat, hamster, dog, monkey, and man. *Cancer Chemother Rep* 1966;50:219–244.
11. Mellett LB. Comparative drug metabolism. *Prog Drug Res* 1969;13:136–169.
12. Dedrick RL. Animal scale-up. *J Pharmacokinet Biopharm* 1973;1:435–461.
13. Dedrick RL, Bischoff KB, Zaharko DS. Interspecies correlation of plasma concentration history of methotrexate (NSC-740). *Cancer Chemother Rep Part 1* 1970;54:95–101.
14. Stahl WR. Similarity analysis of physiological systems. *Perspect Biol Med* 1963;6:291–321.
15. Klotz U, Antonin KH, Bieck PR. Pharmacokinetics and plasma binding of diazepam in man, dog, rabbit, guinea pig and rat. *J Pharmacol Exp Ther* 1976;199:67–73.
16. Weiss M, Sziegoleit W, Forster W. Dependence of pharmacokinetic parameters on the body weight. *Int J Clin Pharmacol Biopharm* 1977;15:572–575.
17. Boxenbaum H. Interspecies variation in liver weight, hepatic blood flow, and antipyrine intrinsic clearance: extrapolation of data to benzodiazepines and phenytoin. *J Pharmacokinet Biopharm* 1980;8:165–176.
18. Boxenbaum H. Interspecies scaling, allometry, physiological time, and the ground plan of pharmacokinetics. *J Pharmacokinet Biopharm* 1982;10:201–227.
19. Boxenbaum H, Ronfeld R. Interspecies pharmacokinetic scaling and the Dedrick plots. *Am J Physiol* 1983;245:R768–R775.

20. Boxenbaum H. Interspecies pharmacokinetic scaling and the evolutionary-comparative paradigm. *Drug Metab Rev* 1984;15:1071–1121.
21. Boxenbaum H, Fertig JB. Scaling of antipyrine intrinsic clearance of unbound drug in 15 mammalian species. *Eur J Drug Metab Pharmacokinet* 1984;9:177–183.
22. Calabrese EJ. Animal extrapolation and the challenge of human heterogeneity. *J Pharm Sci* 1986;75:1041–1046.
23. Mordenti J. Man versus beast: pharmacokinetic scaling in mammals. *J Pharm Sci* 1986; 75:1028–1040.
24. Yates FE, Kugler PN. Similarity principles and intrinsic geometries: contrasting approaches to interspecies scaling. *J Pharm Sci* 1986;75:1019–1027.
25. Hayton WL. Pharmacokinetic parameters for interspecies scaling using allometric techniques. *Health Phys* 1989;57(Suppl 1):159–164.
26. Mordenti J. Forecasting cephalosporin and monobactam antibiotic half-lives in humans from data collected in laboratory animals. *Antimicrob Agents Chemother* 1985;27: 887–891.
27. Swabb EA, Bonner DP. Prediction of aztreonam pharmacokinetics in humans on data from animals. *J Pharmacokinet Biopharm* 1983;11:215–223.
28. Sawada Y, Hanano M, Sugiyama Y, Iga T. Prediction of the disposition of beta-lactam antibiotics in humans from pharmacokinetic parameters in animals. *J Pharmacokinet Biopharm* 1984;12:241–261.
29. Sawada Y, Hanano M, Sugiyama Y, Harashima H, Iga T. Prediction of the volumes of distribution of basic drugs in humans based on data from animals. *J Pharmacokinet Biopharm* 1984;12:587–596.
30. Duthu GS. Interspecies correlation of the pharmacokinetics of erythromycin, oleandomycin, and tylosin. *J Pharm Sci* 1985;74:943–946.
31. Ings RM. Interspecies scaling and comparisons in drug development and toxicokinetics. *Xenobiotica* 1990;20:1201–1231.
32. Bonate PL, Howard D. Prospective allometric scaling: Does the emperor have clothes? *J Clin Pharmacol* 2000;40:665–670; discussion pp 671–676.
33. Mahmood I. Critique of prospective allometric scaling: Does the emperor have clothes? *J Clin Pharmacol* 2000;40:341–344; discussion pp 345–346.
34. Hu TM, Hayton WL. Allometric scaling of xenobiotic clearance: uncertainty versus universality. *AAPS PharmSci* 2001;3:E29.
35. Huxley JS. *Problems of Relative Growth*. London: Methuen; 1932.
36. Schmidt-Nielsen K. *Scaling: Why Is Animal Size So Important?* Cambridge, UK: Cambridge University Press; 1984.
37. Smith RJ. Allometric scaling in comparative biology: problems of concept and method. *Am J Physiol* 1984;246:R152–R160.
38. West GB, Brown JH, Enquist BJ. The fourth dimension of life: fractal geometry and allometric scaling of organisms. *Science* 1999;284:1677–1679.
39. Gillooly JF, Brown JH, West GB, Savage VM, Charnov EL. Effects of size and temperature on metabolic rate. *Science* 2001;293:2248–2251.
40. West GB, Brown JH. The origin of allometric scaling laws in biology from genomes to ecosystems: towards a quantitative unifying theory of biological structure and organization. *J Exp Biol* 2005;208:1575–1592.
41. Heusner AA. Energy metabolism and body size. I. Is the 0.75 mass exponent of Kleiber's equation a statistical artifact? *Respir Physiol* 1982;48:1–12.
42. Gunther B, Morgado E. Dimensional analysis revisited. *Biol Res* 2003;36:405–410.

43. Kleiber M. Body size and metabolic rate. *Physiol Rev* 1947;27:511–541.

44. Feldman HA, McMahon TA. The 3/4 mass exponent for energy metabolism is not a statistical artifact. *Respir Physiol* 1983;52:149–163.

45. Hayssen V, Lacy RC. Basal metabolic rates in mammals: taxonomic differences in the allometry of BMR and body mass. *Comp Biochem Physiol A* 1985;81:741–754.

46. Dodds PS, Rothman DH, Weitz JS. Re-examination of the "3/4-law" of metabolism. *J Theor Biol* 2001;209:9–27.

47. Gould SJ. Allometry and size in ontogeny and phylogeny. *Biol Rev Camb Philos Soc* 1966;41:587–640.

48. Lin JH. Applications and limitations of interspecies scaling and *in vitro* extrapolation in pharmacokinetics. *Drug Metab Dispos* 1998;26:1202–1212.

49. Henderson CJ, Wolf CR. Transgenic analysis of drug-metabolizing enzymes: preclinical drug development and toxicology. *Mol Interv* 2003;3:331–343.

50. Ward KW, Smith BR. A comprehensive quantitative and qualitative evaluation of extrapolation of intravenous pharmacokinetic parameters from rat, dog, and monkey to humans. II. Volume of distribution and mean residence time. *Drug Metab Dispos* 2004;32: 612–619.

51. Ward KW, Smith BR. A comprehensive quantitative and qualitative evaluation of extrapolation of intravenous pharmacokinetic parameters from rat, dog, and monkey to humans. I. Clearance. *Drug Metab Dispos* 2004;32:603–611.

52. Nagilla R, Ward KW. A comprehensive analysis of the role of correction factors in the allometric predictivity of clearance from rat, dog, and monkey to humans. *J Pharm Sci* 2004;93:2522–2534.

53. Jolivette LJ, Ward KW. Extrapolation of human pharmacokinetic parameters from rat, dog, and monkey data: molecular properties associated with extrapolative success or failure. *J Pharm Sci* 2005;74:1467–1483.

54. Ward KW, Nagilla R, Jolivette LJ. Comparative evaluation of oral systemic exposure of 56 xenobiotics in rat, dog, monkey and human. *Xenobiotica* 2005;35:191–210.

55. Nagilla R, Ward KW. Erratum: A comprehensive analysis of the role of correction factors in the allometric predictivity of clearance from rat, dog, monkey to humans. *J Pharm Sci* 2005;94:231–232.

56. Prothero J. Methodological aspects of scaling in biology. *J Theor Biol* 1986;118:259–286.

57. Wilkinson GR. Clearance approaches in pharmacology. *Pharmacol Rev* 1987;39:1–47.

58. Thummel KE, Shen DD. Design and optimization of dosage regimens: pharmacokinetic data. In Limbird LE, Hardman JG, Eds. *Goodman and Gilman's the Pharmacological Basis of Therapeutics*. New York: McGraw-Hill; 2001, pp 1924–2023.

59. Wilkinson GR, Shand DG. Commentary: a physiological approach to hepatic drug clearance. *Clin Pharmacol Ther* 1975;188:377–390.

60. Davies B, Morris T. Physiological parameters in laboratory animals and humans. *Pharm Res* 1993;10:1093–1095.

61. Mahmood I. Interspecies scaling: predicting clearance of anticancer drugs in humans. A comparative study of three different approaches using body weight or body surface area. *Eur J Drug Metab Pharmacokinet* 1996;21:275–278.

62. Mahmood I, Balian JD. Interspecies scaling: predicting clearance of drugs in humans. Three different approaches. *Xenobiotica* 1996;26:887–895.

63. Mahmood I, Balian JD. Interspecies scaling: predicting pharmacokinetic parameters of antiepileptic drugs in humans from animals with special emphasis on clearance. *J Pharm Sci* 1996;85:411–414.
64. Rane A, Wilkinson GR, Shand DG. Prediction of hepatic extraction ratio from *in vitro* measurement of intrinsic clearance. *J Pharmacol Exp Ther* 1977;200:420–424.
65. Calabrese EJ. *Principles of Animal Extrapolation: Predicting Human Responses from Animal Studies*. Chelsea: Lewis Publisher; 1991.
66. Iwatsubo T, Hirota N, Ooie T, Suzuki H, Shimada N, Chiba K, Ishizaki T, Green CE, Tyson CA, Sugiyama Y. Prediction of *in vivo* drug metabolism in the human liver from *in vitro* metabolism data. *Pharmacol Ther* 1997;73:147–171.
67. Lavé T, Schmitt-Hoffmann AH, Coassolo P, Valles B, Ubeaud G, Ba B, Brandt R, Chou RC. A new extrapolation method from animals to man: application to a metabolized compound, mofarotene. *Life Sci* 1995;56:PL473–PL478.
68. Lavé T, Dupin S, Schmitt C, Chou RC, Jaeck D, Coassolo P. Integration of *in vitro* data into allometric scaling to predict hepatic metabolic clearance in man: application to 10 extensively metabolized drugs. *J Pharm Sci* 1997;86:584–590.
69. Lavé T, Coassolo P, Reigner B. Prediction of hepatic metabolic clearance based on interspecies allometric scaling techniques and *in vitro–in vivo* correlations. *Clin Pharmacokinet* 1999;36:211–231.
70. Guarino AM, Anderson JB, Starkweather DK, Chignell CF. Pharmacologic studies of camptothecin (NSC-100880): distribution, plasma protein binding, and biliary excretion. *Cancer Chemother Rep* 1973;57:125–140.
71. Mi Z, Burke TG. Marked interspecies variations concerning the interactions of camptothecin with serum albumins: a frequency-domain fluorescence spectroscopic study. *Biochemistry* 1994;33:12540–12545.
72. Chiou WL, Choi YM. Unbound total (plasma) clearance approach in interspecies pharmacokinetics correlation: theophylline–cimetidine interaction. *Pharm Res* 1995;12: 1238–1239.
73. Stahl WR. Scaling of respiratory variables in mammals. *J Appl Physiol* 1967;48: 1052–1059.
74. Chappell WR, Mordenti J. Extrapolation of toxicological and pharmacological data from animals to humans. In Testa B, Ed. *Advances in Drug Research*, Volume 20. London, UK: Academic Press; 1991, pp 1–116.
75. Mahmood I. Allometric issues in drug development. *J Pharm Sci* 1999;88:1101–1106.
76. Mahmood I. Interspecies scaling: predicting volumes, mean residence time and elimination half-life. Some suggestions. *J Pharm Pharmacol* 1998;50:493–499.
77. Obach RS, Baxter JG, Liston TE, Silber BM, Jones BC, MacIntyre F, Rance DJ, Wastall P. The prediction of human pharmacokinetic parameters from preclinical and *in vitro* metabolism data. *J Pharmacol Exp Ther* 1997;283:46–58.
78. Mordenti J, Chen SA, Moore JA, Ferrailo BL, Green JD. Interspecies scaling of clearance and volume of distribution data for five therapeutic proteins. *Pharm Res* 1991;8: 1351–1359.
79. Mordenti J, Osaka G, Garcia K, Thomsen K, Licko V, Meng G. Pharmacokinetics and interspecies scaling of recombinant human Factor VIII. *Toxico Appl Pharmacol* 1996;4:75–78.
80. Mahmood I, Balian JD. Interspecies scaling: a comparative study for the prediction of clearance and volume using two or more than two species. *Life Sci* 1996;59:579–585.
81. Caldwell GW, Masucci JA, Yan Z, Hageman W. Allometric scaling of pharmacokinetic parameters in drug discovery: can human $CL$, $V_{ss}$ and $t_{1/2}$ be predicted from *in-vivo* rat data? *Eur J Drug Metab Pharmacokinet* 2004;29:133–143.

82. Bachmann K. Predicting toxicokinetic parameters in humans from toxicokinetic data acquired from three small mammalian species. *J Appl Toxicol* 1989;9:331–338.
83. Riviere JE, Martin-Jimenez T, Sundlof SF, Craigmill AL. Interspecies allometric analysis of the comparative pharmacokinetics of 44 drugs across veterinary and laboratory animal species. *J Vet Pharmacol Ther* 1997;20:453–463.
84. Lavé T, Dupin S, Schmitt M, Kapps M, Meyer J, Morgenroth B, Chou RC, Jaeck D, Coassolo P. Interspecies scaling of tolcapone, a new inhibitor of catechol-*O*-methyltransferase (COMT). Use of *in vitro* data from hepatocytes to predict metabolic clearance in animals and humans. *Xenobiotica* 1996;26:839–851.

# INTERRELATIONSHIP BETWEEN PHARMACOKINETICS AND METABOLISM

James W. Paxton
*The University of Auckland, Auckland, New Zealand*

**Contents**

1 Introduction
2 Important Pharmacokinetic Parameters
  2.1 Systemic Clearance ($CL_P$)
  2.2 Oral Bioavailability ($F$)
  2.3 Volume of Distribution ($V$)
  2.4 Elimination half-life ($T_{1/2}$)
3 Approaches to Pharmacokinetic Analysis
  3.1 Compartmental Analysis
  3.2 Noncompartmental Analysis
  3.3 Physiologically Based Analysis
4 Predicting Pharmacokinetic Parameters in Humans
  4.1 Allometric Scaling
  4.2 Microdosing
5 Metabolism Features that May Convey Pharmacokinetic Disadvantages
  5.1 Nonlinear Pharmacokinetics
  5.2 Drug–Drug Interactions
  5.3 Polymorphic Drug Metabolism
  5.4 Formation of Active/Reactive Metabolites
6 Conclusion
  References

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

## 1 INTRODUCTION

For most drugs, the magnitude and duration of their pharmacological effect depend on their concentration–time profile within the body. This profile is largely the result of two opposing processes, the rate of drug input into the body and the rate of drug elimination by the body. The study of the time course of drug in the body is termed pharmacokinetics and incorporates the processes of absorption, distribution, metabolism, and excretion (ADME). For most drugs, absorption and distribution determine the speed of onset of drug effect, while metabolism and excretion terminate the action of a drug by removing the active form (with the exception of prodrugs, which undergo biotransformation to the therapeutically active metabolite). The major pharmacokinetic parameters that describe the ADME processes are clearance ($CL$), volume of distribution ($V$), elimination half-life ($T_{1/2}$), and bioavailability ($F$). These pharmacokinetic parameters are key determinants in the design of a suitable dosage regimen, providing information about the magnitude and frequency of dosing (Fig. 1).

In drug development, it is estimated that approximately 40% of all new chemical entities (NCEs) fail to become clinically successful drugs due to undesirable pharmacokinetic properties [1–4]. The single greatest factor contributing to inappropriate pharmacokinetics is the process of metabolism. Most drugs are fat/lipid soluble, and metabolism plays a crucial role in their deactivation and the determination of their pharmacokinetic profile. For such drugs, the *in vivo* capacity for metabolism is measured by clearance, which has a major impact on the secondary pharmacokinetic parameters, such as half-life and oral bioavailability. A knowledge of the capacity for the metabolism of an NCE and an understanding of the human metabolism pathways and the enzymes involved have now become key components in the selection of compounds for clinical development. This knowledge is important not only for the optimization of the metabolic stability of the NCE but also to avoid

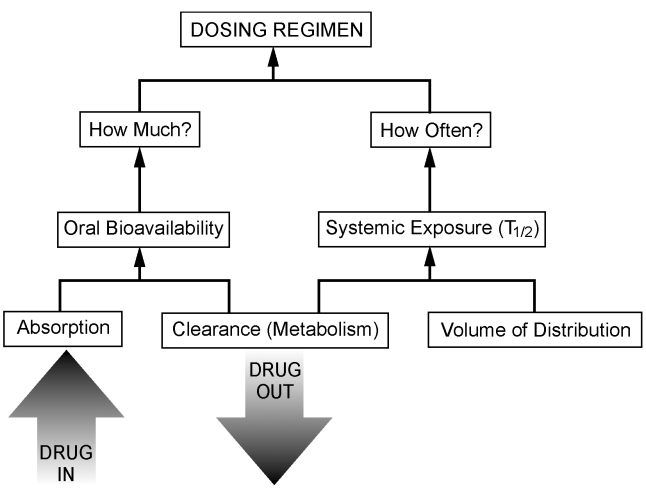


**FIGURE 1** Relationship between metabolism, pharmacokinetic parameters, and dosing regimen. (Adapted from Ref. 4.)

the likelihood that it is subjected to clinically significant genetic polymorphic metabolism, drug–drug interactions (either by induction or inhibition), or the formation of biologically active or reactive metabolites. This chapter appraises the relationship between metabolism and the resulting pharmacokinetic profile of a drug, and reviews the major factors that may have an impact on a drug's metabolism and may have a detrimental effect on its pharmacokinetics and clinical use.

## 2 IMPORTANT PHARMACOKINETIC PARAMETERS

A basic tenet of pharmacokinetics is that the magnitude of both the desired therapeutic response and/or toxicity of a drug is a function of its concentration in the body [5]. If concentrations are too low at the site of action, therapeutic failure may occur, and conversely, if concentrations are too high, adverse effects and toxicity may result. Thus, successful pharmacotherapy is very often dependent on maintaining the drug concentration within a defined range, known as the *therapeutic range* or *window* [5]. The best place to measure drug concentration presumably would be at the site of action, but most often this is not accessible, being in a tissue such as the brain or the heart, and a blood sample (usually plasma or serum) is used instead. At steady-state conditions when the rate of drug administration is balanced by the rate of drug elimination, the plasma concentration for many drugs is a relatively acceptable index of the concentration at the site of action.

Pharmacokinetics is the study of the drug/metabolite concentration–time profiles resulting from ADME, and the factors that may influence these processes. A knowledge of pharmacokinetics is crucial to determine a rational dosing regimen; that is, how much and how often a drug should be given to achieve the desired therapeutic effect. Rational dosing regimens can be established from several simple pharmacokinetic parameters mainly derived from drug plasma concentration–time curves. The key pharmacokinetic parameters to provide information on dose size and dose frequency are clearance, elimination half-life, and bioavailability. The volume of distribution may also be useful to calculate a loading dose at commencement of treatment in order to more rapidly achieve a particular target concentration. Obviously, the greater the therapeutic index (i.e., the separation of the concentration at which adverse effects appear and the minimum concentration required to produce an effective pharmacological response), the less important is the consideration of a drug's pharmacokinetics and the individualization of the dosage regimen.

### 2.1 Systemic Clearance ($CL_P$)

Systemic or total plasma clearance ($CL_P$) is arguably the most important of all pharmacokinetic parameters. Clearance is the most useful parameter for the evaluation of the elimination processes and has the most potential for clinical application. Clearance may be defined as the proportionality constant relating the rate of drug elimination to the plasma concentration ($C$): thus,

$$CL_P = \text{Elimination rate}/C \tag{1}$$

This equation can be re arranged to give

$$\text{Rate of drug elimination} = CL_P \times C \tag{2}$$

Alternatively, the clearance can be defined as the volume of plasma irreversibly cleared of drug per unit time, and usually it has units of L/h or mL/min [5]. Clearance is an index of the capacity for the elimination of an active drug from the body and may involve both metabolism and excretion processes, depending on the physicochemical properties of the drug. For example, a water-soluble drug, such as gentamicin, is eliminated almost entirely unchanged by the kidney, and its clearance will be determined entirely by the ability of the kidney to remove it from the blood flowing through the kidney. However, for most drugs, the total plasma clearance is the sum of clearances by the kidney ($CL_R$) and liver ($CL_H$), plus any other route ($CL_O$), and thus

$$CL_P = CL_R + CL_H + CL_O \tag{3}$$

The balance between these processes depends on the relative efficiencies of each process for a particular drug and the drug's physicochemical properties. However, most drugs are relatively lipid soluble (necessary for acceptable oral absorption and tissue distribution), and consequently, metabolism is the major determinant of elimination, usually converting them into inactive and more excretable forms. Thus, for most drugs, hepatic metabolism is the major component of the total plasma clearance.

For any particular organ, the physiological determinants of drug clearance are the organ blood flow, the inherent ability of the organ to extract the drug from the blood, and the extent of binding to plasma proteins and cells. Thus, an equation for the clearance of an eliminating organ, such as the liver, can be derived using the simplest (and most widely employed) model of hepatic clearance (known as the *well-stirred* or *venous equilibrium model*):

$$CL_H = Q_H \times \left[ \frac{fu \times CL_{int}}{Q_H + (fu \times CL_{int})} \right] \tag{4}$$

where $Q_H$ is the hepatic blood flow, $CL_{int}$ is the intrinsic clearance, and $fu$ is the fraction unbound in blood. This model assumes that the distribution of the drug is so fast in this highly vascular organ that the concentration of the unbound drug in the blood leaving the organ is equal to that within the organ. The full derivation of Eq. 4 can be found in Rowland and Tozer [5].

The $CL_{int}$ represents the intrinsic ability of the liver to remove (metabolize) the drug in the absence of any restrictions imposed by protein binding or blood flow to the liver. The term contained in the square brackets of Eq. 4 is referred to as the hepatic extraction ratio ($ER_H$) and is a measure of the efficiency of hepatic removal of drug and may range from 0 (when no drug is extracted) to 1, when the drug is completely removed during a single pass through the liver. Thus, Eq. 4 can be simplified to

$$CL_H = Q_H \times ER_H \tag{5}$$

Thus, if a drug has a very high intrinsic clearance compared to blood flow, the extraction ratio will tend toward 1, and the drug's hepatic clearance will approach the value of $Q_H$, the hepatic blood flow (but it can never be greater than this blood flow). Such drugs will exhibit blood-flow-dependent elimination kinetics. In this situation, the liver extraction processes or metabolizing enzymes are so active that the liver removes all, or nearly all, of the drug presented to it. Thus, the actual hepatic clearance is determined by the rate of supply of the drug to the liver, that is, the hepatic blood flow. As nearly all of the drug is already being removed, changes in enzyme activity make little or no difference to clearance. Similarly, protein binding is not important for such drugs as the liver has such high affinity/capacity for the drug that even bound drug can be stripped off the plasma proteins in a single pass through the liver. Thus, the hepatic clearance of a drug such as propranolol, lidocaine, and verapamil, for which the liver has a high metabolic capacity, is limited by hepatic blood flow and is uninfluenced by plasma protein binding [5]. The hepatic clearance of such drugs is relatively insensitive to changes in metabolizing activity brought about by induction or inhibition of hepatic metabolizing enzymes. However, a reduction in the rate of drug transport to the liver, such as a decrease in blood flow in heart failure, for example, would result in a decrease in hepatic clearance [5].

At the other extreme, when a drug's intrinsic clearance is very low compared to blood flow, then the hepatic extraction is low, and Eq. 4 for hepatic clearance reduces to

$$CL_H = fu \times CL_{int} \tag{6}$$

The hepatic clearance of such drugs is dependent only on enzyme activity and the unbound fraction in the blood. Thus, for a drug such as theophylline, for which the liver has a low metabolic capacity resulting in a low extraction ratio, its hepatic clearance will be affected by a change in enzyme activity but will not be influenced by changes in hepatic blood flow [5]. Thus, the hepatic extraction ratio may be used to roughly divide drugs into three categories: high ($ER_H > 0.7$), intermediate ($ER_H$ 0.3–0.7), and low ($ER_H < 0.3$) extraction ratio drugs. This classification of a NCE is useful in drug development to predict the influence of blood flow, enzyme induction and inhibition, and protein binding on its hepatic clearance and oral bioavailability (see below).

The extraction ratio across any organ may also be defined as the fraction of the drug removed from blood in one passage across that organ and may be determined by measuring the drug concentration in the plasma entering and exiting the organ. For example, $ER_H$ can be calculated using the following equation:

$$ER_H = (C_{in} - C_{out})/C_{in} \tag{7}$$

where $C_{in}$ is the drug concentration in the plasma entering the liver, and $C_{out}$ is drug concentration in the plasma exiting the liver.

In biochemical terms, the hepatic intrinsic clearance of a particular drug is a measure of how active the liver drug-metabolizing enzymes are with this drug as a substrate. Using the principles of enzyme kinetics and various *in vitro* enzyme systems (e.g., isolated perfused liver, hepatocytes, liver slices, human microsomes, S9

fraction, and various recombinant enzyme expression systems), it is possible to calculate a value for the *in vitro* intrinsic clearance. For example, using a one-site enzyme model and the Michaelis–Menten equation:

$$\text{Rate of metabolism (elimination)} = (V_{\max} \times C)/(K_{\mathrm{m}} + C) \tag{8}$$

where $V_{\max}$ is the maximal velocity of the reaction, $K_{\mathrm{m}}$ is the substrate concentration at which half maximal velocity is attained, and $C$ is the substrate concentration. When substrate concentration is much less than $K_{\mathrm{m}}$, Eq. 8 reduces to

$$\text{Rate of elimination} = (V_{\max}/K_{\mathrm{m}}) \times C \tag{9}$$

This ratio of $V_{\max}/K_{\mathrm{m}}$ estimated from *in vitro* data can be used as an approximation of the expected *in vitro* intrinsic clearance of a drug and usually has units of μL/min/mg of microsomal protein (if using a microsomal preparation). More recently in drug development, it has become popular to use the "depletion" method and measure the disappearance of an NCE over time in an *in vitro* system, and express this as a disappearance half-life, and then calculate intrinsic clearance from this [6–8]. The *in vitro* intrinsic clearance parameter can then be scaled up to the whole liver using various scaling parameters as suggested by a number of investigators [9, 10]. Furthermore, the latter can be extrapolated to predict an expected *in vivo* hepatic clearance value for that particular drug in humans. This can be done using a number of different models for the liver, including the *dispersion, parallel-tube*, or *well-stirred model*. The latter two are currently most often utilized because of their simpler mathematical basis. They differ from each other with regard to their assumption on how drug concentrations decline within the liver: the *well-stirred* model assumes the concentration to be homogeneous throughout the organ and to be reflected by the outflow concentration; whereas the parallel-tube model considers an exponential decline in drug concentration through the organ. The dispersion model is intermediate between those two with respect to drug mixing within the organ but is mathematically more complex. Further details on these models can be found in Refs. 9–11. In a comprehensive comparison of the three models, Ito and Houston [12] observed minimal differences between the models for the prediction of the *in vivo* hepatic clearance and suggested that the well-stirred model was adequate in a screening process to identify an NCE with a high or low clearance. The extrapolation of *in vitro* data to predict *in vivo* hepatic clearance has been extensively reviewed over recent years [13–15].

Acceptable correlations between *in vitro* and *in vivo* clearance values have been reported for a few lipophilic drugs, demonstrating the utility of these models/methods, but in general the quantitative predictions of *in vivo* clearance from *in vitro* metabolism data using existing models has been poor [11, 16]. It has been suggested that this is due to the lack of understanding of the interplay between the intrahepatocyte concentration and plasma and tissue binding, and the part played by drug influx and efflux by hepatocyte transporters [11]. Also, it is increasingly being recognized that many drugs exhibit "atypical" enzyme kinetics *in vitro* that cannot be adequately modeled by the basic Michaelis–Menten equation [17]. In order to improve the prediction of the relevant *in vivo* pharmacokinetic parameters, more complex models need to be applied to accommodate (1) more complex

enzyme kinetics (e.g., autoactivation, substrate inhibition, interactions between several binding sites on the enzyme); (2) incorporation of active transport mechanisms into, and within, the liver; and (3) tissue binding to both blood and liver components. It has been suggested that a physiologically based pharmacokinetic model might be more appropriate, as such models can incorporate more complex physiological and biochemical processes and are capable of predicting drug disposition quantitatively in different tissues in humans [17]. With the increased understanding of the mechanistic processes of ADME from *in vitro* studies, physiologically based pharmacokinetic modeling may prove to be a more valuable tool in the future [18].

In summary, in preclinical drug development, the prediction that a particular NCE has a high hepatic clearance forewarns that (1) it is likely to be subject to flow-dependent elimination (and may require dose reduction in situations of decreased blood flow such as in congestive heart failure), (2) is subject to high first-pass metabolism and thus will have low oral bioavailability (see Section 2.2), and (3) it may be more prone to presystemic drug or food metabolism interactions on oral dosing. In contrast, an NCE with a low hepatic clearance is more susceptible to protein binding changes (but note that the unbound concentration does not change, and thus no change in dose rate is required) and changes in metabolizing enzyme activity due to induction or inhibition.

### 2.2 Oral Bioavailability (*F*)

Orally administered drugs must be absorbed and pass sequentially from the gastrointestinal lumen through the gut epithelia and liver, before entering the systemic circulation. The proportion of an orally administered drug that reaches the systemic circulation intact is referred to as its bioavailability. Hence, bioavailability usually refers to the biologically available portion of the drug dose that may exert its pharmacological actions. Oral bioavailability is a function of the fraction of the dose absorbed, and the fraction of the drug that escapes extrusion back into the gut lumen by active efflux transporter systems and/or metabolism by intestinal and liver enzymes. The latter is referred to as the first-pass metabolism. Although metabolism is largely carried out in the liver, for some drugs, such as cyclosporine, there is evidence that a significant amount also occurs in the enterocytes in the gut wall and contributes to the overall first-pass effect [19, 20]. If a large proportion of a drug is metabolized by the epithelial cells of the gut or by the liver before it reaches the systemic circulation, this will result in a low bioavailability, and consequently, a much greater dose must be administered to compensate for this loss. If we assume all the drug is absorbed intact (i.e., fraction absorbed = 1), then the bioavailability ($F$) depends on the fraction escaping the first-pass extraction, and thus

$$F = 1 - ER_{\mathrm{I+H}} \tag{10}$$

where $ER_{\mathrm{I+H}}$ represents the combined extraction ratio for the intestines and the liver. If we consider a drug with a low extraction ratio, which is poorly metabolized by the liver (and the gut wall), nearly all the drug escapes the first-pass process, and bioavailability is essentially complete (as long as it is well absorbed from the gut). For example, if 3% of a theoretical drug is extracted by the liver, the resulting

bioavailability will be 97%. However, any major alteration in this extraction ratio by induction (e.g., a doubling of $ER_{I+H}$ from 3% to 6%) or inhibition (e.g., $ER_{I+H}$ halved from 3% to 1.5%) will have very little overall effect on bioavailability with a reduction from 97% to 94% on induction, and an increase from 97% to 98.5% with inhibition. Such changes in bioavailability would be insignificant in the clinical situation. In contrast, if a drug has a high $ER_{I+H}$ , most of the dose is extracted on the first pass through the liver, so that only a minor proportion reaches the systemic circulation intact. Thus, although induction or inhibition of the metabolizing enzymes would have only a small effect on $ER_{I+H}$, it would have a significant effect on the proportion escaping extraction and on the bioavailability. For example, a theoretical drug with a high $ER_{I+H}$ of 90% would have a bioavailability of 10%. However, if induction of the metabolizing enzymes increased this $ER_{I+H}$ to 95%, this would result in the halving of the bioavailability from 10% to 5%. Similarly, if the metabolizing enzymes were inhibited, decreasing the $ER_{I+H}$ to 80%, a doubling of the bioavailability would result. For many drugs, such changes in bioavailability would certainly be considered significant. Thus, for drugs with high extraction ratios, hepatic and intestinal enzyme activity are the major determinants of first-pass metabolism, and consequently their oral bioavailability. (However, note that after intravenous administration, hepatic blood flow is the major determinant of the systemic clearance of high extraction drugs.) After oral administration, the plasma concentrations of high extraction drugs, such as propranolol and verapamil, are more variable, both within and between individuals compared to the plasma concentrations of low $ER_H$ drugs, and small changes in the metabolism capacity can cause large changes in oral bioavailability [5, 20]. Obviously, if an NCE experiences a very high first-pass metabolism (i.e., $ER_{I+H} \approx 1$) after oral administration, an alternative route of administration may have to be found or the molecule redesigned to increase its metabolic stability and reduce its first-pass metabolism.

### 2.3 Volume of Distribution ($V$)

Along with plasma clearance, the volume of distribution is the other major independent pharmacokinetic parameter. The concept of distribution volume is used to relate the plasma drug concentration to the dose administered (assuming no elimination has occurred), or to the amount of drug in the body. For most drugs, it does not relate to a real physiological or anatomical space, but provides an estimate of the extent of the drug's distribution through body-fluid compartments and its uptake by tissues. A major determinant is the relative avidity of the drug for tissue components as compared to blood. Obviously, the failure of an NCE to distribute to the site of action in a tissue or a cellular compartment *in vivo* will render the compound inactive and necessitate the redesign of the compound to overcome this problem.

Several volume of distribution terms are used in pharmacokinetics and can be determined in a number of ways. The simplest method is to regard the body as a simple one-compartment model and find the plasma concentration at zero time by back extrapolation of the log concentration–time plot after an intravenous (IV) bolus injection. Many drugs and NCEs adhere to this simple model. However, other two- and three-compartment models may be necessary for an adequate description of the concentration–time profile of some drugs. Alternatively, a steady-state volume of distribution term ($V_{ss}$) can be derived from pharmacokinetic analysis based on

statistical moments [21]. The $V_{ss}$ term essentially corresponds to the plasma and tissues volumes in which a drug is distributed when steady-state conditions have been achieved. Further discussion on compartmental and noncompartmental pharmacokinetic approaches can be found in Section 3.

The volume of distribution is a useful parameter for determining the size of the loading dose, in order to get to the steady-state situation more rapidly after a constant rate infusion or repetitive dosing, but has no influence on the magnitude of the steady-state concentration achieved. This is solely determined by the rate of drug administration and the systemic clearance. Probably of more importance with regard to desirable pharmacokinetic characteristics is the influence of volume of distribution on a drug's elimination half-life, as discussed next.

### 2.4 Elimination Half-life ($T_{1/2}$)

The elimination half-life ($T_{1/2}$) is defined as the time taken for half the amount of drug present in the body to be eliminated by either metabolism or excretion. It has been suggested that the half-life of an NCE is a more important pharmacokinetic parameter than either plasma clearance or volume of distribution, with regard to guiding decisions in the drug discovery and development process [10]. Half-life is important because it is an indicator of duration of action for many drugs and thus determines how often a drug needs to be administered. Dosing frequency of most drugs depends on how long the active drug remains in the circulation, which is determined by its half-life. With a knowledge of half-life, the dosing frequency required to avoid large fluctuations in plasma concentrations over the dosage interval may be determined. For example, if a drug is rapidly absorbed and is given every half-life, the peak-to-trough concentration ratio will be ~2. However, if the half-life is short (e.g., less than 4 hours) and the drug exhibits a narrow therapeutic index (e.g., twofold), it is often difficult to administer frequently enough to avoid toxicity at the peak concentration, or loss of effect at the low concentrations before the next dose. In addition, significant compliance issues occur with a dosing frequency greater than once or twice a day [22].

For most drugs, the half-life is independent of concentration and is a function of its clearance (assuming first-order elimination kinetics) and volume of distribution as follows:

$$T_{1/2} = 0.7V/CL \tag{11}$$

where 0.7 is an approximation of the natural logarithm of 2.

Manipulation of an NCE's volume of distribution has not commonly been used to achieve a more favorable "longer" half-life. For example, it could be suggested that the volume of distribution might be enhanced by increasing the overall lipophilicity of an NCE, thus leading to a longer half-life. However, this improvement in volume of distribution might be offset by increased metabolism and clearance, as the binding sites of the metabolizing enzymes are generally lipophilic in nature and more readily accept lipophilic substrates. Thus, any gains in $T_{1/2}$ time due to a larger volume of distribution may be canceled out by increased metabolic activity. More commonly in drug development, strategies are employed to enhance metabolic stability by structural modifications of the NCE in order to decrease hepatic

clearance, and thus prolong the half-life, and yet maintain the optimal level of potency [23]. A longer half-life should allow lower and less frequent dosing, thus enhancing patient convenience and promoting better compliance. Additional advantages arising from enhanced metabolic stability of the NCE might also include increased oral bioavailability, reduction in the number (and perhaps significance) of possible active metabolites, and lower intra- and interpatient variability in plasma concentration–time profiles (which largely result from differences in metabolic capacity). However, such chemical modifications to improve metabolic stability may in some cases give rise to additional problems with properties such as reduced absorption or increased renal excretion, resulting in a lack of overall improvement in desirable pharmacokinetic properties [24].

An alternative and perhaps easier strategy for reducing dosing frequency for a drug with a short $T_{1/2}$ would be the development of a slow-release formulation [25]. This would slow the rate of drug absorption into the body, so that the magnitude of the concentration fluctuations during a dosing interval would be determined by the absorption rate rather than the elimination rate. A similar approach to lengthen the duration of drug action can make use of a prodrug. The synthesis of a prodrug with a relatively slow metabolic conversion to the active species would result in this conversion step being the rate-determining step in the disappearance of the active species, and would determine its plasma concentration–time profile and thus require less frequent dosing [26].

The elimination half-life of a drug is also useful in predicting the time required to achieve steady-state conditions during a continuous intravenous infusion or repetitive dosing. The maximum effect of many drugs might be expected once this steady-state equilibrium had been achieved during chronic dosing. This is generally accepted to occur after 4–5 half-lives. Similarly, when drug dosing is terminated, it takes 4 half-lives for more than 90% of the drug to be eliminated.

## 3 APPROACHES TO PHARMACOKINETIC ANALYSIS

### 3.1 Compartmental Analysis

The traditional approach to pharmacokinetic analysis involves the use of compartmental analysis, which focuses on the rates of absorption, distribution, and elimination, usually in terms of first-order rate constants (linear kinetics) [27]. This is known as the model-dependent approach. When the exchange of a drug between the plasma and tissue proceeds rapidly compared with the rate of elimination, the whole body may be considered mathematically as a single compartment. This one-compartment model is the simplest model and regards the body as a single, kinetically homogeneous unit, and the decline in plasma concentration after a single bolus intravenous injection can be described by a single exponential function. If the distribution of the drug out of the plasma is so slow that it cannot be disregarded, a model must then be considered, which contains a central compartment and at least one other peripheral compartment. Although these compartments lack physical or anatomical reality, for many drugs, the central compartment may correspond to the plasma or blood volume, together with the extracelular fluid of highly perfused tissues such as the heart, lungs, liver, kidneys, and endocrine glands. It is also pre-

sumed that the drug is eliminated solely from this central compartment. Drugs distribute within a few minutes through this compartment and equilibrium between plasma and tissues is rapidly established. The peripheral compartment is then formed by less perfused tissue such as muscle, skin, or adipose tissue into which the drug may enter more slowly. This combined effect of two compartments gives rise to a biphasic concentration–time curve after an intravenous bolus injection, with two distinct linear portions when drawn on a semilog ordinate scale. Although drug disposition to the peripheral compartment is slow, it is usually much faster than the elimination rate of the drug. The two-compartment model may be expanded to contain additional compartments, which can be described mathematically as the sum of as many individual exponential functions as there are relevant compartments. In practical terms, the sensitivity of the available assay technique for the determination of the drug concentration usually allows the definition of a maximum of three compartments. This approach to pharmacokinetic analysis allows mathematical simulations and the prediction of drug concentrations in plasma (representing the central compartment) and in other deep compartments. These models are usually robust for a variety of conditions, such as altered physiological or pathological conditions and during chronic medication [27].

### 3.2 Noncompartmental Analysis

An alternative and perhaps simpler approach to pharmacokinetic analysis is the use of noncompartmental analysis (the model-independent approach), where the drug pharmacokinetic parameters are determined using statistical moments and are derived from relationships involving the area-under-the-zero-moment-curve (*AUC*) and the area-under-the-first-moment-curve (*AUMC*) of the plasma drug concentration–time curve [27]. No complicated modeling of the disposition of the drug is required, but consequently it lacks the predictive and/or simulative power for drug concentration in the plasma and other tissue compartments [16]. Further discussion and a comparison of these two approaches to pharmacokinetic analysis can be found in Atkinson et al. [21].

### 3.3 Physiologically Based Analysis

Yet another more complex approach to pharmacokinetic analysis involves using physiologically based pharmacokinetic models. This method attempts to take into account the fundamental anatomical and physiological factors (e.g., blood flow to the different organs/tissues, organ uptake and active transport, and tissue binding) that influence drug uptake and disposition [18, 27]. Individual organs or tissue groups are linked in an anatomically correct fashion via the circulatory system, with drug elimination permitted from the relevant organs. This is by far the most realistic view of what is happening to the drug in the body but requires extensive computing power and data inputs (e.g., blood flow, partitioning or tissue binding data for all species). Examples of physiologically based pharmacokinetic models of varying complexity can be found in the review by Gerlowski and Jain [28]. Due to their complexity, these models have only had a limited application to clinical studies to date, but they are expected to have greater importance in the future [17, 27].

## 4 PREDICTING PHARMACOKINETIC PARAMETERS IN HUMANS

As mentioned earlier, very often *in vitro* drug metabolism information in isolation fails to scale up and provide robust predictions of the pharmacokinetics of an NCE in the human population. An alternative strategy for predicting plasma pharmacokinetic parameters of NCEs in humans involves the use of allometric scaling methods plus preclinical animal pharmacokinetic data [29].

### 4.1 Allometric Scaling

Interspecies allometric scaling is based on empirically observed relationships between physiological and biochemical parameters (e.g., basal metabolic rate, creatinine clearance, and cytochrome weight) and body weight among mammals [30]. As pharmacokinetic parameters are a function of biochemical and physiological processes, these have also been shown to scale with body weight across species [31, 32]. This has allowed the prediction of pharmacokinetic parameters, such as $CL$ and $V$, by using a simple power function of the form

$$Y = aB^n \tag{12}$$

where $Y$ is the dependent variable (e.g., clearance) and $B$ is the body weight, with $a$ and $n$ being the allometric coefficient and exponent, respectively. Allometric scaling from preclinical animal studies has proved to be useful for the prediction of the volume of distribution in humans, particularly if plasma protein differences between species are taken into account [33, 34]. However, allometric scaling has not proved to be particularly successful at predicting the hepatic clearance of drugs eliminated mainly by metabolism, especially those with low hepatic clearance. Adjustments to the basic allometric equation using maximum life span potential and brain weight have been proposed, but with limited success [34–36]. More recently, a combination of allometric scaling with *in vitro* human microsomal data has been proposed [36–38]. However, in a recently published study on the comparison of five different models (including allometric scaling with *in vitro* data input) using a dataset of 22 extensively metabolized compounds, Zuegge and colleagues [39] concluded that the most cost-effective and accurate approach for the prediction of hepatic clearance by metabolism was based on *in vitro* data alone, and that the inclusion of *in vivo* preclinical data and allometric scaling did not significantly improve prediction accuracy. They found that the prediction accuracy of the allometric models was at the lower end of all methods compared.

### 4.2 Microdosing

With the development of ultrasensitive analytical technologies (such as accelerator mass spectrometry (AMS), and positron emission tomography (PET)) with the capability of measuring drug and metabolite concentrations in the low picogram to femtogram range, human microdosing has been proposed as a new approach to obtaining crucial human pharmacokinetic data for comparing and selecting drug candidates earlier in their development [40–42]. Both technologies have the necessary sensitivity to follow the fate of trace amounts of radiolabeled drug in the body

after very small doses (e.g., 1% of a therapeutic dose, maximally <100 μg), which would be expected to have no biological activity. PET can provide real-time data on the distribution of a radiolabeled drug in the body and its availability at the site of action; whereas AMS combined with high performance liquid chromatography (HPLC) can quantitate extremely low concentrations of a drug and metabolites in body fluids at time intervals after dosing [43]. The essential aim of microdosing studies is to obtain early human pharmacokinetic information for new drug candidates to assist in the candidate selection process, rather than acquiring efficacy or safety data. The major question with regard to human microdosing is whether the pharmacokinetic parameters after microdoses replicate those observed at therapeutic doses. In a small trial of five diverse drugs eliminated by metabolism, Garner [44] reported a 70% correspondence between the pharmacokinetics after a microdose and a therapeutic dose. However, more in-depth studies are required to judge the predictive abilities of microdosing in humans, and whether it will play a significant role in drug development in the future is not known at present. The exquisite sensitivity and specificity of the combination of HPLC and AMS also raises the possibility of the "cocktail" or cassette-dosing approach, where several NCEs are administered simultaneously in microdoses, and their pharmacokinetics are followed concurrently. This allows for a larger number of compounds to be investigated, and more efficient and rapid identification of the compound with the most desirable pharmacokinetics in humans.

## 5 METABOLISM FEATURES THAT MAY CONVEY PHARMACOKINETIC DISADVANTAGES

As previously indicated, metabolism is responsible for determining the major pharmacokinetic parameters of clearance, bioavailability, and half-life for most drugs. These pharmacokinetic characteristics in turn are very important in defining the pharmacological and toxicological profile of an NCE and indicating to drug developers whether favorable pharmacokinetics can be expected in patients. Desirable pharmacokinetics in patients would include a suitable half-life to allow a convenient dosing frequency of once or twice daily, and low hepatic/intestinal metabolism to allow acceptable and reproducible oral bioavailability. However, it is also important to realize that, although the preclinical *in vitro* and *in vivo* data might predict favorable pharmacokinetics in humans, there are a number of factors (discussed later) that may have a significant impact on an NCE's metabolism, leading to undesirable pharmacokinetic characteristics in patients and ultimately compound failure.

### 5.1 Nonlinear Pharmacokinetics

In the preceding discussion on pharmacokinetics, it has been assumed that the disposition systems operating on the drug obey first-order kinetics. This means that clearance, volume of distribution, bioavailability, and elimination half-life are independent of drug concentration (and dose). Under these conditions, drug accumulation is a linear function of dose, and any change in dose rate will produce a proportional change in the steady-state concentration. Some drugs do not have such linear characteristics. The most common cause of nonlinearity is believed to be the

limited capacity of certain elimination processes, such as the drug-metabolizing enzyme systems. Most drugs are presented to the metabolizing enzymes at concentrations well below their maximal capacity, and they display first-order kinetics. However, for some drugs to be effective, concentrations are required at sufficiently high levels to cause saturation of the eliminating enzyme system, and the elimination occurs at a constant maximal rate, that is, zero-order kinetics. The anticonvulsant drug phenytoin and ethanol are probably the best known and studied examples of such compounds, where saturation of metabolism occurs with pharmacologically active concentrations [5, 45]. For such compounds, drug elimination rate will depend on the plasma concentration (or the magnitude of the dose), and drug accumulation with repetitive dosing will not be readily predictable. Once saturation of the eliminating system has been reached, the plasma concentration will increase disproportionately with subsequent dosing, and the time required to reach steady state will increase with the increasing half-life. If saturation of the elimination mechanisms is observed at pharmacologically active concentrations in the preclinical development of an NCE, structural modification would be recommended to remove this property, as this compound would be very difficult to use safely in patients, especially if it also had a low therapeutic index.

### 5.2 Drug–Drug Interactions

Drug–drug interactions are a major problem for the pharmaceutical industry and have resulted in drugs being withdrawn after being launched onto the worldwide market [46]. For example, terfenadine (a nonsedating antihistamine used for the treatment of rhinitis) was withdrawn after the occurrence of potentially fatal cardiac arrhythmias due to a substantial increase in its bioavailability when given with ketoconazole or grapefruit juice [46, 47]. Both ketoconazole and grapefruit juice are potent inhibitors of CYP3A4, which is the major enzyme involved in the metabolism of terfenadine and the termination of its activity. Indeed, the most common cause of clinically significant drug–drug interactions is primarily associated with the oxidation reactions catalyzed by the cytochrome P450 (CYP) enzyme family [48]. The CYPs in families 1–3 catalyze the metabolism of approximately 60–70% of all clinically used drugs. Of these, CYP3A4 is probably the most important, as it is involved in the metabolism of half of all drugs. Other important CYP isozymes involved in drug metabolism include CYP1A2, CYP2C9, CYP2C19, and CYP2D6. Inhibition of metabolism usually has a rapid onset and most commonly results in a dramatic increase in plasma concentration and an exaggerated response, with increased risk of toxicity and fatalities in extreme cases. The awareness of these adverse consequences has resulted in the development of a battery of automated screens to identify potent inhibitors of the CYP systems [8]. Some examples of drugs most commonly involved in inhibition of the CYP family include the azole antifungal drugs (e.g., ketoconazole), the quinolone and macrolide antibiotics (e.g., erythromycin), the selective serotonin reuptake inhibitors (e.g., fluoxetine, paroxetine, and fluvoxamine), the antiviral protease inhibitors (e.g., ritonavir and indinavir), and the antiulcer drug cimetidine [49].

The potent inhibition of the CYP enzyme family would be a property to be avoided in the development of an NCE, as it could lead to clinically significant drug–drug interactions. However, the decision to withdraw a drug from the market

after evidence of a drug–drug interaction depends on several factors, including the severity of the interaction, the disease being treated, and whether alternative drugs are available. For example, the calcium channel blocker mibefradil, which is a substrate and potent inhibitor of CYP3A4, was withdrawn as there were equally effective drugs available that did not possess this undesirable characteristic. In contrast, a number of protease inhibitors (e.g., indinavir, saquinavir, and ritonavir), which are also potent inhibitors of CYP3A4, were not withdrawn, as there was a lack of effective alternatives in the treatment of HIV infection [50].

Significant drug–drug interactions are not confined to the CYP metabolizing family, as illustrated by the inhibition of dihydropyrimidine dehydrogenase by the antiviral drug sorivudine. Dihydropyrimidine dehydrogenase is a key enzyme in the metabolism (inactivation) of the anticancer drug 5-fluorouracil, and its inhibition by sorivudine caused a reduction in 5-fluorouracil's clearance, leading to elevated concentrations and toxicity. The latter mechanism was believed to have resulted in a number of fatalities in Japan, and consequently sorivudine was withdrawn from the marketplace [51].

Induction of drug-metabolizing enzymes by an NCE may also be problematic in drug development. In contrast to inhibition, induction has a less rapid onset, taking 4–5 days to achieve a peak effect. Most commonly, this results in lower plasma concentrations and treatment failure of comcomitant drugs. Well-known inducing agents include the anticonvulsants (e.g., phenytoin, phenobarbital, and carbamazepine), the antibiotic rifampicin, and the anti-inflammatory agent dexamethasone. Both first-pass metabolism by the intestines and the liver can be dramatically increased by enzyme-inducing agents, resulting in decreased bioavailability and loss of the therapeutic response. The discovery that many compounds induce drug-metabolizing enzymes via an interaction with the nuclear hormone receptors, constitutive androstane receptor (CAR), and the pregnane-X-receptor (PXR) has prompted the development of enhanced throughput systems for identification of CYP inducers, using both PXR radioligand binding assays and also reporter gene technology [52, 53]. The availability of high resolution human PXR ligand-binding domain crystal structures also raises the possibility of *in silico* testing of NCEs for their potential to bind to PXR, as an indication of their induction ability [54].

Drugs likely to be the subject of clinically relevant drug–drug interactions tend to be drugs with a steep concentration–response curve or a narrow therapeutic window [20]. Thus, a relatively small increase in concentration can quickly lead to toxicity, or conversely a decrease can lead to the disappearance of the therapeutic effect. For most drugs, the absence of significant drug–drug interactions may be a crucial determinant of the ultimate clinical success of a drug. Thus, to avoid the potential of drug–drug interactions, it is desirable to develop a new drug candidate that is not a potent CYP inhibitor or inducer, and whose metabolism is not readily inhibited or induced by other drugs. In order to eliminate unfavorable compounds early in development, high throughput *in vitro* systems for identifying drug metabolism pathways, the enzymes involved, and inhibitor/induction potential of new drug candidates are a rapidly evolving area of drug development [8, 55]. However, the information from both induction and inhibition *in vitro* technologies must be integrated with *in silico* approaches to maximize the potential of both [55].

### 5.3 Polymorphic Drug Metabolism

Genetic differences in the expression of many metabolizing enzymes can have an impact on the pharmacokinetic characteristics of a drug in a population of patients and accounts for much of the interindividual variability associated with the efficacy and toxicity of a number of drugs [56]. There are many examples of polymorphisms in genes encoding drug-metabolizing enzymes, which can cause significant differences in drug response within a population (Table 1) [57, 58]. As with drug–drug interactions, most information is available for the CYP system, in particular, the polymorphic enzymes CYP2C9, CYP2C19, and CYP2D6, which catalyze the metabolism of approximately 40% of drugs undergoing phase I oxidation [59, 60]. In general, four phenotypes can be identified: poor metabolizers (PMs), who lack the functional enzyme; intermediate metabolizers (IMs), who are heterozygous for one deficient allele or carry two alleles that cause reduced activity; extensive metabolizers (EMs), who have two normal alleles; and ultrarapid metabolizers (UMs), who have multiple gene copies. Thus, the dosing requirement for a drug such as nortriptyline (which is a substrate for CYP2D6), may differ 10–20-fold among individuals. Consequently, using a "standard" dose in a European population would be expected to elicit no response in the 5% of the population who are UMs (due to too rapid metabolism of the drug), and perhaps elicit an adverse reaction in the 7% of the population who are PMs due to excessively high plasma concentrations [58]. The distribution frequency of these CYP2D6 phenotypes differs between ethnic groups. For example, 10% of Spaniards and 29% of Ethiopians have been reported to be UMs for the CYP2D6 polymorphism [61]. Similarly, polymorphisms in CYP2C19 have been shown to have clinical consequences resulting in toxicity, or lack of efficacy with many drugs, including *S*-mephenytoin, omeprazole, imipramine, and proguanil [58]. Population studies have indicated that CYP2C19 PMs occur in 3–5% of Caucasians and African Americans, 12–23% in most Asian groups, and 38–79% in Polynesian and Micronesian populations [59]. In a Japanese study, patients with peptic or duodenal ulcers were treated with a dual therapy of low dose omeprazole

**TABLE 1 Examples of Genetic Polymorphisms Influencing Drug Metabolism**

| Gene[a] | Substrates |
|---|---|
| CYP2C9 | Tolbutamide, diclofenac, warfarin, phenytoin |
| CYP2C19 | Omeprazole, lansoprazole, diazepam, proguanil |
| CYP2D6 | Antidepressants, codeine, β-blockers |
| CYP2A6 | Coumarin |
| UGT1A1 | Irinotecan |
| NAT2 | Isoniazid, sulfonamide |
| TPMT | 6-Mercaptopurine, 6-thioguanine, azathioprine |
| DPD | 5-Fluorouracil |
| ALDH2 | Acetaldehyde |

[a]UGT1A1, UDP-glucuronosyltransferase 1A1; NAT2, N-Acetyltransferase 2; TPMT, thiopurine S-methyltransferase; DPD, dihydropyrimidine dehydrogenase; ALDH2, aldehyde dehydrogenase 2.

*Source*: Adapted from Ref. 56.

and amoxicillin [62]. Cure rates were highly dependent on CYP2C19 phenotype, with only 25% in EMs, rising to 60% in IMs, and 100% in PMs. The high cure rate in PMs was thought to be due to lack of metabolism, resulting in higher omeprazole plasma concentrations and a longer half-life. These examples illustrate some of the problems that arise with drugs that are eliminated mainly by polymorphic enzymes, especially their use in different ethnic populations. This is important for the global perspective of the pharmaceutical industry, and most companies now take the pharmacogenetic aspects of CYP metabolism into account during drug development. Obviously, an NCE that has a high affinity for a polymorphic metabolizing enzyme that plays a major role in its elimination would be less favorable than one that did not display this characteristic [63]. However, as costs decrease, predictive genotyping of patients may be justified for drugs subject to significant polymorphic metabolism, especially those with a narrow therapeutic index. For example, it has been suggested that the prospective use of the UGT1A1 genotype assay would identify patients with the UGT1A1*28 genotype, which is associated with reduced metabolism of SN-38 (the active metabolite of the anticancer drug irinotecan), and increased incidence of significant toxicity [64]. Such genotyping may help to ensure safer and more effective therapy in the future.

### 5.4 Formation of Active/Reactive Metabolites

The rapid and extensive hepatic metabolism of an NCE may lead not only to pharmacokinetic disadvantages but also to issues with the identification of metabolites and evaluation of their biological activity. In general, *in vivo* metabolism leads to the formation of inactive and more readily excretable metabolites, but in some cases, metabolites with potent biological activity or toxic properties are produced, which may influence the outcome of therapy. The function and fate of these metabolites need to be studied, in order to gain a full insight into the potential biological consequences of exposure to the NCE. Metabolism to reactive intermediates is one of the major mechanisms by which drugs exert toxic effects. Despite a huge investment and many years of research, it is still not possible at present to accurately predict the potential for toxicity of a compound that has been shown to undergo metabolic activation [65]. It is important in drug development to investigate the metabolic fate of an NCE and evaluate the biological properties of possible metabolites. The relative merits of current and potential strategies for dealing with metabolite characterization have recently been reviewed by Nassar and Talaat [66]. The advancement of robotic systems using *in vitro* incubations of an NCE with human microsomes, hepatocytes, or liver slices, accompanied by the analysis of metabolites using techniques such as liquid chromatography–mass spectrometry (LC-MS) and/or liquid chromatography–nuclear magnetic resonance (LC-NMR) has made this more applicable in practice. In particular, significant advances in metabolite identification have been achieved using high performance liquid chromatography coupled with various mass analysers, such as triple quadrupole, quadrupole time-of-flight, ion trap, and quadrupole linear ion trap mass spectrometers [67, 68]. However, with the improved sensitivity and efficiency of NMR technology, the complementary nature of NMR and MS in definitive structure identification is being realized, and combined LC-MS-NMR systems will become crucial in the elucidation and confirmation of metabolite identification in the near

future. Combined LC-MS-NMR systems are in the midst of an evolutionary phase at present that will lead to greatly improved ability and efficiency in metabolite identification [66]. However, although structural information on metabolites can be a considerable asset for enhancing and streamlining the process of NCE development, in order to eliminate potentially toxic compounds early in the development process, it must be integrated with computational technologies to obtain maximum benefits [66, 69, 70]. Comprehensive databases of metabolism and toxicological information are available, and software packages can indicate possible metabolites and identify structural features that may confer adverse properties to molecules [4, 66, 67]. With the advent of improved algorithms and the availability of ever-increasing computing power, the future of drug metabolism and toxicity prediction is envisaged as an integration of structure–activity relationship models based on compound structure with "pattern" databases of tissue or organ response to drugs, and "systems biology" databases of metabolic pathways, genes, and regulatory networks [71, 72].

## 6 CONCLUSION

For many drugs, metabolism plays a central role in defining the concentration–time profile in the body and, consequently, the magnitude and duration of the biological response. We have only a very limited understanding of the relationships between a molecule's structure and its physicochemical properties and its likely metabolic fate in the body. Drug metabolism may occur in many different tissues in the body and may be catalyzed by a large variety of diverse enzymes, which can be influenced by innumerable factors. The capacity for the metabolism of a compound *in vivo* is usually measured by the hepatic clearance, which is arguably the most important pharmacokinetic parameter, with a major influence on the secondary parameters of half-life and oral bioavailability. Clearance, half-life, and bioavailability are key components in the design of a suitable oral dosage regimen, providing information on the magnitude and frequency of dosing. Thus, metabolism is the single most important factor contributing to inappropriate or undesirable pharmacokinetic characteristics in an NCE. This relationship between an NCE's metabolic profile (rate, extent, and multiplicity of resulting metabolites) and the overall systemic clearance in a patient is extremely complex and provides a formidable challenge for the prediction of this parameter in humans from preclinical *in vitro* human data and animal studies. There are many issues concerning the nonphysiological nature of the *in vitro* screening systems, interspecies differences in metabolizing enzymes and their regulation, the part played by active transporters in tissue distribution, and the complexities of *in vivo* pharmacokinetics and toxicity. Nevertheless, some progress has been made over the past ten years with the development of a more diverse range of investigative tools, including superior technology to measure and identify compounds and their metabolites at extremely low concentrations, the automation and expansion of various *in vitro* screening tests to allow more rapid understanding of the ADME fate of a compound, and *in silico* approaches based on molecular modeling, structure–activity relationships, and metabonomics. To fully realize the potential of these advances, there is the need to integrate these technologies into a single comprehensive environment for drug development. A small amount of prog-

ress has been made, but much is yet to be done to attain the ultimate goal of being able to predict human metabolism and pharmacokinetics (and toxicity) based solely on the structure of a novel NCE.

## REFERENCES


1. Prentis RA, Lis Y, Walker SR. Pharmaceutical innovation by seven UK-owned pharmaceutical companies (1964–1985). *Br J Clin Pharmacol* 1988;25:387–391.
2. Thompson TN. Early ADME in support of drug discovery: the role of metabolic stability studies. *Curr Drug Metab* 2000;1:215–241.
3. Palmer A. New horizons in drug metabolism, pharmacokinetics and drug delivery. *Drug News Perspect* 2003;16:57–62.
4. Van de Waterbeemd H, Gifford E. ADMET *in silico* modeling: Towards prediction paradise? *Nat Rev* 2003;2:192–204.
5. Rowland M, Tozer TN. *Clinical Pharmacokinetics: Concepts and Applications*, 3rd ed. Baltimore: Williams and Wilkins; 1995, pp 1–184.
6. Bachmann KA, Ghosh R. The use of *in vitro* methods to predict *in vivo* pharmacokinetics and drug interactions. *Curr Drug Metab* 2001;2:299–314.
7. Jones HM, Houston JB. Substrate depletion approach for determining *in vitro* metabolic clearance: time dependencies in hepatocyte and microsomal incubations. *Drug Metab Dispos* 2004;32:973–982.
8. Riley RJ, Grime K. Metabolic screening *in vitro*: metabolic stability, CYP inhibition and induction. *Drug Discov Today: Technol* 2004;1:365–372.
9. Iwatsubu T, Hiroja J, Ooie T, Suzuki H, Shimada N, Chiba K, Ishizaki T, Green CE, Tyson CA, Sugiyama Y. Prediction of *in vivo* drug metabolism in the human liver from *in vitro* metabolism data. *Pharmacol Ther* 1997;73:147–171.
10. Obach SR, Baxter JG, Liston TE, Silber BM, Jones BC, MacIntyre F, Rance DJ, Wastall P. The prediction of human pharmacokinetic parameters from preclinical and *in vitro* metabolism date. *J Pharmacol Exp Ther* 1997;283:46–59.
11. Masimirembwa CM, Bredberg U, Andersson TB. Metabolic stability for drug discovery and development. *Clin Pharmacokinet* 2003;42:515–528.
12. Ito K, Houston JB. Comparison of the use of liver models for predicting drug clearance using *in vitro* kinetic data from hepatic microsomes and isolated hepatocytes. *Pharm Res* 2004;21:785–792.
13. Houston JB, Galetin A. Progress towards prediction of human pharmacokinetic parameters from *in vitro* technologies. *Drug Metab Rev* 2003;35:393–415.
14. Tracy TS, Hummel MA. Modeling kinetic data from *in vitro* drug metabolism enzyme experiments. *Drug Metab Rev* 2004;36:231–242.
15. Ito K, Houston JB. Prediction of human drug clearance from *in vitro* and preclinical data using physiologically based and empirical approaches. *Pharm Res* 2005;22:103–112.
16. Smith DA, van De Waterbeemd H. Pharmacokinetics and metabolism in early drug discovery. *Curr Opin Chem Biol* 1999;3:373–378.
17. Theil F-P, Guentert TW, Haddad S, Poulin P. Utility of physiologically based pharmacokinetic models to drug development and rational drug discovery candidate selection. *Toxicol Lett* 2003;138:29–49.
18. Schmitt W, Willmann S. Physiology-based pharmacokinetic modeling: ready to use. *Drug Discov Today Technol* 2004;1:449–456.

19. Wacher VJ, Salphati L, Benet LJ. Active secretion and enterocytic drug metabolism barriers to drug absorption. *Adv Drug Deliv Rev* 2001;46:89–102.
20. Gibbs MA, Hosea NA. Factors affecting the clinical development of cytochrome P450 3A substrates. *Clin Pharmacokinet* 2003;42:969–964.
21. Atkinson AJ, Daniels CE, Dedrick RL, Grudzinskas CE, Markey SP. *Principles of Clinical Pharmacology*. San Diego: Academic Press; 2001, pp 29, 75–90.
22. Katzung BG. *Basic and Clinical Pharmacology*, 9th ed. New York: Lange Medical Books/ McGraw-Hill; 2004, pp 1095–1096.
23. Nassar A-EF, Kamel AM, Clarimont C. Improving the decision-making process in the structural modification of drug candidates: enhancing metabolic stability. *Drug Discov Today* 2004;9:1020–1028.
24. Smith D, Schmid E, Jones B. Do drug metabolism and pharmacokinetic departments make any contribution to drug discovery? *Clin Pharmacokinet* 2002;41:1005–1019.
25. Thombre AG. Assessment of the feasibility of oral controlled release in an exploratory development setting. *Drug Discov Today* 2005;10:1159–1166.
26. Testa B. Prodrug research: futile or fertile. *Biochem Pharmacol* 2004;68:2097–2106.
27. Mathiowitz E. *Encyclopedia of Controlled Drug Delivery*. Hoboken, NJ: John Wiley & Sons; 1999, pp 833–851.
28. Gerlowski LE, Jain RK. Physiology-based pharmacokinetic modeling: principles and applications. *J Pharm Sci* 1983;72:1103–1127.
29. Mahmood I, Balian D. Interspecies scaling: predicting clearance of drugs in humans. Three different approaches. *Xenobiotica* 1996;26:887–895.
30. Davidson IWF, Parker JC, Beliles RP. Biological basis for extrapolation across mammalian species. *Regul Toxicol Pharmacol* 1986;6:211–237.
31. Boxenbaum H. Interspecies scaling, allometry, physiological time, and the ground plan of pharmacokinetics. *J Pharmacokinet Biopharm* 1982;10:201–227.
32. Paxton JW. The allometric approach for interspecies scaling of pharmacokinetics and toxicity of anticancer drugs. *Clin Exp Pharmacol Physiol* 1995;22:851–854.
33. Paxton JW, Kim SN, Whitfield LR. Pharmacokinetic and toxicity scaling of the antitumor agents, amsacrine and a new analog (CI-921) in mice, rats, rabbits, dogs and humans. *Cancer Res* 1990;50:2692–2697.
34. Mahmood I, Balian JD. The pharmacokinetic principles behind scaling from preclinical results to Phase I protocols. *Clin Pharmacokinet* 1999;36:1–11.
35. Boxenbaum H, Fertig JB. Scaling of antipyrine intrinsic clearance of unbound drug in 15 mammalian species. *Eur J Drug Metab Pharmacokinet* 1984;9:177–183.
36. Lavé TL, Coassolo P, Reigner B. Prediction of hepatic metabolic clearance based on interspecies allometric scaling techniques and *in vitro–in vivo* correlations. *Clin Pharmacokinet* 1999;36:211–231.
37. Mahmood I. Prediction of clearance in humans from *in vitro* human liver microsomes and allometric scaling. A comparison study of the two approaches. *Drug Metab Drug Interact* 2002;19:49–64.
38. Keldenich J. Prediction of human clearance (CL) and volume of distribution (VD). *Drug Discov Today Technol* 2004;1:389–395.
39. Zuegge J, Schneider G, Coassolo C, Lave T. Prediction of hepatic metabolic clearance: comparison and assessment of prediction models. *Clin Pharmacokinet* 2001;40:553–563.
40. Aboagye EO, Price PM, Jones T. *In vivo* pharmacokinetics and pharmacodynamics in drug development using positron-emission tomography. *Drug Discov Today* 2001;6: 293–302.

41. Lappin G, Garner RC. Big physics, small doses: the use of AMS and PET in human microdosing of development drugs. *Nat Rev Drug Discov* 2003;2:233–240.
42. Wilding IR, Bell JA. Improved early clinical development through human microdosing studies. *Drug Discov Today* 2005;10:890–894.
43. White INH, Brown K. Techniques: the application of accelerator mass spectrometry to pharmacology and toxicology. *Trends Pharm Sci* 2004;25:442–447.
44. Garner RC. Less is more: the human microdosing concept. *Drug Discov Today* 2005;10: 449–451.
45. Winter ME. *Basic Clinical Pharmackinetics*, 4th ed. Philadelphia: Lippincott Williams & Wilkins; 2004, p 321.
46. Li AP. Screening for human ADME/Tox drug properties in drug discovery. *Drug Discov Today* 2001;6:357–366.
47. Dahan A, Altman H. Food–drug interaction: grapefruit juice augments drug bioavailability—mechanism, extent and relevance. *Eur J Clin Nutr* 2004;58:1–9.
48. Lin JH, Lu AYH. Inhibition and induction of cytochrome P450 and the clinical implications. *Clin Pharmacokinet* 1998;35:361–390.
49. Smith DA, Abel SM, Hyland R, Jones BC. Human cytochrome P450s: selectivity and measurement *in vivo*. *Xenobiotica* 1998;28:1095–1128.
50. De Maat MMR, Ekhart GC, Huiteman ADR, Koks CHW, Mulder JW, Beijnen JH. Drug interactions between antiretroviral drugs and comedicated agents. *Clin Pharmacokinet* 2003;42:223–282.
51. Okuda H, Ogura K, Kato A, Takubo H, Watabe T. A possible mechanism of eighteen patient deaths caused by interactions of sorivudine, a new antiviral drug, with oral 5-fluorouracil prodrugs. *J Pharmacol Exp Ther* 1998;287:791–799.
52. Waxman DJ. P450 gene induction by structurally diverse xenochemicals: central role of nuclear receptors CAR, PXR and PPAR. *Arch Biochem Biophys* 1999;369:11–23.
53. Willson T, Kliever SA. PXR, CAR and drug metabolism. *Nat Rev Drug Discov* 2002;1: 259–266.
54. Goodwin B, Redinbo MR, Kliewer SA. Regulation of CYP3A gene transcription by the pregnane X receptor. *Annu Rev Pharmacol Toxicol* 2002;42:1–23.
55. Rostami-Hodjegan A, Tucker G. "In silico" simulations to assess the *in vivo* consequences of *in vitro* metabolic drug–drug interactions. *Drug Discov Today Technol* 2004;1: 441–448.
56. Hiratsuka M, Mizugaki M. Genetic polymorphisms in drug metabolising enzymes and drug targets. *Mol Genet Metab* 2001;73:298–305.
57. Pirmohamed M, Park BK. Genetic susceptibility to adverse drug reactions. *Trends Pharmacol Sci* 2001;22:298–305.
58. Ingelman-Sundberg M. Pharmacogenetics of cytochrome P450 and its applications in drug therapy: the past, present and future. *Trends Pharmacol Sci* 2004;25:193–200.
59. Goldstein J. Clinical relevance of genetic polymorphisms in the human CYP2C subfamily. *Br J Clin Pharmacol* 2001;52:349–355.
60. Kirchheiner J, Brockmoller J. Clinical consequences of cytochrome P450–2C9 polymorphisms. *Clin Pharmacol Ther* 2005;77:1–16.
61. Bondy B, Zill P. Pharmacogenetics and psychopharmacology. *Curr Opin Pharmacol* 2004;4:72–78.
62. Furuta T, Ohashi K, Kamata T, et al. Effect of genetic differences in omeprazole metabolism on cure rates for *Helicobacter pylori* infection and peptic ulcer. *Ann Intern Med* 1998;129:1027–1030.

63. Walker DK. The use of pharmacokinetic and pharmacodynamic data in the assessment of drug safety in early drug development. *Br J Clin Pharmacol* 2004;58:601–608.
64. Kiang TKL, Ensom MHH, Chang TKH. UDP-glucuronosyltransferases and clinical drug–drug interactions. *Pharmacol Ther* 2005;106:97–132.
65. Evans DC, Watt AP, Nicoll-Griffith DA, Baille TA. Drug–protein adducts: an industry perspective on minimizing the potential for drug activation in drug discovery and development. *Chem Res Toxicol* 2004;17:3–16.
66. Nasser A-EF, Talaat RE. Strategies for dealing with metabolite elucidation in drug discovery and development. *Drug Discov Today* 2004;9:317–327.
67. Anari MR, Baille TA. Bridging cheminformatic metabolite prediction and tandem mass spectrometry. *Drug Discov Today* 2005;10:711–717.
68. Rossi DT, Sinz MW. *Mass Spectrometry in Drug Discovery*. New York: Marcel Dekker; 2002, pp 271–336.
69. Yu H, Adedoyin A. ADME-Tox in drug discovery: integration of experimental and computational technologies. *Drug Discov Today* 2003;8:852–861.
70. Nasser A-EF, Kamel AM, Clarimont C. Improving the decision-making process in structural modification of drug candidates: reducing toxicity. *Drug Discov Today* 2004:9: 1055–1064.
71. Bugrim A, Nikolskaya T, Nikolsky Y. Early prediction of drug metabolism and toxicity: systems biology and modeling. *Drug Discov Today* 2004;9:127–135.
72. Whittaker PA. What is the relevance of bioinformatics to pharmacology. *Trends Pharmacol Sci* 2003;24:434–439.

# EXPERIMENTAL DESIGN CONSIDERATIONS IN PHARMACOKINETIC STUDIES

William W. Hope, Vidmantas Petraitis, and Thomas J. Walsh
*National Cancer Institute, National Institutes of Health, Bethesda, Maryland*

**Contents**

1 Introduction
2 Defining the Overall Goal of the Experiments
3 Experimental Platforms
4 Impact of Drug Measurement on Study Design
5 Choosing the Dosage Range and the Number of Dosages Used to Define Pharmacokinetic Relationships
6 Sampling Times
  6.1 Defining the Number and Timing of Informative Sampling Points
  6.2 Peak Concentrations
  6.3 When Precisely to Acquire Pharmacokinetic Data
  6.4 Optimal Sampling and D-Optimal Design
7 Determination of Drug Concentrations in Tissues and Bodily Fluids
8 Conclusion
References

## 1 INTRODUCTION

Pharmacokinetics is the study of the spatial and temporal distribution of drugs. Pharmacokinetic studies are conducted in experimental contexts for two major reasons. The first is to define pharmacokinetic relationships for new drugs that may

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

be beneficial for patients. This process is important for preclinical drug development where one wants to develop candidate molecules with pharmacokinetic profiles that are likely to be favorable for clinical use or that meet predefined developmental targets. A further understanding of the pharmacokinetics in experimental systems is required to obtain prior information before proceeding to human Phase I and II studies. The second reason to conduct pharmacokinetic studies is to further understand the concentration–effect relationships (i.e., pharmacodynamic relationships). For both these purposes, an accurate description of pharmacokinetics is vital in order to obtain appropriate conclusions regarding drug behavior and utility. To achieve this goal, appropriate experimental design is essential.

This chapter outlines some of the issues in the design of experimental models, which enable pharmacokinetic relationships to be established. The objective is to estimate the pharmacokinetic relationships as precisely as possible, at minimum cost, and using the minimum possible number of animals and analytical samples. Although this subject matter is vast, we focus our discussion on antimicrobial pharmacology from which broader principles may be extrapolated.

## 2 DEFINING THE OVERALL GOAL OF THE EXPERIMENTS

The experimental goals and aims will, to a large extent, dictate the design of the pharmacokinetic experiments. One obviously wishes to design the experiment to yield data to address the problems in hand. In this regard, a clear idea as to the overall experimental goal is required. This along with at least some preliminary pharmacodynamic data will determine (1) the concentrations of drug that are likely to induce a pharmacological effect of interest and are therefore relevant for the pharmacokinetic experiments; (2) the length of the experimental period over which one requires pharmacokinetic information; (3) the likely schedules of drug which will be employed; (4) the experimental conditions that will be used (e.g., severity of the model, immune status of the model, or other physiological derangements).

Studying the appropriate dosages and schedules for the right period of time and in the right model are important because (1) the pharmacokinetics may change as a function of dose (i.e., become nonlinear, especially at higher dosages when clearance mechanisms become saturated); (2) one needs to know about the pharmacokinetics over the entire course of the experimental period rather than for a mere fraction; (3) there may be drug accumulation with different schedules of administration, which may have a bearing on the pharmacodynamic relationships; and (4) the pharmacokinetic relationships may be different in different models. The immunological status, model severity, or degree of physiological derangement may significantly alter the pharmacokinetics of any given compound. In the circumstances in which the severity of the model precludes intensive blood sampling, an approach may be to define pharmacokinetic parameters in healthy animals and cross reference data from a select number of informative points from the animals with the disease in question—this is discussed in more detail in Section 6.4.

Thus, the experimental conditions used in the pharmacokinetic model should replicate as closely as possible those used to define the exposure–response relationships. Given the cost of pharmacokinetic experiments, a prudent approach is to

perform only these experiments when one is reasonably sure of the experimental conditions that will be used for other aspects of the study.

## 3 EXPERIMENTAL PLATFORMS

The design of the pharmacokinetic experiment will vary according to the type of experimental platform. In small animals, such as mice, repeated sampling from a single animal is extremely difficult. In this circumstance, pharmacokinetic data can be acquired by designing a serial sacrifice study, in which a cohort of animals receiving a given dose of drug are sacrificed at predefined time points over the course of the experimental period. Blood from mice can be obtained from a terminal cardiac puncture after the administration of a general anesthetic. Alternatively, while technically more challenging, serial blood samples can be obtained in the same mouse from the saphenous vein or from a retro-orbital puncture. The principal disadvantage of such an approach is the relatively small volume of blood that can be procured from each mouse. The number of mice sacrificed at each time point depends on the confidence with which one wishes to estimate the pharmacokinetic relationships. Usually a minimum of three animals is required at each time point, but four to six may be appropriate if the drug concentrations are likely to be variable, or precise estimates of the pharmacokinetic parameters are required. Using such a design, one has point estimates of measures of central tendency (e.g., mean concentration) and the associated dispersions at each time point for each dose. The observed variance from the replicate samples can be used as a weighting function in the pharmacokinetic modeling process. A similar "serial sacrifice" design obviously applies for other experimental platforms in which the experiment is destroyed in the sampling process.

In larger animals, such as rabbits, or *in vitro* models such as hollow fiber systems, multiple samples may be taken from a single animal or via a sampling port. Clearly, however, there is a limit in terms of the volume of blood that can be drawn from an animal over time; in rabbits, for example, this volume is approximately 7 mL/kg. Studying pharmacokinetics in this manner is similar to intensive pharmacokinetic sampling strategies in humans. The data from such an experimental design can be analyzed in one of two ways: (1) the mean and standard deviation for a group of individual animals receiving a given dose of drug in which blood has been drawn at the same time point can be determined; the analysis in this circumstance is similar to murine models, described above, in which the mean drug concentration can be weighted using the inverse of the observed variance at each observation point; (2) a population methodology can be employed; in this case there is no attempt to calculate the mean and standard deviation of the pharmacokinetic data between groups of rabbits. Rather, the data from each rabbit are modeled on an individual basis. In this case, weighting is assumed to be inversely proportional to the estimated assay variance. This relationship can be established by a regression of the mean values and the corresponding standard deviations of samples of known concentration, which encompass the dynamic range of the assay and which have been run in at least quadruplicate. The principal advantage of this approach is that one is able to obtain estimates for both the central tendencies (mean and median values) and dispersions for the population as a whole, *in conjunction* with an optimal set of

pharmacokinetic parameters for each individual animal. Such an approach may be advantageous when one wishes to directly link drug exposure with effect within the same animal. The ability to sample repeatedly from one animal means that fewer large animals are required to establish pharmacokinetic relationships than is the case in smaller animals such as mice. Furthermore, the sacrifice of a larger animal may not be necessary after an appropriate washout period, depending on the compound in question.

## 4 IMPACT OF DRUG MEASUREMENT ON STUDY DESIGN

The practical limitations of drug analysis have an important bearing on the design of experimental pharmacokinetic experiments. Drug levels are most frequently measured using high performance liquid chromatography (HPLC). Other methods such as bioassay or ELISA may be used. Increasingly, however, mass spectrometry is used because of its analytical sensitivity and the possibility of high throughput analysis. One aspect of drug measurement that is critical to the design of pharmacokinetic experiments is the lower limits of detection and quantification. Knowledge of these limits may have an influence on the design of pharmacokinetic experiments in the following ways: (1) there is little point in studying dosages that produce concentrations of drug beneath the limit of detection; (2) as one approaches the limit of quantification, the variance in most analytical assays increases, thus directly influencing the confidence with which the pharmacokinetic parameters can be estimated; and (3) the inability to measure low concentrations of drug may lead to erroneous estimates of the pharmacokinetics of a given drug. This may lead to underestimations of the magnitude of drug exposure that develops following the administration of an otherwise efficacious dose; these concepts are depicted in Fig. 1.

If it is clear that the analytical sensitivity of the assay is suboptimal and precludes the ability to accurately describe the pharmacokinetics, then a number of approaches

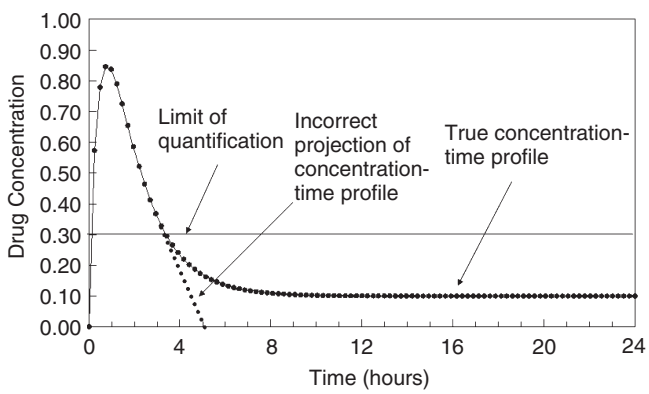


**FIGURE 1** The impact of the limit of quantification on the ability to accurately describe the terminal phase of distribution. The figure shows the incorrect projection of the concentration–time profile at times beneath the limit of quantification, which leads to an underestimate of total drug exposure.

are possible. First, an attempt can be made to improve the limit of quantification via the optimization of the chosen analytical method (e.g., for HPLC analysis changing the mobile phase, detection settings, or the sample injection volume may improve the analytical sensitivity and the limit of quantification). Second, a more sensitive analytical method can be tried—in the majority of cases this means resorting to mass spectrometry. Third, tissue levels of drug can be measured; such an approach may be particularly useful for drugs that undergo extensive tissue distribution and that also exhibit prolonged mean residence times within tissues. By simultaneously comodeling serum and tissue drug concentrations, one may be able to estimate serum concentrations of drug that are well below the limit of detection. Such an approach is exemplified by a recent paper by Louie and colleagues [1] who were able to obtain accurate estimates of the antifungal agent caspofungin, by comodeling serum and kidney drug concentrations.

## 5 CHOOSING THE DOSAGE RANGE AND THE NUMBER OF DOSAGES USED TO DEFINE PHARMACOKINETIC RELATIONSHIPS

The range of dosages used to determine the pharmacokinetic relationships should encompass those that will be used to define the pharmacodynamic relationships. The importance of studying more than one dose is to determine whether the pharmacokinetics are linear or nonlinear. Following the administration of progressively higher dosages, clearance mechanisms may become saturated, and nonlinear pharmacokinetics will be observed. This is especially important if nonlinear kinetics are observed with dosages that are planned for the pharmacodynamic experiments. Ideally, a minimum of three dosages should be studied, but depending on the drug more may be appropriate.

An insight into whether a drug displays linear pharmacokinetics can be achieved in the following ways: (1) the presence of a linear system can be inferred using the principle of superposition; in this case, the drug concentrations following different dosages divided by the administered dose should be superimposable; (2) there should be a linear relationship between the area under the concentration–time curve (AUC) and the different dosages; (3) the pharmacokinetic parameters determined for each dosage group should not vary significantly from dose to dose; and (4) a more sophisticated method involves fitting and comparing both linear and nonlinear pharmacokinetic models to the entire dataset. A fit of one model compared with another, which is superior in a statistically significant manner, can be used to infer whether the pharmacokinetics are linear or nonlinear. An example of this process applied to human data is illustrated in a study by Lodise and colleagues [2], where an attempt was made to distinguish whether the beta-lactam piperacillin was better accounted for by a linear of nonlinear pharmacokinetic model.

## 6 SAMPLING TIMES

### 6.1 Defining the Number and Timing of Informative Sampling Points

The identification of sampling times is a critical component in the ability to obtain confident estimates of the pharmacokinetic parameters. In general, given the

considerable time and resources involved in determining pharmacokinetic relationships, one wishes to sample as little as possible, while still ensuring acceptable estimates of the pharmacokinetic parameters will be obtained. The key in this regard is the identification of informative sampling points. The information content at various time points throughout the dosing interval is not constant. The distribution of information content is a function of the shape of the concentration–time profile of a drug and the structural pharmacokinetic model that is used to obtain parameter estimates. Importantly, the trough concentration, which is frequently used in therapeutic drug monitoring, is the least informative data point for describing pharmacokinetics, since it provides little information regarding the shape of the preceding concentration–time profile. Some idea as to likely informative points can be obtained from prior knowledge of drug behavior, even if acquired in other experimental platforms. Several experiments may be required to fully characterize the concentration–time profile for a drug in early development or to define the pharmacokinetics within a new experimental system. Informative points tend to occur at points of inflection in the concentration–time curve and can be formally identified using optimal design theory (see Section 6.4).

## 6.2 Peak Concentrations

Peak concentrations, unlike troughs, tend to be information rich. In order to capture this information, sampling in the immediate period following drug administration is frequently performed. While this approach is standard and useful, there are some important caveats. (1) The use of compartmental pharmacokinetic modeling techniques requires an assumption of full compartmental mixing; this refers to the presence of homogeneous concentrations of drug throughout the compartment in question. The time required for full mixing to occur, even for drugs that are injected intravenously, is not instantaneous. Sampling should be delayed until full mixing is likely to have occurred; this time depends on both the route of administration and the blood circulation time. A balance must be struck between waiting for mixing to occur, but not waiting too long that important information contained within the period immediately following drug administration is lost. (2) Drug concentrations are changing very rapidly in the early period following drug administration. Consequently, relatively small errors in the time of sampling are likely to have a significant impact on the observed plasma drug concentrations. This may be an issue for experimental reproducibility and fitting pharmacokinetic models to the data obtained immediately following drug administration. (3) In attempting to capture peak concentrations, there may be considerable logistical issues in administering drug to a relatively large number of animals, and then sampling 5–15 minutes later. This is especially the case for mice, where the number of animals tends to be high, and blood samples are usually obtained by cardiac puncture. Even in the hands of experienced personnel, the time taken to obtain a blood sample from an individual mouse is approximately 1 minute. Small departures from this demanding schedule may significantly disrupt the ability to obtain samples at the desired times. Thus collecting data immediately following drug administration requires careful planning, organization, and cooperation among co-workers.

### 6.3 When Precisely to Acquire Pharmacokinetic Data

An issue arises as to whether pharmacokinetic relationships should be defined after the initiation of drug administration or at steady state (if the latter is achieved within the study period). There is no universally correct approach, but there are two guiding principles when planning a pharmacokinetic study. (1) An attempt should be made to collect informative data throughout the study period, regardless of the timing. Consequently, it may be appropriate to sample relatively intensively during the early period following drug administration, again at steady state, and at other select informative times between these periods. Such an approach may help to identify the extent of drug accumulation, and any nonlinearities that may appear with repeated drug administration. (2) If sampling occurs after repeated dosing (e.g., following the sixth dose of drug), the determination of the concentration of drug immediately prior to drug administration should be defined (i.e., a prelevel should be drawn). If this data point is not modeled or the preceding dosages are not recorded as inputs into the system, then accurate estimates for the pharmacokinetic parameters will not be obtained—information regarding prior system inputs and initial conditions cannot simply be ignored.

### 6.4 Optimal Sampling and D-Optimal Design

D-optimal design theory has been applied extensively to humans as a means of acquiring informative data. In circumstances in which intensive sampling is not possible (e.g., neonates, patients with critical illness), D-optimal design allows for samples to be drawn at times when system information is near maximal. Such an approach allows confident estimates of the pharmacokinetic parameters using a minimal dataset. Such an approach can also be applied to experimental models, to minimize the number of samples required to be drawn and analyzed. In our laboratory, rabbits with invasive pulmonary aspergillosis (a fungal infection of the lung) are unable to tolerate intensive plasma sampling. An approach successfully employed by us is to acquire a relatively information-rich dataset in healthy animals and use these data to identify a limited number of informative sampling points in infected rabbits, thus allowing the pharmacokinetic parameters to be estimated while minimizing the burden of repeated blood sampling.

Optimal design can be performed using the sample module of the pharmacokinetic program ADAPT II of D'Argenio and Schumitsky [3]. The mean parameter values and the associated standard deviations are entered, along with the dosages of drug and the number of sampling times that are required. The identification of informative sampling times may also be important in other contexts, in which one wishes to ensure or demonstrate that the administration of a drug results in a concentration–time profile that is comparable to what was defined in a previously similar but subtly different model. Examples include the following: (1) ensuring that there is no pharmacokinetic interaction between drugs given in combination; in this circumstance, one may choose a small number of informative sampling times to demonstrate that the drug concentrations are the same as when the drugs are administered in isolation; (2) if there has been a departure from the conditions in which the original pharmacokinetic relationships were defined, but to a degree that is unlikely to have a significant effect on the pharmacokinetics, one may wish to

obtain a small number of samples in order to confirm that the concentration–time profiles are similar. For example, the original pharmacokinetic relationships may have been defined in an infection model using a particular strain or species; subsequently, one may wish to study a related species, but the experimental conditions are otherwise identical, and a full pharmacokinetic study may not be justified. In this situation it may be possible to sample at several informative points to ensure that the concentration–time profile is the same.

## 7 DETERMINATION OF DRUG CONCENTRATIONS IN TISSUES AND BODILY FLUIDS

Frequently, an assumption is made that there is a linear relationship between the concentrations of drug in the central compartment and concentrations that develop at the effect site (i.e., the site of the drug receptor). In the majority of cases this is true, meaning that serum and tissue drug concentrations do in fact track one-to-one. This enables the response at the effect site to be described in terms of the concentration–time profile within the central compartment. On occasions, however, a better understanding of drug effect can be obtained by measuring the concentration of drug at the effect site itself. Despite a linear relationship, knowledge of the concentrations at the effect site may be important for the following reasons: (1) the shape of the concentration–time profile may be different in the two compartments and for certain drugs and drug classes this may have a bearing on drug effect; (2) there may be differences in protein binding between the serum and the effect site, which could affect the free fraction of drug available to interact with its receptor; (3) there may be dissociation between the concentration–time profiles because of delayed trafficking of drug to the effect site, resulting in hysteresis and leading to an apparent dissociation between serum drug concentrations and the observed effect; (4) despite a linear relationship between the serum and the effect site, there may be an insufficient concentration of drug to elicit a biological response at the effect site; and (5) there may be unique and unappreciated transport mechanisms from the serum to the effect site—an example of this is the concentration that the macrolide group of antibiotics achieve in the epithelial lining fluid (ELF) of the lung, despite extremely low serum drug concentrations. Thus, on occasions, there may be a requirement to measure the concentration of drug in tissues or bodily fluids in order to link drug concentrations at the effect site and the drug effect itself.

The measurement of tissue drug concentrations is not straightforward. The important issues include the following. (1) Tissue from normal animals (i.e., devoid of drug) is required to generate a standard curve, which is used to determine the concentrations of drug within tissues. Ideally the assay needs to be optimized and validated for each tissue in order to define the limit of quantification. (2) In order to obtain an estimate of the time course (pharmacokinetics) of drug within the tissue compartment, serial sampling (if possible) or serial sacrifice experiments are required. The shape of the concentration–time profile may differ from the serum, meaning that different sampling times from those used in serum may be required. Animals should be serially sacrificed throughout the dosing interval. This means the numbers of animals that are required to determine the pharmacokinetics in tissues may be greater than if serum pharmacokinetics alone are determined. This may have

important implications for the cost and feasibility of such a study. (3) One should always bear in mind the complexities of distribution of drug within tissues. The effect of a drug results from an interaction with its receptor, which may only exist within subcompartments. A tissue homogenate, on the other hand, consists of a mixing of all the different compartments within tissue, including the vascular space. This principle is exemplified in a study by Groll et al. [4], where the compartmentalization of intrapulmonary lipid formulations of amphotericin B was characterized in pulmonary alveolar macrophages and epithelial lining fluid. Consequently, drug levels determined from tissue homogenates should be viewed as an approximation of the concentration of drug available to the receptor.

There are a number of additional advantages to simultaneously studying tissue and serum concentrations of drug. The comodeling of serum and tissue data may enable better estimates of pharmacokinetic parameters. A good example of this was provided by Louie et al. [1], who described the pharmacokinetics of the antifungal drug caspofungin in a murine model of invasive candidiasis. Caspofungin is cleared rapidly from the serum and is avidly taken up by tissues. The lower limit of quantification of caspofungin using HPLC is relatively high. The concentrations of drug in the kidney and serum were measured and comodeled. This approach enabled a more accurate estimate of the pharmacokinetics; the caspofungin half-life increased when the serum and tissue data were comodeled. This approach also enabled a circumvention of the problem induced by a relatively high limit of quantification mentioned earlier. Instead of moving to a more sensitive analytical assay to enable a more accurate description of the terminal elimination phase, this was achieved by measuring and comodeling drug concentrations in the kidney. In this circumstance, as drug concentrations in the serum decline beneath the lower limit of quantification, drug is still "seen" returning from peripheral compartments that are acting as a repository of drug (in this case, the kidney).

## 8 CONCLUSION

An accurate description of the pharmacokinetic properties of a drug underpins a further understanding of concentration–effect relationships. The design of pharmacokinetic experiments is critical to ensure that pharmacokinetic relationships can be accurately estimated. If the design of pharmacokinetic experiments is flawed, and the resultant pharmacokinetic models are incorrect, then erroneous conclusions regarding the concentration–effect relationship may result. This chapter summarizes our approach to the design of pharmacokinetic experiments and highlights ways in which pharmacokinetic relationships can be efficiently and accurately estimated.

## REFERENCES

1. Louie A, Deziel M, Liu W, Drusano MF, Gumbo T, Drusano GL. Pharmacodynamics of caspofungin in a murine model of systemic candidiasis: importance of persistence of caspofungin in tissues to understanding drug activity. *Antimicrob Agents Chemother* 2005;49: 5058–5068.

2. Lodise TP Jr, Lomaestro B, Rodvold KA, Danziger LH, Drusano GL. Pharmacodynamic profiling of piperacillin in the presence of tazobactam in patients through the use of population pharmacokinetic models and Monte Carlo simulation. *Antimicrob Agents Chemother* 2004;48:4718–4724.
3. D'Argenio DZ, Schumitzky A. *ADAPT II. A Program for Simulation, Identification, and Optimal Experimental Design. User Manual*. Los Angeles: Biomedical Simulations Resource, University of Southern California; 1997. http://bmsr.esc.edu/.
4. Groll AH, Lyman CA, Petraitis V, Petraitiene R, Armstrong D, Mickiene D, Alfaro RM, Schaufele RL, Sein T, Bacher J, Walsh TJ. Compartmentalized intrapulmonary pharmacokinetics of amphotericin B and its lipid formulations. *Antimicrob Agents Chemother* 2006;50:3418–3423.

# BIOAVAILABILITY AND BIOEQUIVALENCE STUDIES

Alexander V. Lyubimov[1] and Ihor Bekersky[2]

[1]*University of Illinois at Chicago, Chicago, Illinois*
[2]*Consultant, Antioch, Illinois*

**Contents**

1 Introduction
  1.1 Rationale for Bioavailability/Bioequivalence (BA/BE) Studies
  1.2 Need and Aim of BA/BE Studies
  1.3 Role of BA/BE Studies in Drug Development and Nonclinical Studies
  1.4 Regulatory Considerations and Guidelines
  1.5 Importance of Safety Margins
  1.6 Definitions
2 IND-Targeted Nonclinical Development
  2.1 Objective
  2.2 Species in IND-Targeted Toxicology Studies
  2.3 Key Elements of IND-Target Development
  2.4 Use of Animal Models for the Selection of Dosage Forms for Nonclinical Toxicology Studies
  2.5 Species Differences in Bioavailability
3 NDA-Targeted Nonclinical Development
  3.1 Toxicokinetics in Chronic Toxicity Studies
  3.2 Use of Animals to Evaluate the BA/BE of Clinical Dosage Forms
  3.3 Role and Impact of Nonclinical Studies on Clinical Development: Animal to Humans
4 Conclusions
  References

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

# 1 INTRODUCTION

The two questions that confront the pharmaceutical scientist during the process of developing a molecule into a drug are "Is it safe?" and "Is it therapeutic?" The question of safety is an important issue that surfaces very early in development; without a clear understanding of what defines safe use, it is difficult, if not impossible, to continue development. Thus, animal toxicology/safety studies are the primary ingredient in any NDA submission. However, such studies are just as important in the conduct of an IND-enabling program, which allows the conduct of the first human clinical studies.

Today, the process of developing a new chemical entity (NCE) into a drug that is approved for marketing follows a well-defined and regulated pathway. The key issues that will determine approval are safety and efficacy in the Phase III trials. In this context, preclinical animal studies provide both the initial and subsequent "windows" on toxicity safety and thus have a direct impact on decisions as to the first clinical trials and on a dosing regimen in such trials. In fact, concerns about safety and efficacy are evaluated throughout the whole drug development process.

In spite of reservations that can be raised about the extrapolation of animal data to humans, toxicology studies provide the best currently available assessments of what can reasonably be expected in humans. The designs obtained from toxicology studies are thus of particular concern since safety predictions based on the data obtained cannot be found to be unwarranted in retrospect. The challenge to the pharmaceutical scientist therefore is to use available "tools" that provide information on toxicity and drug disposition in the animals used in toxicology/safety studies. Regardless of the stage in development, the systemic exposure (e.g., the AUC) is the metric by which the relationship of drug disposition and toxicity is evaluated.

In this context, the inclusion of biopharmaceutics and pharmacokinetics into preclinical development, specifically into the toxicological/safety studies, is a logical step since drug concentration data defines the extent of exposure to drugs.

## 1.1 Rationale for Bioavailability/Bioequivalence (BA/BE) Studies

Bioavailability and bioequivalence are pharmacokinetic (PK) terms that quickly gained importance in the "early" days of the development of pharmacokinetics as a drug development science. As such, their use was primarily targeted to clinical development. Bioavailability is defined in the *Code of Federal Regulations*, Title 21, 320.1 [1] as the rate and extent to which the active ingredient or active moiety is absorbed from a drug product and becomes available at the site of action. For drug products that are not intended to be absorbed into the bloodstream, bioavailability may be assessed by measurements intended to reflect the rate and extent to which the active ingredient or active moiety becomes available at the site of action.

In clinical use, it is not unusual that several dose strengths or different dosage forms (capsules, tablets) are used; therefore a study(ies) is (are) conducted to ascertain that the same plasma concentration time profile is obtained at the different strengths or dosage forms, thus providing the appropriate drug exposure for patient management—that is, bioequivalence. Bioequivalence is often used in clinical trials for comparison of the reference (pioneer) and generic drugs. The *Code of Federal*

*Regulations*, Title 21, Volume 5 [2] defines bioequivalence as the absence of a significant difference in the rate and extent to which the active ingredient or active moiety in pharmaceutical equivalents or pharmaceutical alternatives becomes available at the site of drug action when administered at the same molar dose under similar conditions in an appropriately designed study. Where there is an intentional difference in rate (e.g., in certain controlled release dosage forms), certain pharmaceutical equivalents or alternatives may be considered bioequivalent if there is no significant difference in the extent to which the active ingredient or moiety from each product becomes available at the site of drug action. This applies only if the difference in the rate at which the active ingredient or moiety becomes available at the site of drug action is intentional and is reflected in the proposed labeling, is not essential to the attainment of effective body drug concentrations on chronic use, and is considered medically insignificant for the drug.

The extent of product bioavailability is estimated by the area under the blood concentration versus time curve (AUC). AUC is most frequently estimated using the linear trapezoidal rule. Other methods for AUC estimation may be proposed and should be accompanied by appropriate literature references during protocol development.

For a single-dose bioequivalence study, AUC should be calculated from time 0 (predose) to the last sampling time associated with quantifiable drug concentration $AUC_{0\text{-}LOQ}$. The comparison of the test and reference product value for this noninfinity estimate provides the closest approximation of the measure of uncertainty (variance) and the relative bioavailability estimate associated with $AUC_{0\text{-}INF}$, the full extent of product bioavailability.

In a multiple-dose study, the AUC should be calculated over one complete dosing interval ($AUC_{0\text{-}t}$). Under steady-state conditions, $AUC_{0\text{-}t}$ equals the full extent of bioavailability of the individual dose $AUC_{0\text{-}INF}$ assuming linear kinetics. For drugs that are known to follow nonlinear kinetics, the sponsor should consult with the Center for Veterinary Medicine (CVM) to determine the appropriate parameters for the bioequivalence determination [7].

The rate of absorption is estimated by the maximum observed drug concentration ($C_{max}$) and the corresponding time to reach this maximum concentration ($T_{max}$). When conducting a steady-state investigation, data on the minimum drug concentration (trough values) observed during a single dosing interval ($C_{min}$) should also be collected. Generally, three successive $C_{min}$ values should be provided to verify that steady-state conditions have been achieved. To determine a steady-state concentration, the $C_{min}$ values should be regressed over time and the resultant slope should be tested for its difference from zero [7].

Confidence intervals should routinely be used to interpret bioequivalence data. The following bioavailability and bioequivalence standards are usually required:

- The 90% confidence interval of the relative mean AUC of the test reference formulation should be within 80–125%.
- The relative mean measured $C_{max}$ of the test to reference formulation should be within 80–125%.
- The relative mean measured $C_{min}$ of the test to reference formulation should be within 80–125%.

### 1.2 Need and Aim of BA/BE Studies

In preclinical toxicology studies, the selection of the dosage form is not as complicated as for clinical use and this will be addressed subsequently. The issue of doses is of importance as the selection of doses for the various types of toxicology studies is an occult process that involves varying amounts of soft data. However, in order to provide a good prospective assessment/projection regarding safety in humans based on systemic exposure, it is clear that BA/BE principles also apply to animal toxicity studies.

In terms of bioavailability, toxicology studies usually entail three or four dose levels; and BA is "relative" as systemic exposure is evaluated by the linearity (or nonlinearity) of eliminating the kinetic at each dose level and/or as a function of time during the toxicity study. Considering that the dose range in toxicology studies is usually a large multiple of the clinical dose, a relative BA is acceptable in data evaluation of drug concentration and toxic effect(s).

Bioequivalence in preclinical animal studies is a more complex issue and in practical terms applies mostly to dog and primate studies where different oral dosage forms may be evaluated (solutions, capsules of neat drug, galenic tablets) to ensure a maximal drug release profile and thus a definition of the "line" between pharmacology and toxicology.

It is important to know that the drug is bioavailable at the selected route of administration, and that the absence of toxic or pharmacologic effect is real and not just a result of the drug not being absorbed.

In short, both concepts have an application to animal toxicology studies as the exposure needs to be determined and confirmed in order to calculate the animal to human safety margins on which clinical development is dependent.

### 1.3 Role of BA/BE Studies in Drug Development and Nonclinical Studies

As stated earlier, the concepts of BA/BE were first used in clinical development as tools to define a clear and consistent therapeutic use. Considering that the objective of toxicology studies is to demonstrate the toxic potential of the drug across a wide dose range, over varying durations of administration, and across different dosage forms and at least two species, the utility of BA/BE is readily apparent as a means to provide the best prospective assessment of animal safety data in terms of a safe exposure in humans. This is especially true in early stages of development where there is limited exposure/safety data in humans available.

### 1.4 Regulatory Considerations and Guidelines

Bioavailability and/or bioequivalence studies play a key role in the drug development period for both new drug products and their generic equivalents. For both, these studies are also important in the postapproval period in the presence of certain manufacturing changes. The article by Chen et al. [3] reviews the regulatory science of bioavailability and bioequivalence and provides the FDA's recommendations for drug sponsors who intend to establish bioavailability and/or demonstrate bioequivalence for their pharmaceutical products during the developmental process

or after approval. Statistic and pharmacokinetic principles of bioavailability and bioequivalence trials are described in Rajaram et al. [4]. This document was prepared as a set of workshop material for the International Congress on Medical and Care Compunetics in June 2004 in Den Hague, The Netherlands.

Toxicokinetics (TK) is the inclusion of pharmacokinetic principles into the toxicology/safety studies conducted in animal species. The objectives of TK are to evaluate drug absorption, distribution, and elimination as a function of dose level; to aid in the selection of realistic dose levels that can be maintained over the duration of the study; to determine the duration of exposure to drug; to determine the effect of chronic drug administration on the absorption and disposition of that drug; and to aid in interpreting and extrapolating the toxicological response in different species.

With few exceptions, toxicology/safety studies are conducted in accordance with the Good Laboratory Practice (GLP) regulations [5]; thus data collection and recording are conducted in a manner that is acceptable to all dossier reviewing regulatory agencies and are subject to verification by a QA unit. In terms of providing commonly acceptable objectives and "suggestions" as to the study design, one can refer to FDA and/or ICH guidelines. A tabulation of these guidelines is presented in Table 1. There are no specific preclinical BA/BE in either FDA or ICH guidelines; however, most of the definitions and approaches applicable for preclinical BA/BE studies can be found in two FDA documents: *Bioavailability and Bioequivalence Studies for Oral Administered Drug Products—General Considerations* [6] and *Bioequivalence Guidance*, Center for Veterinary Medicine [7].

### 1.5 Importance of Safety Margins

As stated earlier, the extent of systemic drug exposure in animal toxicology studies is directly correlated to the determination of safety margins to calculate a safe dose in humans. The first human clinical trials are concerned with administering a safe dose without covert toxicity. The challenge is that only animal toxicity/safety data are available. This may now change with the introduction of Phase 0 studies.

A comprehensive approach to the estimation of the human dose from animal data is available [8]. In short, having NOAELs (mg/kg) from animal studies, one can extrapolate a human equivalent dose based on body surface area, from the tables provided in the guidance [8]; a safety factor is chosen and the human equivalent dose is divided by that factor.

The result is the maximum recommended starting dose. A default safety factor of 10 is normally used, and although higher factors provide more flexibility for the strength of the clinical dose, factors lower than 10 do not necessarily mean a clinical hold if other data/information are available as an aid in the assessment of the projected safety in humans. It should be kept in mind that the NOAEL is not a static number and is redetermined as the duration of the toxicology studies becomes longer in keeping with the clinical objectives and study duration. It should also be pointed out that regardless of the place in development, it is important to know the drug release profile from the dosage forms used in the toxicity/safety studies (i.e., the bioavailability) in order to estimate the animal to human safety margins with assurance.

**TABLE 1 FDA and ICH Guidelines**

| Agency | Title | Date |
|---|---|---|
| FDA | *Bioavailability and Bioequivalence Studies for Oral Drug Products—General Considerations* | March 2003 |
| | *Bioequivalence Guidance* | 9 October 2002 |
| | *Exploratory IND Studies* | 18 January 2006 |
| | *Carcinogenicity Study Protocol Submissions* | 23 May 2002 |
| | *Format and Content of the Nonclinical Pharmacology/Toxicology Section of an Application* | 1 February 1987 |
| | *Immunotoxicology Evaluations of Investigational New Drugs* | 1 November 2002 |
| | *Photosafety Testing* | 7 May 2003 |
| | *Single Dose Acute Toxicity Testing for Pharmaceuticals—Revised* | 26 August 1996 |
| | *Reference Guide for the Nonclinical Toxicity Studies of Antiviral Drugs Indicated for the Treatment of N/A Non-Life Threatening Disease: Evaluation of Drug Toxicity Prior to Phase 1 Clinical Studies* | 1 February 1989 |
| | *Integration of Study Results to Access Concerns About Human Reproductive and Developmental Toxicities* | 13 November 2002 |
| | *Nonclinical Safety Evaluation of Pediatric Drug Products* | 3 February 2003 |
| | *Nonclinical Studies for Development of Pharmaceutical Excipients* | 18 May 2005 |
| | *Statistical Aspects of the Design, Analysis, and Interpretation of Chronic Rodent Carcinogenicity Studies of Pharmaceuticals* | 8 May 2001 |
| ICH | *M3—Nonclinical Safety Studies for the Conduct of Human Clinical Trials for Pharmaceuticals* | 25 November 1997 |
| | *S1A—The Need for Long-Term Rodent Carcinogenicity Studies of Pharmaceuticals* | 1 March 1996 |
| | *S1B—Testing for Carcinogenicity in Pharmaceuticals* | 23 February 1998 |
| | *S1C—Dose Selection for Carcinogenicity Studies of Pharmaceuticals* | 1 March 1995 |
| | *S1C(R)—Dose Selection for Carcinogenicity Studies of Pharmaceuticals: Addendum on a Limit Dose and Related Notes* | 4 December 1997 |
| | *S2A—Specific Aspects of Regulatory Genotoxicity Tests for Pharmaceuticals* | 24 April 1996 |
| | *S2B—Genotoxicity: Standard Battery Testing* | 21 November 1997 |
| | *S3A—Toxicokinetics: The Assessment of Systemic Exposure in Toxicity Studies* | January 1995 |
| | *S4A—Duration of Chronic Toxicity Testing Animals (Rodent and Nonrodent Toxicity Testing)* | 25 June 1999 |
| | *S5B—Detection of Toxicity to Reproduction for Medicinal Products* | 22 September 1994 |
| | *S6—Preclinical Safety Evaluation of Biotechnology Derived Pharmaceuticals* | 18 November 1997 |
| | *S7A—Safety Pharmacology Studies for Human Pharmaceuticals* | 13 July 2001 |
| | *S7B—Safety Pharmacology Studies for Assessing the Potential for Delayed Ventricular Repolarization (QT Interval Prolongation) by Human Pharmaceuticals* | 14 June 2002 (Draft) |
| | *S8—Immunotoxicity Studies for Human Pharmaceuticals* | April 2006 |

**TABLE 2 Acronyms and Definitions**

| Acronym | Definition |
|---|---|
| BA | Bioavailability means the rate and extent to which the active ingredient or active moiety is absorbed from a drug product and becomes available at the site of action. For drug products that are not intended to be absorbed into the bloodstream, bioavailability may be assessed by measurements intended to reflect the rate and extent to which the active ingredient or active moiety becomes available at the site of action (21 CFR 320.1). |
| BE | Bioequivalence means the absence of a significant difference in the rate and extent to which the active ingredient or active moiety in pharmaceutical equivalents or pharmaceutical alternatives becomes available at the site of drug action when administered at the same molar dose under similar conditions in an appropriately designed study. |
| AUC | Area Under the Curve is the integral of drug plasma concentration over a specific time ($AUC_t$) or to infinity ($AUC_{inf}$) and is a measure of drug amount absorbed (systemic absorption) into the systemic circulation. |
| GLP | Good Laboratory Practices Regulation. |
| PK | Pharmacokinetics. |
| NDA | New Drug Application. |
| QA | Quality Assurance. |
| NOAEL | No-Observed-Adverse-Effect-Level. |
| MRSD | Maximum Recommended Starting Dose. |
| HED | Human Equivalent Dose. |
| TK | Toxicokinetics. |
| IND | Investigative New Drug Application. |
| CMC | Controls, Manufacture, and Chemistry. |

### 1.6 Definitions

As is the case with many texts, the use of acronyms is a common way to avoid repetitive use of a concept. A partial tabulation of acronyms and definitions of terms that are germane to the present text is presented in Table 2.

## 2 IND-TARGETED NONCLINICAL DEVELOPMENT

The filing of an IND is the first step in the drug development process. The IND contains nonclinical pharmacology in animal disease models, toxicology in two animal species of duration that does not exceed the duration of the proposed first clinical trial(s), preliminary ADME data, an Investigators Brochure, protocol synopsis of the proposed clinical trial, and CMC information.

### 2.1 Objective

The objective of the IND is thus to obtain FDA "approval" to proceed with the conduct of the proposed clinical trial, that is, study the molecule in the intended model—humans. Prior to the IND filing it is usual to have a pre-IND meeting with the reviewing division, where what the IND will contain is presented and FDA input

is solicited as to its "acceptance." In most cases the first clinical study, being a PK/tolerance study, is in normal volunteers. However, for some indications this study can be conducted in a small number of patients and be of a short duration. In general, the duration of the toxicology studies, the dose ranges used, the NOAELs obtained, and the safety margins that are estimated determine the final clinical protocol for the initial human study(ies).

### 2.2 Species in IND-Targeted Toxicology Studies

Although it is not our purpose to discuss the format or content of an IND dossier, it is useful to present an outline of the IND to indicate the position and role of nonclinical toxicology studies. A general outline of the IND is given in Table 3.

Throughout drug development, toxicology/safety studies are conducted in two species. As most drugs are intended for oral administration, the rat and dog are the primary species used for reasons of availability, costs, and long history of use. However, not all drugs are targeted for oral delivery and other routes (intravenous, dermal, etc.) and species (rabbit for reproductive toxicology, monkey) may be used (Table 4). In general, the current climate supports a "minimal" IND as presented in the recent FDA guidance—*Guidance for Industry, Investigators, and Reviewers: Exploratory IND Studies*. This guidance allows sponsors a great deal of flexibility in the IND content and in the toxicity studies required, provided limited human exposure and no therapeutic intent is the objective of the initial clinical trial. Thus,

**TABLE 3 Outline of an IND**

| |
|---|
| Covering Letter |
| Questions for the FDA |
| Section 1: Form 1571 |
| Section 2: Table of Contents |
| Section 3: Introductory Statement |
| Section 4: General Investigational Plan |
| • Clinical Development Plan |
| • Planned Exposure |
| • Safety/Estimation of Potential Risk |
| Section 5: Investigators Brochure |
| Section 6: Protocol(s) |
| Section 7: Chemistry, Manufacture and Controls (CMC) |
| • Drug Substance |
| • Drug Product |
| Section 8: Nonclinical Pharmacology and Toxicology |
| • Overview of the Nonclinical Testing Strategy |
| • Nonclinical Pharmacology |
| • Nonclinical Pharmacokinetics and Metabolism |
| • Toxicology |
| • Conclusions |
| • References |
| • End of Text Tables: Summary Tabulations |
| • Copies of Literature Cited |
| • Copies of Final Study Reports |

**TABLE 4 Species and Study Types in an Exploratory IND**

| Species | Study Type |
|---|---|
| Mouse | Acute toxicity, subchronic toxicity (2 weeks), genotoxicity (*in vivo* micronucleus) |
| Rat | Acute toxicity, subchronic toxicity (2 weeks), genotoxicity |
| Rat | Safety pharmacology—Irwin test, and CNS effects |
| Rabbit | Reproductive toxicology |
| Dog | Acute toxicity and subchronic toxicity, safety pharmacology—QT interval prolongation |
| Monkey | Safety pharmacology-QT interval prolongation |

a "go/no go" decision can be reached very early and is based on human, albeit limited, experience. It is important to point out that in case the data supports moving forward in development, the traditional toxicology/safety program must be conducted to continue development.

The species and studies that would be contained in such an IND are outlined below. It should also be pointed out that the route of administration in these studies would be the intended route in the clinic.

Not all of the above studies are conducted for an IND filing, and Table 4 presents a "menu" of what can be conducted. In general, acute and subchronic toxicity, the core genotoxicity, and safety pharmacology are required, with the most appropriate species chosen.

### 2.3 Key Elements of IND-Target Development

Throughout all drug development, safety is a factor, perhaps the key factor. The parameters that define safety are presented in the following text. It should be pointed out that as/when development moves forward with the aim of a NSA filing, safety in nonclinical studies continues to be evaluated in studies of longer duration (6 month study in rats, 9 month study in dogs/monkeys, the full panel of reproductive toxicity, carcinogenicity studies in mice and rats). However, the use of toxicokinetics and BA/BE issues remain to support the extent of exposure determinations and data interpretation. Examples of studies/data on file as well as from published reports are presented in the subsequent discussion. In this context it should be pointed out that although numerous studies germane to the present chapter are conducted, the majority are not published—thus the inclusion of data "gleaned" from personal files.

***Toxicology/Safety Profile*** As shown in Table 3, Section 8 of the IND deals with toxicology, and considering that at this time in development, other sections may be rather sparse. As stated earlier, the results of the toxicity studies, in terms of NOAEL and safety margins, are key for the conduct of the first human studies. The other sections provide data that characterize the physicochemical properties of the drug, its stability, and the clinical dosage form for these studies (CMC section); the proposed clinical plan; and the Investigator's Brochure. The toxicological/safety profile is presented in detail and is comprised of nonclinical pharmacology (the rationale

for development), nonclinical pharmacokinetics and metabolism, and toxicology. For all of these components, the text contains the elements of study design, results, discussion and conclusions, and summary tabulations of the data obtained. Copies of QA final study reports are also included in the submission.

***Extent of Absorption*** In practice, the actual (numerical) extent of absorption is not determined in the context of toxicology studies. It is important, however, to demonstrate/ascertain absorption, and this is achieved by the toxicokinetic evaluations; this data is evaluated in a relative, intradose manner and as a function of both dose and time.

Usually the level of absorption (bioavailability) is obtained from the nonclinical PK and ADME studies. Considering that these studies are conducted in the same species and with the same dosage forms as the toxicology studies, the data are thus available to provide a reasonable assessment of the bioavailability of the drug in the toxicity studies, if such an assessment is needed.

An example of a PK study in dogs for bioavailability (relative to an IV injection) of a solution and a solid dosage form (a preliminary clinical capsule) is shown in Fig. 1. The plasma concentration–time profiles clearly show a good oral bioavailability of the drug from the solution, whereas it was virtually quantitative from the preliminary capsule considered for the first clinical trials. The use of this capsule for the toxicity safety studies was considered as it would provide the maximum drug release profile and thus clearly define the toxicological profile and the NOAEL.

***Assessment of Safety Margins Based on Systemic Exposure*** At the IND stage, the primary objective is to provide safety assessments, based on which the first human studies are conducted. These Phase I studies determine single-dose pharmacokinetics/tolerance in normal volunteers with the safe dose obtained in humans from the safety margins calculated from the results of the IND-enabling toxicity studies [9, 10]. A "simple" treatment of safety factor calculations can be derived from the

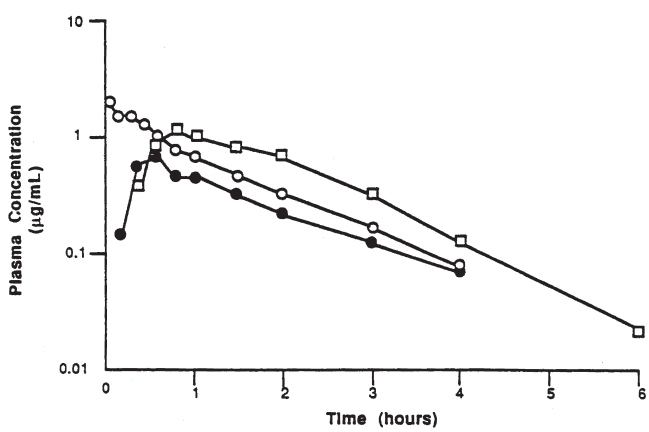


**FIGURE 1** Mean plasma concentrations of the neuroleptic in dogs following a 5 mg/kg dose intravenously (○), as an oral solution (●), and as a clinical capsule (□). (From I. Bekersky, personal data.)

comprehensive methods presented in the guidance [8]. In brief, one can calculate the human dose expected to result in the same effect seen in the animal studies (HED); the safety margin can be estimated on a mg/kg comparison of doses, using the equation $HED = \text{Animal } NOAEL \times (W_{\text{animal}}/W_{\text{human}})^{0.33}$. This calculation takes into account both the weight and the relative body surface area.

A similar equation uses a "conversion factor" ($S$) that addresses the issue of body surface area: ($S$ factor) × ($NOAEL$ dose in mg/kg)/$S$ factor$_{\text{human}}$ = Safe dose in humans in mg/kg (the HED). Knowing the extent of systemic exposure (AUC) from toxicokinetic assessments provides an additional level of confidence to the determination of the NOAEL, and thus to the calculation of the safety margins, and in turn, to setting a safe dose for the initial human trials.

***Relationship of Absorption to Dosage Form(s) and Route of Administration*** The nature of the nonclinical studies that are conducted in drug development depends on the targeted disease, the route of administration in therapeutic use, and the anticipated treatment duration in humans. Thus administration by the oral, intravenous, dermal, inhalation, subcutaneous, and ocular routes in studies ranges from a single dose to a lifetime exposure. As pointed out earlier, although the rat and dog are the most frequently used species, rabbits, mice, mini- and micropigs, and monkeys are also used. The aim is to use the species that are considered to be the most relevant to humans in terms of the intended clinical use.

To discuss and/or present all possible routes of administration is not our intention here. Needless to say, the extent of absorption must be determined relative to a "standard." In this context, examples of some routes of administration are given as illustrative of this principle.

Regardless of species and route of administration, the need is to administer high, but justifiable, doses in order to establish safety margins for human exposure and to determine the target organs of toxicity. Although bioavailability for IV administration is not an issue, it is for other routes of administration. Thus the incorporation of TK into toxicity studies along with assessments of bioavailability ensures not only that the drug is present systemically, but that the extent of systemic exposure is "maximal" in order to obtain the best possible assessment of the toxicological profile.

*Oral Administration* The primary route of administration in many, perhaps most, nonclinical toxicity studies is oral; this is due to the fact that that is the "preferred" route of delivery in the marketplace. However, as stated earlier, other routes—intravenous, subcutaneous, inhalation, topical—are also used, both in the marketplace and thus in the preclinical toxicity studies. The absorption in all cases is determined (TK) and the extent estimated by comparison to PK studies; for example, AUC post-IV to AUC from other routes of administration.

The dosage form itself (i.e., "formulation") can have pronounced effects on the absorption and thus on the selection of the dose. Such effects are well described, especially for oral drugs. Some examples from our own studies and the literature are discussed later in several sections of this chapter.

The rat is the species most commonly used with oral administration the route of choice and is the primary rodent utilized in toxicology studies. Oral administration by gavage as a solution or suspension is the most frequently employed dosage form

and is used in studies of up to 6–9 months duration. The primary consideration is the choice of vehicle. Solutions in water are chosen if the solubility of the drug is sufficient to deliver the highest dose proposed. For drugs of low or limited solubility, a dispersing agent for use in an aqueous milieu is selected. A vehicle of 0.5% aqueous Tween 80 (i.e., polysorbate 80) has a long history of use for a wide range of drugs. Cellulose and its derivatives (e.g., methyl-, carboxy-), starch, and different natural oils (e.g., corn, sesame) are often used as a vehicle for oral gavage formulations. For studies of greater duration than 9 months (i.e., lifetime carcinogenicity studies), oral drugs are usually administered as an admixture in the diet of the rats.

The absorption in all studies is determined from the TK data and evaluated as to whether the drug release profile from the dosage form is sufficient to define the toxicological profile. An example of mean plasma concentration–time data comparing gavage administration to the administration of a similar dose as a dietary admix to rats is shown in Fig. 2. In this example, a composite profile was obtained using three rats per time point using alternate animals. The challenge in rodent studies is the small volume of blood that can be taken; thus an analytical method with a low LOQ is needed. Because of the small amount of blood available, rats are usually alternated through the blood collection time points and in mice blood collection is usually terminal.

Use of the dog, the most commonly utilized nonrodent species, allows a wide range of evaluations of oral dosage forms without the "liability" of limited blood sample volume. An example of comparative absorption profile in dogs is shown in Fig. 3, where an equivalent dose was administered as a solution and a formulated capsule. In Fig. 4 the absorption profile of a highly lipophilic drug formulated as two different salts of Labrafil (a fatty acid complex used to solubilize drugs) is shown. In both cases, the different absorption profile as a function of the dosage form is clearly demonstrated.

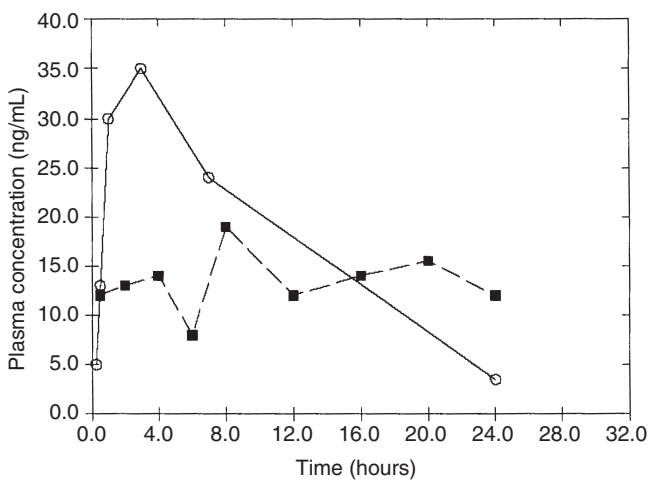


**FIGURE 2** Plasma concentrations of drug in rats after oral administration as a solution/suspension by gavage (○) and as a dietary admix (■). (From I. Bekersky, personal data.)

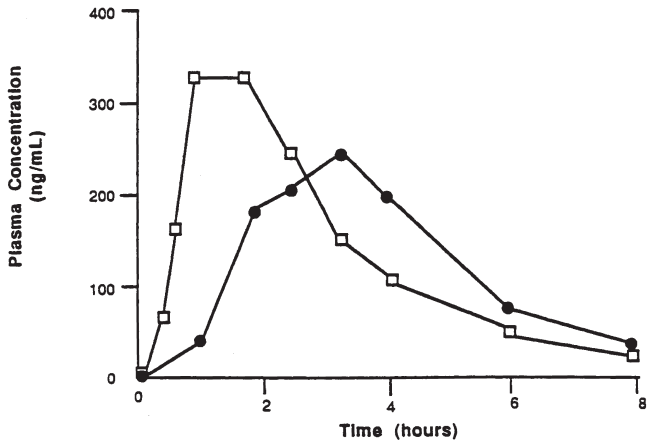


**FIGURE 3** Mean plasma concentrations of drug in dogs after 20 mg/kg oral administrations as a solution in PEG (□) and a formulated capsule (●). (From I. Bekersky, personal data.)

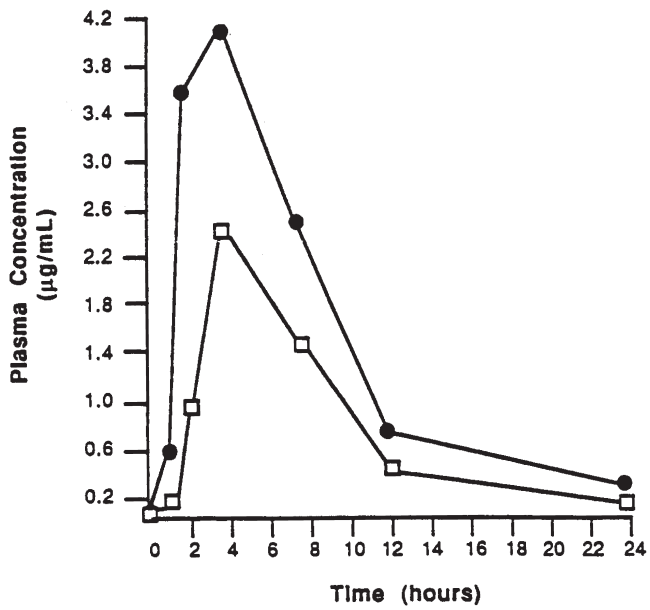


**FIGURE 4** Mean plasma concentrations of drug in dogs after a 10 mg/kg oral administration of capsules of NMG-Larafil (□) and TEA-Labrafil (●). (From I. Bekersky, personal data.)

Low oral bioavailability of some drugs can be improved by including certain enhancers in drug formulation or special coating of the tablet/capsules. Enhancement of drug bioavailability may be studied in several animal species and can be predictable for humans. Coadministration of compound A and indinavir resulted in a 17-fold increase in oral AUC of compound A in rats due to the inhibition of the metabolism of compound A by indinavir, whereas compound A did not affect

indinavir metabolism as indicated by the unchanged indinavir AUC [11]. Similarly, the systemic exposure of compound A in dogs and monkeys was increased substantially following oral coadministration with indinavir by sevenfold and >50-fold, respectively. Enhancement in compound A systemic exposure by indinavir in humans is predicted based on the *in vivo* animal and *in vitro* human studies.

Labrasol was found to improve the intestinal absorption of gentamicin (GM) in rats. Bioavailability of GM in several different formulations containing labrasol was evaluated in beagle dogs [12].

In Wistar rats, 1.0% (w/v) *N*-trimethyl chitosan chloride (quaternized chitosan derivative) solution significantly increased the absorption of the peptide analogue, resulting in a fivefold increase of octreotide bioavailability compared with the controls (octreotide alone). Coadministration of 1.0% (w/v) chitosan hydrochloride did not enhance octreotide bioavailability [13].

In dogs dosed orally, cefoxitin bioavailability increased from 2.4% to 29% with an enteric-coated tablet containing 600 mg PCC, but uncoated tablets containing palmitoyl-DL-carnitine chloride resulted in no improvement of absorption [14]. Glycosylated bile acid analogue was also used in dogs to increase the absorption of gentamicin [15]. The increase in bioavailability depended on the site of administration. Bioavailabilities in controlled and enhanced states, respectively, were 6% and 54% when administered to the ileum, 4% and 23% for jejunal dosing, and 2% and 10% when administered orally.

Dodecylmaltoside (surfactant) increased the colonic absorption of azetirelin 8.7-fold in rats, and increased oral bioavailability in dogs from 15% to 44% when combined with citric acid in an enteric-coated capsule formulation [16]. The extent of absorption enhancement reported ranged from less than twofold to fourfold to fivefold, with maximum bioavailabilities in rats and dogs of approximately 7% and 19%, respectively [17, 18]. Intraduodenal bioavailability of calcein in rats was improved from 2% to 25–37% with microemulsion formulations containing a mixture of caprylic acid or capric acid and their sodium salts, as well as medium chain glycerides [19].

Medium chain glycerides have been delivered orally in several types of formulations. A self-emulsifying w/o microemulsion formulation containing 22% MCG was administered intraduodenally to rats (3.3 mL/kg); the bioavailability of calcein improved from 2.4% to 45%, and the bioavailability of a hydrophilic peptide improved from 0.5% to 27% [20].

To aid delivery to the lower intestine, capsules containing cefmetazole in an MCG solution were enteric coated [21]. Cefmetazole bioavailability in dogs was 65% with the enteric-coated MCG capsules, 21% with uncoated MCG capsules, and 6% with enteric-coated capsules without MCG. Enteric coating of capsules also improved bioavailability of ceftriaxone in monkeys when administered in MCG.

A vehicle containing 1.5% chitosan at pH 6.7 increased the bioavailability of intraduodenally administered buserelin in rats from 0.1% to 5.1% [22]. A 10-fold increase in the oral bioavailability of paclitaxel was observed in mice when coadministered with the P-gp inhibitor SDZ PSC 833 [23]. A similar magnitude of enhanced bioavailability was found in human subjects administered paclitaxel with 15 mg/kg cyclosporin, another P-gp inhibitor [24]. Oral absorption of digoxin in mice was increased by coadministered SDZ PSC 833, and brain concentrations and brain/plasma ratios were also increased by the P-gp inhibitor, raising the possibility that

systemic absorption of P-gp inhibitors could alter the pharmacologic and toxic effects of coadministered drugs [25].

*Intravenous (IV) Administration* As stated earlier, bioavailability is not an issue for IV administration as the total dose is delivered. However, there are other considerations that impact on the dose size and thus on the safety margins obtained. The IV dosage form must be a solution and thus the physicochemical properties of the drug are taken into consideration. Once the formulation is established, its compatibility with blood must be determined to establish venous and perivenous tolerance. In addition, the duration of injection (or infusion) is evaluated to establish the safety margin; however, it should be stated that this factor may not be applicable to human use because of anatomical and physiological difference.

It goes without saying that in drugs administered intravenously in the clinic, the issue of bioavailability is not a consideration. However, the species used can be an important determinant based on other factors that determine drug disposition. Examples are given in Tables 5 and 6, showing that the plasma protein binding parameter was the determinant in the selection of the appropriate species for the toxicological evaluations of ceftriaxone, a parenteral cephalosporin antibiotic. At therapeutic plasma concentrations, the extent of protein binding in human plasma is high (90–95%) and is concentration dependent (Table 5). The plasma protein binding in animals that were used in nonclinical development is dependent on the species [26]. In baboon plasma, protein binding (Table 5) is concentration dependent and the extent of binding is quite similar to that found in human plasma [26]. In dog plasma, however, protein binding is low with no meaningful relationship to drug concentration. Thus the free concentration of cefriaxone varies greatly in human, baboon, and dog over a similar plasma drug concentration range. As a consequence of the plasma protein binding, the pharmacokinetics of ceftriaxone in humans are nonlinear when based on total drug concentration, but linear if based on the free drug concentration (Table 6) [26, 27]. This was the case in baboons, whereas in dogs the pharmacokinetics was the same whether based on total or free drug concentration. These differences played a role in the selection of species for the toxicology studies in early development. In this context, the dog was rejected

**TABLE 5 Protein Binding Profile of Ceftriaxone in Human and Animal Plasma**

| | Human | | Baboon | | Dog | |
|---|---|---|---|---|---|---|
| Drug Concentration (μg/mL) | Bound (%) | Free Concentration (μg/mL) | Bound (%) | Free Concentration (μg/mL) | Bound (%) | Free Concentration (μg/mL) |
| 100 | 92.7 | 7.3 | 95.5 | 4.5 | 19.6 | 80.4 |
| 200 | 88.5 | 23.0 | — | — | 13.9 | 172.2 |
| 240 | — | — | 88.5 | 27.6 | — | — |
| 400 | 78.7 | 35.2 | — | — | 7.1 | 371.6 |
| 560 | — | — | 65.9 | 190.9 | — | — |
| 800 | 64.2 | 64.2 | — | — | 6.9 | 744.8 |
| 1000 | — | — | 51.7 | 483.0 | — | — |

*Source*: Adapted from Ref. 27.

**TABLE 6 Pharmacokinetic Parameters for Ceftriaxone in Humans, Dogs, and Baboon Based on Total (T) And Free (F) Drug Concentrations**

| Species | Dose (mg/kg) | *AUC* (μg·h/mL) | *V* (L) | $CL_p$ (mL/min) | $t_{frac12}$ (h) |
|---|---|---|---|---|---|
| Human | 150 | 268 (T)10 (F) | 7 | 10 | 8.6 |
| | | | 192 | 262 | 8.6 |
| | 1500 | 1978 (T)103 (F) | 9 | 13 | 7.8 |
| | | | 162 | 249 | 7.6 |
| Baboon | 150 | 1573 (T)288 (F) | 4.5 | 11 | 6.1 |
| | | | 11.9 | 66 | 2.6 |
| | 700 | 3913 (T)1632 (F) | 5.7 | 21 | 6.5 |
| | | | 13.1 | 51 | 1.7 |
| Dog | 20 | 79 (T)68 (F) | 3.8 | 40 | 1.2 |
| | | | 5.2 | 52 | 1.3 |

*Source*: Adapted from Refs. 26 and 27.

**TABLE 7 Absolute Bioavailability of Drug Applied Topically to Rats and Micropigs**

| Species | Route | Dose | Application Area | *AUC* (ng·h/mL) | Bioavailabity, *F* (%) |
|---|---|---|---|---|---|
| Rat | IV | 1 mg/kg | | 642 | 100 % |
| | Topical | 0.5% (500 mg/m$^2$) | 10 cm$^2$ | 68 | ≤5% |
| | Topical (abraded) | 0.5% (500 mg/m$^2$) | 10 cm$^2$ | 570 | 62 |
| Micropig | IV | 1 mg/kg | | 3900 | 100 |
| | Topical | 3% (3000 mg/m$^2$) | 400 cm$^2$ | 70 | ≤1 |

*Source*: Adapted from Ref. 28.

based on the different pharmacokinetic profile relative to humans and was replaced by the baboon.

*Dermal Administration* From a toxicological end point, there are two factors that are determined in the development of topical drugs: the local toxicity at the site of application and the potential for systemic toxicity. The skin is a semipermeable barrier and absorption into the systemic circulation can thus be anticipated. In this context, the extent of systemic exposure (i.e., bioavailability) is determined following dermal application. This is determined for both intact and excoriated skin, as in clinical use the skin can be broken (such as in atopic dermatitis) with a subsequent effect on systemic exposure.

An example from a recent development of a dermal product is given in Table 7 [28]. Absolute bioavailability (relative to the respective IV administration) was ≤5% (rat) and ≤1% (micropig) based on AUC comparisons. Abrading the skin in a rat study increased the bioavailability to approximately 62%. The comparison of the

animal data to human IV (1 mg/kg) and topical administration (0.3%, 5000 $cm^2$ application area) indicated that the absolute bioavailability in atopic dermatitis patients was less than 0.5%. Interestingly, after 8 days of treatment, bioavailability decreased slightly (<0.3%), presumably reflecting an improvement in the integrity of the skin in response to treatment.

The cutaneous bioavailability of topical 2% minoxidil solution was verified in live hairless mice—minoxidil and propylene glycol deposition on the skin surface, epidermis, and dermis from the single-dose *in vivo* study [29]. Percutaneous absorption of the drug appeared to be a very small fraction of the applied dose. The pharmacokinetics of methotrexate (MTX) in rabbit skin and plasma after IV bolus and iontophoretic delivery at different current densities was studied [30]. The systemic absorption of MTX increased proportionally with increased current.

*Inhalation and Ocular Administration* These routes, although less frequently used, present challenges and thus warrant discussion. In the rat, the differences of the respiratory tract and the fact that rats breathe through their noses result in an inadequate model for inhalation exposure to humans. A common approach to rat inhalation studies is by exposing the nose of the animal through an opening into a cylinder filled with air containing a known concentration of the drug. Several difficulties are readily apparent: the actual dose administered and the monitoring of the systemic received. In larger animals (dogs, monkeys) inhalation is administered via a nose mask, resulting in a more discreet administration. As in all studies, a blood sampling protocol for TK analysis is included in the study for the assessment of systemic exposure.

Ocular administration is perhaps the most emotive route of administration; toxicity studies are conducted in rats, rabbits, and monkeys, with the monkey frequently used for its similarity in ocular volume to the human eye. As with topical administration, local toxicity is evaluated as well as systemic toxicity if a partitioning from the eye into the systemic circulation is demonstrated.

Preclinical studies are often performed in an attempt to improve bioavailability of nasal or ocular administered drugs. Enhancement of bioavailability of calcitonin formulated with different alkylglycosides following nasal and ocular administration was studied in rats [31]. Enhancement of bioavailability of insulin nanocomplexes in nasal formulation was also studied in rats [32]. Iontophoresis ocular drug delivery and peak concentrations determination were performed for numerous drugs in rabbit studies [33]. Hayden et al. [34] showed that focal delivery of carboplatin using subconjunctival injection or iontophoretic resulted in significantly higher peak concentrations of carboplatin in the choroid, retina, optic nerve, and vitreous body than those obtained after intravenous delivery.

*Comparison of the Different Routes of Exposure* Equivalence of different routes of exposure is not apparent and thus comparison of bioavailability or bioequivalence of the drug administered by the different routes of administration is often studied in various animals. This comparison is illustrated in Table 8 and Fig. 2. Other examples of such studies are presented next.

The pharmacokinetic parameters of midazolam tablets were compared after administration by the sublingual route and intravenous routes in six rabbits to

**TABLE 8 Biopharmaceutic Parameters of a Drug in the Dog from a Dosage Form Evaluation Study**

| Dosage Form | $C_{max}$ (μg/mL) | $T_{max}$ (h) | *AUC* (μg·h/mL) |
|---|---|---|---|
| Intravenous (IV) | | | 10.3 |
| Aqueous solution | 1.5 | 1.4 | 12.5 |
| Capsule[a] | 1.8 | 1.5 | 13.7 |
| Tablet[b] | 1.2 | 1.4 | 10.8 |
| CMC suspension | 0.9 | 1.0 | 5.7 |

[a]Soft gelatin capsule with neat drug.
[b]Tablet formulation for use in first clinical trials.
*Source*: I. Bekersky, personal data.

determine the bioequivalence between these routes [35]. By the sublingual route, midazolam absorption is substantial and fast. The statistical analysis, on data obtained with HPLC, shows no significant difference between pharmacokinetic parameter values calculated after intravenous and sublingual administration (0.5 mg). The absolute bioavailability was close to 100%. 1-Hydroxy-midazolam seems to have a great importance in BZD activity. To estimate the bioequivalence between intravenous and sublingual midazolam administration, it is necessary to take into account the active metabolites.

Brown et al. [36] compare the bioequivalence of the ceftiofur sodium salt in cattle after a single intramuscular (IM) or subcutaneous (SC) dose of 2.2 mg ceftiofur equivalents/kg body weight. The preestablished criterion for equivalence of the $t > 0.2$ for IM and SC administration was satisfied. The equivalence of $AUC_{0\text{-}LOQ}$ and $t > 0.2$ for IM and SC administration of 2.2 mg ceftiofur equivalent (CE)/kg doses of ceftiofur sodium suggest similar therapeutic efficacy and systemic safety for the two routes of administration.

The pharmacokinetic properties of a long-acting formulation of chloramphenicol were determined in six yearling cattle after a single intravenous (IV) administration (40 mg/kg body weight) and after two sequential subcutaneous (SC) or intramuscular (IM) administrations [37]. The two extravascular routes were studied during a crossover trial for a bioequivalence test. Bioavailability was 19.1% after IM injection and 12.4% after SC administration. The extent of absorption from the two routes did not differ significantly. The rate of absorption was significantly lower after SC application than it was after IM injection. The time necessary for the plasma concentration to exceed 5 μg/mL was the same for the two routes. Thus, IM and SC routes were bioequivalent. In the other study [38], the pharmacokinetics of chloramphenicol were studied in sheep after three single intravenous (IV), intramuscular (IM), and subcutaneous (SC) administrations (30 mg/kg) during a crossover trial for a bioequivalence test. The two routes of absorption were not bioequivalent. Using the kinetic values, multidose regimens to maintain the therapeutic chloramphenicol blood level (5 μg/mL) were proposed: 60 mg/kg every 12 hours for 72 hours for the IM administration and 45 mg/kg for SC administration according to the same regimen. Chloramphenicol residues remained at the injection site, and 400 hours

would be necessary to obtain the level of 10 μg/kg. Determination of the creatinine phosphokinase serum values showed that the subcutaneous route induced less damage to muscle than the intramuscular route.

The disposition of spiramycin (a macrolide antibiotic) in plasma and milk after intravenous, intramuscular, and subcutaneous administration was investigated in healthy cows given a single injection of spiramycin at a dose of 30,000 IU/kg by each route [39]. The dose fraction adsorbed after intramuscular or subcutaneous administration was almost 100% and was bioequivalent for the extravascular routes, but the rates of absorption, the maximal concentrations, and the time to obtain them differed significantly between the two routes. Plasma concentrations of doramectin in 40 cattle dosed by subcutaneous (SC) or intramuscular (IM) injection (200 micrograms/kg) were compared to assess the bioequivalence of the two routes of administration [40]. The bioequivalence of the SC and IM formulation has been established.

Pietropaolo et al. [41] performed direct comparisons between injection of myelin basic protein peptide and administration by several nonparenteral routes to determine whether route impacted benefit in the treatment of murine allergic encephalomyelitis, a model for multiple sclerosis. The range of effective peptide doses spanned over 1000-fold, and route of delivery played a major role in determining optimal dose. The oral route of administration was the least effective, requiring at least 50- to 100-fold more antigen than subcutaneous injection, which in turn required at least 10-fold more antigen than delivery of peptide to the lung using an intratracheal instillation. Intratracheal delivery was also considerably more effective than inhalation of peptide and, unlike inhalation, resulted in obvious penetration of delivered material deep into the lung. The increase in therapeutic efficacy did not appear to result from slower systemic delivery of antigen.

Apomorphine (0.5 mg/kg) was administered subcutaneously and percutaneously to rabbit in order to compare the pharmacokinetic data obtained with these two different routes [42]. For the percutaneous administration, an apomorphine gel was prepared by dissolution of apomorphine in a hydroxypropylmethylcellulose gel of medium viscosity. The peak plasma concentration and the area under the curve were significantly greater with the subcutaneous route. The bioequivalence of the percutaneous route was 35% of the subcutaneous administration. One study [43] reports a paradox: the AUC of apomorphine after intravenous injection in rabbits is anomalously lower (statistically significantly) than after subcutaneous injection.

Bergeron et al. [44] examined the effect of route of administration by giving equimolar amounts of two iron chelators—NaHBED and deferoxamine (DFO)—to *Cebus apella* monkeys as either a subcutaneous (SC) bolus or a 20 min intravenous (IV) infusion. By both routes, NaHBED was consistently about twice as efficient as DFO in producing iron excretion. For both chelators at a dose of 150 μmol/kg, SC was more efficient than IV administration.

Acrylamide bioavailability after aqueous gavage was 60–98% and from the diet was 32–44%; however, first-pass metabolism or other kinetic change resulted in much higher (two- to sevenfold) internal exposures to glycidamide (its epoxide metabolite, GA) when compared to the intravenous route. A similar effect on metabolism for GA following oral administration was previously observed under an identical exposure paradigm in mice [45].

### 2.4 Use of Animal Models for the Selection of Dosage Forms for Nonclinical Toxicology Studies

Some of the issues in selecting a dosage form for animal studies have been touched upon previously. Data that is needed prior to the preparation/choosing of a dosage form are solubility, stability, compatibility with excipients, and particle size, keeping in mind that a "maximum" drug release profile is a target.

For rat studies, use of solutions or suspensions administered via gavage has been described previously. Rats are rarely used for comparison of different dosage forms. Of more illustrative value are studies with large animals (dogs, monkeys), as these species allow more flexibility in the kinds of dosage forms administered. An example of a dosage form evaluation study is given in Table 8 where three possible dosage forms for toxicology studies (solution, unformulated capsule, and suspension) as well as a gelenic tablet prepared for the initial clinical trials were tested. The data indicate that relative to an IV administration, the solution, capsule, and galenic tablet were acceptable for use in toxicity studies with a drug release profile equal to the IV Bioavailability of the CMC suspension was approximately half of all the other dosage forms tested.

A biopharmaceutic evaluation that involved the drug particle size and several forms of the drug is given in Table 9. The drug in question had excellent oral activity although the form available at the time (the free base as an oil) resulted in severe GI toxicity, which was in all probability due to the prolonged presence of unabsorbed drug substance. Capsules containing salts of the drug were then tested, indicating that the acetate salt resulted in the greatest extent of absorption relative to the other two forms. The subsequent availability of the free base in crystalline form allowed the drug to be micronized, and a second study was conducted to evaluate the absorption profile. This study indicated that reducing the particle size by micronization resulted in higher $C_{max}$, shorter $T_{max}$, and $AUC$ values that were considerably higher than those in the first study. Interestingly, the more rapid absorption alleviated the GI toxicity found in preliminary studies, in spite of the considerably higher extent of absorption.

The selection of a formulation for nonclinical studies has received a "modest" treatment as being illustrative of the overall study design in the early stages of drug development. In order to attain a maximum drug release profile, in some cases in

**TABLE 9 Biopharmaceutic Parameters (Mean ±SD) from Two Dosage Form Evaluation Studies**

| Dosage Form | $C_{max}$ (μg/mL) | $T_{max}$ (h) | $AUC$ (μg·h/mL) |
|---|---|---|---|
| Nonmicronized | | | |
| Free base (oil) | 1.6 ± 1.4 | 2.7 ± 1.2 | 5.9 ± 5.7 |
| Dihydrochloride | 3.6 ± 1.9 | 0.5 ± 0 | 11.9 ± 5.3 |
| Acetate salt | 4.3 ± 1.9 | 4.0 ± 2.0 | 18.9 ± 8.0 |
| Micronized | | | |
| Free base | 8.4 ± 3.5 | 1.8 ± 0.5 | 70.2 ± 6.6 |
| Dihydrochloride | 4.9 ± 2.4 | 2.0 ± 1.4 | 40.8 ± 21.9 |
| Acetate salt | 5.8 ± 2.3 | 2.1 ± 1.4 | 41.4 ± 21.6 |

spite of undesirable physicochemical properties, and thus clearly to define the line between pharmacology and toxicology, formulation selection has become part of the overall development of drugs. In many cases, formulation evaluation studies are not part of the toxicology/safety studies but are conducted in parallel throughout development, with the information obtained being integrated into all evaluations. Examples from the literature of studies dealing with factors that impact on bioavailability are given next.

Seven kinds of poorly water-soluble compounds as the model compounds were used to compare the plasma concentration profile of the drug following single oral administration of each compound to rats and beagle dogs as a solution, an oily solution, a suspension (or a powder), and oil/water (O/W) microemulsions [46]. Thus for six kinds of the model compounds, except disopyramide, the solubility was from 340 to 98,000 times that in water, and the AUC values in plasma concentration of the compound were equivalent to that of solution or O/W microemulsion administration, or were increased by 1.5–78 times that of suspension administration.

The bioavailability of a modified rafoxanide oral suspension was compared to the original innovator product and a generic formulation in a single-dose, randomized, parallel design study in sheep [47]. The area under the rafoxanide plasma concentration versus time curve (AUC), AUC extrapolated to infinity, and maximum plasma rafoxanide concentrations ($C_{max}$) were significantly smaller for both the modified and generic formulations relative to the original product. There were no significant differences between the modified and generic formulations. In terms of the calculated 90% confidence *t*-intervals of the mean % ratios, the modified and generic formulations were not bioequivalent to the original product, since they were substantially below the accepted range of 80–125%.

### 2.5 Species Differences in Bioavailability

There are many processes that determine bioequivalence of a given drug among species, including absorption from a site of administration; renal, biliary, and intestinal elimination; sequestration (in particular, binding to proteins or other macromolecules); distribution and redistribution; biotransformation; and receptor population density and uniqueness. Short [48] reviewed physiologic and pharmacologic parameters that affect drug distribution, elimination, and metabolism to make comparisons of bioequivalence of drugs between sheep and other ruminants. Interspecies differences are very important for selecting an appropriate species for preclinical studies. The following is a summary of recent literature describing interspecies differences in pharmacokinetics of a wide range of therapeutic agents.

Pharmacokinetics and pharmacodynamics of controlled release gastroretentive dosage forms (CR-GRDF), which improve the bioavailability of medications, were studied for several drugs in rats and dog models [49–52]. The bioavailability of melagatran (active form of ximelagatran) following oral administration of ximelagatran was 5–10% in rats, 10–50% in dogs, and about 20% in humans, with low between-subject variation [53]. The oral bioavailability of levovirin was 29.3% in rats, 51.3% in dogs, and 18.4% in monkeys [54]. Zebularine showed oral bioavailability of 6.7% in mice, 3.1% and 6.1% in rats, and was <1% in monkeys [55].

Extrapolation of pharmacokinetics from animals to humans may be performed based on *in vitro* metabolism studies in hepatocytes. *In vitro* degradation studies of vinpocetine with hepatocytes have shown that the activity of human hepatocytes is about one order of magnitude higher than the activity of dog hepatocytes, and two orders of magnitude higher than that of rat hepatocytes [56]. These differences can explain the differences in bioavailabilities of vinpocetine in the three species (52% in rats, 21.5% ± 19.3% in dogs, and 6.2% ± 1.9% in humans). In dogs and humans, the compound seems to be metabolized exclusively in the liver, whereas in rats extrahepatic metabolism seems also to be important. The *in vivo* clearance predicted from the activity of hepatocytes is in good agreement with the values measured *in vivo* in the case of humans and dogs.

The interspecies absorption, distribution, metabolism, and elimination (ADME) profile of *N*-geranyl-*N*′-(2-adamantyl)ethane-1,2-diamine (SQ109), a new diamine-based antitubercular drug, were characterized after intravenous and oral single administration in rodents and dogs [57]. Based on IV equivalent body surface area dose, the terminal half-life of SQ109 in dogs was longer than that in rodents, reflected by a larger volume of distribution and a higher clearance rate of SQ109 in dogs, compared to that in rodents. The oral bioavailability of SQ109 in dogs, rats, and mice were 2.4–5%, 12%, and 3.8%, respectively. The binding of [$^{14}$C]SQ109 (0.1–2.5 μg/mL) to plasma proteins varied from 6% to 23% depending on the species (human, mouse, rat, and dog). SQ109 was metabolized by rat, mouse, dog, and human liver microsomes, resulting in 22.8%, 48.4%, 50.8%, or 58.3%, respectively, of SQ109 remaining after a 10 min incubation at 37 °C.

Compound A (potent inhibitor of the hepatitis C virus (HCV) NS5B polymerase) exhibited marked species differences in pharmacokinetics [58]. Plasma clearance was 44 mL/min/kg in rats, 9 mL/min/kg in dogs, and 16 mL/min/kg in rhesus monkeys. Oral bioavailability was low in rats (10%) but significantly higher in dogs (52%) and monkeys (26%). Compound A was eliminated primarily by metabolism in rats, with biliary excretion accounting for 30% of its clearance. Metabolism was mainly mediated by cyclohexyl hydroxylation, with N-deethylation and acyl glucuronide formation constituting minor metabolic pathways. Qualitatively, the same metabolites were identified using *in vitro* systems from all species studied, including humans. The low oral bioavailability of compound A in rats was mostly due to poor intestinal absorption. A series of high affinity GABA(A) agonists showed good oral bioavailability in both rat and dog [59].

The pharmacokinetics of DA-6034 in rats and dogs and the first-pass effect in rats were examined [60]. The low bioavailability (approximately 0.136%) of DA-6034 at a dose of 50 mg/kg in rats could be due to considerable intestinal first-pass effect (approximately 69% of oral dose) and unabsorbed fraction from the gastrointestinal tract (approximately 30.5%). The clearance of adaphostin (potential anticancer agent) from plasma, on a square meter ($m^2$) basis, was equivalent for mice and rats but more rapid in dogs [61]. Oxytocin antagonist, 2′,4′-difluorophenyldiketopiperazine derivative 37, has good bioavailability (46%) in the rat and moderate bioavailability (13–31%) in the dog [62], while 5-chloroindole had good oral bioavailability in both rat and dog [63].

BMS-378806 (prototype of novel HIV attachment inhibitors) exhibited species-dependent oral bioavailability, which was 19–24% in rats and monkeys and 77% in dogs [64]. In rats and monkeys, absorption was prolonged, with an apparent terminal

half-life of 2.1 and 6.5 hours, respectively. In rats, linear pharmacokinetics was observed between IV doses of 1 and 5 mg/kg and between PO doses of 5 and 25 mg/kg. The total body clearance was intermediate in rats and low in dogs and monkeys. The steady-state volume of distribution was moderate (0.4–0.6 L/kg), contributing to a short half-life (0.3–1.2 hours) after IV dosing.

The pharmacokinetics and disposition of *N*-(2,6-dichlorobenzoyl)-4-(2,6-dimethoxyphenyl)-L-phenylalanine (TR-14035), a novel a4ss1/a4ss7 antagonist, were investigated in the rat and dog [65]. Results indicated extensive clearance of TR-14035 and low oral bioavailability, 17% and 13% in the rat and dog, respectively, at an oral dose of 10 mg/kg. A species-dependent difference in metabolism was observed. The principal metabolite, *O*-desmethyl TR-14035, observed in rat, dog, and probably human, was further conjugated with sulfate in the rat, but never in dog and human, based on *in vitro* metabolism and *in vivo* metabolite profile studies. Urinary excretion was a minor elimination route, but an interesting species difference was recognized. TR-14035 was reabsorbed from the rat renal proximal tubules and, by contrast, secreted into the tubules in the dog, probably via active transport systems.

Exposure of MGS0028, a potent group II metabotropic glutamate receptor agonist, increased proportionally (1–10 mg/kg PO) in rats, with bioavailability >60% at all doses. However, bioavailability was only approximately 20% in monkeys, and active metabolite was found in relatively high abundance in plasma. In dogs, oral bioavailability was >60%, and the metabolite was not detected. James et al. [66] found that the rat was a better *in vivo* model than the dog for the prediction of human systemic exposure in HIV-1 nonnucleoside reverse transcriptase inhibitors of the diaryltriazine and diarylpyrimidine classes of compounds [67]. The absolute bioavailability of YM466, a new factor Xa inhibitor, was 2.7–4.5% in rats, almost constant regardless of the dose levels investigated, while it was 6.9–24.6% in dogs, indicating nonlinear pharmacokinetics.

The bioavailability of SNI-2011, cevimeline, a novel muscarinic acetylcholine receptor agonist, was approximately 50% and 30% in rats and dogs, respectively. Major metabolites in plasma were both S- and N-oxidized metabolites in rats and only N-oxidized metabolite in dogs, indicating that a large species difference was observed in the metabolism of SNI-2011 [68].

Absolute oral bioavailability of a gavage dose of galantamine was high in rat (77%) and dog (78%). In mice and rats, the bioavailability of galantamine administered via the food was lower than of galantamine administered by gavage [69]. Elimination half-life of galantamine was relatively large in rat and dog and smaller in mouse and rabbit. Galantamine plasma levels after single and repeated administration of 10 mg/kg/day in all species investigated except female rat and rabbit were much higher than mean therapeutic plasma levels of galantamine obtained in humans. The pharmacokinetic profile of galantamine after repeated oral administration in rats was most similar to the profile obtained after repeated administration of 12 mg bid in humans.

Not only species but even a strain within the same species may predispose to pharmacokinetics differences especially when different strains exhibit different metabolism. Systemic exposure to oxycodone and noroxycodone was consistently higher for cytochrome P450 (CYP)2D1/2D2-deficient dark Agouti (DA) rats than for CYP2D1/2D2-replete Sprague–Dawley (SD) rats showing that strain differences predominated over diabetes status [70].

## 3 NDA-TARGETED NONCLINICAL DEVELOPMENT

When the IND is "opened" and the nonclinical and Phase I data are indicative of a "go" decision, the NDA-targeted studies are conducted. Whereas the IND contained acute and subchronic (2 and/or 4 week) toxicology studies, depending on the indication sought, an acceptable NDA will contain chronic toxicity studies in support: 6 month rat, 9 month dog (or monkey), 78 week carcinogenicity study in mice, and 24 month carcinogenicity study in rats. This would be the case for IV, PO, and inhalation drugs. For topical drugs, 52 week photocarcinogenicity and 2 year topical carcinogenicity studies are conducted.

### 3.1 Toxicokinetics in Chronic Toxicity Studies

Considering that the duration of exposure in chronic studies is severalfold that of subchronic studies, it is even more important to assess the extent of exposure. To this end, range-finding studies are conducted to ascertain that the dose range that will be used in the definitive study will provide both a NOAEL and a definition of toxicity. In general, the concepts that have been described previously hold for these studies. Toxicokinetics are an integral part of protocol designs to determine if there are changes with the duration of treatment. It is hoped that the bioavailability of the dosage form has been determined during the pre- and post-IND development, with the assurance of a consistent and demonstrable extent of systemic exposure.

It is worthwhile mentioning that the oral carcinogenicity studies are conducted with the drug administered as a dietary admix. This would require a "bridging study" comparing bioavailability of the dietary admix to a systemic administration to determine the bioavailability. Such studies entail a 5–8 sample profile over 24 hours, whereas the toxicokinetics in the definitive admix studies would consist of a single blood draw, usually at approximately 9 am.

An example of a gavage versus admix comparison using serial sampling is shown in Fig. 2; from such data, exposure and bioavailability based on AUC can readily be determined. Single sample monitoring over the course of a chronic study, however, is useful as accumulation or induction over the study time course can be ascertained and such data will be of help in the study evaluation.

### 3.2 Use of Animals to Evaluate the BA/BE of Clinical Dosage Forms

In the early stages of development, it is important to determine whether the dosage form targeted for clinical use has an acceptable level of bioavailability. To evaluate the dosage form in a clinical study is time consuming and costly. In this case, the clinical form was first evaluated in dogs and subsequently in humans. The results are shown in Table 10. Compared to IV administrations, the bioavailability of the oral capsule formulation in dog was low (27%), and based on data from ADME studies, the low bioavailability was considered to be due to extensive metabolism. Urinary excretion data indicated that only a small fraction of the dose (0.2–0.3%) was eliminated as the intact drug, with the rest excreted as a hydroxyl metabolite. It was considered that even though the bioavailability was low, the level was sufficient to continue with a clinical study. As shown in Table 10, the results in humans were similar: low $C_{max}$ and low bioavailability of the intact drug (approximately

**TABLE 10 Biopharmaceutic Comparison of a Drug in Dog and Human**

| Species | Dosage Form | Dose (mg/kg) | $C_{max}$ (μg/mL) | $AUC_{inf}$ (μg·h/mL) | $F^a$ (%) |
|---|---|---|---|---|---|
| Dog | Capsule | 7.5 | 0.62 | 66 | 27 |
| Human | Tablet | 10 | 0.08 | 142 | 37 |

[a]Relative to IV dose and corrected for dose.

*Source*: Adapted from Ref. 79.

**TABLE 11 Comparison of Drug Release Profile of Drug from Similar Dosage Forms in Dog and Human**

| Species | Dosage Form | Dose (mg/kg) | $T_{max}$ (h) | $C_{max}$ (ng/mL) | $AUC$ (ng·h/mL) |
|---|---|---|---|---|---|
| Dog | PEG solution | 20 | 1 | 320 | 920 |
| | Captex capsule | 20 | 3 | 230 | 928 |
| Human | PEG capsule | 0.2 | 3 | 446 | 1524 |
| | Neobee capsule | 0.2 | 4 | 336 | 1004 |

*Source*: I. Bekersky, personal data.

37%). In this case, the study in dogs indicated a level of oral bioavailability that was acceptable for a clinical evaluation; a clinical study was then conducted with similar results to those found in dogs.

In Table 11, an evaluation of several dosage forms in dogs and subsequently in humans is shown. In dogs, there was little difference between the oral formulations tested, although the PEG solution resulted in a more rapid absorption based on $T_{max}$. In humans, however, the capsules containing PEG formulation appeared to be more bioavailable, an *AUC* of 1524 ng·h/mL versus an *AUC* of 1004 ng·h/mL for the Neobee formulation. Capsules containing drug as solution/suspension in PEG were subsequently developed for use in clinical trials.

In another study, preliminary experiments with an antibiotic indicated a low relative bioavailability. Physicochemical considerations suggested that ester analogues represented an approach that could potentially enhance the oral bioavailability of the parent compound. Ester analogues of the drug were synthesized and tested as solutions in dogs. The results are summarized in Table 12. Relative to the solution, both esters enhanced the oral bioavailability: 26% of the Bac ester, and 22% of the Piv ester; the bioavailability of the parent drug was 10%. On the basis of these results, both esters were administered to humans with results comparable to those "predicted" by the dog study.

Use of the dog to provide insights pertaining to drug absorption is not restricted to situations where information prospective to the first clinical trials is desired. Even when the clinical trials are well advanced, the dog can be utilized to ascertain a situation that may require reformulation of the clinical dosage form. An example of such a case is given in Table 13. The drug was unstable in the presence of oxygen and light and was formulated as a micronized suspension in oil within a soft elastic gelatin (SEG) capsule. On aging, these capsules failed to meet the dissolution speci-

**TABLE 12 Oral Bioavailability of Drug Analogue Solutions in Dog and Human**

| Species | Drug Form | Dose (mg/kg) | $C_{max}$ (μg/mL) | *AUC* (μg·h/mL) | Bioavailability *F* (%) |
|---|---|---|---|---|---|
| Dog | Parent | 20 | 2.6 | 6.9 | 10 |
| | Bac ester | 20 | 9.1 | 18 | 26 |
| | Piv ester | 20 | 5.5 | 15 | 22 |
| Human | Bac ester | 5.3 | 4.4 | 7 | 29 |
| | Piv ester | 5.3 | 3.2 | 6 | 25 |

*Source*: I. Bekersky, personal data.

**TABLE 13 Biopharmaceutic Evaluation of Potential Clinical Formulations in Dog and Human**

| Species | Administration | Approximate Dose (mg/kg) | $C_{max}$ (μg/mL) | $AUC_{inf}$ (μg·h/mL) | $F^a$ (%) |
|---|---|---|---|---|---|
| Dog | Oral solution | 4 | 3.33 | 12 | |
| | "New" SEG capsule | 4 | 1.05 | 4.5 | 39 |
| | "Aged" SEG capsule[b] | 4 | 1.23 | 5.5 | 47 |
| Human | "New" SEG capsule | 1 | 0.26 | 3.8 | 60 |
| | "Aged" SEG capsule | 1 | 0.29 | 3.7 | 59 |

[a]Bioavailability relative to respective oral solution and corrected for dose.
[b]"Aged" capsules were stored for 4 years and failed dissolution specifications.

*Source*: Adapted from Ref. 79.

fications for the product. To determine whether reformulation was required, these "aged" capsules were tested in dogs relative to the identical "new" formulation. In the dog study, no changes in the bioavailability were found due to aging. This was confirmed in a study in humans that gave similar results to those found in the dog study.

The dog can also be used in exploratory research. The drug in question was under development for parenteral delivery only. Oral administration to dogs resulted in very low $C_{max}$ (3–4 μg/mL) with a bioavailability of approximately 4% relative to IV administration. However, since some oral absorption did occur, an exploratory program with dogs was initiated to study the feasibility of developing an oral dosage form for humans. The results are given in Table 14. The intraduodenal (ID) and rectal administration to the dog of two exploratory formulations containing an absorption enhancer indicated a significant increase in absorption as indicated by the $C_{max}$ and *AUC* values. Based on these data, the concept was tested in humans. Although the bioavailability via the different routes in enteral administration was

**TABLE 14 Biopharmaceutic Parameters of a Drug in Dog and Human Following a Single Administration by Different Routes**

| Species | Route | Dose | $C_{max}$ (μg/mL) | $T_{max}$ (h) | $AUC_{inf}$ (μg·h/mL) | $F$ (%) |
|---|---|---|---|---|---|---|
| Dog | IV | 20 mg/kg | — | — | 102 | — |
| | ID[a] | 50 mg/kg | 47 | 0.5 | 82 | 32 |
| | Rectal | 50 mg/kg | 28 | 0.2 | 24 | 10 |
| Human | IV | 500 mg | — | — | 427 | — |
| | ID | 500 mg | 6 | 0.16 | 24 | 6 |
| | Rectal | 500 mg | 23 | 0.75 | 124 | 29 |

[a]ID, intraduodenal.

*Source*: I. Bekersky, personal data.

reversed in humans when compared to dogs, the potential enteral delivery in humans was ascertained.

### 3.3 Role and Impact of Nonclinical Studies on Clinical Development: Animal to Humans

During Phase I–IIa, factors such as a food effect, gender effect, and pre- and systemic metabolism are determined prior to Phase IIb in order to set/justify dose selection for the pivotal Phase III studies. Nonclinical studies have provided "windows" of information prospectively to the conduct of label–targeted human clinical pharmacology studies; examples are given next.

The bioavailability in beagle dogs and the dissolution rates of cyclandelate from five capsule preparations commercially available in Japan were measured [71]. One of the capsules that showed an extremely low bioavailability in humans also showed the lowest bioavailability in beagle dogs, although the difference in bioavailability with the highest preparation was smaller than in humans. A significant correlation was obtained between the results of the studies in humans and beagles. However, the power of the test using beagles was extremely low in comparison with that in the human study. Food enhanced the bioavailability of cyclandelate from the capsules, having the highest and lowest bioavailability in the fasted state in beagles as observed previously in the human study. The bioinequivalence of the cyclandelate capsules detected in the fasted state disappeared in the fed state in the beagle dog study, while the bioinequivalence still remained in the nonfasted state in human subjects. Thus bioequivalence testing in the fed state led to different results in both species.

Coadministration of food increases the rate and extent of cilostazol absorption [72]. After investigating the effect of concomitant food intake on the bioavailability of two nifedipine-containing modified-release dosage forms, authors concluded that a change from taking Slofedipine XL in the fed to the fasted state might result in increased systemic concentrations of nifedipine [73]. In the dog, food decreased the mean area under the plasma concentration–time curve value of compound LY354740, an analogue of glutamic acid, by approximately 34%, hence decreasing the oral bioavailability of the compound [74].

Concomitant intake of levodopa preparations with banana juice, but not with a commercial banana beverage, caused a drug–food interaction reducing levodopa bioavailability in rats [75]. Mean (±SD) peak concentrations of tepoxalin were significantly higher after feeding of low fat (1.08 ± 0.37 μg/mL) and high fat (1.19 ± 0.29 μg/mL) diets than in fasted dogs (0.53 ± 0.20 μg /mL), suggesting that feeding improves oral bioavailability [76].

Bioavailability and pharmacokinetics may be gender dependent. Jann et al. [77] showed that gender may be a significant factor in ondansetrons disposition. Gender differences in the pharmacokinetics of DA-6034 were found after intravenous but not an oral administration at a dose of 50 mg/kg to male and female Sprague–Dawley rats [78]. In rats, plasma levels of galantamine were lower in females than in males, whereas in mice, females showed higher levels than males [69]. No gender differences were observed in dogs. Sex difference was also observed in the pharmacokinetics of cevimeline in rats but not in dogs [68].

## 4 CONCLUSION

Toxicology/safety studies in animals are an integral part of any drug development program, the results of which have a direct impact on both the IND and the NDA. The calculations of safety margins for human doses are based on the measurements of systemic exposure (AUC) and the NOAEL in the animal studies. In this context, the inclusion of pharmacokinetics into the conduct/evaluation of animal toxicity studies (toxicokinetics) has become routine. Toxicokinetic evaluations ascertain that a level of systemic absorption was achieved, the intrastudy relationship of this absorption to the doses administered, and absorption changes with the duration of exposure. The inclusion of bioavailability concepts adds the element of quantity, as the level of absorption relative to a standard is obtained. As a maximum drug release profile is desired, in many cases several dosage forms are tested. Thus bioequivalence determinations of several dosage forms intended for toxicology studies provide assurance that the most appropriate dosage form is used. The bioequivalence is also determined to assess equivalence between the routes of administration, species differences, and drug derivatives preference.

Animal studies for bioavailability/bioequivalence are used both prospectively and retrospectively, and depending on the objective and study design, form an animal to human bridge in drug development.

## REFERENCES

1. *Code of Federal Regulations*, Title 21, 320.1.
2. *Code of Federal Regulations*, Title 21, Volume 5.
3. Chen ML, Shah V, Patnaik R, Adams W, Hussain A, Conner D, Mehta M, Malinowski H, Lazor J, Huang SM, Hare D, Lesko L, Sporn D, Williams R. Bioavailability and bioequivalence: an FDA regulatory overview. *Pharm Res* 2001;18(12):1645–1650.
4. Rajaram L, Roy SK, Skerjanec A. Bioavailability and bioequivalence trials: statistics & pharmacokinetic principles. *Stud Health Technol Inform* 2004;159–179.

5. Good Laboratory Practice Regulations, Title 21 of the *Code of Federal Regulations*, Part 58, effective 20 June, 1979.
6. *Bioavailability and Bioequivalence Studies for Oral Administered Drug Products—General Considerations*. Washington DC: US FDA; 2003.
7. *Bioequivalence Guidance*, Center for Veterinary Medicine. Washington DC: US FDA; 2002.
8. *Estimating the Maximum Safe Starting Dose in Initial Clinical Trials for Therapeutics in Adult Healthy Volunteers*. Washington DC: US FDA; 2005.
9. *Guidance for Industry*, *Investigators*, *and Reviewers*: *Exploratory IND Studies*. Washington DC: US FDA; 2006 .
10. Freireich EJ, Gehan EA, Rall DP, Schmidt LH, Skipper HE. Quantitative comparison of toxicity of anticancer agents in mouse, rat, hamster, dog, monkey, and man. *Cancer Chemother Rep* 1966:50(4):219–244.
11. Jin L, Chen IW, Chiba M, Lin JH. Interaction with indinavir to enhance systemic exposure of an investigational HIV protease inhibitor in rats, dogs and monkeys. *Xenobiotica* 2003;33(6):643–654.
12. Rama Prasad YV, Eaimtrakarn S, Ishida M, Kusawake Y, Tawa R, Yoshikawa Y, Shibata N, Takada K. Evaluation of oral formulations of gentamicin containing labrasol in beagle dogs. *Int J Pharm* 2003;268(1–2):13–21.
13. Thanou M, Verhoef JC, Marbach P, Junginger HE. Intestinal absorption of octreotide: *N*-trimethyl chitosan chloride (TMC) ameliorates the permeability and absorption properties of the somatostatin analogue *in vitro* and *in vivo*. *J Pharm Sci* 2000;89(7): 951–957.
14. Sutton SC, LeCluyse EL, Engle K, Pipkin JD, Fix JA. Enhanced bioavailability of cefoxitin using palmitoylcarnitine. II. Use of directly compressed tablet formulations in the rat and dog. *Pharm Res* 1993;10:1516–1520.
15. Axelrod HR, Kim JS, Longley CB, Lipka E, Amidon GL, Kakarla R, Hui YW, Weber SJ, Choe S, Sofia MJ. Intestinal transport of gentamicin with a novel, glycosteroid drug transport agent. *Pharm Res* 1998;15:1876–1881.
16. Sasaki I, Tozaki H, Matsumoto K, Ito Y, Fujita T, Murakami M, Muranishi S, Yamamoto A. Development of an oral formulation of azetirelin, a new thyrotropin-releasing hormone (TRH) analogue, using *N*-lauryl-*b*(-D-maltopyranoside as an absorption enhancer. *Biol Pharm Bull* 1999;22:611–615.
17. Aungst BJ, Saitoh H, Burcham DL, Huang S-M, Mousa SA, Hussain MA. Enhancement of the intestinal absorption of peptides and non-peptides. *J Control Release* 1996;41: 19–31.
18. Burcham DL, Aungst BJ, Hussain M, Gorko MA, Quon CY, Huang S-M. The effect of absorption enhancers on the oral absorption of the GPIIb/IIIa receptor antagonist, DMP 728, in rats and dogs. *Pharm Res* 1995;12:2065–2070.
19. Constantinides PP, Welzel G, Ellens H, Smith PL, Sturgis S, Yiv SH, Owen AB. Water-in-oil microemulsions containing medium-chain fatty acids/salts: formulation and intestinal absorption enhancement evaluation. *Pharm Res* 1996;13:210–215.
20. Constantinides PP, Scalart J-P, Lancaster C, Marcello J, Marks G, Ellens H, Smith PL. Formulation and intestinal absorption enhancement evaluation of water-in-oil microemulsions incorporating medium-chain glycerides. *Pharm Res* 1994;11:1385–1390.
21. Sekine M, Terashima H, Sasahara K, Nishimura K, Okada R, Awazu S. Improvement of bioavailability of poorly absorbed drugs. II. Effect of medium chain glyceride base on the intestinal absorption of cefmetazole sodium in rats and dogs. *J Pharmacobio-Dyn* 1985;8:286–295.

22. Lueßen HL, de Leeuw BJ, Langemeÿer WE, de Boer AG, Verhoef JC, Junginger HE. Mucoadhesive polymers in peroral peptide drug delivery. VI. Carbomer and chitosan improve the intestinal absorption of the peptide drug buserelin *in vivo*. *Pharm Res* 1996;13:1668–1672.
23. Van Asperen J, van Tellingen O, Sparreboom A, Schinkel AH, Borst P, Nooijen WJ, Beijnen JH. Enhanced oral bioavailability of paclitaxel in mice treated with the P-glycoprotein blocker SDZ PSC 833. *Br J Cancer* 1997;76:1181–1183.
24. Terwogt JMM, Beijnen JH, ten Bokkel Huinink WW, Rosing H, Schellens JHM. Coadministration of cyclosporin enables oral therapy with paclitaxel. *Lancet* 1998;352: 285.
25. Mayer U, Wagenaar E, Dorobek B, Beijnen JH, Borst P, Schinkel AH. Full blockade of intestinal P-glycoprotein and extensive inhibition of blood–brain barrier P-glycoprotein by oral treatment of mice with PSC833. *J Clin Invest* 1997;100:2430–2436.
26. Popick AC, Crouthamel WG, Bekersky I. Plasma protein binding of ceftriaxone. *Xenobiotica* 1987;17:1139–1145.
27. Patel IH, Kaplan SA. Pharmacokinetic profile of ceftriaxone in man. *Am J Med* 1984; 77:17–25.
28. Bekersky I, Lilja H, Lawrence I. Tacrolimus pharmacology and non clinical studies. FK506 to protopic. *Semin Cutan Med Surg* 2001;20:226–232.
29. Tsai JC, Weiner N, Flynn GL, Ferry JJ. Drug and vehicle deposition from topical applications: localization of minoxidil within skin strata of the hairless mouse. *Skin Pharmacol* 1994;7(5):262–269.
30. Stagni G, Shukla C. Pharmacokinetics of methotrexate in rabbit skin and plasma after iv-bolus and iontophoretic administrations. *J Control Release* 2003;93(3):283–292.
31. Ahsan F, Arnold J, Meezan E, Pillion DJ. Enhanced bioavailability of calcitonin formulated with alkylglycosides following nasal and ocular administration in rats. *Pharm Res* 2001;18(12):1742–1746.
32. Simon M, Wittmar M, Kissel T, Linn T. Insulin containing nanocomplexes formed by self-assembly from biodegradable amine-modified poly(vinyl alcohol)-graft-poly(L-lactide): bioavailability and nasal tolerability in rats. *Pharm Res* 2005;22(11):1879–1886.
33. Eljarrat-Binstock E, Domb AJ. Iontophoresis: a non-invasive ocular drug delivery. *J Control Release* 2006;110(3):479–489.
34. Hayden BC, Jockovich ME, Murray TG, Voigt M, Milne P, Kralinger M, Feuer WJ, Hernandez E, Parel JM. Pharmacokinetics of systemic versus focal carboplatin chemotherapy in the rabbit eye: possible implication in the treatment of retinoblastoma. *Invest Ophthalmol Vis Sci* 2004;45:3644–3649.
35. Odou P, Barthelemy C, Chatelier D, Luyckx M, Brunet C, Dine T, Gressier B, Cazin M, Cazin JC, Robert H. Pharmacokinetics of midazolam: comparison of sublingual and intravenous routes in rabbit. *Eur J Drug Metab Pharmacokinet* 1999;24(1):1–7.
36. Brown SA, Chester ST, Speedy AK, Hubbard VL, Callahan JK, Hamlow PJ, Hibbard B, Robb EJ. Comparison of plasma pharmacokinetics and bioequivalence of ceftiofur sodium in cattle after a single intramuscular or subcutaneous injection. *J Vet Pharmacol Ther* 2000;23(5):273–280.
37. Sanders P, Guillot P, Mourot D. Pharmacokinetics of a long-acting chloramphenicol formulation administered by intramuscular and subcutaneous routes in cattle. *J Vet Pharmacol Ther* 1988;11(2):183–190.
38. Dagorn M, Guillot P, Sanders P. Pharmacokinetics of chloramphenicol in sheep after intravenous, intramuscular and subcutaneous administration. *Vet Q* 1990;12(3):166–174.

39. Sanders P, Moulin G, Guillot P, Dagorn M, Perjant P, Delepine B, Gaudiche C, Mourot D. Pharmacokinetics of spiramycin after intravenous, intramuscular and subcutaneous administration in lactating cows. *J Vet Pharmacol Ther* 1992;15(1):53–61.

40. Nowakowski MA, Lynch MJ, Smith DG, Logan NB, Mouzin DE, Lukaszewicz J, Ryan NI, Hunter RP, Jones RM. Pharmacokinetics and bioequivalence of parenterally administered doramectin in cattle. *J Vet Pharmacol Ther* 1995;18(4):290–298.

41. Pietropaolo M, Olson CD, Reiseter BS, Kasaian MT, Happ MP. Intratracheal administration to the lung enhances therapeutic benefit of an MBP peptide in the treatment of murine experimental autoimmune encephalomyelitis. *Clin Immunol* 2000;95(2):104–116.

42. Durif F, Beyssac E, Coudore F, Paire M, Eschalier A, Aiache M, Lavarenne J. Comparison between percutaneous and subcutaneous routes of administration of apomorphine in rabbit. *Clin Neuropharmacol* 1994;17(5):445–453.

43. Ugwoke MI, Agu RU, Kinget R, Verbeke N. Intravenous versus subcutaneous injections of apomorphine in rabbits: a pharmacokinetic paradox. *Boll Chim Farm* 2003;142(8): 315–318.

44. Bergeron RJ, Wiegand J, Brittenham GM. HBED ligand: preclinical studies of a potential alternative to deferoxamine for treatment of chronic iron overload and acute iron poisoning. *Blood* 2002;99(8):3019–3026.

45. Doerge DR, Young JF, McDaniel LP, Twaddle NC, Churchwell MI. Toxicokinetics of acrylamide and glycidamide in Fischer 344 rats. *Toxicol Appl Pharmacol* 2005;208(3): 199–209.

46. Araya H, Nagao S, Tomita M, Hayashi M. The novel formulation design of self-emulsifying drug delivery systems (SEDDS) type O/W microemulsion I: enhancing effects on oral bioavailability of poorly water soluble compounds in rats and beagle dogs. *Drug Metab Pharmacokinet* 2005;20(4):244–256.

47. Swan GE, Botha CJ, Taylor JH, Mulders MS, Minnaar PP, Kloeck A. Differences in the oral bioavailability of three rafoxanide formulations in sheep. *J S Afr Vet Assoc* 1995;66(4): 197–201.

48. Short CR. Consideration of sheep as a minor species: comparison of drug metabolism and disposition with other domestic ruminants. *Vet Hum Toxicol* 1993;35(Suppl 2): 40–56.

49. Klausner EA, Lavy E, Stepensky D, Friedman M, Hoffman A. Novel gastroretentive dosage forms: evaluation of gastroretentivity and its effect on riboflavin absorption in dogs. *Pharm Res* 2002;19:1516–1523.

50. Klausner EA, Eyal S, Lavy E, Friedman M, Hoffman A. Novel levodopa gastroretentive dosage form: *in vivo* evaluation in dogs. *J Control Release* 2003;88:117–126.

51. Stepensky D, Friedman M, Srour W, Raz I, Hoffman A. Preclinical evaluation of pharmacokinetic–pharmacodynamic rationale for oral CR metformin formulation. *J Control Release* 2001;7:107–115.

52. Hoffman A, Stepensky D, Lavy E, Eyal S, Klausner E, Friedman M. Pharmacokinetic and pharmacodynamic aspects of gastroretentive dosage forms. *Int J Pharm* 2004;277(1–2): 141–153.

53. Eriksson UG, Bredberg U, Hoffmann KJ, Thuresson A, Gabrielsson M, Ericsson H, Ahnoff M, Gislen K, Fager G, Gustafsson D. Absorption, distribution, metabolism, and excretion of ximelagatran, an oral direct thrombin inhibitor, in rats, dogs, and humans. *Drug Metab Dispos* 2003;31(3):294–305.

54. Lin CC, Luu T, Lourenco D, Yeh LT, Lau JY. Absorption, pharmacokinetics and excretion of levovirin in rats, dogs and cynomolgus monkeys. *J Antimicrob Chemother* 2003;51(1): 93–99.

55. Holleran JL, Parise RA, Joseph E, Eiseman JL, Covey JM, Glaze ER, Lyubimov AV, Chen YF, D'Argenio DZ, Egorin MJ. Plasma pharmacokinetics, oral bioavailability, and interspecies scaling of the DNA methyltransferase inhibitor, zebularine. *Clin Cancer Res* 2005;11(10):3862–3868.

56. Szakacs T, Veres Z, Vereczkey L. *In vitro–in vivo* correlation of the pharmacokinetics of vinpocetine. *Pol J Pharmacol* 2001;53(6):623–628.

57. Jia L, Noker PE, Coward L, Gorman GS, Protopopova M, Tomaszewski JE. Interspecies pharmacokinetics and *in vitro* metabolism of SQ109. *Br J Pharmacol* 2006;147(5): 476–485.

58. Giuliano C, Fiore F, Di Marco A, Padron Velazquez J, Bishop A, Bonelli F, Gonzalez-Paz O, Marcucci I, Harper S, Narjes F, Pacini B, Monteagudo E, Migliaccio G, Rowley M, Laufer R. Preclinical pharmacokinetics and metabolism of a potent non-nucleoside inhibitor of the hepatitis C virus NS5B polymerase. *Xenobiotica* 2005;35(10–11):1035–1054.

59. Goodacre SC, Street LJ, Hallett DJ, Crawforth JM, Kelly S, Owens AP, Blackaby WP, Lewis RT, Stanley J, Smith AJ, Ferris P, Sohal B, Cook SM, Pike A, Brown N, Wafford KA, Marshall G, Castro JL, Atack JR. Imidazo[1,2-*a*]pyrimidines as functionally selective and orally bioavailable GABA(A)alpha2/alpha3 binding site agonists for the treatment of anxiety disorders. *J Med Chem* 2006;49(1):35–38.

60. Chung HJ, Choi YH, Choi HD, Jang JM, Shim HJ, Yoo M, Kwon JW, Lee MG. Pharmacokinetics of DA-6034, an agent for inflammatory bowel disease, in rats and dogs: contribution of intestinal first-pass effect to low bioavailability in rats. *Eur J Pharm Sci* 2006;27(4):363–374.

61. Li M, Wang H, Hill DL, Stinson S, Veley K, Grossi I, Peggins J, Covey JM, Zhang R. Preclinical pharmacology of the novel antitumor agent adaphostin, a tyrphostin analog that inhibits bcr/abl. *Cancer Chemother Pharmacol* 2006;57(5):607–614.

62. Borthwick AD, Davies DE, Exall AM, Livermore DG, Sollis SL, Nerozzi F, Allen MJ, Perren M, Shabbir SS, Woollard PM, Wyatt PG. 2,5-Diketopiperazines as potent, selective, and orally bioavailable oxytocin antagonists. 2. Synthesis, chirality, and pharmacokinetics. *J Med Chem* 2005;48(22):6956–6969.

63. Ahmed M, Briggs MA, Bromidge SM, Buck T, Campbell L, Deeks NJ, Garner A, Gordon L, Hamprecht DW, Holland V, Johnson CN, Medhurst AD, Mitchell DJ, Moss SF, Powles J, Seal JT, Stean TO, Stemp G, Thompson M, Trail B, Upton N, Winborn K, Witty DR. Bicyclic heteroarylpiperazines as selective brain penetrant 5-HT6 receptor antagonists. *Bioorg Med Chem Lett* 2005;15(21):4867–4871.

64. Yang Z, Zadjura L, D'Arienzo C, Marino A, Santone K, Klunk L, Greene D, Lin PF, Colonno R, Wang T, Meanwell N, Hansel S. Preclinical pharmacokinetics of a novel HIV-1 attachment inhibitor BMS-378806 and prediction of its human pharmacokinetics. *Biopharm Drug Dispos* 2005;26(9):387–402.

65. Tsuda-Tsukimoto M, Ogasawara Y, Kume T. Pharmacokinetics and metabolism of TR-14035, a novel antagonist of a4ss1/a4ss7 integrin mediated cell adhesion, in rat and dog. *Xenobiotica* 2005;35(4):373–389.

66. James JK, Nakamura M, Nakazato A, Zhang KE, Cramer M, Brunner J, Cook J, Chen WG. Metabolism and disposition of a potent group II metabotropic glutamate receptor agonist, in rats, dogs, and monkeys. *Drug Metab Dispos* 2005;33(9):1373–1381.

67. Lewi P, Arnold E, Andries K, Bohets H, Borghys H, Clark A, Daeyaert F, Das K, de Bethune MP, de Jonge M, Heeres J, Koymans L, Leempoels J, Peeters J, Timmerman P, Van den Broeck W, Vanhoutte F, Van't Klooster G, Vinkers M, Volovik Y, Janssen PA. Correlations between factors determining the pharmacokinetics and antiviral activity of HIV-1 non-nucleoside reverse transcriptase inhibitors of the diaryltriazine and diarylpyrimidine classes of compounds. *Drugs R D* 2004;5(5):245–257.
68. Washio T, Kohsaka K, Arisawa H, Masunaga H. Pharmacokinetics and metabolism of the novel muscarinic receptor agonist SNI-2011 in rats and dogs. *Arzneimittelforschung* 2003;53(1):26–33.
69. Monbaliu J, Verhaeghe T, Willems B, Bode W, Lavrijsen K, Meuldermans W. Pharmacokinetics of galantamine, a cholinesterase inhibitor, in several animal species. *Arzneimittelforschung* 2003;53(7):486–495.
70. Huang L, Edwards SR, Smith MT. Comparison of the pharmacokinetics of oxycodone and noroxycodone in male dark Agouti and Sprague–Dawley rats: influence of streptozotocin-induced diabetes. *Pharm Res* 2005;22(9):1489–1498.
71. Kaniwa N, Ogata H, Aoyagi N, Ejima A, Takahashi T, Uezono Y, Imasato Y. Bioavailability of cyclandelate from capsules in beagle dogs and dissolution rate: correlations with bioavailability in humans. *J Pharmacobiodyn* 1991;14(3):152–160.
72. Bramer SL, Forbes WP. Relative bioavailability and effects of a high fat meal on single dose cilostazol pharmacokinetics. *Clin Pharmacokinet* 1999;37(Suppl 2):13–23.
73. Schug BS, Brendel E, Chantraine E, Wolf D, Martin W, Schall R, Blume HH. The effect of food on the pharmacokinetics of nifedipine in two slow release formulations: pronounced lag-time after a high fat breakfast. *Br J Clin Pharmacol* 2002;53(6):582–588.
74. Johnson JT, Mattiuz EL, Chay SH, Herman JL, Wheeler WJ, Kassahun K, Swanson SP, Phillips DL. The disposition, metabolism, and pharmacokinetics of a selective metabotropic glutamate receptor agonist in rats and dogs. *Drug Metab Dispos* 2002;30(1): 27–33.
75. Ogo Y, Sunagane N, Ohta T, Uruno T. Banana juice reduces bioavailability of levodopa preparation. *Yakugaku Zasshi* 2005;125(12):1009–1011.
76. Homer LM, Clarke CR, Weingarten AJ. Effect of dietary fat on oral bioavailability of tepoxalin in dogs. *J Vet Pharmacol Ther* 2005;28(3):287–291.
77. Jann MW, ZumBrunnen TL, Tenjarla SN, Ward ES Jr, Weidler DJ. Relative bioavailability of ondansetron 8-mg oral tablets versus two extemporaneous 16-mg suppositories: formulation and gender differences. *Pharmacotherapy* 1998;18(2):288–294.
78. Yang SH, Bae SK, Kwon JW, Yoo M, Lee MG. Gender differences in the pharmacokinetics of DA-6034, a derivative of flavonoids, in rats. *Biopharm Drug Dispos* 2006;27(1): 47–51.
79. Crouthamel WG, Bekersky I. Preclinical evaluation of new drug candidates and drug delivery systems in the dog. In Crouthamel WG, Sarapu AC, Eds. *Animal Models for Oral Drug Delivery in Man*: *In Vitro and In Vivo Approaches*. Washington DC: American Pharmaceutical Association; 1983, pp 107–123.

# MASS BALANCE STUDIES

Jan H. Beumer, Julie L. Eiseman, and Merrill J. Egorin
*University of Pittsburgh Cancer Institute, Pittsburgh, Pennsylvania*

**Contents**

1 Introduction
2 Definitions
3 Study Design
  3.1 Timing of a Mass Balance Study
  3.2 Guidelines
  3.3 The Radioisotope and Its Position in the Molecule
  3.4 Animals
  3.5 Dosing
4 Methodology
  4.1 Dose Administration
  4.2 Sample Collection and Analysis
  4.3 Metabolic Profiling
5 Results
  5.1 Mass Balance Calculation
  5.2 Plasma Pharmacokinetics
  5.3 Assessment of Absorption
  5.4 Organ Distribution
  5.5 Metabolic Profiling
  References

## 1 INTRODUCTION

A well-designed mass balance study significantly expands information about the pharmacological behavior of an active compound. Among many things, a preclinical

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

mass balance study can elucidate excretory pathways, organ distribution of parent drug and metabolites, and metabolic profiles. Moreover, a preclinical mass balance study can set the stage for its clinical counterpart, the human mass balance study.

**Possible Objectives of a Preclinical Mass Balance Study**

- Identify the excretory pathways of a drug (in general, urine and feces).
- Assess extent and kinetics of metabolism. Is the compound rapidly metabolized and how fast are the produced metabolites eliminated?
- Assess extent and kinetics of excretion. Is all radioactivity eventually excreted, and how long must excreta be collected?
- Elucidate the metabolic fate of a drug. How many metabolites can be observed and/or identified, and how much of the radioactivity do they account for?
- Assess organ distribution. Does (ir)reversible tissue binding occur, explaining incomplete recovery of radioactivity in excreta? Are there specific organs displaying accumulation?
- Provide information to optimally perform the human mass balance study. Is the choice of the isotope and its position in the molecule correct? Generate metabolite reference compounds and predict the radiation exposure of participants in the human mass balance study.

In general, a "mass balance study" is exactly what the term implies. This study involves the "balancing" of quantities of "mass." The "mass" that leaves the system (parent drug and metabolites excreted by various routes like urine and feces) is "balanced" against the known "mass" that was entered into the system (an amount of parent drug administered to the organism). Since most drugs undergo metabolic conversion before being excreted, a radioisotopically labeled variant of the compound is commonly administered to facilitate the quantitation of the sum of parent drug and metabolites in a variety of biological matrices. We propose the following definition for a mass balance study: "A pharmacokinetic study, aimed at characterizing the metabolic and excretory fate of a compound by collecting and analyzing excreta after administration of the compound, employing an analytical technique that allows tracing and quantitation of the compound and its products in spite of metabolic conversion."

Mass balance studies in their simplest form (as often performed in humans) are usually limited to just this particular balance, the amount that is recovered in excreta after an extended collection period, relative to the amount originally administered. However, contrary to a human mass balance study, a preclinical mass balance study allows the calculation of this mass balance at any given time by quantitating radioactivity present in various organs as well as in excreta. Thus, a preclinical mass balance study can provide information about the whereabouts of drug and metabolites inside and outside the organism at any time. This is the unique trait of the preclinical mass balance.

The relevance of the preclinical mass balance study is not limited to only preparing for the human mass balance study. It provides crucial information for the toxicokinetic evaluation of a compound. Toxicokinetic studies are aimed at integrating pharmacokinetics/metabolism and toxicity in nonclinical safety studies. Good agreement of the pharmacokinetics and metabolism of a drug in a specific animal species and humans increases the validity of the safety studies in this species for clinical safety issues [1–4]. Therefore, when there is close agreement between a specific animal species and humans with regard to excretory pathways and metabolic profile (characteristics that are best investigated by the mass balance study), this species is likely to be a better model for safety studies of the respective drug. The importance of metabolites and interspecies comparison is reflected by the recent distribution of a draft guidance for industry by the Center for Drug Evaluation and Research entitled *Safety Testing of Drug Metabolites* [5], which states the following. "Based on data obtained from *in vivo* and *in vitro* metabolism studies, when the metabolic profile of a parent drug is similar qualitatively and quantitatively across species, we can generally assume that potential clinical risks of the parent drug and its metabolites have been adequately characterized during standard non-clinical safety evaluations" [5]. Identification and pharmacologic characterization of individual metabolites is key in comparing results of preclinical studies with those of human studies [4].

The mass balance study is a unique and complicated study. It generally involves the use of radioisotopes and analysis of radioactivity, parent compound, and metabolites in a variety of matrices. As will be shown, a wealth of information can be gained from a mass balance. Fully exploiting a preclinical mass balance study requires careful design and the use of many techniques. In this chapter, the design of preclinical mass balance studies is addressed. The methodologies and expertise required for the conduct of these studies are provided, and we propose ways to express the results. Examples of the usefulness of mass balance studies in drug development are given and used to illustrate the importance of understanding the data obtained from mass balance studies.

## 2 DEFINITIONS

*Mass Balance Study.* A pharmacokinetic study, aimed at characterizing the metabolic and excretory fate of a compound by collecting and analyzing excreta after administration of the compound, employing an analytical technique that allows tracing and quantitation of the compound and its products in spite of metabolic conversion.

*Total Radioactivity.* The radioactivity content of a matrix, which consists of the sum of the parent drug and its metabolites that still contain the radioactive label.

*Metabolite.* Any compound being produced from the handling of a parent compound in the body.

*Radioisotope.* An isotope of an element capable of spontaneously emitting energetic particles by the disintegration of its atomic nuclei.

*Tracer.* The added (radioisotopical) portion of a compound used to follow the course of the metabolic process of that compound.

*Specific Activity.* A measure of the radioactivity per unit mass of a compound.

*Autoradiolysis.* Radiolysis of a radioactive material resulting directly or indirectly from its radioactive decay.

## 3 STUDY DESIGN

### 3.1 Timing of a Mass Balance Study

Performing a mass balance study with one or more animal species can be a costly and time-consuming investment. Hence, there needs to be evidence of activity at specified doses and dose schedules, initial toxicity studies must be satisfactory, and pharmacokinetics of unchanged drug should have been established. As will be argued, rational planning of a mass balance study requires knowledge of *in vitro* metabolic profiles of the compound in the species being used in preclinical safety studies and considered for the mass balance study. This will yield the most relevant information for clinical trials. Metabolic studies in animals may be performed in the late phase of the lead optimization [6]. However, even if the compound has already entered clinical trials, a preclinical mass balance can still be of use. If there are differences in (side) effects between humans and animals, these discrepancies may be explained by the *in vivo* metabolic profiles of the species.

### 3.2 Guidelines

***Role of a Preclinical Mass Balance Study in Complying with Guidelines for Drug Development*** The U.S. FDA [7] and its European counterpart, EMEA [8], have a number of guidelines delineating information that is required to lead a compound through the drug development process. Although there is no guideline specifically for preclinical mass balance studies, a preclinical mass balance study can play a role in obtaining information crucial to drug development. The preclinical mass balance study provides a description of the metabolism of a drug and the contribution of metabolism to overall elimination. This information is crucial in the development of a drug [9]. In determining the maximum recommended starting dose (MRSD) for initial clinical trials from a human equivalent dose, as a default, the most sensitive animal species is used. However, based on differences in absorption, distribution, metabolism, and excretion between species, a different species may be selected, thus increasing the MRSD and reducing the time spent on initial clinical trials [10]. Furthermore, a mass balance study provides the information that is assessed by the (draft) guidance, *Safety Testing of Drug Metabolites* [5].

***Performing a Mass Balance Study*** In performing a preclinical mass balance study, a number of guidelines and regulations must be followed. As regulations may vary by country, one should refer to the locally applicable institutional guidelines, but these guidelines will generally cover: IACUC, an institutional review board (IRB), working with radioactivity in general and in combination with animals specifically. As always in working with radioactivity, contamination must be minimized by working on absorbents, using disposable tools, and performing sweep tests after

every experiment. Surveys with detectors will not suffice because they are not sensitive enough to detect all radioisotopes (including those most commonly used—$^{14}C$ and $^{3}H$).

### 3.3 The Radioisotope and Its Position in the Molecule

The choice of radioisotope and its position in the molecule can have a major impact on chemical and metabolic stability and can thus determine the results of a mass balance study. Use of $^{14}C$ as the radioactive isotope is generally preferable to any other radioisotope. The labeled compound only serves its purpose if it behaves identically to the nonlabeled compound and if the radioisotope is not lost chemically by exchange or during metabolism. Prior knowledge of the metabolism of a compound (from microsomal incubation studies, known analogues, or *in silico* predictions) can assist in determining the optimal radioisotope and its position. Since ideally the same radiolabeled compound is used in both the preclinical and clinical mass balance studies, human metabolic data must be considered as well.

If the studied molecule is large and/or is expected to divide metabolically into two parts, dual labeling may be considered. By labeling one part of the molecule with $^{3}H$ and the other part with $^{14}C$, both parts may be traced independently [11, 12]. Even triple labeling has been reported [13].

***Isotopes*** Several different radioisotopes can be employed to trace the whereabouts of a drug and its metabolites. In choosing the isotope to be used, one should be aware of the type of decay, the decay half-life, and the energy of the emitted particle (see Table 1). The type of radiation that is easiest to handle is $\beta^-$ radiation (electrons). Working with γ-emitters requires use of lead shielding, because the highly energetic γ-radiation easily penetrates solid matter and tissue. Contrary to γ-radiation, the surroundings of a β-source are easily shielded from radiation by a layer of plastic/Plexiglas, and the radiation of, for example, $^{14}C$ will only penetrate up to 50 cm of air [12]. In addition, detection of β-radiation is readily accomplished by liquid scintillation counting. If a radioisotope has a short half-life, care must be taken to minimize the time between drug preparation and usage in the study or the labeled drug will have degraded. The following formula can be used to calculate the percentage of radioisotope remaining after preparation of a radioisotopically labeled compound:

**TABLE 1 Radioisotopes Employed in Preclinical Mass Balance Studies** [12, 14]

| Radioisotope | Type of Decay | Half-life | Energy (MeV) | Comments |
|---|---|---|---|---|
| $^{3}H$ | β– | 12.3 y | 0.019 | Easily lost, low energy |
| $^{14}C$ | β– | 5,730 y | 0.156 | Long half-life |
| $^{32}P$ | β– | 14.3 d | 1.71 | Correction for decay, complicated handling |
| $^{35}S$ | β– | 87.5 d | 0.167 | Correction for decay |
| $^{36}Cl$ | β– | 300,000 y | 0.71 | Long half-life |
| $^{131}I$ | β– | 8.0 d | 0.80–2.16 | Correction for decay |
| | γ | | 0.67, 0.77 | |

$$\%\ \text{Remaining} = 100 \cdot \frac{A_t}{A_0} = 100 \cdot (2)^{-(t/t_{1/2})}$$

where $A_0$ is radioactivity at time $t = 0$, $A_t$ is radioactivity at time $t$, and $t_{1/2}$ is decay half-life. In addition, quantitated radioactivity levels must be corrected for the time expired between administration and quantitation. The following formula can be used to calculate the activity at $t = 0$ from the radioactivity quantitated at time $t$:

$$A_0 = A_t \cdot (2)^{(t/t_{1/2})}$$

Given the atomic composition of drug molecules, the most commonly used radioisotopes are $^{14}C$ (radiocarbon) and $^{3}H$ (tritium), although the use of other radioisotopes has been documented [12]. The radiation emitted by $^{14}C$ is about eightfold more energetic than that of $^{3}H$ (0.156 MeV versus 0.019 MeV). Thus, quantitation of $^{14}C$ is easier and suffers less from matrix effects [14]. Even so, the use of $^{3}H$ may be required when very small amounts of a compound need to be detected, since higher specific activities may be achieved with $^{3}H$ (29.1 Ci/milliatom) than with carbon (62.4 mCi/milliatom) [12].

Although positron emitters can have their use in investigating drug pharmacology, their generally very short half-life (<2 h) limits their use in mass balance studies [15].

On the rare occasion that a drug contains an element that occurs in the body only at trace levels, the use of a radioisotopically labeled compound may be avoided. For example, metal-containing compounds may be quantitated using element-specific techniques such as atomic absorption spectrometry (AAS), inductively coupled plasma atomic emission spectrometry (ICP-AES), or inductively coupled plasma mass spectrometry (ICP-MS) [16, 17], and fluoro compounds can be traced employing $^{19}F$-NMR [18].

Although it would seem obvious that after labeling with the radioisotope the compound must have the same chemical formula, this is not always the case. Occasionally, the disposition of a compound is investigated by substituting with a foreign (radioactive) element [19]. This substitution results in a different compound, and consequently, the new compound likely displays different metabolic and dispositional characteristics.

***Chemical Stability*** A major disadvantage of using $^{3}H$, despite the ease and low costs, is its tendency to be exchanged with hydrogen in aqueous matrices, producing nonlabeled compound and tritiated water. Especially if $^{3}H$ bonded to heteroatoms or α-positioned relative to a double bonded heteroatom, it is likely to be exchanged [20, 21]. A $^{14}C$ atom is not likely to be chemically exchanged.

***Metabolic Stability*** The use of a compound labeled at a chemically stable position does not guarantee that the detected radioactivity adequately represents its metabolism and disposition. The position needs to be metabolically stable as well. For example, positioning a radioisotope in an *N*-methyl moiety or the side chain part of an ester [22] (although easy to synthesize) will not result in a useful compound to trace metabolites. The radioisotope would quickly be lost by oxidative dealkylation and hydrolysis, respectively [23]. The isotope needs to be positioned in a relatively

central part of the compound to enable tracing of its primary metabolites. It is therefore not surprising that the radioisotope is preferentially positioned in an aromatic or alicyclic ring system. Computer-assisted metabolism prediction (CAMP) is available to predict the metabolic stability by means of databases of known metabolic transformations. The suitability of a label may be predicted by *in vitro* (liver microsomal incubation or hepatocytes) testing of the compound [24] or by reviewing the metabolism of analogues. For example, a review of the literature on the metabolism of pyrimidine antimetabolites shows the metabolic loss of the 2-carbon as carbon dioxide, while the 6-carbon is quite stable [25]. Thus, when deciding on the position of the $^{14}C$ in an investigational pyrimidine compound, the 6-position is preferred over the 2-position.

***Kinetic Isotope Effect*** Although a radioisotope is in many regards equal to its stable-isotope counterpart, chemically it may still behave slightly differently. Heavier isotopes will form stronger chemical bonds, and thus reaction kinetics (like metabolic conversion) will be affected. This is called the *kinetic isotope effect* [21]. Thus, the radioisotopically labeled compound may well undergo a different metabolism and display a different disposition from the nonlabeled compound. The replacement of hydrogen ($^{1}H$) with tritium ($^{3}H$) is a tripling in mass, while the substitution of a carbon atom ($^{12}C$) with radiocarbon ($^{14}C$) is only a 17% increase. Therefore, the use of tritiated compounds is more likely to yield results biased by the kinetic isotope effect [21]. However, the kinetic isotope effect is relevant only if the rate-limiting step in the overall metabolic conversion is affected. Generally, cleavage of the carbon–hydrogen bond is the rate-limiting step in the oxidation of aliphatic hydrocarbons. Whether or not this is the case for an aromatic carbon–hydrogen bond depends on the substitution pattern [26]. An example illustrating the kinetic isotope effect is the reduced formation of the genotoxic α-hydroxy metabolite of tamoxifen in rats after exchange of hydrogen for deuterium [27].

### 3.4 Animals

Nonclinical safety testing is usually performed in rats and mice, and a nonrodent mammal [28, 29]. The default rule for starting a first-in-human-study is two mammals, one of which is nonrodent. When dealing with biologicals, species selection will be linked to the presumed mechanism of action and the presence of the target in the available species. In the literature, preclinical mass balance studies are often reported in a rodent species, in dog, and in a primate. However, the choice of species must be a conscious one. Every animal has its own characteristics that influence the metabolic fate of a drug, and the costs and ease of performing the study. Even within species, different strains may have different properties. Inbred strains are the result of several generations of brother × sister mating. Consequently, inbred animals are homozygous at almost all loci, except sex. The genetic homogeneity of present and future generations ensures homogeneity in phenotype and thus more consistent experimental results. Conversely, outbred animals display a more variable genotype and phenotypic behavior.

***Metabolic Differences*** Ideally, the animal species to use for the preclinical mass balance study is the one showing the highest degree of *in vitro* metabolic

concordance with humans. Early identification of human metabolic routes of elimination and metabolites by *in vitro* studies can provide a clear direction for preclinical studies in animals [30, 31]. *In vitro* incubations of the drug with liver slices, hepatocytes, and microsomal and nonmicrosomal fractions can provide useful information to assess this metabolic concordance. It should be noted that these *in vitro* incubations are unable to capture all aspects of the *in vivo* metabolism. First, gastrointestinal, kidney, and lung metabolism and biliary excretion are not assessed in these experiments. Second, phase II metabolism does not necessarily occur in microsomes without addition of the various conjugating moieties, and liver slices have limited enzymatic stability [30].

Obviously, known interspecies differences in metabolism should be taken into account. If the drug has an aromatic amine moiety, it is likely a substrate for *N*-acetyltransferase. Since dogs lack this enzyme [31], to perform a mass balance study with this drug in dogs is not expected to yield information that is relevant to human metabolism, disposition, and safety of this drug. Table 2 lists some reported metabolic differences of animal species relative to humans. In addition to drug-metabolizing enzymes being either present or absent in different species, the organ distribution can also vary widely [32]. With the advent of genetically altered animal strains, some of these differences can be diminished, if not eliminated. For example, esterase activity in rodents greatly exceeds esterase activity in humans. Currently, plasma-esterase-deficient mice have become available, providing a better metabolic model for compounds that are a likely substrate for esterases [33, 34].

***Animal Size*** Animal size has many implications for the mass balance study. A large animal allows repeated blood sampling. It also produces more, more frequent, but likely less concentrated urine that can enable metabolite profiling. On the other hand, large animals are more costly to buy and to keep, are less easily handled,

**TABLE 2 Some Metabolic Differences of Animal Species Relative to Humans**

| Species | Enzyme | Effect | Substrates | References |
|---|---|---|---|---|
| Mouse | Carbonyl reductase | Low activity | Anthracyclines | 31 |
| | Carboxylesterase | High activity | Esters, lactones | 34 |
| | Deoxycytidine kinase | Low activity (10-fold) | Cytidine analogues | 87 |
| Rat | Amidase | High activity | Amides | 88 |
| | Carboxylesterase | High activity | Esters, lactones | 22, 34 |
| | CYP2C8 | Substrate mismatch | Paclitaxel | 31 |
| | Cytidine deaminase | Absent | Cytidine analogues | 32 |
| | *N*-acetyltransferase | High activity | Aromatic amines | 31 |
| Dog | Cytidine deaminase | Low | Cytidine analogues | 89 |
| | *N*-acetyltransferase | Absent | Aromatic amines | 31 |
| All | Microsomal epoxide hydroxylase | Low | Epoxides | 90 |
| | UGTs | Negligible | Various, SN38, AZT | 31 |

and require a larger amount of the (radioisotopically labeled) investigational compound.

***Age*** In general, young but mature adult animals are used for pharmacokinetic studies. Many enzyme activity levels are age dependent. In immature animals, enzyme activity may not have fully developed yet. In older animals, activity of some enzymes may be declining while others may increase. In addition, body composition changes, with an increasing fat percentage over time [35–37]. Most importantly, the results must be comparable with preclinical safety studies, and the age of the animals used in those studies can be used as a guideline.

***Sex*** Sex can be an important determinant of pharmacokinetic and metabolic parameters in animals [38–42]. Sex-dependent differences may be predicted by *in vitro* studies [43]. However, sex-dependent metabolic differences in humans are less dramatic, even though they may be present [44]. Therefore, the sex of the animals to be used may be determined by the methodology of the preceding safety studies.

***Sample Size*** The sample size of the study is dependent on the information that needs to be obtained, and it obviously influences costs and logistics. If only an excretory mass balance study is needed, it will suffice to use few animals (three or four should be a minimum), and to collect excreta for a specified time (e.g., until an acceptable portion of the dose has been recovered). However, if the plasma metabolic profile of the compound is to be investigated, serial sampling is required, which is often destructive in smaller animals. A minimum of three or four animals should be used per time point to allow determination of a standard deviation. The number of time points is dependent on whether the investigation of the metabolic profile is more qualitative in nature, or the complete plasma profile of the drug, total radioactivity, and/or metabolites need to be determined. Usually, larger animals are studied in a smaller sample size for both cost and ethical reasons.

***Genetically Modified Species*** As mentioned earlier with regard to esterase-deficient mice, it is possible to use genetically modified strains of a species in order to better mimic the human metabolic system. However, there is another application for genetically modified strains. Various gene-knockout (KO) and gene-knockin (KI) models have been developed for the mouse, to illustrate the *in vivo* function of a variety of enzymes/efflux pumps, for example, the CYP3A4-KI [45], *mdr1a*(–/–) [46], *mdr1a/b*(–/–) [6, 47], *ABCG2*(–/–) [48], and *SLC22A1*(–/–) [49]. When a mass balance of an IND is performed in these genetically altered strains and their wild-type counterparts, basic information can be obtained pertaining to the excretory function of the efflux pumps. In addition, quantitation of unchanged drug or radioactivity in the organs (especially those known to express efflux pumps like brain, testis, and heart) [47, 50–53] of these animals shows the relevance of efflux pumps to organ disposition. These data can predict possible interactions with other drugs competing for or blocking the efflux pumps. To diminish interspecies differences in substrate specificity of CYP450 enzymes, mice with humanized enzymes are under development [54, 55].

### 3.5 Dosing

***Pharmacological Dose*** Assuming that the compound and its radioisotopically labeled counterpart behave identically, the pharmacological dose consists of the sum of unlabeled and labeled compound. Usually, the contribution of the radioactive part is negligible relative to the unlabeled amount of drug, hence the use of the term "tracer." However, to obtain a sufficient specific activity (Bq/mg) of a potent (dosed in low amounts), high molecular weight drug, a large fraction of the molecules (e.g., 80% [56]) will actually have to be labeled. To be relevant, the pharmacological dose should be comparable to pharmacological or toxicological relevant doses as established in safety studies. Even if linear pharmacokinetics of the parent compound has been shown, the metabolic disposition may be nonlinear. Therefore, the pharmacological dose should not deviate too much from previously established doses.

***Radioactive Dose*** The radioactive dose to be used depends on the specific objectives of the study and the sensitivity of the analytical methods. If the objective is to describe plasma and tissue profiles of the parent drug and metabolites (e.g., by HPLC with radioactivity detection), a higher radioactive dose may be required than for merely describing the excretory fate of a drug. The amount of radioactivity required is also dependent on the pharmacokinetic characteristics of the compound.

A literature search for mass balance studies with chemically related compounds and previous pharmacokinetic studies with the compound under investigation can give an impression of what results may be obtained with a certain dose. Next, a pilot study will provide valuable samples to test an analytical method and to assess the adequacy of the radioactive dose.

The radioactive dose is usually expressed in curies (Ci), as opposed to the SI unit of becquerel (Bq). The conversion of these units is 1 mCi = 37 MBq and 1 Bq = 60 DPM, where DPM means disintegrations per minute. Typical radioactive doses ($^{14}C$) are approximately 1–2 μCi/mouse, 5–20 μCi/rat, 5–100 μCi/dog, and 40–120 μCi/monkey, compared to 100 μCi/human. Recently, the use of accelerator mass spectrometry (AMS) was described in a mass balance study in dogs, requiring only 45 nCi of radioactivity [57]. AMS has received attention for its potential use in microdosing humans [58]. Given the complexity of the technique, its use in preclinical mass balance studies will likely be limited.

***Dosing Schedule*** Most mass balance studies are performed as a single-dose study. However, when single-dose tissue distribution studies suggest that the apparent half-life of the test compound (and/or metabolites) in organs or tissues significantly exceeds the apparent half-life of the elimination phase in plasma and is also more than twice the dosing interval in the toxicity studies, repeated-dose tissue distribution studies may be appropriate [4, 59].

***Route of Administration*** For the mass balance study to be relevant to clinical practice, the route of administration should mimic the anticipated clinical route of administration as much as possible. For comparative and/or scientific purposes, different routes of administration can be studied as well.

## 4 METHODOLOGY

### 4.1 Dose Administration

***Preparation*** Parenteral dosing solutions should be prepared under aseptic conditions. Because radioisotopically labeled compounds may not be available as a sterile product, one should be aware of possible bacterial contamination of the dosing solution. Aseptic filtration may be used, but loss of compound by adsorption to the filter unit must be assessed. The dosing solution should be prepared in excess to enable qualitative and quantitative analysis after completion of the study.

***Administration*** The dose should be administered according to established practice. There are guidelines for maximum volumes that may be administered to animals, which should be followed. Oral administration is usually performed after an overnight fast to assure rapid and reproducible gastrointestinal transit. When dosing multiple groups of animals, stratifying all the animals according to body weight and using one mouse of every stratum per group may achieve a homogeneous weight distribution.

When an infusion device is used, this must be flushed after administration to account for any radioactivity that may have remained in the lines. An organic solvent like methanol may be best suited for this. The amount of radioactivity recovered from the flushing solvent must be subtracted from the calculated dose administered to obtain the actual administered dose [25].

***Quality Control*** Preclinical studies are usually conducted with a nonstandardized formulation of the compound, if it is formulated at all. This is especially true for a mass balance study, because the isotopically labeled compound is often custom synthesized in one batch [60]. In addition, the dose consists of both the isotopically labeled part and the nonlabeled part.

First, for both nonlabeled drug and tracer, the content needs to be determined. This is essential to assure that a relevant pharmacological dose was administered, and to calculate eventually the recovery of the dose in the excreta. In case of a known, established specific activity, the dose of drug can be determined by taking an aliquot for scintillation counting. Using the specific activity of the compound (Bq/g), one can convert the radioactive content to total drug content.

Second, the purity of the dose needs to be established. Given the incidental production of most radioisotopically labeled compounds, it is likely that the purity is not as high as with standard reference compounds and may even be lower than the declaration of the manufacturer [21]. In addition, the radioisotopically labeled compound is likely to be less stable than the nonlabeled compound. Due to the inherent radioactive nature of the compound, energy is released, producing free radicals, which may increase degradation of the compound, a process called *autoradiolysis*. In solution, ethanol, serum albumin, and ascorbic acid have been reported to slow down autoradiolysis [61–63]. In the literature, the purity of the compound used is often not reported. Yet a purity of at least 95% is essential to perform a mass balance study. Impurities can have different pharmacokinetic characteristics and may be toxic. The radiochromatographic system used for metabolic profiling may be employed as a stability-indicating assay for the compound and, afterwards,

the dosing solution. In general (with compounds containing a chromophore), HPLC with UV absorbance or photodiode array UV detection will be able to assess the overall purity of both labeled and nonlabeled compound together.

### 4.2 Sample Collection and Analysis

A mass balance study can consist of merely collecting the excreta after administration of a radiolabeled drug. These studies do not characterize the metabolism, distribution, or disposition of the compound within the body. However, much more information can be obtained from a mass balance study by sampling various matrices in time, as is discussed later.

In general, no stability data are available for the compound and its metabolites in the different matrices (and organs may contain drug-metabolizing enzymes). Consequently, immediate storage at –20 °C or less is advisable, in order to prevent degradation products to be mistakenly identified as metabolites.

***Schedule*** In order to reliably estimate pharmacokinetic parameters, sampling should be adequately planned. In the case of intravenous administration, an early plasma sample is imperative to determine the area under the plasma concentration versus time curve. Thereafter, sampling must be frequent enough to separately characterize the distribution and the elimination phases. The elimination phase is ideally sampled for at least 4 half-lives. In the case of orally administered drugs, frequent plasma sampling should be performed around the expected time of maximum plasma concentration. Sometimes there is no data available to help schedule the sampling. A pilot study in a couple of animals may then be performed to get a rough estimate of the plasma concentration versus time profile, which can also serve to set the required concentration range of the analytical method.

Ideally, excreta are collected until all radioactivity administered has been recovered. However, 100% recovery will usually not be achieved [25]. Under the assumption of instantaneous excretion, 93% of radioactivity will have been excreted after four terminal radioactivity plasma half-lives. Unfortunately, the elimination half-life of radioactivity from plasma is usually not known. Another reported approach uses radioactivity levels in urine and feces for drug and radioactivity [25]. When less than 1% of administered radioactivity is excreted for 2 consecutive days, collection can be stopped. Prolonged collection will not substantially increase recovery. Small animals may be housed together to collect pooled excreta.

***Urine*** Urine is a liquid matrix; it is easily mixed, aliquoted, and quantitatively analyzed. Urine can be collected separately from feces with a metabolic cage (Fig. 1). Because urine is often used for qualitative analysis, it needs to be collected on ice in order to halt any degradation processes. To calculate the total excretion of a compound and its metabolites, the volume of urine excreted in the collection period must be accurately established. This can be accomplished by determining the volume with a graduated cylinder or by weighing, but the latter only if the relative density of the urine is near 1. At the end of the collection period, the collection part of the metabolic cage should be washed with a solvent to remove any adsorbed radioactivity from the device. The solvent used for this wash step must be chosen such that

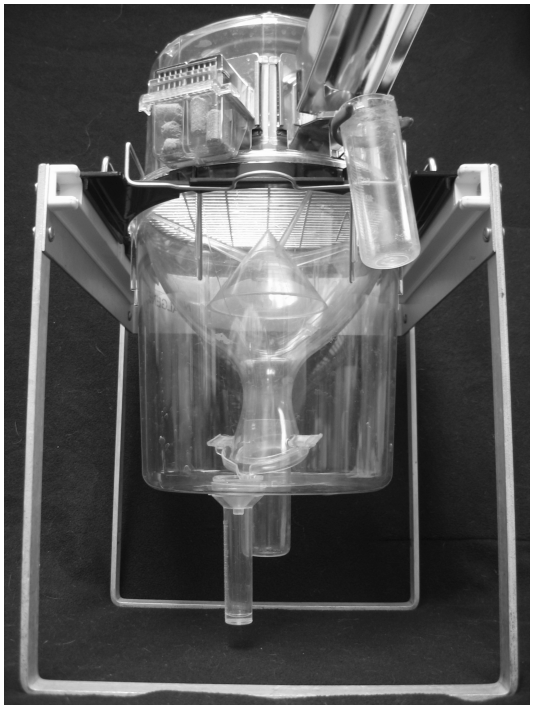

**FIGURE 1** Metabolic cage for the separate collection of urine (lower left collection cone) and feces (middle collection cone) of rodents.

the compound is expected to dissolve easily. The radioactivity found in this wash fluid should be calculated by multiplying the volume of wash solvent that was used (not the volume that was collected) with the concentration of radioactivity. If the investigated compound is known, or suspected, to adsorb to container surfaces, addition of an organic solvent or bovine serum albumin may be considered before subsequent aliquoting and processing.

***Feces*** Feces are collected separately from urine with metabolic cages. They are weighed and then homogenized with a homogenization device to obtain a pipettable, homogeneous matrix. Disposable homogenizer probes are available and may be preferable when processing radioactive samples. Depending on the eventual use of the homogenate (extraction of individual metabolites) and the properties of the compound and its metabolites (solubility, adsorption to container surfaces), homogenization may be done with water, buffer, or bovine serum albumin solution. In general, adsorption to container surfaces will not be an issue with fecal homogenate because the fecal particles will have a high surface area relative to the container surface area. The ratio with which feces are homogenized (varying from 2:1 to 10:1, v/w) will affect pipettability and sensitivity of subsequent analyses. A more dilute homogenate will result in a truly pipettable matrix, with proportionally reduced concentrations of the analytes. More concentrated homogenates may be pipettable, but not accurately so. When determining the radioactivity content, aliquots of

homogenate should be weighed. Because feces is a dark-colored matrix, decoloration prior to liquid scintillation counting may be warranted.

***Bile*** Occasionally, bile is collected to assess the biliary excretion of radioactivity and analytes. Because reabsorption of bile salts is essential for normal bile production (both quantitatively and qualitatively), bile may not be sampled over extended periods of time, that is, longer than 1–2 hours [64, 65]. An elegant but intricate way to deal with this issue is to replenish with blank bile from undosed animals while sampling the bile from the treated animal. Similarly, the contribution of enterohepatic recirculation may be assessed [66].

***Blood and Red Blood Cells*** Blood must be collected from a site distant from that of administration. This is done to prevent residual dosing solution at the administration site from biasing the concentration of the drawn sample. Next, one can determine total radioactivity in either whole blood or in red blood cells (RBCs) after centrifugation and aspiration of plasma. In combination with plasma radioactivity levels, this allows calculation of the blood/plasma ratio or the RBCs/plasma ratio, respectively. These parameters can be interconverted, but such a conversion requires information about the hematocrit [67].

Because red blood cells and blood can be somewhat viscous, an aliquot should be accurately weighted in a sample vial. Liquid scintillation counting of blood or red blood cells is often difficult due to the intense red color, which necessitates decoloration before liquid scintillation counting. A suitable protocol may be: weighting approximately 150 mg of the red blood cell pellet into a 20 mL scintillation vial; adding 500 μL of tissue solubilizer and 500 μL of water, followed by vortexing and incubating for 10 min at 50 °C; adding 100 μL of 0.1 M EDTA and 200 μL of 30% hydrogen peroxide, followed by vortexing of the sample and incubation at 50 °C for 10 min; finally, 20 mL of scintillation fluid may be added prior to scintillation counting.

***Plasma*** Plasma can be aspirated after centrifugation of blood. Because it is a relatively colorless matrix, liquid scintillation counting is straightforward. It consists of mixing an aliquot with liquid scintillation fluid followed by counting.

***Expired Air*** For a number of positions of the radioactive $^{14}C$ label (e.g., 2-position of pyrimidines, *N*-methyl), the expiration of radioactivity as $^{14}CO_2$ is predictable [68]. However, it may not be possible to prevent expiration of radioactivity, especially when there is no information about the metabolic fate of a compound or any analogues. When mass balance is not achieved (i.e., a large part of the radioactive dose cannot be accounted for in the total of excreta, organs, and carcass), one must consider the possibility of loss through expiration as $^{14}CO_2$.

Expired air may be sampled by keeping the animals in an airtight chamber with continuous air replacement and leading the expired air through a basic solution, thus precipitating the $^{14}CO_2$ as the nonvolatile carbonate. A sodium hydroxide [12] or 2-methoxyethanol-ethanolamine (2:1) solution [69] can be used to capture the carbon dioxide. Before air is pumped into the chamber it may need to be led through a similar basic solution to eliminate any natural $^{14}CO_2$, thereby reducing the back-

ground radioactivity counts. Eventually, the precipitated carbonate is counted for its radioactive content.

***Organs*** A preclinical mass balance study employing a radioisotopically labeled drug is usually conducted when the time course, persistence, and potential accumulation of the drug and/or its metabolites in various parts of the animal body needs to be assessed [4].

After dissection, organs should be weighed and frozen as soon as possible. To obtain a pipettable homogeneous matrix for further analysis, homogenization may be required. If no other measurements than total radioactivity determination are intended, direct tissue solubilization may be considered as a faster and easier sample processing method (see below).

A pipettable organ homogenate may be obtained by homogenizing the organ with 3 parts (v/w) of phosphate-buffered saline pH 7.4, water, or 4% bovine serum albumin. The choice of solvent may be dependent on solubility, stability, and adsorption characteristics of the compound under investigation. In addition, the solvent used may interfere with subsequent analytical or pharmacodynamic assays. A fixed ratio of liquid to organ weight is preferable to a fixed volume per organ because the former will assure identical concentrations of organ constituents like proteins, which may be relevant to solubility, stability, and adsorption.

If organ radioactivity concentrations are to be assessed, it is imperative to sample and homogenize the whole organ, and not to sample parts (of different lobes) of the organ. For example, analyte concentrations are reported to vary significantly within a lobe (52-fold) or between lobes of a single liver (25-fold), likely due to differences in regional structure and blood and lymph flows [70]. On the other hand, it may be necessary to determine intratissue distribution of radioactivity. For example, distribution of radioactivity of an anticancer drug to the viable part of a tumor is relevant, but this information will be lost during homogenization [71]. Tissue autoradiography may be employed to identify such a phenomenon.

The organs of the GI tract can be homogenized after removal of the content or as a whole. By removing the content followed by separate analysis, any radioactivity can be more exactly ascribed to GI content (unabsorbed or excreted radioactivity) and organ radioactivity, respectively. This may be especially relevant with orally administered compounds.

***Carcass*** Oftentimes, the radioactivity in excreta does not account for the complete radioactive dose. A compound or its metabolites may exhibit a long half-life or may irreversibly bind to tissue components, or the radioisotope may get incorporated into endogenous compounds. Therefore, it may be required to assess the amount of residual radioactivity in the carcass. Individual organs are easily homogenized or digested and consequently analyzed for radioactivity. The carcass, however, is less easily processed. It contains remaining organs, muscle, connective tissue, and, most importantly, bone, skin, and fur. In order to achieve a homogenate with a high degree of homogeneity, the carcass (with an appropriate liquid phase) is first processed using a blender. Due to the high collagen content, a higher solvent volume to carcass weight ratio (from 4:1) may have to be used for homogenization to prevent coalescing of the homogenate. Next, a tissue solubilizer or a concentrated sodium hydroxide solution may be used to dissolve the tissues as much as possible (allowing it to stand

for several days or applying heating). The latter should be done in closed containers to avoid evaporation of solvents prior to taking an aliquot for scintillation counting.

***Digestion, Decoloration, and Oxidation of Samples*** A number of problems may be encountered in the processing of samples for radioactivity counting. Homogenization is not suitable for all organs (e.g., bone, skin) and is labor intensive, yet it is less destructive (e.g., determination of proteins, DNA, or individual compounds remains a possibility). Digestion of organs (or homogenate aliquots) with a tissue solubilizer is easy and fast, reduces the color of the sample, but is totally destructive. Lastly, oxidation of samples is destructive and labor intensive, but may be essential for processing nonhomogenizable matrices or determination of $^{3}H$ in the presence of $^{14}C$.

Digestion of organs or homogenate aliquots can be performed with so-called tissue solubilizers. Although destructive, this methodology can save a lot of time and effort because it may not be necessary to homogenize the sample beforehand. However, too much of these reagents may adversely affect the scintillation counting process, so that care must be taken to minimize the amount required. Manufacturers of these reagents can provide data on amounts required and the effects on the scintillation counting process.

Usually, the color interfering with scintillation counting is caused by iron-containing components of the sample like hemes. Treatment of the sample (homogenate or digested organ) with peroxide is the common method to reduce the color intensity. A suggested method would be taking a 100 μL aliquot of homogenate (3:1, v/w in water), adding 100 μL of 0.1 M EDTA (chelating metal ions, reducing excessive foaming of the hydrogen peroxide), followed by 200 μL of ~30% hydrogen peroxide. Incubation for 10 min at 50 °C will speed up the decoloration reaction. Next, 5–20 mL of liquid scintillation fluid may be added before counting. Variations of peroxides have been described, such as treatment with solubilizer followed by 0.4 mL of benzene saturated with benzoyl peroxide, or bleaching with glacial acetic acid and hydrogen peroxide [72, 73].

A sample oxidizer consists of a coiled platinum wire that can hold the sample. Additives may be applied to optimize the oxidation process. Subsequently, under a continuous oxygen flow, the sample is oxidized by heating the platinum wire using an electric current, and converting the carbon content to carbon dioxide and the hydrogen content to water, leaving nonvolatile components as ashes. The carbon dioxide containing the $^{14}CO_2$ and the water containing the $^{3}H_2O$ are trapped separately on adsorption columns and eluted into separate counting vials. Transfer efficiencies of over 95% are usually achieved with this technique. Especially in the case of double labeling of a compound with possible false-positive counts of $^{14}C$ in the $^{3}H$ counts, this technique is very suitable. Before bone is oxidized, it may be pulverized first to increase the surface area available for the oxidation process.

Irrespective of the eventual method used for sample processing, it is essential to determine the end efficiency of the procedure. Adding a known amount of radioactivity to a sample, processing it, and determining the liquid scintillation count easily achieves this. One should also prepare an identical sample without the organ component to establish the influence of the reagents used on the counting process. This may identify the need to use less of the reagents and instead increase the

duration of heating the sample. An overall efficiency/accuracy of at least 90% should be aimed for.

***Liquid Scintillation Counting*** For the determination of total radioactivity in samples (the composite of parent drug and metabolites still containing the radioactive label), liquid scintillation counting is often employed [12]. The sample is processed in such a way that it is in a liquid form. An aliquot is mixed with liquid scintillation fluid, a mixture of compounds that are excited by β-radiation and convert the absorbed energy to multiple light pulses. The sample is analyzed by a liquid scintillation counter that converts the light pulses into a number of observed counts per minute (CPM). Because not all the disintegrating nuclei result in a count, the CPM need to be converted to the real number of disintegrations per minute (DPM). The ratio of CPM to DPM is the counting efficiency, which is higher for isotopes with more energetic radiation ($^{32}P > {}^{14}C > {}^{3}H$). Most liquid scintillation counters will be able to determine the counting efficiency of a sample and thus correct the measured count to the actual count. This correction is usually performed by either of two ways. The first is the external standard calibration. A radioactive (γ-) source is placed next to the sample, causing a known amount of additional excitation in the sample. The ratio of the actual counts to the expected counts is a measure of the counting efficiency. The second method is the internal standard calibration. With decreasing efficiency, the signal intensity versus energy distribution shifts to lower energies. Using a prerecorded quench curve, the counting efficiency can be determined from this shift. To maximize the counting efficiency, a sufficiently high scintillation fluid-to-sample ratio must be used (ideally more than 5:1). In addition, the mixture of sample and scintillation fluid must be one homogeneous phase. There are a variety of liquid scintillation fluids available that are suitable for organic, aqueous, and high salt content samples, respectively. Sometimes, especially when the sample is colored, counting efficiency drops, and the correction becomes flawed. This color quenching occurs, in particular, with red samples (blood, liver, kidney, etc.). A suitable sample preparation step involving digestion and decoloration may resolve this problem. Heating of the sample or exposure to sunlight may excite the scintillants (luminescence) and result in an erroneously high count. This can be avoided by letting the samples stand overnight in the dark before counting. The error in the radioactive count is inversely related to the root of the count and the counting time. Increasing the counting time will therefore result in more precise measurements, especially with lower concentrations of radioactivity. In general, a counting time of 5–10 min will suffice. When a double label is used (e.g., labeling a compound simultaneously with both $^{14}C$ and $^{3}H$), the signal of one isotope can cross talk into the signal of the other isotope. The $^{3}H$ signal will also pick up part of the $^{14}C$ signal. Adjusting the settings of the liquid scintillation counter and using corrections may yield the separate counts for both isotopes.

### 4.3 Metabolic Profiling

The overview of the metabolism of a drug is often called the *metabolic profile* [25]. Recently, the FDA posted a draft guidance describing the need for safety testing of metabolites representing more than 10% of systemic exposure human [5]. Thus, early preclinical identification of metabolites of potential relevance to humans is

key. A mass balance study is well suited to elucidate the metabolic profile of a compound because it employs a radioisotopically labeled compound, enabling detection of the parent drug and its metabolites even if the latter are structurally still unknown. After structural elucidation, metabolites may be tested for their pharmacological and toxicological properties. This is important in drug development because the concentrations of the parent drug and/or its active metabolites may provide an important link between drug exposure and desirable and/or undesirable drug effects. For this reason, the development of sensitive and specific assays for a drug and its key metabolites is critical to the study of metabolism and drug–drug interactions [9]. Next, a discussion is provided of the various matrices, sample pretreatment, chromatographic separation, and detection.

***Matrix Selection*** Ideally, metabolites are identified and quantified in plasma, urine, and feces. A matrix is especially interesting when a large proportion of the radioactivity is not explained by unchanged drug (this proportion consists of metabolites). This requires knowledge of the level of the total radioactivity and the unchanged drug in the matrix. Because the metabolites excreted in one matrix may be qualitatively different from those in another matrix, preferably, metabolic profiling is executed in all of these matrices. In addition, metabolites in a target organ (for activity or toxicity) may be relevant. However, quantification of metabolites can be complicated. Plasma may not be available in sufficient amounts to allow analysis of metabolites after analysis of total radioactivity and unchanged drug, especially when it concerns a small animal. Pooling of samples from individual animals may then be considered. It can be difficult to isolate compounds from feces, impeding metabolic profiling. Urine is a concentrated matrix that is suitable and most commonly used for metabolic profiling.

***Sample Pretreatment*** Conventionally, sample pretreatment is aimed at selective extraction and concentration of one or more specific analytes. However, for metabolic profiling one needs to concentrate all metabolites with their heterogeneous physicochemical characteristics, while reducing matrix-derived interference. Consequently, sample pretreatment for metabolic profiling is usually nonspecific and consists of (1) addition of a water-miscible organic solvent, (2) centrifugation to precipitate salts, proteins, and insoluble matter, (3) concentration or evaporation of the supernatant, and (4) reconstitution in a suitable solvent. The efficiency of each of these steps can be monitored and optimized using radioactivity determinations [74]. Radioactivity lost during evaporation may be explained by volatile metabolites or tritiated water.

***Chromatographic Separation*** The high resolution of high performance liquid chromatography (HPLC) makes it the method of choice over other chromatographic methods such as thin layer chromatography. By applying a shallow gradient elution (e.g., C18 column, 5–90% ACN linear, in 60 min), metabolites of a large polarity range can be separated. The use of a guard column can protect the column from high concentrations of matrix-derived constituents [12]. When subsequent mass spectrometric analysis is considered, mobile phase components should be chosen that are compatible with mass spectrometry (i.e., they should allow adequate

ionization of the compounds and must be volatile). In reverse phase chromatography, the parent compound (often less polar than its metabolites) elutes last. The gradient can be designed so that the parent compound elutes at the end of the chromatographic run. An initial run time of 60 min may be necessary to separate all the metabolites, but this may be shortened when the elution times of the metabolites are known and the assay is used more routinely [25].

***Detection Techniques*** The compounds eluting from the HPLC system may be detected with various techniques, used in parallel or in series. Parallel detection techniques involve splitting the column effluent, thereby reducing sensitivity. However, some techniques are destructive in nature and thus cannot be employed in front of another technique. Because mass balance studies employ radioisotopically labeled compounds, radioactivity detection is the primary technique, unambiguously indicating the presence of metabolites at specific retention times. Ultraviolet absorption, preferably using a photodiode array, can be used to indicate changes in the chromophore. Mass spectrometry can be used to acquire additional structural information or to confirm the presence of a suspected metabolite. Finally, more complicated techniques like nuclear magnetic resonance (NMR) may be utilized, especially when investigating fluorinated compounds. Sometimes the radioactive samples may need to be diluted down to an acceptable level of radioactivity to permit their handling in different laboratory areas.

*Radioactivity Detection* Radioactivity detection may be performed either online or offline. Online radioactivity detection is rapid but destructive. It consists of admixing a scintillation fluid with the column effluent (usually in a 3:1, v/v ratio), passing it through a detection cell (a coiled, transparent tube), and recording the counts in time. The mixing of the column effluent with the scintillation fluid excludes any further analytical characterization of the effluent. Furthermore, the online radioactivity counting decreases sensitivity. When employing a flow rate of 1 mL/min, a 3:1 (v/v) ratio of scintillation fluid and a 500 μL detection cell, the residence time of any component in the detection cell is only 7.5 seconds. The offline detection method is more laborious, consisting of fractionation of the column effluent (every 15–60 s). Fractions can be counted for radioactivity, for example, for 5 min each (increasing sensitivity 40-fold relative to offline detection). Offline detection can be performed nondestructively by taking only an aliquot of each fraction for scintillation counting, reserving the residual sample of fractions with high levels of radioactivity for subsequent structural identification. A drawback of the offline radioactivity detection is its labor-intensive nature, precluding rapid sequential sample analysis.

When a more energetic radioisotope is being used (e.g., $^{32}P$), it may be possible to detect the radioactivity without using scintillation fluid, by using Çerenkov detection. This allows collection of unadulterated effluent after radioactivity detection [12].

*Photodiode Array Detection* Absorption measurement using a photodiode array is a nondestructive detection technique. However, it is nonselective, necessitating comparison of the sample chromatogram with a blank. A shift in the spectrum

relative to that of the parent compound indicates a metabolic modification of the chromophore. When applying a gradient elution, the change in mobile phase composition may also cause a shift of the spectrum.

*Mass Spectrometry* More qualitative information may be derived from mass spectrometric analysis of selected fractions after offline profiling or after splitting the column effluent. Changes in the mass spectrum relative to the spectrum of the unchanged compound provide structural information. For example, a mass increase of +16 atomic mass units indicates hydroxylation, +42 atomic mass units indicates acetylation, –14 atomic mass units indicates demethylation, and so on. After tentative structural elucidation of a metabolite, the retention time and mass spectrum may be compared to that of a synthetic standard.

*Nuclear Magnetic Resonance* If there is a fluorine atom present in the compound, $^{19}$F-NMR may be applied to detect, quantitate, and characterize parent compound and metabolites. A large signal for the fluorine ion indicates that the compound is extensively dehalogenated, and those metabolites are thereafter undetectable by $^{19}$F-NMR [18]. Although at present quantitation with $^{19}$F-NMR is hampered by matrix-related effects [75], in the future, more powerful NMR equipment is expected to improve quantitative application, even as to enable *in vivo* monitoring of drug and metabolite distribution [76].

***Additional Characterization*** Besides chromatographic separation of a sample and detection of its metabolic constituents, there is more one can do with the samples. Collection of a particular fraction and repeated chromatographic separation may lead to isolation of a metabolite with sufficient purity to allow determination of the full mass spectrum, absorbance spectrum [77], NMR spectrum, and so on.

In addition, incubation with β-glucuronidase and sulfatase can indicate the presence of glucuronide and sulfate conjugates, respectively, of the parent compound or its phase I metabolites. A decrease in peak area indicates the presence of the conjugate, while an increase indicates the corresponding aglycone counterpart [78].

Finally, metabolites obtained from *in vitro* incubations of the compound of interest, that subsequently have been elucidated structurally, should be used as reference compounds.

## 5 RESULTS

There are a number of useful parameters that may be determined from a mass balance:

- Mass balance calculation
  - Recovery in urine, feces, air, carcass, and total recovery
  - Major excretory pathway
  - Ratio of unchanged drug/radioactivity

- Plasma pharmacokinetics (unchanged drug and radioactivity)
  - Half-life
  - AUC and the unchanged/radioactivity ratio
  - $C_{max}$, $T_{max}$
- Assessment of absorption
  - Unchanged drug $AUC_{PO}/AUC_{IV}$, bioavailability
  - Total radioactivity $AUC_{PO}/AUC_{IV}$, indication of fraction absorbed
- Organ distribution
  - Percentage of radioactivity in organs in time/at end of study
  - Drug and metabolite levels in target organs
  - Organ to plasma radioactivity levels in time
  - Organ AUC to plasma AUC ratios of radioactivity
  - Organ radioactivity ratios of genetically modified to wild-type strain
- Metabolic profiling
  - Number of metabolites (especially those accounting for >10% of radioactivity)
  - Structural information of previously unknown metabolites
  - Pharmacokinetic parameters of metabolites
  - Obtaining reference compounds
  - Concordance in metabolic profile between species

### 5.1 Mass Balance Calculation

The mass balance calculation consists of calculating the amount of radioactivity administered and calculating the amount of radioactivity recovered in urine, feces, and, if applicable, expired air:

$$\text{Recovery} = \frac{\begin{array}{ll} \text{Urine} & \text{volume (mL)} \cdot \text{concentration (DPM/mL)} \\ \text{Feces} & \text{weight (g)} \cdot \text{homogenization factor} \cdot \text{concentration (DPM/mL)} \\ \text{Organs} & \text{weight (g)} \cdot \text{homogenization factor} \cdot \text{concentration (DPM/mL)} \\ \text{Expired} & \text{DPM collected as } ^{14}\text{C-carbonate} \quad + \\ \hline \text{Total} & \end{array}}{\begin{array}{ll} & \text{divided by} \\ \text{Dosed} & \text{Volume dosed (mL)} \cdot \text{concentration (DPM/mL)} \end{array}} \cdot 100\%$$

The total expired radioactivity can be calculated by integrating the $^{14}$C-expiration rate versus time curve or collecting all the expired air. The relative contributions of feces and urine to the recovery of the dose determine the major excretory pathway.

After stopping collection of the excreta, the animals may be sacrificed, and the radioactivity still deposited in the carcass may (after suitable processing) be determined to increase the percentage of radioactive dose accounted for.

The fraction of radioactivity in urine and/or feces that is represented by unchanged drug is an indication of the importance of metabolism relative to unchanged excretion. However, this requires the availability of an analytical method suited to quantitate unchanged drug in excreta.

### 5.2 Plasma Pharmacokinetics

Although the validity of calculating (pseudo) pharmacokinetic parameters based on total or undiscriminated radioactivity levels has been debated, they may have some, if limited, use [79, 80]. Capturing as much of the area under the plasma radioactivity versus time curve increases the usefulness of pseudopharmacokinetic parameters. If a good fit of the terminal elimination phase is achieved and more than 90% of the $AUC_{0\text{-inf}}$ is captured in the $AUC_{0\text{-last}}$, the exposure to the total of the drug and its products is adequately described, even if theoretically the curve consists of multiple monoexponential functions.

Calculation of a half-life for total radioactivity may be useful when comparing such a value with the half-life of the unchanged drug. A much longer half-life for total radioactivity indicates a potential for underestimating the risk of accumulation of metabolites when solely focusing on parent drug. In addition, a pseudo-half-life for radioactivity may aid in determining the usefulness of continued collection of excreta to maximize recovery in the mass balance study.

The ratio of the AUC of unchanged drug over the AUC of total radioactivity is a measure of the importance of metabolism in the elimination of the drug. If this ratio in plasma is different from the ratio of unchanged drug to total radioactivity as determined in the excreta, this may indicate selective excretion or reabsorption of unchanged drug or metabolites, and degradation or metabolism of the drug after glomerular filtration or biliary excretion, respectively.

Finally, $C_{max}$ and $T_{max}$ should be reported, although they will generally coincide with those of the unchanged drug.

### 5.3 Assessment of Absorption

In the case of orally administered drugs that have an incomplete bioavailability, this may be due to incomplete absorption and/or first-pass metabolism by the intestine and the liver. The use of a radioactively labeled drug can provide a measure of the extent of absorption relative to first-pass. The (dose-corrected) ratio of the oral AUC versus the intravenous AUC of the unchanged drug in plasma yields the overall bioavailability. The ratio of the oral AUC versus the intravenous AUC of the total radioactivity in plasma (or the urinary recovery of radioactivity after oral versus intravenous administration) is an indication of the fraction absorbed [20, 81]. Although sometimes these ratios are used quantitatively, this is not correct. The radioactivity in plasma (or urine) after oral administration may consist of different metabolites compared to the radioactivity after intravenous administration. Therefore, distribution characteristics and excretion routes of the total radioactivity may differ between the administration routes and cannot be directly compared. Proper assessment of the contributions of gut and liver to bioavailability can be assessed by measuring plasma AUC of unchanged drug after oral, intraportal, and intravenous administration, respectively [82].

### 5.4 Organ Distribution

Documenting the organ distribution of radioactivity (and unchanged drug) can be informative.

After stopping the collection of excreta, radioactivity in the organs can complete the mass balance to 100%. Any radioactivity still unaccounted for is likely to have been lost through expiration ($^{14}CO_2$ or $^{3}H_2O$).

Following the radioactivity in organs during the study indicates how quickly drug-related material is taken up by an organ (e.g., target organs where it may exert its intended or side effect) and how quickly it is redistributed and excreted again. For example, the organ distribution of lovastatin radioactivity indicated preferential accumulation of radioactivity in liver, the target organ. Further speciation showed that the active hydroxy-acid metabolite constituted the majority of liver radioactivity [22].

Organ/plasma ratios of radioactivity may be a resultant of physicochemical characteristics (lipophilic drugs will accumulate in fat), the activity of drug transporters, and the retention by irreversible binding or uptake of the radioactive label into endogenous substances. Ratios of organ AUC to plasma AUC may be calculated as a measure of relative exposure. These data may also be used to extrapolate radioactivity dosimetry data for the human mass balance [83, 84]. Retention of radioactivity in organs may reflect tissue binding of metabolites, but it may also be caused by a delayed elimination of metabolites. Upon repeated exposure, this may result in accumulation of metabolites to possibly toxic levels.

Drug concentrations are not routinely measured in red blood cells. The ratio of radioactivity in red blood cells relative to plasma (if the ratio deviates from 1.0) may indicate (active) uptake of drug into red blood cells. Knowledge of drug partitioning into red blood cells is useful, so that drug clearances from blood by organs of elimination may be evaluated relative to organ blood flows [4].

Comparison of organ distribution data between wild-type animals and animals defective in a drug transporter can be informative to predict the effects of drugs when they are combined with known inhibitors of these transporters [52, 53].

### 5.5 Metabolic Profiling

Determining the metabolic profile is important given the frequent occurrence of clinically relevant active metabolites [85]. It also enables identification of interspecies differences in metabolic profiles [46]. Reference compounds of possible metabolites can be obtained from prior *in vitro* studies. The *in vivo* study can then confirm the presence of metabolites observed *in vitro*. Consequently, metabolites observed in the preclinical mass balance study are likely to be seen in the human mass balance study, and, therefore, metabolite reference compounds may be isolated from the urine or feces obtained in the preclinical mass balance study.

Metabolites can be quantitated even before the structure has been elucidated. The radioactivity-peak areas are simply converted to molar amounts using the specific radioactivity of the administered compound. If such quantification is possible with adequate precision and accuracy, even pharmacokinetic parameters of these metabolites can be calculated, as shown with formoterol in humans [86].

The presence of metabolites in urine while absent in plasma or the absence of parent drug in urine while present in plasma may indicate renal metabolism or reabsorption processes [69].

Human metabolites that constitute a quantitatively large part of total radioactivity in urine, feces, or plasma, and metabolites that (based on their chemical structure) are likely to have a pharmacological or toxic effect, should be investigated for safety testing [5]. Based on metabolites observed in the preclinical studies, such investigations may be initiated earlier.

## REFERENCES


1. The European Agency for the Evaluation of Medicinal Products. *Note for Guidance on Toxicokinetics: A Guidance for Assessing Systemic Exposure in Toxicology Studies*. The European Agency for the Evaluation of Medicinal Products; 1995. Accessed 3 May 2004.
2. The European Agency for the Evaluation of Medicinal Products. *Note for Guidance on Non-clinical Safety Studies for the Conduct of Human Clinical Trials for Pharmaceuticals*. The European Agency for the Evaluation of Medicinal Products; 2000. Accessed 3 May 2004.
3. Center for Drugs and Biologics, Food and Drug Administration, Department of Health and Human Services. *Guideline for the Format and Content of the Nonclinical Pharmacology/Toxicology Section of an Application*. Center for Drug Evaluation and Research; 1997. Accessed 3 May 2004.
4. Peck CC, Barr WH, Benet LZ, Collins J, Desjardins RE, Furst DE, Harter JG, Levy G, Ludden T, Rodman JH. Opportunities for integration of pharmacokinetics, pharmacodynamics, and toxicokinetics in rational drug development. *Clin Pharmacol Ther* 1992;51(4): 465–473.
5. US Department of Health and Human Services, Food and Drug Administration. *Guidance for Industry—Safety Testing of Drug Metabolites* (draft guidance). Center for Drug Evaluation and Research (CDER): 2005. www.fda.gov/cder/guidance/index.htm. Accessed 16 September 2005.
6. Balani SK, Miwa GT, Gan LS, Wu JT, Lee FW. Strategy of utilizing *in vitro* and *in vivo* ADME tools for lead optimization and drug candidate selection. *Curr Top Med Chem* 2005;5(11):1033–1038.
7. Department of Health and Human Services. US Food and Drug Administration-Center for Drug Evaluation and Research Home Page. http://www.fda.gov/cder/index.html. Accessed 7 January 2006.
8. EMEA. European Medicines Agency–Human Medicines Home Page. http://www.emea.eu.int/index/indexh1.htm. Accessed 8 January 2006.
9. US Department of Health and Human Services, Food and Drug Administration. *Guidance for Industry—In Vivo Drug Metabolism/Drug Interaction Studies—Study Design, Data Analysis, and Recommendations for Dosing and Labeling*. Center for Drug Evaluation and Research (CDER), Center for Biologics Evaluation and Research (CBER); 1999. www.fda.gov/cder/guidance/index.htm. Accessed 16 September 2005.
10. US Department of Health and Human Services, Food and Drug Administration. *Guidance for Industry—Estimating the Maximum Safe Starting Dose in Initial Clinical Trials for Therapeutics in Adult Healthy Volunteers*. Center for Drug Evaluation and Research (CDER); 2005. www.fda.gov/cder/guidance/index.htm. Accessed 9 February 2006.

11. Dalvie D. Recent advances in the applications of radioisotopes in drug metabolism, toxicology and pharmacokinetics. *Curr Pharm Des* 2000;6(10):1009–1028.
12. *Principles and Methods of Toxicology*. Philadelphia: Taylor&Francis; 2001.
13. Fagerholm U, Breuer O, Swedmark S, Hoogstraate J. Pre-clinical pharmacokinetics of the cyclooxygenase-inhibiting nitric oxide donor (CINOD) AZD3582. *J Pharm Pharmacol* 2005;57(5):587–597.
14. Budavari S. *The Merck Index*. Whitehouse Station, NJ: Merck&Co; 1996.
15. Hutchinson OC, Collingridge DR, Barthel H, Price PM, Aboagye EO. Pharmacokinetics of radiolabelled anticancer drugs for positron emission tomography. *Curr Pharm Des* 2003;9(11):917–929.
16. Minami T, Ichii M, Okazaki Y. Comparison of three different methods for measurement of tissue platinum level. *Biol Trace Element Res* 1995;48(1):37–44.
17. Hughes MF, Devesa V, Adair BM, Styblo M, Kenyon EM, Thomas DJ. Tissue dosimetry, metabolism and excretion of pentavalent and trivalent monomethylated arsenic in mice after oral administration. *Toxicol Appl Pharmacol* 2005;208(2):186–197.
18. Desmoulin F, Gilard V, Malet-Martino M, Martino R. Metabolism of capecitabine, an oral fluorouracil prodrug: (19)F NMR studies in animal models and human urine. *Drug Metab Dispos* 2002;30(11):1221–1229.
19. Ferraiolo BL, Moore JA, Crase D, Gribling P, Wilking H, Baughman RA. Pharmacokinetics and tissue distribution of recombinant human tumor necrosis factor-alpha in mice. *Drug Metab Dispos* 1988;16(2):270–275.
20. Pleiss U. [Radioactive-labeled pharmaceutical agents.] Werkzeuge bei der Arzneimittelentwicklung. Radioaktiv markierte pharmazeutische Wirkstoffe. *Pharmazie Zeit* 2005; 34(6):514–519.
21. Dueker SR, Jones AD, Clifford AJ. Protocol development for biological tracer studies. *Adv Exp Med Biol* 1998;445:363–378.
22. Duggan DE, Chen IW, Bayne WF, Halpin RA, Duncan CA, Schwartz MS, Stubbs RJ, Vickers S. The physiological disposition of lovastatin. *Drug Metab Dispos* 1989;17(2): 166–173.
23. Horsmans Y, Saliez A, van d BV, Desager JP, Geubel AP, Pauwels S, Lambotte L. $^{14}$C-Propoxyphene demethylation in the rat. An example of differences between liver and intestinal drug-presystemic metabolism. *Drug Metab Dispos* 1997;25(11):1257–1259.
24. Dalvie D. Recent advances in the applications of radioisotopes in drug metabolism, toxicology and pharmacokinetics. *Curr Pharm Des* 2000;6(10):1009–1028.
25. Beumer JH, Beijnen JH, Schellens JH. Mass balance studies, with a focus on anticancer drugs. *Clin Pharmacokinet* 2006;45(1):33–58.
26. Browne TR, Van Langenhove A, Costello CE, Biemann K, Greenblatt DJ. Pharmacokinetic equivalence of stable-isotope-labeled and unlabeled drugs. Phenobarbital in man. *J Clin Pharmacol* 1982;22(7):309–315.
27. Phillips DH, Potter GA, Horton MN, Hewer A, Crofton-Sleigh C, Jarman M, Venitt S. Reduced genotoxicity of [D5-ethyl]-tamoxifen implicates alpha-hydroxylation of the ethyl group as a major pathway of tamoxifen activation to a liver carcinogen. *Carcinogenesis* 1994;15(8):1487–1492.
28. Toxicokinetics: the assessment of systemic exposure in toxicity studies-ICH topic S3A. ICH; 1995. *CPMP/ICH/385/95*. Accessed 9 February 2006.
29. US Department of Health and Human Services, Food and Drug Administration. *Guidance for Industry—M3 Nonclinical Safety Studies for the Conduct of Human Clinical Trials for Pharmaceuticals*. Center for Drug Evaluation and Research (CDER); 1997. www.fda.gov/cder/guidance/index.htm. Accessed 16 September 2005.

30. US Department of Health and Human Services, Food and Drug Administration. *Guidance for Industry—Drug Metabolism/Drug Interaction Studies in the Drug Development Process: Studies In Vitro*. Center for Drug Evaluation and Research, Center for Biologics Evaluation and Research; 1997. www.fda.gov/cder/guidance.htm. Accessed 16 September 2005.
31. Collins JM. Inter-species differences in drug properties. *Chem-Biol Interact* 2001;134(3): 237–242.
32. Camiener GW, Smith CG. Studies of the enzymatic deamination of cytosine arabinoside. I. Enzyme distribution and species specificity. *Biochem Pharmacol* 1965;14(10):1405–1416.
33. Sharkey EM, ONeill HB, Kavarana MJ, Wang H, Creighton DJ, Sentz DL, Eiseman JL. Pharmacokinetics and antitumor properties in tumor-bearing mice of an enediol analogue inhibitor of glyoxalase I. *Cancer Chemother Pharmacol* 2000;46(2):156–166.
34. Morton CL, Iacono L, Hyatt JL, Taylor KR, Cheshire PJ, Houghton PJ, Danks MK, Stewart CF, Potter PM. Activation and antitumor activity of CPT-11 in plasma esterase-deficient mice. *Cancer Chemother Pharmacol* 2005;56(6):629–636.
35. van Bezooijen CF. Influence of age-related changes in rodent liver morphology and physiology on drug metabolism—a review. *Mech Ageing Dev* 1984;25(1–2):1–22.
36. Leakey JE, Cunny HC, Bazare J Jr, Webb PJ, Feuers RJ, Duffy PH, Hart RW. Effects of aging and caloric restriction on hepatic drug metabolizing enzymes in the Fischer 344 rat. I: The cytochrome P-450 dependent monooxygenase system. *Mech Ageing Dev* 1989; 48(2):145–155.
37. Leakey JA, Cunny HC, Bazare J Jr, Webb PJ, Lipscomb JC, Slikker W Jr, Feuers RJ, Duffy PH, Hart RW. Effects of aging and caloric restriction on hepatic drug metabolizing enzymes in the Fischer 344 rat. II: Effects on conjugating enzymes. *Mech Ageing Dev* 1989;48(2):157–166.
38. Czerniak R. Gender-based differences in pharmacokinetics in laboratory animal models. *Int J Toxicol* 2001;20(3):161–163.
39. Witkamp RF, Yun HI, van't Klooster GA, van Mosel JF, van Mosel M, Ensink JM, Noordhoek J, Van Miert AS. Comparative aspects and sex differentiation of plasma sulfamethazine elimination and metabolite formation in rats, rabbits, dwarf goats, and cattle. *Am J Vet Res* 1992;53(10):1830–1835.
40. Witkamp RF, Lohuis JA, Nijmeijer SM, Kolker HJ, Noordhoek J, van Miert AS. Species- and sex-related differences in the plasma clearance and metabolite formation of antipyrine. A comparative study in four animal species: cattle, goat, rat and rabbit. *Xenobiotica* 1991;21(11):1483–1492.
41. Griffin RJ, Godfrey VB, Kim YC, Burka LT. Sex-dependent differences in the disposition of 2,4-dichlorophenoxyacetic acid in Sprague–Dawley rats, B6C3F1 mice, and Syrian hamsters. *Drug Metab Dispos* 1997;25(9):1065–1071.
42. Kato R, Yamazoe Y. Sex-specific cytochrome P450 as a cause of sex- and species-related differences in drug toxicity. *Toxicol Lett* 1992;64–65(Spec No):661–667.
43. Klecker RW, Cysyk RL, Collins JM. Zebularine metabolism by aldehyde oxidase in hepatic cytosol from humans, monkeys, dogs, rats, and mice: influence of sex and inhibitors. *Bioorg Med Chem* 2006;14(1);62–66.
44. Schwartz JB. The influence of sex on pharmacokinetics. *Clin Pharmacokinet* 2003;42(2): 107–121.
45. van Herwaarden AE, Smit JW, Sparidans RW, Wagenaar E, van der Kruijssen CM, Schellens JH, Beijnen JH, Schinkel AH. Midazolam and cyclosporin a metabolism in

transgenic mice with liver-specific expression of human CYP3A4. *Drug Metab Dispos* 2005;33(7):892–895.

46. Marathe PH, Shyu WC, Humphreys WG. The use of radiolabeled compounds for ADME studies in discovery and exploratory development. *Curr Pharm Des* 2004;10(24):2991–3008.
47. Jonker JW, Wagenaar E, van Deemter L, Gottschlich R, Bender HM, Dasenbrock J, Schinkel AH. Role of blood–brain barrier P-glycoprotein in limiting brain accumulation and sedative side-effects of asimadoline, a peripherally acting analgaesic drug. *Br J Pharmacol* 1999;127(1):43–50.
48. Merino G, Jonker JW, Wagenaar E, van Herwaarden AE, Schinkel AH. The breast cancer resistance protein (BCRP/ABCG2) affects pharmacokinetics, hepatobiliary excretion, and milk secretion of the antibiotic nitrofurantoin. *Mol Pharmacol* 2005;67(5):1758–1764.
49. Jonker JW, Wagenaar E, Mol CA, Buitelaar M, Koepsell H, Smit JW, Schinkel AH. Reduced hepatic uptake and intestinal excretion of organic cations in mice with a targeted disruption of the organic cation transporter 1 (Oct1 [Slc22a1]) gene. *Mol Cell Biol* 2001;21(16):5471–5477.
50. Breedveld P, Pluim D, Cipriani G, Wielinga P, van Tellingen O, Schinkel AH, Schellens JH. The effect of Bcrp1 (Abcg2) on the *in vivo* pharmacokinetics and brain penetration of imatinib mesylate (Gleevec): implications for the use of breast cancer resistance protein and P-glycoprotein inhibitors to enable the brain penetration of imatinib in patients. *Cancer Res* 2005;65(7):2577–2582.
51. Schinkel AH, Wagenaar E, Mol CA, van Deemter L. P-glycoprotein in the blood–brain barrier of mice influences the brain penetration and pharmacological activity of many drugs. *J Clin Invest* 1996;97(11):2517–2524.
52. Schinkel AH, Wagenaar E, van Deemter L, Mol CA, Borst P. Absence of the mdr1a P-glycoprotein in mice affects tissue distribution and pharmacokinetics of dexamethasone, digoxin, and cyclosporin A. *J Clin Invest* 1995;96(4):1698–1705.
53. Schinkel AH, Smit JJ, van Tellingen O, Beijnen JH, Wagenaar E, Van DeemTer L, Mol CA, van der Valk MA, Robanus-Maandag EC, te Riele HP. Disruption of the mouse mdr1a P-glycoprotein gene leads to a deficiency in the blood–brain barrier and to increased sensitivity to drugs. *Cell* 1994;77(4):491–502.
54. Gonzalez FJ, Yu AM. Cytochrome P450 and xenobiotic receptor humanized mice. *Annu Rev Pharmacol Toxicol* 2006;46:41–64.
55. Gonzalez FJ. Cytochrome P450 humanised mice. *Hum Genomics* 2004;1(4):300–306.
56. Beumer JH, Rademaker-Lakhai JM, Rosing H, Lopez-Lazaro L, Beijnen JH, Schellens JH. Trabectedin (Yondelis, formerly ET-743), a mass balance study in patients with advanced cancer. *Invest New Drugs* 2005;23(5):429–436.
57. Rickert DE, Dingley K, Ubick E, Dix KJ, Molina L. Determination of the tissue distribution and excretion by accelerator mass spectrometry of the nonadecapeptide $^{14}$C-Moli1901 in beagle dogs after intratracheal instillation. *Chem-Biol Interact* 2005;155(1–2):55–61.
58. Lappin G, Garner RC. Big physics, small doses: the use of AMS and PET in human microdosing of development drugs. *Nat Rev Drug Discov* 2003;2(3):233–240.
59. *Pharmacokinetics: Guidance for Repeated Dose Tissue Distribution Studies—ICH Topic S3B*. ICH; 2005. *CPMP/ICH/385/95*.
60. Dent NJ. The inspection of drug metabolism and pharmacokinetic studies. *Qual Assur* 1992;1(3):230–236.

61. *Safe and Secure—A Guide to Working Safely with Radiolabelled Compounds*. Amersham; 2005. http://www1.amershambiosciences.com/applic/upp00738.nsf/vLookupDoc/196245642-A125/$file/Safe_Secure%202003.pdf. Accessed 2 November 2005.

62. Kishore R, Eary JF, Krohn KA, Nelp WB, Menard TW, Beaumier PL, Hellstrom KE, Hellstrom I. Autoradiolysis of iodinated monoclonal antibody preparations. *Int J Radiat Appl Instrum B* 1986;13(4):457–459.

63. Kinuya S, Yokoyama K, Tega H, Hiramatsu T, Konishi S, Yamamoto W, Shuke N, Aburano T, Watanabe N, Takayama T, Michigishi T, Tonami N. Rhenium-186-mercaptoacetyltriglycine-labeled monoclonal antibody for radioimmunotherapy: *in vitro* assessment, *in vivo* kinetics and dosimetry in tumor-bearing nude mice. *Japanese J Cancer Res* 1998;89(8): 870–878.

64. Smit MJ, Temmerman AM, Havinga R, Kuipers F, Vonk RJ. Short- and long-term effects of biliary drainage on hepatic cholesterol metabolism in the rat. *Biochem J* 1990;269(3): 781–788.

65. Accatino L, Hono J, Koenig C, Pizarro M, Rodriguez L. Adaptive changes of hepatic bile salt transport in a model of reversible interruption of the enterohepatic circulation in the rat. *J Hepatol* 1993;19(1):95–104.

66. Nezasa K, Takao A, Kimura K, Takaichi M, Inazawa K, Koike M. Pharmacokinetics and disposition of rosuvastatin, a new 3-hydroxy-3-methylglutaryl coenzyme A reductase inhibitor, in rat. *Xenobiotica* 2002;32(8):715–727.

67. Rowland M, Tozer TN, *Clinical Pharmacokinetics—Concepts and Applications*. Baltimore: Williams & Wilkins; 1995.

68. Mukherjee KL, Heidelberger C. Studies on fluorinated pyrimidines. IX. The degradation of 5-fluorouracil-6-C14. *J Biol Chem* 1960;235:433–437.

69. Melgar MD, Zuleski FR, Malbica JO. Metabolism, disposition, and pharmacokinetics of tracazolate in rat and dog. *Drug Metab Dispos* 1984;12(4):396–402.

70. Lee HJ, Chiou WL. Marked heterogeneity in the intrahepatic distribution of quinidine in rats: implications in pharmacokinetics. *J Pharm Sci* 1990;79(9):778–781.

71. Ings RM, Breen M, Devereux K, Gray AJ, Edwards FE, Lucas C, Briggs M, Robinson BV, Campbell DB. The comparative disposition of [$^{14}$C]-fotemustine in non-tumourous and tumourous mice. *Cancer Chemother Pharmacol* 1990;27(2):106–110.

72. Tomiyama Y, Brian JE Jr, Todd MM. Cerebral blood flow during hemodilution and hypoxia in rats: role of ATP-sensitive potassium channels. *Stroke* 1999;30(9):1942–1947.

73. Berger H, Sandow J, Heinrich N, Albrecht E, Kertscher U, Oehlke J. Disposition of the $^{3}$H-labeled gonadotropin-releasing hormone analog buserelin in rats. *Drug Metab Dispos* 1993;21(5):818–822.

74. Slatter JG, Schaaf LJ, Sams JP, Feenstra KL, Johnson MG, Bombardt PA, Cathcart KS, Verburg MT, Pearson LK, Compton LD, Miller LL, Baker DS, Pesheck CV, Lord RS III. Pharmacokinetics, metabolism, and excretion of irinotecan (CPT-11) following I.V. infusion of [($^{14}$)C]CPT-11 in cancer patients. *Drug Metab Dispos* 2000;28(4):423–433.

75. Lenz EM, Wilson ID, Wright B, Partridge EA, Rodgers CT, Haycock PR, Lindon JC, Nicholson JK. A comparison of quantitative NMR and radiolabelling studies of the metabolism and excretion of Statil (3-(4-bromo-2-fluorobenzyl)-4-oxo-3*H*-phthalazin-1-ylacetic acid) in the rat. *J Pharm Biomed Anal* 2002;28(1):31–43.

76. Griffiths JR, Glickson JD. Monitoring pharmacokinetics of anticancer drugs: non-invasive investigation using magnetic resonance spectroscopy. *Adv Drug Deliv Rev* 2000;41(1): 75–89.

77. Sparreboom A, Huizing MT, Boesen JJ, Nooijen WJ, van Tellingen O, Beijnen JH. Isolation, purification, and biological activity of mono- and dihydroxylated paclitaxel metabolites from human feces. *Cancer Chemother Pharmacol* 1995;36(4):299–304.
78. Gilbert PJ, Hartley TE, Troke JA, Turcan RG, Vose CW, Watson KV. Application of $^{19}$F-n.m.r. spectroscopy to the identification of dog urinary metabolites of imirestat, a spirohydantoin aldose reductase inhibitor. *Xenobiotica* 1992;22(7):775–787.
79. Jansen AB. Total radioactivity half-lives. *Drug Metab Dispos* 1979;7(5):350.
80. Di Carlo FJ. Undifferentiated radioactivity revisited. *Drug Metab Dispos* 1980;8(4): 287–288.
81. Lemaire M, Bodoky A, Heberer M, Niederberger W. Quantification of absorption, intestinal and hepatic first pass effect in a chronic dog model. *Eur J Drug Metab Pharmacokinet* 1991;Spec No 3:132–135.
82. Vachharajani NN, Shyu WC, Shah VR, Barbhaiya RH. Pharmacokinetic assessment of the sites of first-pass metabolism of BMS-181101, an antidepressant agent, in rats. *J Pharm Pharmacol* 1998;50(3):275–278.
83. Dain JG, Collins JM, Robinson WT. A regulatory and industrial perspective of the use of carbon-14 and tritium isotopes in human ADME studies. *Pharm Res* 1994;11(6): 925–928.
84. *Code of Federal Regulations*. Prescription drugs for human use generally recognized as safe and effective and not misbranded: drugs used in research. FDA; 2002. Accessed 3 May 2004.
85. Garattini S. Active drug metabolites. An overview of their relevance in clinical pharmacokinetics. *Clin Pharmacokinet* 1985;10(3):216–227.
86. Rosenborg J, Larsson P, Tegner K, Hallstrom G. Mass balance and metabolism of [(3)H]Formoterol in healthy men after combined i.v. and oral administration-mimicking inhalation. *Drug Metab Dispos* 1999;27(10):1104–1116.
87. Collins JM, Grieshaber CK, Chabner BA. Pharmacologically guided phase I clinical trials based upon preclinical drug development. *J Nat Cancer Inst* 1990;82(16):1321–1326.
88. Uchida T, O'Brien RD. Dimethoate degradation by human liver and its significance for acute toxicity. *Toxicol Appl Pharmacol* 1967;10(1):89–94.
89. Dareer SM, Mulligan LT Jr, White V, Tillery K, Mellett LB, Hill DL. Distribution of [$^{3}$H]cytosine arabinoside and its products in mice, dogs, and monkeys and effect of tetrahydrouridine. *Cancer Treat Rep* 1977;61(3):395–407.
90. Kitteringham NR, Davis C, Howard N, Pirmohamed M, Park BK. Interindividual and interspecies variation in hepatic microsomal epoxide hydrolase activity: studies with *cis*-stilbene oxide, carbamazepine 10, 11-epoxide and naphthalene. *J Pharmacol Exp Ther* 1996;278(3):1018–1027.

# PHARMACODYNAMICS

Beom Soo Shin, Dhaval Shah, and Joseph P. Balthasar
*University at Buffalo, The State University of New York, Buffalo, New York*

**Contents**

1 Introduction
2 Overview
3 "Static" Models of Pharmacological Response
  3.1 Drug–Receptor Binding
  3.2 $E_{max}$ Model
  3.3 Sigmoid $E_{max}$ Model (Hill Equation)
4 "Kinetic" Models of Pharmacological Response
  4.1 Direct Effect Models
  4.2 Indirect Response Models
  4.3 Transduction Models
  4.4 Irreversible Effects
  4.5 Tolerance Models
5 Conclusion
  Acknowledgments
  References

## 1 INTRODUCTION

*Pharmacodynamics* refers to the study of relationships between drug concentrations and pharmacological effects. The mathematical description of these relationships, through the use of appropriate *pharmacodynamic models*, facilitates the characterization and comparison of candidate drugs, facilitates the rational selection of appropriate dosing regimens, and assists in the investigation and identification

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

of the mechanisms associated with drug action. This chapter is structured to provide an overview and a catalog of mathematical models that may be employed to characterize drug pharmacodynamics. Section 2 provides a brief overview of the analysis of pharmacodynamic data, Section 3 reviews mathematical models used to relate fixed (e.g., steady-state) drug concentrations to drug effects, and Section 4 covers basic and advanced models that relate the time course of drug exposure to the time course of drug effects. Section 5 provides a brief summary of the chapter.

## 2 OVERVIEW

Pharmacodynamics can be defined as the study of the relationships between the systemic exposure of drug and its biological action. Upon intravascular or extravascular administration, the drug enters the systemic circulation directly or with an absorption phase. Once the drug reaches the systemic circulation, drug is distributed into various organs and then eliminated by metabolism and excretion. *Pharmacokinetics* refers to the investigation of the rates and extent of drug absorption, distribution, metabolism, and excretion, and *pharmacokinetic models* allow description of the relationship between drug dosing and the time course of drug concentrations in the body. The combination of pharmacokinetic models with pharmacodynamic models, which relate drug exposure to drug response, allows for the characterization and prediction of the relations hips between the timing of drug administration and the time course of drug effects (Fig. 1). The main differences between pharmacokinetics and pharmacodynamics are summarized in Table 1 [1].

Pharmacological responses are initiated by binding of drug molecule with proteins (e.g., enzymes, receptors) or, in certain cases, DNA. The intensity of the drug response is altered as a function of the administered dose, with a higher receptor–drug binding generally producing a greater response. At a certain point, the response may be saturated and the relationship between drug concentration and response becomes nonlinear. In 1927, Clark [2] first introduced the drug–receptor binding concept into pharmacodynamics. It was suggested that the quantitative effect of acetylcholine was mediated by a receptor-binding process. Clark's occupancy theory was elaborated by Ariens and De Groot [3] and Stephenson [4] and became a fundamental basis for the field of pharmacodynamics. Upon long-term infusion or multiple dosing, drug concentrations in each area of the body will approach a steady state, where concentrations are constant with time. In such cases, drug response may be assumed to be "static" or time invariant. Section 3 of this chapter discusses several mathematical models that have been developed to describe static drug concentration–response relationships, including the $E_{max}$ and the sigmoid $E_{max}$ model.

Although static models are useful in some situations, "kinetic" models, which relate the time course of drug concentration to the time course of drug response, provide a much more complete characterization of drug pharmacodynamics, and kinetic models often provide great insight into the underlying mechanisms responsible for drug effects. Kinetic models of drug response are reviewed in Section 4. The simplest of the kinetic models, *direct effect models*, assume a direct relationship

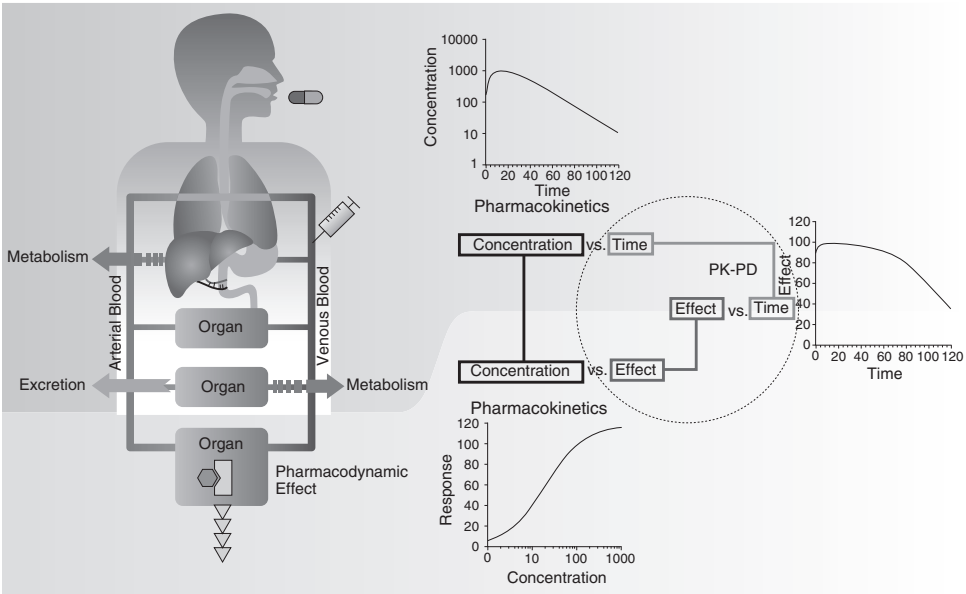


**FIGURE 1** Representative diagram for relationship between pharmacokinetics and pharmacodynamics. Pharmacokinetics mainly deals with the time course of drug absorption (A), distribution (D), metabolism (M), and elimination (E). Pharmacodynamics deals with the drug effect as a function of drug concentration. Pharmacokinetic and pharmacodynamic (PKPD) modeling combines the two processes to allow prediction and characterization of the time course of drug effects.

**TABLE 1 Main Differences Between Pharmacokinetics and Pharmacodynamics**

| Characteristics | Pharmacokinetics | Pharmacodynamics |
|---|---|---|
| Broad definition | What body does to the drug | What the drug does to the body |
| Influence | Affected by interactions throughout the whole body | Affects specific drug receptors/ signaling pathways |
| Determinants | Typically governed by simple physiologic determinants | Typically determined by complex, nonlinear response and control pathways |
| Linearity | Typically linear | Typically nonlinear |
| Initial condition | Baseline concentrations are zero prior to drug dosing | Nonzero baseline response prior to drug dosing |

between drug concentrations in plasma and drug response. In many cases, drug responses are delayed relative to the time course of drug concentrations in plasma. Several models have been developed to describe mechanisms responsible for this delay. For example, the *biophase model* assumes that the delay between the time course of drug concentrations in plasma and the time course of drug effect is due to the kinetics of drug distribution to the site of drug action (i.e., the biophase). *Indirect effect models* assume that the delay is related to the production

or elimination of endogenous substances related to the observed drug effect, and *transduction models* attempt to describe a cascade of signaling events that act as intermediates between drug–receptor binding and drug response. Section 4 also introduces pharmacodynamic models that may be used to characterize irreversible drug responses (e.g., cytotoxicity) and complexities including tolerance and sensitization.

## 3 "STATIC" MODELS OF PHARMACOLOGICAL RESPONSE

Drug response is initiated by the interaction of drug with biomolecules (e.g., receptors, enzymes, DNA) at the biophase. Unfortunately, in most cases, it is difficult or impossible to measure drug concentrations at the biophase and, consequently, pharmacodynamic models typically attempt to relate drug effects to drug concentrations in plasma. However, such relationships are confounded by any disequilibrium that exists between drug concentrations in plasma and drug concentrations at the biophase. Concerns associated with distribution delays may be overcome by the use of *steady-state* drug concentrations. Upon long-term intravenous infusion, a steady state may occur, where drug concentrations in all sites in the body, including plasma and the biophase, remain constant with time. Under these circumstances, a unique, monotonic relationship between drug concentrations in plasma and drug concentrations at the biophase may be assumed. Following administration of drug at a variety of infusion rates, sufficient data may be available to develop relationships between steady state concentrations and the magnitude of drug effect. Due to the nature of the steady state, these relationships may be considered to be time invariant or "static," and these relationships are unable to describe the time course of drug response. This section introduces drug–receptor binding theory and several static pharmacodynamic models, including the $E_{max}$ model, the log-linear $E_{max}$ model, and the sigmoid $E_{max}$ model.

### 3.1 Drug–Receptor Binding

The biological action of a drug is initiated by the interaction of the drug with free receptor, which may be expected to follow the law of mass action:

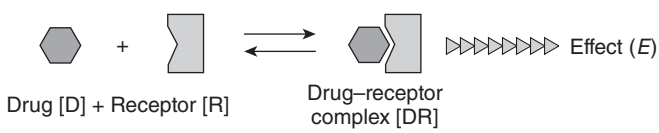


R and D represent free receptor and drug; DR and $E$ represent the drug–receptor complex and the effect of drug, respectively. The brackets [ ] signify the representation of molar concentrations. Typically, the following assumptions are made for this equation:

1. The interaction is reversible.
2. The receptor, drug, and drug–receptor complex are in equilibrium.

3. The receptor contains one binding site for the drug.
4. The drug and receptor interact rapidly to form the drug–receptor complex.

At equilibrium,

$$K_D = \frac{[D][R]}{[DR]} \tag{1}$$

where $K_D$ is the equilibrium dissociation constant for the drug–receptor complex.

The total concentration of receptor, $[R_T]$, is the sum of bound and free receptors.

$$[R_T] = [R] + [DR] \tag{2}$$

By rearranging Eq. 2, we have

$$[R] = [R_T] - [DR] \tag{3}$$

By substituting Eq. 3 into Eq. 1, the following equation is obtained:

$$K_D = \frac{([R_T - [DR]) \times [D]}{[D][R]} = \frac{[R_T][D] - [DR][D]}{[DR]} \tag{4}$$

And by Rearranging Eq. 4, we have

$$[DR] = \frac{[R_T][D]}{K_D + [D]} \tag{5}$$

In 1954, Ariens and De Groot [3] proposed the concept that drug–receptor complex [DR] is linearly proportional to the effect of drug. According to this theory, [DR] can be substituted by effect $E$, and the total receptor concentration $[R_T]$, which will bind with drug and form [DR], can be represented by the maximum response $E_{max}$. The steady-state free drug concentration at the biophase [D] is often assumed to be the same as the steady-state free drug concentration in plasma $[C_p]$. Taken together, Eq. 5 can be rewritten as follows:

$$E = \frac{E_{max} \times [C_p]}{K_D + [C_p]} \tag{6}$$

$K_D$ is in units of concentration (mass/volume). $EC_{50}$ may be defined as the drug concentration associated with the binding of 50% of the drug receptor, where effect is one-half of $E_{max}$ $[E_{max}/2]$.

$$\frac{E_{max}}{2} = \frac{E_{max} \times [EC_{50}]}{K_D + [EC_{50}]} \tag{7}$$

By rearranging Eq. 7, $K_D$ can be represented as the concentration that produces 50% of the maximum response:

$$K_D = EC_{50} \tag{8}$$

Equation 6 can be rewritten as follows:

$$E = \frac{E_{max} \times [C_p]}{EC_{50} + [C_p]} \tag{9}$$

This equation is often referred to as the *maximum effect* ($E_{max}$) *model*.

### 3.2 $E_{max}$ Model

The $E_{max}$ model predicts a hyperbolic relationship between concentration and response, where, at low concentrations, the relationship between concentration and response is linear but, as the dose increases, the response becomes saturated and reaches a plateau (Fig. 2). As indicated in the simple $E_{max}$ model (Eq. 9), drug response will be zero when no drug is present. However, in many cases, the effects produced by endogenous ligands for the receptor may not be differentiated from drug effects, and it is necessary to consider a "baseline" level of response ($E_0$):

$$E = E_0 + \frac{E_{max} \times [C_p]}{EC_{50} + [C_p]} \tag{10}$$

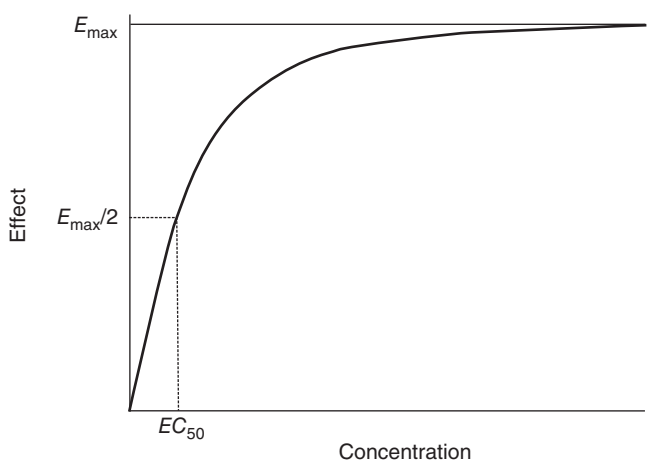


**FIGURE 2** Typical effect versus concentration relationship predicted by the $E_{max}$ model. At low concentrations, there is a linear relationship between drug concentration and response. As the concentration increases, the drug response pathways become saturated and eventually a plateau is achieved. $E_{max}$ represents the maximal effect of the drug and $EC_{50}$ represents the potency of the drug. $EC_{50}$ is defined as the concentration of drug required to produce 50% of the maximum effect.

where $E_0$ represents the baseline effect measured in the absence of drug.

In cases where drug inhibits a physiological process, effect may be characterized by rearrangement of the $E_{max}$ model as follows:

$$E = E_0 - \frac{I_{max} \times [C_p]}{IC_{50} + [C_p]} \tag{11}$$

In the inhibitory model, the maximum inhibitory response and the concentration that produces 50% of the maximum inhibitory response are regarded as $I_{max}$ and $IC_{50}$ instead of $E_{max}$ and $EC_{50}$. In certain cases, response is represented as the percentage change of baseline value. Using the percentage change makes the baseline zero and thus a simple $E_{max}$ model (Eq. 9) may be applied.

***Efficacy, Potency, and Intrinsic Activity*** In Fig. 2, $E_{max}$ represents the *efficacy* of the drug and $EC_{50}$ represents the *potency* of the drug. The efficacy means the strength of response that is generated by a single receptor [4]. The potency is commonly represented by the $EC_{50}$ value. A smaller $EC_{50}$ value means a higher potency. Figure 3 presents effect versus concentration curves for drugs with differing efficacy and potency. In Fig. 3a, both drug A and drug B have the same $E_{max}$ value, although drug A has a smaller $EC_{50}$ value compared with drug B. Thus, to reach the same response, a higher concentration is needed for drug B as compared with drug A, indicating that drug A exhibits a higher potency than drug B. In Fig. 3b, drug A has a smaller $EC_{50}$ value compared with drug B. However, the $E_{max}$ value of drug B is higher than that of drug A. Thus, drug A has a higher potency but lower efficacy compared with drug B. Another important concept of the $E_{max}$ model is the *intrinsic activity* (ε) introduced by Ariens and De Groot [3]. The intrinsic activity can be presented by a relationship between response and occupancy of receptor. Although biological response is directly proportional to receptor occupancy, drugs with different intrinsic activity will generate different levels of response at any given value of receptor occupancy (e.g., as seen with "full agonists" vs. "partial agonists").

$$\varepsilon = \frac{[\mathrm{DR}]}{E} \tag{12}$$

***Linear Model*** At low concentrations of drug, the relationship between concentration and response is assumed to be linear (Fig. 4a):

$$E = mC_p + E_0 \tag{13}$$

where $m$ represents the slope, which is equivalent to $E_{max}/EC_{50}$. However, as the concentration increases, the drug receptors will become saturated and eventually the response reaches the $E_{max}$. Thus, this linearization of the $E_{max}$ model is valid only at low concentrations ($C_p << EC_{50}$).

***Log-Linear Model*** As indicated previously, when plotted on a linear scale, the $E_{max}$ model predicts a hyperbolic relationship between drug concentration and

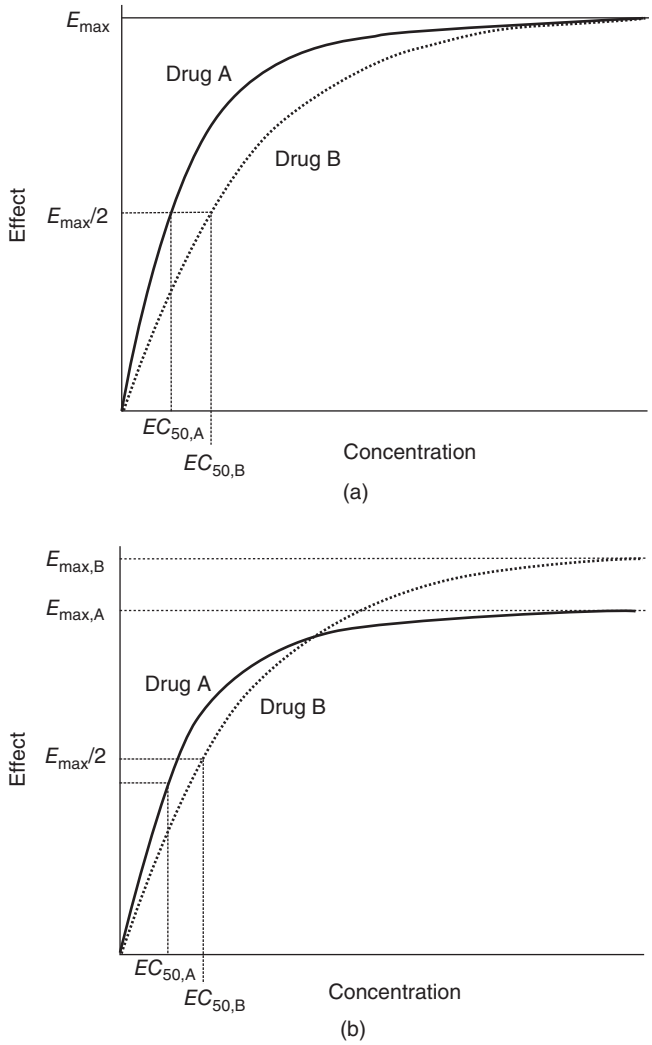


**FIGURE 3** Influence of $E_{max}$ and $EC_{50}$ on effect versus concentration profiles predicted by the $E_{max}$ model. (a) Drug A and drug B have the same efficacy ($E_{max}$) but drug A is more potent compared to drug B ($EC_{50,A} < EC_{50,B}$). Thus, to reach the same response, higher concentrations of drug B are needed compared to drug A. (b) Drug A and drug B show different efficacy and different potency. Drug A has a smaller $EC_{50}$ value compared to drug B, and drug A has a smaller $E_{max}$ value compared to drug B. Thus, drug A shows higher potency but less efficacy compared to drug B.

response. However, when plotted using a log scale for drug concentration, the plot becomes sigmoidal (Fig. 4b), and a log-linear relationship is found between concentration and response over the range of 20–80% of $E_{max}$. This relationship may be represented with the following equation:

$$E = m\log(C_p) + b \tag{14}$$

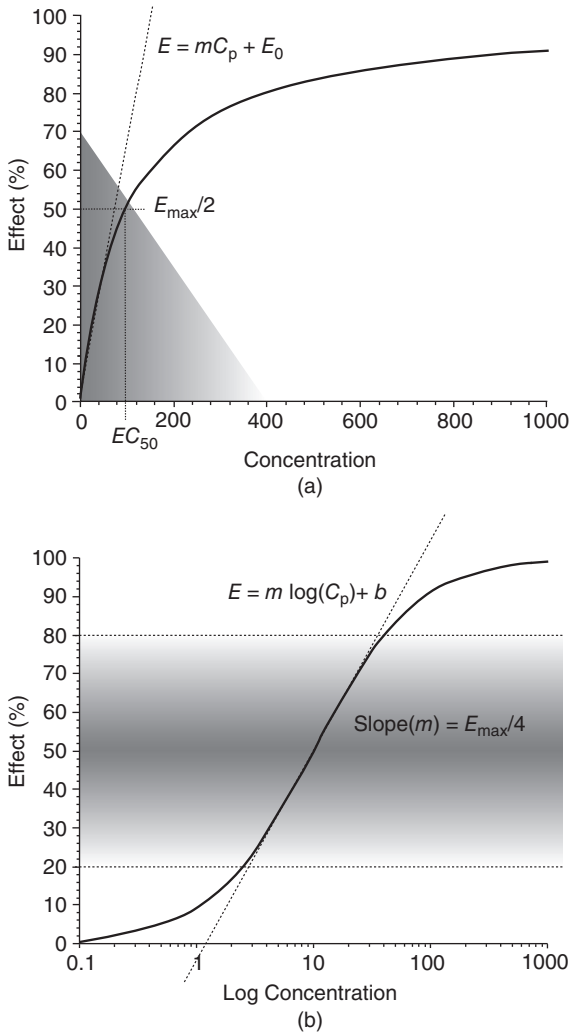


**FIGURE 4** Comparison of rectilinear and semilog plots of effect versus concentration profiles predicted by the $E_{max}$ model. (a) Plotting effect versus concentration profile on a linear scale provides a hyperbolic relationship between response and drug concentration. At very low concentrations of drug, the relationship between response and concentration can be described as linear and can be represented by the equation $E = mC_p + E_0$, where $m$ is the slope and $E_0$ is the intercept. (b) Plotting the effect versus concentration profile on a semilog scale, the relationship between response and concentration becomes sigmoidal. Between 20% and 80% of the maximum response, the relationship between response and concentration is linear and may be described by the equation $E = m \log(C_p) + b$, where $m$ is the slope and $b$ is the intercept. The slope of this linear relationship ($m$) is equal to one-quarter of the maximum response ($E_{max}/4$).

where $m$ and $b$ represent the slope and the intercept, respectively. However, note that in contrast to the linear model, the intercept of the log-linear model does not represent the baseline effect.

The meaning of slope ($m$) may be derived as follows [5]:

$$\text{Slope} = m = \frac{dE}{d\ln C_{\text{p}}} \tag{15}$$

Since $C_{\text{p}}$ can be written as

$$e^{\ln C_{\text{p}}} = C_{\text{p}} \tag{16}$$

Eq. 9 may be rewritten as

$$E = \frac{E_{\max} \times [e^{\ln C_{\text{p}}}]}{EC_{50} + [e^{\ln C_{\text{p}}}]} \tag{17}$$

Solving for the slope, or the derivative of $E$ with respect to $e^{\ln C_{\text{p}}}$, we find

$$\frac{dE}{d\ln C_{\text{p}}} = \frac{E_{\max} C_{\text{p}}(EC_{50} + C_{\text{p}}) - E_{\max} C_{\text{p}}^2}{(EC_{50} + C_{\text{p}})^2} = \frac{E_{\max} EC_{50} C_{\text{p}}}{(EC_{50} + C_{\text{p}})^2} \tag{18}$$

Using Eq. 18, the slope at $C_{\text{p}} = EC_{50}$ can be derived as follows:

$$\frac{dE}{d\ln C_{\text{p}}} = \frac{E_{\max} EC_{50}^2}{(2EC_{50})^2} = \frac{E_{\max}}{4} \tag{19}$$

As shown in Eq. 19, the slope will increase as a function of $E_{\max}$. Additionally, this derivation allows rapid estimation of $E_{\max}$ from the slope of the $E$ versus. ln $C_{\text{p}}$ relationship.

The major benefit of the linear and log-linear models is, of course, the linearization of the relationship. These linearizations allow characterization of pharmacodynamic relationships with the use of simple linear regression, thereby avoiding the need for more computationally intensive analyses (e.g., nonlinear regression) [6, 7]. Although increased availability of nonlinear regression software has minimized the need for application of linearized models for final analyses of datasets, the linear models remain useful for initial analyses and for estimation of initial parameter values (i.e., to facilitate subsequent nonlinear regression analyses).

### 3.3 Sigmoid $E_{\max}$ Model (Hill Equation)

The *sigmoid* $E_{\max}$ *model* was first introduced by Hill to explain the association of oxygen with hemoglobin [8]. Wagner applied this equation to pharmacodynamics in 1968 [9]. Mathematically, this equation is an extension of the $E_{\max}$ model by adding the Hill coefficient ($n$). In the sigmoid $E_{\max}$ model, the relationship between receptor and drug can be represented as

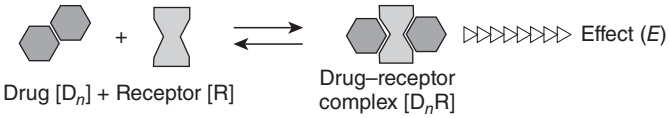


where R and D represent free receptor and drug, respectively. DR and $E$ represent the drug–receptor complex and effect, and $n$ represents the Hill coefficient. At equilibrium,

$$K_D = \frac{[D]^n[R]}{[D_nR]} \tag{20}$$

Free receptor concentration can be described as

$$[R] = [R_T] - [D_nR] \tag{21}$$

Substituting Eq. 21 into Eq. 20, we find

$$K_D = \frac{[D]^n[R_T] - [D_nR][D]^n}{[D_nR]} \tag{22}$$

Rearranging Eq. 22 in terms of $[D_nR]$, we have

$$[D^nR] = \frac{[R_T][D]^n}{K_D + [D]^n} \tag{23}$$

Using a derivation similar to that employed for the $E_{max}$ model (above), we have

$$E = \frac{E_{max} \times [C_p]^n}{EC_{50}^n + [C_p]^n} \tag{24}$$

The Hill coefficient determines the "steepness" of the effect versus concentration curve. As with the $E_{max}$ model, the response range between 20% and 80% of $E_{max}$ is linear when plotted on semilog axes. If $n = 1$, the curve of the sigmoid $E_{max}$ model is identical with the $E_{max}$ model. When $n$ is greater than 1 ($n > 1$), the linear portion of the curve becomes steeper. Conversely, if $n$ is less than 1 ($n < 1$), the linear portion of the curve becomes more shallow. The value of $n$ does not change $EC_{50}$ and $E_{max}$ values and all curves pass through the $EC_{50}$ (Fig. 5).

The steepness of the curve can also be represented by the following equation [5]:

$$\frac{dE}{d\ln C_p} = \frac{nE_{max}}{4} \tag{25}$$

In this equation, the linear portion of the slope is directly proportional to the Hill coefficient. Theoretically, the $n$ value reflects cooperativity in drug binding with

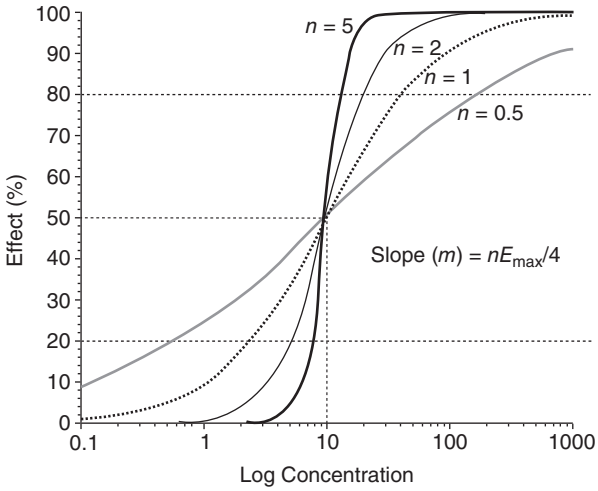


**FIGURE 5** Effect versus concentration profiles predicted by the sigmoidal $E_{max}$ model. If the Hill coefficient, $n$, is equal to 1, the curve is identical to that predicted by the $E_{max}$ model. If $n > 1$, then the linear portion of the curve (between 20% and 80% of the maximum response) becomes steeper; and if $n < 1$, the linear portion becomes more shallow. The slope of the linear portion is equal to the Hill coefficient multiplied by one-quarter of the maximum response ($nE_{max}/4$).

specific receptors. In practical terms, the $n$ value is determined by curve fitting and is typically considered as an empirical parameter with no defined physiological meaning. However, attempts have been made to explain the Hill coefficient pharmacologically and mechanistically. In 1994, Hoffman and Goldberg [10] suggested that the Hill coefficient reflected the extent of heterogeneity in drug receptors. They proposed that the statistical distribution of heterogeneity could affect the steepness of the effect versus concentration curve, where more homogeneous receptor populations would be associated with high values of $n$.

## 4 "KINETIC" MODELS OF PHARMACOLOGICAL RESPONSE

Static pharmacodynamic models are limited to consideration of equilibrium or steady-state conditions and, consequently, static models are unable to describe the time course of pharmacological effect following acute dosing. Evaluation of drug response following acute dosing, particularly where plasma concentrations change rapidly with time, often facilitates the identification of underlying mechanisms associated with drug effect. In this section, "kinetic" models are described that allow characterization of the time course of drug effect. When linked with an appropriate pharmacokinetic model, the combined pharmacokinetic–pharmacodynamic (PKPD) model allows description of the time course of drug effect as a function of the time course of drug administration.

### 4.1 Direct Effect Models

Some drug responses may be characterized as *direct effects*, where interaction of drug and the drug receptor leads to an observed response by direct action of the drug. In practical terms, a drug effect may be categorized as "direct" when all events associated with response proceed very rapidly relative to the kinetics of drug–receptor binding (e.g., consider the vasodilatory and antihypertensive effects of sodium nitroprusside, which appear to be directly and immediately related to the concentrations of sodium nitroprusside in blood). Models appropriate for direct effects are described next.

***Levy's Direct Effect Model*** Following drug administration, the concentration of drug in the body changes with time. Thus, the drug response also changes as a function of time. The first PKPD models, introduced by Gerhard Levy [11, 12], described drugs with direct effects, where it may be assumed that there is no time delay between drug response and drug concentration. Additionally, Levy's initial model assumed rapid equilibration between the drug concentrations in plasma and at the site of action. This model consisted of a one-compartment pharmacokinetic model and log-linear pharmacodynamic model.

Equation 26 represents drug concentration as a function of time for a one-compartment model, where $k$ represents a first-order elimination rate constant:

$$\ln C = \ln C_0 - kt \tag{26}$$

Equation 27 represents the log-linear model, which explains the linear relationship between response and log concentration in the response range between 20% and 80% of $E_{max}$:

$$E = m \ln C + e \tag{27}$$

Equation 27 may be rearranged as

$$\ln C = \frac{E - e}{m} \tag{28}$$

At the initial concentration, $C_0$, the effect can be represented by the initial effect, $E_0$. Thus, another equivalent equation may be derived:

$$\ln C_0 = \frac{E_0 - e}{m} \tag{29}$$

By substituting Eqs. 28 and 29 into Eq. 26, the following equation results:

$$\frac{E - e}{m} = \frac{E_0 - e}{m} - kt \tag{30}$$

Finally, multiplying each term by $m$ and simplifying yields

$$E = E_0 - kmt \tag{31}$$

Interestingly, the model predicts that pharmacological effect will decline linearly with time, whereas drug concentrations decline exponentially with time (Fig. 6) [13]. Although this model only describes the region associated with 20–80% of maximal response, this region includes the entire therapeutic working range for many approved drugs (e.g., theophylline) and, consequently, this simple model has been shown to be extremely useful.

**$E_{max}$ *Model*** Levy's direct response model only covers the linear range (20–80% of maximum response) of the time course of pharmacodynamic responses. Most time courses of pharmacodynamic profiles consist of linear and nonlinear curves. To obtain a complete profile of the time course of pharmacodynamic responses, the

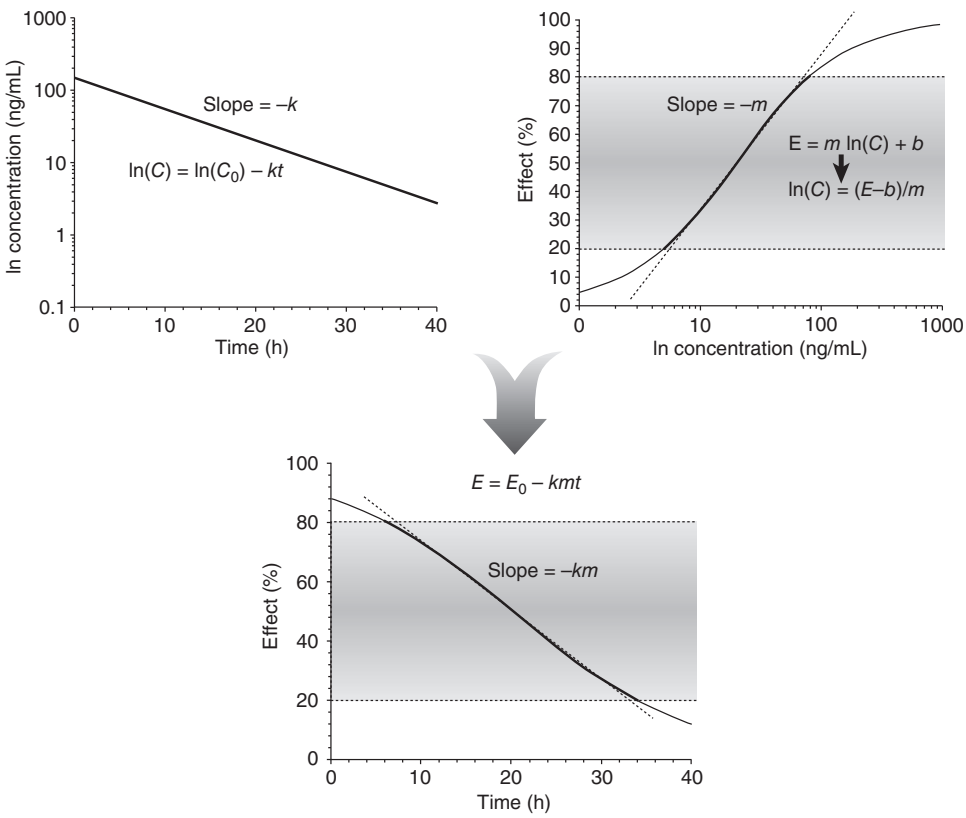


**FIGURE 6** Levy's direct effect model. This model consists of two functions, one is pharmacokinetics (linear, one-compartment) and the other is pharmacodynamics (log-linear; $E_{max}$ model). Note that the model predicts a linear decline in pharmacologic effect with time despite an exponential decline in drug concentration with time. Effect declines with the slope ($-km$), reflecting both pharmacokinetics ($k$) and pharmacodynamics ($m$).

$E_{max}$ or sigmoid $E_{max}$ model may be linked with appropriate pharmacokinetic models. In Fig. 7, the time course of drug concentration and pharmacological response are shown. This decline of drug concentration is predicted from a one-compartment model after intravenous bolus injection (Eq. 32) and the pharmacological response curve is predicted by the $E_{max}$ model linked to the one-compartment pharmacokinetic model (Eq. 33).

$$C = C_0 e^{-kt} \tag{32}$$

$$E = \frac{E_{max} C_0 e^{-kt}}{E_{50} + C_0 e^{-kt}} \tag{33}$$

Immediately following intravenous (IV) bolus injection of a drug, the model predicts a small change in the drug response despite a large change in drug concentration. This prediction is due to the saturation of receptors at high drug concentrations. Within the range of 20–80% of the maximum response, the time course of pharmacological response declines linearly as a function of time, which is consistent with Levy's direct effect model. At low concentrations, below the $EC_{50}$, the drug response declines in parallel with drug concentration as predicted by the log-linear model. The direct effect $E_{max}$ model was first introduced by Wagner in 1968. Using the $E_{max}$ model linked with one-compartment IV bolus injection model, Wagner measured the total response of various dosing intervals by calculating the area under the response–time curve ($AUC_E$) [9]. He demonstrated that the total response was increased substantially by decreasing the dosing interval from 24 h to

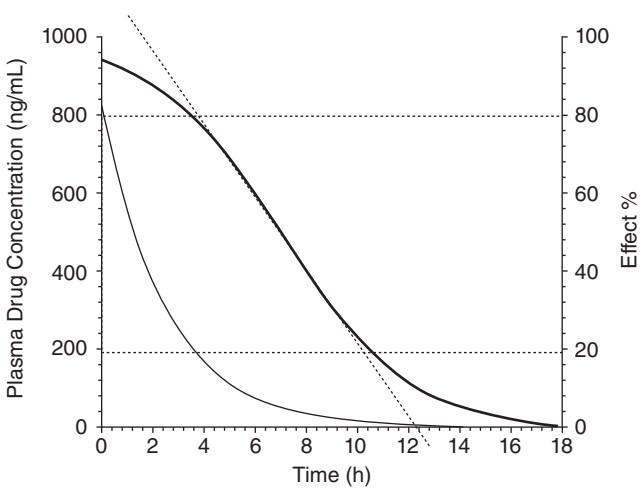


**FIGURE 7** Predictions of a PKPD model defined with a one-compartment pharmacokinetic model and a direct effect $E_{max}$ pharmacodynamic model. Initially, drug effect shows a plateau despite a large change of concentration, and this is due to the saturation of drug receptors at high drug concentrations. Within the range of 20–80% of maximum response, the time course of pharmacological response declines linearly as a function of time. At concentrations less than $EC_{50}$, drug response declines in parallel with drug concentration.

12 h (with simulated administration of the same total dose). With further decreases in the dosing interval (to 6 h or 3 h), the response continued to increase, but in smaller increments. As such, this simulation predicted that drug response could be changed by modulation of the dosing regimen, and high drug response may be achieved by dividing doses into smaller quantities given more frequently. This principle has been demonstrated in clinical studies [14, 15] and provides an early demonstration of the use of modeling and simulation to guide the selection of optimal dosing strategies.

***Biophase Model*** In many cases, the time course of drug effect is delayed relative to the time course of plasma concentrations. This delay between drug effect and concentration may be recognized by comparing the time value associated with maximum plasma concentrations ($T_{max}$) and the time value associated with maximum effect. Additionally, plotting the effect data versus plasma concentration data will often show a nonunique relationship of effect and concentration, where the same value of drug concentration is associated with two different levels of effect (Fig. 8). The plot of effect versus concentration will form a counterclockwise hysteresis loop, where, at a given concentration, the effect value is lower prior to $T_{max}$ relative to effect value observed following $T_{max}$. Delays in the relationship between drug effect and drug concentration may be explained by several mechanisms. Common causes for delays between effect and plasma concentration include (1) distribution delays between plasma and the site of drug action (i.e., the biophase), (2) indirect effect kinetics, where the delay is related to the time course of production or elimination of an endogenous mediator of effect, (3) delays associated with the formation of an active metabolite, and (4) delays associated with transduction mechanisms. Clockwise hysteresis may also be observed in effect versus plasma concentration plots, where effect at a given concentration decreases with time. Clockwise hysteresis occurs much less frequently than counterclockwise hysteresis, and it is typically associated with the development of pharmacological tolerance [16].

The concept of drug distribution to the biophase was first introduced by Furchgott in 1955 [17], and mathematically described by Segre in 1968. Several years later, Sheiner and co-workers [18] introduced a simple model consistent with first-order drug distribution to the biophase, and this *effect site model* rapidly gained wide use. Figure 9 shows a simple effect site PKPD model, where a one-compartment IV bolus injection model is linked with the hypothetical effect site compartment. The differential equations for the rate of change in drug concentration of central and hypothetical effect compartments are

$$\frac{dC_p}{dt} = -k_{el}C_p \tag{34}$$

$$\frac{dC_e}{dt} = k_{1e}C_e - k_{eo}C_e \tag{35}$$

Solving these two differential equations yields

$$C_p = C_0 e^{-kt} \tag{36}$$

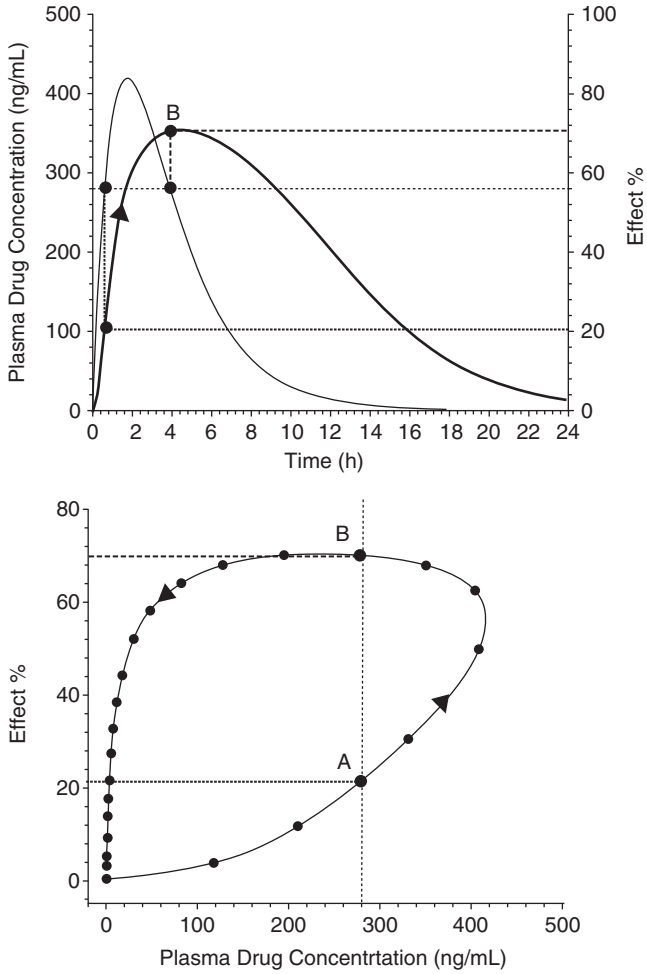


**FIGURE 8** Counterclockwise hysteresis curve of effect versus plasma drug concentration. In case of delayed response (upper panel), plotting of effect versus plasma concentration will show a counterclockwise hysteresis loop (lower panel). Counterclockwise hysteresis loops show a nonunique relationship of effect and concentration, where a given concentration value is associated with two values of drug effect (note ascending (A) and descending (B) concentrations).

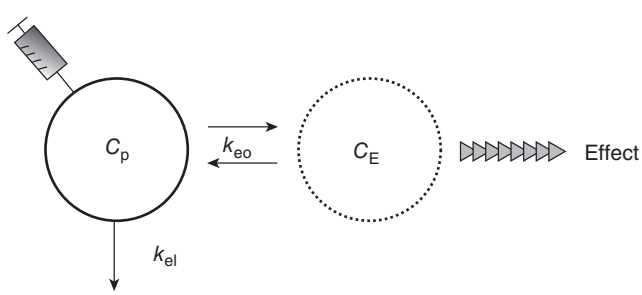


**FIGURE 9** Biophase model. $k_{el}$ represents the elimination rate constant of drug from the central compartment and $k_{eo}$ represents the rate constant for drug distribution between the central and biophase compartments. (Adapted from Ref. 18.)

$$C_e = \frac{k_{1e}C_0}{k_{eo} - k_{el}}(e^{-kt} - e^{-k_{eo}t}) \tag{37}$$

Of note, the effect site model assumes that an insignificant quantity of drug distributes to the biophase and, consequently, the effect site compartment does not influence the plasma pharmacokinetics of the drug. The rate constants $k_{1e}$ and $k_{eo}$ are typically assumed to be identical for model simplification and, with this simplification, Eq. 37 may be rewritten as

$$C_e = \frac{k_{eo}C_0}{k_{eo} - k_{el}}(e^{-kt} - e^{-k_{eo}t}) \tag{38}$$

In most applications of the effect site model, the drug is assumed to produce a direct effect, which is described through the relationship of the effect site concentration, $C_e$, and the pharmacological response with an $E_{max}$ or sigmoid $E_{max}$ model.

$$E = \frac{E_{max}C_e}{E_{50} + C_e} \tag{39}$$

Distribution delays have also been described in PKPD models by the assumption that the biophase is within a peripheral compartment of a mammillary compartmental pharmacokinetic model [19, 20]. Although these models are quite similar to Sheiner's effect site model, some important differences exist. First, the hypothetical effect site compartment of Sheiner's model is not an actual compartment, and a negligible amount of drug is assumed to distribute to the hypothetical effect compartment. This independency between the hypothetical effect compartment and the pharmacokinetic model provides tremendous flexibility and, consequently, Sheiner's model is much more robust than models assuming that effect occurs in a peripheral pharmacokinetic compartment. Second, mammillary, multicompartment pharmacokinetic models predict that drug concentrations in the central and peripheral compartments decline in parallel; however, the biophase model allows values of $k_{eo}$ that may be much smaller than any of the pharmacokinetic rate constants. This facilitates characterization of the time course of effects that dissipate at a rate much slower than predicted by the plasma clearance of drug (Fig. 10). The value of $k_{eo}$ dictates the shape of the hysteresis curve, as shown in Fig. 11, and also influences the magnitude of the response, as shown in Fig. 12. Due to the assumption of first-order kinetics for entry and loss of drug from the biophase, the effect site model predicts that peak concentrations in the effect compartment will occur at the same time for all doses (i.e., from a given dosage regimen), and the effect site model predicts that all doses will lead to peak effects at the same time (relative to drug administration). Following large doses of drug, the effect will plateau at $E_{max}$ and the effect versus time curve will decline linearly over the range of 20–80% of maximum effect. The biophase model is very robust, and the model has been applied successfully to characterize the delayed pharmacological responses of many drugs [21–24].

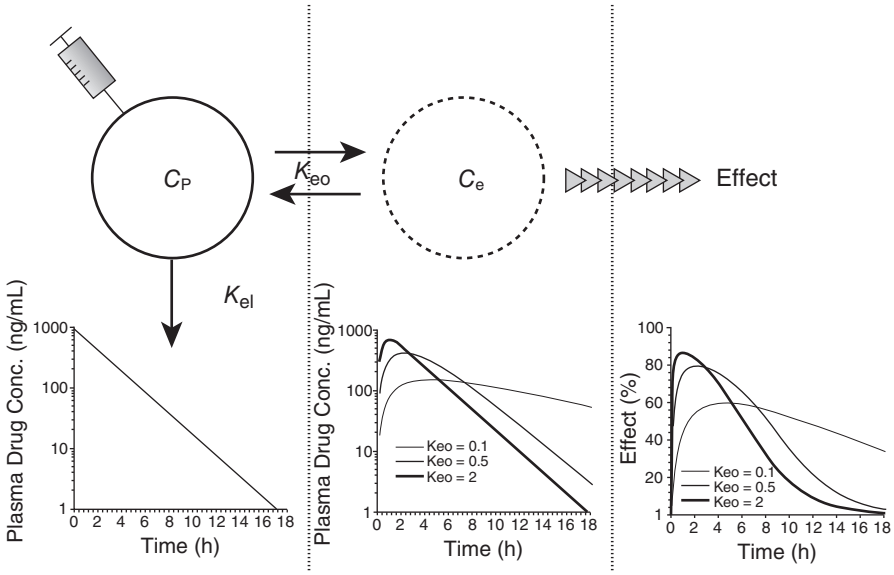


**FIGURE 10** Simulated time course profile of effect site drug concentration and drug effect as a function of $k_{eo}$. Higher values of $k_{eo}$ provide a more rapid rise and fall in effect compartment concentrations and responses that are less delayed. Small values of $k_{eo}$, where $k_{eo} << k_{el}$, leads to a "flip-flop" phenomenon where the time course of effect is very protracted relative to the time course of drug concentrations in plasma.

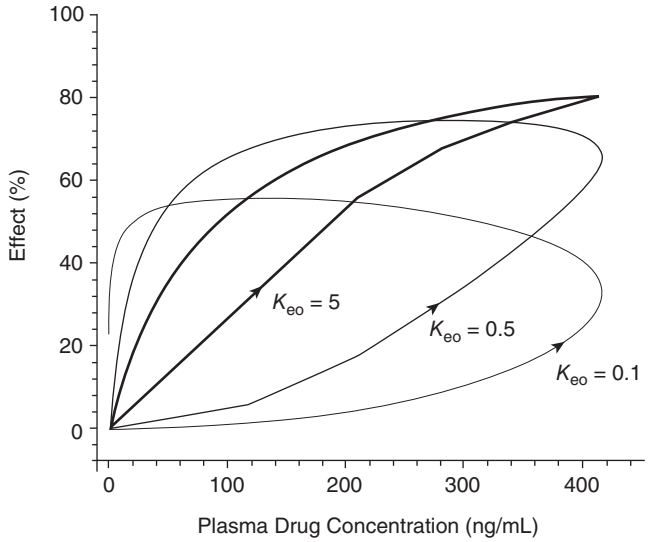


**FIGURE 11** The impact of $k_{eo}$ on the shape of the hysteresis loop. At high values of $k_{eo}$, there is rapid equilibration between plasma and effect compartment drug concentrations, and this leads to little delay in the time course of drug effect relative to drug concentration. As $k_{eo}$ approaches infinity, the hysteresis loop collapses.

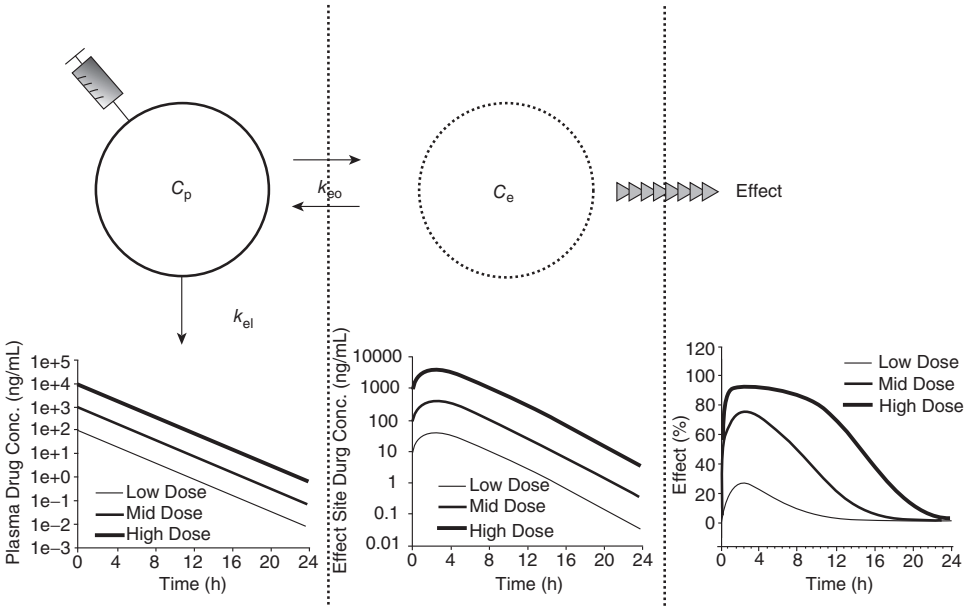


**FIGURE 12** Biophase model simulations: effect of dose. Drug concentration versus time profiles within the hypothetical effect compartment are delayed relative to plasma concentrations; however, the time associated with peak concentrations in the effect compartment are identical for all doses. At large doses, the plateau in the response profile represents $E_{max}$, and the magnitude of effect declines linearly in the range of 20–80% of the maximum response.

## 4.2 Indirect Response Models

Following the introduction of the biophase model [18], many investigators used the biophase model to explain time delays in drug response with little thought as to whether the biophase model was mechanistically appropriate for the given application. Given the simple, robust nature of the biophase model, the model provided good characterization of a wide variety of response data; however, in many cases, the model did not allow simultaneous characterization of data from different doses, and the model did not allow accurate, prospective predictions of drug response. One main reason for these difficulties was the application of the biophase model to drugs that achieve effects indirectly, where the measured drug response occurs not as a direct effect, but due to the effect of drug on the production or elimination of endogenous substances associated with the measured response (Fig. 13).

Areins introduced the concept of indirect effects in 1964 [25], while Nagashima et al. [7] first applied an indirect effect model for PKPD to explain the effect of warfarin on the time course of prothrombin complex activity. Warfarin is an inhibitor of vitamin K epoxidase, which reduces the vitamin K epoxide to vitamin K. Vitamin K is needed for blood coagulation as a cofactor for carboxylation of clotting factors; thus, inhibition of vitamin K production results in the inhibition of coagulation [26]. Due to the indirect nature of the effect of warfarin on the measured response (coagulation), and due to the time delays associated with the elimination

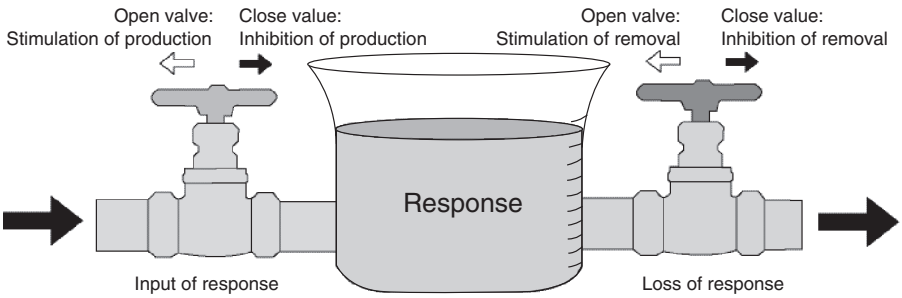


**FIGURE 13** Indirect effects: impact of changes in the production or elimination of response mediators on the measured response. Stimulation of the production or inhibition of the elimination of response mediators will lead to an increase in the measured response. However, inhibition of input or stimulation of the elimination of response mediators will lead to a decrease in the measured response.

of vitamin K and clotting factors, the time course of warfarin effect is dramatically delayed relative to the time course of warfarin exposure, and this delay was well captured by the indirect effect model of Nagashima et al. [7] Jusko and Ko [27] subsequently developed and elaborated a family of models to allow mechanistically appropriate characterization of four common types of indirect effects.

The conceptual basis for indirect response models is provided in the differential equation

$$\frac{dR}{dt} = k_{\text{in}} - k_{\text{out}} R \tag{40}$$

where $R$ refers to the measured response, $k_{\text{in}}$ is the zero-order rate constant for the production of the response, and $k_{\text{out}}$ is the first-order rate constant for loss of the response. Under steady-state conditions, the change of response over time becomes zero; thus, the initial response can be represented as follows:

$$R_0 = \frac{k_{\text{in}}}{k_{\text{out}}} \tag{41}$$

After administration of drug, the measured response ($R$) is changed from the effect of drug (inhibition or stimulation) on the production or dissipation of response, as a function of drug concentration. Four general submodels have been deduced (Fig. 14). Models I and II describe the time course of drug response by inhibition of the rate constant associated with production ($k_{\text{in}}$) or dissipation ($k_{\text{out}}$). The inhibitory function includes a parameter, $I_{\max}$, which describes the maximal extent of inhibition; $I_{\max}$ is always greater than zero and less than or equal to one ($0 < I_{\max} \leq 1$).

$$I(t) = 1 - \frac{I_{\max} C_{\text{p}}}{IC_{50} + C_{\text{p}}} \tag{42}$$

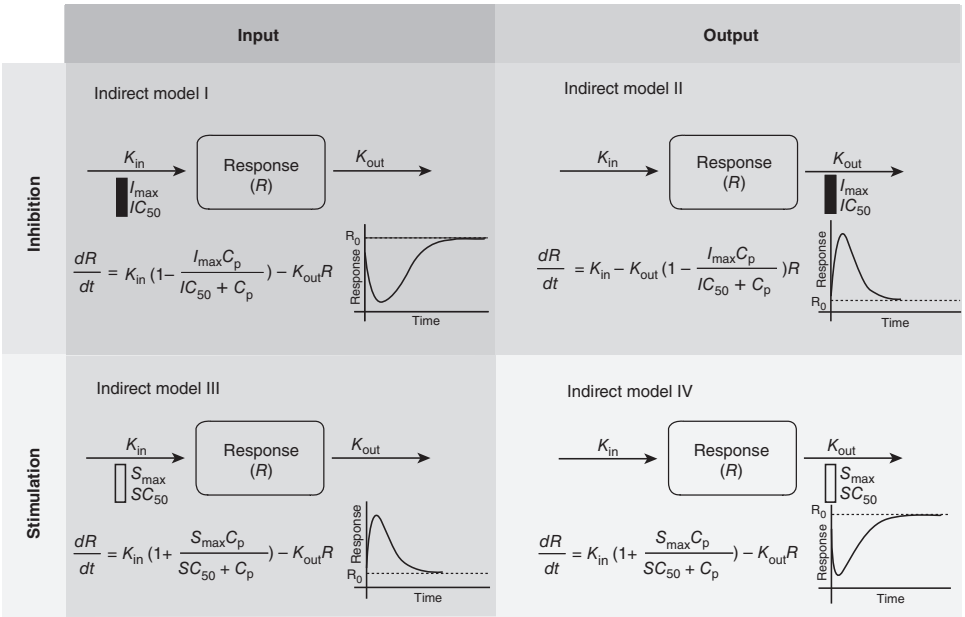


**FIGURE 14** Basic indirect effect models. Shown are the four basic indirect response submodels (I, II, III and IV), and typical response patterns.

$C_p$ represents the plasma concentration and $IC_{50}$ represents the drug concentration needed to produce 50% of the maximum inhibitory response.

Models III and IV include a stimulation function $S(t)$, which affects the production or dissipation of response.

$$S(t) = 1 + \frac{S_{max} C_p}{SC_{50} + C_p} \tag{43}$$

$C_p$ represents the plasma concentration, $SC_{50}$ represents the drug concentration that produces 50% of the maximum stimulation response, and $S_{max}$ represents the maximum stimulation response. Unlike the inhibitory function $I(t)$, the value of $S_{max}$ is any value greater than zero.

In model I, the rate of change in response over time may be represented by Eq. 43, and the results of simulation of drug response after intravenous injection of four different doses are shown in Fig. 15.

$$\frac{dR}{dt} = k_{in}\left(1 - \frac{I_{max} C_p}{IC_{50} + C_p}\right) - k_{out} R \tag{44}$$

The "baseline" response, in the absence of drug, is assumed to be equivalent to the steady-state solution provided in Eq. 41. Immediately after dosing, plasma drug concentrations are high ($C_p >> IC_{50}$), and inhibitory function $I(t)$ becomes $(1 - I_{max})$. If $I_{max} = 1$, then the term $k_{in}$ is eliminated, and the decline of the drug response is

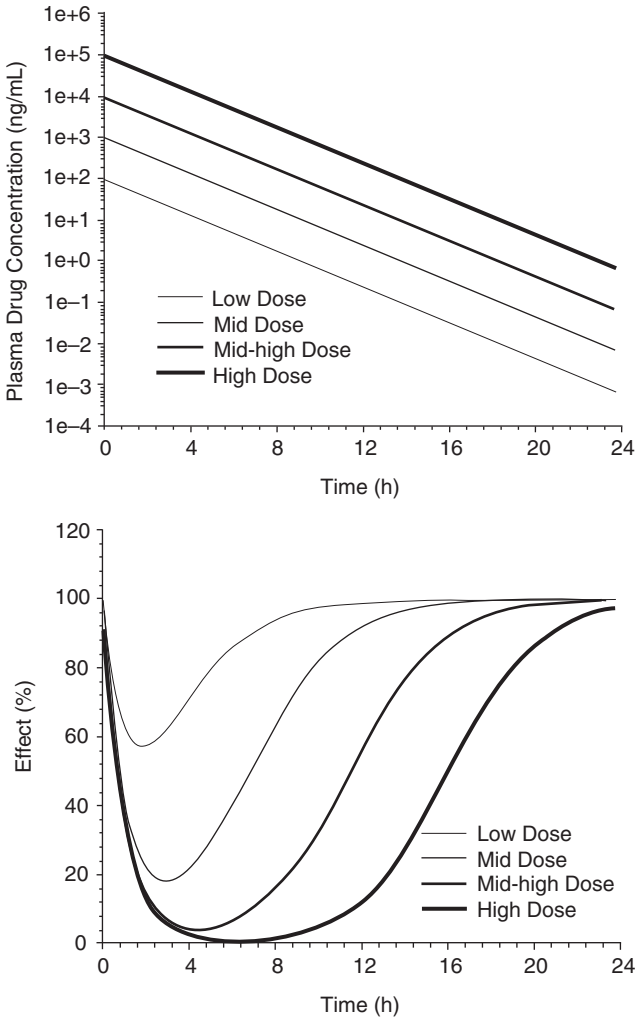


**FIGURE 15** Simulated profile of the indirect response model I, with four different doses. As the amount of dose increases, indirect response model I predicts an increase an response and an increase in the time associated with maximum effect.

governed solely by the first-order rate of $k_{out}$. Thus, initial decline in the slope of the ln response versus time profile becomes $-k_{out}$. As plasma concentrations decrease, the inhibitory effect of drug decreases and, eventually, response begins to increase back to the baseline value. The minimum response value may be defined as the peak effect of drug, $R_{max}$. At $R_{max}$, the change of drug response with time will equal zero, and Eq. 44 may be rearranged to solve for $R_{max}$ as follows:

$$R_{max} = \frac{k_{in}}{k_{out}}\left(1 - \frac{I_{max}C_p}{IC_{50} - C_p}\right) = R_0\left(1 - \frac{I_{max}C_p}{IC_{50} - C_p}\right) \tag{45}$$

When $C_p >> IC_{50}$, the inhibitory function $I(t)$ becomes $(1 - I_{max})$, and Eq. 45 can be rewritten as

$$R_{max} = R_0(1 - I_{max}) \quad \text{model I} \tag{46}$$

In a similar manner, $R_{max}$ for models II, III, and IV can be shown as follows:

$$R_{max} = R_0/(1 - I_{max}) \quad \text{model II} \tag{47}$$

$$R_{max} = R_0(1 + S_{max}) \quad \text{model III} \tag{48}$$

$$R_{max} = R_0/(1 + S_{max}) \quad \text{model IV} \tag{49}$$

After $R_{max}$ has been reached, the response starts to go back to baseline. However, even if the plasma drug concentration is below the $IC_{50}$, time is needed for reequilibration and to return back to baseline. This is the primary difference from the biophase model. In the biophase model, the delayed response is governed by the rate constant from plasma to biophase, $k_{eo}$; thus, the time of maximum effect is not affected by dose. However, in indirect response models, higher doses will result in a longer plateau and increase in the time of maximum effect.

Concerning the shape of the time course of the response profile, four subgroup indirect response models can be divided into two groups, with either a downward trend in response (for models I and IV) or an upward trend in response (for models II and III). In assigning an appropriate model, it is reasonable to start with consideration of the shape of the response profile. Additionally, Sharma and Jusko [28] proposed an infusion study, using multiple dosing rates to allow collection of data to assist with model selection. In many ways, the kinetics represented by the indirect response model equation are similar to those associated with IV infusion in a one-compartment pharmacokinetic model. The zero-order rate constant for the production ($k_{in}$) corresponds to the zero-order infusion rate ($K_o$), and the first-order rate constant ($k_{out}$) for loss of the response corresponds to the elimination rate constant of plasma compartment ($k_{el}$). In the case of infusion, the time to reach steady state ($t_{R_{max}}$) is governed by the elimination rate constant, not the infusion rate. Similarly, in indirect response, the time to reach maximum response is governed by $k_{out}$. In models I and III, $t_{R_{max}}$ is constant (for all drug infusion rates), since $k_{out}$ is not affected by changes in "input." However, in models II and IV, $t_{R_{max}}$ would change as a function of drug infusion rate, because $k_{out}$ is affected by the administration of drug. As such, appropriate selection of indirect models is facilitated by administration of drug over a range of infusion rates.

### 4.3 Transduction Models

The starting point for a pharmacological response is drug–receptor binding. However, in many cases, generation of pharmacological responses requires subsequent signal transduction events. For example, prednisone, a glucocorticoid hormone, binds to cytosolic receptors. The prednisone–receptor complex is then tranlocated to the nucleus, where the complex binds with DNA, modulating the rate of RNA transcription, protein production, and, finally, protein expression [29]. Such cascading steps for signal transduction require time and lead to delayed responses. Ideally, receptor

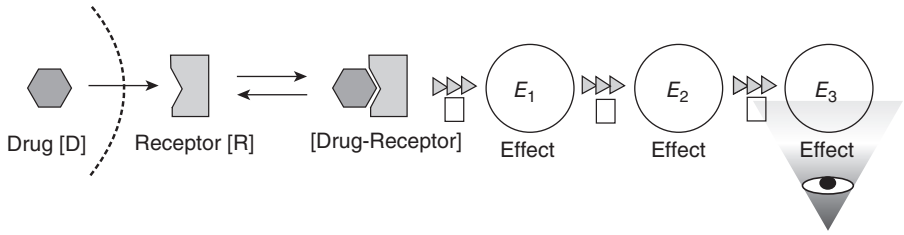


**FIGURE 16** Transit compartment model for signal transduction. Response starts by drug–receptor binding, but the measured effect is achieved via a signal transduction process. τ represents transit time required for the signal to move from compartment to compartment in the transduction cascade. (Adapted from Ref. 30.)

and gene mediated processes should be included in a PKPD model to explain the delayed responses. However, data associated with each step in the signal transduction cascade is difficult to obtain due to limitations of technical feasibility. Although the delayed response may be due to complicated signal transduction pathways, relatively simple transit compartment models have been applied to successfully describe the time course of associated responses *in vivo* [30]. A representative transit compartment model for the signal transduction process is shown in Fig. 16.

As proposed by Ariens and Stephenson, the drug effects may be directly proportional to the concentration of drug–receptor complexes. The initial effective signal in a signal transduction cascade, $E^*$, may be defined as a function of drug–receptor complexes and the intrinsic efficacy of the drug:

$$E^* = \varepsilon \cdot [\mathrm{DR}] \tag{50}$$

where ε represents the efficacy of the drug.

The $E^*$ term may be included within any of the direct or indirect effect models described previously. For example, within the $E_{\max}$ model,

$$E^* = \frac{E_{\max} C_e}{EC_{50} + C_e} \tag{51}$$

where $C_e$ represents the free drug concentration at the effect site.

To characterize signal transduction, a cascade of transit compartments is employed:

$$\frac{dE}{dt} = \frac{1}{\tau}(E^* - E) = \frac{1}{\tau}\left(\frac{E_{\max} C_e}{EC_{50} + C_e} - E\right) \tag{52}$$

where τ represents the transit time.

Additional transit compartments may be employed to allow greater delay and flexibility in the response versus time profile. If the signal transduction process involves three transit compartments, the signal transit step of each transit compartment can be represented by the following series of equations:

$$\frac{dE_1}{dt} = \frac{1}{\tau}(E^* - E_1) = \frac{1}{\tau}\left(\frac{E_{\max}C_e}{EC_{50} + C_e} - E_1\right) \tag{53}$$

$$\frac{dE_2}{dt} = \frac{1}{\tau}(E_1 - E_2) \tag{54}$$

$$\frac{dE_3}{dt} = \frac{1}{\tau}(E_2 - E_3) \tag{55}$$

The transit time between each compartment is often assumed to be equal ($\tau = \tau_1 = \tau_2 = \tau_3$). Simulated profiles for each transit compartment, with different transit time values, are shown in Fig. 17. As shown in this figure, the response associated with the final transit compartment ($E_3$) is more delayed compared with the initial transit compartment ($E_1$). Also, as the transit time ($\tau$) increases, responses become more delayed, and the maximum peak effect is decreased.

Within signal transduction processes, there is opportunity for the amplification or dampening of the signal. Power coefficients may easily be incorporated with transit compartments to account for these effects.

$$\frac{dE_3}{dt} = \frac{1}{\tau}(E_2^{\gamma} - E_3) \tag{56}$$

Simulated data associated with a range of power coefficients are shown in Fig. 18 in terms of simulation. For example, if the power coefficient of the last transit compartment is less than 1, the effect is lower than the unity condition. This indicates that effect is suspended between transit compartments $E_2$ and $E_3$. Conversely, if the power coefficient of the last transit compartment is greater than 1, the effect is higher than the unity condition. This indicates that effect is amplified between transit compartments $E_2$ and $E_3$ [31].

Transit compartment models are typically employed in an empirical manner to allow curve fitting with a minimal number of parameters. However, when detailed mechanistic data are available, this information may easily be added to the structure of the transit compartments to enhance the mechanistic reality of the model [33–35].

### 4.4 Irreversible Effects

Drug effects may be categorized as reversible or irreversible. Examples of irreversible effects include cell killing by action of anticancer or antimicrobial agents, irreversible enzyme inhibition, and formation of covalent drug adducts (e.g., via formation of reactive drug metabolites [1]). PKPD modeling of irreversible effects was introduced by Jusko [36] to explain chemotherapeutic effects.

A typical cell proliferation model of a chemotherapeutic agent is shown in Fig. 19a. The representative equation for the growth and death of cells by the action of a drug is

$$\frac{dR}{dt} = k_s \cdot R - k \cdot C \cdot R \tag{57}$$

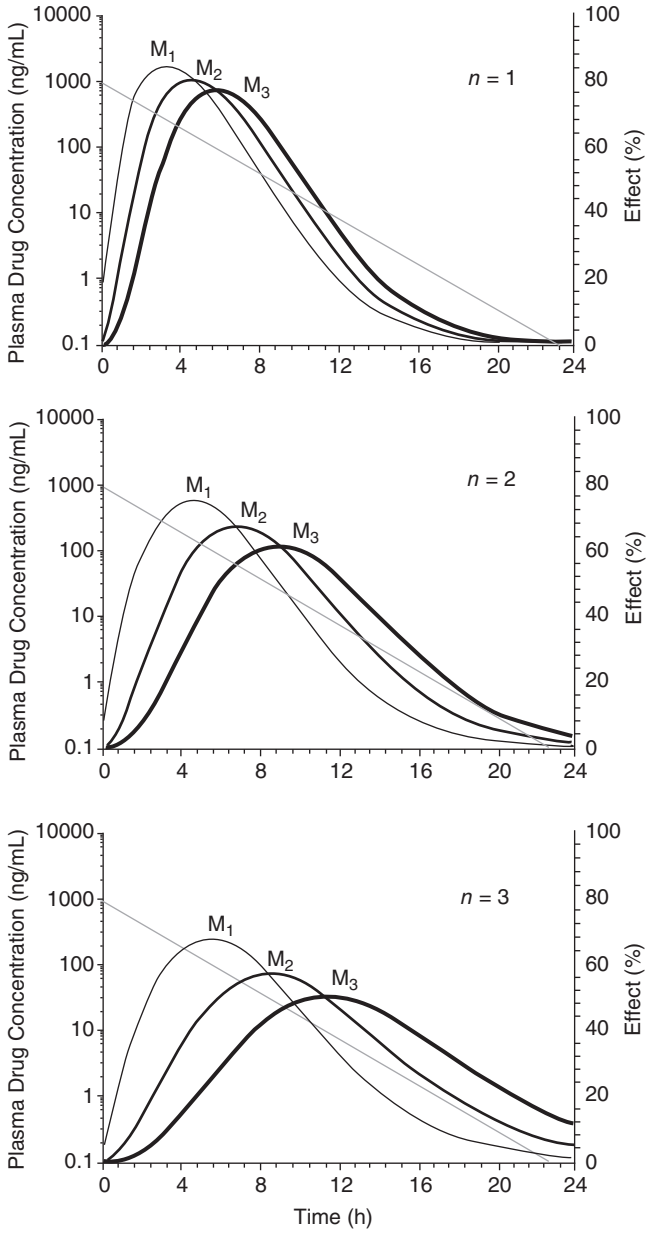


**FIGURE 17** Transit compartment model: effect of transit time. Response in the final compartment ($M_3$) is delayed compared to the first effect compartment ($M_1$). As transit time increases, the response becomes more delayed for all three compartments, and the maximum effect decreases (upper, middle, and lower panels).

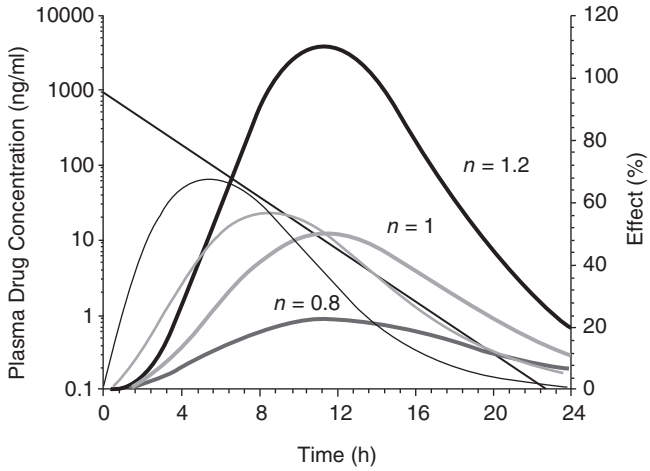


**FIGURE 18** Transit compartment model: effect of the power coefficient. Higher values of the power coefficient ($\gamma > 1$) will amplify response; conversely, power coefficients below 1 will lead to a dampened response.

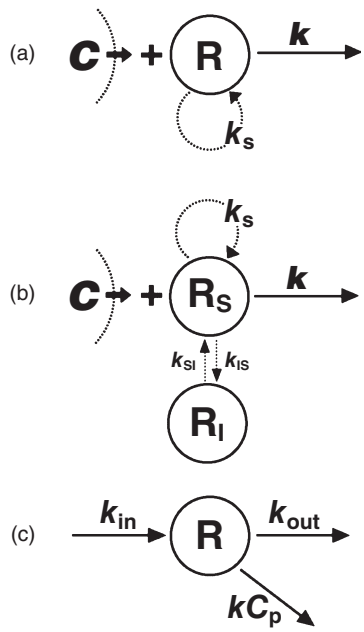


**FIGURE 19** Representative irreversible effect model for chemotherapeutic agents. **(a)** Cell killing model for the effect of a chemotherapeutic agent. (Adapted from Ref. 36.) **(b)** Phase-specific model of cell killing. $k_{SI}$ and $k_{IS}$ are the rate constants for cell movement between compartments associated with sensitivity to drug and with insensitivity to drug. (Adapted from Ref. 38.) **(c)** Application of an indirect response model within a model for irreversible cell killing. (Adapted from Ref. 40.)

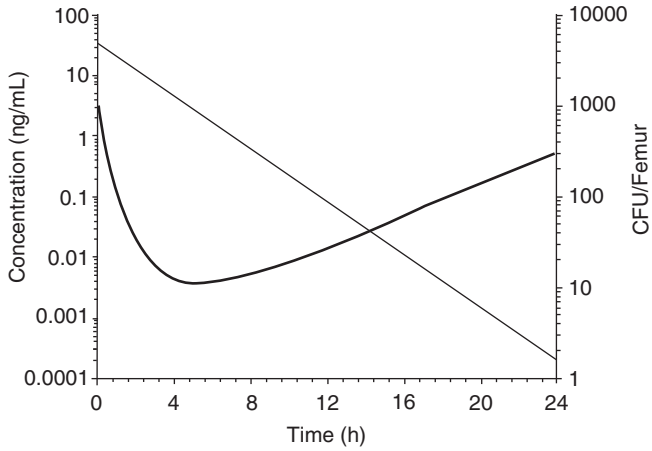


**FIGURE 20** A typical simulated profile for the effect of a chemotherapeutic agent on cell number. The biphasic profile results from the processes of cell proliferation and drug-induced cell killing.

$R$ represents the number of cells, $k_s$ represents the first-order rate constant for growth rate in the natural condition, and $k$ represents the second-order rate constant for cell killing by the action of drug. This equation includes terms for the natural growth and death of cells by the drug effect; thus, the change in the number of cells as a function of time is represented as a biphasic profile that consists of killing and regrowing phases (Fig. 20). The linear term for death of a cell by the drug effect ($kC$) can be substituted by the $E_{max}$ model as

$$\frac{dR}{dt} = k_s \cdot R - \frac{K_{max}C}{KC_{50} + C} \cdot R \tag{58}$$

Using this model, Zhi et al. [37] characterized the effect of piperacillin on the killing and growth kinetic of *Pseudomonas aeruginosa*.

Some chemotherapeutic agents show cell cycle phase-specific effects. To characterize this property, Jusko [38] proposed extension of the cell proliferation model by adding an insensitive cell compartment (Fig. 19b):

$$\frac{dR_S}{dt} = k_S \cdot R_S - k \cdot C_S \cdot R_S - k_{SI}R_S + k_{IS}R_I \tag{59}$$

$$\frac{dR_I}{dt} = k_{SI}R_S - k_{IS}R_I \tag{60}$$

$R_I$ represents the number of insensitive cells and $k_{SI}$ and $k_{IS}$ represent the rate constants for conversion into each cell compartment. The interconversion rate between sensitive and insensitive cells is governed by the rate constants $k_{SI}$ and $k_{IS}$. Another modified form of this model was successfully applied to describe the pharmacodynamic effect of β-lactam antibiotics in *in vitro* bactericidal kinetics [39].

The modeling of irreversible effect can be extended to the indirect response model (Fig. 19c). The typical equation of an indirect response model modified for characterization of an irreversible effect is as follows:

$$\frac{dR}{dt} = k_{\text{in}} - k_{\text{out}} \cdot R - k \cdot C_{\text{p}} \cdot R \tag{61}$$

Acetyl salicylic acid irreversibly inhibits platelet aggregation, and the turnover time of platelets is 7–10 days; thus, a small amount of aspirin shows a long duration of the antiplatelet effect. Using an irreversible effect model, Yamamoto et al. [40] explained the long duration of the antiplatelet effect of acetyl salicylic acid. The extension of this model with addition of precursor compartment was used to explain the prolonged duration of the effect of omeprazole [41].

### 4.5 Tolerance Models

Repeated or continuous exposure to several drugs such as glyceryl trinitrate, morphine, dobutamine, nicotine, cocaine, and benzodiazepines leads to clockwise hysteresis due to pharmacological tolerance [42]. Tolerance may occur due to desensitization, downregulation of receptor, or depletion of precursor. In the $E_{\max}$ model, desensitization and downregulation of drug receptors may be characterized by increasing $EC_{50}$ and decreasing $E_{\max}$ as functions of time ($EC_{50} \cdot e^{\text{kt}}$, $E_{\max} \cdot e^{-\text{kt}}$) [43]. In another mathematical model of tolerance, Porchet et al. [44] proposed the time-dependent development of a hypothetical antagonist to the effects of nicotine (Fig. 21a). Wakelkamp et al. [45] employed an indirect response model to describe acute tolerance to furosemide. In this model, the observed response was used to drive the production of a "modulator," which antagonized the effect of the drug (Fig. 21b). Equations describing this model are

$$\frac{dR}{dt} = k_{\text{in}} - k_{\text{out}} \cdot \left(1 - \frac{I_{\max} \cdot ER}{IC_{50} + ER}\right) \cdot R \cdot (1 + M) \tag{62}$$

$$\frac{dM}{dt} = k_{\text{tol}} \cdot R - k_{\text{tol}} \cdot M \tag{63}$$

where $M$ represents the modulator response and $k_{\text{tol}}$ represents the first-order rate constant of loss and production of modulator responses. Among other models of tolerance is an adaptive pool model presented by Gardmark et al. [46] to describe morphine tolerance.

## 5 CONCLUSION

The relationship between the magnitude and time course of pharmacological response and the time course of drug concentrations is often complex and may be impacted by the time course of drug diffusion to the biophase, the interaction of

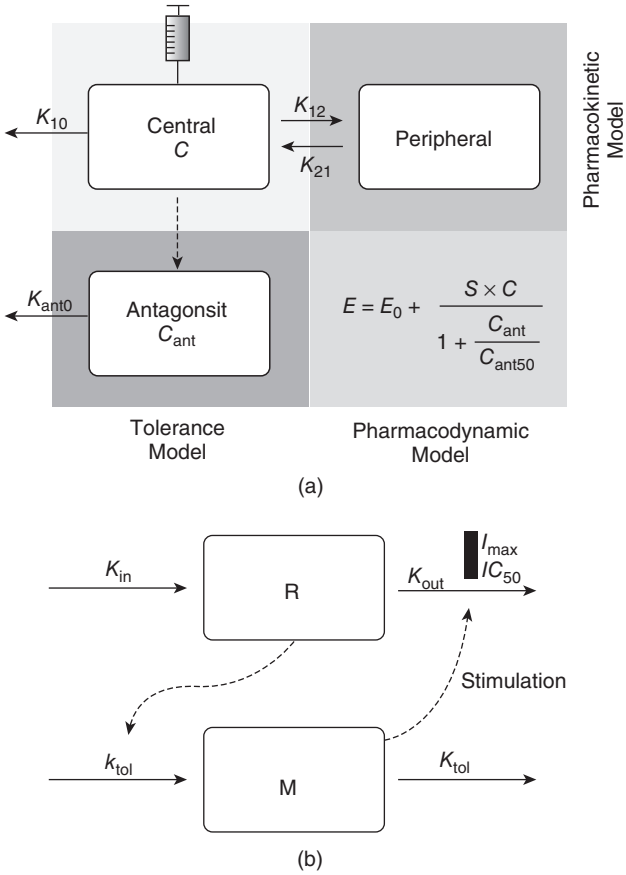


**FIGURE 21** Representative tolerance models. **(a)** Model for acute tolerance to nicotine effect. Tolerance is generated due to the generation of a hypothetical agent that antagonizes drug effect. (Adapted from Ref. 44.) **(b)** Application of an indirect response model to describe tolerance to furosemide. Tolerance results from generation of a modulator that counteracts the effect of furosemide. (Adapted from Ref. 45.)

drug with endogenous mediators of effect, transduction or signaling events, and the development of pharmacological tolerance or sensitization. The main sequences of drug response are summarized in Fig. 22 [47]. As the field of PKPD modeling has progressed, there has been movement from empirical models to models that may be deemed to be more mechanistically appropriate. There is promise that mechanism-based PKPD modeling may accelerate efforts to enhance the therapeutic efficacy and safety of drug therapy, lead to rapid identification of optimal dosing strategies, and assist in testing hypotheses related to the mechanisms responsible for drug action. The catalog of models presented in this chapter offers a starting point for readers interested in using pharmacodynamic analyses to achieve these goals.

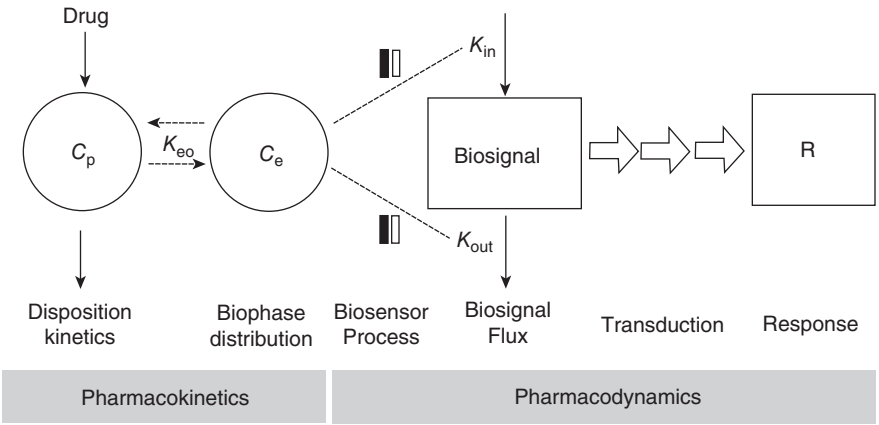


**FIGURE 22** Pharmacokinetics and pharmacodynamics. This scheme summarizes many of the processes associated with the relationship between the time course of drug administration and the time course of drug response. The time course of response is potentially impacted by drug disposition, drug diffusion to the biophase, interaction of drug with processes associated with the production and elimination of endogenous mediators of response, and signaling pathways associated with the measured drug response. (Adapted from Ref. 47.)

## ACKNOWLEDGMENTS

Work on this chapter was supported by Grants AI60687 and CA118213 from the National Institutes of Health. Dr. Beom Soo Shin is supported by a fellowship provided by Pfizer, Inc.

## REFERENCES

1. Jusko WJ. In *Course on Pharmacokinetic–Pharmacodynamic Modeling Concepts and Applications*. Buffalo, NY: The State University of New York; 2006.
2. Clark AJ. The reaction between acetyl choline and muscle cells. *J Physiol* 1926;61: 530–546.
3. Ariens EJ, De Groot WM. Affinity and intrinsic-activity in the theory of competitive inhibition. III. Homologous decamethonium-derivatives and succinyl-choline-esters. *Arch Int Pharmacodyn Ther* 1954;99:193–205.
4. Stephenson RP. A modification of receptor theory. *Br J Pharmacol Chemother* 1956;11: 379–393.
5. Levy G. In *Course of Pharmacokinetic–Pharmacodynamic Modeling: Concepts and Applications*. Buffalo, NY: The State University of New York; 1995.
6. McDevitt DG, Shand DG. Plasma concentrations and the time-course of beta blockade due to propranolol. *Clin Pharmacol Ther* 1975;18:708–713.
7. Nagashima R, O'Reilly RA, Levy G. Kinetics of pharmacologic effects in man: the anticoagulant action of warfarin. *Clin Pharmacol Ther* 1969;10:22–35.
8. Hill AV. The possible effects of the aggregation of the molecules of hemoglobin on its dissociation curves. *J Physiol (London)* 1910;61:530–546.

9. Wagner JG. Kinetics of pharmacologic response. I. Proposed relationships between response and drug concentration in the intact animal and man. *J Theor Biol* 1968;20: 173–201.
10. Hoffman A, Goldberg A. The relationship between receptor–effector unit heterogeneity and the shape of the concentration–effect profile: pharmacodynamic implications. *J Pharmacokinet Biopharm* 1994;22:449–468.
11. Levy G. Relationship between elimination rate of drugs and rate of decline of their pharmacologic effects. *J Pharm Sci* 1964;53:342–343.
12. Levy G. Kinetics of pharmacologic effects. *Clin Pharmacol Ther* 1966;7:362–372.
13. Levy G. Pharmacokinetics of succinylcholine in newborns. *Anesthesiology* 1970;32: 551–552.
14. Alvan G, Paintaud G, Wakelkamp M. The efficiency concept in pharmacodynamics. *Clin Pharmacokinet* 1999;36:375–389.
15. Lalonde RL, Pieper JA, Straka RJ, Bottorff MB, Mirvis DM. Propranolol pharmacokinetics and pharmacodynamics after single doses and at steady-state. *Eur J Clin Pharmacol* 1987;33:315–318.
16. Girard P, Boissel JP. Clockwise hysteresis or proteresis. *J Pharmacokinet Biopharm* 1989;17:401–402.
17. Furchgott RF. The pharmacology of vascular smooth muscle. *Pharmacol Rev* 1955;7: 183–265.
18. Sheiner LB, Stanski DR, Vozeh S, Miller RD, Ham J. Simultaneous modeling of pharmacokinetics and pharmacodynamics: application to *d*-tubocurarine. *Clin Pharmacol Ther* 1979;25:358–371.
19. Kramer WG, Kolibash AJ, Lewis RP, Bathala MS, Visconti JA, Reuning RH. Pharmacokinetics of digoxin: relationship between response intensity and predicted compartmental drug levels in man. *J Pharmacokinet Biopharm* 1979;7:47–61.
20. Wagner JG, Aghajanian GK, Bing OH. Correlation of performance test scores with "tissue concentration" of lysergic acid diethylamide in human subjects. *Clin Pharmacol Ther* 1968;9:635–638.
21. Hammarlund-Udenaes M, Paalzow LK, De Lange EC. Drug equilibration across the blood–brain barrier—pharmacokinetic considerations based on the microdialysis method. *Pharm Res* 1997;14:128–134.
22. Collins JM, Dedrick RL. Distributed model for drug delivery to CSF and brain tissue. *Am J Physiol* 1983;245:R303–R310.
23. Wang Y, Welty DF. The simultaneous estimation of the influx and efflux blood–brain barrier permeabilities of gabapentin using a microdialysis-pharmacokinetic approach. *Pharm Res* 1996;13:398–403.
24. Gupta SK, Ellinwood EH, Nikaido AM, Heatherly DG. Simultaneous modeling of the pharmacokinetic and pharmacodynamic properties of benzodiazepines. I: Lorazepam. *J Pharmacokinet Biopharm* 1990;18:89–102.
25. Ariens EJ. *Molecular Pharmacology*. New York: Academic Press; 1964.
26. Whitlon DS, Sadowski JA, Suttie JW. Mechanism of coumarin action: significance of vitamin K epoxide reductase inhibition. *Biochemistry* 1978;17:1371–1377.
27. Jusko WJ, Ko HC. Physiologic indirect response models characterize diverse types of pharmacodynamic effects. *Clin Pharmacol Ther* 1994;56:406–419.
28. Sharma A, Jusko WJ. Characteristics of indirect pharmacodynamic models and applications to clinical drug responses. *Br J Clin Pharmacol* 1998;45:229–239.
29. Boudinot FD, D'Ambrosio R, Jusko WJ. Receptor-mediated pharmacodynamics of prednisolone in the rat. *J Pharmacokinet Biopharm* 1986;14:469–493.

30. Mager DE, Jusko WJ. Pharmacodynamic modeling of time-dependent transduction systems. *Clin Pharmacol Ther* 2001;70:210–216.
31. Sun YN, Jusko WJ. Transit compartments versus gamma distribution function to model signal transduction processes in pharmacodynamics. *J Pharm Sci* 1998;87:732–737.
32. Oosterhuis B, Braat MC, Roos CM, Wemer J, Van Boxtel CJ. Pharmacokinetic–pharmacodynamic Modeling of Terbutaline bronchodilation in asthma. *Clin Pharmacol Ther* 1986;40:469–475.
33. Nichols AI, Boudinot FD, Jusko WJ. Second generation model for prednisolone pharmacodynamics in the rat. *J Pharmacokinet Biopharm* 1989;17:209–227.
34. Sun YN, DuBois DC, Almon RR, Jusko WJ. Fourth-generation model for corticosteroid pharmacodynamics: a model for methylprednisolone effects on receptor/gene-mediated glucocorticoid receptor down-regulation and tyrosine aminotransferase induction in rat liver. *J Pharmacokinet Biopharm* 1998;26:289–317.
35. Xu ZX, Sun YN, DuBois DC, Almon RR, Jusko WJ. Third-generation model for corticosteroid pharmacodynamics: roles of glucocorticoid receptor mRNA and tyrosine aminotransferase mRNA in rat liver. *J Pharmacokinet Biopharm* 1995;23:163–181.
36. Jusko WJ. Pharmacodynamics of chemotherapeutic effects: dose–time-response relationships for phase-nonspecific agents. *J Pharm Sci* 1971;60:892–895.
37. Zhi JG, Nightingale CH, Quintiliani R. Microbial pharmacodynamics of piperacillin in neutropenic mice of systematic infection due to *Pseudomonas aeruginosa*. *J Pharmacokinet Biopharm* 1988;16:355–375.
38. Jusko WJ. A pharmacodynamic model for cell-cycle-specific chemotherapeutic agents *J Pharmacokinet Biopharm* 1973;1:175–200.
39. Yano Y, Oguma T, Nagata H, Sasaki S. Application of logistic growth model to pharmacodynamic analysis of *in vitro* bactericidal kinetics. *J Pharm Sci* 1988;87:1177–1183.
40. Yamamoto K, Abe M, Katashima M, Sawada Y. Pharmacodynamic analysis of antiplatelet effect of aspirin in literature—modeling based on inhibition of cyclooxygenase in the platelet and the vessel wall endothelium. *Jpn J Hosp Pharm* 1996;22:113–141.
41. Abelo A, Eriksson UG, Karlsson MO, Larsson H, Gabrielsson J. A turnover model of irreversible inhibition of gastric acid secretion by omeprazole in the dog. *J Pharmacol Exp Ther* 2000;295:662–669.
42. Gardmark M, Brynne L, Hammarlund-Udenaes M, Karlsson MO. Interchangeability and predictive performance of empirical tolerance models. *Clin Pharmacokinet* 1999;36:145–167.
43. Lalonde RL. In Burton ME, Shaw LM, Schentag JJ, Evans WE, Eds. *Applied Pharmacokinetics & Pharmacodynamics*. Baltimore: Lippincott Williams & Wilkins; 2006, pp 60–81.
44. Porchet HC, Benowitz NL, Sheiner LB. Pharmacodynamic model of tolerance: application to nicotine. *J Pharmacol Exp Ther* 1988;244:231–236.
45. Wakelkamp M, Alvan G, Gabrielsson J, Paintaud G. Pharmacodynamic modeling of furosemide tolerance after multiple intravenous administration. *Clin Pharmacol Ther* 1996;60:75–88.
46. Gardmark M, Karlsson MO, Jonsson F, Hammarlund-Udenaes M. Morphine-3-glucuronide has a minor effect on morphine antinociception. Pharmacodynamic modeling. *J Pharm Sci* 1998;87:813–820.
47. Jusko WJ, Ko HC, Ebling WF. Convergence of direct and indirect pharmacodynamic response models. *J Pharmacokinet Biopharm* 1995;23:5–8.

# PHYSIOLOGICALLY BASED PHARMACOKINETIC MODELING

Harvey J. Clewell III,[1] Micaela B. Reddy,[2] Thierry Lave,[3] and Melvin E. Andersen[1]

[1]*The Hamner Institutes for Health Sciences, Research Triangle Park, North Carolina*
[2]*Roche Palo Alto LLC, Palo Alto, California*
[3]*F. Hoffmann-La Roche Ltd., Basel, Switzerland*

**Contents**

1 Introduction
  1.1 Noncompartmental, Compartmental, and PBPK Modeling
  1.2 Modeling and Simulation Considerations in the Generation of Drug Metabolism and Pharmacokinetic (DMPK) Data
2 PBPK Modeling Fundamentals
  2.1 Elements of PBPK Model Development
  2.2 Elements of Model Structure
3 PBPK Model Implementation
  3.1 Model Parameterization
  3.2 Numerical Solution of Model Equations
  3.3 Model Evaluation
  3.4 Model Validation
4 Generic PBPK Models
  4.1 User-Friendly Software Packages Bringing PBPK Modeling to the Nonexpert
  4.2 Recent Advances in Generic PBPK Modeling Tools
5 Integrating Preclinical Data Using the PBPK Model
  5.1 Preclinical Data Useful for PBPK Modeling
  5.2 Lead Identification and Optimization Stages
  5.3 Clinical Candidate Selection
  5.4 Supporting Entry into Humans

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

6 Examples of the Use of PBPK Modeling in Drug Development
6.1 Methotrexate
6.2 All-*trans*-Retinoic Acid
6.3 Future Directions
7 Conclusions and Future Perspectives
Acknowledgments
References

## 1 INTRODUCTION

Pharmacokinetics is the quantitative study of factors that control the time course for absorption, distribution, metabolism, and excretion (ADME) of compounds within the body. Pharmacokinetic (PK) models provide sets of equations that simulate the time courses of compounds and their metabolites in various tissues throughout the body. The interest in PK modeling in pharmacology and toxicology arose from the need to relate internal concentrations of active compounds at their target sites with the doses of the compound given to an animal or human subject. The reason, of course, is a fundamental tenet in pharmacology or toxicology that both beneficial and adverse responses to compounds are related to the free concentrations of active compounds reaching target tissues rather than the amounts of compound at the site of absorption. The relationships between tissue dose and administered dose can be complex, especially in high dose toxicity testing studies, with multiple, repeated daily dosing, or when metabolism or toxicity at routes of entry alter uptake processes for various routes of exposure. PK models of all kinds are primarily a tool to assess internal dosimetry at target tissues for a wide range of exposure situations.

In the 1930s, Teorell [1, 2] provided a set of equations for uptake, distribution, and elimination of drugs from the body. These papers are rightly regarded as providing the first physiological model for drug distribution. However, computational methods were not available to solve the sets of equations at that time. Exact mathematical solutions for distribution of compounds in the body could only be obtained for simplified models in which the body was reduced to a small number of compartments that did not correspond directly with specific physiological compartments. Over the next thirty years, PK modeling focused on these simpler descriptions with exact solutions rather than on developing models more concordant with the structure and content of the biological system itself. These approaches are sometimes referred to as *data-based* compartmental modeling, since the work generally took the form of a detailed collection of time-course blood/excreta concentrations at various doses. Time-course curves were analyzed by assuming particular model structures and estimating a small number of model parameters by curve-fitting. In the earliest of these models, all processes for metabolism, distribution, and elimination were treated as first order (i.e., rates changed in direct proportion to the concentration of the chemical species). Two areas of concern that particularly

affected data-based compartmental PK modeling arose in the 1960s and early 1970s: (1) the saturation of elimination pathways and (2) the possibility that blood flow rather than metabolic capacity of an organ might limit clearance. Saturation led to models that were not first order, making it difficult to derive exact solutions to the sets of equations, while blood-flow-limited metabolism in an organ meant that the removal rate constant from a central compartment could not increase indefinitely as the metabolic capacity increased [3].

Scientists trained in chemical engineering and computational methods developed physiologically based pharmacokinetic (PBPK) models for chemotherapeutic compounds [4]. Many of these compounds are highly toxic and have therapeutic efficacy by being slightly more toxic to rapidly growing cells (the cancer cells) than to normal tissues. Initial successes with methotrexate [5] led to PBPK models for other compounds, including 5-fluorouracil [6] and cisplatin [7]. These seminal contributions showed the ease with which realistic descriptions of physiology and relevant pathways of metabolism could be incorporated into PBPK models and paved the way for their more extensive use.

Although PBPK modeling was initially developed in the pharmaceutical industry [8–10], its recent use has primarily been in the area of environmental risk assessment. In the pharmaceutical area, the use of PBPK models was limited in drug development mainly due to the perceived mathematical complexity of the models and the labor-intensive input data required. However, advances in the prediction of hepatic metabolism and tissue distribution from *in vitro* and *in silico* data have made the use of these models more attractive, providing the opportunity to integrate key input data from different sources to not only estimate PK parameters and predict plasma and tissue concentration–time profiles, but also to gain mechanistic insight into the compound's properties. Thus, there is recent evidence of growing interest in applying PBPK models for the discovery and development of drugs [11, 12]. Strategies for PBPK application in drug research are being published [13, 14] and several commercial PBPK simulation tools have recently become available (e.g., GastroPlus™ and PK-Sim®, to name a few).

Some of the inertia to developing PBPK models was also due to the idea that extrapolations were unnecessary since human PK data would eventually be developed in clinical studies. This viewpoint neglects other attributes inherent in PBPK approaches. Some of the capabilities offered by PBPK approaches that have driven their extensive use in environmental risk assessment are: (1) creating models from physiological, biochemical, and anatomical information, entirely separate from collection of detailed concentration time-course curves; (2) evaluating mechanisms by which biological processes govern disposition of a wide range of compounds by comparison of PK results with model predictions; (3) using compounds as probes of the biological processes to gain more general information on the way biochemical characteristics govern the importance of various transport pathways in the body; (4) applying the models in safety assessments; and (5) using annotation of a modeling database as a repository of information on toxicity and kinetics of specific compounds [8]. Due to these perceived advantages, PBPK modeling was adopted as a tool for toxicological investigations and risk assessment in the early 1980s and many advances were made in this area. Recently, PBPK modeling has received new attention from the pharmaceutical industry. Lupfert and Reichel [11] recently wrote:

"The stage is set for a wide penetration of PBPK modeling and simulations to form an intrinsic part of a project starting from lead discovery, to lead optimization and candidate selection, to preclinical profiling and clinical trials."

Here, we present fundamental information necessary to understanding the application of PBPK models in preclinical drug development. Next, we describe the emerging role of PBPK modeling in the pharmaceutical industry in terms of advances in the available tools and increasingly used applications. Additionally, we provide our thoughts on potential applications that have not yet been widely explored. Although some advances have been made in applying PBPK modeling in the development of biotherapeutics (e.g., PBPK modeling for monoclonal antibodies in the mouse [15] and human [16]), this chapter is primarily geared toward small molecules.

## 1.1 Noncompartmental, Compartmental, and PBPK Modeling

Noncompartmental analysis (NCA) is widely used in the pharmaceutical industry for the analysis of single-dose PK (SDPK) data and PK data generated during safety studies. NCA, which involves calculating parameters such as steady-state volume of distribution ($V_{ss}$), clearance ($CL$), area under the plasma concentration curve ($AUC$), and peak plasma concentration ($C_{max}$) based solely on experimental data (i.e., these calculations are model independent), is a convenient way to understand, tabulate, and compare the PK properties of compounds. NCA results are easy to understand and a standard analysis technique. However, NCA results cannot be used to perform simulations or extrapolate to other exposure conditions. Although it is possible to use NCA parameters in a one-compartment model for performing simulations, this method is not likely to result in accurate predictions of pharmacokinetics.

Classical PK models are less commonly used than NCA calculations, but are used to perform simulations (e.g., use data generated for one experimental design to simulate a different experimental design). For example, a compartment model parameterized with intravenous (IV) SDPK data can be used to simulate an IV infusion study, and compartment modeling can be used to predict steady-state pharmacokinetics from SDPK data. Standard software packages often provide flexible tools for developing one-, two-, and three-compartment models from time-course plasma concentrations and provide statistical tools for discriminating between models. Compartment models require more mathematical skill than NCA but are still easy to use. However, the parameters and compartments in compartment models have no physiological meaning and provide no basis for extrapolating from the conditions of the experimental data used to develop the model.

PBPK models differ from classical PK models in that they include specific compartments for tissues involved in exposure, toxicity, biotransformation, and clearance processes connected by blood flow; compartments and blood flows are described using physiologically meaningful parameters, which allows for interspecies extrapolation by altering the physiological parameters appropriately. As with compartment models, PBPK models can be used to simulate a variety of conditions. Because the models have a mechanistic basis, extrapolation to situations differing from the conditions of the data used to calibrate the model is justifiable [17]. PBPK modeling has been used to great effect for interspecies extrapolation, both among animal

models [18] and for predicting human pharmacokinetics based on animal data [13, 19]. The mechanistic basis of PBPK models allows for applications such as understanding species differences in pharmacokinetics, determining if results from different experimental designs are consistent, and exploring possible mechanisms responsible for unexpected or unusual data. These attributes have led to widespread development of PBPK models in recent years [20].

Possibly the primary reason that PBPK models fell out of favor in the pharmaceutical industry is that PBPK models require more than the plasma time-course concentrations that are required by NCA and compartment models. PBPK models require physiological, physicochemical, and biochemical parameters. The physiological, mechanistic basis of the models is both their strength (the mechanistic basis provides exception utility) and their weakness (PBPK models can be expensive and time consuming to construct). However, recent contributions to the literature have demonstrated the effective application of PBPK models using the data normally generated during preclinical development [13, 14, 21, 22]. While NCA and compartment models will always be useful for many routine applications, PBPK modeling can be used for many situations when NCA and compartmental modeling simply cannot be used.

### 1.2 Modeling and Simulation Considerations in the Generation of Drug Metabolism and Pharmacokinetic (DMPK) Data

In preclinical drug development, an array of data is generated to investigate the potential of prospective drug candidates to have drug-like ADME properties in humans (Table 1). Preliminary screens tend to have high throughput and quick turnaround times, but less physiological relevance. As increasingly promising compounds are identified, ADME assays with more physiological relevance, but lower throughput and higher cost, are utilized. In the past, the conceptual model underlying preclinical development could be described as a filtering approach. The medicinal chemists would generate many candidates, and as compounds went through an array of screens for pharmacological, DMPK, and toxicological properties, compounds would be slowly eliminated from consideration until only the best compounds (clinical candidates) remained.

More recently, a different paradigm analogous to a cycle has been adopted. Medicinal chemists generate compounds that are tested for pharmacological activity. In parallel or in series, key ADME screening assays are employed to test for known liabilities of a class of compounds. For example, typical primary screens might be for permeability or metabolic stability, but should drug–drug interactions (DDIs) be a key liability for a given chemical class, then the appropriate DDI screen might be chosen as a primary screen, for example, to avoid the key liability of inhibition and/or induction of CYP3A4 by protease inhibitors for HIV [23]. Screening data is fed back to chemists, who go on to design better molecules based on assay results. The best compounds make it through the screening process to undergo more physiologically relevant, more expensive screens. Regardless of the screening cascade determined to be the best choice for a specific pharmacological target, much of the data generated is useful input for PBPK models. However, to ensure that PBPK modeling can be used to best advantage, a modeling strategy should be considered when making data generation choices.

**TABLE 1 Common Screening Tools Used to Understand ADME Properties in Preclinical Development**

| Absorption | Distribution | Metabolism | Excretion |
|---|---|---|---|
| Physicochemical properties (calculated and measured) | Protein binding | Stability in biological fluids | Urinary excretion |
| Solubility | Equilibrium dialysis | Plasma, blood stability | *In vitro* MDCK and HEK |
| Log *P* | Ultracentrifugation | Gut metabolism | *In situ* isolated, perfused rat kidney |
| Log *D* | Ultrafiltration | Caco-2 stability | *In vivo* urine measurement |
| p$K_a$ | Distribution in blood | Intestinal microsomes | Biliary excretion |
| Polar surface area | Blood-to-plasma ratio | Bile, gut fluids | *In vitro* hepatocytes |
| Permeability | CNS penetration | Hepatic metabolism | *In situ* isolated, perfused rat liver |
| Artificial membrane | Permeability assays | Liver microsomes | *In vivo* bile duct cannulated rats or dogs |
| PAMPA | *In vitro* Caco-2 | Hepatocytes | Transporter assays |
| *In vitro* models | Transporter assays | Liver S-9, cytosol, slices | |
| Caco-2 cell line | *In vivo* time-course tissue concentrations, microdialysis, mdr1a +/+, –/– mice | Metabolic pathway | |
| MDCK cell line | *In vivo* PK | Metabolite ID *in vitro, in vivo*, and *in situ* | |
| $P_{eff}$ with P-gp inhibited | Volume of distribution | Reaction phenotyping | |
| *In vivo* models | | Recombinant CYPs | |
| Bioavailability | | Liver microsomes with P450 inhibitor | |
| Portal vein cannulated animals | | Inhibition | |
| Transporter assays | | CYP, UGT | |
| Cell lines | | Induction | |
| Oocytes | | PXR | |
| | | Hepatocyte RNA, activity | |

The integration of preclinical data is a key to selecting good compounds for clinical development. The most common method for selecting compounds is by eliminating the compounds that fail to meet criteria for DMPK screens, and then picking the best candidate from the remaining compounds based on efficacy and other concerns. But take, for example, a choice between the following compounds: compound A has excellent efficacy in a rat pharmacology bioassay and good bioavailability, but has much higher protein binding in humans than rats, while compound B has good efficacy and good bioavailability, but has protein binding in humans similar to that of rats. Which compound should be chosen to move forward? We suggest integrating all available preclinical data using PBPK modeling, and whenever feasible integrating pharmacology data to develop the PK/PD (PK/pharmacodynamic) relationship, to come up with the most accurate, quantitative interpretation of the dataset and, therefore, the most science-driven selection of the clinical candidate.

## 2 PBPK MODELING FUNDAMENTALS

The basic approach for the development of a PBPK description is illustrated in Fig. 1. The process of model development begins with the identification of the compound and effect of concern. Literature evaluation involves the integration of available information about mechanisms of efficacy and/or toxicity, pathways of biochemical metabolism, the nature of the active compound processes involved in absorption, transport, and excretion, tissue partitioning and binding characteristics, and physiological parameters (e.g., tissue weights and blood flow rates) for the species of interest. Using this information, the investigator develops a PBPK model that expresses mathematically a conception of the animal/chemical system. In the

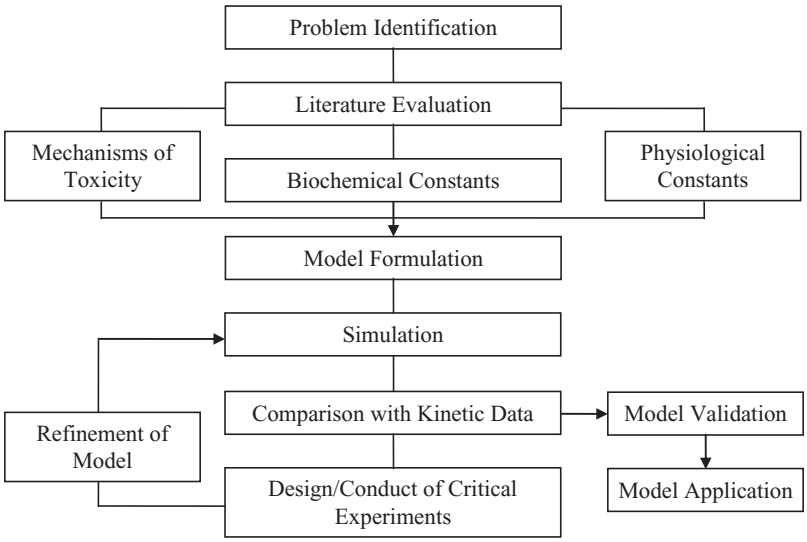


**FIGURE 1** Flowchart of the PBPK modeling process.

model, the various time-dependent biological processes are described as a system of simultaneous differential equations. A mathematical model of this form can easily be written and exercised using commonly available computer software. The specific structure of the model is driven by the need to estimate the appropriate measure of tissue dose under the various exposure conditions of concern in both the experimental animal and the human. Before the model can be used with confidence, it has to be validated against kinetic information and, in many cases, refined based on comparison with the experimental results. The model itself can frequently be used to help design critical experiments to collect data needed for its own validation.

The chief advantage of a PBPK model over an empirical description is its greater predictive power. Since known physiological parameters are used, a different species can be modeled by simply replacing the appropriate constants with those for the species of interest, or by allometric scaling [18, 24, 25]. Similarly, the behavior for a different route of administration can be determined by adding equations that describe the nature of the new input function.

Since measured physicochemical and biochemical parameters are used, the behavior for a different compound can quickly be estimated by determining the appropriate constants. An important result is the ability to reduce the need for extensive experiments with new compounds. The process of selecting the most informative experimental data is also facilitated by the availability of a predictive PK model. Perhaps the most desirable feature of a physiologically based model is that it provides a conceptual framework for employing the scientific method in which the hypothesis can be described in terms of biological processes, predictions can be made on the basis of the description, and the hypothesis can be revised on the basis of comparison with experimental data.

The trade-off against the greater predictive capability of physiologically based models is the requirement for an increased number of parameters and equations. However, values for many of the parameters, particularly the physiological ones, are already available in the literature [4, 26–30], and *in vitro* techniques have been developed for rapidly determining the compound-specific parameters [31]. An important advantage of PBPK models is that they provide a biologically meaningful quantitative framework within which *in vitro* data can be more effectively utilized [32]. There is even a prospect that predictive PBPK models can someday be developed based almost entirely on data obtained from *in vitro* studies.

## 2.1 Elements of PBPK Model Development

This section explores some of the key issues associated with the development of PBPK models. It is meant to provide a general understanding of the basic design concepts and mathematical forms underlying the PBPK modeling process and is not meant to be a complete exposition of the PBPK modeling approach for all possible cases. It must be understood that the specifics of the approach can vary greatly for different types of compounds and for different applications.

Model building is an art and is best understood as an iterative process in the spirit of the scientific method [33–35]. A number of excellent reviews have been written on the subject of PBPK modeling [20, 36, 37], and the literature includes examples of successful PBPK models for a wide variety of compounds that provide a wealth of insight into various aspects of the PBPK modeling process [20]. These

sources should be consulted for further detail on the approach for applying the PBPK methodology in specific cases.

***Tissue Grouping*** The first aspect of PBPK model development that will be discussed is determining the extent to which the various tissues in the body may be grouped together. Although tissue grouping is really just one aspect of model design, which is discussed in the next section, it provides a simple context for introducing the two alternative approaches to PBPK model development: *lumping* and *splitting*. In the context of tissue grouping, the guiding philosophy in the lumping approach can be stated as: "Tissues that are pharmacokinetically and toxicologically indistinguishable may be grouped together." In this approach, model development begins with information at the greatest level of detail that is practical, and decisions are made to combine physiological elements (tissues and blood flows) to the extent justified by their similarity [38]. The common grouping of tissues into richly (or rapidly) perfused and poorly (or slowly) perfused on the basis of their perfusion rate (ratio of blood flow to tissue volume) is an example of the lumping approach. The contrasting philosophy of splitting can be stated as: "Tissues that are pharmacokinetically or toxicologically distinct must be separated." This approach starts with the simplest reasonable model structure and increases the model's complexity only to the extent required to reproduce data on the compound of concern for the application of interest. Lumping (starting with a large number of compartments and then testing whether they can be combined) requires the greater initial investment in data collection and, if taken to the extreme, could paralyze model development. On the other hand, splitting (starting with a small number of compartments and only increasing complexity if the simple model fails) is more efficient but runs a greater risk of overlooking compound-specific determinants of disposition.

There are two alternative approaches for determining whether tissues are kinetically distinct or can be lumped together. In the first approach, the tissue rate constants are compared. The rate constant ($k_T$) for a tissue is similar to the perfusion rate except that the partitioning characteristics of the tissue are also considered:

$$k_T = Q_T/(P_T \cdot V_T)$$

where $Q_T$ is the blood flow to the tissue (L/h), $P_T$ is the tissue : blood partition coefficient for the compound, and $V_T$ is the volume of the tissue (L). Thus, the units of the tissue rate constant are the same as for the perfusion rate, L/h, but the rate constant more accurately reflects the kinetic characteristics of a tissue for a particular compound.

The second, less rigorous, approach for determining whether tissues should be lumped together is simply to compare the performance of the model with the tissues combined and separated. The reliability of this approach depends on the availability of data under conditions where the tissues being evaluated would be expected to have an observable impact on the kinetics of the compound. Sensitivity analysis can sometimes be used to determine the appropriate conditions for such a comparison [39].

***Model Design Criteria*** The alternative approaches to tissue grouping discussed previously are actually reflections of two competing criteria that must be balanced

during model design: *parsimony* and *plausibility*. The principle of parsimony simply states that a model should be as simple as possible for the intended application (but no simpler). This "splitting" philosophy is related to that of Occam's razor: "Entities should not be multiplied unnecessarily." That is, structures and parameters should not be included in the model unless they are needed to support the application for which the model is being designed.

There is no easy rule for determining the structure and level of complexity needed in a particular modeling application. The decision of which elements to include in the model structure for a specific compound and application draws on all of the modeler's experience and knowledge of the animal–chemical system. The wide variability of PBPK model design for different compounds can be seen by comparing the diagram of the PBPK model for methotrexate [5], shown in Fig. 2, with the diagram for the retinoic acid PBPK model [19], shown in Fig. 3.

The desire for parsimony in model development is driven not only by the desire to minimize the number of parameters whose values must be identified, but also by the recognition that as the number of parameters increases, the potential for unintended interactions between parameters also increases. A generally accepted rule of software engineering warns that it is relatively easy to design a computer program that is too complicated to be completely comprehended by the human mind. As a model becomes more complex, it becomes increasingly difficult to validate, even as the level of concern for the trustworthiness of the model should increase.

Countering the desire for model parsimony is the need for plausibility of the model structure. As discussed in Section 1, it is the physiological and biochemical realism of PBPK models that gives them an advantage for extrapolation. The

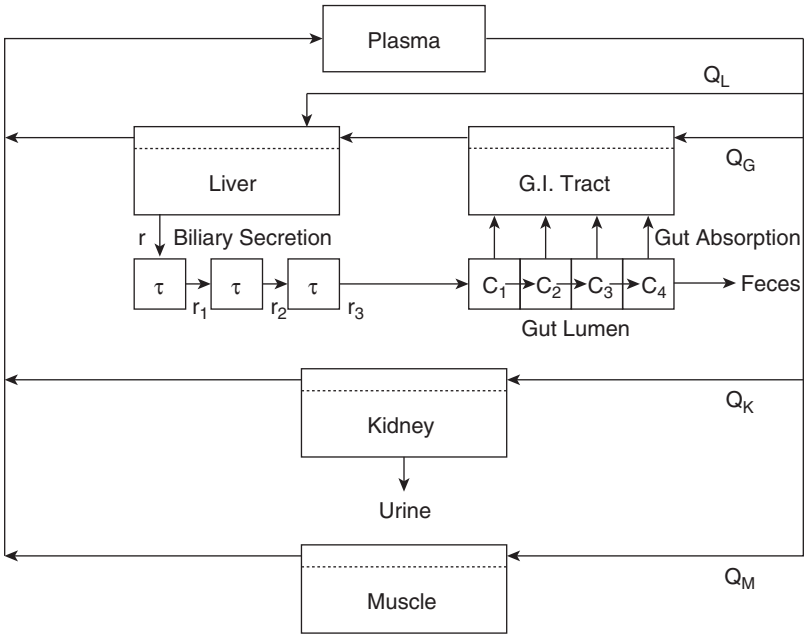


**FIGURE 2** PBPK model for methotrexate [5].

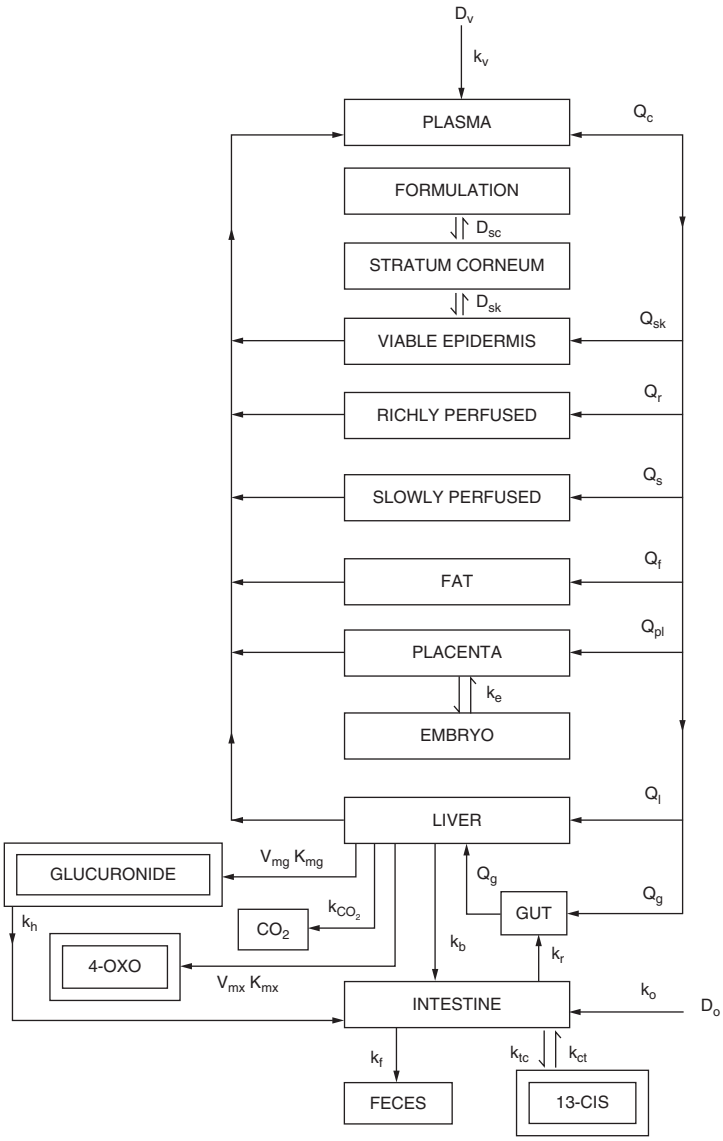


**FIGURE 3** PBPK model for all-*trans*-retinoic acid [19].

credibility of a PBPK model's predictions of kinetic behavior under conditions different from those under which the model was validated rests to a large extent on the correspondence of the model design to known physiological and biochemical structures. In general, the ability of a model to adequately simulate the behavior of a physical system depends on the extent to which the model structure is homomorphic (having a one-to-one correspondence) with the essential features determining the behavior of that system.

***Model Identification*** The process of model identification begins with the selection of those model elements that the modeler considers to be minimum essential determinants of the behavior of the particular animal–chemical system under study, from the viewpoint of the intended application of the model. Comparison with appropriate data, relevant to the intended purpose of the model, can then provide insights into defects in the model that must be corrected either by reparameterization or by changes to the model structure. Unfortunately, it is not always possible to separate these two elements. In models of biological systems, estimates of the values of model parameters will always be uncertain, due both to biological variation and experimental error. At the same time, the need for biological realism unavoidably results in models that are "overparameterized"; that is, they contain more parameters than can be identified from the kinetic data the model is used to describe.

Model identification is the selection of a specific model structure from several alternatives, based on conformity of the models' predictions to experimental observations. The practical reality of model identification in the case of biological systems is that regardless of the complexity of the model there will always be some level of "model error" (lack of homomorphism), which will result in systematic discrepancies between the model and experimental data. This model structural deficiency interacts with deficiencies in the identifiability of the model parameters, potentially leading to misidentification of the parameters or misspecification of structures. This most dangerous aspect of model identification is exacerbated by the fact that, in general, adding equations and parameters to a model increases the model's degrees of freedom, improving its ability to reproduce data, regardless of the validity of the underlying structure. Therefore, when a particular model structure improves the agreement of the model with kinetic data, it can only be said that the model structure is "consistent" with the kinetic data; it cannot be said that the model structure has been "proved" by its consistency with the data. In such circumstances, it is imperative that the physiological or biochemical hypothesis underlying the model structure is tested using nonkinetic data.

## 2.2 Elements of Model Structure

The process of selecting a model structure can be broken down into a number of elements associated with the different aspects of uptake, distribution, metabolism, and elimination. This section treats each of these elements in turn.

***Storage Compartments*** Naturally, any tissues that are expected to accumulate significant quantities of the compound or its metabolites need to be included in the model structure. As discussed earlier, these storage tissues can be grouped together to the extent that they have similar time constants. The muscle tissue in the methotrexate model (Fig. 2) is an example of a storage compartment. The generic mass balance equation for storage compartments such as these is

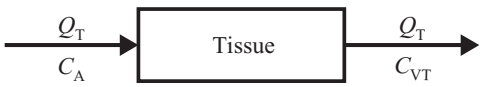


$$dA_T/dt = Q_T \cdot C_A - Q_T \cdot C_{VT}$$

where $A_T$ is the mass of compound in the tissue (mg), $C_A$ is the concentration of compound in the arterial blood reaching the tissue (mg/L), and $C_{VT}$ is the concentration of compound in the venous blood leaving the tissue (mg/L). Thus, this mass balance equation simply states that the rate of change in the amount of compound in the tissue with respect to time ($dA_T/dt$) is equal to the difference between the rate at which compound enters the tissue and the rate at which compound leaves the tissue. We can then calculate the concentration of the compound in the storage tissue ($C_T$) from the amount in the tissue and the tissue volume ($V_T$):

$$C_T = A_T/V_T$$

In PBPK models, it is common to assume *venous equilibration*; that is, that in the time that it takes for the blood to perfuse the tissue, the compound is able to achieve its equilibrium distribution between the tissue and blood. Therefore, the concentration of the compound in the venous blood can be related to the concentration in the tissue by the equilibrium tissue : blood partition coefficient, $P_T$:

$$C_{VT} = C_T/P_T$$

Therefore, we obtain a differential equation in $A_T$:

$$dA_T/dt = Q_T \cdot C_A - Q_T \cdot A_T/(P_T \cdot V_T)$$

If desired, we can reformulate this mass balance equation in terms of concentration:

$$dA_T/dt = d(C_T \cdot V_T)/dt = C_T \cdot dV_T/dt + V_T \cdot dC_T/dt$$

If (and only if) $V_T$ is constant (i.e., the tissue does not grow during the simulation), $dV_T/dt = 0$, and

$$dA_T/dt = V_T \cdot dC_T/dt$$

so we have the alternative differential equation:

$$dC_T/dt = Q_T \cdot (C_A - C_T/P_T)/V_T$$

This alternative mass balance formulation, in terms of concentration rather than amount, is popular in the PK literature. However, in the case of models with compartments that change volume over time (e.g., in a model incorporating growth of a single or multiple tissues) it is preferable to use the formulation in terms of amounts in order to avoid the need for the additional term reflecting the change in volume ($C_T \cdot dV_T/dt$).

Depending on the compound, many different tissues can potentially serve as important storage compartments. The use of a fat storage compartment is typically required for lipophilic compounds. The gut lumen can also serve as a storage site for compounds subject to enterohepatic recirculation, as in the case of methotrexate. Important storage sites for metals, on the other hand, can include the kidney,

red blood cells, intestinal epithelial cells, skin, bone, and hair [40]. Transport to and from a storage compartment does not always occur via the blood, as was described previously; for example, in some cases the storage is an intermediate step in an excretion process (e.g., hair, intestinal epithelial cells). As with methotrexate, it may also be necessary to use multiple compartments in series, or other mathematical devices, to model plug flow (i.e., a time delay between entry and exit from storage).

***Blood Compartment*** The description of the blood compartment can vary considerably from one PBPK model to another depending on the role the blood plays in the kinetics of the compound being modeled. In some cases the blood may be treated as a simple storage compartment, with a mass balance equation describing the summation (Σ) of the venous blood flows from the various tissues and the return of the total arterial blood flow ($Q_C$) to the tissues, as well as any urinary clearance (if, as in the case of glomerular filtration, clearance is described as occurring from the blood compartment):

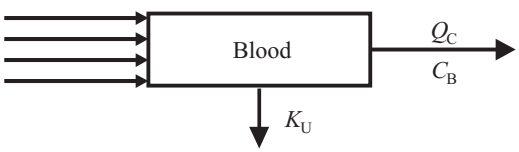


$$dA_B/dt = \sum(Q_T \cdot C_{VT}) - Q_C \cdot C_B - K_U \cdot C_B$$

where $A_B$ is the amount of compound in the blood (mg), $Q_C$ is the total cardiac output (L/h), $C_B$ is the concentration of compound in the blood (mg/L), and $K_U$ is the urinary clearance (L/h). The value for $K_U$ can often be estimated from the unbound fraction in plasma ($f_{ub}$) and the glomerular filtration rate (GFR), unless active transport processes contribute to renal elimination. An alternative method for estimating human renal clearance based on rat renal clearance and incorporating species-specific physiological differences in GFR has been proposed [13, 41].

Most often, drug concentrations are measured in plasma or serum and not in whole blood; hence, plasma or serum is the reference fluid for the derived PK parameters such as clearances and volumes of distribution. However, whole blood, not plasma or serum, is flowing through the vessels of the human body. Therefore, provided that there is evidence in support of fast equilibration of drug between red blood cells and plasma [42], whole blood rather than plasma is the more appropriate reference fluid for calculating and interpreting clearances and volumes of distribution. For this reason, parameterization of the PBPK model is often performed in terms of blood flows. To compare calculated plasma concentrations to experimental data, the calculated blood concentration must then be divided by the blood-to-plasma ratio (BPR), which is often measured experimentally.

In cases where a compound is not taken up by the red blood cells, plasma flow can be used in place of blood flow in the model [5]. For some compounds, where exchange between plasma and red blood cells is slow compared to tissue perfusion,

it may be necessary to model the red blood cells as a storage compartment in communication with the plasma via diffusion-limited transport. Typically, however, exchange between red blood cells and plasma is fast compared to tissue distribution, and the blood can be treated as a single compartment.

If the blood is an important storage compartment for a compound, it may be necessary to carefully evaluate data on tissue concentrations, particularly the richly perfused tissues, to determine whether compound in the blood perfusing the tissue could be contributing to the measured tissue concentration. For other compounds, the amount of compound actually in the blood may be relatively small, in which case only the concentration may be of interest. In this case, instead of having a true blood compartment, a steady-state approximation can be used to estimate the concentration in the blood at any time. Assuming the blood is at steady state with respect to the tissues,

$$dA_B/dt = 0$$

Therefore, solving the blood equation for the concentration:

$$C_B = \sum_T (Q_T \cdot C_{VT})/Q_C$$

***Metabolism/Elimination*** The liver is frequently the primary site of metabolism for a compound. The following equation is an example of the mass balance equation for the liver in the case of a compound that is metabolized through both saturable and nonsaturable components:

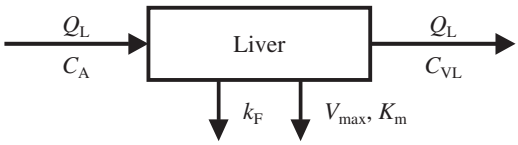


$$dA_L/dt = Q_L \cdot (C_A - C_{VL}) - k_F \cdot C_{Lfree} \cdot V_L - V_{max} \cdot C_{Lfree}/(K_m + C_{Lfree})$$

where $Q_L$ is the total blood flow (arterial and portal) to the liver and $C_{Lfree}$ is the free (unbound) concentration in the liver.

In the equation above, the first term on the right-hand side of the equation represents the mass flux associated with transport in the blood and is essentially identical to the case of the storage compartment described previously. In the case of the liver, however, it should be noted that there are two sources of blood flow: the arterial flow directly to the liver and the portal flow that first perfuses the intestines. When appropriate, these two blood flows can be differentiated in the model (Fig. 2 and 3).

The second term in the equation above describes metabolism by a linear (first-order) pathway with rate constant $k_F$ ($h^{-1}$) and the third term represents metabolism by a saturable (Michaelis–Menten) pathway with capacity $V_{max}$ (mg/h) and affinity $K_m$ (mg/L). If it were desired to model a water-soluble metabolite produced by this

saturable pathway, an equation for its formation and elimination could also be added to the model:

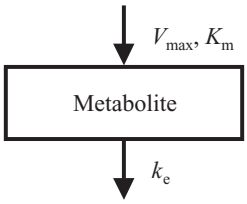


$$dA_M/dt = R_{stoch} \cdot V_{max} \cdot C_{Lfree}/(K_m + C_{LFree}) - k_e \cdot A_M$$

$$C_M = A_M/V_D$$

where $A_M$ is the amount of metabolite in the body (mg), $R_{stoch}$ is the stoichiometric fractional yield of the metabolite (unitless) (times the ratio of its molecular weight to that of the parent compound), $k_e$ is the rate constant for the clearance of the metabolite from the body ($h^{-1}$), $C_M$ is the concentration of the metabolite in the plasma (mg/L), and $V_D$ is the apparent volume of distribution for the metabolite (L).

The definition of free concentration in the liver ($C_{Lfree}$) in these equations is not as straightforward as it may at first appear. In cases where hepatic clearance is relatively low, the free concentration in the liver is often assumed to be equal to the free (unbound) concentration in the blood; that is, $C_{Lfree} = f_u \cdot C_B$. However, when hepatic clearance is high, the free concentration in the liver can be drawn well below the free concentration in the blood, and such an assumption may be inappropriate. An alternative assumption that is often made is that the free concentration in the liver can be estimated by dividing the total liver concentration by the liver:blood partition coefficient; that is, $C_{Lfree} = C_L/P_L$ [5]. This approximation is particularly useful for compounds whose metabolism is limited by hepatic blood flow at low concentrations. If the intent is to use *in vitro* estimates of the metabolic parameters ($V_{max}$, $K_m$, $k_f$) in the model, then the definition of the *in vivo* free concentration should be consistent with the conditions under which the *in vitro* estimates of intrinsic clearance were obtained. In principle, this could require adjusting for differences in binding between the *in vitro* medium and the tissue *in vivo*, although such adjustments are seldom performed in practice.

***Metabolite Compartments*** In principle, the same considerations that drive decisions regarding the level of complexity of the PBPK model for the parent compound must also be applied for each of its metabolites, and their metabolites, and so on. As in the case of the parent compound, the first and most important consideration is the purpose of the model. If the concern is a direct parent compound effect and the compound is inactivated by metabolism, then there is no need for a description of metabolism beyond its role in the clearance of the parent compound. The model for methotrexate (Fig. 2) is an example of a parent compound model.

If one or more of the metabolites are considered to contribute to the efficacy or toxicity of a compound, it may also be necessary to provide a more complete description of the kinetics of the metabolites themselves. For example, in the case

of teratogenicity from all-trans retinoic acid, both the parent compound and several of its metabolites are considered to be toxicologically active; therefore, in developing the PBPK model for this compound it was necessary to include a fairly complete description of the metabolic pathways [19]. Fortunately, the metabolism of xenobiotic compounds often produces metabolites that are relatively water soluble, simplifying the description needed. In many cases, a classical one-compartment description may be adequate for describing the metabolite kinetics. In other cases, however, the description of the metabolite (or metabolites) may have to be as complex as that of the parent compound. If reactive intermediates produced during the metabolism of a compound are responsible for toxicity, a very simple description of the metabolic pathways might be adequate [43].

***Target Tissues*** Typically, a PBPK model used in drug development applications will include compartments for any target tissues for the efficacy of the compound and could also include target tissues of toxicity if the model is to be used for assessing safety. The target tissue description may in some cases need to be fairly complicated, including such features as *in situ* metabolism, binding, and pharmacodynamic processes in order to provide a realistic measure of biologically effective tissue exposure.

A fundamental issue in determining the nature of the target tissue description required is the need to identify the active form of the compound. A compound may produce an effect directly, through its interaction with tissue constituents, or indirectly, through a metabolite. In the case of toxicity, it is often the metabolism of a compound that leads to undesired effects, either through the production of reactive metabolites during metabolism or due to the toxicity of a circulating metabolite.

The specific nature of the relationship between tissue exposure and response depends on the mechanism, or mode of action, involved. Rapidly reversible effects may result primarily from the current concentration of the compound in the tissue, while longer-term effects may depend on both the concentration and duration of the exposure. Although therapeutic effects are generally due to the administered compound, an active metabolite may extend efficacy or cause undesired side effects or toxicity. In general, the appropriate measure of tissue exposure for a toxic side effect of a compound may be different from the appropriate measure for its therapeutic effect. For example, the therapeutic effect of a drug may depend on the prolonged maintenance of a relatively low concentration sufficient to occupy a receptor in the target tissue, while its toxicity may result from transient, high rates of metabolism in the liver occurring shortly after dosing. In such a case, PBPK modeling of the concentration time course in the liver and target tissue for different dosing routes or regimens might help to maximize the therapeutic index for the compound. For a developmental toxicity, the concentration time course might also have to be convoluted with the window of susceptibility for a particular gestational event. The evaluation of the various modes of action for the beneficial and toxic effects of a compound is the most important step in a PK analysis and a principal determinant of the structure and level of detail that will be required in the PBPK model.

***Uptake Routes*** Each of the relevant uptake routes for the compound must be described in the model. Often there are a number of possible ways to describe a

particular uptake process, ranging from simple to complex. As with all other aspects of model design, the competing goals of parsimony and realism must be balanced in the selection of the level of complexity to be used. The following examples are meant to provide an idea of the variety of model code that can be required to describe the various possible uptake processes.

*Intravenous Administration (Bolus Dosing)*

$$A_{B0} = Dose \cdot BW$$

where $A_{B0}$ is the amount of compound in the blood at the time of dosing ($t = 0$), *Dose* is the administered dose (mg/kg), and *BW* is body weight (kg).

*Intravenous Administration (Infusion)* In the case where a steady-state approximation has been used to eliminate the blood compartment, we have

$$C_B = (Q_L \cdot C_{VL} + \cdots + Q_F \cdot C_{VF} + k_{IV})$$

where

$$k_{IV} = \begin{cases} Dose \cdot BW/t_{IV} & (t < t_{IV}) \\ 0 & (t > t_{IV}) \end{cases}$$

where $t_{IV}$ is the duration of time over which the infusion takes place (h). In this case, the model code must be written with a "switch" to change the value of $k_{IV}$ to zero at $t = t_{IV}$.

*Oral Gavage* For a compound that is completely absorbed in the stomach,

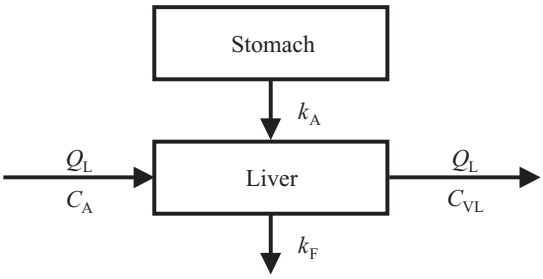


$$A_{ST0} = \text{Dose} \cdot \text{BW}$$

$$dA_{ST}/dt = -k_A \cdot A_{ST}$$

$$dA_L/dt = Q_L \cdot (C_A - C_L/P_L) - k_F \cdot C_L \cdot V_L/P_L + k_A \cdot A_{ST}$$

where $A_{ST0}$ is the amount of compound in the stomach at the beginning of the simulation, $A_{ST}$ is the amount of compound in the stomach at any given time, and $k_A$ is a first-order rate constant ($h^{-1}$) describing uptake from the stomach.

For a compound that is incompletely absorbed,

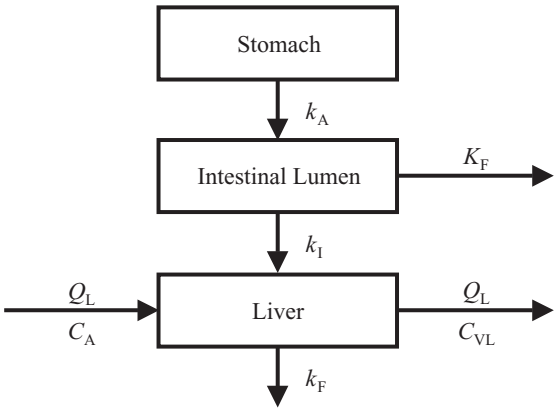


$$A_{ST0} = Dose \cdot BW$$

$$dA_{ST}/dt = -k_S \cdot A_{ST}$$

$$dA_I/dt = k_S \cdot A_{st} - k_I \cdot A_I - K_F \cdot A_I/V_I$$

$$dA_L/dt = Q_L \cdot (C_A - C_{VL}) - k_F \cdot C_{Lfree} \cdot V_L + k_I \cdot A_I$$

where $A_I$ is the amount of compound in the intestinal lumen (mg), $k_I$ is the rate constant for intestinal absorption ($h^{-1}$), $K_F$ is the fecal clearance (L/h), and $V_I$ is the volume of the intestinal lumen (L).

The rate of fecal excretion of the compound is then

$$dA_F/dt = K_F \cdot A_I/V_I$$

The examples described above are highly simplified descriptions. For example, they describe uptake from the stomach or intestinal lumen directly to the liver, when in fact the uptake is into the tissues of the gastrointestinal (GI) tract with subsequent transport to the liver in the portal blood. While this simplification is adequate in some cases, a more accurate description may be necessary in others, such as when metabolism in the gut tissues is important or portal blood flow could limit uptake. The equations shown above also do not describe biliary excretion or other processes that might be important determinants of the intestinal concentration and fecal excretion of a compound over time. Again, these processes would need to be included in the full model description for compounds where these processes are important [19]. Note also that this simple formulation does not consider the plug flow of the intestinal contents and will not reproduce the delay that actually occurs in the appearance of a compound in the feces. Such a delay could be added using a delay function available in common simulation software, or multiple compartments could be used to simulate plug flow, as shown in the diagram of the methotrexate model (Fig. 2).

Although in some cases a simple, first-order model can adequately describe oral absorption, more physiologically representative approaches have been developed and shown to be useful for predicting the fraction absorbed, for example, a

compartmental approach such as the ACAT model [44] and a model describing the GI tract as a tube [45]. Simulation of oral absorption using physiologically based models represents another challenge because of the many processes involved such as the release, dissolution, degradation, metabolism, uptake, and absorption of a compound as it transits through the different segments of the digestive tract [46]. Oral absorption simulations require *in vitro* and *in silico* input data such as solubility, permeability, particle size, partition coefficient ($\log P$), ionization constant ($pK_a$), and dose [47, 48].

*Inhalation* For compounds that are volatile, it is necessary to describe the exchange of vapor between the lung air and blood in the alveolar region. This is true even if the compound is not administered by inhalation, because exhalation can be an important route of clearance for volatile compounds regardless of the dose route.

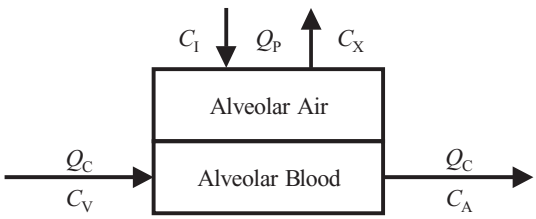


$$dA_{AB}/dt = Q_C \cdot (C_V - C_A) + Q_P \cdot (C_I - C_X)$$

where $A_{AB}$ is the amount of compound in the alveolar blood (mg), $C_V$ is the concentration of compound in the pooled venous blood (mg/L), $C_A$ is the concentration of compound in the alveolar (arterial) blood (mg/L), $Q_P$ is the alveolar (not total pulmonary) ventilation rate (L/h), $C_I$ is the concentration of compound in the inhaled air (mg/L), and $C_X$ is the concentration of compound in the alveolar air (mg/L). Assuming the alveolar blood is at steady state with respect to the other compartments,

$$dA_{AB}/dt = 0$$

Also, assuming lung equilibration (i.e., that the blood in the alveolar region has reached equilibrium with the alveolar air prior to exhalation),

$$C_X = C_A/P_B$$

where $P_B$ is the blood:air partition coefficient.

Substituting into the equation for the alveolar blood and solving for $C_A$, we have

$$C_A = (Q_C \cdot C_V + Q_P \cdot C_I)/(Q_C + Q_P/P_B)$$

Note that the rate of elimination of the compound by exhalation is just $Q_P \cdot C_X$. The alveolar ventilation rate, $Q_P$, does not include the "deadspace" volume (the portion of the inhaled air that does not reach the alveolar region) and is therefore roughly 70% of the total respiratory rate. The concentration $C_X$ represents the

"end-alveolar" air concentration; in order to estimate the average exhaled concentration ($C_{EX}$), the deadspace contribution must be included:

$$C_{EX} = 0.3C_I + 0.7C_X$$

PBPK models including more detailed physiological descriptions of inhalation exposures developed to understand toxicological effects of reactive vapors in the nasal cavity have been reviewed [49].

*Dermal* A simple model can be used to describe dermal absorption from a constant-concentration vehicle on the skin surface.

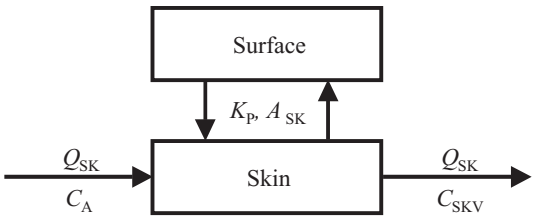


$$dA_{SK}/dt = K_P \cdot A_{SFC} \cdot (C_{SFC} - C_{SK}/P_{SKV})/1000 + Q_{SK} \cdot (C_A - C_{SK}/P_{SKB})$$

where $A_{Sk}$ is the amount of compound in the skin (mg), $K_P$ is the skin permeability coefficient (cm/h), $A_{SFC}$ is the skin surface area (cm$^2$), $C_{SFC}$ is the concentration of compound on the surface of the skin (mg/L), $C_{SK}$ is the concentration of the compound in the skin (mg/L), $P_{SK}$ is the skin:vehicle partition coefficient (i.e., the vehicle containing the compound on the surface of the skin), $Q_{SK}$ is the blood flow to the skin region (L/h), $C_A$ is the arterial concentration of the compound (mg/L), and $P_{SKB}$ is the skin:blood partition coefficient.

Due to adding this compartment, the equation for the blood in the model must also be modified to add a term for the venous blood returning from the skin ($+Q_{SK} \cdot C_{SK}/P_{SKB}$), and the blood flow and volume parameters for the slowly perfused tissue compartment must be reduced by the amount of blood flow and volume for the skin. Approaches used to include the dermal exposure route in PBPK models have been reviewed [50]. Several reviews discuss the wide variety of available compartment models and provide guidance for choosing among them [51, 52].

*Experimental Apparatus* In some cases, in addition to compartments describing the animal–chemical system, it may also be necessary to include model compartments that describe the experimental apparatus in which measurements were obtained.

***Distribution/Transport*** There are a number of issues associated with the description of the transport and distribution of the compound that must be considered in the process of model design. Discussions of the more common ones are included here.

*Diffusion-Limited Transport* Most of the PBPK models in the literature are flow-limited models; that is, they assume that the rate of tissue uptake of the compound is limited only by the flow of the compound to the tissue in the blood. While this assumption appears to be reasonable in general, for some compounds and tissues uptake may be diffusion limited. Examples of tissues for which diffusion-limited transport has often been described include the skin, placenta, brain, and fat. The model compartments described thus far have all assumed flow-limited transport. If there is evidence that the movement of a compound between the blood and a tissue is limited by diffusion, a two-compartment description of the tissue can be used with a "shallow" exchange compartment in communication with the blood and a diffusion-limited "deep" compartment that represents the actual tissue:

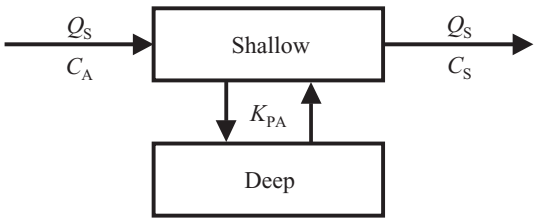


$$dA_S/dt = Q_S \cdot (C_A - C_S) - K_{PA} \cdot (C_S - C_D/P_D)$$
$$dA_D/dt = K_{PA} \cdot (C_S - C_D/P_D)$$

where $A_S$ is the amount of compound in the shallow compartment (mg), $Q_S$ is the blood flow to the shallow compartment (L/h), $C_S$ is the concentration of compound in the shallow compartment (mg/L), $K_{PA}$ is the permeability–area product for diffusion-limited transport (L/h), $C_D$ is the concentration of compound in the deep compartment (mg/L), $P_D$ is the tissue:blood partition coefficient, and $A_D$ is the amount of compound in the deep compartment (mg).

*Saturable Tissue Binding* When there is evidence that saturable binding is an important determinant of the distribution of a compound into a tissue (such as evidence of dose-dependent tissue partitioning), a simple description of tissue binding can be added to the model. In this description, only free (unbound) compound is considered to be available for transport or clearance at any given moment in time. For example, in the case of saturable binding in the liver,

$$dA_L/dt = V_L \; dC_L/dt = Q_L \cdot (C_A - C_{Lfree}) - k_F \cdot C_{Lfree}$$

where $A_L$ is the total (free plus bound) concentration of the compound in the liver (mg), $C_{Lfree}$ is the concentration of free (unbound) compound in the liver (mg/L), and $k_F$ is the rate constant for metabolism ($h^{-1}$).

The apparent complication in adding this equation to the model is that the change in the total amount of compound in the tissue ($dA_L/dt$) is needed for the mass balance, but the determinants of the kinetics are described in terms of the

free concentration in the tissue ($C_{\text{Lfree}}$). To solve for free in terms of total, we note that

$$C_{\text{L}} = C_{\text{Lfree}} + C_{\text{Lbound}}$$

where $C_{\text{Lbound}}$ is the concentration of bound compound in the tissue (mg/L). We can describe the saturable binding with an equation similar to that for saturable metabolism:

$$C_{\text{Lbound}} = B_{\text{m}} \cdot C_{\text{Lfree}}/(K_{\text{d}} + C_{\text{Lfree}})$$

where $B_{\text{m}}$ is the binding capacity (mg/L) and $K_{\text{d}}$ is the binding affinity (mg/L). Substituting this equation into the previous one, we have

$$C_{\text{L}} = C_{\text{Lfree}} + B_{\text{m}} \cdot C_{\text{Lfree}}/(K_{\text{d}} + C_{\text{Lfree}})$$

Rewriting this equation to solve for the free concentration in terms of only the total concentration would result in a quadratic equation, the solution of which could be obtained with the quadratic formula. However, taking advantage of the iterative algorithm by which these PBPK models are exercised (as will be discussed later), it is not necessary to go to this effort. Instead, a much simpler implicit equation can be written for the free concentration (i.e., an equation in which the free concentration appears on both sides):

$$C_{\text{Lfree}} = C_{\text{L}}/(1 + B_{\text{m}}/(K_{\text{d}} + C_{\text{Lfree}}))$$

In an iterative algorithm, this equation can be solved at each time step using the previous value of $C_{\text{Lfree}}$ to obtain the new value! A new value of $C_{\text{L}}$ is then obtained from the mass balance equation for the liver using the new value of $C_{\text{Lfree}}$, and the process is repeated. An underlying assumption for this simple description of tissue binding is that the binding is rapidly reversible compared to the movement or clearance of free compound, such that equilibrium between free and bound chemical can be maintained.

In fact, at least two different computational approaches have been used to describe saturable binding in PBPK models. The concentration of free compound can either be estimated, as described above, from solving a conservation equation for total mass that apportions the total amount of compound between free and bound forms using the equilibrium dissociation constants, $K_{\text{d}}$, and binding maxima, $B_{\text{m}}$ [53], or it can be estimated by explicitly including both "on" (association) and "off" (dissociation) rate constants, $k_{\text{a}}$ and $k_{\text{d}}$ [54]. In the latter case, the dissociation binding constant (with units of concentration) is the ratio of the two rate constants, $k_{\text{d}}/k_{\text{a}}$. The advantage of the rate constant approach is that it does not require the assumption that binding is fast compared to transport and clearance.

While these two computational approaches have usually been applied within tissues, they can also be readily applied to binding in blood if the concentrations and affinities of binding proteins are known. However, in most modeling of drugs, protein binding in blood is linear (not saturable at clinically relevant concentrations) and can be characterized by a single parameter, fraction unbound, rather

than estimates of concentrations of binding sites and their affinities for binding. Nevertheless, there are cases where protein binding is more appropriately described as a nonlinear process [55]. Alternative descriptions of blood binding will be discussed in more detail in the next section.

*Binding in Blood* Protein binding in blood can be a key determinant of drug disposition, affecting compound availability for interaction with therapeutic targets as well as for clearance. A high fraction bound in the blood also gives rise to concerns regarding potential drug–drug interactions that might displace bound drug and produce a transient increase in drug concentrations to potentially toxic levels. Methodologies to estimate drug binding and approaches for the quantitative description of this binding in PK models have been areas of intense interest over the past four to five decades. Consideration of blood binding faces two parallel challenges. First, when drugs are bound in capillary blood, what fraction should be regarded as available for transport into tissue? Second, how does the binding in blood influence blood : tissue partitioning? Historically, these questions have been addressed both in empirical descriptions and in PBPK models, and some care is required to reconcile the different approaches and arrive at a consistent quantitative treatment of blood binding and transport of compound throughout the body in both types of modeling approaches.

In the standard description of clearance of a compound from blood by tissue metabolism, binding in the blood is assumed to be linear and the fraction unbound, $f_u$, is simply multiplied by the intrinsic clearance, leading to a straightforward relationship:

$$CL = Q_T \cdot f_u \cdot CL_{int} / (Q_T + f_u \cdot CL_{int})$$

In this relationship, the maximal tissue clearance, even with a low fraction unbound, is total tissue blood flow. That is, all of the compound in the blood, whether bound or unbound, becomes available for clearance as long as the intrinsic clearance is sufficiently large. Note that in this description the fraction unbound is not a function of the clearance. In fact, the derivation of this equation rests on the assumption that dissociation of bound compound in the blood is fast compared to the rate of tissue clearance. If the uptake of a compound into the tissue is limited by the rate of dissociation of the compound from binding proteins in the blood, the simple formula above will overestimate its clearance.

In PBPK models, blood concentrations of bound and free compound can be described separately, as discussed in the previous section, and only free compound is generally considered to be available to participate in processes such as diffusion, metabolism, tissue reaction, and intercompartmental transfer. As in the case of tissue binding, the relationship of free to bound concentrations in the blood is generally calculated by knowledge of dissociation binding constants ($K_d$) and maximum concentrations of binding proteins ($B_m$), as describe previously. An example of this approach has been described in the case of a PBPK model for estradiol [55]. This approach differs from the simple clearance formula given above in that the fraction unbound can be a nonlinear function of the concentration of the compound in the blood. It shares, however, the assumption that the "on" and "off" rates for binding

are fast compared to rates of clearance. The alternative description of binding described in the previous section, which explicitly includes the binding rate constants, would be preferred in cases where this assumption may be violated.

Protein binding in blood is a critical component of drug action and drug efficacy. PK models need to carefully consider the manner in which binding can be introduced into the basic equations and track free concentrations at sites of action or at least within the plasma. Introduction of blood binding in PBPK models raises some conceptual challenges, especially in comparing conventional empirical approaches with PBPK approaches intended to track thermodynamically free concentrations throughout the body. Although simple descriptions are often adequate, careful consideration of the underlying kinetic processes and their relative rates for the compound and tissues of interest is required to assure that a particular modeling approach is appropriate [56].

*Tissue:Blood Partitioning* Regardless of the manner in which blood binding is implemented in PBPK models, another challenge arises in describing equilibration between blood and tissues. In general, both the blood and the tissues will contain free and bound forms of the compound. For equilibration, only the free compound in the plasma diffuses across the tissue capillary interface into the tissue, and at equilibrium between blood and tissue, the free concentration in the plasma and the tissue is expected to be equal (except in the case of active transport). However, the equilibrium relationship of the concentrations in tissues compared to the blood or plasma is typically described with empirical partition coefficients based on measurements of total concentrations of the compound. Differential binding in plasma and tissue will therefore influence apparent tissue partitioning.

The relationship between apparent tissue:blood partitioning and binding in blood versus tissue can be straightforwardly described, as long as no other factors affect distribution. Assuming that (1) there is no need to adjust for the effect of clearance, and (2) there is no evidence of active transport of the compound between blood and tissue, the free fraction in the tissue, $f_{ut}$, can then be estimated:

$$P_{tb} = P_{tp}/BPR = (f_{up}/f_{ut})/BPR$$

where $P_{tb}$ is the tissue:blood partition coefficient, $P_{tp}$ is the tissue:plasma partition coefficient, $BPR$ is the blood to plasma ratio, $f_{up}$ is the fraction unbound in the plasma, and $f_{ut}$ is the fraction unbound in the tissue.
Then

$$f_{ut} = f_{up}/(P_{tb} \cdot BPR)$$

In fact, there are quite a number of different determinants of the apparent partitioning between blood and tissues:

- Partitioning due to lipophilicity
- Plasma binding
- Tissue binding

- Active transport
- Clearance processes
- Blood to plasma ratio (for converting tissue:plasma partitions to tissue:blood)

A measured or estimated partition coefficient may reflect any combination of these factors, and the modeler must be aware of this potential complexity in attempting to use a particular set of data on a compound. For example, if partition coefficients have been estimated from QSAR, they are likely to primarily reflect lipophilic partitioning and may need to be adjusted for differences in binding in plasma and tissues. Alternatively, the model can be designed to separately describe lipophilic partitioning, using the estimated partition coefficients, and binding, based on other data. *In vitro*-derived partition coefficients, on the other hand, may reflect both thermodynamic partitioning and binding, although disruption of the tissue architecture may alter the binding characteristics of the tissue. It is also necessary to ensure that no metabolic clearance of the compound occurs in either blood or tissue during the measurements. *In vivo*-derived partition coefficients may reflect lipophilicity, binding, and active transport, as well as the effect of any clearance processes [57]. Due to this complexity, it may in some cases be preferable to estimate model parameters for tissue partitioning by fitting the *in vivo* blood and tissue concentration data rather than trying to use it directly to calculate "partition coefficients" for the model.

***Model Diagram*** As described in the previous sections, the process of developing a PBPK model begins by determining the essential structure of the model based on the information available on the compound's toxicity, mechanism of action, and PK properties. The results of this step can usually be summarized by an initial model diagram, such as those depicted in Fig. 2 and 3. In fact, in many cases a well-constructed model diagram, together with a table of the input parameter values and their definitions, is all that an accomplished modeler should need in order to create the mathematical equations defining a PBPK model. In general, there should be a one-to-one correspondence of the boxes in the diagram to the mass balance equations (or steady-state approximations) in the model. Similarly, the arrows in the diagram correspond to the transport or metabolism processes in the model. Each of the arrows connecting the boxes in the diagram should correspond to one of the terms in the mass balance equations for both of the compartments it connects, with the direction of the arrow pointing from the compartment in which the term is negative to the compartment in which it is positive. Arrows only connected to a single compartment, which represent uptake and excretion processes, are interpreted similarly.

The model diagram should be labeled with the names of the key variables associated with the compartment or process represented by each box and arrow. Interpretation of the model diagram is also aided by the definition of the model input parameters in the corresponding table. The definition and units of the parameters can indicate the nature of the process being modeled (e.g., diffusion-limited vs. flow-limited transport, binding vs. partitioning, saturable vs. first-order metabolism).

## 3 PBPK MODEL IMPLEMENTATION

The purpose of developing a PBPK model is to perform simulations and learn about the system (Fig. 1). But first, the model must be parameterized, coded using some type of software package, and tested to make sure it is working properly. Also, sensitivity analysis and other techniques must be used to understand the model and verify that key inputs are known with sufficient accuracy to result in reasonable simulation results.

### 3.1 Model Parameterization

Once the model structure has been determined, it still remains to identify the values of the input parameters in the model.

***Physiological Parameters*** Estimates of the various physiological parameters needed in PBPK models are available from a number of sources in the literature, particularly for the human, monkey, dog, rat, and mouse [4, 26–30]. Estimates for the same parameter often vary widely, however, due both to experimental differences and to differences in the animals examined (age, strain, activity). Ventilation rates and blood flow rates are particularly sensitive to the level of activity [26, 28]. Data on some important tissues are relatively poor, particularly in the case of fat tissue. Table 2 shows typical values of a number of physiological parameters in several species.

***Biochemical Parameters*** One large barrier to the application of PBPK models in preclinical drug development has been that tissue : plasma partition coefficients ($P_{tp}$) are too expensive to routinely measure. Partition coefficients can be determined by *in vivo* or *in vitro* measurements [31, 57, 58] or can be estimated from quantitative structure–activity relationship (QSAR) modeling. Perhaps the driving force behind the recent interest in PBPK modeling in the pharmaceutical industry was reported successes in developing methods for *a priori* predictions of $P_{tp}$ values for volatile organic compounds based on their solubilities in the neutral lipid, phospholipid, and water fractions of tissues as estimated using a vegetable oil : water partition coefficient; the necessary tissue composition data has been tabulated for a variety of tissues [59–61]. This approach was eventually extended to using the mechanistic equations for predicting $P_{tp}$ values for drugs [62]. Briefly, these mechanistic equations assume that the drug distributes homogeneously into the tissue and plasma by passive diffusion, where two processes are accounted for: (1) nonspecific binding to lipids estimated from drug lipophilicity data ($\log P$, $\log D$) and (2) specific reversible binding to common proteins present in plasma and tissue estimated from the fraction unbound in the plasma ($f_{up}$). These $P_{tp}$ values can be used to estimate the volume of distribution [63]. Recent contributions have focused on improving methods for calculating $P_{tp}$ values for moderate-to-strong bases by developing a method incorporating the electrostatic interactions with tissue acidic phospholipids [64] and for acids, very weak bases, neutrals, and zwitterions by incorporating the effect of drug ionization on distribution processes [65].

One of the greatest challenges is typically the characterization of metabolism, which can also make use of both *in vivo* and *in vitro* methods. For compounds

**TABLE 2 Typical Physiological Parameters for PBPK Models**

| Species | Mouse | Rat | Monkey | Human |
|---|---|---|---|---|
| Ventilation | | | | |
| Alveolar (L/h)[a] | 29.0[b] | 15.0[b] | 15.0[b] | 15.0[b] |
| Blood flows | | | | |
| Total (L/h)[a] | 16.5[c] | 15.0[c] | 15.0[c] | 15.0[c] |
| Muscle (fraction) | 0.18 | 0.18 | 0.18 | 0.18 |
| Skin (fraction) | 0.07 | 0.08 | 0.06 | 0.06 |
| Fat (fraction) | 0.03 | 0.06 | 0.05 | 0.05 |
| Liver (arterial) (fraction) | 0.035 | 0.03 | 0.065 | 0.07 |
| Gut (portal) (fraction) | 0.165 | 0.18 | 0.185 | 0.19 |
| Other organs (fraction) | 0.52 | 0.47 | 0.46 | 0.45 |
| Tissue volumes | | | | |
| Body weight (kg) | 0.02 | 0.3 | 4.0 | 80.0 |
| Body water (fraction) | 0.65 | 0.65 | 0.65 | 0.65 |
| Plasma (fraction) | 0.04 | 0.04 | 0.04 | 0.04 |
| RBCs (fraction) | 0.03 | 0.03 | 0.03 | 0.03 |
| Muscle (fraction) | 0.34 | 0.36 | 0.48 | 0.33 |
| Skin (fraction) | 0.17 | 0.195 | 0.11 | 0.11 |
| Fat (fraction) | 0.10[d] | 0.07[d] | 0.05[d] | 0.21 |
| Liver (fraction) | 0.046 | 0.037 | 0.027 | 0.023 |
| Gut tissue (fraction) | 0.031 | 0.033 | 0.045 | 0.045 |
| Other organs (fraction) | 0.049 | 0.031 | 0.039 | 0.039 |
| Intestinal lumen (fraction) | 0.054 | 0.058 | 0.053 | 0.053 |

[a]Scaled allometrically: $QC = QCC \cdot BW^{0.75}$.
[b]Varies significantly with activity level (range: 15.0–40.0).
[c]Varies with activity level (range: 15.0–20.0).
[d]Varies substantially (lower in young animals, higher in older animals).

cleared by hepatic metabolism, intrinsic clearance values determined *in vitro* with hepatocyte or microsomal substrate depletion or kinetic assays are typically used. Many researchers have described physiological methods for scaling hepatocyte and microsome data, including Zuegge et al. [66], Lavé et al. [67], Houston [68], Houston and Carlile [69], Obach [70], and Ito et al. [71]. Recently improved parameters for the physiological scaling of human hepatocyte and microsome data have been published [72].

***Allometry*** The different types of physiological and biochemical parameters in a PBPK model are known to vary with body weight in different ways [24]. Typically, the parameterization of PBPK models is simplified by assuming standard allometric scaling [28], as shown in Table 3, where the scaling factors, $b$, can be used in the following equation:

$$Y = aX^b$$

where $Y$ is the value of the parameter at a given body weight, $X$ (kg), and $a$ is the scaled parameter value for a 1 kg animal.

**TABLE 3 Standard Allometric Scaling for Physiologically Based Pharmacokinetic Model Parameters**

| Parameter Type (units) | Scaling (power of body weight) |
|---|---|
| Volumes | 1.0 |
| Flows (volume per time) | 0.75 |
| Ventilation (volume per time) | 0.75 |
| Clearances (volume per time) | 0.75 |
| Metabolic capacities (mass per time) | 0.75 |
| Metabolic affinities (mass per volume) | 0 |
| Partition coefficients (unitless) | 0 |
| First-order rate constants (inverse time) | –0.25 |

While standard allometric scaling provides a useful starting point, or hypothesis, for cross-species scaling, it is not sufficiently accurate for some applications. In the case of the physiological parameters, the species-specific parameter values are generally available in the literature [4, 26–30] and can be used directly in place of the allometric estimates. For compound-specific parameters, *in vitro* data for metabolism, distribution, or absorption relevant to the species under consideration are used in most of the cases in preference to allometric estimates. However, allometric scaling might provide first parameter estimates when data is lacking.

***Parameter Optimization*** In many cases, important parameter values needed for a PBPK model may not be available in the literature. In such cases it is necessary to measure them in new experiments, to estimate them by QSAR techniques, or to identify them by optimizing the fit of the model to an informative dataset. Even in the case where an initial estimate of a particular parameter value can be obtained from other sources, it may be desirable to refine the estimate by optimization. For example, given the difficulty of obtaining accurate estimates of the fat volume in rodents, a more reliable estimate may be obtained by examining the impact of fat volume on the kinetic behavior of a lipophilic compound. Of course, being able to uniquely identify a parameter from a kinetic dataset rests on two key assumptions: (1) that the kinetic behavior of the compound under the conditions in which the data was collected is sensitive to the parameter being estimated, and (2) that other parameters in the model, which could influence the observed kinetics, have been determined by other means and are held fixed during the estimation process. When it is necessary to estimate multiple parameters from *in vivo* PK data, using sensitivity analysis to verify that sufficient data are available and optimizing on individual parameters before performing a global optimization has been recommended [73].

The actual approach for conducting a parameter optimization can range from simple visual fitting, where the model is run with different values of the parameter until the best correspondence appears to be achieved, to use of a quantitative mathematical algorithm. The most common algorithm used in optimization is the least-squares fit. To perform a least-squares optimization, the model is run to obtain a set of predictions at each of the times a data point was collected. The square of the difference between the model prediction and data point at each time is calculated

and the results for all of the data points are summed. The parameter being estimated is then modified, and the sum of squares is recalculated. This process is repeated until the smallest possible sum of squares is obtained, representing the best possible fit of the model to the data.

In a variation on this approach, the square of the difference at each point is divided by the square of the prediction. This variation, known as relative least squares, is preferable in the case of data with an error structure that can be described by a constant coefficient of variation (i.e., a constant ratio of the standard deviation to the mean). The former method, known as absolute least squares, is preferable in the case of data with a constant variance. From a practical viewpoint, the absolute least-squares method tends to give greater weight to the data at higher concentrations and results in fits that look best when plotted on a linear scale, while the relative least-squares method gives greater weight to the data at lower concentrations and results in fits that look best when plotted on a logarithmic scale.

A generalization of this weighting concept is provided by the extended least-squares method. In the extended least-squares algorithm, the heteroscedasticity parameter can be varied from 0 (for absolute weighting) to 2 (for relative weighting) or can be estimated from the data. In general, setting the heteroscedasticity parameter from knowledge of the error structure of the data is preferable to estimating it from a dataset.

A common example of identifying PBPK model parameters by fitting kinetic data is the estimation of tissue partition coefficients from experiments in which the concentration of compound in the blood and tissues is reported at various time points. Using an optimization approach, the predictions of the model for the time course in the blood and tissues could be optimized with respect to the data by varying the model's partition coefficients. There is little difference in the strength of the justification for estimating the partition coefficients in this way as opposed to estimating them directly from the data (by dividing the tissue concentrations by the simultaneous blood concentration). In fact, the direct estimates would probably be used as initial estimates in the model when the optimization was started.

A major difficulty in performing parameter optimization results from correlations between the parameters. When it is necessary to estimate parameters that are highly correlated, it is best to generate a contour plot of the objective function (sum of squares) or confidence region over a reasonable range of values of the two parameters [74].

***Mass Balance Requirements*** One of the most important mathematical considerations during model design is the maintenance of mass balance. Simply put, the model should neither create nor destroy mass. This seemingly obvious principle is often violated unintentionally during the process of model development and parameterization. A common violation of mass balance, which typically leads to catastrophic results, involves failure to exactly match the arterial and venous blood flows in the model. As described previously, the movement of compound in the blood (in units of mass per time) is described as the product of the concentration of compound in the blood (in units of mass per volume) times the flow rate of the blood (in units of volume per time). Therefore, to maintain mass balance, the sum of the blood flows leaving any particular tissue compartment must equal the sum of the blood flows entering the compartment. In particular, to maintain mass balance in the blood

compartment (regardless of whether it's actually a compartment or just a steady-state equation), the sum of the venous flows from the individual tissue compartments must equal the total arterial blood flow leaving the heart:

$$\sum Q_T = Q_C$$

Another obvious but occasionally overlooked aspect of maintaining mass balance during model development is that if a model is modified by splitting a tissue out of a lumped compartment, the blood flow to the separated tissue (and its volume) must be subtracted from that for the lumped compartment. Moreover, even though a model may initially be designed with parameters that meet the above requirements, mass balance may unintentionally be violated later if the parameters are altered during model execution. For example, if the parameter for the blood flow to one compartment is increased, the parameter for the overall blood flow must be increased accordingly or an equivalent reduction must be made in the parameter for the blood flow to another compartment. Particular care must be taken in this regard when the model is subjected to sensitivity or uncertainty analysis; inadvertent violation of mass balance during Monte Carlo sampling can lead to erroneous sensitivity results [39].

A similar mass balance requirement must be met for transport other than blood flow. For example, if the compound is cleared by biliary excretion, the elimination of compound from the liver in the bile must exactly match the appearance of compound in the gut lumen in the bile. Put mathematically, the same term for the transport must appear in the equations for the two compartments, but with opposite signs (positive vs. negative). For example, the following equation could be used to describe a liver compartment with first-order metabolism and biliary clearance:

$$dA_L/dt = Q_L \cdot (C_A - C_L/P_L) - k_F \cdot C_L \cdot V_L/P_L - K_B \cdot C_L$$

where $K_B$ is the biliary clearance rate (L/h). The equation for the intestinal lumen would then need to include the term: $+K_B \cdot C_L$.

As a model grows in complexity, it becomes increasingly difficult to assure its mass balance by inspection. Therefore, it is a worthwhile practice to check for mass balance by including an equation in the model that adds up the total amount of compound in each of the model compartments, including metabolized and excreted compound, for comparison with the administered or inhaled dose.

### 3.2 Numerical Solution of Model Equations

The previous sections have focused on the process of designing the PBPK model structure needed for a particular application. At this point the model consists of a number of mathematical equations: differential equations describing the mass balance for each of the compartments and algebraic equations describing other relationships between model variables. The next step in model development is the coding of the mathematical form of the model into a form that can be executed on a computer. There are many options available for performing this process, ranging from programming languages such as Fortran, C, and MatLab to more sophisticated simulation software packages such as acslXtreme and Berkeley Madonna.

Mathematically, a PBPK model is represented by a system of simultaneous ordinary differential equations. Each of the differential equations describes the mass balance for one of the "state variables" (compartments) in the model. There may also be additional differential equations to calculate other necessary model outputs, such as the area under the concentration curve (*AUC*) in a particular compartment, which is simply the integral of the concentration over time. The resulting system of equations is referred to as simultaneous because the time courses of the compound in the various compartments are so interdependent that solving the equations for any one of the compartments requires information on the current status of all the other compartments; that is, the equations for all of the compartments must be solved at the same time. This kind of mathematical problem, in which a system is defined by the conditions at time zero together with differential equations describing how it evolves over time, is known as an *initial value problem*, and matrix decomposition methods can be used to obtain the simultaneous solution.

A number of numerical algorithms are available for solving such problems. They all have in common that they are stepwise approximations; that is, they begin with the conditions at time zero and use the differential equations to predict how the system will change over a small time step, resulting in an estimate of the conditions at a slightly later time, which serves as the starting point for the next time step. This iterative process is repeated as long as necessary to simulate the experimental scenario.

The more sophisticated methods, such as the Gear algorithm (named after the mathematician David Gear, who developed it), use a predictor–corrector approach, in which the corrector step essentially amounts to "predicting backwards" after each step forward, in order to check how closely the algorithm is able to reproduce the conditions at the previous time step. This allows the time step to be increased automatically when the algorithm is performing well, and to be shortened when it is having difficulty, such as when conditions are changing rapidly. However, due to the wide variation of the time constants (response times) for the various physiological compartments (e.g., fat vs. richly perfused), PBPK models often represent "stiff" systems. Stiff systems are characterized by state variables (compartments) with widely different time constants, which cause difficulty for predictor–corrector algorithms. The Gear algorithm was specifically designed to overcome this difficulty. It is therefore generally recommended that the Gear algorithm (or an alternative stiff system solver such as the Rosenbrock method) be used for executing PBPK models.

Regardless of the specific algorithm selected, the essential nature of the solution will be a stepwise approximation. However, all of the algorithms made available in computer software are convergent; that is, they can stay arbitrarily close to the true solution, given a small enough time step. On modern personal computers, even large PBPK models can be run to more than adequate accuracy in a reasonable timeframe.

### 3.3 Model Evaluation

This section discusses various issues associated with the evaluation of a PBPK model. Once an initial model has been developed, it must be evaluated on the basis

of its conformance with experimental data. In some cases, the model may be exercised to predict conditions under which experimental data should be collected in order to verify or improve model performance. Comparison of the resulting data with the model predictions may suggest that revision of the model will be required. Similarly, a PBPK model designed for one compound or application may be adapted to another compound or application, requiring modification of the model structure and parameters. It is imperative that revision or modification of a model is conducted with the same level of rigor applied during initial model development, and that structures are not added to the model with no other justification than that they improve the agreement of the model with a particular dataset.

In addition to comparing model predictions to experimental data, model evaluation includes assessing the biological plausibility of the model structures and input parameters, and the resulting confidence that can be placed in extrapolations performed by the model [75]. Both elements of testing the model—kinetic validation and mechanistic validation—are necessary to provide confidence in the model. Unfortunately, there is a temptation to accept kinetic validation alone, particularly when data for mechanistic validation are unavailable. It should be remembered, however, that the simple act of adding equations and parameters to a model will, in itself, increase the flexibility of the model to fit data. Therefore, every attempt should be made to obtain additional experimental data to provide support for the mechanistic hypothesis underlying the model structure.

***Model Documentation*** In cases where a model previously developed by one investigator is being evaluated for use in a different application by another investigator, adequate model documentation is critical for evaluation of the model. The documentation for a PBPK model should include sufficient information about the model so that an experienced modeler could accurately reproduce its structure and parameterization. Usually the suitable documentation of a model will require a combination of one or more "box and arrow" model diagrams together with any equations that cannot be unequivocally derived from the diagrams. Model diagrams should clearly differentiate blood flow from other transport (e.g., biliary excretion) or metabolism, and arrows should be used where the direction of transport could be ambiguous. All tissue compartments, metabolism pathways, routes of exposure, and routes of elimination should be clearly and accurately presented. All equations should be dimensionally consistent and in standard mathematical notation. Generic equations (e.g., for tissue "*i*") can help to keep the description brief but complete. The values used for all model parameters should be provided, with units. If any of the listed parameter values are based on allometric scaling, a footnote should provide the body weight used to obtain the allometric constant as well as the power of body weight used in the scaling.

### 3.4 Model Validation

Internal validation consists of the evaluation of the mathematical correctness of the model [34]. It is best accomplished on the actual model code but, if necessary, can be performed on appropriate documentation of the model structure and parameters, as described earlier (assuming, of course, that the actual model code accurately reflects the model documentation). A more important issue regards the provision

of evidence for external validation (sometimes referred to as verification). The level of detail incorporated into a model is necessarily a compromise between biological accuracy and parsimony. The process of evaluating the sufficiency of the model for its intended purpose, termed *model verification*, requires a demonstration of the ability of the model to predict the behavior of experimental data different from that on which it was based.

Whereas a simulation is intended simply to reproduce the behavior of a system, a model is intended to confirm a hypothesis concerning the nature of the system. Therefore, model validation should demonstrate the ability of the model to predict the behavior of the system under conditions that test the principal aspects of the underlying hypothetical structure. While quantitative tests of goodness of fit may often be a useful aspect of the verification process, the more important consideration may be the ability of the model to provide an accurate prediction of the general behavior of the data in the intended application.

Where only some aspects of the model can be verified, it is particularly important to assess the uncertainty associated with the aspects that are untested. For example, a model of a compound and its metabolites that is intended for use in cross-species extrapolation to humans would preferably be verified using data in different species, including humans, for both the parent compound and the metabolites. If only parent compound data is available in the human, the correspondence of metabolite predictions with data in several animal species could be used as a surrogate, but this deficiency should be carefully considered when applying the model to predict human metabolism. One of the values of biologically based modeling is the identification of specific data that would improve the quantitative prediction of toxicity in humans from animal experiments.

In some cases it is necessary to use all of the available data to support model development and parameterization. Unfortunately, this type of modeling can easily become a form of self-fulfilling prophecy: models are logically strongest when they fail, but psychologically most appealing when they succeed [33]. Under these conditions, model verification can be particularly difficult, putting an additional burden on the investigators to substantiate the trustworthiness of the model for its intended purpose. Nevertheless, a combined model development and verification can often be successfully performed, particularly for models intended for interpolation, integration, and comparison of data rather than for true extrapolation.

***Parameter Verification*** In addition to verifying the performance of the model against experimental data, the model should be evaluated in terms of the plausibility of its parameters. This is particularly important in the case of PBPK models, where the parameters generally possess biological significance and can therefore be evaluated for plausibility independent of the context of the model. The source of each model input parameter value should be identified, whether it was obtained from prior literature, determined directly by experiment, or estimated by fitting a model output to experimental data. Parameter estimates derived independently of tissue time course or dose–response data are preferred. To the extent feasible, the degree of uncertainty regarding the parameter values should also be evaluated. The empirically derived *Law of Reciprocal Certainty* states that the more important the model parameter, the less certain will be its value. In accordance with this principle, the

most difficult, and typically most important, parameter determination for PBPK models is the characterization of the metabolism parameters.

When parameter estimation has been performed by optimizing model output to experimental data, the investigator must assure that the parameter is adequately identifiable from the data [34]. Due to the confounding effects of model error, overparameterization, and parameter correlation, it is quite possible for an optimization algorithm to obtain a better fit to a particular dataset by modifying a parameter that in fact should not be identified on the basis of that dataset. Also, when an automatic optimization routine is employed, it should be restarted with a variety of initial parameter values to assure that the routine has not stopped at a local optimum. These precautions are particularly important when more than one parameter is being estimated simultaneously, since the parameters in biologically based models are often highly correlated, making independent estimation difficult. Estimates of parameter variance obtained from automatic optimization routines should be viewed as lower bound estimates of true parameter uncertainty since only a local, linearized variance is typically calculated. In characterizing parameter uncertainty, it is probably more instructive to determine what ranges of parameter values are clearly inconsistent with the data than to accept a local, linearized variance estimate provided by the optimization algorithm.

It is usually necessary for the investigator to repeatedly vary the model parameters manually to obtain a sense of their identifiability and correlation under various experimental conditions, although some simulation languages include routines for calculating parameter sensitivity and covariance or for plotting confidence region contours. Sensitivity analysis and Monte Carlo uncertainty analysis techniques can serve as useful methods to estimate the impact of input parameter uncertainty on the uncertainty of model outputs. However, care should be taken to avoid violation of mass balance when parameters are varied by sensitivity or Monte Carlo algorithms, particularly where blood flows are affected.

***Sensitivity Analysis*** To the extent that a particular PBPK model correctly reflects the physiological and biochemical processes underlying the pharmacokinetics of a compound, exercising the model can provide a means for identifying the most important physiological and biochemical parameters determining the PK behavior of the compound under different conditions [39]. The technique for obtaining this information is known as *sensitivity analysis* and can be performed by two different methods. Analytical sensitivity coefficients are defined as the ratio of the change in a model output to the change in a model parameter that produced it. To obtain a sensitivity coefficient by this method, the model is run for the exposure scenario of interest using the preferred values of the input parameters, and the resulting output (e.g., brain concentration) is recorded. The model is then run again with the value of one of the input parameters varied slightly. Typically, a 1% change is appropriate. The ratio of the resulting incremental change in the output to the change in the input represents the sensitivity coefficient. It is usually more convenient to use log-normalized sensitivity coefficients, which represent the ratio of the fractional change in output to the fractional change in input. For example, if a 1% increase in an input parameter resulted in a 0.5% decrease in the output, the log-normalized sensitivity coefficient would be –0.5. Log-normalized sensitivity coefficients >1.0 in absolute value represent amplification of input error and would be a cause for concern. An

alternative approach is to conduct a Monte Carlo analysis, as described below, and then to perform a simple correlation analysis of the model outputs and input parameters. This type of approach is often referred to as *global sensitivity analysis* [76]. Both methods have specific advantages. The analytical sensitivity coefficient most accurately represents the functional relationship of the output to the specific input under the conditions being modeled. The advantage of the global sensitivity analysis is that it also reflects the impact of interactions between the parameters during the Monte Carlo analysis.

***Uncertainty and Variability Analysis*** Evaluations of the uncertainty and/or the variability associated with the predictions of a PBPK model are often performed using the Monte Carlo simulation approach [77, 78]. In a Monte Carlo simulation, a probability distribution for each of the PBPK model parameters is randomly sampled, and the model is run using the chosen set of parameter values. This process is repeated a large number of times until the probability distribution for the desired model output has been created. Generally, 1000 iterations or more may be required to ensure the reproducibility of the mean and standard deviation of the output distributions as well as the 1st through 99th percentiles. To the extent that the input parameter distributions adequately characterize the uncertainty in the inputs, and assuming that the parameters are reasonably independent, the resulting output distribution will provide a useful estimate of the uncertainty associated with the model outputs. If simulations are performed so that the probability distribution for PBPK model parameters represents the variability expected in the human population, then the Monte Carlo analysis will result in the simulation of pharmacokinetics expected for a population.

In performing a Monte Carlo analysis, it is important to distinguish uncertainty from variability. As it relates to the impact of pharmacokinetics in risk assessment, uncertainty can be defined as the possible error in estimating the "true" value of a parameter for a representative ("average") person. Variability, on the other hand, should only be considered to represent true interindividual differences. Understood in these terms, uncertainty is a defect (lack of certainty), which can typically be reduced by experimentation, and variability is a fact of life, which must be considered regardless of the risk assessment methodology used. An elegant approach for separately documenting the impact of uncertainty and variability is "two-dimensional" Monte Carlo, in which distributions for both uncertainty and variability are developed and multiple Monte Carlo runs are used to convolute the two aspects of overall uncertainty. Unfortunately, in practice, it is often difficult to differentiate the contribution of variability and uncertainty to the observed variation in the reported measurements of a particular parameter.

Due to its physiological structure, many of the parameters in a PBPK model are interdependent. For example, the blood flows must add up to the total cardiac output and the tissue volumes (including those not included in the model) must add up to the body weight. Failure to account for the impact of Monte Carlo sampling on these mass balances can produce erroneous results. In addition, some physiological parameters are naturally correlated, such as cardiac output and respiratory ventilation rate, and these correlations should be taken into account during the Monte Carlo analysis.

Although sophisticated methods of uncertainty analysis can provide valuable information, very simple techniques can also be used as a way to effectively communicate uncertainty. By simply determining the minimum and maximum values of an uncertain but important model parameter, and performing a simulation for a range of values, the impact of that parameter on PK predictions can be illustrated. When predicting human pharmacokinetics based on preclinical data, there is always uncertainty. Quantifying key uncertainties and providing a range of possible outcomes that could all be reasonable based on the current knowledge of the PK properties of the compound will allow for more informed decision-making.

***Collection of Critical Data*** As with model development, the best approach to model evaluation is within the context of the scientific method. The most effective way to evaluate a PBPK model is to exercise the model to generate a quantitative hypothesis; that is, to predict the behavior of the system of interest under conditions "outside the envelope" of the data used to develop the model (at shorter/longer durations, higher/lower concentrations, different routes, different species, etc.). In particular, if there is an element of the model that remains in question, the model can be exercised to determine the experimental design under which the specific model element can best be tested. For example, if there is uncertainty regarding whether uptake into a particular tissue is flow or diffusion limited, alternative forms of the model can be used to compare predicted tissue concentration time courses under each of the limiting assumptions under various experimental conditions. The experimental design and sampling time that maximize the difference between the predicted tissue concentrations under the two assumptions can then serve as the basis for the actual experimental data collection. Once the critical data has been collected, the same model can also be used to support a more quantitative experimental inference. In the case of the tissue uptake question just described, not only can the *a priori* model predictions be compared with the observed data to test the alternative hypotheses, but the model can also be used *a posteriori* to estimate the quantitative extent of any observed diffusion limitation (i.e., to estimate the relevant model parameter by fitting the data). If, on the other hand, the model is unable to reproduce the experimental data under either assumption, it may be necessary to reevaluate other aspects of the model structure. The key difference between research and analysis is the iterative nature of the former. It has wisely been said, "If we knew when we started what we had to do to finish, they'd call it search, not research."

## 4 GENERIC PBPK MODELS

Despite the clear value of compound-specific PBPK models developed by scientists with experience in kinetic modeling and designed to address kinetic issues specific to that compound, there is also a need for simpler modeling techniques that can readily be applied across diverse compounds by an inexperienced user. The past ten years have seen tremendous advances in the capabilities of generic PBPK models that can simulate pharmacokinetics for humans or preclinical species based on a combination of physicochemical properties and *in vitro* data. The generic PBPK

model has recently become an increasingly important tool in assessing DMPK properties in preclinical development. Recent work has demonstrated the value of applying generic PBPK models in early preclinical development [14, 21, 22, 79].

### 4.1 User-Friendly Software Packages Bringing PBPK Modeling to the Nonexpert

The possible benefits from having PBPK software packages available for scientists who are nonmodelers have been a topic of interest for many years. Such a tool would need to meet many criteria to be broadly useful. This tool would need to be user-friendly, fast, flexible, and requiring little mathematical expertise from the user. The import of data into the tool and the export of results from the tool would need to be fast and automatable. The required input would be clearly defined and difficult to enter incorrectly, making consistent use by scientists of varying backgrounds possible. The tool would ideally be applicable through all stages of preclinical through clinical development. The pharmaceutical industry now has many tools that meet some or all of these criteria. Tools using physiology-based whole body models for the simulation of the PK behavior including absorption, distribution, and elimination include GastroPlus™ (Simulations Plus Inc., www.simulations-plus.com), Simcyp (Simcyp, www.simcyp.com), PK-Sim® (Bayer Technology Services, www.pksim.com), MEDICI-PK (AT Computing in Technology GnbH, www.cit-wulkow.de), and Cloe PK® (Cyprotex, www.cyprotex.com).

### 4.2 Recent Advances in Generic PBPK Modeling Tools

GastroPlus, which is well known in the pharmaceutical industry for its capabilities in predicting oral absorption in preclinical species based on the ACAT model, has recently expanded to include the capability of PBPK modeling. In the most recent version, GastroPlus includes a physiological intestinal model for humans in fed and fasted states as well as models for the rat, dog, cynomolgus monkey, and mouse. Several additional modules are available for parameter optimization, for combining the absorption model with physiological and compartmental PK models, and for PD models. The software also includes clinical trial simulation capabilities and a powerful method of performing sensitivity analysis. Although currently only PBPK modules for the rat and human are available, users can create PBPK models for other species themselves by creating input files with appropriate physiological parameters. The performance of GastroPlus to predict absorption was evaluated by Parrott and Lavé [80]. In this study, GastroPlus was compared to iDEA™, another software to predict absorption that is no longer commercially available.

Another tool for predicting human pharmacokinetics, Simcyp (Simcyp Limited, Sheffield, UK), was previously known as a DDI clinical trial simulator but has expanded to include PBPK modeling capabilities. Simcyp utilizes fundamental scaling procedures described by Houston [68] for the prediction of *in vivo* hepatic clearance ($CL_H$) from *in vitro* metabolism data. These *in vitro* metabolism data can be obtained from individual cytochrome P450s (CYP) in human-expressed recombinant systems. These are used for predicting behavior not only in average individuals but also in whole populations or in special subpopulations. In order to predict

drug–drug interactions involving CYP, Simcyp utilizes the relationship between the inhibitor concentration at the active site *in vivo* and the inhibition constant ($K_i$) determined *in vitro*. Competitive inhibition, induction, and/or mechanism-based inactivation mechanisms can be investigated within this software, according to the principles described by Ito et al. [71] and Tucker et al. [81]. Simcyp not only predicts the mean value but also simulates the extremes in the population by applying a Monte Carlo approach. This software comes with a database of physiological properties for specific populations and of PK information for specific drugs on the market (e.g., the CYP3A4 substrate midazolam, the CYP3A4 inducer rifampicin, and the CYP3A4 inducer and inhibitor ritonavir).

Yet another tool, PK-Sim, developed as a user-friendly tool for PBPK modeling, has PBPK models in the mouse, rat, dog, monkey, and human. In the model, tissue concentrations depend on the blood flow rate to the tissue, the rate of permeation into the tissue, active transport processes into and out of the tissue, and the rate of metabolism within a tissue. Although this detailed description allows significant flexibility and complexity, the user-friendly interface also allows for simple simulations whereby the more complex features are not utilized. In PK-Sim, the GI tract is described as a single tube with physiologically relevant dimensions and spatially varying properties such as pH and surface area. The software has Monte Carlo capabilities, and simulations examining the impact of interindividual variability can be performed using the population PK module. Physiological parameters for people from 0 to 80 years old are provided, allowing simulations of pharmacokinetics in the young and the elderly. Also, the software has the capability of linking PK to PD behavior with a PK/PD module.

In a fresh approach to PBPK modeling, the new software MEDICI-PK was developed using a novel software concept aimed at providing increased flexibility while maintaining user-friendliness. Instead of adopting the traditional approach whereby user-friendliness is made possible by adopting a static model, in MEDICI-PK the system of differential equations is generated automatically when starting a simulation. This software allows for predictions of parent compound and multiple metabolites, and interactions (metabolic and otherwise) between compounds [82]. MEDICI-PK has PBPK modeling capabilities for multiple species (e.g., the human, mouse, and dog) that can be extended by the user, and also contains PD modeling capabilities.

Cloe-PK is a commercial tool aimed at prediction of human pharmacokinetics at the earliest stages of drug discovery [83]. The PBPK model includes 14 organs and tissues, clearance by metabolism and renal excretion, and diffusion-limited uptake in tissues. In each of the organ compartments, the capillary, intracellular, and extracellular spaces are explicitly included. Using *in vitro* and *in silico* data as input, this software can predict human IV and oral *AUC* from clinical trial data within 2-fold and 3.75-fold, respectively (www.cyprotex.com/products/pk_howwell.htm). Studies on the accuracy of the model for predicting PK behavior in the rat and human have been published [21, 22] and illustrate the utility of this software for predicting pharmacokinetics and ranking compounds based on minimal data and *in silico* parameters.

These software packages are exceptionally useful tools, and we expect them to be increasingly used in the pharmaceutical industry. However, their implementation may require a modeling and simulation expert to determine appropriate

applications and methods of use and to perform validation exercises to convince the scientists involved with the preclinical development of drugs that the packages do indeed work. Additionally, in-house PBPK expertise and models will still be necessary because these standard tools do not allow the flexibility to simulate every situation that will arise. For example, currently some generic PBPK tools do not yet have the capability to simulate exposures to a parent compound and a metabolite, which could be of interest for simulating the pharmacokinetics of prodrugs or investigating possible PK interactions between the parent and metabolite. Also, generic PBPK software packages have sophisticated capabilities that may require an expert modeler to determine appropriate ways to use (and not misuse) the features (e.g., GastroPlus has the capability of simulating enterohepatic cycling, and PK-Sim can incorporate saturable and nonsaturable metabolism in multiple tissue compartments). In addition, the availability of such tools simplifies the technical use of PBPK models; however, a good understanding of the models and underlying equations is still mandatory in order to guarantee good interpretation of output.

## 5 INTEGRATING PRECLINICAL DATA USING THE PBPK MODEL

Here, we provide examples of how PBPK modeling is a valuable tool in several stages of drug development. During the lead identification (LI) and lead optimization (LO) stages, limited data is available. During the clinical candidate selection, more data is available. Finally, a rich dataset is available prior to entry into humans. The availability of data and issues that arise at each stage dictate the utility of PBPK modeling.

### 5.1 Preclinical Data Useful for PBPK Modeling

Much of the data typically generated in the preclinical development process used to characterize key aspects of ADME can be used in PBPK model development (Table 1). The first information generated includes *in silico* parameters (e.g., polar surface area, $pK_a$, and $\log P$ values are calculated). Metabolic stability studies (e.g., in microsomes or hepatocytes) are often considered critical to determine if hepatic metabolism is a major route of elimination and if first-pass metabolism might result in unacceptably low bioavailability [84, 85]. Screens for permeability and solubility are often implemented early in the process. Plasma protein binding might be measured to determine if a compound will have a sufficient free concentration for therapeutic efficacy [86]. The blood-to-plasma ratio, *BPR*, might be measured to provide a rational basis for the appropriate biological fluid for assay (e.g., blood or plasma) and to aid in interpreting PK data [42]. Also, screens specific to a disease area might be implemented. For example, a P-glycoprotein (P-gp) screen might be used to determine if a compound will be able to penetrate the central nervous system.

Single-dose pharmacokinetic (SDPK) studies in a rodent species might be performed to determine if a compound has drug-like PK properties and adequate bioavailability. Later studies might include SDPK studies in a nonrodent species. Hepatocyte or microsomal clearance data can be scaled to estimate the *in vivo* metabolic clearance in preclinical species and humans [70, 87]. The scaled clearance

in preclinical species can be compared to clearance in SDPK studies to determine if hepatic metabolism is a major route of elimination in preclinical species. If the clearance seen in SDPK studies is higher than the clearance scaled from hepatocyte or microsome data, urine and/or bile might be collected in SDPK studies to provide additional information on mechanisms of clearance.

These studies provide pharmaceutical scientists with qualitative information on the potential human PK behavior of a compound. Using PBPK modeling, the results of the various preclinical assays can be integrated to provide a quantitative prediction of human pharmacokinetics [88]. During preclinical development, DMPK data are constantly being generated. Therefore, the human PK projection will change as more information becomes available. A PBPK model can act as a repository of current information and is easily updated when new PK or PD data becomes available. An important feature of PBPK models is that they can be used to determine the key data that should be generated.

### 5.2 Lead Identification and Optimization Stages

Drug discovery is increasingly "data rich" with high throughput chemistry generating numerous compounds that are rapidly screened for pharmacological and PK properties. Determination of *in vivo* pharmacokinetics is considerably more costly and slower than *in vitro* screening, and so there is interest in optimizing resources by using simulation to prioritize compounds. Physiologically based pharmacokinetics can be used at this stage to prioritize compounds prior to *in vivo* experimentation—getting an estimate of the expected PK profile of new drug candidates in humans and providing some mechanistic understanding of the compound's properties and of the relevance of the assays used in the screening strategy.

Because of the mechanistic basis of PBPK models, when they do not adequately describe animal PK data, this means that a biological phenomenon affecting pharmacokinetics has not been included in the model and is not represented by the assays used to screen the compounds. Therefore, if a PK issue for a promising compound becomes apparent, PBPK modeling allows you to determine which mechanisms are consistent with the observed data. This information can be used to guide further experimentation to arrive more rapidly at the desired information. Understanding pharmacokinetics in preclinical species using PBPK modeling can also provide a scientific basis for determining which preclinical species are most relevant to humans, thereby improving the scientific rationale of the selected screening strategy.

The emerging technology of generic PBPK modeling has been shown in recent studies to have potential as a valuable tool for assessing the pharmacokinetics of compounds in the early stages of drug development. For example, recent papers reported studies on applying generic PBPK modeling to predict rat pharmacokinetics [21] and human pharmacokinetics [22] following IV administration of test compound. These studies were conducted to evaluate the commercial PBPK modeling package Cloe PK [83]. The PBPK model included 14 organs and tissues, clearance by metabolism and renal excretion, and diffusion-limited uptake in tissues. For these rat and human models, model input included molecular weight, octanol/water partitioning, octanol/water distribution coefficient at a pH of 7.4, hepatic intrinsic metabolic clearance, and the unbound fraction in plasma. Model development involved

the use of a training dataset consisting of 82 compounds for the rat model and 69 compounds for the human model. After the models were developed, they were tested using data from test datasets consisting of 134 compounds in the rat and 18 compounds in the human. Using this approach to rank compounds and select the 25% most likely to have the best exposure (i.e., *AUC*) resulted in 61% of the best compounds being selected, which is a 2.4-fold improvement over random selection [21]. Brightman et al. [22] concluded that "the generic PBPK model is potentially a powerful and cost-effective tool for predicting the mammalian pharmacokinetics of a wide range of organic compounds, from readily available *in vitro* inputs only."

In a similar study, Parrott et al. [79] studied applying generic PBPK modeling to predict rat pharmacokinetics following IV or PO administration of test compound. This rat PBPK model included 11 tissue compartments and clearance by metabolism in the liver; uptake in tissues was assumed to be perfusion limited. GastroPlus was used to simulate oral absorption. For this model, the necessary input included structure-based predictions of lipophilicity, ionization, and protein binding, intrinsic clearance based on rat hepatocyte data, experimentally measured aqueous solubility, and artificial membrane permeability. To improve the prediction of AUC following oral administration of compounds, particularly for compounds with $\log P > 4$, it was necessary to use solubility in simulated intestinal fluid since aqueous solubility apparently underestimates *in vivo* solubility for highly lipophilic compounds. Parrott et al. [79] noted that "overall, this evaluation shows that generic simulation may be applicable for typical drug-like compounds to predict differences in PK parameters of more than twofold based upon minimal measured input data." The authors recommend that, for applying the method for new series of compounds outside the range of the training dataset, the model be verified for several compounds before it is generally applied.

### 5.3 Clinical Candidate Selection

PBPK modeling has a great potential to assist clinical candidate selection where numerous factors must be considered and data related to the pharmacokinetics and pharmacodynamics of a compound needs to be combined and compared in a rational way. This potential was illustrated by Parrott et al. [14], who demonstrated the use of PBPK modeling to select the best compound from among five candidates for the clinical lead. The preclinical data for the five candidates was integrated and the efficacious human doses and associated internal exposures were estimated. For this calculation, the PBPK models were linked to an $E_{max}$ PD model so that the dose resulting in a 90% effect could be identified. This example showed that the PBPK approach facilitates a sound decision on the selection of the optimal molecule to be progressed by integrating the available information and focusing the attention onto the expected properties in human. Importantly, the method can include estimates of variability and uncertainty in the predictions to verify that decisions are based on significant differences between the compounds.

### 5.4 Supporting Entry into Humans

After a clinical candidate has been selected, the first-in-human (FIH) dose must be selected. There are many ways of selecting the FIH dose. For example, in therapeutic

areas other than oncology, FIH dose estimates can be determined by multiplying the no-observed-adverse-effect-level (NOAEL) from preclinical toxicology studies by a safety factor, by using human data for a similar compound when they are available, or by using a PK model based on preclinical data to determine the dose that will result in a systemic exposure identified as therapeutic based on preclinical pharmacology data [89]. When this last approach is adopted, it is often referred to as performing a human PK extrapolation.

The first step of the human PK extrapolation is to identify the internal measure of exposure (e.g., the $C_{min}$, $C_{max}$, or $AUC$) believed to be efficacious in humans based on pharmacology data. Next, a human PK model for IV exposures is developed based on all preclinical data. This PK model can be developed in many ways, including PBPK modeling [13]. The oral exposure route is added to the human IV PK model. Again, there are many ways to model oral absorption, for example, by using an empirical first-order oral absorption rate constant or using the physiological ACAT model. Finally, once the model is developed, the dose that will result in the therapeutic exposure is determined. The estimated efficacious dose and other information (e.g., the $C_{min}$, $C_{max}$, $AUC$(0–24 h), $CL$, $V_{ss}$, $t_{1/2}$) are reported. PK projections can be based on a single or repeated (e.g., steady-state) exposure. At this stage it is important to understand the uncertainty of the FIH dose estimate.

Uncertainty can be illustrated by presenting a range of results (e.g., dose, AUC) from different models, or a range of results for different values of a key parameter that is not known with great certainty. If uncertainty is too great, additional experiments can be designed to aid in narrowing the possible values of key parameters. If uncertainty is extensive and cannot be reduced with additional experimentation, it is perhaps better to simply state that uncertainty is too great to predict human pharmacokinetics than to report a human PK projection that is not likely to be correct. PBPK modeling can be used as a tool to understand the quantitative impact of uncertainty from key knowledge gaps and also helps to minimize uncertainty due to species differences in PK properties.

In its best form, the human PK extrapolation is an integration of all preclinical data (e.g., *in vivo* data for the rodent and nonrodent, and *in vitro* data in the corresponding preclinical species as well as human) to provide a scientifically defensible estimate (extrapolation) of the likely pharmacokinetics of a compound in humans. Luttringer et al. [90] recently examined the ability of PBPK modeling to predict human pharmacokinetics based on preclinical species, using epiroprim as a test compound due to significant species differences in its PK properties. By incorporating information on species differences in PK properties from *in vitro* studies (e.g., in protein binding, the blood-to-plasma ratio, and intrinsic clearance in hepatocytes), PBPK modeling was able to reduce the uncertainty inherent in interspecies extrapolation and to provide a better prediction of human pharmacokinetics than allometric scaling or by direct scaling of hepatocyte data.

Recently, Jones et al. [13] reported similar results in a comparison between empirical and physiologically based approaches for the human PK extrapolation. Human pharmacokinetics of 19 compounds that had entered into humans was predicted using Dedrick plot analysis and PBPK modeling. A strategy for predicting human pharmacokinetics by focusing on methods for understanding the effects of absorption, distribution, and elimination from a physiological perspective was

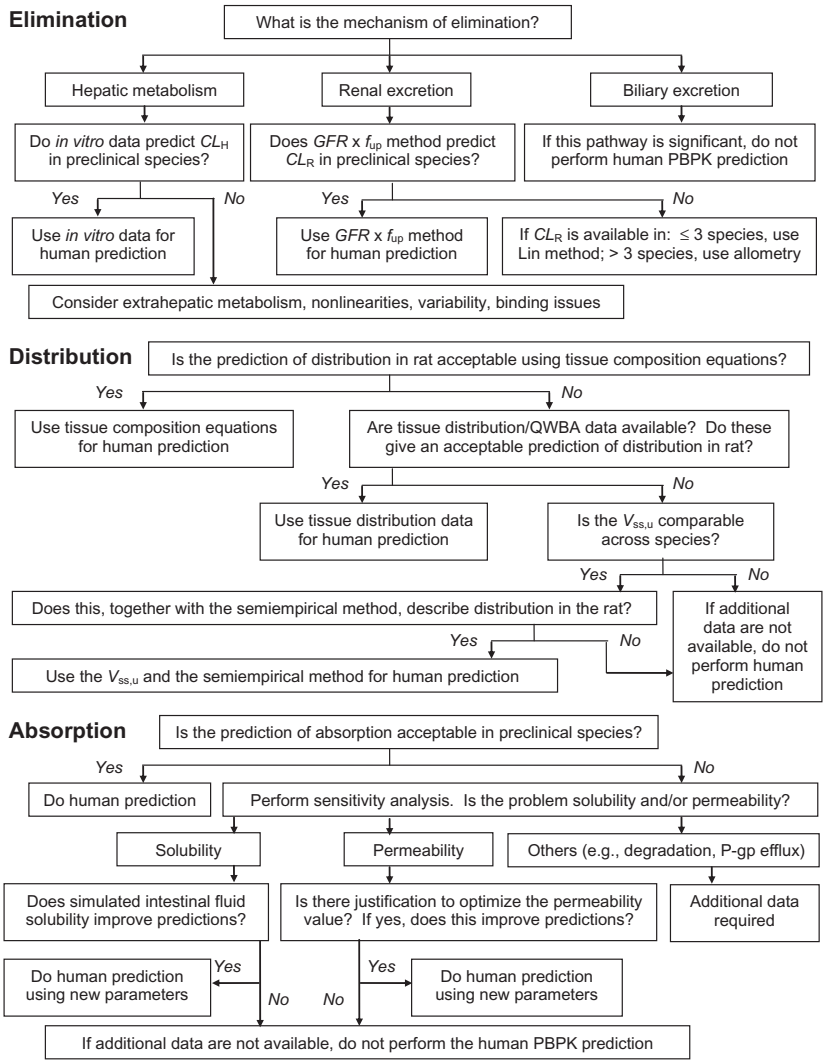


**FIGURE 4** A strategy for applying PBPK modeling to predict human pharmacokinetics prior to FIH. $CL_H$, hepatic clearance; $CL_R$, renal clearance; $f_{up}$, unbound fraction in plasma; *GFR*, glomerular filtration rate; QWBA, quantitative whole-body autoradiography; $V_{ss,u}$, steady-state volume of distribution of unbound drug. (Adapted from Ref. 13.)

developed (Fig. 4). The ability of PBPK modeling to predict pharmacokinetics in preclinical species was tested, and the human pharmacokinetics was only predicted if the model could predict pharmacokinetics in preclinical species. Based on this criterion, for 70% of the 19 compounds included in the study, a human PK prediction could be made. Jones et al. [13] noted: "The prediction accuracy for these compounds in terms of the percentage of compounds with an average-fold error of <2-fold was 83%, 50%, 75%, 67%, 92% and 100% for apparent oral clearance (CL/F), apparent volume of distribution during terminal phase after oral administration

(V-z/F), terminal elimination half-life (t1/2), peak plasma concentration (C-max), area under the plasma concentration-time curve (AUC) and time to reach C-max (t(max)), respectively." Simulated plasma concentrations and PK parameters were compared with human data so that the accuracy of the prediction methods could be assessed. The human PK data were more accurately predicted by the PBPK modeling approach than by the empirical approach. Using the strategy proposed by Jones et al. [13], the model development guides the strategy for gathering the data necessary for a complete understanding of likely human PK behavior.

Poor predictions with PBPK models are often a result of incomplete knowledge resulting in processes not correctly incorporated into the model (e.g., for biliary clearance and enterohepatic recirculation there are limited options for quantitatively predicting the effects on human pharmacokinetics). For such situations, an alternative method of performing the human PK extrapolation (e.g., allometric scaling) might seem attractive. However, allometric scaling is based on the assumption that clearance is proportional to $BW^{0.75}$ and that steady-state volume of distribution is proportional to $BW^{1}$ due to physiological properties (i.e., the same physiological properties that are incorporated in PBPK models). If PBPK modeling does not appear to be a good method for a compound, there is no reason to believe that allometric scaling or other empirical methods will work either. Additionally, developing methods for including increasingly complex PK mechanisms (e.g., first-pass metabolism and transporters in the gut, EH cycling, biliary excretion, and multiple pathways of elimination) is an active area of research, and methods for predicting human pharmacokinetics even with PK complications will soon be possible.

## 6 EXAMPLES OF THE USE OF PBPK MODELING IN DRUG DEVELOPMENT

In this section, two examples are provided of successful PBPK models for drugs that have been developed in the past: methotrexate (a cancer therapeutic) and all-*trans*-retinoic acid (a cancer therapeutic and skin treatment). Two areas in which PBPK modeling will become increasingly important in the future are then discussed: evaluation of PK variability and modeling of pharmacodynamics.

### 6.1 Methotrexate

The PBPK model for methotrexate (MTX) provides an excellent example of the iterative nature of the process of developing a PBPK model for a drug. The stated rationale for the approach used was "basing the model, as much as possible, on established, independently verifiable, physiological concepts" [5]. For the initial model development [91], a single IV dose of 3 mg/kg, which falls in the range of clinical treatment levels, was evaluated in CDF-1 male mice. Key observations regarding the behavior of the kinetic data included an unusually large initial drop in the plasma MTX concentration, indicating rapid uptake, tissue localization, and excretion (MTX is not metabolized extensively in mice). A subsequent asymptotic leveling of the concentration curves for all tissues was taken as evidence for zero-order uptake of MTX from the gut. A rapid early increase in the concentration of

MTX in the small intestine, with a peak gut lumen:plasma concentration ratio of about 100, demonstrated the importance of enterohepatic recycling.

Based on these observations, an initial MTX model was developed that included five compartments: plasma, muscle, liver, kidneys, and GI tract tissue/lumen [91]. The authors also derived reduced model structures with three compartments (plasma/body, liver, gut lumen) or two compartments (plasma/body plus gut lumen) to expedite solution of the associated system of equations. Flow-limited transport was assumed in all versions of the model, except as described below. The model incorporated a time-delay term to account for bile transport and holding time. Biliary clearance rates were estimated by bile duct cannulation experiments. To account for GI travel time and fecal excretion delays, time-delay functions were also applied in the gut lumen. Renal clearance was determined by comparison of the integrated plasma concentration data with cumulative urinary excretion data. Plasma protein binding was estimated at 25%. Comparison of experimental data and model predictions showed good agreement in general, although the model overpredicted gut lumen concentrations for times less than 60 minutes, probably due to the simple time-delay description of biliary transport.

In order to refine their initial model, Bischoff et al. [5] studied a variety of dose levels of MTX in several species, including humans. The additional data reproduced the initial data's features of (1) a rapid drop in MTX plasma concentrations and a corresponding increase in gut lumen concentrations, indicating the importance of tissue uptake and biliary secretion; (2) a peak gut lumen:plasma ratio of approximately 100; and (3) linear binding of MTX in tissues as plasma concentrations increased above 0.1 μg/mL, indicating nonlinear binding at low concentrations. The authors speculated that this phenomenon was probably due to strong binding to dihydrofolate reductase.

A revised PBPK model (Fig. 2) was developed to describe the more extensive data. Renal clearance of MTX was determined by comparing the time integral of plasma concentration with cumulative urine formation after intraperitoneal (IP) or IV administration. Tissue:plasma equilibrium ratios were derived from constant infusion experiments and/or the portion of the IV pulse injection curve after initial redistribution. Constant tissue:plasma equilibrium ratios at high concentrations indicated linear binding, while at lower concentrations, equilibrium ratios were represented by the sum of linear nonspecific binding and strong binding (presumed to be associated with dihydrofolate reductase).

An important feature of the revised model was its use of multicompartment submodels for biliary secretion and GI transport (Fig. 2). The original model used a mathematical time-delay (step function) to simulate bile formation and secretion from the liver. The abrupt introduction of bile into the GI tract lumen at the delay time caused the model to overshoot the data in this time period. In the revised model, biliary secretion was represented by a three-compartment submodel in which a series of discrete compartments was able to produce a smoother "S-shaped" bile concentration efflux curve.

MTX transit through the GI tract was modeled using an approach similar to that used for biliary secretion. Due to the tubular flow nature of the gut lumen, an assumption of uniform mixing would not predict the correct time course for concentrations appearing in the feces. Therefore, the GI tract was divided into four distinct regions with the feces exiting from the last compartment. An assumption of

zero-order absorption from the GI tract that was adequate for the single-dose level described in the initial modeling was not adequate for the wider range of dose levels investigated in the follow-up study. The revised model therefore provided for both saturable and nonsaturable GI tract absorption. In the absence of detailed data indicating otherwise, absorption characteristics were assumed to be the same for all regions, but specific location-dependent absorption characteristics could be included using this approach.

Comparison of model predictions and experimental data demonstrated that a single model structure was able to describe MTX distribution and excretion reasonably well for the mouse, rat, dog, and human, using species-specific parameters. The greatest uncertainty in the model was in the kinetics describing intestinal absorption. The authors concluded that GI tract absorption was not readily saturable, and both saturable and nonsaturable absorption should be included in the model. Based on the observed kinetic behavior at very low doses, the authors concluded that strong, saturable nonlinear binding occurred in the liver and kidney, probably related to binding with dihydrofolate reductase.

### 6.2 All-*trans*-Retinoic Acid

A PBPK model for all-*trans*-retinoic acid (ATRA; tretinoin) was developed in order to provide a coherent description of the absorption, distribution, metabolism, and excretion of the compound and its metabolites across species and routes of administration [19]. The goal of developing such a model was to provide a more biologically relevent dose measure than administered dose for assessing the human teratogenic risk from ATRA.

The PBPK model developed for ATRA provided a full physiological description for ATRA, with compartments for plasma, liver, gut, intestinal lumen, fat, skin, richly perfused tissues, slowly perfused tissues, placenta, and embryo (Fig. 3). Both oxidation (to the 4-oxo derivative) and glucuronidation of ATRA were described with saturable kinetics. Conversion to the 13-*cis* isomer (13-*cis*-RA; isotretinoin) and the subsequent metabolism of that compound were also included (not shown). Simpler compartmental descriptions were used for the metabolites, since there was no evidence that they preferentially partitioned into any of the body tissues. A third metabolic pathway, side chain oxidation producing $CO_2$, was also included. Oral uptake was described by zero-order stomach emptying and first-order uptake from the intestinal lumen. Dermal uptake was described by a two-compartment model, with compartments representing the stratum corneum and viable epidermis. Distribution was assumed to be flow limited except for dermal uptake and transplacental transfer, which were assumed to be diffusion limited. Excretions into urine and feces were modeled as first-order processes, with all chemicals being excreted in the feces and only glucuronides being excreted in the urine. Enterohepatic recirculation of ATRA and its metabolites was also described.

The physiological parameters for adult animals were obtained from Brown et al. [30]. Gestational parameters were based on previous modeling of rats [92, 93]. Partition coefficients were based on distribution studies with mice [94] and with a human placenta [95]. The ATRA volume of distribution calculated using these partitions, 1.1 L/kg, agrees with measured values. Biliary excretion was adjusted on the basis of data from rats with exteriorized bile ducts [96, 97]. The metabolic

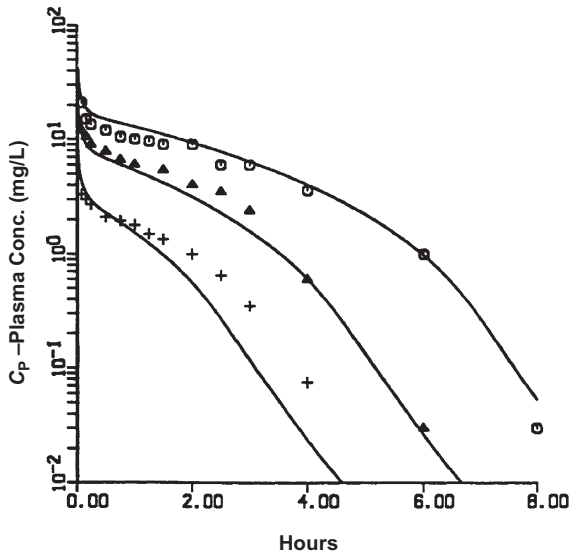


**FIGURE 5** Simulation of plasma concentration of all-*trans*-retinoic acid in the plasma of nonhuman primates following intravenous (IV) dosing at 5, 2.5, and 1 mg/kg.

parameters were adjusted using data for IV administration of ATRA in rats [97] and monkeys [98], as well as from metabolite measurements [96, 99]. A key difference in the metabolism of ATRA across species is the predominance of the oxidative pathway in the rodent [96, 97], in contrast with the predominance of the glucuronide pathway in the primate [99].

The PBPK model for IV dosing of ATRA in the monkey (Fig. 5) was scaled allometrically to predict the kinetics for oral dosing of human leukemia patients with 1.1 mg/kg ATRA [100]. The only parameters adjusted in this extrapolation were those describing the rate of oral uptake. All other parameters were calculated from those for the monkey. The resulting model provided an excellent description of the kinetics for humans with oral dosing (Fig. 6). It was also able to reproduce the observed kinetics for total radioactivity following topical administration of ATRA (Fig. 7). In contrast to the kinetic data, which could only provide an estimate of total exposure to ATRA plus its metabolites, the PBPK model was able to provide separate estimates of internal exposure to ATRA and its active and inactive metabolites. An important result of the modeling was to demonstrate that the low dose rate associated with dermal exposure results in high efficiency clearance of ATRA to its inactive glucuronide metabolites, as opposed to oral dosing, where the higher dose rate leads to a greater proportion of active oxidative metabolites.

The model was used to simulate oral clinical dosing with ATRA, as well as minimal teratogenic doses of ATRA in the primate and rodent. Based on literature data on the teratogenic potential of the various chemical species, the most appropriate dose surrogate could be either $C_{max}$ or *AUC* for the total concentration of active retinoids (ATRA + 4-oxo-ATRA + 13-*cis*-RA + 4-oxo-13-*cis*-RA). The glucuronides are not included in this dose measure since they do not cross the placenta. On

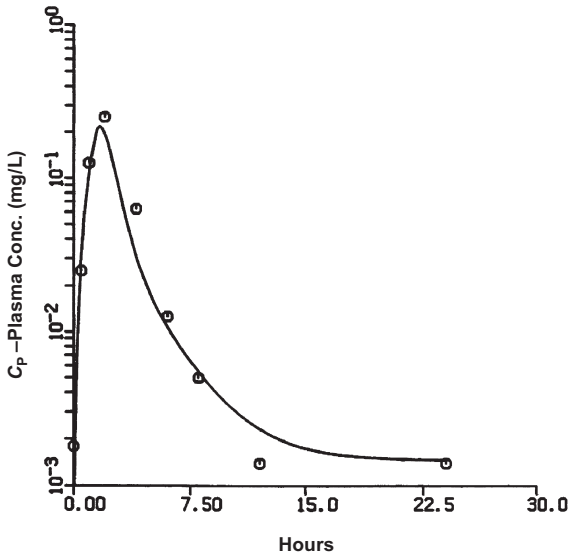


**FIGURE 6** Simulation of all-*trans*-retinoic acid concentration in plasma of human subject receiving an oral dose of 1.1 mg/kg.

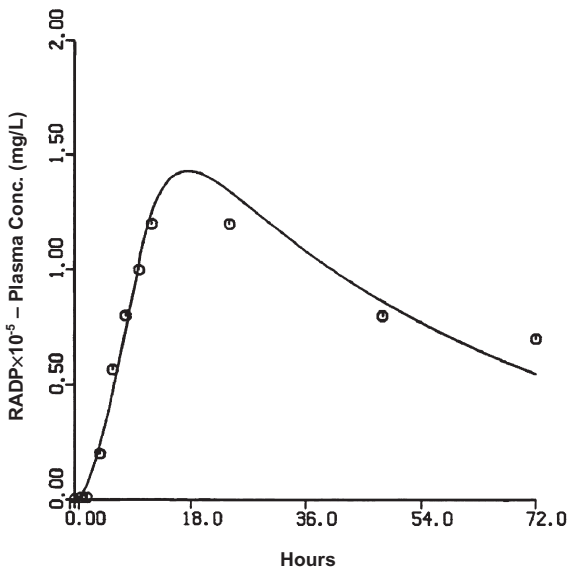


**FIGURE 7** Simulation of total radiation in plasma following topical treatment of human subject with 50 micrograms of all-*trans*-retinoic acid.

the other hand, based on activity at retinoic acid receptors it could be argued that the most appropriate dose surrogate could be the peak concentration or area under the curve for ATRA alone. In either case, maternal plasma levels were used as a surrogate for fetal levels, based on evidence from animal studies that concentrations achieved in maternal and fetal plasma are similar. The calculated

**TABLE 4 Comparison of Dose Surrogates for Retinoic Acid Teratogenicity**

| | | | All-*trans*-Retinoic Acid | | Total Active Retinoids | | Total Retinoids |
|---|---|---|---|---|---|---|---|
| | Route | Dose (mg/kg) | $C_{max}$ (ng/mL) | *AUC* (ng·h/mL) | $C_{max}$ (ng/mL) | *AUC* (ng·h/mL) | $C_{max}$ (ng·eq/mL) |
| **Minimal teratogenic doses** (all-*trans*-RA) | | | | | | | |
| Mouse | Oral | 4.0 | 1,048 | 1,852 | 2,681 | 12,658 | 3,107 |
| Rat | Oral | 2.5 | 943 | 2,029 | 1,918 | 13,554 | 2,051 |
| Monkey | Oral | 5.0 | 1,830 | 3,714 | 2,294 | 5,962 | 4,983 |
| **Clinical doses** (human) | | | | | | | |
| ATRA | Oral | 1.1 | 218 | 695 | 264 | 1,088 | 1,237 |
| 13-*Cis* | Oral | 1.1 | 654[a] | 3,154[a] | 1,033 | 7,103 | 1,148 |

[a]Reflects concentration of 13-*cis*-RA rather than all-*trans*-RA.

dose surrogates, as well as the estimated peak plasma level for total retinoids (which includes the glucuronides), for several species are shown in Table 4.

One test of a useful dose surrogate is that similar values should be calculated at similar effect levels across species. For a minimal teratogenic effect, all of the dose surrogates in Table 4 (including administered dose) are within a factor of 2–3 across species. The most consistent dose surrogate is the $C_{max}$ for total active retinoids, which is essentially constant across species. Based on this dose measure, the internal exposure of patients receiving oral ATRA treatment for cancer is below the threshold for teratogenic effects by about a factor of 7–10. However, this comparison assumes that the maternal plasma concentration profile is representative of fetal exposure. If one takes into consideration the longer period of organogenesis in the human (around 35 days) as compared to the rodent (around 10 days) and assumes, as a worst case, fetal exposure at the maximum maternal concentration throughout the entire period, the margin of safety could be as low as 2–3. It is of interest that the kinetics of 13-*cis*-RA in the human are considerably different from those of ATRA. 13-*cis*-RA has a much longer half-life than ATRA in the human and oxidation, rather than glucuronidation, is the dominant form of metabolism. The calculated plasma concentrations and *AUC* values for total active retinoids following oral treatment with 13-*cis*-RA are in the same range as those causing teratogenic effects in animals. This result is consistent with the observation of teratogenic effects associated with the human use of 13-*cis*-RA (isotretinoin, Table 4). This PBPK model was used by the FDA in its evaluation of the safety of a topical skin treatment containing ATRA.

### 6.3 Future Directions

***Interindividual Variability*** Because of the heterogeneity of the human population, it is generally expected that there will be a broad range of responses to the biological effects of drugs. This heterogeneity is produced by interindividual variations in physiology, biochemistry, and molecular biology, reflecting both genetic and environmental factors, and results in differences among individuals in the biologically effective tissue dose associated with a given administered dose

(pharmacokinetics) as well as in the response to a given tissue dose (pharmacodynamics). Often it is possible to distinguish specific classes of individuals, such as infants or the elderly, who appear to be more susceptible to a specific effect. PBPK modeling provides the capability to quantitatively describe the potential impact of PK factors on the variability of individual response. Because the parameters in a PBPK model have a direct biological correspondence, they provide a useful framework for determining the impact of observed variations in physiological and biochemical factors on the population variability in dosimetry.

For example, PBPK models can be used to determine the impact of differences in key metabolism enzymes, whether due to multiple genotypic expression, such as polymorphisms, or just due to normal variation in enzyme activities within the general population. Other potential modulators of sensitivity that can be addressed quantitatively with a PBPK model include physical condition, level of activity, disease states, age, hormonal status, and interactions with other chemicals and drugs. In each case, the PBPK model provides a quantitative structure for determining the effect of these various factors on the relationship between the administered dose and the internal (biologically effective) target tissue exposure. When coupled with Monte Carlo analysis, the PBPK model provides a method to assess the quantitative impact of these sources of variability on individual response (as opposed to average population response) by comparing model predicted internal doses over the distribution of individual parameter values.

It is useful to consider the total variability among humans in terms of three contributing sources: (1) the variation across a population of "normal" individuals at the same age (e.g., young adults); (2) the variation across the population resulting from their different ages (e.g., infants or the elderly); and (3) the variation resulting from the existence of subpopulations that differ in some way from the "normal" population (e.g., due to genetic polymorphisms). A fourth source of variability, health status, should also be considered. To the extent that the variation in physiological and biochemical parameters across these population dimensions can be elucidated, PBPK models can be used together with Monte Carlo methods to integrate their effects on the *in vivo* kinetics of a compound and predict the resulting impact on the distribution of responses (as represented by target tissue doses) across the population.

There has sometimes been a tendency in drug development to use information on the variability of a specific parameter, such as the *in vitro* activity of a particular enzyme, as the basis for expectations regarding the variability in dosimetry for *in vivo* exposures. However, whether or not the variation in a particular physiological or biochemical parameter will have a significant impact on *in vivo* dosimetry is a complex function of interacting factors. In particular, the structures of physiological and biochemical systems frequently involve parallel processes (e.g., blood flows, metabolic pathways, excretion processes), leading to compensation for the variation in a single factor. Moreover, physiological constraints may limit the *in vivo* impact of variability observed *in vitro*. For instance, high affinity intrinsic clearance can result in essentially complete metabolism of all the compound reaching the liver in the blood; under these conditions, variability in amount metabolized *in vivo* would be more a function of variability in liver blood flow than variability in metabolism *in vitro*. Thus, it is often true that the whole (the *in vivo* variability in dosimetry) is less than the sum of its parts (the variability in each of the PK factors).

The overall PK variability across a population is a function of many compound-specific, genetic, and physiological factors. Due to the complex interactions among these factors, speculation regarding the extent of population variability on the basis of the observed variation in a single factor can be highly misleading. Analysis using PBPK modeling and Monte Carlo techniques provides a more reliable approach for estimating population PK variability. PBPK modeling can also be useful in a more qualitative sense, to determine whether there is reason for concern regarding a particular age-group that might be more sensitive due to PK differences. Similar analyses can be performed to determine whether exposure during special life stages, such as gestation or lactation, represents a significant concern.

***Modeling of Pharmacodynamics*** The growing popularity of the PBPK modeling approach represents a movement from simpler kinetic models toward more biologically realistic descriptions of the determinants that regulate disposition of drugs in the body. To a large extent, the application of these PBPK models to study the time courses of compounds in the body is simply an integrated systems approach to understanding the biological processes that regulate the delivery of drugs to target sites. Many PBPK models integrate information across multiple levels of organization, especially when describing interactions of compounds with molecular targets, such as reversible binding of ligands to specific receptors, as in the case of methotrexate binding to dihydrofolate reductase [5, 101]. In such cases, the PBPK models integrate molecular, cellular, organ level, and organism-level processes to account for the time courses of compounds, metabolites, and bound complexes within organs and tissues in the body. The system under scrutiny in PBPK models is more the integrated physiological system, appropriate enough for a discipline defined as physiologically based modeling.

The main goal of PBPK models is quite simple—to predict the target tissue dose of compounds and their metabolites at target tissues and, in some cases, to describe interactions in target tissues. PBPK models once developed are extensible. They can be used to extrapolate to various other conditions because of their biological fidelity. While the goal in applying these models is to predict dosimetry, it is important to remember that the overall goal of using PBPK modeling in efficacy and safety assessment with drugs is broader than simply estimating tissue dose, regardless of the level of detail provided in the interactions of compounds with tissue constituents. The goal in the larger context is to understand the relationship between dose delivered to target tissues and the biological sequelae of the exposure of target tissues to compounds. The specific steps that lead from these dose metrics to tissue, organ, and organism-level responses have usually been considered part of the pharmacodynamic (PD) process. In general, PD models used in drug evaluation have been more empirical, utilizing simple effect compartments correlated with blood or tissue concentrations of active compound. Another inexorable development will be expansion of the systems approaches into the PBPD arena. This latter area will represent a systems biology approach for describing perturbations of biological systems by compounds and the exposure/dose conditions under which these perturbations become sufficiently large to pose significant health risks or sufficiently large to achieve specific therapeutic outcomes.

The systems biology approach (Fig. 8) focuses on normal biological function and the perturbations associated with exposure to compounds. Perturbations of

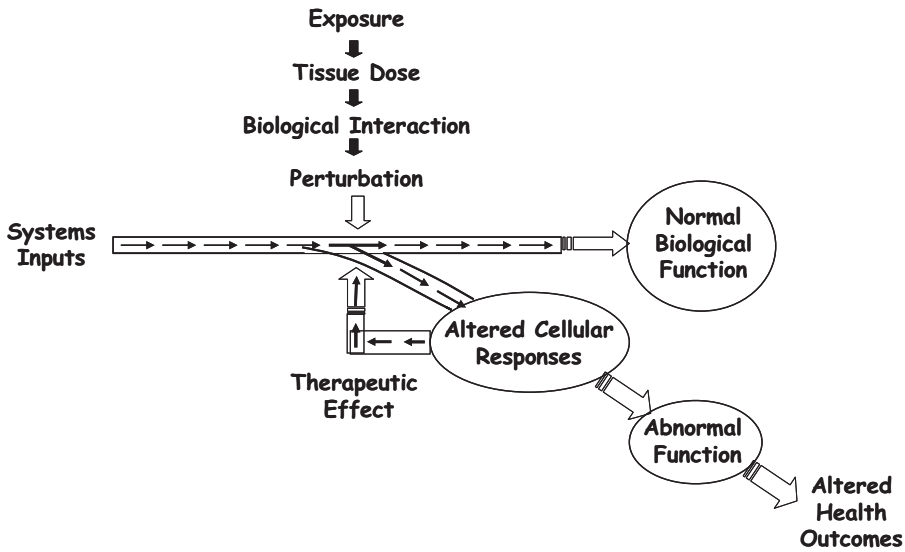


**FIGURE 8** Diagram of the approach for understanding the pharmacological and toxicological effects of compounds in terms of their interaction with the biological system. The horizontal axis represents the biological component of the interaction, and the vertical axis represents the chemical component. Drugs can both restore altered biological function (efficacy) and produce altered function (toxicity).

biological processes by compounds lead to either adverse responses (toxicity) or restoration of normal function to a compromised tissue (efficacy). The effects of compounds, whether for good or ill, can best be described by PBPK approaches linked through PBPD models of responses of cellular signaling networks. Toxicity and efficacy are then defined by an intersection of compound action with the biological system. Toxicology and pharmacology are disciplines at the interface of chemistry/pharmacokinetics (primarily embedded in the vertical component) and biology/pharmacodynamics (primarily captured by the horizontal chain). Clearly, the main differences in the next generation of systems approaches in PK and PD modeling will be the increasingly detailed descriptions of biology afforded by new technologies and the expansion of modeling tools available for describing the effects of compounds on biological signaling processes.

## 7 CONCLUSIONS AND FUTURE PERSPECTIVES

It is ironic that PBPK modeling of drugs now appears to lag behind modeling of environmental contaminants [102], since the concept of PBPK modeling was first described in the context of drug disposition, in the seminal work of Teorell [1, 2]. In the pharmaceutical industry, noncompartmental "model-free" PK analyses are the standard practice, despite the fact that these methods are useful only for summarizing the characteristics of a dataset (e.g., volume of distribution, mean retention time, terminal half-life) and do not provide the advantages inherent in PBPK modeling

(hypothesis testing, prediction, etc.). Their principal advantages are that they are rapid and inexpensive to perform.

However, in recent years there has been increasing pressure on the pharmaceutical industry to accelerate the drug development process, and PBPK modeling has repeatedly been identified as one of the technologies that could prove useful to this end [12, 102–104]. Because the collection of ADME data on a drug is a required element of drug development, much of the data necessary for developing a PBPK model is often available. The PBPK model offers the opportunity to make better use of these data, by serving as a structured repository for quantitative information on the compound, a conceptual framework for hypothesis testing, and a quantitative platform for prediction.

The rapid development of computational chemistry [105], genomics [106], and high throughput screening [107] has brought increasing attention to the discovery phase of drug development, including growing interest in "discovery toxicology" [108]. PBPK modeling can play a complementary role to two other technologies that are finding increasing use in drug discovery: QSAR analysis and genomics. QSAR can be used to estimate compound-specific parameters for the PBPK model, while genomic data can provide mode of action insights that drive model structure decisions. In this scenario, the PBPK model provides a quantitative biological framework for integrating the physicochemical characteristics of the drug candidate, together with *in vitro* data on its ADME and toxicity, within the constraints of the fundamental physiological and biochemical processes governing compound behavior *in vivo*. This approach is particularly effective when used consistently during drug development, because the information gained from modeling of previous candidate compounds can greatly facilitate model development for new compounds with similar structures or properties.

As the drug development process proceeds, the model can be iteratively informed and refined on the basis of the data being collected, and in turn the model can serve as a platform for *in silico* hypothesis testing and experimental design. As the model becomes more robust, it can also be employed for extrapolation (e.g., selection of first human dose) and for simulation (e.g., Monte Carlo analysis for design of clinical trials). The physiological structure of the PBPK model makes it a perfect platform for conducting evaluations of alternative dosing methods and regimens, and for investigating the impact of genetic-, age-, and disease-related differences in physiology and metabolism on drug kinetics. Of particular importance, by using a PBPK model the pharmacodynamic effects of a drug can be investigated more directly, relating the effects to the concentration in the tissue (e.g., the brain) where the compound interacts with the biological system, rather than attempting to elucidate a potentially indirect relationship with a concentration in the central (e.g., plasma) compartment. By obtaining quantitative information on the dose–responses for both the efficacy and toxicity of the compound, the PBPK model can be exercised to evaluate the potential to increase the efficacy/toxicity ratio of the drug through manipulation of dose rate and dose route using novel drug delivery systems. These and other attributes of PBPK models for organizing and interpreting diverse datasets, with the specific goals of understanding efficacy and toxicity, are reviving interest in applying these tools in drug development and evaluation [109–113]. The availability of increasingly sophisticated tools that can make predictions even with

early preclinical data will encourage increasing application of PBPK modeling in preclinical development.

## ACKNOWLEDGMENTS

The authors thank Dr. Carl Brown, Dr. Werner Rubas, and Dr. Phil Worboys for reviewing the manuscript for this chapter and providing valuable discussion and insight.

## REFERENCES

1. Teorell T. Kinetics of distribution of substances administered to the body. I. The extravascular mode of administration. *Arch Int Pharmacodyn* 1937;57:205–225.
2. Teorell T. Kinetics of distribution of substances administered to the body. I. The intravascular mode of administration. *Arch Int Pharmacodyn* 1937;57:226–240.
3. Andersen ME. Saturable metabolism and its relation to toxicity. *Crit Rev Toxicol* 1981;9:105–150.
4. Bischoff KB, Brown RG. Drug distribution in mammals. *Chem Eng Prog Symp Ser* 1966;62(66):33–45.
5. Bischoff KB, Dedrick RL, Zaharko DS, Longstreth JA. Methotrexate pharmacokinetics. *J Pharm Sci* 1971;60:1128–1133.
6. Collins JM, Dedrick RL, Flessner MF, Guarino AM. Concentration dependent disappearance of fluorouracil from peritoneal fluid in the rat: experimental observations and distributed modeling. *J Pharm Sci* 1982;71:735–738.
7. Farris FF, Dedrick RL, King FG. Cisplatin pharmacokinetics: applications of a physiological model. *Toxicol Lett* 1988;43:117–137.
8. Andersen ME, Yang RSH, Clewell HJ, Reddy MB. Introduction: a historical perspective of the development and application of PBPK models. In Reddy MB, Yang RSH, Clewell HJ III, Andersen ME, Eds. *Physiologically Based Pharmacokinetic Modeling: Science and Applications*. Hoboken, NJ: John Wiley & Sons; 2005, pp 1–18.
9. McMullin TS. Drugs. In Reddy MB, Yang RSH, Clewell HJ III, Andersen ME, Eds. *Physiologically Based Pharmacokinetic Modeling: Science and Applications*. Hoboken, NJ: John Wiley & Sons; 2005, pp 273–296.
10. Pott O'Brien W. Antineoplastic agents. In Reddy MB, Yates RSH, Clewell HJ III, Andersen ME, Eds. *Physiologically Based Pharmacokinetic Modeling: Science and Applications*. Hoboken, NJ: John Wiley & Sons; 2005, pp 297–317.
11. Lupfert C, Reichel A. Development and application of physiologically based pharmacokinetic-modeling tools to support drug discovery. *Chem Biodiversity* 2005;2:1462–1486.
12. Nestorov I. Whole body pharmacokinetic models. *Clin Pharmacokinet* 2003;42:883–908.
13. Jones HM, Parrott N, Jorga K, Lave T. A novel strategy for physiologically based predictions of human pharmacokinetics. *Clin Pharmacokinet* 2006;45:511–542.
14. Parrott N, Jones H, Paquereau N, Lavé T. Application of full physiological models for pharmaceutical drug candidate selection and extrapolation of pharmacokinetics to man. *Basic Clin Pharmacol Toxicol* 2005;96:193–199.

15. Baxter LT, Zhu H, Mackensen DG, Jain RK. Physiologically-based pharmacokinetic model for specific and nonspecific monoclonal-antibodies and fragments in normal-tissues and human tumor xenografts in nude-mice. *Cancer Res* 1994;54:1517–1528.
16. Baxter LT, Zhu H, Mackensen DG, Butler WF, Jain RK. Biodistribution of monoclonal-antibodies—scale-up from mouse to human using a physiologically-based pharmacokinetic model. *Cancer Res* 1995;55:4611–4622.
17. Aarons L. Editors' view. Physiologically based pharmacokinetic modeling: a sound mechanistic basis is needed. *Br J Clin Pharm* 2005;60(6):581–583.
18. Dedrick RL, Bischoff KB. Species similarities in pharmacokinetics. *Fed Proc* 1980; 39:54–59.
19. Clewell HJ, Andersen ME, Wills RJ, Latriano L. A physiologically based pharmacokinetic model for retinoic acid and its metabolites. *J Am Acad Dermatol* 1997;36: S77–S85.
20. Reddy MB, Yang RSH, Clewell HJ III, Andersen ME, Eds. *Physiologically Based Pharmacokinetic Modeling: Science and Applications*. Hoboken, NJ: John Wiley & Sons; 2005.
21. Brightman FA, Leahy DE, Searle GE, Thomas S. Application of a generic physiologically based pharmacokinetic model to the estimation of xenobiotic levels in rat plasma. *Drug Metab Dispos* 2006;34:84–93.
22. Brightman FA, Leahy DE, Searle GE, Thomas S. Application of a generic physiologically based pharmacokinetic model to the estimation of xenobiotic levels in human plasma. *Drug Metab Dispos* 2006;34:94–101.
23. Malaty LI, Kuper JJ. Drug interactions of HIV. Protease inhibitors. *Drug Safety* 1999;20:147–169.
24. Adolph EF. Quantitative relations in the physiological constitutions of mammals. *Science* 1949;109:579–585.
25. Dedrick RL. Animal scale-up. *J Pharmacokinet Biopharm* 1973;1:435–461.
26. Astrand P, Rodahl K. *Textbook of Work Physiology*. New York: McGraw-Hill; 1970, pp 157–160, 206–211.
27. International Commission on Radiological Protection (ICRP). Report of the Task Group on Reference Man. ICRP Publication 23; 1975, pp 228–237, 280–285, 325–327.
28. Environmental Protection Agency (EPA). Reference physiological parameters in pharmacokinetic modeling. EPA/600/6–88/004. Washington, DC: Office of Health and Environmental Assessment; 1988.
29. Davies B, Morris T. Physiological parameters in laboratory animals and humans. *Pharm Res* 1993;10:1093–1095.
30. Brown RP, Delp MD, Lindstedt SL, Rhomberg LR, Beliles RP. Physiological parameter values for physiologically based pharmacokinetic models. *Toxicol Indust Health* 1997;13(4):407–484.
31. Jepson GW, Hoover DK, Black RK, McCafferty JD, Mahle DA, Gearhart JM. A partition coefficient determination method for nonvolatile chemicals in biological tissues. *Fundam Appl Toxicol* 1994;22:519–524.
32. Clewell HJ. Coupling of computer modeling with *in vitro* methodologies to reduce animal usage in toxicity testing. *Toxicol Lett* 1993;68:101–117.
33. Yates FE. Good manners in good modeling: mathematical models and computer simulations of physiological systems. *Am J Physiol* 1978;234:R159–R160.
34. Carson ER, Cobelli C, Finkelstein L. *The Mathematical Modeling of Metabolic and Endocrine Systems. Model Formulation, Identification, and Validation*. Hoboken, NJ: John Wiley & Sons; 1983, pp 23–45, 113–127, 217–231.

35. Rescigno A, Beck JS. Perspectives in pharmacokinetics. The use and abuse of models. *J Pharmacokinet Biopharm* 1987;15(3):327–344.
36. Himmelstein KJ, Lutz RJ. A review of the application of physiologically based pharmacokinetic modeling. *J Pharmacokinet Biopharm* 1979;7:127–145.
37. Gerlowski LE, Jain RK. Physiologically based pharmacokinetic modeling: principles and applications. *J Pharm Sci* 1983;72:1103–1126.
38. Nestorov IA, Aarons LJ, Arundel PA, Rowland M. Lumping of whole-body physiologically based pharmacokinetic models. *J Pharmacokinet Biopharm* 1998;26(1):21–46.
39. Clewell HJ, Lee T, Carpenter RL. Sensitivity of physiologically based pharmacokinetic models to variation in model parameters: methylene chloride. *Risk Anal* 1994;14: 521–531.
40. Campain J. Metals and inorganic compounds. In Reddy MB, Yang RSH, Clewell HJ III, Andersen ME, Eds. *Physiologically Based Pharmacokinetic Modeling: Science and Applications*. Hoboken, NJ: John Wiley & Sons; 2005, pp 239–270.
41. Lin JH. Applications and limitations of interspecies scaling and *in vitro* extrapolation in pharmacokinetics. *Drug Metab Dispos* 1998;26:1202–1212.
42. Hinderling PH. Red blood cells: a neglected compartment in pharmacokinetics and pharmacodynamics. *Pharmacol Rev* 1997;49:279–295.
43. Andersen ME, Clewell HJ, Gargas ML, Smith FA, Reitz RH. Physiologically-based pharmacokinetics and the risk assessment for methylene chloride. *Toxicol Appl Pharmacol* 1987;87:185–205.
44. Agoram B, Woltosz WS, Bolger MB. Predicting the impact of physiological and biochemical processes on oral drug bioavailability. *Adv Drug Deliv Rev* 2001;50(Suppl 1): S41–S67.
45. Willmann S, Schmitt W, Keldenich J, Dressman JB. A physiologic model for simulating gastrointestinal flow and drug absorption in rats. *Pharm Res* 2003;20:1766–1771.
46. Semino G, Lilly P, Andersen ME. A pharmacokinetic model describing pulsatile uptake of orally-administered carbon tetrachloride. *Toxicology* 1997;117(1):25–33.
47. Chaturvedi PR, Decker CJ, Odinecs A. Prediction of pharmacokinetic properties using experimental approaches during early drug discovery. *Curr Opin Chem Biol* 2001;5(4): 452–463.
48. Cai H, Stoner C, Reddy A, Freiwald S, Smith D, Winters R, Stankovic C, Surendran N. Evaluation of an integrated *in vitro–in silico* PBPK (physiologically based pharmacokinetic) model to provide estimates of human bioavailability. *Int J Pharm* 2006;308(1–2): 133–139.
49. Dennison JE, Andersen ME. Reactive vapors in the nasal cavity. In Reddy MB, Yang RSH, Clewell HJ III, Andersen ME, Eds. *Physiologically Based Pharmacokinetic Modeling: Science and Applications*. Hoboken, NJ: John Wiley & Sons; 2005, pp 375–387.
50. Reddy MB. PBPK modeling approaches for special applications: dermal exposure models. In Reddy MB, Yang RSH, Clewell HJ III, Andersen ME, Eds. *Physiologically Based Pharmacokinetic Modeling: Science and Applications*. Hoboken, NJ: John Wiley & Sons; 2005, pp 375–387.
51. McCarley KD, Bunge AL. Pharmacokinetic models of dermal absorption. *J Pharm Sci* 2001;90(11):1699–1719.
52. Roberts MS, Anissimov YG, Gonsalvez RA. Mathematical models in percutaneous absorption (Reprinted from *Percutaneous Adsorption* 1999;3–55). *J Toxicol-Cutaneous Ocular Toxicol* 2001;20:221–270.

53. Andersen ME, Mills JJ, Gargas ML, Kedderis L, Birnbaum LS, Neubert D, Greenlee WF. Modeling receptor-mediated processes with dioxin: implications for pharmacokinetics and risk assessment. *Risk Anal* 1993;13(1):25–36.
54. Kohn MC, Lucier GW, Clark GC, Sewall C, Tritscher AM, Portier CJ. A mechanistic model of effects of dioxin on gene expression in the rat liver. *Toxicol Appl Pharmacol* 1993;120(1):138–154.
55. Plowchalk DR, Teeguarden J. Development of a physiologically based pharmacokinetic model for estradiol in rats and humans: a biologically motivated quantitative framework for evaluating responses to estradiol and other endocrine-active compounds. *Toxicol Sci* 2002;69(1):60–78.
56. Mendel CM. The free hormone hypothesis: distinction from the free hormone transport hypothesis. *J Androl* 1992;13(2):107–116.
57. Lam G, Chen M, Chiou WL. Determination of tissue to blood partition coefficients in physiologically-based pharmacokinetic studies. *J Pharm Sci* 1982;71(4):454–456.
58. Lin JH, Sugiyama Y, Awazu S, Hanano M. *In vitro* and *in vivo* evaluation of the tissue-to-blood partition coefficients for physiological pharmacokinetic models. *J Pharmacokinet Biopharm* 1982;10(6):637–647.
59. Pelekis M, Poulin P, Krishnan K. An approach for incorporating tissue composition data into physiologically based pharmacokinetic models. *Toxicol Ind Health* 1995;11: 511–522.
60. Poulin P, Krishnan K. A biologically-based algorithm for predicting human tissue—blood partition coefficients of organic chemicals. *Hum Exp Toxicol* 1995;14:273–280.
61. Poulin P, Krishnan K. An algorithm for predicting tissue:blood partition coefficients of organic chemicals from *n*-octanol:water partition coefficient data. *J Toxicol Environ Health* 1995;46:117–129.
62. Poulin P, Theil FP. *A priori* prediction of tissue:plasma partition coefficients of drugs to facilitate the use of physiologically-based pharmacokinetic models in drug discovery. *J Pharm Sci* 2000;89:16–35.
63. Poulin P, Theil F. Prediction of pharmacokinetics prior to *in vivo* studies. 1. Mechanism-based prediction of volume of distribution. *J Pharm Sci* 2002;91:129–156.
64. Rodgers T, Leahy D, Rowland M. Physiologically based pharmacokinetic modeling 1: Predicting the tissue distribution of moderate-to-strong bases. *J Pharm Sci* 2005;94: 1259–1276.
65. Rodgers T, Rowland M. Physiologically based pharmacokinetic modelling 2: Predicting the tissue distribution of acids, very weak bases, neutrals and zwitterions. *J Pharm Sci* 2006;95:1238–1257.
66. Zuegge J, Schneider G, Coassolo P, Lavé T. Prediction of hepatic metabolic clearance—comparison and assessment of prediction models. *Clin Pharmacokinet* 2001;40:553–563.
67. Lavé T, Coassolo P, Reigner B. Prediction of hepatic metabolic clearance based on interspecies allometric scaling techniques and *in vitro in vivo* correlations. *Clin Pharmacokinet* 1999;36:211–231.
68. Houston JB. Relevance of *in-vitro* kinetic-parameters to *in-vivo* metabolism of xenobiotics. *Toxicol In Vitro* 1994;8:507–512.
69. Houston JB, Carlile DJ. Prediction of hepatic clearance from microsomes, hepatocytes, and liver slices. *Drug Metab Rev* 1997;29:891–922.
70. Obach RS. Prediction of human clearance of twenty-nine drugs from hepatic microsomal intrinsic clearance data: an examination of *in vitro* half-life approach and nonspecific binding to microsomes. *Drug Metab Dispos* 1999;27:1350–1359.

71. Ito K, Iwatsubo T, Kanamitsu S, Nakajima Y, Sugiyama Y. Quantitative prediction of *in vivo* drug clearance and drug interactions from *in vitro* data on metabolism, together with binding and transport. *Annu Rev Pharmacol Toxicol* 1998;38:461–499.
72. Barter ZE, Bayliss MK, Beaune PH, Boobis AR, Carlile DJ, Edwards RJ, Houston JB, Lake BG, Lipscomb JC, Pelkonen OR, Tucker GT, Rostami-Hodjegan A. Scaling factors for the extrapolation of *in vivo* metabolic drug clearance from *in vitro* data: reaching a consensus on values of human microsomal protein and hepatocellularity per gram of liver. *Curr Drug Metab* 2007;8:33–45.
73. Reddy MB, Andersen ME, Morrow PE, Dobrev ID, Varaprath S, Plotzke KP, Utell MJ. Physiological modeling of inhalation kinetics of octamethylcyclotetrasiloxane in humans during rest and exercise. *Toxicol Sci* 2003;72:3–18.
74. Blau GE, Neely WB. Dealing with uncertainty in pharmacokinetic models using SIMUSOLV. In *Drinking Water and Health*, Vol. 8. Washington, DC: National Research Council; 1987, pp 185–207.
75. Cobelli C, Carson ER, Finkelstein L, Leaning MS. Validation of simple and complex models in physiology and medicine. *Am J Physiol* 1984;246:R259–R266.
76. Gueorguieva I, Nestorov IA, Rowland M. Reducing whole body physiologically based pharmacokinetic models using global sensitivity analysis: diazepam case study. *J Pharmacokinet Pharmacodyn* 2006;33(1):1–27.
77. Clewell HJ. The use of physiologically based pharmacokinetic modeling in risk assessment: a case study with methylene chloride. In Olin S, Farland W, Park C, Rhomberg R, Scheuplein L, Starr T, Wilson J, Eds. *Low-Dose Extrapolation of Cancer Risks: Issues and Perspectives*. Washington, DC: ILSI Press; 1995.
78. Clewell HJ, Andersen ME. Use of physiologically-based pharmacokinetic modeling to investigate individual versus population risk. *Toxicology* 1996;111:315–329.
79. Parrott N, Paquereau N, Coassolo P, Lavé T. An evaluation of the utility of physiologically based models of pharmacokinetics in early drug discovery. *J Pharm Sci* 2005;94: 2327–2343.
80. Parrott N, Lavé T. Prediction of intestinal absorption: comparative assessment of GASTROPLUS (TM) and IDEA (TM). *Eur J Pharm Sci* 2002;17:51–61.
81. Tucker GT, Houston JB, Huang SM. Optimizing drug development: strategies to assess drug metabolism/transporter interaction potential—toward a consensus. *Pharm Res* 2001;18:1071–1080.
82. Telgmann R, von Kleist M, Huisinga W. Software supported modelling in pharmacokinetics. *CompLife* 2006;216–225.
83. Leahy DE. Integrating *in vitro* ADMET data through generic physiologically based pharmacokinetic models. *Expert Opin Drug Metab Toxicol* 2006;2:619–628.
84. Lavé T, Dupin S, Schmitt C, Valles B, Ubeaud G, Chou RC, Jaeck D, Coassolo P. The use of human hepatocytes to select compounds based on their expected hepatic extraction ratios in humans. *Pharm Res* 1997;14:152–155.
85. Thompson TN. Early ADME in support of drug discovery: the role of metabolic stability studies. *Curr Drug Metab* 2000;1:215–241.
86. Poggesi I. Predicting human pharmacokinetics from preclinical data. *Curr Opin Drug Discov Dev* 2004;7:100–111.
87. Riley RJ, McGinnity DF, Austin RP. A unified model for predicting human hepatic, metabolic clearance from *in vitro* intrinsic clearance data in hepatocytes and microsomes. *Drug Metab Dispos* 2005;33:1304–1311.
88. Poulin P, Theil FP. Prediction of pharmacokinetics prior to *in vivo* studies. II. Generic physiologically based pharmacokinetic models of drug disposition. *J Pharm Sci* 2002;91:1358–1370.

89. Reigner BG, Blesch KS. Estimating the starting dose for entry into humans: principles and practice. *Eur J Clin Pharmacol* 2002;57:835–845.
90. Luttringer O, Theil F-P, Poulin P, Schmitt-Hoffmann AH, Guentert TW, Lavé T. Physiologically based pharmacokinetic (PBPK) modeling of disposition of epiroprim in humans. *J Pharm Sci* 2003;92:1990–2007.
91. Bischoff KB, Dedrick RL, Zaharko DS. Preliminary model for methotrexate pharmacokinetics. *J Pharm Sci* 1970;59:149–154.
92. Fisher JW, Whittaker TA, Taylor DH, Clewell H III, Andersen ME. Physiologically based pharmacokinetic modeling of the pregnant rat: a multiroute exposure model for trichloroethylene and its metabolite, trichloroacetic acid. *Toxicol Appl Pharmacol* 1989;99: 395–414.
93. O'Flaherty EJ, Scott W, Schreiner C, Beliles RP. A physiologically based kinetic model of rat and mouse gestation: disposition of a weak acid. *Toxicol Appl Pharmacol* 1992;112:245–256.
94. Kalin JR, Starling ME, Hill DL. Disposition of all-*trans*-retinoic acid in mice following oral doses. *Drug Metab Dispos* 1981;9:196–201.
95. Asai M, Faber W, Neth L, di Sant'Agnese PA, Nakanishi M, Miller RK. Human placental transport and metabolism of all-*trans*-retinoic acid *in vitro*. *Placenta* 1991;12: 367–368.
96. Zile MH, Inhorn RC, DeLuca HF. Metabolites of all-*trans*-retinoic acid in bile: identification of all-*trans*- and 13-*cis*-retinoyl glucuronides. *J Biol Chem* 1982;257:3537–3543.
97. Swanson BN, Frolik CA, Zaharevitz DW, Roller PP, Sporn MB. Dose-dependent kinetics of all-*trans*-retinoic acid in rats. Plasma levels and excretion into bile, urine, and faeces. *Biochem Pharmacol* 1981;30:107–113.
98. Adamson PC, Balis FM, Smith MA, Murphy RF, Godwin KA, Poplack DG. Dose-dependent pharmacokinetics of all-*trans*-retinoic acid. *JNCI* 1992;84:1332–1335.
99. Kraft JC, Slikker W Jr, Bailey JR, Roberts LG, Fischer B, Wittfoht W, Nau H. Plasma pharmacokinetic and metabolism of 13-*cis*- and all-*trans*-retinoic acid in the cynomolgus monkey and the identification of 13-*cis*- and all-*trans*-retinoyl–glucuronides. *Drug Metab Dispos* 1991;19:317–324.
100. Muindi JRF, Frankel SR, Huselton C, DeGrazia F, Garland WA, Young CW, Warrell RP Jr. Clinical pharmacology of oral all-*trans* retinoic acid in patients with acute promyelocytic leukemia. *Cancer Res* 1992;52:2138–2142.
101. Dedrick RL. Pharmacokinetic and pharmacodynamic considerations for chronic hemodialysis. *Kidney Int* 1975;7(Suppl 2):S7–S15.
102. Rowland M, Balant L, Peck C. Physiologically based pharmacokinetics in drug development and regulatory science: a workshop report (Georgetown University, Washington, DC, 29–30, May 2002). *AAPS PharmSci* 2004;6(1):E6.
103. Peck CC, Barr WH, Benet LZ, Collins J, Desjardins RE, Furst DE, et al. Opportunities for integration of pharmacokinetics, pharmacodynamics, and toxicokinetics in rational drug development. *J Pharm Sci* 1992;81(6):605–610.
104. Charnick SB, Kawai R, Nedelman JR, Lemaire M, Niederberger W, Sato H. Perspectives in pharmacokinetics. Physiologically based pharmacokinetic modeling as a tool for drug development. *J Pharmacokinet Biopharm* 1995;23:217–229.
105. Jorgensen WL. The many roles of computation in drug discovery. *Science* 2004;303(5665): 1813–1818.
106. Ricke DO, Wang S, Cai R, Cohen D. Genomic approaches to drug discovery. *Curr Opin Chem Biol* 2006;10(4):303–308.

107. Lahoz A, Gombau L, Donato MT, Castell JV, Gomez-Lechon MJ. *In vitro* ADME medium/high-throughput screening in drug preclinical development. *Mini Rev Med Chem* 2006;6(9):1053–1062.
108. van de Waterbeemd H, Gifford E. ADMET *in silico* modelling: towards prediction paradise? *Nat Rev Drug Discov* 2003;2(3):192–204.
109. Blesch KS, Gieschke R, Tsukamoto Y, Reigner BG, Burger HU, Steimer JL. Clinical pharmacokinetic/pharmacodynamic and physiologically based pharmacokinetic modeling in new drug development: the capecitabine experience. *Invest New Drugs* 2003; 21(2):195–223.
110. Lin JH, Lu AYH. Role of pharmacokinetics and metabolism in drug discovery and development. *Pharmacol Rev* 1997;49:403–449.
111. Lin J, Sahakian DC, de Morais SMF, Xu JJ, Polzer RJ, Winter SM. The role of absorption, distribution, metabolism, excretion and toxicity in drug discovery. *Curr Top Med Chem* 2003;3:1125–1154.
112. Poulin P, Schoenlein K, Theil FP. Prediction of adipose tissue:plasma partition coefficients for structurally unrelated drugs. *J Pharm Sci* 2001;90:436–447.
113. Theil FP, Guentert TW, Haddad S, Poulin P. Utility of physiologically based pharmacokinetic models to drug development and rational drug discovery candidate selection. *Toxicol Lett* 2003;138:29–49.

# MATHEMATICAL MODELING AS A NEW APPROACH FOR IMPROVING THE EFFICACY/TOXICITY PROFILE OF DRUGS: THE THROMBOCYTOPENIA CASE STUDY

Zvia Agur,[1,2] Moran Elishmereni,[1] Yuri Kogan,[1] Yuri Kheifetz,[1] Irit Ziv,[2] Meir Shoham,[2] and Vladimir Vainstein[1,2]

[1]*Institute for Medical Biomathematics (IMBM), Bene-Ataroth, Israel*
[2]*Optimata Ltd., Ramat-Gan, Israel*

**Contents**

1 Introduction
  1.1 Biomathematics and Its Use for Rationalizing Drug Treatment Strategies: A Short History
  1.2 Aims of This Chapter
2 Thrombopoiesis
  2.1 Biology
  2.2 Mathematical Modeling
3 Thrombocytopenia
  3.1 Pathophysiology
  3.2 Current Therapy
  3.3 Novel Cytokine Therapy

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

4 Modeling Drug Effects on Thrombopoiesis
4.1 Toxicity Modeling: Interleukin-11 (IL-11)
4.2 Efficacy Modeling: Thrombopoietin (TPO)
4.3 Modeling Combination Therapy: Chemotherapy and TPO Support
5 Using the Validated Model for Predicting Improved Drug Efficacy/Toxicity Profile
5.1 Model Validation by Retrospective Human Study Results: IL-11 and TPO
5.2 Model Validation in Prospective Animal Trials: TPO Applied to the Virtual Mouse and Virtual Monkey
5.3 Predicting the Optimal TPO Efficacy/Toxicity Ratio in Monkeys
6 Using the Thrombopoiesis Model for Predicting an Unknown Animal Toxicity Mechanism
7 Transition to Phase I
8 Conclusion
References

## 1 INTRODUCTION

Despite recent innovations, many life-threatening diseases still lack effective treatments. In March 2004, the U.S. Food and Drug Administration (FDA) agency issued a major report that identifies both the problems and the potential solutions for bringing more breakthroughs in medical science to patients, as quickly and efficiently as possible. The report looks at the development processes for drugs, biologics, and medical devices and calls for a joint effort of industry, academic researchers, product developers, patient groups, and the FDA to identify key problems and to develop solutions: "A new product development toolkit—containing powerful new scientific and technical methods such as computer-based predictive models . . . is urgently needed to improve predictability and efficiency along the critical path from laboratory concept to commercial product" [1]. The role of the present chapter is to demonstrate how the science of biomathematics can be harnessed for improving the solution of safety issues in the critical path of drug development, in general, and in the transition from the preclinical phases to Phase I, in particular.

### 1.1 Biomathematics and Its Use for Rationalizing Drug Treatment Strategies: A Short History

A new synthesis of ecology and "hard" biology, called biomathematics, emerged in the second half of the 20th century in the scientific community. Its role was to rationalize complex biological processes. Thus, in contrast to experimental biologists, who work at the microscopic cellular level and develop analytical tools that are analogous to the still camera in photography, biomathematicians develop formulae, which effectively animate these shots, allowing us to understand the dynamics of the complex process we are investigating. Biomathematics has enabled the development of a range of new theories dealing with significant problems of disease progression and control, hitherto beyond the reach of "snapshot", experimental biology.

But in spite of these efforts, until recently the biomedical research community was unanimous in viewing biological systems as too complex to enable accurate retrieval by mathematical models. This declarable mistrust in the power of biomathematics left the biomedical sciences lagging behind other sciences, as an immature sequel of experimental observations. In recent years, the somewhat disillusioned postgenomic era has rediscovered biological complexity, renaming it "system biology." It seems that, finally, biomathematicians are becoming legitimate and almost beloved children of the scientific community.

In the 1980s, notions from population dynamics have been used by Agur and colleagues to develop relatively simple formulae describing the growth patterns of interwoven populations of healthy and cancerous cells, formulae that have resulted in new drug regimens where the toxicity of chemotherapy has been significantly reduced [2, 3, 63]. The power of mathematics here was to prove the universality of the theoretical results and therefore to justify and encourage collaboration with cancer research experimentalists. The latter not only verified the theory in the laboratory, but also pinpointed the feasibility and strength of the theory-to-lab arrow [4, 5].

In the above mentioned mathematical models, populations are subjected to a loss process due to randomly occurring environmental disturbances, which are effective only during a portion of the life cycle. The models are studied over a large range of time scales of the environmental change and for different degrees of variance in the system parameters. Analysis shows that the expected survival time of the population has a strong nonmonotonic dependence on the relation between the duration of the disturbance-resistant life stages and the periodicity of the environmental disturbance. This effect, termed *resonance phenomenon*, was found at integer and fractional multiple of this relation. Interestingly, persistence in all harshly varying environments is shown to depend on the degree of resonance in the environmental and population processes [2, 6–8, 64].

The universality of the resonance phenomenon, and its implication that the frequency of environmental disturbances determines population growth, seemed to offer attractive possibilities for the control of cancer. Thus, Agur and colleagues have shown, both theoretically and experimentally, how the resonance phenomenon could be employed for choosing cytotoxic drug regimens that allow discrimination between different target cell populations. More specifically, it was asserted that one could determine the period of drug pulses, so as to maximize the elimination of malignant cells and, at the same time, to minimize the destruction of drug-susceptible host cells.

In the first stage, it was necessary to investigate the generality of the resonance phenomenon for the dynamics of different host cell populations. To allow analytical tractability, and hence parameter-independent conclusions, deliberately crude mathematical models of the cancer patient were constructed, making simplified assumptions that both tumor and target host tissues are cells that vary only in cell-cycle parameters. The spatial arrangement of these cells was hence ignored. The interest was focused on comparing the overall behavior of the tumor to that of the host target tissues under different chemotherapy regimens of cell-cycle phase-specific drugs, and on finding regimens that are selective to cancer cells. The conviction was that only a simplified model can clarify the underlying properties of the biomedical scenario. Model analysis showed that the resonance phenomenon is valid when

considering drug-induced myelotoxicity and tumor progression. Specifically, the mathematical analysis suggested that the efficacy/toxicity ratio of cell-cycle phase-specific drugs would be maximized at an integer or fractional multiple of the mean cycle time of the host target populations (this being denoted the *Z-method*). This applies either when mean cancer cell-cycle parameters differ from those of the target host cell population, or when those are similar but the variation in the distribution of their cell cycle is larger [3, 9, 10].

It seemed mandatory, already at this level of qualitative understanding of the system, to carefully examine the plausibility of applying the Z-method to cancer and host cells. To this end, a series of laboratory experiments was launched, first examining the effect of drug-pulsing rhythm on cell proliferation, *in vitro*, and then checking how varying the pulsing rhythm of different drugs will affect disease progression on the level of the whole organism.

Applying an antimetabolite chemotherapeutic drug, cytarabine (ara-C), to lymphoma cells *in vitro*, it was experimentally shown that when total drug dose and total treatment duration were kept constant, drug schedules exerting significantly higher cell growth were those possessing periodicity that was an integer multiple of the average population cell-cycle time, as evaluated by rates of DNA synthesis. In equivalent *in vivo* experiments, mice were treated by short ara-C pulses, having different distributions of the interdosing intervals, including stochastically determined periodicity. Toxicity was evaluated by spleen weight, differential peripheral blood measurements, kinetic measurements of bone marrow cell proliferation (with bromodeoxiuridine labeling and flow cytometry analysis), and by overall animal survival. Results of these trials proved the superiority of the Z-method in controlling cell proliferation. In particular, it was shown that when the interdosing interval is an exact multiple of the intermitotic time of bone marrow stem and progenitor cells, cell-cycle kinetics are less affected, myelotoxicity is minimal, and survival time of tumor-bearing mice is maximal [3–5, 11].

The above experiments demonstrate that regimens employing a drug period, which coincides with the inherent periodicity of bone marrow cells, protect the bone marrow rather well, leading to increased life expectancy of cancer-bearing mice. In contrast, other regimens, notably those employing intervals that are randomly distributed around the same mean periodicity, are extremely toxic. In general, this work is a proof of the concept that by using a rational drug scheduling, based on the Z-method, it is feasible to control drug toxicity exerted on the bone marrow of cancer-bearing mice, while at the same time delaying their tumor progression. This conclusion has been further supported under more realistic and complex tumor growth models [12, 13] and under physiologically based mathematical models of human hematopoiesis [14].

The above described mathematical models added important new insights on drug treatment strategies, by considering fundamental effects on cellular dynamics. Yet, to enable mathematical tractability, drug-related characteristics were treated very simplistically. For this reason, only qualitative conclusions can be drawn from these model analyses. Now that the proof of concept has been provided, the models are becoming increasingly instrumental in planning concrete treatment strategies. To this end, modeling of the underlying pathological and physiological processes in a patient is wedded with the detailed relevant pharmacokinetics (PK) and pharmacodynamics (PD) [15, 65]. By uniting in one framework the pharmacological properties of the drug with the dynamic properties of the cells in the target tissues, one

obtains an improved understanding of how the timing of drug administration—the interdosing interval, drug fractionation, and other aspects of treatment scheduling—may affect the patient. Biomathematics has the power of integrating the vast biomedical and pharmaceutical knowledge into a concise formal language, which enables intensive calculations of the outcomes of complex processes. These calculations, being by far more comprehensive and accurate than any human intuition can be, justify their use for rational treatment design at any stage of the critical path, especially when partial preclinical results collected in several species are to be used for predicting the human response.

### 1.2 Aims of This Chapter

In this chapter, we discuss the application of mathematical modeling tools to elucidate the hematological disorder thrombocytopenia in different animal species, and to accurately describe the efficacy and safety of the thrombocytopenia-alleviating drugs, thrombopoietin and interleukin-11. Model validation in mice, monkeys, and humans, by the use of diverse experimental information, is discussed, as well as the use of animal models for predicting human safety and for improving patient treatment.

## 2 THROMBOPOIESIS

Any mathematical model describing a biological process is based on a verbal model of the process, that is, on an algorithm concisely describing its essential physiological "flow." We devote this section to a brief presentation of the verbal model of thrombopoiesis, and to the subsequent description of a mathematical model, constituting the essence of platelet development.

### 2.1 Biology

Platelets are components of the hematopoietic system, being the nucleus-lacking derivatives of megakaryocytes (MKs). Platelets function primarily in processes of thrombus formation and vascular repair, but also have roles in natural immunity and metastatic tumor biology. Platelet counts in healthy humans average 150–400 billion cells per liter of blood, a homeostasis that is crucial to maintain. Low counts may lead to severe hemorrhage, renal failure, and other disorders, while high counts indicate poor prognosis with respect to cardiovascular diseases and malignancies. Thrombopoiesis, the formation of blood platelets, has been studied for the last century, although critical advances in our understanding of the unique processes that accompany it are only of the last decade. We refer the reader to Fig. 1 for an illustration of the process and to the literature for a more elaborate description [16–19].

Normal platelet formation begins with bone marrow progenitors of the lineage, which descend from hematopoietic stem cells and proliferate in response to certain growth factors. The next stages in the sequence involve the immature forms of MKs (promegakaryoblasts and megakaryoblasts), which then develop into larger, mature MKs. As the cells mature into MKs, they gradually lose their ability to divide but

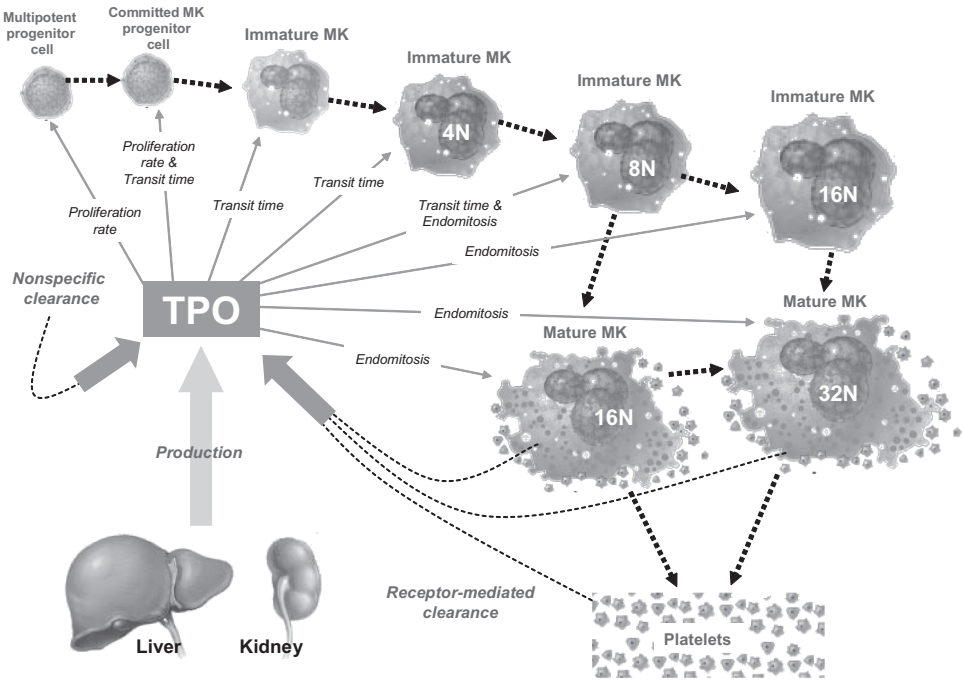


**FIGURE 1** The process of thrombopoiesis. Uncommited progenitors in the bone marrow (hematopoietic stem cells) proliferate and develop into commited progenitors of the megakaryocyte (MK) lineage, which further develop into immature forms of MKs (promegakaryoblasts and megakaryoblasts). These then develop into larger, mature MKs. Maturation of cells is coupled with loss of the mitotic ability, but endomitosis and polyploidity are maintained until the early stages of MK fragmentation and systemic distribution of platelets. Thrombopoietin (TPO), derived from the liver and kidney constitutively, is a primary mediator of the platelet formation process and is sequestered by platelets and MKs, as well as nonspecifically.

maintain the function of DNA synthesis, thereby undergoing endomitosis and attaining a polyploid state. Finally, MK fragmentation occurs, resulting in release of platelets, first to bone marrow sinusoids and then to the systemic circulation. Radiolabeling studies indicate that the amount of platelets normally consumed by hemostasis is very small and constant. The highly regulated process of platelet consumption is induced by their adhesion to von Willebrand factor (vWF) and extracellular matrix components of damaged blood vessel wall underlying the endothelium. Following platelet adhesion, a clot is formed and the vessel wall is recovered. With a life span of approximately 10 days, unconsumed platelets that senesce are removed by macrophages of the reticuloendothelial system, mostly in the spleen.

Thrombopoietin (TPO), a glycoprotein produced by the liver and kidney constitutively and by the bone marrow upon demand, is characterized as the primary regulator of platelet formation. Although the focus was set on TPO in the late 1960s, it was successfully cloned only in 1992. TPO acts in all stages of thrombopoiesis, its receptors being expressed on each cell type in the lineage. Interestingly, systemic and bone marrow levels of TPO are inversely correlated to platelet numbers: due

to high affinity platelet-bearing TPO receptors, increased amounts of platelets result in high internalization of TPO, forming successful autoregulation of the thrombopoiesis. Slightly less significant cytokines, namely, the interleukins 3, 6, 11, and stem cell factor (SCF), stromal cell-derived factor 1 (SDF-1), and granulocyte macrophage colony-stimulating factor (GM-CSF), were also identified as modulators in normal platelet formation, each acting in different developmental phases, and some acting in synergy with others. Studies indicate that surrounding tissue, comprised of fibroblasts, endothelial cells, and macrophages, is also capable of augmenting thrombopoiesis, although these environmental factors are not thought to be critical since platelets are easily produced also *in vitro*.

The manipulation of molecular and genetic techniques in the last two decades has greatly boosted the understanding of biological pathways and mechanisms involved in platelet formation. Nonetheless, at the current stage, the regulation of thrombopoiesis, the central molecular components taking part in it, and their significance for therapeutic intervention in platelet-associated illnesses are yet to be clarified.

## 2.2 Mathematical Modeling

Although the intricacies of thrombopoiesis are far from being fully elucidated, current knowledge enabled the verbalization and subsequently the mathematical formalization of a basic thrombopoiesis-describing model. Schematized in Fig. 2, and elaborated elsewhere, the model divides the entire thrombopoietic lineage into 17 compartments. Every compartment, except for the first, is represented by time-dependent age distribution, where the number of cells is given as a function of discrete age (time spent in the compartment) and time. The number of cells in the first compartment is a function of time only. There are six groups of compartments in the model, in accordance with the above-mentioned structure of the thrombopoietic lineage. The subdivision of the six groups is as follows: (1) one compartment of lineage-uncommitted progenitors (UCP), (2) one compartment of lineage-committed progenitors (CP), (3) six compartments of endomitotic progenitors (EP), (4) four compartments of developing progenitors (DP), (5) four compartments of MKs, and (6) one compartment of blood-circulating platelets.

The first compartment, UCP, refers to all bone marrow hematopoietic progenitor cells that can differentiate into more than one line (e.g., pluripotent stem cells, CFU-GEMM). We assume one homogeneous UCP population, as its division into various subpopulations is not feasible, since kinetic data regarding the various subpopulations of the UCP compartment (i.e., rates of proliferation, maturation, and self-renewal) is rather scarce. In our model, cells of this compartment proliferate at a certain rate ($\alpha_{UCP}$) and differentiate into MKs or other precursors. Based on previous studies showing that the probabilities of stem cell differentiation into any given hematopoietic lineage are constant, it was assumed that a fixed proportion of mature uncommitted progenitor cells flows into the thrombopoietic lineage ($\varphi_{UCP}$). Apoptotic cell death may have a significant effect on cell numbers in the proliferating compartments. Thus, we included its effect together with the effect of cell proliferation in the total amplification of cell number in a specific compartment. Since no information is currently available concerning apoptosis in the nonproliferating MK compartments, this issue was disregarded in these compartments.

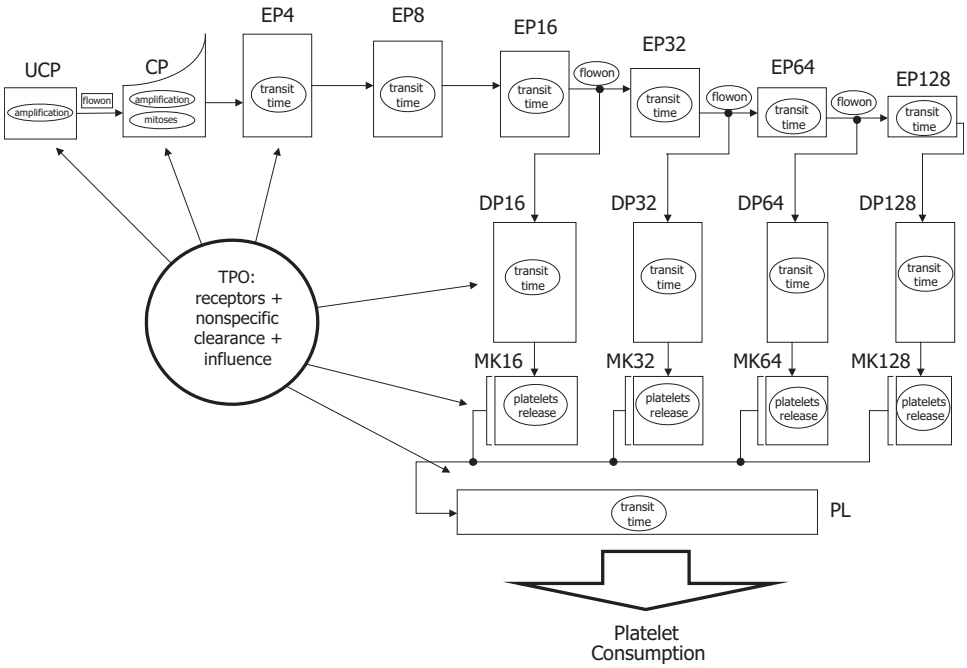


**FIGURE 2** A scheme of the thrombopoiesis mathematical model. The model consists of 17 compartments, which can be grouped into six clusters: (1) one compartment of lineage-uncommitted stem cell progenitors (UCP), (2) one compartment of lineage-committed progenitors (CP), (3) six compartments of endomitotic progenitors (EP) of ploidies 4N, 8N, 16N, 32N, 64N, and 128N, (4) four compartments of developing progenitors (DP) of ploidies 16N, 32N, 64N, and 128N, (5) four compartments of megakaryocytes (MKs) of ploidies 16N, 32N, 64N, and 128N, and (6) one compartment of individual, blood-circulating platelets. Relevant key parameters included in the kinetic calculations of the different compartments are amplification rate, flow-on fraction (flowon), transit time, and rate of platelet release. The arrows indicate the direction of intercompartmental cell flow.

The CP compartment stands for an age distribution of progenitor cells already committed to the thrombopoietic lineage, but still capable of proliferation. The compartment is characterized by a transit time ($\tau_{CP}$) and a number of mitoses ($N_m$). The latter represents the average number of possible cell divisions at that stage. This number may be noninteger, considering that certain cells can be quiescent and undergo fewer mitoses than others. Cells leaving CP enter the first of the six EP compartments, which comprises cells that have lost proliferation ability and yet are not sufficiently mature to release platelets. Thus, this subgroup of compartments, namely, EP4, EP8, EP16, EP32, EP64, and EP128, are formed according to the biologically known sequence of cells of the MK lineage bearing ploidy of 4N, 8N, 16N, 32N, 64N, and 128N, respectively. These endomitotic precursors cannot divide but continue endomitosis. The cells that enter the EP4 compartment are of ploidy 2N, at the exit they have ploidy 4N, and so on for subsequent EP compartments. It is assumed that MKs can release platelets only after reaching the 16N-ploidy phase. Hence, to arrive at platelet production in our model, the EP16 compartment must

be entered. After cells spend their designated transit time ($\tau_{EP}$) in an EP compartment, they can move to the next EP compartment. Fractions of cells that continue endomitosis after moving through EP16, EP32, or EP64 are designated $\varphi_{EP16}$, $\varphi_{EP32}$, and $\varphi_{EP64}$, respectively, where in normal conditions $\varphi_{EP64} = 0$.

Alternatively, cells at the EP16, EP32, or EP64 stage may stop the endomitotic process by entering the relevant DP compartment, which is represented by four subgroups of cells. DP cells of 16N, 32N, 64N, and 128N ploidy increase their cytoplasmic contents prior to platelet production. Cells that have stopped endomitosis after exiting EP16 compartment enter into DP16; cells that have stopped endomitosis after exiting EP32 compartment enter into DP32, and so forth. After spending a certain amount of time ($\tau_{DP}$) in the DP compartment, cells pass to the corresponding MK compartment. The four subcompartments of this stage signify cells capable of releasing platelets. MK16 compartment receives cells exiting from DP16, MK32 receives cell exiting from DP32, and so on. As in previous compartments, here too the cells are distributed over age (time spent in the compartment), yet there is no predefined transit time. Instead, the new cells entering MK are assumed to have some initial releasing capacity, indicating the number of potential platelets, or total platelets volume, to be released until the cell is exhausted. This capacity is assumed to be larger for cells of larger ploidy. Cells release platelets continuously, and their releasing capacity is consumed until it drops to zero, thereupon disappearing from the compartment.

The last compartment, PL, represents platelet counts distributed by age (time spent in blood after the release). Distribution of mean platelet volume (MPV) is also computed in order to allow calculations of platelet degradation. The platelets are readily consumed by the body in an age-independent manner, and in terms of volume (platelet number multiplied by MPV). There is also a maximal life span for old platelets, so that all nonconsumed platelets disappear upon reaching a maximal age (denoted by $\tau_{PL}$). New platelets appear at every time step, as released by MK cells. This release is also described in terms of volume, and at every time step the amount and MPV are evaluated for the new platelets, taking into account the relative contributions of MK cells of different ploidies.

The formal description of the cellular compartments is supplemented by the model for the endogenous TPO, which regulates the dynamics of the process. A one-compartment PK model for TPO, with a constant production and a linear nonspecific clearance, is applied. In addition, there is an alternative, specific clearance of TPO, achieved through the consumption by the cells (receptor binding and internalization). To include this in the model, distributions of receptors per cell in all the line-committed compartments (CP and the groups that follow) are incorporated. Formulae are constructed for receptor numbers distributed over cell age. Receptors, constantly produced by the cells, excluding platelets that have no nucleus, bind free TPO and are internalized. To reflect the nonspecific clearance of TPO, Michaelis–Menten behavior is assumed for TPO-receptor kinetics.

The major stabilizing feedback in the model is exerted via the effects of TPO on the cells throughout the lineage. Kinetic parameters of the model's compartments are functions of TPO concentrations. In the UCP compartment, the amplification rate is an increasing function of the TPO concentration. Similarly, the number of mitoses in the CP compartment is an increasing function of TPO. In contrast, the transit times in CP, EP, and DP, and the EP flow-on (fraction of cells that continue

endomitosis) are decreasing functions of the TPO concentration. Finally, the MPV of released platelets (and consequenty the loss of cell volume) is an increasing function of the TPO concentration in the MK compartments. For all the dependencies listed above, sigmoid functions with saturation for large values of TPO concentration are used. In addition to this feedback, a TPO-independent feedback on cell proliferation in the model, or a "crowding effect," is included as well. The amplification rate in UCP and the number of mitoses in CP are decreasing functions of cell numbers in these compartments, respectively. This represents internal regulation of cell proliferation in these stages. Such functions are also assumed to have a sigmoid form.

The model is mathematically formalized according to the above description, using discrete difference equations. The variables of the model (e.g., cell numbers) are represented as functions of discrete time points, with a certain step $\Delta t$. The set of the model equations describes the recursive calculation of the system state for the current time point, using the state of the system at previous time points. For the specific clearance model, steady-state approximations are used, as these processes are on a much faster time scale than all the rest of the model. The model is simulated using a C++ code, and its parameters are estimated either directly from experimental measurements or by fitting the model to the published experimental data. The validation of the model is further described in Section 5.

## 3 THROMBOCYTOPENIA

Setting the essential rules for depicting thrombopoiesis, as explained in Section 2, it is possible to evaluate not only the homeostatic behavior of the system but also its irregularities. Diversion from normal steady-state thrombopoiesis can arise from various reasons and can have several physiological consequences. In this section, we discuss some of the main mechanisms leading to low platelet count, termed thrombocytopenia (TP). Both current and future therapeutic avenues will be illustrated.

### 3.1 Pathophysiology

The concise summary of TP pathophysiology, presented in this section, is based on extensive reviews of this pathology in the current hematology literature [16, 20, 21]. The clinical significance of TP, characterized by a blood platelet count of less than 150,000 cells per microliter, stems from the major role of platelets in the process of normal hemostasis. Low platelet counts lead to diminished ability of hemostasis and, consequently, increased bleeding propensity. Yet there are important exceptions to this rule. First, clinically significant bleeding occurs regularly only when the amount of platelets decreases below 10,000–20,000/μL. Second, exact clinical manifestation of this disorder depends on the underlying pathophysiological mechanisms. While TP induced by low platelet production leads to increased risk of bleeding, TP caused by excessive platelet consumption may result in an elevated risk of thromboembolism.

TP can be caused by one or more of the following mechanisms (elaborated in the following section): (1) decreased platelet production, (2) increased sequestration of platelets, or (3) their accelerated destruction. TP induced by the first mechanism,

impaired production of platelets, accompanies a variety of bone marrow pathologies. Congenital hereditary decreased production TP, one of these cases, is very rare. It can be caused by malfunctioning TPO receptor signaling (e.g., amegakaryocytic TP) or by impaired MK fragmentation into mature platelets. The more common syndrome is acquired decreased production TP. This form of TP can be a result of pathological bone marrow infiltration in hematopoietic malignancies (leukemias and lymphomas) or in massive metastases of extramedullary solid tumors. Alternatively, this could occur in myelodysplastic syndromes and vitamin $B_{12}$ deficiency, when intrinsic maturation defects arise in the early hematopoietic progenitors, with involvement of 1–3 hematopoietic lines. Extrinsic inhibitors of hematopoietic progenitor proliferation, such as (1) cytotoxic drugs (anticancer chemotherapy), (2) circulating autoantibodies to MKs (immune thrombocytopenic purpura, systemic lupus erythematosus, etc.), (3) proinflammatory cytokines (sepsis), or (4) various viral infections (EBV, CMV, HIV, etc.), can induce this TP form as well.

Increased sequestration TP, the second mechanism, is a common finding in splenomegaly of any cause. A frequent consequence of splenic enlargement is increased pressure in the portal vein, which causes draining of blood from the spleen into the liver. This syndrome, referred to as portal hypertension, is present in a variety of chronic liver diseases and in cases of portal vein thrombosis due to hypercoagulation states. In portal hypertension, increased sequestration is often the only cause of TP. Other diseases with splenomegaly generally produce TP of a combined etiology. For example, splenomegaly is associated with conditions of leukemia and lymphoma, in which bone marrow infiltration results in decreased platelet production. Similar combinations of TP mechanisms appear in several infectious disorders (e.g., brucellosis and visceral leishmaniasis) and inflammatory diseases (e.g., Felty's syndrome and systemic juvenile rheumatoid arthritis).

Pathologically accelerated platelet destruction (the third mechanism of TP) can be induced by either increased consumption or physical factors. Physical platelet destruction occurs upon increased blood shear stress near valvular and intravascular prosthetic devices. There are varying disorders of increased platelet consumption by pathological uncontrolled coagulation. Thrombotic thrombocytopenic purpura (TTP) and hemolytic uremic syndrome (HUS) are associated with thrombosis, rather than bleeding, despite low platelet count. Recent progress in TTP research reveals that the underlying pathology involves decreased or absent function of a protease known as a disintegrin and metalloproteinase with a thrombospondin type 1 motif, member 13 (ADAMTS13), responsible for specific cleavage of vWF. As a consequence, large multimers of vWF are formed, and they initiate bouts of platelet adhesion and activation. Conversely, pathogenesis of HUS is incompletely understood. Since this disorder appears typically after certain bacterial infection, it is assumed that cross-reactive antibodies are responsible for platelet activation in this syndrome. Disseminated intravascular coagulation (DIC), distinct from the latter two disorders by its simultaneous induction of both thrombosis and bleeding, is a common feature of several severe illnesses including overwhelming sepsis, poisoning due to snake bites, amniotic fluid embolism, and fat embolism. The pathogenesis of DIC includes uncontrolled activation of intravascular coagulation pathways, leading to increased platelet consumption.

Accelerated destruction TP can also be attributed to immunological disorders. The most recognized disorder of this type is immune (idiopathic) thrombocytopenic

purpura (ITP). The underlying pathology consists of formation of autoantibodies against specific platelet surface antigens. Platelets covered with autoantibodies are effectively and prematurely destroyed by the phagocytes of the reticuloendothelial system. In the minority of cases, presence of autoantibodies to MKs is also suspected, which results in TP of mixed etiology—both decreased production and accelerated destruction. Other diseases with autoimmune platelet destruction include systemic lupus erythematosus and antiphospholipid antibody syndrome. Drug-induced immune TP was observed upon the use of many exogenous factors, including heparin, gold, antibiotics, and anti-inflammatory drugs. In these cases, antibodies are formed against the complex of a drug with platelet surface antigens. In heparin-induced TP, thrombotic complications are common (similar to those in TTP), since antibodies lead to excessive platelet activation in addition to increased clearance.

To examine the risk of TP in new drugs, or in yet to be tried drug regimens, the PK and PD of the drug should be simulated superimposed on a mathematical model of the underlying TP mechanism. However, how to determine the specific TP mechanism in each particular case is a very challenging task. While certain clues can be received from the patient's medical history, physical examination, and peripheral blood smear, bone marrow biopsy is frequently required in order to evaluate the number of MKs. Low numbers are indicative of decreased production, while increased numbers are associated with accelerated destruction. Some cases remain undefined, even after thorough laboratory investigation and manipulation of extensive tools. The type of TP guides further clinical workup for specific underlying pathology as well as possible treatment options.

### 3.2 Current Therapy

In light of the different mechanisms for TP, treatment is diverse and highly dependent on the underlying cause. Generally, we can divide current therapy into treatments directed against pathophysiological mechanisms, and symptomatic treatment of TP-induced bleeding/thrombosis. In the majority of the cases, platelet counts do not drop below 10,000–20,000 per microliter of blood, so that serious bleeding complications do not develop. Treatment of the underlying causes includes withdrawal of the offending stimulus (e.g., discontinuation of an immunogenic drug), removal of vWF complexes by plasma exchange in TTP, replacement of mechanical prosthesis by a biological one, and specific treatment of malignancies and infections.

It is imperative to identify accelerated platelet destruction by the reticuloendothelial system, for which splenectomy and/or administration of intravenous immunoglobulins (IVIGs) is indicated in refractory cases. Splenectomies lead to decreased autoantibody production and remove a large fraction of phagocytes responsible for platelet destruction. IVIGs, on the other hand, saturate antibody-binding receptors on the phagocytes and suppress their interaction with antibody-covered platelets and MKs. Therefore, theoretically, IVIGs should be more effective in cases of MK-specific antibodies. This hypothesis, however, has not been examined clinically, due to technical difficulties in identifying antibody specificity in particular patients. Importantly, neither a splenectomy nor IVIG therapy is effective in cases of direct complement-mediated antibody toxicity to platelets and MKs.

Symptomatic treatment of TP-associated bleeding disorders consists mainly of transfusion of donor platelets. Platelet transfusions are commonly used, with approximately 9 million units yearly applied in the United States alone. Although repeated platelet transfusions may aid in preventing bleeding, they can transmit both viral and bacterial infections and can cause alloimmunization, which requires HLA-matched donors. Health care costs and inconvenience to patients also pose a concern. Furthermore, this treatment must be repeated on a regular and frequent basis, due to the short half-life of donor platelets (3–4 days), and since the underlying causes of TP are left untreated. Platelet transfusions are of especially low efficacy in immunologically mediated TP that involve an even shorter platelet half-life.

### 3.3 Novel Cytokine Therapy

Over the past fifteen years, attempts to find new therapies to replace platelet transfusions in the treatment of thrombocytopenic disorders have naturally been focused on the growth factors involved in normal thrombopoiesis (see Section 2). Interleukin-11 (IL-11), discovered in 1990 to be a thrombopoietic growth factor secreted from bone marrow stromal cells, was the first candidate for this purpose. The genetically engineered form of this protein, recombinant human IL-11 (rhIL-11), was evaluated for treatment in patients with nonmyeloid malignancies that suffered from chemotherapy-induced thrombocytopenia and did, in fact, produce a dose-dependent increase in platelets. RhIL-11, approved in 1997 for prevention of severe drug-induced thrombocytopenia, is currently the only cytokine licensed in the United States for this purpose. However, with its only modest thrombopoietic activity and its use often associated with intolerable side effects, IL-11 has not satisfied the demands for an efficacious thrombopoietic agent to be applied in the clinical setting [22–24].

In light of the central role of TPO in the proliferation and survival of MKs and increased platelet production, its potential to replace standard platelet transfusion therapy and improve platelet harvesting efficacy in donors seems even greater than that of IL-11 [22, 25]. Recombinant human TPO (rhTPO) has been examined for exogenous application in patients with primary thrombocytopenic disorders, as well as for treatment of chemotherapeutically induced myelosuppression in cancer patients. However, Phase I studies with rhTPO in healthy individuals remain disappointing. In a trial that was aimed at improving platelet yields in donors, some of them developed immune-mediated TP due to appearance of autoantibodies to native TPO. Evaluation of rhTPO therapy has also encountered difficulties due to delayed peak platelet response and neutralizing antibodies formed against the pegylated molecule. It is believed that most complications are caused by significant antigenic differences between the native and recombinant TPO proteins. As a result, this molecule currently remains in its developmental stage and is not yet approved. Small TPO analogues and synthetic TPO-derived agents are currently being developed to bypass immunogenicity-related problems [26–34].

Mathematical modeling of thrombopoiesis (as described in Section 2) and drug-induced TP (as will be shown in Section 6) can aid in disentangling disease dynamics and in deciphering pathophysiological mechanisms underlying the disease. In particular, mathematical models of TPO-affected thrombopoiesis and of IL-11-induced side effects enable one to propose improved treatment schedules for both drugs,

obtaining maximum efficacy and substantially improved safety (see Sections 4 and 5). In turn, preclinical validation of these improved regimens can give new hope for clinical application of these cytokines in their multiple arenas.

## 4 MODELING DRUG EFFECTS ON THROMBOPOIESIS

Innovative cytokine-based therapeutics of IL-11 and TPO (see Section 3) developed for application in thrombopoiesis-related syndromes and for efficiently harvesting donor platelets have been extensively studied in the last decade or so. Yet, even today, toxicity-related limitations of each of these molecules prevent their clinical applicability. The following section is dedicated to mathematical models focusing on both beneficial and detrimental consequences of these therapeutic avenues, and on the manner by which such models may be used to suggest improved treatment strategies.

### 4.1 Toxicity Modeling: Interleukin-11 (IL-11)

Although IL-11 has been approved for treating drug-induced TP for almost a decade, its induction of intolerable toxic effects has impeded its therapeutic use. The cytokine has been shown to elicit a wide range of side effects, mainly associated with blood volume expansion (BVE). These include complications of various grades of edema, anemia, pleural effusion, and cardiac arrhythmia [23].

An initial and generalized perspective on this issue required a mathematical model of IL-11-induced BVE. However, the definitive mechanism underlying IL-11-mediated water accumulation is yet to be discovered, and the incomplete data [23, 35] restrict the possibility to produce a concrete and reliable model for this problem. Instead, a novel biomathematical approach was developed so that not one but several mathematical models, each representing an alternative mechanism for BVE following IL-11 therapy, were devised and evaluated [36]. The underlying assumption here was that by concurrently simulating multiple models, which formally describe alternative mechanisms, but all producing the same fluid-retentive behavior, it is possible to faithfully depict this effect despite the obscurity of the actual mechanism and related data. Accordingly, convergence of all models to yield unified BVE dynamics was the characteristic by which their success and predictive ability was measured. Furthermore, since the range of different models that capture the behavior of stable biological systems (e.g., body water homeostasis) is assumed to be relatively narrow, the chances of arriving at a reliable model was considered even greater in the drug-afflicted BVE setting [36].

Three closely related biologically based models were constructed using ordinary differential equations (ODEs) and standard Michaelis–Menten kinetics. The models incorporated basic water regulation properties (blood volume, pressure, and vascular compartments) and endocrinal feedback effects (compartments for BVE-upregulating and BVE-downregulating hormones), together with the IL-11-induced perturbation of the system. Notably, a short-term neuronal influence or long-term permanent structural changes were not considered, restricting the modeling to the therapeutic time scale. The variation between the different models was expressed in (1) the component directly affected by IL-11, being either blood volume or blood

vessels, and (2) the components affected by the endocrinal hormones, being both vessels and volume (full effect) or only volume (partial effect). Thus, model 1 assumes IL-11-induced vessel enlargement and full endocrinal feedback; model 2 assumes IL-11-induced volume expansion and full endocrinal function; and model 3 assumes IL-11-induced volume expansion and partial endocrinal function. The models were identical in all other assumptions.

Dynamics of subcutaneously administered IL-11 were represented by a one-compartment PK model, with first-order absorption from the external tissue. IL-11 concentrations in the tissue ($I_T$) and plasma ($I_P$) are given by

$$\dot{I}_T = -c_1 I_T \tag{1}$$

and

$$\dot{I}_P = f c_1 I_T - c_2 I_P \tag{2}$$

where $I_T$ ($t = 0$) is the administered IL-11 concentration and $I_P$ ($t = 0$) = 0. The coefficient $c_1$ denotes the rate of drug transport from the primary tissue to the plasma, $c_2$ is the drug clearance rate, and the fraction of transferred drug, an indicator of bioavailability, is $f$.

Blood pressure was described by considering its classical correlation to heart rate, stroke volume, and systemic vascular resistance. As heart rate variability was disregarded in experiments, and an inverse relationship between vascular resistance and vascular capacity was assumed, blood pressure was described to be a function of blood volume and the vascular capacity (expressed in terms of vessel surface):

$$P(t) = \left( \frac{V(t)}{S(t)^{3/2}} \right)^{n_0} \tag{3}$$

The function $P(t)$ is blood pressure at time $t$, $S(t)$ denotes overall vessel surface at time $t$, $V(t)$ is blood volume at time $t$, and $n_0$ is a parameter. Steady-state values of pressure and overall vessel surface and volume were given by $P_{st}$, $S_{st}$, and $V_{st}$, respectively.

Dynamics of the BVE-upregulating hormones ($H_U$) and the BVE-downregulating hormones ($H_D$) were subjected to blood pressure variations and were therefore given by

$$\dot{H}_U = \alpha_1 g(P_{st} - P(t)) - d_1 H_U(t) \tag{4}$$

and

$$\dot{H}_D = \alpha_2 g(P(t) - P_{st}) - d_2 H_D(t) \tag{5}$$

respectively. The description is general and does not refer to hormonal concentrations per se, but rather to effects associated with endocrinal deviations. In Eqs. 4 and 5, formation rates ($\alpha_1, \alpha_2$) and degradation rates ($d_1, d_2$) of the endocrinal effects

were taken as constants, and blood pressure deviations from steady state allowed formation of each factor using the function

$$g(x) = \begin{cases} x & \text{if} \quad x > 0 \\ 0 & \text{if} \quad x \leq 0 \end{cases} \tag{6}$$

The vascular structure was described in two-dimensional units, for example, by relating to the average surface of the inner endothelium layer of blood vessels. IL-11-induced vessel surface enlargement (in model 1) and $H_U$-induced vasoconstriction or $H_D$-induced vasodilation (in models 1 and 2) were described by biologically compatible nonlinear functions. As described earlier, model 3 does not enable vascular resizing in response to IL-11 or hormones. All three models, however, incorporate simple intrinsic feedback effects to vessel steady state in blood vessel dynamics. Thus, the equations for models 1, 2, and 3 are given by Eqs. 7, 8, and 9, respectively:

$$\dot{S} = \frac{\beta_0 (I_P(t))^{r_0}}{k_0 + (I_P(t))^{r_0}} - \frac{\beta_1 (H_U(t))^{r_1}}{k_1 + (H_U(t))^{r_1}} + \frac{\beta_2 (H_D(t))^{r_2}}{k_2 + (H_U(t))^{r_2}} + \beta_3 (S_{st} - S(t)) \tag{7}$$

$$\dot{S} = -\frac{\beta_1 (H_U(t))^{r_1}}{k_1 + (H_U(t))^{r_1}} + \frac{\beta_2 (H_D(t))^{r_2}}{k_2 + (H_U(t))^{r_2}} + \beta_3 (S_{st} - S(t)) \tag{8}$$

and

$$\dot{S} = \beta_3 (S_{st} - S(t)) \tag{9}$$

Vascular modification rates in response to IL-11, $H_U$, and $H_D$ are denoted as $\beta_0$, $\beta_1$ and $\beta_2$, respectively, and $\beta_3$ is the coefficient of direct vessel surface regulation. Parameters $k_0$, $k_1$, $k_2$ and $r_0$, $r_1$, $r_2$ are positive.

Volume dynamics were initially characterized by water absorption–excretion differences. Absorption rates were assumed not to deviate from steady-state values, whereas excretion rates were taken as susceptible to (1) systemic blood pressure deviations, known to specifically act on renal water excretion, and (2) the IL-11-induced effects. The function of blood volume was therefore described by the difference between absorption rates, which were constant ($A_{st}$), and excretion rates, which were set as the pressure-related difference between normal steady-state excretion ($E_{st}$) and the excretion in the IL-11 scenario ($E_{drug}$):

$$\dot{V} = A_{st} - \left(\frac{P(t)}{P_{st}}\right)^{n_1} (E_{st} - E_{drug}(t)) \tag{10}$$

The power coefficient $n_1$ sets the strength of the dependence of the volume changes on the pressure variation. Assuming equal steady-state absorption and excretion rates ($A_{st} = E_{st}$) under normal conditions, Eq. 10 turns into

$$\dot{V} = E_{st}\left(1 - \left(\frac{P(t)}{P_{st}}\right)^{n_1}\right) + \left(\frac{P(t)}{P_{st}}\right)^{n_1} (E_{drug}(t)) \tag{11}$$

Excretion of volume in the therapeutic scenario (i.e., the term $E_{drug}$) was assumed to be affected by the hormones in all three models. As described above, in models 2 and 3 it was additionally assumed that the drug directly influences volume excretion. Both of these effects (IL-11 and hormone mediated) were assumed to be nonlinear and were introduced into the volume dynamics given in Eq. 11. For model 1, this gave

$$\dot{V} = E_{st}\left(1-\left(\frac{P(t)}{P_{st}}\right)^{n_1}\right)+\left(\frac{P(t)}{P_{st}}\right)^{n_1}\left(\frac{\gamma_1(H_U(t))^{q_1}}{m_1+(H_U(t))^{q_1}}-\frac{\gamma_2(H_D(t))^{q_2}}{m_2+(H_D(t))^{q_2}}\right) \tag{12}$$

while for models 2 and 3, this gave

$$\dot{V} = E_{st}\left(1-\left(\frac{P(t)}{P_{st}}\right)^{n_1}\right)+\left(\frac{P(t)}{P_{st}}\right)^{n_1}\left(\frac{\gamma_0(I_P(t))^{q_0}}{m_0+(I_P(t))^{q_0}}+\frac{\gamma_1(H_U(t))^{q_1}}{m_1+(H_U(t))^{q_1}}-\frac{\gamma_2(H_D(t))^{q_2}}{m_2+(H_D(t))^{q_2}}\right) \tag{13}$$

The IL-11-, $H_U$-, and $H_D$-afflicted rates of water excretion are denoted $\gamma_0$, $\gamma_1$, and $\gamma_2$, respectively. Parameters $m_0$, $m_1$, $m_2$, and $q_0$, $q_1$, $q_2$ are positive constants.

Subsequent calibration of the models was accomplished by obtaining and utilizing data from a human study of IL-11 safety, and their predictive ability was evaluated under different regimens of IL-11 therapy. These stages of the work are further described in Section 5.

### 4.2 Efficacy Modeling: Thrombopoietin (TPO)

As illustrated in Section 3, platelet deficiencies hamper the development of improved therapeutic regimens for hematology and oncology patients, as well as in their clinical management [22, 27]. The possible use of TPO to abrogate these disorders has also been debated (Section 3). The PK of rhTPO therapy was mathematically modeled in terms of ODEs and solved numerically using the Euler method. This extension was simulated in conjunction with the thrombopoiesis model described in Section 2 (see also Refs. 37 and 38). The independent PK model for rhTPO was similar in its structure to that of endogenous TPO, although the kinetic parameters were assumed to be different. The reason for this lies in the fact that exogenously administered synthetic molecules have been shown to differ from the endogenous analogues in a number of characteristics [32]. The model extension also assumes competitive binding of TPO and rhTPO to cell receptors. When describing the PD of the drug, its effects were considered similar to those of the endogenous protein, the latter already existing in the model. Parameter calibration of the model, as well as its retrospective and prospective validation in different species, is detailed in Section 5.

### 4.3 Modeling Combination Therapy: Chemotherapy and TPO Support

The consideration of TPO as a potential candidate for alleviating chemotherapy-induced TP in patients with malignant disorders (see Section 3) introduces the issue

of combination therapy. The addition of TPO to a cytotoxic-drug-affected scenario, allowing restoration of the damaged thrombopoietic system, can be described via minor manipulation of the basic mathematical model. To do this, one should take into account the effect of chemotherapeutic drugs on the thrombopoietic lineage. Chemotherapy nonspecifically damages all dividing cells, particularly progenitor cells of the thrombopoietic lineage. These targets are represented by the UCP and CP compartments of the TPO model (Section 2). Accordingly, the model assumes that the cytotoxic drug eliminates dividing cells in a concentration-dependent manner and, possibly (depending on the drug), in a cell-cycle phase-specific manner. The model incorporates the PK of the drug, to obtain the time course of drug concentrations in blood and bone marrow. Subsequently, a PD model for this drug is applied: the apoptosis probability of a dividing cell is calculated as a function of its position in the cycle and of the current drug concentration.

After constructing such a model and validating it, drug developers and treating physicians can explore the unlimited space of possible conjugate treatments, in which the side effects of chemotherapy are suppressed by use of exogenous TPO. The model allows examining different treatment regimens, predicting their possible outcomes, and selecting the best possible treatment strategies, while avoiding adverse effects, such as those observed in clinical trials of TPO. This identification of improved treatment strategies can be done for the general population, as well as on a patient-specific basis, the latter requiring the specific model parameters of the treated patient, which can be evaluated using the individual platelet profiles post drug intervention.

## 5 USING THE VALIDATED MODEL FOR PREDICTING IMPROVED DRUG EFFICACY/TOXICITY PROFILE

As explained in Section 3, clinical evaluation of IL-11 and TPO was obstructed due to safety problems. It was therefore imperative to investigate the mathematical models of TPO-mediated thrombopoiesis and IL-11-induced fluid retention (Sections 2 and 4), in order to suggest how to overcome the limitations involved in the development of these drugs. We show in this section how species-specific model parameters can be evaluated and the means by which retrospective and prospective validation of the mathematical models is achieved. Thereafter we discuss model use for predicting improved chemotherapy and supportive administration regimens.

### 5.1 Model Validation by Retrospective Human Study Results: IL-11 and TPO

Since modeling IL-11-mediated BVE has employed the specially developed multiple modeling approach (see Section 4), the previously depicted diverse models are considered validated and of predictive capability only upon their arrival at unified volume dynamics. The study therefore consisted of two essential steps: experimentally based calibration and parallel assessment of model simulation-derived BVE behavior. These procedures are described next [36].

In the first stage, model parameters were successfully calibrated to retrospectively retrieve clinically evaluated blood volume measurements following daily IL-11 administration of 25 μg/kg doses to healthy volunteers for a week [35]. The data

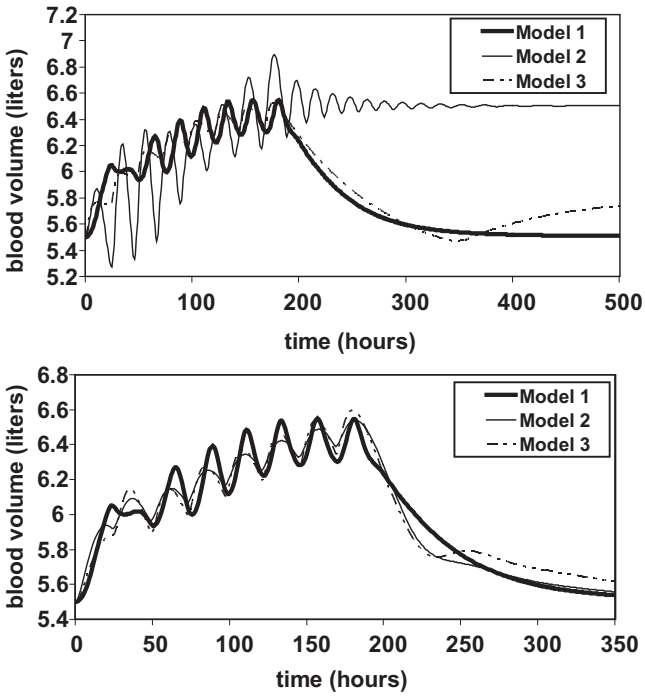


**FIGURE 3** Comparison of alternative mathematical models for BVE dynamics following calibration by results of a clinical IL-11 study. Blood volume (liters) measured during one week of IL-11 therapy (7 daily doses of 25 μg/kg) was taken from a previous study in healthy humans. Three models, calibrated with data from either the therapeutic days (upper panel) or both the therapeutic and post-therapeutic windows (lower panel), were simulated under the described experimental treatment schedule. (Data adapted from Ref. 36.)

that served for approximation were collected during the first week, that is, within the therapeutic time frame. However, subsequent simulations of the models under the same regimen failed to generate unified behavior of BVE dynamics (Fig. 3, upper panel). In contrast, when simulations were preceded by calibration of the models using data from stages both within therapy (the first week) and up to one month following treatment termination, the output dynamics were highly similar between the models (Fig. 3, lower panel). In fact, the observed high correlation between the numbers of experimental points used for model calibration, and the models' increased similarity (data not shown), strengthened the possibility that by using more data points, model similarity and thus reliability and clinical predictability can be improved.

In the second stage of the work, the predictability of the models was evaluated in a wider therapeutic spectrum. The models were simulated under IL-11 treatment strategies of various durations, various interdosing intervals, and various dose intensities, where model-derived dynamics were again compared to assess the model robustness in portraying BVE dynamics. Highest values of similarity between models, as shown by the difference index, were observed when the average IL-11 dose per day of the tested treatment was close to that of the experimental treatment

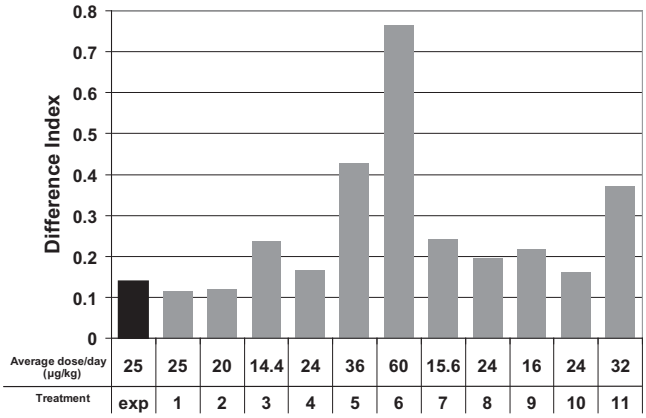


**FIGURE 4** Model-derived BVE dynamics in therapeutic regimens that are close to experimental schedules in average dose per day values. Model simulations under the experimental schedule (treatment exp) and other regimens of various durations, doses, and interdosing intervals (treatments 1–11), following calibration with clinical measurements, are compared using the difference index. (Data adapted from Ref. 36.)

of daily 25 μg/kg doses (see Fig. 4 as an example). This suggests that in this example of an obscure toxicity mechanism, one can trust the model-derived BVE predictions for various IL-11 treatments, provided their average daily dose of IL-11 is in proximity to that used in the original clinical experiment.

The mathematical method applied in the case of IL-11 shows that given preliminary experimental data from various stages of therapy, beginning at phases within the therapeutic window and ending at the recuperated steady state, the different models reliably predict dynamics of water retention within a certain span of IL-11-administration strategies. Considering the recommended IL-11 levels (10–100 μg/kg/day), and the above model-derived conclusions, clinical studies evaluating BVE within concentrations of 10, 20, 30, . . . μg/kg/day are required to form a complete predictability frame of BVE within all relevant treatment scenarios. This would then allow appraisal of the clinical applicability of each of these IL-11 therapy schedules, as well as others.

As in the above-presented process of IL-11 model evaluation, clinical information was used for the TPO case. A thrombopoiesis computational model was designed on the basis of the mathematical description of TPO-induced thrombopoiesis (elaborated in Sections 2 and 4), to enable simulation and evaluation of *in vivo* platelet formation and to suggest therapeutic TPO regimens of improved efficacy and reduced immunogenicity [38]. The computer model was adjusted and validated for murine, simian, and human cases, as will be elaborated later.

Currently, human validation of the thrombopoiesis computer model is possible only in retrospective, as the drug has not yet undergone the appropriate regulatory procedures enabling human administration. To retrospectively validate the model, it was calibrated using a set of biologically realistic parameter values, so that homeostatic behavior (i.e., a normative count of 320,000 platelets/μL blood) was retrieved in untreated conditions. Next, simulation of the computerized model under

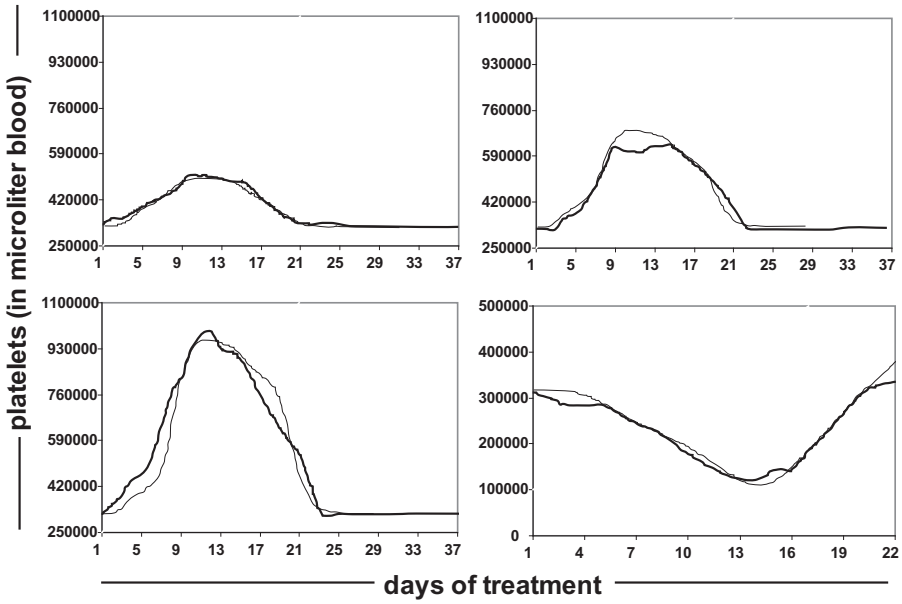


**FIGURE 5** Model predictions of platelet profiles in patients treated by TPO and chemotherapy compared to clinical results. Average platelet counts in human patients are plotted, following a single IV bolus rhTPO administration of 0.3 μg/kg (upper left panel), 1.2 μg/kg (upper right panel), or 2.4 μg/kg (lower left panel), or an IV infusion of doxorubicin for 3 days, the total dose being 90 mg/m$^2$ (lower right panel). Simulation-generated curves are shown by thick lines; clinical results are shown by thin lines.

recombinant human TPO (rhTPO) administration was carried out. Data from a clinical trial [39], which included PK information (i.e., rhTPO concentration in blood) and accurate platelet counts in response to different TPO dosages, served the calibration process: three patients were assigned to a single IV bolus administration of each of the four dosages 0.3, 0.6, 1.2, or 2.4 μg/kg. The three platelet count curves for each dosage were averaged and a single parameter set, which enabled the model to accurately retrieve platelet dynamics observed under the various treatment regimens, as shown in Fig. 5, was identified. This parameter set was determined as the representative for the "average patient."

While TPO therapies pose many questions, such as its effects on elevated platelet counts in clonal thrombocytosis, the opposite problem, namely, that of reduced platelet counts, introduces equally significant challenges. Since TP is a common dose-limiting side effect of cancer chemotherapeutics, as discussed in Sections 3 and 4, accurate predictions of the chemotherapeutically reduced platelet dynamics are of importance when planning treatments. To allow predicting the effect of cytotoxics on thrombopoiesis in individual patients, the average "Virtual Patient" was tested for its ability to retrieve the outcomes of a standard chemotherapeutic treatment by doxorubicin. The model was successfully calibrated in reference to clinical published experimental data [40], as shown by Figure 5 (lower right panel) and improved TPO regimens were suggested. Prospective validation of the generality and clinical predictability of the model is warranted. A different

biomathematical approach, elaborated in Section 7, uses data from human and animal cell cultures and from preclinical studies that are available in the pharmaceutical industry, in order to develop therapeutic strategies with minimal suppression of thrombopoiesis.

In conclusion, the thrombopoiesis computer model was successfully calibrated to accurately retrieve relevant clinical information reflecting TPO therapy, as well as chemotherapy, but it is yet to undergo prospective human validation. Importantly, the models described here for different thrombopoietic growth factors are each targeted toward opposing aspects of therapy: while the IL-11 mathematical model attempts to define the toxicity associated with drug-induced alleviation of TP, the computerized TPO model is aimed at accurately predicting the direct beneficial results of such therapy on platelet restoration.

### 5.2 Model Validation in Prospective Animal Trials: TPO Applied to the Virtual Mouse and Virtual Monkey

To use the model predictions for clinical decision making, one must still substantiate the superiority of the model-suggested improved TPO regimens. This task must be carried out in animals first, not only due to safety issues, but also because typically in such cases, the agent in question is not yet approved for the treatment of patients. Once a proof of concept in animals is provided for its power to suggest improved regimens, the model can become instrumental during the clinical phases of drug development, for avoiding excessive and unnecessary toxicity and for reducing the number of patients in Phase I of the clinical development. These steps are expected to pave the road for routine clinical use of the computer-based prediction tool.

The underlying assumption in this work was that improved TPO schedules are those employing the smallest drug doses required to achieve the desired efficacy. This assumption is based on a biomathematical study that examined the effect of dose intensities on drug immunogenicity. Results of this work suggest that for a given immunoreactive system, there exists a critical drug dose for which the relevant immunogenic threshold is expected to be exceeded, thus affecting patient health [41].

To validate the model's predictions in preclinical experiments with BALB/c mice and rhesus monkeys, the thrombopoiesis model was calibrated to reflect murine or simian thrombopoiesis. This was achieved by using literature-derived data. The species-adapted models, Virtual Mouse and Virtual Monkey, were subsequently used to adjust the improved TPO treatment to the species in question and test this treatment in animal experiments. This work is described in Ref. 38.

Parameter calibration of the thrombopoiesis computer model was performed for fine-tuning the model to accurately describe thrombopoiesis in each species. This work involved definition of a biologically realistic range of values for each parameter in murine thrombopoiesis [28, 42–47] and simian thrombopoiesis [22, 48–54], which were then entered as the initial parameter values, subsequently to be fine-tuned according to the specific tested population. Ability of the model to retrieve platelet profiles in each animal group was then tested.

Next, the prediction accuracy of the thrombopoiesis computer model was validated in mice. Specifically, by simulating the model it was predicted that platelet counts, similar to those achieved with the accepted TPO administration schedule, can also be generated under different schedules of appreciably reduced TPO doses (Fig. 6, upper panel). Hence, two different administration regimens were

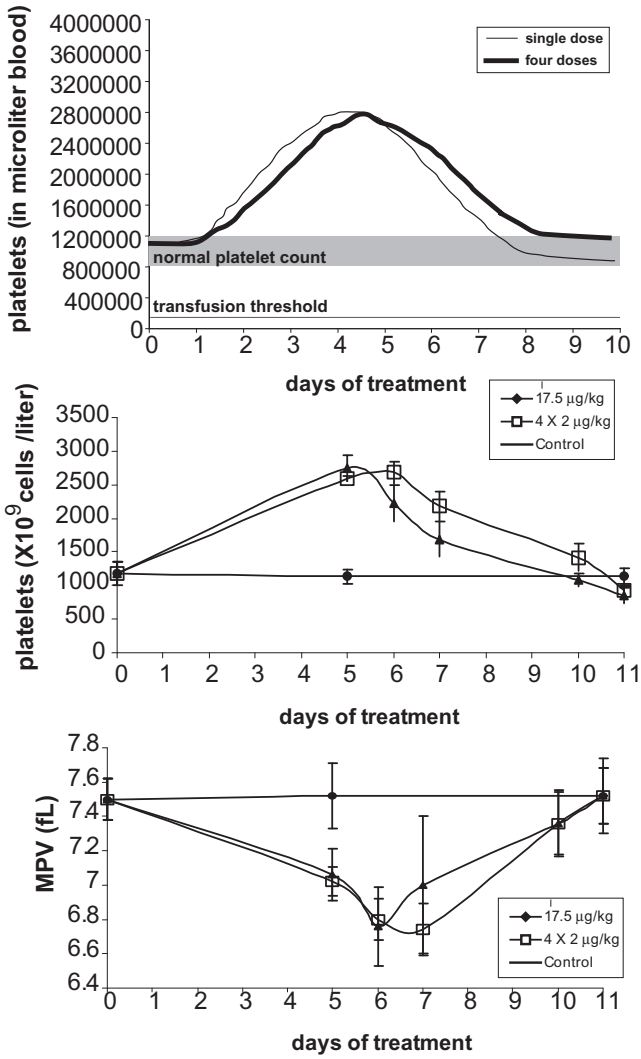


**FIGURE 6** Model predictions of platelet profiles in mice treated by TPO as compared to experimental results. (Upper panel) The thrombopoiesis model-generated predictions indicating similar efficacy in platelet formation between a "conventional" single-dose TPO treatment of 17.5 µg/kg, and a model-suggested reduced TPO treatment of four daily doses of 2 µg/kg each, comprising only 10–45% of the standard regimen. Experimental results of platelet formation (middle panel) and MPV measurements (lower panel) following conventional rmTPO therapy (black diamonds) and the alternative therapy (white squares) in mice, according to the above schedules, validate the model predictions. Treatments are initiated at day 0. Average ± SD of five mice are shown, per each entry. (Data adapted from Refs. 37 and 38.)

experimentally applied to mice clustered in two experimental arms, and their subsequent responses were compared. Mice of the first arm (A) received a standard regimen consisting of a single injection of 17.5 μg/kg recombinant murine TPO (rmTPO), which was well below the reported saturating level [42]. The second arm (B) tested the model-generated proposition that the same platelet yields can be obtained in mice receiving a total dose of 8 μg/kg rmTPO, divided over four equal daily injections, whereas the null hypothesis here was that a significantly smaller total dose of rmTPO would be less efficient in elevating the platelet counts.

The results of the experiments, presented in Fig. 6 (middle panel), clearly show that the platelet profile of arm A (a regimen of one 17.5 μg/kg rmTPO injection) is similar to the profile of arm B (applying a regimen of four daily 2 μg/kg rmTPO doses). The differences are statistically insignificant, as evaluated by the Student's t-test. The average profiles of both arm A and B peaked at fairly the same time (day 5 vs. day 6, respectively) and at similar mean platelet counts of $2741 \pm 193 \times 10^9$/L and $2685 \pm 164 \times 10^9$/L, respectively. The latter schedule resulted in a slightly extended thrombocytosis. MPV, known to decrease in platelet production induced by low dose TPO in mice [55], was compared between arm A and arm B as well (Fig. 6, lower panel). Results show that the familiar phenomenon of MPV reduction also occurred similarly in both groups. Once again, arm A had preceded arm B in reaching its nadir by about 24 hours (6.76 fL on day 6, vs. 6.74 fL on day 7, respectively). Specific adverse effects were not observed following either treatment.

In order to evaluate in rhesus monkeys the efficacy of the treatment regimen, already validated in mice, we first checked the ability of the simian thrombopoiesis model to predict individual monkey responses to different TPO treatments. Next, the simian response to the model-suggested schedule was simulated and the resulting predictions were compared with the platelet counts that were experimentally obtained under the model-suggested schedule.

To this end, model parameters were first evaluated to fit the empirical platelet profile of a single rhesus monkey. The quantitatively adequate simulations of thrombopoiesis response to treatment were apparent from their resemblance to the empirical results (Fig. 7). This calibrated model was further validated in several other monkeys, each receiving a different drug schedule (see schedule in Fig. 8, top panel).

Results in Fig. 9 demonstrate that the model predictions run remarkably close to the empirical data points. Note that as the platelet counts were monitored in these monkeys only once weekly, a mismatch is introduced into the comparison between the empirical results and the daily predicted counts. Thus, there is no reason to assume that the peaks in the predicted responses on days that were not tested empirically did not actually take place.

### 5.3 Predicting the Optimal TPO Efficacy/Toxicity Ratio in Monkeys

Once verified, the Virtual Rhesus model was simulated for identifying an improved TPO administration regimen. The model-based regimen was then evaluated for verifying its improved efficacy and safety profile. This was done in two monkeys (see Fig. 8, bottom panel). The resulting elevation in platelet counts peaked at 1700–

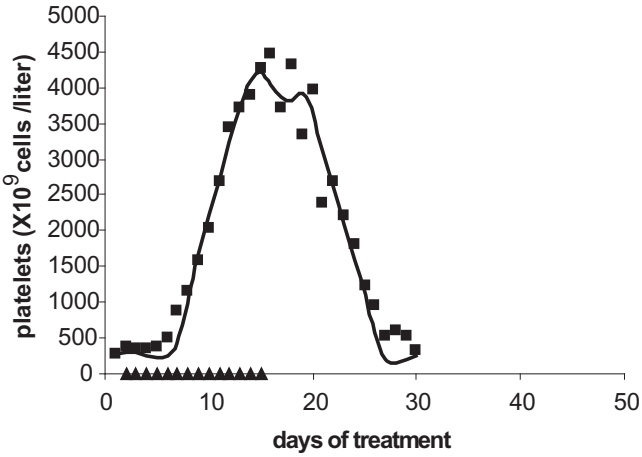


**FIGURE 7** Comparison of model predictions to empirical data of TPO therapy in a rhesus monkey. Daily experimental measurements (rectangles) are shown as well as model simulations (solid line) of platelet counts of one monkey treated by TPO. Fourteen daily doses of 5 μg/kg TPO were applied, administration days marked by triangles. (Data adapted from Refs. 37 and 38.)

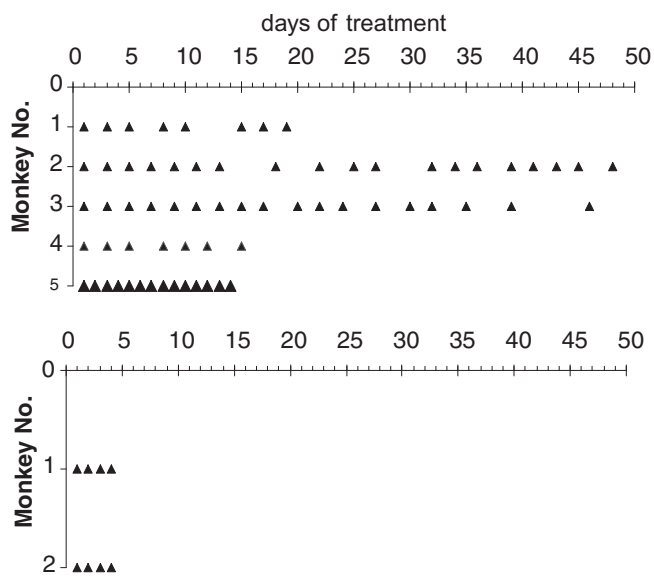


**FIGURE 8** TPO administration schedules applied in simians for model validation. (Upper panel) Five different schedules, comprising various combinations of 5 μg/kg doses of recombinant full-length rhesus monkey TPO (each dose indicated by triangles), were individually applied to the studied monkeys. (Lower panel) Model-suggested TPO schedules (also using doses of 5 μg/kg) applied to two monkeys.

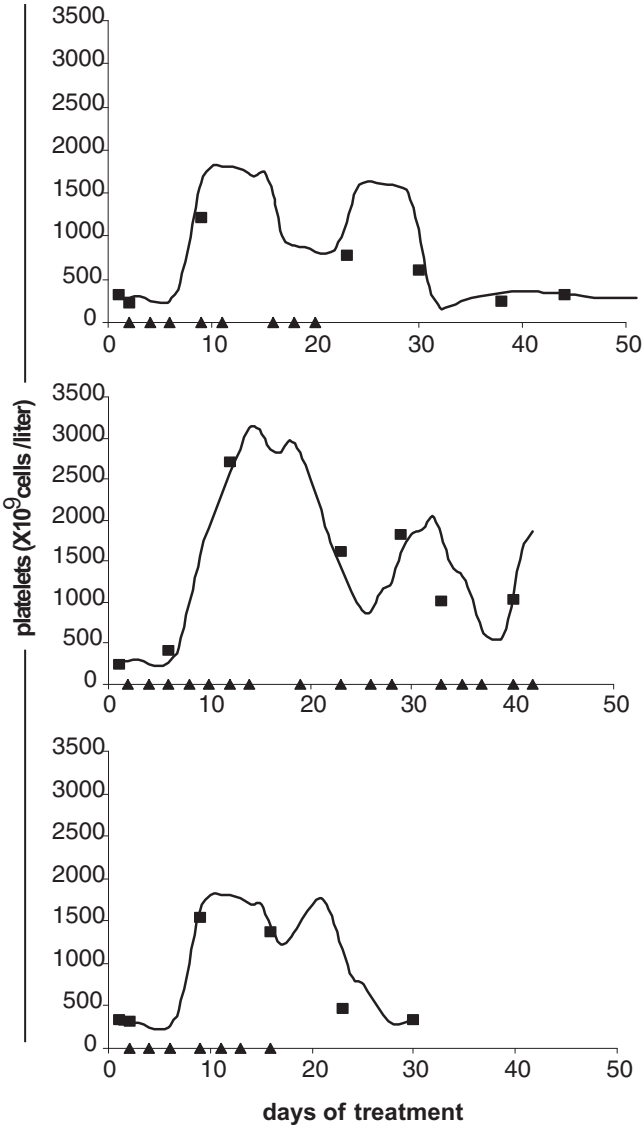


**FIGURE 9** Comparison of model predictions to experimental platelet profiles of rhesus monkeys under various TPO treatment schedules. Three monkeys (upper, middle, and lower panels) were subjected to individual TPO regimens, consisting of multiple dosing treatments of 5 μg/kg TPO, administration days marked by triangles. Experimental measurements of platelet counts in blood were performed weekly or biweekly (squares) and are plotted against model simulations (solid line).

$2400 \times 10^9$/L following each treatment cycle. The results of a second cycle of the same treatment were similar to those of the first one in showing no neutralizing antibodies. Only when the dosing cycle was applied for the third time were low antibody titers detectable, which did not jeopardize TPO's efficacy and did not result in decreased platelet counts. The individual empirical responses of the two monkeys

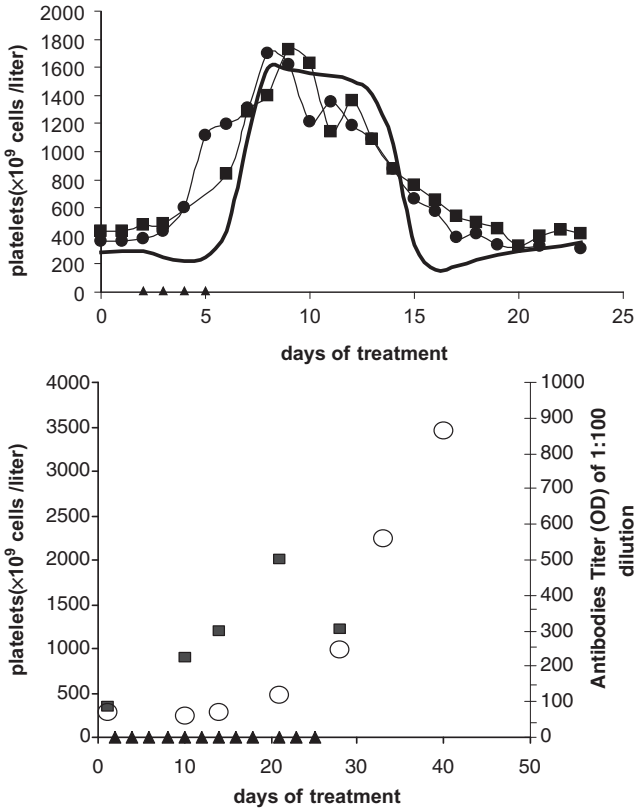


**FIGURE 10** Superiority of the model-suggested TPO treatment schedule validated in rhesus monkeys, as compared to other regimens. (Upper panel) Empirical platelet profiles of monkeys under the model-suggested TPO regimen of four daily 5 μg/kg doses (solid squares and circles) are plotted; administration days are marked by triangles. (Lower panel) Antibody titer and platelet profiles of a monkey undergoing an arbitrary treatment by 12 TPO doses.

to the first treatment cycle are shown in Fig. 10 (top panel) (to be compared with a highly immunogenic response of a monkey who underwent another TPO treatment, Fig. 9, bottom panel).

These results show that the model-predicted administration regimen of smaller TPO doses with an interval of 24 hours yields platelet counts that are equal in effect to the more than double dose regimen, administered as a bolus injection. This experiment joins previous *in vivo* experiments with other drugs in verifying a mathematical theory discussed in Section 1.1, which essentially suggests that treatment efficacy can be decisively influenced by modulating the interdosing interval according to the internal cell kinetics. The monkey experiment supports the model prediction of an equally efficient and nonimmunogenic treatment, if dose is fractionated by an appropriate dosing interval. Importantly, this schedule, adopted for improving safety of TPO administered for platelet harvesting in monkeys, has been in use for the last five years or so. No adverse effects have been associated with this model-suggested treatment.

In general terms, the experimental results in both animal species supported and validated the predictions of the mathematical model that more efficacious treatment with TPO can be implemented. Although intuitively it may not seem beneficial for the patient to replace a single injection by multiple doses, it is clearly of significance for safety considerations.

The computerized mathematical model is capable of successfully predicting treatment scenarios not yet tested experimentally, as well as of identifying safer regimens. This tool is expected to be of aid in suggesting improved drug strategies for an individual or for a patient population. Human trials are necessary for testing the suggested improved TPO strategies, possibly in conjunction with chemotherapy.

## 6 USING THE THROMBOPOIESIS MODEL FOR PREDICTING AN UNKNOWN ANIMAL TOXICITY MECHANISM

Several pathophysiological mechanisms can account for drug-induced TP (see Section 3), and the actual underlying mechanism in each particular case can have important therapeutic implications. Yet, no easily applicable noninvasive laboratory tests are available for distinguishing between the different mechanisms for TP. In this section, we describe the manner by which mathematical modeling can help in understanding the dynamics and etiology of drug-induced TP.

A monoclonal antibody-based drug, developed against a platelet-unrelated antigen and currently under preclinical evaluation, was found to possess cross-reactive effects eliciting severe TP in monkeys. Despite routine laboratory testing and histological examination of bone marrow of affected monkeys, the drug's mechanism for inducing TP has not been elucidated. In an attempt to decipher the underlying mechanism, we used the mathematical model for TPO-regulated thrombopoiesis [37, 38], which was previously formed (as described in Section 2) and validated *in vivo* (as depicted in Sections 4 and 5). This model was extended to include the effect of the drug on MKs and platelets. Since the mechanism is still unknown, several biologically plausible alternatives were examined. Antibodies can exert their effect either by antibody-assisted phagocytosis or by complement-mediated direct toxicity [56]. In the first mechanism, antibody-covered platelets or platelet-forming MKs would be efficiently removed by macrophages that bear specific antibody-recognizing receptors. In the latter case, platelet-bound antibodies activate the complement pathway that leads to cell membrane perforation and cell destruction (Fig. 11).

Mathematical implementation of both antibody-mediated killing mechanisms into the PD-extended model was performed as follows. For complement-mediated toxicity, an "accumulated hits" model, according to which each platelet suffers from cumulative damage ("hits") induced by antibody-mediated complement binding, was suggested. Upon accumulation of a certain number of "hits," the platelet is permanently damaged. The resulting probability density function, $P(t)$, describing a fraction of platelets with a given number of "hits" in time $t$ postexposure, is given by the gamma function

$$P(t) = \frac{(\alpha t)^m e^{-\alpha t} \alpha}{m!}$$

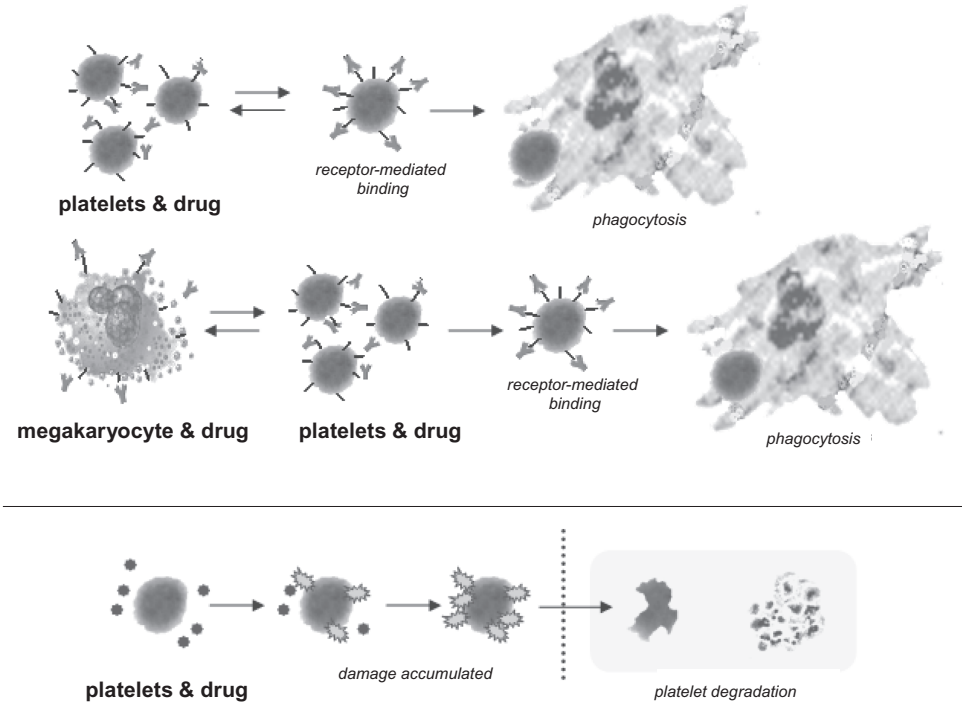


**FIGURE 11** The "phagocytosis" putative mechanism and the "hits" putative mechanism of drug-induced thrombocytopenia. In the "phagocytosis" mechanism (upper panel), drug molecules bind to platelet and megakaryocyte surface antigens, leading to their phagocytosis by the reticuloendothelial system. In the "hits" mechanism (lower panel), drug molecules bind to platelet surface antigens, leading to complement activation and direct destruction of the cells.

where $m$ is the number of hits needed for platelet destruction, and $\alpha$ is the rate of hit formation. The rate of hits is given by a linear function of the antibody concentration in blood:

$$\alpha = aX + b$$

where $X$ is antibody concentration, and $a$ and $b$ are parameters.

For the case of antibody-mediated phagocytosis, the "phagocytosis" model, it was assumed that the rate of platelet "capture" by macrophages has a Hill's function form, depending on the number of antibodies that are bound to the platelet's surface. This is described as follows:

$$B(i) = \frac{B_{\max} i^z}{K^z + i^z}$$

where $B(i)$ is the rate of capture of platelets that possess $i$ bound antibodies, and $K$ and $z$ are parameters. Since binding and dissociation of antibodies from the

platelet surface are stochastic processes, the distribution of platelets with respect to the number of bound antibodies, $V_i$, was explicitly calculated, so that the rate of antibody-induced platelet elimination is given by

$$R = \sum_i V_i B_i$$

Antibodies can act against mature platelets, bone marrow-derived MKs, or both. Hence, several combinations of these suggested drug-toxicity mechanisms were implemented in the extended model. Retrospective results of real-life experiments in a number of individual rhesus monkeys [37, 38] were used for comparison with the simulated outcomes of these various alternative models. For each combination, the PD-associated parameters of the extended model were adjusted to best fit to these experimental results. Only one of the combinations was expected to consistently retrieve experimental data, and thus be identified as the preferable mechanism underlying this antibody-mediated TP. Systemic platelet profiles in individual monkeys following administration of multiple or single doses of the drug were compared with the best-fit profiles produced by models with different drug effect mechanisms (Fig. 12). The analysis of these simulations indicated that the "hits" model failed to predict dynamics of the drug-induced TP, while the "phagocytosis" model coincided with experimental data. Model results also show that by assuming toxicity to both platelets and MKs, rather than to just platelets, more precise predictions can be yielded.

We conclude that the developed computerized thrombopoiesis model superimposed by mathematical models for putative toxicity mechanisms can retrieve adverse thrombocytopenic effects recorded empirically in simians. Different models of drug-toxicity mechanisms yield different profiles of blood platelets, yet the antibody-mediated phagocytosis of platelets and MKs seems the most plausible mechanism of the drug-induced TP in this situation. In this manner, mathematical modeling can help decipher the underlying mechanisms of other disorders.

## 7 TRANSITION TO PHASE I

Drug development is a costly and time-consuming endeavor, due to its requirement for enormous collections of data concerning the efficacy and safety of the proposed drug. The ability to extrapolate toxicity and efficacy data from animal to human is therefore of great importance for efficiently advancing development of pharmaceuticals. Yet, despite the efforts invested in this direction, and the great advances achieved in recent years, extrapolation of data from animals to humans is still considered unreliable [57–59]. In this section, we provide a detailed description of an algorithm, which in conjunction with data from cell cultures and animal experiments, can be used for the modeling of clinically stimulated (drug-induced) TP in humans, and we show how this method enables good clinical predictability of this disorder.

As discussed in Section 3, TP can result from therapeutic intervention, such as anticancer chemotherapy. The mathematical model of thrombopoiesis, described in Section 2, was used as a basic model for retrieving clinical cytotoxic drug-induced

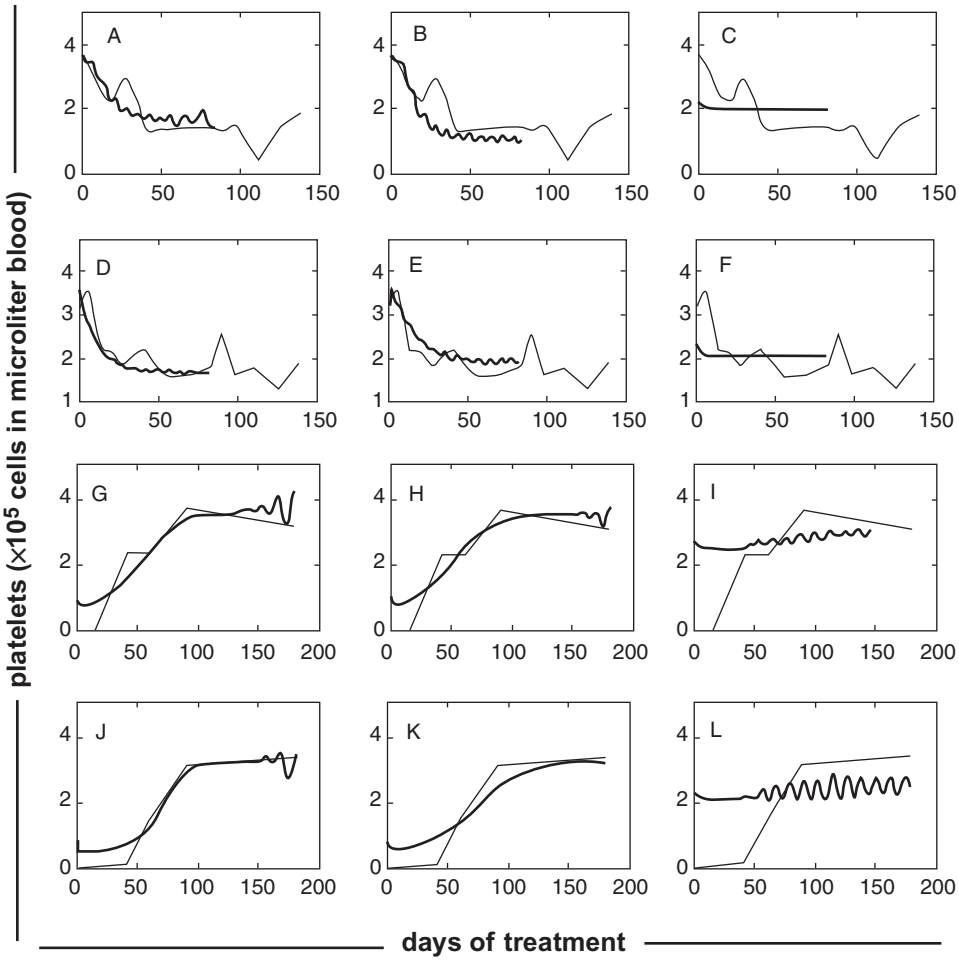


**FIGURE 12** Model retrieval of drug-induced thrombocytopenia in rhesus monkeys. Platelet dynamics derived from experiments in four monkeys, receiving a treatment of multiple doses (A–F) or a single high dose (G–L) of the drug. The effects of each treatment on the monkey's platelet profile were simulated by the three mathematical models of alternative drug-toxicity mechanisms. Model predictions were then statistically compared with the experimental results. Simulation results are for the "phagocytosis" model assuming that the drug targets platelets alone (left column panels), for the "phagocytosis" model assuming that the drug targets both platelets and megakaryocytes (middle column panels), or for the "hits" model assuming that the drug targets only platelets (right column panels). Panels A–C, D–F, G–I, and J–L correspond to data from each of the four monkeys. Empirical data (thin lines) is shown versus simulation-derived dynamics (thick lines).

TP. Yet to obtain a reasonable approximation of the predicted toxicity in humans, based on preclinical data, additional assumptions and simplifications are required. These are detailed next.

First, translation of drug PD from the preclinical to the clinical stage is necessary. One must consider (1) differences between *in vitro* and *in vivo* data and

(2) differences between PD parameters of different species. Both variation types involve built-in inaccuracies. With respect to the comparison between *in vitro* and *in vivo* experiments, a few issues must be taken into account. Cells *in vitro* are more sensitive to a drug, and thus more susceptible to drug-induced apoptosis. They have higher mitotic indices due to less contact inhibition and are usually exposed to constant levels of drug concentration over time. Cells in the physiological, whole-organism context, however, are of lower sensitivity to a drug; the surrounding microenvironment, consisting of high proliferating and less drug-susceptible cells, provides a constant influx of healthy cells into the total pool, thereby limiting the damage caused by the drug [60]. Differences in exposure to growth factors between the two experimental settings exist as well. Drug sensitivity differs not only between *in vitro* and *in vivo* settings, but also in various species, which may differ in size of cells, mitotic indices of various compartments and differential stages of cells, apoptotic indices, transit times, and sensitivity to hormones [61].

In order to replicate clinical use of the drug, the basic mathematical model is extended to incorporate an additional formula for the drug PD. It is assumed that bone marrow cells are exposed to systemic drug concentration at any time, as the tissue is highly vascularized. The model also assumes that cells at different developmental stages, denoted by a subscript $X$ (e.g., UCP, CP, EP), are of diverse sensitivity to the drug. The functional relation between the concentration of the cytotoxic drug in blood ($CD_{\mathrm{bl}}$) and the cytotoxic effect ($CE$) is given by the equation

$$CE_X(t) = f(CD_{\mathrm{bl}}(t)) \tag{14}$$

where $CE_X$ denotes the fraction of bone marrow thrombopoietic cells of a differentiation stage $X$, eliminated following exposure to $CD_{\mathrm{bl}}$ at time $t$ ($CD_{\mathrm{bl}}(t)$). The dependence is represented by function $f$, where $0 \leq f\,(CD_{\mathrm{bl}}(t)) \leq 1$ for any $t$.

Most cytotoxic drugs target mitotic phases of the cell cycle and are thereby more harmful to proliferating, rather than resting, cells. Accordingly, by further subdividing the model compartments into cell-cycle phases (G0-G1 and S-G2-M subcompartments), a more accurate prediction of drug toxicity can be obtained. For example, in the above model, representation of UCP and CP compartments (see Section 2) uses cell-cycle length instead of amplification rate. This allows us to calculate, not only the total number of cells remaining at every time point after drug application but also the number of cells at every age and cell-cycle phase. In EP compartments, the age and the position in the endomitotic cycle are synchronized; thus there is no need for additional distribution. The rest of the compartments, consisting of nondividing cells, are not considered susceptible to the standard cytotoxic drug and are therefore not subdivided in this manner. Cytotoxic drugs that selectively target these stages can be addressed after subdivision of those compartments. In this complex form of the model, parameters (e.g., compartments, sizes, transit times) are evaluated differently for all the species examined, while the model equations themselves retain the same structure.

Following adaptation of the model equations to include the PD of the drug, parameters are calibrated using the available experimental information. Preclinical toxicity data to be used by the model consist of the platelet measurements over time in one animal of a chosen species (S), which has shown (in previous *in vitro* studies) higher sensitivity to the drug than other species. Various sets of data are obtained

for a wide range of drug doses, so that different thrombocytopenic responses, from low toxicity to grade 3–4 toxicities, are considered.

The model also exploits data from cell culture assays. For example, for the estimation of the cytotoxic drug concentration, the model uses $IC_{20}$, $IC_{50}$, and $IC_{90}$ (the concentrations of the drug leading to inhibition of 20%, 50%, and 90% of *in vitro* cell growth, respectively, as compared to untreated cells). These are obtained from toxicity tests of the platelet lineage-committed bone marrow cells (the CFU-Meg assay), in both human and species S-derived cells [62]. In the complex form of the model, one may also use experimental tools of cell-cycle analysis, such as standard cell-sorting assays.

Based on this data, three functions $f$ (Eq. 14), denoted as $f_{\text{inVitroHu}}$, $f_{\text{inVitroS}}$, and $f_{\text{inVivoS}}$ for human cell cultures, species S cell cultures, and preclinical S species studies, respectively, can be evaluated. In estimating the function $f_{\text{inVivoS}}$, it is necessary to consider only the drug's killing effect on bone marrow cells of compartment $X$, the depletion of which is mainly responsible for platelet nadir (see Eq. 14). $X$ is identified by the time of the nadir after drug application, taking into account the transit times of the different bone marrow compartments. As a first approximation, only one compartment $X$ of drug-sensitive cells is assumed.

The parameters of function $f$ in the clinical scenario ($f_{\text{inVivoHu}}$), which cannot be estimated as in the other three cases due to lack of data, are thereby calculated according to their counterparts. First, it is assumed that between different species, there is a linear relation between the drug killing effect observed *in vivo* and *in vitro* (i.e., the $f_{\text{inVivo}}/f_{\text{inVitro}}$ ratio is constant between diverse species). The function $f_{\text{inVivoHu}}$ is thereby obtained by shifting the IC values on the concentration axis according to this relation, using the *in vitro* data of $IC_{50}$ ($IC_{50\text{inVitroHu}}$) and $IC_{90}$ ($IC_{90\text{inVitroHu}}$). The relation itself is found by comparing the *in vitro* and *in vivo* data for species S. Then the shift of $f_{\text{inVivoS}}$ to $f_{\text{inVivoHu}}$ on the concentration axis for $IC_{50}$ is computed by the equation

$$IC_{50\,\text{inVivoHu}} = \frac{IC_{50\,\text{inVivoS}} \cdot IC_{50\,\text{inVitroHu}}}{IC_{50\,\text{inVitroS}}} \tag{15}$$

The same method can be applied for $IC_{90}$.

The last step requires simulations of the model with the human parameters, under different regimens, applying the $f_{\text{inVivoHu}}$ effect on the same type of bone marrow cells as in species S. The clinical toxicity profile of the drug—that is, the induction and severity of drug-induced TP in humans—is therefore predictable. Importantly, this estimation is a first approximation and a safety factor should be considered prior to initiation of clinical trials.

## 8 CONCLUSION

- Mathematical models are instrumental in increasing the understanding of intricate drug–patient interactions. Specifically, a mathematically based method, termed the Z-method, has been devised for increasing the efficacy/toxicity ratio of chemotherapeutics. The method suggests that safety of cell-cycle-specific drugs can be increased by resonating drug-pulsing with the internal periodicity

of target host cells replication. Experimental verification of the Z-method in mice provides a proof-of-concept for the utility of biomathematics in drug development.

- Increasing model complexity, a prerequisite for the quantitative precision of its conclusions, was performed in order to provide a practical tool for predicting drug effects on platelet counts. To this end, thrombopoiesis models of various mammal species have been developed. Models of TPO supportive treatment have been prospectively validated in mice and monkeys.
- Toxicity effects of IL-11 therapy and chemotherapy have been modeled and various drug regimens were tested. Model validation under data limitation has been examined and the need to collect patient's recovery data has been emphasized. The range of the treatment schedules under which the models could reliably predict the outcomes of drug therapy under obscured biological knowledge has been defined.
- The thrombopoiesis model has been employed for suggesting TPO administration schedules, which would maximize drug safety while maintaining its efficacy. Dose fractionation, having a well-calculated, species-dependent dosing interval, preserves drug effect on platelet counts, yet reduces its immunologic adverse effects below the clinically relevant threshold.
- The above schedule, adjusted by the murine thrombopoiesis model, was validated in mice for efficacy. The monkey-adjusted schedule derived from simulations of the simian thrombopoiesis model, was validated in rhesus monkeys for both efficacy and toxicity. Now the human model may be employed for improving safety of TPO mimetics and other thrombocytopenia-alleviating drugs.
- To the best of our knowledge, this is the first instance of a mathematical model generating quantitative predictions, the precision of which was validated in the preclinical setting.
- The thrombopoiesis model can be used, in conjunction with alternative mathematical models, for deciphering unknown drug-toxicity mechanisms in different species. This can be achieved by comparing preclinical study results with the simulated drug-affected platelet profiles, generated for different models of putative drug-toxicity mechanisms.
- In order to facilitate the transition from preclinical studies to Phase I, the mathematical models for the effect of investigational drugs on different animal species can be adapted accordingly and simulated in the human patient model for predicting parameters, the early estimation of which can economize Phase I clinical trials.

## REFERENCES

1. FDA White Paper on Innovation and Stagnation in Drug Industry, Challenges and Opportunities on the Clinical Paths to New Medical Products, March 2004.
2. Agur Z. The effect of drug schedule on responsiveness to chemotherapy. *Ann NY Acad Sci* 1986;504:274–277.
3. Agur Z, Arnon R, Schechter B. Reduction of cytotoxicity to normal tissues by new regimes of cell-cycle phase-specific drugs. *Math Biosci* 1988;92:1–15.

4. Agur Z, Arnon R, Sandak B, Schechter B. The effect of the dosing interval on myelotoxicity and survival in mice treated by cytarabine. *Eur J Cancer* 1992;28A(6/7):1085–1090.
5. Ubezio P, Tagliabue G, Schechter B, Agur Z. Increasing 1-beta-D-arabinofuranosylcytosine efficacy by scheduled dosing interval based on direct measurement of bone marrow cell kinetics. *Cancer Res* 1994;54:6446–6451.
6. Agur Z. Persistence in uncertain environments. In Freedman HI, Strobeck C, Eds. *Population Biology, Lecture Notes in Biomathematics*. Heidelberg: Springer-Verlag; 1982, pp 125–132.
7. Agur Z. Randomness, synchrony and population persistence. *J Theor Biol* 1985;112: 677–693.
8. Agur Z, Deneubourg JL. The effect of environmental disturbances on the dynamics of marine intertidal populations. *Theor Population Biol* 1985;27(1):75–90.
9. Cojocaru L, Agur Z. A theoretical analysis of interval drug dosing for cell-cycle-phase-specific drugs. *Math Biosci* 1992;109:85–97.
10. Kheifetz Y, Kogan Y, Agur Z. Long-range predictability in models of cell populations subjected to phase-specific drugs: growth-rate approximation using properties of positive compact operators. *Math Models Methods Appl Sci* 2006;16(7):1–18.
11. Agur Z, Arnon R, Sandak B, Schechter B. Zidovudine toxicity to murine bone marrow may be affected by the exact frequency of drug administration. *Exp Hematol* 1991; 19:364–368.
12. Arakelyan L, Vainstein V, Agur Z. A computer algorithm describing the process of vessel formation and maturation, and its use for predicting the effects of anti-angiogenic and anti-maturation therapy on vascular tumor growth. *Angiogenesis* 2002;5:203–214.
13. Ribba B, Alarcon T, Marron K, Maini PK, Agur Z. Doxorubicin treatment efficacy on non-Hodgkin's lymphoma: computer model and simulations. *Bull Math Biol* 2005;67: 79–99.
14. Vainstein V, Ginosar Y, Shoham M, Ranmar D, Ianovski A, Agur Z. The complex effect of granulocyte on human granulopoiesis analyzed by a new physiologically-based mathematical model. *J Theor Biol* 2005;234(3):311–327.
15. Agur Z, Ziv I, Shohat R, Wick M, Webb C, Hankins D, Arakelyan L, Sidransky D. Using a novel computer technology for tailoring targeted and chemotherapeutic drug schedules to the individual patient. *First AACR International Conference on Molecular Diagnostic in Cancer Therapeutic Development. Chicago IL*, 2006.
16. Hoffman R, Benz EJ, Shattil SJ, Furie B, Cohen HJ, Silberstein LE, McGlave P. *Hematology*, 3rd ed. London: Churchill Livingstone; 2000.
17. Kaushansky K. The molecular mechanisms that control thrombopoiesis. *J Clin Invest* 2005;115:3339–3347.
18. Kaushansky K. Lineage-specific hematopoietic growth factors. *N Engl J Med* 2006; 354:2034–2045.
19. Schulze H, Shivdasani RA. Mechanisms of thrombopoiesis. *J Thromb Haemost* 2005; 3:1717–1724.
20. Cines DB, Bussel JB, McMillan RB, Zehnder JL. Congenital and acquired thrombocytopenia. In *Hematology, the American Society of Hematology Education Program Book*: 2004, pp 390–406.
21. McCrae KR, Bussel JB, Mannucci PM, Cines DB. Platelets: an update on diagnosis and management of thrombocytopenic disorders. In *Hematology, the American Society of Hematology Education Program Book*: 2001, pp 282–305.
22. Demetri GD. Targeted approaches for the treatment of thrombocytopenia. *Oncologist* 2001;6(Suppl 5):15–23.

23. Schwertschlag US, Trepicchio WL, Dykstra KH, Keith JC, Turner KJ, Dorner AJ. Hematopoietic, immunomodulatory and epithelial effects of interleukin-11. *Leukemia* 1999; 13(9):1307–1315.
24. Tepler I, Elias L, Smith JW 2nd, Hussein M, Rosen G, Chang AY, Moore JO, Gordon MS, Kuca B, Beach KJ, Loewy JW, Garnick MB, Kaye JA. A randomized placebo-controlled trial of recombinant human interleukin-11 in cancer patients with severe thrombocytopenia due to chemotherapy. *Blood* 1996;87(9):3607–3614.
25. Haznedaroglu IC, Goker H, Turgut M, Buyukasik Y, Benekli M. Thrombopoietin as a drug: biologic expectations, clinical realities, and future directions. *Clin Appl Thromb/Hemost* 2002;8:193–212.
26. Basser R. The impact of thrombopoietin on clinical practice. *Curr Pharm Design* 2002; 8(5):369–377.
27. Elting LS, Rubenstein EB, Martin CG, Kurtin D, Rodriguez S, Laiho E, Kanesan K, Cantor SB, Benjamin RS. Incidence, cost, and outcomes of bleeding and chemotherapy dose modification among solid tumor patients with chemotherapy-induced thrombocytopenia. *J Clin Oncol* 2001;19:1137–1146.
28. Kaushansky K, Lin N, Grossman A, Humes J, Sprugel KH, Broudy VC. Thrombopoietin expands erythroid, granulocyte-macrophage, and megakaryocytic progenitor cells in normal and myelosuppressed mice. *Exp Hematol* 1996;24:265–269.
29. Li J, Yang C, Xia Y, Bertino A, Glaspy J, Roberts M, Kuter DJ. Thrombocytopenia caused by the development of antibodies to thrombopoietin. *Blood* 2001;98(12):3241–3248.
30. Vadhan-Raj S. Recombinant human TPO: clinical experience and *in vivo* biology. *Semin Hematol* 1998;35:261–268.
31. Vadhan-Raj S. Clinical experience with recombinant human thrombopoietin in chemotherapy-induced thrombocytopenia. *Semin Hematol* 2000;37(2 Suppl 4):28–34.
32. Vadhan-Raj S. Recombinant human thrombopoietin in myelosuppressive chemotherapy. *Oncology (Williston Park)* 2001;15(7 Suppl 8):35–38.
33. Vadhan-Raj S, Cohen V, Bueso-Ramos C. Thrombopoietic growth factors and cytokines. *Curr Hematol Rep* 2005;4(2):137–144.
34. Wolff SN, Herzig R, Lynch J, Ericson SG, Greer JP, Stein R, Goodman S, Benyunes MC, Ashby M, Jones DV Jr, Fay J. Recombinant human thrombopoietin (rhTPO) after autologous bone marrow transplantation: a phase I pharmacokinetic and pharmacodynamic study. *Bone Marrow Transplant* 2001;27(3):261–268.
35. Dykstra KH, Rogge H, Stone A, Loewy J, Keith JC Jr, Schwertschlag US. Mechanism and amelioration of recombinant human interleukin-11 (rhIL-11)-induced anemia in healthy subjects. *J Clin Pharmacol* 2000;40(8):880–888.
36. Kheifetz Y, Elishmereni M, Horowitz S, Agur Z. Fluid-retention side-effects of the chemotherapy-supportive treatment interleukin-11: mathematical modelling as affected by data availability. *Comput Math Methods Med* 2006;7(2–3):71–84.
37. Skomorovski K, Agur Z. A new method for predicting and optimizing thrombopoietin (TPO) therapeutic protocols in thrombocytopenic patients and in platelet donors [abstract]. *Hematol J* 2001;1(Suppl 1):185.
38. Skomorovski K, Harpak H, Ianovski A, Vardi M, Visser TP, Hartong S, Van Vliet H, Wagemaker G, Agur Z. New TPO treatment schedules of increased safety and efficacy: preclinical validation of a thrombopoiesis simulation model. *Br J of Haematol* 2003; 123(4):683–691.
39. Vadhan-Raj S, Murray LJ, Bueso-Ramos C, Patel S, Reddy SP, Hoots WK, Johnston T, Papadopolous NE, Hittelman WN, Johnston DA, Yang TA, Paton VE, Cohen RL, Hellmann SD, Benjamin RS, Broxmeyer HE. Stimulation of megakaryocyte and platelet

production by a single dose of recombinant human thrombopoietin in patients with cancer. *Ann Intern Med* 1997;126:673–681.

40. Vadhan-Raj S, Patel S, Bueso-Ramos C, Folloder J, Papadopolous N, Burgess A, Broemeling LD, Broxmeyer HE, Benjamin RS. Importance of predosing of recombinant human thrombopoietin to reduce chemotherapy-induced early thrombocytopenia. *J Clin Oncol* 2003;21:3158–3167.
41. Castiglione F, Selitser V, Agur Z. The analyzing hypersensitive to chemotherapy in a cellular automata model of the immune system. In Preziosi L. Ed. *Cancer Modelling and Simulation*. Boca Raton, FL: CRC Press; 2003.
42. Arnold JT, Daw NC, Stenberg PE, Jayawardene D, Srivastava DK, Jackson CW. A single injection of pegylated murine megakaryocyte growth and development factor (MGDF) into mice is sufficient to produce a profound stimulation of megakaryocyte frequency, size, and ploidization. *Blood* 1997;89:823–833.
43. Ebbe S, Stohlman F. Megakaryocytopoiesis in the rat. *Blood* 1965;26:20–35.
44. Luoh S, Stephanich E, Solar G, Steinmetz H, Lipari T, Pestina TI, Jackson CW, de Sauvage FJ. Role of the distal half of the c-Mpl intracellular domain in control of platelet production by thrombopoietin *in vivo*. *Mol Cell Biol* 2000;20:507–515.
45. Schermer S. *Blood Morphology of Laboratory Animals*. Philadelphia: FA Davis Co; 1967.
46. Siegers MP, Feinendegen LE, Lahiri SK, Cronkite EP. Relative number and proliferation kinetics of hemopoietic stem cells in the mouse. *Blood Cells* 1979;5:211–236.
47. Ulich TR, del Castillo J, Senaldi G, Cheung E, Roskos L, Young J, Molineux G, Guo J, Schoemperlen J, Munyakazi L, Murphy-Filkins R, Tarpley JE, Toombs CF, Kaufman S, Yin S, Nelson AG, Nichol JL, Sheridan WP. Megakaryocytopoiesis: the prolonged hematologic effects of a single injection of PEG-rHuMGDF in normal and thrombocytopenic mice. *Exp Hematol* 1999;27:117–130.
48. De Serres M, Yeager RL, Dillberger JE, Lalonde G, Gardner GH, Rubens CA, Simkins AH, Sailstad JM, McNulty MJ, Woolley JL. Pharmacokinetics and hematological effects of the pegylated thrombopoietin peptide mimetic GW395058 in rats and monkeys after intravenous or subcutaneous administration. *Stem Cells* 1999;17:316–326.
49. Farese AM, Hunt P, Boone T, MacVittie TJ. Recombinant human megakaryocyte growth and development factor stimulates thrombopoiesis in normal nonhuman primates. *Blood* 1995;86:54–59.
50. Harker LA, Hunt P, Marzec UM, Kelly AB, Tomer A, Hanson SR, Stead RB. Regulation of platelet production and function by megakaryocyte growth and development factor in nonhuman primates. *Blood* 1996;87:1833–1844.
51. Harker LA, Marzec UM, Hunt P, Kelly AB, Tomer A, Cheung E, Hanson SR, Stead RB. Dose–response effects of pegylated human megakaryocyte growth and development factor Of platelet production and function in nonhuman primates. *Blood* 1996;88: 511–521.
52. Neelis KJ, Hartong SC, Egeland T, Thomas GR, Eaton DL, Wagemaker G. The efficacy of single-dose administration of thrombopoietin with coadministration of either granulocyte/macrophage or granulocyte colony-stimulating factor in myelosuppressed rhesus monkeys. *Blood* 1997;90:2565–2573.
53. Sola MC, Christensen RD, Hutson AD, Tarantal AF. Pharmacokinetics, pharmacodynamics, and safety of administering pegylated recombinant megakaryocyte growth and development factor to newborn rhesus monkeys. *Pediatr Res* 2000;47:208–214.
54. Wagemaker G, Hartong SC, Neelis KJ, Egeland T, Wognum AW. *In vivo* expansion of hemopoietic stem cells. *Stem Cells* 1998;16(Suppl 1):185–191.

55. Kabaya K, Akahori H, Shibuya K, Nitta Y, Ida M, Kusaka M, Kato T, Miyazaki H. *In vivo* effects of pegylated recombinant human megakaryocytes growth and development factor on hematopoiesis in normal mice. *Stem Cells* 1996;14:651–660.
56. Janeway CA, Travers P, Walport M, Shlomchik M. *Immunobiology*, 6th ed. New York: Garland Science; 2005.
57. Al-Dabbagh SG, Smith RL. Species differences in oxidative drug metabolism: some basic considerations. *Arch Toxicol Suppl* 1984;7:219–231.
58. Kararli TT. Comparison of the gastrointestinal anatomy, physiology, and biochemistry of humans and commonly used laboratory animals. *Biopharm Drug Dispos* 1995;16(5): 351–380.
59. Lin JH. Species similarities and differences in pharmacokinetics. *Drug Metab Dispos* 1995;23(10):1008–1021.
60. Phillips RM, Bibby MC, Double JA. A critical appraisal of the predictive value of *in vitro* chemosensitivity assays. *J Nat Cancer Inst* 1990;82:1457–1468.
61. Harker LA, Finch CA. Thrombokinetics in man. *J Clin Invest* 1969;48:963–974.
62. Ratajczak J, Machalinski B, Samuel A, Pertusini E, Majka M, Czajka R, Ratajczak MZ. A novel serum free system for cloning human megakaryocytic progenitors (CFU-Meg). The role of thrombopoietin and other cytokines on bone marrow and cord blood CFU-Meg growth under serum free conditions. *Folia Histochem Cytobiol* 1998;36(2) 61–66.
63. Agur Z. Population dynamics in harshly varying environments; evolutionary, ecological and medical aspects. In Hallam TG, Gross LJ, Levin SA, Eds. *Mathematical Ecology*. River Edge, NJ: World Scientific; 1988, pp 440–454.
64. Agur Z. Resonance and anti-resonance in the design of chemotherapeutic protocols. *J Theor Biol* 1998;1:237–245.
65. Vainstein V, Ginosar Y, Shoham M, Ianovski A, Rabinovich A, Kogan Y, Selitser V, Ariad S, Chan S, Agur Z. Clinical validation of a physiologically-based computer model of human granulopoiesis and its use for improving cancer therapy by doxorubicin and granulocyte colony-stimulating factor (G-CSF). *Annual Meeting American Society of Hematology (ASH), Orlando, FL*, 2006.

# REGULATORY REQUIREMENTS FOR INDs/FIH (FIRST IN HUMAN) STUDIES

SHAYNE COX GAD
*Gad Consulting Services, Cary, North Carolina*

**Contents**

1 Introduction
2 Overview of U.S. Regulations
  2.1 Regulations: General Considerations
  2.2 Regulations: Human Pharmaceuticals
  2.3 Regulations: Environmental Impact
  2.4 Regulations: Antibiotics
  2.5 Regulations: Biologics
  2.6 Regulations Versus Law
3 Organizations Regulating Drugs in the United States
4 Process of Pharmaceutical Product Development and Approval
5 Testing Guidelines
  5.1 Toxicity Testing: Traditional Pharmaceuticals
  5.2 General or Systematic Toxicity Assessment
  5.3 Genetic Toxicity Assessment
  5.4 Toxicity Testing: Biotechnology Products
  5.5 Toxicity/Safety Testing: Cellular and Gene Therapy Products
6 Toxicity Testing: Special Cases
  6.1 Oral Contraceptives
  6.2 Life-Threatening Disease (Compassionate Use)
  6.3 Optical Isomers
  6.4 Special Populations: Pediatric and Geriatric Claims
  6.5 Orphan Drugs
  6.6 Botanical Drug Products
7 International Pharmaceutical Regulation and Registration
  7.1 International Conference on Harmonization
  7.2 Other International Considerations
  7.3 Safety Pharmacology

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

8 Conclusion
References and Suggested Reading

## 1 INTRODUCTION

The safety of pharmaceutical agents, medical devices, and food additives is the toxicology issue of the most obvious and longest-standing concern to the public. A common factor among the three is that any risk associated with a lack of safety of these agents is likely to affect a very broad part of the population, with those at risk having little or no option as to undertaking this risk. Modern drugs are essential for life in today's society, yet there is a consistent high level of concern about their safety.

This chapter examines the regulations that establish how the safety of human pharmaceutical products is evaluated and established in the United States and the other major international markets prior to initial trials in human beings harmonization requirements between countries has proceeded markedly over the last fifteen years [1].

## 2 OVERVIEW OF U.S. REGULATIONS

### 2.1 Regulations: General Considerations

The U.S. federal regulations that govern the testing, manufacture, and sale of pharmaceutical agents are covered in Chapter 1, Title 21 of the *Code of Federal Regulations* (21 CFR). These comprise nine 6 in. × 8 in. (printing on both sides of the pages) volumes that stack 8 in. high. This title also covers foods, veterinary products, and cosmetics. As these topics will be discussed elsewhere in this book, here we briefly review those parts of 21 CFR that are applicable to human health products and medicinal devices [2].

Of most interest to a toxicologist working in this arena would be Chapter 1, Subchapter A (parts 1–78), which cover general provisions, organization, and so on. The Good Laboratory Practices (GLPs) are codified in 21 CFR 58.

Table 1 is a checklist of required data to support the development of a drug (in the general case), as well as the expected time to conduct each study through to an audited draft report.

General regulations that apply to drugs are in Subchapter C (Parts 200–299). This covers topics such as labeling, advertising, commercial registration, manufacture, and distribution. Of most interest to a toxicologist would be a section on labeling (Part 201, Subparts A–G, which covers Sections 201.1 through 201.317 of the regulations) as much of the toxicological research on a human prescription drug goes toward supporting a label claim. For example, specific requirements on content and format of labeling for human prescription drugs are covered in Section 201.57. Directions for what should be included under the "Precautions" section of a label are listed in Section 201.57(f). This included Section 201.57(f)(6), which covers categorization of pregnancy risk, and the reliance upon animal reproduction studies in making these categorizations are made quite clear. For example, a drug is given a

**TABLE 1 General Case Safety Evaluation Requirements [3][a]**

| Test Requirement (if Approval Sought as Pharmaceutical Ingredient) | Time[b] |
|---|---|
| *Initial Clinical Trial/IND Requirements* | |
| Regulatory Status as a Pharmaceutical Component | |
| 1. Acute toxicity in rodents | 6 weeks |
| 2. Acute toxicity in nonrodents | 6 weeks |
| 3. Acute parenteral toxicity in rodents | 6 weeks |
| 4. Genotoxicity: mutagenicity (bacterial) | 4 weeks |
| 5. Genotoxicity: *in vitro* clastogenicity | 6 weeks |
| 6. Genotoxicity: *in vivo* | 4 weeks |
| 7. Safety pharmacology: CV-hERG[c] | 4 weeks |
| 8. Safety pharmacology: CV-Purkinje fiber[c] | 4 weeks |
| 9. Safety pharmacology: CV *in vivo* | 6 weeks |
| 10. Safety pharmacology: CNS (FOB/Irwin) | 4 weeks |
| 11. Safety pharmacology: respiratory—rodent | 6 weeks |
| 12. Pivotal/repeat dose in rodents | 3–4 months |
| 13. Pivotal/repeat dose in nonrodents | 3–4 months |
| 14. Pyrogenicity (if parenteral) | 2 weeks |
| 15. Local irritation: ocular | 1 month |
| 16. Local irritation: dermal | 1 month |
| 17. Immunotoxicity[c] | 4 weeks |
| 18. Pivotal/repeat dose in rodents[c] | 9 months |
| 19. Pivotal/repeat dose in nonrodents[c] | 9 months |
| 20. Reproductive/developmental[c] | 6 months |
| 21. Tumorigenicity/carcinogenicity[c] | 30 months |

[a]All studies described must be performed under GLP.
[b]From study start to draft report.
[c]Not required prior to investigational new drug/first-in-human (IND/FIH) studies but recommended.

Pregnancy Category B if "animal reproduction studies have failed to demonstrate a risk to the fetus." The point here is not to give the impression that the law is most concerned with pregnancy risk. Rather, we wish to emphasize that much basic toxicological information must be summarized on the drug label (or package insert). This section of the law is quite detailed as to what information is to be presented and the format of the presentation. Toxicologists working in the pharmaceutical arena should be familiar with this section of the CFR.

## 2.2 Regulations: Human Pharmaceuticals

The regulations specifically applicable to human drugs are covered in Subchapter D, Parts 300–399. The definition of a new drug is covered in Part 310(g): "A new drug substance means any substance that when used in the manufacture, processing or packaging of a drug causes that drug to be a new drug but does not include intermediates used in the synthesis of such substances."

The regulation then goes on to discuss "newness with regard to new formulations, indications, or in combinations." For toxicologists, the meat of the regulations can be found in Section 312 (INDA) and Section 314 (applications for approval to market a new drug or antibiotic drug or NDA). The major focus for a toxicologist working in the pharmaceutical industry is on preparing the correct toxicology

"packages" to be included to "support" these two types of applications. (The exact nature of these packages is covered later.)

In a nutshell, the law requires solid scientific evidence of safety and efficacy before a new drug will be permitted in clinical trials or (later) on the market. The INDA (covered in 21CFR 310) is for permission to proceed with clinical trials on human subjects. Once clinical trials have been completed, the manufacturer or "sponsor" can then proceed to file an NDA (covered in 21 CFR 314) for permission to market the new drug.

As stated in Section 321.21: "A sponsor shall submit an IND if the sponsor intends to conduct a clinical investigation with a new drug . . . [and] shall not begin a clinical investigation until . . . an IND . . . is in effect." Similar procedures are in place in other major countries. In the United Kingdom, for example, a Clinical Trials Certificate (CTC) must be filed or a CTX (clinical trial exemption) obtained before clinical trials may proceed. Clinical trials are divided into three phases, as described in Section 312.21. Phase I trials are initial introductions into healthy volunteers primarily for the purposes of establishing tolerance (side effects), bioavailability, and metabolism. Phase II clinical trials are "controlled studies . . . to evaluate effectiveness of the drug for a particular indication or disease." The secondary objective is to determine common short-term side effects; hence, the subjects are closely monitored. Phase III studies are expanded clinical trials. It is during this phase that the definitive, large-scale, double-blind studies are performed.

The toxicologist's main responsibilities in the IND process are to design, conduct, and interpret appropriate toxicology studies (or "packages") to support the initial IND and then design the appropriate studies necessary to support each additional phase of investigation. Exactly what may constitute appropriate studies are covered elsewhere in this chapter. The toxicologist's second responsibility is to prepare the toxicology summaries for the (clinical) investigator's brochure (described in Section 312.23(a)(8)(ii)). This is an integrated summary of the toxicological effects of the drug in animals and *in vitro*. The U.S. Food and Drug Administration (FDA) has prepared numerous guidance documents covering the content and format of INDs. It is of interest that in the *Guidance for Industry* [4] an in-depth description of the expected contents of the Pharmacology and Toxicology sections was presented. The document contains the following self-explanatory passage: "Therefore, if final, fully quality-assured individual study reports are not available at the time of IND submission, an integrated summary report of toxicological findings based on the unaudited draft toxicologic reports of the completed animal studies may be submitted."

If unfinalized reports are used in an initial IND, the finalized report must be submitted within 120 days of the start of the clinical trial. The sponsor must also prepare a document identifying any differences between the preliminary and final reports, and the impact (if any) on interpretation.

Thus, while the submission of fully audited reports is preferable, the agency does allow for the use of incomplete reports.

Once an IND or CTC/X is opened, the toxicologists may have several additional responsibilities. First, to design, conduct, and report the additional tests necessary to support a new clinical protocol or an amendment to the current clinical protocol (Section 312.20). Second, to bring to the sponsor's attention any finding in an ongoing toxicology study in animals "suggesting a significant risk to human subjects, including any finding of mutagenicity, teratogenicity or carcinogenicity," as described

**TABLE 2 Composition of Standard Investigational New Drug Application (INDA)[a]**

1. IND cover sheets (form FDA-1571)
2. Table of contents
3. Introductory statement
4. General (clinical) investigation plan
5. (Clinical) investigators brochure
6. (Proposed) clinical protocol(s)
7. Chemistry, manufacturing, and control information
8. Pharmacology and toxicology information (includes metabolism and pharmacokinetic assessments done in animals)
9. Previous human experience with the investigational drug
10. Additional information
11. Other relevant information

[a]Complete and thorough reports on all pivotal toxicological studies must be provided with the application.

in 21 CFR 312.32. The sponsor has a legal obligation to report such findings within 10 working days. Third, to prepare a "list of the preclinical studies . . . completed or in progress during the past year" and a summary of the major preclinical findings. The sponsor is required (under Section 312.23) to file an Annual Report (within 60 days of the IND anniversary date) describing the progress of the investigation. INDs are never "approved" in the strict sense of the word. Once filed, an IND can be opened 30 days after submission, unless the FDA informs the sponsor otherwise. The structure of an IND is outlined in Table 2.

If the clinical trials conducted under an IND are successful in demonstrating safety and effectiveness (often established at a Pre-NDA meeting, described in 21 CFR 312.47(b)(2)), the sponsor can then submit an NDA. Unlike an IND, the NDA must be specifically approved by the FDA. The toxicologist's responsibility in the NDA/Marketing Authorization Application (MAA) process is to prepare an integrated summary of all the toxicology and/or safety studies performed and be in a position to present and review the toxicology findings to the FDA or its advisory bodies. The approval process can be exhausting, including many meetings, hearings, and appeals. The ground rules for all of these are described in Part A of the law. For example, all NDAs are reviewed by an "independent" (persons not connected with either the sponsor or the FDA) Scientific Advisory Panel, which will review the findings and make recommendations as to approval. MAAs must be reviewed by and reported on by an expert recognized by the cognizant regulatory authority. Final statutory approval in the United States lies with the Commissioner of the FDA. It is hoped that few additional studies will be requested during the NDA review and approval process. When an NDA is approved, the FDA will send the sponsor an approval letter and will issue a Summary Basis of Approval (SBA)(312.30), which is designed and intended to provide a public record on the FDA's reasoning for approving the NDA while not revealing any proprietary information. The SBA can be obtained through Freedom of Information and can provide insights into the precedents for which types of toxicology studies are used to support specific types of claims.

### 2.3 Regulations: Environmental Impact

Environmental Impact Statements, while once important only for animal drugs, must now accompany all MDAs. This assessment must also be included in the Drug Master File (DMF). The procedures, formats, and requirements are described in 21 CFR 2531. This requirement has grown in response to the National Environmental Policy Act, the heart of which required that federal agencies evaluate every major action that could affect the quality of the environment. In the INDs, this statement can be a relatively short section claiming that relatively small amounts will post little risk to the environment. The EEC has similar requirements for drug entities in Europe, although data requirements are more strenuous. With NDAs, this statement must be more substantial, detailing any manufacturing and/or distribution process that may result in release into the environment. Environmental fate (e.g., photohydrolysis) and toxicity (e.g., fish, daphnia, and algae) studies will be required. While not mammalian toxicology in the tradition of pharmaceutical testing, preparing an environmental impact statement will clearly require toxicological input. The FDA has published a technical bulletin covering the tests it may require [5].

### 2.4 Regulations: Antibiotics

The NDA law (safety and effectiveness) applies to all drugs, but antibiotic drugs were treated differently until the passage of FDAMA in 1997. Antibiotic drugs had been treated differently by the FDA since the development of penicillin revolutionized medicine during World War II. The laws applicable to antibiotic drugs were covered in 21 CFR 430 and 431. Antibiotics such as penicillin or doxorubicin are drugs derived (in whole or in part) from natural sources (such as molds or plants), which have cytotoxic or cytostatic properties. They were treated differently from other drugs as the applicable laws required a batch-to-batch certification process. Originally passed into law in 1945 specifically for penicillin, this certification process was expanded by the 1962 amendment (under Section 507 of the Food, Drug, and Cosmetic Act) to require certification of all antibiotic drugs, meaning that the FDA would assay each lot of antibiotic for purity, potency, and safety. The actual regulations were covered in 21 CFR Subchapter D, Parts 430–460 (over 600 pages), which describes the standards and methods used for certification for all approved antibiotics. Section 507 was repealed by FDAMA (Section 125). As a result of the repeal of Section 507, the FDA is no longer required to publish antibiotic monographs. In addition, the testing, filing, and reviewing of antibiotic applications are now handled under Section 505 of the Act like any other new therapeutic agent. The FDA has published a guidance document to which the reader is referred for more details [6, 7].

### 2.5 Regulations: Biologics

Biological products are covered in Subchapter F, Parts 600–680. As described in 21 CFR 600.3(h), "biological product means any virus, therapeutic serum, toxin, antitoxin or analogous product applicable to the prevention, treatment or cure of diseases or injuries of man." In other words, these are vaccines and other protein

products derived from animal sources. Clearly, the toxicological concerns with such products are vastly different from those involved with low molecular weight synthetic molecules. There is little rational basis, for example, for conducting a one year, repeated dose toxicity study with a vaccine or a human blood product. The FDA definition for safety with regard to these products is found in 21 CFR 603.1(p): "Relative freedom from harmful effect to persons affected, directly or indirectly, by a product when prudently administered." Such safety consideration has more to do with purity, sterility, and adherence to good manufacturing standards than with the toxicity of the therapeutic molecule itself. The testing required to show safety is stated in the licensing procedures 21 CFR 601.25(d)(1): "Proof of safety shall consist of adequate test methods reasonably applicable to show the biological product is safe under the prescribed conditions." Once a license is granted, each batch or lot of biological product must be tested for safety, and the methods of doing so are written into the law. A general test for safety (i.e., required in addition to other safety tests) is prescribed using guinea pigs as described in Section 610.11. Additional tests are often applied to specific products. For example, 21 CFR 630.35 describes the safety tests required for measles vaccines, which includes tests in mice and *in vitro* assays with tissue culture. Many new therapeutic entities produced by biotechnology are seeking approval as biologics with the results being FDA approval of a Product License Application (PLA). Table 3 presents general guidance for the basis of deciding if an individual entity falls under the Center for Drug

**TABLE 3 Product Class Review Responsibilities**

| |
|---|
| *Center for Drug Evaluation and Review* |
| Natural products purified from plant or mineral sources |
| Products produced from solid tissue sources (excluding procoagulants, venoms, blood products, etc.) |
| Antibiotics, regardless of method of manufacture |
| Certain substances produced by fermentation |
| Disaccharidase inhibitors |
| HMG-CoA inhibitors |
| Synthetic chemicals |
| Traditional chemical synthesis |
| Synthesized mononuclear or polynuclear products including antisense chemicals |
| Hormone products |
| *Center for Biologics Evaluation and Review* |
| Vaccines, regardless of manufacturing method |
| *In vivo* diagnostic allergenic products |
| Human blood products |
| Protein, peptide, and/or carbohydrate products produced by cell culture (other than antibiotics and hormones) |
| Immunoglobulin products |
| Products containing intact cells or microorganisms |
| Proteins secreted into fluids by transgenic animals |
| Animal venoms |
| Synthetic allergens |
| Blood banking and infusion adjuncts |

Evaluation and Research (CDER) or the Center for Biologic Evaluation and Research (CBER) authority for review.

The International Conference on Harmonization has published its document *S6 Preclincial Safety Evaluation of Biotechnology-Derived Pharmaceuticals*. The FDA (both the Center for Drug Evaluation and Research, and the Center for Biologics Evaluation and Research jointly) has published the document as a *Guidance for Industry* [8, 9].

A current list of regulatory documents available by email (including the most recent PTCs, or points to consider) can be found at DOC_LIST@A1.FDA.GOV.

### 2.6 Regulations Versus Law

A note of caution must be inserted here. The law (the document passed by Congress) and the regulations (the documents written by the regulatory authorities to enforce the laws) are separate documents. The sections in the law do not necessarily have numerical correspondence. For example, the regulations on the NDA process is described in 21 CFR 312, but the law describing the requirement for an NDA process is in Section 505 of the FDCA. Because the regulations rather than the laws themselves have a greater impact on toxicological practice, greater emphasis is placed on regulation in this chapter. For a complete review of FDA law, the reader is referred to the monograph by Food and Drug Law Institute [2].

Laws authorize the activities and responsibilities of the various federal agencies. All proposed laws before the U.S. Congress are referred to committees for review and approval.

## 3 ORGANIZATIONS REGULATING DRUGS IN THE UNITED STATES

The agency formally charged with overseeing the safety of drugs in the United States is the FDA. It is headed by a commissioner who reports to the Secretary of the Department of Health and Human Services (DHHS) and has a tremendous range of responsibilities. Drugs are overseen primarily by the CDER (although some therapeutic or health care entities are considered biologics and are overseen by the corresponding CBER). Figure 1 presents the organization of CDER. The organization of CBER is shown in Fig. 2.

Most of the regulatory interactions of toxicologists are with the two offices of Drug Evaluation, which have under them a set of groups focused on areas of therapeutic claim (cardiorenal, neuropharmacological, gastrointestinal and coagulation, oncology and pulmonary, metabolism and endocrine, antiinfective and antiviral). Within each of these are chemists, pharmacologists/toxicologists, statisticians, and clinicians. When an INDA is submitted to the offices of Drug Evaluation, it is assigned to one of the therapeutic groups based on its area of therapeutic claim. Generally, it will remain with that group throughout its regulatory approval "life." INDs, when allowed, grant investigators the ability to go forward into clinical (human) trials with their drug candidate in a predefined manner, advancing through various steps of evaluation in human (and in additional preclinical or animal studies)

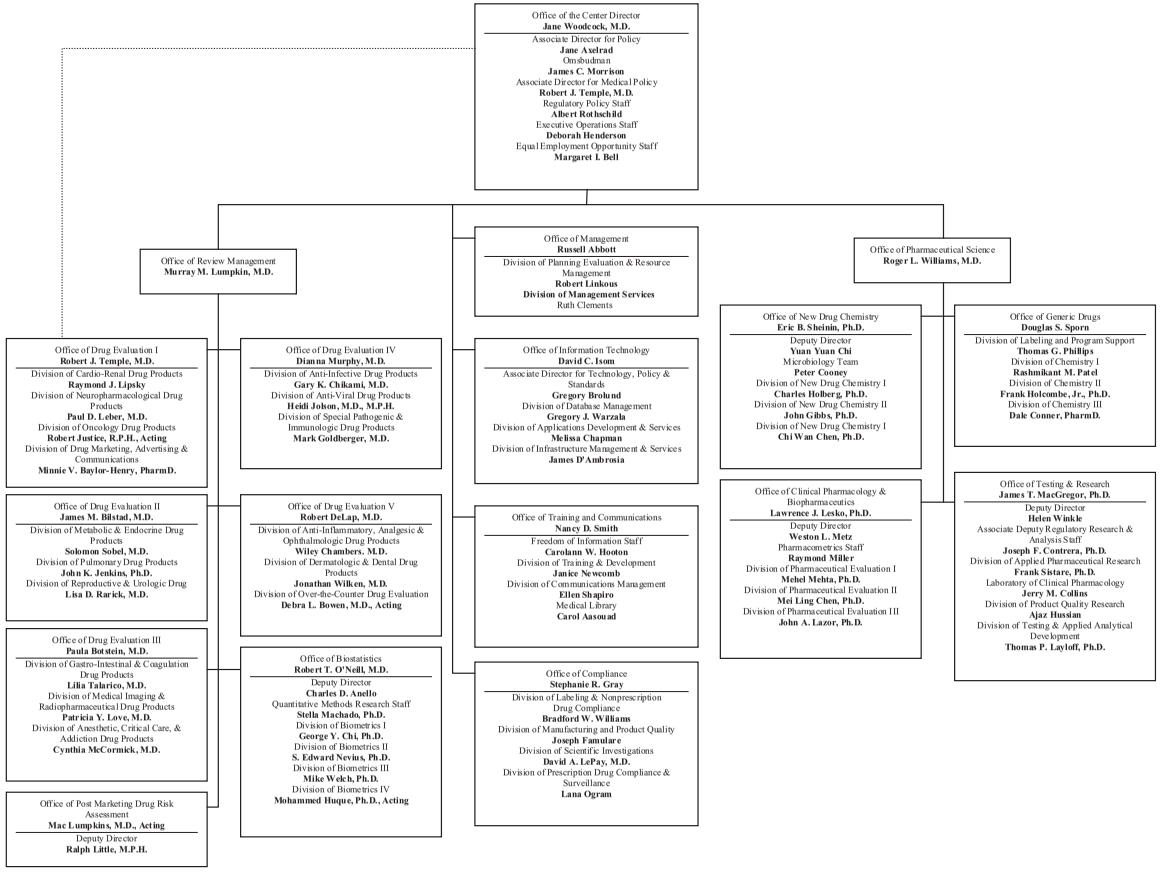


**FIGURE 1** Organizational chart for CDER.

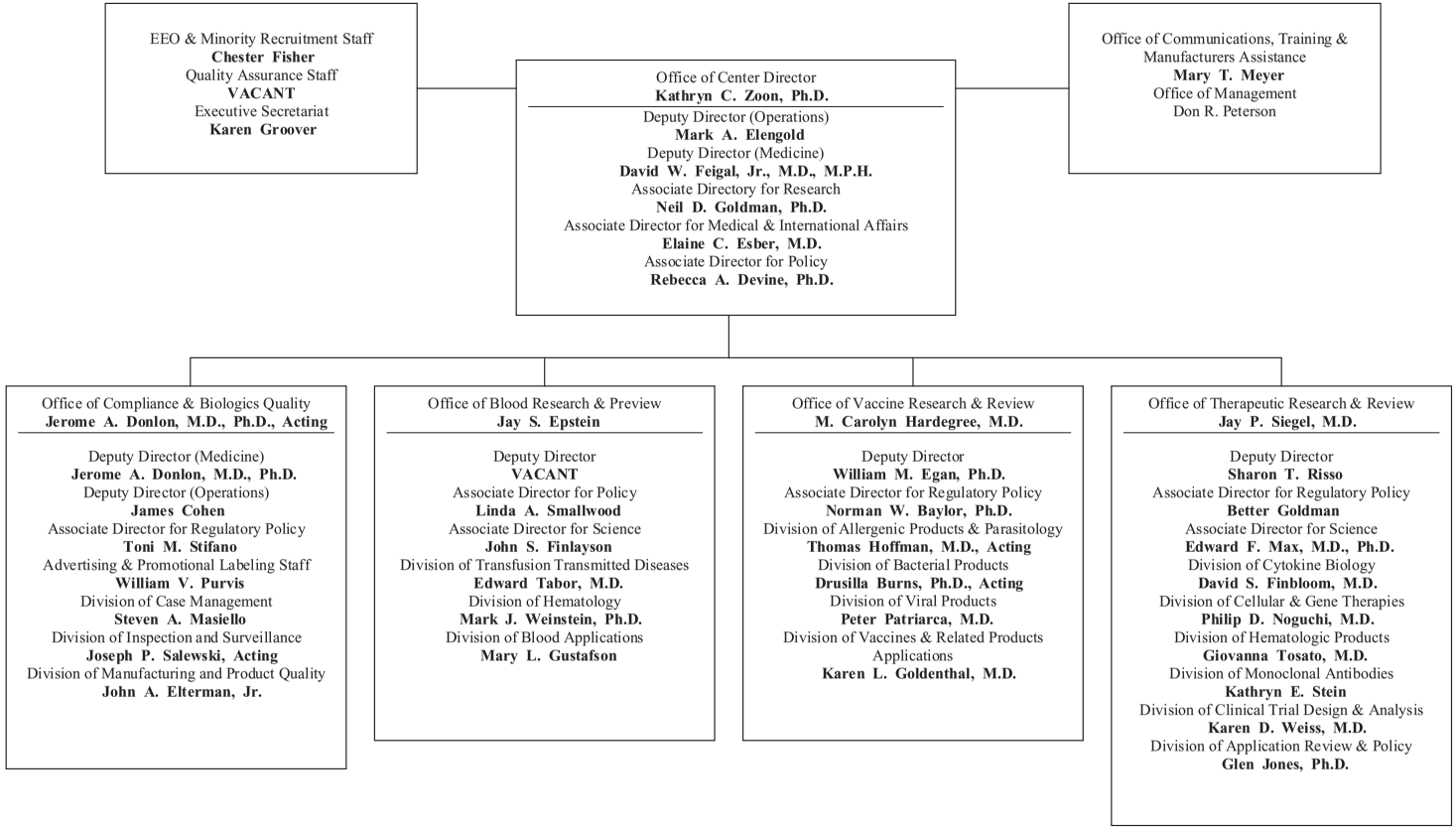


**FIGURE 2** Organizational chart for CBER.

until an NDA can be supported, developed, and submitted. Likewise for biological products, the PLA or other applications (INDA, IND) are handled by the offices of Biological Products Review of the CBER.

For drugs, there is at least one nongovernmental body that must review and approve various aspects—the USP (United States Pharmacopeia, established in 1820)—which maintains (and revises) the compendia of the same name, as well as the National Formulary, which sets drug composition standards [10]. This volume sets forth standards for purity of products in which residues may be present and tests for determining various characteristics of drugs, devices, and biologics.

## 4 PROCESS OF PHARMACEUTICAL PRODUCT DEVELOPMENT AND APPROVAL

Except for a very few special cases (treatments for life-threatening diseases such as cancer or AIDS), the safety assessment of new drugs as mandated by regulations proceeds in a rather fixed manner. The IND is filed to support this clinical testing. An initial set of studies (typically, studies of appropriate length by the route intended for humans are performed in both a rodent (typically rat) and a nonrodent (usually a dog or a primate)) are required to support Phase I clinical testing. Such Phase I testing is intended to evaluate the safety ("tolerance" in clinical subjects), pharmacokinetics, and general biological effects of a new drug and is conducted in normal volunteers (almost always males).

Successful completion of Phase I testing allows, with the approval of the FDA, progression into Phase II clinical testing. Here, selected patients are enrolled to evaluate therapeutic efficacy, dose ranging, and more details about the pharmacokinetics and metabolism. Longer-term systemic toxicity studies must be in conformity with the guidelines that are presented in the next section. Once a sufficient understanding of the actions, therapeutic dose response, and potential risk-to-benefit ratio of the drug is in hand (once again, with FDA approval), trials move into Phase III testing.

Phase III tests are large, long, and expensive. They are conducted using large samples of selected patients and are intended to produce proof of safety and efficacy of the drug. Two studies providing statistically significant proof of the claimed therapeutic benefit must be provided. All the resulting data from preclinical and clinical animal studies are organized in a specified format in the form of an NDA, which is submitted to the FDA.

By the time that Phase III testing is completed, some additional preclinical safety tests must also generally be in hand. These include the three separate reproductive and developmental toxicity studies (Segments I and III in the rat, and Segment II in the rat and rabbit) and carcinogenicity studies in both rats and mice (unless the period of therapeutic usage is intended to be very short). Some assessment of genetic toxicity will also be expected.

The ultimate product of the pharmaceutical toxicologist will thus generally be the toxicology summaries of the IND and NDA (or PLA). For medical devices, the equivalents are the IDE and Product Development Notification (PDN). Data required to support each of these documents is specified in a series of guidelines, as discussed later.

Acceptance of these applications is contingent not only upon adherence to guidelines and good science, but also to GLPs.

## 5 TESTING GUIDELINES

### 5.1 Toxicity Testing: Traditional Pharmaceuticals

Although the 1938 Act required safety assessment studies, no consistent guidelines were available [11]. Guidelines were first proposed in 1949 and published in the *Food, Drug and Cosmetic Law Journal* that year [12]. Following several revisions, these guidelines were issued as *The Appraisal Handbook* in 1959. While never formally called a guideline, it set the standard for preclinical toxicity test design for several years. The current basic guidelines for testing required for safety assessment in support of the phases of clinical development of drugs were first outlined by Goldenthal [13, 14] and later incorporated into a 1971 FDA publication entitled *FDA Introduction to Total Drug Quality* [15].

### 5.2 General or Systematic Toxicity Assessment

Table 4 presents an overview of the current FDA toxicity testing guidelines for human drugs. Table 5 presents the parallel ICH guidelines [16], which are now largely supplanting the FDA guidelines. They are misleading in their apparent simplicity, however. First, each of the systemic toxicity studies in these guidelines must be designed and executed in a satisfactory manner. Sufficient animals must be used to have confidence in finding and characterizing any adverse drug actions that may be present. In practice, as the duration of the study increases, small doses are administered and larger numbers of animals must be employed per group. These two features—dosage level and group size—are critical to study designs. Table 6 presents general guidance on the number of animals to be used in systemic studies.

Note that for novel chemical structures, the FDA may still require 12 month nonrodent toxicity studies.

The protocols discussed thus far have focused on general or systemic toxicity assessment. The agency and, indeed, the lay public have a special set of concerns with reproductive toxicity, fetal/embryo toxicity, and developmental toxicity (also called *teratogenicity*). Collectively, these concerns often go by the acronyms *DART* (developmental and reproductive toxicity) or *RTF* (reproduction, teratogenicity, fertility). Segment II studies are more designed to detect developmental toxicity. Only pregnant females are dosed during the critical period of organogenesis. Generally, the first protocol DART test (exclusive of range-finding studies) is a Segment I study of rats in fertility and general reproductive performance. This is generally done while the drug is in Phase II clinical trials. Alternatively, many companies are now performing the Segment II teratology study in rats before the Segment I study because the former is less time and resource intensive. One or both should be completed before including women of child-bearing potential in clinical trials. The FDA requires teratogenicity testing in two species—a rodent (rat or mouse) and the rabbit. The use of the rabbit was instituted as a result of the finding that thalidomide was a positive teratogen in the rabbit but not in the rat. On occasion, when a test

**TABLE 4 Synopsis of General Guidelines for Animal Toxicity Studies for Drugs**

| Category | Duration of Human Administration | Clinical Phase | Subacute or Chronic Toxicity | Special Studies |
|---|---|---|---|---|
| Oral or parenteral | Several days | I, II, III, NDA | Two species; 2 weeks | For parenterally administered drugs |
| | Up to 2 weeks | IIIIII, NDA | Two species; 4 weeksTwo species; up to 4 weeks<br>Two species; up to 3 months | |
| | Up to 3 months | I, IIIIINDA | Two species; 4 weeksTwo species; 3 months<br>Two species; up to 6 months | Compatibility with blood where applicable |
| | Six months to unlimited | I, IIIIINDA | Two species; 3 months<br>Two species; 6 months or longer<br>Two species; 9 months (nonrodent) and 6 months (rodent)<br>+Two rodent species for CA; 18 months (mouse)<br>May be replaced with an allowable transgenic mouse study; 24 months (rat) | |
| Inhalation (general anesthetics) | | I, II, III, NDA | Four species; 5 days (3 h/day) | |
| Dermal | Single application | I | One species; single 24 hour exposure followed by 2 week observation | Sensitization |
| | Single or short-term application | II | One species; 20 day repeated exposure (intact and abraded skin) | |
| | Short-term application | III | As above | |
| | Unlimited application | NDA | As above, but intact skin study extended up to 6 months | |
| Ophthalmic | Single application | I | | Eye irritation tests with graded doses |
| | Multiple applications | I, II, IIINDA | One species; 3 weeks daily applications, as in clinical use<br>One species; duration commensurate with period of drug administration | |
| Vaginal or rectal | Single application<br>Multiple applications | II, II, III, NDA | Two species; duration and number of applications determined by proposed use | Local and systematic toxicity after vaginal or rectal application in two species |
| Drug combinations | (4) | III, III, NDA | Two species; up to 3 months | Lethality by appropriate route, compared to components run concurrently in one species |

**TABLE 5 Duration of Repeated Dose Toxicity Studies to Support Clinical Trials and Marketing[a]**

| Duration of Clinical Trials | Minimum Duration of Repeated Dose Toxicity Studies[b] | | Duration of Clinical Trials | Minimum Duration of Repeated Dose Toxicity Studies[c] | |
|---|---|---|---|---|---|
| | Rodents | Nonrodents | | Rodents | Nonrodents |
| Single dose | 2 weeks[d] | 2 weeks | Up to 2 weeks | 1 month | 1 month |
| Up to 2 weeks | 2 weeks[d] | 2 weeks | Up to 1 month | 3 months | 3 months |
| Up to 1 month | 1 month | 1 month | Up to 3 months | 6 months | 3 months |
| Up to 6 months | 6 months | 6 months[b] | >3 months | 6 months | Chronic[d] |
| >6 months | 6 months | Chronic[b] | | | |

[a]In Japan, if there are no Phase II clinical trials of equivalent duration to the planned Phase III trials, conduct of longer duration toxicity studies is recommended.[b]Data from 6 months of administration in nonrodents should be available before the initiation of clinical trials longer than 3 months. Alternatively, if applicable, data from a 9 month nonrodent study should be available before the treatment duration exceeds that which is supported by the available toxicity studies.
[c]The table also reflects the marketing recommendations in the three regions except that a chronic nonrodent study is recommended for clinical use >1 month.
[d]In the United States, as an alternative to 2 week studies, single-dose toxicity studies with extended examinations can support single-dose human trials.

**TABLE 6 Numbers of Animals per Dosage Group in Systemic Toxicity Studies**

| Study Duration (per sex) | Rodents (per sex) | Nonrodents |
|---|---|---|
| 2–4 weeks | 5 | 3 |
| 13 weeks | 20[a] | 6 |
| 26 weeks | 30 | 8 |
| Chronic | 50 | 10 |
| Carcinogenicity | 60[b] | Applies only to contraceptives |
| Bioassays | | Applies only to contraceptives |

[a]Starting with 13 week studies, one should consider adding animals (particularly to the high dose) to allow evaluation of reversal of effects.
[b]In recent years there have been decreasing levels of survival in rats on 2 year studies. What is required is that at least 20–25 animals/sex/group survive at the end of the study. Accordingly, practice is beginning to use 70 or 75 animals per sex, per group.

article is not compatible with the rabbit, teratogenicity data in the mouse may be substituted. There are also some specific classes of therapeutics (e.g., the quinalone antibioticsexample) where Segment II studies in primates are effectively required prior to product approval. Both should be completed before entering Phase III clinical trials. The most complicated of the DART protocols—Segment III—is generally commenced during Phase III trials and should be part of the NDA. There are differences in the different national guidelines (as discussed later with international considerations) regarding the conduct of these studies. The large multinational drug companies try to design their protocols to be in compliance with as many of the

guidelines as possible to avoid duplication of testing while allowing the broadest possible approval and marketing of therapeutics.

### 5.3 Genetic Toxicity Assessment

Genetic toxicity testing generally focuses on the potential of a new drug to cause mutations (in single-cell systems) or other forms of genetic damage. The tests, generally short in duration, often rely on *in vitro* systems and generally have a single endpoint of effect (point mutations, chromosomal damage, etc.). It is of interest that the FDA has no standard or statutory requirement for genetic toxicity testing but generally expects to see at least some such tests performed and will ask for them if the issue is not addressed. If one performs such a study, any data collected, of course, must be sent to the FDA as part of any INDA, PLA, or NDA. These studies have yet to gain favor with the FDA (or other national regulatory agencies) as substitutes for *in vivo* carcinogenicity testing. However, even with completed negative carcinogenicity tests, at least some genetic toxicity assays are generally required. Generally, pharmaceuticals in the United States are evaluated for mutagenic potential (e.g., the Ames' Assay) or for chromosomal damage (e.g., the *In Vivo* Mouse Micronucleus Test). In general, in the United States, pharmaceutical companies apply genetic toxicity testing in the following fashion:

- *As a Screen.* An agent that is positive in one or more genetic toxicity tests may be more likely than one that is negative to be carcinogenic and therefore may not warrant further development.
- *As an Adjunct.* An agent that is negative in carcinogenicity testing in two species and also negative in a genetic toxicity battery is more likely than not to be noncarcinogenic in human beings.
- *To Provide Mechanistic Insight.* For example, if an agent is negative in a wide range of genetic toxicity screens but still produces tumors in animals, then one could hypothesize that an epigenetic mechanism was involved.

While not officially required, the FDA does have the authority to request, on a case-by-case basis, specific tests it feels may be necessary to address a point of concern. A genetic toxicity test could be part of such a request. In general, therefore, companies deal with genetic toxicity (after "screening") on a case-by-case basis, dictated by good science. If more than a single administration is intended, common practice is to perform the tests prior to submitting an IND.

### 5.4 Toxicity Testing: Biotechnology Products

As mentioned, the regulation of traditional pharmaceuticals (small molecules such as aspirin or digitalis) and biologicals (proteins such as vaccines and antitoxins derived from animal sources) have very different histories. See Section 2.5. Until 1972, the NIH (or its forerunning agency, the Hygienic Laboratory of the Department of the Treasury) was charged with the responsibilities of administering the Virus Act of 1902. With the passage of the Food and Drug Laws of 1906, 1938, and 1962, there was recurring debate about whether these laws applied or should apply

to biologicals. This debate was resolved when the authority for the regulation of biologics was transferred to the FDA's new Bureau of Biologics (now the CBER) in 1972. Since then, there appears to have been little difference in the matter of regulation for biologics and pharmaceuticals. The FDA essentially regulates biologics as described under the 1902 act, but then uses the rule-making authority granted under the Food and Drug Act to "fill in the gaps."

The Bureau of Biologics was once a relatively "sleepy" agency, primarily concerned with the regulation of human blood products and vaccines used for mass immunization programs. The authors of the 1902 law could hardly have foreseen the explosion in biotechnology that occurred in the 1980s. New technology created a welter of new biological products, such as recombinant-DNA-produced proteins (e.g., tissue plasminogen activator), biological response modifiers (cytokinins and colony-stimulating factors), monoclonal antibodies, antisense oligonucleotides, and self-directed vaccines (raising an immune response to self-proteins such as gastrin for therapeutic reasons). The new products raised a variety of new questions on the appropriateness of traditional methods of evaluating drug toxicity that generated several points-to-consider documents [17–19]. For the sake of brevity, this discussion focuses on the recombinant DNA proteins. Some safety issues have been raised over the years [20]:

- The appropriateness of testing a human-specific peptide hormone in nonhuman species.
- The potential that the peptide could break down due to nonspecific metabolism, resulting in products that had no therapeutic value or even a toxic fragment.
- The potential sequelae to an immune response (formation of neutralizing antibodies, provoking an autoimmune or a hypersensitivity response), and pathology due to immune precipitation and so on.
- The presence of contamination with oncogenic virus DNA (depending on whether a bacterial or mammalian system was used on the synthesizing agent) or endotoxins.
- The difficulty interpreting the scientific relevance of response to supraphysiological systemic doses of potent biological response modifiers.

The intervening last few years have shown that some of these concerns were more relevant than others [21]. The "toxic peptide fragment" concern, for example, has been shown to be without merit. The presence of potentially oncogenic virus DNA and endotoxins is a quality assurance concern and is not truly a toxicological problem. Regardless of the type of synthetic pathway, all proteins must be synthesized in compliance with Good Manufacturing Practices. Products must be as pure as possible, not only free of rDNA but also free of other types of cell debris (endotoxin). Batch-to-batch consistency with regard to molecular structure must also be demonstrated using appropriate methods (e.g., amino acid). The regulatory thinking and experience over the last fifteen years has come together in the document *S6 Preclincial Safety Evaluation of Biotechnology-Derived Pharmaceuticals* prepared by the International Conference on Harmonization. The FDA (both the Center for Drug Evaluation and Research, and the Center for Biologics Evaluation

and Research jointly) has published the document as a *Guidance for Industry* [5, 6]. The document intended to provide basic guidance for the preclinical evaluation of biotechnology-derived products, including proteins and peptides, produced by cell culture using rDNA technology, but did not cover antibiotics, allergenic extracts, heparin, vitamins, cellular drug products, vaccines, or other products regulated as biologics. Items covered are summarized as follows [22]:

*Test Article Specifications.* In general, the product that is used in the definitive pharmacology and toxicology studies should be comparable to the product proposed for the initial clinical studies.

*Animal Species/Model Selection.* Safety evaluation should include the use of relevant species, in which the test article is pharmacologically active due, for example, to the expression of the appropriate receptor molecule. These can be screened with *in vitro* receptor binding assays. Safety evaluation should normally include two appropriate species, if possible and/or feasible. The potential utility of gene knockout and/or transgenic animals in safety assessment is discussed.

*Group Size.* No specific numbers are given, but it does state that a small sample size may lead to failure to observe toxic events.

*Administration.* The route and frequency should be as close as possible to that proposed for clinical use. Other routes can be used when scientifically warranted.

*Immunogenicity.* It has also been clearly demonstrated in the testing of rDNA protein products that animals will develop antibodies to foreign proteins. This response has been shown to neutralize (rapidly remove from circulation) the protein, but no pathological conditions have been shown to occur as a sequelae to the immune response. Bear in mind, however, that interleukins have powerful effects on immune response, but these are due to their physiological activity and not to an antigen–antibody response. The first has to do with "neutralizing antibodies." That is, is the immune response so great that the test article is being removed from circulation as fast as it is being added? If this is the case, does long-term testing of such a chemical make sense? In many cases, it does not. The safety testing of any large molecule should include the appropriate assays for determining whether the test system has developed a neutralizing antibody response. Depending on the species, route of administration, intended therapeutic use, and development of neutralizing antibodies (which generally takes about 2 weeks), it is rare for a toxicity test on an rDNA protein to be longer than 4 weeks duration. However, if the course of therapy in humans is to be longer than 2 weeks, formation of neutralizing antibodies must be demonstrated or longer-term testing performed. The second antigen–antibody formation concern is that a hypersensitivity response will be elicited. Traditional preclinical safety assays are generally adequate to guard against this if they are 2 weeks or longer in duration and the relevant endpoints are evaluated.

*Safety Pharmacology.* It is important to investigate the potential for unwanted pharmacological activity in appropriate animal models and to incorporate monitoring for these activities in the toxicity studies.

*Exposure Assessment.* Single- and multiple-dose pharmacokinetics, toxicokinetics, and tissue distribution studies in relevant species are useful. Proteins are not given orally; demonstrating absorption and mass balance is not typically a primary consideration. Rather, this segment of the test should be designed to determine half-life (and other appropriate pharmacokinetic descriptor parameters), the plasma concentration associated with biological effects, and potential changes due to the development of neutralizing antibodies.

*Reproductive Performance and Developmental Toxicity Studies.* These will be dictated by the product, clinical indication, and intended patient population.

*Genotoxicity Studies.* The S6 document states that the battery of genotoxicity studies routinely conducted for traditional pharmaceuticals are not appropriate for biotechnology-derived pharmaceuticals. In contrast to small molecules, genotoxicity testing with a battery of *in vitro* and *in vivo* techniques of protein molecules has not become common U.S. industry practice. Such tests are not formally required by the FDA but, if performed, have to be reported. They are required by European and Japanese regulatory authorities. This has sparked a debate as to whether or not genotoxicity testing is necessary or appropriate for rDNA protein molecules. It is the author's opinion that such testing is, scientifically, of little value. First, large protein molecules will not easily penetrate the cell wall of bacteria or yeast, and (depending on size, charge, lipophilicity, etc.) penetration across the plasma lemma of mammalian cells will be highly variable. Second, if one considers the well-established mechanism(s) of genotoxicity of small molecules, it is difficult to conceive of how a protein can act in the same fashion. For example, proteins will not be metabolized to be electrophilic active intermediates that will crosslink guanine residues. In general, therefore, genotoxicity testing with rDNA proteins is a waste of resources. It is conceivable, however, that some proteins, because of their biological mechanism of action, may stimulate the proliferation of transformed cells. For example, it is a feasible hypothesis that a colony-stimulating factor could stimulate the proliferation of leukemic cells (it should be emphasized that this is a hypothetical situation, presented here for illustrative purposes). Again, this is a question of a specific pharmacological property, and such considerations should be tested on a case-by-case basis.

*Carcinogenicity Studies.* These are generally inappropriate for biotechnology-derived pharmaceuticals; however, some products may have the potential to support or induce proliferation of transformed cells . . . possibly leading to neoplasia. When this concern is present, further studies in relevant animal models may be needed.

These items are covered in greater detail in the S6 guidance document and in a review by Ryffel [23].

So, given the above discussion, what should the toxicology testing package of a typical rDNA protein resemble? Based on the products that have successfully wound their way through the regulatory process, the following generalizations can be drawn.

- The safety tests look remarkably similar to those for traditional tests. Most have been done on three species: the rat, the dog, or the monkey. The big difference

has to do with the length of the test. It is rare for a safety test on a protein to be more than 13 weeks long.

- The dosing regimens can be quite variable and at times very technique intensive. These chemicals are almost always administered by a parenteral route—normally intravenously or subcutaneously. Dosing regimens have run the range from once every 2 weeks for an antihormone "vaccine" to continuous infusion for a short-lived protein.
- As reviewed by Ryffel [24], most side effects in humans of a rDNA therapy may be predicted by data from experimental toxicology studies, but there are exceptions. IL-6, for example, induced a sustained increase in blood platelets and acute phase proteins, with no increase in body temperature. In human trials, however, there were increases in temperature.
- The S6 document also mentions monoclonal antibody products. Indeed, many of the considerations for rDNA products are also applicable to monoclonal antibodies (including hybridized antibodies). With monoclonal antibodies, there is the additional concern of crossreactivity with nontarget molecules.

As mentioned, the rapid development in the biotechnology industry has created some confusion as to what arm of the FDA is responsible for such products. In October 1992, the two major reviewing groups, CBER and CDER, reached a series of agreements to explain and organize the FDA's position on products that did not easily fall into its traditional classification schemes. CDER will continue to have responsibility for traditional chemically synthesized molecules as well as those purified from mineral or plant sources (except allergenics), antibiotics, hormones (including insulin and growth hormone), most fungal or bacterial products (disaccharidase inhibitors), and most products from animal or solid human tissue sources. CBER will have responsibility for products subject to licensure (BLA), including all vaccines, human blood or blood-derived products (as well as drugs used for blood banking and transfusion), immunoglobulin products, products containing intact cells, fungi, viruses, proteins produced by cell culture or transgenic animals, and synthetic allergenic products. This situation was further simplified by the introduction of the concept of *well-characterized biologics*. When introduced during the debate on FDA reform in 1996, the proposed section of S-1447 stated that "biological products that the secretary determines to be well-characterized shall be regulated solely under the Federal Food, Drug and Cosmetic Act." Under this concept, highly purified, well-characterized therapeutic rDNA proteins would be regulated by CDER, regardless of therapeutic target [25].

### 5.5 Toxicity/Safety Testing: Cellular and Gene Therapy Products

Human clinical trials of cellular and gene therapies involve administration to patients of materials considered investigational biological, drug, or device products. Somatic cell therapy refers to the administration to humans of autologous, allogenic, or xenogenic cells, which have been manipulated or processed *ex vivo*. Gene therapy refers to the introduction into the human body of genes or cells containing genes foreign to the body for the purposes of prevention, treatment, diagnosing, or curing disease.

Sponsors of cellular or gene therapy clinical trials must file an Investigational New Drug Application (INDA) or in certain cases an Investigational Device Exemption (IDE) with the FDA before initiation of studies in humans. It is the responsibility of the Center of Biologics Evaluation and Research (CBER) to review the application and determine if the submitted data and the investigational product meet applicable standards. The critical parameters of identity, purity, potency, stability, consistency, safety, and efficacy relevant to biological products are also relevant to cellular and gene therapy products.

In 1991, the FDA first published *Points to Consider* in *Human Somatic Cell and Gene Therapy* [18]. At this time virtually all gene therapies were retroviral and were prepared as *ex vivo* somatic cell therapies. This was subsequently reviewed by Kessler et al. [26]. While the data for certain categories of information such as the data regarding the molecular biology were defined in previous guidance documents relating to recombinant DNA products, the standards for preclinical and clinical development were less well defined. Over the past five years, the field has advanced to include not only new vectors but also novel routes of administration. The *Points to Consider in Human Somatic Cell Therapy and Gene Therapy* [27] has thus been recently amended to reflect both the advancements in product development and more importantly the accumulation of safety information over the past five years.

FDA regulations state that the sponsor must submit, in the IND, adequate information about pharmacological and toxicological studies of the drug including laboratory animals or *in vitro* studies on the basis of which the sponsor has considered that it is reasonably safe to conduct the proposed clinical investigation. For cellular and gene therapies, designing and conducting relevant preclinical safety testing has been a challenge to both the FDA and the sponsor. For genes delivered using viral vectors, the safety of the vector system per se must be considered and evaluated.

The preclinical knowledge base is initially developed by designing studies to answer fundamental questions. The development of this knowledge base is generally applicable to most pharmaceuticals, as well as biopharmaceuticals, and includes data to support (1) the relationship of the dose to the biological activity, (2) the relationship of the dose to the toxicity, (3) the effect of route and/or schedule on activity or toxicity, and (4) identification of the potential risks for subsequent clinical studies. These questions are considered in the context of indication and/or disease state. In addition, there are often unique concerns relative to the specific category or product class.

For cellular therapies, safety concerns may include development of a database from studies specifically designed to answer questions relating to growth factor dependence, tumorigenicity, local and systemic toxicity, and effects on host immune responses including immune activation and altered susceptibility to disease. For viral-mediated gene therapies, specific questions may relate to the potential for overexpression of the transduced gene, transduction of normal cells/tissues, genetic transfer to germ cells and subsequent alterations to the genome, recombination/rescue with endogenous virus, reconstitutions of replication competence, potential for insertional mutagenesis/malignant transformation, altered susceptibility to disease, and/or potential risk(s) to the environment.

To date, cellular and gene therapy products submitted to the FDA have included clinical studies indicated for bone marrow marking, cancer, cystic fibrosis, AIDS, and

inborn errors of metabolism and infectious diseases. Of the current active INDs, approximately 78% have been sponsored by individual investigators or academic institutions and 22% have also been industry sponsored. In addition to the variety of clinical indications the cell types have also been varied. Examples include tumor infiltrating lymphocytes (TILs) and lymphocyte activated killer (LAK) cells, selected cells from bone marrow, and peripheral blood lymphocytes (e.g., stem cells, myoblasts, tumor cells), and encapsulated cells (e.g., islet cells and adrenal chromaffin cells).

***Cellular Therapies*** Since 1984 the CBER has reviewed close to 300 somatic cell therapy protocols. Examples of the specific categories include manipulation, selection, mobilization, tumor vaccines, and other.

*Manipulation*—autologous, allogenic, or xenogenic cells that have been expanded, propagated, manipulated, or had their biological characteristics altered *ex vivo* (e.g., TIL or LAK cells; islet cells housed in a membrane).

*Selection*—products designed for positive or negative selection if autologous or allogenic cells are intended for therapy (e.g., purging of tumor from bone marrow, selection of CD34+ cells).

*Mobilization*—*in vivo* mobilization of autologous stem cells intended for transplantation.

*Tumor vaccines*—autologous or allogenic tumor cells that are administered as vaccine (e.g., tumor cell lines, tumor cell lysates, primary explant). This group also includes autologous antigen presenting cells pulsed with tumor-specific peptides or tumor cell lysates.

*Other*—autologous, allogenic, and xenogenic cells that do not specifically fit the above; this group includes cellular therapies such as extracorporeal liver assist devices.

***Gene Therapies*** The types of vectors that have been used, or proposed, for gene transduction include retrovirus, adenovirus, adeno-associated viruses, other viruses (e.g., herpes, vaccinia), and plasmid DNA. Methods for gene introduction include *ex vivo* replacement, drug delivery, and marker studies and others and *in vivo*, viral vectors, plasmid vectors, and vector producer cells.

*Ex Vivo*

*Replacement*—cells transduced with a vector expressing a normal gene in order to correct or replace the function of a defective gene.

*Drug delivery*—cells transduced with a vector expressing a gene encoding a therapeutic molecule that can be novel or native to the host.

*Marker studies*—cells (e.g., bone marrow, stem cells) transduced with a vector expressing a marker or reporter gene used to distinguish it from other similar host tissues.

*Other*—products that do not specifically fit the above (e.g., tumor vaccines in which cells are cultured or transduced *ex vivo* with a vector).

*In Vivo*

*Viral vectors*—the direct administration of a viral vector (e.g., retrovirus, adenovirus, adeno-associated virus, herpes, vaccinia) to patients.

*Plasmid vectors*—the direct administration of plasmid vectors with or without other vehicles (e.g., lipids) to patients.

*Vector producer cells*—the direct administration of retroviral vector producer cells (e.g., murine cells producing HTK vector) to patients.

***Preclinical Safety Evaluation*** The goal of preclinical safety evaluation includes the recommendation of an initial safe starting dose and safe dose-escalation scheme in humans, identification of potential target organ(s) of toxicity, identification of appropriate parameters for clinical monitoring, and identification of "at risk" patient population(s). Therefore, when feasible, toxicity studies should be performed in relevant species to assess a dose-limiting toxicity. General considerations in study design include selection of the model (e.g., species, alternative model, animal model, or disease), dose (e.g., route, frequency, and duration), and study endpoint (e.g., activity and/or toxicity).

The approach to preclinical safety evaluation of biotechnology-derived products, including novel cellular and gene therapies, has been referred to as the *case-by case* approach. This approach is science-based, data-driven, and flexible. The major distinction from past practices from traditional pharmaceuticals is that the focus is directed at asking specific questions across various product categories. Additionally, there is a consistent reevaluation of the knowledge base to reassess real or theoretical safety concerns and hence reevaluation of the need to answer the same questions across all product categories. In some cases there may even be conditions that may not need specific toxicity studies: for example, when there is a strong efficacy model that is rationally designed to answer specific questions and/or there is previous human experience with a similar product with respect to dose and regimen.

***Basic Principles for Preclinical Safety Evaluation of Cellular and Gene Therapies (Biotechnology-Derived Products in General)***

- Use of product in animal studies that is comparable or the same as the product proposed for clinical trial(s).
- Adherence to basic principles of GLP to ensure quality of the study including a detailed protocol prepared prospectively.
- Use of the same or similar route and method of administration as proposed for clinical trials (whenever possible).
- Determination of appropriate doses delivered based on preliminary activity obtained from both *in vitro* and *in vivo* studies (i.e., finding a dose likely to be effective and not dangerous, a no observed adverse effect level, and a dose-causing dose-limiting toxicity).

- Selection of one or more species sensitive to the endpoint being measured (e.g., infections or pathologic sequelae) and/or biological activity or receptor binding.
- Consideration of animal model(s) of disease as better to assess the contribution of changes in physiologic or underlying physiology to safety and efficacy.
- Determination of effect on host immune response.
- Localization/distribution studies—evaluation of target tissue, normal surrounding tissue, and distal tissue sites and any alteration in normal or expected distribution, and local reactogenicity.

*Additional Considerations for Cellular Therapies*

- Evaluation of cytopathogenicity.
- Evaluation of signs of cell transformation/growth factor dependence—effect on animal cells, normal human cells, and cells prone to transform easily.
- Determination of alteration in cell phenotype, altered cell products, and/or function.
- Tumorigenicity.

*Additional Considerations for Gene Therapies*

- Determination of phenotype/activation state of effector cells.
- Determination of vector/transgene toxicity.
- Determination of potential transfer to germ line.
- *In vitro* challenge studies—evaluation of recombination or complementation, potential for "rescue" for subsequent infection with wild-type virus.
- Determination of persistence of cells/vector.
- Determination of potential for insertional mutagenesis (malignant transformation).
- Determination of environmental spread (e.g., viral shedding).

## 6 TOXICITY TESTING: SPECIAL CASES

On paper, the general case guidelines for the evaluation of the safety of drugs are relatively straightforward and well understood. However, there are also a number of special case situations under which either special rules apply or some additional requirements are relevant. The more common of these are summarized next.

### 6.1 Oral Contraceptives

Oral contraceptives are subject to special testing requirements. These have recently been modified so that in addition to those preclinical safety tests generally required, the following are also required [28]:

- A 3 year carcinogenicity study in beagles (this is a 1987 modification in practice from earlier FDA requirements and the 1974 publication).
- A rat reproductive (Segment I) study including a demonstration of return to fertility.

### 6.2 Life-Threatening Diseases (Compassionate Use)

Drugs to treat life-threatening diseases are not strictly held to the sequence of testing requirements because the potential benefit on any effective therapy in these situations is so high [29]. In the early 1990s, this situation applied to AIDS-associated diseases and cancer. The development of more effective HIV therapies (protease inhibitors) has now made cancer therapy more the focus of these considerations. Although the requirements for safety testing prior to initial human trials is unchanged, subsequent requirements are flexible and subject to negotiation and close consultation with the FDA's Division of Oncology (within CDER) [29, 30]. The more recent thinking on anticancer agents has been reviewed by DeGeorge et al. [31]. The preclinical studies that will be required to support clinical trials and marketing of new anticancer agents will depend on the mechanism of action and the target clinical population. Toxicity studies in animals will be required to support initial clinical trials. These studies have multiple goals: (1) determine a starting dose for clinical trials, (2) identify target organ toxicity and assess recovery, and (3) assist in the design of clinical dosing regimens.

The studies should generally confirm to the protocols recommended by the National Cancer Institute, as discussed by Greishaber [32]. In general, it can be assumed that most antineoplastic cytotoxic agents will be highly toxic. Two studies are essential to support initial clinical trials (IND phase) in patients with advanced disease. These are studies of 5–14 days in length, but with longer recovery periods. A study in rodents is required that identifies those doses that produce either life-threatening or non-life-threatening toxicity. Using the information from this first study, a second study in nonrodents (generally the dog) is conducted to determine if the tolerable dose in rodents produces life-threatening toxicity. Doses are compared on a $mg/m^2$ basis. The starting dose in initial clinical trails is generally one-tenth of that required to produce severe toxicity in rodents (STD10) or one-tenth the highest dose in nonrodents that does not cause severe irreversible toxicity. While not required, information on pharmacokinetic parameters, especially data comparing the plasma concentration associated with toxicity in both species, is very highly regarded. Special attention is paid to organs with high cell-division rates—bone marrow, testes, lymphoid tissue, and GI tract. As these agents are almost always given intravenously, special attention needs to be given relatively early in development to intravenous irritation and blood compatibility study. Subsequent studies to support the New Drug Application will be highly tailored, depending on the following:

- Therapeutic indication and mechanism of action.
- The results of the initial clinical trials.
- The nature of the toxicity.
- Proposed clinical regimen.

Even at the NDA stage, toxicity studies with more than 28 days of dosing are rarely required. While not required for the IND, assessment of genotoxicity and developmental toxicity will need to be addressed. For genotoxicity, it will be important to establish the ratio between cytotoxicity and mutagenicity. *In vivo* models (e.g., the mouse micronucleus test) can be particularly important in demonstrating the lack of genotoxicity at otherwise subtoxic doses. For developmental toxicity, ICH stage C–D studies (traditionally known as Segment II studies for teratogenicity in rat and rabbits) will also be necessary.

The emphasis of this discussion has been on purely cytotoxic neoplastic agents. Additional considerations must be given to cytotoxic agents that are administered under special circumstances: those that are photoactivated, delivered as liposomal emulsions, or delivered as antibody conjugates. These types of agents will require additional studies. For example, a liposomal agent will need to be compared to the free agent and a blank liposomal preparation. There are also studies that may be required for a particular class of agents. For example, anthracyclines are known to be cardiotoxic, so comparison of a new anthracycline agent to previously marketed anthracyclines will be expected.

In addition to antineoplastic, cytotoxic agents, there are cancer therapeutic or preventative drugs that are intended to be given on a chronic basis. This includes chemopreventatives, hormonal agents, and immunomodulators. The toxicity assessment studies on these will more closely resemble those of more traditional pharmaceutical agents. Chronic toxicity, carcinogenicity, and full developmental toxicity (ICH A–B, C–D, E–F) assessments will be required. For a more complete review, the reader is referred to DeGeorge et al. [31].

### 6.3 Optical Isomers

The FDA (and other regulatory agencies, as reviewed by Daniels et al. [33]) has become increasingly concerned with the safety of stereoisomeric or chiral drugs. Stereoisomers are molecules that are identical to one another in terms of atomic formula and covalent bonding, but differ in the three-dimensional projections of the atoms. Within this class are those molecules that are nonsuperimposable mirror images of one another. These are called enantiomers (normally designated as *R*- or *S*-). Enantiometric pairs of a molecule have identical physical and chemical characteristics except for the rotation of polarized light. Drugs have generally been mixtures of optical isomers (enantiomers), because of the difficulties in separating the isomers. It has become apparent in recent years, however, that these different isomers may have different degrees of both desirable therapeutic and undesirable toxicologic effects. Technology has also improved, to the extent that it is now possible to perform chiral-specific syntheses, separations, and/or analyses. It is now highly desirable from a regulatory [29, 30, 34, 35] basis to develop a single isomer unless all isomers have equivalent pharmacological and toxicologic activity. The FDA has divided enantiometric mixtures into the following categories:

- Both isomers have similar pharmacologic activity, which could be identical, or they could differ in the degree of efficacy.
- One isomer is pharmacologically active, while the other is inactive.
- Each isomer has completely different activity.

During preclinical assessment of an enantiometric mixture, it may be important to determine to which of these three classes it belongs. The pharmacological and toxicological properties of the individual isomers should be characterized. The pharmacokinetic profile of each isomer should be characterized in animal models with regard to disposition and interconversion. It is not at all unusual for each enantiomer to have a completely different pharmacokinetic behavior.

If the test article is an enantiomer isolated from a mixture that is already well characterized (e.g., already on the market), then appropriate bridging guides need to be performed which compare the toxicity of the isomer to that of the racemic mixture. The most common approach would be to conduct a subchronic (3 month) and a Sement II type teratology study with an appropriate "positive" control group that received the racemate. In most instances, no additional studies would be required if the enantiomer and the racemate did not differ in toxicity profile. If, on the other hand, differences are identified, then the reasons for this difference need to be investigated and the potential implications for human subjects need to be considered.

### 6.4 Special Populations: Pediatric and Geriatric Claims

Relatively few drugs marketed in the United States (approximately 20%) have pediatric dosing information available. Clinical trials had rarely been done specifically on pediatric patients. Traditionally, dosing regimens for children have been derived empirically by extrapolating on the basis of body weight or surface area. This approach assumes that the pediatric patient is a young adult, which simply may not be the case. There are many examples of how adults and children differ qualitatively in metabolic and/or pharmacodynamic responses to pharmaceutical agents. In their review, Schacter and DeSantis [36] state that "the benefit of having appropriate usage information in the product label is that health care practitioners are given the information necessary to administer drugs and biologics in a manner that maximizes safety, minimizes unexpected adverse events, and optimizes treatment efficacy. Without specific knowledge of potential drug effects, children may be placed at risk. In addition, [in] the absence of appropriate proscribing information, drugs and biologics that represent new therapeutic advances may not be administered to the pediatric population in a timely manner." In response to the need for pediatric information, the FDA had developed a pediatric plan. This two-phase plan called first for the development of pediatric information on marketed drugs. The second phase focused on new drugs. The implementation of the plan was to be coordinated by the Pediatric Subcommittee of the Medical Policy Coordinating Committee of CDER. The Pediatric Use Labeling Rule was a direct result of phase 1 in 1994 [7]. Phase 2 resulted in 1997 from a proposed rule entitled *Pediatric Patients: Regulations Requiring Manufacturers to Assess the Safety and Effectiveness of New Drugs and Biologics*. Soon after this rule was proposed, the FDA Modernization Act of 1997 was passed. FDAMA contained provisions that specifically addressed the needs and requirements for the development of drugs for the pediatric population.

The FDAMA bill essentially codified and expanded several regulatory actions initiated by the FDA during the 1990s. Among the incentives offered by the bill, companies would be given an additional six months of patent protection for performing pediatric studies (clinical trials) on already approved products. In fact, the

FDA was mandated by FDAMA to develop a list of over 500 drugs for which additional information would produce benefits for pediatric patients. The FDA is supposed to provide a written request for pediatric studies to the manufacturers [37].

In response to the pediatric initiatives, the FDA has published policies and guidelines and conducted a variety of meetings. CDER has established a web site (http//www.fda.gov/cder/pediatric), which lists three pages of such information. Interestingly, the focus has been on clinical trials, and almost no attention has been given to the preclinical toxicology studies that may be necessary to support such trials. There are three pages of documents on the pediatric web site. None appear to address the issue of appropriate testing. This is a situation that is just now being addressed and is in a great deal of flux.

In the absence of any guidelines from the FDA for testing drugs in young or "pediatric" animals, one must fall back on the maxim of designing a program that makes the most scientific sense. As a guide, the FDA designated levels of postnatal human development and the approximate equivalent ages (in the author's considered opinion) in various animal models are given in Table 7. The table is somewhat inaccurate, however, because of differences in the stages of development at birth. A rat is born quite underdeveloped when compared to a human being. A one-day old rat is not equivalent to a one-day old full-term human infant. A four-day old rat would be more appropriate. In terms of development, the pig may be the best model of those listed; however, one should bear in mind that different organs have different developmental schedules in different species.

Table 7 can be used as a rough guide in designing toxicity assessment experiments in developing animals. In designing the treatment period, one needs to consider not only the dose and the proposed course of clinical treatment, but also the proposed age of the patient, and whether or not an equivalent dosing period in the selected animal model covers more than one developmental stage. For example, if the proposed patient population is human infants, initiating a toxicity study of the new pharmaceutical agent in 3-day-old rats is not appropriate. Furthermore, if the proposed course of treatment in adult children is 2 weeks, it is unlikely that this would cross over into a different developmental stage. A 2-week treatment initiated in puppies, however, might easily span two developmental stages. Thus, in designing an experiment in young animals, one must carefully consider the length of the treatment period, balancing the developmental age of the animal model and the proposed length of clinical treatment. Where appropriate (infant animals), one needs to also assess changes in standard developmental landmarks, (e.g., eye opening, pinae eruption, external genitalia development) as well as the more standard

**TABLE 7 Comparison of Postnatal Development Stages**

| Stage | Human | Rat | Dog | Pig |
|---|---|---|---|---|
| Neonate | Birth to 1 month | Birth to 1 week | Birth to 3 weeks | Birth to 2 weeks |
| Infant | 1 month to 2 years | 1–3 weeks | 3–6 weeks | 2–4 weeks |
| Child | 2–12 years | 3–9 weeks | 6 weeks to 5 months | 4 weeks to 4 months |
| Adolescent | 12–16 years | 9–13 weeks | 5–9 months | 4–7 months |
| Adult | Over 16 years | Over 13 weeks | Over 9 months | Over 7 months |

indicators of target organ toxicity. The need for maintaining the experimental animals past the dosing period, perhaps into sexual maturity, to assess recovery or delayed effects needs also to be carefully considered.

To summarize, the current status of assessment of toxicity in postnatal mammals, in response to the pediatric initiatives covered in FDAMA, is an extremely fluid situation. One needs to carefully consider a variety of factors in designing the study and should discuss proposed testing programs with the appropriate office at the CDER.

Drugs intended for use in the elderly, like those intended for the very young, may also have special requirements for safety evaluation, but geriatric issues were not addressed in the FDAMA of 1997. The FDA has published a separate guidance document for geriatric labeling [7]. As was the case with pediatric guidance, this document does not address preclinical testing. With the elderly, the toxicological concerns are quite different from the developmental concerns associated with pediatric patients. With the elderly, one must be concerned with the possible interactions between the test article and compromised organ function. The FDA had previously issued a guidance for examining the clinical safety of new pharmaceutical agents in patients with compromised renal and/or hepatic function [38]. The equivalent ICH guideline (S5A) was issued in 1994. Whether this type of emphasis will require toxicity testing in animal models with specifically induced organ insufficiency remains to be seen. In the interim, we must realize that there is tacit evaluation of test article-related toxicity in geriatric rodents for those agents that undergo 2 year carcinogenicity testing. As the graying of America continues, labeling for geriatric use may become more of an issue in the future.

### 6.5 Orphan Drugs

The development of sophisticated technologies, coupled with the rigors and time required for clinical and preclinical testing, has made pharmaceutical development very expensive. In order to recoup such expenses, pharmaceutical companies have tended to focus on therapeutic agents with large potential markets. Treatment for rare but life-threatening diseases have been "orphaned" as a result. An orphan product is defined as one targeted at a disease that affects 200,000 or fewer individuals in the United States. Alternatively, the therapy may be targeted for more than 200,000 but the developer would have no hope of recovering the initial investment without exclusivity. The Orphan Drug Act of 1983 was passed in an attempt to address this state of affairs. Currently applicable regulations were put in place in 1992 [39]. In 1994, there was an attempt in Congress to amend the Act, but it failed to be passed into law. The current regulations are administered by the Office of Orphan Product Development (OPD). The Act offers the following incentives to encourage the development of products to treat rare diseases:

- Seven years exclusive market following the approval of a product for an orphan disease.
- Written protocol assistance from the FDA.
- Tax credits for up to 50% of qualified clinical research expenses.
- Available grant to support pivotal clinical trials.

As reviewed by Haffner [40], other developed countries have similar regulations.

The ODA did not change the requirements of testing drug products. The nonclinical testing programs are similar to those used for more conventional products. They will undergo the same FDA review process. A major difference, however, is the involvement of the OPD. A sponsor must request OPD review. Once OPD determines that a drug meets the criteria for orphan drug status, it will work with the sponsor to provide the assistance required under the Act. The ODA does not review a product for approval. The IND/NDA process is still handled by the appropriate reviewing division (e.g., Cardiovascular) for formal review. The Act does not waive the necessity for submission of an IND and for the responsibility of toxicological assessment. As always, in cases where there is ambiguity, a sponsor may be well served to request a pre-IND meeting at the appropriate division to discuss the acceptability of a toxicology assessment plan.

### 6.6 Botanical Drug Products

There is an old saying, "What goes around, comes around," and so it is with botanicals. At the beginning of the 20th century, most marketed pharmaceutical agents were botanical in origin. For example, aspirin was first isolated from willow bark. These led the way in the middle part of the century, for reasons having to do with patentability, manufacturing costs, standardization, selectivity, and potency. The dawning of the 21st century has seen a grass-roots return to botanical preparations (also sold as herbals or dietary supplements). These preparations are being marketed to the lay public as "natural" supplements to the nasty synthetic chemicals now proscribed as pharmaceutical products. In 1994, the Dietary Supplement Health and Education Act was passed, which permitted the marketing of dietary supplements (including botanicals) with limited submissions to the FDA [33]. If a producer makes a claim that an herbal preparation is beneficial to a specific part of the body (e.g., enhanced memory), then it may be marketed after a 75 day period of FDA review but without formal approval. On the other hand, if any curative properties are claimed, then the botanical will be regulated as a drug and producers will be required to follow the IND/NDA process. In 1997 and 1998 combined, some 26 INDs were filed for botanical products [41].

The weakness in the current regulation has to do with its ambiguity. The line between a beneficial claim and a curative claim is sometimes difficult to draw. What is the difference, for example, between an agent that enhances memory and one that prevents memory loss? Given the number of products and claims hitting the shelves every day, this situation will probably demand increased regulatory scrutiny in the future.

## 7 INTERNATIONAL PHARMACEUTICAL REGULATION AND REGISTRATION

### 7.1 International Conference on Harmonization

The International Conference on Harmonization (ICH) of Technical Requirements for Registration of Pharmaceuticals for Human Use was established to make the

drug-regulatory process more efficient in the United States, Europe, and Japan. The U.S. involvement grew out of the fact that the United States is party to the General Agreement on Tariffs and Trade, which included the Agreement on Technical Barriers to Trade, negotiated in the 1970s, to encourage reduction of nontariff barriers to trade [42]. The main purpose of the ICH is, through harmonization, to make new medicines available to patients with a minimum of delay. More recently, the need to harmonize regulation has been driven, according to the ICH, by the escalation of the cost of R&D. The regulatory systems in all countries have the same fundamental concerns about safety, efficacy, and quality, yet sponsors had to repeat many time-consuming and expensive technical tests to meet country-specific requirements. Secondarily, there was a legitimate concern over the unnecessary use of animals. Conference participants include representatives from the drug-regulatory bodies and research-based pharmaceutical industrial organizations of three regions: the European Union, the United States, and Japan, where over 90% of the world's pharmaceutical industries are located. Representation is summarized in Table 8. The biennial conference has met four times, beginning in 1991, rotating between sites in the United States, Europe, and Japan.

The ICH meets its objectives by issuing guidelines for the manufacturing, development, and testing of new pharmaceutical agents that are acceptable to all three major parties. For each new guideline, the ICH Steering Committee establishes an expert working group with representation from each of the six major participatory ICH bodies. So far, the ICH has proposed or adopted over 40 safety, efficacy, and quality guidelines (listed in Table 9) for use by the drug-regulatory agencies in the United States, Europe, and Japan. The guidelines are organized under broad categories: the "E" series having to do with clinical trials, the "Q" series having to do with quality (including chemical manufacturing and control as well as traditional GLP issues), and the "S" series having to do with safety. Guidelines may be obtained from the ICH secretariat, c/o of IFPMA, 30 rue de St.-Jean, PO Box 9, 1211 Geneva 18, Switzerland, or may be downloaded from a web site set up by Ms. Nancy McClure (http://www.mcclurenet.com/index.html). They are also published in the *Federal Register*. The guidelines of the "S" series will have the most impact on

**TABLE 8 ICH Representation**

| Country/Region | Regulatory[a] | Industry[a] |
|---|---|---|
| European Union | European Commission (2) | European Federation of Pharmaceutical Industries Associations (2) |
| Japan | Ministry of Health, Labor and Welfare (2) | Japanese Pharmaceutical Manufacturers Association (2) |
| United States | Food and Drug Administration (2) | Pharmaceutical Research and Manufacturers of America (2) |
| Observing organizations | World Health Organization, European Free Trade Area, Canadian Health Protection Branch | International Federation of Pharmaceutical Manufacturers Associations (2): also provides the secretariat |

[a]( ) = number of representatives on the ICH Steering Committee.

**TABLE 9 International Conference on Harmonization Guidelines**

| Reference | Guideline | Date Issued |
|---|---|---|
| E1 | The Extent of Population Exposure to Assess Clinical Safety | Oct 94 |
| E2A | Clinical Safety Data Management: Definitions and Standards for Expedited Reporting | Oct 94 |
| E2B(M) | Maintenance of the Clinical Safety Data Management including: Data Elements for Transmission of Individual Case Safety Reports | Nov 00 |
| E2B(R) | Revision of the E2B(M) ICH Guideline on Clinical Safety Data Management Data Elements for Transmission of Individual Case Safety Reports | May 05 |
| E2C | Clinical Safety Data Management: Periodic Safety Update Reports for Marketed Drugs | Nov 96 |
| E2D | Post-Approval Safety Data Management: Definitions and Standards for Expedited Reporting | Nov 03 |
| E2E | Pharmacovigilance Planning | Nov 04 |
| E3 | Structure and Content of Clinical Study Reports | Nov 95 |
| E4 | Dose Response Information to Support Drug Registration | Mar 94 |
| E5 | Ethnic Factors in the Acceptability of Foreign Clinical Data | Feb 98 |
| E6 | Good Clinical Practice: Consolidated Guideline; Notice of Availability | May 96 |
| E6A | GCP Addendum on Investigator's Brochure | Mar 95 |
| E6B | GCP: Addendum on Essential Documents for the Conduct of a Clinical Trial | Oct 94 |
| E7 | Studies in Support of Special Populations: Geriatrics | Jun 93 |
| E8 | Guidance on General Considerations for Clinical Trials; Notice | July 97 |
| E9 | Draft Guideline on Statistical Principles for Clinical Trials; Notice of Availability | Feb 98 |
| E10 | Choice of Control Group and Related Issues in Clinical Trials | July 00 |
| E11 | Clinical Investigation of Medicinal Products in the Pediatric Population | July 00 |
| E12A | Principles for Clinical Evaluation of New Antihypertensive Drugs | Mar 00 |
| E14 | Clinical Evaluation of QT/QTc Interval Prolongation and Proarrhythmic Potential for Non-Antiarrhythmic Drugs | May 05 |
| M2 | Electronic Transmission of Individual Case Safety Reports Message Specification | Nov 00 |
| M3 | Non-Clinical Safety Studies for the Conduct of Human Clinical Trials for Pharmaceuticals | July 97 |
| M4 | Organization of the Common Technical Document | Oct 05 |
| M5 | Data Elements and Standards for Drug Dictionaries | May 05 |
| Q1A(R) | Stability Testing of New Drug Substances and Products (2nd Revision) | May 03 |
| Q1B | Photostability Testing of New Drug Substances and Products | Nov 96 |
| Q1C | Stability Testing for New Dosage Forms | Nov 96 |
| Q1D | Bracketing and Matrixing Designs for Stability Testing of Drug Substances and Drug Products | Feb 02 |
| Q1E | Evaluation of Stability Data | Feb 03 |
| Q1F | Stability Data Package for Registration Applications in Climatic Zones III and IV | Feb 03 |
| Q2A | Validation of Analytical Procedures: Definitions and Terminology | Oct 94 |
| Q2B | Validation of Analytical Procedures: Methodology | Nov 96 |
| Q3A(R) | Impurities in New Drug Substances Revised Guideline | Feb 02 |

**TABLE 9** ***Continued***

| Reference | Guideline | Date Issued |
|---|---|---|
| Q3B(R) | Impurities in New Drug Products Revised Guideline | Feb 03 |
| Q3C | Impurities: Guideline for Residual Solvents | July 97 |
| Q3C(M) | Impurities: Guideline for Residual Solvents (Maintenance) | Sept 02 |
| Q5A | Viral Safety Evaluation of Biotechnology Products Derived from Cell Lines of Human or Animal Origin | Mar 97 |
| Q5B | Quality of Biotechnological Products: Analysis of the Expression Construct in Cells Used for Production of r-DNA Derived Protein Products | Nov 95 |
| Q5C | Q5C: Quality of Biotechnological Products: Stability Testing of Biotechnological/Biological Products | Nov 95 |
| Q5D | Q5D: Derivation and Characterisation of Cell Substrates Used for Production of Biotechnological/Biological Products | July 97 |
| Q5E | Comparability of Biotechnological/Biological Products Subject to Changes in Their Manufacturing Process | Nov 04 |
| Q6A | Specifications: Test Procedures and Acceptance Criteria for New Drug Substances and New Drug Products: Chemical Substances | Oct 99 |
| Q6B | Specifications: Test Procedures and Acceptance Criteria for Biotechnological/ Biological Products | May 99 |
| Q7A | Good Manufacturing Practice Guide for Active Pharmaceutical Ingredients | Nov 00 |
| Q8 | Pharmaceutical Development | Nov 05 |
| Q9 | Quality Risk Management | Nov 05 |
| S1A | Guideline on the Need for Carcinogenicity Studies of Pharmaceuticals | Nov 95 |
| S1B | Draft Guideline on Testing for Carcinogenicity of Pharmaceuticals | July 97 |
| S1C | Dose Selection for Carcinogenicity Studies of Pharmaceuticals | Oct 94 |
| S1C(R) | Guidance on Dose Selection for Carcinogenicity Studies of Pharmaceuticals: Addendum on a Limit Dose and Related Notes; Availability; Notice | July 97 |
| S2A | Genotoxicity: Guidance on Specific Aspects of Regulatory Genotoxicity Tests for Pharmaceuticals | Jul 95 |
| S2B | Genotoxicity: A Standard Battery for Genotoxicity Testing of Pharmaceuticals; Availability; Notice | July 97 |
| S3A | Toxicokinetics: Guidance on the Assessment of Systemic Exposure in Toxicity Studies | Oct 94 |
| S3B | Pharmacokinetics: Guidance for Repeated Dose Tissue Distribution Studies | Oct 94 |
| S4 | Single Dose Toxicity Testing; Revised Guidance | Aug 96 |
| S4A | Duration of Chronic Toxicity Testing in Animals (Rodent and Non-Rodent Toxicity Testing) | Sept 98 |
| S5A | Detection of Toxicity to Reproduction for Medicinal Products | Jun 93 |
| S5B(M) | Maintenance of the ICH Guideline on Toxicity to Male Fertility: An Addendum to the Guideline on Detection of Toxicity to Reproduction for Medicinal Products | Nov 95 |
| S6A | Guidance on Preclinical Safety Evaluation of Biotechnology-Derived Pharmaceuticals; Availability | Nov 97 |
| S7A | Safety Pharmacology Studies for Human Pharmaceuticals | Nov 00 |
| S7B | The Nonclinical Evaluation of the Potential for Delayed Ventricular Repolarization (QT Interval Prolongation) by Human Pharmaceuticals | May 05 |
| S8 | Immunotoxicology Studies for Human Pharmaceuticals | Sept 05 |

toxicologists. The biggest changes having to do with toxicological assessment are summarized next.

***Carcinogenicity Studies*** Carcinogenicity studies are covered in Guidelines S1A, S1B, and S1C. The guidelines are almost more philosophical than they are technical. In comparison to the EPA guidelines, for example, the ICH guidelines contain little in the way of concrete study criteria (e.g., the number of animals, the necessity for clinical chemistry). There is discussion on when carcinogenicity studies should be done, whether two species are more appropriate than one, and how to set dosages on the basis of human clinical PK data. The following major changes are being wrought by these guidelines:

- Only one 2 year carcinogenicity study should generally be required. Ideally, the species chosen should be the one most like humans in terms of metabolic transformations of the test article.
- The traditional second long-term carcinogenicity study can be replaced by a shorter-term alternative model. In practical terms, this guideline is beginning to result in sponsors conducting a 2 year study in the rat and a 6 month study in an alternative mouse model, such as the P53 or the TG.AC genetically manipulated mouse strains.
- In the absence of target organ toxicity with which to set the high dose at the maximally tolerated dose, the high dose can be set at the dose that produces an area under the curve (AUC). This is 25-fold higher than that obtained in human subjects.

***Chronic Toxicity*** Traditionally, chronic toxicity of new pharmaceuticals in the United States was assessed in studies of 1 year duration in both the rodent and the nonrodent species of choice. The European view was that studies of 6 months are generally sufficient. The resulting guideline (S4A) was a compromise. Studies of 6 months duration were recommended for the rodent, as rodents would also be examined in 2 year studies. For the nonrodent (dog, nonhuman primate, and pig) studies of 9 months duration were recommended.

***Developmental and Reproductive Toxicity*** This was an area in which there was considerable international disagreement and the area in which the ICH has promulgated the most technically detailed guidelines (S5A and S5B). Some of the major changes include the following:

- The traditional Segment I, II, and III nomenclature has been replaced with different nomenclature, as summarized in Table 10.
- The dosing period of the pregnant animals during studies on embryonic development (traditional Segment II studies) has been standardized.
- New guidelines for fertility assessment (traditional Segment I) studies have shortened the premating dosing schedule (e.g., in male rats from 10 weeks to 4 weeks). There has been an increased interest in assessment of spermatogenesis and sperm function.

**TABLE 10 Comparison of Traditional and ICH Guidelines for Reproductive and Developmental Toxicology**

| Traditional Protocol | Stages Covered | ICH Protocol | Dosing Regimen |
|---|---|---|---|
| Segment I (rats) | A. Premating to conception<br>B. Conception to implantation | Fertility and early embryonic development, including implantation | Males: 4 weeks premating, mating (1–3 weeks), plus 3 weeks postmating<br>Females: 2 weeks premating, mating through day 7 of gestation |
| Segment II (rabbits) | C. Implantation to closure of hard palate<br>D. Closure of hard palate to the end of pregnancy | Embryo–fetal development | Female rabbits: day 6 to day 20 of pregnancy |

| Study Title | Termination | Endpoints: In-life | Endpoints: Postmortem |
|---|---|---|---|
| Fertility and early embryonic development, including implantation | Females: day 13–15 of pregnancy<br>Males: day after completion of dosing | Clinical signs and mortality<br>Body weights and feed in-take<br>Vaginal cytology | Macroscopic exam + histological exam of gross lesions<br>Collection of reproductive organs for possible histology<br>Quantitation of corpora lutea and implantation sites<br>Seminology (count, motility, and morphology) |
| Embryo–fetal development | | Clinical signs and mortality<br>Body weights and changes<br>Feed in-take | Macroscopic exam + histological exam of gross lesions<br>Quantitation of corpora lutea and implantation sites<br>Fetal body weights<br>Fetal abnormalities |
| Pre- and postnatal development, including maternal function | | Clinical signs and mortality<br>Body weights and changes<br>Feed in-take<br>Duration of pregnancy<br>Parturition | Macroscopic exam + histological exam of gross lesions<br>Implantation<br>Abnormalities (including terata)<br>Live/dead offspring at birth<br>Pre- and postweaning survival and growth ($F_1$)<br>Physical development ($F_1$)<br>Sensory functions and reflexes ($F_1$)<br>Behavior ($F_1$) |

- The new guidelines allow for a combination of studies in which the endpoints typically assessed in the traditional Segment II and Segment III studies are now examined under a single protocol.

For a more complete review of the various study designs, the reader is refereed to the review by Manson [43].

While they were not quite as sweeping in approach as the aforementioned guidelines, a toxicologist working in pharmaceutical safety assessment should become familiar with all the other ICH guidelines in the S series.

In an interesting recent article, Ohno [44] discussed not the harmonization of nonclinical guidelines, but the need to harmonize the timing of nonclinical tests in relation to the conduct of clinical trials. For example, there are regional differences in the inclusion of women of childbearing potential in clinical trials. In the United States, including women in such trials is becoming more important, and therefore evaluation of embryo–fetal development will occur earlier in the drug development process than in Japan. Whether or not such timing or staging of nonclinical tests becomes part of an ICH guideline in the near future remains to be established.

### 7.2 Other International Considerations

The United States is the single largest pharmaceutical market in the world. But the rest of the world (particularly but not limited to the second and third largest markets, Japan and the European Union) represents in aggregate a much larger market, so no one develops a new pharmaceutical for marketing just in the United States. The effort at harmonization (exemplified by the International Conference on Harmonization, or ICH) has significantly reduced differences in requirements for these other countries, but certainly not obliterated them. Although a detailed understanding of their regulatory schemes is beyond this chapter, the bare bones and differences in toxicology requirements are not.

***European Union*** The standard European Union toxicology and pharmacologic data requirements for a pharmaceutical include the following

- Single-dose toxicity
- Repeat-dose toxicity (subacute and chronic trials)
- Reproduction studies (fertility and general reproductive performance, embryotoxicity, and peri/postnatal toxicity)
- Mutagenic potential (*in vitro* and *in vivo*)
- Carcinogenicity
- Pharmacodynamics
  Effects related to proposed drug indication
  General pharmacodynamics
  Drug interactions
- Pharmacokinetics
  Single dose
  Repeat dose

Distribution in normal and pregnant animals
Biotransformation
- Local tissue tolerance
- Environmental toxicity

In general, the registration process in the European Union allows one to apply to either an overall medicines authority or to an individual national authority. Either of these steps is supposed to lead to mutual recognition by all the individual members.

***Japan*** In Japan, the Ministry of Health, Labor and Welfare is the national regulatory body for new drugs [45].

The standard $LD_{50}$ test is no longer a regulatory requirement for new medicines in the United States, the European Union, or Japan. The Japanese guidelines were the first to be amended in accordance with this agreement, with the revised guidelines becoming effective in August 1993. The Japanese may still anticipate that single-dose (acute) toxicity studies should be conducted in at least two species, one rodent and one nonrodent (the rabbit is not accepted as a nonrodent). Both males and females should be included from at least one of the species selected: if the rodent, then a minimum of 5 per sex; if the nonrodent, at least 2 per sex. In nonrodents, both the oral and parenteral routes should be used, and normally the clinical route of administration should be employed. In nonrodents, only the intended route of administration need be employed; if the intended route of administration in humans is intravenous, then use of this route in both species is acceptable. An appropriate number of doses should be employed to obtain a complete toxicity profile and to establish any dose–response relationship. The severity, onset, progression, and reversibility of toxicity should be studied during a 14 day follow-up period, with all animals being necropsied. When macroscopic changes are noted, the tissue must be subjected to histological examination.

Chronic and subchronic toxicity studies are conducted to define the dose level, when given repeatedly, that causes toxicity, and the dose level that does not lead to toxic findings. In Japan, such studies are referred to as repeated-dose toxicity studies. As with single-dose studies, at least two animal species should be used, one rodent and one nonrodent (rabbit not acceptable). In rodent studies, each group should consist of at least 10 males and 10 females; in nonrodent species, 3 of each sex are deemed adequate. Where interim examinations are planned, however, the numbers of animals employed should be increased accordingly. The planned route of administration in human subjects is normally explored. The duration of the study will be dictated by the planned duration of clinical use (Table 11).

At least three different dose groups should be included, with the goals of demonstrating an overtly toxic dose and a no-effect dose, and establishing any dose–response relationship. The establishment of a nontoxic dose within the framework of these studies is more rigorously adhered to in Japan than elsewhere in the world. All surviving animals should also be necropsied, either at the completion of the study or during its extension recovery period, to assess reversal of toxicity and the possible appearance of delayed toxicity. Full histological examination is mandated on all nonrodent animals used in a chronic toxicity study; at a minimum, the highest-

**TABLE 11 Determination of Duration of Studies**

| Duration of Dosing in Toxicity Study | Duration of Human Exposure |
|---|---|
| 2 weeks | Single dose or repeated dosage not exceeding 12 days |
| 1 month | Single dose or repeated dosage not exceeding 4 weeks |
| 3 months | Repeated dosing exceeding 1 week and to a maximum of 12 weeks |
| 6 months (rodent) | Repeated dosing exceeding 13 weeks and to a maximum of lifetime |
| 9/12 months[a] (nonrodent) | Repeated dosing exceeding 6 months or where this is deemed to be appropriate |

[a]Where carcinogenicity studies are to be conducted, the Koseisho had agreed to forego chronic dosage beyond 6 months.

*Source*: New Drugs Division Notification No. 43, June 1992.

dose and control groups of rodents must be submitted to a full histological examination.

While the value of repeated-dose testing beyond 6 months has been questioned [46], such testing is a regulatory requirement for a number of agencies, including the U.S. FDA and the Koseisho. In Japan, repeated-dose testing for 12 months is required only for new medicines expected to be administered to humans for periods in excess of 6 months [47]. At the First International Conference on Harmonization held in Brussels, the consensus was that 12 month toxicity studies in rodents could be reduced to 6 months where carcinogenicity studies are required. While not yet adopted in the Japanese guidelines, 6 month repeated-dose toxicity studies have been accepted by the agencies of all three regions. Japan—like the European Union—accepts a 6 month duration if accompanied by a carcinogenicity study. The United States still requires a 9 month nonrodent study.

With regard to reproductive toxicology, as a consequence of the first ICH, the United States, European Union, and Japan agreed to recommend mutual recognition of their respective current guidelines. A tripartite, harmonized guideline on reproductive toxicology has achieved ICH Step 4 status and should be incorporated into the local regulations of all three regions soon [48]. This agreement represents a very significant achievement that should eliminate many obstacles to drug registration.

***Preclinical Male Fertility Studies*** Before conducting a single-dose male volunteer study in Japan, it is usually necessary to have completed a preclinical male fertility study (Segment I) that has an in-life phase of 10 or more weeks (i.e., 10 weeks of dosing, plus follow-up). Although government guidelines do not require this study to be completed before Phase I trials begin, the responsible Institutional Review Board or the investigator usually imposes this condition. Japanese regulatory authorities are aware that the Segment I male fertility study is of poor predictive value. The rat, which is used in this study, produces a marked excess of sperm. Many scientists therefore believe that the test is less sensitive than the evaluation of testicular weight and histology that constitute part of the routine toxicology assessment.

***Female Reproductive Studies*** Before entering a female into a clinical study, it is necessary to have completed the entire reproductive toxicology program, which consists of the following studies:

- Segment 1: Fertility studies in the rat or mouse species used in the Segment 2 program.
- Segment 2: Teratology studies in the rat or mouse, and the rabbit.
- Segment 3: Late gestation and lactation studies in a species used in the Segment 2 studies.

Such studies usually take approximately 2 years. Although the U.S. regulations state the need for completion of Segments 1 and 2, and the demonstration of efficacy in male patients, where appropriate, before entering females into a clinical program, the current trend in the United States is toward relaxation of the requirements to encourage investigation of the drug both earlier and in a larger number of females during product development. Growing pressure for the earlier inclusion of women in drug testing may encourage selection of this issue as a future ICH topic. The trend in the United States and the European Union toward including women earlier in the critical program has not yet been embraced in Japan, however.

The three tests required in Japan for genotoxicity evaluation are a bacterial gene mutation test, *in vitro* cytogenetics, and *in vivo* tests for genetic damage. The Japanese regulations state these tests to be the minimum requirement and encourage additional tests. Currently, Japanese guidelines do not require a mammalian cell gene mutation assay. Harmonization will likely be achieved by the Koseisho recommending all four tests, which will match requirements in the United States and European Union. At present, this topic is at Step 1 in the ICH harmonization process. The mutagenicity studies should be completed before the commencement of Phase II clinical studies.

Guidelines presented at the second ICH are likely to alter the preclinical requirements for registration in Japan; they cover toxicokinetics and when to conduct repeated-dose tissue distribution studies. The former document may improve the ability of animal toxicology studies to predict possible adverse events in humans; currently, there are no toxicokinetic requirements in Japan, and their relevance is questioned by many there. Although there is general agreement on the registration requirement for single-dose tissue distribution studies, implementation of the repeated-dose study requirement has been inconsistent across the three ICH parties.

### 7.3 Safety Pharmacology

Japan was the first major country to require extensive pharmacological profiling on all new pharmaceutical agents as part of the safety assessment profile. Prior to commencement of initial clinical studies, the drug's pharmacology must be characterized in animal models. In the United States and Europe, these studies have been collectively called safety pharmacology studies. For a good general review of the issues surrounding safety pharmacology the reader is referred to Hite [1]. The Japanese guidelines for such characterizations were published in 1991 (New Drugs Division Notification No. 4, January 1991). They include the following:

- Effects on general activity and behavior.
- Effects on the central nervous system.
- Effects on the autonomic nervous system and smooth muscle.
- Effects on the respiratory and cardiovascular systems.
- Effects on the digestive system.
- Effects on water and electrolyte metabolism.
- Other important pharmacological effects.

In the United States, pharmacological studies in demonstration of efficacy have always been required, but specific safety pharmacological studies have never been required. Special situational or mechanistic data would be requested on a case-by-case basis. This is a situation that is changing. In the United States the activities of the Safety Pharmacology Discussion Group, for example, have helped bring attention to the utility and issues surrounding safety pharmacology data. In 1999 and 2000, the major toxicological and pharmacological societal meetings had symposia on safety pharmacological testing. Many major U.S. pharmaceutical companies are in the process of implementing programs in safety pharmacology. The issue has been taken up by the ICH and the draft guideline is currently at the initial stages of review. This initial draft (Guideline S7) includes core tests in the assessment of CNS, cardiovascular, and respiratory function. Studies will be expected to be performed under GLP guidelines.

## 8 CONCLUSION

In summary, we have touched upon the regulations that currently control the types of preclinical toxicity testing done on potential human pharmaceuticals. This chapter has reviewed the history, the law, the regulations themselves, the guidelines, and common practices employed to meet regulatory standards.

## REFERENCES AND SUGGESTED READING

1. Hite M. Safety pharmacology approaches. *Int J Toxicol* 1997;16:23–31, 37.7.3.
2. FDLI. *Compilation of food and drug laws, Vol. I–III and supplement*. Washington, DC: FDLI; 1995–1998.
3. Gad SC. *Drug Safety Evaluation*. Hoboken, NJ: John Wiley & Sons; 2002.
4. Lumpkin M. *Guidance for Industry: Content and Format of Investigational New Drug Applications* (INDs) *for Phase 1 Studies of Drugs, Including Well-characterized, Therapeutic, and Biotechnology-Derived Products*. 1995. http://www.fda.gov/cder/guidance/phase1.pdf.
5. FDA. *Environmental Assessment Technical Assistance Handbook*. NTIS No. PB87-175345; 1987.
6. *Annual Report of the Pharmaceutical Research and Manufacturer Association* (Priority #2: Improved FDA Regulation of Drug Development). 1998.
7. *Guidance for Industry: Content and Format for Geriatric Labeling (Draft Guidance)*. 1998. http://www.fda.gov/cder/guidance/2527dft.pdf.

8. *Guidance for Industry and Reviewers: Repeal of Section 507 of the Federal Food, Drug and Cosmetic Act*. 1997. http://www.fda.gov/cder/guidance/index.htm.
9. *Points to Consider in the Manufacture and Testing of Monoclonal Antibody Products for Human Use*. 1997. http://www.fda.gov/cber/cberftp.html.
10. *United States Pharmacopeia XXIV* and *National Formulary 19*. Philadelphia, PA: USP; 2000.
11. Anderson O. *The Health of a Nation; Harvey Wiley and the Fight for Pure Food*. Chicago: University of Chicago; 1988.
12. Burns J. Overview of safety regulations governing food, drug and cosmetics. In Homberger F, Ed. *The United States in Safety and Evaluation and Regulation of Chemicals 3: Interface Between Law and Science*. New York: Karger; 1983.
13. Goldenthal E. Current view on safety evaluation of drugs. *FDA Papers* May 1968; 13–18.
14. Blakeslee D. Thalidomide, a background briefing. *JAMA HIV/AIDS Information Center Newsline*. 1998. http://www.ama-assn.org/special/hiv/newsline/briefing/briefing.htm.
15. FDA. *Introduction to Total Drug Quality*. Washington, DC: US. Government Printing Office; 1971.
16. ICH. *Non-Clinical Safety Studies for the Conduct of Human Clinical Trials for Pharmaceuticals*. 2000.
17. *Points to Consider in the Production and Testing of New Drugs and Biologicals Produced by Recombinant DNA Technology*. Washington, DC: FDA; 1985.
18. *Points to Consider in Human Somatic Cell Therapy and Gene Therapy*. Washington, DC: FDA; 1991.
19. *Points to Consider in the Characterization of Cell Lines Used to Produce Biologicals*. Washington, DC: FDA; 1993.
20. Wessinger J. Pharmacologic and toxicologic considerations for evaluating biologic products. *Reg Tox Pharmacol* 1989;10:255–263.
21. Hayes T, Ryffel B. Symposium in writing: safety considerations of recombinant protein therapy; introductory comments. *Clin Immunol Immunother* 1997;83:1–4.
22. Buesing, Mary McSweegan. Submitting biologics applications to the Center for Biologics Evaluation and Research Electronically. *Drug Inf J* 1999;33:1–15.
23. Ryffel B. Safety of human recombinant proteins. *Biomed Environ Sci* 1997;10:65–71.
24. Ryffel B. Unanticipated human toxicology of recombinant proteins. *Arch Toxicol Suppl* 1996;18:333–341.
25. Well-Characterized Biologics Would Be Regulated Under FD&C Act By Kassebaum FDA Reform Bill. *-F-D-C Reports (pink sheets)* 1996;58:11–12.
26. Kessler D, Siegel J, Noguchi P, Zoon K, Feidon K, Woodcock J. Regulation of somatic-cell therapy and gene therapy by the Food and Drug Administration. *N Engl J Med* 1993;329:1169–1173.
27. *Points to Consider in Human Somatic Cell Therapy and Gene Therapy*. Washington, DC: FDA; 1996.
28. Berliner VR. US Food and Drug Administration requirements for toxicity testing of contraceptive products. In Briggs MH, Diczbalusy E, Eds. *Pharmacological Models in Contraceptive Development. Acta Endocrinol (Copenhagen)* 1974;185:240–253.
29. FDA. Investigational new drug, antibiotic, and biological drug product regulations, procedures for drug intended to treat life-threatening and severely debilitating illnesses. *Fed Reg* 1988;53(No. 204, 198):41516–41524.
30. FDA. Good Laboratory Practices, *CFR* 21 Part 58, Apr 1988.

31. DeGeorge J, Ahn C, Andrews P, Bower M, Giorgio D, Goheer M, Lee-Yam D, McGuinn W, Schmidt W, Sun C, Tripathi S. Regulatory considerations for the preclinical development of anticancer drugs. *Cancer Chemother Pharmacol* 1998;41:173–185.
32. Grieshaber C. Prediction of human toxicity of new antineoplastic drugs from studies in animals. In Powis G, Hacker M, Eds. *The Toxicity of Anticancer Drugs*. New York: Pergamon Press; 1991, pp 10–24.
33. Daniels J, Nestmann E, Kerr A. Development of stereoisomeric (chiral) drugs: a brief review of the scientific and regulatory considerations. *Drug Inf J* 1997;639–646.
34. FDA. Perspective on development of stereoisomers. *AAPS*, 16 May 1988.
35. FDA's Policy Statement for the Development of New Stereoisomeric Drugs. 1992. http://www.fda.gov/cder/guidance/stereo.htm.
36. Schacter E, DeSantis P. Labeling of drug and biologic products for pediatric use. *Drug Inf J* 1998;32:299–303.
37. Hart C. Getting PSArt the politics and paperwork of pediatric drug studies. *Modern Drug Discov* 1999;2:15–16.
38. CDER. Guideline for the study of drugs likely to be used in the elderly. *FDA*, Nov 1989.
39. Orphan Drug Regulations 21 CFR PART 316. 1992. http://www.fda.gov/orphan/about/odreg.htm.
40. Haffner M. Orphan drug development—international program and study design issues. *Drug Inf J* 1998;32:93–99.
41. Wu K-M, DeGeorge J, Atrachi A, Barry E, Bigger A, Chen C, Du T, Freed L, Geyer H, Goheer A, Jacobs A, Jean D, Rhee H, Osterburg R, Schmidt W, Farrelly J. Regulatory science: a special update from the United States Food and Drug Administration. Preclinical issues and status of investigations of botanical drug products in the United States. *Toxicol Lett* 2000;111:199–202.
42. Barton B. International Conference on Harmonization—good clinical practices update. *Drug Inf J* 1998;32:1143–1147.
43. Manson J. Testing of pharmaceutical agents for reproductive toxicity. In Kimmel C, Buelke-Sam J, Eds. *Developmental toxicology*. New York: Raven Press; 1994; pp 379–402.
44. Ohno Y. Harmonization of the timing of the non-clinical tests in relation to the conduct of clinical trials. *J Control Release* 1999;62:57–63.
45. Currie WJC. *New Drug Approval in Japan*. Waltham, MA: Parexel; 1995.
46. Lumley CE, Walker SE. An international appraisal of the minimum duration of chronic animal toxicity studies. *Hum Exp Toxicol* 1992;11:155–162.
47. Yakuji Nippo, Ltd. *Guidelines for Toxicity Study of Drugs Manual*. Tokyo: Yakuji Nippo Ltd; 1994.
48. ICH. *International Conference on Harmonization Safety Steps 4/5 Documents*. Buffalo Grove, IL: Interpharm Press; 1997.

# DATA ANALYSIS

Jayesh Vora[1] and Pankaj B. Desai[2]
[1]*PRTM Management Consultants, Mountain View, California*
[2]*University of Cincinnati Medical Center, Cincinnati, Ohio*

**Contents**

1 Introduction and Scope
2 Systemic Concentration–Time Profiles and Urinary Excretion
  2.1 Single Intravenous Dose
  2.2 Multiple Intravenous Doses
  2.3 Single Oral Dose
  2.4 Other Routes of Administration
3 Pharmacokinetic Analysis
  3.1 Example 1: Intravenous Dosing
  3.2 Example 2: Oral Dosing
4 Implications of Pharmacokinetics on Drug Discovery and Development
References

## 1 INTRODUCTION AND SCOPE

Pharmacokinetic analysis of systemic (or blood or plasma) concentration–time data, or amount excreted in urine over various time intervals, represents a critical activity in the preclinical development of drug candidates. This chapter describes the common methods for analysis of such data in support of drug development efforts, with an emphasis on nonclinical drug discovery and development. Theoretical principles underlying pharmacokinetic data analysis are presented elsewhere in other chapters of this volume.

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

Concentration–time or urinary excretion data are typically analyzed to generate estimates of bioavailability, clearance, apparent volume of distribution, and fraction of dose converted to a specific metabolite. These parameters are vital for determination of appropriate dose(s), administration route(s), and dosing regimens in preclinical efficacy and toxicology studies.

Noncompartmental or compartmental methods are generally employed to perform these analyses [1, 2]. At their core, noncompartmental methods, based on statistical moment theory, require estimation of the area under the concentration–time curve. Compartmental methods assume that the body is composed of a series of compartments, even though such compartments have no physiologic relevance, with first-order (concentration-independent) transfer or elimination rate constants between compartments. A one-compartment model, for instance, depicts the body as a single, homogeneous unit, with first-order elimination.

Many modern software applications are available for rapid, reliable, and pharmacokinetic analyses by noncompartmental or compartmental methods (e.g., WinNonlin®, Kinetica®, PKAnalyst®, NCOMP) [3].

This chapter discusses noncompartmental methods of pharmacokinetic analysis with primary emphasis on data analysis following intravenous and oral dosing. This chapter is meant to provide an introductory discussion on pharmacokinetic data analysis with a few "hands-on" examples. A comprehensive discussion of all current and emerging pharmacokinetic data analysis techniques is documented in several excellent books [1–5].

## 2 SYSTEMIC CONCENTRATION–TIME PROFILES AND URINARY EXCRETION

### 2.1 Single Intravenous Dose

Following a single intravenous bolus administration, the test article (or drug) is assumed to instantaneously mix within the components of the circulating blood (red and white cells, plasma water, and plasma proteins). Concentration of drug in blood is then gradually diluted through distribution to tissues and organs. Concurrently, the drug may be metabolized or degraded by circulating enzymes (such as esterases) or by metabolic enzymes in the liver (or lungs) or may be excreted unchanged in urine through the kidneys. This complex combination of distribution, degradation, metabolism, and excretion results in a gradual decline in blood or systemic concentrations of drug over time. Whereas individual processes may not be generally discernible, their combination is generally evident as a monoexponential decline in systemic concentrations of drug over time. In instances when one or more processes are relatively slow, a multiexponential decline in systemic concentrations over time is evident. In any event, the apparent terminal monoexponential portion of concentration–time profile is assumed to represent elimination of drug from systemic circulation. A schematic of monoand multiexponential systemic concentration–time profiles following intravenous administration of a drug is depicted in Fig. 1.

The complex interplay between distribution and elimination processes observed in systemic circulation following an intravenous dose is also reflected in urine. As the systemic concentration of drug declines, the corresponding amount of unchanged

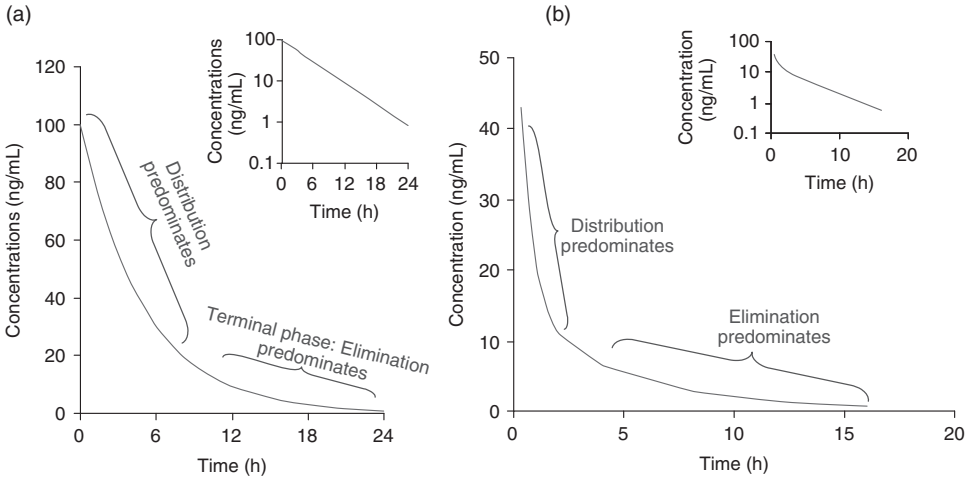


**FIGURE 1** Schematic of (a) monoexponential and (b) multiexponential concentration–time curves following a single intravenous dose of a drug. (Inset: Semilog concentration–time plot.)

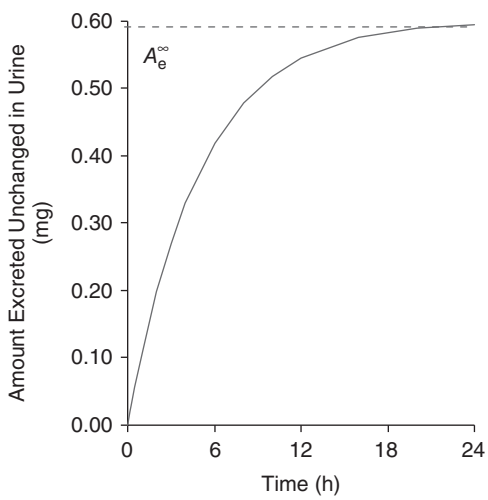


**FIGURE 2** Schematic of urinary excretion–time profile following a single intravenous dose of a drug.

(unmetabolized) drug excreted in urine increases, reaching a maximum value referred to as the maximum amount of drug excreted unchanged in urine ($A_e^\infty$). In cases when systemic concentration of an administered drug cannot be determined due to inaccessibility of appropriate sampling sites (such as in pediatric or infant patients), measurement of urinary excretion serves as a convenient tool for determination of corresponding pharmacokinetic parameters. A schematic of the urinary excretion–time profile of an intravenously administered drug is presented below in Fig. 2. This approach to pharmacokinetic characterization is utilized even for agents

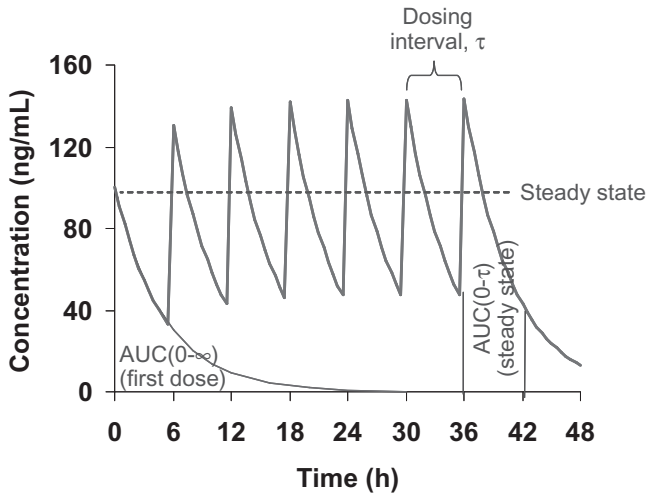


**FIGURE 3** Schematic of systemic concentration–time profile following multiple intravenous administration of a drug.

that are fully metabolized and some of the metabolites excreted in urine. The reader is referred to the literature for further discussions [5].

## 2.2 Multiple Intravenous Doses

Multiple drug administration generally results in increasing systemic exposure when concentrations from prior doses are not completely eliminated, reaching "steady-state" levels (Fig. 3). Such an increase in systemic exposure is commonly referred to as *accumulation*.

If additional doses are administered following "almost complete" elimination of prior doses then accumulation is minimal. The implications of accumulation (or lack thereof) are discussed elsewhere in this book and in the literature [5].

## 2.3 Single Oral Dose

Following an oral (or other nonparenteral) route of administration, the drug must be released from its dosage form and be absorbed from the site of administration into the systemic circulation. Distribution, degradation, metabolism, and excretion processes occur in parallel with drug absorption. At least initially, absorption rate predominates over other processes, resulting in increasing systemic concentrations of drug with time. As the amount of dose remaining to be absorbed declines at the site of administration, the rate of absorption declines and systemic concentrations reach a peak before declining thereafter in mono- or multiexponential fashion with time, as described previously. A schematic of systemic concentration–time profile following oral (or nonparenteral) administration is presented in Fig. 4. As discussed earlier, urinary excretion–time profiles provide an additional window into the absorption, distribution, and elimination processes following oral administration of a drug.

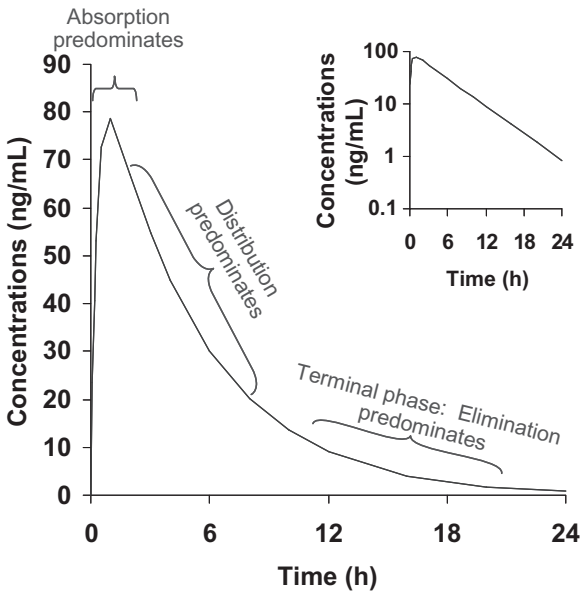


**FIGURE 4** Schematic of monoexponential concentration–time curve following a single oral (or nonparenteral) dose of a drug. (Inset: Semilog concentration–time plot.)

## 2.4 Other Routes of Administration

The test article may be administered to animals and humans during drug development by a number of other routes, including subcutaneous, intraperitoneally, and transdermally. Such routes result in systemic concentration–time profiles that mimic those following oral dosing (Fig. 4). Each includes absorption of the test article from the site of administration to systemic circulation with concurrent distribution, metabolism, and elimination processes. In such cases, drug absorption depends on the amount of administered dose remaining at the administration site. As the amount of administered dose decreases with time, the absorption rate declines as well. Such an absorption process is referred to as *first-order absorption*.

In contrast to changing absorption rates over time, a *zero-order* drug administration maintains constant rate of absorption. Such constant absorption rate is generally achieved by a constant rate intravenous infusion or through osmotic pumps. Typically, drugs with narrow therapeutic window (i.e., low safety margins) are administered in this fashion. Systemic concentrations increase steadily, reaching "steady state" when the rate of drug "input" equals the rate of drug "output" (i.e., removal of drug from systemic circulation). Once the infusion is terminated (or drug delivery device, such as an osmotic pump, is removed), systemic concentrations decline in exponential manner, reflecting either the mono- or multiexponential characteristics described earlier. A schematic of systemic concentration–time profile following a constant rate infusion is presented in Fig. 5.

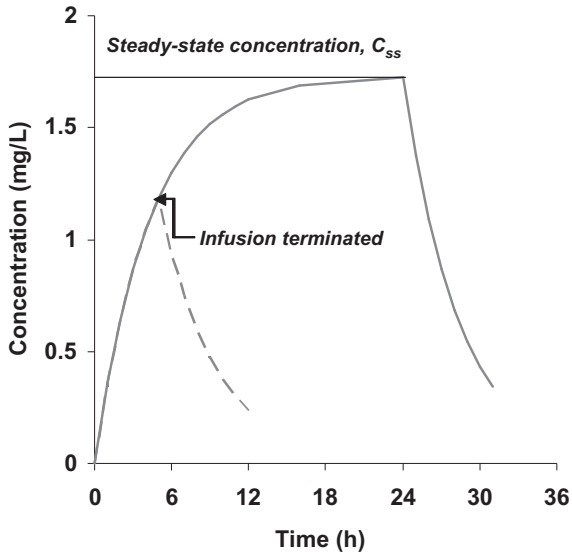


**FIGURE 5** Schematic of systemic concentration–time profile following intravenous infusion of drug. Infusion was terminated (a) prior to achievement of steady state or (b) following steady state.

## 3 PHARMACOKINETIC ANALYSIS

Pharmacokinetic analysis of systemic concentration–time data is typically initiated by noncompartmental methods resulting, at least initially, in estimation of the apparent terminal elimination half-life ($t_{1/2}$) and area under the concentration–time curve (*AUC*) (systemic exposure). The terminal elimination rate constant is calculated from the slope of the terminal $\log_{10}$(concentration)–time data:

$$\text{Terminal elimination rate constant, } k_e = -Slope \times 2.303 \tag{1}$$

where, *Slope* is the slope of the terminal portion of $\log_{10}$(concentration)–time data

$$\text{Terminal elimination half-life, } t_{1/2} = \frac{0.693}{k_e} \tag{2}$$

*AUC* values are generally estimated using the trapezoidal rule, which is the sum of the area of all trapezoids between two sampling points. The area of the curve beyond the last sampling point is estimated using the ratio of the corresponding concentration value and the terminal elimination rate constant.

$$AUC(0\text{–}t_{\text{last}}) = \sum_{t=0}^{t_n} \frac{(C_{n-1} + C_n)}{2} \cdot (t_n - t_{n-1}) \tag{3}$$

$$AUC(0\text{–}\infty) = AUC(0\text{–}t_{\text{last}}) + AUC_{\text{terminal}} \tag{4}$$

where

$$AUC_{\text{terminal}} = \frac{C_{\text{last}}}{k_{\text{e}}} \tag{5}$$

### 3.1 Example 1: Intravenous Dosing

Following an intravenous dose of 1 mg of Drug X to human volunteers, blood samples were collected for 24 hours. Plasma concentration–time data are summarized in Table 1.

***Elimination* $t_{1/2}$ *and AUC Values*** The terminal elimination rate constant ($k_e$) and terminal elimination half-life ($t_{1/2}$) may be estimated by examination of $\log_{10}$(concentration)–time profile to identify the terminal elimination phase and estimate the slope of the corresponding line (Eq. 1) (Table 2).

The application of Eqs. 3–5 for estimation of *AUC* values is exemplified in Table 3.

***Clearance, Mean Residence Time, and Volume of Distribution*** Estimation of apparent terminal half-life ($t_{1/2}$) and *AUC* values is directly useful in estimation of other key pharmacokinetic parameters, such as clearance (*CL*), mean residence time (*MRT*), and volume of distribution ($V_{ss}$).

$$\text{Systemic clearance } (CL) = \frac{Dose_{\text{IV}}}{AUC(0-\infty)} \tag{6}$$

**TABLE 1 Plasma Concentration–Time Profile Following a Single Intravenous Dose of 1 mg Drug X to Human Volunteers**

| Time (h) | Plasma Concentration (ng/mL) |
|---|---|
| 0 | 100 |
| 0.0833 | 98.3 |
| 0.167 | 96.7 |
| 0.25 | 95.1 |
| 0.5 | 90.5 |
| 1 | 81.9 |
| 2 | 67.0 |
| 3 | 54.9 |
| 4 | 44.9 |
| 6 | 30.1 |
| 8 | 20.2 |
| 10 | 13.5 |
| 12 | 9.07 |
| 16 | 4.08 |
| 20 | 1.83 |
| 24 | 0.82 |

**TABLE 2 Estimation of Apparent Terminal Elimination Rate Constant and Half-life Following a Single Intravenous Dose of 1 mg of Drug X to Human Volunteers**

| Time (h) | Plasma Concentration (ng/mL) | $\log_{10}$(Concentration) |
|---|---|---|
| 0 | 100 | |
| 0.0833 | 98.3 | |
| 0.167 | 96.7 | |
| 0.25 | 95.1 | |
| 0.5 | 90.5 | |
| 1 | 81.9 | |
| 2 | 67.0 | |
| 3 | 54.9 | |
| 4 | 44.9 | |
| 6 | 30.1 | |
| 8 | 20.2 | |
| 10 | 13.5 | |
| 12 | 9.07 | 0.958 |
| 16 | 4.08 | 0.610 |
| 20 | 1.83 | 0.263 |
| 24 | 0.823 | –0.0846 |
| | Slope = –0.0869 | |
| | $k_e$ (Eq. 1) = 0.2 $h^{-1}$ | |
| | $t_{1/2}$ (**Eq. 2**) = **3.46 h** | |

**TABLE 3 Estimation of Area Under Plasma Concentration–Time Curve (*AUC*) Following a Single Intravenous Dose of 1 mg Drug X to Human Volunteers**

| (1) Time (h) | (2) Plasma Concentration (ng/mL) | (3) $\frac{(C_{n-1}+C_n)}{2}$ | (4) $(t_n - t_{n-1})$ | (5) *AUC* Value of Each Trapezoid (Column 3 × Column 4) | (6) Cumulative *AUC* Values [$AUC(0–t_{last})$] |
|---|---|---|---|---|---|
| 0 | 100 | 99.2 | 0.0833 | 8.26 | 8.26 |
| 0.0833 | 98.3 | 97.5 | 0.0833 | 8.13 | 16.4 |
| 0.167 | 96.7 | 95.9 | 0.0833 | 7.99 | 24.4 |
| 0.25 | 95.1 | 92.8 | 0.25 | 23.2 | 47.6 |
| 0.5 | 90.5 | 86.2 | 0.5 | 43.1 | 90.7 |
| 1 | 81.9 | 74.5 | 1 | 74.5 | 165 |
| 2 | 67.0 | 61.0 | 1 | 61.0 | 226 |
| 3 | 54.9 | 49.9 | 1 | 49.9 | 276 |
| 4 | 44.9 | 37.5 | 2 | 75.1 | 351 |
| 6 | 30.1 | 25.2 | 2 | 50.3 | 401 |
| 8 | 20.2 | 16.9 | 2 | 33.7 | 435 |
| 10 | 13.5 | 11.3 | 2 | 22.6 | 458 |
| 12 | 9.07 | 6.57 | 4 | 26.3 | 484 |
| 16 | 4.08 | 2.95 | 4 | 11.8 | 496 |
| 20 | 1.83 | 1.33 | 4 | 5.31 | 501 |
| 24 | 0.823 | | | | |
| | | | | $AUC(0–t_{last})$ (Eq. 3) = 501 ng·h/mL | |
| | | | | $AUC_{terminal}$ (Eq. 5) = 4.11 ng·h/mL | |
| | | | | $AUC(0–\infty)$ (**Eq. 4**) = **505 ng·h/mL** | |

$$\text{Mean residence time } (MRT) = \frac{AUMC}{AUC} \tag{7}$$

where *AUMC* is the area under the first-moment time curve, calculated in similar fashion to *AUC*, using the following formula:

$$AUMC = \left\{ \sum_{t=0}^{t_n} \frac{(t_{n-1} \cdot C_{n-1} + t_n \cdot C_n)}{2} \cdot (t_n - t_{n-1}) \right\} + \left\{ \frac{C_{\text{last}} \cdot t_{\text{last}}}{k_e} + \frac{C_{\text{last}}}{k_e^2} \right\} \tag{8}$$

The apparent volume of distribution at steady state ($V_{ss}$) is the product of *CL* and *MRT* after a single intravenous dose of a drug:

$$V_{ss} = CL \cdot MRT \tag{9}$$

For Drug X, clearance is calculated using Eq. 6:

$$CL = 1\,\text{mg}/(505\,\text{ng} \cdot \text{h/mL}) = 1979\,\text{mL/h}$$

Calculation of *MRT* requires estimation of *AUMC* (Eq. 8) as illustrated in Table 4.

**TABLE 4 Estimation of Area Under First-Moment Time Curve (*AUMC*)**

| (1) Time (h) | (2) Plasma Concentration (ng/mL) | (3) $t \cdot C$ (Column 1 × Column 2) | (4) $\frac{(t_{n-1}C_{n-1} + t_nC_n)}{2}$ | (5) $(t_n - t_{n-1})$ | (6) *AUMC* (Column 4 × Column 5) | (7) Cumulative *AUMC* Values |
|---|---|---|---|---|---|---|
| 0 | 100 | 0 | 4.10 | 0.0833 | 0.341 | 0.341 |
| 0.0833 | 98.3 | 8.20 | 12.2 | 0.0833 | 1.01 | 1.35 |
| 0.167 | 96.7 | 16.1 | 20.0 | 0.083 | 1.66 | 3.02 |
| 0.25 | 95.1 | 23.8 | 34.5 | 0.250 | 8.63 | 11.6 |
| 0.5 | 90.5 | 45.2 | 63.6 | 1 | 31.8 | 43.4 |
| 1 | 81.9 | 81.9 | 108 | 1 | 108 | 151 |
| 2 | 67.0 | 134 | 149 | 1 | 149 | 301 |
| 3 | 54.9 | 165 | 172 | 1 | 172 | 473 |
| 4 | 44.9 | 180 | 180 | 2 | 360 | 833 |
| 6 | 30.1 | 181 | 171 | 2 | 342 | 1176 |
| 8 | 20.2 | 162 | 148 | 2 | 297 | 1472 |
| 10 | 13.5 | 135 | 122 | 2 | 244 | 1717 |
| 12 | 9.07 | 109 | 87.0 | 4 | 348 | 2065 |
| 16 | 4.08 | 65.2 | 50.9 | 4 | 204 | 2269 |
| 20 | 1.83 | 36.6 | 28.2 | 4 | 113 | 2381 |
| 24 | 0.823 | 19.8 | | | | |
| | | | | | $AUMC\ (0\text{–}t_{\text{last}}) = 2381$ | |
| | | | | | $(C_{\text{last}} \cdot t_{\text{last}})/k_e = 98.7$ | |
| | | | | | $C_{\text{last}}/k_e^2 = 20.6$ | |
| | | | | | $AUMC(0\text{–}\infty)$ (**Eq. 8**) = **2501 ng·h²/mL** | |

Therefore, using Eq. 7, mean residence time ($MRT$) is calculated as

$$MRT = 2501/505 = 4.95\,\text{h}$$

The apparent volume of distribution at steady state, $V_{ss}$, is calculated using Eq. 9:

$$V_{ss} = CL \cdot MRT = (1979\,\text{mL/h}) \cdot 4.95\,\text{h} = 9797\,\text{mL}$$

***Urinary Excretion*** Assume that urine samples were collected over various time intervals following a single intravenous dose of Drug X. The volume of urine collected over each interval was recorded and the corresponding concentration of unchanged Drug X was determined, as summarized in Table 5.

These apparent terminal elimination half-life ($t_{1/2}$) values as well as the fraction of dose excreted unchanged in urine ($f_e$) may be estimated by plotting the excretion rate ($dA_e/dt$) against the "midpoint" of sample collection ($t_{midp}$) (Fig. 6, Table 6). The theoretical basis for this derivation is beyond the scope of this chapter and is exhaustively discussed in the literature [2, 3, 5].

$$\log \frac{A_e}{dt} = \log(f_e \cdot Dose) - \frac{k_e t}{2.303} \tag{10}$$

where $f_e$, is the fraction of administered dose excreted unchanged in urine.

$$Slope = -k_e/2.303$$

$$t_{1/2} = 0.693/k_e = 3.46\,\text{h}$$

The intercept of the semilogarithmic plot provides an estimate of the fraction of administered dose excreted unchanged in urine at infinity ($f_e \cdot Dose$), 0.123 mg.

**TABLE 5 Urinary Concentration of Unchanged Drug Following Single Intravenous Administration of 1 mg Drug X to Healthy Volunteers**

| Time Interval (h) | Urine Volume (mL) | Concentration in Urine (ng/mL) |
|---|---|---|
| 0–0.5 | 100 | 571 |
| 0.5–1 | 74 | 698 |
| 1–2 | 88 | 1012 |
| 2–4 | 150 | 884 |
| 4–6 | 170 | 523 |
| 6–8 | 250 | 238 |
| 8–12 | 340 | 196 |
| 12–16 | 500 | 59.9 |
| 16–20 | 520 | 25.9 |
| 20–24 | 450 | 13.4 |

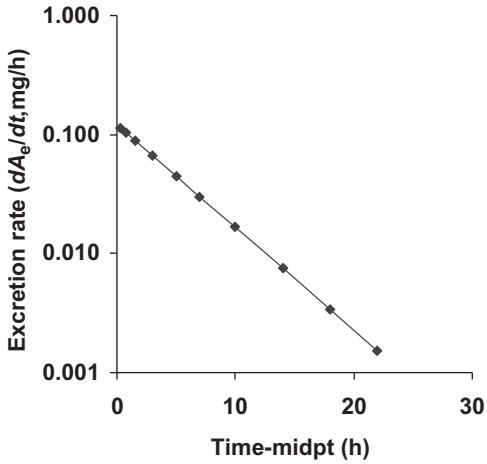


**FIGURE 6** Semilogarithmic plot of urinary excretion rate versus $t_{\text{midpt}}$.

**TABLE 6 Urinary Excretion Analysis Following Intravenous Administration of 1 mg Drug X to Healthy Volunteers**

| (1) Time Interval (h) | (2) Urine Volume (mL) | (3) Concentration in Urine (ng/mL) | (4) Amount Excreted in Urine ($A_e$, mg) (Column 2 × Column 3) | (5)[a] $t_{\text{midpta}}$ | (6)[b] $dA_e/dt$ |
|---|---|---|---|---|---|
| 0–0.5 | 100 | 571 | 0.0571 | 0.25 | 0.114 |
| 0.5–1 | 74 | 698 | 0.0517 | 0.75 | 0.103 |
| 1–2 | 88 | 1012 | 0.0890 | 1.5 | 0.0890 |
| 2–4 | 150 | 884 | 0.133 | 3 | 0.0663 |
| 4–6 | 170 | 523 | 0.0889 | 5 | 0.0444 |
| 6–8 | 250 | 238 | 0.0596 | 7 | 0.0298 |
| 8–12 | 340 | 196 | 0.0667 | 10 | 0.0167 |
| 12–16 | 500 | 59.9 | 0.0300 | 14 | 0.00749 |
| 16–20 | 520 | 25.9 | 0.0135 | 18 | 0.00337 |
| 20–24 | 450 | 13.4 | 0.00605 | 22 | 0.00151 |

[a] $t_{\text{midpt}}$ is the midpoint of the urine collection interval.
[b] $dA_e/dt$ is calculated as $(dA_{e(n+1)}-dA_e n)/(t_{n+1}-t_n)$.

***Multiple Intravenous Dosing*** For a drug following *linear* (dose- or concentration-independent) *pharmacokinetics*, *AUC* values at steady state equal those following the first dose:

$$AUC(0-\tau)_{\text{steady state}} = AUC(0-\infty)_{\text{first dose}} \tag{11}$$

where $\tau$ is the dosing interval. $AUC(0-\tau)_{\text{sleady state}}$ is calculated using the trapezoidal rule as illustrated in Table 1.

**TABLE 7 Estimation of Pharmacokinetic Parameters Following a Single 1 mg Oral Dose to Human Volunteers**

| (1) Time (h) | (2) Plasma Concentration (ng/mL) | (3) $\text{Log}_{10}$ (Concentration) | (4) $\frac{(C_{n+1}+C_n)}{2}$ | (5) $(t_{n+1}+t_n)$ | (6) *AUC* Value of Each Trapezoid | (7) $t \cdot C$ | (8) $(t_{n-1}C_{n-1}-t_nC_n)/2$ | (9) *AUMC* Values |
|---|---|---|---|---|---|---|---|---|
| 0 | 0 | | 1.29 | 0.0833 | 0.107 | 0 | 0.107 | 0.00896 |
| 0.0833 | 2.58 | | 3.78 | 0.0833 | 0.315 | 0.215 | 0.523 | 0.0436 |
| 0.167 | 4.99 | | 6.12 | 0.0833 | 0.510 | 0.832 | 1.321 | 0.110 |
| 0.25 | 7.24 | | 10.2 | 0.25 | 2.55 | 1.81 | 4.19 | 1.05 |
| 0.5 | 13.1 | | 17.4 | 0.5 | 8.68 | 6.56 | 14.1 | 7.04 |
| 1 | 21.6 | | 25.6 | 1 | 25.6 | 21.6 | 40.3 | 40.3 |
| 2 | 29.5 | | 30.1 | 1 | 30.1 | 59.1 | 75.6 | 75.6 |
| 3 | 30.7 | | 29.7 | 1 | 29.7 | 92.0 | 103 | 103 |
| 4 | 28.7 | | 25.3 | 2 | 50.6 | 115 | 123 | 246 |
| 6 | 21.9 | | 18.7 | 2 | 37.4 | 131 | 128 | 255 |
| 8 | 15.5 | | 13.1 | 2 | 26.1 | 124 | 115 | 230 |
| 10 | 10.6 | | 8.91 | 2 | 17.8 | 106 | 96.3 | 193 |
| 12 | 7.20 | 0.857 | 5.23 | 4 | 20.9 | 86.4 | 69.2 | 277 |
| 16 | 3.26 | 0.513 | 2.36 | 4 | 9.44 | 52.1 | 40.7 | 163 |
| 20 | 1.46 | 0.166 | 1.06 | 4 | 4.25 | 29.3 | 22.5 | 90.2 |
| 24 | 0.658 | –0.182 | | | | 15.8 | | |
| | Slope = –0.0866<br>$k_e = 0.199\,h^{-1}$<br>$\mathbf{t_{1/2} = 3.48\,h}$ | | | $AUC(0\text{–}t_{last}) = 264$<br>$C_{last}/k_e = 3.30$<br>$\mathbf{AUC(0\text{–}\infty) = 267\,ng{\cdot}h/mL}$<br>$\mathbf{F = 52.9\%}$ | | | $AUMC(0\text{–}t_{last}) = 1682$<br>$(C_{last} \cdot t_{last})/k_e = 79.2$<br>$C_{last}/k_e^2 = 16.6$<br>$\mathbf{AUMC(0\text{–}\infty) = 1778\,ng{\cdot}h^2/mL}$<br>$\mathbf{MRT = 0.150\,h}$ | |

**TABLE 8 Implications of Pharmacokinetic Data Analyses on Drug Discovery and Development**

| Key Pharmacokinetic Information | Methodological Approach | Drug Discovery and Development Implication |
|---|---|---|
| Systemic exposure and peak systemic concentrations | Analysis of systemic concentration–time profile: determination of the area under the curve (*AUC*) and $C_{max}$ | Potential correlation with drug response (efficacy and/or toxicity) |
| Rate of drug absorption | Analysis of systemic concentration–time profile: time to $C_{max}$ ($T_{max}$) | Choice of an appropriate route of administration and drug formulation |
| Systemically available fraction of administered dose | Bioavailability assessment | Choice of an appropriate route of administration and dose |
| Presystemic or first-pass elimination following oral dosing | Comparison of nonparenteral *AUC*, and other routes such as portal administration with *AUC* after intravenous administration | Factors affecting bioavailability |
| Drug persistence in the body | Apparent terminal elimination half-life ($t_{1/2}$) | Frequency of drug administration and designing "acceptable" dosing regimen |
| Drug accumulation following multiple dosing (parenteral or nonparenteral) | Analysis of systemic concentration–time profiles at steady state | Relationship of the drug levels to pharmacologically relevant drug levels (therapeutic window) |
| Route of drug clearance | Rate and extent of urinary excretion and extent of drug metabolism (including *in vitro* methodologies) | Factors affecting drug elimination; extent of intersubject variability in drug disposition; and drug–drug interactions |

### 3.2 Example 2: Oral Dosing

The plasma concentration–time data following a single oral dose of 1 mg Drug X is summarized in Table 7. Apparent terminal elimination $t_{1/2}$, *AUC* values, *CL*, and *MRT* values are calculated as indicated in Eqs. 1–2. The corresponding calculations are included in Table 7. The maximum plasma concentration ($C_{max}$) and time to reach $C_{max}$ ($T_{max}$) are generally determined by observation. Therefore, in this instance,

$$C_{max} = 30.7\,\text{ng/mL}$$

$$T_{max} = 3\,\text{h}$$

Oral bioavailability is defined by the following equation:

$$F(\%) = \frac{AUC_{\text{oral}}/Dose_{\text{oral}}}{AUC_{\text{IV}}/Dose_{\text{IV}}} \tag{12}$$

## 4 IMPLICATIONS OF PHARMACOKINETICS ON DRUG DISCOVERY AND DEVELOPMENT

This chapter describes the typical data analysis methodologies used to characterize the pharmacokinetics of a drug following single or multiple intravenous or oral dosing using systemic (plasma) concentration–time and urinary excretion–time data. The implications of such analyses on drug discovery and development are summarized in Table 8.

## REFERENCES

1. Gibaldi M, Perrier D. *Pharmacokinetics*, 2nd ed. New York: Marcel Dekker; 1982.
2. Boroujerdi M. *Pharmacokinetics: Principles and Application*. New York: McGraw-Hill; 2002.
3. Bourne DWA. Pharmacokinetic Software Resource. http://www.boomer.org/pkin/soft.html. Accessed on 26 February 2007.
4. Gabrielsson J, Weiner D. *Pharmacokinetic and Pharmacodynamic Data Analysis, Concepts and Applications*, 3rd ed. Stockholm, Sweden: Swedish Pharmaceutical Press; 2001.
5. Rowland M, Tozer TN. *Clinical Pharmacokinetics: Concepts and Applications*, 3rd ed, Baltimore: Lippincott Williams & Wilkins; 1995.

# PRECLINICAL DRUG DEVELOPMENT PLANNING

Nirmala Bhogal, Robert Combes, and Michael Balls
*FRAME, Nottingham, United Kingdom*

**Contents**

1 Introduction
  1.1 Overview of Objectives
  1.2 Drug Development Models
  1.3 Information Required Prior to Drug Authorization/Approval
2 Finding New Drug Targets
  2.1 Background
  2.2 Impact of New Technologies on Target Discovery
  2.3 Data Mining
3 Traditional Approaches to Drug Discovery and Development
  3.1 Hit to Lead
  3.2 Pharmacokinetics
4 Special Considerations for Novel Therapeutic Classes
  4.1 New Classes of Therapeutic Agents: A New Drug Development Strategy?
  4.2 Protein-Based Therapeutics
  4.3 Gene Therapeutics
  4.4 Nanomedicines
5 *In Vitro* Assays: Applications in Safety Pharmacology
  5.1 Background
  5.2 QT Prolongation
  5.3 ICH Guidelines

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

6 Systems Biology
6.1 Omics-Based Technologies
6.2 Microfluidics Systems
6.3 Enabling Technologies for Multiparameter Studies
7 Computational Methodology Used in Early Preclinical Drug Development
7.1 Combinatorial Chemistry and SAR
7.2 *In Silico* Prediction of ADME and Toxicity
7.3 Prediction of Tissue-Specific Exposure: PBPK Modeling
8 Animal Models Used in Preclinical Testing of Pharmaceuticals
8.1 Selection of a Suitable Test Species
8.2 Experimental Design
9 The Use of Prior Information
9.1 Sources of Prior Knowledge
9.2 Standardization of Data Collection and Meta-Analysis
10 Conclusions
Acknowledgment
References

## 1 INTRODUCTION

### 1.1 Overview of Objectives

It is well recognized that productivity in drug development has been disappointing over the last decade, despite the steady increase in R&D investment [1] and advances in techniques for producing potentially new candidate molecules. The principal problems appear to be a lack of efficacy and/or unexpected adverse reactions, which account for the majority of drug withdrawals and drugs undergoing clinical testing being abandoned. This high attrition rate could be dramatically reduced by improving the preclinical testing process, particularly by taking account of multidisciplinary approaches involving recent technologies, and by improving the design of preclinical projects to facilitate the collection and interpretation of relevant information from such studies, and its extrapolation to the clinical setting.

The objective of this chapter is to provide an overview of the early drug discovery and development processes. The main focus is the use of *in vitro* and *in silico* methods. This is because these techniques are generally applied during the earliest stages to identify new targets (target discovery) and lead compounds (drug discovery), as well as for subsequent drug development. They are also used to resolve equivocal findings from *in vivo* studies in laboratory animals, to guide selection of the most appropriate preclinical *in vivo* models, and to help define the mechanistic details of drug activity and toxicity. However, the use of animals in preclinical testing is also considered, since animal data form part of new medicine dossiers submitted to regulatory bodies that authorize clinical trials and the marketing of new products. The drug development process that will be considered is shown in Fig. 1. Definitions of the terminology and abbreviations/acronyms used in this chapter are listed in Table 1.

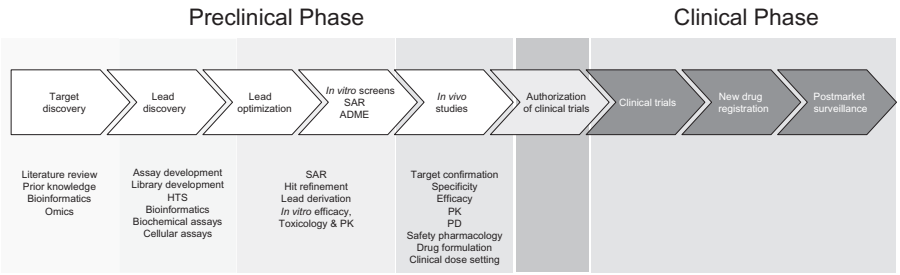


**FIGURE 1 The key stages of drug discovery and development.** A typical series of methods and strategies uses preclinical phases. Note that some of the studies may not be required and the process can be iterative. Refer also to Fig. 2 for a more detailed description of toxicity testing planning.

## 1.2 Drug Development Models

An essential part of drug development is the selection of the most appropriate animal, *ex vivo*, *in vitro*, or *in silico* systems, to allow the collection of information that can be interpreted in terms of the effects of a new therapeutic agent in humans or in one or more subpopulations of humans. There are several deciding factors that guide model selection. During early drug discovery screening, the main consideration is whether the chosen model can cope with large libraries of potentially bioactive molecules. It is generally accepted that, while nonanimal models generally lack the sophistication of studies on vertebrate animals and are based on nonclinical endpoints, they are a useful means of filtering out poor candidates during early drug discovery. The possibility of false hits during this stage is accepted as a trade-off, but it is also recognized that data from the use of several techniques and prior information can assist with the weeding out of false hits. The drug development process involves a more extensive evaluation using *in vitro* and *in silico* approaches and preclinical studies in vertebrate animals on a limited number of potential therapeutic agents.

The drive toward the use of systems biology approaches that take into account the roles of multiple biological and physiological body systems earlier in the drug development process has prompted a dramatic change in the way that data from cell-based studies are used. In many instances, data from several tests can be assembled and analyzed by using *in silico* models to gain a systems biology overview of drug ADMET and activity. Advances in comparative genomics have also opened up the scope for using zebra fish (*Brachydanio rerio*) and invertebrate organisms, such as nematode worms (*C. elegans*) and the fruit fly, *Drosophila melanogaster*, during the early stages of drug development. Likewise, advances in information mining, bioinformatics, data interpretation, the omics technologies, cell culture techniques, and molecular biology have the potential to greatly enhance the drug development process. Ironically, up to now, few of these methodologies has been standardized, formally validated, and accepted for regulatory use. Indeed, *in vitro* data are generally considered supplementary to animal data, rather than as an alternative source of information that is useful and applicable in its own right. Nevertheless, *in vitro* approaches provide information about the mechanisms of action

**TABLE 1 Terminology and Abbreviations**

| Term | Definition |
|---|---|
| 2D heteronuclear NMR | Two radionuclides are used to construct a two-dimensional map of a binding site by NMR. |
| Agglomeration | The process of particle attraction and adhesion. |
| Algorithm | A set of rules to assist with problem solving. |
| Allometric scaling | The process by which size, blood volume, and anatomical features of an organism are taken into account during extrapolation of information from animals to humans. |
| Analogue-based minimization | The process of using information about variants of the natural ligand for a target to derive a minimum number of features required of a smaller substance, so that binding affinity, efficacy, and/or specificity for the target in question are retained. |
| Antisense | A piece of genetic material that is the exact opposite of the natural messenger RNA that encodes a potential protein. |
| Bioaccumulation | The buildup of a drug or its metabolite(s) in a particular tissue or cell type. |
| Bioavailability | A measure of the amount of an administered drug that reaches its intended site of action. |
| Bioinformatics | The management and analysis of information, in order to use computer-based processes to understand biological events. |
| Biokinetic | Describes the key physiological processes that follow the exposure of an organism to a chemical or drug. |
| Biomarker | A molecular indicator of a biological event. |
| Biotechnology product | Replacement therapeutics or recombinant protein or DNA products isolated from or produced by using GM animals, cell cultures, plants, or microorganisms. |
| Biotransformation | The process by which a substance is chemically or functionally modified within the body, which usually involves the action of specific enzymes. |
| Combinatorial library | Large libraries of chemicals generated by a combination of acquisition and understanding of the requirements for recognition of a particular target. |
| Comparative genomics | The study of human genetics by reference to the genetics of other organisms as a means of deciphering human gene organization and function. |
| Cytotoxicity | A measure of the ability of a substance to damage or kill a cell. |
| Decision tree | A support tool for selection among competing choices and their possible consequences. |
| DNAzymes | A DNA-modifying enzyme. |
| Drug mimetic | A drug or drug-like molecule with a structure or modulatory activity that resembles that of a substance found within the body. |
| Druggable genome | The sum of the genes, their encoded disease-related proteins, or gene expression regulatory elements, which can functionally be modulated by drugs and drug-like molecules. |

| | |
|---|---|
| Druggable proteins | Proteins that bind drugs with a binding affinity below 10 μM. |
| Drug discovery | The identification of a potential therapeutic agent. |
| Drug development | The progress of a lead from drug discovery toward a marketable drug. |
| Drug-like compound | A compound that has a molecular weight typical of a drug (around 500 daltons) and a structure that indicates it may have pharmaceutical properties. |
| Efficacy | The capacity of an agent to cause the desired biological effect. |
| Endpoint | The measurable effect of a substance on a biological system. |
| Epitope | The recognition site on a molecule for a particular molecule or class of molecules. |
| Eukaryotic | Describes organisms whose cells possess a nucleus and other membrane-bound vesicles, including fungi, plants, and animals. |
| *Ex vivo* | Literally, "out of the living"—used to refer to experiments that are conducted on tissues or cells isolated directly from a living organism. |
| Gene silencing | The process of preventing a gene from being expressed. |
| Genome | The entire genetic makeup of an organism. |
| Genomics | The study of the genetic makeup of an organism. |
| Genotoxicity | The adverse effects of a substance on the genetic makeup of a cell or organism. |
| Glucuridonation | The process of conjugating the uronic acid of glucose to substances, to detoxify or inactivate them. |
| Hapten | A substance that must combine with a carrier, in order to induce specific antibody production. |
| Hematotoxicity | The adverse effects of a substance on blood cells or on the cells or processes that produce specific types of blood cells. |
| Hit | The product of the high-throughput screening of large libraries of drug-like compounds, fragments, peptides, or proteins, identified by predominantly one-shot affinity, activity, or *in silico* methods. |
| Homeobox | DNA sequences found throughout the genome of most organisms that regulate gene expression, particularly during early development. |
| Homolog | A molecule with corresponding structures or functions in two or more species. |
| Humanized | The product of a process that is aimed to confer more human-like properties on a molecule, cell, or living organism. |
| Hydrophobicity | The tendency of a molecule to repel or exclude water molecules. (Means the same as lipophilicity.) |
| Immunogenicity | The ability of a substance to stimulate an immune response. |
| Immunohistochemistry | The testing of the ability of a tissue to be stained with an antibody. |
| Immunoprecipitation | The ability of an antibody–molecule complex to pull a second molecule out of solution as a result of interactions between the antibody recognizing molecule and secondary molecule. |
| Indels | Insertional or deletion mutations in DNA. |
| *In silico* | Using computer-based methods and virtual systems. |
| *In vitro* | Literally, "in glass"—used to refer to maintenance of tissues, cells, or cell fractions outside the body from which they were derived. |

**TABLE 1** ***Continued***

| Term | Definition |
|---|---|
| *In vivo* | Literally, "within the living"—used to refer to experiments conducted on intact living organisms. |
| Isozyme | Variants of enzymes that catalyze the same reaction(s) but differ from each other in primary structure and/or electrophoretic mobility. |
| Karyotype | The chromosomal complement of an organism. |
| Lead compound | A compound identified by hit generation that has suitable physicochemical and functional properties to serve as a starting point for the development of a potentially marketable drug. |
| Lipophilicity | The affinity of a molecule for a lipophilic environment. |
| Log *P* | The octan-1-ol/water partition coefficient—used to express lipophilicity. |
| Macroparticle | Particulate matter of a crystalline nature, generally exceeding a 10 nm diameter. |
| Margin of safety (MOS) | A ratio of the maximum amount of a substance that causes no effect in animals and the actual exposure (intended or otherwise) of the human population. |
| Meta-analysis | A statistical process for combining information from different sources. |
| Metabolic competence | The ability of a system to metabolize. |
| Metabonomics | The study of metabolic responses to drugs and chemicals. |
| Microfluidics | Small-scale systems comprised of chambers connected by a fluid matrix. |
| Molecular dynamics | Computer simulations of the movement of atoms, based on changes in the energy required to maintain certain conformations. |
| Monte Carlo simulation | A statistical method for studying systems, especially those with large numbers of coupled degrees of freedom. |
| Nanomedicines | Therapeutic agents based on the use of nanoparticles. |
| Nanoparticle | A microscopic particle with a unit size not exceeding 100 nm. |
| Oligonucleotide | A short stretch of synthetic DNA. |
| Omics | Technologies relating to the study of the genome, proteome, or metabolic responses of cells, tissues, and organisms. |
| Organotypic | An *in vitro* system designed to preserve or reconstitute the 3D structure of a tissue or organ, to mimic the *in vivo* situation. |
| Patch clamping | A process for measuring electrical activity across a living membrane by using electrodes. |
| Permeability | The ability to cross a living membrane. |
| Phage display | A system whereby a protein is displayed on the surface of a bacterial virus (a bacteriophage). |
| Phagocytic | Describes the engulfing of a molecule, a microorganism or part of an organism, by leukocytes (a type of white blood cell). |
| Pharmacokinetic | Describes the uptake, biotransformation, and distribution of a pharmaceutical agent and its metabolites in the tissues, and their subsequent elimination. |
| Pharmacophore | A collection of electrical and molecular features that define interactions between a molecule and its binding site on its target. |

| | |
|---|---|
| Plasma clearance rate | The speed at which a substance is removed from the blood. |
| Polymorphisms | Genetic differences within the population that occur for a given gene at a frequency of 1% or more. |
| Posttranslational modification | The process by which a protein is altered after it is synthesized, by the additional or removal of specific moieties. |
| Potency | The comparative ability of a drug to induce the desired effect. |
| Prokaryote | Cellular organisms that lack a distinct nuclear membrane or membrane-bound organelles (e.g., bacteria). |
| Proteome | The total protein complement encoded by the genome of an organism. |
| Proteomics | The study of protein expression patterns in specific cells, tissues, or organisms. |
| Quantum dot | Nanocrystals comprised of a semiconductor metal core. |
| Reactive oxygen species | Oxygen radicals or super-radicals that are capable of causing cellular damage. |
| Recombinant DNA technology | The process of DNA manipulation in an artificial environment. |
| Redox | The process of loss of oxygen or gain of hydrogen by one molecule accompanied by the gain of oxygen or loss of hydrogen by another molecule. |
| Reporter gene | A gene that is expressed in response to an upstream biochemical event, which can be used to monitor that event. |
| Reverse pharmacology | The screening of a library of compounds against one particular target to identify a lead for drug development. |
| RNA aptamers | RNA-based molecules that bind to enzymes. |
| RNAi | RNA interference—process of silencing or dampening protein expression. |
| Signal-to-noise ratio | Measure of the signal strength (change being observed) against the background within an experiment. |
| Therapeutic agent | A chemical, protein/peptide, DNA, stem cell, natural product, or biotechnology product that forms the active component of a finished pharmaceutical product. |
| Therapeutic index | The ratio between the toxic dose and the therapeutic dose of a drug, which is related to the MOS. |
| Three Rs | The principles of *replacement* of animal experiments, *reduction* of the number of animals used in a given study, or *refinement* of the procedures used, in order to minimize suffering and distress. |
| Toxicogenomics | The use of genomics and bioinformatics to identify and characterize mechanisms of action, based on changes in gene expression as monitored by the production of mRNA transcripts. |
| Toxicokinetic | Describes the uptake, biotransformation, distribution, and effects of a directly or indirectly toxic substance and its metabolites in the tissues, and their subsequent elimination. |
| Transcription | The process of messenger RNA production from a gene. |
| Transgene | A gene or variant of a gene that is inserted into the genetic makeup of an organism. |
| Vector DNA/RNA | Carrier DNA/RNA that may also facilitate the expression and/or cellular uptake of foreign genetic material by cells and tissues. |
| Xenobiotic | A chemical or other substance that is not a natural component of the makeup of the organism exposed to it. |

**TABLE 1** ***Continued***

| Abbreviation/Acronym | Full Name |
|---|---|
| ADME(T) | absorption, Distribution, Metabolism, Elimination (Toxicity) |
| ADR | Adverse Drug Reaction |
| BBB | Blood–Brain Barrier |
| BCS | Biopharmaceutics Classification System |
| BRET | Bioluminescent Resonance Energy Transfer |
| cAMP | Cyclic Adenosine Monophosphate |
| CASE | Computer Automated Structure Evaluation |
| CBER | Center for Biologics, Evaluation and Research |
| CDER | Center for Drugs, Evaluation and Research |
| CoMFA | Comparative Molecular Field Analysis |
| COMPACT | Computerized Optimized Parametric Analysis of Chemical Toxicology |
| CRE | Cyclic-amp Responsive Element |
| CYP | Cytochrome P450 |
| CYP450-DMO | Cytochrome P450-Dependent Monooxygenase |
| DEREK | Deduction of Risk from Existing Knowledge |
| ECVAM | European Centre for the Validation of Alternative Methods |
| ELISA | Enzyme-linked Absorbance Assay |
| EMEA | European Medicines Agency |
| EPA | Environmental Protection Agency |
| ERE | Estrogen Responsive Element |
| FACS | Fluorescence-Activated Cell Sorting |
| FDA | Food and Drug Administration |
| FRET | Fluorescent Resonance Energy Transfer |
| GFP | Green Fluorescent Protein |
| GPCR | G-Protein-Coupled Receptor |
| HESC | Human Embryonic Stem Cell |
| HTS | High-Throughput Screening |
| IAM | Immobilized Artificial Membrane |
| ICCVAM | Interagency Coordinating Committee on the Validation of Alternative Methods |
| ICH | International Conference for Harmonization |
| IND | Investigational New Drug |
| LC-MS/MS | Liquid Chromatography and Tandem Mass Spectrometry |

| | |
|---|---|
| $LD_{50}$ | Lethal Dose that kills 50% of a test group (of animals) |
| Log $P$ | Octan-1-ol/water partition coefficient |
| MAP kinase | Mitogen-Activated Protein Kinase |
| MCASE | Multi-CASE |
| MDCK | Madin–Darby Canine Kidney |
| MOS | Margin of Safety |
| MS | Mass Spectrometry |
| MTD | Maximum Tolerated Dose |
| NCE | New Chemical Entity |
| NCTR | FDA National Center for Toxicological Research |
| NMR | Nuclear Magnetic Resonance |
| OECD | Organization for Economic Co-operation and Development |
| PAMPA | Parallel Artificial Membrane Permeation Assay |
| PBPK | Physiologically Based Pharmacokinetic |
| PCR | Polymerase Chain Reaction |
| PK | Pharmacokinetic |
| $pK_a$ | The acid-ionization constant |
| PTFE | Polytetrafluoroethylene |
| QSAR | Quantitative |
| QSAR–ES | Quantitative Structure–Activity Relationship—Expert System |
| QT Interval | The time between the start of the Q wave and the end of the T wave in the heart's electrical cycle |
| SAR | Structure–activity Relationship |
| SPA | Scintillation Proximity Assay |
| SPR | Surface Plasmon Resonance |
| TOPKAT | The Open Practical Knowledge Acquisition Toolkit |
| UFAW | Universities Federation for Animal Welfare |

of a drug that is vital for the design of *in vivo* animal studies and can add substantial weight to the product dossier submitted to regulatory bodies.

Increasingly, predictions about the ways in which a particular chemical is likely to interact with its desired cellular target are made by undertaking *in silico* modeling. These results are used to filter out poor candidate molecules according to chemical class and structural or functional features during drug discovery. However, filtering of this kind is sometimes impossible, so lead identification still relies to some extent on serendipitous finds from random libraries, rather than on rational lead discovery. For instance, for new chemical entities (NCEs) for which there are no data, i.e., are first-in-class, *in silico* screenings are difficult to handle, particularly where there is also limited knowledge of the structure of the active site of the target. Also, there might be a lack of important information for other compounds. For example, predicting drug effects can be seriously compromised when ADME data on the behavior of a molecule in different tissues and species are lacking. This is confounded by the reality that this kind of information for different individuals will always be limited. Both of the above situations are most evident in the case of large molecules, such as (1) peptides and proteins with complex structures and multiple conformations, (2) humanized products that could be differentially immunogenic in different species, and (3) nanoparticle formulations.

### 1.3 Information Required Prior to Drug Authorization/Approval

Once a new therapeutic candidate has been successfully identified from preclinical studies, the next stage involves the authorization of clinical studies. The information required prior to the authorization of any clinical trial is crucial for the design and execution of preclinical studies, irrespective of whether the aim is to define drug action or provide safety information. Such information includes (1) manufacturing quality, (2) physicochemical properties, (3) efficacy, (4) proposed mechanism of action, (5) selectivity, (6) ADME, and (7) possible adverse effects in humans.

In the United States, the Food and Drug Administration (FDA) handles drug approvals. The FDA has fast tracked this process for treatments for serious diseases where no therapies currently exist [2]. Drug developers are required to submit an Investigational New Drug (IND) Application, in which evidence from preclinical studies is provided for review by the FDA. The FDA decides whether it is reasonably safe for the company to test the drug in humans. Under the FDA's jurisdiction, the Center for Drugs, Evaluation and Research (CDER) and the Center for Biologics, Evaluation and Research (CBER) are responsible for reviewing different types of therapeutic agent applications (Table 2). Note that these changes in jurisdiction mean that biological products, the testing of which was at one point based on limited animal tests (because of their poor predictivity), are likely to require more stringent testing under the CDER [3].

The FDA has exclusive executive control over decisions regarding drug approvals in the United States. However, in Europe, it is possible to have a drug approved by a number of different routes. This is because companies can apply either via the EMEA (European Medicines Agency) for pan-European approval or via one or more national agencies. However, since November 2005, all new drugs for the major diseases, including AIDS, cancer, diabetes, and neurodegenerative disorders, and

**TABLE 2 CDER and CBER:[a] Review of New Therapeutic Agent Applications**

**CDER**

- Traditional small molecule therapeutics
- Growth hormone, insulin, and other endocrine peptide therapeutics
- Monoclonal antibodies
- Proteins (e.g., cytokines, enzymes, and other novel proteins), except those specifically assigned to the CBER, namely, vaccines and blood products that are assigned to CBER
- Immunomodulatory agents (but not vaccines)
- Growth factors intended to modulate hematopoiesis *in vivo*
- Combination products where the primary mode of action is that of an agent assigned to the CDER

**CBER**

- Products composed of human, bacterial, or animal cells or fragments of cells, for use as preventative or therapeutic vaccines
- Gene therapy products
- Vaccines
- Allergenic extracts used for the diagnosis and treatment of allergic diseases
- Antitoxins, antivenins, and venoms
- Blood and blood products from humans or animals
- Combination products where the primary mode of action is that of an agent assigned to the CBER

[a]The CDER and CBER are afforded jurisdiction by the U.S. FDA.

medicinal products developed by means of biotechnological processes must be approved via the EMEA.

With the globalization of the pharmaceutical industry, the International Conference on Harmonization (ICH) guidelines have, since 1990, set out to standardize drug applications in terms of their content and format. Japan, the United States and the European Union (EU) comply with these requirements for the quality, safety, and efficacy assessment of new drugs. These guidelines operate alongside national requirements. Quality assessment guidelines are provided to standardize the assessment of drug stability (shelf-life), and the management of risks due to impurities, such as residual solvents and infectious agents, such as viruses (which can be present when a drug is isolated from plants, animals, humans, or cell lines). The guidelines also require the standardization of cell lines, test procedures, acceptance criteria, and procedures for formulation and development. Efficacy guidelines are also provided, to standardize the conduct, interpretation, and reporting of clinical trials.

There are some important practical considerations that should be borne in mind when conducting preclinical studies. The most comprehensive guidelines are those provided for drug safety testing, which cover a number of toxicological endpoints, including carcinogenicity, genotoxicity, reproductive and developmental toxicity, and immunotoxicity. Some of the guidelines apply generically to all new drugs, while others focus on specific types of therapeutic agents, such as biotechnology products. These guidelines are essential reading for researchers engaged in drug development and are considered in more detail throughout the remainder of this chapter.

Another important source of reference is the Organization for Economic Co-operation and Development (OECD). By ratifying the convention of the OECD, many European countries, Australia, Japan, New Zealand and the United States have agreed to abide by a set of test guidelines for assessing the human health effects of chemicals [4], which apply equally to the testing of therapeutic agents. Later, we refer to a number of nonanimal methods and refinements of animal procedures accepted by the member countries of the OECD.

## 2 FINDING NEW DRUG TARGETS

### 2.1 Background

Until relatively recently, drug development focused on a limited number of targets, against which NCEs with a desired effect could be selected. These "druggable" targets were once most extensively investigated by using animal models. However, greater access to recombinant DNA technology means that most early screens are now conducted primarily by using different genetically engineered cell lines expressing putative targets that can be arrayed into high density plastic plate formats suitable for interactions between the targets and potential lead chemicals (for methods, see later discussion).

Overington et al. [5] derived a consensus figure for the number of therapeutic drug targets for the FDA-approved drugs that were available in 2005. They identified 324 drug targets for all classes of approved therapeutic agents, which were targeted by in excess of 1357 drugs, of which 1204 were small molecules and 166 were biologicals. Cell surface receptors and channels represented the targets for >50% of all the FDA-approved drugs. A further 10% of the drugs, including monoclonal antibodies, also target other cell surface proteins. Most of the remaining targets were enzymes, nuclear receptors, DNA, or ribosomes. These targets represent a minute fraction of the genome, and a mere 3% (266 proteins) of the predicted proteome.

According to this survey, on average 5.3 new druggable targets are discovered each year. This means that many more potential drug targets remain to be discovered. Whether a potential drug target will be a good therapeutic target, however, depends on whether (1) it plays a key role in gene regulation, (2) it is selectively expressed in certain disease states or tissues, and (3) it has a definable and unique binding site.

Often, a further important piece of information is the nature or identity of the endogenous modulator. For example, >1000 G-protein-coupled receptors (GPCRs) have been cloned from various species, including 160 distinct human subtypes with known ligands, although these represent only a limited set of targets for current therapeutic agents. A further 100 or so are orphan receptors, for which there is currently no known natural ligand. In such cases, the starting point is the gene, from which the protein receptor can be expressed and used to screen large combinatorial libraries of chemicals in the search for a modulator. Such a reverse pharmacology strategy uses the orphan receptor as a "hook" for screening libraries and hit generation, where little is known about the natural ligand. In many cases, receptor models use the crystal structure of rhodopsin as a template, as this is the only GPCR whose structure has been resolved. The importance of GPCRs is emphasized by the fact that, although >20% of the top 200 current best-selling drugs interact with these cell

surface receptors, they generate worldwide sales of drugs such as cimetidine, losartan, and ropinerole of over $20 billion (U.S.) [6].

### 2.2 Impact of New Technologies on Target Discovery

Comparative genetics can provide much relevant information, particularly with regard to the role of human-specific genes and the suitability of animal models for drug development. The application of microarray techniques, standards, and resources that permit the comparison of gene expression patterns across species and between cell types and tissues has started to provide some insight into the metabolic and biochemical differences between health and disease states. A good example of this is the Cancer Genome Anatomy Project (www.ncbi.nlm.nih.gov/CGAP) [7], in which mutational sites in cancer cells have been identified.

A cursory examination of the 373 completed genome sequences for archeal, prokaryote, and eukaryote [8] species suggests that, although genome size increases from archea through prokaryotes to eukaryotes, genome size is not directly linked to the number of genes within the functional genomes, nor with evolutionary status. It is, however, clear that, as the complexity of organisms increases, so does the complexity of gene regulation and the level of genetic redundancy—the ability of several genes to rescue loss-of-function of another gene. Nevertheless, for highly conserved genes, such as those that are involved in early development, and homeobox genes, studies on early life stages of species such as zebra fish and invertebrate models can indicate the roles of genes. However, in general, such studies are more relevant to safety pharmacology than to mechanistic and efficacy studies. It is worth bearing in mind that computational predictions and statistical analyses have suggested that the bacterial *Escherichia coli* and human genomes account for 35 common metabolic pathways, namely, those that are important in biosynthesis and in degradation and respiratory processes [9], and that, possibly as a result of bacterial infection, a number of bacterial genes have become permanently integrated in the human genome [9, 10]. This opens up the possibility of using bacterial studies to decipher a limited number of biochemical pathways affected by drugs, as well as for genotoxicity testing.

Unicellular eukaryotes, such as yeast, share remarkable genetic and functional similarities with multicellular eukaryotes. The most useful yeast strain in terms of dissecting protein and gene interactions is *Saccharomyces cerevisiae*. At 12,100 kilobases, the *S. cerevisiae* genome is much smaller than the human genome. However, because its gene density is 50 times greater than that of the human genome, genes found in the *S. cerevisiae* genome resemble around 30% of the genes associated with diseases in humans [11]. Since the entire genome of *S. cerevisiae* encodes no more than 6000 proteins, it is relatively straightforward to investigate gene function in yeast and make genome-wide microarray measurements. Such data, together with information from other sources, have made it possible to identify a number of putative drug targets [12] and protein–protein interactions [13], thereby facilitating the development of extensive maps of protein and gene interactions. Such studies in *S. cerevisiae* have been particularly useful in neurodegenerative and ageing research and in studies on diseases that arise as a consequence of mitochondrial DNA damage. One example is the observation that yeast mutants for α-synuclein result in a large change in yeast sexual reproduction, as well as causing cytotoxicity,

both endpoints of which are suited to high-throughput screening assays for new treatments for Parkinson disease [14].

Subsequent studies on yeast-based models of Parkinson disease have suggested that there is substantial scope for using yeast for the high-throughput screening of chemicals for drug discovery [15]. For example, *S. cerevisiae* possesses three distinct G-protein-coupled receptors (GPCRs), which are involved in pheromone (Ste2 and Ste3 receptors) and glucose sensing (Gpr1) [16]. These receptors are related, albeit to a limited extent, to the vastly expanded human GPCR repertoire. By coupling heterologously expressed human GPCRs to the yeast MAP kinase pathway (associated with yeast mating and growth arrest), in yeasts where the MAP kinase pathway is linked to reporter gene expression [17], it is possible to monitor receptor recognition and activation by simple growth or colorimetric reporter assays.

*Caenorhabditis elegans* is another organism that can be used in early drug discovery. This nematode worm is transparent, has a short life span, is a mere 1 mm in length and 80 μM in diameter, reproduces every 3 days by self-fertilization to produce over 300 offspring, and is a multicellular organism composed of exactly 959 somatic cells. It displays many of the basic features of higher eukaryotes, including the possession of muscle, excretory cells, and neural cells, and has been extensively used to increase understanding of the mechanisms of gene regulation and gene function. Antisense knock-out or knock-down of gene expression can be achieved simply by feeding the worm with *E. coli* bacteria transformed with plasmid DNA containing antisense DNA. More recently, RNA interference (RNAi) has been used to manipulate the genomes of organisms such as *C. elegans*, although the possibility of transmission of RNA silencing to subsequent generations can occur [18]. Like all multicellular organisms, *C. elegans* exhibits programmed cell death (apoptosis) [19], in a way that is very similar to that seen in higher organisms as part of ageing and disease processes. Similarities between the signaling pathways involved in the regulation of cell proliferation in *C. elegans* and humans suggest that this organism might provide information on the regulation of cell proliferation, which will be of relevance to cancer therapeutics. The entire 302-cell nervous system of this worm has been mapped by electron microscopy, and although the average human possesses somewhere in the order of 100 billion neurons, it seems that neurotransmission is similar in the two species. Thus, *C. elegans* possesses the major classes of ion channels, receptors, transporters, and neurotransmitters that make it a suitable candidate for some forms of drug screening, such as the discovery of new dopaminergic drugs. Similarly, *D. melanogaster* shares much of its basic neurobiology with higher organisms, including humans. It possesses the same neurodegenerative states, neurotransmission mechanisms, and receptor homolog that are found in humans as key targets for neurally active therapeutic agents, making studies with these organisms useful for the development of treatments for conditions such as Parkinson's disease [20].

### 2.3 Data Mining

Novel drug targets can also be found in other ways, including data mining. This involves analyzing the literature, to determine the biochemistry underlying particular human diseases, and human physiology. In addition, human population genetics studies can be undertaken, to determine the roles of human genes, how they interact, the consequences of population differences at the gene level, and, ultimately, the complete physiology of the human body. In the last-named case, since the

possibilities for human studies are limited, most of the information gathered comes from fundamental research that examines modes of interaction of specific substances with any given novel targets, and the modulation of their physiological roles, by combining several approaches, including *in vivo* studies.

The next step is to define whether a newly discovered potential drug target is a feasible target, by identifying the binding site of the proposed molecular target. In this respect, the potential for data mining has been greatly enhanced by the recent development of a druggable-protein database. This can provide information that is useful for deriving rules for the computational identification of drug binding sites. Indeed, there are now algorithms designed specifically for this purpose [21]. Some analyses relate to the identification of pockets within the binding site that serve as potential specific drug targets. However, this approach can be complicated, since the binding pocket that is targeted by an endogenous or natural modulator of target function might include only part of the binding site, or might lack it altogether. A recently described approach to this problem, in which 2D heteronuclear NMR is used to screen drug-like and fragment libraries for interactions with proteins, generates additional reliable data than is obtainable from conventional high-throughput screens. While such information can be used for computational application, including the refinement of protein models, it is limited by the number of protein structures that are currently available. An exception to this are quantitative structure–activity relationships (QSARs) generated by computational techniques such as CoMFA, which rely on molecular descriptors for molecules that are specific for a target, in order to generate a set of conformers that can be used to predict the ability to bind to a protein.

## 3 TRADITIONAL APPROACHES TO DRUG DISCOVERY AND DEVELOPMENT

### 3.1 Hit to Lead

The current attrition rate for NCEs can be gauged from the fact that, on average, for every 7 million molecules screened, only one product is marketed [22]. These odds have resulted in the concentration by pharmaceutical companies on refining, rather than expanding, their chemical libraries and methods. A further important factor that determines the success of early drug screening is the choice of methodologies used to identify hits and to screen potential leads and their derivatives. In this section, we describe the key stages and methodologies used for hit generation, hit confirmation, lead, identification and lead characterization (Table 3).

Before 1980, nearly all drugs were small molecules of around 50 to 1000 times smaller than the size of a typical protein at around 500 daltons, or smaller. Extensive combinatorial libraries of small molecules are generated in-house by all large pharmaceutical companies, often by diversity-oriented synthesis, in which small molecular building blocks are randomly combined in all possible spatial orientations. Screening libraries can consist of thousands of chemicals and rely on an appropriate hit generation and lead characterization strategy. The chemicals concerned must meet certain purity, molecular weight, lipophilicity (log $P$), and functional conformer criteria.

Schreiber [23] first used diversity-oriented synthesis to generate bead-attached libraries of target-oriented and diversity-oriented chemicals. This approach involves

**TABLE 3 Key Methods Used During Hit Generation and Lead Optimization**[a]

| Methods | Assay Principles | Advantages | Limitations |
|---|---|---|---|
| *Affinity-based biophysical methods* | | | |
| Mass spectrometry | Relies on the affinity of a compound for a protein to cause mass/charge shifts. | Can handle large drug-like/ fragment mixtures. | Not truly an HTS platform; poor at resolving mixtures; false hits. |
| NMR | Monitors the location of radionuclides in the target–ligand complex and is used to probe the active site of folded/*in situ* proteins/DNA. A number of new higher resolution techniques (e.g., magic angle spinning NMR) do not require high purity target proteins. | Provides structural information for *in silico* platforms; suited to screening large fragment libraries. | Does not provide SAR data; false hits; weakly potent fragment hits are poorly detected. |
| X-ray crystallography | X-ray diffraction by crystallized protein/ protein–ligand complexes. | Provides structural information; HTS platform. | Weakly potent fragment hits are poorly detected; erroneous assumption about structural similarity can lead to some compounds being discarded; there are not crystal structures available for all target proteins. |
| *Biochemical screens* | | | |
| Scintillation proximity assay | Monitors energy transfer changes as an indicator of binding interactions. | Provides kinetic data | High background; limited plate format; not easily correlated to physiological effect. |
| Radiometric binding assays | Uses radioactive tracing of target–tracer/molecule interactions. | Direct measurement of binding interactions; adaptable for a wide range of possible target-based screens. | Relatively expensive to generate suitable tracer; health and safety considerations; not real-time measurements. |

**TABLE 3** ***Continued***

| Methods | Assay Principles | Advantages | Limitations |
|---|---|---|---|
| SPR | Commonly based on the target being immobilized on a chip and the compound mixture being passed over it. Interactions are monitored as an electrical readout. | Permits kinetic measurements; can be used to identify hits from complex mixtures. | Chip preparation and availability; requires relatively large amounts of materials; more suited to detailed mechanistic studies than HTS. |
| Nonradioactive assays | Includes colorimetric/absorbance-based assays (such as ELISA), luminescence-based assays, and fluorescence-based assays (e.g., FRET, real-time fluorimetry, fluorescence correlation spectroscopy), as generally used in conjunction with cell-based assays (see below). | Generate quantitative data suited to SAR; can give real-time data; can provide mechanistic information; suitable to HTS formats. | Often more suited to later stages of lead discovery. |
| *Cell-based assays* | | | |
| Reporter gene assays | Involves the use of genes such as those encoding GFP, luciferase, and β-galactosidase coupled to a biochemical pathway modulated by a substance to monitor the extent or modulation. | Generates quantitative data suited to SAR; minimum resources needed. | Not truly HTS; can give equivocal data; false hits; not well suited for fragment screens. |
| FRET | Monitors energy transfer between a fluorescent energy donor and acceptor as a measure of the proximity between the two groups, commonly found on the target and a tracer. | Suitable for high density formats; provides mechanistic information; broad range of applications; real-time monitoring of interactions. | High incidence of false hits; prone to fluorescence quenching. |

**TABLE 3** ***Continued***

| Methods | Assay Principles | Advantages | Limitations |
|---|---|---|---|
| BRET | Similar principles to FRET. | Suitable for medium density formats; provides mechanistic information; suitable for monitoring protein–protein interactions; real-time monitoring of interactions. | Some limitations on application; involves protein engineering of the target. |
| Reporter gene | Based on recombinant protein engineering and expression technology to couple an endogenous pathway to the expression and/or activity of a protein from a transgene in response to drug modulation of a target. | Several commercially available plasmids (e.g., with cAMP, calcium, and estrogen responsive elements); sensitive high throughput assay formats. | High incidence of false hits; long incubation times; indirect correlation with target modulation. |
| Electrical readout | Includes biosensor-based methods and patch clamping. | Suitable for monitoring channel activity. | Not truly suited to HTS; limited utility |
| Second messenger assays | Based on a direct measurement of one of more downstream changes in signal mediators in response to drug modulation of a target. Includes assays such as those that measure changes in intracellular calcium (FLIPR/ Aequroscreen), cAMP, and many more. | Direct measurement of the effects of a substance. | Only suited to some types of targets (e.g., receptor, channels, enzymes); time consuming. |

**TABLE 3** ***Continued***

| Methods | Assay Principles | Advantages | Limitations |
|---|---|---|---|
| Fluorophore and chromophore-based methods | Rely on the use of an ion-sensitive dye to detect intracellular changes in ion content. | Suitable for monitoring increases in intracellular calcium, potassium, and sodium ions. | Sensitivity dependent on dye chemistry. |
| Cell proliferation assays | Includes methods such as dye or radioisotope uptake, protein estimations, cell counting, and oxygen sensor measurements to monitor the competence, viability, and growth rate of cells. | Minimum resources needed; generic application; quantitative data can be obtained. | Difficulty equating to physiological endpoint. |
| *In silico methods* | | | |
| Protein modeling | *Ab initio* or homology-based protein structure modeling based on amino acid sequence analysis and biophysical/biochemical data. | Binding site identification and pharmacophore modeling. | Need experimental confirmation of findings. |
| Molecular docking/SAR/combinatorial chemistry | Molecular dynamics simulations and energetic calculations. | Virtual screening prior to chemical synthesis. | Need experimental confirmation of findings. |
| PBPK modeling | Mathematical prediction of the fate of a drug. | Can be used to identify the sites of action of a drug and to estimate likely internal dose. | Reliant on large amounts of data; can involve considerable mathematical expertise. |

[a]A number of different approaches are used during drug discovery and development. Here, a list of methods applicable to hit generation and lead development are listed alongside the main advantages and limitations of each method or group of methods. HTS, high-throughput screen.

the use of fragments—small chemicals—of around 120–250 kDa. Generally, these fragments display lower (10 μM to millimolar) affinities for a target than do more complex, drug-sized chemicals (affinities within the nanomolar range). It is therefore necessary to complement fragment screens by using sensitive analytical techniques, such as protein-detected or ligand-detected NMR [24], MS [25], X-ray crystallography [26], and SPR [27] (although the last named is generally more applicable for hit confirmation; see later discussion). These techniques are preferable to

bioassays, such as cell-based binding or functional assays, or to the step-wise combination of hit fragments either by chemical synthesis or by combining pharmacophores [28]. Despite the fact that the method used to screen fragments affects the success of such screens, the hit rate for fragment-based lead discovery is substantially higher than that for drug-like screens, there being an apparent inverse relationship between chemical complexity and target complementarity. Indeed, a screen of <1000 fragments might identify several useful hits for lead development.

A "library in tube" method is being developed for large mixtures of chemicals, which has been adapted from a concept put forward by Brenner and Lerner in 1992 [29]. This technique involves coding each chemical with a DNA tag, in order to identify the attached chemical by PCR, such that mixtures of chemicals can be panned against a target. This approach has much potential for diversity-oriented hit generation (see Ref. 30 for a review).

Biochemical screening can be performed by using several types of readout, including those reviewed in Ref. 31. Whatever the assay used, it should display good signal-to-noise ratios and should also be reproducible. The two most commonly used screening formats are radiometric and nonradiometric assays, both of which are suitable for intact cell or tissue-based studies. Radiometric assays include filtration-based methods, where the unbound radioactive probe (generally the radioligand specific for the target) competes for ligand binding with the unlabeled screen compound, after which it is removed in readiness for scintillation counting, or for scintillation proximity assays (SPAs), where β-particle emissions from isotopes with short β-particle path lengths (namely, $^{3}H$ and $^{125}I$) are measured *in situ* by using scintillant-impregnated microspheres. The amount of reduction of the radiolabel signal intensity due to competition is measured. The use of the former isotope renders the method amenable to a 384-plate format, while the latter is generally more suited to a 96-well format.

Nonradiometric assays include those based on colorimetric, fluorescent, luminescent, or electrical changes. Commonly used methods include proximity-based fluorescent resonance energy transfer (FRET), which can be used to monitor interactions between a fluorescent donor and an acceptor on the target, and to screen chemicals. This technique is suited for both monitoring a wide range of molecular interactions and to 1536-well formats. One example of how FRET may be useful is in the screening of enzyme inhibitors [32]. The drawbacks of this method are the high incidence of false positives and problems with fluorescent quenching. Bioluminescent resonance energy transfer (BRET) is another proximity-based screen. This method, while being prone to quenching, requires the use of proteins such as *renilla reinformis* luciferase donor and green fluorescent protein (GFP) acceptor, in the presence of coelenterazine a (luciferase substrate). BRET is generally more useful for screening interactions between large molecules, such as proteins, due to the bulky nature of acceptor and donor groups, luciferase, and GFP. Nevertheless, it can also be used to screen for chemicals that perturb such interactions, and indeed, BRET has been proposed as a screen for HIV-1 protease inhibitors [33]. The sensitivity of both FRET and BRET is dramatically improved when there is a large difference between the emission spectra prior to and following energy transfer from the donor to the acceptor group.

Other commonly used screens rely on the expression of a reporter gene (e.g., β-galactosidase or luciferase) in response to the activation of a specific pathway.

However, many more screening techniques are specific for the targets in question, as is the case for GPCRs [34] and HIV-1 [35]. An example of the usefulness of electrical readouts is the examination of the interaction between DNA and metallocompounds. In this case, the DNA is immobilized on electrodes, and interactions with the drug can alter the electrical output [36]. Generally, these functional assays (with the notable exception of SPA) can provide a mechanistic overview of drug action. However, further insight can be gained by using surface plasmon resonance (SPR). SPR is a real-time monitoring system based on change in mass, in which microgram amounts of the target are immobilized on a chip and exposed to the test chemical. The flow rate and wash rate can be varied, such that not only can the individual chemicals in a mixture be resolved according to rank order of affinity, but also the on–off rates of binding can be monitored. Membrane protein targets, however, are difficult to isolate and refold into the chip matrix, so SPR is far more useful for the screening of drugs that target soluble proteins and DNA [37].

As a typical screen of 1 million chemicals can take 6 months to complete, there is interest in expediting hit generation by using higher density plate formats or by chemical pooling. Increasing the assay density by increasing the well density is feasible, but is highly dependent on the nature of the screen. Chemical pooling involves placing multiple chemicals into each well of a plate, with a single chemical overlap between two wells. This can reduce the screening time to a matter of weeks. However, factors such as the possibility that two of the compounds in the same well will cancel the effects of each other or will act synergistically, can result in false negatives and positives, respectively. It is also general practice to include pairs of structurally related chemicals in each screen.

A new drug can also be developed as a result of rational drug design, particularly when there is extensive knowledge of the structure and function of the target protein, as well as available computer models and the capability to dock virtual compounds into the active site. In many cases, however, the original first-in-class compound was designed by modification of the endogenous ligand for the target. The classical example of this is the design of small nonpeptide antagonists that target neuropeptide receptors (e.g., neurokinin receptors) by gradual structural minimization and constraint of the natural endogenous receptor ligands [28]. In general, the design of these smaller nonpeptide ligands, based on knowledge of the natural ligand, requires extensive peptide analogue generation and screening for efficacy and activity, so as to identify the key interactions and functional groups on the peptide that determine specificity and activity. In the above example of neurokinin receptor binding, the key interactions were identified as being with the terminal Phe-X-Gly-Leu-Met-$NH_2$ motif. Indeed, all ligands that retain neurokinin receptor affinity contain aromatic rings and amine groups that fit into the receptor pocket.

The latter analogue-based minimization of the natural ligand for a target protein is particularly relevant, given that larger molecules such as peptides and proteins are increasingly being investigated as clinical agents. Currently, more than 40 peptides are marketed worldwide, with some 700 more at various stages of development as drug leads. Similarly, there are some 120 antibody-, hormone-, and enzyme-based therapeutics currently on the global market. Many of these therapeutics are more specific and more active than their small molecule counterparts, and they accumulate less readily in tissues, with generally lower oral bioavailability and less stability.

They are all potentially immunogenic and are relatively expensive to manufacture. These molecules are also not generally amenable to rational design strategies and are often developed by *de novo* routes with limited *in silico* approaches, in view of the difficulties associated with docking flexible peptides and proteins into the target protein.

Screening for peptide, polypeptide, and protein therapeutic leads presents a problem, in that large libraries are generally not amenable to chemical synthesis. One solution to this problem is to use systems in which the peptide is linked to the DNA that encodes it. Phage display, for instance, is a technique that allows one or more genes encoding any number of protein variants to be expressed in an anchored form amenable to affinity probing. The genes of interest are inserted into the genome of a nonlytic phage, which is introduced into bacteria. The proteins encoded by the genes are expressed (displayed) on a defined coat protein of the respective phage. Phage display libraries of over a billion different peptide or protein sequences can be prepared, the only limitation being the efficiency with which the bacteria are infected. By using the molecular target as a probe to isolate hits from this library, it is possible to undertake successive rounds of optimization until the most specific hits are identified. Phage display, and the similar, more recent ribosomal display systems [38], can be used to screen for protein and hapten hits for drug development and have proved particularly useful with respect to the development of specific antibodies [39]. However, the need for folded proteins has led to the development of a yeast-display technology, whereby proteins are presented in their folded form on the yeast cell wall. These anchored systems all facilitate miniaturized screening and, in the case of the yeast-display libraries, FACS [40].

The techniques used for developing genetically based therapeutics share some similarities with more traditional drug discovery approaches. Genetically based therapeutics include plasmids containing transgenes for gene therapy, oligonucleotides for antisense applications, DNAzymes, RNA aptamers, and small interfering RNAs for RNAi [41]. So far, two such products have been approved for clinical use and many more are in the course of development, so this important group of therapeutics requires specific consideration in the context of preclinical planning. Very little is currently understood about the suitability of many genetically-based therapeutics. It is known, however, that the design of the vector crucially determines delivery and nuclear uptake, and also that the promoter used will determine the expression levels of the transgene and the efficiency of gene silencing (reviewed in Ref. 41). Since uptake is a key determinant of efficacy, the development of these therapeutic agents must be used together with an evaluation of DNA delivery techniques, such as microinjection, electroporation, viral delivery systems, and carrier molecules that either promote cellular endocytosis (e.g., cationic lipids or amines) or facilitate uptake (e.g., carbon nanotubes) (see Ref. 42 for a review and Section 4.4). Equally, the expression of the encoded DNA is reliant on the precise nucleotide sequence, with codon use often resulting in changes in the expression of the encoded protein product and, in some cases, to its cellular fate.

Whatever the discovery route for a lead compound from drug-like libraries or fragment libraries, it is clear that most of the drugs that are currently marketed are highly similar to the leads from which they were derived [43]. This makes lead discovery a crucial step in the drug discovery process. The most widely used approach to confirming leads is affinity-based screening [44], where qualitative (e.g., rank

order) or quantitative ($K_d$, $IC_{50}$) measurements are used to monitor interactions between compound libraries and protein, RNA, or DNA targets, by using approaches such as standard binding assays, NMR, SPR, or X-ray crystallography. Other approaches involve the use of changes in biochemical events that have been identified from target modulation or predicted by *in silico* screening. A combination of all three approaches has the advantage over using biochemical techniques alone, of reducing the number of false hits while allowing higher screening throughputs. For instance, experimentally based screening may result in false hits, because of (1) nonspecific interactions (predominantly hydrophobic in nature), (2) aggregation or poor solubility of the drug, and (3) purities, reactive groups, or chemical stability that are not readily discernible from *in silico* predictions. MS-based methods result in fewer false positives because of nonspecific hydrophobic interactions, poor solubility, impurities, and reactive functional groups. In practice, however, the method used for hit generation is dependent on the resources available.

In the case of *in vitro* biochemical and cellular assays, miniaturized formats can be used to screen around 1 million drug-like molecules, by using 1–50 μM concentrations and a 30–50% activity cutoff between potential hits and failures [45]. Where fragment libraries are used, activity might only be detectable at substantially higher concentrations, and by using more-sensitive techniques. As a result of these selection criteria, the rate of false hits (and failures) is also relatively high.

Hit confirmation generally involves biochemical assays to confirm that the observed activity is linked to the desired mechanism of action. The choice of methodologies is important, since it is at this stage that eliminating false-positive hits becomes most important and depends on the necessary properties of the final drug. It is also at this stage that hits begin to be ranked according to specificity, activity, and suitability to be used for lead development. Indeed, data from hit confirmation studies are often amenable to structure–function analysis by using *in silico* methods that may ultimately guide decisions as to the most favorable leads.

This process is developed further during the hit-to-lead stage, in which potency is no longer considered to be the deciding factor, but selectivity, the feasibility of chemical synthesis and modification, the mechanisms of target interaction and modulation, pharmacokinetics, and patentability of the final drug have become increasingly important. Many of these issues are considered later. It is important to note, however, that determining whether individual fragment hits fulfill these criteria is much more problematical. The ability to chemically modify a hit lends itself to the three main ways of generating a lead compound from initial promising hits and subsequently derivatizing and modifying the lead to give the final drug, namely, by using biophysical or biochemical methods, cell-based screens, or *in silico* predictions.

It is at the above stage of development that the possible risks associated with a new drug candidate begin to be addressed. The affinity and specificity of the drug candidate for the desired target can often dictate whether it will be discarded at an early stage. For instance, if there is a difference of several orders of magnitude in affinities for selected targets and off-targets, the drug is less likely to have predictable side effects. That is, it is possible that a drug may have a desirable effect within one concentration range, above which it causes toxicity. The relationship between the desired therapeutic effects of a drug and its adverse effects is expressed as a margin of safety (MOS; also referred to as therapeutic index)—being the difference between the effective dose and that which gives rise to toxicity.

Two important sources of information can contribute to a widening of the MOS during lead optimization. The first is a fundamental understanding of the mechanisms of interaction with the desired target and off-targets. The second is information from combinatorial chemistry and rapid *in vitro* screens to determine the relationship between structure and activity, which can then be applied to developing computational analysis techniques. This is a fundamental principle of rational drug design, where the original lead is often structurally related to the endogenous substance that modulates target activity. On a final note, however, rational drug design is not applicable in all circumstances, and a great deal of drug discovery still relies heavily on the serendipitous discovery of new drugs by empirical screening of various chemical classes.

## 3.2 Pharmacokinetics

***Introduction*** Lead derivation and optimization are guided by three predominant factors: efficacy, specificity, and pharmacokinetics. Pharmacokinetics is the study of the time course of drug absorption, distribution, metabolism, and excretion (ADME), and how ADME relates to the therapeutic and toxic effects of a drug. The key parameters and methods used in ADME studies are listed in Table 4. During the 1990s, it was noticed that many drug candidates were abandoned during clinical trials due to poor pharmacokinetics [46]. This, in part, reflects problems with

**TABLE 4 Key ADME Parameters and Methodologies[a] for Early Studies**

| |
|---|
| **Physicochemical properties** |
| Chemical stability and degradation |
| Solubility |
| $pK_a$ |
| Lipophilicity ($\log P$) |
| **Binding target screens** |
| Plasma protein binding |
| Nonspecific interactions/binding studies |
| **Absorption and distribution** |
| Passive transport into the systemic circulation system—Caco-2 MDCK cells |
| P-gp substrate/transporter assays |
| Absorption screening—models of the blood–brain, placental/reproductive, epithelial, and, corneal barriers |
| PBPK modeling |
| **Metabolism and excretion** |
| CYP metabolism |
| CYP inhibition/induction |
| Glucuronidation |
| Nuclear receptor activation |
| Regulation of lipid and cholesterol metabolism |
| Aromatase inhibition |
| Metabolite stability |
| Kidney cells and tissue preparations |

[a]These approaches are increasingly being used by pharmaceutical companies in an attempt to reduce drug attrition rates.

species extrapolation, allometric scaling, and the selection of suitable preclinical models.

***Absorption and Distribution*** For ease of administration, oral delivery is the most favored route of administration. Orally delivered drugs need to possess good gut absorption and clearance, as well as good metabolic stability; this is to ensure adequate systemic exposure. This is expressed as oral bioavailability, which is the fraction of the ingested drug that is available systemically, depends on both absorption and elimination, and which can be estimated from the equation:

$$\text{Oral bioavailability} = \text{Fraction absorbed across the intestinal wall} \times \text{Fraction that is not cleared by the gut} \times \text{Fraction not cleared by the liver} \times \text{Fraction not cleared by the lungs}$$

The fraction absorbed from the gut is dependent on lipophilicity (log $P$), namely, the hydrophobicity, molecular size, and hydrogen bonding potential of the drug, and its permeability, which is dictated by van der Waals forces that are due to nitrogen and oxygen atoms, which influence the polar surface area of a drug molecule. Highly lipophilic drugs are likely to be absorbed directly from the gut into the lacteals and enter the lymphatic system before the general circulation, thus avoiding first-pass metabolism in the liver. On the other hand, small, moderately lipophilic drugs are likely to be passively or actively transported (depending on electrical charge, structure, and intermolecular interactions) across the intestinal barrier into the hepatic portal vein and are destined for the hepatobiliary system. Lipinski's rule-of-five [47] is commonly used to predict the permeability of a compound according to the rule that >5 hydrogen bonds, coupled with a molecular mass of >500 Da, a $c \log P > 5$, and >10 nitrogen and oxygen atoms are indicators of poor absorption. The full list of criteria for passive absorption through biological membranes is given in Table 5.

The biopharmaceutics classification system (BCS) [48] is a scientific framework for classifying drug substances based on their aqueous solubility and intestinal permeability. It takes into account three major determinants of the rate and extent of absorption following the administration of a solid oral dosage—(1) dissolution, (2) solubility, and (3) intestinal permeability—and can be used to avoid some animal studies. According to the BCS, drug substances are classified as follows:

**TABLE 5 Criteria[a] for Absorption from the Gut for Drug-like Compounds**

| |
|---|
| **Rule-of-five:** |
| • MW ≤ 500 |
| • $c \log P \leq 5$ |
| • H-bond donors ≤ 5 |
| • H-bond acceptors (N and O atoms) ≤ 10 |
| **Additional criteria** |
| • Polar surface area ≤140 Å$^2$ or Σ H-bond donors and acceptors ≤12 |
| • Rotatable bonds ≤10 |

[a]These include Lipinski's criteria and experimentally determined criteria for the gut absorption of orally administered drugs.

Class I: High Solubility–High Permeability
Class II: Low Solubility–High Permeability
Class III: High Solubility–Low Permeability
Class IV: Low Solubility–Low Permeability

In the case of Class 2 and 4 drugs, dissolution is the predominant factor, and this can be estimated by using *in vitro* data that correlate well with *in vivo* results. The rate-limiting step for the absorption of a class 1 drug is gastric emptying, whereas permeability is the most important factor in the case of class 3 drugs. On a cautionary note, however, many drug-like compounds are exceptions to this rule, due to active transport mechanisms, the involvement of carriers, and possible biotransformation in the gut, and because the rule only applies to orally administered drugs. The rule is more likely to be applicable where the drug is a mimetic of a natural product. Furthermore, good oral bioavailability is shown by many drugs that are larger than 500 Da, but conformationally constrained, and/or that have reduced polar surfaces. This means that many potentially useful leads are discarded, if predictions are based solely on physicochemical indicators, without additional studies.

Thus, *in vitro* screens for the determination of bioavailability are indispensable during early drug development. *In vitro* systems range from relatively simple subcellular fractions, tissue slices or perfused organ preparations, through primary cultures and cell lines, grown either as mono cultures or cocultures, to three-dimensional organotypic cultures, which include reconstructed tissue models (see Ref. 49 for a review). For instance, *in vitro* approaches to determine permeability include the use of animal tissues or cell lines. Where a drug is absorbed by simple passive diffusion, rat everted intestinal rings [50], single-pass perfused intestinal preparations [51], and the human adenocarcinoma cell line Caco-2 [52] are often used to assess intestinal permeability. The common problem of accounting for differences in biotransformation can be addressed by supplementing these assays with a metabolic component. Everted tissue rings, in particular, can suffer from problems related to tissue viability, while the use of Caco-2 cells suffers from (1) the lack of mucus that coats the luminal surface of intestinal cells *in situ* and (2) the possibility that metabolic properties and other essential properties of the cells are lost during repeated passaging. The need for long culture times has been met by using Madin–Darby canine kidney (MDCK) cells, which, like Caco-2 cells, form a columnar epithelium with tight junctions, but which only require 3 days in culture. MDCK cells express fewer transporters, however, and are thus more suited to situations where there is little indication of the involvement of transporters, where the data from MDCK and Caco-2 cells are comparable [53]. However, where first-pass metabolism occurs, cells expressing CYP enzymes are more useful (see later discussion). A comparison of cell-based and tissue-based permeability data indicates that there are good correlations between data from such studies, at least in terms of their ability to rank drugs according to permeability [54]. However, both methods tend to underestimate the absorption of drugs that are actively transported, and to overestimate the absorption of drugs that are subject to efflux pump transport [55].

Two further methods for the assessment of absorption are (1) the use of immobilized membranes or artificial membrane (IAM) [56] chromatography and

(2) the parallel artificial membrane permeation assay (PAMPA) [57]. IAM and PAMPA are both relatively rapid screens, the latter also being amenable to a microtiter plate format. Hence, both are suitable for screening large libraries of drugs for passive diffusion. Again, these systems are only really suited to absorption studies where there is little indication that a drug might be subject to transporter/efflux protein binding, as determined by direct binding assays. For instance, the drug export pump, P-glycoprotein (P-gp), which protects the brain from toxic substances by transporting them back into the blood, and which is, incidentally, also found in tissues such as the kidney and the liver, has been used as the basis of a model system composed of porcine brain capillary endothelial cells. This system allows the ability of drugs to inhibit the efflux of a fluorescent P-gp substrate out of the cells to be readily monitored in a 96-well format [58]. This model may also be a useful supplement to many of the studies on more complex tissue and cell culture systems, given the key role of efflux proteins in determining internal drug concentrations. Indeed, the internal dose that reaches other tissues may have a crucial bearing on the use of the final marketed drug and the nature of the target patient group. Despite the fact that the placental transfer of drugs and other xenobiotics is known to occur, the process is poorly understood. What is known is that it involves efflux proteins, and that P-gp binding capacity may assist with decisions as to whether reproductive and developmental toxicity testing is likely to be required.

Dermal absorption can be determined by using organotypic skin models comprising stratified layers of epidermal cells, with each layer exhibiting morphological and functional differentiation. This has given rise to several commercially available organotypic and reconstructed human skin *in vitro* culture models, including EPISKIN™ (http://www.loreal.com) and EpiDerm™, and its fibroblast-supported version, Full Thickness EpiDerm™ (http://www.mattek.com). Multilayered models of the tracheobronchial tract are also available and permit squamous metaplasia [59], mucin production, and mucociliary clearance to be analyzed for making respiratory toxicity predictions of inhaled drug preparations [60]. Of particular relevance to the development of many cell-based organotypic models is the use of microporous substrates, which have led to physiologically more relevant culture conditions for studies of transcellular transport and cell–tissue interactions. Also, blood–brain barrier (BBB) function can be monitored by using coculture systems and reconstructed models in which sufficiently tight cell–cell junctions are formed (reviewed in Ref. 61), which might be more appropriate when it is envisaged that a delivery system might be used. For example, in models of the BBB, brain microcapillary endothelial cells can be cultured with astrocytes, glial cells, or neurones [62, 63] and "whole-brain" spheroid culture systems have been used to model absorption in neurotoxicology [64].

The nasal route is also widely used for topical and systemic targeting and has recently been considered as a suitable route for active peptide and protein administration, since the nasal mucosa has a higher permeability than the intestinal epithelial layer, because of its porous and neural pH endothelial basement membrane, and since first-pass hepatic and intestinal metabolism is largely avoided. However, mucociliary clearance is relatively fast, and the dose volume is smaller than that administered by other routes. These factors should be considered during the early development of therapeutics designed for nasal administration. *In vitro* models of human nasal absorption are particularly important, given that there are substantial

differences in nasal structure and function between many laboratory animal species and humans, which complicates data extrapolation [65]. Human nasal epithelial tissues can be obtained noninvasively, but the resulting heterogeneous primary cell population is difficult to culture and maintain. This has prompted the development of a nasal absorption model, based on Calu-3 human lung adenocarcinoma cells, derived from the upper airway of the lung. The cells can be grown at the air–liquid interface, as confluent and polarized sheets, with tight junctions, which secrete mucin, while also possessing cytochrome P450 (CY)PA1 and CYP2B6 activities and transport functions [66].

As illustrated earlier, because whole perfused organs, tissue slices, tissue isolates, organ fragment cultures, and other organotypic preparations can have limited lifetimes in culture, a wide variety of cell lines are used instead of primary cells and tissues. However, it is important for the cell lines to be thoroughly characterized, since, although use of human tissue and cells is advantageous as it obviates the need for interspecies extrapolation, the more commonly used animal cell lines may also provide data of use during drug development.

***Metabolism*** Metabolism is a major determinant of drug efficacy and toxicity and has a major influence on drug pharmacokinetics. The metabolism of a drug will depend not only on its structure but also on the presence and expression levels of biotransformation enzymes. Metabolism is one of the main factors that determines not only how well a drug is absorbed into the systemic system but also how it is then transported to, and taken up by, specific tissues, as well as how it is eliminated. Susceptibility to biotransformation can dictate the effective dose of a drug at its intended site of action. Hence, predicting the likely routes and consequences of biotransformation is an essential part of early drug screening.

Phase I drug biotransforming enzymes include the CYP enzymes, flavin-containing monooxygenases, alcohol and aldehyde dehydrogenases, aldehyde oxidase, and peroxidase. Phase II pathways involve conjugation reactions, such as glucuronidation, glutathione conjugation, sulfation, methylation, coenzyme A conjugation, and phosphorylation. The enzymes involved are polymorphic and exist as a large number of isozyme forms that have wide substrate specificity and vary in nature and activity according to tissue and species. This is also the case for other metabolic enzymes and, indeed, for individual differences in drug response and metabolism. It is known, for instance, that there are at least 30 variants of the human CYP2D6 [67], which is responsible for the metabolism of almost one-third of all current therapeutic drugs. Phase I biotransformation is the main source of toxic intermediates or active drugs from innocuous parent chemicals (prodrugs). The main redox reaction, catalyzed by CYP-dependent monooxygenases (CYP-DMO), yields more polar and therefore more readily excretable metabolites. These by-products can exhibit biological effects that exceed, or differ from, those elicited by the parent molecule. Hence, a xenobiotic might appear to lack efficacy in an *in vitro* system, due to the absence of a biotransformation pathway necessary for its conversion into an active form.

Most of the systems described so far are not specifically designed with metabolic competence in mind. One way to circumvent problems with metabolic competence is by adding subcellular or cellular metabolizing systems and assessing the production of known metabolites [68]. For instance, by using rat liver microsomes,

it is possible to monitor the biotransformation of small amounts of drugs by using capillary electrophoresis [69]. Metabolic cellular systems can be divided into three main categories: (1) metabolically competent indicator cells (e.g., hepatocytes), (2) coculture systems comprising noncompetent indicator cells (e.g., fibroblasts) mixed with metabolicallycompetent cells, and (3) genetically engineered cell lines that can simultaneously act as indicators of both selected metabolic pathways and toxicity.

Since the liver is the major site of drug metabolism, the use of primary hepatocytes has become well established for studying drug metabolism and drug interactions. Hepatocyte models composed of primary cells assembled into a double collagen layer sandwich provide an *in vivo*-like environment, which can retain some important liver functions long enough for useful data to be obtained [70]. More recently, a mini bioreactor scaffold has been devised for biotransformation studies, comprised of a polycarbonate scaffold in 6-, 24-, and 96-well formats, which supports a gas-permeable PTFE membrane, suitable for medium throughput screening [71]. Animal or human cells can be used, although the latter are often in short supply and rapidly lose their viability and phase I and II biotransforming capacities. However, using animal cells does not overcome potential problems with interspecies differences. HepG2 cells have been used, since they allegedly retain human phase I and II drug-metabolizing activities, as well as bile acid and albumin synthesizing capabilities that are normally lost during hepatocyte culture. In addition, these cells can be transformed with specific genes (e.g., CYP genes), in order to bolster their metabolic competence. Drug interactions with efflux proteins such as P-gp can also be monitored with HepG2 cells, to gauge the extent of drug uptake into the liver [72].

Krebsfaenger et al. [73] constructed a panel of Chinese hamster cell lines that stably expressed variants of the human CYP2D6 gene, or in the case of the genetically engineered and metabolically competent V79 Cell Battery™, tissue-specific human phase I or II metabolic enzymes, including a number of CYP variants, glutathione *S*-transferases, and *N*-acetyltransferases. Since the cell battery also contains cell lines that express equivalent enzymes from other species, there is some scope for resolving species differences in metabolism. This system may not be able to account for the consequences of the overexpression of enzymes, or account for the onslaught of long-range biochemical pathways that may regulate metabolism. Nevertheless, the use of polymorphic cell lines has clear advantages over the use of a genetically homogeneous cell line.

A major problem, particularly when trying to model chronic exposure effects, is simulating *in vivo* perfusion rates by using cell-based systems of biotransformation to (1) ensure the removal of metabolites, and of other products such as reactive oxygen species; and (2) avoid cofactor depletion [74, 75]. Systems based on cell or cell coculture models have been designed to address these problems. Canova et al. [76] have recently described a flat membrane bioreactor, in which primary rat hepatocytes were cultured as collagen sandwiched monolayers on a polycarbonate plate. Since this system permits continuous perfusion, the cells exhibited a high level of metabolic competence and extended viability, as compared to cells grown as adherent monolayers.

Drugs can also inhibit CYP activity without being themselves subject to CYP metabolism, thereby causing the accumulation of toxic substances by modeling the interactions between the drug and the enzyme. Some basic information about the

chemical classes that are subject to biotransformation by these key enzymes is already known and can be used for such predictions. For example, CYP1A1 is very active in metabolizing polycyclic aromatic hydrocarbons, CYP2B1 and CYP1A2 preferentially metabolize aromatic amines, and CYP2E1 metabolizes low molecular weight chemicals. Perhaps the most important enzyme, CYP3A4, metabolizes larger molecules, and only six isoenzymes account for the metabolism of almost half of all the drugs in clinical use, namely, CYP1A2, CYP2C9, CYP2C19, CYP2D6, CYP3A3, and CYP3A4 [77, 78]. These enzymes are all polymorphic and (with the notable exception of CYP2D6) are inducibly expressed. Since there are a number of X-ray diffraction structures for CYPs, which can also serve as templates for the homology modeling of other isoenzymes, it is possible to use molecular dynamics simulations and molecule docking to determine whether a novel chemical class of drug candidate is likely to be subjected to CYP metabolism or act as an inhibitor [79]. The advantages of such approaches include the fact that these screens can be used to eliminate poor drug candidates prior to synthesis of the compounds, and undesirable metabolic labile sites can be modified to alter the effective dose and metabolically linked contraindications between medicines can be reduced. Some such *in silico* prediction methods rely on knowledge of the structure and physicochemical properties of the drug in question to serve as alerts. For instance, Singh et al. [80] developed a rapid and semiquantitative model for identifying CYP3A4-labile groups on the basis of exposure and the energetics of hydrogen bonds. This has been used to accurately predict CYP3A metabolism for a number of chemicals [80] (for a review, see also Ref. 81).

The most common technique used to identify metabolic stability, rate of drug biotransformation, and likely metabolic fate of a drug is LC-MS/MS. This technique has recently been automated for high-throughput automated metabolic and protein binding screening [82]. Established cell lines and cell batteries (such as the V79 Cell Battery described earlier) that are genetically engineered to express various phase I and phase II enzymes, either singly or in combination, are particularly important in this respect, since they permit the contributions of specific isozymes to metabolism to be investigated [73] during relatively fast metabolic screening.

## 4 SPECIAL CONSIDERATIONS FOR NOVEL THERAPEUTIC CLASSES

### 4.1 New Classes of Therapeutic Agents: A New Drug Development Strategy?

Over the last decade or so, there has been a dramatic change in the nature of targets, with new protein, nanomedicines, and gene-based and cell-based therapeutics being investigated at an unprecedented rate. These biotechnology-derived pharmaceuticals are mainly used for the diagnosis, prevention, and treatment of serious and chronic diseases and range from blood and blood components and antitoxins to monoclonal antibodies, growth factors, vaccines, gene transfer products, and, potentially, cell-based therapeutics, and they may incorporate nanoparticulate delivery systems. As discussed below, the complex pharmacokinetic behavior of these therapeutics is only just starting to be understood, and attempts to group the preclinical testing of them has been attempted on a very *ad hoc* basis. Nevertheless, there are

some specific considerations for each type of therapeutic that should be borne in mind during early therapeutic development.

### 4.2 Protein-Based Therapeutics

Protein-based therapeutics can be subjected to a broad range of posttranslational modifications, including phosphorylation, glycosylation, and intermolecular and intramolecular bond formation, which can vary, depending on the cell or organism used to generate the product. In the case of replacement therapeutics, it is important to mimic the posttranslational modification seen in humans as precisely as possible. For novel therapeutics, the consequences of using recombinant techniques to produce a biological product, such as a monoclonal antibody, should be evaluated, to ensure that the finished product is sufficiently humanized, has folded correctly; and exhibits the type of posttranslation modification required for its proper function. In all cases, the primary aim is to avoid having the product elicit immunogenic responses that reduce its effectiveness or give rise to adverse effects.

The pharmaceutical industry has developed ways of scaling up the cell-based manufacture of biological products in serum-free and protein-free chemically defined media, in order to minimize the abnormal glycosylation of proteins. This is to comply with a regulatory requirement for the glycosylation profile of a protein to be maintained. This is particularly important, since products such as monoclonal antibodies can elicit antibody-dependent cell cytotoxicity, depending on their glycosylation state [83]. Indeed, protein-based therapeutics are often developed and characterized by mass (e.g., SPR, MS), activity (in cellular or biochemical assays), and immunogenicity assays (e.g., antibody-dependent cell-mediated cytotoxicity and immune cell proliferation assays), more so than in small chemical therapeutics. These measurements are generally obtained much earlier in the drug development process.

Other problems encountered during the development of protein-based therapeutics are caused by the fact that most of the products are human specific, potentially limiting the relevance of preclinical animal-based studies. In the case of the immunomodulatory monoclonal antibody TGN1412, a surrogate anti-rat CD28 antibody (JJ316), specific for the equivalent epitope to that recognized by TGN1412 in humans, was used to probe the mechanism of action of the therapeutic. However, subsequent studies on TGN1412 were conducted on macaques, on the basis that the CD28 antibody in this species possesses an identical TGN1412 recognition epitope to that found on the human protein. None of these preclinical models were able to predict the very serious contraindications that later materialized during Phase I trials [84]. This casts some doubt over the utility of testing surrogate molecules in a surrogate species with regard to the estimation of human safety when applied to humanized products. In the United States, the review of biological products, including monoclonal antibodies and other therapeutic agents with novel mechanisms of action, was until relatively recently conducted by the CBER. However, the CBER recognized that the utility of animal studies in the development of such therapeutic agents is often limited, due to species differences between the molecular targets. As a consequence, the jurisdiction for testing such products has now been transferred to the CDER, and it is almost certain that there will be a requirement that biological products conform to the two species test criteria adopted for other pharmaceuticals.

### 4.3 Gene Therapeutics

Advances in the use of viral vectors (e.g., adenoviruses, lentiviruses, and RNA viruses) and nonviral vectors (e.g., cells, liposomes, and DNA) have resulted in a growing interest in developing gene therapeutics for targeting single gene deficiencies, as well as cancers and neurodegenerative and tissue repair diseases. Nucleotide-based therapeutics will almost certainly be taken up by different cell types in the body via the same mechanisms. However, the key questions are the longevity of these treatments, whether they will result in incorporation into the genetic makeup of specific cell types, and the long-term consequences for the genetic makeup of germline and somatic cells. These products may therefore require early screening for reproductive and developmental toxicity, potentially via the use of fish larval forms and developmental assays with *D. melanogaster or C. elegans*. Genotoxicity screens that involve cell-based screens may also be relevant. Incidentally, the FETAX (frog embryo teratogenesis assay—*Xenopus*) test [85], a 4-day whole embryo developmental toxicity test, has been available for use for several years, but the apparent *status quo* with regard to its evaluation is consistent with concerns over its relevance to humans. The requirement for animal-based developmental, reproductive, or genotoxicity testing can then be assessed on a case-by-case basis, depending on the class of vector, transgene, and delivery method. In the absence of suitable long-term *in vitro* models, chronic toxicity studies in animals might also be justified, in which DNA integration is monitored by sensitive PCR-based techniques, although the two-year rodent assay may not be suitable for this purpose [86].

### 4.4 Nanomedicines

Nanomedicines (and ultrafines) consist of, or contain, organic or inorganic nanomaterials of variable dimensions, which, according to the U.S. Patent Office, are of 100 nm or less in size. The two main classes are carbon-based liposomes, dendrimers, fullerenes, nanotubes, nanowires, nanorods and metal oxide-based quantum dots and PEG-polyester systems. Nanomedicines, such as PEG-modified versions of existing drugs (e.g., PEG-granulocyte colony stimulating factor), are already on the market. Many other nanotechnology-based applications, such as carbon nanotube-based RNA, DNA, protein, and drug delivery systems, are being developed [87]. Indeed, between 2004 and 2005 there was a 60% increase in the number of such products in the development pipeline.

There is currently little regulatory guidance on how nanomedicines should be characterized, and each nanomedicine is generally assessed on an individual basis. It is also clear that information for macroparticle equivalents is often of limited relevance. According to the FDA, toxicity screening of nanomedicines [88] should involve (1) an assessment of physicochemical characteristics, including core and surface chemistry; (2) *in vitro* studies to determine absorption by the intended, and potentially additional, routes of exposure, binding studies, bioaccumulation/cellular uptake studies, and cytotoxicity screening; and (3) *in vivo* understanding of biokinetics and toxicity.

Nanomedicines display physical and chemical properties that are often distinct from their macroparticle counterparts. For instance, the small particle size corresponds with a high surface area: mass ratio means that not only are nanoparticles

likely to be more reactive because of their quantum properties [89], they are also more likely to agglomerate in a way that makes absorption difficult to predict by using biokinetic modeling alone. This would require information about the extent of agglomeration and the proportion of each type of particle in heterogeneous mixtures of particles, as well as a fundamental understanding of how agglomeration affected different reactivities. Agglomeration is dependent on surface chemistry and charge, and this will also have a crucial bearing on uptake into different tissues [90].

Nanoparticles may be able to escape phagocytic activity and may not only have extended biological half-lives but may also elicit immunogenic and inflammatory responses, especially since they may be able to access lung tissues and be able to cross the BBB, the intestinal mucosa, and other physical barriers more readily than macroparticles. Indeed, the size of a nanoparticle is generally that of a typical protein. Note that BBB passage requires nanoparticle dimensions of 20–50 nm, whereas 70-nm particles are able to cross into the pulmonary system, and spherical fullerenes appear to accumulate in the liver [91]. However, at present, there is little information about how nanoparticles are able to interact with the immune system and with specific tissues.

Examining the tendency of nanomedicines to bioaccumulate is an important step in the nanomedicine development strategy, so the need for long-term toxicity testing cannot be overemphasized. There is already some indication as to how readily different types of carbon-based nanomaterials accumulate in different tissues [92], as well as about how physicochemical properties such as size, shape, tendency to agglomerate, surface fictionalization, and chemical composition relate to the absorption, distribution, and biokinetics of nanoparticles (see Ref. 93 for a review).

## 5 *IN VITRO* ASSAYS: APPLICATIONS IN SAFETY PHARMACOLOGY

### 5.1 Background

Many surrogate assays have been developed for the detection of toxicological endpoints. Many of those used specifically for lead development and early ADME characterization have already been considered. Here, we consider *in vitro* assays that are relevant to the early characterization of toxicologically relevant endpoints. A list of methods that are currently available for toxicological endpoints that may be more suited to high-throughput safety pharmacology have been developed and/or validated, some of which have already been afforded OECD Test Guideline status (Table 6). These and other such tests may eventually expedite the safety testing of pharmaceutical products. At present, however, they are most commonly used to provide evidence to supplement animal test data. Their potential application to toxicology testing is illustrated in Fig. 2. Their validation status will have a bearing on the acceptability of specific tests as part of a medicine's dossier for different regulatory agencies, and there is generally a delay of several years between validation and regulatory acceptance. It should be noted that in the EU, under the terms of *Directive 86/609/EEC*, once an *in vitro* alternative to an animal test has been deemed to be reasonably and practicably available and has been validated for a specific purpose, it must be used instead of the equivalent animal test.

**TABLE 6 Status of Nonanimal Methods that Are of Relevance to Drug Development**[a]

| Test Method | Test System | Endpoint | OECD TG Reference and Comments |
|---|---|---|---|
| *In Vitro Test Methods for Which There Are OECD Health Effects Test Guidelines (Including Draft Guidelines Under Review for Acceptance) http://www.oecd.org/home/* | | | |
| Transcutaneous electrical resistance test (TER) | Monitors changes in electrical resistance as a measure of loss of corneum integrity and barrier function; involves skin disks from euthanized rats | Skin corrosion (topical agents) | TG 430 |
| Human skin models (EpiDerm™, EPISKIN™) | Reconstructed human epidermal equivalent (commercial system) used to assess cell viability, involving the MTT reduction test | Skin corrosion (topical agents) | TG 431 |
| 3T3 NRU phototoxicity test | BALB/c 3T3 (murine) cell line cytotoxicity based on Neutral Red uptake to measure cell viability; not a direct replacement alternative, as there is no *in vivo* equivalent test | Phototoxicity | TG 432 |
| Corrositex™ membrane barrier test | An artificial barrier system coupled to a pH-based chemical detection system | Skin corrosion (topical agents) | Draft TG 435 |
| Bacterial Reverse Mutation test (Ames) | Revertant bacteria detected by their ability to grow in the absence of the amino acid | Genotoxicity | TG 471 |
| *In vitro* mammalian chromosome aberration test | Microscopic detection of chromosomal damage to cells in culture | Genotoxicity | TG 473 |
| *In vitro* mammalian cell gene mutation test | Functional bioassays to monitor mutations in enzyme encoding genes | Genotoxicity | TG 476 |
| Sister chromatid exchange assay | Cells in culture are examined after two rounds of division by metaphase arrest and chromosomal preparation; chromatid exchange is monitored by microscopy | Genotoxicity | TG 479 |
| Gene mutation assay in yeast | *Saccharomyces cerivisiae* exposed to the test substance are grow under different culture conditions used to monitor mutagenic potential (cf. Ames test) | Genotoxicity | TG 480 |
| Mitotic recombination assay in yeast | Crossover or gene conversion following exposure of yeast to the test substance; relies on different growth requirements of mutated and wild-type yeast strains | Genotoxicity | TG 481 |

| | | | |
|---|---|---|---|
| Unscheduled DNA synthesis in mammalian cells | Measures the DNA repair synthesis after deletions caused by the test substance; based on the incorporation of radioactive nucleotides into the newly synthesized DNA | Genotoxicity | TG 482 |
| *In vitro* micronucleus test | Cell-based assay; supplement to TG 474 (*in vivo* micronucleus test); detection of chromosome damage and formation of micronuclei | Genotoxicity | Draft TG 487 |
| Sex-linked recessive lethal test | *Drosophila* are exposed to the test substance. Germline transmission of mutations is monitored through two successive generations | Reproductive toxicity | TG 477 |
| | *Validated Methods that Are Yet to Be Introduced into Regulatory Use* | | |
| EpiOcular™ | Human keratinocyte derived model of the corneal epithelium barrier function | Eye irritation (topical application) | Retrospective (weight-of-evidence) validation (ECVAM) |
| *In vitro* micronucleus test | CHL/IU, CHO, SHE, or V79 cell lines are commonly used, with or without metabolic activation, to monitor damage and formation of micronuclei in interphase | Mutagenicity | Retrospective (weight-of-evidence) validation (ECVAM) |
| Embryonic stem cell test | 3T3 cell cytotoxicity and differentiation of embryonic stem murine cell lines used to examine teratogenic potential | Developmental toxicity | Endorsed as screening test (EU) |
| Postimplantation rat whole embryo test | Morphological assessment of rat embryos | Developmental toxicity | Endorsed as screening test (EU) |
| Micromass test | Micromass cultures of rat limb are bud monitored for inhibition of cell proliferation and differentiation | Developmental toxicity | Endorsed as screening test (EU) |
| | *Methods Undergoing Validation* | | |
| EPISKIN™ | Reconstructed human skin system used with MTT assay to monitor barrier function | Skin irritation | Report stage in EU |
| EpiDerm™ | Similar to EPISKIN™ | Skin irritation | Report stage in EU |

**TABLE 6** ***Continued***

| Test Method | Test System | Endpoint | OECD TG Reference and Comments |
|---|---|---|---|
| | *Prevalidated Methods* | | |
| SkinEthic eye model | Epithelial corneal cell line used for cytotoxicity testing based on the MTT reduction assay | Eye irritation | Appraisal stage in EU |
| | *Methods Undergoing Development, Prevalidation, or Evaluation* | | |
| Tissue culture models | Neutral Red release and silicon microphysiometry or fluorescein leakage bioassays with human keratinocytes and MDCK cells, respectively; red blood cell (RBC) hemolysis test | Eye irritation | Being reviewed by ECVAM for possible retrospective (weight-of-evidence) validation |
| Organotypic models | Bovine corneal opacity and permeability (BCOP) assay, with postmortem corneas; hen's egg test on the chorioallantoic membrane (HET-CAM assay); isolated rabbit and chicken eye tests (IRE and ICE) | Eye irritation | Being reviewed by ICCVAM for possible retrospective (weight-of-evidence) validation |
| Cell transformation assay | With SHE and BALB/c 3T3 cell lines | Carcinogenicity | |
| Modified Leydig cell line | Analysis of progesterone production as a measure of the test substance effects on steroid hormone production | One/two generation study | For use as part of test battery |
| Testis slices | Assessment of steroid production capacity of the Leydig cells upon exposure of *ex vivo* rat tissue to toxicants | One/two generation study | For use as part of test battery |
| Human adrenocortical carcinoma cell line | Assay to allow entire steroid pathway effects to be mapped | One/two generation study | For use as part of test battery |
| Placental microsomal aromatase assay | Monitors the ability of substances to affect steroid production; a subcellular microsomal assay is used industrially | One/two generation study | For use as part of test battery |

[a]This is a comprehensive list of methods that have been validated or that are at various stages of development for toxicity testing. More information about these methods and how they can be applied is available from: http://www2.defra.gov.uk/research/project_data/more.asp?!=CB01067&M=KWS&V=reach&scope=0.

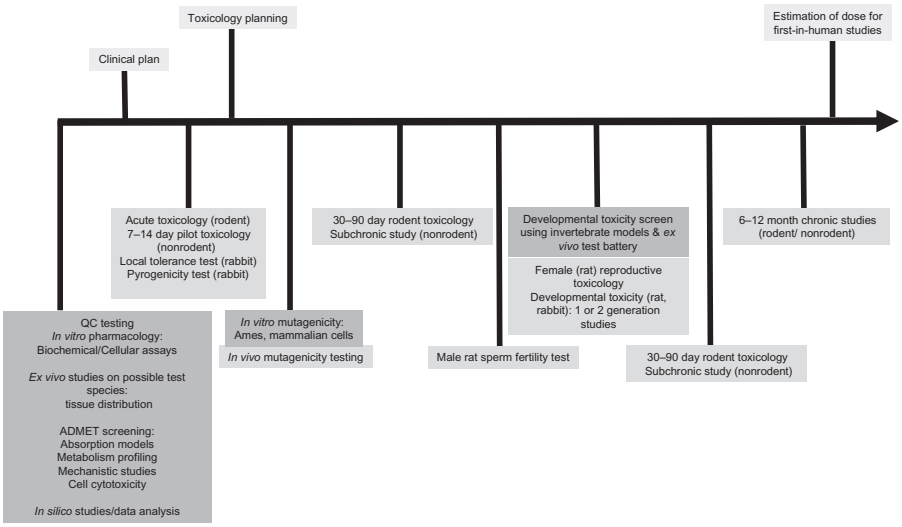


**FIGURE 2 A typical toxicity plan from formulation of a clinical plan to first-in-human studies.** The figure summarizes the general regulatory requirements and the role of *in vitro* and animal studies in preclinical safety pharmacology.

## 5.2 QT Prolongation

The recently developed ICH S7B guideline relates to the prolongation of the QT interval, the time between the start of the Q wave and the end of the T wave in the electrical cycle of the heart. Drugs that prolong the QT interval do so by blocking the activity of the human ether-a-go-go-related gene (hERG) channel found on cardiomyocytes, which plays a role in phase 3 depolarization. This can, but does not always, lead to the potentially fatal, although rare, tachyarrhythmia, Torsades des pointes. Electrophysiological recordings in cells and tissue explants can be used to screen for hERG activity. The most commonly used nonanimal assays involve studies on rabbit left ventricular wedge or perfused heart preparations [94].

Several cell-based and *in silico* prediction tools could also be used effectively for early screening. It is clear that while several classes of drugs, including antihistamines, antiemetics, antibacterials, and neuroleptics, have been associated with QT prolongation [95], not all the drugs of a particular class will cause this effect. Hence, although mutational analysis of hERG and pharmacophore modeling has helped to elucidate how several drugs block channel activity (reviewed in Ref. 96), the prediction of QT prolongation by using *in silico* methods alone is problematical. As a consequence, the first realistic step in screening for hERG-blocking activity generally depends on the use of cell-based studies. It should also be remembered that hERG activity is not the sole determinant of whether a drug is likely to cause cardiac arrhythmias, since the expression and activity of other ion channels can either mitigate or exacerbate the effects of a drug on QT interval prolongation. Metabolism and species differences in ion channel distribution and activity must also be taken into account. Species differences can be avoided by using cell lines such as Chinese hamster ovary or human embryonic kidney cells heterologously expressing hERG, since mammalian cells are the ideal system for studying potassium channel activity at physiologically relevant temperatures (see Ref. 96 for a review). This is the case

even despite the fact that while patch clamping is generally more readily achievable for larger cell types, such as *Xenopus* oocytes expressing hERG, the lower temperature required for their culture can complicate data extrapolation to humans. More recently, human embryonic stem cell (HESC)-derived cell types with primary cardiomyocyte-like properties have become available [97, 98]. However, although these cells act as a suitable surrogate for cardiomyocytes, the use of HESCs raises ethical issues. Adult stem cells derived from tissues such as bone marrow have therefore also been investigated as primary cardiomyocyte surrogates [99].

### 5.3 ICH Guidelines

Pyrogenicity testing forms part of the quality control of pharmaceuticals, particularly of biotechnological products that may be contaminated with gram-negative bacteria. Five *in vitro* methods have been proposed and evaluated as possible alternatives to pyrogenicity testing in animals (usually rabbits). These are based on the use of human whole blood or blood subcellular fractions linked to interleukin-based screens. These methods have been validated under the auspices of ECVAM, and the ECVAM Scientific Advisory Committee has endorsed their use. Draft recommendations regarding the applicability of these methods were published by ICCVAM in December 2006 [100].

The ICH guidelines for the safety testing of pharmaceutical products for human use are listed in Table 7. In general, tests are conducted by the same route of exposure as that anticipated for clinical use. Where there is more than one route of exposure, at least two routes should be investigated. Where the intended route of human exposure is impracticable for animal studies, the route of exposure in animal studies will need to be different, although the aim is to establish whether the level of systemic exposure is similar to that via the clinical route. In the case of ocular administration, systemic exposure is not automatically assumed, so some toxicological studies, such as carcinogenicity testing, may not be required, unless there is compelling evidence to suggest otherwise. Carcinogenicity testing is not usually required for peptide and protein replacement therapeutics, unless they are significantly different from their natural counterparts. The dose used is generally derived from the MTD from 3-month toxicity studies, as well as from a consideration of toxicokinetic data and the known or predicted consequences of biotransformation of the chemical. Carcinogenicity testing may also be guided by the outcome of genotoxicity and cell transformation studies, the majority of which involve cell-based screens for gene mutation or chromosomal aberration. The ICH guidelines require three tests for genotoxicity: (1) a test for gene mutation in bacteria, (2) an *in vitro* mammalian cell or mouse lymphoma assay for chromosomal damage, and (3) an *in vivo* test for the detection of chromosomal damage in hematopoietic cells in the bone marrow of rodents.

Determination of the acute toxicity of a single dose of a pharmaceutical no longer requires an $LD_{50}$ determination. Instead, it is most commonly derived from the maximum tolerated dose (MTD). The ICH guidelines recommend that acute toxicity is determined in a rodent and in a nonrodent species. However, in many cases, it may be possible to justify the use of a single species and/or incorporation of the MTD estimation into a dose escalation study from longer term/repeat-dose studies. The justification for conducting repeat-dose studies is partly based on the outcome of single-dose studies. This is especially the case when single-dose studies indicate

**TABLE 7 ICH Guidelines[a] for Safety Pharmacology**

| ICH Guideline for Safety Pharmacology http://www.ich.org/ | Description of Content |
|---|---|
| *Acute Toxicity Testing* | |
| M3(R1)—Maintenance of the ICH Guideline on Non-clinical Safety Studies for the Conduct of Human Clinical Trials for Pharmaceuticals | S4—single dose $LD_{50}$ tests for pharmaceuticals have been abandoned; the guideline sets out the duration of acute, repeat dose, and chronic tests in animals, and the criteria for reproductive, developmental, fertility, genotoxicity, and carcinogenicity testing, as well as the times during development that each type of safety test should be performed |
| *Carcinogenicity Testing* | |
| S1A—Guideline on the Need for Carcinogenicity Studies of Pharmaceuticals<br>S1B—Testing for Carcinogenicity of Pharmaceuticals<br>S1C(R1)—Dose Selection for Carcinogenicity Studies of Pharmaceuticals | Long-term studies, with dosing at the MTD level or, in the United States, a dose 100-fold higher than the recommended human daily dose, are conducted in a single rodent species (the rat unless evidence suggests it is not an appropriate species), for a duration of no less than 6 months; additional rodent studies of short-term or medium-term duration may be required (generally in mice/GM mice); a weight-of-evidence approach is used to determine whether there is a likely cause for concern based on genotoxicity tests, the results of cell transformation studies, effects of same-in-class/SAR, preneoplastic lesions in repeat-dose studies, and/or long-term tissue retention of the drug or its metabolites |
| *Genotoxicity Testing* | |
| S2A—Guidance on Specific Aspects of Regulatory Genotoxicity Tests for Pharmaceuticals<br>S2B—Genotoxicity: A Standard Battery for Genotoxicity Testing of Pharmaceuticals | Provides recommendations as to the bacterial strains and mammalian cell lines or human blood cells that could be used to detect various types of DNA damage; defines the concentrations to be used in each type of test and how to make decisions as to whether to proceed with *in vivo* animal studies |
| *Toxicokinetic Studies* | |
| S3A—Note for Guidance on Toxicokinetics: The Assessment of Systemic Exposure in Toxicity Studies<br>S3B—Pharmacokinetics: Guidance for Repeated Dose Tissue Distribution Studies | Toxicokinetic studies are normally integrated with repeat dose/3-month toxicity studies and provide multiple-dose pharmacokinetic data |
| *Chronic Toxicity Testing* | |
| S4—Duration of Chronic Toxicity Testing in Animals (Rodent and Non Rodent Toxicity Testing) | Performed for a 6-month duration in rodents and a 9-month duration in nonrodent species |

**TABLE 7** ***Continued***

| ICH Guideline for Safety Pharmacology http://www.ich.org/ | Description of Content |
|---|---|
| *Reproductive Toxicity Testing* | |
| S5(R2)—Detection of Toxicity to Reproduction for Medicinal Products<br>Toxicity to Male Fertility: An Addendum to the ICH Tripartite Guideline on Detection of Toxicity to Reproduction for Medicinal Products | one- or two-generation studies (generally in rodents) may be needed, depending on the intended use, form, and route of administration of a drug; incorporates embryotoxicity studies (usually on rabbits/rodents), but also considers the merits of other test systems, such as tissue, organ, and organism cultures |
| *Biotechnological Products* | |
| S6—Preclinical Safety Evaluation of Biotechnology-Derived Pharmaceuticals | Pharmaceutical products from cell cultures and transgenic plants/animals, such as proteins, hormones, antibodies, and proteins extracted from human tissues, are primarily subject to these guidelines; normally, two species are used, but in some cases only a single species is relevant; if no species is relevant, the creation of a GM model is considered |
| *Miscellaneous Safety Pharmacology* | |
| S7A—Safety Pharmacology Studies for Human Pharmaceuticals<br>S7B—The Non-clinical Evaluation of the Potential for Delayed Ventricular Repolarization (QT Interval Prolongation) By Human Pharmaceuticals | Defines the way in which toxicity of the major organ systems can be monitored by using animal and *in vitro* methods; one of the criteria for determining whether safety pharmacology needs to be performed is whether a drug is able to enter into the systemic system, so many topically administered drugs are excluded from full safety evaluation<br>QT prolongation: criteria are given for *in vitro* and *in vivo* animal tests |
| *Immunotoxicity Testing* | |
| S8—Immunotoxicology Studies for Human Pharmaceuticals | Addresses the use of local tolerance/allergy tests at the site of injection (generally in rabbits); and the use of assays that monitor humoral and cell-mediated immunity and immunomodulation; the requirement for immunotoxicity testing is based on a weight-of-evidence approach |

[a]The ICH website has additional information about the testing criteria and study design for each endpoint. Guidelines for quality assessment are also available from this site.

that the half-life of the drug (and/or its metabolites) in organs or tissues significantly exceeds the apparent half-life of the elimination phase in plasma, and when the steady-state levels of a drug are higher in repeat-dose studies than in single-dose studies. Predicted tissue-specific retention that leads to tissue-specific toxicity is also a justification for repeat-dose studies. Pharmaceuticals developed for site-specific delivery might also require repeat-dose studies.

Determining the acute dose toxicity of biotechnology products in animals is likely to be of limited relevance to safety assessment. This is because such products (e.g., recombinant proteins) are highly human specific and are likely to give rise to immunogenicity. However, the latter effect can be investigated by using the murine local lymph node assay (LLNA). This test has been accepted in the United States and the European Union as an alternative to the guinea pig maximization test for assessing allergic contact dermatitis caused by pharmaceuticals, and is now part of the OECD Health Effects Test Guidelines (TG 429).

The possibility of using *in vitro* systems to permit the high-throughput assessment of chronic toxicity is currently very limited, predominantly due to the problems inherent in undertaking repeat exposures and maintaining long-term cultures. However, one system developed from stem cells consists of a neurosphere with an outer layer of cells surrounding a growing core [64] that can be maintained in culture for up to a year. This system might be amenable for development into a long-term assay for assessing neurotoxicological endpoints (see Section 3.2).

## 6 SYSTEMS BIOLOGY

Many of the screens and assays described thus far are focused on measuring a single parameter. However, although they provide information relevant to the *in vivo* assessment of new drugs, they lack some of the complexities of animal studies. Nevertheless, this can sometimes be an advantage, particularly as the effects observed during *in vivo* studies can be due to a multitude of unknown mechanisms, and this can confound data interpretation and extrapolation. As illustrated next, recent developments in the omics-based technologies and microfluidics may provide a way of addressing this caveat against the use of existing cell-based and tissue-based screens.

### 6.1 Omics-Based Technologies

***Biomarkers*** Biomarkers discovered from, and subsequently used in, nonclinical studies can play several useful roles in drug development. During lead discovery, a biomarker can assist with the screening of chemicals, based on the ability of the candidate molecule to modulate the activity of a specific process, such as a biochemical pathway. In some cases, these biomarkers can be used to develop reporter gene screens, such as the CRE and ERE activation-linked expression of GFP or enzymes. In this way, the application of biomarkers can provide information about target interactions in support of the feasibility of a given therapeutic approach, as well as to assist in the development of high-throughput screening systems to expedite drug discovery.

Biomarkers are also relevant for selecting the most appropriate test species, since the absence of a significant marker is likely to be indicative of differences in the

activities of a drug in humans and the test species. The use of biomarkers as surrogate endpoints can identify the commonality between *in vitro* and *in vivo* effects, leading to dramatic improvements in the power of meta-analysis and of other approaches, where the goal is to assemble and evaluate value of information from various sources, such as animal experiments, human studies, and *in vitro* studies. This is, of course, subject to the very careful selection of the biomarkers to be used. One consideration is a need to readily detect changes in levels of biomarkers and to be able to relate such changes to drug-related efficacy or toxicity. In the interests of animal welfare and experimental convenience, metabolites and other small molecules that can be detected noninvasively in blood, stool, hair, or urine samples are preferred. This is because such biomarkers can be continually monitored during an experiment, and large numbers of samples can be collected, thereby allowing statistically reliable data to be generated from a small number of animals. However, most importantly, such biomarkers can also be readily analyzed in body fluid and stool samples taken from participants in clinical trials and can be used in the conduct of postmarket surveillance of large cohorts of patients.

***Omics Toward Biomarker Discovery*** *Genomics* can be used to identify new targets, since it is now possible to analyze changes in the transcription of >20,000 genes in various cell types and tissues in multiparameter experiments. There are two key approaches: (1) the detection of cDNAs that correspond to the entire protein-encoding gene; and (2) the use of RNAi to determine the relevance of mRNA levels in terms of protein expression and gene function. Recently, these approaches have been used to assess the influence of noncoding RNAs on specific biochemical pathways in mammalian cells [101]. The application of genomics has already proved useful in fundamental research and for gaining a mechanistic understanding of drug activity. However, so far, it has made only limited contributions to drug discovery [1]. Similarly, *toxicogenomic* approaches, in which global changes in gene expression resulting from exposure of specific cells or tissues to a test substance are identified, have been of limited use up to now. Nevertheless, the microarray or serial analysis of gene expression, by using commercially derived or custom-made gene chips, such as those from Affymetrix (www.affymetrix.com/), is a technique that is already proving useful for the identification of markers of hematotoxicity and carcinogenicity [102, 103].

*Proteomics* is an approach that involves measuring total cellular protein and determining the posttranslational modification and fate of proteins. The most applicable techniques involve spotting cell lysates onto arrayed antibody wells to search for potential biomarkers and to profile molecular pathways. In this way, much useful information on specific protein–protein interactions has been gained by using yeast-2-hybrid systems, and coimmunoprecipitation or other affinity-based methodologies. The antibodies required for this purpose are being generated and characterized by the Human Proteome Organization (HUPO) [104]. The reason why data from proteomics can be more relevant than the information from genomic analysis is that changes in mRNA levels do not necessarily correlate with changes in protein expression. Furthermore, it is clear that a drug might alter the protein recycling/degradation/synthesis rates and/or the extent, as well as the nature, of posttranslational modification. However, proteomics is limited by the fact that some classes of proteins are difficult to resolve (e.g., membrane proteins) and it is currently not amenable to high-throughput formats.

*Metabonomics* is defined as "the quantitative measurement of the dynamic multiparametric metabolic response of living systems to pathophysiological stimuli or genetic modification" [105]. It has traditionally required the use of analytical methods such as MS and NMR, in conjunction with separation techniques such as gas or liquid chromatography, which are primarily suited to the resolution of low molecular weight metabolites. Although NMR is generally less sensitive than MS-based methods, newer techniques such as magic angle spinning NMR cater for the direct metabolic profiling of as little as 20 mg of tissue or biofluids [106, 107]. Typically, drug testing by metabolic profiling involves the collection of body fluids from animals or human volunteers and is conducted alongside drug metabolism and PK studies.

In the case of animal experiments, many variables have complicated the interpretation of data from metabonomic studies in animals. Such variables include the species and strain of animal, gender, age, and the influence of diurnal variations and environmental factors and stressors. Other factors include the choice of dosing vehicle and the acclimatization period prior to the conduct of studies in metabolism cages. Nevertheless, the development of increasingly sophisticated analytical techniques (see later discussion) has permitted the use of smaller group sizes and reductions in the amounts of body fluids required, and reductions in the need for invasive studies, particularly where metabolic profiling is conducted on urine samples.

The main advantage of metabonomics over proteomics and genomics is that there is a much stronger link to a tangible physiological response to a drug, because it is not reliant on potentially delayed effects on protein expression, but rather on the measurement of biochemical changes that result from the modulation of specific pathways.

### 6.2 Microfluidics Systems

A microscale cell culture system that consists of a series of chambers each containing mammalian cells connected by a fluid network can be used to assess drug dynamics as a quasi *in vivo* surrogate system [108]. The development of robotic DNA printing systems [109] that allow cells to be transformed *in situ*, and adaptations that facilitate the growth and maintenance of such microfluidics systems, without the need for plating, can, as a result, be used for longer term as well as real-time studies [110]. The use of human cells instead of animal cells greatly increases the relevance of such systems to human safety.

### 6.3 Enabling Technologies for Multiparameter Studies

All the approaches discussed generate such a large amount of information that data analysis and interpretation can be very complicated, so complex analytical methods are needed to integrate and analyze the results. One of the most useful of these methods is pattern recognition. Pattern recognition algorithms differ widely in the ways in which data are analyzed and used to develop new hypotheses. The further development and application of these methods (reviewed in Ref. 111) could transform the ways in which systems biology approaches will facilitate the use of information from omic-based studies, data mining, pathway informatics, and other types of study, in order to gain an overview of drug action. At present, more than 150 databases, data analysis tools, and software suites are already available for this purpose [112]. These bioinformatics platforms mainly comprise commercially available clustering techniques that enable the extraction of exposure-related infor-

**TABLE 8 Pathway Signal Databases and Generation Tools for Systems-Based Drug Development**[a]

| |
|---|
| **Pathway Tools** |
| MetaCore http://metacore.com/web/index.php |
| A pathway modeling database |
| Pathway Studio www.ariadnegenomics.com |
| Assists with the interpretation of omics data and the generation of pathways that can be updated with information from other sources |
| Pathway Analyst http://path-a.cs.ualberta.ca/ |
| Predicts pathways in which a target protein may be involved |
| Gepasi http://www.gepasi.org/ |
| Software for modeling biochemical pathways; incorporates biokinetic simulation tools |
| Jdesigner http://sbw.kgi.edu/software/jdesigner.htm |
| A system that allows biochemical networks to be drawn |
| **Whole cell biochemical pathways** |
| Genomatica http://www.genomatica.com/index.shtml |
| A collection of models for whole cells and organisms that displays gene, protein, and biochemical networks |
| E-Cell http://www.e-cell.org/software |
| A simulation, modeling, and analysis tool for application to complex biological systems |
| **Whole tissue/organ pathway tools** |
| Human Physiome Project https://www.bioeng.auckland.ac.nz/physiome/physiome_project.php |
| A computational database that aims to build integrative models of biological systems, and whole organ/tissue structure and functions |
| **Clinical data** |
| Entelos http://www.entelos.com/ |
| For virtual patient simulations based on normal and disease physiology, to assist in deciphering disease-related pathways |

[a]The rapidly growing number of resources is impossible to list in full. However, these are some key resources that can be used during drug discovery and development.

mation on changes in proteins, in metabolites, or in gene expression, and can overlay information from different sources to find common hotspots or to identify biochemical networks. The length of the branch corresponds to the similarities between datasets. To further reduce the number of false-positive biomarkers, a common problem with all omics-based technologies, condition-specific algorithms, modeling approaches, and data-mining software packages have recently been developed (reviewed in Ref. 113; Table 8). There are also several tools that assist with the weighting of information from these varied approaches (Table 8).

## 7 COMPUTATIONAL METHODOLOGY USED IN EARLY PRECLINICAL DRUG DEVELOPMENT

### 7.1 Combinatorial Chemistry and SAR

By avoiding the incorporation of particular chemical groups into a drug, the likelihood of toxicity can be greatly reduced. Computational methods are used at the

very early stages of drug development to analyze the behavior of virtual chemical libraries, in terms of target specificity and affinity, or in terms of transport across physiological barriers, metabolism, or the causation of off-target modulation [114]. These predictions can rely on either a local set of data for a limited number of structurally related chemicals or global models. The latter require a large training set for refining a computational predictive tool and are generally able to identify alerts that may result in toxicity with greater success than locally trained tools (i.e., those based on a small set of known compounds). Affinity receptor binding data are very useful for the development of predictive computational models of closely related analogues of a drug molecule, as is information on the natural ligand derived from *in vitro* target–ligand binding studies.

An ability to generate an accurate model of the target protein is also very useful. Such models can be created *ab initio* or by homology-based modeling. In fact, the protein target models themselves are generated by using several approaches, the most popular of which combines primary sequence analysis for structural signatures that indicate probable protein folding with X-ray diffraction and crystallographic studies on a protein of the same class. A variety of molecular docking programs can be used, and most algorithms are based on an assumption that the target (generally a protein) is a rigid structure, while the ligand is flexible. The notable exceptions are force-field-based methods, such as molecular dynamics and Monte Carlo simulations, which allow for flexibility in both the target and the ligand. This approach accommodates the proposed "induced fit" model for most peptide and protein ligands. Fragment-based incremental methods are similarly suitable for assembling ligands from fragments that are incrementally docked into the target.

QSAR modeling is useful when each member of the same class of chemicals (sharing key structural elements and physicochemical properties) acts by a common mechanism. It involves generating rules from a training set of molecules based on their structures and known biological activities. These rules are then used to predict the biological activity of a novel candidate molecule with structural features that fall within the applicability domain of the model. For each docking application, a scoring system to identify virtual hits and a rule base that can accommodate experimental SAR data, are essential. The reader is referred to an excellent review on this topic [115] (see also Table 9).

### 7.2 *In Silico* Prediction of ADME and Toxicity

Adverse drug reactions (ADRs) remain one of the major reasons why a drug that has passed through preclinical testing will eventually be dropped following clinical trials. Efforts have recently been made to develop computer-based prediction tools and other such expert systems, for use in assessing the ADR potentials of drugs. Although there are four general types of expert systems, all of them are built up from experimental toxicity data. Hence, the early prediction of whether a drug is likely to have adverse effects can be used to refine compound libraries prior to high-throughput hit generation and lead development. DEREK (Deduction of Risk from Existing Knowledge) is a knowledge-based system, which focuses on molecular substructures (or "alerts") associated with known toxicological endpoints [116, 117]. It is able to predict three different endpoints, including mutagenicity, for which it was found to be 84% correctly predictive for a set of 226 chemicals. CASE

**TABLE 9 Examples of QSARs that Can Be Used During Lead Discovery and Optimization**[a]

| Database | Type of Information |
|---|---|
| ChemTree http://www.goldenhelix.com/chemtreesoftware.html | Statistical analysis QSAR for chemical library refinement |
| QikProp http://www.schrodinger.com | ADME prediction QSAR for lead generation and optimization |
| Catalyst http://www.accelrys.com/products/catalyst/ | 3D QSAR and query-based database management tool |
| CoMFA (Comparative Molecular Field Analysis) http://www.chem.ac.ru/Chemistry/Soft/COMFA.en.html | 3D QSAR technique based on data from known active molecules; applicable when the structure of the target protein is unknown |
| VolSurf http://www.moldiscovery.com/soft_volsurf.php | Produces 2D molecular descriptors from 3D molecular interaction energy grid maps, for the optimization of *in silico* PK properties (eADME or IS-DMPK) |

[a]More information about these methods and how they can be applied is available from http://www2.defra.gov.uk/research/project_data/more.asp?!=CB01067&M=KWS&V=reach&scope=0. Website links accessed 30 January 2007.

(Computer Automated Structure Evaluation) is an automated rule induction system [118]. It generates its own structural alerts by breaking down each chemical into all its possible fragments that are then classified as biophores (associated with toxicity) or biophobes (not associated with toxicity). CASE has been adapted by the FDA to provide MCASE QSAR-ES, which is able to predict carcinogenicity with an accuracy of 75% [119]. TOPKAT is an automated QSAR system [120] that uses large amounts of data for diverse compounds to build a set of descriptors. Decision tree-based methods are typified by OncoLogic [121], which, as its name suggests, is also used to predict carcinogenicity. The toxicological endpoints predicted by these and other such computer-based tools are summarized in Table 10. However, it should be noted that the predictivity of these tools is strictly determined by the quality of the molecular descriptors and the training set used to generate the prediction algorithm. In many cases, it may be possible to increase the predictive power of these and other such expert systems, by using more than one expert system for each endpoint. For instance, when 14 carcinogens were submitted to COMPACT and HazardExpert, used separately they predicted 10 and 8 actual carcinogens as carcinogenic, respectively, but when the two methods were used together, all 14 carcinogens were identified [122]. Tools such as INVDOCK, which combine molecular docking with structural alert profiling, may also be useful tools [123].

### 7.3 Prediction of Tissue-Specific Exposure: PBPK Modeling

A fundamental problem when using hazard data for risk assessment is the need to relate the effects detected at the dose level administered to a test system (the external dose) with the effects that would be caused by the dose that actually reaches the target in humans (the internal dose). The internal target organ dose can be predicted by undertaking toxicokinetic studies. Physiologically based biokinetic

**TABLE 10 Examples of Computer-Based Expert Systems Available for ADME and Toxicity Predictions**

| Name | Type of Expert System | Website | Some of the Endpoints Predicted[a] | |
|---|---|---|---|---|
| CASE/ MCASE/ CASETOX | Automated rule based | www.multicase.com | • Carcinogenicity<br>• Teratogenicity<br>• Mutagenicity | • Acute toxicity<br>• Chronic toxicity |
| DEREK | Knowledge based | www.chem.leeds.ac.uk/luk | • Teratogenicity<br>• Mutagenicity<br>• Respiratory sensitization | • Carcinogenicity<br>• Skin irritation<br>• Skin sensitization |
| HazardExpert | Knowledge based | www.compudrug.com | • Oncogenicity<br>• Mutagenicity<br>• Teratogenicity<br>• Immunotoxicity<br>• Neurotoxicity | • Membrane irritancy/ sensitivity<br>• Bioavailability<br>• Bioaccumulation |
| TOPKAT | Automated rule-based QSAR | www.accelrys.com | • Carcinogenicity<br>• Mutagenicity<br>• Developmental toxicity<br>• Skin sensitization | • Eye irritancy<br>• Biodegradability<br>• Acute toxicity<br>• Chronic toxicity |
| OncoLogic® | Decision tree approach | http://www.epa.gov/oppt/cahp/actlocal/can.html | • Carcinogenicity | |
| COMPACT | Knowledge-based QSAR | www.surrey.ac.uk/SBMS | • Carcinogenicity via CYP1A-related and CYP2E-related metabolic activation | |

[a]The endpoints that can be predicted using each expert system are based on the current status of each system. More information about these methods and how they can be applied is available from http://www2.defra.gov.uk/research/project_data/more.asp?!=CB01067&M=KWS&V=reach&scope=0. Website links accessed 30 January 2007.

(PBPK) modeling is a way of predicting ADME *in vivo* by combining results from the literature and from computational techniques [124], and by extrapolating data from *in vitro* studies between species. Some key PBPK methods are listed in Table 11. Plasma protein binding can have a crucial bearing on the internal dose, since it determines the free (unbound) concentration (this also has a bearing on the composition of cell culture media) [125]. Differential equations can be derived, which, when solved, provide information of relevance to humans. A better, biologically based, dose–response model of *in vivo* toxicity can then be developed from external dose data.

Significant advances are currently being made in biokinetic modeling, including the development of software programs and databases for the rapid generation of new models (http://www.hsl.gov.uk/capabilities/pbpk-jip.htm). These will improve the usefulness of this approach for evaluating large numbers of chemicals and will assist with the interpretation of *in vitro* hazard predictions for risk assessment purposes.

**TABLE 11 Examples of Programs for the Prediction of Biokinetic Properties**

| Name | Supplier | Website | Properties Predicted[a] |
|---|---|---|---|
| Cloe PK® | Cyprotex | www.cyprotex.com | • Potential exposure<br>• Absorption from GI tract<br>• Plasma, tissue, and organ concentrations<br>• Renal excretion<br>• Hepatic metabolism |
| iDEA pkEXPRESS™ | LION Bioscience | www.lionbioscience.com | • Absorption from GI tract<br>• Systemic circulation<br>• Bioavailability<br>• Plasma concentration<br>• Elimination |
| Megen100 | Health & Safety Lab | www.hsl.gov.uk/capabilities/pbpk.htm | • Oral and intravenous absorption<br>• Concentration–time profiles for plasma and major organs and tissues<br>• Hepatic metabolism |
| PK-Sim® | Bayer Technology Services | www.bayertechnology.com | • Oral absorption<br>• Concentration–time profiles for plasma and major organs and tissues<br>• Bioavailability<br>• Renal and billiary excretion |

[a]More information about these methods and how they can be applied is available from http://www2.defra.gov.uk/research/project_data/more.asp?!=CB01067&M=KWS&V=reach&scope=0.

One key issue is the role of transporter proteins in the absorption and uptake of a drug or its metabolites, as this will determine the internal dose to which any particular organ is exposed. Problems with crystallizing membrane proteins have made it inherently difficult to generate 3D models of many important transporter proteins. Homology and comparative modeling can, however, be used to generate models by reference to experimental data, the structure of related proteins, and more fundamental predictions of protein folding and tertiary structure, made from the primary amino acid sequences of relevant proteins. For instance, three ATP-binding cassette (ABC) transporters of bacterial origin have been crystallized, and from these structures, comparative modeling has been used to generate models of ABC transporters that play a key role in drug efflux in humans. Another approach is to generate a pharmacophore from experimental data for known transporter binding molecules, as in the case for monoamine transporters, then to use QSAR systems to make

predictions as to whether a novel drug is likely to be a transported. These approaches are discussed in more detail in Ref. 126.

## 8 ANIMAL MODELS USED IN PRECLINICAL TESTING OF PHARMACEUTICALS

### 8.1 Selection of a Suitable Test Species

Although not an absolute regulatory requirement, it is still widely accepted that drug development should involve initial studies conducted in a rodent species, followed by studies in a nonrodent species. In addition, further studies may be conducted in other species, such as the rabbit (e.g., for local tolerance and for pyrogenicity). There is often flexibility in the nature of the tests required by most regulators for a new medicine's dossier. For instance, where only a single species can be shown to be relevant, regulators may be willing to consider information from nonanimal studies to decide whether the two-species testing requirement can be waived.

The Universities Federation for Animal Welfare (UFAW) has published a handbook that lists the major characteristics of laboratory species, to which the reader is referred [127]. Perhaps of more general importance is the use of existing preclinical, clinical, and basic research data in the species selection process. Information from mechanistic and early *in vitro* studies is particularly important, as is understanding differences between species in the spatial and temporal expression of the drug target and in target modulation and activity. The starting point is understanding any significant differences between the pharmacology of the human protein and the possible test species. Sequence homology and structural homology do not guarantee the functional equivalence of species homologs. Once functional equivalence has been established by using cell expression systems, the next stage is to understand the tissue distribution of the target in the selected test species. This can involve *in vivo* studies with target-specific probes such as fluorescent or radiolabeled antibodies or ligands. In general, however, this information is more readily attainable from tissue distribution studies, including immunohistochemistry. A good illustration of how this can be important was made in a recent paper, which highlighted differences between the human and rat versus the mouse in the distribution and pharmacology of serotonin receptors, which showed that the mouse is a poor preclinical model for some classes of antipsychotic agents [128]. The rat was a more suitable rodent model, although the mouse has already been used in some preclinical studies for CNS-active serotoninergic drugs. With the greater availability of genomic information, species selection (or, indeed, the selection or generation of an appropriate GM animal model) can be made on the basis of amino acid sequence, as is the case when developing humanized antibodies. In some cases, this may obviate the need to conduct extensive studies in higher order vertebrates, such as primates.

In addition to differences in target expression, differences in plasma clearance rate and routes of excretion can also have a crucial bearing on whether a particular species or strain is used [129]. Differences between anatomical and functional properties associated with the intended route of administration (as in the case of nasal absorption, see Section 3.2), functional differences between species homologs of

efflux and transporter proteins such as P-gp [130], and differences between the distribution and activities of key metabolic enzymes predicted to be involved in the biotransformation of a drug may limit the choice of test species.

Where more than one species is likely to be used for preclinical studies, the choice of species may also involve a consideration of the need for allometric scaling. This allows the prediction of human pharmacokinetics from the pharmacokinetics of a drug in a test species, by taking body weight into account. A recent study [131] indicates that although allometric scaling is most reliable when PK parameters are available for five species, certain three-species combinations (such as mouse/rat, monkey, and dog) are adequate, whereas other three-species combinations (such as rabbit, monkey, and dog) are significantly worse. In general, two-species combinations are poorer for making allometric predictions, and certain combinations, such as mouse and rat, rabbit and monkey, and dog and monkey, are particularly poor. However, in practice, two-species combinations are commonly used, which usually involve a rodent and a nonrodent species.

Species selection is also dependent on the feasibility of a study and on the availability and cost of acquiring and caring for the animals concerned. Developmental toxicity, for instance, could not be reasonably or ethically conducted on primates or some other vertebrate species and is generally conducted in rodents, because of their small size, shorter life spans, and the larger litter sizes produced at each generation. Preliminary studies may be conducted with fish larval forms and invertebrates.

## 8.2 Experimental Design

The ICH guidelines advocate the principles of the three Rs. That is, prior to the use of vertebrate animals, scientists are required to consider: (1) whether there are alternatives to using animals that can provide information that is as valid and acceptable to regulators as animal data; (2) where there are no replacement alternatives, how prior information and information from nonanimal experiments could be used to reduce the number of animals used; and (3) how animal experiments could be refined in order to reduce suffering (e.g., by using more humane endpoints).

The first step in designing an animal experiment is to clearly define its objectives in terms of the nature of information sought or the hypothesis to be tested. The correct grouping of animals according to their treatment (nature and frequency of intervention, etc.) and husbandry regime (light–dark cycle, handling, monitoring, etc.) reduces variation in experimental data. Standard operating procedures should be developed that take into account the scientific objectives and animal welfare issues associated with each experiment. For instance, all the animals (control, vehicle control, and test animals) should be handled with the same frequency, for the same periods of time and by the same technician, and subjected to identical procedures. Other causes of variation stem from inadvertent infections, which can be minimized by good laboratory practice, routine health surveillance, and using suitably ventilated cages. Whether food and fluid control (i.e., limiting daily supply), timed feeding and water supply, or *ad libitum* feeding and drinking are appropriate can also be an important consideration. The effects of weight, age, sex, and/or strain on the experimental outcome can have important implications for experimental design and can be accounted for in various ways, for example, by appropriate grouping of animals

(some examples are given in http://oslovet.veths.no/compendia/LAS/KAP28.pdf). This is because different strains of rats can vary dramatically in their clinical responses to a drug, so selecting the most appropriate isogenic strain can be difficult and it might sometimes be appropriate to use more than one strain in each experiment. There are a number of useful resources on this topic (e.g., Ref. 132).

The number of possible variables highlights the importance of a pilot study. Pilot studies that are designed on the basis of existing information can be used to identify logistical, animal welfare, and scientific problems and to address specific scientific questions, prior to the conduct of larger animal studies [133]. Whether an experiment entails multiple endpoints, different sexes, and/or different strains, the interpretation of the data might be compromised without prior and careful statistical planning, or in the worst case scenario, be impossible, resulting in the need to repeat the experiment.

The ICH guidelines recommend carcinogenicity testing, where the weight of evidence suggests it may be required or where there is insufficient evidence to rule out the possibility of carcinogenic hazard. Such testing normally should be conducted only in rats, as there is evidence to suggest that the rat bioassay is more predictive of human risk Cancer-associated biomarkers have an enormous potential to streamline preclinical drug development and can also be used to devlop humane endpoints and facilitate temporal studies in small numbers of animals [134].

## 9 USE OF PRIOR INFORMATION

### 9.1 Sources of Prior Knowledge

Information retrieval is an important aspect of drug development, and a list of databases that provide information that can be used is given in Table 12. The tremendous rate at which information is increasing in volume and is diversifying requires that computational tools are developed to support the extraction and ranking of information according to its relevance and reliability. In addition, models of how different elements of a biochemical pathway interact can be derived mathematically, and such models have been used to successfully construct biochemical networks of relevance to drug discovery [135].

Some special sources of prior information include data suites that collate information about specific biochemical pathways, specific diseases, human genetics, and human sub-populations. Because most diseases have complex etiologies, particularly diseases such as cancer, heart disease, stroke, and diabetes, which arise from a combination of lifestyle, environmental, and genetic factors, several large-scale population studies are being undertaken (some examples are given in Table 12). These studies may contribute to a greater understanding of the complex basis of such diseases, leading to improvements in drug discovery and target selection where patterns that link genetic differences to drug effects or disease can be established.

Perhaps the most directly relevant information is that available from preclinical and clinical studies on existing therapeutic products. Usually, this information is not publicly available, but some confidential information is being incorporated into many databases. PharmaPendium™ [136] is an example of a recent public resource that includes drug safety data (preclinical and clinical) for FDA-approved drugs. Such resources are likely to prove invaluable for researchers and regulators alike.

**TABLE 12 Information Resources**[a]

| | |
|---|---|
| *Omics Databases and Resources* | |
| SRS http://srs.ebi.ac.uk/srsbin/cgi-bin/wgetz?-page+srsq2+-noSession | A gene/protein sequence retrieval system that can be used to browse various biological sequence and literature databases |
| ToxExpress http://www.genelogic.com/genomics/toxexpress/ | A toxicogenomic profiling suite that can be used in biomarker discovery |
| MIAME http://www.mged.org/Workgroups/MIAME/miame.html | Minimum Information About a Microarray Experiment: criteria needed to permit the interpretation of the results of the experiment unambiguously and, potentially, to reproduce the experiment |
| Array Track http://www.fda.gov/nctr/science/centers/toxicoinformatics/ArrayTrack/ | Developed by FDA National Center for Toxicological Research (NCTR); an integrated suite designed to manage, analyze, and interpret microarray data |
| KEGG http://www.genome.jp/kegg/ | Kyoto Encyclopedia of Genes and Genomes provides a complete computer representation of the cell, the organism, and the biosphere |
| Unigene http://www.unigene.com | Provides an organized view of the transcriptome |
| SNP http://www.ncbi.nlm.nih.gov/projects/SNP/ | The Single Nucleotide Polymorphism database |
| SPAD http://www.grt.kyushu-u.ac.jp/spad/ | The Signaling PAthway Database (SPAD): an integrated database for genetic information and signal transduction systems |
| *Pharmacology, Toxicology, and Biological Systems* | |
| BIOPRINT http://www.cerep.fr/cerep/users/pages/Collaborations/bioprint.asp | A pharmacology and ADME database that contains *in vitro* pharmacology profiles, ADR, PK and clinical data for over 2500 marketed drugs, failed drugs, and reference compounds |
| BioRS http://biors.gsf.de:8111/searchtool/searchtool.cgi | A biological data retrieval system |
| CEBS http://www.niehs.nih.gov/cebs-df/index.cfm | Chemical Effects in Biological Systems: a knowledge base for information and resource exchange |
| BIND http://bond.unleashedinformatics.com | Biomolecular Interaction Network Database: designed to store full descriptions of interactions, molecular complexes and pathways |
| BioCarta http://www.biocarta.com/ | Provides interactive graphic models of molecular and cellular pathways |
| BRENDA www.brenda.uni-koeln.de | A collection of enzyme functional data |
| CSNDB http://geo.nihs.go.jp/csndb/ | Cell Signaling Networks Database: a database and knowledge base for signaling pathways of human cells |
| SwissProt http://expasy.org/sprot/ | A protein sequence database with descriptions of the function of proteins, protein structure, posttranslational modifications, variants, etc. |
| TransPath http://www.biobase-international.com/pages/index.php?id=transpath | Provides information about (mostly mammalian) signal transduction molecules and reactions, focusing on signaling cascades that change the activities of transcription factors and thus alter the gene expression profiles of a cells |

**TABLE 12** ***Continued***

| | |
|---|---|
| PathArt http://jubilantbiosys.com/ppa.htm | A database of biomolecular interactions with tools for searching, analysis and visualization of data |
| DSSTox http://www.epa.gov/ncct/dsstox/index.html | The EPA's Distributed Structure-Searchable Toxicity Database for improved structure–activity and predictive toxicology capabilities |
| TOXNET http://toxnet.nlm.nih.gov/ | A database on toxicology, hazardous chemicals, environmental health, and toxic releases |
| PharmGKB http://www.pharmgkb.org/ | A database on relationships among drugs, diseases, and genes |
| | *Literature Database* |
| PubMed www.pubmed.com | A database that includes over 16 million citations from MEDLINE and other life science journals for biomedical articles back to the 1950s; PubMed includes links to full text articles and other related resources |
| Bio-Frontier P450/CYP http://www.fqs.pl/ | A database for testing CYP interactions |
| | *Human Population Genetics and Toxicity Databases and Resources* |
| The Collaborative on Health and the Environment http://database.healthandenvironment.org/ | A searchable database of links between chemical contaminants and human diseases |
| The Personalized Medicine Research Project (Marshfield Project) http://www.marshfieldclinic.org/chg/pages/default.aspx | A human population genetic database to understand the interplay of human genetics, diseases, and environmental factors |
| Medgene$^{SM}$ database http://hipseq.med.harvard.edu/MEDGENE/login.jsp | A database of disease-associated genes |
| CARTaGENE http://www.cartagene.qc.ca/ | A source of information on the genetic variation of a large population |
| Latvian Genome Project http://bmc.biomed.lu.lv/gene/ | Large-scale human population genetic project to discover disease linkages |
| Estonian Genome Project http://www.geenivaramu.ee/ | A source of information on the genetic variation of a large population |
| The United Kingdom Biobank (UK Biobank) http://www.ukbiobank.ac.uk/ | Genetic and medical information is being collected for 500,000 UK volunteers |
| Translational Genomic Research in the African Diaspora (TgRIAD) http://www.genomecenter.howard.edu/TGRIAD.htm | A database to understand disease, genetics, and environmental factor linkage in people of African descent |
| Obesity gene map database http://obesitygene.pbrc.edu/ | A database of genetic markers associated with obesity |
| COGENE the Craniofacial and Oral Gene Expression Network http://hg.wustl.edu/cogene/ | A consortium that looks at the genetics of early development, in particular, craniofacial disorders |

**TABLE 12** ***Continued***

| | |
|---|---|
| Human genome variation database http://hgvbase.cgb.ki.se/ | Contains links to a number of single nucleotide polymorphism databases for particular diseases |
| The International HapMap Project http://www.hapmap.org/ | A consortia aimed at finding genes associated with human disease and response to pharmaceuticals |
| GenomEUtwin http://www.genomeutwin.org/ | A database of human population genetics aimed specifically at finding genetic and lifestyle linkages to disease that involves studies on twins |
| Public Population Project in Genomics (P3G) http://www.p3gconsortium.org/ | A consortium that aims to develop a human population genetics database |

[a]This is a list of resources that are applicable to understanding the output of omics-based studies and compiling a systems biology view of diseases and drug effects. Websites accessed 30 January 2007.

Other databases include the Adverse Drug Effects database [137], which stores information on approved drugs, including the severity and incidence of adverse effects, which is relevant to the discovery and design of new clinical products.

### 9.2 Standardization of Data Collection and Meta-Analysis

The quality and completeness of the available toxicological data will significantly affect the level of confidence in the preclinical data. The application of Good Laboratory Practice (GLP) should increasingly help to standardize the way in which experiments are designed, conducted, and reported, thereby improving the quality of the information available for guiding subsequent studies.

Meta-analysis is a statistical approach, which is used to combine data from different sources, but it needs to be applied with great caution. It is particularly difficult to use meta-analysis when different datasets contradict one another. Nevertheless, the use of surrogate endpoints, such as biomarkers, may dramatically improve the power of meta-analysis, since appropriate biomarkers can be used to increase the credibility of animal and human cell-based and tissue-based preclinical studies, and to facilitate extrapolation between such *in vitro* studies and preclinical *in vivo* studies, in animals and in humans.

Several potential biomarkers of exposure and toxicity can be considered. For example, such a scheme was originally proposed by Sobels [138] for the extrapolation of data on genetic damage from animals to humans, and was subsequently modified by Sutter [139] to permit *in vitro–in vivo* extrapolation. In some cases, threshold doses can be set, solely on the basis of *in vitro* tests (e.g., for some genotoxins). This parallelogram approach can then be used to extrapolate preclinical data to effects on humans, according to the paradigm:

$$\text{Human } (\textit{in vivo}) \text{ toxicity} = \frac{[\text{Rodent } (\textit{in vivo}) \text{ toxicity} \times \text{Human } (\textit{in vitro}) \text{ toxicity}]}{\text{Rodent } (\textit{in vitro}) \text{ toxicity}}$$

The concept assumes that the ratio of *in vitro* toxicity to *in vivo* toxicity for any particular endpoint is broadly comparable across species. Up to now, this approach has been used for extrapolating data on genetic damage but has proved to be less useful for extrapolating other forms of toxicity data to humans, because of the relatively complex mechanisms of toxicity that are involved. Nevertheless, in a recent paper [140], the concept was applied to a comparison of rat and human skin penetration rates *in vitro* and to predicting the *in vivo* effects of topically applied substances. As key biomarkers for drug effect and toxicological endpoints become available, the applicability of this approach is likely to expand to other areas of drug development and safety pharmacology.

## 10 CONCLUSIONS

Despite decades of research and development, the issue of adverse drug reactions that result in drug withdrawals remains a significant problem. This problem is confounded by the fact that information from clinical studies on human volunteers and patients is often kept from public scrutiny. Indeed, only a small number of pharmaceutical companies post their clinical trials information on publicly available registers. A recent study indicates that target organ toxicities in humans are not always predicted reliably by preclinical tests in animals. The predictivity of cardiovascular, hematopoietic, and gastrointestinal toxicity is around 80% but is lower for toxicity to the liver, skin, and nervous system [141]. Hence, there is an urgent need for a new approach to drug development, which involves the targeted use of new and advanced technologies that are based on defined cell systems, either as standalone alternatives to animal studies or as tools to assist with the extrapolation of animal data to humans.

Indeed, in 2004, the FDA produced a report that suggested that the fall in drug development returns was due largely to the failure to use the new technologies such as genomics, proteomics, and bioinformatics platforms to detect safety problems that cannot be identified in the more traditional animal-based methods. These newer, and often systems biology-based, approaches hold enormous potential in this respect but are very much in their infancy. One of the most significant problems is the difficulty in standardizing and validating these new technologies, in order to ensure that the quality of data and the quality of data analysis form a suitable basis for safety assessments. These systems are being developed at an unprecedented rate. For example, a consortium of global pharmaceutical giants has been assembled to put forward biomarkers and screening assays for consideration by the FDA. It remains to be seen whether this initiative will reduce the current drug attrition rate. Preclinical planning must look at both the existing regulatory requirements and the scope for cost- and time-effective studies that make the maximum use of the new and exciting technologies.

## ACKNOWLEDGMENT

The authors would like to thank Dr Michelle Scrivens for her valued assistance in the preparation of this chapter.

## REFERENCES


1. Challenge and Opportunity on the Critical Path to New Medical Products. US FDA March 2004. Available from http://www.fda.gov/oc/initiatives/criticalpath/whitepaper.html (accessed 29 January 2007).
2. *Guidance for Industry: Fast Track Drug Development Programs—Designation, Development, and Application Review*. Rockvill, MD: US Department of Health and Human Services Food and Drug Administration, Center for Drug Evaluation and Research (CDER), Center for Biologics Evaluation and Research (CBER); January 2006. Procedural Revision 2. Available from http://www.fda.gov/cber/gdlns/fsttrk.pdf (accessed 29 January 2007).
3. Schwieterman WD. Regulating biopharmaceuticals under CDE versus CBER: an insider's perspective. *Drug Discov Today* 2006;11(19–20):945–951.
4. Organisation for Economic Co-operation and Development. *Chemcials Testing Guidelines. Section 4: Health Effects*. Available from http://www.oecd.org/document/55/0,2340,en_2649_34377_2349687_1_1_1_1,00.html (accessed 29 January 2007).
5. Overington JP, Al-Lazikani B, Hopkins AL. How many drug targets are there? *Nat Rev Drug Discov* 2006;5(12):993–996.
6. Wilson S, Bergsma D. Orphan G-protein coupled receptors: novel drug targets for the pharmaceutical industry. *Drug Des Discov* 2000;17(2):105–114.
7. Cancer Genome Anatomy Project: National Cancer Institute. Available from www.ncbi.nlm.nih.gov/CGAP (accessed 29 January 2007).
8. Genomes on-line database version 2. Available from http://www.genomesonline.org/ (accessed 29 January 2007).
9. Romero P, Wagg J, Green ML, Kaiser D, Krummenacker M, Karp PD. Computational prediction of human metabolic pathways from the complete human genome. *Genome (Biol)* 2004;6(1): http://genomebiology.com/2004/6/1/R2 (accessed 29 January 2007).
10. Salzberg SL, White O, Peterson J, Eisen JA. Microbial genes in the human genome: lateral transfer or gene loss? *Science* 2001;292(5523):1903–1906.
11. Steinmetz LM, Scharfe C, Deutschbauer AM, Mokranjac D, Herman ZS, Jones T, Chu AM, Giaever G, Prokisch H, Oefner PJ, Davis RW. Systematic screen for human disease genes in yeast. *Nat Genet* 2002;31:400–404.
12. Hughes TR, Marton MJ, Jones AR, Roberts CJ, Stoughton R, Armour CD, Bennett HA, Coffey E, Dai H, He YD, Kidd MJ, King AM, Meyer MR, Slade D, Lum PY, Stepaniants SB, Shoemaker DD, Gachotte D, Chakraburtty K, Simon J, Bard M, Friend SH. Functional discovery via a compendium of expression profiles. *Cell* 2000;102:109–126.
13. Kumar A, Agarwal S, Heyman JA, Matson S, Heidtman M, Piccirillo S, Umansky L, Drawid A, Jansen R, Liu Y, Cheung K-H, Miller P, Gerstein M, Roeder GS, Snyder M. Subcellular localization of the yeast proteome. *Genes Dev* 2002;16(6):707–719.
14. Outeiro TF, Lindquist S. Yeast cells provide insight into alpha-synuclein biology and pathobiology. *Science* 2003;302:1772–1775.
15. Griffioen G, Duhamel H, Van Damme N, Pellens K, Zabrocki P, Pannecouque C, van Leuven F, Winderickx J, Wera S. A yeast-based model of α-synucleinopathy identifies compounds with therapeutic potential. *Biochem Biophys Acta* 2006;1762:312–318.
16. Versele M, Lemaire K, Thevelein JM. Sex and sugar in yeast: two distinct GPCR systems. *EMBO Rep* 2001;2:574–579.

17. Pausch MH, Lai M, Tseng E, Paulsen J, Bates B, Kwak S. Functional expression of human and mouse P2Y12 receptors in Saccharomyces cerevisiae. *Biochem Biophys Res Commun* 2004;324:171–177.
18. Fire A, Xu S, Montgomery MK, Kostas SA, Driver SE, Mello CC. Potent and specific genetic interference by double stranded RNA in *Caenorhabditis elegans*. *Nature* 1998;391:806–811.
19. Metzstein MM, Stanfield GM, Horvitz HR. Genetics of programmed cell death in *C. elegans*: past, present and future. *Trends Genet* 1998;14(10):410–416.
20. Nichols CD. *Drosophila melanogaster* neurobiology, neuropharmacology, and how the fly can inform central nervous system drug discovery. *Pharmacol Ther* 2006;112(3): 677–700.
21. Kellenberger E, Muller P, Schalon C, Bret G, Foata N, Rognan D. Sc-PDB: an annotated database of druggable binding sites from the protein data bank. *J Chem Inf Model* 2006;46:717–727.
22. Tapolczay D, Chorlton A, McCubbin Q. Probing drug structure improves the odds. *Curr Opin Drug Discov Dev* 2000;April:30–33.
23. Schreiber SL. Target-oriented and diversity-oriented organis synthesis in drug discovery. *Science* 2000;287(5460):1964–1969.
24. Villar HO, Yan J, Hansen MR. Using NMR for ligand discovery and optimization. *Curr Opin Chem Biol* 2004;8(4):387–391.
25. Swayze EE, Jefferson EA, Sannes-Lowery KA, Blyn LB, Risen LM, Arakawa S, Osgood SA, Hofstadler SA, Griffey RH. SAR by MS: a ligand-based technique for drug lead discovery against RNA targets. *J Med Chem* 2002;45(18):3816–3819.
26. Hartshorn MJ, Murray CW, Cleasby A, Frederickson M, Tickle IJ, Jhoti H. Fragment-based lead discovery using X-ray crystallography. *J Med Chem* 2005;48(2):403–413.
27. Lofas S. Optimizing the hit-to-lead process using SPR analysis. *Assay Drug Dev Technol* 2004;4:407–415.
28. Rees DC, Congreve M, Murray CW, Carr R. Fragment-based lead discovery. *Nat Rev Drug Discov* 2004;3:660–672.
29. Brenner S, Lerner RA. Encoded combinatorial chemistry. *Proc Natl Acad Sci USA* 1992;89:5381–5383.
30. Scheuermann J, Dumelin CE, Melkko S, Neri D. DNA-encoded chemical libraries. *J Biotechnol* 2006;126:568–581.
31. Landro JA, Taylor ICA, Stirtan WG, Osterman DG, Kristie J, Hunnicutt EJ, Rae PMM, Sweetnam PM. HTS in the new millennium. The role of pharmacology and flexibility. *J Pharmacol Toxicol Methods* 2000;44:273–289.
32. Kokko L, Johansson N, Lovgren T, Soukka T. Enzyme inihibitor screening using a homogenous proximity-based immunoassay for estradiol. *J Biomol Screen* 2005;10:348–354.
33. Kimberly H, Clement, JF, Abrahamyan L, Strebel K, Bouvier M, Kleinman L, Mouland AJ. A human immunodeficiency virus type 1 protease biosensor using bioluminescence resonance energy transfer. *J Virol Methods* 2005;128:93–103.
34. Thomsen W, Frazer J, Unett D. Functional assays for screening GPCR targets. *Curr Opin Biotechnol* 2005;16:655–665.
35. Westby M, Nakayama GR, Butler SL, Blair WS. Cell-based and biochemical screening approaches for the discovery of novel HIV-1 inihibitors. *Antiviral Res* 2005;67:121–140.

36. Mascini M, Bagni G, Di Pietro ML, Ravera M, Baracco S, Osella D. Electrochemical biosensor evaluation of the interaction between DNA and metallo-drugs. *Biometals* 2006;19:409–418.
37. Huber W. New strategy for improved secondary screening and lead optimization using high-resolution SPR characterization of compound-target interactions. *J Mol Recognition* 2005;18:273–281.
38. Hanes J, Pluckhun A. *In vitro* selection and evolution of functional proteins by using ribosomal display. *Proc Natl Acad Sci USA* 1997;94:4937–4942.
39. Silacci M, Brack S, Schirru G, Marlind J, Ettorre A, Merlo A, Viti F, Neri D. Design, construction, and characterization of a large synthetic human antibody phage display library. *Proteomics* 2005;5:2340–2350.
40. Boder ET, Wittrup KD. Yeast surface display for screening combinatorial polypeptide libraries. *Nat Biotechnol* 1997;15:553–557.
41. Patil SD, Rhodes DG, Burgess DJ. DNA-based therapeutics and DNA delivery systems: a comprehensive review. *AAPS* 2005;7(1):E61–E77.
42. Klumpp C, Kostarelos K, Prato M, Bianco A. Functionalized carbon nanotubes as emerging nanovectors for the delivery of therapeutics. *Biochim Biophys Acta* 2006;1758:404–412.
43. Proudfoot JR. Drugs, leads and drug-likeness: an analysis of some recently launched drugs. *Bioorg Med Chem Lett* 2002;12(12):1647–1650.
44. Makara GM, Athanasopoulos J. Ligand affinity binding in improving success-rates for lead generation. *Curr Opin Biotechnol* 2005;16:666–673.
45. Golebiowski A, Klopfenstein SR, Portlock DE. Lead compounds discovered from libraries: part 2. *Curr Opin Chem Biol* 2003;7:308–325.
46. Kola I, Landis J. Can the pharmaceutical industry reduce attrition rates? *Nat Rev Drug Discov* 2004;3:711–715.
47. Lipinski CA, Lambardo F, Dominy BW, Feeney PJ. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv Drug Deliv Rev* 1997;23:3–25.
48. Amidon GLH, Lennernäs H, Shah VP, Crison JR. A theoretical basis for a biopharmaceutic drug classification: the correlation of *in vitro* drug product dissolution and *in vivo* bioavailability. *Pharm Res* 1995;12(3):413–420.
49. Schmeichel KL, Bissell MJ. Modeling tissue-specific signaling and organ function in three dimensions. *J Cell Sci* 2003;116(12):2377–2388.
50. Leppert PS, Fix JA. Use of everted intestinal rings for *in vitro* examination of oral absorption potential. *J Pharm Sci* 1994;8:976–981.
51. Amidon GL, Sinko PJ, Fleisher D. Estimating human oral fraction dose absorbed: a correlation using rat intestinal membrane permeability for passive and carrier-mediated compounds. *Pharm Res* 1988;10:651–654.
52. Hidalgo IJ. Cultured intestinal epithelial cell models. *Pharm Biotechnol* 1996;8:35–50.
53. Irvine JD, Takahashi L, Lockhart K, Cheong J, Tolan JW, Selick HE, Grove JR. MDCK (Madin–Darby canine kidney) cells: a tool for membrane permeability screening. *J Pharm Sci* 1999;88:28–33.
54. Stewart BH, Chan OH, Lu RH, Rener EL, Schmid HL, Hamilton HW, Steinbaugh BA, Taylor MD. Comparison of intestinal permeabilities determined in multiple *in vitro* and *in situ* models: relationship to absorption in humans. *Pharm Res* 1995;12:693–699.
55. White RE. High-throughput screening in drug metabolism and pharmacokinetic support of drug discovery. *Annu Rev Pharmacol Toxicol* 2000;40:133–157.

56. Pidgeon C, Ong S, Munroe J, Hornback WJ, Kasher JS, Grunz L, Liu H, Qiu X, Pidgeon M, Dantzig A. IAM chromatography: an *in vitro* screen for predicting drug membrane permeability. *Med Chem* 1995;38:590–594.
57. Kansy M, Senner F, Gubernator K. Physicochemical high throughput screening: parallel artificial membrane permeation assay in the description of passive absorption processes. *J Med Chem* 1998;41:1007–1010.
58. Bubik M, Ott M, Mahringer A, Fricker G. Rapid assessment of P-glycoprotein–drug interactions at the blood–brain barrier. *Anal Biochem* 2006;358:51–58.
59. Gray AC, Malton J, Clothier RH. The development of a standardised protocol to measure squamous differentiation in stratified epithelia, by using the fluorescein cadaverine incorporation technique *Altern Lab Anim* 2004;32(2):91–100.
60. Gray TE, Guzman K, Davis CW, Abdullah LH, Nettesheim P. Mucociliary differentiation of serially passaged normal human tracheobronchial epithelial cells. *Am J Respir Cell Mol Biol* 1996;14(1):104–112.
61. Deli MAM, Abraham CS, Kataoka Y, Niwa M. Permeability studies on *in vitro* blood–brain barrier models: physiology, pathology and pharmacology. *Cell Mol Neurobiol* 2005;25:59–126.
62. Combes RD, Balls M, Bansil L, Barratt M, Bell D, Botham P, Broadhead C, Clothier R, George E, Fentem J, Jackson M, Indans I, Loizou G, Navaratnam V, Pentreath V, Phillips B, Stemplewski H, Stewart J. The Third FRAME Toxicity Committee: Working Toward Greater Implementation of Alternatives in Toxicity Testing. *Altern Lab Anim* 2002;32(S1):635–642.
63. Gumbleton M, Audus KL. Progress and limitations in the use of in vitro cell cultures to serve as a permeability screen for the blood–brain barrier. *J Pharm Sci* 2001;90:1681–1698.
64. Atterwill CK, Purcell WM. Brain spheroid and other organotypic culture systems in *in vitro* neurotoxicology. In Pentreath VW, Ed., *Neurotoxicology In vitro* Washington DC: Taylor and Francis; 1999, pp 213–238.
65. Harkema JR. Comparative structure, function and toxicity of the nasal airways. Predicting human effects from animal studies. In Gardner DE, Crapo JD, McCellan RO, Eds. *Toxicology of the Lung*. Washington DC: Taylor and Francis; 1999, pp 55–83.
66. Dimova S, Brewster ME, Noppe M, Jorissen M, Augustijns P. The use of human nasal *in vitro* cell systems during drug discovery and development. *Toxicol In Vitro* 2005;19:107–122.
67. Daly AK. Nomenclature for human CYP2D6 alleles. *Pharmacogenetics* 1996;6:193–201.
68. Coecke S, Rogiers V, Bayliss M, Castell JM, Doehmer J, Fabre G, Fry J, Kern A, Westmoreland C. The use of long-term hepatocyte cultures for detecting induction of drug metabolising enzymes: the current status: ECVAM Hepatocytes and Metabolically Competent Systems Task Force Report 1. *Altern Lab Anim* 1999;27(4):579–638.
69. Kim HS, Wainer IW. On-line drug metabolism in capillary electrophoresis. 1. Glucuronidation using rat liver microsomes. *Anal Chem* 2006;78:7071–7077.
70. Dunn JC, Tompkins RG, Yarmush ML. Long-term *in vitro* function of adult hepatocytes in a collagen sandwich configuration. *Biotechnol Prog* 1991;7:237–245.
71. Scmitmeier S, Langsch A, Bader A. A membrane-based small-scale bioreactor for accelerated *in vitro* drug screenings with primary hepatocytes. *Desalination* 2006;199:258–260.
72. Fearn RA, Hirst BH. Predicting oral drug absorption and hepatobiliary clearance: human intestinal and hepatic *in vitro* cell models. *Environ Toxicol Pharmacol* 2006;21:168–178.

73. Krebsfaenger N, Murdter TE, Zanger UM, Eichelbaum MF, Doehmer J. V79 Chinese hamster cells genetically engineered for polymorphic cytochrome P450 2D6 and their predictive value for humans. *ALTEX* 2003;20(3):143–154.

74. Boonstra J, Post JA. Molecular events associated with reactive oxygen species and cell cycle progression in mammalian cells. *Gene* 2004;337:1–13.

75. Prabhakaran K, Sampson DA, Hoehner JC. Neuroblastoma survival and death: an *in vitro* model of hypoxia and metabolic stress. *J Surg Res* 2004;116:288–296.

76. Canova N, Kmonickova E, Lincova D, Vitek L, Farghali H. Evaluation of a flat membrane hepatocyte bioreactor for pharmacotoxicological applications: evidence that inhibition of spontaneously produced nitric oxide improves cell functionality. *Altern Lab Anim* 2004;32:25–35.

77. Wong JM, Harper PA, Meyer UA, Bock KW, Morike K, Lagueux J, Ayotte P, Tyndale RF, Sellers EM, Manchester DK, Okey AB. Ethnic variability in the allelic distribution of human aryl hydrocarbon receptor codon 554 and assessment of variant receptor function *in vitro*. *Pharmacogenetics* 1991;1:66–67.

78. Schwarz UI. Clinical relevance of genetic polymorphisms in the human CYP2C9 gene. *Eur J Clin Invest* 2003;33:23–30.

79. Refsgaard HHF, Jensen BF, Christensen IT, Hagen N, Brockhoff PB. *In silico* prediction of cytochrome P450 inihibitors. *Drug Dev Res* 2006;67:417–429.

80. Singh SB, Shen LQ, Walker MJ, Sheridan RP. A model for predicting likely sites of CYP3A4-mediated metabolism on drug-like molecules. *J Med Chem* 2003;46: 1330–1336.

81. Crivori P, Pogessi I. Computational approaches for predicting CYP-related metabolism properties in the screening of new drugs. *Eur J Med Chem* 2006;41:795–808.

82. Chovan LE, Black-Schaefer C, Dandliker PJ, Lau YY. Automatic mass spectrometry method development for drug discovery: application in metabolic stability assays. *Rapid Commun Mass Spectrom* 2004;18:3105–3112.

83. Jefferis R. Glycosylation of recombinant antibody therapeutics. *Biotechnol Prog* 2005;21:11–16.

84. Bhogal N, Combes R. TGN1412: time to change the paradigm for the testing of new pharmaceuticals. *Altern Lab Anim* 2006;34(2):225–239.

85. Fort DJ, Paul RR. Enhancing the predictive validity of Frog Embryo Teratogenesis Assay—*Xenopus* (FETAX). *J Appl Toxicol* 2002;22:185–191.

86. International Conference on Harmonisation of technical requirements for registration of pharmaceuticals for human use (25 October 2006). *ICH Considerations: General Principles to Address the Risk of Inadvertent Germline Integration of Gene Therapy Vectors*. Available from http://www.ich.org/LOB/media/MEDIA3363.pdf (accessed 29 January 2007).

87. Alonso MJ. Nanoparticles for overcoming biological barriers. *Biomed Pharmacother* 2004;58:168–172.

88. Sadrieh N. Nanaotechnology: regulatory perspective for drug development http://www.fda.gov/nanotechnology/powerpoint_conversions/regulatory_perspective_files/outline/index.html (accessed 29 January 2007).

89. Oberdörster G, Maynard A, Donaldson K, Carter J, Karn B, Kreyling W, Lai D, Olin S, Monteiro-Riviere N, Warheit S, Yang H. Principles of characterizing the potential health effects from exposure to nanomaterials: elements of a screening strategy. A report from ILSI Research Foundation/Risk Science Instate Nanomaterial Toxicity Screening Working Groups. *Particle Fibre Toxicol* 2005;2:1–36.

90. Labhesetwar V, Song C, Humphrey W, Shebucki R, Levy RJ. Arterial uptake of biodegradable nanoparticles: effects of surface modifications. *J Pharm Sci* 2000;87: 1229–1234.
91. Environmental Toxicity Council (ETC) Group: Size matters! The case for a global moratorium. 2003/04. Occasional paper series vol 7 no 1. Available from http://etcgroup.org/en/materials/publications.html?id=165 (accessed 10 September 2007).
92. Lacerda L, Bianco A, Prato M, Kostarelos K. Carbon nanotubes as nanomedicines: from toxicology to pharmacology. *Adv Drug Dev Rev* 2006;58:1460–1470.
93. Nel A, Xia T, Maedler L, Li N. Toxic potential of materials at the nanolevel. *Science* 2006;311:622–627.
94. Lawrence CL, Pollard CE, Hammond TG, Valentin JP. Nonclinical proarrhythmia models: predicting torsades des pointes. *J Pharmacol Toxicol Methods* 2005;52:46–59.
95. De Pont FI, Poluzzi E, Montanaro N. Organising evidence on QT prolongation and occurrence of torsades des pointes with non-antiarrhythmic drugs: a call for consensus. *Eur J Clin Pharmacol* 2001;57:185–209.
96. Recanatini M, Poluzzi E, Masetti M, Cavalli A, De Ponti F. QT prolongation through hERG $K^+$ channel blockade: current knowledge and strategies for the early prediction during drug development. *Med Res Rev* 2005;25:133–166.
97. Lavon N, Bevenisty N. Differentiation and genetic manipulation of human embryonic stem cells and the analysis of the cardiovascular system. *Trends Cardiovasc Med* 2003;13:47–52.
98. He JQ, Ma Y, Lee Y, Thompson JA, Kamp T. Human embryonic stem cells develop into multiple types of cardiac myocytes: action potential characterisation. *Circ Res* 2003;93:1–9.
99. Makino S, Fukuda K, Miyoshi S, Konishi F, Kodama H, Pan J, Sano M, Takahashi T, Hori S, Abe H, Hata J, Umezawa A, Ogawa S. Cardiomyocytes can be generated from marrow stromal cells in vitro. *J Bioeng Biomembr* 1999;28:379–385.
100. Draft ICCVAM test method recommendations: *In Vitro Pyrogenicity Test Methods* (December 2006). Available from http://iccvam.niehs.nih.gov/methods/pyrodocs/supp/PWGrec12016.pdf (accessed 29 January 2007).
101. Willingham AT, Orth OP, Batalov S, Eters EC, Wen BG, Aza-Blanc P, Hogenesch JB, Sculz PG. A strategy for probing the function of noncoding RNAs finds a repressor of NFAT. *Science* 2005;309:1570–1573.
102. Yoon BI, Li GX, Kitada K, Kawasaki Y, Igarashi K, Kodama Y, Inoue T, Kobayashi K, Kanno J, Kim DY, Inoue T, Hirabayashi Y. Mechanisms of benzene-induced hematotoxicity and leukemogenicity: cDNA microarray analyses using mouse bone marrow tissue. *Environ Health Perspect* 2003;111:1411–1420.
103. Kramer JA, Curtiss SW, Kolaja KL, Alden CL, Blomme EA, Curtiss WC, Davila JC, Jackson CJ, Bunch RT. Acute molecular markers of rodent hepatic carcinogenesis identified by transcription profiling. *Chem Res Toxciol* 2004;17:463–470.
104. Nilsson P, Paavilainen L, Larsson K, Odling J, Sundberg M, Andersson AC, Kampf C, Persson A, Al-Khalili Szigyarto C, Ottosson J, Bjorling E, Hober S, Wernerus H, Wester K, Ponten F, Uhlen M. Towards a human proteome atlas: high-throughput generation of mono-specific antibodies for tissue profiling. *Proteomics* 2005;5:4327–4337.
105. Nicholson JK, Lindon JC, Holmes E. "Metabonomics": understanding the metabolic responses of living systems to pathophysiological stimuli via multivariate statistical analysis of biological NMR spectroscopic data. *Xenobiotica* 1999;29:1181–1189.
106. Bollard ME, Murray AJ, Clarke K, Nicholson JK, Griffin JL. A study of metabolic compartmentation in the rat heart and cardiac mitochondritiss using high resolution magic angle spinning $^1$H NMR spectroscopy. *FEBS Lett* 2003;553:73–78.

107. Bollard ME, Stanley EG, Lindon JC, Nicholson JK, Holmes E. NMR-based metabonomic approaches for evaluating physiological influences on biofluid composition. *NMR Biomed* 2005;18:143–162.
108. Viravaidya K, Sin A, Shuler ML. Development of a microscale cell culture analog to probe naphthalene toxicity. *Biotechnol Prog* 2004;20:316–323.
109. Ziauddin J, Sabatini DM. Microarrays of cells expressing defined cDNAs. *Nature* 2001;411:107–110.
110. Hung PJ, Lee PJ, Sabouchi P, Lin R, Lee P. Continuous perfusion microfluidic cell culture array for high-throughput cell-based assays. *Biotechnol Bioeng* 2005;89:1–8.
111. Keun HC. Metabonomic modelling of drug toxicity. *J Pharmacol Exp Ther* 2006;109:92–106.
112. Butcher EC, Berg EL, Kunkel EJ. System biology in drug discovery. *Nat Biotechnol* 2004;22:1253–1259.
113. Ekins S, Nikolsky Y, Nikolskaya T. Techniques: application of systems biology to absorption, distribution, metabolism, excretion and toxicity. *Trends Pharm Sci* 2005;26:202–209.
114. Wilson AG, White AC, Mueller RA. Role of predictive metabolism and toxicity modelling in drug discovery—a summary of some recent advancements. *Curr Opin Drug Discov Dev* 2003;9:74–83.
115. Rester U. Dock around the clock—current status of small molecule docking and scoring. *QSAR Comb Sci* 2006;25:605–615.
116. Barratt M, Langowski J. Validation and development of the DEREK skin sensitisation rulebase by anaylsis of the BgVV list of contact allergens. In Balls M, Zeller AM, Halder M, Eds. *Progress in the Reduction Refinement and Replacement of Animal Experimentation*. New York: Elsevier; 2000, pp 493–512.
117. Zinke S, Gerner I, Schlede E. Evaluation of a rule base for identifying contact allergens by using a regulatory database: comparison of data on chemicals notified in the European Union with "structural alerts" used in the DEREK expert system. *Altern Lab Anim* 2002;30:285–298.
118. Dearden JC, Barratt MD, Benigni R, Bristol DW, Combes RD, Cronin MTD, Judson PN, Payne MP, Richard AM, Tichy N, Worth AP, Yourick JJ. The development and validation of expert systems for predicting toxicity. *Altern Lab Anim* 1997;25:223–252.
119. Matthews EJ, Contrera JF. A new highly specific method for predicting the carcinogenic potential of pharmaceuticals in rodents using enhanced MCASE QSAR-ES software. *Regul Toxicol Pharmacol* 1998;28:242.
120. Enslein K, Gombar VK, Blake BW. International Commission for Protection Against Environmental Mutagens and Carcinogens. Use of SAR in computer-assisted prediction of carcinogenicity and mutagenicity of chemicals by the TOPKAT program. *Mutat Res* 1994;305:47.
121. Woo YT, Lai D, Argus A, Arcos JC. Development of structure–activity relationship rules for predicting carcinogenic potential of chemicals. *Toxicol Lett* 1995;79:219.
122. Lewis DFV, Bird MG, Jacobs MN. Human carcinogens: an evaluation study via the COMPACT and HazardExpert procedures. *Hum Exp Toxicol* 2002;21:115.
123. Ji ZL, Wang Y, Yu L, Ha LH, Zheng CJ, Chen YZ. *In silico* search of putative adverse drug reaction related proteins as a potential tool for facilitating drug adverse effect prediction. *Toxicol Lett* 2006;164:104–112.
124. Blaauboer BJ. The integration of data on physico-chemical properties, *in vitro*-derived toxicity data and physiologically based kinetic and dynamics modelling as a tool in hazard and risk assessment: a commentary. *Toxicol Lett* 2003;138:161–171.

125. Deglmann CJ, Ebner T, Ludwig E, Happich S, Schildberg FW, Koebe HG. Protein binding capacity *in vitro* changes metabolism of substrates and influences the predictability of metabolic pathways *in vivo*. *Toxicol In Vitro* 2004;18:835–840.
126. Chang C, Ray A, Swaan P. *In silico* strategies for modelling membrane transporter function. *Drug Discov Today* 2005;10:663–671.
127. Poole TB, Ed. *The UFAW Handbook on the Care and Management of Laboratory Animals*. London: UFAW; 1999.
128. Hirst WD, Abrahamsen B, Blaney FE, Calver AR, Aloj L, Price GW, Medhurst AD. Differences in the central nervous system distribution and pharmacology of the mouse 5-hydroxytryptamine-6 receptor compared with rat and human receptors investigated by radioligand binding, site-directed mutagenesis, and molecular modelling. *Mol Pharmacol* 2003;64 (6):1295–1308.
129. Morton DM. Importance of species selection in drug toxicity testing. *Toxicol Lett* 1998;102–103:545–550.
130. Tang-Wai DF, Kajiji S, diCapua F, de Graff D, Roninson IB, Gros P. Human (MDR1) and mouse (mdr1, mdr3) P-glycoproteins can be distinguished by their respective drug resistance profiles and sensitivity to modulators. *Biochemistry* 1995;34:32–39.
131. Tang H, Mayersohn M. Accuracy of allometrically predicted pharmacokinetic parameters in humans: role of species selection. *Drug Metab Dispos* 2005;33:1288–1295.
132. Festing M, Ed. *The Design of Animal Experiments: Reducing the Use of Animals in Research Through Better Experimental Design*. London: Royal Society of Medicine Press Ltd, 2002.
133. Morris T, Goulet S, Morton D. The International Symposium on Regulatory Testing and Animal Welfare: recommendations on best scientific practices for animal care in regulatory toxicology. *ILAR J* 2002;43(Suppl):S123–S125.
134. Gold LS, Bernstein L, Magaw R, Slone TH. Inter-species extrapolation in carcinogenesis: prediction between rats and mice. *Environ Health Perspect* 1989;81:211–219.
135. Krauthammer M, Kaufmann CA, Gilliam TC, Rzhietsky A. Molecular triangulation: bridging linkage and molecular-network information for identifying candidate genes in Azheimer's disease. *Proc Natl Acad Sci USA* 2004;101:15148–15153.
136. PharmaPendium™. Elsevier MDL. Available from http://www.mdl.com/products/knowledge/pharmapendium/index.jsp (accessed 29 January 2007).
137. Adverse Drug Effects Database™. Wolters-Kluwer Health. Medi-Span Available from http://www.medispan.com/Products/index.aspx?id=18 (accessed 29 January 2007).
138. Sobels FH. Evaluating the mutagenic potential of chemicals: the minimal battery and extrapolation problems. *Arch Toxicol* 1980;46:21–30.
139. Sutter TR. Molecular and cellular approaches to extrapolation for risk assessment. *Environ Health Perspect* 1995;103:386.
140. Van Ravenzwaay B, Leibold E. The significance of *in vitro* rat skin absorption studies to human risk assessment. *Toxicol In Vitro* 2004;18:219–225.
141. Greaves P, Williams A, Eve M. First dose of potential new medicines to humans: how animals help. *Nat Rev Drug Discov* 2004;3:226–236.

# USE OF PROJECT TEAMS IN PRECLINICAL DEVELOPMENT

Dorothy M. K. Keefe, Joanne M. Bowen, and Rachel J. Gibson
*Royal Adelaide Hospital Cancer Centre and University of Adelaide, Adelaide, South Australia*

**Contents**

1 General History of Preclinical Development
  1.1 What Is Preclinical Development?
  1.2 Why Is It Important?
  1.3 Elements Associated with Preclinical Development
2 What Is a Project Team?
  2.1 Differing Models for Project Teams
  2.2 Leadership of a Project Team
3 Benefits of Project Teams in Preclinical Development
  3.1 Individual Responsibilities Within the Project Team
  3.2 General Responsibilities of the Project Team
4 Special Considerations of Academic Project Teams
5 Using the Dark Agouti (DA) Rat Model of Mucositis to Illustrate
  5.1 Choice of Cytotoxic Agent
  5.2 Time Course of Small Intestinal Mucositis in the DA Rat Model
  5.3 Tumor Bearing Versus Tumor Native?
  5.4 Advantages of Our Rat Model of Mucositis
6 Our DA Rat Model of Mucositis and Preclinical Development

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

7 How Do We Generate Questions/Avenues to Investigate for Preclinical Development?
7.1 Questions Are Formulated by What We Need/Want to Know
7.2 Structure of Our Preclinical Team
7.3 Team Meeting
8 How Do We Present the Results and How Do We Determine What They Mean?
9 The Future for Preclinical Project Teams
References

## 1 GENERAL HISTORY OF PRECLINICAL DEVELOPMENT

### 1.1 What Is Preclinical Development?

As the name suggests, preclinical development of a drug is that part of drug development occurring before the drug enters human trials. During this time, safety, activity, and mechanism of action studies can occur. In fact, even once clinical trials have begun, further preclinical studies can be performed to answer developing questions. Any new therapeutic agent that has shown activity in *in vitro* and *in vivo* models still needs to be thoroughly investigated in order for it to move successfully from the laboratory to the clinic. Commonly, these are referred to as translational activities and may involve a number of different studies including scale-up synthesis of the therapeutic agent, development of analytical assays, development and manufacture of dosage formulations, and animal research studies including pharmacology and toxicology. Regulatory authorities around the world have requirements for new drugs before they can begin human studies, and companies need to comply with these. (http://www3.niaid.nih.gov/research/topics/HIV/therapeutics/intro/preclinical_drug_dev). In order for this to happen effectively, there needs to be a project team in place to manage the overall development.

### 1.2 Why Is It Important?

There are many new drugs investigated every year that are found to be active in various ways in either cell culture or animal model systems. However, few of these translate successfully to human use. This may be due to unacceptable toxicity, species variation in effect, cost–benefit ratio in production, inability to scale production sufficiently to produce quantities required in human use, and financial limitations. Having a standard set of procedures that need to be followed makes the process easier, although practical experience contributes a great deal of value.

### 1.3 Elements Associated with Preclinical Development

This includes *in vitro* and *in vivo* studies of efficacy, safety, pharmacokinetics, and pharmacodynamics, in order to produce pharmaceutical grade materials, and to make effective submission for investigational new drug status with the U.S. Food and Drug Administration (FDA) and other regulatory authorities. By the end of

preclinical development there needs to be the ability to produce sufficient quantities of drug for human use, analytical assays to measure the drug and its metabolites, appropriate dosage formulations, and animal toxicology and pharmacology. All of this needs to comply with FDA and other regulatory requirements. The vast majority of drugs that enter preclinical development do not end up as marketable items. However, those that do are often able to radically improve patients' quality of life.

## 2 WHAT IS A PROJECT TEAM?

A project team in this setting is the group of people charged with taking the drug through the preclinical phase, from the time of decision regarding its potential for human use, up to the time of clinical trial. It is made up of all the key people who can organize the studies that need to be conducted and needs to include adequate administrative support, record keeping, time-line planning, and recording of milestones. A project team is vital for the success of any drug/compound undergoing preclinical development. It is able to successfully implement all aspects of drug development projects, including planning, conducting, regulatory aspects, clinical aspects, administration, ethics approval, administrative support, and reporting of results. More generically, however, a project team is a group of people brought together with a common purpose (in this case to prepare a drug for clinical development).

### 2.1 Differing Models for Project Teams

In the pharmaceutical industry, project teams can be aligned in different ways. For example, it may be appropriate to have one team that looks after the entire development of a drug, set up along lines historically used by the company. There may be standard operating procedures (SOPs) for a given company that will be applied to all new teams that are required. This team may then oversee the bringing of a new drug from basic research right through animal trials to human study. Or, there may be SOPs for preclinical study teams that are completely separate from Clinical Teams, with a hand-over of responsibility between teams once human studies are commenced. Or a new team for a new drug may be allowed (or in the case of new companies, may need) to set up a project team from scratch for a new drug. In this case, a clear idea of the aims of the project is vital. Alternatively, development of a drug may come under the responsibility of several different teams, each responsible for a certain part of the overall project; for example, there could be different project teams along department lines, doing all of the pharmacology, molecular pharmacology, bioanalysis, medicinal chemistry, business development, or marketing for all drugs under development by that particular company. There would then be a requirement for higher level liaison between groups regarding each single drug.

### 2.2 Leadership of a Project Team

Each project team is assigned a project leader who has the responsibility to run the project forward—that is, maintain an updated project plan, meet milestones, ensure

personnel resources are sufficient by communicating with departmental heads, and keep management and the company updated.

This is like leadership in any situation: it contributes enormously to the success of a team, and if not present is sorely missed. There are many theories about leadership, but Griffiths [1] lists eleven competencies that are required: establishing focus, influencing others, drive to achieve, focus on customers, building relationships, fostering teamwork, attention to information, interpersonal awareness, improving performance, developing others, and empowering others. And these are all important in any project team, whatever the goal of that team. It should be noted, however, that a project team never works in isolation but needs to be aware (to a certain degree) of its position within its institution, organizational unit, and program, before it reaches the project or individual level. There also needs to be trust between a team leader and the team members for optimum success [1].

There is an increasing literature on building high performance teams and on project teams that is outside the scope of this chapter.

## 3 BENEFITS OF PROJECT TEAMS IN PRECLINICAL DEVELOPMENT

The major benefit of project teams in preclinical development is the ability to coordinate the activities of all investigators/staff in order to get the drug to clinical trial, or make a decision that it will not work, in the shortest possible time. The team acts as a repository for all information regarding the drug and ensures that all regulatory requirements are met in a timely manner so that nothing is omitted that will delay application for an investigational new drug (IND) approval.

Members of the preclinical project team include the team leader, laboratory head, scientists, students, and administrative support staff. All of these people bring together the skills that are required for the effective management of the team. These skills include budgeting, legal issues, protocol writing, report writing, publication writing, presentation of results, and building of the IND file. Mostly, the project team will be run out of a pharmaceutical company, but many academic institutions will have their own project teams working on specific issues and liaising with industry as appropriate. All members of the project team play a crucial role in any preclinical development.

There are a number of different terminologies that are often used by project teams. We have listed below some of the more common ones.

*Investigational Product.* A pharmaceutical form of an active ingredient or placebo being tested or used as a reference in a clinical trial, including a product with a marketing authorization used or assembled (formulated or packaged) in a way different from the approved form, or when used for an unapproved indication, or when used to gain further information about an approved use.

*Investigator.* Any person responsible for the conduct of the clinical trial at a trial site. If a trial is conducted by a team or individuals at a trial site, the investigator is the responsible leader of the team and may be called the principal investigator.

*Sponsor–Investigatory.* An individual who both initiates and conducts, alone or with others, a clinical trial, and under whose immediate direction the investigational product is administered to, dispersed to, or used by a subject. The term does not include any person other than an individual (e.g., it does not include a corporation or an agency).

*Subinvestigator/Coinvestigator.* Any individual member of the clinical trial team designated and supervised by the investigator at a trial site to perform critical trial-related procedures and/or to make important trial-related decisions (e.g., associates, residents, research fellows).

*Animal Research Ethics Committee.* An independent body constituted of veterinary, scientific, and nonscientific members, whose responsibility is to ensure the protection of the rights, safety, and well-being of animals involved in research by, among other things, reviewing, approving, and providing continuing review of trial protocol and amendments and of the methods and materials to be used in each research project.

### 3.1 Individual Responsibilities Within the Project Team

***Principal Investigator*** The principal investigator has a number of specific roles to play in a project team. First, the he/she is responsible for the signing of appropriate forms including regulatory forms and, if applicable, disclosing of financial interests and arrangements. The principal investigator is also responsible for signing contracts with sponsors and third parties as applicable. Other key responsibilities include supervision of each member of the project team, documentation of the delegation of responsibilities, ensuring the safety and welfare of animals (where an animal trial is being undertaken), hiring and training of members of the project team, meeting sponsor requirements, meeting with sponsors to discuss planned and ongoing studies, and finally meeting with auditors (internal, sponsor, and regulatory authorities) at the conclusion of their audits to review findings.

***Providing Investigator Qualifications and Agreements*** This is one of the key roles that a principal investigator has in a project team. It is essential that he/she continues to keep an up-to-date curriculum vitae, detailing experience including the proper education and training to undertake the research project or clinical trial. Furthermore, the principal investigator is responsible for signing the protocols and external contracts and documenting the financial aspects of the trial.

***Assuring Protocol Compliance*** By having a detailed understanding of the requirements of each research trial and clinical trial, the principal investigator can ensure that all other staff on the project team are aware of and undertaking their roles and responsibilities accurately. The principal investigator can also ensure that staff are not deviating from the protocol without prior agreements from the sponsor and/or the human and/or animal research ethics committees.

***Liaisons with Human and/or Animal Research Ethics Committees*** It is the role of the principal investigator of any project team to prepare animal research ethics committee applications. This involves the initial written application to conduct

animal research, notifying the committee if there are any changes that need to be made to the protocols, providing the committees with results (either positive or negative), and providing the committees with annual reports and end-of-project reports.

***Support Staff*** All support staff of the project teams are required to fulfill those job responsibilities specific to that job title according to regulations and guidelines as well as to the appropriate standard operating procedures (SOPs) and maintain study files and archives.

***Administrative Responsibilities*** The administrative responsibilities of the project team lie with the principal investigator. Briefly, he/she is responsible for the hiring and training of other members of the project team, management of the business aspects of the studies (including the development and negotiation of study budgets and contracts), and overseeing the management of documents (including the filing and archiving).

### 3.2 General Responsibilities of the Project Team

***Principal Investigator, Coinvestigator, Data Manager, Study Pharmacist, and Support Staff*** All members of the project team are required to conduct all studies according to applicable regulations and guidelines and SOPs of the site, and according to the policies and procedures of the institution as appropriate. They are required to ensure that the principal investigator is informed in a timely manner of all study-related activities through appropriate mechanisms, to ensure the safety and welfare of animals by thoroughly understanding ongoing study protocols and to be knowledgeable about investigational products. For studies that are conducted under U.S. IND, all investigators and research personnel must comply with regulations governing disclosure of personal, professional, or financial interest in a research study that may impact upon its conduct, evaluation, or outcome. Finally, all project team members are required to comply with study protocol and procedures. (See Table 1.)

## 4 SPECIAL CONSIDERATIONS OF ACADEMIC PROJECT TEAMS

The project team required to complete a preclinical study varies considerably depending on the research avenue and outcomes sought. However, a reasonable model is illustrated in Fig. 1. The principal investigator instigates the preclinical project. He/she will then liaise with the senior research staff to adequately design a research plan to address the gaps in knowledge and questions to be answered. This is often quite a time-consuming exercise, as ideas are bounced back and forth and an appropriate course of action is decided upon. Will the study involve humans or animals? Will the study be observational or require intervention? Will the study be using novel treatments? These are just a few of the questions that need to be addressed before the study commences. The project team has a crucial role to play in this initial development of the project. More importantly, the project team will also decide whether other expertise will be required for the study. For example, will

**TABLE 1 Major Components of Project Team Activity**

| Administration | Project Management | Investigator-Sponsored Trials (If Applicable) |
|---|---|---|
| Contract negotiations, sign contracts | Ethics committee submissions and communications | Quality assurance/quality control |
| Fiscal management | Regulatory files creation and maintenance | Data handling |
| Facilities management/availability | Data management | Record keeping |
| Strategic planning | Adverse event reports | Financing |
| Database development | Protocol implementation and management | Notification to regulatory authorities |
| Performance tracking | Organizational tools | Information on investigational product(s) |
| Quality control and assurance (including SOPs) | Office staff training | Supply and handling of investigational product(s) |
| Insurance and indemnity | Storing study documents | Safety information |
| | | Monitoring |
| | | Audit and noncompliance |
| | | Premature termination or suspension of a trial |
| | | Study reports |

pharmacists, nurses, technical officers, and students be required? In many cases of preclinical development, biotechnology and pharmaceutical industries are involved, not only in aiding in funding but also in providing technology and reagents. Ethical relations need to be maintained, to ensure that there is no bias/pressure from industry. This guarantees that there is independence and transparency, which ultimately benefits everyone.

Once the research project has commenced, the tiers of the project team contribute to the daily running of the project under direct supervision from the principal investigator of the project team. Constant communication with the principal investigator is essential. Students can be involved in the studies and often play a critical role in any project team.

## 5 USING THE DARK AGOUTI (DA) RAT MODEL OF MUCOSITIS TO ILLUSTRATE

In academic preclinical research, the drivers may be slightly different from those in industry, although the structure and activity of a project team still follows the same ideals. Using our laboratory as an example, we have developed an animal model of chemotherapy- and radiotherapy-induced gastrointestinal mucositis. This is not a model used for developing a single drug for registration, rather a model we have developed to study the pathobiology and treatment of mucositis. However, we do

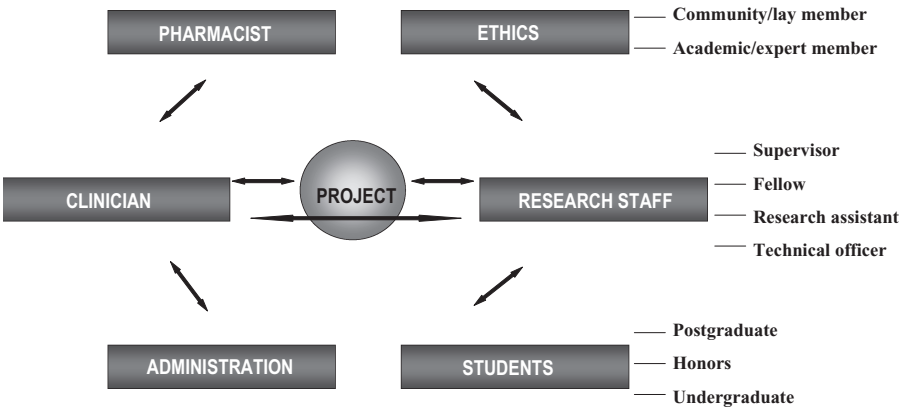


**FIGURE 1** A diagrammatic representation of a project team in the research setting. The project team required to complete a preclinical study varies depending on research avenue and outcomes sought, however, a reasonable model includes those shown. The clinician is often the first to recognize a need for research into a particular area and instigates the preclinical project. He/She then liases with research staff to adequately design a research plan to address the gaps in knowledge and questions to be answered. The combination of clinician and senior research staff allows for multiple principal investigators to be named on any grant application. The research plan must be submitted to an ethics committee for scrutiny. Each committee is comprised of both lay and expert community members to provide a broad view of the project, allowing comments on its significance as well any potential improvements required. In many cases, the biotechnology and pharmaceutical industries are to be used in the project, by providing technology and agents, and the clinician is primarily responsible for opening dialogue with these professions. The administration department within the hospital in the form of record keeping and patient control is involved in supplying background information readily as needed. Once the project is under way, the tiers of laboratory staff contribute to the daily running of the research under supervision from the head of the lab. Constant communication with the clinician as well as data management, tissue collection, and manuscript preparation are all the responsibility of the research staff. Students are also involved through placement in certain parts of the experiment following training by technical staff.

have our own project team to run the studies, and the principles are similar so it will serve as an illustration.

An appropriate animal model was required to investigate further the mechanisms underlying alimentary mucositis. The model had few requirements: it had to have value commercially, it had to be able to be used in true translational research (i.e., research that goes from the laboratory to the clinic and from the clinic to the laboratory), and the overriding goal for the model had to be (and still is today) to improve patient outcomes. It needed to be able to be interrogated in various ways and using various different potential drugs in a standard fashion.

A model was therefore developed using subcutaneously implanted isogeneic rat breast cancer in the female dark Agouti (DA) rat. The mammary adenocarcinoma arose spontaneously in the 1970s and has been propagated ever since by passage through female rats. Female rats are used because the tumor passages more effectively through females than males. The tumor is injected (as a cell suspension) subcutaneously into both flanks and is harvested between 10 and 14 days and processed

for further passage or experimental use. This model had previously been used for studies of malnutrition following chemotherapy [2, 3] and for studies of neuroprotection by glutamate [4], but had not been previously used to assess the alimentary tract following chemotherapy. The model has allowed an attempt at ameliorating intestinal toxicities following chemotherapy [5] and more recently fractionated radiotherapy, while guarding against tumor growth. More recently, the model has been refined to using tumor-naive rats [6]. The advantage of the nontumor model is a time and cost saving, whereas the advantage of the tumor model is the interaction between tumor and host.

### 5.1 Choice of Cytotoxic Agent

Initially, the detailed time course of intestinal damage and repair in the DA rat model needed to be established. In order to commence studies, an appropriate chemotherapeutic agent was required. The cytotoxic agent had to be commonly used and known to cause mucositis in the clinic. Methotrexate (MTX) is a commonly used cytotoxic agent; the mode of action is in the inhibition of dihydrofolate reductase (DHFR) [7–9]. MTX is also known to cause diarrhea and anorexia, accompanied with malabsorption, malnutrition, and dehydration in patients. Furthermore, MTX is known to inhibit epithelial proliferation and enterocyte function as well as increase the risk of gut-associated sepsis due to disruption of the mucosal barrier. This has led to gastrointestinal toxicity being the major dose-limiting factor for the use of MTX [9–11]. Therefore, MTX was deemed an appropriate choice of cytotoxic agent to begin our studies.

### 5.2 Time Course of Small Intestinal Mucositis in the DA Rat Model

Rats were assigned to experimental or control groups; experimental animals received 1.5 mg/kg MTX on days designated 0 and 1, while control animals received saline only [12]. The dose of MTX was chosen from pilot experiments, which indicated that it caused significant, nonlethal small intestinal mucositis [5]. Experimental rats were killed on days 1 (6 h following the second dose of MTX) through 5, with control rats being killed on days 1 and 5 only. This small study detailed the time course of MTX-induced small intestinal mucositis and found that the small intestine is indeed the predominant site of damage. Apoptosis increased 28-fold in the small intestinal crypts 6 h after the second dose and indicated that crypt hypoplasia and villus atrophy occurred between days 2 and 4 following treatment. Following this initial injury, the epithelium enters a rapidly proliferating state, repairing and regenerating the intestine [9, 12–14]. This small animal experiment confirmed that the time course of small intestinal mucositis in the DA rat model [12] was similar to that observed in our initial human studies [5, 15], indicating that the animal model was an appropriate model to continue developing. Furthermore, these results allowed us to continue to understand the mechanisms underlying mucositis.

### 5.3 Tumor Bearing Versus Tumor Naive?

The DA rat model was initially developed using subcutaneously implanted isogeneic rat breast cancers in the female dark Agouti rat. However, it has been well docu-

mented that patients with malignancies will have alterations in their immune function [16], including impaired delayed-type hypersensitivity, decreased lymphocytic function, and decreased lymphocyte proliferation response [16, 17]. In addition to this, tumors are known to be responsible for the secretion of different mediators, which decrease the efficiency of immunological integrals [18]. Until recently, there was no evidence on whether the presence of a tumor burden will affect the gastrointestinal response (i.e., mucositis) to chemotherapy. However, a study conducted by our laboratory found that response to chemotherapy was worse in tumor-bearing rats than tumor-naive rats, indicating that the presence of a tumor load adds to the comorbidity following chemotherapy [6].

### 5.4 Advantages of Our Rat Model of Mucositis

The DA rat model of mucositis has a number of key advantages over other animal models. First, the rats have a rapid growth and development phase. Second, there is homogeneity among all animals, meaning that we can accurately assess what damage is occurring and when. Finally, we are able to examine the entire gastrointestinal tract from mouth to anus and from epithelium to muscle. This has enabled us to develop a three-dimensional model of mucositis.

***How Do We Generate Questions/Avenues to Investigate in the DA Rat Model?*** Many of the questions and avenues that have been investigated in the DA rat model of mucositis have stemmed from questions that need to be answered in the clinic. Our first studies in mucositis were done in humans, but it is very hard to take biopsies in real patients throughout the GIT and throughout the course of cytotoxic treatment.

## 6 OUR DA RAT MODEL OF MUCOSITIS AND PRECLINICAL DEVELOPMENT

***How Do We Use the Model for Preclinical Development?*** The DA rat model has been used to assess the pathobiology of mucositis following various chemotherapeutic agents. Using a standard trial protocol and measuring at different levels of the GIT at different time points, we have built up a clear picture of the damage caused by various agents. We can then add any potential antimucotoxic drug of interest and assess, in the same way, whether it reduces mucosal toxicity. We have previously done this with IL-11, glutamine, Palifermin, Velafermin, and VSL#3.

***How Do We Design an Appropriate Trial?*** Having set up the model to measure damage throughout the GIT at various time points, it is possible to decide which time points and areas of the gut are of particular interest for a given study. In small pilot studies, mortality and diarrhea can be used as endpoints. If the drug of interest shows promise in the pilot setting, then we can examine more specific time points and different areas of the GIT and make measurements including histology and apoptosis in larger, more detailed studies.

***How Do We Determine Variables to Change?*** As with any experiment, the number of variables to be changed at each given time should be as close to one as possible.

Which one to change depends on the question being asked. For example, do we wish to compare three different doses? Or perhaps three different administration schedules? This is done by discussion with the team and the company and depends on pilot results as well as other information from earlier studies.

***What Happens When Things Don't Go According to Plan?*** This is a frequent occurrence in a laboratory. First, it is important not to panic! Every step of the protocol needs to be rechecked (this is one of the reasons that good record keeping is vital). Decisions need to be made with respect to altering the experiment midway, abandoning it, or restarting. No hasty decisions should be made, as they are often wrong.

## 7 HOW DO WE GENERATE QUESTIONS/AVENUES TO INVESTIGATE FOR PRECLINICAL DEVELOPMENT?

### 7.1 Questions Are Formulated by What We Need/Want to Know

There are several ways in which questions can be generated for our model. First, a pharmaceutical company may approach us to ask us to try its agent in our DA rat model. Alternatively, through careful reading of the current literature, or through scientific research discussions, we come across a question that we cannot currently answer and that we feel it would be useful to investigate. Other ways in which we can generate questions are that we learn about a new drug and seek it out for testing in our model or we seek to expand what the model can do. For example, we can try and test new chemotherapeutic agents, or test the development of the radiotherapy model. A research avenue to investigate these questions is then formulated in the rat model, using the principles of changing as few variables as possible, using minimal numbers of animals, and using adequate controls.

Following on from this, an appropriate trial is designed (in consultation with the project team) in the rat model of mucositis to answer our questions. Our questions are answered by a variety of techniques: the rat's response to treatment, histology, histopathology, intestinal morphometry, apoptosis, Western blotting, RT-PCR, and microarrays (with diarrhea and mortality in pilots).

From these experiments, we are then able to generate results that lead to potential mechanisms and invariably more questions. All of these lead to an attempt to understand mechanisms in the rat model and translation into the clinic.

With proper planning for storage of samples, further questions can be answered as and when they become important. The development of new technologies allows for this, as with the development of tissue microarray technology. Properly stored tissues can have RNA extracted for assessment.

### 7.2 Structure of Our Preclinical Team

The Mucositis Research Group head is a clinician scientist, allowing for understanding of both basic and clinical issues to ensure that the research carried out by our group is clinically relevant. There is then a laboratory manager who is responsible for day-to-day running of the laboratory, occupational health and safety, task allocation, record keeping, stock ordering, and routine finances, as well as playing a lead

role in manuscript planning, mentoring of junior staff, and medium-term planning. We have a technical officer who works under supervision of the laboratory manager on the above tasks, and the rest of the staff are postgraduate students whose projects all fit into the overall goal of the group, which is to understand and then minimize gastrointestinal toxicity from anticancer treatments.

### 7.3 Team Meetings

There is an annual direction planning meeting, which looks at overall strategy: funding from partners and grant applications, priorities of research projects, student numbers and progress, manuscript planning, abstract submission plans, and conferences to be attended. A fortnightly team meeting looks at individual progress but in the group setting with reports of activity against targets, checking of laboratory record books, and updates on tasks (writing, abstracts). On alternate fortnights the laboratory head and laboratory manager meet each other person separately to assess progress in a more confidential setting. An open-door policy with regard to communication is maintained, and the overall effect is a happy, hard-working, and high-achieving workforce. The environment is supportive and optimistic, with all members of the team recognized and valued for their contributions and their differences. Turnover of staff is low, and this is a reflection of the atmosphere. In addition, *ad hoc* meetings with visiting experts are conducted as required. Team building exercises, such as the occasional lunch, the annual trip to the Multinational Association of Supportive Care Symposium, and afternoon tea for special celebrations, help to keep the team functioning well.

This is the model that has developed over the years in the Mucositis Research Group, but it is not the only possibility. There are many different organizational models, and different ones may suit different companies. This can be a problem during company mergers, as the "personality" of one group of workers may not fit with that of another! This is due to organizations having their own definite structural, humanistic, political, and cultural models. Of critical importance, however, is periodic review of goals, orientation, and general direction. This can be particularly important in a research laboratory, where the temptation can be to head off in any direction that seems interesting at the time, without stopping to think how that fits in with the overall aims of the research group.

## 8 HOW DO WE PRESENT THE RESULTS AND HOW DO WE DETERMINE WHAT THEY MEAN?

It is a principle of good science that all results should be presented and published, whether positive or negative, and it is our duty to ensure that this happens with our studies. However, it is harder in the commercial world to do this due to problems with commercial confidentiality. The importance of being transparent is essential for good science and this cannot be overstated. However, again it is somewhat easier in academia than in industry.

***Different Scenarios Can Lead to Different Outcomes*** Sometimes it can be the protocol that leads to a drug not working rather than the drug itself: the dose or

schedule can be wrong, or an ineffective batch of drug can be used. These problems need to be checked. A timely reminder is that a single experiment always needs repeating before it can be said to be valid.

***The Importance of Being Transparent*** While improving company profits is obviously one of the major aims of any industry, the overriding issue in drug research and development is to produce drugs that really work, and that will improve quality and quantity of life, with minimal adverse effects. This may sound like unnecessarily stating the obvious, but it is an important component to a project team's work: getting the right answer is the goal, not getting an answer that will drive up profits, especially if it is wrong. Putting an unsafe or ineffective drug onto the market will get found out and will result in much more trouble down the track.

***What Happens If the Drug Doesn't Work?*** We need to check the experiment, look at potential confounders, consider changes to dose, route, or scheduling, and if possible try these. Sometimes a drug can be removed from development because the company determines not to try altering the protocol due to cost blow-outs. It is the responsibility of the project team to try to ensure the protocol chosen for the "do or die" study is the one most likely to work. However, if the drug doesn't work, then it should *not* be marketed.

***Interface with Clinical Development*** During the progression of preclinical research into the clinical setting, a number of questions need to be answered. Can we find the answers by studying humans? If yes, then we begin research in humans. If not, further animal or cell model studies are required to be designed so that this can be achieved. Ultimately, any research in humans will lead to further questions. Then we must again consider if these questions can be answered by looking at humans. If not, then we must go back to preclinical models of research before we can continue (Fig. 2).

## 9 THE FUTURE FOR PRECLINICAL PROJECT TEAMS

***Potential Improvements*** With the ongoing development of teams and companies, SOPs are revised, improved, and developed with the experience of the organization. This should lead to a continued refinement of project team theory as well as practical implementation within a company/academic institution. The increasing crossover between business and research should enable constant improvement, as long as the differences between science and business are respected.

***New Difficulties*** The same crossover can also have negative impacts, with business decisions interfering in the practice of research. Obviously, there has to be a balance, as financial imperatives have to be heeded to a certain degree. The size of a team can become unwieldy and the structure of an organization can impose dysfunction on a team. Poor leadership and management of a team can lead to failure, and it can be hard to distinguish between failure of the drug and failure of the team on some occasions.

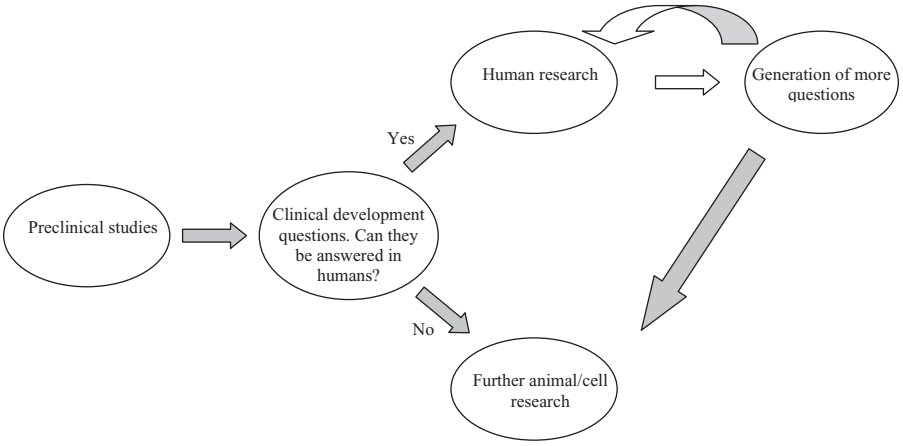


**FIGURE 2** Development from preclinical to clinical investigations. During the progression of preclinical research into the clinical setting a number of questions need to be answered. Can we find the answers by studying humans? If yes, then we begin research in humans. If not, further animal or cell model studies are required to be designed so that this can be achieved. Ultimately, any research in humans will lead to further questions. Then we must again consider if these can be answered by looking at humans. If not, then we must go back to preclinical models of research before we can continue.

***Summary*** A good preclinical project team is compact, goal-oriented, well-resourced both financially and with skilled personnel, and has clear time-lines for meeting its objectives. It undergoes regular assessment against its tasks and is able to adapt to results in a positive, timely manner.

## REFERENCES

1. Griffiths X. *The Art of Leading a Team*. Adelaide, Australia: KRG Consultants; 2002.
2. Rofe AM, Bourgeois CS, Washington JM, Philcox JC, Coyle P. Metabolic consequences of methotrexate therapy in tumour-bearing rats. *Immunol Cell Biol* 1994;72:43–48.
3. Rofe AM, Bais R, Conyers RA. Ketone-body metabolism in tumour-bearing rats. *Biochem J* 1986;233:485–491.
4. Boyle FM, Wheeler HR, Shenfield GM. Glutamate ameliorates experimental vincristine neuropathy. *J Pharmacol Exp Ther* 1996;279:410–415.
5. Keefe DMK. *The Effect of Cytotoxic Chemotherapy on the Mucosa of the Small Intestine*. Adelaide: Department of Medicine, University of Adelaide; 1998.
6. Gibson RJ, Bowen JM, Alvarez E, Finnie JW, Keefe DMK. Establishment of a single dose irinotecan model of gastrointestinal mucositis. *Chemotherapy* 2007;53:360–369.
7. Huennekens FM. The methotrexate story: a paradigm for development of cancer chemotherapeutic agents. *Adv Enzyme Regul* 1994;34:397–419.
8. Tenenbaum L. *Cancer Chemotherapy and Biotherapy*. Philadelphia: WB Saunders; 1994, pp 7–8.
9. Bowen JM, Gibson RJ, Cummins AG, Keefe DMK. Intestinal mucositis: the role of the Bcl-2 family, p53 and caspases in chemotherapy-induced damage. *Support Care Cancer* 2006;17:713–731.

10. Papaconstantinou HT, Xie C, Zhang W, Ansari NH, Hellmich MR, Townsend CM, Ko TC. The role of caspases in methotrexate-induced gastrointestinal toxicity. *Surgery* 2001;859–865.
11. Verburg M, Renes IB, Meijer HP, Taminiau JAJM, Buller HA, Einerhand AWC, Dekker J. Selective sparing of goblet cells and Paneth cells in the intestine of methotrexate-treated rats. *Am J Physiol* 2000;279:G1037–G1047.
12. Gibson RJ, Keefe DMK, Thompson FM, Clarke JM, Goland GJ, Cummins AG. Effect of interleukin-11 on ameliorating intestinal damage after methotrexate treatment of breast cancer in rats. *Dig Dis Sci* 2002;47:2751–2757.
13. Howarth GS, Francis GL, Cool JC, Xu X, Byard RW, Read LC. Milk growth factors enriched from cheese whey ameliorate intestinal damage by methotrexate when administered orally to rats. *J Nutr* 1996;126:2519–2530.
14. Taminiau JAJM, Gall DG, Hamilton JR. Response of the rat small-intestine epithelium to methotrexate. *Gut* 1980;21:486–492.
15. Keefe DM, Brealey J, Goland GJ, Cummins AG. Chemotherapy for cancer causes apoptosis that precedes hypoplasia in crypts of the small intestine in humans. *Gut* 2000;47: 632–637.
16. Fittkau M, Voigt W, Holzhausen HJ, Schmoll HJ. Saccharic acid 1.4-lactone protects against CPT-11-induced mucosa damage in rats. *J Cancer Res Clin Oncol* 2004;130(7):388–394.
17. Evans TRJ. Vaccine therapy for cancer—fact or fiction? *Proc R Coll Physicians Edinburgh* 2001;31:9–16.
18. Broder S, Waldmann TA. The suppressor-cell network in cancer (first of two parts). *N Engl J Med* 1978;299:1281–1284.

# RELATIONSHIP BETWEEN ANIMAL MODELS AND CLINICAL RESEARCH: USING MUCOSITIS AS A PRACTICAL EXAMPLE

Rachel J. Gibson, Joanne M. Bowen, and Dorothy M. K. Keefe
*Royal Adelaide Hospital Cancer Centre, and University of Adelaide, Adelaide, South Australia*

**Contents**

1 General History of the Use of Animal Models
2 Benefits of Animal Models in Mucositis Research
  2.1 Oral Mucositis Animal Models
  2.2 Gastrointestinal Animal Models
3 Difficulties of Animal Models in Mucositis Research
4 The Dark Agouti (DA) Rat Mammary Adenocarcinoma Model of Mucositis
  4.1 Choice of Cytotoxic Agent
  4.2 Time Course of Small Intestinal Mucositis in the DA Rat Model
  4.3 Tumor Bearing Versus Tumor Naive
  4.4 Advantages of Our Rat Model of Mucositis
5 Generation of Research Question/Avenues for Investigation in the DA Rat Model
6 History of Mucositis in Humans
7 Use of Potential Antimucotoxics
  7.1 Interleukin-11
  7.2 Palifermin
  7.3 Velafermin
  7.4 VSL#3
  7.5 Other Potential Antimucotoxics
8 Irinotecan: A Special Case?
9 Radiotherapy: A New Application for the DA Rat Model of Mucositis

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

10 New Advances
10.1 Introduction
10.2 Continuum of Mucosal Injury Research
10.3 Emerging Tools in Cancer Treatment and Diagnosis Improvement Requiring Mucositis Research
11 Future Use of Animal Models in Mucositis Research
12 Conclusion
Acknowledgment
References

## 1 GENERAL HISTORY OF THE USE OF ANIMAL MODELS

Animals have been used in "research" by humans for hundreds of years with the first recorded use of animals for research purposes by Erasistratus in Alexandria in the third century BC [1]. In the second century AD, Galen of Pergamum used animals to further understand anatomy; however, it was not until the beginning of the 19th century that "true" animal research began [2]. With the lack of appropriate anesthesia, some scientists proposed principles for animal research: experiments needed to be necessary, needed to have a clearly defined and attainable objective, should not be repeated unnecessarily, and needed to be conducted with the least possible infliction of pain on the animals [2]. This formed the basis of animal research for the next century [2].

By the late 1800s, basic anesthesia had been developed, leading to major advances in medical research. This necessitated developments in animal research [2]. Following World War II, animal research grew at a rapid rate as research into cancer, cardiovascular diseases, digestive diseases, and ageing began in earnest [2]. The development of appropriate animal models has allowed for major advances to be made for human benefit. We have improved our knowledge and understanding of many diseases including but not exclusive to AIDS, cholera, diabetes, and spinal cord injuries [3, 4]. Furthermore, animal models have allowed us to understand biology, including the physiology of reproductive biology, which has led to the development of the oral contraceptive pill [3, 4].

Today, animal research is tightly regulated, with researchers adhering to Codes of Practice [5]. The purposes of these codes are to ensure the ethical and humane care and use of animals used for scientific purposes [5]. Animal models continue to lead the way with advances in medical research.

## 2 BENEFITS OF ANIMAL MODELS IN MUCOSITIS RESEARCH

Mucositis is a major clinical problem in oncology, caused by the cytotoxic effects of cancer chemotherapy and radiotherapy. It can affect the mucosa of the oral cavity and gastrointestinal tract, causing mouth pain, ulceration, abdominal pain, bloating, vomiting, and diarrhea, depending on the target tissue [6, 7]. For many years, mucositis received little attention. However, the development of animal models has

allowed considerable work to be undertaken on mucositis in the oral cavity [8–15] and in the gastrointestinal tract (GIT) [16–20]. Despite the advances, however, these investigations have separated the GIT into the oral and esophageal mucosa and the remainder of the tract, with no investigations to date having investigated the entire GIT from mouth to anus [7, 21]. There are at least two major reasons for this approach to past studies of mucositis. First, while both the upper and lower GIT are lined with mucosa, the tissue types are different. The upper GIT consists of a renewing stratified squamous mucosa, whereas the lower GIT is primarily columnar epithelium. As a consequence, the kinetics of mucositis in the two areas are vastly different, as are the clinical endpoints. Second, the ease with which the oral mucosa can be inspected and studied in *in vivo* models has been an asset compared to the remainder of the tract [7, 21]. However, it is now recognized that the entire GIT has the same embryological route of development with the differences seen being due to cellular differentiation at various sites in order to conduct specialized functions [7, 21]. With the development of appropriate animal models, we can investigate this hypothesis.

### 2.1 Oral Mucositis Animal Models

While there are a number of different animal models in oral mucositis research, two in particular have been studied extensively. Wolfgang Dorr and colleagues [8–12] have developed a radiation model in mice, which involves irradiating the tongue and snout. This model has also been able to extensively investigate the effects of combined chemotherapy and radiotherapy [22]. In addition, it has enabled detailed studies to be conducted on the effects of single dose and fractionated radiotherapy on the head. The second model, which has advanced the understanding of the mechanisms of oral mucositis, is the hamster model, which has been used extensively by Stephen Sonis and colleagues [15]. Briefly, this model of mucositis uses male golden Syrian hamsters, as unlike other rodents, they have a buccal cheek pouch that is susceptible to chemotherapy. Mucositis is able to be induced by the administration of 5-fluorouracil (5-FU) at 60 mg/kg on three days (days 0, 5, and 10). The buccal pouch mucosa is superficially irritated (mechanically scratched) on days 1, 2, and 3, resulting in mucositis in most of the animals [15].

### 2.2 Gastrointestinal Animal Models

Few animal models exist for investigating the remainder of the GIT. Howarth and colleagues [23] have developed an animal model for utilizing the sucrose breath test and also for investigating potential antimucotoxics [24, 25]. However, very few other models exist that are able to investigate the underlying mechanisms of gastrointestinal mucositis. The dark Agouti (DA) rat model of mucositis (see Section 4) fills this need.

The development of appropriate animal models allows us to ask specific questions about any region of the GIT, develop an appropriate line of questioning, and then investigate it. We can also measure our response to these questions by changing only one variable at a time. Other advantages of animal models include the ability to work with a homogeneous population, where all of the animals are inbred. This reduces the variability between each animal and ensures that we are getting an

accurate reflection of what is occurring. Standard questions are able to be investigated in each experiment, which generate standard answers. All of these combine to reduce the time of drug development, meaning that drugs move from the research laboratory to the clinic in a more timely fashion.

## 3 DIFFICULTIES OF ANIMAL MODELS IN MUCOSITIS RESEARCH

While animal models undoubtedly have benefits, they also have difficulties and limitations. One of the first animal models for investigating mucositis was established by Sonis and colleagues [15]. While this model has enabled hypotheses to be developed for mucositis mechanisms, there is the confounding issue of wound healing. Hamsters have cheek pouches and mucositis is able to be induced by either chemotherapy (5-FU) [15, 26, 27] or radiotherapy [28]. However, following administration of the chemotherapy, the cheek pouch needs to be "mechanically" scratched or irritated in order to induce ulcerated lesions. In humans, however, the oral mucosa does not need to be superficially irritated in order to induce mucositis, and so this model cannot be compared with the clinical setting. Additionally, superficial irritation may also result in wound healing mechanisms being initiated. This highlights the fact that, despite similarities, animal models are never identical to humans, and there will always be issues with translation from animal to human research. This does not, however, devalue animal research; it just adds an appropriate note of caution. With the hamster cheek-pouch model, some agents that have appeared promising in the animals have not translated to success in humans with mucositis. One factor here may be the component of wound healing. With the need for mechanical scratching of the mucosa to allow chemotherapy to cause an ulcer, the insult to the mucosa is composed of both the scratch and the chemotherapy, and this may alter both the development of the ulcer and the recovery. In humans, no scratch is necessary. If a given agent is acting to heal the component induced by the scratch, but has no effect on the development of a chemotherapy-induced ulcer, then it would not be expected to work for chemotherapy-induced mucositis in humans. However, dose and scheduling issues are also important and cannot be overlooked. The doses used in rats do not automatically translate to humans: there may be species differences in susceptibility to different agents, and the traditional mg/kg dosing of rodents is not often used in humans, where we tend to use (for reasons that are not always logical) body surface area dosing.

An added difficulty with animal models has been introduced with the development of monoclonal antibodies for treatment of human disease. Fully humanized monoclonal antibodies may not be active in animal models and toxicities may not develop until translation occurs to the human situation.

Difficulties also arise in the DA rat model of mucositis. Unlike the hamster, in the rat, we are unable to successfully induce visible oral mucositis due to the highly keratinized nature of the epithelium (D. Wilson and D. Keefe, personal communication). This makes it difficult to successfully investigate oral mucositis. Furthermore, higher doses of chemotherapy are required to induce mucosal injury in animal models, due to the resilience of the rat GIT. Another difference is the presence of squamous epithelium in the rat stomach, which can lead to reduction in oral intake when keratinocyte growth factor, a stimulator of epithelial growth, is used. We also

know that rats do not vomit, and since some vomiting is a manifestation of mucosal injury, this is a disadvantage.

The route of administration of chemotherapy has important implications for drug metabolism. In the DA rat model of mucositis, intravenous administration of chemotherapeutic drugs is extremely difficult, with administration into the tail vein being made especially difficult due to the skin pigmentation. This means that investigation into mucositis induced by drugs administered via this route is not routinely performed. Although all chemotherapeutic drugs cause damage [29, 30], the mechanisms of how they do this may be different.

Other contributing factors also cause difficulties in animal research. These include stresses in the animals from isolation due to experimental procedures, the need to anesthetize animals on a regular basis, and the effect that this has on mucosal homeostasis, and the efficacy of any investigative drugs on tumor load.

## 4 THE DARK AGOUTI (DA) RAT MAMMARY ADENOCARCINOMA MODEL OF MUCOSITIS

An appropriate animal model was required to investigate further the mechanisms underlying alimentary mucositis. The model had few requirements: it had to be capable of modeling the changes that occur in the human GIT following insult with chemotherapy; it had to have commercial value; it had to be able to be used in true translational research (research that goes from the laboratory to the clinic and from the clinic to the laboratory); and the overriding goal for the model had to be (and still is today) to improve patient outcomes. It had additional advantages of being able to be used either with or without tumor and to allow study of any organ of choice. It has subsequently also been used for a radiotherapy-induced mucositis study, and in the future combination studies could be performed.

A model was therefore developed using subcutaneously implanted isogeneic rat breast cancer in the female dark Agouti (DA) rat. The mammary adenocarcinoma arose spontaneously in the 1970s and has been propagated ever since by passage through female rats. Female rats are used because the tumor passages more effectively through females than males. It is worth noting that the tumor does not grow in culture; otherwise it would be preferential to study it using these methods. The tumor is injected (as a cell suspension) subcutaneously into both flanks and is harvested between 10 and 14 days and processed for further passage or experimental use. This model had previously been used for studies of malnutrition following chemotherapy [31, 32] and for studies of neuroprotection by glutamate [33], but had not been used to assess the alimentary tract following chemotherapy. The model has allowed attempts at ameliorating intestinal toxicities following chemotherapy [34] and more recently fractionated radiotherapy [35], while guarding paradoxically against tumor growth. More recently, the model has been refined to using tumor-naive rats [36].

### 4.1 Choice of Cytotoxic Agent

Initially, the detailed time course of intestinal damage and repair in the DA rat model needed to be established. In order to commence studies, an appropriate

chemotherapeutic agent was required. The cytotoxic agent had to be commonly used and known to cause mucositis in the clinic. Methotrexate (MTX) is a commonly used cytotoxic agent; the mode of action is in the inhibition of dihydrofolate reductase (DHFR) [6, 37, 38]. MTX is also known to cause diarrhea and anorexia, accompanied with malabsorption, malnutrition, and dehydration in patients. Furthermore, MTX is known to inhibit epithelial proliferation and enterocyte function as well as increase the risk of gut-associated sepsis due to disruption of the mucosal barrier. This has led to gastrointestinal toxicity being the major dose-limiting factor for the use of MTX [6, 39, 40]. Therefore, MTX was deemed an appropriate choice of cytotoxic for our initial studies.

### 4.2 Time Course of Small Intestinal Mucositis in the DA Rat Model

Rats were assigned to experimental or control groups; experimental animals received 1.5 mg/kg MTX on days designated 0 and 1, while control animals received saline only [20]. The dose of MTX was chosen from pilot experiments, which indicated that it caused significant, nonlethal small intestinal mucositis [34]. Experimental rats were killed on days 1 (6 h following the second dose of MTX) through 5, with control rats being killed on days 1 and 5 only. This small study detailed the time course of MTX-induced small intestinal mucositis and found that the small intestine is indeed the predominant site of damage. Apoptosis increased 28-fold in the small intestinal crypts 6 h after the second dose and indicated that crypt hypoplasia and villus atrophy occurred between days 2 and 4 following treatment. Following this initial injury, the epithelium enters a rapidly proliferating state, repairing and regenerating the intestine [6, 20, 25, 41]. This small animal experiment confirmed that the time course of small intestinal mucositis in the DA rat model [20] was similar to that observed in our initial human studies [34, 42], indicating that the animal model was an appropriate model to continue developing. Furthermore, these results allowed us to continue to understand the mechanisms underlying mucositis.

### 4.3 Tumor Bearing Versus Tumor Naive

The DA rat model was initially developed using subcutaneously implanted isogeneic rat breast cancers in the female dark agouti (DA) rat. However, it has been well documented that patients with malignancies will have alterations in their immune function [43], including impaired delayed-type hypersensitivity, decreased lymphocytic function, and decreased lymphocyte proliferation response [43, 44]. In addition to this, tumors are known to be responsible for the secretion of different mediators, which decrease the efficiency of immunological integrals [45]. Until recently, there was no evidence regarding whether the presence of a tumor burden will affect the gastrointestinal response (i.e., mucositis) to chemotherapy. A study conducted by our laboratory found that the response to chemotherapy was worse in tumor-bearing rats than tumor-naive rats, indicating that the presence of a tumor load adds to the comorbidity following chemotherapy [36].

### 4.4 Advantages of Our Rat Model of Mucositis

The DA rat model of mucositis has a number of key advantages over other animal models. First, the rats have a rapid growth and development phase. Second, there is

homogeneity between all animals, allowing accurate assessments of what damage is occurring and when it is occurring, in relation to chemotherapy administration and healing times. Finally, we are able to examine the entire GIT from mouth to anus and from epithelium to muscle. This has enabled us to develop a three-dimensional model of mucositis. The relative simplicity of the DA rat model of mucositis, combined with the homogeneity of the tumor, ease of interrogation of the model, investigation of other organs in the body, the reproducibility of the model, and its ability to be translated to other laboratories makes it a highly effective animal model. Furthermore, there is the ability to develop simple study protocols that allow investigation of mucosal damage as well as looking at different interventions of drugs at varying time points throughout the study, including prechemotherapy, postchemotherapy, or a combination of both. This similarity of protocols strengthens the model. A wide variety of tests are also able to be performed, from relatively simple mortality and diarrhea assessments through to the more complex gene expression studies.

## 5 GENERATION OF RESEARCH QUESTIONS/AVENUES FOR INVESTIGATION IN THE DA RAT MODEL

Many of the questions and avenues that have been investigated in the DA rat model of mucositis have stemmed from questions that need to be answered in the clinic; but with mucositis there are a vast number of questions that have never been asked, leading to an almost infinite list of potential future studies. The use of the model for mucositis occurred as a direct result of preliminary human studies into chemotherapy-induced injury to the small intestine in humans. Although the original model had been used for the study of immunology [31, 32] and to study neurotoxicity of chemotherapy, this was not carried out in tumor-bearing animals due to time constraints with respect to tumor growth and lead time to neurotoxicity [33].

## 6 HISTORY OF MUCOSITIS IN HUMANS

Mucositis is a major oncological problem caused by the cytotoxic effects of cancer chemotherapy and radiotherapy. By definition, mucositis suggests that inflammation is present; however, recent studies have shown that this is not necessarily the case [34, 42]. Rather, the name simply refers to the damage that occurs to the mucous membranes of the body [34, 42]. Mucositis affects the entire alimentary tract (AT) from mouth to anus and causes pain and ulceration in the mouth and small and large intestines. In addition, it causes abdominal bloating, vomiting, and diarrhea [6, 7, 46, 47]. The potentially severe nature of mucositis can have some further devastating effects, including a reduction or cessation of treatment (which may decrease the chance for remission) and increased stays in hospitals, leading to increased costs of treatment and use of opioids for pain management [46, 48]. Mucositis occurs in approximately 40% of patients after standard doses of chemotherapy and in 100% of patients undergoing high dose chemotherapy and stem cell or bone marrow transplantation [6, 7, 46, 47]. The frequency and severity of mucositis varies depending on the type of cancer (and therefore the treatment

regimen) and on the patient's age, with the very young and the very elderly being the most affected [49].

One of the major problems with mucositis is that the mechanisms behind the development of mucositis are not fully understood, making it hard to target treatment. Recent work has shown that the entire GIT has the same embryological route of development, with the differences seen being due to cellular differentiation at various sites in order to conduct specialized functions [7]. Based on this theory, it is likely that mucositis is the same throughout the GIT, with the local differences in manifestations being due to localized specialized differentiation necessary for specific function [7]. It may be that these specialized differences in function explain why different regions of the tract, such as the small intestine, are more susceptible to "early" mucositis than other regions, such as the oral mucosa. To date, there are no investigations that have examined the entire GIT response to chemotherapy; this is most likely due to the difficulties in appropriate models.

Mucositis is a field that has rapidly evolved over the last decade. We have seen advances leading from basic clinical science questions able to be investigated in humans, to conducting detailed microarray experiments in animal models. The first step in the mucositis matrix started with relatively simple investigations of the severity and time course of changes in intestinal permeability following high dose chemotherapy and autologous blood stem cell transplantation [50]. This study involved 35 patients and found that maximum sugar permeability occurred in the small intestine 14 days after the start of chemotherapy, returning to normal 35 days after chemotherapy. Furthermore, this abnormality was found to correlate with the time frame that the patients suffered from other gastrointestinal symptoms, including anorexia and nausea [50]. This first insight into damage after chemotherapy led to further questions being asked about exactly how chemotherapy damages the small intestine.

Our second study into understanding mucositis assessed the frequency, duration, and severity of both oral and gastrointestinal symptoms after chemotherapy and also involved a clinical study. Sixty patients in total were assessed, including those with a newly diagnosed malignancy and those undergoing high dose chemotherapy and autologous peripheral blood stem cell transplantation. Patients were assessed nutritionally, completed a symptom questionnaire, performed an intestinal sugar permeability test, and underwent blood testing for serum endotoxin. A small number of patients also underwent a series of breath tests before and after chemotherapy [34]. Results from this study indicated that patients experienced gastrointestinal symptoms 3–10 days after chemotherapy, and these returned to normal by day 28. In contrast to this, patients did not experience oral symptoms until day 7 after chemotherapy, remaining until day 14 before returning to normal by day 28 [34].

These results were the first to show that chemotherapy does alter the small intestine. A third patient-based study was conducted to assess small intestinal mucosal histology. Twenty-three patients were recruited for the study and underwent an endoscopy with duodenal biopsy before and at varying time points after chemotherapy. Biopsies were assessed using a variety of techniques, including the apoptosis specific assay (TUNEL), enterocyte height, and electron microscopy [42]. We found from this study that there was a sevenfold increase in apoptosis in intestinal crypts 1 day after chemotherapy (Fig. 1), followed by a reduction in intestinal morphometry 3 days after chemotherapy. Values had returned to normal by 16 days

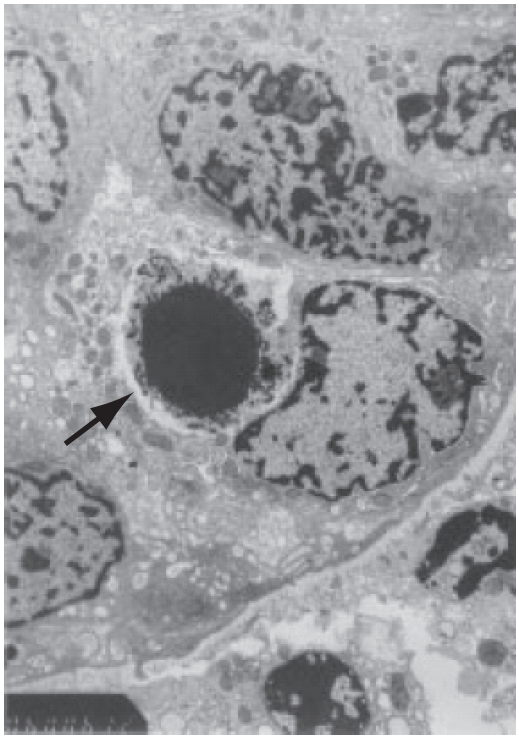

**FIGURE 1** Electron micrograph of an apoptotic cell in the epithelial layer of the small intestinal crypt at 1 day after chemotherapy. The arrow indicates an apoptotic cell with a pyknotic nucleus (original magnification 12,000×). (Used with permission from Ref. 42.)

after chemotherapy [42]. All of these findings from the three clinical studies indicated that cancer chemotherapy does indeed damage the small intestine. Each study answered a question, but also led to more questions that needed answering. However, in order to fully understand the mechanisms of how this damage occurs, an appropriate animal model needed to be developed. While it was possible to perform upper GI endoscopy and biopsy in a small number of patients, additional more detailed investigations would have been extremely difficult and time-consuming for patients, who are heterogeneous in so many ways.

## 7 USE OF POTENTIAL ANTIMUCOTOXICS

### 7.1 Interleukin-11

One of the first potential antimucotoxics to be investigated in great detail was interleukin-11 (IL-11). IL-11 was first cloned as a stromal-cell derived multifunctional cytokine [51], and over the last decade has been the subject of much research, including early clinical trials for patients with varying malignancies [52–54]. IL-11 has also been the focus of much research into the protection and recovery of stem cells found within the hematopoietic system [55–57]. IL-11 stimulates proliferation

of hematopoietic stem cells [55] as well as stimulating hematopoietic recovery following high dose chemotherapy treatment [56]. Additionally, IL-11 has been shown to have nonhematological effects and is produced in varying locations throughout the body, including the alveolar cells of the lung, chondrocytes, synoviocytes, and thyroid carcinoma cells. Thus, IL-11 may have a role in the stem cells found in the small intestinal crypts.

IL-11 and its receptor IL-11Rα are expressed within the gastrointestinal epithelium [57–59]. Animal studies have highlighted the pleiotrophic nature of IL-11 in numerous small intestinal pathologies, including those induced by cytotoxic treatment. Keith and colleagues [60] used two different models of gastrointestinal pathology and concluded that there was a strong beneficial effect of IL-11, which often led to the reduction or resolution of the pathology, although they made no attempt to elucidate the mechanisms behind this. Further studies conducted by Du and colleagues [57, 61, 62] examined the functional significance of IL-11 in mice treated with combined chemotherapy/radiotherapy. The results from these studies clearly demonstrated that mice receiving IL-11 had an increased survival rate compared with their control counterparts, as well as an elevated mitotic index within the crypt epithelial cells, indicating a more rapid recovery time [57, 61, 62]. In another study, IL-11 increased the survival of mice treated with combined chemotherapy/radiotherapy, aided the recovery of small intestinal villi, and increased cell proliferation in crypts [63]. This was a comprehensive study and analyzed different administration schedules of IL-11 including pre, pre and during, and postcytotoxic exposure. However, this study did not evaluate escalating doses of IL-11 and was unable to ascertain whether increasing the dose resulted in corresponding increases in protection [63]. Furthermore, none of the above studies used tumor-bearing animals and were therefore unable to evaluate the effect of IL-11 on the tumor, nor the confounding effect of IL-11 and chemotherapy/radiotherapy on the tumor.

In the DA rat model of mucositis, the time course of mucositis and the effect of IL-11 on preventing intestinal damage were investigated. The effect of IL-11 on untreated animals was firstly ascertained prior to investigating the effect of IL-11 on improving mucositis following treatment with the cytotoxic MTX [20]. Biological efficacy was confirmed by an increased peripheral platelet count. Small intestinal mucositis was assessed using apoptosis and intestinal morphometry and it was found that IL-11 did not reduce apoptosis following chemotherapy, but instead increased it [20]. Surprisingly, this increased level of apoptosis did not lead to increased intestinal damage, and observed intestinal morphometry levels were less severe. These findings suggest that small intestinal apoptosis itself does not inevitably lead to mucositis [20]. IL-11 has now been approved by the Food and Drug Administration for treatment of adults with solid tumors and lymphomas with severe thrombocytopenia induced by chemotherapy [64].

### 7.2 Palifermin

Palifermin (keratinocyte growth factor, KGF or Kepivance®) is a member of the heparin-binding growth factor family [65, 66] and is different from other family members in that it has a specific trophic action for epithelial cells [65, 66]. Palifermin mediates its function in a paracrine–epitheliomesechymal manner by binding to its

receptor located in the epithelial region of the GIT and other tissues after synthesis by mesenchymal cells located in adjacent epithelium [67–70].

Over recent years it has been the subject of many animal and clinical studies, with particular emphasis on its interactions within the GIT. Farrell and colleagues [71] have demonstrated that palifermin is able to protect against intestinal mucositis in some animal models by improving weight loss and crypt survival. Palifermin has also been shown to be effective in protecting the oral mucosa of mice following radiation therapy [10, 11, 72–74]. These studies have clearly demonstrated that palifermin is an effective antimucotoxic.

Further evidence for the antimucotoxic effect displayed by palifermin came from a recent study in our laboratory. Using the DA rat model of mucositis, we aimed to determine whether palifermin was effective in ameliorating intestinal mucositis following chemotherapy. Our results demonstrated that palifermin pretreatment significantly reduces diarrhea and mortality after chemotherapy. This indicates that palifermin is an effective antimucotoxic for gastrointestinal mucositis [75]. These findings have led to palifermin being the first approved antimucotoxic drug used clinically in Australia, Europe, and North America from December 2004 onwards, for patients receiving stem cells transplantation, hyperfractionated radiation, and high dosage chemotherapy. However, palifermin has yet to be approved for standard cycles of chemotherapy.

### 7.3 Velafermin

Velafermin (FGF-20) is also a member of the fibroblast growth factor family [76–78], and members of this family have been implicated in the normal growth and development of the GIT [76–78]. Receptors for FGFs have been located in the epithelial region of the GIT [79, 80]. Velafermin has previously been shown to have activity in models of ulcerative colitis [77] and oral mucosal damage following cytotoxic insult [81]. Recent results from a Phase I clinical trial investigating velafermin have indicated that it is well tolerated following intravenous administration [82] and these results have led to Phase II clinical trials, which are now under way.

Further evidence for the role of velafermin has come in recent months. Using the DA rat model of mucositis, multiple doses and schedules of administration of velafermin to determine its efficacy as an antimucotoxic in gastrointestinal mucositis were assessed. While some of the investigated doses of velafermin were found to be less effective in reducing the severity of diarrhea and mortality, others appeared to increase the diarrhea and mortality. However, evidence of importance of both the dose scheduling and duration of the dosing had previously been reported by Alvarez and colleagues [81]. In their study, they reported that administration of velafermin on varying days and for varying durations led to different efficacies of the antimucotoxic, and they hypothesized that this was due to the multifactorial elements in mucositis progression. However, in our study, we identified that 16 mg/kg velafermin was a potentially effective dose. This dose of velafermin delayed the onset and reduced the duration of severe diarrhea and improved survival rates. These encouraging results indicate that velafermin has the potential to be an effective antimucotoxic, and further animal studies are now warranted in order to understand the mechanisms of action.

### 7.4 VSL#3

The probiotic compound VSL#3 (VSL Pharmaceuticals, Italy) is a new high potency preparation of highly concentrated freeze-dried living bacteria. Each sachet contains *Streptococcus thermophilus* and several species of *Lactobacillus* and *Bifidobacteria*. The compound has been examined previously in models of inflammatory bowel disease (IBD) and has level 1 evidence of effectiveness in ameliorating pouchitis and Crohn disease [83]. Being a "probiotic" means that VSL#3 is capable of exerting good effects on the host organism by improving the balance of intestinal flora and by ameliorating the growth of possible pathogenic microbes [84]. The mechanism of action appears to be through protective, trophic, and anti-inflammatory effects on bowel mucosa [85]. As such, it seems sensible that VSL#3 would also be effective in ameliorating mucositis, specifically chemotherapy-induced diarrhea (CID), which has a number of overlapping pathologies with inflammatory bowel disease (IBD), including upregulation of inflammatory mediators [86–88]. Recently, VSL#3 has been investigated within our laboratory using the DA rat mucositis model in a study comparing schedules for effectiveness of diarrhea prevention following chemotherapy. Promising results have come from the initial series of experiments indicating its potential antimucotoxic ability.

### 7.5 Other Potential Antimucotoxics

There are numerous other antimucotoxics that have been tested in animal models. These include transforming growth factor-β [89], insulin-like growth factor-1 [24, 90], whey growth factor [25], glutamine [34, 91], and epidermal growth factor [92], to name a few. It is important to continue to study these potential antimucotoxics in animal models prior to use in humans.

## 8 IRINOTECAN: A SPECIAL CASE?

Irinotecan hydrochloride (CPT-11) is a relatively new cytotoxic agent used primarily to treat colorectal carcinoma [6, 17, 75]. The mode of action is to inhibit DNA topoisomerase I [93–96]. One of the biggest problems with using this cytotoxic agent is the severe and frequent gastrointestinal toxicities, particularly diarrhea [6, 17, 75, 93–96]. The severe nature of the diarrhea, coupled with the fact that the underlying mechanisms of its development remain unknown, means that the use of irinotecan is limited. This makes an otherwise very effective drug very difficult to use clinically [75]. Although chemotherapy-induced diarrhea (CID) is well recognized [97–99], the underlying mechanisms have received very little attention. Much of the information in the published literature is based on clinical observations with very little basic science. It is therefore important that we gain a better understanding of the mechanisms behind irinotecan-induced diarrhea, as its continued use will be questionable if the dose-limiting and often life-threatening diarrhea associated with administration cannot be prevented [100].

Utilizing a number of different animal models, including the DA rat model of mucositis, the way in which the GIT responds to irinotecan has been somewhat elucidated; however, the exact mechanism of induction remains unclear [6, 17, 75, 93–96].

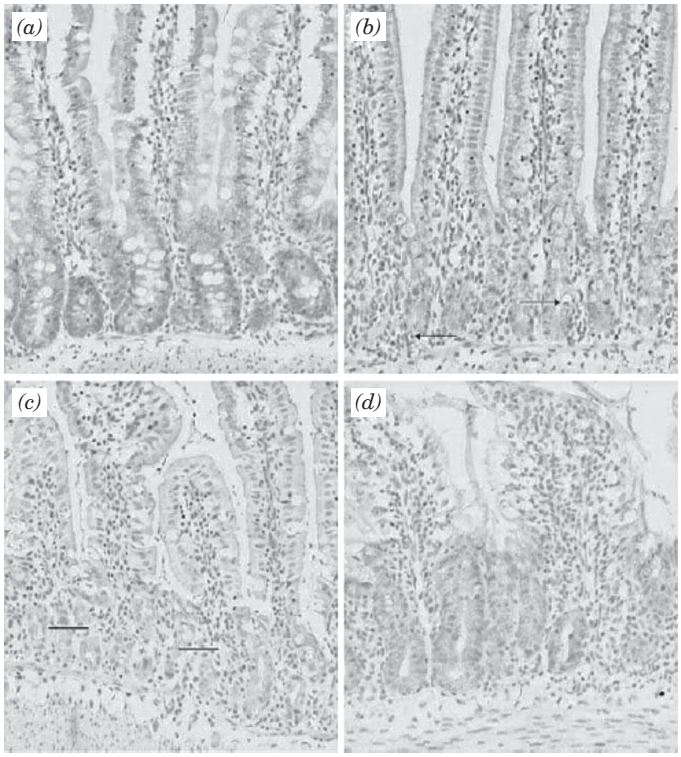


**FIGURE 2** Photomicrographs of rat jejunum stained with hematoxylin and eosin: (A) Untreated control, (B) 6 h, (C) 24 h, and (D) 48 h. Irinotecan treatment caused changes in morphometry, which included apoptosis (arrows) and crypt degeneration (bars) at early time points. Crypt hyperplasia and villus atrophy occurred at 48 h. Original magnfication 100×. (Used with permission from Ref. 114.)

Using the DA rat model of mucositis, our laboratory has, for the first time, been able to establish the changes that occur in the GIT. We first demonstrated a clear time course of gastrointestinal mucositis, caused by irinotecan (Fig. 2). Furthermore, we demonstrated that irinotecan causes severe colonic damage (with apoptosis) (Figs. 3 and 4) and accompanying excessive mucus secretion, suggesting that this increased level of apoptosis, histopathological changes, and goblet cell changes may cause changes in absorption rates leading to diarrhea [17, 47]. Further animal models have investigated the effect of irinotecan on the mouse ileum and cecum. These studies have suggested that the diarrhea induced by irinotecan is the result of the malabsorption of water and electrolytes and mucin hypersecretion [95].

More recently, evidence has arisen that suggests irinotecan-induced diarrhea is a result of interactions between drug metabolism and bacterial microflora [95, 96, 101–103]. One such series of animal experiments has suggested that irinotecan-induced diarrhea is in fact due to changes in bacterial microflora [104, 105]. These studies concluded that irinotecan treatment causes changes in the microflora of the stomach, jejunum, colon, and feces of rats, and that these changes are associated with the development of diarrhea. Furthermore, these changes in microflora may

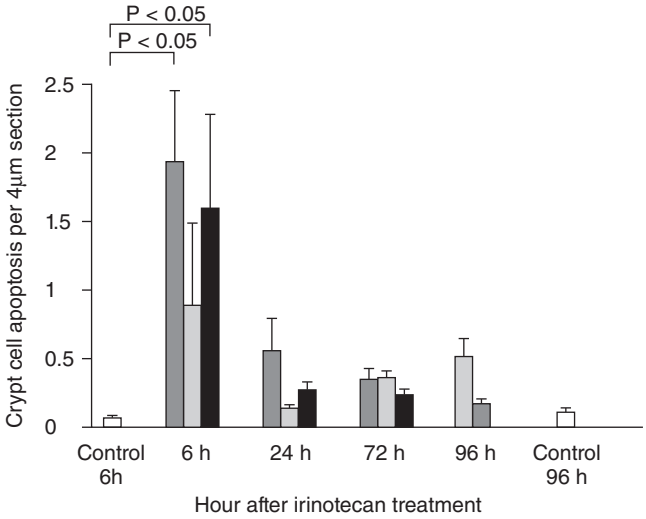


**FIGURE 3** Effect of 100 mg/kg, 150 mg/kg and 200 mg/kg irinotecan given intraperitoneally (IP) over 2 days on jejunal crypt cell apoptosis in rats with breast cancer. Apoptosis was maximal in all doses 6 h following treatment (100 mg/kg and 200 mg/kg $P < 0.05$). White bar, control; diagonal stripes, 100 mg/kg irinotecan; dots, 150 mg/kg irinotecan; black bar 200 mg/kg irinotecan. Results are shown as mean ± SEM. (Used with permission from Ref. 17.)

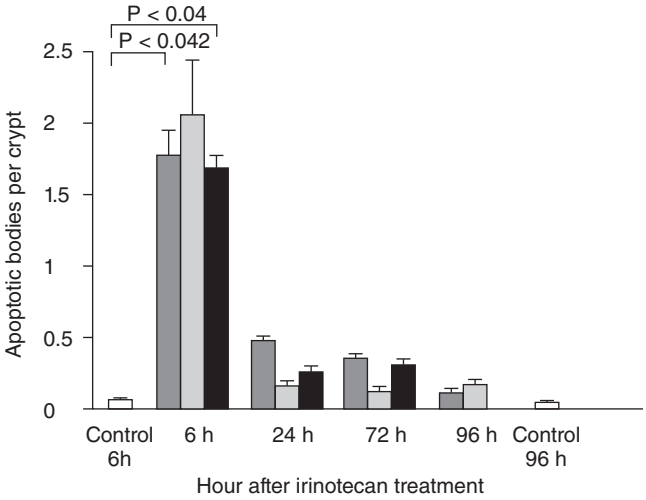


**FIGURE 4** Effect of 100 mg/kg, 150 mg/kg, and 200 mg/kg irinotecan given IP over 2 days on colon crypt cell apoptosis in rats with breast cancer. Apoptosis was maximal in all doses 6 h following treatment (100 mg $P < 0.042$; 200 mg/kg $P < 0.04$). White bar, control; diagonal stripes, 100 mg/kg irinotecan; dots, 150 mg/kg irinotecan; black bar, 200 mg/kg irinotecan. Results are shown as mean ± SEM. (Used with permission from Ref. 17.)

have systemic effects and, in particular, may contribute to the development of chemotherapy-induced mucositis [104, 105]. These results from recent animal studies have been able to be translated back to the clinical setting. We are currently investigating if microflora change in patients following chemotherapy. This highlights the true translational research capacity of an appropriate animal model.

## 9 RADIOTHERAPY: A NEW APPLICATION FOR THE DA RAT MODEL OF MUCOSITIS

Radiation therapy for head and neck cancers results in 30% and 60% of patients developing oral mucositis, while pelvic radiotherapy can lead to acute radiation damage to the anorectal region in up to 75% of patients [106]. These symptoms can be severe enough to interrupt the planned course of treatment in around 10% of patients [107]. Intestinal epithelial injury is one of the major dose-limiting factors in radiation oncology, as acute radiation mucositis severely negates the ability of patients to tolerate higher doses of radiation, which produce greater antitumor effects than lower doses [108]. Despite this, there is currently no appropriate animal model for determining fractionated radiation damage to the GIT, and we have been unable to assess how multiple fractions/doses of radiation therapy cause damage, not only to the epithelium but to all layers of the GIT. It is important to develop such a model as fractionated radiation therapy is a common treatment option in the clinic, leading to severe side effects.

The DA rat model of mucositis has been used extensively in the study of chemotherapy-induced mucositis and results have clearly indicated that mucositis induced by this model is similar to that seen in humans. Currently, there is no animal model for fractionated radiotherapy to study damage to the GIT, with the majority of animal models involving irradiation utilizing a single dose of whole body irradiation [22, 109] or fractionated doses to the head and neck (snout) regions only [72]. The research conducted by Dorr and colleagues have shown that palifermin is effective in reducing the acute oral mucositis induced by radiation therapy [10, 22, 72–74, 109], with the positive effects seeming to be based on changes occurring to the epithelial proliferation and differentiation processes [10]. It is now of vital importance to translate this research from a head and neck (snout) only radiation model to an animal model that is able to study damage to the entire GIT. Many patients receive this modality of treatment and in order to understand how fractionated doses of radiation therapy cause mucositis in the clinic, we need an appropriate animal model. There are differences between chemotherapy, single-dose radiation therapy, and fractionated radiation therapy that relate to the continued, cumulative damage of daily radiation.

## 10 NEW ADVANCES

### 10.1 Introduction

Clinical research has documented the occurrence of mucositis in cancer patients, highlighting its prevalence and importance; however, animal research has primarily

been responsible for elucidating the mechanisms underpinning its development. The overall goal of preclinical or basic mucositis research is to reduce the burden of disease in cancer patients. Effective realization of this goal will be achieved through an enhanced focus on identifying patients who are at a greater risk of toxic effects and those who are less likely to respond to treatment with potentially life-threatening agents [110]. Increased understanding of complex disease pathways and application of a number of new technologies in preclinical models have created a renewed ability to move forward in this area.

## 10.2 Continuum of Mucosal Injury Research

The most significant advance in mucositis research in the past five years has been the development and characterization of the five-phase model of mucositis to explain its unique pathobiology [13, 86, 88]. The model includes the sequential, but not necessarily discrete, five biological stages of damage induced by cancer treatment.

1. *Initiation*. This stage occurs rapidly following administration of cancer treatment and is activated by the simultaneous effects of both DNA and non-DNA damage within epithelial and submucosal cells and generation of reactive oxygen species (ROSs).
2. *Upregulation and Message Generation*. Here DNA strand breaks and ROSs activate multiple transcription factors, including p53 and NFκB, which regulate a great number of downstream genes involved in mucosal toxicity. At the same time, non-DNA damage-associated events occurring include production of ceramide and matrix metalloproteinases, which lead to apoptosis and tissue breakdown.
3. *Signaling and Amplification*. Due to the activation of transcription factors, proinflammatory cytokines accumulate and target the tissues of the submucosa, with this damage also acting as a positive feedback signal to amplify the reaction. Furthermore, certain cytokines are capable of activating the initial transcription factors further and enhancing upregulation of metalloproteinases and ceramide.
4. *Ulceration*. This phase is considered the most significant in mucositis and is characterized by a loss of mucosal integrity with superficial bacterial colonization of the subsequent lesions. The consequences of this include activation of infiltrating mononuclear cells to release additional proinflammatory cytokines and the upregulation of proapoptotic genes, potentiating tissue injury.
5. *Healing*. Generally, mucositis resolves once cancer treatment has ceased and is likely carried out in a similar fashion to other types of mucosal healing, with epithelial cell migration, proliferation, and differentiation of healing tissue.

This model was discovered gradually, with the questions answered from a preclinical animal model of oral mucositis [15, 27, 28, 111, 112], which was further evidenced in other animal models [16, 17, 20, 104, 113–115] and confirmed in human biopsy samples [18, 116, 117]. It changed the historical thinking of mucositis as a solely epithelial-driven injury to one involving the entire mucosa and all cell types

contained within. This new model of mucositis has allowed opportunities to target steps in development of injury. A definitive pathway for the disease is critical for designing experiments for testing interventions in the few mucositis-specific animal models that exist. A particular example of this is anti-inflammatory agents. Since it is now recognized as an important mechanism involved in mucosal injury, currently used agents in the treatment of other inflammatory disorders may prove useful in prevention and treatment of mucositis and should be tested in preclinical models [118].

Toxicities associated with cancer treatment include those that are localized or regional (ulcers, xerostomia, abdominal pain, malabsorption) and those that are more generalized or systemic (fatigue, lack of appetite, nausea) [119]. The recent realization of concurrent tissue-based and systemic toxicities has resulted in the new paradigm of toxicity clustering [120]. New research initiatives are now under way to further investigate the common aspects of pathogenesis in all cancer treatment-related toxicities (D. Keefe and S. Sonis, personal communication), a significant advance in the approach to this oncological problem. Interestingly, the proof of principle testing for this new way of thinking was carried out in cancer patients [120]. Translational research in the laboratory using animal testing is now occurring to examine in greater details some of the initial findings.

## 10.3 Emerging Tools in Cancer Treatment and Diagnosis Improvement Requiring Mucositis Research

***Biological Response Modifiers*** Although biological response modifiers (BRMs) were initially studied in the 1960s for their potential use in cancer therapy, poor trial results caused a loss in momentum in research [121, 122]. Now refreshed with renewed interest, it has become a rapidly advancing area in cancer treatment. BRMs modify the relationship between the tumor and the patient and can be divided into three main categories: (1) agents that restore, augment, or modulate the patients' normal immunological mechanisms; (2) agents that have direct antitumor effects; and (3) agents that have other biologic effects including interference with a tumor cell's ability to metastasize, promotion of cell differentiation, or interference with transformation of cells (Oncolink website). BRM agents currently being evaluated include monoclonal antibodies, interferons, interleukin 2, and tumor necrosis factor. Collectively, this form of treatment has been named immunotherapy due to its immune system regulatory properties. Side effects associated this these treatments include fever, chills, rash, vomiting, hypotension, and allergic reactions and can be severe [123, 124]. As such, patients need to be carefully monitored following administration of drugs. Although initially developed through animal-based research, it is obvious that to improve the therapeutic use of these agents, further preclinical research needs to be conducted to elucidate mechanisms for averting toxicity. The rat mucositis model is an appropriate animal model for this.

A special case is the targeted anticancer therapies, including monoclonal antibodies and small molecules [125]. Both of these are gaining a significant and expanding role in the therapy of cancer, but the use of targeted therapy remains in combination with conventional cancer treatments (chemotherapy and radiotherapy) for optimal response. A recent review of the literature surrounding the targeted anticancer therapies identified mucosal toxicities as a significant component of the treatments

that was not well studied [125]. These toxicities included diarrhea and mouth ulcers [125]. Furthermore, economic considerations of these toxicities need to be addressed. The combination treatment of monoclonal antibodies with other therapeutic agents, including other antibodies, can cost upwards of $10,000–$25,000 (U.S.) a month. This has staggering financial implications for patients and challenges the healthcare system. Considering that not all patients are responsive (most likely due to differences in targeted receptors), molecular testing to assign drugs on an individual patient basis needs to quickly become part of the paradigm of biological therapy [126]. Thus, there is an obvious need to examine the mucosal toxicities as well as other systemic toxicities in order to understand the mechanism(s) of damage as well as to target appropriate prevention or treatments. Animal models may provide the most appropriate way of investigating this.

***Gene Therapy*** The emerging field of gene therapy is quite broad and promises a number of innovative treatments in the cancer treatment setting as well as the prevention of disease. In animal models of lung cancer, survival benefits have been demonstrated using gene therapy to create cancer vaccines, target viruses to cancer cells for lysis, decrease the blood supply to the tumor, and introduce genes into the cancer cells that either cause death or restore a normal phenotype [127, 128]. This is the next generation in cancer therapeutics thanks to the introduction of genetic engineering. However, caution must be used when applying the therapy. Current gene therapy vaccines rely on modulation of the immune system and stimulation of proinflammatory cytokines to destroy tumor cells. This could lead to perturbation of mucositis, where activation of the inflammatory cascade results in mucosal tissue damage. Continued development of gene therapy approaches in animal models will provide the necessary platform to address this and many other concerns relating to its future use in the clinic.

***Pharmacogenetics*** Expanding knowledge of the molecular basis of cancer and its associated diseases, such as mucositis, has shown that differences in gene expression patterns can guide therapy. The field of research concerning genetic variability in response to drugs is termed pharmacogenetics. Technologies involved in this type of research include DNA sequencing, single nucleotide polymorphism (SNP) discovery, and genotyping [129]. Clinical research dominates in this area, with the observation of patient response heterogeneity triggering interest. However, the application of the knowledge garnered from patient results to design preclinical experiments should prove valuable in the future. Realization that certain drug metabolism pathways can be manipulated to modulate cancer therapy response should be investigated in appropriate animal models.

***Pharmacogenomics*** Pharmacogenomics is the study of inherited differences in interindividual drug disposition and effects. Pharmacogenomics is especially important for oncology, as severe systemic toxicity and unpredictable efficacy are hallmarks of cancer therapies [130]. The recent completion of the human genome combined with emerging genomic and proteomic technologies has advanced this field considerably in the past five years. Arguably the most important discovery platform now in use is microarrays. The types of microarray in oncology molecular diagnostics are currently cDNA, oligonucleotide, and tissue based. For effective

explanation of the huge amount of measurements generated by these microarrays, significant computer-based data analysis systems have been designed. The four predominant methods of data analysis have emerged as hierarchic clustering, self-organizing maps, multidimensional scaling, and pathway associations [131]. Collectively, they can be used for tumor classification and subtyping, drug and biomarker discovery, and the powerful tool of transcription profiling of whole genome mRNA expression to establish profiles of gene expression and biological pathway activation [129]. This is possible for multiple species, so application of this technology will allow a great number of questions to be answered, and of course generated, using animal models of disease. Mucositis research using the DA rat model has recently introduced microarray experiments into routine laboratory tests and generated a wealth of new directives for elucidating mucosal toxicity mechanisms and risk factors [6].

An extension of pharmacogenomics is molecular cartography, which is the science of identifying the interaction of genes and gene products that characterize the function and specialization of each individual cell in the context of cell–cell interaction, tissue and organ function, and system's biology [132]. Molecular cartography or "meta-genomics" will be an exciting area for research in the future. Again, animal models will provide the platform needed to investigate the intricate and multifaceted interactions of genes, so as to elucidate pathological processes, including mucositis, and in the development of novel or improved modes of treatment intervention.

***Toxicogenomics*** Of particular importance in mucositis research is the developing field of toxicogenomics. This is the study of gene expression patterns using high throughput microarrays, automated real-time PCR, nuclear magnetic resonance, and proteomic strategies designed to detect up- and downregulation of genes associated with drug toxicity risk [133]. Toxicogenetic markers for adverse side effects identified in animal models will influence selection and optimization of lead compounds for human studies. An example of this is the discovery that inherited forms of the long QT syndrome can be caused by high affinity drug blockade associated with mutations in the HERG potassium channel regulator [134]. The true value of toxicogenomics is currently unrealized; however, with continued advances in data mining and expression profiling technologies, this will occur in the near future.

***Nanomedicine*** The National Cancer Institute has described nanotechnology as a significant research initiative that could transform cancer research and clinical approaches to cancer care (NCI website 2006). There are a growing number of examples of how nanotechnology tools and nanomachines are transforming the way we approach disease diagnosis, treatment, and prevention. Each has developed through animal models and many remain in the preclinical stages [134]. For example, nanopore sequencing is an ultrarapid method of sequencing based on pore nano-engineering and assembly used for the detection of SNPs and for gene diagnosis of pathogens; microneedles are micromachined needles and lancets engineered from single crystal silicon to be used for painless drug infusion, cellular injection, and a number of diagnostic procedures; and drug delivery microchips are microfabricated devices that incorporate micrometer-scale pumps, valves, and flow channels to allow controlled release of single or multiple drugs on command, especially useful in the long-term treatment of conditions requiring pulsatile release of drug following

implantation. Each of these technologies will have particular and individual side effects of treatment.

The issue of toxicity is a particular concern with all new technologies, such as nanomedicine, but often is ignored. Therefore, to ensure efficient application of these technologies, it is essential that fundamental research is carried out to address this issue. In this respect, the use of a preclinical animal model is vital to development of the treatment. In the situation of mucositis, new drug delivery systems have the potential to be truly effective in reducing tumor burden but may induce considerable mucosal toxicity and should be examined fully in the model. In this respect, it should become a priority for the NIH and NCI in their respective nanotechnology funding programs.

## 11 FUTURE USE OF ANIMAL MODELS IN MUCOSITIS RESEARCH

With the constant application of drug screening and agent testing for potential cancer treatments and supportive agents to use in the clinic, the animal model of mucositis will continue to be a highly valuable tool. We are currently in an age where the biotechnology and pharmaceutical industries are progressing rapidly, offering exceptional new drugs into development. However, it remains that proper rigorous preclinical testing in appropriate and truly representative models needs to be carried out with each new antimucotoxic treatment to ensure that innovative treatment approaches are not introduced before the technology or its understanding has matured sufficiently to extract maximum benefit.

Over the next twenty years, due to the ageing population, the global incidence of cancer will greatly increase and with it mucositis. There will also be an increase in consumerism in medicine, with better informed and assertive patients seeking out novel therapies. For these reasons, the continued development of clinically useful therapies for mucositis is essential. However, the combination of complex factors including technological success, society's willingness to pay, and future healthcare delivery systems will undoubtedly influence how preclinical models are designed and implemented [135].

## 12 CONCLUSION

To summarize, this chapter has described the use of animal models for research of the clinical problem of mucositis. The strong relationship in this example, between animal investigation and translation to the clinical setting, has provided a unique opportunity to conduct effective research in what continues to be a significant oncological problem. With continuing dedication to research and drug development in this area, we will undoubtedly see further improvements in treatment.

## ACKNOWLEDGMENTS

Dr. Rachel Gibson is supported by a research fellowship from The Cancer Council South Australia. Dr. Joanne Bowen is supported by a research fellowship from The Royal Adelaide Hospital.

## REFERENCES


1. Straight W. Man's debt to laboratory animals. *Miami Univ Med School Bull* 1962;16:106.
2. Rowan AN. *Of Mice, Models, and Men: A Critical Evaluation of Animal Research*. Albany: State University of New York Press; 1984, p 323.
3. Smith JA, Boyd KM. *Lives in the Balance: The Ethics of Using Animals in Biomedical Research*. New York: Oxford University Press; 1991, p 352.
4. Paton W. *Man and Mouse: Animals in Medical Research*. New York: Oxford University Press; 1993, p 288.
5. National Health and Medical Research Council. *Australian Code of Practice for the Care and Use of Animals for Scientific Purposes*, 7th ed. Australian Government; 2004, pp 1–85.
6. Bowen JM, Gibson RJ, Cummins AG, Keefe DMK. Intestinal mucositis: the role of the Bcl-2 family, p53 and caspases in chemotherapy-induced damage. *Support Care Cancer* 2006;17:713–731.
7. Keefe DM, Gibson RJ, Hauer-Jenson M. Gastrointestinal mucositis. *Semin Oncol Nurs* 2004;20:38–47.
8. Dorr W, Brankovic K, Hartmann B. Repopulation in mouse oral mucosa: changes in the effect of dose fractionation. *Int J Radiat Biol* 2000;76:383–390.
9. Dorr W, Emmendorfer H, Weber-Frisch M. Tissue kinetics in mouse tongue mucosa during daily fractionated radiotherapy. *Cell Proliferation* 1996;29:495–504.
10. Dorr W, Spekl K, Farrell CL. Amelioration of acute oral mucositis by keratinocyte growth factor: fractionated irradiation. *Int J Radiat Oncol Biol Phys* 2002;54:245–251.
11. Dorr W, Spekl K, Farrell CL. The effect of keratinocyte growth factor on healing of manifest radiation ulcers in mouse tongue epithelium. *Cell Proliferation* 2002;35: 86–92.
12. Dorr W, Spekl K, Martin M. Radiation-induced mucositis in mice: strain differences. *Cell Proliferation* 2002;35:60–67.
13. Sonis ST. The pathobiology of mucositis. *Nat Rev Cancer* 2004;4:277–284.
14. Sonis ST, Fey EG. Oral complications of cancer therapy. *Oncology* 2002;16:680–695.
15. Sonis ST, Tracey C, Shklar G, Jenson J, Florine D. An animal model for mucositis induced by cancer chemotherapy. *Oral Surg Oral Med Oral Pathol* 1990;69:437–443.
16. Gibson RJ, Bowen JM, Cummins AG, Keefe DMK. Relationship between chemotherapy and subsequent apoptosis and damage in the gastrointestinal tract of the rat with breast cancer. *Clin Exp Med* 2005;4:188–195.
17. Gibson RJ, Bowen JM, Inglis MR, Cummins AG, Keefe DMK. Irinotecan causes severe small intestinal damage as well as colonic damage in the rat with implanted breast cancer. *J Gastroenterol Hepatol* 2003;18:1095–1100.
18. Gibson RJ, Cummins AG, Bowen JM, Logan RM, Healey T, Keefe DMK. Apoptosis occurs early in the basal layer of the oral mucosa following cancer chemotherapy. *Asia Pacific J Clin Oncol* 2006;2:39–49.
19. Gibson RJ, Keefe DMK, Clarke JM, Regester GO, Thompson FM, Goland GJ, Edwards BE, Cummins AG. The effect of keratinocyte growth factor on tumour growth and small intestinal mucositis after chemotherapy in the rat with breast cancer. *Cancer Chemother Pharmacol* 2002;50:53–58.
20. Gibson RJ, Keefe DMK, Thompson FM, Clarke JM, Goland GJ, Cummins AG. Effect of interleukin-11 on ameliorating intestinal damage after methotrexate treatment of breast cancer in rats. *Digest Dis Sci* 2002;47:2751–2757.

21. Keefe DM. Gastrointestinal mucositis: a new biological model. *Support Care Cancer* 2004;12:6–9.
22. Dorr W, Bassler S, Reichel S, Spekl K. Reduction of radiochemotherapy-induced early oral mucositis by recombinant human keratinocyte growth factor (palifermin): experimental studies in mice. *Int J Radiat Oncol Biol Phys* 2005;62:881–887.
23. Howarth GS, Tooley KL, Davidson GP, Butler RN. A non-invasive method for detection of intestinal mucositis induced by different classes of chemotherapy drugs in the rat. *Cancer Biol Ther* 2006;5:1189–1195.
24. Howarth GS, Cool JC, Bourne AJ, Ballard FJ, Read LC. Insulin-like growth factor-I (IGF-I) stimulates regrowth of the damaged intestine in rats, when administered following, but not concurrent with, methotrexate. *Growth Factors* 1998;15:279–292.
25. Howarth GS, Francis GL, Cool JC, Xu X, Byard RW, Read LC. Milk growth factors enriched from cheese whey ameliorate intestinal damage by methotrexate when administered orally to rats. *J Nutr* 1996;126:2519–2530.
26. Sonis ST, Van Vugt AG, McDonald J, Dotoli E, Schwertschlag U, Szklut P, Keith J. Mitigating effects of interleukin 11 on consecutive courses of 5-fluorouracil-induced ulcerative mucositis in hamsters. *Cytokine* 1997;9:605–612.
27. Sonis ST, Van Vugt AG, Brien JP, Muska AD, Bruskin AM, Rose A, Haley JD. Transforming growth factor-beta 3 mediated modulation of cell cycling and attenuation of 5-fluorouracil induced oral mucositis. *Oral Oncol* 1997;33:47–54.
28. Sonis ST, Peterson RL, Edwards LJ, Lucey CA, Wang L, Mason L, Login G, Ymamkawa M, Moses G, Bouchard P, Hayes LL, Bedrosian C, Dorner AJ. Defining mechanisms of action of interleukin-11 on the progression of radiation-induced oral mucositis in hamsters. *Oral Oncol* 2000;36:373–381.
29. Ijiri K, Potten CS. Response of intestinal cells of differing topographical and hierarchical status to ten cytotoxic drugs and five sources of radiation. *Br J Cancer* 1983;47: 175–185.
30. Ijiri K, Potten CS. Further studies on the response of intestinal crypt cells of different hierarchical status to eighteen different cytotoxic agents. *Br J Cancer* 1987;55:113–123.
31. Rofe AM, Bourgeois CS, Washington JM, Philcox JC, Coyle P. Metabolic consequences of methotrexate therapy in tumour-bearing rats. *Immunol Cell Biol* 1994;72:43–48.
32. Rofe AM, Bais R, Conyers RA. Ketone-body metabolism in tumour-bearing rats. *Biochem J* 1986;233:485–491.
33. Boyle FM, Wheeler HR, Shenfield GM. Glutamate ameliorates experimental vincristine neuropathy. *J Pharmacol Exp Ther* 1996;279:410–415.
34. Keefe DMK. The effect of cytotoxic chemotherapy on the mucosa of the small intestine. MD Thesis, Department of Medicine, University of Adelaide.
35. Yeoh ASJ, Gibson RJ, Bowen JM, Yeoh E, Stringer AM, Giam K, Logan RM, Keefe DMK. Radiation therapy-induced mucositis: relationships between short and long-term fractionated radiation, NF-κB, Cox-1 and Cox-2. *Cancer Treatment Reviews* 32:645–651.
36. Gibson RJ, Bowen JM, Alvarez E, Finnie JW, Keefe DMK. Establishment of a single dose irinotecan model of gastrointestinal mucositis. *Chemotherapy* 2006;53:360–369.
37. Huennekens FM. The methotrexate story: a paradigm for development of cancer chemotherapeutic agents. *Adv Enzyme Regul* 1994;34:397–419.
38. Tenenbaum L. *Cancer Chemotherapy and Biotherapy*. Philadelphia: WB Saunders; 1994, pp 7–8.
39. Papaconstantinou HT, Xie C, Zhang W, Ansari NH, Hellmich MR, Townsend CM, Ko TC. The role of caspases in methotrexate-induced gastrointestinal toxicity. *Surgery* 2001;859–865.

40. Verburg M, Renes IB, Meijer HP, Taminiau JAJM, Buller HA, Einerhand AWC, Dekker J. Selective sparing of goblet cells and Paneth cells in the intestine of methotrexate-treated rats. *Am J Physiol* 2000;279:G1037–G1047.
41. Taminiau JAJM, Gall DG, Hamilton JR. Response of the rat small-intestine epithelium to methotrexate. *Gut* 1980;21:486–492.
42. Keefe DM, Brealey J, Goland GJ, Cummins AG. Chemotherapy for cancer causes apoptosis that precedes hypoplasia in crypts of the small intestine in humans. *Gut* 2000;47:632–637.
43. Fittkau M, Voigt W, Holzhausen HJ, Schmoll HJ. Saccharic acid 1.4-lactone protects against CPT-11-induced mucosa damage in rats. *J Cancer Res Clin Oncol* 2004; 130:388–394.
44. Evans TRJ. Vaccine therapy for cancer—fact or fiction? *Pro R Coll Physicians Edinburgh* 2001;31:9–16.
45. Broder S, Waldmann TA. The suppressor-cell network in cancer (first of two parts). *N Eng J Med* 1978;299:1281–1284.
46. Keefe DMK. Mucositis management in patients with cancer. *Support Cancer Ther* 2006;3:154–157.
47. Gibson RJ, Keefe DMK. Cancer chemotherapy-induced diarrhoea and constipation: mechanisms of damage and prevention strategies. *Support Care Cancer* 2006;14: 890–900.
48. Elting LS, Cooksley C, Chambers M, Cantor SB, Manzullo E, Rubenstein EB. The burdens of cancer therapy, clinical and economic outcomes of chemotherapy-induced mucositis. *Cancer* 2003;98:1531–1539.
49. Pico J-L, Avila-Garavito A, Naccache P. Mucositis: its occurrence, consequences, and treatment in the oncology setting. *Oncologist* 1998;3:446–451.
50. Keefe DM, Cummins AG, Dale BM, Kotasek D, Robb TA, Sage RE. Effect of high-dose chemotherapy on intestinal permeability in humans. *Clin Sci* 1997;92:385–389.
51. Paul SR, Bennet F, Calvetti JA, Kelleher KG, Wood CR, Leary AC, Sibley B, Clark SC, Williams DA, Yang Y-C. Molecular cloning of a cDNA encoding interleukin 11, a stromal cell-derived lymphopoietic and haematopoietic cytokine. *Proc Natl Acad Sci* 1990;87: 7512–7516.
52. Tepler I, Elias L, Smith JW 2nd, Hussein M, Rosen G, Chang AY, Moore JO, Gordon MS, Kuca B, Beach KJ, Loewy JW, Garnick MB, Kaye JA. A randomized placebo-controlled trial of recombinant human interleukin-11 in cancer patients with severe thrombocytopenia due to chemotherapy. *Blood* 1996;87:3607–3614.
53. Gordon MS. Thrombopoietic activity of recombinant human interleukin 11 in cancer patients receiving chemotherapy. *Cancer Chemother Pharmacol* 1996;38:S96–S98.
54. Gordon MS, McCaskill-Stevens WJ, Battiato LA, Loewy J, Loesch D, Breeden E, Hoffman R, Beach KJ, Kuca B, Kaye J, Sledge GW Jr. A phase I trial of recombinant human interleukin-11 (neumega rhIL-11 growth factor) in women with breast cancer receiving chemotherapy. *Blood* 1996;87:3615–3624.
55. Musashi M, Yang Y-C, Paul SR, Clark SC, Sudo T, Ogawa M. Direct and synergistic effects of interleukin-11 on murine hemopoiesis in culture. *Proc Natl Acad Sci* 1991;88:765–769.
56. Leonard JP, Quinto CM, Kozitza MK, Neben TY, Goldman SJ. Recombinant human interleukin-11 stimulates multilineage hematopoietic recovery in mice after a myelosuppressive regimen of sublethal irradiation and carboplatin. *Blood* 1994; 83:1499–1506.

57. Du XX, Doerschuk CM, Orazi A, Williams DA. A bone marrow stromal-derived growth factor, interleukin-11, stimulates recovery of small intestinal mucosal cells after cytoablative therapy. *Blood* 1994;83:33–37.
58. Du XX, Williams DA. Interleukin-11: a multifunctional growth factor derived from the hematopoietic microenvironment. *Blood* 1994;83:2023–2030.
59. Peterson RL, Trepicchio WL, Bozza MM, Wang L, Dorner AJ. G1 growth arrest and reduced proliferation of intestinal epithelial cells induced by rhIL-11 may mediate protection against mucositis. *Blood* 1994;86:311a.
60. Keith JC Jr, Albert A, Sonis ST, Pfeiffer CJ, Schaub RG. IL-11, a pleiotrophic cytokine: exciting new effects of IL-11 on gastrointestinal mucosal biology. *Stem Cells* 1994;12: 79–90.
61. Du XX, Liu Q, Yang Z, Orazi A, Rescorla FJ, Grosfield JL, Williams DA. Protective effects of interleukin-11 in a murine model of ischemic bowel necrosis. *Am J Physiol* 1997;272:G545–G552.
62. Du XX, Neben DT, Goldman S, Williams DA. Effects of recombinant human interleukin-11 on hematopoietic reconstitution in transplant mice: acceleration of recovery of peripheral blood neutrophils and platelets. *Blood* 1993;81:27–34.
63. Potten CS. Interleukin-11 protects the clonogenic stem cells in murine small-intestinal crypts from impairment of their reproductive capacity by radiation. *Int J Cancer* 1995;62:356–361.
64. Cairo MS, Davenport V, Bessmertny O, Goldman SC, Berg SL, Kreissman SG, Laver J, Shen V, Secola R, van de Ven C, Reaman GH. Phase I/II dose escalation study of recombinant human interleukin-11 following ifosfamide, carboplatin and etoposide in children, adolescents and young adults with solid tumours or lymphoma: a clinical, haematological and biological study. *Br J Haematol* 2005;128:49–58.
65. Bansal GS, Cox HC, Marsh S, Gomm JJ, Yiangou C, Luqmani Y, Coombes RC, Johnston CL. Expression of keratinocyte growth factor and its receptor in human breast cancer. *Br J Cancer* 1997;75:1567–1574.
66. Brauchle M, Madlener M, Wagner AD, Angermeyer K, Lauer U, Hofschneider PH, Gregor M, Werner S. Keratinocyte growth factor is highly overexpressed in inflammatory bowel disease. *Am J Pathol* 1996;149:521–528.
67. Chailler P, Basque JR, Corriveau L, Menard D. Functional characterization of the keratinocyte growth factor system in human fetal gastrointestinal tract. *Pediatr Res* 2000;48:504–510.
68. Meropol NJ, Somer RA, Gutheil J, Pelley RJ, Modiano MR, Rowinsky EK, Rothenberg ML, Redding SW, Serdar CM, Yao B, Heard R, Rosen LS. Randomized phase I trial of recombinant keratinocyte growth factor plus chemotherapy: potential role as a mucosal protectant. *J Clin Oncol* 2003;21:1452–1458.
69. Rubin JS, Bottaro DP, Chedid M, Miki K, Cunha GR, Finch PW. Keratinocyte growth factor as a cytokin that mediates mesenchymal–epithelial interaction. *EXS* 1995;74: 191–214.
70. Borges L, Rex KL, Chen JN, Wei P, Kaufman S, Scully S, Pretorius JK, Farrell CL. A protective role for keratinocyte growth factor in a murine model of chemotherapy and radiotherapy-induced mucositis. *Int J Radiat Oncol Biol Phys* 2006;66:254–262.
71. Farrell CL, Bready JV, Rex KL, Chen JN, DiPalma CR, Whitcomb KL, Yin S, Hill DC, Wiemann B, Starnes CO, Havill AM, Lu ZN, Aukerman SL, Pierce GF, Thomason A, Potten CS, Ulich TR, Lacey DL. Keratinocyte growth factor protects mice from chemotherapy and radiation-induced gastrointestinal injury and mortality. *Cancer Res* 1998; 58:933–939.

72. Dorr W, Noack R, Spekl K, Farrell CL. Amelioration of radiation-induced oral mucositis by keratinocyte growth factor (rhKGF): experimental studies. *Int J Radiat Oncol Biol Phys* 2000;46:729.
73. Dorr W, Noack R, Spekl K, Farrell CL. Modification of oral mucositis by keratinocyte growth factor; single radiation exposure. *Int J Radiat Biol* 2001;77:341.
74. Dorr W. Modification of acute radio (chemo) therapy effects in squamous epithelia by keratinocyte growth factor. *Radiother Oncol* 2001;60:S8.
75. Gibson RJ Bowen JM, Keefe DM. Palifermin reduces diarrhoea and increases survival following irinotecan treatment in tumour-bearing DA rats. *Int J Cancer* 2005;116: 464–470.
76. Stern J, Ippoliti C. Management of acute cancer treatment-induced diarrhea. *Semin Oncol Nurs* 2003;19:11–16.
77. Jeffers M, McDonald WF, Chillakuru RA, Yang M, Nakase H, Deegler LL, Sylander ED, Rittman B, Bendele A, Sartor RB, Lichenstein HS. A novel human fibroblast growth factor treats experimental intestinal inflammation. *Gastroenterology* 2002;123:1151–1162.
78. Jeffers M, Shimkets R, Prayaga S, Boldog F, Yang M, Burgess C, Fernandes E, Rittman B, Shimkets J, LaRochelle WJ, Lichenstein HS. Identification of a novel human fibroblast growth factor and characterization of its role in oncogenesis. *Cancer Res* 2001;61: 3131–3138.
79. Hughes SE. Differential expression of the fibroblast growth factor receptor (FGFR) multigene family in normal human adult tissues. *J Histochem Cytochem* 1997;45: 1005–1019.
80. Housley RM, Morris CF, Boyle W, Ring B, Blitz R, Tarpley JE, Aukerman SL, Devine PL, Whitehead RH, Pierce GF. Keratinocyte growth factor induces proliferation of hepatocytes and epithelial cells throughout the rat gastrointestinal tract. *J Clin Invest* 1994;94:1764–1777.
81. Alvarez E, Fey EG, Valax P, Yim Z, Peterson JD, Mesri M, Jeffers M, Dindinger M, Twomlow N, Ghatpande A, LaRochelle WJ, Sonis ST, Lichenstein HS. Preclinical characterization of CG53135 (FGF-20) in radiation and concomitant chemotherapy/radiation-induced oral mucositis. *Clin Cancer Res* 2003;9:3454–3461.
82. Schuster MW, Shore TB, Greenberg J, Jalilizeinali B, Possley S, Halvorsen Y-D, Annino V, Hahne W. Phase I trial of CG53135-05 to prevent mucositis in patients undergoing high-dose chemotherapy (HDCT) and autologous peripheral blood stem cell transplantation (PBSCT). *J Clin Oncol* 2005;23:6656.
83. Reid G, Hammond JA. Probiotics. Some evidence of their effectiveness. *Can Fam Physician* 2005;51:1487–1493.
84. Delia P, Sansotta G, Donato V, Messina G, Frosina P, Pergolizzi S, De Renzis C, Famularo G. Prevention of radiation-induced diarrhea with the use of VSL#3, a new high-potency probiotic preparation. *Am J Gastroenterol* 2002;97:2150–2152.
85. Petrof EO, Kojima K, Ropeleski MJ, Musch MW, Tao Y, De Simone C, Chang EB. Probiotics inhibit nuclear factor-kappaB and induce heat shock proteins in colonic epithelial cells through proteasome inhibition. *Gastroenterology* 2004;127:1474–1487.
86. Sonis ST. A biological approach to mucositis. *J Support Oncol* 2004;2:21–32; discussion 35–36.
87. Sonis ST. The pathobiology of mucositis. *Nat Rev Cancer* 2004;4:277–284.
88. Sonis ST. Pathobiology of mucositis. *Semin Oncol Nurs* 2004;20:11–15.
89. Sonis ST, Lindquist L, Van Vugt A, Stewart AA, Stam K, Qu GY, Iwata KK, Haley JD. Prevention of chemotherapy-induced ulcerative mucositis by transforming growth factor beta 3. *Cancer Res* 1994;54:1135–1138.

90. Howarth GS, Fraser R, Frisby CL, Schirmer MB, Yeoh EK. Effects of insulin-like growth factor-I administration on radiation enteritis in rats. *Scand J Gastroenterol* 1997; 32:1118–1124.
91. Huang E-Y, Wan Leung S, Wang C-J, Chen H-C, Sun L-M, Fang F-M, Yeh S-A, Hsu H-C, Hsuing C-Y. Oral glutamine to alleviate radiation-induced oral mucositis: a pilot randomized trial. *Int J Radiat Oncol Biol Phys* 2000;46:535–539.
92. Fehrmann A, Dorr W. Effect of EGFR-inhibition on the radiation response of oral mucosa: experimental studies in mouse tongue epithelium. *Int J Radiat Biol* 2005; 81:437–443.
93. Araki E, Ishikawa M, Iigo M, Koide T, Itabashi M, Hoshi A. Relationship between development of diarrhea and the concentration of SN-38, and active metabolite of CPT-11, in the intestine and the blood plasma of athymic mice following intraperitoneal administration of CPT-11. *Jpn J Cancer Res* 1993;84:697–702.
94. Cao S, Black JD, Troutt AB, Rustum YM. Interleukin 15 offers selective protection from irinotecan-induced intestinal toxicity in a preclinical animal model. *Cancer Res* 1998;58:3270–3274.
95. Ikuno N, Soda H, Watanabe M, Oka M. Irinotecan (CPT-11) and characteristic changes in the mouse ileum and cecum. *J Natl Cancer Inst* 1995;87:1876–1883.
96. Takasuna K, Hagiwara T, Hirohashi M, Kato M, Nomura M, Nagai E, Yokoi T, Kamataki T. Involvement of B-glucuronidase in intestinal microflora in the intestinal toxicity of the antitumor camptothecin derivative irinotecan hydrochloride (CPT-11) in rats. *Cancer Res* 1996;56:3752–3757.
97. Gwede CK. Overview of radiation- and chemoradiation-induced diarrhea. *Semin Oncol Nurs* 2003;19:6–10.
98. Engelking C, Rutledge DN, Ippoliti C, Neumann J, Hogan CM. Cancer-related diarrhea: a neglected cause of cancer-related symptom distress. *Oncol Nurs Forum* 1998;25: 859–860.
99. Viele CS. Overview of chemotherapy-induced diarrhea. *Semin Oncol Nurs* 2003;19: 2–5.
100. Govindarajan R. Irinotecan/thalidomide In metastatic colorectal cancer. *Oncology (Williston Park)* 2002;16:23–26.
101. Alimonti A, Gelibter A, Pavese I, Satta F, Cognetti F, Ferretti G, Rasio D, Vecchione A, Di Palma M. New approaches to prevent intestinal toxicity of irinotecan-based regimens. *Cancer Treat Rev* 2004;30:555–562.
102. Ma MK, McLeod HL. Lessons learned from the irinotecan metabolic pathway. *Curr Med Chem* 2003;10:41–49.
103. Takasuna K, Hagiwara T, Hirohashi M, Kato M, Nomura M, Nagai E, Yokoi T, Kamataki T. Inhibition of intestinal microflora beta-glucuronidase modifies the distribution of the active metabolite of the antitumor agent, irinotecan hydrochloride (CPT-11) in rats. *Cancer Chemother Pharmacol* 1998;42:280–286.
104. Stringer AM, Gibson RJ, Logan RM, Bowen JM, Yeoh ASJ, Burns J, Finnie JW, Keefe DMK. Chemotherapy-induced diarrhoea is associated with changes in the luminal environment in the DA rat. *Exp Biol Med* 2006;232:96–107.
105. Stringer AM, Gibson RJ, Bowen JM, Logan RM, Yeoh ASJ, Keefe DMK. Chemotherapy-induced mucositis: the role of gastrointestinal microflora and mucins in the luminal environment. *J Support Oncol* 2006;5:259–267.
106. Kushwaha RS, Hayne D, Vaizey CJ, Wrightham E, Payne H, Boulos PB. Physiologic changes of the anorectum after pelvic radiotherapy for the treatment of prostate and bladder cancer. *Dis Colon Rectum* 2003;46:1182–1188.

107. Yeoh EK, Russo A, Botten R, Fraser R, Roos D, Penniment M, Borg M, Sun WM. Acute effects of therapeutic irradiation for prostatic carcinoma on anorectal function. *Gut* 1998;43:123–127.
108. Egan LJ, Eckmann L, Greten FR, Chae S, Li ZW, Myhre GM, Robine S, Karin M, Kagnoff MF. IkappaB-kinasebeta-dependent NF-kappaB activation provides radioprotection to the intestinal epithelium. *Proc Nat Acad Sci USA* 2004;101:2452–2457.
109. Dorr W, Reichel S, Spekl K. Effects of keratinocyte growth factor (palifermin) administration protocols on oral mucositis (mouse) induced by fractionated irradiation. *Radiother Oncol* 2005;75:99–105.
110. Sharma R, Tobin P, Clarke SJ. Management of chemotherapy-induced nausea, vomiting, oral mucositis, and diarrhoea. *Lancet Oncol* 2005;6:93–102.
111. Sonis ST. The biologic role for nuclear factor-kappaB in disease and its potential involvement in mucosal injury associated with anti-neoplastic therapy. *Crit Rev Oral Biol Med* 2002;13:380–389.
112. Sonis ST. Mucositis as a biological process: a new hypothesis for the development of chemotherapy-induced stomatotoxicity. *Oral Oncol* 1998;34:39–43.
113. Gibson RJ, Bowen JM, Keefe DM. Palifermin reduces diarrhea and increases survival following irinotecan treatment in tumor-bearing DA rats. *Int J Cancer* 2005;116: 464–470.
114. Bowen JM, Gibson RJ, Tsykin A, Cummins AG, Keefe DMK. Irinotecan changes gene expression profiles in the small intestine of the rat with breast cancer. *Cancer Chemother Pharmacol* 2006;59:337–348.
115. Logan RM, Gibson RJ, Sonis ST, Keefe DM. Nuclear factor-kappaB (NF-kappaB) and cyclooxygenase-2 (COX-2) expression in the oral mucosa following cancer chemotherapy. *Oral Oncol* 2006;43:395–401.
116. Yeoh AS, Bowen JM, Gibson RJ, Keefe DM. Nuclear factor kappaB (NFkappaB) and cyclooxygenase-2 (Cox-2) expression in the irradiated colorectum is associated with subsequent histopathological changes. *Int J Radiat Oncol Biol Phys* 2005;63: 1295–1303.
117. Lalla RV, Schubert MM, Bensadoun RJ, Keefe D. Anti-inflammatory agents in the management of alimentary mucositis. *Support Care Cancer* 2006;14:558–565.
118. Hickok JT, Morrow GR, Roscoe JA, Mustian K, Okunieff P. Occurrence, severity, and longitudinal course of twelve common symptoms in 1129 consecutive patients during radiotherapy for cancer. *J Pain Symptom Manage* 2005;30:433–442.
119. Sonis S, Haddad R, Posner M, Watkins B, Fey E, Morgan TV, Mookanamparambil L, Ramoni M. Gene expression changes in peripheral blood cells provide insight into the biological mechanisms associated with regimen-related toxicities in patients being treated for head and neck cancers. *Oral Oncol* 2006;43:289–300.
120. Nadler LM, Stashenko P, Hardy R, Kaplan WD, Button LN, Kufe DW, Antman KH, Schlossman SF. Serotherapy of a patient with a monoclonal antibody directed against a human lymphoma-associated antigen. *Cancer Res* 1980;40:3147–3154.
121. Miller RA, Maloney DG, Warnke R, Levy R. Treatment of B-cell lymphoma with monoclonal anti-idiotype antibody. *N Eng J Med* 1982;306:517–522.
122. Bonner JA, Harari PM, Giralt J, Azarnia N, Shin DM, Cohen RB, Jones CU, Sur R, Raben D, Jassem J, Ove R, Kies MS, Baselga J, Youssoufian H, Amellal N, Rowinsky EK, Ang KK. Radiotherapy plus cetuximab for squamous-cell carcinoma of the head and neck. *N Eng J Med* 2006;354:567–578.
123. Kiewe P, Hasmuller S, Kahlert S, Heinrigs M, Rack B, Marme A, Korfel A, Jager M, Lindhofer H, Sommer H, Thiel E, Untch, M. Phase I trial of the trifunctional anti-HER2

× anti-CD3 antibody ertumaxomab in metastatic breast cancer. *Clin Cancer Res* 2006;12:3085–3091.

124. Keefe DMK, Gibson RJ. Mucosal injury from targeted anti-cancer therapy. *Support Care Cancer* 2006;15:483–490.
125. Sharkey RM, Goldenberg DM. Targeted therapy of cancer: new prospects for antibodies and immunoconjugates. *CA Cancer J Clin* 2006;56:226–243.
126. Cross D, Burmester JK. Gene therapy for cancer treatment: past, present and future. *Clin Med Res* 2006;4:218–227.
127. Vattemi E, Claudio PP. Gene therapy for lung cancer: practice and promise. *Ann Ital Chir* 2004;75:279–289.
128. Ross JS, Schenkein DP, Kashala O, Linette GP, Stec J, Symmans WF, Pusztai L, Hortobagyi GN. Pharmacogenomics. *Adv Anat Pathol* 2004;11:211–220.
129. Watters JW, McLeod HL. Cancer pharmacogenomics: current and future applications. *Biochim Biophys Acta* 2003;1603:99–111.
130. Zhang MQ. Large-scale gene expression data analysis: a new challenge to computational biologists. *Genome Res* 1999;9:681–688.
131. Chiappelli F. The molecular immunology of mucositis: implications for evidence-based research in alternative and complementary palliative treatments. *Evidence Based Complementary Alternative Med* 2005;2:489–494.
132. Hamadeh H, Amin P, Paules R, Afshari C. An overview of toxicogenomics. *Curr Issues Mol Biol* 2002;4:45–56.
133. Mitcheson JS, Chen J, Lin M, Culberson C, Sanguinetti MC. A structural basis for drug-induced long QT syndrome. *Proc Nat Acad Sci USA* 2000;97:12329–12333.
134. Moghimi SM, Hunter AC, Murray JC. Nanomedicine: current status and future prospects. *FASEB J* 2005;19:311–330.
135. Shikora K. The impact of future technology on cancer care. *Clin Med* 2002:2:560–568.

# BACTERIAL MUTATION ASSAY

PREMKUMAR KUMPATI
*University of Pittsburgh, Pittsburgh, Pennsylvania*

**Contents**

1 Introduction
2 Historical Perspective
3 Principles and Strategies of the Assay
4 Significance and Limitations of the Assay
5 Assay Procedures and Experimental Design
  5.1 Procedural Outline
  5.2 Bacteria
  5.3 Storage and Reisolation
  5.4 Medium
  5.5 Metabolic Activation and Preparation of S9 Mix
  5.6 Controls
  5.7 Test Material and Solvents
6 Experimental Procedure/Design of Experiments
7 Critical Factors in the Procedure
8 Troubleshooting
  8.1 Background Lawn
  8.2 Spontaneous Reversion
  8.3 Criteria for Checking Strain Characteristics
9 Interpretation of Data and Reporting
10 Discussion
11 Safety Considerations
12 Conclusion
  Acknowledgment
  References

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

## 1 INTRODUCTION

The population at large has been exposed to a variety of existing/new chemicals and drugs that have subsequently been shown to trigger mutations (genetic damage) and cancers, in addition to other genetic diseases. Hence, genetic toxicity testing has been an important aspect and essential component for assessing the mutagenic/carcinogenic potential of such agents and is indispensable for product development and registration. It is no wonder that numerous chapters, reviews, and books have been devoted to the development, standardization, and usage of predictive tests for mutagenicity/carcinogenicity. International guidelines indicate that a standard battery of genotoxicity tests, including the bacterial mutation assay, is normally required for new chemicals to assess their ability to cause genetic damage. During the past thirty years, bacterial systems have been extensively employed and modified for their ability to detect genotoxic potential of chemicals and drugs. The bacterial mutation assay is a reliable and widely used short-term assay for initial screening of chemicals for their potential mutagenic/carcinogenic activity. The assay has been recommended and accepted by various regulatory agencies for registration and acceptance of many chemicals. Although short-term tests are useful, they have their own advantages and limitations. This chapter provides an overall outline on the utilization of bacterial systems for rapidly screening/detecting mutagenic/carcinogenic chemicals, by their ability to induce single gene mutations. In addition, it provides procedures and recommendations for conducting the tests.

## 2 HISTORICAL PERSPECTIVE

Chemical mutagenesis and carcinogenesis have been known for many years and are well documented. Previously, potential carcinogenic/mutagenic chemicals have been identified based on their known association with human cancers and hereditary disorders and also by their ability to produce tumors in laboratory animals. Due to the lengthy, complex, and expensive nature of these studies and limited resources, attention has been focused on the development of "short-term tests" for detecting chemical carcinogens, using biological systems rather than whole mammals. Numerous short-term tests have been developed and are being used; such assays need less time than classical long-term studies.

The bacterial mutation assay is a widely used short-term test based on the principle of detecting chemicals that induce reverse mutations and restore the functional competence of a bacterial strain that is defective in synthesizing a vital amino acid. Chemical mutagens and/or carcinogens act directly on DNA, inducing irreversible genetic damage and cancers. However, the majority of chemicals are not biologically active until they are metabolized or enzymatically converted into reactive molecules, generally in the liver. Since bacteria are unable to metabolize chemicals as in mammals and other vertebrates, early attempts to demonstrate the potential mutagenicity/carcinogenicity of chemicals using bacterial assays failed to show responsiveness to various chemical mutagens.

In the 1950s, chemicals were tested for mutagenicity in *Escherichia coli* using suspension and plate (spot) tests or its variation. Due to the relative insensitivity of *E. coli* mutation in reverting from streptomycin dependence to streptomycin

independence and lack of knowledge of metabolic activation of mutagens, this test was not well recognized. In suspension assay, cells are treated in suspension. In the spot test, a small amount of the test chemical is applied directly to the center of a selective agar medium plate seeded with the test organism. As the chemical diffuses into the agar, a concentration gradient is formed and will give rise to a ring of revertant colonies surrounding the area where the chemical is applied [1].

In 1971, Ames [2] initially used the same spot test method for screening mutagens using a large collection of *Salmonella typhimurium* histidine mutant strains. At the same time, Yanofsky [3] presented a number of *E. coli* tryptophan mutants that could also be used for detection of mutagens. The concept of metabolic activation of chemicals by mammalian liver was introduced by Miller and co-workers [4–6]. This was supported by Heinrich Malling's [7] report on the *in vitro* activation of dimethylnitrosamine to a mutagen for *Salmonella*, by incorporating a mouse liver homogenate with the cells in a suspension assay. This was further supported by a publication by Garner et al. [8] showing that liver microsomes could activate aflatoxin B1 to a product that was toxic to the *Salmonella* tester strains. Following the concepts of liver homogenate, Ames et al. [9] prepared a supernatant of liver homogenate by centrifugation at 9000*g*, called it S9, and incorporated it into a top agar containing a tester bacterial strain and a test chemical, thereby producing the plate test, which is more sensitive and quantitative than the spot test. The first comprehensive validation of bacterial tests for detecting carcinogens was conducted by Ames and colleagues [10], using a combination of bacteria (*Salmonella*) and mammalian microsomal enzymes. A series of studies from Ames's laboratory and other laboratories have demonstrated that *Salmonella* mutagens were rodent carcinogens, and almost all carcinogens were mutagens [11, 12]. Green and Muriel [13] first reported the $trp^+$ (tryptophan) reversion assay using *E. coli* WP2 strain for routine screening of chemicals. It detects $trp^-$ to $trp^+$ reversion at a site blocking a step in the biosynthesis of tryptophan prior to the formation of anthranilic acid. This assay is used by some laboratories in combination with the *Salmonella* assay to screen chemicals for mutagenic activity. Over the years, modifications to the standard plate incorporation assay have been developed by different researchers and additional mutations were engineered into the strains of *Salmonella* and *E. coli* that enhanced the sensitivity of the test and allowed testing of a wider range of chemicals. It is noteworthy that multiple tester strains are necessarily used in the bacterial mutation assay because of the differences in their sensitivity.

## 3 PRINCIPLES AND STRATEGIES OF THE ASSAY

The bacterial mutagenicity assay is a widely used short-term genetic toxicology assay designed to screen a wide range of chemical substances and drugs for their mutagenic or carcinogenic potential [10, 14]. It is also known as the bacterial reverse mutation test. In principle, the test employs several amino-acid-requiring strains of *Salmonella typhimurium/Escherichia coli (E. coli)*, carrying a different mutation in various genes involved in the histidine/tryptophan biosynthetic pathway [13, 15]. Hence, they are dependent on amino acid supplementation in culture medium in order to grow. Chemicals that induce mutations revert those mutations already present in the bacteria (back mutation) and restore the functional capacity of the

bacteria to synthesize a particular amino acid, which is detected by the bacteria's ability to grow and form colonies in the absence of that amino acid. The number of spontaneously induced revertant colonies per plate is relatively constant. However, when a mutagen is added to the plate, the number of revertant colonies per plate is increased, usually in a dose-related manner. Figure 1 shows control and mutagen-treated bacterial plates.

Bacterial test systems fall into three main classes: namely, those that detect backward mutations, those that detect forward mutations, and those that rely on a DNA repair deficiency. However, the most widely exploited method is the induction of backward or reverse mutations in *Salmonella typhimurium* and less frequently in *Escherichia coli*. The objective of this assay is to evaluate the mutagenic potential of test chemicals by studying their effect on one or more histidine (auxotrophic) requiring strains of *Salmonella typhimurium* in the absence and in the presence of a liver metabolizing system. When the cultures are exposed to a mutagen, some of the bacteria undergo genetic changes due to chemical interactions, resulting in reversion of the bacteria to a non-histidine-requiring state. The reverted bacteria will then grow in the absence of exogenous histidine, thus providing an indication of the potential of the chemical to cause mutation.

Bacteria are single-celled organisms. They divide rapidly and are relatively easy to grow in large numbers in a few hours. However, mutations occur naturally at a

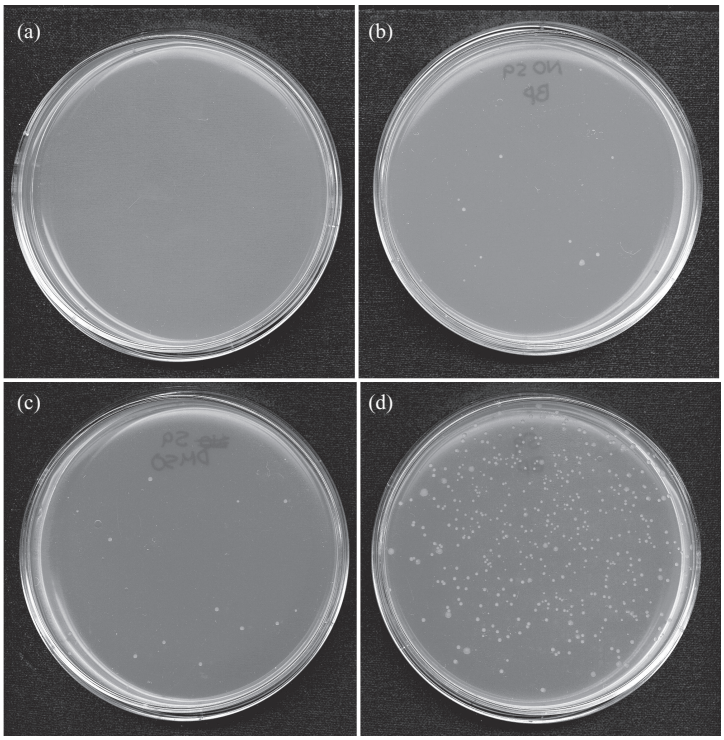


**FIGURE 1** (**a**) Control, (**b**) benzopyrene treated without S9 mix, (**c**) solvent (DMSO) control with S9 mix, and (**d**) benzopyrene treated with S9 mix.

specific locus in less than one in a million bacteria at each cell division. As the biochemistry and genetics of bacteria are well established, it has been possible to develop special strains that are sensitive to a wide range of mutagens. Since different strains are mutated by different classes of compounds, multiple tester strains are necessarily used. Different amino acid requiring bacterial strains could also be used; however, the basis of the test is very similar, except that the bacteria have a requirement for a different amino acid.

To increase the test's sensitivity, bacterial strains have been modified genetically and made more susceptible to mutagens by changing the structure of their cell walls, so that they become more permeable to large fat-soluble molecules. Another genetic manipulation introduces plasmids (small DNA molecules) that carry genes that interfere with DNA repair, making the host bacteria even more susceptible to the mutagenic effects of chemicals. The possibility of a chemical mutagen hitting the mutational site of the tester strain in order to reverse mutate is increasingly achieved by using several strains of the same species of bacterium, each carrying a different preexisting mutation (targets) in the same amino acid gene. For example, there are several strains of *S. typhimurium*, each carrying different mutations in the histidine gene but commonly sharing other traits, like DNA-repair defects and cell-wall defects that improve the strain's sensitivity.

As mentioned previously, unlike mammalian cells, bacteria lack the ability to metabolize chemicals; hence, the test requires use of an exogenous source of metabolic activation [16]. The enzymes are therefore added in the form of a liver extract prepared from laboratory animals, usually rats. The rats are given chemicals ("inducers") that increase the amount of metabolic-activation enzymes in the liver. Subsequently, the liver homogenate is subjected to centrifugation at a speed of 9000*g* so as to collect the supernatant containing the metabolic enzymes and this is called S9 (short for "9000*g* supernatant"). The most widely used inducer is Aroclor 1254, a mixture of polychlorinated biphenyls. Phenobarbital plus 5,6-benzoflavone is also used for induction.

In the bacterial mutation assay, the amino acid deficient bacteria, $his^-$ bacteria, are mixed with S9 and several doses of the test chemical with very limited supply of histidine and are allowed to divide. A chemical that induces mutation either by itself or by its metabolites (by the action of S9) is detected by its ability to transform bacteria into those that can grow into a colony even when the supply of histidine has been used up. These transformed bacteria/revertants restore the ability to synthesize histidine from inorganic nitrogen and can grow without supplementation of histidine. Results are expressed as the number of revertant colonies per plate (Fig. 2).

## 4 SIGNIFICANCE AND LIMITATIONS OF THE ASSAY

The bacterial mutation assay is a rapid, reliable, and economical method for screening many types of substances for their potential genotoxic activity. It is widely used as an early screening tool, due to its procedural simplicity, and has demonstrated a positive correlation between bacterial mutagenicity and carcinogenicity in mammals. Many chemicals that are positive in this test are genotoxic in other tests as well. Bacterial tests have the advantage of testing millions of cells with a relatively short

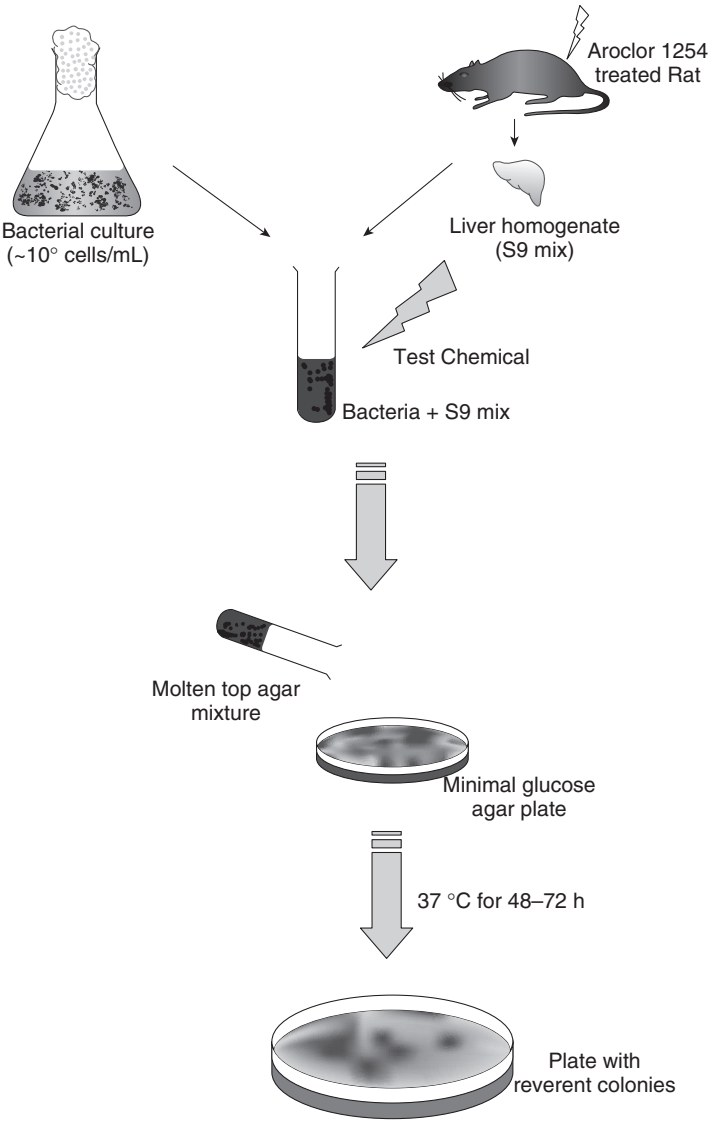


**FIGURE 2** Diagrammatic procedural outline: bacterial mutation assay.

generation time. The rapidly dividing nature of a bacterial system and the rare occurrence of natural mutations at a specific locus makes it suitable for mutagenicity testing. Unlike higher organisms, in which the DNA is organized into complex chromosomal structures, bacteria contain a single circular molecule of DNA that is readily accessible to chemicals that can penetrate the cell wall. The biochemical and genetic nature of the bacteria is well recognized and has led to the possibility of developing special strains of bacteria that are sensitive to a wide range of mutagens. Each strain is specific and provides information on the types of mutations induced by genotoxic agents. Bacterial mutation assays are used in a large number of labo-

ratories throughout the world. A database of the results for a wide variety of chemical structures is available for bacterial reverse mutation tests. The results obtained may be combined and correlated with the results of other tests (e.g., *in vitro* mammalian cell assays and *in vitro* cytogenetic assays) to evaluate the possible mutagenic risk of a test chemical. Thus, bacterial mutation assays give the best overall performance, producing reliable results in a large number of laboratories, and are confirmed as the first choice as an initial screening test. Furthermore, bacterial tests have been validated in far greater detail than any other tests currently used in genetic toxicology.

Although bacterial mutation tests have lots of advantages, they also have some limitations. Bacteria, being prokaryotic cells, differ from mammalian cells in factors such as uptake, metabolism, chromosome structure, and DNA-repair processes. Only gene mutations can be detected using a bacterial system. Bacteria do not or insufficiently possess enzyme systems responsible for metabolic processes. Hence, the test requires the use of an exogenous source of metabolic activation (S9), which does not represent the complicated enzyme network in mammalian tissue. The information obtained through the test, on the mutagenic and carcinogenic potency of a substance in mammals, is not always promising. The test may not provide sufficient information on the genotoxicity potential of compounds that are excessively toxic to bacteria, such as antibiotics, and compounds that are known to interfere with the mammalian cell replication system (e.g., topoisomerase inhibitors, nucleoside analogues, or inhibitors of DNA metabolism). The test also cannot detect chemical carcinogens that act by nongenotoxic (without causing DNA damage) mechanisms (e.g., asbestos, nickel, arsenic, and hormone-like chemicals such as diethylstilbestrol). Despite these limitations, bacterial mutation tests have been found to be extremely valuable as the first in a series of tests for screening chemicals for potential mutagenic and carcinogenic activity.

## 5 ASSAY PROCEDURES AND EXPERIMENTAL DESIGN

### 5.1 Procedural Outline

The bacterial mutation assay is used to assess the chemical's ability to induce point mutations in the bacteria, which are deficient in synthesizing an amino acid (due to the mutation present in the amino acid synthesizing gene), thus reverting the mutation and restoring the capacity to synthesize the amino acid and to grow without amino acid supplementation. In general, bacterial cultures are exposed to the test substance in the presence and absence of an exogenous metabolic activation system and are plated onto minimal agar medium. After 2–3 days of incubation, revertant colonies are counted and compared with that of the negative or solvent control plates. Statistical analysis of the counts are carried out, and the results for mutagenicity are assessed. This is one of several genotoxicity assays required for product safety testing.

Numerous procedures have been described for performing a bacterial mutation assay. The two basic and commonly used methods are the plate incorporation method and the preincubation method. The plate incorporation method is the most commonly used method; however, the preincubation method is recommended for some chemical classes such as azo dyes, aliphatic *N*-nitroso compounds, and

alkaloids [17]. Other procedures include the fluctuation method and the suspension method. In the plate incorporation method, bacterial cultures are exposed to the test chemical in either the presence or absence of a metabolic activation system, combined with molten overlay agar, and plated immediately onto minimal agar medium [18–20]. In the case of the preincubation method, the treatment mixture is preincubated at 37 °C for 20–60 minutes and mixed with the overlay agar before plating onto minimal agar medium [19]. Minimal agar medium contains salts and a carbon source that is sufficient to support growth. The top agar/overlay agar contains a trace amount of amino acid that is sufficient to promote the limited growth of the nonrevertant bacteria. This is important for several reasons, which are discussed in later sections. Establishment of consistent methods for every phase of the experiment is a prerequisite [21]; hence, the parameters and the procedural details that influence the specificity and fidelity of the assay are discussed next.

### 5.2 Bacteria

Healthy growing bacterial cultures with high titer of viable bacteria should be used for the experiments. Viable cell numbers could be determined by a plating experiment. Bacterial cultures should be grown up to the late exponential or early stationary phase of growth (approximately $10^9$ cells/mL). It is noteworthy that cultures in late stationary phase should not be used. For overnight cultures, excessive aeration and shaking exceeding 120 rpm should be avoided [30]. The culture temperature should be maintained at 37 °C. Although all *Salmonella* strains are histidine dependent, various strains of bacteria are available that differ in their sensitivity to various classes of mutagens. Some chemicals are strain specific, whereas others may be detected in several strains. The five recommended combination and commonly used strains of *S. typhimurium* are TA100, TA1535, TA102, TA98, and TA1537 or TA97. These strains, except TA102, have GC base pairs at the primary reversion site and detect a majority of mutagens [17] but not certain oxidizing mutagens, crosslinking agents, and hydrazines. Such substances may be detected by DNA repair-proficient strains of *E. coli* (e.g., *E. coli* WP2 or *E. coli* WP2 (pKM101)). *Escherichia coli* WP2 or *S. typhimurium* TA102 have an AT base pair at the primary reversion site [12]. Thus, on the basis of the chemical class or structure, alternative strain(s) should be considered.

Furthermore, additional mutations/genetic alterations have been introduced to increase the sensitivity of the tester strains to chemical mutagens: for example, *uvrA* gene mutation in *E. coli* and *uvrB-bio* gene mutation in *S. typhimurium*, which eliminate the accurate excision repair mechanism, thereby allowing more DNA lesions to be repaired by the error-prone DNA-repair mechanism. Furthermore, the deletion through the biotin gene makes the bacteria biotin dependent. *rfa* mutations lead to a defective lipopolysaccharide (LPS) layer that coats the bacterial surface, making the bacteria more permeable to bulky chemicals [16]. Introduction of plasmid pKM101 in strains enhances chemical and UV-induced mutagenesis via an increase in the error-prone recombinational DNA-repair pathway [22]. The plasmid confers ampicillin resistance, which is a convenient marker to detect the presence of the plasmid. Introduction of multicopy plasmid pAQ1 with a *hisG428* mutation amplifies the number of target sites.

In order to maintain experimental quality and reliability, bacterial strains should regularly be checked for their characteristic genetic traits, including amino acid requirement, background mutation, induced mutation with reference mutagens, presence of plasmids where appropriate, and presence of cell-wall and DNA-repair mutations. Furthermore, it is also necessary that the strains should yield spontaneous revertant colony counts within the frequency ranges expected from the laboratory's historical control data and preferably within the range reported in the literature. Stock cultures should be stored at a temperature below –70 °C. Overnight cultures for routine assays should be prepared by inoculation from stock cultures—never from a previously used overnight culture. The overnight culture should contain at least $10^9$ viable bacteria/mL and should be freshly prepared for each experiment.

### 5.3 Storage and Reisolation

Frozen stocks of strains are stored at –80 °C. They are prepared from fresh overnight cultures to which 10% (v/v) DMSO or glycerol is added as a cryoprotective agent. Usually fresh oxoid nutrient broth culture is grown to a density of $1–2 \times 10^9$ bacteria/mL. For each 1 mL of culture, 0.09 mL of high purity grade DMSO is added. For preparing multiple frozen tubes, combine the culture and DMSO in a sterile tube, flask, or bottle, mix gently until the DMSO is dissolved, and distribute the culture aseptically into sterile 1.2 mL cryotubes that have been labeled appropriately. All the tubes are placed in a bed of crushed dry ice until the cultures are frozen solid and subsequently transferred to –80 °C freezer/liquid nitrogen.

Tester strains are reisolated from the frozen stock by streaking the bacteria on minimal glucose agar plates enriched with histidine and biotin. Dip a sterile wooden stick into a stock culture (or scrape the bacteria from the surface of the frozen culture) and make a single streak across the surface onto the agar plate. Cross streak the culture using sterile platinum wire. Incubate the plate for 48 h at 37 °C; pick a well isolated colony for overnight growth in oxoid nutrient broth. It should be noted that frequent thawing and refreezing of tester strains may result in reduced viability.

### 5.4 Medium

An appropriate minimal agar medium containing Vogel–Bonner minimal medium E, 0.5–2% glucose [23], and 1.5% agar (w/v) is generally used and is referred to as glucose minimal agar medium. Minimal agar plates contains 25 mL of solidified medium and are stored in sealed plastic bags at 4 °C either in a refrigerator or cold room, to avoid dehydration. Overlay agar or top agar containing 0.6% agar and 0.6% NaCl with a trace amount of amino acid such as histidine or biotin at a concentration of 0.05 mM (for *S. typhimurium*) or tryptophan (for *E. coli*), which is sufficient to promote a few cell divisions, should be used. The *uvrB* deletion extends through the biotin gene—hence the need for biotin in those strains possessing the *uvrB* mutation. These essential components can be added to minimal glucose agar before the test plates are poured or they can be applied to the surface of minimal glucose agar plates and incorporated into the plates using a glass spreader.

### 5.5 Metabolic Activation and Preparation of S9 Mix

Unlike mammalian cells, bacteria lack the ability to metabolize chemicals. To compensate for this, bacteria should be exposed to the test substance in both the presence and absence of an appropriate exogenous metabolic activation system. The most commonly used system is a cofactor-supplemented postmitochondrial fraction (S9) prepared from the livers of rodents (usually rats) treated with enzyme-inducing agents such as Aroclor 1254 (500 mg/kg body weight). Other inducers are phenobarbital and β-naphthoflavone [14, 24]. The animals should be free of disease and infection, kept at a reasonable temperature, and should not be stressed by careless handling. Dosing with inducing agents should be consistent from one batch of animals to the next. Animals should be killed humanely and the livers removed and chilled as soon as possible. A 9000*g* supernatant (S9) is prepared from pooled, homogenized livers of the treated animals. S9 should be stored at or below −70 °C in aliquots in order to avoid contamination. In order to maintain the activity of the S9 fraction, it should always be kept on ice during the experiment or kept frozen when not in use.

There are two widely accepted methods of using S9 mix. (1) S9 mix is mixed with the top agar, bacteria, and test substance, and the whole mixture is immediately poured onto the surface of the bottom agar. (2) In the preincubation method, the test substance, bacteria, and S9 mix are mixed and incubated for 30 min; top agar is then added, and the mixture is poured onto the bottom agar. The postmitochondrial supernatant fraction is usually used at concentrations in the range of 10–30% v/v in the S9 mix. The choice and concentration of a metabolic activation system may depend on the class of chemical being tested. In some cases, it may be appropriate to utilize more than one concentration of postmitochondrial fraction. Nevertheless, for optimal mutagenesis with a particular compound, the amount of S9 per plate is critical. Too much as well as too little S9 can drastically lower the sensitivity. Each new batch of S9 prepared should be routinely checked with known mutagenic compounds such as benzo[*a*]pyrene and aflatoxin B1. The optimum S9 level for a given compound should therefore be checked. The amount of S9 per plate is best expressed as milligram liver protein per plate calculated from the protein concentration of the S9.

Rodent liver homogenate should be supplemented with NADP and cofactor for NADPH-supported oxidation and with FMN for reductive metabolism of chemicals [23]. A reductive metabolic activation system may be more appropriate for azo dyes and diazo compounds. In addition, glucose-6-phosphate (for providing energy) and phosphate buffer (to maintain pH), including magnesium and potassium salts, are added. Liver S9 should be prepared using aseptic techniques so that subsequent filter sterilization could be avoided. Filtration of the S9 or S9 mix may lead to loss of enzyme activity. Each batch of S9, whether produced by the testing laboratory or obtained commercially, should be tested for sterility and discarded if contaminated.

To prepare S9 mix, rats weighing 200 g are exposed to a single intraperitoneal (IP) injection of Aroclor 1254 at a concentration of 500 mg/kg body weight. Aroclor is diluted in corn oil to a concentration of 200 mg/mL and is administered to each rat for 5 days before being sacrificed. Rats are provided with unlimited drinking water and chow until 12 h before killing. After the treatment period, rats are killed

and the livers must be removed aseptically using sterile surgical tools. All the steps of the procedure are carried out at 0–4 °C using cold sterile solutions and glassware. The freshly excised livers are washed several times in fresh chilled 0.15 M KCl solution to ensure sterile preparation and to remove hemoglobin. Washed livers are placed in preweighed beakers containing approximately 3 mL of chilled 0.15 M KCl/g of wet liver and are minced with sterile scissors and homogenized using a homogenizer. The liver homogenate is centrifuged for 10 min at 9000*g* and the supernatant is decanted and assayed. Sterility of the preparation is determined by plating 0.1 mL on minimal agar containing histidine and biotin. Generally, the protein concentration of S9 is ~40 mg/mL. S9 mix consists of 1 mL of S9 fraction, 1 mL NADP (40 mM), 1 mL of glucose-6-phosphate (40 mM), 1 mL of $MgSO_4$ (70 mM), 5 mL of 0.1 M phosphate buffer (pH 7.4), and 1 mL water. This mixture is filtered through a millipore filter (0.45 μm) kept at 0 °C. Next, 2.5 mL of S9 mix is added to 10 mL of diluted bacterial cultures. Different sources of S9 and conditions of incubation can lead to qualitative and quantitative differences. To avoid frequent freezing and thawing cycles and contamination, it is advisable to aliquot into 1–2 mL and store at −80 °C.

### 5.6 Controls

Each experiment or assay should include a negative control (solvent control) in order to check the background mutation and positive controls (reference/known mutagen control). Possibly, the chemicals selected as positive controls should be structurally related to the compound under test. Concentrations that demonstrate the effective performance of each assay should be selected. Concurrent strain-specific positive and negative (solvent or vehicle) controls, both with and without metabolic activation, should be included in each assay. Some of the examples of strain-specific positive controls are: for strains TA1535 and TA100, sodium azide; for strain TA98, 2-nitrofluorene (without activation) and 2-anthramine (with activation); for TA1537, 9-aminoacridine; and for TA98, benzopyrene (with activation). Negative controls, consisting of solvent or vehicle alone, should be included without test substance and otherwise treated in the same way as the treatment groups. In addition, untreated controls should also be used unless there is historical control data demonstrating that no deleterious or mutagenic effects are induced by the chosen solvent.

### 5.7 Test Material and Solvents

Solid test substances should be dissolved or suspended in appropriate solvents or vehicles and diluted prior to treatment of the bacteria. Liquid test substances may be added directly to the test systems and/or diluted prior to treatment. Fresh preparations should be employed unless stability data demonstrate the acceptability of storage; unused portions should be discarded. Sterile distilled water is the good choice of solvent. However, chemicals that do not dissolve in water should be dissolved in dimethyl sulfoxide (DMSO). It dissolves numerous different kinds of chemicals, is miscible with water, and, at the amount used in the test (0.1 mL or less), is not toxic to bacteria. In cases where it is unsuitable, other solvents may be used, such as acetone, ethyl alcohol, diethyl formamide, and methyl ethyl ketone. It is

noteworthy that the solvent/vehicle should not be suspected of chemical reaction with the test substance and the concentration used should be compatible with the survival of the bacteria and the S9 activity [25].

All data available on the substance to be tested should be provided and recorded, including its lot or batch number, physical appearance, chemical structure, purity, solubility, reactivity in aqueous and nonaqueous solvents, temperature and pH stability, and sensitivity to light. A sample of each substance to be assayed for mutagenicity should be retained for reference purposes. The proposed uses of the test substance should be known, since antibiotics, surfactants, preservatives, and biocides pose special problems in bacterial mutation assays. Gaseous or volatile substances should be tested by appropriate methods, such as in sealed vessels or in glass desiccators [26, 27].

Among the criteria to be taken into consideration when determining the highest amount of test substance to be used are cytotoxicity and solubility in the final treatment mixture. Cytotoxicity may be detected by a reduction in the number of revertant colonies or by a clearing or attenuation of the background lawn. The cytotoxicity of a substance may be altered in the presence of metabolic activation systems. Insolubility should be assessed as precipitation in the final mixture under the actual test conditions. The recommended maximum test concentration for soluble noncytotoxic substances is 5 mg/plate or 5 μL/plate. For noncytotoxic substances that are not soluble at 5 mg/plate or 5 μL/plate, one or more concentrations tested should be insoluble in the final treatment mixture. Test substances that are cytotoxic below 5 mg/plate or 5 μL/plate should be tested up to a cytotoxic concentration. For volatile liquids, 0.5–5 mL should be used as the low and high dose, respectively. At least five to eight different analyzable concentrations of the test substance should be used with approximately half-log (i.e., $\sqrt{10}$ ) intervals between test points for an initial experiment. For most mutagens, there is a linear dose–response over a certain concentration range. The highest concentration of test substance is desired to show evidence of significant toxicity. If precipitate is present on any of the plates, it may interfere with automatic counting of the colonies. In such a situation, all plates in that series of doses and controls should be counted by hand.

## 6 EXPERIMENTAL PROCEDURE/DESIGN OF EXPERIMENTS

In the plate incorporation method, each bacterial strain is grown in nutrient broth, supplemented with antibiotics for 15–18 h, and contains about $1 \times 10^9$ bacteria/mL. One hundred microliters (0.1 mL) of bacterial culture is added to 2.0 mL of 45 °C molten top agar, containing histidine and trace amounts of biotin, 0.05–0.1 mL of the test compound (a range of doses) or solvent, and 0.5 mL of sodium phosphate buffer (1 M, pH 7.4) or S9 mix. S9 mix consists of S9 (usually between 4% and 30% by volume) [15, 25] to which has been added nicotinamide-adenine dinucleotide phosphate (NADP) and glucose-6-phosphate (which together provide energy for metabolism), phosphate buffer to maintain pH, and salts of magnesium and potassium. A set of tubes is also prepared without S9. This is to check whether the test chemical can cause mutation without the need for metabolic activation. Chemicals of this type are directly acting mutagens: certain directly acting mutagens can be made nonmutagenic by S9. The additions of bacteria, test chemical, and S9 mix are

made in rapid succession, in order to avoid the potentially harmful effects of the rather high temperature (45 °C) necessary to keep the soft agar molten. As soon as possible after mixing, the molten top agar mixture is poured onto 25–30 mL of solid 1.5% minimal glucose agar plates. The plate is shaken to distribute the top agar in a thin, even layer over the bottom agar and is placed on a level surface. As the top agar cools and solidifies, the plates are inverted and incubated for 48–72 hours at 37 °C; revertant colonies are then counted either manually or by an automatic colony counter. During the first few hours of incubation, all the *his*$^-$ bacteria will grow, since there is a trace of histidine present. When all the histidine has been used up, the bulk of bacteria will stop dividing, and a thin, visible confluent lawn of bacteria will be formed in the soft agar. However, bacteria that have sustained DNA damage leading to a mutation with the effect of reverting *his*$^-$ gene to *his*$^+$ will continue to divide, since they can now synthesize their own histidine from the ammonium salts in the bottom agar. Several tubes are set aside to act as "controls," that is, tubes that will receive the solvent but not the test chemical and will therefore indicate the background (spontaneous) level of mutation. If the spontaneous revertant colonies are more than 200 colonies/plate or chemical discolors the agar or when precipitate is present on the plate, it is better to use hand counting rather than automatic counting. All chemicals should be tested twice with triplicate plates for all doses and solvent controls for an adequate estimate of variation. Test dose levels are selected based on the known literature of the chemical or a structurally related substance. Several doses ranging from submicrogram to milligram levels of the test chemical should be tested at intervals, together with positive and solvent controls. The experiment should be repeated until a consistent picture emerges.

For the preincubation method, the test substance/test solution (usually 0.05 or 0.1 mL) is preincubated in liquid environment with the tester strain (0.1 mL, containing approximately $10^9$ viable cells) and sterile buffer (0.5 mL) or the metabolic activation system (0.5 mL) usually for 20 min or more at 30–37 °C prior to mixing with the overlay agar (2.0 mL) and pouring onto the surface of a minimal agar plate. Tubes are usually aerated during preincubation by using a shaker.

For testing gases, the tester strains are precultured with the nutrient broth. The reaction mixture, containing phosphate buffer, S9, cofactors, and the preculture strain, is poured into the agar plates and placed upside down, without their lids, in a plate holder. The plate holder is then placed in a 10 L tedlar bag/desiccator through an opening on one side of the bag made by scissors. The bag is then closed by folding the opening two or three times and sealing it with adhesive tape. Air in the bag is removed and then the gas is released at a fixed amount per plate. The bacterial plates in the bag are kept at 37 °C for a time period. After termination of the exposure, the test substance (gas) in the bag is removed. The tape is peeled off and the bag is left to stand for 30 min, then covered with a lid and incubated for 24 h. Next, the revertant colonies are counted [17]. The results are expressed as number of revertants per plate based on moles of gas used. Volatile materials can be tested at reduced temperatures or at raised pressure. Alternatively, if the chemicals are volatile, the plates are exposed to known concentrations (v/v) of the compound and are incubated in closed systems or glass desiccators. However, bactericidal substances may be studied using the liquid suspension assay method. A treat-and-wash bacterial mutation test is employed to evaluate the mutagenicity of soluble biological material, capable of releasing histidine and tryptophan. Histidine and tryptophan

can cause overgrown background bacterial lawns. In such cases, the procedure is modified to include a longer preincubation step for 90 min followed by a washing step to remove the amino acids prior to plating so as to prevent overgrown lawns. After the 90 min of preincubation, 15 mL of wash solution of oxoid nutrient broth in phosphate buffered saline (1 : 7) is added. The washed bacterial pellet is collected by centrifugation at 2000*g* for 30 min. Supernatant is discarded and the bacterial pellet is resuspended in the residual supernatant (~0.7 mL of the supernatant) prior to plating [28].

It should be noted that all plates in a given assay should be incubated at 37 °C for 2 or 3 days before being scored for the number of revertant colonies per plate. It is important to ensure that volatile test compounds and gases are incubated in closed systems or glass desiccators. After the incubation period, it is essential to check the background lawn of both treated and control plates for toxic effects (thinning of the lawn) or excess growth, which may indicate the presence of amino acids in the test material. All plates within a given experiment should be incubated for the same time period. If a clear positive response is obtained in the first experiment, then a second or confirmatory experiment may not be required. If a negative response or ambiguous response occurs, it may be necessary to perform a second experiment or additional experiments. The result of a bacterial mutation assay, whether positive or negative, is expressed as mean revertants per plate and standard deviation. Statistics are used to support the interpretation of the data.

## 7 CRITICAL FACTORS IN THE PROCEDURE

The bacterial mutation test forms the basis of all screening programs. There are several conditions that must be met in order to ensure a satisfactory test. It is essential that a baseline protocol should be written before starting a screening program. Methods for the preparation and storage of S9 and bacterial strains and other procedures should be thoroughly checked by performing assays with reference mutagens and authenticated bacterial strains, under conditions prescribed by the chosen protocol. Bacterial strains should be checked regularly for their characteristic genetic traits, including amino acid requirement, background mutation, induced mutation with reference mutagens, presence of plasmids where appropriate, and presence of cell-wall and DNA-repair mutations [29, 30]. Advice should be sought from experienced investigators. The strains should also yield spontaneous revertant colony counts within the frequency ranges expected from the laboratory's historical control data and preferably within the range reported in the literature. Deviations from standard procedures need to be scientifically justified. The following technical details are not intended as a defined recommended protocol, but represent good current practice and good criteria for successful bacterial tests.

## 8 TROUBLESHOOTING

### 8.1 Background Lawn

A faint background lawn of bacteria indicates the presence of trace amounts of histidine in the agar, which allows all the bacteria to undergo several divisions,

producing a faint background lawn of bacteria. Since dividing cells are more suitable for mutation during DNA replication, the background lawn serves as a good indicator. If the amount of histidine in the agar is increased, mutagenesis may be enhanced but will also cause a heavy growth of the background lawn, obscuring the revertants. Plate without background lawn, but with colonies, may not contain true revertants, as they arise due to the surviving bacteria that live on the histidine present in the agar and therefore should not be scored. If the background lawn is thin compared to the control, it indicates toxicity of the chemical to the bacteria. To check whether colonies are true revertants, they should be streaked out onto minimal plates containing biotin but no histidine. True revertants will be able to grow, whereas microcolonies arising due to toxicity will not.

### 8.2 Spontaneous Reversion

Spontaneous reversion (to histidine independence) can be measured and expressed as the number of spontaneous revertants per plate. The number of spontaneous revertants is dependent on two populations: (1) pre-existing mutants in the inoculum and (2) mutants arising following divisions on the plate. Each tester strain should revert spontaneously at a frequency that is characteristic of the strain. Also, each laboratory should establish its own historical range, as these ranges are likely to vary from laboratory to laboratory. Any deviation from historical range may be due to altered characteristics of the strain and should be tested. It should also be noted that the number of spontaneous revertants varies between experiments; therefore, it is recommended that at least three control plates be included for each strain in a mutagenicity assay.

Abnormal high spontaneous reversion may indicate contamination or the accumulation of back mutations by repeated subculturing. If this occurs, the strain should be recovered by reisolation from the frozen master copy. Spontaneous reversions are influenced by histidine concentration in agar and also by the mutagens in the environment of the bacteria, such as the reagents used to perform the assay—for example, ethylene oxide used to sterilize plastic ware. If a test substance is derived from biological material that causes an increase in mutant colonies in a bacterial mutagenicity test, it is possible that such an increase may be due to the presence of amino acids or peptides in the test substance, at levels that interfere with standard procedure. In such situations, testing of amino-acid-free extracts of the test substance, with appropriate controls, is necessary. Abnormal test results may also be due to changes in the strain's characteristics.

### 8.3 Criteria for Checking Strain Characteristics

*His⁻ Character.* The *his⁻* character of the tester strains should be confirmed by demonstrating the histidine requirement for growth on selective agar plates.

*TrpE Marker.* Trypton auxotrophy of *E. coli* strain should be confirmed by streaking a overnight culture across the surface of glucose minimal defined agar plates with and without excess tryptophan (10–20 μg/mL). After overnight incubation at 37 °C, no growth should be observed on the plate without tryptophan, while full growth should be observed on the plate supplemented with tryptophan.

*uvr⁻*. *uvrA* mutation in *E. coli* and *uvrB* mutation in *S. typhimurium* should be confirmed by demonstrating sensitivity to UV light. The *uvrB⁻* deletion extends through the biotin gene, hence the need for biotin in those strains possessing the *uvrB* mutation. A sample of the cultures is streaked across a biotin control plate and then a histidine–biotin plate. The plates are incubated overnight at 37 °C and examined for growth on the histidine–biotin plates. There should be no growth on the control plates.

*rfa Character.* The *rfa* mutation permits large molecules, such as crystal violet, to enter and kill the bacteria. The *rfa* character of the tester strains is confirmed by determining crystal violet sensitivity. Crystal violet, added to the center of a sterile paper disk, is pressed lightly against the seeded plates to embed. The plate is then inverted and incubated. A clear zone of inhibition appears around the disk, indicating the presence of the *rfa* mutation.

*R factor.* Ampicillin resistance is a convenient marker enabling testing for the presence of the plasmid. The ampicillin resistant R factor is somewhat unstable and can be lost from the bacteria and hence should be checked routinely. This could be done by cross-streaking the culture to be tested against an ampicillin solution coated nutrient agar plate and allowing it to grow by incubating for 12–24 hours at 37 °C. Strains that contain the R factor will grow.

## 9 INTERPRETATION OF DATA AND REPORTING

If the test chemical causes an increase in revertant colonies, by inducing mutations in the genome of either *Salmonella typhimurium* and/or *Escherichia coli*, it indicates that the results are positive or the chemical is mutagenic in the species tested. In contrast, no colonies or very few colonies indicates the results are negative or the chemical is non-mutagenic. If the test chemical under investigation is found to be positive, there is no need to verify by additional testing. However, if it shows marginal or weak positive response, it should be further verified by additional testing.

Most experiments will give clear positive or negative results. However, in a few cases it is difficult to judge the nature of the test chemical. Hence, in the case of ambiguous results, further testing is necessary using different experimental conditions, such as modifying the concentrations or method of treatment (plate incorporation or liquid preincubation) or by adjusting metabolic activation conditions (source and concentration of S9 in the mix). Nevertheless, if the results are ambiguous or questionable even after repeated testing with modified protocols, it is advisable to carry out additional testing protocols using different test systems. Furthermore, negative results could also be confirmed by repeated testing using modified protocols. A test substance for which the results do not meet the above criteria is considered non-mutagenic in this test.

In general, data should be presented as the number of revertant colonies per plate. Individual plate counts, the mean number of revertant colonies per plate, and the standard deviation (mean revertant colonies per plate ± the SD or SEM) should be presented for each dose of the test substance, with positive (known mutagen) and negative (untreated and/or solvent) controls. Appropriate statistical methods may be used as an aid in evaluating the test results. There are several criteria for

determining a positive result, one of which is a statistically significant dose-related increase in the number of revertants.

Minimum fold increase is another method, expressed as a two- or threefold increase of revertants. However, this is insensitive for strains with relatively high reversion frequencies and is too sensitive for chemicals with low reversion frequencies [31]. Another criterion is based on the mutant frequency, which is expressed as the quotient of the number of revertant colonies over the number of colonies in the negative control. A mutagenic potential of a test item is assumed if the mutant frequency is 2.0 or higher. A possible mutagenic potential is assumed if the quotient ranges between 1.7 and 1.9. No mutagenic potential is assumed if all quotients range between 1.0 and 1.6.

## 10 DISCUSSION

Although the bacterial mutation assay is known for its simplicity, there are many critical factors that can influence the validity of the data, and close attention to every aspect of the experimental procedure is required. Establishment of consistent methods for every phase of the experiment is a prerequisite for obtaining reliable results and for running large numbers of tests in routine screening programs. A deficiency in just one area will jeopardize the whole enterprise. A bacterial mutagenicity assay simply determines whether the substance under investigation is or is not a bacterial mutagen in the presence and/or absence of an exogenous metabolizing system derived from a mammal (S9). Such a test cannot determine whether the test substance is mutagenic and/or carcinogenic in any other species. However, it may be concluded that a substance found to be mutagenic in properly conducted bacterial mutation assays should be regarded as potentially mutagenic or carcinogenic in mammals (including humans) until further evidence indicates otherwise.

## 11 SAFETY CONSIDERATIONS

As mentioned in earlier sections, the bacterial mutation assay is a simple and easy to perform assay. However, there are two important factors to be considered while performing such assays or in designing laboratories for bacterial assays: (1) prevention of contamination of cultures by other microorganisms and (2) protection of staff against exposure to hazardous test chemicals. Hence, experimental procedures should be conducted in appropriate biological safety cabinets in which a curtain of filter-sterilized air protects the worker from chemical exposure and the cultures from contamination.

Mutagen contamination should be avoided. Assays should be performed in a well-ventilated fume hood designated solely for this purpose. Air from the cabinets should be extracted outside the building through appropriate filters to prevent environmental contamination. Incubators should have precise temperature control, and those used for testing purposes should be in an area where the ventilation system removes any hazardous vapors from volatile test chemicals, when incubator doors are opened.

Culture media may be either purchased as ready-poured plates or prepared in the laboratory from basic ingredients. In the latter case, a clean working area must be available for pouring and drying plates. Using disposable test tubes, petri plates, and micropipetters with disposable tips and autoclaving of used materials before washing is highly recommended to avoid recirculation of contamination. During experimentation, the wearing of gowns, eye glasses, and gloves is advisable. Either manual or electronic devices are used for counting bacterial colonies.

Most importantly, a safe means of disposal of cultures should be provided; for example, they should be sealed in plastic or paper sacks in the laboratory and then incinerated. Furthermore, it is essential to devise operating procedures that minimize the hazards from storage, handling, weighing, pipetting, and disposing of mutagens/carcinogens and that deal with accidental contamination. Laboratories performing the test should follow the guidelines. Closed chambers or desiccators are recommended for testing highly volatile chemicals and gases [32].

## 12 CONCLUSION

Point mutations are the cause of many human genetic diseases and there is substantial evidence that point mutations in oncogenes and tumor suppressor genes of somatic cells are involved in tumor formation in humans and experimental animals. The bacterial mutation assay utilizes bacterial strains to detect point mutations. Therefore, chemicals that are mutagenic in bacteria are more likely to cause cancer than chemicals that are not mutagenic. Furthermore, the test has proved to be a valuable component for assessing mutagenic/carcinogenic potential.

## ACKNOWLEDGMENT

Grateful thanks to Dr. Mayumi Ishizuka for generously providing the photographs.

## REFERENCES

1. Iyer VN, Szybalski W. Two simple methods for the detection of chemical mutagens. *Appl Microbiol* 1958;6:23–29.
2. Ames BN. The detection of chemical mutagens with enteric bacteria. In *Chemical Mutagens: Principles and Methods for Their Detection*. New York: Plenum Press; 1971, Vol 1, pp 267–282.
3. Yanofsky C. Mutagenesis studies with *Escherichia coli* mutants with known amino acid (and base-pair) changes. In *Chemical Mutagens: Principles and Methods for Their Detection*. New York: Plenum Press; 1971, Vol 1, pp 283–287.
4. Gelboin HV, Miller JA, Miller EC. The *in vitro* formation of protein-bound derivatives of aminoazo dyes by rat liver preparations. *Cancer Res* 1959;19:975–985.
5. Miller EC, Miller JA. Mechanisms of chemical carcinogenesis: nature of proximate carcinogens and interactions with macromolecules. *Pharmacol Rev* 1966;18:805–838.

6. Miller EC, Miller JA. Studies on the mechanism of activation of aromatic amine and amide carcinogens to ultimate carcinogenic electrophilic reactants. *Ann NY Acad Sci* 1969;163:731–750.
7. Malling HV. Dimethylnitrosamine: formation of mutagenic compounds by interaction with mouse liver microsomes. *Mutat Res* 1971;13:425–429.
8. Garner RC, Miller EC, Miller JA, Garner JV, Hanson RS. Formation of a factor lethal for *S. typhimurium* TA1530 and TA1531 on incubation of aflatoxin B1 with rat liver microsomes. *Biochem Biophys Res Commun* 1971;45:774–780.
9. Ames BN, Durston WE, Yamasaki E, Lee FD. Carcinogens are mutagens: a simple test system combining liver homogenates for activation and bacteria for detection. *Proc Natl Acad Sci USA* 1973;70:2281–2285.
10. McCann J, Choi E, Yamasaki E, Ames BN. Detection of carcinogens in the *Salmonella*/microsome test: assay of 300 chemicals. *Proc Natl Acad Sci USA* 1975;72:5135–5139.
11. Sugimura T, Sato S, Nagao M, Yahagi T, Matsushima T, Seino Y, Takeuchi M, Kawachi T. Overlapping of carcinogens and mutagens. In *Fundamentals of Cancer Prevention*. Baltimore: University Park Press; 1976, pp 191–215.
12. Purchase IF, Longstaff E, Ashby J, Styles JA, Anderson D, Lefevre PA, Westwood FR. An evaluation of 6 short-term tests for detecting organic chemical carcinogens. *Br J Cancer* 1978;37:873–903.
13. Green MH, Muriel WJ. Mutagen testing using $TRP^+$ reversion in *Escherichia coli*. *Mutat Res* 1976;38:3–32.
14. McCann J, Ames BN. Detection of carcinogens as mutagens in the *Salmonella*/microsome test: assay of 300 chemicals: discussion. *Proc Natl Acad Sci USA* 1976;73:950–954.
15. Ames BN, McCann J, Yamasaki E. Methods for detecting carcinogens and mutagens with the *Salmonella*/mammalian microsome mutagenicity test. *Mutat Res* 1975;31:347–364.
16. Ames BN, Lee FD, Durston WE. An improved bacterial test system for the detection and classification of mutagens and carcinogens. *Proc Natl Acad Sci USA* 1973;70:782–786.
17. Yahagi T, Degawa M, Seino Y, Matsushima T, Nagao M. Mutagenicity of carcinogenic azo dyes and their derivatives. *Cancer Lett* 1975;1:91–96.
18. Kier LD, Brusick DJ, Auletta AE, Von Halle ES, Brown MM, Simmon VF, et al. The *Salmonella typhimurium*/mammalian microsomal assay: a report of the U.S. Environmental Protection Agency Gene-Tox Program. *Mutat Res* 1986;168:69–240.
19. Gatehouse D, Haworth S, Cebula T, Gocke E, Kier L, Matsushima T, et al. Recommendations for the performance of bacterial mutation assays. *Mutat Res* 1994;312:217–233.
20. Aeschbacher HU, Wolleb U, Porchet LJ. Liquid preincubation mutagenicity test for foods. *Food Safety* 1987;8:167–177.
21. OECD. *Guidelines for Testing of Chemicals*. Paris: Organization for Economic Co-operation and Development; 1997, pp 1–11.
22. Shanabruch WG, Walker GC. Localization of the plasmid (pKM101) gene(s) involved in $recA^+$ $lexA^+$—dependent mutagenesis. *Mol Gen Genet* 1980;179:289–297.
23. Vogel HJ, Bonner DM. Acetylornithinase of *E. coli*; partial purification and some properties. *J Biol Chem* 1956;218:97–106.
24. Elliott BM, Combes RD, Elcombe CR, Gatehouse DG, Gibson GG, Mackay JM, Wolf RC. Alternatives to Aroclor 1254-induced S9 in *in vitro* genotoxicity assays. *Mutagenesis* 1992;7:175–177.
25. Maron D, Katzenellenbogen J, Ames BN. Compatibility of organic solvents with the *Salmonella*/microsome test. *Mutat Res* 1981;88:343–350.

26. Prival MJ, Bell SJ, Mitchell VD, Peiperl MD, Vaughan VL. Mutagenicity of benzidine and benzidine-congener dyes and selected monoazo dyes in a modified *Salmonella* assay. *Mutat Res* 1984;136:33–47.
27. Araki A, Noguchi T, Kato F, Matsushima T. Improved method for mutagenicity testing of gaseous compounds by using a gas sampling bag. *Mutat Res* 1994;307:335–344.
28. Thompson C, Morley P, Kirkland D, Proudlock R. Modified bacterial mutation test procedures for evaluation of peptides and amino acid-containing material. *Mutagenesis* 2005;20:345–350.
29. Zeiger E, Pagano DA, Robertson IGC. A rapid and simple scheme for confirmation of *Salmonella* tester strain phenotype. *Environ Mutagen* 1981;3:205–209.
30. Maron DM, Ames BN. Revised methods for the *Salmonella* mutagenicity test. *Mutat Res* 1983;113:173–215.
31. Cariello NF, Piegorsch WW. The Ames test: the twofold rule revisited. *Mutat Res* 1996;369:23–31.
32. Hughes TJ, Simmons DM, Monteith LG, Claxton LD. Vaporization technique to measure mutagenic activity of volatile organic chemicals in the Ames/*Salmonella* assay. *Environ Mutagen* 1987;9:421–441.

# *IN VITRO* MAMMALIAN CELL MUTATION ASSAYS*

C. Anita H. Bigger,[1] Martha M. Moore,[2] and Robert H. Heflich[2]

[1]*U.S. Food and Drug Administration, Center for Drug Evaluation and Research, Silver Spring, Maryland*

[2]*U.S. Food and Drug Administration, National Center for Toxicological Research, Jefferson, Arkansas*

**Contents**

1 Introduction
2 Regulatory Guidances and Protocol Guidelines for *In Vitro* Mammalian Cell Mutation Assays
  2.1 Regulatory Guidances
  2.2 Protocol Guidelines
3 Endpoints and Unique Characteristics of *In Vitro* Mammalian Cell Mutation Assays
  3.1 Differences in the Spectra of Genetic Events Detected by *In Vitro* Mammalian Cell Mutation Assays
  3.2 Differences in the Spectra of Genetic Events Detected by the MLA, the *In Vitro* Chromosomal Aberrations Assays, and the Ames Test
  3.3 Correlation of *In Vitro* Mammalian Cell Mutation Assays with Rodent Bioassays
4 Criticisms of *In Vitro* Mammalian Cell Assays
  4.1 False Positives, High Concentrations, and Cytotoxicity Criteria
  4.2 Hazard Identification and Risk Assessment
5 Protocols for *In Vitro* Mammalian Cell Mutation Assays
  5.1 Commonalities and Differences in Protocols for *In Vitro* Mammalian Cell Mutation Assays
  5.2 Detailed Protocols
References

*The views expressed in this chapter do not necessarily reflect FDA policy.

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

## 1 INTRODUCTION

The *in vitro* mammalian cell mutation assays used for preclinical safety testing detect forward mutations at the following target genes: (1) the thymidine kinase gene (*Tk*), (2) the hypoxanthine–guanine phosphoribosyl transferase gene (*Hprt*), and (3) the *E. coli* xanthine–guanine phosphoribosyl transferase gene (*xprt* or *gpt*), integrated into mammalian cell DNA as a stable bacterial transgene [1]. All three target genes detect base substitutions, frameshifts, deletions, insertions, and rearrangements within the locus; however, assays employing the *Tk* gene and the *xprt* transgene offer the additional advantage of detecting mutations involving multilocus deletion. The *Tk* gene also detects translocation, as well as mutations requiring two alleles, such as mitotic recombination, gene conversion, and chromosome nondisjunction [1–11].

Commonly used cell lines for *in vitro* mammalian cell mutation assays are L5178Y mouse lymphoma cells, Chinese hamster ovary (CHO), AS52, and V79 cell lines derived from Chinese hamsters, and TK6 human lymphoblastoid cells [1]. At this time, by far the most frequently used *in vitro* cell mutation assay in drug development is the Mouse Lymphoma Assay (MLA), utilizing L5178Y mouse lymphoma cells and the *Tk* target gene. This is primarily because the MLA, in addition to being capable of detecting the broadest spectrum of types of genetic damage of the *in vitro* genetic toxicology assays [1–11], is one of the assays in the standard battery of genotoxicity tests agreed upon by the International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH) [12]. The ICH *Guidance for Industry S2B Genotoxicity: A Standard Battery for Genotoxicity Testing of Pharmaceuticals* (ICH S2B) [12] also describes conditions under which the *Hprt* or *xprt* target genes in any of the cell lines mentioned above might be appropriate for assessment of the genotoxic potential of pharmaceuticals.

This chapter provides the following: (1) discussion of regulatory guidance for the use of *in vitro* mammalian cell mutation assays, as well as the role of *in vitro* mammalian cell mutation assays in the evaluation of pharmaceuticals for genetic toxicity; (2) comments on the validated and internationally accepted protocol guidelines available for these assays; (3) discussion of the advantages and disadvantages of the assays for determining the nature of the genotoxicity elicited by pharmaceuticals; and (4) study designs for the MLA, CHO/*Hprt*, and AS52/*xprt* assays, including recent changes to the MLA protocol, test criteria, and evaluation of results.

## 2 REGULATORY GUIDANCES AND PROTOCOL GUIDELINES FOR *IN VITRO* MAMMALIAN CELL MUTATION ASSAYS

### 2.1 Regulatory Guidances

The two primary ICH guidances devoted to genetic toxicology testing of pharmaceuticals are *Guideline for Industry: Specific Aspects of Regulatory Genotoxicity Tests for Pharmaceuticals* (ICH S2A) [13] and *Guidance for Industry S2B Genotoxicity: A Standard Battery for Genotoxicity Testing of Pharmaceuticals* (ICH S2B) [12].

ICH S2A provides guidance and recommendations on conducting the tests and evaluating the test results. The S2A defines the top concentration, cytotoxicity levels,

and a strategy for testing poorly soluble compounds with *in vitro* mammalian cell mutation assays. ICH S2A also provides general guidance for the evaluation of assay results.

ICH S2B defines the standard test battery for pharmaceuticals as: (1) a test for gene mutation in bacteria; (2) an *in vitro* test for the cytogenetic evaluation of chromosomal damage in mammalian cells or an *in vitro* MLA measuring *Tk* gene mutation; and (3) an *in vivo* test for chromosomal damage using rodent hematopoietic cells. Other regulatory agencies also recommend this battery of tests for screening chemicals [14]. During the early ICH discussions of the standard battery, some parties preferred a four-test battery that included both of the *in vitro* mammalian assays that are listed in the three-test battery as alternates. The ICH Expert Working Group (EWG), however, reviewed additional studies, some of which were specifically designed to address this issue. The EWG compared results in the two *in vitro* mammalian cell tests for compounds that were considered genotoxic but gave negative results in the bacterial test. These studies demonstrated a high level of congruence between the *in vitro* mammalian cell assays, leading the EWG to conclude that the MLA and the *in vitro* mammalian cell chromosomal damage test could be used interchangeably in the standard battery (see ICH S2B III(ii), Notes 3 and 4). Based on this evidence, the EWG compromised with a three-test battery in which either the *in vitro* mammalian cell cytogenetics assay or the *in vitro* MLA satisfied the second element of the battery.

The ICH S2B also allows modifications of the three-test battery (ICH S2B IV) to include *in vitro* mammalian cell mutation assays other than the MLA. For example, for compounds that are extremely toxic to bacteria or for compounds that interfere with the mammalian cell replication system, the guidance suggests employing two *in vitro* mammalian cell tests. The two tests should use different cell lines and different endpoints, that is, gene mutation and chromosomal damage (ICH S2B IV A). ICH S2B (Note 1) also lists the currently accepted target loci and cell lines for *in vitro* mammalian cell mutation assays: (1) for the *Tk* gene, mouse lymphoma L5178Y cells or human lymphoblastoid TK6 cells; (2) for the *Hprt* gene, CHO cells, V79 cells, or L5178Y cells; and (3) for the *xprt* gene, AS52 cells.

The ICH S2B defines standard procedures for *in vitro* mammalian cell mutation tests other than the MLA (ICH S2A V) and indicates that a range-finding test can substitute for a complete repeat of the assay, as long as certain criteria are met (ICH S2B Note 10). For the MLA, the guidance states that the protocol must include testing with and without exogenous metabolic activation, with appropriate controls, and with treatment for 3 or 4 hours. If the short treatment without metabolic activation is negative, then that arm of the test should be repeated with treatment for 24 hours. An acceptable MLA also includes positive controls that induce small colonies and colony sizing for both positive and negative controls and at least one positive test compound dose, if present.

Both ICH S2A and ICH S2B were finalized approximately a decade ago. At the present time, there is discussion of revising these two guidances through the ICH Maintenance Procedure. That process, however, is likely to require years of discussion to achieve an agreed upon, finalized guidance to replace S2A and S2B. In the interim, possible changes to the guidances are discussed at international genetic toxicology meetings and conclusions are disseminated through publications in the literature. For example, the ICH S2B states that the microwell (or microtiter)

method for the MLA is preferred over the agar method for use in the standard battery (ICH S2B Note 4). After extensive discussion, an informal agreement was reached between all the parties to the ICH EWG for genetic toxicology that both versions of the assay are acceptable for regulatory purposes. This information was made public through several publications [15, 16].

Other ICH guidances that govern the regulatory use of *in vitro* mammalian cell mutation assays are: (1) *ICH M3 Guidance: Non-clinical Safety Studies for the Conduct of Human Clinical Trials for Pharmaceuticals* [17]; (2) *ICH S1B Guidance: Testing for Carcinogenicity of Pharmaceuticals* [18]; (3) *ICH Q3A Guidance: Impurities in New Drug Substances* [19]; and (4) *ICH Q3B Guidance: Impurities in New Drug Products* [20].

ICH M3 states that, prior to the first human exposure to a pharmaceutical, *in vitro* assays should be performed to evaluate the ability of the pharmaceutical to induce mutations and chromosomal damage. The MLA is one option for addressing this recommendation.

ICH S1B states that additional genetic toxicology testing may be warranted in cases where the compound was negative in the standard battery but was positive in a carcinogenicity test with no evidence of an epigenetic mechanism. Additional tests could include changing the metabolic activation system used in the *in vitro* assays.

The decision trees for ICH Q3A and Q3B include a minimum screen for the genotoxic potential of impurities. The minimum screen consists of two *in vitro* studies, one to detect point mutations and one to detect chromosomal aberrations.

In addition to ICH guidances, regulatory agencies also may provide guidance to industry. The U.S. Food and Drug Administration (FDA) recently issued *Guidance for Industry and Review Staff: Recommended Approaches to Integration of Genetic Toxicology Study Results* [21]. The purpose of this guidance is to inform industry how the FDA Center for Drug Evaluation and Research (CDER) views positive genetic toxicology findings. The guidance gives recommendations on how to proceed with clinical studies when the results of genetic toxicology assays indicate the pharmaceutical may be capable of inducing genetic damage and cancer. The guidance recommends a number of options when one or more of the three assays in the ICH standard battery of genetic toxicology tests are positive. With respect to *in vitro* mammalian cell mutation assays, if any one of the three assays in the standard battery is positive, the guidance recommends first performing the fourth test in the standard battery and then applying a weight of evidence evaluation of the data. For example, a positive finding in an *in vitro* chromosomal aberration assay should be corroborated by positive findings in the MLA using the same exposure regimen. If the results of the MLA assay are negative, then doubt is cast on the significance of the positive finding in the *in vitro* chromosomal aberration assay. The guidance also emphasizes the need to carefully control cytotoxicity levels in the *in vitro* assays in order to avoid a secondary genotoxic effect that is not directly related to treatment with the test compound.

### 2.2 Protocol Guidelines

The following protocol guidelines are available for *in vitro* mammalian cell mutation assays: (1) *OECD Guideline for the Testing of Chemicals: In Vitro Mammalian Cell Gene Mutation Test* [22]; (2) *EPA Health Effects Test Guidelines OPPTS 870.5300:*

*In Vitro Mammalian Cell Gene Mutation Test* [23]; (3) Reports from the MLA Workgroup of the International Workshop on Genotoxicity Testing (IWGT) from meetings in Washington, District of Columbia, in 1999 [16], New Orleans, Louisiana, in 2000 [24], Plymouth, England in 2002 [25], Aberdeen, Scotland in 2003 [26], and in San Francisco, California, in 2005 [27]; and (4) *FDA CFSAN Redbook 2000: IV.C.1.c. Mouse Lymphoma Thymidine Kinase Gene Mutation Assay* [28].

The Organization for Economic Cooperation and Development (OECD) and the U.S. Environmental Protection Agency (EPA) guidelines on the *in vitro* mammalian cell gene mutation test are similar and both give general recommendations covering all the target genes and cell lines used for *in vitro* mammalian cell gene mutation testing. However, there are significant differences between the MLA and the other *in vitro* mammalian cell mutation assays employing the *Hprt* and *xprt* genes, and these guidelines do not address those differences in adequate detail for practical use.

The IWGT convened the MLA Workgroup to address a number of issues concerning the MLA. Through a series of meetings beginning in 1999, the Workgroup reached consensus on five areas: (1) acceptable versions of the assay, the soft agar [29], and microwell versions [30, 31]; (2) a common cytotoxicity measure for both versions of the assay; (3) acceptability of a 24 hour treatment arm; (4) microwell sizing and colony counting; and (5) criteria for data acceptance and approaches for data evaluation. At the first meeting in Washington, DC, the Workgroup agreed that both versions of the assay are equally acceptable and formulated a strategy for reaching consensus on the other four issues. Since the Workgroup agreed that more data were necessary before decisions could be made, data were gathered in ten laboratories for analysis and discussion at the subsequent meetings. Since the first meeting, the MLA Workgroup has reached consensus for all areas of focus and has published its recommendations for conducting the assay in five publications [16, 24–27].

The FDA CFSAN Redbook 2000 published a chapter on the MLA in October 2001. This chapter presented a detailed protocol for the assay; however, because it was published before the IWGT MLA Workgroup had completed its work, the chapter did not contain all of the MLA Workgroup recommendations. A recent update of the Redbook 2000 (April 2006) includes the IWGT MLA Workgroup recommendations through reference to the MLA Workgroup publications.

## 3 ENDPOINTS AND UNIQUE CHARACTERISTICS OF *IN VITRO* MAMMALIAN CELL MUTATION ASSAYS

### 3.1 Differences in the Spectra of Genetic Events Detected by *In Vitro* Mammalian Cell Mutation Assays

As indicated previously, the *in vitro* mammalian cell assays employing the *Tk, Hprt*, and *xprt* genes all detect base pair substitutions, frameshifts, deletions, insertions, and rearrangements within the locus; however, the MLA offers the additional advantage of being able to detect mutations involving the loss of multiple genes and recombination between chromosome homologs. This is because the *Tk* gene in the MLA cell line (L5178Y $Tk^{+/-}$-3.7.2C) is on an autosome (mouse chromosome 11) and it is heterozygous; that is, one of its alleles is wild-type while the other is a

nonfunctional mutant. The same is true of the *TK* gene in TK6 cells, although the *TK* gene lies on chromosome 17 in human cells, and the inactivating mutations in the negative alleles of L5178Y $Tk^{+/-}$ and TK6 cells are different. In contrast, the hemizygous (or functionally hemizygous) location of the *Hprt* gene on the X chromosome makes it unlikely that mutations involving multilocus deletion or recombination will be recovered in assays employing this target gene [4].

In order to facilitate the identification of mammalian cells that have undergone a mutation, gene targets have been chosen so that single mutations can confer a selectable phenotype to the cells. One way to do this is to choose a target gene that is located on the X chromosome. This was the strategy used for the development of the CHO/*Hprt* assay and other assays employing the *Hprt* gene. Another approach, and the one used for the development of the MLA, is to create a cell strain in which one of the alleles of an autosomal target gene is a nonfunctional mutant. This is the case for the L5178Y $Tk^{+/-}$-3.7.2C strain. A major (theoretical) difference between these two types of mutational targets relates to their ability to detect multilocus deletions. The function of genes vital to the survival of the cell that flank the *Tk* gene can be provided by their homologs on the $Tk^{-}$ chromosome; thus, mouse lymphoma cells will survive even if the $Tk^{+}$ allele and vital surrounding genes are deleted from the $Tk^{+}$ chromosome. In the case of the *Hprt* gene, large deletions may never be seen because there is no back-up in the cell to provide the function of vital genes eliminated along with the *Hprt* gene; there is only one functional copy of each gene on the X chromosome, even in female cells. The presence of two functional alleles allows the MLA to detect mitotic recombination/gene conversion and those chromosomal events that typically are detected by cytogenetics assay, such as large structural deletions, translocations, and aneuploidy. None of these are detected by the CHO/*Hprt* assay or other assays employing the *Hprt* gene [7–11]. These theoretical considerations form the mechanistic rationale for treating the MLA and *in vitro* chromosomal damage assays as equal alternatives in regulatory testing batteries.

An additional feature of the MLA is that the mutant colonies that are detected in the assay fall into two size categories: small colonies with a slow growth rate and large colonies with a wild-type growth rate. The molecular and cytogenetic nature of small and large colonies has been studied extensively over the years [7–9, 32, 33], but the reasons underlying these growth phenotypes still are incompletely understood. In general, small colony mutants are induced by chemicals that are regarded as clastogens; that is, they have the ability to break chromosomes [34]. In contrast, large colony mutants are more likely to be induced by chemicals that induce small genetic lesions, such as point mutations. Thus, the mutagenic mechanism of a test article can be inferred from the proportion of small and large colonies induced by that test article.

The AS52 cell assay uses a CHO cell line that has a nonrevertable *Hprt* deletion mutation, but that carries a single, expressed copy of the bacterial *xprt* (or *gpt*) gene on one of its autosomes [35, 36]. The *xprt* gene is the functional equivalent of the *Hprt* gene so that *xprt* mutants are selected by 6-thioguanine (TG) in the AS52 assay similarly to how *Hprt* mutants are selected in the CHO/*Hprt* assay (see below for a description of mutant selection). The AS52 assay has some of the theoretical advantages of the MLA for detecting mutations: because the *xprt* gene is on an autosome, multilocus deletions that include the *xprt* gene are not necessarily lethal. Comparative analyses of the *Hprt* and AS52 cell assays indicate that a higher

percentage of mutations with deletions are detected by the AS52 cell assay and that AS52 cells are more sensitive to the mutagenicity of clastogenic agents [37–41]. Because AS52 cells contain only one copy of the *xprt* gene, however, the AS52 cell assay does not detect mutations requiring two alleles (e.g., mitotic recombination) as does the MLA.

### 3.2 Differences in the Spectra of Genetic Events Detected by the MLA, the *In Vitro* Chromosomal Aberrations Assays, and the Ames Test

The MLA detects mutations that do not compromise the viability of the cell; that is, the assay detects only heritable changes in DNA sequence. Besides relatively small changes within the *Tk* gene, these mutations include large alterations (e.g., translocations, deletions) that also can be detected cytogenetically. The *in vitro* chromosomal aberration assay detects only microscopically visible (i.e., relatively large) chromosome alterations, many of which are not heritable to daughter and granddaughter cells because they are lethal events. Like the MLA, the Ames test detects only heritable alterations in DNA. However, the Ames test employs a series of specifically designed mutant tester strains, and each tester strain detects a limited spectrum of either base pair substitution or frameshift mutations. Because sequence context can cause striking differences in mutagenic frequencies at a given site [42], the specificity of the mutational target and the use of a bacterial chromosome rather than eukaryotic chromosomes as the target for mutation severely restricts the spectrum of point mutations detected in the Ames test. Unlike the MLA, the Ames test is incapable of detecting mutations larger than frameshifts, including mutations requiring chromosome homologs and sequence alterations generally considered chromosome-type mutations. Finally, the mutations that are detected by both the MLA and the Ames test alter the phenotype of the test cell. For the MLA, only mutations that inactivate the $Tk^+$ allele will be identified, while only mutations that restore the wild-type function of the target sequence will be scored as mutants in the Ames test. The only limitation on scoring sequence changes in cytogenetic assays is that they be visible microscopically.

### 3.3 Correlation of *In Vitro* Mammalian Cell Mutation Assays with Rodent Bioassays

Early in the development of the field of genetic toxicology, the various assays were claimed to be short-term tests that could be used to identify carcinogens. Over the years, a number of studies have been conducted to assess the ability of the assays to predict whether a chemical is a carcinogen. A very influential study conducted by the National Toxicology Program (NTP) used data generated by contractors specifically for the analysis [43]. More recently, Kirkland et al. [44, 45] and Matthews et al. [46, 47] have used extensive genetic toxicology databases compiled from a variety of sources with the goal of obtaining a more current assessment. All of these studies report that *in vitro* mammalian mutation assays have relatively low predictive ability for rodent carcinogenicity.

Unfortunately, the assays, particularly the MLA and the *in vitro* aberration assays, have evolved over the years, not only with respect to their protocols, but also in the acceptance criteria and the evaluation of data. This has resulted in a large number

of assays that do not meet the current guidelines being included in the databases used for these studies of assay predictive ability. With the exception of the EPA Gene-Tox Expert Committee reports conducted in the 1980s [48], there has been no expert panel review of the databases. It should be noted that the MLA Gene-Tox panel placed a large number of chemicals (particularly from the NTP studies) into a category of insufficient data to determine whether the response was positive or negative [48]. The major reasons for this situation were that the dose selection process used for the NTP studies often did not provide data in the critical portion of the dose–response curve and that the NTP used a unique set of criteria for their positive and negative calls.

Because there currently is no adequate database for the MLA and chromosome aberration assay, it is unclear as to how well these assays actually predict the carcinogenicity of chemicals. Furthermore, in interpreting the data from genotoxicity studies, including *in vitro* mammalian cell mutation assays, it is important to keep in mind that cancer is a multistep process, having multiple modes of action, and that mutation is only one of the possible key events involved in the etiology of tumors.

## 4 CRITICISMS OF *IN VITRO* MAMMALIAN CELL ASSAYS

### 4.1 False Positives, High Concentrations, and Cytotoxicity Criteria

There is currently a lively debate in the regulatory community concerning the issue of too many "false positives" in the *in vitro* mammalian assays, specifically the MLA and the chromosome aberration test. This statement often is made based on the fact that the Ames test and the *in vivo* micronucleus (MN) assays for a test agent are negative while the *in vitro* mammalian assays are positive. There are several factors that may contribute to the generation of these "false positives."

First, as already indicated, it is expected that there will be chemicals that are positive in the *in vitro* mammalian assays and negative in the Ames test. In particular, chemicals that exclusively cause chromosomal damage (clastogens) will only be detected in assays capable of detecting chromosomal alterations; they would be expected to be negative in the Ames test. Furthermore, the *in vivo* MN assay detects only those events that result in micronuclei, which are not representative of the full spectrum of genetic damage. In addition, the rodent hematopoietic system target in this assay may not be reflective of other tissues. Finally, adequate concentrations of the test agent may not reach the bone marrow target for MN formation.

Second, the term "false positive" is often applied when a chemical is positive in a genetic toxicology assay and negative in the cancer bioassay. This should not be an unanticipated finding; the induction of mutation is only one of the possible modes of action for tumor induction. It would be more appropriate to judge the *in vitro* mutation assays for their ability to predict the induction of *in vivo* mutation.

Third, *in vitro* assays are criticized for the use of test agent concentrations higher than those that can be achieved *in vivo*. It is important to understand that by design the *in vitro* assays have properties that are unique to the *in vitro* test conditions. These properties make it possible to use high concentrations of test article and very short treatment times compared to those used *in vivo* (e.g., cancer bioassays typi-

cally are conducted using 2-year exposures). High test article concentrations and short treatment times are consistent with the role of *in vitro* assays in risk assessment, which is hazard identification. The use of high doses for short durations is intended to reveal potential test-article-induced toxicities before the drug is administered *in vivo* for longer durations at lower doses. In addition, high concentrations may be necessary to ensure assay validity. During the development of the MLA, a panel of known carcinogens was tested to determine what concentrations would be necessary to achieve a positive result in this assay [49]. This study revealed that the lowest effective concentration ranged from less than 1 μg/mL to over 10 mg/mL. Recent studies [44, 45], using reevaluated MLA data, found a similar broad range (<1 to >10,000 μM) for producing a positive response in the MLA. Importantly, these studies reported that, if positive results were included only when obtained below 10 mM, currently the internationally agreed upon limit dose for MLA [13], 5 out of 29 rodent carcinogens would not be found positive in the MLA.

Fourth, the *in vitro* assays are criticized for using high cytotoxicity levels. It is important to note that the criteria for cytotoxicity levels in the *in vitro* mammalian cell assays were established in the same manner as those establishing the maximum high concentration: that is, tests of panels of known carcinogens established appropriate cytotoxicity levels [49]. The ICH S2A (II.A.2.b.iii) [13] recommendation that the highest concentration should produce at least 80% toxicity reflects the high levels of cytotoxicity necessary to detect some carcinogens in these assays.

However, high doses of the test materials can result in nonphysiological alterations of osmolality and pH, leading to false-positive responses [13, 29]. Secondary effects from too much cytotoxicity also can cause increases in mutant frequency or chromosome aberrations. This raises the issue of biologically irrelevant responses (false positives) that may be obtained *in vitro*. Such positive responses would be expected to be observed only at the "higher" cytotoxicity levels, levels that may not be reached *in vivo*. In order to reduce/minimize the number of these biologically irrelevant "positive" responses, the ICH S2A (II.A.2.b.iii) [13] recommends caution with positive results obtained at levels of survival lower than 10% and requires demonstrating that osmolality and pH are maintained at appropriate levels during testing. For this reason, the IWGT MLA Workgroup set the maximum cytotoxicity level for the MLA, based on relative total growth (RTG), at 10%. As explained in more detail later, RTG is calculated by multiplying the relative suspension growth by the relative plating efficiency at the time of cloning for mutant selection. It is important to note that this DOES NOT represent killing 90% of the cells. For example, if a treated culture grows 50% as well as the negative control over the 2 day suspension growth period and plates 30% as well as the negative control during mutant selection, the RTG will be 15% (50% × 30%).

Despite the rationale for high concentrations and cytotoxicity levels and criteria for assay validity that set limit levels for both concentration and cytotoxicity, there is a "gray" zone where the concentration and/or cytotoxicity are close to the limit levels and the positive response in the assay is very weak. The biological relevance of such results to hazard identification is questionable. It is this "gray" zone that is driving much of the current debate on whether or not concentrations and cytotoxicity levels in the *in vitro* assays should be lowered and, if so, by what criteria. Since Kirkland et al. [44, 45] have shown that the current high concentration of 10 mM will not detect some rodent carcinogens, there also is concern that altering the test

criteria by further lowering the top concentration or reducing the cytotoxicity limit will compromise safety assessments of drugs in development.

### 4.2 Hazard Identification and Risk Assessment

For preclinical safety evaluation, the *in vitro* mammalian (and other) genotoxicity assays are used to identify compounds that pose a hazard to exposed humans, specifically a hazard to the genetic material. The assays also are used as a part of the weight of the evidence evaluation for whether a chemical is a carcinogen and to make other safety assessments as a part of the drug approval process. Currently, the *in vitro* mammalian cell mutation assays are not used for establishing a safe human dose because the *in vitro* conditions of these tests do not sufficiently mimic *in vivo* physiological conditions, such as endogenous metabolism and detoxification systems, and do not provide information on the targeting of specific tissues by a drug. Although the drug approval process generally does not include a quantitative risk assessment for compounds that are genotoxicants, it should be noted that genotoxicity assays play a key role in such evaluations in other regulatory contexts (e.g., the setting of allowable limits for drinking water, air, and food contaminants), where the public has little choice in exposure to hazardous substances [50].

## 5 PROTOCOLS FOR *IN VITRO* MAMMALIAN CELL MUTATION ASSAYS

### 5.1 Commonalities and Differences in Protocols for *In Vitro* Mammalian Cell Mutation Assays

All of the assays include (1) cell preparation; (2) a preliminary cytotoxicity test to determine initial doses for the definitive assays; (3) the use of an exogenous metabolic activation system in one arm of treatment; (4) an expression period that allows fixation of DNA damage generated by the test article and depletion of the wild-type enzyme produced by the target gene; (5) a selection period during which the cells are treated with an agent that is toxic to nonmutant cells; (6) collection of data on the number of mutant versus wild-type cells, the level of cytotoxicity, and the cloning efficiency; (7) data analyses to determine cytotoxicity, cloning efficiency, and mutant frequency; and (8) evaluation of the validity of the test and the response to the test article.

The assays differ procedurally in a number of ways: the methods used for culturing and cloning of cells, the length of the expression period, the agents used for selection of mutants, and the manner of data collection. For example, the CHO/*Hprt* assay uses attached monolayer cultures and single-cell cloning of attached cells, while the MLA uses cultures of unattached cells in suspension and cloning either by immobilization in agar or by plating at limiting-dilutions in 96-well microtiter plates. However, the major difference in the assays stems from the differences in the target genes for mutation detection, as exemplified by the two most commonly used mammalian cell mutation assays, the MLA and the CHO/*Hprt* assay.

The *Tk* target gene used for the MLA is located on mouse chromosome 11 and codes for thymidine kinase (Tk), an enzyme involved in thymidine triphosphate synthesis [49, 51]. Thymidine triphosphate can be generated through metabolism of

deoxyuridine monophosphate by thymidylate synthetase (*de novo* pathway) or by phosphorylation of exogenous thymidine by Tk (pyrimidine salvage pathway). The pyrimidine analogue trifluorothymidine (TFT) is also a substrate for Tk. However, phosphorylation of TFT makes cells vulnerable to its cytotoxic and cytostatic effects. Thus, TFT can be used as a selective agent for mouse lymphoma cells that have been mutated (either spontaneously or by the the test article) from $Tk^{+/-}$ (Tk-proficient) to $Tk^{-/-}$ (Tk-deficient) [52]. Mutation of the $Tk^{+}$ allele to $Tk^{-}$ renders the cell resistant to the toxic effects of TFT; the cells remain viable because DNA synthesis can still be carried out by the *de novo* synthetic pathway. When test-article-exposed cells subsequently are grown in medium containing TFT, only $Tk^{-/-}$ cells can grow and the mutant frequency can be determined from the number of clones appearing in TFT-containing medium. Because cells in culture do not clone with 100% efficiency, the mutant frequencies are calculated by adjusting the number of clones observed in the presence of TFT with the cloning efficiency of cells from the same culture in medium without TFT.

The X-chromosome-linked *Hprt* target gene used for the CHO/*Hprt* assay codes for an enzyme in another salvage pathway, one involved in purine nucleotide synthesis [53, 54]. The Hprt enzyme catalyzes the transfer of a phosphoribosyl moiety to guanine or hypoxanthine, thereby creating a purine mononucleotide. The purine analogue 6-thioguanine (TG) is also a substrate for Hprt, generating 6-thioguanine monophosphate (TGMP). TGMP inhibits purine nucleotide biosynthesis through a feedback mechanism and is, in addition, toxic to the cell through incorporation into DNA and RNA. Thus, as was the case with TFT and the MLA, TG can be used to select cells in which the enzyme coded by the *Hprt* target gene has been mutated and rendered inactive. Mutated cells remain viable in the presence of toxic concentrations of TG because purine nucleotides continue to be produced via an endogenous *de novo* pathway that does not involve Hprt.

As indicted earlier, the product of the *xprt* transgene of AS52 cells is the functional equivalent of Hprt, the only difference being that xprt will phosphorylate xanthine as well as guanine and hypoxanthine [35]. From a methodological standpoint, TG can be used to select xprt-deficient AS52 cell mutants using the same methods developed for selecting Hprt-deficient mutants in the CHO/*Hprt* assay [55].

## 5.2 Detailed Protocols

### *MLA*

1. *Principle of the Assay.* The purpose of the MLA is to determine the ability of a test article to induce forward mutations in the L5178Y $Tk^{+/-}$ mouse lymphoma cell line. The mutations are detected by growth of colonies in the presence of TFT.

2. *Soft Agar Versus Microwell Versions of the MLA*. There are two equally acceptable methods for conducting the MLA: (1) the soft agar method [29] and the microwell (or microtiter) method [30, 31]. The principal difference between these two methods is the technique used for cloning. In the first method, single cells are immobilized in soft agar medium in petri dishes, allowed to grow into clones, and counted. In the microwell method, cells are cloned by limiting dilution in 96-well

microwell plates. Over the years, some additional differences between the two methods developed, mainly involving the calculation of mutant frequency and the determination of cytotoxicity. Since 1999, the IWGT MLA Workgroup has been bringing the two methods into alignment with the goal of developing an internationally harmonized guideline for the conduct and interpretation of the assay. The consensus agreements [16, 24–27] for protocol changes in the MLA are included below. Additional updated descriptions of the protocol can be found in the FDA CFSAN RedBook 2000 [28] and in Chen and Moore [56].

3. *Cells.* It is important that the $Tk^{+/-}$-3.7.2C subline of L5178Y mouse lymphoma cells, the subline used to develop and validate the assay, be used for the MLA. Cells should be checked periodically for mycoplasma contamination and karyotype stability. Both a banded and spectral karyotype for the cell line have been published [10, 11, 57]. Attention should be paid to population doubling time (between 8 and 10 hours) when developing master stocks. Cells should be maintained in log-phase culture for no longer than approximately 3 months; at which time, a new working stock culture should be initiated. Some laboratories prefer to use a freshly thawed vial of cells for each experiment. The cells are used for the assay after they have recovered from the thawing and are growing normally.

4. *Media and Culture Conditions.* Both Fischer's Medium for Leukemic Cells of Mice and RPMI 1640 medium are appropriate for the MLA assay but these two media are used somewhat differently, as detailed in Chen and Moore [56]. It is important that the culture conditions promote optimal growth during the expression period and maximize colony forming ability for both mutant and nonmutant cells. In addition, the culture conditions must ensure optimal growth of small colony *Tk* mutants. In particular, different lots of horse serum should be tested for their ability to support optimal growth in suspension culture, high plating efficiency, and high small colony mutant recovery.

Maintenance of pH and osmolality in the physiological range (pH 7.0 ± 0.4 and 300 ± 20 mOsm) is critical for avoiding false-positive results [29].

5. *Preparation of Cultures.* Cells from stock cultures are seeded into culture medium and incubated at 37 °C. Before use in the assay, the cells are cleansed of preexisting mutants by growth in a medium containing methotrexate, thymidine, hypoxanthine, and guanine so that Tk-deficient cells are inhibited while Tk-competent cells grow optimally [29].

6. *Metabolic Activation.* Cells lose metabolic competence after a few passages as a cell line. Since many chemicals require metabolism to show genotoxic potential, test articles must be tested in the presence and absence of an exogenous metabolic activation system. The most commonly used metabolic activation system contains a postmitochondrial supernatant (S9) prepared from the livers of Aroclor-1254-induced male rats supplemented with a standard mixture of cofactors. The amount of S9 varies from 1% to 10% v/v in the final test medium. The standard S9 is available commercially or can be produced by published methods [29]. Alternative metabolic activation systems such as the use of S9 from another species or the use of hepatocytes [58, 59] should be justified scientifically.

7. *Solvent/Vehicle.* The solvent/vehicle is chosen principally on the basis of its ability to dissolve the test article. The solvent/vehicle also must not reduce cell survival or S9 activity or react with the test article. An aqueous solvent (water or culture

medium) is preferred, but organic solvents, such as dimethyl sulfoxide (DMSO), acetone, or ethyl alcohol, can be used in the MLA when necessary to achieve sufficient solubility of the test article.

8. *Dose Selection.* Dose selection is based on the cytotoxicity and solubility of the test article, and the criteria for determining the highest dose described in ICH S2A [13]. For *in vitro* mammalian cell mutation assays, ICH S2A (II.2) defines the top concentration for freely soluble, nontoxic compounds as 5 mg/mL or 10 mM, whichever is lower. ICH S2A recognizes that some genotoxic compounds are active in *in vitro* mammalian cell mutation assays only at cytotoxic doses and that excessive cytotoxicity can induce secondary genotoxic events unrelated to the drug being tested. Keeping these facts in mind, the ICH EWG agreed that *in vitro* mammalian cell mutation assays should produce at least 80% cytotoxicity (i.e., no more than 20% survival), but that positive results obtained at survival levels below 10% are suspect and generally do not serve as a basis for the evaluation of test results. The IWGT MLA Workgroup also agreed on circumstances under which a test article could be deemed nonmutagenic when there is no test culture with an RTG (a measure of cytotoxicity; see below for definition) between 10% and 20% [24]: (1) there is no evidence of mutagenicity in a series of data points falling between 20% and 100% RTG and there is one negative data point between 20% and 25% RTG; (2) there is no evidence of mutagenicity in data points falling between 25% and 100% RTG and there is a negative data point between 1% and 10% RTG.

The IWGT MLA Workgroup addressed the issue that the two versions of the assay (agar and microwell) had evolved to use different cytotoxicity measures [16, 24]. The cytotoxicity measure commonly used for the microwell method was the relative survival (RS), determined by cloning cells immediately after treatment. This measurement was not made in laboratories using the soft agar method. The cytotoxicity measure for the soft agar method was the relative total growth (RTG), which is a combination of cell loss after treatment, reduction in growth rate during the expression period, and reduction in the cloning efficiency determined when the cells are plated for mutant selection. Another cytotoxicity measure, relative suspension growth (RSG), also was widely used for the selection of doses for the full mutation assay. The situation was further confounded by the use of two slightly different protocols for the microwell method, which resulted in different calculations of the RS, RSG, and RTG, such that a comparison of results between the two methods was not reliable. After additional data analysis and discussion, the Workgroup reached a consensus that the RTG would be the cytotoxicity measure for both the soft agar and microwell methods and that an adjustment—that includes any cytotoxicity that occurs during the treatment phase of the assay—would be applied to one of the two microwell method protocols. This adjustment makes the RS, RSG, and RTG values from both microwell methods and the soft agar version comparable.

The ICH S2A [13] also addressed the problems associated with testing insoluble compounds with *in vitro* mammalian mutagenesis assays. While recognizing that there is evidence that insoluble compounds may induce dose-related genotoxic activity in the insoluble range, precipitates may interfere with the scoring of the endpoint, render dose exposure uncontrollable, and make the compound unavailable to the cells. In the case of the MLA, cells are grown in suspension and are

normally pelleted by centrifugation at the end of the treatment period and resuspended in fresh medium to remove the test article. If the test article precipitates, it will be pelleted with the cells and taken up with them into fresh medium, thereby extending the treatment time and possibly interfering with scoring mutant colonies. The ICH S2A recommends the following approach for testing insoluble compounds. If evidence of cytotoxicity is missing, the lowest precipitating concentration in the final treatment medium should be designated the high concentration as long as it does not exceed the 5 mg/mL or 10 mM maximum for testing with mammalian cell lines.

A dose range-finding experiment may be used to determine appropriate doses to achieve optimum cytotoxicity levels in the definitive tests. For this preliminary assay, cells are treated in the presence and absence of exogenous metabolic activation for 3 or 4 hours. Sufficient doses (e.g., nine or more) are used to cover a wide dose range (e.g., 4-log dose range) up to the limit of solubility or to the limit dose (5 mg/mL or 10 mM). After suspension growth for 24 hours, cells from test-article-treated and solvent control cultures are counted and diluted. After another 24 hours of suspension growth, cells from all cultures are again counted. The cell counts are used to determine cytotoxicity based on total suspension growth over the 48 hours of culture as compared to the solvent control cultures. The results are used to select doses for the definitive assay.

9. *Controls.* Positive and negative controls should be included in each experiment. The negative control should be exposed to the solvent/vehicle in the same manner as cultures receiving the test article. Untreated controls also should be included if there are no data on the responses produced by the solvent/vehicle in the assay.

The positive control should demonstrate the induction of small colony mutants and one positive control should require metabolic activation to serve as a control for S9 activity. Cyclophosphamide, benzo[*a*]pyrene, and 3-methylcholanthrene are appropriate controls for the metabolic activation arm of the test. Methyl methanesulfonate is appropriate for the arm without metabolic activation.

10. *Treatment with Test Article.* For the definitive assay, single or duplicate cultures may be used, as long as a sufficient number of concentrations are used to ensure that the appropriate dosing range has been covered. Logarithmically growing cells in suspension are treated with the test article in the presence and absence of metabolic activation for 3 or 4 hours. If the test article is particularly insoluble, a longer time may be warranted. ICH S2B (V and Note 4) [12] recommends a 24 hour treatment in the absence of metabolic activation for all test articles testing negative with the 3 or 4 hour treatment. ICH S2B also states that the detection of aneugens is enhanced by the use of a 24 hour treatment using the microwell method. The IWGT MLA Workgroup considered the usefulness of this requirement and published a consensus report on its findings [27]. After analyzing 990 datasets, the group found that less than 2% of the chemicals were uniquely positive at 24 hours. They found that the 24 hour treatment was effective for insoluble chemicals that could not be tested at adequate levels of cytotoxicity. The group also found evidence that some (but not all) aneugens require the longer treatment time. Based on these data, the group agreed to continue supporting the ICH recommendation that a 24 hour treatment be used when short treatment results are negative or equivocal.

11. *Expression Time and Measurement of Mutant Frequency.* After treatment, the cells are pelleted by centrifugation and washed to remove the test article by suspension in fresh media and recentrifugation. Cultures are grown for 2 days to allow expression of the mutant phenotype [60]. Cells are then cloned in the presence and absence of TFT [52], using either the soft agar or microwell method [16], in order to determine the cloning efficiency of cells in the presence and absence of mutant selection. Plates with seeded cells are incubated for 11–14 days, followed by colony counting and sizing.

The two cloning efficiencies (with and without TFT) are used to calculate the mutant frequency for each treatment. Mutant frequencies are calculated differently when using the soft agar method and the microwell method. For the soft agar method, mutant frequency (MF) is calculated by dividing the number of TFT-resistant colonies by the number of cells plated for selection, corrected for the plating efficiency (PE) of the cells. PE is determined from the cloning efficiency of cells from the same culture grown in the absence of TFT. This calculation can be expressed as MF = (number of mutants/number of cells plated) × PE. For the microwell method, the Poisson distribution is used to calculate PEs for cells cloned without TFT and with TFT selection. From the zero term of the Poisson distribution, the probable number of clones/well ($P$) is equal to –ln(negative wells/total wells) and PE = $P$/(Number of cells plated per well). Mutant frequency is then calculated as MF = (PE(TFT selected)/PE(unselected)). Mutant frequency usually is expressed as TFT-resistant mutants per $10^6$ clonable cells or per $10^6$ survivors.

12. *Mutant Colony Sizing.* For the soft agar method, colonies are counted and sized using an automatic colony counter capable of evaluating the size of the colonies.

For the microwell method, colonies are identified by eye or low power microscope. Small colonies are less than one-quarter of the diameter of the well, while large colonies are more than one-quarter of the diameter of the well (standard 96-well flat-bottomed culture plate).

If the test article is positive, colony sizing should be performed on mutant colonies from at least one of the treated cultures, preferably from the highest acceptable concentration, and on mutant colonies from the negative and positive controls. If the test article is negative, colony sizing should be performed on mutant colonies from the negative and positive controls. The assay, and a negative evaluation, is not valid unless the positive control demonstrates the ability to induce small colony mutants. Colony sizing of the negative control will demonstrate that large colonies are able to grow appropriately.

13. *Assay Validity Criteria.* Data should include pH and osmolality determinations, information on test article solubility, test article concentrations in dosing solutions, PE and cell count determinations, colony counts and mutant frequencies for control and treated cultures, RS, RTG, and RSG determinations, and the results of colony sizing of mutant colonies from at least one concentration of the test article for a positive response and of mutant colonies from the negative and positive controls regardless of outcome. These data must demonstrate that appropriate physiological conditions were maintained during the test, that the test concentrations were appropriate in that either a positive mutant frequency response or an 80% reduction in RTG or a persistent precipitate was achieved, and that the assay was able to

recover small colony mutants. In addition, acceptance criteria (see below) for the negative/vehicle and positive controls must be met.

*Criteria for Negative/Vehicle and Positive Controls* The IWGT MLA Workgroup [26] reached consensus on acceptance criteria for the soft agar and microwell method negative/vehicle control parameters using the 3 or 4 hour treatment, as shown in the following table.

| Parameter | Soft Agar Method | Microwell Method |
|---|---|---|
| Mutant frequency | $35–140 \times 10^{-6}$ | $50–170 \times 10^{-6}$ |
| Cloning efficiency | 65–120% | 65–120% |
| Suspension growth | 8–32-fold | 8–32-fold |

In establishing these criteria, the Workgroup decided to include a measure of how well the negative/vehicle control cells grow during the 2 day expression phase. Suspension growth reflects the number of times the cell number increases from the starting cell density. Cells in stock cultures generally should undergo a 5-fold increase in number every 24 hours. Thus, over a 2 day period, there would theoretically be a 25-fold increase. The group was particularly concerned about poorly growing cultures and therefore set a minimum of an 8-fold increase. The maximum of 32-fold provides for reasonable errors in cell dilution and for a slightly variable rate in cell number increases.

The IWGT MLA Workgroup [27] also reached consensus on acceptance criteria for the soft agar and microwell method negative/vehicle control parameters for the 24 hour treatment, as shown in the following table. Suspension growth values were changed for the longer treatment time because it involves 3 days of suspension growth, rather than the 2 days for the 3 or 4 hour treatment.

| Parameter | Soft Agar Method | Microwell Method |
|---|---|---|
| Mutant frequency | $35–140 \times 10^{-6}$ | $50–170 \times 10^{-6}$ |
| Cloning efficiency | 65–120% | 65–120% |
| Suspension growth | 32–180-fold | 32–180-fold |

The IWGT MLA Workgroup [27] reached consensus on two equally appropriate criteria for an acceptable positive control response for both the short and long treatment times. (1) The positive control produces an induced mutant frequency (total mutant frequency minus concurrent negative control mutant frequency) of at least $300 \times 10^{-6}$ with at least 40% of the colonies being small colonies. (2) The positive control produces an induced small colony mutant frequency of at least $150 \times 10^{-6}$. For both approaches, the RTG must be greater than 10%. Some laboratories prefer to use more than one dose of their positive control. In particular, they prefer to use a dose that gives only small increases in mutant frequency. These laboratories can meet the positive control criteria by selecting at least two doses for the positive control (one of which meets the criteria).

*Confirmatory tests* There is no requirement for verification of a clear positive response in the MLA. Experiments that do not provide enough information to

determine whether the chemical is positive or negative should be clarified by further testing, preferably by modifying the test concentrations or metabolic activation conditions. Negative results achieved using the 3 or 4 hour treatment should be confirmed by repeat testing using 24 hour treatment in the absence of S9 activation (ICH S2B (V and Note 4)) [12]. However, performing only a 24 hour treatment (i.e., with no 3 or 4 hour treatment) in the absence of S9 activation is not acceptable unless the highest concentration is a precipitating concentration or the limit concentration. The reason for performing a 3 or 4 hour treatment is that higher concentrations can be used than are generally tolerated when cells are treated for 24 hours.

14. *Assay Evaluation Criteria.* There have been several different approaches to defining a positive result in the MLA. All involve a concentration-dependent or reproducible increase in mutant frequency, but the absolute increase in mutant frequency that differentiates a positive response from a negative response varies. In some laboratories, a twofold increase in mutant frequency over background is considered a positive response. In other laboratories, the approach developed by the U.S. EPA MLA Gene-Tox Workgroup [48] is used as guidance in interpreting data. Also, the United Kingdom Environmental Mutagen Society developed a statistical program for determining a positive or negative result in the microwell version of the assay [61].

Recently, the IWGT MLA Workgroup recommended the adoption of a global evaluation factor (GEF) and an appropriate statistical trend analysis to define positive and negative results in the MLA [24–27]. This approach also takes into account the recommendation in previous guidance documents [22, 28] that biological relevance should be a factor in data evaluation.

To arrive at a GEF, the Workgroup analyzed distributions of negative/vehicle mutant frequencies for the agar and microwell versions of the MLA that they gathered from ten laboratories. The Workgroup defined the GEF as the mean of the negative/vehicle mutant frequency distribution plus one standard deviation. Applying this definition to the data, the Workgroup arrived at a GEF of 90 for the agar method and of 126 for the microwell method. The GEF is applied in the following manner. If the mutant frequency of the negative control in a soft agar assay is $100 \times 10^{-6}$, then one treatment group must have a mutant frequency of at least $190 \times 10^{-6}$ (100 plus GEF of 90) to be called positive. In addition to evaluating the data against the GEF, the Workgroup recommended that the data be evaluated for the presence of a positive trend.

Following these recommendations, a test article is positive in the MLA if both the induced mutant frequency meets or exceeds the GEF and a trend test using the data is positive. A test article is negative if the induced mutant frequency is below the GEF and the trend test is negative.

Situations where either the GEF test or the trend test is positive should be considered on a case-by-case basis. Additional testing may be helpful to better define the assay response, particularly if the RTG is in the 10–30% range. While additional testing, which includes refining the dose range selection, will generally provide data for a definitive call, there will be some situations where the response will either vary between well-designed experiments or fluctuate within an experiment. Fluctuation within an experiment can occur when closely spaced doses are used to provide

extensive coverage of an important part of the dose response for a test article. Either of these outcomes can result in an "equivocal" (equal voice) determination for the MLA. In such a situation, the MLA results should be evaluated in the context of other information as a part of a weight of the evidence assessment of the potential hazard posed by the test chemical.

***CHO/*Hprt *Assay*** Compared with the extensive guidance available for conducting and interpreting the MLA, there is very little published material that specifically relates to the use of the CHO/*Hprt* assay (and the AS52 assay) for preclinical testing. This most likely is due to the secondary status accorded the CHO/*Hprt* assay in the ICH test battery and the subsequent lack of emphasis placed on the assay by the IWGT. The following protocols for the CHO/*Hprt* and AS52 assays were derived from original research reports using the assays, literature reviews of assay performance, including that of the EPA Gene-Tox Program for the CHO/*Hprt* assay, and the authors' experience with these assays in their own laboratories.

1. *Principle of the Assay.* The purpose of the CHO/*Hprt* assay is to determine the ability of a test article to induce forward mutations in the *Hprt* gene of CHO cells. The mutations are detected by growth of colonies in the presence of TG (or, in rare instances, 8-azaguanine (AG)).

2. *Cells.* CHO-$K_1$-$BH_4$ cells [62] are the standard cells used in this assay. This cell line is a proline auxotroph with a modal chromosome number of 20 and a population doubling time of 12–14 hours [53]; some CHO cell lines may have slightly longer doubling times. Cells should be checked periodically for mycoplasma contamination, genetic markers, and karyotypic stability.

3. *Media and Culture Conditions.* Ham's nutrient mixture F12 containing 5–10% heat-inactivated fetal bovine serum (FBS) is used for growing stock cultures. The medium is adjusted to pH 6.8–7.2 before the addition of cells. While most parts of the assay are performed with F12 plus FBS, hypoxanthine-free nutrient mixture F12 is used for selection of TG-resistant mutants. In addition, the medium used for selection is often prepared with dialyzed FBS. Osmolality should be maintained in the physiological range (300 ± 20 mOsm) when cells are treated with a test article.

Cells are typically plated at a concentration of $1 \times 10^6$/100-mm dish or $5 \times 10^5$/25-$cm^2$ T-flask and subcultured every 2–3 days to maintain exponential growth. Trypsin (0.05%) is generally used for cell detachment for subculturing.

4. *Preparation of Cultures.* Frozen cell stocks are prepared from exponentially growing cells that were cleansed of preexisting *Hprt* mutants in medium supplemented with hypoxanthine, aminopterin, and thymidine. Freshly thawed cells are plated in growth medium and incubated at 37 °C for 16–24 hours to allow time for cell attachment and growth to approximately $1 \times 10^6$ cells/25-$cm^2$ surface area.

5. *Metabolic Activation.* The rationale for using exogenous metabolic activation in the assay is the same as for the MLA. Also like the MLA, the most commonly used metabolic activation system uses an S9 prepared from the livers of Aroclor-1254-induced male rats. The standard S9 is available commercially or can be produced by published methods [29]. The S9 is supplemented with a standard mixture of cofactors as described by Machanoff and co-workers [63] and used at a concentration of 1–10% v/v in the final test medium. The concentration of S9 is determined

by preliminary tests of the cytotoxicity of the S9 and its ability to activate known metabolism-dependent mutagens (see below for activation-dependent controls). Alternative metabolic activation systems such as the use of S9 from another species or the use of hepatocytes should be justified scientifically.

6. *Solvent/Vehicle.* The solvents appropriate for this test system include but are not limited to the following, in order of preference: FBS-free F12, distilled water, DMSO, ethanol, and acetone. Nonaqueous solvent concentrations should not be greater than 1% and should be identical in all test cultures. The choice of the solvent principally depends on its ability to dissolve the test article. However, the suitability of the proposed solvent also should be evaluated by determining the effect of the dissolved agent on the pH and osmolality of the treatment medium, as well as by determining if the solvent reacts with the test article, the culture vessel, or the cells.

7. *Dose Selection.* As described for the MLA, dose selection is based on the cytotoxicity and solubility of the test article, and the criteria for determining the highest test dose described in ICH S2A [13]. For *in vitro* mammalian cell mutation assays, ICH S2A (II.2) defines the top concentration for freely soluble, nontoxic compounds as 5 mg/mL or 10 mM, whichever is lower. Thus, a preliminary cytotoxicity test should include a range of doses with a high dose of 5 mg/mL or 10 mM, unless limited by pH, osmolality, or solubility. Cells are seeded and allowed to grow for 18–24 hours followed by treatment with the solvent or the test article in the presence and absence of exogenous metabolic activation for 5 hours. Depending on how the mutagenicity assay is conducted (see later discussion), either immediately after the treatment or on the next day, the cells are trypsinized and reseeded into new plates at cloning density (100–200 cells/25-$cm^2$ T-flask or 60-mm plate). After incubation for 7–10 days, the colonies are fixed with absolute methanol, stained with methylene blue, and counted. Cytotoxicity is expressed as the cloning efficiency of test-article-treated cells relative to the cloning efficiency of the solvent control. When possible, the high dose is chosen to give a cell survival of 10–20%.

8. *Controls.* Appropriate negative controls are culture medium for the untreated control and solvent/vehicle for the solvent control. Appropriate positive controls are ethyl methanesulfonate or *N*-methyl-*N′*-nitro-*N*-nitrosoguanidine for direct-acting mutagens and 7,12-dimethylbenz[*a*]anthracene, benzo[*a*]pyrene, dimethylnitrosamine, or 2-acetylaminofluorene for promutagens requiring metabolic activation.

9. *Treatment with Test Article.* Exponentially growing cells are seeded into plates at a density of approximately $1 \times 10^6$ cells/55 $cm^2$ and incubated for 18–24 hours to achieve an approximate doubling of cell number. Next, the plates are washed with FBS-free F12 or Hank's Balanced Salt Solution (HBSS) and exposed to the test article or controls for 5 hours, generally with and without S9 activation. Duplicate cultures and three to five concentrations of the test article typically are used. After the treatment, the media are aspirated and the cells are washed with HBSS.

10. *Expression Time and Measurement of Mutant Frequency.* After the treatment, most laboratories culture the cells in complete F12 medium for 16–24 hours (overnight) before subculturing. This allows the various steps of the assay to be completed easily within an 8 hour working day. Alternatively, the cells can be subcultured immediately after treatment, a procedure viewed as giving more "accurate"

cytotoxicity data [54]. The posttreatment subculture involves establishing two sets of cultures from each treatment: one (100–200 cells/60-mm dish in triplicate) is incubated 7–10 days and used to determine cytotoxicity, while the second (no more than $1 \times 10^6$ cells/100-mm dish) is cultured for the expression of mutants. During the 7–9 day expression period, the cells are subcultured at 2–3 day intervals. At the end of the expression period, the cells are again split into two subcultures. One subculture is designated for the selection of mutant cells: $2 \times 10^5$ cells are plated in each of five (or more) 100-mm dishes using hypoxanthine-free F12 medium containing 10 μM TG. The other subculture is used for the determination of cloning efficiency: 100–200 cells are plated in 100-mm dishes using medium free of TG. After incubating both sets of dishes 7–10 days, the colonies are fixed, stained, and counted.

Cytotoxicity is determined from the plates seeded at cloning density up to 1 day after the treatment and is expressed as the cloning efficiency of the treated cultures relative to the solvent control. Mutant frequency is calculated by dividing the total number of mutant colonies by the total number of cells seeded into selection plates, corrected by the cloning efficiency of the cells seeded at the time of selection into plates without TG. Mutant frequency usually is expressed as TG-resistant mutants per $10^6$ clonable cells or per $10^6$ survivors.

11. *Assay Validity Criteria.* The cloning efficiency of the untreated and solvent controls must be greater than 50%. Spontaneous mutant frequency in the solvent and untreated controls may vary in different laboratories but should fall within 0–20 mutants per $10^6$ clonable cells [54, 64]. Assays with slightly higher spontaneous mutant frequencies may be acceptable, especially when the test agent produces a clear increase in mutant frequency; higher spontaneous mutant frequencies, however, may prevent the detection of weak mutagens. The positive control must induce a mutant frequency sufficiently greater than the mutant frequency of the solvent control—for example, three times that of the solvent control—and should exceed the minimum mutant frequency considered to be a positive result. The high dose for noncytotoxic, soluble test articles must be 5 mg/mL or 10 mM (whichever is lower). Alternatively, the high dose may be limited by either the test agent's solubility in the treatment medium or by a maximum cytotoxicity of approximately 10–20% relative cloning efficiency measured up to one day after exposure.

12. *Assay Evaluation Criteria.* Data should be reproducible for both positive and negative responses. However, an abbreviated preliminary or confirmatory assay coupled with a full (definitive) assay may be sufficient for demonstrating reproducibility [54]. ICH S2B V [12] states that a range-finding test may be a satisfactory substitute for a complete repeat of an *in vitro* mammalian cell mutation assay other than the MLA, if the range-finding test is performed with and without metabolic activation, with appropriate positive and negative controls, and with quantification of mutants. Positive compounds should exhibit a concentration-related increase over at least 2 or 3 concentrations. Because the spontaneous mutant frequency can vary from 0 to 20 mutants per $10^6$ cells, evaluation of the response in terms of a fold-increase in mutant frequency relative to the background mutant frequency is not reliable. Some laboratories use a minimum mutant frequency for a response to be considered positive, for example, greater than 40 or 50 mutants per $10^6$ clonable

cells [65]. A statistical method may also be employed, although no single method has been agreed upon [53].

***AS52/*xprt *Assay*** The procedure for conducting the AS52/*xprt* assay is nearly identical to that described above for the CHO/*Hprt* assay. Both assays are based on the same parent cell line: AS52 cells are derived from CHO-K1-$BH_4$ cells [35], and hence CHO cells and AS52 cells have similar growth and cloning characteristics. In addition, the expression and selection of *xprt* mutant AS52 cells and *Hprt* mutant CHO cells can be conducted using the same protocol [55].

A major difference between the two assays is that AS52 cells should be grown in MPA medium (nutrient mixture F12 supplemented with 5% FBS, 250 μg/mL xanthine, 25 μg/mL adenine, 50 μM thymidine, 3 μM aminopterin, and 10 μM mycophenolic acid) right up to the time of treating cells in the assay. MPA medium not only cleanses the cultures of *xprt* mutants (minimizing the background mutant frequency), but also ensures that cells that have lost their transgene (and will be selected as mutants) are not included in the assay. Assays conducted with MPA cleansing have spontaneous mutant frequencies only somewhat higher than that seen in the CHO/*Hprt* assay [37, 41, 55, 66], but culturing cells in the absence of MPA cleansing can result in background mutant frequencies of several hundred per $10^6$ cells [67].

## REFERENCES


1. Moore MM, DeMarini DM, DeSerres FJ, Tindall KR, Eds. *Banbury Report 28: Mammalian Cell Mutagenesis*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory; 1987.
2. Chu EHY, Malling HV. Mammalian cell genetics. II. Chemical induction of specific locus mutations in Chinese hamster cells *in vitro*. *Proc Natl Acad Sci USA* 1968;61: 1306–1312.
3. Liber HL, Thilly WG. Mutation assay at the thymidine kinase locus in diploid human lymphoblasts. *Mutat Res* 1982;94:467–485.
4. Moore MM, Harrington-Brock K, Doerr CL, Dearfield KL. Differential mutant quantitation at the mouse lymphoma TK and CHO HGPRT loci. *Mutagenesis* 1989;4:394–403.
5. Aaron CS, Stankowski LF Jr. Comparison of the AS52/XPRT and the CHO/HPRT assays: evaluation of six drug candidates. *Mutat Res* 1989;223:121–128.
6. Aaron CS, Bolcsfoldi G, Glatt HR, Moore M, Nishi Y, Stankowski L, Theiss J, Thompson E. Mammalian cell gene mutation assays working group report. Report of the International Workshop on Standardisation of Genotoxicity Test Procedures. *Mutat Res* 1994;312:235–239.
7. Applegate ML, Moore MM, Broder CB, Burrell A, Juhn G, Kasweck KL, Lin P, Wadhams A, Hozier JC. Molecular dissection of mutations at the heterozygous thymidine kinase locus in mouse lymphoma cells. *Proc Natl Acad Sci USA* 1990;87:51–55.
8. Hozier J, Sawyer J, Moore M, Howard B, Clive D. Cytogenetic analysis of the L5178Y/$TK^{+/-} \rightarrow TK^{-/-}$ mouse lymphoma mutagenesis assay system. *Mutat Res* 1981;84: 169–181.
9. Moore MM, Clive D, Hozier JC, Howard BE, Batson AG, Turner NT, Sawyer J. Analysis of trifluorothymidine-resistant ($TFT^r$) mutants of L5178Y/$TK^{+/-}$ mouse lymphoma cells. *Mutat Res* 1985;151:161–174.

10. Sawyer JR, Moore MM, Hozier JC. High-resolution cytogenetic characterization of the L5178Y $TK^{+/-}$ mouse lymphoma cell line. *Mutat Res* 1989;214:181–193.
11. Sawyer J, Moore MM, Clive D, Hozier J. Cytogenetic characterization of the L5178Y $TK^{+/-}$ 3.7.2C mouse lymphoma cell line. *Mutat Res* 1985;147:243–253.
12. ICH Guidance for Industry S2B: Genotoxicity: A standard battery for genotoxicity testing of pharmacuticals (July 1997). *Federal Register, 62 FR 62472* (http://www.fda.gov/cder/guidance/index.htm).
13. ICH Guideline for Industry S2A: Specific aspects of regulatory genotoxicity tests for pharmaceuticals (April 1996). *Federal Register, 61 FR 18199* (http://www.fda.gov/cder/guidance/index.htm).
14. Dearfield KL, Auletta AE, Cimino MC, Moore MM. Considerations in the U.S. Environmental Protection Agency's testing approach for mutagenicity. *Mutat Res* 1991;258:259–283.
15. Müller L, Kikuchi Y, Probst G, Schechtman L, Shimada H, Sofuni T, Tweats D. ICH-harmonised guidances on genotoxicity testing of pharmaceuticals: evolution, reasoning and impact. *Mutat Res* 1999;436:195–225.
16. Moore MM (Chair), Honma M (Co-Chair), Clements J (Rapporteur), Awogi T, Bolcsfoldi G, Cole J, Gollapudi B, Harrington-Brock K, Mitchell A, Muster W, Myhr B, O'Donovan M, Ouldelhkim M-C, San R, Shimada H, Stankowski LF Jr. The mouse lymphoma thymidine kinase locus (tk) gene mutation assay: International Workshop on Genotoxicity Test Procedures (IWGTP) Workgroup Report. *Environ Mol Mutagen* 2000;35:185–190.
17. ICH M3 Guidance for Industry: Nonclinical safety studies for the conduct of human clinical trials for pharmaceuticals (July 1997). *Federal Register, 62 FR 62922* (http://www.fda.gov/cder/guidance/index.htm).
18. ICH S1B Guidance for Industry: Testing for carcinogenicity of pharmaceuticals (July 1997). *Federal Register, 63 FR 8983* (http://www.fda.gov/cder/guidance/index.htm).
19. ICH Q3A Guidance for Industry: Impurities in new drug substances (February 2003). (http://www.fda.gov/cder/guidance/index.htm).
20. ICH Q3B (R2) Guidance for Industry: Impurities in new drug products (July 2006). (http://www.fda.gov/cder/guidance/index.htm).
21. FDA CDER Guidance for Industry and Review Staff: Recommended approaches to integration of genetic toxicology study results (January 2006). (http://www.fda.gov/cder/guidance/index.htm).
22. OECD Guideline for the testing of chemicals: *In vitro* mammalian cell gene mutation test. OECD Guideline 476, July 1997.
23. EPA Health effects test guidelines OPPTS 870.5300: *In vitro* mammalian cell gene mutation test. EPA Guideline 870.5300, August 1998.
24. Moore MM, Honma M, Clements J, Harrington-Brock K, Awogi T, Bolcsfoldi G, Cifone M, Collard D, Fellows M, Flanders K, Gollapudi B, Jenkinson P, Kirby P, Kirchner S, Kraycer J, McEnaney S, Muster W, Myhr B, O'Donovan M, Oliver J, Ouldelhkim M-C, Pant K, Preston R, Riach C, San R, Shimada H, Stankowski L Jr. Mouse lymphoma thymidine kinase gene mutation assay: follow-up International Workshop on genotoxicity test procedures, New Orleans, Louisiana, April 2000. *Environ Mol Mutagen* 2002; 40:292–299.
25. Moore MM, Honma M, Clements J, Bolcsfoldi G, Cifone M, Delongchamp R, Fellows M, Gollapudi B, Jenkinson P, Kirby P, Kirchner S, Muster W, Myhr B, O'Donovan M, Ouldelhkim M-C, Pant K, Preston R, Riach C, San R, Stankowski L Jr, Thakur A, Wakuri S, Yoshimura I. Mouse lymphoma thymidine kinase locus gene mutation assay: International Workshop (Plymouth, England) on genotoxicity test procedures workgroup report. *Mutat Res* 2003;540:127–140.

26. Moore MM, Honma M, Clements J, Bolcsfoldi G, Burlinson B, Cifone M, Clarke J, Delongchamp R, Durward R, Fellows M, Gollapudi B, Hou S, Jenkinson P, Lloyd M, Majeska B, Myhr B, Omori T, Riach C, San R, Stankowski L Jr, Thakur A, Van Goethem F, Wakuri S, Yoshimura I. Mouse lymphoma thymidine kinase gene mutation assay: follow-up meeting of the International Workshop on genotoxicity tests—Aberdeen, Scotland, 2003—Assay acceptance criteria, positive controls, and data evaluation. *Environ Mol Mutagen* 2006;47:1–5.
27. Moore MM, Honma M, Clements J, Bolcsfoldi G, Burlinson B, Cifone M, Clarke J, Clay P, Doppalapudi R, Fellows M, Gollapudi B, Hou S, Jenkinson P, Muster W, Pant K, Kidd D, Lorge E, Lloyd M, Myhr B, O'Donovan M, Riach C, Stankowski L Jr, Thakur A, Van Goethem F. Mouse lymphoma thymidine kinase gene mutation assay: meeting of the International Workshop on Genotoxicity Testing—San Francisco 2005—Recommendations for 24-h treatment. *Mutat Res*, 2007;627:36–40.
28. FDA CFSAN Redbook 2000: IV.C.1.c. Mouse lymphoma thymidine kinase gene mutation assay. http://www.cfsan.fda.gov/guidance.html.
29. Turner NT, Batson AG, Clive D. Procedures for the L5178Y/TK$^{+/-}$ to TK$^{-/-}$ mouse lymphoma assay. In Kilbey BJ, Legator M, Nichols W, Ramel C, Eds. *Handbook of Mutagenicity Test Procedures*, 2nd eds. Amsterdam: Elsevier; 1984, pp 239–268.
30. Cole J, Arlett CF, Green MHL, Lowe J, Muriel W. A comparison of the agar cloning and microtitration techniques for assaying cell survival and mutation frequency in L5178Y mouse lymphoma cells. *Mutat Res* 1983;111:371–386.
31. Cole J, Muriel WJ, Bridges BA. The mutagenicity of sodium fluoride to L5178Y (wild-type and TK$^{+/-}$ 3.7.2C) mouse lymphoma cells. *Mutagenesis* 1986;1:157–167.
32. Hozier J, Sawyer J, Clive D, Moore M. Cytogenetic distinction between the TK$^{+}$ and TK$^{-}$ chromosomes in the L5178Y TK$^{+/-}$ 3.7.2C mouse-lymphoma cell line. *Mutat Res* 1982;105:451–456.
33. Hozier J, Sawyer J, Clive D, Moore MM. Chromosome 11 aberrations in small colony L5178Y TK$^{-/-}$ mutants early in their clonal history. *Mutat Res* 1985;147:237–242.
34. Moore MM, Doerr CL. Comparison of chromosome aberration frequency and small-colony TK-deficient mutant frequency in L5178Y/TK$^{+/-}$-3.7.2C mouse lymphoma cells. *Mutagenesis* 1990;5:609–614.
35. Tindall KR, Stankowski LF Jr, Machanoff R, Hsie AW. Detection of deletion mutations in pSV2*gpt*-transformed cells. *Mol Cell Biol* 1984;4:1411–1415.
36. Michaelis KC, Helvering LM, Kindig DE, Garriott ML, Richardson KK. Localization of the xanthine guanine phosphoribosyl transferase gene (*gpt*) of *E. coli* in AS52 metaphase cells by fluorescence *in situ* hybridization. *Environ Mol Mutagen* 1994;24:176–180.
37. Stankowski LF Jr, Tindall KR. Characterization of the AS52 cell line for use in mammalian cell mutagenesis studies. In Moore MM, DeMarini DM, DeSerres FJ, Tindall KR, Eds. *Banbury Report 28*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory; 1987, pp 71–79.
38. Hsie AW, Xu ZD, Yu YJ, Sognier MA, Hrelia P. Molecular analysis of reactive oxygen-species-induced mammalian gene mutation. *Teratog Carcinog Mutagen* 1990;10: 115–124.
39. Hsie AW, Xu Z, Yu Y, An J, Meltz ML, Schwartz JL, Hrelia P. Quantitative and molecular analyses of genetic risk: a study with ionizing radiation. *Environ Health Perspect* 1993;101(Suppl 3):213–218.
40. Li AP, Aaron CS, Auletta AE, Dearfield KL, Riddle JC, Slesinski RS, Stankowski LF Jr. An evaluation of the roles of mammalian cell mutation assays in the testing of chemical genotoxicity. *Regul Toxicol Pharmacol* 1991;14:24–40.

41. Oberly TJ, Huffman DM, Scheuring JC, Garriott ML. An evaluation of 6 chromosomal mutagens in the AS52/XPRT mutation assay utilizing suspension culture and soft agar cloning. *Mutat Res* 1993;319:179–187.
42. Bigger CAH, Ponten I, Page JE, Dipple A. Mutational spectra of polycyclic aromatic hydrocarbons. *Mutat Res* 2000;450:75–93.
43. Tennant RW, Margolin BH, Shelby MD, Zeiger E, Haseman JK, Spalding J, Caspary W, Resnick M, Stasiewicz S, Anderson B, Minor R. Prediction of chemical carcinogenicity in rodents from *in vitro* genetic toxicity assays. *Science* 1987;236:933–941.
44. Kirkland D, Aardema M, Henderson L, Müller L. Evaluation of the ability of a battery of three *in vitro* genotoxicity tests to discriminate rodent carcinogens and non-carcinogens I. Sensitivity, specificity and relative predictivity. *Mutat Res* 2005;584:1–256.
45. Kirkland D, Aardema M, Müller L, Hayashi M. Evaluation of the ability of a battery of three *in vitro* genotoxicity tests to discriminate rodent carcinogens and non-carcinogens II. Further analysis of mammalian cell results, relative predictivity and tumour profiles. *Mutat Res* 2006;608:29–42.
46. Matthews EJ, Kruhlak NL, Cimino MC, Benz RD, Contrera JF. An analysis of genetic toxicity, reproductive and developmental toxicity, and carcinogenicity data: I. Identification of carcinogens using surrogate endpoints. *Regul Toxicol Pharmacol* 2006;44:83–96.
47. Matthews EJ, Kruhlak NL, Cimino MC, Benz RD, Contrera JF. An analysis of genetic toxicity, reproductive and developmental toxicity, and carcinogenicity data: II. Identification of genotoxicants, reprotoxicants, and carcinogens using *in silico* methods. *Regul Toxicol Pharmacol* 2006;44:97–110.
48. Mitchell AD, Auletta AE, Clive D, Kirby PE, Moore MM, Myhr BC. The L5178Y/tk$^{+/-}$ mouse lymphoma specific gene and chromosomal mutation assay: a phase III report of the US Environmental Protection Agency Gene-Tox Program. *Mutat Res* 1997;394: 177–303.
49. Clive D, Johnson KO, Spector JFS, Batson AG, Brown MMM. Validation and characterization of the L5178Y TK$^{+/-}$ mouse lymphoma mutagen assay system. *Mutat Res* 1979;59:61–108.
50. Dearfield KL, Moore MM. Use of genetic toxicology information for risk assessment. *Environ Mol Mutagen* 2005;46:236–245.
51. Clive D, Caspary W, Kirby PE, Krehl R, Moore M, Mayo J, Oberly TJ. Guide for performing the mouse lymphoma assay for mammalian cell mutagenicity. *Mutat Res* 1987;189:143–156.
52. Moore-Brown MM, Clive D, Howard BE, Batson AG, Johnson KO. The utilization of trifluorothymidine (TFT) to select for thymidine kinase-deficient (TK$^{-/-}$) mutants from L5178Y/TK$^{+/-}$ mouse lymphoma cells. *Mutat Res* 1981;85:363–378.
53. Hsie AW, Casciano DA, Couch DB, Krahn DF, O'Neill JP, Whitfield BL. The use of Chinese hamster ovary cells to quantify specific locus mutation and to determine mutagenicity of chemicals. A report of the Gene-Tox Program. *Mutat Res* 1981;86:193–214.
54. Li AP, Carver JH, Choy WN, Hsie AW, Gupta RS, Loveday KS, O'Neill JP, Riddle JC, Stankowski LF Jr, Yang LL. A guide for the performance of the Chinese hamster ovary cell/hypoxanthine-guanine phosphoribosyl transferase gene mutation assay. *Mutat Res* 1987;189:135–141.
55. Stankowski LF Jr, Tindall KR, Hsie AW. Quantitative and molecular analyses of ethyl methanesulfonate- and ICR 191-induced mutation in AS52 cells. *Mutat Res* 1986;160:133–147.
56. Chen T, Moore MM. Screening for chemical mutagens using the mouse lymphoma assay. In Yan Z, Caldwell GW, Eds. *Methods in Pharmacology and Toxicology Optimization in Drug Discovery: In Vitro Methods*. Totowa, NJ: Humana Press; 2004, pp 337–352.

57. Sawyer JR, Binz RL, Wang J, Moore MM. Multicolor spectral karyotyping of the L5178Y $TK^{+/-}$-3.7.2C mouse lymphoma cell line. *Environ Mol Mutagen* 2006;47:127–131.
58. Brock KH, Moore MM, Oglesby LA. Development of an intact hepatocyte activation system for routine use with the mouse lymphoma assay. *Environ Mutagen* 1987;9:331–341.
59. Oglesby LA, Harrington-Brock K, Moore MM. Induced hepatocytes as a metabolic activation system for the mouse-lymphoma assay. *Mutat Res* 1989;223:295–302.
60. Moore MM, Clive D. The quantitation of $TK^{-/-}$ and HGPRT-mutants of L5178Y/$TK^{+/-}$ mouse lymphoma cells at varying times post treatment. *Environ Mutagen* 1982;4:499–519.
61. Robison WD, Green MHL, Cole J, Healy MJR, Garner RC, Gatehouse D. Statistical evaluation of bacterial/mammalian fluctuation tests. In Kirkland DJ, Ed. *Statistical Evaluation of Mutagenicity Test Data*. Cambridge, UK: Cambridge University Press; 1989, pp 102–140.
62. Hsie AW, Brimer PA, Mitchell TJ, Gosslee DG. The dose–response relationship for ethyl methanesulfonate-induced mutations at the hypoxanthine–guanine phosphoribosyl transferase locus in Chinese hamster ovary cells. *Somatic Cell Genet* 1975;1:247–261.
63. Machanoff R, O'Neill JP, Hsie AW. Quantitative analysis of cytotoxicity and mutagenicity of benzo[*a*]pyrene in mammalian cells (CHO/HGPRT system). *Chem-Biol Interact* 1981;34:1–10.
64. Li AP, Gupta RS, Heflich RH, Wassom JS. A review and analysis of the Chinese hamster ovary/hypoxanthine guanine phosphoribosyl transferase assay to determine the mutagenicity of chemical agents; a report of Phase III of the U.S. Environmental Protection Agency Gene-Tox Program. *Mutat Res* 1988;196:17–36.
65. Putman DL, San RHC, Bigger CAH, Jacobson-Kram D. Genetic toxicology. In Derelanko MJ, Hollinger MA, Eds. *CRC Handbook of Toxicology*. New York: CRC Press; 1995, pp 337–355.
66. Tindall KR, Stankowski LF Jr, Machanoff R, Hsie AW. Analysis of mutation in PSV2*gpt*-transformed CHO cells. *Mutat Res* 1986;160:121–131.
67. Li AP. Spontaneous frequency of 6-thioguanine resistant mutants in CHO-AS52 cells after prolonged culturing in the absence of selective agents. *Environ Mol Mutagen* 1990;15:214–217.

# *IN VITRO* MAMMALIAN CYTOGENETIC TESTS

R. Julian Preston

*National Health and Environmental Effects Research Laboratory, U.S. Environmental Protection Agency, Research Triangle Park, North Carolina*

**Contents**

1 Purpose of Cytogenetic Tests
2 History of Cytogenetic Testing Protocols
3 Description of Endpoints
  3.1 Chromosomal Structural Alterations
  3.2 Chromosomal Numerical Alterations (Aneuploidy)
  3.3 Sister Chromatid Exchanges (SCEs)
  3.4 Micronuclei
4 Description of Cell Test Systems
  4.1 Established Cell Lines
  4.2 Human Lymphocyte Cultures
5 Requirements for Cytogenetic Assays
6 Description of Assays
  6.1 Cell Growth and Maintenance
  6.2 Metabolic Activation
  6.3 Control Groups
  6.4 Test Chemicals
  6.5 Exposure Concentrations
  6.6 Treatment with Test Substances
  6.7 Culture Harvest Times
  6.8 Chromosome Preparations
  6.9 Analysis of Cytogenetic Alterations
  6.10 Statistical Analysis and Evaluation of Results
  6.11 Test Conclusions
  6.12 Test Report
7 Future Directions
Acknowledgments
References

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

## 1 PURPOSE OF CYTOGENETIC TESTS

The first step in any cancer risk assessment process for exposures to environmental chemicals is *hazard identification*. Thus, in order for new products to be registered for the first time or to be reregistered, an assessment of whether or not they present a potential hazard has to be conducted. This applies equally to products for agricultural, pharmaceutical, or consumer use. In the present context, a hazard is considered to be represented by genetic alterations assessed as gene mutations or chromosomal alterations. In particular, this chapter concentrates on the induction of chromosomal alterations, both numerical and structural ones.

## 2 HISTORY OF CYTOGENETIC TESTING PROTOCOLS

The analysis of chromosomal alterations in cells exposed to environmental agents, particularly ionizing radiation, has a long history, starting in a quantitative fashion with the pioneering work of Karl Sax [1] using pollen grains of *Tradescantia*. These studies provided information on the dose response for X-ray- and neutron-induced chromosome aberrations and the relative effectiveness of acute, chronic, and fractionated exposures. These types of studies were expanded over the next decade or so to include the assessment of the cytogenetic effects produced by chemicals. The types of assay available for cytogenetic analysis were greatly expanded by the ability to grow mammalian cells *in vitro*, either as primary cultures or as transformed cell lines. This capability was further enhanced to include the use of readily available human cells following development of the *in vitro* culture of human lymphocytes by Moorhead et al. [2]. This method relied on the use of phytohemagglutinin to stimulate peripheral lymphocytes to reenter the cell cycle, thereby allowing for metaphase cells to be obtained. At this time, it was necessary to obtain metaphase cells because all cytogenetic assays relied on microscopic evaluation of chromosomes at metaphase. The use of colcemid to inhibit cells at metaphase and of hypotonic solutions to swell the cells for easy analysis of chromosomal alterations also made cytogenetic tests much more reliable, repeatable, and technically straightforward [3].

The standardization of approaches for visualizing chromosomes for assessing structural and numerical alterations following exposure to radiation or chemicals led to the development of cytogenetic assays for testing chemicals for potential hazard to humans. A plethora of different test systems became available, and there was unique merit to many of these. With fairly extensive validation studies, the assays that were considered to be the most straightforward to conduct and provided the more reliable predictions of clastogenicity (chromosome breakage) for carcinogens and noncarcinogens were incorporated into testing guidelines. This history is concisely described in Chapter 1 of the *Evaluation of Short-Term Tests for Carcinogens: Report of the International Collaborative Program* [4]. These *in vitro* cytogenetic assays incorporate chromosome alterations, sister chromatid exchanges (SCEs), or micronuclei as endpoints, and permanent cell lines (especially Chinese hamster) as the test cells. Based on the selection of the highest performing assays as regards carcinogen/noncarcinogen predictions, a number of national and international organizations developed test batteries that could further enhance predictive value. Typi-

cally, these include a bacterial reverse mutation assay, an *in vitro* cytogenetic assay in mammalian cell cultures and/or an *in vitro* gene mutation assay in mammalian cell cultures, and an *in vivo* cytogenetic assay in rodent bone marrow cells.

This standard battery of genotoxicity tests can satisfy the requirements of various global regulatory bodies. For example, several national and international regulatory agencies have presented minimum data requirements for regulatory reviews of commercial chemicals, and these include the use of a standard battery of genotoxicity tests as described above [5]. Although there have been some modifications to these basic assays for improving overall sensitivity and predictivity, the basic principles underlying the assays remain the same, as discussed in Sections 5 and 6.

Recent advances in chromosomal imaging techniques, based on the development of sophisticated fluorescence *in situ* hybridization (FISH) methods, have significantly enhanced the understanding of the mechanisms of chromosome aberration induction and the ability to identify specific aberration types (e.g., reciprocal translocations). These techniques have generally remained in the purview of mechanistic studies and have not been incorporated into cytogenetic tests for assessing carcinogens.

## 3 DESCRIPTION OF ENDPOINTS

### 3.1 Chromosomal Structural Alterations

A broad range of structural alterations can be observed in metaphase cells using either conventional cytogenetic techniques or fluorescence *in situ* hybridization (FISH). The types of structural alterations produced by treatment with chemicals and radiation are the same as those observed in untreated cells. Thus, treatment with a clastogenic agent enhances the frequencies of chromosomal alterations but does not produce a different spectrum of types. A comprehensive description of the classes of structural chromosomal alterations can be found in Savage [6]. In general terms, structural alterations can be subdivided into *unstable* (generally nontransmissible) and *stable* (transmissible to daughter cells at division). Within these two broad categories, an important distinction is made between chromosomal alterations produced prior to DNA replication ($G_1$) and those produced during (S) or after ($G_2$) replication. The former are called *chromosome-type* aberrations because they involve both chromatids of a chromosome and the latter are called *chromatid-type* because they involve one of the two chromatids. An exception is the so-called isochromatid aberration that is produced in S or $G_2$ of the cell cycle but involves both chromatids. Within each of these two classes, chromosomal alterations can be described as *deletions* (loss of chromosomal material) and *exchanges* (between or within chromosomes). Deletions are generally unstable alterations and exchanges can be stable (reciprocal exchanges) or unstable (dicentrics). For the purposes of most mammalian cytogenetic tests that are used for hazard identification, this level of definition of the range of structural alterations is generally sufficient.

### 3.2 Chromosomal Numerical Alterations (Aneuploidy)

There are chemicals that can alter the normal process of cell division either resulting in chromosomes failing to separate at metaphase or preventing one or more

chromosomes attaching to the mitotic spindle. The results are that daughter cells will have either additional chromosomes beyond the normal diploid number (hyperploidy) or fewer chromosomes than the diploid number (hypoploidy). Because of possible artifacts produced during the preparative stages of cells for cytogenetic analysis that can result in chromosome loss, the general approach for assessing aneuploidy is to place greater emphasis on hyperploidy.

### 3.3 Sister Chromatid Exchanges (SCEs)

During the course of DNA replication, errors can occur that lead to the reciprocal exchange between the two sister chromatids of a chromosome. The result is the so-called sister chromatid exchange. While such exchanges had been identified by genetic methods for decades and using radioisotopes in the 1950s, it was not until the 1970s that a simple cytogenetic method was developed for direct observation of SCE [7, 8]. This method relied in general terms on differential staining of the sister chromatids—one being lightly stained with Giemsa and the other darkly stained. Exchanges between the sister chromatids could readily be seen in metaphase cells as light–dark switches. The principle behind the method was the preferential elution of DNA from one chromatid based on incorporation of bromodeoxyuridine (BrdU), binding of a BrdU-sensitive fluorescence dye to asymmetrically BrdU-labeled chromatids, and the subsequent preferential breakage of the DNA containing the dye by fluorescent light and its elution by warm saline sodium citrate. The chromatid with more DNA stains dark with Giemsa and that with less DNA stains lighter [7]. More recently, fluorescent antibodies to BrdU have been used for sister chromatid differentiation [9].

It has been demonstrated that a wide range of chemicals can increase the frequency of SCEs in a broad range of cell types [10]. There is a background level of SCEs that varies with cell type and concentration of BrdU; these SCEs are a consequence of the BrdU incorporation into the DNA. The analysis of SCEs can be a useful component of hazard identification, although it needs to be noted that there is no adverse phenotypic outcome known to be associated with the induction of SCEs. There is, however, a high frequency of SCEs in cells from individuals with Bloom syndrome that is a reflection of error-prone DNA replication of the BrdU-containing template [11]. Thus, SCE increases probably reflect an enhancement of errors in DNA replication, induced, for example, by a clastogen that produces DNA adducts. An advantage of the SCE assay is its high level of sensitivity of detecting chromosomal effects that are produced by S-phase-dependent mechanisms.

### 3.4 Micronuclei

The analysis of structural chromosomal alterations by standard metaphase analysis requires a high level of expertise and is time consuming. For hazard identification purposes, the analysis of micronuclei can be equally informative. Micronuclei are membrane-bounded structures that form either from acentric pieces of DNA (deletions) or whole chromosomes that are not incorporated into daughter nuclei. It is possible to distinguish between acentric fragments and whole chromosomes using an antibody to the kinetochore component of the centromere; a positive stain represents a micronucleus containing a whole chromosome [12]. Thus, the micronucleus

assay can be used for the assessment of structural and numerical chromosomal alterations. While the micronucleus assay does not provide the detail for all types of chromosomal alterations induced by a clastogen that would be obtained by metaphase analysis, it is a robust assay for the identification of a clastogen.

## 4 DESCRIPTION OF CELL TEST SYSTEMS

### 4.1 Established Cell Lines

It is possible to use any permanent cell line that is cycling for cytogenetic tests. However, it is considerably easier for analysis of the data to use cell lines that have a stable karyotype that is close to the diploid number for the particular species. Chinese hamster cell lines (CHO and V79) have frequently been used because of their low near-diploid number of chromosomes ($2n$ is normally 22). Primary human cell lines or virally transformed ones can be used, but they have a relatively low mitotic index that makes analysis more burdensome than with rapidly dividing cells. Thus, any cell line with a stable karyotype and a high mitotic index is most suitable for cytogenetic tests. Human lymphocytes can be virally transformed to produce lymphoblastoid cell lines that generally have a stable karyotype close to the diploid number, but have the advantage of a relatively high mitotic index.

### 4.2 Human Lymphocyte Cultures

Human peripheral blood lymphocytes are nondividing cells but can be stimulated to reenter the cell cycle by use of a mitogen such as phytohemagglutinin. The stimulated cells move somewhat synchronously through the first cell cycle, so that cells can be treated in $G_1$, S, and/or $G_2$ either by selection of the appropriate treatment time after stimulation or by selection of an appropriate time after treatment for cell sampling and fixation. Mitogen stimulated cells have a relatively high mitotic index, especially at the first metaphase, and being primary cultures they have a stable diploid karyotype when sampled from a normal individual. A concern with their routine use is that blood samples have to be obtained from suitable volunteers and this requires some form of informed consent. However, the process for meeting the consent requirement can quite readily be met.

## 5 REQUIREMENTS FOR CYTOGENETIC ASSAYS

While it is possible to use any cycling cell system for cytogenetic analysis, it is highly advantageous to use one that has a stable diploid or near-diploid karyotype. Again, as noted previously, a high mitotic index is also greatly advantageous, if nothing else because this reduces the time to search for analyzable metaphase cells. Primary human lymphocyte cultures are very suitable for meeting both of these needs.

The great majority of chemical clastogens produce structural aberrations by inducing errors of DNA replication (e.g., from an induced DNA adduct) and so treatment with a test chemical has to be for cells in the S phase or that pass through the S phase between treatment and analysis at metaphase [13]. For maximum sensitivity, treatment in S phase is required such that repair of induced DNA damage

prior to replication is minimized. Also, for maximal sensitivity of the assay for structural changes, analysis should be conducted on cells at their first metaphase after treatment. As noted in Section 3, a number of structural alterations are not transmitted to daughter cells, and many are cell lethal events. Thus, analysis at the second or subsequent metaphase after treatment will be much less sensitive than at the first metaphase.

For SCE, aneuploidy, and micronucleus assays, analysis has to be at the second or subsequent metaphase after treatment (SCE and aneuploidy) or during the second cell cycle after treatment (micronucleus) because either cell division produces the event (aneuploidy and micronuclei) or the assay itself relies on two cell cycles for observation of the endpoint (SCE). The maximum frequency for induced micronuclei will be observed during the second cell cycle after the start of treatment, and so a method has been developed to ensure that this requirement will be met. Cytochalasin B inhibits the process of cytokinesis (cell division) and so at division the cell itself will not divide in the presence of cytochalasin B but its nucleus will, and so a binucleate cell will be produced. Thus, micronuclei can be exclusively analyzed in binucleate cells to maximize the sensitivity of the assay [14].

There are other features of assays that more generally fall under the category of personal laboratory preferences rather than requirements. These modifications can be found in individual research papers or study reports [15].

## 6 DESCRIPTION OF ASSAYS

There are, of course, a wide range of protocols available for the conduct of *in vitro* cytogenetic assays. Some of these provide a required format for regulatory purposes (e.g., U.S. Environmental Protection Agency (EPA), U.S. Food and Drug Administration (FDA), and the European Union (EU) regulatory requirements). Others follow best practices and serve as guidelines (e.g., Organization for Economic Co-operation and Development (OECD) and the International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH)). These all address the requirements described in Section 5 but do differ in some of the specific technical details of the tests. These specific details can be found in the individual publications and reports from these organizations. Pertinent information to be used as a starting point for testing guidelines can be found in the review by Muller et al. [16]. Many of the components of *in vitro* assays are the same for the different cytogenetic endpoints whereas some are endpoint specific. In this latter case, the requirements for each assay are presented separately in the following sections.

### 6.1 Cell Growth and Maintenance

Any mammalian cell type that can be grown in tissue culture medium *in vitro* is acceptable for cytogenetic testing. As noted previously, a stable diploid or near-diploid karyotype is advantageous, as is a high mitotic index. Stocks of cells should be stored in liquid nitrogen until needed for a test rather than maintained in culture all the time. At the time of use for a test, cells should be in logarithmic growth phase. Cells should be checked routinely for mycoplasma contamination and not used if

contaminated. Appropriate culture medium recommended for the routine growth of specific cell lines should be used. Culture conditions should be standardized for $CO_2$ concentration, temperature, and humidity. Besides routinely checking the karyotype of the cells to be used, a routine check of cell-cycle time under the standard culture conditions should be made.

### 6.2 Metabolic Activation

A broad range of chemicals require some form of metabolic activation to convert the parent compound into its active metabolites. The majority of cultured cells do not have endogenous metabolic capability and so some form of exogenous metabolic activation needs to be incorporated into the assay. A test is typically conducted with and without metabolic activation to establish whether or not such activation is required for clastogenicity. The most frequently used metabolic activation system is a cofactor-supplemented postmitochondrial fraction (S9) prepared from the livers of rodents treated with metabolism enzyme-inducing agents such as Aroclor 1254 [17]. The choice of concentration of S9 can be dependent on the particular chemical class under test. There have been proposals to use an appropriate species for the formulation of the S9 dependent on the species of the test cells used. For routine testing this has not been a general recommendation. In addition, specific cell lines have been engineered to express specific metabolic activating enzymes, providing an endogenous metabolizing system [18]. This has some potential advantages but knowledge of the metabolism of the test chemical is required so that the appropriate transgenic cell line can be used in a test. There clearly is room for improving the process whereby metabolic activation is incorporated into an assay, but for the present the use of S9 is reasonable.

### 6.3 Control Groups

The inclusion of appropriate control groups is an essential component of any assay system. At a minimum, it is necessary to have a control sample that does not include the test chemical but does include the solvent or vehicle used for the test chemical. These vehicle controls should include samples with and without the metabolic activation system. In addition, a negative control should be included for every cell sampling time used in an assay.

A positive control must be used and should be a known clastogen at treatment concentration levels that are known to give a reproducible and detectable increase over the historic background level for the cytogenetic endpoint being assayed. In this way the positive control demonstrates the sensitivity of the test system at the particular time of the test. If metabolic activation is used, then the positive control should be a clastogen that requires metabolic activation for its effectiveness.

It has been quite common for positive controls to be used at high concentrations that give a very potent positive response. However, this usually means that the slide analyzer will immediately recognize the positive control samples and this can provide a potential bias to the analysis. Positive control concentrations should produce a clear positive effect but not an excessive amount of chromosomal damage.

Several positive controls are used quite frequently and include the following:

| Metabolic Activation Condition | Chemical | CAS Number |
|---|---|---|
| Absence of exogenous metabolic activation | Methyl methanesulfonate | [66-27-3] |
| | Ethyl methanesulfonate | [62-50-0] |
| | Ethylnitrosourea | [759-73-9] |
| | Mitomycin C | [50-07-7] |
| | 4-Nitroquinoline-*N*-oxide | [56-57-5] |
| Presence of exogenous metabolic activation | Benzo[*a*]pyrene | [50-32-8] |
| | Cyclophosphamide | [50-18-01] |

### 6.4 Test Chemicals

Solid test substances need to be dissolved or suspended in appropriate solvents or vehicles. If feasible, they may be prepared in tissue culture medium (without serum). In general, it is recommended that such chemical preparation be performed shortly prior to use in a test. In specific instances where stability of the test chemical in its dissolved state has been established, longer term storage might be used. The treatment solution or suspension should be diluted to the desired concentrations for test use. The vehicle itself should be at an amount in the cell culture that does not affect cell viability or growth rate.

### 6.5 Exposure Concentrations

The chemical should be tested over a range of concentrations that will allow for the assessment of a concentration–response relationship. If feasible, the highest test substance concentration tested with and without metabolic activation should show evidence of cytotoxicity or reduced mitotic activity. In addition, the highest concentration selected should be consistent with solubility issues and should not result in significant changes in pH or osmolality. The concentration range selected is based on preliminary cytotoxicity data and should include concentrations that induce substantial cytotoxicity (about 60%), intermediate cytotoxicity, and low or no cytotoxicity. In those cases where cytotoxicity is not observed in preliminary studies, the highest concentration should be of the order of 0.01 M, 5 mg/mL, or 5 μL/mL. Gases or volatile liquids should be tested in sealed vessels.

### 6.6 Treatment with Test Substances

Duplicate cultures should be used for each dose level, harvest time, or other experimental subdivision. For cell lines and strains, cells in the exponential phase should be treated with the test substance, with and without an exogenous metabolizing system. Human lymphocytes that have been stimulated to reenter the cell cycle with a mitogen can be similarly treated.

For chromosome aberration studies, there are a number of different rationales covering the timing of the start of treatment and the duration of treatment [17]. Given the broad need for cells to be treated at some point of the total treatment in the S phase of the cell cycle, an acute treatment of about 3 h with sampling some 3–6 h afterwards for the great majority of cell lines and 6–12 h for stimulated human lymphocytes will meet this requirement. For lymphocytes, treatment should start about 36 h after mitogenic stimulation, or earlier if longer treatment times are

desired. The only exception would be for high concentrations when extensive cell cycle delay is induced. In this case, a repeat experiment using longer times between the end of treatment and sampling might be necessary. Variations in treatment duration and sampling time are recommended for different protocols. Continuous treatment throughout one cell cycle can also be used, provided a balance is maintained between cytotoxicity and mitotic index to obtain a sufficient number of analyzable metaphase cells.

For SCE or micronucleus studies, longer treatment times are recommended based on the requirement for a cell division (micronuclei) or two DNA replications (SCE) between the start of treatment and observation. For cell lines, treatment can be continuous for about two cell cycles in duration, or it can be more acute (3 h) with a sampling of about two cell cycles from the start of treatment. For the micronucleus assay, cytochalasin B should be added during the first cell cycle of treatment with cells being harvested before the second mitosis. For lymphocytes, treatment can start at about 24 h after mitotic stimulation and last for about two cell cycles until harvesting. Alternatively, an acute treatment of about 12 h can be used starting also at 24 h after mitotic stimulation. For the micronucleus assay, cytochalasin B is typically added at about 44 h after mitotic stimulation with cell harvesting at 72 h after stimulation.

### 6.7 Culture Harvest Times

The harvest times for established cell lines and for human lymphocytes are basically covered in Section 6.6, because these times are related to the start and end of treatment. An additional major requirement is that the majority of cells to be analyzed at metaphase (chromosome aberrations and SCEs) or as binucleate cells (micronuclei) have spent a significant proportion of the total treatment time in the S phase of the cell cycle. To ensure that this occurs and also to ensure sufficient mitotic cells or binucleate cells for analysis, either multiple sampling times or a single harvest time with supporting data for its selection can be used.

A spindle inhibitor, such as colchicine or colcemid, is added to the culture 2–3 h prior to harvest to facilitate accumulation of metaphase cells and to produce contracted metaphase chromosomes that are particularly suitable for analysis purposes.

### 6.8 Chromosome Preparations

There is a fairly wide selection of methods available for obtaining cells suitable for chromosome aberration, SCE, or micronucleus analysis. There are some essential components and some that are individual preferences. For chromosome aberration and SCE analysis, cells need to be swollen with a hypotonic solution and gently fixed (e.g., with glacial acetic acid and alcohol) to avoid cell rupture. Ideal preparations for analysis have well-spread metaphase chromosomes with no overlaps but with minimal probability of chromosome loss as a result of the chromosome spreading technique. For micronuclei, the cytoplasm should be maintained (a hypotonic solution is not generally used) for the most reliable identification of micronuclei and for identification of binucleate cells.

The slides can be stained by a variety of methods with the requirement that clear definition of centromeres and differential chromatid staining are obtained where appropriate. Giemsa staining is generally the standard for metaphase analysis for chromosome aberrations and SCE. More recently, chromosome-specific fluorescent probes have been used for the assessment of reciprocal translocations but this is not generally a part of a standard battery of tests for hazard identification. For micronucleus analysis, fluorescent DNA-specific dyes are preferred, because these are able to facilitate the detection of even very small micronuclei. Anti-kinetochore antibodies or FISH with pancentromeric DNA probes can be used to establish whether not micronuclei contain whole chromosomes or chromosomal acentric fragments [12].

### 6.9 Analysis of Cytogenetic Alterations

A very important component of analysis is that it be conducted without the microscopist having knowledge of the treatment status of the material being analyzed. This can be accomplished by coding all the slides prior to microscopic analysis.

The number of cells to be analyzed for the different cytogenetic endpoints varies somewhat among the different guidelines and protocols: the following is a general guidance. For chromosome aberrations, it is reasonable for subsequent statistical analysis to analyze 100 cells from each of two duplicate cultures for each treatment group. For SCE analysis, 25 cells per duplicate sample per treatment group is reasonable and for micronuclei, 1000 binucleate cells from each of the duplicate cultures for each treatment is appropriate. For all endpoints, an assessment of cytotoxicity using a cell proliferation index or mitotic index is appropriate.

As general guidance, for human lymphocytes, chromosome aberration and SCE analysis should be performed for cells that have a centromere number of 46. For cell lines, it is frequently recommended that cells with ±2 centromeres from the modal number can be analyzed. This perhaps indicates a less than stable karyotype and it might be more appropriate to use a cell line that has a tight modal number of centromeres so that ±1 from this mode would be the recommended range for analysis.

### 6.10 Statistical Analysis and Evaluation of Results

There are many different statistical approaches that have been applied to *in vitro* cytogenetic data and it is not possible to discuss these and their relative merits in this short chapter. The reader is directed to two comprehensive discussions of the issues to be addressed by statistical analysis of these types of data and approaches for the conduct of the appropriate analysis. The reviews are by Margolin et al. [19] and Kim et al. [20].

Similarly, there has been much guidance on the criteria for establishing a positive or a negative result for *in vitro* cytogenetic data. The following is taken from the ICH Guidance presented in Muller et al. [16] since this provides a very informative set of questions to be addressed for evaluating datasets. (Note: Reference to other parts of the manuscript have been deleted from this quote)

2.3.1. *Guidance on the evaluation of in vitro test results*

2.3.1.1. *In vitro positive results*
The scientific literature gives a number of conditions which may lead to a positive in vitro result of questionable relevance. Therefore, any in vitro positive test result should be evaluated for its biological relevance taking into account the following considerations (this list is not exhaustive, but is given as an aid to decision-making):
(i) Is the increase in response over the negative or solvent control background regarded as a meaningful genotoxic effect for the cells?
(ii) Is the response concentration-related?
(iii) For weak/equivocal responses, is the effect reproducible?
(iv) Is the positive result a consequence of an in vitro specific metabolic activation pathway/active metabolite?
(v) Can the effect be attributed to extreme culture conditions that do not occur in in vivo situations, e.g., extremes of pH; osmolality; heavy precipitates especially in cell suspensions?
(vi) For mammalian cells, is the effect only seen at extremely low survival levels?
(viii) Is the positive result attributable to a contaminant (this may be the case if the compound shows no structural alerts or is weakly mutagenic or mutagenic only at very high concentrations)?
(viii) Do the results obtained for a given genotoxic endpoint conform to that for other compounds of the same chemical class?

2.3.1.2. *In vitro negative results*
For in vitro negative results, special attention should be paid to the following considerations (the examples given are not exhaustive, but are given as an aid to decision-making): Does the structure or known metabolism of the compound indicate that standard techniques for in vitro metabolic activation (e.g., rodent liver S9) may be inadequate? Does the structure or known reactivity of the compound indicate that the use of other tests methods/systems may be appropriate?

Based on adherence to a standard protocol and the availability of experience with a particular assay, it would be expected that a test can be described as clearly positive or clearly negative. However, this is certainly not always the case: test results sometimes do not fit the predetermined criteria for a positive or negative decision and therefore have to be described as "equivocal." It is thus assumed that neither biological interpretation nor statistical methods can resolve this situation. In such cases, further testing is usually required to resolve equivocal results. Such tests might involve a different concentration range, treatment times, and/or harvest times depending on available data that might indicate the appropriate experimental design.

## 6.11 Test Conclusions

It is important that the conclusions from *in vitro* cytogenetic assays be stated within the specific confines of the particular assay. For example, *positive results in the in vitro cytogenetic assay indicate that under the test conditions the test substance induces chromosomal aberrations* (or *SCE* or *micronuclei*) *in cultured mammalian somatic cells* (or *human lymphocytes*). Similarly, *negative results indicate that under the test*

*conditions the test substance does not induce chromosomal aberrations* (or *SCE* or *micronuclei*) *in cultured mammalian somatic cells* (or *human lymphocytes*).

### 6.12 Test Report

The OECD Guideline for the Testing of Chemicals Draft Proposal for a New Guideline 487: *In Vitro* Micronucleus Test (found at http://www.oecd.org/dataoecd/60/28/32106288.pdf) provides a comprehensive summary of the information that constitutes a complete report for an *in vitro* cytogenetic test. This guideline is for the micronucleus assay but generally applies to other *in vitro* cytogenetic tests. The information to be included is covered in Sections 6.1–6.11.

## 7 FUTURE DIRECTIONS

The field of cytogenetics has changed quite significantly over the past decade or so with the development of a range of FISH techniques [21]. These can be used to measure specific events such as reciprocal translocations and genomic alterations associated with tumors, as examples. While these can provide significant input for understanding the underlying mechanisms of disease formation, particularly for tumors, and for use in quantitative risk assessments, they have not significantly enhanced the process of hazard identification. Given the added cost of FISH methods, their use in standard *in vitro* cytogenetic tests for hazard identification remains fairly limited.

There is an increasing interest in the role of stem cells in the initiation of disease processes and at the same time an increased availability of long-term adult stem cell *in vitro* cultures [22]. It is generally desirable, where possible, to use the most relevant cell types for *in vitro* cytogenetic assays, and so it is anticipated that human adult stem cell cultures will be incorporated into protocols.

There might also be further development of automated analysis systems for chromosomal alterations, SCEs, and micronuclei. Certainly, flow cytometry has been employed for micronucleus analysis with some success and is gaining in popularity [23]. Automated chromosome analysis systems are less reliable and very much in the development phase. Their improvement to join the arsenal of automated analysis tools would be a significant step in regard to time and cost savings.

## ACKNOWLEDGMENTS

I wish to thank Dr. Andrew Kligerman and Dr. James Allen for their valuable review of this chapter. This chapter has been reviewed by the U.S. Environmental Protection Agency and approved for publication but it does not necessarily reflect Agency policy.

## REFERENCES

1. Sax K. Induction by X-rays of chromosome aberrations in *Tradescantia* microspores. *Genetics* 1938;23:494–516.

2. Moorehead PS, Nowell PC, Mellman WJ, Battips DM, Hungerford DA. Chromosome preparations of leukocytes cultured from human peripheral blood. *Exp Cell Res* 1996;20:613–616.
3. Hsu TC. *Human and Mammalian Cytogenetics: An Historical Perspective*. New York: Springer-Verlag; 1979.
4. deSerres FJ, Ashby, J Eds. *Evaluation of Short-Term Tests for Carcinogens: Report of the International Collaborative Program. Progress in Mutation Research*, Volume 1. New York: Elsevier/North Holland; 1981.
5. Gollapudi BB, Krishna, G. Practical aspects of mutagenicity testing strategy: an industrial perspective. *Mutat Res* 2000;455:21–28.
6. Savage JR. Classification and relationships of induced chromosomal structural changes. *J Med Genet* 1976;13:103–122.
7. Latt SA. Microfluorometric analysis of deoxyribonucleic acid replication kinetics and sister chromatid exchanges in human chromosomes. *J Histochem Cytochem* 1974;22: 478–491.
8. Perry P, Evans HJ. Cytological detection of mutagen–carcinogen exposure by sister chromatid exchange. *Nature* 1975;258:121–125.
9. Tucker JD, Christensen ML, Strout CL, Carrano AV. Determination of the baseline sister chromatid exchange frequency in human and mouse peripheral blood lymphocytes using monoclonal antibodies and very low doses of bromodeoxyuridine. *Cytogenet Cell Genet* 1986;43:38–42.
10. Tucker JD, Auletta A, Cimino MC, Dearfield KL, Jacobson-Kram D, Tice RR, Carrano AV. Sister-chromatid exchange: second report of the Gene-Tox program. *Mutat Res* 1993;297:101–180.
11. Shiraishi Y, Ohesuki Y. SCE levels in Bloom-syndrome cells at very low bromodeoxyuridine (BrdU) concentrations: monoclonal anti-BrdU antibody. *Mutat Res* 1987; 176:157–164.
12. Schriever-Schwemmer G, Adler ID. Differentiation of micronuclei in mouse bone marrow cells: a comparison between CREST staining and fluorescent *in situ* hybridization with centromeric and telomeric probes. *Mutagenesis* 1994;9:333–340.
13. Preston RJ. Mechanisms of induction of chromosomal alterations and sister chromatid exchanges: presentation of a generalized hypothesis. In Li AP, Heflich RH, Eds. *Genetic Toxicology: A Treatise*. Telford Press: Caldwell, NJ 1991; Chap 3, pp 41–66.
14. Fenech M. The *in vitro* micronucleus techniques. *Mutat Res* 2000;455:81–95.
15. Bonassi S, Fenech M, Lando C, Lin YP, Ceppi M, Chang WP, Holland N, Kirsch-Volders M, Zeiger E, Ban S, Barale R, Bigatti MP, Bolognesi C, Jia C, DiGiorgio M, Ferguson LR, Fucic A, Lima OG, Hrelia P, Krishnaja AP, Lee TK, Migliore L, Mikhalevich L, Mirkova E, Mosesso P, Muller WU, Odagiri Y, Scarffi MR, Szabova E, Vorbtsova I, Vral A, Zijno A. HUman MicroNucleus project: international database comparison for results with the cytokinesis-block micronucleus assay in human lymphocytes: I. Effect of laboratory protocol, scoring criteria, and host factors on the frequency of micronuclei. *Environ Mol Mutagen* 2001;37:31–45.
16. Muller L, Kikuchi Y, Probst G, Schechtman L, Shimada H, Sofuni T, Tweats, D. ICH-Harmonized guidances on genotoxicity testing of pharmaceuticals: evolution, reasoning and impact. *Mutat Res* 1999;436:195–225.
17. Galloway SM, Aardema MJ, Ishidate M Jr, Ivett JL, Kirkland DJ, Morita T, Mosesso P, Sofuni T. Report from working group on *in vitro* tests for chromosomal aberrations. *Mutat Res* 1994;312:241–261.

18. Doherty AT, Ellard S, Perry EM, Parry JM. An investigation into the activation and deactivation of chlorinated hydrocarbons to genotoxins in metabolically competent human cells. *Mutagenesis* 1996;11:247–274.
19. Margolin BH, Resnick MA, Rimpo JY, Archer P, Galloway SM, Bloom AD, Zeiger E. Statistical analyses for *in vitro* cytogenetic assays using Chinese hamster ovary cells. *Environ Mutagen* 1986;8:183–204.
20. Kim BS, Zhao B, Kim HJ, Cho M. The statistical analysis of the *in vitro* chromosome aberration assay using Chinese hamster ovary cells. *Mutat Res* 2000;469:243–252.
21. Speicher MR, Carter NP. The new cytogenetics: blurring the boundaries with molecular biology. *Nat Rev Genet* 2005;6:782–792.
22. Li L, Neaves WB. Normal stem cells and cancer stem cells: the niche matters. *Cancer Res* 2006;66:4553–4557.
23. Dertinger SD, Bishop ME, McNamee JP, Hayashi M, Zuzuki T, Asano N, Nakajima M, Moore M, Torous DK, Macgregor JT. Flow cytometric analysis of micronuclei in peripheral blood reticulocytes: I. Intra- and inter-laboratory comparison with microscopic scoring. *Toxicol Sci* 2006;94:83–91.

# *IN VIVO* GENOTOXICITY ASSAYS

Andreas Hartmann,[1] Krista L. Dobo,[2] and Hans-Jörg Martus[1]
[1]*Novartis Pharma AG, Basel, Switzerland*
[2]*Pfizer Global R&D, Groton, Connecticut*

**Contents**

1 Introduction
2 Standard *In Vivo* Test for Chromosomal Damage in Rodent Hematopoietic Cells
  2.1 *In Vivo* Micronucleus Assessment in Rodent Hematopoietic Tissue
  2.2 *In Vivo* Rodent Bone Marrow Chromosomal Aberration Assay
3 Supplemental *In Vivo* Genotoxicity Assays
  3.1 *In Vivo* Genotoxicity Tests for the Assessment of Primary DNA Lesions
  3.2 Unscheduled DNA Synthesis (UDS) Assay for the Detection of DNA Repair
  3.3 *In Vivo* Gene Mutation Assays
4 *In Vivo* Genotoxicity Tests for Investigating Germline Cells
5 Conclusion
  References

## 1 INTRODUCTION

The preclinical safety assessment of drug candidates includes conduct of assays intended to identify genotoxic hazards. Experience with genetic toxicology testing over the past several decades has demonstrated that no single assay is capable of detecting all genotoxic effects. Therefore, the potential for a chemical to cause genotoxicity is typically determined through a battery of tests conducted *in vitro* and *in vivo*. Genotoxicity tests are typically conducted early in the development of

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

a drug candidate as they are relatively short in duration and inexpensive and provide an early means to identify potential genotoxic carcinogens, which otherwise would not be detected until the completion of 2 year cancer bioassays. Current internationally harmonized guidance [1] recommends the conduct of both *in vitro* and *in vivo* genotoxicity assays, which together detect the potential of a drug candidate to induce small genetic mutations and chromosomal scale damage.

Compared to *in vitro* testing, an *in vivo* test system provides a model that takes into account important factors such as absorption, distribution, metabolism, and excretion that may influence the genotoxic activity of a compound. In addition, an *in vivo* test may detect chemicals that only act *in vivo*, although experience has shown that such compounds are rare. A small number of genotoxic carcinogens, including procarbazine, hydroquinone, urethane, and benzene, are reliably detected using bone marrow tests for chromosomal damage, whereas *in vitro* tests yield negative, weak, or conflicting results [2]. Thus, data from one *in vivo* test is valuable in that it can provide additional reassurance that a drug candidate does not possess genotoxic potential, beyond initial conclusions derived from *in vitro* testing.

Internationally harmonized guidance, ICH M3 [3], outlines which nonclinical safety studies are required to support the various stages of clinical development. Since Phase I clinical trials are usually limited to a few volunteers for a short duration of treatment, it is considered acceptable to limit genotoxicity testing for Phase I studies to the generally more sensitive *in vitro* tests of the standard battery consisting of a bacterial mutagenicity test and a clastogenicity test in cultured mammalian cells. Expansion of clinical investigations into larger populations with longer treatment duration, that is, Phase II and III studies, requires an *in vivo* cytogenetic assessment prior to the initiation of such trials. This *in vivo* assay can either be an analysis of chromosomal aberrations in bone marrow cells or an analysis of micronuclei in bone marrow or peripheral blood erythrocytes. Only under specific circumstances, that is, for compounds that are not systemically absorbed and do not provide for adequate target tissue exposure, may an assessment be based on *in vitro* data alone [1].

It is important to note that there are specific circumstances under which the ICH S2B guidance [1] recommends both *in vivo* and *in vitro* genotoxicity testing be complete prior to initiating clinical trials, including the following: (1) prior to clinical testing in women of childbearing potential, pregnant women, or in children; and (2) for compounds bearing structural alerts for genotoxic activity. Structurally alerting compounds as defined in the literature [4] or with *in silico* systems may have to be further investigated using appropriate tests if the standard tests yield negative results. (3) In the case of positive or equivocal findings *in vitro, in vivo* genotoxicity testing is required prior to clinical Phase I studies for the purpose of risk assessment [5]. A negative result in a standard *in vivo* test for chromosomal damage in rodent hematopoietic cells will not provide sufficient data to conclude that the compound is inactive in somatic cells *in vivo*. Rather, the *in vivo* relevance of positive *in vitro* genotoxicity data needs to be assessed in at least one additional *in vivo* genotoxicity test utilizing a different tissue.

Finally, an alternative approach to assessing genotoxicity prior to the initiation of clinical investigations has been outlined recently in an FDA draft guidance on exploratory investigational new drug (exploratory IND) applications. In the case of clinical investigations in which volunteers will receive a low number of doses of an

investigational drug at relatively low doses, with the intent of collecting limited but critical information to determine feasibility of drug development (e.g., pharmacokinetics), a more limited nonclinical safety assessment is considered acceptable. With regard to genetic toxicology testing, it is acceptable to conduct a standard bacterial mutation assay as well as a test for chromosomal aberrations either *in vitro* or *in vivo*. Furthermore, according to the draft guidance it is acceptable to perform the *in vivo* cytogenetic assessment in conjunction with the repeated dose toxicity study in the rodent [6].

This chapter provides an overview of the standard *in vivo* cytogenetic assays that are commonly used as part of the genotoxicity test battery intended for hazard identification. In addition, exploratory *in vivo* assays are discussed that are utilized to follow up on positive genotoxicity results obtained in the standard battery or to assess a contribution of genotoxicity to preneoplastic/neoplastic changes detected in long-term tests in rodents.

## 2 STANDARD *IN VIVO* TESTS FOR CHROMOSOMAL DAMAGE IN RODENT HEMATOPOIETIC CELLS

The most commonly applied *in vivo* cytogenetic test is the evaluation of micronucleus formation in bone marrow or peripheral blood erythrocytes. Micronuclei are small, round remnants of nuclear chromatin that result from the induction of chromosome breakage or aberrant mitotic division. With the expulsion of the main nucleus during erythrocyte maturation, micronuclei are easily detected in the otherwise anucleate erythrocytic cells (Fig. 1). A less commonly applied, but equally acceptable approach for *in vivo* cytogenetic assessment is metaphase chromosome

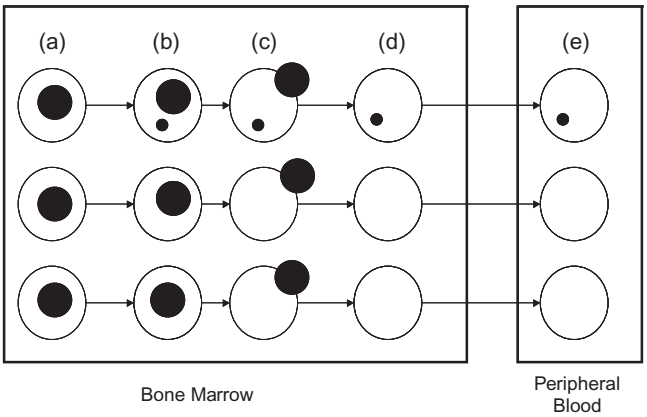


**FIGURE 1** Micronucleated erythrocyte formation. (**a**) Rapidly dividing erythroblast precursor cells in the bone marrow are exposed to test article capable of causing chromosome breakage or abnormal chromosomal segregation. (**b**) In the subsequent interphase erythroblast cells that were subject to chromosome breakage or loss contain micronuclei in addition to the main nucleus. (**c**) Expulsion of the main nucleus during erythrocyte maturation. (**d**) Micronuclei formed in susceptible erythroblast precursor cells remain in otherwise anucleate erythrocytes. (**e**) Migration of normal and micronucleated erythrocytes from the bone marrow to the peripheral blood.

aberration analysis in rodent bone marrow. Using this assay, the frequency and types of structural chromosomal damage observed in bone marrow cells are assessed. In general, metaphase analysis is much more tedious and time consuming and requires a highly skilled cytogenetic experimenter to complete. Furthermore, unlike the *in vivo* micronucleus assay that can detect clastogenicity and aneugenicity, metaphase analysis is limited to assessment of chromosome breakage.

ICH guidelines for genotoxicity testing [1, 6] provide general recommendations regarding the use of *in vivo* cytogenetic assays for the purposes of detecting genotoxic hazards. In addition, OECD guidelines 474 [7] and 475 [8] provide an overview of major methodological considerations for the design, conduct, analysis, and reporting of *in vivo* cytogenetic tests. Subsequent sections of this chapter review key aspects of study design, conduct, interpretation, and reporting of standard *in vivo* cytogenetic assays in the context of use for hazard identification.

### 2.1 *In Vivo* Micronucleus Assessment in Rodent Hematopoietic Tissue

Developed as a rapid alternative to chromosome aberration assessment in rodent bone marrow [9], the rodent micronucleus (MN) assay in erythrocytes is currently the most widely utilized *in vivo* genotoxicity test for assessing induction of cytogenetic damage. The basis of the *in vivo* MN assay is that actively dividing cells exposed to a test article capable of causing chromosome breakage or loss will induce the formation of micronuclei, which are readily identified in progeny cells (i.e., cells in the subsequent division). Immature or newly formed erythrocytes provide an ideal cell population for MN assessment as the erythroblast precursor cells are rapidly dividing in the bone marrow, with the nucleus extruded from the cell a few hours after final mitosis. Providing that there is sufficient time between exposure to the test article and recovery of erythrocytes for analyses, micronuclei that result from induction of chromosomal breakage or malsegregation in erythroblast precursor cells can easily be detected in the newly formed anucleate erythrocyte population. Furthermore, assessment of MN induction in the immature erythrocyte population provides a high assay sensitivity, as these cells were susceptible to chromosome damage during the time of test article treatment.

The MN assay was originally developed in the mouse bone marrow [9] and later it was demonstrated that assessment of MN induction in peripheral blood of mice was equally sensitive [10]. However, given that the rat is typically utilized in preclinical toxicology assessments of acute and subchronic duration, it is more practical to use the rat for *in vivo* cytogenetic safety assessment. Consequently, more recently, greater effort has been applied to investigations and developments using the rat.

Characteristics of the drug candidate that are established in initial toxicology studies, which help facilitate the design of an acceptable *in vivo* micronucleus test, include drug formulation, route of administration, toxicokinetics, and toxicity. Careful consideration of all available data for the drug candidate at the time of protocol design will help maximize the value of the *in vivo* MN study. In the absence of data from prior studies, a dose-range finding study may be conducted to gain this experience.

There is a large body of scientific data contained in the public literature regarding the design and conduct of *in vivo* MN studies. Major methodological considerations of *in vivo* MN study design have been the topic of numerous meetings of the Inter-

national Workshop on Genotoxicity Test Procedures (IWGTP) [11, 12]. In addition, the Collaborative Study Group for the Micronucleus Test (CSGMT) has coordinated investigations directed at understanding the importance of specific elements of study design and methodology [13–19]. These efforts, in addition to contributions from a large number of investigators, have established current practices considered acceptable for the design and conduct of *in vivo* MN assays. Some specific aspects studied in detail include rodent species [18], sex [13, 20], strain [14, 21], treatment route [22], treatment regimen [15, 19, 23, 24], sampling times, sensitivity of blood versus bone marrow [18, 19, 25–27], and development and validation of automated analysis [28–35]. Important methodological aspects of the *in vivo* MN study are reviewed in the following sections. However, more detailed information may be found in the relevant cited references.

***Species, Strain, Sex, and Number of Animals*** In preclinical drug development the most commonly applied species of animals for the conduct of *in vivo* cytogenetic assessment are rat and mouse. In selecting a species and strain of test animal, several considerations should be made. First, the availability of historical control data for the strain and species intended for use is important for the selection of statistical methods and the interpretation of data. Another factor to consider is the availability of drug candidate toxicity data in a rodent species and strain that can be employed for the MN assay, as this can greatly facilitate the design of the study. Furthermore, if the same strain of animal is used for the conduct of long-term toxicology studies or carcinogenicity studies, the MN data may be valuable in assessing the relevance of other findings later in the development of the drug candidate. Given that the rat is the primary rodent species utilized in preclinical toxicology studies, the rat is often the species of choice for the conduct of *in vivo* MN assessment. In the absence of any special requirements for a particular age, young adult animals should be used.

The use of one sex is generally accepted based on studies that have demonstrated that sensitivity to MN induction is similar between male and female rodents [13, 20]. However, if there are data that suggest significant differences in toxicity, toxicokinetics, or metabolism between male and female animals, then both sexes should be utilized. In the case of MN studies that are integrated with toxicology studies, both male and female rodents will be on test by default. In such cases, it is recommended that samples from both sexes be prepared for analysis. However, it is considered acceptable to limit the conduct of MN analyses to a single gender [12]. Each treatment and control group should include at least five analyzable animals. Therefore, if mortality is expected at the high dose, additional animals should be included. Additional animals may also be included for the negative and test treatment groups for plasma drug concentration determination.

***Treatment Regimen and Route of Administration*** The route and schedule of administration of test articles should be determined based on consideration of the bioavailability and pharmacokinetics of the drug candidate being evaluated. It is imperative that the study be designed such that the hematopoietic tissue is exposed to significant concentrations of the drug candidate, otherwise the conduct of the study will add little value to the nonclinical safety assessment. In general, test substance should be administered at least once each day unless the half-life justifies otherwise. Therefore, in the case of drug candidates with short half-lives, multiple

daily treatments may be necessary to maintain sufficiently high blood levels. Doses or formulations that require large volume administration to deliver the test article may also justify the need for multiple daily treatments. Challenges associated with drug candidates having short half-lives or limited oral bioavailability may be overcome by considering administration by continuous infusion or the parenteral route, respectively. Ultimately, the route and schedule of administration should be designed keeping in mind that the hematopoietic tissue is the target of exposure and subsequent evaluation.

***Acute Treatment Protocols and Tissue Sampling*** Studies of acute duration (i.e., 1 or 2 days) are typically designed as independent protocols. The application of acute treatment protocols in the mouse with assessment of MN induction in bone marrow or peripheral blood has been widely utilized, as well as in the rat with MN induction being evaluated in bone marrow. Due to early investigations in the rat demonstrating the ability of the spleen to capture and destroy micronucleated erythrocytes from the peripheral blood [36], assay sensitivity in this tissue has been a concern. Although the efforts of numerous investigators have resulted in the compilation of a significant body of data demonstrating that rat peripheral blood is also sensitive to the detection of aneugens and clastogens [18, 19, 25–27, 37] its use is not widely accepted.

For 1 day treatment protocols, tissue samples (i.e., bone marrow and/or blood) are typically collected and evaluated at two time points. For bone marrow, samples should be collected at ~24 and 48 hours post-treatment, whereas blood samples should be collected at ~48 and 72 hours post-treatment. In the case of 2 day treatment, tissues are typically sampled at a single time point, with bone marrow and/or blood collected at ~24 hours following the final administration.

***Subchronic Treatment Protocol and Tissue Sampling*** The incorporation of MN assessment into subchronic general toxicity studies is an attractive alternative to the acute protocol, as it can reduce animal usage with genotoxicity data generated in parallel with general toxicity and pharmacokinetic data. Treatment protocols that specify up to 4 weeks of treatment are considered acceptable study designs for both mouse and rat [12]. This is based on the outcome of studies in which repeated daily treatments with subchronic duration of exposures were shown to produce similar results (magnitude of response) in comparison to acute treatment in the majority of cases [19, 23, 38–41]. The use of a subchronic versus acute treatment protocol should be given special consideration for drug candidates that induce a positive response in vitro. In such cases an acute protocol should be used if higher systemic exposures can be achieved. As is the case for short duration studies, both bone marrow and blood are acceptable tissues for MN assessment in mouse and bone marrow for MN assessment in rat. For subchronic studies, tissue sampling (bone marrow and/or blood) should take place at one time point, approximately 24 hours following the last treatment.

***Negative Controls*** A concurrent negative control group should be included for each gender and for every sampling time. Negative control animals should be treated with the solvent or vehicle formulation, delivered via the same route of administration as test article, with animals receiving a volume equivalent to the test-article-treated animals. This control is used to help discriminate test-substance effects from

effects of the solvent or vehicle alone (the vehicle should not produce toxicity at the doses used to administer drug). In addition, inclusion of a concurrent untreated control group should also be considered. In the absence of historical or published data demonstrating lack of confounding or mutagenic effects of the solvent or vehicle being used to deliver the test article, untreated control animals should be included in the study. If MN induction is being assessed in peripheral blood, untreated control data can be generated by collecting blood samples immediately prior to the initiation of the study (i.e., prior to the first dose of test article).

***Positive Controls*** The requirement for including a concurrent positive control group is dependent on a number of other aspects of the study design. However, at a minimum, positive control samples must be included as part of microscopic analysis to control for staining and scoring procedures. Omission of a concurrent positive control group may be considered acceptable if (1) verification of test article concentration and stability is conducted according to Good Laboratory Practice (GLP) guidance, and (2) systemic exposure is demonstrated under the experimental conditions used for the study, especially in the case of test articles that are expected to be relatively nontoxic *in vivo* or have demonstrated clastogenicity *in vitro*. If these conditions are not met, a concurrent positive control group should be included. Several examples of positive control substances include cyclophosphamide, mitomycin C, ethyl methanesulfonate, ethyl nitrosourea, and triethylenemelamine. The dose of positive control administered should create an increase in micronucleated cells that is clearly elevated, but not immediately evident to a technician conducting microscopic analysis. The route of administration of the positive control article may differ from that of the test article and a single treatment is acceptable with cells harvested at the appropriate time post-treatment. For example, in a bone marrow MN study conducted with test article treatment for 14 days, positive control animals may receive a single administration of cyclophosphamide on day 14, with cell harvest conducted approximately 24 hours post-treatment, concurrently with the harvest of samples from negative control and test-article-treated groups.

***Dose Selection*** For drug candidates that induce toxicity, the highest dose selected should be the maximum tolerated dose (MTD), that is, the dose that produces signs of toxicity, such that higher doses would result in lethality. Bone marrow toxicity, as evidenced by a reduction in the ratio of polychromatic erythrocytes (PCEs) to normochromatic erythrocytes (NCEs), can also be used to define the upper dose limit and provides an excellent confirmation that the target cells were exposed to the test article. In the absence of test compound toxicity, the highest dose may be limited to the maximum practical dose (MPD), which is defined as 2000 mg/kg/day for studies of acute duration (14 days or less) and 1000 mg/kg/day for studies greater than 14 days duration [11]. Three dose levels should be used and ideally cover a range from clear toxicity to little or no toxicity. Twofold dose spacing [11, 42] and log-type dose spacing [43] have been recommended. However, consideration should be given to all available toxicity data in selection of doses.

***Sample Preparation and Analysis*** Methodology for the preparation of hematopoietic tissue for microscopic analyses is detailed in the scientific literature [44, 45]. In brief, if MN assessment is to be conducted in bone marrow cells, the samples should be collected from the femurs and/or tibia of rodents immediately following

sacrifice. Cells from bone marrow are flushed out using fetal serum. Subsequently, bone marrow samples are concentrated by centrifugation and then spread onto standard microscope slides. A column filtration method may be used to concentrate anucleated cells [46]. Once dry, the slides are fixed for several minutes in absolute methanol. Peripheral blood samples can be obtained from the tail vein or any other blood vessel.

Samples are then stained such that immature or polychromatic erythrocytes (PCEs), which contain RNA, can be differentiated from mature or normochromatic erythrocytes (NCEs), which are devoid of RNA. There are a number of staining methods that differentiate these cell populations while also facilitating the visualization of micronuclei. Blood samples may be stained supravitally by placing a small sample of blood onto a slide that has been previously coated with acridine orange [47]. Alternatively, blood smears may be prepared, air-dried, and fixed in absolute methanol. Fixed samples of bone marrow and blood can be stained using DNA-specific stains including acridine orange (Fig. 2) [48] or Hoechst 33258 and pyronin Y [49]. The non-DNA-specific stain Giemsa may also be utilized; however, it is not recommended for the preparation of bone marrow samples obtained from rat, due to the presence of mast cell granules, which can be mistaken for micronuclei.

All slides including positive and negative controls are coded prior to microscopic analysis. As a measure of toxicity to the hematopoietic tissue, the proportion of immature or polychromatic erythrocytes (PCEs) among total erythrocytes is determined for each animal by scoring 200 erythrocytes in the bone marrow or 1000

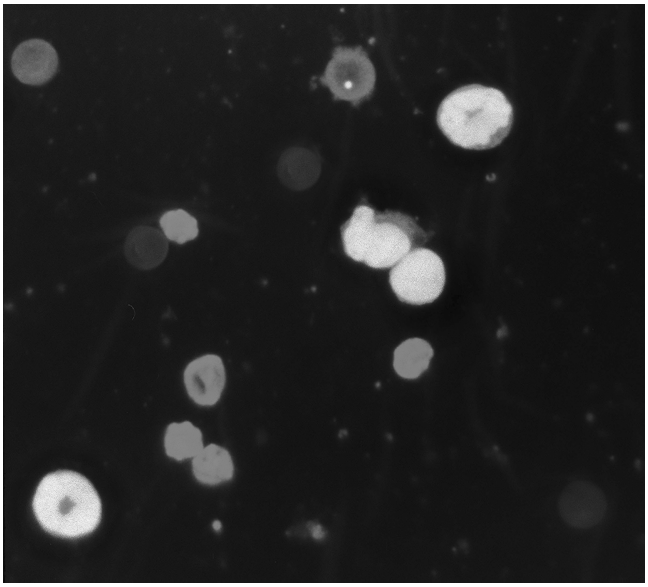

**FIGURE 2** Rat bone marrow smear stained with acridine orange. Immature erythrocytes (PCEs) are anucleate, contain residual RNA, and appear bright red using fluorescent microscopy. Mature erythrocytes (NCEs) contain no RNA and do not stain with acridine orange, yet are readily visible. A micronucleated PCE appears as an anucleate red cell, containing a small, round, yellow nuclear fragment (top center). Nucleated cells with large yellow nuclei are also present in the bone marrow smear.

erythrocytes in peripheral blood. Two thousand PCEs per animal are scored for the frequency of micronucleated cells. In the mouse, supplementary information may also be obtained by scoring micronuclei in the mature erythrocyte population, as these cells have a relatively long lifetime and are not removed by the spleen.

Due to the tedious and time-consuming nature of microscopic analysis, a number of automated methods for MN quantification have been developed, including automated microscope slide reading [28] or flow cytometry based approaches [29–34]. This is a significant advancement, as the high analysis rates of flow cytometry allow tens of thousands of erythrocytes to be analyzed in several minutes. Both OECD and ICH guidances indicate that automated methods of MN assessment are acceptable as long as there is sufficient validation [1, 7]. The most widely investigated and validated flow cytometric method available to date is that developed by Dertinger et al. [31, 32]. This method uses anti-CD71-FITC labeling to differentiate immature (CD71+) from mature (CD71–) erythrocytes. In addition, RNA degradation and propidium iodide staining of DNA allows discrimination of erythrocytes with and without micronuclei. By using a single laser to excite both fluorochromes, the immature and mature erythrocytes with and without micronuclei are then readily differentiated into four discrete populations (Fig. 3). Typically, when flow cytometry is used to quantify MN frequencies, 20,000 immature erythrocytes are evaluated for the frequency of micronucleated cells. Similar to microscopic analysis, the proportion of immature erythrocytes among total erythrocytes is also determined and utilized as a measure of toxicity to the hematopoietic tissue.

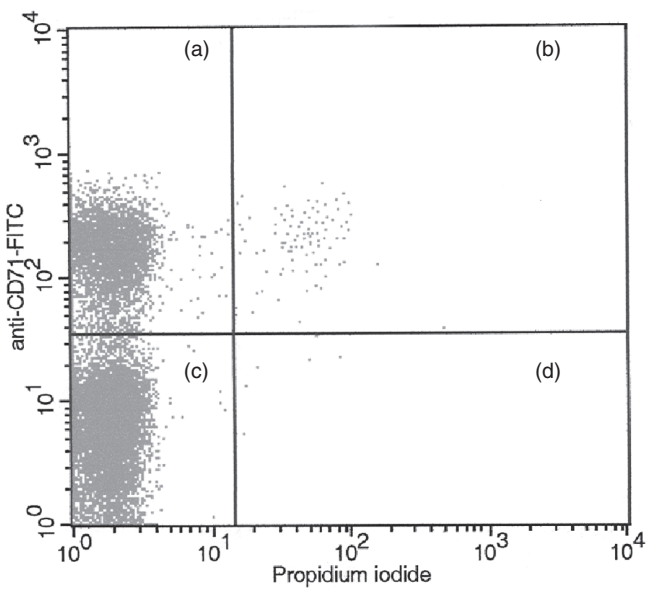


**FIGURE 3** Dot plot from a rat treated with positive control article, demonstrating application of MicroFlow® flow cytometric method for enumeration of micronucleated erthyrocytes. (a) Immature erythrocytes, without micronuclei (CD71+, no PI stain). (b) Immature erythrocytes, with micronuclei (CD71+, containing PI stain). (c) Mature erythrocytes, without micronuclei (CD71–, no PI stain). (d) Mature erythrocytes, with micronuclei (CD71–, containing PI stain).

***Reporting of Data*** Details of study design, study conduct, tissue preparation, sample analysis, and results should be reported in accordance with expectations outlined in OECD guidance 474 [7]. Data obtained from microscopic or cytometric analysis should be presented in tabular form and listed separately for each animal. Data points should include the numbers of immature erythrocytes scored, the numbers of micronucleated immature erythrocytes observed, and the numbers of immature erythrocytes among total erythrocytes. If micronuclei are scored in the mature erythrocyte population, then this data should be presented as well. Calculated values should also be presented, including the proportion of immature erythrocytes among total erythrocytes, the percentage of micronucleated immature erythrocytes, group mean values, and standard errors.

***Data Analysis and Interpretation of Results*** Data should be analyzed using statistical methods [11, 50–53]; however, the biological relevance of statistically significant results should be taken under consideration when interpreting data and making a conclusion. Both trend tests and pairwise comparison tests are appropriate for the statistical analysis of results obtained from *in vivo* MN assessment, with numerous methods considered equally acceptable [50].

The use of negative historical control data in the evaluation and interpretation of study results is also valuable. For example, concurrent negative control data can be compared to the historical negative control data to confirm the acceptability of the study. The historical data can also be used to understand the biological relevance of statistically significant increases in micronuclei. For example, consider study results in which all concurrent negative control and test-article-treated animals generate MN frequencies within the historical control range, yet the trend test indicates statistical significance (Tables 1 and 2). In this case, comparison of the study data to the negative historical control data indicates that the statistically significant result produced by the trend test is not indicative of a biologically relevant response to drug treatment. With regard to biological relevance and interpretation of results, it is also worth noting that a number of confounding factors have been demonstrated to induce increases in micronucleated erythrocytes including hypothermia [54, 55], induction of hematopoiesis [56], and malnutrition [57].

The criteria for a positive, negative, and equivocal response should be clearly defined prior to initiation of the study (i.e., in the study protocol). A positive

**TABLE 1 Representative Historical Control Data Compiled from Bone Marrow Micronucleus Studies Conducted in Male Rats**

| Control Type | Number[a] | % PCE Mean ± SD[b] | Range (%)[c] | % MN-PCE Mean ± SD[d] | Range (%)[e] |
|---|---|---|---|---|---|
| Negative | 412 | 64.8 ± 13.2 | 34.0–94.7 | 0.11 ± 0.09 | 0.00–0.60 |
| Positive | 115 | 54.4 ± 15.6 | 16.4–89.0 | 4.18 ± 2.28 | 0.50–9.30 |

[a]The total number of animals evaluated.
[b]The mean percentage of PCEs among total erythrocytes observed in control animals ± the standard deviation.
[c]The range of percent PCE frequencies observed among control animals.
[d]The mean percentage of micronucleated PCEs observed in control animals ± the standard deviation.
[e]The range of percent micronucleated-PCE frequencies observed among control animals.

**TABLE 2 Representative Example of Summary Study Data from Rat Micronucleus Assay (Male Rat)**

| Compound | Animal Number | % PCE[a] | % MN PCE[b] |
|---|---|---|---|
| Vehicle control | 1 | 65.8 | 0.00 |
| | 2 | 60.5 | 0.05 |
| | 3 | 67.9 | 0.05 |
| | 4 | 60.0 | 0.10 |
| | 5 | 62.6 | 0.10 |
| | Mean | *63.4* | *0.06* |
| | SD ± | *3.4* | *0.04* |
| Positive control | 7 | 39.3 | 3.95 |
| | 8 | 51.4 | 3.45 |
| | 9 | 46.9 | 3.15 |
| | 10 | 42.0 | 3.90 |
| | 11 | 58.6 | 4.40 |
| | Mean | *47.6* | *3.77* |
| | SD ± | *7.7* | *0.48* |
| Test article, 500 mg/kg | 19 | 66.0 | 0.00 |
| | 20 | 54.6 | 0.00 |
| | 21 | 51.8 | 0.05 |
| | 22 | 61.4 | 0.15 |
| | 23 | 42.2 | 0.00 |
| | Mean | *55.2* | *0.04* |
| | SD ± | *9.2* | *0.07* |
| Test article, 1000 mg/kg | 25 | 56.2 | 0.10 |
| | 26 | 59.8 | 0.05 |
| | 27 | 58.6 | 0.10 |
| | 28 | 61.1 | 0.05 |
| | 29 | 52.3 | 0.15 |
| | Mean | *57.6* | *0.09* |
| | SD ± | *3.5* | *0.04* |
| Test article, 2000 mg/kg | 31 | 40.5 | 0.25 |
| | 32 | 36.0 | 0.05 |
| | 33 | 60.8 | 0.20 |
| | 34 | 54.9 | 0.05 |
| | 35 | 42.9 | 0.15 |
| | Mean | *47.0* | *0.14* |
| | SD ± | *10.4* | *0.09* |

[a]Percentage of PCEs among total erythrocytes.
[b]Percentage of micronucleated PCEs.

response may be defined by a number of criteria. For example, test results may be considered positive when statistical significance is observed from trend test analysis, with concurrent observation of MN frequencies outside the historical control range for at least one treatment group. Alternatively, a clear increase in micronucleated cells in a single treatment group at a single sampling time may also define a positive response. The observation of a positive response in the *in vivo* MN assay indicates that the drug candidate has the potential to induce clastogenic or aneugenic effects in the hematopoietic tissue. Negative results indicate that the drug candidate does not induce such effects under the test conditions utilized. When the results of an

*in vivo* MN assay are equivocal, conduct of a subsequent study to clarify the finding should be considered, preferably using a modification to the original experimental design (e.g., refined dose selection or treatment regimen).

### 2.2 *In Vivo* Rodent Bone Marrow Chromosomal Aberration Assay

An equally accepted [1, 5] but much less utilized method for identifying the potential of a drug candidate to induce clastogenicity *in vivo* is the evaluation of chromosomal aberrations in rodent bone marrow metaphase cells. Bone marrow is the target tissue for this assay as the cells are rapidly dividing, thereby making the tissue susceptible to the genotoxic effects of chemical treatment. In contrast to the quantification of micronuclei in erythrocytes, the assessment of chromosomal damage via aberration analysis in metaphase cells is tedious and requires a highly skilled experimenter to complete. In addition, chromosome analysis does not provide ready analysis of aneugenic mechanisms, as in the case of the MNT recommendations regarding major methodological aspects of study design, conduct, analysis, and reporting published in OECD guidance 475 [8]. Additional detail regarding minimal requirements for the design and conduct of scientifically valid and practical studies has also been published [58]. The previous discussions regarding optimal study design for *in vivo* MN assays are largely applicable to the design of bone marrow aberration assays. Therefore, the following overview focuses on those aspects of the aberration assay study design, conduct, analysis, and reporting that differ from the MN assay or are unique to the chromosomal aberration assay.

***Overview of the Study Design of the* In Vivo *Chromosome Aberration Assay*** Mouse and rat are the most commonly utilized rodent species for the conduct of the bone marrow chromosome aberration assay, based on their widespread use in general toxicity studies. Similar to the micronucleus (MN) assay, young adult animals should be utilized with a minimum of 5 analyzable animals per treatment group. A single sex may be tested unless data demonstrate differences in toxicity or pharmacokinetics between male and female animals. Concurrent negative vehicle control groups should be included for each sex and sampling time. A concurrent positive control group should be included for each sex as well, but a single sampling time is acceptable. The positive control substances suggested for use in the *in vivo* MN assay are also applicable for the aberration assay.

***Acute and Subchronic Treatment Protocols and Sampling Times*** Important considerations for selecting route of administration and treatment regimen, such that exposure to the hematopoietic tissue is optimized, were previously discussed for *in vivo* MN study design and are applicable to the design of a bone marrow chromosome aberration assay. Both acute and subchronic treatment protocols can be utilized. Acute treatment protocols of 1 day should be followed by two sample collections. Sample collection is scheduled such that cells are in the first metaphase after exposure to drug. Therefore, the first sample collection should occur between 12 and 18 hours following treatment, which corresponds to approximately 1.5 cell cycles. The second collection should occur 24 hours following the first. The second sampling collection is intended to account for treatment-related cell cycle delay or slow rates of drug uptake and metabolism. In order to enrich the population of

metaphase cells in the bone marrow prior to harvest, animals are administered a metaphase arresting compound (e.g., colchicine) intraperitoneally about 3–5 hours prior to sacrifice. For studies in which drug is administered for multiple days, only one sample collection is necessary and should occur at 12–18 hours following the final treatment.

***Dose Selection*** The recommendations for dose selection are similar to those of the *in vivo* MN assay. The highest dose tested should be the maximum tolerated dose (MTD), or one that induces bone marrow toxicity, such that there is a significant reduction of the mitotic index in test-article-treated animals in comparison to the concurrent negative controls. Alternatively, for nontoxic drug candidates, the maximum practical dose (MPD) can be used to define the upper dose limit (2000 mg/kg/day for 1–14 days; 1000 mg/kg/day for greater than 14 days). Three test-article-treated groups should be included with twofold or log-type dose spacing.

***Sample Preparation and Analysis*** Bone marrow is collected from the femur and/or tibia immediately following sacrifice. Samples are prepared by exposure to a hypotonic solution and then fixative. Fixed samples are dropped onto slides and stained. Slides from all treatment groups are blind coded prior to microscopic evaluation. Sample analysis includes determination of the mitotic index by scoring 1000 cells/animal for the frequency of metaphase cells, and quantification of chromosome aberrations by scoring a minimum of 100 metaphase cells/animal. In selecting cells for chromosome aberration analysis, it is considered acceptable to score metaphase cells with a total centromer count of $2n \pm 2$, where $n$ is equal to the haploid number of chromosomes for the species being tested. During analysis, the number and types of aberrations observed in each cell should be recorded. The minimal classes of aberrations that should be differentiated when collecting data include chromatid-type and chromosome-type gaps, breaks, and rearrangements. It is also recommended that a polyploidy index be determined for all bone marrow samples by assessing 200 metaphase cells for the frequency of polyploidy and endoreduplicated cells [58]. Increases in polyploid or endoreduplicated cells in treated animals may be indicative of aneugenic potential.

***Reporting of Data*** Details of study design, study conduct, tissue preparation, sample analysis, and results should be reported in accordance with expectations outlined in OECD guidance 475 [8]. Data should be reported in tabular form for each animal including the number of cells scored, the number of aberrations per cell, and the percentage of cells with structural aberrations. Structural chromosome aberrations should be listed according to their classification with the number and frequency for treated and control groups. Gaps should also be reported in tabular form but not included in the tally of aberrant cells.

***Data Analysis and Interpretation of Results*** The use of statistical analysis in the evaluation and interpretation of chromosome aberration assay data is recommended, along with consideration of historical control data and biological relevance [58]. As with the *in vivo* MN assay, commonly utilized and acceptable methods of statistical analysis include the application of trend and/or pairwise comparison tests. For example, data can be analyzed to assess the occurrence of a statistically

significant dose–response relationship by conducting a trend test such as the Cochran–Armitage trend test. Alternatively, or in addition, pairwise comparisons may also be conducted (e.g., Fisher's exact test) to assess significant increases at each individual dose group. Historical control data is also valuable for evaluation and interpretation of data in that it can be used to assess the acceptability of the response observed in the concurrent control, and also for considering the biological relevance of responses that are flagged as statistically significant.

Several criteria can be used for concluding a result is positive; a reasonable example is observation of a dose-related increase in aberrant cells with a concurrent statistically significant increase in at least one dose group. The observation of equivocal results should be clarified by conducting further testing, preferably using a modification of the experimental conditions used in the first study. Positive results in the chromosomal aberration assay indicate that the drug candidate induces clastogenic effects under the test conditions, whereas negative results indicate that the drug candidate is not clastogenic. Finally, increases in polyploidy may be indicative of aneugenic potential; however, polyploidy can also be the result of cytotoxicity.

## 3 SUPPLEMENTAL *IN VIVO* GENOTOXICITY ASSAYS

Supplemental *in vivo* genotoxicity studies are used (1) to follow up on positive findings in one or more tests of the standard genotoxicity battery; (2) to elucidate a potential contribution of genotoxicity to the induction of preneoplastic and/or neoplastic changes detected in long-term tests in rodents; and (3) to elucidate mechanisms of micronucleus formation to differentiate clastogenic from aneugenic effects. Since aneugenicity is well accepted to result from mechanisms of action for which thresholds exist, demonstration that micronucleus formation is a result of chromosome loss should allow an acceptable level of human exposure to be defined [59]. No matter the trigger for conducting supplemental *in vivo* genotoxicity testing, it is critical that the approach utilized—for example, the endpoint and target tissue assessed—is scientifically valuable, such that the results will aid in interpreting the relevance of the initial finding of concern. Ultimately, the goal of supplemental genotoxicity testing is to determine if a genotoxic risk is posed to patients under intended conditions of treatment.

*Follow-up Testing of Drug Candidates Positive in the Standard Genotoxicity Test Battery* It has been reported that approximately 30% of pharmaceuticals produce positive genotoxicity results *in vitro* [60]. In contrast, results from bone marrow cytogenetic assays are frequently negative, even for those compounds that produce positive results *in vitro*. This discrepancy may result from a number of major differences that exist when testing in cultured cells versus intact animals. For example, differing metabolic pathways can exist *in vitro* and *in vivo*, metabolic inactivation can occur in the intact animal, parent compound or active metabolite may not reach the target cell *in vivo*, rapid detoxification and elimination may occur, or plasma levels *in vivo* may not be comparable to concentrations that generate positive responses in the *in vitro* assay, which is often accompanied by high levels of cytotoxicity. It is also worth noting that positive results generated *in vitro* may be secondary to effects, such as cytotoxicity, which may never be achieved under *in vivo*

exposure conditions. Data from *in vivo* experiments are therefore essential before definitive conclusions are drawn regarding the potential mutagenic hazard to humans from chemicals that produce positive results in one or more *in vitro* tests.

*Follow-up Testing of Tumorigenic Drug Candidates Negative in the Standard Genotoxicity Test Battery* Carcinogenicity testing of pharmaceutical drug candidates negative in the standard *in vitro* and *in vivo* genotoxicity assays may yield evidence of a tumorigenic response in rodents. The ICH guidance S2B [1] stipulates that such compounds shall be investigated further in supplemental genotoxicity tests, if rodent tumorigenicity is not clearly based on a nongenotoxic mechanism. Typically, supplemental *in vivo* genotoxicity tests should be performed with cells of the respective tumor target organ to distinguish between genotoxic and nongenotoxic mechanisms of tumor induction.

*Endpoints Assessed in Supplemental Assays* In section 3.1, commonly applied test systems are described that are used as supplemental genotoxicity assays. These assays differ with respect to the endpoints assessed:

1. Induction of primary DNA lesions, that is, measurement of exposure, uptake, and reactivity to DNA via the comet assay or $^{32}$P-postlabeling assay.
2. Measurement of the repair of DNA lesions using the unscheduled DNA synthesis (UDS) test.
3. Measurement of induction of genetic changes using transgenic animal assays for point mutations or the mouse spot test.

### 3.1 *In Vivo* Genotoxicity Tests for the Assessment of Primary DNA Lesions

Primary DNA lesions are detected with so-called indicator tests. These tests do not directly measure consequences of DNA interaction (i.e., mutation) but do detect effects related to the process of mutagenesis, such as DNA damage, recombination, and repair. Results from indicator tests can provide additional useful information in the context of extended genotoxicity testing. However, primary DNA lesions may be repaired error-free and do not necessarily result in formation of mutations. The most commonly utilized assays in pharmaceutical development are the $^{32}$P-postlabeling assay and the comet assay. Further methods such as the alkaline elution or unwinding assays will not be described here. However, comprehensive data for these tests and descriptions of the assays have been published [61, 62]. A comparison of different aspects of the methods described in the text is depicted in Table 3. Basic aspects regarding optimal study design for *in vivo* micronucleus assays are largely applicable to the design of supplemental *in vivo* assays. Specific or unique aspects on study protocols are described more extensively where appropriate.

***Comet Assay for the Detection of DNA Damage*** The *in vivo* comet assay (single-cell gel electrophoresis) is increasingly being used as a supplemental genotoxicity test for drug candidates [63, 64]. There are general review articles on the comet assay [65, 66] and a general guideline for test conductance has been published as a result of the International Workshop on Genotoxicity Test Procedures (IWGTP) [67].

**TABLE 3 Overview on Key Aspects of Exploratory In Vivo Genotoxicity Assays**

| Aspect | Comet Assay | DNA Adducts | UDS Test (Liver) | Transgenic Gene Mutation |
|---|---|---|---|---|
| Test definition (accepted protocol) | Yes | No | Yes | Yes |
| Regulatory acceptance/use | Yes | Yes | Yes | Yes |
| Relevance of endpoint | Moderate | Moderate | Moderate | High |
| Technical demands | Low–moderate | Moderate–high | Moderate | High |
| Widespread use | Yes | No | Yes | No |
| Applicable to most tissues? | Yes | Yes | No | Yes |
| Dependence of cell turnover | No | No | No | Yes |
| Cost | Low–moderate | Moderate–high | Low | High |

More specific recommendations with the goal of gaining more formal regulatory acceptance of the comet assay were published following the 4th International Comet Assay Workshop [68]. An updated position paper on specific aspects of test conditions and data interpretation was prepared following the IWGT in 2005 [69].

*Principle of the Method* The basic principle of the comet assay is the migration of DNA in an agarose matrix under electrophoretic conditions (Fig. 4). When viewed through the microscope, a cell has the appearance of a comet, with a head (the nuclear region) and a tail containing DNA fragments or strands migrating in the direction of the anode (Fig. 5). Among the various versions of the comet assay, the alkaline (pH of the unwinding and electrophoresis buffer ≥13) method enables detection of the broadest spectrum of DNA damage and is therefore recommended (in the first instance) for regulatory purposes [67, 68]. The alkaline version detects DNA damage such as strand breaks, alkali-labile sites (ALS), and single strand breaks associated with incomplete excision repair. Under certain conditions, the assay can also detect DNA–DNA and DNA–protein crosslinking, which (in the absence of other kinds of DNA lesions) appears as a relative decrease in DNA migration compared to concurrent controls. In contrast to other DNA alterations, crosslinks may stabilize chromosomal DNA and inhibit DNA migration [70, 71]. Thus, reduced DNA migration in comparison to the negative control (which should show some degree of DNA migration) may indicate the induction of crosslinks, which are relevant lesions with regard to mutagenesis and should be further investigated. Increased DNA migration indicates the induction of DNA strand breaks and/or ALS. Furthermore, enhanced activity of excision repair may result in increased DNA migration. DNA excision repair can influence comet assay effects in a complex way [72, 73]. While DNA repair generally reduces DNA migration by eliminating DNA lesions, ongoing excision repair may increase DNA migration due to incision-related DNA strand breaks. Thus, the contribution of excision repair to the DNA

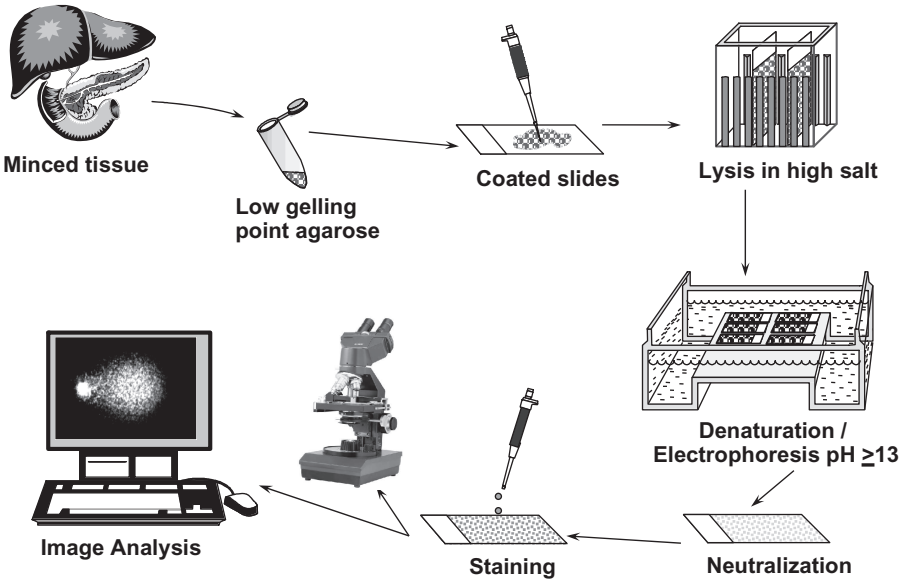


**FIGURE 4** Flow diagram for performing a comet assay.

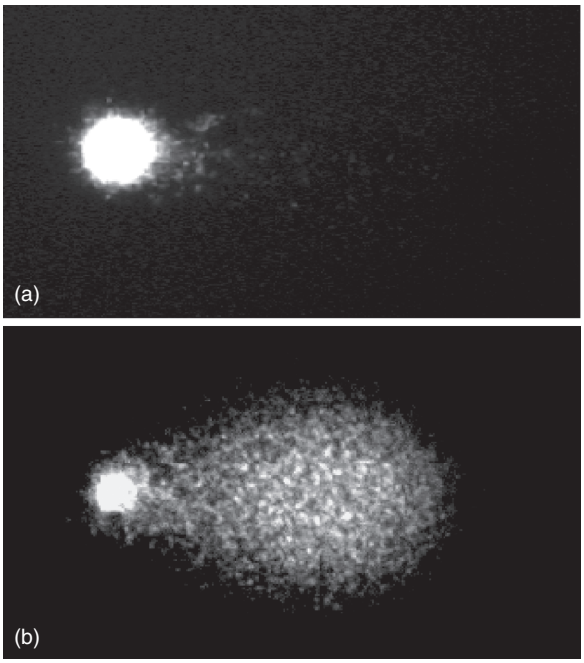


**FIGURE 5** Photomicrographs of rat hepatocytes after processing in the comet assay: (a) hepatocyte of vehicle control animal and (b) hepatocyte of rat dosed once orally with 40 mg/kg methyl methanesulfonate, exhibiting increased DNA migration.

effects seen in the comet assay depends on the types of induced primary DNA damage and the time point of analysis [73].

*Test Procedure* Aspects regarding test animals, test substance, use of concurrent negative and positive control animals, as well as dose selection for the design of a cytogenetic assay, as described in detail previously, are largely applicable to the design of an *in vivo* comet assay. In addition, more specific details can be found in an earlier publication [68]. A single treatment or repeated treatments (generally at 24 hour intervals) are equally acceptable. In both experimental designs, the study is acceptable as long as a positive effect has been demonstrated or, for a negative result, as long as an appropriate level of animal or tissue toxicity has been demonstrated or the limit dose with appropriate tissue exposure has been used. For repeated treatment schedules, dosing must be continued until the day of sampling. On a daily basis, test substances may be administered as a split dose (i.e., two treatments separated by no more than a few hours), to facilitate administering a large volume of material. The test may be performed in two ways. If animals are treated with the test substance once, then tissue/organ samples are obtained at 2–6 hours and 16–26 hours after dosing. The shorter sampling time is considered sufficient to detect rapidly absorbed as well as unstable or direct acting compounds. In contrast, the late sampling time is intended to detect compounds that are more slowly absorbed, distributed, and metabolized. When a positive response is identified at one sampling time, data from the other sample time need not be collected. Alternatively, if multiple treatments at 24 hour intervals are used, tissue/organ samples need be collected only once. The sampling time should be 2–6 hours after the last administration of the test substance. Alternative sampling times may be used when justified on the basis of toxicokinetic data.

*Selection of Tissues and Cell Preparation* In principle, any tissue of the experimental animal, provided that a high quality single cell/nucleus suspension can be obtained, can be used for a comet assay. Selection of the tissue(s) to be evaluated should be based, wherever possible, on data from absorption, distribution, metabolism, excretion studies, and/or other toxicological information. A tissue should not be evaluated unless there is evidence of, or support for, exposure of the tissue to the test substance and/or its metabolite(s). In the absence of such information and unless scientifically justified, two tissues should be examined. Recommended tissues are liver, which is the major organ for the metabolism of absorbed compounds, and a site of first contact tissue—for example, gastrointestinal for orally administered substances, respiratory tract for substances administered via inhalation, or skin for dermally applied substances. Which tissue is evaluated first is at the discretion of the investigator and both tissues need not be evaluated if a positive response is obtained in the first tissue evaluated.

Single-cell suspensions can be obtained from solid tissue by mincing briefly with a pair of fine scissors [74], incubation with digestive enzymes such as collagenase or trypsin [75], or by pushing the tissue sample through a mesh membrane. In addition, cell nuclei can be obtained by homogenization [76, 77]. During mincing or homogenization, EDTA can be added to the processing solution to chelate calcium/magnesium and prevent endonuclease activation. In addition, radical scavengers (e.g., DMSO) can be added to prevent oxidant-induced DNA damage. Any cell dissocia-

tion method is acceptable as long as it can be demonstrated that the process is not associated with inappropriate background levels of DNA damage.

*Cytotoxicity—A Potential Confounding Factor* A general issue with DNA strand break assays such as the comet assay is that indirect mechanisms related to cytotoxicity may lead to enhanced strand break formation. However, since DNA damage in the comet assay is assessed on the level of individual cells, dead or dying cells may be identified on microscopic slides by their specific image. Necrotic or apoptotic cells can result in comets with small or nonexistent head and large, diffuse tails [78] as observed *in vitro* upon treatment with cytotoxic, nongenotoxic articles [79–81]. However, such microscopic images are not uniquely diagnostic for apoptosis or necrosis since they may also be detected after treatment with high doses of radiation or high concentrations of strong mutagens [82]. For the *in vivo* comet assay, only limited data are available to establish whether cytotoxicity results in increased DNA migration in tissues of experimental animals. Despite necrosis or apoptosis in target organs of rodents such as kidneys [83], testes [84], liver, or duodenum [63], no elevated DNA migration was observed. However, enhanced DNA migration was seen in homogenized liver tissue of mice dosed with carbon tetrachloride [85] when histopathological examination showed evidence of necrosis in the liver. Therefore, to avoid potential false-positive effects resulting from cytotoxicity, recommendations regarding a concurrent assessment of target organ toxicity have been made, including dye viability assays, histopathology, and a neutral diffusion assay [67, 68].

*Biological Significance of Lesions Detected* DNA lesions leading to effects in the comet assay can be strand breaks which may be relevant to the formation of chromosome aberrations or DNA modifications such as abasic sites (AP sites) with relevance to the induction of gene mutations. However, primary lesions detected by the comet assay may also be correctly repaired without resulting in permanent genetic alterations. Neither the magnitude of DNA migration in the comet assay nor the shape of the comet can reveal the types of DNA damage causing the effect or their biological significance, that is, their mutagenic potential. Therefore, conclusions regarding the mutagenicity of a test compound cannot be made solely on the basis of comet assay effects. There are a few limitations of the comet assay with regard to its application and interpretation of test results. For example, short-lived primary DNA lesions such as single strand breaks, which may undergo rapid DNA repair, could be missed when using inadequate sampling times. However, an appropriate study design including one early preparation time point (i.e., at 3–6 h) is considered sufficient to ensure that these lesions are captured—in particular at higher dose levels, where DNA repair may be significantly delayed or even overwhelmed. In any case, it should be kept in mind that a negative comet result can be considered as a strong indicator for the absence of a mutagenic potential.

*Advantages* The advantages of this assay for use in genotoxicity testing of drug candidates include its applicability to various tissues and/or special cell types, its sensitivity for detecting low levels of DNA damage, its requirement for small numbers of cells per sample, the general ease of test performance, the short time needed to complete a study, and its relatively low cost. The comet assay can be

applied to any tissue in the given *in vivo* model, provided that a single cell/nuclei suspension can be obtained. Therefore, the comet assay has potential advantages over other *in vivo* genotoxicity test methods, which are reliably applicable to rapidly proliferating cells only or have been validated preferentially in a single tissue only. The comet assay may detect a broader spectrum of primary DNA lesions, including single strand breaks and oxidative base damage, which may not be detected in the UDS test because they are not repaired by nucleotide excision repair [86]. The advantages of the comet assay over the alkaline elution test include the detection of DNA damage on a single-cell level and the requirement for only small numbers of cells per sample. In contrast, when using the alkaline elution assay, large quantities of cells are necessary for the determination of genotoxic effects, and, therefore, only a limited number of organs/tissues can be evaluated using this technique. In particular, this seems important for investigation of suspected tissue-specific genotoxic activity, which includes "site-of-contact" genotoxicity (cases of high local versus low systemic exposure).

*Limitations* Experimental variability is an important issue and should be kept to a minimum to ensure reliable interpretation and comparability of the data obtained with other *in vivo* comet experiments. Experimental variability may result from shortcomings with regard to number of doses tested, number of animals per dose, number of slides per animal, number of cells analyzed, lack of sufficient DNA migration in cells of concurrent controls, and deviation from minimum time for treatment of slides with alkaline buffer. Shortcomings regarding technical details apply, for example, to the most comprehensive overview on *in vivo* comet assay test results published so far [87]. This paper provides data on more than 200 compounds tested in rodents, all of which was produced before specific recommendations on assay conduct [67] were available. Therefore, although a valuable resource with very relevant data, the paper by Sasaki et al. [87] has areas where technical aspects of the assay differ from the current minimal requirements recommended [67, 68]. Considering these discrepancies, the data of the comprehensive study [87], as well as other study reports not in agreement with the current recommendations, should be interpreted with caution. This point was highlighted recently in a position paper on the use and status of the *in vivo* comet assay in genotoxicity testing [64], which critically assessed published data produced under test conditions not fully in agreement with the minimal requirements for an acceptable test [67, 68]. For example, it was noted that positive comet assay data were published for compounds that have been assessed before to be neither genotoxic nor carcinogenic, such as food additives [88]. Such isolated positive comet assay results should be critically evaluated in the light of current recommendations [67, 68] to exclude methodological shortcomings and potential artifacts. In cases where negative carcinogenicity data are already available and the *in vivo* comet assay result represents an isolated positive finding in the context of existing genotoxicity data, the biological significance of the effect seen in the comet assay should be assessed with caution.

***Determination of DNA Adduct Formation*** The assessment of DNA adducts in target organs of toxicity is an effective molecular dosimeter of genotoxic exposure and may facilitate differentiation of genotoxic and nongenotoxic agents [89]. The role of DNA adducts in mutagenesis is well documented [90]. However, the relation-

ship between DNA adduct formation, mutagenesis, and subsequent carcinogenesis is complex. For example, comparative investigations have shown that there is not always a clear relationship between DNA adduct levels as an early event and the induction of mutations in marker genes [91, 92]. Thus, some DNA adducts are of greater significance with respect to mutagenesis or carcinogenesis with some adducts being efficiently removed, while other types, or the same adducts at different DNA locations, remain unrepaired. In some cases, the presence of unrepaired adducts may lead to mutations via mispairing during replication. Alternatively, the removal of adducts, particularly if DNA resynthesis is involved, may not be error-free and may lead to mutations. Attempts to elucidate the relationship between specific primary DNA alterations and mutations are limited and influenced by confounding factors [93, 94].

*The $^{32}$P-Postlabeling Assay* The $^{32}$P-postlabeling assay has widely been used to measure covalent DNA–xenobiotic adducts [95–97]. In the $^{32}$P-postlabeling assay, DNA is hydrolyzed enzymatically to 3′-monophosphates and DNA adducts are enriched by the selective removal of normal nucleotides [98]. The DNA adducts are then labeled with [$^{32}$P]phosphate and the resulting $^{32}$P-labeled DNA adducts are usually separated by thin-layer chromatography or high performance liquid chromatography (HPLC). Radioactivity of DNA adducts is detected by autoradiography and liquid scintillation counting, imaging analysis, or a liquid scintillation analyzer.

*Advantages and Limitation* The 32P-postlabeling technique is the most sensitive method for the detection of a wide range of large hydrophobic compounds bound to DNA and can potentially detect one DNA adduct, such as those derived from polycyclic aromatic compounds in $10^9$–$10^{10}$ bases. As a result of an IWGTP conference in 1999, recommendations on the use of the postlabeling assay have been issued [96]. The assay allows the detection of adducts from different chemicals with diverse chemical structures, such as polycyclic aromatic hydrocarbons, aromatic amines, heterocyclic amines, alkenylbenzene derivatives, benzene and its metabolites, styrene, mycotoxins, simple alkylating agents, unsaturated aldehydes from lipid peroxidation, pharmaceuticals, reactive oxygen species, and UV radiation. One unique strength of the assay is that it has been useful for the detection of adducts from complex mixtures [99, 100]. The assay can be applied to measure adducts after multiple dosing in any tissue or to examine the removal of adducts after cessation of exposure. However, one of the major challenges or limitations of using the postlabeling assay for investigating new drug candidates is selection of appropriate methods in the absence of knowing what adducts are or are not formed. The chances of adduct detection may be severely limited as applied methods of chromatography may not be appropriate for the adduct being formed [96]. The assay could give false-negative results due to loss of adducts. In addition, false-negative results are more likely to occur with nonaromatic versus aromatic adducts due to the lower sensitivity of the assay for nonaromatic adducts. Furthermore, polynucleotide kinase can label non-nucleic-acid components, such as some hydroxylated metabolites, leading to false-positive results [101]. Finally, endogenous DNA adducts called I-compounds [102] can interfere with the detection of adducts formed from a test compound exposure. Some I-compounds are present at levels of 1 adduct in $10^7$ or

more DNA nucleotides and migrate on chromatography plates similar to adducts derived from aromatic carcinogens [102].

*Mass Spectrometry* Mass spectrometry methods for measuring DNA adducts have been introduced more recently. Of the physicochemical methods used to detect DNA binding, mass spectrometry has the greatest potential because of its high chemical specificity that allows for unequivocal characterization of the DNA binding products [103]. Thus, genuine DNA adducts derived from the chemical being tested may be distinguished from unrelated adducts and from products of endogenous DNA damage, which is not possible with the postlabeling assay. Detailed structural information on DNA adducts may also be obtained. Accelerator mass spectrometry (AMS) has been used to measure radiocarbon isotope with attomole ($10^{-18}$ mol) sensitivity [104]. Because of the extraordinary sensitivity of AMS, only trace levels of radioactivity are required. However, this technique requires that the isolated DNA be devoid of noncovalently bound radioactivity to assure accurate estimates of DNA adduct levels. Although structural information is not provided by this technique, the use of HPLC in combination with AMS provides a greater degree of confidence in analyte identity. Soft ionization techniques such as matrix-assisted laser desorption ionization (MALDI) and electrospray ionization (ESI) have emerged as techniques to detect nonvolatile and thermally labile compounds [105]. The online coupling of HPLC with tandem mass spectrometry provides structural information of the adducts, and the incorporation of stable, isotopically labeled internal standards into the assay assures precision and accurate quantification of the DNA adducts [106]. Although tandem mass spectrometry is not as sensitive as those of $^{32}$P-postlabeling or AMS, DNA adduct detection limits have been reported to range from 1 adduct per $10^7$ to $10^9$ bases using 100–500 μg of DNA [107].

### 3.2 Unscheduled DNA Synthsis (UDS) Assay for the Detection of DNA Repair

The unscheduled DNA synthesis assay is a widely used method to investigate chemically induced DNA excision repair. The induction of repair mechanisms is presumed to have been preceded by DNA damage. Measuring the extent to which DNA synthesis occurs offers indirect evidence of the DNA damaging ability of a test chemical. The recommended methodology is described in OECD test guideline 486 [108].

*Principle of the Method* The UDS assay measures DNA synthesis induced for the purposes of repairing an excised segment of DNA containing a region damaged by a test chemical. DNA synthesis is measured by detecting tritium-labeled thymidine ($^3$H-TdR) incorporation into DNA, preferably using autoradiography. The liver is generally used for analysis because, under normal circumstances, there is a low proportion of primary hepatocytes in the S phase of the cell cycle. Therefore, an increase in DNA synthesis can readily be attributed to repair of induced DNA damage, rather than DNA synthesis supporting normal cell division. The liver is also the site of first-pass metabolism for chemicals administered orally or by intraperi-

toneal injection. The larger the number of nucleotides excised and repaired, the greater is the amount of detectible $^3$H-TdR incorporated into DNA. For this reason, the UDS assay is more sensitive in detecting DNA damage that is repaired through nucleotide excision repair (removal of up to 100 nucleotides) as compared to base excision repair (removal of 1–3 nucleotides) [108]. Test chemicals more prone to inducing nucleotide excision repair, such as those that form bulky DNA adducts, have a greater potential to cause detectible UDS. However, the UDS assay does not, in itself, indicate if a test chemical is mutagenic because it provides no information regarding the fidelity of DNA repair, and it does not identify DNA lesions repaired by mechanisms other than excision repair. The UDS assay is usually conducted using rats, although other species may be used if justified. Two dose levels are selected on the basis of preliminary toxicity testing, with the highest dose defined as that causing toxicity such that higher levels would be expected to produce lethality. At least three animals per group (typically males only) are administered the test chemical once by gavage. Intraperitoneal injection is not recommended because it could potentially expose the liver directly to the chemical. A group of animals is sacrificed at 2–4 h and another at 12–16 h after treatment. Cultures of hepatocytes are prepared and incubated for 3–8 h in $^3$H-TdR. Slides are prepared and processed for autoradiography using standard techniques. At least 100 cells per animal are examined and both nuclear and cytoplasmic grains should be counted to determine the net nuclear grain count (cytoplasmic grains subtracted from nuclear grains). Chemicals inducing a significant increase in net nuclear grain count for at least one treatment group are considered to have induced UDS [109].

*Advantages and Limitations* Basically, any tissue with a low proportion of cells in the S phase can be used for analysis. However, the *in vivo* UDS test has been validated only in the liver, whereas alternative tissues have been investigated to a limited extent [110]. Because UDS is measured in the whole genome, it is potentially much more sensitive than assays examining only specific loci. However, the extent of UDS gives no indication of the fidelity of the repair process. For that reason, UDS does not provide specific information on the mutagenic potential of a test chemical, but only information suggesting it does or does not induce excision repair. Despite this limitation, a positive response in the *in vivo* liver UDS assay has very high correlation with rodent hepatocarcinogenicity. However, an important limitation of the UDS test is the comparatively low sensitivity for DNA lesions being repaired by base excision repair [108].

### 3.3 *In Vivo* Gene Mutation Assays

***Transgenic Rodents for the Analysis of Somatic Mutations*** Few endogenous genes lend themselves to the analysis of gene mutations *in vivo*. Due to the fact that vertebrates are diploid organisms, only hemizygous genes are amenable to the detection of mutations occurring in one chromosome, which normally is only the case with genes residing on sex chromosomes. Furthermore, to detect mutations, which are rare events, the necessity exists to propagate cells in order to select phenotypically for mutants among a vast excess of unmutated cells. This limits the use of endogenous genes to only a few cell types or developmental stages, such as color coat genes in the case of the mouse spot test, or the *hprt* gene of lymphoid cells.

Finally, endogenous genes generally serve a biological purpose, so that in many cases the mutation of such a gene may result in a selective disadvantage for the affected cells, leading to the risk of elimination of mutations with an obvious probability to underestimation of true mutation frequencies.

The use of shuttle vectors as mammalian mutagenesis test systems was first proposed *in vitro* by Glazer et al. [111], and later by Malling and Burkhart [112], *in vivo*, where they described the use of λsufF and ΦX174 or shuttle vectors, respectively, as mammalian mutation systems. The first assay that came into wider use was that of Gossen et al. [113], who placed the bacterial *lacZ* gene, encoding for β-galactosidase, in a λgt10 vector. This model is now commercially available as Muta™-Mouse (Covance, www.covance.com). Shortly thereafter, similar transgenic mice were introduced by Kohler et al. [114, 115], utilizing the bacterial *lacI* gene embedded in a λLIZ vector, which codes for the repressor of the *lacZ* gene, as the mutational target *in vivo*. This model, which is now commercially available as BigBlue®, can also be obtained on a rat background (Stratagene www.stratagene.com). Those novel lambda vectors share the advantages of being forward mutation systems with hundreds of inactivating target sequences, which, in contrast to the reverse mutation ΦX174 type, allows the detection of many kinds of different mutagenic events.

All these models utilize the so-called shuttle vector principle for mutation analysis, which describes a method in which the transgenes are mutated in the animal body, whereas mutation analysis occurs after specific retrieval of the transgene and subsequent transfer of the mutational target gene into suitable bacterial hosts. Mutation analysis is carried out by extracting high molecular weight genomic DNA from the tissue of interest, packaging the lambda shuttle vector *in vitro* into lambda phage heads, and testing for mutations that arise in the transgene sequences following infection of an appropriate bacterial strain. Packaging extract, which contains the catalytic and structural proteins needed for excising single transgene copies from the genomic DNA and packaging them individually into phage heads, can either be obtained commercially or prepared with standard laboratory methods. A key prerequisite of those systems has been the use of methylation-restriction-deficient host strains such as *E. coli* C, which allow the rescue of the nontranscribed and therefore highly methylated transgenic DNA without any host defense [116].

Later, accumulating evidence indicated that those lacZ or lacI bacteriophage lambda models are insufficiently sensitive for events that involve large genomic rearrangements such as deletions, insertions, or recombinations [117, 118]. Therefore, models were developed to overcome these limitations, such as the pUR288 or the gpt-delta mouse. Those models are described later in more detail.

One key assumption in this approach is that the bacterial target gene behaves similar with respect to mutation induction to endogenous genes, which are causally related to tumor initiation. Several differences between transgenes and native genes can be assumed or have been experimentally demonstrated, which have the potential to affect the relative magnitude of the mutagenic response observed. Thus, prokaryotic transgenes possess attributes that differ from most mammalian genes, which include higher GC content, higher density of the dinucleotide "CpG" and associated 5-methylcytosine, which is a mutation hotspot after spontaneous desamination [119], and existence of the transgenic cluster as a nontranscribed multicopy, head-to-tail concatemer, which is hypermethylated and lacks transcription-coupled repair. While these differences do not preclude the utility of transgenic assays as

general mutation analysis tools, they should be taken into consideration when interpreting assay results. On the other hand, it should be kept in mind that every mutation assay, whether transgenic or not, has its own spectrum of detectable mutations, and, accordingly, some differences are to be expected between any two assays compared. For the transgenic systems, a large body of evidence has been accumulated, which demonstrates that mutation induction in those transgenes strongly resembles the one in endogenous genes [120] and that those models therefore will deliver relevant and valid results.

A number of transgenic systems have been generated and used for mutation experiments *in vivo*, which are described in detail later. Generally, these systems offer the opportunity to study gene mutations, which is considered to be a relevant toxicological endpoint indicative of a tumorigenic potential, in all organs or tissues of mice and rats from which high molecular weight genomic DNA can be obtained. A huge body of experimental data exists that demonstrates the suitability of these systems to study mutation induction *in vivo*, and that shows the relevance of the results obtained (for a detailed review see Ref. 121). However, of the models available, mostly the two commercially available bacteriophage lambda models Muta™Mouse and Big Blue®, and with restrictions the pUR288 and the gpt-delta mouse, can be considered sufficiently validated to be recommended for regulatory purposes. Commercial models have the advantage that there is professional support in the case of experimental problems, and that there is a high quality supply of all necessary material, which may be a problem for the less frequently used systems. Standards for the conductance of transgenic assays for regulatory purposes have been defined in international workshops [122, 123]. Generally, repeated dosing of at least 28 days is advised in order to allow for the detection of weak effects. Different tissues should be analyzed, guided by information such as metabolism or distribution of the test compound and also known biological mechanisms or toxicity target organs. Generally, it can be concluded that at present those systems represent the most reliable method of analyzing *in vivo* gene mutations and therefore can be considered to be the systems of choice for the *in vivo* clarification of gene mutagenic effects observed in other (*in vitro*) systems.

*Bacteriophage Lambda-Based Models (lacZ or lacI)* The two most popular and widely used transgenic models are Muta™Mouse and Big Blue®, employing the bacterial *lacZ* and *lacI* genes, respectively, as mutational targets [131, 114, 122, 123]. In those animals a recombinant bacteriophage lambda is used as the transgenic shuttle vector. Genomic DNA is extracted from any organ or tissue of interest by proteinase K digestion, followed by phenol/chloroform extraction, ethanol precipitation of the total genomic DNA, and subsequent resuspension in the appropriate buffer. For all bacteriophage lambda models, highly intact unfragmented genomic DNA is an essential element for a sufficiently high rescue efficiency, so that alternative DNA extraction methods such as columns have not been shown to be feasible so far. However, dialysis procedures have been described to yield good quality genomic DNA [124]. Subsequently, bacteriophage particles containing single copies of the recombinant phages are produced by *in vitro* lambda packaging using a commercial or self-made packaging extract. Phage DNA is brought into suitable host bacteria by infection with the reconstituted lambda particles, after which the cells are plated either on plates containing a chromogenic substrate (lacI or the earlier

lacZ method) or using a P-Gal selection system (lacZ). Advantages of both methods are the robustness of the experimental procedure, the vendor support that is provided as for other commercially available systems, which can be very helpful in the case of experimental problems, and the extensive body of experimental data and experience in the literature and the scientific community. Disadvantages of the models are that experiments are relatively expensive and laborious to perform, which is, however, the case for all gene mutation assays *in vivo*. Furthermore, in all lambda models a relatively high quality high molecular genomic DNA is essential for an efficient rescue of the transgenic phage, and both have shown to be not very sensitive for mutations involving large DNA rearrangements, such as deletions.

*Muta™Mouse* Transgenic lacZ mice were produced by microinjection of ~150 copies of the monomeric λgt10lacZ vector into the male pronucleus of fertilized eggs of (BALB/c × DBA/2) CD2F1 mice [113]. From the offspring, strain 40.6 was used to generate the commercially available Muta™Mouse model. In this strain, ~40 copies of the transgene per haploid genome are integrated at a single site on chromosome 3 [125]. As basically always, the transgene is integrated as a concatemeric cluster, in a head-to-tail fashion. The *lacZ* gene, which serves as the mutational target, has a size of ~3100 base pairs, whereas the lambda vector is about 47 kb large. To assess mutation, the λgt10lacZ shuttle vectors are excised from genomic DNA and packaged into phage heads by using an *in vitro* packaging extract, as described earlier. The resultant reconstituted phage particles are used to infect *E. coli* C ($lacZ^-$) cells. Originally, bacteria were plated onto agar plates containing the chromogenic X-Gal (a substrate for β-galactosidase that yields an insoluble blue reaction product). Blue plaques containing wild-type *lacZ* genes were distinguished visually from white plaques containing mutant *lacZ* genes. Subsequently, a simpler and faster selective system was developed in which an *E. coli* C ($galE^-lacZ^-$) host is used for phage infection [126]. In this system, a small proportion of the bacterial suspension is on nonselective titer plates, whereas the majority is seeded on selection plates containing P-Gal medium. Phospho-P-Gal accumulates to toxic concentrations in $galE^-$ cells that express a functional *lacZ* gene; thus, only phage that harbors a mutated *lacZ* will be able to form plaques on P-Gal medium. The *lacZ* mutant frequency is determined by calculating the proportion of plaques on both plate types [127]. As usual, genomic DNA is isolated from the tissue of interest by phenol/chloroform extraction of proteinase K-digested tissue followed by ethanol precipitation, and as for all bacteriophage lambda models, it is critical to obtain high molecular DNA, which has not been overly sheared during preparation. One major advantage of the lacZ model over the lacI model lies in the availability of a functional selection system, which allows the analysis of a high number of rescued copies with relative ease, whereas the lacI system needs a visual identification of colored plaques among the excess of colorless plaques (see later discussion). On the other hand, if sequencing of mutants is a goal, the ~3 kb *lacZ* gene is more laborious to sequence than the *lacI* gene, which is roughly one-third in size.

*Big Blue®* The Big Blue mouse and rat transgenic systems are based on the bacterial *lacI* gene. The λLIZα shuttle vector, which carries the bacterial *lacI* gene (1080 bp) as a mutational target, together with the *lacO* operator sequences and *lacZ* gene, was injected into a fertilized oocyte of C57BL/6 mice. The transgenic

C57Bl/6 A1 line was also crossed with an animal of the C3H line to produce a transgenic B6C3F1 mouse with the same genetic background as the National Toxicology Program bioassay test strain. The 45.6 kb λlacI construct is integrated as approximately 40 copies per genome, with integration occurring at a single locus on chromosome 4, in a head-to-tail arrangement. In the meantime, a lacI transgenic rat was produced in a Fisher 344 background [128]. Mutations arising in the rodent genome *in vivo* are scored in *E. coli* SCS-8 cells (lacZΔM15) following *in vitro* packaging of the λLIZα phage. As the BigBlue vector utilizes the lac repressor (*lacI*) gene as the mutational target, white (colorless) plaques will arise from phage bearing wild-type lacI when the SCS-8 host is plated on X-Gal medium. Mutations inactivating lacI will produce a lac repressor that is unable to bind to the lac operator so that *lacZ* transcription is derepressed and β-galactosidase will cleave X-Gal, producing a blue plaque. The proportion of blue plaques among the total number of plaques is a measure of mutant frequency. To date, there has not been an effective positive selection method for $lacI^-$ mutants developed for the Big Blue mouse or rat systems. Due to its smaller size, the *lacI* gene is easier to sequence than the three times bigger *lacZ* gene.

*cII System* For all bacteriophage lambda models, an assay has been developed that utilizes the *cII* gene within the lambda molecule. As discussed, both the *lacI* and *lacZ* genes are of considerable size (~1 and 3 kb, respectively), which makes sequencing in order to obtain molecular information on mutagenic events relatively laborious. The *cII* gene is a component of the lysogenic life cycle of bacteriophage lambda, and phages containing an inactivating forward mutation in this gene are selected for in $hft^-$ bacterial host strains [129, 130], which are commercially available. This gene, with a size of 294 base pairs, is substantially easier to sequence than *lacI* or *lacZ*, so that this model shows advantage over the traditional MutaMouse or Big Blue when sequencing is an experimental goal. However, this target gene has not been as extensively validated as *lacI* or *lacZ*.

*Conductance of a Transgenic Study* Of the previously described models, MutaMouse, Big Blue mouse and rat, and, with some restrictions, the lacZ plasmid mouse and the gpt-delta mouse have a sufficient quantity of experimental data associated with them to refer to them as sufficiently validated for drug safety objectives. Comprehensive recommendations for transgenic studies for regulatory purposes were formulated during the IWGT (International Workshops on Genotoxicity Tests) held 1999 in Washington, DC, and 2001 in Plymouth, UK. In these workshops, advice was given on how a transgenic mutation study should be designed in order to provide the maximum confidence in a negative result, that is, exclusion of a mutagenic potential under the chosen experimental conditions [122, 123]. By their nature, transgenic mutation assays are comparably insensitive to biological artifacts, which may mimic a mutagenic response, as happens for other endpoints, due to the fact that the endpoint assessed, the gene mutation, is the result of specific sequential events that are not mimicked by cytotoxic effects, as occurs with other endpoints. A number of experimental artifacts may exist, such as the carryover of DNA adducts into bacteria with subsequent *ex vivo* mutagenesis in bacteria. However, the large body of evidence indicates that the mutations measured in these systems arise predominantly in the rodent tissues [122]. Also, one of the main confounders to

contribute to the apparent mutant frequency is the spontaneous desamination of methylated cytosine, which occurs preferentially at GpC sites within the gene. Therefore, the spontaneous mutant frequency is relatively high, which means that it is basically difficult to detect weak increases. However, gene mutations, particularly those in selectionally neutral transgenes, accumulate linearly in most tissues unless cells are eliminated by cell turnover, which should occur at the same rate for mutated and unmutated cells. Therefore, a repeated (daily) dosing is recommended for most situations, and cessation of treatment should be at or shortly before tissue sampling (see later discussion).

A number of influence factors should be considered when designing a transgenic study. Some of those, such as species (if available), route of administration, or gender, follow the same lines of argumentation as would be used for other *in vivo* assays. Thus, species- or gender-specific metabolism or other PK parameters may guide the choice of species or gender, where it has to be kept in mind that, apart from Big Blue or gpt-delta rats, the mouse will be the species of choice in all cases. Information from other toxicity studies can be used for decision making, and as a general rule the use of one gender is considered sufficient if no toxicity or pharmacokinetic differences are observed between males and females. Thus, unless scientific evidence commands otherwise, the following features should characterize a transgenic rodent study used for regulatory purposes [122, 123]: 5 animals per gender (if a gender difference in factors such as metabolism or toxicity is known or suspected), or 7 animals of one gender, normally males, are used. Scenarios such as a gender-specific therapeutic use of a test substance may guide the choice of gender. Generally, the animals are treated daily for 28 consecutive days and tissues are sampled 3 days after the final treatment. Also, slightly differing recommendations have been formulated [131], and one may deviate from this scheme in cases where the human treatment schedule is strongly different. However, there is consensus that, unless there is a strong mutagenic effect to be expected, a repeated daily dosing for 4 weeks is sufficient to detect most mutagens. Obviously, this recommendation could be modified by reducing or increasing the number of treatments or the length of the treatment period, when scientifically justified. Normally, at least one rapidly proliferating and one slowly proliferating tissue should be sampled. Scientific arguments such as metabolism, accumulation or disposition, or site-of-contact considerations should guide the choice of tissue to be analyzed. For statistical reasons, 124,000–300,000 plaques or colonies should be analyzed per tissue sample, and three doses are used, whereas not all doses need to be analyzed, should the top dose yield a negative result. At sampling, tissues should be flash frozen, after which they can be stored at –80 °C for prolonged times. Likewise, isolated genomic DNA can be stored at 4 °C, although not as long as deep frozen tissues. Although sequencing data are not normally required, they might provide useful additional information in specific circumstances. However, they are mainly used to identify and correct for clonal expansion (i.e., an artificially elevated mutant frequency), and in some cases to determine a mechanism associated with a positive response, although owing to the significant variation in background mutant spectra the power of this tool may be limited. A schematic overview of a transgenic study is given in Fig. 6.

*Advantages and Limitations* Thus, in summary, the advantages of transgenic mutation assays lie in the analysis of gene mutations, an endpoint that can be considered

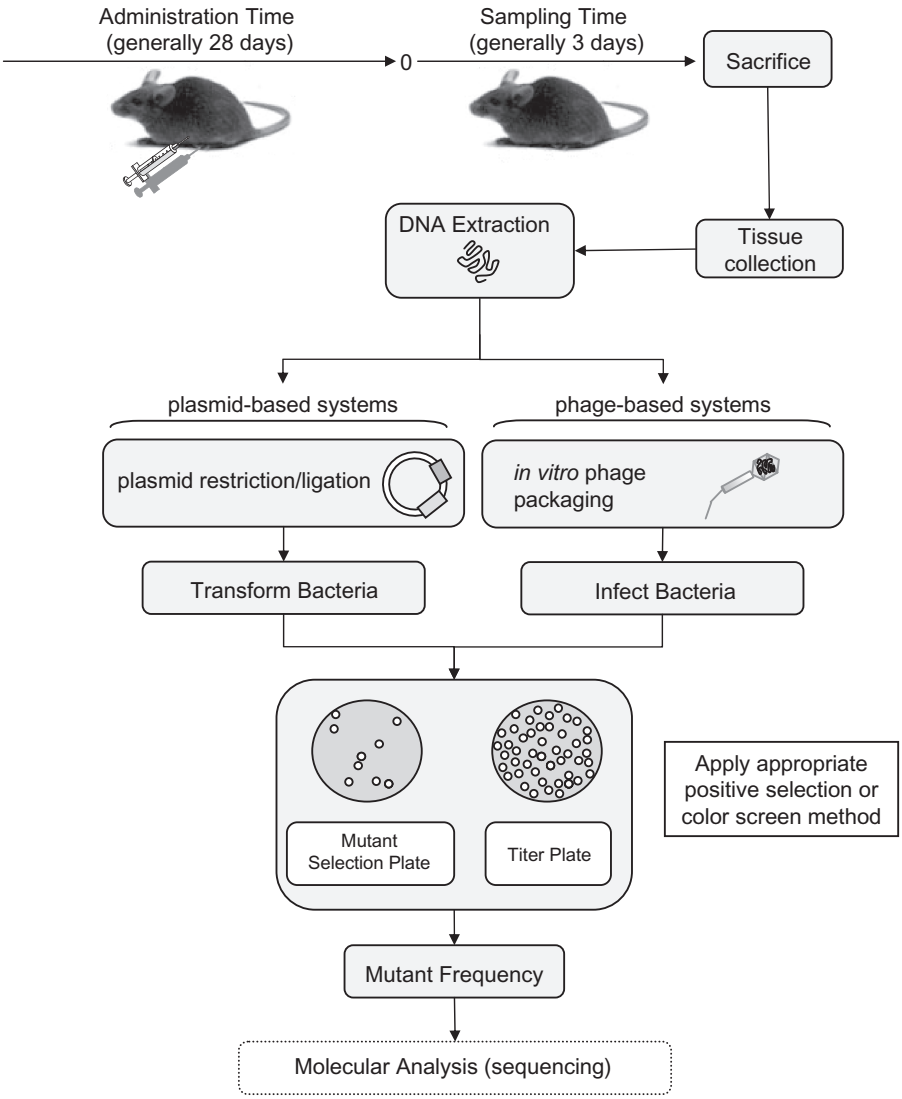


**FIGURE 6** Flow diagram for performing a transgenic animal experiment.

as very relevant for tumor induction, in basically any tissue of interest in mice and some rat models. Obviously, the analysis is limited to those strains where specific models have been generated. The fixed gene mutations are definitive endpoints that accumulate in most organs so that an increased sensitivity can be achieved by repeated dosing. Thus, when using the recommended 4 week treatment schedule, it can be expected that even weak mutagens are detected reliably [121–123]. Also, gene mutation induction in those models is not subject to artifactual conditions such as cellular toxicity, so that a positive result can be regarded as a very relevant indicator of an *in vivo* mutagenicity. These assays permit the screening of a large number of copies of a locus and provide statistically reliable results on the frequency of

mutations in an animal or tissue. Furthermore, the mutants can be sequenced so that the molecular nature of the mutation can be analyzed. The transgenic loci are expected to be genetically neutral, which avoids the influence of selective pressures *in vivo* on the mutant frequency. Therefore, transgenic models provide at this moment the best validated models for investigating tissue-specific induction of gene mutations and the clarification of *in vitro* gene mutagenicity findings. The disadvantage of these assays lies in a relatively laborious and therefore expensive procedure with a recommended 4 week treatment time. However, the effort needed, particularly for the postmortem phase, depends on the know-how of the respective laboratory, and, once used routinely, experience shows that the assays can be conducted with relative ease and high efficiency.

### *Other Transgenic Models*

*pUR288 Plasmid Mouse* In order to overcome some of the limitations of bacteriophage lambda models, such as the insensitivity to deletion mutations or the dependence on very high molecular genomic DNA, a pUR288 plasmid-based transgenic mouse was introduced [132, 133]. This model carries the plasmid pUR288, which utilizes the same lacZ sequence as the MutaMouse as the mutational target, and therefore the P-Gal selection system can be used. The mice were made on a C57Bl/6J background, and plasmid copies are found in line 60, the line that has been used for most experiments, integrated on chromosomes 3 and 4 [134]. Plasmid rescue from total genomic DNA, which can be prepared by classical phenol/chloroform extraction followed by ethanol precipitation, or alternatively by using commercially available DNA isolation columns, is done by digesting the DNA with HindIII. From this digest, plasmid copies are specifically retrieved by using a lacI/lacZ fusion protein, of which the lacI domain, which is the repressor for the *lacZ* gene located on the plasmid, specifically binds to the operator sequence of lacZ. Rescue is done with anti-β-galactosidase-coated ferromagnetic beads, which bind the lacZ domain of the fusion protein and thus couple it to the surface of the beads. After removing the excess of unbound genomic DNA, plasmid copies are eluted from the beads with the lacZ inducer IPTG (isopropyl thiogalactoside), relegated, and electroporated into galE and lacZ-deficient *E. coli* C host bacteria. Then 0.1% of the transformants are plated on titer plates to assess the overall rescue efficiency, whereas 99.9% are seeded on P-Gal containing selection plates on which only mutants can grow. Mutant frequency is determined by the colony ratio on both plates times the dilution factor. Experiments using this model have shown that indeed it is able to detect mutations induced by typical clastogens such as X-ray [132] or age-dependent mutation accumulation, and that even large mutations ranging into the mouse genome are recovered, owing to the existence of HindIII sites in the mouse genomic DNA [134]. Other data show that part of the spontaneous mutation spectrum is generated by a residual star activity of HindIII, but generally it is assumed that this effect does not contribute significantly to the overall mutant frequency [135].

*gpt-Delta Mouse and Rat* Another transgenic system, which was constructed to overcome the relative insensitivity of bacteriophage lambda models of the *lacI* and *lacZ* type for clastogenic events, is the *gpt*-delta mouse [136]. This model carries the

bacteriophage lambda EG10 as the transgene. Similar to lambda systems, the phage is rescued from total genomic DNA by an *in vitro* packaging method. The novel feature of this system is the use of two distinct selection methods, namely, 6-thioguanine for the selection of point mutations and Spi$^-$ selection for larger rearrangements such as deletions. The 6-thioguanine selection used the *gpt* gene of *E. coli* as the selection marker, which is known to be sensitive for point mutations and, due to its relatively small size of 456 base pairs, is also convenient to sequence. The second selection principle uses a system that relies on the loss of two distinct genes, *red* and *gam*, to allow for a positive selection by plaque formation on P2 lysogens of *E. coli*. Spi$^-$ selection utilizes the restricted growth of wild-type lambda phage in P2 lysogens [137, 138]. Only mutant lambda phages that are deficient in the functions of both the *gam* and *redBA* genes can grow well in P2 lysogens and display the Spi$^-$ phenotype. Simultaneous inactivation of both the *gam* and *redBA* genes, the prerequisite for plaque formation, happens generally only in the case of large deletions. Experiments with this model indeed have shown that typical point mutagens (e.g., ethylnitrosourea) or clastogens (e.g., gamma irradiation, mitomycin C) induce the expected effects [139]. In between, this model is also available on a rat background [140]. Obviously, this system is somewhat more laborious than a system that uses only one selective principle. However, without sequencing, the two parallel selection methods reveal some fundamental information about possible mutagenic mechanisms.

*ΦX174 Mouse* The ΦX174 am3 model was presented by Burkhardt et al. [141]. The bacteriophage ΦX174, which is about 5 kb in size, is chromosomally integrated at about 50 copies per haploid genome of inbred C57Bl/6 mice. Phage recovery from the mouse genome is done from proteinase K digested, phenol/choloroform extracted, and ethanol precipitated total genomic DNA by a PstI and Sau3A restriction digest, followed by filtration on Sephacryl® S-1000 gel filtration columns to remove genomic DNA and RNA. The DNA is ligated and electroporated into competent cells, where mutant frequency determination is done by a single burst method using *E. coli* CQ2 (sup$^+$) cells to determine the titer of totally recovered phage copies, whereas mutations at the A:T target site at the nonsense mutation in the am3 stop codon are selected for in *E. coli* C (sup$^-$) cells. In this system, reversion of the am3 target codon can occur via two transversions and one transition, which means that only certain types of mutations can be detected. On the other hand, the specificity of this site prevents a major type of mutation, which, in the bacteriophage lambda systems, stems from spontaneous desamination of 5-methylcytosin at GpC sites, which in those models contributes significantly to the background mutation load.

A few other transgenic systems have been described. One of the earlier models introduced was a supF gene-based bacteriophage lambda mouse, which, after *in vitro* lambda packaging from mouse genomic DNA and infection of suitable host bacteria, mutations in the supF gene are selected for by growth on lacI(amber) *E. coli* in the presence of X-Gal [111]. Another model used the plasmid pML4 as the shuttle vector, present at relatively high copy numbers (several hundred per genome, depending on the mouse line [142]). After restriction digest and direct ligation of genomic DNA, followed by electroporation of kanamycin-resistant and streptomycin-sensitive *E. coli* host cells, mutations in the *rspL* gene are detected by

plating on kanamycin- and streptomycin-containing plates. Also, an *in situ* model was presented that held the promise of mutation detection *in situ*, which would allow an analysis at the single-cell level [143]. However, those systems have so far not advanced to a state where more than one or a few laboratories have used them, so that the basis for experimental data is insufficient to draw conclusions on their performance.

***Mouse Spot Test*** The mouse spot test is an *in vivo* assay capable of detecting genetic effects of several kinds, including intragenic mutations, minute deficiencies, deletions (through breakage or nondisjunction) of various amounts of chromosomal material, and somatic crossing-over [144, 145]. The assay involves treatment of mouse embryos, which are heterozygous for different recessive coat color genes and are treated *in utero* at gestation day 9–11 with the test substance. The exposed embryo at gestation day 10 contains about 150–200 melanoblasts and each melanoblast has 4 coat color genes that are studied [145, 146]. The *in utero* exposure may result in an alteration or loss of a specific wild-type allele in a pigment precursor cell, resulting in a color spot in the coat of the adult animal. The frequency of spots is compared with the frequency in negative controls [145, 146]. In the mouse spot test there are four possible mechanisms by which the recessive coat color alleles can be expressed: gene mutation in the wild-type allele, deficiency (large or small) of a chromosomal segment involving the wild-type allele, nondisjunctional (or other) loss of the chromosome carrying the wild-type allele, and somatic recombination (using a marker gene that then becomes homozygous after recombination). Mutagenic but also clastogenic effects are detected by this test system [147, 148].

The mouse spot test is not a very frequently employed genotoxicity assay for the detection of somatic mutations; however, an OECD guideline exists [149]. The test was suggested as a complementary *in vivo* test to the bacterial mutagenicity assay for detection of mutagenic substances and as a confirmatory test for the identification of carcinogens [145]. However, the infrequent use of the spot test may partly be due to considerations of cost effectiveness and the number of animals needed for testing and also for toxicological considerations. The usefulness of the mouse spot test in toxicology is limited by restrictions in toxicokinetics, sensitivity, target cell/organ, and molecular genetics. The test may be used to further investigate specific issues such as compounds demonstrating effects in the *in vitro* bacterial mutation assay but not in clastogenicity tests *in vitro* and *in vivo*. One such example is the antiviral pryrimidine nucleoside analogue 5-(2-chloroethyl)-2′-deoxyuridine [150]. However, for the detection of gene but not chromosome mutations *in vivo*, the mouse spot test should likely be replaced by the transgenic mouse assays, which were shown to be of comparable sensitivity [151].

***Other Assays for Somatic Cell Gene Mutations*** Various other assays exist that investigate endogenous genes. These tests are rarely used for testing purposes because they are cumbersome and generally not suitable for routine use. Some promising mutation tests using endogenous genes have been developed. These endogenous gene mutation assays include RetinoBlast, Hprt, Aprt, $Tk^{+/-}$, and Dlb-1. A comprehensive review outlining the current status of these test has recently been published [121].

## 4 *IN VIVO* GENOTOXICITY TESTS FOR INVESTIGATING GERMLINE CELLS

Thus far, all established germ cell mutagens have been shown also to produce positive results in bone marrow assays. There is no current evidence for germ-cell-specific mutagens [152]; however, the reverse is not true. Not all somatic cell mutagens can be demonstrated to be germ cell mutagens, which indicates mechanisms exist that protect germ cells against genotoxic exposure of induction of DNA damage. However, for most compounds identified as *in vivo* somatic cell mutagens, no specific genotoxicity testing in germ cells will be needed since these compounds will be assumed to be both potential genotoxic carcinogens and potential germ cell mutagens, unless the exposure of germ cells can be reliably excluded, which will be difficult in most cases. Data on the mutagenic effects in germ cell DNA are needed before any definite conclusions can be drawn relating to the mutagenic hazard of a chemical specifically to germ cells. There may be justified cases in which germ cell studies will be undertaken for the specific purpose of demonstrating whether an identified *in vivo* somatic cell mutagen is a germ cell mutagen as well. However, none of these assays provide conclusive information as to whether the effects seen are heritable in future generations. The methods available for such risk assessments, such as the mouse heritable translocation test and the mouse specific locus test, involve the investigation of effects in subsequent generations bred from treated animals. In view of the very large number of animals used, these are not practical options and should only be considered in exceptional circumstances.

In principle, the methods described earlier in this chapter may be applied to investigate genotoxic effects in germ line cells. The transgenic animal systems may be used to investigate effects in germ cells [153], micronucleus induction may be detected in spermatocytes using appropriate staining methods, and induction of DNA lesions in germ cell DNA may be obtained from UDS [154–156], comet [157, 158], or postlabeling assays [159].

Specific germline cell tests are not very frequently applied; however, they are listed in ICH guidelines. A brief summary on the specific areas of use, advantages, and disadvantages of these test systems are provided. In general, the available methods involve measuring effects in the gonads of male rodents. Methods for investigating clastogenicity in mammalian spermatogonial cells are well established. There is an internationally recognized guideline (OECD No. 483) for this approach [160].

The dominant lethal assay may also be used to investigate clastogenicity or aneugenicity in germ cells [161]. There is an internationally recognized guideline, OECD No. 478, for this assay [162]. For this method, the endpoint is the production of embryo-lethal genetic changes measured as death of the conceptus as a blastoma or soon afterwards. Dominant lethal mutations are believed to be primarily due to structural or numerical chromosome aberrations. There are essentially two different dosing regimes that may be used in this assay. In one case, this involves repeated dosing of the males for a period covering spermatogenesis followed by mating with untreated females and examining the latter for dominant lethals after an appropriate period of gestation. In the other regime, single dosing is followed by sequential mating of females. The latter provides information on the various stages of the germ cell cycle that may be affected but uses very many more animals and the need for this additional information is rarely justified.

## 5 CONCLUSION

The assessment of a genotoxic potential of a drug candidate is typically determined through a battery of tests conducted *in vitro* and *in vivo*. Genotoxicity tests provide a means to identify potential genotoxic carcinogens relatively early in the development of a drug candidate. According to current internationally accepted guidance, the most appropriate initial *in vivo* test will be a bone marrow micronucleus assay unless the initial considerations indicate otherwise. Techniques for the assessment of whole chromosomes are appropriate if there is evidence of aneugenicity. If a genotoxic hazard is observed in any *in vitro* tests, further investigations are conducted to assess whether genotoxic activity observed *in vitro* can be expressed in somatic cells *in vivo*. For this purpose, supplemental tests utilizing other tissue(s) are conducted. The type of study needs to be considered on a case-by-case basis, taking into account the available information on the compound including the results from earlier tests. Studies that may be appropriate include, but are not limited to, the liver UDS assay, comet assay, $^{32}$P-postlabeling assay, or assays using transgenic animals. Specific germline genotoxicity assays usually are not required since it can be assumed that drug candidates demonstrating *in vivo* genotoxicity will be both potential genotoxic carcinogens and germ cell mutagens. However, in some cases, germ cell studies may be undertaken to demonstrate that a somatic cell mutagen is not a germ cell mutagen, unless there is convincing evidence that the chemical does not dispose into the germ cell compartment. In such a case, most of the above listed supplemental assays can be applied to germline cells. However, it is important to note that none of these assays provide conclusive information as to whether genotoxic effects will be seen in future generations. Assays that facilitate such risk estimates include the heritable translocation test and the mouse specific locus test. However, these assays are not practical in view of the very large number of animals needed. Ultimately, the choice and design of *in vivo* genotoxicity assays needs to be scientifically justified and conducted in a way that adds value to the nonclinical safety assessment either in terms of initial genotoxic hazard identification or with regard to understanding the risk and relevance of positive results obtained from *in vitro* genotoxicity tests or rodent cancer bioassays.

## REFERENCES

1. ICH S2B. *A Standard Battery for Genotoxicity Testing of Pharmaceuticals*. 1997. http://www.ich.org/.
2. Müller L, Blakey D, Dearfield KL, Galloway S, Guzzie P, Hayashi M, Kasper P, Kirkland D, MacGregor JT, Parry JM, Schechtman L, Smith A, Tanaka N, Tweats D, Yamasaki H. Strategy for genotoxicity testing and stratification of genotoxicity test results—report on initial activities of the IWGT Expert Group. *Mutat Res* 2003;540:177–181.
3. ICH M3(M). *Maintenance of the ICH Guideline on Non-clinical Safety Studies for the Conduct of Human Clinical Trials for Pharmaceuticals*. 2000. http://www.ich.org/.
4. Ashby J, Tennant RW. Definitive relationships among chemical structure, carcinogenicity and mutagenicity for 301 chemicals tested by the U.S. NTP. *Mutat Res* 1991; 257:229–308.
5. ICH S2A. *Specific Aspects of Regulatory Genotoxicity Tests*. 1995. http://www.ich.org/.

6. Food and Drug Administration. *Guidance for Industry, Investigators, and Reviewers. Exploratory IND Studies*. Draft Guidance. 2005.
7. OECD Guideline 474. *Guideline for the Testing of Chemicals. Mammalian Erythrocyte Micronucleus Test*. OECD; 1997.
8. OECD Guideline 475. *Guideline for the Testing of Chemicals. Mammalian Bone Marrow Chromosome Aberration Test*. OECD; 1997.
9. Heddle J. A rapid *in vivo* test for chromosome damage. *Mutat Res* 1973;18:187–190.
10. MacGregor JT, Wehr CM, Gould DH. Clastogen-induced micronuclei in peripheral blood erythrocytes: the basis of an improved micronucleus test. *Environ Mutagen* 1980;2:509–514.
11. Hayashi M, Tice RR, MacGregor JT, Anderson D, Blakey DH, Kirsch-Volders M, Oleson FB Jr, Pacchierotti F, Romagna F, Shimada H, Sutou S, Vannier B. *In vivo* rodent erythrocyte micronucleus assay. *Mutat Res* 1994;312:293–404.
12. Hayashi M, MacGregor JT, Gatehouse DG, Adler I-D, Blakey DH, Dertinger SD, Krishna G, Morita T, Russo A, Sutou S. *In vivo* rodent erythrocyte micronucleus assay. II. Some aspects of protocol design including repeated treatments, integration with toxicity testing, and automated scoring. *Environ Mol Mutagen* 2000;35:234–252.
13. CSGMT. Sex differences in the micronucleus test. *Mutat Res* 1986;172:151–163.
14. CSGMT. Strain differences in the micronucleus test. *Mutat Res* 1988;204:307–316.
15. CSGMT. Single versus multiple dosing in the micronucleus test: the summary of the fourth collaborative study by CSGMT/JEMS-MMS. *Mutat Res* 1990;234:205–222.
16. CSGMT. Micronucleus test with mouse peripheral blood erythrocytes by acridine orange supravital staining: the summary report of the 5th collaborative study by CSGMT/JEMS-MMS. *Mutat Res* 1992;278:83–98.
17. CSGMT. Protocol recommended by the CSGMT/JEMS-MMS for the short-term mouse peripheral blood micronucleus test. *Mutagenesis* 1995;10:153–159.
18. Wakata A, Miyamae Y, Sato S, Suzuki T, Morita T, Asano N, Awogi T, Kondo K, Hayashi M. Evaluation of the rat micronucleus test with bone marrow and peripheral blood: summary of the 9th collaborative study by CSGMT/JEMS-MMS. *Environ Mol Mutagen* 1998;13:347–356.
19. Hamada S, Sutou S, Morita T, Wakata A, Asanami S, Hosoya S, et al. Evaluation of the rodent micronucleus assay by a 28-day treatment protocol: summary of the 13th collaborative study by the Collaborative Study Group for the Micronucleus Test (CSGMT)/Environmental Mutagen Society of Japan (JEMS)–Mammalian Mutagenicity Study Group (MMS). *Environ Mol Mutagen* 2001;37:93–110.
20. Hamada S, Nakajima K, Namiki C, Serikawa T, Hayashi M. Sex differences in the chemical induction of micronuclei in the rat. *Environ Mutagen Res* 2003;25:33–37.
21. Hamada S, Yamasaki K-I, Nakanishi S, Omori T, Serikawa T, Hayashi M. Evaluation of the general suitability of the rat for the micronucleus assay: the effect of cyclophosphamide in 14 strains. *Mutat Res* 2001;495:127–134.
22. Hayashi M, Sutou S, Shimada H, Sato S, Sasaki Y-F, Wakata A. Difference between intraperitoneal and oral gavage application in the micronucleus test, the 3rd collaborative study by CSGMT/JEMS-MMS. *Mutat Res* 1989;23:329–344.
23. MacGregor JT, Wehr CM, Henika PR, Shelby MD. The *in vivo* erythrocyte micronucleus test: measurement at steady state increases assay efficiency and permits integration with toxicity studies. *Fundam Appl Toxicol* 1990;14:513–522.
24. Garriott ML, Brunny JD, Kindig DEF, Parton JW, Schwier LS. The *in vivo* rat micronucleus test: integration with a 14-day study. *Mutat Res* 1995;342:71–76.

25. Asanami S, Shimono K, Sawamoto O, Kurisu K, Uejima M. The suitability of rat peripheral blood in subchronic studies for the micronucleus assay. *Mutat Res* 1995;347:73–78.
26. Holder HE, Majeska JB, Studwell D. A direct comparison of mouse and rat bone marrow and blood as target tissues in the micronucleus assay. *Mutat Res* 1997; 391:87–89.
27. Torous DK, Hall NE, Murante FG, Gleason SE, Tometski CR, Dertinger SD. Comparative scoring of micronucleated reticulocytes in rat peripheral blood by flow cytometry and microscopy. *Toxicol Sci* 2003;74:309–314.
28. Frieauff W, Romagna F. Technical aspects of automatic micronucleus analysis in rodent bone marrow assays. *Cell Biol Toxicol* 1994;10:283–289.
29. Grawé J, Zetterbert G, Amneus H. Flow-cytometric enumeration of micronucleated polychromatic erythrocytes in mouse peripheral blood. *Cytometry* 1992;13:750–758.
30. Tometsko AM, Torous DK, Dertinger SD. Analysis of micronucleated cells by flow cytometry. 3. Advanced technology for detecting clastogenic activity. *Mutat Res* 1993;292:145–153.
31. Dertinger SD, Torous DK, Tometsko KR. Simple and reliable enumeration of micronucleated reticulocytes with a single-laser flow cytometer. *Mutat Res* 1996;371:283–292.
32. Dertinger SD, Torous DK, Tometsko KR. Flow cytometric analysis of micronucleated reticulocytes in mouse bone marrow. *Mutat Res* 1997;390:257–262.
33. Criswell K, Krishna G, Zielinski D, Urda G, Theiss J, Juneau P, Bleavins M. Use of acridine orange in flow cytometric assessment of micronuclei induction. *Mutat Res* 1998;414:63–75.
34. Torous DK, Hall NE, Dertinger SD, Diehl MS, Illi-Love AH, Cederbrant K, et al. Flow cytometric enumeration of micronucleated reticulocytes: high transferability among 14 laboratories. *Environ Mol Mutagen* 2001;38:59–68.
35. Torous DK, Hall NE, Illi-Love AH, Diehl MS, Cederbrant K, Sandelin K, et al. Interlaboratory validation of a CD71-based flow cytometric method (MicroFlow®) for scoring of micronucleated reticulocytes in mouse peripheral blood. *Environ Mol Mutagen* 2005;45:44–55.
36. Schlegel R, MacGregor JT. The persistence of micronucleated erythrocytes in the peripheral circulation of normal and splenectomized Fischer 344 rats: implications for cytogenetic screening. *Mutat Res* 1984;127:169–174.
37. Torous D, Dertinger S, Hall N, Tometsko C. Enumeration of micronucleated reticulocytes in rat peripheral blood: a flow cytometric study. *Mutat Res* 2000;465:91–99.
38. Schlegel R, MacGregor JT. A rapid screen for cumulative chromosomal damage in mice. Accumulation of circulating micronucleated erythrocytes. *Mutat Res* 1982;113:481–487.
39. Schlegel R, MacGregor JT. The persistence of micronuclei in peripheral blood erythrocytes: detection of chromic chromosome breakage in mice. *Mutat Res* 1983;104: 367–369.
40. Choy WN, MacGregor JT, Shelby MD, Maronpot RR. Induction of micronuclei by benzene in B6C3F1 mice: retrospective analysis of peripheral blood smears from the NTP carcinogenesis bioassay. *Mutat Res* 1985;143:55–59.
41. Krishna G, Urda G, Theiss J. Principles and practices of integrating genotoxicity evaluation into routine toxicology studies: a pharmaceutical industry perspective. *Environ Mol Mutagen* 1998;32:115–120.
42. MacGregor JT, Heddle JA, Hite M, Margolin BH, Ramel C, Salamone MF, Tice RR, Wild D. Guidelines for the conduct of micronucleus assays in mammalian bone marrow erythrocytes. *Mutat Res* 1987;189:103–112.

43. Mackay JM, Elliott BM. Dose-ranging and dose-settings for *in vivo* genetic toxicology studies. *Mutat Res* 1992;271:97–99.
44. Schmid W. The micronucleus test for cytogenetic analysis. In Hollaender A, Ed. *Chemical Mutagens*, Vol 4. New York: Plenum Press; 1976, pp 31–53.
45. Salamone MF, Heddle JA. The bone marrow micronucleus assay: rationale for a revised protocol. In Serres FJ, Ed. *Chemical Mutagens*, Vol 8. New York: Plenum Press; 1983, pp 111–149.
46. Romagna F. Improved method of preparing bone marrow micronucleus slides: fractionation of a pure PE and NE population from rodent bone marrow. *Mutat Res* 1988;206:307–309.
47. Hayashi M, Morita T, Kodama Y, Sofuni T, Ishidate M Jr. The micronucleus assay with mouse peripheral blood reticulocytes using acridine orange-coated slides. *Mutat Res* 1990;245:245–249.
48. Hayashi M, Sofuni T, Ishidate M. An application of acridine orange fluorescent staining to the micronucleus test. *Mutat Res* 1983;120:241–247.
49. MacGregor JT, Wehr CM, Langlois RG. A simple fluorescent staining procedure for micronuclei and RNA in erythrocytes using Hoechst 33258 and pyronin Y. *Mutat Res* 1983;120:269–275.
50. Kirkland DJ. Statistical evaluation of mutagenicity test data: recommendations of the UKEMS. *Environ Health Persp Suppl* 1994;102:43–47.
51. Yoshimura I, Sofuni T, Ishidate M Jr, Hayashi M. A procedure for data analysis of the rodent micronucleus test involving a historical control. *Environ Mol Mutagen* 1989;13:347–356.
52. Lovell DP. Anderson D, Albanese R, Amphlett GE, Clare G, Ferguson R, Richold M, Papworth DG, Savage JRK. Statistical analysis of *in vivo* cytogenetic assays. In Kirkland DJ, Ed. *Statistical Evaluation of Mutagenicity Test Data. UKEMS Sub-Committee on Guidelines for Mutagenicity Testing, Report, Part III*. New York: Cambridge University Press; 1989, pp 184–232.
53. Adler I-D, Bootman J, Favor J, Hook G, Schriever-Schwemmer G, Welzl G, Whorton E, Yoshimura I, Hayashi M. Recommendations for statistical designs of *in vivo* mutagenicity tests with regard to subsequent statistical analysis. *Mutat Res* 1998;417:19–30.
54. Asanami S, Shimono K. Hypothermia induces micronuclei in mouse bone marrow cells. *Mutat Res* 1997;393:91–98.
55. Asanami S, Shimono K. Effects of chemically- and environmentally-induced hypothermia on micronucleus induction in rats. *Mutat Res* 2000;471:81–86.
56. Hamada S, Nakajima K, Serikawa T, Hayashi M. The effect of aging on the results of the rat micronucleus assay. *Mutagenesis* 2003;18:273–275.
57. Ortiz R, Medina H, Rodriguez L, Gonzalez-Marquez H, Cortés E. Spontaneous and mitomycin C-induced micronuclei in peripheral blood reticulocytes from severely malnourished rats. *Environ Mol Mutagen* 2004;43:179–185.
58. Tice RR, Shelby MD. International Workshop on Standardisation of Genotoxicity Test Procedures. Report of *in vivo* subgroup. *Mutat Res* 1994;312:287–292.
59. Bentley KS, Kirkland D, Murphy M, Marshall R. Evaluation of thresholds for benomyl- and carbendazim-induced aneuploidy in cultured human lymphocytes using fluorescence *in situ* hybridization. *Mutat Res* 2000;464:41–51.
60. Kirkland DJ, Müller L. Interpretation of the biological relevance of genotoxicity test results: the importance of thresholds. *Mutat Res* 2000;464:137–147.
61. Ahnstrom G. Techniques to measure DNA single-strand breaks in cells: a review. *Int J Radiat Biol* 1988;54:695–707.

62. Elia MC, Storer RD, McKelvey TW, Kraynak AR, Barnum JE, Harmon LS, DeLuca JG, Nichols WW. Rapid DNA degradation in primary rat hepatocytes treated with diverse cytotoxic chemicals: analysis by pulsed field gel electrophoresis and implications for alkaline elution assays. *Environ Mol Mutagen* 1994;24:181–191.
63. Hartmann A, Schumacher M, Plappert-Helbig U, Lowe P, Suter W, Mueller L. Use of the alkaline *in vivo* comet assay for mechanistic genotoxicity investigations. *Mutagenesis* 2004;19:51–59.
64. Brendler-Schwaab S, Hartmann A, Pfuhler S, Speit G. The *in vivo* comet assay: use and status in genotoxicity testing. *Mutagenesis* 2005;20:245–254.
65. Tice RR. The single cell gel/comet assay: a microgel electrophoretic technique for the detection of DNA damage and repair in individual cells. In Phillips DH, Venitt S, Eds. *Environmental Mutagenesis*. Oxford, UK: Bios Scientific Publishers; 1995, pp 315–339.
66. Speit G, Hartmann A. The comet assay (single-cell gel test)—a sensitive genotoxicity test for the detection of DNA damage and repair. In Henderson DS, Ed. *Methods in Molecular Biology Volume 113: DNA Repair Protocols: Eucaryotic Systems*, 2nd ed. Totowa, NJ: Humana Press; 2005, pp 275–286.
67. Tice RR, Agurell E, Anderson D, Burlinson B, Hartmann A, Kobayashi H, Miyamae Y, Rojas E, Ryu J-C, Sasaki YF. The single cell gel/comet assay: guidelines for *in vitro* and *in vivo* genetic toxicology testing. *Environ Mol Mutagen* 2000;35:206–221.
68. Hartmann A, Agurell E, Beevers C, Brendler-Schwaab S, Burlinson B, Clay P, Collins AR, Smith A, Speit G, Thybaud V, Tice RR. Recommendations for conducting the *in vivo* alkaline comet assay. *Mutagenesis* 2003;18:45–51.
69. Burlinson B, Tice RR, Speit G, Agurell E, Brendler-Schwaab SY, Collins AR, Escobar P, Honma M, Kumaravel TS, Nakajma M, Sasaki YF, Thybaud V, Uno Y, Vasquez M, Hartmann A. In Vivo Comet Assay Workgroup, part of the Fourth International Workgroup on Genotoxicity Testing. *Mutat Res* 2007;627:31–35.
70. Pfuhler S, Wolf HU. Detection of DNA-crosslinking agents with the alkaline comet assay. *Environ Mol Mutagen* 1996;27:196–201.
71. Merk O, Speit G. Detection of crosslinks with the comet assay in relationship to genotoxicity and cytotoxicity. *Environ Mol Mutagen* 1999;33:167–172.
72. Speit G, Hartmann A. The contribution of excision repair to the DNA-effects seen in the alkaline single cell gel test (comet assay). *Mutagenesis* 1995;10:555–559.
73. Collins AR, Duthie SJ, Dobson VL. Direct enzymic detection of endogenous oxidative base damage in human lymphocyte DNA. *Carcinogenesis* 1993;14:1733–1735.
74. Tice RR, Andrews PW, Hirai O, Singh NP. The single cell gel (SCG) assay: an electrophoretic technique for the detection of DNA damage in individual cells. In Witmer CR, Snyder RR, Jollow DJ, Kalf GF, Kocsis JJ, Sipes IG, Eds. *Biological Reactive Intermediates IV, Molecular and Cellular Effects and Their Impact on Human Health*. New York: Plenum Press; 1991, pp 157–164.
75. Brendler-Schwaab SY, Schmezer P, Liegibel U, Weber S, Michalek K, Tompa A, Pool-Zobel BL. Cells of different tissues for *in vitro* and *in vivo* studies in toxicology: compilation of isolation methods. *Toxicol in Vitro* 1994;8:1285–1302.
76. Miyamae Y, Yamamoto M, Sasaki YF, Kobayashi H, Igarashi Soga M, Shimol K, Hayashi M. Evaluation of a tissue homogenization technique that isolates nuclei for the *in vivo* single cell gel electrophoresis (comet) assay: a collaborative study by five laboratories. *Mutat Res* 1998;418:131–140.
77. Sasaki YF, Tsuda S, Izumiyama F, Nishidate E. Detection of chemically induced DNA lesions in multiple mouse organs (liver, lung, spleen, kidney, and bone marrow) using the alkaline single cell gel electrophoresis (comet) assay. *Mutat Res* 1997;388:33–44.

78. Olive PL, Frazer G, Banath JP. Radiation-induced apoptosis measured in TK6 human B lymphoblast cells using the comet assay. *Radiat Res* 1993;136:130–136.
79. Henderson L, Wolfreys A, Fedyk J, Bourner C, Windebank S. The ability of the comet assay to discriminate between genotoxins and cytotoxins. *Mutagenesis* 1998;13:89–94.
80. Hartmann A, Kiskinis E, Fjaellman A, Suter W. Influence of cytotoxicity and compound precipitation on test results in the alkaline comet assay. *Mutat Res* 2001;497:199–212.
81. Kiffe M, Christen P, Arni P. Characterization of cytotoxic and genotoxic effects of different compounds in CHO K5 cells with the comet assay (single-cell gel electrophoresis assay). *Mutat Res* 2003;537:151–168.
82. Rundell MS, Wagner ED, Plewa MJ. The comet assay: genotoxic damage or nuclear fragmentation? *Environ Mol Mutagen* 2003;42:61–67.
83. Mensing T, Welge P, Voss B, Fels LM, Fricke HH, Bruning T, Wilhelm M. Renal toxicity after chronic inhalation exposure of rats to trichloroethylene. *Toxicol Lett* 2002;128:243–247.
84. Scassellati-Sforzolini G, Pasquini R, Moretti M, Villarini M, Fatigoni C, Dolara P, Monarca S, Caderni G, Kuchenmeister F, Schmezer P, Pool-Zobel BL. *In vivo* studies on genotoxicity of pure and commercial linuron. *Mutat Res* 1997;390:207–221.
85. Sasaki YF, Saga A, Akasaka M, Ishibashi S, Yoshida K, Su YQ, Matsusaka N, Tsuda S. Detection of *in vivo* genotoxicity of haloalkanes and haloalkenes carcinogenic to rodents by the alkaline single cell gel electrophoresis (comet) assay in multiple mouse organs. *Mutat Res* 1998;419:13–20.
86. Purschke M, Jacobi H, Witte I. Differences in genotoxicity of $H_2O_2$ and tetrachlorohydroquinone in human fibroblasts. *Mutat Res* 2002;513:159–167.
87. Sasaki YF, Sekihashi K, Izumiyama F, Nishidate E, Saga A, Ishida K, Tsuda S. The comet assay with multiple mouse organs: comparison of comet assay results and carcinogenicity with 208 chemicals selected from the IARC monographs and U.S. NTP Carcinogenicity Database. *Crit Rev Toxicol* 2000;30:629–799.
88. Sasaki YF, Kawaguchi S, Kamaya A, Ohshita M, Kabasawa K, Iwama K, Taniguchi K, Tsuda S. The comet assay with 8 mouse organs: results with 39 currently used food additives. *Mutat Res* 2002;519:103–119.
89. Phillips DH. Use of $^{32}$P-postlabelling to distinguish between genotoxic and nongenotoxic carcinogens. In Vainio HJ, Magee PN, McGregor DB, McMichael JA, Eds. *Mechanisms of Carcinogenesis in Risk Identification*. Lyon, France: IARC Scientific Publications No. 116; 1992, pp 211–224.
90. Beland FA, Poirier MC. DNA adducts and their consequences. In Tardiff RG, Lohmann PHM, Wogan GN, Eds. *Methods to Assess DNA Damage and Repair: Interspecies Comparisons*. Chichester, UK: John Wiley & Sons; 1994, pp 29–55.
91. Vodicka P, Bastlova T, Vodickova L, Peterkova K, Lambert B, Hemminki K. Biomarkers of styrene exposure in lamination workers: levels of *O*6-guanine DNA adducts, DNA strand breaks and mutant frequencies in the hypoxanthine guanine phosphoribosyltransferase gene in T-lymphocytes. *Carcinogenesis* 1995;16:1473–1482.
92. Bastlova T, Vodicka P, Peterkova K, Hemminki K, Lambert B. Styrene oxide-induced HPRT mutations, DNA adducts and DNA strand breaks in cultured human lymphocytes. *Carcinogenesis* 1995;16:2357–2362.
93. Weinstein IB. The origins of human cancer: molecular mechanisms of carcinogenesis and their implications for cancer prevention and treatment. *Cancer Res* 1988;48:4135–4143.
94. Perera FP, Whyatt RM. Biomarkers and molecular epidemiology in mutation and cancer research. *Mutat Res* 1994;313:117–129.

95. Beach AC, Gupta RC. Human biomonitoring and the $^{32}$P postlabeling assay. *Carcinogenesis* 1992;13:1053–1074.
96. Phillips DH, Farmer PB, Beland FA, Nath RG, Poirier MC, Reddy MV, Turteltaub KW. Methods of DNA adduct determination and their application to testing compounds for genotoxicity. *Environ Mol Mutagen* 2000;35:222–233.
97. Randerath K, Reddy MV, Gupta RC. $^{32}$P-labeling test for DNA damage. *Proc Natl Acad Sci USA* 1981;78:6126–6129.
98. Reddy MV, Randerath K. Nuclease P1-mediated enhancement of sensitivity of $^{32}$P-postlabelling test for structurally diverse DNA adducts. *Carcinogenesis* 1986;7:1543–1551.
99. Randerath K, Randerath E. $^{32}$P-postlabelling methods for DNA adduct detection: overview and critical evaluation. *Drug Metab Rev* 1994;26:67– 85.
100. Phillips DH. Detection of DNA modifications by the $^{32}$P-postlabelling assay. *Mutat Res* 1997;378:1–12.
101. Masento MS, Hewer A, Grover PL, Phillips DH. Enzyme-mediated phosphorylation of polycyclic hydrocarbon metabolites: detection of non-adduct compounds in the $^{32}$P-postlabelling assay. *Carcinogenesis* 1989;10:1557–1559.
102. Randerath K, Li D, Moorthy B, Randerath E. I-compounds: endogenous DNA markers of nutritional status, ageing, tumour promotion and carcinogenesis. In Phillips DH, Castegnaro M, Bartsch H, Eds. *Postlabelling Methods for Detection of DNA Damage*. Lyon, France: IARC; 1993, pp 157–165.
103. Sweetman GMA, Shuker DEG, Glover RP, Farmer PB. Mass spectrometry in carcinogenesis research. *Adv Mass Spectrom* 1998;14:343–376.
104. Turteltaub KW, Dingley KH. Application of accelerated mass spectrometry (AMS) in DNA adduct quantification and identification. *Toxicol Lett* 1998;102–103:435–439.
105. Turesky RJ, Vouros P. Formation and analysis of heterocyclic aromatic amine–DNA adducts *in vitro* and in vivo. *J Chromatogr B* 2004;802:155–166.
106. Andrews CL, Vouros P, Harsch A. Analysis of DNA adducts using high-performance separation techniques coupled to electrospray ionization mass spectrometry. *J Chromatogr A* 1999;856:515–526.
107. Paehler A, Richoz J, Soglia J, Vouros P, Turesky RJ. Analysis and quantification of DNA adducts of 2-amino-3,8-dimethylimidazo[4,5-*f*]quinoxaline in liver of rats by liquid chromatography/electrospray tandem mass spectrometry. *Chem Res Toxicol* 2002; 15:551–561.
108. *OECD Guideline 486. Guidelines for Testing of Chemicals: Unscheduled DNA Synthesis (UDS) Test with Mammalian Liver Cells in Vitro*. OECD; 1997.
109. Madle S, Dean SW, Andrae U, Brambilla G, Burlinson B, Doolittle DJ, Furihata C, Hertner T, McQueen CA, Mori H. Recommendations for the performance of UDS tests *in vitro* and *in vivo*. *Mutat Res* 1994;312:263–285.
110. Burlinson B, Morriss S, Gatehouse DG, Tweats DJ, Jackson MR. Uptake of tritiated thymidine by cells of the rat gastric mucosa after exposure to loxtidine or omeprazole. *Mutagenesis* 1991;6:11–18.
111. Glazer PM, Sarker SN, Summer WC. Detection and anaylsis of UV-induced mutations in mammalian cell DNA using a λ phage shuttle vector. *Proc Natl Acad Sci USA* 1986;83:1041–1044.
112. Malling HV, Burkhart JG. Use of ΦX17 as a shuttle vector for the study of *in vivo* mammalian mutagenesis. *Mutat Res* 1989;212:11–21.
113. Gossen JA, de Leeuw WJF, Tan CHT, Lohman PHM, Berends F, Knook DL, Zwarthoff EC, Vijg J. Efficient rescue of integrated shuttle vectors from transgenic mice: a model for studying mutations *in vivo*. *Proc Natl Acad Sci USA* 1989;86:7981–7985.

114. Kohler SW, Provost GS, Kretz PL, Dycaico MJ, Sorge JL, Short JM. Develepoment of a short-term, *in vivo* mutagenesis assay: the effects of methylation on the recovery of a lambda phage shuttle vector from transgenic mice. *Nucl Acids Res* 1990;18:3007–3013.

115. Kohler SW, Provost GS, Kretz PL, Fieck A, Sorge JA, Short JM. The use of transgenic mice for short-term, *in vivo* mutagenicity testing. *Genet Anal Tech Appl* 1990; 7:212–218.

116. Gossen JA, Vijg J. *Escherichia coli* C: a convenient host strain for rescue of highly methylated DNA. *Nucl Acids Res* 1988;16:9343.

117. Suzuki T, Hayashi M, Sofuni T, Myhr BC. The concomitant detection of gene mutations and micronucleus induction by mitomycin C *in vivo* using lacZ transgenic mice. *Mutat Res* 1993;285:219–224.

118. Tao KS, Urlando C, Heddle JA. Comparison of somatic mutation in a transgenic versus host locus. *Proc Natl Acad Sci USA* 1993;90:10681–10685.

119. Duncan BK, Miller JH. Mutagenic deamination of cytosine residues in DNA. *Nature* 1980;287:560–561.

120. Casentino L, Heddle JA. Differential mutation of transgenic and endogenous loci *in vivo*. *Mutat Res* 2000;454:1–10.

121. Lambert IB, Singer T, Boucher S, Douglas GR. Detailed review of transgenic rodent gene mutation assays. *Mutat Res Rev* 2005;290:1–280.

122. Heddle JA, Dean J, Nohmi T, Boerrigter M, Casciano D, Douglas GR, Glickman BW, Gorelick NJ, Mirsalis JC, Martus HJ, Skopek TR, Thybaud VR, Tindall KR, Yajima N. *In vivo* transgenic mutation assays. *Environ Mol Mutagen* 2000;35:253–259.

123. Thybaud VS, Dean S, Nohmi T, deBoer J, Douglas GR, Glickman BW, Gorelick NJ, Heddle JA, Heflich RH, Lambert I, Martus HJ, Mirsalis JC, Suzuki T, Yajima N. *In vivo* transgenic mutation assays. *Mutat Res* 2003;540:141–151.

124. Winegar RA, Carr G, Mirsalis JC. Analysis of the mutagenic potential of ENU and MMS in germ cells of male C57BL/6 lacI transgenic mice. *Mutat Res* 1997;388:175–178.

125. Swiger RR, Myhr B, Tucker JD. The LacZ transgene in MutaMouse maps to chromosome 3. *Mutat Res* 1994;325(4):145–148.

126. Gossen JA, Molijn AC, Douglas GR, Vijg J. Application of galactose-sensitive *E. coli* strains as selective hosts for lacZ- plasmids. *Nucl Acids Res* 1992;20:32–54.

127. Vijg J, Douglas GR. Bacteriophage lambda and plasmid lacZ transgenic mice for studying mutations *in vivo*. In Pfeifer GP, Ed. *Technologies for Detection of DNA Damage and Mutations*. New York: Plenum Press; 1996, pp 391–410.

128. Dycaico MJ, Provost GS, Kretz PL, Ransom SL, Moores JC, Short JM. The use of shuttle vectors for mutation analysis in transgenic mice and rats. *Mutat Res* 1994;307:271–281.

129. Jakubczak JL, Merlino G, French JE, Muller WJ, Paul B, Adhya S, Garges S. Analysis of genetic instability during mammary tumor progression using a novel selection-based assay for *in vivo* mutations in a bacteriophage lambda transgene target. *Proc Natl Acad Sci USA* 1996;93:9073–9078.

130. Swiger RR. Quantifying *in vivo* somatic mutation using transgenic mouse model systems. In Keohavong P, Grant SG, Eds. *Methods in Molecular Biology, Volume 291, Molecular Toxicology Protocols*. Totowa NJ: Humana Press; 2005, pp 145–153.

131. Heddle JA, Martus HJ, Douglas GR. Treatment and sampling protocols for transgenic mutation assays. *Environ Mol Mutagen* 2003;41:1–6.

132. Gossen JA, Martus HJ, Wei JY, Vijg J. Spontaneous and X-ray-induced deletion mutations in a LacZ plasmid-based transgenic mouse model. *Mutat Res* 1995;331:89–97.

133. Boerrigter METI, Dolle MET, Martus HJ, Gossen JA, Vijg J. Plasmid-based transgenic mouse model for studying *in vivo* mutations. *Nature* 1995;377:657–659.

134. Dolle MET, Giese H, Hopkins CL, Martus H-J, Hausdorff JM, Vijg J. Rapid accumulation of genome rearrangements in liver but not in brain of old mice. *Nat Genet* 1997;17:431–434.

135. Dolle ME, Snyder WK, van Orsouw NJ, Vijg J. Background mutations and polymorphisms in LacZ-plasmid transgenic mice. *Environ Mol Mutagen* 1999;34:112–120.

136. Nohmi T, Katoh M, Suzuki H, Matsui M, Yamada M, Watanabe M, Suzuki M, Horiya N, Ueda O, Shibuya T, Ikeda H, Sofuni T. A new transgenic mouse mutagenesis test system using spi- and 6-thioguanine selections. *Environ Mol Mutagen* 1996;28:465–470.

137. Ikeda H, Shimizu H, Ukita T, Kumagai M. A novel assay for illegitimate recombination in *Escherichia coli*: stimulation of lambda biotransducing phage formation by ultraviolet light and its independence from RecA function. *Adv Biophys* 1995;31:197–208.

138. Shimizu H, Yamaguchi H, Ikeda H. Molecular analysis of lambda biotransducing phage produced by oxolinic acid-induced illegitimate recombination *in vivo*. *Genetics* 1995;140:889–896.

139. Nohmi T, Masumura K. Molecular nature of intra-chromosomal deletions and base substitutions induced by environmental mutagens. *Environ Mol Mutagen* 2005; 45:150–161.

140. Hayashi H, Kondo H, Masumura K, Shindo Y, Nohmi T. Novel transgenic rat for *in vivo* genotoxicity assays using 6-thioguanine and spi-selection. *Environ Mol Mutagen* 2003;41:253–259.

141. Burkhart JG, Winn RN, VanBeneden R, Malling HV. The ΦX174 bacteriophage shuttle vector for the study of mutagenesis. *Environ Mol Mutagen* 1992;19(Suppl 20):8.

142. Gondo Y, Shioyama Y, Nakao K, Katsuki M. A novel positive detection system of *in vivo* mutation in rpsL (strA) transgenic mice. *Mutat Res* 1996;360:1–14.

143. Deprimo SE, Stambrook PJ, Stringer JR. Human placental alkaline phosphatase as a histochemical marker of gene expression in transgenic mice. *Transgen Res* 1996; 5:459–466.

144. Fahrig R. A mammalian spot test: induction of genetic alterations in pigment cells of mouse embryos with X-rays and chemical mutagens. *Mol Gen Genet* 1975; 138:309–314.

145. Styles JA, Penman MG. The mouse spot test. Evaluation of its performance in identifying chemical mutagens and carcinogens. *Mutat Res* 1985;154:183–204.

146. Fahrig R. The mammalian spot test (Fellfleckentest) with mice. *Arch Toxicol* 1977;38:87–98.

147. Russell LB. Validation of the *in vivo* somatic mutation method in the mouse as a prescreen for germinal point mutations. *Arch Toxicol* 1977;38:75–85.

148. Russell LB, Selby PB, von Halle E, Sheridan W, Valcovic L. Use of the mouse spot test in chemical mutagenesis: interpretation of past data and recommendations for future work. *Mutat Res* 1981;86:355–379.

149. *OECD Guideline 484. Guideline for the Testing of Chemicals. Genetic Toxicology: Mouse Spot Test*. OECD; 1986.

150. Suter W, Plappert-Helbig U, Glowienke S, Poetter-Locher F, Staedtler F, Racine R, Martus HJ. Induction of gene mutations by 5-(2-chloroethyl)-2′-deoxyuridine (CEDU), an antiviral pyrimidine nucleoside analogue. *Mutat Res* 2004;568:195–209.

151. Wahnschaffe U, Bitsch A, Kielhorn J, Mangelsdorf I. Mutagenicity testing with transgenic mice. Part II: Comparison with the mouse spot test. *J Carcinogen* 2005;4:4.

152. Shelby MD. Selecting chemicals and assays for assessing mammalian germ cell mutagenicity. *Mutat Res* 1996;352:159–169.
153. Douglas GR, Jiao J, Gingerich JD, Gossen JA, Soper LM. Temporal and molecular characteristics of mutations induced by ethylnitrosourea in germ cells isolated from seminiferous tubules and in spermatozoa of lacZ transgenic mice. *Proc Natl Acad Sci USA* 1995;92:7485–7489.
154. Sega GA. Unscheduled DNA synthesis in the germ cells of male mice exposed *in vivo* to the chemical mutagen ethyl methanesulfonate. *Proc Natl Acad Sci USA* 1974;71:4955–4959.
155. Sega GA. Unscheduled DNA synthesis (DNA repair) in the germ cells of male mice—its role in the study of mammalian mutagenesis. *Genetics* 1979;92:49–58.
156. Working PK, Butterworth BE. An assay to detect chemically induced DNA repair in rat spermatocytes. *Environ Mutagen* 1984;6:273–286.
157. Anderson D, Dobrzynka MM, Jackson LI, Yu TW, Brinkworth MH. Somatic and germ cell effects in rats and mice after treatment with 1,3-butadiene and its metabolites, 1,2-epoxybutene and 1,2,3,4-diepoxybutane. *Mutat Res* 1997;391:233–242.
158. Dobrzynska MM. The effects in mice of combined treatments to X-rays and antineoplastic drugs in the comet assay. *Toxicology* 2005;207:331–338.
159. Anderson D, Bishop JB, Garner RC, Ostrosky-Wegman P, Selby PB. Cyclophosphamide: review of its mutagenicity for an assessment of potential germ cell risks. *Mutat Res* 1995;330:115–181.
160. *OECD Guideline 483. Mammalian Spermatogonial Chromosome Aberration Test*. OECD; 1997.
161. Holmstrom LM, Palmer AK, Favor J. The rodent dominant lethal assay. In *Supplementary Mutagenicity Tests. UKEMS Report*. Cambridge, UK: Cambridge University Press; 1993, pp 129–156.
162. *OECD Guideline 478. Guideline for the Testing of chemicals. Genetic Toxicology: Rodent Dominant Lethal Test*. OECD; 1984.

# REPEAT DOSE TOXICITY STUDIES

Shayne Cox Gad

*Gad Consulting Services, Cary, North Carolina*

**Contents**

1 Introduction
2 Objectives
3 Regulatory Considerations
  3.1 Good Laboratory Practices (GLPs)
  3.2 Animal Welfare Act
  3.3 Regulatory Requirements for Study Design
4 Study Design and Conduct
  4.1 Animals
  4.2 Routes and Setting Doses
  4.3 Parameters to Measure
  4.4 Other In-Life Endpoints for Evaluation
  4.5 Histopathology
  4.6 Study Designs
5 Study Interpretation and Reporting
  References

## 1 INTRODUCTION

In the broadest sense, subchronic and chronic studies for pharmaceutical products can incorporate any of the routes used to administer a therapeutic agent, use any of a number of animal models, and conform to a broad range of experimental designs. They can be two weeks long (what used to be called "subacute" studies because they were conducted at dose levels below those employed for single dose or acute studies) or last up to a year. Another name for these studies is repeat dose studies [1, 2]—that is, those studies whereby animals have a therapeutic agent

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

administered to them on a regular and repeated basis by one or more routes over a period of one year or less. There is great flexibility and variability in the design of such studies.

The first thing to note about repeat dose studies and their design are the prinicples of dose response—a governing general principle of toxicology:

**The Three Dimensions of Dose Response**

As dose increases:

- Incidence of responders in an exposed population increases.
- Severity of response in affected individuals increases.
- Time to occurrence of response or of progressive stage of response decreases.
- An implication is that as duration of dosing increases, NOEL/NOAEL decreases.

This chapter seeks to provide a firm grasp of the objectives for repeat dose studies, the regulatory requirements governing them, the key factors in their design and conduct, and the interpretation of their results.

## 2 OBJECTIVES

As with any scientific study or experiment (but especially for those in safety assessment), the essential first step is to define and understand the reason(s) for the conduct of the study—that is, its objectives. There are three major (scientific) reasons for conducting repeat dose toxicity studies, but the basic characteristic and objective of all but a few repeat studies needs to be understood. The repeat dose study is (as are most other studies in whole animal toxicology) a broad screen. It is not focused on a specific endpoint; rather, it is a broad exploration of the cumulative biological effects of the administered agent over a range of doses: so broad an exploration, in fact, that it can be called a "shotgun" study.

The objectives of the typical repeat dose study fall into three categories. The first is to broadly define the toxicity (and, if one is wise, the pharmacology and hyperpharmacology) of repeated doses of a potential therapeutic agent in an animal model [3, 4, 5]. This definition is both qualitative (what are the target organs and the nature of the effects seen) and quantitative (at what dose levels, or, more importantly, at what plasma and tissue levels, are effects definitely seen and not seen). It seeks to characterize the cumulative toxicity of the drug under study. Such cumulative toxicity is rarely predicted by single dose toxicity.

The second objective (and the one that in the pharmaceutical industry usually compels both timing and compromising of design and execution) is to provide support for the initiation of and/or continued conduct of clinical trials in humans [6, 7]. As such, repeat dose studies should provide not only adequate clearance (therapeutic margin) of initial dose levels and duration of dosing, but also guidance for any special measures to be made or precautions to be taken in initial clinical trials. Setting inappropriate dose levels (either too low or too high) may lead to the failure of a study. A successful study must both define a safe or "clean" dose level

**TABLE 1 Duration of Treatment Supported by Preclinical Studies**

| Animal Study Length | Generally Allowed Human Dosing |
|---|---|
| 2 weeks | Up to 3 doses (but could be as much as 12 consecutive doses) |
| 1 month | Up to 30 days (though usually 28) |
| 3 months | 3 months |
| 6 months (rodent); 9–12 months (dog)(U.S.)/6 months (EEC and Japan) | Unlimited |

(one that is as high as possible, to allow as much flexibility as possible in the conduct of clinical studies) and demonstrate and/or characterize signs of toxicity at some higher dose. The duration of dosing issue is driven by a compromise between meeting established regulatory guidelines (as set out in Table 1) and the economic pressure to initiate clinical trials as soon as possible.

The third objective is one of looking forward to later studies. The repeat dose study must provide sufficient information to allow a prudent setting of doses for later, longer studies (including, ultimately, carcinogenicity studies). At the same time, the repeat dose study must also provide guidance for the other (than dose) design features of longer-term studies (such as what parameters to measure and when to measure them, how many animals to use, and how long to conduct the study).

These objectives are addressed by the usual repeat dose study. Some repeat dose studies, however, are unusual in being conceived, designed, and executed to address specific questions raised (or left unanswered) by previous preclinical or early clinical studies. Such a special purpose is addressed separately. Common study durations are 2, 4, and 13 weeks (all species), 26 weeks (rodent chronic toxicity), 39 or 52 weeks (nonrodent chronic toxicity), and 2 years (rodent carcinogenicity).

Chronic studies (those that last six or nine months or a year) may also be conducted for the above purposes but are primarily done to fulfill registration requirements for drugs that are intended for continuous long-term (lifetime) use or frequent intermittent use.

## 3 REGULATORY CONSIDERATIONS

Much of what is done (and how it is done) in repeat dose studies is a response to a number of regulations. Three of these have very broad impact. These are the Good Laboratory Practices requirements, Animal Welfare Act requirements, and regulatory requirements that actually govern study design.

### 3.1 Good Laboratory Practices (GLPs)

Since 1978, the design and conduct of preclinical safety assessment studies for pharmaceuticals in the United States (and, indeed, internationally) have been governed and significantly influenced by GLPs. Strictly speaking, these regulations cover qualifications of staff and facilities, training, record-keeping, documentation, and actions required to ensure compliance with and the effectiveness of these steps. Although the initial regulations were from the U.S. FDA [8], they have always

extended to cover studies performed overseas [9]. Most other countries have adopted similar regulations. A discussion of these regulations is beyond the scope of the current chapter, but several aspects are central to this effort. Each technique or methodology to be employed in a study (such as animal identification, weighing and examination, blood collection, and data recording) must be adequately described in a standard operating procedure (SOP) before the study begins. Those who are to perform such procedures must be trained in them beforehand. The actual design of the study, including start date and how it is to be ended and analyzed, plus the principal scientists involved (particularly the study director), must be specified in a protocol that is signed before the study commences. Any changes to these features must be documented in amendments once the study has begun. It is a good idea that the pathologist who is to later perform or oversee histopathology be designated before the start of the study, and that the design be a team effort involving the best efforts of the toxicologist, pathologist, and (usually for subchronic studies) the drug metabolism scientist [10].

### 3.2 Animal Welfare Act

Gone are the days when the pharmaceutical scientist could conduct whatever procedures or studies that were desired using experimental animals. The Animal Welfare Act [11] (and its analogues in other countries) rightfully requires careful consideration of animal usage to ensure that research and testing uses as few animals as possible in as humane a manner as possible. As a start, all protocols must be reviewed by an Institutional Animal Care and Use Committee. Such review takes time but should not serve to hinder good science. When designing a study or developing a new procedure or technique, the following points should be kept in mind:

1. Will the number of animals used be sufficient to provide the required data yet not constitute excessive use? (It ultimately does not reduce animal use to utilize too few animals to begin with and then have to repeat the study.)
2. Are the procedures employed the least invasive and traumatic available? This practice is not only required by regulations but is also sound scientific practice, since any induced stress will produce a range of responses in test animals that can mask or confound the chemically induced effects.

### 3.3 Regulatory Requirements for Study Design

The first consideration in the construction of a study is a clear statement of its objectives, which are almost always headed by meeting regulatory requirements to support drug development and registration. Accordingly, the relevant regulatory requirements must be analyzed, which is complicated by the fact that new drugs are no longer developed for registration and sale in a single-market country. The expense is too great, and the potential for broad international sales too appealing. While each major country has its own requirements as to study designs and studies required (with most of the smaller countries adhering to the regulations of one of the major players), harmonization has done much to smooth these differences [12]. Meeting these regulatory requirements is particularly challenging for several reasons. First, the only official delineation of general requirements in the Untied States is dated

[13], and recently special cases have arisen (anti-HIV agents, biotechnologically derived agents, therapeutic agents for neonates and the very elderly, etc.) that try the utility of these requirements. These needs have led to a stream of points-to-consider, which seek to update requirements. Second, the term "guidelines" means different things in different countries (in the United States it means "requirements," and in Japan, "suggestions").

Agents intended to treat or arrest the progress of rapidly spreading life-threatening diseases (such as AIDS) are subject to less stringent safety assessment requirements prior to initial clinical evaluations than are other drugs. However, even though approval (if clinical efficacy is established) for marketing can be granted with preclinical testing still under way, all applicable safety assessments (as with any other class of drugs) must still be completed [14].

Drugs intended for use in either the elderly or the very young have special additional requirements for safety evaluation, in recognition of the special characteristics and potential sensitivities of these populations. For the elderly, these requirements call for special consideration of renal and hepatic effects [15]. Likewise, drugs intended for the young require special studies to be performed in neonates and juvenile animals (usually of 2–4 weeks' duration in rats).

In the last five to six years, a number of potentially important drugs have been produced by recombinant DNA technology. These biomacromolecules, which are primarily endogenously occurring proteins, present a variety of special considerations and concerns, including the following:

- Because they are endogenously occurring molecules, assessing their pharmacokinetics and metabolism presents special problems.
- One must determine if the externally commercially produced molecule is biologically equivalent to the naturally occurring one.
- Since they are proteins, there is the question of whether they are immunogenic or provoke neutralizing antibodies that will limit their usefulness.
- Because they are available only in very small quantities, the use of traditional protocols (such as those that use ever-increasing doses until an adverse effect is achieved) is impractical.
- Agents with such specific activity in humans may not be appropriately evaluated in rodents or other model species.

Each of these points must be addressed in any safety testing plan [16]. The requirements set out in this chapter are designed to do this (for repeat dose testing).

## 4 STUDY DESIGN AND CONDUCT

### 4.1 Animals

In all but a few rare cases, for pharmaceutical safety assessment, separate studies in at least two species are required. Regulations require that both species be mammalian, and one of these must be a nonrodent; practice and economics dictate that the other species will be a rodent. With extremely rare exception, the rodent species employed is the rat (although the mouse also sees significant use). There is consider-

ably more variability in the nonrodent species, with a range of factors determining whether the dog (most common choice), a primate species (typically the rhesus or cynomolgus, although some others are used in particular cases), the pig (particularly in Europe), or some other animal (e.g., the ferret) is selected. The factors that should and do govern species selection are presented in detail in Gad and Chengelis [17]. The use of multiple species is a regulatory requirement arising from experience and the belief (going back to 1944 at least) that it will provide a better chance of detecting the full range of biological responses (adverse and otherwise) to the new molecular entity being evaluated. This belief has come under fire in recent years [18], but is unlikely to be changed soon. Along the same lines, unless an agent is to be used by only one sex or the other of humans, equal numbers of both sexes of an animal species are utilized in the studies, with the sexes being treated as unrelated for purposes of statistical analysis. Also except in rare cases, the animals used are young, healthy adults in the logarithmic phase of their growth curve. (The FDA specifies that rodents be less than 6 weeks of age at the initiation of dosing.)

Numbers of animals to be used in each dose group of a study are presented in Table 2. Although the usual practice is to use three different dose groups and at least one equal-sized control group, this number is not fixed and should be viewed as a minimum (see Section 4.6). Use of more groups allows for a reduction in the risk of not clearly defining effects and establishing the highest possible safe dose at a modest increase in cost. There must be as many control animals as are in the largest-size test group to optimize statistical power.

Animals are assigned to groups (test and control) by one or another form of statistical randomization. Prior to assignment, animals are evaluated for some period of time after being received in-house (usually at least 1 week for rodents and 2 weeks for nonrodents) to ensure that they are healthy and have no discernible abnormalities. The randomization is never pure; it is always "blocked" in some form or another (by initial body weight, at least) so that each group is not (statistically) significantly different from the others in terms of the "blocked" parameters.

Proper facilities and care for test animals are not only a matter of regulatory compliance (and a legal requirement), but also essential for a scientifically sound and valid study.

Husbandry requires clean cages of sufficient size and continuous availability of clean water and food (unless the protocol requires some restriction on their availability). Environmental conditions (temperature, humidity, and light–dark cycle) must be kept within specified limits. All of these must, in turn, be detailed in the

**TABLE 2 Numbers of Animals for Chronic and Subchronic Study per Test Group**

| Study Length | Rats per Sex | Dogs per Sex | Primates per Sex |
|---|---|---|---|
| 2–4 weeks | 5–10 | 3–4 | 3 |
| 3 months[a] | 20 | 6 | 5 |
| 6/9/12 months | 30 | 8 | 5 |
| Carcinogenicity (24 months) | 50 | 10 | 10 |

[a]Starting with 1 month studies, one should consider adding animals (particularly to the high dose and control groups) to allow evaluation of reversal (or progression) of effects. The FDA is becoming more invested in this approach—to evaluate both recovery from effects and possible progression after dosing.

protocols of studies. The limits for these conditions are set forth in relevant NIH and USDA publications.

### 4.2 Routes and Setting Doses

Route (how an agent is administered to a test animal) and dose (how much of and how frequently an agent is administered) are inseparable in pharmaceutical safety assessment studies and really cannot be defined independently. The selection of both begins with an understanding of the intended use of the drug in humans. The ideal case is to have the test material administered by the same route, at the same frequency (once a day, three times a day, etc.), and for the same intervals (continuously, if the drug is an intravenously infused agent, for example) as the drug's eventual use in people. Practical considerations such as the limitations of animal models (i.e., there are some things you can't get a rat to do), limitations on technical support,[1] and the like, and regulatory requirements (discussed below as part of dose setting) frequently act or interact to preclude this straightforward approach.

Almost 30 routes exist for administration of drugs to patients, but only a handful of these are commonly used in preclinical safety studies [19]. The most common deviation from what is to be done in clinical trials is the use of parenteral (injected) routes such as IV (intravenous) and SC (subcutaneous) deliveries. Such injections are loosely characterized as bolus (all at once or over a very short period, such as 5 minutes) and infusion (over a protracted period of hours, days, or even months). The term *continuous infusion* implies a steady rate over a protracted period, requiring some form of setup such as an implanted venous catheter or infusion port.

It is rare that the raw drug itself is suitable (in terms of stability, local tissue tolerance, and optimum systemic absorption and distribution) for direct use as a dosage form. Either it must be taken into a solution or suspension in a suitable carrier, or a more complex formulation (a prototype of the commercial form) must be developed. Gad and Chengelis [20] should be consulted for a more complete discussion of dose formulation for animals or humans. One formulation or more must be developed (preferably the same one for both animals and humans) based on the specific requirements of preclinical dosage formulation. For many therapeutic agents, limitations on volumes that can be administered and concentrations of active ingredient that can be achieved impact heavily on dose setting.

Setting of doses for longer-term toxicity studies is one of the most difficult tasks in study design. The doses administered must include one that is devoid of any adverse effect (preferably of *any* effect) and yet still high enough to "clear" the projected clinical dose by the traditional or regulatory safety factors. Such safety or uncertainty factors are currently predictions on the FDA HED guidance [21]. At the same time, if feasible, at least one of the doses should characterize the toxicity profile associated with the agent (for some biotechnologically derived agents, particularly those derived from endogenous human molecules, it may only be possible to demonstrate biological effects in appropriate disease models, and impossible to

[1] Many antiviral agents, particularly some anti-HIV agents, have rather short plasma half-lives, which requires frequent oral administration of the agent. Thirteen week studies have been conducted with tid (3x/day) dosing of rats and monkeys, requiring around-the-clock shift work for technical staff of the laboratory.

demonstrate toxicity). Because of limitations on availability of protodrugs, it is generally undesirable to go too high to achieve this second (toxicity) objective.

Traditionally, studies include three or more dose groups to fulfill these two objectives. Based on earlier results (generally single dose or two week studies), doses are selected. It is, by the way, generally an excellent idea to observe the "decade rule" in extrapolation of results from shorter to longer studies; that is, do not try to project doses for more than an order-of-magnitude-longer study (thus the traditional progression from single dose to 14 day to 90 day studies). Also, one should not allow the traditional use of three dose groups plus a control to limit designs. If there is a great deal of uncertainty, it is much cheaper in every way to use four or five dose groups in a single study than to have to repeat the entire study. Finally, remember that different doses may be appropriate for the different sexes.

It should also be kept in mind that formulating materials may have effects of their own, and a "vehicle" control group may be required in addition to a negative control group.

### 4.3 Parameters to Measure

As stated earlier, repeat dose studies are "shotgun" in nature; that is, they are designed to look at a very broad range of endpoints with the intention of screening as broadly as possible for indications of toxicity. Meaningful findings are rarely limited to a single endpoint—rather, what typically emerges is a pattern of findings. This broad search for components of toxicity profile is not just a response to regulatory guidelines intended to identify potentially unsafe drugs. An understanding of all the indicators of biological effect can also frequently help one to understand the relevance of findings, to establish some as unrepresentative of a risk to humans, and even to identify new therapeutic uses of an agent.

Parameters of interest in the repeat dose study can be considered as sets of measures, each with its own history, rationale, and requirements. It is critical to remember, however, that the strength of the study design as a scientific evaluation lies in the relationships and patterns of effects that are seen in looking at each of these measures (or groups) not as independent findings but rather as integrated profiles of biological effects.

***Body Weight*** Body weight (and the associated calculated parameter of body weight gain) is a nonspecific, broad screen for adverse systemic toxicity. Animals are initially assigned to groups based on a randomization scheme that includes each group varying insignificantly from one another in terms of body weight. Weights are measured prior to the initial dose, then typically 1, 3, 5, 7, 11, and 14 days thereafter. The frequency of measurement of weights goes down as the study proceeds; after 2 weeks, weighing is typically weekly through 6 weeks, then every other week through 3 months, and monthly thereafter. Because the animals used in these studies are young adults in the early log phase of their growth, decreases in the rate of gain relative to control animals is a very sensitive (albeit nonspecific) indicator of systemic toxicity [22].

***Food Consumption*** Food consumption is typically measured with one or two uses in mind. First, it may be explanatory in the interpretation of reductions

(either absolute or relative) in body weight. In cases where administration of the test compound is via diet, it is essential to be able to adjust dietary content so as to accurately maintain dose levels. Additionally, the actual parameter itself is a broad and nonspecific indicator of systemic toxicity. Food consumption is usually measured over a period of several days, first weekly and then on a once-a-month basis. Water consumption, which is also sometimes measured, is similar in interpretation and use [23, 24].

***Clinical Signs*** Clinical signs are generally vastly underrated in value, probably because insufficient attention is paid to care in their collection. Two separate levels of data collection are actually involved here. The first is the morbidity and mortality observation, which is made twice a day. This generally consists of a simple cage-side visual assessment of each animal to determine if it is still alive, and, if so, whether it appears in good (or at least stable) health. Historically, this regulatory required observation was intended to ensure that tissues from intoxicated animals were not lost for meaningful histopathologic evaluation due to autolysis [25].

The second level of clinical observation is the detailed hands-on examination analogous to the human physical examination. It is usually performed against a checklist (e.g., see Ref. 20) and evaluation is of the incidence of observations of a particular type in a group of treated animals compared to controls. Observations range from being indicative of nonspecific systemic toxicity to fairly specific indicators of target organ toxicity. These more detailed observations are typically taken after the first week of a study and on a monthly basis thereafter.

Ophthalmologic examinations are typically made immediately prior to initiation of a study (and thus serve to screen out animals with preexisting conditions) and toward the end of a study.

Particularly when the agent under investigation either targets or acts via a mechanism likely to have a primary effect on a certain organ for which functional measures are available, an extra set of measurements of functional performance should be considered. The organs or organ systems that are usually of particular concern are the kidneys, liver, and cardiovascular, nervous, and immune systems. Special measures (such as creatinine clearance as a measure of renal function) are combined with other data already collected (organ weights, histopathology, clinical pathology, etc.) to provide a focused "special" investigation or evaluation of adverse effects on the target organ system of concern. In larger animals (dogs and primates) some of these measures (such as ECGs) are made as a matter of course in all studies.

***Clinical Pathology*** Clinical pathology covers a number of biochemical and morphological evaluations based on invasive and noninvasive sampling of fluids from animals that are made periodically during the course of a subchronic study. These evaluations are sometimes labeled as clinical (as opposed to anatomical) pathology determinations. Table 3 presents a summary of the parameters measured under the headings of clinical chemistry, hematology, and urinalysis, using samples of blood and urine collected at predetermined intervals during the study. Conventionally, these intervals are typically at three points evenly spaced over the course of the study, with the first being 1 month after study initiation and the last being immediately prior to termination of the test animals. For a 3 month study, this means that

**TABLE 3 Clinical Pathology Measures**

| Clinical Chemistry | Hematology | Urinalysis |
|---|---|---|
| Albumin | Erythrocyte count (RBC) | Chloride |
| Alkaline phosphatase (ALP) | Hemoglobin (HGB) | Bilirubin |
| Blood urea nitrogen (BUN) | Hematocrit (HCT) | Glucose |
| Calcium | Mean corpuscular hemoglobin (MCH) | Occult blood |
| Chloride | Mean corpuscular volume (MCV) | pH |
| Creatine | Platelet count | Phosphorus |
| Creatine phosphokinase (CPK) | Prothrombin time | Potassium |
| Direct bilirubin | Reticulocyte count | Protein |
| γ-Glutamyltransferase (GGT) | White cell count (WBC) | Sodium |
| Globulin | White cell differential count | Specific gravity |
| Glucose | | Volume |
| Lactic dehydrogenase (LDH) | | |
| Phosphorus | | |
| Potassium | | |
| Serum glutamic-oxaloacetic trans-aminase (SGOT) | | |
| Serum glutamic-pyruvic transaminase (SGPT) | | |
| Sodium | | |
| Total bilirubin | | |
| Total cholesterol | | |
| Total protein | | |
| Triglycerides | | |

samples of blood and urine would be collected at 1, 2, and 3 months after study initiation (i.e., after the first day of dosing of the animals). There are some implications of these sampling plans that should be considered when the data are being interpreted. Many of the clinical chemistry (and some of the hematologic) markers are really the result of organ system damage that may be transient in nature (see Table 4 for a summary of interpretations of clinical chemistry findings and Table 5 for a similar summary for hematologic findings). The samples on which analysis is performed are from fixed points in time, which may miss transient changes (typically increases) in some enzyme levels.

***Pharmacokinetics and Metabolism*** Pharmaceutical subchronic toxicity studies are always accompanied by a parallel determination of the pharmacokinetics of the material of interest administered by the same route as that used in the safety study. This parallel determination consists of measuring plasma levels of the administered agent and its major metabolites either in animals that are part of the main study or in a separate set of animals (in parallel with the main study) that are dosed and evaluated to determine just these endpoints. The purpose of these determinations is both to allow a better interpretation of the findings of the study and to encourage the most accurate possible extrapolation to humans. The first data of interest are the absorption, distribution, and elimination of the test material, but a number of other types of information can also be collected [26, 27, 28]. For nonparenteral

**TABLE 4 Association of Changes in Biochemical Parameters with Actions at Particular Target Organs**

| | | | | | | *Organ System* | | | |
|---|---|---|---|---|---|---|---|---|---|
| Parameter | Blood | Heart | Lung | Kidney | Liver | Bone | Intestine | Pancreas | Notes |
| Albumin | | | | ↓ | ↓ | | | | Produced by the liver. Very significant reductions indicate extensive liver damage. |
| ALP (alkaline phosphatase) | | | | | ↑ | ↑ | ↑ | | Elevations usually associated with cholestasis. Bone alkaline phosphatase tends to be higher in young animals. |
| Bilirubin (total) | ↑ | | | | ↑ | | | | Usually elevated due to cholestasis either due to obstruction or hepatopathy. |
| BUN (blood urea nitrogen) | | | | ↑ | ↓ | | | | Estimates blood-filtering capacity of the kidneys. Doesn't become significantly elevated until kidney function is reduced 60–75%. |
| Calcium | | | | ↑ | | | | | Can be life threatening and result in acute death. |
| Cholinesterase | | | | ↑ | ↓ | | | | Found in plasma, brain, and RBCs. |
| CPK (creatine phosphokinase) | | ↑ | | | | | | | Most often elevated due to skeletal muscle damage but can also be produced by cardiac muscle damage. Can be more sensitive than histopathology. |
| Creatine | | | | ↑ | | | | | Also estimates blood-filtering capacity of kidney as BUN does. More specific than BUN. |

**TABLE 4** ***Continued***

| Parameter | Blood | Heart | Lung | Kidney | Liver | Bone | Intestine | Pancreas | Notes |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | *Organ System* | | | | |
| Glucose | | | | | | | | ↑ | Alterations other than those associated with stress are uncommon and reflect an effect on the pancreatic islets or anorexia. |
| GGT (γ-glutamyltransferase) | | | | | ↑ | | | | Elevated in cholestasis. This is microsomal enzyme and levels often increase in response to microsomal enzyme induction. |
| HBDH (hydroxybutyric dehydrogenase) | | ↑ | | | ↑ | | | | Most prominent in cardiac muscle tissue. |
| LDH (lactic dehydrogenase) | | ↑ | ↑ | ↑ | ↑ | | | | Increase usually due to skeletal muscle, cardiac muscle, and liver damage. Not very specific unless isozymes are evaluated. |
| Protein (total) | | | | ↑ | ↑ | | | | Absolute alterations are usually associated with decreased production (liver) or increased loss (kidney). |
| SGOT (serum glutamic-oxaloacetic trans-aminase); also called AST (asparate aminotransferase) | | ↑ | | ↑ | ↑ | | | ↑ | Present in skeletal muscle and heart and most commonly associated with damage to these. |
| SGPT (serum glutamic-pyruvic transaminase); also called ALT (alanine aminotransferase) | | | | | ↑ | | | | Evaluations usually associated with hepatic damage or disease. |
| SDH (sorbitol dehydrogenase) | | | | | ↑ or ↓ | | | | Liver enzyme that can be quite sensitive but is fairly unstable. Samples should be processed as soon as possible. |

**TABLE 5 Some Probable Conditions Affecting Hematological Changes**

| Parameter | Elevation | Depression |
|---|---|---|
| Red blood cells | 1. Vascular shock<br>2. Excessive diuresis<br>3. Chronic hypoxia<br>4. Hyperadrenocorticism | 1. Anemias<br>(a) Blood loss<br>(b) Hemolysis<br>(c) Low RBC production |
| Hematocrit | 1. Increased RBCs<br>2. Stress<br>3. Shock<br>(a) Trauma<br>(b) Surgery<br>4. Polycythemia | 1. Anemias<br>2. Pregnancy<br>3. Excessive hydration |
| Hemoglobin | 1. Polycythemia (increase in production of RBCs) | 1. Anemias<br>2. Lead poisoning |
| Mean cell volume | 1. Amemias<br>2. Vitamin $B_{12}$ deficiency | 1. Iron deficiency |
| Mean corpuscular hemoglobin | 1. Reticulocytosis | 1. Iron deficiency |
| White blood cells | 1. Bacterial infections<br>2. Bone marrow stimulation | 1. Bone marrow depression<br>2. Cancer chemotherapy<br>3. Chemical intoxication<br>4. Splenic disorders |
| Platelets | | 1. Bone marrow depression<br>2. Immune disorder |
| Neutrophilis | 1. Acute bacterial infections<br>2. Tissue necrosis<br>3. Strenuous exercise<br>4. Convulsions<br>5. Tachycardia<br>6. Acute hemorrhage | 1. Viral infections |
| Lymphocytes | 1. Leukemia<br>2. Malnutrition<br>3. Viral infections | |
| Monocytes | 1. Protozoal infections | |
| Eosinophils | 1. Allergy<br>2. Irradiation<br>3. Pernicious anemia<br>4. Parasitism | |
| Basophils | 1. Lead poisoning | |

routes it is essential to demonstrate that systemic absorption and distribution of the test material did occur; otherwise, it is open to question whether the potential safety of the agent in humans has been adequately addressed (not to mention the implication for potential human therapeutic efficacy).

Both clinical chemistry, pathology, and pharmacokinetics have in common the requirement to collect blood. There are species-specific limitations on how this may be done and how much can be collected without compromising the test animals. These are summarized in Tables 6 and 7.

**TABLE 6 Blood Sample Requirements**

| Species | Body Weight (kg) | Blood Volume (mL) | Sample / Time Point (mL) |
|---|---|---|---|
| Mouse | 0.03 | 2 | 0.075 |
| Rat | 0.3 | 20 | 1 |
| Dog | 12 | 1000 | 50 |
| Cynomolgus monkey | 3 | 200 | 10 |
| Rabbit | 3 | 200 | 10 |

**TABLE 7 Blood Sampling Techniques**

| Species | Techniques |
|---|---|
| Mouse | Orbital sinus, cardiac puncture |
| Rat | Sublingual, orbital sinus, jugular vein, cardiac puncture, tail vein |
| Dog | Jugular vein, cephalic vein |
| Cynomolgus monkey | Femoral vein/artery, cephalic vein |
| Rabbit | Jugular vein, marginal ear vein, cardiac puncture |

### 4.4 Other In-Life Endpoints for Evaluation

***Ophthalmology*** Ophthalmological examination of all animals in study (particularly nonrodents) should be performed both before study initiation and at the completion of the period for which the drug is administered. This should be performed by an experienced veterinary ophthalmologist.

***Cardiovascular Function*** Particularly in light of recent concerns with drug-induced arrhythmias, careful consideration must be given to incorporating adequate evaluation of drug-induced alterations on cardiovascular function. This is usually achieved by measuring blood pressure, heart rate, and an EKG prestudy and periodically during the course of the study (usually at least one intermediate period and at the end of the study) in the nonrodent species being employed.

***Neurotoxicology*** Table 8 presents the FDA's current draft criteria [3, 4] for endpoints to be incorporated in studies as a screen for neurotoxicity. In practice, a functional observation battery is employed at several endpoints (usually 1 and 3 months into the study) to fill these requirements.

***Immunotoxicology*** In response to concerns about potential effects of drugs on the immune system, the FDA [4] has proposed that a basic set of criteria (Table 9) be evaluated and considered in standard subchronic and chronic studies. It should be noted that most of these endpoints are already collected in traditional subchronic designs.

***Pharmacokinetics*** All subchronic and chronic toxicity studies now incoprorate (either in the study itself or in a parallel study) evaluation of the basic pharmacokinetics of a compound [29].

**TABLE 8 FDA Draft Criteria for a Neurotoxicity Screen as a Component of Short-Term and Subchronic Studies**

- Histopathological examination of tissues representative of the nervous system, including the brain, spinal cord, and peripheral nervous system.
- Quantitative observations and manipulative test to detect neurological, behavioral, and physiological dysfuntions. These may include:
  General appearance
  Body posture
  Incidence and severity of seizure
  Incidence and severity of tremor, paralysis, and other dysfunction
  Level of motor activity and arousal
  Level of reactivity to stimuli
  Motor coordination
  Strength
  Gait
  Sensorimotor response to primary sensory stimuli
  Excessive lacrimation or salivation
  Piloerection
  Diarrhea
  Ptosis
  Other signs of neurotoxicity deemed appropriate

**TABLE 9 FDA Draft Recommendations for Type I Immunotoxicity Test that Can Be Included in Repeat Dose Toxicity Studies**

Hematology
- White blood cell counts
- Differential white blood cell counts
- Lymphocytosis
- Lymphopenia
- Eosinophilia

Histopathology
- Lymphoid tissues
- Spleen
  Lymph nodes
  Thymus
  Peyer's patches in gut
  Bone marrow
- Cytology (if needed)[a]
  Prevalence of activated macrophages
  Tissue prevalence and location of lymphocytes
  Evidence of B-cell germinal centers
  Evidence of T-cell germinal centers
- Necrotic or proliferative changes in lymphoid tissues

Clinical chemistry
- Total serum production
- Albumin
- Albumin-to-globulin ratio
- Serum transaminases

[a]More comprehensive cytological evaluation of the tissues would not be done unless there is evidence of potential immunotoxicity from the preceding evaluations.

### 4.5 Histopathology

Histopathology is generally considered the single most significant portion of data to come out of a repeat dose toxicity study. It actually consists of three related sets of data (gross pathology observations, organ weights, and microscopic pathology) that are collected during the termination of the study animals. At the end of the study, a number of tissues are collected during termination of all surviving animals (test and control). Organ weight and terminal body weights are recorded at study termination, so that absolute and relative (to body weight) values can be statistically evaluated.

These tissues, along with the organs for which weights are determined, are listed in Table 10. All tissues collected are typically processed for microscopic observation, but only those from the high dose and control groups are necessarily evaluated microscopically. If a target organ is discovered in the high dose group, then successively lower dose groups are examined until a "clean" (devoid of effect) level is discovered [30].

In theory, all microscopic evaluations should be performed blind (without the pathologist knowing from which dose group a particular animal came), but this is difficult to do in practice and such an approach frequently degrades the quality of the evaluation. Like all the other portions of data in the study, proper evaluation benefits from having access to all data that addresses the relevance, severity, timing, and potential mechanisms of a specific toxicity. Blind examination is best applied in peer review or consultations on specific findings.

In addition to the "standard" set of tissues specified in Table 10, observations during the course of the study or in other previous studies may dictate that additional tissues be collected or special examinations (e.g., special stains, polarized light or electron microscopy, immunocytochemistry, or quantitative morphometry) be

**TABLE 10 Tissues for Histopathology**

| | |
|---|---|
| Adrenals[a] | Mainstream bronchi |
| Body and cervix | Major salivary glands |
| Brain, all three levels[a] | Mesenteric lymph nodes |
| Cervical lymph nodes | Ovaries and tubes |
| Cervical spinal cord | Pancreas |
| Duodenum | Pituitary |
| Esophagogastric junction | Prostate |
| Esophagus | Skeletal muscle from proximal hind limb |
| Eyes with optic nerves | Spleen[a] |
| Femur with marrow | Sternebrae with marrow |
| Heart | Stomach |
| Ileum | Testes with epididymides[a] |
| Kidneys[a] | Thymus and mediastinal contents[a] |
| Large bowel | Thyroid with parathyroid[a] |
| Larynx with thyroid and parathyroid | Trachea |
| Liver[a] | Urinary bladder |
| Lungs[a] | Uterus including horns |

[a]Organs to be weighed.

undertaken to evaluate the relevance of, or to understand the mechanisms underlying, certain observations.

Histopathology testing is a terminal procedure, and, therefore, sampling of any single animal is a one-time event (except in the case of a tissue collected by biopsy). Because it is a regulatory requirement that the tissues from a basic number of animals be examined at the stated end of the study, an assessment of effects at any other time course (most commonly, to investigate recovery from an effect found at study termination) requires that satellite groups of animals be incorporated into the study at start-up. Such animals are randomly assigned at the beginning of the study, and otherwise treated exactly the same as the equivalent treatment (or control) animals.

### 4.6 Study Designs

The traditional design for a repeat dose toxicity study is very straightforward. The appropriate number of animals of each sex are assigned to each of the designated dose and control groups. Unfortunately, this basic design is taken by many to be dogma, even when it does not suit the purposes of the investigator. There are many possible variations to study design, but four basic factors should be considered: controls, the use of interval and satellite groups, balanced and unbalanced designs, and staggered starts.

Classically, a single control group of the same size as each of the dose groups is incorporated into each study. Some studies incorporate two control groups (each the same size as the experimental groups) to guard against having a statistically significant effect due to one control group being abnormal for one or more parameters (a much more likely event when laboratory animals were less genetically homogeneous than they are now). The belief is that a "significant" finding that differs from one (but not both) of the concurrent control groups, and does not differ from historical control data, can be considered as not biologically significant. This is, however, an indefensible approach. Historical controls have value, but it is the concurrent control group(s) in a study that is of concern.

Interval or satellite groups have been discussed at two earlier points in this chapter. They allow measurement of termination parameters at intervals other than at termination of the study. They are also useful when the manipulation involved in making a measurement (such as the collection of an extensive blood sample), while not terminal, may compromise (relative to other animals) the subject animals. Another common use of such groups is to evaluate recovery from some observed effect at study termination.

Usually, each of the groups in a study is the same size, with each of the sexes being equally represented. The result is called a balanced design, with statistical power for detection of effects optimized for each of the treatment groups. If one knows little about the dose–toxicity profile, this is an entirely sound and rational approach. However, there are situations when one may wish to utilize an unbalanced design—that is, to have one or more dose groups larger than the others. This is usually the case when either greater sensitivity is desired (typically in a low dose group), or an unusual degree of attrition of test animals is expected (usually due to mortality in a high dose group), or as a guard against a single animal's idiopathic response being sufficient to cause "statistical significance."

As it is normal practice to have a balanced design, it is also traditional to initiate treatment of all animals at the same time. This may lead to problems at study termination, however. It is a very uncommon toxicology laboratory that can "bring a study down" on a single day. In fact, there are no labs that can collect blood and perform necropsies in a single day on even the 48–80 dogs involved in a study, much less the 160–400+ rats in the rodent version. Starting all animals in a study on the same day presents a number of less than desirable options. The first is to terminate as many animals as can be done each day, continuing to dose (and therefore further affect) the remaining test animals. Assuming that the animals are being terminated in a random, balanced manner, this means that the last animals terminated will have received from 3 to 10 additional days of treatment. At the least, this is likely to cause some variance inflation (and therefore both decrease the power of the study design and possibly confound interpretation). If the difference in the length of treatment of test animals is greater than 3% of the intended length of the study, one should consider alternative designs.

An alternative approach to study design that addresses this problem employs one of several forms of staggered starts. In these, distinct groups of animals have their dosing initiated at different times. The most meaningful form recognizes that the two sexes are in effect separate studies anyway (they are never compared statistically, with the treatment groups being compared only against the same-sex control group). Thus, if the termination procedure for one sex takes 3–5 days, then one sex should be initiated on dosing one week and the other on the following week. This maximizes the benefits of common logistical support (such as dose formulation) and reduces the impact of differential length of dosing on study outcome.

Another variation recognizes that any repeat dose study using both males and females is really two studies. The different sexes are considered separately and, except for ending up in a single report, are treated as separate studies. Accordingly, each sex can be started on a different day without increasing variability.

A further variation on this is to stagger the start-up either of different dose groups, or of the satellite and main study portions of dose groups. The former is to be avoided (it will completely confound study outcome), while the latter makes sense in some cases (pharmacokinetics and special measures) but not others (recovery and interval sacrifice).

## 5 STUDY INTERPRETATION AND REPORTING

For a successful repeat dose study, the bottom line is the clear demonstration of a no-effect level, characterization of a toxicity profile (providing guidance for any clinical studies), enough information on pharmacokinetics and metabolism to scale dosages to human applications, and at least a basic understanding of the mechanisms involved in any identified pathogenesis. The report that is produced as a result of the study should clearly communicate these points—along with the study design and experimental procedures, summarized data, and their statistical analysis—and it should be GLP compliant, suitable for FDA submission format.

Interpretation of the results of a study should be truly scientific and integrative. It is elementary to have the report state only each statistically and biologically significant finding in an orderly manner. The meaning and significance of each in rela-

tion to other findings, as well as the relevance to potential human effects, must be evaluated and addressed.

The author of the report should ensure that it is accurate and complete, but also that it clearly tells a story and concludes with the relevant (to clinical development) findings.

## REFERENCES


1. Ballantyne B. Repeated exposure toxicity. In Ballantyne B, Marrs T, Syversen T, Eds. *General and Applied Toxicology*. London: MacMillan; 2000, pp 55–66.
2. Wilson NH, Hardisty JF, Hayes JR. Short-term, subchronic and chronic toxicity studies. In Hayes AW, Ed. *Principles and Methods of Toxicicology*, 4th ed. Philadelphia: Taylor & Francis; 2001, pp 917–958.
3. FDA. *Toxicological Principles for the Safety Assessment of Direct Food Additives and Color Additives Used in Food*, Redbook II, p 86. Washington, DC: Center for Food Safety and Applied Nutrition, FDA; 1993.
4. FDA. *Toxicological Principles for the Safety of Food Ingredients*, Redbook 2000. Washington, DC: Center for Food Safety and Applied Nutrition, FDA; 2000.
5. OECD. *Guidance Notes for Analysis and Evaluation of Chronic Toxicity and Carcinogenicity Studies* 1998.
6. O'Grady J, Linet OI. *Early Phase Drug Evaluation In Man*. Boca Raton, FL: CRC Press; 1990.
7. Smith CG. *The Process of New Drug discovery and Development*. Boca Raton, FL: CRC Press; 1992.
8. Food and Drug Administration (FDA). Good Laboratory Practices for Nonclinical Laboratory Studies. *CFR* 21 Part 58, March 1983.
9. Food and Drug Administration (FDA). Good Laboratory Practices. *CFR* 21 Part 58, April 1988.
10. OECD. *Principles of Good Laboratory Practice* 1997.
11. Animal and Plant Health Inspection Service (APHIS). United States Department of Agriculture (31 August 1989). *Fed Reg* 1989;54(168):36112–36163.
12. Alder S, Zbinden G. *National and International Drug Safety Guidelines*. Zollikon, Switzerland: M.T.C. Verlag; 1988.
13. Food and Drug Administration (FDA). *FDA Introduction to Total Drug Quality*. Washington, DC: US Government Printing Office; 1971.
14. Food and Drug Administration (FDA). Investigational new drug, antibiotic, and biological drug product regulations, procedures for drugs intended to treat life-threatening and severely debilitating illness. *Fed Reg* 1988;53(204):41516–41524.
15. Center for Drug Evaluation and Research (CDER). *Guideline for the Study of Drugs Likely to Be Used in the Elderly*. FDA; November 1989.
16. Weissinger J. Nonclinical pharmacologic and toxicologic considerations for evaluating biologic products. *Regul Toxicol Pharmacol* 1989;10:255–263.
17. Gad SC, Chengelis CP. *Animal Models in Toxicology*. New York: Marcel Dekker; 1992.
18. Zbinden G. The concept of multispecies testing in industrial toxicology. *Regul Toxicol Pharmacol* 1993;17:84–94.
19. Gad SC. Routes in safety assessment studies. *J Am Coll Toxicol* 1994;13:1–17.

20. Gad SC, Chengelis CP. *Acute Toxicology*, 2nd ed. San Diego, CA: Academic Press; 1999.
21. CBER. *Estimating the Safe Starting Dose in Clinical Trials for Therpeutics in Adult Healthy Volunteers*. 2005. http://www.fda.gov/cber/gdlns/dose.htm.
22. Allaben WT, Hart RW. Nutrition and toxicity modulation: the impact of animal body weight on study outcome. *Int J Toxicol* 1998;17(Suppl 2):1–3.
23. CPMS/SWP/2877/00 *Note for Guidance on Carcinogenic Potential*.
24. Ellaben WT, Hart RW. Nutrition and toxicity modulation: the impact of animal body weight on study outcome. *Int J Toxicol* 1998;17(Suppl 2):1–3.
25. Arnold DL, Grice HC, Krawksi DR. *Handbook of In Vivo Toxicity Testing*. San Diego, CA: Academic Press; 1990.
26. Yacobi A, Skelly JP, Batra VK. *Toxicokinetics and New Drug Development*. New York: Pergamon Press; 1989.
27. Tse FLS, Jaffe JM. *Preclinical Drug Disposition*. New York: Marcel Dekker; 1991.
28. EMEA. *Guideline on the Evaluation of Control, Samples for Toxicokinetics Parameters in Toxicology Studies: Checking for Contamination with the Test Substance*. 2004.
29. CPMP/ICH/384/95; ICH S3A. *Toxicokinetics: A Guidance Assessing Systematic Exposure in Toxicology Studies*.
30. Haschek WM, Rousseaup CG. *Handbook of Toxicology Pathology*. San Diego, CA: Academic Press; 1991.

# IRRITATION AND LOCAL TISSUE TOLERANCE STUDIES IN PHARMACETICAL SAFETY ASSESSMENT

SHAYNE COX GAD

*Gad Consulting Services, Cary, North Carolina*

**Contents**

1 Introduction
2 Primary Dermal Irritation Test
3 Other Nonparenteral Route Irritation Tests
4 Factors Affecting Irritation Responses and Test Outcome
5 Problems in Testing (and Their Resolutions)
6 Ocular Irritation Testing
7 Vaginal Irritation
  7.1 Acute Primary Vaginal Irritation Study in the Female Rabbit
  7.2 Repeated Dose Vaginal Irritation Study in the Female Rabbit
  7.3 Repeated Dose Vaginal Irritation Study in the Ovariectomized Rat
8 Parenteral Irritation/Tolerance
  8.1 Parenteral Routes
  8.2 Bolus Versus Infusion
  8.3 Test Systems for Parenteral Irritation
  8.4 Acute Intramuscular Irritation in the Male Rabbit
  8.5 Acute Intravenous Irritation in the Male Rabbit
  8.6 Alternatives
9 Phototoxicity
  9.1 Theory and Mechanisms
  9.2 Factors Influencing Phototoxicity/Photosensitization
  9.3 Predictive Tests for Phototoxicity
  9.4 Pyrogenicity
  9.5 Blood Compatibility
References

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

## 1 INTRODUCTION

Both irritation and local tolerance studies assess the short-term hazard of pharmaceutical agents in the immediate region of their application or installation. In particular, these studies are generally performed to assess the potential for topically or parenterally administered drugs to cause damage at its site of application or administration. The point of administration is where tissues in patients or test animals' bodies experience the highest concentration exposure and therefore are subject to the greatest physiochemical, chemical, and pharmacologic challenge [1, 2].

Topical local tolerance effects are generally limited to irritation but may also include hematocompatability and pyrogenicity. Alhough this usually means dermal irritation, it can also be intracutaneous, mucosal, penile, perivascular, venous, vaginal, muscle, bladder, rectal, nasal, or ocular, depending on the route of drug administration (Table 1). All but ocular irritation use some version of a common subjective rating scale (see Table 2) to evaluate responses—the outcome of all of these tests, which primarily evaluate the response of the first region of tissue (which is exposed to the highest concentration) to an administered drug substance. In general, any factor that enhances absorption through this tissue is likely to decrease tissue tolerance.

**TABLE 1 Irritation Tests—Types**

| Intracutaneous | Venous |
|---|---|
| Dermal | Penile |
| Ocular | Rectal |
| Mucosal | Bladder |
| Vaginal | Perivascular |
| Muscle (injection site) | Nasal |

**TABLE 2 Evaluation of Local Tissue Reactions in Tissue Irritation Studies**

| Skin Reaction | Value |
|---|---|
| Erythema and eschar formation | |
| No erythema | 0 |
| Very slight erythema (barely perceptible) | 1 |
| Well-defined erythema | 2 |
| Moderate to severe erythema | 3 |
| Severe erythema (beet redness) to slight eschar formation (injuries in depth) | 4 |
| Necrosis (death of tissue) | +N |
| Eschar (sloughing or scab formation) | +E |
| Edema formation | |
| No edema | 0 |
| Very slight edema (barely perceptible) | 1 |
| Slight edema (edges of area welldefined by definite raising) | 2 |
| Moderate edema (raised approximately 1 mm) | 3 |
| Severe edema (raised more than 1 mm and extending beyond the area of exposure) | 4 |
| Total possible score for primary irritation | 8 |

For the skin, this scale is used in the primary dermal irritation test, which is performed for those agents that are to be administered to patients by application to the skin. As with all local tolerance tests, it is essential that the material be evaluated in "condition of use"—that is, in the final formulated form, applied to test animals in the same manner that the agent is to be used clinically.

## 2 PRIMARY DERMAL IRRITATION TEST

### A. Rabbit Screening Procedure

1. A group of at least 8–12 New Zealand white rabbits are screened for the study.
2. All rabbits selected for the study must be in good health; any rabbit exhibiting sniffles, hair loss, loose stools, or apparent weight loss is rejected and replaced.
3. One day (at least 18 h) prior to application of the test substance, each rabbit is prepared by clipping the hair from the back and sides using a small animal clipper. A size No. 10 blade is used to remove long hair and then a size No. 40 blade is used to remove the remaining hair.
4. Six animals with skin sites that are free from hyperemia or abrasion (due to shaving) are selected. Skin sites that are in the telogen phase (resting stage of hair growth) are used; those skin sites that are in the anagen phase (stage of active growth, indicated by the presence of a thick undercoat of hair) are not used.

### B. Study Procedure

1. As many as four areas of skin, two on each side of the rabbit's back, can be utilized for sites for administration.
2. Separate animals are not required for an untreated control group. Each animal serves as its own control.
3. Besides the test substance, a positive control substance (a known skin irritant—1% sodium lauryl sulfate in distilled water) and a negative control (untreated patch) are applied to the skin. When a vehicle is used for diluting, suspending, or moistening the test substance, a vehicle control patch is required—especially if the vehicle is known to cause any toxic dermal reactions or if there is insufficient information about the dermal effects of the vehicle.
4. The intact (free of abrasion) sites of administration are assigned a code number. Up to four sites can be used, as follows: (#1) test substance, (#2) negative control, (#3) positive control, and (#4) vehicle control (if required).
5. Application sites should be rotated from one animal to the next to ensure that the test substance and controls are applied to each position at least once.
6. Each test or control substance is held in place with a 1 in. × 1 in. 12-ply surgical gauze patch. The gauze patch is applied to the appropriate skin

site and secured with 1 in. wide strips of surgical tape at the four edges, leaving the center of the gauze patch nonoccluded.

7. If the test substance is a solid or a semisolid, a 0.5 g portion is weighed and placed on the gauze patch. The test substance patch is placed on the appropriate skin site and secured. The patch is subsequently moistened with 0.5 mL of physiological saline.
8. When the test substance is in flake, granule, powder, or other particulate form, the weight of the test substance that has a volume of 0.5 mL (after compacting as much as possible without crushing or altering the individual particles, such as by tapping the measuring container) is used whenever this volume is less than 0.5 g. When applying powders, granules, and the like, the gauze patch designated for the test sample is secured to the appropriate skin site with one of the four strips of the tape at the most ventral position of the animal. With one hand, the appropriate amount of sample measuring 0.5 mL is carefully poured from a glycine weighing paper onto the gauze patch that is held in a horizontal (level) position with the other hand. The patch containing the test sample is then carefully placed into position onto the skin and the remaining three edges are secured with tape. The patch is subsequently moistened with 0.5 mL of physiological saline.
9. If the test substance is a liquid, a patch is applied and secured to the appropriate skin site. A 1 mL tuberculin syringe is used to measure and apply 0.5 mL of test substance to the patch.
10. The negative control site is covered with an untreated 12-ply surgical gauze patch (1 in. × 1 in.).
11. The positive control substance and vehicle control substance are applied to a gauze patch in the same manner as a liquid test substance.
12. The entire trunk of the animal is covered with an impervious material (such as Saran Wrap) for a 24 h period of exposure. The Saran wrap is secured by wrapping several long strips of athletic adhesive tape around the trunk of the animal. The impervious material aids in maintaining the position of the patches and retards evaporation of volatile test substances.
13. An Elizabethan collar is fitted and fastened around the neck of each test animal. The collar remains in place for the 24 h exposure period. The collars are utilized to prevent removal of wrappings and patches by the animals, while allowing the animals food and water *ad libitum*.
14. The wrapping is removed at the end of the 24 h exposure period. The test substance skin site is wiped to remove any test substance still remaining. When colored test substances (such as dyes) are used, it may be necessary to wash the test substance from the test site with an appropriate solvent or vehicle (one that is suitable for the substance being tested). This is done to facilitate accurate evaluation for skin irritation.
15. Immediately after removal of the patches, each 1 in. × 1 in. test or control site is outlined with indelible marker by dotting each of the four corners. This procedure delineates the site for identification.

### C. Observations

1. Observations are made of the test and control skin sites 1 h after removal of the patches (25 h postinitiation of application). Erythema and edema are evaluated and scored on the basis of designated values presented earlier in Table 2.
2. Observations are again performed 48 and 72 h after application and scores are recorded.
3. If necrosis is present or the dermal reaction is unusual, the reaction should be described. Severe erythema should receive the maximum score (4), and +N should be used to designate the presence of necrosis and +E the presence of eschar.
4. When a test substance produces dermal irritation that persists 72 h postapplication, daily observations of test and control sites are continued on all animals until all irritation caused by the test substance resolves or until day 14 postapplication.

### D. Evaluation of Results

1. A *Subtotal Irritation Value* for erythema or eschar formation is determined for each rabbit by adding the values observed at 25, 48, and 72 h postapplication.
2. A *Subtotal Irritation Value* for edema formation is determined for each rabbit by adding the values observed at 25, 48, and 72 h postapplication.
3. A *Total Irritation Value* is calculated for each rabbit by adding the subtotal irritation value for erythema or eschar formation to the subtotal irritation value for edema formation.
4. The *Primary Dermal Irritation Index* is calculated for the test substance or control substance by dividing the sum of the Total Irritation Scores by the number of observations (3 days × 6 animals = 18 observations).
5. The categories of the Primary Dermal Irritation Index (PDII) are as follows (this categorization of dermal irritation is a modification of the original classification described by Draize et al. [3]):

| | |
|---|---|
| PDII = 0.0 | Nonirritant |
| > 0.0–0.5 | Negligible irritant |
| > 0.5–2.0 | Mild irritant |
| > 2.0–5.0 | Moderate irritant |
| > 5.0–8.0 | Severe irritant |

Other abnormalities, such as atonia or desquamation, should be noted and recorded.

## 3 OTHER NONPARENTERAL ROUTE IRRITATION TESTS

The design of vaginal, rectal, and nasal irritation studies is less formalized, but follows the same basic pattern as the primary dermal irritation test. The rabbit is the preferred species for vaginal and rectal irritation studies, but the monkey and

dog have also been used for these [4]. Both the rabbit and rat have commonly seen use for nasal irritation evaluations. Defined quantities (typically 1.0 mL) of test solutions or suspensions are instilled into the orifice in question. For the vagina or rectum, inert bungs are usually installed immediately thereafter to continue exposure for a defined period of time (usually the same period of hours as future human exposure). The orifice is then flushed clean, and 24 h after exposure it is examined and evaluated (graded) for irritation using the scale in Table 2.

## 4 FACTORS AFFECTING IRRITATION RESPONSES AND TEST OUTCOME

The results of local tissue irritation tests are subject to considerable variability due to relatively small differences in test design or technique. Weil and Scala [5] arranged and reported on the best known of several intralaboratory studies to clearly establish this fact. Although the methods presented earlier have proved to give reproducible results in the hands of the same technicians over a period of years [6] and contain some internal controls (the positive and vehicle controls in the PDI) against large variabilities in results or the occurrence of either false positives or negatives, it is still essential to be aware of those factors that may systematically alter test results. These factors are summarized next.

A. In general, any factor that increases absorption through the stratum corneum, mucous membrane, or other outer barrier layers will also increase the severity of an intrinsic response. Unless this factor mirrors potential exposure conditions, it may, in turn, adversely affect the relevance of test results.

B. The physical nature of solids must be carefully considered both before testing and in interpreting results. Shape (sharp edges), size (small particles may abrade the skin due to being rubbed back and forth under the occlusive wrap), and rigidity (stiff fibers or very hard particles will be physically irritating) of solids may all enhance an irritation response.

C. Solids frequently give different results when they are tested dry than if wetted for the test. As a general rule, solids are more irritating if moistened (going back to Item A, wetting is a factor that tends to enhance absorption). Care should also be taken as to moistening agent—some (few) batches of U.S. Pharmacopeia physiological saline (used to simulate sweat) have proved to be mildly irritating to the skin and mucous membrane on their own. Liquids other than water or saline should not be used.

D. If the treated region on potential human patients will be a compromised skin surface barrier (e.g., if it is cut or burned) some test animals should likewise have their application sites compromised. This procedure is based on the assumption that abraded skin is uniformly more sensitive to irritation. Experiments, however, have shown that this is not necessarily true; some materials produce more irritation on abraded skin, while others produce less [6, 7].

E. The degree of occlusion (in fact, the tightness of the wrap over the test site) also alters percutaneous absorption and therefore irritation. One important quality control issue in the laboratory is achieving a reproducible degree of occlusion in dermal wrappings.

F. Both the age of the test animal and the application site (saddle of the back vs. flank) can markedly alter test outcome. Both of these factors are also operative in humans, of course [8], but in dermal irritation tests, the objective is to remove all such sources of variability. In general, as an animal ages, sensitivity to irritation decreases. For the dermal test, the skin middle of the back (other than directly over the spine) tends to be thicker (and therefore less sensitive to irritations) than that on the flanks.

G. The sex of the test animals can also alter study results, because both regional skin thickness and surface blood flow vary between males and females.

H. Finally, the single most important (yet also most frequently overlooked) factor that influences the results and outcome of these (and in fact most) acute studies is the training of the staff. In determining how test materials are prepared and applied and in how results are "read" against a subjective scale, both accuracy and precision are extremely dependent on the technicians involved. To achieve the desired results, initial training must be careful and all-inclusive. As important, some form of regular refresher training must be exercised—particularly in the area of scoring results. Use of a set of color photographic standards as a training and reference tool is strongly recommended; such standards should clearly demonstrate each of the grades in the Draize dermal scale.

I. It should be recognized that the dermal irritancy test is designed with a bias to preclude false negatives and therefore tends to exaggerate results in relation to what would happen in humans. Findings of negligible irritancy (or even in the very low mild irritant range) should therefore be of no concern unless the product under test is to have large-scale and prolonged dermal contact.

## 5 PROBLEMS IN TESTING (AND THEIR RESOLUTIONS)

Some materials, by either their physicochemical or toxicological natures, generate difficulties in the performance and evaluation of dermal irritation tests. The most commonly encountered of these problems are presented next.

A. *Compound Volatility.* One is sometimes required or requested to evaluate the potential irritancy of a liquid that has a boiling point between room temperature and the body temperature of the test animal. As a result, the liquid portion of the material will evaporate off before the end of the testing period. There is no real way around the problem; one can only make clear in the report on the test that the traditional test requirements were not met, although an evaluation of potential irritant hazard was probably achieved (for the liquid phase would also have evaporated from a human that it was spilled on).

B. *Pigmented Material.* Some materials are strongly colored or discolor the skin at the application site. This makes the traditional scoring process difficult or impossible. One can try to remove the pigmentation with a solvent; if successful, the erythema can then be evaluated. If use of a solvent fails or is unacceptable, one can (wearing thin latex gloves) feel the skin to determine if there is warmth, swelling, and/or rigidity—all secondary indicators of the irritation response.

C. *Systemic Toxicity.* On rare occasions, the dermal irritation study is begun only to have the animals die very rapidly after test material is applied.

## 6 OCULAR IRRITATION TESTING

Ocular irritation is significantly different from the other local tissue irritation tests on a number of grounds. For the pharmaceutical industry, eye irritation testing is performed when the material is intended to be put into the eye as a means or route of application for ocular therapy. There are a number of special tests applicable to pharmaceuticals or medical devices that are beyond the scope of this volume, since they are not intended to assess potential acute effects or irritation. In general, however, it is desired that an eye irritation test that is utilized by this group be both sensitive and accurate in predicting the potential to cause irritation in humans. Failing to identify human ocular irritants (lack of sensitivity) is to be avoided, but of equal concern is the occurrence of false positives.

The primary eye irritation test was originally intended to predict the potential for a single splash of chemical into the eye of a human being to cause reversible and/or permanent damage. Since the introduction of the original Draize test 50 years ago [3], ocular irritation testing in rabbits has both developed and diverged. Indeed, clearly there is no longer a single test design that is used and different objectives are pursued by different groups using the same test. This lack of standardization has been recognized for some time and attempts have been made to address standardization of at least the methodological aspects of the test, if not the design aspects.

One widely used study design, which begins with a screening procedure as an attempt to avoid testing severe irritants or corrosives in animals, goes as follows:

### A. Test Article Screening Procedure

1. Each test substance will be screened in order to eliminate potentially corrosive or severely irritating materials from being studied for eye irritation in the rabbit.
2. If possible, the pH of the test substance will be measured.
3. A primary dermal irritation test will be performed prior to the study.
4. The test substance will not be studied for eye irritation if it is a strong acid (pH of 2.0 or less) or strong alkali (pH of 11.0 or greater), and/or if the test substance is a severe dermal irritant (with a PDII of 5–8) or causes corrosion of the skin.
5. If it is predicted that the test substance does not have the potential to be severely irritating or corrosive to the eye, continue to Section B, Rabbit Screening Procedure.

### B. Rabbit Screening Procedure

1. A group of at least 12 New Zealand white rabbits of either sex are screened for the study. The animals are removed from their cages and placed in rabbit restraints. Care should be taken to prevent mechanical damage to the eye during this procedure.
2. All rabbits selected for the study must be in good health; any rabbit exhibiting sniffles, hair loss, loose stools, or apparent weight loss is rejected and replaced.

3. One hour prior to installation of the test substance, both eyes of each rabbit are examined for signs of irritation and corneal defects with a hand-held slit lamp. All eyes are stained with 2.0% sodium fluorescein and examined to confirm the absence of corneal lesions. *Fluorescein Staining:* Cup the lower lid of the eye to be tested and instill one drop of a 2% (in water) sodium fluorescein solution onto the surface of the cornea. After 15 seconds, thoroughly rinse the eye with physiological saline. Examine the eye, employing a hand-held long-wave ultraviolet illuminator in a darkened room. Corneal lesions, if present, appear as bright yellowish-green fluorescent areas.
4. Only 9 of the 12 animals are selected for the study. The 9 rabbits must not show any signs of eye irritation and must show either a negative or minimum fluorescein reaction (due to normal epithelial desquamation).

## C. Study Procedure

1. At least 1 h after fluorescein staining, the test substance is placed in one eye of each animal by gently pulling the lower lid away from the eyeball to form a cup (conjunctival cul-de-sac) into which the test material is dropped. The upper and lower lids are then gently held together for 1 second to prevent immediate loss of material.
2. The other eye remains untreated and serves as a control.
3. For testing liquids, 0.01 mL of the test substance is used.
4. For solids or pastes, 100 mg of the test substance is used.
5. When the test substance is in flake, granular, powder, or other particulate form, the amount that has a volume of 0.01 mL (after gently compacting the particles by tapping the measuring container in a way that will not alter their individual form) is used whenever this volume weighs less than 10 mg.
6. For aerosol products, the eye should be held open and the substance administered in a single, 1 second burst at a distance of about 4 inches directly in front of the eye. The velocity of the ejected material should not traumatize the eye. The dose should be approximated by weighing the aerosol can before and after each treatment. For other liquids propelled under pressure, such as substances delivered by pump sprays, an aliquot of 0.01 mL should be collected and instilled in the eye as for liquids.
7. The treated eyes of six of the rabbits are not washed following the instillation of the test substance.
8. The treated eyes of the remaining three rabbits are irrigated for 1 minute with room temperature tap water, starting 20 seconds after instillation.
9. To prevent self-inflicted trauma by the animals immediately after instillation of the test substance, the animals are not immediately returned to their cages. After the test and control eyes are examined and graded at 1 h post-exposure, the animals are returned carefully to their respective cages.

## D. Observations

1. The eyes are observed for any immediate signs of discomfort after instilling the test substance. Blepharospasm and/or excessive tearing are indicative of

**TABLE 3 Scale of Weighted Scores for Grading the Severity of Ocular Lesions**[a]

| Reaction Criteria | Score |
|---|---|
| I. Cornea | |
| A. Opacity degree of density (area that is most dense is taken for reading) | |
| 1. Scattered or diffuse area, details of iris clearly visible | 1 |
| 2. Easily discernible translucent area, details of iris slightly obscured | 2 |
| 3. Opalescent areas, no details of iris visible, size of pupil barely discernible | 3 |
| B. Area of cornea involved | |
| 1. One-quarter (or less) but not zero | 1 |
| 2. Greater than one-quarter, less than one-half | 2 |
| 3. Greater than one-half, less than whole area | 3 |
| 4. Greater than three-quarters up to whole area | 4 |
| Scoring equals A × B × 5; total maximum = 80[b] | |
| II. Iris | |
| A. Values | |
| 1. Folds above normal, congestion, swelling, circumcorneal ingestion (any one or all of these or combination of any thereof), iris still reacting to light (sluggish reaction is possible) | 1 |
| 2. No reaction to light, hemorrhage, gross destruction (any one or all of these) | 2 |
| Scoring equals A × B (where B is the area of the iris involved, graded as "under cornea"); total maximum = 10 | |
| III. Conjunctivae | |
| A. Redness (refers to palpebral conjunctivae only) | |
| 1. Vessels definitely injected above normal | 1 |
| 2. More diffuse, deeper crimson red, individual vessels not easily discernible | 2 |
| 3. Diffuse beefy red | 3 |
| B. Chemosis | |
| 1. Any swelling above normal (includes nictitating membrane) | 1 |
| 2. Obvious swelling with partial eversion of the lids | 2 |
| 3. Swelling with lids about half closed | 3 |
| 4. Swelling with lids about half closed to completely closed | 4 |
| C. Discharge | |
| 1. Any amount different from normal (does not include small amount observed in inner canthus of normal animals) | 1 |
| 2. Discharge with moistening of the lids and hair just adjacent to the lids | 2 |
| 3. Discharge with moistening of the lids and considerable area around the eye | 3 |
| Scoring (A + B + C) × 2; total maximum = 20 | |

[a]The maximum total score is the sum of all scores obtained for the cornea, iris, and conjunctivae. [b]All A × B = Σ (1–3) × Σ (1–4) for six animals.

irritating sensations caused by the test substance, and their duration should be noted. Blepharospasm does not necessarily indicate that the eye will show signs of ocular irritation.

2. Grading and scoring or ocular irritation are performed in accordance with Table 3. The eyes are examined and grades of ocular reactions are recorded.
3. If signs or irritation persist at day 7, readings are continued on days 10 and 14 after exposure or until all signs of reversible toxicity are resolved.

4. In addition to the required observation of the cornea, iris, and conjunctiva, serious effects (such as pannus, rupture of the globe, or blistering of the conjunctivae) indicative of a corrosive action are reported.
5. Whether or not toxic effects are reversible depends on the nature, extent, and intensity of damage. Most lesions, if reversible, will heal or clear within 21 days. Therefore, if ocular irritation is present at the 14 day reading, a 21 day reading is required to determine whether the ocular damage is reversible or nonreversible.

## 7 VAGINAL IRRITATION

Few if any products are administered via the vagina that are intended for systemic absorption. Thus, this route has not been as widely studied and characterized as others. On the other hand, large numbers of different products (douches, spermicides, antiyeast agents, etc.) have been developed that require introduction into the vagina in order to assert their localized effects. Increased research into different birth control and antiviral prophylaxis will result in more vaginal products in the future. All these must be assessed for vaginal irritation potential, and this serves as an example of the other tissue tolerance issues.

Considerable research [4, 9] has indicated that the rabbit is the best species for assessing vaginal irritation. There are those investigators, however, who consider the rabbit too sensitive and recommend the use of ovariectomized rats. Ovariectomy results in a uniformly thin, uncornified epithelium, which is more responsive to localized effects. This model is used when the results from a study with rabbits are questionable [9]. The routine progression of studies consists of first doing an acute primary vaginal irritation study, then a 10 day repeated dose study in rats. These protocols are summarized next. Longer-term vaginal studies have been conducted in order to assess systemic toxicity of the active agents, when administered by these routes (while the intended effects may be local, one cannot assume that there will be no systemic exposure).

### 7.1 Acute Primary Vaginal Irritation Study in the Female Rabbit

1. *Overview of Study Design.* One group of six adult rabbits receive a single vaginal exposure and are observed for 3 days with periodic examination (1, 24, 48, and 72 hours postdosing) of the genitalia. Animals are then euthanized and the vagina is examined macroscopically.
2. *Administration.*
   (a) Route: The material is generally introduced directly into the vagina using a lubricated 18 French rubber catheter attached to a syringe for quantification and delivery of the test material. Gentle placement of the catheter is important because one needs to ensure complete delivery of the dose without mechanical trauma. For rabbits, the depth of insertion is about 7.5 cm, and the catheter should be marked to about that depth. After delivery is completed, the tube is slowly withdrawn. No attempt is made to close or seal the vaginal orifice. Alternative methods may be used to

administer more viscous materials. The most common is to backload a lubricated 1 mL tuberculin syringe, then warm the material to close to body temperature. The syringe is then inserted into the vagina and the dose administered by depressing the syringe plunger.

(b) Dosage: The test material should be one (concentration, vehicle, etc.) that is intended for human application.

(c) Frequency: Once.

(d) Duration: 1 day.

(e) Volume: 1 mL per rabbit.

3. *Test System.*

(a) Species, age, and weight range: Sexually mature New Zealand white rabbits are generally used, weighing between 2 and 5 kg. The weight is not as important as the fact that the animals need to be sexually mature.

(b) Selection: Animals should be multiparous and nonpregnant. Animals should be healthy and free of external genital lesions.

(c) Randomization: Because there is only one group of animals, randomization is not a critical issue.

4. *In-Life Observations.*

(a) Daily observations: At least once daily for clinical signs.

(b) Detailed physical examination: Once during the week prior to dosing.

(c) Body weight: Day of dosing.

(d) Vaginal irritation: Scored at 1, 24, 48, and 72 hours postdosing. Scoring criteria are shown in Table 4.

5. *Postmortem Procedures.* Rabbits are euthanized by lethal dose of a barbiturate soon after the last vaginal irritation scores are collected. The vagina is opened by longitudinal section and examined for evidence of mucosal damage such as erosion and localized hemorrhage. No other tissues are examined. No tissues are collected. After the macroscopic description of the vagina is recorded, the animal is discarded.

### 7.2 Repeated Dose Vaginal Irritation Study in the Female Rabbit

1. *Overview of Study Design.* Four groups of five adult rabbits each receive a single vaginal exposure daily for 10 days. The genitalia are examined daily. Animals are then euthanized and the vagina is examined macroscopically and microscopically.

2. *Administration.*

(a) Route: The test materials are introduced directly into the vagina using a lubricated 18 French rubber catheter, using the techniques described previously for acute studies.

(b) Dosage: The test material should be one (concentration, vehicle, etc.) that is intended for human application. There will also be a sham-negative control (catheter in place but nothing administered), a vehicle control, and a positive control (generally 2% nonoxynol-9).

(c) Frequency: Once daily.

**TABLE 4 Scoring Criteria for Vaginal Irritation**

| Value | Irritation |
|---|---|
| | *Erythema* |
| 0 | No erythema |
| 1 | Very slight erythema (barely perceptible) |
| 2 | Slight erythema (pale red in color) |
| 3 | Moderate to severe erythema (definite red in color) |
| 4 | Severe erythema (beet or crimson red) |
| | *Edema* |
| 0 | No edema |
| 1 | Very slight edema (barely perceptible) |
| 2 | Slight edema (edges of area well defined by definite raising) |
| 3 | Moderate edema (raised approximately 1 mm) |
| 4 | Severe edema (raised more than 1 mm and extending beyond area of exposure) |
| | *Discharge* |
| 0 | No discharge |
| 1 | Very slight discharge |
| 2 | Slight discharge |
| 3 | Moderate discharge |
| 4 | Severe discharge (moistening of considerable area around vagina) |

(d) Duration: 10 days.

(e) Volume: 1 mL per rabbit for each material.

3. *Test System.*

(a) Species, age and weight range: Sexually mature New Zealand white rabbits are generally used, weighing between 2 and 5 kg. The weight is not as important as the fact that the animals need to be sexually mature.

(b) Selection: Animals should be nulliparous and nonpregnant. Animals should be healthy and free of external genital lesions.

(c) Randomization: At least 24 animals should be on pretest. Randomization to treatment groups is best done using a computerized blocking by body weight method or a random number generation method.

4. *In-Life Observations.*

(a) Daily observations: At least once daily for clinical signs.

(b) Detailed physical examination: Once during the week prior to dosing and immediately prior to necropsy.

(c) Body weight: First, fifth, and last day of dosing.

(d) Vaginal irritation: Scored once daily. Scoring criteria are shown in Table 4.

5. *Postmortem Procedures.* Rabbits are euthanized by lethal dose of a barbiturate soon after the last vaginal irritation scores are collected. The vagina is isolated using standard prosection techniques and then opened by longitudinal section and examined for evidence of mucosal damage such as erosion and

**TABLE 5 Microscopic Scoring Procedure for Vaginal Sections**

| Section | Value |
|---|---|
| Epithelium | |
| Intact—normal | 0 |
| Cell degeneration or flattening of the epithelium | 1 |
| Metaplasia | 2 |
| Focal erosion | 3 |
| Erosion or ulceration, generalized | 4 |
| Leukocytes | |
| Minimal—<25 per high power field | 1 |
| Mild—25–50 per high power field | 2 |
| Moderate—50–100 per high power field | 3 |
| Marked—>100 per high power field | 4 |
| Injection | |
| Absent | 0 |
| Minimal | 1 |
| Mild | 2 |
| Moderate | 3 |
| Marked with disruption of vessels | 4 |
| Edema | |
| Absent | 0 |
| Minimal | 1 |
| Mild | 2 |
| Moderate | 3 |
| Marked | 4 |

*Source*: From Ref. 2.

localized hemorrhage. No other tissues are examined. The vagina and cervix are collected and fixed in 10% neutral buffered formalin. Standard hematoxylin/eosin stained, paraffin-embedded histologic glass slides are prepared by routing methods. Three levels of the vagina (low, middle, and upper) are examined and graded using the scoring system shown in Table 5. Each level is scored separately and an average is calculated. Irritation is rated as follows:

| Score | Rating |
|---|---|
| 0 | Nonirritating |
| 1–4 | Minimal irritation |
| 5–8 | Mild irritation |
| 9–11 | Moderate irritation |
| 12–16 | Marked irritation |

The score for each rabbit is then averaged and acceptability ratings are given as follows:

| Average Score | Acceptability Ratings |
|---|---|
| 0–8 | Acceptable |
| 9–10 | Marginal |
| 11 or greater | Unacceptable |

### 7.3 Repeated Dose Vaginal Irritation Study in the Ovariectomized Rat

This study is very similar in design to that described previously for rabbits, with the following (sometimes obvious) exceptions. Mature ovariectomized female rats can be obtained from a commercial breeder. A 15% surplus should be obtained. Ten animals per group should be used (40 total for the study). The vaginal catheter is placed to a depth of approximately 2.5 cm and the treatment volume should be 0.2 mL.

## 8 PARENTERAL IRRITATION/TOLERANCE

There are a number of special concerns about the safety of materials that are routinely injected (parenterally administered) into the body [10]. By definition, these concerns are all associated with materials that are the products of the pharmaceutical and (in some minor cases) medical device industries. Such parenteral routes include three major ones—IV (intravenous), IM (intramuscular), and SC (subcutaneous)—and a number of minor routes (such as intraarterial) that are not considered here.

These unusual concerns include irritation (vascular, muscular, or subcutaneous), pyrogenicity, blood compatibility, and sterility [11]. The background of each of these along with the underlying mechanisms and factors that influence the level of occurrence of such an effect are briefly discussed next.

***Irritation*** Tissue irritation upon injection, and the accompanying damage and pain, is a concern that must be addressed for the final formulation, which is to be either tested in humans or marketed, rather than for the active ingredient. This is because most irritation factors are either due to or influenced by aspects of formulation design (see Refs. 10 and 12 for more information on parenteral preparations). These factors are not independent of the route (IV, IM, or SC) that will be used and, in fact (as discussed later), are part of the basis for selecting between the various routes.

The lack of irritation and tissue damage at the injection site is sometimes called *tolerance*. Some of the factors that affect tolerance are not fully under the control of an investigation and are also unrelated to the material being injected. These include body movement, temperature, and animal age. Factors that can be controlled, but that are not inherent to the active ingredient, include solubility, tonicity, and pH. Finally, the active ingredient and vehicle can have inherent irritative effects and factors such as solubility (in the physiological milieu into which they are being injected), concentration, volume molecular size, and particle size must be considered. Gray [13] and Ballard [14] discuss these factors and the morphological consequences that may occur if they are not addressed.

***Pyrogenicity*** Pyrogenicity is the induction of a febrile (fever) response induced by the parenteral (usually IV or IM) administration of exogenous material. Pyrogenicity is usually associated with microbiological contamination of a final formulation, but it is now of increasing concern because of the growing interest in biosynthetically produced materials. Generally, ensuring sterility of product and process will guard against pyrogenicity for traditional pharmaceuticals. For biologically

produced products, the FDA has promulgated the general guideline that no more than 5.0 units of endotoxin may be present per milligram of drug substance [38].

***Blood Compatibility*** It is important that cellular components of the blood are not disrupted and that serum- or plasma-based responses are not triggered by parenteral administration. Therefore, two mechanisms must be assessed regarding the blood compatibility of component materials. These include the material's effect on cellular components that cause membrane destruction and hemolysis and the activation of the clotting mechanism resulting in thromboembolism.

Many of the nonactive, ingredient-related physicochemical factors that influence irritation (e.g., tonicity, pH, and particle size) also act to determine blood compatibility. But the chemical features of a drug entity itself—its molecular size and reactivity—can also be of primary importance.

***Sterility*** Sterility is largely a concern to be answered in the process of preparing a final clinical formulation, and it is not addressed in detail in this chapter. However, it should be clear that it is essential that no viable microorganisms are present in any material to be parenterally administered (except for vaccines).

### 8.1 Parenteral Routes

There are at least 13 different routes by which to inject material into the body, including the following: (1) intravenous, (2) subcutaneous, (3) intramuscular, (4) intraarterial, (5) intradermal, (6) intralesional, (7) epidural, (8) intrathecal, (9) intracisternal, (10) intracardiac, (11) intraventricular, (12) intraocular, and (13) intraperitoneal.

Only the first three are discussed in any detail here. Most of these routes of administration place a drug directly or indirectly into systemic circulation. There are a number of these routes, however, by which the drug exerts a local effect; in which case, most of the drug does not enter systemic circulation (e.g., intrathecal, intraventricular, intraocular, intracisternal). Certain routes of administration may exert both local and systemic effects, depending on the characteristics of the drug and excipients (e.g., subcutaneous) [15].

The choice of a particular parenteral route will depend on the required time of onset of action, the required site of action, and the characteristics of the fluid to be injected, among other factors.

### 8.2 Bolus Versus Infusion

Technically, for all the parenteral routes (but in practice only for the IV route), there are two options for injecting a material into the body. The bolus and infusion methods are differentiated on the single basis of rate of injection, but they actually differ on a wide range of characteristics.

The most commonly exercised option is the bolus, or "push," injection, in which the injection device (syringe or catheter) is appropriately entered into the vein and a defined volume of material is introduced through the device. The device is then

removed. In this operation, it is relatively easy to restrain an experimental animal and the stress on the animal is limited. Although the person doing the injection must be skilled, it takes only a short amount of time to become so. And the one variable to be controlled in determining dosage is the total volume of material injected (assuming dosing solutions have been properly prepared).

There are limitations and disadvantages to the bolus approach, however. Only a limited volume may be injected, which may prohibit the use of bolus when volumes to be introduced are high (e.g., due to low active compound solubility or a host of other reasons). Only two devices (syringe and catheter) are available for use in the bolus approach. If a multiple-day course of treatment is desired (say, every day for 15 days), separate injections must be made at discreet entry sites.

The infusion approach involves establishing a fixed entry point into the vein, then slowly passing the desired test material through that point over a period of time (30 minutes is about minimum, while continuous around-the-clock treatment is at least therapeutically possible). There are a number of devices available for establishing their entry point; catheter, vascular port [16], or osmotic pump [17]. Each of these must, in turn, be coupled with a device to deliver the dosing solution at a desired rate. The osmotic pump, which is implanted, is also its own delivery device. Other options are gravity driven "drips," hand–held syringes (not practical or accurate over any substantial period of time), or syringe pumps. Very large volumes can be introduced by fusion over a protracted period of time, and only a single site need be fitted with an entry device.

However, infusion also has its limitations. Skilled labor is required, and the setup must be monitored over the entire period of infusion. Larger animals must be restrained, while there are special devices that make this requirement unnecessary for smaller animals. Restraint and protracted manipulation are very stressful on animals. Over a period of time, one must regularly demonstrate patency of a device—that is, entry into the vascular system continues to exist. Finally, one is faced with having to control two variables in controlling the dose—total volume and rate.

When are the two approaches (bolus and infusion) interchangeable? And why select one over the other? The selection of infusion is usually limited to two reasons: (1) when a larger volume must be introduced than is practical in a bolus injection or (2) tolerance is insufficient if the dose is given all at once (i.e., an infusion will "clear" a higher daily dose than will a bolus injection). For safety studies, when a bolus can be used to clear a human infusion, dosing is a matter of judgment. If the planned clinical infusion will take less than a half hour, practicality dictates that the animal studies be accomplished by bolus. In other situations, pharmacokinetics (in particular, the half-life of the drug entity) should be considered in making the decision.

### 8.3 Test Systems for Parenteral Irritation

There are no regulatory guidelines or suggested test methods for evaluating agents for muscular or vascular irritation. Since such guidelines are lacking, but the evaluation is necessary, those responsible for these evaluations have tried to develop and employ the most scientifically valid procedures.

Hagan [18] first suggested a method for assessing IM irritation. His approach, however, did not include a grading system for evaluation of the irritation, and the method used the sacrospinalis muscles, which are somewhat difficult to dissect or repeatedly inject.

Shintani et al. [19] developed and proposed the methodology that currently seems to be more utilized. It uses the lateral vastus muscle and includes a methodology for evaluation, scoring, and grading of irritation. Additionally, Shintani and coworkers investigated the effects of several factors such as pH of the solution, drug concentration, volume of injection, the effect of repeated injections, and the time to maximum tissue response.

## 8.4 Acute Intramuscular Irritation in the Male Rabbit [20]

1. *Overview of Study Design.* Each rabbit is injected as follows:

| Site | Treatment |
|---|---|
| (musculus vastus lateralis) | (1.0 mL/site) |
| Left | (Test article) |
| Right | (Vehicle) |

Day 1: Injection of all treatment groups—9 rabbits.

Day 2: Sacrifice and evaluation: 24 h posttreatment group—3 rabbits.

Day 3: Sacrifice and evaluation: 48 h posttreatment group—3 rabbits.

Day 4: Sacrifice and evaluation: 72 h posttreatment group—3 rabbits.

2. *Administration.*
   (a) Route: The test article is injected into the musculus vastus lateralis of each rabbit.
   (b) Dose: The dose selected is chosen to evaluate the severity of irritation and represents a concentration that might be used clinically. This volume has been widely used in irritation testing.
   (c) Frequency: Once only.
   (d) Duration: 1 day
   (e) Volume: 1.0 mL per site.
3. *Test System.*
   (a) Species, age, and weight range: Male New Zealand white rabbits weighing 2–5 kg are used. The New Zealand white rabbit has been widely used in muscle irritation research for many years and is a reasonable sized, even-tempered animal that is well adapted to the laboratory environment.
   (b) Selection: Animals to be used in the study are selected on the basis of acceptable findings from physical examinations and body weights.
   (c) Randomization: Animals are ranked by body weight and assigned a number between one and three. The order of number assigned (e.g., 1-3-2)

**TABLE 6 Scoring of Muscle Injection Site Irritation**

| Reaction Criteria | Score |
|---|---|
| No discernible gross reaction | 0 |
| Slight hyperemia | 1 |
| Moderate hyperemia and discoloration | 2 |
| Distinct discoloration in comparison with the color of the surrounding area | 3 |
| Brown degeneration with small necrosis | 4 |
| Widespread necrosis with an appearance of "cooked meat" and occasionally an abscess involving the major portions of the muscle | 5 |

is chosen from a table of random numbers. Animals assigned number 1 are in the 24 h posttreatment group; those assigned number 2 are in the 48 h posttreatment group; and those assigned number 3 are in the 72 h posttreatment group.

4. *In-Life Observations.*
   (a) Daily observations: Once daily following dosing.
   (b) Physical examinations: Once within the 2 weeks before the first dosing day.
   (c) Body weight: Should be determined once before the start of the study.
   (d) Additional examinations may be done by the study director to elucidate any observed clinical signs.
5. *Postmortem Procedures.* Irritation is evaluated as follows. Three rabbits are sacrificed by a lethal dose of barbiturate at approximately 24, 48, or 72 h after dosing. The left and right lateral vastus muscules of each rabbit are excised. The lesions resulting from injection are scored for muscle irritation on a numerical scale of 0–5 as described in Table 6 [19]. The average score for the nine rabbits is calculated, and a category of irritancy assigned based on the following table:

| Average Score | Grade |
|---|---|
| 0.0–0.4 | None |
| 0.5–1.4 | Slight |
| 1.5–2.4 | Mild |
| 2.5–3.4 | Moderate |
| 3.5–4.4 | Marked |
| 4.5 or greater | Severe |

### 8.5 Acute Intravenous Irritation in the Male Rabbit

The design here is similar to the intramuscular assay, except that injections are made into the veins in specific muscle masses.

1. *Overview of Study Design.* Rabbits are injected as follows:

| Group | Number of Animals | Treatment Site | Evaluation |
|---|---|---|---|
| 1 | 2 | Musculus vastus lateralis (left) and cervicodorsal subcutis (left) | 24 h |
| | | Musculus vastus lateralis (right) and cervicodorsal subcutis (right) | 24 h |
| 2 | 2 | Musculus vastus lateralis (left) and cervicodorsal subcutis (left) | 72 h |
| | | Musculus vastus lateralis (right) and cervicodorsal subcutis (right) | 72 h |
| 3 | 2 | Auricular vein (left) | 24 and 72 h evaluations |
| | | Auricular vein (right) | |

Day 1: Injection of all groups (6 rabbits).

Day 2: Evaluation of Group 3 (2 rabbits). Sacrifice and evaluation of Group 1 (2 rabbits).

Day 4: Evaluation of Group 3 (2 rabbits). Sacrifice and evaluation of Group 2 (2 rabbits).

2. *Administration.*
   (a) Intramuscular: Musculus vastus lateralis.
   (b) Subcutaneous: Cervicodorsal subcutis.
   (c) Intravenous: Auricular vein.
   (d) Dose: The doses and concentration selected are chosen to evaluate the severity of irritation. The dose volumes have been widely used in irritation testing.
   (e) Frequency: Once only.
   (f) Duration: 1 day.
   (g) Volume: Musculus vastus lateralis and cervicodorsal subcutis: 1.0 mL per site; auricular vein: 0.5 mL per site.
3. *Test System.*
   (a) Species, age, and weight range: Male New Zealand white rabbits, weighing 2–5 kg, are used.
   (b) Selection: Animals to be used in the study are selected on the basis of acceptable findings from physical examinations.
   (c) Randomization: Animals are ranked by body weight and assigned a number between 1 and 3. The order of numbers assigned (e.g., 1-3-2) is chosen from a table of random numbers. Animals assigned number 1 are in Group 1; those assigned number 2 are in Group 2; and those assigned number 3 are in Group 3.

**TABLE 7 Muscle Irritation Scoring**

| Reaction Criteria | Score |
|---|---|
| No discernible gross reaction | 0 |
| Slight hyperemia and discoloration | 1 |
| Moderate hyperemia and discoloration | 2 |
| Distinct discoloration in comparison with the color of the surrounding area | 3 |
| Small areas of necrosis | 4 |
| Widespread necrosis, possibly involving the underlying muscle | 5 |

4. *In-Life Observations.*
   (a) Daily observations: Once daily following dosing.
   (b) Physical examinations: Once within the 2 weeks before the first dosing day.
   (c) Body weight: Determined once before the start of the study.
   (d) Additional examinations may be done by the study director to elucidate any observed clinical signs.
5. *Postmortem Procedures.*
   (a) Intramuscular irritation is evaluated as follows. Rabbits are sacrificed by lethal dose of barbiturate approximately 24 and 72 h after dosing. The left and right lateral vastus muscles of each rabbit are excised. The reaction resulting from injection is scored for muscle irritation using the scale shown in Table 6.
   (b) Subcutaneous irritation is evaluated as follows. Rabbits are scarified by a lethal dose of barbiturate approximately 24 and 72 h after dosing. The subcutaneous injection sites are exposed by dissection, and the reaction is scored for irritation on a scale of 0–5 as in Table 7 [19].

| Average Score per Site | Irritancy Grade |
|---|---|
| 0.0–0.4 | None |
| 0.5–1.4 | Slight |
| 1.5–2.4 | Mild |
| 2.5–3.4 | Moderate |
| 3.5–4.4 | Marked |
| 4.5 or greater | Severe |

   (c) Intravenous irritation is evaluated as follows. Rabbits are sacrificed by a lethal dose of barbiturate following the 72 h irritation evaluation. The injection site and surrounding tissue are grossly evaluated at approximately 24 and 72 h after dosing on a scale of 0–3 as in Table 8.

| Average Score per Site | Irritancy Grade |
|---|---|
| 0.0 to 0.4 | None |
| 0.5 to 1.4 | Slight |
| 1.5 to 2.4 | Moderate |
| 2.5 or greater | Severe |

   (d) Additional examinations may be done by the study director to elucidate the nature of any observed tissue change.

**TABLE 8 Venous Irritation Scoring**

| Reaction Criteria | Score |
|---|---|
| No discernible gross reaction | 0 |
| Slight erythema at injection site | 1 |
| Moderate erythema and swelling with some discoloration of the vein and surrounding tissue | 2 |
| Severe discoloration and swelling of the vein and surrounding tissue with partial or total occlusion of the vein | 3 |

### 8.6 Alternatives

Intramuscular (IM) and intravenous (IV) injection of parenteral formulations of pharmaceuticals can produce a range of discomfort including pain, irritation, and/or damage to muscular or vascular tissue. These are normally evaluated for prospective formulations before use in humans by histopathologic evaluation of damage in intact animal models, usually the rabbit. Attempts have been made to make this *in vivo* methodology both more objective and quantitative based on measuring the creatine phosphokinase released in the tissue surrounding the injection site [21]. Currently, a protocol utilizing a cultured skeletal muscle cell line (L6) from the rat as a model has been evaluated in an interlaboratory validation program among 11 pharmaceutical laboratories. This methodology [22] measures creatine kinase levels in media after exposure of the cells to the formulation of interest and predicts *in vivo* IM damage based on this endpoint. It is reported to give excellent rank-correlated results across a range of antibiotics [23]. The current multilaboratory evaluation covers a broader structural range of compounds and has shown a good quantitative correlation with *in vivo* results for antibiotics and a fair correlation for a broader range of parenteral drug products. Likewise, Kato et al. [24] have proposed a model that uses cultured primary skeletal muscle fibers from the rat. Damage is evaluated by the release of creatine phosphokinase. An evaluation using six parenterally administered antibiotics (ranking their $EC_{50}$ values) showed good relative correlation with *in vivo* results.

Another proposed *in vitro* assay for muscle irritancy for injectable formulations is the red blood cell hemolysis assay [25]. Water-soluble formulations in a 1:2 ratio with freshly collected human blood are gently mixed for 5 min. The percentage of red blood cell survival is then determined by measuring differential absorbance at 540 nm; this value is then compared to values for known irritants and nonirritants. Against a very small group of compounds (four), this assay reportedly accurately predicts muscle irritation.

## 9 PHOTOTOXICITY

The potential for sunlight (or selected other light frequencies) to transform a drug or product is both a useful tool for activating some drugs and a cause of significant adverse effects for others (such as the quinolone antibiotics [26, 27]).

### 9.1 Theory and Mechanisms

The portion of the solar spectrum containing the biologically most active region is from 290 to 700 mm.

The ultraviolet (UV) part of the spectrum includes wavelengths from 200 to 400 nm. Portions of the UV spectrum have distinctive features from both the physical and biological points of view. The accepted designations for the biologically important parts of the UV spectrum are UVA, 400–315 nm; UVB, 315–280 nm; and UVC, 280–220 nm. Wavelengths less than 290 nm (UVC) do not occur at the earth's surface because they are absorbed, predominantly by ozone in the stratosphere. The most thoroughly studied photobiological reactions that occur in skin are induced by UVB. The quinolones, for example, absorb light strongly in the 300–400 nm wavelength range. Although UVB wavelengths represent only approximately 1.5% of the solar energy received at the earth's surface, they elicit most of the known chemical phototoxic and photoallergic reactions. The visible portions of the spectrum, representing about 50% of the sun's energy received at sea level, includes wavelengths from 400 to 700 nm. Visible light is necessary for such biological events as photosynthesis, the regulation circadian cycles, vision, and pigment darkening. Furthermore, visible light in conjunction with certain chromophores (e.g., dyes, drugs, and endogenous compounds that absorb light and therefore "give" color) and molecular oxygen induces photodynamic effects.

Understanding the toxic effects of light impinging on the skin requires knowledge of both the nature of sunlight and the skin's optical properties. Skin may be viewed as an optically heterogeneous medium, composed of three layers that have distinct refractive indices, chromophore distributions, and light-scattering properties. Light entering the outermost layer, the stratum corneum, is in part reflected—4–7% for wavelengths between 250 and 300 nm [28]—due to the difference in refractive index between air and the stratum corneum. Absorption by urocanic acid (a deamination product of histidine), melanin, and proteins containing the aromatic amino acids tryptophan and tyrosine in the stratum corneum produces further attenuation of light, particularly at shorter UV wavelengths. Approximately 40% of UVB is transmitted through the stratum corneum to the viable epidermis. The light entering the epidermis is attenuated by scattering and, predominantly, absorption. Epidermal chromophores consist of proteins, urocanic acid, nucleic acids, and melanin. Passage through the epidermis results in appreciable attenuation of UVA and UVB radiation. The transmission properties of the dermis are largely due to scattering, with significant absorption of visible light by melanin, β-carotene, and the blood-borne pigments bilirubin, hemoglobin, and oxyhemoglobin. Light traversing these layers of the skin is extensively attenuated, most drastically for wavelengths less than 400 nm. Longer wavelengths are more penetrating. It has been noted that there is an "optical window"—that is, greater transmission—for light at wavelengths of 600–1300 nm, which may have important biological consequences.

Normal variations in the skin's optical properties frequently occur. The degree of pigmentation may produce variations in the attenuation of light, particularly between 300 and 400 nm, by as much as 1.5 times more in blacks than in Caucasians [29]. Alterations in the amount or distribution of other natural chromophores account for further variations in skin optical properties. Urocanic acid, deposited on the skin's surface during perspiration [28], and UV-absorbing lipids, excreted in

sebum, may significantly reduce UV transmission through the skin. Epidermal thickness, which varies over regions of the body and increases after exposure to UVB radiation, may significantly modify UV transmission.

Certain disease states also produce alterations in the skin's optical properties. Alterations of the skin's surface, such as by psoriatic plaques, decrease transmitted light. The effect may be lessened by application of oils whose refractive index is similar to that of skin [28]. Disorders such as hyperbilirubinemia, porphyrias, and blue skin nevi result in increased absorption of visible light due to accumulation or altered distribution of endogenous chromophoric compounds.

The penetration of light into and through dermal tissues has important consequences. This penetration is demonstrated in Fig. 1. Skin, as the primary organ responsible for thermal regulation, is overperfused relative to its metabolic requirements [28].

It is estimated that the average cutaneous blood flow is 20–30 times that necessary to support the skin's metabolic needs. The papillary boundaries between epidermis and dermis allow capillary vessels to lie close to the skin's surface, permitting the blood and important components of the immune system to be exposed to light. The equivalent of the entire blood volume of an adult may pass through the skin, and potentially be irradiated, in 20 min. This corresponds to the time required to receive 1 or 2 MEDs (the MED is defined as the minimal dose of UV irradiation that produces definite, but minimally perceptible, redness 24 h after exposure). The accessibility of incident radiation to blood has been exploited in such regimens as phototherapy of hyperbilirubinemia in neonates, where light is used as a therapeutic agent. However, in general, there is a potential for light-induced toxicity due to irradiation of blood-borne drugs and metabolites.

### 9.2 Factors Influencing Phototoxicity/Photosensitization

There are a number of factors that can influence an agent acting either as a phototoxin or a photoallergen.

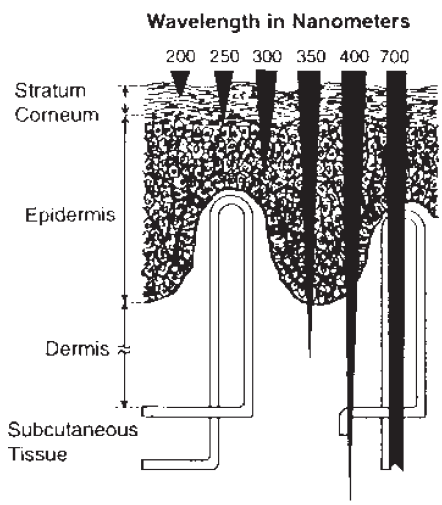


**FIGURE 1** Schematic penetration of light of varying wavelengths into the skin.

1. The quantity and location of photoactive material present in or on the skin.
2. The capacity of the photoactive material to penetrate into normal skin by percutaneous absorption as well as into skin altered by trauma, such as maceration, irritation, and sunburn.
3. The pH, enzyme presence, and solubility conditions at the site of exposure.
4. The quantity of activating radiation to which the skin is exposed.
5. The capacity of the spectral range to activate the materials on or within the skin.
6. The ambient temperature and humidity.
7. The thickness of the horny layer.
8. The degree of melanin pigmentation of the skin.
9. The inherent "photoactivity" of the chemical. Does it weakly or strongly absorb light?

Basically, any material that has both the potential to absorb ultraviolet light (in the UVA or UVB regions) and the possibility of dermal exposure or distribution into the dermal region should be subject to some degree of suspicion as to potential phototoxicity. As shown in Table 9, a large number of agents have been identified as phototoxic or photoallergenic agents. Of these, tetrachlorosalicylanilide (TCSA) is the most commonly used as a positive control in animal studies.

### 9.3 Predictive Tests for Phototoxicity

Before we start on our description of the different methods, we first cover some basics on light dosimetry. The intensity of the irradiation used in phototoxicity testing is determined with a light meter, which provides output as watts per meter$^2$. The shelves on which the animals rest during the exposure periods are normally adjustable in order to control the dose of light to the exposure area. The irradiation from fluorescent lights will vary somewhat from day to day, depending on temperature, variations in line current, and so on. The dose the animals receive is generally represented as joules/centimeter$^2$. A joule is equal to 1 watt/second. Therefore, the dose of light is dependent on the time of exposure. For example, in their review, Lambert et al. [27] discuss dosages of light of 9 or 10 J/cm$^2$ in the UVA spectral region. If the irradiation from the light is found to be 20 W/m$^2$ at the exposure site, then the time of exposure required to obtain the target dose of light (in joules) is calculated as

$$\text{Time of exposure} = \frac{\text{W}\cdot\text{s}}{\text{J}}\,\frac{9\,\text{J}}{\text{cm}^2}\,\frac{\text{m}^2}{20\,\text{W}}\,\frac{10^4\,\text{cm}^2}{\text{m}^2}\,\frac{\text{min}}{60\,\text{s}} = 75\,\text{min}$$

If, with the same set of lights, 2 weeks later the irradiation is determined to be 19 W/m$^2$, then the exposure period would have to be 79 min.

There are three recommended protocols for assessing topical phototoxicity potential—rabbit, guinea pig, and mouse. Only the rabbit and guineapig are described.

**TABLE 9 Known Phototoxic Agents**

| In Humans | | In Animals | |
|---|---|---|---|
| Compounds | Route | Compound | Route |
| Aminobenzoic acid derivatives | Topical | Acradine | Topical |
| | | Amiodarone | Oral |
| Amyldimethylamino benzoate, mixed *ortho* and *para* isomers | Topical | Anthracine | Topical |
| | | Bergapten (5-methoxypsolaren) | Topical |
| Anthracene acridine | Topical | Bithionol | Topical |
| Bergapten (5-methoxypsoralen) | Topical | Chlorodiazepoxide | Intraperitoneal |
| Cadmium sulfide | Tattoo | Chlorprothiazide | Intraperitoneal |
| Chlorothiazides | Oral | Chlorpromazine | Topical |
| Coal tar (multicomponent) | Topical | Demeclocycline | Intraperitoneal |
| Dacarbazine | Infusion | Griseofulvin | Intraperitoneal |
| Disperse blue 35 (anthaquinone-based dye) | Topical | Kynuremic acid | Oral |
| Nalidixic acid | Oral | Nalidixic acid | Oral |
| Padimate A or Escolol 506 (amyl-*p*-dimethylamino benzoate) | Topical | Prochlorperazine | Intraperitoneal |
| | | Quinokine methanol | Intraperitoneal |
| | | Quinolone (antibacterial) | Oral |
| Psoralens | Oral, topical | Tetracyclines | Intraperitoneal, topical |
| Quinolone (antibacterial) | Oral | Xanthotoxin (8-methoxypsoralen) | Oral, intraperitoneal, intramuscular |
| Tetracyclines | Oral | | |
| Xanthotoxin (8-methoxypsoralen) | Topical, oral | | |

***Rabbit*** The traditional methodology for a predictive test for phototoxicity has been an intact rabbit test [30]. This test is conducted as follows (and illustrated diagrammatically in Fig. 2):

**A. Animals and Animal Husbandry**

1. Strain/species: Female New Zealand white rabbits.
2. Number: 6 rabbits per test; 2 rabbits for positive control.
3. Age: Young adult.
4. Acclimation period: At least 7 days prior to study.
5. Food and water: Commercial laboratory feed and water are freely available.

**B. Test Article**

1. A dose of 0.5 mL of liquid or 500 mg of a solid or semisolid is applied to each test site.

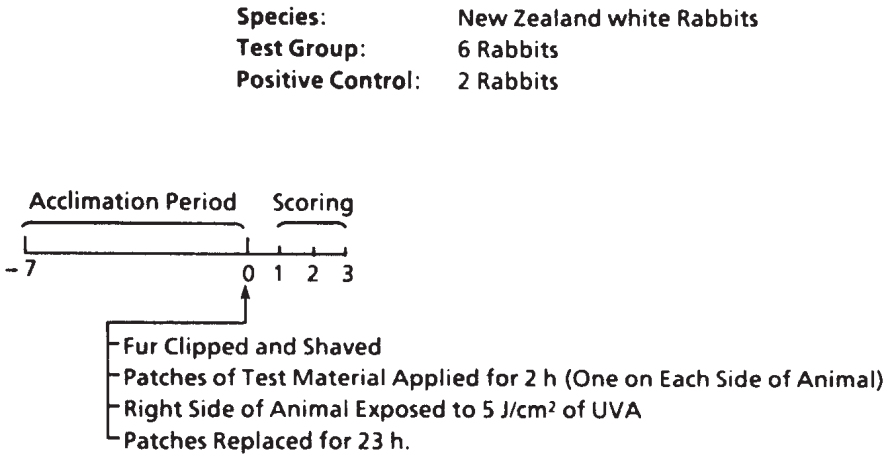


**FIGURE 2** Line chart for the design and conduct of phototoxicity assay using rabbits.

2. Liquid substances are used undiluted.
3. For solids, the test article is moistened with water (500 mg test article/0.5 mL water or another suitable vehicle) to ensure good contact with the skin.
4. The positive control material is a lotion containing 1% 8-methoxypsoralen.

## C. Experimental Procedures

1. Animals are weighed on the first day of dosing.
2. On the day prior to dosing, the fur of the test animals is clipped from the dorsal area of the trunk using a small animal clipper, then shaved clean with a finer bladed clipper.
3. On the day of dosing, the animals are placed in restraints.
4. One pair of patches (approximately 2.5 cm × 2.5 cm) per test article is applied to the skin of the back, with one patch on each side of the backbone.
5. A maximum of two pairs of patches may be applied to each animal and the patches must be at least 2 in. apart.
6. The patches are held in contact with the skin by means of an occlusive dressing for the 2 h exposure period.
7. After the 2 h exposure period, the occlusive dressing and the patches on the right side of the animal are removed (aluminum foil).
8. The left side of the animal is covered with opaque material.
9. The animal is then exposed to approximately 5 J/cm$^2$ of UVA (320–400 nm) light.
10. After exposure to the UVA light, the patches on the right side of the animal, as well as the occlusive dressing, are replaced.
11. The dressing is again removed approximately 23 h after the initial application of the test article. Residual test article is carefully removed, where applicable, using water (or another suitable vehicle).

12. Animals are examined for signs of erythema and edema and the responses are scored at 24, 48, and 72 h after the initial test article application according to the Draize reaction grading system previously presented in this volume.
13. Any unusual observation and mortality are recorded.

### D. Analysis of Data

The data from the irradiated and nonirradiated sites are evaluated separately. The scores from erythema and eschar formation and edema at 24, 48, and 72 h are added for each animal (six values). The six values are then divided by 3, yielding six individual scores. The mean of the six individual animal irritation scores represents the mean primary irritation score (maximum score = 8, as in the primary dermal irritation study). This method was developed after a human model had been developed.

***Guinea Pig*** Recently, a standardized protocol for using the guinea pig for phototoxicity testing has been proposed [31], which has been the subject of an international validation exercise. This is detailed in Fig. 3.

### A. Animals and Animal Husbandry

1. Strain/species: Male Hartley guinea pigs.
2. Number: At least 10 (two groups). Irradiation control: 4 animals; test material treated: 6 animals.

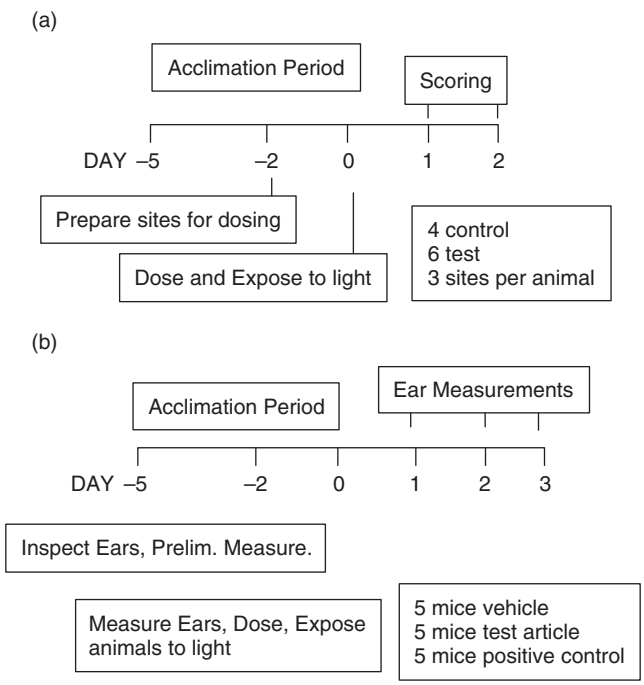


**FIGURE 3** (a) Guinea pig and (b) mouse for phototoxicity testing.

3. Age: Young adult, 300–500 g.
4. Acclimation period: At least 5 days.
5. Feed/water: *Ad libitum*.

**B. Test Material**

1. Vehicle: Test assumes that material will be in solution. Use the most volatile, nonirritating organic solvent possible, for example, ethanol, acetone, dimethylacetamide, or some combination.
2. Treatment: There can be up to four sites per animal, each measuring 1.5 cm × 1.5 cm (2.25 $cm^2$). In general, one side should be for a vehicle control, and another for a positive control (8-methoxypsoralen (8-MOP), 0.005% in ethanol).
3. Dosage: A dose of 0.025–0.050 mL is applied using a micropipette to each site.

**C. Experimental Procedure**

1. Animals are weighed on the first day of dosing.
2. Preparation: Approximately 48 h prior to treatment, remove the hair from a 6 cm × 8 cm area on the back with a fine clipper. On the day of dosing, animals are dosed as described previously. Tests are situated as to prevent mixing of test solutions after application. No patches or wraps are used.
3. Immediately after the dose application, the animals are placed in a restraint while keeping the test sites uncovered. Prior to irradiation the heads are covered to prevent ocular damage from the light exposure.
4. Thirty minutes after dosing, animals are exposed to a nonerythmogenic dose of light in the UVA band (should have peak intensity between 335 and 365 nm). The dose of light should be 9 or 10 J/$cm^2$ for UVA and 0.1–0.3 J/$cm^2$ for UVB.
5. Immediately after light exposure, the animals are wiped clean if necessary and returned to their home cages.
6. Animals are inspected and scored at 24 and 48 h postexposure according to the following:

| | |
|---|---|
| 0 | No reaction |
| 1 | Slight erythema |
| 2 | Moderate erythema |
| 3 | Severe erythema, with or without edema |

The reader should note that this scoring scheme is the same one used for dermal sensitization scoring, whereas the scoring method for the rabbit model discussed previously is that used for dermal irritation studies.

7. Any unusual clinical signs noted during exposure should be documented. The following descriptive parameters can be calculated from the data.

$$\text{Phototoxic Irritation Index (PTII)} = \frac{(\text{Number of positive sites} \times 100)}{(\text{Number of exposure sites})}$$

$$\text{Phototoxicity Severity Index (PSI)} = \frac{\text{(Total of scores)}}{\text{(Total of observations)}}$$

Lovell and Sanders [32] had previously proposed a similar model of assessing topical phototoxicity potential in the guinea pig. Their model differed from the one proposed by Nilson et al. [31] with regard to the following:

Only one test site per animal was used.
Test sites were smaller (about 1.6 $cm^2$).
Amounts applied were less (about 10 μL).
Light intensity was set at 15 $J/cm^2$.
Their paper made no reference to the use of a restrainer.
Assessments were conducted at 4, 24, 48, and 72 h.
The scoring system was as follows:

0 Normal
2 Faint/trace erythema
4 Slight erythema
6 Definite erythema
8 Well-developed erythema

(Intermediate scores were indicated by odd numbers.)
They recommended the use of acidine (weak phototoxin) or anthracene (strong phototoxin) for positive controls.

## 9.4 Pyrogenicity

The United States Pharmacopeia describes a pyrogen test using rabbits as a model. This test, which is the standard for limiting risks of a febrile reaction to an acceptable level, involves measuring the rise in body temperature in a group of three rabbits for 3 h after injection of 10 mL of test solution.

*Apparatus and Diluents* Render the syringes, needles, and glassware free of pyrogens by heating at 250 °F for not less than 30 min or by any other suitable method. Treat all diluents and solutions for washing and rinsing of devices or parenteral injection assemblies in a manner that will ensure that they are sterile and pyrogen-free. Periodically perform control pyrogen tests on representative portions of the diluents and solutions that are used for washing or rinsing of the apparatus.

*Temperature Recording* Use an accurate temperature-sensing device, such as a clinical thermometer or thermistor or similar probe, that has been calibrated to ensure an accuracy of ±0.1 °C and has been tested to determine that a maximum reading is reached in less than 5 min. Insert the temperature-sensing probe into the rectum of the test rabbit to a depth of not less than 7.5 cm and, after a period of time not less than that previously determined as sufficient, record the rabbit's temperature.

*Test Animals* Use healthy, mature rabbits. House the rabbits individually in an area of uniform temperature (between 20 and 23 °C) free from disturbances likely to excite them. The temperature should vary no more than ±3 °C from the selected temperature. Before using a rabbit for the first time in a pyrogen test, condition it for not more than 7 days before use by a sham test that includes all of the steps as directed under Procedure, except injection. Do not use a rabbit for pyrogen testing more frequently than once every 48 h, nor prior to 2 weeks following a maximum rise in its temperature of 0.6 °C or more while being subjected to the pyrogen test, or following its having been given a test specimen that was determined to be pyrogenic.

*Procedure* Perform the test in a separate area designated solely for pyrogen testing and under environmental conditions similar to those under which the animals are housed. Withhold all food from the test rabbits during the period of the test. Access to water is allowed at all times but may be restricted during the test. If probes measuring rectal temperature remain inserted throughout the testing period, restrain the rabbits with loose-fitting Elizabethan collars that allow the rabbits to assume a natural resting posture. Not more than 30 min prior to the injection of the test dose, determine the "control temperature" of each rabbit; this is the base for the determination of any temperature increase resulting from the injection of a test solution. In any one group of test rabbits, use only those rabbits whose control temperatures do not vary by more than 1 °C from each other, and do not use any rabbit having a temperature exceeding 39.8 °C.

Unless otherwise specified in the individual protocol, inject 10 mL of the test solution per kilogram of body weight into an ear vein of each of three rabbits, completing each injection within 10 min after the start of administration. The test solution is either the product, constituted if necessary as directed in the labeling, or the material under test. For pyrogen testing of devices or injection assemblies, use washings or rinsings of the surfaces that come in contact with the parenterally administered material or with the injection site or internal tissues of the patient. Ensure that all test solutions are protected from contamination. Perform the injection after warming the test solution to a temperature of 37 ± 2 °C. Record the temperature at 1, 2, and 3 h subsequent to the injection.

*Test Interpretation and Continuation* Consider any temperature decreases as zero rise. If no rabbit shows an individual rise in temperature of 0.6 °C or more above its respective control temperature, and if the sum of the three individual maximum temperature rises does not exceed 1.4 °C, the product meets the requirements for the absence of pyrogens. If any rabbit shows an individual temperature rise of 0.6 °C or more, or if the sum of the three individual maximum temperature rises exceeds 1.4 °C, continue the test using five other rabbits. If not more than three of the eight rabbits show individual rises in temperature of 0.6 °C or more, and if the sum of the eight individual maximum temperature rises does not exceed 3.7 °C, the material under examination meets the requirements for the absence of pyrogens.

**In Vitro *Pyrogenicity*** *In vitro* pyrogenicity testing (or bacterial endotoxin testing) is one of the great success stories for *in vitro* testing. Some fifteen years ago, the limulus amebocyte lysate (LAL) test was developed, validated, and accepted

as an *in vitro* alternative [33, 34] to the rabbit test. An *in vitro* test for estimating the concentration of bacterial endotoxins that may be present in or on a sample of the article(s) to which the test is applied uses LAL that has been obtained from aqueous extracts of the circulating amebocytes of the horseshoe crab, *Limulus polyphemus*, and that has been prepared and characterized for use as an LAL reagent for gel-clot formation. The test's limitation is that it detects only the pyrogens of gram-negative bacteria. This is generally not significant, since most environmental contaminants that gain entrance to sterile products are gram-negative [35].

Where the test is conducted as a limit test, the specimen is determined to be positive or negative to the test judged against the endotoxin concentration specified in the individual monograph. Where the test is conducted as an assay of the concentration of endotoxin, with calculation of confidence limits of the result obtained, the specimen is judged to comply with the requirements if the result does not exceed (1) the concentration limit specified in the individual monograph and (2) the specified confidence limits for the assay. In either case the determination of the reaction endpoint is made with parallel dilutions of redefined endotoxin units.

Since LAL reagents have also been formulated to be used for turbidimetric (including kinetic) assays or colorimetric readings, such tests may be used if shown to comply with the requirements for alternative methods. These tests require the establishment of a standard regression curve and the endotoxin content of the test material is determined by interpolation from the curve. The procedure includes incubation for a preselected time of reacting endotoxin and control solutions with LAL reagent and reading the spectrophotometric light absorbance at suitable wavelengths. In the case of the turbidimetric procedure, the reading is made immediately at the end of the incubation period. In the kinetic assays, the absorbance is measured throughout the reaction period and rate values are determined from those readings. In the colorimetric procedure, the reaction is arrested at the end of the preselected time by the addition of an appropriate amount of acetic acid solution prior to the readings. A possible advantage in the mathematical treatment of results, if the test is otherwise validated and the assay suitably designed, could be the confidence interval and limits of potency from the internal evidence of each assay itself.

## 9.5 Blood Compatibility

The standard test (and its major modifications) currently used for this purpose is technically an *in vitro* one, but it requires a sample of fresh blood from a dog or other large donor animal. The test was originally developed by the National Cancer Institute for use in evaluating cancer chemotherapeutic agents [36] and is rather crude, though definitive.

The variation described here is one commonly utilized. It uses human blood from volunteers, eliminating the need to keep a donor colony of dogs. The test procedure is described next.

1. *Test System.* Human blood. Collect 30 mL heparinized blood for whole blood and plasma (three tubes) and 30 mL clotted blood for serum (two tubes) from each of six donors.

2. *Precipitation Potential.*
   (a) For each donor, set up and label eight tubes 1 through 8.
   (b) Add 1 mL serum to tubes 1 through 4.
   (c) Add 1 mL plasma to tubes 5 through 8.
   (d) Add 1 mL formulation to tubes 1 through 5.
   (e) Add 1 mL vehicle to tubes 2 and 6.
   (f) Add 1 mL physiological saline to tubes 3 and 7 (negative control).
   (g) Add 1 mL 2% nitric acid to tubes 4 and 8 (positive control).
   (h) Observe tubes 1 through 8 for qualitative reactions (e.g., precipitation or clotting) before and after mixing.
   (i) If a reaction is observed in the formulation tubes (tubes 1 and/or 5), dilute the formulation in an equal amount of physiological saline ($\frac{1}{2}$ dilution) and test 1 mL of the dilution with an equal amount of plasma and/or serum. If a reaction still occurs, make serial dilutions of the formulation in saline ($\frac{1}{4}$, $\frac{1}{8}$, etc.).
   (j) If a reaction occurs in the vehicle tubes (tubes 2 and/or 6), repeat in a manner similar to that in Step (i).

3. *Hemolytic Potential.*
   (a) For each donor, set up and label eight tubes 1 through 8.
   (b) Add 1 mL whole blood to each tube.
   (c) Add 1 mL formulation to tube 1.
   (d) Add 1 mL vehicle to tube 2.
   (e) Add 1 mL of 1/2 dilution of formulation in saline to tube 3.
   (f) Add 1 mL of 1/2 dilution of vehicle in saline to tube 4.
   (g) Add 1 mL of 1/4 dilution of formulation in saline to tube 6.
   (h) Add 1 mL of 1/4 dilution of vehicle in saline to tube 6.
   (i) Add 1 mL of physiological saline to tube 7 (negative control).
   (j) Add 1 mL of distilled water to tube 8 (positive control).
   (k) Mix by gently inverting each tube three times.
   (l) Incubate tubes for 45 min at 37 °C.
   (m) Centrifuge 5 min at 1000 *g*.
   (n) Separate the supernate from the sediment.
   (o) Determine hemoglobin concentrations to the nearest 0.1 g/dL on the supernate (plasma).
   (p) If hemoglobin concentrations of the above dilutions are 0.2 g/dL (or more) greater than the saline control, repeat the procedure, adding 1 mL of further serial dilutions ($\frac{1}{8}$, $\frac{1}{16}$, etc.) of formulation or vehicle to 1 mL of blood until the hemoglobin level is within 0.2 g/dL of the saline control.

There are two proposed, true *in vitro* alternatives to this procedure [37, 38], but neither has been widely evaluated or accepted.

## REFERENCES


1. Gad SC, Chengelis CP. *Acute Toxicity: Principles and Methods*, 2nd ed. San Diego: CA: Academic Press; 1998.
2. Gad SC. Irritation and local tissue tolerance in pharmaceutical safety assessment. In *Drug Safety Evaluation*. Hoboken NJ: John Wiley & Sons; 2002, pp 367–403.
3. Draize JH, Woodard G, Calvery HO. Method for the study of irritation and toxicity of substances applied topically to the skin and mucous membranes. *J Reprod Fertil* 1944;20:85–93.
4. Eckstein P, Jackson M, Millman N, Sobero A. Comparison of vaginal tolerance test of spermical preparations in rabbits and monkeys. *J Reprod Fertil* 1969;20:85–93.
5. Weil CS, Scala RA. Study of intra- and interlaboratory variability in the results of rabbit eye and skin irritation tests. *Toxicol Appl Pharmacol* 1971;19:276–360.
6. Gad SC, Walsh RD, Dunn BJ. Correlation of ocular and dermal irritancy of industrial chemicals. *Ocular Dermal Toxicol* 1986;5(3):195–213.
7. Guillot JP, Gonnet JF, Clement C, Caillard L, Trahaut R. Evaluation of the cutaneous-irritation potential of 56 compounds. *Fundam Chem Toxicol* 1982;201:563–572.
8. Mathias CGT. Clinical and experimental aspects of cutaneous irritation. In Margulli FM, Maiback HT, Eds. *Dermatotoxicology*. New York: Hemisphere Publishing; 1983, pp 167–183.
9. Auletta C. Vaginal and rectal administration. *J Am Coll Toxicol* 1994;13:48–63.
10. Jochims K, et al. Local tolerance testing of parenteral drugs: how to put into practice. *Regul Toxicol Pharmacol* 2003;38:166–182.
11. Avis KE. Parenteral preparations. In Gennaro AR, Ed. *Remington's Pharmaceutical Sciences*. Easton, PA: Mack Publishing Company; 1985, pp 1518–1541.
12. Bacterial endotoxins test. *United States Pharmacopeia*. Rockville, MD: USP Convention; 1995, pp 1696–1697.
13. Gray JE. Pathological evaluation of injection injury. In Robinson J, Ed. *Sustained and Controlled Release Drug Delivery Systems*. New York: Marcel Dekker; 1978, pp 351–405.
14. Ballard BE. Biopharmaceutical considerations in subcutaneous and intramuscular drug administration. *J Pharm Sci* 1968;57:357–378.
15. Intracutaneous test. *United States Pharmacopoeia*. Rockville, MD: USP Convention; 1995, pp 1201–1202.
16. Garramone JP. *Vascular Access Port Model SLA. Users Manual*. Skokie, IL: Norfolk Medical Products; 1986.
17. Theeuwes F, Yum SI. Principles of the design and operation of generic osmotic pumps for the delivery of semisolid or liquid drug formulations. *Ann Biomed Eng* 1976;4:343–353.
18. Hagan EC. *Appraisal of the Safety of Chemicals in Foods, Drugs and Cosmetics*. Austin, TX: Association of Food and Drug Officials of the United States; 1959, p 19.
19. Shintani S, Yamazaki M, Nakamura M, Nakayama I. A new method to determine the irritation of drugs after intramuscular injection in rabbits. *Toxicol Appl Pharmacol* 1967;11:293–301.
20. Intramuscular irritation test. *United States Pharmacopeia*. Rockville, MD: USP Convention; 1985, pp 1180–1183.
21. Sidell FR, Calver DL, Kaminskis A. Serum creatine phosphokinase activity after intramuscular injection. *JAMA* 1974;228:1884–1887.

22. Young MF, Tobretta LD, Sophia JV. Correlative *in vitro* and *in vivo* study of skeletal muscle irritancy. *Toxicologist* 1986;6(1):1225.
23. Williams PD, Masters BG, Evans LD, Laska DA, Hattendorf GH. An *in vitro* model for assessing muscle irritation due to parenteral antibiotics. *Fundam Appl Toxicol* 1987;9:10–17.
24. Kato I, Harihara A, Mizushima Y. An *in vitro* method for assessing muscle irritation of antibiotics using rat primary cultured skeletal muscle fibers. *Toxicol Appl Pharmacol* 1992;117:194–199.
25. Brown S, Templeton D, Prater DA, Potter CC. Use of an *in vitro* hemolysis test to predict tissue irritancy in an intramuscular formation. *J Parenter Sci Technol* 1989;43:117–120.
26. Horio T, Kohachi K, Ogawa A, Inoue T, Ishirara M. Evaluation of photosensitizing ability of quinolones in guinea pigs. *Drugs* 1995;49(Suppl 2):283–285.
27. Lambert L, Warmer W, Kornhauser A. Animal models for phototoxicity tesing. *Toxicol Methods* 1996;2:99–114.
28. Anderson RR, Parrish JA. The optics of skin. *J Invest Dermatol* 1981;77:13–19.
29. Pathak MA. Photobiology of melanogenesis: biophysical aspects. In Montagna W, Uh F, Eds. *Advances in Biology of Skin, Volume 8, The Pigmentary System*. New York: Pergamon; 1967, pp 400–419.
30. Marzulli FM, Maibach HK. Perfume phototoxicity. *J Soc Cosmet Chem* 1970;21:685–715.
31. Nilsson R, Maurer T, Redmond N. A standard protocol for phototoxicity testing: results from an interlaboratory study. *Contact Dermatitis* 1993;28:285–290.
32. Lovell WW, Sanders DJ. Phototoxicity testing in guinea pigs. *Food Chem Toxicol*. 1992;30:155–160.
33. Cooper JF. Principles and applications of the limulus test for pyrogen in parenteral drugs. *Bull Parenter Drug Assoc* 1975;3:122–130.
34. Wearly M, Baker B. Utilization of the limulus amebocyte lysage test for pyrogen testing of large-volume parenterals, administration sets and medical devices. *Bull Parenter Drug Assoc* 1977;31:127–133.
35. Develeeshouwer MJ, Cornil MF, Dony J. Studies on the sensitivity and specificity of the limulus amebocyte lysate test and rabbit pyrogen assays. *Appl Environ Microbiol* 1985;50:1509–1511.
36. Prieur DJ, Young DM, Davis RD, Cooney DA, Homan ER, Dixon RL, Guarino AM. Procedures for preclinical toxicologic evaluation of cancer chemotherapeutic agents: protocols of the laboratory of toxicology. *Cancer Chemother Rep* 1973;4(Part 3):1–30.
37. Mason RG, Shermer RW, Zucker WH, Elston RC, Blackwelder WC. An *in vitro* test system for estimation of blood compatibility of biomaterials. *J Biomed Mater Res* 1974;8:341–356.
38. Kambic HE, Kiraly RJ, Yukihiko N. A simple *in vitro* screening test for blood compatibility of materials. *J Biomed Mater Res Sympos* 1976;7:561–570.

# SAFETY ASSESSMENT STUDIES: IMMUNOTOXICITY

JACQUES DESCOTES

*Poison Center and Claude Bernard University, Lyon, France*

**Contents**

1 Introduction
2 The Immune System: An Overview
  2.1 Lymphoid Organs
  2.2 Immunocompetent Cells
  2.3 Innate Immunity
  2.4 Specific Immunity
  2.5 External Influences on the Immune System
3 Immunotoxic Effects
  3.1 Clinical Consequences of Immunosuppression
  3.2 Clinical Consequences of Immunostimulation
  3.3 Hypersensitivity Reactions
  3.4 Autoimmune Reactions
4 Preclinical Assessment of Immunosuppression
  4.1 Histology
  4.2 Hematology and Immunology
5 Preclinical Assessment of Immunostimulation
  5.1 Histology
  5.2 Hematology and Immunology
  5.3 Immune Function Assays
  5.4 Host Resistance Models
6 Preclinical Assessment of Hypersensitivity
  6.1 Models of IgE-Mediated Hypersensitivity Reactions
  6.2 Models of Contact Sensitization
  6.3 Other Animal Models of T Lymphocyte-Mediated Hypersensitivity Reactions
  6.4 Models of Pseudoallergic Reactions

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

7 Preclinical Assessment of Autoimmunity
8 Regulatory Aspects and Risk Assessment
8.1 Guidelines
8.2 Good Laboratory Practices in Immunotoxicity Evaluation
8.3 Strategy for the Evaluation of Unintended Immunosuppression
8.4 Strategies for the Nonclinical Evaluation of Other Immunotoxic Effects
8.5 Immunotoxicity Risk Assessment
8.6 Clinical Immunotoxicity Assessment
9 Conclusion
References

## 1 INTRODUCTION

Over the last three decades, immunotoxicology slowly grew into a significant component of the preclinical safety evaluation of xenobiotics. In the mid-1960s, infections and neoplasia were identified as complications of potent immunosuppressive drugs shortly after their introduction for clinical use in transplant patients [1, 2]. These clinical findings served as an impetus to conduct experimental studies that almost exclusively focused on immunosuppression. Most early studies, however, lacked sufficient awareness that conditions of exposure or adequately selected doses are absolutely essential to generate relevant results for extrapolation from animal to human. Even though immunotoxicology was seemingly for the very first time the sole theme of a workshop held in 1974 [3], the birth of this new discipline was heralded in 1979 [4]. The shift from immunology to toxicology led to the introduction of toxicological concepts in immunotoxicity evaluation as summarized in a special issue of the journal *Drug and Chemical Toxicology* [5]. A major advance was the concept of tiered protocols, which consisted of a screening phase (tier I) including a matrix of *in vitro* and *in vivo* assays to be selected initially, and then a mechanism phase (tier II) based on a list of additional assays to be conducted depending on the results of the first tier [6]. This rationalized approach culminated in the conduct of interlaboratory validation studies of which the most significant were the B6C3F$_1$ mouse immunotoxicology study under the auspices of the U.S. National Toxicology Program (NTP) [7], the Fisher 344 cyclosporine validation study [8], the modified 28-day rat study [9], and the ICICIS rat validation study [10].

After many years almost entirely devoted to immunosuppression—often misleadingly used as a synonym for immunotoxicity even today—the scope of immunotoxicology expanded to include autoimmunity, then allergy, and finally immunostimulation. The importance of autoimmunity was first highlighted by the withdrawal from the market of the β-blocking drug practolol, which was the cause of the so-called oculomucocutaneous syndrome [11]. Thereafter, two major epidemics, the Spanish toxic oil syndrome, which affected more than 20,000 persons in 1981–1982 and killed over 300 [12], and the eosinophilia-myalgia syndrome (EMS) in users of preparations containing L-tryptophan [13], showed that chemically induced autoimmunity can be a significant health issue. Despite the routine use of guinea pig models to assess contact sensitization—a prototypic immunoallergic

reaction [14]—there was an implicit but unfounded consensus that allergic reactions cannot be predicted in animal studies. However, when drug allergies turned out to be a major cause of concern, efforts began to be paid to designing predictive models for use during drug development [15]. Last but not least, the introduction of the therapeutic cytokines demonstrated that enhancement of immune responses, that is, (unspecific) immunostimulation, is a relevant aspect of immunotoxicology [16].

Immunotoxic effects have often been subdivided into two broad categories, namely, immunosuppression and immunostimulation, the latter encompassing allergy and autoimmunity. As suggested long ago [17], categorization into immunosuppression, immunostimulation, hypersensitivity, and autoimmunity should be preferred as it more closely reflects the distinction between clinically significant immunotoxic effects that involve different effector mechanisms and require different animal models and assays for identification during nonclinical immunotoxicity evaluation.

## 2 THE IMMUNE SYSTEM: AN OVERVIEW

The immune system is a complex network of cells and molecules that ensures the detection of self from nonself in order to defend the integrity of the body [18]. To achieve this goal, a wide variety of mechanisms, either nonspecific (innate or natural immunity) or specific (adaptive or acquired immunity), come into play. Coordinated immune responses require interactions between specialized immunocompetent cells and biologically active molecules [19]. Immunocompetence is under the control of many mechanisms with redundant effects or conflicting outcome. Redundancy is an essential feature, which contributes to the functional reserve capacity of the immune system. Thus, it may prove difficult to anticipate the consequences of isolated morphological changes or functional alterations induced by medicinal products because of compensatory mechanisms. Last but not least, immune responses can be detrimental in certain circumstances, such as responses against self constituents of the host that result in autoimmune diseases, or reactivity against "innocent" foreign antigens (e.g., penicillin), leading to life-threatening systemic allergies.

### 2.1 Lymphoid Organs

Immunocompetent cells are found in the blood and almost all organs or tissues of the body, but they concentrate into specific anatomical structures, the lymphoid organs, which include central and peripheral lymphoid organs [20].

***Central Lymphoid Organs*** Central (or primary) lymphoid organs, where the production and/or maturation of lymphocytes take place, are lymphoepithelial structures that develop early during organogenesis and independently of antigenic stimulations.

*Bone Marrow* The bone marrow is the main hematopoietic organ in mammals and the equivalent of the bursa of Fabricius in birds. All blood cell types at various stages of maturation (except mature T lymphocytes) and their progenitors can be found in the medullary cavity of bones. A common multipotent progenitor cell gives rise

to lymphoid and myeloid progenitor cells that follow differentiation pathways to produce a variety of blood cell types. Soluble factors released by the stroma or accessory cells, including interleukin-3 (IL-3), the macrophage-colony stimulating factor (M-CSF), granulocyte-macrophage-colony stimulating factor (GM-CSF), and granulocyte-colony stimulating factor (G-CSF), exert critical influences on these complex processes. B lymphocytes, which derive from lymphoid progenitor cells, mature in the bone marrow.

*Thymus* The thymus is the main central lymphoid organ because T lymphocytes mature within the thymus. T lymphocytes at different stages of development can be identified from the expression of surface molecules (markers) denoted CD for cluster of differentiation. Triple negative bone marrow-derived thymocytes express none of the CD3, CD4, and CD8 major T lymphocyte markers. One subsequent step is the DP (double positive) thymocytes in the thymic cortex, which express both CD4 and CD8. Then, DP thymocytes migrate to the medulla to become single-positive T lymphocytes, either $CD4^+CD8^-$ or $CD4^-CD8^+$. They mature further and leave the thymus to populate organs and lymphoid tissues. T lymphocyte maturation is regulated by the epithelial microenvironment and thymic hormones. In the cortex, more than 95% of immature DP thymocytes are eliminated by apoptosis. Only those immature thymic T cells expressing receptors (TCRs) that recognize major histocompatibility complex (MHC) class I or class II molecules and associated self-peptides with moderate affinity escape from apoptosis and continue to differentiate (positive selection).

***Peripheral Lymphoid Organs*** Antigens and immunocompetent cells come into contact in the peripheral (or secondary) lymphoid organs, which fully develop only after repeated antigenic stimulations. The lymphoid follicles are small clusters of B lymphocytes, either primary follicles prior to antigenic stimulation, or secondary follicles containing a germinative center where precursors of plasma cells, memory B lymphocytes, $CD4^+$ T lymphocytes, and macrophages can be seen.

*Spleen* The spleen is a filter in the circulation and the site of immune responses against blood-borne antigens (e.g., opsonized microbes). The red pulp is the most prominent compartment containing macrophages, lymphocytes, and plasma cells. About 25% of all lymphocytes are located in the white pulp of the spleen, which consists of periarteriolar lymphoid sheaths with many T lymphocytes and adjacent follicles containing B lymphocytes. The white pulp plays a major role in immune responses against blood-borne antigens.

*Lymph Nodes* The lymph nodes are grouped at sites connecting blood and lymph vessels to ensure direct contact between antigens and immunocompetent cells. They are comprised of cortical, paracortical, and medullary areas. Antigens are processed by macrophages in the medulla. In the paracortex, free or processed antigens are presented to T lymphocytes. The cortical area is the site where B lymphocytes encounter antigens.

*MALT* The mucosa-associated lymphoid tissue (MALT) is located underneath the epithelium of the gut (gut-associated lymphoid tissue or GALT), the bronchopul-

monary tract (bronchi-associated lymphoid tissue or BALT), and the upper airways (nasal-associated lymphoid tissue or NALT). GALT includes the Peyer's patches in the duodenum and jejunum, and the appendix of the large intestine. In addition to IgA-producing plasma cells and intraepithelial T lymphocytes, MALT contains the M cells, a particular type of epithelial cells that take up inhaled or ingested antigens by pinocytosis. Because of its role in mucosal immunity, MALT is considered to be a first line of the host's defense.

## 2.2 Immunocompetent Cells

All immune cells originate in the bone marrow from multipotent stem cells that mature along different pathways. The lymphoid cell lineage gives rise to lymphocytes, and the myeloid cell lineage to neutrophils, basophils and eosinophils, monocytes and macrophages, dendritic cells, and mast cells (Fig. 1).

***Lymphocytes*** Lymphocytes [21] constitute 10% of blood cells in pigs, 20–40% in humans and dogs, 40–65% in monkeys, 50–70% in mice, and >80% in rats. There are marked interindividual variations depending on the age, sex, and strain of the animals. As lymphocytes display no cytological or ultrastructural characteristics, surface markers are used to distinguish between lymphocyte populations. Lymphocytes include T and B lymphocytes and natural killer (NK) cells.

*T Lymphocytes* Approximately 55–70% of circulating lymphocytes are T lymphocytes in humans and most mammals. They express the T-cell receptor (TCR), and CD2 and CD3 surface markers. T lymphocytes are subdivided into $CD4^+$ and $CD8^+$ T lymphocytes. $CD4^+$ (helper) T lymphocytes make up two-thirds of T lymphocytes in humans, rats, and dogs, but only one-half in monkeys. All $CD4^+$ T lymphocytes release IL-3, GM-CSF, and the tumor-necrosis factor-α (TNF-α). Depending on the profile of cytokines they secrete in certain circumstances, $CD4^+$ T lymphocytes are further divided into $T_H1$ and $T_H2$ lymphocytes, which primarily release IL-2 and interferon-γ (INF-γ), or IL-4, IL-5, and IL-10, respectively. The dichotomy between $T_H1$ and $T_H2$ lymphocytes is not a preexisting characteristic. Resting $T_H0$ lymphocytes are thought to be polarized to mount a $T_H1$ or a $T_H2$ response by ill-understood mechanisms. $CD4^+$ T lymphocytes control immune responses. $CD8^+$ (cytotoxic/effector) T lymphocytes recognize antigens presented by antigen-presenting cells (APCs) in association with MHC class I molecules. Activated $CD8^+$ T lymphocytes lyse target cells. Similar patterns of cytokine secretion have been shown so that $CD8^+$ T lymphocytes can also be subdivided into $T_C1$ and $T_C2$ lymphocytes.

*B Lymphocytes* Approximately 10–20% of peripheral blood lymphocytes are B lymphocytes in all mammals. They develop in the fetal liver, then in the bone marrow. They are identified by the surface immunoglobulins IgM, IgD, IgG, IgA, or IgE, and various surface cell markers including CD19, CD22, CD23, and CD37. Following antigen encounter, B lymphocytes proliferate and this results in the formation of germinal centers in the spleen and lymph nodes. As a general rule, B lymphocytes cannot be activated by the antigen alone. A second accessory signal is normally required for B lymphocytes to be activated (thymus-dependent antigens).

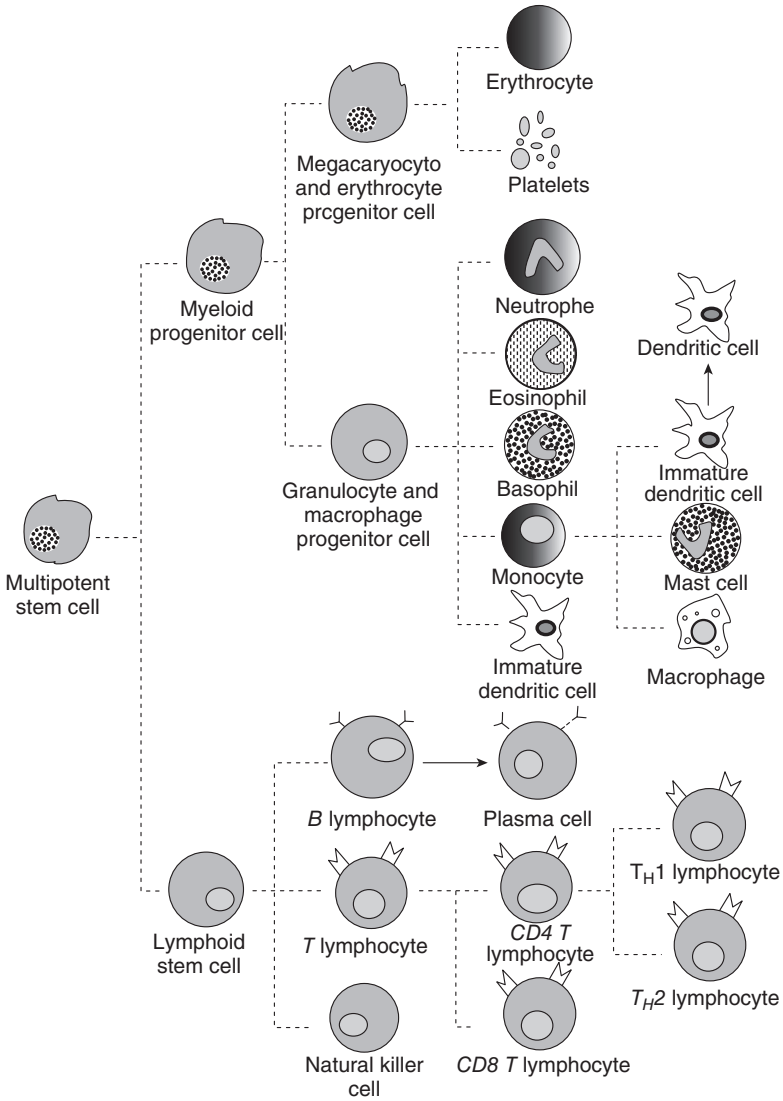


**FIGURE 1** Cells of the immune system and differentiation pathways. All cells of the immune system derive from a common multipotent bone marrow progenitor. Differentiation pathways lead to the myeloid and lymphoid lineages, and then well-characterized cells.

At their final stage of maturation, B lymphocytes form antibody-secreting plasma cells.

*Memory Lymphocytes* Lymphocytes have been estimated to produce $10^{15}$ different antibodies and TCRs. Antigen-binding specificities correspond to the antigenic repertoire whose diversity is achieved by extensive gene rearrangements. Normally, only a few thousand lymphocytes are specific for a given antigen. Memory T and B lymphocytes can persist in the body in a resting state for many years to ensure a

more vigorous and effective immune response upon reexposure to a previously encountered antigen.

*NK Cells* Natural killer (NK) cells are large granular lymphocytes that account for 15% of all leukocytes in humans, but only 5% in rats [22]. They are $CD3^{-}CD56^{+}CD16^{+}$ cells. However, murine NK cells do not express CD56, and CD56 is not a suitable marker in monkeys. Most NK cells are non T cells, but a subset of T lymphocytes—the NKT cells—express characteristic markers of NK cells along with a TCR. The role of NK cells is primarily to kill virus-infected or tumor cells without prior sensitization through the release of stored cytoplasmic cytotoxic mediators (e.g., perforin and granzymes). Activation of NK cell cytotoxicity is the result of a balance between killer-activating and killer-inhibiting receptors that recognize HLA molecules. Normal nucleated cells express MHC class I molecules on their surface, but virus-infected cells and tumor cells lose this ability. As the inhibitory signals are lacking, NK cells can kill the abnormal target cells and contribute to the host's defense against viral infections and cancers.

***Phagocytes*** Phagocytosis is the process by which cells can ingest and eradicate microbial pathogens and foreign particles whose size exceeds 1 μm. The most effective phagocytozing cells are neutrophils and monocytes/ macrophages.

*Neutrophils* Neutrophils constitute one-half of circulating white blood cells in most species and over 95% of PMNLs, but only 5–35% of rat leukocytes are neutrophils. Neutrophils contain two types of cytoplasmic granules: the primary (azurophil) granules that contain lytic enzymes (lysozyme, myeloperoxidase-derived substances, and cationic proteins), and the secondary (specific) granules that contain enzymes, such as lactoferrin and collagenases. Their half-life is 4–8 hours and because of this rapid turnover, they are particularly susceptible to toxic or immune-mediated insults. The primary function of neutrophils is phagocytosis so that patients with an innate or acquired decrease in neutrophil number or function can develop severe infections [23].

*Monocytes/Macrophages* Monocytes constitute 5% of circulating leukocytes. Their life span in the circulation is between 8 and 70 hours; then, they migrate to tissues where they differentiate into macrophages. The vast majority of macrophages are found in tissues. In humans and pigs, but not in rodents and dogs, there are intravascular macrophages fixed to the capillary bed of the lungs and their *in situ* activation can result in lung injury. There is a wide morphological and functional polymorphism of macrophages depending on their location [24]. Monocytes and macrophages are phagocytozing cells, but they also play the role of APCs. Activated monocytes and macrophages release cytokines (IL-1, IL-6, TNF-α), and this can result in target-organ toxicity (e.g., hepatotoxicity).

***Other Immunocompetent Cells*** A number of other cell types are involved in immune responses.

*Dendritic Cells* Immature dendritic cells (DCs) are scattered throughout the body. Because of their low phagocytic capacity, they degrade antigens by endocytosis to

small peptides subsequently presented to T lymphocytes. So-called danger signals can activate DCs via specific receptors to endogenous or exogenous substances. DCs then mature, attract and cluster with T lymphocytes to present antigenic peptides. Mature DCs play a major role in immune regulation [25]. They direct antigen-specific T lymphocytes either to die by apoptosis or to become anergic, memory or effector T lymphocytes. They also trigger T lymphocyte proliferation. When mature DCs release IL-12, they skew the immune response toward a $T_H1$ response. DCs also regulate antigen-specific $CD8^+$ T cell cytotoxicity and NK cell activity.

*Polymorphonuclear Leukocytes (PMNLs)* In addition to neutrophils, PMNLs include eosinophils and basophils [26]. Eosinophils account for 1–3% of peripheral blood leukocytes in humans and most animal species. Their cytoplasm contains specific granules where cationic proteins are stored. IL-5 controls the production of eosinophils in the bone marrow, and overproduction of IL-5 results in hypereosinophilia. Their half-life in blood is 8–18 hours. They distribute to tissues, especially the lungs and gut. Activated eosinophils release proinflammatory mediators including cationic proteins, newly synthesized eicosanoids, and cytokines. The major basic protein (MBP) accounts for 50% of all stored cationic proteins and is directly toxic to epithelial cells in the lungs. CD69 is a marker of activated eosinophils. Hypereosinophilia is typically noted in allergic diseases, inflammatory disorders, and helminth infections. Less than 1% of circulating leukocytes are basophils. They are stained by basic dyes, such as toluidine blue. After maturation in the bone marrow under the influence of IL-3, basophils circulate in the blood with a life span of several days. They express the high affinity receptor for IgE (FcεRI). Activation of basophils is induced by cross-linking of two adjacent FcεRI following binding of specific divalent antigen. CD63 is a marker of basophil activation. Activated basophils release preformed mediators stored in cytoplasmic granules including histamine, MBP, and tryptase. They also release newly formed mediators, such as leukotriene $LTC_4$, IL-4, and IL-13.

*Mast Cells* Mast cells are long-lived cells preferentially found near blood vessels and under epithelial, serous, and synovial membranes in the skin, the lungs, and the gut [26]. Like basophils, they are characterized by dense cytoplasmic granules stained by toluidine blue. Mast cells do not circulate in blood. They express a variety of surface receptors including FcεRI and the C3a and C5a receptors. The main preformed mediators stored in cytoplasmic granules are histamine, the neutral proteases tryptase, chymase, carboxypeptidase, and cathepsin D, and the proteoglycans heparin and chondroitin sulfate E. Newly synthesized mediators including the eicosanoids prostaglandin D2 and leukotriene LTC4, and TNF-α are also released by activated mast cells. The mechanism of mast cell activation is similar to that of basophils.

### 2.3 Innate Immunity

Immune responses are divided into innate (nonspecific) and adaptive (specific) immunity. Innate immunity, the first line of defense of the body against invading pathogens, involves a number of components and mechanisms [27].

### *Components of Innate Immunity*

*Epithelial Barriers* The body is protected by epithelial barriers in the skin and the respiratory, gastrointestinal, and urogenital tracts. Epithelial defenses include mechanical mechanisms (flow of air or mucus to entrap invading bacteria), chemical mechanisms (secretion of enzymes, such as lysozyme, or antimicrobial peptides), and microbial mechanisms (gut flora). When an epithelial barrier is breached by an invading pathogen, other mechanisms are brought into play.

*Soluble Factors* Antimicrobial peptides are small polypeptides divided into three main families: the defensins, histatins, and cathelicidins [28]. The isolated peptides are very heterogeneous in length, sequence, and structure, but most of them are small, cationic, and amphipathic. They exhibit broad-spectrum activity against bacteria, yeasts, fungi, and enveloped viruses.

*Complement* The complement is a complex system of more than 30 proteins with marked interspecies differences [29]. Several proteins are normally present in the circulation and tissues as functionally inactive molecules. Complement activation functions as an enzymatic cascade: inactive proteins are cleaved and activated, and then cleave and activate downstream proteins. Three activation pathways have been identified: the classical, alternative, and lectin-mediated pathways. The complement system serves important functions, as shown by the development of various pathological conditions in patients with primary or secondary complements deficiencies. Whatever the activation pathway, complement convertases cleave component C3 to C3a and C3b, which can bind covalently to the surface of pathogens. C3b-bound (opsonized) pathogens are more easily taken up and destroyed by phagocytes that bind C3b via the surface receptor CR1 (CD35). The small fragments C3a, C4a, and C5a—the anaphylatoxins—are formed during complement activation and can release inflammatory mediators from mast cells and basophils, which results in pseudoallergic reactions. The binding of the classical or alternative C3 convertase to C3b leads to the sequential cleavage of C5, C6, C7, and C8, and the polymerization of C9 to form the membrane-attack complex, which induces the formation of transmembrane pores in pathogens. The membrane-attack complex is also involved in the immunoallergic destruction of blood cells. The removal of circulating immune complexes is enhanced by complement activation. When immune complexes are not removed, they enlarge and precipitate at the basement membrane of small blood vessels, especially in the kidney glomeruli, a common feature in patients with systemic lupus erythematosus due to frequently associated deficiencies in the early complement components.

*Acute-Phase Proteins* Acute-phase proteins are molecules whose concentrations increase in response to inflammation [30]. In addition to several complement components, they include C-reactive protein, serum amyloid A, fibronectin, antiproteases ($\alpha$1-protease inhibitor, $\alpha$1-antichymotrypsin), and coagulation proteins (fibrinogen, plasminogen, urokinase). Following infection, trauma, or stress, increased plasma concentrations of acute-phase proteins are associated with fever, behavioral changes (anorexia, somnolence), stimulation of the adrenal cortex, and anemia. Clinically, the most widely used indicator of an acute-phase response is C-reactive

protein plasma concentration. The protective role of acute-phase proteins is still hypothetical.

*Chemokines* Chemokines are small cytokines that facilitate the migration (chemotaxis) of lymphocytes, monocytes, neutrophils, and eosinophils [31]. Over 50 different chemokines have been identified. Depending on their structure, they are classified into CC, CXC, C, and CX3C chemokines. Over 20 distinct receptors are known. Functionally, chemokines are either inducible chemokines that recruit leukocytes in response to stress, or constitutive chemokines involved in the trafficking of leukocytes and the architecture of secondary lymphoid organs. Chemokines play a major role in cell migration, stimulate leukocyte degranulation, and promote angiogenesis. Recent data suggest that chemotaxis is not their sole function. Indeed, chemokines have been shown to play a significant role in various pathological conditions including cardiovascular diseases, allergy, transplantation, neuroinflammation, and cancer.

*Proinflammatory Cytokines* In addition to their role in adaptive immunity, cytokines defend the body against infectious invaders. IFN-α and IFN-β produced by virus-infected cells inhibit virus replication and enhance antiviral defenses and NK cell activity. Activated macrophages release IL-1, IL-6, and TNF-α with a resulting increase in the production of acute-phase proteins.

***Mechanisms of Innate Immunity*** Inflammation defends the host against aggressions produced by physical or chemical agents, microbial pathogens, and various immune reactions. It is primarily characterized by the migration (chemotaxis) of leukocytes and the extravasation of serum proteins from the circulation. Thus, leukocytes can play their phagocytic role at the site of inflammation. The most effective phagocytozing cells are neutrophils and monocytes/macrophages. Under the influence of chemokines, phagocytes migrate to their targets and adhere to their surface nonspecifically (via lectins) or following opsonization. Then, phagocytes engulf their prey within phagosomes, where they release catalytic enzymes and free radicals. Soluble mediators involved in inflammatory reactions include free radicals, such as oxygen and nitrogen reactive species that induce direct tissue and cell injury or promote chain reactions, such as lipid peroxidation, leading to further tissue damage. Membrane lipids are the source of potent mediators, such as the prostaglandins and leukotrienes. Tissue injury is associated with the release of proteases that cause further tissue damage. Activation of the complement cascade can contribute to the inflammatory reaction via the formation of the membrane-attack complex or release of anaphylatoxins. Other important soluble mediators include the kinins, proinflammatory cytokines, and acute-phase proteins.

Recent evidence suggests that innate immunity facilitates specific immune responses. When pathogens breach the epithelial barrier, receptors, especially the toll-like receptors (TLRs) [32], recognize molecular patterns characteristic of particular classes of microorganisms as they are not found on self tissues. TLRs are mainly expressed by macrophages, neutrophils, and epithelial or endothelial cells. Recognition of pathogen-associated molecular patterns results in the activation of signaling pathways that drive and control specific immune responses.

## 2.4 Specific Immunity

Specific (adaptive) immunity comprises immune responses that are antigen-specific in contrast to innate immunity. Specific immune responses are subdivided into humoral and cell-mediated responses, both requiring close cell–cell interactions.

***Humoral Immunity*** Humoral immunity is the production of antigen-specific antibodies following contact with an antigen. Antibodies are heterodimeric glycoproteins, the immunoglobulins (Ig). In humans and most mammals, five classes of Ig have been identified, namely, the IgG, IgM, IgA, IgE, and IgD. The IgG class is further divided into four subclasses, $IgG_{1\text{-}4}$ in humans; $IgG_1$, $IgG_{2a}$, $IgG_{2b}$, and $IgG_{2c}$ in rodents. An Ig consists of two heavy (H) and two light (L) chains linked by disulfide bridges, which define domains ($V_H$, $V_L$, $C_{1L}$, $C_{1H}$, $C_{2H}$, and $C_{3H}$) in the amino acid sequence of the variable (V) and constant (C) regions of the H and L chains. The antibody site is formed by the variable domains. Each Ig molecule can be divided into two Fab fragments, which bind the antibody, and one Fc fragment involved in complement activation, opsonization, and placental transfer.

Antibody production requires the activation of B lymphocytes expressing the antigen-specific membrane Ig. Activated B lymphocytes proliferate and are transformed into lymphoblasts that give rise to memory B lymphocytes and antibody-secreting plasma cells. Following a first encounter with the antigen (primary response), antibody production starts after a latency period of 3–4 days, depending on the route of antigen entry and the dose and nature of the antigen (Fig. 2). Antibody titers subsequently grow exponentially and then decrease totally or to a residual level. In primary responses, IgM are first detected, then IgG. After a second encounter with the antigen (secondary response), antibody production is much earlier and more marked. Antibodies are primarily IgG with markedly higher affinity for the antigen and sustained antibody response.

***Cell-Mediated Immunity*** The immune system is a complex network where T lymphocytes play a pivotal role. Intercellular communications include the processing then the presentation of the antigen to T lymphocytes under strict genetic

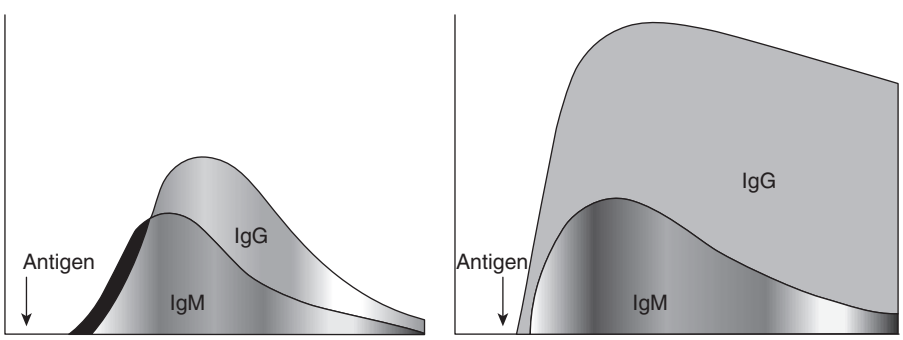


**FIGURE 2** Primary and secondary antibody response. Primary antibody responses are characterized by a latency period of 3–4 days. Then IgM and IgG titers grow exponentially and decrease totally or to a residual level. In secondary responses, antibody (primarily IgG) production is much earlier and more marked.

control of MHC class I or II molecules. Activation of T lymphocytes is under the control of IL-2. For optimal activation, two signals are required: a TCR signal upon binding to the peptide–MHC on an antigen-presenting cell (APC), and an antigen-independent signal involving costimulatory molecules (Fig. 3). If both signals are provided, T lymphocytes can proliferate and secrete cytokines. In the absence of costimulatory signals, T lymphocytes become anergic or apoptotic. Surface immunoglobulins bound to B lymphocytes can recognize a wide range of antigens in their native form, but TCRs can only recognize antigens as peptides bound to MHC molecules at the surface of APCs. Thus, antigens must first be processed into small peptides to be recognized. When T lymphocytes encounter a peptide–MHC complex, the signal delivered through the TCR is insufficient to trigger activation. A second costimulatory signal is required via ligation of costimulatory molecules, for example, CD28-B7, CD40L-CD40, CD2-LFA-3 (CD48, CD58), and LFA-1-ICAM-1. The term immunological synapse is used to refer to these close interactions between APCs and T lymphocytes [33].

Cellular immune responses include delayed-type hypersensitivity (DTH) reactions, systemic responses to microbial pathogens, contact dermatitis, granulomatous reactions, allograft rejection, and graft-versus-host reactions. They are mediated by $CD4^+$ or $CD8^+$ T lymphocytes. $CD4^+$ T lymphocytes are mainly cytokine-secreting lymphocytes that trigger and control immune responses. A subset of $CD4^+$ T lymphocytes, the $CD4^+CD25^+$ (regulatory) T cells, control immune homeostasis [34]. $CD8^+$ T lymphocytes are primarily involved in T lymphocyte-mediated cytotoxicity to kill target cells. Cytokines, of which more than 150 have been identified to date, are polypeptides released by a variety of cells, which play a pivotal role in many responses including inflammation or specific immunity [35]. They include

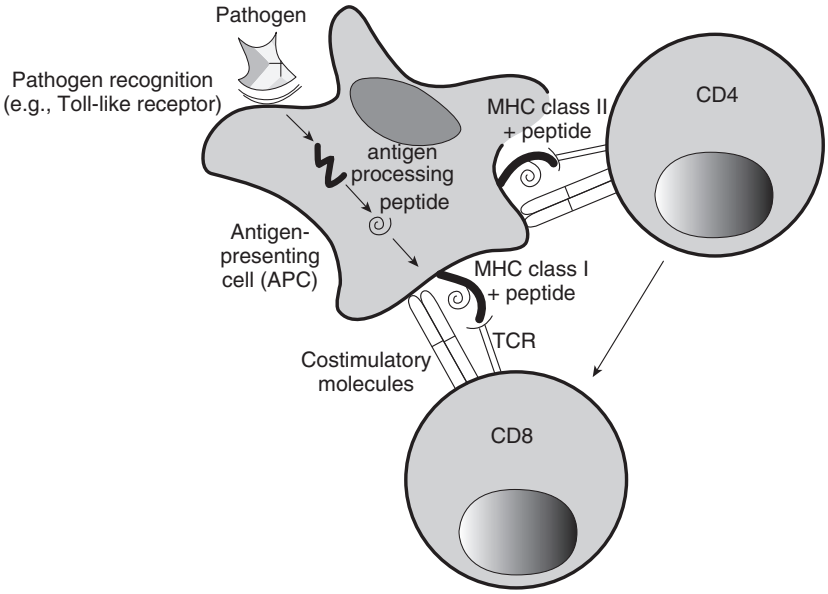


**FIGURE 3** Schematic representation of interactions between an antigen-presenting cell and $CD4^+$ or $CD8^+$ T lymphocytes.

interleukins, chemokines, growth factors, and many other endogenous substances. A brief description of the main cytokines is given in Table 1. The cytokine network is extremely complex. One given cytokine often exerts effects on various target cell populations (pleiotropy) and several cytokines can exert similar effects on the same target cells (redundancy).

### 2.5 External Influences on the Immune System

External factors can interfere with immune responses. They include the central and autonomic nervous system, the neuroendocrine system, nutrition, and age. The role of external factors should not be overlooked when interpreting immune findings in

**TABLE 1 Main Sources and Effects of Major Cytokines**

| Cytokine | Main Sources | Main Effects |
|---|---|---|
| **IL-1α/β** | Monocytes/macrophages, B cells, dendritic cells | Activation of T lymphocytes and APCs; induction of IL-2 production; synthesis of acute phase proteins |
| **IL-2** | $T_H1$ cells | Stimulation of T cells, B cells, and NK cells |
| **IL-3** | $T_H$ cells, NK cells | Growth factor of bone marrow precursor cells; activation of mast cells |
| **IL-4** | $T_H2$ cells | Secretion of $IgG_1$ and IgE by B cells; inhibition of $T_H1$ cytokine production; proliferation of T cells |
| **IL-5** | $T_H2$ cells | Growth factor of eosinophils; proliferation and differentiation of B cells; synthesis of IgA |
| **IL-6** | B and $T_H2$ cells, fibroblasts, macrophages, endothelial cells | Differentiation of B cells into plasma cells; induction of acute-phase reaction |
| **IL-7** | Bone marrow and thymus stroma cells | Differentiation of stem cells, B and T cells, progenitor cells |
| **IL-8** | Macrophages, endothelial cells | Chemotactic factor of neutrophils |
| **IL-10** | $T_H2$ cells | Activation and proliferation of B cells, thymocytes, and mast cells; inhibition of IFN-γ production |
| **IL-12** | Macrophages, B cells | Stimulation of IFN-γ production; activation and differentiation of $T_H1$ cells; cytotoxicity |
| **GM-CSF** | $T_H$ cells | Growth and differentiation of monocytes and dendritic cells |
| **IFN-α** | Leukocytes | Inhibition of viral replication; expression of MHC class I |
| **IFN-γ** | $T_H1$ and $T_C$ cells, NK cells | Inhibition of viral replication; expression of MHC; proliferation of $T_H2$ cells |
| **TGF-β** | T cells, monocytes | Chemotaxis of monocytes/macrophages; IgA synthesis; activation of macrophages (IL-1 secretion) |
| **TNF-α** | Macrophages, NK cells | Cytokine expression; induction of tumor cell death |

nonclinical studies. Immune responses can be altered by acute and chronic stress so that it can be difficult to differentiate immunosuppressive effects related to stress from a direct impact of the test article in animal studies. The thymus, spleen, lymph nodes, and MALT are innervated with neurotransmitter-containing fibers and many types of receptors have been demonstrated on immunocompetent cells including noradrenaline, adrenaline, dopamine, and serotonin receptors. There is a large body of evidence that cytokines exert effects on nonimmune tissues, such as the nervous and endocrine systems, and most hormones have immunomodulatory properties as well. Immune functions decline with age and aging-related changes contribute to more frequent diseases in elderly people. Both humoral and cellular immunity are affected, but cellular immunity is more markedly impaired due to physiologic thymic involution, which starts at puberty.

## 3 IMMUNOTOXIC EFFECTS

Even though this chapter is intended to review animal models and assays applicable to nonclinical immunotoxicity evaluation, a description of immunotoxic effects is helpful to select the most appropriate methods and interpret results adequately.

### 3.1 Clinical Consequences of Immunosuppression

Immunosuppression corresponds to decreased immune responsiveness. The difference between a total or partial inhibition of immune responsiveness—that is, immunosuppression and immunodepression, respectively—has never gained much attention despite obvious consequences in terms of risk assessment. Infections and virus-induced malignancies are the predominant clinical adverse effects associated with immunosuppression.

***Infectious Complications*** Immunocompromised human subjects typically present with more frequent, often more severe and relapsing, and sometimes atypical (opportunistic) infections. All pathogens including bacteria, viruses, fungi, and parasites can be involved. Even though all sites of the body can be affected, infections of the respiratory and gastrointestinal tracts, central nervous system, and skin are more frequent. Infectious complications have been described in patients treated with various immunosuppressive drug regimens, for instance, to prevent graft rejection [36], with corticosteroids to treat asthma, especially children [37], or with anti-TNF-α drugs used to treat rheumatoid arthritis and other autoimmune diseases [38]. Opportunistic infections, because they are caused by atypical pathogens and are located in uncommon sites, are generally easily attributed to marked immunosuppression. Clinically and microbiologically unremarkable infections may also develop in human subjects treated with less potent immunosuppressive (immunodepressive) drugs. Because these infections are unremarkable, only dedicated epidemiological studies can demonstrate an increased incidence in treated human subjects.

***Virus-Associated Malignancies*** Organ transplant patients are at a fourfold to 500-fold greater risk of developing malignancies, the most frequent being skin and

lip cancers, lymphoproliferative disorders, and Kaposi's sarcomas [39]. The more profound is immunosuppression, the more frequent are malignancies. However, lymphoproliferative disorders have also been described in patients treated with low-dose methotrexate [40] or anti-TNF-α drugs [41], suggesting a mild-to-moderate level of immunosuppression may suffice. A latent viral infection due to a human herpesvirus or the Epstein–Barr virus has been suggested to be involved, which can explain the possible regression of lymphoproliferative disorders in those patients where discontinuation of the offending drug is acceptable.

### 3.2 Clinical Consequences of Immunostimulation

Adverse effects associated with the therapeutic use of recombinant cytokines are the best examples of the clinical consequences of immunostimulation [16]. In this context, immunostimulation refers not to an antigen-specific but to a "globally" stimulated immune response, so that the term unspecific immunostimulation (or immunoactivation) is sometimes used.

***Flu-Like Reactions*** Mild-to-moderate fever (38–38.5 °C) with chills, arthralgias, and malaise can occur following initiation of treatment with immunostimulatory drugs or vaccination [42]. The reaction can, however, be much more severe with marked pyrexia (>40 °C), cardiovascular collapse or myocardial ischemia, obnubilation, and seizures. The suggested mechanism is the acute release of hypothalamic fever-promoting factors secondary to overproduction of IL-1 and TNF-α by activated monocytes/macrophages.

***Autoimmune Diseases*** A variety of autoimmune diseases similar to spontaneous diseases are more frequent in patients treated with several recombinant cytokines, such as the interferons-α and rIL-2 [43]. The underlying mechanism is not elucidated in many instances. Preexisting pathological conditions, such as hepatitis C infection, may be a risk factor. Enhanced expression of major histocompatibility complex (MHC) class I and class II antigens and the immunomodulatory effects of cytokines are thought to play a critical, even though ill-understood role.

***Hypersensitivity Reactions to Unrelated Allergens*** Treatments with immunostimulatory drugs can be associated with more frequent hypersensitivity reactions to unrelated allergens (e.g., asthma, rhinitis, or eczema). Interestingly, controlled clinical studies evidenced a greater risk of developing hypersensitivity reactions to radiocontrast media in rIL-2-treated cancer patients when compared to matched non-rIL-2-treated patients [44].

***Inhibition of Drug-Metabolizing Enzymes*** A poorly recognized adverse effect of immunostimulatory drugs is their potential to inhibit drug-metabolizing enzymes [45]. Human studies showed a negative impact of the interferons-α, several viral vaccines, or BCG on drug-metabolizing enzymes at therapeutic doses, sometimes leading to clinically significant drug interactions. Loss of cytochromes P450 was shown to predominate at the level of gene expression.

### 3.3 Hypersensitivity Reactions

The term allergy is often used misleadingly, hence the preferred use of hypersensitivity. Hypersensitivity reactions are considered to be the most frequent immunotoxic effects of drugs in humans, even though their actual incidence is not known accurately [46]. A variety of mechanisms are suspected, but only few have been fully elucidated. This may explain current difficulties in the diagnosis of drug-induced immune-mediated hypersensitivity reactions [47]. After years of debate, it is widely accepted that hypersensitivity reactions are caused by either immune-mediated or nonimmune-mediated mechanisms [48].

***Immune-Mediated Hypersensitivity Reactions*** Immune-mediated hypersensitivity (immunoallergic) reactions result from an antigen-specific immune response against the parent drug molecule or its metabolites. Nearly every organ or tissue (e.g., the skin, blood cells, lungs, kidneys, and liver) can be affected [49]. The time course of the reaction can be acute or chronic. Specific IgE, IgG, IgM, or IgA antibodies as well as T lymphocytes can be involved. The Gell and Coombs classification into four pathogenic mechanisms (Types I–IV) is no longer recommended as only a limited number of drug-induced immune-mediated reactions are adequately covered by this classification [50].

Large foreign molecules, such as recombinant cytokines, therapeutic proteins, and monoclonal antibodies, can act as direct immunogens and trigger a specific immune response. In contrast, the majority of medicinal products are small-molecular-weight chemicals that can only induce sensitization by playing the role of haptens, which means they have to bind to carrier proteins to form immunogenic complexes. In most instances, they have minimal or no intrinsic chemical reactivity, and therefore highly reactive intermediate metabolites instead of the parent molecule are thought to be involved [51]. Recently, however, direct binding of several drugs or metabolites to T lymphocytes has been shown to occur, thus bypassing the haptenization step [52].

***Nonimmune-Mediated Hypersensitivity Reactions*** A prior sensitizing contact is an absolute prerequisite for any drug-induced immune-mediated hypersensitivity reaction to develop. Because patients with no history of prior contact can nevertheless develop clinically similar reactions, nonimmune-mediated (pseudoallergic) mechanisms have been hypothesized to be involved. Clinical similarities are linked to the involvement of the same mediators as in immune-mediated reactions, but their release is due to pharmacotoxicological instead of immunological mechanisms [53]. Direct (nonspecific) histamine release and complement activation are characterized by clinical features more or less reminiscent of an anaphylactic (IgE-mediated) reaction, but flush, redness of the skin, abdominal pain, cough, mild-to-moderate tachycardia, and blood pressure changes are often predominant. Two additional mechanisms are cyclo-oxygenase inhibition with the compensatory overproduction of leukotrienes, resulting in intolerance to aspirin and NSAIDs, and the accumulation of bradykinin normally cleaved into inactive fragments by the angiotensin-converting enzyme (ACE), leading to angioedema caused by ACE inhibitors and angiotensin II receptor antagonists.

### 3.4 Autoimmune Reactions

Autoimmune diseases are collectively frequent in the general population, but drug-induced autoimmune reactions seem to be rare [54]. Autoimmune diseases that develop more frequently in patients treated with certain immunoactivating drugs (see earlier discussion) must be clearly differentiated from drug-induced autoimmune reactions. In contrast to the first situation, where nearly every autoimmune disease can be observed, one given drug is usually associated with only one type of autoimmune reaction: for example, autoimmune hemolytic anemia and α-methyl dopa, or the lupus syndrome and hydralazine. In addition, the clinical and immunobiological features of systemic autoimmune reactions, such as the lupus syndrome, are dissimilar from those of the spontaneous disease, that is, systemic lupus erythematosus (SLE).

## 4 PRECLINICAL ASSESSMENT OF IMMUNOSUPPRESSION

Immunosuppression has been the main if not only focus of most immunotoxicologists for many years. Thus, quite a few animal models and assays have been designed, standardized, and validated so that unintended immunosuppressive drugs can reasonably be expected to be detectable provided adequate nonclinical studies are performed in due time.

### 4.1 Histology

Histological examination of the major lymphoid organs is routinely performed. Lymphoid organs are sites where immunocompetent cells, mainly lymphocytes and monocytes/macrophages, are found in an organized microenvironment. In fact, immunocompetent cells are disseminated in nearly every organ and tissue of the body. For practical purposes, histological examination is restricted to some key lymphoid organs and tissues.

***Histological Technique*** Typically, the lymphoid organs submitted to histological examination include the thymus, spleen, bone marrow, lymph nodes, and Peyer's patches [55]. Lymphoid organs must be carefully dissected and weighed unfixed. Although lymphoid organ weights have often been used as an index of immunotoxicity, there is considerable variation so that statistical significance between controls and treated animals is rarely attained. Decreases in the relative weight of the thymus often characterize a potential for immunosuppression, whereas the relevance of decreases in spleen weight is debatable. Changes in lymph node weight are rarely useful due to marked interindividual variability. The bone marrow collected from the sternum or the femur is processed to smears stained with conventional dyes. Cell suspensions can be prepared for cytological or flow cytometry analysis. GALT is obtained from rolled intestinal segments or "Swiss rolls" that are either fixed or frozen. BALT is best examined from lung sections. Sampling of blood and nonlymphoid organs including the skin, lungs, kidneys, and liver is also needed for comprehensive histological examination. Tissues and organs are fixed as quickly as possible after necropsy. Formalin fixation, paraffin embedding, and staining with hematoxylin and eosin are recommended to identify changes in the architecture of lymphoid

tissues and the morphology of most immunocompetent cells. Electron microscopy requires special equipment and glutaraldehyde/paraformaldehyde fixatives. A description of best practice for routine pathology evaluation of the immune system was recently published by the Society of Toxicologic Pathology [56]. Interspecies variability must be taken into account when interpreting the results of histological examination [57].

Frank lesions, such as necrosis, fibrosis, and inflammation, are easily recognized, whereas changes in cell density, germinal center development, or prominence of high endothelial venules are often more subtle and require a skilled pathologist to differentiate abnormal findings from normal range. Structural changes are expected to be associated with modifications in the function of the affected organ, but it is as yet not established whether altered immune functions do correlate with structural changes. The term "enhanced pathology" was coined in the mid-1990s when standard histological examination of the lymphoid organs in rodents was first suspected to lack sufficient sensitivity. A number of new techniques have been introduced in toxicological pathology and are being considered for inclusion in nonclinical immunotoxicity evaluation. Immunohistochemistry detects proteins in tissue samples using specific polyclonal or monoclonal antibodies [58]. A secondary marker (e.g., a fluorochrome-conjugated antibody or streptavidin-biotin label) is used to visualize the binding of a primary antibody to specific epitopes. Nonfluorescent markers, such as immunoperoxidase and immunogold conjugates, are suitable for bright-field microscopy. A major limitation of immunohistochemistry is that it cannot be used for quantitative measurement, except cell proliferation. When antibodies are available commercially, a careful prevalidation is absolutely essential. When no antibodies are available, cross-reactivity across species is possible. In order to avoid diffusion in tissues, target molecules must be immobilized. Freezing following mild fixation (e.g., with acetone) is the preferred method as formalin-fixed, paraffin embedded tissue may not be appropriate due to antigen loss. Immunohistochemistry is used to detect changes in cell surface markers in response to inflammation or immunomodulation. Changes in the number or distribution of cell surface markers may be observed despite a lack of apparent changes in lymphoid tissue. However, the observed changes do not provide information on cellular function. Importantly, misinterpretation is possible due to the expression of the same cell surface marker by different subsets of immune cells. As standard histological examination can only provide a semiquantitative assessment of morphological changes, quantitative analysis may be necessary utilizing image analysis and stereological techniques that generate numerical values. *In situ* hybridization is based on the principle that labeled sequences of nucleic acids (probes) can specifically hybridize with cellular mRNA or DNA (gene) in tissues [59]. RNA probes offer the advantage of higher sensitivity, but they are labile and require technical skill. Labeling is achieved by radioisotopic methods using $^{35}$S- and $^{33}$P-labeled probes, or nonradioisotopic methods where probes are labeled with fluorochromes or the enzymes horseradish peroxidase and alkaline phosphatase. *In situ* hybridization is a valuable tool for the study of gene expression following drug administration, but its role in nonclinical immunotoxicity evaluation remains to be established.

***Histology of the Bone Marrow*** The bone marrow is distributed throughout the cavities of the skeleton. The hematopoietic tissue is progressively limited to the

proximal epiphyses of long bones, central skeleton, and skull after birth. Fat constitutes 70–80% of the bone marrow in rodents and only 50% in larger animals. Hematopoiesis takes place in the convoluted extravascular spaces of sinuses lined by a single layer of endothelial cells interdigitating with other cell types including reticular and dendritic cells. Niches are formed in the extravascular spaces where hematopoietic and myelopoietic cells can be generated. Bone marrow evaluation from sections of the femur or sternum is based on the cytological examination of smears to measure cellularity of the erythroid, myeloid, lymphoid, and megakaryocytic lines, the maturation of cells, and the ratio of myeloid to nucleated erythroid cells [60]. In most species, the myeloid:erythroid ratio has a slight myeloid predominance. The interpretation of results requires a peripheral blood cell count. From the limited published database, histological examination of the bone marrow is more likely to serve as a tool to detect myelotoxic drugs, which should be subsequently considered for extensive immunotoxicological evaluation.

***Histology of the Thymus*** The thymus is located in the anterosuperior mediastinum, except in the guinea pig, where it is located in the neck. It consists of two independent lobes attached by connective tissue. Each lobe is divided into lobules. Lobules are the basic thymic functional units enclosed by a fibrous capsule (Fig. 4). Three main areas can be identified: the subcapsular zone containing the earliest progenitor cells; the cortical area, which is packed with immature thymocytes undergoing selection; and the medulla, which is less densely populated by mature lymphocytes ready to leave. The thymus architecture is grossly similar across species

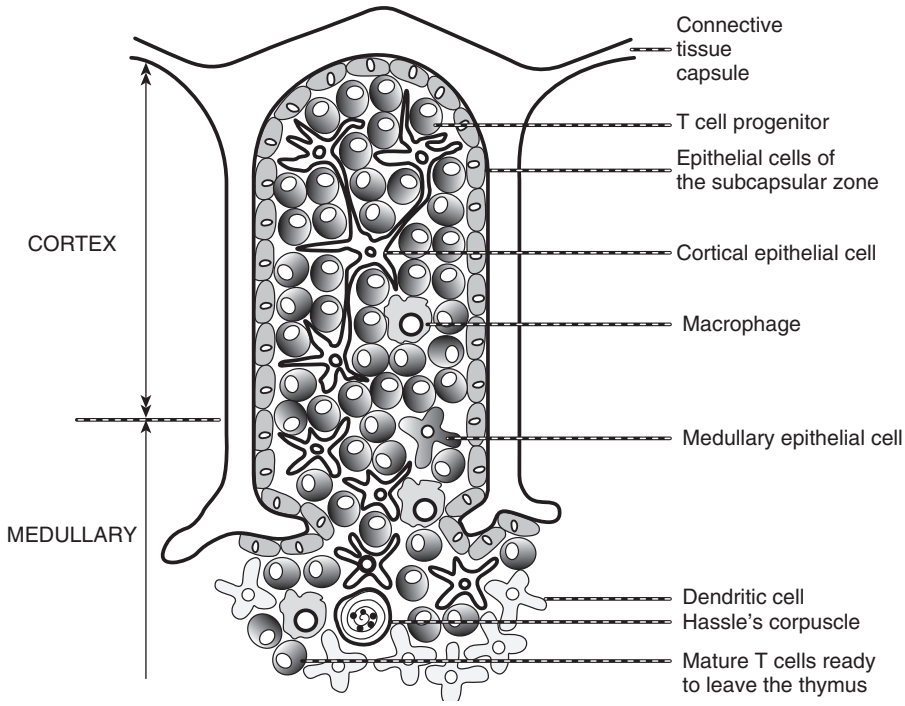


**FIGURE 4** Histology of the thymus.

[57]. Blood vessels and closely related nerves enter the lobules at the corticomedullary junction and plunge into the cortex. The thymus develops early in fetal life and the peripheral immunologic repertoire is fully established at birth. The thymus size increases considerably after birth, peaks before puberty, and then involutes. Most of the parenchyma is progressively replaced by adipose tissue so that only sparse lymphoid islets are seen in aged animals. The reason for aging-associated thymus atrophy is not known. Involution of the thymus in response to stress can occur and is primarily due to an increased production of corticosteroid hormones. Acute involution is seen in the first few days after stress and regeneration starts very rapidly with proliferation of large thymocytes in the outer cortex. The size of the cortex gradually increases with a considerable reduction in size of the medulla. Transient hypertrophy can be seen. Systemic toxicity resulting in decreased food consumption, or stressful housing and handling of animals are possible causes of thymic involution. Most potent immunosuppressive drugs induce thymic atrophy in rodents.

***Histology of the Spleen*** The spleen is surrounded by a capsule of dense connective tissue that forms branched trabeculae extending into the parenchyma [61]. The red pulp is the most prominent compartment with large, irregular, thin-walled blood vessels—the splenic sinusoids—interposed between sheets of reticular connective tissue—the splenic cords of Billroth—containing many macrophages. In mice and to a lesser extent in rats, the red pulp is the site of extramedullary hematopoiesis. The white pulp consists of a central arteriole surrounded by periarteriolar lymphoid sheaths with many T cells—the thymo-dependent area—and adjacent follicles containing B cells. The marginal zone, easily identified in rats but absent in humans, is located at the border of the red and white pulps. Spleen weight is a relatively insensitive correlate of immunosuppression. Published examples of significant changes in the histology of the spleen are very scarce. Atrophy of splenic T cell areas is generally associated with thymus atrophy. Decreased cellularity and germinal center development in splenic follicular areas are considered to be sensitive parameters.

***Histology of the Lymph Nodes*** Lymph nodes have a bean-shaped structure (Fig. 5). They are usually grouped at sites connecting blood and lymph vessels. Rodents have fewer lymph nodes than larger animals or humans, but the general architecture is relatively consistent across species despite interindividual differences and pathogen-free breeding [57]. The structure of lymph nodes consists of the cortical, paracortical, and medullary areas. Antigens enter lymph nodes by the afferent lymphatic vessels, which end in the medulla. Free or processed antigens are presented to T lymphocytes in the paracortex and B lymphocytes encounter antigens in the cortical area containing primary follicles, which are aggregates of small resting B lymphocytes, and secondary follicles containing B lymphoblasts within germinal centers. Upon antigen stimulation, the size of lymph nodes is very rapidly increased and signs of hyperactivity with well-developed germinal centers appear. In the lymph nodes, histological changes induced by immunosuppressants are generally similar and concomitant to changes observed in other lymphoid organs (e.g., the thymus).

***Histology of Peyer's Patches*** Peyer's patches are areas of lymphoid tissue beneath the epithelium of the duodenum and jejunum. The organization of Peyer's patches and other mucosa-associated lymphoid tissue (MALT) is similar to that of lymph nodes [62]. A characteristic of MALT is that IgA is the predominant antibody pro-

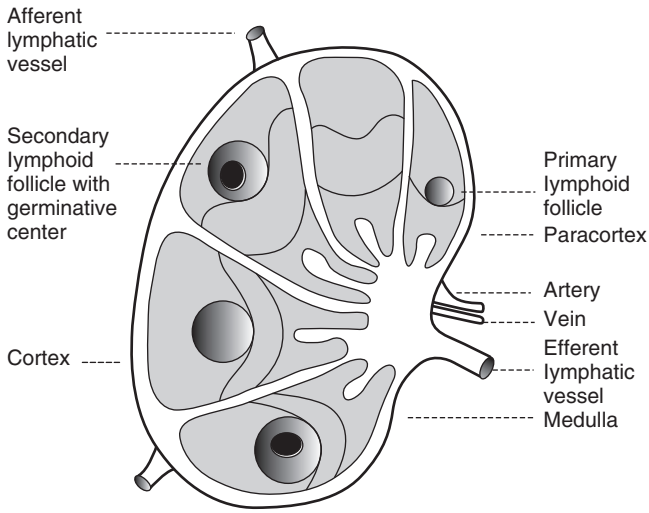


**FIGURE 5** Histology of lymph nodes.

duced by mucosal B lymphocytes. Although histological examination of the Peyer's patches is often required, relevant endpoints need to be defined and very few examples of consistent histological changes induced by medicinal products have been reported.

### 4.2 Hematology and Immunology

Hematology endpoints are included in every toxicity study and some changes may be used as predictors of immunosuppression. Specific immunological endpoints can also be measured.

***Hematology*** Differential leukocyte counts enumerate neutrophils, lymphocytes, monocytes, eosinophils, and basophils in the peripheral blood. In general, statistically significant, dose-related changes of 10–20% from baseline values are considered relevant. Leukopenia, lymphopenia, neutropenia, and eosinophilia are suggestive of an immunosuppressive effect, but they may also be due to stress. Indeed, leukocytosis can be seen in excited or frightened animals. Neutrophilia is more frequent in dogs and lymphocytosis in rats. This should be carefully avoided by adequate handling of the animals. Neutrophilia can also reflect an inflammatory response to the technical procedure, such as catheter implantation or repeated injections. Neutropenia is generally a direct toxic effect of the test article. In contrast, lymphopenia, especially in rats where the number of lymphocytes is normally high, can suggest an immunosuppressive effect [63]. Hematopoietic cells from the bone marrow are exquisitely sensitive to cytotoxic effects. Bone marrow toxicity is primarily evidenced by colony-forming cell assays, but flow cytometry techniques are increasingly used [64]. A number of *in vitro* assays are available to investigate the effects of medicinal products on specific progenitor cell populations of rodents, dogs, and monkeys including CFU-E (erythroid), CFU-GM (granulocytic-macrophage), and CFU-MK (megakaryocytic) clonogenic assays.

### *Immunology*

*Serum Immunoglobulins* All classes of serum immunoglobulins can be assayed by ELISA in animals. In most instances, however, changes in serum immunoglobulins are not associated with consistent changes in other parameters.

*Lymphocyte Subset Analysis* Flow cytometry is a powerful technique to analyze lymphocyte subsets [65]. Large populations of 10,000 or more single cells can be analyzed within a short period of time. Individual cells are passed in a fluid stream to be positioned through the path of a laser beam. The emitted light is gathered by detectors and then amplified by photomultiplier tubes. Information on cell size and density is obtained from the light scattered from each cell. Forward-angle light scatter is an approximate indicator of cell size used to distinguish viable from dead cells that exhibit lower forward-angle light scatter, whereas activated cells exhibit increased forward-angle light scatter. Side-angle light scatter correlates with granularity of the cell. Cell subsets can be recognized within complex cell populations using a combination of forward-angle and side-angle scatter. Fluorochromes (or fluorescent dyes), which interact with cells, enhance the power of flow cytometry. For instance, cells can be labeled with fluorochromes specific for surface markers (CDs). The resulting fluorescence is recorded cell by cell and quantified. The most recent cytometers can measure several fluorescence parameters simultaneously (multiparameter analysis). For lymphocyte subset analysis (immunophenotyping), lymphocytes are treated with a panel of monoclonal antibodies that are covalently bound to fluorochromes. In rodents, splenocytes are more commonly used than peripheral blood lymphocytes, but the latter may be more sensitive [66]. The use of lymphocyte subset analysis in nonclinical immunotoxicity evaluation largely derives from results obtained in the NTP interlaboratory study in B6C3F1 mice [67]. However, the conclusion of a panel of experts was that immunophenotyping has not been sufficiently validated to be recommended as a routine assay [68]. A major limitation is the current use of a small panel of lymphocyte subsets allowing the enumeration of total B and T lymphocytes, $CD4^+$ T lymphocytes and $CD8^+$ T lymphocytes. An important avenue of research is the use of activation surface markers [69], but no extensive validation studies have so far been conducted to compare changes in the expression of activation surface markers with the results of immune function assays.

*Measurement of Humoral Immunity* Besides the measurement of serum immunoglobulin levels, humoral immunity can be assessed by antibody response toward a given antigen injected either a few days before or at the end of drug treatment. In rodents, the duration of treatment should be at least 21–28 days to take into account the half-life of immunoglobulins. The kinetics of antibody production is accelerated and the magnitude of the response is increased in secondary as compared to primary responses. IgM antibodies are initially and predominantly produced in primary responses, whereas IgG of greater affinity are the predominant antibodies in secondary responses. Anyway, primary responses are most often used. Despite obvious differences between primary and secondary responses, no comparative studies on their respective predictive potency have seemingly been performed. Most often, antibody responses are under the control of T lymphocytes because T-dependent

antigens are by far the most commonly encountered antigens. Commonly used T-dependent antigens include sheep erythrocytes (SRBCs), bovine serum albumin (BSA), ovalbumin, tetanus toxoid, and keyhole limpet hemocyanin (KLH). SRBCs have long been the preferred antigen. Because they are not standardized and cannot be used in all animal species, they tend to be replaced by KLH. The use of T-independent antigens, such as DNP-Ficoll, TNP-*Escherichia coli* LPS, polyvinylpyrrolidone, and flagellin, is limited to mechanistic studies. Two different approaches can be used to assess humoral immunity, either the enumeration of antigen-specific antibody-producing spleen cells or the measurement of antigen-specific antibody serum levels. Rare comparative studies have been performed [70–72], so that the respective sensitivity of each approach is not known.

PLAQUE-FORMING CELL (PFC) ASSAY The PFC assay measures the production of specific antibodies by antibody-producing cells following immunization in rodents [73]. SRBCs are the most widely used antigens, but KLH or tetanus toxoid can also be used after coupling to SRBCs using tannic acid, chromic chloride, or carbodiimide. Because of considerable variations depending on the antigen and the experimental protocol, only SRBCs are used in the direct PFC assay for nonclinical immunotoxicity evaluation (Fig. 6). Due to variable antigenicity, the same commercial source of SRBCs should be used even though batch-to-batch variability will still exist. SRBCs can be stored at +4 °C for several weeks. Typically, (1–5) $\times 10^8$ SRBCs are injected intravenously. The intraperitoneal route can also be used. The animals are sacrificed after 4–5 days. A suspension of splenocytes is incubated with SRBCs and guinea pig complement for 3 hours over a slide according to Cunningham's technique. At the end of the incubation period, pale areas (plaques) corresponding to the lysis of red blood cells around an antibody-producing cell are counted with the naked eye or at a very small magnification. The number is usually given for $10^5$ or $10^6$ splenocytes. Based on the results of the NTP interlaboratory validation study [67], the direct PFC assay has long been considered to be the first-line function assay for nonclinical immunotoxicity evaluation. Because SRBCs are not properly standardized antigens, there is marked inter- and intralaboratory variability. In addition, cell counting cannot be automated and results vary depending on the technician's skill. Finally, this is a time-consuming assay that must be

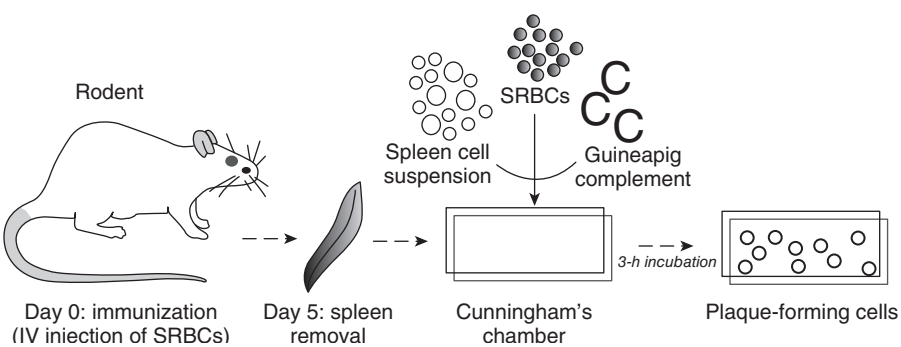


**FIGURE 6** Plaque-forming cell (PFC) assay.

performed shortly after sacrifice of the animals, which requires appropriate human resources.

ELISA Enzyme-linked immunoadsorbent assays (ELISA) are increasingly preferred to measure humoral immunity in nonclinical immunotoxicity evaluation. They offer the advantage of measuring a large number of samples in a relatively short period of time using an automated and properly standardized technique. Samples stored at +4 °C or frozen can be measured at a later date, and antibody levels can be measured on several occasions in the same animal. ELISA typically follows a five-step procedure: (1) coating of the microtiter plate wells with antigen; (2) blocking of all unbound sites to prevent false-positive results; (3) addition of the antibody to the wells; (4) addition of anti-IgG conjugated to an enzyme; and (5) reaction of a substrate with the enzyme (e.g., horseradish peroxidase or alkaline phosphatase) to produce a colored product indicative of a positive response. The color intensity is measured by a spectrophotometer. Results are expressed as absorbance values, optical density ratios (by comparing the optical density of the test sample with reference negative samples), or comparison to a standard curve. ELISA consists of relatively complex systems that require careful selection of the technique, adequate reagents and equipment, and technical skill to assure the quality, reproducibility, sensitivity, and accuracy of results. ELISA was first used to measure antibody responses to SRBCs [74], but today KLH is increasingly the most popular antigen [75–77]. As antibody responses vary depending on the antigen, differences in sensitivity can be expected. Decision on the most appropriate antigen as regards sensitivity and predictability can only be made when adequate data are available.

*Measurement of Cellular Immunity* *In vivo* animal models or *in vitro* assays can be used to measure cellular immunity in nonclinical immunotoxicity evaluation.

*IN VIVO* ANIMAL MODELS *In vivo* models are easy to perform and offer the advantage of taking into account the possible influence of other systems, such as the nervous and endocrine systems that may be negatively affected by the test article, which in turn may result in immune changes. Either the classical delayed-type hypersensitivity (DTH) or contact hypersensitivity model can be used [78]. Whatever the model, the experimental protocol consists of the sensitization (or induction) phase during which the animals receive one or several administrations of the reference immunogen, or contact sensitizer, then a rest period of 7–14 days required for the immune system to mount the expected immune response, and finally the eliciting (or challenge) phase, where the animals are administered the same immunogen, or contact sensitizer as previously. The selection of the immunogen or contact sensitizer, the dose and number of sensitizing injections, the route of administration during the sensitization phase, the duration of the rest period, and the dose and route of administration during the induction phase is actually not based on objective data, but instead on the technical skill and experience of the testing facilities. Although results with potent immunosuppressive drugs are roughly similar, it is not known to what extent variations in the experimental protocol can influence the relevance of the measured effects.

DTH models are the most commonly used models [79, 80]. They are as sensitive as *in vitro* assays. Commonly used T-dependent antigens include SRBCs, KLH, BSA,

ovalbumin, and tetanus toxoid. SRBCs are typically preferred in rodents, whereas KLH and tetanus toxoid are preferred in nonrodents. The antigen is injected either subcutaneously or intradermally during the sensitization phase. In rodents, the preferred sites of injection are the back, the basis of the tail, or the footpad for sensitization, and the footpad for elicitation. Hind-footpad thickness is most commonly used to measure the magnitude of the DTH response. The use of a dial caliper is a simple technique and more sophisticated methods, such as $^{125}$I diffusion, footpad weight, or footpad dye accumulation, are not more sensitive [81]. Footpad thickness is measured immediately prior to elicitation, and then 24–72 hours later on one or several occasions. The reaction normally peaks at 48 hours. Nonrodents are sensitized either subcutaneously or intradermally in one site and elicited in another site. DTH is assessed after intradermal injection of the antigen, which induces a skin reaction similar to that seen in human skin testing [82]. The magnitude of the reaction is measured semiquantitatively from the size of the erythema and the presence of edema.

Contact hypersensitivity models are primarily used in mice and guinea pigs. The experimental protocol is very close to that of contact sensitization assays [83, 84]. Strong contact sensitizers, such as picryl chloride, oxazolone, and DNFB in mice, and DNCB in guinea pigs, can be used. They are applied topically on the shaved abdomen or interscapular area to sensitize the animals that are elicited topically on the shaved abdomen (guinea pig, mouse) or the ear (mouse) after a short rest period. In guinea pigs, the response (erythema, edema) is measured semiquantitatively. Ear thickness measurement using a dial caliper is the preferred method in mice. Ear thickness is measured immediately prior to elicitation, and again 24–48 hours later. Immunosuppressive drugs decrease ear swelling.

### *IN VITRO* ASSAYS

*Lymphocyte Proliferation* Lymphocyte proliferation assays are based on the ability of cultured lymphocytes to proliferate. At termination of the drug treatment, animals are killed to collect lymphocytes, either from the peripheral blood, the spleen, or lymph nodes. Lymphocytes are typically cultured for 72 hours. The addition of a mitogen induces the nonspecific proliferation of either T (concanavalin-A, phytohemagglutinin) or B (LPS) lymphocytes, or both (pokeweed mitogen). Proliferation can also be induced by alloantigens. The animals are then sensitized *in vivo* by an antigen, such as tuberculin, and lymphoproliferation is induced *in vitro* by the addition of the mycobacterial extract PPD to lymphocyte cultures. Finally, lymphocyte proliferation can be triggered in a mixed lymphocyte response assay, where a cell suspension derived from the spleen of treated animals is incubated in the presence of pooled stimulator splenocytes from control histoincompatible animals after inactivation by mitomycin C [85]. Whatever the selected assay, tritiated thymidine is added 4–24 hours before culture termination for incorporation into proliferating lymphocytes. The incorporated radioactivity measured using a liquid scintillation counter is a correlate of lymphocyte proliferation and hence of cellular immunity. Colorimetric methods, using MTT or XTT, or BrdU incorporation can also be used to quantitate lymphocyte proliferation [86], but proliferation indices are at least 10 times less than with thymidine so that lower sensitivity or accuracy can be suspected [87].

*T Lymphocyte Cytotoxicity* Cytotoxic $CD8^+$ T lymphocytes are specifically cytotoxic to target cells. The *in vitro* induction of T lymphocyte cytotoxicity from splenocytes of treated animals is used to evaluate cellular immunity [87]. Splenocytes of treated and control animals are cocultured for 5 days with histoincompatible target cells, for example, P815 mastocytoma cells in mice and Fu-G1 tumor cells in rats, and inactivated by mitomycin C. Four hours before culture termination, the splenocytes are added to fresh $^{51}$Cr-labeled target cells. Cytotoxicity is measured from the radioactivity released into the supernatant. Cell-mediated cytotoxicity can also be measured by flow cytometry [88].

*Cytokine Assays* Because cytokines play pivotal roles in the triggering, nature, and control of immune responses, cytokine assays are deemed to become essential components of nonclinical immunotoxicity evaluation [89]. A number of hurdles, however, should be overcome for optimal performance of these assays: (1) very low concentrations of cytokines exert pleiotropic effects on numerous target cells; (2) cytokines can be found under multiple molecular forms with different assay behaviors; (3) a number of endogenous cytokine inhibitors have been described; and (4) the half-life of cytokines is usually less than 10 minutes. The conditions of sampling and storage of biological fluids for cytokine measurements are therefore of crucial importance. Sterile sampling on EDTA, but not heparin, is recommended for immunoassays. Sera must be decanted immediately and frozen at –80 °C. Bioassays measure cytokine levels in a biological system, usually a cell line [90]. Results are calculated by comparing with a standard dose–response curve obtained with known graded concentrations of the cytokine. Bioassays lack specificity and sensitivity and are time consuming. Immunoassays [91], either radioimmunoassays or ELISA, are much more specific and sensitive, but they are based on the binding of antibodies to the cytokine under scrutiny irrespective of its functional role. Many commercial kits are available, but results obtained with one kit may not be reproducible with another kit. Cytokines were initially measured in supernatants of cells stimulated by mitogens or anti-CD3 antibodies in order to achieve measurable levels, but subtle changes in cytokine production may not be detected under these conditions and observed changes may differ from those observed without stimulation. More recent techniques can measure cytokine levels in whole blood [92] or intracellular cytokines [93].

*Measurement of Innate Immunity* Although innate immunity plays a major role in the protection of the host against pathogens, measurement of the components and mechanisms involved during nonclinical immunotoxicity evaluation has not been paid as much attention as those involved in adaptive immunity.

PHAGOCYTOSIS As phagocytosis is a complex process involving different steps, a number of *in vivo* and *in vitro* techniques have been proposed. Most assays have been developed for clinical immunology purposes to investigate primary (hereditary) defects. Overall, they have not been properly standardized and validated from a toxicological point of view. Clearance assays measure the elimination of a substance by phagocytic cells *in vivo*. The standard assay is the *Listeria monocytogenes* clearance assay [94]. Mice or rats are injected intravenously with a suspension of *L.monocytogenes* and then killed at day 1 or 2 postinjection. After spleen removal, serial dilutions are plated to determine the number of colony-forming units. The

*L.monocytogenes* can also be inoculated intratracheally and the number of viable bacteria in the lungs is counted [95]. Ingestion assays measure the number of pathogens ingested per neutrophil or macrophage *in vitro*. They are not considered to be reliable tools to measure phagocytosis and in addition they are time consuming. The respiratory burst consists of metabolic changes leading to increases in oxygen uptake, oxidation of glucose uptake via the hexose monophosphate shunt, production of hydrogen peroxide and superoxide anion, generation of photons of light, and reduction of tetrazolium dyes. Metabolic assays measure either one of these changes [96]. Superoxide production can be measured from cytochrome-*c* reduction using a spectrophotometer. Chemiluminescence corresponds to light emission by stimulated cells, for example, phagocytes in the early phase of phagocytosis. As the quantity of light emitted is extremely low, amplifying agents, such as luminol, are used. In the presence of hydrogen peroxide or superoxide anion, luminol is converted to aminophthalic acid and photons of light are emitted, which are measured using a chemiluminometer. Chemiluminescence assays are time consuming and require technical skill and dedicated expensive equipment. Interestingly, they can be used *ex vivo* or *in vivo* and comparable results have been obtained in humans, rats, monkeys, and dogs [97]. Flow cytometry is another technique for the evaluation of phagocytosis [98]. The combination of forward-angle light scatter (size) and side-angle light scatter (granularity) allows for the discrimination between lymphocytes and polymorphonuclear leukocytes (PMNLs). Nonlymphocytes are gated to forward-angle light scatter and analyzed for associated fluorescence related to bacteria, fungi, or beads labeled with a fluorescent dye, most often fluorescein isothiocyanate. The oxidative burst can also be measured from the conversion of nonfluorescent substrates into fluorescent products by flow cytometry [99].

CHEMOTAXIS Infections in patients with congenital disorders of phagocyte motility demonstrate the importance of leukocyte motility and chemotaxis in innate immunity. Several techniques are available [100]. Boyden's technique measures the distance traveled by leukocytes *in vitro* through a polycarbonate or nitrocellulosis filter covered with a chemoattractant substance, such as formyl-methionyl-leucyl-phenylalanine (FMLP). In the agarose technique, migration is measured in wells of 2–3 mm in diameter dug in agarose. Wells contain the leukocyte suspension, the chemoattractant, or a control substance. Chemotaxis is measured by comparing the distance traveled by leukocytes toward the well containing the chemoattractant and the well containing the control substance.

NK CELL ACTIVITY The gold standard assay to measure NK cell activity is the $^{51}$Cr release assay [22]. Target cells, either YAK-1 murine lymphoma cells in rodents or K562 human myeloma cells in monkeys, are labeled with $^{51}$Cr. Thereafter, splenocytes or peripheral blood mononuclear cells are incubated with labeled target cells at 37 °C. Different effector:target cell ratios are used, for example, 200 : 1, 100 : 1, 50 : 1, or 25 : 1, as comparison of the percent lysis at a single effector:target ratio does not provide a quantitative assessment of potency between groups. Radioactivity of the supernatant is measured using a liquid scintillation counter. NK cell activity is calculated by comparing the radioactivity released by target cells in the presence of effector cells with the maximal radioactivity released by target cells in the presence of HCl or Triton X100. Spontaneous release is also taken into account. Flow cyto-

metry is an alternative to the $^{51}$Cr release assay. Target cells are labeled with 5-(6)-carboxy-fluorescein succinimidyl ester (CFSE). Subsequent labeling with propidium iodide is used to identify dead target cells. Results are expressed as the percentage of dead targets on a cell-to-cell basis. Flow cytometry has been used in rats and monkeys, and preliminary results suggest sensitivity similar to that of the $^{51}$Cr release assay [101]. As compared to the $^{51}$Cr release assay, flow cytometry offers the advantage of avoiding the use of radioisotopes, combining the measurement of NK cell activity with NK cell counting in the same animal and measuring NK cell activity on several occasions. The relevance of NK cell activity in nonclinical immunotoxicity evaluation is debatable. Potent immunosuppressive drugs, such as cyclosporine or azathioprine, exerted no effect in most studies. In fact, only a few drugs and chemicals (e.g., some morphine derivatives) were shown to inhibit NK cell activity reproducibly in rodents.

COMPLEMENT Exploration of the complement system is very rarely included in nonclinical immunotoxicity evaluation and then often is limited to the measurement of CH50 or C3 and C4 component serum levels that are endpoints neither sensitive nor comprehensive enough to provide an in-depth analysis of the effects of medicinal products on the complement system. A comprehensive study performed in a highly specialized laboratory is recommended when the complement system is a suspected target.

*Host Resistance Models* As changes in the histology of lymphoid organs and/or immune function may only be mild to moderate, or inconsistent, host resistance models can be helpful to determine whether the observed changes can be associated with impaired resistance to experimental microbial infections or implanted tumors [102]. In the early days of immunotoxicology, host resistance models were considered to be pivotal assays [103]. Their use, however, declined after several studies showed correlations between immune function changes and resistance to infection [104, 105], and also because of the growing concern about animal welfare issues. The use of host resistance models nevertheless is still an open question [106]. In experimental infection models, the pathogenicity of the infectious agent should ideally be as close as can be to that of the spontaneous human disease. It is also essential that the selected infectious agent can be used in conditions ensuring an adequate level of safety and compliance to good laboratory practices. Highly virulent pathogens are a significant biohazard risk and thus require appropriate safety measures. Finally, the immunological mechanisms involved in the infectious model should match with the immunotoxicity profile of the test article. Experimental infections where humoral defense mechanisms are primarily involved (e.g., *Streptococcus pneumoniae* or *Klebsiella pneumoniae* infections) are irrelevant to assess medicinal products that mainly alter cellular immunity. On the other hand, infections involving cellular defense mechanisms, such as *L.monocytogenes* infections, are not helpful to investigate medicinal products that impair humoral immunity. However, most immunotoxic drugs impair more than one aspect of immune functions and several host defense mechanisms are usually involved in fighting pathogens. The selection of an adequate infectious model is therefore difficult.

A variety of viral infection models have been proposed. Two rat models deserve particular attention, namely, infections with cytomegalovirus [104] and influenza virus [107]. NK cell activity was shown to play a major role in the defense against

cytomegalovirus, and humoral immunity in the defense against influenza virus. Infection with *L. monocytogenes* is a prototypic bacterial infectious model [108]. In normal animals, most injected *Listeria* are removed by macrophages. Surviving *Listeria* multiply in the liver, but the infection stops within 5–6 days when cell-mediated immunity is not otherwise impaired. Mortality or the number of viable pathogens in the spleen and liver of inoculated animals are the measured endpoints. Infections with *Streptococcus pneumoniae* and *zooepidemicus*, where the complement and humoral immunity play a major role, are also widely used [109]. Parasitic infections offer the advantage of a more consistent virulence and endpoints other than lethality can easily be measured. The *Trichinella spiralis* model, where cellular immunity plays a pivotal role in addition to enhanced parasite elimination by specific antibodies, is the most popular parasitic infection model [110]. An orally administered suspension of larvae induces infection, manifesting after one week with worms in the gut of 50–70% of the animals that are sacrificed on day 14 postinfestation to count adult worms. Immunosuppressed animals have a much higher number of encysted larvae than normal animals. Specific serum IgE levels can also be measured as a correlate to anti-infectious resistance.

Because of their low incidence, spontaneous tumors cannot be used as an indicator of unexpected immunosuppression in nonclinical studies. A number of implanted tumor models have been proposed of which fibrosarcoma PYB6 and melanoma B16F10 were more often used [102]. All tumor models include the injection of a suspension of tumor cells, a rest period to allow for tumor progression, the onset of a palpable tumor, and finally death of the animal. A syngenic tumor must be used to avoid tumor rejection by specific immune mechanisms. A small number of tumor cells to induce tumors in 10–30% of control animals is preferred so that associated drug-induced immunosuppression can be expected to result in increased incidence of tumors in treated animals. Measured endpoints include the incidence of tumors or animals with one or several tumors, the delay between inoculation of tumor cells and the appearance of palpable tumors, the time of tumor growth, the number of nodules, mortality, or the delay to death.

## 5 PRECLINICAL ASSESSMENT OF IMMUNOSTIMULATION

Relatively limited data are available regarding the nonclinical evaluation of (unexpected) immunostimulation as immunotoxicologists have long focused most of their efforts on immunosuppression and potent immunostimulatory drugs have only recently been introduced into the clinical setting.

### 5.1 Histology

Like unexpected immunosuppression, standard histological examination of lymphoid organs in repeated dose toxicity studies can be used to detect unintended immunostimulatory drugs. Technical aspects and the interpretation of results are also globally similar (see Section 4). Immunostimulation has been reported to be associated with increased spleen weight, even though no histological changes may be detectable; hyperplasia of the bone marrow, red and white pulps of the spleen, and lymph nodes, with mononuclear cell infiltration and more numerous germinal

centers; and mononuclear cell infiltration in the liver, lungs, and adrenal glands [111–113]. Extramedullary hematopoiesis, particularly in the spleen, has been observed in mice [114]. Histology of the thymus does not seem to be affected.

### 5.2 Hematology and Immunology

Mild-to-moderate hyperleukocytosis and lymphocytosis may occur [113, 115]. Depending on the mechanism of action, acute-phase response reflecting immune activation can result in decreased albumin and total protein serum levels [116]. A slight increase in liver weight, liver enzyme serum levels, and bilirubin and fibrinogen levels is possible. Increased serum Ig levels and increased B lymphocyte counts, and decreased $CD8^+$ T lymphocyte counts have rarely been observed.

### 5.3 Immune Function Assays

The immune function assays used to detect unintended immunostimulatory drugs are those used to detect immunosuppressive drugs. Enhanced responses have been reported using the PFC assay [117], antitetanus toxoid antibody levels [118], DTH to SRBCs [119] or *L.monocytogenes* [120], mitogen-stimulated lymphocyte proliferation [121], NK cell activity [122], or resistance to *T.spiralis* experimental infections [123]. Whether the same experimental conditions are appropriate to detect unintended immunostimulatory effects as reliably as immunosuppressive effects has rarely if ever been addressed. Indeed, animal models and assays designed to detect immunosuppression involve the use of optimal doses of the antigen to mount a humoral or cellular response, or optimal mitogen concentration in lymphoproliferation assays. From a theoretical point of view at least, suboptimal antigen doses or mitogen concentrations can be expected to result in improved detection of unintended immunostimulation. *In vitro* screens using various cell lines have been proposed to predict the cytokine-releasing potency of immunostimulatory drugs [124, 125].

### 5.4 Host Resistance Models

Because host resistance models are intended to assess the immunotoxicological relevance of observed immune changes, models applicable for the prediction of adverse consequences related to unexpected immunostimulation differ from those applicable to immunosuppressive drugs. Thus, experimental infection models are intended to detect enhanced resistance and it would be logical to use microbial challenge expected to kill more control animals than in experimental infection models, which are used to predict immunosuppression associated with decreased resistance. However, experimental infection models using the same design as in the evaluation of immunosuppressive drugs have been used [126].

More specific animal resistance models are those intended to predict whether immunostimulatory drugs trigger, accelerate, and/or aggravate autoimmune processes [127]. A variety of genetically prone mouse and rat strains develop frequent autoimmune diseases. NZB × NZW $F_1$ hybrid female mice develop a syndrome similar to SLE with anti-DNA autoantibodies as early as 2 months of age, and immune complex glomerulonephritis with proteinuria resulting in shorter life expec-

tancy [128]. The IgG and C3 deposits in the mesangium are evidenced by immunofluorescence from 5 months of age. Severe cortical thymic atrophy is a consistent finding, whereas spleen and lymph node hyperplasia is of variable intensity. An accelerated time course of the disease in this model reflects the detrimental effects of immunostimulatory drugs [129]. In the MRL-lpr/lpr mouse, the spontaneous disease is characterized by massive generalized lymphadenopathy as early as 8 weeks of age, immune glomerulonephritis, degenerative lesions of the coronary arteries and myocardium, and a high incidence of autoantibodies sometimes associated with polyarthritis, autoantibodies, and circulating immune complexes [130]. Life expectancy is less than 6 months. Clinical features similar to human type-I insulin-dependent diabetes mellitus (IDDM) are noted in nonobese diabetic (NOD) mice with a progressive onset of hyperglycemia, glycosuria, ketonuria, polydipsia, polyuria, and polyphagia [131]. Insulitis consisting of an infiltration of B and T lymphocytes in pancreatic islets resulting in loss of β-cells is evidenced by histological examination. Diabetes is best detected from blood glucose levels, but glycosuria can be used to avoid frequent bleeding. The incidence of diabetes is generally 80–100% in females and 30–50% in males. Female mice are permanently hyperglycemic at 3–6 months of age versus 5–8 months in males. Mice normally die within 5–7 weeks after reaching peak blood glucose levels. The incidence and course of the disease can be altered by immunomodulating drugs [132]. Two autoimmunity-prone rat strains can also be used. Mercuric chloride was shown to induce B lymphocyte polyclonal activation with a massive and rapid increase in IgE levels resulting in immune complex formation and glomerulonephritis in the Brown-Norway (BN) rat [133]. Although mercury has never been convincingly demonstrated to induce autoimmunity in exposed human subjects, the BN rat model is used to investigate the potential of drugs and chemicals for inducing or facilitating autoimmunity [134]. Typically, 60–80% of BB (Bio-Breeding) rats develop IDDM between 60 and 120 days of age, and facilitation of the disease has been reported following treatment with low doses of the interferon inducer poly(I:C) [135]. In addition to genetically autoimmunity-prone animals, a variety of experimental autoimmune diseases can be induced in laboratory animals. Arthritis induced by subcutaneous injections of Freund's complete adjuvant (FCA) or collagen type II in rodents is commonly used [136]. Lesions vary depending on the strain. The progress and severity of the disease are assessed from footpad swelling and joint deformity, histological examination of the joints, and various imaging techniques. Experimental autoimmune encephalomyelitis (EAE) is a model of multiple sclerosis [137]. Following immunization with myelin proteins, rodents develop clinical manifestations, such as loss of tail tonus and forelimb paralysis, which can easily be quantified. Overall, conflicting results have been obtained with genetically prone animal models and extremely few data are available regarding the use of experimental autoimmune disease models for nonclinical immunotoxicity evaluation.

## 6 PRECLINICAL ASSESSMENT OF HYPERSENSITIVITY

In sharp contrast to the growing concern about drug-induced hypersensitivity reactions, only limited aspects can be predicted and most available animal models and assays have at best poor predictability [138, 139].

### 6.1 Models of IgE-Mediated Hypersensitivity Reactions

As most medicinal products are not directly immunogenic because of their small size, they have to bind to carrier macromolecules to induce sensitization. As already mentioned, direct interaction with T lymphocytes bypassing hapten formation has been shown to occur and result in hypersensitivity reactions, but no generally applicable animal models are available to date. Because most medicinal drugs are devoid of the chemical reactivity that is required for covalent binding, reactive metabolites are widely considered to be involved instead of the parent molecule. The site of binding and degree of conjugation are critical factors [140]. Depending on the system used, IgG, IgM, or IgE are raised against the carrier, the hapten, or both. Attempts to reproduce spontaneously occurring haptenization have so far been unsuccessful. The selection of the optimal route of administration is another unsolved issue. The oral route—the most common route of therapeutic administration in humans—typically leads to tolerance instead of sensitization. Sensitization via the intranasal or intratracheal route is possible, but has so far only been documented with highly reactive chemicals [141]. Therefore, caution is required when interpreting positive results with highly reactive chemicals, such as industrial chemicals, intended to support the predictability of models or assays supposedly applicable to medicinal products as well as (false) negative results obtained with nonreactive medicinal products.

Whatever the model used, the induction of drug-specific IgE is a critical step. Although guinea pigs produce more $IgG_1$ than IgE when sensitized to an allergen, the detection of specific IgE is used for the evaluation of drug allergenicity [142, 143]. Adjuvants can be injected with the test article as a mixture or at different sites in order to increase sensitization. The use of Freund's complete adjuvant (FCA) is illogical as it favors a $T_H1$ response in contrast to aluminum hydroxide which induces a $T_H2$ response [144]. Specific IgE levels can be quantitated by passive cutaneous anaphylaxis (PCA) or ELISA. In PCA models, guinea pigs or mice are sensitized by several subcutaneous or intradermal injections of the test article alone or conjugated to a carrier [145]. Serum from sensitized animals is then injected intradermally into nonsensitized (naïve) guinea pigs, mice, or rats, and finally the test article mixed in a dye, such as blue Evans, is injected intravenously. When specific antibodies are present in the dermis, activation of mast cells ensues with subsequent release of vasoacting mediators. Due to increased capillary permeability, extravasation of the dye in tissues surrounding the injection site results in the formation of a blue spot, the size of which reflects the intensity of the reaction. ELISA can detect lower levels of specific IgE than PCA. It is noteworthy that most published results were obtained with penicillins or cephalosporins, and it is unknown to what extent they can be extrapolated to other classes of medicinal products. The induction of systemic anaphylaxis is another approach to evidence the production of specific IgE. During the first phase (elicitation or sensitization), guinea pigs are injected subcutaneously or intradermally on several occasions, and then after a rest phase of variable duration (usually 7–21 days), elicitation or challenge consists of the intravenous or intradermal injection of the test article. Severe acute reactions are characterized by markedly decreased respiratory rate and convulsions, followed by the death of the animal within 10–30 minutes. In mild-to-moderate reactions, clinical signs usually resolve within 30 minutes. Although systemic anaphylaxis guinea pig models have been

widely used, especially by Japanese authors, few validation studies have been conducted. A correlation was found between guinea pig results and the clinical experience with a panel of microbial extracts and vaccines [146]. A major limitation, however, is that macromolecules of human origin or humanized molecules consistently induce false-positive responses in this model, which may explain why their use is not recommended by ICH Guideline S6 on the preclinical safety evaluation of biotechnology-derived pharmaceuticals [147].

### 6.2 Models of Contact Sensitization

Although contact sensitization is an immune-mediated reaction, none of the current guidelines include contact sensitization in immunotoxicity studies of medicinal products. Following the pioneer work of Landsteiner and Draize, guinea pig models were once the most popular methods for the detection of contact sensitizers until mouse models, particularly the local lymph node assay (LLNA), emerged as alternative models forty years later [148].

***Guinea Pig Contact Sensitization Models*** At least seven guinea pig models have been accepted at some time point by regulatory authorities, but a number of other models have been proposed [14]. No general agreement has ever been reached on the optimal model or relative sensitivity of available models. Adjuvant-free models are generally considered to be less sensitive compared to adjuvant models that were sometimes claimed to be too sensitive and produce false-positive results. Similar results across guinea pig strains and sex have been reported. The technical skill and in-house historical data are in fact probably more important than the selected experimental procedure. All guinea pig contact sensitization models include: (1) an induction phase during which a mildly to moderately irritating concentration of the test article is applied epidermally or intradermally (when the test article is absolutely nonirritating, a slightly irritating solvent/vehicle, e.g. sodium lauryl sulfate, can be used); (2) a rest period of variable duration (e.g., 10–21 days); and (3) an eliciting (challenge) phase where a nonirritating concentration of the test article is applied topically on the shaved skin. Concentrations of the test article for the induction and challenge phases are determined during a preliminary study. The magnitude of the response is usually assessed by visual scoring including four grades (0 = no visible change; 1 = discrete or patchy erythema; 2 = moderate and confluent erythema; 3 = intense erythema and edema). Positive reference contact sensitizers are normally not included, but the selected assay should be performed at least every 6 months with positive controls to ensure the consistency of results in the testing facilities. There is no general consensus regarding the selection of positive controls.

The occlusive patch test was originally proposed by Buehler [149]. It consists of 1 or 3 weekly applications of the test article using induction patches to the same area of the shaved dorsal surface for 6 hours, and then occlusive patches maintained for 3–9 hours. After a rest period of 10–14 days, 6 hour challenge patches with a nonirritating concentration of the test article are applied to a naive clipped skin site. Reactions are graded for erythema 24 and 48 hours after patch removal. The test article is considered positive when sensitization is noted in at least 15% of animals. Validation studies showed a relatively good but inconsistent correlation between results in the Buehler test and clinical data [150]. One main advantage of the

Buehler test is the possibility to assess the final concentration or the whole formulation to be used commercially. The guinea pig maximization test (GPMT) was initially proposed by Magnusson and Kligman [151]. It consists of paired 0.1 mL intradermal injections of FCA, the test article mixed in a suitable vehicle, and a mixture of the test article dissolved or suspended in FCA into the clipped and shaved shoulders. On day 7, an occlusive patch is applied for 48 hours on the clipped shoulders. On day 21, challenge using a 4 $cm^2$ skin site on the left flank of the shaved abdomen is performed by application of a 24 hour occlusive patch with a nonirritating concentration of the test article in a suitable vehicle. The challenge site is observed at 24 and 48 hours after removal of the occlusive patch. The GPMT is a very sensitive test. The ranking of contact sensitizers used to be based on a five-grade scale: grade I (sensitization rate: 0–8%, weak); grade II (sensitization rate: 9–28%, mild); grade III (sensitization rate: 29–64%, moderate); grade IV (sensitization rate: 65–80%, strong); grade V (sensitization rate: 81–100%, extreme). Other guinea pig contact sensitization tests include the Freund's complete adjuvant test, in which the induction phase consists of three intradermal injections of a 1 : 1 mixture of FCA and the test article at a concentration not exceeding 5% into the suprascapular region, and the challenge phase consists of epicutaneous applications of several concentrations of the test article on 2 $cm^2$ areas of the clipped flank skin. The optimization test consists of 10 intracutaneous applications of a 0.1% solution or suspension of the test article every other day. The test article is incorporated in FCA at the same concentration during the second and third weeks. The animals are challenged intradermally 14 days after the last application, and then epicutaneously after a rest period of 10 days. The split adjuvant test was designed to allow assessing formulations following epidermal exposure. A 2 $cm^2$ area of shaved suprascapular skin is exposed to dry ice for 5 seconds prior to application of the test article on day 0. Occlusion is maintained for 48 hours until the next application. A total of four applications is used. Two intradermal injections of 0.1 mL of FCA are administered immediately before the fourth application on each side of the application site. The challenge consists of an open or occlusive application of the test article on day 21 and then the reaction is read after 24–72 hours. The open epicutaneous test consists of daily epicutaneous applications of the test article for 5–7 days and then 3 : 1 progressive dilutions up to a total of 20 applications on the intact skin of the flank. The animals are challenged on the contralateral clipped side of the flank with the test article or a nonirritating concentration on days 21 and 29 and the reactions are read at 24, 48, and 72 hours. Finally, the epicutaneous maximization test was developed to allow for assessing end-use products. The induction consists of topical applications of the test article on days 0, 2, 4, 7, 9, 11, and 14 using a 48 hour occlusive patch on the shaved suprascapular region. On day 0, 0.1 mL of FCA is injected intradermally. The challenge phase consists of the topical application of a nonirritating concentration of the test article on a naive skin site using a 48 hour occlusive patch on day 28.

***Mouse Contact Sensitization Models*** In contrast to guinea pig models, mouse contact sensitization models were developed rather recently after ear swelling had become a popular endpoint to measure murine contact hypersensitivity. The Mouse Ear Swelling Test (MEST) was the first attempt to propose a murine alternative to guinea pig contact sensitization models using a very similar approach [152]. Initially,

the MEST design included two intradermal injections of 0.05 mL of FCA into the abdomen immediately prior to the first topical application of the test article. Mice are dosed with 100 μL of a minimally irritating concentration of the test article or the vehicle topically on a previously shaved and tape-stripped abdominal area for 4 consecutive days. Seven days after the last application, 20–40 μL of the test article is applied to the left ear and 20–40 μL of the vehicle to the right ear. The thickness of both ears is measured with a micrometer at 24 and 48 hours postchallenge under light anesthesia. Of 49 known contact sensitizers tested, only vanillin, a mild sensitizer in guinea pigs and humans, was negative in the MEST and all 23 nonsensitizers were negative. Subsequent refinement included a diet supplemented with 250 IU/g vitamin A acetate starting 21 days before and maintained throughout the experiment to enhance ear swelling [153]. Results from several laboratories, however, showed that MEST is not a reliable model to detect weak-to-moderate sensitizers and a poorer predictor of contact sensitizers than the GPMT. MEST is no longer considered to be an alternative to guinea pig models.

Following pioneering work published as early as 1986 [154] and since then extensive interlaboratory validation studies [155, 156], the LLNA is widely accepted by regulatory authorities [157]. In contrast to all other contact sensitization models, the LLNA measures changes in the draining lymph node (i.e., the auricular lymph node) elicited in previously sensitized animals, instead of skin response [158]. Young adult female CBA/Ca or CBA/J mice are recommended as they proved to give more robust results. The test article can be dissolved in a suitable solvent or vehicle, for example, acetone/olive oil (4:1), *N,N*-dimethyl-formamide, methyl ethyl ketone, propylene glycol, or dimethylsulfoxide, and diluted if necessary. Several concentrations of the test article are generally used, but the highest achievable concentration should induce neither overt systemic toxicity nor excessive local irritation. For each mouse, 25 μL/ear of the test article is applied to the dorsum of both ears on days 1, 2, and 3. Then, mice are injected with 250 μL of sterile PBS containing 20 μCi of [$^3$H]-methyl thymidine via the tail vein on day 6. The auricular lymph node of each ear is excised 5 hours later and pooled in PBS for each animal. A single cell suspension of lymph node cells is prepared for each animal (Fig. 7). A stimulation index >3 is considered a positive response. Various modifications of the original protocol have been produced and include the measurement of proliferation by immunohistochemistry [159], IL-2 production [160], or BrdU incorporation [161]. Other endpoints including IL-6, IL-12, and IFN-γ production have been used, but none were shown to enhance sensitivity [162, 163]. Overall, the LLNA proved to be a sensitive assay to detect human contact allergens [164]. However, discrimination of primary irritants is a debated issue [165]. The percentage of B220$^+$ B lymphocytes in lymph nodes measured by flow cytometry was increased by contact sensitizers, but not by primary irritants [166].

### 6.3 Other Animal Models of T Lymphocyte-Mediated Hypersensitivity Reactions

Because T lymphocytes are involved in a number of immune-mediated reactions, it is tempting to use DTH or contact sensitization models to assess drug allergenicity. The subcutaneous injection of both benzylpenicillin conjugated to human serum albumin and FCA induced footpad swelling in mice 24 hours after injection of the

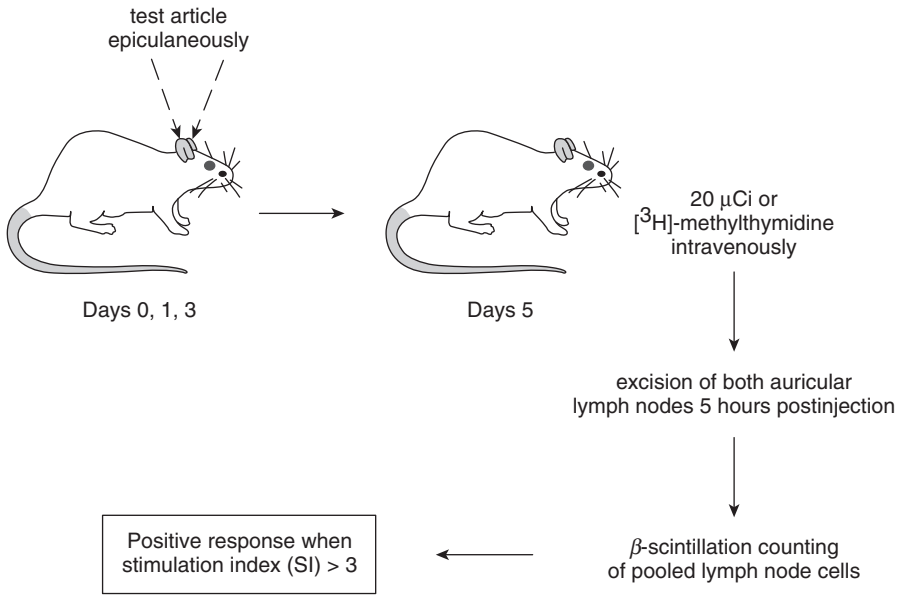


**FIGURE 7** Local lymph node assay (LLNA).

conjugate into the footpad of previously sensitized mice [167]. Similar findings were obtained with nonconjugated cephalothin. Footpad swelling induced by sulfamethoxazole mixed with S9 mix was obtained in mice previously sensitized by subcutaneous injections in the back, showing the role of metabolites in the induction of DTH by sulfamethoxazole [168]. A comparison of published GPMT results with clinical experience showed a good, or relatively good, correlation for 59 drugs and chemicals and conflicting results for 11 [169]. In contrast, only 14 of 83 drugs reported to the U.S. FDA for inducing systemic hypersensitivity induced a positive guinea pig test [170].

### 6.4 Models of Pseudoallergic Reactions

Only rare animal models are available to predict the risk of drug-induced pseudoallergic reactions. Intravenous administration to dogs that are particularly sensitive to nonspecific degranulation of mast cells and basophils can be helpful to evidence the histamine-releasing properties of a test substance [171, 172]. *In vitro* or *ex vivo* assays, such as histamine release [173], basophil degranulation [174], or complement activation [175] as used for the diagnosis of clinical reactions, may be more valuable.

## 7 PRECLINICAL ASSESSMENT OF AUTOIMMUNITY

Two major hurdles in the preclinical assessment of drug-induced autoimmune reactions are the poor understanding of underlying mechanisms and the role of individual and environmental predisposing factors [176]. As already mentioned, more

frequent autoimmune diseases can be associated with immunostimulatory drugs and these should be distinguished from drug-induced autoimmune reactions, which are addressed later. To date, no animal models are available to predict for the risk of drug-induced organ-specific autoimmune reactions. Early studies described autoantibodies in mice, rats, or dogs treated with a variety of drugs, but these findings either were obtained under conditions that did not mimic human treatment or could not be reproduced. Few studies searched for autoantibodies in the course of standard toxicity testing and results were either negative or conflicting [177]. Histological findings suggestive of an autoimmune process have sometimes been described [178]. Inflammation is the hallmark of autoimmunity in laboratory animals. However, mononuclear cell infiltration is not a reliable indicator of autoimmunity as it can be seen in autoimmune diseases as well as other pathological conditions. A predominantly $CD4^+$ T lymphocyte infiltrate and follicular hyperplasia in the spleen and lymph nodes may be more reliable indicators. Because of the key role of genetics in autoimmunity, autoimmunity-prone animal species are often considered for use in nonclinical immunotoxicity evaluation. However, the usefulness of the Brown-Norway (BN) rat model to assess drug-induced autoimmunity is debatable [134] and the same conclusion applies to autoimmunity-prone mouse models [179].

The popliteal lymph node assay (PLNA) may be the only currently available assay that could prove a predictor of systemic autoimmune reactions [180]. The original assay—the direct PLNA—is simple and inexpensive (Fig. 8). The test article is injected subcutaneously into one mouse or rat hind footpad, and the same volume of the vehicle (usually 50 μL) into the contralateral footpad. Seven days later, both popliteal lymph nodes are removed and immediately weighed to calculate the treated:control weight ratio (weight index). A weight index >2 is considered to be a positive response; indices >10 are rarely observed. The selection of the vehicle is critical. Saline is the first-choice vehicle, but in case of poor solubility, 20% DMSO or 25% ethanol seems suitable. Acetone, methyl ethyl ketone, and pure

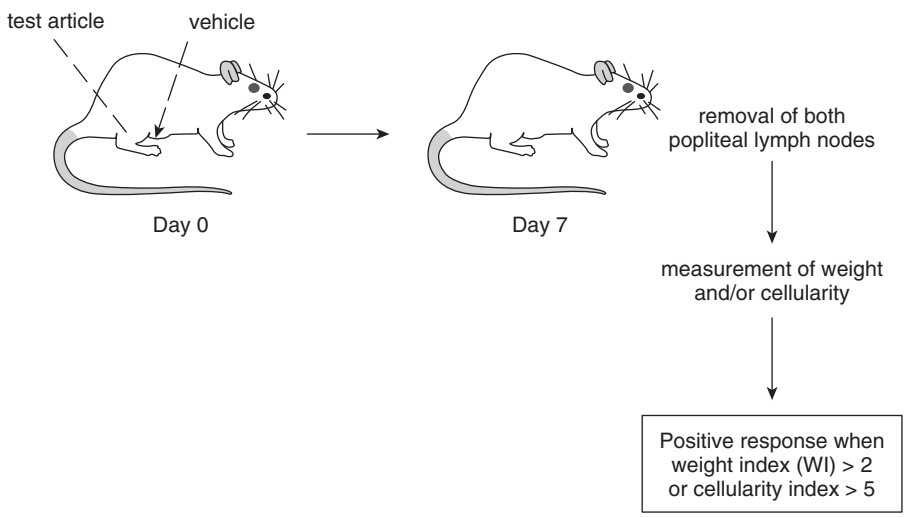


**FIGURE 8** Popliteal lymph node assay (PLNA).

ethanol because of primary irritancy, and corn oil, paraffin oil, and 1,2-propylene glycol because of excessive viscosity should be avoided as they cause false-positive responses. The dose is selected arbitrarily (usually 1 mg in mice and 5 mg in rats). Similar results have been obtained using various strains of mice and rats. A stronger response has sometimes been observed with PLN removal on days 9–11, instead of day 7. To enhance PLNA sensitivity, additional endpoints have been proposed. The cellularity index (CI) is calculated from the ratio of treated:control cellularity of lymph nodes carefully crushed immediately after weighing. A CI > 5 is considered to be a positive response and indices >50 have been reported. CI has been suggested to enhance PLNA accuracy as well as sensitivity. In contrast, thymidine incorporation 5 hours after the intravenous administration of [$^3$H]methylthymidine or lymphocyte subset analysis did not enhance PLNA sensitivity. Standard histological examination shows changes reminiscent of a graft-versus-host (GvH) reaction with blurring of the lymph node architecture due to an intermingling of the cortical and paracortical areas, numerous immunoblasts in paracortical areas, high-grade maturation of germinal centers, and numerous plasma cells in medullary areas. Results on more than 130 compounds with the direct PLNA have been published and most often there was a parallelism between human and animal data. One interlaboratory study showed consistent blind results with a limited panel of compounds [181]. False-positive responses, however, have been reported. They can be due to sensitization or immunogenicity, but then histological examination shows a conserved architecture of the lymph node and numerous germinal centers. Nonspecific inflammation due to primary irritation is another cause of false-positive responses. Standard histological examination is not helpful in contrast to cytokine production, most notably IL-6 and IFN-γ that are increased by streptozotocin, but not acetone or ethanol [182]. Depletion in CD8$^+$ T lymphocytes by administration of an anti-CD8 monoclonal antibody prior to performing the PLNA is another possibility to rule out false-positive responses due to primary irritants [183]. False-negative responses can also be seen. They are primarily due to the involvement of metabolites. Increasing the dose can positive the response, but systemic toxicity can then occur. *In vitro* preincubation with S9 mix and pretreatment with an enzyme-inducing agent, such as β-naphthoflavone or phenobarbital, are helpful alternatives. As the direct PLNA does not demonstrate whether the reaction is antigen-specific, variations of the original technique have been proposed. The secondary PLNA consists of a first phase where the test article is administered as in the direct PLNA, or by the intranasal, oral, or intravenous route, and then a dose insufficient to induce a primary positive response is injected into one hind footpad of the same animals after a rest period of 4–6 weeks [184]. In the adoptive transfer PLNA, T lymphocytes from a previously exposed donor are injected subcutaneously into the hind footpad of a naive recipient and the next day, a suboptimal dose of the test article is injected subcutaneously at the same site [185]. The modified PLNA was proposed more recently [186]. In this assay, the test article is injected with the reported antigens TNP-Ficoll or TNP-ovalbumin into the footpad and the specific antibody response is measured on day 7 by ELISPOT. Compounds that induce inflammation elicit an IgG response to TNP-ovalbumin, but not to TNP-Ficoll, whereas those inducing a T-lymphocyte-mediated reaction elicit an IgG response to both TNP-ovalbumin and TNP-Ficoll. The measurement of $IgG_1$, and $IgG_{2a}$ and IgE, can differentiate $T_H2$ and $T_H1$ responses, respectively. Even though the available database is still small, the

modified PLNA may prove more suitable to identify sensitizing drugs, whereas the direct PLNA seems to be more appropriate to detect drugs with potential for inducing autoimmune reactions [187, 188].

## 8 REGULATORY ASPECTS AND RISK ASSESSMENT

### 8.1 Guidelines

The evolution of concepts and techniques applicable to the nonclinical immunotoxicity evaluation of drugs was accompanied by regulatory documents [189]. The first step was the revision of OECD guideline 407 adopted on 27 July 1995, which requires histological examination of the main lymphoid organs in the control and high dose groups of 28 day rat toxicity studies [190]. The next year, the International Conference on Harmonization (ICH) S6 guideline emphasized the role of immunotoxicity in the safety evaluation of biotechnology-derived products [191]. In 2000, the European Medicines Evaluation Agency (EMEA) issued a note for guidance on repeated dose toxicity including a long annex on immunotoxicity testing [192], and two years later, the U.S. FDA released a guidance on the immunotoxicology evaluation of investigational new drugs [193]. Finally, the ICH S8 guideline on immunotoxicity studies for human pharmaceuticals was adopted on 15 September 2005 [194].

### 8.2 Good Laboratory Practices in Immunotoxicity Evaluation

Nonclinical immunotoxicity evaluation studies should adhere to GLP rules, but this issue has very rarely been addressed [195]. Different commercial sources of animals of the same strain can result in variable immune responsiveness. Transport conditions can also affect immune function, but changes are usually of short duration. Factors of animal environment including light, noise, humidity, temperature, bedding, or cage size can impact on immunocompetence. A critical issue may be the influence of changes in housing conditions. Different immune responses have indeed been described, even though inconsistently, between singly and group-housed animals, or between animals enjoying or not enjoying environmental enrichment. Handling stress is often associated with a certain degree of immunosuppression. The poor standardization of reagents is another issue. Most immunological reagents have not been carefully standardized. SRBCs are a typical example: there is marked variability of results, which is partly due to the use of different animal sources and the time lag between samplings in the same animal. When species-specific reagents are not available, cross-reacting reagents may be used, but careful validation is absolutely required prior to performing the assay. It is also important to bear in mind that similar reagents or kits may generate variable results depending on the commercial source.

### 8.3 Strategy for the Evaluation of Unintended Immunosuppression

The potential for unintended immunosuppression is typically investigated in rodent short-term repeated-dose studies.

***Species Selection*** General toxicity data should preferably be available in the species selected for nonclinical immunotoxicity assessment to ensure a sound interpretation of histological and/or functional immune changes. Even though the largest interlaboratory study by far was conducted in the $B6C3F_1$ mouse, a hybrid strain bred from C57Bl/6N female and DNA/2N male mice [7], the rat is currently the most commonly used species. It is noteworthy that most current concepts or assays derive from the results of mouse studies and that the available database in rats is much smaller than in mice, and often restricted to a few potent immunosuppressive drugs. The selection of an outbred instead of an inbred strain has been the matter of long debate, but as outbred strains showed no greater interindividual variability, they are currently preferred. It is essential to bear in mind that interindividual variability is a major issue in the interpretation of immunotoxicity studies. Animals as well as human subjects are either good or poor immune responders. As it is absolutely impossible to predict whether a given animal is likely to be a poor or good responder, unexpected difficulties for the interpretation of results may arise when one group of animals (e.g., control animals) obviously includes a greater percentage of poor or good responders. Even though rodents are commonly used, a nonrodent species may be preferred, such as monkeys for the evaluation of biotechnology-derived products with no detectable biological activity in rats. Surprisingly, other nonrodent species, such as dogs or mini-pigs, have so far very rarely been used.

***Route and Duration of Administration*** The test article is typically administered daily for 28 days using the intended route of administration in human subjects. There is seemingly no evidence that exposure of longer duration can be associated with either exacerbated or delayed immunosuppression. Few data are available for routes other than the oral route, except for the intravenous route.

***Dose Selection*** Only doses that do not cause overt toxicity should normally be used. It is often recommended that the highest dose should not result in a decrease in weight gain of more than 10% as compared to control animals. Indeed, doses either overtly toxic or associated with exaggerated pharmacodynamic effects can induce stress-related immune changes due to the release of corticosteroid hormones or other mediators, such as increases in circulating neutrophils, decreases in circulating lymphocytes and thymus weight and thymic cortical cellularity, and changes in spleen and lymph node cellularity [194].

***Measured Endpoints*** Strict tiered protocols [6] are no longer recommended and current strategies to detect unexpectedly immunosuppressive drugs are based primarily on the histological examination of the lymphoid organs as required by the ICH S8 guideline. There is an ongoing debate whether standard histology of the main lymphoid organs is sensitive enough. A number of pathologists claimed that standard histological examination is a reliable tool [196–198], but there was also a wide belief that immune function assays are nevertheless needed [199]. So far, only one study carefully addressed this issue [200]: the critical importance of the pathologist's skill and the criteria selected for scoring histological changes was highlighted, and the conclusion was that histopathology is not as sensitive as immune function assays, even though it may provide a reasonable level of accuracy

to identify immunosuppressive chemicals. When an immune function assay is considered to be necessary, for example, because clinical signs or histological changes suggestive of immunosuppression have been noted, or depending on the pharmacological mechanism of action, or the intended therapeutic use (weight of evidence), it is widely agreed that a T-cell-dependent antibody response assay is the first-line immune function assay in additional animals to avoid possible changes caused by antigenic stimulation even though there is no indication that this might happen. Indeed, antibody response assay offers the advantage to assess antigen presentation by APCs, T lymphocyte control, and antibody production by B lymphocytes, simultaneously. As already mentioned, ELISA to measure anti-KLH antibodies is increasingly the preferred method. Second-line assays, such as lymphocyte subset analysis, NK cell activity measurement, delayed-type hypersensitivity or lymphocyte proliferation assays, and macrophage/neutrophil function assays, are only recommended when standard histology and a T-cell-dependent antibody response assay suggest possible immunosuppressive effects. The selection of assays will depend on the type and magnitude of changes observed initially and the weight of evidence.

### 8.4 Strategies for the Nonclinical Evaluation of Other Immunotoxic Effects

No general strategy can be proposed for the prediction of unintended immunostimulation due to the very limited published database. Although medicinal products with unintended immunostimulatory properties have been very infrequently identified, it is not known whether they are actually rare or have been overlooked because of inadequate testing modalities. The prediction of the sensitizing potential of most medicinal products is nearly impossible at the present time, except for contact sensitization. No reliable animal models or assays are available to predict the risk of drug-induced autoimmune reactions.

### 8.5 Immunotoxicity Risk Assessment

Risk assessment is a systematic scientific characterization of adverse health effects potentially resulting from exposure to hazardous agents or situations. No study seemingly ever focused on specific issues related to the immunotoxicity risk assessment of medicinal products. Immunotoxicity risk assessment has often been limited to the extrapolation of animal data for the prediction of unexpected immunosuppression. In the parallelogram concept (Fig. 9), results from *in vivo* studies and *in vitro* assays in animals are compared with results from *in vitro* assays using human cells to extrapolate potential *in vivo* changes in human subjects. A major difficulty in immunotoxicity risk assessment is dose response that may follow an atypical pattern. Although dose response is often linear, a hockey cross-type curve reminiscent of the hormesis concept [201] may characterize compounds that exert immunosuppressive properties at high doses, immunoenhancing properties at low doses, and are inactive at mid-doses. In such a situation, totally different health consequences can be expected depending on the level of human exposure. Dose response is also atypical in hypersensitivity reactions. A bell-shaped curve suggests that no reactions may be expected at concentrations below or above an optimal range of concentrations that, in addition, is likely to vary across individuals.

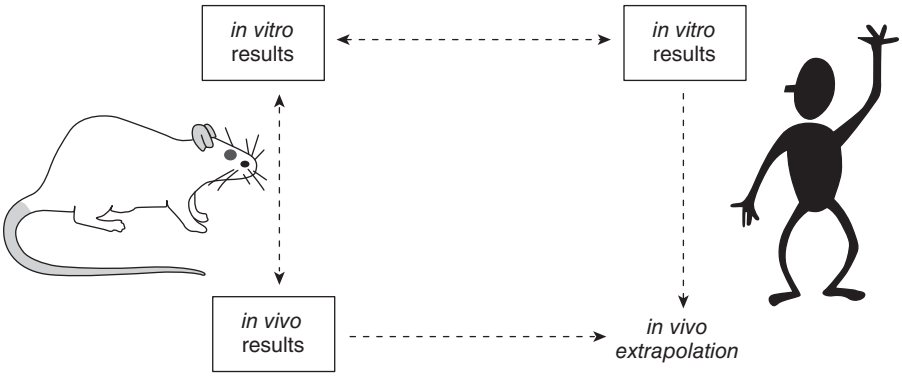


**FIGURE 9** The parallelogram concept for risk assessment.

### 8.6 Clinical Immunotoxicity Assessment

Clinical immunotoxicology is a neglected area of immunotoxicology and our current ignorance of immunotoxic effects in humans is a hurdle to the development and acceptability of animal models and assays for nonclinical immunotoxicity evaluation. Indeed, so little is known of the impact of most medicinal drugs on the immune system of treated humans that correlation between animal and human results is impossible in the vast majority of cases. The inclusion of immunological endpoints in clinical trials has very rarely been addressed specifically [202]. Immune function assays similar to those used in rodent and nonrodent studies for the identification of unintended immunosuppression or immunostimulation are proposed. However, it is important to bear in mind that such assays are often unable to predict the efficacy of immunosuppressive drugs to prevent organ transplant rejection, or immunoactivating drugs, such as therapeutic cytokines. Therefore, it may be possible that, in the future, modalities to evaluate the immunotoxicity of medicinal products in animal studies as well as clinical trials are based on an entirely different approach. For the time being, recent guidelines emphasize the need for a better prediction of the immunotoxicity of medicinal products, which led to various strategies as briefly reviewed earlier. With nonclinical immunotoxicity evaluation becoming a routine component of safety evaluation, it can be assumed that immunological endpoints will have to be included in clinical trials. For the sake of comparison between animal and human studies, it is logical to recommend that endpoints as close as possible are measured in both situations. Thus, lymphocyte subset analysis, which is not considered to be a first-line assay in nonclinical studies [68], can nevertheless be a helpful endpoint as it is often used in human studies. In contrast, KLH that is increasingly used as a T-cell-dependent antigen in antibody response assays should not be used in humans because of the risk of cross-allergies with seafood.

## 9 CONCLUSION

Current procedures for assessing the immunotoxicological safety of medicinal products are far from perfect because of our poor understanding of most underlying

mechanisms, and insufficient standardization and validation. The validation of available animal models and assays has often been restricted to few potent immunosuppressive compounds and there is an urgent need for large interlaboratory validation studies to refine current strategies. In addition, nonclinical immunotoxicity evaluation is currently limited to unintended immunosuppression, even though other immunotoxic effects, in particular, hypersensitivity, are far more critical safety issues in relation to the therapeutic use of medicinal products.

## REFERENCES


1. Meylers L. Cytostatic drugs. In Meylers L, Ed. *Side-Effects of Drugs*, 2nd ed. Amsterdam: Excerpta Medica; 1966, pp 472–494.
2. Leibowitz S, Schwartz RS. Malignancy as a complication of immunosuppressive therapy. *Adv Intern Med* 1971;17:95–123.
3. Groupement Lyonnais de Lutte Contre les Intoxications. *L'Immunotoxicologie*. Lyon: Lacassagne; 1975.
4. Abrutyn E. The infancy of immunotoxicology. *Ann Intern Med* 1979;90:118–119.
5. Special issue. *Drug Chem Toxicol* 1979;2:1–89.
6. Dean JH, Padarathsingh ML, Jerrells TR. Assessment of immunobiological effects induced by chemicals, drugs and food additives. I. Tier testing and screening approach. *Drug Chem Toxicol* 1979;2:5–17.
7. Luster MI, Munson AE, Thomas PT, Holsapple MP, Fenters JD, White KL, et al. Development of a testing battery to assess chemical-induced immunotoxicity: National Toxicology Program's guidelines for immunotoxicity evaluation in mice. *Fundam Appl Toxicol* 1988;10:2–19.
8. White KL, Grennings C, Murray MJ, Dean JH. Summary of an international methods validation study, carried out in nine laboratories, on the immunological assessment of cyclosporine A in the Fischer 344 rat. *Toxicol In Vitro* 1994;8:957–961.
9. Richter-Reichhelm HB, Dasenbrock CA, Descotes G, Emmendorffer AC, Ernst HU, Harleman JH, et al. Validation of a modified 28-day rat study to evidence effects of test compounds on the immune system. *Regul Toxicol Pharmacol* 1995;22:54–56.
10. The ICICIS Group Investigators. Report of validation study of assessment of direct immunotoxicity in the rat. *Toxicology* 1998;125:183–201.
11. Behan PO, Behan WMH, Zacharias FJ, Nicholls JT. Immunological abnormalities in patients who had the oculomucocutaneous syndrome associated with practolol therapy. *Lancet* 1976;2:984–987.
12. Posada de la Paz M, Philen RM, Borda AI. Toxic oil syndrome: the perspective after 20 years. *Epidemiol Rev* 2001;23:231–247.
13. Belonga EA, Mayeno AN, Osterholm MT. The eosinophilia-myalgia syndrome and tryptophan. *Annu Rev Med* 1992;12:235–256.
14. Kimber I, Maurer T. *Toxicology of Contact Hypersensitivity*. London: Taylor and Francis; 1996.
15. Ratajczak HV. Drug-induced hypersensitivity: role in drug development. *Toxicol Rev* 2004;23:265–280.
16. Vial T, Choquet-Kastylevsky G, Descotes J. Adverse effects of immunotherapeutics involving the immune system. *Toxicology* 2002;174:3–11.

17. Luster MI, Ackermann MF, Germolec DR, Rosenthal GJ. Perturbations of the immune system by xenobiotics. *Environ Health Perspect* 1989;81:157–162.
18. Metzhitov R, Janeway CA. How does the immune system discriminate self from non-self? *Semin Immunol* 2000;12:185–188.
19. Chaplin DD. Overview of the human immune response. *J Allergy Clin Immunol* 2006;117:S430–S435.
20. Kuper CF, De Heer E, Van Loveren H, Vos JG. Immune system. In Haschek MW, Rousseaux CG, Wallig MA, Eds. *Handbook of Toxicologic Pathology, Volume 2*, 2nd ed. San Diego, CA: Academic Press; 2002, pp 585–646.
21. Alam R, Gorska M. Lymphocytes. *J Allergy Clin Immunol* 2003;111:S476–S485.
22. Descotes J, Ravel G. Role of NK cells in immunotoxicity: an update. *Exp Rev Clin Immunol* 2005;1:605–610.
23. Segal AW. How neutrophils kill microbes. *Annu Rev Immunol* 2005;23:197–223.
24. Gordon S, Taylorn PR. Monocyte and macrophage heterogeneity. *Nat Rev Immunol* 2005;5:953–964.
25. Wallet MA, Sen P, Tisch R. Immunoregulation of dendritic cells. *Clin Med Res* 2005;3:166–175.
26. Prussin C, Metcalfe DD. IgE, mast cells, basophils, and eosinophils. *J Allergy Clin Immunol* 2003;111:S486–S494.
27. Tosi MF. Innate immune responses to infection. *J Allergy Clin Immunol* 2005;116: 241–249.
28. De Smet K, Contreras R. Human antimicrobial peptides: defensins, cathelicidins and histatins. *Biotechnol Lett* 2005;27:1337–1347.
29. Rus H, Cudrici C, Niculescu F. The role of the complement system in innate immunity. *Immunol Res* 2005;33:103–112.
30. Ceciliani F, Giordano A, Spagnolo V. The systemic reaction during inflammation: the acute-phase proteins. *Protein Pept Lett* 2002;9:211–223.
31. Le Y, Zhou Y, Iribarren P, Wang J. Chemokines and chemokine receptors: their manifold roles in homeostasis and disease. *Cell Mol Immunol* 2004;1:95–104.
32. Schnare M, Rollinghoff M, Qureshi S. Toll-like receptors: sentinels of host defence against bacterial infection. *Int Arch Allergy Immunol* 2006;139:75–85.
33. Dustin ML. The immunological synapse. *Arthritis Res* 2002;4:S119–S125.
34. Chatila TA. Role of regulatory T cells in human diseases. *J Allergy Clin Immunol* 2005;116:949–959.
35. http://microvet.arizona.edu/Courses/MIC419/Tutorials/cytokines.html.
36. Fishman JA, Rubin RH. Infection in organ-transplant recipients. *N Engl J Med* 1998;338:1741–1751.
37. Klein NC, Go CH, Cunha BA. Infections associated with steroid use. *Infect Dis Clin North Am* 2001;15:423–32.
38. Giles JT, Bathon JM. Serious infections associated with anticytokine therapies in the rheumatic diseases. *J Intensive Care Med* 2004;19:320–334.
39. Vial T, Descotes J. Immunosuppressive drugs and cancer. *Toxicology* 2003;185:229–240.
40. Georgescu L, Paget SA. Lymphoma in patients treated with rheumatoid patients. What is the evidence of link with methotrexate? *Drug Saf* 1999;20:475–487.
41. Brown SL, Greene MH, Gershon SK, Edwards ET, Braun MM. Tumor necrosis factor antagonist therapy and lymphoma development: twenty-six cases reported to the Food and Drug Administration. *Arthritis Rheum* 2002;46:3151–3158.

42. Descotes J, Vial T. The flu-like syndrome and cytokines. In House RV, Descotes J, Eds. *Cytokines in Human Health: Immunotoxicology, Pathology and Therapeutic Applications*. Totowa, NJ: Humana Press; 2006.
43. Krause I, Valesini G, Scrivo R, Shoenfeld Y. Autoimmune aspects of cytokine and anti-cytokine therapies. *Am J Med* 2003;115:390–397.
44. Shulman KL, Thompson JA, Benyunes MC, Winter TC, Fefer A. Adverse reactions to intravenous contrast media in patients treated with interleukin-2. *J Immunother* 1993; 13:208–212.
45. Renton KW. Hepatic drug metabolism and immunostimulation. *Toxicology* 2000; 142:173–178.
46. Gomes ER, Demoly P. Epidemiology of hypersensitivity drug reactions. *Curr Opin Allergy Clin Immunol* 2005;5:309–316.
47. Choquet-Kastylevsky G, Vial T, Descotes J. Drug allergy diagnosis in humans: possibilities and pitfalls. *Toxicology* 2001;158:1–10.
48. Johansson SG, Hourihane JO, Bousquet J, Bruijnzeel-Koomen C, Dreborg S, Haahtela T, et al. A revised nomenclature for allergy. An EAACI position statement from the EAACI nomenclature task force. *Allergy* 2001;56:813–824.
49. Solensky R. Drug hypersensitivity. *Med Clin North Am* 2006;90:233–260.
50. Descotes J, Choquet-Kastylevsky G. Gell and Coombs classification: is it still valid? *Toxicology* 2001;158:43–49.
51. Naisbitt DJ, Williams DP, Pirmohamed M, Kitteringham NR, Park BK. Reactive metabolites and their role in drug reactions. *Curr Opin Allergy Clin Immunol* 2001; 1:317–325.
52. Britschgi M, Von Greyerz S, Burkhart C, Pichler WJ. Molecular aspects of drug recognition by specific T cells. *Curr Drug Targets* 2003;4:1–11.
53. Descotes J, Payen C, Vial T. Pseudoallergic drug reactions. *Perspect Exp Clin Immunotoxicol* 2006;1:40–49.
54. Vial T, Descotes J. Autoimmunity and toxic exposures. *Perspect Exp Clin Immunotoxicol* 2006;1:2–13.
55. Kuper CF, Harleman JH, Richter-Reichelm HB, Vos JG. Histopathologic approaches to detect changes indicative of immunotoxicity. *Toxicol Pathol* 2000;28:454–466.
56. Haley P, Perry R, Ennulat D, Frame S, Johnson C, Lapointe JM, et al. STP position paper: best practice guideline for the routine pathology evaluation of the immune system. *Toxicol Pathol* 2005;33:404–407.
57. Haley PJ. Species differences in the structure and function of the immune system. *Toxicology* 2003;188:49–71.
58. Haines DM, Westn KH. Immunohistochemistry: forging the links between immunology and pathology. *Vet Immunol Immunopathol* 2005;108:151–156.
59. Kadkol SS, Gage WR, Pasternack GR. *In situ* hybridization—theory and practice. *Mol Diagn* 1999;4:169–183.
60. Hyun BH, Gulati GL, Ashton JK. Bone marrow examination: techniques and interpretation. *Hematol Oncol Clin North Am* 1988;2:513–523.
61. Chadburn A. The spleen: anatomy and anatomical function. *Semin Hematol* 2000;37: S13–S21.
62. MacDonald TT. The mucosal immune system. *Parasite Immunol* 2003;25:235–246.
63. Gergely P. Drug-induced lymphopenia: focus on $CD4^+$ and $CD8^+$ cells. *Drug Saf* 1999;21:91–100.

64. Wierda D, Irons RD. Bone marrow and immunotoxicity. In Smialowicz RJ, Holsapple MP, Eds. *Experimental Immunotoxicity*. Boca Raton, FL: CRC Press; 1996, pp 101–117.

65. Marti GE, Stetler-Stevenson M, Bleesing JJ, Fleisher TA. Introduction to flow cytometry. *Semin Hematol* 2001;38:93–99.

66. Nygaard UC, Lovik M. Blood and spleen lymphocytes as targets for immunotoxic effects in the rat—a comparison. *Toxicology* 2002;174:153–161.

67. Luster MI, Portier C, Pait DG, White KL Jr, Gennings C, Munson AE, et al. Risk assessment in immunotoxicology. I. Sensitivity and predictability of immune tests. *Fundam Appl Toxicol* 1992;18:200–210.

68. Immunotoxicology Technical Committee. Application of flow cytometry to immunotoxicity testing: summary of a workshop. *Toxicology* 2001;163:39–48.

69. Burchiel SW, Lauer FT, Gurule D, Mounho BJ, Salas VM. Uses and future applications of flow cytometry in immunotoxicity testing. *Methods* 1999;19:28–35.

70. Temple L, Kawabata TT, Munson AE, White KL. Comparison of ELISA and plaque-forming cell assays for measuring the humoral response to SRBC in rats and mice treated with benzo[*a*]pyrene or cyclophosphamide. *Fundam Appl Toxicol* 1993;21:412–419.

71. Johnson CW, Williams WC, Copeland CB, DeVito MJ, Smialowicz RJ. Sensitivity of the SRBC PFC assay versus ELISA for detection of immunosuppression by TCDD and TCDD-like congeners. *Toxicology* 2000;156:1–11.

72. Condevaux F, Horand F, Ruat C, Descotes J. Relative sensitivities of the plaque forming cell (PFC) assay and anti-KLH antibody ELISA in the same rats treated with cyclophosphamide. *Toxicol Lett* 2001;123:S58 (abstract).

73. Jerne NK, Henry C, Nordin AA, Fuji H, Koros AM, Lefkovits I. Plaque forming cells: methodology and theory. *Transplant Rev* 1974;18:130–191.

74. Vos JG, Krajnc EI, Beekhof P. Use of the enzyme-linked immunosorbent assay (ELISA) in immunotoxicity testing. *Environ Health Perspect* 1982;43:115–121.

75. Gore ER, Gower J, Kurali E, Sui J-L, Bynum J, Ennulat D, Herzyk D. Primary antibody response to keyhole limpet hemocyanin in rat as a model for immunotoxicity evaluation. *Toxicology* 2004;197:23–35.

76. Ulrich P, Paul G, Perentes E, Mahl A, Roman D. Validation of immune function testing during a 4-week oral toxicity study with FK506. *Toxicol Lett* 2004;149:123–131.

77. Piccotti JR, Alvey JD, Reindel JF, Guzman RE. T-cell-dependent antibody response: assay development in cynomolgus monkeys. *J Immunotoxicol* 2005;2:191–196.

78. Luster MI, Dean JH, Boorman GA. Cell-mediated immunity and its application to toxicology. *Environ Health Perspect* 1982;43:31–36.

79. Descotes J, Tedone R, Evreux JC. Effects of cimetidine and ranitidine on delayed-type hypersensitivity. *Immunopharmacology* 1983;6:31–35.

80. Henningsen GM, Koller LD, Exon JH, Talcott PA, Osborne CA. A sensitive delayed-type hypersensitivity model in the rat for assessing *in vivo* cell-mediated immunity. *J Immunol Methods* 1984;70:153–165.

81. Ponton J, Regulez P, Cisterna R. Comparison of three tests for measuring footpad swelling in the mouse. *J Immunol Methods* 1983;63:139–143.

82. Thilsted JP, Shifrine M. Delayed cutaneous hypersensitivity in the dog: reaction to tuberculin purified protein derivative and coccidioidin. *Am J Vet Res* 1978;39: 1702–1705.

83. Descotes J, Tedone R, Evreux JC. Immunotoxicity screening of drugs and chemicals: value of contact hypersensitivity to picryl chloride in the mouse. *Methods Find Exp Clin Pharmacol* 1985;7:303–305.
84. Descotes J, Evreux JC. Effect of chlorpromazine on contact hypersensitivity to DNCB in the guinea-pig. *J Neuroimmunol* 1982;2:21–25.
85. Smialowicz RJ. *In vitro* lymphocyte proliferation assays: the mitogen-stimulated response and the mixed-lymphocyte reaction in immunotoxicity testing. In Burleson GR, Dean JH, Munson AE, Eds. *Methods in Immunotoxicology, Volume 1*. Hoboken, NJ: Wiley-Liss; 1995, pp 197–210.
86. Wemme H, Pfeifer S, Heck R, Muller-Quernheim J. Measurement of lymphocyte proliferation: critical analysis of radioactive and photometric methods. *Immunobiology* 1992;185:78–89.
87. House RD, Dean JH. Studies on structural requirements for suppression of cytotoxic T lymphocyte induction by polycyclic compounds. *Toxicol In Vitro* 1987;1:149–162.
88. Hoppner M, Luhm J, Schlenke P, Koritke P, Frohn C. A flow-cytometry based cytotoxicity assay using stained effector cells in combination with native target cells. *J Immunol Methods* 2002;267:157–163.
89. House RV. Theory and practice of cytokine assessment in immunotoxicology. *Methods* 1999;19:17–27.
90. Meager A. Measurement of cytokines by bioassays: theory and application. *Methods* 2006;38:237–252.
91. De Jager W, Rijkers GT. Solid-phase and bead-based cytokine immunoassay: comparison. *Methods* 2006;38:294–303.
92. Langezaal I, Coecke S, Hartung T. Whole blood cytokine response as a measure of immunotoxicity. *Toxicol In Vitro* 2001;15:313–318.
93. Pala P, Hussell T, Openshaw PJ. Flow cytometric measurement of intracellular cytokines. *J Immunol Methods* 2000;243:107–124.
94. Vos JG, De Klerk A, Krajnc EI, Kruizinga W, Van Ommen B, Rozing J. Toxicity of bis(tri-*n*-butyltin)oxide in the rat. II. Suppression of thymus-dependent immune responses and of parameters of nonspecific resistance after short-term exposure. *Toxicol Appl Pharmacol* 1984;75:387–408.
95. Antonini JM, Roberts JR, Clarke RW, Yang HM, Barger MW, et al. Effect of age on respiratory defense mechanisms: pulmonary bacterial clearance in Fischer 344 rats after intratracheal instillation of *Listeria monocytogenes*. *Chest* 2001;120:240–249.
96. Rodgers K. Measurement of the respiratory burst of leukocytes for immunotoxicologic analysis. In Burleson GR, Dean JH, Munson AE, Eds. *Methods in Immunotoxicology*, Volume 2. Hoboken, NJ: Wiley-Liss; 1995, pp 67–77.
97. Verdier F, Condevaux F, Tedone R, Virat M, Descotes J. *In vitro* assessment of phagocytosis. Interspecies comparison of chemiluminescence response. *Toxicol In Vitro* 1993;7:317–320.
98. Lehmann AK, Sornes S, Halstensen A. Phagocytosis: measurement by flow cytometry. *J Immunol Methods* 2000;243:229–242.
99. Cretinon C, Condevaux F, Horand F, Descotes J. Comparison of the phagocytic activity in rats and monkeys using two commercial kits. *Toxicol Sci* 2003;72:S105 (abstract).
100. Wilkinson PC. Assays of leukocyte locomotion and chemotaxis. *J Immunol Methods* 1998;216:139–153.
101. Cederbrant K, Marcusson-Stahl M, Condevaux F, Descotes J. NK-cell activity in immunotoxicity drug evaluation. *Toxicology* 2003;185:241–250.

102. Thomas PT, Sherwood RL. Host resistance models in immunotoxicology. In Smialowicz RJ, Holsapple MP, Eds. *Experimental Immunotoxicology*. Boca Raton, FL: CRC Press; 1996, pp 29–45.
103. Dean JH, Luster MI, Boorman GA, Luebke RW, Lauer LD. Application of tumor, bacterial, and parasite susceptibility assays to study immune alterations induced by environmental chemicals. *Environ Health Perspect* 1982;43:81–88.
104. Selgrade MK, Daniels MJ, Dean JH. Correlation between chemical suppression of natural killer cell activity in mice and susceptibility to cytomegalovirus: rationale for applying murine cytomegalovirus as a host resistance model and for interpreting immunotoxicity testing in terms of risk of disease. *J Toxicol Environ Health* 1992; 37:123–137.
105. Luster MI, Portier C, Pait DG, Rosenthal GJ, Germolec DR, et al. Risk assessment in immunotoxicology. II. Relationships between immune and host resistance tests. *Fundam Appl Toxicol* 1993;21:71–82.
106. Wierda D. Can host resistance assays be used to evaluate the immunotoxicity of pharmaceuticals? *Hum Exp Toxicol* 2000;19:244–245.
107. Lebrec H, Burleson GR. Influenza virus host resistance models in mice and rats: utilization for immune function assessment and immunotoxicology. *Toxicology* 1994;91: 179–188.
108. Meade BJ, Hayes BB, Klykken PC. Development and validation of a *Listeria monocytogenes* host resistance model in female Fisher 344 rats. *Toxicol Methods* 1998;8:45–57.
109. Bradley SG. Streptococcus host resistance model. In Burleson GR, Dean JH, Munson AE, Eds. *Methods in Immunotoxicology, Volume 2*. Hoboken, NJ: Wiley-Liss; 1995, pp 159–168.
110. De Waal EJ, De Jong WH, Van der Stappen AJ, Verlaan B, Van Der Loveren H. Effects of salmeterol on host resistance to *Trichinella spiralis* in rats. *Int J Immunopharmacol* 1999;21:523–529.
111. Henry SP, Taylor J, Midgley L, Levin AA, Kornbrust DJ. Evaluation of the toxicity of ISIS 2302, a phosphorothioate oligonucleotide, in a 4-week study in CD-1 mice. *Antisense Nucleic Acid Drug Dev* 1997;7:473–481.
112. Anderson TD, Hayes TJ. Toxicity of human recombinant interleukin-2 in rats. Pathologic changes are characterized by marked lymphocytic and eosinophilic proliferation and multisystem involvement. *Lab Invest* 1989;60:331–346.
113. Cesario TC, Vaziri ND, Ulich TR, Khamiseh G, Oveisi F, et al. Functional, biochemical, and histopathologic consequences of high-dose interleukin-2 administration in rats. *J Lab Clin Med* 1991;118:81–88.
114. Sparwasser T, Hultner L, Koch ES, Luz A, Lipford GB, Wagner H. Immunostimulatory CpG-oligodeoxynucleotides cause extramedullary murine hemopoiesis. *J Immunol* 1999;162:2368–2374.
115. Helfand SC, Soergel SA, MacWilliams PS, Hank JA, Sondel PM. Clinical and immunological effects of human recombinant interleukin-2 given by repetitive weekly infusion to normal dogs. *Cancer Immunol Immunother* 1994;39:84–92.
116. Ryffel B, Weber M. Preclinical safety studies with recombinant human interleukin 6 (rhIL-6) in the primate *Callithrix jacchus* marmoset: comparison with studies in rats. *J Appl Toxicol* 1995;15:19–26.
117. Cleveland RP, Kumar A. Enhancement of anti-sheep erythrocyte plaque-forming cell levels by cimetidine *in vivo*. *Immunopharmacology* 1987;14:145–150.
118. Eastcott JW, Holmberg CJ, Dewhirst FE, Esch TR, Smith DJ, Taubman MA. Oligonucleotide containing CpG motifs enhances immune response to mucosally or systemically administered tetanus toxoid. *Vaccine* 2001;19:1636–1642.

119. Descotes J, Tedone R, Evreux JC. Effects of cimetidine and ranitidine on delayed-type hypersensitivity. *Immunopharmacology* 1983;6:31–35.

120. De Waard R, Claassen E, Bokken GC, Buiting B, Garssen J, Vos JG. Enhanced immunological memory responses to *Listeria monocytogenes* in rodents, as measured by delayed-type hypersensitivity (DTH), adoptive transfer of DTH, and protective immunity, following *Lactobacillus casei* Shirota ingestion. *Clin Diagn Lab Immunol* 2003; 10:59–65.

121. Ershler WB, Hacker MP, Burroughs BJ, Moore AL, Myers CF. Cimetidine and the immune response. I. *In vivo* augmentation of nonspecific and specific immune response. *Clin Immunol Immunopathol* 1983;26:10–17.

122. Iho S, Yamamoto T, Takahashi T, Yamamoto S. Oligodeoxynucleotides containing palindrome sequences with internal 5′-CpG-3′ act directly on human NK and activated T cells to induce IFN-gamma production *in vitro*. *J Immunol* 1999;163:3642–3652.

123. Nomoto K, Miake S, Hashimoto S, Yokokura T, Mutai M, et al. Augmentation of host resistance to *Listeria monocytogenes* infection by *Lactobacillus casei*. *J Clin Lab Immunol* 1985;17:91–97.

124. Robinet E, Morel P, Assossou O, Revillard JP. TNF-alpha production as an *in vitro* assay predictive of cytokine-mediated toxic reactions induced by monoclonal antibodies. *Toxicology* 1995;100:213–223.

125. Oshiro Y, Morris DL. TNF-alpha release from human peripheral blood mononuclear cells to predict the proinflammatory activity of cytokines and growth factors. *J Pharmacol Toxicol Methods* 1997;37:55–59.

126. Herzyk DJ, Ruggieri EV, Cunnigham L, Polsky R, Herold C, et al. Single-organism model in host defense against infection: a novel immunotoxicologic approach to evaluate immunomodulatory drugs. *Toxicol Pathol* 1997;25:351–362.

127. Cohen IR, Miller A. *Autoimmune Disease Models. A Guidebook*. San Diego, CA: Academic Press; 1994.

128. Burnett R, Ravel G, Descotes J. Clinical and histopathological progression of lesions in lupus-prone (NZB × NZW) F1 mice. *Exp Toxicol Pathol* 2004;56:37–44.

129. Adam C, Thoua Y, Ronco P, Verroust P, Tovey M, Morel-Maroger L. The effect of exogenous interferon: acceleration of autoimmune and renal diseases in (NZB/W) F1 mice. *Clin Exp Immunol* 1980;40:373–382.

130. Griffin JM, Blossom SJ, Jackson SK, Gilbert KM, Pumford NR. Trichloroethylene accelerates an autoimmune response by Th1 T cell activation in MRL+/+ mice. *Immunopharmacology* 2000;46:123–137.

131. Letter EH, Prochazka M, Coleman DL. The non-obese diabetic (NOD) mouse. *Am J Pathol* 1987;128:380–383.

132. Baxter AG, Healey D, Cooke A. Mycobacteria precipitate autoimmune rheumatic disease in NOD mice via an adjuvant-like activity. *Scand J Immunol* 1994;39:602–606.

133. Pelletier L, Hirsch F, Rossert J, Druet E, Druet P. Experimental mercury-induced glomerulonephritis. *Springer Semin Immunopathol* 1987;9:359–369.

134. White KL, David DW, Butterworth LF, Klykken PC. Assessment of autoimmunity-inducing potential using the Brown-Norway rat challenge model. *Toxicol Lett* 2000; 112–113:443–451.

135. Ewel CH, Sobel DO, Zeligs BJ, Bellanti JA. Poly I:C accelerates development of diabetes mellitus in diabetes-prone BB rat. *Diabetes* 1992;41:1016–1021.

136. Bendele A, McComb J, Gould T, McAbee T, Sennello G, et al. Animal models of arthritis: relevance to human disease. *Toxicol Pathol* 1999;27:134–142.

137. Goverman J, Brabb T. Rodent models of experimental allergic encephalomyelitis applied to the study of multiple sclerosis. *Lab Anim Sci* 1996;46:482–492.
138. Choquet-Kastylevski G, Descotes J. Value of animal models for predicting hypersensitivity reactions to medicinal products. *Toxicology* 1998;129:27–35.
139. Bala S, Weaver J, Hastings KL. Clinical relevance of preclinical testing for allergic side effects. *Toxicology* 2005;209:195–200.
140. Park BK, Kitteringham NR. Drug–protein conjugation and its immunological consequences. *Drug Metab Rev* 1990;22:87–144.
141. Pauluhn J. Predictive testing for respiratory sensitisation. *Toxicol Lett* 1996;86:177–185.
142. Koizumi K, Suzuki S, Fukuba S, Tadokoro K, Hirai K, Muranaka M. Antigenicity of semisynthetic penicillin preparations to evoke systemic anaphylactic reactions in animal models. *Allergy* 1980;35:657–664.
143. Hattori H, Yamaguchi F, Wagai N, Kato M, Nomura M. An assessment of antigenic potential of beta-lactam antibiotics, low molecular weight drugs, using guinea pig models. *Toxicology* 1997;123:149–160.
144. Shibaki A, Katz S. Induction of skewed Th1/Th2 T-cell differentiation via subcutaneous immunization with Freund's adjuvant. *Exp Dermatol* 2002;11:126–134.
145. Verdier F, Chazal I, Descotes J. Anaphylaxis models in the guinea-pig. *Toxicology* 1994;93:55–61.
146. Brutzkus B, Coquet B, Danve B, Descotes J. Systemic anaphylaxis in guinea pigs: intralaboratory validation study. *Fundam Appl Toxicol* 1997;36 (Suppl):192 (abstract).
147. *ICH Guideline S6. Preclinical Safety Evaluation of Biotechnology-Derived Pharmaceuticals*. ICH; 1997. Document available at http://www.ich.org/LOB/media/MEDIA503.pdf.
148. Maurer T. Past, present and future of contact sensitization. *Perspect Exp Clin Immunotoxicol* 2007;1:in press.
149. Buehler EV. Delayed contact hypersensitivity in the guinea pig. *Arch Dermatol* 1965;91:171–177.
150. Basketter DA, Gerberick GF. An interlaboratory evaluation of the Buehler test for the identification and classification of contact sensitizers. *Contact Dermatitis* 1996;35:146–151.
151. Magnusson B, Kligman AM. The identification of contact allergens by animal assay, the guinea pig maximization test method. *J Invest Dermatol* 1969;52:268–276.
152. Gad SC, Dunn BJ, Dobbs DW, Reilly C, Walsh RD. Development and validation of an alternative dermal sensitization test: the Mouse Ear Swelling Test (MEST). *Toxicol Appl Pharmacol* 1986;84:93–114.
153. Gad SC. The Mouse Ear Swelling Test (MEST) in the 1990s. *Toxicology* 1994;93:33–46.
154. Kimber I, Mitchell JA, Griffin AC. Development of a murine local lymph node assay for the determination of sensitization potential. *Fundam Chem Toxicol* 1986;24:585–586.
155. Kimber I, Hilton J, Botham PA, Basketter DA, Scholes EW, Miller K, Robbins MC, Harrison PT, Gray TJ, Waite SJ. The murine local lymph node assay: results of an interlaboratory trial. *Toxicol Lett* 1991;55:203–213.
156. Loveless SE, Ladics GS, Gerberick GF, Ryan CA, Basketter DA, Scholes EW, House RV, Hilton J, Dearman RJ, Kimber I. Further evaluation of the local lymph node assay in the final phase of an international collaborative trial. *Toxicology* 1996;108:141–152.

157. OECD. Skin sensitisation: Local Lymph Node Assay. *OECD Guideline for the Toxicity Testing of Chemicals 429*. Paris: OECD; 2002.

158. Kimber I, Dearman RJ, Basketter DA, Ryan CA, Gerberick GF. The local lymph node assay: past, present and future. *Contact Dermatitis* 2002;47:315–328.

159. Boussiquet-Leroux C, Durand-Cavagna G, Herlin K, Holder D. Evaluation of lymphocyte proliferation by immunohistochemistry in the local lymph node assay. *J Appl Toxicol* 1995;15:465–475.

160. Hatao M, Hariya T, Katsumura Y, Kato S. A modification of the local lymph node assay for contact allergenicity screening: measurement of interleukin-2 as an alternative to radioisotope-dependent proliferation assay. *Toxicology* 1995;98:15–22.

161. Suda A, Yamashita M, Tabei M, Taguchi K, Vohr HW, Tsutsui N, Suzuki R, Kikuchi K, Sakaguchi K, Mochizuki K, Nakamura K. Local lymph node assay with non-radioisotope alternative endpoints. *J Toxicol Sci* 2002;27:205–218.

162. Dearman RJ, Scholes EW, Ramdin LS, Basketter DA, Kimber I. The local lymph node assay: an interlaboratory evaluation of interleukin 6 (IL-6) production by draining lymph node cells. *J Appl Toxicol* 1994;14:287–291.

163. Dearman RJ, Hilton J, Basketter DA, Kimber I. Cytokine endpoints for the local lymph node assay: consideration of interferon-gamma and interleukin 12. *J Appl Toxicol* 1999;19:149–155.

164. Ryan CA, Gerberick GF, Cruse LW, Basketter DA, Lea L, Blaikie L, Dearman RJ, Warbrick EV, Kimber I. Activity of human contact allergens in the murine local lymph node assay. *Contact Dermatitis* 2000;43:95–102.

165. Ulrich P, Streich J, Suter W. Intralaboratory validation of alternative endpoints in the murine local lymph node assay for the identification of contact allergic potential: primary ear skin irritation and ear-draining lymph node hyperplasia induced by topical chemicals. *Arch Toxicol* 2001;74:733–744.

166. Gerberick GF, Cruse LW, Ryan CA, Hulette BC, Chaney JG, Skinner RA, Dearman RJ, Kimber I. Use of a B cell marker (B220) to discriminate between allergens and irritants in the local lymph node assay. *Toxicol Sci* 2002;68:420–428.

167. Shiho O, Nakagawa Y, Kawaji H. Delayed type hypersensitivity for penicillin in mice. I. Induction and characterization of delayed type hypersensitivity for penicillin in mice. *J Antibiot* 1981;34:452–458.

168. Choquet-Kastylevsky G, Santolaria N, Tedone R, Aujoulat M, Descotes J. Induction of delayed-type hypersensitivity to sulfamethoxazole in mice: role of metabolites. *Toxicol Lett* 2001;119:183–192.

169. Vial T, Descotes J. Contact sensitization assays in guinea-pigs: are they predictive of the potential for systemic allergic reactions? *Toxicology* 1994;93:63–75.

170. Weaver JL, Staten D, Swann J, Armstrong G, Bates M, Hastings KL. Detection of systemic hypersensitivity to drugs using standard guinea pig assays. *Toxicology* 2003; 193:203–217.

171. Eschalier A, Lavarenne J, Burtin C, Renoux M, Chapuy E, Rodriguez M. Study of histamine release induced by acute administration of antitumor agents in dogs. *Cancer Chemother Pharmacol* 1988;21:246–250.

172. Guedes AG, Rude EP, Rider MA. Evaluation of histamine release during constant rate infusion of morphine in dogs. *Vet Anaesth Analg* 2006;33:28–35.

173. Lorenz W. Histamine release in man. *Agents Actions* 1975;5:402–416.

174. Sanz ML, Maselli JP, Gamboa PM, Oehling A, Dieguez I, De Weck AL. Flow cytometric basophil activation test: a review. J Invest Allergol *Clin Immunol* 2002;12:143–154.

175. Szebeni J, Muggia FM, Alving CR. Complement activation by Cremophor EL as a possible contributor to hypersensitivity to paclitaxel: an *in vitro* study. *J Natl Cancer Inst* 1998;90:300–306.
176. Descotes J. Autoimmunity and toxicity testing. *Toxicol Lett* 2000;112–113:461–465.
177. Verdier F, Patriarca C, Descotes J. Autoantibodies in conventional toxicity testing. *Toxicology* 1997;119:51–58.
178. Kuper CF, Schuurman H, Bos-Kuijpers M, Bloksma N. Predictive testing for pathogenic autoimmunity: the morphological approach. *Toxicol Lett* 2000;112–113:433–442.
179. Pollard KM, Pearson DL, Hultman P, Hildebrandt B, Kono DH. Lupus-prone mice as models to study xenobiotic-induced acceleration of systemic autoimmunity. *Environ Health Perspect* 1999;107 (Suppl 5):729–735.
180. Ravel G, Descotes J. The popliteal lymph node assay: facts and perspectives. *J Appl Toxicol* 2005;25:451–458.
181. Vial T, Carleer J, Legrain B, Verdier F, Descotes J. The popliteal lymph node assay: results of a preliminary interlaboratory validation study. *Toxicology* 1997;122:213–218.
182. Choquet-Kastylevsky G, Descotes J. Popliteal lymph node responses to ethanol and acetone differ from those induced by streptozotocin. *Arch Toxicol* 2004;74:649–654.
183. Choquet-Kastylevsky G, Tedone R, Descotes J. Positive responses to imipramine in the popliteal lymph node assay are due to primary irritation. *Hum Exp Toxicol* 2001;20: 591–596.
184. Nagata N, Hurtenbach U, Gleichmann E. Specific sensitization of Lyt-$1^{+}2$ T-cells to spleen cells modified by the drug D-penicillamine or a stereoisomer. *J Immunol* 1986;136:136–142.
185. Kubicka-Muryani M, Goebels R, Goebel C, Uetrecht JP, Gleichmann E. T lymphocytes ignore procainamide, but respond to its reactive metabolites in peritoneal cells: demonstration by the adoptive transfer popliteal lymph node assay. *Toxicol Appl Pharmacol* 1993;122:88–94.
186. Albers R, Broeders A, Van Der Pijl A, Seinen W, Pieters R. The use of reporter antigens in the popliteal lymph node assay to assess immunomodulation by chemicals. *Toxicol Appl Pharmacol* 1997;143:102–109.
187. Pieters R, Ezendam J, Bleumink R, Bol M, Nierkens S. Predictive testing for autoimmunity. *Toxicol Lett* 2002;127:83–91.
188. Gutting BW, Updyke LW, Amacher DE. Investigating the TNP-OVA and direct popliteal lymph node assays for the detection of immunostimulation by drugs associated with anaphylaxis in humans. *J Appl Toxicol* 2002;22:177–183.
189. Descotes J. Regulatory aspects of immunotoxicity evaluation. In Descotes J, Ed. *Immunotoxicology of Drugs and Chemicals. An Experimental and Clinical Approach. Vol. 1. Principles and Methods of Immunotoxicology*. Amsterdam: Elsevier Science; 2004, pp 257–268.
190. *OECD Guideline for the Testing of Chemicals. Repeated Dose 28-day Oral Toxicity Study in Rodents OECD*; 1995. Document available at http://www.oecd.org/dataoecd/17/52/1948386.pdf.
191. *Harmonized Tripartite Guideline S6: Preclinical Safety Evaluation of Biotechnology-Derived* Pharmaceuticals. ICH; 1996. Document available at http://www.ich.org/MediaServer.jser?@_ID=503&@_MODE=GLB.
192. *Note for Guidance on Repeated Dose Toxicity*. EMEA; 2000. Document available at http://www.emea.eu.int/pdfs/human/swp/104299en.pdf.
193. *Immunotoxicology Evaluation of Investigational New Drugs*. FDA; 2002. Document available at http://www.fda.gov/cder/guidance/4945fnl.pdf.

194. *Harmonized Tripartite Guideline S8: Immunotoxicity Studies for Human Pharmaceuticals*. ICH; 2005. Document available at http://www.ich.org/MediaServer.jser?@_ID=1706&@_MODE=GLB.
195. Descotes J. Nonclinical strategies of immunotoxicity evaluation and risk assessment. In Descotes J, Ed. *Immunotoxicology of Drugs and Chemicals. An Experimental and Clinical Approach. Vol. 1. Principles and Methods of Immunotoxicology*. Amsterdam: Elsevier Science; 2004, pp 269–293.
196. Basketter DA, Bremmer JN, Buckley P, Kammuller ME, Kawabata T, Kimber I, et al. Pathology considerations for, and subsequent risk assessment of, chemicals identified as immunosuppressive in routine toxicology. *Food Chem Toxicol* 1995;33:239–243.
197. Gopinath C. Pathology of toxic effects on the immune system. *Inflamm Res* 1996; 2 (Suppl):74–78.
198. Crevel RW, Buckley P, Robinson JA, Sanders IJ. Immunotoxicological assessment of cyclosporin A by conventional pathological techniques and immune function testing in the rat. *Hum Exp Toxicol* 1997;16:79–88.
199. Putman E, Van Loveren H, Bode G, Dean J, Hastings K, Nakamura K, Verdier F, Van Der Laan JW. Assessment of the immunotoxic potential of human pharmaceuticals: a workshop report. *Drug Inf J* 2002;36:417–427.
200. Germolec DR, Kashon M, Nyska A, Kuper CF, Portier C, Kommineni C, Johnson KA, et al. The accuracy of extended histopathology to detect immunotoxic chemicals. *Toxicol Sci* 2004;82:504–514.
201. Calabrese EJ. Hormetic dose–response relationships in immunology: occurrence, quantitative features of the dose response, mechanistic foundations, and clinical implications. *Crit Rev Toxicol* 2005;35:89–295.
202. Buhles WC. Application of immunologic methods in clinical trials. *Toxicology* 1998; 129:73–89.

# IMMUNOTOXICITY TESTING: ICH GUIDELINE S8 AND RELATED ASPECTS

Hans-Gerd Pauels[1] and John Taylor[2]

[1]*Dr. Pauels–Scientific and Regulatory Consulting, Münster, Germany*
[2]*ProPhase Development Ltd, Harrogate, United Kingdom*

**Contents**

1 Introduction
2 Categories of Immunotoxicity
  2.1 Immunosuppression
  2.2 Unintended Immunostimulation
  2.3 Immunogenicity
  2.4 Allergenicity
  2.5 Autoimmunity
3 Regulatory Considerations
  3.1 History of Regulatory Guidelines
  3.2 Current Developments for Pharmaceuticals
  3.3 The Common Spirit of Regulatory Immunotoxicity Guidelines
4 Immunotoxicity Testing Strategies
  4.1 Initial Considerations
  4.2 Timing of Immunotoxicity Studies
  4.3 Choice of Test Species
  4.4 Immunotoxicity Alerts
  4.5 Decision Tree
5 Immunotoxicity Testing Protocols: Immunosuppression
  5.1 Basic Testing
  5.2 Additional Testing
  References

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

## 1 INTRODUCTION

The physiological role of the immune system is to fight viral, bacterial, and parasitic infections, and perhaps also cancer. The latter, however, is a matter of debate because, with the exception of some rare tumors [1–3], no clear-cut correlation exists between immunosuppression and tumor incidence. Toxicological research over the past decades has indicated that the immune system is a potential target for toxic damage due to xenobiotics, and immunotoxicology has evolved as a discipline concerned with adverse interactions of chemicals and drugs with the immune system.

The immunotoxic effects of chemicals encompass five categories: (1) unintended immunosuppression, (2) unintended immunostimulation, (3) immunogenicity, (4) induction of hypersensitivity, and (5) induction of autoimmunity. The last three categories, however, are variations of the same theme and are discussed together in this chapter.

Immunotoxicology as a discipline has evolved mainly from the 1980s with human health immunotoxicity risk assessments of chemicals, pollutants, and pesticides. Such risk assessments have focused on ensuring that adequate safety factors are applied to dose–response effects such that immunotoxicological risks are negligible for these substances. Various agencies/institutes have been concerned particularly from this period onward in developing immunotoxicity testing strategies and risk assessment processes, as outlined in Section 3. Immunotoxicity risk assessment has been integrated in a more general way for pharmaceuticals up until the late 1990s when guidelines were initiated in the ICH regions, culminating in the development of the ICH S8 guideline entitled *Immunotoxicity Studies for Human Pharmaceuticals*, which came into operation in Europe and in the United States in spring 2006. Since immunotoxicity plays a role in some drug-induced adverse effects, resulting in morbidity and mortality, as noted for drugs such as cyclosporine and anticancer cytotoxic agents and more recently for TNF-α antagonists, it is recognized that immunotoxicity risk assessment is an important element of drug safety assessment. The introduction of guidelines including ICH S8 for drugs further supports the clinical risk–benefit assessments that are made for drugs.

Immunotoxicologists have long focused on immunosuppression, but other aspects of immunotoxicity like immunogenicity, induction of autoimmunity, or hypersensitivity have become increasingly important for drugs, especially with the introduction of biotechnology-derived drugs into the pharmaceutical pipeline. A number of well standardized and validated animal models exist for the nonclinical assessment of immunosuppression. Furthermore, some predictive *in vitro* and *in vivo* assays are available for the assessment of unintended immunostimulation. In both cases, many additional techniques can be used to extend the assessment of immunotoxic mechanistic effects on a case-by-case basis. The situation is less satisfactory for the assessment of immunogenicity, allergenicity, and autoimmunity. The main problem, in such cases, is the limited predictive value of animal models for the outcome of a drug treatment in humans.

Over the past decades, several guidelines have been issued by regulatory agencies to integrate immunotoxicity assessment into the risk assessment of chemicals, pollutants, and drugs. To date, these guidelines take a different position whether to make immunotoxicity tests a mandatory step during preclinical safety assessment,

or whether the decision on immunotoxicity testing should be based on a weight of evidence basis. Furthermore, different recommendations are given with regard to the kinds of assays that should be employed.

In the pharmaceutical sector, all three ICH regions have made strong efforts to harmonize the immunotoxicity risk assessment for investigational new drugs. These efforts culminated in the release of the ICH S8 guideline, which was adopted by the CHMP in October 2005 and came into operation in April 2006. It was furthermore adopted by the FDA and published in the Federal Register in April 2006, as well as by the MHLW in April 2006. According to this current guideline, initial immunotoxicity assessment should be based on the evaluation of data already available from standard toxicity studies and other characteristics of the drug substance like pharmacological properties, the intended patient population, known drug class effects, clinical data for the drug, and its disposition. The need for additional immunotoxicity testing should be decided on the basis of a weight of evidence assessment, taking into account all available information. Testing should thus be a tiered approach, triggered and determined by concerns from the weight of evidence assessment.

Immunotoxicity is defined in the ICH S8 guideline as unintended immunosuppression or enhancement. It must be noticed that drug-induced hypersensitivity and autoimmunity are not in the scope of ICH S8. Furthermore, the ICH S8 guideline excludes biological and biotechnology-derived products. Nevertheless, these aspects are also addressed in this chapter.

## 2 CATEGORIES OF IMMUNOTOXICITY

### 2.1 Immunosuppression

The immune system is composed of a multitude of cooperating organ systems, cell types, and soluble factors, which are distributed over the entire body, and which normally are in a state of delicate balance. Furthermore, the immune system encompasses a plethora of different signal transduction pathways. As a consequence, multiple entry points exist for unintended noxious actions, and most compounds that interfere with certain signal transduction pathways are likely to find some target structures somewhere in the immune system. Furthermore, since the immune system requires a substantial portion of the body's energy and resources, all metabolic inhibitors are likely to cause an effect on the immune system. In addition, the immune system is closely interrelated with the peripheral and central nervous system [4]. Thus, care must be taken to differentiate noxious toxicological actions on the immune system from secondary effects related to general stress [5].

Immune defense is based on two arms—the evolutionary old arm, called the innate immune system, and the evolutionary younger arm, called the adaptive immune system. In vertebrates, both arms of the immune system are in close interplay, and the adaptive immune system could not work without the sentinel and preparative work of its innate counterpart. A potent innate immune system is found in invertebrates and in vertebrates. Receptors of the innate arm are designed to

recognize invariable molecular patterns of microorganisms. A good example of such evolutionary old receptors is the family of toll-like receptors, which have already been identified to be involved in immune defense in *Drosophila*, and which are still in use in humans for the same purposes. In a wider sense, the innate immune system encompasses all physical, chemical, and cellular barriers that protect the individual from microbial infections without the need to learn to discriminate self from nonself. The body protects itself from dangerous actions of the innate immune system by the lack of expression of molecular patterns of microorganisms and by the abundant expression of inhibitors. Most components of the innate immune system can work independently and in parallel to destroy microorganisms (Fig. 1). Thus, due to its redundancy, the innate immune system is rather robust regarding its afferent and efferent actions.

The adaptive immune system is only found in vertebrates. Receptors of the adaptive arm are generated by random recombination and mutation processes and sub-

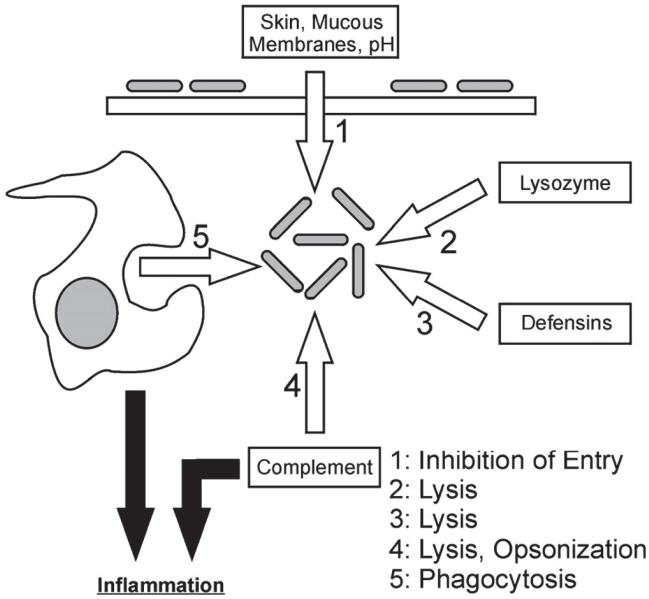


**FIGURE 1** Parallel actions of the innate immune system during antibacterial defense. Innate defense against bacteria is composed of at least five independent components. (1) Invasion of bacteria from external body surfaces is hampered by keratinized skin, low dermal pH, ciliated epithelia, or mucus layers. (2) Bacterial murein cell walls are digested by the enzymatic activity of the ubiquitous enzyme lysozyme. (3) Bacteria are attacked by antibiotic peptides like defensins (e.g., by pore formation). (4) Bacteria are attacked by the alternative and lectin pathway of the complement cascade, leading to pore formation and 3Cb deposition on the cell wall. (5) Bacteria are recognized by macrophages via receptors for microbial molecular patterns (e.g., LPS receptor, mannose-fucose receptor, scavenger receptor), or via complement receptors detecting deposited C3b. Bacteria are eliminated by macrophages via phagocytosis and intracellular digestion. Both activation of macrophages and activation of the complement system are associated with the release of proinflammatory factors like TNF-α (macrophages) or C3a and C5a (complement).

sequent selection. The adaptive immune system is designed to learn to recognize any foreign antigen that may be expressed by a microorganism. The fact that the adaptive arm of the immune system depends on delicate tolerance mechanisms to shape the repertoire of the randomly produced antigen receptors makes this part of the immune system susceptible for the induction of autoimmunity. In contrast to innate immune responses, induction of an adaptive immune response depends on a well regulated temporal and spatial sequence of activation events, which requires the cooperation of several cell types, and the migration of cells between different organs (Fig. 2). As a consequence, activation of adaptive immune responses is a rather slow process, as compared with the activation of innate defense mechanisms. All cellular interactions during the induction of an adaptive immune response are based on differentiated intercellular communications pathways involving a plethora of cytokines and surface receptors. Thus, the four basic types of adaptive immune responses—antigen-specific antibody responses, cytotoxic T cell (CTL) responses, as well as $CD4^+$ inflammatory ($T_H1$) and helper T cell ($T_H2$) responses—offer a variety of entry points for immunotoxic actions. A simplified overview of possible entry points for immunotoxic actions, as exemplified by the induction of an antibody response, is given in Table 1. As a general rule, the adaptive immune system is more sensitive to the action of immunotoxicants than the innate immune system.

The term immunosuppression refers to the impairment of any component of the immune system resulting in decreased immune function. Immunosuppression may be an intended primary pharmacodynamic action of certain pharmaceuticals (immunosuppressants such as cyclosporine, anti-inflammatory drugs such as corticoids). In such cases, a substantial body of data relating to immunotoxic actions may already be available from basic pharmacodynamic studies. Thus, the need for additional formal immunotoxicity studies may be discussed on a case-by-case basis. The real focus of immunotoxicity assessment is the detection, description, and quantitation of unintended or unforeseen immunotoxic effects.

A critical factor for the risk assessment of potential immunosuppressive compounds is the duration of suppression. The major adverse consequences of long-term immunosuppression are infectious complications and virus-induced malignancies. Short episodes of immunosuppression can be well tolerated by otherwise healthy individuals. However, unless sufficient anti-infective comedication can be provided, even short-term immunosuppression must be avoided in situations of preexisting infectious disease such as parasitoses, systemic mycoses, tuberculosis, and especially viral infections. Therefore, for every drug, the potential to induce unintended immunosuppression should be known, and assessment of potential suppressive effects on the immune system must be regarded as an important component of the overall evaluation of drug safety.

The most robust methods available for practical use in the assessment of adverse drug effects on immune function are those designed to detect and evaluate immunosuppression. For most small molecules, nonclinical assessment of immunosuppression can be performed in small rodent species. Additional general toxicological assessments are also performed using a second (nonrodent) species, and indicators of immunotoxicity are assessed also in the second species. In the case of biotechnology-derived pharmaceuticals like recombinant cytokines or chemokines, however, species specificity of the respective receptors and ligands, as well as immunogenicity

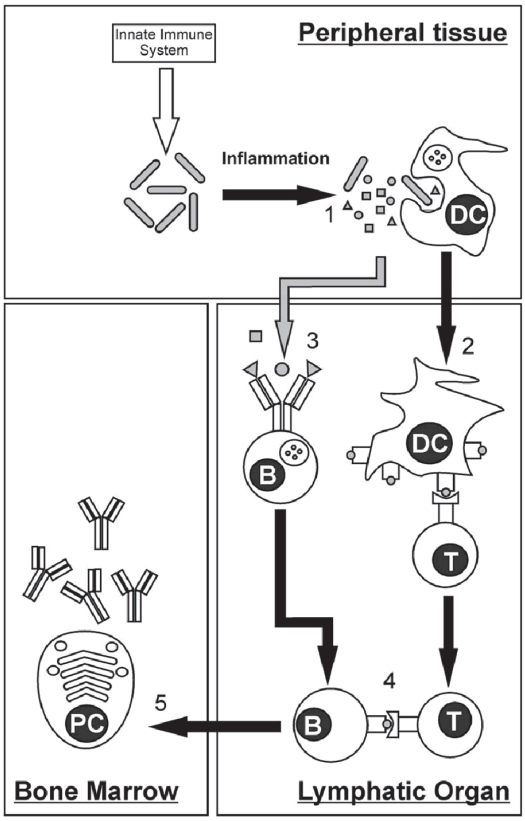


**FIGURE 2** Temporal and spatial sequence of events leading to the induction of an antibody response. At the first instance, the innate immune system must associate invaded microorganisms with "danger" and cause inflammation. (1) Inflammatory signals instruct immature dendritic cells (DCs) to ingest, process, and finally present external microbial antigens with MHC class I and class II molecules. The immature DCs subsequently differentiate into mature DCs and migrate from the site of antigen exposure to proximal lymphatic organs. (2) In lymphatic organs mature DCs present antigen to resting T cells and stimulate them for proliferation and differentiation into activated T effector cells. (3) B cells have to bind antigens with their membrane immunoglobulins, internalize and process the ingested antigens, and finally present the processed antigens with MHC class II molecules. (4) Inside the lymphatic organ antigen-specific B cells and $CD4^+$ T helper cells (specific for the nominal antigen in processed form) have to enter into direct cellular cross-talk via membrane receptors and cytokines. Depending on the kind of information exchanged between both cells, the B cell gets instructed to proliferate and differentiates into an antibody secreting plasma cell (PC). (5) For long-term antibody production, the PC has to leave the lymphatic organ and find its way to the bone marrow. Note that this picture is still incomplete: antigen presentation to B cells by interdigitating DCs in lymphatic organs is required for affinity maturation of the antibody response; also, antibodies themselves are marker molecules rather than effector molecules. Bound antibodies identify microorganisms for subsequent destruction by the complement cascade via the classical pathway: for antibody-dependent cellular cytotoxicity (ADCC) by NK cells via Fcγ receptors; for phagocytosis by neutrophils or macrophages via Fcγ receptors; or for destruction (e.g., parasitic worms) by eosinophils via Fcε receptors.

**TABLE 1 Summary of Key Events Required for the Induction of a T-Cell-Dependent Humoral Immune Response and Possible Entry Points for Immunotoxic Actions**

| Organ | APC | B Cell | $CD4^+$ T Cell | Possible Immunotoxic Entry Point | Possible Immunotoxic Outcome |
|---|---|---|---|---|---|
| Bone marrow | Differentiation from pluripotent stem cells | Differentiation from pluripotent stem cells | Differentiation from pluripotent stem cells | • Inhibition of cell proliferation<br>• Interference with signal transduction<br>• Genotoxicity | • Leukopenia<br>• Malignancies |
| | | Rearrangement of antigen receptors | Rearrangement of antigen receptors | • Interference with gene recombination<br>• Interference with DNA repair | • Lymphopenia |
| | | Central repertoire selection | | • Interference with signal transduction<br>• Interference with apoptosis | • Selective lymphopenia<br>• Autoimmunity |
| | | Homing, antibody production | | • Interference with signal transduction<br>• Inhibition of protein synthesis | • Impaired adaptive immune response |
| Thymus | | | Proliferation and positive/negative repertoire selection | • Interference with signal transduction<br>• Interference with apoptosis | • Selective lymphopenia<br>• Autoimmunity |
| Peripheral tissue | Antigen uptake and processing | | | • Inhibition of cell motility<br>• Interference with phagolysosomal degradation<br>• Interference with intracellular transport | • Lack of antigen presentation<br>• Impaired adaptive immune response |
| | Maturation | | | • Inhibition of cell motility<br>• Interference with signal transduction | • Lack of antigen presentation<br>• Impaired adaptive immune response |
| Lymphatic organ | Antigen presentation | Antigen uptake and presentation | Antigen recognition (APC) and activation; proliferation; differentiation | • Inhibition of cell motility<br>• Interference with phagolysosomal degradation or intracellular transport<br>• Interference with signal transduction | • Impaired adaptive immune response |
| | | Cross-communication with T cell | Cross-communication with B-cell | • Inhibition of cytokine release<br>• Interference with signal transduction | • Impaired adaptive immune response<br>• Autoimmunity; allergy |
| | | Activation, proliferation, differentiation | | • Inhibition of cell proliferation<br>• Interference with signal transduction | • Impaired adaptive immune response |

of the compound itself, must be considered when choosing an appropriate test system.

### 2.2 Unintended Immunostimulation

As for immunosuppression, immunostimulation may be an intended primary pharmacodynamic action of drugs. Immunostimulation may be a well justified therapeutic measure in situations of selective impairment of the immune system, or in situations of life-threatening tumor disease, where the potential treatment benefits surpass the risk of unwanted side effects. In general, however, immunostimulation is a double-edged sword. The nonaffected immune system is in a continuous state of homeostatic equilibrium. Disturbing this equilibrium, even by selective activation of single components of the immune system, may lead to severe immunotoxic side effects encompassing hypersensitivity, autoimmunity, or even hematological malignancies.

Some drug substances (or their impurities) may interfere with signal transduction pathways or may be ligands for activating receptors. Such substances may cause a general or selective activation of components of the immune system. Consequences of an unintended activation of the immune system may be fever, influenza-like symptoms, an increased frequency of autoimmune diseases or hypersensitivity reactions, as well as the inhibition of drug-metabolizing cytochrome P450 (CYP) enzymes [6]. Most of these consequences are related to a cytokine burst induced by the activating compound [7]. To make the situation even more complicated, activation of single components of the immune system may cause depression of others, leading to an overall immunosuppression. In general, unintended immunostimulation must be considered as nothing but dysregulation of an otherwise perfectly working immune system. Therefore, for every drug, the potential to induce unintended immunostimulation should be known.

Assessment of unintended immunostimulation *in vivo* is often associated with complex endpoints indicative of lymphoproliferative disorders, inflammation, hypersensitivity, or autoimmunity [8]. Thus, follow-up assessment of immunostimulation should ideally be done *in vitro* using defined cell populations. A number of well established procedures exist to assess immunostimulation *in vitro*, and especially in the case of complex biotechnology-derived pharmaceuticals, species specificity of the respective receptors and ligands can be much better reflected using *in vitro* approaches.

### 2.3 Immunogenicity

Adaptive immune responses fall into four major categories:

1. Antibody responses are based on B cells, which express immunoglobulins as antigen receptors on their cell surfaces, and which differentiate into antibody producing plasma cells after appropriate stimulation. Antibodies bind antigens in their native form.
2. $CD4^+$ inflammatory ($T_H1$) responses are characterized by $CD4^+$ effector T cells secreting the cytokines IL-2 and IFN-γ, but not secreting IL-4, IL-5, and IL-6.

The physiological role of $T_H1$ cells is to induce inflammation in peripheral tissues and to stimulate macrophages for microbicidal activity [9].

3. $CD4^+$ helper ($T_H2$) responses are characterized by $CD4^+$ effector T cells secreting the cytokines IL-4, IL-5, and IL-6, but not secreting IL-2 and IFN-γ. The physiological role of $T_H2$ cells is to activate B cells for antibody production [9]. Both $T_H1$ and $T_H2$ cells bind small peptide fragments from extracellular protein antigens, which are processed and presented by professional antigen presenting cells in the context of MHC class II molecules.
4. $CD8^+$ cytotoxic T cell (CTL) responses are characterized by $CD8^+$ effector T cells containing the cytolytic proteins perforin and granzyme. The physiological role of CTLs is to kill virus-infected target cells and tumor cells by direct cell contact. CTLs bind small peptide fragments from intracellular protein antigens, which are processed and presented by all nucleated cells in the context of MHC class I molecules.

There are two major concerns associated with drug immunogenicity: drug allergenicity and the ability of antidrug immune responses to alter the pharmacokinetic and pharmacodynamic profile of a drug substance. Allergy, in this context, is nothing but a special pathological manifestation of immunity, and any drug that induces an allergic reaction must first induce an immune response. Thus, immunogenicity and allergenicity of drugs have to be discussed in the same context.

The immune system is programmed to mount specific immune responses against all kinds of foreign macromolecules like proteins or polysaccharides. Furthermore, even small molecules that may act as "haptens" after binding to carrier proteins can induce immune responses. Good examples of such pharmacologically relevant haptens are penicillin and sulfonamides. When addressing the question of drug immunogenicity, one must keep in mind that almost everything can be made immunogenic, provided the appropriate physical conditions and route of administration are used. Therefore, the question is not whether a drug substance is potentially immunogenic; the question is rather whether a drug substance in its given formulation, dosage, posology, and route of administration can induce an immune response.

Assessment of immunogenicity requires a complete and intact immune system. No *in vitro* procedure is currently in sight to determine the immunogenic potential of a drug, or to replace animal experimentation in this context. However, it must be noted that even the currently available nonclinical *in vivo* immunogenicity tests have only a limited predictive value for the human situation. The ability of a substance to induce an immune response in a given species or even a given individual depends on a variety of factors encompassing the following:

- The chemical nature and dimension of the molecule.
- The ability to form covalent or noncovalent complexes with carrier molecules in the recipient (relevant for small "hapten" molecules).
- The ability to be processed and to be presented with molecules encoded by the recipient's major histocompatibility complex (relevant for proteins and peptides).

- The similarity of the molecule with "self"-antigens of the recipient, and thus the depth of preexisting immunological tolerance of the recipient.
- The site, dose, and frequency of administration.
- The physical form and formulation of the molecule.
- The biological activity of the molecule.
- The genetic background of the recipient.

It is a well known phenomenon that a certain molecule, when administered subcutaneously or intraperitoneally, may cause a vigorous immune response, while the same molecule, when administered via the oral route, causes a state of immunological tolerance. It is furthermore well known to immunologists that mice of a given strain can be good "responders" to a certain antigen, whereas mice of a different strain are "nonresponders" to the same antigen. Thus, when an immunogenicity test in one species gives a negative result, the test should be repeated in at least one further appropriate species to provide a reasonable assessment of immunogenic potential.

As a very rough rule, the predictive value of a nonclinical immunogenicity test increases with increasing "xenogenicity" of the molecule to be assessed. Thus, a molecule that is as foreign to a mouse or guinea pig as it is to humans—for example, an antibiotic or a protein from snake venom—is likely to give predictive results for the human situation in nonclinical immunogenicity investigations. On the other hand, injecting a recombinant human protein into a guinea pig would be a rather senseless approach to assess its potential immunogenicity in humans. The guinea pig is likely to mount an immune response against this foreign protein in such circumstances. Using the recombinant guinea pig homolog to the human protein would be the more informative study design in this case.

### 2.4 Allergenicity

Induction of hypersensitivity resulting in allergic reactions is among the most frequent reasons for terminating a drug treatment. Allergies encompass immune-mediated allergic, as well as non-immune-mediated pseudoallergic, reactions, which can be divided into five categories, depending on the effector molecules and effector mechanisms involved.

*Type I* (immediate type) hypersensitivity is mediated by antibodies of the IgE isotype, leading to mast cell degranulation after passive adsorption to Fcε receptors on mast cells and cross-linking by allergens. There are two general subtypes of Type I reactions: systemic hypersensitivity (e.g., anaphylaxis, urticaria) and respiratory hypersensitivity (e.g., asthma).

*Type II* hypersensitivity is mediated predominantly by IgG antibodies, which bind to haptenized cell surfaces and elicit cytolytic or inflammatory reactions following binding of natural killer cells or neutrophils via Fcγ receptors.

*Type III* hypersensitivity is mediated by immune complex deposition and subsequent inflammation due to the activation of the complement system.

*Type IV* hypersensitivity (delayed type hypersensitivity, DTH) is mediated by antigen- or hapten-specific $T_H1$ cells inducing strong inflammatory reactions

after binding antigen on professional antigen presenting cells especially in the skin. A contact allergy, for example against nickel, is a typical manifestation of a Type IV hypersensitivity reaction.

*Pseudoallergies* are not immunologically mediated, but mediated by active substances that induce or facilitate histamine release from mast cells or basophils. In contrast to real allergies, pseudoallergies show a clear dose–response relationship.

Despite its high clinical relevance, evaluation of drugs for allergenic potential is extremely unreliable in nonclinical toxicology studies. First, a drug that has been proved to be immunogenic in a laboratory animal species may not necessarily be immunogenic in humans, and vice versa. Second, the subtle factors that determine whether an individual responds to an antigen with an IgG or an IgE response can hardly be extrapolated from one species to another. Thus, an immune response that manifests as an allergic response in a laboratory animal may not necessarily manifest as an allergic response in humans, and vice versa.

With regard to antibody-mediated hypersensitivity, three methods have been used extensively to assess the induction of drug-specific (Type I) anaphylactic reactions:

- The passive cutaneous anaphylaxis (PCA) assay.
- The active cutaneous anaphylaxis (ACA) assay.
- The active systemic anaphylaxis (ASA) assay.

All three assays are normally conducted in guinea pigs, which is the only rodent species that actually develops symptoms of severe anaphylactic reactions and even fatal allergic shock. However, the usefulness of these assays for the safety assessment of drugs is considered limited. Since IgE as well as IgG antibodies can cause anaphylactic reactions in guinea pigs, a positive result in any of the three assays can only be weighed as proof of immunogenicity, but not allergenicity of a drug. The PCA, ACA, and ASA assays are therefore not requested or recommended for the routine evaluation of allergenicity of investigational new drugs by any regulatory agency.

The situation is as unsatisfactory for the prediction of a Type II and Type III allergenic potential of drugs. Although there are examples of drugs that are associated with Type II and Type III hypersensitivity reactions, there are no standard nonclinical methods for predicting these effects. Manifestations of both kinds of immunopathies are often indistinguishable from direct, non-immune-mediated drug toxicity. Thus, in some instances of hemolytic anemia, vasculitis, or glomerulonephritis, which may be observed during standard toxicity studies, follow-up studies should be considered to determine if antibody-mediated immune mechanisms are involved.

Since all available nonclinical assays to assess the antibody-based allergenic potential of drugs have a limited predictivity for the human situation [10], detection of drug immunogenicity should already be considered to be a potential safety alert. Whether proven immunogenicity in a nonclinical test does in fact lead to allergic complications in patients can only be convincingly demonstrated (or excluded) in

clinical trials or even later still during postmarketing surveillance of the approved drug.

The most robust and predictive procedures available for assessment of allergenicity are those measuring the skin sensitizing potential of topically administered drug substances. In these cases, a drug has to permeate the keratinized skin, bind to MHC molecules of dermal antigen presenting cells, and stimulate $CD4^+$ T cells for proliferation and $T_H1$ differentiation. Any drug that is able to induce the above sequence of events will inevitably induce an inflammatory reaction, a so-called delayed type hypersensitivity reaction, after subsequent challenge exposure to the skin. Thus, in this special situation of dermal sensitization, a proof of immunogenicity is also a proof of allergenicity.

When a drug is intended for topical administration (dermal, ocular, vaginal, rectal), the skin sensitizing potential of the drug should be determined using an appropriate assay based on sensitization and challenge as part of nonclinical safety evaluation. The most common methods for evaluating the dermal sensitizing potential of drugs have been the Buehler assay (BA) and the guinea pig maximization test (GPMT). Both *in vivo* guinea pig based methods are reliable and have demonstrated a high correlation with known human skin sensitizers. Techniques using mice, like the mouse ear swelling test, which uses an induction and challenge pattern similar to the traditional guinea pig tests, or the murine local lymph node assay (LLNA), correlate well with traditional guinea pig tests [11, 12]. Especially the LLNA, which is designed to detect lymphoproliferation in draining lymph nodes of the exposition area instead of inflammation following challenge, gives quantitative results. Furthermore, the assay is now accepted by most regulatory agencies with regard to reduction, refinement, and replacement of animal experimentation.

Pseudoallergic (anaphylactoid) reactions, which are independent of antigen-specific immune responses, result from direct drug-mediated histamine release or complement activation. Anaphylactoid reactions can be differentiated from true IgE-mediated anaphylaxis by *in vitro* testing of drug-induced histamine release from mast cell lines [13], or by the detection of activated complement products in serum of animals showing signs of anaphylaxis [14].

### 2.5 Autoimmunity

Chemical- or drug-induced autoimmune reactions are rather rare events. However, such reactions have been described for heavy metal ions [15] and for vaccines [16]. Drug substances or their impurities may induce autoimmunity by two major mechanisms. Immunogenic drug substances or vaccines may induce autoimmune responses by cross priming for self-antigens. Other drug substances or their impurities may cause an unintended nonspecific activation of the immune system, leading to a break of peripheral self-tolerance. In analogy to allergies, autoimmune reactions may be categorized according to the effector mechanisms involved:

*Type II* autoimmunity is mediated predominantly by IgG antibodies, which bind to cell surfaces and elicit cytolysis or block cellular functions (e.g., hemolytic anemia, myasthenia gravis).

*Type III* autoimmunity is mediated by immune complex deposition, activating the complement system (e.g., lupus erythematosus).

*Type IV* autoimmunity is mediated by self-antigen-specific T cells (e.g., multiple sclerosis, type I diabetes).

There is no Type I (IgE-mediated) autoimmune disease known so far.

An assessment of the potential of a drug (or vaccine) to induce autoimmunity should always be considered when a drug substance, or its impurities, has proved to be immunogenic. The question for autoimmunity should be further addressed when the final formulation of a drug (or vaccine) contains immune response modifiers or adjuvants. However, there are no predictive standard methods for determining the potential of experimental drugs to produce autoimmune reactions [17, 18]. In such situations, special attention should be devoted to signs of glomerulonephritis, lupus-like syndrome, hemolytic anemia, and vasculitis during standard toxicity studies. Furthermore, an assessment of a basic panel of antinuclear or anticytoplasmic autoantibodies should be considered for inclusion in the study design [19].

## 3 REGULATORY CONSIDERATIONS

### 3.1 History of Regulatory Guidelines

Immunotoxicology regulatory guidelines have evolved from the 1980s. As detailed in Table 2, an overview is provided on the chronology of key programs and guidelines that have influenced the development of the science of immunotoxicology and the development of more recent guidelines for pharmaceuticals. It can be seen from this table that the U.S. Environmental Protection Agency and the U.S. National Toxicology Program have been heavily involved in the formative years in the development of tiered immunotoxicology testing guidance. Separate efforts were made in the 1980s in Europe to provide guidance on the prospective assessment of the immunotoxicological potential of chemicals including pharmaceuticals.

The spirit of these immunotoxicology programs and guidelines has been adopted for subsequent testing guidelines issued in the drug-related sector, as detailed later in the text. In particular, tiered approaches to testing have been used for many years, and recent developments for immunotoxicity testing of pharmaceuticals have focused particularly on optimizing tiered testing strategies.

### 3.2 Current Developments for Pharmaceuticals

The developing awareness of the possible unwanted immune-modulating effects of pharmaceuticals has resulted in the release of a number of regulatory guidance documents for immunotoxicity testing in all three ICH regions. Within the various guidelines that have been developed, it is clear that evaluation of the potential adverse effects of human pharmaceuticals on the immune system should be incorporated into standard drug development. Final guidelines for identifying the potential immunotoxicity of new chemical entities have been released by the European Union (EU) Committee for Proprietary Medicinal Products (CPMP) in 2000 [20] and the U.S. Food and Drug Administration Center for Drug Evaluation and Research (CDER) in 2002 [21].

**TABLE 2 Key Immunotoxicity Programs and Guidelines Prior to Contemporary Immunotoxicity Guidelines for Pharmaceuticals**

| Key Immunotoxicology Program/Guideline | Main Features of Program/Guideline |
|---|---|
| U.S. Environmental Protection Agency Office of Pesticide Program guideline for the immunotoxicology evaluation of pesticides (Subdivision M; Pesticide Assessment Guidelines), 1982 | New pesticides should be evaluated according to Tier I and Tier II immunotoxicology evaluation. |
| Council of the European Communities, Official Journal, No. L332/11, 1983 [35] | Recommendation for assessment of immunotoxicity of new medicinal products, with particular emphasis on histopathological assessment of immune system. |
| National Institute of Public Health and Environmental Protection (RIVM) in The Netherlands [36] | Tiered approach for testing based essentially on OECD Guideline 407 at Tier I. Tier II testing (cell-mediated immunity, humoral immunity, macrophage function, NK function, host resistance) performed where required to further define immunotoxic effects. |
| U.S. Environmental Protection Agency proposal for first draft revision of 1982 guideline [37] | New pesticides should be evaluated by repeat dose (>30 days) toxicity testing at Tier I (includes functional assays). Tier II would be performed if positive or uninterpretable immunotoxicity results were obtained from Tier I or if other sources indicate immunotoxicity. |
| U.S. National Toxicology Program [38–40] | Interlaboratory immunotoxicology (repeat dose) validation study in mice of 5 test substances (including diethylstilbestrol and cyclophosphamide). The study later included over 50 test substances. Tier I testing was for general toxicity and pathology and functional endpoints. Tier II testing was performed to further define immunotoxic effects. |
| Organisation for Economic Development and Cooperation (OECD), Revised Guideline 407, 1995 | Repeat dose toxicity guideline, including revised specific guidance on immunotoxicity (pathological) investigations. |
| U.S. Environmental Protection Agency Toxic Substances Control Act (TSCA) [41] | Update of the pesticides guidelines on immunotoxicity testing covering histopathological and functional endpoints. Tier I testing was for general toxicity and pathology and humoral immunity. Innate immunity (NK assay) may be performed. Tier II testing further defines immunotoxic effects. |

The current immunotoxicity guidelines focus on immunosuppression by small chemical entities and exclude biotechnology-derived pharmaceuticals. The latter are presently covered by the ICH S6 guideline, *Preclinical Safety Evaluation of Biotechnology-Derived Pharmaceuticals*, released in 1997 [22]. This guideline considers both the immunogenicity and immunotoxicity of biotechnology-derived pharmaceuticals, many of which are intended to stimulate or suppress the immune system, thus potentially modulating humoral and cell-mediated immunity. The ICH S6 guideline

excludes immunotoxicity testing of vaccines, which is covered by a number of specific guidelines focusing on the testing of vaccines or adjuvants released in the three ICH regions. For the United States, the appropriate website is www.fda.gov/cber. For the European Union (EU), the appropriate website is www.emea.europa.eu, and this contains guidance on adjuvants and vaccines. In 2003 the WHO [23] published comprehensive guidelines on the nonclinical evaluation of vaccines, with detailed information on immunotoxicity assessments.

There is still much debate about the best approach for assessing the immunotoxicity potential of new chemical entities. Much of this debate is focused on subtle differences between the FDA and CPMP guidelines with regard to the best approach for determining the risk of immunotoxicity. Efforts to harmonize immunotoxicity testing programs have culminated in development of the ICH S8 guideline, *Immunotoxicity Studies for Human Pharmaceuticals*, which was adopted by the CHMP in October 2005 and came into operation in April 2006. It was furthermore adopted by the FDA and published in the Federal Register in April 2006, as well as by the MHLW in April 2006.

Table 3 compares the new as well as the still operating guidelines for immunotoxicity testing issued by the ICH, the FDA CDER (U.S.), and the CPMP (EU).

**TABLE 3 Comparison of Current ICH, and former EU, and U.S. Immunotoxicity Guidelines**

| | ICH S8 (In operation in all ICH regions since 2006) [24] | U.S. FDA CDER (Still in operation since 2002) [21] | EU CPMP (Still in operation since 2000) [20] |
|---|---|---|---|
| Specific immunotoxicity guideline | Yes | Yes | No, included in guidance on repeat dose toxicity. |
| Drug-induced hypersensitivity, immunogenicity, and autoimmunity excluded | Yes | No, these categories are included in the guideline. | Yes. (Note: Skin sensitizing potential addressed in CPMP Note for Guidance on Non-Clinical Local Tolerance Testing, 2001 [42]). |
| Screening study(ies) required | Yes, the initial screen for potential immunotoxicity involves Standard Toxicity Studies (STSs) from short-term to chronic repeat dose studies in rodents and nonrodents. | Yes, including all standard repeat dose toxicology studies that have been performed. | Yes, screening required for all new active substances in at least one repeat dose toxicity study (duration ideally should be 28 days). Rats or mice are species of choice. |

**TABLE 3** ***Continued***

| | ICH S8 (In operation in all ICH regions since 2006) [24] | U.S. FDA CDER (Still in operation since 2002) [21] | EU CPMP (Still in operation since 2000) [20] |
|---|---|---|---|
| Screening study(ies) immunotoxicity parameters | Changes in hematology, lymphoid organ weights, histopathology of immune system, and serum globulins and increased incidences of infections and tumors should be evaluated for signs of immunotoxic potential in the STSs. | Changes in hematology, lymphoid organ weights, gross pathology and histopathology of immune system, and serum globulins and increased incidences of infections and tumors should be evaluated for signs of immunotoxic potential. | Hematology, lymphoid organ weights, histopathology of lymphoid tissues, bone marrow cellularity, distribution of lymphocyte subsets, and NK cell activity (if latter two unavailable, primary antibody response to T-cell-dependent antigen). |
| Other factors to consider in evaluation of potential immunotoxicity and the need for additional immunotoxicity studies | Pharmacological properties of drug; patient population; structural similarities to known immunomodulators; drug disposition; clinical data. | Patient population; known drug class effects (including SARs); drug pharmacokinetics; clinical data. If drug intended for HIV, immune function studies required. | None specifically included in the guideline. |
| "Follow-on"/ "Additional" immunotoxicity studies | "Additional" studies may be required depending on the "weight of evidence review" of STSs and "Other Factors." "Additional" studies addressed in 3.2, 3.3, and Appendix of guideline. | "Additional" immune function studies (Sections III.B and III.C of guideline) may be required depending on "weight of evidence review" of effects in toxicity studies and "Other Factors." | "Follow-on" functional immunotoxicity studies (Appendix B of guideline) warranted on a case-by-case basis. |
| Timing of "Follow-on"/"Additional" immunotoxicity testing in relation to clinical studies | "Additional" immunotoxicity testing, if required, should be completed before clinical Phase III, or earlier depending on the effect or the patient population. | Not specified. | Not specified. |

The U.S. Food and Drug Administration CDER [21] publication also covers developmental immunotoxicity, which should be addressed where a drug has shown immunosuppressive potential in adult animal studies. More recently, a group of immunotoxicology experts from the United States and European Union has proposed a testing framework for developmental immunotoxicity [25]. The major conclusions are that the rat is the preferred model and that validated developmental immunotoxicity methods should be incorporated into standard developmental and reproductive toxicity protocols where possible.

### 3.3 The Common Spirit of Regulatory Immunotoxicity Guidelines

All current immunotoxicity guidelines for new drug entities are based on a tiered testing approach with a basic set of parameters to be assessed during standard toxicological studies (STSs), and "Additional" or "Follow-up" investigations initiated by alerts from basic toxicological studies. As outlined in the ICH S8 guideline [24] and in the U.S. FDA CDER guideline [21], in addition to the alerts from the STSs, other causes for concern that might prompt additional immunotoxicity studies include the pharmacological properties of the drug, the intended patient population, known drug class effects, the disposition of the drug, and clinical data.

All guidelines from the different ICH regions request that basic immunotoxicity screening should encompass hematology, lymphoid organ weights, and histopathology of the lymphoid organs. All of these parameters are assessed in standard repeat dose toxicity studies. Discrepancy remains in the ICH regions on the need for a functional assay. Currently, only Europe requests functional assays in routine screening. In Europe [20], the routine screening of every compound should include distribution of lymphocyte subsets (by phenotyping) and natural killer cell activity. If these are not available, primary antibody response to a T-cell-dependent antigen may be assessed as an alternative.

The FDA CDER [21] document advocates a case-by-case approach to the need for functional assays. This approach is similar to the position taken by the ICH S8 guideline [24]. Irrespective of whether a functional assay for immunotoxicity was already integrated into the basic toxicological tests (as in Europe), a decision whether additional immunotoxicity follow-up studies are appropriate should be determined by a weight of evidence review of cause(s) for concern. It is generally agreed that a follow-up immunotoxicity testing program must be designed on a case-by-case basis to allow for a sufficient flexibility reflecting the different effects a drug may exert on the immune system. All relevant guidelines only recommend testing methods for follow-up assessment of immunotoxicity. It is at the discretion of the investigator to choose appropriate models, test methods, and protocols for follow-up immunotoxicity studies.

Preclinical programs designed to assess adverse effects of new drugs on the immune system should be based on generally accepted as well as technically and biologically validated methods. The program should furthermore follow the general rules of toxicology by reflecting the following parameters:

- Statistical and biological significance of the changes
- Severity of the effects
- Dose dependency

- Safety factor above the expected clinical dose
- Study duration
- Number of species and endpoints affected
- Changes that may occur secondarily to other factors (e.g., stress)
- Possible cellular targets and/or mechanism of action
- Doses that produce these changes in relation to doses that produce other toxicities
- Reversibility of effect(s)

It must be kept in mind that the availability of biomarkers for evaluating the immune system in clinical trials is limited, and therefore, increased emphasis should be put on preclinical data as is the case for the testing of carcinogenic potential. Preclinical toxicity data support the design of clinical trials and may trigger enhanced screening for certain immune-related effects in clinical trials. The importance of preclinical immunotoxicity risk assessment should not be underestimated, since the findings may affect the indication, patient management, and the labeling of the drug product.

## 4 IMMUNOTOXICITY TESTING STRATEGIES

### 4.1 Initial Considerations

There are some points to consider before collecting data relating to immunotoxicity from standard toxicity studies and before entering into follow-up immunotoxicity testing. The major focus of immunotoxicology is detection and evaluation of undesired effects of substances on the immune system. The prime concern is to assess the importance of these interactions with regard to human health. The basic decision for the selection of assays to be used in evaluating compounds for immunotoxicity must therefore be that the assays are standardized and recognized as validated, meaningful, and predictive for the human situation. Rationales for choosing a certain assay must reflect these criteria. A collection of recommended assays is given in the appendix of the ICH S8 guideline. However, these assays are neither binding nor necessarily suitable to answer all questions related to immunotoxicity.

As stipulated in the ICH S8 guideline, the decision on immunotoxicity testing is driven by alerts from standard toxicity studies, especially from 28 day repeat dose toxicity studies with full assessment of hematology, clinical chemistry, and histopathology. Other causes for concern that might prompt additional immunotoxicity studies include the pharmacological properties of the drug, the intended patient population, known drug class effects, the disposition of the drug, and available clinical data. When no alerts or concerns can be identified during a proper weight of evidence based risk assessment, immunotoxicity testing beyond the parameters already tested during the initial immunotoxicity screening can be omitted. However, the occurrence of immunotoxic potential later in development during clinical trials may require follow-up preclinical investigations.

The immune system is flexible and often able to utilize alternative factors and mechanisms to compensate for deficiencies in a particular immune function. For this

reason, tests in appropriate animal models will normally provide a more accurate picture of immune system competence and a more relevant indication of immunotoxic potential than *in vitro* tests in which compensatory alternative mechanisms may be lacking. However, if a sound scientific rationale can be provided as to why *in vitro* tests will suffice, then their use is encouraged to minimize *in vivo* investigations. Marked differences in the immune system do exist across species. It would therefore be advisable to use more than one species for preclinical immunotoxicity evaluation to confirm that negative as well as positive results in one species are also negative or positive in another species. The use of a rodent and a nonrodent species that are routinely required for standard repeat dose toxicity tests allows for such assessments in two species by assessment of hematology, clinical chemistry, and pathology. However, it must be noted that validated immunotoxicity follow-up testing protocols and methods in nonrodent species are, with a few exceptions, not yet available.

Immunotoxicity studies are expected to be performed in compliance with Good Laboratory Practice (GLP). This is at least binding for the initial immunotoxicity screening during standard toxicology studies, as well as for the assessment of dermal or respiratory sensitization. It is recognized, however, that some specialized assays for follow-up immunotoxicity assessment might not comply fully with GLP. Deviation from GLP rules, or assay performance in a noncertified laboratory, can be justified in such cases. From an analytical point of view, complex biological assays like those used for follow-up immunotoxicity assessment are difficult to validate with regard to accuracy and precision, as well as interassay or interlaboratory reproducibility. Thus, results should ideally not be presented as primary measurements (e.g., optical densities or counts per minute), but rather in terms of more robust derivatives like stimulation indices or percent inhibition data. Furthermore, it is generally recommended to have all necessary positive and negative controls included in the same measurement series. As a matter of course, the weak points and variabilities of an assay should be known and taken into account during the design of a study protocol and final data analysis. Dose comparisons to clinical use should ideally be based on relative body surface areas, as opposed to body weight.

### 4.2 Timing of Immunotoxicity Studies

Some data on the immunotoxic potential of the drug substance or drug product will be known before clinical investigations in humans are initiated, since repeat dose standard toxicity studies are required before human exposure. If required, additional (follow-up) immunotoxicity data should be available before a large number of patients will be treated with an investigational new drug. This means that follow-up immunotoxicity studies, when triggered by an initial weight of evidence risk analysis or by alerts from early clinical trials, are normally performed in parallel with clinical Phase I or Phase II studies. The respective data should be available before entry into clinical Phase III. Nevertheless, a need for additional immunotoxicity studies may also be triggered by unexpected findings during later stages of clinical development or postmarketing pharmacovigilance. An earlier timing of follow-up immunotoxicity studies must be considered when the intended patient group is immunocompromised, suffering from chronic or latent infections, or is affected by allergic or autoimmune disorders.

### 4.3 Choice of Test Species

As a general rule, the same species that gave immunotoxicity alerts during standard toxicity testing should also be employed for follow-up immunotoxicity testing. However, there is only limited availability of validated immunotoxicity test procedures or suitable tools (e.g., monoclonal antibodies, recombinant cytokines, ELISA systems) for otherwise common test species like dogs. In such cases a limitation to rodent tests may be justifiable. The situation is often much better for nonhuman primates. In the latter case, many of the tools designed for human use are also fully compatible with primate material. Thus, nonhuman primates, especially cynomolgus monkeys, may become the "second species" of choice for future nonclinical immunotoxicity assessment. However, in view of the necessary limitations on the use of primates for preclinical testing, primates should only be considered where nonprimate species are not acceptable for human risk assessment.

***Assessment of Immunotoxicity of Small Molecules*** Given the complexity of the immune system, a species without primary pharmacodynamic activity of the drug (due to different receptors or different drug metabolism) should not be the species of choice for immunotoxicity testing. Concordant with the general rule of drug toxicity assessment, the species should be responsive to the intended primary pharmacodynamic effects of the compound. The same holds for the expected secondary pharmacodynamic effects. For small molecules, the species of choice for immunotoxicity assessment, encompassing immunosuppression and unintended immunostimulation, is mostly the mouse or the rat. In the case of small molecules a good degree of predictivity for the human situation is mostly given. Currently, a preference is given to the rat for mechanistic immunotoxicity testing in Europe, and the mouse in the United States. This, however, has a more historic background and is neither binding nor sustained by a superiority of one system over the other. It should be pointed out, however, that the rat is predominantly used as the rodent species for standard repeat dose toxicity tests and for reproductive toxicity and developmental toxicity testing.

All guidelines in the three ICH regions, as well as the ICH S8 guideline, recommend testing for immunotoxicity in a rodent species in repeat dose toxicity studies. Initial immunotoxicity screening (hematology, clinical chemistry, and pathology) is incorporated into the standard toxicity test(s) also performed in the nonrodent species. The need for additional immunotoxicity testing would be assessed from the results of the standard toxicity tests and other factors as discussed in the ICH S8 guideline. If the preclinical test program is performed in a nonrodent species only (i.e., for a biotechnology drug), additional immunotoxicity testing should then be considered in this species if required based on the initial screen performed in the standard toxicity test(s) using this species.

In contrast to the assessment of immunosuppression and unintended immunostimulation, responsiveness to the pharmacodynamic effects of the drug substance is advisable but not mandatory for immunogenicity or allergenicity testing. However, it must be noted that the kind and strength of an immune response is to a very high degree dependent on the route of antigen administration. Different routes of antigen administration (e.g., oral vs. subcutaneous) may have a totally different outcome with regard to the induced immune response. Thus, the design of an immunogenicity or allergenicity test should always include the intended route of clinical administration.

General humoral immunogenicity of small molecules may be tested in rats and mice, provided a suitable assay system (e.g., ELISA, ELISPOT, PFC assay) is available to detect drug-specific antibodies with sufficient sensitivity. The same holds for immunogenicity assessment with regard to T cells. In this case an *ex vivo* assay for T cell activation following drug administration (e.g., determination of antigen-specific cytokine production or proliferation) may be employed. If such *ex vivo* or *in vitro* assays are not available or technically feasible, the more traditional guinea pig based *in vivo* assays for the determination of Type I and Type IV hypersensitivity (ASA, ACA, PCA, GPMT) may provide a good alternative. Just to reiterate this point, a positive result in any of the above guinea pig assays can be taken as a proof of immunogenicity of a drug but must not be weighed as a proof of allergenicity in humans. In any case, a lack of immunogenicity during nonclinical testing does not necessarily exclude immunogenicity or allergenicity in humans, and vice versa. Thus, preclinical data on immunogenicity or allergenicity of drugs in experimental animals have to be interpreted with the utmost caution.

***Assessment of Immunotoxicity and Immunogenicity/Allergenicity of Biotechnology-Derived Drugs*** The decision on a suitable species for preclinical immunotoxicity assessment of biotechnology-derived drugs must be made on a case-by-case basis. In any case, the limits of predictivity should be clearly stated in the rationales for choosing a certain assay protocol. The biological activity together with species and/or tissue specificity of many biotechnology-derived pharmaceuticals (e.g., recombinant cytokines, therapeutic antibodies) often preclude standard toxicity testing designs in commonly used species (e.g., rats and dogs). The same holds for immunotoxicity testing. The design of immunotoxicity testing programs for biotechnology-derived drugs should include the use of relevant species. A relevant species is one in which the test material is pharmacologically active due to the expression of the receptor or an epitope (in the case of monoclonal antibodies). A variety of techniques (e.g., immunochemical or functional *in vitro* tests) can be used to identify a relevant species. In some cases nonhuman primates may be the only suitable species available. When no relevant species exists, the use of transgenic mice expressing the human receptor or epitope may be accepted by regulatory agencies. The information gained from use of a transgenic mouse model expressing the human receptor is optimized when the interaction of the product and the humanized receptor has similar physiological consequences to those expected in humans.

In other cases, the use of the homologous animal protein instead of the human counterpart may be considered. While useful information may also be gained from the use of homologous proteins, it should be noted that the production process, range of impurities/contaminants, pharmacokinetics, and exact pharmacological mechanism(s) may differ between the homologous form and the product intended for clinical use. Thus, from a formalistic point of view, the test item used in such protocols is not identical to the drug substance to be assessed. Results from such studies can therefore only be weighed as "supportive data." In such situations, it is highly recommended to discuss the testing strategy with the responsible regulatory agency for scientific advice. Where it is not possible to use transgenic animal models or homologous proteins, it may be advisable to assess certain aspects of potential immunotoxicity *in vitro* using human material like PBMCs, monocyte-derived macrophages, or long-term cultivated cell lines of hematopoetic origin.

Most biotechnology-derived pharmaceuticals intended for human use are *per se* immunogenic in animals. The induction of antibody formation in animals is therefore not predictive of a potential for antibody formation in humans. In this regard, the results of, for example, a guinea pig anaphylaxis test, which is usually positive for xenogenic protein products, is not predictive for reactions in humans. Such studies are therefore considered of little value for the routine evaluation of these types of products. It must be kept in mind that even humanized proteins may be immunogenic in humans. In most cases, reliable information on immunogenicity of a biotechnology-derived drug can therefore only be obtained during clinical studies. However, immunogenicity studies in animals using biotechnology-derived drugs may yield valuable information when comparing the immunogenic potential of a test compound with a biosimilar reference compound, or between different production batches.

Even if immunogenicity assessment of biotechnology-derived pharmaceuticals has limited predictivity for the human situation, measurement of antibodies associated with the administration of biotechnology-derived drugs should always be included in the design of a repeat dose toxicity study [26]. Antibody responses should be characterized with regard to titer, number of responding animals, and neutralizing or nonneutralizing antibodies. Furthermore, the detection and quantitation of antibodies should be correlated with any pharmacological and/or toxicological changes. Specifically, the effects of antibody formation on pharmacokinetic/pharmacodynamic parameters, incidence and/or severity of adverse effects, complement activation, or the emergence of new toxic effects should be considered when interpreting the data. Attention should also be paid to the evaluation of possible pathological changes related to immune complex formation and deposition.

### 4.4 Immunotoxicity Alerts

According to the ICH S8 guideline, the decision on "Additional" immunotoxicity studies or a "Follow-up" immunotoxicity program depends on the results of a weight of evidence review of cause(s) for concern of immunotoxicity. This immunotoxicity risk assessment must reflect alerts from basic toxicological studies, pharmacological properties of the drug, the intended patient population, known drug class effects, the disposition of the drug, and clinical data (if available). Results of the entire dataset should be evaluated as to whether sufficient data are available to reasonably determine the risk of immunotoxicity. If the weight of evidence review of cause(s) for concern suggests that the risk of immunotoxicity is acceptable, then no follow-up testing might be called for.

Typical indicators of immunotoxicity, which may be observed during standard short- to long-term repeat dose toxicity studies (STSs), are summarized in Table 4. Note that all indicators can also result from non-immune-mediated toxicity of drugs and must not be weighed as a proof of immunotoxicity. Nevertheless, if such indicators are found, additional immunotoxicity studies are needed to identify and assess human risk or exclude immunotoxic mechanisms.

Drug substances encompassing primary pharmacodynamic immunosuppressants, cytostatic and cytotoxic antineoplastic drugs, and steroidal and nonsteroidal anti-inflammatory drugs, as well as primary pharmacodynamic immune response modi-

**TABLE 4 Typical Indicators of Immunotoxicity, Which May Be Observed During Standard Short- to Long-Term Repeat Dose Toxicity Studies**

| Findings | Possible Indicator of |
|---|---|
| **During the In-Life Phase** | |
| Increased frequencies of infectious disease | Immunosuppression |
| Increased frequencies of tumors in long-term studies in the absence of genotoxicity or nongenotoxic indicators of tumorigenicity (e.g., endocrine) | Immunosuppression |
| Unexpected pathological symptoms or deaths shortly after administration | Hypersensitivity |
| Strong inflammatory reactions at the site of administration | Hypersensitivity |
| **Gross Necropsy** | |
| Significant increase or decrease of size and weight of lymphatic organs | Unintended immunostimulation or immunosuppression |
| **Hematology** | |
| Changes in total or differential blood counts | Unintended immunostimulation or immunosuppression |
| Anemia | Type II hypersensitivity |
| Altered frequencies of lymphocyte subsets (flow cytometry)[a] | Unintended immunostimulation or immunosuppression |
| **Clinical Chemistry** | |
| Altered total globulin levels or albumin : globulin ratio | Unintended immunostimulation or immunosuppression |
| Changes of immunoglobulin isotype levels[a] | Unintended immunostimulation or immunosuppression |
| Reduction of hemolytic complement activity[a] | Unintended immunostimulation or Type III hypersensitivity |
| Antinuclear or anticytoplasmic antibodies[a] | Unintended immunostimulation or autoimmunity |
| **Histopathology** | |
| Changes of cellularity and/or microanatomy of lymphatic organs | Unintended immunostimulation or immunosuppression |
| Vasculitis, glomerulonephritis | Type III hypersensitivity |

[a]This parameter is normally not measured during standard toxicity studies but may be integrated when a focus is drawn on immunotoxicity assessment.

fiers like recombinant cytokines or chemokines, can be assumed to have secondary pharmacodynamic effects on the immune system that might result in immunotoxicity. For all the above drug substances, immunotoxicity studies are mandatory to assess the exact mode of immune interference, as well as the severity, duration, and reversibility of effects. However, in such cases a comprehensive body of data relating to immunotoxic actions may already be available from basic pharmacodynamic

studies, and the extent of additional formal immunotoxicity studies may be discussed on a case-by-case basis. It must be reiterated that vaccines and/or adjuvants for vaccines, even if they are not in the scope of immunotoxicity guidelines, may have severe immunotoxic side effects related to hypersensitivity and autoimmunity. Thus, immunotoxicity testing with regard to the latter parameters is stipulated in all vaccine guidelines.

Additional immunotoxicity testing is further mandatory for all drugs intended for the treatment of patients with an increased susceptibility to immunotoxic effects or with an increased risk to be harmed by immunotoxic drugs. These patient populations encompass patients with congenital and acquired immunodeficiency, as well as patients with compromised immune function due to concurrent medical treatment (e.g., antineoplastic chemotherapy, whole body irradiation, or organ/bone marrow transplantation).

Furthermore, additional immunotoxicity studies should be undertaken when drug compounds are structurally similar to compounds with known immunosuppressive properties, with compounds that show strong protein binding properties (testing in this case is required especially with regard to immunogenicity), and with compounds that tend to accumulate in cells or organs of the immune system.

Finally, unexpected clinical findings suggestive of immunotoxicity in patients exposed to the drug should call for additional nonclinical immunotoxicity assessment to identify the degree and mechanisms of immunotoxicity.

### 4.5 Decision Tree

If signs of immunotoxicity are observed in STSs and/or one of the factors relating to pharmacodynamic properties—intended patient population, known drug class effects, pharmacokinetic properties, or clinical findings indicative of immunotoxicity of the drug—apply, it is recommended that the sponsor conducts additional nonclinical studies of drug effects on immune function or provides justification (including clinical risk–benefit analysis) for not performing these evaluations. The decision tree, as recommended by the ICH S8 guideline (Fig. 3), starts with a weight of evidence review performed on information from all the factors outlined above. The aim of the review is to determine whether a cause for concern exists. A finding of sufficient magnitude in a single area should trigger additional immunotoxicity studies. Findings from two or more factors, each one of which would not be sufficient on its own, could trigger additional studies. If additional immunotoxicity studies are not performed at the appropriate stage of clinical development, the sponsor will likely be asked by regulatory agencies to provide justification. If this cannot be provided, delays in clinical development and/or regulatory approval for marketing may occur.

## 5 IMMUNOTOXICITY TESTING PROTOCOLS: IMMUNOSUPPRESSION

### 5.1 Basic Testing

The first assessment phase of immunotoxicity testing is the collection and evaluation of toxicological signs indicative of immunotoxicity from standard toxicity studies. These data are regularly generated during 28 day (or similar duration) repeat dose

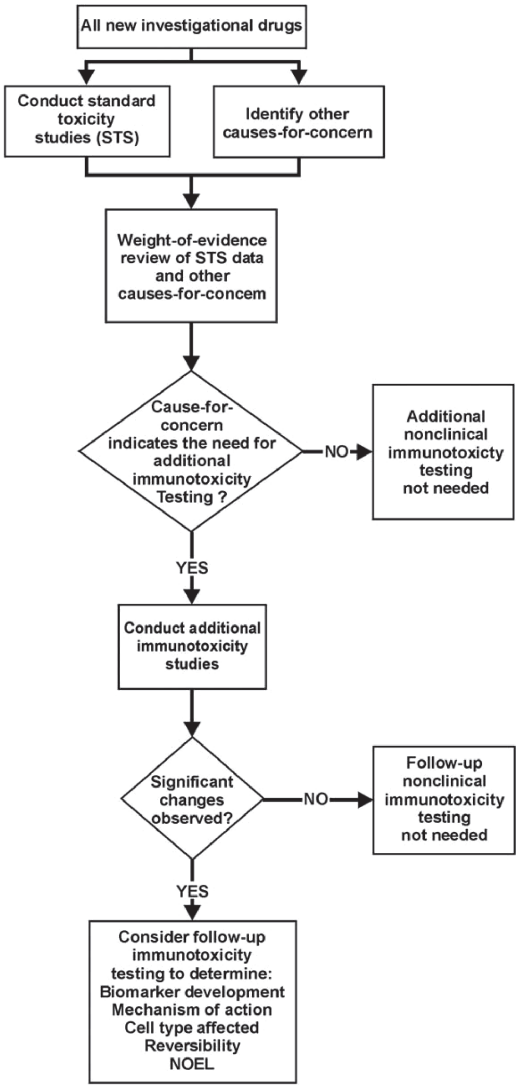


**FIGURE 3** Decision tree for the conduction of additional and follow-up immunotoxicity studies.

toxicity studies and may be further derived from longer term chronic exposure including carcinogenicity studies. Table 5 lists the relevant parameters that must be evaluated in standard toxicity studies for signs of immunotoxicity.

Total and differential leukocyte counts are highly predictive parameters to identify immunotoxicity and to obtain an idea about the involved mechanism. For example, cell loss in peripheral blood resulting from damage to bone marrow cells follows a time course that reflects the half-life of the cell type. Thus, the timing of the onset of any change in blood counts should be carefully evaluated, if possible. In the case of damage to an early stem cell, granulocytopenia is likely to be observed first, followed by thrombocytopenia. Anemia will appear much later, reflecting the

**TABLE 5 Relevant Parameters that Must Be Evaluated in Standard Toxicity Studies for Signs of Immunotoxicity**

| Mandatory Parameter | Specific Component |
|---|---|
| Hematology | Total and differential blood counts |
| Clinical chemistry | Globulin levels and A/G ratios |
| Gross pathology | Lymphoid organs/tissues |
| Organ weights | Thymus, spleen (optional: lymph nodes) |
| Histology | Thymus, spleen, draining lymph node, and at least one additional lymph node, bone marrow, Peyer's patch[a]; bronchus-associated lymphoid tissue (BALT)[b]; nasal-associated lymphoid tissue (NALT)[b] |

[a]In case of oral administration only.
[b]In case of inhalation or nasal route only.

long lifetime (approximately 120 days) of red blood cells. If the loss of a cell type is inconsistent with bone marrow damage, direct attack on mature cells might be indicated.

Changes in serum globulins can be an indication that there are changes in serum immunoglobulins. Thus, when decreased serum globulin level is observed, the protein components affected should be determined using appropriate assays. Although serum immunoglobulins are an insensitive indicator of immunosuppression, changes in immunoglobulin levels can be useful in certain situations in order to better understand target cell populations or mechanism of action.

All lymphoid tissues should be evaluated for gross changes at necropsy. However, this can be difficult for the Peyer's patches of rodents due to their small size. Spleen and thymus weights should be recorded. To minimize variability of spleen weights in dogs and monkeys, bleeding the animals thoroughly at necropsy is recommended. Atrophy of the thymus with aging can preclude obtaining accurate thymus weight.

Histopathological changes of the spleen and thymus should be evaluated as a further highly predictive indicator of systemic immunotoxicity [27]. The lymphoid tissue that drains or contacts the site of drug administration, and therefore is exposed to the highest concentration of the drug, should also be examined. These sites include the Peyer's patches and mesenteric lymph nodes for orally administered drugs, bronchus-associated lymphoid tissue (BALT) for drugs administered by the inhalation route, nasal-associated lymphoid tissue (NALT) for drugs administered by the inhalation or nasal route, and the most proximal regional draining lymph nodes for drugs administered by the dermal, intramuscular, intradermal, or subcutaneous routes. For intravenously administered drugs, the spleen can be considered the draining lymphoid tissue. It is recommended that a "semiquantitative" description of changes in compartments of lymphoid tissues be used in recording changes and reporting treatment-related changes in lymphoid tissues (e.g., –, no changes; +, mild-to-moderate changes; ++, marked changes; +++, severe changes). The parameters that should be evaluated in immunorelevant organs with regard to immunotoxic changes are summarized in Table 6.

Other indicators of immunosuppression in nonclinical toxicology studies include treatment-related infections and increased incidences of tumors. Increased treatment-related infections observed in nonclinical toxicology studies, especially infections caused by weakly pathogenic organisms, could be an important indicator

**TABLE 6 Parameters that Should Be Evaluated in Immunorelevant Organs with Regard to Immunotoxic Changes**

| Immunorelevant Organ | Parameters |
|---|---|
| Thymus | • Increased/decreased grade of corticomedullary ratio (related only to area)<br>• Increase of starry sky cells<br>• Changes of cellular density in the cortex<br>• Changes of cellular density in the medulla |
| Spleen | • Altered cellular composition of follicles<br>• Changes of the cellularity of PALS, lymphoid follicles, marginal zone, red pulp<br>• Altered number of germinal centers<br>• Extramedullary hematopoiesis<br>• Congestion |
| Lymph nodes | • Changes in the cellularity of follicles, interfollicular area, paracortical area, medulla<br>• Altered number of germinal centers<br>• Changes in the cellularity of sinus<br>• Altered cellular composition of paracortex<br>• Altered number of germinal centers<br>• Hyperplasia of high endothelial venules |
| Peyer's patches | • Changes of the cellularity of follicles (including mantle zone and germinal centers)<br>• Changes of the cellularity of interfollicular area |
| Bone marrow | • Changes of the cellularity<br>• Changes of the myeloid:erythroid ratio |

of unintended immunosuppression. It must be noted that no clear-cut correlation exists between the immunosuppressed state and the incidence of common tumors. The relationship between immunosuppression and cancer is controversial, and with the exception of rare lymphoproliferative, reticuloendothelial, or virus-induced tumors, convincing evidence for increased cancer rates due to immunosuppression have not been found. Under most circumstances, when increased incidences of common tumors are observed in standard long-term chronic/carcinogenicity rodent bioassays (or in other nonclinical toxicology studies), the effects are likely to be related to genotoxicity, hormonal effects, or other relatively well understood mechanisms that will require careful analysis and possibly further investigation. However, for some investigational drugs, the cause of tumor findings in nonclinical studies might not be apparent. In those situations, the potential role of immunosuppression should be considered.

It must be kept in mind that, with standard toxicity studies, doses near or at the maximum tolerated dose can result in changes to the immune system related to stress (e.g., by exaggerated pharmacodynamic action). Commonly observed stress-related immune changes include increases in circulating neutrophils, decreases in circulating lymphocytes, decreases in thymus weight, decreases in thymic cortical cellularity and associated histopathologic changes, and changes in spleen and lymph node cellularity. Thymic weight decreases in the presence of clinical signs, such as decreased body weight and physical activity, are too often attributed to stress. These findings on their own should not be considered sufficient evidence of stress-related

**TABLE 7 Additional Tests that Might Be Technically Integrated into a Standard Repeat Dose Toxicity Study**

| Parameter | Additional Component |
|---|---|
| Hematology | • Analysis of lymphocyte subsets by flow cytometry |
| Clinical chemistry | • Measurement of total hemolytic complement activity<br>• Quantitation of immunoglobulin isotypes<br>• Detection of antinuclear (ANA) and anticytoplasmic (ACA) autoantibodies |
| Histology | • Immunohistological or flow cytometric differentiation of lymphocyte subset in lymphatic organs |
| Functional *ex vivo* tests | • Determination of NK activity in peripheral blood or spleen cells<br>• Determination of mitogen responsiveness of peripheral blood or spleen cells<br>• Determination of phagocytic activity of peritoneal macrophages<br>• Determination of oxidative burst of phagocytes |

immunotoxicity. The evidence of stress should be compelling (e.g., by determination of stress-related blood hormone levels [28]), in order to justify not conducting additional immunotoxicity studies.

The minimum parameters that have to be assessed during repeat dose toxicity studies are well defined by regulatory testing guidelines. Nevertheless, the catalog of assays that can be additionally integrated into such studies can be expanded without the need to modify the overall study design. Some additional tests that might give additional valuable information with regard to immunotoxic effects are summarized in Table 7. When immunotoxicity assessment is regarded as especially important during preclinical development of certain drug substances, using such additional tests may substantially improve the available database for the weight of evidence review of concern(s) for additional immunotoxicity testing. By the way, such an approach might also help to justify an omission of additional immunotoxicity studies.

Most of the possible additional components that can be integrated into a standard repeat dose toxicity study (Table 7) are recommended by regulatory immunotoxicity guidelines as "Additional" or "Follow-up" immunotoxicity assays. Nevertheless, these parameters can already be assessed in a standard repeat dose toxicity study, provided that sufficient amounts of blood can be obtained, and lymphatic organ samples can be obtained under aseptic conditions. Sufficient amounts of blood for the functional assessment of PBMCs can generally not be obtained from mice. For larger animals (including rats, following cardiac puncture), however, obtaining sufficient amounts of blood is not generally a problem.

### 5.2 Additional Testing

If the weight of evidence review of concern(s) indicates that additional immunotoxicity studies are needed, there are a number of animal models that can be used. If there are changes in standard toxicity testing data suggesting immunosuppression, the type of additional immunotoxicity testing that is appropriate will depend on the

nature of the immunological changes observed and concerns raised by the class of compound.

To assess drug-induced immunotoxicity, the generally accepted study design is a 28 day study with consecutive daily dosing. Mice or rats are the species of choice (usually the rat since this is routinely used as the rodent toxicity species), but nonrodents should be used if the drug substance to be investigated shows no primary pharmacodynamic activity in rodents. Furthermore, the species, strain, dose, duration, and route of administration used in additional immunotoxicity studies should be consistent, where possible, with the standard toxicity study in which an adverse immune effect was observed. In addition, nonrodents should be used, if possible, to confirm immunotoxicity findings in rodents in a second species. Usually both sexes should be investigated except for studies in nonhuman primates when required. A rationale should be given when only one sex is used in testing. As a central requirement for immunotoxicity studies, the high dose should be above the no observed adverse effect level (NOAEL) but below a level inducing changes secondary to stress. Multiple dose levels are recommended in order to determine dose–response relationships and the dose at which no immunotoxicity is observed.

In concordance with other regional guidelines, the ICH S8 guideline does not stipulate certain tests for additional immunotoxicity testing, but merely gives recommendations. However, due to the dynamic nature of the immune system, immune effects should ideally be demonstrated using a functional immune assay. Where a specific target is not identified, an immune function study such as the T-cell-dependent antibody response (TDAR) is recommended. If specific cell types are affected in STSs or have been identified by immunophenotyping of leukocyte populations, more specific assays that measure function of that specific cell type might be preferable.

In general, the selected immunotoxicity test should be widely accepted and have been demonstrated to be adequately sensitive and specific for known immunosuppressive agents. The limitations of the applied test encompassing intra- and interassay precision, technician-to-technician precision, limit of quantitation, and linear region of quantitation should be known. In addition, assay sensitivity to known immunosuppressive agents should be established. The test should be embedded in an appropriate quality assurance program; for example, the laboratory should conduct a positive control study periodically in order to demonstrate proficiency of performance. However, when extensive validation might have not been completed or the assay might not be widely used, a scientific rationale for use of the assay should be given. Furthermore, a sufficient number of appropriate positive and negative controls should be incorporated.

***T-Cell-Dependent Antibody Response (TDAR)*** The TDAR should be performed using a recognized T-cell-dependent antigen like sheep red blood cells (SRBCs), bovine serum albumin (BSA), or keyhole limpet hemocyanin (KLH) that result in a robust antibody response. For the SRBC assay, IgM measurement is considered the most appropriate endpoint, whereas IgG measurement is considered to be most appropriate for BSA or KLH. Antibody can be measured by using an ELISA or other immunoassay method. One advantage of ELISA over the traditional plaque forming cell (PFC) assay [29] is that samples can be collected serially during the study, if necessary. Since immunization is likely to have effects on hematology, clini-

cal chemistry, and histology of lymphatic organs, TDAR studies should always be performed as separate studies, or at least in satellite groups of repeat dose toxicity studies.

A protocol for a KLH-based TDAR in mice is depicted in Table 8. Brief details of the assay are as follows.

**Treatment**

1. Use a suitable SPF mouse strain like BALB/c or C57BL/6 × C3H F1 (B6C3F1).
2. Allocate 120 animals in six groups of 10 males and 10 females each according to Table 8. The group sizes may be reduced to 5 male and 5 female animals per group, when a substantial immunosuppressive effect can be expected. The larger group size should be chosen when immunosuppression shall be excluded, since otherwise the statistical power of the assay might be insufficient to prove a lack of immunosuppression.
3. Allow acclimatization for 7 days before sampling of pretest serum (day –7) from the test groups, and allow an additional 7 days of rest before first dosing. Pretest serum of recovery groups may be taken on day 35, which is one week postdosing.
4. Administer the test substance and vehicle daily over a period of 28 days (day 1 to day 28) to all animals using an appropriate route of administration.

**TABLE 8 Design of a Possible Protocol for a KLH-Specific TDAR Study in Mice, Based on a 28 Day Dosing Regimen**

| Dose Level | Group Size | Start of Acclimatization (day) | Dosing of Vehicle and Test Substance (days) | Immunization with KLH (100 μg IP) Without Adjuvant (day) | Serum Sampling (days) |
|---|---|---|---|---|---|
| **Test group** | | | | | |
| Vehicle | 10 male/ 10 female | –14 | 1–28 | 14 | –7; 29 |
| Low dose | 10 male/ 10 female | –14 | 1–28 | 14 | –7; 29 |
| Intermediate dose | 10 male/ 10 female | –14 | 1–28 | 14 | –7; 29 |
| High dose | 10 male/ 10 female | –14 | 1–28 | 14 | –7; 29 |
| **Recovery group** | | | | | |
| Vehicle | 10 male/ 10 female | –14 | 1–28 | 42 | 35; 57 |
| High dose | 10 male/ 10 female | –14 | 1–28 | 42 | 35; 57 |

5. Use a low, intermediate, and high dose level, whereby the high dose level should be above the NOAEL and below a dose level that causes stress, if possible. The intermediate dose (or low dose) level should ideally represent the intended clinical dose level.
6. Immunize all mice of the test groups on day 14, and mice of the recovery groups on day 42 by intraperitoneal injection of 100 μg KLH per mouse without the use of adjuvant.
7. Sample immune serum from all animals of the test groups on day 29, and from all of the recovery groups on day 57.
8. Store serum at –20 °C until ELISA testing.

**ELISA**

1. Coat ELISA plates overnight with 10 μg/mL KLH in a carbonate coating buffer, pH 9.6, and block with 1% BSA in PBS or other appropriate blocking agents like casein or dry milk.
2. Distribute pretest and immune serum samples to the ELISA plates in serial or fixed dilutions. The chosen antiserum dilutions should ideally fall into the linear range of the ELISA. Use KLH-specific monoclonal or polyclonal standard antibodies to calculate antibody concentrations, if necessary. Use at least duplicate measurements (triplicates are preferred).
3. Develop the ELISA using biotinylated or enzyme-linked anti-mouse IgG secondary antibodies and appropriate substrate systems.
4. Record direct optical densities or calculate derivatives like antibody concentrations, titers, stimulation indices, percent inhibition, or whatever is found to be appropriate.

The ELISA may be "tuned-up" by using immunoglobulin isotype-specific secondary antibodies to obtain a more detailed picture of the humoral immune response with regard to possible isotype shifts. Such isotype shifts may be indicative of a possible $T_H1/T_H2$ bias and other cytokine-related modulations of the immune response.

The details of ELISA performance and evaluation are at the discretion of the investigator/sponsor. It must, however, be pointed out that the ELISA, when used in a GLP study, should be thoroughly validated with reference to applicable quality guidelines as, for example, ICH Q2A and Q2B guidelines.

***Immunophenotyping*** Immunophenotyping is the identification and enumeration of leukocyte subsets using antibodies. Immunophenotyping can be done either by flow cytometric analysis or by immunohistology. With flow cytometry, the percentage and absolute counts of a specific cell type or activation markers can be determined with high precision from large numbers of cells analyzed. When immunophenotyping studies are used to characterize or identify alterations in specific leukocyte populations, the choice of the lymphoid organs to be evaluated should be based on changes observed.

Flow cytometric analysis should at least assess the absolute and relative frequency of the major lymphocyte subsets. No specific recommendations can be made

regarding the panel of antibodies to be employed, but at least antibodies specific for T cells (CD3, CD4, CD8), B cells (B220 in mice and CD19 or CD20 or sIg in other species), and NK cells (NK 1.1/CD161 in mice and rats, and CD16 in other species) as well as monocytes/macrophages (Mac-1/CD11b in mice and rats, and CD14 in other species) should be included in the test design. It is at the discretion of the investigator/sponsor to expand the panel by further subset markers for γ, δ-T cells and NK-T cells, by activation markers like CD25, by intracellular cytokine staining, or whatever might make sense to answer special questions with respect to the immunotoxicity assessment. The decision might also be influenced by the capability of the applied flow cytometer with regard to simultaneous multicolor analysis.

Immunophenotyping studies may be designed as separate or satellite studies. As a separate study, the study design should ideally comprise repeated dosing over 28 days; vehicle; low, intermediate, and high dose levels; a sufficient number of animals; and, if possible, both sexes and recovery groups. However, since no immunization or other forms of antigen challenge are required, immunophenotyping can also be easily included in the main groups of a standard repeat dose GLP toxicology study to evaluate changes in lymphocyte subsets. When peripheral blood is used for immunophenotyping, changes can be followed at several time points during the dosing phase and periods without drug exposure (reversal period), if necessary.

Flow cytometry is normally employed to enumerate specific cell populations. It is not a functional assay in this case. However, flow cytometry can be used to measure antigen-specific immune responses of lymphocytes when using fluorescent antigen conjugates for B cells or novel tools like MHC tetramers for T cells. Flow cytometry can also be used for functional endpoints of immune competence like lymphocyte activation, cytokine release, phagocytosis, apoptosis, oxidative burst, and natural killer cell activity [30]. These parameters could be added if the mechanism of action of the drug targets suggests involvement of a particular function of immune cells.

Note that flow cytometry is difficult to validate and hardly meets the criteria of an analytical assay. Thus, to generate an acceptable dataset for immunotoxicity assessment by flow cytometry, a broad set of positive controls (e.g., known marker-positive cells) and negative controls (e.g., isotype-control antibodies) should be included in a test design.

***Natural Killer Cell Activity Assays*** Natural killer (NK) cell activity assays should be conducted if immunophenotyping studies demonstrate a decrease in NK cell number, or if STSs demonstrate increased viral infection rates, or in response to other causes for concern. In general, all NK cell assays are *ex vivo* assays in which tissues (e.g., spleen) or blood are obtained from animals that have been treated with the test compound. Cell preparations are traditionally coincubated with target cells that have been labeled with chromium. Nevertheless, new methods that involve LDH release from target cells, or labeling of target cells with CFSE and propidium iodide followed by flow cytometry, can be used if adequately validated. Different effector to target cell ratios should be evaluated for each assay to obtain a sufficient picture of the level of cytotoxicity.

As for immunophenotyping, assessment of NK cell activity may be designed as a separate or satellite study. As a separate study, the study design should ideally

comprise repeated dosing over 28 days; vehicle; low, intermediate, and high dose levels; a sufficient number of animals; and, if possible, both sexes and recovery groups. However, since no immunization or other forms of antigen challenge are required, assessment of NK cell activity can also be easily included in the main groups of a standard repeat dose GLP toxicology study, provided the technical personnel are able to provide sufficient amounts of blood, and to obtain lymphatic organ samples under aseptic conditions. Sufficient amounts of peripheral blood for the functional assessment of NK cell activity can generally not be obtained from mice. When peripheral blood is used for the assessment of NK cell activity in larger laboratory animals, changes can be followed at several time points during the dosing phase and periods without drug exposure (reversal period), if necessary.

A brief protocol for the assessment of NK cell activity in murine spleen cells during a standard GLP toxicity study could be as follows:

1. Remove spleens under aseptic conditions during necropsy and determine spleen weights (e.g., by using a preweighed tissue culture tube filled with sterile PBS).
2. Remove about one-quarter of the spleen for later histopathology using a sterile scalpel or sterile scissor. Prepare a (sterile) single-cell suspension from the remaining spleen by mechanical disaggregation using surgical pincettes, wire meshes, or whatever method is available.
3. Harvest YAC-1 lymphoma cells (ATCC No. TIB-160) from tissue culture flasks. Label YAC-1 cells in a small volume for 1 h with $^{51}$Cr-sodium chromate (~100 μCi per $10^6$ cells).
4. Cocultivate $10^4$ YAC-1 target cells and spleen cells at effector : target ratios of 100 : 1, 50 : 1, 25 : 1, and 12 : 1 in 200 μL of cell culture medium containing 10% FCS on round bottomed 96 well microculture plates. Setup of triplicate cultures is recommended. Include at least six 200 μL control cultures, each containing target cells alone (spontaneous $^{51}$Cr release) or target cells treated with 1 N NaOH or detergent (maximum $^{51}$Cr release).
5. After 6 h of coculture, draw a defined volume (50 or 100 μL) of the culture supernatants and determine radioactivity as counts per minute (cpm) in the samples using a gamma counter.
6. Calculate specific cytotoxicity according to the following formula:

$$\%\ \text{Specific release} = \frac{(\text{Mean experimental release} - \text{Mean spontaneous release (cpm)})}{(\text{Mean maximum release} - \text{Mean spontaneous release (cpm)})} \times 100$$

The YAC-1 cells must be harvested from exponentially growing cultures to be in sufficiently good condition. Furthermore, YAC-1 cells must be free from mycoplasms. Since a $^{51}$Cr release assay is difficult to validate due to high interassay variability (which is to a high degree related to the actual condition of the YAC-1 cells), all measurements should be conducted in a single measurement series.

***Host Resistance Studies*** Host resistance studies involve challenging groups of mice or rats treated with different doses of a drug substance with defined doses of a pathogen (bacterial, viral, parasitic) or tumor cells. Infectivity of the pathogens or tumor burden observed in vehicle versus test substance treated animals is used to determine if the test compound is able to alter host resistance [31]. Table 9 summarizes the most widely used infectious and tumor models, including endpoints and the defense mechanisms involved. Since the host defense mechanisms in these models are relatively well understood, it is recommended to choose a host resistance assay suitable for a specific target if known.

In conducting host resistance studies, the investigator should carefully consider the direct or indirect (non-immune-mediated) effects of the test compound on the growth and pathogenicity of the organism or tumor cell. For example, compounds that inhibit the proliferation of certain tumor cells can seem to increase host resistance. The same can be expected when using a bacterial challenge model for the immunotoxicity assessment of an antibiotic. In any case, appropriate positive control groups of test animals, dosed with a known immunosuppressive compound, should be included with each host resistance system used.

*Listeria Monocytogenes* Resistance to *L. monocytogenes* involves early clearance by neutrophils, followed by the activation of T cells; the latter is required to activate

**TABLE 9 Widely Used Host Resistance Models, Including Endpoints and Involved Defense Mechanisms**

| Resistance Model | Endpoints | Defense Mechanisms |
|---|---|---|
| Bacterial models | | |
| *Listeria monocytogenes* | Spleen and lung clearance (rats); mortality (B6C3F1 mice) | Phagocytosis by macrophages and T-cell-dependent lymphokine production |
| *Streptococcus* species | Mortality | Complement, neutrophils, antibody |
| Viral models | | |
| Influenza | Mortality | Cell-mediated immunity, interferon, antibody |
| Cytomegalovirus | Clearance from salivary gland | Cytotoxic T-cells, NK cells, antibody |
| Parasite models | | |
| *Plasmodium yoelii* | Parasitemia | Antibody, macrophages, T cells |
| *Trichinella spiralis* | Muscle larvae counts and worm expulsion | T cells (inflammation of the gut), eosinophils, antibodies |
| Tumor models | | |
| B16F10 melanoma | Lung burden | NK cells, cytotoxic T cells |
| PYB6 sarcoma | Tumor incidence | NK cells, cytotoxic T cells |

macrophages that finally eliminate the remainder of the bacteria. Humoral immunity is not relevant in protection against infection in this model. In rats, clearance of *Listeria* after infection via the intravenous or the intratracheal route can be assessed at various times after infection by determining the numbers of colony forming units in the spleen or lungs, respectively, using serial dilutions of homogenates of the organs. Differences in the numbers of bacteria retrieved from the organs are a measure of the clearance of the bacteria by the host. Histopathology after a *Listeria* infection can also be valuable, since pulmonary infection with *Listeria* induces histopathological lesions characterized by foci of inflammatory cells, such as lymphoid and histiocytic cells, accompanied by local cell degeneration and influx of granulocytes.

When C3H or B6C3F1 mice are challenged with *Listeria*, mortality is the usual endpoint monitored. However, clearance and organ bacterial colony counts can also be determined. In mice the resistance to the organism is genetically regulated, and some strains like C57BL/6 are not useful for assays due to resistance. In studies in susceptible mice, a nonlethal challenge level with *L. monocytogenes* of approximately $10^4$ CFU may be given for the enumeration of bacteria in organs, or a 20% lethal challenge level of approximately $5 \times 10^4$ CFU for assessment of mortality. Mortality is recorded daily for 14 days. Treatment groups consisting of 12 mice per group have been found to be useful for obtaining statistically meaningful data on host resistance. This assay is extremely reproducible when the organism is administered intravenously. Since *Listeria* is a human pathogen, appropriate precautions are needed in conducting the assay.

*Streptococcus Pneumoniae* The first line of defense against *S. pneumoniae* is the complement system. Activation of the complement system can result in direct lysis of certain strains of *S. pneumoniae*. However, due to the nature of their cell wall, some strains may be resistant to lysis by complement. Deposition of complement factor C3b on the surface of the bacteria triggers phagocytosis predominantly by neutrophils. In the later stages of the infection, antigen-specific antibody plays a major role in controlling the infection. Thus, compounds that affect complement, polymorphonuclear leukocytes, B cell maturation and proliferation, or the production of antibody can be evaluated in this system. *Streptococcus pneumoniae* is therefore an excellent model for evaluating immunotoxicity, since it elicits multiple immune components that participate in host resistance, each of which can be a potential target for an adverse effect of an immunotoxic drug.

Routinely, B6C3F1 mice are used. Since the $LD_{50}$ of group B streptococci is approximately 5 living bacteria in this strain, mice are vaccinated twice with $2 \times 10^5$ heat inactivated bacteria during the dosing period, the last vaccination given 7 days before the challenge with a lethal dose of living bacteria. Owing to the rapid onset of infection, mortality is recorded twice daily for 7 days during the second half of a 28 day dosing regimen. A nonimmunized control group should be included in each experiment to demonstrate that the challenge dose was lethal to most mice that were not immunized.

*B16F10 Melanoma Cells* The B16F10 melanoma line forms tumor metastases almost exclusively in the lungs of C57Bl/6 mice following IV administration, and it is widely used to assess host resistance to nonimmunogenic or minimally immuno-

genic tumors in mice. NK cells and possibly Tc are involved in resistance to this tumor. Mice are challenged by IV administration of $1 \times 10^5$ B16F10 tumor cells on day 12 of a 28 day dosing regime, and lungs are removed for enumeration of experimental metastases using a dissecting microscope 1 day after the last dose.

*Mouse and Rat Cytomegalovirus* Cytomegaloviruses are classified as members of the family Herpesviridae. The roles of several arms of the immune system in resistance to cytomegalovirus have been identified, including antibodies, cytotoxic T cells, and NK cells. Antibodies in this case contribute to neutralization and antibody-dependent cell-mediated cytotoxicity. NK cell activity appears to be the most effective during the initial stages of infection [32].

Experimentally, rodents can be inoculated intraperitoneally with a species-specific cytomegalovirus, and the concentration of the virus in tissue can be determined in a plaque-forming assay using susceptible adherent cell lines. Homogenates of different organs (salivary gland, lung, kidney, liver, spleen) may be obtained at various times after infection for this purpose. Plaques are counted under a stereoscopic microscope. In rats, cytomegalovirus is detectable 8 days after infection, although the virus load is much higher on days 15–20.

***Macrophage Function*** *In vitro* macrophage function assays can be used for several species. These assays assess macrophage function of cells either derived from untreated animals and exposed to the test compound *in vitro* or obtained from animals treated with the test compound as an *ex vivo* assay [33].

Macrophage function assays may be designed as separate or satellite studies. As a separate study, the study design should ideally comprise repeated dosing over 28 days; vehicle; low, intermediate, and high dose levels; a sufficient number of animals; and, if possible, both sexes and recovery groups. However, since no immunization or other forms of antigen challenge are required, macrophage function assays can also be easily included in the main groups of a standard repeat dose GLP toxicology study to evaluate changes in lymphocyte subsets.

An *in vivo* assay could also be used to assess the effects on the reticuloendothelial cells to phagocytize chromium-labeled sheep red blood cells. In this case animals should be treated with the test compound and injected with chromium-labeled sheep red blood cells. Animals should be necropsied and tissues (e.g., liver, spleen, lung, kidney) removed, followed by a quantitation of specific radioactivity incorporated.

A macrophage phagocytosis assay in mice, measuring the ability of peritoneal macrophages to phagocytize latex beads, and assessment by flow cytometry, might by conducted as follows:

1. Collect peritoneal macrophages from mice immediately after euthanasia and before the peritoneal cavity is opened for gross necropsy.
2. Carefully cut the abdominal skin with scissors and pull it back from the abdomen without hurting the abdominal wall.
3. Inject approximately 5 mL of ice cold $Ca^{2+}$ and $Mg^{2+}$ free PBS containing 10 IU/mL sodium heparin intraperitoneally using a thin needle.

4. Gently massage the peritoneum and aspire the peritoneal wash with a separate syringe and a larger needle. Repeat the procedure twice, if possible.
5. Pellet the peritoneal exudate cells by centrifugation and determine cell viability by trypan blue dye exclusion.
6. Adjust cells to $10^6$ cells/mL with complete cell culture medium containing 10% FCS, and add carboxyfluorescein modified latex microspheres at a ratio of 100:1 (beads:cells).
7. Incubate aliquots of the cell suspension at either 37 °C or 4 °C for 120 min, subjected to gentle agitation.
8. After 120 min, centrifuge the cells and resuspend in 100 μL PBS containing 0.05% formalin, followed by flow cytometric analysis.
9. Identify and gate macrophages according to forward and side scatter properties, and record data from the fluorescein channel.
10. Analyze recorded histograms with the first fluorescent peak corresponding to a single ingested fluorescent bead, and each successive fluorescent peak corresponding to the phagocytosis of an additional bead.
11. Calculate the frequency of active macrophages with one or more ingested microspheres, and the mean number of microspheres ingested by active macrophages.
12. For each data point subtract the respective measurements at 4 °C to correct for simple adherence of the beads.

This assay may be modified by using antibody-labeled fluorescent erythrocytes instead of latex microbeads or by simple (but time-consuming) microscopic enumeration.

***Assays to Measure T-Cell-Mediated Immunity*** Three types of assays have been widely used to assess T-cell-mediated immunity. Mostly, delayed type hypersensitivity (DTH) reactions are used, which are based on an immunization with SRBCs, KLH, or BSA, followed by a subsequent challenge with the antigen into the pinna of the ear or the hind footpad, and determination of the degree of inflammation-induced swelling [34]. The above assay may be modified by assessing the *ex vivo* recall response using lymphatic cells derived from spleen or draining lymph nodes. In this case the *in vitro* proliferation of primed T cells in response to the nominal antigen (using [$^3$H]thymidin or BrdU incorporation) or cytokine production is measured. The *ex vivo* modification of the assay is technically more challenging but usually results in less variability of the results.

Instead of vaccination with SRBCs or protein antigens, topical sensitization and challenge with a known contact sensitizer like oxazolone or picrylic acid may be used. The induced contact allergic response is also a T-cell-mediated immune response that may be affected by immunosuppressive drugs.

Alternatively, cytotoxic T cell response can be generated in mice/rats by vaccination with inactivated tumor cells or viruses, followed by an *ex vivo* assessment of the resulting cytotoxic T cell response using a specific $^{51}$Cr release assay.

Since immunization is likely to have effects on hematology, clinical chemistry, and histology of lymphatic organs, assays to measure T-cell-mediated immunity

should always be performed as separate studies, or at least in satellite groups of repeat dose toxicity studies.

## REFERENCES


1. Penn I. Tumors of the immunocompromised patient. *Annu Rev Med* 1988;39:63–73.
2. Penn I. The role of immunosuppression in lymphoma formation. *Springer Semin Immunopathol* 1998;20:343–355.
3. Rygaard J, Povlsen CO. The nude mouse vs. the hypothesis of immunological surveillance. *Transplant Rev* 1976;28:43–61.
4. Ader R, Felten D, Cohen N. Interactions between the brain and the immune system. *Annu Rev Pharmacol Toxicol* 1990;30:561–602.
5. Khansari DN, Murgo AJ, Faith RE. Effects of stress on the immune system. *Immunol Today* 1990;11:170–175.
6. Renton KW, Gray JD, Hall RI. Decreased elimination of theophylline after influenza vaccination. *Can Med Assoc J* 1980;123:288–290.
7. Mannering GJ, Renton KW, el Azhary R, Deloria LB. Effects of interferon-inducing agents on hepatic cytochrome P-450 drug metabolizing systems. *Ann NY Acad Sci* 1980;350:314–331.
8. Farine JC. Animal models in autoimmune disease in immunotoxicity assessment. *Toxicology* 1997;119:29–35.
9. Mosmann TR, Coffman RL. Heterogeneity of cytokine secretion patterns and functions of helper T cells. *Adv Immunol* 1989;46:111–147.
10. Choquet-Kastylevsky G, Descotes J. Value of animal models for predicting hypersensitivity reactions to medicinal products. *Toxicology* 1998;129:27–35.
11. Gad SC, Dunn BJ, Dobbs DW, Reilly C, et al. Development and validation of an alternative dermal sensitization test: the mouse ear swelling test (MEST). *Toxicol Appl Pharmacol* 1986;84:93–114.
12. Kimber I, Hilton J, Botham PA, Basketter DA, et al. The murine local lymph node assay: results of an inter-laboratory trial. *Toxicol Lett* 1991;55:203–213.
13. Baxter AB, Lazarus SC, Brasch RC. *In vitro* histamine release induced by magnetic resonance imaging and iodinated contrast media. *Invest Radiol* 1993;28:308–312.
14. Szebeni J. Complement activation-related pseudoallergy caused by liposomes, micellar carriers of intravenous drugs, and radiocontrast agents. *Crit Rev Ther Drug Carrier Syst* 2001;18:567–606.
15. Bagenstose LM, Salgame P, Monestier M. Cytokine regulation of a rodent model of mercuric chloride-induced autoimmunity. *Environ Health Perspect* 1999;107(Suppl 5):807–810.
16. Ravel G, Christ M, Horand F, Descotes J. Autoimmunity, environmental exposure and vaccination: Is there a link? *Toxicology* 2004;196:211–216.
17. Pieters R, Albers R. Screening tests for autoimmune-related immunotoxicity. *Environ Health Perspect* 1999;107(Suppl 5):673–677.
18. Descotes J. Autoimmunity and toxicity testing. *Toxicol Lett* 2000;112–113:461–465.
19. Verdier F, Patriarca C, Descotes J. Autoantibodies in conventional toxicity testing. *Toxicology* 1997;119:51–58.
20. Committee for Proprietary Medicinal Products (CPMP). *Note for Guidance on Repeated Dose Toxicity* (CPMP/SWP/1042/99 corr.; London, 27 July 2000).

21. US Food and Drug Administration, Center for Drug Evaluation and Research. *Guidance for Industry: Immunotoxicology Evaluation of Investigational New Drugs*. October 2002.
22. *ICH S6 Harmonised Tripartite Guideline. Preclinical Safety Evaluation of Biotechnology-Derived Pharmaceuticals* (Step 4, 1997).
23. *WHO Guidelines on Nonclinical Evaluation of Vaccines. Annex 1*. Adopted by the 54th Meeting of the WHO Expert Committee on Biological Standardization, 17–21 November 2003.
24. *ICH S8 Harmonised Tripartite Guideline. Immunotoxicity Studies for Human Pharmaceuticals*. (Recommended for Adoption at Step 4 of ICH Process on 15 September 2005).
25. Holsapple MP, Burns-Naas LA, Hastings KL, Ladics GS, et al. A proposed testing framework for developmental immunotoxicology (DIT). *Toxicol Sci* 2005;83:18–24.
26. Wierda D, Smith HW, Zwickl CM. Immunogenicity of biopharmaceuticals in laboratory animals. *Toxicology* 2001;158:71–74.
27. Schuurman HJ, Kuper CF, Vos JG. Histopathology of the immune system as a tool to assess immunotoxicity. *Toxicology* 1994;86:187–212.
28. Pruett SB, Ensley DK, Crittenden PL. The role of chemical-induced stress responses in immunosuppression: a review of quantitative associations and cause–effect relationships between chemical-induced stress responses and immunosuppression. *J Toxicol Environ Health* 1993;39:163–192.
29. Jerne NK, Henry C, Nordin AA, Fuji H, et al. Plaque forming cells: methodology and theory. *Transplant Rev* 1974;18:130–191.
30. Burchiel SW, Kerkvliet NL, Gerberick GF, Lawrence DA, et al. Assessment of immunotoxicity by multiparameter flow cytometry. *Fundam Appl Toxicol* 1997;38:38–54.
31. Dean JH, Luster MI, Boorman GA, Leubke RW, et al. Application of tumor, bacterial and parasite susceptibility assays to study immune alterations induced by environmental chemicals. *Environ Health Perspect* 1982;43:81–88.
32. Selgrade MK, Daniels MJ, Dean JH. Correlation between chemical suppression of natural killer cell activity in mice and susceptibility to cytomegalovirus: rationale for applying murine cytomegalovirus as a host resistance model and for interpreting immunotoxicity testing in terms of risk of disease. *J Toxicol Environ Health* 1992;37:123–137.
33. Hubbard AK. Effects of xenobiotics on macrophage function: evaluation *in vitro*. *Methods* 1999;19:8–16.
34. Holsapple MP, Page DG, Bick PH, Shopp GM. Characterization of the delayed hypersensitivity response to a protein antigen in the mouse—I. Kinetics of reactivity and sensitivity to classical immunosuppressants. *Int J Immunopharmacol* 1984;6:399–405.
35. Council of the European Communities. *Official Journal of the European Communities*, No. L332/11, 26 October 1983. Brussels.
36. Vos JG, van Loveren H. Immunotoxicity testing in the rat. In Burger EJ, Tardiff RE, Bellanti JA, Eds. *Environmental Chemical Exposures and Immune System Integrity*. Princeton, NJ: Princeton Scientific Publishing; 1987, pp 167–180.
37. Sjoblad RD. Potential future requirements for immunotoxicology testing of pesticides. *Toxicol Ind Health* 1988;4:391–395.
38. Luster MI, Munson AE, Thomas PT, Holsapple MP, et al. Development of a testing battery to assess chemical-induced immunotoxicity: National Toxicology Program's guidelines for immunotoxicity evaluation in mice. *Fundam Appl Toxicol* 1988;10:2–19.
39. Luster MI, Portier C, Pait DG, White KL Jr, et al. Risk assessment in immunotoxicology. I. Sensitivity and predictability of immune tests. *Fundam Appl Toxicol* 1992;18:200–210.

40. Luster MI, Portier C, Pait DG, Rosenthal GJ, et al. Risk assessment in immunotoxicology. II. Relationships between immune and host resistance tests. *Fundam Appl Toxicol* 1993;21:71–82.
41. Toxic Substances Control Act. Test Guidelines. Final Rule. *Fed Reg* 1997;62: 43819–43864.
42. Committee for Proprietary Medicinal Products (CPMP). *Note for Guidance on Non-Clinical Local Tolerance Testing of Medicinal Products* (CPMP/SWP/2145/00; London, 1 March 2001).

# REPRODUCTIVE AND DEVELOPMENTAL TOXICOLOGY

Ronald D. Hood[1,2] and Robert M. Parker[3]
[1]*Ronald D. Hood & Associates, Toxicology Consultants, Tuscaloosa, Alabama*
[2]*The University of Alabama, Tuscaloosa, Alabama*
[3]*Hoffmann-LaRoche Inc., Nutley, New Jersey*

**Contents**

1 Regulatory Agency Guidance
2 Flexibility in Developing a Testing Strategy
3 Fertility and General Reproductive Performance Study (Segment I)
  3.1 Rationale for the Test
  3.2 Male-Only, Female-Only, and Combined Segment I Studies
  3.3 Animal Species and Strain Selection
  3.4 Procurement and Acclimation of the Test Animals
  3.5 Selection of Dosage Groups, Controls, and Route of Administration
  3.6 Treatment Timing and Test Duration
  3.7 Mating Procedures, Detection of Mating, and Housing of Mated Females
  3.8 In-Life Observations: Parental Body Weight, Feed Consumption, Estrous Cycles, and Clinical Signs
  3.9 Female Reproductive Evaluation
  3.10 Male Reproductive Evaluation
  3.11 Analysis of Fertility Data
  3.12 Interpretation of the Data
4 Prenatal Development Toxicity Study (Segment II)
  4.1 Rationale for the Test
  4.2 Dose Range-Finding Study
  4.3 The Definitive Study
  4.4 Interpretation of the Data
  4.5 Reporting the Results of Developmental Toxicity Studies
5 Perinatal and Pastnatal Study (Segment III)
  5.1 Rationale for the Test
  5.2 Animal Species and Strain Selection

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

5.3 Procurement and Acclimation of the Test Animals
5.4 Housing of Female Rats
5.5 Selection of Dosage Groups, Controls, and Route of Administration
5.6 Treatment Timing and Test Duration
5.7 $F_0$ Generation
5.8 $F_1$ Generation
5.9 Analysis of Maternal and Litter/Pup Data
6 Two-Generation Reproduction Toxicity Studies
6.1 Rationale for the Test
6.2 Principle of the Test Method
6.3 Animal Species and Strain Selection and Rationale
6.4 Procurement and Acclimation of the Test Animals
6.5 Number of Test Animals
6.6 Selection of Dosage Groups, Controls, and Route of Administration
6.7 Selection of Dosage Levels
6.8 Treatment Timing and Test Duration
6.9 Mating Procedures, Detection of Mating, and Housing of Mated Females
6.10 $F_0$ Generation
6.11 $F_1$ Generation
6.12 $F_2$ Generation
6.13 Analysis of Maternal and Litter/Pup Data
7 Juvenile Toxicity Testing
7.1 Rationale for the Test
7.2 Study Justification
7.3 Species of Choice
7.4 Age at Start and End of Treatment
7.5 Route of Administration
7.6 Frequency of Administration and Duration of Treatment
7.7 Range-Finding Studies
7.8 Juvenile Animal Study Design
7.9 Timing of the Juvenile Animal Study
8 Conclusion
References

## 1 REGULATORY AGENCY GUIDANCE

This chapter addresses testing protocols for two general categories of toxicity: (1) reproductive toxicity, which includes effects on fertility, parturition, and lactation; and (2) developmental toxicity, which addresses effects on the developing offspring, namely, mortality, dysmorphogenesis (structural alteration), altered growth, and functional abnormality. It is fortunate that in recent years the United States, the European Community, and Japan have worked together to harmonize their test guidelines via the International Conferences on Harmonization (ICH), thus reducing the regulatory burden inherent in their previously widely differing requirements. According to the ICH S5A document, *Guideline for Industry, Detection of Toxicity to Reproduction for Medicinal Products* [1], a testing regimen should be selected that would "allow exposure of mature adults and all stages of development from conception to sexual maturity. To allow detection of immediate and latent effects

of exposure, observations should be continued through one complete life cycle, i.e., from conception in one generation through conception in the following generation." The document further suggests that an integrated test sequence can be subdivided into the following stages:

- *Stage A*. Premating to conception (adult male and female reproductive functions, development and maturation of gametes, mating behavior, fertilization).
- *Stage B*. Conception to implantation (adult female reproductive functions, preimplantation development, implantation).
- *Stage C*. Implantation to closure of the hard palate (adult female reproductive functions, embryonic development, major organ formation).
- *Stage D*. Closure of the hard palate to the end of pregnancy (adult female reproductive functions, fetal development and growth, organ development and growth).
- *Stage E*. Birth to weaning (adult female reproductive functions, neonate adaptation to extrauterine life, preweaning development and growth).
- *Stage F*. Weaning to sexual maturity (postweaning development and growth, adaptation to independent life, attainment of full sexual function).

Nevertheless, in most cases those six stages are assessed by means of just three studies: (1) the fertility and general reproductive performance study (Segment I, Stages A–B); (2) the prenatal developmental toxicity study (Segment II, Stages C–D); and (3) the perinatal and postnatal study (Segment III, Stages C–F).

In addition, in recent years increased emphasis has been placed on direct toxicity studies in juvenile animals [2, 3]. All of these study types are discussed in some detail later in this chapter. And in some cases, some of the above studies can be omitted. For example, in establishing the safety of veterinary drug residues in human food, a multigeneration study in at least one rodent species is recommended because human exposure may be long term, and the segment studies listed above are not viewed as essential [4].

## 2 FLEXIBILITY IN DEVELOPING A TESTING STRATEGY

The "bad news" about regulatory agency guidelines (i.e., "guidances" according to the FDA) for nonclinical studies is that they are not intended to be specific blueprints that can be followed exactly in every case. That is also the good news regarding such strategies for reproductive and developmental toxicity testing, as they allow for flexibility and use of scientific judgment in those cases in which a "one size fits all" approach would not yield the most useful data. That has led to use of the term "guidance," as opposed to "guideline," by the U.S. FDA, thus implying a less rigid stance. Implementation of flexible test guidelines can be somewhat problematic, however, in that many of the individuals currently designing, implementing, or making regulatory decisions regarding developmental and reproductive toxicology study protocols have had little or no formal education or background knowledge in the disciplines that are especially critical to the field. Of course, a reasonable degree

of regulatory guidance flexibility is essential, because no immutable guideline can be designed that would cover all possible eventualities for these complex studies, and guidelines are not typically revised rapidly enough to keep abreast of advances in the relevant testing methodology. It then becomes essential to utilize the training and advice available in relevant references that provide more detail than the current chapter (e.g., Refs. 5 and 6).

Nonclinical testing strategies may require modification to be more predictive of the hazard potential of the particular product under development. Thus, it is important to discuss and consider possible "variations in test strategy according to the state-of-the-art and ethical standards in human and animal experimentation" [7] prior to finalizing study protocols for drug candidates and other products with unique characteristics. For example, the test protocol for new vaccines should be specifically tailored to the particular vaccine under consideration. With regard to reproduction and development, vaccine testing is typically restricted to pre- and postnatal developmental toxicity studies, as no fertility or postweaning studies are conducted for most vaccine products. Such vaccine testing is commonly conducted in only one species, preferably one in which an immune response is elicited, and test agent administration should generally be episodic, rather than by the daily dosing regimen most commonly used for testing other products [8]. As another example of deviation from the usual protocol, for agents typically administered only once, such as diagnostics and medicines used during surgery, dosing repeatedly may be impractical. In such cases, use of high doses for brief treatment periods may be necessary [7]. Also, if an adverse effect is detected, in some cases it can be useful to conduct further studies to more fully characterize the nature of the effect and perhaps provide an explanation for its basis. This can prove valuable if, for example, it is determined that the mechanism for the toxic effect does not apply to humans or would not occur at the likely human equivalent dose.

## 3 FERTILITY AND GENERAL REPRODUCTIVE PERFORMANCE STUDY (SEGMENT I)

### 3.1 Rationale for the Test

Fertility studies (i.e., Segment I studies) test for toxic effects resulting from treatment of both sexes with the test article before mating and during cohabitation and mating, and continuing to treat the females until implantation on gestation day 6 (GD6). This comprises evaluation of ICH Stages A and B of the reproductive process. For females, this should detect effects on the estrous cycle (ovarian and uterine effects), tubal transport, libido, development of preimplantation stages of the embryo, and implantation. For males, it will permit detection of functional effects (e.g., on libido, mating behavior, capacitation, epididymal sperm maturation) that may not be detected by histological examinations of the male reproductive organs.

### 3.2 Male-Only, Female-Only, and Combined Segment I Studies

The fertility study can be performed as male-only [9] (for paternally mediated effects) and female-only [1] (for maternally mediated effects) studies or as a com-

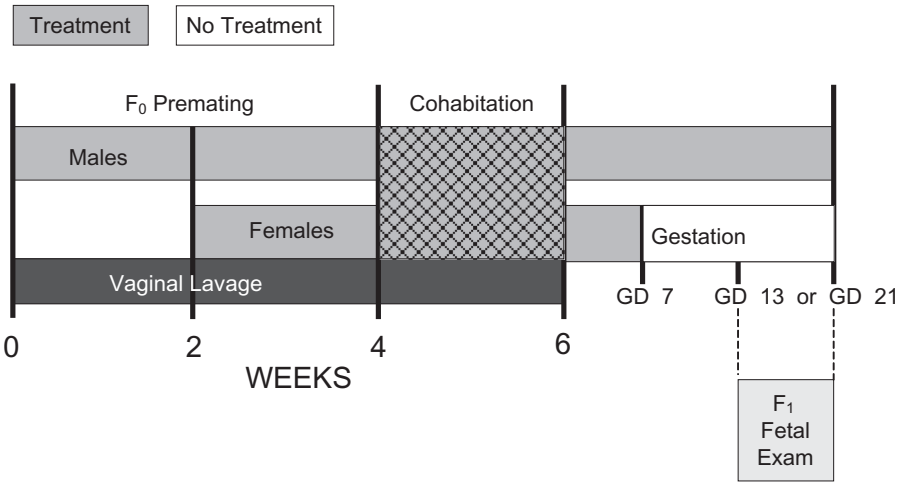


**FIGURE 1** Fertility study schematic (FDA Segment I and ICH S5 and S5A).

bined male/female study (for couple-mediated effects) (Fig. 1). Most regulatory reproduction and fertility guidelines specify cohabitation of treated male rats with treated females, complicating the resolution of gender-specific influences. If male or female reproductive organ toxicity has been observed in subchronic studies, then performance of the male-only and female-only study would allow for greater resolution of gender-based toxicities.

Males are treated with the test article for 2, 4, or 10 weeks (depending on guideline requirements, see later discussion) prior to cohabitation and mating, and continuing until termination. The males are routinely treated until after the female necropsy has been completed; then the males are euthanized and necropsied. Females are treated 2 weeks prior to cohabitation, through cohabitation and mating, and until implantation (usually GD 6). The dam is euthanized on either GD 13 or 21 and necropsied, and the conceptuses are evaluated for viability.

***Male-Specific Endpoints and Paternally Mediated Effects*** Data on the potential male reproductive toxicity of a test article may be collected from subchronic, chronic, and fertility studies. Male-specific endpoints that are determined in these studies are listed in Table 1. Several of these parameters are not directly observed (e.g., mating behavior, secretion of hormones), but perturbations to reproductive organs may suggest alterations in these parameters.

Male-only treatment with a variety of agents has been shown to produce adverse effects in offspring, including pre- and postimplantation loss, growth and behavioral deficits, and malformations [10–14]. Many of the chemicals reported to cause paternally mediated effects display genotoxic activity and may cause transmissible genetic alterations. However, other mechanisms of induction of male-mediated effects are also possible, including nongenetic (e.g., presence of drug in seminal fluid) or epigenetic (e.g., an effect on imprinted gene expression from the paternal alleles, resulting in alteration of gene function). Additional studies will be needed to clarify the mechanisms of paternal exposure associated with adverse

**TABLE 1 Male-Specific Endpoints in Fertility Studies**

| |
|---|
| Body weight |
| Mating behavior (not observed) |
| Libido |
| Mounting |
| Erection |
| Ejaculation |
| Reproductive organ weights |
| Histopathology |
| Testes |
| Epididymides |
| Seminal vesicles |
| Prostate |
| Pituitary |
| Sperm parameters |
| Quality |
| Count |
| Viability |
| Morphology |
| Motility |
| Transport |
| Maturation |
| Storage in the epididymis |
| Production of seminal fluid |
| Production and secretion of hormones in the pituitary–hypothalamus–gonadal axis |

effects on offspring. If a test article is identified as causing a paternally mediated adverse effect on offspring in the test species, the effect should be considered an adverse reproductive effect.

***Female-Specific Endpoints and Maternally Mediated Effects*** The female-specific endpoints in fertility studies are focused on estrous cyclicity, mating, pregnancy, and offspring survival and development. Additional reproductive system endpoints can also be useful in evaluating reproductive toxicity. Adverse alterations in the nonpregnant female reproductive system have been reported at treatment levels below those that result in reduced fertility or produce other effects on pregnancy or pregnancy outcomes. Female-specific endpoints of reproductive toxicity evaluated in the fertility study are listed in Table 2. Female mating behavior and tubal transport are not directly observed, but perturbations to reproductive organs may suggest alterations in these parameters.

Unlike the male, the status of the adult female reproductive system fluctuates. In nonpregnant female rats, the ovarian and uterine structures (and other reproductive organs) change throughout the estrous cycle. Although not cyclic, other changes accompany the progression of pregnancy, lactation, and return to cyclicity after lactation. It is important to be aware of the reproductive status of the female at necropsy, including estrous cycle stage. These normal fluctuations may affect or confound the evaluation of female reproductive endpoints. This facilitates interpretation of

**TABLE 2 Female-Specific Endpoints in Fertility Studies**

| |
|---|
| Body weight |
| Estrous cycle |
| Onset |
| Length |
| Patterns |
| Mating behaviors |
| Reproductive organ weights |
| Histopathology |
| Ovary |
| Uterus |
| Vagina |
| Ovulation |
| Number of corpora lutea |
| Tubal transport |
| Implantation |
| Pre- and postimplantation losses |

effects with endpoints such as uterine weight and histopathology of the ovary, uterus, and vagina.

It is well established that female-only treatment can indirectly produce adverse effects in offspring (i.e., maternally mediated effects). If a test article is identified as causing a maternally mediated adverse effect on offspring, the effect should be considered an adverse reproductive effect.

***Couple-Mediated Endpoints*** Couple-mediated endpoints are reproductive endpoints where both parents may have a contributory role. Couple-mediated endpoints from fertility studies include mating behaviors (mating rate, time to mating), pregnancy rate, number of implantation sites, number of conceptuses (total and live), and preimplantation and postimplantation loss. Most regulatory reproduction and fertility guidelines specify cohabitation of treated male rats with treated female rats, complicating the resolution of gender-specific influences.

### 3.3 Animal Species and Strain Selection

***Rationale for Species and Strain Selection*** It is generally desirable to use the same species and strain as in other toxicological studies. Reasons for using rats as the predominant rodent species are practicality, comparability with other results obtained in this species, and the large amount of accumulated background knowledge. The disadvantages of different species are described in Note 5 of the ICH S5A guideline [1].

The animals used must be well defined with respect to their health, fertility, fecundity, prevalence of abnormalities, and embryo–fetal deaths, and their consistency among studies. Within and between studies, animals should be of comparable age, weight, and parity at the start.

When nonhuman primates or other large animal species are used, it is recommended that the protocol design be discussed with the regulatory agencies prior to

the conduct of the study to ensure that there is consensus regarding the sensitivity of assay. Because of the normal low pregnancy rate, nonhuman primate fertility studies do not routinely include mating. The treatment effects on menstrual cycle and reproductive hormones are evaluated. Alterations in normal menstrual cycle or reproductive hormone patterns may be considered as adverse.

***Number, Sex, and Age of Test Animals*** The number of animals per sex per group should be sufficient to allow meaningful interpretation of the data. Routinely, 20 rats per sex per group are evaluated.

### 3.4 Procurement and Acclimation of the Test Animals

The test animals should be young but sexually mature and from the same source and strain, and the females should be nulliparous and as nearly the same age and weight as is practicable. They should be obtained from reputable suppliers that have programs ensuring the genetic integrity of their animals, in addition to their freedom from diseases and parasites. The test animals should be acclimated to the laboratory conditions prior to study start.

### 3.5 Selection of Dosage Groups, Controls, and Route of Administration

Most fertility studies are designed with a vehicle control and at least three dosage groups, all of which should be concurrent and identically housed and handled. Males and females should be assigned by a random procedure to assure that mean body weights are comparable among all test groups. The test compound or vehicle should be administered by the route that most closely approximates the pattern of expected human exposure.

Selection of dosages is one of the most critical issues in design of the reproductive toxicity study. Data from all available studies (pharmacology, acute, subchronic, and chronic toxicity, and kinetic studies) should be used to select the high dose level. A repeated dose toxicity study of about 2–4 weeks' duration provides a close approximation to the duration of treatment in segmental designs of reproductive studies. Using similar doses in the reproductive toxicity studies as in the repeated dose toxicity studies allows interpretation of any potential effects on fertility in context with general systemic toxicity. Some minimal toxicity should be induced in the highest dosage group.

According to the specific compound, factors limiting the high dosage determined from repeated dose toxicity studies could include (1) increased or reduced body weight gain; (2) hematology, clinical chemistry, and histopathology; (3) specific target organ toxicity; (4) exaggerated pharmacological response; (5) physicochemical property limitations in the amount that can be administered; and (6) pharmacokinetic profile.

After deciding on the highest dosage, lower dosages should be selected in a descending sequence, the intervals depending on kinetic and other toxicity factors. The strategy should also result in a no observed adverse effect level (NOAEL) for reproductive aspects.

### 3.6 Treatment Timing and Test Duration

The usual frequency of administration is once daily, but consideration should be given to using either more or less frequent administration, taking kinetics into account.

***Males*** The duration of treatment selected should be adequate for determining effects on spermatogenesis. The duration of spermatogenesis (approximately 63 days in the rat) was the reason that most fertility and reproductive studies require treatment of the males for approximately 10 weeks prior to mating. The premating treatment duration will permit evaluation of functional effects on male fertility that may not be detected by histopathological examination in shorter duration repeated dose toxicity studies.

However, the premating treatment interval may be reduced to 4 weeks in ICH fertility studies if no male reproductive toxicity effects were found in a repeated dose toxicity study (of at least 4 weeks' duration). Validation studies [15, 16] have indicated that a 2 week premating treatment period is adequate to detect male reproductive toxicity. Therefore, treatment of the males for 2 weeks before mating is also acceptable in ICH fertility studies. Justification for the length of the premating treatment period should be stated in the protocol. In fertility studies, treatment should continue throughout the cohabitation period and until termination for males.

***Females*** Provided that no female reproductive toxicity effects were found in a repeated dose toxicity study (of at least 4 weeks' duration), a 2 week premating treatment interval should be used for females in ICH fertility studies. Justification for the length of the premating treatment period should be stated in the protocol. In fertility studies, treatment should continue throughout the cohabitation period and until implantation for females.

### 3.7 Mating Procedures, Detection of Mating, and Housing of Mated Females

***Mating Procedure*** A mating ratio of 1 : 1 is advisable, and procedures should allow identification of both parents of a litter. This is the safest option for obtaining good pregnancy rates and avoiding incorrect interpretation of results. The female is introduced into the male's cage in the afternoon, and the female is checked for evidence of mating each morning until mating is detected.

***Detection of Mating*** In standard reproductive toxicology studies, the evidence of sexual receptivity and successful mating is based on the presence of copulatory plugs or sperm-positive vaginal lavage. When the male rat ejaculates, the semen quickly coagulates, forming a copulatory plug. Copulatory plugs can remain *in situ* (in the vagina), or they can be dislodged from the vagina and fall into the cage pan (cage plugs). While in cohabitation, the females and their cages should be checked daily for the presence of copulatory plugs. If a copulatory plug is observed *in situ* (in the vagina), a vaginal lavage sample is not necessary. The presence of a copulatory plug in the cage pan is not sufficient evidence that mating has occurred. If only a cage pan plug is observed or no evidence for mating is observed, then a vaginal lavage

should be performed and checked for sperm. When sperm are seen, a positive (+) mating is recorded. When no sperm are present in the vaginal smear, a negative (–) mating is recorded.

However, these *post hoc* findings do not demonstrate that male sexual activity resulted in adequate sexual stimulation of the female rat. Failure to provide adequate stimulation to the female has been shown to impair sperm transport in the genital tract of female rats, thereby reducing the probability of successful fertilization [17]. The reduced fertility (as calculated in the Fertility Index) might be erroneously attributed to an effect on the spermatogenic process in the male or on fertility of the female, rather than on altered mating behavior. Effects on sexual behavior should be considered as adverse reproductive effects. However, impairment of sexual behavior that is secondary to more generalized physical debilitation (e.g., intoxication, impaired motor activity, or general lethargy) should not be considered an adverse reproductive effect; such conditions represent adverse systemic effects.

***Housing of Mated Females*** Once the females have mated, they should be housed individually to allow for monitoring of feed and, if necessary, water consumption. Feed and water should be provided *ad libitum*.

### 3.8 In-Life Observations: Parental Body Weight, Feed Consumption, Estrous Cycles, and Clinical Signs

During study, clinical signs and mortalities should be assessed at least once daily. Body weight should be recorded at least twice weekly. Feed and, if necessary, water intake should be measured at least once weekly (except during cohabitation, when feed consumption is not recorded). Vaginal smears should be collected and evaluated daily for at least 2 weeks during the pretreatment period (baseline), daily during the 2 week treatment period prior to mating, and during the 2 week mating period (for estrous cyclicity and confirmation of mating). Other observations that have proved of value in other toxicity studies may be recorded.

### 3.9 Female Reproductive Evaluation

***Assessment of the Dam, Uterine Contents, and Conceptuses at Necropsy*** Any mated females that appear moribund should be euthanized appropriately. Any mated females found dead should be necropsied to investigate the cause of death. In all such cases, any evidence of pregnancy and embryonic development should be evaluated and recorded.

Dams are routinely necropsied on GD 13 (early in the second half of gestation). The dam is weighed and euthanized, and the ovary and uterus are examined. Uterine contents (number of implantation sites) are recorded. The dam is subjected to a complete necropsy, and the reproductive organs and other tissues specified in the protocol are weighed and saved for possible histopathological evaluation.

*Inspection of Maternal Viscera* The gravid uterus should be removed and weighed. The remaining maternal abdominal and thoracic organs should be inspected grossly for evidence of toxicity, and protocol-specified target organs should be removed and

weighed. If gross organ lesions are noted, the affected organs should be preserved for possible histopathologic examination.

The uterus should be examined to determine if the animal was pregnant, and the relative numbers of live and dead conceptuses and their relative positions in the uterus should be recorded. If the uterus shows no signs of implantation, it should be treated with a reagent, such as 10% ammonium sulfide, to reveal any early resorptions (partially autolyzed embryos) that may not have been obvious [18].

*Enumeration of Corpora Lutea* Each corpus luteum appears as a round, slightly pink, discrete swelling on the ovarian surface. The corpora lutea can be counted with the naked eye or under a dissecting microscope to determine the number of eggs ovulated. The number of corpora lutea should be recorded for each ovary. In a litter with no obvious signs of resorptions, the number of corpora lutea should be equal to or exceed the number of total implantation sites. Corpora lutea of rats are relatively obvious; however, some corpora lutea may obscure others, making counting difficult. Corpora lutea can be discounted when the count is less than the number of implants, or the number of corpora lutea can be set to correspond to the number of implants to correct for involuting corpora lutea in partially resorbed litters. The corpora lutea count is used in determining preimplantation losses. A reduction in the number of corpora lutea or an increase in the preimplantation loss should be considered an adverse reproductive effect.

*Uterine Evaluation* In fertility studies where the pregnant female rodents are necropsied around GD 13, the conceptuses are categorized as live (normal appearance, heart beat, etc.) or dead (discolored, undergoing autolysis). An implanted conceptus that is undergoing autolysis is termed a resorption (see Table 3). An early resorption is usually visualized by staining with a 10% aqueous solution of ammonium sulfide by the Salewski method or by observation through pressed glass plates. The number and uterine position of each conceptus is recorded.

In some fertility studies where the necropsy is performed near term, the fetuses are classified as live (fetus exhibits normal appearance, spontaneous movement, heart beat, etc.) or dead (nonliving full size fetus with discernible digits, with a body weight greater than 0.8 g for rats). Resorptions are classified as early or late. An

**TABLE 3 Categorization Scheme for Offspring from Prenatal Developmental Toxicity Studies**

| Category | Distinguishing Characteristics |
|---|---|
| Early resorption | Only dark tissue derived from the placenta remains; if very early, will be visible only with ammonium staining or similar aid to visualize metrial glands and decidual changes in the uterine lining |
| Mid resorption | Embryonic structures visible but typically show signs of autolysis; may have limb buds but no discernible digits; embryonic remains are generally pale in appearance |
| Late resorption | Dead fetus that is partially autolyzed |
| Dead fetus | Dead fetus that shows no signs of autolysis; may be gray in color |
| Live fetus | Exhibits spontaneous movement, or movement can be elicited by touch; typically pink in color because of circulating blood |

early resorption is defined as a conceptus that has implanted but has no recognizable embryonic characteristics evident upon examination. Late resorptions are defined as fetal remains or tissues that have recognizable fetal characteristics (such as limb buds with no discernible digits present) and are undergoing autolysis or maceration.

*Vaginal Lavage Methodology* Vaginal lavage is an easy, noninvasive method of collecting vaginal cells for determination of the estrous cycle or to confirm mating (presence of sperm). Vaginal lavage for the determination of the stage of estrus can be performed at any time of the day; however, it is important that the samples be collected at approximately the same time each day over the course of a study. A pre- and postcheck for contamination of the saline should be made. The female rat is removed from the cage, and her identification is verified. One or two drops (approximately 0.25 mL) of physiological (0.9%) saline are withdrawn from the saline aliquot into a new, clean dropping pipette, and the tip of the pipette is gently inserted into the vaginal canal. The pipette bulb is firmly but gently depressed to expel the saline from the pipette into the vagina. The saline is gently drawn back into the dropping pipette, and the pipette is removed from the vaginal canal. If the flush is very clear, the lavage should be repeated. Care must be taken not to insert the tip too far into the vaginal canal, because stimulation of the cervix can result in pseudopregnancy (prolonged diestrus), and the female is returned to her cage. The contents of the pipette are then delivered onto either a clean glass slide or a ring slide containing designated areas for the placement of each sample. Each slide should be numbered, a top and bottom of each slide should be clearly identified, and a corresponding record should be made that indicates the animal number and location for each vaginal smear. All slides should be handled carefully to avoid mixing vaginal lavage samples from different animals. If samples become mixed, they are discarded, and the sampling procedure is repeated for those animals.

*The Estrous Cycle* The estrous cycle [19, 20] is separated into distinct phases: metestrus (also called diestrus I), diestrus (also called diestrus II), proestrus, and estrus. The metestrus/diestrus (diestrus I and II) duration is usually 2–3 days, proestrus is usually 1 day, and estrus is also about 1 day.

- Metestrus vaginal smears contain leukocytes, little or few polynucleated epithelial cells, and a variable number of cornified epithelial cells (some of these cells are beginning to roll).
- Diestrus vaginal smears contain a mixture of cell types, mainly leukocytes, a variable number of polynucleated epithelial cells, and few to no cornified epithelial cells. The diestrus period is a time of low serum estradiol, as new follicles mature in the ovary. The transition from diestrus to proestrus is caused by increased estradiol secretion that reaches a peak around the midday of proestrus.
- During the proestrus period, vaginal epithelial cells are proliferating and sloughing off into the vaginal lumen. The proestrus vaginal smear contains numerous round polynucleated epithelial cells (in clumps or strands), a variable number of leukocytes, and few to no cornified epithelial cells. Serum estradiol falls rapidly on the evening of proestrus, as the ovarian follicles leutinize.

- Ovulation occurs early in the morning of vaginal estrus. Proliferation of the vaginal epithelium decreases, and cornified epithelial cells predominate. The estrus vaginal smear contains no leukocytes, few or no polynucleated epithelial cells, and many flat cornified cells.

ESTROUS CYCLE STAGING Each vaginal lavage sample (fresh or stained) is examined microscopically (low power, 100×). Cellular characteristics of vaginal smears reflect structural changes of vaginal epithelium. The stage of the estrous cycle is determined by recognition of the predominant cell type in the smear. Metestrus smears contain leukocytes and cornified epithelial cells. Diestrus smears contain predominantly leukocytes. The proestrus smears contain predominantly round, polynucleated epithelial cells (in clumps or strands). The estrus smears contain predominantly flat cornified cells. The stage of the estrous cycle is recorded for each female.

ESTROUS CYCLE EVALUATION Measurement of estrous cycle length is done by selecting a particular stage and counting until the recurrence of the same stage (usually day of first estrus, day of first diestrus, or day of proestrus). The estrous cycle lengths should be calculated for each female. It is often easier to visualize the estrous cycle by plotting the data. The group mean estrous cycle length should be calculated, and the number of females with a cycle length of 6 days or greater should be calculated. A single cycle with a diestrus period of 4 days or longer or an estrus period of 3 days or longer is generally considered aberrant.

Females with two or more days of estrus during most cycles are classified as showing *persistent* estrus. Females with four or more days of diestrus during most cycles are classified as showing *persistent* or prolonged diestrus. Pseudopregnancy is a persistent diestrus (approximately 14 days in duration). Constant estrus and diestrus, if observed, should be reported. Persistent diestrus indicates at least temporary and possibly permanent cessation of follicular development and ovulation, and thus temporary infertility. Persistent diestrus or anestrus may be indicative of interference with follicular development, depletion of the pool of primordial follicles, or perturbation of the gonadotropin support of the ovary.

The presence of regular estrous cycles after treatment does not necessarily indicate that ovulation occurred, because luteal tissue may form in follicles that have not ruptured. However, lack of ovulation should be reflected in reduced fertility parameters. Altered estrous cyclicity or complete cessation of vaginal cycling in response to toxicants should be considered an adverse female reproductive effect. Subtle changes of cyclicity can occur at dosages below those that alter fertility. However, subtle changes in cyclicity without associated changes in reproductive or hormonal endpoints should not be considered adverse.

*Oocyte Quantification* Histopathological examination should detect qualitative depletion of the primordial follicle population. However, reproductive study guidelines suggest that a quantitative evaluation of primordial follicles of the $F_0$ females be performed. Recently, the Society of Toxicologic Pathology [21] has recommended against performing this evaluation. Ovarian histology, oocyte quantification, and differential follicle count methodology have been described in several publications [22–27].

The number of animals, ovarian section selection method, and section sample size should be statistically appropriate for the evaluation procedure used. The guidelines cite ovarian follicular counting methodologies that use a differential count of the primordial follicles, growing follicles, and antral follicles. The current guidelines require evaluation of ovaries from ten randomly selected females from the highest treatment and control groups. Five 6 μm ovarian sections are made from the inner third of each ovary, at least 100 μm apart. The number of primordial follicles (and small growing follicles) is counted, and the presence or absence of growing follicles and corpora lutea is confirmed for comparison of highest dosage group and control ovaries. If statistically significant differences occur, the next lower dosage group is evaluated, and so forth.

### 3.10 Male Reproductive Evaluation

***Assessment of the Males at Necropsy*** In the event of an equivocal mating result, males could be mated with untreated females to ascertain the male's fertility. Males are routinely necropsied after the outcome of mating is known. The males treated as part of the study may also be used for evaluation of toxicity to the male reproductive system if dosing is continued beyond mating and sacrifice is delayed. Any males that are found dead or appear moribund should be necropsied to investigate the cause of death. The male is weighed, euthanized, and subjected to a complete necropsy.

***Inspection of Male Viscera*** The testes and the accessory sex organs are removed and weighed. The remaining abdominal and thoracic organs should be inspected grossly for evidence of toxicity, and target organs should be removed and weighed. If gross organ lesions are noted, the affected organs should be preserved for possible histopathological examination. Although not mandated, evaluation of the histopathology of the testes and epididymides is generally performed, and an evaluation of sperm for motility, count, concentration, and morphology can be performed if that was not done in previous subchronic or chronic studies.

#### *Histopathology*

*Testis* The most sensitive endpoint for the determination of testicular toxicity is histopathology. Testicular histology [28] and sperm parameters are linked, and alterations in testicular structure are usually accompanied by alterations in testicular function. However, some functional changes may occur in the absence of detectable structural changes in the testis. Alterations in the testicular structure and function are also linked to alterations in epididymal structure and function, but changes in epididymal function may occur in the absence of testicular or epididymal structural changes.

Histopathological evaluation of the testis can determine if the germinal epithelium is severely depleted or degenerating, if multinucleated giant cells are present, or if sloughed cells are present in the tubule lumen. The rete testis should be examined for dilation caused by obstruction or disturbances in fluid dynamics and for the presence of proliferative lesions and rete testis tumors. Both spontaneous and chemically induced lesions in the rete testis often appear as germ cell degeneration

and depletion. More subtle lesions, such as retained spermatids or missing germ cell types, can significantly affect the number of sperm being released into the tubule lumen.

Several methodologies for qualitative or quantitative assessment of testicular tissues are available that can assist in the identification of less obvious lesions, including use of the technique of "staging spermatogenesis." A detailed morphological description of spermatogenesis and spermiogenesis, with light and electron microscopic photographs of these processes, is available in a book by Russell et al. [29].

*Epididymides, Accessory Sex Glands, and Pituitary* Compared with the testes, less is known about structural changes associated with exposure of the epididymides and accessory sex glands to toxicants. Histopathological evaluation should include the caput, corpus, and cauda segments of the epididymis. A longitudinal section of the epididymis that includes all epididymal segments should be obtained for histopathological examination. Presence of debris and sloughed cells in the epididymal lumen indicates damage to the germinal epithelium or to the excurrent ducts. The presence of lesions, such as sperm granulomas, leukocytic infiltration (inflammation), or absence of clear cells in the cauda epididymal epithelium, should be noted. The density of sperm and the presence of germ cells in the lumina reflect time-dependent events in the testis. Sperm in the caput were released from the testis only a few days earlier, while the sperm in the cauda were released approximately 14 days earlier. When decreased copulatory plugs are observed during cohabitation, a careful examination of the accessory sex glands is indicated.

***Sperm Evaluation*** The sperm quality parameters are sperm number, sperm motility, and sperm morphology [30–32]. Although effects on sperm production can be reflected in other measures, such as testicular spermatid count or cauda epididymis weight, no other measures are adequate to assess effects on sperm motility and velocity. The Computer Assisted Sperm Analysis (CASA) System [33] calculates sperm count, concentration, motility, progressive velocity, track speed, path velocity, straightness, linearity, cross beat frequency, and amplitude of lateral head displacement. For detailed methodologies for epididymal and testicular sample preparations and manual and CASA analysis of epididymal sperm counts and motility and testicular homogenization-resistant spermatid counts, respectively, refer to Parker [34] or the CASA manuals.

While changes in endpoints that measure effects on spermatogenesis and sperm maturation have been related to fertility in several test species, the prediction of infertility from these data (in the absence of actual fertility data) may not be reliable. Fertility is dependent not only on having adequate numbers of sperm, but also on the normality of the sperm. When rat sperm quality is good, rat sperm number must be substantially reduced to affect fertility. Decrements in sperm number or motility may not cause infertility in rats but can be predictive of infertility.

### 3.11 Analysis of Fertility Data

Maternal and paternal toxicity may be evidenced by indicators such as body weight or weight gain, feed and water consumption, organ weights, gross or microscopic

**TABLE 4 Indices Used in Fertility Study Assessments**

$$\text{Male Mating Index} = \frac{\text{Number of males with confirmed matings}}{\text{Number of males cohabitated}} \times 100$$

$$\text{Male Fertility Index} = \frac{\text{Number of males impregnating a female}}{\text{Number of males cohabitated}} \times 100$$

$$\text{Female Mating Index} = \frac{\text{Number of females with confirmed matings}}{\text{Number of females cohabitated}} \times 100$$

$$\text{Female Fertility Index} = \frac{\text{Number of pregnant females}}{\text{Number of females cohabitated}} \times 100$$

$$\text{Fecundity Index} = \frac{\text{Number of pregnant females}}{\text{Number of females with confirmed matings}} \times 100$$

$$\text{Implantation Index} = \frac{\text{Number of implants}}{\text{Number of corpora lutea}} \times 100$$

$$\text{Preimplantation Loss} = \frac{\text{Number of corpora lutea} - \text{Number of implants}}{\text{Number of corpora lutea}} \times 100$$

$$\text{Postimplantation Loss} = \frac{\text{Number of implants} - \text{Number of viable fetuses}}{\text{Number of implants}} \times 100$$

$$\text{Gestation Index} = \frac{\text{Number of females with live born}}{\text{Number of females with evidence of pregnancy}} \times 100$$

Precoital Interval = average number of days after the initiation of cohabitation required for each pair to mate

Mean Gestational Length = Duration from GD 0 to parturition

organ pathology, clinical signs of toxicity, and mortality. Maternal toxicity may be evidenced by indicators such as alterations in the estrous cycle or mating behavior, pre- and postimplantation losses, decreased corpora lutea or ovarian follicle counts, and altered weights and histopathology of the reproductive organs. Paternal toxicity may be evidenced by indicators such as mating behavior, fertility, altered weights and histopathology of the reproductive organs, and abnormal sperm motility and morphology. Fertility data can be analyzed by using the formulas listed in Table 4. For further descriptions of each parameter, see Parker [34].

### 3.12 Interpretation of the Data

***Importance of Historical Control Data*** Of particular concern in reproductive toxicity testing is the development and maintenance of a historical control database [35]. The data should be derived from animals from the same supplier and strain and should be kept current, as animal characteristics can change over time. This database of male and female reproductive information is useful in situations where the data are not otherwise obvious. Published databases are useful for comparison, but they do not take the place of internal historical controls. One must always be

aware, however, that even data from animals from a single source may vary unexpectedly or drift over time.

***The Need for Expert Judgment and Specialized Technical Training*** To assess Segment I studies, the toxicologist must be proficient in reproductive toxicology. To become proficient in reproductive toxicology requires training and experience in a number of different fields, including reproductive anatomy, physiology, endocrinology, pathology, toxicology, and behavior. Fertility is a complex issue encompassing male and female reproductive system function, mating, conception, and implantation.

***Reporting the Results of Reproductive Toxicity Studies*** Guidelines for reporting the results of reproductive toxicity studies are contained in U.S. FDA *Redbook 2000* [36]. More complete information regarding pathology and statistics can be found in *Redbook* Chapters IVB3c and IVB4a, respectively.

Reproductive toxicity studies are conducted in compliance with Good Laboratory Practice Regulations [37], as specified in Section 21 of the U.S. Code of Federal Regulations (CFR) 58, or, if the study was not conducted in compliance with GLPs, a signed statement including a list of the specific area(s) of noncompliance and the reason(s) for the noncompliance must be included with the study report. Each study report should also include a signed record of periodic inspections conducted by the Quality Assurance Unit (QAU), showing the date of inspection, the phase or segment of the study inspected, and the date the findings were reported to management.

The study report should include all information necessary to provide a complete and accurate description and evaluation of the study procedures and results. The following sections should be included:

- Protocol and Amendments.
- Summary and Conclusions. This section of the study report should contain a brief description of the methods, summary of the data, analysis of the data, and a statement of the conclusions drawn from the analysis.
- Materials. The materials section of the study report should include, but not be limited to, the following information: test substance, storage, animal data, feed and water, diet and availability, facilities, caging, number of animals per cage, bedding material, ambient temperature, humidity, and lighting conditions.
- Methods. The methods section of the study report should include, but not be limited to, the following: deviations from guidelines, specification of test methods, statistical analyses, dosage administration, duration of treatment and study, and method and frequency of observation of the test animals.
- Results. Each study report must include individual animal data and results and tabulation of data must be provided in sufficient detail to permit independent evaluation of the results. The following information should be included for each test animal, where appropriate: time of first observation of each abnormal sign and its subsequent course; time of death during the study; feed consumption data (and water consumption data, if the test compound is administered in the drinking water); body weights and body weight changes; hematology, clinical

chemistry, and other clinical findings; gross necropsy findings including description and incidence of gross lesions; and histopathology findings, including description and incidence of microscopic lesions.

- Data also should be summarized in tabular form, organized by sex and dosage group. When appropriate, data should also be organized by litter. When numerical means are presented, they should be accompanied by an appropriate measure of variability, such as the standard error. For each summarized parameter, the following information should be included: the number of animals at the beginning of the study, the number of animals evaluated for each parameter, the time when animals were evaluated for each parameter, and the number and percentage of animals positive and/or negative for each parameter in comparison to controls.
- All numerical results should be evaluated by an appropriate statistical method. Evaluation of the results should include the following: statements about the nature of relationships, if any, between exposure to the test substance and the incidence and severity of all general and specific toxic effects (such as organ weight effects and mortality effects); statements about the relationships between clinical observations made during the course of a study and postmortem findings; and an indication of the dosage level at which no toxic effects attributable to the test substance were observed.
- References. This section of the study report should include the following information: availability and location of all original data, specimens, and samples of the test substance. The literature or references should include, when appropriate, references for test procedures and statistical and other methods used to analyze the data; compilation and evaluation of results; and the basis on which conclusions were reached.

## 4 PRENATAL DEVELOPMENTAL TOXICITY STUDY (SEGMENT II)

### 4.1 Rationale for the Test

In the developmental toxicity test (i.e., Segment II test, teratology study), pregnant laboratory mammals are exposed to a candidate drug to assess the potential for adverse effects on the developing offspring. This test assesses only prenatal growth, morphological development, and survival, as exposure occurs during gestation and offspring evaluation is carried out prior to the anticipated time of parturition.

### 4.2 Dose Range-Finding Study

A dose range-finding or range-characterization study is a useful preliminary to the developmental toxicity test, and this is especially true for rabbits, as adequate prior toxicity data are seldom available for them. Dosing small groups of dams (generally 5–10 per dosage group) in a pilot study is helpful in selecting appropriate dosages for the definitive study. Although dose selection for this study should be based on any available toxicity and pharmacokinetics data, dosing during gestation in the range-finder controls for physiological differences between pregnant and nonpregnant animals that might cause the choice of inappropriate dosages if selection were

based only on findings from the latter. Furthermore, such a study can provide useful preliminary information regarding the test compound's potential for effects on the conceptus. Caution should be used, however, in interpreting the results of a range-finding study, as the small sample size of each treatment group can result in atypical results and inaccurate statistical measures (both Type I and Type II errors), and typically fewer endpoints are examined.

### 4.3 The Definitive Study

***Animal Species and Strain Selection*** Most commonly, developmental toxicity tests are conducted in two species, rats and rabbits, with mice sometimes used in place of the rat. Other species may be used if they are deemed more relevant in terms of extrapolation of test findings to humans, on the basis of pharmacokinetic or metabolism data or other known attributes of the test compound, or if there are specific toxicity concerns. For example, rabbits are not ideal test animals for antibiotics, as they are dependent on the presence of viable microbial flora in the gut to aid digestion.

Although use of the most relevant species is often said to be desirable, no single animal species has been shown to be an obviously superior model of the human response to developmental toxicants. Thus, rats, rabbits, and mice are used far more often than alternative test species, primarily because of considerations of practicality. They are also used because of regulatory agency recommendations for at least one rodent and one nonrodent species [1]. These animal models offer ready availability, low cost, ease of handling, high fertility, ease of breeding, large litters, short gestation lengths, ease of determination of mating time, low rates of spontaneous deaths and developmental abnormalities, ease with which their fetuses can be examined, and availability of large amounts of information on their reproduction, development, physiology, and response to developmental toxicants [38].

***Procurement and Acclimation of the Test Animals*** The test animals should be young but fully sexually mature and from the same source and strain, and the females should be nulliparous and as nearly the same age and weight as is practicable. They should be obtained from reputable suppliers that have programs ensuring the genetic integrity of their animals, in addition to their freedom from diseases and parasites. The test animals should be acclimated to the laboratory conditions prior to mating.

Enough females should be used to provide approximately 20 pregnant animals at the time of sacrifice. That is considered to be a minimal number of females to provide adequate data for a developmental toxicity evaluation. Mating 25 rats or mice and 22 rabbits will generally provide at least 20 pregnant dams per test group. Procuring additional female rodents (30–50%) should provide enough matings to result in around 25 mated per test group within a 3–4 day period, whereas an additional two does per group should yield 20 pregnant rabbits at sacrifice [39].

***Mating Procedures, Detection of Mating, and Housing of Mated Females*** Upon receipt from a supplier, female rodents can be group housed. Once they are mated, females should be housed individually to avoid harassment by dominant individuals and to allow for monitoring of feed consumption. Feed and water should be

provided *ad libitum*, except in the case of newly received rabbits, as they may initially benefit from decreased feed availability [39].

It is typical to mate only one or two females to each male, and sibling matings are to be avoided [40]. The number of females to expose to males is limited by the number of litters that can be processed on a given day. As female rodents go through 4–5 day estrous cycles, only a portion will be in estrus and mate on any given day. Thus, it is useful to check female rats for estrous cycle stage by vaginal lavage with physiological (0.9%) saline followed by microscopic examination of the cells thus obtained [34]; only females found to be in estrus or late proestrus should then be placed with males. Care should be taken, however, not to stimulate the cervix when obtaining the cell sample, so as not to induce pseudopregnancy. If estrous cycle stage is not determined, considerably more females must be placed with males than are required to mate on a given day.

Female mice can be influenced by pheromones. If groups of females are housed together, they may cease cycling regularly, but if a male in a wire cage or bedding from a male's cage is placed in a cage housing females, this can synchronize their estrous cycles. A large percentage of female mice can be expected to mate 3 days later if placed with males after such an exposure. Because rabbits are induced ovulators [41], the number mating on a given day is more easily controlled.

Typical ages at mating for rodents and rabbits are shown in Table 5, and they differ somewhat among laboratories according to preference. Female rodents to be mated are placed in the male's cage. Rats can be checked each morning by vaginal lavage for the presence of sperm, while mice are examined for the presence of a copulation plug in the vagina as evidence of mating. Alternatively, if rats are kept in wire bottom cages the pan below each cage can be inspected for a copulation plug, as the plug falls out intact by about 6–8 h after mating. Copulation plugs in mice remain in the vagina and are often readily visible. However, in some cases it is necessary to use a small probe to expand the vaginal opening to facilitate detection of the plug, and even with experienced examiners, a few matings can go unrecognized. It is also possible to purchase timed-mated rats and mice, but this practice does not allow for an acclimation/quarantine period, and shipping stress can cause implantation failure. The day on which mating is detected in rodents is, by convention, most commonly referred to as gestation day 0 (GD 0).

Ovulation in rabbits can be induced by natural mating or by injection of luteinizing hormone. Hormone injection is used in conjunction with artificial insemination, employing semen collected via an artificial vagina. Semen can be collected by enticing the buck with an infertile "teaser" doe or by means of a rabbit skin sleeve on the collecting technician's arm. Artificial insemination allows examination of the

**TABLE 5 Typical Ages and Weights of Laboratory Animals at Time of Mating for Use in Developmental Toxicity Tests**

| | Age at Mating (weeks) | | |
|---|---|---|---|
| Test Species | Females | Males | Weight of Females at Mating |
| Rat | 8–12 | 10–25 | 200 g |
| Mouse | 8–12 | 10–25 | 25 g |
| Rabbit | 5–6 | 29–104 | 2.5–5 kg |

semen prior to use, and more than one female can be inseminated with a single ejaculate. However, care must be taken to eliminate any bucks that produce abnormal offspring from untreated females [39]. It is also possible to purchase timed-pregnant rabbits if the vendor is based in close proximity to the test facility. The day on which mating is observed or the day on which the doe is inseminated is designated as GD 0. GD 0 body weights provided by vendor are often used for randomization.

***Selection of Dosage Groups and Route of Administration*** Typically, developmental toxicity studies are designed with a vehicle control and at least three dosage groups, all of which should be concurrent and housed and handled identically. Mated females should be assigned by a random procedure in a manner that assures that mean maternal body weights are comparable among all test groups and that mating days are similarly represented across dosage groups. The test compound or vehicle must be administered by the route that most closely approximates the pattern of expected human exposure. The highest dosage group should be selected to induce observable maternal toxicity, but without an excessive incidence of maternal mortality (often defined as no more than 10% maternal deaths) [42] or of complete litter loss.

For a given route of administration, the physicochemical characteristics of the test article or dosage formulation may impose practical limitations on the amount that can be successfully administered. In the absence of observable maternal toxicity, a high dose of 1000 mg/kg, a so-called limit dose, is generally acceptable [7]. Alternatively, a high dose that is at least 1000 times the highest expected human dose may be adequate. Ideally, the low dose should be selected so as to provide an NOAEL and to provide a margin of safety [42]. The intermediate dose should be a threshold dose, and the high dose should approximate the maximum tolerated dose (MTD). In practice, this ideal is difficult to achieve with only three dosage levels, making the added information from a well-conducted preliminary dose-characterization study of considerable value. Also, it must be kept in mind that developmental toxicity dose–response curves are often relatively steep, and thus wide separation of dosages above the threshold dose should generally be avoided [43]. Moreover, the dosage intervals should be close enough so as to reveal possible dose-related trends [7].

***Treatment Timing and Test Duration*** The originally recommended developmental toxicity test (Segment II) study design called for dosing to extend from the time of embryonic implantation until closure of the hard palate of the specific test species (Figs. 2 and 3), thereby encompassing the period of major organogenesis. Thus, treatments were administered as shown in Table 6. However, recently the FDA has recommended extending the dosing period until the day before the scheduled necropsy and fetal examination [40]. The FDA also considers treatment beginning on GD 0, rather than after implantation, an acceptable alternative. Starting the treatment period prior to implantation is particularly desirable for drug candidates that require prolonged exposure to achieve a steady-state concentration in the maternal system [43]. The test compound is usually administered once daily, but pharmacokinetic considerations may suggest an alternative schedule. Dosing should take place at approximately the same time each day for all dosage

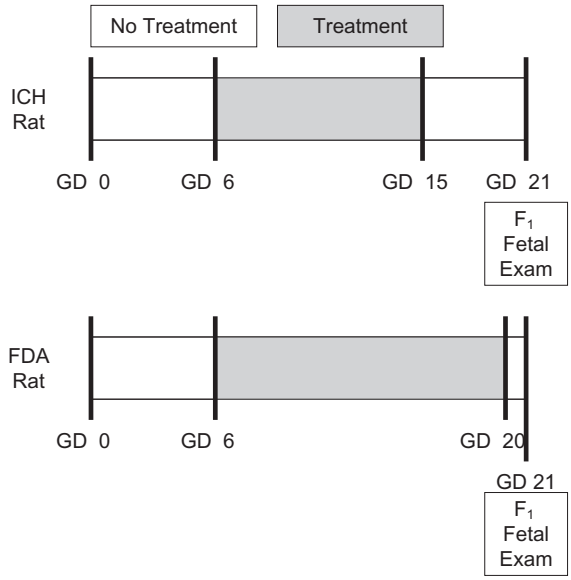


**FIGURE 2** Rat developmental toxicity study schematics (FDA Segment II and ICH OPPTS 870.3700 and OECD 414).

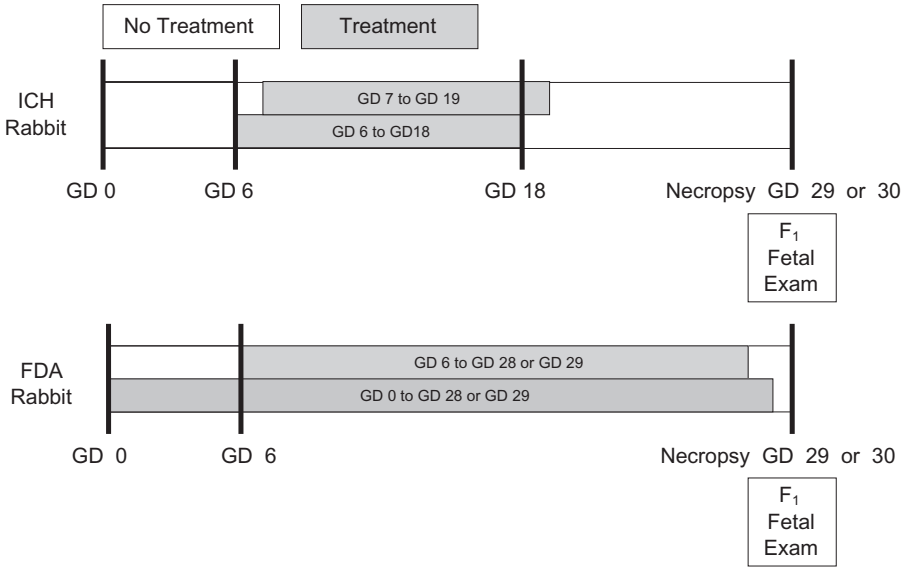


**FIGURE 3** Rabbit developmental toxicity study schematics (FDA Segment II and ICH S5A).

groups. Aside from considerations of diurnal maternal physiological differences that may influence toxicity response in the dam, development of the offspring is rapid and exact stage of development at the time of toxic insult can influence any prenatal effects.

**TABLE 6 Timing of Maternal Treatment and Necropsy in Developmental Toxicity Tests**

| | Guideline and Dosing Period | | |
|---|---|---|---|
| Test Species | ICH[a] | FDA[b] | Necropsy Day |
| Rat | 6–15 | 6 (or 0)–19 or 20 | 20 or 21 |
| Mouse | 6–15 | 6 (or 0)–17 or 18 | 17 or 18 |
| Rabbit | 6–18 | 6 (or 0)–28 or 29 | 29 or 30 |

[a]From Ref. 1.
[b]From Ref. 40.

**TABLE 7 Typical Endpoints for Assessment of Maternal Toxicity in Developmental Toxicity Studies**

Mortality
Mating index [(Number with seminal plugs or sperm/Number mated) × 100]
Fertility index [(Number with implants/Number of matings) × 100]
Gestation length (useful when animals are allowed to deliver pups)
Body weight
  Gestation day 0
  During gestation
  On day of necropsy
Body weight change
  Throughout gestation
  During treatment (including increments of time within treatment period)
  Posttreatment to sacrifice
  Corrected (body weight change throughout gestation minus gravid uterine weight or litter weight at sacrifice)
Organ weights (in cases of suspected target organ toxicity and especially when supported by adverse histopathology findings)
  Absolute
  Relative to body weight
  Relative to brain weight
Food and water consumption (where relevant)
Clinical evaluations
  Type, incidence, degree, and duration of clinical signs
  Clinical chemistries
Gross necropsy and histopathology

*Source*: Adapted from Ref. 62.

***Maternal and Fetal Endpoints Measured*** Typical maternal and fetal endpoints are shown in Tables 7 and 8, and selected aspects of the test protocols are discussed in the following sections.

*Clinical Signs, Maternal Body Weight, and Feed Consumption* The FDA suggests that mated females be observed at least twice daily for clinical signs of toxicity and mortality and advises that the first examination be a thorough examination, while the second can be limited to observing the animals in their cages [40]. If possible, the second observation should be timed appropriately after dosing to allow for detection of the onset and progression of any toxic or pharmacologic effects, and

**TABLE 8 Typical Endpoints for Assessment of Developmental Toxicity**

| |
|---|
| *Litters with Implants* |
| Number of implantation sites/dam |
| Number of corpora lutea (CL)/dam |
| Percent preimplantation loss: (CL − Implantations) × 100/CL |
| Number and percent live fetuses/litter |
| Number and percent resorptions/litter |
| Number and percent litters with resorptions |
| Number and percent late fetal deaths/litter |
| Number and percent nonlive (late fetal deaths + resorptions) implants/litter |
| Number and percent litters with nonlive implants |
| Number and percent affected (nonlive + malformed) implants/litter |
| Number and percent litters with affected implants |
| Number and percent litters with total resorptions |
| Number and percent stillbirths/litter |
| Number and percent litters with live fetuses |
| *Litters with Live Fetuses* |
| Sex ratio/litter |
| Mean fetal body weight/litter |
| Mean male or female fetal body weight/litter |
| Number and percent fetuses with malformations/litter |
| External |
| Visceral |
| Skeletal |
| Number and percent malformed fetuses/litter |
| Number and percent litters with malformed fetuses |
| Number and percent fetuses with variations/litter |
| External |
| Visceral |
| Skeletal |
| Number and percent fetuses with variations/litter |
| Number and percent litters having fetuses with variations |
| Types and incidence of individual fetal malformations |
| Types and incidence of individual fetal variations |
| Individual fetuses with malformations or variations (grouped by litter and dose) |

*Source*: Adapted from Ref. 62.

to permit moribund or dead animals to be removed and preserved prior to excessive tissue deterioration.

All dams should be weighed on GD 0 and daily or at least every 3 days during the dosing period and at necropsy. If dosing does not commence until GD 6, the dams should also be weighed on GD 3. Gains and losses of body weight are often comparatively sensitive indicators of maternal toxicity. Body weight changes are more sensitive than body weights as indicators of effects, and at times they may be the only means of demonstrating toxicity at the high dose. Such body weight changes should be calculated during any preexposure period, for each interval for which weights were recorded, and for the entire period of exposure. This allows detection of temporary effects and compensatory weight gains.

Regulatory guidelines [8, 40] suggest measuring feed consumption weekly as a minimum, but taking such measurements on the same days as body weights are obtained will give more useful data in those instances in which maternal toxicity alters test diet intake. Feed consumption can be calculated as grams/animal/day or as grams/kg/day to adjust for relative maternal size. Water consumption should be measured if suggested by prior test results on the current or similar test compounds.

*Maternal Toxicity and Its Significance* Maternal toxicity may be evidenced by indicators such as body weight or weight gain, feed and water consumption, organ weights, gross or microscopic organ pathology, clinical signs of toxicity, and mortality. Of course, interpretation of putative signs of toxicity should consider the timing of their onset or disappearance in relation to the period of dosing in cases where treatment does not span the entire time between mating and necropsy.

In addition to fulfilling the requirement for evidence of maternal toxicity at the high dose, signs of such toxicity have a special significance in developmental toxicity testing. That is because these tests involve two organisms concurrently, the dam and the conceptus (the products of conception, i.e., the embryo or fetus and its membranes), and the conceptus is dependent on the dam for its survival and well-being. Deleterious effects on the embryo/fetus in the absence of discernible maternal toxicity are viewed as especially troublesome, as they suggest that the developing organism is especially sensitive to the test compound. However, if adverse effects on the offspring are observed only at maternally toxic dosages, their interpretation can be confounded by the presence of said maternal toxicity. The default position of the U.S. FDA [40] is that developmental effects occurring in the presence of minimal maternal toxicity are considered to be evidence of developmental toxicity unless it can be shown that those effects are clearly maternally mediated. If developmental effects are seen only at substantially maternally toxic dose levels, then the possibility that said effects were maternally mediated should be investigated before a valid assessment can be made regarding the hazard potential of the test compound. The FDA's position appears to be valid, in that numerous examples of relatively severe maternal toxicity without concomitant developmental effects have been reported, although other instances have been noted in which mechanisms have been found to support the presumption of maternally related causation of embryo–fetal toxicity or teratogenicity [44]. An example of the latter was the finding that diflunisal-associated skeletal defects in rabbits were apparently caused by maternal hemolytic anemia [45]. Additional discussion of the relationship between maternal and developmental toxicity can be found in the review by Hood and Miller [46].

*Assessment of the Dam, Uterine Contents, and Fetuses at Necropsy* Any mated females that appear moribund or that appear to have aborted or prematurely delivered should be euthanized appropriately and necropsied to investigate the cause of death. Animals found dead should also be necropsied. In all such cases any evidence of pregnancy and of embryonic or fetal development [47] should be evaluated as well. All surviving females should be humanely euthanized prior to term (see Table 6 for timing), laparohysterectomized, and examined as described in the following.

INSPECTION OF MATERNAL VISCERA, ENUMERATION OF CORPORA LUTEA, AND ASSESSMENT OF PRENATAL MORTALITY The gravid uterus should be removed and weighed, and the female's carcass should be weighed to allow determination of the net maternal weight at sacrifice (the maternal body weight without the uterine weight). The remaining maternal abdominal and thoracic organs should be inspected grossly for evidence of toxicity, and the liver and any known target organs should be removed and weighed. If gross organ pathology is noted or anticipated, the affected organs should be preserved for possible histopathologic examination. The numbers of corpora lutea should be counted and recorded. Corpora lutea of rats and rabbits are relatively obvious, but those of mice must be counted with the aid of a dissecting microscope. Such counts aid in assessing preimplantation mortality, keeping in mind that evidence of preimplantation mortality is only treatment related if dosing began prior to implantation.

The uterus should be examined to determine if the animal was pregnant, and the relative numbers of live, dead, and resorbed offspring and their relative positions in the uterus should be recorded. If the uterus shows no signs of implantation, it should be treated with a reagent, such as 10% ammonium sulfide, to reveal any early resorptions (partially autolyzed embryos) that may not have been obvious [18]. Each remaining uterus should be cut open so that the contents can be examined and characterized. Different characterization schemes can be used, but all bear similarity to the classification scheme shown in Table 9. Once fetuses are removed from the uterus and determined to be alive, they should be anesthetized or euthanized prior to weighing or any further examination.

ASSESSMENT OF FETAL GROSS MORPHOLOGY AND BODY WEIGHT Generally, fetuses from the control and all dosage groups are weighed and examined grossly for developmental defects. In some cases inspection of the low dose and mid-dose fetuses for visceral or skeletal anomalies can be omitted if no differences are found between the control and high dose groups, but in such cases, fixation and storage of the unexamined fetuses is advisable.

Once each fetus is removed from the uterus, its umbilical cord should be cut, and the fetus should be blotted on absorbent paper to remove blood and amniotic fluid. Especially in the case of mouse fetuses, the paper should be slightly damp lest the fetal skin adhere to it and become torn. Then each fetus is carefully inspected for externally visible anatomical abnormalities. This examination should be conducted in a standardized, repeatable manner for each fetus, and the results recorded as either normal or abnormal, with any deviations from normal specifically described. In addition, the sex of each rat or mouse fetus should be assessed by careful measurement of the anogenital distance, which is roughly twice as great in males as in females. Rabbit fetuses must be sexed internally. Fetal inspections of rats, and especially of mice, should be carried out under magnification. Additional endpoints, such as fetal crown-to-rump distance, can also be measured if desired.

The placenta should also be grossly evaluated. Some laboratories weigh the placentas or preserve them in fixative for possible histopathological evaluation. Although not required by regulatory guidelines, placental weights can be useful in cases in which the test chemical may have estrogenic properties.

Each fetus should be numbered individually, weighed, and placed initially in a multi compartmented container, so that all fetuses can be individually identifiable

**TABLE 9 Comparison of Advantages and Disadvantages of the Major Visceral Examination Methods Used in Developmental Toxicity Studies**

| Fresh Dissection (Staples' Method) | Free-Hand Sectioning (Wilson's Method) |
|---|---|
| The same fetus can be examined for both soft tissue and skeletal changes, if desired, increasing the reliability of the study findings and statistical analyses. | Fetuses cannot be examined for skeletal defects once they have been sectioned, so only half of the total available fetuses can be examined for either endpoint. |
| Training of technicians is simpler, as the viscera are examined *in situ*; that is, structures are visualized as they are learned in anatomy courses. | Technicians must be trained in the normal appearance of two-dimensional serial cross sections and in visualizing anatomical alterations in each such section. |
| Allows the use of the same general technique for rodents and nonrodents. | Wilson's method cannot be used for nonrodents. |
| Decreases the likelihood of overlooking clefts, breaches, or absence of a structure (e.g., a heart valve). | Certain anomalies are more likely to be missed. |
| Avoids the difficulty of sectioning every fetus in exactly the same way, as is needed to ensure comparability of views among fetuses. | Requires relatively precise and consistent sectioning of each fetus, including fetuses of different sizes. |
| Allows viewing fresh coloration of organs, vessels, and tissues; detection of color changes; and evaluation of vascular tissue perfusion by coloration. | All structures may be colored more or less alike. |
| The probability of creating artifacts is relatively low. | Artifacts may be created by improper sectioning or fixation. |
| No chemicals are required except for fixing fetal heads, although a method is now available that uses quick freezing of the heads with dry ice [53]. | Has traditionally required use of picric acid, which is a strong irritant, an allergen, and explosive, although other fixation methods have now been described [52]. |
| Organs can be saved for histopathology if desired, although this is rarely done. | Organs can't be saved intact, but all sections can be saved for future re-inspection if needed. |
| Allows shorter study turnaround. | Requires extra time for fixation and decalcification of fetuses (approximately 1 week). |
| Requires that more technicians are available to process the fetuses in a timely manner. | Fetuses can be stored until technicians are available to examine them. |

throughout the assessment process. Individual identification is then assured by means such as use of chemical-resistant tags. Rabbits can be tagged or numbered on top of the head with a permanent marker (assuming that the area is normally intact). Individual fetal identification is important so that all findings, such as body weight and gross, visceral, and skeletal examinations, can be related to a specific individual.

FETAL VISCERAL EVALUATION Many fetal abnormalities are only detectable internally, so it is imperative that both the soft tissues and the skeleton are examined.

All nonrodent fetuses, being larger, are examined for visceral defects by fresh dissection, with only the heads subjected to sectioning. Typically half of each litter of rodent fetuses is randomly selected and examined by either a fresh dissection (Staples') method [39, 48] or the free-hand sectioning (Wilson's) method [49, 50]. There are positive and negative attributes of both methods (Table 9) but there are more positives for the fresh dissection method if enough technicians are available to examine all fetuses in a reasonable amount of time. Even if availability of technical staff is limiting, rodent fetuses can be refrigerated for a day or so prior to visceral examination if necessary. Visceral examination is often best carried out under magnification, and detailed instructions for visceral exams can be found in the references listed above, and in other sources, such as Christian [51].

Whole fetuses and fetal heads to be examined by Wilson's method must be preserved with an appropriate decalcifying fixative prior to sectioning. It should be noted that alternate fixatives, such as Davidson's [52], may be preferable to Bouin's fixative for that method. They avoid the use of picric acid, and they may produce superior contrast and enhanced definition of organs and vessels, as well as allowing the sections to remain moist for a greater length of time. Another method for preparing fetal heads for sectioning has been described for rat fetuses by Astroff et al. [53]. This method employs freezing with dry ice, avoids the use of fixative chemicals entirely, and avoids the delay of several days necessitated by use of chemical fixatives.

FETAL SKELETAL EVALUATION All rabbit fetuses and at least half of each litter of rodent fetuses should be processed and examined for skeletal defects. There are two general skeletal staining techniques, with several published variations of each. For staining only the ossified structures, rodent fetuses are first fixed in 70% ethanol. Then each fetus is cut open along the ventral midline, and the viscera are removed. The soft tissues are macerated with a strong alkaline solution and simultaneously stained with alizarin red S, and then the soft tissues are "cleared" (i.e., made transparent), by use of glycerin-containing solutions. Rabbit fetuses are processed similarly, following their visceral examination, but should be skinned first. Alternatively, the cartilage is stained blue with alcian blue, and the mineralized bone is stained red with alizarin red S. To facilitate penetration of the alcian blue dye, rats and rabbits should be skinned prior to this "double-staining" process. Moreover, although the original double-staining technique called for use of glacial acetic acid, hazards posed by exposure to that substance make its use problematic. According to Webb and Byrd [54], "replacement of the glacial acetic acid with potassium hydrogen phthalate eliminates these hazards without compromising the quality of the stained specimen."

Although there is disagreement on the matter [51], there have been published comparisons claiming that double-staining was to be preferred [55]. Detailed instructions for macerating and staining fetuses can be found in a number of sources [54, 56–60]. Such methods can generally be adapted for use with automated staining systems designed for the purpose, and use of such systems can decrease processing time and labor. However, it must be kept in mind that unusually small fetuses, especially those of mice, can disintegrate if kept too long in strong alkali solutions. Maceration times may have to be adjusted in such cases, and this is especially of concern when automatic processors are employed.

***Analysis of Maternal and Fetal Data*** Obviously, both the maternal and fetal data must be subjected to appropriate statistical analysis. The FDA [40] has suggested that maternal body weights be compared by analysis of covariance, with adjustment for initial body weight, followed by analysis by protected least significant difference tests. The FDA also suggested comparison of fetal weight by nested ANOVA, fetal abnormalities by Fisher's exact test, and fetal mortality and litter incidence of anomalies by arcsine transformation followed by ANOVA. There has been considerable interest in exactly how fetal data should be analyzed; as the data may not be "normally distributed," they tend to be interrelated in various ways, and their endpoints range from relatively continuously distributed to the highly categorical. Fetal findings are commonly influenced by the so-called "litter effect" (i.e., fetuses within a litter tend to be affected more similarly), giving rise to the admonition that the litter, rather than the fetus or neonate, should be the experimental unit [61]. Moreover, inappropriately calculating fetal malformation data is a common error. The percentage of affected fetuses should be determined for each litter, allowing calculation of an overall mean of the litter means [43]. Complexities such as the foregoing make it imperative to consult with an experienced statistician, beginning at the study design stage.

### 4.4 Interpretation of the Data

***Importance of Historical Control Data*** Of particular concern in developmental toxicity testing is the development and maintenance of a historical control database for each test species. The data should be derived from animals from the same supplier and strain and should be kept current, as animal characteristics can change over time. This database of litter and fetal characteristics is invaluable in situations in which it is not otherwise obvious whether a finding, such as an apparent increase in resorptions (evidence of prenatal demise) or of specific malformations or developmental variations, is treatment related. Although comparisons with data from the same species and strain from other suppliers or collected by other laboratories may be of some value for comparison purposes, *such data cannot take the place of data collected in-house and from the same source*. One must always be aware, however, that even animals from a single source may vary unexpectedly over time.

***Possible Adverse Effects on the Offspring*** There are four major categories of toxicity effects on the conceptus: mortality, structural alteration, altered growth, and functional deficit. Prenatal mortality can be revealed as pre- or peri-implantation loss (which may be treatment related if dosing was initiated prior to implantation), early or late resorption, abortion, or fetal death. Manifestations of abnormal fetal anatomical development (dysmorphology) are often categorized as malformations or variations. Malformations can be defined as permanent structural changes that may adversely affect survival, development, or function, whereas variations may be regarded as deviations beyond the usual range of structural constitution that may not adversely affect survival or health [62]. However, the division between variations and malformations is somewhat artificial, as there is a continuum of responses from the normal to the extremely abnormal, and there are numerous classification schemes for malformations and variations [63]. Growth alteration is almost always

seen as diminished growth, although there can be other possibilities, such as localized overgrowth or altered anogenital distance, and increased fetal weight can be secondary to generalized edema. Functional deficits can be manifested as abnormal physiological or biochemical parameters, although these are not routinely measured. Behavioral parameters are sometimes affected, but they can only be assessed postnatally.

***Evaluation of Low Incidence Effects*** The U.S. FDA [64] has defined "rare event" as "an endpoint that occurs in less than 1% of the control animals in a study and in historical control animals." When such small numbers of adverse fetal findings, such as malformations, are observed, it is often difficult to ascertain whether they are spontaneous background events or the result of test article toxicity. The incidences of malformations in typical guideline studies are generally too low to be considered statistically significantly elevated above the control incidence, thus increasing the difficulty in separating such events from normal biological variation. Nevertheless, appearance of two or three malformed fetuses in a test group, especially if they occur in different litters, is cause for concern that they may be treatment related.

Detailed guidance for evaluation of the biological significance of low incidence adverse events can be found in the discussion of the issue by Holson et al. [43]. They advise that careful examination of the test facility's historical control database is the logical initial step in assessing whether a small, statistically nonsignificant alteration in an endpoint should be considered a possible treatment effect. They further advise that reliable estimation of the incidence of a low incidence developmental defect in the test population requires approximately 100 control datasets, thus rendering the concurrent controls relatively unhelpful for discerning treatment-related versus spontaneous occurrences for low incidence developmental toxicity endpoints.

Distinguishing between spontaneous adverse findings and those caused with low frequency by a test compound is a daunting task for developmental toxicologists, because the usual insights provided by statistical analyses are generally unavailable. Thus, Holson et al. [43] suggest reliance on additional tools, such as "probabilistic theory, intuition for developmental processes and patterns, and arrays and timing of effects."

***The Need for Expert Judgment and Specialized Technical Training*** It may perhaps be appreciated from the foregoing that developmental toxicology is a complex subject. It requires knowledge of a number of different fields, including embryology, anatomy, physiology, reproductive biology, endocrinology, laboratory animal science, biochemistry, cell biology, statistics, and, of course, toxicology. Such knowledge is important because development of the embryo and fetus encompasses a highly complex series of many events occurring concurrently at the organ, tissue, cellular, and molecular levels, and these events are also continually changing over time. An additional level of complexity is added by the relationship of the embryo/fetus to the mother via the placenta and the extraembryonic membranes. That relationship is further influenced by the mother's physiology, which is changing throughout pregnancy. If postnatal survival and development are assessed, maternal behaviors and lactation are also influential.

Unfortunately, as affirmed by Kimmel et al. [65], "many of the current graduate programs have an academic/basic science focus that gives the student neither an understanding of the complex interplay of anatomy with physiological and biochemical processes, nor a perspective of more applied areas of toxicology." Relatively few students are currently being exposed to relevant courses, such as developmental anatomy, or even to basic anatomy and physiology, and they may have never even worked with live animals [66]. In-depth principles of developmental toxicology are not often taught in college or graduate school, even in toxicology programs, and few graduate students are engaged in research that prepares them to conduct or interpret guideline compliant studies, such as the prenatal developmental toxicity study. New entrants to the field are often life sciences majors. They are intelligent, but when they have not been exposed to the relevant disciplines much of their training must be on-the-job training. And then they are asked to design, perform, and interpret some of the most technically and numerically complex studies in the entire discipline of toxicology!

To make matters even more challenging, a large number of endpoints are examined in developmental toxicity assessments, and it is not unusual for at least a few of the findings to be problematic to interpret. In part, that is because some of these findings may be differences from the controls that are deemed statistically significant merely because of the combination of biological variability and the large number of comparisons that must be made. Others may fall into the category of "rare events," as described earlier. All of these factors make expert judgment of considerable importance in interpreting developmental toxicity test findings, and especially in evaluating potential human risk. They also highlight the relevance of organized approaches to interpretation of the data, as detailed in Holson et al. [43].

### 4.5 Reporting the Results of Developmental Toxicity Studies

Study reports should include complete tabulated data on the dams and individual fetuses, as well as tables summarized by litters. The data should be calculated by litter incidence and by the number and percent of litters with a specific endpoint. The dosages employed should be reported as mg/kg/day, that is, milligrams of test compound per kilogram of body weight per day [40]. It is also extremely important to report all relevant historical control data, as this is critical in allowing reviewers to evaluate the biological significance of findings, such as malformations, variations, resorptions, and the like. The historical control data should include all significant factors, such as the dates when they were obtained, the strain of animals used (which should preferably be the strain that was used in the current study), the group sizes, and the range as well as the mean of the values for specific parameters.

The most common problems seen when developmental toxicity studies are reviewed by the FDA are failure to use adequate numbers of pregnant dams in each test group, not using random selection procedures for allocating dams to test groups, and use of the individual fetus rather than the litter as the experimental unit for statistical analysis [40]. It is also all too common to see errors in test protocol descriptions, and these often arise whenever "boilerplate" text is taken from previous reports and put together in new combinations.

## 5 PERINATAL AND POSTNATAL STUDY (SEGMENT III)

The Segment III study or perinatal/postnatal evaluation was designed to evaluate the effects of a test article on gestation, parturition, and lactation. The perinatal/postnatal study is generally performed with treatment of pregnant female rats beginning on GD 7 and continuing through gestation, parturition, and through a 21 day lactation period. This combines embryo, fetal, and perinatal/postnatal exposures in one study. Postnatal changes may be induced by insult at any time during treatment. Parturition and pup survival and development are monitored with observations for prolonged gestation, dystocia, impaired maternal nesting or nursing behavior, and postnatal mortality. If indicated by the results of this study or other studies, a cross-fostering procedure may be included to demonstrate whether findings in the pups are direct effects of exposure during gestation and/or lactation or if they are associated with abnormalities in maternal reproductive behavior or health. The peri/postnatal evaluation is not required by the FDA *Redbook*, OECD, or U.S. EPA guidelines. Rather, these evaluations are part of the multigeneration studies performed to meet the needs of these regulatory bodies, which consider this design an appropriate overall evaluation of reproductive capacity. The ICH guideline acknowledges the overlap in the methods used to test chemicals and pharmaceuticals for adverse effects on the reproductive process, but recommends that segmented study designs generally be used because humans usually take a pharmaceutical product for limited intervals, rather than an entire lifetime. However, the ICH document [2] also notes that lifetime exposure to a drug sometimes occurs, and that there may be drugs that can be more appropriately tested by exposure throughout the entire reproductive life (i.e., through use of a multigeneration study design) (see later discussion).

Rather than a standard "checklist" approach, the ICH guideline [1] emphasizes flexibility in developing the testing strategy. Although the most probable options are identified in the guideline, the development of the testing strategy is to be based on the following points:

1. Anticipated drug use, especially in relation to reproduction.
2. Form of the test article and route(s) of administration intended for humans.
3. Use of any existing data on toxicity, pharmacodynamics, kinetics, and similarity to other compounds in structure/activity.

The notes in the ICH guideline describe some methods appropriate for use in specific portions of the tests; however, it emphasizes that the individual investigator is at liberty to identify the tests used, and that the ICH guideline is truly a guideline and not a set of rules. The overall aim of the described reproductive toxicity studies is to identify any effect of an active substance on mammalian reproduction, with subsequent comparison of this effect with all other pharmacologic and toxicologic data. The objective is to determine whether human risk of effects on the reproductive process is the same as, increased, or reduced in comparison with the risks of other toxic effects of the agent. The document clearly states that, for extrapolation of results of the animal studies, other pertinent information should be used, including human exposure considerations, comparative kinetics, and the mechanism of the toxic effect.

### 5.1 Rationale for the Test

The Segment III study or perinatal/postnatal evaluation was designed to detect adverse effects following exposure of the female from implantation through weaning (Fig. 4). Since manifestations of effect induced during this period may be delayed, observations should be continued through sexual maturity (i.e., Stages C–F). The adverse effects assessed in a perinatal/postnatal study include enhanced toxicity relative to that in nonpregnant females, maternal nursing behavior (including milk production), pre- and postnatal death of offspring, altered growth and development, and functional deficits in offspring, including behavior, maturation (puberty), and reproduction.

### 5.2 Animal Species and Strain Selection

It is generally desirable to use the same species and strain as in other toxicological studies. Reasons for using rats as the predominant rodent species are practicality, comparability with other results obtained in this species, and the large amount of background knowledge accumulated. Within and between studies, animals should be of comparable age, weight, and parity at the start of the study.

***Number of Test Animals*** The number of animals per group should be sufficient to allow meaningful interpretation of the data. Routinely, 20–25 pregnant female

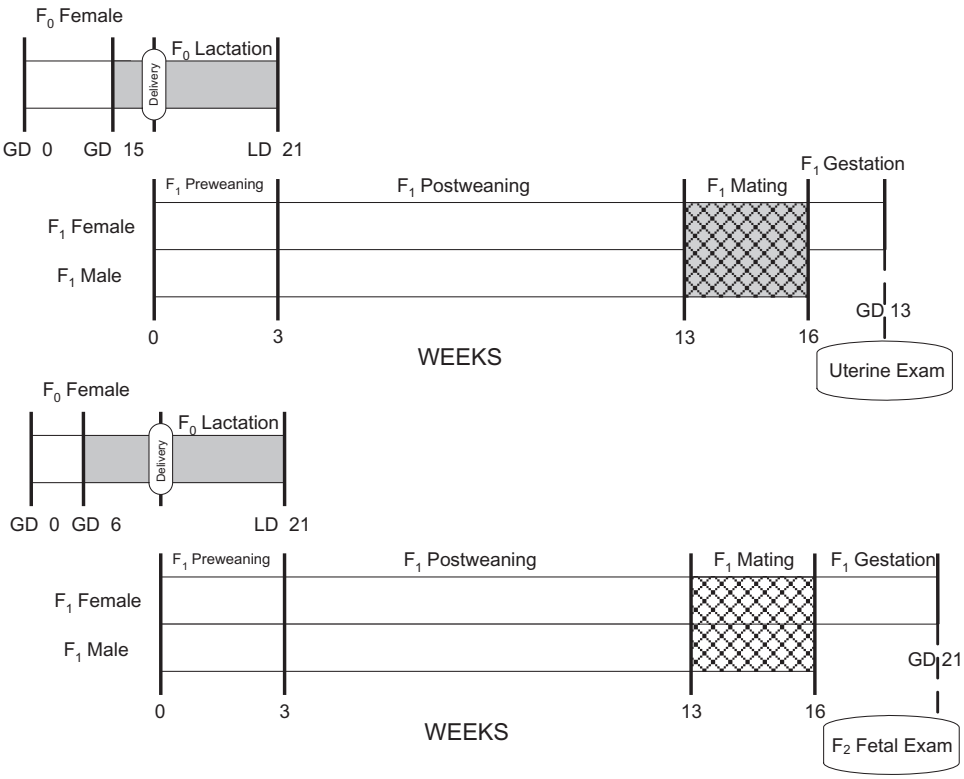


**FIGURE 4** Segment III study (perinatal/postnatal toxicity study).

rats per group are used. If lethality is expected, then an increased number of animals should be mated. Usually 16–20 litters per group are required for evaluation. With less than 16 litters per group, the study results can become inconsistent, while greater than 20–24 litters per group does not greatly enhance the group consistency.

***Statistical Considerations*** When using statistical procedures, the experimental units of comparison should be the dam and the litter, not the individual offspring.

The endpoints observed in reproductive tests are usually nonnormal and extend from the continuous to the categorical. The interpretation of the data *must* be based on biological plausibility. To assume that a difference from control values is not biologically relevant simply because it is not "statistically significant" is unfounded, while the assumption that a "statistically significant" difference must be biologically relevant is also unwise. Statistical tests should be used only as a support for the interpretation of results. The statistical tests and statistical programs used should be stated in the protocol and their use justified.

### 5.3 Procurement and Acclimation of the Test Animals

The test animals should be young but sexually mature and from the same source and strain, and the females should be nulliparous and as nearly the same age and weight as is practicable when mated. However, the availability of timed-mated pregnant rats from vendors has reduced the need to maintain a breeding colony to deliver the requisite number of pregnant rats. The pregnant rats can be ordered to arrive at the testing facility with staggered days of conception, allowing the dosing of subsets of each group to begin on different days (usually four). A 4 day staggered start will allow five pregnant females per group as a subset, allowing the pup evaluation workload to be distributed over several days for each task. The acclimation period may vary from the usual laboratory acclimatization period when timed-mated pregnant rats are used. If pregnant rats are ordered to be delivered to arrive at the testing facility on GD 2, 3, 4, and 5, and if treatment begins on GD 7, then the acclimation period would range from 2 to 5 days. This should be stated in the protocol.

### 5.4 Housing of Female Rats

Upon receipt from a supplier, nonpregnant females can be group housed. Once they are mated, dams should be housed individually to avoid harassment by dominant individuals and to allow for monitoring of feed and, if necessary, water consumption. Feed and water should be provided *ad libitum*. Alternately, upon receipt from a supplier, timed-mated pregnant dams can be individually housed.

### 5.5 Selection of Dosage Groups, Controls, and Route of Administration

Most perinatal/postnatal studies are designed with a vehicle control and at least three dosage groups, all of which should be concurrent and housed and handled identically. Dams should be assigned by a random procedure to assure that mean body weights are comparable for all test groups. The test compound or vehicle

should be administered by the route that most closely approximates the pattern of expected human exposure.

### 5.6 Treatment Timing and Test Duration

Females are exposed to the test article daily from implantation (GD 7) to the end of lactation (LD 21). Dosing should take place at approximately the same time each day for all dosage groups. $F_1$ generation pups are not directly given the test article and/or vehicle, but may possibly be exposed to the test article during gestation (*in utero* exposure) or via milk during the lactation period.

### 5.7 $F_0$ Generation

***Maternal Endpoints Measured***

*Maternal Body Weight, Feed Consumption, Behavior, and Clinical Signs* Clinical signs and mortalities should be recorded at least once daily. Body weight values should be recorded at least twice weekly but are routinely collected daily or at protocol-specified intervals (e.g., GD 0, 7, 9, 12, 15, 18, and 20, and LD 0, 7, 14, and 21). Food consumption should be measured at least once weekly up to LD 14 at the same interval as the body weight collection. Pups begin feeding on their own at this time, so food consumption measurement on LD 21 would not be reliable. Duration of pregnancy, parturition, and nursing behaviors, and any additional observations that have proved of value in other toxicity studies are also recorded.

*Assessment of the Dam at Necropsy* Dams are routinely necropsied on LD 21. The dam is weighed and euthanized, and the ovary and uterus are examined. Any mated females that appear moribund should be euthanized appropriately, and mated females that are found dead should be necropsied to investigate the cause of death. Any evidence of pregnancy and of embryonic development should be recorded.

*Inspection of Maternal Viscera* If necropsy occurs prior to parturition, then the gravid uterus should be removed and weighed. The uterus should be examined to determine if the animal was pregnant, and the relative numbers of live and dead conceptuses and their relative positions in the uterus should be recorded. If the uterus shows no signs of implantation, it should be treated with a reagent, such as 10% ammonium sulfide, to reveal any early resorptions [18]. The remaining maternal abdominal and thoracic organs should be inspected grossly for evidence of toxicity, and target organs should be removed and weighed. If gross organ lesions are noted, the affected organs should be preserved for possible histopathologic examination.

If necropsy occurs as scheduled or after parturition, then the uterus should be removed and weighed. The remaining maternal abdominal and thoracic organs should be inspected grossly for evidence of toxicity, and target organs should be removed and weighed. If gross organ lesions are noted, the affected organs should be preserved for possible histopathologic examination.

*Enumeration of Corpora Lutea* The number of corpora lutea should be recorded for each ovary as described earlier.

***Parturition*** Dams should be housed in litter boxes with bedding material no later than GD 20 (two days prior to expected parturition, earlier if required by protocol). The dams are observed periodically throughout the day for stretching, visual uterine contractions, vaginal bleeding, and/or placentas in the nesting box. These signs are strong indicators of parturition onset and indicate that dams should be closely monitored for the delivery of the first pup. As soon as the first pup is found, parturition is considered initiated. Once delivery has begun, the dams should be checked periodically for evidence of difficulty of labor or delivery (e.g., dam cold to touch or pale, partially delivered pup). Dams should not be treated during delivery.

The calendar date when delivery is completed is designated as LD 0, PND 0 (postnatal day 0), and day 0 of age. The following maternal behaviors are indicators that parturition is complete: removal of the amniotic sacs, placentas, and umbilical cords, and grooming of the pups by the dam; self-grooming; nesting behavior; and nursing. However, parturition may be complete without all of these behaviors present. Individual pup observations should not be made until delivery is completed. Apparently dead pups should be removed from the nesting box to preclude cannibalization by the dam.

***Milk Collection (if Required)*** Prior to milk collection, the pups are withheld from the dam for approximately 4 hours. At least 5 minutes prior to milk collection, 0.05 mL of oxytocin (20 units/mL) is administered intravenously to the dam. The time that the oxytocin is administered should be documented on the appropriate form. The dam is then held in a position that facilitates access to the teats. Milk is expressed from one teat at a time by applying gentle digital pressure. Initially the pressure is applied at the base of the teat and then is directed distally. The milk is collected into a microtainer or other suitable container (labeled with the protocol and animal number, date, identification of sample and test system, i.e., rat milk). The time of collection and volume of milk collected are documented, and the dam is returned to her litter.

### 5.8 $F_1$ Generation

***Litter Evaluations*** The general appearance of the entire litter and maternal and pup nesting behavior are observed. The following items should be checked: (1) that there is evidence of a nest, (2) that the pups are grouped together, (3) that the pups are warm and clean, and (4) that the pups are nursing or have milk present in their stomachs (milk spot).

After the litter observations, the bedding in the nesting box can be replaced with clean bedding or a new box with clean bedding can be obtained. The dam is weighed and observed for clinical signs, the feed jar is weighed (if appropriate), and the dam and pups are placed into the clean box.

*Gender Determination* The pups are removed from the nesting box, and the gender of each pup is determined. Anogenital distance (the length between the anus and the genital tubercle) is used to determine gender. Beginning on DG 17 and continuing through PND 21, anogenital distances are greater for male than for female rodents. The increased growth of this region occurs in response to testosterone.

Anogenital distances for rat fetuses (DG 20 or 21) or pups on PND 0 to 3 are measured with a micrometer and a stereomicroscope [43]. Pups that have been cannibalized to the extent that their gender cannot be determined are recorded as "sex unable to be determined because of cannibalization." If a malformation precludes sex determination, this is recorded (e.g., "tail absent, no apparent anus, sex could not be determined"). If the malformed pup is dead, the sex is determined internally.

*Viability* The viability of each pup is determined, and dead pups are removed from the nesting box to prevent cannibalization. The dead pup is necropsied, the trachea is tied off with string, and the lungs are removed and placed in a container of water. If the dead pup's lungs float, the pup is assumed to have breathed, and the viability is recorded as "born alive, but found dead." If the lungs sink, the pup is assumed to have never taken a breath, and the viability is recorded as "stillborn." Dead pups can be cannibalized to the point where viability cannot be determined. The sex and viability status are recorded for each pup (including dead, cannibalized, and malformed pups) in all litters.

*External Alterations* Each pup in the litter is evaluated for external alterations. The following items are evaluated for each pup: the general shape of the head; presence/completeness of features; trunk continuity; presence of bruises, lesions, or depressions; length and shape of limbs (digits are counted); length and diameter of the tail; presence of an anus; milk in the stomach; and whether partial cannibalization has occurred. All live pups are weighed.

*Missing Pups* If a pup cannot be found during the daily litter count, the pups are sorted by gender in order to determine the sex of the missing pup(s) by process of elimination. The missing pup is recorded as "missing, presumed cannibalized." If the pups have been tattooed or otherwise identified, the actual identification number of the missing pup(s) should be recorded.

*Pups Found Dead* All found dead pups are necropsied. Stomach contents of all dead pups should be recorded (i.e., "stomach contains/does not contain milk"). Unless otherwise dictated by the protocol, dead pups with gross lesions found on LD 0 to 4 should be fixed in Bouin's fixative. Gross lesions from pups found dead on LD 5 through weaning are fixed in 10% neutral buffered formalin. Because of postmortem autolysis, dead pups are not routinely weighed.

*Dam Dies Before Scheduled Sacrifice or All Pups in a Litter Die* If a dam with a litter dies before her scheduled sacrifice, the pups are sacrificed and necropsied. If all the pups in a litter die, the dam may remain on study until her scheduled sacrifice date, or the dam may be sacrificed at the discretion of the study director. The sacrifice and necropsy observations are recorded.

***Pup Evaluations***

*Pup Body Weights* Individual pup weights should always be measured on the day of parturition. The statistical unit is the litter; therefore, data from the individual

pups are reported as the entire litter weight and as weight by sex per litter. Birth weights are influenced by maternal nutritional status, intrauterine growth rates, litter size, and gestation length. Individual pups in small litters tend to be larger than pups in larger litters; conversely, individual pups in large litters tend to be smaller. Therefore, reduced birth weights that can be related to large litter size or increased birth weights that can be related to small litter size should not be considered an adverse effect unless the altered litter size is considered treatment related and survivability and/or development of the offspring have been compromised.

*Postnatal Pup Body Weights* Postnatal weights are dependent on litter size and on pup gender, birth weight, suckling ability of the offspring, and normality of the individual, as well as on maternal milk production. With large litters, small or weak offspring may not thrive and may show further impairments in growth or development. Because one cannot determine whether growth retardation or decreased survival rate was due solely to the increased litter size, these effects are usually considered adverse developmental effects. Conversely, pup weights in very small litters may appear comparable to or greater than control weights and therefore may mask decreased postnatal weights in other litters.

*Crown–Rump Length* Crown–rump lengths for fetuses are measured with a caliper. Fetuses may be euthanized before measuring the crown–rump length. Place each fetus on its side on a flat surface, and place the caliper with one prong touching the top of the head and the opposite prong touching the base of the tail. Do not stretch or compress the fetus. Record the measurements in millimeters.

***Standardization of Litter Sizes*** Animals should be allowed to litter normally and rear their offspring to weaning. Standardization of litter sizes is optional. If standardization is performed, the procedure should be described in detail in the protocol or Standard Operating Procedure. On PND 4, the size of each litter may be adjusted by eliminating extra pups by random selection to yield, as nearly as possible, four males and four females per litter or five males and five females per litter. Selective elimination of pups based on body weight, size, and/or activity is not appropriate. Whenever the number of male or female pups prevents having four (or five) of each sex per litter, partial adjustment (e.g., five males and three females, or four males and six females) is acceptable. Adjustments are not appropriate for litters of eight pups or less.

***Survival Indices*** The following survival indices should be calculated and reported: Live Birth Index and Survival Index on PND 4, 7, 14, and 21. These indices are primary endpoints in a reproduction study, reflecting the offspring's ability to survive postnatally to weaning. If litter standardization or culling of the litter is performed on PND 4, then the PND 4 *Survival Index* should be based on the preculling number of offspring. All subsequent *Survivability Index* calculations will use the number of offspring remaining after standardization as the denominator in the equation. If litter standardization was performed, the PND 21 survival index is known as the *Lactation Index*. If litter standardization was not performed, the PND 21 survival index is known as the *Weaning Index*. Pup survival endpoints can be affected by the toxicity of the test article, either by direct effects on the offspring (e.g., low birth

weight, impaired suckling ability, structural or functional developmental defects), indirectly through effects on the dam (e.g., neglect, poor milk production, acute intoxication), or by a combination of maternal and offspring effects. Pup survival may also be impaired by nontreatment-related factors, such as litter size or improper husbandry. The methods of calculation of the survival indices are as follows:

$$\text{Live Birth Index} = \frac{\text{Number of pups born alive}}{\text{Total number of pups born}} \times 100$$

$$\text{Survival Index} = \frac{\text{Total number of live pups (at designated timepoint)}}{\text{Number of pups born}} \times 100$$

***Sex Ratio*** Offspring gender in mammals is determined at fertilization of an ovum by sperm with an X or a Y chromosome. Altered sex ratios may be related to several factors, including selective loss of male or female offspring, sex-linked lethality (genetic germ cell abnormalities), abnormal production of X- or Y-chromosome-bearing sperm, or hormonal alterations that result in intersex conditions. The calculations should include both live and dead pups, because sex can easily be determined (except in the case of cannibalized pups or in some sexually ambiguous offspring). Treatment effects on sex ratio can indicate an adverse reproductive effect, especially if embryonic or fetal loss is observed. Because of the altered anogenital distance, visual determination of sex is often incorrect when the pups have an intersex condition (masculinized females or feminized males). The correct diagnosis is usually determined later during lactation when the condition becomes obvious or at the postmortem evaluation, when the reproductive organs are examined. The data are usually presented as a ratio, using the following formula:

$$\text{Sex ratio} = \frac{\text{Number of male offspring}}{\text{Number of female offspring}}$$

Data are often presented as a percentage of male offspring, using the following formula:

$$\text{Percentage of male offspring} = \frac{\text{Number of male offspring}}{\text{Total number of offspring}} \times 100$$

***Weaning*** Weaning is usually performed at LD 21. At weaning, one male and one female offspring per litter are selected for rearing to adulthood and for mating with a non-littermate from the same dosage group to assess reproductive competence.

***Assessment of $F_1$ Weanlings Not Selected for Continuation*** At the time of termination or death during the study, when the litter size permits, pups should be examined macroscopically for any structural abnormalities or pathological changes. Special attention should be paid to the reproductive organs. Dead pups or pups that are terminated in a moribund condition should be examined for possible defects and/or cause of death.

***Developmental Landmarks*** The onset of various developmental landmarks can be used to assess postnatal development. Sexual maturation landmarks (balanopreputial separation and vaginal patency) are required in perinatal/postnatal studies, where the offspring are raised to adulthood. Additional developmental landmarks can be assessed, including pinna detachment, hair growth, incisor eruption, eye opening, nipple development, and testes descent. The development of the following reflexes can also be evaluated: surface righting reflex, cliff avoidance, forelimb placing, negative geotaxis, pinna reflex, auditory startle reflex, hindlimb placing, air righting reflex, forelimb grip test, and pupillary constriction reflex. For a description of these developmental landmarks and reflexes, see Parker [34]. The examinations must begin prior to the landmark's historical day of onset and continue daily until each animal in the litter meets the criterion [67]. Data for each day's testing should be expressed as the number of pups that have achieved the criterion for each developmental landmark, divided by the total number of pups tested in the litter. Forelimb grip test and pupillary constriction tests are conducted on PND 21 only. Data for the forelimb grip test and pupillary constriction tests should be expressed as the number of pups that have achieved the criterion, divided by the total number of pups tested in the litter.

*Balanopreputial Separation* In rats, balanopreputial separation is considered to result from membrane cornification, which leads to the detachment of the prepuce from the glans penis in the rat. Preputial separation occurs dorsolaterally and then ventrally on the penis and down the shaft of the penis. The prepuce remains attached to the glans penis on its ventral surface by the frenulum of the prepuce of the penis. Male rats are examined for balanopreputial separation beginning on PND 35 to 40. Published preputial separation age ranges from PND 32 to 42 in Japanese studies [68–70] to PND 41 to 46 in American studies, depending on the observation criterion.

Because manipulation of the prepuce can accelerate the process of preputial separation, the males must be examined gently. Each male is removed from its cage and held in a supine position. Gentle digital pressure is applied to the sides of the rodent's prepuce, and the criterion is met when the prepuce completely retracts from the head of the penis. The foreskin can be attached along the shaft of the penis, but it cannot be attached to the opening of the urethra. Each male is examined daily until acquisition or until PND 55, whichever is earliest. Body weight should be recorded on the day the criterion is met.

*Vaginal Patency* After vagina canalization occurs in rats, the vaginal opening remains covered by a septum or membrane. The age of vaginal patency (when the septum is no longer evident) occurs around the time of puberty in rats, and it is their most readily determined marker for puberty. Female rats are examined beginning a few days prior to the expected age of maturation (e.g., PND 25 to 28) and examinations are continued until the criterion for patency has been achieved or until PND 43, whichever comes first.

The female is removed from the cage and held in a supine position, exposing the genital area. Gentle pressure is applied to the side of the vaginal opening to determine if it is patent. When the membrane is present, the area has a slight "puckered" appearance; however, when the membrane has broken, the vaginal opening is about

the size of a pinhead. The criterion has been met when the vaginal opening is observed. Body weight should be recorded on the day the criterion is met.

***$F_1$ Mating Procedure*** At approximately 90 days of age, the $F_1$ rats are cohabitated for a period of 21 days. A mating ratio of 1:1 from the same dosage group but from different litters allows identification of both parents of a litter. Females with sperm observed in the vaginal lavage and/or having a copulatory plug will be presumed pregnant and at day 0 of gestation and will be housed individually. Females that do not mate within 14 days will be cohabitated with an alternate male animal from the same dosage group that has mated.

### *$F_1$ Female Observations*

*$F_1$ Female Body Weight, Feed Consumption, Behavior, and Clinical Signs* Clinical signs and mortalities should be recorded at least once daily. Body weight values should be recorded on first day of dosing and at least weekly thereafter. Food consumption should be measured at least once weekly, at the same interval as the body weight collection. Water consumption should be measured weekly if the test article is administered in the drinking water.

### *$F_1$ Maternal Observations*

*Maternal Body Weight, Feed Consumption, Behavior, and Clinical Signs* Maternal body weight, feed consumption, behavior, and clinical signs should be recorded at the same intervals as for the $F_0$ maternal observations.

*Laparohysterectomy* Female $F_1$ rats will be laparohysterectomized on GD 21. The dams should be examined for numbers and distribution of corpora lutea, implantation sites, placentas, live and dead fetuses, and early and late resorptions. The location of each fetus should be recorded. Each fetus will be examined for gender and gross external alterations (representative photographs of alterations may be taken). Fetuses will be tagged with individual identification number, litter number, uterine distribution, and fixative. Approximately one-half of the fetuses in each litter should be retained in Bouin's solution; the remaining fetuses should be retained in alcohol. If a Segment II study has already been performed, then the fetuses do not require further evaluation but should be retained for possible future evaluation. If a Segment II or other supplementary study has not been performed, then evaluation of the viscera and skeleton should be performed as described earlier (Section 4.3).

### *$F_1$ Paternal Observations*

*Paternal Body Weight, Feed Consumption, Behavior, and Clinical Signs* Clinical signs, mortalities, and body weight values should be recorded at the same intervals and in the same manner as for the $F_1$ females.

### *Assessment of the $F_1$ Males at Necropsy*

*Unscheduled Deaths* Any males that are found dead or appear moribund are subjected to a complete necropsy, and the viscera are examined to investigate the cause of death.

*Scheduled Deaths* Males are routinely euthanized and necropsied after the outcome of mating is known. In the event of an equivocal mating result, males can be mated with untreated females to ascertain their fertility or infertility. The males treated as part of the study may also be used for evaluation of toxicity to the male reproductive system if dosing is continued beyond mating and sacrifice is delayed. Routinely, after day 21 of cohabitation, the male is weighed, euthanized, and subjected to a complete necropsy.

*Inspection of $F_1$ Male Viscera* The thoracic, abdominal, and pelvic organs should be inspected grossly for evidence of toxicity. If gross organ lesions are noted, the affected organs should be preserved for possible histopathological examination. The testes and epididymides should be excised and their weights recorded (total weights paired and individual organ weights). The epididymides should be stored in neutral buffered 10% formalin. The testes should be fixed in Bouin's solution for 48–96 hours and then retained in 10% neutral buffered formalin.

### 5.9 Analysis of Maternal and Litter/Pup Data

***Interpretation of the Data*** Values from control and test groups of animals should be compared statistically. The following techniques may be used, but others may be substituted if they are appropriate. The fertility and gestation indices may be analyzed by a one-tailed Fisher exact test. For the sex ratio index, a two-tailed Fisher exact test may be used. Data for the viability and weaning indices may be transformed by the Freeman–Tukey arcsine transformation for binomial proportions. The transformed data may then be analyzed by an analysis of variance (ANOVA) followed by a one-tailed protected least significant difference (LSD) test to compare the control with the treated groups if the ANOVA yields $p < 0.10$. The average litter size and the number of viable pups throughout the reproduction phase may be analyzed by ANOVA followed by a protected LSD test (one-tailed). For the growth (weight gain) and organ weight analyses, an analysis of covariance may be used, followed by a protected LSD test (two-tailed) to compare the control and treated groups.

***Importance of Historical Control Data*** Historical control data should be used to enhance interpretation of study results. Historical control data, when used, should be compiled, presented, and evaluated in an appropriate and relevant manner. In order to justify its use as an analytical tool, information such as the dates of study conduct, the strain and source of animals, and the vehicle and route of administration should be included.

## 6 TWO-GENERATION REPRODUCTIVE TOXICITY STUDIES

### 6.1 Rationale for the Test

The multigeneration study (two-generation reproduction testing guideline, OPPTS 870.3800 [71] and OECD 416 [72]) was designed to detect adverse effects on the integrity and performance of the male and female reproductive systems, including

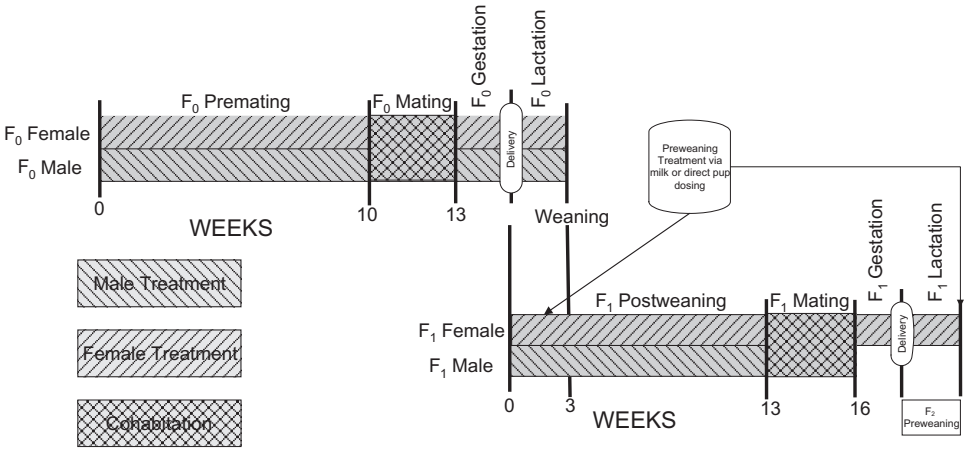


**FIGURE 5** Multigeneration study.

gonadal function, estrous cycle, mating behavior, conception, gestation, parturition, lactation, and weaning, and on the growth and development of the offspring (Fig. 5). The adverse effects assessed in a multigeneration study also include neonatal mortality, mortality, target organs in offspring, and functional deficits in offspring, including behavior, maturation (puberty), and reproduction.

### 6.2 Principle of the Test Method

The test article is administered to the parental animals ($F_0$) prior to and during mating, during gestation, parturition, and lactation, and through weaning of the $F_1$ offspring. The test article is administered to selected $F_1$ offspring from weaning through adulthood, during cohabitation and mating, and through gestation, parturition, and lactation until the $F_2$ generation is weaned (PND 21).

### 6.3 Animal Species and Strain Selection and Rationale

The rat is the commonly used species for multigeneration studies. It is generally desirable to use the same species and strain as in other toxicological studies. If another mammalian species is used, then justification for its use should be provided in the protocol. Within and between studies, animals should be of comparable age, weight, and parity at the start of the study. In OECD 416, the weight variation at the commencement of study should be minimal and not exceed 20% of the mean weight of each sex.

### 6.4 Procurement and Acclimation of the Test Animals

The test animals should be young (4–8 weeks of age) upon arrival. The test animals should be from the same source and strain, and the females should be nulliparous, nonpregnant, and approximately the same age and weight. The test animals should be acclimated to the laboratory conditions at least 5 days prior to study start. Each

animal should be assigned a unique identification number. This is done just prior to initiation of dosing for the $F_0$ generation and just after weaning for the $F_1$ generation.

### 6.5 Number of Test Animals

The number of animals per group should be sufficient to yield not less than 20 pregnancies. Routinely, 25 male and 25 female rats per group are used. Each test group should contain a similar number of mating pairs.

### 6.6 Selection of Dosage Groups, Controls, and Route of Administration

Most multigeneration studies are designed with a vehicle control and at least three dosage groups, all of which should be concurrent and housed and handled identically. Males and females should be assigned by a randomization procedure that assures that mean body weights are comparable among all test groups. The test compound and vehicle should be administered by the route that most closely approximates the pattern of expected human exposure. If the route of administration differs from the expected human exposure, a justification will be required in the protocol.

### 6.7 Selection of Dosage Levels

The dosage levels should be spaced to produce a gradation of toxic effects. Unless limited by physiochemical or biological properties, the highest dose should be chosen to produce reproductive and/or systemic toxicity. For the $F_0$ generation, the highest dose selected should not have caused greater than 10% mortality in previous toxicity studies. The intermediate dose should only produce minimal observable toxic effects. The low dose should not produce any evidence of either systemic or reproductive toxicity (NOAEL). Two- or fourfold intervals are often used for spacing of the dosage levels. For dietary studies, the dosage interval should not exceed threefold (OECD 416). Metabolism and pharmacokinetics of the test article should be available from previously conducted studies to demonstrate the adequacy of the dosing regimen.

A concurrent control group (either untreated, sham treated, or vehicle treated) should always be used. If a vehicle control group is used, then the dosing volume should be equal to the highest volume used. If the test article is administered in the diet and causes reduced dietary intake, then the use of a pair-fed control group may be necessary.

### 6.8 Treatment Timing and Test Duration

The $F_0$ animals are dosed daily beginning at 5–9 weeks of age. Daily dosing of the selected $F_1$ generation offspring commences at weaning. For both generations ($F_0$ and $F_1$) dosing should be continued for at least 10 weeks prior to mating. Daily dosing of the $F_0$ and $F_1$ animals should continue until termination.

### 6.9 Mating Procedures, Detection of Mating, and Housing of Mated Females

A mating ratio of 1:1 from the same dosage group allows identification of both parents of a litter. The mating pair is cohoused until evidence of copulation is observed or either three estrous periods or 2 weeks have elapsed. Animals should be separated as soon as evidence of copulation is observed (as described earlier). Once they are pregnant, females should be housed individually to allow for monitoring of feed and, if necessary, water consumption. Feed and water should be provided *ad libitum*. Pregnant animals should be provided with nesting materials near the end of gestation.

### 6.10 $F_0$ Generation

#### *$F_0$ Maternal Observations*

*Maternal Body Weight, Feed Consumption, Behavior, and Clinical Signs* Clinical signs and mortalities should be recorded at least once daily throughout the test period. Body weight values should be recorded on the first day of dosing and at least weekly, but during gestation and lactation body weights are routinely collected at protocol-specified intervals (e.g., GD 0, 7, 14, and 21, and LD 0, 4, 7, 14, and 21). Food consumption should be measured at least once weekly and at the same interval as the body weight collection through LD 14. Water consumption should be measured weekly if the test article is administered in the drinking water. Estrous cycle length and pattern should be determined from vaginal smears (see earlier discussion) during a minimum of 3 weeks prior to mating and throughout cohabitation.

*$F_0$ Parturition* Dams should be housed in litter boxes with bedding material no later than DG 20 (2 days prior to expected parturition, earlier if required by protocol). The dams are observed periodically for signs of parturition onset. Individual pup observations should not be made until delivery is completed. Dead pups should be removed from the nesting box to preclude cannibalization by the dam.

#### *Assessment of the $F_0$ Dam at Necropsy*

*Unscheduled Deaths* Any mated females that appear moribund should be euthanized appropriately. Any mated females that are found dead should be necropsied to investigate the cause of death. Any evidence of pregnancy and of embryonic development should be evaluated. A vaginal smear should be taken and the stage of estrus determined.

*Scheduled Deaths* All $F_0$ females Should be terminated when they are no longer needed for assessment of reproductive effects. Dams are routinely necropsied on LD 21. A vaginal smear should be taken and the stage of estrus determined. The dam is weighed and euthanized, her ovaries and uterus are examined, and she is subjected to a complete necropsy.

#### *Inspection of $F_0$ Maternal Viscera*

*Unscheduled Deaths* If necropsy occurs prior to parturition, then the gravid uterus should be removed and weighed. The uterus should be examined to determine if

the animal was pregnant, and the relative numbers of live and dead conceptuses and their relative positions in the uterus should be recorded. If the uterus shows no signs of implantation, it should be treated with a reagent, such as 10% ammonium sulfide, to reveal any early resorptions. The number of corpora lutea should be recorded for each ovary, as described earlier. The remaining maternal abdominal and thoracic organs should be inspected grossly for evidence of toxicity. The following organs should be weighed from all $F_0$ females: uterus with oviducts, ovaries, brain, pituitary, liver, kidneys, adrenal glands, spleen, and any known target organs. The vagina, uterus with oviducts, cervix, ovaries, pituitary, and adrenal glands should be removed, fixed, and stored in appropriate medium for histopathological examination. If gross organ lesions are noted, the affected organs should be preserved for possible histopathological examination.

*Scheduled Deaths* If necropsy occurs as scheduled, then the uterus should be removed and weighed. The number of corpora lutea should be recorded for each ovary as described earlier. The remaining maternal abdominal and thoracic organs should be inspected grossly for evidence of toxicity. The organs for organ weight collection and possible histopathological examination are listed below. If gross organ lesions are noted, the affected organs should be preserved for possible histopathological examination.

*Histopathology for $F_0$ Females* The vagina, uterus with oviducts, cervix, ovaries, pituitary and adrenal glands, and any target organs and gross lesions should be fixed and stored in a suitable medium for histopathological examination. Complete histological evaluation is performed on the listed organs from ten randomly selected control and high dose $F_0$ females. Organs demonstrating treatment-related changes should be examined from the remainder of the high dose and control animals, and for all $F_0$ females in the low and mid-dose groups. Additionally, reproductive organs of the low and mid-dose females suspected of reduced fertility (those that failed to mate, conceive, or deliver healthy offspring, or for which estrous cyclicity was affected) should be subjected to histopathological evaluation as described previously (Segment I). Histopathological examination should detect qualitative depletion of the primordial follicle population; however, reproductive study guidelines require a quantitative evaluation of primordial follicles of the $F_0$ females (see Segment I).

### *$F_0$ Paternal Observations*

*Paternal Body Weight, Feed Consumption, Behavior, and Clinical Signs* Clinical signs and mortalities should be recorded at least once daily. Body weight values should be recorded on first day of dosing and at least weekly thereafter. Food consumption should be measured at least once weekly at the same interval as the body weight collection. Water consumption should be measured weekly if the test article is administered in the drinking water.

### *Assessment of the $F_0$ Males at Necropsy*

*Unscheduled Deaths* Any males that are found dead or appear moribund are subjected to a complete necropsy, and their visceral contents are examined to investigate the cause of death.

*Scheduled Deaths* Males are routinely euthanized and necropsied after the outcome of mating is known. In the event of an equivocal mating result, males can be mated with untreated females to ascertain their fertility or infertility. The males treated as part of the study may also be used for evaluation of toxicity to the male reproductive system if dosing is continued beyond mating and sacrifice is delayed. The male is weighed, euthanized, and subjected to a complete necropsy.

*Inspection of $F_0$ Male Viscera* The testes, epididymides (paired and individually), seminal vesicles (with coagulating glands and their fluids), prostate, brain, pituitary, liver, kidneys, adrenal glands, spleen, and any known target organs are removed and weighed. For all $F_0$ males, sperm from one testis and one epididymis should be collected for enumeration of homogenization-resistant sperm and caudal epididymal sperm reserves, respectively. In addition, sperm from one cauda epididymis (or vas deferens) should be collected for evaluation of sperm motility and morphology. The remaining abdominal and thoracic organs should be inspected grossly for evidence of toxicity. If gross organ lesions are noted, the affected organs should be preserved for possible histopathological examination.

*Histopathology for $F_0$ Males* One testis (preserved in Bouin's fixative or a comparable preservative), one epididymis, seminal vesicles, prostate, coagulating, pituitary, and adrenal glands, and any target organs and gross lesions should be fixed and stored in a suitable medium for histopathological examination. Complete histological evaluation is performed on the listed organs from ten randomly selected control and high dose $F_0$ males. Organs demonstrating treatment-related changes should be examined from the remainder of the high dose and control animals, and from all $F_0$ males in the low and mid-dose groups. Additionally, reproductive organs of the low and mid-dose males suspected of reduced fertility (those that failed to mate or impregnate or had abnormal sperm number, motility, or morphology) should be subjected to histopathological evaluation as described earlier (Segment I).

*$F_0$ Sperm Evaluation* The sperm quality parameters are sperm number, sperm motility, and sperm morphology. Sperm evaluation will be performed on all control and high dose group $F_0$ males unless treatment-related effects have been observed; in that case, the lower dosage groups should be evaluated. The sperm evaluation methodologies were described earlier (Section 3.10). A morphological evaluation of epididymal (or vas deferens) sperm sample should be performed. If sperm evaluation parameters have already been evaluated as part of a systemic toxicity study of at least 90 days' duration, the evaluations need not be repeated in this study.

### 6.11 $F_1$ Generation

***$F_1$ Litter Evaluations*** The general appearance of the entire $F_1$ litter and maternal and pup nesting behavior is observed as described earlier for Segment III.

***$F_1$ Pup Evaluations*** The litter is evaluated as soon as possible after delivery to establish the number and sex of pups, stillbirths, live births, and the presence of gross anomalies. Pups that are found dead should be examined for possible defects and

cause of death. Pups that have been cannibalized to the extent that their gender cannot be determined are recorded as "sex unable to be determined due to degree of cannibalization" (see earlier discussion).

Live pups should be counted, sexed, and weighed individually at birth, and on LD 4, 7, 14, and 21. If a pup cannot be found during the daily litter count, the pups are sorted by gender in order to determine the sex of the missing pup(s) by process of elimination. The missing pup is recorded as "missing, presumed cannibalized." If the pups have been tattooed or otherwise identified, the actual identification number of the missing pup(s) should be recorded.

The pup survival indices (e.g., Live Birth Index, Survival Index, and Sex Ratio) should be calculated as described previously (Section 5.8).

***Standardization of $F_1$ Litter Sizes*** Animals should be allowed to litter normally and rear their offspring to weaning. Standardization of litter sizes at LD 4 is optional (see Section 5.8).

***Weaning*** Weaning is usually performed on LD 21. At weaning, one male and one female offspring per litter are selected for rearing to adulthood and mating to assess reproductive competence.

***Test Article Administration*** The test article is administered to selected $F_1$ offspring from weaning through adulthood, during cohabitation and mating, and through gestation, parturition, and lactation until the $F_2$ generation is weaned.

***$F_1$ Postweaning Developmental Landmarks*** Sexual maturation landmarks (balanopreputial separation and vaginal patency) are required in multigeneration studies where the selected offspring are raised to adulthood. The OECD 416 guideline recommends that additional developmental parameters be evaluated for supplementary information (e.g., eye and ear opening, tooth eruption, hair growth). Other functional investigations are also recommended before and/or after weaning, such as motor activity, sensory function, and reflex ontogeny. These observations should not be performed on the pups selected for mating. These landmarks are described in Parker [34].

***$F_1$ Mating Procedure*** A mating ratio of 1 : 1 from the same dosage group but from different litters allows identification of both parents of a litter. The mating pair is cohoused as described above. Pregnant animals should be provided with nesting materials near the end of gestation. In instances of poor reproductive performance in the controls or in the event of a treatment-related alteration in the litter size, the adults of that generation may be remated to produce an $F_{1b}$ litter. The dams should not be remated until approximately 1–2 weeks following weaning of the $F_{1a}$ litter.

***$F_1$ Maternal Observations***

*Maternal Body Weight, Feed Consumption, Behavior, and Clinical Signs* Maternal body weight, feed consumption, behavior, and clinical signs should be recorded at the same intervals as the $F_0$ maternal observations. Estrous cycle length and pattern

should be determined from vaginal smears during a minimum of 3 weeks prior to mating and throughout cohabitation. Observations that have proved of value in other toxicity studies, as well as duration of pregnancy, parturition, and nursing behaviors, are recorded.

*$F_1$ Parturition* Dams should be housed in litter boxes with bedding material no later than DG 20 (2 days prior to expected parturition, earlier if required by protocol). The dams are observed periodically throughout the day for signs of parturition onset. As soon as the first pup is found, parturition is considered initiated and should be observed as described earlier (Section 5.7).

***Assessment of $F_1$ Weanlings Not Selected for Continuation*** At the time of termination or death during the study, when the litter size permits, at least three pups per sex per litter should be examined macroscopically for structural abnormalities or gross lesions, paying special attention to the reproductive organs. Dead pups or pups that are terminated in a moribund condition should be examined for possible defects and cause of death. The following organs should be weighed from one randomly selected pup per sex per litter: brain, spleen, and thymus. The following organs should be fixed and stored for histopathological examination: gross lesions and any target organs.

***Assessment of the Parental $F_1$ Female at Necropsy*** Same as assessment of the parental $f_0$ female at necropsy, as described in Section 5.7.

***$F_1$ Male Observations*** Same as $F_0$ male observations, as described in Section 5.8.

### 6.12 $F_2$ Generation

***$F_2$ Litter and Pup Evaluations*** The general appearance of the entire $F_2$ litter and maternal and pup nesting behavior should be observed as described earlier for the $F_1$ pups. The litter is evaluated as soon as possible after delivery to establish the number and sex of pups, stillbirths, live births, and the presence of gross anomalies as described earlier for the $F_1$ pups. Live $F_2$ pups should be counted, sexed, and weighed individually at birth, and on LD 4, 7, 14, and 21 as described for $F_1$ pups. Pup survival indices should be calculated and reported as described earlier (Section 5.8).

***Assessment of $F_2$ Pups at Weaning*** At weaning or death during the study, when the litter size permits, at least three pups per sex per litter should be examined macroscopically for any structural abnormalities or pathological changes, paying special attention to the reproductive organs. Dead pups or pups that are terminated in a moribund condition should be examined for possible defects and/or cause of death. The following organs should be weighed from one randomly selected pup per sex per litter: brain, spleen, and thymus. The following organs should be fixed and stored for histopathological examination: any gross lesions and target organs.

### 6.13 Analysis of Maternal and Litter/Pup Data

Values from control and test groups of animals should be compared statistically as described in Section 5.9. Historical control data should be used to enhance interpretation of study results, as described in Section 5.9.

## 7 JUVENILE TOXICITY TESTING

### 7.1 Rationale for the Test

Drug development programs have used safety data from preclinical and clinical studies in adults to support the use of a drug in pediatric patients. This method assumes that the pediatric population has a similar disease progression and responds to the therapy in a similar fashion as the adult. Unfortunately, adult clinical trials and supportive preclinical animal studies may not always assess drug effects on developmental processes that are pediatric age-group specific. Pharmacological and pharmacodynamic processes in the pediatric population may differ from those of the adult population, resulting in adverse effects that may not be detected in adult preclinical or clinical studies. Juvenile animal studies can be designed to characterize postnatal developmental toxicities that would not be detected by use of the routine perinatal/postnatal toxicity study design. See Brent [73] for a review of the utilization of juvenile animal studies to determine human effects.

In 1997, the FDA [75] stated that the lack of pediatric labeling information posed a significant risk for children. The lack of pediatric safety information in product labeling exposes pediatric patients to the risk of age-specific adverse reactions unexpected from adult experience. These reactions could be avoided by development of appropriate pediatric dosage information. In addition to adverse effects due to overexposure, the absence of pediatric testing and labeling may also expose pediatric patients to ineffective treatment through underdosing, or may deny pediatric patients therapeutic advances because physicians choose to prescribe existing, less effective medications in the face of insufficient pediatric information about a new medication. Failure to develop a pediatric formulation of a drug or biological product, where younger pediatric populations cannot take the adult formulation, may also deny pediatric patients access to important new therapies, or may require pediatric patients to take the drug in formulations that may be poorly or inconsistently bioavailable. Also in 1997, the ICH M3 guidance, *Nonclinical Safety Studies for the Conduct of Human Clinical Trails for Pharmaceuticals* [74], states that juvenile animal studies should be considered on an individual basis when previous animal data and human safety data are insufficient.

In February 2006, the U.S. FDA Center for Drug Evaluation and Research (CDER) issued a *Guidance for Industry: Nonclinical Safety Evaluation of Pediatric Drug Products* [75]. It discusses the conditions when a juvenile animal study should be conducted, describes the design of juvenile animal studies, and makes recommendations on nonclinical testing (e.g., the timing and utility of juvenile animal studies in relation to phases of clinical development). This guidance is available on the FDA website (http://www.fda.gov/cder/guidance/5671fnl.htm).

### 7.2 Study Justification

Before performing a juvenile animal toxicity study, several questions should be addressed. The first question to address is the justification of the juvenile animal study. To answer this question, several additional questions must be answered.

- What is the age of the targeted pediatric population?

  Consideration should be given to the age of the intended population and its stage of postnatal development. The endpoints assessed in the nonclinical juvenile studies should be tailored to address concerns for a particular pediatric population. In addition, the disease indication may influence the type, duration, and timing of the testing.
- Is the target organ or system developing during the juvenile period in the species of choice?

  The toxic effects of drugs on postnatal development are believed most likely to occur in those organs and tissues that are undergoing significant postnatal development. The nervous, reproductive, pulmonary, renal, skeletal, and immune systems undergo considerable postnatal growth and development. Exposure to the drug must occur during the appropriate periods of growth and development in the animal species. A comparison of the times of target organ or system development between the species of choice and humans should be reviewed, usually by literature mining. Anatomical development as well as physiological development must be considered. For example, the timing of the presence of functional metabolic enzymes varies among species, as well as during the developmental period for each species. The FDA guidance has tables that give basic information; however, a series of articles on comparative organ system development are also available that provides more in-depth information (many of these are collected in Hood [38, Appendix C]): male and female reproductive system [76, 77], lung [78, 79], renal [80, 81], heart [82], skeletal system [83], behavioral aspects of the central nervous system [84], immune system [85, 86], brain [87], and gastrointestinal system [88].
- Is there enough preexisting data from juvenile animal studies with a compound that is structurally similar (e.g., same class of chemicals) or pharmacologically similar (same mechanism of action) that this study may not be required?

  Many potential pediatric drugs have established adult efficacy and safety profiles and there may be limited data available from pediatric patients 12 years or older. For marketed drugs that have already undergone extensive adult clinical trials, nonclinical pharmacology and toxicology data are available, including studies of general, reproductive, and genetic toxicity, as well as carcinogenicity and other special toxicology studies. Studies in juvenile animals such as perinatal/postnatal studies may be available. Target organs of toxicity in animals and adult humans have already been identified in these studies. A careful evaluation of these data should enable the toxicologist to (1) identify potential safety concerns for the intended pediatric population, (2) judge the adequacy of the nonclinical studies, and (3) identify gaps in the data that could be addressed by juvenile animal testing. Based on this assessment, a juvenile animal study might not be warranted.

- Are the metabolic profiles similar during the juvenile period between the animal species and humans?

  Metabolic profiles should be considered in determining the selection of an appropriate species. The adult human metabolic pathways are not necessarily the same as the pediatric metabolic pathways, and the pediatric metabolic pathways may differ from juvenile animal metabolic pathways. The presence or absence of functional metabolic enzymes in the juvenile animal species may alter the pharmacokinetics and toxicology of the therapeutic agent. By altering drug metabolism, increased drug accumulation or elimination can occur that affects exposure and duration. In addition, new reactive metabolites can be created or fail to be metabolized.
- Is the pediatric formulation different from the adult formulation?

  If, after pharmacokinetic bridging studies, it is determined that the pediatric formulation is pharmacokinetically (e.g., altered absorption, clearance, $C_{max}$, $T_{max}$, AUC) sufficiently different from the adult formulation, a juvenile animal study should be performed using the pediatric formulation.
- If the juvenile animal study is justified, what are the key experimental design issues—for example, species, age at start and end of treatment, route of administration (maternal and fetal/neonatal through adulthood), duration of exposure, endpoints, and general or specific histopathology?

  When determining the appropriate species, the following factors should be considered: (1) pharmacology, pharmacokinetics, and toxicology of the compound; (2) comparative developmental stage of major target organs; and (3) sensitivity of the selected species to a particular toxicity.

### 7.3 Species of Choice

The rat is generally the species of choice, even though the rat may not be most sensitive or the most relevant species. This preference is based on the fact that appropriate developmental windows in the rat cover all postnatal developing target organs and because of the large rat postnatal database. There will be cases where the rat is not an appropriate species (e.g., when the target organ toxicity is not observed in the rat but is observed in the second species) and another species should be considered. The usual second species of choice is the dog (although nonhuman primates, pigs, and minipigs may be appropriate when drug metabolism in the rat differs significantly from that of the human).

### 7.4 Age at Start and End of Treatment

Age at start and end of treatment may vary depending on the target pediatric population and the timing of the corresponding developmental stage in the animal model. The specific age ranges for pediatric populations are presented in the ICH E11 guidline, *Clinical Investigation of Medicinal Products in the Pediatric Population* [89].

### 7.5 Route of Administration

The route of administration is usually the route in human adults. However, pharmacologic/pharmacodynamic differences in the juvenile animal model may require

an alternate route. If multiple routes of administration are planned for the different pediatric populations and/or indications, then multiple routes of exposure should also be tested. Assessment of differential drug pharmacokinetics should be made prior to the study according to established guidelines (ICH S3A guideline, *Note for Guidance on Toxicokinetics: Assessment of Systemic Exposure in Toxicity Studies* [90].

### 7.6 Frequency of Administration and Duration of Treatment

The frequency of administration should be relevant to the intended use of the drug. However, sometimes this is not feasible because of technical considerations or metabolic or pharmacokinetic differences. The duration of treatment should include the relevant period of postnatal development for the selected species and may be expanded to cover a longer duration, based on chronic pediatric administration. The assessment of delayed toxicity may require continuation of exposure through organ maturity or until adulthood, based on the safety concern. Addition of a treatment-free period should also be considered to assess potential reversibility of adverse effects, thereby distinguishing short-term pharmacodynamic effects from frank organ developmental toxicity.

### 7.7 Range-Finding Studies

Range-finding studies (including toxicokinetic/pharmacokinetic parameters) are highly recommended for establishment of a maximum tolerated dose, tolerability, target organs, and subsequent dose level selection. These studies can also be used to address specific questions and develop methodologies. For the definitive juvenile animal study, the highest dosage should cause toxicity (developmental or general), while the lowest dosage should produce little to no toxicity and should establish a NOAEL.

### 7.8 Juvenile Animal Study Design

There are two options for juvenile animal study design. The study can be (1) designed to address a specific safety concern (i.e., known or anticipated effects based on adult studies or drug class effects), or (2) a modified perinatal/postnatal developmental toxicity study designed to evaluate most potential outcomes. The target organ toxicity directed specific study design can be developed using a modified repeat dose toxicity study design. The design is modified for species considerations, age of the animal at initiation of exposure, duration of treatment, and endpoints evaluated. Routinely, both sexes are evaluated. When the magnitude of the biological effect is small, greater numbers of animals per group may be required.

The second design is a modified Segment III study (as described earlier) that assures adequate exposure during the target organ(s) development and assessment of specific organ requirements (e.g., additional developmental landmarks or reflex ontogeny). The routine Segment III design consists of maternal dosing from GD 6 through PND 20, with weaning on PND 21. If the drug or active metabolite is not transferred through the milk, then direct dosing of the pup throughout the prewean-

ing phase may be necessary. Selected pups are raised to adulthood and mated after 80 days of age, and their offspring are evaluated. Developmental and behavioral assessments (e.g., reflex ontogeny, vaginal patency, preputial separation, reactivity, sensorimotor function, motor activity, and learning and memory assessments) are performed. Additional developmental and behavioral landmarks may need to be assessed, based on target organs. Additional measurements of overall growth are required (growth velocity over time and tibial length). Additional endpoints should be considered on a case-by-case basis. Juvenile animal studies are conducted on a case-by-case basis, and the study design should be discussed with the regulatory agency.

### 7.9 Timing of the Juvenile Animal Study

If a juvenile animal study is justified, when will the study be conducted relative to clinical trial? The timing of the juvenile animal study used to support the pediatric clinical trials depends on the duration of the juvenile animal study and the subsequent timing of the pediatric clinical trial. A juvenile animal study using the Segment III design will require approximately 5 months for in-life data collection and up to six additional months for data analysis and reporting. Therefore, the juvenile animal study should be initiated at least 1 year prior to the pediatric clinical study. If toxicities identified in the juvenile animal studies are likely to occur in pediatric patients, cannot be clinically monitored, and do not have adequate safety margins, then, after risk–benefit analysis for the therapeutic indication, it may not be possible to safely conduct pediatric clinical trials.

## 8 CONCLUSION

The design and conduct of developmental and reproductive toxicity (DART) studies and juvenile animal studies for pediatric drugs are among the most difficult in toxicology. These studies are run under FDA/ICH Good Laboratory Practice and Good Automated Laboratory Practice regulations, which also require computer database validation and validation of behavioral testing paradigms. These are monumental tasks requiring significant expenditures of personnel and money. Behavioral testing requires that the apparatus and data collection system has been validated and the conduct of positive control studies using drugs that achieve both ceiling and floor effects. Apparatuses or methods used in adult animal behavioral assessments may not be sensitive enough or small enough to test pups. DART studies require highly trained and skilled technicians to properly evaluate the reproductive and developmental aspects, and now many such studies require behavioral evaluations calling for behaviorally trained technicians and scientists. Interobserver variations must be minimized, which requires significantly more training and evaluation than routine clinical observations. In order to evaluate the data and report the results, trained reproductive and behavioral toxicologists are needed. Furthermore, while Segment I, II, and III protocols do not routinely need review by regulatory agencies prior to conduct, it is recommended that regulatory review be carried out prior to commencing juvenile animal studies to support pediatric clinical trials.

## REFERENCES


1. US Food and Drug Administration. International Conference On Harmonisation; Guideline on Detection of Toxicity to Reproduction for Medicinal Products. *Fed Reg Part IX* 1994;59(183):48746–48752.
2. Hurtt ME, Daston G, Davis-Bruno K, Feuston M, Silva Lima B, Makris S, McNerney ME, Sandler JD, Whitby K, Wier P, Cappon GD. Workshop summary: juvenile animal studies: testing strategies and design. *Birth Defects Res (Part B)* 2004;71:281–288.
3. US Food Drug Administration. *Guideline for Industry. Nonclinical Safety Evaluation of Pediatric Drug Products*. CDER; 2006. http://www.fda.gov/cder/guidance/5671fnl.pdf.
4. US Food and Drug Administration. *Guidance for Industry. Safety Studies for Veterinary Drug Residues in Human Food: Reproduction Studies*. VICH GL22. Rockville, MD: US Department of Health and Human Services; 2000.
5. Slikker W, Chang LW, Eds. *Handbook of Developmental Neurotoxicology*. San Diego: Academic Press; 1998.
6. Hood RD, Ed. *Developmental and Reproductive Toxicology, A Practical Approach*. Boca Raton, FL: CRC Press; 2006.
7. US Food Drug Administration. Guideline for industry, detection of toxicity to reproduction for medicinal products. *Fed Reg* 1994;59(183). (Same as ICH Guideline for Industry S5A.)
8. US Food and Drug Administration. *Guidance for Industry. Considerations for Reproductive toxicity Studies for Preventive Vaccines for Infectious Disease Indications*. Rockville, MD: US Department of Health and Human Services; 2000.
9. International Conference on Harmonisation. ICH harmonized tripartite guideline: Maintenance of the ICH guideline on toxicity to male fertility, An addendum to the ICH tripartite guideline on detection of toxicity to reproduction for medicinal products, S5B(M). (Amended 9 November, 2000).
10. Davis DL, Friedler G, Mattison D, Morris R. Male-mediated teratogenesis and other reproductive effects: biologic and epidemiologic findings and a plea for clinical research. *Reprod Toxicol* 1992;6:289–292.
11. Colie CF. Male mediated teratogenesis. *Reprod Toxicol* 1993;7:3–9.
12. Savitz DA, Sonnenfeld NL, Olshan AF. Review of epidemiologic studies of paternal occupational exposure and spontaneous abortion. *Am J Ind Med* 1994;25:361–383.
13. Trasler JM, Doerksen T. Teratogen update: paternal exposures—reproductive risks. *Teratology* 1999;60:161–172.
14. Ulbrich B, Palmer AK. Detection of effects on male reproduction—a literature survey. *J Am Coll Toxicol* 1995;14:293–327.
15. Takayama S, Akaike M, Kawashima K, Takahashi M, Kurokawa Y. Collaborative study in Japan on optimal treatment period and parameters for detection of male fertility disorders in rats induced by medicinal drugs. *J Am Coll Toxicol* 1995;14:66–292.
16. Sakai T, Takahashi M, Mitsumori K, Yasuhara K, Kawashima K, Mayahare H, Ohno Y. Collaborative work to evaluate toxicity on male reproductive organs by 2-week repeated dose toxicity studies in rats. Overview of the studies. *J Toxicol Sci* 2000;25:1–21.
17. Adler NT, Toner JP. The effect of copulatory behavior on sperm transport and fertility in rats. In Komisaruk BR, Siegel HI, Chang MF, Feder HH, Eds. *Reproduction: Behavioral and Neuroendocrine Perspective*. New York: New York Academy of Science; 1986, pp 21–32.
18. Salewski E. Färbemethode zum makroskopischen nachweis von implantationsstellen am uterus der ratte. *Naunyn-Schmiedebergs Arch Exp Pathol Pharmakol* 1964;247:367.

19. Long JA, Evans HM. The oestrous cycle in the rat and its associated phenomena. *Mem Univ Calif* 1922;6:1–111.
20. Cooper RL, Goldman JM, Vandenbergh JG. Monitoring of the estrous cycle in the laboratory rodent by vaginal lavage. In Heindel JJ, Chapin RE, Eds. *Methods in Toxicology: Female Reproductive Toxicology*. San Diego: Academic Press; 1993, pp 45–56.
21. STP Working Group. STP Position Paper: ovarian follicular counting in the assessment of rodent reproductive toxicity. *Toxicol Pathol* 2005;33:409–412.
22. Plowchalk DR, Smith BJ, Mattison DR. Assessment of toxicity to the ovary using follicle quantitation and morphometrics. In Heindel JJ, Chapin RE, Eds. *Methods in Toxicology: Female Reproductive Toxicology*. San Diego: Academic Press; 1993, pp 57–68.
23. Pedersen T, Peters HJ. Proposal for the classification of oocytes and follicles in the mouse ovary. *Reprod Fertil* 1968;17:555–557.
24. Bolon B, Bucci TJ, Warbritton AR, Chen JJ, Mattison DR, Heindel JJ. Differential follicle counts as a screen for chemically induced ovarian toxicity in mice: results from continuous breeding bioassays. *Fundam Appl Toxicol* 1997;39:1–10.
25. Bucci TJ, Bolon B, Warbritton AR, Chen JJ, Heindel JJ. Influence of sampling on the reproducibility of ovarian follicle counts in mouse toxicity studies. *Reprod Toxicol* 1997;11:689–696.
26. Heindel JJ, Thomford PJ, Mattison DR. Histological assessment of ovarian follicle number in mice as a screen of ovarian toxicity. In Hirshfield AN, Ed. *Growth Factors and the Ovary*. New York: Plenum; 1989, pp 421–426.
27. Christian MS, Brown WR. Control primordial follicle counts in multigeneration studies in Sprague–Dawley ("gold standard") rats. *Toxicologist* 2002;66(No. 1-S).
28. Russell LD, Ettlin R, Sinha Hikim AP, Clegg ED. The classification and timing of spermatogenesis. In Russell LD, Ettlin R, Sinha Hikim AP, Clegg ED, Eds. *Histological and Histopathological Evaluation of the Testis*. Clearwater, FL: Cache River Press; 1990, Chap 2.
29. Russell LD, Ettlin R, Sinha-Hikim AP, Clegg ED, Eds. *Histological and Histopathological Evaluation of the Testis*. Clearwater, FL: Cache River Press; 1990.
30. Chapin RE, Filler RS, Gulati D, Heindel JJ, Katz DF, Mebus CA, Obasaju F, Perreault SD, Russell SR, Schrader S, Slott V, Sokol RZ, Toth G. Methods for assessing rat sperm motility. *Reprod Toxicol* 1992;6:267–273.
31. Toth GP, Stober JA, Zenick H, Read EJ, Christ SA, Smith MK. Correlation of sperm motion parameters with fertility in rats treated subchronically with epichlorohydrin. *J Androl* 1991;12:54–61.
32. Toth GP, Read EJ, Smith MK. Utilizing Cryo Resources CellSoft computer-assisted sperm analysis system for rat sperm motility studies. In Chapin RE, Heindel JJ, Eds. *Methods in Toxicology: Male Reproductive Toxicology*. San Diego: Academic Press; 1993, pp 303–318.
33. UK Industrial Reproductive Toxicology Discussion Group (IRDG). Computer-Assisted Sperm Analysis (CASA) Group, rat sperm morphological assessment, guideline document, edition 1, October 2000. http://www.irdg.co.uk/IRDG_spermmorphology.doc.
34. Parker RM. Testing for reproductive toxicity. In Hood RD, Ed. *Developmental and Reproductive Toxicology, A Practical Approach*. Boca Raton, FL: CRC Press; 2006, pp 425–487.
35. MARTA and MTA. Historical Control Data (1992–1994) for developmental and reproductive toxicity studies using the Crl:CD®(SD)BR Rat. 1996. http://www.criver.com/flex_content_area/documents/rm_rm_r_tox_studies_crlcd_sd_br_rat.pdf.

36. US Food Drug Administration. *Redbook 2000*. Toxicological principles for the safety assessment of food ingredients, section IV.B.2. Guidelines for reporting the results of toxicity studies. CFSAN; 2000. http://www.cfsan.fda.gov/~redbook/red2ivb2.html.

37. US Food Drug Administration. Good Laboratory Practice Regulations, Final Rule. 21 CFR 58, 33768. 1987.

38. Hood RD. Animal models of effects of prenatal insult. In Hood RD, Ed. *Developmental Toxicology—Risk Assessment and the Future*. New York: Van Nostrand Reinhold; 1990, pp 183–188.

39. Tyl RW, Marr MC. Developmental toxicity testing—methodology. In Hood RD, Ed. *Developmental and Reproductive Toxicology, A Practical Approach*. Boca Raton, FL: CRC Press; 2006, pp 201–261.

40. US Food and Drug Administration. *Redbook 2000*. Toxicological principles for the safety assessment of food ingredients. Rockville, MD: US Department of Health and Human Services, Food and Drug Administration, Center for Food Safety and Applied Nutrition; 2000.

41. Johnson M, Everitt B. *Essential Reproduction*. Oxford: Blackwell Scientific Publications; 1984.

42. Collins TFX, Sprando RL, Shackelford ME, Hansen DK, Welsh JJ. Food and Drug Administration proposed testing guidelines for developmental toxicity studies. *Regul Toxicol Pharmacol* 1999;30:39–44.

43. Holson JF, Nemec MD, Stump DG, Kaufman LE, Lindstrom P, Varsho BJ. Significance, reliability, and interpretation of developmental and reproductive toxicity study findings. In Hood RD, Ed. *Developmental and Reproductive Toxicology, A Practical Approach*. Boca Raton, FL: CRC Press; 2006, pp 3–14.

44. Hood RD. Maternal vs. developmental toxicity. In Hood RD, Ed. *Developmental Toxicology—Risk Assessment and the Future*. New York: Van Nostrand Reinhold; 1990, pp 67–75.

45. Clark RL, Robertson RT, Minsker DH, Cohen SM, Tocco DJ, Allen HL, James ML, Bokelman DL. Diflunisal-induced maternal anemia as a cause of teratogenicity in rabbits. *Teratology* 1984;30:319–332.

46. Hood RD, Miller DB. Maternally mediated effects on development. In Hood RD, Ed. *Developmental and Reproductive Toxicology, A Practical Approach*. Boca Raton, FL: CRC Press; 2006, pp 93–124.

47. Solecki R, Bergmann B, Burgin H, Buschmann J, Clark R, Druga A, Van Duijnhoven EA, Duverger M, Edwards J, Freudenberger H, Guittin P, Hakaite P, Heinrich-Hirsch B, Hellwig J, Hofmann T, Hubel U, Khalil S, Klaus A, Kudicke S, Lingk W, Meredith T, Moxon M, Muller S, Paul M, Paumgartten F, Rohrdanz E, Pfeil R, Rauch-Ernst M, Seed J, Spezia F, Vickers C, Woelffel B, Chahoud I. Harmonization of rat fetal external and visceral terminology and classification. Report of the Fourth Workshop on the Terminology in Developmental Toxicology, Berlin, 18–20 April 2002. *Reprod Toxicol* 2003;17:625–637.

48. Stuckhardt JL, Poppe SM. Fresh visceral examination of rat and rabbit fetuses used in teratogenicity testing. *Teratog Carcinog Mutagen* 1984;4:181–188.

49. Wilson JG, Warkany J. *Teratology: Principles and Techniques* (Lectures and demonstrations Given at the first workshop in teratology, University of Florida, 2–8 February, 1964). Chicago: University of Chicago Press; 1965.

50. Wilson JG. *Environment and Birth Defects*. New York: Academic Press; 1973, Appendix III, pp 228–232.

51. Christian MS. Test methods for assessing female reproductive and developmental toxicology. In Hayes AW, Ed. *Principles and Methods of Toxicology*, 4th ed. Philadelphia: Taylor & Francis; 2001, pp 1301–1381.
52. Jonassen H, Lucacel C, Agajanov T, Gondek H, Schoenfeld HA. Comparison of modified Davidson's fixative and Bouin's solution in fixation of fetal soft tissue for evaluation using the Wilson's technique. *Teratology* 2002;65:332.
53. Astroff AB, Ray SE, Rowe LM, Hilbish KG, Linville AL, Stutz JP, Breslin WJ. Frozen-sectioning yields similar results as traditional methods for fetal cephalic examination in the rat. *Teratology* 2002;66:77–84.
54. Webb GN, Byrd RA. Simultaneous differential staining of cartilage and bone in rodent fetuses: an alcian blue and alizarin red S procedure without glacial acetic acid. *Biotech Histochem* 1994;69:181–185.
55. Menegola E, Broccia ML, Di Renzo F, Giavina E. Comparative study of sodium valproate-induced skeletal malformations using single or double staining methods. *Reprod Toxicol* 2002;16:815–823.
56. Crary DD. Modified benzyl alcohol clearing of alizarin stained specimens without loss of flexibility. *Stain Technol* 1962;37:124–125.
57. Kimmel CA, Trammel C. A rapid procedure for routine double staining of cartilage and bone in fetal and adult animals. *Stain Technol* 1981;56:271–273.
58. Narotsky MG, Rogers JM. Examination of the axial skeleton of fetal rodents. In Taun RS, Lo CW, Eds. *Methods in Molecular Biology. Developmental Biology Protocols, Volume 1*. Totowa, NJ: Humana Press; 2000, pp 139–150.
59. Young AD, Phipps DE, Astroff AB. Large-scale double-staining of rat fetal skeletons using alizarin red S and alcian blue. *Teratology* 2000;61:273–276.
60. Tyl RW, Marr MC. Developmental toxicity testing—methodology. In Hood RD, Ed. *Developmental and Reproductive Toxicology, A Practical Approach*. Boca Raton, FL: CRC Press; 2006, pp 201–261.
61. Chen JJ. Statistical analysis for developmental and reproductive toxicologists. In Hood RD, Ed. *Developmental and Reproductive Toxicology, A Practical Approach*. Boca Raton, FL: CRC Press; 2006, pp 697–711.
62. US Environmental Protection Agency. Guidelines for developmental toxicity risk assessment; notice. *Fed Reg* 1991;56(234):63798–63826.
63. Hood RD. Malformations vs. variations and "minor defects." In Hood RD, Ed. *Developmental Toxicology—Risk Assessment and the Future*. New York: Van Nostrand Reinhold; 1990, pp 178–182.
64. US Food Drug Administration. *Reviewer Guidance: Integration of Study Results to Assess Concerns About Human Reproductive and Developmental Toxicities* (draft). Rockville, MD: US Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research; 2001.
65. Kimmel G, Harris S, Tassinari M, Knudsen T, Cunny H, Tyl R, Carlton B, Holson J, Slikker W. Concern over decreased training in embryology and developmental/reproductive toxicology. *Birth Defects Res (Part B)* 2004;71:191–192.
66. Hood RD. Principles of developmental toxicology revisited. In Hood RD, Ed. *Developmental and Reproductive Toxicology, A Practical Approach*. Boca Raton, FL: CRC Press; 2006, pp 3–14.
67. Bates HK, Cunny HC, Kebede GA. Developmental neurotoxicity testing methodology. In Hood RD, Ed. *Handbook of Developmental Toxicology*. Boca Raton, FL: CRC Press; 1997, pp 291–324.

68. Kai S, Kohmura H. Comparison of developmental indices between Crj:CD(SD)IGS rats and Crj:CD(SD) rats. In Matsuzawa T, Inoue H, Eds. *Biological Reference Data on CD(SD)IGS-1999*. Yokohama: CD(SD)IGS Study Group; 1999, pp 157–167.
69. Ohkubo Y, Furuya K, Yamada M, Akamatsu H, Moritomo A, Ishida S, Watanabe K, Hayashi Y. Reproductive and developmental data on Crj:CD(SD)IGS rats. In Matsuzawa T, Inoue H, Eds. *Biological Reference Data on CD(SD)IGS-1999*. Yokohama: CD(SD)IGS Study Group; 1999, pp 153–156.
70. Ema M, Fujii S, Furukawa M, Kiguchi M, Ikka T, Harazano A. Rat two-generation reproductive toxicity study of bisphenol A. *Reprod Toxicol* 2001;15:505–523.
71. US Environmental Protection Agency. *Health Effects Test Guidelines OPPTS 870.3800, Reproduction and Fertility Effects*. Washington, DC: EPA 712-C-96-208; 1996, pp 1–11.
72. Organization for Economic Cooperation and Development. *OECD 416: Two-Generation Reproduction Toxicity Study* (updated guideline, adopted 22 January 2001). Paris: OECD; 2001.
73. Brent R. Utilization of juvenile animal studies to determine the human effects and risks of environmental toxicants during postnatal developmental stages. *Birth Defects Res (Part B)* 2004;71:303–320.
74. US Food and Drug Administration, Center for Drug Evaluation Research (CDER), Center for Biologics Evaluation and Research (CBER). *Guidance for Industry: M3 Nonclinical Safety Studies for the Conduct of Human Clinical Trials for Pharmaceuticals*. 1997.
75. US Food and Drug Administration, Center for Drug Evaluation and Research (CDER). *Pharmacology and Toxicology, Guidance for Industry: Nonclinical Safety Evaluation of Pediatric Drug Products*. 2006. http://www.fda.gov/cder/guidance/5671fnl.pdf.
76. Beckman DA, Feuston M. Landmarks in the development of the female reproductive system. *Birth Defects Res (Part B)* 2003;68P:137–143.
77. Marty MS, Chapin R, Parks L, Thorsrud B. Development and maturation of the male reproductive system. *Birth Defects Res (Part B)* 2003;68:125–136.
78. Zoetis T, Hurtt ME. Species comparison of lung development. *Birth Defects Res (Part B)* 2003;68:121–124.
79. Burri PH. Structural aspects of prenatal and postnatal development and growth of the lung, lung growth and development. In McDonald JA, Ed. *Lung Growth and Development*. New York: Marcel Dekker; 1997, pp 1–35.
80. Zoetis T. Species comparison of anatomical and functional renal development. *Birth Defects Res (Part B)* 2003;68:111–120.
81. Radde IC. Mechanism of drug absorption and their development. In Macleod SM, Radde IC, Eds. *Textbook of Pediatric Clinical Pharmacology*. Littleton, MA: PSG Publishing Co; 1985, pp 17–43.
82. Hew KW, Keller KA. Postnatal anatomical and functional development of the heart: a species comparison. *Birth Defects Res (Part B)* 2003;68:309–320.
83. Zoetis T, Tassinari MS, Bagi C, Walthall K, Hurtt ME. Species comparison of postnatal bone growth and development. *Birth Defects Res (Part B)* 2003;68:86–110.
84. Wood SL, Beyer BK, Cappon GD. Species comparison of postnatal CNS development: functional measures. *Birth Defects Res (Part B)* 2003;68:391–407.
85. Holsapple MP, West LJ, Landreth KS. Species comparison of anatomical and functional immune system development. *Birth Defects Res (Part B)* 2003;68:321–334.
86. Miyawaki T, Moriya N, Nagaoki T, Taniguchi N. Maturation of B-cell differentiation and T-cell regulatory function in infancy and childhood. *Immunol Rev* 1981;57:61–87.

87. Rice D, Barone S Jr. Critical periods of vulnerability for the developing nervous system: evidence from human and animal models. *Environ Health Perspect* 2000;108(Suppl 3):511–533.
88. Walthall K, Cappon GD, Hurtt ME, Zoetus T. Postnatal development of the gastrointestinal system: a species comparison. *Birth Defects Res (Part B)* 2005;74:132–156.
89. US Food Drug Administration, Center for Drug Evaluation and Research (CDER), Center for Biologics Evaluation and Research (CBER). ICH, Guidance for Industry: E11. *Clinical* investigation of medicinal products in the pediatric population. *Fed Reg* 2000;65:19777–19781. Docket No. 00D-1223.
90. ICH Harmonised Tripartite Guideline. *Guideline for Industry S3A, Note for Guidance on Toxicokinetics: Assessment of Systemic Exposure in Toxicity Studies*. ICH; 1994. http://www.ich.org/LOB/media/MEDIA495.pdf.

# CARCINOGENICITY STUDIES

SHAYNE COX GAD
*Gad Consulting Services, Cary, North Carolina*

**Contents**

1 Introduction
2 Regulatory Requirements and Timing
3 Species and Strain
4 Animal Husbandry
5 Dose Selection
  5.1 Number of Dose Levels
  5.2 Number of Control Groups
  5.3 Criteria for Dose Selection
6 Group Size
7 Route of Administration
8 Study Duration
9 Survival
10 Endpoints Measured
11 Statistical Analysis
12 Transgenic Mouse Models
  12.1 The Tg·AC Mouse Model
  12.2 The Tg.rasH2 Mouse Model
  12.3 The TSP $p53^{+/-}$ Mouse Model
  12.4 The $XPA^{-/-}$ Mouse Model
13 Interpretation of Results
  13.1 Criteria for a Positive Result
  13.2 Statistical Analysis
  13.3 Methods to Be Avoided
  13.4 Use of Historical Controls
  13.5 Relevance to Humans
14 Conclusion
  References

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

## 1 INTRODUCTION

In the experimental evaluation of substances for carcinogenesis based on experimental results of studies in a nonhuman species at some relatively high dose or exposure level, an attempt is made to predict the occurrence and level of tumorigenesis in humans at much lower levels. Examined here are the assumptions involved in this undertaking and a review of the aspects of design and interpretation of traditional long-term (lifetime) animal carcinogenicity studies as well as some of the alternative short-term models.

When are carcinogenicity studies required?

- Studies should be performed for pharmaceuticals with continuous clinical use for at least 6 months. (Most compounds indicated for 3 months of treatment would likely be used for 6 months and longer.)
- Studies are needed for pharmaceuticals used frequently in an intermittent manner. For example, carinogenicity studies would be needed for drugs used to treat allergic rhinitis, depression, and anxiety; while no carcinogenicity studies would be needed for anesthetics and radiolabeled imaging compounds.
- Additional criteria for defining cases of concern include:
  - Previous demonstration of a carcinogenic potential in the product class.
  - Structure–activity relationship suggesting carcinogenic risk.
  - Evidence of preneoplastic lesions in repeat-dose toxicity studies.
  - Long-term tissue retention of compound or metabolites resulting in local tissue reactions.
- Unequivocally genotoxic compounds do not need carcinogenicity studies.
- Carcinogenicity studies have the following timing considerations:
  - Carcinogenicity studies should be completed before application for marketing approval (i.e., upon NDA filing).
  - In the case of certain serious (life-threatening or limiting) diseases, studies may (should) be conducted postapproval.
  - Complete rodent carcinogenicity studies are not needed in advance of the conduct of large scale clinical trials, unless there is a special concern.
  - Where patient life expectancy is short (i.e., less than 2–3 years), no carcinogenicity studies are required.
  - Pharmaceuticals in tumor *adjuvant* therapy usually do need studies.

At least in a general way, we now understand what appear to be most of the mechanisms underlying chemical- and radiation-induced carcinogenesis. The most recent regulatory summary on identified carcinogens [1] lists 44 agents classified as "Known to be Human Carcinogens." Several hundred other compounds are also described as having lesser degrees of proof. A review of these mechanisms is not germane to this chapter (readers are referred to Miller and Miller [2] for a good short review), but it is now clear that cancer as seen in humans is the result of a multifocal set of causes.

Mechanisms and theories of chemical carcinogenesis include (1) *genetic* (all due to some mutagenic event), (2) *epigenetic* (no mutagenic event), (3) *oncogene*

*activation*, (4) *two-step* (induction/promotion), and (5) *multistep* (combination of 1–4) [3].

Considered in another way, the four major carcinogenic mechanisms are DNA damage, cell toxicity, cell proliferation, and oncogene activation. Any effective program to identify those drugs that have the potential to cause or increase the incidence of neoplasia in humans must effectively screen for these mechanisms [4].

The single most important statistical consideration in the design of bioassays in the past was based on the point of view that what was being observed and evaluated was a simple quantal response (cancer occurred or it didn't), and that a sufficient number of animals needed to be used to have reasonable expectations of detecting any meaningful risks of such an effect. Although the single fact of whether or not the simple incidence of neoplastic tumors is increased due to an agent of concern is of interest, a much more complex model must now be considered. The time-to-tumor, patterns of tumor incidence, effects on survival rate, and age of first tumor all must now be captured in a bioassay and included in an evaluation of the relevant risk to humans [5].

The task of interpreting the results of any of the animal-based bioassays must be considered from three different potential perspectives as to organ responsiveness:

Group I. Those organs with high animal and low human neoplasia rates.

Group II. Those organs with high neoplasia rates in both animals and humans.

Group III. Those organs with low animal but high human neoplasia rates.

Note that not considered is the potential other case—where the neoplasia rates are low for both animals and humans. This is a very rare case and one for which our current bioassay designs probably lack sufficient power to be effective.

In group I, the use of animal cancer data obtained in the liver, kidney, forestomach, and thyroid gland are perceived by some as being hyperresponsive, too sensitive, and of limited value and utility in the animal cancer data obtained in group I organs. The liver is such a responsive and important organ in the interpretation of cancinogenesis data that the discussion of this subject area has been broken up into three sections for human, rat, and mouse data. Peroxisome proliferation in the liver, particularly in mice, is an area of interpretive battle as in many cases the metabolism and mechanisms involved are not relevant to humans.

Group II organs (mammary gland, hematopoietic, urinary bladder, oral cavity, and skin) are less of an interpretive battleground than group I organs. For group II organs, all four major mechanisms of carcinogenesis (electrophile generation, oxidation of DNA, receptor–protein interactions, and cell proliferation) are known to be important. The high cancer rates for group II organs in both experimental animals and humans may at first give us a false sense of security about how well the experimental animal models are working. As we are better able to understand the probable carcinogenic mechanism(s) in the same organ in the three species, we may find that the important differences between the three species are more numerous than we suspect. This is particularly true for receptor-based and for cell-proliferation-based carcinogenic mechanisms.

Animal cancer data for group III organs are the opposite of group I organs. Group III organs have low animal cancer rates and high human cancer rates. In

contrast to the continuing clamor and interpretive battleground associated with group I organs, there is little debate over group III organs. Few voices have questioned the adequacy of the present-day animal bioassay to protect the public health from possible cancer risks in these group III organs. Improved efforts must be made toward the development of cancer-predictive systems or short-term tests for cancer of the prostate gland, pancreas, colon/rectum, and cervix/uterus.

Carcinogenicity bioassays are the longest and most expensive of the extensive battery of toxicology studies required for the registration of pharmaceutical products in the United States and in other major countries. In addition, they are often the most controversial with respect to interpretation of their results. These studies are important because, as noted by the International Agency for Research on Cancer [6], "in the absence of adequate data on humans, it is biologically plausible and prudent to regard agents for which there is sufficient evidence of carcinogenicity in experimental animals as if they presented a carcinogenic risk to humans."

In this chapter, we consider the major factors involved in the design, conduct, analysis, and interpretation of carcinogenicity studies as they are performed in the pharmaceutical industry.

## 2 REGULATORY REQUIREMENTS AND TIMING

The prior FDA guidance on the need for carcinogenicity testing of pharmaceuticals presented a dual criteria: such studies were required to support registration of a drug that was to be administered for a period of 3 months or more (in Japan and Europe this was stated to be 6 months or more); and such testing had to be completed before filing for registration in such cases. ICH guidelines [5] now fix this human exposure period that triggers the need for testing at 6 months, excluding agents given infrequently through a lifetime or for shorter periods of exposure unless there is reason for concern (such as positive findings in genotoxicity studies, structure–activity relationships suggesting such risk, evidence of preneoplastic lesions in repeat dose studies, or previous demonstration of carcinogenic potential in the product class that is considered relevant to humans). Such studies are still only required to be completed before filing for registration. Most developers conduct carcinogenicity studies in parallel with Phase III clinical studies.

Endogenous peptides, protein substances, and their analogues are generally not required to be evaluated for carcinogenicity. There are three conditions that call the need into question, however:

- Products where there are significant differences in biological effects to the natural counterparts.
- Products where modifications lead to significant changes in structure compared to the natural substance.
- Products resulting in humans having a significant increase over the existing local or systemic concentration.

The ICH and EMEA have also given guidance on design, dose selection, statistical analysis, and interpretation for such studies [7–11]. The FDA has also offered guidance; the most recent form is a 44 page document available online [12].

There has been extensive debate and consideration on the relevance and value of the traditional long-term rodent bioassays. The FDA looked at rat and mouse studies for 282 human pharmaceuticals, resulting in the conclusion that "sufficient evidence is now available for some alternative *in vivo* carcinogenicity models to support their application as complimentary studies *in combination with a single 2-year carcinogenicity study* [italics added] to identify trans-species tumorigens" [13].

The Europeans, meanwhile, have focused on the need for better care in study design, conduct, and interpretation [14], aiming to incorporate these in the revision of the Center for Proprietary Medicinal Products (CPMP) carcinogenicity guidelines.

## 3 SPECIES AND STRAIN

Two rodent species are routinely used for carcinogenicity testing in the pharmaceutical industry—the mouse and the rat. Sprague–Dawley derived rats are most commonly used in American pharmaceutical toxicology laboratories. However, the Wistar and Fischer 344 strains are favored by some companies, while the Long Evans and CFE (Carworth) strains are also used but rarely [15].

With respect to mice, the CD-1 is by far the most commonly used strain in the pharmaceutical industry. Other strains used less frequently are the B6C3F1, CF-1, NMRI, C57B1, BALB/c, and Swiss [15, 16]. "Swiss" is the generic term since most currently used inbred and outbred strains were originally derived from the "Swiss" mouse.

If either the mouse or the rat is considered to be an inappropriate species for a carcinogenicity study, the hamster is usually chosen as the second species.

The use of two species in carcinogenicity studies is based on the traditional wisdom that no single species can be considered an adequate predictor of carcinogenic effects in humans. Absence of carcinogenic activity in two different species is thought to provide a greater level of confidence that a compound is "safe" for humans than data derived from a single species.

One may question this reasoning on the basis that data from two "poor predictors" may not be significantly better than data from a single species. It is also reasonable to expect that the ability of one rodent species to predict a carcinogenic effect in a second rodent species should be at least equal to, if not better than, its ability to predict carcinogenicity in humans. The concordance between mouse and rat carcinogenicity data has been investigated and a summary of the results is presented in the next paragraph.

A review of data from 250 chemicals found an 82% concordance between results of carcinogenicity testing in the mouse and the rat [17]. Haseman et al. [18] reported a concordance of 73% for 60 compounds studied in both species. However, 30–40% of 186 National Cancer Institute (NCI) chemicals were found to be positive in one species and negative in the other [19]. It is reasonable to conclude that neither rodent species will always predict the results in the other rodent species or in humans, and that the use of two species will continue until we have a much better understanding of the mechanisms of carcinogenesis.

The choice of species and strain to be used in a carcinogenicity study is based on various criteria including susceptibility to tumor induction, incidence of sponta-

neous tumors, survival, existence of an adequate historical database, and availability.

Susceptibility to tumor induction is an important criterion. There would be little justification for doing carcinogenicity studies in an animal model that did not respond when treated with a "true" carcinogen. Ideally, the perfect species/strain would have the same susceptibility to tumor induction as the human. Unfortunately, this information is usually unavailable, and the tendency has been to choose animal models that are highly sensitive to tumor induction to minimize the probability of false negatives.

The incidence of spontaneous tumors is also an important issue. Rodent species and strains differ greatly in the incidence of various types of spontaneous tumors. The Sprague–Dawley strain, although preferred by most pharmaceutical companies, has a very high incidence of mammary tumors in aging females, which results in substantial morbidity during the second year of a carcinogenicity study. If one chooses the Fischer 344 (F344) strain, the female mammary tumor incidence will be lower, but the incidence of testicular tumors will be higher (close to 100%) than that in Sprague–Dawley rats.

A high spontaneous tumor incidence can compromise the results of a carcinogenicity study in two ways. If a compound induces tumors at a site that already has a high spontaneous tumor incidence, it may be impossible to detect an increase above the high background "noise." Conversely, if a significant increase above control levels is demonstrated, one may question the relevance of this finding to humans on the basis that the species is "highly susceptible" to tumors of this type.

The ability of a species/strain to survive for an adequate period is essential for a valid assessment of carcinogenicity. Poor survival has caused regulatory problems for pharmaceutical companies and is therefore an important issue [15]. The underlying concept is that animals should be exposed to the drug for the greater part of their normal life span to make a valid assessment of carcinogenicity. If animals on study die from causes other than drug-induced tumors, they may not have been at risk long enough for tumors to have developed. The sensitivity of the bioassay would be reduced and the probability of a false-negative result would be increased.

The availability of an adequate historical database is often cited as an important criterion for species/strain selection. Historical control data can sometimes be useful in evaluating the results of a study. Although such data are not considered equal in value to concurrent control data, they can be helpful if there is reason to believe that the concurrent control data are "atypical" for the species/strain.

Although outbred stocks (e.g., Sprague–Dawley rats and CD-1 mice) are generally favored in the pharmaceutical industry, inbred strains are also used (e.g., Fischer 344 rats and B6C3F1 mice). Inbred strains may offer greater uniformity of response, more predictable tumor incidence, and better reproducibility than outbred strains. However, their genetic homogeneity may also result in a narrower range of sensitivity to potential carcinogens than exists in random-bred animals. In addition, extrapolation of animal data to humans is the ultimate goal of carcinogenicity studies, and the human population is anything but genetically homogeneous.

The ideal species for carcinogenicity bioassays should absorb, metabolize, and excrete the compound under study exactly as humans do. Unfortunately, because of the small number of species that meet the other criteria for selection, there is

limited practical utility to this important scientific concept, as applied to carcinogenicity studies.

Before concluding this discussion of species/strain selection, it may be worthwhile to take a closer look at the animals preferred by pharmaceutical companies to determine to what extent they meet the conditions described earlier. Advantages of the CD-1 mouse are (1) a good historical database including various routes of exposure, (2) demonstrated susceptibility to induction of tumors, and (3) relatively low spontaneous incidence of certain tumors to which other strains are highly susceptible, especially mammary and hepatic tumors. Disadvantages are (1) lack of homogeneity, (2) relatively low survival, (3) moderate to high incidence of spontaneous pulmonary tumors and leukemias, and (4) high incidence of amyloidosis in important organs, including the liver, kidney, spleen, thyroid, and adrenals [20].

There was a reduction in survival of Sprague–Dawley rats and rats of other strains during the 1990s [21]. This reduction may have been the result of *ad libitum* feeding, as preliminary results suggest that caloric restriction may improve survival. Leukemia appears to be the major cause of decreasing survival in the F344 rat. While this reduced survival trend has flattened out, a resumption of the problem of reduced survival may necessitate a reevaluation of the survival requirements for carcinogenicity studies by regulatory agencies.

## 4 ANIMAL HUSBANDRY

Because of the long duration and expense of carcinogenicity studies, the care of animals used in these studies is of paramount importance. Various physical and biological factors can affect the outcome of these studies. Some important physical factors include light, temperature, relative humidity, ventilation, atmospheric conditions, noise, diet, housing, and bedding [22]. Biological factors include bacteria and viruses that may cause infections and diseases.

The duration, intensity, and quality of light can influence many physiological responses, including tumor incidence [23, 24]. High light intensity may cause eye lesions, including retinal atrophy and opacities [25, 26]. Rats housed in the top row and the side columns of a rack may be the most severely affected.

The influence of light on the health of animals may be managed in several ways. The animals may be randomly assigned to their cages on a rack such that each column contains animals of a single dose group. The location of the columns on the rack may also be randomized so that the effect of light is approximately equal for all dose groups. In addition, the cages of each column of the rack may be rotated from top to bottom when the racks are changed.

Room temperature has been shown to influence the incidence of skin tumors in mice [27]. Changes in relative humidity may alter food and water intake [28]. Low humidity may cause "ringtail," especially if animals are housed in wire mesh cages [29].

Diets for rodents in carcinogenesis studies should ideally be nutritionally adequate while avoiding excesses of nutrients that may have adverse effects.

Types of caging and bedding have been shown to affect the incidence and latency of skin tumors in mice. In a study by DePass et al. [30], benzo[a]pyrene-treated mice were housed either in stainless steel cages or polycarbonate shoebox cages with

hardwood bedding. The mice housed in shoebox cages developed tumors earlier and with higher frequency than those housed in steel cages.

Housing of rats in stainless steel cages with wire mesh floors may result in decubitous ulcers on the plantar surfaces. This condition may be a significant problem associated with high morbidity and may affect survival of the animals if euthanasia is performed for humane reasons. Ulcers are particularly frequent and severe in older male Sprague–Dawley rats, perhaps because of their large size and weight compared with females and rats of other strains.

Common viral infections may affect the outcome of carcinogenicity studies by altering survival or tumor incidence. Nevertheless, viral infections did not cause consistent adverse effects on survival or tumor prevalence in control F344 rats from 28 NCI-NTP studies, although body weights were reduced by Sendai and pneumonia viruses of mice [31]. The probability of such infections can be minimized by using viral-antibody-free animals, which are readily available.

## 5 DOSE SELECTION

### 5.1 Number of Dose Levels

For pharmaceuticals, most carcinogenicity studies employ at least three dose levels in addition to the controls, but four levels have occasionally been used [15]. The use of three or four dose levels satisfies regulatory requirements [32] as well as scientific and practical considerations. If a carcinogenic response is observed, information on the nature of the dose–response relationship will be available. If excessive mortality occurs at the highest dose level, a valid assessment of carcinogenicity is still possible when there is adequate survival at the lower dose levels.

### 5.2 Number of Control Groups

Pharmaceutical companies have frequently favored the use of two control groups of equal size [15]. A single control group of the same size as the treated groups is also used and, less frequently, one double-sized control group may be used. The diversity of study designs reflects the breadth of opinion among toxicologists and statisticians on this issue.

Use of two control groups has the advantage of providing an estimate of the variation in tumor incidence between two groups of animals in the absence of a drug effect. If there are no significant differences between the control groups, the data can be pooled, and the analysis is identical to that using a single, double-sized group. When significant differences occur between the control groups, one must compare the data from the drug-treated groups separately with each control group.

There will be situations in which the incidence of a tumor in one or more drug-treated groups is significantly higher than that of one control group but similar to that of the other control group. In such a situation, it is often helpful to compare the tumor incidences in the control groups to appropriate historical control data. One may often conclude that, for this tumor, one of the control groups is more "typical" than the other and should therefore be given more weight in interpreting the differences in tumor incidence.

In spite of its popularity in the pharmaceutical industry, the use of two control groups is opposed by some statisticians on the grounds that a significant difference between the two groups may indicate that the study was compromised by excessive, uncontrolled variation. Haseman et al. [33], however, analyzed tumor incidence data from 18 color additives tested in rats and mice and found that the frequency of significant pairwise differences between the two concurrent control groups did not exceed that which would be expected by chance alone.

The use of one double-sized group is sometimes preferred because it may provide a better estimate of the true control tumor incidence than that provided by a smaller group. Nevertheless, more statistical power would be obtained by assigning the additional animals equally to all dose groups rather than to the control group only, if power is a primary consideration.

### 5.3 Criteria for Dose Selection

Dose selection is one of the most important and difficult activities in the design of a toxicology study. It is especially critical in carcinogenicity studies because of their long duration. Whereas faulty dose selection in an acute or subchronic toxicity study can easily be corrected by repeating the study, this situation is much less desirable in a carcinogenicity study, especially since such problems may not become evident until the last stages of the study.

The information used for dose selection usually comes from subchronic toxicity studies, but other information about the pharmacological effects of a drug and its metabolism and pharmacokinetics may also be considered. The maximum recommended human dose (MRHD) of the drug may be an additional criterion, if this is known when the carcinogenicity studies are being designed. An additional consideration is that for U.S. FDA submissions, protocols must be reviewed and dose selections justified (prior to study start) by the Carcinogenicity Advising Committee (CAC) [34].

For most pharmaceutical companies, the doses selected are as follows. The highest dose is selected to be the estimated maximum tolerated dose (MTD). The lowest dose is usually a small multiple (one to five times) of the MRHD, and the mid-dose approximates the geometric mean of the other two doses [15, 35].

The MTD is commonly estimated to be the maximum dose that can be administered for the duration of the study that will not compromise the survival of the animals by causes other than carcinogenicity. It should be defined separately for males and females. ICH [6] states that the MTD is "that dose which is predicted to produce a minimum toxic effect over the course of the carcinogenicity study, usually predicted from the results of a 90-day study." Factors used to define minimum toxicity include no more than a 10% decrease in body weight gain relative to controls, target organ toxicity, and/or significant alterations in clinical pathology parameters. If the MTD has been chosen appropriately, there should be no adverse effect on survival, only a modest decrement in body weight gain and minimal overt signs of toxicity. The procedures for dose selection described above are generally consistent with major regulatory guidelines for carcinogenicity studies[1] [21, 32]. There are,

[1]Note that the FDA *Redbook 2000* applies, strictly speaking, only to food additives. It is cited here because it is a well-known toxicology guideline routinely applied to animal pharmaceuticals to which humans may be exposed. The *Redbook* has recently been updated by the FDA [16].

however, exceptions to the general approach described above. For example, for nontoxic drugs, the difference between the high and the low doses may be many orders of magnitude, if the high dose is set at the estimated MTD and the low dose is a small multiple of the clinical dose. Some guidelines suggest that the low dose be no less than 10% of the high dose [32]. In this situation, it may be acceptable to set the high dose at 100 times the MRHD, even if the MTD is not achieved [32]. Similarly, when a drug is administered in the diet, the highest concentration should not exceed 5% of the total diet, whether or not the MTD is achieved [36].

**MTD (Maximum Tolerated Dose) Concept**

- Criteria of MTD
  - Dose to elicit signs of minimum toxicity without altering the animal's normal life span due to effects other than carcinogenicity.
  - No more than 10% decrease in body weight gain.
  - Target organ toxicity and/or alterations in clinical pathological parameters.
- Criticism: For compounds with low toxicity, extremely high unrealistic doses may result.
- Threshold concept for nongenotoxic compounds: alterations of normal physiology.
- The high dose should allow an adequate safety margin over the human therapeutic exposure and should be tolerated without significant physiological dysfunction or increased mortality.

**High Dose Selection**

- Toxicity endpoints
  - Traditionally, carcinogenicity studies have relied on the MTD (maximum tolerated dose) concept; no other toxicological endpoints are agreed upon.
  - Pharmacokinetic endpoints may determine the high dose (you reach a saturation of absorption plateau, where additional dosing does not achieve any higher plasma levels (systemic exposure).
  - Systemic exposure represents a large multiple of the human dose (minimum 25×) on a basis of $mg/m^2$ body surface.
- Pharmacodynamic endpoints: receptor selectivity, specific argumentation.
- Maximally feasible or practicable dose (i.e., you cannot give any more): 5% of the diet or 1500 mg/kg/day (only for human doses up to 500 mg/day).
- Additional endpoints: must be scientifically justified.

Metabolism and/or pharmacokinetic data, when available, should also be considered in the dose selection process. It is desirable that a drug not be administered at such a high dose that it is excreted in a different manner than at lower doses, such as the MRHD. Similarly, the high dose should not lead to the formation of metabolites other than those formed at lower (clinical) doses. If data show that a given dosage produces maximum plasma levels, administration of higher doses should be

unnecessary. These considerations may be very useful when interpreting the results of the study or attempting to extrapolate the results to humans.

## 6 GROUP SIZE

The minimum number of animals assigned to each dose group in pharmaceutical carcinogenicity studies is 50 of each sex [15]. Most companies, however, use more than the minimum number, and some use up to 80 animals per sex per group. The most important factor in determining group size is the need to have an adequate number of animals for a valid assessment of carcinogenic activity at the end of the study. For this reason, larger group sizes are used when the drug is administered by daily gavage because this procedure may result in accidental deaths by perforation of the esophagus or aspiration into the lungs. Larger group sizes are also used when the carcinogenicity study is combined with a chronic toxicity study in the rat. In this case, serial sacrifices are performed at 6 and 12 months to evaluate potential toxic effects of the drug.

In the final analysis, the sensitivity of the bioassay for detecting carcinogens is directly related to the sample size. Use of the MTD has often been justified based on the small number of animals at risk compared to the potential human population, in spite of the difficulties inherent in extrapolating effects at high doses to those expected at much lower clinical doses. A reasonable compromise may be the use of doses lower than the MTD combined with a larger group size than the 50 per sex minimum accepted by regulatory agencies.

## 7 ROUTE OF ADMINISTRATION

In the pharmaceutical industry, the two most common routes of administration are via diet and gavage [15]. Some compounds are given by drinking water, topical (dermal) application, or injection, depending on the expected clinical exposure route, which is the primary criterion for determining the route of administration in carcinogenicity studies. When more than one clinical route is anticipated for a drug, the dietary route is often chosen for practical reasons.

Dietary administration is often preferred over gavage because it is far less labor intensive. Another advantage is that the MTD has rarely been overestimated in dietary studies, whereas it has often been overestimated in gavage studies, according to data from the NTP [37]. The dietary route is unsuitable for drugs that are unstable in rodent chow or unpalatable. The dietary route is also disadvantaged by the fact that dosage can only be estimated based on body weight and food intake data, in contrast with gavage by which an exact dose can be given. Disadvantages of gavage testing are the likelihood of gavage-related trauma, such as puncture of the trachea or esophagus, and possible vehicle (e.g., corn oil) effects.

When doing studies by the dietary route, the drug may be administered as a constant concentration at each dose level, or the concentration may be increased as body weight increases to maintain a constant dose on a milligram per kilogram basis. The latter method allows greater control of the administered dose and avoids

age- and sex-related variations in the dose received, which occur with the former method. Both methods are acceptable to regulatory agencies.

## 8 STUDY DURATION

The duration of carcinogenicity studies for both rats and mice is 2 years in most pharmaceutical laboratories [15]. Occasionally, rat studies are extended to 30 months, while come companies terminate mouse studies at 18 months. The difference in duration between mouse and rat studies is based on the belief that rats have a longer natural life span than mice. Recent data indicate, however, that this is not the case. The most commonly used strains, the Sprague–Dawley rat and the CD-1 mouse, have approximately equal survival at 2 years, based on industry data [15]. The same is true for the most popular inbred strains, the Fischer 344 rat and the B6C3F1 mouse [15]. Data from NCI studies confirm that the 2 year survival of the B6C3F1 mouse is at least equal to, if not greater than, that of the Fischer 344 rat [38].

## 9 SURVIVAL

As stated earlier, adequate survival is of primary importance in carcinogenicity studies because animals must be exposed to a drug for the greater part of their life span to increase the probability that late-occurring tumors can be detected. Early mortality, resulting from causes other than tumors, can jeopardize the validity of a study because dead animals cannot get tumors.

In general, the sensitivity of a carcinogenicity bioassay is increased when animals survive to the end of their natural life span, because weak carcinogens may induce late-occurring tumors. The potency of a carcinogen is often inversely related to the time to tumor development. By analogy, as the dose of a carcinogen is reduced, the time to tumor occurrence is increased [30, 39].

Why do we not allow all animals in a carcinogenicity study to live until they die a natural death if by so doing we could identify more drugs as carcinogens? In fact, the sensitivity of a bioassay may not be improved by allowing the animals to live out their natural life span because the incidence of spontaneous tumors tends to increase with age. Thus, depending on the tumor type, the ability of the bioassay to detect a drug-related increase in tumor incidence may actually decrease, rather than increase, with time. Therefore, the optimum duration of a carcinogenicity study is that which allows late-occurring tumors to be detected but does not allow the incidence of spontaneous tumors to become excessive.

Reduced survival in a carcinogenicity study may or may not be drug related. Sometimes, the MTD is exceeded and increased mortality occurs at the highest dose level and occasionally at the mid-dose level as well. This situation may not necessarily invalidate a study; in fact, the protocol may be amended to minimize the impact of the drug-induced mortality. For example, cessation of drug treatment may enhance the survival of the animals in the affected groups and allow previously initiated tumors to develop. As shown by Littlefield et al. [39] in the CNTR ED01 study, liver tumors induced by 2-acetylaminofluorene, which appeared very late in the

study, were shown to have been induced much earlier and not to require the continuous presence of the carcinogen to develop. By contrast, bladder tumors that occurred in the same study were dependent on the continued presence of the carcinogen.

Whether drug treatment is terminated or not, drug-related toxicity may also be managed by performing complete histopathology on animals in the lower dose groups rather than on high dose and control animals only. If there is no increase in tumor incidence at a lower dose level that is not compromised by reduced survival, the study may still be considered valid as an assessment of carcinogenicity.

When reduced survival is related to factors other than excessive toxicity, the number of animals at risk for tumor development may be inadequate, and the validity of the study may be compromised even in the absence of a drug effect on survival. Obviously, the adjustments described above for excessive, drug-related toxicity are not relevant to this situation.

There is no unanimity of opinion among regulatory agencies as to the minimum survival required to produce a valid carcinogenicity study or as to the best approach for dealing with survival problems. Even within a single agency such as the FDA, different opinions exist on these issues. For example, the recently issued FDA *Redbook II Draft Guideline* [21] requires that rats, mice, or hamsters be treated for 24 months. Early termination due to decreased survival is not recommended. The EEC guidelines differ in that they suggest termination of the study when survival in the control group reaches 20%, while Japanese guidelines suggest termination at 25% survival in the control or low dose groups [32]. These provisions make good sense in that they do not request termination of the study when drug-related mortality may be present only at the highest dose.

## 10 ENDPOINTS MEASURED

A carcinogenicity study is more focused than a chronic toxicity study—fewer endpoints are evaluated and as such it is a simpler study. The key endpoints are actually few:

- Pathology (limited to neoplastic and preneoplastic tissue transformations).
- Body weight (to ensure that toxicity is not so great as to invalidate the assays—and also that it is just sufficient to validate the assay).
- Survival (key to determining when to terminate the study).
- Clinical pathology (limited to evaluating the morphology of white blood cells, and usually this is actually deferred until there are indications that such data is needed).
- Food consumption (actually measured to ensure that dietary administration doses are accurate).

Only pathology will be considered in detail.

The primary information for carcinogenicity evaluation is generated by pathologists. Table 1 lists the tissues normally collected, processed, and evaluated (see also Table 2). These professionals, like any other group of professionals, vary in

**TABLE 1 Standard Tissue List**

| | | |
|---|---|---|
| Kidney | Urinary bladder | Aorta |
| Heart | Trachea | Lungs |
| Liver | Gallbladder | Pancreas |
| Fat | Salivary gland | Spleen |
| Cervical lymph node | Mesenteric lymph node | Thymus |
| Tongue | Esophagus | Stomach |
| Duodenum | Jejunum | Ileum |
| Cecum | Colon | Mammary gland |
| Skin | Skeletal muscle | Sciatic nerve |
| Parathyroid | Thyroid | Adrenal gland |
| Pituitary | Prostate | Seminal vesicles |
| Testes | Epididymides | Ovaries |
| Oviducts | Uterine horns | Uterine body |
| Cervix | Vagina | Brain |
| Spinal cord | Sternum | Rib/bone |
| Eyes | Harderian glands | BM smear |
| Nares | Clitoral/preputial gland | Zymbal's gland |
| Gross lesions | | |

**TABLE 2 Current Global Carcinogenicity Study Requirements**

| Guide | Eye | Blood Smear | RBC | WBC | Differential | Clinical Chemistry | Urine | Organ | Histopathology |
|---|---|---|---|---|---|---|---|---|---|
| EPA 1998 | – | + | – | – | + C+H | – | – | + | C+H |
| Redbook 1993; 2000 | – | – | + | + | + | + | + | + | All? |
| Japan 1999 | – | + | + | + | – | – | – | – | C+H |
| CHMP 2003 | + | – | + | + | + | + | + | + | All |
| CDER 1993 | – | + | + | + | – | – | – | – | C+H |

their training and experience, and these are characteristics that may influence the evaluation in a number of ways.

1. Differences in terminology may be important when considering controversial lesions.
2. Lack of consistency throughout a study is likely when a pathologist has only recently become involved with rodent carcinogenicity. Training is often in a clinical situation (especially in Europe), where each animal or person is unique, and a rodent carcinogenicity study can consist of 500 animals.
3. Unfamiliarity with the observed lesion in a particular species may cause problems in interpretation.

Possible bias introduced by knowledge of treatment can be corrected in several ways, but the use of a two-stage process would seem to be most efficient:

1. An initial evaluation is performed with full knowledge of the animal's history, including treatment.
2. A second evaluation of specific lesions is then carried out. This should be done blind, either by the same pathologist or, preferably, by the same and a second pathologist.

Differences in evaluation between pathologists should always be discussed by them to resolve the differences; they may be due to subtle differences in diagnosis and do not indicate incompetence in one of the pathologists. It is unacceptable for a study sponsor to shop around until a pathologist is found who gives (for whatever reason) the result the sponsor is looking for without giving an opportunity for interaction with all of the other evaluators. Sometimes these diagnoses are given years apart, during which time understanding of the pathogenesis of lesions may change, and even the first pathologist may not arrive at the same conclusion as he/she did some years ago.

Evaluation of the data is not purely a statistical exercise. A number of important factors should be considered: (1) dose–effect relationship, (2) a shift toward more anaplastic tumors in organs where tumors are common, (3) earlier appearance of tumors, and (4) presence of preneoplastic lesions.

The language used to describe the carcinogenic response has masked its complexity and presents a stumbling block to its understanding among nonhistopathologists. Benign or malignant neoplasms do not arise without some precursor change within normal tissue. An important concept in carcinogenicity evaluation is that of neoplastic progression, which was derived from studies on skin tumors and expanded to a number of other tissues. There is, on many occasions, a far from clear distinction between hyperplastic and "benign" neoplasia and between benign and malignant neoplasia.

Hyperplasia and benign and malignant neoplasia are convenient medical terms with prognostic significance. Hyperplasia can occur either as a regenerative response to injury, with no neoplastic connotations, or as a sustained response to a carcinogenic agent. It is an increase in the number of normal cells retaining normal intercellular relationships within a tissue. This normally may break down, resulting in altered growth patterns and altered cellular differentiation—a condition that may be described as atypical hyperplasia or presumptively as preneoplastic lesions. Possible sequelae to hyperplasia are (1) persistence without qualitative change in either structure or behavior, (2) permanent regression, (3) regression with later reappearance, and (4) progression to develop new characteristics indicating increased probability of malignancy. The last of these in the least likely to occur in experimental multistage models, such as in mouse skin or rat liver, where large numbers of hyperplastic lesions may occur, but notably fewer carcinomas develop from them.

Benign neoplasms in most rodent tissues apparently arise in hyperplastic foci, for example, squamous cell papillomas of the skin and forestomach. Furthermore, these papillomas seldom demonstrate autonomous growth and even fewer progress to squamous cell carcinomas. This decisive progression to carcinoma, when it occurs,

provides powerful evidence for the multistage theory of carcinogenesis: the new, malignant cells arising as a focus within the papilloma or even in an area of hyperplasia, since the papilloma is not a necessary intermediate stage. In other organs, benign neoplasia is usually characterized by well differentiated cell morphology, a fairly uniform growth pattern, clear demarcation from surrounding tissues, and no evidence of invasion. The progression toward malignancy involves anaplasia (loss of differentiation) and pleomorphism (variety of phenotypic characteristics within the neoplasm). These changes may be focal in an otherwise benign neoplasm and may vary in degree and extent. Evidence of invasion into surrounding tissues or of metastasis is not an essential characteristic of malignancy, although it strengthens the diagnosis.

The grouping together of certain tumor types can aid statistical analysis, but it must be done carefully, with full appreciation of the biology and whatever is known of the pathogenesis of the lesions. Grouping for analysis of all animals showing neoplasia, irrespective of the tumor type, is inappropriate because the incidence in most treatment control groups can be very high and, in U.S. National Toxicology Program studies, approaches 100% in rats and 50–70% in mice (Table 3).

There may be similar incidences of tumors in aging people, but the real prevalence of tumors in human populations is uncertain. In the United States, where autopsies are uncommon, over one-third reveal previously undiagnosed cancers when they are conducted. A single type of neoplasm, renal adenoma, is present in 15–20% of all adult kidneys, although it is unclear whether these 2–6 mm foci of proliferating tubular and papillary epithelium represent small carcinomas or benign precursors of renal cell carcinomas. Irrespective of the significance of these lesions in human pathology, the presence of similar foci in a rodent carcinogenicity experiment would trigger the recording of renal tumor-bearing animals and hence their consideration in the statistical and pathological evaluation processes. Evaluation is further complicated by the increased background incidences of tumors as animals get older.

**TABLE 3 Tumor-Bearing Animals in Control Groups from Rodent Studies**

| Control Animals for 2 year NTP Bioassay | Number of Animals | Those (%) with Tumors | | |
|---|---|---|---|---|
| | | Malignant | Benign | Total |
| B6C3F1 mice | | | | |
| Male | 1692 | 42 | 35 | 64 |
| Female | 1689 | 45 | 33 | 64 |
| F344 rats | | | | |
| Male | 1596 | 55 | 95 | 98 |
| Female | 1643 | 38 | 76 | 88 |
| Osborne–Mendel rats | | | | |
| Male | 50 | 26 | 68 | 78 |
| Female | 50 | 12 | 80 | 88 |
| Sprague–Dawley rats | | | | |
| Male | 56 | 9 | 36 | 39 |
| Female | 56 | 30 | 68 | 79 |

*Source*: J. K. Haseman (unpublished summary of U.S. NTP data).

The independent analysis of every different diagnosis in rodent studies would also mask significant effects in many cases, while enhancing them in others. Benign and malignant neoplasms of a particular histogenesis are often grouped because the one is seen as a progression from the other. However, this grouping may result in a nonsignificant difference from the controls because there has been an acceleration of progression toward malignancy, the incidence of benign neoplasms decreasing while the malignant neoplasms increase. Guidelines are available for "lumping" or "splitting" tumor types, but in using them, the basis for the classification of neoplastic lesions should be clarified, especially when data generated over several or many years are coupled, since diagnostic criteria and ideas regarding tumor histogenesis may have changed. Reliance on tabulated results alone can lead to serious misinterpretation by those not closely connected with a particular study. For this very important reason, the pathology and toxicology narrative should be full and clear. If it is not, then there will always be doubts about future interpretations, even if these doubts are not, in reality, justified.

## 11 STATISTICAL ANALYSIS

Irrespective of the specific protocols used, all carcinogenicity studies end with a statistical comparison of tumor proportions between treated and control groups. This analysis is necessary because the control incidence of most tumor types is rarely zero. In the unlikely case that a type of tumor is found in treated animals but not in concurrent or appropriate historical controls, it is reasonable to conclude that the tumor is drug-related without statistical analysis.

Most pharmaceutical companies analyze tumor data using mortality-adjusted methods [15, 40]. The Peto/International Agency for Research on Cancer (IRC) methodology is most commonly used, perhaps because this method is currently favored by the FDA [41]. The use of life-table methods is most appropriate for "lethal" tumors, that is, those that cause the death of the animals. Various statistical methods are available for analyzing the incidence of the lethal and nonlethal tumors [42–45]. These methods are especially useful when there are drug-related differences in mortality rates. When there is no drug effect on survival, unadjusted methods will generally give the same results.

As a general approach, most pharmaceutical statisticians begin by testing for the presence of a dose-related trend in tumor proportions. If the trend test is significant, that is, the $p$ value is less than or equal to 0.05, pairwise comparisons are performed between the treated and control groups. Trend and pairwise analyses may be adjusted for mortality as stated earlier, or performed without mortality adjustment using such simple methods as chi-square or Fisher's exact tests.

Although in most cases the use of trend tests is appropriate since most biological responses are dose related, there are exceptions to this rule. Certain drugs, especially those with hormonal activity, may not produce classical dose responses and may even induce inverse dose–response phenomena. In these cases, a pairwise comparison may be appropriate in the absence of a significant positive trend.

Most (70%) pharmaceutical companies use one-tailed comparisons, and a substantial number use two-tailed methods [15]. Since regulatory agencies are primarily interested in identifying carcinogenic drugs, as opposed to those that inhibit carci-

nogenesis, the use of one-tailed tests is generally considered more appropriate. Some companies prefer two-tailed comparisons because, in the absence of a true carcinogenic effect, there is an equal probability of seeing significant decreases as well as significant increases by chance alone.

One of the most important statistical issues in the analysis of carcinogenicity data is the frequency of "false positives" or Type I errors. Because of the multiplicity of tumor sites examined and the number of tests employed, there is concern that non-carcinogenic drugs may be erroneously declared carcinogens. If a $p < 0.05$ increase in tumor incidence is automatically regarded as a biologically meaningful result, then the false-positive rate may be as high as 47–50% [33].

Several statistical procedures designed to correct for the multiplicity of significance tests have been published (and reviewed by Haseman [46]). One approach to the problem of multiple tumor site/type testing is a procedure attributed to Tukey by Mantel [47]. This method is used to adjust a calculated $p$ value based on the number of tumor types/sites for which there are a minimum number of tumors in the particular study. The reasoning here is that, for tumor sites, the number of tumors found is so small that it is impossible to obtain a significant result for that tumor site no matter how the tumors might have been distributed among the dose groups. Only those sites for which a minimum number of tumors are present can contribute to the false-positive rate for a particular study.

A method proposed by Schweder and Spjotvoll [48] is based on a plot of the cumulative distribution of observed $p$ values. Farrar and Crump [49] have published a statistical procedure designed not only to control the probability of false-positive findings but also to combine the probabilities of a carcinogenic effect across tumor sites, sexes, and species.

Another approach to controlling the false-positive rate in carcinogenicity studies was proposed by Haseman [50]. Under this "rule," a compound would be declared a carcinogen if it produced an increase that was significant at the 1% level in a common tumor or an increase that was significant at the 5% level in a rare tumor. A rare neoplasm was defined as a neoplasm that occurred with a frequency of less than 1% in control animals. The overall false-positive rate associated with this decision rule was found to be not more that 7–8%, based on control tumor incidences from NTP studies in rats and mice. This false-positive rate compares favorably with the expected rate of 5%, which is the probability at which one would erroneously conclude that a compound was a carcinogen. The method is notable for its simplicity and deserves serious consideration by pharmaceutical statisticians and toxicologists. Without resorting to sophisticated mathematics, this method recognizes the fact that tumors differ in their spontaneous frequencies and therefore in their contribution to the overall false-positive rates in the carcinogenicity studies. False-positive results are much less likely to occur at tissue sites with low spontaneous tumor incidences than at sites with high frequencies.

As a final point that has special relevance to pharmaceutical carcinogenicity studies, one may question whether the corrections for multiple comparisons and their effect on the overall false-positive rate are appropriate for all tumor types. For example, if a compound is known to bind to receptors and produce pharmacological effects in a certain organ, is it justified to arbitrarily correct the calculated $p$ value for the incidence of tumors in that organ, using the methods described above? It is difficult to justify such a correction considering that the basis for correcting the

calculated $p$ value is that the true probability of observing an increased incidence of tumors at any site by chance alone may be much higher than the nominal alpha level (usually 0.05). It is reasonable to expect that, when a drug has known pharmacological effects on a given organ, the probability of observing an increased tumor incidence in that organ by chance alone is unlikely to be higher than the nominal 5% alpha level.

Although most pharmaceutical statisticians and toxicologists agree on the need to control the probability of false-positive results, there is no consensus as to which method is most appropriate or most acceptable to regulatory agencies. The FDA and other such agencies will accept a variety of statistical procedures but will often reanalyze the data and draw their own conclusions based on their analyses.

## 12 TRANSGENIC MOUSE MODELS

Sine the early 1970s, the standard for adequate evaluation of the carcinogenic potential of a candidate pharmaceutical has been the conduct of lifetime, high dose assays in two species—almost always the rat and the mouse.

The relevance (and return on investment) for the bioassays preformed in mice has been questioned for some time. In 1997, the ICH opened the possibility for the substitution of some form of short- or medium-term mouse test as an alternative for the traditional lifetime mouse bioassay. The FDA has subsequently stated that it would accept "validated" forms of a set of medium-term mouse studies based on transgenic models, and significant effort has since gone into such validation.

The huge advances made in molecular biology since the late 1980s have provided the possibility of evaluating chemicals and potential drugs for carcinogenic potential by approaches that are different, less expensive, and take a shorter period of time than traditional long-term bioassays. This work has also been stimulated by dissatisfaction with the performance of traditional test systems.

The traditional long-term bioassays use highly inbred animals, developed with the goal of reducing the variability in background tumor incidences as a means of increasing the statistical sensitivity of the bioassays. This inbreeding has led to narrowing of the allelic pool in the strains of animals that are currently used for testing, as opposed to the wild-type populations (of humans) that the tests are intended to protect [51]. Transgenic models should serve to improve the identification of carcinogens by providing the gene-specific mechanistic data, minimizing the influence of spontaneous tumors and strain-specific effects, and reducing time required and cost and animal usage [52].

As it has become possible to transfer new or engineered genes to the germlines of mammals, the results have been transgenic mice that can be used in shorter term *in vivo* assays for carcinogenicity and that are also useful for research into the characterization of genotoxic events and mechanisms in carcinogenesis. By coupling reporter phenotypes (such as papillomas in the Tg·AC mouse), the task of "reading" results in test animals is made much less complex.

There are four transgenic mouse models that have been broadly evaluated—the Tg·AC, the Tg.rasH2, the TSP $p53^{+/-}$ and the $XPA^{-/-}$. Each of these has its own characteristics, and each of these merits some consideration. They are each made by either zygote injection or specific gene targeting in embryonic cells [53, 54].

### 12.1 The Tg·AC Mouse Model

This was the earliest of the models to be developed, and its use in mouse skin carcinogenicity studies was first reported in 1990. The mice have four copies of the v-H-ras oncogene in tandem on chromosome 11, and the transgene is fused with a fetal ζ-globin gene that acts as a promoter. The transgene codes for a switch protein, which is permanently "on," and this results in the mice having genetically initiated skin cancer. The application of tumor promoters to the surface of the skin causes the rapid induction of pedunculate papillomas that arise from the follicular epithelium. This is despite the fact that the transgene is not expressed in the skin, although it is present in the papillomas that form, and also in the focal follicular hyperplastic areas that are the precursors to the papillomas. In about 40% of the mice, the papillomas become malignant skin tumors—mainly squamous cell carcinomas and sarcomas.

The first assessments of this model as an alternative to traditional carcinogenicity studies were performed by the U.S. NIEHS and NTP, and the results with over 40 chemicals have been published. The majority of studies were performed by skin painting, regardless of whether the product was a dermal or systemic carcinogen. However, a good correlation was found with the known carcinogenicity of the test compounds, and both mutagenic and nonmutagenic were identified. It was found that great care had to be taken with the skin, because damage could also induce papillomas, which means that these animals cannot be identified using transponder chips. This sensitivity may also explain some of the false-positive results that have occurred with resorcinol and rotenone. Of more concern is that there have been false negatives with known carcinogens, namely, ethyl acrylate and *N*-methyl-*o*-acrylamide. The model was designed for use in the context of the two-stage model of carcinogenesis with the underlying mechanistic pathway involving specific transcription factors, hypomethylation, and cell-specific expression of the results. Along with the p53 model, this model has seen the widest use and evaluation (in terms of number of agents evaluated) so far. The carrier mouse strain employed, the FVB/N, is not commonly employed in toxicology and is prone to sound-induced seizures. It may be that the dermal route is not suitable for all systemic carcinogens, and this is the reason that in the ILSI program, both the dermal and systemic routes are being investigated in this model.

Another problem with this model was the occurrence of a nonresponder genotype to positive control agents. This was found to be attributable to a rearrangement of the ζ-globin promoter region, but it is claimed that this problem has been resolved. However, this has considerably delayed the ILSI studies with this model [52, 53]. It is already clear that the model gives a robust response to the positive control agent, 12-*o*-tetradecanoylphorbol 13-acetate (TPA).

### 12.2 The Tg.rasH2 Mouse Model

This model was developed at CIEA in Japan, and the first information about the mouse was published in 1990. The mice have five or six copies of the human H-ras proto-oncogene inserted in tandem into their genome surrounded by their own promoter and enhancer regions. This transgene has been very stabile, with no loss of responsiveness since the model was developed. The transgene codes

for a molecular switch protein in the same way as the previous model, but the transgene is expressed in all organs and tissues. Thus, the response endpoint is not primarily dermal.

The initial studies with this model revealed a rapid appearance of forestomach papillomas with *N*-methyl-*N*-nitrosourea (MNU), and this compound has already been used as the positive control agent in subsequent studies with this strain. A study duration of 6 months is sufficient to obtain a positive response, and longer periods should be avoided because the mice start to develop various spontaneous tumors, such as splenic hemangiosarcomas, forestomach and skin papillomas, lung and Harderian gland adenocarcinomas, and lymphomas. The model has a high level of constitutive expression and some spontaneous tumors even when the animals are younger. It is, however, very responsive to carcinogens—one gets a rapid onset after exposure and a higher response incidence than with the other models. The underlying mechanism is still not certain.

A large number of studies have been run in this strain in Japan in advance of the ILSI program. The model is sensitive to both mutagenic and nonmutagenic carcinogens, although cyclophosphamide and furfural have given equivocal results in each category, respectively. The majority of noncarcinogens have also been identified correctly, although again, there are a small number of compounds that have given equivocal results.

## 12.3 The TSP *p53*$^{+/-}$ Mouse Model

The TSP *p53*$^{+/-}$ (hereafter referred to as the *p53*—the designation of the tumor suppressor gene involved) is a heterozygous knockout with (up to seven or so months of age) a low spontaneous tumor incidence. It is responsive to the genotoxic carcinogens by a mechanism based on the fact that many (but not all) tumors show a loss of the wild-type allele. The *p53* has been extensively worked on by Tennant's group at NIEHS [55, 56]. This model was developed in the United States and carries a hemizygous knockout of the *p53* gene, which was developed by integrating a mutated copy of the gene into the genome of mice. The *p53* gene is known as a tumor-suppressor gene, and it is the most commonly mutated gene in human malignancies. It searches for a protein transcription factor, which activates multiple genes when damage to DNA strands occurs, and this in turn leads to either the arrest of the cell cycle while DNA repair occurs, or to apoptosis (programmed cell death), which removes the damaged cell. The heterozygote is used because homozygotes show a very high incidence of spontaneous tumors within a few months of birth. The heterozygotes have a low background incidence of tumors up to 12 months, but during this time there is a high chance of a second mutagenic event occurring—following exposure to a carcinogen, for example—and this would result in a loss of suppressor function or an increase in transforming activity.

The initial studies with this model as an alternative in traditional carcinogenicity testing were performed at the U.S. NIEHS, and these suggested that it was sensitive to mutagenic carcinogens such as benzene and *p*-cresidine within 6 months. Nonmutagenic carcinogens were negative in the assay, as were mutagenic noncarcinogens. However, subsequent studies and some parts of the ILSI program have shown clear indications that a 6 month duration may be insufficient. In particular, benzene

has given negative or equivocal results within 6 months, although positive results have been obtained by extending the study to 9 months. It will be very important to assess the results of the ILSI program when deciding the best study duration for this model. This is the most popular model in the United States.

## 12.4 The XPA$^{-/-}$ Mouse Model

This was the last of the models to be developed and was created using a knockout technique after the *XPA* gene had been cloned. The first data were published by RIVM in The Netherlands in 1995. Both alleles of the *XPA* gene have been inactivated by a homologous recombination in ES cells, resulting in a homozygous deletion of the gene spanning exons 3 and 4. The protein coded by this gene is essential for the detection and repair of DNA damage, using the nucleotide excision repair (*NER*) pathway. This model only has between 2% and 5% of residual *NER* activity.

The initial studies at RIVM demonstrated that exposure of these mice to IV-B radiation or 7,12-dimethylbenz[*a*]anthracene resulted in the rapid induction of skin tumors. It was also shown that various internal tumors could be induced following oral administration of mutagenic carcinogens such as benzo[*a*]pyrene (B[*a*]P) and 2-acetylaminofluorine (2-AAF). The early studies suggested that this response could occur within 6 months, but further experience has indicated that a 9 month treatment period is essential in order to obtain a response with positive control agents such as B[*a*]P,2-AAF, and *p*-cresidine.

It is generally proposed that while such models can improve the identification of carcinogens in three ways (providing gene-specific mechanistic data, minimizing the influence of spontaneous tumors and strain-specific effects, and reducing the time, cost, and animal usage involved), they have two potential uses in pharmaceutical development. These are either in lieu of the mouse 2 year cancer bioassay or in subchronic toxicity assessments prior to making a decision to commit to a pair of 2 year carcinogenicity bioassays.

As performance data has become available on these strains, the ICH [8] has incorporated their use into pharmaceutical testing guidelines in lieu of the second rodent species tests (i.e., to replace the long-term mouse bioassay when the traditional rat study has been performed). The FDA has stated that it would accept such studies when "performed in a validated mode." In fact, the CBER has accepted such studies as a sole carcinogenicity bioassay in some cases where there was negative traditional genotoxicity data and strong evidence of a lack of a mechanistic basis for concern.

A joint ILSI and HESI validation program has been completed, looking at the results of the four prime candidate models in identifying carcinogens as compared to the results of traditional long-term rodent bioassays. This validation program involved 51 different laboratories and imposed protocol standards to allow comparison of results. Three dose levels were studied per chemical, with 15 males and 15 females being used for each dose group. A vehicle and high dose control in wild-type animals was also included, with information from NTP bioassays and 4 week range-finding assays being used to help set doses. Animals were dosed for 26 weeks. The issues in and coming out of these validation programs bear consideration [56].

- Is human or rodent bioassay data the proper comparator data for evaluating performance? It should be kept in mind that there are sets of rodent bioassay data (particularly those involving liver tumors in mice) that are widely accepted as irrelevant in the prediction of human risk.
- How will the data from these assays be incorporated into any weight-of-evidence approach to assessing human health risk?
- What additional mechanistic research needs to be undertaken to improve our understanding of the proper incorporation and best use of the data from these assays?
- How can the results of current validation efforts be best utilized in the timely evaluation of the next generation of assays?
- Given that, at least under some conditions, assays using these models tend to "blow up" (have high spontaneous tumor rates) once the animals are more than 8 or 9 months of age, how critical are age and other not currently apprehended factors to optimizing both sensitivity and specificity?
- How wide and unconditional will FDA (and other regulatory bodies) acceptance be of these models in lieu of the traditional 2 year mouse bioassay?

## 13 INTERPRETATION OF RESULTS

### 13.1 Criteria for a Positive Result

There are three generally accepted criteria for a positive result in a carcinogenicity study. The first two are derived directly from the results of the statistical analysis: (1) a statistically significant increase in the incidence of a common tumor and (2) a statistically significant reduction in the time to tumor development. The third criterion is the occurrence of very rare tumors, that is, those not normally seen in control animals, even if the incidence is not statistically significant.

### 13.2 Statistical Analysis

The actual statistical techniques used to evaluate the results of carcinogenicity bioassays basically utilize four sets of techniques, three of which have been presented earlier in this book. These methods are *exact tests, trend tests, life tables* (such as log rank techniques), and *Peto analysis*. These are then integrated into the decision-making schemes discussed earlier in this chapter. The methods themselves and alternatives are discussed elsewhere in detail [11, 57].

***Exact Tests*** The basic forms of these (the Fisher exact test and chi-squared) have previously been presented, and the reader should review these. Carcinogenicity assays are, of course, conducted at doses that are at least near those that will compromise mortality. As a consequence, one generally encounters competing toxicity producing differential mortality during such a study. Often, particularly with certain agricultural chemicals, latency of spontaneous tumors in rodents may shorten as a confounded effect of treatment with toxicity. Because of such happenings, simple tests on proportions, such as $\chi^2$ and Fisher–Irwin exact tests on contingency tables,

**TABLE 4 Trend Versus Heterogeneity**

| Number at Risk | Number with Tumor | Dose Level |
|---|---|---|
| 50 | 2 | 0 |
| 50 | 4 | 1 |
| 50 | 6 | 2 |
| 50 | 7 | 3 |

may not produce optimal evaluation of the incidence data. In many cases, however, statisticians still use some of these tests as methods of preliminary evaluation. These are unadjusted methods without regard for the mortality patterns in a study. Failure to take into account mortality patterns in a study sometimes causes serious flaws in interpretation of the results. The numbers at risk are generally the numbers of animals histopathologically examined for specific tissues.

Some gross adjustments on the numbers at risk can be made by eliminating early deaths or sacrifices by justifying that those animals were not at risk to have developed the particular tumor in question. Unless there is dramatic change in tumor prevalence distribution over time, the gross adjusted method provides fairly reliable evidence of treatment effect, at least for nonpalpable tissue masses.

***Trend Tests*** Basic forms of the trend tests (such as that of Tarone [58]) have previously been presented in this text.

Group comparison tests for proportions notoriously lack power. Trend tests, because of their use of prior information (dose levels), are much more powerful. These are illustrated in Table 4. Also, it is generally believed that the nature of true carcinogenicity (or toxicity for that matter) manifests itself as dose response. Because of these facts, evaluation of trend takes precedence over group comparisons. In order to achieve optimal test statistics, many people use ordinal dose levels $(0, 1, 2, \ldots)$ instead of the true arithmetic dose levels to test for trend. However, such a decision should be made *a priori*. The following example demonstrates the weakness of homogeneity tests.

COCHRAN–ARMITAGE TEST FOR TREND

| | Calculated $\chi^2$ Subgroup | DF | Alpha | Two-Tail Probability |
|---|---|---|---|---|
| Trend | 3.3446 | 1 | 0.0500 | 0.0674 |
| Departure | 0.0694 | 2 | 0.0500 | 0.9659 |
| Homogeneity | 3.4141 | 3 | 0.0500 | 0.3321 |

ONE-TAIL TESTS FOR TREND

| Type | Probability |
|---|---|
| Uncorrected | 0.0337 (direction = +) |
| Continuity corrected | 0.0426 (direction = +) |
| Exact | 0.0423 (direction = +) |

MULTIPLE PAIRWISE GROUP COMPARISONS BY FISHER–IRWIN EXACT TEST

| Groups compared | Alpha | One-Tail Probability |
|---|---|---|
| 1 vs. 2 | 0.0500 | 0.33887 |
| 2 vs. 3 | 0.0500 | 0.13433 |
| 1 vs. 4 | 0.0500 | 0.07975 |

As is evident from this example, often group comparison tests will fail to identify significant treatment but a trend test will. The same arguments apply to survival adjusted tests on proportions as well. In an experiment with more than one dose group ($K > 1$), the most convincing evidence for carcinogenicity is given by tumor incidence rates that increase with increasing dose. A test designed specifically to detect such dose-related trends is Tarone's [52] trend test.

Letting $\mathbf{d} = (O, d_1, d_2, \ldots, d_k)^T$ be the vetor of dose levels in *all* $K + 1$ groups and letting

$$(\mathbf{O} - \mathbf{E}) = (O_0 - E_0, \ldots, O_k - E_k)^T \quad \text{and} \quad \mathbf{V} = \begin{pmatrix} \mathbf{V}_{00} & \cdots & \mathbf{V}_{0K} \\ \vdots & & \vdots \\ \mathbf{V}_{K0} & \cdots & \mathbf{V}_{KK} \end{pmatrix}$$

contain elements as described in the previous section but for *all* $K + 1$ *groups*, the trend statistic is given by

$$\mathbf{X}_T^2 = \frac{[d^T(\mathbf{O} - \mathbf{E})]^2}{d^T \mathbf{V} \mathbf{d}}$$

The statistic $\mathbf{X}_T^2$ will be large when there is evidence of a dose-related increase or decrease in the tumor incidence rates, and small when there is little difference in the tumor incidence between groups or when group differences are not dose related. Under the null hypothesis of no differences between groups, $\mathbf{X}_T^2$ has approximately a chi-squared distribution with one degree of freedom.

Tarone's trend test is most powerful at detecting dose-related trends when tumor onset hazard functions are proportional to each other. For more power against other dose-related group differences, weighted versions of the statistic are also available; see Breslow [59] or Crowley and Breslow [60] for details.

These tests are based on the generalized logistic function [61]. Specifically, one can use the Cochran–Armitage test (or its parallel, the Mantel–Haenszel version) for monotonic trend as the heterogeneity test.

***Life-Table and Survival Analysis*** These methods are essential when there is any significant degree of mortality in a bioassay. They seek to adjust for the differences in periods of risk that individual animals undergo. Life-table techniques can be used for those data where there are observable or palpable tumors. Specifically, one should use Kaplan–Meier product limit estimates from censored data graphically,

Cox–Tarone binary regression (log-rank test), and Gehan–Breslow modification of Kruskal–Wallis tests [62] on censored data.

The Kaplan–Meier estimates produce a step function for each group and are plotted over the lifetimes of the animals. Planned, accidentally killed, and lost animals are censored. Moribund deaths are considered to be treatment related. A graphical representation of Kaplan–Meier estimates provide excellent interpretation of survival adjusted data except in the cases where the curves cross between two or more groups. When the curves cross and change direction, no meaningful interpretation of the data can be made by any statistical method because the proportional odds characteristic is totally lost over time. This would be a rare case where treatment initially produces more tumor or death and then, due to repair or other mechanisms, becomes beneficial.

For Cox–Tarone binary regression [58, 62], censored survival and tumor incidence data are expressed in a logistic model as dose over time. The log-rank test [63] is based on the Weibull distribution, and the Mantel–Haenszel test [64] is very similar to this test when there are no covariates or stratifying variables in the design. The logistic regression based Cox–Tarone test is preferable because one can easily incorporate covariates and stratifying variables, which one cannot do in the IARC methods.

Gehan–Breslow modification of the Kruskal–Wallis test is a nonparametric test on censored observations. It assigns more weight to early incidences compared to the Cox–Tarone test.

For survival adjusted tests on proportions, as mentioned earlier, instead of having a single $2 \times k$ table, one has a series of such $2 \times k$ tables across the entire lifetime of the study. The numbers at risk for such analyses will depend on the type of tumor one is dealing with:

1. *Palpable or Lethal Tumors*. Number at risk at time $t$ = number of animals surviving at the end of time $t - 1$.
2. *Incidental Tumors*. Number at risk at time $t$ = number of animals that either died or were sacrificed whose particular tissue was examined histopathologically.

The methods of analyzing the incidences, once the appropriate numbers at risk are assigned for these tumors, are rather similar, either binary regression based or by pooling evidence from individual tables [43].

***Peto Analysis*** The Peto method of analysis of bioassay tumor data is based on careful classification of tumors into five different categories, as defined by the IARC.

1. Definitely Incidental.
2. Probably Incidental.
   *Comment:* Combine (1) and (2).
3. Probably Lethal.
4. Definitely Lethal.
   *Comment:* These categories may be combined into one (otherwise it requires a careful cause of death determination).

5. Mortality Independent (such as mammary, skin, and other observable or superficial tumors).

*Interval Selection for Occult (Internal Organ) Tumors*

1. FDA: 0–50, 51–80, 81–104 Weeks, Interim Sacrifice, Terminal Sacrifice.
2. NTP: 0–52, 53–78, 79–92, 93–104 Weeks, Interim Sacrifice, Terminal Sacrifice.
3. IARC: *Ad hoc* Selection Method [41].

   *Comment*: Any of the above may be used. There are problems with the IARC selection method, however, Two sexes or two or more strains will have different intervals for the same compound. Different interval selection methods will produce different statistical significance levels. This may produce bias and requires an isotonic tumor prevalence for ready analysis.

*Logistic Regression Method for Occult (Internal Organ) Tumors [65]*

Tumor prevalence is modeled as a logistic function of dose and polynomial in age.

*Comment:* The logistic tumor prevalence method is unbiased, requires maximum likelihood estimation, and allows for covariates and stratifying variables. It may be time consuming and have convergence problem with sparse tables (low tumor incidences) and clustering of tumors.

### 13.3 Methods to Be Avoided

The following methods and practices should be avoided in evaluation of carcinogenicity:

1. Use of only the animals surviving after 1 year in the study.
2. Use of a two-strata approach—separate analyses for animals killed during the first year of the study and the ones thereafter.
3. Exclusion of all animals in the study that died on test and analysis of only the animals that are sacrificed at the end of the study.
4. Exclusion of interim sacrifice animals from statistical analyses.
5. Evaluation of number of tumors of all sites as opposed to the number of animals with tumors for specific sites of specific organs.

Another issue is subjectivity in slide reading by most pathologists who do not want to read them in a coded fashion, whereby they will not know the dose group an animal is coming from. This is not under statisticians' control but they should be aware of that in any case.

Often a chemical being tested is both toxic as well as potentially carcinogenic. When competing toxicity causes extreme differences in mortality or there is a clustering effect in tumor prevalence in a very short interval of time, none of the adjusted methods works. One then must use biological intuition to evaluate the tumor data.

Use of historical control incidence data for statistical evaluation is controversial. There are too many sources of variation in these data. For example, different pathologists use different criteria for categorizing tumors (in fact, the same pathologist may change opinion over time); there is laboratory-to-laboratory variation; there may be genetic drift over time; location of suppliers may make a difference; and finally, these data are not part of the randomized concurrent control. Regulatory agencies and pathologists generally use these data for qualitative evaluation. My personal view is that that is where they belong.

## 13.4 Use of Historical Controls

When the study is over, the data analyzed, and the $p$ values corrected, as appropriate, one may find that one or more tumor types increased in drug-treated groups relative to concurrent controls. Although the FDA and other regulatory agencies play down the importance of historical control data, it is common practice in the pharmaceutical industry to use historical data in the interpretation of tumor findings. The first and most appropriate comparison of a treated group is with concurrent control group(s), but it is of interest to see how tumor incidences in the treated groups compare with the historical incidence, and such a comparison is an accepted practice in toxicology and biostatistics [42, 66, 67]. A treated group may have a tumor incidence significantly higher than that of the concurrent control groups, but comparable to or lower than the historical incidence. Occasionally, a small number of tumors may be found in a treated group and the incidence may be significant because of the absence of this tumor in the concurrent controls. Review of appropriate historical control data may reveal that the low tumor incidence in the treated group is within the "expected" range for this tumor.

The role of historical control data in interpreting carcinogenicity findings depends on the "quality" of the historical data. Ideally, the data should be derived from animals of the same age, sex, strain, and supplier, housed in the same facility, and the pathology examinations should have been performed by the same pathologist or using the same pathological criteria for diagnosis. Since genetic drift occurs even in animals of a given strain and supplier, recent data are more useful than older data. The value of historical control data is directly proportional to the extent to which these conditions are fulfilled.

Although methods are available for including historical control data in the formal statistical analysis [68, 69], this is usually not done and for good reason. The heterogeneity of historical data requires that they be used qualitatively and selectively to aid in the final interpretation of the data, after completion of the formal statistical analysis. Table 5 presents a summary of background tumor incidences for the most commonly employed rodent strains.

## 13.5 Relevance to Humans

After statistical analyses have been performed and historical data consulted, the final interpretation may be that a drug appears to cause tumors at one or more tissue sites in the mouse or the rat. But what does this mean for the species to which the drug will be administered, namely, the human? Extrapolation of rodent carcinogenicity data to humans remains one of the greatest challenges of modern toxicol-

**TABLE 5 Comparative Percent Incidence of Pertinent Neoplasia in Different Strains of Rats and Mice (104 Weeks Old)**

| | F344 Rats | | S-D Rats | | Wistar Rats | | $B6C3F_1$ Mice | | CD-1 Mice | |
|---|---|---|---|---|---|---|---|---|---|---|
| Types of Neoplasia | Males | Females | Males | Females | Males | Females | Males | Females | Males | Females |
| Hepatocellular adenoma | 4 | <1 | 5 | <1 | 1 | 2 | 29 | 30 | 26 | 5 |
| Hepatocellular carcinoma | 2 | 0 | 2 | 0 | <1 | <1 | 26 | 16 | 10 | 1 |
| Pancreas islet adenoma | 12 | 2 | 8 | 9 | 4 | 2 | 2 | 0 | <1 | <1 |
| Pancreas islet carcinoma | 3 | 0 | <1 | 5 | <1 | <1 | 0 | 0 | 0 | 0 |
| Pancreas acinar adenoma | 6 | 0 | 1 | 0 | 13 | <1 | 2 | 0 | <1 | 0 |
| Pheochromocytoma | 21 | 4 | 23 | 5 | 10 | 2 | 0 | 2 | <1 | <1 |
| Adrenocortical adenoma | 0 | 2 | 3 | 0 | 8 | 9 | <1 | 0 | 1 | <1 |
| Pituitary adenoma | 49 | 42 | 62 | 85 | 34 | 55 | 2 | 8 | 0 | 5 |
| Thyroid C-cell adenoma | 17 | 8 | 7 | 6 | 6 | 8 | 0 | 0 | 0 | 0 |
| Thyroid follicular adenoma | 0 | 0 | 4 | 2 | 2 | 1 | 2 | 6 | 1 | <1 |
| Mammary gland fibroadenoma | 4 | 57 | 2 | 54 | 3 | 36 | 0 | 0 | <1 | 1 |
| Mammary gland carcinoma | 0 | 4 | <1 | 26 | 1 | 13 | 0 | 0 | 0 | 6 |
| Skin fibroma | 10 | 2 | 2 | <1 | 5 | 1 | 1 | 2 | <1 | <1 |
| Skin papilloma | 6 | 0 | 2 | 0 | 2 | <1 | 0 | 0 | <1 | 0 |
| Pulmonary adenoma | 4 | 4 | <1 | <1 | <1 | 0 | 22 | 6 | 15 | 15 |
| Preputial gland neoplasia | 10 | NA | >1 | NA | <1 | NA | <1 | NA | <1 | NA |
| Leydig cell neoplasia | 89 | NA | 7 | NA | 11 | NA | 0 | NA | 1 | NA |
| Clitoral gland neoplasia | NA | 14 | NA | <1 | NA | <1 | NA | <1 | NA | 0 |
| Uterine polyps | NA | 14 | NA | 6 | NA | 16 | NA | 1 | NA | <1 |
| Ovarian neoplasia | NA | 6 | NA | 1 | NA | 8 | NA | 6 | NA | 1 |
| Mononuclear cell leukemia | 62 | 42 | 0 | 0 | <1 | <1 | 0 | 0 | 2 | 2 |
| Lymphoma | 0 | 0 | 2 | 1 | 3 | 5 | 14 | 24 | 8 | 22 |
| Forestomach papilloma | 0 | 2 | <1 | <1 | 0 | <1 | 4 | 2 | <1 | <1 |
| Scrotal mesothelioma | 5 | NA | 1 | NA | 2 | NA | 0 | NA | 0 | NA |

*Note*: F344, Fischer 244 rats; S-D, Sprague–Dawley rats; B6C3F1, mice, (C57BL/6N+C3H/HeN)F1; CD-1, 1CRCr: CD-1 mice; NA, nonapplicable; the average number used by species/strain/gender was in excess of 750 animals.

ogy. There is no simple formula, and each case must be evaluated on its own merits. Very generally speaking, the FDA and other major regulatory agencies consider compounds that are tumorigenic in one or more animal species to be "suspect" tumorigens in humans. The actual impact of this conclusion on the approval of a drug depends on the target population and the indication. For example, even a suspicion of carcinogenic activity may be fatal for a potential contraceptive drug intended for use in a very large population of healthy people. By contrast, clear evidence of carcinogenic activity may be overlooked in a drug being considered for use in a restricted population with a life-threatening disease.

Regardless of the target population and indication, the FDA and other agencies have, in recent years, attempted to consider the mechanism of tumor induction in rodents and its relevance for humans. If a drug is known to cause tumors in a rodent via a mechanism that does not exist in humans, the importance of the tumor findings may be markedly reduced. For example, drugs that cause tumors by a secondary hormonal mechanism shown to be inapplicable to humans may be given special consideration. It is the sponsor's responsibility to provide pertinent data on the mechanism of tumor induction and its relevance, or irrelevance, for humans. If the sponsor can show that an apparently drug-related tumor is species specific, the importance of the tumor in the overall evaluation of the drug will be greatly minimized. Table 5 presents a list of neoplastic /tumorigenic responses seen in rodents, which have limited relevance to human safety. Part of the consideration must also be a recognition of the main characteristics of nongenotoxic carcinogens. These are recognized to be dose-dependent responses with operative thresholds. The major characteristics [14] are:

- Specificity (of species, sex, and organ).
- A threshold is operative and must be exceeded for cell proliferation and tumor development to occur.
- There is a step wise dose–response curve/relationship between exposure, cell proliferation, and tumor development.
- The response is reversible with a cessation of dosing unless a point of no return has been passed.

Not all tumors seen in animals are revelant to predicting human risk, particularly these in rodents. Table 6 summarizes such situations.
Interpretation of the data and results from a carcinogenicity study is generally not simply a yes or no (or counting turnovers) matter. Rather, one must also consider effects on survival. An approach to this is shown in Table 7.

## 14 CONCLUSION

The design, conduct, and interpretation of carcinogenicity studies is one of the major challenges for the pharmaceutical toxicologist, pathologist, biostatistician, and regulator. This is a rapidly changing field generating more questions than answers. The biggest question continues to be the extrapolation of rodent data to humans,

**TABLE 6 Examples of Neoplastic Effects in Rodents with Limited Significance for Human Safety**

| Neoplastic Effect | Pathogenesis (Agents) |
|---|---|
| Renal tubular neoplasia in male rats | $\alpha_2\mu$-Globulin nephropathy/hydrocarbons (*d*-limonene, *p*-dichlorobenzene) |
| Hepatocellular neoplasia in rats and mice | Peroxisome proliferation (clofibrate, phthalate esters, phenoxy agents)<br>Phenobarbital-like promotion |
| Urinary bladder neoplasia in rats | Crystalluria, carbonic anhydrase inhibition, urine pH extremes, melamine, saccharine, carbonic anhydrase inhibitors, dietary phosphates |
| Hepatocellular neoplasia in mice | Enzymatic-metabolic activation (in part unknown)/phenobarbital-like promotion |
| Thyroid follicular cell neoplasia in rats | Hepatic enzyme induction, thyroid enzyme inhibition/axazepam, amobarbital, sulfonamides, thioureas |
| Gastric neuroendocrine cell neoplasia mainly in rats | Gastric secretory suppression, gastric atrophy induction (climetidine, omeprazole, butachlor) |
| Adenohypophysis neoplasia in rats | Feedback interference/neuroleptics (dopamine inhibitors) |
| Mammary gland neoplasia in female rats | Feedback interference/neuroleptics, antiemetics, antihypertensives (calcium channel blockers), serotonin agonists, anticholinergics, exogenous estrogens |
| Pancreatic islet cell neoplasia in rats | Feedback interference/neuroleptics |
| Harderian gland neoplasia in mice | Feedback interference/misoprostol ($PGE_1$), nalidixic acid, aniline dyes |
| Adrenal medullary neoplasia in rats | Feedback interference (lactose, sugar alcohols) |
| Forestomach neoplasia in rats and mice | Stimulation of proliferation/butylated hydroxyanisole, phthalate esters, proprionic acid |
| Lymphomas in mice | Immunosuppression/cyclosporine |
| Mononuclear cell leukemia in rats (mainly F344) | Immunosuppression (in part unknown)/furan, iodinated glycerol |
| Splenic sarcomas in rats | Methemoglobinemia (in part unknown)/dapsone |
| Osteomas in mice | Feedback interference/lactose, sugar alcohols, H2 antagonists, carbamazepine, vidarabine, isradipine, dopaminergics, finasteride |
| Leydig cell testicular neoplasia in mice | Feedback interference (proestrogens, finasteride, methoxychlor, cadmium) |
| Endometrial neoplasia in rats | Feedback interference (proestrogens, dopamine agonists) |
| Uterine leiomyoma in mice | Feedback interference ($\beta_1$-antagonists) |
| Mesovarial leiomyoma in rats (occasionally in mice) | Feedback interference ($\beta_2$-agonists) |

**TABLE 7 Interpretation of the Analysis of Tumor Incidence and Survival Analysis (Life Table)**

| Outcome Type | Tumor Association with Treatment[a] | Mortality Association with Treatment | Interpretation[b] |
|---|---|---|---|
| A | – | + | Unadjusted test may underestimate tumorigenicity of treatment. |
| B | + | + | Unadjusted test gives valid picture of tumorigenicity of treatment. |
| C | + | – | Tumors found in treated groups may reflect longer survival of treated groups. Time adjusted analysis is indicated. |
| D | – | + | Apparent negative findings on tumors may be due to the shorter survival in treated groups. Time adjusted analysis and/or a retest at lower doses is indicated. |
| E | – | 0 | Unadjusted test gives a valid picture of the possible tumor-preventive capacity of the treatment. |
| F | – | – | Unadjusted test may underestimate the possible tumor-preventive capacity of the treatment. |
| G | 0 | + | High mortality in treated groups may lead to unadjusted test missing a possible tumorigen. Adjusted analysis and/or retest at lower doses is indicated |
| H | 0 | 0 | Unadjusted test gives a valid picture of lack of association with treatment. |
| I | 0 | – | Longer survival in treated groups may mask tumor-preventive capacity of treatment. |

[a]+, Yes; –, No; and 0, No bearing on discussion.
[b]The unadjusted test referred to here is a contingency table type of analysis of incidence, such as Fisher's exact test.

especially when data on mechanisms of tumor induction are unavailable or controversial. Much has been written on the difficulties inherent in extrapolating results from rodents treated with MTDs of a compound to humans who will be exposed to much lower doses and often for shorter periods. A discussion of these and other aspects of carcinogenic risk assessment is beyond the scope of this chapter.

Regulatory agencies are very aware of these challenges and deserve credit for attempting to respond to changes in the state of knowledge, while still discharging their responsibility to protect the public health. For example, the latest version of the Japanese guidelines [27] acknowledges that the highest dose in a carcinogenicity study may be set at 100 times the clinical dose, instead of requiring that the MTD be achieved. It is also noteworthy that the FDA Center for Drug Evaluation has announced the formation of a Carcinogenicity Assessment Committee representing all drug review divisions. This group will advise all the divisions on issues related to carcinogenicity. Creation of such a group reflects the importance that the FDA places on carcinogenicity data in evaluating the safety of new drug candidates.

## REFERENCES


1. NIH *9th Report on Carcinogens*. Washington, DC: US Department of Health and Human Services; 2000.
2. Miller EC, Miller JA. Mechanisms of chemical carcinogenesis. *Cancer* 1981;47:1055–1064.
3. Powell CJ, Berry CL. Non-genotoxic or epigenetic carcinogenesis. In Ballantyne B, Marrs T, Syversen T, Eds. *General and Applied Toxicology*, 2nd ed. London: MacMillan; 2000.
4. Kitchin KT. *Carcinogenicity*. New York: Marcel Dekker; 1999.
5. Williams GM, Iatropoulos MJ. and Enzhmann HG. Principles of testing for carcinogenic activity. In Hayes AW, Ed. *Principles and Methods of Toxicology*, 5th ed. Philadelphia: Taylor and Francis; 2007.
6. International Agency for Research on Cancer. IARC Monographs on the Evaluation of Carcinogenic Risks to Humans—Preamble. *IARC Internal Technical Report 87/001*. Lyon: IARC; 1987.
7. *The Need for Long-Term Rodent Carcinogenicity Studies of Pharmaceuticals*. ICH; 1996.
8. *Testing for Carcinogenicity of Pharmaceuticals*. ICH; 1997.
9. *Dose Selection for Carcinogenicity Studies of Pharmaceuticals*. ICH; 1997.
10. *CPMP Note for Guidance on Carcinogenic Potential*. EMEA; July 2002.
11. Weaver RJ, Brunden MN. The design of long-term carcinogenicity studies. In Chow S, Lide J, Eds. *Design and Analysis of Animal Studies in Pharmaceutical Development*. New York: Marcel Dekker; 1998.
12. FDA. *Guidance for Industry: Statistical Aspects of the Design, Analysis and Interpretation of Chronic Rodent Carcinogenicity Studies of Pharmaceuticals*. Washington, DC: USDHEW; 2001.
13. Contrera JF, Jacobs AC, DeGeorge JJ. Carcinogenicity testing and the evaluation of regulatory requirements for pharmaceuticals. *Regul Toxicol Pharmacol* 1997;25:130–145.
14. Spindler P, Loan J, Ceuppens P, Harling R, Eittlin R, Lima BS. Carcinogenicity testing of pharmaceuticals in the European Union: a workshop report. *Drug Inf J* 2000; 34:821–828.
15. Pharmaceutical Manufacturers Association. *Results of a Questionnaire Involving the Design of and Experience with Carcinogenicity Studies*. PMA; 1988. This document has not been published in the open literature but is widely available within the pharmaceutical industry. It may be obtained by writing to the Pharmaceutical Manufacturers Association.

16. Rao GN, Birnbaum LS, Collins JJ, Tennant RW, Skow LC. Mouse strains for chemical carcinogenicity studies: overview of a workshop. *Fundam Appl Toxicol* 1988;10:385–394.
17. Purchase IFH. Interspecies comparisons of carcinogenicity. *Br J Cancer* 1980;41: 454–468.
18. Haseman JK, Crawford DD, Huff JE, Boorman GA, McConnell EE. Results for 86 two-year carcinogenicity studies conducted by the National Toxicology Program. *J Toxicol Environ Health* 1984;14:621–639.
19. Gold LS, Sawyer CB, Magaw R, Blackman G, deVeciana M, Levenson R, Hooper NK, Havender WR, Bernstein L, Peto R, Pike MC, Ames BN. A carcinogenic potency data base of the standardized results of animal bioassays. *Environ Health Perspect* 1984;58:9–319.
20. Sher SP, Jensen RD, Bokelman DL. Spontaneous tumors in control F344 and Charles River CD rats and Charles River CD-1 and B6C3F1 mice. *Toxicol Lett* 1982;11: 103–110.
21. Food and Drug Administration (FDA). Toxicological principles for the safety assessment of direct food additives and color additives used in food. In *Redbook II (draft)*. Washington, DC: Food and Drug Administration; 1993, pp 111–115.
22. Rao GN, Huff J. Refinement of long-term toxicity and carcinogenesis studies. *Fundam Appl Toxicol* 1990;15:33–43.
23. Greenman DL, Kodell RL, Sheldon WG. Association between cage shelf level and spontaneous induced neoplasms in mice. *J Natl Cancer Inst* 1984;73:107–113.
24. Wiskemann A, Sturm E, Klehr NW. Fluorescent lighting enhances chemically induced papilloma formation and increases susceptibility to tumor challenge in mice. *J Cancer Res Clin Oncol* 1986;112:141–143.
25. Bellhorn RW. Lighting in the animal environment. *Lab Anim Sci* 1980;30:440–450.
26. Greenman DL, Bryant P, Kodell RL, Sheldon W. Influence of cage shelf level on retinal atrophy in mice. *Lab Anim Sci* 1982;32:353–356.
27. Weisbrode SE, Weiss HS. Effect of temperature on benzo[*a*]pyrene induced hyperplastic and neoplastic skin lesions in mice. *J Natl Cancer Inst* 1981;66:978–981.
28. Fox JG. Clinical assessment of laboratory rodents on long term bioassay studies. *J Environ Pathol Toxicol* 1977;1:199–226.
29. Flynn RJ. Studies on the aetiology of ringtail in rats. *Proc Anim Care Panel* 1960;9:155–160.
30. DePass LR, Weil CS, Ballantyne B, Lewis SC, Losco PE, Reid JB, Simon GS. Influence of housing conditions for mice on the results of a dermal oncogenicity bioassay. *Fundam Appl Toxicol* 1986;7:601–608.
31. Rao GN, Haseman JK, Edmondson J. Influence of viral infections on body weight, survival, and tumor prevalence in Fisher 344 rats on two-year studies. *Lab Anim Sci* 1989;39:389–393.
32. Speid LH, Lumley CE, Walker SR. Harmonization of guidelines for toxicity testing of pharmaceuticals by 1992. *Regul Toxicol Pharmacol* 1990;12:179–211.
33. Haseman JK, Winbush JS, O'Donnell MW. Use of dual control groups to estimate false positive rates in laboratory animal carcinogenicity studies. *Fundam Appl Toxicol* 1986;7:573–584.
34. *Guidance for Industry: Carcinogenicity Study Protocol Submissions*. CDER; May 2002.
35. McGregor D. Carcinogenicity and genotoxic carcinogens. In Ballantyne B, Marrs T, Syversen T, Eds. *General and Applied Toxicology*, 2nd ed. London: MacMillan; 2000.

36. Japanese Ministry of Health and Welfare (MHW). *Revised Guidelines for Toxicity Studies Required for Application for Approval of Manufacturing/Importing Drugs*. Tokyo: Ministry of Health and Welfare; 1989, pp 37–48.
37. Haseman JK. Issues in carcinogenicity testing: dose selection. *Fundam Appl Toxicol* 1985;5:66–78.
38. Cameron TP, Hickman RL, Kornreich MR, Tarone RE. History, survival, and growth patterns of B6C3F1 mice and F344 rats in the National Cancer Institute Carcinogenesis Testing Program. *Fundam Appl Toxicol* 1985;5:526–538.
39. Littlefield NA, Farmer JH, Gaylor DW, Sheldon WG. Effects of dose and time in a long-term, low-dose carcinogenic study. In Staffa JA, Mehlman MA, Eds. *Innovations in Cancer Risk Assessment (ED01Study)*. Chicago, Illinois: Pathotox Publishers; 1979, pp 17–34.
40. Gaylor DW, Kodell RL. Dose–response trend tests for tumorigenesis adjusted For differences in survival and body weight across doses. *Toxicol Sci* 2001;59:219–225.
41. Peto R, Pike MC, Day NE, Gray RG, Lee PN, Parish S, Peto J, Richards S, Wahrendorf J. Guidelines for simple, sensitive significance tests for carcinogenic effects in long-term animal experiments. In *IARC Monographs on the Evaluation of the Carcinogenic Risk of Chemicals to Humans*. Lyon: International Agency for Research on Cancer; 1980.
42. Gart JJ, Chu KC, Tarone RE. Statistical issues in interpretation of chronic bioassay tests for carcinogenicity. *J Natl Cancer Inst* 1979;62:957–974.
43. Gart JJ, Krewski D, Lee PN, Tarone RE, Wahrendorf J. The design and analysis of long-term animal experiment. In *Statistical Methods in Cancer Research*, *Volume III*. IARC Scientific Publication No. 79. Lyon: International Agency for Research on Cancer; 1986.
44. Dinse GE, Lagakos SW. Regression analysis of tumor prevalence data. *Appl Stat* 1983;32:236–248.
45. Portier CJ, Bailer AJ. Testing for increased carcinogenicity using a survival-adjusted quantal response test. *Fundam Appl Toxicol* 1989;12:731–737.
46. Haseman JK. Use of statistical decision rules for evaluating laboratory animal carcinogenicity studies. *Fundam Appl Toxicol* 1990;14:637–648.
47. Mantel N. Assessing laboratory evidence for neoplastic activity. *Biometrics* 1980; 36:381–399.
48. Schweder T, Spjotvoll E. Plots of $p$-values to evaluate many tests simultaneously. *Biometrika* 1982;69:493–502.
49. Farrar DB, Crump KS. Exact statistical tests for any carcinogenic effect in animal bioassays. *Fundam Appl Toxicol* 1988;11:652–663.
50. Haseman JK. A reexamination of false-positive rates for carcinogenicity studies. *Fundam Appl Toxicol* 1983;3:334–339.
51. Festing MW. Properties of inbreed strains and outbreed stocks, with special reference to toxicity testing. *J Toxicol Environ Health* 1979;5:53–68.
52. Eastin WC, Haseman JK, Mahler JF, Bucher JR. The National Toxicology Program evaluation of genetically altered mice predictive models for identifying carcinogens. *Toxicol Pathol* 1998;26:461–584.
53. McAnulty PA. Transgenic mouse models in carcinogenicity testing. *European Pharmaceutical Contractor*; 2000.
54. French JE, Spalding JW, Dunnick JK, Tice RR, Furedi-Marchacck M, Tennant RW. The use of transgenic animals in cancer testing. *Inhalation Toxicol* 1999;11:541–544.
55. Tennant RW, French JE, Spalding JW. Identification of chemical carcinogens and assessing potential risks in short-term bioassays using transgenic mouse models. *Environ Health Perspect* 1995;103:942–950.

56. Tennant RW, Stasiewicz S, Mennear J, French JE, Spalding JW. Genetically altered mouse models for identifying carcinogens. In McGregor DB, Rice JM, Venith S, Eds. *The Use of Short and Medium-Term Tests for Carcinogens and Data on Genetic Effects in Carcinogenic Hazard Evaluation*. IARC Science Publication No. 146. Lyon, France: IARC; 1999, pp 23–148.
57. Gad SC. *Statistics and Experimental Design for Toxicologists*, 4th ed. Boca Raton, FL: CRC Press; 2007.
58. Tarone RE. Tests for trend in life table analysis, *Biometrika* 1975;62:679–682.
59. Breslow N. Comparison of survival curves. In Buyse ME, Staquet MJ, Sylvester RJ, Eds. *Cancer Clinical Trials: Methods and Practice*. Oxford, UK: Oxford University Press; 1984, pp 381–406.
60. Crowley J, Breslow N. Statistical analysis of survival data. *Annu Rev Public Health* 1984;5:385–411.
61. Cox DR. Regression models and life-tables. *J R Stat Soc* 1972;34B:187–220.
62. Thomas DG, Breslow N, Gart JJ. Trend and homogeneity analyses of proportions and life table data. *Comput Biomed Res* 1977;10:373–381.
63. Peto R. Guidelines on the analysis of tumor rates and death rates in experimental animals. *Br J Cancer* 1974;29:101–105.
64. Mantel N, Haenszel W. Statistical aspects of the analysis of data from the retrospective studies of disease. *J Natl Cancer Inst* 1952;22:719–748.
65. Dinse GE. Estimating tumor prevalence, lethality and mortality. Presented at the Symposium on Long-Term Animal Carcinogenicity Studies: A Statistical Perspective, 4–6 March 1985, Bethesda.
66. Hajian G. Statistical issues in the design and analysis of carcinogenicity bioassays. *Toxicol Pathol* 1983;11:83–89.
67. Haseman JK, Huff J, Boorman GA. Use of historical control data in carcinogenicity studies in rodents. *Toxicol Pathol* 1984;12:126–135.
68. Tarone RE. The use of historical control information in testing for a trend in proportions. *Biometrics* 1982;38:215–220.
69. Dempster AP, Selivyn MR, Weeks BJ. Combining historical and randomized controls for assessing trends in proportions. *J Am Stat Assoc* 1983;78:221–227.

# TOXICOKINETICS: AN INTEGRAL COMPONENT OF PRECLINICAL TOXICITY STUDIES

Sonu Sundd Singh
*Nektar Therapeutics India Private Limited, Secunderabad, India*

**Contents**

1 Introduction
2 Objectives of Toxicokinetic Study
3 Toxicokinetics in Various Toxicity Studies for Pharmaceuticals
  3.1 Safety Pharmacology
  3.2 Single Dose-Ranging Toxicity Studies
  3.3 Rising Dose Toxicity Studies
  3.4 Repeated Dose Toxicity Studies (4 Weeks)
  3.5 Repeated Dose Toxicity Studies (6 and 12 Months)
  3.6 Genotoxicity Studies
  3.7 Carcinogenicity Studies
  3.8 Reproductive Toxicity Studies
4 Toxicokinetic Evaluation of Biopharmaceuticals
5 Factors that Affect the Measurement and Interpretation of Toxicokinetic Data
  5.1 Measuring Unbound Versus Bound Drug
  5.2 Racemic Compound Versus Enantiomer
  5.3 Nonlinear Dose Kinetics
  5.4 Measurement of Metabolites
  5.5 Measurement of Produgs in Biological Fluids and Tissues
  5.6 Measurements of Labile Reactive Metabolites
  5.7 Human Metabolites that Are Absent in Animal Models
  5.8 Species Differences in Pharmacokinetics
  5.9 Gender Differences in Pharmacokinetics
  5.10 Interpretation of Drug Plasma Concentrations
  5.11 Choice of Biological Fluid
  5.12 Combination Drugs May Cause Drug–Drug Interactions

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

6 Toxicokinetics and Good Laboratory Practice (GLP)
6.1 Overview of Federal Register (21 CFR Part 58)
6.2 Overview of OECD Principles of GLP
7 Selection and Extrapolation of Dose
7.1 Dose Levels
7.2 Interspecies Scaling: Extrapolation of Dose
8 Route of Dose Administration
9 Sampling Approaches
9.1 Serial Sampling Approach
9.2 Sparse Sampling Approach
10 Analytical Methodology
10.1 Bioanalytical Method
10.2 Bioanalytical Method Development
10.3 Bioanalytical Method Validation
10.4 Method for the Formulation
10.5 Method Validation for the Formulation
10.6 Method Development Issues for Biotech Products
11 Conclusion
Acknowledgment
References

## 1 INTRODUCTION

The establishment of preclinical safety data for an investigational new drug (IND) is a global regulatory requirement. Recently, it has been universally recognized that efficacy and safety of a drug can be better correlated with systemic drug exposure rather than directly with dose [1]. Drug discovery scientists now believe that toxicological experiments should include the assessment of the concentration–time course of the investigated agent for proper interpretation of the toxicological data. Several regulatory guidelines recommend toxicokinetic evaluation of an investigational drug. Toxicokinetics has emerged as an area distinct from pharmacokinetics, where kinetic data for a new compound are generated at subtherapeutic dose levels. The Toxicokinetics guideline of the International Conference on Harmonization (ICH S3A) defines toxicokinetics [1, 2] as "the generation of pharmacokinetic data, either as an integral component of non-clinical toxicity studies or in specially designed supportive studies, to assess systemic exposure." In other words, kinetics and tissue distribution studies performed on the same animals (large animals) or a satellite group of animals (small animals such as rodents) during a nonclinical toxicity study to augment the interpretation of toxicological findings are known as *toxicokinetics*. Here the term "satellite group of animals" refers to a group of animals included in the design and conduct of a toxicity study, which are treated and housed under similar conditions as those of the main toxicity study but used for generating toxicokinetic data.

Toxicokinetics is a relatively small but important component of preclinical animal toxicology studies, which form the foundation of a preclinical program leading to

the filing of an IND application. Therefore, the sole purpose of conducting toxicokinetic studies is to add value to the existing toxicological data. Incorporation of toxicokinetics in the design of conventional toxicity studies helps in ratifying the design of the study with respect to pharmacokinetic and metabolic principles by providing a proof of concept that the animals were really exposed to appropriate systemic levels of administered test compound and/or its metabolites. Toxicokinetics brings in better understanding of toxicity data and better extrapolation of nonclinical data for clinical use. Most importantly, toxicokinetic evaluation makes for good science [2]. The term "pharmacokinetics" should not be confused with "toxicokinetics" as the parameters obtained from the former are used for the characterization of the drug, whereas "toxicokinetics" serves as a fundamental tool for assessing the safety profile of a drug at toxic dose levels. The toxicokinetic data thus obtained are utilized in co-relating the plasma/tissue levels to toxicological findings and their relevance to clinical safety. Toxicokinetic profiling of toxicity studies should establish the level of exposure that has been achieved during the course of the study and may also serve as an alert to toxicologists toward nonlinear, dose-related changes in exposure that may have occurred. Toxicokinetics allows better interspecies comparisons than simple dose/body weight (or surface area) comparisons. The extrapolation of data between animals and humans for the purpose of establishing safety margin, when solely based on doses (mg/kg), can give very misleading results due to vast pharmacokinetic differences between species and gender. Invariably, there appears to be much greater safety margin when plasma concentration data, which represent actual drug exposure, are used [2].

Toxicokinetic (TK) evaluation can be performed during safety pharmacology, single dose and dose escalation studies, repeated dose toxicity studies, reproductive toxicity, genotoxicity, and carcinogenicity studies. TK data at "no observed toxic effect dose" provide a suitable safe starting and upper dose for the clinical investigator in designing a Phase I clinical trial. Since new drug development is a dynamic process and continuous feedback between clinical and nonclinical studies is required for proper safety assessment, flexible procedures based solely on pure scientific judgments are recommended [1]. TK evaluation should conform to Good Laboratory Practice (GLP) as described in 21 CFR Part 58 [3] and OECD GLP guidelines 1997 (F) [4]. Meaningful TK data should be based on reliable and well validated analytical methods [5].

## 2 OBJECTIVES OF TOXICOKINETIC STUDY

The prime objective of a toxicokinetic study is to estimate the systemic exposure of a drug achieved at different dose levels in a toxicity study and to establish a dose versus exposure relationship. The secondary objectives may be to correlate the exposure achieved with the toxicological findings and to contribute to the assessment of the relevance of these findings to clinical safety. The toxicokinetic data in conjugation with toxicity findings also help in identifying the right toxicity species and in designing the subsequent nonclinical toxicity studies. The measurement of various toxicokinetic parameters results in meeting all the above objectives. The parameters usually measured are maximum plasma concentration ($C_{max}$), time point of maximum plasma concentration ($T_{max}$), area under the plasma concentration–time curve

(AUC), and half-life of drug elimination during the terminal phase ($T_{1/2}$). These parameters are usually calculated from the concentration/time data, using a curve prediction package (e.g., Win Nonlin from Pharsight: http://www.pharsight.com).

## 3 TOXICOKINETICS IN VARIOUS TOXICITY STUDIES FOR PHARMACEUTICALS

### 3.1 Safety Pharmacology

These studies comprise a battery of tests for assessing the CNS, cardiovascular, and respiratory functions. The ICH guideline for safety pharmacology [6] does not specifically mention the requirement of toxicokinetic measurement during these studies. Nevertheless, the estimation of systemic exposure helps immensely in correlating any adverse effects to plasma concentrations.

### 3.2 Single Dose-Ranging Toxicity Studies

These studies are carried out primarily in rodents, generally at early stages of development. As a common practice, toxicokinetics is not routinely incorporated due to nonavailability of validated bioanalytical methods. However, some samples are taken and stored for future analysis without any insight into the stability of the samples. Ideally, the kinetic data for the single dose studies should be generated initially for the intended broad dose range as a separate study on the same animal models. This would give a fairly good idea of dose linearity range, that is, the dose versus exposure (AUC) relationship. In addition, it would also provide information about the presence of metabolites, the highest possible dose with no mortality, the exhibition of some toxic effects, the dose administration constraints due to absorption saturation, and formulation-related problems (at very high concentrations the permissible volumes may not be possible). Single dose-ranging studies designed and conducted with the background of this data would certainly provide meaningful results.

### 3.3 Rising Dose Toxicity Studies

These studies are carried out in nonrodents or higher animal models (e.g., dogs, monkeys). Here toxicokinetics can easily be incorporated into the main study, as there is no constraint on the number of sampling points due to the large blood volume. Moreover, these studies are usually carried out after the rodent single dose studies. Thus, by that time, the bioanalytical methods are usually available. Here toxicokinetics is performed on the same animal at various time points for each new dose level. Such an evaluation gives very good information about the dose linearity range and clinical symptoms can be readily co-related with plasma concentration of the parent compound or the metabolites.

### 3.4 Repeated Dose Toxicity Studies (4 Weeks)

Single/rising dose toxicity studies are followed by repeated dose studies in both rodent and nonrodent species. These studies are carried out at three dose

levels—low, intermediate, and high—along with a vehicle control group. The low dose lies near the efficacy dose, whereas the high dose falls near the end of the dose response–linearity curve and this dose is expected to exhibit some signs of toxicity. The control samples should be analyzed along with the treated samples [1]. The selection of dose levels and the treatment regimen (dosage, formulation, route of administration, and dosing frequency) are based on the results of previous single dose and rising dose toxicity studies. Usually a duration of 4 weeks is recommended for supporting Phase I clinical trials. A satellite group of animals is generally used for rodent studies due to the sampling constraint of low blood volumes, as only 10% of the total blood volume can be withdrawn [7]. Samples are withdrawn on day 1, day 14, and day 28. Sampling times are based on earlier pharmacokinetic data or on the toxicokinetic data obtained for single dose/rising dose studies. Typically, three or more animals are used for each dose level. An average of drug concentrations at each time point is taken and toxicokinetic parameters are also presented as the mean ± SD. In some instances, pooling of equal volumes of plasma samples for each time point is also done prior to sample analysis. Tissue distribution is also incorporated in the design of the study and organs of interest are removed after sacrifice and analyzed for the drug and metabolites. A typical design of a 4 week toxicokinetic study is shown in Table 1. A comparison of systemic exposure in terms of plasma and tissue concentrations and AUC is done in order to interpret the outcome of the study. Proper designing of a toxicokinetic study would provide valuable information about the accumulation of drug and/or metabolites in body fluids or organs/tissues. There is also a possibility of reduction in the overall exposure after repeated dose administration, which may be attributed to autoinduction of drug metabolizing enzymes, saturated absorption, or increased clearance [2]. At low doses, blood levels may indicate nonaccumulation but tissue distribution could reveal accumulation of drug in specific organs.

### 3.5 Repeated Dose Toxicity Studies (6 and 12 Months)

In order to assess longer exposure, repeated dose studies are conducted for 6 and 12 months in rodents and nonrodents, respectively. Toxicokinetics can be incorporated in a similar manner as 4 week studies, covering the entire span of the study.

**TABLE 1 Animal Grouping for a 4 Week Toxicokinetic Study**

| | | Day of Sampling | | | | | |
|---|---|---|---|---|---|---|---|
| | | Day 1[a] | | Day 14[b] | | Day 28 | |
| | | Animal Numbers | | | | | |
| Group | Dose (mg/kg) | Male | Female | Male | Female | Male | Female |
| I | 10.0 | 1–3 | 19–21 | 4–6 | 22–24 | 7–9 | 25–27 |
| II | 30.0 | 10–12 | 28–30 | 13–15 | 31–33 | 16–18 | 34–36 |

[a]Animals were bled on day 1 for toxicokinetics data and dosing of these animals was continued for 28 days and on day 29 the animals were sacrificed for tissue distribution after 24 hours from the last dosing.
[b]These animals would be sacrificed after day 14 for toxicokinetics data.

### 3.6 Genotoxicity Studies

Generally, two *in vitro* tests and one *in vivo* tests are recommended for genotoxicity testing [8, 9]. A rodent micronucleus (bone marrow or peripheral erythrocytes) test or chromosomal aberration test is carried out for *in vivo* investigation of genotoxicity. The ICH toxicokinetics guideline [1] recommends that for "negative results of *in vivo* geno toxicity studies, it may be appropriate to have demonstrated systemic exposure in the species used or to have characterized exposure in indicator organs."

### 3.7 Carcinogenicity Studies

Carcinogenicity studies are generally conducted to identify a tumorigenic potential of a drug in animals and to assess the relevant risk in humans. Pharmaceuticals that are proposed for continuous use for at least 6 months of clinical use should be considered for carcinogenicity testing. Pharmaceuticals administered for a short duration of exposure, such as anesthetics, critical care medications, and radiolabeled imaging agents, do not need carcinogenic assessment. Certain formulations or delivery systems that prolong the exposure of the test compound need carcinogenicity studies. Carcinogenicity studies are not needed for endogenous substances like peptides, proteins, and their analogues. Carcinogenicity studies should be completed before filing an NDA (New Drug Application). Complete rodent carcinogenicity studies are not needed in advance for conducting clinical trials. The following ICH guidelines address the various issues concerning nonclinical carcinogenicity studies:

S1A: Guidance on the Need for Carcinogenicity Studies of Pharmaceuticals, March 1996 [10].

This guideline defines the need and conditions under which carcinogenicity studies should be conducted in order to avoid unnecessary usage of animals.

S1B: Testing for Carcinogenicity of Pharmaceuticals, July 1997 [11].

This guideline describes the experimental approaches for the evaluation of carcinogenic potential of pharmaceuticals. This also outlines the necessity of two routine long-term rodent studies in rats and mice.

S1C: Dose Selection for Carcinogenicity Studies of Pharmaceuticals, March 1995 [12].

S1C (R): Addendum to Dose Selection for Carcinogenicity Studies of Pharmaceuticals: Addition of a Limit Dose and Related Notes, July 1997.

These guidelines address the issues of dose selection for long-term carcinogenicity studies.

Additionally the Center for Drug Evaluation and Research of the U.S. FDA in May 2002 issued the guidance *Carcinogenicity Protocol Submission*.

The basic scheme consists of a dose range-finding study, generally of 90 days' duration. Dose selection requires an in-depth knowledge of the pharmacology, repeated dose toxicology, toxicokinetics, and human pharmacokinetics of the test compound. All regulatory authorities recommend that a high dose be set with no mortality for the study period. There are six different methods for selection of high

dose for range–finding studies: (1) toxicity based endpoints, (2) pharmacokinetic endpoints, (3) saturation of absorption, (4) pharmacodynamic endpoints, (5) maximum feasible dose, and (6) additional endpoints.

This is followed by two main studies—(a) long-term (2 year) study in rats and another long-term or short-term study that provides additional information that is not readily available from the first long-term study. This is either a long-term study in a second rodent species (e.g., mouse) or a short- or medium-term study in alternative models of carcinogenicity. There are four short- or medium-term transgenic models in common use: (1) activated oncogene (Tg.AC model), (2) activated oncogene (rasH2 model), (3) inactivated tumor suppressor gene ($p53^{+/-}$ model), and (4) inactivated DNA repair gene ($XPA^{-/-}$ model). The short-term studies can provide better answers for differences in the range of susceptible target tissues, in which tumors develop, and knowledge of a compounds' ADME profile for human risk assessment. There is a regulatory requirement for information on the assessment of systemic exposure to the parent compound and its metabolites [1]. For the toxicokinetic study, blood samples are collected from the caudal vein in the unanesthetized state from satellite animals (~5–10 per sex/group) at various time intervals such as 4, 12, 26, 52, and 103 weeks (morning and afternoon). These satellite animals are sacrificed after final sampling without any further investigation.

### 3.8 Reproductive Toxicity Studies

Although no specific guidance [1, 13] is available, regulatory authorities still expect toxicokinetic studies to be conducted on pregnant animals. The toxicokinetic data are generated for fertility studies in rats—embryo and fetal development. Reproductive toxicity studies became very important after the tragic thalidomide story, where the drug caused rare birth defects in pregnant women. The tragedy led to the passage of the Harris–Kefauver Amendment to the Federal Pure Food and Drug Act in the United States in 1962 to ensure that approved drugs have proof of safety and efficacy [14]. Data from nonpregnant animals is used to set the dose levels for the reproductive toxicity study. The toxicokinetic evaluation should be carried out in embryo–fetal studies at the beginning and end of gestation in the animal study. However, it can also be conducted in preliminary studies or in the main study with a satellite group of animals.

## 4 TOXICOKINETIC EVALUATION OF BIOPHARMACEUTICALS

The toxicokinetic assessment of biotechnology products is a difficult exercise. However, some information on the absorption, disposition, and clearance in relevant models is essential in order to predict safety prior to clinical investigation. Therefore, wherever possible, systemic exposure should be monitored during toxicity studies [1, 15]. The species difference in pharmacokinetics would help in assigning dose–response relationships in toxicity studies. Single- and multiple-dose pharmacokinetic, toxicokinetic, and tissue distribution studies have proved useful in the interpretation of safety data. However, mass balance studies may not yield useful results. A common problem encountered during the measurement of biotechnology products is the presence of neutralizing antibodies raised in animals against human

proteins. The neutralizing antibodies alter the pharmacokinetic profile due to immune-mediated clearance and affect the interpretation of toxicity data. Some biotechnology products may exhibit a significant delay or prolongation in the expression of pharmacodynamic effect relative to pharmacokinetic profile/plasma levels. While measuring radioactivity it is important that the radiolabeled linkage is stable. The assessment of radioactivity becomes difficult due to rapid *in vivo* metabolism of radiolabel. Care should be taken in the interpretation of studies using radioactive tracers incorporated into specific amino acids because of recycling of amino acids into non-drug-related proteins/peptides. The biotech products are metabolized into small peptides and individual amino acids. Since the metabolic pathways are generally understood, classical biotransformation studies are not required. Selection of assay methods should be based on a proper understanding of the behavior of the biopharmaceutical in the biological matrix such as plasma, serum, whole blood, and cerebrospinal fluid. Protein binding is an important factor that can affect the assay. One quantitative assay should be sufficient, provided it is specific for the analyte, and the effect of plasma protein binding and antibodies on the assay should be predetermined.

## 5 FACTORS THAT AFFECT THE MEASUREMENT AND INTERPRETATION OF TOXICOKINETIC DATA

### 5.1 Measuring Unbound Versus Bound Drug

Plasma protein binding of a drug influences its pharmacokinetic parameters [16–23] (i.e., distribution, clearance, and elimination half-life) and pharmacodynamic parameters (i.e., efficacy and toxicity) [24]. A drug may bind to several components/macromolecules, for example, albumin, α1-acid glycoprotein (AGP), lipoproteins, immunoglobulins (IgG), and erythrocytes within the blood. The formation of a drug–protein complex is termed drug–protein binding. Most drugs bind to proteins in a reversible manner by means of weak chemical bonds such as ionic, van der Waals, hydrogen, and hydrophobic bonds with the hydroxyl, carboxyl, or other reversible sites available in the amino acids that constitute the protein chain. The major contribution to drug binding in the plasma is made by albumin [25], which is synthesized in the liver and constitutes about half of total plasma proteins. The molecular weight of albumin ranges between 65,000 and 69,000 daltons. Acidic drugs are known to bind strongly [25] to human serum albumin (HSA), which has two ligand-specific binding sites [26, 27], namely, site I and site II. The ligand selectivity is comparatively broader for these two sites, allowing a range of drug molecules to bind at these sites. This broad selectivity is considered to be a consequence of the significant allosteric effects in HSA [28] and drug molecules can also interact non-specifically with HSA. HSA is responsible for maintaining the osmotic pressure of the blood and is a carrier of many molecules [29], such as free fatty acids, bilirubin, and various hormones (e.g., cortisone, aldosterone, thyroxin).

α1-Acid glycoprotein (AGP) is a relatively low molecular weight (approximately 40,000 daltons) protein. Its concentration in plasma is about 40–100 mg/100 mL and primarily binds to basic (such as lidocaine, propranolol, imipramine, and quinidine) and neutral drugs [25] in addition to some acidic drugs [30]. In general, gyrase inhibi-

tors [31] and NSAID (nonsteroidal anti-inflammatory drugs) [32] bind to albumin and antiretroviral compounds [33] bind to AGP. Binding of drugs to lipoproteins, red blood cells, and other membranes is not a true binding reaction but is similar to dissolving of the drugs in the lipids of the membrane. Very lipophilic drugs partition preferentially into the membrane lipids rather than the plasma water. Some drugs bind strongly to particular tissue components such as DNA (e.g., some anticancer drugs) and melanin-rich tissues (e.g., chloroquine, amiodarone).

Free drug concentration in plasma is responsible for the observed pharmacological effect or therapeutic response [34–36]. The drug bound to plasma protein is not available for distribution, hepatic metabolism, and renal elimination. The drug–protein complex does not permeate phospholipid bilayers, including capillary membranes, glomerular membranes in the nephrons, and the blood–brain barrier. Bound drugs are also less available to the cytochrome P450 and other enzymes involved in first-pass metabolism. The driving force for drug excretion in the kidney is the free drug concentration in the plasma. The glomerular capillaries permit the passage of free drug molecules but restrict the passage of plasma proteins and drug–protein complexes and, as a consequence, only free or unbound drug is filtered. After the metabolic and excretory processes have cleared much of the free drug, the reversible drug–protein complex serves as a depot to replenish the concentration *in vivo* [35]. Drugs with high protein binding tend to have a greater elimination half-life compared to those with low binding. The prolonged pharmacological activity resulting from these factors may be desirable or may promote the emergence of undesirable side effects. Therefore, estimation of the extent of drug–protein binding is crucial for clinical drug development. Ideally, determination of both total and unbound plasma drug concentrations is necessary to obtain an understanding of drug available for pharmacological effect.

In the past, several techniques have been explored for quantitative determination of drug–protein binding *in vitro*. Among those, equilibrium dialysis, gel filtration, ultrafiltration, and ultracentrifugation [31, 37–41] have been conventionally and most commonly used. These conventional methods suffer from long analysis time. Different *in vitro* methods yield different results; for example, equilibrium dialysis indicated 23% plasma protein binding for fleroxacin whereas ultrafiltration indicated 47% plasma protein binding [31]. A lot of interlaboratory variation in plasma protein binding data obtained via the same technique leading to a broad range for ciprofloxacin (20–40%), ofloxacin and norfloxacin (8–30%), and enoxacin (30–50%) [31] has been reported. Although the ultrafiltration method is believed to be comparable to *in vivo* processes, such as ultrafiltration of drug in the kidney [31], it still suffers from the limitation of nonspecific binding of the drug to the membrane and leakage of drug from the membrane.

In comparison to conventional methods, chromatography-based methods [37, 42–46] employing columns immobilized with plasma proteins have gained popularity over the years because of their simplicity, specificity, and speed.

### 5.2 Racemic Compound Versus Enantiomer

Enantiomers must be considered as essentially different chemical compounds, as they usually differ in their pharmacokinetic and pharmacodynamic properties due to stereoselective interaction with biological macromolecules. Early evaluation of

stereoselective pharmacokinetics, pharmacodynamics, and toxicity of individual enantiomers is essential to decide whether to develop the racemate or an individual enantiomer. Absorption from the gastrointestinal tract, tissue distribution, and renal excretion are passive processes for most drugs, where the extent and rate are mainly governed by physiological properties of the drug. Since the physiological properties of stereoisomers are the same, stereoselectivity is not expected for these processes unless active transport is involved. On the contrary, stereoselective plasma protein binding and metabolism are common and well documented [47, 48]. In addition, both plasma protein binding and metabolism of different enantiomers are species specific due to structural differences in plasma proteins and drug metabolizing enzymes across species. The plasma protein binding of MK-571 [49] is extensive, stereoselective, and species dependent. The *R* enantiomer was bound to rat plasma to a greater extent as compared to the *S* enantiomer, whereas in dog and monkey the reverse was true [50]. The elimination of the enantiomers was directly related to plasma protein binding in a particular species and the unbound enantiomer was more rapidly cleared. Mephenytoin exhibits stereoselective metabolism, which is also species specific. In *in vitro* studies using liver microsomes [51], it was observed that the rate of 4-hydroxylation was twofold to sixfold higher for *R* enantiomer as compared to *S* enantiomer in rabbit, dog, and rat, whereas 4′-hydroxylation of *S* enantiomer was 5–15 times higher as compared to *R* enantiomer in human and monkeys. Similarly, the *S* enantiomer of propranolol had a higher intrinsic clearance in liver microsomes of dog as compared to *R* enantiomer. On the contrary, human microsomes exhibited higher clearance for *R* enantiomer [52]. Flurbiprofen, an aryl propionic acid NSAID, is marketed as a racemic compound. Its anti-inflammatory activity is believed to reside with the *S* enantiomer [53, 54]. *In vivo*, the *R* enantiomer gets irreversibly converted to the *S* enantiomer. The extent of this unidirectional conversion is 100% in guinea pigs, 40% in dog, and ~5% in rat, gerbil, and human [55, 56]. Ketoprofen, another 2-aryl propionic acid NSAID, also exhibits *in vivo* chiral inversion. The extent of inversion was high in rat [57], low in rabbit [58], and insignificant in human [48].

The large differences in pharmacokinetic profile of enantiomers result in stereoselective safety and efficacy. Thalidomide, the hypnotic that caused birth defects called phocomelia, was administered as a racemic compound, where both the enantiomers had equal sedating effect. After its withdrawal, studies in pregnant mice and rats revealed that the *S* enantiomer of thalidomide was teratogenic, whereas the *R* enantiomer was devoid of any such symptoms [59, 60]. Following intraperitoneal administration of *S* enantiomer (200 mg/kg/day) to pregnant animals, 30% of mice and 50% of rats had deformed fetuses; however, there was no deformity reported after the administration of *R* enantiomer. On the other hand, both *R* and *S* enantiomers were equally teratogenic in rabbit: 40% of deformed fetuses were recorded from pregnant rabbits that were administered racemic thalidomide, whereas 16–17% when *R* or *S* enantiomer was administered individually [61, 62]. Thalidomide is a stark example of stereoselective toxicity.

After the administration of a racemic drug, it becomes important to monitor the respective enantiomers *in vivo*. Development of a more potent and safer enantiomer instead of the racemic compound not only reduces the dose but also addresses the concerns of toxicity. For the racemic drug, it becomes important to demonstrate that the administered enantiomer is stable *in vivo* and does not get converted into

the undesired enantiomer. In other words, a sensitive bioanalytical method for monitoring the *in vivo* chiral conversion of such a drug candidate during toxicokinetic analysis is desirable.

### 5.3 Nonlinear Dose Kinetics

When the dose of a drug is increased, we expect that the concentration at steady state will increase proportionately: that is, if the dose rate is increased or decreased, say, twofold, the plasma drug concentration will also increase or decrease twofold. This is known as linear dose kinetics or first-order kinetics. The rate constants of such models do not change with concentration in the body. However, for some drugs, the plasma drug concentration changes either more or less than would be expected from a change in dose rate. Such drugs do not obey first-order kinetics. In such cases, the first-order kinetics get transformed into a mixture of first-order and zero-order rate processes. This is known as nonlinear pharmacokinetic behavior and can cause problems when adjusting doses. For such drugs that behave nonlinearly within the therapeutic range (e.g., ascorbic acid and naproxen), it is difficult to predict the plasma concentrations achieved. Nonlinear kinetic behavior may occur due to various reasons, such as saturation of first-pass metabolism causing an increase in bioavailability; saturation of protein binding sites causing a change in fraction of drug unbound in plasma; saturation of elimination mechanisms causing a change in intrinsic clearance; and absorption being rate limited.

### 5.4 Measurement of Metabolites

Although the primary concern of toxicokinetics is to determine the plasma and tissue levels of the parent compound in the toxicology species, there are instances when measurement of metabolite concentrations in plasma or other body fluids may become essential for the conduct of toxicokinetic studies. For a "prodrug," the primary active metabolite is monitored. If the administered compound undergoes extensive metabolism, the measurement of plasma or tissue concentrations of a major metabolite is the only practical means of estimating exposure. The metabolites are also monitored in addition to the parent compound when the compound is metabolized to one or more pharmacologically or toxicologically active metabolites that could significantly contribute to pharmacological responses. Measurement of metabolites, however, cannot account for toxic reactive intermediate metabolites, which are known to cause idiosyncratic reactions.

### 5.5 Measurement of Produgs in Biological Fluids and Tissues

The term prodrug was first introduced by Albert in 1958 [63] to describe compounds that undergo *in vivo* biotransformation before exhibiting their pharmacological effect. Prodrugs are precursors of active principles that can be used to modify the pharmacokinetics of a drug. This may lead to improvement in bioavailability, increased drug stability, enhancement of patient acceptance, and compliance by minimizing taste and odor problems or elimination of pain on injection. Prodrugs do not possess any pharmacological effect, but their metabolite is pharmacologically

active. Piva ampicillin, tala ampicillin, and baca ampicillin are ester prodrugs of ampicillin. The absorption of these prodrugs is complete as compared to 50% absorption of ampicillin [64, 65]. Enalapril, a prodrug of enaprilate, is a popular ACE inhibitor with an oral absorption of 100%, whereas the parent compound is poorly absorbed (~10%) [66, 67]. Acyclovir, a potent antiherpes drug, is poorly absorbed after oral administration. Deoxyacyclovir is the prodrug of acyclovir, which provides superior delivery of acyclovir [68–70]. γ-Glutamyldopa is a kidney-specific prodrug of levodopa [71], responsible for receptor-mediated vasodilation in the kidney. It is a precursor of the neurotransmitter dopamine, which plays an important role in the CNS. During the toxicokinetic evaluation of a prodrug, the analytical method should be sensitive enough to distinguish between the prodrug and the active metabolite in biological matrices of interest.

### 5.6 Measurement of Labile Reactive Metabolites

Following an oral administration in humans, the plasma levels of clopidogrel are extremely low due to extensive metabolism and are difficult to quantify. The active metabolite (thiol) is highly labile and remains undetected in plasma. The main circulating metabolite (carboxylic acid derivative) is pharmacologically inactive. Since the parent compound and the active metabolite cannot be detected, the estimation of the major inactive metabolite is the only way of assessing systemic exposure [72, 73]. The half-life of acetaminophen is two- to threefold longer in rat than in mouse, it is toxic in mice at a low dose of 200–300 mg/kg but well tolerated in rats up to a dose of 1500 mg/kg. The hepatotoxicity caused by acetaminophen is due to a reactive toxic metabolite, which is labile and too small in quantity to be detected in plasma [74]. Therefore, the parent compound and its sulfate conjugate are monitored *in vivo*, which is of little relevance in predicting hepatotoxicity.

### 5.7 Human Metabolites that Are Absent in Animal Models

Major metabolic differences between the animal and human studies reduce the confidence in toxicity studies as predictors of safety. A method of comparing *in vitro* metabolic profiles of various animal models with that of humans using recombinant liver microsomes and/or recombinant drug metabolizing enzymes [75] has recently come to the rescue of toxicologists. An early *in vitro* comparison of the metabolic profile of test species with humans helps in selecting the relevant animal models for the assessment of potential human risk.

### 5.8 Species Differences in Pharmacokinetics

Due to ethical reasons, the safety of all pharmaceutical drugs cannot be directly evaluated on humans but should be tested on rodents and nonrodents prior to approval for clinical investigation. The extrapolation of risk assessment from animal to human is a challenge, as it is far from straightforward to extrapolate animal toxicity to human toxicity. This is because the drug-induced toxicity is often species dependent and to date no single animal model resembles humans completely. Among rodents, there are vast differences in response to the same drug. Most common antibiotics are lethal in guinea pigs and hamsters but may be tolerated by

mice, rats, and gerbils. Fleming had used the rat as the animal model for penicillin; had he used guinea pigs or hamsters in his research, the wonder drug would have been dropped at an early stage. When there is so much difference among various rodent species, one can imagine the amount of variability that could be expected between rodents and humans. Aspirin causes birth defects in most rodents but not in humans. Cortisone causes birth defects in mice but not in rats and humans. Aspartame, an artificial sweetener, has been shown to cause lymphoid cancer in rats, whereas there is no clear evidence of its toxicity in humans, and has been in use for a long time. A drug may cause cancer in development in certain animal species but not in humans and vice versa.

Acetaminophen caused hepatotoxicity [74] in mice at the low dose of 200–300 mg/kg but was tolerated in rats up to a high dose of 1500 mg/kg. In humans, it is well tolerated within the therapeutic range but is toxic at a high dose of 10–15 g (150–200 mg/kg). Dichloromethane, a common industrial chemical, causes lung and liver cancer in mice after chronic inhalation exposure but not in rats and hamsters under similar exposure [76]. Treatment with perfluoro decanoic acid, a potent peroxisome proliferator, led to the development of pronounced hepatomegaly in rats but not in guinea pigs [77]. The sulfonamide, sulfomethoxazole, produced thyroid nodules in rats at 50 mg/kg/day dose for 1 year, whereas rhesus monkeys treated with a higher dose (300 mg/kg/day) for the same time period exhibited no increase in thyroid weight and no morphological alterations [78].

Indomethacin, an anti-inflammatory agent used widely for the treatment of rheumatoid arthritis, caused ulcers in rats and dogs but not in humans [79].

Species differences in toxicity of a drug can be interpreted and explained in terms of species differences in the pharmacokinetics or ADME of a drug. Species differences in metabolism, which is the major cause of diversity, arise due to structural differences in drug metabolizing enzymes. Since drug-induced toxicity is directly related to systemic exposure to drug and its metabolites, the species differences in metabolism of the drug are perhaps the most important factor in explaining the observed differences in toxicity. The nucleotide and amino acid sequence of CYP450 varies in different species. Even small changes in the amino acid sequence can give rise to a large difference in substrate specificity [80]. Similarly, uridine diphosphoglucuronyl transferases (UDPGTs), responsible for glucuronidation, also exhibit species differences in structure. Comparison of 10 rat UDPGTs and 8 human UDPGTs showed that they have a common C terminal domain but the N terminal is quite variable. As a result, their substrate specificities are different, although they still have overlapping substrate specificities [81]. Carboxyl esterases hydrolyze drugs containing ester bonds and amide linkages. They exist as multiple isoforms and exhibit significant species differences in their activities. The carboxyl esterases across various species (i.e., mouse, hamster, rat, guinea pig, rabbit, and monkey) exhibit high homology in amino acid sequence but have different N terminal amino acids, resulting in different substrate specificity [82, 83].

A comparison of the *in vitro* metabolic profile of losartan, a potent nonpeptide angiotensin II receptor antagonist, in hepatic slices obtained from rat, monkey, and human provides clear evidence of species differences in metabolism [84]. Rat liver slices primarily oxidize losartan to monohydroxylated and carboxylic metabolites. Glucuronidation of tetrazole moiety predominated in monkey, whereas in human both the oxidized and glucuronidated metabolites were equally present.

The metabolic clearance of dihydropyridine calcium channel blockers amlodipine, felodipine, nicardine, nisoldipine, nitrendipine, and nilvadipine was compared in rat, dog, and human. For all the compounds, the metabolism in humans was quantitatively similar to dogs, whereas rats show a much higher capacity of metabolism [85]. Azidothymidine (AZT), an HIV reverse transcriptase inhibitor, is extensively metabolized in humans, but not in rats [86, 87]. Acetaminophen is primarily sulfated in rats but glucuronidation is more prominent in humans [88, 89]. There is a marked species difference in enzyme induction. For example, the gastric acid-suppressing drug omeprazole is a CYP1A2 enzyme inducer in humans but has no such inductive effect in mice or rabbits [90, 91]. Furthermore, members of the CYP3A subfamily in rats are inducible by the steroidal agent pregnenolone-16α-carbonitrile, but not by the antibiotic rifampin. The opposite is true in rabbits and humans [92, 93]. Therefore, the assumption that xenobiotics that do not induce drug-metabolizing enzymes in animal models would not induce human enzymes as well and vice versa is totally baseless. To overcome this drawback, both *in vitro* (human hepatocytes) and *in vivo* (probe drugs for certain human cytochrome P450s) techniques have become available and have increasingly been used by investigators to evaluate the potential induction of human cytochrome P450s by a variety of therapeutic agents. The above examples demonstrate that extrapolation of drug metabolism from animals to human is difficult, if not impossible.

### 5.9 Gender Differences in Pharmacokinetics

Sex-related differences in metabolism have been known for more than sixty years. Recent studies have shown that gender differences in xenobiotic metabolism are more pronounced in rat than any other species. Early studies designed to examine the mechanism of sex-dependent differences in response to specific barbiturates [94] demonstrated that female rats had higher and more prolonged serum concentration of parent compound due to lower rate of metabolism as compared with male rats. Female rats have 10–30% less total cytochrome P450 (CYP450) as compared to male rats. This explains why female rats in general metabolize many drugs and compounds more slowly than male rats. Cocaine, diazepam, hexobarbital, indinavir, morphine, phenobarbital, and tolbutamide are some of the drugs and chemicals that have a higher rate of metabolism in male rats as compared to female rats [95]. Although male rats generally have a higher rate of metabolism, there are instances where female rats exhibit a higher metabolic rate than males. There are sex-dependent differences in the expression of microsomal CYP isoforms, which could be attributed to the fact that the CYP isoforms are very different in male and female rats. Immunological data have shown that CYP2C11 is expressed only in male rats, whereas CYP2C12 expression is specific to female rats. CYP2A2 and CYP3A2 are male dominant whereas CYP2A1 and CYP2C7 are female dominant enzymes [96–99]. Sprague–Dawley rats treated with phenobarbital (1,3, or 20 mg/kg) for 6 days showed an increase in hexobarbital hydroxylase activity and aminopyrine N demethylation in hepatic microsomes prepared from male rats but not in those prepared from female rats [100]. Sex-dependent differences have also been observed in the expression of conjugative enzymes such as sulfotransferase, glutathione S transferases, and glucuronyl transferases [95]. In general, male rats have higher drug metabolizing enzyme activities than female rats. Sex-related differences in metabolism are

not unique to rats but have been observed in mice [101], ferrets [102], dogs [103], and humans [104]. However, the magnitude of differences in these species is less pronounced as compared to rats.

After rat, xenobiotic metabolism has been best characterized in mouse. Sex-related differences in mice are often strain dependent [101]. Males have a high rate of metabolism in some strains while females have a high rate in other strains. In general, female mice more commonly have a higher rate of metabolism than males [101]. The magnitude of sex-dependent differences in metabolism is subtle compared to rats. For example, a male rat can have enzyme activities as much as fivefold greater than a female rat, whereas the greatest degree of difference in mice is twofold. Sex-dependent differences in the xenobiotic metabolism of rabbit has been reported in the flavine containing monooxygenase flavo proteins that oxidize nitrogen and sulfur containing compounds [105].

Indinavir, a potent HIV protease inhibitor, exhibits sex-related differences in clearance in rats and dogs but not in monkeys and humans. Rats exhibited a clearance of 89 mL/min/kg for male and 41 mL/min/kg for female. Dogs exhibited just the opposite trend, with higher rate of clearance in female dogs (26 mL/min/kg) as compared to male dogs (15 mL/min/kg) [106]. The *in vitro* liver microsome data was consistent with the *in vivo* data, where male rats had substantially higher metabolizing activity toward indinavir as compared to female rats. Hepatic microsomes from female dogs exhibited a higher metabolic rate than male dogs. Monkey and human liver microsomes exhibited no differences in the metabolism of indinavir. The functional activity of CYP3A when measured in liver microsomes exhibited significant gender differences in rat and dog [106].

In general, when sex-dependent differences in pharmacokinetics of xenobiotics are observed in humans, females have higher plasma concentrations compared to males. This difference has been reported with drugs like acetaminophen, aspirin, chlorophenicol, diazepam, lidocaine, mephobarbital, nortriptyline, oxazepam, propranolol, rifampicin, and tetracycline [95]. Unlike the rat, the sex-related differences in the pharmacokinetics in humans are not related to differences in CYP isoforms. This lack of sex-dependent isoform in humans indicates that the rat, the most popular animal model so far, is not an accurate model for prediction of sex-dependent pharmacokinetic differences in humans. Furthermore, these differences in humans are not as significant as those in rodents. In humans, interindividual differences in metabolism due to genetic polymorphism outweigh any differences regulated by sex-related factors [104]. Therefore, metabolism is not the sole factor for the gender differences in humans: the basic differences in physiology and body composition are responsible for it.

Absorption of certain drugs from the gastrointestinal tract may be affected by the fact that both gastric acid secretion and gastric emptying are lower in women as compared to men [107]. The difference in rates of gastric absorption cause men to achieve peak sodium salicylate plasma concentration more quickly than women. Moreover, the volume of distribution of certain drugs can be affected by the fact that lean body mass is greater in men while adipose tissue content is greater in women [107]. For example, intramuscular injection of drugs is handled differently between men and women because of sex differences in the differences of gluteal fat. Because of this, lipophilic chemicals can have a greater volume of distribution in woman as compared to men. Hormonal regulation also plays a role in metabolism

in humans and the manipulation of normal levels of circulating steroid hormones can alter the way men and women metabolize xenobiotics. There is evidence that the phase of a woman's menstrual cycle can affect the pharmacokinetics of a number of xenobiotics by altering their distribution and clearance. There are changes in gastric emptying rate and acidity of stomach contents at about day 14 of a 28 day menstrual cycle. As progesterone rises, ovulation increases the gastric emptying rate and the secretion of acid in the stomach. Therefore, the bioavailability of a compound may change depending on the phase of a woman's menstrual cycle. The phase of the menstrual cycle has also been shown to affect the volume of distribution and half-life of various drugs, including diazepam and acetaminophen [108].

### 5.10 Interpretation of Drug Plasma Concentrations

Higher plasma concentration in the later stage of a repeated dose toxicokinetic study is caused by the accumulation of drug. The accumulation of drug occurs due to long plasma half-life, reduced clearance, inhibition of drug metabolizing enzymes, or enterohepatic recirculation. Clearance can be affected by impaired hepatic function or capacity-limited elimination. Lower plasma concentrations at a later stage of a study as compared to initial levels could be related to high first-pass metabolism or autoinduction of drug metabolizing enzymes. At times low plasma concentrations may be interpreted as no accumulation but exceptionally high drug levels may be present in the tissues. Therefore, toxicokinetic evaluation should be carried out after careful comparison of both plasma and tissue levels. Lack of dose proportionality occurs due to either autoinduction of drug metabolizing enzymes or saturation of mechanisms such as intestinal absorption, enzyme metabolizing systems, and tubular reabsorption.

### 5.11 Choice of Biological Fluid

Plasma is the most common matrix of choice. The less common ones are serum, whole blood, or cerebrospinal fluid. Whole blood estimation is generally carried out for compounds that tend to accumulate in blood cells. CNS drugs are monitored in cerebrospinal fluid.

### 5.12 Combination Drugs May Cause Drug–Drug Interactions

Marketed drugs that are intended for combination therapy need special toxicokinetic evaluation for possible drug–drug interactions. Pharmacokinetic drug interactions are consequences of altered levels of exposure to the drugs or its metabolites through one or more of the following mechanisms: altered absorption by chelation or complex formation, altered distribution due to the competition of two coadministered drugs for the same plasma protein binding site, altered transport inhibition of transporter proteins, induction and/or inhibition of CYP450, altered drug excretion because of competition for renal anion or cation transport systems, changes in urinary pH, or inhibition of renal metabolism. Among these, "inhibition" and 'induction' of cytochrome P450 is of prime importance and should be extensively explored in the preclinical stage, as they have been known to be notoriously associated with the withdrawal of several drugs (suruvidine, terfinadine, mibefradil, astemizole, cisapride, levacetylmethadol, cerivastatin, and repaglinide) in the past.

When administered in combination, the drug metabolizing enzymes, especially cytochrome P450, may be transcriptionally activated or induced by one of the drugs, resulting in increased production of the enzyme. The pharmacological consequences of enzyme induction could lead to faster metabolism, reduced efficacy, and/or formation of toxic metabolites, resulting in treatment failure and/or toxicity of drugs, which is the basis of undesirable drug–drug interactions. Evaluation of drug–drug interactions in animal models is less relevant due to the structural differences in the drug metabolizing enzymes across species. For instance, omeprazole induces CYP1A2 in humans but not in mice and rabbits [90, 91]. Rifampicin induces CYP3A in human and rabbit but not in rat. Pregnenolone-16-α-carbonitrile induces rat CYP3A but not human [92, 93].

When the enzyme responsible for metabolizing a drug is inhibited by the coadministered drug, there is usually an increased plasma concentration of the former, which may lead to toxicity—another major reason of drug–drug interaction. The enzyme inhibition potential of drugs also differs across species. For instance, furafylline is a more potent inhibitor of phenacetin O-deethylation catalyzed by CYP1A2 in humans as compared to rats [109]. Quinidine is a more potent inhibitor of debrisoquine 4-hydroxylation in humans than in rats [110]. To overcome these obstacles, the enzyme inductive or inhibitory potential of combination drugs should first be evaluated *in vitro* on recombinant enzymes from human and relevant animal species [75]. If the *in vitro* enzyme inductive or inhibitory effect of the animal model is comparable to humans, only then should an *in vivo* study be designed. *In vitro* drug interaction studies should be conducted at concentrations similar to those attained *in vivo*.

Drugs that are metabolized by more than one drug metabolizing enzyme generally have a decreased likelihood of clinically significant drug interactions due to the availability of compensatory metabolic pathways if one is inhibited. Drug interactions are likely to be particularly important when the elimination of a drug occurs primarily through a single metabolic pathway, the drug is a potent inhibitor or inducer of a drug metabolizing enzyme, one or both of the interacting drugs have a steep dose–response curve, one or both of the interacting drugs have a narrow therapeutic range, inhibition of the primary metabolic enzyme or induction of a secondary metabolic enzyme results in diversion of the drug into an alternative pathway that generates a metabolite having toxic or modified pharmacodynamic activity, the drug has nonlinear pharmacokinetics, or the interaction results in conversion from linear to nonlinear pharmacokinetics.

## 6 TOXICOKINETICS AND GOOD LABORATORY PRACTICE (GLP)

There are two main guidelines of GLP.

### 6.1 Overview of Federal Register (21 CFR Part 58)

In 1976, the FDA proposed GLP guidelines; they were finalized in 1978 and became effective in 1979. The GLP guidelines continue to evolve with additional changes made in September 1987 and updates in April 2006.

The FDA GLP guidelines are divided into nine subparts, dealing with general provisions, organization and personnel, facilities, equipment, testing facility operations, test and control articles, protocol for and conduct of a nonclinical laboratory study, records and reports, and disqualification of testing facilities.

21 CFR PART 58

Subpart A—General Provisions

- 58.1 Scope
- 58.3 Definitions
- 58.10 Applicability to studies performed under grants and contracts
- 58.15 Inspection of a testing facility

Subpart B—Organization and Personnel

- 58.29 Personnel
- 58.31 Testing facility management
- 58.33 Study director
- 58.35 Quality assurance unit

Subpart C—Facilities

- 58.41 General
- 58.43 Animal care facilities
- 58.45 Animal supply facilities
- 58.47 Facilities for handling test and control articles
- 58.49 Laboratory operation areas
- 58.51 Specimen and data storage facilities

Subpart D—Equipment

- 58.61 Equipment design
- 58.63 Maintenance and calibration of equipment

Subpart E—Testing Facilities Operation

- 58.81 Standard operating procedures
- 58.83 Reagents and solutions
- 58.90 Animal care

Subpart F—Test and Control Articles

- 58.105 Test and control article characterization
- 58.107 Test and control article handling
- 58.113 Mixture of articles with carriers

Subpart G—Protocol for and Conduct of a Nonclinical Laboratory Study

- 58.120 Protocol
- 58.130 Conduct of a nonclinical laboratory study

Subparts H, I—[Reserved]

Subpart J—Records and Reports

- 58.185 Reporting of nonclinical laboratory study results
- 58.190 Storage and retrieval of records and data
- 58.195 Retention of records

Subpart K—Disqualification of Testing Facilities
58.200 Purpose
58.202 Grounds for disqualification
58.204 Notice of and opportunity for hearing on proposed disqualification
58.206 Final order on disqualification
58.210 Actions upon disqualification
58.213 Public disclosure of information regarding disqualification
58.215 Alternative or additional actions to disqualification
58.217 Suspension or termination of a testing facility by a sponsor
58.219 Reinstatement of a disqualified testing facility

***Subpart A—General Provisions*** This part prescribes the scope of good laboratory practices for conducting nonclinical laboratory studies that support or are intended to support applications for research or marketing permits for products regulated by the Food and Drug Administration, including food and color additives, animal food additives, human and animal drugs, medical devices for human use, biological products, and electronic products. The Definitions describe the meaning of certain terms used in the regulation such as Act, Test Article, Control Article, Nonclinical Laboratory Study, Sponsor, Testing Facility, Application for Research or Marketing Permit, Person, Test System, Specimen, Raw Data, Quality Assurance Unit, Study Director, Batch, Study Initiation Date, and Study Completion Date.

Applicability to studies performed under grants and contract fall under the provision of this regulation. An authorized FDA employee can inspect a laboratory's facility, records, and specimens at any time and without warning.

***Subpart B—Organization and Personnel*** Each personnel engaged in the conduct of or responsible for supervision of a study must possess appropriate education, training, and experience to perform the assigned tasks; the education, training and experience must be documented. Job descriptions must be available and current for each individual. The laboratory must have available the appropriate number of personnel required to perform each task as stated in the standard operating procedures (SOPs). Personnel must take all necessary precautions to avoid contamination of any testing materials.

The management of the laboratory testing facility must designate a study director; create a quality assurance unit; assure that test materials are assayed as required; make available personnel, resources, facilities, equipment, materials, and methodologies as scheduled; assure that the testing personnel clearly understand their duties; and assure that any deviations from protocols and SOPs are communicated, corrected, and documented.

The study director has the overall responsibility for the conduct of the study and is the single point of study control. The responsibilities of the study director include assuring that the protocol is approved and followed, all data are recorded and verified, unforeseen circumstances are noted and corrective action is taken and documented, test systems are specified in the protocol, GLP is followed, and documentation is archived.

GLP regulations mandate that the testing facility has a QA (Quality Assurance) unit that is separate from and independent of the personnel engaged in the conduct and direction of each study. The duties of the QA unit are extensive and include monitoring each study; assuring that the facility, equipment, personnel, methods, practices, records, and controls conform to GLP; maintaining a master schedule of duties being performed; maintaining copies of all protocols; periodically completing internal audits and submitting status reports, which include descriptions of problems noted and any corrective actions; submitting written status reports to management and the study director; ensuring no deviations occur from the approved protocols or SOPs without appropriate authorization and documentation; reviewing the final study reports; enclosing the QA report in the final study report; and preparing and maintaining QA SOPs and making documentation available for regulatory inspections.

***Subpart C—Facilities*** Each testing facility should be of suitable size and construction to facilitate the proper conduct of nonclinical laboratory studies. A separate laboratory operations area is needed. Storage space to archive specimens and data must be adequate. To avoid contamination and mix-ups between test and control articles, separate areas are needed for receipt, storage, and preparation of materials and the storage of products.

***Subpart D—Equipment*** GLP regulation mandates that equipment used in the generation, measurement, or assessment of data and equipment used for facility environmental control should be of appropriate design and adequate capacity to function according to the protocol and should be suitably located for operation, inspection, cleaning, and maintenance. Detailed written SOPs are needed for each instrument. All instrumentation must be adequately operated, inspected, cleaned, maintained, tested, calibrated, and standardized. Equipment performance, use, and maintenance are documented.

***Subpart E—Testing Facilities Operation*** A testing facility should have standard operating procedures in writing setting forth nonclinical laboratory study methods that management is satisfied are adequate to ensure the quality and integrity of the data generated in the course of a study. All deviations in a study from standard operating procedures should be authorized by the study director and should be documented in the raw data. Significant changes in established standard operating procedures should be properly authorized in writing by management. The GLP SOP states exactly how the procedure is to be done each time, every time. The SOPs must be current, clearly written, and immediately available to the staff, adhered to and authorized. A historical file of SOPs is maintained. The regulations list those procedures requiring SOPs.

All reagents and solutions in the laboratory areas should be labeled to indicate identity, titer or concentration, storage requirements, and expiration date. Deteriorated or outdated reagents and solutions should not be used.

There should be standard operating procedures for the housing, feeding, handling, and care of animals. Detailed guidelines governing animal care are mentioned in the regulation.

***Subpart F—Test and Control Articles*** Test and control materials are handled in such a way as to ensure receipt documentation, proper identification, appropriate storage, and adequate distribution processes to avoid contamination, deterioration, or damage. For each batch of test and control materials, the identity, strength, purity, composition, and stability must be determined and documented. Methods of synthesis for the materials are documented. Each batch is labeled with the material name, CAS number, batch number, and expiration date and storage conditions. If test or control articles are mixed with carriers, the uniformity, concentration, stability, and expiration date of the mixture must be determined.

***Subpart G—Protocol for and Conduct of a Nonclinical Laboratory Study*** Each study should have an approved written protocol that clearly indicates the objectives and all methods for the conduct of the study. A minimum applicable requirement of study protocol is described in the regulation. All changes in or revisions of an approved protocol and the reasons thereof should be documented, signed by the study director, dated, and maintained with the protocol.

The nonclinical laboratory study should be conducted in accordance with the protocol and documented accordingly. All data generated during the conduct of a nonclinical laboratory study, except those that are generated by automated data collection systems, should be recorded directly, promptly, and legibly in ink. All data entries should be dated on the date of entry and signed or initialed by the person entering the data. Any change in entries should be made so as not to obscure the original entry, should indicate the reason for such change, and should be dated and signed or identified at the time of the change.

***Subpart H and I—[Reserved]***

***Subpart J—Records and Reports*** The testing laboratory's final report must state the laboratory's name and address, the start and end dates of the study, the objectives and procedures of the protocol, statistical methods used, and the identification of test and control articles.

The final report should be signed and dated by the study director. Corrections or additions to a final report should be in the form of an amendment by the study director. The amendment should clearly identify that part of the final report that is being added to or corrected and the reasons for the correction or addition, and should be signed and dated by the person responsible.

Archiving is an assigned responsibility because access to the archives is restricted to authorized personnel only. Archiving must be timely and all archived data are indexed to permit expedient retrieval.

***Subpart K—Disqualification of Testing Facilities*** This part of the regulation describes the purpose and grounds for disqualification of a nonclinical testing facility. An opportunity for hearing on disqualification should be provided, followed by a final order for disqualification. Once a testing facility has been disqualified, each application for a research or marketing permit, whether approved or not, containing or relying on any nonclinical laboratory study conducted by the disqualified testing facility may be examined to determine whether such study was or would be essential to a decision.

A testing facility that has been disqualified may be reinstated as an acceptable source of nonclinical laboratory studies, to be submitted to the Food and Drug Administration if the Commissioner determines, upon an evaluation of the submission of the testing facility, that the facility can adequately assure that it will conduct future nonclinical laboratory studies in compliance with the good laboratory practice regulations set forth in this part and, if any studies are currently being conducted, that the quality and integrity of such studies have not been seriously compromised.

## 6.2 Overview of OECD Principles of GLP

The OECD published GLP principles in 1981. These were readily accepted by 30 member industrial countries including the European Union, the United States, and Japan. As a result, it became popular internationally. The OECD guidelines are on the same lines of 21 CFR Part 58. The document has been divided into two sections. Section I gives the introduction to GLP principles. It consists of a preface, scope, and definitions of important terms. Section II describes GLP principles.

### *Section I*

*Scope* The principles of GLP are applicable to the testing of chemicals for characterizing their properties and assessing their safety with respect to human safety.

*Definition of Terms* This section defines important terms concerning the organization of the test facility, the study, and test substances. The term "GLP is concerned with the organizational process and the conditions under which laboratory studies are planned performed, monitored, recorded and reported." The "test facility" means persons, premises, and operational units that are necessary for conducting the study. Study director, quality assurance program, standard operating procedures, and sponsor are some of the important terms concerning the organization of a facility.

The term "study" has been defined as an experiment or a set of experiments in which a test substance is examined to obtain data on its properties and/or its safety with respect to human health and the environment. The other important terms related to the study are study plan, OECD test guidelines, test system, specimen, and raw data. "Raw data" means all original laboratory records and documentation or verified copies thereof, which are the result of original observations and activities in a study. As per the OECD definition, "test substance" is a chemical substance or a mixture that is under investigation. The other important terms concerning the test system are reference standard, batch, vehicle, and sample.

***Section II*** Section II has been further subdivided into 10 subsections: (1) test facility organization and personnel, (2) quality assurance program, (3) facilities, (4) apparatus, material, and reagents, (5) test systems, (6) test and reference substances, (7) standard operating procedures, (8) performance of the study, (9) reporting of study results, and (10) storage and retention of records and material.

*Test Facility Organization and Personnel* This section of OECD principles of GLP describes the responsibilities of management, the study director, and personnel. It

is the responsibility of management to ensure that the principles of GLP are complied with in a test facility.

The management must designate a study director, establish a quality assurance unit, and assign personnel responsible for managing the archives. It should appoint a sufficient number of qualified personnel and provide appropriate facilities, equipment, and material. The management should ensure that the persons involved in the study are properly trained and the records of their qualification, training experience, and job description are maintained. It should ensure that all safety and health precautions are taken as per national regulations. It should also ensure that appropriate SOPs are established and followed, and amendments to the study plan are agreed upon and documented. It is also the responsibility of management to maintain copies of the study plan and historical files of SOPs.

The study director is an individual responsible for the overall technical conduct of the study, including interpretation, analysis, documentation, report generation, archiving of documents/reports, and GLP compliance. The study director should ensure that the study is conducted in accordance with the study plan. The study plan is a document that defines the entire scope of the study. It is the director's responsibility to see that any amendment/modification is done after proper authorization and is documented, stating the reason for the change. The study director is supposed to sign the final report in order to certify the validity of data and GLP compliance.

The personnel conducting the study are expected to follow safety precautions in the laboratory and handle hazardous chemicals cautiously. They should ensure health precautions as well to minimize risk to themselves and integrity of the study. Diseased and sick personnel likely to have an adverse effect on the study should be excluded from the study or a particular sensitive operation.

*Quality Assurance (QA) Program* The QA program is an internal control system designed to ascertain that the study is in compliance with the principles of GLP. This subsection deals with the general description of a quality program and the responsibilities of QA personnel. The QA program ensures that the studies performed conform to the principles of GLP. QA personnel are independent entities designated directly by the management and are expected to be familiar with the test procedures. They are supposed to report their observations directly to the management and the study director.

QA personnel should ensure that the study plan and standard operating procedures (SOPs) are accessible to the study personnel and should conduct frequent inspections of the facility and audit the study in progress to ensure that the study plan and SOPs are being adhered to. Proper documentation and records of the inspections and/or audits should be maintained by the QA unit. Any unauthorized deviation from the final plan or SOPs should be immediately brought to the attention of the study director. The QA unit should review the final report to assure that the methods, procedures, and observations are correct and that the reported results are in agreement with the raw data generated during the study. The final report should contain a statement prepared and signed by the QA unit, specifying the dates of inspection and dates of findings reported to the management and study director.

*Facilities* In general, it is required that the test facility should be of a suitable size, construction, and location to meet the study requirements. The design of the study should allow separation of different activities to avoid cross contamination and mix-ups and in turn preserve the integrity of the study. The facilities can be categorized into test system facilities, facilities for handling test and reference substances, archive facilities, and waste disposal facilities. The test system facilities should have enough rooms or separate areas to allow isolation of various test systems based on species, strain, and individual projects involving biohazardous substances. The health of the test system is an important aspect that can affect the outcome of the study, and therefore it is mandatory for the test facility to have separate areas for diagnosis, treatment, and control of disease in test systems.

There should be separate storage areas for supplies and equipment. Perishable materials should be kept under refrigeration. The storage areas should be distinctly separated from animal housing areas to prevent contamination and infestation.

It has been recommended that the test and reference substance should be isolated from the test system housing area and should be such that their identity, purity, concentration, and stability are maintained. There should be a provision for safe storage of hazardous substances. The facility should have a separate archival area for the purpose of retrieval of raw data, reports, and specimens.

The waste should be handled and disposed of in such a manner that the ongoing studies are not affected in any way. The waste disposal procedures should be consistent and meet regulatory requirements. This would include provision for collection, storage, and disposal facilities and decontamination and transportation procedures. The waste disposal/ handling activities should be properly documented and records maintained.

*Apparatus, Material, and Reagents* The apparatus used for analysis and data generation should be of suitable design and capacity and appropriately located. There should be a system of periodic calibrations, maintenance, and performance checks of all the apparatus and proper records should be maintained. These procedures should be written down in the form of SOPs. All reagents should be properly labeled indicating the source, identity, concentration, date of preparation, expiration date, and storage conditions.

*Test Systems* Any animal, plant, microbial, or other cellular, subcellular, chemical, or physical system or a combination used in the study is referred to as the test system. Test systems have broadly been categorized into physical/chemical and biological test systems. The physical or chemical data generation should be performed using well calibrated apparatus of appropriate design and capacity. The integrity of physical and chemical test systems should be verified against a highly pure reference substance. Biological systems should be housed and cared for under appropriate conditions in accordance with national regulatory requirements for the import, collection, care, and use of animals. There should be a systematic health checkup program for newly arrived systems, until which time they should be kept in isolation. If the newly arrived animals are diseased or exhibit unusual mortality or morbidity, they should be destroyed as per regulatory guidelines. Upon new arrival of animals, the source, date of arrival, and health condition upon arrival should be documented and records maintained. For obtaining reliable data, the test system should be prop-

erly acclimatized prior to study initiation. The containers housing the test systems should bear all information related to the identity of the test systems. The diagnosis and treatment of any disease before or during the study should be properly documented.

*Test and Reference Substances* A test substance is a chemical substance or a mixture under investigation. A reference substance is any well defined chemical substance or any mixture other than the test substance used to provide a basis for comparison with the test substance. Proper records of characterization, date of receipt, quantity received, and consumption should be maintained after the receipt of the substance. The procedures for handling, sampling, and storage of these substances should be laid down in relevant SOPs so as to maintain the homogeneity and stability of the substances and avoid possible contamination and mix-ups. The storage container should be labeled with the information on identity, expiration date, and storage conditions. The characterization of each substance should be done with respect to code, name, and chemical abstract number (CAS). The reference substance used in each study should also have a known identity, batch number, purity, composition, and concentration. The stability of both the test and reference substances along with the storage conditions should be known. The stability and homogeneity of the formulation in the vehicle should be established before the initiation of a study. Here the vehicle is an agent that serves as a carrier used to mix, disperse, or solubilize the test or reference substance to facilitate the administration to the test system.

*Standard Operating Procedures* Standard operating procedures (SOPs) are written procedures that describe how to perform certain routine laboratory tests or activities normally not specified in detail in study plans or guidelines. In order to ensure the quality and integrity of data, each test facility should have written SOPs approved by management. The SOPs should be immediately available in each laboratory. Published textbooks, articles, and journals may be used as supplements to these SOPs. In general, the written approved SOPs are required in the following areas:

a. Receipt, identification, labeling, handling, sampling, and storage of test and reference substances.
b. Use, maintenance, cleaning, calibration of measuring apparatus, and preparation of reagents.
c. Coding of studies, data collection, report preparation, indexing systems, handling of data including computerized data systems are some important SOPs in the area of record keeping, reporting, storage, and retrieval of data.
d. Test systems should have SOPs on (1) room preparation and environmental room conditions for the test system; (2) procedures for receipt, transfer, proper placement, characterization, identification, and care of test system; (3) test system preparation, observations, and examinations, before, during, and at termination of the study; (4) handling of test system individuals found moribund or dead during the study; (5) collection, identification, and handling of specimens including necropsy and histopathology.
e. Operation of quality assurance personnel in performing and reporting study audits, inspections, and final study report reviews.

f. SOPs on health and safety precautions as required by national and/or international legislation or guidelines.

*Performance of the Study* This section has been divided into study plan, content of study plan, and conduct of study.

The study plan is a document that defines the entire scope of the study. It should be written before the initiation of a study and retained as raw data. It should contain all changes, modifications, and revisions including justifications as agreed to, signed and dated by the study director.

The study plan should consist of the following minimum information:

a. The study should be identified by a descriptive title and a statement that reveals the nature and purpose of the study and the test substance and reference substance should be identified by code or name (IUPAC, CAS number, etc.).
b. Name and address of the sponsor, test facility, and study director.
c. The date of agreement to the study plan by signature of the study director and, when appropriate, of the sponsor and/or the test facility management along with the proposed starting and completion dates.
d. Test methods should have reference to OECD test guidelines or other test guidelines to be used.
e. Study plan should address issues (where applicable) such as the justification for selection of the test system, characterization of the test system, such as the species, strain, substrain, source of supply, number, body weight range, sex, age, and other pertinent information.
f. It should also contain the method of administration, and the reason for its choice, the dose levels and/or concentration(s), frequency, duration of administration, detailed information on the experimental design, including a description of the chronological procedure of the study, all methods, materials, and conditions, type and frequency of analysis, measurements, observations, and examinations to be performed.
g. A list of records to be retained.

Each study should be identified by a separate identification number and all study items should bear this number. The study should be conducted as per the previously compiled and approved study plan. The raw data should be entered directly, promptly, and accurately and legibly signed or initialed and dated by study personnel. The changes or corrections in the raw data should be made in such a way that the previous entry is not obscured. A reason for the change must be indicated along with the date and signature of the person making the change. The data generated directly into the computer should also indicate the person responsible for data input. Corrections in such case should be entered separately with the reason for change and the date and identity of the individual making the change.

*Reporting of Study Results* The study results should be compiled in the form of a final report, which is signed and dated by the study director. If reports from principal

scientists from various disciplines are included in the final report, they should be signed and dated by them. Corrections can be made in the final report in the form of amendments clearly specifying the reason for the change and signed and dated by the study director and principal scientists.

The final report should contain the following minimum information:

1. Identification of the Study and the Test and Reference Substances
   (a) A descriptive title
   (b) Identification of the test substance by code or name (IUPAC, CAS number, etc.)
   (c) Identification of the reference substance by chemical name
   (d) Characterization of the test substance including purity, stability, and homogeneity
2. Information Concerning the Test Facility
   (a) Name and address
   (b) Name of the study director
   (c) Name of other principal personnel having contributed reports to the final report
3. Dates
   (a) Dates on which the study was initiated and completed
4. Statement
   (a) A quality assurance statement certifying the dates inspections were made and any findings were reported to management and to the study director
5. Description of Materials and Test Methods
   (a) Description of methods and materials used
   (b) Reference to OECD test guidelines or other test guidelines
6. Results
   (a) A summary of results
   (b) All information and data required in the study plan
   (c) A presentation of the results, including calculations and statistical methods
   (d) An evaluation and discussion of the results and, where appropriate, conclusions
7. Storage
   (a) The location where all samples, specimens, raw data, and the final report are to be stored

*Storage and Retention of Records and Material* The archive should be equipped to securely store the study plan, raw data, final report, records of laboratory inspection and study audits, samples, and specimens. All archived materials should be indexed to ensure orderly storage and rapid retrieval. Movement of material into and out of the archive should be properly recorded and only persons authorized by management should have access to the archive.

## 7 SELECTION AND EXTRAPOLATION OF DOSE

### 7.1 Dose Levels

Dose levels for repeated dose studies are largely governed by toxicological findings in dose range-finding studies. Toxicokinetics is generally conducted at three dose levels [1]—low dose, intermediate dose, and high dose.

*Low Dose Level.* This is preferably the "no toxic effect" dose or "no observed adverse effect" dose level. At this dose pharmacological response is observed but no adverse effect is observed. Here the exposure in animal models should ideally be equal to or just exceed the maximum exposure in human patients.

*Intermediate Dose Level.* Exposure at the intermediate dose level should normally represent an appropriate multiple or fraction of exposure of lower or higher dose level.

*High Dose Level.* This is the highest dose that is chosen with the aim of inducing toxic effects but no mortality or severe suffering. There may be certain dose-related constraints in choosing the high dose. The toxicokinetic data may indicate that the absorption of the compound limits exposure. Under these circumstances, it should be ensured that absorption is the rate-limiting step and not increased clearance. Hence, the lowest dose producing maximum exposure should be considered as the highest dose. The limits for acceptable volumes [7] that can be administered to animals at times constrain the dose levels for solutions and suspensions.

### 7.2 Interspecies Scaling: Extrapolation of Dose

The ultimate aim of interspecies scaling of dose is to derive the maximum recommended starting dose (MRSD) for "first in human" studies. The U.S. FDA guidance [111] for "estimating safe starting dose in clinical trials for therapeutics in adult healthy volunteers" describe sit as a five step process:

*Step 1.* No observed adverse effect level (NOAEL) determination.
*Step 2.* Human equivalent dose (HED) calculation.
*Step 3.* Most appropriate species selection.
*Step 4.* Application of safety factor.
*Step 5.* Consideration of the pharmacologically active dose (PAD).

NOAEL is the highest dose level achieved in an animal toxicity study that dose not produce a significant increase in adverse effect. The NOAEL is generally the accepted benchmark for safety and serves as a starting point for the calculation of a safe starting human dose, which is subsequently converted to HED. It is a usual assumption that after systemic administration of therapeutics to animals, toxic endpoints such as maximum tolerated dose (MTD) and NOAEL scale well between various species when doses are normalized to body surface area ($mg/m^2$). The following equation is applied for interspecies scaling:

$$\text{HED} = \text{Animal NOAEL} \times (W_{\text{animal}}/W_{\text{human}})^{1-b}$$

The derivation of the above equation is not within the scope of this chapter and is reported elsewhere [111]. Conventionally, for body surface area (mg/m$^2$), normalization of the value of the allometric exponent $b$ would be 0.67 as originally derived by Freireich et al. [112] and Schein et al. [113]. Although subsequent analysis showed that $\boldsymbol{W}^{0.75}$ was a better normalization factor, the use of body surface area ($\boldsymbol{W}^{0.67}$) and normalization to body surface area has remained a widespread practice for extrapolation of animal dose to humans. The guidance document recommends the conversion of NOAEL dose to HED based on the body surface area correction factor $\boldsymbol{W}^{0.67}$ for selection of a starting dose in initial studies on healthy human volunteers. An analysis conducted to address the effect of body weight on actual body surface area correction factor (BSA CF) demonstrated that it provides a reasonable estimate of HED over a broad range of human and animal weights. Table 2 shows the standard interspecies conversion factors as recommended by the CDER and CBER. Any deviation from the body surface area approach needs a proper justification. An alternative approach for determining initial clinical dose would be the use of animal pharmacokinetics and modeling rather than body surface area. However, in the initial stages of an IND study, animal data are not sufficient to construct a scientifically valid model. There are occasions when scaling based on body weight (mg/kg)

**TABLE 2 Conversion of Animal Doses to Human Equivalent Doses (HEDs) Based on Body Surface Area**

| Species | To Convert Animal Dose in mg/kg to Dose in mg/m$^2$, Multiply by | To Convert Animal Dose in mg/kg to HED[a] in mg/kg, either: | |
|---|---|---|---|
| | | Divide Animal Dose by | Multiply Animal Dose by |
| Human | 37 | — | — |
| Mouse | 3 | 12.3 | 0.08 |
| Hamster | 5 | 7.4 | 0.13 |
| Rat | 6 | 6.2 | 0.16 |
| Ferret | 7 | 5.3 | 0.19 |
| Guinea pig | 8 | 4.6 | 0.22 |
| Rabbit | 12 | 3.1 | 0.32 |
| Dog | 20 | 1.8 | 0.54 |
| *Primates* | | | |
| Monkeys[b] | 12 | 3.1 | 0.32 |
| Marmoset | 6 | 6.2 | 0.16 |
| Squirrel monkey | 7 | 5.3 | 0.19 |
| Baboon | 20 | 1.8 | 0.54 |
| Micro-pig | 27 | 1.4 | 0.73 |
| Mini-pig | 35 | 1.1 | 0.95 |

[a]Assumes 60 kg human. For species not listed or for weights outside the standard ranges, human equivalent dose can be calculated from the formula:

$$\text{HED} = (\text{Animal dose in mg/kg}) \times (\text{Animal weight in kg/human weight in kg})^{0.33}.$$

[b]For example, cynomolgus, rhesus, and stumptail.

instead of body surface area may be appropriate. In order to consider mg/kg scaling, the available data should show that NOAEL occurs at a similar dose across species. If only two NOAELs from toxicity studies in separate species are available, one of the following criteria should be met:

1. After oral dose administration, the dose is limited by local toxicity.
2. The toxicity in humans is dependent on the exposure parameter (i.e., $C_{max}$) that is highly correlated with dose across species.
3. Other pharmacological and toxicological endpoints such as MTD and lowest lethal dose also scale between species by mg/kg.

However, mg/kg scaling gives a 12-, 6-, and 2-fold higher HED than the default $mg/m^2$ approach for mouse, rat, and dog, respectively. Therefore, if the above stated factors are not complied with, the body surface area scaling would be a safer approach.

There are other exceptions to the body surface area approach such as:

1. Therapeutics administered by alternative routes, such as topical, intranasal, subcutaneous, and intramuscular, for which dose is limited by local toxicity, should be normalized to concentration (mg/area of application) or amount (mg) of drug at application site.
2. Administration of drugs into anatomical compartments such as intrathecal, intranasal, intraocular, intrapleural, and intraperitoneal, with little distribution outside the compartment, should be normalized according to the volume of the compartment and the concentration of the drug.
3. Intravenously administered biological products with molecular weight >100,000 daltons should be mg/kg.

When there is no knowledge on the relevant species, by default the most sensitive species with the lowest HED should be chosen for MRSD calculation. Following are factors that influence the choice of appropriate species instead of using the most sensitive species.

1. Differences in ADME between species. While determining MRSD for the Phase I clinical trial, the ADME for humans is not known. But comparative *in vitro* data might be available. These data are relevant when there are marked differences in both the *in vivo* metabolic profiles and HEDs in animals.
2. Class experience may indicate a particular model is predictive of human toxicity.
3. There is limited biological cross-species pharmacologic reactivity of the therapeutics.
4. If a species is most sensitive but is different in terms of physiology as compared to humans, the species may not be the most appropriate one for MRSD calculations.

The application of a safety factor provides a margin of safety for protection of human subjects receiving the initial clinical dose. It protects from uncertainties due

to enhanced sensitivity to therapeutic activity in humans versus animals, difficulties in detecting certain toxicities in animals (e.g., headache, myalgias, mental disturbances), differences in receptor densities or affinities, unexpected toxicities, and interspecies differences in absorption, distribution, metabolism, and excretion of the therapeutic. These differences may be accommodated by lowering the human starting dose from the HED of the selected species NOAEL.

A default safety factor of 10 is generally recommended for adequate protection of human volunteers. The safety factor can be raised in case of increased safety concerns such as:

*Steep Dose–Response Curve.* Steep dose–response curves for significant toxicities in the most appropriate species or in multiple species may indicate a greater risk to humans.

*Severe Toxicities.* Qualitatively severe toxicities or damage to an organ system (e.g., central nervous system (CNS)) indicate increased risk to humans.

*Nonmonitorable Toxicity.* Nonmonitorable toxicities may include histopathologic changes in animals that are not readily monitored by clinical pathology markers.

*Toxicities Without Prodromal Indicators.* If the onset of significant toxicities is not reliably associated with premonitory signs in animals, it may be difficult to know when toxic doses are approached in human trials.

*Variable Bioavailability.* Widely divergent bioavailability in several species, with poor bioavailability in the test species used to derive the HED, suggests a greater possibility for underestimating the toxicity in humans.

*Irreversible Toxicity.* Irreversible toxicities in animals suggest the possibility of permanent injury in human trial participants.

*Unexplained Mortality.* Mortality that is not predicted by other parameters raises the level of concern.

*Large Variability in Doses or AUC Levels Eliciting Effect.* When doses or exposure levels that produce a toxic effect differ greatly across species, the ability to predict a toxic level in humans is reduced and a greater safety factor may be called for.

*Questionable Study Design or Conduct.* Poor study design or conduct casts doubt on the accuracy of the conclusions drawn from the data. For instance, few dose levels, wide dosing intervals, or large differences in responses between animals within dosing groups may make it difficult to characterize the dose–response curve.

*Novel Therapeutic Targets.* Therapeutic targets that have not been previously clinically evaluated may increase the uncertainty of relying on the nonclinical data to support a safe starting dose in humans.

*Animal Models with Limited Utility.* Some classes of therapeutic biologics may have very limited interspecies cross reactivity or pronounced immunogenicity, or may work by mechanisms that are not known to be conserved between (nonhuman) animals and humans; in these cases, safety data from any animal studies may be very limited in scope and interpretability.

A safety factor lower than 10 can be used for drugs of well characterized class, same route of administration, same schedule and duration of administration, similar metabolic profile and bioavailability, and similar toxicity across species including human. A smaller safety factor might also be used when toxicities of drugs are easily monitored, reversible, and predictable and exhibit a moderate to shallow dose–response relationship with toxicities that are qualitatively and quantitatively consistent across the tested species. When NAOEL is based on safety studies of longer duration than proposed clinical schedule, a greater margin of safety is built into NOAEL and hence a factor lower than 10 is justified.

## 8 ROUTE OF DOSE ADMINISTRATION

The route of administration in animals should be similar to the intended clinical route [1]. Adoption of an alternative route should be based on pharmacokinetic considerations. A comparison of systemic exposure and metabolic profile of the investigational compound should be made by both routes of administration. If there is a significant change in pharmacokinetic parameters, a rigorous safety evaluation is required. When similar metabolism and exposure can be demonstrated by the alternative route, then nonclinical toxicity studies may focus only on local toxicity.

## 9 SAMPLING APPROACHES

The number of sampling time points should be frequent enough to estimate exposure but not so frequent as to cause physiological stress to animals. The justification of sampling time points should be based on kinetic data obtained from previous pilot studies or dose range-finding studies in similar animal models. Usually three or more animals are used to assess the concentration at each time point. The time points can be analyzed individually or after pooling.

### 9.1 Serial Sampling Approach

Systemic exposure is generally determined in terms of area under plasma/serum concentration versus time curve (AUC) [114] and $C_{max}$. This requires withdrawal of 10–20 samples [115] from the animal at different time points. Withdrawal of multiple samples from the same experimental animal at different time points for drug exposure assessment is reffered to as serial sampling. Withdrawal of multiple samples from the same animal is feasible in large animals such as dog and monkey. However, blood sampling in rodents is restricted due to low blood volume [116]. Moreover, the trauma and stress caused by frequent venipuncture and blood loss could cause adverse changes in the animal physiology and ultimately the pharmacokinetics of the drug. A satellite group of animals (4–5 animals/dose) is recommended for rodent toxicokinetics [1], and these animals undergo exactly the same treatment as those in the main study. The approach requires more animals, greater amount of drug, and increased personnel for dose administration and veterinary care. In addition, there may be lack of correlation in plasma drug concentration and toxicity findings.

### 9.2 Sparse Sampling Approach

To overcome these problems, several investigators [116–118] suggested the use of the sparse sampling approach (SSA) within the toxicity study itself. The basic characteristic of the SSA is that a complete profile is not sampled in every animal; that is, serial sampling is not done. Rather, each animal contributes to the profile with just one or two samples. Hence, blood samples can be generated directly from animals in toxicity studies as one or two samples will not effect the toxicity findings [116, 119], thereby precluding the need for separate toxicokinetic studies. This also enables direct correlation of plasma drug concentrations and toxicity findings within the same study and hence leads to a more efficient preclinical drug development program.

## 10 ANALYTICAL METHODOLOGY

All analytical methods should be developed and validated before the initiation of a toxicokinetic study. The methods should possess the desired sensitivity, precision, and accuracy and should be selective for the analyte with no interference from the endogenous components of the matrix. It is also important to study the stability of the analyte in the matrix.

### 10.1 Bioanalytical Method

The choice of biological should be continuously reviewed and methods should be developed in the adequate matrices—blood, plasma, serum, urine, and so on. If tissue distribution is also incorporated in the study, then it is required to develop methods in all organs of interest—for example, brain, heart, liver, kidney, adipose tissue, and skeletal muscles—and properly validate them. There should be clarity about the type of analyte to be analyzed—be it the parent compound, the active or inactive metabolites, or the enantiomers. Generally, the administered dose range in a toxicity study is broad, ranging from a lower dose to a higher toxic dose level. Therefore, the linearity curve should be capable of covering all the doses, or there can be two linearity curves—one for the low dose levels and another for the higher dose levels. For higher dose levels for which the concentrations falls beyond the upper limit of quantitation, a dilution integrity test should be done and, if convenient, the samples should be diluted appropriately to bring them within the linearity range. Later, these concentrations can be back calculated based on the dilution factor.

### 10.2 Bioanalytical Method Development

This general strategy is applicable to high-pressure liquid chromatography bioanalytical methods performed on drugs and metabolites present in biological matrices, such as blood, serum, plasma, or urine.

**Formula Goals to Be Achieved**

1. LLOQ. Minimum quantifiable concentration of drug and/or metabolites in biological fluid (approximately 20-fold less than reported $C_{max}$ values).

2. Concentration range. Linearity is preferred: a concentration range from LLOQ to a value greater than $C_{max}$ is expected in the study (preferably 10-fold more than the value of $C_{max}$ in reported values). Calibration standards in the 1, 2, 5, 10, 20, 50, … pattern of concentration are preferable. The shape of the calibration curve should be defined in mathematical terms.
3. Number of samples for calibration curve. At least seven or eight nonzero standards.
4. Strategy.
   a. Collect information on physicochemical properties of the drug and/or metabolites—structure, solubility, and stability in solution (including photostability).
   b. Check literature for information on sample preparation and chromatographic methods of compounds with closely related chemical structure.
   c. Select appropriate internal standard that is structurally similar.
   d. Select an appropriate detector—variable wavelength detector, photodiode array detector, fluorescence detector, and or electrochemical detector, depending on the sensitivity and selectivity required.
   e. Chromatography.
      Select suitable detecting wavelengths (the most sensitive wavelength of the drug is preferable—more than 250 nm in the case of UV—and the most sensitive excitation and emission wavelengths in the case of fluorescence).
      Check columns with different physical properties and/or from different suppliers (preferably with $C_{18}$ or $C_8$).
      Check for pH effects (pH range 2–7).
      Optimize flow rate.
      Optimize mobile phase composition to obtain optimal retention (5–14 min) and peak shape.
      Select appropriate injection volume (20–100 μL).
   f. Sample preparation. Select an appropriate method—dilution, protein precipitation, liquid–liquid extraction, or solid phase extraction.
      *Protein precipitation.* Suitable for highly water-soluble drugs. Precipitating agents used are perchloric Acid (5–28% solutions), trichloroacetic acid (5–25% solutions), and acetonitrile (two- to fivefold of sample volume).
      *Liquid–liquid extraction.* Based on distribution coefficient of analytes, the preferable organic solvents employed are methanol, acetonitrile, diethyl ether, dichloromethane, chloroform, hexane, *t*-butyl methyl ether, and toluene, and/or their appropriate mixtures.
      *Solid phase extraction.* Based on the partition coefficient of analytes, a cartridge having suitable sorbent is selected ($C_{18}$, $C_8$, etc.). A typical solid phase extraction (SPE) procedure consists of the following steps: (i) condition and equilibrate the cartridge bed, (ii) load the sample, (iii) wash off interferences, (iv) elute analyte(s) of interest, and (v) dry the elution solvent in a hot water bath under a stream of nitrogen air.

g. Prepare a spiked sample. Elution or reconstitution solvent must be compatible with HPLC mobile phase.
h. Check recovery and optimize sample preparation procedure.
i. Reoptimize chromatography, if necessary.
j. Assay performance.

Analyze a solution of analyte(s) and internal standard in the mobile phase, $n \geq 5$, to determine the robustness of the spectrometric detection. Prepare one spike sample and analyze at least three times to determine the robustness of the chromatographic method. Prepare at least five spike samples of the same concentration and analyze to determine the robustness of the total method. Partly validate the method for linearity of spiked and nonspiked standards and their recovery.

## 10.3 Bioanalytical Method Validation

In order to define the various prestudy validation parameters for a validating bioanalytical [5] method for toxicokinetic studies, the following parameters should be assessed during prestudy validation:

- Specificity,
- Calibration curve
- Accuracy and precision
- Recovery (of analyte and the internal standard)
- Stability

This procedure is applicable to high-pressure liquid chromatography analytical methods performed on drugs and/or metabolites obtained from biological matrices such as blood, serum, plasma, or urine.

***Specificity*** At least six randomly selected control blank biological matrix samples should be chromatographed to determine the extent to which endogenous biological matrix components may contribute to interference at the retention time of analyte and internal standard. Any interference should be compared with the response of the LLOQ and internal standard. Chromatograms should be evaluated to make sure that there is no interference at the retention times of analyte and internal standard. Any interference is compared with the analyte response of the LLOQ. For hyphenated mass spectrometry based methods, testing six independent matrices for interference may not be important. Any significant interference peak corresponding to analyte and internal standard peak should not be present. The significance of the interference peak should be considered if it meets the following criteria:

- The retention time of the interference peak corresponds to the analyte peak and the internal standard peak.

- The response of interference peak is greater than or equal to 20% of the response of LLOQ.
- The response of the interference peak is greater than or equal to 5% of the response of the internal standard.

***Linearity*** For the establishment of linearity, a series of calibration samples should be prepared, by spiking a known concentration of drug with a working standard solution. The calibration curve should consist of a blank sample (matrix sample processed without IS and analyte), a zero sample (matrix sample processed only with IS), and six to eight nonzero samples covering the range of expected concentrations.

A minimum of four linearity curves containing not less than six nonzero concentrations should be analyzed.

The factors calculated from the calibration curve should match the following acceptance criteria:

- The deviation of the LLOQ from nominal concentration should not be more than 20%.
- The deviation of the standards, other than LLOQ, from nominal concentration should not be more than 15%.
- A minimum of six nonzero standards should meet the above criteria, including the LLOQ and the calibration standard at the highest concentration.
- The correlation coefficient should be more than 0.95.

***Accuracy and Precision*** The precision of an analytical method describes the closeness of individual measures of an analyte when the procedure is applied repeatedly to multiple aliquots of a single homogeneous volume of biological matrix.

- The accuracy of an analytical method describes the closeness of test results obtained by the method to the true value of the analyte.
- Within-batch accuracy and precision evaluations should be performed by replicate analysis of the same concentration of drug in biological matrix. The run should consist of a calibration curve and six replicates of each LLOQ, low, medium, and high quality control samples, that is, precision and accuracy (PA) batches.
- The between-batch accuracy and precision are assessed by analysis of a minimum of four precision and accuracy batches.
- For the evaluation of precision, the %CV of each concentration level should not be more than 15%, except for the LLOQ, for which it should not be more than 20%. The same is true for accuracy: the mean value should not deviate by more than ±15% of the actual concentration, except for the LLOQ, where it should not deviate by more than ±20% of the actual concentration. At least 67% of QC samples should be within 15% of their respective nominal (theoretical) values; 33% of the QC samples (not all replicates at the same concentration) can be outside ±15% of the nominal value.

***Recovery*** The recovery of an analyte in an assay is the detector response obtained from an amount of the analyte added to and recovered from the biological matrix, compared to the detector response obtained for the pure authentic standard. Recovery pertains to the extraction efficiency of an analytical method within the limits of variability.

Recovery of drug in biological matrix should be evaluated by comparing the mean peak responses of a minimum of three samples of low, medium, and high quality control samples, prepared in biological matrix, to mean peak responses of a minimum of three samples prepared in diluent. Recovery of internal standard in the biological matrix should be evaluated by comparing the mean peak responses of a minimum of three quality control samples, prepared in biological matrix, to mean peak responses of a minimum of three samples of the same quality control samples prepared in diluent. Recovery of the analyte need not be 100%, but the extent of recovery of an analyte and of the internal standard should be consistent, precise, and reproducible.

*Dilution Integrity* A minimum of three replicates of each low and high QC sample should be used to prepare the required concentration by multiple aliquoting of their actual concentration. Then the samples should be diluted with biological matrix to meet their actual concentration. The samples should be analyzed and the results should be compared against the theoretical values. The percent change in concentration of the prepared samples from the theoretical value should not be more than ±15%.

*Interaction with Anticoagulant* Interaction of anticoagulant with the drug is established as the vacutainers are coated with anticoagulant. A minimum of three replicates of each low and high QC sample should be prepared in the subject sample collection culture tube having an anticoagulant coating and be kept in a deep freezer at –70 °C until completion of the stability testing period. After completion of the stability testing period, the samples should be processed along with a precision and accuracy batch and compared against the theoretical values. The percent change in concentration of the prepared samples from the theoretical value should not be more than ±15%.

***Stability*** Drug stability in a biological fluid is a function of the storage conditions, the chemical properties of the drug, the matrix, and the container system.

*Freeze and Thaw Stability* Testing for freeze and thaw cycles on stability should be determined after three freeze and thaw cycles. At least three aliquots of each of the low and high concentrations should be stored at the same storage conditions as the study samples for 24 hours and thawed unassisted at room temperature. When completely thawed, the samples should be refrozen for 12–24 hours under the same conditions. The freeze–thaw cycle should be repeated two more times and then analyzed on the third cycle, and results should be compared against the theoretical value of QC sets. If there is a remarkable instability, the test should be repeated with one or two cycles. The percent change in concentration of the stability samples should not be more than ±15% from the theoretical value.

*Short-Term Stability of Samples* A minimum of three aliquots each of the low and high QC samples should be kept at room temperature (25 ± 5 °C) for 4–24 hours after spiking into biological matrix. After completion of the stability testing period, the samples should be analyzed and compared against the theoretical value of QC sets. If there is a remarkable instability, the test should be performed for a suitably shorter duration. The percent change in concentration of the stability samples should not be more than ±15% from the theoretical value.

*Long-Term Stability* The storage time in long-term stability evaluation should exceed the time difference between the date of first sample collection and the date of the last sample analysis. Long-term stability should be determined by storing at least three aliquots of each of the low and high concentrations or near to them under the same conditions as the study samples. The samples should be analyzed and compared to the theoretical value of QC sets. The test should be performed periodically to establish long-term stability from suitable duration to optimum duration. The percent change in concentration of the stability samples should not be more than ±15% from the theoretical value.

*Autosampler Stability* The stability of processed samples in the autosampler should be determined at the autosampler temperature over the anticipated run time for the batch size to be used in studies. The minimum of three QC sets of low and high concentrations should be kept in the autosampler for the expected duration and analyzed. The drug concentrations obtained should be compared against their theoretical value. The percent change in concentration of the stability samples should not be more than ±15% from the theoretical value.

*Stock Solution Stability* The stock solutions of drug and the internal standard stability under suitable conditions such as ambient temperature (20–30 °C), or refrigerated (2–8 °C) should be established by comparing the response or ratio of drug and internal standard from the system suitability chromatogram area of the first day of preparation of the solution and the system suitability chromatogram area of the last day of validation or any other relevant analysis. The duration of stock solution stability can be established, if required, by periodic analysis of system suitability. The percent change in area or the area ratio of the drug and internal standard of the stability samples should not be more than ±15% from the theoretical value.

*Dry Extract Stability* Dry extract stability should be determined by storing at least three aliquots of each of the low and high concentrations, after processing without reconstitution under the same conditions as the study samples (at about –70 °C in a deep freezer). After completion of the anticipated stability testing period, the samples should be reconstituted and analyzed and results should be compared against the theoretical value of the QC set. The percent change in concentration of the stability samples should not be more than ±15% from the theoretical value.

***Partial Validation*** Partial validations are modifications of already validated bioanalytical methods. Partial validation should be performed after analyzing a minimum of one precision and accuracy batch. Partial validation can range from as

little as one PA batch to nearly full validation based on requirement. Some typical bioanalytical method changes that require partial validation are:

- Bioanalytical method transfers between laboratories or analysts.
- Change in analytical methodology (e.g., change in detection systems).
- Change in anticoagulant in harvesting biological fluid.
- Change in matrix within species (e.g., human plasma to human urine).
- Change in sample processing procedures.
- Change in species within matrix (e.g., rat plasma to mouse plasma).
- Change in relevant concentration range.
- Changes in instruments and/or software platforms.
- Limited sample volume (e.g., pediatric study).
- Rare matrices.
- Selectivity demonstration of an analyte in the presence of concomitant medications.
- Selectivity demonstration of an analyte in the presence of specific metabolites.

***Data Handling*** The regression equation and the correlation coefficient should be calculated by performing a linear regression analysis according to the least squares method of the peak area response versus analyte concentration.

## 10.4 Method for the Formulation

In order to ensure proper dose administration, it is very important to establish the stability data for the formulation used in the study. At the preclinical stage, the formulation is generally very simple with minimum excipients. Several strengths are required to cover the broad range of dose levels. Therefore, scientists prefer to prepare them more frequently depending on the requirement. The minimum stability period should cover the time required for dosing all the animals on one particular day. In the case where a longer stability period is achieved, the formulation can be prepared accordingly. The formulation cannot be used beyond the established stability period, or if stored in conditions (temperature, humidity, and packing/container) other than established during stability studies. The formulation should be assayed on the first and last day of the study with some intermediate points in order to ascertain that the right dose was administered during the study. Therefore, it is mandatory to develop and validate a sensitive stability indicating method for the formulation.

## 10.5 Method Validation for the Formulation

Validation [120, 121] should be performed for each chemical species to be measured in the placebo/vehicle. Typical performance parameters that should be assessed during method validation include: accuracy, linearity range, limit of quantitation, precision, ruggedness, robustness, and specificity.

***Accuracy*** Accuracy is the measure of how close the experimental value is to the true value. Accuracy studies for drug substance and drug product are performed at the 80%, 100%, and 120% levels of label claim. For the drug product, this is performed frequently by the addition of known amounts of drug by weight or volume (dissolved in diluent) to the placebo formulation working in the linear range of detection of the analyte. This would be a true recovery for liquid formulations. This test evaluates the specificity of the method in the presence of the excipients under the chromatographic conditions used for the analysis of the drug product. It will pick up recovery problems that could be encountered during the sample preparation and the chromatographic procedures. Recovery data should be assayed at least in triplicate, at each level (80%, 100% and 120% of label claim). The RSD of the replicates will provide the analysis variation or how precise the test method is. The mean of the replicates (expressed as % label claim) indicates how accurate the test method is.

***Linearity Range*** The linear range of detectability that obeys Beer's law is dependent on the compound analyzed and the detector used. The working sample concentration and samples tested for accuracy should be in the linear range. The linearity range for an assay method will be evaluated by spiking the test compound to placebo formulation with 40%, 60%, 100%, 120%, and 140% of label claim. The regression coefficient ($r$) should be about 0.999. Intercept and slope should be indicated.

***Limit of Quantitation (LOQ)*** Quantitation limit is the lowest concentration of analyte in a sample that can be determined with acceptable precision and accuracy under the stated experimental conditions. The limit of quantitation is expressed as the concentration of the analyte in the sample. LOQ is the concentration at which the signal-to-noise ratio is 10:1.

***Precision*** Precision is the measure of how close the data values are to each other for a number of measurements under the same analytical conditions.

- *Injection Repeatability.* The precision as measured by multiple injections or a system suitability test of a homogeneous sample (prepared solution) indicates the performance of the HPLC instrument under the chromatographic conditions and day tested. An RSD of 1% RSD for precision of the system suitability tests for at least five injections ($n = 5$) for the active drug product is desirable.
- *Analysis Repeatability.* Determination, expressed as the RSD, consists of multiple measurements of a sample by the same analyst under the same analytical conditions on the same day (intraday). An RSD of 2% RSD for precision of the system suitability tests for 80%, 100%, and 120% levels of label claim for at least two determinations is desirable. For practical purposes, it is often combined with accuracy and carried out as a single study.

***Ruggedness*** This attribute evaluates the reliability of the method in a different environment other than that used during development of the method. The objective is to ensure that the method will provide the same results when similar samples are

analyzed once the method development phase is over. Depending on time and resources, the method will be tested by conducting the same assay on multiple days, with multiple analysts and instruments. Intermediate precision in the test method can be partly assured by good system suitability specifications.

***Robustness*** Robustness is a measure of the method's capability to remain unaffected by small but deliberate variations in method parameters. Robustness can be partly assured by good system suitability specifications. Thus, it is important to set tight, but realistic, system suitability specifications. Robustness of the method will be evaluated by varying some or all conditions, for example, columns, flow rate, slight change in mobile phase composition, pH of buffer in mobile phase, and reagents.

***Specificity*** The analyte should have no interference from other extraneous components and should be well resolved from them. A representative HPL chromatogram or profile should be generated to show that the extraneous peaks from either addition of known compounds or samples from stress testing (thermal degradation at 55 °C for 3 hours) are baseline resolved from the parent analyte. The photodiode array detector will determine peak purity.

***Acceptance Criteria*** An analytical method can be considered fully validated when it meets the following criteria:

- Accuracy: RSD should not be more than 2% at all three concentrations.
- Linearity: The regression coefficient ($r$) should be about 0.999 for the linearity range.
- Limit of quantitation (LOQ) : Signal-to-noise ratio should be 10:1.
- Precision: Injection repeatability—RSD of 5 replicate injections should not be more than 1%. Analysis repeatability—RSD for two determinations on the same day at 80%, 100%, and 120% levels of label claim should be no more than 2%.
- Ruggedness: RSD for two determinations on two different days at 80%, 100%, and 120% levels of label claim by different analysts and on two different instruments should be no more than 2%.
- Robustness: Changes in column, flow rate, or a slight change in mobile phase should not significantly affect the % Assay of the same formulation.
- Specificity: There should be no interfering peak below the main peak for stressed and nonstressed samples. Peak purity should not be less than 0.99.

## 10.6 Method Development Issues for Biotech Products

Many bioanalytical validation parameters are also applicable to microbiological and ligand-binding assays for biotech products also. However, these assays possess some selectivity and quantitation issues that need to be considered along with method development and validation.

***Selectivity Issues*** Similar to chromatographic methods, selectivity is important for microbiological and ligand-binding assays also. Here selectivity needs to be shown

in terms of (1) interference from substances physicochemically similar to the analyte and (2) interference from matrix components that are unrelated to the analyte.

1. Interference from Substances Physicochemically Similar to the Analyte
   (a) Cross-reactivity of metabolites, concomitant medications, or endogenous compounds should be evaluated individually and in combination with the analyte of interest.
   (b) When possible, the immunoassay should be compared with a validated reference method (such as LC-MS) using already analyzed samples and predetermined criteria for agreement of accuracy of the immunoassay and the reference method.
   (c) The dilution linearity to the reference standard should be assessed using study (incurred) samples.
   (d) Separation steps can be incorporated prior to the immunoassay in order to improve selectivity.

2. Matrix Effects Unrelated to the Analyte

   (a) The standard curve should be prepared in biological fluids and should be compared with standard in buffer to rule out matrix effects.
   (b) Parallelism of diluted study samples should be evaluated with diluted standards to rule out any matrix effects.
   (c) Nonspecific binding should be determined.

***Quantification Issues*** Microbiological and imunoassay standard curves are inherently nonlinear and, in general, more concentration points may be required to define the standard curve [5]. In addition, the response–error relationship for immunoassay standard curves is a nonconstant function of the mean response. Therefore, a minimum of six nonzero calibrator concentrations, run in duplicate, is recommended. The concentration–response relationship is most often fitted to a four- or five-parameter logistic model, although others may be used with suitable validation. The overall curve fit can be improved by employing anchoring points in the asymptotic high- and low-concentration ends of the standard curve. Generally, these anchoring points will be at concentrations that are below the established LLOQ and above the established ULOQ. Calibrators should be prepared in the same matrix as the study samples or in an alternate matrix of equivalent performance. Both ULOQ and LLOQ should be defined by acceptable accuracy, precision, or confidence interval criteria based on the study requirements. The accuracy can be improved by the use of replicate samples. In the case where replicate samples should be measured during the validation to improve accuracy, the same procedure should be followed as for unknown samples.

The following recommendations apply to quantification issues:

- If separation is used prior to assay for study samples but not for standards, it is important to establish recovery and use it in determining results. Efficiency and reproducibility of recovery can be achieved by (1) the use of radiolabeled tracer analyte (quantity too small to affect the assay), (2) the advance establish-

ment of reproducible recovery, and (3) the use of an internal standard that is not recognized by the antibody but can be measured by another technique.

- Key reagents, such as antibody, tracer, reference standard, and matrix, should be characterized appropriately and stored under defined conditions.
- Assessments of analyte stability should be conducted in true study matrix and not in a matrix from which endogenous interference has been removed.
- Acceptance criteria: At least 67% (4 out of 6) of QC samples should be within 15% of their respective nominal value; 33% of the QC samples (not all replicates at the same concentration) may be outside 15% of nominal value. In certain situations, wider acceptance criteria may be justified.
- Assay reoptimization or revalidation may be important when there are changes in key reagents, as follows:

  *Labeled Analyte (Tracer).* Binding should be reoptimized and performance should be verified with standard curve and QCs.

  *Antibody.* Key cross-reactivities should be checked and tracer experiments above should be repeated.

  *Matrix.* Tracer experiments above should be repeated and method development experiments should include a minimum of six runs conducted over several days,with at least four concentrations (LLOQ, low, medium, and high) analyzed in duplicate in each run.

## 11 CONCLUSION

Toxicokinetic evaluation is a small but vital part of the drug development process and plays a crucial role in deciding the fate of a drug candidate. Reliable toxicokinetic evaluation should consist of a correct study design, adequate sampling, sensitive and robust analytical methodology, accurate raw data, proper interpretation of toxic response versus exposure of drug and/or metabolites, and overall GLP compliance.Toxicokinetics provides a proof of concept for animal exposure to appropriate concentrations of administered drug or its metabolites. It brings in better understanding of toxicity data in terms of pharmacokinetic and metabolic principles and facilitates better extrapolation of preclinical data for clinical use. It is important to generate toxicokinetic data in safety pharmacology, repeated dose toxicity studies, genotoxicity studies, carcinogenicity studies, and reproductive toxicity studies. The primary concern of generating toxicokinetic data is to determine the plasma and tissue levels of the parent compound in the toxicology species; however, measurement of metabolite concentrations in plasma or other body fluids may be required for "prodrug," extensively metabolized drugs, or when one or more pharmacologically or toxicologically active metabolites are formed. The drug bound to plasma proteins is not available for distribution, hepatic metabolism, and renal elimination and is therefore pharmacologically inactive. Drug-induced toxicity is often species dependent due to species differences in pharmacokinetics, which can reduce the confidence in preclinical toxicity data. Other factors to be considered are sex differences in kinetics. Prolonged plasma half-life, reduced clearance, inhibition of drug metabolizing enzymes, or enterohepatic recirculation cause accumulation of drug

after repeated administration. On the other hand, first-pass metabolism or autoinduction of drug metabolizing enzymes may cause reduced drug levels. Lack of dose proportionality also has clinical relevance and may occur either due to autoinduction of drug metabolizing enzymes or because of saturation of mechanisms such as intestinal absorption, enzyme metabolizing systems, and tubular reabsorption.

## ACKNOWLEDGMENT

The authors is thankful to Shriprakash Singh, group leader at Ranbaxy Laboratories, Gurgaon, India, for his valuable contribution in compiling the GLP guidelines and in editing of manuscript for this chapter.

## REFERENCES

1. ICH Topic S3A, CPMP/ICH/384/95: Operational June 1995. *Toxicokinetics: A Guidance for Assessing Systemic Exposure in Toxicology Studies*. (See http://www.emea.eu.int.)
2. Baldrich P. Toxicokinetics in preclinical evaluation. *Drug Discov Today* 2003;8(3): 127–133.
3. *Federal Register*—FDA 21 CFR Part 58.
4. OECD GLP guidelines 1997.
5. *Guidance for Industry: Bioanalytical Method Validation*. US Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research (CDER), Center for Veterinary Medicine (CVM), May 2001. www.fda.gov/cder/guidance/4252fnl.htm.
6. ICH Topic S7A, CPMP/ICH/539/00: Operational June 2001. *Safety Pharmacology Studies for Human Pharmaceuticals*. (See http://www.emea.eu.int.)
7. Diehl KH, Hull R, Morton D, Pfister R, Rabemampianina Y, Smith D, Vidal JM, van de Vorstenbosch C. A good practice guide To the administration of substances and removal of blood, including routes and volumes. *J Appl Toxicol* 2001;21:15–23.
8. ICH Topic S2B, CPMP/ICH/174/95: Operational March 1998. *Genotoxicity: A Standard Battery of Genotoxicity Testing of Pharmaceuticals*. (See http://www.emea.eu.int.)
9. ICH Topic S2A, CPMP/ICH/141/95: Operational April 1996. *Genotoxicity: Guidance on Specific Aspects of Regulatory Genotoxicity Tests for Pharmaceuticals*. (See http://www.emea.eu.int.)
10. ICH Topic S1A. *Guideline on the Need for Carcinogenicity Studies of Pharmaceuticals*. International Conference on Harmonization on Technical Requirements for Registration of Pharmaceuticals for Human Use. Harmonized Tripartite Guideline; 1995.
11. ICH Topic S1B: *Testing for Carcinogenicity of Pharmaceuticals*. International Conference on Harmonization on Technical Requirements for Registration of Pharmaceuticals for Human Use. Harmonized Tripartite Guideline; 1996.
12. ICH S1C: *Dose Selection for Carcinogenicity Studies of Pharmaceuticals*. International Conference on Harmonization on Technical Requirements for Registration of Pharmaceuticals for Human Use. Harmonized Tripartite Guideline; 1994.
13. ICH Topic S5A: CPMP/ICH/386/95, Operational March 1994. *Reproductive Toxicology: Detection of Toxicity to Reproduction for Medicinal Products*. (See http://www.emea.eu.int.)

14. Blaschke TF, Nies AS, Mamelok RD. Principles of therapeutics. In Gilman AF, Goodman LS, Rall TW, Murad F, Eds. *The Pharmacological Basis of Therapeutics*, 7th ed. New York: Macmillan Publishing; 1985, pp 49–65.
15. ICH Topic S6: CPMP/ICH/302/95, Operational March 1998. *Preclinical Safety Evaluation of Biotechnology Products Derived Pharmaceuticals*. (See http://www.emea.eu.int.)
16. van de Waterbeemd H. *Curr Opin Drug Discov Dev* 2002;5:33.
17. Viswanathan VN, Balan C, Hulme C, Cheetham JC, Yaxiong Sun Y. *Curr Opin Drug Discov Dev* 2002;5:400.
18. Chaturvedi PR, Decker CJ, Odinecs A. *Curr Opin Chem Biol* 2001;5:452.
19. Banik GM. *Curr Drug Discov* 2004;31.
20. Segall MD. *Future Drug Discov* 2004;81.
21. van de Waterbeemd H, Gifford E. *Nat Rev in Drug Discov* 2003;2:192.
22. Clark DE, Grootenhuis PD. *Curr Opin Drug Discov Dev* 2002;5:382.
23. Fichtl B, Nieciecki A, Walter K. *Adv Drug Res* 1991;20:117.
24. Olson RE, Christ DD. *Annu Rep Med Chem* 1996;31:327.
25. Naranjo CA, Sellers EM. *Drug–Protein Binding*. New York: Praeger; 1986, pp 233–251.
26. Carter DC, He XM, Munson SH, Twigg PD, Gernert KM, Broom MB, Miller TY. *Science* 1994;244:1195.
27. Curry S, Mandelkow H, Brick P, Franks N. *Nat Struct Biol* 1998;5:827.
28. Diaz N, Suarez D, Sordo TL Jr, Merz KM. *J Med Chem* 2001;44:250.
29. Herve F, Urien S, Albengres E, Duche JC, Tillement JP. *J Clin Pharmacokinet* 1994;26:44–58.
30. Kremer JM, Wilting J, Janssen LH. *Pharmacol Rev* 1988;40:1.
31. Zlotos G, Bucker A, Kinzig-Schippers M, Sorgel F, Holzgrabe U. *J Pharm Sci* 1998;87:215.
32. Borga O, Borga B. *J Pharmacokinet Biopharm* 1997;25:63.
33. Hoggard PG, Owen A. *J Antimicrob Chemother* 2003;51:493.
34. Goodman A, Gilman AG. *The Pharmacological Basis of Therapeutics*, 9th ed. New York: McGraw-Hill; 1996, pp 1712–1792.
35. Kratochwil NA, Huber W, Muller F, Kansy M, Gerber PR. *Biochem Pharmacol* 2002;64(9):1355.
36. Ebert SC. *Pharm Ther* 2004;29(4):244.
37. Cheng Y, Ho E, Subramanyam B, Tseng JL. *J Chromatogr B* 2004;809:67.
38. Kwong TC. *Clin Chim Acta* 1985;151:17.
39. Melten JW, Witterbrood AJ, Willems HJJ, Faber GH, Wemer J, Faber DB. *J Pharm Sci* 1985;74:692.
40. Machinist JM, Kukulka MJ, Bopp BA. *Clin Pharmacokinet* 1995;29:34.
41. Fung EN, Chen Y-H, Lau YY. *J Chromatogr B* 2003;795:187.
42. Beaudry F, Coutu M, Brown NK. *Biomed Chromatogr* 1999;13:401.
43. Bertucci C, Bartolini M, Gotti R, Andrisano V. *J Chromatogr B* 2003;797:111.
44. Kim HS, Hage DS. *J Chromatogr B* 2005;816:57.
45. Chen J, Ohnmacht C, Hage DS. *J Chromatogr B* 2004;809:137.

46. Bertucci C, Andrisano V, Gotti R, Cavrini V. *J Chromatogr B* 2002;768:147.
47. Lee EJD, Williams KM. Chirality: clinical pharmacokinetic and pharmacodynamic considerations. *Clin Pharmacokinet* 1990;18:339–345.
48. Jamali F, Mehvar R, Pasutto FM. Enantioselective aspects of drug action and disposition: therapeutic pitfalls. *J Pharm Sci* 1989;78:695–715.
49. Lin JH, deLuna FA, Ulm EH, Tocco DJ. Species-dependent enantioselective plasma protein binding of MK-571, a potent leukotriene D4 antagonist. *Drug Metab Dispos* 1990;18:484–487.
50. Tocco DJ, deLuna FA, Duncan AEW, Hsieh JH, Lin JH. Interspecies differences in stereoselective protein binding and clearance of MK-571. *Drug Metab Dispos* 1990;18:388–392.
51. Yasumori T, Chen L, Nagata K, Yamazoe Y, Kato R. Species differences in stereoselective metabolism of mephenytoin by cytochrome P-450 (CYP 2C and CYP 3A). *J Pharmacol Exp Ther* 1993;264:89–94.
52. Silber B, Holford NHG, Riegelman S. Stereoselective disposition and glucuronidation of propranolol in humans. *J Pharm Sci* 1982;71:699–704.
53. Wechter WJ, Loughead DG, Reischer RJ, Van Giessen GJ, Kaiser DG. Enzymatic inversion at saturated carbon: nature and mechanism of the inversion of R(−)-p-isobutylhydratropic acid. *Biochem Biophys Res Commun* 1974;61:833–837.
54. Hutt AJ, Caldwell J. The metabolic chiral inversion of 2-arylpropionic acids: a novel route with pharmacological consequences. *J Pharm Pharmacol* 1983;35:693–704.
55. Manzel-Soglowek S, Geisslinger G, Beck WS, Brune K. Variability of inversion of (*R*)-flurbiprofen in different species. *J Pharm Sci* 1992;81:888–891.
56. Jamali F, Berry BW, Tehrani MR, Russell SA. Stereoselective pharmacokinetics of flurbiprofen in humans and rats. *J Pharm Sci* 1988;77:666–669.
57. Foster RT, Jamali F. Stereoselective pharmacokinetics of ketoprofen in the rat: influence of route of administration. *Drug Metab Dispos* 1988;16:623–626.
58. Abas A, Meffin PJ. Enantioselective disposition of 2-arylpropionic acid nonsteroidal anti-inflammatory drugs: IV—ketoprofen disposition. *J Pharmacol Exp Ther* 1987;240:637–641.
59. Blaschke TF. Chromatographic resolution of racemates. *Angew Chem Int Ed Engl* 1980;19:13–24.
60. Blaschke TF, Kraft HP, Fickentscher K, Köhler F. Chromatographische racemattrennung von thalidomid und teratogene wirkung der enantiomere. *Drug Res* 1979;29: 1640–1642.
61. Fabro S, Smith RL, Williams RT. Toxicity and teratogenicity of optical isomers of thalidomide (Comment). *Nature (London)* 1967;215:296.
62. Simonyi M. On chiral drug action. *Med Res Rev* 1984;4:359–413.
63. Albert A. Chemical aspects of selective toxicity. *Nature (London)* 1958;182:421–426.
64. Loo JCK, Foltz EL, Wallick H, Kwan KC. Pharmacokinetics of pivampicillin and ampicillin in man. *Clin Pharmacol Ther* 1974;16:35–43.
65. Clayton JP, Cole M, Elson SW, Ferres H. BRL 8988 (talampicillin): a well absorbed oral form of apicillin. *Antimicrob Agents Chemother* 1975;5:670–671. Bodin NO, Ekström B, Forsgren U, Jular LP, Magni L, Ramsey CH, Sjöberg B. An orally well-absorbed derivative of ampicillin. *Antimicrob Agents Chemother* 1978;8:518–525.
66. Ulm EH, Hichens M, Gomez HJ, Till AE, Hand E, Vassil TC, Biollaz J, Brunner HR, Schelling JL. Enalapril maleate and a lysine analogue (MK-521) disposition in man. *Br J Clin Pharmacol* 1982;14:357–362.

67. Tocco DJ, deLuna FA, Duncan AEW, Vassil TC, Ulm EH. The physiological disposition and metabolism of enalapril maleate in laboratory animals. *Drug Metab Dispos* 1982;10:15–19.
68. Rees PJ, Selby P, Prentice HG, Whiteman PD, Grant DM. A prodrug of acyclovir with increased bioavailability. *J Antimicrob Chemother* 1986;18(Suppl B):215–222.
69. Krasny HC, Petty BG. Metabolism of desciclovir, a prodrug of acyclovir, in humans after multiple oral dosing. *J Clin Pharmacol* 1987;27:74–77.
70. Krenitsky TA, Hall WW, de Miranda P, Beauchamp LM, Schaeffer MJ, Whiteman PD. 6-Deoxyacyclovir: a xanthine oxidase-activated prodrug of acyclovir. *Proc Natl Acad Sci USA* 1984;81:3209–3213.
71. Wilk S, Mizoguchi H, Orlowski M. γ-Glutamyl dopa: a kidney-specific dopamine precursor. *J Pharmacol Exp Ther* 1978;206:227–232.
72. Clarke TA, Waskell LA. *Drug Metab Dispos* 2003;31(1):53.
73. Pereillo JM, Maftouh M, Andrieu A, Uzabiaga MF, Fedeli O, Savi P, Pascal M, Herbert JM, Mafford JP, Picard C. *Drug Metab Dispos* 2002;30:1288–1295.
74. Gillette JR. Problems in extrapolations from animals to man. In Kato R, Estabrook RW, Cayen MN, Eds. *Xenobiotic Metabolism and Disposition*. London: Taylor & Francis; 1989, pp 209–216.
75. *Guidance for Industry: Drug Metabolism/Drug Interaction Studies in the Drug Development Process: Studies In Vitro*. Department of Health and Human Services, US Food and Drug Administraion, Center for Drug Evaluation and Research, Center For Biologics Evaluation and Research; April 1997. http://www.fda.gov/cder/guidance.htm.
76. Burek DJ, Nitschke KD, Bell TJ, Wackerle DL, Childs RC, Beyer JE, Dittenber DA, Rampy LW, McKenna MJ. Methylene chloride: a two year inhalation toxicity and oncogenicity study in rats and hamsters. *Fundam Appl Toxicol* 1984;4:30–47.
77. Chinje E, Kentish P, Jarnot B, George M, Gibson G. Induction of the CYP 4A subfamily by perfluorodecanoic acid: the rat and the guinea pig as susceptible and non-susceptible species. *Toxicol Lett* 1994;71:69–75.
78. Swarm RL, Roberts GKS, Levy AC, Hines LR. Observations on the thyroid gland in rats following the administration of sulfamethoxazole and trimethoprim. *Toxicol Appl Pharmacol* 1973;24:351–363.
79. Duggan DE, Hooke KF, Noll RM, Kwan KC. Enterohepatic circulation of indomethacin and its role in intestinal irritation. *Biochem Pharmacol* 1975;24:1749–1754.
80. Lindberg LP, Negishi M. Alternation of mouse cytochrome P450 substrate specificity by mutation of a single amino-acid residue. *Nature (London)* 1989;339:632–634.
81. Clarke DJ, Burchell B. The uridine diphosphate glucuronosyltransferase multigene family: function and regulation. In Kauffman FC, Ed. *Conjugation–Deconjugation Reactions in Drug Metabolism and Toxicity*. Berlin: Springer-Verlag; 1994, pp 3–43.
82. Satoh T. Role of carboxylesterases in xenobiotic metabolism. In Hodgson E, Bend JR, Philot RM, Eds. *Review in Biochemical Toxicology*, Vol 8. Amsterdam: Elsevier; 1987, pp 155–181.
83. Hosokawa M, Maki T, Satoh T. Characterization of molecular species of liver microsomal carboxylesterases of several animal species and human. *Arch Biochem Biophys* 1990;277:219–227.
84. Stearns RA, Miller RR, Doss GA, Chakravarty PK, Rosegay A, Gatto GJ, Chiu S-HL. The metabolism of Dup 753, a nonpeptide angiotensin II receptor antagonist by rat, monkey and human liver slices. *Drug Metab Dispos* 1992;20:281–287.
85. Smith DA. Species differences in metabolism and pharmacokinetics: are we close to an understanding? *Drug Metab Rev* 1993;23:355–373.

86. Blum MR, Liao SHT, Good SS, deMiranda P. Pharmacokinetics and bioavailability of zidovudine in man. *Am J Med* 1988;85:189–196.
87. Resetar A, Spector T. Glucuronidation of 3′-azido-3′-deoxythymidine: human and rat enzyme specificity. *Biochem Pharmacol* 1989;38:1389–1393.
88. Lin JH, Levy G. Effect of prevention of inorganic sulfate depletion on the pharmacokinetics of acetaminophen in rats. *J Pharmacol Exp Ther* 1986;239:94–98.
89. Slattery JT, Levy G. Acetaminophen kinetics in acutely poisoned patients. *Clin Pharmacol Ther* 1979;25:185–195.
90. McDonnell WM, Scheiman JM, Traber PG. Induction of cytochrome P450 IA genes (CYP 1A) by omeprazole in the human alimentary tract. *Gastroenterology* 1992;103:1509–1516.
91. Diaz D, Fabre I, Daujat M, Saint Aubert B, Bories P, Michel H, Maurel P. Omeprazole is an aryl hydrocarbon-like inducer of human hepatic cytochrome P-450. *Gastroenterology* 1990;99:737–747.
92. Strolin Benedetti M, Dostert P. Induction and autoinduction properties of rifamycin derivatives: a review of animal and human studies. *Environ Health Perspect* 1994;102(Suppl 9):101–105.
93. Nebert DW, Gonzalez FJ. The P450 gene superfamily. In Ruckpaul K, Rein H, Eds. *Principles, Mechanisms and Biological Consequences of Induction*. London: Taylor & Francis; 1990, pp 35–61.
94. Holk HGO, Munir AK, Mills LM, Smith EL. Studies upon the sex differences in rats in tolerance to certain barbiturates and to nicotine. *J Pharmacol Exp Ther* 1973;60:323–346.
95. Gregory L, Kedderis GL, Mugford CA. Sex dependent metabolism of xenobiotics. *Drug Metab Rev* 1998;30:441–498.
96. Kobliakov V, Popova N, Rossi L. Regulation of the expression of the sex-specific isoforms of cytochrome P-450 in rat liver. *Eur J Biochem* 1991;195:585–591.
97. Bandiera S. Expression and catalysis of sex-specific cytochrome P-450 isozymes in rat liver. *Can J Physiol Pharmacol* 1990;68:762–768.
98. Legraverend C, Mode A, Wells T, Robinson I, Gustafsson J-A. Hepatic steroid hydroxylating enzymes are controlled by the sexually dimorphic pattern of growth hormone secretion in normal and dwarf rats. *FASEB J* 1992;6:711–718.
99. Waxman DJ, Dannan GA, Guengerich FP. Regulation of rat hepatic cytochrome P450: age-dependent expression, hormonal imprinting and xenobiotic inducibility of sex-specific isoenzymes. *Biochemistry* 1985;24:4409–4417.
100. Shapiro BH. Sexually dimorphic response of rat hepatic monooxygenases to low dose phenobarbital. *Biochem Pharmacol* 1986;35:1766–1768.
101. Macleod JN, Sorensen MP, Shapiro BH. Strain independent elevation of hepatic monooxygenase enzyme in female mice. *Xenobiotics* 1987;17:1095–1102.
102. Ioannides C, Sweatman B, Richards R, Parke DV. Drug metabolism in the ferret: effects of age, sex and strain. *Gen Pharmacol* 1977;8:243–249.
103. Dogterom P, Rothuizen J. A species comparison of tolbutamide metabolism in precision-cut liver slices from rats and dogs: qualitative and quantitative sex differences. *Drug Metab Dispos* 1993;21:705–709.
104. Hunt CM, Westerkam WR, Stave GM. Effect of age and gender on the activity of human hepatic CYP3A. *Biochem Pharmacol* 1992;44:275–283.
105. Tynes RE, Philpot RM. Tissue and species dependent expression of multiple forms of flavine monooxygenases. *Mol Pharmacol* 1987;31:569–574.

106. Lin JH, Chiba M, Balani SK, Chen I-W, Kwei GY-S, Vastag KJ, Nishime JA. Species differences in the pharmacokinetics and metabolism of indinavir, a potent HIV protease inhibitor. *Drug Metab Dispos* 1996;24:1111–1120.
107. Gludicelli JF, Tillement JP. Influence of sex on drug kinetics in man. *Clin Pharmacokinet* 1977;2:157–166.
108. Macleod SM, Giles HG, Bengeret B, Lui FF, Sellers EM. Age and gender related differences in diazepam pharmacokinetics. *J Clin Pharmacol* 1979;19:15–19.
109. Sesardic D, Boobis AR, Murray P, Murray S, Segura J, de la Torre R, Davies DS. Furafylline is a potent and selective inhibitor of cytochrome P450 IA2 in man. *Br J Clin Pharmacol* 1990;29:651–663.
110. Kobayashi S, Murray S, Watson D, Sesardic D, Davies DS, Boobis AR. The specificity of inhibition of debrisoquine 4-hydroxylase activity in quinidine and quinine in the rat is the reverse of that in man. *Biochem Pharmacol* 1989;38:2795–2799.
111. *Guidance for Industry and Reviewers: Estimating the Safe Starting Dose in Clinical Trials for Therapeutics in Adult Healthy Volunteers*. US Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research (CDER), Center for Biologics Evaluation and Research (CBER); December 2002. Pharmacology and Toxicology, http://www.fda.gov/cber/guidelines.htm.
112. Freireich EJ, Gehan EA, Rall DP, Schmidt LH, Skipper HE. Quantitative comparison of toxicity of anticancer agents in mouse, rat, hamster, dog, monkey, and man. *Cancer Chemother Rep* 1966;50:219–244.
113. Schein PS, Davis RD, Carter S, Newman J, Schein DR, Rall DP. The evaluation of anticancer drugs in dogs and monkeys for the prediction of qualitative toxicities in man. *Clin Pharmacol Ther* 1970;11:3–40.
114. Cayen MN. Consideration in the design of toxicokinetic studies. *Toxicol Pathol* 1995;23:148–157.
115. Gibaldi M, Perrier D, Eds. *Pharmacokinetics*, 2nd ed. New York: Marcel Dekker; 1982.
116. Dahlem AM, Allerheiligen SR, Vodicnik MJ. Concomitant toxicokinetics techniques for interpretation of exposure data obtained during the conduct of toxicology studies. *Toxicol Pathol* 1995;2:170–178.
117. Bree JV, Nedelman J, Steimer JL, Tse F, Robinson W, Niederberger W. Application of sparse sampling approaches in rodent toxicokinetics: a prospective view. *Drug Inf J* 1994;28:263–279.
118. Campbell DB, Jochemsen R. Nonclinical pharmacokinetics and toxicokinetics. In Cartwright AC, Mathews BR, Eds. *International Pharmaceutical Product Registration, Aspects of Quality, Safety, and Efficacy—Chemistry, Pharmacy and Manufacturing*. Chichester: Ellis Horwood; 1994, pp 560–627.
119. Tse F, Nedelman J. Serial versus sparse sampling in toxicokinetic studies. *Pharm Res* 1996;13:1105–1108.
120. *Reviewer Guidance: Validation of Chromatographic Methods*. Center for Drug Evaluation and Research (CDER), November 1994.
121. *United States Pharmacopoeia* 2000;24:⟨1225⟩.

# *IN VITRO* TOXICOKINETICS AND DYNAMICS: MODELING AND INTERPRETATION OF TOXICITY DATA

ARIE BRUININK

*Materials–Biology Interactions, Materials Science & Technology (EMPA), St. Gallen, Switzerland*

**Contents**

1 Introduction
2 Benchmark Dose
3 Toxicokinetics *In Vivo*
4 Basics of *In Vitro* Toxicodynamics
  4.1 Test Substance Binds Predominatly to One Important Effectual Cell Protein
  4.2 Test Substance Binds to Two Important Effectual Proteins with Different Affinities
  4.3 Test Substance Binds to Many Important Effectual Proteins with Approximately Equal Affinity
5 Rationale Describing *In Vitro* Toxicokinetics and Dynamics and Aim of the Present Study
6 *In Vitro* Toxicokinetics and Dynamics
  6.1 Effects Based on Peak Concentrations (One-Compartment Models)
  6.2 Effects Based on Accumulation (Two-Compartment Models)
7 Concluding Remarks Regarding the Eight Models of Toxicity and the Discrimination Between Them
Acknowledgment
References

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

## 1 INTRODUCTION

If the dose of a bioactive test compound (drug, chemical, protein, etc.) exceeds a certain concentration and/or if the exposure duration at subtoxic doses is lengthened, adverse effects may develop in organisms. Knowledge about toxicokinetics and toxicodynamics is of crucial importance for various areas of research: this ranges from risk assessment of compounds, to the development of new drugs and new biomedical materials. Toxicokinetics is generally regarded as an area of science dealing with kinetics of exposure, absorption, distribution, metabolism, and excretion of test compounds and metabolites. Toxicodynamics, however, focuses on studying dose–response relationships. Depending on the application of test compound, the *no adverse effect level* (NOAEL)—the dose or dose/concentration at which the parameter of interest is reduced by 50% (e.g., *lethal dose* 50 ($LD_{50}$))—is of importance. Whereas numerous publications have described approximations of the fate of test compounds in the intact body and dose–effect relationships, no such information is currently available for *in vitro* systems.

This chapter describes *in vitro* concentration–effect relationships, taking variations in toxic compound concentration into account. For this, a set of equations are presented that are based on known first-order equations describing uptake and elimination-based fluctuations in compound blood levels [1] and equations describing receptor–ligand interaction relationships [2]. Furthermore, special attention is paid to determine inhibitory concentrations reducing the parameter of interest by 50% ($IC_{50}$) or 5% ($IC_5$; benchmark dose 5%).

## 2 BENCHMARK DOSE

All compounds are toxic under certain exposure conditions. The change of the occurrence of these adverse effects can be minimized by limiting the dose or exposure period. A basis for hazard assessment of chemicals is the no observed adverse effect level (NOAEL). For assessment of the NOAEL, only animal toxicity data are used. It represents the highest dose at and below which no significant adverse effects are seen. In reality, it is defined by the *lowest adverse effect level* (LOAEL)—the dose at which a significant effect is seen [3]. Instead of NOAEL, recent alternative approach is to use the *benchmark dose* [4]. In contrast to the NOAEL, in this approach all data obtained are taken into account. A line is fitted through the data using a mathematical model for the dose–response relationship. With the help of this fitted relationship, we can calculate the dose at which the incidence or frequency of a toxic adverse effect is increased (in the case of the benchmark dose by 5%). In addition, the statistical confidence limits of this dose are estimated. Both values are dependent on the mathematical model for the dose–response relationship used. Still, no general model for the dose–response relationship exists against which the line should be fitted. In most cases, a log-probit extrapolation of concentration–response data to the 95% lower confidence limit on the toxic concentration is employed. The practical use of the benchmark dose has been evaluated by applying dose–response models to an extensive historical database [5]. It was found that the lower confidence limit on the 5% benchmark dose is comparable to the NOAEL for most datasets. In this chapter a mathematical base is given to calculate the effec-

tive concentration of a test compound that reduces the parameter of interest by 5% and 50%.

## 3 TOXICOKINETICS *IN VIVO*

One simplified model that has gained wide acceptance in toxicokinetics is that of describing the system in terms of a single compartment in which the test compound is homogeneously distributed after administration. This is the case if (1) the distribution equilibrium between test compound in blood and tissues is reached before a significant amount of test compound has been eliminated, and (2) the equilibrium constant is such that the concentration ratio between blood and tissues remains stable during the period of elimination. The one-compartment model is a strong simplification of the *in vivo* situation but, to a large extent, is applicable for several types of test compound.

After intravenous (IV) application, the serum level exponentially decreases due to elimination processes. In the intact animal, the administered compounds are eliminated from the circulation by various routes. By assuming one homogeneous compartment, the serum level of the compound can be described mathematically [1] by

$$y_i = y_0 e^{-k_{el} t_i} \tag{1}$$

where $y_i$ is the blood concentration of the compound at time $t_i$ after addition at $t_0$, $y_0$ is the blood concentration of the compound just after administration at $t_0$, and $k_{el}$ is the elimination constant.

The elimination velocity is characterized by the elimination half-life ($t_{1/2}$). The elimination half-life is defined as the period that is needed to reduce the serum level by a factor of 2 ($y_i = 0.5 y_0 = y_0 e^{-k_{el} t_{1/2}}$). Thus,

$$t_{1/2} = \frac{\ln 2}{k_{el}} \tag{2}$$

After an extravascular administration, the compound first has to enter the circulation before it is eliminated from this compartment. The delay and magnitude of the peak plasma concentration following extravascular administration are a function of the kinetics of absorption as well as, in the case of inhalation and oral administration, the fraction ($f$) of the absorbed unchanged compound (Fig. 1). The value of $f$ normally equals 1 for an extravascular injection of the compound (IM, SC, IP) but is lower than 1 after an oral intake or after inhalation. The volume of distribution $V$ is strongly compound dependent and may generally vary between 0.1 and 10 L/kg body weight.

The equation for the concentration of the test compound $y_i$ in the blood serum assuming one compartment corresponds to the Bateman equation [1]:

$$y_i = \left(\frac{Df}{V}\right) k_{abs} \left(\frac{e^{-k_{el} t_i} - e^{-k_{abs} t_i}}{k_{abs} - k_{el}}\right) \tag{3}$$

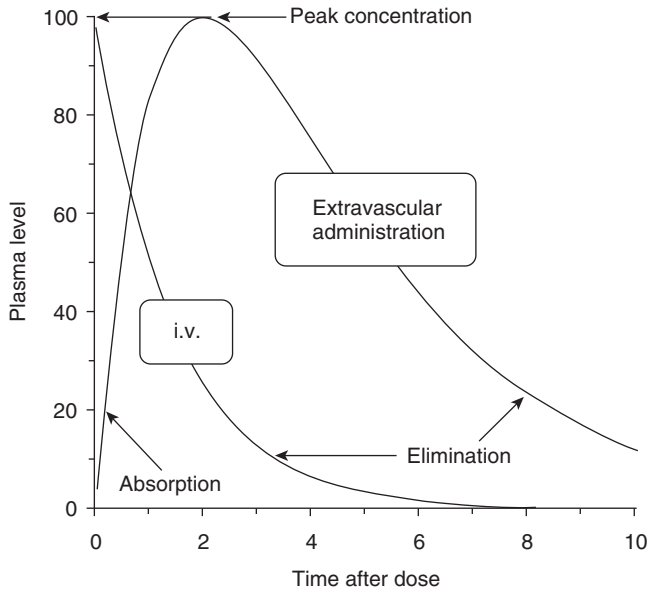


**FIGURE 1** Simulated plasma concentration–time curve assuming the one-compartment model. These relationships are characterized by an elimination and, in the case of extravascular administration, an absorption phase. The AUC is the area below each curve.

in which $k_{abs}$ is the absorption constant, $D$ is the dose, $f$ is the fraction of the test compound dose taken up (extent of absorption), and $V$ is the distribution volume.

Independent of the route of administration, a similar plasma concentration–time curve is obtained for the metabolite(s) of the test compound.

One assumption in toxicology is that the cumulative dose at the sensitive site determines the reaction. A simplification can be made: the extent of adverse effects is correlated to the *area under the curve* (AUC) of the concentration–time relationship of the active form of the compound at the site of action (Fig. 1). Based on the latter yet unproven assumption, other theories (e.g., linear systems theory) have been hypothesized to characterize toxicokinetics and toxicodynamics [6]. Later on, it will be shown that, based on the premises made in this chapter, this assumption is only correct in some selected cases.

## 4 BASICS OF *IN VITRO* TOXICODYNAMICS

In the *in vivo* situation, multiple relatively large compartments are present in which the administered compounds can be distributed, metabolized, and redistributed. Various physiologically based toxicokinetic models have been developed as tools to describe blood levels after exposure [7–9]. Normally, the various compartments of the *in vitro* systems are composed of large (the medium part of the culture) and negligibly small (the cells) parts. Models to be developed can assume one compartment, that is, the sensitive target cell component is directly accessible, or two compartments, that is, the culture medium and the compartment within the cell that the

toxic compound first has to reach and develop a certain concentration before adverse effects are seen.

In *in vivo* studies, benchmark dose, NOAEL, LOAEL, as well as cumulative dose are generally used to describe the concentration–effect relationships of a test compound; equivalent expressions for describing *in vitro* effects do not exist in the case of the benchmark dose or are only rarely used. *In vitro*, the potency of a test compound in affecting cell functions is defined by the *effective culture medium concentration* (EC) or *inhibiting culture medium concentration* (IC). In general, the expressions $EC_{50}$ and $IC_{50}$ are used. They represent the concentrations of a test compound that modify or diminish, respectively, the value of a given measured parameter (cell function, concentration of certain proteins, etc.) by 50% relative to control. In analogy to the NOAEL, its *in vitro* equivalent, the *no observed effect medium concentration* (NOEC), is strongly dependent on the choice of the test compound concentration and reproducibility of the measurements of the given parameter. In contrast, like the *in vivo* benchmark dose approach, all results are taken into account for the estimation of $EC_{50}$ and $IC_{50}$ values. Thus, taking a certain additional safety factor into account, the use of $EC_{50}$ and $IC_{50}$ values for risk analysis seem to be justified. However, the appearance of biphasic effects makes it very difficult to assess these $EC_{50}$ and $IC_{50}$ values correctly. Furthermore, the differences between medium concentrations of the test compound, which only negligibly affect the parameter of interest, and $EC_{50}$ or $IC_{50}$ levels are determined by the shape of the medium concentration–effect relationships and are cell culture, compound, and parameter dependent. This means that an approach equivalent to *in vivo* benchmark dose 5% would be more adequate for defining *in vitro* threshold medium concentrations. For the estimation of both the benchmark dose 5% in *in vivo* studies and the *benchmark culture medium concentration 5%* ($BMC_5$) in *in vitro* studies, a mathematical model for the dose–response relationship is needed, according to which the lines are fitted.

The description of the dose–response curves obtained from *in vivo* and *in vitro* experiments is generally based on a linear regression analysis of a near linear portion of the concentration–response curve, plotting the logarithm of the dose on the horizontal axis and the response on the vertical axis. This method permits the calculation of an effective dose. However, the latter value can be manipulated by the choice of data points. Complex dose–response curves, such as those showing stimulation at low concentrations but inhibition at high concentrations, cannot be described. Moreover, the slope of the curve, another important index characterizing the effects of a compound, also depends on the data points chosen and is nearly always neglected.

In 1992, I proposed a new mathematical approach to analyze *in vitro* toxicity data [3]. Since the book in which this approach is described is no longer available and because the present paper is built on this, the approach is briefly described here. The statistical method is based on models of enzyme kinetics and the law of mass action. This method allows both the effective concentration and the slope of the dose–response curve to be reliably estimated, as well as the $IC_5$ and the benchmark dose 5% [10–13]. The values of the various parameters in the following formulas can be calculated using available spreadsheet programs with macro possibilities; the contribution from each data point is weighted according to the standard error of the mean. Based on this model, the toxic effects of several neurotoxic compounds

have been evaluated [14–18]. Recently, other very similar approaches have been suggested [19].

The description of the test compound concentration/dose–effect relationship is based on the primary toxic event. The exposure of cells to a test compound or its toxic metabolite generally first results in the test compound binding to one or more of the proteins in the outer membrane of the cells, where it may intercalate into the membrane or diffuse (transported) into the cell and bind to one or more intracellular proteins (Fig. 2). Binding of the test compound induces a range of effects, such as stimulation or inhibition of synthesis or degradation of various cellular molecules directly or mediated indirectly by secondary processes. It is reasonable to assume that the effect on the various cell functions will be directly related to the first event, by the fraction of target protein sites occupied by the test substance. The measurement of binding is amenable to mathematical analysis and therefore also its evoked effects. The effects on the measured cell function(s) can be positive or negative depending on the compound and the compound concentration.

### 4.1 Test Substance Binds Predominantly to One Important Effectual Cell Protein

Generally, duration of exposure to the test compound is sufficient to ensure that, within the most sensitive cell compartment, binding to cell effectual protein involved in cell function reaches equilibrium. If equilibrium is established and degradation of the test compound is negligible, the law of mass action is valid as described by

$$\frac{[\mathrm{SB}]}{[\mathrm{B}_{\max}]} = \frac{[\mathrm{S}]}{K_{\mathrm{d}} + [\mathrm{S}]} \tag{4}$$

where S is the test compound, B is the unbound effectual protein, and SB is the test compound–effectual protein complex. In this equilibrium state, only the complex SB induces, directly or indirectly, the changes in cell function. $K_{\mathrm{d}}$ represents the

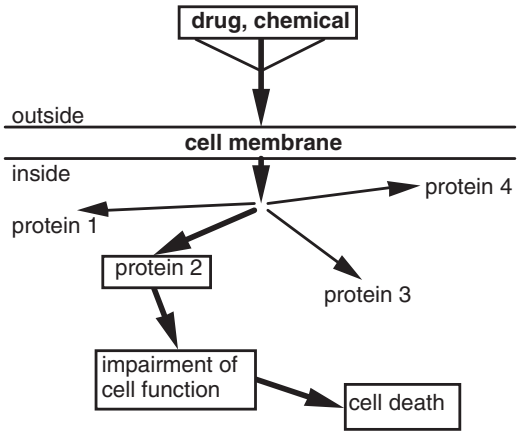


**FIGURE 2** Exposure of a cell with a test compound or chemical: schematic representation.

equilibrium dissociation constant. The total amount of protein $[B_{max}]$ is the sum of [B] and [SB]. $[SB]/[B_{max}]$ is the fraction of effectual protein sites to which the test compound is bound. If the relationship between the fraction of bound sites and the changes in the cell function(s) being measured is linear, then Eq. 4 can be rewritten to incorporate the cell function:

$$\frac{E_i}{E_u} = \frac{[S]}{K_r + [S]} \tag{5}$$

where $E_u$ is the experimental value of the cell function in untreated cells, $E_i$ is the reduction in $E_u$ caused by the test substance, and $K_r$ is the response constant of the cell for the test substance.

The value of cell function of interest in the presence of the test substance as a fraction of the value of the cell function in untreated cells ($E_t/E_u$) is, therefore

$$\frac{E_t}{E_u} = \frac{K_r}{K_r + [S]} \tag{6}$$

where $E_t$ is the measured value of the cell function in treated cells and $E_u = E_t + E_i$.

Generally, the toxic effects are expressed by $IC_{50}$ or $BMC_5$. With this equation, $IC_{50}$ and $BMC_5$ can be calculated. As defined previously, $IC_{50}$ is the concentration of the test substance when the measured value of cell function is reduced to half its maximum value: $E_t/E_u = 0.5$ (see Fig. 3: $N_c = 1$). Thus,

$$0.5 = \frac{K_r}{K_r + IC_{50}} \quad \text{or} \quad IC_{50} = K_r \tag{7}$$

The $BMC_5$ is defined as the concentrations where $E_t/E_u = 0.95$. Thus,

$$BMC_5 = \frac{K_r}{19} \tag{8}$$

If in Eq. 8 $K_r$ is replaced by $IC_{50}$, then according to Eq. 7 $BMC_5$ can be related to $IC_{50}$ as follows:

$$BMC_5 = \frac{IC_{50}}{19} \tag{9}$$

In most cases, however, the effect on the cell function(s) of the binding of the test substance to the relevant effectual protein is not linearly related to the fraction of bound effectual protein sites ($[SB]/[B_{max}]$ ). Generally, as the number of occupied sites increases, the effect on the measured cell function will be enhanced with every additional occupied site. Thus, as [S] increases, $K_r$ decreases. In other words, a positive cooperativity analogous to the binding of oxygen to hemoglobin as described by Hill [2] is observed. Therefore,

$$\frac{E_t}{E_u} = \frac{1}{i}\left(\frac{K_{r1}}{K_{r1}+[S]} + \frac{K_{r2}}{K_{r2}+[S]} + \cdots + \frac{K_{ri}}{K_{ri}+[S]}\right)$$

in which $i$ is the number of effectual protein molecules or, more generally,

$$\frac{E_t}{E_u} = \frac{K_c}{K_c + [S]^{N_c}} \tag{10}$$

where $K_c$ is the response constant under cooperativity conditions, and $N_c$ is the cooperativity constant. Examples of concentration–effect relationships with varying $N_c$ values are shown in Fig. 3.

Equation 6 is a special case of Eq. 10, where $N_c = 1$, and because of that, $K_c = K_r$. The concentration of the substance where the measured value of the cell parameter is half the maximum value is

$$\begin{aligned} 0.5 &= \frac{K_c}{K_c + (IC_{50})^{N_c}} \Rightarrow \\ IC_{50} &= K_c^{1/N_c} \quad (\text{or } K_c = (IC_{50})^{N_c}) \end{aligned} \tag{11}$$

In analogy, the $BMC_5$ can be calculated:

$$\begin{aligned} 0.95 &= \frac{K_c}{K_c + (BMC_5)^{N_c}} \Rightarrow \\ BMC_5 &= \left(\frac{K_c}{19}\right)^{1/N_c} \end{aligned} \tag{12}$$

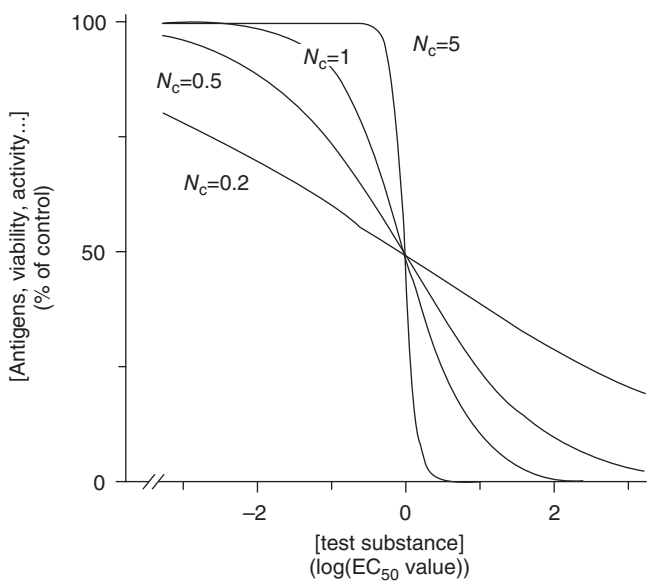


**FIGURE 3** Relationship between the cell parameter and the concentration of the test substance calculated from Eq. 10 for various values of the cooperativity constant $N_c$, assuming that $K_c = IC_{50}$ and $E_u = 100$ (arbitrary units).

Small variations of $N_c$ strongly affect the $BMC_5$ value. In analogy to Eq. 9, by combining Eqs. 11 and 12, the $BMC_5$ in relation to $N_c$ and $IC_{50}$ can be described by the following equation:

$$BMC_5 = \frac{IC_{50}}{19^{1/N_c}} \tag{13}$$

By reducing $N_c$ values, the ratio $IC_{50}/BMC_5$ increases exponentially. Thus, small variations of $N_c$ value affect the $BMC_5$ value maximally at very small $N_c$.

The distribution of the logarithm of the $IC_{50}$ and $BMC_5$ values found under experimental conditions is (as stated for the *in vivo* situation [20]) log-normal Gaussian around the mean of the logarithm of the $IC_{50}$ and $BMC_5$ values, respectively, for each experiment and not (as suggested by others [19]) normal at the nonlogarithmic linear scale. The same appears to be true for the $N_c$ values.

### 4.2 Test Substance Binds to Two Important Effectual Proteins with Different Affinities

After application of a compound under the present assumption, two different main effects on cells can be discriminated depending on the applied concentration. In this case, $E_t/E_u$ represents a combination of the effects induced by effectual protein 1 with parameters $E_{u1}$, $K_{c1}$, and $N_{c1}$ and by effectual protein 2 with parameters $E_{u2}$, $K_{c2}$, and $N_{c2}$. Thus,

$$\frac{E_t}{E_u} = \frac{\dfrac{E_{u1}K_{c1}}{K_{c1}+[S]^{N_{c1}}} + \dfrac{E_{u2}K_{c2}}{K_{c2}+[S]^{N_{c2}}}}{E_{u1}+E_{u2}} \tag{14}$$

where

$$E_u = E_{u1} + E_{u2} \tag{15}$$

For both phases of the medium concentration–effect relationship, $BMC_5$ values can be calculated. Binding of the test substance to the effectual proteins can stimulate or inhibit cell function. In case that binding to effectual protein 1 stimulates cell function and binding to effectual protein 2 inhibits cell function, the value of $E_{u2}$ will represent the sum of the control value and the increase induced by binding to effectual protein 1 (Fig. 4) where $E_{u1}$ has a negative value. Instead of $IC_{50}$, the more general term *effective concentration 50%* ($EC_{50}$) is used to describe the concentration resulting in a 50% stimulation or inhibition. Examples of both situations, namely, where both $E_u$ values are positive or one is positive and one is negative, are shown in Figs. 4 and 5 (line D in Fig. 5 is the sum of B and C, and line E is the sum of A and C). Equation 14 also applies if a test substance in the concentration range investigated only affects a fraction of the measured functional parameter (Fig. 4 in the concentration range 0–1000$EC_{50,1}$, and Fig. 5 lines A and B). In this case, the $IC_{50}$ value for binding to the second protein is outside the tested concentration range and the cooperativity constant $N_c$ must be set to $10^{-\infty}$.

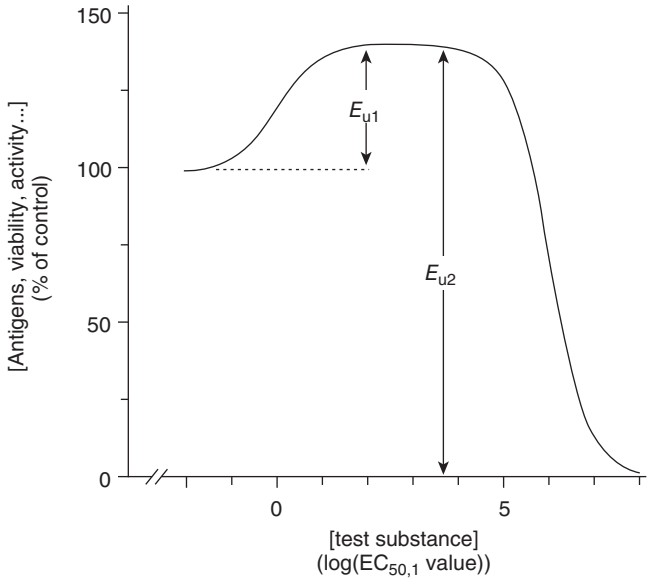


**FIGURE 4** Relationship between the cell parameter and the concentration of the test substance calculated from Eq. 14 where the test substance binds to two of the effectual proteins involved in the measured cell function ($E_{u1} = -40$, $K_{c1}$=$EC_{50,1}$, $N_{c1} = 1$; $E_{u2} = 140$, $K_{c2} = 100{,}000EC_{50,1}$, and $N_{c2} = 1$) (arbitrary units).

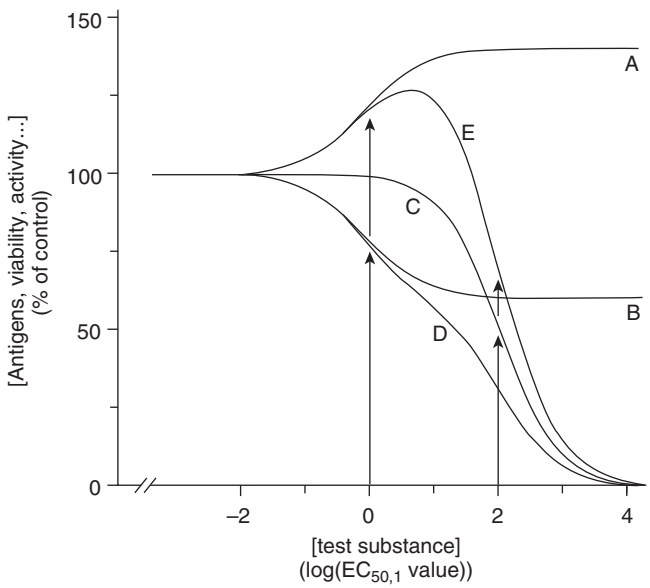


**FIGURE 5** Relationship between the cell parameter and the concentration of the test substance calculated from Eq. 10 (line C: $K_c = 100EC_{50,1}$, $E_u = 100$, $N_c = 1$) and Eq. 14 (line A: $K_{c1} = EC_{50,1}$, $E_{u1} = -40$, $N_{c1} = 1$, $K_{c2} = \infty$, $E_{u2} = 140$, $N_{c2} = 10^{-\infty}$; line B: $K_{c1} = EC_{50,1}$, $E_{u1} = 40$, $N_{c1} = 1$, $K_{c2} = \infty$, $E_{u2} = 60$, $N_c = 10^{-\infty}$; line D: $K_{c1} = EC_{50,1}$, $E_{u1} = 40$, $N_{c1} = 1$, $K_{c2} = 100EC_{50,1}$, $E_{u2} = 60$, $N_{c2} = 1$; line E: $K_{c1} = EC_{50,1}$, $E_{u1} = -40$, $N_{c1} = 1$, $K_{c2} = 100EC_{50,1}$, $E_{u2} = 140$, $N_{c2} = 1$) (arbitrary units).

For certain indices of *in vivo* response, animals react to test compound treatment mainly by stimulation/increase at low doses and inhibition at high doses (e.g., behavioral response to benzodiazepine receptor ligands [21]). Using Eq. 14, the $BMC_5$ and $EC_{50}$ for the low concentration effect and the $BMC_5$ and $IC_{50}$ of the inhibitory effect observed at the high compound concentration range may be determined correctly.

### 4.3 Test Substance Binds to Many Important Effectual Proteins with Approximately Equal Affinity

In this case, the effect on the cell function induced by binding of the test substance to various effectual proteins involved in cell function is cumulative. If the relationship between the change in the cell function and the fraction of bound sites is linear for each of the proteins (no cooperativity), the value of the cell function in the presence of the test substance as a fraction of the control value ($E_t/E_u$) is

$$\frac{E_t}{E_u} = \frac{\dfrac{E_{u1}K_{c1}}{K_{c1}+[S]^{N_{c1}}} + \dfrac{E_{u2}K_{c2}}{K_{c2}+[S]^{N_{c2}}} + \cdots + \dfrac{E_{ui}K_{ci}}{K_{ci}+[S]^{N_{ci}}}}{E_{u1}+E_{u2}+\cdots+E_{ui}} \tag{16}$$

As the number of different proteins ($i$) increases and assuming that the distribution of the logarithm of the affinities to the various proteins is Gaussian, the generalized form of Eq. 16 will approach the situation described by Eq. 10. If, as described for the one effector site model, there is a cooperative effect for each single effector protein, this equation will still hold, but the values of $N_c$ and $K_c$ will change. Thus, if there is a cooperative effect or the test substance binds to many proteins that are involved in the measured cell functions, the results are similar. Depending on the precise values of $EC_{50}$, $N_c$, and the standard deviations of the values of the measured cell function, the phenomenon can even take place for the two effector site model (compare Fig. 5 line D and Fig. 3).

## 5 RATIONALE DESCRIBING *IN VITRO* TOXICOKINETICS AND DYNAMICS AND AIM OF THE PRESENT STUDY

As in *in vivo*, the *in vitro* toxicity of compounds and test compounds is determined by toxicokinetic parameters in addition to the sensitivity of the cells. Although the concentrations of the test compounds are in most cases not or relatively slowly modified in *in vitro* systems, medium has to be partly replaced from time to time, especially in long-term cultures. The medium replacement may affect the medium test compound concentration. Furthermore, the compound may need an accumulation within the target compartment (e.g., cytoplasm, lysosomes) before toxic effects are induced. As a result, the toxicity for two different situations can be discriminated: toxicity of a compound and/or of its metabolite(s) is determined by peak concentrations or by the treatment period times concentration. It depends on the compound tested which of the two possibilities is crucial for the toxicity. So far, no mathematical approach exists for the data analysis of *in vitro* pharmaco/toxicoki-

netic data. In Sections 6.1 and 6.2 eight different hypothetical extreme *in vitro* scenarios are discussed. In Models 1–4, toxicity is solely determined by the $C_{max}$ (Section 6.1) and in Models 5–8 toxicity is determined by the treatment period (Section 6.2). Concerning the stability of the compound, in Models 1 and 5 the compound is stable; in Models 2 and 6 the toxic test compound is inactivated, similar to the *in vivo* situation after IV administration; in Models 3 and 7 the nontoxic test compound is modified into a toxic compound; and in Models 4 and 8 the toxic metabolite formed (Models 3 and 7) is thereafter inactivated (Fig. 6). This latter case is similar to an *in vivo* extravascular administration of a test compound. It must be noted that Models 4 and 8 are only mentioned for the sake of completeness and will only be discussed briefly. Furthermore, a scheme is presented that may help select the model according to how the given test compound behaves (Section 7). In this chapter, it is assumed that test compound concentration–effect relationship characteristics can be described using the equations in Section 4. Furthermore, in this chapter the equations are simplified: it is assumed that the effects are induced according to the single effector site model (Section 4.1). One must be aware that Models 1–8, as mathematical models for describing *in vivo* serum concentrations, simplify the real existing situation found in the culture dish and are based on the premises made. However, the discussed models may be of help to optimize experiments, to provide additional data from the obtained results, and to understand the collected results.

The outcome of experiments, in which concentration–effect relationships are investigated, is generally expressed in calculated compound concentrations, taking the known concentrations of the test compound in the stock solution just after dissolving in the culture medium as a starting point; however, concentration during the experiment may vary and the modified form of the test compound is responsible

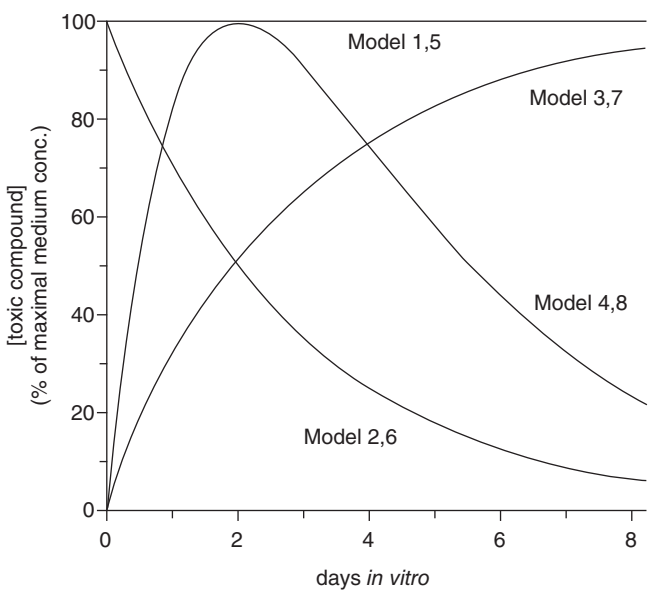


**FIGURE 6** Concentration of the toxic compound in the culture medium in the eight hypothetical models after various periods of time.

for the observed adverse effects. Because of these reasons, instead of concentration–effect relationship, the term *initial outer compartment (cell culture medium) test compound concentration–effect relationship* (OCER) is used. Furthermore, the calculated *initial (starting) concentration* of the compound causing certain effects is based on the compound concentration in the stock solution and is called the *effective initial concentration* (EIC). In $EIC_z$, $z$ stands for the observed effect relative to control. Here, $z$ is 50 if at the end of the treatment period, the value of the parameter of interest is exactly between the value of (untreated) controls and the maximal possible change. $EIC_{50}$ is thereby the starting (initial) concentration at which, at the end of the treatment period, this 50% reduction (or stimulation) occurs. In analogy, in $BMC_5$ $z$ is 5 and represents a 5% change relative to control values. One possibility to determine the velocity of activation or inactivation of a test compound in the cell cultures is to modify the treatment schedules in such a way that the $C_{max}$ (Models 1–4) and AUC (Models 5–8) of the toxic compound or of its toxic metabolite are changed. These modifications in the treatment schedule include **(a)** a partial replacement of the culture medium by fresh culture medium not containing any compound (*compound dilution*) during the treatment period; **(b)** a partial replacement of the culture medium during the treatment period by fresh culture medium *containing* the compound at the initial concentration (*compound replacement*); and **(c)** a preincubation of the compound of interest at nearly the final concentration before adding to the cell cultures (*compound preincubation*).

As mentioned above, schedules (a) and (b) normally occur if cultures are treated or kept for prolonged periods of time. In some labs, stock solutions of test compounds are made and kept until use. In that case, treatment schedule (c) takes place even if the person involved is unaware of it. This chapter takes these situations into account and even makes use of it. Models 1–4 are based on the assumption that the test compound does not penetrate into the cells (or, in the case where it is taken up, the test compound does not interact with intracellular components) and exerts its influence by interacting with components of the cell membrane: that is, a change in $C_{max}$ will result in a change in $EIC_z$ values (one-compartment models). In contrast, Models 5–8 are based on the assumption that the test compound first has to penetrate into the cells and exert its influence by interacting with an intracellular component; that is, the AUC and/or shape of the AUC will result in a change in $EIC_z$ values (two-compartment models). The observed differences in $EIC_z$ values in relation to experiments in which no compound dilution, replacement, and preincubation are performed (termed *control experiments*) are used as a base for the assessment of the test compound activation ($k_{met}$) and/or inactivation ($k_{el}$) constants. In this case test compound activation means modification of the test compound in a way that the resulting modified test compound is toxic. The opposite is meant with inactivation.

## 6 *IN VITRO* TOXICOKINETICS AND DYNAMICS

### 6.1 Effects Based on Peak Concentrations (One-Compartment Models)

Certain classes of compound either cannot overcome the cell membrane or they specifically interact with components of the outer cell membrane like receptors and

ion channels. In this case, the reaction occurs if enough test compound binds to the receptive membrane component. The threshold treatment period is assumed to be the time needed to reach equilibrium. The presumption is made that the latter period is negligible and is therefore not taken into account in Models 1–4.

***Model 1: The Compound Is Stable*** In this model the concentration of the test compound does not vary during the treatment period. The medium compound concentration at the end of the treatment period corresponds to the initial concentration at the beginning of the treatment period. Since the maximum medium concentration is assumed to be the effective compound concentration, a compound dilution, such as occurs as a result of feeding the cultures by replacing old medium with fresh culture medium without test compound during the treatment period, or a compound preincubation before addition will not affect or shift the effective initial medium compound concentration (EIC)–response curves (OCERs).

***Model 2: The Toxic Compound Is Inactivated*** *In vitro*, compounds are inactivated (eliminated) on the one hand by the cultured cells, or on the other hand extracellularly by decomposition, by reducing the cellular accessibility, such as binding to released products of the cultured cells or by interaction with medium constituents (mainly proteins). The half-life of the compound due to intracellular inactivation and extracellular inactivation as a result of the interaction with released products is determined by the cell density and the cellular activity. Since the volume of the culture medium relative to the number of cells is extremely large, of the three inactivation pathways, the effect of inactivation by cells or their products is mainly negligible. The inactivation due to decomposition is exponential with a constant half-life. The half-life of the test compound that is inactivated by interaction with medium constituents depends on the medium concentrations of these constituents. A nearly constant half-life will only be found at small test compound concentrations relative to the concentration of these constituents; otherwise the half-life will increase exponentially with time.

Due to the inactivation of the test compound, its maximum concentration is present at the moment of addition of the test compound to the cultures. Therefore, a dilution of the culture medium after addition will not affect the OCERs. The maximum medium concentration will only be reduced in the case of compound preincubation, but only if the test compound is inactivated by decomposition or by interaction with medium constituents and not by cell-related processes.

Assuming one compartment and a single treatment of the cultures, the culture medium concentration of the toxic compound ($y_i$) follows the OCER as the one described for blood concentrations after an IV administration *in vivo* under the same assumptions. Thus, Eqs. 1 and 2 are also applicable for the present model for describing the culture medium concentration and the inactivation half-life, respectively.

After preincubation, less toxic compound will be present, resulting in an apparent shift in the OCER in which the effect is related to the test compound at the moment before preincubation was started (initial or starting concentration, EIC). Comparing $EIC_{50}$ values as calculated by Eq. 11, an inactivation of 50% of the compound implies that, before preincubation, twice the concentration had to be present to give the same concentrations at the moment that the dilutions are added to the cultures and thus in concentrations giving the same effect. The presence of equal effects implies

that at the beginning of the treatment period equal toxic compound concentrations were present in the cell cultures. An increase in EICz as result of preincubation can be explained by a partial conversion of the toxic compound into a non-toxic substance. For example if 50% is inactivated during the preincubation period twice the concentration is needed to have at start of the treatment the same concentration of toxic compound. This can be described by the following equation,

$$\mathrm{EIC}_{z1} = \mathrm{EIC}_{z2} \cdot e^{-k_{\mathrm{el}} t_{\mathrm{p}}} \tag{17}$$

where $k_{\mathrm{el}}$ is the elimination constant in the culture medium, $t_{\mathrm{p}}$ is the preincubation duration (difference between the time point at which the preincubation was started and $t_0$), $\mathrm{EIC}_{z1}$ is the obtained $\mathrm{EIC}_z$ in experiments without compound preincubation, and $\mathrm{EIC}_{z2}$ is the $\mathrm{EIC}_z$ in experiments with compound preincubation.

The $k_{\mathrm{el}}$ can be calculated with the following equation, which is obtained by rearranging Eq. 17:

$$k_{\mathrm{el}} = \left(\frac{-1}{t_{\mathrm{p}}}\right) \ln\left(\frac{\mathrm{EIC}_{z1}}{\mathrm{EIC}_{z2}}\right) \tag{18}$$

***Model 3: The Nontoxic Compound Is Modified into a Toxic Compound*** This model describes the concentration and effects of those test compounds that are bioactivated as a result of an interaction with components of the culture medium. Effects of biotransformation by a rate-limiting process inside the cells or at the cell surface are not described by this model but are by Model 5 (Section 6.2). Although a nontoxic test compound may give rise to a range of products, each of them having different toxic effects, generally one product dominates in its toxic effects on the cell culture. The formation of this main important toxic product is subjected to the same kind of processes as described for inactivation of the test compound (Model 2). The only difference is that in Model 3 the product is the toxic compound and in Model 2 the product is the original test compound. If one molecule of test compound gives rise to one toxic molecule, the concentration of the toxic product can be calculated by subtracting the concentration of the test compound at time $t_i$ from the initial test compound concentration at time $t_0$. Otherwise, the test compound concentrations need to be multiplied by a factor representing the largest number of toxic molecules that are obtained from one molecule of test compound. The concentration of toxic product ($A_i$) increases with treatment period ($t_i$) (see Eq. 1) and is

$$A_i = A_0 (1 - e^{-k_{\mathrm{met}} t_i}) \tag{19}$$

where $A_0$ is the final concentration of the toxic product when all test compound is bioactivated, $A_i$ is the concentration of the toxic product at $t_i$, and $k_{\mathrm{met}}$ is the rate constant of activation (*met*abolism) [1].

Like the elimination half-life, the rate of bioactivation can also be expressed in a "half-life period" defined as the time needed to activate half of the test compound. Thus, $A_i = 0.5A_0 = A_0(1 - e^{-k_{\mathrm{met}} t_{1/2}})$ and, in analogy to Eq. 2,

$$t_{1/2} = \frac{\ln(2)}{k_{\mathrm{met}}} \tag{20}$$

Although the nontoxic test compound only evokes toxic effects by its toxic product(s), the extent of toxic effects is related to the concentration of the nontoxic compound added to the culture medium. This characteristic can be used to assess $k_{met}$ experimentally without measuring the concentrations in the culture medium. The $k_{met}$ can be calculated if results from the control experiment are compared with data from experiments using compound dilution, compound replacement, or compound preincubation.

COMPOUND DILUTION According to Eq. 19, the concentration of the toxic product at the time point just before medium dilution ($t_1$) corresponds to

$$A_1 = A_0(1 - e^{-k_{met}t_1}) \tag{21}$$

and at the end of the treatment period after dilution of the medium by a factor $x$ to

$$A_i = \frac{A_0(1 - e^{-k_{met}t_i})}{x} \tag{22}$$

By comparing data obtained from dilution and control experiments, the following scenarios may occur:

1. The medium is diluted by test compound-free medium after full activation of the test compound in the cell cultures. Under these circumstances, the concentration–response curve will not be affected by the dilution.
2. The medium is diluted by test compound-free medium before the test compound is fully activated in the cell cultures.
   (a) The dilution factor and/or $t_1$ is large enough that $A_1 > A_i$. Under these conditions, identical maximum concentrations and thus effects are obtained if

$$A_{01}(1 - e^{-k_{met}t_i}) = A_{02}(1 - e^{-k_{met}t_1})$$

where

$A_{01}$ is the initial, starting concentration of the test compound in experiments without dilution and $A_{02}$ is the initial, starting concentration of the test compound in experiments with dilution.

If the maximum concentrations that are reached in culture medium during cell treatment are directly correlated with the resulting cellular effects,

$$\frac{A_{01}}{A_{02}} = \frac{\mathrm{EIC}_{z1}}{\mathrm{EIC}_{z2}} \tag{23}$$

Thus,

$$\frac{\mathrm{EIC}_{z1}}{\mathrm{EIC}_{z2}} = \frac{1 - e^{-k_{met}t_1}}{1 - e^{-k_{met}t_i}} \tag{24}$$

$k_{met}$ can be calculated after filling in $\mathrm{EIC}z_1$, $\mathrm{EIC}z_2$, $t_1$, and $ti$.

(b) The most common situation is that $x$ is large and so $A_1 < A_i$. Under these circumstances, identical maximum concentrations are obtained if

$$A_{01}(1-e^{-k_{\text{met}}t_i}) = \frac{A_{02}(1-e^{-k_{\text{met}}t_i})}{x}$$

where again $x$ is the dilution factor (see Eqs. 21 and 22). Also, in combination with Eq. 23 and after some rearrangement,

$$\frac{\text{EIC}_{z1}}{\text{EIC}_{z2}} = \frac{1}{x}$$

This implies that under these circumstances $k_{\text{met}}$ *cannot* be assessed.

COMPOUND REPLACEMENT An analogue situation as described earlier occurs if the replacing medium contains the test compound at an identical concentration as present at $t_0$. The $\text{EIC}_z$ in experiments with compound replacement will increase in comparison to control experiments. This will occur because at $t_1$ not only is the test compound replaced but the toxic product formed in the cultures is diluted by a factor $x$. No change or only a negligible change in $\text{EIC}_z$ will occur if the activation half-life is very small and at $t_1$ or $t_i$ the maximum concentration of the toxic metabolite is already present. A similar situation occurs if both $x$ and $t_1$ are very large ($t_1 >> 0.5t_i$). The increase in $\text{EIC}_z$ value is maximum in the case where the concentrations of the toxic product at $t_1$ and $t_i$ are the same (i.e., after complete replacement $t_1 = 0.5t_i$). In the latter situations, the maximum concentration of the toxic product is reached in the culture just before the compound replacement and Eq. 24 can be used for calculating $k_{\text{met}}$. Otherwise, the following correlation between $A_{01}$ (Eq. 21) and $A_{02}$ (see Eq. 22) exists:

$$A_{01}(1-e^{-k_{\text{met}}t_1}) = \frac{A_{02}(1-e^{-k_{\text{met}}t_i})}{x} + \frac{(x-1)A_{02}(1-e^{-k_{\text{met}}(t_i-t_1)})}{x}$$

or in combination with Eq. 23

$$\frac{\text{EIC}_{z1}}{\text{EIC}_{z2}} = \frac{x-e^{-k_{\text{met}}t_i}+(1-x)e^{-k_{\text{met}}(t_i-t_1)}}{x(1-e^{-k_{\text{met}}t_1})}$$

Again, $k_{\text{met}}$ can be calculated after filling in $\text{EIC}_{z1}$, $\text{EIC}_{z2}$, $x$, $t_1$, and $t_i$.

COMPOUND PREINCUBATION If no full activation is achieved, the final concentration $A_i$ of the activated compound is $A_0(1-e^{-k_{\text{met}}(t_i+t_p)})$ instead of $A_0(1-e^{-k_{\text{met}}t_i})$ after inclusion of a preincubation period. To obtain identical concentrations at the end of the treatment period, the starting concentration must be comparably higher to the extent that

$$A_{01}(1-e^{-k_{\text{met}}t_i}) = A_{02}(1-e^{-k_{\text{met}}(t_i+t_p)})$$

By assuming that the effects seen are directly dependent on the final concentration of the bioactivated compound, we find

$$\frac{\mathrm{EIC}_{z1}}{\mathrm{EIC}_{z2}} = \frac{1 - e^{-k_{\mathrm{met}}(t_{\mathrm{p}}+t_i)}}{1 - e^{-k_{\mathrm{met}}t_i}}$$

in which $\mathrm{EIC}_{z1}$ and $\mathrm{EIC}_{z2}$ are the concentrations needed to obtain similar effects ($\mathrm{EIC}_{z1}$ (controls) and $\mathrm{EIC}_{z2}$ (experiment with compound preincubation) according to Eq. 19 and in combination with Eq. 23).

***Model 4: The Nontoxic Test Compound Is Transformed into a Toxic Product that Afterwards Is Inactivated*** This model finds its analogue in the *in vivo* situation after extravascular administration of a toxic compound, when absorption and elimination are first-order processes as described in Section 3. This means that the pattern of the toxic product $A_i$ in the culture medium can be described by the Bateman equation (Eq. 3, also like *in vivo*). For the *in vitro* situation, the absorption constant ($k_{\mathrm{abs}}$) has to be replaced by the metabolization constant ($k_{\mathrm{met}}$) and the ratio ($Df/V$) by $A_0$, representing the maximum concentration of toxic product in the case $k_{\mathrm{el}} = 10^{-\infty}$, or

$$A_i = A_0 k_{\mathrm{met}} \left( \frac{e^{-k_{\mathrm{el}}t_i} - e^{-k_{\mathrm{met}}t_i}}{k_{\mathrm{met}} - k_{\mathrm{el}}} \right) \tag{25}$$

where $k_{\mathrm{met}}$ is the activation constant and $k_{\mathrm{el}}$ is the inactivation (elimination) constant.

Although the nontoxic test compound only evokes toxic effects by its toxic products(s), the extent of toxic effects is directly related to and can be expressed in the concentration of the nontoxic compound like in Model 3. Therefore, $k_{\mathrm{met}}$ and $k_{\mathrm{el}}$ can also be expressed in nontoxic test compound concentrations. Important for the observed effects in Models 1–4 are the peak concentration of the toxic product and the time at which this peak concentration is present in the medium. The time point ($t_{\mathrm{max}}$) in this model at which the peak concentration is reached is

$$t_{\mathrm{max}} = \frac{\ln(k_{\mathrm{met}}/k_{\mathrm{el}})}{k_{\mathrm{met}} - k_{\mathrm{el}}} \tag{26}$$

and the peak concentration ($A_{\mathrm{max}}$) is [1]

$$A_{\mathrm{max}} = A_0 \left( \frac{k_{\mathrm{met}}}{k_{\mathrm{el}}} \right)^{k_{\mathrm{el}}/(k_{\mathrm{el}} - k_{\mathrm{met}})} \tag{27}$$

Up to the time point $t_{\mathrm{max}}$ a direct relationship exists between $\mathrm{EIC}_z$ and $y_i$ of Eq. 25. After $t_{\mathrm{max}}$ the $\mathrm{EIC}_z$ is constant and directly related to $A_{\mathrm{max}}$ (Eq. 27).

COMPOUND DILUTION If the culture medium is replaced by medium without a compound (compound dilution) at a time point after $t_{\mathrm{max}}$, then the $\mathrm{EIC}_z$ value is not affected. A dilution before $t_{\mathrm{max}}$ is similar to the situation described by Model 3: the

maximum concentration of toxic product is present in the culture medium at the end of the treatment period. Comparing data obtained from dilution and control experiments, the same scenarios may occur as in Model 3. However, $EIC_z$ correlates with Eq. 25 and not with Eq. 19. Since before $t_{max}$ $A_i$ can nearly be mimicked by Eq. 19, in practice it will not be possible to discriminate between Model 3 and 4 under these circumstances using compound dilution experiments.

COMPOUND PREINCUBATION In analogy to compound dilution, the value of $t_{max}$ is decisive as to which effect a compound preincubation has on the $EIC_z$ value. Of the four different scenarios that can be discriminated, three are possible in practice: (1) The $t_{max}$ in both cases (with and without preincubation) has a value between $t0$ and $t_i$. In this scenario a compound preincubation has no effect on $EIC_z$. (2). The $t_{max}$ is reached during the preincubation period. In this case a preincubation of the compound dilutions results in an increase in the $EIC_z$ value. The compound behaves very similarly to that described in Model 2. (3) The $t_{max}$ is greater than $t_i$, irrespective of the value of $t_p$, or occurs only after preincubation. Under these circumstances, a compound preincubation results in a decrease in the $EIC_z$ value. The compound behaves as described in Model 3. Since under possibilities (2) and (3) the OCER can be mimicked fairly well by the equations describing Models 2 and 3, respectively, it is not possible to discriminate between these models and Model 4 under experimental conditions.

### 6.2 Effects Based on Accumulation (Two-Compartment Models)

Several compounds of different compound classes seem not to have an acute effect or only after extreme doses. Their toxicity seems to be dependent on the total cumulative dose received and not merely on the treatment schedule. Therefore, their effects seem to be related with the *in vivo* area under the serum concentration–time curve (AUC) (Fig. 1) and not by a single dose or by the maximum serum concentration.

The reason why one toxic compound at a given blood concentration is effective at a short exposure time while other compounds require a prolonged period of time to be effective is assumed to be the difference in the accessibility of the target protein(s) for the toxic test compound. Whereas some compounds introduced into a test system are obviously able to interact directly and freely with the target protein(s) resulting in acute effects, others are not. These latter compounds first have to overcome some kind of "barrier," which is not only determined by diffusion velocity. This barrier consists of saturable processes that limit transfer into the target compartment, such as carrier dependent transport, or saturable processes in which certain enzymes play a key role (formation of toxic product S*; Fig. 7). In the latter case, a complex or a metabolite may be formed. In these kinds of barrier, accumulation is restricted by the number of interaction sites (binding sites). Binding site-limited processes can be described by the Michaelis–Menten relationship [22]:

$$\frac{v_0}{V_{max}} = \frac{[S]}{K_m + [S]} \tag{28}$$

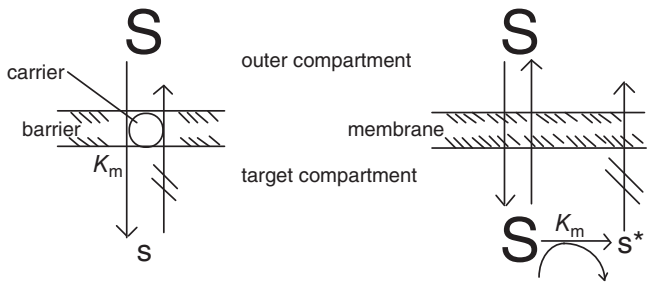


**FIGURE 7** Entry of a toxic compound into the target compartment: example of a carrier-mediated inward transport and of an enzyme-dependent accumulation.

where [S] is the external concentration of the test compound (termed the outer compartment test compound concentration), $v_0$ is the compound transfer velocity per time unit at [S], $V_{max}$ is the maximum compound transfer velocity per time unit, and $K_m$ is the Michaelis–Menten constant representing the [S] at which $v_0 = 0.5V_{max}$.

Besides $K_m$, the stability of the test compound plays an important role in how much of the toxic product finally reaches the target compartment. Previously, we described what occurs if the compound is **(a)** stable (Model 5), **(b)** modified into a nontoxic compound (Model 6), **(c)** modified into a toxic product, the test compound being nontoxic (Model 7), and **(d)** in addition to (c) the resulting toxic product is thereafter modified into a nontoxic compound (Model 8). For these four models, it is assumed that the toxic test compound, test compound complex, or product (termed toxic compound in the target compartment) cannot be released from the target compartment. In other words, the process of accumulation is assumed to be unidirectional. Furthermore, it is assumed that the effects are induced according to the single effector site model (see Section 4.1). For model 5, we discuss the possibility that the effects are due to an interaction with two different effector sites. The compound concentration outside the target compartment is assumed to be homogeneous and is termed the outer compartment. The following models are simplified with the restriction that the delay, which may occur between reaching the toxic compartment concentration and the onset of effects, is not taken into account.

***Model 5: The Test Compound Concentration Is Stable*** Assuming that [S] is constant, the concentration $s$ of the toxic compound within the target compartment equals

$$s = t_i S_f \left( \frac{v_0}{V_{max}} \right) \tag{29}$$

where $t_i$ is the period between the addition and the time of interest, $S_f$ is the maximum concentration increase per time unit of the toxic compound within the target compartment, and $V_0/V_{max}$ is the fraction of compound transferred relative to the maximum.

The relationship between the outer compartment and target compartment concentrations can be obtained by replacing the ratio $v_0/V_{max}$ in Eq. 29 by Eq. 28:

$$s = t_i S_f \left( \frac{[S]}{K_m + [S]} \right) \tag{30}$$

Thus, at a given outer compartment concentration, the maximum target compartment concentration is determined by only $t_i$, $K_m$ (Fig. 8a), and $S_f$. The transfer of the compound into the target compartment may be altered under certain circumstances, such as modified culture conditions. Effects on $S_f$ can be mimicked by changes in $t_i$ (doubling of $S_f$ has the same effect as doubling $t_i$ (see Eq. 30). In contrast, a change in $K_m$ value does not affect the maximum compound concentration within the target compartment after a given treatment period but only affects the concentration at which the transfer into the target compartment reaches maximum values (Fig. 8b). The slope of the outer compartment–target compartment compound concentration relationship remains unchanged under these conditions. By testing different compound concentrations, it is possible to discriminate between effects on $S_f$ and $K_m$ on the target compartment concentration.

After a certain period of time, the concentration of toxic compound/product within the target compartment reaches a critical level: the compound/product starts to noticeably interfere with cell functions (Fig. 9a). It is reasonable to assume that these interferences are directly related with binding of the toxic compound to effector molecules within the target compartment. If the induced effects are concentration dependent according to the one effector site model (Section 4.1), [S] of Eq. 10 has to be changed to $s$ (or $s^*$). The variables $s$ and $s^*$ represent the concentrations of the toxic test compound/product within the target compartment. In the next step, $s$ (or $s^*$) is replaced by Eq. 30, or

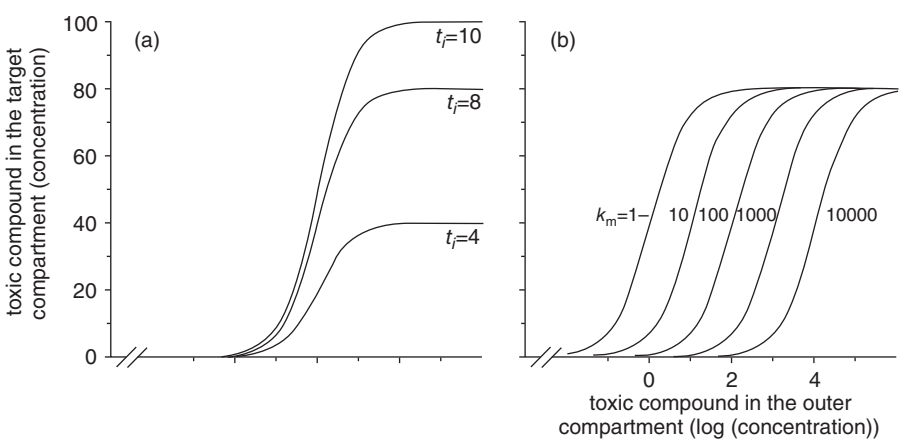


**FIGURE 8** Correlation between the concentration of the toxic compound in the outer compartment and the concentration in the target compartment according to Eq. 30: effects of variat$_i$ons in (a) the treatment period $t_i$ ($S_f$ = 10, $K_m$ = 100) and (b) the Michaelis–Menten constant $K_m$ ($t_i$ = 8; $S_f$ = 10) (arbitrary units).

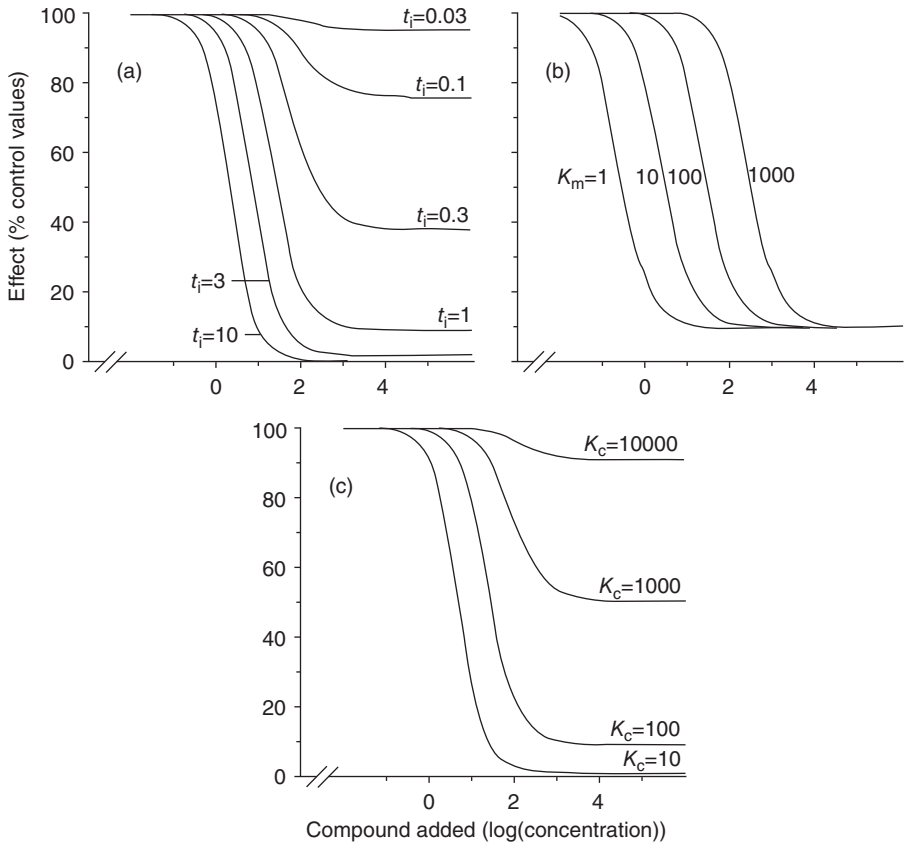


**FIGURE 9** Outer compartment compound concentration–effect relationships (OCERs) according to Eq. 31. $N_c$ = 1.5, $K_c$ = 100, $t_i$ = 1, $S_f$ = 100, and $K_m$ = 100 (arbitrary units) or as mentioned in the figure. Effects on the outcome are shown for variations in (a) the treatment period $t_i$, (b) the Michaelis–Menten constant $K_m$, and (c) the cooperativity constant $K_c$.

$$\frac{E_t}{E_u} = \frac{K_c}{K_c + \left( t_i S_f \left( \frac{[S]}{K_m + [S]} \right) \right)^{N_c}} \tag{31}$$

or

$$\frac{E_t}{E_u} = \frac{1}{1 + \frac{(t_i S_f)^{N_c}}{K_c} \left( \frac{[S]}{K_m + [S]} \right)^{N_c}} \tag{32}$$

**Thus**, *the response constant* $K_c$ *and the slope* $N_c$ in Eq. 31 and 32 *are characterized by the target compartment and not by the outer compartment concentration! In addition, the shape of the present relationship is not the same as the one described by* Eq. 10, *although the obtained relationships are apparently very similar.*

In Eq. 32, $(t_i S_f)^{N_c}/K_c$ behaves like a single constant. This implies that modifications in the $K_c$, $t_i$, or $S_f$ value can be mimicked by changes in one of the other parameters. At a given test compound outer compartment concentration, the parameters $K_c$, $t_i$, and $S_f$ determine the maximum effect level. The maximum effect level is reached when [S] is large, and therefore when the ratio $[S]/(K_m + [S])$ reaches the value 1. Thus, under the assumptions of Model 5, the maximum effect level is *independent* of $K_m$. As clearly demonstrated by Eq. 33, a decrease in $K_m$ value by a factor of $x$ can be mimicked by an increase in test compound in the outer compartment by the same factor.

$$\frac{x[S]}{K_m + x[S]} = \frac{[S]}{(K_m/x) + [S]} \tag{33}$$

Thus, modifications in $K_m$, in contrast to $t_i$, $S_f$, and $K_c$, shift the OCER (Fig. 9b) without effect on its shape. It must be noted that a change in the $K_m$, $K_c$, or $S_f$ value influences the shape but not the slope of the OCER. In analogy to $s$ (or $s^*$), modifications of the compound toxicity due to changes in culture conditions can be based on modifications of $S_f$, $K_m$, and/or $K_c$ values. By measuring the effect of various compound concentrations under different culture conditions and assessing the effects on the maximum effect level and location of the OCER, it is possible to discriminate between effects on $S_f$ or $K_c$ and $K_m$ but not between $S_f$ and $K_c$. To discriminate between the effect on $S_f$ and $K_c$, the concentration of the compound in the target compartment has to be measured and related to the observed effects. The latter, however, will in most cases not be possible.

For calculation of the $EIC_{50}$ value, two reference levels are needed: control values and a reference value for maximum effects. For maximum effects, two different reference lines can be considered: (1) the 0% activity line, corresponding to the effect of the compound at maximum concentration and treatment duration, and (2) the activity percentage reached at large [S] values at $t_i$.

The 0% activity is taken as the reference for maximum effects. Under this assumption, the $EIC_{50}$ is defined as the outer compartment compound concentration at which at the chosen treatment period $E_t = 0.5E_u$. In Eq. 31, if we replace $E_t/E_u$ by 0.5 and [S] by $EIC_{50}$, the following equation can be obtained after some rearrangements:

$$EIC_{50} = \frac{K_m\{(K_c)^{1/N_c}/S_f\}}{t_i - \{(K_c)^{1/N_c}/S_f\}}$$

or

$$EIC_{50} = \frac{K_m}{t_i S_f (K_c)^{-1/N_c} - 1} \tag{34}$$

For calculation of $BMC_5$, where $E_t/E_u = 0.95$, $K_c$ must be replaced by $K_c/19$, or

$$BMC_5 = \frac{K_m}{t_i S_f (K_c/19)^{-1/N_c} - 1} \tag{35}$$

Using this equation, $N_c$, $K_m$, and $S_f$ must be calculated by taking into account the activity percentage reached at large [S] values *at the given* $t_i$. An example of such a relationship between $EIC_{50}$ and $t_i$ according to Eq. 34 is shown in Fig. 10. A positive, fictitious value for $EIC_{50}$ is only obtained if $t_i S_f (K_c)^{-1/N_c} > 1$. In this equation, $(K_c)^{-1/N_c}$ represents the target compartment $EIC_{50}$ (see Eq. 11). Only if this target compartment concentration is reached, can measurement of a 50% reduction be possible.

At short $t_i$ values, when $t_i S_f$ values are near $(K_c)^{1/N_c}$, a doubling of $t_i$ will reduce $EIC_{50}$ much *more* than 50% due to the subtraction of 1 from $t_i S_f (K_c)^{-1/N_c}$. *Thus, the toxicity is not directly related to the AUC*. If the compound concentration in the outer compartment is stable, the AUC after the treatment period $t_i$ is defined by

$$AUC_{t_i} = (AUC_{t_0 \Rightarrow t_i}) = [S]t_i$$

where $t_i$ is the difference between the time at the end of the treatment period and the time at the beginning of the treatment period ($t_0$).

At large $t_i$ the subtraction of 1 from $t_i S_f (K_c)^{-1/N_c}$ will only marginally influence the outcome. Only under these circumstances does a doubling of $t_i$ bisect the $EIC_{50}$ value or, in other words, $EIC_{50}/t_i$ is constant. This implies that only *at large* $t_i$ *can an inverse linear relationship exist between toxicity expressed by* $EIC_{50}$ *value and AUC*.

If the results of two experiments are compared in which only $t_i$ was varied, an estimate of $K_m$ can be made if an inhibition of the parameter of interest of at least

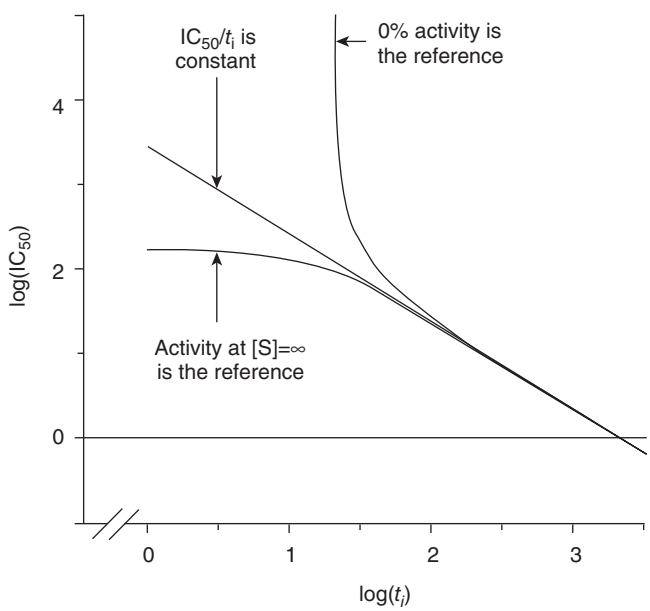


**FIGURE 10** Correlation between $EIC_{50}$ defined by 0% activity or by the activity percentage reached if [S] = ∞ and treatment period $t_i$. The effects are simulated according to the one effector site model (Eq. 34 and 39, respectively) ($S_f$ = 100, $K_c$ = 100,000, $K_m$ = 100, $N_c$ = 1.5) (arbitrary units).

$z$% at maximum [S] is reached. The two different treatment periods are termed $t_{i1}$ and $t_{i2}$. As a new variable, $r$ is introduced representing the ratio $r = (t_{i1} - t_{i2})/t_{i2}$, or $t_{i1} = t_{i2}(1 + r)$. In Models 5–8, the effects after $t_{i1}$ and $t_{i2}$ are the same if $s_1 = s_2$, or

$$(1+r)t_{i2}S_{\mathrm{f}}\left\{\frac{\mathrm{EIC}_{z1}}{K_{\mathrm{m}}+\mathrm{EIC}_{z1}}\right\} = t_{i2}S_{\mathrm{f}}\left\{\frac{\mathrm{EIC}_{z2}}{K_{\mathrm{m}}+\mathrm{EIC}_{z2}}\right\}$$

where $\mathrm{EIC}_{z1}$ and $\mathrm{EIC}_{z2}$ are the initial outer compartment concentration after $t_{i1}$ and $t_{i2}$, respectively, resulting in a reduction of the parameter of interest by $z$%. Or after rearrangement,

$$K_{\mathrm{m}} = \frac{r\cdot\mathrm{EIC}_{z1}\cdot\mathrm{EIC}_{z2}}{\mathrm{EIC}_{z2}-(r+1)\cdot\mathrm{EIC}_{z1}} \tag{36}$$

Besides the possibility of calculating $K_{\mathrm{m}}$, this equation also indicates that a relation between treatment period and $\mathrm{EIC}_{z1}$ and $\mathrm{EIC}_{z2}$ exists, independent of $S_{\mathrm{f}}$, $K_{\mathrm{c}}$, and $N_{\mathrm{c}}$.

The reference for maximum effects is defined by the activity percentage reached at large [S] values. As mentioned before, at large [S] values the ratio $[S]/(Km + [\mathrm{S}])$ in Eq. 31 reaches the value 1. Thus, the maximum possible effect level $(E_{\mathrm{tmax}}/E_{\mathrm{u}})$ after the given treatment period $t_i$ equals

$$\frac{E_{\mathrm{t\,max}}}{E_{\mathrm{u}}} = \frac{K_{\mathrm{c}}}{K_{\mathrm{c}}+(t_iS_{\mathrm{f}})^{N_{\mathrm{c}}}} \tag{37}$$

The $\mathrm{EIC}_{50}$ at the treatment period chosen under the present assumptions is defined as the outer compartment concentration at which

$$1-\frac{E_{\mathrm{t}}}{E_{\mathrm{u}}} = 0.5\left(1-\frac{E_{\mathrm{t\,max}}}{E_{\mathrm{u}}}\right) \tag{38}$$

In Eq. 38, $(E_{\mathrm{t}}/E_{\mathrm{u}})$ can be replaced by Eq. 31 and $(E_{\mathrm{tmax}}/E_{\mathrm{u}})$ by Eq. 37. After some rearrangements, the following is obtained:

$$\mathrm{EIC}_{50} = \frac{K_{\mathrm{m}}}{\{2+(t_iS_{\mathrm{f}})^{N_{\mathrm{c}}}/K_{\mathrm{c}}\}^{1/N_{\mathrm{c}}}-1} \tag{39}$$

In analogy, the $\mathrm{BMC}_5$ can be calculated:

$$1-\frac{E_{\mathrm{t}}}{E_{\mathrm{u}}} = 0.05\left(1-\frac{E_{\mathrm{t\,max}}}{E_{\mathrm{u}}}\right)$$

or

$$\mathrm{BMC}_5 = \frac{K_{\mathrm{m}}}{\{20+19(t_iS_{\mathrm{f}})^{N_{\mathrm{c}}}/K_{\mathrm{c}}\}^{1/N_{\mathrm{c}}}-1} \tag{40}$$

As in the previous case, in which the 0% activity is taken as a reference, only at large $t_i$ can doubling of $t_i$ bisect the $EIC_{50}$ value. However, in contrast to the previous case, at very small $t_i$ a doubling of $t_i$ will reduce the $EIC_{50}$ value by *less* than 50%. This is due to the presence of the constant with a value of 2 in the denominator. Thus, again toxicity is inversely linearly correlated with the AUC *only* at large treatment periods (Fig. 10). Roughly, one can say that a linear relationship between AUC and $EIC_{50}$ exists as far as Eqs. 34 and 39 give similar $EIC_{50}$ values.

COMPOUND DILUTION If during the treatment period at a time $t_1$ the medium of the cultures is partially replaced by fresh medium without the test compound, the remaining compound is diluted, corresponding to the ratio of remaining old medium divided by total medium volume (factor $x$). As a result, the increase in $s$ per time unit is reduced after $t_1$. Mathematically, the obtained OCER can be described by Eq. 32, including an adaptation for the dilution of [S] at $t_1$:

$$\frac{E_{\text{tdil}}}{E_{\text{u}}} = \frac{K_{\text{c}}}{K_{\text{c}} + \left( t_1 S_{\text{f}} \left\{ \frac{[\text{S}]}{K_{\text{m}} + [\text{S}]} \right\} + (t_i - t_1) S_{\text{f}} \left\{ \frac{x[\text{S}]}{K_{\text{m}} + x[\text{S}]} \right\} \right)^{N_{\text{c}}}} \tag{41}$$

where $E_{\text{tdil}}$ and $E_{\text{u}}$ are values of the parameter of interest at the end of the treatment period, in which a dilution or no dilution step was carried out, respectively; $t_1 S_{\text{f}}\{[\text{S}]/(K_{\text{m}} + [\text{S}])\}$ is the concentration of the test compound in the target compartment at the moment of dilution ($t_1$), and $(t_i - t_1)S_{\text{f}}\{x[\text{S}]/(K_{\text{m}} + x[\text{S}])\}$ is the increase of test compound in the target compartment from $t_1$ to the end of the treatment period ($t_i$).

At large [S], however, the ratio $x[\text{S}]/(K_{\text{m}} + x[\text{S}])$ is still near 1 except if $x$ is extremely small (see Fig. 8 and Eq. 30). This implies that the *same* maximum effect level is obtained with or without compound dilution. Depending on the location of the maximum effect level at $t_1$, two different situations can be discriminated after a compound dilution.

(a) *The parameter of interest is completely inhibited at $t_1$ at maximum [S] values.* With decreasing $x$ values, the OCER found at $t_i$ will shift asymptotically toward the OCER measured at $t_1$. No swallowed OCER will be obtained (Fig. 11). Independent of the value of $t_i$, if $x[\text{S}]$ is relatively small in comparison to $K_{\text{m}}$, then in $(t_i - t_1)S_{\text{f}}\{x[\text{S}]/(K_{\text{m}} + x[\text{S}])\}$, a decrease in the $x$ value can be mimicked by an increase in $t_i$ and by that of $t_i - t_1$.

(b) *Even at very large* [S], *s (or s*) at $t_1$ is still too low to inhibit completely the parameter of interest.* Only the part of the curve that represents the effect due to the treatment period to $t_1$ will behave in a similar way as described under (a). The other part will be shifted corresponding to the dilution. A dilution of 0.1 will shift the latter part of the relationship by a factor of 10, a dilution of 0.01 will result in a shift of a factor 100, and soon (see Fig. 11). At relatively large $x$ values (0.01–1), only a change in the apparent $N_{\text{c}}$ will be observed and a swallowed OCER cannot be recognized. A complete removal of the compound at $t_1$ ($x = 10^{-\infty}$) will result in an OCER corresponding to the effect of a treatment to $t_1$. Thus, only under these extreme situations (complete removal) can a dilution of the compound be mimicked by a reduction in the total treatment period without dilution!!

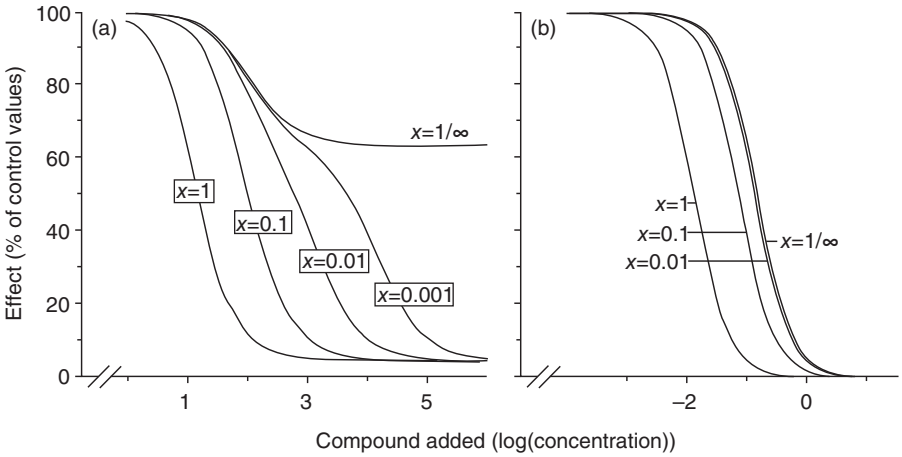


**FIGURE 11** Simulation of initial outer compartment concentration–effect relationship (OCER) according to Eq. 41: effects on the outcome due to variations of dilution factor $x$ ($K_m$ = 100, $K_c$ = 100, $N_c$ = 1.5 (arbitrary units)). (a) $t_1$ = 0.15 and $t_i$ = 1.65, (b) $t_1$ = 150 and $t_i$ = 1650.

The $K_m$ value can be estimated if an inhibition of the parameter of interest is present by more than $z$%, at least at maximum [S]. The reference for maximum effects is 0% (complete inhibition or absence of the parameter of interest). Comparing the effective outer compartment concentrations giving the same effect $z$ in experiments with and without compound dilution: $E_t/E_u$ (Eq. 31) = $E_{tdil}/E_u$ (Eq. 41), or

$$\frac{K_c}{K_c + \left((1+r)t_1 S_f \left\{\frac{\mathrm{EIC}_{z1}}{K_m + \mathrm{EIC}_{z1}}\right\}\right)^{N_c}} = \frac{K_c}{K_c + \left(t_1 S_f \left\{\frac{\mathrm{EIC}_{z2}}{K_m + \mathrm{EIC}_{z2}}\right\} + r t_1 S_f \left\{\frac{x\mathrm{EIC}_{z2}}{K_m + x\mathrm{EIC}_{z2}}\right\}\right)^{N_c}}$$

where $\mathrm{EIC}_{z1}$ and $\mathrm{EIC}_{z2}$ are the $\mathrm{EIC}_z$ values measured in experiments without and with compound dilution, respectively, at $t_1$, and $r$ is the ratio $(t_i - t_1)/t_1$.

After some rearrangements, the following equation can be obtained:

$$K_m = \frac{\mathrm{EIC}_{z2}\{x\mathrm{EIC}_{z2}(1+r) - \mathrm{EIC}_{z1}(x+r)\}}{\mathrm{EIC}_{z1}(1+r) - \mathrm{EIC}_{z2}(1+rx)} \tag{42}$$

At extremely small $x$ values ($x \Rightarrow 10^{-\infty}$), Eq. 42 resembles Eq. 36.

COMPOUND REPLACEMENT AND PREINCUBATION Compound replacement and compound preincubation have no effect, since the concentration of the test compound in the cultures and within the outer compartment will not be modified due to these treatments.

*Special Cases: Two or More Effector Sites Are Present* If two or more effector sites are present, these various sites may be located in the same or in different target compartments.

**(a)** *Both sites are located within the target compartment.* The interaction between test compound and the main effector site need not always result in cell death. When numerous effector sites play a role in the observed effects, the resulting OCER can be described by Eq. 16, replacing [S] by the $s$ of Eq. 30. The resulting equation is indistinguishable from Eq. 31. Interactions of the test compound with two effector sites (the main and another effector site causing cell death) will result in swallowed or at least a less steep OCER. Mathematically, this OCER can be described by Eq. 14, replacing [S] by the $s$ (or $s^*$) of Eq. 30. The latter substitution results in the following equation:

$$\frac{E_t}{E_u} = \frac{\dfrac{K_{c1}E_{u1}}{K_{c1} + \left(t_i S_f \left\{\dfrac{[S]}{K_m + [S]}\right\}\right)^{N_{c1}}} + \dfrac{K_{c2}E_{u2}}{K_{c2} + \left(t_i S_f \left\{\dfrac{[S]}{K_m + [S]}\right\}\right)^{N_{c2}}}}{E_{u1} + E_{u2}} \tag{43}$$

After a short treatment period, the maximum test compound concentration in the target compartment will be low. After reaching a certain concentration, the compound will interact with the high-affinity, in the present example (Fig. 12) stimulatory, effector site. Parallel with a prolongation of the treatment period, the maximum compound concentration within the target compartment will increase.

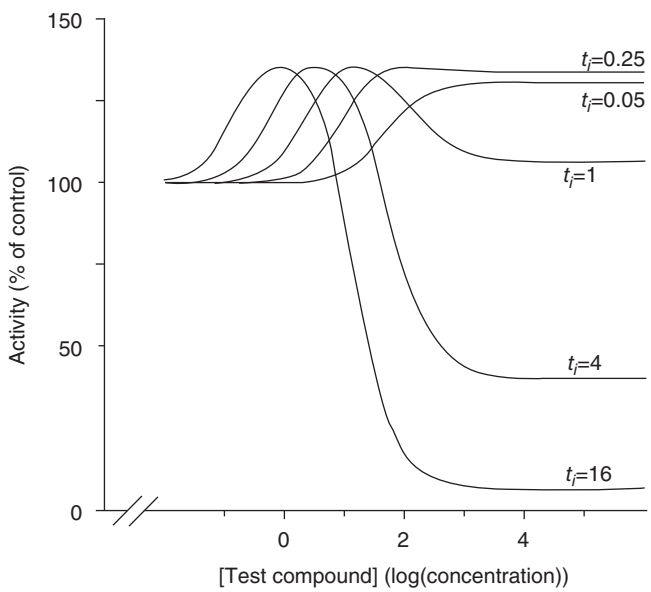


**FIGURE 12** Simulation of outer compartment compound concentration–effect relationships according to Eq. 43 in relation to the treatment period $t_i$. ($E_{u1} = -40$, $N_{c1} = 1.5$, $K_{c1} = 0.1$, $E_{u2} = 140$, $N_{c2} = 1.5$, $K_{c2} = 100$, $S_f = 10$, and $K_m = 100$ (arbitrary units)).

The OCER of the first effector site will shift to the left. In addition, other cell functions will be affected due to the interaction with the second effector site. With further prolongation of the treatment period, both OCERs will shift to the left. Since only the combined effects can be detected, this shift will be reflected in a change in the overall shape of the outer culture medium concentration–response curve (Fig. 12).

**(b)** *The two effector sites are located in different compartments*. One of these compartments may represent a directly and freely accessible compartment, that is, the compartment to which the compound is added, and is termed the "outer" compartment. The second compartment represents a compartment outside the latter compartment. The compound first has to be transferred to this compartment and it is termed the "target" compartment. This implies that, as a result of an increase of the treatment period, only that part of the OCER will be shifted which is the result of the effector site being located within the target compartment. The other part of the OCER remains unaffected by a variation of the treatment period. If the various effector sites induce the *same* effect(s), their contributions for the observed effect(s) will change with the increase in treatment period. Otherwise, if the effector sites affect *different* functional parameters, the characteristics of the altered functional state of the cell will depend on the treatment period. *With prolonged treatment periods, the effect induced by the effector site located in the outer compartment will be totally overshadowed by the effect induced by the other effector site(s) inside the target compartment*. Often, two effects are characterized by different $N_c$ values. In this case, a treatment-period-dependent change in $N_c$ values will be observed. The sum of both effects can be characterized mathematically by multiplying Eq. 10, which describes the treatment-period-independent effects, by Eq. 31 or 43, which describes the transfer-dependent effects. Since transfer-dependent and acute effects can be described according to the one site effector model, the OCER can be characterized by the following equation:

$$\frac{E_t}{E_u} = \left\{ \frac{K_{c1}}{K_{c1} + \left( t_i S_f \left\{ \frac{[S]}{K_m + [S]} \right\} \right)^{N_{c1}}} \right\} \left\{ \frac{K_{c2}}{K_{c2} + [S]^{N_{c2}}} \right\} \tag{44}$$

in which effector site 1 (characterized by $K_{c1}$ and $N_{c1}$) is responsible for the transfer-dependent effects and effector site 2 (characterized by $K_{c2}$ and $N_{c2}$) is responsible for transfer-independent effects. In analogy, the same transition can be made if the transfer-dependent effects can be described according to the two effector site model (not shown).

The presence of such a mixture of effects, in which one effector site is located in the outer compartment and the other(s) is not, is typical for toxic compounds, with different $EIC_{50}$ values for acute and long-term exposure (Fig. 13). For acute effects, no transfer into a certain target compartment is needed. Mainly the effective concentration is high and the OCER is steep ($N_c$ is large) relative to the transfer-dependent effects. To avoid intermediate effects, representing a mixture of transfer-dependent and transfer-independent effects, prolonged treatment periods are preferred.

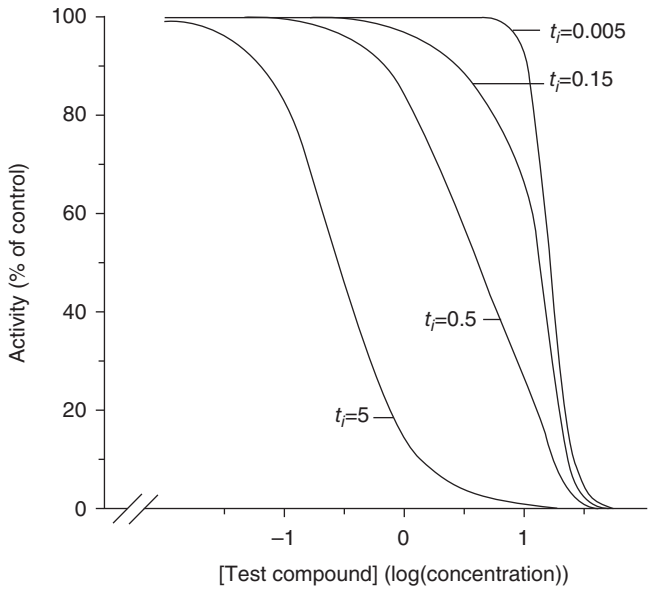


**FIGURE 13** Simulation of an outer compartment compound concentration–effect relationship (OCER) according to Eq. 44 in relation to the treatment period $t_i$ ($E_u$ = 100). Target compartment effector site: $N_{c1}$ = 1.5, $K_{c1}$ = 0.1, $S_f$ = 10 and $K_m$ = 10. Outer compartment effector site: $N_{c2}$ = 5, $K_{c2}$ = 100,000; (all in arbitrary units).

*Special Cases: The Accumulation Is Limited* In some cases accumulation of the test compound in the target compartment seems to be limited. Apparently only during the first part of the treatment period does a correlation between toxic effects and treatment period exist at a constant outer compartment test compound concentration.

An explanation for this phenomenon is the presence of a retro transport of the test compound (or its metabolite) out of the target compartment (Fig. 14). As a result, the accumulation in the target compartment is limited. The retrotransport may be due to diffusion or may be the result of a carrier-dependent retrotransport of the test compound. The increase of the target compartment concentration is maximum at small $s_i$ (Fig. 14). A maximum concentration is reached the moment that the outward flux is equivalent to the inward flux. It must be noted that test compounds that are metabolized or activated inside the target compartment and the metabolite or activated test compounds that are able to leave the target compartment in one way or another behave identically.

The target compartment concentration $s$ is dependent on the rat of influx (Fig. 14 left) or conversion rate (Fig. 14 right) and efflux rate. This influx/conversion is constant due to a constant outer compartment concentration (Eq. 30). The efflux, however, is dependent on the target compartment concentration and thus changes during the treatment period. If the efflux is carrier dependent, this may be described, in analogy to Eq. 4, as follows:

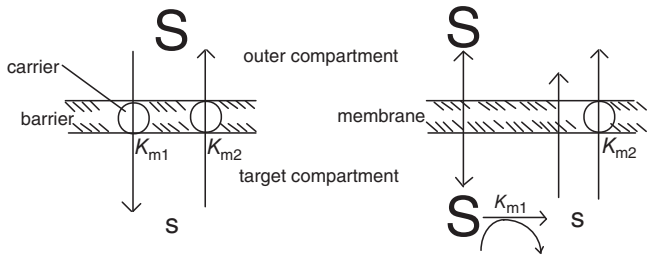


**FIGURE 14** Entry of a toxic compound into the target compartment: example of a carrier mediated inward and outward transport (left) and of an enzyme-dependent accumulation with diffusion or carrier-mediated outward flux (right).

$$s_{\text{out},i} = S_{\text{f}}\left(\frac{s_i}{K_{\text{m}} + s_i}\right) \tag{45}$$

where $s_{\text{out},i}$ is the quantity of compound transported to the outer compartment expressed in concentration units of the target compartment, $S_{\text{f}}$ is the maximum concentration decrease per time unit of the toxic compound within the target compartment, $s_i$ is the target compound concentration at $t_i$, and $K_{\text{m}}$ is the Michaelis–Menten constant representing the $s_i$ at which efflux velocity $v_0 = 0.5V_{\max}$.

The concentration $s_i$ can be described as the difference between the quantity that is transported into the target compartment (Eq. 30) and the quantity that is transported out of the target compartment during the treatment period (see Eq. 45), or

$$s_i = t_i S_{\text{f1}}\left(\frac{[\text{S}]}{K_{\text{m1}} + [\text{S}]}\right) - S_{\text{f2}}\int_{\tau=0}^{\tau=t_i}\left(\frac{s_\tau}{K_{\text{m2}} + s_\tau}\right)d\tau \tag{46}$$

in which $s_\tau$ is the target compound concentration at $t = \tau$, $S_{\text{f1}}$ and $S_{\text{f2}}$ are the maximum concentration increase and decrease, respectively, per time unit of the toxic compound within the target compartment, and $K_{\text{m1}}$ and $K_{\text{m2}}$ are the Michaelis–Menten constants representing $s_i$ at which, respectively, inward and outward flux velocity $v_0 = 0.5V_{\max}$.

Since the target compartment concentration is present on the left and right sides of Eq. 46, this equation can be simplified not toward $t_i$ but only toward $s_i$. After differentiation of Eq. 46 the following can be obtained:

$$S_{\text{f1}}\left(\frac{[\text{S}]}{K_{\text{m1}} + [\text{S}]}\right) - S_{\text{f2}}\left(\frac{s}{K_{\text{m2}} + s}\right) = \frac{ds}{dt} \tag{47}$$

or after some rearrangements

$$t_i = \frac{s_i}{a - S_{\text{f2}}} + \left(\frac{K_{\text{m2}}}{a - S_{\text{f2}}}\right)\left(1 - \frac{a}{a - S_{\text{f2}}}\right)\log\left(1 + \frac{s_i(a - S_{\text{f2}})}{aK_{\text{m2}}}\right) \tag{48}$$

in which a stands for $S_{\text{f1}}[\text{S}]/(k_{\text{m1}} + [\text{S}])$

For describing the OCER, for each $S$ and $t_i$ the $s_i$ must be estimated using a numerical approach. An example of a relationship between $s_i$ and $t_i$ is presented in Fig. 15. Thereafter, [S] of Eq. 10 is replaced by the estimated $s_i$. The characteristics of the OCER can be estimated as described above.

A dilution of the test compound during the treatment period will result in an adjustment of the target compartment concentration asymptotically to the new equilibrium according the new outer compartment concentration. A dilution will have no effect on the OCER if the maximum target concentration is reached at the time of dilution. Thus, it appears that the treatment period is unimportant for the toxicity and important only for the peak concentration (Model 1). Therefore, the system has characteristics in-between Model 1 and Model 5 without outward transport. Compound replacement and compound pretreatment will have no effects on toxicity, as mentioned for Model 5 in the absence of outward transport.

***Model 6: The Toxic Compound Is Inactivated*** In analogy to Model 2, it is assumed that the compound is inactivated within the outer compartment during the treatment period. The concentration just after the addition of the compound ($S_0$) is reduced with increase of the treatment period according to Eq. 1. [S] at time point $t_i$ ($S_i$) is equal to $S_0 e^{-k_{el} t_i}$. Not only the concentration in the outer compartment, per time unit ($t - t_{-1}$), but also the ratio $[S_i]/(K_m + [S_i])$ is slowly reduced during the treatment period. Therefore, the quantity of the test compound that is transferred from the outer compartment into the target compartment is continuously reduced. No compound is transferred at the moment that no compound is left in the outer compartment. The final concentration within the target compartment represents the

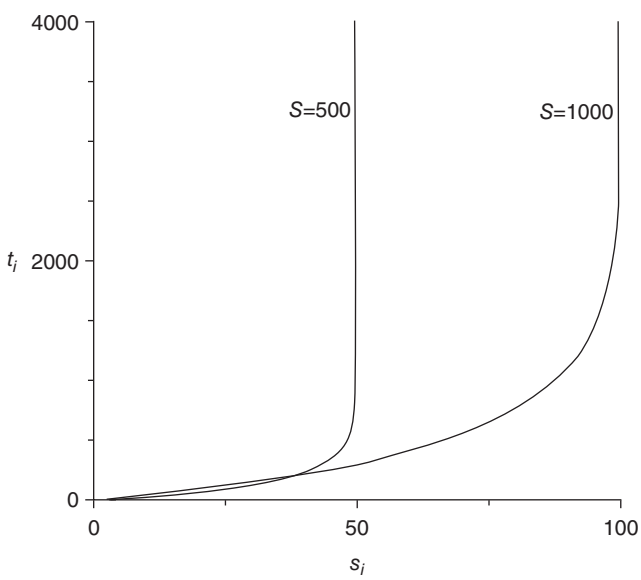


**FIGURE 15** Correlation between the concentration of the toxic compound in the target compartment and the treatment period according to Eq. 48: effects of variations in the outer compartment concentration $S$ ($S_{f1}$ = 10, $S_{f2}$ = 10, $K_{m1}$ = 100, $K_{m2}$ = 100 (arbitrary units)).

quantity of compound that is transferred into the target compartment over the whole treatment period. Thus, the concentration within the target compartment is

$$s_i = \sum_{t=0}^{i} (t - t_{-1}) S_f \left\{ \frac{S_0 e^{-k_{el}t}}{S_0 e^{-k_{el}t} + K_m} \right\}$$

or

$$s_i = S_f \int_{t=0}^{i} \left\{ \frac{S_0 e^{-k_{el}t}}{S_0 e^{-k_{el}t} + K_m} \right\} dt$$

Thus,

$$s_i = \left( \frac{S_f}{k_{el}} \right) \ln \left\{ \frac{K_m + S_0}{K_m + S_0 e^{-k_{el}t_i}} \right\} \tag{49}$$

At large $S_0$ the ratio $(K_m + S_0)/(K_m + S_0 e^{-k_{el}t_i})$ has the value $e^{k_{el}t_i}$. After introducing this factor into Eq. 49, the following equation is obtained: $s_i = S_f t_i$. This equation is identical to Eq. 30. This means that *at large $S_0$ the maximum concentration level is time dependent and not dependent on $K_m$ and $k_{el}$ (see also Fig. 16). Furthermore, this implies that at large $S_0$ the same maximum s level is found within the target compartment independent of the stability of the toxic test compound!!*

$S_0 e^{-k_{el}t_i}$ will be 0 at the moment that all compound is inactivated. At this time point, the maximum target compartment concentration ($s_{max}$) is reached.

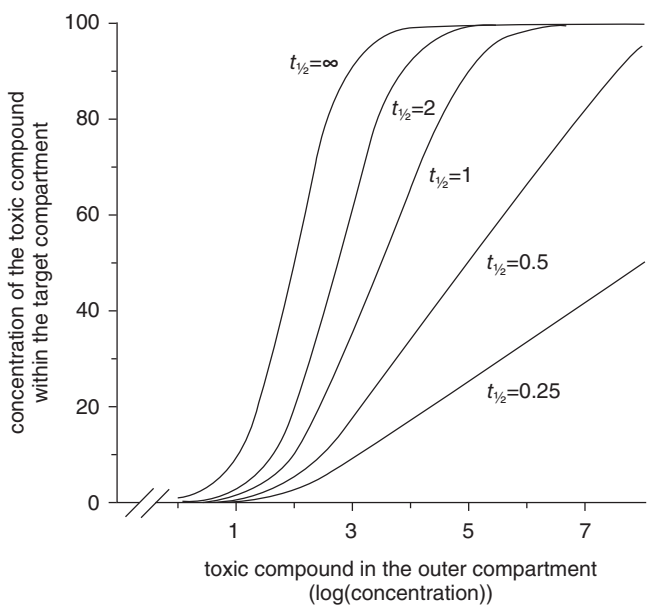


**FIGURE 16** Correlation between the concentration of the toxic compound in the outer compartment and the concentration in the target compartment according to Eq. 49: effects of variations in the elimination half-life ($t_i$ = 10; $S_f$ = 10; $K_m$ = 100 (arbitrary units)).

$$s_{\max} = \left(\frac{S_f}{k_{el}}\right)\ln\left\{\frac{K_m + S_0}{K_m}\right\} \tag{50}$$

If the induced effects are target compartment concentration-dependent, the OCER can be simulated in analogy to Model 5. Since the effects can be described according to the one effector site model, [S] of Eq. 10 can be replaced by Eq. 49, or

$$\frac{E_t}{E_u} = \frac{K_c}{K_c + \left(\left(\frac{S_f}{k_{el}}\right)\ln\left\{\frac{K_m + S_0}{K_m + S_0 e^{-k_{el}t_i}}\right\}\right)^{N_c}} \tag{51}$$

An example of such a relationship is shown in Fig. 17. Please note that *the maximum effect level is only dependent on $t_i$ (Fig. 17a) and not dependent on $k_{el}$* (*Fig. 17b*).

Taking the 0% activity as a reference for maximum effects, the $EIC_{50}$ value at the given treatment period can be calculated in analogy to Eq. 34 by $E_t = 0.5E_u$. By replacing $S_0$ by $EIC_{50}$ and $E_t/E_u$ by Eq. 51, the following equation can be obtained after some rearrangements:

$$EIC_{50} = \frac{K_m(v-1)}{1 - ve^{-k_{el}t_i}} \tag{52}$$

in which $v$ stands for

$$v = \exp\left(\frac{k_{el}(K_c)^{1/N_c}}{S_f}\right) \tag{53}$$

For calculation of $BMC_5$, $K_c$ must be replaced by $K_c/19$.

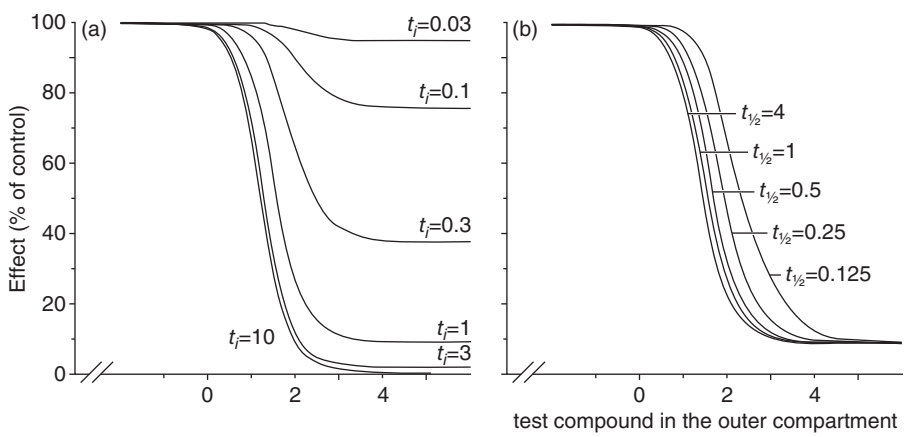


**FIGURE 17** Simulation of outer compartment compound concentration–effect relationships (OCERs) according to Eq. 51 ($N_c = 1.5$, $K_c = 100$, $t_i = 1$, $S_f = 100$, and $K_m = 100$ (arbitrary units) or as mentioned in the figure): effects of variations in (a) the treatment period $t_i$ and (b) the elimination half-life ($t_{1/2}$).

The reference value for maximum effects can also be defined as the activity percentage reached at large [S] values. In analogy to Eq. 39, Eq. 38 is used for the calculation of the $EIC_{50}$. In this case, the $EIC_{50}$ equals Eq. 52; however, in this case $v$ stands for

$$v = \exp\left\{\left(\frac{k_{el}}{S_f}\right)\left(\frac{2K_c}{1+K_c/(t_i S_f)^{N_c}} - K_c\right)^{1/N_c}\right\} \tag{54}$$

At short treatment periods, the relationship between $EIC_{50}$ and the duration of the treatment period is similar to the one described for Model 5 (compare Figs 10 and 18). Because no test compound is present in the outer compartment at large $t_i$, no change in $EIC_{50}$ value is observed with increasing $t_i$ (Figs. 17 and 18). *As a result, only during a short time period does a nearly linear relationship between $t_i$ and $EIC_{50}$ exist!* In contrast to Model 5, a linear relationship between AUC and $EIC_{50}$ may not automatically be assumed if Eq. 52 in combination with Eq. 54 and Eq. 52 in combination with Eq. 53 give similar $EIC_{50}$ values!

COMPOUND DILUTION A dilution of the compound during the treatment period will have less effect on the OCER as in the previous case (Model 5), since during the period before the dilution some test compound is already inactivated. To measure identical effects in cultures with or without compound dilution, the concentrations within the target compartment(s) need to be the same at $t_i$. Thus, $s_1$(without compound dilution) = $s_2$(with compound dilution). If in the cultures without compound dilution at $t_i$ no compound is present, then

$$\left(\frac{S_f}{k_{el}}\right)\ln\left\{\frac{K_m + EIC_{z1}}{K_m}\right\} = \left(\frac{S_f}{k_{el}}\right)\ln\left\{\frac{K_m + EIC_{z2}}{K_m + EIC_{z2}e^{-k_{el}t_1}}\right\} + \left(\frac{S_f}{k_{el}}\right)\ln\left\{\frac{K_m + xEIC_{z2}e^{-k_{el}t_1}}{K_m}\right\} \tag{55}$$

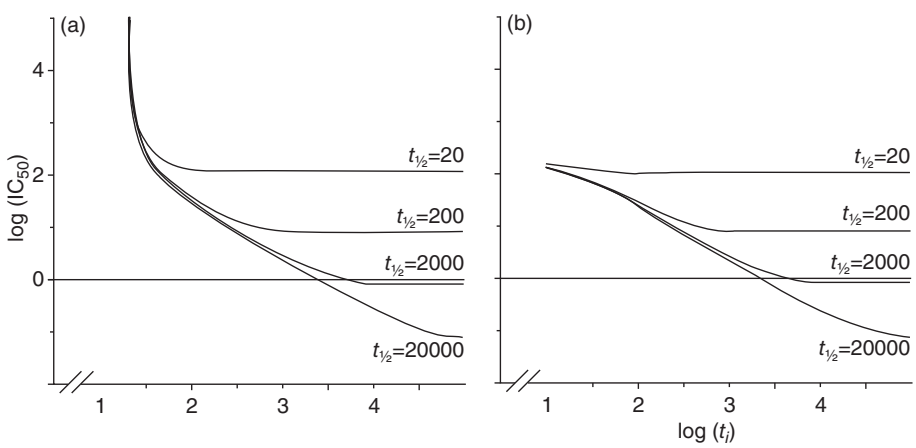


**FIGURE 18** Correlation between $EIC_{50}$ defined by 0% activity (a) or by the activity percentage reached if [S] = ∞ and treatment period $t_i$ (b). The effects are simulated according to the one effector site model (Eq. 52 and with $v$ defined by Eqs. 53 and 54, respectively) ($S_f$ = 100, $K_c$ = 100,000, $K_m$ = 100, $N_c$ = 1.5 (arbitrary units)).

in which $EIC_{z1}$ is the obtained $EIC_z$ in experiments without dilution, $EIC_{z2}$ is the $EIC_z$ in experiments in which the compound was diluted during the treatment period, $(S_f/k_{el})\ln\{(k_m + EIC_{z1})/k_m\}$ is the test compound concentration in the target compartment at $t_i$ in experiments without dilution according to Eq. 48, $(S_f/k_{el})\ln\{(K_m + EIC_{z2})/(K_m + EIC_{z2}e^{-k_{el}t_1})\}$ is the test compound concentration in the target compartment at $t1$ according to Eq. 47, and $(S_f/k_{el})\ln\{(K_m + xEIC_{z2}e^{-k_{el}t_1})/K_m\}$ is the increase of target compartment concentration after dilution by a factor $x$ over the period $t_1$ to $t_i$ (according to Eq. 50).

After a rearrangement of Eq. 55, we find

$$k_{el} = \left(\frac{-1}{t_1}\right)\ln\left(\frac{K_m(EIC_{z2} - EIC_{z1})}{EIC_{z2}\{K_m(1-x) + EIC_{z1} - EIC_{z2}x\}}\right) \tag{56}$$

or

$$K_m = \frac{(xEIC_{z2} - EIC_{z1})EIC_{z2}e^{-k_{el}t_1}}{(1-x)EIC_{z2}e^{-k_{el}t_1} + EIC_{z1} - EIC_{z2}} \tag{57}$$

The same equations can also be used if experiments with different treatment periods are compared. In this case, $x = 0$ and $t_1$ and $t_i$ represent the different treatment periods. The disadvantage of Eq. 56 and 57 is that $K_m$ has to be known in order to calculate $k_{el}$. It may be possible to estimate $k_{el}$ and $K_m$ by including within the same experiment the effects of a dilution at different time points to obtain Eq. 57 with other values for $EIC_{z2}$ and $t_1$. Since in both obtained equations the value for $K_m$ is the same, $k_{el}$ can theoretically be resolved.

A more complex situation exists if at the end of the treatment period not all test compound is inactivated. In this case, a correction needs to be included for the remaining compound. Therefore, Eq. 50 may not be used. Under the present assumptions, Eq. 55 is adapted to

$$\left(\frac{S_f}{k_{el}}\right)\ln\left\{\frac{K_m + EIC_{z1}}{K_m + EIC_{z1}e^{-k_{el}t_i}}\right\} = \left(\frac{S_f}{k_{el}}\right)\ln\left\{\frac{K_m + EIC_{z2}}{K_m + EIC_{z2}e^{-k_{el}t_1}}\right\} + \left(\frac{S_f}{k_{el}}\right)\ln\left\{\frac{K_m + xEIC_{z2}e^{-k_{el}t_i}}{K_m + xEIC_{z2}e^{-k_{el}t_1}e^{-k_{el}(t_i - t_1)}}\right\}$$

or after rearrangement,

$$\begin{aligned} K_m = \{&(1-x)EIC_{z1}EIC_{z2}e^{-k_{el}(t_i-t_1)} + (xEIC_{z2} - EIC_{z1})EIC_{z2} + \\ &(EIC_{z2} - EIC_{z1})EIC_{z2}e^{-k_{el}(t_i-t_1)}\}/ \\ &\{(1-x)EIC_{z2} + (xEIC_{z2} - EIC_{z1})e^{-k_{el}(t_i-t_1)}\} \end{aligned} \tag{58}$$

Again, by changing the time point of dilution as mentioned before, $k_{el}$ and $K_m$ can theoretically be calculated.

COMPOUND REPLACEMENT Compound replacement during treatment maximally results in a doubling of the target compartment concentration if all compound is inactivated at the moment of replacement and in addition at the moment that the

measurement is done again all compound is inactivated. If not all compound is inactivated at $t_1$ or at $t_i$, the change will be much less. Because of this, an effect on $EIC_{50}$ will hardly be measurable under experimental conditions. Thus, the description of models 6–8 for the effect of compound replacement is only briefly discussed.

COMPOUND PREINCUBATION During the preincubation period ($t_p$), the toxic compound is inactivated as described by Eq. 1. Thus, at $t_0$ only $S_0\exp(-k_{el}t_p)$ is left of the test compound. The presence of identical effects in cultures with and without compound preincubation implies that at the time of measurement the concentrations within the target compartment are the same ($s_1$ (without) = $s_2$ (with)). Thus,

$$\left(\frac{S_f}{k_{el}}\right)\ln\left\{\frac{K_m + EIC_{z1}}{K_m + EIC_{z1}e^{-k_{el}t_i}}\right\} = \left(\frac{S_f}{k_{el}}\right)\ln\left\{\frac{K_m + EIC_{z2}e^{-k_{el}t_p}}{K_m + EIC_{z2}e^{-k_{el}t_p}e^{-k_{el}t_i}}\right\} \tag{59}$$

where $EIC_{z1}$ is the obtained $EIC_z$ in experiments without compound preincubation, and $EIC_{z2}$ is the $EIC_z$ in experiments with compound preincubation.

*After rearrangement of Eq. 59, Eq. 18 is obtained! Thus, from preincubation experiments, Models 2 and 6 cannot be discriminated!* Furthermore, $k_{el}$ can be calculated without knowing $K_m$ and $t_i$. This implies that the outcome is not affected by the transfer velocity into the target compartment. *Moreover, the outcome is not influenced by the remaining unchanged test compound at time $t_i$ at which the effects are measured!*

***Model 7: The Nontoxic Compound Is Modified into a Toxic Compound*** In analogy to Model 3, in the present model the test compound in the outer compartment is modified into a toxic compound by interaction with the culture medium or by cells at a constant rate. The target compartment concentration can be described in analogy to Model 6 by

$$a_i = \sum_{t=0}^{i}(t - t_{-1})S_f\left(\frac{A_0(1-\exp(-k_{met}t))}{A_0(1-\exp(-k_{met}t)) + K_m}\right)$$

or

$$a_i = S_f t_i - S_f \int_{t=0}^{i}\left\{\frac{1}{1 + (A_0/K_m) - (A_0/K_m)e^{-k_{met}t}}\right\}dt$$

resulting in

$$a_i = \left(\frac{S_f}{A_0 + K_m}\right)\left\{A_0 t_i - \left(\frac{K_m}{k_{met}}\right)\ln\left[1 + \left(\frac{A_0}{K_m}\right)(1 - e^{-k_{met}t_i})\right]\right\} \tag{60}$$

If, as assumed, the induced effects are only dependent on $s$, the OCER can be simulated in analogy to Models 5 and 6. Since the effects may be described according to the one effector site model, [S] of Eq. 10 can be replaced by Eq. 60,

or

$$\frac{E_t}{E_u} = \frac{K_c}{K_c + \left(\left(\frac{S_f}{A_0 + K_m}\right)\left\{A_0 t_i - \left(\frac{K_m}{k_{met}}\right)\ln\left\{1 + \left(\frac{A_0}{K_m}\right)(1 - e^{-k_{met} t_i})\right\}\right\}\right)^{N_c}} \tag{61}$$

Since the transfer from the outer compartment into the target compartment (or the enzymatic modification of $S$ to $s^*$; see Fig. 7) is a rate-limiting process, only at low concentrations does the rate of transfer parallel the outer compartment concentration of the toxic product (Fig. 19). At large concentrations, this transfer is constant and is not or only negligibly changed by an increase in the outer compartment concentration. If the treatment period is increased, a disproportionate increase in target compartment concentration is seen if the $t_{1/2}$ value relative to the treatment period is large (slow activation; Fig. 20). Under testing conditions, the shift in OCER of the toxic product is larger than expected if erroneously Model 5 is assumed. Thus a compound dilution affects the $EIC_z$ value more, as expected if we assume Model 5. At small $t_{1/2}$ values, no difference between the present model and Model 5 can be detected in dilution and preincubation experiments under experimental conditions. The equations describing the effects of compound dilution, compound replacement, and compound preincubation are extremely complex. Since a suspicion of compound activation implies retesting with the toxic product, a more detailed description is not needed. However, if under some circumstances such a description is necessary, the equations can easily be derived from Eq. 60.

***Model 8: The Nontoxic Test Compound Is Transformed into a Toxic Product that Afterwards Is Inactivated*** For the sake of completeness, Model 8 is mentioned in

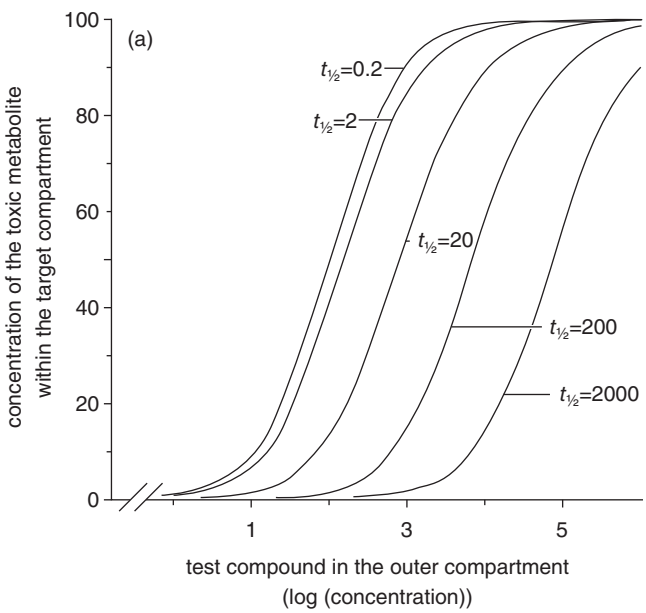


**FIGURE 19** Correlation between the concentration of the toxic compound in the outer compartment and the concentration in the target compartment according to Eq. 60: effects of variations in the elimination half-life ($t_i$ = 10; $S_f$ = 10; $K_m$ = 100 (arbitrary units)).

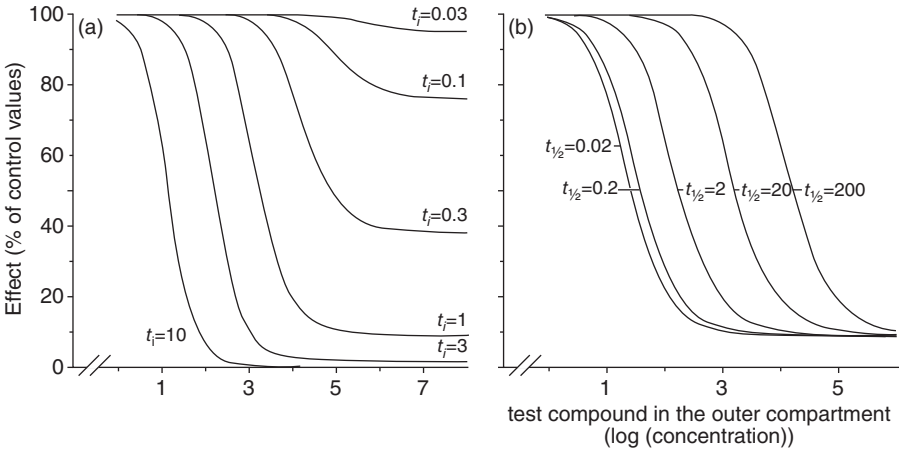


**FIGURE 20** Outer compartment compound concentration–effect relationships (OCERs) according to Eq. 61. (a) Effects of variations in the treatment period $t_i$. (b) Effects of variations in the elimination half-life ($t_{1/2}$). In the present example, $N_c = 1.5$, $K_c = 100$, $t_i = 1$, $S_f = 100$ and $K_m = 100$ (arbitrary units) or as mentioned in the figure.

which, like Model 4, a nontoxic test compound is added to the cultures. Due to an interaction with components of the outer compartment, this test compound is activated into a toxic product, which is inactivated afterwards. The concentration of the toxic product inside the target compartment can be described by the following equation:

$$s_i = S_f \int_{t_0}^{t_i} \left( \frac{S_0 * k_{met} \left( \frac{e^{-k_{el} * t} - e^{-k_{met} * t}}{k_{met} - k_{el}} \right)}{S_0 * k_{met} * \left( \frac{e^{-k_{el} * t} - e^{-k_{met} * t}}{k_{met} - k_{el}} \right) + K_m} \right) dt$$

The target compartment concentration can only be calculated using a numerical approach by employing computer programs such as MatLab or MAPLE.

Variations in treatment period, inclusion of dilution, or inclusion of preincubation steps will result in a mixture of the effects observed in Models 6 and 7. Which of these effects dominate depends on the location of $t_{max}$ (see Eq. 26) in relation to $t_i$ (see also Model 4). Under experimental conditions, only having knowledge of the fate of the test compound in the culture medium ($k_{el}$ and $k_{met}$ are known), an estimate can be made for compound is probably the most toxic component: the test compound itself (behavior like Model 6), an active intermediate of the test compound (behavior like Model 8), or a stable product or complex of the test compound (behavior like Model 7).

## 7 CONCLUDING REMARKS REGARDING THE EIGHT MODELS OF TOXICITY AND THE DISCRIMINATION BETWEEN THEM

The OCERs that are defined by multiple data points, each with very small standard deviations, can be used to distinguish Models 1–8. Mostly, the behavior of the test compound in the medium and the location and number of target proteins are

unknown before experiments are started. Generally, OCERs can roughly be described by Eq. 15–26. The following series of experiments should be performed to distinguish roughly among the eight models:

- Standard experiments (no dilution, no preincubation of the test compound).
- Experiments in which the dilutions of the test compound before addition to the cells are preincubated for a certain period (e.g., 24 hours) under the same conditions as the cell cultures.
- Experiments in which the treatment period is reduced due to nearly complete dilution of the test compound shortly after addition (e.g., after a few hours).

If the results of these experiments are compared as described in Fig. 21, the most probable model may be assessed according to which toxicity is induced.

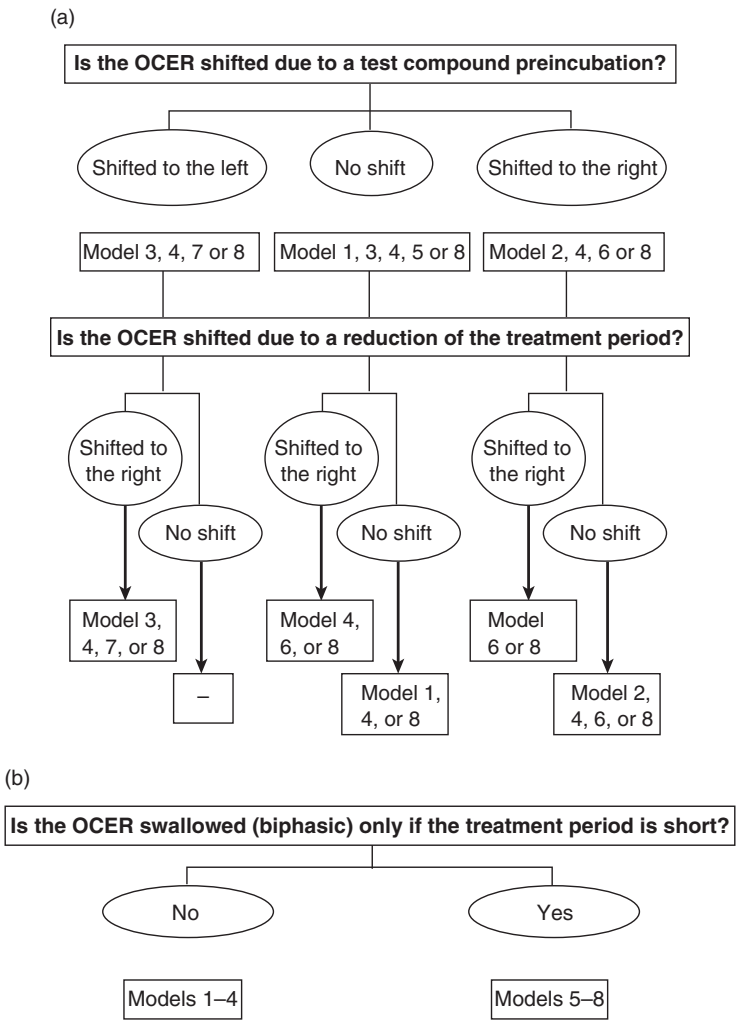


**FIGURE 21** Schematic representation to discriminate Models 1–8. (a) Comparison of the location of the OCERs. (b) Comparsion of the course of the OCERs.

Furthermore, the data may help to decide which further experiments, like measurement of the test compound concentration and/or its metabolite(s) at the end of the treatment period, are needed to further specify the obtained characteristics.

In summary, the present chapter describes tools to elucidate the toxicomechanisms under *in vitro* conditions. For this, equations were developed to describe *in vitro* toxicokinetics and toxicodynamics under various conditions and assumptions.

## ACKNOWLEDGMENT

The present study was supported by the European Commission through the specific targeted research project CellForce (Contract No. NMP4-CT-2005-016626).

## REFERENCES

1. Dost FH. *Grundlagen der Pharmakokinetik*. Stuttgart: Thieme; 1968, p 450.
2. Hill AV. The possible effects of the aggregation of the molecules of haemoglobin on its oxygen dissociation curve. *J Physiol (London)* 1913;40:IV–VII.
3. Bruinink A. Serum-free monolayer cultures of fetal chick brain and retina: immunoassays of developmental markers, mathematical data analysis, and establishment of optimal culture conditions. In Zbinden G, Ed. *The Brain in Bits and Pieces*. Zollikon: MTC Verlag; 1992, pp 23–50.
4. Filipsson AF, Sand S, Nilsson J, Victorin K. The benchmark dose method—review of available models, and recommendations for application in health risk assessment. *Crit Rev Toxicol* 2003;33:505–542.
5. Auton TR. Calculation of benchmark doses from teratology data. *Regul Toxicol Pharmacol* 1994;19:152–167.
6. Links JM, Schwartz BS, Simon D, Bandeen-Roche K, Stewart WF. Characterization of toxicokinetics and toxicodynamics with linear systems theory: application to lead-associated cognitive decline. *Environ Health Perspect* 2001;109:361–368.
7. Nakayama Y, Kishida F, Nakatsuka I, Matsuo M. Simulation of the toxicokinetics of trichloroethylene, methylene chloride, styrene and *n*-hexane by a toxicokinetics/toxicodynamics model using experimental data. *Environ Sci* 2005;12:21–32.
8. Bois FY, Gelman A, Jiang J, Maszle DR, Zeise L, Alexeef G. Population toxicokinetics of tetrachloroethylene. *Arch Toxicol* 1996;70:347–355.
9. Haddad S, Krishnan K. Physiological modeling of toxicokinetic interactions: implications for mixture risk assessment. *Environ Health Perspect* 1998;106(Suppl 6):1377–1384.
10. Bruinink A, Rasonyi T, Sidler C. Reduction of ochratoxin A toxicity by heat-induced epimerization. *In vitro* effects of ochratoxins on embryonic chick meningeal and other cell cultures. *Toxicology* 1997;118:205–210.
11. Bruinink A, Sidler C. The neurotoxic effects of ochratoxin-A are reduced by protein binding but are not affected by l-phenylalanine. *Toxicol Appl Pharmacol* 1997;146: 173–179.
12. Bruinink A, Rasonyi T, Sidler C. Differences in neurotoxic effects of ochratoxin A, ochracin and ochratoxin-alpha *in vitro*. *Natural Toxins* 1998;6:173–177.
13. Bruinink A, Faller P, Sidler C, Bogumil R, Vasak M. Growth inhibitory factor and zinc affect neural cell cultures in a tissue specific manner. *Chemi-Biol Interact* 1998;115: 167–174.

14. Müller JP, Bruinink A. Neurotoxic effects of aluminium on embryonic chick brain cultures. *Acta Neuropathol* 1994;88:359–366.
15. Bruinink A, Birchler F. Effects of cisplatin and ORG.2766 in chick embryonic brain cell cultures. *Arch Toxicol* 1993;67:325–329.
16. Bruinink A, Reiser P, Müller M, Gähwiler BH, Zbinden G. Neurotoxic effects of bismuth *in vitro*. *Toxicol In Vitro* 1992;6:285–293.
17. Bruinink A, Zimmermann G, Riesen F. Neurotoxic effects of chloroquine *in vitro*. *Arch Toxicol* 1991;65:480–484.
18. Bruinink A, Sidler C, Birchler F. Neurotrophic effects of transferrin on embryonic chick brain and neural retina cell cultures. *Int J Dev Neurosci* 1996;14:785–795.
19. Goldoni M, Vettori MV, Alinovi R, Caglieri A, Ceccatelli S, Mutti A. Models of neurotoxicity: extrapolation of benchmark doses in vitro. *Risk Anal* 2003;23:505–514.
20. *Progress Report of the Ecological Committee on FIFRA Risk Assessment Methods: V. Terrestrial Effects Assessment*. US EPA; 2005.
21. Stephens DN, Voet B. Differential effects of anxiolytic and non-anxiolytic benzodiazepine receptor ligands on performance of a differential reinforcement of low rate (DRL) schedule. *Behav Pharmacol* 1994;5:4–14.
22. Michaelis L, Menten ML. Die Kinetik der Invertinwirkung. *Biochem Z* 1913;49: 333–369.

# TOXICOLOGIC PATHOLOGY

Paul B. Tchounwou[1] and José A. Centeno[2]

[1]*Jackson State University, Jackson, Mississippi*
[2]*Armed Forces Institute of Pathology, Washington, DC*

**Contents**

1 Introduction
2 Toxicologic Pathology of the Cell
  2.1 Introduction
  2.2 Cellular Responses to Insult
  2.3 Cell Death
  2.4 Carcinogenesis
3 Chemical-Induced Toxicologic Pathology and Organ-Specific Responses
  3.1 Arsenic-Induced Health Effects
  3.2 Mechanisms of Arsenic-Induced Toxicologic Pathology
  3.3 Arsenic-Induced Keratosis
  3.4 Arsenic-Induced Skin Cancer
  3.5 Arsenic-Induced Liver Cancer
  3.6 Arsenic-Induced Kidney Cancer
  3.7 Arsenic-Induced Bladder Cancer
  3.8 Arsenic-Induced Lung Cancer
  3.9 Arsenic-Induced Gastrointestinal Cancer
  3.10 Arsenic-Induced Brain Cancer
4 Radiation-Induced Toxicologic Pathology
  4.1 Ionizing Radiation and Ultraviolet Radiation
  4.2 Nonionizing Radiation
5 Conclusion
  Acknowledgments
  References

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

# 1 INTRODUCTION

Toxicologic pathology is defined as a scientific discipline that integrates two important areas of scientific endeavors: toxicology and pathology. Toxicology is the study of the deleterious effects of chemical compounds on biologic systems and the evaluation of the probability of their occurrence. Pathology, on the other hand, is the study of the nature and magnitude of diseases and the evaluation of structural and functional changes in cells, tissues, and organs in response to the exposure to physical, chemical, and/or biological agents. Hence, the new discipline of toxicologic pathology requires knowledge of toxicology and pathology, as well as other related disciplines, such as molecular and cellular biology, biochemistry, histology, chemistry, microbiology, bioengineering, anatomy, and homeostasis, to provide a scientific understanding of the biologic significance of the biochemical, functional, cellular, and morphological changes arising from various environmental exposures. Understanding the biologic significance of health-related conditions at the cellular, molecular, and morphological levels is critical to assessing the health risk associated with environmental exposures, for developing safety standards, and for developing cost-effective management strategies. This chapter discusses important aspects of toxicologic pathology at the molecular and cellular levels and uses arsenic and radiation as model toxicants to provide relevant information on chemically induced toxicologic pathology at the cellular and target organ levels.

# 2 TOXICOLOGIC PATHOLOGY OF THE CELL

## 2.1 Introduction

Toxicologic pathology has been defined as the study of the molecular, cellular, tissue, and/or organ responses of the living organism when exposed to injurious agents or deprivations. These responses represent a spectrum of cellular changes ranging from cell death to malignant transformations; tissue and organ responses, including regeneration, inflammation, and organization; and the overall response as identified by clinical changes and alterations in body fluids [1]. The cell constitutes the basic unit of life. Hence, morphologic changes in organs and tissues as a result of injury begin with the responses of underlying cells to the toxic insult. A thorough evaluation and understanding of related pathology must logically begin at the cellular level. Several cellular components whose alterations have been reported to be critically associated with cell injury include the plasma membrane, site of osmotic, electrolyte, and water regulation, as well as signal transduction; the mitochondrion, site of energy storage and aerobic respiration; the endoplasmic reticulum, site of much protein synthesis, as well as calcium storage in some cases; and the nucleus, which contains the genetic material and where transcription of the genetic code takes place [2].

Biochemical changes such as enzyme induction and gene expression occur at the earlier stages of the exposure–disease continuum. The degree of cellular injury depends on the metabolic rate. Cells with high metabolic rates such as neurons, myocardial cells, and renal proximal convoluted tubule epithelial cells (PTCs) may suffer from injury more quickly than low metabolizing ones. These high metabolism cells depend on a continuous flow of oxygen to perform aerobic metabolism neces-

sary to provide required energy in the form of ATP for the maintenance of membrane polarity and membrane integrity (neurons), for the continual muscular contraction/relaxation and calcium transport (myocardium), and for the transport of fluids, electrolytes, and metabolites (renal PTCs). Hence, any depletion in oxygen supply is likely to have a significant impact on their survival. In contrast, cells with low metabolic activity such as fibroblasts and adenocytes are less affected by the low supply of oxygen, and hence their prominent role in regeneration and scarring [2].

### 2.2 Cellular Responses to Insult

Homeostasis is one of the most remarkable and most typical properties of highly complex biologic systems. It defines the ultimate environment under which cells maintain the physio- chemical conditions (intracellular pH, cytosolic osmolarity, ion gradients) necessary to perform their biologic functions. In biologic systems, homeostasis is maintained by means of a multiplicity of dynamic equilibriums rigorously controlled by interdependent regulation mechanisms. Hence, the homeostatic system reacts to changes or disturbances in response to various insults by exerting a series of modifications or adjustments to maintain the internal balances and conditions within tolerable limits.

In recent years, the hypothalamus has been identified as one of the critical sites involved in energy homeostasis. Earlier experimental studies with rodents have pointed out that food intake and body weight are under the control of specific hypothalamic subregions. A recent review on the hormonal regulation of the neuronal circuits of the arcuate nucleus of the hypothalamus also reported that the central nervous system (CNS) and the hypothalamus in particular are key players in the regulation of food intake and energy expenditure by peripheral signals such as hormones [3]. Accumulating evidence indicates that the melanocortin (MC) system plays a key role in the regulation of food intake and energy balance. This evidence includes findings that either spontaneous genetic mutations or targeted gene deletions that impair melanocortin signaling cause disrupted food intake and body-weight control. In addition, expression of the mRNA that encodes the endogenous agonists and antagonists for CNS melanocortin receptors is regulated by changes in energy balance and body-adiposity signals [4]. Regulation of signaling by melanocortin 3 and melanocortin 4 receptors in the CNS is controlled via neuronal cell bodies in the arcuate nucleus of the hypothalamus that synthesize melanocortin receptor agonists such as alpha-melanocyte-stimulating hormone (alpha-MSH) or antagonists such as Agouti-related protein (AgRP). The activity of these two populations of neurons is reciprocally regulated by a number of peripheral and central systems that influence energy balance [5].

Deviations from homeostasis may lead to cell injury and eventually cell death, depending on the degree of alterations. Biologic systems must be able to sense and respond rapidly to changes in their environment in order to maintain homeostasis and survive. The induction of heat shock proteins (Hsps) is a common cellular defense mechanism for promoting survival in response to various stress stimuli. Heat shock factors (HSFs) are transcriptional regulators of Hsps, which function as molecular chaperones in protecting cells against proteotoxic damage. Mammals have three different HSFs, which have been considered functionally distinct; HSF1

is essential for the heat shock response and is also required for developmental processes, whereas HSF2 and HSF3 are important for differentiation and development. Recent evidence, however, suggests a functional interplay between HSF1 and HSF2 in the regulation of Hsp expression under stress conditions [6]. Given the broad cytoprotective properties of the heat shock response, there has been a strong interest in discovering and developing pharmacological agents capable of inducing the heat shock response [7].

Cellular adaptation defines the mechanisms by which cells attempt to regulate their internal environment. Two types of adaptation have been reported: atrophy and hypertrophy. Atrophy is an adaptation in which there is a reduction in cell, organ, or tissue mass. At the cellular level, atrophy is often a response to decreased demand for the specialized functions of a particular cell. At the most basic level, atrophied cells undergo a catabolism, often reflected ultrastructurally by an overt breakdown of cellular organelles, including mitochondria, endoplasmic reticulum, microtubules, and microfilaments [2]. Hypertrophy, on the other hand, is an adaptation in which there is an increase in cell, tissue, or organ mass without cellular proliferation. In general, it is a response to increased metabolic activity/demand for the particular cell. At the ultrastructural and histological levels, this translates into an increase in volume of the cytoplasm and an increase in the number of cell organelles, microfilaments, microtubules, and other specialized structures [2].

In a review of recent experimental insights into the molecular mechanisms underlying organellar protein quality control processes, Leidhold and Voos [8] reported that organellar protein homeostasis is maintained by a combination of specific protein biogenesis processes and protein quality control mechanisms that together guarantee the functional state of the organelle. Organelles like mitochondria, chloroplasts, or the endoplasmic reticulum (ER) are essential subcompartments of eukaryotic cells that fulfill important metabolic tasks. Based on their endosymbiotic origin, mitochondria and chloroplasts contain internal protein quality control systems that consist of a cooperative network of molecular chaperones and proteases. In contrast, the ER employs the main cytosolic degradation machinery, the proteasome, for the removal of damaged or misfolded proteins [8]. In all eukaryotic cells, the ER is the site where folding and assembly occurs for proteins destined to the extracellular space, plasma membrane, and the exo/endocytic compartments. The ER is exquisitely sensitive to alterations in homeostasis and provides stringent quality control systems to ensure that only correctly folded proteins transit to the Golgi and unfolded or misfolded proteins are retained and ultimately degraded. A number of biochemical and physiologic stimuli, such as perturbation in calcium homeostasis or redox status, elevated secretory protein synthesis, expression of misfolded proteins, sugar/glucose deprivation, altered glycosylation, and overloading of cholesterol, can disrupt ER homeostasis, impose stress on the ER, and subsequently lead to accumulation of unfolded or misfolded proteins in the ER lumen [9].

In a study of metabolic homeostasis and tissue renewal, Cheung et al. [10] reported that Golgi beta1,6*N*-acetylglucosaminyltransferase V (Mgat5) produces beta1,6GlcNAc-branched N-glycans on glycoproteins, which increases their affinity for galectins and opposes loss from the cell surface to constitutive endocytosis. Oncogenic transformation increases Mgat5 expression, beta1,6GlcNAc-branched N-glycans on EGF, and TGF-beta receptors, and enhances sensitivities to ligands,

cell motility, and tumor metastasis [10]. From another recent study, Wellen et al. [11] pointed out that metabolic and inflammatory pathways crosstalk at many levels, and, while required for homeostasis, interaction between these pathways can also lead to metabolic dysregulation under conditions of chronic stress. This study also demonstrated that the six-transmembrane protein STAMP2 is a critical modulator of this integrated response system of inflammation and metabolism in adipocytes. Lack of STAMP2 in adipocytes resulted in aberrant inflammatory responses to both nutrients and acute inflammatory stimuli; indicating that STAMP2 plays a key role in systemic metabolic homeostasis [11].

There is accumulating evidence that reactive oxygen species (ROSs) are not only toxic but play an important role in cellular signaling and in the regulation of gene expression. Biochemical and physiologic stimuli, such as perturbation in redox status, expression of misfolded proteins, altered glyc(osyl)ation and glucose deprivation, overloading of products of polyunsaturated fatty acid peroxidation (hydroxynonenals, HNE), or cholesterol oxidation and decomposition, can disrupt redox homeostasis, have been found to impose stress, and subsequently lead to accumulation of unfolded or misfolded proteins in brain cells. Alzheimer disease (AD), Parkinson disease (PD), Huntington disease (HD), amyotrophic lateral sclerosis (ALS), and Friedreich ataxia (FRDA) are major neurological disorders associated with production of abnormal proteins and, as such, belong to the "protein conformational diseases." The brain response to detect and control metabolic or oxidative stress is accomplished by a complex network of "longevity assurance processes" integrated to the expression of genes termed vitagenes [12]. Nitric oxide and other reactive nitrogen species appear to play crucial roles in the brain such as neuromodulation, neurotransmission, and synaptic plasticity but are also involved in pathological processes such as neurodegeneration and neuroinflammation. Acute and chronic inflammation result in increased nitrogen monoxide formation and nitrosative stress. It has been well documented that NO and its toxic metabolite, peroxynitrite, can inhibit components of the mitochondrial respiratory chain leading to cellular energy deficiency and, eventually, to cell death [13]. Gestational diabetes mellitus (GDM) and type 2 diabetes mellitus (DM2) have been suggested to be caused by the same metabolic disorder, and defects in gut hormone-dependent regulation of beta-cell function (enteroinsular axis) have been proposed to contribute to the pathogenesis of DM2 [14].

### 2.3 Cell Death

Severe disruption in homeostasis can lead to an irreversible injury to the cell, resulting in cell death in which the normal cell function can no longer be restored. Two distinct types of cell death have been reported: oncotic or accidental necrosis, and apoptotic/programmed cell death.

***Oncotic/Accidental Necrosis*** Necrotic cell death is characterized by several ultrastructural changes including cell swelling with rarefaction of the cytosol due to the influx of water, loss of plasma membrane integrity, dilation of the cisternae of the endoplasmic reticulum, loss or deformation of other specialized surface features, and subsequent loss of cellular integrity resulting in recruitment of cells that cause inflammation and phagocytes and cellular fragments. Hence, necrotic death has

been described as a toxic process in which the cell is not a targeted cell but a passive victim that does not sustain the energy level for an orderly process of cell destruction and assassination [2, 15]. Changes in plasma membrane characterized by the loss of surface specialization including the disappearance of microvilli and swelling of cilia are among the first necrotic changes. Other changes include intercellular attachment breakdown leading to neighboring cell detachment because of separation of gap junctions, dissolution of the terminal web, and degradation of maculae densae and zonulae adherentes. Cytoplasmic blebs or outpocketing may form on the surface of injured cells. Mitochondria are the most dramatically affected organelles from the depletion of ATP, leading to necrotic cell death [2]. Morphological changes in the nucleus are associated with changes in the structure of chromatin. Gross changes associated with oncotic necrosis are variable and very much dependent on the tissue in which the injury has occurred and the etiology of the injury in the response of surrounding tissues [2].

***Apoptotic Cell Death*** Apoptosis plays a critical role in maintaining tissue homeostasis and represents a normal function to eliminate excess or dysfunctional cells. Cell death typically follows one of two patterns: oncotic necrosis and apoptosis. Necrosis is typically the consequence of acute metabolic perturbation with ATP depletion, as occurs in ischemia/reperfusion and acute drug-induced hepatotoxicity. Apoptosis, in contrast, represents the execution of an ATP-dependent death program often initiated by death ligand/death receptor interactions, such as Fas ligand with Fas, which leads to a caspase activation cascade. A common event leading to both apoptosis and necrosis is mitochondrial permeabilization and dysfunction, although the mechanistic basis of mitochondrial injury may vary in different settings [16].

For many years cell death was considered to be an uncontrolled, degenerative, and catastrophic failure of homeostasis in response to cellular injury, and thus of little scientific interest. However, scientific discovery over the past two and a half decades has pointed out that cells within multicellular organisms have the ability to undergo a programmed cell death termed "apoptosis." From a toxicological point of view, it is believed that cells can be killed by exposure to various toxicants and hazardous conditions, and especially if the chemical dose is high enough to induce lethality through oncotic necrosis or cell "murder." However, it is now recognized that it is more common for chemically exposed cells to be destroyed by apoptosis or cell "suicide" [17]. Apoptosis is now believed to be a pathophysiological cell death that is much more common than necrosis and is a desirable mechanism by which biologic systems dispose or get rid of cells that have been damaged by mutagenic chemicals or irradiation without perturbing the homeostatic balance of its environment.

Apoptosis has been reported to play a role in the pathogenesis of several health conditions. This process of programmed cell death has been implicated in diseases such as cancer and AIDS [18], neurologic disorders including ischemic brain injury, Alzheimer disease, amyotrophic lateral sclerosis or Lou Gehrig disease, Huntington disease, and Parkinson disease [19], renal diseases including congenital dysplastic kidneys, diabetic nephropathy, glomerulonephritis, glomerulous sclerosis hemolytic uremic syndrome, HIV nephropathy, obstructive uropathy and hydronephrosis polycystic kidney disease, and renal cancer, reproduction abnormalities including

the disruption of spermatogenesis and male infertility by toxicants and hormone exposure [20] and the induction of ovarian toxicity in females [21].

Apoptotic cell death has been described as occurring in three phases: first, a receptor-mediated commitment to the cell death, which is termed induction phase, followed by an effector phase characterized by distinct morphological changes in cell structure with modification of mitochondrial functions including change in the status of ATP and other intracellular constituents, and finally a degradation phase with extensive break up of the targeted cell. Figure 1 presents the morphological features of necrosis and apoptosis [22].

Utrastructurally, apoptotic cell death has unique features that are quite distinct from necrotic cell death (Table 1). Its execution phase has striking morphological characteristics including the loss of cell contact with neighboring viable cells, chromatin condensation to form dense compact masses, and breakdown of the cytoskeletal scaffolding. The cell surface membrane, which is already contracted, begins to bleb. This is followed by breakdown of the cell into membrane-bound fragments or "apoptotic bodies" [17]. In contrast to necrosis, the plasma membrane and cytoplas-

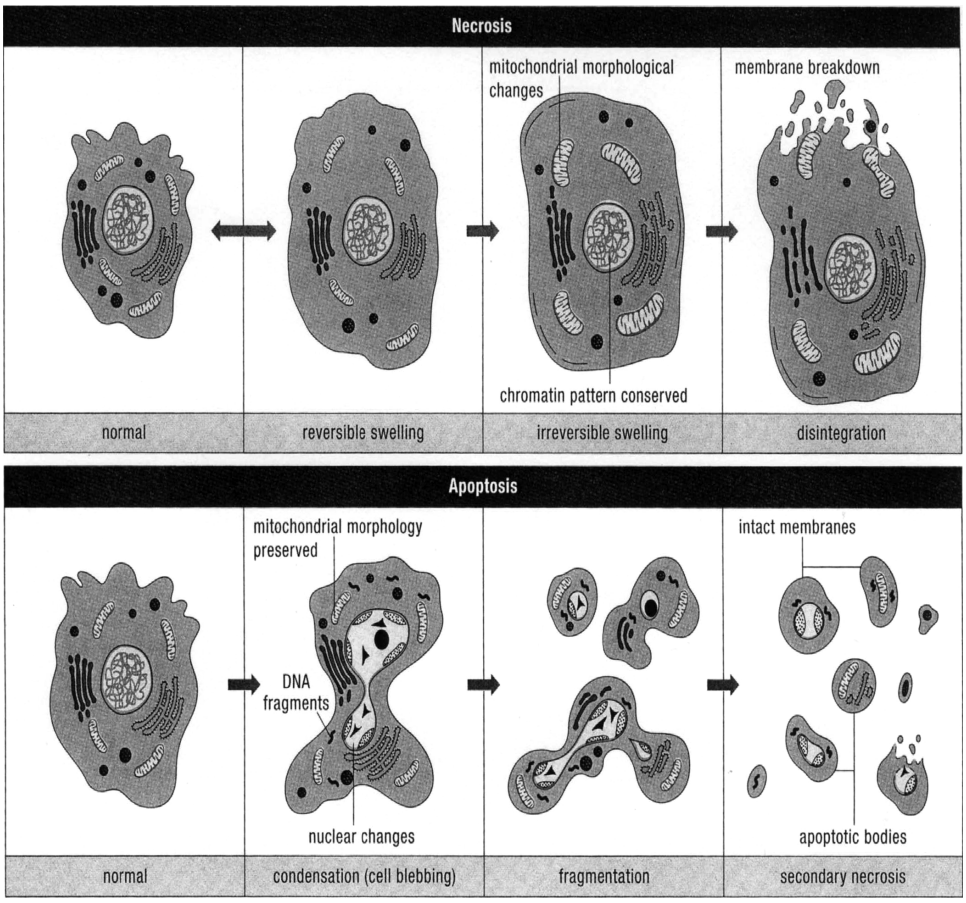


**FIGURE 1** Morphological features of necrosis and apoptosis [22].

**TABLE 1 Differential Characteristics of Apoptotic and Necrotic Cell Death**[a]

| Characteristic | Apoptosis | Necrosis |
|---|---|---|
| Histological changes | Individual cells scattered throughout the affected tissue<br>Heperbasophilia or hyperiosinophilia<br>Formation of round bodies, often within a "halo"<br>Chromatin condensation into "caps" or "crescents," with round nuclear bodies; preservation of nuclear envelope | Whole fields of cells affected<br>Hyperiosinophilia<br>Loss of cell border with irregular fragmentation<br>Irregular chromatin clumping, pyknosis, karyorhesis, and/or karyolysis; rupture of nuclear envelope |
| Ultrastructural changes | Condensation followed by rapid "zeoisis" (budding)<br>Condensation of cytoplasm followed by rarefaction after ingestion by phagocytes<br>Preservation of organellar integrity<br>Preservation of mitochondrial ultrastructure<br>Preservation of internal and external membranes with preservation of membranes around apoptotic bodies<br>Migration of uniformly degraded chromatin to margin of nuclear envelope; preservation of nuclear envelope | Swelling and loss of surface structures with "blebbing" and loss of apical portions of cytoplasm<br>Rarefaction of cytoplasm followed by condensation after death<br>Swelling and loss of organellar integrity<br>Low amplitude swelling of mitochondria followed by high amplitude swelling and rupture<br>Rupture and degradation of internal and external membranes with bursting of the cell<br>Irregular clumping and degradation of chromatin; rupture of nuclear envelope |
| Morphological changes | Membrane blebbing, but no loss of integrity; aggregation of chromatin at nuclear membrane<br>Begins with shrinking of cytoplasm and condensation of nucleus<br>Ends with fragmentation of cell into smaller bodies<br>Formation of membrane-bound apoptotic bodies; mitochondria become leaky | Loss of membrane integrity<br>Begins with swelling of cytoplasm and mitochondria<br>Ends with total cell lysis<br>Disintegration/swelling of organelles |
| Gross changes | Minimal or atrophy without scarring | Grossly evident with disruption of normal tissue structure and detail, scarring if long term |

**TABLE 1** ***Continued***

| Characteristic | Apoptosis | Necrosis |
|---|---|---|
| Biochemical features | Tightly regulated process involving activation and enzymatic steps<br>Energy (ATP) dependent; active process<br>Nonrandom mono- and oligonucleosomal length fragmentation of DNA (ladder pattern after agarose gel electrophoresis)<br>Prelytic DNA fragmentation<br>Release of cytochrome-*c*, AIF from the mitochondria into the cytoplasm<br>Activation of caspase cascade<br>Alteration in membrane asymmetry; phosphatidylserine translocation | Loss of regulation of ion homeostasis<br>No energy requirement; passive process<br>Random digestion of DNA (smear of DNA after agarose gel electrophoresis)<br>Postlytic DNA fragmentation (late event of death) |
| Sequelae | Retention of intracellular enzymes within the apoptotic bodies<br>No release of proinflammatory products<br>Ingestion by adjacent cells or by tissue macrophages<br>Atrophy with stromal collapse but no scarring | Release of intracellular enzymes into extracellular milieu<br>Release of proinflammatory cell breakdown products<br>Ingress of neutrophils followed by macrophages<br>Active inflammation with scarring |

[a]See Refs. 2 and 22.

mic organelles of the apoptotic cells remain intact. The biochemical effect most commonly associated with apoptosis is the cleavage of chromosomal DNA, resulting in nucleosomal fragments recognizable as a DNA ladder on agarose gel electrophoresis. This DNA fragmentation induced preferentially in the linker region between nucleosomes is generally believed to be catalyzed by an apoptotic specific $Ca^{2+}$/$Mg^{2+}$-dependent endonuclease [23]. An increase in intracellular $Ca^{2+}$ has been reported in early apoptosis to occur in many cell types. This process may activate a nucleolus apoptotic specific endonuclease that cleaves genomic DNA to nucleosomal oligomers [23]. It has been reported that the apoptotic process requires energy from ATP and sometimes depends on new gene expression for completion.

The molecular mechanisms of apoptosis involve a signal transduction in plasma membranes, during which second messengers including CAMP, inositol-1,4,5 triphosphate ($IP_3$), diacyl glycerol, and CGMP are produced through the signal transduction systems, such as adenylase cyclase, phospholipase C, guanylate cyclase, and ion channel systems, which are essential in living systems. Although it is unclear what happens downstream of the apoptosis-specific Fas/APO-1 (CD 95) and tumor necrosis factor (TNF), the crosstalk of the second messengers produced results in

the expression of apoptosis-regulator genes, the activation of preexisting apoptotic gene products, and/or suppression of the expression of survival genes [23].

Several oncogenes such as c-myc, p53, bcl2, c-fos, c-gun, and Rb, have been shown to modulate apoptosis. Changes in these gene products can be responsible for initiation of apoptosis that is dependent on induction of the *de novo* synthesis of either RNA or protein or on suppression. The tumor suppressor oncogene p53 has been shown to have dramatic effect on apoptosis. Several studies have shown that DNA damaging agents such as alkylating agents and UV radiation require p53 for apoptosis induction. Mutations of p53 have been observed in many human cancers [23]. On the other hand, overexpression of bcl2 has been linked to oncogenic activation and the prevention of apoptosis induced by irradiation, glucocorticoid or c-myc. The c-myc protooncogene, known to be important for cell proliferation control, plays a part in the regulation of apoptosis. Its ability to regulate apoptosis results from the coordinated activation of c-myc and several protein partners (such as Max) that facilitate DNA binding and activate transcription. C-myc can regulate apoptosis induced by a variety of insults such as hypoxia, glucose deprivation, cancer chemotherapeutics, and DNA damage [24].

Programmed cell death is essential for the development and maintenance of cellular homeostasis of the immune system. B-cell lymphoma-2 (Bcl2) family members have been demonstrated to play a crucial role in the regulation of apoptosis as mediators between the apical stimuli sensing steps and the executory mechanisms of apoptosis. Deregulation of their role may subvert the homeostasis of a given tissue and collaborate in the genesis of a myriad of diseases characterized by exacerbated or insufficient apoptosis, including neurodegenerative diseases and cancer [25]. The Bcl2 family of proteins comprises both proapoptotic and antiapoptotic members. They are characterized by the presence of conserved sequence motifs, known as Bcl2 homology (BH) domains. Antiapoptotic members share all four BH domains, designated as BH1–4; the multidomain proapoptotic members contain BH1–3 domains, whereas another subgroup of proapoptotic members only have a BH3 domain. The BH3-only proteins act as sensors for distinct apoptosis pathways, whereas multidomain proapoptotic Bax and Bak are executioners of death orders relayed by the BH3-only proteins. Antiapoptotic Bcl2 family members appear to function, at least in part, by interacting with and antagonizing proapoptotic family members [26]. The functional balance of proapoptotic versus antiapoptotic influences, which operates at organelles, determines whether cells will live or undergo apoptosis [27]. BH3-only molecules, often working in concert, compete for downstream multidomain pro- and antiapoptotic Bcl2 members to control serial stages of lymphocyte development and homeostasis [28]. In a study of the modulation of proapoptotic (Bax) and antiapoptotic (Bcl2) gene expression in isolated porcine hepatocytes perfused within a radial-flow bioreactor after low-temperature storing, Mischiati et al. [29] demonstrated that the loss in Bax expression was paralleled with the increase in Bcl2 expression, indicating that the proapoptotic activity seems to balance the antiapoptotic activity. Accumulated evidence suggests that apoptosis helps to maintain cellular homeostasis during the menstrual cycle by eliminating senescent cells from the functional layer of the uterine endometrium during the late secretory and menstrual phase of the cycle. The Bcl2 family and the Fas/FasL system have been extensively studied in human endometrium and endometriotic tissues [30].

The family of nuclear factor kB (NF-kB) nuclear transcription factors also regulates cell survival by inhibiting or promoting apoptosis in several cell types. The antiapoptotic role of NF-kB has been demonstrated in several models of apoptosis including Ras-induced apoptosis of NIH 3T3 cells, nitric-oxide induced apoptosis, and anti-IgM-induced apoptosis of lymphoma cells [24].

Cell cycle regulation is essential for proper cellular homeostasis. In hematopoietic cells, cycling dependent kinase-2 (Cdk2), a major regulator of S phase entry, is activated by mitogenic cytokines and has been suggested to be involved in antigen-induced apoptosis of T lymphocytes. However, a recent report indicates that Cdk2 is not required for proliferation and differentiation of hematopoietic cells *in vivo*, although *in vitro* analyses consider Cdk2 a main player for the proliferation and apoptosis in these cells, and a potential target for therapy [31].

## 2.4 Carcinogenesis

It is estimated that cancer causes at least one million deaths worldwide each year, and exposure to carcinogenic compounds constitutes a major risk factor. Cancer is not a single disease but a large group of diseases, all of which can be characterized by the uncontrolled growth of an abnormal cell to produce a population of cells that have acquired the ability to multiply and invade surrounding and distant tissues [32].

***Stages of Carcinogenesis*** Carcinogenesis is a complex multistep process that involves epigenetic events such as the inappropriate expression of certain cellular genes, and genetic events that include the mutational activation of oncogenes and the inactivation of tumor suppressor genes. Multistep models of carcinogenesis have been useful for defining events in the neoplastic process and form the cornerstone of current hypotheses of the biological mechanisms of carcinogenesis [33]. This process has been divided into at least three stages known as initiation, promotion, and progression. Table 2 presents its morphological and biological features [34].

During the initiation phase, a normal cell undergoes an irreversible change characterized by an intrinsic capacity for autonomous growth. However, this capacity remains latent for a period of time during which the initiated cell is morphologically similar to the normal cell. Initiation involves alteration of cellular DNA, and this effect may be triggered by exposure to a known carcinogen or to an ultimate carcinogen generated by metabolic activation of a procarcinogen. Initiators interact with host cellular macromolecules and nucleic acids in a specific pattern typically involving the generation of reactive oxygen species (electrophiles) esters or free radicals that bind covalently to nucleophilic sites in critical cellular molecules [13]. Initiation is hence a genotoxic or DNA-damaging event, one in which an alteration in the DNA sequence is produced.

In the promotion stage of carcinogenesis, specific agents known as promoters enhance the development of neoplasm in initiated cells. Promoters do not cause cancer by themselves. The temporal sequence of exposure is important in characterizing promotion, a process that is considered as a nongenotoxic or epigenetic event. If the agent potentiates carcinogenesis from coexposure with an initiation, it is called a cocarcinogen rather than a promoter. Tumor promotion has been found to be modulated by various factors including age, gender, and hormonal status [32, 33].

**TABLE 2 Morphological and Biological Features of the Stages of Carcinogenesis**[a]

| Initiation | Promotion | Progression |
|---|---|---|
| Irreversible | Operationally reversible, both at the level of gene expression and at the cellular level | Irreversible |
| Initiated stem cell not morphologically identifiable | Promoted cell population existence dependent on continued administration of promoting agent | Morphologically discernible alteration in cellular genomic structure resulting from karyotypic instability |
| Efficiency is sensitive to xenobiotic and other chemical factors | Efficiency is sensitive to aging and dietary and hormonal factors | Growth of altered cells sensitive to environmental factors during early phase |
| Spontaneous (endogenous) occurrence of initiated cells | Endogenous promoting agents may effect spontaneous promotion | Benign or malignant neoplasms observed at this stage |
| Requirement for cell division for fixation | Dose–response exhibits measurable threshold and maximum effect | Progressor agents advance promoted cells into this stage |
| Dose–response not exhibiting a readily measurable threshold | Relative potency of promoters measured by their effectiveness in causing an expansion of the initiated cell population | |
| Relative potency of initiators dependent on quantitation of preneoplastic lesions after defined period of promotion | | |

[a]See Ref. 14.

The third stage of carcinogenesis progression involves additional genetic damage and enhances the development of malignant tumors from benign tumors [32, 33]. Associated with progression is the development of an increased degree of karyotypic instability and of aneuploidy leading to chromosomal rearrangements, especially in leukemia. Table 3 presents a comparison of specific features of benign and malignant neoplasms [34].

***Regulation of Carcinogenesis*** Both oncogenes and tumor-suppressor genes have been involved in the regulation of carcinogenesis. Oncogenes are dominant-acting structural genes that encode for protein products capable of transforming the phenotype of a cell. They are known to encode for growth factors, growth factor receptors, regulatory proteins in signal transduction, nuclear regulatory proteins, and protein kinases. The activation of these gene products including c-onc, c-myc, c-abl, N-myc, K-ras, C-H-ras, C-U-ras, C-K-ras, myc, myb, erb B, mos, Ink1, Ink2, and Jun contributes to the neoplastic process. Tumor suppressor genes, on the other hand, are regulatory genes that function to suppress or limit growth by inhibiting the activity of structural genes that are responsible for cellular growth. While proto-oncogenes have to be activated to influence carcinogenesis, tumor suppressor genes have to be inactivated for the transformed phenotype to be expressed. Hence, several human cancers have been associated with mutations in various tumor suppressor genes including p53, RB1, APC, WT-1, NF-1, p16INK4, BRCA1, BRCA2, TSC2, MSH2, MLH1, VHL, and PTC [32].

**TABLE 3 Comparative Features of Benign and Malignant Neoplasms**[a]

| Characteristic | Benign Neoplasms | Malignant Neoplasms |
|---|---|---|
| General effect on the host | Little, usually do not cause death | Will always kill the host if untreated |
| Rate of growth | Slow; may stop or regress | More rapid, but slower than repair tissue; autonomous; never stop or regress |
| Histological features | Encapsulated; remain localized at primary site | Infiltrate or invade; metastasize |
| Mode of growth | Usually grow by extension, displacing surrounding normal tissue | Invade, destroy, and replace surrounding normal tissue |
| Metastasis | Do not metastasize | Most can metastasize |
| Architecture | Encapsulated; have complex stroma and adequate blood supply | Not encapsulated; usually have poorly developed stroma; may become necrotic at center |
| Danger to host | Most are without lethal signature | Always ultimately lethal unless removed or destroyed *in situ* |
| Injury to host | Usually negligible but may become very large and compress or obstruct vital tissue | Can kill host directly by destruction of vital tissue |
| Radiation sensitivity | Sensitivity near that of half of normal parent cell; rarely treated with radiation | Sensitivity increases in rough proportion to malignancy; often treated with radiation |
| Behavior in tissue | Cells are cohesive and inhibited by mutual contact | Cells do not cohere; frequently not inhibited by mutual contact |
| Resemblance to tissue | Cells and architecture resemble tissue of origin | Cells are atypical and pleomorphic; disorganized bizarre architecture |
| Mitotic figures | Mitotic figures are rare and normal | Mitotic figures may be numerous and abnormal in polarity and configuration |
| Shape of nucleus | Normal and regular; show usual stain affinity | Irregular; nucleus frequently hyperchromatic |
| Size of nucleus | Normal; ratio of nucleus to cytoplasm near normal | Frequently large; nucleus to cytoplasm ratio increased |
| Nucleolus | Not conspicuous | Hyperchromatic and larger than normal |

[a]See Ref. 34.

The ubiquitin-proteasome system (UPS) has been reported to play a significant role in cell fate and carcinogenesis. It is comprised of a multiunit cellular protease system that controls several of the proteins that regulate the cell cycle, apoptosis, angiogenesis, adhesion, and cell signaling. Inhibition of this UPS has been found to be a prerequisite for apoptosis and is already clinically exploited with the proteasome inhibitor bortezomib in multiple myeloma [35].

The cyclooxygenase (COX) enzymatic system has also been implicated in carcinogenesis and apoptosis. It includes two isoenzymes, COX1 and COX2, that convert arachidonic acid to prostaglandins. COX1 is constitutively expressed and considered to be a housekeeping gene, while COX2 is not usually detectable in normal tissues

but can readily be induced in processes like inflammation, reproduction, and carcinogenesis [36]. The mechanisms by which COX2 is thought to be involved in carcinogenesis include resisting apoptosis, increasing cell proliferation, stimulating angiogenesis, and modulating the invasive properties of cancer cells [36]. Buskens et al. [37] reviewed the mechanisms by which COX2 contributes to carcinogenesis, its role in prognosis, and the possible place of selective COX2 inhibitors in the prevention and treatment of gastrointestinal malignancies, focusing particularly on esophageal cancer. Inhibition of COX2 by aspirin and other nonsteroidal anti-inflammatory drugs (NSAIDs) has been found to inhibit proliferation of colorectal cancer cells and in epidemiologic studies has been shown to reduce colon polyp formation in genetically predisposed populations and in the general population [35]. Epidemiological studies have also reported that long-term use of NSAIDs is associated with a reduced incidence of breast, colorectal, gastric, and esophageal cancers, while experimental and clinical studies have demonstrated that treatment with NSAIDs causes a statistically significant reduction in both the number and the size of polyps in familial adenomatous polyposis patients [36, 37].

Disruptions of the control mechanisms of the cell cycle can initiate carcinogenesis and play a role in progression to cancer. A recent study by Horree et al. [38] has shown that the increase in cell proliferation during endometrial carcinogenesis paralleled the progressive derailment of cyclin B1, cyclin D1, cyclin E, p16, p21, p27, p53, and cdk2, indicating the importance of these cell cycle regulators in endometrial carcinogenesis.

Inactivation of the p53 tumor suppressor gene through a point mutation and/or loss of heterozygosity is one of the most common genetic changes found in various types of human tumors. Mitsudomi et al. [39] investigated the relationship between the presence of p53 gene mutations and survival of patients with non-small-cell lung cancer (NSCLC) of all stages who underwent surgery with preoperative curative intent as a routine therapeutic intervention. The prognostic significance of parameters like sex, age, tumor histology, and the stage of the disease was also evaluated. Overall, the p53 mutations did not correlate with sex, age, or the clinical stage of the disease but showed frequent association with tumors of squamous cell histology. It was concluded that the occurrence of p53 mutations in some NSCLC tumors may be independently associated with a shortened overall survival and may be of somewhat more prognostic significance in patients with advanced stage than in those with early stage of the disease [39].

## 3 CHEMICAL–INDUCED TOXICOLOGIC PATHOLOGY AND ORGAN–SPECIFIC RESPONSES

Understanding the mechanisms of chemical-induced toxicologic pathology is pivotal to evaluating the health-related risks of drugs and other chemical compounds. A high degree of toxicologic pathology has been associated with exposure to many types of chemical substances including heavy metals, pesticides, pharmaceuticals, and various industrial compounds. The toxicologic pathology of these compounds, and especially their impact on human health, has received a great deal of attention. This attention is mostly due to the availability of new data indicating the need to reevaluate the health effects, and to revise the existing criteria and standards. The

following discussion focuses on arsenic as an example to illustrate the toxicologic pathology associated with chemical exposure on various organ systems.

### 3.1 Arsenic-Induced Health Effects

A review of biological properties and toxic effects of arsenic indicates that this chemical is a systemic toxicant capable of causing a significant number of health effects, including cardiovascular disease, peripheral vascular disease, developmental effects, neurologic and neurobehavioral effects, diabetes, hearing loss, portal fibrosis of the liver, lung fibrosis, hematologic effects (anemia, leukopenia, and eosinophilia), and carcinogenic effects [40–42]. Table 4 presents a summary of health effects associated with arsenic exposure [43].

Research has also pointed to significantly higher standardized mortality rates for cancers of the bladder, kidney, skin, and liver in many areas of arsenic pollution. The severity of adverse health effects is related to the chemical form of arsenic and is also time and dose dependent [44, 45]. Although the evidence of carcinogenicity of arsenic in humans seems strong, the mechanism by which it produces tumors in humans is not completely understood [43].

### 3.2 Mechanisms of Arsenic-Induced Toxicologic Pathology

One of the major mechanisms by which arsenic exerts its toxic effect is through impairment of cellular respiration by the inhibition of various mitochondrial enzymes, and the uncoupling of oxidative phosphorylation. Most toxicity of arsenic results from its ability to interact with sulfhydryl groups of proteins and enzymes, and to substitute phosphorus in a variety of biochemical reactions [46]. Arsenic *in vitro* reacts with protein sulfhydryl groups to inactivate enzymes, such as dihydrolipoyl dehydrogenase and thiolase, thereby producing inhibited oxidation of pyruvate and betaoxidation of fatty acids [47]. Although the evidence of carcinogenicity of arsenic in humans seems strong, the mechanism by which it produces tumors in humans is not completely understood. In contrast to most other human carcinogens, it has been difficult to confirm the carcinogenicity of arsenic in experimental animals. Research evaluating the chronic effects in laboratory animals exposed to inorganic and organic arsenic compounds by the oral route or skin contact has not shown any potential for initiation or promotion of carcinogenicity [48].

In the absence of animal models, *in vitro* studies become particularly important in providing information on the carcinogenic mechanisms of arsenic toxicity. Arsenic and arsenical compounds are toxic and induce morphological changes in cultured cells [49]. Studies in our laboratory have indicated that arsenic is cytotoxic and able to transcriptionally activate a significant number of stress genes in transformed human liver cells [50, 51]. Recently, we have also demonstrated that the toxicity of arsenic depends on its chemical form, the inorganic form being more toxic than the organic one [45]. Arsenic and arsenic containing compounds have also been shown to be potent clastogens both *in vivo* and *in vitro*. They have been reported to induce sister chromatid exchanges and chromosome aberrations in both human and rodent cells in culture [52, 53] and in cells of exposed humans [54]. Arsenical compounds have also been shown to induce gene amplification, arrest cells in mitosis, inhibit DNA repair, and induce expression of the *c-fos* gene and the oxidative stress protein

**TABLE 4 Clinical and Pathological Manifestations of Acute and Chronic Arsenic Poisoning**[a]

| Organ/Tissue Level | Acute Effects | Chronic Effects |
|---|---|---|
| Dermatologic | • Capillary flush<br>• Contact dermatitis<br>• Folliculitis<br>• Hair loss | • Melanosis<br>• Bowen disease<br>• Facial edema<br>• Palmoplantar hyperkeratosis<br>• Cutaneous malignancies<br>• Hyperpigmentation<br>• Desquamation |
| Neurologic | • Hyperpyrexia<br>• Convulsions<br>• Tremor/coma<br>• Disorientation | • Encephalopathy<br>• Headache<br>• Peripheral neuropathy<br>• Axonal degeneration |
| Gastrointestinal/ hepatic | • Abdominal pain<br>• Dysphagia<br>• Vomiting<br>• Blood/rice water diarrhea<br>• Garlicky odor to breath<br>• Mucosal erosions<br>• Fatty liver<br>• Cholangitis<br>• Cholecystitis | • Nausea<br>• Vomiting<br>• Diarrhea<br>• Anorexia<br>• Weight loss<br>• Hepatomegaly<br>• Jaundice<br>• Pancreatitis<br>• Cirrhosis<br>• Liver cancer |
| Renal | • Tubular damage<br>• Glomerular damage<br>• Oliguria<br>• Uremia | • Nephritis<br>• Proteinuria |
| Hematologic | • Anemia<br>• Thrombocytopenia | • Bone marrow hypoplasia<br>• Anemia<br>• Thrombocytopenia<br>• Basophilic stippling<br>• Karyorrhexis |
| Cardiovascular | • Ventricular fibrillation<br>• Tachycardia | • Arrhythmias<br>• Pericarditis<br>• Acrocyanosis<br>• Raynaud's gangrene |
| Respiratory | • Pulmonary edema<br>• Bronchial pneumonia<br>• Tracheobronchitis | • Cough<br>• Pulmonary fibrosis<br>• Lung cancer |

[a]See Ref. 44.

heme oxygenase in mammalian cells [55], and have been implicated as promoters and comutagens for a variety of agents [56].

Various hypotheses have been proposed to explain the carcinogenicity of inorganic arsenic [57]. Nevertheless, molecular mechanisms by which this arsenical induces cancer are still poorly understood. Results of previous studies have indicated that inorganic arsenic does not act through classic genotoxic and mutagenic mechanisms, but rather may be a tumor promoter that modifies signal transduction

pathways involved in cell growth and proliferation [58]. Inorganic arsenic has been shown to modulate expression and/or DNA-binding activities of several key transcription factors, including nuclear factor kappa B, tumor suppressor protein (p53), and activating protein-1 (AP-1) [59, 60]. Mechanisms of AP-1 activation by trivalent arsenic include stimulation of the mitogen-activated protein kinase (MAPK) cascade with a consequent increase in the expression and/or phosphorylation of the two major AP-1 constituents, *c-jun* and *c-fos* [60].

Although much progress has recently been made in the area of arsenic's possible mode(s) of carcinogenic action, a scientific consensus has not yet been reached. A recent review discusses nine different possible modes of action of arsenic carcinogenesis: induced chromosomal abnormalities, oxidative stress, altered DNA repair, altered DNA methylation patterns, altered growth factors, enhanced cell proliferation, promotion/progression, suppression of p53, and gene amplification [57]. Presently, three modes (chromosomal abnormality, oxidative stress, and altered growth factors) of arsenic carcinogenesis have shown a degree of positive evidence, both in experimental systems (animal and human cells) and in human tissues. The remaining possible modes of carcinogenic action (progression of carcinogenesis, altered DNA repair, p53 suppression, altered DNA methylation patterns, and gene amplification) do not have as much evidence, particularly from *in vivo* studies with laboratory animals, *in vitro* studies with cultured human cells, or human data from case or population studies. Thus, the mode-of-action studies suggest that arsenic might be acting as a cocarcinogen, a promoter, or a progressor of carcinogenesis.

### 3.3 Arsenic-Induced Keratosis

Arsenical keratosis is often regarded as premalignant and the clinical syndrome comprises several pathological features including hyperkeratosis, parakeratosis, arsenical pigmentation, and squamous cell carcinoma *in situ* (Bowen disease). There are two gross variations; the first being multiple, punctate, hard, yellow, corn-like papules arising in areas of friction such as on the soles and palms, and the second being slightly elevated, scaly, erythematous or pigmented patches appearing on non-sun-exposed areas.

Arsenical bowenoid lesions are noninvasive squamous cell carcinomas (*in situ*) that can be found in both sun-exposed and nonexposed areas of the skin. Atypical epithelial cells occupy the full thickness of the surface epithelium. Arsenical keratosis and Bowen disease tend to persist for many years before becoming invasive squamous cell carcinoma. Gross pigmentary changes in chronic arseniasis include hyperpigmentation (darker color) and hypopigmented (lighter color) raindrop-like macules ranging in size from 1 to 2 millimeters. Figure 2 presents three examples of arsenic-induced keratotic lesions [61].

### 3.4 Arsenic-Induced Skin Cancer

The skin manifestations are most diagnostic for chronic arsenic toxicity as internal organ cancers have no characteristic histologic features suggestive of arsenical etiology. Some studies of arsenic toxicity suggest that internal cancers behave differently and may have a different mechanism from skin cancers as related to concentration and length of exposure [62].

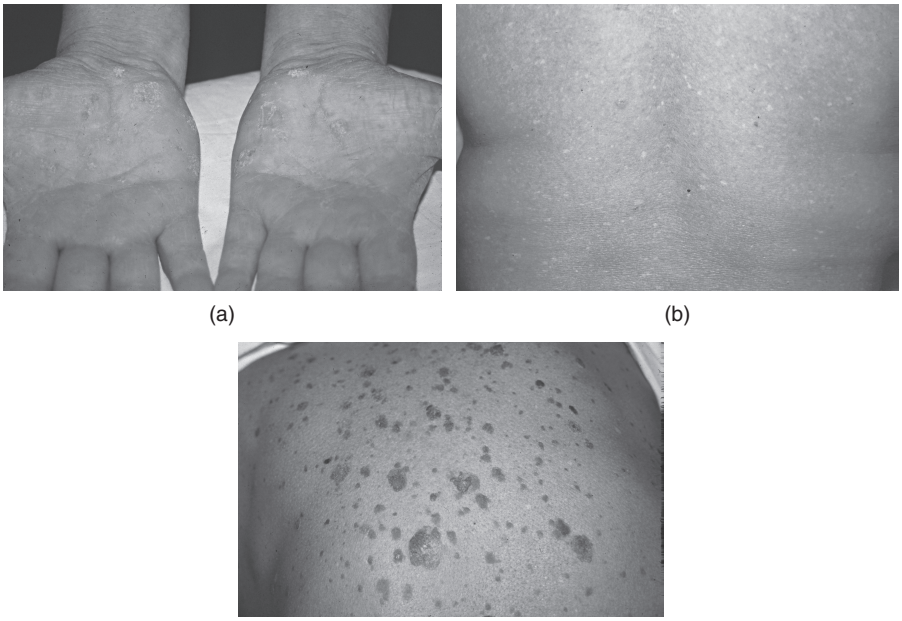


**FIGURE 2** Arsenic-induced skin lesions: (a) the palms show multifocal hyperkeratotic lesions; (b) skin of back with "rain-drop" like hyperpigmented and hypopigmented macules; and (c) skin of back with chronic arsenical keratotic lesions.

Squamous cell carcinomas (SCCs) are invasive skin tumors that arise from the surface epidermal layer. Histologically, they resemble normal surface squamous cells but with atypical features and growth patterns. SCCs typically arise in sun-exposed areas of the body, but in chronic arsenic ingestion, they are usually found in sun-protected areas.

Basal cell carcinomas (BCCs) that are arsenic-induced most often arise from normal tissue, are almost always multiple, and frequently occur on the trunk. Malignant cells in the lower part of the epidermis show a variety of growth patterns including adenoid, reticulated, trichoepitheliomatous, and hyperpigmented. In one study, arsenical BCC was associated with multinucleated giant cells, some of which had a bizarre appearance and atypical mitoses [63]. Figure 3 illustrates two cases of arsenic-induced basal cell carcinoma [61].

Malignant melanomas arise from pigment making cells, melanocytes, that are scattered along the base of the squamous epithelium. Several conflicting reports tying arsenic exposure to the development of malignant melanoma are found in the literature. One report on art glassware workers exposed to arsenic in Italy found one adenocarcinoma, one melanoma, and one squamocellular carcinoma in the sinonasal tract [64]. However, the National Cancer Registry Program in Taiwan contained information on cell types of arsenic-related skin cancers in 2369 patients, and all were either associated with squamous cell carcinoma or basal cell carcinoma, but none with malignant melanoma [65].

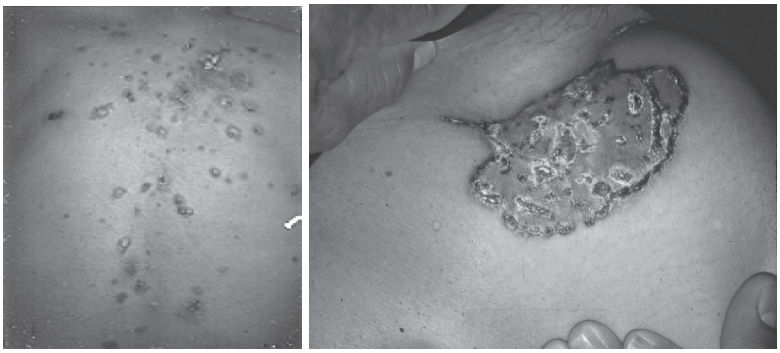

(a) (b)

**FIGURE 3** Photograph demonstrating patients with basal cell carcinoma: (a) multifocal basal cell carcinoma of the back; and (b) patient with a single lesion of the left groin.

### 3.5 Arsenic-Induced Liver Cancer

The liver is the main organ for detoxification of substances entering the body. Exposure to toxic agents may result in a variety of hepatic lesions that may resolve when the toxin is removed, or evolve to hepatic failure, malignancy, or death. Chronic arsenic exposure causes liver injury when it accumulates and damages the cells' organelles: these include macrovesicular steatosis, phospholipidosis, cholestatic lesions, steatohepatitis, granulomatous reactions, fibrosis, cirrhosis, vascular lesions, and neoplasms. Arsenic has been reported to enhance hepatic morphological and biochemical changes in phenobarbital-pretreated rats. Hydrophic degeneration, total loss of glycogen, necrosis in some centrolobular zones, and an increase in lipid vacuoles around the periportal area have been observed [66]. Subchronic exposure to arsenic has also been associated with hepatic histopathological changes in rats including the disruption of hepatic cord, sinusoidal dilation, fatty infiltration, and altered expression of cyclin D1, p27, JLK, PTEN, and beta-catenin in the liver [67]. In a study evaluating the combined effect of a mixture of chemicals (arsenic, benzene, chloroform, chromium, lead, phenol, and tetrachloroethylene) on Fischer-344 rats, Constan et al. [68] reported a direct correlation between apoptosis and changes in cell proliferation associated with low level exposure.

Liver angiosarcomas have been associated through medical treatment with inorganic trivalent arsenic compounds, particularly Fowler's solution. Angiosarcoma of the liver is the malignant transformation of the sinusoidal endothelial cells and is linked to exposure to inorganic arsenic as well as vinyl chloride, anabolic steroids, and Thorotrast [69]. Histologically, angiosarcomas exhibit two growth patterns—cavernous and solid. The cavernous areas contain plump, atypical endothelial cells that line the dilated sinusoids, while solid areas show sheets of anaplastic-spindled tumor cells and often have areas with hemorrhage and necrosis. The cells are immunopositive for the endothelial markers CD31, CD34, and Factor VIII-related antigen [70].

Hepatocellular carcinoma (HCC) constitutes a malignant transformation of hepatocytes. Recent studies have documented its association with ingested inorganic arsenic [42]. Large populations exposed to elevated levels of arsenic in drinking

water have been studied in Japan, Mexico, Chile, Germany, Argentina, Taiwan, China, India, and Bangladesh. In Taiwan, Japan, and Germany, a dose–response relationship between inorganic arsenic and HCC has been reported. This has also been linked to the interaction of arsenic with other risk factors, including hepatitis B and C infection, aflatoxin exposure, alcohol abuse, and genetic hemochromatosis [71, 72]. Histologically, hepatocellular carcinoma resembles normal liver, and a well-differentiated HCC requires close examination for lack of acinar architecture, growth pattern, and positivity for sinusoidal staining with CD34 (Q-Bend-10) immunostain, which will be increased in neoplastic liver [73]. Moderately differentiated HCCs are easier to diagnose, as they still resemble liver but have increased nuclear pleomorphism, mitoses, and growth pattern abnormalities (thickened trabeculae) and lack normal acinar architecture. However, poorly differentiated HCCs need clinical investigations to exclude metastatic disease. Figure 4 shows the histological sections of a normal liver (A), and neoplasia associated with chronic exposure to arsenic, including hepatocellular carcinoma (B) and liver angiosarcoma (C).

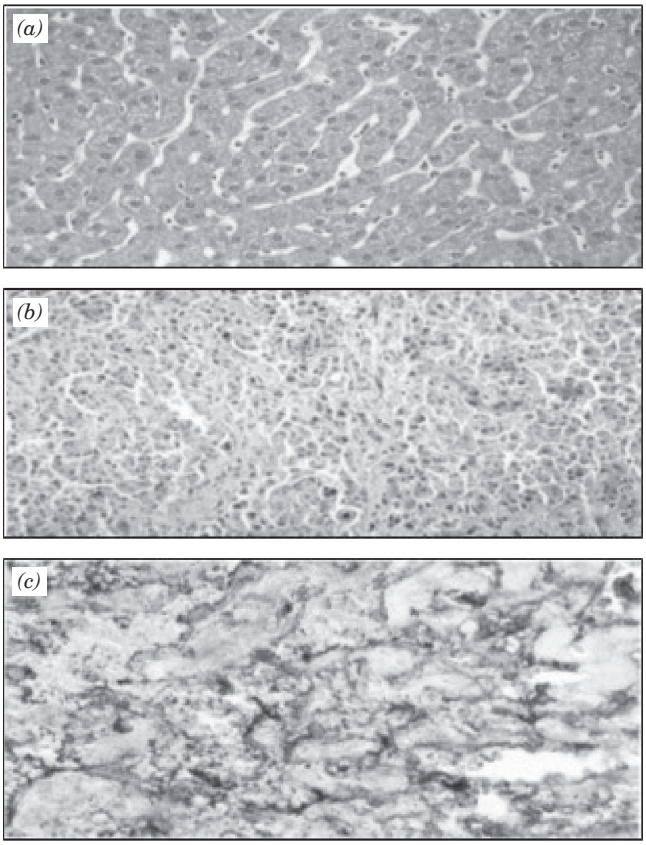


**FIGURE 4** Histological sections of (a) normal liver, and neoplasia associated with chronic exposure to arsenic, including (b) hepatocellular carcinoma and (c) liver angiosarcoma.

### 3.6 Arsenic-Induced Kidney Cancer

Chronic arsenic exposure has been associated with kidney cancer. A 30% increased incidence of renal carcinoma has been reported in smelter workers in Tacoma, Washington (USA), who were chronically exposed to arsenic as a by-product of smelting nonferrous metal ores [43]. Mining workers and copper smelters not only develop renal carcinoma but also have increased rates of lung cancer, gastrointestinal cancer, and hematolymphatic malignancies. The development of carcinoma appears to be a clear dose-related response. Histologically, the renal tumor is not renal cell carcinoma, but a malignancy of the lining epithelium of the renal pelvis, or transitional cell carcinoma, similar to that of the urinary bladder (see Section 3.7).

### 3.7 Arsenic-Induced Urinary Bladder Cancer

Transitional cell carcinoma (TCC) from arsenic exposure in drinking water has been reported in Argentina, Chile, and Taiwan [71, 74, 75]. It has an 8–20 year latent period and appears to be dose related. This tumor develops in the lining of the renal pelvis, ureters, and urinary bladder. Urothelial carcinomas comprise two distinct diseases, both biologically and molecularly: a low-grade papillary tumor, which frequently recurs; and a high-grade malignancy, which can present as dysplasia or carcinoma *in situ* but frequently presents as invasive disease. Histologically, more than seven layers of urothelium cover the papillae. The cells vary from almost normal in appearance to highly atypical. The cytoplasm is pale eosinophilic and usually moderate in amount but can vary and be more abundant, densely eosinophilic, or clear. There are three grades using cytologic criteria that correlate with prognostic significance. Grades 1 through 3 increase with the number of layers of urothelial cells, cytologic atypia, and mitotic figures present. Invasive TCC can arise from both papillary and sessile precursors. The histologic appearance is variable and there are several specific variants. The tumor frequently grows in nests, small clusters, and single cells, dispersed in the lamina propria or invading into the muscularis propria. The tumor cells are usually medium in size with modest eosinophilic cytoplasm, but in some cases the cytoplasm is abundant, densely eosinophilic, or clear. The urothelial characteristics of the tumor cells are usually appreciable, except in cases of high-grade tumors [76].

### 3.8 Arsenic-Induced Lung Cancer

Recent evidence from both epidemiological and experimental studies have indicated that arsenic and cigarette smoking exposure act synergistically to increase the incidence of lung cancer, through induction of oxidative stress leading to an increase in DNA oxidation and reduction of glutathione levels in the lung [77]. Arsenic-related lung cancers reported in smelter workers tend to be adenocarcinomas, followed by small cell carcinoma [78]. Bronchogenic "non-small-cell" carcinoma is the most common type of primary lung cancer. The incidence is clearly higher in smokers and is increased with the number of "pack-years" of smoking. Other etiologic factors include exposure to radiation, asbestos, radon, and inhaled substances, such as nickel, chromates, and arsenic, and genetic factors [79]. It is estimated that

1.5 million workers in the United States are occupationally exposed to arsenic in industries that manufacture glass, pigments, pesticides, and paints. Studies have shown a dose-related association between arsenic in drinking water with both kidney and lung cancers [71].

Primary pulmonary adenocarcinoma is a malignant epithelial neoplasm with features of glandular differentiation. It tends to arise in the periphery of the lung and is more common in women. The tumor histology may vary from well-differentiated glands, to poorly differentiated single cells that are only recognized as adenocarcinoma based on the demonstration of intracellular mucin production. In well-differentiated adenocarcinoma, the glands are lined by tall columnar or mucinous epithelium with ample cytoplasm, basal nuclei, prominent nucleoli, and frequent mitoses. As differentiation decreases, there is increased nuclear atypia and mitosis [80].

Squamous cell carcinoma (SCC) is a malignant epithelial neoplasm with features of squamous differentiation. It arises in the bronchus, close to the hilum. Grading depends on the presence of intercellular bridges and keratinization, and higher grades have more pronounced cytologic atypia, increased mitoses, and frequent foci of necrosis and hemorrhage. *In situ* lesions are recognized in this type of tumor, but there is no *in situ* counterpart for adenocarcinoma. Histologically, the cells show a plate-like growth pattern, have ample eosinophilic cytoplasm, round to oval nuclei, prominent nucleoli, and well-defined cell borders with intercellular bridges [80]. Occasionally, the cytoplasm appears water-clear due to glycogen accumulation, but the tumor still behaves like conventional SCC. Other variants include spindle cell squamous carcinoma, basaloid carcinoma, and lymphoepithelioma-like carcinoma [81].

Small cell carcinoma is one of the least differentiated neuroendocrine neoplasms of the lung and is often centrally located. Patients typically present with symptoms of airway obstruction or may have signs of one of the paraneoplastic syndromes, that is, Cushing syndrome, inappropriate antidiuretic hormone secretion, and Eaton–Lambert syndrome. Histologically, the tumor is composed of small, round to oval, primitive appearing cells, growing in haphazard sheets and nests with scant cytoplasm, round to oval nuclei, dense granular chromatin, and inconspicuous nucleoli. There is brisk mitotic activity and cellular "molding." Distortion "crush" artifact is commonly seen in biopsy material, as well as basophilic deposition of DNA around blood vessels (Azzopardi phenomenon). In Nakajo, Japan, small cell carcinoma was reported in patients exposed to arsenic in well-water with an exposure of five years [82]. Neuroendocrine differentiation may be difficult to demonstrate immunohistochemically, but sometimes positive reactions can be seen using chromogranin A or synaptophysin [81].

### 3.9 Arsenic-Induced Gastrointestinal Cancer

Conflicting reports linking carcinoma of the gastrointestinal tract and arsenic ingestion are scattered in the literature [83]. Most cases are of multiple cancers developing many years after initial exposure. A Japanese study of 839 copper smelters showed significantly increased mortalities from lung and colon cancers [84], while gold miners in Ontario, Canada had excess mortality from *stomach cancer* [85]. Swedish glass-blowers exposed to high levels of lead, arsenic, antimony, and man-

ganese had increased risk of dying from cancer of the stomach, lung, and colon [86]. Although histologic types are not listed, adenocarcinoma would be the most common in the gastrointestinal tract.

### 3.10 Arsenic-Induced Brain Cancer

A Swedish study of occupational exposure to chemicals including arsenic, mercury, and petroleum products found an increased incidence of gliomas in men and an increased incidence of both gliomas and meningiomas in women [87]. About half of all primary brain tumors and about one-fifth of all primary spinal cord tumors are gliomas. Gliomas grow from glial cells within the brain and usually occur in the cerebral hemispheres but may also strike other areas, especially the optic nerve, the brain stem, and, particularly among children, the cerebellum. Gliomas are classified into several groups because there are different kinds of glial cells.

Astrocytomas, the most common type of glioma, are also the most common type of primary brain tumor. They develop from star-shaped glial cells called astrocytes. Well-differentiated or low-grade (grade I or II) astrocytomas contain cells that are relatively normal and are less malignant than the other two. They grow relatively slowly and may sometimes be completely removed through surgery. In some cases these tumors can also progress or recur as higher grade tumors. Anaplastic (or high-grade) astrocytomas (grade III) grow more rapidly than lower grade astrocytomas and contain cells with malignant traits.

## 4 RADIATION-INDUCED TOXICOLOGIC PATHOLOGY

### 4.1 Ionizing Radiation and Ultraviolet Radiation

Both ionizing radiation and ultraviolet radiation appear to be important in the evaluation of radiation effects on humans. Human exposure to ionizing radiation involves external exposures from radioactive materials in the environment, and internal exposures from naturally occurring radionuclides deposited in the body. Diagnostic X-rays used for medical purposes are the second largest source of human exposure for ionizing radiation. The sun is the most common source of ultraviolet radiation (UV) exposure for humans and other animals [88]. UV exposure from the sun is also considered as the most important cause of skin cancer. It has been reported that sunburns and excessive exposures cause cumulative damage, which induces immunosuppression and skin cancers [89]. Although the relationship for melanoma and basal cell carcinoma (BCC) is complex, there exists strong evidence implicating the sun as a potent inducer of squamous cell carcinoma (SCC) [85].

It has been reported that intermittent sun exposure is important in the pathogenesis of BCC, whereas cumulative exposure is important for both BCC and SCC [90, 91]. Other studies have pointed out that sunlight and psoralen and UV-A are risk factors for the development of SCC and, to a lesser extent, BCC. UV-A irradiation has been associated with free radical formation, including peroxides, superoxide anion, and hydroxyl radical, and macromolecule damage, including DNA, RNA, proteins, and lipids. Adverse cellular effects include apoptotic cell death, cell cycle arrest, mutation, and altered gene expression [88]. Most target organs of toxicity

include the skin, with skin cancer being the most critical effect, the eyes, and the immune system. UV-B used in the treatment of psorasis might also increase the risk of these tumors [92].

The biologic effect of ionizing radiation has been attributed to its direct interaction with macromolecules or its indirect effect associated with free radical generation and subsequent induction of oxidative stress. Most of the significant effects of ionizing radiation from cell killing to carcinogenesis appear to be related to DNA damage. Ionizing radiation has been implicated in the induction of large scale structural DNA damage, resulting in the loss of entire genes or groups of genes from aberrations such as deletions, rearrangements, and recombination. Radiation-induced genomic instability is characterized by alterations in the mammalian genome and is recognized as a hallmark of cancer cells and carcinogenesis [88]. Studies of childhood thyroid cancers in the territories affected by the Chernobyl nuclear accident have demonstrated a significant increase in these cancers associated with the presence of I-131 and other short-lived isotopes of idione released into the environment. Also, the mutational spectrum of these cancers demonstrates that gene rearrangements that lead to the activation of protein kinase signaling seem to have a pivotal role; point mutations are rare [93].

Alteration in cell cycle progression and chromosome damage has also been reported in radiation pathology. Ionizing radiation is considered as being a complete carcinogen because it has been fully implicated in all three stages of carcinogenesis including initiation, promotion, and progression. It is also considered to be a systemic toxicant capable of causing injury to various organ systems/tissues including the hematopoietic and lymphoid systems, the respiratory system, the dermatologic system, the gastrointestinal system, the skeletal system, the ocular system, the cardiovascular system, the urogenital/reproductive system, and the integumentary system [87].

### 4.2 Nonionizing Radiation

Nonionizing radiation (NIR) includes electromagnetic energy distributed as near-ultraviolet and visible light, infrared radiation, microwaves, radio frequencies, and very low frequency and extremely low frequency alternating electric and magnetic fields. Almost every member of modern society is exposed to it in some form. Usually the intensity of exposure is low in the general population but can be greatly increased in the workplace [94]. However, the biological effects of NIR remain to be elucidated [95]. Although sunlight contains all UV wavelengths, only the UV-A and a portion of the UV-B are of biologic significance because the Earth's atmosphere screens out UV wavelengths shorter than 280 nm [88].

## 5 CONCLUSION

The science of toxicologic pathology has been critical for the evaluation and understanding of adverse effects of physical, chemical, and biologic agents on biologic systems. This emerging science is very important in health risk assessment and management as it provides public health officials with a strong scientific basis for informed decision making. It serves as a tool for the diagnosis of disease status and

the evaluation of the toxicity induced by various environmental exposures. The examination of tissues at the molecular, cellular, target organ, and organismal levels provides an excellent opportunity to describe the adverse effects that are associated with environmental exposure, and to understand the pathological manifestations of toxicity. This includes the description of adaptive changes in response to perturbations of homeostasis, and degenerative lesions and alterations induced by toxic exposures.

## ACKNOWLEDGMENTS

This work has been supported in part by the NIH-RCMI Center for Environmental Health (Grant No. 1G12RR13459) and the United States Department of the Army Chemical Materials and Computational Modeling Cooperative Agreement No. W912H2-04-2-0002 to Jackson State University (Jackson, Mississippi, USA), and in part by the Armed Forces Institute of Pathology (Washington DC, USA).

## REFERENCES

1. Rousseaux CG, Haschek WM, Wallig MA. Toxicologic pathology: an introduction. In Haschek WM, Rousseaux CG, Wallig MA, Eds. *Handbook of Toxicologic Pathology*. New York: Academic Press; 2002, pp 31–114.
2. Wallig MA. Morphologic manifestations of toxic cell Injury. In Haschek WM, Rousseaux CG, Wallig MA, Eds. *Handbook of Toxicologic Pathology*. New York: Academic Press; 2002, pp 39–65.
3. Coppola A, Diano S. Hormonal regulation of the arcuate nucleus melanocortin system. *Front Biosci* 2007;12:3519–3530.
4. Seeley RJ, Drazen DL, Clegg DJ. The critical role of the melanocortin system in the control of energy balance. *Annu Rev Nutr* 2004;24:133–149.
5. Benoit S, Schwartz M, Baskin D, Woods SC, Seeley RJ. CNS melanocortin system involvement in the regulation of food intake. *Horm Behav* 2000;37(4):299–305.
6. Akerfelt M, Trouillet D, Mezger V, Sistonen L. Heat shock factors at a crossroad between stress and development. *Ann NY Acad Sci* 2007 (in press).
7. Calabrese V, Lodi R, Tonon C, D'Agata V, Sapienza M, Scapagnini G, Mangiameli A, Pennisi G, Stella AM, Butterfield DA. Oxidative stress, mitochondrial dysfunction and cellular stress response in Friedreich's ataxia. *J Neurol Sci* 2005;233(1–2):145–162.
8. Leidhold C, Voos W. Chaperones and proteases: guardians of protein integrity in eukaryotic organelles. *Ann NY Acad Sci* 2007 (in press).
9. Zhang K, Kaufman RJ. The unfolded protein response: a stress signaling pathway critical for health and disease. *Neurology* 2006;66(2):S102–S109.
10. Cheung P, Pawling J, Partridge EA, Sukhu B, Grynpas M, Dennis JW. Metabolic homeostasis and tissue renewal are dependent on {beta}1,6GlcNAc-branched N-glycans. *Glycobiology* 2007;17(8):828–837.
11. Wellen KE, Fucho R, Gregor MF, Furuhashi M, Morgan C, Lindstad T, Vaillancourt E, Gorgun CZ, Saatcioglu F, Hotamisligil GS. Coordinated regulation of nutrient and inflammatory responses by STAMP2 is essential for metabolic homeostasis. *Cell* 2007;129(3):537–548.

12. Calabrese V, Guagliano E, Sapienza M, Mancuso C, Butterfield DA, Stella AM. Redox regulation of cellular stress response in neurodegenerative disorders. *Ital J Biochem* 2006;55(3–4):263–282.
13. Calabrese V, Boyd-Kimball D, Scapagnini G, Butterfield DA. Nitric oxide and cellular stress response in brain aging and neurodegenerative disorders: the role of vitagenes. *In Vivo* 2004;18(3):245–267.
14. Cypryk K, Vilsboll T, Nadel I, Smyczynska J, Holst JJ, Lewinski A. Normal secretion of the incretin hormones glucose-dependent insulinotropic polypeptide and glucagon-like peptide-1 during gestational diabetes mellitus. *Gynecol Endocrinol* 2007;23(1):58–62.
15. Reed DJ. Mechanisms of chemically induced cell injury and cellular protection mechanisms. In Hodgson E, Smart RC, Eds. *Introduction to Biochemical Toxicology*, 3rd ed. Hoboken NJ: Wiley; 2001, Chap 10, pp 221–253.
16. Malhi H, Gores GJ, Lemasters JJ. Apoptosis and necrosis in the liver: a tale of two deaths? *Hepatology* 2006;43(2):S31–S44.
17. Gill JH, Dive C. Apoptosis: basic mechanisms and relevance to toxicology. In Roberts R, Ed. *Apoptosis in Toxicology*. New York: Taylor & Francis; 2000, Chap 1, pp 1–19.
18. Cosulich SC, Roberts RA. Perturbation of apoptosis as a mechanism of action of non genotoxic carcinogens. In Roberts R, Ed. *Apoptosis in Toxicology*. New York: Taylor & Francis, 2000, pp 169–185.
19. Gibson RM. Role of apoptosis in neuronal toxicology. In Roberts R, Ed. *Apoptosis in Toxicology*. New York: Taylor & Francis; 2000, pp 145–167.
20. Woolveridge I, Morris I. Apoptosis in male reproductive toxicology. In Roberts R, Ed. *Apoptosis in Toxicology*. New York: Taylor & Francis; 2000, pp 71–93.
21. Tilly JL. Apoptosis in female reproductive toxicology. In Roberts R, Ed. *Apoptosis in Toxicology*. New York: Taylor & Francis; 2000, pp 95–115.
22. Wyllie A. Cell death–apoptosis and necrosis. In *Apoptosis and Cell Proliferation*, 2nd ed. Mannheim, Germany: Boehringer Mannheim Gmbh; 1998, pp 1–61.
23. Tanuma SI. Molecular mechanisms of apoptosis. In Sluyser M, Ed. *Apoptosis in Normal Development and Cancer*. New York: Taylor & Francis; 1996, pp 39–59.
24. Davis MA, Jeffery EH. Organelle biochemistry and regulation of cell death. In Haschek WM, Rousseaux CG, Wallig MA, Eds. *Handbook of Toxicologic Pathology*, 2nd ed. New York: Academic Press; 2002, pp 67–81.
25. Roset R, Ortet L, Gil-Gomez G. Role of Bcl-2 family members on apoptosis: what we have learned from knock-out mice. *Front Biosci* 2007;1(12):4722–4730.
26. Chan SL, Yu VC. Proteins of the bcl-2 family in apoptosis signaling: from mechanistic insights to therapeutic opportunities. *Clin Exp Pharmacol Physiol* 2004;31(3):119–128.
27. Lanave C, Santamaria M, Saccone C. Comparative genomics: the evolutionary history of the Bcl-2 family. *Gene* 2004;333:71–79.
28. Opferman JT, Korsmeyer SJ. Apoptosis in the development and maintenance of the immune system. *Nat Immunol* 2003;4(5):410–415.
29. Mischiati C, Puviani AC, Brogli M, Guarniero S, Sereni A, Breda L, Ricci D, Galavotti D, Morsiani E, Gambari R. Modulation of pro-apoptotic (Bax) and anti-apoptotic (Bcl-2) gene expression in isolated porcine hepatocytes perfused within a radial-flow bioreactor after low-temperature storing. *Int J Artif Organs* 2003;26(2):139–148.
30. Harada T, Taniguchi F, Izawa M, Ohama Y, Takenaka Y, Tagashira Y, Ikeda A, Watanabe A, Iwabe T, Terakawa N. Apoptosis and endometriosis. *Front Biosci* 2007;12:3140–3151.
31. Berthet C, Rodriguez-Galan MC, Hodge DL, Gooya J, Pascal V, Young HA, Keller J, Bosselut R, Kaldis P. Hematopoiesis and thymic apoptosis are not affected by the loss of Cdk2. *Mol Cell Biol* 2007;2714:5079–5089.

32. Smart RC, Akunda JK. Carcinogenesis. In Hodgson E, Smart RC, Eds. *Introduction to Biochemical Toxicology*, 3rd ed. Hoboken, NJ: Wiley; 2001, pp 343–395.
33. Mostorides S, Maronpot RR. Carcinogenesis. In Haschek WM, Rousseaux CG, Wallig MA, Eds. *Handbook of Toxicologic Pathology*. New York: Academic Press; 2002, pp 3–122.
34. Pilot HC, Dragan YP. Chemical carcinogenesis. In Klaassen CD, Ed. *Cassarett & Doull's Toxicology: The Basic Science of Poisons*. New-York: McGraw-Hill; 1996, pp 201–267.
35. Voutsadakis IA. Pathogenesis of colorectal carcinoma and therapeutic implications: the roles of the ubiquitin-proteasome system and Cox-2. *J Cell Mol Med* 2007; 11(2):252–285.
36. Moran EM. Epidemiological and clinical aspects of nonsteroidal anti-inflammatory drugs and cancer risks. *Environ Pathol Toxicol Oncol* 2002;21(2):193–201.
37. Buskens CJ, Ristimaki A, Offerhaus GJ, Richel DJ, van Lanschot JJ. Role of cyclooxygenase-2 in the development and treatment of oesophageal adenocarcinoma. *Scand J Gastroenterol* Suppl 2003;239:87–93.
38. Horree N, van Diest PJ, van der Groep P, Sie-Go DM, Heintz AP. Progressive derailment of cell cycle regulators in endometrial carcinogenesis. *J Clin Pathol* 2007 (in press).
39. Mitsudomi T, Oyama T, Kusano T, Osaki T, Nakanishi R, Shirakusa T. Mutations of the p53 gene as a predictor of poor prognosis in patients with non-small-cell lung cancer. *J Natl Cancer Inst* 1993;85(24):2018–2023.
40. NAS (National Academy of Science). *Arsenic in Drinking Water*. Washington, DC: National Academies Press; 1999, pp 299–301.
41. NRC (National Research Council). *Arsenic in Drinking Water: Update*. Washington, DC: National Academies Press; 2001.
42. Tchounwou PB, Wilson B, Ishaque A. Important considerations in the development of public health advisories for arsenic and arsenic-containing compounds in drinking water. *Rev Environ Health* 1999;14(4):211–229.
43. Tchounwou PB, Centeno JA, Patlolla AK. Arsenic toxicity, mutagenesis, and carcinogenesis—a health risk assessment and management approach. *Mol Cell Biochem* 2004; 255(1-2):47–55.
44. Tchounwou PB, Patlolla AK, Centeno JA. Carcinogenic and systemic health effects associated with arsenic exposure—a critical review. *Toxicol Pathol* 2003;31(6):575–588.
45. Tchounwou PB, Wilson BA, Abdelgnani AA, Ishaque AB, Patlolla AK. Differential cytotoxicity and gene expression in human liver carcinoma (HepG2) cells exposed to arsenic trioxide and monosodium acid methanearsonate (MSMA). *Int J Mol Sci* 2002;3:1117–1132.
46. Li JH, Rossman TC. Inhibition of DNA ligase activity by arsenite: a possible mechanism of its comutagenesis. *Mol Toxicol* 1989;2:1–9.
47. Belton JC, Benson NC, Hanna ML, Taylor RT. Growth inhibition and cytotoxic effects of three arsenic compounds on cultured Chinese hamster ovary cells. *J Environ Sci Health* 1985;20A:37–72.
48. Goyer RA. Toxic effects of metals. In Klaassen CD, Ed. *Cassarett & Doull's Toxicology: The Basic Science of Poisons*. New York: McGraw-Hill; 1996, pp 691–736.
49. Lee TC, Oshimura M, Barrett JC. Comparison of arsenic-induced cell transformation, cytotoxicity, mutation and cytogenetic effects in Syrian hamster embryo cells in culture. *Carcinogenesis* 1985;6(10):1421–1426.
50. Tchounwou PB, Wilson BA, Ishaque A, Schneider J. Atrazine potentiation of arsenic trioxide-induced cytotoxicity and gene expression in human liver carcinoma cells (HepG2). *Mol Cell Biochem* 2001;222:49–59.

51. Tchounwou PB, Wilson BA, Schneider J, Ishaque A. Cytogenetic assessment of arsenic trioxide toxicity in the Mutatox, Ames II, and CAT-Tox assays. *Metal Ions Biol Med* 2000;6:89–91.
52. Vega L, Gonsebatt ME, Ostrosky-Wegman P. Aneugenic effect of sodium arsenite on human lymphocytes *in vitro*: an individual susceptibility effect detected. *Mutat Res* 1995;334(3):365–373.
53. Patlolla A, Tchounwou PB. Cytogenetic evaluation of arsenic trioxide toxicity in Sprague–Dawley rats. *Mutat Res Genet Toxicol Environ Mutagen* 2005;587(1–2):126–133.
54. Gonsebatt ME, Vega L, Salazar AM, Montero R, Guzman P, Blas J, Del Razo LM, Garcai-Vargas G, Albores A, Cebrian ME, Kelsh M, Ostrosky-Wegman P. Cytogenetic effects in human exposure to arsenic. *Mutat Res* 1997;386:219–228.
55. Jingbo Pi, Hiroshi Y, Yoshito K, Guifan S, Takahiko Y, Hiroyuki A, Claudia HR, Nobuhiro S. Evidence for induction of oxidative stress caused by chronic exposure of Chinese residents to arsenic contained in drinking water. *Environ Health Perspect* 2002;110(4): 331–336.
56. Cavigelli M, Li WW, Lin A, Su B, Yushioka K, Karin M. The tumor promoter arsenite stimulates AP-1 activity by inhibiting a JNK phosphatase. *EMBO J* 1996;15:6269–6279.
57. Kitchin K. Recent advances in arsenic carcinogenesis: modes of action, animal model systems, and methylated arsenic metabolites. *Toxicol Appl Pharmacol* 2001;172: 249–261.
58. Simeonova PP, Luster MI. Mechanisms of arsenic carcinogenicity: genetic or epigenetic mechanisms? *J Environ Pathol Toxicol Oncol* 2000;19:281–286.
59. Barchowsky A, Dudek EJ, Treadwell MD, Wetterhahn KE. Arsenic induces oxidant stress and NF-kB activation in cultured aortic endothelial cells. *Free Radic Biol Med* 1996;21:783–790.
60. Simeonova PP, Wang S, Toriumi W, Kommineni C, Matheson J, Unimye N, Kayama F, Harki D, Ding M, Vallyathan V. Arsenic mediates cell proliferation and gene expression in the bladder epithelium: association with AP-1 transactivation. *Cancer Res* 2000;60: 3445–3453.
61. Centeno JA, Tchounwou PB, Patlolla AK, Mullick FG, Murakat L, Meza E, Gibb H, Longfellow D, Yedjou CG. Environmental pathology and health effects of arsenic poisoning: a critical review. In Naidu R, Smith E, Smith J, Bhattacharya P, Eds. *Managing Arsenic In the Environment: From Soil to Human Health*. Adelaide, Australia: CSIRO Publishing Corp; 2006.
62. Centeno JA, Martinez L, Ladich ER, Page NP, Mullick FG, Ishak KG, Zheng B, Gibb H, Thompson C, Longfellow D. *Arsenic-Induced Lesions*. Washington, DC: Armed Forces Institute of Pathology, American Registry of Pathology; 2000.
63. Schwartz RA. Arsenic and the skin. *Int J Dermatol* 1997;36:241–250.
64. Battista G, Bartoli D, Iaia TE, Dini F, Fiumalbi C, Giglioli S, Valiani M. Art glassware and sinonasal cancer: report of three cases. *Am J Ind Med* 1996;30(1):31–35.
65. Guo HR, Yu HS, Hu H, Monson RR. Arsenic in drinking water and skin cancers: cell-type specificity. *Cancer Causes Control* 2001;12(10):909–916.
66. Albores A, Cebrian ME, Garcia-Vargas GC, Connelly JC, Price SC, Hinton RH, Bach PH, Bridges JW. Enhanced arsenite-induced hepatic morphological and biochemical changes in phenobarbital-pretreated rats. *Toxicol Pathol* 1996;24(20):172–180.
67. Cui X, Li S, Shraim A, Kobayashi T, Hayakawa T, Kanno S, Yamamoto M, Hirano S. Subchronic exposure to arsenic through drinking water alters expression of cancer-related genes in rat liver. *Toxicol Pathol* 2004;32(1):64–72.

68. Constan AA, Benjamin SA, Tessari JD, Baker DC, Yang RS. Increased rate of apoptosis correlates with hepatocellular proliferation in Fischer-344 rats following long-term exposure to a mixture of groundwater contaminants. *Toxicol Pathol* 1996;24(3):315–322.
69. Tsai S, Wang T, Ko Y. Mortality for certain diseases in areas with high levels of arsenic in drinking water. *Arch Environ Health* 1999;54:186–193.
70. Centeno JA, Mullick FG, Martinez L, Page NP, Gibb H, Longfellow D, Thompson C, Ladish ER. Pathology related to chronic arsenic exposure. *Environ Health Perspect* 2002;110(5):883–886.
71. Hopenhayn-Rich C, Biggs ML, Smith AH. Lung and kidney cancer mortality associated with arsenic in drinking water in Cordoba, Argentina. *Int J Epidemiol* 1998;27:561–569.
72. Chen CJ, Yu MW, Liaw YF. Epidemiological characteristics and risk factors of hepatocellular carcinoma. *J Gastroenterol Hepatol* 1997;12(9–10):S294–S308.
73. Di Carlo I, Fraggetta F, Lombardo R, Azzarello G, Vasquez E, Puleo S. CD 34 expression in chronic and neoplastic liver diseases. *Panminerva Med* 2002;44(4):365–367.
74. Johansson SL, Cohen SM. Epidemiology and etiology of bladder cancer. *Semin Surg Oncol* 1997;13(5):291–298.
75. Chiou HY, Chiou ST, Hsu YH, Chou YL, Tseng CH, Wei ML, Chen CJ. Incidence of transitional cell carcinoma and arsenic in drinking water: a follow-up study of 8,102 residents in an arseniasis-endemic area in northeastern Taiwan. *Am J Epidemiol* 2001;153(5):411–418.
76. Royce RK, Ackerman LV. Carcinoma of the bladder: therapeutic and pathologic aspects of 135 cases. *J Urol* 1951;65:66–86.
77. Hays AM, Srinivasan D, Witten ML, Carter DE, Lantz RC. Arsenic and cigarette smoke synergistically increase DNA oxidation in the lung. *Toxicol Pathol* 2006;34(4):396–404.
78. Wicks MJ. Arsenic exposure in a copper smelter as related to histological type of lung cancer. *Am J Ind Med* 1981;2:25–31.
79. Reger RB, Morgan WK. Respiratory cancers in mining. *Occup Med* 1993;8(1):185–204.
80. Travis WD, Travis LB, Deveesa SS. Lung cancer. *Cancer* 1995;75:191–202.
81. Travis WD. Pathology of lung cancer. *Clin Chest Med* 2002;23(1):65–81.
82. Nakadaira H, Endoh K, Katagiri M, Yamamoto M. Elevated mortality from lung cancer associated with arsenic exposure for a limited duration. *J Occup Environ Med* 2002;44(3):291–299.
83. Murata K, Iwazawa T, Takayama T, Yamashita K, Okagawa K. Quadruple cancer including Bowen's disease after arsenic injections 40 years earlier: report of a case. *Surg Today* 1994;24(12):1115–1118.
84. Tokudome S, Kuratsune M. A cohort study on mortality from cancer and other causes among workers at a metal refinery. *Int J Cancer* 1976;17(3):310–317.
85. Kusiak RA, Ritchie AC, Springer J, Muller J. Mortality from stomach cancer in Ontario miners. *Br J Ind Med* 1993;50(2):117–126.
86. Wingren G, Axelson O. Mortality in the Swedish glassworks industry. *Scand J Work Environ Health* 1987;13(5):412–416.
87. Navas-Acien A, Pollan M, Gustavsson P, Plato N. Occupation, exposure to chemicals and risk of gliomas and meningiomas in Sweden. *Am J Ind Med* 2002;42(3):214–227.
88. Benjamin SA, Powers BE, Haln FF, Kusewitt DF. Radiation and heat. In Haschek WM, Rousseaux CG, Wallig MA, Eds. *Handbook of Toxicologic Pathology*, 2nd eds. New York: Academic Press; 2002, pp 529–594.
89. Saladi RN, Persaud AN. The cause of skin cancer: a comprehensive review. *Drugs Today* 2005;41(1):37–53.

90. English DR, Armstrong BK, Kricker A, Fleming C. Sunlight and cancer. *Cancer Causes Control* 1997;8(3):271–283.
91. Almahroos M, Kurban AK. Ultraviolet carcinogenesis in nonmelanoma skin cancer, part II: review and update on epidemiologic correlations. *SkinMed* 2004;3(3):132–139.
92. Lim JL, Stern RS. High levels of ultraviolet B exposure increase the risk of non-melanoma skin cancer in psoralen and ultraviolet A-treated patients. *J Invest Dermatol* 2005;124(3):505–513.
93. Yamashita S, Saenko V. Mechanisms of disease: molecular genetics of childhood thyroid cancers. *Nat Clin Pract Endocrinol Metab* 2007;3(5):422–429.
94. Yost MG. Occupational health effects of nonionizing radiation. *Occup Med* 1992;7(3):543–566.
95. Hietanen M. Health risks of exposure to non-ionizing radiation—myths or science-based evidence. *Med Lav* 2006;97(2):184–188.

# SECONDARY PHARMACODYNAMIC STUDIES AND *IN VITRO* PHARMACOLOGICAL PROFILING

Duncan Armstrong,[1] Jacques Migeon,[2] Michael G. Rolf,[1] Joanne Bowes,[1] Mark Crawford,[2] and Jean-Pierre Valentin[1]

[1]*AstraZeneca R&D, Alderley Park, Macclesfield, Cheshire, United Kingdom*
[2]*Cerep, Redmond, Washington*

**Contents**

1 Introduction
  1.1 Definitions
  1.2 Historical and Regulatory Perspectives
2 Drivers for Secondary Pharmacodynamic Studies
  2.1 Regulatory Requirements
  2.2 Economic Environment
3 Approaches to Pharmacological Profiling
  3.1 Factors Influencing the Approach to Secondary Pharmacodynamics
  3.2 *In Vivo* and *In Vitro* Approaches to Secondary Pharmacodynamics
  3.3 Strategies for *In Vitro* Pharmacological Profiling
4 Technical Considerations for *In Vitro* Profiling
  4.1 Assay Technologies
  4.2 *In Vitro* Panel Design
  4.3 ADME/PK Profiling
  4.4 Data Analysis and Interpretation

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

5 Impact of Secondary Pharmacodynamic Profiling: Case Studies
5.1 Case Study 1. Ranolazine: Secondary Pharmacodynamic and Nonclinical Safety Pharmacology Studies Supporting Regulatory Approval
5.2 Case Study 2. Muscarinic $M_3$ Receptors and Mydriasis: Linking a Molecular Target to a Physiological Effect
5.3 Case Study 3. Chlorpromazine and Reprofiling: The Application of Pharmacological Profiling Beyond Secondary Pharmacodynamics
6 Conclusion
References

## 1 INTRODUCTION

### 1.1 Definitions

Pharmacological investigations during the drug discovery process may be divided into three distinct categories: primary pharmacodynamic, secondary pharmacodynamic, and safety pharmacological studies. Primary pharmacodynamic studies are those in which the mode of action of a substance is investigated in relation to its desired (therapeutic) effect. In contrast, secondary pharmacodynamic studies are defined as those that "investigate the mode of action and/or effects of a substance not related to its desired therapeutic target" [1]. Finally, safety pharmacological studies are those that "investigate the potential undesirable pharmacodynamic effects of a substance on physiological functions in relation to exposure in the therapeutic range and above" [1].

This chapter discusses drivers for secondary pharmacodynamic studies and different strategic approaches to such studies, focusing specifically on *in vitro* pharmacological profiling techniques, data analysis, and interpretation. Some examples of the impact of pharmacological profiling on the drug discovery and development process are provided.

### 1.2 Historical and Regulatory Perspectives

In Japan, the Ministry of Health and Welfare (MHW), now referred to as the Ministry of Health, Labour, and Welfare, promulgated comprehensive guidance for organ function testing as early as 1975 (see Table 1). These guidelines describe which organ systems should be included in a first tier evaluation (Category A studies) and make specific recommendations regarding study design (including description of models, criteria for dose selection, and which endpoints would be included in the investigation). These guidelines also describe a second phase of studies (Category B) to be conducted, as necessary, in the light of the results of the Category A studies (Table 1 [2]). As the Japanese guidelines were the most comprehensive of their time, they became the *de facto* foundation for organ function testing throughout the pharmaceutical industry [3–5]. The organ function studies included in Categories A and B were intertwined with studies whose aim was to identify functions and activities in addition to the primary pharmacological function/activity. Kinter et al. [6] distinguished two subgroups of objectives embedded in the Japanese

studies as safety pharmacology and pharmacological profiling. This concept was enlarged upon by the International Conference on Harmonization (ICH) safety pharmacology Expert Working Group (EWG) to define three categories of pharmacology studies: primary and secondary pharmacodynamic studies, and safety pharmacology studies [1, 5, 7] (Table 1).

The origin of the term secondary pharmacodynamic studies is obscure. It first appeared in the draft guidelines of the ICH M3 and S6 (see Table 1 [8, 9]). ICH S6 stated that secondary pharmacodynamic studies aimed to "investigate the mode of action and/or effects of a substance not related to its desired therapeutic target." The ICH S7 EWG began its work in the first quarter of 1999 and a harmonized safety pharmacology guidance was finalized and adopted by the regional regulatory authorities over 2000–2001 (ICH S7A [1]). The ICH S7A guidance defines the objectives and general principles of secondary pharmacodynamic studies, including the timing of these investigations in relation to the clinical development program and the requirement for Good Laboratory Practice (GLP), where applicable [1, 10].

## 2 DRIVERS FOR SECONDARY PHARMACODYNAMIC STUDIES

### 2.1 Regulatory Requirements

The ICH S7A guideline [1] provides a definition for secondary pharmacodynamic studies (see Section 1.1); however, unlike safety pharmacological studies, secondary pharmacodynamic studies are not required as part of the core assessment of the safety of a new drug substance. The guidance suggests that secondary pharmacodynamic studies performed early in the drug discovery process may contribute to the design of safety pharmacological studies, providing information relevant to, for example, dose selection.

The ICH S7A guidance suggests that in some cases secondary pharmacodynamic findings may contribute to the safety evaluation of a substance and may be considered along with the findings of safety pharmacology studies. In cases where the secondary pharmacodynamic findings are pivotal to the safety evaluation, it may be necessary to conduct secondary pharmacodynamic studies to the Good Laboratory Practice standard [1, 10].

The ICH M4(R3) [14] guideline describes an acceptable format and content for a Common Technical Document (CTD) suitable for submission to regulatory agencies in all three ICH regions. Secondary pharmacodynamic studies should be submitted as part of the pharmacology section of the CTD.

### 2.2 Economic Environment

Drug development is a high-risk venture. Estimates for the full cost of bringing a new drug to market are U.S. $800M for compounds approved in the late 1990s, U.S. $1.1B in 2001, and expected to rise to U.S. $1.9B in 2013 [15]. It is estimated that 70% of this figure supports projects that do not lead to a marketed product [16]. Adding to the development risk is the fact that some compounds succeed in clinical development, reach the market, and, only after widespread use, are found to have unexpected adverse effects. The consequences of these postapproval failures are

**TABLE 1 Regulatory Guidance Documents Referring to Secondary Pharmacodynamic Studies for Pharmacological Profiling**[a]

| Document and Source | Comments | Reference |
|---|---|---|
| *Japanese Guidelines for Non-clinical Studies of Drugs Manual 1995—Guidelines for General Pharmacology Studies* | Defines a list of studies to be conducted on all NCE (Category A) and a list of studies to be conducted on a case-by-case basis (Category B). Recommends the use of *in vitro* assays such as isolated organs and tissues. Indicates that ligand-binding methods should be positively adopted. Guidance to be superseded by ICH S7A and S7B. | 1, 2, 11 |
| Committee for Proprietary Medicinal Products—*Points to Consider: The Assessment of QT Interval Prolongation by Non-cardiovascular Medicinal Products* | Recommends a set of nonclinical (including *in vitro* assays) and clinical studies to address the risk for a noncardiovascular drug to slow cardiac repolarization and prolong the QT/QTc interval. Considers *in vitro* assays aimed at assessing activity at cardiac ion channels. | 12 |
| ICH M3—*Timing of Non-clinical Safety Studies for the Conduct of Human Clinical Trials for Pharmaceuticals* | Provides guidance on the timing of safety pharmacology studies in relation to clinical development. Establishes that safety pharmacology studies should be conducted prior to first administration to humans. | 8 |
| ICH S7A—*Note for Guidance on Safety Pharmacology Studies for Human Pharmaceuticals* | Provides clear definitions for primary, secondary, and safety pharmacology studies. Provides the general framework for *in vitro* and *in vivo* safety pharmacology studies. Ligand-binding or enzyme assays data suggesting a potential for adverse effects should be considered in selecting and designing safety pharmacology studies. | 1 |
| ICH S7B—*Note for Guidance on the Non-clinical Evaluation for Delayed Ventricular Repolarization (QT Interval Prolongation) by Human Pharmaceuticals* | Describes a nonclinical testing strategy for assessing the potential of a test substance to slow ventricular repolarization. Includes information concerning nonclinical assays and integrated risk assessment. Effects on $I_{Kr}$ current and other cardiac ion channels should be considered as part of "core battery" and "follow-up" studies, respectively. | 3 |
| *Draft Guideline on the Non-clinical Investigation of the Dependence Potential of Medicinal Products* | Provides guidance on the need for testing of dependence potential in animals during the development of medicinal products and indicates what type of information is expected as part of a MAA. Receptor-binding studies of parent drug and metabolites should be considered to reveal signals pointing toward dependence potential. | 13 |
| ICH M4—*Common Technical Document* | Describes an acceptable format and content for submission to regulatory agencies. Secondary pharmacodynamics studies should be submitted as part of the pharmacology section. | 14 |

[a]ICH, International Conference on Harmonization; IND, investigational new drug; MAA, Marketing Authorization Application; NCE, new chemical entity.

enormous from the perspective of individual patients, public health, and the pharmaceutical industry. Thus, there is a high value in all technologies that can identify potential clinical liabilities at an early stage.

At this time technologies are not available to accurately predict, through preclinical experiments, the many factors related to clinical safety and efficacy. To understand the full scope of this problem, it must be recognized that:

1. The patient population is heterogeneous [17]. Patients differ at the genetic level with regard to drug transport, drug metabolism, expression of disease, and many other factors not yet understood. Humans change with age [18] and are changed by disease [19].
2. Early testing using tissues, cells, and subcellular components is only modestly predictive; it is not yet possible to reduce human physiology to an array of *in vitro* tests.
3. Preclinical testing using animal models is only modestly predictive of effects in humans.

The pharmaceutical industry has long recognized the magnitude of this problem and made significant investments to address it. As one example of success [20], in 1991 ~40% of attrition was due to poor pharmacokinetics/bioavailability. By 2000, attrition due to these factors was reduced to ~10%. During this time period, most pharmaceutical companies had implemented sophisticated panels of early *in vitro* absorption, distribution, metabolism, and excretion (ADME) assays in the discovery/development process. These included intestinal permeability assays, hepatic metabolism assays, and physicochemical characterization. Many in the industry believe that the use of this early testing strategy is the basis for the profound reduction in attrition due to poor pharmacokinetics.

At this time the major factors leading to clinical attrition are safety/toxicology issues and lack of efficacy that is only apparent at the clinical level. Unanticipated safety and efficacy issues each account for approximately 30% of current attrition [20]. This chapter focuses on some of the key *in vitro* technologies that can improve early prediction of clinical safety and efficacy. It is anticipated that, over time, these methods will have a profound impact on clinical attrition and on the economics of drug development.

## 3 APPROACHES TO PHARMACOLOGICAL PROFILING

### 3.1 Factors Influencing the Approach to Secondary Pharmacodynamics

In addition to regulatory requirements, several factors will influence the strategy for secondary pharmacology in any given pharmaceutical organization. The main factors are presented in Table 2. Thus, the pharmaceutical industry is beset with a number of significant challenges to achieve high quality, high throughput, and predictive secondary pharmacological studies, during the early stages of the drug discovery process. As well as satisfying project demands, scientific efficacy, selectivity and safety questions, international regulatory guidelines, and increased patient awareness, secondary pharmacology is increasingly being used to enable informed

**TABLE 2 Factors Influencing the Approach to Secondary Pharmacodynamic Studies**

| Increasing | Decreasing |
|---|---|
| Number, complexity, and stringency of regulatory requirements | Supply of compound during the early discovery stages |
| Number and novelty of new chemical entities | Availability of scientific and technical expertise in key areas (e.g., Integrative Physiology and Pharmacology) |
| Number, diversity, and novelty of molecular targets and approaches | Late stage attrition |
| Throughput of *in vitro* versus *in vivo* assays | Discovery and development time |
| Risk identification initiatives during early discovery stages (e.g., "front loading" initiatives) | Numbers of animals used in the pharmaceutical industry |
| Patient awareness and expectations | |
| Predictive value of *in vivo* and *in vitro* nonclinical assays with respect to human safety | |

decision making. One of the key factors influencing the approach to secondary pharmacology is the rapidly evolving scientific and technological knowledge. A recent example is the evolution from conventional labor-intensive patch-clamp electrophysiology to medium throughput electrophysiology-based platforms to assess activities at ion channels [21].

### 3.2 *In Vivo* and *In Vitro* Approaches to Secondary Pharmacodynamics

Two complementary approaches to secondary pharmacodynamic studies may be considered: *in vivo* observational methods and *in vitro* pharmacological profiling.

An *in vivo* strategy may be employed to measure the integrated response to drug challenge in the whole animal, where a variety of factors such as genetics, biochemistry, metabolism, and environment combine to influence the observed effect. The advantage of such an approach is that the effect of the test substance is determined in an intact physiological system; of course, this offers its own challenge in interpretation and translating the results to other species.

The complexity of study design, numbers of animals, and quantity of compound required to conduct *in vivo* studies require a rationally designed, hypothesis-driven approach. Prior knowledge of adverse events associated with a related target class or chemical structure is invaluable in designing effective studies. Selection of the most appropriate animal model is also important: studies may be conducted in specific disease models, or in knock-out animals in order to increase the predictive power of the chosen model.

Interpretation of data from *in vivo* studies is complex as the molecular target or targets responsible for mediating a given response are unknown and it may be difficult to distinguish the effects of the drug due to action at the primary target from effects mediated through secondary targets.

*In vivo* secondary pharmacodynamic studies are clearly not suited to a high throughput approach, which would make it difficult to apply an *in vivo* approach to

a feedback system in early drug discovery. However, this approach may be more applicable in later stages of compound development, where refinement of predictions from *in vitro* studies on small numbers of compounds becomes important.

A complementary approach to *in vivo* studies is the more recent and now more common *in vitro* method of pharmacological profiling in which the compound of interest is screened in a battery of *in vitro* assays to determine its affinity at a range of receptors, ion channels, transporters, and enzymes. Determination of the potency and efficacy of the drug in functional assays may generate useful information in defining whether it is an agonist or antagonist at a given target. The technologies available for these kinds of analyses are discussed in detail in Section 4.1.

*In vitro* approaches offer a high throughput method for studying secondary pharmacodynamics; integration of *in vitro* secondary pharmacodynamic analyses into early drug discovery is therefore relatively simple as capacity and turnaround time are likely to be equivalent to primary pharmacologic assays. No prior knowledge of target class or chemical structure is required in the experimental design, although such knowledge can be used to enhance the power of the design. Assay panel design is discussed further in Section 4.2.

The key limitation of the pharmacological profile approach is in the translation of simple quantitative pharmacological data (affinity or efficacy values) from *in vitro* systems to whole organism physiology, although one method to address this is discussed in Section 4.4.

It should be noted that the *in vitro* and *in vivo* approaches discussed are not mutually exclusive and may indeed be complementary: *in vivo* experiments may be carefully designed around the results of *in vitro* panel screening to investigate the effect of any significant activity in an intact physiological system; *in vitro* panels may be employed to deconvolute an observed physiological response in order to identify the underlying molecular target(s) involved.

### 3.3 Strategies for *In Vitro* Pharmacological Profiling

***Stepwise Approach*** Pharmacological profiling may be applied in a stepwise, or tiered, process, where compounds are assayed for activity at an increasing number of molecular targets as the chemical series progresses through the drug discovery and development process (Fig. 1). The screens applied in the early stages of discovery consist of a small panel of key targets that mediate functional effects of the vital physiological systems: cardiovascular, central nervous, and respiratory systems. This early screening may aid selection between multiple lead series. During the optimization phases of discovery, the profile is expanded to a greater number of assays, adding targets that mediate effects on other organs systems such as the gastrointestinal system, enabling refinement of compound selectivity alongside primary target affinity/efficacy. As the drug evolves into a candidate for the development phase, a greater number of targets covering all organ systems are screened. The result is the generation of a comprehensive pharmacological profile of the drug before it is first administered to humans.

The assay methodologies applied are firstly to perform *in vitro* radioligand binding assays and enzyme activity assays, where possible using recombinant human targets. The drug is initially screened at an appropriate single concentration and any

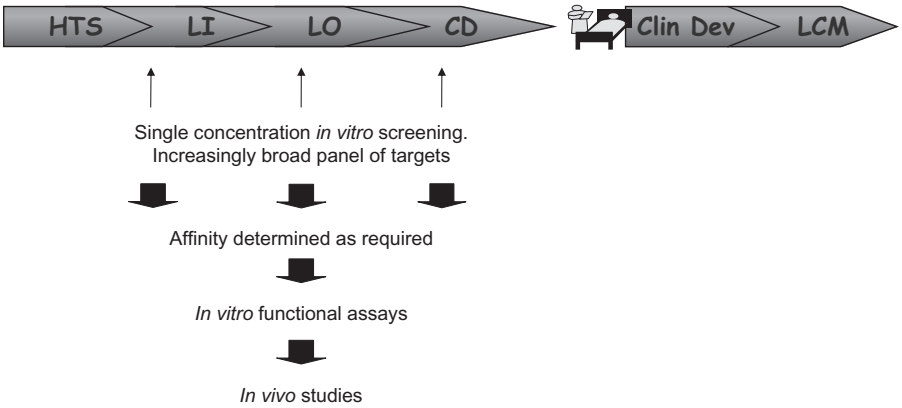


**FIGURE 1** Stepwise pharmacological profiling in drug discovery. Compounds are assayed in single concentration screens for activity at an increasing number of molecular targets as the chemical series progresses through the drug discovery and development process. In the early stages of discovery, the profiling panel may be small and focussed toward those targets of greatest impact, allowing fast and efficient generation of data, perhaps to select between lead series. As compounds progress, the number of secondary targets screened increases to broaden the biological space searched for potential undesirable effects. At any stage, identification of significant activity at a secondary target may trigger determination of affinity in order to estimate selectivity compared with the primary target. Further investigation for targets of concern may be determination of functional activity: agonism or antagonism, which may dictate the likely effect of the compound *in vivo*. Finally, specific *in vivo* studies may be necessary to fully explore the activity of the compound at secondary targets. (HTS = high throughput screening; LI = lead identification; LO = lead optimization; CD = candidate drug; Clin Dev = clinical development; LCM = life cycle management)

targets that show significant activity are retested in concentration–response studies to provide a quantitative analysis of the activity (Fig. 1). The test concentration selected may be based on the affinity of the drug at the therapeutic target, and it is recommended that the test concentration should be a multiple of the primary therapeutic target affinity or efficacy value, in order to use the data to assess the selectivity of the compound for the primary target. Alternatively, a therapeutic margin (the ratio of the dose at which desired therapeutic activity occurs to the dose at which unacceptable adverse events occur) may be considered by comparing the affinity of the compound for secondary targets with the measured (or predicted) maximum exposure *in vivo*. For example, a compound in the early phases of discovery may be tested at concentration 10-fold higher than its affinity for the primary target, but for compounds in later phases it may be appropriate to test at 100-fold higher than the primary target potency or affinity.

Once the activity of the drug at the secondary molecular target has been quantified in a radioligand binding assay or enzyme activity assay, the mode of action may be explored in *in vitro* functional assays to determine whether the drug is an agonist or antagonist at that target. Functional effects in relation to dose and plasma exposure of the drug may then be explored in specifically designed *in vivo* studies (Fig. 1).

***Reverse Sequential Approach*** Some pharmaceutical companies have initiated intensive profiling programs to define chemical space that is most suitable for the use of discovering lead compounds against important target classes [22, 23]. This approach requires the ability to test large series of molecules in large panels of assays at low cost in order to build a dataset correlating diverse areas of chemical space with pharmacological profiles. This database will form the foundation for new drug discovery programs, allowing focused libraries of compounds likely to have affinity for the target of interest, and limited secondary target affinity, to be designed *in silico*. These libraries may be screened by traditional high throughput methods, resulting in high quality lead compounds. Lead optimization, screening compounds for both primary target efficacy and secondary profile, will cycle until a suitable candidate drug can be selected, but this process should not operate in isolation: data generated should be added to the database, and analysis and predictions from alignment of lead compounds with data already in the database should allow rapid optimization of the compound. This reverse sequential drug discovery process is described in Fig. 2.

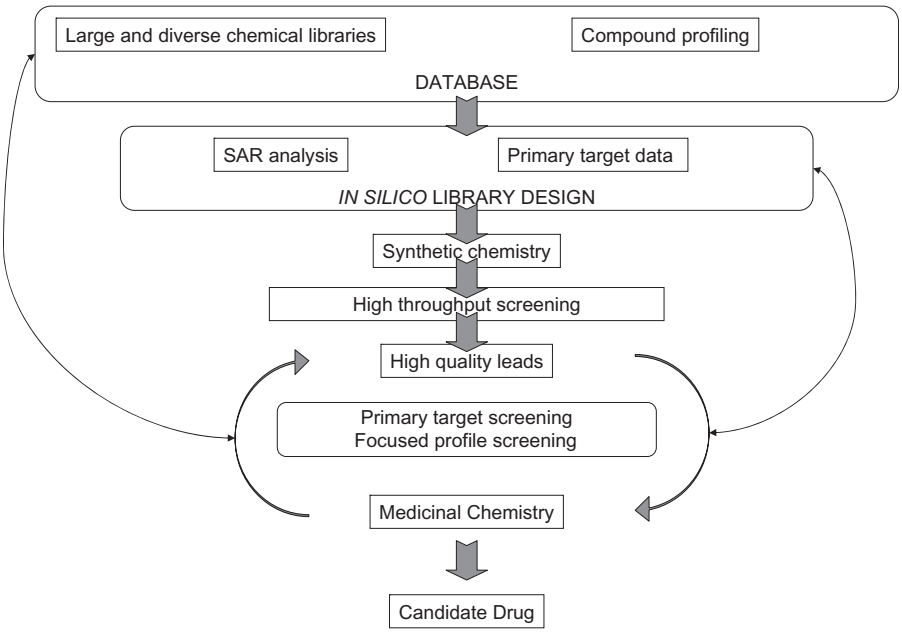


**FIGURE 2** Reverse sequential drug discovery. A database containing pharmacological profiles for large and diverse chemical libraries is generated. *In silico* analysis of these data allow prediction of structures likely to have affinity at the primary target of interest, and limited off-target affinities. Small, focused libraries based on these analyses may be synthesized and screened to generate high quality leads. Optimization of these leads for primary target efficacy and lack of secondary target effects along with other key properties, including ADME and PK, will cycle until a suitable candidate drug may be selected. Data generated in this process should be captured in the original database to enhance its power for future use.

***Problem Solving*** *In vitro* profiles of compounds may be generated in an attempt to find the molecular mechanism behind observed *in vivo* effects of a new drug substance. These observations may come from animal models used in early efficacy testing, toxicological studies, or core battery safety pharmacology studies; or they may be observations from clinical trials of a candidate drug. A broad spectrum profile may be generated in order to attempt to find a molecular target responsible for some physiological effect, or perhaps more focused screening of key molecular targets involved in specific physiological systems may be employed to determine rapidly whether a particular pathway is involved in the observed physiology. An example of this approach is given in Section 5.2.

## 4 TECHNICAL CONSIDERATIONS FOR *IN VITRO* PROFILING

### 4.1 Assay Technologies

A wide variety of assay technologies are available for compound profiling. These range from assays using highly purified single protein reagents, to subcellular fractions, to whole cells, to tissues, and finally to animals. Within each level there are a variety of detection techniques available. Here we focus on the underlying assay parameters rather than the chemistry, physics, and engineering behind various signal detection systems.

Most profiling is done in a tiered fashion with the first tier consisting primarily of assays using highly purified protein targets or subcellular fractions. Compounds with activity in these primary assays are then progressed to more physiologically relevant functional assays. Primary assays generally measure competition between the test compound and either a receptor–ligand or enzyme–substrate interaction. These types of assays have the advantage of using reagents that can be well characterized and easily produced. For this reason, these biochemical assays generally give consistent results over time and are suitable for the rapid testing of large numbers of compounds.

It is important to recognize that the interaction of test compounds with the assay method and technology will have a profound effect on the results. Compounds must be soluble in the assay buffer at all test concentrations. Test compounds that either fluoresce or have strong absorbance can interfere with optical detection systems. Compounds with high protein binding will give variable results depending on the assay protein content. Many test compounds show nonspecific binding to plastics used in pipette tips, assay plates, and filtration media; thus, the actual concentration available to interact with the target may be substantially lower than the theoretical test concentration. Changing consumables or taking steps to block nonspecific binding will impact assay results. Test results for lipophilic compounds will vary with amount of membrane fraction in receptor binding and cytochrome P450 assays. These types of physicochemical interactions can generally be anticipated based on a prior understanding of the nature of the test compounds (solubility, optical properties, protein binding, lipophilicity).

Some assay technologies are optimized for throughput, good signal-to-noise ratio, and signal stability in ways that severely constrain the concentration or type of key assay components. To optimize assay robustness and signal-to-noise ratio, high affin-

ity antagonists will often be chosen as the radioligand in receptor binding assays. In some cases, these ligands are not effective for detecting agonists and important interactions may be missed. There are many kinase assay technologies that require substrate or cofactor concentrations well above $K_m$ and/or require the use of substrates with no physiological relevance. Many assays are run using receptors expressed at high density, which may exist in abnormal states and thus exhibit aberrant pharmacology. Many assays use the active site fragment of an enzyme or receptor, which may yield results different from those obtained using the entire protein. In many pharmacology profiles, ion channel targets are tested in binding rather than functional assays; these binding assays are known to miss some channel-blocking compounds. There are a number of cases of assay format variations leading to significant and consistent changes in test results that are more difficult to anticipate.

In many cases, both human and animal forms of important targets are available. Not surprisingly, there are examples of species-specific differences in the affinities for some compounds [24]. For this reason, there has been a great effort to introduce human forms of the target of interest into all profiles. There are, however, good reasons to retain assays based on relevant animal-derived targets, as most models for *in vivo* safety and efficacy are in animals. It is important to anticipate these differential results in animal models and expected clinical outcome. There are also several examples of important targets, where the human form is not yet cloned or is unavailable due to blocking patents. Over time, most early pharmacology profiles have incorporated, whenever available, the human form of important targets with animal targets primarily used in focused studies to support a full understanding of *in vivo* animal models.

The difficulties outlined in the preceding paragraphs highlight the importance of informed assay design and a full understanding of the relationship between an *in vitro* assay result and the *in vivo* outcome that it is designed to predict. This understanding can come only from full characterization of each assay using as many known effectors as possible. For example, if a cardiovascular safety profile contains a group of adrenergic receptor subtype assays, it is important to know how all reference compounds with well characterized *in vivo* outcomes perform across the panel. These reference compounds would include marketed drugs, compounds that have been characterized in animal studies, and include a selection of agonists and antagonists. Only with this data as context can there be an informed interpretation of data from new test compounds.

Once a compound has been shown to have significant activity in an *in vitro* assay, the next step or tier of assays is designed to define the likely functional outcome of the compound–target interaction. For enzyme targets, this would normally be definition of the mechanism (competitive, noncompetitive, uncompetitive). For receptor targets, this would normally be differentiation between agonists, partial agonists, antagonists, and inverse agonists. For transporters, this could be differentiation between substrates and inhibitors. In some cases, these assays can be run using the same target reagent as the initial screening assay. However, in many cases (primarily receptors, ion channels, and transporters), the relevant experiments can only be done in living cells or isolated organs. These more biologically complex assays open up a new area of difficulty in assay design and data interpretation. Some of these difficulties can be addressed by a good understanding of the physicochemical prop-

erties of each test compound: Is it likely to reach an intracellular target? Is it soluble at the highest test concentration? Will results be influenced by serum proteins? Some of the difficulties must be addressed by having a good understanding of the additional factors in cells and tissues that will influence assay results. For example, is the signal transduction between a highly expressed receptor and the measured signal a good representation of normal physiology? Can interactions between the test compound and endogenous nontarget proteins in the cell line or tissue influence the assay results? Thus, once again, it is essential to fully characterize each assay by having tested all known effectors. This reference dataset is the context in which data on new compounds can be fully understood.

The third tier of testing often involves animal experiments to explore fully the functional effect of an activity discovered in the early tiers of profiling. These models are based on specific organ functions known to be associated with the earlier observed *in vitro* assay results. Testing at this level may be the first time a compound has been used *in vivo* and brings up questions of formulation, dosage, route of administration, and pharmacokinetics. Consideration of these factors is beyond the scope of this chapter; however, once again, it is key to have a strong set of reference compound data in each model, which can be used as the context in which to interpret data on new compounds.

In conclusion, regardless of the particular assay technology chosen, profiling assays should meet two basic requirements. First, the relationship between the assay result and one or more clinical outcomes should be well understood; and second, the assay should produce consistent results over time.

### 4.2 *In Vitro* Panel Design

***Introduction*** *In vitro* assay panels can be designed using two major strategies. Assays can be selected one-by-one based on known safety issues associated with those assays [25] or the assay panel can be chosen on the basis of assay diversity as determined by a variety of different methods [22]. In reality, most assay panels are designed using a combination of these two strategies. With greater technological advancement, a large number of cell lines and expression systems have been engineered to allow development of assays using human targets. As a result, a wide range of *in vitro* pharmacology assays have been developed to act as models for predicting *in vivo* effects. With such a great diversity of assays available to choose from, efforts have been made to determine the most clinically important side effects and to link those side effects with assays [26].

Growing experience with profiling has facilitated the selection of assay panels. As pharmaceutical companies screen an increasing number of drug candidates, this experience, in the form of data, can be applied to better understand the assays and to fine-tune the assay selections. In-house, public, and commercial profiling databases offer unique opportunities to analyze the relationships between targets as well as to access the profiles of drugs and drug-like molecules (e.g., see Refs. 22 and 23). Standard statistical methods such as clustering and correlation can be applied to these datasets to elucidate the sometimes counterintuitive relationships between assays. In this section, the application of the knowledge from long-running screening programs to assay panel design is discussed.

***Assay-by-Assay Choice: Beyond Known Pharmacology*** The most intuitive methods of assay selection are based on known pharmacology and clinical experience. Experience from long-running screening programs and commercial drug information databases can be applied to enhance these methods of selection. For example, associations between assays and adverse drug reactions (ADRs) can be generated using a database of *in vitro* pharmacology data for drugs and the accompanying *in vivo* ADR data. In this example, uniformly generated *in vitro* pharmacology data is coupled with *in vivo* ADR data gathered from drug labels and mined for significant correlations between the *in vitro* pharmacology and *in vivo* clinical effects [26]. In this analysis, the natural language of drug label ADRs were translated into COSTART (Coding Symbols for Thesaurus of Adverse Reaction Terms, U.S. Food and Drug Administration) ADR terms and loaded into a database. Each COSTART ADR term was compared with each pharmacology assay to identify significant correlations. The hits in a given assay were binned by strength, and among the drugs in a given bin, the percentage of those drugs that report each of the 800 ADR terms was determined. Assay bin/ADR pairs were flagged when the percentage of compounds in that bin that list the ADR on the label was significantly higher than the baseline level for that ADR across all the drugs in the dataset. Further analysis looked for an *in vitro* activity–ADR dose response across the hit strength bins for a given assay or target. In other words, as the *in vitro* potency against a target goes up, the ADR should be increasingly present on the labels of the drugs in those bins.

An example of the results of this analysis can be seen in Fig. 3. This figure illustrates the association of mydriasis with the $M_3$ muscarinic receptor. The baseline frequency of this ADR is 5.99%, while nearly half of the drugs with $M_3$ hits in the <100 nM $IC_{50}$ bin list this ADR on the drug label.

Over 5000 significant *in vitro* assay–ADR associations were found covering over 100 assays/targets. Close examination of this set of associations yields many results

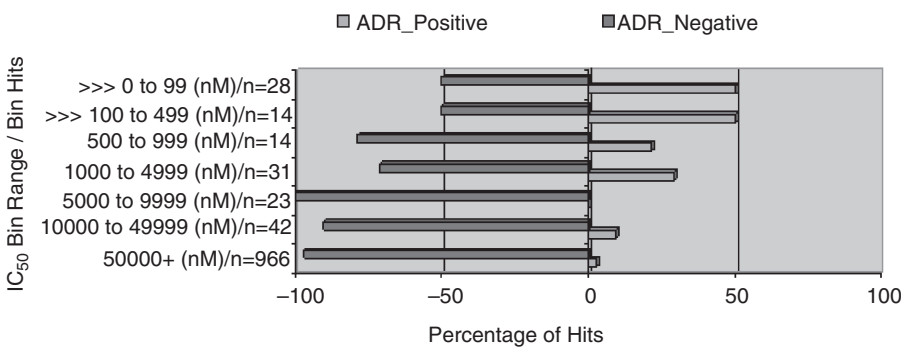


**FIGURE 3** Significant statistical association between in vitro binding to the $M_3$ muscarinic receptor and clinical mydriasis. Seven $IC_{50}$ activity bins are found on the y-axis ranging from the 0-99 nM bin to the greater-than-50000 nM bin which is the default bin for the non-hits. The number of compounds in each bin is indicated. The bars shown with each bin represent the percentage of compounds in that bin that are positive (light bars on the right) or negative (dark bars on the left) respectively for the ADR in question. The baseline percentage for that ADR is 5.99%. The result of a Spearman rank correlation of this association is a robust 0.82. Data from BioPrint (Cerep SA).

that are consistent with known pharmacology and many others that appear to be novel. Another useful feature of these associations is that they aid in interpretation of *in vitro* data because one can identify the hit strength bins where the association becomes significant. For instance, in the case of $M_3$ and mydriasis, the association only becomes of concern in the 1000–4999 nM bin.

***Diversity: Beyond Phylogenetics*** Phylogenetic or pharmacological diversity can be applied to the selection of assays for profiles. Phylogenetic and pharmacological descriptions of how the members of a receptor or enzyme family are related are not necessarily in agreement. Here too, the availability of large *in vitro* datasets is useful, allowing analysis of interassay overlap, correlation, and clustering.

The muscarinic acetylcholine receptor can act as an interesting test case for assay selection. Among the five different muscarinic receptor subtypes, which and how many of these receptors should be included in a profile? While all five may eventually be tested, which ones might prove most informative for the earliest testing? Several different kinds of information can be brought to bear in making this evaluation. While clustering compounds by an *in vitro* profile is now a common practice, there is less recognition of the utility of using those same datasets to cluster the assays. In the case of compound clustering, the assay data is used to probe the relationships of the compounds to one another. Similarly, the compounds can be used as probes to define how the *in vitro* assays relate to one another. From the clustering of receptors by pharmacological activity, it can be seen that the $M_1$ and $M_2$ are most closely related to one another followed by the pair of $M_3$ and $M_4$ with $M_5$ most distantly related to both pairs (Fig. 4).

Assuming that it is desirable to cover muscarinic pharmacology as completely as possible without having to run all five assays, it would follow that one might choose to run three assays, one chosen from each pair of related assays and $M_5$. To help choose between these pairs, the overlap of hits on each of these assays can be examined (Fig. 5) as well as the number of ADR associations that have been found for each of these targets (Table 3). The overlap matrix shows what proportion of the hits (defined as >30% inhibition at 10 μM) on the target defining each column are also hits on the target defining each row. For instance, looking at the first row in Fig. 5, 88% of the hits at $M_2$ are also hits at $M_1$, 96% of the hits on $M_3$ are also hits at $M_1$, 85% of hits at $M_4$ are also hits at $M_1$, and 86% of hits at $M_5$ are also hits at $M_1$.

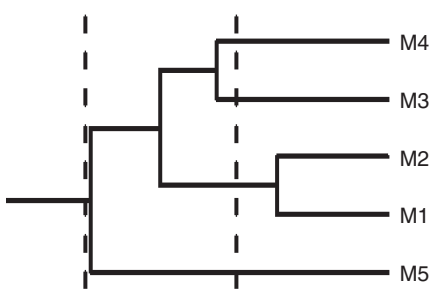


**FIGURE 4** Clustering by Pearson correlation using the complete linkage. A dataset of nearly 2500 molecules tested on all five subtypes was used to define the relationship between muscarinic acetylcholine receptor targets. Data from BioPrint (Cerep SA).

| | M1 | M2 | M3 | M4 | M5 |
|---|---|---|---|---|---|
| M1 | 1 | 0.88 | 0.96 | 0.85 | 0.86 |
| M2 | 0.88 | 1 | 0.96 | 0.85 | 0.84 |
| M3 | 0.96 | 0.96 | 1 | 0.71 | 0.77 |
| M4 | 0.85 | 0.85 | 0.71 | 1 | 0.78 |
| M5 | 0.86 | 0.84 | 0.77 | 0.78 | 1 |

**FIGURE 5** Overlap matrix for the five muscarinic acetylcholine receptors. The matrix shows the overlap frequency between hits on the individual mAChRs. For example, 62% of hits on $M_1$ are also hits on $M_3$. Conversely, 96% of hits on $M_3$ are also hits on $M_1$. Data from BioPrint (Cerep SA).

**TABLE 3 Distribution of ADR Association Across the mAChRs**[a]

| Assays | Number of ADR Associations that Are Unique to the Assay or Group of Assays |
|---|---|
| $M_1$ alone | 24 |
| $M_5$ alone | 20 |
| $M_4$ alone | 4 |
| $M_1$ and $M_4$ | 9 |
| $M_1$ and $M_2$ | 4 |
| $M_1$, $M_2$, and $M_3$ | 2 |
| $M_1$, $M_2$, $M_3$, and $M_4$ | 36 |
| $M_1$, $M_2$, and $M_4$ | 1 |
| $M_1$, $M_3$, and $M_4$ | 1 |

[a]For example, there are 24 ADR associations unique to $M_1$ and 9 associations shared by $M_1$ and $M_4$. Data from BioPrint (Cerep SA).

Table 3 shows the number of ADR associations found for each receptor subtype. As the five subtypes are highly correlated, they share many of the same ADR associations; for example, there are 36 ADR associations common to $M_1$, $M_2$, $M_3$, and $M_4$. Based on the information presented in the table, $M_1$ should be chosen over $M_2$ because although the two are highly correlated, $M_1$ has the greater overlap with other subtypes and has many unique ADR associations while $M_2$ has no unique associations. $M_3$ and $M_4$ cluster together even though $M_3$ is more highly correlated with $M_1$. Most of the hits at $M_3$ (96%) are also hits at $M_1$. $M_4$ is the better choice here because it is less correlated and has less overlap with $M_1$ and $M_2$. $M_4$ also has more unique ADR associations than $M_3$. $M_5$ should be included because it differs most from the other family members by correlation, clustering, and overlap and has a large number of ADR associations uniquely associated with it. Based on this analysis, the best assays to include would be $M_1$, $M_4$, and $M_5$.

It is important to recognize that this kind of analysis is limited by the dataset used to generate associations. For example, there are currently very few muscarinic

receptor subtype-selective compounds on the market, which may explain some of the overlap and correlations described here. It is important, therefore, that the analyses used to define profiling panels are reviewed regularly, especially as novel compounds are profiled, in order to ensure that the relationships still hold. Likewise, new scientific knowledge of targets and effects should be incorporated in order to maximize the value of the profiling panel.

### 4.3 ADME/PK Profiling

***Introduction*** While the pharmacological screening of development compounds has been widely accepted by the pharmaceutical industry, *in vitro* ADME/PK screening has been adopted more slowly. Nevertheless, ADME/PK profiling is now widely performed at early stages in lead development. Selection of leads with favorable *in vitro* ADME properties is important because it is much harder to design them in after a compound with the appropriate pharmacology has been identified. *In vitro* ADME assay panels are chosen for their potential to predict *in vivo* pharmacokinetics. Some of the *in vitro* assays measure properties that contribute to the *in vivo* bioavailability of the new drug candidate. These include aqueous solubility, log *D*, and physicochemical properties known to play a role in drug absorption. More direct assays such as immobilized artificial membrane (IAM) columns and parallel artificial membrane permeation assay (PAMPA) are also gaining popularity as measures of passive permeability. Results from these types of assay are useful by themselves but interpretation of these data is enhanced by the use of large *in vitro* and *in vivo* datasets that act as a context in which to understand these *in vitro* results.

In Fig. 6, a histogram shows the distribution of the log *D* values for compounds with poor (<20%) and good (>80%) oral absorption. The distributions are overlapping but offset and as such log *D* alone cannot be used to make a clear determination of potential permeability. Nevertheless, information like this can be taken into

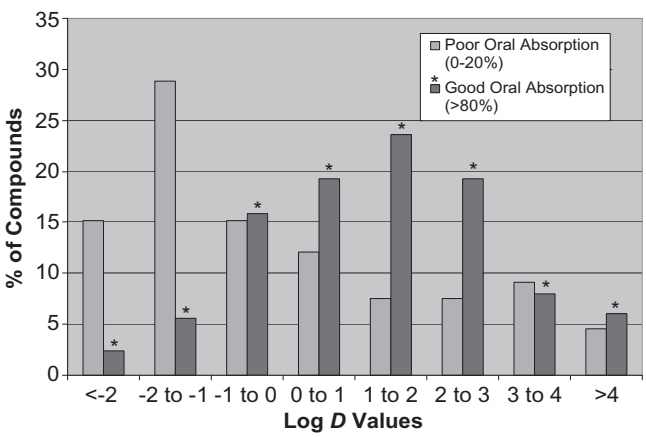


**FIGURE 6** Comparison of the distribution of log *D* values for compounds with good (>80%, bars with asterisk) and poor (<20%, plain bars) oral absorption. Data from BioPrint (Cerep SA).

account for library design by trying to target compound libraries to have log *D* values that fall into the range preferred for good oral absorption. In-house or commercially available software can be used for *in silico* log *D* or log *P* predictions for potential libraries. Early screening of drug candidates can allow selection of those with the best possible *in vitro* ADME properties.

Permeability assays using cell lines like Caco-2, MDCK, or HT-29 may be employed to make permeability determinations that are able to take into account the diverse transporters and efflux pumps that play a significant role in drug permeability *in vivo*. In the case of the Caco-2 assay, it is most informative to run the assay in both the apical (A) to basolateral (B) and the basolateral to apical directions in order to identify compounds that might be differentially transported in one direction versus another. The P-glycoprotein (P-gp) inhibition (of the B-to-A flux of P-gp transporter substrate digoxin) and P-gp ATPase assays are used further to identify whether a compound is likely to be effluxed *in vivo*. Figure 7 plots percent oral absorption versus Caco-2 cell permeability data. This view of the data shows that while the general trends are as expected—drugs with low permeability are poorly absorbed and drugs with high permeability are well absorbed—this is not the complete answer. As a rule of thumb, an apparent permeability value ($P_{app}$) of 5 or greater indicates that a compound is not likely to be permeability-limited; however, there are clearly many exceptions.

*In vitro* metabolic stability assays using liver microsome preparations and isolated cytochrome P450s allow a qualitative assessment of the likely metabolic fate of the test compound. Figure 8 shows that there is no direct correlation between *in vitro* metabolic stability and oral bioavailability. However, it is worth noting that there are very few compounds having both poor metabolic stability (<20% remaining) and high oral bioavailability (>80).

***Drug Interactions at Cytochrome P450*** Panels of cytochrome P450 inhibition assays can aid in identifying compounds that might be susceptible to P450-mediated

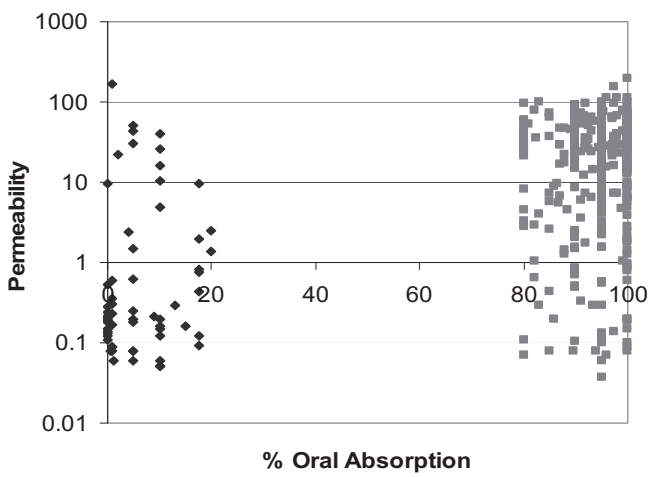


**FIGURE 7** Comparison of the distribution of apical to basolateral apparent permeability values ($P_{app}$ $1 \times 10^{-6}$ cm/s) for compounds with good (>80%) and poor (<20%) oral absorption. Data from BioPrint (Cerep SA).

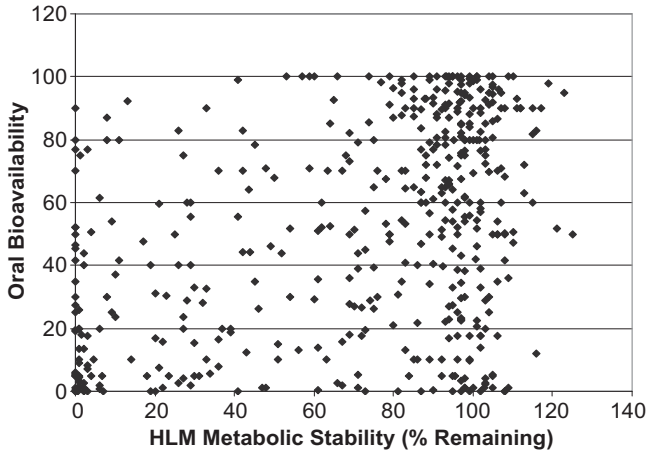


**FIGURE 8** Plot of human liver microsome metabolic stability (% remaining) versus human oral bioavailability (%); $n$ = 468 compounds. Data from BioPrint (Cerep SA).

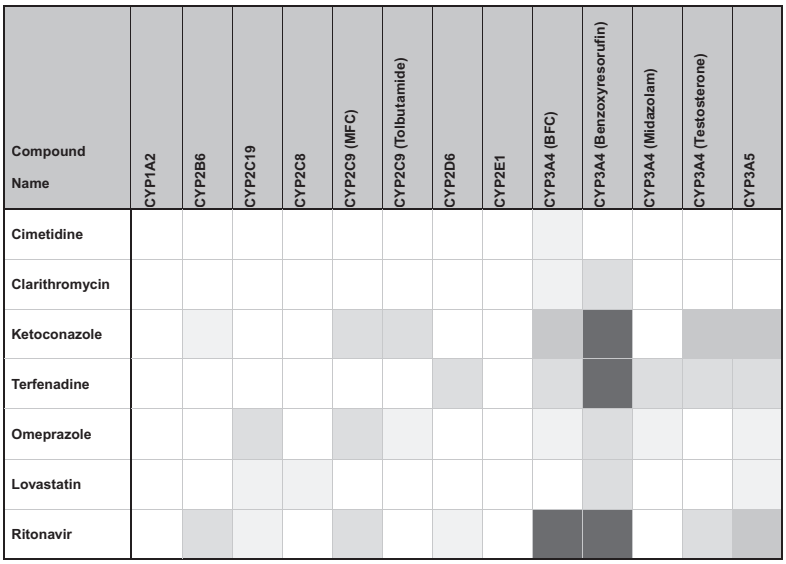


**FIGURE 9** Heat map of P450 profiles of a selection of compounds known to be involved in clinically relevant drug–drug interactions: individual drug data is shown in the rows, while the assays are named at the top of the columns. Data from BioPrint (Cerep SA).

drug–drug interactions. Figure 9 shows the P450 interactions of a selection of drugs known to be vulnerable to drug–drug interactions as either substrates or inhibitors. For instance, lovastatin is a CYP3A4 substrate susceptible to interactions with potent CYP3A4 inhibitors such as ketoconazole [27]. Terfenadine is metabolized by CYP3A4 [28] and has been involved in adverse interactions with inhibitors such as ritonavir.

***Use of an ADME Profile*** All together, these assays can be used to create an ADME profile that is useful both as individual results and as a collective ADME profile. The use of the collective profile may help to overcome the limitations of the individual assays. Very low solubility as determined in the solubility assay might explain why compounds like albendazole or niclosamide that have good *in vitro* permeability characteristics are poorly absorbed. Standard statistical methods can be used to identify which marketed drugs have ADME profiles similar to that of a test compound of interest. The well understood clinical PK properties of these known compounds can be used to help predict the *in vivo* properties of the clinical candidate. Figure 10 shows the *in vitro* ADME profiles for two sets of drugs and some corresponding pharmacokinetic data. Both sets of drugs share similar results in the four permeability assays, but differ in metabolic stability and log *D* values. Both sets of compounds are generally well absorbed orally, but the set with low metabolic stability understandably has low oral bioavailability. When comparing the *in vitro* ADME profile of a drug candidate to those of known drugs, one can have greater confidence in extrapolating from the neighbors (compounds with similar *in vitro* ADME profiles) when most of the neighbors have homogeneous *in vivo* ADME properties. If the neighbors have very diverse properties, there is little basis for extrapolation.

## 4.4 Data Analysis and Interpretation

***Introduction*** Can *in vitro* pharmacological profiling capture information relevant to drug biology and drug development? The heat map in Fig. 11 shows the profiles of tricyclic antidepressants versus the newer second and third generation antidepressants. The antidepressants are divided into three tiers in the figure. In the upper tier are the older tricyclic antidepressants, the second tier are the popular second generation antidepressants, and some of the newer antidepressants are found in the

| Compound Name | Permeability A to B pH 7.4/7.4 (Papp 10e-6 cm/s) | Permeability B to A pH 7.4/7.4 (Papp 10e-6 cm/s) | Permeability A to B pH 6.5/7.4 (Papp 10e-6 cm/s) | Permeability B to A pH 7.4/6.5 (Papp 10e-6 cm/s) | Aqueous Solubility (10e-6 M) | HLM Metabolic Stability (% Remaining) | LogD (octanol) | LogD (Cyclohexane) | Oral Bioavailibility (%) | Oral Absorption (%) |
|---|---|---|---|---|---|---|---|---|---|---|
| Nefazodone | 96 | 19 | 36 | 15 | 72 | 0 | 3.53 | 1.68 | 19 | 100 |
| Lovastatin | 14 | 7 | 26 | 10 | 13 | 0 | 4.08 | 1.28 | 5 | 10 |
| Ergotamine l-Tartrate | 20 | 17 | 26 | 23 | 41 | 1 | 3.57 | 0.71 | 4 | 55 |
| Astemizole | 14 | 9 | 26 | 16 | 42 | 3 | 4.05 | 2.1 | 7 | 90 |
| Benzarone | 38 | 8 | 36 | 8 | 31 | 6 | 4.04 | 1.1 | 1 | 73 |
| Sildenafil | 42 | 23 | 35 | 41 | 32 | 19 | 2.8 | 0.65 | 40 | 92 |
| | | | | | | | | | | |
| Quinidine | 43 | 18 | 25 | 51 | 200 | 69 | 1.96 | -1.5 | 74 | 80 |
| Yohimbine | 43 | 21 | 31 | 47 | 196 | 69 | 1.98 | na | 28 | |
| Clozapine | 32 | 14 | 26 | 31 | 203 | 78 | 2.98 | 1.23 | 53 | 93 |
| Amitriptyline | 25 | 78 | 26 | 51 | 175 | 79 | 2.77 | 1.7 | 48 | 100 |
| Tramadol | 56 | 19 | 38 | 44 | 200 | 90 | 1.24 | 1.12 | 77 | 75 |
| Pantoprazole | 31 | 28 | 35 | 23 | 166 | 91 | 1.93 | -1.1 | 79 | |

**FIGURE 10** Compounds with similar *in vitro* ADME profiles and corresponding *in vivo* pharmacokinetics. Data from BioPrint (Cerep SA).

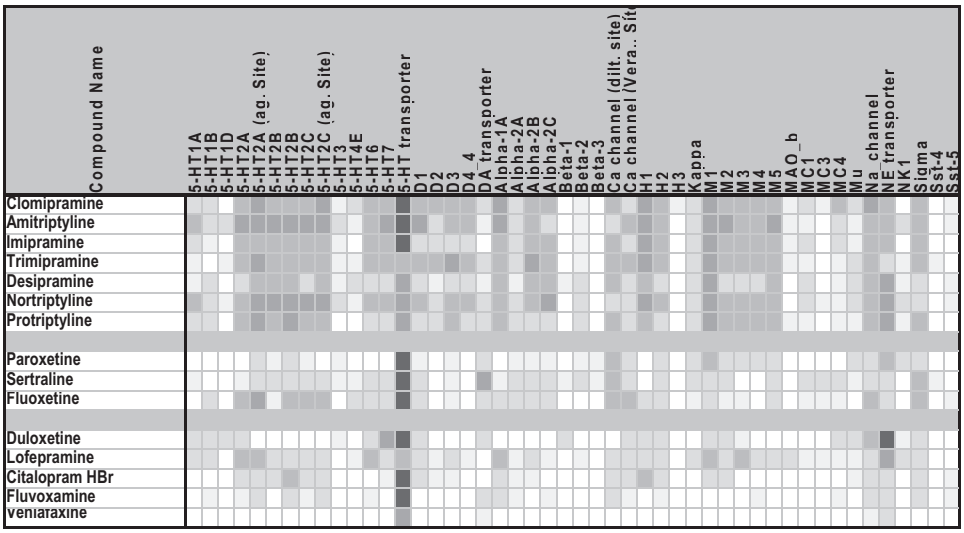


**FIGURE 11** Heat map of antidepressant profiles. In this heat map, the individual drugs are shown in the rows, while the assays are named at the top of the columns. Data from BioPrint (Cerep SA).

third tier. Nearly all the drugs share strong interactions with the 5-HT transporter. Further visual inspection shows that the tricyclics also share hits on the muscarinic, histamine, and adrenergic targets, all of which are known to be associated with the commonly reported side effects of these drugs. Each succeeding generation of drugs has fewer and weaker hits on these targets. Thus, in this case, profiling clearly captures the differences between these generations of drugs.

The profile similarities shared by the many antidepressants make great sense both from the standpoints of shared therapeutic targets such as the serotonin and norepinephrine transporters and of the shared muscarinic and histaminic adverse effects. Nevertheless, it is not necessary that a drug's target class be included in the profile for that class to cluster together. An example of this is shown in Fig. 12, where a large group of the "conazole" family of antifungals are found to cluster together despite the absence of the therapeutic target in the assay panel used to cluster the compounds. These drugs are clustering together based on rich profiles of weak hits.

***Profile Interpretation on a Hit-by-Hit Basis*** The most obvious way to interpret profile data is on a hit-by-hit basis. When following up on hits, the first consideration is the strength of the off-target effect relative to the therapeutic target. The hit strength and agonist/antagonist status can be determined by $IC_{50}$ or $K_i$ and functional assay follow-up. Having obtained an $IC_{50}$ for a hit in question, the value may be compared with available plasma exposure data or with the projected therapeutic concentration of the drug candidate. If there is good (>100-fold) separation between these values, the hit may be of little concern.

Another line of investigation involves exploring the ADR associations discussed previously in Section 4.2. ADR associations for the assay of interest can be retrieved

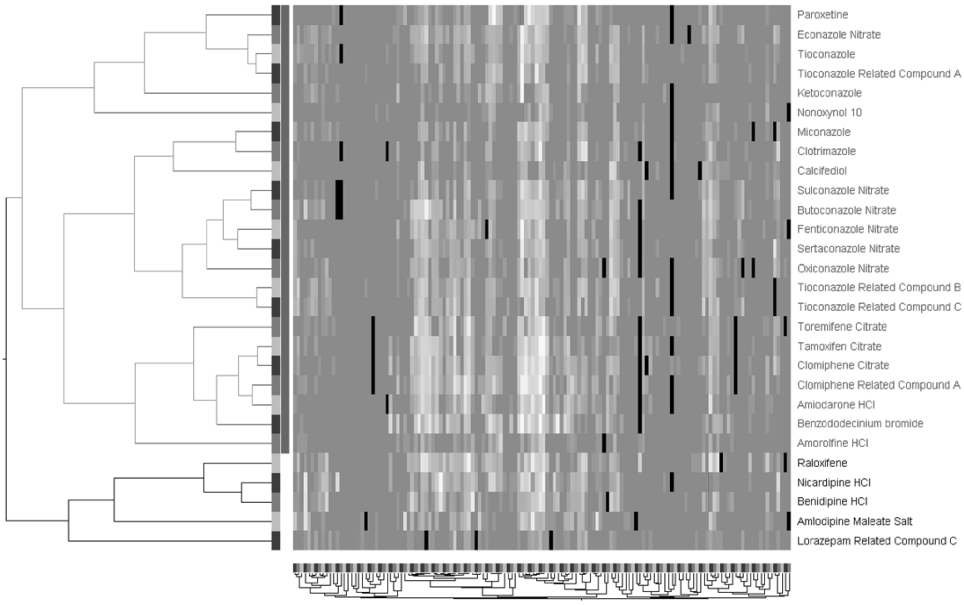


**FIGURE 12** Clustered heat map. Individual drugs are clustered on the *y*-axis while the pharmacology assays are clustered on the *x*-axis. Clustering of both the compounds and the assays is performed using the conventional Pearson correlation and complete linkage. This figure only shows a subcluster of an original clustering of over 2000 compounds. Data from BioPrint (Cerep SA).

and examined. The ADR associations only become significant at a certain hit level ($IC_{50}$ bin). This information can be used to estimate a hit strength threshold at which the risk of a given ADR becomes significant from the clinical point of view. These data can be considered in the context of other compounds, in particular, marketed pharmaceuticals that have similar levels of activity at the target of interest. It can be considered whether or not the ADRs associated with the target of interest proved to be a liability for these marketed drugs.

While binding assays make up a large part of a pharmacology safety profile, it is also important to be aware of the limitations of these assays. First, radioligands can be either agonist or antagonist and the choice can influence the ability to detect agonists. For instance, the neurotransmitters serotonin, dopamine, and norepinephrine have broad ranges of potency on the diverse members of their own and related receptor families (see Fig. 13). Second, binding assays for ion channels and enzymes using a single ligand are not able to assess fully the many ways a test compound might interact with the normal functioning of the native protein. These binding assays only monitor drug interactions at a specific site.

***Using the Entire Profile*** The previous section discussed profile analysis based on individual assay results. Using standard statistical clustering techniques, compounds with profiles similar to the compound of interest can be identified. Examples of this have been provided, where even profiles of weak hits not including the therapeutic target cause a group of antifungal drugs to cluster together (Fig. 12). The grouping

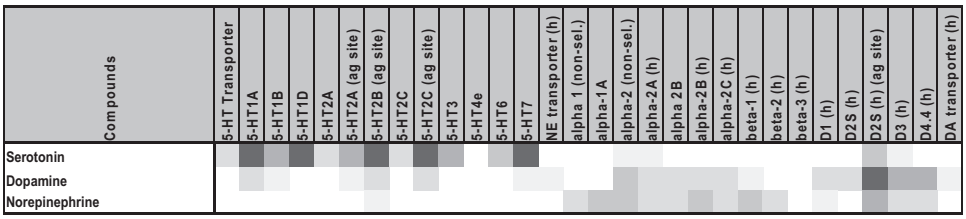


**FIGURE 13** Binding assay data for serotonin, dopamine, and norepinephrine across a panel of aminergic targets. In the case of serotonin, the potency of the endogenous ligand varies greatly ranging from no hit on 5-$HT_{4E}$ to a nM hit on the 5-$HT_{2C}$ agonist site assay. Data from BioPrint (Cerep SA)

of the *in vitro* pharmacological profiles of the antidepressants also makes the point that compounds with similar profiles have similar side effects (Fig. 11). The *in vivo* properties (ADRs) of the compounds with similar profiles can be tabulated. Investigation of the ADRs common to the group of similar compounds may help in predicting the potential of ADRs for the profiled compound. It is also worth noting that this technique allows identification of ADRs that might be associated with the drug class but which are not captured by the individual assay pharmacology.

Thus, the profile data can be interpreted at a variety of levels and interpretation is greatly aided by access to large datasets of *in vitro* and *in vivo* data on compounds with well documented *in vivo* history. Interpretation of individual hits is facilitated by comparison to drugs or benchmark compounds with similar activities. Assay/ADR associations mined from these large datasets can help to assess the risk associated with the off-target hits of a drug candidate. Analysis of the profile as a whole allows one to identify compounds or drugs with similar profiles with the added benefit of introducing drug class ADRs that are not captured by the profile assays or our understanding of the *in vivo* biology from those assays.

## 5 IMPACT OF SECONDARY PHARMACODYNAMIC PROFILING: CASE STUDIES

In this section, three case studies are presented to offer a context for the application of secondary pharmacodynamic studies and *in vitro* pharmacological profiling.

### 5.1 Case Study 1. Ranolazine: Secondary Pharmacodynamic and Nonclinical Safety Pharmacology Studies Supporting Regulatory Approval

Ranolazine (Raxena™) is a new anti-anginal agent approved in the United States for use as combination therapy when chronic stable angina is not adequately controlled with other anti-anginal agents (for review, see Ref. 29). While the exact mechanism of action of ranolazine is not known, its anti-anginal and anti-ischemic effects do not appear to depend on changes in blood pressure or heart rate. Pur-

ported primary mechanisms of action include inhibition of the late sodium ($I_{Na}$) current and partial inhibition of fatty acid oxidation. In patients, ranolazine has been associated with a slight but dose-dependent prolongation of the rate corrected QT interval of the surface electrocardiogram (QTc; mean increases of 4–12 ms at 500–1500 mg dose, which equates to plasma exposures of ~2–6 μM).

It is recognized that some drug-induced sudden deaths are secondary to the development of an arrhythmia called torsades de pointes (TdP). Recent advances in the understanding of this issue indicate that the primary event is likely to be inhibition of the rapid component of the delayed rectifier potassium current ($I_{Kr}$) by such drugs. Since $I_{Kr}$ plays a key role in repolarization of the cardiac action potential, its inhibition slows repolarization and this is manifested on the electrocardiogram as a prolongation of the QT interval. While QT interval prolongation is not a safety concern *per se*, in a small percentage of people it is associated with TdP and degeneration into ventricular fibrillation (for review see Refs. 30 and 31).

Since the cardiac action potential is generated by a complex interplay between several voltage-gated ion channel types and also involves transporters, a series of experiments were carried out to assess the effects of ranolazine on cardiac electrophysiology using electrophysiology-based techniques. In dog myocytes, ranolazine inhibits $I_{Kr}$, late $I_{Na}$, and $I_{Ca}$ at clinically relevant therapeutic concentration ($IC_{50}$ of 12, 6, and 50 μM, respectively) [32]. Ranolazine also blocks the peak $I_{Ca}$ and $I_{Na/Ca}$ exchange although with lower potencies (296 and 91 μM, respectively) [32]. Subsequently, the effects of ranolazine in *in vitro* and *in vivo* repolarization assays were evaluated. In the canine wedge preparation, ranolazine caused a slight prolongation of action potential duration (APD) in epicardial cells [33]. Furthermore, in an atrioventricular block dog model, ranolazine produced modest, nonsignificant prolongation of QT interval [34]. Since ranolazine has the ability to block $I_{Kr}$ and to prolong, although minimally, the APD and the QT interval, its proarrhythmogenic potential was evaluated to refine further its cardiac safety profile. In the canine wedge preparation, ranolazine, in contrast to selective $I_{Kr}$ blockers, causes a shortening of the APD in M cells but a slight prolongation of APD in epicardial cells [33]. This leads to a rate-independent reduction in transmural dispersion of repolarization, a key arrhythmogenic factor. In an atrioventricular block dog model, ranolazine, in contrast to D-sotalol, produces modest, nonsignificant prolongation of QT interval that was not associated with reverse use dependency [33]. Furthermore, ranolazine abolishes sea anemone toxin-induced early after depolarization (EADs) in guinea pig ventricular myocytes and suppresses sotalol-induced EADs in Purkinje fiber preparations; ranolazine *per se* failed to induce EADs in these preparations [35]. Furthermore, in the isolated female rabbit Langendorff heart, ranolazine *per se* did not induce TdP but suppressed TdP-induced by $I_{Kr}$ blockers and $I_{Na}$ activators [33]. Finally, in a dog in an atrioventricular block dog model, ranolazine, in contrast to D-sotalol or epinephrine, failed to induce TdP [34]. Overall, ranolazine failed to demonstrate a proarrhythmic profile and showed some antiarrhythmic properties in nonclinical models.

The overall polypharmacology of ranolazine may explain in part its overall modest ability to prolong the QT interval and its lack of arrhythmogenic potential. The nonclinical experiments described above were instrumental in supporting the approvability by regulatory agencies of ranolazine.

### 5.2 Case Study 2. Muscarinic $M_3$ Receptors and Mydriasis: Linking a Molecular Target to a Physiological Effect

Compound AZ123 was profiled in a panel of 150 *in vitro* radioligand binding and enzyme assays at a single concentration of 10 μM as a routine part of the drug discovery process. Significant activity, defined as greater than 50% inhibition at 10 μM, was detected in one radioligand binding assay, the human recombinant muscarinic $M_3$ receptor, where 75% inhibition of the binding of [$^3$H]-*N*-methylscopolamine to the human muscarinic $M_3$ receptor was detected. No significant activity was detected in any of the other muscarinic receptor radioligand binding assays (human recombinant $M_1$, $M_2$, $M_4$, or $M_5$). A concentration–effect curve was constructed in the human $M_3$ receptor binding assay and the affinity ($K_i$) was determined to be 0.3 μM.

The compound was subsequently tested in the core battery of safety pharmacology studies to explore its effects on physiological functions. The doses selected achieved free plasma levels greater than that predicted to drive efficacy at the primary molecular target. In the rat functional observational battery (FOB) test, used to assess effects on nervous system function, AZ123 caused a dose-dependent increase in pupil diameter (mydriasis) under ambient lighting, an impairment of the pupillary light reflex and a reduction in body weight gain over 24 h. The free plasma level at which the effect was observed was greater than 1 μM. Literature investigations indicated these effects could be linked to antagonist action at the muscarinic $M_3$ receptor, as there is good evidence that reflex contraction of the iris in response to light is under parasympathetic nervous system control [36]. The muscarinic $M_3$ receptor is expressed on the iris sphincter pupillae and parasympathetic innervation of this sphincter regulates pupil diameter to constrict and reduce pupil diameter in response to light [37]. In addition, parasympathetic innervation of the gastrointestinal tract causes contraction of the gastrointestinal tract and this is mediated by acetylcholine action at muscarinic $M_3$ receptors expressed on smooth muscle. Antagonists of the muscarinic $M_3$ receptor inhibit the contraction of gastrointestinal tract smooth muscle and reduce spontaneous gastrointestinal motility and gastric emptying [38]. Muscarinic $M_3$ receptor antagonists are used in the clinic to prevent the gastrointestinal tract peristalsis when performing radiological diagnosis of gastrointestinal disorders [39].

AZ123 was subsequently tested in an *in vitro* functional assay using the guinea pig isolated ileum preparation to explore its functional effects at muscarinic $M_3$ receptors. AZ123 alone had no effect on the guinea pig ileum preparation but inhibited the contraction caused by methacholine in a concentration-dependent manner with an $IC_{50}$ of 9.5 μM suggesting that AZ123 was a functional antagonist of the muscarinic $M_3$ receptor. It can therefore be concluded that the effects of AZ123 observed in the rat are likely to be mediated by antagonist activity at the muscarinic $M_3$ receptor. These data can be used to make predictions of the potential effects in the clinic. AZ123 has the potential to cause the classical side effects caused by other drugs that have antagonist activity at the muscarinic $M_3$ receptor, such as dry mouth, effects on pupil diameter, and subsequent effects on vision as well as effects on gastrointestinal motility (constipation) [40] if free plasma levels are achieved that exceed the $K_i$ for AZ123 at the human muscarinic $M_3$ receptor. Pharmacological profiling has enabled explanation of *in vivo* observations from

an animal model and allowed predictions to be made of potential effects in the clinic.

### 5.3 Case Study 3. Chlorpromazine and Reprofiling: The Application of Pharmacological Profiling Beyond Secondary Pharmacodynamics

The development of novel uses for known compounds, called "drug repurposing" or "drug repositioning," is becoming increasingly important in the pharmaceutical industry. Although life-cycle management of compounds by large companies to extend the use of a current drug into new indications is nothing new, a number of recent publications [41–44] and the emergence of biotechnology companies focused on drug repurposing highlight the increasing interest in finding new therapeutic indications for existing compounds.

Although there are several examples of repurposed drugs, where a new indication is to be treated using the same primary mechanism of action, there are fewer examples of drugs that have been reused with an entirely different primary target for a new indication. However, there may be one historical lesson worthy of consideration.

In the late 1940s and early 1950s, Rhône-Poulenc synthesized a series of antihistaminergic compounds, including the successfully marketed promethazine. These compounds had high affinity for, and were antagonists at, histamine receptors. However, effects of promethazine were also observed in the central nervous system [45].

Chlorpromazine, a derivative of promethazine, was found to have a wide range of pharmacological actions, including cardiovascular effects, antiemetic action, and antagonism of alpha adrenoceptors [46]. However, it had a lower potency than promethazine as an antihistamine and so was perhaps of little interest to Rhône-Poulenc.

Smith, Kline and French subsequently negotiated an agreement with Rhône-Poulenc to study the efficacy of chlorpromazine as an antiemetic but found that although it prevented apomorphine-induced vomiting in the dog, side effects included sedation and hypotension, and so chlorpromazine was dismissed as unsuitable as an antiemetic [47].

The value of chlorpromazine began to be realized when it was used in a "lytic cocktail" to enhance surgical anesthesia [48, 49] and later suggested that chlorpromazine alone be used in psychiatry [50]. Subsequently, the successful use of chlorpromazine in the treatment of psychiatric patients was described [51, 52], marking the beginning of the era of antipsychotic drugs.

Chlorpromazine is an example of a compound for which (by modern standards) a limited pharmacological profile was developed. Clearly, the compound was of little use as an antihistamine due to low affinity, while undesired side effects limited its use as an antiemetic. However, in retrospect, it is easy to see that a high affinity for dopamine $D_2$ receptors may explain its undesirable side effects as an antihistamine or antiemetic and its subsequent use as an antipsychotic [53]. In the modern era with the ability to develop rapidly detailed profiles for compounds at the molecular level, pharmacological profiling may have a key role to play in drug repurposing by identifying new targets for which old compounds have high affinity.

## 6 CONCLUSION

The study of secondary pharmacodynamics for candidate drugs is now essential, as embodied in the International Committee for Harmonisation guidelines for drug development (ICH S7A [1]). Both *in vivo* and *in vitro* approaches to secondary pharmacodynamic studies have been discussed, but it is clear that, at the present time, *in vitro* pharmacological profiling is becoming the method of choice for many of the large pharmaceutical companies. The ability to interpret such profile data is rapidly being enhanced with the advent of databases containing pharmacological profiles linked to *in vivo* observations and adverse drug reactions.

In the short term, the application of *in vitro* pharmacological profiling to both secondary pharmacodynamics and to other areas in the pharmaceutical industry is likely to expand. The ability to interpret *in vitro* data to predict *in vivo* effects is likely to increase and be refined. In the medium term, it is likely that the generation of very large datasets linking chemical structures to *in vitro* profiles will enable *in silico* prediction of a chemical's likely profile, and thus its potential physiological effects, without ever having to synthesize or to test the compound.

## REFERENCES

1. Anon. ICH S7A: *Note for Guidance on Safety Pharmacology Studies for Human Pharmaceuticals*. CPMP/ICH/539/00. 2000.
2. Anon. *Guidelines for General Pharmacology Studies—Japanese Guidelines for Non-clinical Studies of Drugs Manual*. Tokyo, Japan: Nippo Yakuji; 1995, pp 71–80, pp 125–129.
3. Kinter LB, Valentin JP. Safety pharmacology and risk assessment. *Fundam Clin Pharmacol* 2002;16:175–182.
4. Redfern WS, Wakefield ID, Prior H, Hammond TG, Valentin JP. Safety pharmacology—a progressive approach. *Fundam Clin Pharmacol* 2002; 16:161–173.
5. Bass AS, Kinter LB, Williams P. Origins, practices and future of safety pharmacology. *J Pharmacol Toxicol Methods* 2004;49:145–151.
6. Kinter LB, Gossett KA, Kerns WD. Status of safety pharmacology in the pharmaceutical industry—1993. *Drug Dev Res* 1994;32:208–216.
7. Bass AS, Williams PD. Status of international regulatory guidelines on safety pharmacology. InWilliams PD, Bass AS, Eds. *Safety Pharmacology A Practical Guide*. TherImmune Research Corporation, USA; 2003, pp 9–20.
8. Anon. ICH M3: *Timing of Non-clinical Safety Studies for the Conduct of Human Clinical Trials for Pharmaceuticals*. CPMP/ICH/286/95. 1997.
9. Anon. ICH S6: *Preclinical Safety Evaluation of Biotechnology-Derived Pharmaceuticals*. CPMP/ICH/302/95. 1997.
10. Anon. *The GLP Pocket Book: The Good Laboratory Practice Regulations 1999 and Guide to UK GLP Regulations*. London: MCA Publications; 1999, pp 1–75.
11. Anon. ICH S7B: *The Non-clinical Evaluation of the Potential for Delayed Ventricular Repolarization (QT Interval Prolongation) by Human Pharmaceuticals*. CPMP/ICH/423/02. 2005.
12. Anon. Committee for Proprietary Medicinal Products. *Points to Consider: The Assessment of QT Interval Prolongation by Non-cardiovascular Medicinal Products*. CPMP/986/96. 1997.

13. Anon. *Draft Guideline on the Non-clinical Investigation of the Dependence Potential of Medicinal Products*. EMEA/CHMP/SWP/94227/2004. 2005.
14. Anon. ICH M4: *Organization of the Common Technical Document for the Registration of Pharmaceuticals for Human Use*. CPMP/ICH/2887/99 rev.2. 1999.
15. DiMasi JA, Hansen RW, Grabowski HG. The price of innovation: new estimates of drug development costs. *J Health Econ* 2003;22:151.
16. Kola I, Rafferty M. Oral communication: new technologies that may impact drug discovery in the 5–10 year time frame workshop. Presented at Biomed Expo, Ann Arbor, Michigan, USA, 2002. Data derived from Prous Science, Drug News Prospect.
17. Shastry BS. Pharmacogenetics and the concept of individualized medicine. *Pharmacogenetics J* 2006;6:16–21.
18. Mangoni AA. Cardiovascular drug therapy in elderly patients: specific age-related pharmacokinetic, pharmacodynamic, and therapeutic considerations. *Drugs Aging* 2005; 22:913–941.
19. Feldman DS, Carnes CA, Abraham WT, Bristow MR. Mechanisms of disease: β-adrenergic receptors—alterations in signal transduction and pharmacogenomics in heart failure. *Nat Clin Practice Cardiovasc Med* 2005;2:475–483.
20. Kola I, Landis J. Can the pharmaceutical industry reduce attrition rates? *Nat Rev Drug Discov* 2004;3:711.
21. Bridgland-Taylor MH, Hargreaves AC, Easter A, Orme A, Henthorn DC, Ding M, Davis AM, Small BG, Heapy CG, Abi-Gerges N, Persson F, Jacobson I, Sullivan M, Albertson N, Hammond TG, Sullivan E, Valentin JP, Pollard CE. Optimisation and validation of a medium-throughput electrophysiology-based hERG assay using IonWorks HT. *J Pharmacol Toxicol Methods* 2006;54:189–199.
22. Fliri AF, Loging WT, Thadeio PF, Volkmann RA. Biological spectra analysis: linking biological activity profiles to molecular structure. *Proc Nat Acad Sci USA* 2005; 102:261–266.
23. Fliri AF, Loging WT, Thadeio PF, Volkmann RA. Analysis of drug-induced effect patterns to link structure and side effects of medicines. *Nat Chem Biol* 2005;1:389–391.
24. Maemoto T, Finlayson K, Olverman HJ, Akahane A, Horton RW, Butcher SP. Species differences in brain adenosine A1 receptor pharmacology revealed by use of xanthine and pyrazolopyridine based antagonists. *Br J Pharmacol* 1997;122:1202–1208.
25. Whitebread S, Hamon J, Bojanic D, Urban L. Keynote review: *in vitro* safety pharmacology profiling: an essential tool for successful drug development. *Drug Discov Today* 2005;10:1421–1433.
26. Krejsa CM, Horvath D, Rogalski SL, Penzotti JE, Mao B, Barbosa F, Migeon JC. Predicting ADME properties and side effects: the BioPrint approach. *Curr Opin Drug Discov Dev* 2003;6:470–480.
27. Law M, Rudnicka AR. Statin safety: a systematic review. *Am J Cardiol* 2006;17: S52–S60.
28. Zhou S, Yung Chan S, Cher Goh B, Chan E, Duan W, Huang M, McLeod HL. Mechanism-based inhibition of cytochrome P-450 3A4 by therapeutic drugs. *Clin Pharmacokinet* 2005;44:279–304.
29. Siddiqui MAA, Keam SJ. Ranolazine: a review of its use in chronic stable angina pectoris. *Drugs* 2006;66:693–710.
30. Antzelevitch C. Arrhythmogenic mechanisms of QT prolonging drugs: is QT prolongation really the problem? *J Electrocardiol* 2004;37(Suppl):15–24.
31. Belardinelli L, Antzelevitch C, Vos MA. Assessing predictors of drug-induced torsades de pointes. *Trends Pharmacol Sci* 2003;24:619–625.

32. Antzelevitch C, Belardinelli L, Zygmunt AC, Burashnikov A, Di Diego JM, Fish JM, Cordeiro JM, Thomas G. Electrophysiological effects of ranolazine, a novel anti-anginal agent with anti-arrhythmic properties. *Circulation* 2004;110:904–910.
33. Antzelevitch C, Belardinelli L, Wu L, Fraser H, Zygmut AC, Burashnikov A, Diego JM, Fish JM, Cordeiro JM, Goodrow RJ Jr , Scornik F, Perez G. Electrophysiologic properties and anti-arrhythmic actions of a novel anti-anginal agent. *J Cardiovasc Pharmacol Ther* 2004;9(Suppl 1):S65–S83.
34. Schram G, Zhang L, Derakhchan K, Ehrlich JR, Belardinelli L, Nattel S. Ranolazine: ion-channel blocking actions and in vivo electrophysiological effects. *Br J Pharmacol* 2004;142:1300–1308.
35. Song Y, Skrycok JC, Wu L, Belardinelli L. Antagonism by ranolazine of the pre-arrhythmic effects of increasing late $I_{Na}$ in guinea-pig ventricular myocytes. *J Cardiovasc Pharmacol* 2004;44:192–199.
36. Jackson PC. Innervation of the iris by individual parasympathetic axons in the adult mouse. *J Physiol* 1986;378:485–495.
37. Gupta N, Drance SM, McAllister R, Prasad S, Rootman J, Cynader MS. Localization of $M_3$ muscarinic receptor subtype and mRNA in the human eye. *Ophthalmic Res* 1994;26:207–123.
38. Lin S, Kajimura M, Takeuchi K, Kodaira M, Hanai H, Kaneko E. Expression of muscarinic receptor subtypes (effect of M3 selective antagonist on gastric motility and in rat gastric smooth muscle emptying). *Dig Dis Sci* 1997;42:907–914.
39. Ross WA. Premedication for upper gastrointestinal endoscopy. *Gastrointest Endosc* 1989;32:120–126.
40. Zinner N, Susset J, Gittelman M, Arguinzoniz M, Rekeda L, Haab F. Efficacy, tolerability and safety of darifenacin, an M3 selective receptor antagonist: an investigation of warning time in patients with OAB. *Int J Clin Practice* 2006;60:119–126.
41. Carley DW. Drug repurposing: identify, develop and commercialize new uses for existing or abandoned drugs. Part I. *IDrugs* 2005;8:306–309.
42. Carley DW. Drug repurposing: identify, develop and commercialize new uses for existing or abandoned drugs. Part II. *IDrugs* 2005;8:310–313.
43. Bradley D. Why big pharma needs to learn the three "R"s. *Nat Rev Drug Discov* 2005;4:446.
44. Ashburn TT, Thor KB. Drug repositioning: identifying and developing new uses for existing drugs. *Nat Rev Drug Discov* 2004;3:673–683.
45. Winter C, Flataker L. The effect of anti-histaminic drugs on the performance of trained rats. *J Pharmacol Exp Ther* 1951;101:156–162.
46. Courvoisier S. Pharmacodynamic basis for the use of chlorpromazine in psychiatry. *J Clin Exp Psychopathol* 1956;17:25–37.
47. Domino EF. History of modern psychopharmacology: a personal view with an emphasis on anti-depressants. *Psychosom Med* 1999;61:591–598.
48. Laborit H, Huguenard P. L'hibernation artificielle par MoyEnes pharmacodynamiques et physiques En chirurgie. *J Chir* 1951;67:631–641.
49. Laborit H, Huguenard P. L'hibernation artificielle par moyenes pharmacodynamiques et physiques. *La Presse Med* 1951;59:13–29.
50. Laborit H, Huguenard P, Alluaume R. Un nouveau stabilisateur vegetatif (le 4560 RP). *La Presse Med* 1952;60:206–208.
51. Delay J, Deniker P. Le traitement des psychoses par une methode neurolytque derive de l'hibernotherapie (Le 4560 R.P. utilise seul. En cure prolongee et continue.) *CR Congress Med Alien Neruol France* 1952;50:497–502.

52. Deniker P. Introduction of neuroleptic chemotherapy into psychiatry. In Ayd FJ, Blackwell B, Eds. *Discoveries in Biological Psychiatry*. Philadelphia: Lippincott; 1970, pp 155–164.
53. Carlsson A, Lindquist M. Effect of chlorpromazine or haloperidol on the formation of 3-methoxytyramine and normetanephrine in mouse brain. *Acta Pharmacol Toxicol* 1963;20:140–144.

# CURRENT PRACTICES IN SAFETY PHARMACOLOGY

Alan S. Bass,[1] Peter K. S. Siegl,[2] Gary A. Gintant,[3]
Dennis J. Murphy,[4] and Roger D. Porsolt[5]

[1]*Schering-Plough Research Institute, Kenilworth, New Jersey*
[2]*Merck Research Laboratories, West Point, Pennsylvania*
[3]*Abbott Laboratories, Abbott Park, Illinois*
[4]*GlaxoSmithKline Pharmaceuticals, King of Prussia, Pennsylvania*
[5]*Porsolt & Partners Pharmacology, Boulogne-Billancourt, France*

**Contents**

1 Introduction
2 Regulatory Guidelines Governing Safety Pharmacology
  2.1 Introduction
  2.2 Regulatory Strategy for Safety Pharmacology
  2.3 Recommended Core Safety Pharmacology Studies
  2.4 Follow-Up or Supplemental Safety Pharmacology Studies
  2.5 Conditions Under Which Safety Pharmacology Studies Are Unnecessary
  2.6 Application of Good Laboratory Practice (GLP)
  2.7 Dose Levels or Concentrations of Test Substance
  2.8 Timing of Safety Pharmacology Studies in Relation to Clinical Development
  2.9 Test Systems and Route of Administration
  2.10 Duration of Studies
  2.11 Integrated Risk Assessment
  2.12 Summary
3 *In Vivo* Cardiovascular Safety Pharmacology
  3.1 Introduction
  3.2 Consideration of Study Design

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

3.3 Interpretation of Results: An Integrated Assessment of Cardiovascular Risk of Human Exposure to the Test Substance
3.4 Summary
4 *In Vitro* Cardiac Safety Pharmacology (Cardiac Electrophysiology)
4.1 Introduction
4.2 Ionic Current Assays
4.3 Repolarization (APD) Assays
4.4 Isolated Tissue or Cell Preparations Used for Repolarization Studies
4.5 Isolated Heart Preparations Used for Repolarization Studies
4.6 *In Vitro* ECG (QT) Assays
4.7 *In Vitro* Proarrhythmia Models
4.8 Conclusions
5 Respiratory Safety Pharmacology
5.1 Introduction
5.2 Functional Disorders of the Respiratory System
5.3 Techniques for Measuring Changes in Respiratory Function
5.4 Design of Respiratory Function Safety Studies
6 Central Nervous System (CNS) Safety Pharmacology
6.1 Introduction
6.2 Core Battery CNS Safety Pharmacology Procedures
6.3 Follow-up CNS Safety Pharmacology Procedures
6.4 General Conclusions
Acknowledgments
References

## 1 INTRODUCTION

In the latter half of the 1960s and through the 1970s, it was recognized that contemporary approaches to toxicity testing of new pharmaceuticals were insensitive to detection of many of the severe pharmacodynamic properties, which unfortunately were ultimately realized following clinical exposure to the promising new drug. Professor Gerhard Zbinden, arguably the father of the field of safety pharmacology, said: "The adverse drug reactions which the standard toxicological test procedures do not aspire to recognize include most of the functional side-effects. Clinical experience indicates, however, that these are much more frequent than the toxic reactions due to morphological and biochemical lesions" [1]. Up to the early 1980s, pharmacodynamic testing was conducted ad hoc on selected organ systems taking into consideration the known pharmacology of the therapeutic class or concerns related to the chemical structure [2]. At that time, study protocols were not well defined and the selection of which organ systems would be evaluated did not follow a systematic approach. By the early to late 1980s, safety pharmacology programs began to appear in the pharmaceutical industry [2]. However, it was during the 1990s that the industry saw the greatest growth of these programs, with most institutions turning to the Japanese *Guidelines for General Pharmacology Studies* [3, 4] as guidance for the selection of which organ systems to evaluate and protocols to follow. By the latter part of the 1990s, regulatory bodies from around the world began to recognize the need for specific guidelines for the pharmacodynamic study of the safety of new chemical entities. This led to the preparation of draft guidelines in the

United States, Europe, and Japan [2]. Part of the drive to prepare guidelines by the regulatory bodies was the recognition of adverse events emerging during the postmarketing surveillance period with several new pharmaceuticals [5, 6]. These adverse events had not been identified during the development of these molecules based on contemporary methods of safety testing. Combining the general principles of each of these regional recommendations into a common document, a draft guideline was generated under the auspices of the International Conference on Harmonization (ICH) following the adoption of an initiative on Safety Pharmacology (Topic S7) by the ICH Steering Committee in 1998 [2]. The Expert Working Group (EWG) of Topic S7, made up of members from the regulatory community and pharmaceutical industry worldwide, deliberated for three years over the content and wording of the guidance recommendations. The document finally achieved Step 4 and was adopted by EWG members in 2000 and implemented by regulatory authorities worldwide in the summer of 2001 [7]. The suffix "A" was added to Topic ICH 7 (ICH S7A) to allow for a new initiative, which during the three years was met with a great deal of discussion and debate, Topic S7B. ICH S7B addressed the preclinical study of drug effects on cardiac ventricular repolarization as a surrogate biomarker of torsades de pointes proarrhythmic risk. This important topic is considered in detail in a later section of this chapter, but suffice it to say that this effort by the EWG members was met with a great deal of controversy, disagreement, and compromise because of the lack of definitive data informing this area of study [8–10]. In spite of these challenges, the ICH S7B guidelines were developed and revised over a period of five years, achieving Step 4 in May 2005 and being implemented in the United States and Europe in the latter part of that year. As of this publication, ICH S7B has not been implemented in Japan. [11].

As a field of study, safety pharmacology is a relatively young discipline. In many ways this is an exciting time for safety pharmacologists given the prospects for a future filled with challenges and opportunities that are clearly illustrated by the great progress that has been made during the field's brief history. Safety pharmacology is described in the ICH S7A guidelines as "those studies that investigate potential undesirable pharmacodynamic effects [of new chemical entities or biologics] on physiological functions in relation to exposure in the therapeutic range and above." More broadly, the field of safety pharmacology allows for the study of new and marketed therapeutic agents with the objectives of identifying their pharmacodynamic properties and, through detailed analysis and investigation, understanding the human risk posed by exposure to the promising new drug. These nonclinical studies can stand alone and provide critical data demonstrating the pharmacodynamic safety of a potential therapeutic agent before it progresses into clinical development. Alternatively, the results of safety pharmacology studies may serve a complementary role: providing additional data from directed studies, offering important perspectives of findings from other nonclinical or clinical investigations. In some cases, these nonclinical investigations may be immediately critical to understanding an observation of specific concern. Safety pharmacology studies may also provide data to validate a clinical strategy, which involves an intervention that may mitigate the risk to humans exposed to a new therapy (e.g., effectiveness of anticonvulsant therapy when there is a concern of seizurogenic properties). Although many examples will be cited in this chapter of the ways in which safety pharmacology serves to support the discovery and development of a new pharmaceutical

agent, by no means can we be inclusive of all situations with which the reader will be faced. In fact, newly identified human health-related issues, discovery of drugs that are uniquely directed at target sites for which the safety database may be limited or completely unknown, introduction of new technologies that expand the opportunities to explore the biology and pharmacology of promising pharmaceuticals, and enhancements to technologies that improve the degree of sensitivity and specificity with which we can relate the nonclinical pharmacodynamic findings to the human response, in many ways describe the opportunities that are afforded safety pharmacology to align the growth of this emerging field with these advancements in science.

Given the broad scope of safety pharmacology to identify and further understand the potential pharmacodynamic safety or risk posed by a promising new molecule, the design of *in vitro* or *in vivo* studies may evolve from an unlimited number of decisions (e.g., species, gender, formulation, route) [7]. The ICH S7A guidelines describe three types of studies: the *core battery studies*, which include the cardiovascular, respiratory, and central nervous systems; *supplemental studies*, which include the other major organ systems (e.g., gastrointestinal, renal, immune); and *follow-up studies* of the core battery systems, which are more detailed and investigational and directed toward further defining an observation made in the core battery study or another study of a chemical agent. The focus of this chapter is on the safety pharmacology core battery systems, since these organ systems are highlighted in the ICH S7A as essential to sustaining life [7]. That is, an abrupt interruption of function of any of these organ systems may have a profound, acute, and potentially disastrous outcome, which may under extreme conditions include death. This principle articulated in the ICH S7A guidelines is not meant to diminish the importance of studying other major organ systems. The reader is referred to Refs. 12–16, which describe the study of those systems encompassed in the conduct of supplemental studies.

In the discourse of each of the presentations in this chapter, we provide guidance for many of the choices and decisions that will be made in devising an appropriate and rational study design. We focus on such issues as which models are available for studying a particular organ system, those factors that are taken into consideration in developing a study rationale, and our own personal perspective of how data from these investigations may be viewed in developing an integrated assessment of human risk. These are by no means the only approaches to the study of new pharmaceutical compounds [17–21]. In fact, several approaches to the study of each organ system presented by the different authors. That way, the reader is introduced to a diversity of approaches for considering those factors that are important in developing a rational study design. Although this chapter provides a comprehensive overview of the current technologies and approaches of investigating pharmacodynamic activity of potentially novel new drugs, as suggested earlier, the scientific fields that intersect with safety pharmacology are continuously evolving. In many ways, these advancements may complement the "established" practices of this field, but in some cases these advancements may supplant the common approaches utilized by safety pharmacologists [22]. Based on this vision of the future of safety pharmacology, the reader is encouraged to continuously view the information provided in this chapter in context with the advancements emerging from those interconnecting scientific fields.

## 2 REGULATORY GUIDELINES GOVERNING SAFETY PHARMACOLOGY

### 2.1 Introduction

Safety pharmacology studies were first referenced in the International Conference guidelines: *Safety Evaluation of Biotechnology-Derived Pharmaceuticals* and *Nonclinical Safety Studies for the Conduct of Human Clinical Trials for Pharmaceuticals* [23, 24]. In these guidelines, three categories of pharmacology studies are described for regulatory applications: primary pharmacodynamic studies, secondary pharmacodynamic studies, and safety pharmacology studies. Secondary pharmacodynamic studies (also referred to as general pharmacology studies) are described in the *Guideline for General Pharmacology* issued by the Japanese Ministry of Health and Welfare (MHW; currently referred to as the PMDA) in 1991 [3, 4]. Until 2000 there were no recognized guidelines for safety pharmacology studies.

Nonclinical toxicology studies performed to support safe conduct of clinical trials with drug candidates and registration of new drugs have traditionally used endpoints of mortality, physical signs, histopathology, and hematology/clinical chemistry. The addition of pharmacological endpoints to assess direct effects of drug candidates on major organ function (cardiovascular, respiratory, and central nervous system) complements the traditional toxicology package and provides more complete information for predicting human risk. ICH guidelines S7A and S7B were developed to provide specific information for conducting safety pharmacology studies. Compared to the earlier Japanese guidelines, the ICH safety pharmacology studies are more focused on whole organ system function and *in vivo* assays, with a prerequisite of identifying the safety or risk posed by exposure of subjects or patients to a drug candidate. Now that the ICH S7A and S7B guidelines have been adopted by all regions [7, 11], it has been assumed by some that the ICH guidelines will supersede the requirement for the PMDA general pharmacology studies for drug registration in Japan. However, data from both safety pharmacology and PMDA general pharmacology (secondary pharmacodynamic) studies to support new drug registration in Japan should be considered on a case-by-case basis and may be requested by Japanese regulatory reviewers.

Guidelines for human pharmaceuticals are created to provide both sponsors and regulators with a common reference to facilitate drug development and approval. While guidelines are not rules or absolute requirements, effective guidelines reflect current best practices and recommend approaches that regulatory agencies consider acceptable for addressing concerns about the safety of drug candidates. In developing guidelines in the ICH process, there is participation by representatives from both regulatory agencies and the pharmaceutical industry. The regulatory point of view is valuable because regulators have the opportunity to see many different types of drug applications, while the pharmaceutical industry point of view is important because sponsors provide the practical and business (resources) points of view. The Expert Working Group members from both regulatory agencies and industry provide scientific expertise. Representation from the three major regions (United States, Europe, and Japan) enables a global perspective and should result in a guideline that helps sponsors to build a new drug application package that is acceptable in all three regions. Despite efforts to harmonize the guidelines and ensure

consistency in wording throughout the document, there can be regional differences in the application of the guidelines due to variations in how the regulators choose to interpret the specific text.

The two safety pharmacology ICH guidelines, S7A and S7B, share the following common objectives: (1) to provide general principles, definitions, and recommendations; (2) to minimize use of animals and resources; and (3) to provide information that is a framework which facilitates the use of scientifically valid and internationally accepted approaches for drug development. ICH S7A was created to define the safety pharmacology studies referenced in the earlier ICH guidelines and to provide specific recommendations for the series of studies that constitute a core evaluation in safety pharmacology. Adoption of ICH S7B followed that of ICH S7A with the objective to define an approach that specifically addresses the risk of a delay in cardiac ventricular repolarization (QT interval prolongation) by drug candidates. The two guidelines should be considered together since the basic principles of ICH S7A apply to ICH S7B.

## 2.2 Regulatory Strategy for Safety Pharmacology

Members of the pharmaceutical industry and regulatory community share a common goal of making safe and effective new drugs available to patients. Therefore, bringing both groups together to develop ICH guidelines is appropriate. In addition to optimizing the safety of drug candidates, industry would like to minimize failures of these promising drugs, particularly those that advance to late stage clinical development, increasing the probability that there will be an acceptable return on resource investments. Regulators look to balance clinical safety with approval of drugs that provide true benefit to patients. It is valuable for those working in the pharmaceutical industry to consider the regulator's point of view and understand that both parties share this important responsibility to patients. Providing sound rationale for the choice of assays and an integrated summary of data as it relates to human safety is very helpful for the regulators. When there are questions or concern about a strategy or data package that varies from the approach recommended in the ICH guidelines, the sponsor should engage in discussions with regulatory agencies as early as possible. The most effective strategy takes into account the concerns and needs of the regulators, engages in discussions with regulators in advance when necessary, and ensures that issues are addressed clearly and with appropriate scientific rigor.

Since safety pharmacology data contribute to the overall benefit to risk assessment of drugs, the value of these data is enhanced by inclusion of comparative data with drugs from the same class or mechanism. Results with positive controls and reference agents are very useful in supporting the data with a new drug candidate because it allows the regulators to evaluate the sensitivity of the assays and comparative pharmacodynamic profile. This is particularly important when there are no adverse effects observed with the new drug candidate since using positive controls will demonstrate the sensitivity of the assay to detect a significant pharmacodynamic change had one been detected with the new drug. As allowed within the ICH S7A guidelines, the sensitivity to a positive control can be determined at the point of establishing the *in vivo* model system or with each study of a new molecular entity when an *in vitro* model is chosen.

### 2.3 Recommended Core Safety Pharmacology Studies

In both ICH S7A and ICH S7B, there are recommended studies. In ICH S7A, assays assessing cardiovascular, respiratory, and central nervous system (CNS) effects are referred to as *core battery studies*. Although this term is not used in S7B, the *in vitro* $I_{Kr}$ *assay* and the *in vivo QT assay* are considered equivalent to core battery assays. In most cases, these are the minimum studies that the regulators will expect to see in regulatory dossiers. There may be cases where the safety of the test substance in one or more of the organ systems is assessed with pharmacology studies used during early development (i.e., during selectivity profiling in primary or secondary pharmacodynamic studies). In most cases, these studies are not performed according to GLP. According to ICH S7A, these studies will not need to be repeated for the safety pharmacology package; however, the scientific strength of the data and conclusions from the non-GLP studies should be clearly documented (see Section 2.4). There may be cases where one or more of the core battery assays are not practical because unwanted effects of the test substance make interpretation of the results of the studies of very limited value. Examples mentioned in the guideline include test substances that cause tremors, changes in behavior, and cardiovascular parameters. The absence of core battery studies of a specific organ system may be acceptable if their noninclusion is clearly explained and justified.

### 2.4 Follow-Up or Supplemental Safety Pharmacology Studies

Follow-up studies are intended "to provide greater depth of understanding than, or additional knowledge to, that provided by the core battery" (ICH S7A Section 2.8.1 and ICH S7B Section 2.3.5), while supplemental studies are intended "to evaluate potential adverse pharmacodynamic effects on organ systems functions not addressed by the core battery" (ICH S7A Section 2.8.2). It is the responsibility of the sponsor to determine if follow-up or supplemental studies will provide useful information when considering results from evaluation of the test substance in core battery studies, primary or secondary pharmacodynamic studies, and clinical studies in subjects or the targeted patient population. While these studies are not required unless there is serious concern regarding the safety of the subject or patient, they provide an opportunity for the sponsor to obtain additional information that will support conclusions and strategic decisions concerning both safety and further development of the test substance. This additional information can also enable regulators to make more informed decisions regarding continuation of a clinical program.

### 2.5 Conditions Under Which Safety Pharmacology Studies Are Unnecessary

In ICH S7A, it is acknowledged that there are conditions where safety pharmacology studies "may not be needed" (ICH S7A Section 2.9). This applies to both ICH S7A and ICH S7B studies. Examples cited include locally applied agents, where there is no or very low systemic exposures, cytotoxic agents for treatment in end-stage cancer patients, and biotechnology-derived products that achieve highly specific receptor targeting. There could be other conditions not listed in the guideline where safety pharmacology studies are not needed. The key text here is "may not

be needed," indicating that the absence of data from safety pharmacology studies in the regulatory dossier should be acknowledged and scientifically justified. In other words, the guideline provides flexibility to not do safety pharmacology studies if the results will be of limited value. For example, if a sponsor is developing a new cytotoxic drug for cancer, it is not appropriate to omit these data without a comment. In such a case, the sponsor should document the lack of safety pharmacology data and provide justification of how the data will not contribute to the safety profile of the test substance. In cases where discrete safety pharmacology studies are not performed, every effort should be made to include safety pharmacology parameters in the multiple dose toxicology studies.

When there are questions about the need for one or more safety pharmacology studies, discussions with the regulatory agencies will provide clarification. It is important to note that negotiating the need for, or justifying the omission of, studies with a regulatory agency in one region may not apply to all regions since there could be regional differences in the perceived need for a particular study.

### 2.6 Application of Good Laboratory Practice (GLP)

ICH S7A Section 2.11 discusses the application of GLP when performing safety pharmacology studies. The specific requirements to satisfy GLP are available from regulatory agencies in each region and have generally been harmonized [2]. Performing studies according to GLP is a requirement for nonclinical toxicity studies that support clinical trials in most regions [25]. The ICH S7A guideline recommends that the core battery safety pharmacology studies be performed in compliance with GLP and "any study or study component not conducted in compliance should be adequately justified and the potential impact on evaluation of the safety pharmacology endpoints should be explained." The purpose of GLP is to recommend procedures that will help to ensure quality and reliability of data. Components of GLP include (1) accountability (Study Director and Quality Assurance); (2) documentation with standard operating procedures (SOPs), protocols, and deviations; (3) data archiving procedures to enable accurate reconstruction of the study; and (4) independent validation of data and procedures by a Quality Assurance group. It should be pointed out that satisfying GLP requirements does not by itself guarantee scientific quality and validity. It is important for sponsors to employ scientifically robust approaches including data from positive control substances to assure assay validation and facilitate the interpretation of results. Keep in mind that the purpose of safety pharmacology studies is to generate data with conclusions that can be supported in a regulatory environment. One should follow GLP to the greatest extent possible but also be practical in the use of animals and resources. For example, there may be cases where safety pharmacology studies are performed before GMP or GLP drug substance is available. In this case, one can provide documentation according to GLP and if the study file is scientifically defendable, the assay need not be repeated merely to satisfy all GLP requirements for the drug substance. As indicated earlier, the burden of proof that the data support the conclusions is on the sponsor and one should not expect the regulatory agencies to accept the scientific interpretation of the data at face value (even if performed according to GLP). There may be some regional differences in the interpretation of need for complete compliance with GLP, and when this is the case, it might be prudent to defer repli-

cating the studies until just prior to approval. This will minimize the use of experimental animals and resources in cases when the drug candidate is dropped from development.

In ICH S7B, there is specific reference to application of GLP for the recommended *in vitro* $I_{Kr}$ and *in vivo* QT assays (Section 1.4), indicating that "when performed for regulatory submission, [these assays] should be conducted in compliance with GLP." This has led to confusion since one generally assumes that all safety pharmacology studies are performed for regulatory submission. The wording in ICH S7B was crafted to accommodate one region that has challenged the value of data from nonclinical studies for predicting risk of QT interval prolongation in humans. Regulators in this region prefer to rely predominantly on clinical data for this risk assessment (see ICH E14 guideline) [26] and, at present, are reluctant to consider nonclinical data to obviate the need for a thorough clinical QT/QTc study. If a sponsor chooses to rely solely on clinical data for assessment of QT safety, then the nonclinical studies do not need to conform with GLP. This approach may be complicated by regional differences in the need for data prior to initiation of clinical studies (see Section 2.8) and therefore availability of clinical data to justify not performing safety pharmacology studies in accordance with GLP. In practice, the sponsor should perform these assays to a sufficiently high scientific standard that they can support the clinical safety of the drug candidate. This includes documentation, computer validation [27], and archiving of data in a manner that satisfies GLP requirements. This regional difference in GLP recommendation reflects the complexity and controversy surrounding different opinions on the value of nonclinical assays to detect risk of delayed repolarization in humans. To help address this, the ICH S7B guideline recommends using an integrated risk assessment (IRA) that considers all relevant data (nonclinical and clinical) to support conclusions for predicting risk in humans. In other words, this approach allows for revision and refinement of the IRA as new data emerge to provide a different perspective of nonclinical or clinical results.

In Section 2.11 of ICH S7A, it is recommended that follow-up and supplemental studies "should be conducted in compliance with GLP to the greatest extent feasible." There is an assumption that each of these studies will be designed to address specific issues, and therefore full documentation via all aspects of GLP and 21CFRPart 11, as noted earlier, may not be practical. The guideline provides flexibility to further characterize drug candidates; however, the general principle of conducting and documenting studies in a scientifically defendable manner is expected. As mentioned earlier, the burden of proof that the data are adequately documented and strong enough to support conclusions is on the sponsor. Scientific quality and integrity are paramount.

### 2.7 Dose Levels or Concentrations of Test Substance

Recommended dose levels for *in vivo* studies and concentrations for *in vitro* studies are discussed in ICH S7A (Section 2.4) and apply to assays described in ICH S7B. The guidelines recommend defining a dose–response relationship for the adverse effects by using doses/concentrations in the therapeutic range and above. In the absence of a pharmacodynamic response on the organ system of interest, it is recommended that doses/concentrations be increased to a level that produces a

moderately adverse effect or to use the maximum feasible doses. The guidelines allow exclusion of a maximum dose, where the pharmacodynamic/toxic properties may confound measurement or interpretation of safety pharmacology data (e.g., tremors interfering with interpretation of the electrocardiogram). There may be regional differences in acceptability of safety margins (ratio of exposure at a no-observed-adverse-effect level and the projected exposure at a therapeutic dose) when the maximum tolerated or maximum feasible dose is not used. In general, exploration of as large a dose level/concentration as feasible is a sound approach to a risk assessment. From a regulatory standpoint, it is preferable to identify pharmacodynamic toxicity at a defined, large margin over anticipated clinical exposure than to show no toxicity at a small multiple of exposure at a clinical dose.

## 2.8 Timing of Safety Pharmacology Studies in Relation to Clinical Development

Consistent with the stated objective of the ICH S7A guideline, "to help protect clinical trial participants and patients," the value of safety pharmacology data is greatest when it is obtained prior to first administration to humans. In ICH S7A (Section 2.10.1), it is recommended that the core assays and follow-up or supplemental studies to address specific safety issues be conducted prior to first administration to humans. When clinical studies with adequate data to address the safety of the test substance on the organ systems of interest are available, one can consider whether the nonclinical safety pharmacology studies are necessary (Section 2.10.3). There can be regional differences in the value of nonclinical data even when clinical data are available. Therefore, in cases where the sponsor chooses to propose that safety pharmacology studies are not necessary, a discussion with representatives from the regulatory agency in advance is prudent to avoid potential delays in review of the regulatory package.

Concerning the timing of ICH S7B nonclinical studies (Section 2.4), the wording is more flexible than in ICH S7A. In ICH S7A (Section 2.10.1), the recommendation is that the safety pharmacology core battery *should* be conducted prior to first administration to humans, while in ICH S7B it is stated that "conduct of S7B nonclinical studies . . . prior to first administration to humans *should be considered*." As described in Section 2.6, this greater flexibility for timing of ICH S7B studies was included to accommodate the regional differences in regulatory positions on the value of nonclinical data in the risk assessment for ventricular repolarization or QT interval prolongation in humans and their value in precluding the need for a thorough clinical QT/QTc study. Regardless of these regional differences, the value of the nonclinical data is greatest when obtained prior to initiation of clinical studies because the results can be used to help in the design of clinical studies and interpretation of the results. It is also recognized that nonclinical data can provide an important perspective in situations where clinical findings suggest an issue of concern requiring resolution. Therefore, regional differences can be considered in the timing of ICH S7B nonclinical studies (in particular, the hERG assay since the *in vivo* cardiovascular study, including evaluation of QT/QTc interval, is required prior to first-in-human clinical trials in accordance with ICH S7A). However, also considered is the scientific value of having data early in development versus the commitment of animals and resources to obtaining these data.

### 2.9 Test Systems and Route of Administration

The experimental conditions and protocols should be selected to ensure that scientifically valid information is obtained. There is a preference for use of unanesthetized, unrestrained animal models and the clinical route of administration. While these conditions best replicate the clinical situation, one should consider test systems and routes of administration that can most effectively assess endpoints that reflect the safety of the test substance. For example, if the clinical route is oral but the exposure following oral dosing in the test system is poor, an intravenous infusion may yield more informative results. In some cases, conditioning and restraining animals may increase the sensitivity of the assay but may require greater resources. The guidelines provide recommendations but do allow flexibility so that sponsors can design experiments that most effectively assess the safety of test substances.

### 2.10 Duration of Studies

The recommended duration of safety pharmacology studies is generally a single-dose administration (ICH S7A Section 2.5). This approach will detect acute effects of the test substance on the organ system of interest. The rationale is that one will evaluate the direct actions of the test substance without confounding effects that are secondary to toxicities and histopathological changes that appear with repeated doses. Repeat-dose safety pharmacology studies should be considered when the data can more effectively address a specific concern that is associated with the pharmacological class, revealed in repeat-dose nonclinical studies and/or clinical studies. Good scientific judgment should guide the decision to include repeat-dose studies in the safety pharmacology package.

### 2.11 Integrated Risk Assessment

There is a well recognized, inherent uncertainty in quantitatively predicting responses to drugs in humans using nonclinical data and, as such, there is usually no single nonclinical assay or type of data that fully captures all dimensions of the clinical setting. In the ICH S7B guideline, it is recommended that the sponsor develop an integrated risk assessment for QT interval prolongation using all available data; that is, an assessment of the potential risk that a new pharmaceutical will prolong QT interval in the clinical trial subject or patient taking into account a composite analysis of all existing data (e.g., pharmacodynamic, pharmacokinetic/toxicokinetic, toxicology). This is similar to the integrated assessment of safety using toxicology data in applications for approval of new drugs. It is applicable to data from ICH S7A as well as ICH S7B studies. The integrated risk assessment is based on basic scientific principles, where results from the various nonclinical and clinical studies should complement each other and, when there are inconsistencies among the data, additional studies or information should be obtained. The integrated risk assessment is a dynamic tool and therefore should be updated as new data become available. In the ICH S7B guideline, it is suggested that the integrated risk assessment for QT interval prolongation be included in the Clinical Investigator Brochure and other regulatory documents concerning the safety of a drug candidate. The integrated risk

assessment is a tool to help sponsors more effectively present all relevant data and communicate conclusions. Specifically, this will aid regulators to assess the safety of drug candidates with an integrated evaluation of available data.

### 2.12 Summary

The ICH S7A and S7B guidelines provide information and recommendations that assist both the pharmaceutical industry and regulators to bring safe and effective drugs to patients. The guidelines were designed to provide flexibility such that the assessments can best address the needs of the drug development candidate or disease as well as encourage sponsors to provide the types of data, with a level of scientific rigor, that will enable regulators to make effective decisions regarding the progression of promising new drug candidates.

## 3 *IN VIVO* CARDIOVASCULAR SAFETY PHARMACOLOGY

### 3.1 Introduction

The nonclinical study of cardiovascular function in safety pharmacology can employ both *in vitro* and *in vivo* models. *In vitro* models primarily serve to facilitate a greater understanding of potential mechanisms and sites of pharmacodynamic activity. Many institutions use *in vitro* models to prioritize predevelopment candidates during lead selection [17, 28, 29]. In contrast, *in vivo* models allow hypotheses and questions of the integrated pharmacodynamic properties, mechanisms of action, influence of indirect modulation (i.e., neuronal or hormonal), and the role of major metabolites (i.e., human metabolites) to be addressed. In addition, pharmacokinetic/pharmacodynamic modeling and an indication of exposure multiples can be obtained from results of *in vivo* cardiovascular studies [30, 31]. The first level of a strategy for the comprehensive evaluation of the cardiovascular system includes an assessment of the major functional endpoints (i.e., heart rate, blood pressures (systolic, diastolic, and mean), and both intervals (RR, PR, QRS, and QT) and morphology of the electrocardiogram (including screening for the presence of cardiac arrhythmias)). Study in unanesthetized, unrestrained animals is a preferred methodology in the ICH guidelines and is feasible based on the availability of current radiotelemetry technologies [7, 32–35]. However, study of either restrained or anesthetized animals is also valuable under certain circumstances, as described later. Ideally, the information that is available at the time of designing the study is a complete and comprehensive package. Those factors considered may include an understanding of the physical properties of the active pharmaceutical ingredient (e.g., solubility, stability); the pharmacokinetics or toxicokinetics of the test agent in the animal species and at the doses considered for study; the metabolic profile of the test substance in various species, including human microsomes and identification of any potential major metabolites; the primary pharmacology of the test agent, including pharmacodynamic–pharmacokinetic modeling based on nonclinical data; and information regarding overt tolerance to the test agent at the doses being considered and the target organs identified in multiple-dose toxicology studies. However, as eluded to earlier in this chapter, safety pharmacology studies may be conducted at any stage

of discovery (e.g., lead optimization) through the early to late clinical development phases and into the postmarketing period; as such, the amount of information available to make decisions on study design may be more limited or complete than those that have just been described.

The assays that apply to study of cardiovascular function *in vivo*, including assessment of drug effects on cardiac ventricular repolarization, are considered in this section. *In vitro* models applied to the study of drug effects on repolarization of the cardiac cycle will be deferred to next section of this chapter. As mentioned previously, the important concern of identifying drug effects on cardiac ventricular repolarization is the subject of the ICH S7B guidelines [11]. QT interval prolongation, which is a biomarker for delayed cardiac ventricular repolarization, serves as one of several risk factors for the potentially self-limiting or fatal cardiac arrhythmia, torsades de pointes [5, 8, 9]. Further consideration of this important topic will follow in this and the following section on *in vitro* cardiovascular safety pharmacology.

### 3.2 Consideration of Study Design

***General Overview*** This section focuses on the principles and design options considered in developing a study of cardiovascular function using current methodologies and models. The reader is directed to Refs. 36–41 for a basic understanding of cardiovascular physiology and pharmacology.

At a basic first tier level, monitoring the key organ system parameters such as heart rate, arterial blood pressures (systolic, diastolic, mean) and electrocardiographic parameters (RR, PR, QRS, QT intervals) and morphology of the electrocardiogram, including surveying for the presence of arrhythmias, is essential for a general assessment of the cardiovascular properties of a promising new drug. In the absence of cardiovascular findings on these endpoints in the safety pharmacology and repeat-dose toxicity studies, there may not be a need for additional nonclinical investigation of cardiovascular risk. However, significant results from the core battery cardiovascular study or other nonclinical or clinical studies may prompt a more detailed assessment of cardiovascular function focusing on hemodynamic or electrophysiologic endpoints. At this advanced second tier level, there is a desire for greater understanding of the physiologic changes underlying these events. A more detailed evaluation may include measurements of systemic or regional blood flow (i.e., cardiac output that is the product of heart rate and stroke volume), systemic or regional vascular resistances (equivalent to systemic or regional blood pressure divided by systemic or regional flow, respectively), intracardiac dynamics (i.e., left or right atrial or ventricular pressures during systole or diastole, dynamic change in pressure ($dP/dt$, volume–pressure loops)), a temporal assessment of intracardiac and vascular changes employing echocardiography, or detailed evaluation of electrical conduction through the heart facilitated by SA- and AV-nodal recordings, His-bundle recordings, or measurement of monophasic action potentials and effective refractory period; the latter measures of electrophysiologic parameters enable a more detailed assessment concerning effects of a test substance on cardiac conduction and ventricular repolarization [38–47].

One important note to consider in incorporating many of these methodologies in the nonclinical investigation of a novel test agent is that the same technologies may also be adapted to clinical investigations. The relationship of nonclinical

findings to the clinical population is an important principle that must be considered in assuring one's ability to assess potential for translation of the activity to humans. That is, critical decisions of progressing or terminating the development of a promising new drug are made based on nonclinical safety data, and so the ability to monitor the same endpoints in animals and humans and to relate these findings to an assessment of risk in humans is significantly important.

Consideration of study design is based on the available data describing the properties of the test agent being evaluated and an understanding of the target population in which this drug will be used. This information may provide some indication of whether to anticipate effects on the cardiovascular system (known pharmacologic class effects or effects related to the chemical structure) or whether modifications to the study design are in order based on the properties of the drug or the target population in which it will be used.

***Selection of Gender*** Consideration should be given to whether a drug will be used in a single gender or in both males and females. However, thought should also be given to whether significant exposure of the nontargeted gender is possible based on environmental exposure or the possibility of a new indication. In the case where both males and females, may be exposed to the test substance targeted for a single gender, study of that test agent in both males and females may be appropriate; although, in general, unique cardiovascular pharmacodynamic effects in males or females are not likely to occur.

On the other hand, the toxicology, pharmacokinetic/toxicokinetic, or pharmacodynamic properties of the drug may differ in males and females (i.e., higher exposure in males, greater tolerance in females) and these properties should also be taken into consideration when selecting the gender for study. If there are no preconditional factors that favor the study of one gender over the other, then the study of cardiovascular function in a single gender (males or females) should be acceptable.

***Selection of Species*** Species typically include mice, rats, guinea pigs, ferrets, rabbits, dogs, pigs, and nonhuman primates (i.e., cynomolgus monkeys). Selection of the species should be based on the objective of the investigation and assurance that the biology being studied in animal models can be translated to understanding the potential pharmacodynamic properties of a drug in humans. As a good example of biological differences between humans and animals, the ion channels associated with cardiac ventricular repolarization, in particular, the delayed rectifier channel, $I_{Kr}$ [is encoded by the human-ether-a-go-go-related-gene (hERG)], are important contributors to repolarization in guinea pigs, rabbits , dogs, and monkeys, but not in mice or rats [9, 11]. Blockade of repolarizing current $I_{Kr}$ is a prominent mechanism through which many drugs delay cardiac ventricular repolarization and cause QT interval prolongation. Therefore, evaluation of ventricular repolarization or ECGs in mice or rats will have limited value in identifying drugs with a potential to produce QT interval prolongation by this mechanism. In other words, if risk of QT interval prolongation by the test article is one of the primary objectives of the study, then the guinea pig, rabbit, dog, or nonhuman primate are more suitable species; taking into account other properties of the test agent that may allow further

prioritization, such as the relationships of metabolism of the test agent to that of humans.

Given the objectives of the core battery cardiovascular study, the dog and nonhuman primate are most often employed as the test species. Some investigators advocate the use of mini-pigs in these studies; however, broad utilization of this species awaits further validation. One benefit of selecting either dogs or nonhuman primates is that they are an acceptable species and widely used in repeat-dose toxicology studies. In that case, data from the toxicology study (i.e., toxicokinetics, tolerance, generation of major metabolites) may complement the design and interpretation of results for the cardiovascular study. Furthermore, collective results of the preclinical studies may help position an integrated assessment of human risk.

***Study Conditions: Unrestrained or Chemical or Physical Restraint*** In most situations with few exceptions, study of a drug in unanesthetized, unrestrained animals is preferred since this allows the investigator to follow the clinical route of administration and eliminates the potential influence of the anesthetic agent or restraint procedure on the pharmacology and pharmacokinetic properties (absorption, distribution, metabolism, and excretion) of the drug. Study in unanesthetized, unrestrained animals is practical today as a result of contemporary technologies that allow collection of cardiovascular parameters through radiotelemetry [7, 32–35]. Based on this technology, continuous monitoring of cardiovascular endpoints for hours, days, and weeks is possible with the only limitation being the duration of the battery life and the stability of the model (e.g., drifting of the blood pressure signal from baseline, occlusion of the blood pressure catheter, loss of a quality electrocardiographic signal, physical condition of the animal). However, for most core battery cardiovascular studies of a relatively short duration, the limitation posed by the battery life should not be a significant impediment.

One of the drawbacks to studying unanesthetized, unrestrained animals is the confounding influences if the drug has pronounced pharmacodynamic effects (e.g., central nervous system pharmacology) and is not well tolerated at the high doses intended for study. Other factors that may potentially confound interpretation of data include those associated with the study environment (i.e., noise and activity in the animal facility). As an example, a drug that stimulates locomotor behavior may produce an increase in heart rate secondary to the increase in activity; affect the quality of the electrocardiogram (e.g., increased presence of electrocardiogram-signal artifact); and lead to sporadic alterations in blood pressure. This animal's behavior may also influence the level of activity and associated cardiovascular parameters of its study mates. These confounding changes in the cardiovascular parameters may not be easily distinguished from the pharmacodynamic activity of the drug. Thus, careful consideration of the known properties of the test agent should be made in selecting the appropriate condition for the study; including consideration of physical restraint or potential use of anesthesia.

The use of anesthetized animals may be an appropriate decision in cases where the drug is known to have profound pharmacodynamic behavioral effects that may potentially interfere with the interpretation of results. Alternatively, the behavioral properties of a drug may be the basis for defining the upper dose in the study [7], and by selecting lower doses free of these confounding effects, a study in conscious animals may be a preferred design.

However, there will be situations where chemical restraint using one of a variety of anesthetic agents is a desired approach to conducting the cardiovascular study in animals. In addition to eliminating or reducing those factors that could confound the interpretation of the results, anesthesia may be desirable where a more elaborate preparation of the animal is needed (i.e., acute placement of flow probes and intracardiac catheters) and recovery of the animal at the end of cardiovascular monitoring is not feasible. Chemical restraint also produces a more stable condition for cardiovascular monitoring, without the changes in cardiovascular parameters imparted by an animal's spontaneous activity. There may be conditions where this stability is desired to increase the sensitivity to detect small drug-related changes.

The reader is referred to Refs. 48 and 49 for a list of typical anesthetics and their advantages and disadvantages. Careful consideration should be given to the effects of the anesthetic agent on normal cardiovascular parameters (e.g., ECG intervals, including prolongation of QT/QTc interval; heart rate; arterial blood pressures), as well as normal physiologic hemodynamic reflexes, so that the investigator is cognizant of the limitations created by the anesthetic agent that may affect or alter the study results [48–55]. The reader is also directed to consult with the veterinary staff with regard to the preferred methodology for induction and maintenance of anesthesia, taking into consideration the specific objectives of the study. During the study it will be important that the study team carefully monitor the basal physiologic conditions of the animal (i.e., blood pressure, heart rate, electrocardiogram, blood gases, blood pH, and bicarbonate levels), adjusting those parameters (e.g., depth of anesthesia, physiologic fluids) that will assure maintenance of stable preparation through the course of the study.

There may be situations where the investigator wishes to study a test agent in conscious animals, but the parameters monitored are not accessible through radiotelemetry technology. This situation requires instrumentation of the animal and externalization of the cable and connectors through which the physiological signal will emanate to the monitoring devices [56]. Between experiments, all exteriorized cables and connectors would be protected within secure jackets. The animals will be accustomed to physical restraint procedures (i.e., restraint tubes for rodents, upright in slings, left or right lateral recumbency for dogs, and dorsal recumbency or chairs for nonhuman primates) for a period that will extend beyond the period of time intended for monitoring in the study. However, the need for training may be avoided if the technology employed in a study can be fully supported by a tethering device; in which case the animal will be free to move about the cage. The investigator should recognize that the level of complexity and number of resources and amount of time required to conduct a study using physical restraint are significantly increased over that of the other procedures that have thus far been described. This would include routine monitoring and maintenance of the hardware exiting the body surface and surgical sites.

***Selection of Technologies*** Introduction of radiotelemetry technologies for the study of cardiovascular function has significantly advanced the field of cardiovascular safety pharmacology and the ability to monitor core cardiovascular parameters continuously over the course of hours to days to weeks in freely moving animals. While the application of these technologies across species continues to be refined, monitoring of basic functions such as heart rate and blood pressure (systolic,

diastolic, and mean) and electrocardiograms is possible in both rodent and nonrodent species and monitoring of systemic flows and cardiac dynamics in nonrodent species is also an established technology [57–59]. These technologies require surgical proficiency to implement in the laboratory; but once animals are instrumented, they can be used repeatedly over the period of years. After implantation of the radiotelemetry device, scheduled monitoring is required to assure the stability of cardiovascular parameters and electronic function of the transmitter and, importantly, the health of the animal. The latter is particularly important in advance of an upcoming study to assure that there are an adequate number of animals to assign to the study and that there were no lingering effects of a prior study agent.

With some of the technologies that require an internal implantation of vascular catheters, ECG leads, and the transmitting units, there will be a limitation of battery life and drifting of the cardiovascular signal, both indications of a deterioration of the transmitter unit. Other technologies that allow external calibration of the blood pressure and blood flow signals and the exchange of expended batteries from an external pack obviate the problems inherent in internally implanted transmitters. However, as noted earlier, there is greater commitment of resources devoted to maintaining animals that are instrumented with radiotelemetry devices that have external cables and leads.

Technologies are also available that allow monitoring of the electrocardiogram by radiotelemetry using transmitters and other hardware retained in specially designed protective jackets for dogs and nonhuman primates [60, 61]. While limited to the evaluation of the electrocardiogram, there are times when focusing on the electrocardiogram is warranted; such as a desire to expand monitoring to a larger population of animals, monitor over extended periods of time in free moving nonrodents, or alter the lead configuration.

Limitations on the number of animals studied using radiotelemetry technologies are based on the number of instrumented animals with accurately functioning transmitters in a colony, the configuration of the data acquisition system(s) (e.g., capable of simultaneously monitoring up to six to eight animals with two to four parameters per animal), the size of the technical staff required for dosing and bleeding in a relatively abbreviated period of time (i.e., completing all tasks in 15 minutes from start to finish), and practical consideration of the number of resources required for data processing and analysis. There are a number of computer-based data acquisitions technologies available on the market for monitoring cardiovascular and electrocardiographic parameters [58, 60, 62–66]. The reader is directed to the vendor websites to review the technologies and determine which system most closely aligns with the requirements of his/her program.

As described earlier, study of conscious animals in the presence of physical restraint (i.e., restraint tubes for rodents, upright in slings or left or right lateral recumbency for dogs, and dorsal recumbency or chairs for nonhuman primates) is a preferred methodology in certain situations. Once the investigator has accepted the increased complexity, the number of resources, and the amount of time devoted to a study employing physical restraint, there are several technologies that are available that allow a more detailed evaluation of the cardiovascular state. Measurement of parameters such as intracardiac dynamics, systemic blood flows, and the standard parameters (heart rate and blood pressures (systolic, diastolic, mean)) are all possible using conscious restrained procedures [67–70].

Animals that are physically restrained can be studied for hours or days, either continuously based on the length of the training or for brief periods intermittently over a longer period of time. During data collection, the requirement of close monitoring of the animal's condition limits the number of animals that can be studied at the same time.

Alternatively, the investigator may employ a tether system to acquire cardiovascular data from instrumented animals that are able to move relatively freely around their cages. However, this procedure also requires maintenance of the cables and exit wounds and training to the tether system.

The same technologies available for monitoring freely moving animals or animals under physical restraint can also be applied in an acute study under chemical restraint. In addition, the anesthetized animal allows the introduction of other technologies that may not be amendable to the chronically instrumented models, such as cardiac catheterization, echocardiography, and other more advanced procedures that allow a detailed assessment of cardiovascular function [70, 71].

Chemical restraint procedures also allow monitoring of electrophysiologic parameters such as the length of the monophasic action potential, the duration of the effective refractory period, or conduction through the SA-node, AV-node, His-bundle, and Purkinje fiber systems [43–45, 51, 72].

Chemical restraint procedures also allow for extrinsic manipulation of the cardiovascular system: for example, stimulation of pre- or postganglionic autonomic fibers innervating the heart or systemic vasculature to study the site of test article action, *in vivo* electrophysiologic techniques to maintain a constant heart rate by cardiac pacing during the evaluation of the proarrhythmic potential of a test substance that is confounded by its property of lowering heart rate, or study of the propensity for proarrhythmic activity of a test agent facilitated by programmed electrical stimulation [73–76].

***Dose Selection*** The ICH S7A guidelines define the basis for dose selection in the core battery cardiovascular study. The selection of doses in other cases will depend on the objective of the study and the availability of data to assist in this important decision. Both the quality and volume of data forming the basis for a rational dose selection will improve incrementally with the phase in which the study will ensue (i.e., early to late discovery; early to late development).

As described in the ICH S7A guidelines, "doses should include and exceed the primary pharmacodynamic or therapeutic range. In the absence of an adverse effect on safety pharmacology parameter(s) … , the highest tested dose should be a dose that produces moderate adverse effects in this or in other studies of similar route and duration." However, as mentioned earlier, the ICH S7A guidelines allow for doses producing biologic responses that had the potential to confound interpretation of results to be identified as a limiting dose: "In practice, some effects in the toxic range [e.g., tremors or fasciculation during ECG recording] may confound the interpretation of the results and may also limit dose levels" [7]. In this situation, the investigator will justify selection of a high dose as being absent of confounding clinical effects.

Selection of the low dose would be based on the desire to study the test substance in the therapeutic range, that is, at a onefold or greater exposure multiple of the projected or measured peak plasma exposure achieved at the clinically efficacious

dose(s). The exposure multiple is most typically determined in safety pharmacology studies from comparison of peak plasma concentrations ($C_{max}$), but may also be based on the *area under* the plasma-drug-concentration–time *curve* [AUC]). The reader is referred to Refs. 77–79 to learn more about the pharmacokinetic/toxicokinetic characterization of new molecules. Defining an observed-effect-level or a no-observed-effect-level based on $C_{max}$ and/or AUC is a topic of considerable interest. Selection of a dose from $C_{max}$ would be based on knowledge that a pharmacodynamic response is the result of achieving a threshold concentration (e.g., a concentration needed to activate a specific receptor subtype, thereby eliciting downstream intracellular biochemical changes culminating in a pharmacodynamic response of the target organ). On the other hand, if it is known or suspected that the pharmacodynamic effect of the drug is the result of accumulation of a test substance in the target tissue following repeated dosing to steady state or secondary to toxicity of the tissue parenchyma, then AUC may be the more appropriate basis for projecting the exposure multiple used in the selection of the low dose. However, the design of the safety pharmacology study usually follows a single dose, and, therefore, a steady-state condition is typically not achieved. A repeat-dose study would be conducted in those situations where there is a specific question or concern. If the toxicity is related to continuous intracellular exposure, but resulting from a mechanism that is dependent on a threshold-related effect, then the use of both $C_{max}$ and AUC may be appropriate for determining the exposure multiples between a dose exposure being evaluated in the nonclinical study and the anticipated exposure achieved in the clinic. Similar considerations of dose selection should be given to assuring the evaluation of major metabolites (particularly major human metabolites) formed in the animal species. After having selected the low and high doses for the investigation, the choice of the mid-dose may be based on the geometric mean between the other two dose levels.

ICH S7A, which requires the core battery cardiovascular safety study to be conducted according to good laboratory practice (GLP), provides for a study in situations where the high dose of the experimental agent is devoid of pharmacodynamic effects and, in such a case, only a vehicle and a single dose are required for testing (Section 2.4.1): "Testing of a single group at the limiting dose as described above may be sufficient in the absence of an adverse effect on safety pharmacology endpoints in the test species" [7]. Application of this design may pose a risk of not being able to define a no-observed-effect-level and, therefore, a decision to follow this simplified design requires prior knowledge about the test substance suggesting a probable negative study outcome. On the other hand, this may be an acceptable level of risk for a selected test agent, given the reduced number of resources and time devoted to the conduct and reporting of an abbreviated study.

***Formulation*** There are several factors to consider in selection of the formulation for the cardiovascular safety pharmacology study. First, consideration is given to the properties of the formulation excipients and their ability to have an effect on cardiovascular endpoints. It will be important in the design of a cardiovascular study using a pharmacodynamically active excipient that you are able to distinguish a placebo effect from that of the active pharmaceutical ingredient.

In situations where novel excipients are employed in the formulation of the test article product, there may be a need to evaluate the cardiovascular safety of the

placebo formulation by comparing its pharmacodynamic properties to an inert negative control agent (i.e., isotonic saline, isotonic dextrose). The required evaluation of a unique vehicle formulation may also occur in situations where the product formulation is being developed for oral administration in early drug safety and clinical investigations, but the design of the cardiovascular study requires a unique formulation in order to accommodate a parenteral route (i.e., for intravenous administration). Most often, for cardiovascular safety pharmacology studies, however, the formulation will be the same as that intended for early clinical administration or that devised to support the early drug safety studies.

***Route of Administration*** The preferred route of administration in the cardiovascular safety pharmacology study is the same route as that intended for use in the clinic. Assuming that all pharmacokinetic properties are the same between animals and humans, this approach allows the study of the test agent under similar conditions of exposure to the test substance.

However, there may be times when an alternative route of administration to that intended for clinical application of the test substance is desirable; for example, in cases where the peak plasma concentration ($C_{max}$) to a test substance may be enhanced by the alternate route (i.e., a test substance administered intravenously achieving a higher $C_{max}$ than by the oral route). This may be desirable when wanting to assess cardiovascular parameters following test substance administration at peak levels greater than those achieved by the oral route.

In the study design where chemical restraint is employed, the possible routes of administration are generally limited to the parenteral or intraduodenal routes, keeping in mind in the case of intraduodenal administration that the duration of study may be important to determining the peak blood levels of the active pharmaceutical which can be achieved. As suggested in an earlier section, the duration of the study using chemical restraint will be determined by the ability to maintain stable cardiovascular parameters through the control of hydration, acid–base status, and the level of oxygenation. When administering the test article via the intravenous route, the use of an infusion, as opposed to administration as a bolus, may provide a concentration–time relationship that more closely models the plasma-concentration curve following the oral route. However, a novel development candidate with a slow rate of elimination (thereby maintaining a peak plasma concentration over a prolonged period of time) may be amenable to a study where animals are first dosed orally and then chemically restrained once peak plasma concentrations have been achieved.

Even in the case of studying cardiovascular function in conscious animals, careful consideration of the route of administration should be given. For drugs that achieve rapid peak exposures following oral administration, the early period of monitoring may be lost in the animal's cardiovascular-related stress response to the physical restraint required for dosing. In the same manner, the cardiovascular responses to bolus intravenous injection of the test substance may be lost in this early period of stress-related cardiovascular instability. For these situations, training animals to accommodate to physical restraint as described in an earlier section and maintaining this restraint position through the baseline and postdose periods will accomplish stability during which drug-related changes can be discerned from normal background variation. This same model may be used in studying drugs during a

prolonged period of infusion, where having stable baseline and postdose periods is desirable.

The cardiovascular study of inhalation products is an area of limited general experience in safety pharmacology. The conditions for exposure of animals to the test substance can be the same as those used in the repeat-dose toxicology study, with the monitoring of cardiovascular and electrocardiographic parameters facilitated by radiotelemetry. Alternative routes of administration (e.g., IV dosing) may also be a means of evaluating the cardiovascular properties of the test agent. However, the unique conditions of exposure to the test substance following the inhalation route (e.g., preferential, sustained, and significantly greater exposure of the heart to the test article following administration into the lungs) may not be obtainable by the alternative route [80, 81].

***Dose Regimen and Dose Groups*** The objective of the core battery cardiovascular study is to identify those pharmacodynamic properties of a new test agent which may pose a risk in preparing for the early clinical trials in healthy volunteers or patients. The design of subsequent clinical studies will not only be based on the results of nonclinical safety studies but may also be tailored according to the observations (e.g., pharmacokinetics, tolerance, and biomarkers of efficacy, where possible) made during the early clinical phase. Based on this development strategy, the design of a typical core battery study uses a single-dose regimen. However, the duration of dosing may be extended to multiple daily doses or dosing over several days (1) when the objective of the nonclinical study is satisfied by achieving a higher steady-state plasma concentration of parent drug or major human metabolite(s), (2) to facilitate tissue accumulation based on a known propensity of the test substance to distribute to cardiovascular tissues or other organs or tissues affecting cardiovascular functions (i.e., central nervous system), (3) when this dosing regimen is needed to avert a dose-limiting toxicity that can be attenuated by a dose escalation design (e.g., emesis), or (4) when more than a single dose is part of the protocol design to evaluate a hypothesis or reproduce study conditions in which cardiovascular toxicity was observed following multiple doses of the experimental substance.

The number of dose groups and dosing design will be based on the objectives of the study and the pharmacokinetic, pharmacodynamic, and toxicity profile of the drug, and the statistical power of the study design (see later section on analysis of data). In the case of the core battery cardiovascular safety pharmacology study, the possible study designs may be an escalating dose or crossover design or a study in discrete groups [82–85].

When the information about tolerance to a drug is limited, the escalating dose design allows the investigator to understand the potentially adverse properties of a drug before moving to the next dose level. The dose escalation design is where one group of animals is given escalating doses of test substance and an optional parallel group is administered vehicle at each of the study intervals. This dosing design can either follow a single dose or multiple doses and decisions about the time interval between each dosing phase may be made based on the information available (e.g., pharmacokinetic/toxicokinetic, tolerance, pharmacodynamic properties). For certain drugs, where there is limited information regarding tolerance or where there is a prolonged half-life, the escalating dose design may be desirable, accepting in the

latter example that there will be accumulation of drug substance with each of the subsequent doses.

The crossover design, most often a balanced Latin square crossover design [83–85], is where each of the animals is given vehicle and test substance at each of the dose levels on separate days, allowing for a washout period between each day of dosing. Animals are randomly assigned to individual dosing schedules, which in the Latin square crossover design follows an established sequence of doses (e.g., high, low, vehicle, mid-dose given on separate days to one animal; low, mid, high, vehicle dose given on separate days to another animal; and so on) such that there is a balance across the study with respect to the combinations of dosing sequences. The washout period should be at least 5 half-lives to minimize the residual drug levels remaining before administration of the next dose. The advantage of this design is that the same animal will receive each of the treatment levels on different days and the potential influences of environment are spread across groups (e.g., at least one subject receives each dose on any particular day). Using such a design, the statistical power and sensitivity to detect changes in cardiovascular parameters are enhanced when compared to the statistical power of the other designs. However, this approach cannot be used with drugs that have lingering toxicities (i.e., cytotoxic agents being developed for oncology) and may not be practical for promising drug candidates with a long half-life.

For drugs that have a long half-life, making the escalating dose and crossover designs impractical, or for those promising agents that cannot be given repeatedly to the same animal (e.g., due to residual toxicity), a study in discrete groups becomes necessary (i.e., each subject is challenged with only one dose). However, for a drug with a prolonged half-life, the long washout period between studies means that this pool of animals will be unavailable for another study for a matter of some time.

Another situation where a study in discrete groups may also be desirable, particularly in comparison to the crossover design, is where there is a need to have the entire dosing phase completed in days, rather than in weeks, which are required for the washout period in the crossover design of many test agents. The disadvantage of a study in discrete groups is the number of animals needed for this design is greater than the number needed for a study using the crossover design (e.g., vehicle and 3 treatment levels: n = 4/group, total of 16 animals for a study in discrete groups; n = 4/dose level, total of 4 animals for a study using the crossover design) and loss of the ability to look for differences in response at different dose levels within an individual subject. This also means that the size of the animal pool and the associated resources required for surgery, longitudinal monitoring, and maintenance will be greater for the animal pool required to support consecutive studies by the discrete group design.

Irrespective of the study design selected, the inclusion of a concurrent negative control (e.g., vehicle, saline, dextrose) is essential to account for time-, environment-, or individual behavior-related changes in the cardiovascular parameters. However, where the anticipated outcome of drug administration is well beyond normal variation for a significant parameter pivotal to understanding the test substance's effects in an animal species (e.g., profound increase or decrease in arterial blood pressure), the need for a concurrent negative control may not be necessary.

***Pretest Monitoring*** Pretest monitoring of cardiovascular parameters in chronically instrumented animals occurs at the juncture in preparing for a study where there is a need to know the animal's suitability, which is judged based on the fidelity of the electrical signal (e.g., arterial blood pressure and electrocardiogram), the relationship of the cardiovascular parameters to the baseline in that animal at the time of instrumentation and changes that have evolved over the course of the animal's residence, and the health of the animal according to a veterinary exam and evaluation of clinical pathology parameters. The duration of cardiovascular data collection during the pretest monitoring may mirror the period of data collection on the day of dosing. In addition to allowing for an assessment of the suitability of the animal for inclusion in the study, these data provide a basis for comparison should questions of preexisting conditions arise during the analysis of the study results. If animals remain unused in a study for prolonged periods of time (e.g., 2 months), periodic monitoring of cardiovascular parameters is recommended to assure that there are a minimal number of animals viewed as acceptable for a full study in the existing pool. This would include at least an additional two to four animals serving as spares.

***Duration of Monitoring*** Careful consideration is given to the duration of data collection during the dosing phase. Generally, data are collected prior to dosing to establish a baseline (e.g., 1–2 hours) and for a sufficient period following dosing to meet the study objectives. At a minimum, the period of data collection should be adequate to encompass the peak plasma concentration ($C_{max}$) of the active pharmaceutical ingredient and any of its major metabolites—particularly any major human metabolites identified in the clinical studies. In some cases, multiple days of dosing may be required to achieve higher steady-state concentrations of the test article and/or its major metabolites than that which can be achieved with a single dose. If knowledge of recovery from possible effects is an important objective, then continuous monitoring for at least 5 half-lives may provide a basis for selecting the period of postdose data collection. An understanding of the onset and duration of activity following dosing will be used to assess the safety of a test substance. In particular, knowing the duration of activity and that an observed pharmacodynamic outcome is reversible are important to devising a plan to manage a potential adverse event should this same event be observed in subjects or patients. With the ability to collect continuous cardiovascular waveforms through radiotelemetry and data acquisition technologies, an investigation can encompass continuous collection of cardiovascular parameters and post hoc on-screen review to identify discrete periods of interest for a more detailed analysis of data generated in a study.

***Parameters*** As described earlier in this section, *in vivo* study of conscious or anesthetized animals allows a detailed evaluation of a number of parameters that will form the basis for characterization of the cardiovascular effects of a potential new drug. Essential to this review is an understanding of the effects of the test substance on heart rate, arterial blood pressure (systolic, diastolic, mean), ECG intervals (RR, PR, QRS, QT), and ECG morphology, including knowledge related to the potential for cardiac arrhythmias. The challenge to the investigator in assessing these critical parameters is to have knowledge of the normal background incidence and variation in each of the animal populations [86–88]. This would include variations in the

particular species selected for study as well as the historical control data generated within the institution. This is particularly important in distinguishing test article-related changes from artifacts associated with movement or a normal background incidence in a specific species, for example, second degree atria-ventricular blockade is a common finding in dogs [86–88]. To better understand the numerical range of normal values in the selected species, investigators are encouraged to understand the confidence limits for each cardiovascular parameter in their institutional datasets.

The study of cardiovascular parameters in addition to those already described may be facilitated by their incorporation into the safety pharmacology core battery study, when there is a known pharmacodynamic property of the test agent or its therapeutic/chemical class for which there is a concern. Additional pharmacodynamic measures may also be added to follow-up studies in order to address questions directed at better defining the conditions for safely administering a drug or understanding an underlying mechanism of action for an observed event. For example, a test substance that raises arterial blood pressure may do so by increasing systemic blood flow (cardiac output) and/or systemic vascular resistance (equivalent to arterial blood pressure divided by systemic or regional blood flow). These changes result in increased cardiac work, which may pose a significant risk to certain patient populations, such as those individuals with compromised cardiac function (e.g., congestive heart failure). In contrast, individuals with reduced blood flow to an organ (e.g., kidney) may be able to tolerate an increase in systemic flow, but not an increase in peripheral vascular resistance if this action further compromises the vulnerable organ. Understanding the hemodynamic changes elicited by a promising new drug may not only define the safety of the molecule in the particular patient population but may also form the basis for conducting a study to evaluate potential underlying mechanism(s) of action.

Drug-related changes in conduction of an electrical impulse through the heart, from the atrium to the ventricle, may also have significant life-threatening consequences. Alterations in the morphology of the electrocardiogram (or pseudo-electrocardiogram) associated with test article administration (assessed by electrodes placed on the body surface, subcutaneously, within the intrathoracic cavity around the heart, or directly on the heart) may be a harbinger of proarrhythmia and provide directions for further study. Furthermore, an understanding of cardiac depolarization and repolarization-related changes following exposure to a potential new drug, as reflected in ECG interval measurements (PR, QRS, QT, and rate corrected QT (QTc) intervals), and the test substance's ability to induce automaticity and reentrant cardiac electrical behaviors will help define the potential risks associated with its clinical exposure [8, 39, 40, 86, 87, 89].

***Analysis of Data*** The reader is referred to Refs. 82–85 to select a statistical application that is most appropriate to the design of the study that has been chosen. Careful attention should be paid to avoiding type 1 and 2 errors in the selection of the statistical application. Discussion with institutional biostatistics colleagues will help facilitate the selection of methodology. An analysis of the power offered by a statistical application and study design and determination of the confidence limits of the historical database will guide investigators in understanding the sensitivity of their models to detect individual changes in specific cardiovascular parameters. As

suggested earlier in this chapter, this information will be of value in providing the appropriate perspective for developing an integrated assessment of risk.

Graphical views (minute-by-minute) of cardiovascular parameters over the course of the predosing and postdosing periods is easily accomplished with contemporary data acquisition technology and may identify periods that warrant a more detailed visual interrogation. Careful consideration should be given to the relationship of pharmacokinetic exposure to the drug and any coincident pharmacodynamic changes. This temporal relationship may aid in distinguishing effects related to the test agent and/or its major metabolites from those resulting from spontaneous activity. The hysteresis of this relationship may also identify findings associated with tissue accumulation, a previously unknown novel metabolite, or a mechanism that manifests long after the peak exposure to the test substance has resolved [30, 31].

Complementary to the graphical review of study findings are tabular presentations of cardiovascular parameters at specific intervals (e.g., every 5, 15, 30, or 60 minutes or a combination of intervals), depending on the route of administration and associated pharmacokinetic/toxicokinetic profile of the test agent and/or its major metabolites. Tabular data may be presented as absolute or change from baseline values. Experience has shown that baseline-corrected parameters have a lower degree of variance than data presented as absolute values and, as a result, baseline-corrected values retain a greater degree of statistical power. The reader is encouraged to work with the biostatistics group to define the best approach to the analysis of data collected under the conditions of study within his/her institution.

As described earlier in this chapter, the subject of drug-induced delay in cardiac ventricular repolarization (QT interval prolongation) has received significant attention by both the medical community and international regulators since QT interval prolongation has been identified as an important substrate contributing to the cardiac arrhythmia, torsades de pointes [5, 8, 9, 11, 90]. Measurement of cardiac ventricular repolarization (*in vivo*) on the electrocardiogram is reflected in the QT interval parameter. However, the QT interval is subject to changes in heart rate independent of other underlying mechanisms that also affect cardiac repolarization. An increase in heart rate leads to a decrease in QT interval and a decrease in heart rate leads to an increase in QT interval under normal physiologic conditions. Thus, many investigators have attempted to devise a mathematical approach to normalizing this relationship (QTc = heart rate corrected QT interval), but these attempts have met with mixed results [91–98]. At this time, it appears that correction formulas that are established specifically for the pool of animals studied or those based on the heart rate–QT interval relationships in individual animals have proved to be the best approaches to normalization, particularly for drugs that have an effect on heart rate. However, there are also laboratories that advocate collecting sufficient amounts of data such that application of rate correction formulas becomes unnecessary and that simply looking at the relationship of QT interval over a range of heart rates (binning method) becomes practical [98, 99]. Alternatively, a beat-to-beat analysis of the relationship of QT interval to TQ interval (the principle of restitution) or RR interval to QT interval have also been suggested as alternative methods of identifying new test substances that may pose a cardiac risk [100]. It should be recognized that there is no ideal method for normalizing the relationship between heart rate and QT interval and the inaccuracies of the current correction formulas appear to increase proportionally with the change in heart rate.

A discussion of rate correction formulas would not be complete without consideration of two formulas most commonly used in animal studies: that proposed by Bazett ($QTc = QT/\sqrt{RR}$) and the one proposed by Fridericia ($QTc = QT/\sqrt[3]{RR}$) in 1920 for use in clinical studies [97, 101]. Both of these formulas have an inherent problem of consistently overcorrecting QTc interval when heart rates are relatively high (tending to show a greater QTc value than that which is normal in the absence of drug-related effects) and consistently undercorrecting QTc interval when heart rates are relatively slow (tending to show a lower QTc value than that which is normal in the absence of drug-related effects) [98]. In this regard, in many preparations, the error produced by the Bazett formula is significantly greater than that derived from the Fridericia formula. As a result, several authors have suggested that the Bazett formula should not be used in animal studies [98]. On the other hand, the fact that the Fridericia formula also demonstrates a similar tendency toward overcorrection or undercorrection suggests that this method of rate correction should be used together with other formulas (i.e., pooled or individual rate correction formulas) to draw a comparison in defining the safety of a novel new drug. An additional word of caution is to establish a formula as appropriate to the conditions of a study in your laboratory and to the specific species being studied. A formula devised for one species under specific study conditions may not be applicable to another. This evaluation should be conducted at specific intervals (e.g., once per year) and the methodology revised as appropriate based on the analysis of new data.

***Establishing In Vivo Cardiovascular Models by the Use of Positive Control Agents*** According to ICH S7A guidelines, "appropriate negative and positive control groups should be included in the experimental design." However, the guidelines go on to say that "in well-characterized *in vivo* test systems, positive controls may not be necessary. The exclusion of controls from studies should be justified" [7]. Since the pharmacodynamic effects of a test article in the multitude of possible cardiovascular parameters may change in multiple directions, the guideline requirement for "well-characterized *in vivo* test systems" is by far the only practical approach to demonstrating sensitivity and specificity of the *in vivo* model. The positive control agents chosen for such a study will accordingly be based on the parameters of interest (e.g., heart rate, arterial blood pressure, and ECG intervals) and the multitude of possible directions in which they may change, and whether this study is establishing a model that will be used in the standard "core battery cardiovascular" study or whether this study is critical to comparing the test agent to a marketed drug, for example, as an objective of a "follow-up" cardiovascular study [7]. In either situation, the positive control agent should be recognized as having been well characterized for the effect of interest by the international scientific community. The reader is directed to Ref. 37 for examples of possible positive control agents across a wide class of cardiovascular targets.

***Discrete Blood Level Sampling*** Maintaining a controlled study environment is of paramount importance in assuring the sensitivity of the model to detect discrete cardiovascular changes. As a result of this concern, the ability to enter a study room in the midst of cardiovascular data collection in order to obtain blood samples should not be followed without careful planning as noted later. This limitation does

not apply to a study conducted in the presence of chemical or physical restraint, where the investigator is continuously in close proximity to the animal. In those situations, multiple sampling of blood from the test species is possible as long as the procedures of blood collection do not interfere with the acquisition of data.

In a study in unrestrained animals, the collection of blood samples can be used to confirm exposure, as long as the time at which blood is collected does not coincide with the peak levels of either the test article or any of its major metabolites. As long as the variability of the pharmacokinetic/toxicokinetic properties of a test substance is small, as demonstrated in a formal study with a full panel of 5–12 blood samples, then the data generated from limited sampling in the cardiovascular study can be compared to the data from the full pharmacokinetic/toxicokinetic study forming the basis for an estimation of the peak plasma concentration achieved in the cardiovascular study. In order to obtain the blood sample without interfering with acquisition of cardiovascular data, the sample(s) are collected after the estimated peak concentration of the test substance and the major metabolites had been achieved; typically one or more samples are collected after cardiovascular data collection has been completed and concurrent with the elimination phase of the plasma concentration–time profile. There can be situations, however, where cardiovascular data are collected over a prolonged period following test substance administration, in which case it may become necessary to collect blood samples intermittently throughout the period of data collection, but avoiding those times in which it is critical to maintain a stable, undisturbed laboratory environment.

### 3.3 Interpretation of Results: An Integrated Assessment of Cardiovascular Risk of Human Exposure to the Test Substance

It will not be practical to comprehensively review all of those factors that can be taken into consideration in forming an integrated assessment of risk in this chapter. In the same manner, the decision to continue or discontinue development of a new molecular entity based on adverse findings in preclinical studies may differ from one institution to another based on the availability of other data (nonclinical and clinical) and the strategic interests of the organization. Because of these variables, this section does not necessarily address all situations with which an investigator may be faced or recommend a simple course of action. Rather, the purpose of this section is to list many of the critical questions that are considered in formulating a decision of the benefit–risk relationship for a drug candidate. More broadly, this careful assessment will contribute to the decision of whether a promising new drug can be safely advanced through the clinical phases and on to registration. Those pharmacodynamic properties considered in devising an integrated risk assessment may include the following.

***Magnitude of Effect in the Animal Studies*** Based on these data, a judgment can be made of whether a similar effect in humans might be tolerated. For example, there is currently significant discussion regarding the pro-torsadegenic risk posed by a test agent that produces an increase of QT/QTc interval having a 95% confidence interval upper bound of 10 ms in the patient population, such that changes of less than 10 ms would be considered nonproarrhythmic [26]. A similar criterion does not currently exist with regard to nonclinical studies; however, there is opportunity

for questioning the relationship between findings in animals and those in humans based on a careful analysis of the outcome of clinical studies.

***Dose (or Concentration)–Response Relationship*** This relationship (e.g. none, shallow, or steep) is important in several respects. The absence of a dose–response relationship may suggest that the effect observed is an artifact of the experimental model, unrelated to the test substance, or that a maximum effect has been achieved. In the latter situation, administering higher doses would not be associated with a greater response and if the magnitude of the effect observed can be tolerated in the clinical studies, progression of the test article may be acceptable. On the other hand, a steep dose–response relationship is problematic as this profile suggests that an adverse event may occur without forewarning once a threshold effect has been achieved during the phase of dose escalation. This situation may warrant more extensive monitoring of the normal volunteers or patients. In extreme cases where life-threatening cardiovascular toxicity is a concern, an institution may decide to discontinue development of the promising new drug candidate.

***Margin of Safety*** The margin of safety (e.g., plasma concentrations eliciting an adverse cardiovascular event in the nonclinical study compared to plasma concentrations required for efficacy in clinical studies) is important in understanding the probability of achieving adequate drug level in patients to produce efficacy without placing an individual at risk of untoward cardiovascular events.

***Pharmacokinetic/Toxicokinetic Properties*** Several questions may emerge from understanding the pharmacokinetic/toxicokinetic properties of a new test substance. For example, was the cardiovascular event coincident with the peak levels of the test substance or its major metabolites? If the cardiovascular activity is related to concentration of the major metabolites, are these formed at concentrations in humans adequate to pose a risk? Could the effects observed in the preclinical species be attributed to either tissue (e.g., cardiac, vascular) accumulation of the test substance or the accumulation of major metabolites with repeated dosing? If these questions remain after careful review of the nonclinical and clinical data, there may be reason to consider the conduct of an additional, follow-up cardiovascular study with more extensive monitoring of pharmacokinetic/toxicokinetic parameters.

***Relevance of the Nonclinical Model to Understanding Human Safety*** Understanding the biology of the system and whether it is operational in humans is of paramount importance in deciding whether a test article-related observation in animals is an indication of a human risk. This topic was described earlier with regard to the differences in the contribution of repolarizing currents operative in rodents when compared to those that are operative in humans [11]. This situation may also apply to novel therapeutic targets, where there is limited data describing the relevance of a novel target to the cardiovascular system in either animals or humans. In such a situation, further nonclinical studies may be needed to understand mechanism-based effects.

***Relationship of Comparable In Vitro and In Vivo Cardiovascular Findings in Terms of Concentration–Response*** A cardiovascular response from one study, which is corroborated by results of another study, provides a strong indication of a potential for a cardiovascular event in humans, assuming that adequate exposures

to the test substance or major metabolite(s) are achieved in the nonclinical study of the test substance. In addition, the results of the *in vitro* study may also provide a basis for a more detailed understanding of the potential underlying mechanism of the adverse event or whether a result may be produced in another species. In contrast, conflicting data between *in vitro* and *in vivo* studies may be the basis for raising questions of whether an observation will occur in humans.

***Distinguishing Direct Pharmacodynamic Effects of the Test Substance from Indirect Effects Due to Toxicities*** Recognizing whether a drug produces a cardiovascular adverse event as a result of its pharmacodynamic properties or as an indirect secondary event resulting from abnormal clinical or anatomic pathology is important in formulating an analysis of human risk and devising a plan for monitoring risk factor(s) in the clinical study of the test agent: for example, QT prolongation due to a direct effect of a compound on the cardiac potassium rectifying current, $I_{Kr}$, versus that resulting from toxicity associated with profound hypokalemia. In addition, understanding the underlying etiology of cardiovascular toxicity provides a basis for directly evaluating reversibility of the untoward effects.

***Recognizing that the Targeted Patient Population May Be Highly Susceptible to a Drug-Related Toxicity*** As suggested earlier in this section, ultimate decisions of the safe progression of a test article with undesirable cardiovascular pharmacodynamic properties will be based, in part, on consideration of the particular susceptibility of the patient population in which that test agent would be used (e.g., torsades de pointes in heart failure patients). In this regard, it is important to recognize that the animal models used in safety pharmacology are typically healthy animals and do not represent the underlying condition of susceptibility that may exist in that targeted patient population.

### 3.4 Summary

The objective of the cardiovascular study of a new molecular entity is to have sufficiently characterized its pharmacodynamic properties in preclinical models that emulate human physiology in order to anticipate the potential risk posed to subjects and patients following clinical exposure. A comprehensive cardiovascular study of a new test agent at the minimum would include monitoring of arterial blood pressures (systolic, diastolic, and mean), heart rate, electrocardiographic intervals, and the electrocardiogram for the presence of cardiac arrhythmias [7]. The evaluation of test article effects on QT interval *in vivo* as recommended in ICH S7B will be monitored in the core battery safety pharmacology study [11]. Additional cardiovascular endpoints may be added to the core battery study based on a cause for concern for human safety. Alternatively, the promising new test substance may be evaluated in a follow-up cardiovascular study for a more detailed understanding of test-article-induced changes in cardiovascular function. Collectively, the composite profile of preclinical and clinical data is assimilated into an integrated assessment of the potential risk posed by human exposure to the promising new drug. In addition, these data inform the clinical investigator where more detailed monitoring of cardiovascular function is needed during the course of dose escalation in early clinical trials. These data also inform the organization that is continuously judging the value and safety of continuing the development of the new molecular entity.

## 4 *IN VITRO* CARDIAC SAFETY PHARMACOLOGY (CARDIAC ELECTROPHYSIOLOGY)

### 4.1 Introduction

A number of noncardiovascular drugs from different therapeutic classes have been linked to delayed ventricular repolarization and sudden cardiac death resulting from a rare polymorphic ventricular tachyarrhythmia known as torsades de pointes [102]. The rare occurrence of TdeP for noncardiac drugs that affect ventricular repolarization (incidence likely too low to reliably detect in clinical trials) and potentially lethal consequences of this arrhythmia necessitate the use of nonclinical surrogate markers to detect drug-induced delayed repolarization (also termed acquired long QT syndrome, or aLQTS). In fact, drug-induced delay in cardiac ventricular repolarization serves not only as a surrogate marker, but also as one of several risk factors (see below) placing susceptible individuals at risk for TdeP. Numerous *in vitro* models have been described that evaluate potential proarrhythmia on subcellular, cellular, tissue, and organ levels. However, none have proved to be fully predictive of either QT prolongation or torsadogenic potential in humans. This likely reflects the extent of integration of more complex systems as one progresses from individual ionic currents to electrophysiologic interactions between different regions of the ventricular myocardium. In addition, several other risk factors (including female gender, hypokalemia, bradycardia, heart disease, excessive drug concentrations, pharmacodynamic interactions, and ion channel defects) have been identified and linked to aLQTS. These findings support the "multiple hit hypothesis" that states that the concomitant appearance of multiple risk factors may be necessary to set the stage for the initiation and perpetuation of TdeP. Thus, the successful identification (and avoidance) of aLQTS and TdeP requires a careful assessment of models of varying complexity and an appreciation of their interactions with additional risk factors. This section discusses the more widely accepted *in vitro* approaches for evaluating proarrhythmic risk of evolving compounds and their organization based on increasing levels of integration. Ionic current assays are discussed first, followed by action potential duration/repolarization studies that reflect responses of tissues, representing a higher level of integration. Finally, evaluations employing *ex vivo* whole heart preparations are described.

### 4.2 Ionic Current Assays

***hERG Current*** The outwardly directed delayed rectifier current $I_{Kr}$ plays a prominent role in initiating and defining cardiac repolarization in humans and larger mammalian species. In humans, the pore forming portion of the channel that carries the $I_{Kr}$ current is encoded by the *hERG* gene (the acronym for *human ether-a-go-go related gene*, also known as KCNH2 or Kv 11.1 9) [103, 104]. Further evidence that the hERG channel is involved in drug-induced delayed repolarization is provided by the finding that mutations of the hERG gene on chromosome 7 are associated with the congenital long QT2 syndrome [105]. Thus, block of hERG current has become an accepted surrogate marker for QT interval prolongation and a risk factor for cardiac proarrhythmia. Indeed, the evaluation of hERG current block plays a pivotal role in the nonclinical evaluation of potential proarrhythmic risk and is sug-

gested in the ICH S7B guidance [11]. While $I_{Kr}$ can be recorded in isolated cardiac myocytes, it is difficult to measure due to its relatively small amplitude and the presence of other time- and voltage-dependent currents that overlap the smaller $I_{Kr}$ current. Thus, drug effects on this current are more easily assessed using human embryonic kidney (HEK) or Chinese hamster ovary (CHO) cells stably transfected with hERG [106, 107]. While transiently transfected *Xenopus* oocytes have been used, concentration–response relationships obtained with this preparation may be confounded and difficult to interpret due to drug accumulation into the yolk sac, thereby reducing drug concentrations at the vestibule of the hERG channel on the intracellular sarcolemmal surface [108].

hERG current is most easily assessed using voltage clamp techniques in which the transmembrane potential is controlled (or clamped) and changes in ionic current flowing across the cell membrane are recorded (Fig. 1). Voltage-dependent hERG channels are activated upon depolarization and conduct outward (repolarizing) current during each action potential. This current reflects the flow of $K^+$ ions through hERG channels, mimicking repolarizing $I_{Kr}$ current flowing in native cardiac myocytes. Drug effects on hERG current are usually assessed from the reductions in the tail current amplitude (measured upon repolarization); tail currents reflect the deactivation (or decay) of hERG current activated by the preceding depolarizing (activating) voltage pulse. Voltage clamp protocols resembling square waves are usually used, although repolarizing ramp clamp pulses as well as waveforms resembling the cardiac action potential have been used when evaluating hERG current block. While no clamp protocol appears to provide significant advantages, a combined step-ramp pattern may provide a conservative estimate of hERG inhibition within a diverse panel of drugs [109].

While prior electrophysiologic experiments have employed traditional patch clamp techniques, the advent of planar patch methods (coupled with automation) holds the promise for greater throughput for functional current assays [110]. This technique allows for the simultaneous recording of currents from individual cells with gigaohm seals from multiple wells. However, issues related to the use of physiologic temperatures and validation of drug concentrations in experimental wells need to be addressed to secure this technology as a definitive safety pharmacology assay.

It is preferable to evaluate effects on hERG current at physiologic temperatures, as reducing the temperature from 35–37 °C to room temperature (22 °C) has been shown to affect potency of block for six of fifteen compounds in one experimental series (less potent for four of the six compounds, with greatest change being approximately sevenfold) [109, 112]. It is also recommended that bath concentrations be measured, as drug levels may be significantly less than anticipated due to adsorption to perfusion apparatus [11, 113].

$IC_{50}$ values for test substance block derived from hERG current studies can be compared with anticipated clinical exposures in order to provide an assessment of relative risk. It is also important to test known hERG-blocking comparator compounds to benchmark the sensitivity of the assay. Redfern and colleagues [114] have suggested that for most compounds a 30-fold margin between hERG $IC_{50}$ values and free $C_{max}$ may provide an acceptable safety margin in humans for TdeP with a low risk of false positives, recognizing that drug effects on other cardiac channels as well as potential benefits of the test agent must also be considered.

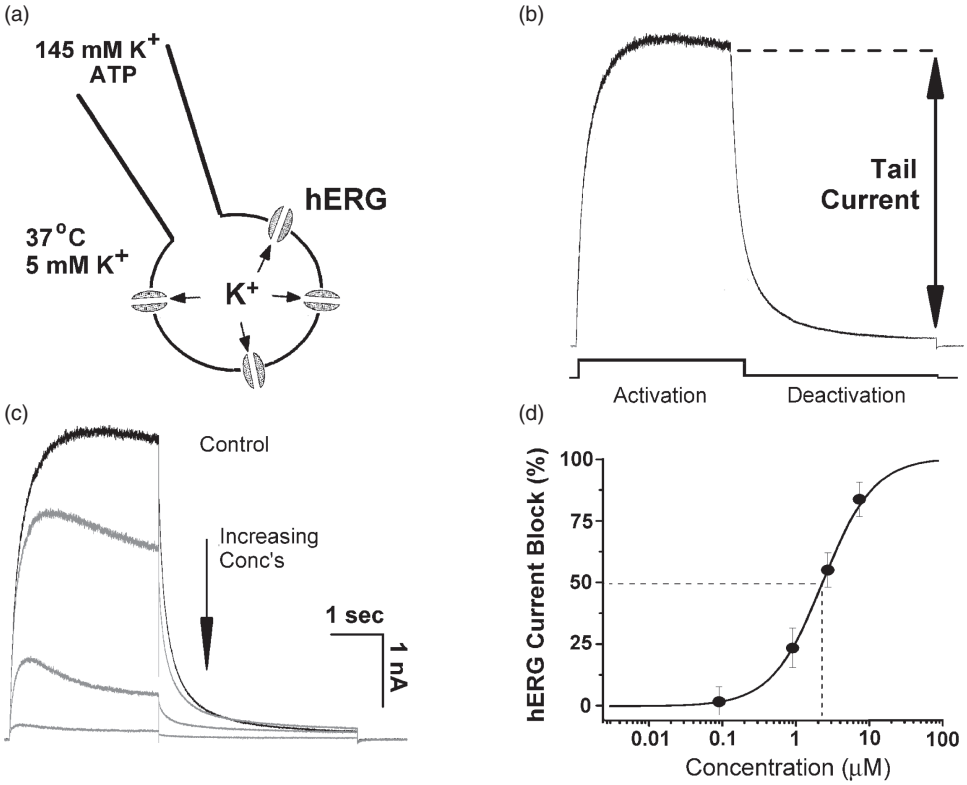


**FIGURE 1** *In vitro* recordings of hERG current. (a) A patch clamp electrode is attached to cell stably transfected with hERG gene and expressing $I_{Kr}$. Typically, a glass electrode is applied to the cell membrane, forming a high resistance gigaohm seal essential for high fidelity current recordings. A pulse of suction is then applied to rupture the cell membrane under the electrode tip, providing electrical access to the cell interior through the pipette solution equilibrating with the cytoplasm. High $K^+$ concentrations in the cell and pipette solutions provide the driving force for outward (repolarizing) current flow. Circuitry connected with the pipette is used to control the membrane potential and record the transmembrane current in response to changes in membrane potential. (b) Records of voltage-dependent hERG current activated upon depolarization and deactivated upon repolarization. The declining outward current upon repolarization (referred to as tail current) reflects the deactivation of hERG channels opened during the prior depolarizing test pulse. A square wave voltage clamp protocol is shown in this example; typical depolarizing pulses range near 0 mV, with tail current measured near –50 mV. (c) Recordings of drug-induced block of hERG current obtained with increasing concentrations of a hERG blocking agent. While activating and tail currents are both reduced, the reduced amplitude of tail current (downward arrow) is typically measured to evaluate hERG current block. (d) $IC_{50}$ values characterizing potency of hERG current block derived from Hill equation fit to experimental data. (From Ref. 111.)

Binding studies utilizing high-affinity radiolabeled ligands [115], atomic absorption measurement of rubidium efflux [116], and fluorescence detection of voltage-sensitive dyes [117] are used in initial screening for hERG channel block. However, these assays often demonstrate lesser sensitivity and do not directly measure functional current. Consequently, they are generally used as higher throughput screening

assays to prioritize advancing candidates earlier in drug discovery rather than to provide a final evaluation of hERG channel inhibition liability.

It is also recognized that some drugs (such as fluoxetine and verapamil) are relatively potent hERG blockers, yet are not typically associated with either delayed repolarization or torsades de pointes arrhythmias. This apparent discrepancy is explained by concomitant block of the inwardly directed L-type cardiac calcium current that may mitigate the consequences of reduced hERG current [118, 119]. In such cases, changes in the action potential obtained with repolarization assays may be used to infer additional counterbalancing drug effects (see below). Drug effects on multiple cardiac ion channels (termed multichannel block) [120] emphasize the fact that hERG represents only one of multiple cardiac ionic currents that may influence cardiac repolarization.

***HERG Trafficking*** Recent reports have suggested that a minority of drugs may delay cardiac repolarization and predispose to TdeP not by direct block of hERG current, but rather by preventing the normal processing and cellular trafficking of the nascent hERG channel protein from the endoplasmic reticulum and reducing the surface membrane expression of functional channels. Such drug candidates may include the cancer chemotherapeutic drug arsenic trioxide [121, 122] and the antiprotozoal drug pentamidine [123, 124]. Other "conventional" hERG channel blocking drugs may also carry an additional risk of hERG trafficking inhibition [125], which may occur over the course of hours (rather than minutes as typically associated with simple block of hERG current). Further studies are necessary to define the role of assays evaluating indirect effects of drugs on ion channel protein trafficking and expression in preclinical screening.

***Evaluation of Drug Effects on Other Cardiac Ion Channels*** Studies of congenital long QT syndromes have also linked reduction of slowly activating repolarizing current ($I_{Ks}$, carried by KVLQT1/minK channels) or defective inactivation of the inwardly directed sodium current (SCN5A) to torsades de pointes proarrhythmia (long QT syndromes 1 and 3, respectively) [126, 127]. While screening of drug effects on these additional channels may be prudent, such studies are not recommended in present guidelines [11], possibly due to the likely limited number of drugs affecting these channels compared to hERG. The presumption of a testing strategy for new molecule testing is that drug-related effects on these other channels will be revealed in the *in vivo* cardiovascular studies as described in Section 3.

### 4.3 Repolarization (APD) Assays

$I_{Kr}$ current (carried by hERG channels) represents only one of multiple ionic currents that are involved in cardiac repolarization. Indeed, it is the complex interplay of multiple ionic currents, pumps, and exchangers that is responsible for the generation of the action potential. Figure 2a illustrates a ventricular action potential—a recording of transmembrane potential versus time. From a resting membrane more negative than –80 mV, stimulation elicits a rapid upstroke (termed phase 0), with an amplitude greater than 100 mV. Repolarization ensues, consisting of three sequential phases: phase 1 represents rapid repolarization that occurs immediately following the upstroke, phase 2 defines the action potential plateau sustained by the inward

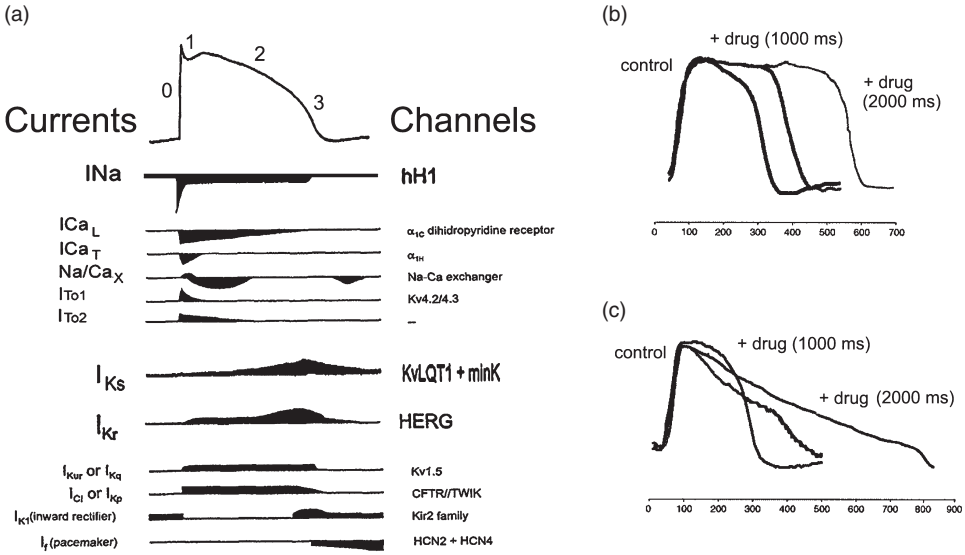


**FIGURE 2** Repolarization assays. (a) The upper traces show a representative intracellular action potential with the phases of the action potential (0–3) labeled. Shown below are the currents that activate and inactivate during the action potential; currents are labeled on left, with corresponding channels indicated on right. Inward currents are represented by downward deflections, while outward currents are shown as upward deflections. Mutations of three channels linked to congenital long QT syndromes ($I_{Na}$, $I_{Ks}$, and $I_{Kr}$) are shown in bold letters. (Adapted from Ref. 243). (b, c) Monophasic action potential recordings from left ventricular epicardium of a rabbit heart Langendorff preparation. Panel (b) demonstrates drug-induced action potential prolongation without triangulation. Reverse use-dependent effects of this drug are evident, with greater prolongation at the slower (2000 ms) versus more rapid (1000 ms) pacing cycle length. Panel (c) demonstrates triangulation of the action potential with another drug; greater triangulation is evident at the slower stimulation rate. See text for further discussion. (Adapted from Ref. 130.)

calcium current, and phase 3 delineates terminal repolarization, where increasing outward currents contributed by $I_{Kr}$, $I_{Ks}$, and $I_{K1}$ return the membrane potential to the resting potential. Reduction of these outward currents (or increased inward current through $I_{Na}$) reduces the "repolarization reserve" of the myocardium, predisposing to proarrhythmia [128].

Transmembrane action potentials are recorded using sharp or patch microelectrodes (the former for tissues or isolated cardiac myocytes, the latter primarily for cardiac myocytes) that provide intracellular access and allow for voltage measurements across the cell membrane (between the cell interior and a reference electrode in the recording chamber). These action potentials arise from negative potentials and have upstrokes that exceed 0 mV, with amplitudes typically greater than 100 mV.

It is also possible to record action potentials from the myocardial surface using two blunt extracellular electrodes, with one electrode depolarizing or injuring the underlying myocardium, and the second electrode recording extracellular voltages

from the uninjured extracellular surface. Such recordings are termed monophasic action potentials (MAPs) [129]. MAPs are smaller in amplitude than transmembrane action potentials and do not exceed 0 mV (hence the term monophasic). Typically recorded from epicardial or endocardial surfaces, MAPs may be subject to greater motion artifacts than intracellular recordings. While MAPs represent the general time course of repolarization, they cannot be used to evaluate changes in the resting membrane potential or upstroke characteristics (the latter reflecting changes in fast inward sodium current). MAPs may be recorded from isolated Langendorff hearts, but also from intact anesthetized animals as noted in Section 3.

The lower portion of Fig. 2a displays the prominent currents involved in generating a ventricular action potential. Inward, depolarizing currents (grouped toward the top) are indicated by downward deflections, while outward, repolarizing currents (grouped toward the bottom) are indicated by upward deflections. The complexity is evident. Of particular note is the time course of activation of the outward repolarizing current $I_{Kr}$ (hERG) during phase 2 to initiate terminal repolarizaiton of the action potential. Variations in the relative contributions of different currents are responsible for regional heterogeneity of the action potential configuration and duration across the ventricular wall. These differences also contribute to the electrical gradient responsible for the T wave of the electrocardiogram. Furthermore, differences in the configuration and duration of the action potential affect the extent of drug-induced action potential prolongation recorded from different regions of the myocardium (e.g., drug-induced delayed repolarization is typically greatest in Purkinje fibers and ventricular midmyocardium compared to epicardium or endocardium during slow stimulation) [131, 132]. Thus, the selection of tissues needs to be taken into consideration when evaluating drug effects on repolarization, and effects of known standards should be included to benchmark the sensitivity of the preparation. Selection of species is also important; for example, $I_{Kr}$ current contributes minimally to repolarization in mouse and rat ventricular myocardium [133–136] and would be poor choices to use to detect drugs that reduced $I_{Kr}$. As in all scientific studies, drug effects on repolarization are most readily interpreted when parallel, vehicle-control studies demonstrate no effects on the action potential configuration over time.

Various indices of repolarization have been used to predict delayed repolarization and increased risk of torsades de pointes. Based on the link between QT interval prolongation and TdeP, prolongation of the action potential duration (APD) has been used as a measure of proarrhythmia risk. This measure is typically reported as $APD_{90}$, the time required for 90% repolarization. Most drugs that block $I_{Kr}$ also demonstrate reverse use dependence, defined as greater drug-induced prolongation of the action potential at slower versus more rapid stimulation rates. An abnormal form of triggered electrical activity (termed early afterdepolarizations, or EADs) may be observed in repolarization assays during slow stimulation and the presence of QT prolonging drugs. Early afterdepolarizations are single or multiple premature depolarizations that arise during the action potential plateau. This form of triggered activity has been attributed to reactivation of inward L-type calcium current that arises during the action potential to initiate TdeP [137]. EADs are enhanced during slow stimulation and are thus considered harbingers of proarrhythmia.

More recently, drug-induced changes in the action potential configuration from a "rectangular" shape to a more triangular-shaped configuration (termed triangulation) have been implicated in proarrhythmia risk [130]. Measured as either the difference in time between 30% and 90% repolarization (or alternatively, normalized using the ratio of these values), triangulation may increase the risk of proarrhythmia in the absence of changes in the action potential duration for some drugs [130, 138, 139]; a notable exception to this rule is the hERG and calcium channel blocking drug verapamil. Other electrophysiologic indices linked to proarrhythmia risk, but not necessarily linked to altered repolarization or TdeP, may also be gleaned from intracellular recordings of the action potential; these include effects on upstroke (related to changes in conduction velocity) and changes in the resting membrane potential (related to block of the inwardly rectifying potassium current $I_{K1}$). Proarrhythmia associated with congenital short QT intervals has been described, linked to mutations in $I_{Kr}$, $I_{Ks}$, and $I_{K1}$ channels [140]. Further work is necessary to evaluate the proarrhythmic risk of drugs that may mimic the short QT syndrome, and the utility of repolarization assays to evaluate clinical risk.

### 4.4 Isolated Tissue or Cell Preparations Used for Repolarization Studies

Repolarization studies have been conducted on a wide variety of isolated tissues employing intracellular recording techniques that record transmembrane potential. The mechanistic interpretation of repolarization assays may be complex, due to the multiple ionic currents contributing to the genesis of the action potential. The following sections discuss some repolarization assays in order of ascending level of complexity and integration.

***Isolated Cardiac Myocytes*** The cardiac myocyte represents the smallest integrated unit to evaluate ventricular repolarization. In contrast to tissues, intracellular recordings from myocytes can be attained using either sharp microelectrodes or patch pipettes. Due to intracellular dialysis associated with traditional patch pipettes, the action potential may be directly altered by the technique used, making interpretation of drug effects more difficult; this concern can be minimized through the use of the perforated patch technique [141]. Time to attain steady-state effects of drugs with isolated myocytes is typically more rapid than with superfused syncytial preparations due to diffusional barriers present in isolated tissues. However, isolated myocytes also demonstrate greater beat-to-beat variability compared to syncytial preparations, an effect attributable to a loss of cell-to-cell coupling in isolated myocytes [142]. Attention to the anatomical origin of isolated myocytes, beat-to-beat variability, as well as time and vehicle effects are necessary to properly interpret the results of repolarization assays using isolated cellular preparations. Representative repolarization studies are provided by Refs. 143–146.

***Purkinje Fibers and Papillary Muscles*** Purkinje fibers and papillary muscles are established, well-characterized preparations that have been widely used for electrophysiologic studies; intracellular recordings are typically obtained from these superfused preparations (see earlier discussion). Purkinje fibers are specialized conducting tissue in the ventricle. In ungulates and most mammals, fibers are free running across the ventricular surface, allowing for easy harvesting of tissues. In addition, paucity

of contractile elements makes fiber impalement easy to maintain. Studies have suggested that the assay using Purkinje fibers is able to detect most drugs linked to delayed repolarization and proarrhythmia [147]. Furthermore, the Purkinje fiber assay using fibers harvested from rabbits may be more sensitive than fibers harvested from dogs [147, 148]. Some drugs (such as cisapride) elicit bell-shaped concentration–response curves, with low to intermediate concentrations eliciting prolongation, and higher concentrations eliciting lesser (or shortening) of the APD [146, 147]; these studies highlight the complexity of effects on cardiac repolarization and the need to evaluate concentration–response curves in the context of clinical exposures. For drugs highly plasma protein bound, calculated free concentrations have been used during *in vitro* studies to approximate clinical exposures. However, this approach may not be realistic, particularly if the drug is readily exchanged from plasma into the myocardium, where it may accumulate. It is possible to superfuse tissues *in vitro* with solutions containing plasma proteins to more closely approximate drug exposures *in vitro*.

Guinea pig papillary muscles have also been used to evaluate drug effects on cardiac electrical activity [149]. These preparations are typically thinner than other cardiac muscle preparations, minimizing anoxia that may occur in the center of larger syncytial preparations during bath superfusion. Studies suggest a strong correlation between hERG channel blockade and APD prolongation for a validation dataset of 21 compounds [150].

***Ventricular "Wedge" Preparations*** Wedge preparations are created from arterially perfused, tangentially sliced transmural sections of ventricular free wall, providing exposed transmural surfaces for electrode impalements. This preparation allows for a comparison of drug effects from different regions of the myocardium (e.g., epicardium, midmyocardium, and endocardium of the left ventricular freewall of canine or rabbit [151, 152]. Due to the vigorous activity of these contracting preparations, flexibly mounted ("floating") microelectrodes are typically used to transiently record transmembrane potentials. Wedge preparations allow for the evaluation of delayed repolarization and dispersion of repolarization between myocardial regions—factors implicated in both the genesis of TdeP as well as waveforms resembling TdeP from ECG-like traces [153]. A recent study demonstrated that despite eliciting greater APD prolongation with azimilide compared to d,l-sotalol, EADs elicited by azimilide often failed to propagate transmurally, likely due to diminished dispersion with azimilide [154]. These preparations require a relatively long equilibration period and are challenging to maintain impalements, resulting in generally low throughput as screening assays.

### 4.5 Isolated Heart Preparations Used for Repolarization Studies

While it is difficult to obtain stable intracellular recordings from the epicardium of beating hearts, it is possible to measure changes in the action potential configuration recorded from isolated, perfused (Langendorff model) hearts using monophasic action potential techniques (see Section 4.3). Briefly, these preparations utilize hearts rapidly excised and retrogradely perfused through the aorta with buffered salt solutions [155]; the atrioventricular node is typically ablated to allow for pacing control of ventricular rate. However, the preparation may only be stable for a few

hours, especially if protein-free perfusates are used and hearts become gradually edematous [155]. Using isolated rabbit hearts, Hondeghem and colleagues [130, 156, 157] evaluated the effects of numerous compounds (many linked to proarrhythmia and TdeP) on changes in the MAP recordings from the epicardium. Based on this model, they proposed a set of repolarization parameters useful in detecting proarrhythmic risk that included *T*riangulation, *R*everse use dependence, *I*nstability, and *D*ispersion (referred to by the acronym TRIaD). Instability is defined as the extent of beat-to-beat variation in the action potential duration and was considered the most potent predictor of proarrhythmia in rabbit heart [157]. Dispersion of repolarization can be measured using multiple MAP electrodes, or (alternatively) indirectly using multiple volume-conducted ECG recordings. In blinded experiments, the SCREENIT model (system) with its multiple parameter evaluation provided a predictive qualitative assessment of potential proarrhythmia risk [156]. At present, only one laboratory routinely performs this TRIaD assay for proarrhythmic drugs; independent validation would be beneficial. In general, TRIaD is most useful for prioritizing drugs according to their probable liability for proarrhythmia, and not necessarily predicting a plasma concentration at which this liability may be manifested in the clinic.

### 4.6 *In Vitro* ECG (QT) Assays

Additional *in vitro* assays have been described that focus primarily on QT interval changes recorded from acutely isolated Langendorff perfused hearts *in vitro*. Pseudo-ECGs may be recorded with electrodes placed within the organ chamber in a configuration resembling that of the traditional ECG lead system. Alternatively, electrogram recordings may be obtained from electrodes held lightly against the epicardium of left and right ventricles to generate bipolar transcardiac electrograms [158]. These approaches afford direct measures of changes in the duration of the QT interval as well as discriminating changes in the T wave configuration to further discern proarrhythmic risk (an area of continuing research). Epicardial and endocardial MAPs may also be recorded, providing further information on repolarization dispersion. It is also possible to control heart rate by external pacing, thus providing the ability to evaluate QT interval changes at selected heart rates (thus avoiding the use of correction formulas, see Section 3). A number of published studies have demonstrating the sensitivity and specificity of isolated perfused hearts from rabbit [159, 160] and guinea pig hearts [158] to drugs that prolong the QT interval clinically.

### 4.7 *In Vitro* Proarrhythmia Models

While the proarrhythmia model of methoxamine-treated female rabbit hearts are generally recognized as the more sensitive model, other recognized risk factors may also be applied to exaggerate drug effects (e.g., hypokalemia). Using these torsadogenic models, the relationship between the specificity and magnitude of drug effects with human proarrhythmia is uncertain. Hearts from smaller mammals (rats, mice) may be inappropriate for these studies due to the minimal role of $I_{Kr}$/hERG current in facilitating ventricular repolarization in these species as well as the rapid heart rates compared to humans.

### 4.8 Conclusions

Numerous *in vitro* models representing various levels of complexity and integration have been described that detect a drug's ability to affect and delay ventricular repolarization, a surrogate marker of proarrhythmia and TdeP. Block of hERG current remains a cornerstone in this collection of *in vitro* cardiac safety testing. It should be recognized that the sensitivity of this assay to detect a positive signal of QT risk is high, while specificity may not be sufficiently high to avoid false positives (e.g., drugs such as verapamil that produce multichannel blockade). Repolarization assays provide evidence of more integrated responses, providing further insights into multichannel block and torsadogenic effects such as EADs, instability, triangulation, and reverse use dependence, as well as the increased dispersion of repolarization resulting from a heterogeneous electrophysiologic substrate. Finally, ECG recordings from isolated hearts provide for evaluation of direct effects on the duration (and configuration) of the QT interval. The challenge of the present is to improve the ability of preclinical studies to predict torsadogenic risk through the use and balanced interpretation of multiple models. The challenge of the future is to develop preclinical models that provide further mechanistic insights into the initiation and perpetuation of TdeP, as a direct measure of torsadogenic risk.

## 5 RESPIRATORY SAFETY PHARMACOLOGY

### 5.1 Introduction

A study designed to evaluate the potential for new drugs to produce adverse effects on respiratory function must first determine whether the drug can produce a change in respiratory function and then establish whether this change is a liability. Such changes can result either from the primary or secondary pharmacological properties of a drug or from organ dysfunction resulting from its toxicological properties. Drugs from a variety of pharmacological classes are known to adversely affect the respiratory system (see Tables 1 and 2). Most drug-induced effects on the lung involve the parenchyma, or alveolar region, and produce changes involving alveolar proteinosis, interstitial and intra-alveolar edema, as well as interstitial pneumonitis and fibrosis [161, 162]. Drugs can also influence the pleura and airways, producing effects that include pleural effusion, pneumothorax, bronchospasm, and bronchiolitis obliterans [161, 162]. Drugs from at least 23 different pharmacological classes are also known to influence the neuromuscular control of pulmonary ventilation. Such drugs can be classified as respiratory stimulants or depressants and produce their effects primarily through mechanisms involving the central nervous system [163, 164], peripheral arterial chemoreceptors [165, 166], or pulmonary chemo- and mechanoreceptors [167].

Agencies that regulate drug testing and marketing have recognized the importance of conducting respiratory function safety studies. The International Conference on Harmonization (ICH), which represents regulatory agencies from the United States, Europe, and Japan, has established guidelines for the safety testing of pharmaceuticals on organ functions. This document (ICH S7A), entitled *Safety Pharmacology Studies for Human Pharmaceuticals*, states that the respiratory system is a vital organ system whose function is critical for life and, as such, must

**TABLE 1 Drugs Known to Cause Bronchoconstriction and/or Pulmonary Injury**

| | |
|---|---|
| Chemotherapeutics | Analgesics |
| Carmustine (BCNU) | Heroin |
| Azathioprine | Morphine |
| Bleomycin | Methadone |
| Busulfan | Naloxone |
| Cyclophosphamide | Ethchlorvynol |
| Mitomycin | Propoxyphene |
| Nitrosoureas | Salicylates |
| Procarbazine | Cardiovascular drugs |
| Lomustine | Amiodarone |
| Methotrexate | ACE inhibitors |
| Cytosine arabinoside | Endothelin antagonists |
| Mercaptopurine | Anticoagulants |
| Procarbazine | Beta-blockers |
| Vinblastine | Fibrinolytics |
| Immunosuppressives | Dipyridamole |
| Cyclosporine | Protamine |
| Interleukin-2 | Tocainide |
| Antivirals | Adenosine agonists |
| Interferons | Miscellaneous drugs |
| Antibiotics | Muscarinic ($M_3$) agonists |
| Ampicillin | Serotonin agonists |
| Amphotericin B | Histamine ($H_1$) agonists |
| Cephalosporines | Ergotamine |
| Cyclines | Bromocriptine |
| Furazolidone | Paracetamol |
| Nitrofurantoin | Polymyxins |
| Sulfasalazine | Myorelaxants |
| Sulfonamides | Dantrolene |
| Pentamidine | Hydrochlorothiazide |
| Anti-inflammatories | Methylsergide |
| Gold salts | Oral contraceptives |
| NSAIDs | Tocolytic agents |
| Penicillamine | Tricyclics |
| Beclomethasone | Tazanolast |
| Acetylsalicyclic acid | |

be evaluated in safety pharmacology studies. As part of the core battery of safety pharmacology studies, the ICH guideline requires that the effect of drugs on respiratory function be evaluated prior to first human administration.

The objective of this section is to provide a brief overview of the functional disorders of the respiratory system and to present the strategy and techniques considered to be most appropriate for detecting and characterizing drug-induced respiratory disorders in nonclinical safety studies.

**TABLE 2 Drugs Known to Influence Ventilatory Control**

| | |
|---|---|
| Depressants | Tranquilizers/analgesics |
| Inhaled anesthetics | Chlorpromazine |
| Aspartic acid analogues | Rompum (xylazine) |
| Barbiturates | Nalorphine |
| Benzodiazepines | Antibiotics |
| Diazepam | Macrolides |
| Temazapan | Aminoglycosides |
| Flunitrazepam | Endothelins |
| Chlordiazepoxide | Nitric oxide synthase inhibitors |
| Serotonin analogues | Stimulants |
| Methoxy(dimethyl)tryptamine | Alkaloids |
| Dopamine analogues | Nicotine |
| Apomorphine | Lobeline |
| Adenosine analogues | Piperidine |
| 2-Chloroadenosine | Xanthine analogues |
| R-Phenylisopropyl-adenosine (R-PIA) | Theophylline |
| *N*-ethylcarboxamide (NECA) | Caffeine |
| β-Adrenergic antagonists | Theobromine |
| Timolol maleate | Analeptics |
| GABA analogues | Doxapram |
| Muscimol | Salicylates |
| Baclofen | Progesterone analogs |
| Calcium channel blockers | Almitrine |
| Daurisoline | Glycine analogues |
| Antimalarials | Strychnine |
| Chloroquine | GABA antagonists |
| Opiates | Picrotoxin |
| Morphine | Bicuculline |
| Codeine | Serotonin synthesis inhibitors |
| Methadone | ρ 18-Chlorophenylalanine |
| Meperidine | Reserpine |
| Phenazocine | |
| Anticholinergics | |
| $M_1$ selective | |
| $M_2$ selective | |

## 5.2 Functional Disorders of the Respiratory System

The respiratory system can be divided functionally into a pumping apparatus and a gas exchange unit. The pumping apparatus includes those components of the nervous and muscular systems that are responsible for generating and regulating breathing patterns, whereas the gas exchange unit consists of the lung with its associated airways, alveoli, and interstitial area that contains blood and lymph vessels and an elastic fibrous network.

***Pumping Apparatus*** The pumping apparatus of the respiratory system is responsible for generating and regulating the pleural pressures needed to inflate and deflate the lung. Normal gas exchange between the lung and blood requires breath-

ing patterns that ensure appropriate alveolar ventilation. Ventilatory disorders that alter alveolar ventilation are defined as hypoventilation or hyperventilation syndromes. Hypoventilation is defined as an increase in the partial pressure of arterial $CO_2$ ($Pa_{CO_2}$) above normal limits and can result in acidosis, pulmonary hypertension, congestive heart failure, headache, and/or disturbed sleep [168]. Hyperventilation is defined as a decrease in $Pa_{CO_2}$ below normal limits and can result in alkalosis, syncope, epileptic attacks, reduced cardiac output, and/or muscle weakness [169, 170].

Mechanisms responsible for ventilatory disorders may involve muscle and/or neural dysfunction. Those dysfunctions involving the nervous system are categorized as either central or peripheral. Central mechanisms involve the neurological components of the pumping apparatus that are located within the central nervous system and include the medullary central pattern generator (CPG) and the integration centers that regulate the output of the CPG. Integration centers are located in the medulla, pons, hypothalamus, and cortex of the brain [171]. The major neurological inputs from the peripheral nervous system that influence the CPG are arterial chemoreceptors [171] and respiratory tract chemo- and mechanoreceptors [161]. Drugs can stimulate or depress ventilation by selective interaction with the central nervous system [163, 164, 172], arterial chemoreceptors [165, 166], and/or espiratory tract chemo- and mechanoreceptors [161].

***Gas Exchange Unit*** The primary function of the respiratory system is the exchange of gases ($O_2$ and $CO_2$) between the environment and blood. Functional disorders of the respiratory system are defined, therefore, as changes in the respiratory system that reduce the efficiency of gas exchange. Airway narrowing or blockade can reduce the efficiency of gas exchange by decreasing or preventing airflow between the upper or central airways and the alveoli. Functional changes of this type are called obstructive disorders [173]. Airway narrowing can result from airway smooth muscle contraction or hypertrophy, accumulation of mucus or edema fluid in the airway, luminal cell hyperplasia, or breakdown in elastic support tissue [174]. An obstructive disorder also creates a resistive load during ventilation and, by increasing the work of breathing, can lead to respiratory fatigue and failure [175].

A reduction in the normal elasticity, or a "stiffening," of the lung increases the work (or pressure) required to expand the lung. This decrease in lung compliance generally produces a rapid shallow breathing pattern [175], which can reduce ventilatory efficiency by lowering alveolar ventilation and, ultimately, can lead to respiratory fatigue and failure [174]. Functional changes of this type are called restrictive disorders and can result from changes involving interstitial thickening due to edema, hemorrhage, cellular infiltration, or fibrosis, or from intra-alveolar changes caused by edema, cellular infiltration, or abnormal surfactant production [176]. These interstitial and intra-alveolar changes also affect lung function by increasing the diffusion barrier for $CO_2$ and $O_2$ between the pulmonary blood vessels and the alveolar surface [176].

### 5.3 Techniques for Measuring Changes in Respiratory Function

***Ventilatory Disorders*** A plethysmograph chamber or facemask with a pneumotachometer is generally used to measure ventilatory patterns in humans and con-

scious animal models. Plethysmograph chambers that most accurately measure ventilatory patterns are designed to directly monitor lung volume changes or airflows generated by thoracic movements. Such chambers isolate openings to the upper airways (mouth and nose) from the thorax and are referred to as "head-out" or "head-enclosed" chambers (Fig. 3). The pressure changes in these chambers are a direct measure of lung airflow and, consequently, can accurately measure both the duration and volume of each breath (Fig. 4). These measurements are required for determining the various parameters needed to characterize breathing patterns (see below). Whole body (barometric) plethysmograph chambers are often used to monitor ventilatory patterns and offer the advantage that the subject does not have to be restrained [177] (Fig. 5b). This method, however, is limited in that the volume of each breath is determined indirectly from the relatively minor increase in chamber pressure resulting from the expansion of inhaled gases due to a decrease in intrapulmonary pressure and to increases in intrapulmonary temperature and humidity. Furthermore, the reliability of this estimated tidal volume is rather low since it is highly dependent on a number of variables including breathing rate and pattern, chamber temperature and humidity, and nasal air temperature [178]. Altering these variables can produce significant effects. For example, increasing the breathing rate of a rat from 40 to 70 breaths per minute caused an underestimation of actual tidal volume by up to 30% [178]. Other indirect methods for estimating lung volume changes in animal models include the use of induction, bioimpedence, or sonomicrometry. These methods utilize induction coils, biopotential electrodes, or piezoelectric crystals either placed in a strap on the skin surface or implanted subcutaneously to measure the linear expansion of the thorax and abdomen during breathing. These methods can accurately measure respiratory rate; however, they are limited in that lung volume is generally estimated from a single dimensional measurement. As such, these estimated volume measurements are highly dependent of body position (e.g., standing, sitting, or recumbent), which can alter the three-

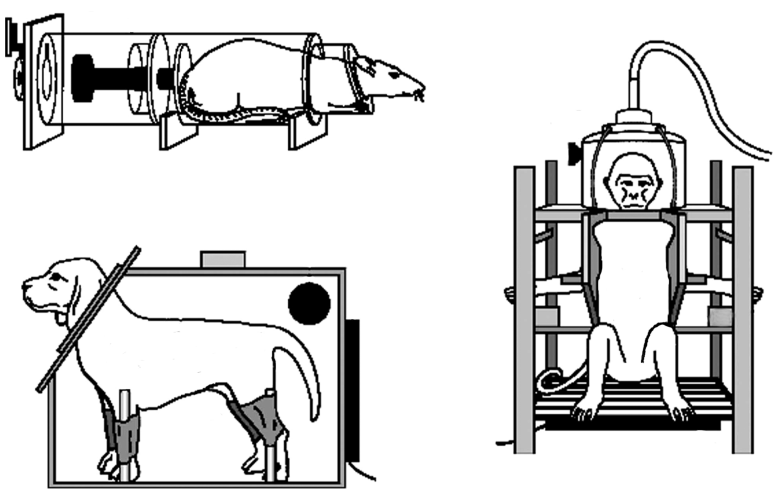

**FIGURE 3** Plethysmograph chambers for the direct measurement of ventilatory parameters. The rat and dog are in "head-out" chambers, while the monkey is in a "head-enclosed" chamber. For all chambers, an airtight seal is made around the neck using a rubber collar.

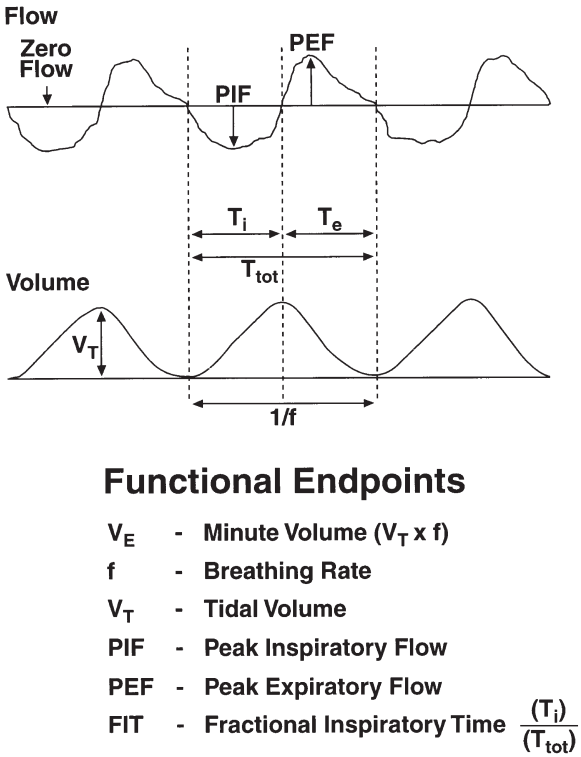


**FIGURE 4** Tracings of lung airflow and lung volume changes during spontaneous breathing in a conscious rat. Airflow was measured directly using a "head-out" plethysmograph chamber. The functional endpoints are automatically calculated for each breath using an analog/digital converter and a data acquisition and analysis software system.

dimensional shape of the thorax or abdomen [179]. The accuracy and reliability of these methods have not yet been demonstrated in animal models.

Because most anesthetics, analgesics, and sedatives alter ventilatory reflexes and drive, studies evaluating the effects of drugs on ventilation should use conscious animals. It is also essential that animals be appropriately acclimated to the experimental apparatus to ensure normal (resting) ventilatory patterns. Proper characterization of a breathing pattern must include measurements of respiratory rate, tidal volume, and minute volume. Other parameters such as inspiratory flow (peak or mean), expiratory flow (peak or mean), inspiratory time, expiratory time, and total breath time can be used to help define mechanisms of ventilatory change (Fig. 4).

Respiratory rate, tidal volume, and minute volume are essential measurements since normal ventilation requires that the pumping apparatus provide both adequate total pulmonary ventilation (minute volume) and the appropriate depth (tidal volume) and frequency of breathing. The depth and frequency of breathing required for alveolar ventilation are determined primarily by the anatomic dead space of the lung. In general, a rapid shallow breathing pattern (tachypnea) is less efficient than a slower deeper breathing pattern that achieves the same minute volume. Thus, any

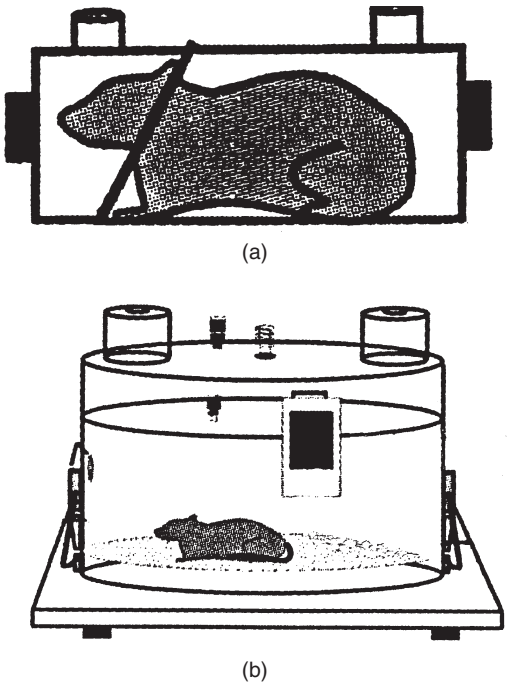


**FIGURE 5** Plethysmograph chambers for obtaining indirect measures of airway resistance: (a) a dual-chamber (Pennock box) that separates the head and thorax into two separate compartments; (b) a whole-body (barometric) chamber.

change in minute volume, tidal volume, or the rate of breathing can influence the efficiency of ventilation [175].

An additional reason for measuring both respiratory rate and tidal volume is that these parameters are independently controlled [171] and can be selectively altered by drugs. Thus, monitoring changes in tidal volume or respiratory rate alone cannot be used to evaluate drug-induced effects on pulmonary ventilation. For example, the phosphodiesterase inhibitor and adenosine antagonist theophylline increases minute volume by increasing tidal volume without affecting respiratory rate [180], whereas the opioid morphine depresses minute volume by reducing respiratory rate without affecting tidal volume [181, 182]. Other drugs such as adenosine agonists depress respiratory rate but have no effect on minute volume since there is a compensatory increase in tidal volume [181].

The inspiratory and expiratory phases of individual breaths are also independently controlled [171] and, by characterizing changes in these parameters, mechanisms responsible for ventilatory changes can be determined. Thus, by measuring changes in the airflow rate and duration of each of these phases, mechanisms responsible for changes in tidal volume or respiratory rate can be identified [175]. For example, a selective decrease in airflow during inspiration (the active phase) is generally indicative of a decrease in respiratory drive, whereas a selective decrease in airflow during expiration (the passive phase) is generally indicative of an obstructive disorder. Furthermore, a change in the ratio of inspiratory to expiratory time

can be used to differentiate effects on stretch and irritant pulmonary receptors [171]. The examples presented above clearly demonstrate that measurements of both respiratory rate and tidal volume are critical for the evaluation of drug-induced effects on pulmonary ventilation and that airflow and time measurements of individual breaths are needed to help define mechanisms of change in ventilatory control.

Moderate to marked changes in pulmonary ventilation can result in hypo- or hyperventilation syndromes. In conscious rats and monkeys, decreases of approximately 30% or greater in total pulmonary ventilation can produce hypoventilation [181, 183]. To establish the occurrence of these syndromes, $Paco_2$ is measured. In humans and large animals, this can be accomplished readily by collecting arterial blood with a catheter or needle and analyzing for $Paco_2$ using a blood gas analyzer. In conscious rodents (or other small animals), however, obtaining arterial blood samples by needle puncture or catheterization is generally disruptive of normal breathing patterns and often invalidates the measurement. An alternative and non-invasive method for monitoring $Paco_2$ is the measurement of peak expired (end-tidal) $CO_2$ concentrations. This technique has been used successfully in humans [184], pigs [185], and conscious rats [181].

### *Obstructive Disorders*

*Forced Maneuvers* An obstructive disorder is defined functionally by the American Thoracic Society (ATS) as "a disproportionate reduction of maximum airflow from the lung with respect to the maximum volume that can be displaced from the lung" [186]. In human clinical studies, obstructive disorders are detected by measuring maximum (forced) expiratory airflows following maximal inspiration. Spirometry is used to measure volume and flow changes, and the functional endpoints include forced expiratory volume after 1 second (FEV1), peak expiratory flow (PEF), forced expiratory flows at 25 and 75% of expired volume ($FEF_{25\%}$ and $FEF_{75\%}$), mean mid-expiratory flow (MMEF) and total expired gas volume or forced vital capacity (FVC). Decreases in $FEV_1$, PEF, and $FEF_{25\%}$ indicate central (or large) airway narrowing, while $FEF_{75\%}$ and MMEF are used to detect peripheral (or small) airway narrowing [186]. Since airflow is dependent on the starting volume, the flow parameters are normalized to FVC. A disproportionate reduction in flow relative to lung volume distinguishes an obstructive from a restrictive disorder [186].

Experimental systems that can simulate the flow–volume maneuver performed by humans have been developed for use in animals. A pressure panel has been developed that controls lung inflation and deflation rates of anesthetized animals (Buxco Electronics, Inc., Sharon, CT). The animal's lung is inflated to a maximum volume or inspiratory capacity (IC), as determined by the inflation pressure (generally 30 cm $H_2O$), and then is rapidly deflated by switching to a vacuum reservoir (generally at –50 cm $H_2O$) using a solenoid valve. A plethysmograph chamber is required to measure lung volume and airflow changes, and a pressure-sensitive catheter is required to measure airway or transpulmonary (i.e., airway pressure – pleural pressure) pressure changes. Details of the equipment used for conducting this maneuver in rats and dogs have been described [187, 188]. Analyzers are also available (Buxco Electronics, Inc.) that will calculate IC, FVC, FEV, PEF, MMEF, and FEF values for each flow–volume maneuver. Evaluating obstructive disorders

in animals using the flow–volume maneuver offers a high degree of sensitivity since changes are measured at maximum attainable flow rates and lung volumes. In addition, this procedure provides functional endpoints similar to human measurements. The disadvantages of this procedure are that it generally requires an anesthetized/paralyzed animal and is limited to a single time point measurement. The latter limitation excludes this as a procedure for measuring acute dynamic changes in lung airflow as might occur following administration of a drug given intravenously or by inhalation.

*Dynamic Measurements—Direct* Drug-induced lung airflow disorders can also be evaluated in animal models by measuring total pulmonary resistance or conductance during tidal breathing or mechanical ventilation (i.e., dynamically) (Fig. 6). Total pulmonary resistance defines the change in pleural, airway, or transpulmonary pressure ($\Delta P$) required to produce a defined change in lung airflow ($\Delta F$) and is calculated as $\Delta P/\Delta F$, while conductance is calculated as $\Delta F/\Delta P$. To determine dynamic resistance or conductance, the $\Delta P$ and $\Delta F$ are measured for each breath at the same lung volume during inspiration and expiration (usually between 50% and 70% of tidal volume). By selecting isovolumetric points, the dependence of $\Delta P$ on the elastic component of the lung is removed, leaving $\Delta P$ dependent on resistance to the flow of gas in the airways. This procedure requires the use of a plethysmograph chamber

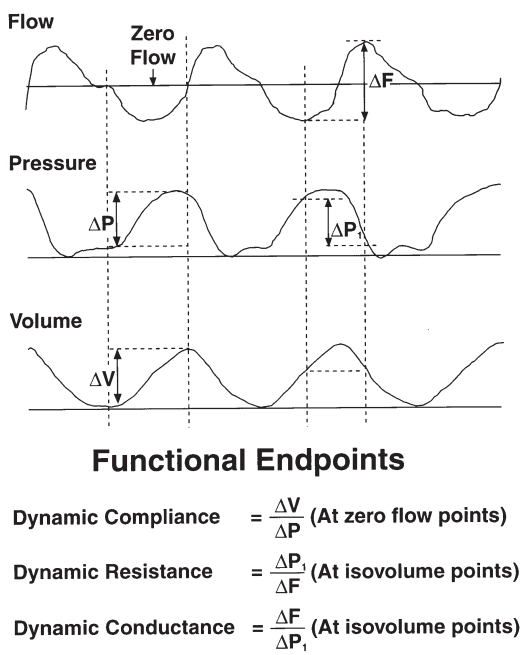


**FIGURE 6** Tracings of lung airflow, transpulmonary pressure, and lung volume changes during spontaneous breathing in a rat. Airflow was measured directly using a "head-out" plethysmograph chamber, while pleural pressure was measured using a pressure-sensitive catheter placed into the esophagus within the thoracic cavity. The functional endpoints are automatically calculated for each breath using an analog/digital converter and a data acquisition and analysis software system.

or facemask with pneumotachometer for volume and flow measurements, and a pressure-sensitive catheter for the measurement of pleural, airway, or transpulmonary pressure (Fig. 6). The methods for conducting these measurements have been established in guinea pigs [189], rats [190], and dogs [191].

Because dynamic measurements of airflow resistance in a spontaneously breathing animal require a pressure-sensitive catheter to be inserted into the esophagus or pleural cavity, animals are generally anesthetized and the measurements are collected over a period of minutes to hours. Attempts to develop a method for measuring pleural pressure chronically in conscious animals have involved surgical implantation of pressure-sensitive catheters directly into the pleural cavity. The success of these techniques has been limited by lung damage [189] and/or tissue growth and encapsulation of the pressure-sensitive catheter with a damping or loss of the signal [189, 192, 193]. To overcome these limitations, Murphy et al. [194, 195] developed a novel surgical procedure for placement of a pressure-sensitive catheter beneath the pleural surface. The catheter (attached to a radiotelemetry transmitter) is surgically implanted beneath the serosal layer of the esophagus within the thoracic cavity (Fig. 7). Using this technique, the pleural pressure changes have remained constant for at least 1 year in rats and 2 years in monkeys following surgery (personal observation). By combining this procedure with the use of a plethysmograph chamber, dynamic measurements of lung airflow resistance can

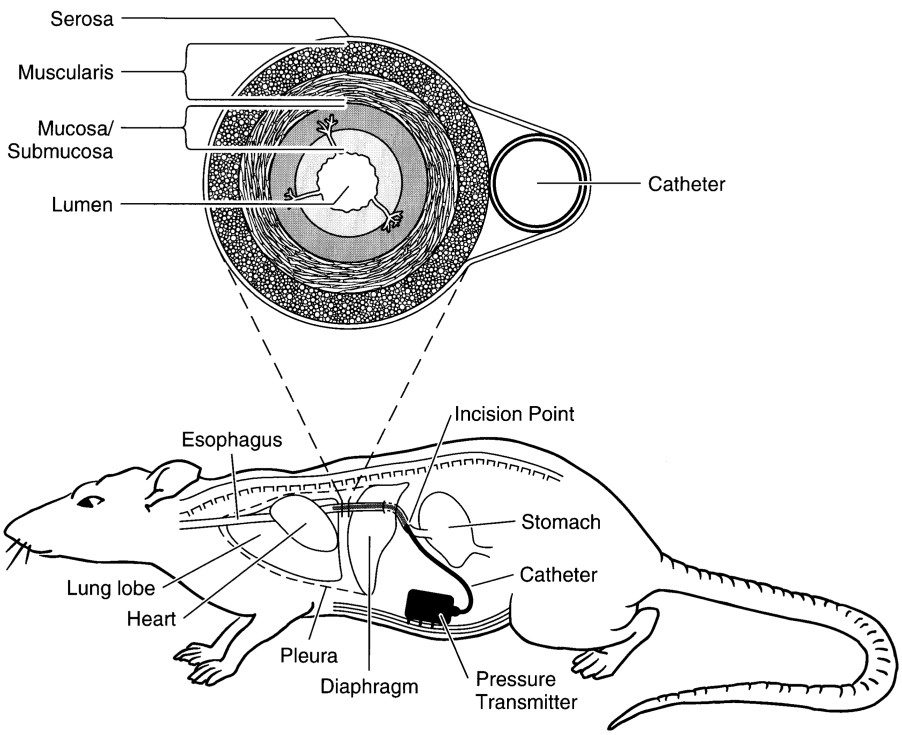


**FIGURE 7** Drawing of a rat showing placement of the pressure-sensitive subpleural catheter and radiotelemetry transmitter for chronic measurement of pleural pressure in conscious animals. The enlargement is a cross section through the esophagus showing the position of the catheter between the serosal and muscularis layers.

be obtained repeatedly and chronically in conscious animals (Fig. 3). This procedure has been used successfully in rats [194], monkeys [195], and dogs (personal experience, data not published).

*Dynamic Measurements—Indirect* Noninvasive procedures have been developed that can detect changes in resistance to lung airflow indirectly in conscious animals by measuring changes in airflow patterns. A plethysmograph chamber that separates the head and trunk into separate compartments has been used to measure the time delay between thoracic and nasal airflows in a spontaneously breathing animal [196] (Fig. 5a). A prolongation of this delay is related to the product of airway resistance and thoracic gas volume (specific airway resistance). The primary limitation of this procedure is that it is an indirect measure of airway resistance with the estimated value (specific airway resistance) dependent on a number of variables including respiratory rate, lung volume, and the temperature and humidity differences in the head and trunk chambers. Furthermore, the reliability of specific airway resistance measurements appears to decrease as resistance to lung airflow increases [196].

A procedure using barometric whole-body plethysmography has been developed for noninvasively detecting changes in the resistance to lung airflow of conscious, nonrestrained animals [197] (Fig. 5b). Estimated airway resistance (airway responsiveness) is determined for each breath by calculating an expiratory pause, which is dependent on differences in peak inspiratory and expiratory flows, and the duration of the early and late phases of expiration. This measurement is not influenced significantly by changes in respiratory rate and lung volume and appears to be a reliable method for measuring acute, severe bronchoconstriction. However, the scientific validity of this method has been challenged [198]. Furthermore, the method appears to be less sensitive than direct measurements of lung airflow resistance [197], false-positive results can occur for drugs that alter flow patterns by influencing respiratory drive, and the value of this procedure for chronic (repeated) measurements of mild to moderate lung airflow changes has yet to be established.

***Restrictive Disorders*** A restrictive disorder is defined functionally either as a reduction in total lung capacity (TLC) or as a decrease in lung compliance [176, 186, 199]. In human clinical studies, TLC is determined either by using a gas dilution technique or by combining the volumes obtained for inspiratory capacity (IC) and functional residual capacity (FRC) [200]. IC is the increase in volume of gas in the lung following inflation from resting expiration to maximum inspiration and is measured using spirometry. FRC is the volume of gas remaining in the lung at resting expiration and is obtained by a maneuver that involves breathing against an occluded airway at resting expiration and recording the pressure and volume changes generated. Boyle's law ($PV = P_1V_1$) is used to calculate the volume of gas at end-expiration [201]. Lung compliance is a measure of the change in airway or transpulmonary pressure ($\Delta P$) resulting from a given change in lung volume ($\Delta V$) and is calculated as $\Delta V/\Delta P$. A pressure-sensitive catheter is required to measure airway or transpulmonary pressure, and a plethysmograph chamber or facemask with pneumotachometer is used to measure lung volume.

*Forced Maneuvers* Forced maneuver procedures are available for determining TLC [41], FRC [202], IC [188], and quasistatic lung compliance [188] in animal

models. A pressure panel has been developed (Buxco Electronics, Inc., Sharon, CT) that can perform a pressure–volume maneuver in anesthetized animals. The animal's lung is inflated to IC, as determined by the inflation pressure (generally 30 cm $H_2O$), and then slowly deflated at a constant rate using a solenoid valve and a vacuum reservoir. During deflation, lung volumes and airway or transpulmonary pressures are measured. This technique also requires a plethysmograph chamber for measuring lung volumes and a pressure-sensitive catheter for measuring airway or transpulmonary pressures. Details of the equipment used for conducting this maneuver in rats and dogs have been described [187, 188]. Analyzers are also available (Buxco Electronics, Inc., Sharon, CT) that will calculate IC, FRC, TLC, and quasistatic lung compliance values during each pressure–volume maneuver.

Evaluating restrictive disorders in animals using the pressure–volume maneuver offers a high degree of sensitivity since changes are measured at maximum attainable lung volumes and transpulmonary pressures. In addition, this method provides functional endpoints similar to human measurements (e.g., IC, FRC, TLC, and compliance). As described under the section on resistance measurements, the limitations of this technique are that it generally requires the use of anesthetized/paralyzed animals and is a single time point measurement. The latter limitation excludes this as a procedure for measuring acute dynamic changes in lung compliance as might occur following administration of a drug given intravenously or by inhalation.

*Dynamic Measurements* Restrictive disorders can also be evaluated in animal models by measuring dynamic lung compliance during tidal breathing or mechanical ventilation. Dynamic compliance is obtained by measuring the differences in airway or transpulmonary pressure ($\Delta P$) and volume ($\Delta V$) that occur at the beginning and end of each inspiration (i.e., at zero flow points) (Fig. 6). By selecting zero flow points, the dependence of $\Delta P$ on tissue and airflow movement is removed, leaving $\Delta P$ dependent on the elastic component of the lung. This procedure requires the use of a plethysmograph chamber or facemask with pneumotachometer for volume measurements and a pressure-sensitive catheter for airway or transpulmonary pressure measurements. Methods for measuring dynamic lung compliance acutely have been established in guinea pigs [29], rats [190], and dogs [191]. Murphy et al. [194, 195] recently developed a procedure that allows for chronic measurement of dynamic compliance in conscious animals.

Dynamic lung compliance measurements are required for evaluating rapid, transient changes in compliance. This procedure can be used in conscious animals, which also allows for repeated measurements over prolonged periods of time. Changes in dynamic lung compliance, however, need to be interpreted with caution since changes in small airway patency can alter dynamic compliance measurements by changing the time constant for alveolar filling [187].

## 5.4 Design of Respiratory Function Safety Studies

### *General Approach*

*Tier 1 Evaluation (Core Battery Safety Pharmacology Studies)* Because the respiratory system consists of two functional units—the pumping apparatus and the gas exchange unit—a study designed to evaluate the potential for new drugs to produce

adverse effects on respiratory function must evaluate both of these components. Evaluating dose–response relationships, determining the degree of change in specific respiratory parameters, and characterizing changes in arterial blood gases and pH are used to establish the liability associated with the functional changes. Because the safety profile defined by this type of study can have a significant impact on the successful development of new therapeutic agents, it is important that the techniques and assays used in this type of study minimize the occurrence of false-negative and false-positive results. For this reason, techniques that provide direct measures of respiratory parameters should be used. A direct measure is one that provides the endpoint of interest, in contrast to an indirect measure that provides a surrogate endpoint, which requires certain assumptions and/or calculations to estimate the true endpoint. Furthermore, because most drugs are intended for use in conscious patients, and because most anesthetics, analgesics, and sedatives can alter ventilatory reflexes, respiratory drive, and airway reactivity, nonclinical studies evaluating the effects of drugs on respiratory function should utilize conscious animal models. Nonclinical safety studies evaluating the potential adverse effects of new therapeutic agents on respiratory function should be conducted prior to human exposure since acute changes in respiratory function can be life threatening [7].

The core (Tier 1) test for evaluating drug-induced effects on the neuromuscular components of the pumping apparatus is the measurement of ventilatory patterns in conscious animals (Fig. 8). Parameters to be initially measured should include respiratory rate, tidal volume, and minute volume. If a change in these parameters occurs, inspiratory flow (mean or peak), expiratory flow (mean or peak), fractional inspiratory time (inspiratory time/total breath time), and time between breaths (expiratory pause or apnea) should be evaluated to help define the mechanism. Measurement of arterial gases ($Pa_{O_2}$, $Pa_{CO_2}$) and arterial pH should also be conducted to help define the extent or severity of the ventilatory change.

The Tier 1 test for evaluating drug-induced effects on the gas exchange unit is the measurement of the mechanical properties of the lung (Fig. 8). This is most effectively accomplished by obtaining the dynamic measurements of total pulmonary resistance and lung compliance. The advantages of this method are that measurements can be obtained repeatedly and continuously in conscious animals and that both ventilatory and lung function parameters can be evaluated simultaneously. Evaluating a drug-induced effect on both ventilation and lung mechanics will establish whether a drug can alter respiratory function and will determine whether a ventilatory change results from pulmonary or extrapulmonary factors.

*Tier 2 Evaluation (Follow-up Studies)* Follow-up (Tier 2) respiratory function tests should be initiated to further investigate mechanisms or liabilities (Fig. 8) [7]. Studies to investigate mechanisms of ventilatory change should initially determine whether the effect was due to respiratory muscle or neural function change. This can be accomplished by assessing the effect of drug treatment on efferent phrenic nerve activity [203, 204]. Changes in phrenic nerve activity (action potential burst frequency or intensity) are indicative of an effect on the nervous system. Direct stimulation of the phrenic nerve and assessment of its effect on respiratory muscle function (e.g., transdiaphragmatic function) can also be used to detect muscle defects [205]. A neurological effect is further classified as either central or peripheral. Peripheral nervous system effects are generally mediated by peripheral arterial

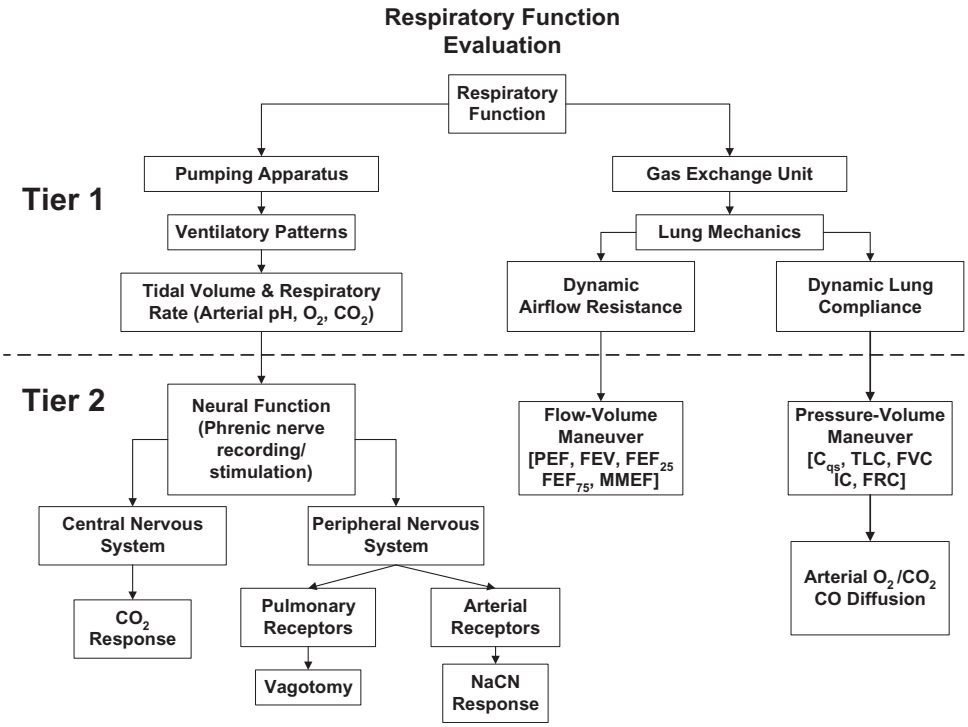


**FIGURE 8** Flowchart showing the strategy for evaluating drug-induced effects on respiratory function. Tier 1 tests are conducted to detect and quantify functional changes of the respiratory system, while Tier 2 tests are conducted to investigate mechanisms or further characterize effects associated with a functional change. Abbreviations are described in the text.

chemoreceptors or respiratory tract chemo- and mechanoreceptors. Exposure to $CO_2$ gas stimulates ventilation primarily through a central mechanism [206], while injection of NaCN produces a transient stimulation of ventilation through a mechanism that involves selective stimulation of peripheral chemoreceptors [166]. Based on these responses, a noninvasive method for distinguishing central from peripheral nervous system effects in conscious animals has been developed [181]. Cervical vagotomy can be used to determine whether pulmonary chemo- or mechanoreceptors are involved in chemical-induced effects on ventilatory control (e.g., see Ref. 48).

Tier 2 studies to further investigate lung mechanical changes include forced maneuver procedures that can distinguish central from peripheral airway obstruction [161], identify the presence of gas trapping (FRC) [164], establish changes in lung compliance [163], and provide measures of lung capacities [164]. Acute drug-induced changes in lung compliance often result from pulmonary edema and/or interalveolar septal thickening [161, 162]. Procedures for identifying these effects include measurement of CO diffusion from alveoli to blood or changes in the ratio of arterial $P_{O_2}$ to $P_{CO_2}$ [188, 190, 207]. Decreases in CO diffusion or the ratio of arterial $P_{O_2}/P_{CO_2}$ are generally indicative of pulmonary edema, thickening of interalveolar septa, or the loss of normal alveolar architecture [161, 162].

***Dosing and Dose Level Selection*** Nonclinical safety studies of respiratory function are designed to support initial administration of a new chemical entity to humans. The primary safety concerns in the initial single rising dose human studies are acute life-threatening changes in major organ systems. As such, single-dose studies are generally appropriate for addressing these concerns. However, respiratory disorders that are associated with multiple-dose treatment can occur and should be addressed in a multiple-dose safety study. Since pharmacological and toxicological effects can be route specific, the test drugs should be given by the intended clinical route. Furthermore, to evaluate dose–response relationships, three or more dose levels and a vehicle control should be used. If possible, selection of the high dose should be based on acute toxicological findings and should be a dose that produces evidence of toxicity. This dose level, however, must not be associated with toxicological changes in secondary organ systems, which may compromise the respiratory function measurements. For example, effects such as emesis, hyperactivity, tremors, or muscle hypertonia would be unacceptable, whereas a slight increase in serum ALT or a slight decrease in body weight is generally acceptable. Selecting the maximum possible dose level is important for safety evaluation since secondary pharmacological effects are often associated with nonspecific receptor interactions where affinity for the drug may be 100–1000-fold lower than the intended target (e.g., receptor, ion channel, or enzyme). The middle and low doses should generally be log or half-log decrements of the high dose, with the low dose no less than that required to produce systemic exposure similar to that at the primary pharmacological or anticipated human dose. In the absence of toxicology data, multiples of the pharmacological dose (e.g., $ED_{50}$) should be used. If toxicity is not limiting, the highest multiple should be at least 100 times the pharmacologically active dose.

***Species Selection*** The rat is considered an appropriate species for general use in studies designed to evaluate the potential adverse effects of drugs on respiratory function. The physiology of the respiratory system has been well characterized in this species [208, 209] and much of the information on drug-induced effects on ventilatory control and mechanisms of airway disease has been obtained in the rat [164, 210]. The rat is also readily available from animal vendors, is easy to handle and train, has a relatively stable breathing pattern, and has the appropriate temperament for conscious respiratory measurements. Furthermore, the techniques for measuring respiratory functions in rats are well established, and the rat is commonly used in toxicology studies. For safety pharmacology studies that are conducted after a compound has been selected for development, selecting a species that is used in toxicology studies provides additional supportive information including pharmacokinetic data [161] that can be used to define the test measurement intervals, acute toxicity data [162] that can be used to select the appropriate high dose, and toxicology/pathology findings [163] that can be used to help define the mechanism of the functional changes measured in safety pharmacology studies.

Alternate species may be required to address specific study requirements. For example, the development of humanized monoclonal antibodies or other biotechnology-derived proteins requires species that have homologous target proteins and do not produce an antigenic response to the drug. In such cases, transgenic mice or nonhuman primates are generally required. Similarly, if the test compound has known histamine or leukotriene activity, the guinea pig should be considered.

## 6 CENTRAL NERVOUS SYSTEM (CNS) SAFETY PHARMACOLOGY

### 6.1 Introduction

For many years the principal official recommendations for safety pharmacology studies were contained in the Japanese guidelines, the most recent version of which was published in 1995 [4]. The latest guidelines are those proposed by the *European Agency for the Evaluation of Medicinal Products*, ICH S7A [211], which have been in operation in Europe since June 2001 and have also been adopted by the United States and Japan. These guidelines describe a core battery of CNS studies that can be considered mandatory and include "motor activity, behavioral changes, coordination, sensory/motor reflex responses and body temperature" with the remark that "the central nervous system should be assessed appropriately." Core battery studies should be carried out before proceeding to first studies in humans (Phase I) and it is recommended that they be conducted in compliance with internationally recognized standards of Good Laboratory Practice (GLP). A further series of studies, follow-up or supplementary studies, should be carried out where indicated and include "behavioral pharmacology, learning and memory, ligand-specific binding, neurochemistry, visual, auditory and/or electrophysiology examinations." In general, follow-up studies are conducted prior to product approval and, because of their unique characteristics, require only assurance of "data quality and integrity" rather than full compliance with GLP.

The following sections describe CNS studies of both kinds. For further discussion the reader is referred to other publications [212, 213] on this subject.

### 6.2 Core Battery CNS Safety Pharmacology Procedures

Core battery CNS procedures are typically simple tests, using traditional techniques, which can be carried out rapidly in a routine fashion. They are the first techniques to be employed in CNS safety assessment. Indeed, they are frequently applied at the very beginning of the discovery process as a screen to eliminate substances with a potential for CNS risk. There is, however, a difference in emphasis between early CNS safety screening and core battery CNS safety tests as defined by ICH S7A. Initial screening studies are more exploratory in nature, investigate a wider range of doses using a limited number of procedures, and are seldom conducted according to GLP. In contrast, pre-Phase I core battery studies investigate a more restricted dose range selected in relation to the doses active in the primary indication. Whereas early screening is frequently conducted in the mouse because smaller quantities of test substance are required, pre-Phase I core battery studies are usually conducted in the rat to provide data more comparable with that obtained in other domains (toxicology, pharmacokinetics, biochemistry). At this stage of drug development, the required quantity of test substance is not a primary consideration.

***ICH S7A requirements*** ICH S7A recommends that a core battery should cover a range of simple tests measuring the effects of test substances on motor activity, behavior, coordination, sensory/motor reflex responses and body temperature. These effects can be covered by the battery of tests described below. In addition, we believe that it is useful, for little extra cost, to evaluate the effects of test substances on the convulsive threshold and their interaction with a standard CNS depressant

such as a barbiturate. The former is primarily useful for detecting proconvulsive activity whereas the later is very sensitive for unmasking latent sedative or psychostimulant effects. Both kinds of test were included in the Japanese guidelines but were dropped in ICH S7A.

*General Behavioral Observation* A first approach is to assess the global behavioral profile of a test substance using either a systematic observation procedure following that originally described in the mouse by Irwin [214] or the more recently described functional observation battery (FOB) [215]. Both are specifically mentioned in ICH S7A, but the FOB is more frequently used for assessing neurotoxicity and is less suitable for serial observations over several hours. As psychopharmacologists rather than toxicologists, we have preferred to use the Irwin test for initial screening of novel substances because of its rapidity and simplicity. In screening applications, the test substance is usually investigated over a wide dose range (6–8 doses) with few animals per dose ($N = 3$) investigated sequentially, starting at a high and potentially lethal dose and working upwards or downwards according to a bracketing prodecure to determine the first lethal dose, the active dose range, and the principal behavioral and physiological effects. In core battery applications, the test substance is usually evaluated over a smaller predetermined dose range (3–4 doses) with more animals per dose ($N = 4$–10), ranging from the dose known to be active in tests for the primary indication and ascending to a given multiple (30–100 times the low dose depending on feasibility). In a core battery evaluation, the test substance is rarely administered at potentially lethal doses because these are already known from available toxicological data.

In the Irwin procedure, rats or mice are given the test substance and are repeatedly observed over a 3–4 hour period followed by daily observations up to 72 hours using a standardized observation grid containing most or all of the following items: mortality, sedation, excitation, stereotypies, aggressiveness, reaction to touch, pain sensitivity, muscle relaxation, loss of righting reflex, changes in gait and respiration, catalepsy, ptosis, corneal reflex, pupil diameter, and rectal temperature.

The effects of a standard benzodiazepine (diazepam) in the Irwin test in the rat are shown in Table 3.

*Locomotor Activity Tests* Locomotor activity can be quantified in rodents in a variety of ways (interruptions of photoelectric beams, activity wheels, changes in electromagnetic fields, Doppler effects, video-image analysis, telemetry) [216]. Animals are administered the test substance and are placed in standardized enclosures for a limited observation period (10–40 minutes). Unlike the Irwin test, where animals are observed continuously from the moment of administration, locomotor activity tests are usually conducted at a fixed time after administration, the time being selected as a function of the test substance kinetics to approximate maximum exposure during the test. This avoids interpretational difficulties related to interactions between the onset of drug action and behavioral habituation to the test environment. In contrast to the Irwin test, which is labor intensive, most locomotor activity tests are automated and permit several animals to be tested simultaneously. Although different authors quantify different aspects of locomotion (small displacements, large displacements, rearing), the basic information yielded by activity meter tests is whether a test substance increases or decreases locomotion. Data obtained should be correlated with that obtained from direct observation (Irwin

**TABLE 3 Effects of Diazepam (Oral Dose) in the Primary Observation (Irwin) Test in the Rat**[a]

| 1 (mg/kg) | 2 (mg/kg) | 4 (mg/kg) | 8 (mg/kg) | 16 (mg/kg) | 64 (mg/kg) |
|---|---|---|---|---|---|
| **No change** | **Hypothermia** **+** at 30 min | **Sedation** + ↓ **Traction** **Hypothermia** (++ → 30 min) (+ at 60 and 120 min) | **Sedation +** ↓ **Traction** **Hypothermia** (++ → 60 min) (+ at 120 min) | **Sedation ++** ↓ **Traction** ↓ Reactivity to touch ↓ Muscle tone Abnormal gait **Hypothermia** ++ (+++ at 30 min) (+ at 120 min) | **Sedation** ++ ↓ **Traction** ↓ Reactivity to touch ↓ Muscle tone Abnormal gait **Hypothermia** (+++ → 120 min) (++ at 180 min) (+ at 24 h) |

[a]Note the dose-dependent increase in sedation and decreases in reactivity to touch, traction, and muscle tone accompanied by abnormal gait. $N = 4$ for each treatment group. + = slight; ++ = moderate; +++ = marked. Observations were performed at 15, 30, 60, 120, and 180 minutes after administration. Symptoms that did not necessitate handling were also observed up to 15 min immediately following administration.

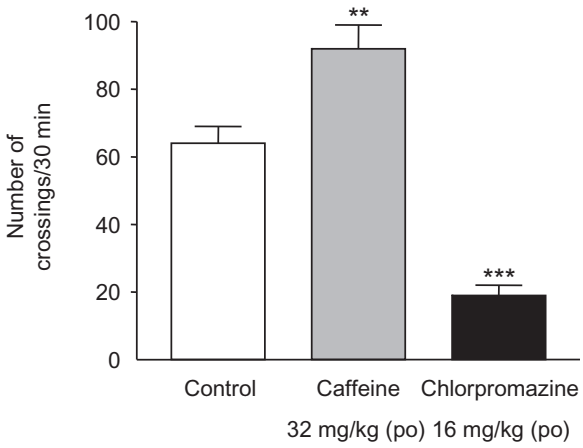


**FIGURE 9** Effects of caffeine and chlorpromazine (oral administration) on locomotion in the rat.

test) to ensure that apparent decreases are not due to motor incapacity or even to overexcitation.

The effects of a standard psychostimulant (caffeine) and a standard CNS depressant (chlorpromazine) in the activity meter test in the rat are shown in Fig. 9.

*Motor Coordination Tests* Locomotor coordination is most commonly assessed using a rotarod [217]. Rats or mice are placed onto a rod rotating either at a fixed speed or at a constantly increasing speed. The time taken for the animal to fall off the rod, or the number of animals remaining on the rod over a set duration, is mea-

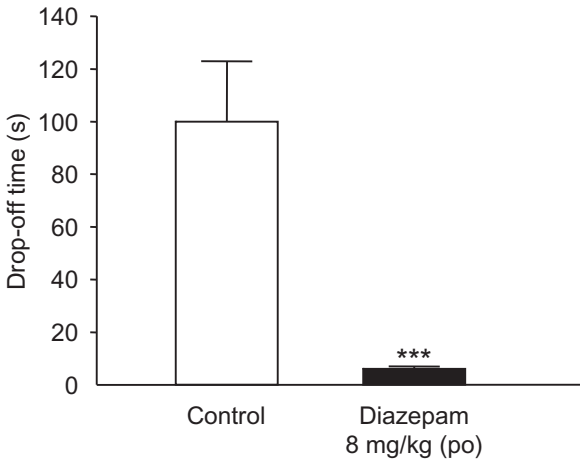


**FIGURE 10** Effects of diazepam (oral administration) on rotarod performance in the rat.

sured. To decrease test variability, the animals are usually given prior habituation to the rotarod before receiving the test substance. With most setups, several animals are tested simultaneously on the same rod, separated physically and visually by partitions with automated scoring of performance. When used in conjunction with locomotor activity tests, the rotarod test provides a useful quantification of the margin of safety between doses that alter spontaneous activity and those that disturb motor function. Unlike locomotor activity tests, which show bi-directional drug effects, rotarod tests are seldom capable of showing facilitation of motor performance.

The effects of a standard benzodiazepine (diazepam) in the rotarod test in the rat are shown in Fig. 10.

*Effects on the Convulsive Threshold* Test substances can induce frank convulsions when given at sufficiently high doses, but can also lower the convulsive threshold without inducing frank convulsions. Surprisingly, the ICH S7A core battery does not include tests for the convulsive threshold, although such tests were included in the earlier Japanese guidelines. We have criticized this absence [218] because we consider the existence of proconvulsant activity an important CNS risk. Furthermore, although anticonvulsant activity is not in itself a risk, it is known that several anticonvulsants (benzodiazepines, NMDA antagonists) cause memory disturbance. Thus, the presence of anticonvulsant activity could constitute a first marker for cognition impairment, with obvious implications for CNS safety. Both kinds of activity can be seen with chemically induced convulsions such as with pentylenetetrazole (PTZ), where a shortened or lengthened latency or a higher or lower frequency of convulsions/deaths can be observed with pro- and anticonvulsants, respectively [219]. A variant of the electroconvulsive shock (ECS) method, whereby successive animals are subjected to increasing or decreasing intensities of ECS depending on the occurrence of convulsions with the preceding shock level, permits a sensitive measure of the electroconvulsive threshold. It is important for safety pharmacology purposes that convulsive threshold tests can vary in both directions.

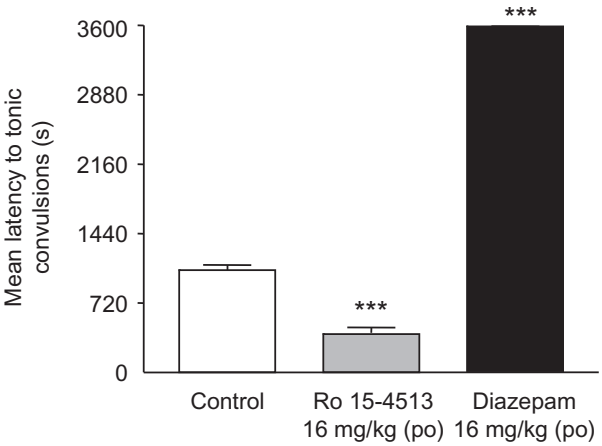


**FIGURE 11** Effects of Ro 15-4513 and diazepam (oral administration) on the latency to tonic convulsions induced by PTZ (120 mg/kg subcutaneous) in the rat.

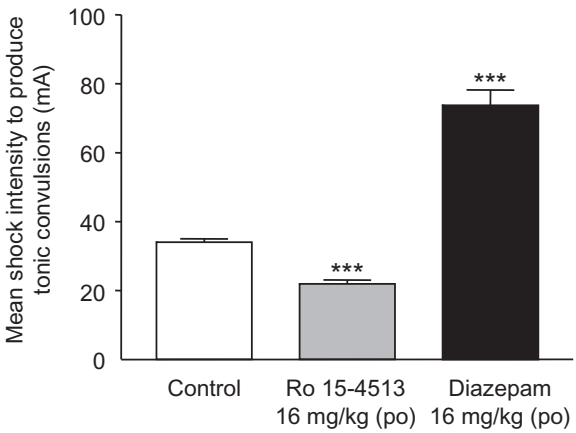


**FIGURE 12** Effects of Ro 15-4513 and diazepam (oral administration) on the mean shock intensity required to induce tonic convulsions in the rat.

The effects of a standard proconvulsant benzodiazepine (Ro13–4513) and an anticonvulsant benzodiazepine (diazepam) in the PTZ test in the rat are shown in Fig. 11 and in the ECS threshold test in the rat in Fig. 12.

*Interaction with Hypnotics* Another risk factor that is not life threatening in itself but can have life threatening consequences is the induction of drowsiness or sleep. Sleep induction in rodents can be evaluated by loss of the righting reflex. Intrinsic sleep-inducing activity can be observed during the Irwin test or FOB (see earlier discussion). A more sensitive index of sleep enhancement can, however, be obtained by observing the interaction of the test substance with sleep induced by standard hypnotics such as barbiturates [220]. Such procedures can also unmask psychostimulant activity not always evident from direct observation (decrease in barbiturate-induced sleep). Barbiturate interaction studies therefore provide a useful complement to studies of general behavior and spontaneous locomotion.

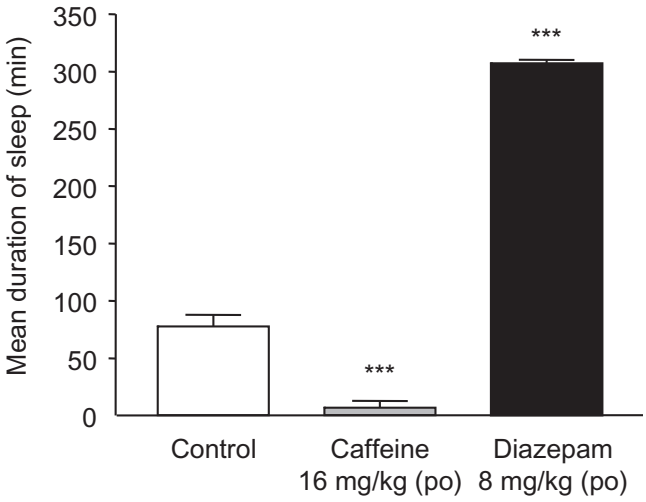


**FIGURE 13** Effects of caffeine and diazepam (oral administration) on the duration of sleep induced by barbital (150 mg/kg intraperitoneal) in the rat.

The effects of a standard psychostimulant (caffeine) and a standard CNS depressant (diazepam) in the barbital interaction test in the rat are shown in Fig. 13.

### 6.3 Follow-up CNS Safety Pharmacology Procedures

Follow-up CNS safety pharmacology involves more complex and time-consuming experimental procedures and includes methods for evaluating drug effects on cognitive function, the electroencephalogram (EEG), and the potential for the new substance to cause drug abuse or dependence.

***Cognitive Function*** Included under the term cognitive function are learning, memory, attention, and general intellectual activity. New drugs should be devoid of impairing effects on these functions, whatever their indication. Indeed, package inserts for many kinds of drugs contain warnings about the use of the drug while driving, working, or engaging in various daily occupations, referring mainly to an attentional deficit. Drug-induced impairment of learning and memory can also be troublesome side effects of certain agents, even if not life threatening. It is therefore important that CNS safety pharmacology provide procedures for evaluating these effects in animals. The following section provides an indication of the kinds of experiment that could be undertaken. The list is by no means exhaustive.

*Passive Avoidance* One of the simplest procedures for looking for adverse effects on learning/memory is the one-trial passive avoidance task [221]. There are several variants of this procedure, but the basic principle is that a rat or a mouse receives an aversive stimulation in a recognizable environment and on a later occasion shows it has remembered by not going there (passive avoidance). For example, a rat placed into the light compartment of a two-compartment box will explore the apparatus and eventually enter the dark compartment, where it receives a brief electric shock [222]. When placed again into the light compartment 24 or 48 hours later, the animal will avoid going into the dark compartment despite a natural preference for the dark. Amnesia-inducing drugs (benzodiazepines, anticholinergics, NMDA

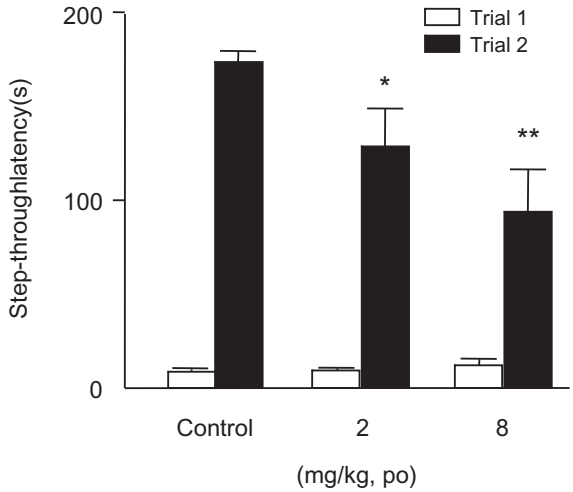


**FIGURE 14** Effects of diazepam (oral administration) on passive avoidance behavior in the rat. Note the absence of effects of diazepam on the step-through latency at Trial 1 and the presence of a dose-dependent decrease in step-through latency at Trial 2 in the absence of diazepam.

antagonists), administered before the first exposure, will attenuate the animal's memory for the shock as shown by a decreased latency to enter the dark compartment on the test day.

The effects of a standard benzodiazepine (diazepam) in the passive avoidance test in the rat are shown in Fig. 14.

Passive avoidance is a fairly simple and rapid procedure because it involves learning obtained in a single trial. It is therefore suitable as a first screen for potential cognition impairing activity. On the other hand, it is not clearly interpretable in terms of the cognitive function involved as it is not possible to distinguish a drug's effects on learning or memory, or whether the drug influenced performance by impairing attention or even pain sensitivity during the learning trial.

*Morris Maze* Another procedure, which allows a greater degree of interpretation, is the Morris water maze [223]. In this test a rat or a mouse is placed into a circular tank containing opacified water and has to find an escape platform just beneath the surface and therefore invisible to the animal. After swimming around for a certain time (usually limited to 2 minutes), the animal will eventually come across the hidden platform and climb onto it to escape from the water. When placed again in the water on subsequent occasions, the animal will find the platform with increasing rapidity, indicating that it has learned the position of the platform.

Morris maze learning involves spatial navigation and the capacity of the animal to attend to extra-maze cues. Furthermore, the procedure can be modulated to examine the effects of drugs on both short-term memory (within-day learning) and longer term memory (day-to-day learning).

The effects of different neuroleptics (tiapride, haloperidol, and risperidone) on short-term memory and on longer term memory in the Morris maze test in the rat are shown in Figs. 15 and 16.

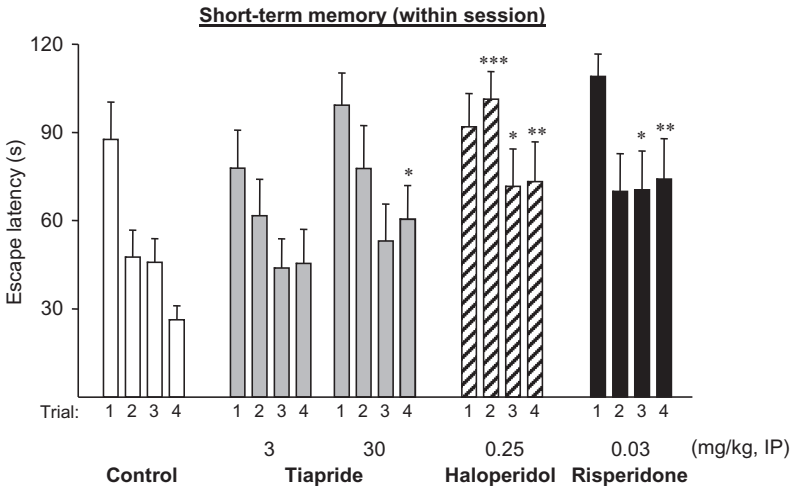


**FIGURE 15** Effects of three neuroleptics (intraperitoneal administration), tiapride, haloperidol, and risperidone, on escape latencies during the first session of the Morris maze task in the rat. Note the clear decrease in escape latencies in the vehicle control group (short-term memory) and the absence of effect of the three substances on performance at the first trial (absence of intrinsic effects on swimming performance). Tiapride slightly but dose-dependently attenuates the decrease in escape latency from trial to trial, whereas haloperidol and risperidone clearly attenuate this decrease (perturbing effect on short-term memory).

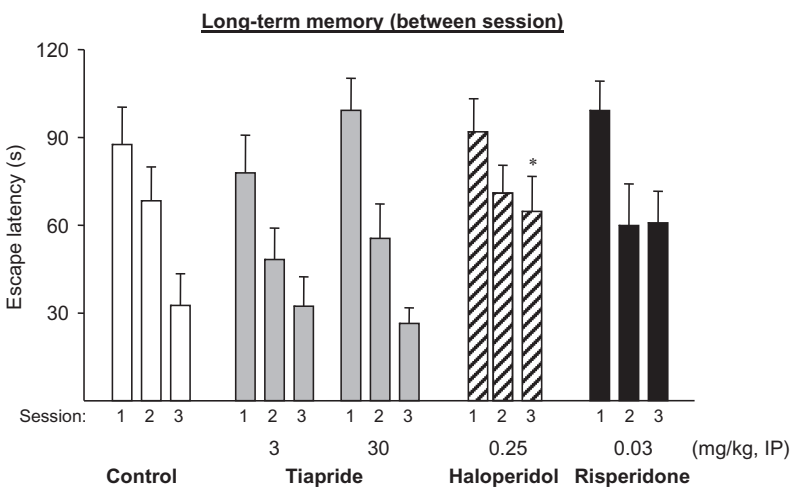


**FIGURE 16** Effects of three neuroleptics (intraperitoneal administration), tiapride, haloperidol, and risperidone, on escape latencies during the same Morris maze task on three consecutive days (first trial of each day). Note the clear decrease in escape latencies in the vehicle control group from day to day (long-term memory). Tiapride has no effect on the decrease in the first trial escape latency from day to day (absence of effect on long-term memory), whereas haloperidol clearly attenuates this decrease with a similar tendency for risperidone (perturbing effect on long-term memory).

*Radial Maze* Another task in which short-term memory can be differentiated from long-term memory is the radial maze [224]. This apparatus consists of a central platform with arms radiating from it like the spokes of a wheel. A hungry rat or a mouse is placed in the center of the maze and can find food at the end of each arm. Before drug testing, animals are usually given initial maze training to learn that food is available at the end of each arm. Once trained in the maze, an ideal performance during a trial would be for the animal to visit each arm once to collect food. This task therefore requires that, during a trial, the animal remembers the arms already visited (short-term or "working" memory). Longer term memory can also be investigated by arranging that only some of the arms are baited, but always the same ones, during repeated sessions. In this case the animal, in addition to always choosing a new arm ("working" memory), must visit only the arms previously associated with food (long-term or "reference" memory). Amnesia-inducing drugs, such as anticholinergics or benzodiazepines, clearly impair radial maze performance, in particular, "working" memory.

With the radial maze, other aspects of performance can also be measured. These include running times and response choice strategy. The strategy parameter assesses the degree to which the animal uses behavioral routines for selecting arms. For example, when animals are highly trained in the maze, they will tend to select subsequent arms always in a particular direction (left or right) and thereby minimize errors without using working memory at all. Assessing the effects of test substances on the different parameters can therefore permit a more accurate interpretation of the behavioral changes observed.

The effects of MK-801 (dizocilpine) in the radial maze test in the rat are shown in Fig. 17.

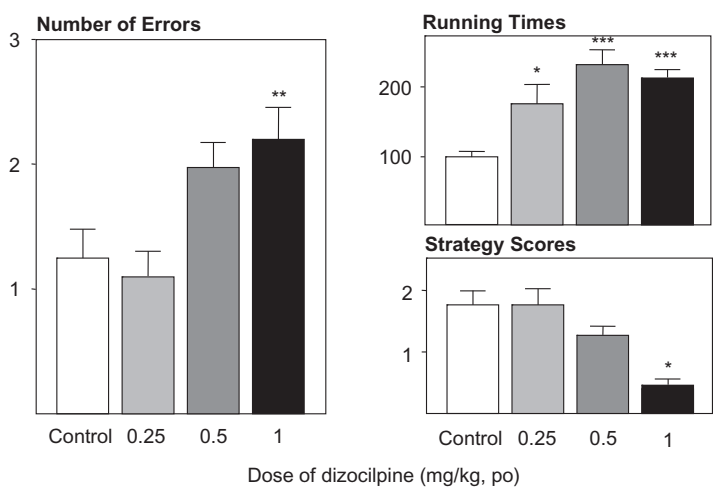


**FIGURE 17** The effects of MK-801 (dizocilpine) (oral administration) on the three parameters measured (errors, running times, response strategy) during the radial maze task in the rat (mean of three sessions). Note the dose-dependent increase in the number of errors with no effect on running times. A decrease in response strategy occurs only at the highest dose. Taken together, these data suggest a specific impairment of "working" memory by MK-801.

The radial maze requires considerable training and is therefore more time consuming than either passive avoidance or the Morris maze. On the other hand, by means of the different parameters measured, it offers data that are more clearly interpretable. Furthermore, the radial maze is based on a different motivational system (food reward). Radial maze data can thus usefully complement those obtained in the other procedures.

*Operant Behavior Techniques: Delayed Responding* In delayed responding procedures, the animal is required to retain information over a short period (usually seconds) and then to show, by pressing an appropriate lever, whether it has correctly remembered it [225]. In a typical example, a rat in a Skinner box is presented with a lever, either on the left or the right side of the food dispenser. The rat presses the lever and the lever is withdrawn. Five seconds later, two levers are presented and the rat has to press the lever opposite to that which was presented previously to obtain a food reward (delayed alternation, delayed nonmatching to sample). If the rat presses the same lever as that previously presented, the lever will be withdrawn but no food will be given. In this fashion, the animal can be trained to retain a piece of spatial information (position of a lever), and thereby demonstrate its short-term memory capacity over time. Using the same apparatus, it is possible to measure the animal's reaction times, both to the single lever presentations (simple reaction time) and to the two-lever presentations (choice reaction time), and thereby provide additional information about other aspects of cognitive performance more related to attention and information processing speed.

The effects of a standard benzodiazepine (diazepam) and a standard anticholinergic (scopolamine) in the delayed alternation test in the rat are shown in Fig. 18.

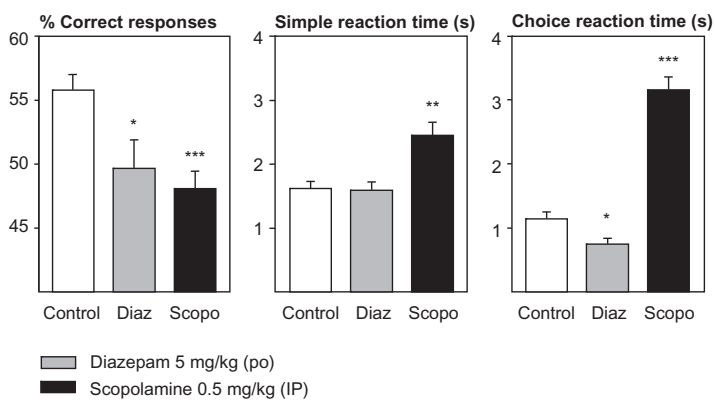


**FIGURE 18** Effects of diazepam (oral administration) and scopolamine (intraperitoneal administration) on the 3 parameters measured during the acquisition of a delayed alternation task in the rat (mean performance over 10 sessions). Note that both diazepam and scopolamine significantly decrease the number of correct responses (impairment of learning/memory). Diazepam has no effect on simple reaction times, whereas scopolamine significantly increases it (impairment of attention). Diazepam significantly decreases choice reaction times, whereas choice reaction times are clearly increased by scopolamine. The increased choice reaction times with scopolamine probably reflect an impairment of information processing speed, whereas the decrease observed with diazepam probably reflects the disinhibitory effects of diazepam.

Although most of the published work with delayed responding concerns studies performed in the rat [226], similar tasks can be performed in monkeys [227] and in humans [228].

***EEG Studies*** Electroencephalogram (EEG) investigations are recommended as possible follow-up studies in both ICH S7A and the previous Japanese guidelines. An EEG provides a useful means of evaluating the effects of a test substance on brain activity in freely moving conscious animals. As far as CNS follow-up safety pharmacology is concerned, three major EEG applications (EEG trace monitoring, quantified EEG (QEEG), and the sleep/wake cycle) appear to be of particular relevance.

*EEG Trace Monitoring* If initial experiments on the convulsive threshold suggest potential proconvulsant activity, follow-up studies using visual monitoring of the EEG traces can be employed to determine whether paroxysmal activity (spikes) can be detected [229]. Certain substances induce clear paroxysmal activity on the EEG at doses that have no observable effects or before these effects become clearly apparent in behavior. It is thus possible to confirm the existence of proconvulsant activity from the EEG traces and even to identify the locus of such activity from differential placement of the EEG electrodes within the brain, for example, cortex or hippocampus.

Trace monitoring studies, for economic reasons, are performed mainly in the rat but can also be conducted in the dog using telemetric devices. Dog studies are particularly useful if the primary therapeutic activity of the test substance, for example, effects on the cardiovascular system, have been established in the same species.

The effects of a standard convulsant (PTZ) on the cortical EEG trace in the dog are shown in Fig. 19.

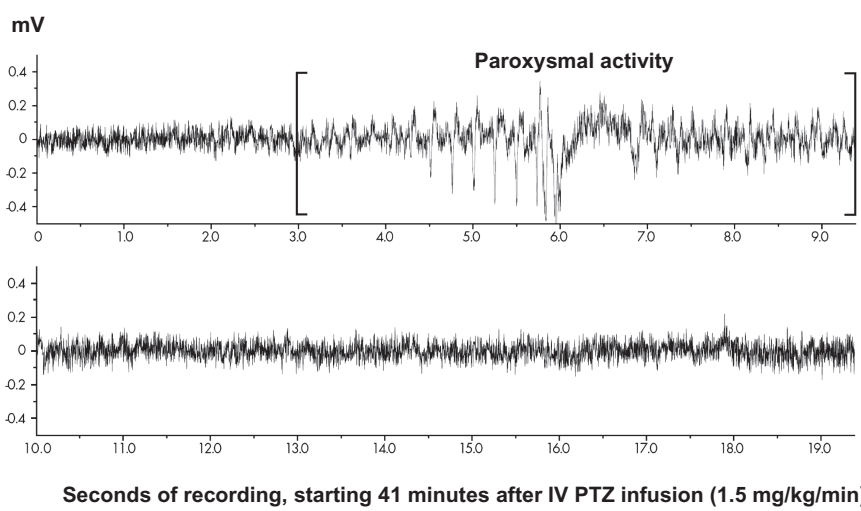


**FIGURE 19** Effects of an intravenous infusion of PTZ (1.25 mg/kg/min) on the EEG trace in the conscious freely moving beagle dog measured via telemetry. Note the sequence of paroxysmal activity (marked by the square brackets). This occurred in the absence of overt convulsive phenomena.

*QEEG* QEEG makes use of the Fourier theorem to analyze the amount of electrical power in the different frequency bands of the EEG [230]. Animals, with electrodes fixed to the skull (cortical leads) and sometimes also implanted stereotaxically in selected structures (e.g., hippocampus, striatum), are exposed to brief measurement periods (1–2 hours) during which they are either left free to move spontaneously or activated by means of a treadmill to ensure a stable heightened level of vigilance. The EEG is recorded and is subjected, via appropriate software, to a fast Fourier transformation.

While there is debate as to whether QEEG, by virtue of the different profiles observed, is capable of identifying specific classes of psychotropic agents, there is general consensus that the QEEG can detect basic stimulant, sedative, or even convulsant activity [231]. This could have clear relevance for drug safety by corroborating, in terms of brain activity, data obtained from behavioral observation. A more clearly safety pharmacology application is to determine the highest dose that can be administered without affecting the QEEG.

The effects of a standard psychostimulant (amphetamine) on the QEEG power spectrum in the rat are shown in Fig. 20.

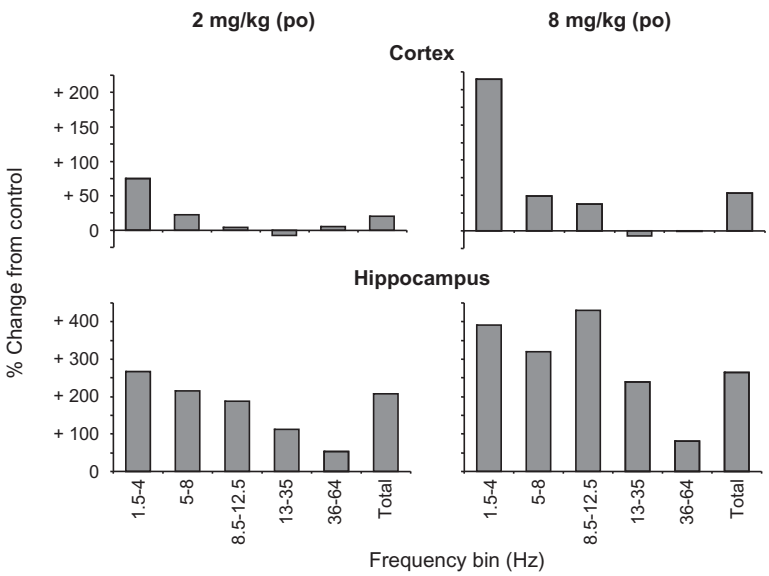


**FIGURE 20** Effects of d-amphetamine (oral administration) on the power spectrum of the QEEG. Note that amphetamine dose-dependently increases power over the whole frequency range, more markedly so at low frequencies and particularly in the hippocampus. Similar profiles are observed with other psychostimulants such as methylphenidate or nomifensine (data not shown).

*Sleep/Wake Cycle* Studies of the sleep/wake cycle are to be distinguished from QEEG by the fact that spontaneous as opposed to induced changes are studied over longer periods of time [232]. Such studies determine whether the test substance induces alterations in the architecture of natural sleep [233]. Although sleep patterns differ between rodents and humans, rodents being mostly awake during the

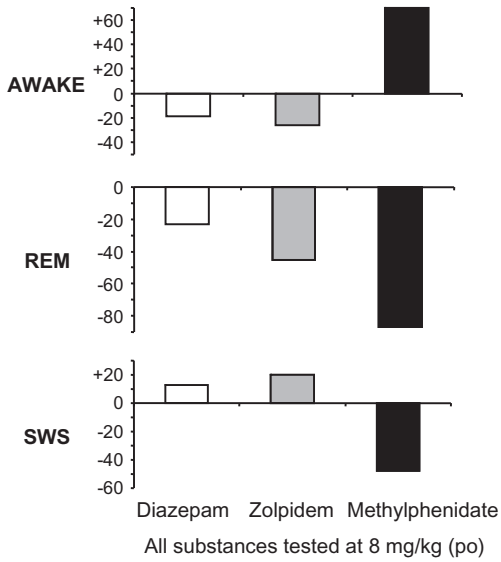


**FIGURE 21** Effects of diazepam, zolpidem, and methylphenidate (oral administration) on the sleep/wake cycle in the rat during the first 4 hours after drug administration. Note that both diazepam and zolpidem decrease the amount of wakefulness and increase the amount of slow wave sleep, whereas opposite effects are observed with methylphenidate. All substances decrease the amount of REM sleep, with the most marked effects being observed on this parameter with methylphenidate.

dark phase of the light/dark cycle, drug effects appear to be highly comparable between the species. From a safety pharmacology standpoint, as for QEEG, the aim of this type of study is to determine up to which dose the test substance is devoid of effects on sleep function.

The effects of two standard sleep-enhancing agents (diazepam and zolpidem) and a psychostimulant (methylphenidate) on the sleep/wake cycle in the rat are shown in Fig. 21.

***Drug dependence and abuse*** Another subject, clearly within the notion of safety, is the assessment of a novel substance's potential to induce problems of dependence or abuse. These are mentioned briefly in ICH S7A ("other organ systems"), but more complete guidelines have been proposed recently in Europe [234] and North America [235].

Although dependence and abuse frequently occur together, they are not synonymous [236]. Drug dependence refers to the inability of the dependent individual to function normally without the drug. Drug abuse, as its name implies, refers to misuse or overuse of a drug and is generally covered under the notion of drug seeking behavior. Dependence can be both physical and psychological and is usually manifested on drug discontinuation. Psychological dependence is indicated by craving for the drug, whereas physical dependence is demonstrated more objectively by signs ranging from changes in body temperature to life-threatening conditions such as status epilepticus or delirium tremens. Withdrawal symptoms are readily demonstrated in animal models, whereas drug craving is more problematic. The

difference between dependence and abuse can be illustrated pharmacologically. For example, marihuana and LSD are frequently abused, but it is difficult to show any clear signs of withdrawal on drug discontinuation. In contrast, selective serotonin reuptake inhibitors (SSRIs) do induce discontinuation signs, but have not been subject to abuse. On the other hand, drugs such as heroin and cocaine cause both abuse and marked signs of physical and psychological dependence.

*Drug Dependence* Tests for drug dependence consist mainly of repeated treatment studies, followed by drug discontinuation, where withdrawal symptoms occur either spontaneously (nonprecipitated withdrawal) or after administration of specific antagonists, for example, naloxone (opiates) or flumazenil (benzodiazepines).

In the absence of a known mechanism, the most convincing evidence of drug dependence liability comes from studies where the test substance is repeatedly administered at high doses and is then simply discontinued. Under these conditions, clear withdrawal signs (weight loss, decreased food consumption, hyperthermia) can be observed after a variety of substances, for example, benzodiazepines, morphine, and cocaine [237]. If it can be shown that a test substance can be repeatedly administered at maximally tolerated doses and then be discontinued without observable symptoms under conditions where similar treatment with an appropriately chosen reference substance induces clear withdrawal signs, then it seems reasonable to conclude that the test substance is unlikely to cause physical dependence.

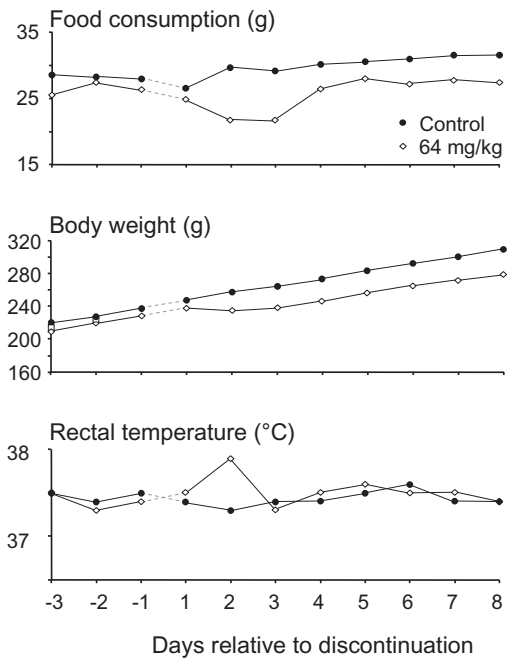


**FIGURE 22** Effects of drug discontinuation after 10 days of twice daily treatment with chlordiazepoxide (oral administration) on food consumption, body weight gain, and rectal temperature in the rat. Note a decrease in food consumption and an increase in rectal temperature during the first 2 days after drug discontinuation with recovery toward vehicle control values thereafter. Similar but less marked effects are observed on body weight.

Withdrawal effects in the rat after discontinuing repeated treatment with a standard benzodiazepine (chlordiazepoxide) are shown in Fig. 22.

*Drug Abuse* Drug abuse represents a more serious social problem than drug dependence because it involves a greater variety of substances. Tests for abuse are generally of two kinds. The first kind assesses whether a new substance resembles a known drug of abuse. The second kind assesses whether the test substance possesses positive reinforcing properties and can thereby sustain drug seeking behavior.

### SIMILARITY TO KNOWN DRUGS OF ABUSE

*Direct Observation Tests* A first approach to assessing similarity is simply to compare the behavioral effects of the test substance with those of a known drug of abuse using an Irwin-style observation procedure.

The effects of two NMDA antagonists (PCP and MK-801) in the mouse Irwin test are shown in Fig. 23.

**Irwin Test in the Mouse**

**MK 801 (p.o.)**

| 0.125 (mg/kg) | 1 (mg/kg) | 4 (mg/kg) | 32 (mg/kg) |
|---|---|---|---|
| No change | Excitation +++ | Excitation +++ | Sedation ++ |
| | Straub | Straub | Straub |
| | Stereotypies | Stereotypies | Stereotypies |
| | Convulsions | Convulsions | Convulsions |
| | Hyperthermia + | Hyperthermia + | Tremor |
| | | | ↓ Reactivity to touch |
| | | | Motor incoordination |
| | | | ↓ Muscle tone |

**Phencyclidine (p.o.)**

| 1 (mg/kg) | 4 (mg/kg) | 16 (mg/kg) | 64 (mg/kg) |
|---|---|---|---|
| ↓ Reactivity | Excitation ++ | Excitation +++ | Convulsions |
| | ↓ Traction | Stereotypies | Death (3/3) |
| | ↓ Reactivity to touch | Convulsions | |
| | | Tremor | |
| | | ↓ Fear | |
| | | ↓ Traction | |
| | | Motor incoordination | |
| | | ↓ Muscle tone | |
| | | ↑ Respiration | |

**FIGURE 23** Effects of PCP and MK-801 (dizocilpine) in the Irwin test in the mouse. Note the large overlap in behavioral signs (excitation, stereotypies, motor incoordination, convulsions) between the two substances.

*Drug Discrimination Tests* A more sensitive approach to assessing drug similarity is the use of drug discrimination procedures where animals, by pressing on one of two levers in a Skinner box, can show that they can discriminate the presence of a particular drug of abuse (training drug) from vehicle [238]. They are then given the new substance and can indicate by their choice of lever, whether the new substance resembles the training drug. A test substance that is perceived as being subjectively similar to a drug of abuse is also likely to be abused. An advantage of drug discrimination paradigms is that similar methodology can be used in monkeys [239] and even humans [240].

An interesting variant of the drug discrimination procedure is to use the test substance as the training drug. If the test substance can be shown to serve as a discriminative stimulus, it can be supposed that the test substance induces identifiable subjective effects. If the substance cannot be discriminated, it is less likely to be abused because of its subjective effects. On the other hand, the existence of a drug-induced discriminative stimulus with a test substance is no proof that substance will be abused. For example, naloxone can clearly be discriminated but is never abused. If clear discrimination can be established with the test substance, using the test substance as the training drug becomes particularly useful because the same trained animals can subsequently be tested for generalization to a range of drugs of abuse. This represents a more economical approach than using independent groups of animals, each trained to recognize a particular class of drug. It must be remembered, however, that drug discrimination procedures represent only an indirect approach toward predicting whether a test substance will be abused. They cannot predict the abuse potential of a drug inducing totally novel sensations.

The effects of three standard psychostimulants (amphetamine, cocaine, and methylphenidate) in rats trained to discriminate amphetamine are shown in Fig. 24.

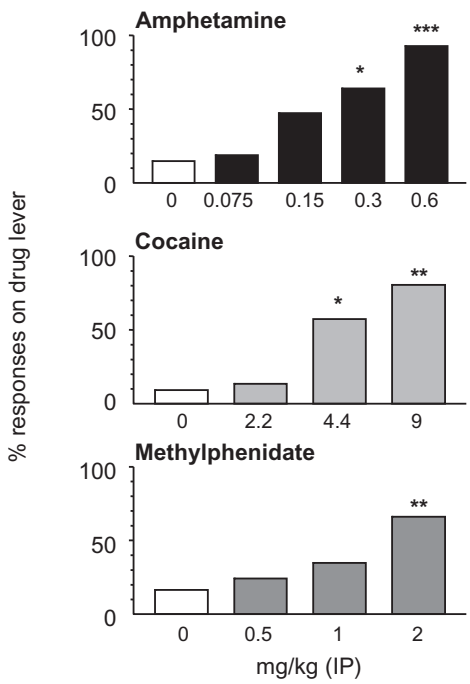


**FIGURE 24** Effects of d-amphetamine, cocaine, and methylphenidate (intraperitoneal administration) on drug discrimination in rats trained to discriminate 0.6 mg/kg (IP) d-amphetamine from saline. Note the dose-dependent generalization of lower doses of amphetamine toward responding on the lever associated with the training dose of amphetamine, with similar generalization curves for cocaine and methylphenidate.

TESTS FOR POSITIVE REINFORCING PROPERTIES

*Place Preference Tests* A relatively simple paradigm is conditioned place preference, where it can be shown that an animal, which has been repeatedly exposed to an identifiable environment in the presence of a drug of abuse, will show preference for that environment when later given a choice [241].

The effects of two standard drugs of abuse (cocaine and morphine) in the conditioned place preference test in the rat are shown in Fig. 25.

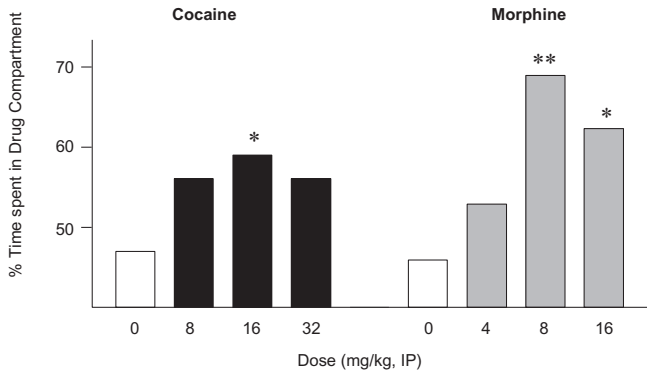


**FIGURE 25** Effects of cocaine and morphine (intraperitoneal administration) in the conditioned place preference test in the rat.

*Self-Administration Procedures* The most direct tests of abuse potential are self-administration procedures, where the animal, by pressing on a lever, causes itself to receive an IV infusion of the test substance via a pump system connected to an indwelling catheter [242]. Parallel methods exist in both rats and monkeys and provide substantially similar findings. Virtually all drugs that are abused in humans will induce or maintain self-administration behavior in animals.

Generally, two basic procedures are used. The most common is the substitution procedure, where an animal is first trained to self-administer a known drug of abuse such as cocaine or heroin. The animal is then tested to see whether a test substance will substitute for the training drug, that is, whether the animal will continue to self-administer the test substance. Another procedure is to assess whether a test substance can initiate self-administration behavior in drug-naive animals. The substitution procedure is probably the more sensitive for detecting self-administration potential, because the test animal can already be considered a "drug abuser." A test substance that does not maintain self-administration in this procedure is therefore unlikely to do so in any other variant and thus can be considered devoid of abuse potential. Another advantage of the substitution procedure is that it permits a quantification of the self-administration level in individual animals before assessing the test substance. This is important because there exist marked individual differences between animals, particularly monkeys, in their inclination to self-administer drugs, as there are in humans. A final advantage of the substitution procedure is that it permits the animal, particularly the monkey, to be used repeatedly for drug evaluation programs, which represents an important economic consideration for these time-consuming and costly procedures.

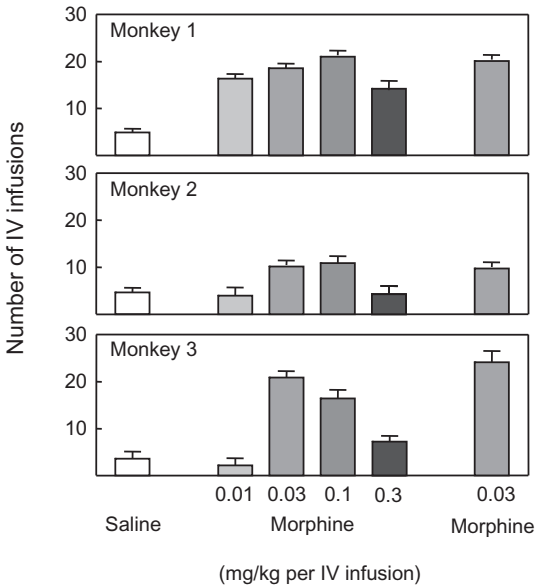


**FIGURE 26** Effects of morphine and a standard SSRI (fluoxetine) in three rhesus monkeys habituated to self-administer 0.1 mg/kg/infusion morphine (substitution procedure). Fluoxetine does not maintain self-administration in contrast to heroin.

The effects of a standard drug of abuse (morphine) and a standard SSRI (fluoxetine) in a self-administration procedure in the rhesus monkey are shown in Fig. 26.

### 6.4 General Conclusions

This section has provided an overview of the kinds of studies that can be carried out to evaluate the risk of adverse effects of a novel substance on CNS function. The basic routines useful for core battery CNS safety assessment, mainly in the rodent, were covered. Additional and more complex procedures were also included to provide a more comprehensive review relevant to CNS safety prior to drug registration. For this reason, an important part of the present section was devoted to methods for evaluating higher CNS function (cognitive behavior and electrical brain activity), because adverse effects at this level could radically affect the safety of many substances in everyday use. Methods for evaluating drug abuse and dependence were also included, despite the fact that these problems generally concern only CNS-active drugs.

A final point is that no guidelines can provide a complete recipe for the studies that should be undertaken to establish a drug's CNS safety. The considerable latitude of ICH S7A is a reflection of changing official attitudes toward defining the precise methods to be employed. It is now up to the individual drug developers to decide and justify which specific procedures are included. The present section has attempted to illustrate some of the possibilities beyond routine core battery studies, to ensure that a novel substance is devoid of risk on higher CNS function.

## ACKNOWLEDGMENTS

We thank Dr. Paul Moser, Dr. Vincent Castagné, Dr. Niklaus Dürmüller, and Ms. Julie Presher for their comments on the contents of this chapter.

## REFERENCES

1. Zbinden G. *Pharmacological Methods in Toxicology*. Elmsford, NY: Pergamon Press; 1979.
2. Bass AS, Kinter L, Williams P. Origins, practices and future of safety pharmacology. *J Pharmacol Toxicol Methods* 2004;49:145–151.
3. Anon. *Guidelines for General Pharmacology Studies*. Tokyo, Japan: Ministry of Health and Welfare; 1991.
4. Anon. *General Pharmacology Japanese Guidelines for Nonclinical Studies of Drugs Manual*, Tokyo, Japan: Yakuji Nippo Limited; 1995.
5. Shah RR. Drug-induced QT interval prolongation: regulatory perspectives and drug development. *Ann Med* 2004;36 (Suppl 1):47–52.
6. Gardin JM, Schumacher D, Constantine G, Davis KD, Leung C, Reid CL. Valvular abnormalities and cardiovascular status following exposure to dexfenfluramine or phentermine/fenfluramine. *Multicenter Study JAMA* 2000;283:1703–1709.
7. Anon. *ICH Guidance for Industry—S7A Safety Pharmacology Studies for Human Pharmaceuticals*. Washington, DC: US Department of Health and Human Services, Food and Drug Administration; 2001.
8. Bass AS, Tomaselli G, Bullingham R III, Kinter LB. Drug effects on ventricular repolarization: a critical evaluation of the strengths and weaknesses of current methodologies and regulatory practices. *J Pharmacol Toxicol Methods* 2005;52:12–21.
9. Kinter LB, Siegl PK, Bass AS. New preclinical guidelines on drug effects on ventricular repolarization: safety pharmacology comes of age. *J Pharmacol Toxicol Methods* 2004;49:153–158.
10. Baillie M, Bass AB, Darpo B, Engwall M, Garnes D, Gintant G, Hammond T, Hanson LA, Mittelstadt S, Rampe D, Sarazan RD, Siegl P, Thomas K. International Life Sciences Institute/Health & Environmental Sciences Institute Cardiovascular Risk Assessment, Cardiovascular Risk Assessment Workshop Proceedings; 2007.
11. Anon. *ICH Guidance for Industry—S7B Nonclinical Evaluation of the Potential for Delayed Ventricular Repolarization (QT Interval Prolongation) by Human Pharmaceuticals*. Washington, DC: US Department of Health and Human Services, Food and Drug Administration; 2005.
12. Harrison AP, Erlwanger KH, Elbrond VS, Andersen NK, Unmack MA. Gastrointestinal-tract models and techniques for use in safety pharmacology. *J Pharmacol Toxicol Methods* 2004;49:187–99.
13. Kinter LB. Safety pharmacology of the renal and gastro-intestinal systems. In: Williams PD, Bass AS, Eds. *Safety Pharmacology: A Practical Guide*. Gaithersburg, MD: TherImmune Research Corporation; 2003.
14. Gray LE. Jr, Kelce WR, Wiese T, Tyl R, Gaido K, Cook J, Klinefelter G, Desaulniers D, Wilson E, Zacharewski T, Waller C, Foster P, Laskey J, Reel J, Giesy J, Laws S, McLachlan J, Breslin W, Cooper R, Di Giulio R, Johnson R, Purdy R, Mihaich E, Safe S, Sonnenschein C, Welshons W, Miller R, McMaster S, Colborn T. Endocrine Screening Methods Workshop report: detection of estrogenic and androgenic hormonal and

antihormonal activity for chemicals that act via receptor or steroidogenic enzyme mechanisms. *Reprod Toxicol* 1997;11:719–750.

15. Emeigh Hart SG. Assessment of renal injury *in vivo*. *J Pharmacol Toxicol Methods* 2005;52:30–45.
16. Descotes J. Developments in biological standardization. *Immunotoxicol Immunomodulators* 1992;77:99–102.
17. Whitebread S, Hamon J, Bojanic D, Urban L. *In vitro* safety pharmacology profiling: an essential tool for successful drug development. *Drug Discov Today* 2005;10:1421–1433.
18. Friedrichs GS, Patmore L, Bass AS. Non-clinical evaluation of ventricular repolarization (ICH S7B): results of an interim survey of international pharmaceutical companies. *J Pharmacol Toxicol Methods* 2005;52:6–11.
19. Kinter LB, Valentin JP. Safety pharmacology and risk assessment. *Fundam Clin Pharmacol* 2002;16:175–182.
20. Matsuzawa T, Hashimoto M, Nara H, Yoshida M, Tamura S, Igarashi T. Current status of conducting function tests in repeated dose toxicity studies in Japan. *J Toxicol Sci* 1997;22:375–382.
21. Gad SC. *Safety Pharmacology in Pharmaceutical Development and Approval*. Boca Raton, FL: CRC Press; 2004.
22. Bass AS, Darpo B, Breidenbach A, Bruse K, Feldman HS, Garnes D, Hammond T, Haverkamp W, January C, Koerner J, Lawrence C, Leishman D, Roden D, Valentin JP, Vos MA, Zhou YY, Thomas K, Sager P. Recommendations for areas of investigation from the Workshop Breakout Sessions. Workshop: Moving Towards Better Predictors of Drug-Induced Torsades de Pointes (TdP); 2005.
23. Anon. *ICH Guidance for Industry—S6 Safety Evaluation of Biotechnology-Derived Pharmaceuticals*. Washington, DC: US Department of Health and Human Services, Food and Drug Administration; 1997.
24. Anon. *ICH Guidance for Industry—M3 NonClin Safety Studies for the Conduct of Human Clinical Trials for Pharmaceuticals*. Washington, DC: US Department of Health and Human Services, Food and Drug Administration; 1997.
25. Anon. Title 21 Code of Federal Regulations (21 CFR Part 58). *Good Laboratory Practice for NonClinical Laboratory Studies*. Washington, DC: US Department of Health and Human Service, Food and Drug Administration; revised 1 April 2006.
26. Anon. *ICH Guidance for Industry—E14 Clinical Evaluation of QT/QTc Interval Prolongation and Proarrhythmic Potential for Non-Antiarrhythmic Drugs*. Washington, DC: US Department of Health and Human Services, Food and Drug Administration; 2005.
27. Anon. Title 21 Code of Federal Regulations (21 CFR Part 11). *Electronic Records; Electronic Signatures*. Washington, DC: US Department of Health and Human Service, Food and Drug Administration, 2000.
28. Wakefield ID, Pollard C, Redfern WS, Hammond TG, Valentin JP. The application of *in vitro* methods to safety pharmacology. *Rev Fundam Clin Pharmacol* 2002;16:209–218.
29. Suter W. Predictive value of *in vitro* safety studies. *Curr Opin in Chem Biol* 2006;10:362–366.
30. Ollerstam A, Visser SA, Persson AH, Eklund G, Nilsson LB, Forsberg T, Wiklund SJ, Gabrielsson J, Duker G, Al-Saffar A. Pharmacokinetic-pharmacodynamic modeling of drug-induced effect on the QT interval in conscious telemetered dogs. *J Pharmacol Toxicol Methods* 2006; 53:174–183.
31. Sheiner LB, Steimer JL. PharmacokInetic/pharmacodynamic ModelIng In drug development. *Annu Rev Pharmacol Toxicol* 2000;40:67–95.

32. Hanson LA, Bass AS, Gintant G, Mittelstadt S, Rampe D, Thomas K. ILSI-HESI cardiovascular safety subcommittee initiative: evaluation of three non-clin models of QT prolongation. *J Pharmacol Toxicol Methods*. 2006;54:116–129.
33. Klumpp A, Trautmann T, Markert M, Guth B. Optimizing the experimental environment for dog telemetry studies. *J Pharmacol Toxicol Methods* 2006;54:141–149.
34. Chaves AA, Keller WJ, O'Sullivan S, Williams MA, Fitzgerald LE, McPherson HE, Goykhman D, Ward PD, Hoe CM, Mixson L, Briscoe RJ. Cardiovascular monkey telemetry: sensitivity to detect QT interval prolongation. *J Pharmacol Toxicol Methods* 2006;54:150–158.
35. Swissa M, Zhou S, Paz O, Fishbein MC, Chen LS, Chen PS. Canine model of paroxysmal atrial fibrillation and paroxysmal atrial tachycardia. *Am J Physiol Heart Circ Physiol* 2005;289:H1851–H1857.
36. Acosta D Jr, Ed. *Cardiovascular Toxicology*. London: Taylor & Francis; 2001.
37. *Goodman and Gilman's The Pharmacological Basis of Therapeutics*. New York: McGraw-Hill; 2006.
38. Berne RM, Levy MN. *Cardiovascular Physiology*, 8th ed. St Louis, MO: Mosby; 2001.
39. De Mello WC, Janse MJ. Heart cell coupling and impulse propogation [sic] in health and disease. Boston: Kluwer Academic Pubulishers; 2002.
40. Levy MN, Pappano AJ. *Cardiovascular Physiology*, 9th ed. Philadelphia, PA: Mosby Elsevier; 2007.
41. Opie LH. *Heart Physiology: From Cell to Circulation*, 4th ed. Philadelphia: Lippincott Williams & Wilkins; 2004.
42. Pypendop BH, Verstegen JP. Hemodynamic effects of medetomidine in the dog: a dose titration study. *Vet Surg* 1998;27:612–622.
43. Gomez De Segura IA, Vazquez Moreno-Planas I, Benito J, Galiano A, De Miguel E. Electrophysiologic cardiac effects of the new local anesthetic IQB-9302 and of bupivacaine in the anesthetised dog. *Acta Anaesthesiol Scand* 2002;46:666–673.
44. Beckman KJ, Parker RB, Hariman RJ, Gallastegui JL, Javaid JI, Bauman JL. Hemodynamic and electrophysiological actions of cocaine. Effects of sodium bicarbonate as an antidote in dogs. *Circulation* 1991;83:1799–1807.
45. Bicer S, Nakayama T, Hamlin RL. Effects of chronic oral amiodarone on left ventricular function, ECGs, serum chemistries, and exercise tolerance in healthy dogs. *J Vet Intern Med* 2002;16:247–540.
46. Lee JD, Tajimi T, Patritti J, Ross J Jr. Preload reserve and mechanisms of afterload mismatch in normal conscious dog. *Am J Physiol* 1986;250(3 Pt 2):H464–H473.
47. Boon JA. *Manual of Veterinary Echocardiology*. Philadelphia: Lippincott Williams & Wilkins; 1998.
48. Kohn DF, Wixson SK, White WJ, Benson GJ. *Anesthesia and Analgesia in Laboratory Animals*. San Diego, CA: Academic Press; 1997.
49. Muir WW. III, Hubbell JAE, Skarda RT, Bednarski RM. *Handbook of Veterinary Anesthesia*, 3rd ed. St Louis, MO: Mosby; 2000.
50. Riley DC, Schmeling WT, al-Wathiqui MH, Kampine JP, Warltier DC. Prolongation of the QT interval by volatile anesthetics in chronically instrumented dogs. *Anesth Analg* 1988;67:741–749.
51. Gallagher JD. Effects of halothane and quinidine on intracardiac conduction and QTc interval in pentobarbital-anesthetized dogs. *Anesth Analg* 1992;75:688–695.
52. Shimizu W, McMahon B, Antzelevitch C. Sodium pentobarbital reduces transmural dispersion of repolarization and prevents torsades de pointes in models of acquired and congenital long QT syndrome. *J Cardiovasc Electrophysiol* 1999;10:154–164.

53. Takahara A, Sugiyama A, Satoh Y, Wang K, Honsho S, Hashimoto K. Halothane sensitizes the canine heart to pharmacological $I_{Kr}$ blockade. *Eur J Pharmacol* 2005;507:169–177.
54. Shimosato S, Etsten BE. Effect of anesthetic drugs on the heart: a critical review of myocardial contractility and its relationship to hemodynamics. *Clin Anesth* 1969;3:17–72.
55. Picker O, Scheeren TWL, Arndt JO. Inhalation anaesthetics increase heart rate by decreasing cardiac vagal activity in dogs. *Br J Anaesth* 2001;87:748–754.
56. Picker O, Schindler A, Scheeren TW. Accuracy and reproducibility of long-term implanted transit-time ultrasound flow probes in dogs. *Intensive Care Med* 2000;26:601–607.
57. www.dissdata.com.
58. Data Sciences International (www.datasci.com).
59. www.transonic.com.
60. Emka Technologies (www.emkatech.com).
61. www.vivometrics.com.
62. Buxco Electronics, Inc. (www.buxco.com).
63. Data Integrated Scientific Systems (www.itstelemetry.com).
64. LDS Test and Measurement (PoNeMah; http://www.lds-group.com).
65. Modular Instruments Inc. (www.mi2.com).
66. Notocord Systems (www.notocord.com).
67. Miura T, Miyazaki S, Guth BD, Indolfi C, Ross J Jr. Heart rate and force-frequency effects on diastolic function of the left ventricle in exercising dogs. *Circulation* 1994;89:2361–2368.
68. Kingma JG Jr, Armour JA, Rouleau JR. Left ventricular intramyocardial pressure determination using two different solid-state micromanometric pressure sensors. *Can J Physiol Pharmacol* 1996;74(6):701–705.
69. Hamlin B, Guth B, Sarazan D, Picker O, Schindler A, Scheeren TW. Accuracy and reproducibility of long-term implanted transit-time ultrasound flow probes in dogs. *Intensive Care Med* 2000;26:601–607.
70. Hamlin RL, Nakayama T, Nakayama H, Carnes CA. Effects of changing heart rate on electrophysiological and hemodynamic function in the dog. *Life Sci* 2003;72:1919–1930.
71. Oyama MA, Sisson DD, Bulmer BJ, Constable PD. Echocardiographic estimation of mean left atrial pressure in a canine model of acute mitral valve insufficiency. *J Vet Intern Med* 2004;18:667–672.
72. Goldberger JJ, Kadish AH, Johnson D, Qi X. New technique for vagal nerve stimulation. *J Neurosci Methods* 1999;91:109–114.
73. Bass AS, Robie NW. Stereoselectivity of S- and R-sulpiride for pre- and postsynaptic dopamine receptors in the canine kidney. *J Pharmacol Exp Ther* 1984;229:67–71.
74. Bass AS, Murphy MB, Kohli JD, Goldberg LI. Potentiation by dopexamine of the cardiac responses to circulating and neuronally released norepinephrine: a possible mechanism for the therapeutic effects of the drug. *J Cardiovasc Pharmacol* 1989;13:667–671.
75. Nolan ER, Feng MR, Koup JR, Liu J, Turluck D, Zhang Y, Paulissen JB, Olivier NB, Miller T, Bailie M/B. A novel predictive pharmacokinetic/pharmacodynamic model of repolarization prolongation derived from the effects of terfenadine, cisapride and E-4031 in the conscious chronic AV node—ablated, His bundle-paced dog. *J Pharmacol Toxicol Methods* 2006;53:1–10.

76. Toshida N, Hirao K, Yamamoto N, Tanaka M, Suzuki F, Isobe M. Ventricular echo beats and retrograde atrioventricular nodal exits in the dog heart: multiplicity in their electrophysiologic and anatomic characteristics. *J Cardiovasc Electrophysiol* 2001; 12:1256–1264.
77. Smith DA, van de Waterbeemd, Walker DK. *Pharmacokinetics and Metabolism in Drug Design*, 2nd ed. Weinheim: Wiley-VCH; 2006.
78. Schoenwald RD, ed. *Pharmacokinetics in Drug Discovery and Development*. Boca Raton, FL: CRC Press; 2002.
79. Boroujerdi M. *Pharmacokinetics: Principles and Applications*. New York: McGraw-Hill Medical Publications Division; 2001.
80. Markert M, Klumpp A, Trautmann T, Guth B. A novel propellant-free inhalation drug delivery system for cardiovascular drug safety evaluation in conscious dogs. *J Pharmacol Toxicol Methods* 2004;50:109–119.
81. Petruska JM, Beattie JG, Stuart BO, Pai S, Walters KM, Banks CM, Lulham GW, Mirro EJ. Cardiovascular effects after inhalation of large doses of albuterol dry powder in rats, monkeys, and dogs: a species comparison. *Fundam Appl Toxicol* 1997;40:52–62.
82. Gad SC. *Statistics and Experimental Design for Toxicologists and Pharmacologists*, 4th ed. Boca Raton, FL: CRC Press; 2005.
83. Jones B, Kenward MG. *Design and Analysis of Cross-over Trials*. London: Chapman & Hall; 1989.
84. Winer BJ, Brown DR, Michels KM. *Statistical Principles in Experimental Design*, 3rd ed. New York: McGraw Hill; 1991.
85. Snedecor GW, Cochran WG. *Statistical Methods*, 6th ed. Ames: Iowa State University Press; 1967.
86. Detweiler DK. Electrocardiographic monitoring in toxicological studies: principles and interpretations. *Adv Exp Med Biol* 1983;161:579–607.
87. Tilley LP, Burtnick NL. *ECG for the Small Animal Practitioner*. Jackson WY: Teton NewMedia; 1999.
88. Hanton G, Rabemampianina Y. The electrocardiogram of the beagle dog: reference values and effect of sex, genetic strain, body position and heart rate. *Lab Anim* 2006;40:123–136.
89. Roden DM, Anderson ME. Proarrhythmia. In Starke FK, Ed., *Handbook of Experimental Pharmacology*, Vol. 171. New York; Springer 2006, pp 73–97.7.
90. Fenichel RR, Malik M, Antzelevitch C, Sanguinetti M, Roden DM, Priori SG, Ruskin JN, Lipicky RJ, Cantilena LR. Independent Academic Task Force. Drug-induced torsades de pointes and implications for drug development. *J Cardiovasc Electrophysiol* 2004;15:475–495.
91. Soloviev MV, Hamlin RL, Shellhammer LJ, Barrett RM, Wally RA, Birchmeier PA, Schaefer GJ. Variations in hemodynamic parameters and ECG in healthy, conscious, freely moving telemetrized beagle dogs. *Cardiovasc Toxicol* 2006;6:51–62.
92. Watanabe H, Miyazaki H. A new approach to correct the QT interval for changes in heart rate using a nonparametric regression model in beagle dogs. *J Pharmacol Toxicol Methods* 2006;53:234–241.
93. Tattersall ML, Dymond M, Hammond T, Valentin JP. Correction of QT values to allow for increases in heart rate in conscious beagle dogs in toxicology assessment. *J Pharmacol Toxicol Methods* 2006;53:11–19.
94. Batey AJ, Doe CP. A method for QT correction based on beat-to-beat analysis of the QT/RR interval relationship in conscious telemetred beagle dogs. *J Pharmacol Toxicol Methods* 2002;48:11–19.

95. Raunig D, Depasquale MJ, Huang CH, Winslow R, Fossa AA. Statistical analysis of QT interval as a function of changes in RR interval in the conscious dog. *J Pharmacol Toxicol Methods* 2001;46:1–11.
96. Malik M, Camm AJ. Evaluation of drug-induced QT interval prolongation: implications for drug approval and labelling. *Drug Safety* 2001;24:323–351.
97. Fridericia LS. Die systolendauer im elektrokardogramm bei normalen menchen und bei Herzkranken. *Acta Med Scand* 1920;53:469–486.
98. Spence S, Soper K, Hoe CM, Coleman J. The heart rate-corrected QT interval of conscious beagle dogs: a formula based on analysis of covariance. *Toxicol Sci* 1998;45:247–258.
99. Osborne BE, Leach GDH. The beagle electrocardiogram. *Fundam Cosmet Toxicol* 1971;9:2857–2864.
100. Fossa AA, Wisialowski T, Crimin K. QT prolongation modifies dynamic restitution and hysteresis of the beat-to-beat QT-TQ interval relationship during normal sinus rhythm under varying states of repolarization. *J Pharmacol Exp Ther* 2006;316:498–506.
101. Bazett HC. An analysis of the time-relations of electrocardiograms. *Heart* 1920;7:353–370.
102. Dessertenne F. La tachycardia ventriculaire a deux foyers opposes variables. *Arch Mal Coeur* 1996;59:263–272.
103. Sanguinetti MC, Jiang C, Curran ME, Keating MT. A mechanistic link between and inherited and acquired cardiac arrhythmia: HERG encodes the iKr potassium channel. *Cell* 1995;81:299–307.
104. Trudeau MC, Warmke JW, Ganetzky B, Robertson GA. HERG, a human inward rectifier in the voltage-gated potassium channel family. *Science* 1995;269:92–95.
105. Curran ME, Splawski I, Timothy KW, Vincent GM, Green ED, Keating MT. A molecular basic for cardiac arrhythmia: HERG mutations cause long QT syndrome. *Cell* 1995;80:795–803.
106. Zhou Z, Gong Q, Ye B, Fan Z, Makielski JC, Robertson GA, January CT. Properties of HERG channels stably expressed in HEK293 cells studied at physiological temperature. *Biophys J* 1998;74:230–241.
107. Walker BD, Valenzuela SM, Singleton CB, Tie H, Bursill JA, Wyse KR, Qiu MR, Breit SN, Campbell TJ. Inhibition of HERG channels stably expressed in a mammalian cell line by the antianginal agent perhexiline maleate. *Br J Pharmacol* 1999;127:243–251.
108. Witchel HJ, Milnes JT, Mitcheson JS, Hancox JC. Troubleshooting problems with *in vitro* screening of drugs for QT interval prolongation using HERG K+ channels expressed in mammalian cell lines and Xenopus oocytes. *J Pharmacol Toxicol Methods* 2002;48:65–80.
109. Kirsch GE, Trepakova ES, Brimecombe JC, Sidach SS, Erickson HD, Kochan MC, Shyjka LM, Lacerda AE, Brown AM. Variability in the measurement of hERG potassium channel inhibition: effects of temperature and stimulus pattern. *J Pharmacol Toxicol Methods* 2004;50:93–101.
110. Guo L, Guthrie H. Automated electrophysiology in the preclinical evaluation of drugs for potential QT prolongation. *J Pharmacol Toxicol Methods* 2005;52(1):123–135.
111. Gintant GA, Su Z, Martin RL, Cox BF. Utility of hERG assays as surrogate markers of delayed cardiac repolarization and QT safety. *J Toxicol Pathol* 2006;34:81–90.
112. Yao J-A, Du X, Lu D, Baker RL, Daharsh E, Atterson P. Estimation of potency of HERG channel blockers: impact of voltage protocol and temperature. *J Pharmacol Toxicol Methods* 2005;52:146–153.

113. Herron W, Towers C, Templeton A. Experiences in method development for the analysis of *in vitro* study solutions for content. *J Pharmacol Toxicol Methods* 2004;49:211–216.
114. Redfern WS, Carlsson L, Davis AS, Lynch WG, MacKenzie I, Pelethorpe Sl, Siegl PKS, Strang I, Sullivan AT, Wallis R, Camm AJ, Hammond TG. Relationships between preclinical cardiac electrophysiology, clinical QT interval prolongation and torsades de pointes for a broad range of drugs: evidence for a provisional safety margin in drug development. *Cardiovasc Res* 2003;58:32–45.
115. Diaz GJ, Daniell K, Leitza ST, Martin RL, Su Z, McDermott JS, Cox BF, Gintant GA. The [$^3$H] dofetilide binding assay is a predictive screening tool for hERG blockade and proarrhythmia: comparison of intact cell and membrane preparations and effects of altering $[K^+]_o$. *J Pharmacol Toxicol Methods* 2004;50:187–199.
116. Chaudhary KW, O'Neal JM, Mo Z-L, Fermini B, Gallavan RH, Bahinski A. Evaluation of the rubidium efflux assay for preclinical identification of hERG blockade. *Assay Drug Dev Technol* 2006;4(1):73–82.
117. Baxter DF, Kirk M, Garcia AF, Raimondi A, Holmqvist MH, Flint KK, Bojanic D, Distefano PS, Curtis Rl, Xie Y. A novel membrane potential-sensitive fluorescent dye improves cell-based assays for ion channels. *J Biomol Screen* 2002;7(1):79–85.
118. Bril A, Gout B, Bonhomme M, Landais L, Faivre J, Linee P, Poyser RH, Ruffolo RR Jr. Combined potassium and calcium channel blocking activities as a basis for antiarrhythmic efficacy with low proarrhythmic risk: experimental profile of BRL-32872. *J Pharmacol Exp Ther* 1996;509:129–137.
119. Zhang S, Zhou Z, Gong Q, Makielski JC, January CT. Mechanism of block and identification of the verapamil binding domain to HERG potassium channels. *Circ Res* 1999;84:989–998.
120. Martin RL, McDermott JS, Salmen HJ, Palmatier J, Cox BF, Gintant GA. The utility of hERG and repolarization assays in evaluating delayed cardiac repolarization: Influence of multi-channel block. *J Cardiovasc Pharmacol* 2004;43:369–379.
121. Ficker El, Kuryshev YA, Dennis AT, Oberjero-Paz C, Wang L, Hawryluk P, Wible BA, Brown AM. Mechanisms of arsenic-induced prolongation of cardiac repolarization. *Mol Pharmacol* 2004;66:33–44.
122. Drolet B, Simard C, Roden DM. Unusual effects of a QT-prolonging drug, arsenic trioxide, on cardiac potassium currents. *Circulation* 2004;109:26–29.
123. Kuryshev YA, Ficker E, Wang L, Hawryluk P, Dennis AT, Wible BA, Brown AM, Kang J, Chen XL, Sawamura K, Reynolds Wl, Rampe D. Pentamidine-induced long QT syndrome and block of HERG trafficking. *J Pharmacol Exp Ther* 2005;312(1):316–323.
124. Cordes JS, Sun Z, Lloyd DB, Bradley JA, Opsahl AC, Tengowski MW, Chen X, Zhou J. Pentamidine reduces hERG expression to prolong the QT interval. *Br J Pharmacol* 2005;145(1):15–23.
125. Wible BA, Hawryluk P, Ficker E, Kuryshev YA, Kirsch G, Brown AM. HERG-lite: a novel comprehensive high throughput screen for drug-induced hERG risk. *J Pharmacol Toxicol Methods* 2005;52(1):136–145.
126. Glaaser IW, Kass RS, Clancy CE. Mechanisms of genetic arrhythmias: from DNA to ECG. *Prog Cardiovasc Dis* 2003;46(3):259–270.
127. Roden DM, Viswanathan PC. Genetics of acquired long QT syndrome. *J Clin Invest* 2005;115(8):2025–2032.
128. Roden DM. Long QT syndrome: reduced repolarization reserve and the genetic link. *J Intern Med* 2006;259:59–69.
129. Franz MR. Current status of monophasic action potential recording: theories measurements and interpretations. *Cardiovasc Res* 1999;41:25–40.

130. Hondeghem LM, Carlsson L, Duker G. Instability and triangulation of the action potential predict serious proarrhythmia but APD prolongation is antiarrhythmic. *Circulation* 2001;103:2004–2013.

131. Strauss HC, Bigger JT Jr, Hoffman BF. Electrophysiological and beta-receptor blocking effects of MJ 1999 on dog and rabbit cardiac tissue. *Circ Res* 1970;26(6):661–678.

132. Sosunov EA, Anyukhovsky EP, Rosen MR. Effects of quinidine on repolarization in canine epicardium midmyocardium, and endocardium. I. *In vitro* study. *Circulation* 1997;96:4011–4018.

133. Wang L, Feng Z-P, Kondo CS, Sheldon RS, Duff HJ. Developmental changes in the delayes rectifier K+ channels in mouse heart. *Circ Res* 1996;79:79–85.

134. Liu GX, Zhou J, Nattel S, Koren G. Single-channel recordings of a rapid delayed rectifier current in adult mouse ventricular myocytes; basic properties and effects of divalent cations. *J Physiol (London.)* 2004;556(2):401–413.

135. Wymore RS, Gintant GA, Wymore RT, Dixon JE, McKinnon D, Cohen IS. Tissue and species distribution of mRNA for the iKr-like $K^+$ channel, erg. *Circ Res* 1997;80:261–268.

136. McDermott JS, Salmen HJ, Cox BF, Gintant GA. Importance of species selection in arrhythmogenic models of QT interval prolongation. *Antimicrob Agents Chemother* 2002;46(3):938–939.

137. January CT, Riddle JM. Early afterdepolarizations: mechansim of induction and block. A role for L-type Ca2+ current. *Circ Res* 1999;64(5):977–990.

138. Lu HR, Vlaminckx E, Van Ammel K, De Clerck F. Drug-induced long QT in isolated rabbit Purkinje fibers; importance of action potential duration, triangulation and early afterdepolarizations. *Eur J Pharmacol* 2002;452(2):183–192.

139. Hondeghem LM. TRIad; foundation for proarrhythmia (triangulation, reverse use dependence and instability). *Novartis Found Symp* 2005;266:235–244.

140. Brugada R, Hong K, Cordeiro JM, Dumaine R. Short QT syndrome. *CMAJ* 2005;173(11):1349–1354

141. Horn R, Marty A. Muscarinic activation of ionic currents measured by a new whole-cell recording method. *J Gen Physiol* 1988;92:145–159.

142. Zaniboni M, Pollard AE, Yang L, Spitzer KW. Beat-to-beat repolarization variability in ventricular myocytes and its suppression by electrical coupling. *Am J Physiol Heart Circ Physiol* 2000;278(3):H677–H687.

143. Sawanobori T, Adaniya H, Namiki T, Hiraoka M. Rate-dependent effects of sematilide on action potential duration in isolated guinea pig ventricular myocytes. *J Pharmacol Exp Ther* 1994;271(1):302–310.

144. Salata JJ, Jurkiewicz NK, Wallace AA, Stupienski RF 3rd, Guinosso PJ Jr, Cardiac electrophysiological actions of the histamine H1-receptor antagonists astemizole and terfenadine compared with chlorpheniramine and pyrilamine. *Circ Res* 1995;76:110–119.

145. Gintant GA. Azimilide causes reverse rate-dependent block while reducing both components of delayed rectifier current in canine ventricular myocytes. *J Cardiovasc Pharmacol* 1998;31:945–953.

146. Davie C, Pierre-Valentin J, Pollard C, Standen N, Mitcheson J, Alexander P, Thong B. Comparative pharmacology of guinea pig cardiac myocyte and cloned hERG (IKr) channel. *J Cardiovasc Electrophysiol* 2004;15:1302–1309.

147. Gintant GA, Limberis JT, McDermott JS, Wegner CD, Cox BF. The canine Purkinje fiber: an *in vitro* model system for acquired long QT syndrome and drug-induced arrhythmogenesis. *J Cardiovasc Pharmacol* 2001;37:607–618.

148. Lu HR, Vlaminckx E, Teisman A, Gallacher DJ. Choice of cardiac tissue plays an important role in the evaluation of drug-induced prolongation of the QT interval *in vitro* in rabbit. *J Pharmacol Toxicol Methods* 2005;52(1):90–105.

149. Varro A, Nakaya Y, Elharrar V, Surawicz B. The effects of amiodarone on repolarization and refractoriness of cardiac fibers. *Eur J Pharmacol* 1988;154(1):11–18.

150. Kii Y, Hayashi S, Tabo M, et al. QT PRODACT. Evaluation of the potential of compounds to cause QT interval prolongation by action potential assays using guinea-pig papillary muscles. *J Pharmacol Sci* 2006;99(5):449–457.

151. Yan G-X, Shimizu W, Antzelevitch C. Characteristics and distribution of M cells in arterially perfused canine left ventricular wedge preparations. *Circulation* 1998;98:1921–1927.

152. Yan G-X, Rials SJ, Wu Y, Liu T, Xu X, Marinchak RA, Kowey PR. Ventricular hypertrophy amplifies transmural repolarization dispersion and induces early afterdepolarization. *Am J Physiol Heart Circ Physiol* 2001;281:H1968–H1975.

153. Di Diego JM, Belardinelli L, Antzelevitch C. Cisapride-induced transmural dispersion of repolarization and torsades de pointes in the canine left ventricular wedge preparation during epicardial stimulation. *Circulation* 2003;108:1027–1033.

154. Yan G-X, Wu Y, Liu T, Wang J, Marinchak RA, Kowey PR. Phase 2 early afterdepolarizations as a trigger of polymorphic ventricular tachycardia in acquired long-QT syndrome. Direct evidence from intracellular recordings in the intact left ventricular wall. *Circulation* 2001;103:2851–2856.

155. Sutherland FJ, Hearse DJ. The isolated blood and perfusion fluid perfused heart. *Pharmacol Res* 2000;41(6):613–627.

156. Hondeghem LM, Hoffmann P. Blinded test in isolated female rabbit heart reliably identifies action potential duration prolongation and proarrhythmic drugs: importance of triangulation, reverse use-dependence, and instability. *J Cardiovasc Pharmacol* 2003;41:14–24.

157. Shah RR, Hondeghem LM. Refining detection of drug-induced proarrhythmia: QT interval and TRIaD. *Heart Rhythm* 2005;2:758–772.

158. Hamlin RL, Cruze CA, Mittelstadt SW, Kijtawornrat A, Keene BW, Roche BM, Takayama T, Nakayama H, Hamlin DM, Arnold T. Sensitivity and specificity of isolated perfused guinea pig heart to test for drug-induced lengthening of QTc. *J Pharmacol Toxicol Methods* 2004;49:15–23.

159. Eckardt L, Haverkamp W, Mertens H, Johna R, Clague JR, Borggrefe M, Beirthardt G. Drug-related torsades de pointes in the isolated rabbit heart: comparison of clofilium D,L-sotalol, and erythromycin. *J Cardiovasc Pharmacol* 1998;32:425–434.

160. Johna R, Mertens H, Haverkamp W, Eckardt L, Niederboker T, Borggrefe M, Breithardt G, Clofilium in the isolated perfused rabbit heart: a new model to study proarrhythmia induced by class III antiarrhythmic drugs. *Basic Res Cardiol* 1998;93:127–135.

161. Akoun GM, Milleron BJ, Mayaund CM, Francois TJ. Natural history of drug-induced pneumonitis. In Akoun GM, White PJ, Eds. Drug Induced Disorders, Vol. 3. Elsevier Scientific Publishers; 1989, pp 3–9.

162. Rosenow EC, Myers JL, Swensen SJ, Pisani RJ. Drug induced pulmonary disease: an update. *Chest* 1992;102:239–250.

163. Eldridge FL, Millhorn DE. Central regulation of respiration by endogenous neurotransmitters and neuromodulators. *Annu Rev Physiol* 1981;43:121–135.

164. Mueller RA, Lundberg DBA, Breese GR, Hedner J, Hedner T, Jonason J. The neuropharmacology of respiratory control. *Pharmacol Rev* 1982;34:255–285.

165. Heymans C. Action of drugs on carotid body and sinus. *Pharmacol Rev* 1955;7:119–142.
166. Heymans C, Neil E. The effects of drugs on chemoreceptors. In Heymans C, Neil E, eds. *Reflexogenic Areas of the Cardiovascular System*. London: Churchill Ltd; 1958; pp 192–199.
167. Sant'Ambrogio G. Nervous receptors of the tracheobronchial tree. *Annu Rev Physiol* 1987;49:611–627.
168. Phillipson EA. Disorders in the control of breathing: hypoventilation syndromes. In Murray JF, Nadal JA, eds. *Textbook of Respiratory Medicine*. Philadelphia: Saunders; 1988, pp 1831–1840.
169. Cherniack NS. Disorders in the control of breathing: hyperventilation syndromes. In Murray JF, Nadal JA, eds. *Textbook of Respiratory Medicine*. Philadelphia: Saunders; 1988, pp 1861–1866.
170. Laffey JG. Kavavagh BP. Hypocapnia. *N Engl J Med* 2002;347(1):43–53.
171. Boggs DF. Comparative control of respiration. In Parent RA, ed. *Comparative Biology of the Normal Lung*, Vol I. Boca Raton FL: CRC Press; 1992, pp 309–350.
172. Keats AS. The effects of drugs on respiration in man. *Annu Rev Pharmacol Toxicol* 1985;25:41–65.
173. McCarthy DS. Airflow obstruction. In Kryger MH, eds. *Pathophysiology of Respiration*. Hoboken, NJ: Wiley; 1981, pp 7–42.
174. Rochester DF, Arora NS. Respiratory muscle failure. *Med Clin Studies North Am* 1983;67:573–597.
175. Milic-Emili J. Recent advances in clinical assessment of control of breathing. *Lung* 1982;160:1–17.
176. Warren CPW. Lung restriction. In Kryger MH, ed. *Pathophysiology of Respiration*. Hoboken, NJ: Wiley; 1981, pp 43–69.
177. Drorbaugh JE, Fenn WO. A barometric method for measuring ventilation in newborn infants. *Pediatrics* 1955;16:81–86.
178. Jacky JP. Barometric measurement of tidal volume: effects of pattern and nasal temperature. *J Appl Physiol Respir Environ Exere Physiol* 1980;49(2):319–325.
179. Zimmerman PV, Connellan SJ, Middleton HC, Tabona MV, Goldman MD, Pride N. Postural changes in rib cage and abdominal volume-motion coefficients and their effect on the calibration of a respiratory inductance plethysmograph. *Am Rev Respir Dis* 1983;127:209–214.
180. Murphy DJ, Joran ME, Renninger JE. Effects of adenosine agonists and antagonists on pulmonary ventilation in conscious rats. *Gen Pharmacol* 1993;24(4):943–954.
181. Murphy DJ, Joran ME, Grando JC. Microcapnometry: a non-invasive method for monitoring arterial $CO_2$ tension in conscious rats. *Toxicol Methods* 1994;4(3):177–187.
182. Murphy DJ, Joran ME, Grando JC. A non-invasive method for distinguishing central from peripheral nervous system effects of respiratory depressant drugs in conscious rats. *Gen Pharmacol* 1995;26(3):569–575.
183. Watchko JF, Standaert TA, Mayock DE, Twiggs G, Woodrum DE. Ventilatory failure during loaded breathing: the role of central neural drive. *J Appl Physiol* 1988;65(1):249–255.
184. Nuzzo PF, Anton WR. Practical applications of capnography. *Respir Therapy* 1986;16:12–17.
185. Lucke JN, Hall GM, Lister D. Porcine malignant hyperthermia I: metabolic and physiologic changes. *Br J Anesthesiol* 1976;48:297–304.

186. Becklake M, Crapo RO. American Thoracic Society Statement: Lung function testing: selection of reference values and interpretative strategies. *Am Rev Respir Dis* 1991;144:1202–1218.

187. Diamond L, O'Donnell M. Pulmonary mechanics in normal rats. *J Appl Physiol Respir Environ Exerc Physiol* 1977;43:942–948.

188. Mauderly JL. Effect of inhaled toxicants on pulmonary function. In McClellan RO, Henderson RF, eds. *Concepts in Inhalation Toxicology*. New York: Hemisphere Publishing; 1989, pp 347–401.

189. Amdur MO, Mead J. Mechanics of respiration in unanesthetized guinea pigs. *Am J Physiol* 1958;192:364–368.

190. O'Neil JJ, Raub JA. Pulmonary function testing in small laboratory animals. *Environ Health Perspect* 1984;56:11–22.

191. Mauderly JL. The influence of sex and age on the pulmonary function of the beagle dog. *J Gerontol* 1974;29:282–289.

192. Douglas JS, Dennis MW, Ridgeway P, Bouhuys A. Airway dilation and constriction in spontaneously breathing guinea pigs. *J Pharmacol Exp Ther* 1971;180:98–109.

193. Kruger JJ, Bain T, Patterson JL Jr. Elevation gradient of intrathoracic pressure. *J Appl Physiol* 1961;16:465–468.

194. Murphy DJ, Renninger JP, Gossett KA. A novel method for chronic measurement of pleural pressure in conscious rats. *J Pharmacol Toxicol Methods* 1998;39:137–141.

195. Murphy DJ, Renninger JP, Coatney RW. A novel method for chronic measurement of respiratory function in the conscious monkey. *J Pharmacol Toxicol Methods* 2001;46:13–20.

196. Pennock BE, Cox CP, Rogers RM, Cain WA, Wells JH. A noninvasive technique for measurement of changes in specific airway resistance. *J Appl Physiol* 1979;46:39–406.

197. Hamelmann E, Schwarze J, Takeda K, Oshiba A, Larsen GL, Irvin CG, Gelfand EW. Noninvasive measurement of airway responsiveness in allergic mice using barometric plethysmography. *Am J Respir Crit Care Med* 1997;156:766–775.

198. Lundblad LKA, Irvin CG, Adler A, Bates JHT. A re-evaluation of the validity of unrestrained plethysmography in mice. *J Appl Physiol* 2002;93:1198–1207.

199. Celli BR. American Thoracic Society Statement: Standards for the diagnosis and care of patients with chronic obstructive pulmonary disease. *Am J Respir Crit Care Med* 1995;152 (5) Suppl:S78–S121.

200. Brown LK, Miller A. Full lung volumes: functional residual capacity, residual volume and total lung capacity. In Miller A, ed. *Pulmonary Function Tests: A Guide for the Student and House Officer*. New York: Grune & Stratton; 1987, pp 53–58.

201. King TKC. Measurement of functional residual capacity in the rat. *J Appl Physiol* 1966;21:233–236.

202. Palecek F. Measurement of ventilatory mechanics in the rat. *J Appl Physiol* 1969;27:149–156.

203. Smolders FDJ, Folgering HTh, Bernards JA. Ventilation estimated from efferent phrenic nerve activity in the paralyzed cat. *Pflugers Arch* 1975;359:157–169.

204. Chitravanshi VC, Sapru HN. Phrenic nerve responses to chemical stimulation of the sub regions of ventral medullary respiratory neuronal group in the rat. *Brain Res* 1999;821:443–460.

205. Syabbalo N. Assessment of respiratory muscle function and strength. *Postgrad Med J* 1998;74:208–215.

206. Borison HL. Central nervous system depressants: control-systems approach to respiratory depression. *Pharmacol Ther B* 1997;3:211–226.

207. Trenchard D, Gardner D, Guz A. Role of pulmonary vagal afferent nerve fibers in the development of rapid shallow breathing in lung inflammation. *Clin Sci* 1972;42:251–263.

208. Drazen JM. Physiological basis and interpretation of indices of pulmonary mechanics. *Environ Health Perspect* 1984;56:3–9.

209. Costa DL, Raub JA, Tepper JS. Comparative respiratory physiology of the lung. In Parent RA, ed. *Comparative Biology of the Normal Lung*, Vol I. Boca Raton, FL: CRC Press; 1992, pp 175–402.

210. Hakkinen PJ, Witschi HP. Animal models. In Witschi HP, Brain JD, eds. *Handbook of Experimental Pharmacology*, Vol 75. New York: Springer-Verlag; 1985, pp 95–114.

211. Anon. *ICH S7A: Safety Pharmacology Studies for Human Pharmaceuticals*. European Agency for the Evaluation of Medicinal Products. Evaluation of Medicines for Human Use. CPMP/ICH/539/00, London, 16 November 2000.

212. Porsolt RD, Lemaire M, Dürmüller N, Roux S. New perspectives in CNS safety pharmacology. *Fundam Clin Pharmacol* 2002;16:197–207.

213. Porsolt RD. Central nervous system (CNS) safety pharmacology studies. In Vogel GH, Hock FJ, Maas J, Mayer D. eds. *Drug Discovery and Evaluation: Safety and Pharmacokinetic Assays*. Heidelberg: Springer Verlag; 2006, pp 15–60.

214. Irwin S. Comprehensive behavioral assessment: a systematic quantitative procedure for assessing the behavioral and physiologic state of the mouse. *Psychopharmacologia* 1968;13:222–257.

215. Mattson JL, Spencer PJ, Albee RR. A performance standard for clinical and functional observational battery examination of rats. *J Am Coll Toxicol* 1996;15:239–250.

216. Reiter LR, McPhail RC. Motor activity: a survey of methods with potential use in toxicity testing. *Neurobehav Toxicol* 1979;1 (Suppl):53–66.

217. Dunham NW, Miya TS. A note on a simple apparatus for detecting neurological deficit in mice and rats. *J Am Pharm Assoc* 1957;46:208–209.

218. Porsolt RD, Picard S, Lacroix P. International safety pharmacology guidelines (ICH S7A and S7B): where do we go from here? *Drug Dev Res* 2005;64:83–89.

219. Krall E. Antiepileptic drug development: anticonvulsant drug screening. *Epilepsia* 1978;19:409–428.

220. Simon P, Chermat R, Doaré L, Bourin M, Farinotti R. Interactions imprévues de divers psychotropes avec les effets du barbital et du pentobarbital chez la souris. *J Pharmacol (Paris)* 1982;13:241–252.

221. Bammer C. Pharmacological investigations of neurotransmitter involvement in passive avoidance responding: a review and some new results. *Neurosci Biobehav Rev* 1982;6:247–296.

222. Glick SD, Zimmerberg D. Amnesic effects of scopolamine. *Behav Biol* 1972;7:2445–2454.

223. Morris RGM. Spatial localization does not require the presence of local cues. *Learn Motiv* 1981;12:239–260.

224. Olton DS. The radial maze as a tool in behavioral pharmacology. *Physiol Behav* 1986;40:793–797.

225. Dunnett SB, Evenden JL, Iversen SD. Delay-dependent short-term memory deficits in aged rats. *Psychopharmacology* 1988;96:174–180.

226. Porsolt RD, Roux S, Wettstein JG. Animal models of dementia. *Drug Dev Res* 1995;35:214–229.
227. Prendergast MA, Jackson WJ, Terry AV, Decker MW, Arneric SA, Buccafusco JJ. Central nicotinic receptor agonists ABT-418, ABT-089, and (–)-nicotine reduce distractibility in young-adult monkeys. *Psychopharmacology* 1998;136:50–58.
228. Bartus RT, Flicker C, Dean RL. Logical principles for the development of animal models of age-related memory impairments. In Crook T, Bartus RT, Ferris S, Gershon S, eds. *Assessment in Psychopharmacology*. Madison, WI: Mark Powley Associates; 1983, pp 263–299.
229. Dürmüller N, Guillaume P, Lacroix P, Porsolt RD, Moser P. Use of dog EEG in safety pharmacology. *J Pharmacol Toxicol Methods* 2007.
230. Itil TM. The discovery of psychotropic drugs by computer-analyzed cerebral bioelectrical potentials (CEEG). *Drug Dev Res* 1981;1:373–407.
231. Van Riezen H, Glatt AF. Introduction and history of the use of electroencephalography in animal drug studies. *Neuropsychobiology* 1993;28:118–125.
232. Jouvet M. Biogenic amines and the states of sleep. *Science* 1969;163:32–41.
233. Ruigt GSF, van Proosdij JN, van Delft AML. A large scale, high resolution automated system for rat sleep staging. *EEG Clin Neurophysiol* 1989;73:52–71.
234. Anon. *Guideline on the Non-clinical Investigation of the Dependence Potential of Medicinal Products*. EMEA/CHMP/SWP/94227/2004, London, 23 March 2006.
235. Balster RL, Bigelow GE. Guidelines and methodological reviews concerning drug abuse liability assessment. *Drug Alcohol Assess* 2003;70 (Suppl 3):S13–S40.
236. Balster RL. Drug abuse potential evaluation in animals. *Br J Addiction* 1991;86:1549–1558.
237. Goudie AJ, Harrison AA, Leathley MJ. Evidence for a dissociation between benzodiazepine withdrawal signs. *NeuroReport* 1993;4:295–299.
238. Lal H, ed. *Discriminative Stimulus Properties of Drugs*. New York: Plenum Press; 1977.
239. Gerak LR, France CP. Discriminative stimulus effects of nalbuphine in rhesus monkeys. *J Pharmacol Exp Ther* 1996;276:523–531.
240. Foltin RW, Fischman MW. The cardiovascular and subjective effects of intravenous cocaine and morphine combinations in humans. *J Pharmacol Exp Ther* 1992;261:623–632.
241. Schechter MD, Calcagnetti DJ. Trends in place preference conditioning, with a cross-indexed bibliography 1957–1991. *Neurosci Biobehav Rev* 1993;17:21–41.
242. Brady JV, Fischman MW. Assessment of drugs for dependence potential and abuse liability: an overview. In Seiden LS, Balster RL, eds. *Behavioral Pharmacology: The Current Status*. New York: Alan R Liss; 1985, pp 361–382.
243. Fozzard HA, Task Force of the Working Group on Arrhythmias of the European Society of Cardiology. The Sicilian gambit: A new approach to the classification of antiarrhthmic drugs based on their actions on arrhythmogenic mechanisms. *Circulation* 1991;84:1831–1851.

# SAFETY ASSESSMENT OF BIOTECHNOLOGY-DERIVED THERAPEUTICS

Mary Ellen Cosenza
*Amgen Inc., Thousand Oaks, California*

**Contents**

1 Introduction
2 Species Specificity and Selection
  2.1 Species Selection
  2.2 Nonhuman Primates
  2.3 Surrogates and Other Alternatives
3 Immunogenicity
  3.1 Types of Antibodies to Biotechnology-Derived Pharmaceuticals
  3.2 Study Design Issues
4 Special Studies
  4.1 Genotoxicity
  4.2 Reproductive Toxicology
  4.3 Carcinogenicity
  4.4 Safety Pharmacology
  4.5 Local Tolerance
5 Future Trends
  References

## 1 INTRODUCTION

Over the course of the last 20 years, the biotechnology field has grown and is now a large part of pharmaceutical therapeutics development. According to a 2006 report from the Pharmaceutical Research and Manufacturers of America (PhRMA)

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

**TABLE 1 Categories of Biotechnology-Derived Pharmaceuticals**

| |
|---|
| Proteins |
| Hormones |
| Blood products |
| Cytokines |
| Growth factors |
| Ligands and receptors |
| Modified human proteins |
| Antibodies |
| Murine |
| Chimeric |
| Humanized |
| Fully human |
| Fragments |
| Gene therapies |
| Cellular therapies |
| Engineered tissue products |
| Vaccines |

[1], approximately 125 biotechnology-derived pharmaceuticals have been approved and are on the market. In addition, over 418 biotechnology-derived pharmaceuticals are in development for over 100 different diseases, and many of these potential therapeutics are being developed for grievous illnesses such as cancer, autoimmune diseases, and AIDS/HIV. Therapeutics derived from biotechnology (often referred to simply as "biologics") fall into several categories, including growth factors, hormones, enzymes, receptors, monoclonal antibodies, cytokines, vaccines, and gene and cell therapies (Table 1).

Many of the early biotechnology-derived therapeutics were replacement molecules, that is, recombinant-derived proteins largely unaltered from the endogenous counterparts. These included recombinant forms of human insulin, human growth hormone, interferon, erythropoietin, clotting factors, and colony stimulating factors (G-CSF and GM–CSF). Enzymes such as Activase (TPA) and Pulmozyme (DNAase) were developed in the late 1980s to early 1990s. Over time, modified versions of these proteins have been developed (e.g., pegylated proteins) as well as humanized and fully human monoclonal antibodies. More complicated molecules have been developed as well, such as fusion molecules (e.g., Enbrel, a peptide fused to a human IgG Fc) and conjugated antibodies (e.g., Mylotarg, a recombinant humanized antibody conjugated to a cytotoxic antibiotic) (Table 2). Early monoclonal antibodies were murine derived and chimerized; with the advent of more advanced technologies, monoclonal antibodies recently approved or currently in development are more humanized or fully human. These molecules bind to a specific antigen or epitope on a cell receptor. In addition, biotechnology-derived molecules are being designed to address more complicated diseases. New molecules are being developed to subtly alter immune function and to target specific ligand and receptor interactions. The background disease states are also more complicated and many molecules will be used in combinations with traditional pharmaceuticals or other biologics.

The regulation of these products differs by region and by type of product (e.g., replacement proteins versus DNA vaccines). Several years ago, a guidance from the

**TABLE 2 Examples of Marketed Biotechnology Products**[a]

| Year of Approval | Approved Biologics | Trade Name |
|---|---|---|
| 1982 | Insulin | Humulin® |
| 1985 | Growth hormone | Protropin® |
| 1986 | Interferon alfa | Roferon®, Intron A® |
| 1986 | Muromonab-CD3 | Orthoclone OKT® 3 |
| 1987 | Alteplase (TPA) | Activase® |
| 1989 | Epoetin alfa | Epogen® |
| 1990 | Interferon gamma | Actimmune® |
| 1991 | Filgrastim (G-CSF) | Neupogen® |
| | Sargramostim (GM-CSF) | Leukine® |
| 1992 | Aldesleukin (IL-2) | Proleukin® |
| 1993 | Interferon beta | Betaseron® |
| | Dornase alfa (DNase) | Pulmozyme® |
| 1994 | Imiglucerase (β-glucocerebrosidase) | Cerezyme® |
| | Abciximab | ReoPro® |
| 1997 | Rituximab | Rituxan® |
| | Daclizumab | Zenapax® |
| | Oprelvekin (IL-11) | Neumega® |
| 1998 | Trastuzumab | Herceptin® |
| | Infliximab | Remicade® |
| | Basiliximab | Simulect® |
| | Palivizumab | Synagis® |
| 2001 | Pegfilgrastim (pegylated G-CSF) | Neulasta® |
| | Darbepoetin alfa | Aranesp® |
| | Alemtuzumab | Campath® |
| 2003 | Alefacept | Amevive® |
| | Tositumomab | Bexxar® |
| | Efalizumab | Raptiva® |
| 2004 | Bevacizumab | Avastin® |
| | Natalizumab | Tysabri® |
| | Palifermin (KGF) | Kepivance® |

[a]DNase, deoxyribonuclease I; G-CSF, granulocyte colony-stimulating factor; GM-CSF, granulocyte-macrophage colony-stimulating factor; IL-2, interleukin 2; IL-11, interleukin 11; KGF, keratinocyte growth factor; TPA, tissue plasminogen activator.
*Source*: www.fda.gov.

International Congress on Harmonisation was developed to address the preclinical development and safety issues of most of these types of products (ICH S6 *Preclinical Safety Evaluation of Biotechnology-Derived Pharmaceuticals*). ICH S6 defined biotechnology-derived pharmaceuticals as "products derived from characterized cells through the use of a variety of expression systems including bacteria, yeast, insect, plant, and mammalian cells." The scope of this guidance included proteins, peptides, derivatives of these, or products of which they are components. It also states that these principles may apply to "recombinant DNA protein vaccines, chemically synthesized peptides, plasma derived products, endogenous proteins extracted from human tissues, and oligonucleotide drugs," but does not cover "antibiotics, allergenic extracts, heparin, vitamins, cellular blood components, conventional bacterial or viral vaccines, DNA vaccines, or cellular and gene therapies."

Regulatory agencies throughout the world define some of these products differently, which may affect how they are regulated. The main theme of the ICH S6 guidance is to create a case-by-case, science-driven approach to biotechnology preclinical product development. Each molecule should be fully examined for both its physical and pharmacologic properties. Its pharmacologic attributes need to be understood before embarking on a preclinical safety assessment development plan. The goal of any preclinical development plan is the same: to provide information for designing and conducting clinical trials. The biggest challenge is to use all of the information available and to conduct the best preclinical, *in vitro* and *in vivo*, experiments to best predict what will happen in humans. A rational, science-based approach is the best way to develop biotechnology-derived products [2].

Many traditional small-molecule studies are not very appropriate for biologics [3]: some for "technical" reasons (the protein may be too large to enter an ion channel) and other studies for scientific reasons (monoclonal antibodies would not be expected to have direct genotoxic effects on DNA or chromosomes). Carcinogenicity studies are also not generally conducted for biotechnology-derived therapeutics unless there is a "cause for concern." In the early days of biotechnology development, there was greater concern for the technical issues of conducting these types of studies. These concerns included immunogenicity of human proteins in rodents and the technical challenge of parenteral administration in rats and mice over a 2 year time period. Over time some of these concerns have been diminished as immunogenicity can sometimes be addressed via dosing changes.

Early concerns for biotechnology products focused on "quality" issues [4] or process-related impurities. The concerns at that time were for carry-over of DNA or other cellular proteins, endotoxins, chemical contaminants, and viruses. Of course, these concerns still exist, but methods for purification and assays for evaluation of clearance have alleviated the need for the safety assessment scientist to focus on contaminants; current safety assessments focus on the pharmacologic activity of the molecules. This chapter therefore does not address the standard assays and methods (e.g., viral clearance) used to address purity issues. There are several guidances that address these issues, especially the FDA's Guidance for Industry on content and format of investigational new drug (IND) applications for Phase I studies of drugs, including well-characterized therapeutic, biotechnology-derived products [5] and the ICH's Q6B Specifications: *Test Procedures and Acceptance Criteria for Biotechnological/Biological Products* [6]. Other product-related impurities that do need to be considered by the safety assessment scientist include genetic variants, aggregate forms, chemical linkers, and differences in glycosylation patterns.

The assessment of safety in the preclinical development of biotechnology-derived therapeutics differs from that of traditional small molecules primarily in the species specificity and the potential for immunogenicity of the former. In addition, clinical indications, length of exposure, route of administration, and dosing intervals can be very different between biologics and traditional small-molecule pharmaceuticals. Determining safety margins can be challenging in a situation where "toxicity" is mainly due to exaggerated pharmacology. It is also not unusual to have trouble establishing "no effect" (NOEL) doses with these highly selective and potent molecules. The development plans for traditional pharmaceuticals are more routine and driven by well-established guidelines, while the development plans for biologics are

**TABLE 3 Differences Between Small-Molecule and Biotechnology-Derived Pharmaceuticals**

| Parameter | Small-Molecule Pharmaceuticals | Biotechnology-Derived Pharmaceuticals |
|---|---|---|
| Size | <500 daltons | >1 kD (macromolecules) |
| Immunogenicity | Nonimmunogenic | Potential for immunogenicity |
| Metabolism | Metabolized | Degraded |
| Frequency of dosing | Daily | Variable |
| Toxicity | Often structurally based | Exaggerated pharmacology |
| Half-life | Short | Long |
| Route of administration | Oral, some IV | Parenteral |
| Species specificity | Active in several species | Species specific |
| Synthesis | Organic chemistry (synthesized) | Genetic engineering (derived from living material) |
| Structure | Well defined | Often not fully characterized |

**TABLE 4 Typical Toxicology Studies for Small-Molecule and Protein Therapeutics**

| Small-Molecule Therapeutics | Protein Therapeutics |
|---|---|
| Screening studies | Range-finding studies |
| Range-finding studies | One month studies |
| Acute GLP studies | Three/six month studies |
| One month studies | Safety pharmacology |
| Safety pharmacology | Reproductive studies |
| Mutagenicity studies | Tissue cross-reactivity studies |
| Three month studies | Irritation/tolerance |
| Reproductive studies | Others as needed |
| Six month–-rat | |
| One year–-dog/monkey | |
| Industrial toxicology | |
| Diet RF studies | |
| Carcinogenicity studies | |
| *Total cost*: 5.0–6.5 million | *Total cost*: 2.5–3.0 million |
| *Linear time*: 4.5–5 years | *Linear time*: 2–2.5 years |

more driven by the pharmacology of the molecule under investigation (Tables 3 and 4).

This chapter focuses on the standard approaches to safety assessment and toxicity evaluation of biotechnology-derived therapeutics and current issues in the field.

## 2 SPECIES SPECIFICITY AND SELECTION

Perhaps the greatest challenge in the preclinical development of biotechnology-derived therapeutics is species specificity. As therapeutics have become more specific over the years, this has become an even greater challenge. Unlike the pharmaceutical development of traditional small molecules, one cannot assume that a biotechnology-derived therapeutic will be active in a standard species used for

toxicity testing. Rats and dogs are often selected for the assessment of safety of small molecules without regard to pharmacologic activity, the focus being on off-target toxicity due to structural effects. Species selection for small-molecule toxicity studies is often based on ADME considerations, particularly metabolite formation (i.e., which species has a metabolite profile most like humans). Only in recent years, with companies developing both traditional pharmaceuticals and biologics, has more emphasis been placed on testing for pharmacologic activity in species for standard toxicity assessments. This has always been an issue for biologics, for which lack of pharmacologic activity in a species leads to no effects in that species. In addition, the use of standard studies in standard species for biologics had little purpose and was of questionable use (except for potential effects of containments). The early work performed on interferons is a classic example; studies performed in nonrelevant species gave misleading information [7–9]. Several studies were conducted in rodents with recombinant human interferons. No evidence of toxicity was noted in many of these studies; these results later proved to be not predictive of what was to occur in humans. Activity in monkeys was more predictive, but there was a relative difference in the level of activity between primates and humans, indicating again how complex these assessments can be. Further work was performed on many of the interferons using homologous (surrogate) molecules.

## 2.1 Species Selection

There are several ways to determine whether a preclinical species is appropriate for safety assessment. If the therapeutic target is a replacement protein, the sequencing of the protein across species is a good place to start. Sequencing of related targets (e.g., ligands and receptors) can also be compared. If the therapeutic target is a receptor, receptor or ligand distribution between species can be investigated even before developing or testing the molecule. Often, *in vitro* studies are done first to look at cross-reactivity. Receptor binding can be a start, but binding itself is not sufficient. There are human or humanized molecules that demonstrate binding in primate cells but have no bioactivity. Receptor affinity and then occupancy or binding should be compared across species. Often there are log-fold differences between humans and nonhuman primates or lower species, especially in the immune system [10].

Assays for biologic activity should be available to test potency and specificity. Assays used for characterization in Good Laboratory Practice (GLP) test species should be validated. Assays can be cell culture-based or biochemical in nature. They may measure physiologic responses or enzymatic reaction rates. *In vitro* assays can be used to compare the potency and specificity across species and to predict what will happen in humans by correlating the responses *in vitro* with *in vivo* pharmacodynamic studies. There are instances when *in vitro* assay data does not correlate with *in vivo* data; this is sometimes the case when a protein is pegylated (i.e., is conjugated to polyethylene glycol (PEG) and activity is related to pharmacokinetics [11]. In these cases, the *in vitro* activity of the molecule is actually reduced due to steric hindrance from the large (PEG) moiety added to the protein. The *in vitro* results would lead one to believe that the potency of the linked molecule is less than the unlinked molecule. In actuality, because of the extended *in vivo* pharmacokinetic half-life of the linked molecule, the *in vivo* potency is greater than the original molecule.

Tissue cross-reactivity studies have traditionally been performed on monoclonal antibodies as described in the FDA guidance *Points to Consider in the Manufacture and Testing of Monoclonal Antibody Products for Human Use* [12]. These studies are usually conducted with tissues from three unrelated human donors. Comparison of the binding in tissues from different species can be helpful in determining the most relevant animal species for *in vivo* toxicology studies. These studies can help determine if an antibody binds to tissues in different species, and whether the distribution (which tissues show significant binding) is similar between human tissues and preclinical species. Unintended binding can also be evaluated and may give clues of what to look for in future *in vivo* studies (e.g., which target organs). These types of studies have several limitations, and better information can often be derived from studies examining mRNA across tissues and species. As stated in the FDA document [12], newer technologies should be employed as they become available and validated.

At one time, lack of binding in these studies led to the consideration of not performing any *in vivo* studies with the clinical molecule. If the target is not found in any appropriate animal species, then surrogates or other approaches may need to be considered. Lack of binding may also reflect a technical issue, and other methods to determine if the test species are relevant should be employed as well. Many targets in recent years are soluble receptors and traditional immunohistochemistry studies do not detect these receptors well (low receptor density, or the receptor is not easily accessible) [13]. Differences in target expression and binding in different species may be compensated for by altering doses, but these differences need to be identified before starting the safety assessment studies.

*In vivo* studies also have applicability in the selection of test species. If there are disease models in the potential safety assessment species and activity can be determined with the molecule intended for humans, safety assessment plans become more straightforward. The doses with activity can help determine margins of safety from the toxicology studies. Comparisons of *in vitro* data from human cells/tissues with data from the pharmacologically active species can greatly aid in the dose selection for early human studies. Animal models of disease often have provided useful information for safety assessment. When there are no disease models, pharmacologic activity can be measured using biomarkers or surrogate markers in normal animals. If this is not feasible, then receptor occupancy *in vitro* may be compared to *in vitro* activity. Other strategies may be needed if the molecule does not cross-react with an appropriate species. This is where exaggerated pharmacology and target liability (i.e., toxicity inherent to the receptor, ligand, enzyme, etc. being targeted) may need to be evaluated with a surrogate molecule, knock-out animals, or knock-in models (in which the human receptor/target is spliced into the genome of the test species) [13].

### 2.2 Nonhuman Primates

Many biologics are only active in primates; this is especially common with monoclonal antibodies. There are many challenges for programs that are limited to testing in nonhuman primates. These include the availability of nonhuman primates and the availability of laboratories and historical databases. This is particularly true for reproduction toxicology (see Section 4.2). Although more safety pharmacology

(see Section 4.4) parameters are now validated in non-human primates, there are some that are not testable in this model. This is particularly true for developmental CNS/behavioral measurements that are well validated in rodents. Other challenges are in the immunotoxicology arena, where intended or unintended effects on the immune system are measured using flow cytometry, cytokine measurements, and functional assays. These have been validated in rodents over the years, and only a few laboratories have been able to validate these assays in primates. The historical databases for T-cell-dependent antibody response (TDAR) and keyhole limpet hemocyanin (KLH) assays in nonhuman primates are still small.

Marmosets were gaining popularity a few years ago due to their small size, hence reducing the amount of compound needed for chronic studies. This size advantage can be outweighed by other factors, such as the small amount of blood available in these animals. In addition, marmosets are sensitive to stress [14]. Few laboratories have much experience with these animals and there is only a small historical database available.

Some molecules are even more specific and are only active in humans and chimpanzees (*Pan troglodytes*), severely limiting the types of testing that can be conducted preclinically. This situation will lead the safety assessment scientist to consider the use of surrogate molecules and/or the use of transgenic or knock-out animals. The use of chimpanzees for safety assessment is often impractical [15] because of the limitations of the testing laboratories. There are very few such laboratories available, and they are not resourced sufficiently to run multiple complicated studies concurrently. These laboratories also are not equipped to run toxicity studies at the same standards for GLPs as traditional toxicology laboratories. Chimpanzees are a protected and endangered species and cannot be bred. Few tests and bleedings can be conducted as each of these requires sedation of the animals. They are also very expensive studies to conduct, they are nonterminal (i.e., animals are not euthanized at the end of the study), and only a small number of animals can be separately housed at any time, making long-term monitoring difficult. Because chimpanzees often weigh more than 45 kg, they require large amounts of test material [16]. Many of the chimps available in the United States have been infected with HIV, which can complicate test results especially for compounds that have effects on the immune system. Special studies such as reproductive toxicology and carcinogenicity cannot be conducted in chimpanzees for numerous ethical, technical, and logistical reasons.

### 2.3 Surrogates and Other Alternatives

Surrogates or homologous proteins have their own issues. They can be difficult to develop in the desired species and by definition will be different from the molecules being developed for humans. They may have different affinity for receptors, different epitopes, different half-lives, and so on, all of which may change the dynamics and alter the pharmacologic and toxicologic effects of the molecules. They can delay the development of other programs because they will use the same manufacturing facilities. Surrogates provide little or no information on the final human product contaminants and degradation products. Overall, the decision whether to investigate a surrogate or homologous protein is the balance between the time, cost, and delay to other programs versus getting target liability information. In general,

any of these molecules need to undergo the same rigorous evaluation as the clinical agent [17].

Other methods of getting target liability information include literature reviews and knock-out animals. Surrogate antibodies can help elucidate the mechanism of action [18]. Other tactics include the use of homologous ligands or the use of transgenics expressing human receptors. This allows the use of human-specific molecules but may not address off-mechanism toxicities. Knock-out models also have been used to mimic the suppression of endogenous cytokines or proteins. Disease models have also been used for safety assessment [19], especially when compound availability is a concern.

All of these strategies have their challenges, advantages, and disadvantages. In evaluating all of these data, it is important to bear in mind the goals of preclinical safety of biologics: to determine the pharmacologic and toxicologic attributes and activities of the molecule to be tested in humans, to determine the target organs, and to assess the potential similarities and differences. This information will be used to predict appropriate doses (safe starting doses as well as maximal doses) for clinical testing and to determine what parameters may need to be measured in early clinical trials. In some cases, one species may be sufficient if there is only one species that has activity or if the biologic activity or the molecule and its target are well understood.

## 3 IMMUNOGENICITY

It is generally well accepted that immunogenicity is not well predicted across species. It has also been well documented that many biologic compounds intended for human use are immunogenic in animals [20]. ICH S6 states that "most biotechnology-derived pharmaceuticals intended for humans are immunogenic in animals." Traditional antigenicity studies or guinea pig anaphylaxis studies are not useful for predicting immunogenicity in humans and are now generally recognized (ICH S6) as not being appropriate studies for biologics. When these studies were conducted years ago, at the request of regulators, they were generally positive and led to adverse effects in animals. Since these studies have no predictive value, they are no longer considered appropriate and may be an unnecessary use of animals.

### 3.1 Types of Antibodies to Biotechnology-Derived Pharmaceuticals

All preclinical (and clinical) studies with biologics should include measurements of total incidence of antibodies and whether these antibodies are neutralizing. Assays to determine antibodies have become more sophisticated over the years, and the newest technologies allow detection at lower levels than were achievable with traditional enzyme-linked immunoassays (ELISAs) [21]. The technology exists today to measure and characterize the production of antibodies and to evaluate their effects on pharmacokinetics and pharmacodynamics. Clinically relevant antibodies are those that clear, sustain, neutralize, or cross-react with an endogenous protein. It is important to screen for the presence and development of antibodies to the test molecule throughout development. Since the consequence of these antibodies may range from no clinically significant effects to serious safety effects, assays need to

be developed to determine if the antibodies that appear are able to block the biologic activity of the test compound. This may occur either by direct binding to an epitope with activity or by blocking the active site through steric hindrance after binding to a site in close proximity [21].

A "clearing antibody" is one that results in more rapid clearance of test molecule: that is, the binding of the antibody to the test molecule causes the test molecule to be removed more quickly from the circulation. This type of antibody can be detected by comparing the pharmacokinetic profiles of the test molecule in animals with and without antibodies, or by comparing the pharmacokinetic profile of the first dose (at which time no antibodies are expected) with a subsequent dose (see Fig. 1). A clearing antibody may or may not have a direct neutralizing effect, but the consequence of the rapid clearance may eradicate any pharmacologic activity of the test compound. This type of antibody can make the toxicology studies invalid; if there is not adequate exposure to the test molecule, then the lack of toxicity does not give assurance of lack of effect from the pharmacologic activity of the molecule itself.

A "sustaining antibody" is one that delays the clearance of the test molecule: that is, the binding of the antibody to the test molecule keeps the test molecule in the circulation longer. This type of antibody also can be determined by measuring the serum or blood concentrations in animals with and without antibodies (see Fig. 2). The binding of a sustaining antibody to the test compound may or may not be neutralizing. When there are sustaining antibodies, the recovery period of the toxicology studies may need to be extended to see true reversibility of effects. These types of antibodies can lead to even greater exaggeration of pharmacology, and dose levels may need to be altered to account for unanticipated accumulation.

Both clearing and sustaining antibodies may also be "neutralizing antibodies": that is, the binding of the antibody to the test molecule changes the ability of the test molecule to produce its pharmacologic activity. The most worrisome type of antibodies are those that cross-react with an endogenous protein (see Fig. 3). In this case, the individual's own endogenous proteins can no longer function pharmacologically because the antibody response is binding and neutralizing its

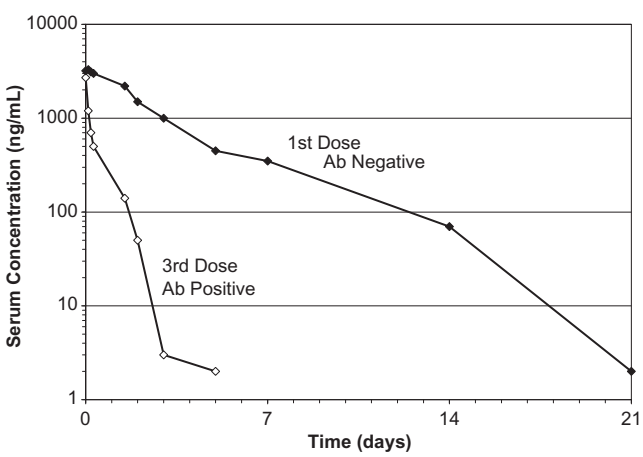


**FIGURE 1** Clearing antibody.

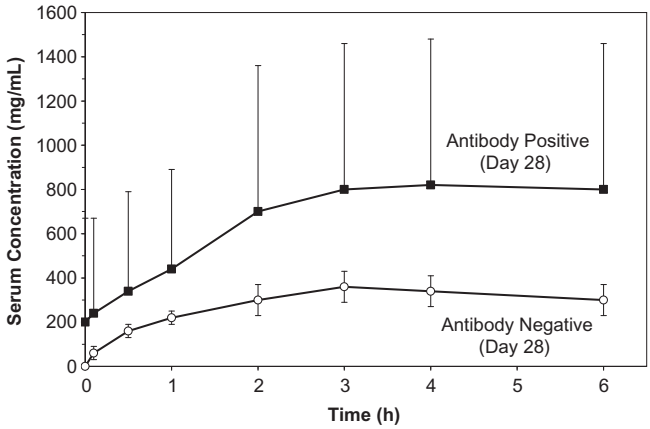


**FIGURE 2** Sustaining antibody.

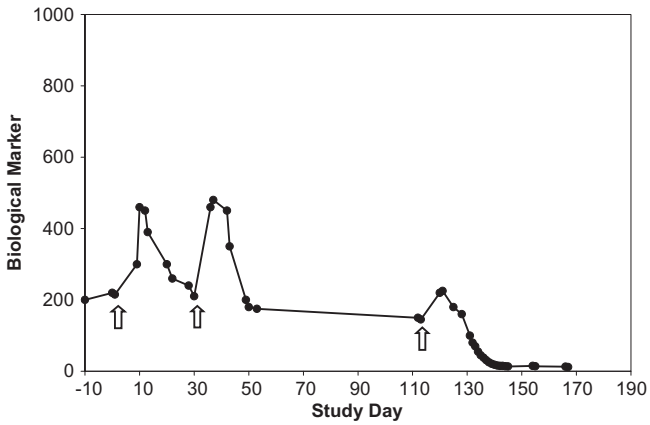


**FIGURE 3** Neutralizing antibody.

activity as well. These types of antibodies can have devastating effects in the clinic [22–24].

There are ways to alter the incidence and robustness of antibodies in preclinical studies. Altering the dosing regimen and route is the most obvious manner. The subcutaneous and intramuscular routes are the most immunogenic with the intravenous route being the least immunogenic [25]. This may be due to aggregation at the injection sites. The frequency and duration of dosing also can affect immunogenicity. The longer the dosing period, the greater the probability of an immunogenic response and the greater the response may be. The dosing frequency may also have an effect, but there is a balance between more immunogenicity due to greater exposure versus a frequency that mimics a pulse effect [23].

One of the most controversial topics in this area is the ability or lack of ability of animal immunogenicity results to predict immunogenicity in humans. Generally, it has been accepted that an antibody response in animals cannot directly predict such a response in humans. But recent studies have shown that animal studies may

be able to predict relative immunogenicity of different compounds (e.g., modified versions of a protein) and the potential subsequent effects of antibody production [20, 26]. There is also evidence that glycosylated proteins are less immunogenic than nonglycosylated proteins [22, 27]. Other effects that need to be assessed include those due to antibody–drug complexes. These complexes can deposit in tissues and organs [28] and cause glomerulonephritis [29]. Infusion reactions and anaphylactic responses also need to be monitored [24].

In summary, immunogenicity is a substantial complication for preclinical safety assessment studies. It can invalidate the model species, but antibody production alone should not stop the conduct of these studies. The antibodies need to be measured with reproducible assays [30]. Longer recovery periods are often needed for these studies to wash out the drug from the test animals. This can greatly increase the ability of the assays to detect antibodies to the drug. If there are active drug levels in the blood samples, they can interfere with the ability to measure true antibody levels [21]. The effects on pharmacokinetics and pharmacodynamics need to be measured and evaluated. The potential consequences of the antibodies on endogenous molecules need to be evaluated as well. The secondary effects, such as antibody deposition, should be measured. The lack of ability to predict human immunogenicity and the relative potential to predict this should be considered on a case-by-case basis.

### 3.2 Study Design Issues

Unlike most traditional small molecules, biologics are not given once or twice a day, orally, in a pill or capsule. They are almost all dosed via a parenteral route, are not given (or "taken") daily, and some are given in hospital settings or in clinician offices (e.g., monoclonal antibodies for the treatment of solid tumors such as Avastin® and Erbitux®). Several are now self-administered on a weekly or even less frequent basis (e.g., TNF inhibitors such as Enbrel® and Remicade®). The dose regimen for the safety assessment studies should reflect the dosing regimen plan for the human studies. The dosing interval in animals may be different from that planned for humans, depending on the half-life of the molecule in animals versus humans. Also, the development of antibodies in animals may alter the pharmacokinetics, and modifying the dosing interval may reduce the incidence of antibodies.

Other design issues include dose level selection. Often when the dose-limiting toxicity is related to the pharmacology (often referred to as "exaggerated pharmacology") it is difficult to establish a large margin of safety. The difference between the intended level of effect and an effect that leads to toxicity, even if it is the same effect, may be very small.

Other factors that need to be considered in dose selection are those that fall under the category of pharmacokinetics and drug metabolism. These factors will not be considered in any great depth in this chapter, but a few points to consider are listed here:

- Route of dosing (IV, SQ, infusion)
- Half-life of the molecule (e.g., IgG isotypes)
- PK/PD modeling (very useful for growth factors such as Epogen® and Neupogen®)

- Dosing regimen in the clinic (weekly, monthly, once per cycle)
- Length of intended clinical use (acute vs. chronic)

One study type generally not deemed relevant for biologics is metabolism studies. Mass balance studies have not proved useful when performed with proteins. Radiolabeled studies have limited usefulness as these labeled molecules undergo rapid metabolism and the label is often unstable. Protein biologics undergo proteolytic degradation into smaller peptides and then into individual amino acids. Distribution is viewed more from the perspective of target distribution, except for gene and cell therapies or vectors.

Toxicokinetics must be assessed in these studies, as with traditional pharmaceuticals. These data are necessary to prove exposure (which may differ with route) and to monitor the potential effects of antibodies on the exposure levels over time. One of the greatest differences between small molecules and biologics is the potential for immunogenicity, which can greatly affect pharmacokinetics, as previously discussed.

Recovery periods are very important in studies with biologics for several reasons. As already discussed, it is important to have a wash-out period of the drug to properly test for antibody levels. This works best with samples that are clear of the test compound itself (assay interference).

## 4 SPECIAL STUDIES

### 4.1 Genotoxicity

It is generally accepted now that these studies are not applicable for biotechnology-derived products, unless there is a "chemical" linker or toxic conjugate to the molecule. This is clearly stated in ICH S6 as well as the FDA document for monoclonal antibodies [12].

### 4.2 Reproductive Toxicology

When a biologic cross-reacts with traditional reproductive toxicology species, then these studies should be conducted as appropriate for the intended clinical population. The main consideration for these studies, if conducted in rodents and rabbits, is the dosing regimen and immunogenicity. The regimen should be carefully chosen to ensure pharmacokinetic coverage during the pivotal parts of gestation. The frequency of dosing in these studies might need to be closer to ensure that there is adequate exposure. Also, this may address the potential immunogenicity as discussed earlier. The exposure period for the Segment II studies in rodents and rabbits is usually short enough to avoid a strong immunogenic response, but samples should be taken to measure drug levels and antibodies in the dose range-finding studies.

The more complicated situation is with those molecules that only work in nonhuman primates. Here a decision must be made as to whether to conduct these studies in primates or to create a rodent surrogate. The decision should be based on good science as well as the practical and technical issues. There is no right or wrong answer

to this question, but there are several factors that need to be considered. There are differences between non-human primate subspecies to consider as well. For example, rhesus monkeys are seasonal breeders and cynomolgus monkeys are not. Cynomolgus monkeys also have a similar duration of menstrual cycle and placental morphology and physiology similar to humans. Human IgGs are known to cross the placental barrier, and their response to teratogens has been shown to be similar to humans [31]. On the other hand, compared with rodents, nonhuman primates have a small number of offspring (usually just one offspring per mother), they are very expensive, and their gestation period is 150 days long. They have a low conception rate and a high spontaneous abortion rate. In addition, there are only a small number of laboratories that have the capabilities of conducting primate reproductive studies. The historical database is not that large.

As for specific studies, male fertility is easier to test in cynomolgus monkeys [32], and these assays (e.g., hormones, sperm count) can often be incorporated into standard toxicology studies [33]. Mating studies are much more difficult, but fertility parameters can be measured in females as well (e.g., menstrual cycle, FSH, LH). Segment II studies are usually dosed during organogenesis (days 20 to 50), and the mothers are C-sectioned on day 100 [34]. Placental transfer can also be measured with continuous dosing.

### 4.3 Carcinogenicity

Unless there is a specific "cause for concern," these studies are generally considered inappropriate for biologics (ICH S6). This issue is also addressed in ICH S1A [35], which states that these studies are not generally needed for molecules that are developed as replacement therapy (e.g., insulin and certain hormones). This guidance also recognizes that neutralizing antibodies may "invalidate" the results. As with the S6 document, ICH S1A suggests that these studies be considered depending on the clinical indication, patient population, and treatment duration. Further reasons to consider these studies might include modifications made to the molecule or indications of potential effects based on subchronic studies (e.g., unexpected hyperplasia).

For molecules that are only pharmacologically active in non-human primates, there are additional technical issues. Carcinogenicity studies would not be feasible in non-human primates due to the long life span, nor are they practical due to the large number of animals that would be needed for statistics. The alternative would be to develop a surrogate that would be active in rodents. This would carry the same issues as already discussed. In addition, the longer time span needed for testing would only increase the chance for neutralizing antibody production. Alternative approaches, such as *in vitro* proliferation studies or receptor binding studies, may be informative but still not definitive. Another complication is with molecules that are immunosuppressive, as these might increase the incidence of any background tumors.

### 4.4 Safety Pharmacology

Often, parameters assessed in safety pharmacology studies can be incorporated into standard toxicology studies, including single-dose studies. For molecules that cross-react in traditional species (rodents and dogs), study designs can be more flexible,

while study designs for molecules that are active only in nonhuman primates have more challenges. Non-human primates can be telemeterized for cardiovascular assessments if there are concerns and even many CNS evaluations can be performed in non-human primates as well now.

### 4.5 Local Tolerance

Local tolerance assessments can usually be conducted as part of subchronic or chronic toxicity studies, by taking sections from injection sites. They are often used to test formulation changes as well.

## 5 FUTURE TRENDS

In the future, biotechnology products are going to become more complex. We will see more conjugated monoclonal antibodies, oligonucleotides, fusion molecules, and gene or plural: cell therapies. The need to address the basic questions with these products will remain the same, but the challenges will be greater. There are new tools becoming available to the toxicologist to address these concerns. Global development strategies for development of biotechnology products will be more science driven, more problem focused, and with additional regulatory challenges. The overall goal should still be to evaluate preclinically the potential effects of these molecules in humans in the diseases of interest. The safety assessment scientist should develop a plan with a strong scientific rationale, using all of the necessary tools and good justification for the animal model chosen.

## REFERENCES

1. Pharma Org. www.pharma.org.
2. Cavagnaro JA. Preclinical safety evaluation of biotechnology-derived pharmaceuticals. *Nat Rev* 2002;1:469.
3. Griffiths SA, Lumley CE. Non-clinical safety studies for biotechnologically-derived pharmaceuticals: conclusions from an International Workshop. *Hum Exp Toxicol* 1998;17:63.
4. Dayan AD. Safety evaluation of biological and biotechnology-derived medicines. *Toxicology* 1995;105:59.
5. FDA IND Guidelines. www.fda.gov.
6. ICH Q6B Specifications: *Test Procedures and Acceptance Criteria for Biotechnological/Biological Products* 1999.
7. Green JD, Terrell TG. Utilization of homologous proteins to evaluate the safety of recombinant human proteins—case study: recombinant human interferon gamma (rhIFN-γ). *Toxicol Lett* 1992;64/65:321.
8. Serabian MA, Pilaro AM. Safety assessment of biotechnology-derived pharmaceuticals: ICH and beyond. *Toxicol Pathol* 1999;27:2.
9. Ryan AM, Terrell TG. Biotechnology and its products. In Haschek W, Ed. *Handbook of Toxicologic Pathology*, 2nd ed. London: Academic Press; 2002, p 479.
10. Hart TK, et al. Preclinical efficacy and safety of pascolizumab (SB 240683): a humanized anti-interleukin-4 antibody with therapeutic potential in asthma. *Clin Exp Immunol* 2002;130:93.

11. Elliott S, Lorenzini T, Asher S, et al. Enhancement of therapeutic protein in vivo activities through glycoengineering. *Nat Biotechnol* March 2003;21:414–421.
12. *Points to Consider in the Manufacture and Testing of Monoclonal Antibody Products for Human Use*. Washington, DC: FDA, CBER; 1997.
13. Bolon B, et al. Genetic engineering and molecular technology. In Krinke G, Ed. *The Laboratory Rat*. London: Academic Press; 2000, p 603.
14. Zuhlke U, et al. The common marmoset (*Callithrix jacchus*) as a model in biotechnology. In Korte R, Vogel F, Weinbauer GF, Eds. *Primate Models in Pharmaceutical Drug Development*. Munster, Germany: Waxmann; 2002, p 119.
15. Klingbeil C, Hsu D. Pharmacology and safety assessment of humanized monoclonal antibodies for therapeutic use. *Toxicol Pathol* 1999;27:1.
16. Tony Ndifor Presentation at CRL Biotechnology Symposium 2003.
17. Clarke J, et al. Evaluation of a surrogate antibody for preclinical safety testing of an anti-CD11a monoclonal antibody. *Regul Toxicol Pharmacol* 2004;40:219.
18. Treacy G. Using an analogous monoclonal antibody to evaluate the reproductive and chronic toxicity potential for a humanized anti-TNFa monoclonal antibody. *Hum Exp Toxicol* 2000;19:226.
19. Sato K, et al. Efficacy of intracoronary versus intravenous FGF-2 in a pig model of chronic myocardial ischemia. *Ann Thorac Surg* 2000;70:2113–2118.
20. Bugelski PJ, Treacy G. Predictive power of preclinical studies in animals for the immunogenicity of recombinant therapeutic proteins in humans. *Curr Opin Mol Ther* 2004;6:10.
21. Gupta S, et al. Recommendations for the design, optimization, and qualification of cell-based assays used for the detection of neutralizing antibody responses elicited to biological therapeutics. *J Immunol Methods* 2006;321:1–18.
22. Koren E, Zuckerman LA, Mire-Sluis AR. Immune responses to therapeutic proteins in humans—clinical significance, assessment and prediction. *Curr Pharm Biotechnol* 2002;3:349.
23. Schellekens H. Immunogenicity of therapeutic proteins: clinical implications and future prospects. *Clin Ther* 2002;24:1720.
24. Rosenburg AS. Immunogenicity of biological therapeutics: a hierarchy of concerns. In Brown F, Mire-Sluis A, Eds. *Immunogenicity of Therapeutic Biological Products*. Basel: Karger AG; 2003, p 15.
25. Working PK. Potential effects of antibody induction by protein drugs. In Ferraiolo BL, Mohler MA, Gloff CA, Eds. *Protein Pharmacokinetics and Metabolism*. New York: Plenum Press; 1992, p 73.
26. Wierda D, Smith HW, Zwickl CM. Immunogenicity of biopharmaceuticals in laboratory animals. *Toxicology* 2001;158:71.
27. Gribben JG, et al. Development of antibodies to unprotected glycosylation sites on recombinant human GM-CSF. *Lancet* 1990;335:434.
28. Henck JW, et al. Reproductive toxicity testing of therapeutic biotechnology agents. *Teratology* 1996;53:185.
29. Terrell TG, Green JD. Comparative pathology of recombinant murine interferon-γ in mice and recombinant human interferon-γ in cynomolgus monkeys. In Richter GW, Solez K, Ryffel B, Eds. *International Review of Experimental Pathology: Cytokine-Induced Pathology Part A: Interleukins and Hematopoietic Growth Factors*. San Diego: Academic Press; 1993, p 73.

30. Mire-Sluis AR. Progress in the use of biological assays during the development of biotechnology products. *Pharm Res* 2001;18:1239.
31. Weinbauer GF. The nonhuman primate as a model in developmental and reproductive toxicology. In Korte R, Weinbauer GF, Eds. *Primate Models in Pharmaceutical Drug Development*. Munster, Germany: Waxmann; 2002, p 49.
32. Vogel F. How to design male fertility investigations in the cynomolgus monkey. In Korte R, Weinbauer GF, Eds. *Towards New Horizons in Primate Toxicology*. Munster, Germany: Waxmann; 2000, p 43.
33. Weinbauer GF, Cooper TG. Assessment of male fertility impairment in the macaque model. In Korte R, Weinbauer GF, Eds. *Towards New Horizons in Primate Toxicology*. Munster, Germany: Waxmann; 2000, p 13.
34. Weinbauer GF. The nonhuman primate as a model in developmental and reproductive toxicology. In Korte R, Weinbauer GF, Eds. *Primate Models in Pharmaceutical Drug Development*. Munster, Germany: Waxmann; 2002, p 49.
35. ICH S1A: *The Need for Long-Term Rodent Carcinogenicity Studies of Pharmaceuticals* 1995.

# PRECLINICAL DEVELOPMENT OF PROTEIN PHARMACEUTICALS: AN OVERVIEW

Dipankar Das and Mavanur R. Suresh
*University of Alberta, Edmonton, Alberta, Canada*

**Contents**

1 Introduction
2 Macromolecular Drugs
3 Genes, Directed Evolution, and Gene Shuffling
4 Molecular and Biochemical Characterization of Protein Drugs
5 Influence of Carbohydrate in Protein Drugs
6 Antibodies
7 Hybridoma-Derived Monoclonals
  7.1 Immunization
  7.2 Isolation of B-Cells from the Spleen
  7.3 Fusion and Postfusion
  7.4 Cloning
  7.5 Cell Line Characterization and Cryopreservation
  7.6 Small Scale Bioreactor Culture
  7.7 Purification
  7.8 Applications
  7.9 Immunogenicity Issues and Humanization
8 Bispecific Monoclonal Antibodies
9 Recombinant Antibody
  9.1 Phage Display
  9.2 Yeast Display
  9.3 Ribosome and mRNA Display
10 High Level Mammalian Expression of Recombinant Antibodies and Proteins by Gene Amplification System
11 Special Considerations

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

12 Supramolecular therapeutics: Cells, Tissues, and Organs
Acknowledgments
References

## 1 INTRODUCTION

Protein pharmaceuticals can broadly be classified into three main categories—namely, recombinant proteins and enzymes, monoclonal antibodies (MAbs) and their derivatives, and synthetic peptides. The development of hybridoma technology for the generation of monospecific antibodies by Kohler and Milstein [1] in 1975 was a landmark technological achievement to harness unique specific antibodies. This technology and the development and refinement of recombinant DNA technology also in the early 1970s form the two pillars for the emergence of protein pharmaceuticals. Another important related technique is the solid phase peptide synthetic approach for production of pure shorter peptides. This synthetic strategy is often limited by the size, secondary modifications, and lack of glycosylation. However, it has the unique advantage of being able to incorporate synthetic analogues and derivatives that cannot be developed biologically. This chapter deals primarily with antibodies and recombinant protein pharmaceuticals. Readers are directed to several comprehensive reviews and books on solid phase synthetic strategies, a technique invented by Merrifield [2, 3]. The majority of currently approved pharmaceuticals and generics are small molecular weight (MW) compounds. The major differences between small MW pharmaceuticals and the larger proteins are summarized in Table 1. Biogeneric protein pharmaceuticals are a new category of therapeutics that are likely to be a new class in the next few years.

## 2 MACROMOLECULAR DRUGS

Two major technological advances ushered in the era of large molecular weight protein drugs. In the early 1970s advances in the development of recombinant DNA

**TABLE 1 Differences Between Traditional Drugs and Protein Pharmaceuticals**

| Small Molecular Weight Pharmaceuticals | Protein Pharmaceuticals |
|---|---|
| 1. Mostly synthetic | 1. Biology-based production |
| 2. Uniquely pure entities (with exception of enantiomers) | 2. Microheterogeneity in glycosylation and other posttranslation modifications |
| 3. Rarely form aggregates | 3. Protein aggregates are common |
| 4. Nonimmunogenic | 4. Immunogenic to different degrees |
| 5. Rarely an issue with adventitious agents | 5. Adventitious agents can be an issue |
| 6. Generally stable | 6. Susceptible to denaturation and physical and chemical instabilities |
| 7. Unique metabolic pathways for each class of drugs | 7. Universal protease-based cleavage, metabolism, and reabsorption |
| 8. Generally lack targeting properties | 8. Often with targeting capabilities |

technology allowed the expression and purification of mammalian proteins in prokaryotic systems. *Escherichia coli* has been used extensively to generate recombinant proteins for a variety of biopharmaceutical productions. Expression in this simple platform, however, has two major drawbacks. First, this prokaryote can only express nonglycosylated proteins; and second, the bulk of the recombinant protein is obtained as denatured aggregate of inclusion bodies. Structural and functional information is either lost or distorted as a result and relatively laborious extraction and refolding strategies are required to renature the protein from the inclusion bodies [4]. Recent advances have resulted in the development of secreted recombinant proteins from *E. coli* with substantial native activity [5]. The limitations of the prokaryotic expression platforms have led to the development of yeast, insect cell, and mammalian expression platforms [6–11]. A number of biopharmaceuticals are produced in yeast including human insulin and hepatitis A vaccine, providing therapeutic proteins to millions of people around the world. The advantages of the yeast platform include the high level production of the recombinant protein and the wealth of large scale fermentation expertise in the brewery industry, which was the natural starting point for bulk production of biopharmaceuticals. However, yeast and insect cells impart a different glycosylation footprint on the expressed protein. For example, the recombinant proteins expressed in yeast are of the high mannose glycoprotein product. In a pharmaceutical context, the pharmacokinetics of aglycosyl and glycosylated proteins are significantly different, with the latter being cleared rapidly by the liver asialoglycoprotein receptor [12].

The most ideal platform for expression of biopharmaceuticals is the mammalian cell. The CHO (Chinese hamster ovary) cells have been one of the platforms exploited for protein expression. This platform provides a biopharmaceutical product that closely mimics the natural protein with minimal differences compared to the other prokaryotic, yeast, and insect platforms. Production of biopharmaceuticals in mammalian platforms is expensive and generally provides comparatively lower yield of product. However, unique amplification strategies using the dihydrofolate reductase (DHFR) or the glutamine synthetase (GS) systems have provided yields that are significantly higher for commercial viability.

Almost all of the current approved biopharmaceuticals are therapeutic proteins and vaccines. Several other proteins are used in nonpharmaceutical applications, such as in industrial uses, including detergents, leather tanning, meat and cheese processing, beverage production, and the baking industry. Currently, there are several hundred protein pharmaceuticals in various stages of clinical trials. The excellent safety and therapeutic properties have resulted in choices made by patients and health care professionals to prefer biopharmaceuticals over small molecular weight drugs. A case in point in therapeutic options in our province (Alberta, Canada) is the use of Herceptin, a monoclonal antibody targeting the Her-2neu receptor positive aggressive breast cancer. A few years ago the provincial body approved reimbursement or use of the MAb only as a third line therapy after anthracycline treatment of recurrent metastatic breast cancer. This year all women in the Province of Alberta can access MAb treatment for Her-2neu receptor positive breast cancer as the primary first line therapy. While proteins dominate biopharmaceuticals as the largest category, several other promising large MW drugs are in clinical trials or in early preclinical development, including DNA vaccines, antisense oligonucleotides, RNAi, polysaccharides, cells, and organs as well as novel delivery

**TABLE 2 Classification of Biopharmaceuticals**

| Category | Example |
|---|---|
| 1. Monoclonal antibody | 1. Rituxan, Synagis, Simulect (18 approved by FDA) |
| 2. Polyclonal antibody | 2. Respigam, Thymoglobulin |
| 3. Antisense oligonucleotide | 3. Vitravene |
| 4. Hormones | 4. Insulin |
| 5. Enzymes | 5. Asparaginase, Raspuricase |
| 6. Interferons | 6. Interferons α, β, γ |
| 7. Interleukin | 7. Interleukin 2 |
| 8. Tumor necrosis factor | 8. TNF-α |
| 9. Vaccines | 9. Hepatitis B surface antigen |
| 10. Hematopoietic growth factor | 10. Erythropoeitin, colony stimulating factor |
| 11. Blood factors | 11. Factors VIII, IX |
| 12. Nonrecombinant proteins | 12. Sucraid, Wellferon |
| 13. Thrombolytic agents | 13. Tissue plasminogen activator |

systems such as nasal therapeutics and even skin patch vaccines. Table 2 has a summary of the categories of approved biopharmaceutical drugs.

## 3 GENES, DIRECTED EVOLUTION, AND GENE SHUFFLING

The advent of genomics and proteomics along with the complete sequence information of over several hundred organisms including *Homo sapiens* has provided new insights and opportunities for the future of biopharmaceuticals. The 2000 version of the human genome sequence had estimated that we harbored ~40,000–50,000 genes. A more recent version [13] of the human genome has narrowed this to about 25,000–30,000 genes. Classical recombinant strategies of cloning and expression of proteins have been refined with the emergence of new gene modifications for overexpression of pharmaceuticals as well as redesigning the natural proteins for new and improved properties. At least four innovative strategies can be recognized in the context of optimization of biopharmaceutical production. They are codon optimization, directed evolution, gene shuffling, and enhanced expression platforms.

Codon optimization is a strategy that has been introduced to exploit two technological developments. The first relates to the refinement of automated synthesis and sequencing of genes, including large DNA molecules in a relatively short time. We recently obtained the large synthetic spike gene of SARS-coronavirus consisting of ~4000 nucleotides in a few weeks from a commercial outfit (www.Geneart.com). This is a significant technological resource that has evolved from the very first aminoacyl t-RNA synthetase gene constructed manually by a group of several scientists over a period of one year [14]. Similarly, on the sequencing end we are now routinely able to sequence genes with the help of automated gene sequencers within days, a far cry from the first gene sequence of a phage gene in the 1970s [15, 16]. In the recent SARS virus outbreak in several countries, the combined efforts of two laboratories in Canada resulted in the complete genome sequence of the whole virus within a week [17].

The second innovation toward codon optimizing genes is the detailed understanding of the patterns of t-RNA composition in the most common host expression systems. Many a gene expression fails to generate the desired product or barely produces the recombinant protein. The causes of such diminished production are several and key among them is the paucity of certain species of t-RNA in the cell used for expression. As a result, the cellular translation machinery is less efficient in production of the recombinant biopharmaceutical. Hence, modifying the specific codons in the synthetic gene to match the abundant species of t-RNA permits efficient protein synthesis. In a recent study, we have successfully utilized the codon optimized nucleoprotein gene of SARS-CoV for 10–20-fold overexpression at ~30–55 mg/L of *E. coli* culture (Fig. 1) [18].

Another major gene improvement strategy is rational design of neobiopharmaceuticals, but it requires in-depth understanding of structure, function, and mechanisms. The technique of site-directed mutagenesis is an elegant method of specific sequence modification of biopharmaceuticals for improved characteristics, including function, biodistribution, and immunogenicity. Random site-directed mutagenesis of proteins has evolved into directed evolution and gene shuffling to make improved biopharmaceuticals. Unlike rational design, which requires a thorough understanding of the protein, the directed evolution approach requires only minimal knowledge of its structure and function [19]. The strategy in directed evolution is based on either random mutagenesis of the gene or recombination of homologous gene segments followed by specific screening based on desired functional characteristics.

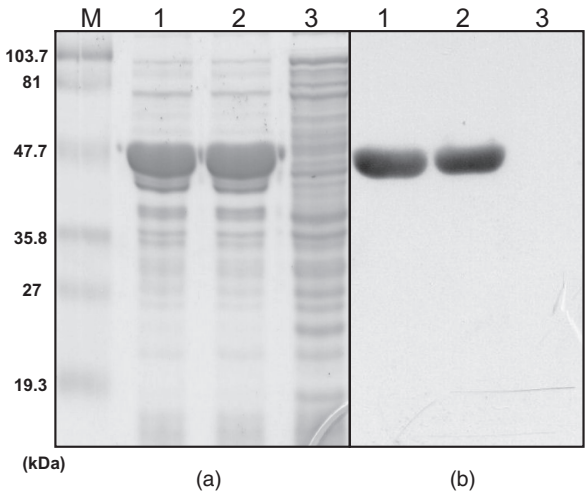


**FIGURE 1** Analysis of codon optimized SARS-CoV NP gene expression in *E. coli*. Total cell protein analysis by SDS-PAGE and Western blot. The 10% polyacrylamide gel was stained with Coomassie Brilliant Blue (a). Proteins transferred to nitrocellulose membrane were subsequently probed with anti-$His_6$ MAb followed by goat anti-mouse-HRPO (b). Lane M: standard prestained molecular weight markers; lanes 1 and 2: induced culture; lane 3: uninduced culture. Note the overproduction of the protein and its detection by Western blot.

This is akin to Darwinian evolution, which requires generation of diversity by mutagenesis and selection of exploiting assays for improved protein characteristics. A particularly useful technique in the above approach is DNA shuffling. A whole range of gene fragments are generated following initial incorporation of point mutations into a gene. The new variant is generated by reassembly of the fragments and functional selection for desired property. Exploiting the above methods, many proteins and biopharmaceuticals with improved catalytic activity, enzyme stability, thermostability, enzyme substrate specificity, new substrate activity, and stereoselectivity have been developed. Altering the *in vivo* biological properties of a therapeutic protein is also an important objective, which includes pharmacokinetics and immunogenicity. Retavase is a modified form of tPA (tissue plasminogen activator), which is used during myocardial infarction to initiate the activation of plasminogen to promote fibrinolytic reactions. This modified protein has only two of the five structural domains of native tPA along with increased half-life in blood.

## 4 MOLECULAR AND BIOCHEMICAL CHARACTERIZATION OF PROTEIN DRUGS

The initial identification and development of a protein pharmaceutical starts from early seminal discoveries on a potential protein candidate exploiting classical methods or using genomics and proteomics. Biomarker identification from the latter methods is an exciting protein mining prospect today. Classical approaches rely on the identification of candidate proteins with specific properties such as hormones, interleukins, monoclonals, toxins, and other effectors. Early laboratory data from investigators is often the starting point for systematic characterization of a single candidate pharmaceutical. Table 3 enumerates the FDA requirements for characterizing a protein pharmaceutical. Genomic and proteomic chip or microarray-based

**TABLE 3 FDA Requirements for Physicochemical Characterization of Biologicals**

1. Amino acid analysis
2. Amino acid sequencing, entire sequence, or C- and N-terminal sequences
3. Peptide mapping
4. Determine disulfide linkage if any
5. SDS-PAGE (sodium dodecyl sulfate–polyacrylamide gel electrophoresis) reduced and nonreduced
6. Isoelectric focusing
7. Conventional and high pressure liquid chromatography (HPLC), for example, reverse phase, size exclusion, ion-exchange chromatography
8. Mass spectroscopy
9. Assays to detect product-related proteins including deamidated, oxidized, cleaved, and aggregated forms and other variants, for example, amino acid substitutions and adducts/derivatives
10. Assays to detect residual host proteins, DNA, and reagents
11. Immunochemical analyses
12. Assays to quantitate bioburden and endotoxin

*Source*: http://www.fda.gov/.

identification often differentiates several potential candidates. However, even among the few candidate markers, one would have to be selected based on the efficacy and the feasibility of pharmaceutical development. Classical approaches of marker identification will often provide some degree of characterization of the protein. *In vitro* bioassays for the protein or *in vivo* animal-based therapeutic endpoints are often developed to test the potency and efficiency. One of the first objectives in preclinical development and characterization requires the purification of the protein from bulk raw materials. In the last few decades the refinement of techniques in recombinant DNA cloning and expression of pure proteins has revolutionized protein pharmaceuticals including monoclonal antibodies. Enumerated below is a series of preclinical steps to be undertaken although the order of some of the steps may be altered. Table 4 lists the properties of proteins and methods to obtain the parameters.

1. *Gene Sequence.* Knowledge of the entire human genome provides the sequence of the gene for any human protein for expression in a heterologous system. Monoclonal antibodies useful for human therapeutics can also be analyzed to obtain the primary gene sequence of the variable H and L chains from hybridomas. Hence, one of the first steps is to decide the gene or gene segment to be cloned and the choice of the expression system. The gene of interest could be synthesized using codon optimization to maximize the yield.

2. *Expression Platform and Vectors.* An early choice must be made for the type of plasmid vector appropriate to the expression platform. The most common expression systems are based on *E. coli*, yeast, and mammalian CHO cells. The choices of plasmid vectors are numerous and require careful consideration for the type of drug resistant selection, promoters, and enhancers along with the genes of interest. An important practical issue in biopharmaceutical development is the practical yields of the protein in a given expression system. Two well developed approaches for

**TABLE 4 Some Key Analytical Methods for Preclinical Studies**

| |
|---|
| *Characterization of Proteins* |
| Structure, amino acid composition, prosthetic group, disulfide and free SH group, metals, confirmation, p*I*, aggregation, mass, two-dimensional fragmentation pattern, fluorescence, light scattering, sedimentation pattern, charge and electrophoretic properties, UV extinction coefficient, enzymatic activity, biological assays and potency, immunological properties and stability (denaturation, aggregation, oxidation), carbohydrate composition and sequence. |
| *Methods for Protein Analysis* |
| Spectroscopy (UV-Visible), fluorescence spectroscopy, affinity, avidity, size exclusion chromatography, circular dichroism (CD), high performance liquid chromatography, SDS-PAGE, IEF, Fourier transform—infrared, nuclear magnetic resonance (NMR), light scattering, ELISA, MS or LC/MS (Q-TOF), bioassays, stress tests–shake test, surfactant test, freeze–thaw test, heat stability. |

*Source*: Adapted from Ref. 20.

enhanced expression in mammalian systems are the DHFR- and GS-based amplification strategies. The preclinical development of cells expressing the desired protein could be optimized early by the above strategies. This can be further enhanced in process development for further enhancement of yield.

3. *Purification Strategies.* The small scale purification methodology is often the starting point for future scale-up of the protein. The myriad of protein purification strategies can broadly be classified into two categories—namely, general protein chromatography techniques that are applicable to all types of proteins and affinity chromatography methods that apply to specific proteins or to certain groups of proteins. General chromatography techniques include methods based on charge hydrophobicity or size of the protein. The addition of sophisticated instrumentation and protein monitoring methods along with purification supports provide the means to purify almost any protein in several steps. Affinity chromatography can be divided broadly into the group-specific and ligand-specific methods. Both have distinct advantages over the multistep traditional chromatography approaches in being able to substantially purify specific proteins in one step. Group-specific affinity chromatography utilizes affinity ligands that can bind and purify specific classes of proteins [21]. The protein G affinity support can purify all types of monoclonal IgG irrespective of the specificity of the antibody. This is due to the ability of the immobilized protein G to bind to the Fc part of all subclasses of IgG. The popular immobilized metal affinity chromatography (IMAC) columns can bind to metalloproteins and recombinant hexa histidine (His) tag incorporated peptides. Ligand-specific affinity columns are more specific and can purify a specific receptor protein, providing almost pure protein in one step.

4. *Preclinical Protein Characterization.* An in-depth description of the biochemical, biophysical, and molecular preclinical characterization of proteins is an essential part of the development of new biopharmaceuticals. This is an extensive systematic study of the future product prior to formulation and early Phase I and II clinical trials. The key data obtained from this preclinical study will also be the basis of future specifications of the key raw materials and the protein product derived from the early production process. The FDA website provides the detailed analytical testing to be done for a candidate prepharmaceutical (http://www.fda.gov/cber/gdlns/cmcdna.pdf). Table 3 lists the data required for the drug substance defined as the unformulated active ingredient. Based on this data, the finished drug product is developed with the addition of excipients.

A protein drug can be a simple polypeptide consisting of only amino acids without any higher order structure or with homo or hetero multimeric quarternary structure. More complex proteins may have prosthetic or coenzyme moieties as well as posttranslational modifications including glycosylation, sulfation, phosphorylation, and/or formylation. Conjugation of some of the categories of biopharmaceuticals with radioisotope, drugs, or toxins is also common. Thus, unique biologicals will require the design of special analyses as well as assays to confirm integrity of the pharmaceutical in later stages of development. In addition, data on the optimal storage conditions and data on aggregation, stability, and degradation are also required. This could include both accelerated and real-time stability that would impact on product development, shipping, and marketing strategies. Technological

advancements in protein identification and characterization have expanded the scope of preclinical analysis listed in Table 2. A detailed enumeration of each is beyond the scope of this chapter and is discussed extensively in a recent book chapter [20].

It is also important to recognize that the boundaries are not rigid between preclinical characterization of biopharmaceuticals and formulation studies. Thus, detailed understanding of the properties of the biopharmaceutical at an early stage is an advantage in the eventual formulation of the final product. Addition of excipients is common and can include human albumin, amino acids, carbohydrates, detergents, chelates, and other chemicals and preservatives.

## 5 INFLUENCE OF CARBOHYDRATE IN PROTEIN DRUGS

Many protein pharmaceuticals are glycoproteins and hence have special properties and pharmacokinetic profiles compared to nonglycosylated proteins. Glycosylated proteins are expressed in yeast, insect cells, or mammalian platforms. The carbohydrate chains are linked to specific amino acids and include N-linked asparagine and the O-linked carbohydrate chains linked to either serine or threonine. The presence of carbohydrates can have an impact on its biodistribution in the body depending on the route of administration and the lack of sialic acid in the side chain. The asialoglycoprotein receptor is a major sink for glycoproteins that lack sialic acid or have lost their sialic acid as a result of degradation or senescence. In some of our early studies using radioactively labeled proteins in mice and rabbits, we demonstrated that intra-venous (IV) injection of asialoglycoproteins results in them being taken up by the liver almost on the first pass, presumably by the hepatocytes [12]. It is estimated that ~40,000 copies of such receptors are on the surface of each hepatocyte and hence the liver has an enormous capacity to clear glycoproteins. Modulation of protein drugs for enhanced delivery to target cells and organs can also be achieved by altering the carbohydrate modeling composition of the glycan chains. This is elegantly shown in the case of enhanced targeting of β-glucocerebrosidase to the macrophages of the reticuloendothelial system to treat Gaucher disease, which is characterized by the accumulation of glucosylceramide. The therapeutic recombinant enzyme (Cerezyme) incorporates the carbohydrate modeling process to generate mannose terminated oligosaccharides to enhance the uptake of the enzyme by the macrophages of the reticuloendothelial system in which Gaucher disease is manifested.

## 6 ANTIBODIES

Antibodies are a group of glycoproteins produced by B-lymphocytes in response to foreign antigens. Natural or deliberate immunization of foreign antigens results in specific antibody production recognizing several epitopes of the antigens. Immunogenicity of the foreign antigen dictates the magnitude of the immune responses in

an animal. Antibodies produced in this way are referred to as polyclonal antibodies. Antitoxin or venom antidotes are in this category, which is available commercially. They exhibit heterogeneity in binding avidity, specificity, immunoglobulin class, and isotype. Repeated immunizations or natural exposure to the antigens results in an improved high affinity immunoglobulin response as a result of maturation of the B-cell immune responses. A single B-cell isolated from the polyclonal humoral response and immortalized results in a monoclonal colony secreting identical pure monoclonal antibodies. Each B-cell has the capacity to produce an antibody that recognizes a single epitope. Conferring immortality to a B-cell is accomplished by fusion to a myeloma cell (immortal B-cell cancer), which results in a hybridoma line combining the key properties of the two parental cells of unlimited cell division and secreting a single antibody species.

In the last 30 years since this method was introduced, monoclonal antibodies have become invaluable tools in diagnostic immunoassays and therapeutics of various human diseases [22–26]. Most immunodiagnostic kits now use monoclonal antibodies instead of the polyclonal antibodies. Recombinant DNA technology has also contributed to the development of a new generation of antibody pharmaceuticals for treatment of human diseases and various therapeutic proteins. Most protein pharmaceuticals today are produced recombinantly using *E. coli*, yeast, insect, or mammalian cells. They are safe, convenient, and less expensive to produce, and an abundant supply can be produced within a short period of time. Antibodies and their derivative fragments have also been used as tools in a variety of applications, such as fundamental research work, as well as in diagnosis and human therapeutics. Antibody pharmaceuticals constitute at least 25% of the proteins in different phases of clinical trials [27, 28]. Utilization of antibodies as drug delivery vehicles or as triggers for human immune responses in cancer therapy is clearly promising in future applications [29].

**Steps in Hybridoma Production**

1. Immunization of mice with antigens.
2. Check of antibody titer in mouse blood.
3. Spleen removal and preparation of a single cell suspension.
4. Myeloma cell (SP2/0) preparation.
5. Fusion of spleen cells and myeloma cells with polyethylene glycol (PEG).
6. Hybridoma selection in selective (HAT) medium.
7. Observation and maintenance of different hybridoma clones (7–15 days postfusion).
8. Screen for specific antibody secretion by ELISA.
9. Expansion of positive clones.
10. Recloning of all positive hybridoma clones by limiting dilution technique.
11. To ensure monoclonality, recloning to >90% positive clones and storage of ~10 vials in liquid $N_2$.
12. Isotype determination.
13. Small scale cell culture and purification on protein G.
14. Characterization.

## 7 HYBRIDOMA-DERIVED MONOCLONALS

### 7.1 Immunization

Immunization of the animals with the desired antigen is the first step in monoclonal antibody development. Conventionally, this is done by the immunization of BALB/c mice with the antigen followed by repeated booster injections until a suitable serum titer has been obtained. Typically, 6–8 week old female mice are initially immunized and the fusion is performed within 6 months. The planned immunization schedule is very important for the development of the antibody. Collecting a small amount of the serum prior to immunization and storage at –20 °C is standard practice. This should be used as a negative control for screening assays. At least three to four injections with a maximum of 100–200 μL of volume per mouse are administered, with 1–2 week intervals between the injections. Approximately 20–50 μg of antigen, emulsified with complete Freund's adjuvant (CFA) for primary immunization, should be used for subcutaneous (SC) or intraperitoneal (IP) administration. Subsequent IP injections of 10–25 μg antigen should be emulsified with incomplete Freund's adjuvant (IFA). A final boost 2–3 days prior to fusion should be administered with the antigen in phosphate buffer saline (PBS). An intravenous or intrasplenic injection may be given to maximize the amount of immunogen reaching the spleen and activating B-cells [30–34]. Adequate immunization time is required for affinity maturation and isotype class switching to occur. The first exposure of a naive immune system to a specific immunogen will primarily result in production of IgM antibodies. Subsequent repeated immunizations induce the B-cells to undergo a class switch of the immunoglobulin isotype. The interval between the immunizations can also greatly affect the affinity of antibody produced. As the interval is increased, the antigen level decreases and only higher affinity B-cells remain ready to respond to a subsequent boost. Therefore, the immune response is mostly directed against the more immunogenic antigen or domain. The use of insufficient or excessive amounts of antigen can lead to immune tolerance. Excessive immunization can also block lymphocyte antigen receptors and produce immune paralysis. Overimmunization has also been implicated in the production of relatively low avidity antisera by selecting for low affinity B-lymphocyte clones.

It is apparent that mice consistently do not respond well to certain antigens or small peptides, or antigens that are not immunogenic. In order to generate an immune response to small molecules, also known as haptens, conjugation to larger carrier proteins is essential. Keyhole limpet hemocyanin (KLH) is mostly used as a carrier protein to conjugate synthetic peptides or small molecules and subsequently used as the immunogen to inject mice [35–38].

### 7.2 Isolation of B-Cells from the Spleen

The immunized mouse exhibiting high antibody titers against the antigen is euthanized in a $CO_2$ chamber. The blood is quickly collected from the heart and the serum is separated and stored at –20 °C as a positive control. The mouse is completely submerged in a beaker containing 70% alcohol to maintain aseptic conditions and the beaker is placed in a biosafety hood. The mouse is placed on a petri dish so that its left side is facing upward. The area is washed with 70% alcohol or betadine and

cut open to isolate the spleen. Using sterile scissors and forceps, the connective tissue and fat are trimed off from around the spleen. The spleen is placed in the petri dish containing 10 mL of RPMI-1640. The spleen is injected with culture medium to release the cells. Using a syringe and a cell strainer, a single cell suspension is obtained. The cell pellets are resuspended in 5 mL of ice cold lysing buffer (0.15 M NaCl, 10 mM $KHCO_3$, 0.1 mM EDTA pH 7.3) to remove the red blood cells. The cell pellet should be erythrocyte-free and the spleen cells are now ready for fusion with myeloma cells.

### 7.3 Fusion and Postfusion

The development of the drug selection procedure employing HAT (hypoxanthine, aminopterin, and thymidine) in the early 1960s by Littlefield's group [39] and the construction of appropriate myeloma cells devoid of HGPRT (hypoxanthine guanine phosphoribosyl transferase) were critical to development of hybridoma technologies to usher in the monoclonal revolution as a new class of protein therapeutics. Currently, the SP2/0 myeloma cell deficient in the HGPRT enzyme is one of the common choices for fusion with spleenocytes. Several other myeloma cells have been made including rat and human myelomas. While mouse and rat myelomas have been used successfully to generate hybridoma secreting mouse and rat antibodies, respectively, the human myelomas have been problematic, including hybridoma yield and stability. An inherent and key limitation of cell fusion hybridoma technology is that the fusion is random and the frequency is low (<1% to 5%) [40]. As a result, many a good, specific, high affinity antibody secreting B-cell may not fuse with the myeloma cell. This problem is mitigated in phage display and other technologies described in subsequent sections. The myeloma cells should be in the log phase of growth when used for the fusion. It is also essential to check the HGPRT deficiency of the SP2/0 cells by growing them in media with 8-azaguanine. Since the SP2/0 cells are HGPRT deficient, SP2/0 cells should not die in 8-azaguanine.

Spontaneous fusions of cultured cells rarely occur; however, the addition of certain viruses or chemical fusogens can markedly increase the rate at which they fuse. Electrofusion methods have also been developed as alternative strategies. PEG (polyethylene glycol) is the chemical fusogen most commonly used for cell fusion. Cell membranes adhere because of hydrophobic interactions between membrane phospholipids and a small proportion of them break down and reform so that fused cells are formed. Early techniques employed addition of peritoneal macrophages as feeder cells for promoting the growth of hybridoma cells. Alternatively, the HAT medium is supplemented with 50 μg/mL of LPS and 20 μg/mL of dextran sulfate. These agents stimulate the spleen cells to respond well to these mitogens. In about 7–10 days hybridoma clones start to appear.

Following the hybridoma fusion, the first 10 days appear to be the most critical to cell survival in HAT selection medium. The cell population consists of fused and unfused B and myeloma cells. Although the unfused lymphocytes die rapidly, it is necessary to introduce a selective mechanism to separate unfused myeloma cells from fused heterohybrid cells. HGPRT is the key enzyme of the salvage pathway for DNA synthesis. Fusion of myeloma cells with the B-cells compensates for their deficiency of HGPRT. Aminopterin blocks the *de novo* DNA synthesis pathway and since unfused myeloma cells cannot use the *de novo* pathway for DNA synthesis,

they subsequently die due to the lack of the salvage pathway HGPRT enzyme. Heterohybrid (myeloma fused with lymphocytes) fused cells are able to bypass the *de novo* pathway blockage due to the functional HGPRT in the lymphocyte and survive only through the salvaged pathways. In this selection mechanism, only fused cells are able to grow in the HAT medium.

Due to great variation in growth at this critical time, it is extremely important to care for the hybrids within individual wells. Each plate should be microscopically checked often following fusion for colony formation as well as potential contamination problems. Many wells may not show any growth and others may contain more than one hybrid cell. Nonproducing hybrids usually outgrow the producing hybrids in the same well and eventually completely overtake them; therefore, care must be taken to avoid this with early recloning at limited dilutions of <1 cell per well. Nonproducers are hybrid cells that may have received enough genetic information from both cells to survive selection and remain viable, but lack antibody production. Since immunoglobulin synthesis represents ~30% of the total B-cell protein synthesis, nonsecretory hybrid cells have lesser metabolic burden and hence can have a selective advantage to outgrow in the wells.

Some hybrids grow slower than others and may not initially produce antibodies within the detection limits of the screening method. A well that initially appears to be slightly positive and then becomes negative is most likely due to the production of immunoglobulin from unfused B-cells before they die. When a third of the well is covered with clones or the color of medium changes to orange or yellow, clones are ready for screening using ELISA. Solid phase ELISA is the most common method for selecting desired clones. This method of screening involves coating the ELISA plate with the desired antigen (1 μg/well). Following a blocking step, an aliquot of the clone supernatant (100 μL) is incubated with the antigen coated plate and the antigen–antibody complex is detected using an enzyme-labeled antibody capable of binding to the Fc region of the mouse antibody. Clones showing ELISA values five times higher than the negative control are considered positive. The serum of unimmunized mice and irrelevant antibody are used as negative controls. A positive serum from immunized mice is generally used as positive control. The strong positive clones are cultured and expanded in 24 well tissue culture plates. All the positive clones should be cultured for at least 10 days in HT medium (hypoxanthine and thymidine) to utilize the salvage pathway for synthesizing nucleic acids and to ensure that all traces of aminopterin have been removed from the wells.

### 7.4 Cloning

When a microtiter well tests positive for the antibody of interest, it is important to reclone as soon as possible to avoid the potential loss of the positive clone due to overgrowth by nonsecreting cells. Cloning is the isolation and propagation of a hybridoma cell at limiting dilution of less than one hybridoma per well to reach monoclonality of the culture. The application of the MAb technique can only be realized if the cell culture producing the antibody is truly monoclonal, consisting solely of cells derived from a single precursor hybrid cell. Successive recloning two to four times can ensure monoclonality and the isolation of the stable clone. In the soft agar method, the heterogeneous population is dispersed in warm, molten agar and poured into petri dishes. As the cells begin to divide, individual colonies can be

selected and transferred to tissue culture plates for expansion and screening. Cloning by limiting dilution is less labor intensive and more widely used, since the primary hybridoma is merely diluted in media and aliquoted at less than one cell per well. Limiting dilution recloning requires an accurate cell count of the viable hybridoma cells suspended in the medium of the microtiter well in which the cells are growing. If less than 80% of the clones tested are positive in the second recloning step, the hybridoma should be recloned until 90–100% cloning efficiency is reached. The two best clones with the highest ELISA values are chosen and a cell bank is prepared for long-term storage in liquid $N_2$. Valuable therapeutic hybridoma clones should also be protected by storage at the ATCC, USA which is an excellent repository for a variety of biologicals. Alternatively, advances in DNA sequencing by automated machines allow rapid determination of the variable heavy and light chains of the monoclonal antibody for long-term protection of valuable clones. Such sequence information can utilize total synthesis of the gene and cloning for possible resurrection of the valuable antibodies by recombinant DNA techniques.

### 7.5 Cell Line Characterization and Cryopreservation

Once a hybridoma has been cloned and recloned, it is necessary to characterize the line to determine the conditions for optimal growth and good antibody secretion. Mycoplasma contamination of any line needs to be tested. A number of commercial tests are available to detect mycoplasma. These include general mycoplasma tests to specific PCR detection of individual species. The master hybridoma cell lines are cryopreserved to allow for long-term storage. Generally, a freezing medium (90% FBS and 10% DMSO) is used for long-term storage in liquid $N_2$. A simple commercial device is available (Freezing container, Nalgene Inc.) to gradually cool the hybridoma cells to –80 °C for 24–48 h, and then transfer to liquid $N_2$ is essential. Thawing of stored hybridoma cells in primary or secondary banks is a rapid procedure. Unlike slow cryopreservation of cell lines, thawing must occur very quickly to minimize damage to the cells. The stored cell line is thawed in a 37 °C water bath for 1 min and subsequently cultured for expanding the clone.

### 7.6 Small Scale Bioreactor Culture

The thawed hybridoma cell line is cultured and expanded for generating small amounts of the MAb. Early methods routinely injected hybridomas into pristane primed mice to generate ascites fluid to purify the MAb. Such methods are currently discouraged due to animal care concerns. In addition, the ascites could also harbor small amounts of endogenous mouse polyclonal antibodies that copurify with the monoclonal antibody. This defeats the very purpose of the development of monoclonal antibodies. Hence, *in vitro* culture of hybridomas in any one of the commerical small scale bioreactors is recommended to prepare pure MAb for preclinical characterization of the biopharmaceutical.

The I-MAb™ bioreactor bag is a monoclonal antibody production system we routinely use to make laboratory amounts of MAbs (~25–50 mg). This is based on the feature of high surface area/gas exchange of $CO_2/O_2$ in a closed environment for increased cell growth, metabolism, and decreased risk of contamination. The bag

is incubated until the cell viability falls to 10–20%. This is usually reached at the mid 3–4 week incubation period. Antibody production continues during the declining phase of growth and is often maximal during this phase. This is supported by two theories: viable cells secrete antibodies and dead cells release the antibodies. After an incubation period of 25–30 days, the I-MAb bag is harvested and clarified by centrifugation. We have successfully reutilized the bioreactors for several cycles to minimize the cost of antibody production.

### 7.7 Purification

Generally, the desired species of MAbs are of the IgG class. Often the subclass of the IgG is determined by the subtype of specific commercial reagents. Purifying MAbs from a small scale bioreactor can be performed by a number of methods. Classical biochemical methods exploited ethanol precipitation, salt precipitation, PEG precipitation, caprylic acid precipitation, ion exchange chromatography, thiophilic gel chromatography, or affinity chromatography. Among these methods, only affinity chromatography gives high purity of IgG in one simple step. Affinity chromatography is of two broad types—namely, group-specific and ligand-specific strategies. In the former, the purification is based on the binding of the Fc domain to affinity supports such as protein G [41] and protein A [42]. The second strategy exploits the specific antigen or mimetic as ligand to bind the antibody via the antigen combining site. Protein G is the most common affinity matrix and is derived from a cell wall protein of type G streptococci that has an affinity for the Fc region of immunoglobulin G (IgG). Protein G also has a binding site for albumin; however, it has been deleted in recombinant forms of protein G so that it only has affinity for IgG. Protein A does not have a comparable binding affinity or diversity in subclass binding that is found in protein G; many mouse IgG subclass antibodies will not bind to protein A. Protein L is a newer isolated affinity matrix from *Peptostreptococcus magnus*. It binds to the kappa light chain of all antibody classes including IgG. Hence, it is the most versatile of the three proteins [43]. protein L is especially useful for purification of Fab and scFv fragments. Another advantage of protein L is that it does not bind to bovine antibody, which is found in trace amounts in extensively used fetal bovine serum (FBS). The technique of affinity chromatography relies on the ability of proteins to recognize and bind to specific ligands in a reversible manner. Affinity chromatography uses an adsorbent comprised of a matrix to which the ligand is attached. The attachment or bonding is performed so that the immobilized ligand is still able to interact with the protein. An affinity separation is then performed by passing the impure protein over the adsorbent. The target protein is adsorbed, and contaminants pass through. Following adsorption, the adsorbent is washed to remove residual contaminants and the bound protein is eluted in a pure form. Elution is normally achieved by changing the buffer or salt composition so the protein can no longer interact with the immobilized ligand. MAbs differ in their biochemical and ligand characteristics, making it necessary to develop individual protocols for their purification; however, due to the existence of common constant domains (Fc and light chains), group-specific affinity chromatography can be utilized for MAb purification. The immobilized ligands commonly used are protein G, protein A, and protein L. A variation of affinity chromatography is the technique of affinity cochromatography. In this technique, antibody is purified

indirectly [21, 44, 45]. The antigen is captured on an affinity support and this complex is used to isolate antibody specific to the antigen.

## 7.8 Applications

Monoclonal antibodies (MAbs) are used extensively in most immunoassays, such as radioimmunoassays, enzyme linked immunosorbant assay (ELISA), immunohistology, and flow cytometry for *in vitro* diagnosis. Several MAbs are also used for *in vivo* diagnostic imaging and immunotherapy of different human diseases. Currently, applications of MAbs fall within four main areas: diagnosis, imaging, therapy, and affinity purification.

*Disease Diagnosis* Polyclonal antibodies have been used in the diagnosis of diseases for several decades; however, they have inherent limitations due to their polyvalent nature of cross-reacting with irrelevant antigens to some extent. MAbs are particularly useful in diagnostics, where monospecificity is the most important criterion [26]. The use of MAbs in the analysis of cell surface antigens is also important in tissue typing for transfusion and qualitative and quantitative analysis of specific cell populations in blood in normal and disease states.

*Molecular Imaging* Imaging is the *in vivo* use of specific or antibody probes to identify and localize disease. Imaging can be considered intermediate to diagnosis and therapy. Its purpose is to search for and localize antigens such as tumor location within the body, for treatment decisions to be made. Imaging antibodies need to meet a stringent set of criteria: they must be stable during labeling and after administration and be of sufficiently high affinity to localize in small tumors. Several radiolabled antibodies are approved by the FDA for *in vivo* imaging. A more recent innovation is fluorescence-based imaging.

*Therapy* A variety of protein and carbohydrate molecules on the surface of cancer cells are potential targets for antibody-directed therapies [22, 24, 25, 46–49]. Clinically useful targets may be uniquely expressed by cancer cells or expressed at higher levels by cancer cells than by normal cells. Consequently, antibodies will selectively bind to or accumulate in cancer cells. The antibody can act in several ways including the induction of apoptosis, the blocking of growth factor receptors, induction of cell- or complement-mediated cytotoxicity, or the blocking of angiogenesis. Targeting strategies can be an antibody linked to a radioactive particle, drug loaded liposome immunotoxin, or a chemotherapeutic agent. Targeting a receptor that internalizes upon antibody binding will lead to intracellular delivery of the toxic agents. By attaching liposomes containing cytotoxic drugs to MAbs or smaller fragments, anticancer drugs can be delivered to the cell interior [29]. The treatment of solid tumors with such relatively large molecules as immunoglobulin presents certain challenges, especially since direct action requires the antibody to penetrate the tumor mass. Penetration may be facilitated by the use of antibody fragments, which can be made by recombinant technology.

The potential use of MAbs in therapeutics was quickly recognized and the first MAb was approved for therapeutic use in 1986. This was a mouse monoclonal to CD3 antigen for the prevention of transplantation rejection. The U.S. FDA has since

approved 18 different MAbs for treating different human diseases [50]. Current sales of the antibody-based protein pharmaceuticals exceed $10 B and is expected to grow to $16.7 B in 2008 [51] and to $26B by 2010 (http://www.drugresearcher.com/news/ng.asp?id=60408-worldwide-antibody-market).

### 7.9 Immunogenicity Issues and Humanization

Following the discovery of hybridoma technology, murine monoclonal antibodies have proved useful in *in vitro* diagnostics of various diseases; however, when used *in vivo* for the treatment of patients with various ailments, their effect is not always sustained [22, 25, 26]. When mouse antibodies are administered to patients, human immune systems elicit human antibodies against the foreign mouse domains. The specific response of human antibodies directed against mouse antibodies is referred to as HAMA (human anti-mouse antibody) [52, 53]. The HAMA response can complex and neutralize the mouse antibodies by rapid clearance from the blood, thus preventing the mouse antibody from binding to its target. In addition, mouse antibodies only weakly recruit the human immune system components that may be necessary to clear the targeted antigens or tumors. Rodent antibodies tend to provoke strong HAMA responses, which restrict their usefulness for repeated application in the same patient. Hence, current MAb pharmaceuticals would have to be at least a chimeric antibody with a substantial human constant domain.

The first chimeric antibodies were developed in 1984 to solve the problem of the HAMA response [54]. Chimeric antibodies are antibodies in which the constant regions (H and L chains) of a mouse or rat antibody are replaced with human constant regions by recombinant DNA technology. This provides the antibody with human effector functions and also reduces immunogenicity (HAMA) caused by the rodent Fc region. While chimeric antibodies solved the problem of efficient interaction with the human immune system, these antibodies, which have about 30% mouse and 70% human protein sequence, could still elicit a HACA (human anti-chimeric antibody) response against the chimeric mouse part, therefore limiting the use of MAbs as human therapeutics [53, 55].

The next step in MAb evolution was the further humanization beyond the chimeric antibodies. The technology of antibody humanization was developed in 1986 in which only those parts of a mouse antibody that are directly involved in binding to its specific target are transplanted from the mouse antibody to a human antibody framework or CDR grafting [56]. It was thought that humanization would remove much of the immunogenicity remaining with chimeric antibodies; however, human anti-human antibody (HAHA) responses were also observed, although with a reduced incidence compared with HAMA and HACA [53]. An optimal humanized antibody would contain the minimum number of mouse domains required to reproduce the binding strength and specificity of the original mouse antibody and can still be capable of recruiting components of the human immune system. Humanization or fully human antibodies can be developed in several ways: (1) humanized/CDR grafted/reshaped antibodies, (2) human antibodies from immune donors, (3) fully human antibodies from phage libraries, (4) humouse technology, and (5) deimmunization [57].

The ongoing immunogenicity problems of using MAbs as human therapeutics have continued to be an issue [58]. The evolution of MAb technology that began with the generation of rodent chimeric and then humanized MAbs has culminated

with the development of fully human MAbs [59]. Human B-cells are immunized *ex vivo* in the presence of human antigens and then immortalized by means of cell fusion. The group has generated MAbs specific to a number of human antigens and some of them show neutralizing activity against human proteins [60].

## 8 BISPECIFIC MONOCLONAL ANTIBODIES

Bispecific monoclonal antibodies (BsMAbs) have distinct binding specificities to two different antigens incorporated in the same molecule. The early bispecific antibodies were designed by chemically coupling two distinct polyclonal antibodies [61]. Cell fusion of two hybridomas to generate hybrid hybridomas or quadroma secreting bispecific antibodies was an alternate convenient strategy [62]. Triomas are BsMAb secreting cells derived by fusing a hybridoma with spleen cells derived from a mouse immunized with a second antigen. The coexpression of two different IgGs in triomas and quadromas result in the assembly of bispecific MAbs as well as unwanted MAb species resulting from random pairing of heavy and light chains. These unwanted contaminants compete with the activity of the bispecific antibody and must be removed from the mixture, posing a significant challenge for purification. Exploiting elegant protein engineering techniques, it is now possible to engineer the inter-CH3 domain disulfide bonds of heavy chains so that they preferentially form heterodimers over homodimers [63]. The light chain mispairing problem is resolved by using an identical light chain for each arm of the bispecific antibodies. By these methods, nine unwanted H and L chain pairings were eliminated from the population and only one pure product was obtained for further use. Recombinant DNA strategies have also been used to generate bispecific antibodies. The heavy and light chains of the two immunoglobulin binding regions are independently linked by a $(\text{Gly-Gly-Gly-Gly-Ser})_3$ linker and two single chain antibodies are linked by a different peptide to generate a bispecific single chain MAb.

BsMAbs are unique heterobifunctional nano cross-linkers that remove chemical conjugation steps to cross-link any two desired antigens or cells in a predetermined fashion. Such probes have a variety of applications in immunohistology, immunoassays, immunoimaging, immunotherapy, and even retargeting of cytotoxic or dendritic cells to cancer sites [64–67]. Full length bispecific antibodies forming heterodimers based on the knobs-in-holes concept has also been described for therapeutic applications [68].

The use of bispecific molecules in therapy is based on the selective recruitment of an effector mechanism against a defined disease-related biomarker. Hence, bispecific molecules serve as bridges or mediators between an effector and a target. Bispecific molecules are capable of recruiting effector molecules (toxins, drugs, prodrugs, cytokines, and radionuclides), effector cells (NK cells, macrophages, cytotoxic T lymphocytes, or dendritic cells), and the retargeted carrier systems (viral vectors for gene therapy). Therapeutically useful bispecific antibodies should also have the following criteria for *in vivo* uses: (1) be nonimmunogenic to avoid neutralizing immune responses, (2) have a defined structure and bind to the effector cells to induce activation only after binding to the target cells, (3) have a minimum size to allow penetration into the tumor tissues and exhibit sufficiently long half-lives to

induce prolonged therapeutics effects, and (4) be flexible if possible to accommodate a variety of biologicals.

## 9 RECOMBINANT ANTIBODY

### 9.1 Phage Display

In classical hybridoma technology, hetero fusion efficiency between a B-cell and myeloma is low. Consequently, many of the valuable B-cells secreting high affinity and high specificity antibodies fail to form viable hybridomas. Attempts have been made to improve fusion efficiency with polyethylene glycol, electrofusion, and affinity strategies, but these approaches do not encompass the entire antibody repertoire generated by the mammalian immune system upon immunization with a specific antigen. To overcome this limitation and to minimize the immunogenicity issues of full length murine antibodies, recombinant DNA technology introduced phage display technology for the surface expression of the Fab or scFv libraries derived from total splenic mRNA.

Phage display technology is a very powerful technique for generating recombinant libraries containing millions of different peptides, proteins, or small molecules. Phage display has been used to define epitopes for monoclonal antibodies, select enzyme substrates, and screen cloned antibody repertoires for affinity screening of combinatorial peptide libraries to identify ligands for peptide receptors [69]. Phage display technology provides a very effective means to identify proteins that bind to a target molecule of interest [69–73]. A large variety of phage display systems have been developed. The most common is the phagemid vector-based system, where the gene that codes for the protein of interest is genetically cloned in frame into the viral gene 3 protein (g3p). The proteins encoded by the library are expressed on the surface of the phage and designed to select those that bind to the target molecule of interest by a simple panning step. Phages that bind the specific target molecule contain the gene for the protein and have the ability to replicate and propagate in the proper host cell.

Bacteriophage display of combinatorial antibody libraries can select monoclonal antibody reductants of a desired specificity without the use of conventional hybridoma technology. Two general methods have been used to generate phage display libraries. One technique utilizes immunized animals, while the other uses naive Ab genes. With the first technique, an animal is immunized with a target molecule of interest and a phage display library is constructed using the total complement of sequences from the antibody mRNA in the lymphocytes of the immunized animal. The isolation of specific antibodies from a cloned immunological repertoire requires a large, diverse library, as well as an efficient selection procedure. To achieve this goal, a natural strong immune response must be generated in mice and the quality of the total splenic mRNA and cDNA from which the library is constructed must be high. A high serum antibody titer against the immunizing antigen usually indicates that there is enrichment for antigen-specific clones. The second method to generate a phage display library uses germline antibody genes to construct a naive phage display library [74, 75]. This library is not initially enriched with antibody sequences that bind to a particular molecule; therefore, it is more challenging to

isolate the antibodies of interest. The advantage of a naive library is that it only needs to be synthesized once and can be used to isolate a clone and improve the binding affinity by random and selective modifications. The applications of phage technology include the creation and screening of libraries to discover novel therapeutic targets and methods for selection of biologically active ligands.

### 9.2 Yeast Display

*Saccharomyces cerevisae* is useful as a host for the recombinant eukaryotic protein expression, since it allows the proper folding and glycosylation of expressed heterologous proteins [76–78]. The rigid structure of the cell makes yeast suitable for several genetic applications especially cell surface display of foreign proteins. Yeast display exploits the agglutinin adhesion receptor protein to display recombinant protein on the surface of the yeast [79, 80]. The receptor consists of two proteins: Aga1 and Aga2. Aga1 is secreted from the cell and becomes covalently attached to β-glucan in the extracellular matrix of the yeast cell wall. Aga2 is also secreted by the cell and binds Aga1 through two disulfide bonds and remains attached to the cell wall via Aga1. The yeast display system takes advantage of the association between Aga1 and Aga2 by fusing heterologous proteins to the C terminus of the Aga2 protein within a yeast display vector. The resulting construct is then transformed into the yeast strain, which only contains the Aga1 gene. Expression of both proteins results in the association of Aga2 fusion protein to Aga1. Consequently, the tagged heterologous molecule is displayed on the cell surface at 10,000–100,000 copies per cell and can be monitored by fluorescent labeled antibodies recognizing the tags encoded by the display vector.

### 9.3 Ribosome and mRNA Display

Unlike the phage display and yeast display technology, *in vitro* display technology can be used to generate and select the ligands by evolutionary strategies. In the phage display system, the whole library is transformed into *E. coli*, generating the recombinant phage particles displaying the target heterologous molecules. In contrast, DNA library molecules are first transcribed into mRNA *in vitro* and then translated by the ribosome machinery [81]. In this system, the library size is small, which represents the number of different full length protein molecules coupled to the encoded mRNA molecules in the complex. In both the ribosome and mRNA display systems, the functional library size is limited by the amount of *in vitro* translation mixture. The difference between the two methods is how the protein is connected to the mRNA molecules.

In ribosome display, the DNA library is transcribed first to mRNA and then translated *in vitro* with a stoichiometric quantity of ribosomes [82]. Since the DNA has no stop codon, the translation process continues to the end of the mRNA molecule, which results in the nascent polypeptide being bound to the ribosome. The phenotype of the complex is represented by the protein, while the genotype is represented by the mRNA. In order to bind the nascent protein to the target, it has to be properly folded while still attached to the ribosome. Exploiting recombinant DNA technology, the C-terminal region of the DNA library is connected to the ribosome, subsequently providing flexibility for the protein to be folded and be

available as a free unit for binding to the target. The ribosomal complex is captured by the target and the mRNA is isolated by destabilizing the complex. RT-PCR and PCR provide the DNA template for further characterization.

In mRNA display [83–85], the DNA library, which is free of stop codons, is transcribed *in vitro* to mRNA. Each mRNA molecule is then ligated to a short DNA linker containing the adaptor molecule. The modified mRNA is then translated into protein and the translation process will stall at the mRNA–DNA junction. During this step, a peptide bond will form between the adaptor molecule and the nascent peptide by the complex translational machinery. The resulting mRNA and protein complex is purified from the ribosome and subjected to reverse transcription (RT-PCR). The cDNA/mRNA–protein complex is then subjected to selective binding to the target. The isolated enriched complex is hydrolyzed by NaOH and the single stranded DNA is recovered. The recovered DNA is then PCR amplified for the next round of selection and further characterization.

## 10 HIGH LEVEL MAMMALIAN EXPRESSION OF RECOMBINANT ANTIBODIES AND PROTEINS BY GENE AMPLIFICATION SYSTEM

The selection of a candidate preclinical antibody or protein product also requires bulk production for eventual characterization and clinical trials. At least two novel mammalian expression systems have been developed that can amplify the production/secretion of the recombinant antibodies. The range of secretion varies from 100 to 1000 mg/L of mammalian culture. The dihydrofolate reductase (DHFR) gene is used as the amplification vector. It catalyzes the chemical reaction of dihydrofolate to tetrahydrofolate and, at a much slower rate, the conversion of folate to tetrahydrofolate. It has been found that chemical compounds such as methotrexate, pyrimethamine, and trimethoprim inhibit the catalysis. The dihydrofolate reductase system is exploited for high level expression of recombinant antibodies [86]. The most widely used mammalian expression system is the gene amplification procedure utilizing DHFR deficient Chinese hamster ovary (CHO) cells [87–89]. Methotrexate binds to the DHFR enzyme, but DHFR deficient CHO cells, which have taken up mammalian expression vector containing the DHFR gene, can develop resistance to methotrexate by amplifying the DHFR gene. Since the antibody gene of interest is also colinked with the DHFR gene in the mammalian expression vector, it is subsequently coamplified. When the cells are subjected to successive rounds of selection in medium containing stepwise increments in methotrexate levels, recombinant CHO cells with specific antibody productivity also increase. The magnitude of specific productivity of the recombinant CHO cells progressively increases with gene copy number [88]. It has been reported [90] that using the DHFR amplification system with CHO cells resulted in specific antibody productivity as high as 100 μg/$10^6$ cells/day.

Another gene amplification system for high levels of protein biopharmaceutical is the glutamine synthetase based amplification system [91–93]. The CHO cells growing in culture require glutamine for their growth and survival. Glutamine can either be obtained from the medium or be synthesized by the cells from glutamate and ammonia through catalysis by the enzyme glutamine synthetase. The CHO cells grown in a selection medium that is glutamine free and contains methionine sulf-

oximine (MSX is an inhibitor of GS) do not survive. Specific antibody gene expression can be achieved by plating the CHO cells in MSX solution medium and then transfecting the cells with a mammalian expression vector that coded for the GS gene along with the foreign gene of interest. Only those cells that incorporate the vector and express GS at levels sufficiently high to overcome the MSX block will survive. Expression of the foreign gene can be amplified by growing the cells in successively higher concentrations of MSX. The specific gene of interest that is colinked to the GS gene in the same vector or adjacently in the host chromosome is coamplified. As a result of these two amplification strategies based on DHFR or GS vectors, significantly higher levels of antibodies have been obtained for preclinical and eventual clinical use.

## 11 SPECIAL CONSIDERATIONS

Many protein pharmaceuticals are formulated naturally occurring human proteins such as hormones, cytokines, and enzymes. However, several therapeutic proteins are also derived from nonhuman sources such as microbes and animals. Examples include several monoclonals and enzymes such as urate oxidase from *Saccharomyces cerevisiae*, L-asparaginase from *E. coli*, and streptokinase from *Streptococcus hemolyticus*. These foreign therapeutic proteins can elicit significant immunogenic responses upon administration to humans. These include strong humoral and cell-mediated immune responses. As a result, two types of biomedical responses can occur. First, such proteins can elicit allergic reactions, compromising the safety of the patient. Second, the antibodies generated against the nonhuman protein therapeutic can complex and inactivate the pharmaceutical. Two strategies have been adopted to mitigate the immunogenicity of such proteins. The first approach with therapeutic enzymes incorporates the attachment of polyethylene glycol chains on the surface of the therapeutic proteins. This imparts several beneficial characteristics such as reduced immunogenicity, toxicity, clearance, and proteolysis as well as enhanced bioavailability with slower clearance.

Early therapeutic monoclonal antibodies were mouse proteins and hence generated human anti-mouse antibodies (HAMA) against the constant mouse regions as well as against the variable regions. This led to the development of chimeric antibodies incorporating human constant regions in place of the mouse domains. These chimeric antibodies reduced the immunogenicity of such therapeutics. The introduction of fully human therapeutic antibodies has further minimized the immunogenicity but has not eliminated it altogether. A new technique of deimmunization has been introduced to selectively replace the T-cell and B-cell epitopes of preclinical pharmaceutical candidates to mitigate immune responses by rational design [57]. This new approach to designing therapeutic proteins is emerging due to an improved understanding of T-cell modulation of the immune response and new methods for modeling T-cell epitopes using different bioinformatics tools. Potential T-cell epitopes present in the antibody or in any protein are identified using bioinformatics software (EpiMatrix) to analyze the protein sequence for known human HLA class I and/or II epitopes. Once these epitopes are identified, substitution of key amino acids in the T-cell epitopes may alter or attenuate the immunogenicity of the protein. Modification of the key amino acids can abolish binding to human class II MHC

molecules and presentation of the peptides, in the context of MHC, to T-helper cells. Following substitution of the key amino acids, immunogenicity of the modified protein can also be evaluated *in vitro*.

## 12 SUPRAMOLECULAR THERAPEUTICS: CELLS, TISSUES, AND ORGANS

A unique new category of biopharmaceuticals is the use of whole cells as therapeutics, which has already became a reality. The U.S. FDA has approved the use of a magnetic device for the isolation and purification of CD34 antigen positive bone marrow cells for bone marrow therapy. This device is an affinity column that has the anti-CD34 antibody coated with paramagnetic particles as the affinity reagent to isolate highly purified bone marrow cells using a magnet free of contaminating cancer and other immune cells for transplantation.

Pioneering work in our university has isolated and purified human islets from both cadaver and live donor pancreas can be transplanted to diabetics [94–97]. Several patients injected with the islet cells into the liver via the portal vein have been free of daily injections of insulin. Advances in tissue therapeutics have also made artificial skin for grafting in burn injury as well as regeneration and reconstruction of the artificial bladder using the three-dimensional growth of bladder cells. In the future, the horizons of therapeutic possibilities will merge with several novel nonbiological materials including nanotechnologies. A case in point is the cardiovascular stents coated with biologicals to prevent clotting, encapsulation of islet cells in microbeads, implantable devices, targeted nanodevices, and biologicals.

## ACKNOWLEDGMENTS

This work was supported by NIH grant # 5U01AI061233-02 and an NSERC strategic grant.

## REFERENCES

1. Kohler G, Milstein C. Continuous cultures of fused cells secreting antibody of predefined specificity. *Nature* 1975;256(5517):495–497.
2. Merrifield RB. Solid-phase peptide synthesis. 3. An improved synthesis of bradykinin. *Biochemistry* 1964;3:1385–1390.
3. Merrifield RB. Automated synthesis of peptides. *Science* 1965;150(693):178–185.
4. Das D, Kriangkum J, Nagata LP, Fulton RE, Suresh MR. Development of a biotin mimic tagged ScFv antibody against western equine encephalitis virus: bacterial expression and refolding. *J Virol Methods* 2004;117(2):169–177.
5. Zhang G, Brokx S, Weiner JH. Extracellular accumulation of recombinant proteins fused to the carrier protein YebF in *Escherichia coli*. *Nat Biotechnol* 2006;24(1):100–104.
6. Bendig MM. The production of foreign proteins in mammalian cells. *Genet Eng* 1988;7:91–127.

7. Colosimo A, Goncz KK, Holmes AR, Kunzelmann K, Novelli G, Malone RW, Bennett MJ, Gruenert DC. Transfer and expression of foreign genes in mammalian cells. *Biotechniques* 2000;29(2):314–318, 320–322, 324 passim.
8. Hinnen A, Buxton F, Chaudhuri B, Heim J, Hottiger T, Meyhack B, Pohlig G. Gene expression in recombinant yeast. *Bioprocess Technol* 1995;22:121–193.
9. Hinnen A, Meyhack B, Heim J. Heterologous gene expression in yeast. *Biotechnology* 1989;13:193–213.
10. Jarvis DL. Foreign gene expression in insect cells. *Bioprocess Technol* 1993;17:195–219.
11. Kost TA, Condreay JP, Jarvis DL. Baculovirus as versatile vectors for protein expression in insect and mammalian cells. *Nat Biotechnol* 2005;23(5):567–575.
12. Boniface GR, Selvaraj S, Suresh MR, Longenecker BM, Noujaim AA. Synthetic glycosylated-HSA analogues for hepatocyte receptor imaging. In Billinghurst MW, Ed. *Current Applications in Radiopharmacology*. Oxford: Pergamon; 1986, pp 157–164.
13. Finishing the euchromatic sequence of the human genome. *Nature* 2004;431(7011): 931–945.
14. Agarwal KL, Buchi H, Caruthers MH, Gupta N, Khorana HG, Kleppe K, Kumar A, Ohtsuka E, Rajbhandary UL, Van de Sande JH, Sgaramella V, Weber H, Yamada T. Total synthesis of the gene for an alanine transfer ribonucleic acid from yeast. *Nature* 1970;227(5253):27–34.
15. Sanger F, Air GM, Barrell BG, Brown NL, Coulson AR, Fiddes CA, Hutchison CA, Slocombe PM, Smith M. Nucleotide sequence of bacteriophage phi X174 DNA. *Nature* 1977;265(5596):687–695.
16. Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci USA* 1977;74(12):5463–5467.
17. Marra MA, Jones SJ, Astell CR, Holt RA, Brooks-Wilson A, Butterfield YS, Khattra J, Asano JK, Barber SA, Chan SY, Cloutier A, Coughlin SM, Freeman D, Girn N, Griffith OL, Leach SR, Mayo M, McDonald H, Montgomery SB, Pandoh PK, Petrescu AS, Robertson AG, Schein JE, Siddiqui A, Smailus DE, Stott JM, Yang GS, Plummer F, Andonov A, Artsob H, Bastien N, Bernard K, Booth TF, Bowness D, Czub M, Drebot M, Fernando L, Flick R, Garbutt M, Gray M, Grolla A, Jones S, Feldmann H, Meyers A, Kabani A, Li Y, Normand S, Stroher U, Tipples GA, Tyler S, Vogrig R, Ward D, Watson B, Brunham RC, Krajden M, Petric M, Skowronski DM, Upton C, Roper RL. The genome sequence of the SARS-associated coronavirus. *Science* 2003;300(5624):1399–1404.
18. Das D, Suresh MR. Copious production of SARS-CoV nucleocapsid protein employing codon optimized synthetic gene. *J Virol Methods* 2006;137(2):343–346.
19. McGrath BM, Walsh G, Eds. *Directory of Therapeutic Enzymes*. London: CRC/Taylor & Francis; 2005.
20. Niazi SK. *Handbook of Preformulation: Chemical, Biological, and Botanical Drugs*. London: Taylor & Francis/CRC Press; 2006.
21. Gupta S, Suresh M. Affinity chromatography and co-chromatography of bispecific monoclonal antibody immunoconjugates. *J Biochem Biophys Methods* 2002;51(3):203–216.
22. Adams GP, Weiner LM. Monoclonal antibody therapy of cancer. *Nat Biotechnol* 2005;23(9):1147–1157.
23. Carter P. Improving the efficacy of antibody-based cancer therapies. *Nat Rev Cancer* 2001;1(2):118–129.
24. Gatto B. Monoclonal antibodies in cancer therapy. *Curr Med Chem Anticancer Agents* 2004;4(5):411–414.
25. von Mehren M, Adams GP, Weiner LM. Monoclonal antibody therapy for cancer. *Annu Rev Med* 2003;54:343–369.

26. Waldmann TA. Monoclonal antibodies in diagnosis and therapy. *Science* 1991;252(5013):1657–1662.
27. Glennie MJ, Johnson PW. Clinical trials of antibody therapy. *Immunol Today* 2000;21(8):403–410.
28. Hudson PJ. Recombinant antibody fragments. *Curr Opin Biotechnol* 1998;9(4):395–402.
29. Allen TM. Ligand-targeted therapeutics in anticancer therapy. *Nat Rev Cancer* 2002;2:750–763.
30. Nilsson BO, Larsson A. Intrasplenic immunization with minute amounts of antigen. *Immunol Today* 1990;11(1):10–12.
31. Spitz M. "Single-shot" intrasplenic immunization for the production of monoclonal antibodies. *Methods Enzymol* 1986;121:33–41.
32. Spitz M, Spitz L, Thorpe R, Eugui E. Intrasplenic primary immunization for the production of monoclonal antibodies. *J Immunol Methods* 1984;70(1):39–43.
33. Svalander PC, Andersson J, Nilsson BO. Intrasplenic immunization for production of monoclonal antibodies against mouse blastocysts. *J Immunol Methods* 1987; 105(2):221–227.
34. Thorpe R, Perry MJ, Callus M, Gaffney PJ, Spitz M. Single shot intrasplenic immunization: an advantageous procedure for production of monoclonal antibodies specific for human fibrin fragments. *Hybridoma* 1984;3(4):381–385.
35. Erlanger BF. The preparation of antigenic hapten-carrier conjugates: a survey. *Methods Enzymol* 1980;70(A):85–104.
36. Harris JR, Markl J. Keyhole limpet hemocyanin (KLH): a biomedical review. *Micron* 1999;30(6):597–623.
37. Harris JR, Markl J. Keyhole limpet hemocyanin: molecular structure of a potent marine immunoactivator. A review. *Eur Urol* 2000;37(Suppl 3):24–33.
38. Kerr DE, Garrigues US, Wallace PM, Hellstrom KE, Hellstrom I, Senter PD. Application of monoclonal antibodies against cytosine deaminase for the *in vivo* clearance of a cytosine deaminase immunoconjugate. *Bioconjug Chem* 1993;4(5):353–357.
39. Littlefield JW, Gould EA. The toxic effect of 5-bromodeoxyuridine on cultured epithelial cells. *J Biol Chem* 1960;235:1129–1133.
40. Coligan JE, Kruisbeek AM, Margulies DH, Shevach EM, Strober W. *Current Protocols in Immunology*. New York: Greene Publishing/Wiley-Interscience; 1992.
41. Bjorck L, Kronvall G. Purification and some properties of streptococcal protein G, a novel IgG-binding reagent. *J Immunol* 1984;133(2):969–974.
42. Forsgren A, Sjoquist J. "Protein A" from *S. aureus*. I. Pseudo-immune reaction with human gamma-globulin. *J Immunol* 1966;97(6):822–827.
43. Kastern W, Sjobring U, Bjorck L. Structure of peptostreptococcal protein L and identification of a repeated immunoglobulin light chain-binding domain. *J Biol Chem* 1992; 267(18):12820–12825.
44. Husereau DR, Suresh MR. A general affinity method to purify peroxidase-tagged antibodies. *J Immunol Methods* 2001;249(1–2):33–41.
45. Xu D, Leveugle B, Kreutz FT, Suresh MR. Mimetic ligand-based affinity purification of immune complexes and immunoconjugates. *J Chromatogr B Biomed Sci Appl* 1998; 706(2):217–229.
46. Alpaugh K, Von Mehren M. Monoclonal antibodies in cancer treatment: a review of recent progress. *Biodrugs* 1999;12(3):209–236.
47. Batra JK, FitzGerald D, Gately M, Chaudhary VK, Pastan I. Anti-Tac(Fv)-PE40, a single chain antibody *Pseudomonas* fusion protein directed at interleukin 2 receptor bearing cells. *J Biol Chem* 1990;265(25):15198–15202.

48. Cheng JD, Rieger PT, von Mehren M, Adams GP, Weiner LM. Recent advances in immunotherapy and monoclonal antibody treatment of cancer. *Semin Oncol Nurs* 2000;16(4 Suppl 1):2–12.
49. Suresh MR. *Handbook of Immunoassay: Cancer Markers*, 2nd ed. Nature Publishing Group, UK; 2005, pp 635–663.
50. Carter PJ. Potent antibody therapeutics by design. *Nat Rev Immunol* 2006;6(5):343–357.
51. Pavlou AK, Belsey MJ. The therapeutic antibodies market to 2008. *Eur J Pharm Biopharm* 2005;59(3):389–396.
52. Brekke OH, Sandlie I. Therapeutic antibodies for human diseases at the dawn of the twenty-first century. *Nat Rev Drug Discov* 2003;2(1):52–62.
53. Mirick GR, Bradt BM, Denardo SJ, Denardo GL. A review of human anti-globulin antibody (HAGA, HAMA, HACA, HAHA) responses to monoclonal antibodies. Not four letter words. *Q J Nucl Med Mol Imaging* 2004;48(4):251–257.
54. Morrison SL, Johnson MJ, Herzenberg LA, Oi VT. Chimeric human antibody molecules: mouse antigen-binding domains with human constant region domains. *Proc Natl Acad Sci USA* 1984;81(21):6851–6855.
55. Hanauer SB. Review article: safety of infliximab in clinical trials. *Aliment Pharmacol Ther* 1999;13 (Suppl 4):16–22; discussion 38.
56. Jones PT, Dear PH, Foote J, Neuberger MS, Winter G. Replacing the complementarity-determining regions in a human antibody with those from a mouse. *Nature* 1986; 321(6069):522–525.
57. Presta LG. Engineering of therapeutic antibodies to minimize immunogenicity and optimize function. *Adv Drug Deliv Rev* 2006;58(5–6):640–656.
58. Amin T, Carter G. Immunogenicity issues with therapeutic proteins. *Curr Drug Discov* 2004;20–24.
59. Trikha M, Yan L, Nakada MT. Monoclonal antibodies as therapeutics in oncology. *Curr Opin Biotechnol* 2002;13(6):609–614.
60. Li J, Sai T, Berger M, Chao Q, Davidson D, Deshmukh G, Drozdowski B, Ebel W, Harley S, Henry M, Jacob S, Kline B, Lazo E, Rotella F, Routhier E, Rudolph K, Sage J, Simon P, Yao J, Zhou Y, Kavuru M, Bonfield T, Thomassen MJ, Sass PM, Nicolaides NC, Grasso L. Human antibodies for immunotherapy development generated via a human B cell hybridoma technology. *Proc Natl Acad Sci USA* 2006;103(10):3557–3562.
61. Nisonoff A, Rivers MM. Recombination of a mixture of univalent antibody fragments of different specificity. *Arch Biochem Biophys* 1961;93:460–462.
62. Cao Y, Suresh MR. Bispecific antibodies as novel bioconjugates. *Bioconjug Chem* 1998;9(6):635–644.
63. Merchant AM, Zhu Z, Yuan JQ, Goddard A, Adams CW, Presta LG, Carter P. An efficient route to human bispecific IgG. *Nat Biotechnol* 1998;16(7):677–681.
64. Cao Y, Lam L. Applications of bispecific antibodies in therapeutics. *Drugs Future* 2002;27(1):33–41.
65. Cao Y, Lam L. Bispecific antibody conjugates in therapeutics. *Adv Drug Deliv Rev* 2003;55(2):171–197.
66. Kontermann RE. Recombinant bispecific antibodies for cancer therapy. *Acta Pharmacol Sin* 2005;26(1):1–9.
67. Segal DM, Weiner GJ, Weiner LM. Bispecific antibodies in cancer therapy. *Curr Opin Immunol* 1999;11(5):558–562.
68. Carter P. Bispecific human IgG by design. *J Immunol Methods* 2001;248(1–2):7–15.

69. Griffiths AD, Duncan AR. Strategies for selection of antibodies by phage display. *Curr Opin Biotechnol* 1998;9(1):102–108.
70. Hill HR, Stockley PG. Phage presentation. *Mol Microbiol* 1996;20(4):685–692.
71. McGregor D. Selection of proteins and peptides from libraries displayed on filamentous bacteriophage. *Mol Biotechnol* 1996;6(2):155–162.
72. Rader C, Barbas CF 3rd. Phage display of combinatorial antibody libraries. *Curr Opin Biotechnol* 1997;8(4):503–508.
73. Winter G, Griffiths AD, Hawkins RE, Hoogenboom HR. Making antibodies by phage display technology. *Annu Rev Immunol* 1994;12:433–455.
74. Gram H, Marconi LA, Barbas CF 3rd, Collet TA, Lerner RA, Kang AS. *In vitro* selection and affinity maturation of antibodies from a naive combinatorial immunoglobulin library. *Proc Natl Acad Sci USA* 1992;89(8):3576–3580.
75. Nissim A, Hoogenboom HR, Tomlinson IM, Flynn G, Midgley C, Lane D, Winter G. Antibody fragments from a "single pot" phage display library as immunochemical reagents. *EMBO J* 1994;13(3):692–698.
76. Boder ET, Wittrup KD. Yeast surface display for screening combinatorial polypeptide libraries. *Nat Biotechnol* 1997;15(6):553–557.
77. Boder ET, Wittrup KD. Optimal screening of surface-displayed polypeptide libraries. *Biotechnol Prog* 1998;14(1):55–62.
78. Boder ET, Wittrup KD. Yeast surface display for directed evolution of protein expression, affinity, and stability. *Methods Enzymol* 2000;328:430–444.
79. Feldhaus MJ, Siegel RW. Yeast display of antibody fragments: a discovery and characterization platform. *J Immunol Methods* 2004;290(1–2):69–80.
80. Kondo A, Ueda M. Yeast cell-surface display—applications of molecular display. *Appl Microbiol Biotechnol* 2004;64(1):28–40.
81. Lipovsek D, Pluckthun A. *In-vitro* protein evolution by ribosome display and mRNA display. *J Immunol Methods* 2004;290(1–2):51–67.
82. Zahnd C, Amstutz P, Pluckthun A. Ribosome display: selecting and evolving proteins *in vitro* that specifically bind to a target. *Nat Methods* 2007;4(3):269–279.
83. Nemoto N, Miyamoto-Sato E, Husimi Y, Yanagawa H. *In vitro* virus: bonding of mRNA bearing puromycin at the 3′-terminal end to the C-terminal end of its encoded protein on the ribosome *in vitro*. *FEBS Lett* 1997;414(2):405–408.
84. Roberts RW. Totally *in vitro* protein selection using mRNA–protein fusions and ribosome display. *Curr Opin Chem Biol* 1999;3(3):268–273.
85. Roberts RW, Szostak JW. RNA–peptide fusions for the *in vitro* selection of peptides and proteins. *Proc Natl Acad Sci USA* 1997;94(23):12297–12302.
86. Kim SJ, Kim NS, Ryu CJ, Hong HJ, Lee GM. Characterization of chimeric antibody producing CHO cells in the course of dihydrofolate reductase-mediated gene amplification and their stability in the absence of selective pressure. *Biotechnol Bioeng* 1998; 58(1):73–84.
87. Gu X, Wang DI. Improvement of interferon-gamma sialylation in Chinese hamster ovary cell culture by feeding of *N*-acetylmannosamine. *Biotechnol Bioeng* 1998;58(6):642–648.
88. Pendse GJ, Karkare S, Bailey JE. Effect of cloned gene dosage on cell growth and hepatitis B surface antigen synthesis and secretion in recombinant CHO cells. *Biotechnol Bioeng* 1992;40(1):119–129.
89. Van Hove JL, Yang HW, Wu JY, Brady RO, Chen YT. High-level production of recombinant human lysosomal acid alpha-glucosidase in Chinese hamster ovary cells which

targets to heart muscle and corrects glycogen accumulation in fibroblasts from patients with Pompe disease. *Proc Natl Acad Sci USA* 1996;93(1):65–70.

90. Page MJ, Sydenham MA. High level expression of the humanized monoclonal antibody Campath-1H in Chinese hamster ovary cells. *Biotechnology* (*NY*) 1991;9(1):64–68.
91. Blochberger TC, Cooper C, Peretz D, Tatzelt J, Griffith OH, Baldwin MA, Prusiner SB. Prion protein expression in Chinese hamster ovary cells using a glutamine synthetase selection and amplification system. *Protein Eng* 1997;10(12):1465–1473.
92. Brown ME, Renner G, Field RP, Hassell T. Process development for the production of recombinant antibodies using the glutamine synthetase (GS) system. *Cytotechnology* 1992;9(1–3):231–236.
93. Cockett MI, Bebbington CR, Yarranton GT. High level expression of tissue inhibitor of metalloproteinases in Chinese hamster ovary cells using glutamine synthetase gene amplification. *Biotechnology* (*NY*) 1990;8(7):662–667.
94. Shapiro AM, Lakey JR, Ryan EA, Korbutt GS, Toth E, Warnock GL, Kneteman NM, Rajotte RV. Islet transplantation in seven patients with type 1 diabetes mellitus using a glucocorticoid-free immunosuppressive regimen. *N Engl J Med* 2000;343(4):230–238.
95. Shapiro AM, Ricordi C, Hering BJ, Auchincloss H, Lindblad R, Robertson RP, Secchi A, Brendel MD, Berney T, Brennan DC, Cagliero E, Alejandro R, Ryan EA, DiMercurio B, Morel P, Polonsky KS, Reems JA, Bretzel RG, Bertuzzi F, Froud T, Kandaswamy R, Sutherland DE, Eisenbarth G, Segal M, Preiksaitis J, Korbutt GS, Barton FB, Viviano L, Seyfert-Margolis V, Bluestone J, Lakey JR. International trial of the Edmonton protocol for islet transplantation. *N Engl J Med* 2006;355(13):1318–1330.
96. Shapiro AM, Ryan EA, Lakey JR. Clinical islet transplant—state of the art. *Transplant Proc* 2001;33(7–8):3502–3503.
97. Shapiro AM, Ryan EA, Lakey JR. Pancreatic islet transplantation in the treatment of diabetes mellitus. *Best Pract Res Clin Endocrinol Metab* 2001;15(2):241–264.

# THE PHARMACOGENOMICS OF PERSONALIZED MEDICINE

Ronald E. Reid

*University of British Columbia, Vancouver, British Columbia, Canada*

**Contents**

1 Introduction
2 Pharmacogenomics
3 Personalized Medicine
4 Molecular Profiling Technologies
  4.1 Genotyping Technologies—Genomics
  4.2 Gene Expression Profiling—Transcriptomics
  4.3 Protein Analysis—Proteomics
  4.4 Metabolomics
  4.5 Enviromics
5 Associative Analysis
6 Conclusion
  Acknowledgments
  References

## 1 INTRODUCTION

According to the European Agency for Evaluation of Medicinal Products (EMEA, http://www.emea.eu.int/), *pharmacogenetics* is defined as "the study of inter-individual variations in DNA sequence related to drug response." This definition emphasizes the studies on allelic variation of metabolic enzymes involved in drug metabolism and points to the early successes in correlating the alterations in drug metabolism with variations in the genes responsible for production of the metabolic enzymes.

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

A first indication of the individuality of disease biochemistry appeared in an article by A. E. Garrod in 1902 in which he described the high concentrations of homogentisic acid in the urine of those patients suffering from alkaptonuria. Garrod believed this could arise from a congenital defect in a particular enzyme at some stage in the metabolism of the amino acids tyrosine and phenylalanine [1]. Garrod also suggested that drug idiosyncracies might also be attributed to inherited differences, thus setting the stage for the scientific investigation of gene–drug interactions.

Some 50 years later A. Forbat, the Anaesthetic Registrar at St. James Hospital in London, reported an incident where a patient undergoing surgery was treated with the muscle relaxant succinyldicholine and showed "no sign of returning respiration at the conclusion of the operation." He concluded that the patient's sensitivity to succinyldicholine was due to a lowered pseudocholinesterase level and he suggested that "further cases of sensitivity to succinyldicholine should be published, and if a low pseudocholinesterase level is found which cannot be explained by disease, malnutrition, or poisoning, their families should be investigated for a hereditary or racial basis of abnormally low pseudocholinesterase level" [2]. Subsequent studies confirmed this conclusion [3–5] and publications on similar "familial" related abnormal responses to primaquine therapy of malaria [6] and isoniazid therapy of tuberculosis [5] established the gene–drug interaction in disease therapy. R. J. Williams's book on biochemical individuality [7] and the introduction of the term "pharmacogenetics" in papers by A. Motulsky [8] and F. Vogel [9], describing genetically determined variability in the effects of drugs on animal species, initiated a philosophy of genetic biomarkers that would be valuable in decisions on specific drug therapies based on the genetic makeup of the individual patient.

Watson and Crick's seminal paper on the structure of deoxyribonucleic acid (DNA) [10] led to a plethora of research into the structure of DNA and the mechanisms by which information was stored and transmitted. Eventually, the analysis of genes and gene products led to the principles that one gene produces one protein. It also led to the principle that the genetic code links the sequence of bases in DNA to the sequence of amino acids in proteins and hence to the structure of proteins which is instrumental in determining protein function. The "Central Dogma" propounded the directional nature of the flow of information in the genome from DNA to messenger RNA to proteins and it became obvious that informational errors in DNA in the form of mutations would show up in the structure of proteins and ultimately the function of these proteins in the cell. Therefore, it appears logical that the normal or abnormal structure of proteins is related to the normal or abnormal function of a cell, which in turn is related to the normal or diseased function of an organism.

The Human Genome Project has demonstrated that the genome sequence is 99.9% identical in all humans. The remaining 0.1% genetic variation provides the basis upon which individuals differ in physical characteristics, susceptibility to disease, and response to drug therapy. Genetic variations or mutations can occur in coding or noncoding regions of the genome. The coding region variations (Fig. 1) include (1) *silent mutations* that produce no change in the amino acid sequence of the expressed protein; (2) *conservative missense mutations* that result in a codon specifying an amino acid with a side chain that has similar physicochemical properties to the amino acid that appears in the normal (*wild-type*) sequence; (3) *noncon-*

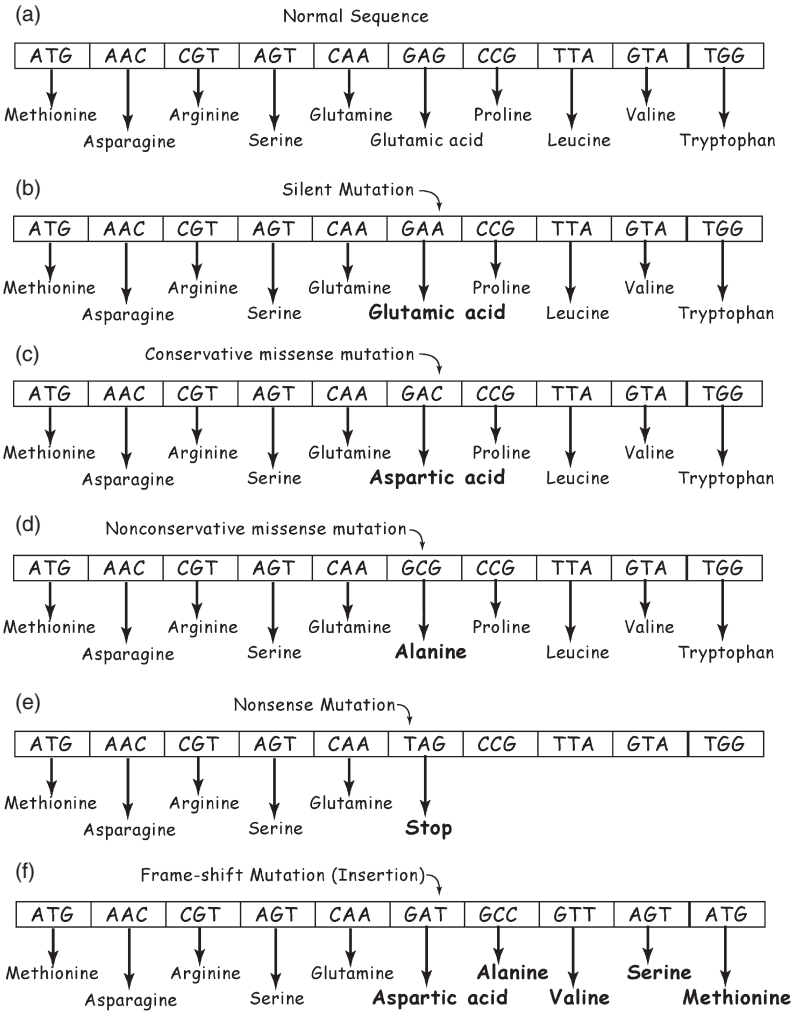


**FIGURE 1** The effects of different types of single nucleotide polymorphisms: (a) the normal genomic DNA sequence and the protein sequence encoded; (b) a silent mutation; (c) a conservative missense mutation where glutamic acid is replaced by aspartic acid through an A/G mutation but both amino acids are acidic residues; (d) a nonconservative missense mutation where glutamic acid (an acidic residue) is replaced by alanine (a hydrophobic residue) through an A/C mutation; (e) a missense mutation where a G/T mutation results in a stop codon, thus truncating the protein; (f) an insertion of a T between the A and G of the GAG glutamic acid codon, thus shifting the reading frame so that all amino acids downstream (including the glutamic acid) are changed. (Adapted from Ref. 11.)

*servative missense mutations* that result in a codon specifying an amino acid with a side chain that has distinctly different physicochemical properties to the wild-type amino acid; (4) *nonsense mutations* that result in a stop codon leading to truncation of the protein; and (5) *frame-shift mutations*, where the addition or deletion of a number of bases that are not a multiple of three (in this example, one base) causes a shift in the reading frame of the gene, thus changing the amino acid sequence of

the protein from the point of insertion onwards, and usually creates a stop codon somewhere in the sequence. Noncoding variations usually occur in the 5′- or 3′-untranslated regions of the genome, causing a possible decrease or increase in expression levels, or in other noncoding regions such as the introns where alternate splice junctions could alter the final protein.

The common forms of "informational errors" found in the human genome are (1) indels, (2) microsatellites or tandem repeats, and (3) single nucleotide polymorphisms. *Single nucleotide polymorphisms* (SNPs), defined as single base pair positions in genomic DNA at which different sequence alternatives (alleles) exist wherein the least frequent allele has an abundance of 1% or greater, are the most commonly occurring genetic variations [12]. These SNPs have two important applications in genome research. First, a polymorphism by nature is a change in the DNA that might alter amino acid coding, gene splicing, or gene expression and hence directly alter phenotype. Second, the nature of SNPs makes them powerful markers for the variants that affect phenotype (see Section 2, HapMap discussion).

The Human Genome Project (HGP) was built on the potential for the application of genetic information to the human condition that would go a long way to opening up the black boxes of disease diagnosis and therapy. Completion of the HGP [13] has changed the focus of genomic research from that of gene identification to that of examining gene structure and function. The relation of gene function and functional variation to disease and drug action and the application of this relationship to disease diagnosis (*D*X) and therapy (*R*X) at the molecular level is the major medical related outcome expected from the billions of dollars spent on sequencing the human genome.

Current pharmacogenetic research has established the importance of genetic information in drug therapy involving the genes expressing drug metabolizing enzymes in the liver. It is now well established that genetic differences between individuals can explain some of the variability seen in the pharmacokinetics of drug therapy, which in turn alters the efficacy and toxicity of certain drugs [14]. Many of these variations are highly penetrant, single gene traits involving the phase I (oxidation) and phase II (conjugation) metabolic enzymes [15] (*penetrance* is the frequency with which individuals who carry a given gene variant will show the manifestations associated with that variant; if penetrance of an allele is 100% then all individuals carrying that allele will express the associated disorder).

In principle, one is able to manipulate the therapeutic outcome of a patient population by separating the patients who will not respond normally to a particular drug therapy from those who will, based on an analysis of established genetic biomarkers related to the drug's metabolic profile (Fig. 2). The patients homozygous for the genetic biomarkers associated with poor metabolism of the recommended therapeutic agent may, depending on the therapeutic window of the drug, exhibit a response resembling an overdose to the "normal" drug dose. These individuals will require a reduced dose to achieve therapeutic levels (see Patient A in Fig. 3). Meanwhile those patients carrying biomarkers associated with extensive metabolism may exhibit a lack of efficacy at "normal" doses and therefore would require an increased dose of the drug to achieve therapeutic activity (see Patient D in Fig. 3). The patients who are homozygous or heterozygous for the normal (wild-type) genetic biomarker (see Patients B and C, respectively, in Fig. 3) may require no alteration of the "normal" dose of the drug for adequate therapeutic outcome.

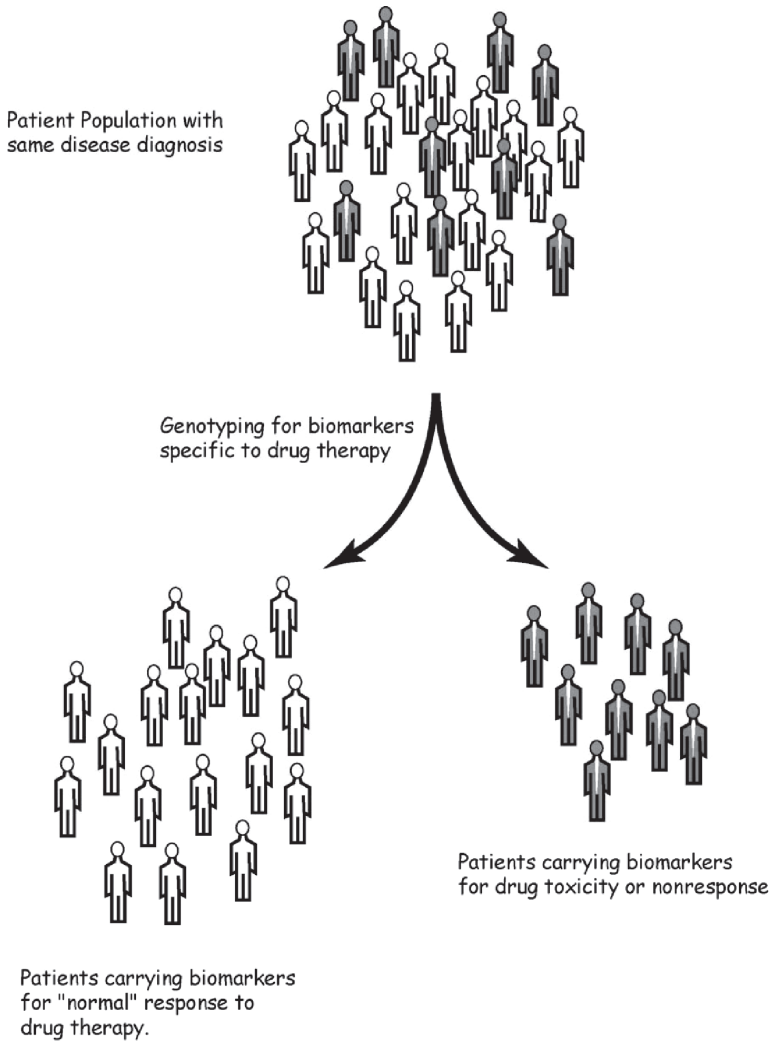


**FIGURE 2** Manipulating therapeutic outcomes through pharmacogenetic analysis of a population with the same disease diagnosis to eliminate those patients with genetic biomarkers indicating an adverse response to a particular drug therapy.

Documented success in correlating the therapeutic activity of a number of drugs with allelic differences in the genes for such common metabolic enzymes as cytochrome P450 2D6 (CYP2D6), CYP2C19, CYP2C9, thiopurine methyltransferase (TPMT), and *N*-acetyltransferase type 2 (NAT2) (Fig. 4) has led to the appearance of pharmacogenetic information on drug labels for aripiprazole, atomoxetin, and celecoxib. The U.S. Food and Drug Administration (USFDA) has set up guidelines for the drug industry to submit voluntary pharmacogenetic data on drugs for which they are seeking approval (http://www.fda.gov/cber/gdlns/pharmdtasub.pdf). In response, a number of companies have developed instrumentation and assays for pharmacogenetic tests, including the analytical instrument division of Roche, which

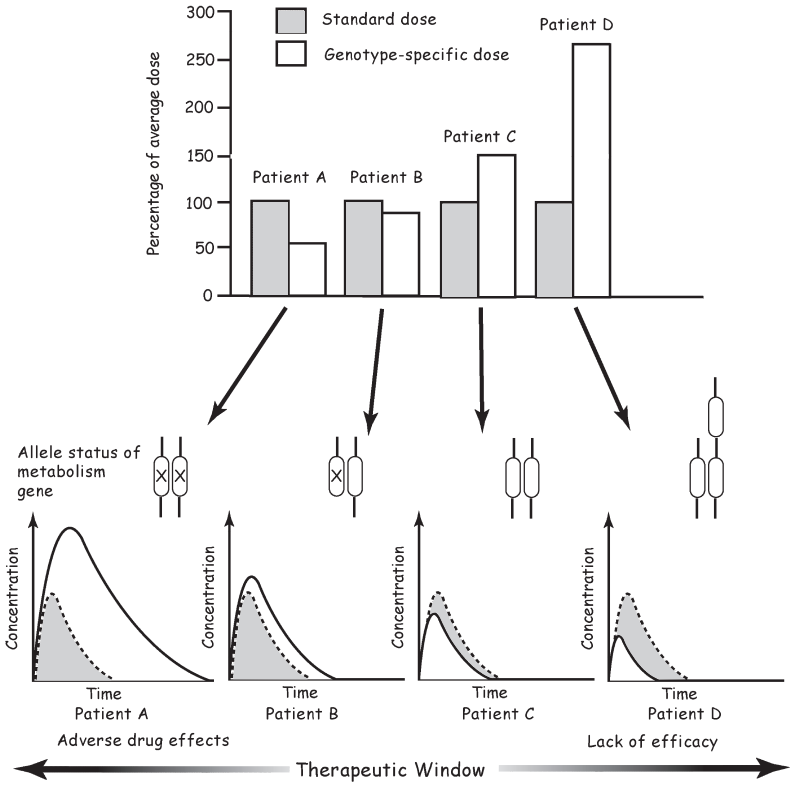


**FIGURE 3** Schematic representation of the potential benefit of adjusting dose to genotype. The dashed line in the concentration–time curves for Patients A–D is the theoretical "normal" drug concentration–time curve, while the solid line is the theoretical "genetic" drug concentration–time curve for each patient. The allele status for each patient is shown above the theoretical plots. The oval containing the X indicates the presence of the detrimental allelic variation. The extra oval in Patient D indicates a duplication of one of the normal alleles. The top bar graph indicates the normal doses (shaded) compared to the dose adjustments that should be made based on the patient's allele status. (Adapted from Ref. 16 with permission of the author and Nature Publishing Group.)

is marketing a gene chip for the analysis of CYP2D6 genetic polymorphisms; Applied Biosystems, which has developed a TaqMan™ assay targeting 220 genes in phase I and II metabolic enzymes and drug transport proteins; and Third Wave, which has recently received FDA clearance for *in vitro* diagnostic testing of UGT1A1 polymorphisms using their Invader™ assay. Finally, it has been proposed that drug dosing adjustments based on genotype could be beneficial if the pharmacokinetic differences in the genetic variants fall outside the acceptable range for stating bioequivalence (*bioequivalence* of two formulations of a given drug occurs if the area under the curve, maximum blood concentrations, and times of maximum blood concentrations differ by no more than 80–125%), unless sufficiently powerful studies can show that the pharmacokinetic differences do not affect the clinical outcome [17].

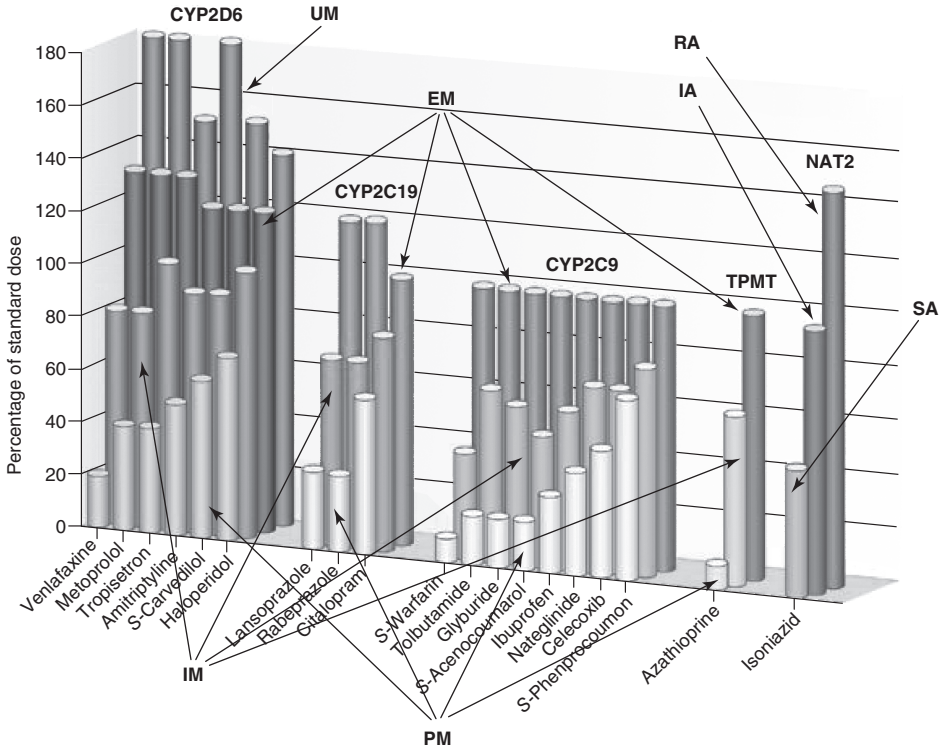


**FIGURE 4** Examples of dose adjustments based on pharmacogenetic information. The influence of genetic polymorphisms in CYP2D6, CYP2C19, CYP2C9, TPMT, and NAT2 is expressed as subpopulation-specific dosages, according to the difference in pharmacokinetic parameters from clinical studies. Substantial adjustments need to be made to the drug dose to achieve the same level of drug exposure in individuals with different genotypes. EM, extensive metabolizer; IA, intermediate acetylator; IM, intermediate metabolizer; PM, poor metabolizer; RA, rapid acetylator; SA, slow acetylator; UM, ultrarapid metabolizer. (From Ref. 17 with permission of the author and Nature Publishing Group.)

## 2 PHARMACOGENOMICS

The failure of pharmacogenetics to make major inroads in clinical applications has led many to question whether or not pharmacogenetics is more a product of hype than science. In spite of the success of the single gene approach (Fig. 5) in correlating the genotype of some drug metabolic enzymes with the dose of drug necessary for effective therapeutic activity, it is a rare occasion that a discrete biological activity such as drug response can be attributed to an individual molecule such as a metabolic enzyme.

Drug action is a reflection of the multifactorial polygenic pharmacokinetic and pharmacodynamic parameters of the drug (Fig. 6). The pharmacokinetic parameters determine the eventual distribution of the drug in the body including the concentration of the drug at the site of pharmacodynamic action. The enzyme action of drug metabolism has already been discussed and is the cornerstone of pharmacogenetics. Other pharmacokinetic parameters include drug absorption,

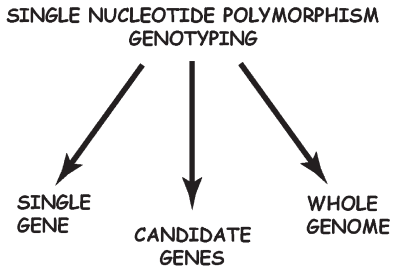


**FIGURE 5** Approaches to genotyping for drug therapy. The single gene approach has been used successfully in pharmacogenetic studies of a number of phase I and II metabolic enzymes. The candidate gene approach involves the selection of the genes for proteins that are demonstrated to be involved in a particular drug action including metabolic enzymes, transport proteins, drug targets, and proteins involved in modular drug response pathways. The whole genome approach utilizes single nucleotide polymorphisms throughout the genome.

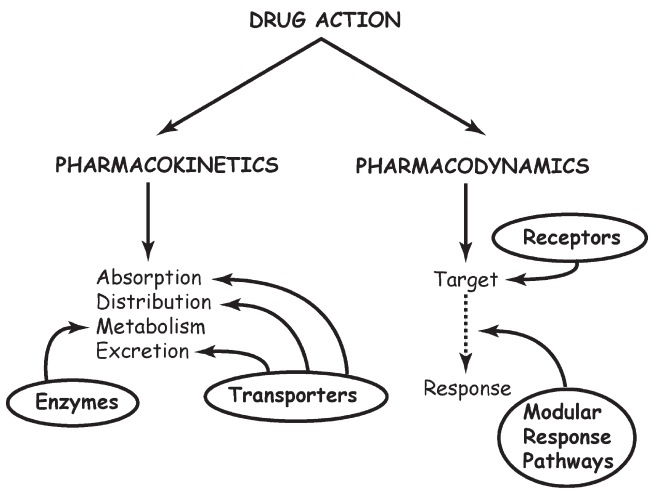


**FIGURE 6** Schematic representation of the multigenic *therapeutic drug pathway*.

distribution, and excretion, all of which may be affected by transport proteins to some extent.

The pharmacodynamics of drug action is a description of the effect of the drug on the body and is a result of complex interactions involving drug targets and the subsequent biochemical reactions, which eventually produce a biological response. The modular nature of biochemical systems lends itself to application in the drug response mechanism in that it is possible to describe the response of many drug–target interactions through linked modules of biochemical systems. The genes involved in drug metabolism, drug transport, and drug responses are considered the *candidate genes* for the therapeutic drug pathway.

The suggestion of genome-wide searches for gene effects using large scale testing of single nucleotide polymorphisms (SNPs) has led to considerable discussion of the efficiency of different approaches [18]. The candidate gene approach [19] suggested a study of coding and/or promoter variants with potential functional significance in disease diagnosis and drug therapy. The whole genome approach [20] suggested that a high density of coding and noncoding SNPs evenly spaced throughout the genome could be used to determine optimal drug therapy. The use of *linkage disequilibrium* (LD; the observation that two or more alleles, usually at loci that are physically close together on a chromosome, are not inherited independently but are observed to occur together more frequently than predicted under Mendel's Law of Independent Assortment) to map SNP loci that may be relevant to disease therapy found that the extent of LD in different parts of the genome is highly variable and those fragments of the genome that appear to be inherited in blocks, indicating a high degree of LD, are known as *haplotype blocks* [21, 22]. Investigations into this haplotype diversity have culminated in the HapMap Project (www.hapmap.org), which will map all haplotype blocks in the human genome and validate the map in four populations: the Yoruba in Nigeria; the Han Chinese in Beijing, China; the Japanese in Tokyo, Japan; and residents in Utah, USA with ancestry from northern and central Europe [23]. It has been demonstrated that, while a haplotype block may contain several hundred SNPs, only a few of the SNPs in a given haplotype block are needed to identify that particular block. These SNPs are known as *haplotype tag SNPs (htSNPs)* and tend to capture most of the variation of a particular region of the genome [24, 25]. This concept of htSNPs has been further extended in the pharmacogenetics of candidate genes such that the number of SNPs needed to characterize a therapeutic drug pathway is reduced by limiting the htSNP variations to those SNPs in genes involved in the therapeutic drug pathway [26–30].

Although recent studies indicate that LD is structured into discrete blocks in the human genome [21, 31, 32], the range and distribution of LD in different populations is largely unknown. Furthermore, the effect of genotype on disease phenotype varies between disorders and populations owing to genetic and environmental heterogeneity [33]. For these reasons, it is difficult to estimate the number of SNPs or the number of samples that would be required for a successful genome-wide LD study. The estimates that have been made vary widely, from 1 million [34] or 0.5 million [35] to as little as 30,000 SNPs [36]. If 0.5 million SNPs were to be analyzed in, for example, 1000 individuals, and if the project was to be carried out in a year, then approximately 1.5 million SNP genotypes per day would need to be produced. At genotyping costs as low as 10 cents per SNP, this example project would cost $50 million (U.S.).

The individual's genotype is only part of the equation for predicting individual drug response and other factors play a contributory role in the individual's response to drug therapy. If one takes the complexity of disease diagnosis, the environment, and drug therapy into consideration, a new definition of the pharmacogenetics concept is required (Fig. 7).

Since *genomics* is defined as a study of the full complement of genetic information both coding and noncoding in an organism (genome), this holistic view of pharmacogenetics is termed *pharmacogenomics*, which is defined as the application of information from the human genome, transcriptome, proteome, metabolome, and envirome to drug design, discovery, and clinical development, reflecting the state of

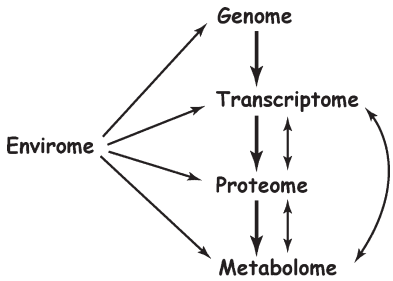


**FIGURE 7** A schematic representation of the biological organization of the "-omes" describing an individual phenotype. The classical view of biological organization is to consider the flow of information from the genome to the transcriptome, to the proteome and then the metabolome (heavy arrows). However, each tier of the organization depends on the other, so a perturbation in one network can affect another (bidirectional light arrows). Furthermore, the environment has a crucial impact on not only expression and concentrations of transcripts, proteins, and metabolites, but also on the genome by selecting from adaptive changes in subpopulations of cells (light unidirectional arrows). Metabolomics can potentially probe much more than classical metabolism and metabolic disorders: it can also be used to monitor changes in the genome or to measure the effects of downregulation or upregulation of a specific gene transcript. (Adapted from Ref. 37 with permission of Nature Publishing Group.)

drug responses at cellular, tissue, individual, or population levels. Pharmacogenomics is a specialized branch of functional genomics reflecting the application of structural genomics to drug therapy. In this context, *structural genomics* is defined as the construction of high resolution genetic, physical, and transcript maps of an organism [38] and *functional genomics* refers to the development and application of global (genome-wide or system-wide) experimental approaches to assess gene function by making use of the information and reagents provided by structural genomics. Functional genomics is characterized by high throughput or large scale experimental methodologies combined with statistical and computational analysis of the results. The fundamental strategy in a functional genomics approach is to expand the scope of biological investigation from studying single genes, proteins, or metabolites to studying all genes, proteins, or metabolites at once in a systematic fashion [38]. Pharmacogenomics, then, aims to study all genes, proteins, and metabolites relevant to drug therapy and its successful outcome.

## 3 PERSONALIZED MEDICINE

Evidence-based medicine is predicated on a standardized application of the scientific principles of unbiased observation and experiment to disease diagnosis and drug therapy. The techniques involved are derived from randomized and double-blind clinical trials and are inconsistent with the treatment of an individual patient [39]. The application of statistical information derived from clinical trials on large populations results in a standard dose range for the population,

which either overdoses or underdoses a small but significant portion of that population. The failure to recognize patients as individuals is likely a factor in adverse drug and toxic drug–drug interactions that account for 100,000 patient deaths, 2 million hospitalizations, and $100 billion in health care costs in the United States yearly [40–43].

With the completion of the Human Genome Project, there has been as shift in research from gene identification to the study of gene structure and function [44]. The early successes in the application of genetic information to disease diagnosis and the estimation of drug efficacy and toxicity have evolved into the concept of *personalized medicine*. This is defined as the utilization of molecular biomarkers from an individual's genome, transcriptome, proteome, and metabolome under the influence of the envirome in the assessment of that individual's health and treatment choices. This includes the predisposition to disease, screening and early diagnosis of disease, assessment of prognosis, pharmacogenomic prediction of therapeutic drug efficacy and risk of toxicity, and monitoring the illness until the final therapeutic outcome is determined (Fig. 8) [39, 43, 45–49].

A *biomarker* is defined as a characteristic that is objectively measured and evaluated as an indicator of normal biologic processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention [50]. A *molecular biomarker* is defined as a specific subset of small molecule (metabolite) or large molecule (protein or DNA) biomarkers that may be discovered using genomic or proteomic technologies as opposed to *clinical biomarkers* arising from imaging technologies [51]. The U.S. Food and Drug Administration elaborates on a *valid biomarker* as a biomarker that is measured in an analytical test system with well-established performance characteristics and for which there is an established scientific framework or body of evidence that elucidates the physiologic, toxicologic, pharmacologic, or clinical significance of the test results [52].

*Molecular profiling* (the application of genomic, transcriptomic, proteomic, and metabolomic technologies to the development of *in vitro* molecular biomarkers) of

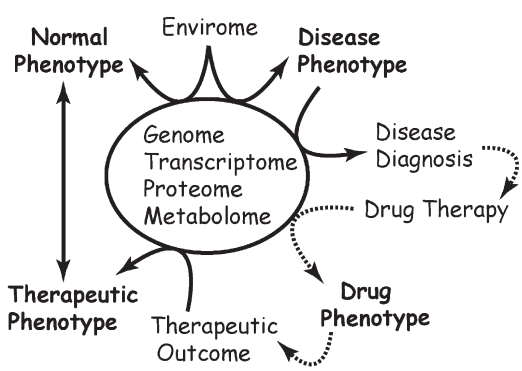


**FIGURE 8** Schematic representation of the concept of molecular profiling for personalized medicine. The dashed arrows indicate the critical components of pharmacogenomics: disease diagnosis, drug therapy, drug phenotype, and therapeutic outcome.

a patient's phenotype (Fig. 8) utilizes molecular biomarkers of the genome, transcriptome, proteome, and metabolome to provide a molecular basis for the normal or healthy patient phenotype in a particular envirome and thereby provides many opportunities for improving the choice of therapy and the prediction of drug response as well as disease pathology [53]. Alteration of molecular biomarkers in the diseased state compared to the normal state (characterized by the normal molecular phenotype) provides a molecular characterization of the disease phenotype, which is utilized in the diagnosis. The disease phenotype is then utilized to choose the best drug therapy, which will alter the molecular biomarkers to a drug therapy phenotype (Fig. 8). The difference between the biomarkers used in the diagnosis and drug phenotype descriptions will be the result of drug therapy. The molecular biomarkers present at assessment of the therapeutic outcome will provide a therapeutic phenotype for that particular drug profile.

In principle, if a drug therapy has been successful in restoring the patient's phenotype to normal, the diagnostic biomarkers for the disease that appear in the drug phenotype, therapeutic phenotype, and normal phenotype should be similar. Attainment of this ideal outcome is heavily dependent on the disease diagnosis, patient biomarker variance, and therapeutic outcome assessment. It is suggested here that the failure of many pharmacogenomic-based therapies is likely the result of inconsistencies in the disease diagnosis and/or assessment of therapeutic outcomes through a lack of sufficient molecular biomarkers or lack of proper interpretation of the biomarkers.

Finally, the interplay between complex genetic factors and the equally complex environmental factors (envirome) is an important yet poorly understood area of research in pharmacogenomics. The genetic regulation of metabolic pathways is also complicated by the fact that they are polygenic or oligogenic, with many different genes interacting in the regulatory process. These genes may also exhibit allelic heterogeneity with many different predisposing risk alleles. Different individuals may also exhibit locus heterogeneity with different or overlapping sets of genes being important in a particular pathway, which is additionally complicated if the gene interactions are additive, synergistic, or epistatic. A further complication arises with individuals when resulting pathologies are qualitative rather than all or nothing traits, with thresholds determining clinical manifestations. It is quite obvious that the application of personalized medicine in the health care environment will require extensive education and training of health care professionals as well as politicians and the public in general.

Walter Fierz has developed the concept of personalized medicine through six dimensions, which I have listed and taken liberty with in my presentation (Fig. 8) [39].

1. *Genome, Transcriptome, Proteome, and Metabolome.* An individual is characterized by a *phenotype*, which is defined as the observable characteristics of an organism, either in total or with respect to one or more particular named characteristics and the phenotype is a manifestation of gene expression in that organism under the influence of the environmental conditions [54]. The observable characteristics of an organism are a product of the molecular state of the individual's genome, transcriptome, proteome, and metabolome in the envirome. The application of molecular biomarkers and molecular profiling of an individual's genome through

SNP analysis and htSNP determination; an individual's transcriptome through mRNA microarray analysis, an individual's proteome through microcharacterization, differential display, and protein–protein interactions; and an individual's metabolome through metabolite profiling provides an accurate molecular description of the individual's phenotype (either normal or diseased), the preferred drug therapy, or therapeutic outcome.

2. *Envirome.* The transcriptome, proteome, and metabolome are a direct result of genome expression; however, modifications to these components and to the genome occur as a result of effects from the individual's environment. It is assumed here that a complete description of the effects of the envirome on the genome, transcriptome, proteome, and metabolome through molecular biomarker assessment, as far as technology will allow, is then a description of the individual's molecular phenotype from which all other higher order characteristics such as disease diagnosis and therapeutic outcomes are derived. The envirome includes the individual's geographical location, lifestyle, and the total ensemble of environments, both current and in earlier life, that may affect the occurrence of disease and drug therapy. This includes predisposing risk factors like type of neighborhood, family income, intrauterine exposure to teratogens such as maternal cocaine misuse, and exposure to radiation, and provoking environmental factors that can act as triggers of various disorders, such as crises in personal relationships and social stressors [55]. The consistently accurate assessment of such diverse factors is a true challenge to the success of pharmacogenomics and personalized medicine.

3. *Disease Diagnosis.* Individuals are differentially predisposed and susceptible to disease and disease will progress differently depending on personal factors. The application of molecular markers to the pathology and etiology of disease complicates the already difficult process of accurate disease diagnosis and treatment. The application of laboratory tests to disease diagnosis is not a new concept; however, the application of molecular biomarkers provides a plethora of new tests that must be considered in a description of the pathology and etiology of the disease. Personalized medicine will not be successful without a consistent, accurate application of the information in the genome, transcriptome, proteome, and metabolome to the disease diagnosis.

4. *Medication.* The *therapeutic drug index* (TI) is equal to the effective dose (drug efficacy)/toxic dose (drug toxicity) and is an indicator of the drug's *therapeutic window* (the smaller the TI, the wider the window). Obviously, the ability to robustly predict efficacy and toxicity will go a long way to improving the drug development process in terms of success versus dollars spent. The development of specific, reliable biomarkers that can predict efficacy and toxicity should reduce the time, size, and cost of clinical trials. The identification of biomarkers that predict a response to drug therapy will provide tests for predicting an individual patient's response to a particular drug therapy. The major problem with biomarkers is that their usefulness falls off rapidly with their lack of accessibility. Clinically useful biomarkers should be obtained in the least invasive manner possible [51].

The importance of molecular biomarkers for the selection and dose of therapeutic agent is where we began this chapter. The application of genome biomarkers in the form of SNPs in the genes for metabolic enzymes provides the success that has been attained in pharmacogenetics to date, by predicting the efficacy and toxicity

of therapeutic agents (Table 1). Pharmacogenomics expands this early approach to include therapeutic drug pathways that are made up of the proteins involved in the pharmacokinetics and pharmacodynamics of drug therapy (Fig. 6). Considerable expertise will be required to obtain and apply the information from molecular biomarkers to the choice and delivery of the correct therapeutic agent to the patient at the proper dose at the right time. Assessing the outcome of therapy will also be critical to determining the correct application of molecular biomarkers. Pharmacists, with specialized knowledge and expertise, will be critical participants in the health care arena and the resulting partnership between physicians and pharmacists will create an environment for optimal application of personalized medicine.

5. *Health Care.* The United States Pharmacopoeial Convention held a meeting in 1992 to discuss the future of health care technologies, focusing on the design of an optimal system of health care delivery and what that system might mean to drug standards-setting and drug information development [56]. The participants very perceptively compiled a list of five forces that would be driving health care technology development between 1992 and 2020: (1) globalization, (2) cost containment and outcomes, (3) advanced personalized therapies, (4) advances in information infrastructure, and (5) the predict-and-manage paradigm in health care, all of which are beginning to surface in the development of personalized medicine.

The international effort to complete the Human Genome Project has now led to international efforts to study gene function and will only contribute more to the international harmonization of drug/technology regulation and standards. The cost of tests for molecular biomarkers is at odds with the potential that these tests have to reduce the cost of health care. Yet as technology progresses, these tests will be well within the realm of routine use. The application of genomic data to drug therapy has already emphasized the advance of personalized medicine in this genomic age and this will only become a greater part of health care as the developing technologies force novel approaches to genetic counseling, patient education, evaluating risk profiles, medical decision making, monitoring treatment, privacy and regulatory issues, and patient empowerment [39]. The individual's access to information through home computers and the Internet has already changed the way we shop and entertain ourselves and will also bring tremendous advances in the individual's access to medical information, thus empowering people to take responsibility for their own health care. The advances in biomedical knowledge and information systems will shift the emphasis of health care from the reactive response to illness to the proactive minimization of unnecessary morbidity over an individual's lifetime [56].

6. *Information Management.* The development of information infrastructure capable of managing the genomic, transcriptomic, proteomic, and metabolomic data that will arise through personalized medical therapy as well as managing the environmental influences on each of these groups of data is critical to the success of personalized medicine. Personalized medicine is all about connecting patient-specific information with knowledge-based information [39] and requires a management system that can be readily accessed and interpreted by health care professionals. The success of personalized medicine depends on information management systems (both bioinformatic and healthinformatic) that are rapid and user friendly. This also

**TABLE 1 Specific Examples of the Application of Allelic Variation in Drug Therapy**

| Gene | Action | Drug | Therapy | Efficacy/Toxicity |
|---|---|---|---|---|
| Cytochrome P450 2D6 | Drug metabolism | Codeine | Pain | Efficacy |
| Cytochrome P450 2C9 | Drug metabolism | Warfarin | Coagulation | Toxicity (hemorrhage) |
| Thiopurine *S*-methyltransferase | Drug metabolism | Mercaptopurine | Acute lymphoblastic leukemia | Toxicity (myelotoxic) |
| HER2/neu | Drug target | Herceptin | Breast cancer | Efficacy |
| bcr/abl | Drug target | Gleevec | Chronic myelogenous leukemia | Efficacy |
| EFGR | Drug target | Erbitux | Colon cancer | Efficacy |
| Apolipoprotein E-4 | Marker | Tacrine | Alzheimer disease | Efficacy |
| Cholesteryl ester transferase | Marker | Statins | Atherosclerosis | Efficacy |
| ATP binding cassette B1 | Drug transport | Saqinavir, indinavir, ritonavir; daunorubicin, etoposide | HIV; leukemia | Efficacy |
| UDP-glucuronosyltransferase 1A1 | Drug metabolism | Irinotecan | Cancer | Toxicity (myelotoxic) |
| *N*-acetyltransferase 2 | Drug metabolism | Isoniazid | Tuberculosis | Toxicity (hepatotoxic) |
| Pseudocholinesterase | Drug metabolism | Suxamethonium | Muscle relaxation during surgery | Toxicity (prolonged apnea) |

means the storage and transmission of personal data, which brings ethical, social, and legal issues to bear on health care development.

## 4 MOLECULAR PROFILING TECHNOLOGIES

### 4.1 Genotyping Technologies—Genomics

*Genomics* [53, 57–60] is defined as the study of the full complement of genetic information, both coding and noncoding, in an organism (genome).

Most genotyping assays currently used to screen for polymorphic variations involve amplification by the polymerase chain reaction (PCR) and specific methods for discriminating and detecting the amplified product. SNP detection can involve either scanning DNA sequences for previously unknown polymorphisms or screening individuals for known polymorphisms. The latter is commonly known as genotyping (Fig. 9).

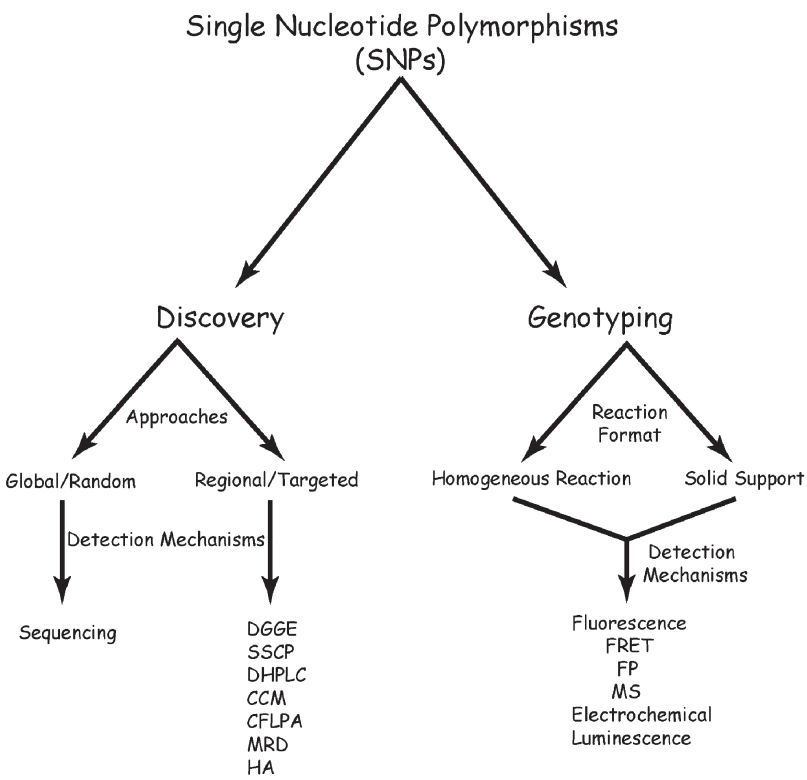


**FIGURE 9** SNP analysis technologies. DGGE, denaturing gradient gel electrophoresis; SSCP, single strand conformation polymorphism; DHPLC, denaturing high performance chromatography; CCM, chemical cleavage of mismatch; CFLPA, cleavage fragment length polymorphism analysis; MRD, mismatch repair detection; HA, heteroduplex analysis; FRET, fluorescence resonance energy transfer; FP, fluorescence polarization; MS, mass spectrometry.

***DNA Scanning for New Polymorphisms*** DNA scanning for new polymorphisms is further subcategorized into a global, or random, approach and the regional, or targeted, approach [58]. The random global approach is typified by the variation in restriction enzyme site sequences resulting in restriction fragment length polymorphisms that are detected by Southern blotting. However, whole genome shotgun sequencing was the best technique for global polymorphism detection and discovery [60, 61].

While the global, random approach to SNP discovery seeks high density markers throughout the genome, the targeted, regional approach focuses on identification of all SNPs in particular candidate genes [62, 63]. The technologies developed for the targeted approach exploit the difference between mismatched heteroduplex DNA and the perfectly matched homoduplex DNA and include such techniques as denaturing gradient gel electrophoresis (DGGE), single strand conformation polymorphism (SSCP), and denaturing high performance liquid chromatography (DHPLC). These techniques will detect the presence of sequence variation but will not determine the base change; sequencing of the product is required. These and many other techniques for targeted approach SNP analysis are described in a more specialized article [58].

***Genotyping Known SNPs*** The application of molecular biomarkers such as single nucleotide polymorphisms toward *in vitro* diagnostic assays is developing into a critical technology in disease therapy [57, 58, 64–66]. These molecular biomarkers will inform physicians about subtypes of disease at the molecular level that require differential treatment. They will provide the pharmacist with information for choice of the best therapeutic methodology for effectively managing the disease and can indicate which individual patients are at the highest risk of experiencing either an adverse reaction or a lack of response to a given drug, thereby identifying those patients likely to respond optimally to a given therapy [53].

Selection of a genotype assay is critical if one is to advance from a preclinical to a clinical setting. The assay must be cheap, high throughput (highly multiplexed), easily and reliably developed from sequence information, robust (specific genotyping of all SNPs under similar reaction conditions), automated, accurate, reproducible, homogeneous (self-contained assay requiring minimal handling for sample analysis), small footprint for the hardware (must easily fit on a lab bench), reliable and easy to use software, and simple (or not complex) in terms of performing the assay (designing primers and probes) as well as interpreting the results.

The general genotyping procedure consists of PCR amplification of the region of interest, allele discrimination, and detection of the discrimination product. Allele discrimination generally falls into one of five techniques illustrated in Fig. 10 [57, 58, 64–67]: differential hybridization, allele-specific PCR, oligonucleotide ligation assay (OLA), nuclease cleavage (flap cleavage, 5′-nuclease cleavage, restriction enzyme cleavage), and minisequencing or primer extension. Each detection reaction may be carried out in homogeneous reaction medium, on a solid support, or one may use technology that will carry out the detection reaction in a homogeneous medium and then utilize a solid support to capture the product.

*Amplification* Most methods used for genotyping SNPs in large diploid genomes depend on PCR amplification of the genomic regions that span the

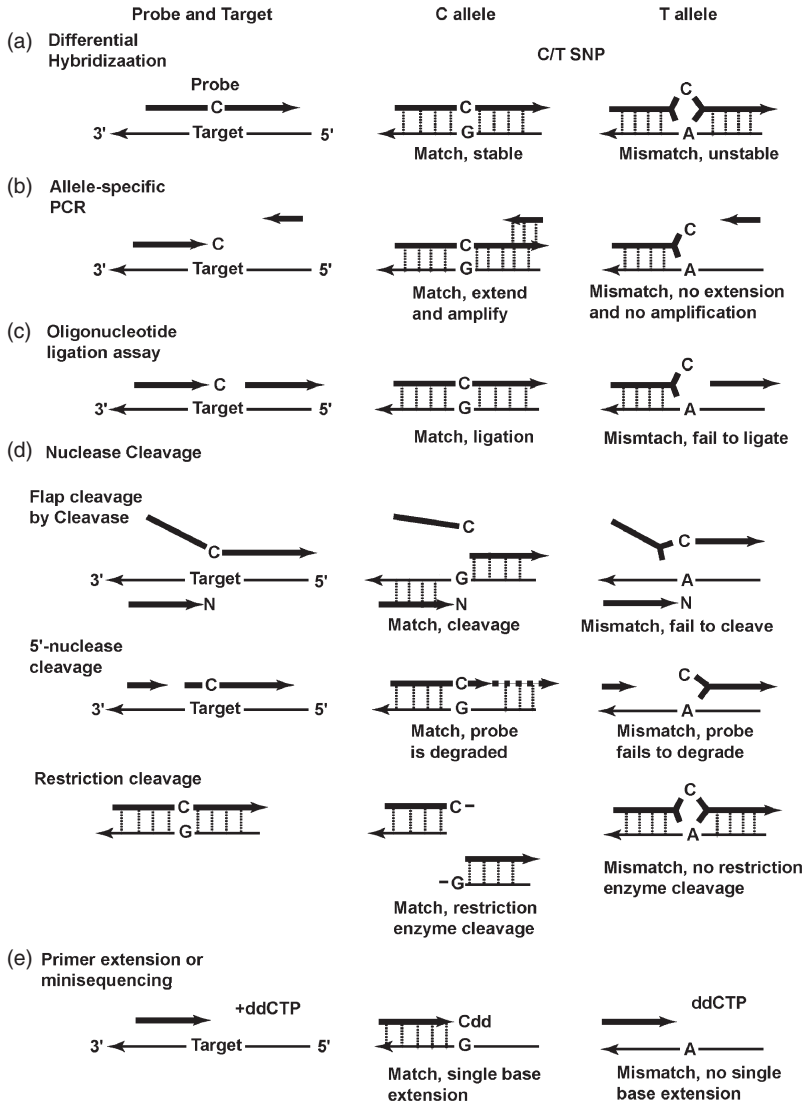


**FIGURE 10** Allele discrimination strategies for the detection of the C allele of a C/T transition: a summary of the five main biochemical reaction principles that underlie the SNP genotyping technologies. The bold arrows indicate the probes while the thinner arrows indicate the target. (Adapted from Refs. 64 and 66.)

SNPs before the actual genotyping or detection reaction is implemented [68]. The PCR provides the required sensitivity and specificity for distinguishing between heterozygous and homozygous SNP genotypes in large, complex genomes and the difficulties in designing and carrying out multiplex PCR reactions is a critical factor limiting the throughput of the current SNP-genotyping assays [66].

*Allele Discrimination* The allele discrimination (Fig. 10 and Table 2) component of the genotyping technology can be divided into five general categories based on the biochemical reaction principles that underlie the technology: differential hybridization, allele-specific PCR (AS-PCR), the oligonucleotide ligation assay (OLA), the cleavage assays, and the primer extension or minisequencing assay.

1. Differential hybridization using allele-specific oligonucleotides (ASOs) (Fig. 10a) was first demonstrated in 1979 [127] and used in 1983 [128] to detect the sickle-cell mutation in the β-globin gene using Southern blot hybridization to genomic DNA. Differential hybridization relies on the differential interaction between the amplified target (amplicon) and a specific labeled oligonucleotide (probe).

The use of ASOs in solution as hybridization probes is considered homogeneous because no purification steps are required and monitoring is done in real time during the amplification process. The most common homogeneous ASO hybridization techniques employ SNP detection in real time during the PCR amplification process in homogeneous solution using fluorescence detection. Applied Biosystems' TaqMan (http://www.appliedbiosystems.com/) and Public Health Research Institute's (PHRI) Molecular Beacons (http://www.molecular-beacons.org/) provide technology for the real-time PCR detection of SNPs. The Molecular Beacon probes have a tendency to adopt a stem–loop structure that destabilizes mismatched hybrids, thereby increasing their power of allele discrimination compared with linear ASO probes [74, 129, 130]. Other techniques for improving the discrimination power of homogeneous reactions include the TaqMan probes modified with minor groove-binder molecules that increase their affinity for the target [131] and the use of two probes, each labeled with a different reporter fluorophore, which allows both SNP alleles to be detected in a single tube [66].

The homogeneous reactions are robust, flexible, and not labor intensive. The thermal stability of a hybrid between an ASO probe and its SNP-containing target sequence is determined by the stringency of the reaction conditions as well as by the nucleotide sequence that flanks the SNP and the secondary structure of the target sequence. Prediction of the reaction conditions or the sequence of the ASO probe that will allow the optimal distinction between two alleles that differ at a single nucleotide position using ASO hybridization is extremely difficult. These parameters must be established separately for each SNP. As a result, there is no single set of reaction conditions that would be optimal for genotyping all SNPs, which makes the design of multiplex assays based on hybridization with ASO probes in solution an almost impossible task [66].

A common technique to circumvent some of the design problems of the homogeneous reaction is to use microarrays where either the probe or target oligonucleotide is immobilized on a solid support. There is some inconsistency in nomenclature of the probe and target in the microarray technology. Some authors name the target as the immobilized oligonucleotide and the probe as the labeled oligonucleotide, while others give the impression that the probe is immobilized on the solid support while the target is hybridized to the probe array [114, 132, 133]. I will use the latter definition (Fig. 11).

The reactions on solid support are carried out on glass slides, silicon chips, latex beads, or on the walls of a microtiter plate. The dot blot [134] is a form of differential hybridization where the probe oligonucleotide is labeled and the target or amplicon

**TABLE 2 Genotyping Technologies**

| Technique[a] | Mechanism of Allele Discrimination | Reaction Format | Detection Mechanism | References |
|---|---|---|---|---|
| Reverse dot blot | Hybridization | Solid support | Colorimetric | 69 |
| Microarray | Hybridization | Solid support | Fluorescence | 12, 70, 71 |
| DASH | Hybridization | Solid support | Fluorescence | 72, 73 |
| Molecular Beacons | Hybridization | Homogeneous reaction | FRET | 74, 75 |
| 5′-Nuclease cleavage (TaqMan) | Hybridization | Homogeneous | FRET or FP | 76, 77 |
| Cleavase (Invader assay) | Hybridization | Homogeneous reaction | FRET, FP, MS | 78–82 |
| Biosensor Microchip | Hybridization | Solid support | Electrochemical and fluorescence | 83–89 |
| AS-PCR | Primer extension | Homogeneous reaction | Fluorescence or FRET | 90–95 |
| AS-PE | Primer extension | Solid support | Fluorescence | 96 |
| APEX | Primer extension | Solid support | Fluorescence | 97–100 |
| FP-TDI | Primer extension | Homogeneous reaction | FRET or FP | 101–104 |
| MALDI-TOF MS | Primer extension | Homogeneous reaction | MS | 105–108 |
| Pyrosequencing | Primer extension | Homogeneous reaction or solid support | Luminescence | 109–113 |
| BioBeads and TagArrays | Primer extension | Homogeneous reaction and solid support capture | Fluorescence | 114–117 |
| OLA | Ligation | Homogeneous reaction and solid support | Colorimetric and fluorescence | 118–120 |
| RCA | Ligation | Solid support | Fluorescence | 121–124 |
| Closed tube | Ligation | Homogeneous reaction | FRET | 125 |
| Microsphere/ microarray ligation | Ligation | Homogeneous reaction | Fluorescence | 115, 126 |

[a]Abbreviations: DASH, dynamic allele-specific hybridization; AS-PCR, allele-specific polymerase chain reaction; AS-PE, allele-specific primer extension; APEX, arrayed primer extension; FP-TDI, fluorescence polarization template directed dye terminator incorporation; MALDI-TOF-MS, matrix assisted laser desorption/ionization time-of-flight mass spectrometry; OLA, oligonucleotide ligation assay; RCA, rolling circle amplification.

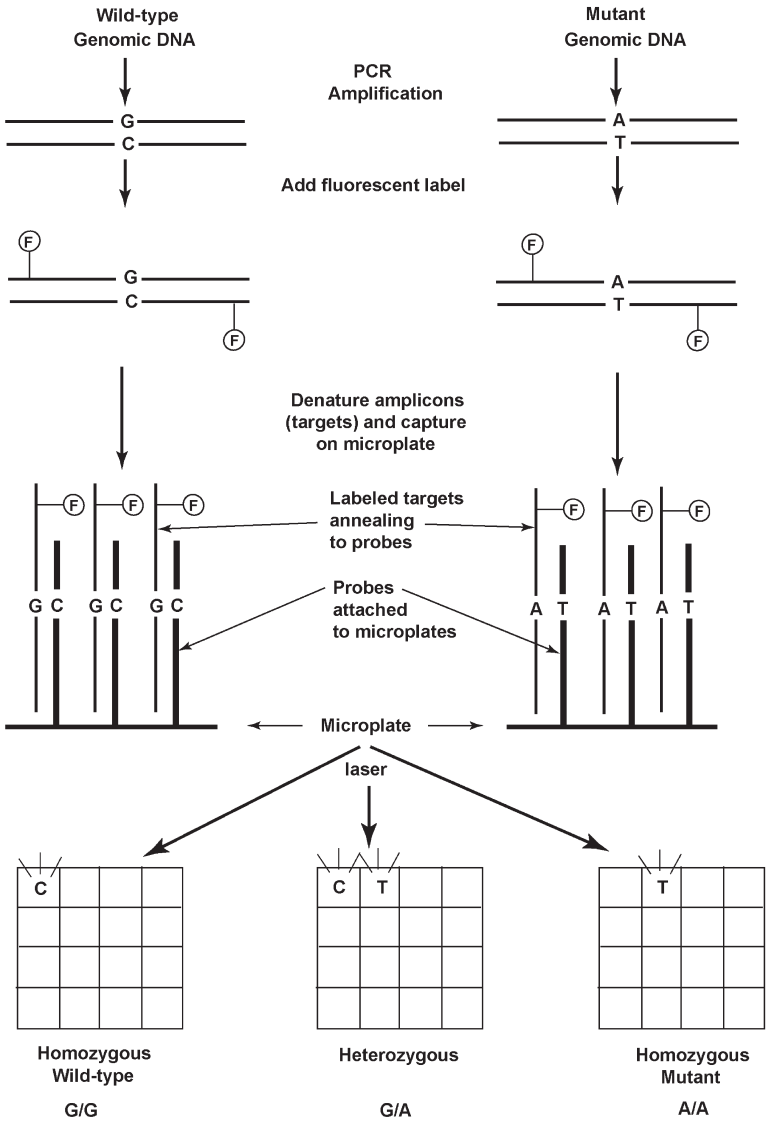


**FIGURE 11** Schematic representation of the reverse blot microarray technology. The amplicons or "targets" are represented by thin lines while the synthetic "probes" covalently linked to the microplate are represented by thick lines. Note in this figure that the probes select for one strand of the fluorescently labeled amplified targets.

is immobilized on a solid support. Gene chips are a form of reverse dot blot where the amplicon is fluorescently tagged and hybridized to an immobilized array of oligonucleotide probes (Fig. 11) [69]. Because the oligonucleotide sequences are SNP specific, each chip has a unique design, and the cost and effort involved in the manufacture of unique chips can be prohibitive. The identity of an oligonucleotide at a particular location on the microarray is known and the genotypes are inferred by determining which immobilized oligonucleotide is associated with a positive signal (Fig. 11). The GeneChip from Affymetrix (http://www.affymetrix.com/index.affx)

uses this principle. The probes are synthesized on the glass chip by photolithography or solid-phase DNA synthesis [135].

The major advantage of performing genotyping reactions on solid supports is that many labeled targets can be probed at the same time. Along with saving time and reagents, performing numerous reactions in parallel also decreases the probability of sample/result mix-ups. The main drawback of performing genotyping reactions on solid support is that design of the oligonucleotide arrays and optimization of the multiplex reactions require substantial investment of time and capital [58].

Other approaches using microarray ASO hybridization include dynamic allele-specific hybridization (DASH) [72, 136] and the use of electric field strength instead of temperature to denature the probe–target hybrids [83] as well as the use of peptide nucleic acids (PNAs) [137, 138] or locked nucleic acids (LNAs) [139, 140] with high affinity for complementary DNA.

2. Allele-specific PCR (AS-PCR) (Fig. 10b) uses PCR primers with the 3′-end complementary to either of the nucleotides of a SNP in combination with a common reverse PCR primer to selectively amplify the SNP alleles [141]. The simplest approach for monitoring the formation of allele-specific PCR products using a homogeneous assay format is to include a fluorescent dye such as SYBR Green that intercalates with the double-stranded PCR products in the reaction mixture [92]. Other detection techniques include fluorescence resonance energy transfer (FRET) and fluorescence [91, 94, 95]. Roche's LightCycler™ (http://www.roche-applied-science.com/sis/rtpcr/lightcycler/), Perkin Elmer's AlphaScreen (http://las.perkinelmer.com/), and adaptations of the Molecular Beacon are three techniques using this technology. In general, a limitation of all the variations of AS-PCR is similar to the ASO differential hybridization in solution in that the reaction conditions or primer design for selective allele amplification must be optimized empirically for each SNP. Therefore, homogeneous allele-specific PCR methods like TaqMan, Molecular Beacon, and LightCycler assay are best suited for the analysis of a limited number of SNPs (10–100) in large sample collections [66].

3. The oligonucleotide ligation assay (OLA) (Fig. 10c) uses DNA ligase to discriminate between mismatches at the ligation site of two adjacent hybridized oligonucleotides as the basis for SNP genotyping (Fig. 12).

One of the hybridized allele-specific nucleotides carries a donor dye while the other carries an acceptor dye [125, 142, 143]. In order to detect both alleles, one donor dye (e.g., FAM or 5-carboxy-fluorescine) and two acceptor dyes (e.g., ROX or 6-carboxy-X-rhodamine and TAMRA or *N*,*N*,*N*′, *N*′-tetramethyl-6-carboxyrhodamine) are required (Fig. 12). A 5′-donor dye-labeled common probe is designed to terminate one base immediately upstream of the polymorphic site. The 5′-phosphorylated probes carrying the 3′-acceptor dyes are designed to anneal with the allelic bases at the 5′-end. The PCR primers are designed to anneal with high melting temperatures while the 5′- and 3′-labeled probes are designed to have a matching but low melting temperature. The reaction is carried out with a thermostable DNA polymerase with no 5′-nuclease activity and a thermostable DNA ligase. These reactions can be carried out in a single sealed tube since the PCR amplification is carried out at high temperature where the ligation probes will not anneal and be removed from the ligation reaction. When the amplification is complete, the temperature is lowered and the ligation reaction is carried out. Fluorescence reso-

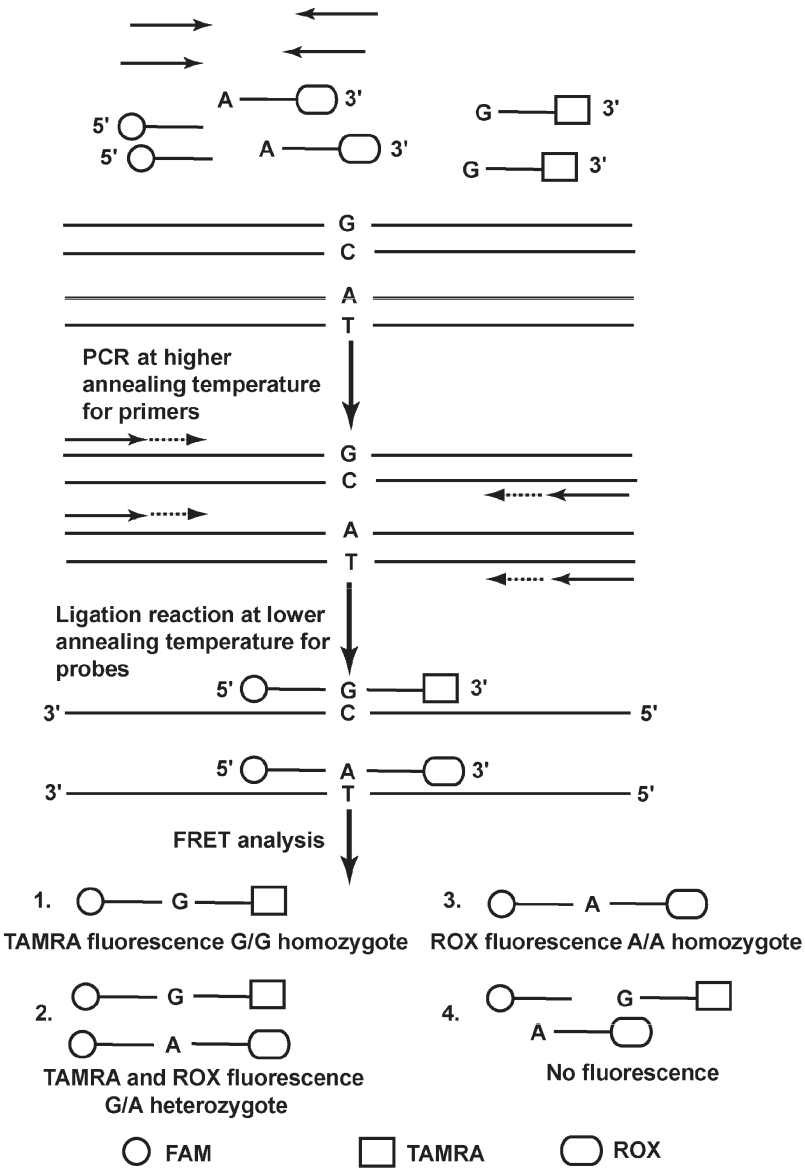


**FIGURE 12** Oligonucleotide ligation assay (OLA).

nance energy transfer (FRET) is measured in real time and the genotypes are determined by measuring the rate of emergence of fluorescence for the two dyes found on the ligation probes.

Variations on this theme include colorimetric detection in microtiter plate wells [118, 120], multiplex detection using fluorescently labeled ligation probes with different electrophoretic mobilities that can be analyzed in a sequencing instrument [119], one immobilized ligation probe [122], immobilized single stem–loop probes [144], padlock probes and the rolling circle amplification [121, 123, 145–147], and homogeneous reaction with tag-capture [126, 148].

4. The cleavage assays (Fig. 10d) include the 5′-nucelase TaqMan assay (discussed in the hybridization category because it is sometimes considered a hybridization technique), the cleavase or Invader assay, and the restriction fragment length polymorphism (RFLP) assay.

The Invader assay from Third Wave Molecular Diagnostics (http://www.twt.com/) is based on a three-dimensional structure formed by the target DNA sequence, a signaling probe with an allele-specific 3′-sequence and a 5′-sequence that is noncomplementary to the target and forms a "flap" (Fig. 10d), and an invader probe oligonucleotide that binds to the 3′-region of the target sequence. When the allele-specific probe is perfectly matched with the SNP region on the target DNA, a complex is formed with the target sequence, the allele-specific probe, and the invader probe that is a substrate for a 5′-endonuclease called cleavase or FLAP endonuclease. Cleavage of this substrate by the enzyme releases the 5′-noncomplementary sequence of the signaling probe that is detected by a variety of methods [78, 80–82, 149–151].

The restriction enzyme cleavage technology is a classical method where the PCR amplified target region is cleaved with restriction enzymes and the resulting fragments are separated by gel electrophoresis. Allelic variations that either create or remove restriction sites can be detected using this technology; however, the need for gel separation of fragments and the requirement that the alleles being analyzed be located at restriction sites are not amenable to high throughput analysis.

5. The primer extension/minisequencing technique (Fig. 10e) has become the reaction principle of choice for high throughput genotyping due to the advantages of using the high specificity of DNA-polymerase to distinguish between SNPs and the fact that primer extension allows genotyping of various SNPs at similar reaction conditions, thereby reducing the problems of optimization and assay design. The versatility of this technique is reflected in the large number of commercially available systems utilizing the minisequencing primer extension technology. These include Beckmann-Coulter's SNPStream [96, 114, 117] (http://www.beckmancoulter.com), Biotage's Pyrosequencing [112, 113] (http://www.pyrosequencing.com/), Sequenom's MassArray [108] (http://www.sequenom.com/), Asper Biotech's Apex [98, 99] (http://www.asperbio.com/), and PerkinElmer's FP-TDI [104] (http://las.perkinelmer.com/content/snps/fp-tdi.asp#).

The primer extension technique has been augmented with the 5′-tag technology to permit the purification of the extended products via specific tags, which in turn allows for versatile high throughput genotyping (Fig. 13). The genomic DNA is amplified by PCR and single base extension with tagged allele-specific primers and fluorescent-labeled ddNTPs is carried out in solution to detect the SNP. The resulting labeled and tagged fragments are captured by complementary tags on microtiter plates or color-coded beads. Fluorescence resulting from laser excitation of the plates or beads is read by a detector and fed to the computer database for analysis and storage. Most companies providing the single base extension (SBE) tag-capture technology also provide a primer design service so that users can take advantage of multiplexing limited only by the capacity to design effective primers for the amplification and detection steps.

The choice of genotyping technology just described will be governed by the nature of the analysis being carried out. If one is interested in a small number of

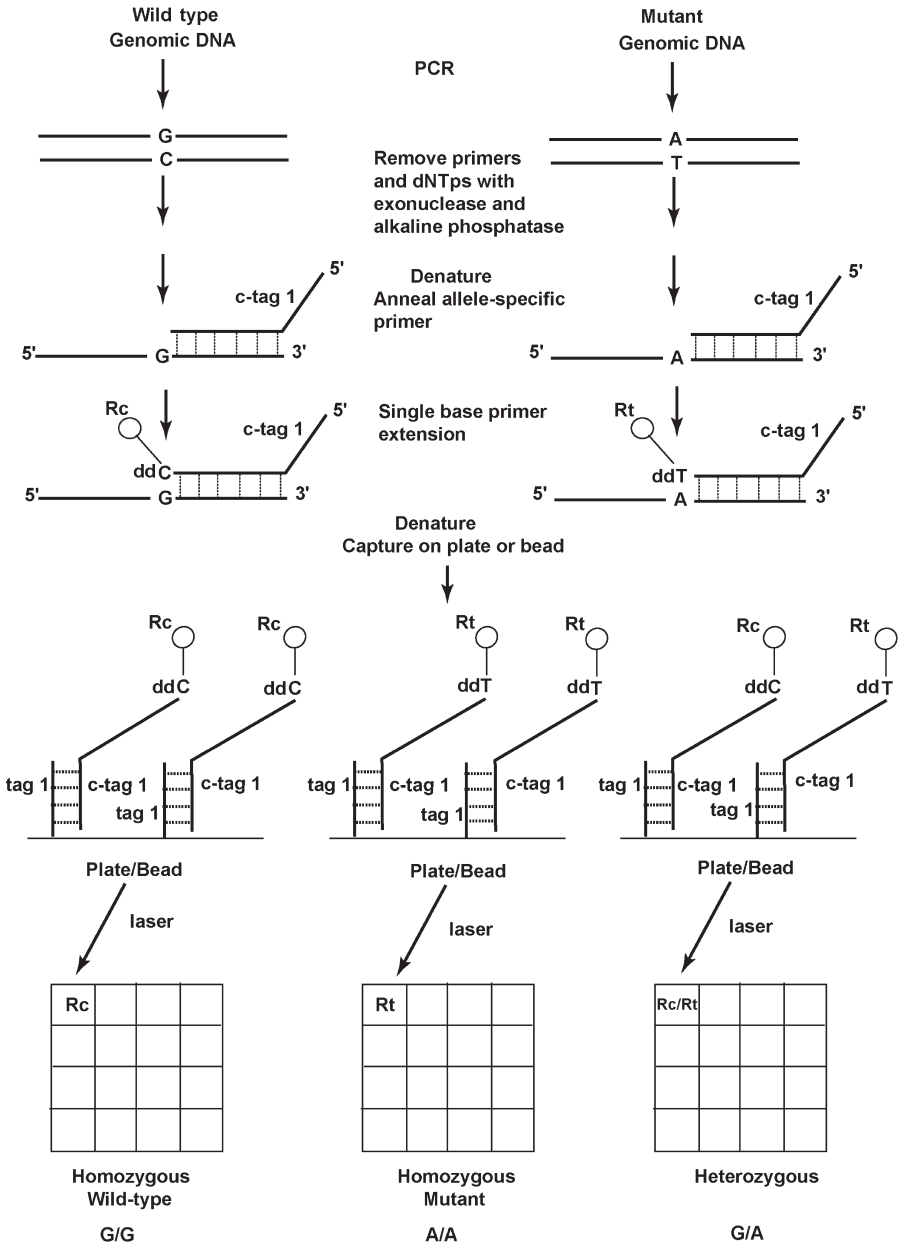


**FIGURE 13** Tag-array and bead technology. R, reporter fluorescent label. The tag 1 is specific for the G/A mutation at a particular position in the gene under investigation. Other tags can be designed for other SNPs to permit multiplexing in the same well of the microtiter plate to produce an "array of arrays" [96].

biomarkers, then PCR analytical techniques like ABI's TaqMan assay or Roche's LightCycler assay are useful. When more complex assays that can analyze large populations for dozens of polymorphisms in a single multiplex analysis are required, the 12- and 48-plex technology of Beckman-Coulter's SNPStream microtiter plate tag-array technology, the bead-based tag-array technology of Luminex and Illumina, the Affymetrix oligonucleotide arrays, or ABI's PinPoint and Sequenom's MassExtend mass spectrometry technology are more accommodating.

*Detection* The different formats for detection [64] of the allele discrimination product include gel electrophoresis, microtiter plate or closed tube, and the microarray bead or chip format. Historically, the allele discrimination product was detected using radioisotopic methods but the technique has been almost completely replaced by monitoring light emitted by the product (fluorescence, luminescence, fluorescence resonance energy transfer, fluorescence polarization; Fig. 14 and Table 2) or a change in mass (mass spectrometry) or electrochemical conductivity of the product (Table 2).

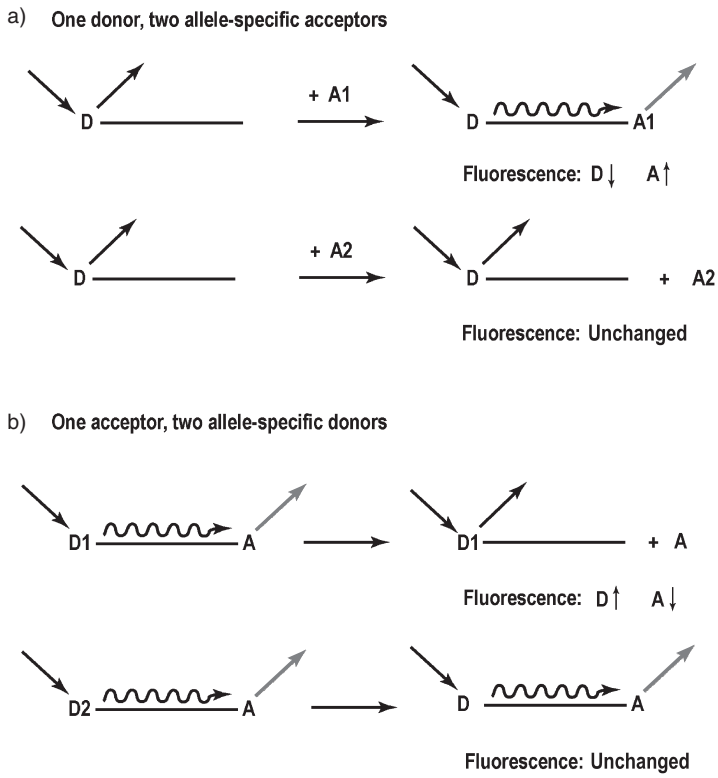


**FIGURE 14** Fluorescence-based techniques for genotype scoring. The alleles can be detected using allele-specific acceptors or allele-specific donors: (a) two allele-specific acceptors with one donor, examples of which occur in minisequencing and the oligonucleotide ligation assay; and (b) two allele-specific donors with one acceptor, examples of which occur in the TaqMan assay, the Invader assay, and Molecular Beacons. (Adapted from Ref. 64.)

### 4.2 Gene Expression Profiling—Transcriptomics

*Transcriptomics* is defined as the study of the full complement of RNA in the cell, including mRNA transcripts (transcriptome).

Technologies that examine gene expression profiles are at the core of functional genomics and differential gene expression (DGE) is at the center of the functional genomics technology. Genes that are expressed or transcribed from genomic DNA are collectively known as the "transcriptome" and are considered to be a major determinant of cellular phenotype and function. Methods of expression profiling are classified into two types: open and closed. Open systems do not require any advance knowledge of the sequence of the genome being examined. Closed systems require some advance knowledge of the genes being examined and usually involve the use of gene chip technology.

***Closed Gene Expression Profiling—Microarray Technologies*** The most common methods of closed system analyses [152] are oligonucleotide or cDNA array hybridization technologies and quantitative polymerase chain reaction. The arrays detect mRNA levels that can be confirmed by RTQ-PCR (real-time quantitative polymerase chain reaction) or *in situ* hybridization [51]. Closed system profiling technologies are used for analysis of a known subset of genes, where there is a solid understanding of such global genetic complexities as splice variants, polymorphisms, and RNA editing [153].

cDNA arrays (Fig. 15) are prepared by spotting gene-specific PCR products including full-length cDNAs, collections of partially sequenced cDNAs (or expressed sequence tags—ESTs), or randomly chosen cDNAs from any library of interest onto a matrix consisting of glass, silicon, nylon or nitrocellulose membranes, gels, or beads. Glass is particularly suitable because of its durable and nonporous properties and low background fluorescence. The glass-based matrix is coated with polylysine, amino silanes, or amino-reactive silanes to aid in the attachment of the cDNA sample. PCR products are spotted onto the matrix by robots through contact printing or noncontact piezo or ink-jet devices. The probe DNA is cross-linked to the matrix by UV irradiation and rendered single stranded by heat or alkali treatment. The array is used to analyze mRNA expression. Total mRNA from both a test and reference sample is reverse transcribed to produce cDNA, during which the product is individually labeled with either fluorescent Cye3- or Cye5-dUTP. The fluorescent targets are pooled and allowed to hybridize under stringent conditions to the clones on the array. Laser excitation of the incorporated targets yields an emission with a characteristic spectra, which is measured using a scanning confocal laser microscope. The fluorescence analysis of the test and reference sample allows comparison of the gene expression of one to the other, thus identifying up- and downregulated genes. The rather large amount of RNA required per hybridization (50–200 μg of target per array) is a limiting factor in the use of this technology [132]. Whole genome high density tiling arrays (WGAs) are a variation of the microarray technology that provides a universal data capture platform for a variety of genomic information [154].

Data collection is obviously very complex and computer algorithms for analyzing the data fall into four methods: hierarchic clustering, self-organizing maps, multidimensional scaling, and pathway associations. Cluster analysis of microarrays consists

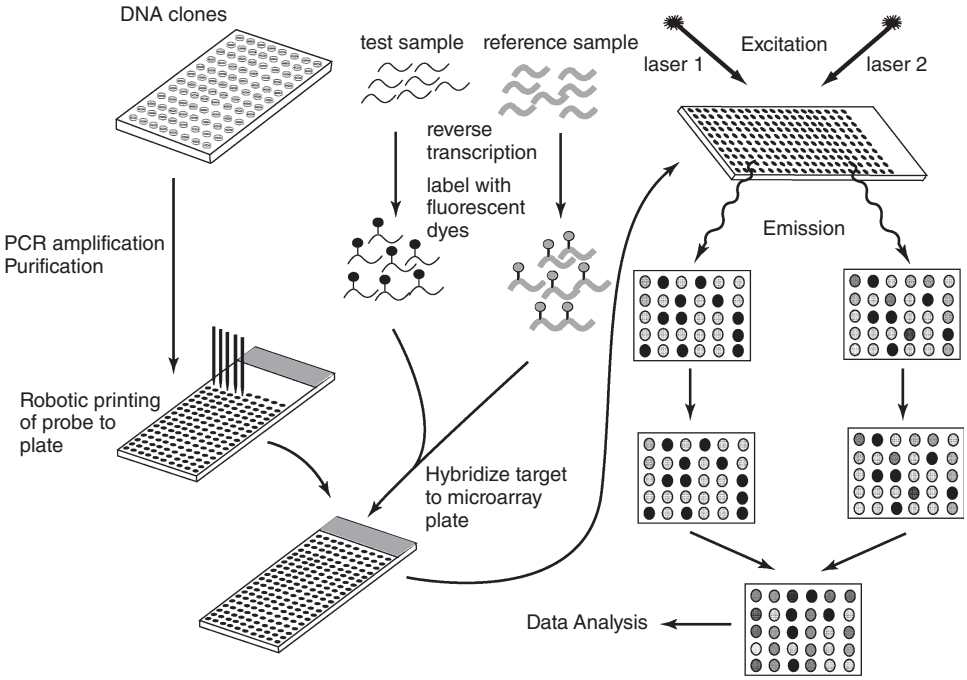


**FIGURE 15** Schematic representation of a cDNA microarray. Templates for genes of interest are obtained from DNA clones and amplified by PCR. Following the purification and quality control, aliquots are printed on coated glass microscope slides using a computer-controlled, high speed robot. The array is used to analyze mRNA expression. Total mRNA from both a test and reference sample is fluorescently labeled with either Cye3- or Cye5-dUTP using a single round of reverse transcription. The fluorescent targets are pooled and allowed to hybridize under stringent conditions to the clones on the array. Laser excitation of the incorporated targets yields an emission with a characteristic spectra, which is measured using a scanning confocal laser microscope. (From Ref. 132 with permission of the author and Nature Publishing Group.)

of arranging the hybridization data into a cluster order based on color-coded intensities, which in turn provides clues as to which genes are being regulated in groups [155].

Comparisons of test and control expression profiles provide useful insights into the molecular pathologies of a variety of diseases and drug therapies [49]. It will not, however, deliver the kind of intimate understanding of the highly interrelated control circuitry that is necessary to achieve true understanding of genome function [132]. Duggan et al. [132] have suggested that we should reconsider our perception of transcriptional control as a simple on–off switch to a model, whereby control is analogous to a highly gated logic circuit, where numerous, often contradictory, inputs are summed to produce a final response.

There has been considerable enthusiasm in the scientific community in promoting the application of DNA microarray technology for the analysis of genetic variation aimed at identifying molecular biomarkers for application to personalized medicine and drug development. The U.S. Food and Drug Administration has issued

a draft regarding guidance on the submission of pharmacogenomic data for drug development by the pharmaceutical industry (http://www.fda.gov/cber/gdlns/pharmdtasub.pdf). However, variations in the technical, instrumental, computational, and interpretative aspects of microarray technologies are causing serious concerns and the lack of reproducibility and accuracy of microarray studies within and between facilities are just some of the challenges and pitfalls facing the microarray community and regulatory agencies before microarray data can be reliably applied to support regulatory decision making. Problems in intralaboratory data quality have been observed, and the microarray community and regulatory agencies must work together to identify critical factors affecting data quality, and to optimize and standardize microarray procedures to make certain that biologic interpretation and decision making are based on reliable data [156]. These fundamental issues must be adequately addressed before microarray technology can be transformed from a research tool to clinical practice.

***Open Gene Expression Profiling*** The open systems for detecting differential gene expression [153] are distinct from the closed systems in that no preexisting biological or sequence information is necessary. These technologies will not be discussed in detail here since they are not extensively used in the molecular biomarker field of pharmacogenomics. A brief description of some of the major open system technologies is given in Table 3 with appropriate references.

### 4.3 Protein Analysis—Proteomics

*Proteomics* [166–168] is defined as the study of the total protein complement of a genome, or the complete set of proteins expressed by a cell, tissue, or organism (proteome).

The detection of an open reading frame (ORF) is an indication of the presence of a gene but does not indicate transcription, RNA editing, or translation. Verification of a gene product by proteomic technology is important to confirm the above as well as determine posttranslational modifications and the presence of protein isoforms. Transcriptome analysis does not indicate alteration in protein levels by proteolysis, recycling, and sequestration of proteins; therefore, proteomics is important in determining the levels of proteins in a cell as well as protein localization, protein expression, and protein–protein interaction. Hence, an understanding of the complex relationship between proteins and genes on a genomic scale is vital to understanding protein function in disease and disease therapy.

Proteomics can be divided into three main areas [166]: (1) protein microcharacterization for large scale identification of proteins and their posttranslational modifications; (2) "differential display" proteomics for comparison of protein levels with potential application in a wide range of diseases; and (3) studies of protein–protein interactions using techniques such as protein chips, the mass spectrometry ICAT technology, or the yeast two-hybrid system.

Routine microcharacterization of proteins using gel electrophoresis will only detect the most abundant proteins and a purification strategy is required to reduce the complexity of the proteome and allow visualization of the less abundant proteins. This is especially important when analyzing the signaling and regulatory proteins that are of critical importance to understanding drug action.

**TABLE 3 Comparison of Different Open System Technologies**

| Expression Analysis Technology | Amount of Template Required (μg) | Sensitivity | *N*-fold Difference Detection Limit | Genes Covered (%)[a] | References |
|---|---|---|---|---|---|
| Differential display | 0.5 (mRNA) | NA | NA | 96 | 157 |
| Arbitrarily Primed Polymerase Chain Reaction (AP-PCR) | 0.012 (total RNA) | NA | NA | NA | 158 |
| Serial Analysis of Gene Expression (SAGE) | 5 (total RNA) | <1:10,000 | NA | 92 (of transcripts at ≥3/cell) | 159, 160 |
| Representational Difference Analysis (RDA) | 0.02 (mRNA estimated from 10,000 cells) | <1:300,000 | NA | N/A | 161 |
| Gene Calling® | 0.02 (mRNA) | 1:125,000 | 1.5 | ~95 | 162 |
| Restriction Enzyme Analysis of Differentially Expressed Sequences (READS™) | 10 (total RNA; 100 pg/reaction) | 1:100,000 | NA | 90 | 163 |
| Total Gene Expression Analysis (TOGA) | NA | <1:100,000 | 60–98 | 60% coverage using one enzyme, 98% using four (4 × 256 reactions) | 164 |
| Tandem Arrayed Ligation of Expressed Sequence Tags (TALEST) | 2 (mRNA) | 1:200,000 | NA | 85 | 165 |

[a]For many of the techniques the reported coverage percent is not correlated to mRNA abundance. Coverage of rare transcripts may be considerably lower. NA, information not available.

***Differential Gel Electrophoresis*** Proteins isolated from cells or tissues that have been subjected to different conditions (experimental and control) are separated by their "isoelectric point" in the first dimension and by size in the "second dimension" (Fig. 16) [51, 166, 167, 169, 170]. The technology relies on excising the spots from the gels, proteolytically digesting the protein, and extracting the

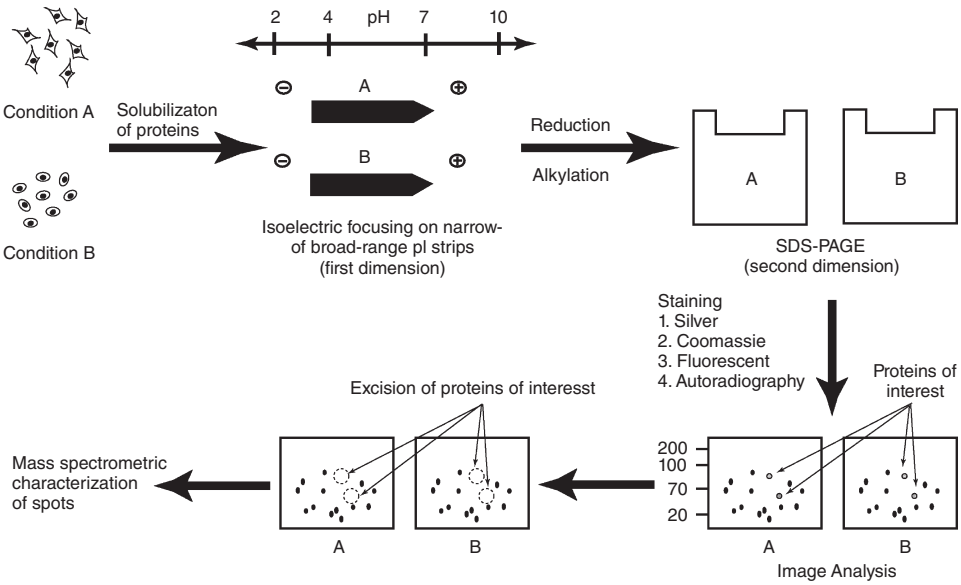


**FIGURE 16** A schematic representation of the two-dimensional (2D) gel approach to protein separation. Cells (or tissues) derived from two different conditions, A and B, are harvested and the proteins solubilized. The crude protein mixture is then applied to a "first dimension" gel strip that separates the proteins based on their isoelectric points. After this step, the strip is subjected to reduction and alkylation and applied to a "second dimension" SDS-PAGE gel where proteins are denatured and separated on the basis of size. The gels are then fixed and the proteins visualized by silver or Coumassie staining or by fluorescence (if the proteins had been fluorescently labeled) or autoradiography (also depending on the nature of the label). After visualization, the resulting protein spots are recorded and quantified. Image analysis requires sophisticated software. Finally, the spots of interest are excised from the gel and subjected to mass spectrometric analysis. (Adapted from Ref. 166 with permission of the author and Nature Publishing Group.)

peptide fragments produced. The peptide fragments produced are analyzed by mass spectrometry or tandem mass spectrometry and the data derived is correlated with database information of protein sequence, genomic sequence, or expressed sequence tags.

***Protein Chips*** The protein chip approach to simplification of the protein mixture utilizes a variety of "bait" molecules that are ligands, such as proteins, peptides, antibodies, antigens, allergens, and small molecules immobilized in high density on modified surfaces to form functional and analytical protein microarrays (Fig. 17) [171–174]. The cell lysate is passed over the surface of the solid support immobilizing the bait molecules to allow it to capture its counterpart in the cell lysate. The most common bait protein used in the protein chip technology is the antibody [175–182]. The protein chip technique has been coupled with a direct matrix assisted laser desorption/ionization (MALDI) mass spectrometric detection of the material bound to the bait protein [183–186].

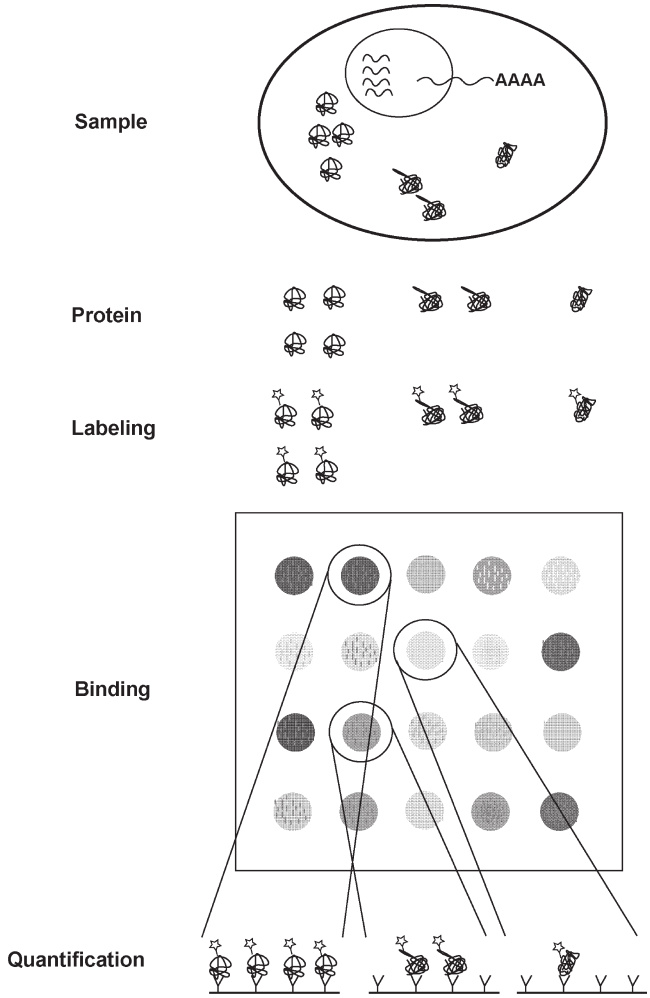


**FIGURE 17** Parallel quantification of proteins by antibody microarray. Different proteins labeled with fluorochromes can be detected in parallel with a microarray-based assay. Specific capture antibodies immobilized in an array interact with their respective target proteins present in the solution. The resulting signal intensity correlates with the amount of captured target. Within each microarray, different kinds of control spots can be included, such as positive and negative control spots and/or internal calibration spots. This will allow accurate signal quantification. (From Ref. 171. Copyright © 2002, with permission from Elsevier.)

***Isotope Coded Affinity Tag (ICAT)*** ICAT is a chromatography–mass spectrometry-based technology where two samples are labeled with tags that differ in their incorporation of a heavy label (deuterium for hydrogen or $^{13}C$ for $^{12}C$) (Fig. 18). The labeling reagent is typically a thiol-specific reagent with an alkyl linker containing either nine $^{12}C$ or nine $^{13}C$ atoms and a biotin moiety for avidin chromatography [187–190].

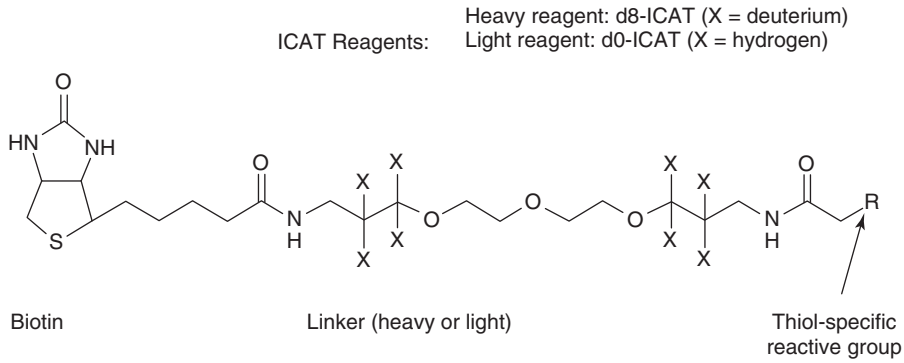


**FIGURE 18** Isotope coded affinity tag labeling reagent. (From Ref. 171 with permission of the author and Nature Publishing Group.)

Two protein mixtures representing two different cell states are treated with the isotopically labeled light and heavy ICAT reagents (Fig. 18), respectively; an ICAT reagent is covalently attached to each cysteinyl residue in every protein. The protein mixtures are combined and proteolyzed to peptides, and ICAT-labeled peptides are isolated utilizing the biotin tag. These peptides are separated by micro-capillary high performance liquid chromatography. A pair of ICAT-labeled peptides is chemically identical and is easily visualized because they essentially coelute and there is an 8 Da mass difference (depending on the tag) measured in a scanning mass spectrometer (four $m/z$ unit difference for a doubly charged ion) (Fig. 19). The ratios of the original amounts of proteins from the two cell states are strictly maintained in the peptide fragments. The relative quantification is determined by the ratio of the peptide pairs. Every other scan is devoted to fragmenting and then recording sequence information about an eluting peptide (tandem mass spectrum). The protein is identified through a computer search of the recorded sequence information against large protein databases [166].

***Yeast Two-Hybrid System*** The yeast two-hybrid system (Fig. 20) [191] is one of the best *in vitro* approaches to mapping protein–protein interactions. Using this technique, the protein of interest or "bait" protein is fused to a protein containing a DNA-binding domain. Proteins being screened for interaction with the bait protein are fused to proteins containing a transcriptional activating domain. If the bait protein binds to the protein being screened, the DNA-binding domain is brought into proximity of the transcriptional activating domain and the activation of transcription of a reporter construct occurs resulting in reporter gene product (Fig. 20).

This technology can be used to identify compounds and peptides that disrupt protein–protein interactions [192] and is amenable to automation and high throughput. One of the main consequences of the forward and reverse techniques shown in Fig. 20 is that once a positive interaction is detected (in either the forward or reverse systems), simple sequencing of the relevant clone identifies the open reading frame and thereby identifies the proteins interacting (forward) or those whose interaction has been disrupted (reverse).

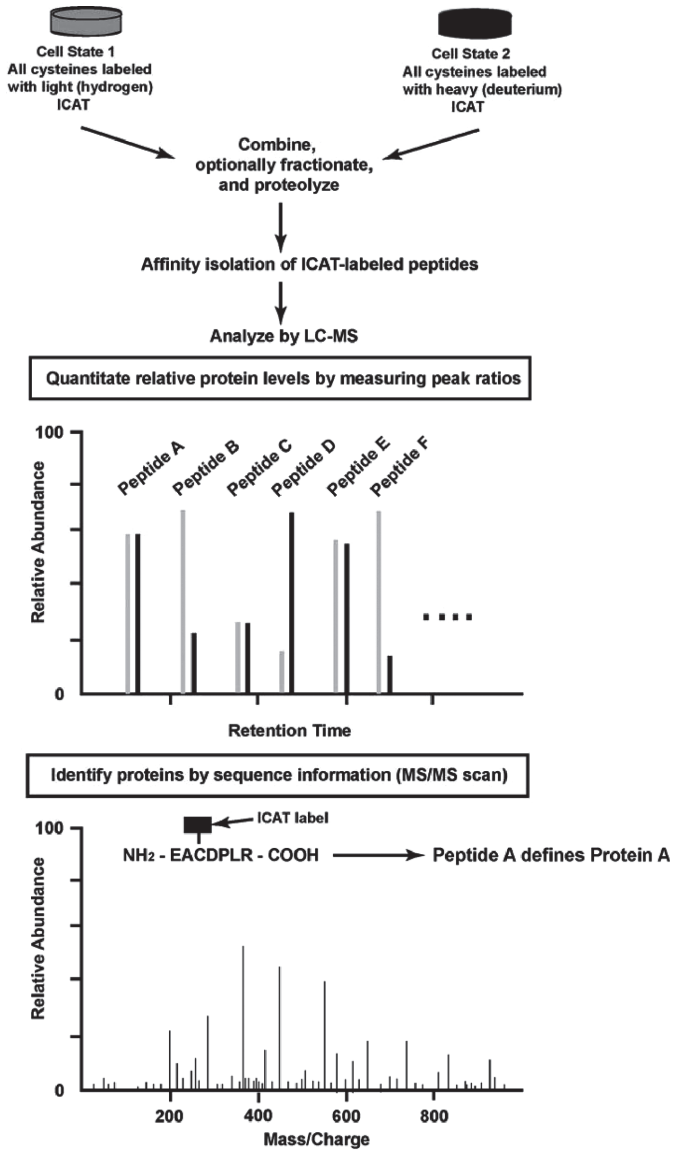


**FIGURE 19** The ICAT strategy for quantifying differential protein expression. (Adapted from Ref. 166 with permission of the author and Nature Publishing Group.)

## 4.4 Metabolomics

*Metabolomics* is defined as the measurement of metabolite concentrations, fluxes, and secretions in cells and tissues in which there is a direct connection between genetic activity (transcriptome), protein activity (proteome), and the metabolic activity itself [193, 194]. There is some multiplicity in the nomenclature in this area and one will also encounter "metabonomics" [195] as well as "metabolic profiling" [196].

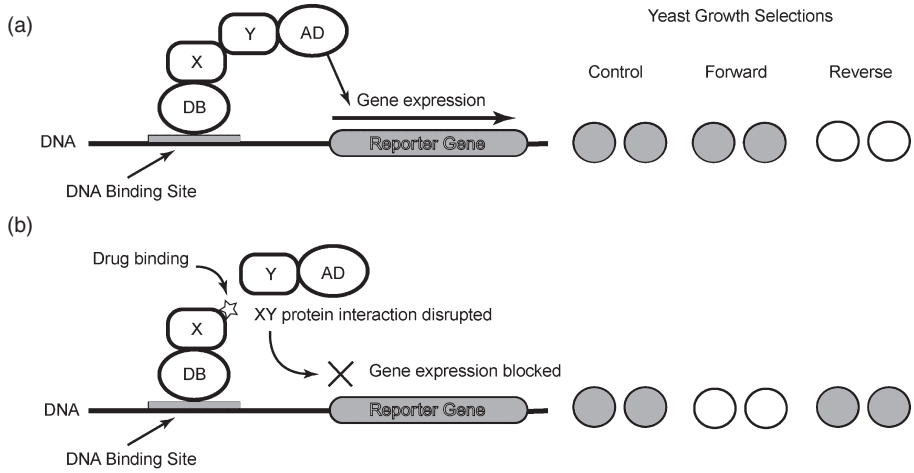


**FIGURE 20** Schematic representations of the way two-hybrid systems work. DB, DNA binding domain; AD, activating domain; X and Y are the proteins under investigation for interaction. The grey and white patches on the right-hand side represent growing and non-growing yeast cells, respectively; under normal conditions (control), the yeast cells grow whether or not a two-hybrid interaction takes place. In the forward two-hybrid selections, potential interactions (a) are identified by the transcriptional activation of a reporter gene required for growth, which confers a selective advantage. In the reverse two-hybrid selections, the interaction activates the expression of a "toxic gene" and so prevention of the interaction (b) provides a selective advantage. (From Ref. 192. Copyright © 1999, with permission from Elsevier.)

In our effort to precisely link the organism's gene function to phenotype, we arrive at the complex analysis of the metabolome. Metabolites have been characterized as "the ultimate gene product" [197] and are at the cross-roads of genotype–phenotype interactions. Under our definition of phenotype [54], metabolomics is considered here to be a form of molecular phenotyping. If we consider the organism's phenotype to be composed of cellular phenotypes and the cellular phenotype to be composed of metabolic networks, then the interrelationship of metabolic pathways is considered the fundamental component of an organism's phenotype [198].

It is the description of the metabolic profile of the organism that is critical to a phenotypic description of that organism's physiology and pathophysiology. Phenotype cannot rely solely on a description dictated by predicted proteins encoded by the genome of the organism but must include metabolic transformations that are stoichiometrically, thermodynamically, and kinetically possible (Fig. 21) [198]. This is the level at which we see the implications of allelic variations occurring in the genome. The identification of an allelic variation in a gene producing an enzyme that has an altered activity is the basis of functional genomics but the question remains: Does the allelic variation alter phenotype and, if so, how? The complex flow of metabolites in an organism has a robustness that arises from the extensive redundancy in metabolic pathways. There are often a number of isozymes that catalyze the same reaction and the alteration of one isozyme at the genetic level may

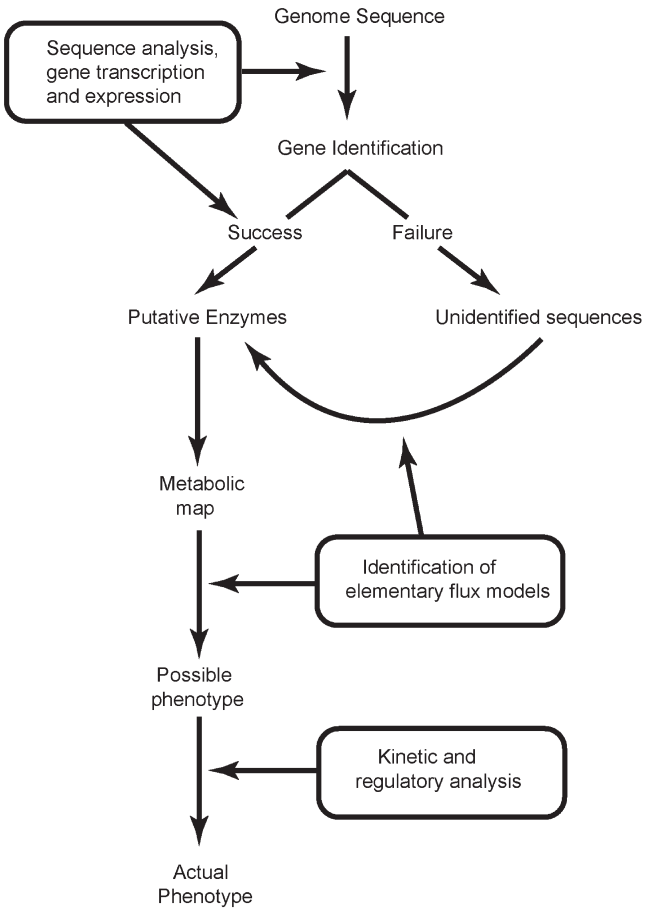


**FIGURE 21** The necessary steps to convert a genome sequence into a phenotypic description include not only the early stages of identifying the genes and investigating their transcription and expression, but also later events that include assembling the enzyme-catalyzed reactions into a metabolic map, determining what kinds of metabolic activity are stoichiometrically possible, and, finally, determining which of these are kinetically possible. Identification of elementary flux modes is especially useful for defining the stoichiometric structure of the map, but it can also aid in identifying the roles of orphan genes remaining after bioinformatics studies have permitted the identification of other genes by comparison with the sequences of known enzymes. (Adapted from Ref. 198 with permission of Nature Publishing Group.)

not alter the metabolic system to any great extent. Alternatively, the allelic variation may result in a metabolic flux in another closely linked pathway that is totally unexpected and not seen until an identification and comparison of the metabolic profiles of the wild type and mutant is made (Fig. 22).

Distinguishing features of the metabolome include the fact that metabolites are not directly linked to the genetic code but are actually products of a concerted action of several networks of enzymes and other regulatory proteins [199]. Metabolites are also structurally far more complex than the four-letter code describing

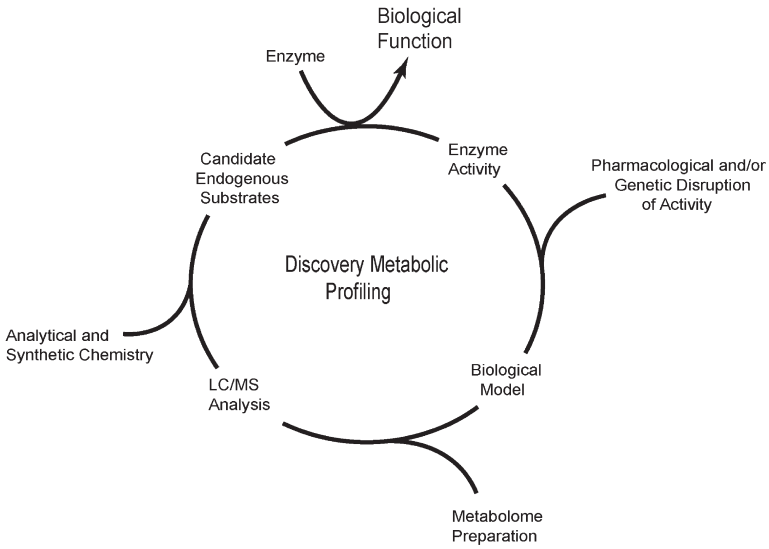


**FIGURE 22** Connecting the proteome and metabolome through discovery metabolite profiling (DMP). Assigning the enzyme function using DMP begins by disruption of the enzyme activity of interest through either pharmacological or genetic means. Differences in the metabolome between wild-type and enzyme-disrupted systems are then identified using untargeted LC/MS methods and converted into chemical structures using more detailed analytical methods (tandem MS, NMR) and chemical synthesis. These compounds are then directly examined as substrates for the enzyme of interest. (From Ref. 199. Copyright © 2005, with permission from Elsevier.)

genes or the 20-letter code describing proteins [194]. As a result, metabolites cannot be sequenced like genes or proteins and their characterization requires that the structure of each metabolite be determined individually through a description of elemental composition and stereochemical orientation of functional groups.

The metabolome is complicated by the interactive nature of the biochemical systems that produce the metabolites, and the environmental conditions under which a metabolomic experiment is carried out need to be accurately specified. Also, it is obvious that metabolites are not just steps in a metabolic cycle but are also sensors and regulators in a complex series of molecular interactions in the whole organism [200]. Therefore, metabolic studies are severely limited by the knowledge of the biochemical systems of the organism under investigation as well as the "state" of the specific individual under investigation. This individuality further complicates the metabolomic study since individuals are not only genetically unique but are also environmentally unique. It is also important to note that the complexity of the metabolome may mean that variation in specific genes reflected in the metabolome will not be reflected by a concomitant change in the morphological or physiological phenotype (a phenotypically silent variation) until it is accompanied by additional variations in the genome, transcriptome, proteome, metabolome, or envirome [201]. Some unique biochemical pathways will be affected by allelic variation in the genes responsible for the regulators of the pathway, but those metabolic reaction pathways that can be accommodated by other pathways will not necessarily be altered by an

allelic variation until a combination of variations in both pathways occurs. The study of metabolomics provides an important "time frame" into which the analysis of genomic and proteomic variations can be placed to maximize the probability of observing biological transitions that predict functional outcomes of the variations [195].

Approaches to metabolomic investigations are classified in a manner similar to the single gene, candidate gene, and whole genome classification of genomic studies (see Fig. 5) [194]. As the metabolic/genetic investigation increases in complexity so does the protocol and technology used. The strategy of *target analysis* is to directly investigate the primary metabolic effect of a genetic alteration. The analysis is usually limited to the substrate and/or direct product of the protein expressed by the altered gene. The *metabolite profiling* strategy limits the study to a number of predefined metabolites in a biological sample and is usually narrowed down to metabolites in a particular metabolic pathway. *Metabolomics* defines the strategy for a comprehensive analysis, including identification and quantification, of all the metabolites in a biological system and is designed to investigate the possible multiple effects of genetic variation on many different biochemical pathways. The technology behind metabolomics will often be called upon to identify unknown metabolites as well as handle the complex datasets of multiple interacting biochemical systems. Finally, the strategy of *metabolic fingerprinting* is to screen a large number of metabolites for biological relevance in diagnostics or other analytical procedures where it is important to rapidly distinguish between individuals in a population.

The main analytical technologies used in metabolomic investigations are nuclear magnetic resonance (NMR) and mass spectrometry (MS) alone or in combination with liquid (LC-MS) or gas chromatographic (GC-MS) separation of metabolites [37, 193–197, 199, 201–207]. Other techniques include thin layer chromatography [208], Fourier transform infrared (FT-IR) spectrometry [209], metabolite arrays [210], and Raman spectroscopy [211].

### 4.5 Enviromics

The analysis of gene–environment interactions is relatively new and mainly in the area of disease diagnosis. The term enviromics was first used in the field of psychiatric epidemiology and defined the study of "the total ensemble of environments, both current and in earlier life, that affect the occurrence of mental and behavioural disturbances" [55]. A more generalized definition of enviromics that would be applicable to pharmacogenomics would be the total ensemble of environments both current and in earlier life that can be identified to affect disease development and therapeutic outcomes. This would include such variables as physical environment for work and relaxation, lifestyle, social stressors, patient demographics, diet, and exposure to other drugs and toxic agents including pathogens. The critical problem to development of enviromics in the description of pharmacogenomics is to accurately describe and adequately quantify the environmental factors under consideration in the drug study. *Genetic epidemiology* is the discipline that seeks to explain the extent to which genetic and environmental factors that people are exposed to influence their risk of disease, by means of population-based investigations.

Some relatively simple examples of gene–environment interactions and their correlation with particular diseases are listed in Table 4. The concept of pharmacogenetics could be considered a special example of the binary nature of gene–environment interaction, where the environmental exposure is a drug. The allelic variation of genes involved in the metabolism of the particular drug are used to identify individuals who are at risk for adverse drug reactions or treatment failure. The population can then be divided into those who should avoid taking the drug or those who should have their dose modified [15]. However, it would be naive to assume that the drug is the only environmental factor to play a role during disease therapy. Indeed, under the holistic view of pharmacogenomics, there are a number of complicating factors leading to a successful drug therapy.

Enviromics is a relatively new and undocumented area in pharmacogenomics and considerable research must be done, selecting critical environmental factors affecting personalized medicine and examining their interaction with the genome, transcriptome, proteome, and metabolome in the production of molecular biomarkers. Without a thorough understanding of the effects of the envirome, one will have only limited success with the application of the concept of personalized medicine.

## 5 ASSOCIATIVE ANALYSIS

Ultimately there are four phenotypes that must be described to adequately assess the affect of the envirome, genome, transcriptome, proteome, and metabolome on disease development, initiation of therapy, and therapeutic outcome: the disease and normal phenotypes, the drug phenotype, and the therapeutic phenotype (see Fig. 8). The molecular profile of the disease will play a major role in the selection of therapies that are available to the physician. The relationship between the molecular profiles of the normal phenotype and the disease phenotype will provide a valuable source of detailed molecular information for the predict-and-manage paradigm of health care, outlined by the United States Pharmacopoeial Convention [56]. It should also be mentioned here that the genetic basis of complex phenotypes are often "modifier" variations in that changes in expression levels of multiple genes contribute to the phenotypic outcome.

The complexity of gene–environment interactions and their effect on the disease phenotype is summed up in a quote from D. J. Weatherall [224]: "A central problem for the medical sciences in general, and for clinical genetics in particular, is the extent to which it will be possible to relate findings at the molecular level to clinical phenotype."

The association of genetic variation with environmental influences is a rather new field for drug therapy but it has been extensively studied with respect to the disease phenotype, although the effect of environment has usually been limited to a single variable. An example of the complexity that exists in genetic influences on the disease phenotype alone has been impressively demonstrated for the β-thalassemias (Fig. 23) [224]. Weatherall [224] has pointed out that the mechanisms underlying the phenotypic diversity of the β-thalassemias can be classified as follows: *primary causes* that are the many genetic variations that occur in the β-globin genes that underlie the β-thalassemias; *secondary causes* that result from variations at other genetic loci involved in globin synthesis; and *tertiary causes* resulting from variations

**TABLE 4 Selected Examples of Gene–Environment Interactions Observed in at Least Two Studies**

| Gene Symbol[a] | Variant(s) | Environmental Exposure | Outcome and Nature of Interaction | References |
|---|---|---|---|---|
| Genes for skin pigmentation (e.g., MC1R) | Variants for fair skin color | Sunlight or ultraviolet light B | Risk of skin cancer is higher in people with fair skin color who are exposed to higher amounts of sunlight | 212 |
| CCR5 | Δ-32 deletion | HIV[a] | Carriers of the receptor deletion have lower rates of HIV infection and disease progression | 213 |
| MTHFR | Ala222Val polymorphism | Folic acid intake | Homozygotes for the low activity Ala222Val variant are at different risk of colorectal cancer and adenomas if nutritional folate status is low | 214 |
| NAT2 | Rapid versus slow acetylator SNPs | Heterocyclic amines in cooled meat | Red meat intake is more strongly associated with colorectal cancer among rapid acetylators | 215 |
| F5 | Leiden prothrombotic variant | Hormone replacement | Venous thromboembolism risk is increased in Factor V Leiden carriers who take exogenous steroid hormones | 216 |
| UGT1A6 | Slow metabolism SNPs | Aspirin | Increased benefit of prophylactic aspirin use in carriers of the slow metabolism variants | 217 |
| APOE | E4 allele | Cholesterol intake | Exaggerated changes in serum cholesterol in response to dietary cholesterol changes in APOE4 carriers | 218 |
| ADH1C | γ-2 alleles | Alcohol intake | Inverse association between ethanol intake and myocardial infarction; risk is stronger in carriers of slow-oxidizing γ-2 alleles | 219 |
| PPARG2 | Pro12Ala | Dietary fat intake | Stronger relation between dietary fat intake and obesity in carriers of the Pro12Ala allele | 220 |
| HLA-DPB1 | Glu69 | Occupational beryllium | Exposed workers who are carriers of the Glu69 allele are more likely to develop chronic beryllium lung disease | 221 |
| TPMT | Ala154Thr and Tyr240Cys | Thiopurine drugs | Homozygotes for the low-activity alleles of TPMT are likely to experience severe toxicity when exposed to thiopurine drugs | 222 |
| ADRB2 | Arg16Gly | Asthma drugs | Arg16Gly homozygotes have a greater response in the airway to albuterol | 223 |

[a]Abbreviations: ADH1C, alcohol dehydrogenase 1C (class 1), γ-polypeptide; ADRB2, adrenergic β-2 receptor; CCR5, chemokine (C–C motif) receptor 5; APOE, apolipoprotein E; F5, coagulation factor V; HIV, human immunodeficiency virus; HLA-DPB1, major histocompatibility complex, class II, DP β-1; MC1R, melanocortin receptor 1; MTHFR, 5,10-methylenetetrahydrofolate reductase (NADPH); NAT2, *N*-acetyltransferase 2; PPARG2, peroxisome proliferative activated receptor-γ; TPMT, thiopurine-*S*-methyltransferase; UGT1A6, UDP-glycosyltransferase-1 family, polypeptide A6.
*Source*: From Ref. 15.

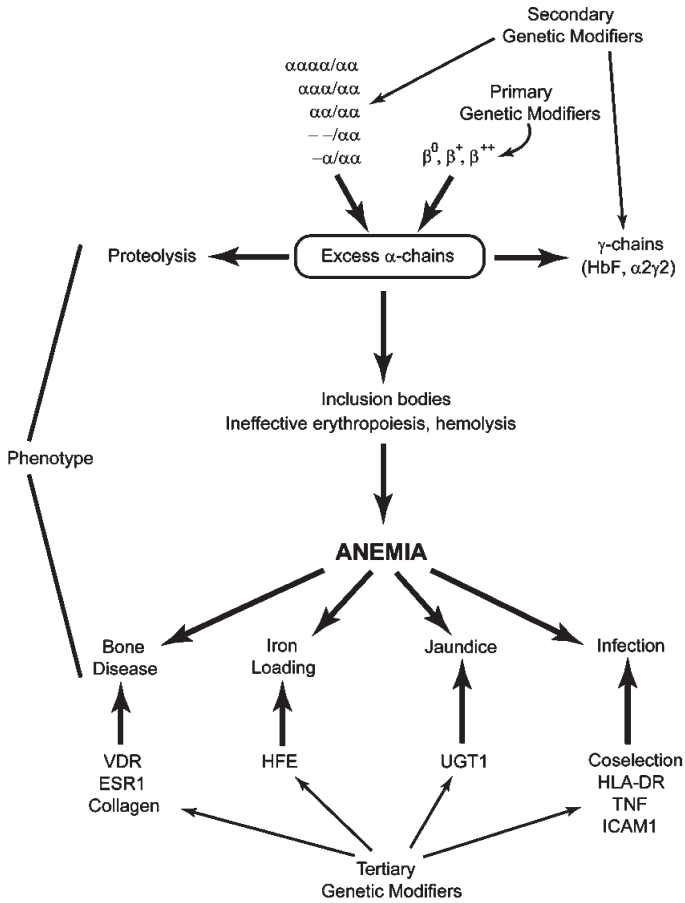


**FIGURE 23** Summary of the genetic contributions to phenotypic diversity of the β-thalassemias. The secondary modifiers, the α- and γ-globin genes, are shown at the top of the figure. The tertiary modifiers are shown at the bottom of the figure. VDR, vitamin D receptor; ESR1, estrogen receptor; collagen, several genes determined in collagen synthesis; HFE, the locus for hereditary hemochromatosis; UGT1, UGT glucuronyltransferase involved in bilirubin glucuronidation; HLA-DR, major histocompatibility complex locus; TNF, tumor necrosis factor-α; ICAM1, intercellular adhesion molecule 1. (Adapted from Ref. 224 with permission of the author and Nature Publishing Group.)

at genetic loci that are not involved in globin production but that might modify the complications of the disease in many different ways (Fig. 23).

Well over 200 different mutations have been identified in the β-globin genes of β-thalassemia patients with the majority being point mutations or the deletion of one or two bases [225, 226]. The primary genetic cause of the resulting phenotypes is described by the complete loss of β-globin gene product ($\beta^o$) and the moderate or marked reduction in the normal output of β-globin gene product ($\beta^+$ and $\beta^{++}$, respectively) (Fig. 23). The location of the mutation in the genomic DNA as well as the degree of recessive or dominant mode of inheritance of these mutations and

the stability of the gene products play major roles in the primary cause of the final phenotype seen in the clinic [224].

Clinical studies have shown that there is a wide diversity in the severity of disease among individuals with the same β-thalassemia genotype. Many of these unusual examples of phenotypic diversity can be explained by the secondary action of products of the α- and γ-globin gene loci involved in globin synthesis [224].

Finally, variability at gene loci that have nothing to do with globin chain synthesis may have important effects on variation in the disease phenotype, particularly with respect to the complications of the disease [224]. These tertiary genetic causes of phenotypic diversity may come from such examples as (1) jaundice resulting from the alteration of bilirubin levels caused by a polymorphism in the promoter region of the gene for UDP-glucuronosyltransferase (UGT1); (2) bone disease resulting from allelic variation in genes involved in bone metabolism, including the vitamin D receptor gene, collagen, and the estrogen receptor; (3) damage from variations in iron loading resulting from polymorphisms in the HFE locus and causing such complications as cardiac disease, hepatic disease, and diabetes; and (4) phenotypic variation due to coselection reflecting a β-thalassemia heterozygote advantage against malaria. The varying susceptibility to malaria is also reflected by polymorphic variation in the major histocompatibility complex loci (HLA-DR), tumor necrosis factor-α (TNF-α), and intercellular adhesion molecule 1 (ICAM1).

Molecular profiles of the primary, secondary, and tertiary genetic variables for the β-thalassemias are obviously complex and mechanistically incomplete. Therefore, the major task of including enviromic data in these molecular profiles is still a daunting undertaking and will be a major obstacle to overcome in the associative analyses leading to a complete description of disease, drug, and therapeutic phenotypes [224].

A similar classification of genetic factors could be devised for the general drug phenotype category in pharmacogenomics. The candidate gene approach and the modular biology of drug action both lend themselves to a division of genetic factors into primary, secondary, and tertiary causes. Polymorphisms altering drug metabolism, transport, or target function could be considered the primary factors.

The association of primary genetic factors with phenotype in pharmacogenomics is currently restricted to polymorphic variation in drug metabolizing enzymes in accessible tissues like erythrocytes and leukocytes [59]. Phenotyping involves direct measurement of enzyme activity or the determination of the metabolic ratio of a probe drug (Table 5). Since determination of enzyme activity can be problematic, the administration of probe drugs to the patient and determination of drug and metabolite levels resulting from metabolism of the probe drug in the liver is the common technique for determination of drug phenotype [59]. The majority of phenotyping studies have analyzed the levels and polymorphic variations of metabolizing enzymes, and there have been few such studies on the effects of variation in drug transport proteins or receptor proteins. Application of the candidate gene approach and the modular biology concept in the therapeutic drug pathway to pharmacogenomics will begin to correct this shortfall.

The interaction between genes and the environment in which they exist is not only a cornerstone to evolutionary theory but plays a critical role in disease pathology and drug therapy. A study of the effect of genes and environment on disease development and drug therapy requires a serious effort to assess environmental risk

**TABLE 5 Phenotyping Methods for Analysis of Pharmacogenetic Polymorphisms**

| Gene[a] | Method or Probe Drug | Type of Sample | References |
|---|---|---|---|
| Cholinesterase | Enzyme inhibition | Plasma | 227 |
| COMT | Enzyme levels | Erythrocytes | 228 |
| CYP1A1 | Induced enzyme activity | Lymphocytes | 229 |
| CYP1A2 | Caffeine | Blood, urine, and saliva | 230, 231 |
| CYP2A6 | Coumarin | Urine | 232 |
| CYP2C9 | Tolbutamide | Urine | 233 |
| CYP2C19 | Mephenytoin | Urine | 234 |
| CYP2C19 | Omeprazole | Urine | 235 |
| CYP2D6 | Debrisoquine | Urine | 236 |
| CYP2D6 | Sparetine | Urine | 237 |
| CYP2D6 | Dextromethorphan | Urine | 238 |
| CYP2D6 | Metoprolol | Urine | 239 |
| CYP2E1 | Chlorzoxazone | Blood | 240 |
| CYP2E1 | Direct measurement | Lymphocytes | 241 |
| CYP 3A4 | Midazolam | Blood | 242 |
| CYP 3A4 | 6-Hydroxycortisol | Urine | 243 |
| CYP 3A4 | Erythromycin breath test | Breath test | 244 |
| FMO3 | Trimethylamine | Urine | 245 |
| GSTM1 | Enzyme levels | Leukocytes | 246 |
| GSTT1 | Enzyme levels | Erythrocytes | 247 |
| NAT1 | 4-Aminobenzoate | Urine | 248 |
| NAT2 | Caffeine | Urine | 249 |
| Paraoxonase | Enzyme levels | Plasma | 250 |
| Sulfotransferase | Enzyme levels | Platelet | 251, 252 |
| TPMT | Enzyme levels | Erythrocytes | 253 |
| P-glycoprotein | Digoxin | Blood | 254 |

[a]Abbreviations: COMT, catechol-*O*-methyltransferase; CYP, cytochrome P450; FMO, flavin-linked monooxygenase; GST, glutathione *S*-transferase; NAT, *N*-acetyltransferase; TPMT, thiopurine methyltransferase.
*Source*: Adapted from Ref. 59 with kind permission of the author and Springer Science and Business Media.

factors and not to simply treat them as a "binary indicator variable for 'exposed or not' or as a black box of correlated residual variation. It entails serious consideration of exogenous exposures as complex, multivariate, time-dependent variables, along with an understanding of the metabolic pathways that influence the internal environment and the genes that regulate these pathways" [255]. Obtaining high quality information on environment and lifestyle in conjunction with biological samples to assess genetic variants in genes responsible for producing the proteins of a therapeutic drug pathway, as well as a consistent and accurate evaluation of therapeutic outcomes, will be a difficult and complex, but crucial, task in the assessment of gene–environment interactions for drug therapy [15].

Well designed studies of gene–environment interactions are assessed using either a qualitative (Table 6) [15, 256–258] or statistical model [15, 259]. However, the complexity of screening a large number of potential gene–environment interactions in populations with a large number of genotypes and many variables of environmental exposure greatly increases the chance of finding false-positive results at

**TABLE 6 Some Patterns of Relative Risk in Gene–Environment Interactions for Three Classical Genetic Diseases that Have an Environmental Component, Assuming Dichotomous Genetic Susceptibility and Environmental Exposure.**

| Gene Variant | Environmental Exposure | Relative Risk (XP)[a] | Relative Risk (PKU)[b] | Relative Risk (Emphysema)[c] |
|---|---|---|---|---|
| Absent | Absent | 1.0 | 1.0 | 1.0 |
| Present | Absent | ~1.0 | 1.0 | Modest |
| Absent | Present | Modest | 1.0 | Modest |
| Present | Present | Very high | Very high | High |

[a]In the XP (xeroderma pigmentosum) example, exposure to ultraviolet light increases the risk of developing skin cancer in noncarriers of XP mutations. The combination of these mutations and exposure to ultraviolet light vastly increases the risk of skin cancer. In theory, if individuals with XP mutations completely avoid ultraviolet light, their risk of skin cancer becomes close to the background risk.
[b]In the PKU (phenylketonuria) example, only individuals with recessive mutations in the causative gene (phenylalanine hydroxylase) who are exposed to phenylalanine in the diet are susceptible to PKU.
[c]In the emphysema example, with a deficiency in the α-1 antitrypsin gene, both nonsmokers who are at genetic risk and smokers who are not at genetic risk have an increased risk of developing emphysema, and the combination (smokers who are at genetic risk) is associated with the highest risk.
*Source*: From Ref. 15. Data from Refs. 256 and 257.

conventional levels of statistical significance [15]. Most studies are not powerful enough to detect modest interactions and demanding small $p$ values to counteract this problem will result in a lower probability of declaring true-positive interactions as "significant." A plausible option is to restrict the search for gene–environment interactions to those that involve gene products and environmental exposures that interact in the same biological pathways. Further restricting analysis to gene variants that alter gene function within the biological pathways may also reduce the complexity of the analysis, although for variants that affect gene regulation this science is in its infancy [260]. Therefore, the application of the candidate gene approach to pharmacogenomic analysis may prove useful in the analysis of gene–environment interactions as well, especially when one utilizes the "modular biology" concept to simplify the biochemical pathways of drug response [261, 262].

## 6 CONCLUSION

The preclinical development of genetic data for application in the pharmacogenomics of disease therapy and management has become far more complex than the earlier studies had suggested. The Mendelian single gene pharmacogenetic approach to personalized medicine will be limited to a small number of phase I (CYP2D6, CYP2C9, CYP2C19) and phase II (TPMT, UGT) metabolic enzymes that can be directly linked to toxicity or efficacy problems through a demonstration of a genetic link to "poor metabolism" or "ultra metabolism" of a particular drug. The next stages beyond the single gene approach are the candidate genes and the whole genome approaches to pharmacogenomics. While the application of linkage disequilibrium to SNP analysis and the development of the HapMap in the whole genome approach to pharmacogenomics is conceptually the most comprehensive application of allelic variation to disease diagnosis and drug therapy in personalized medicine, the number of SNPs that must be determined for each patient may be a prohibitive

factor. The current application of htSNPs in candidate genes, determined through the development of a drug therapeutic pathway using the modular biology approach to assess the pharmacokinetic and pharmacodynamic parameters, may be the technique that reduces the number of SNPs that must be determined to a level that is financially feasible and sufficiently reproducibly accurate to apply to the clinical environment of personalized medicine.

The correlation of genetic variation in the genes involved in a drug's pharmacokinetics of distribution in the body and pharmacodynamics of action on the body is not likely to provide sufficient information for successful personalized medicine. One must also accept the fact that gene transcription, translation, and protein function involve genetically independent mechanisms that will modify the correlation of genotype with phenotype. As a result, it will be important to augment the genetic biomarkers with transcriptomic, proteomic, and metabolomic biomarkers in order to assure a successful application of personalized medicine in a clinical environment.

While the technology for the high throughput genotyping required of candidate gene analysis is highly developed, there is room for cheaper high throughput methods with more effective elimination of false positives while maintaining sensitivity. Also, extensive technology development will be required to get the application of high throughput in transcriptomics, proteomics, and metabolomics to acceptable standards for the clinic. The main hurdle looming on the pharmacogenomic horizon is the envirome. The genotype–phenotype correlation in disease diagnosis and therapy cannot escape interaction with the environment. And nowhere will the effects of environment be felt greater than in disease diagnosis and assessment of therapeutic outcome. Health care professionals will become more adept at selecting and accurately assessing critical environmental factors and their interactions with the many genomic, transcriptomic, proteomic, and metabolomic molecular biomarkers in the diagnosis of disease, choice of therapeutic agent (or not), and assessment of therapeutic outcome or proactive predict-and-manage approach.

This endeavor will require the education of health care professionals in the application of these technologies and education of the public on what it means to have a genetic test. It will also require the education of politicians and policy makers on the efficient utilization of the technology to both reduce health care costs, over the long term, and at the same time protect the individual's personal information and right to privacy. There will also be a large requirement for health informatics to handle the huge amount of data that will arise from the various clinical studies in developing the pharmacogenomics database, as well as a demand for user friendly access to the information by health care professionals who do not have the time nor the inclination to spend hours learning a computer program. Some of the biggest challenges for pharmacogenomics lie in handling the information. Innovation and technology tend to outstrip the development of good algorithms and software to complement them. There is a need for better software to collect, organize, and interpret data. Since pharmacogenomics is a combination of several different areas of research, we need to work on sharing and combining different types of biological information, rather than trying to generate it faster. The development of pharmacogenomics for effective and efficient disease diagnosis and therapy in the clinic will depend heavily on the cooperation and collaboration of a number of different scientific communities involved in the many aspects of genome research.

## ACKNOWLEDGMENTS

I am deeply in debt to a number of individuals who have taken time out of their busy schedules to read and comment on this chapter. The quality of the content and presentation of the information in this chapter was greatly improved by the criticisms and corrections provided by Dan Koboldt, David Wishart, Zhming Jiang, and Tom Pfeifer, for which I thank them profusely. In the final judgment, however, the responsibility for any remaining errors and omissions is mine.

## REFERENCES

1. Garrod AE. The incidence of alkaptonuria: a study in chemical individuality. *Lancet* 1902;2:1616–1620.
2. Forbat A, Lehmann H, Silk E. Prolonged apoena following injection of succinyldicholine. *Lancet* 1953;Nov 21:1067–1068.
3. Kalow W. Familial incidence of low pseudocholinesetrase level. *Lancet* 1956;Sept 15:576–577.
4. Lehmann H, Ryan E. The familial incidence of low pseudocholinesterase level. *Lancet* 1956;July 21:124.
5. Price Evans DA, Manley KA, McKusick VA. Genetic control of isoniazid metabolism in man. *Br Med J* 1960;Aug 13:485–491.
6. Carson PE, Flanagan CL, Ickes CE, Alving AS. Enzymatic deficiency in primaquine-sensitive erythrocytes. *Science* 1956;124:484–485.
7. Williams RJ. *Biochemical Individuality. The Basis of the Genetotrophic Concept*. Austin: University of Texas Press;1956.
8. Motulsky AG. Drug reactions, enzymes and biochemical genetics. *JAMA* 1957;165:835–837.
9. Vogel F. Moderne probleme der humangenetik. *Ergebnisse der Innere Medizinische und Kinderheilkunde* 1959;12:52–62.
10. Watson JD, Crick FHC. Molecular structure of nucleic acids. *Nature* 1953;171:737–738.
11. Guttmacher AE, Collins FS. Genomic medicine—a primer. *N Engl J Med* 2002;347:1512–1520.
12. Wang J, Fan JB, Siao CJ, Berno A, Young P, Saplosky R, Ghandour G, Perkins N, Winchester E, Spenser J, Kruglyak L, Stein L, Hsie L, Topaloglou T, Hubbell E, Robinsion E, Mittmann M, Morris MS, Shen N, Kilburn D, Rioux J, Nusbaum C, Rozen S, Hudson TJ, Lander ES. Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome. *Science* 1998;280:1077–1082.
13. HGP. The human genome. Science genome map. *Science* 2001;291:1218.
14. Brockmoller J, Kirchheiner J, Meisel C, Roots I. Pharmacogenetic diagnostics of cytochrome P450 polymorphisms in clinical drug development and in drug treatment. *Pharmacogenomics* 2000;1:125–151.
15. Hunter DJ. Gene–environment interactions In human diseases. *Nat Rev Genet* 2005;6:287–298.
16. Kirchheiner J, Fuhr U, Brockmoller J. Opinion: Pharmacogenetics-based therapeutic recommendations—ready for clinical practice? *Nat Rev Drug Discov* 2005;4:639–647.
17. Kirchheiner J, Brockmoller J. Clinical consequences of cytochrome P450 2C9 polymorphisms. *Clin Pharmacol Ther* 2005;77:1–16.

18. Roses AD. Pharmacogenetics and the practice of medicine. *Nature* 2000;405:857–865.
19. Risch N, Merikangas K. The future of genetic studies of complex human diseases. *Science* 1996;273:1516–1517.
20. Collins FS, Guyer MS, Charkravarti A. Variations on a theme: cataloging human DNA sequence variation. *Science* 1997;278:1580–1581.
21. Daly MJ, Rioux JD, Schaffner SF, Hudson TJ, Lander ES. High-resolution haplotype structure in the human genome. *Nat Genet* 2001;29:229–232.
22. Gabriel SB, Schaffner SF, Nguyen H, Moore JM, Roy J, Blumenstiel B, Higgins J, DeFelice M, Lochner A, Faggart M, Liu-Cordero SN, Rotimi C, Adeyemo A, Cooper R, Ward R, Lander ES, Daly MJ, Altshuler D. The structure of haplotype blocks in the human genome. *Science* 2002;296:2225–2229.
23. Consortium TIH. The International HapMap Project. *Nature* 2003;426:789–796.
24. Taillon-Miller P, Saccone SF, Saccone NL, Duan S, Kloss EF, Lovins EG, Donaldson R, Phong A, Ha C, Flagstad L, Miller S, Drendel A, Lind D, Miller RD, Rice JP, Kwok PY. Linkage disequilibrium maps constructed with common SNPs are useful for first-pass disease association screens. *Genomics* 2004;84:899–912.
25. Johnson GC, Esposito L, Barratt BJ, Smith AN, Heward J, Di Genova G, Ueda H, Cordell HJ, Eaves IA, Dudbridge F, Twells RC, Payne F, Hughes W, Nutland S, Stevens H, Carr P, Tuomilehto-Wolf E, Tuomilehto J, Gough SC, Clayton DG, Todd JA. Haplotype tagging for the identification of common disease genes. *Nat Genet* 2001;29: 233–237.
26. Ahmadi KR, Weale ME, Xue ZY, Soranzo N, Yarnall DP, Briley JD, Maruyama Y, Kobayashi M, Wood NW, Spurr NK, Burns DK, Roses AD, Saunders AM, Goldstein DB. A single-nucleotide polymorphism tagging set for human drug metabolism and transport. *Nat Genet* 2005;37:84–89.
27. Goldstein DB, Ahmadi KR, Weale ME, Wood NW. Genome scans and candidate gene approaches in the study of common diseases and variable drug responses. *Trends Genet* 2003;19:615–622.
28. Deloukas P, Bentley D. The HapMap project and its application to genetic studies of drug response. *Pharmacogenomics J* 2004;4:88–90.
29. Lin M, Aquilante C, Johnson JA, Wu R. Sequencing drug response with HapMap. *Pharmacogenomics J* 2005;5:149–156.
30. Kamatani N, Sekine A, Kitamoto T, Iida A, Saito S, Kogame A, Inoue E, Kawamoto M, Harigai M, Nakamura Y. Large-scale single-nucleotide polymorphism. (SNP) and haplotype analyses, using dense SNP Maps, of 199 drug-related genes in 752 subjects: the analysis of the association between uncommon SNPs within haplotype blocks and the haplotypes constructed with haplotype-tagging SNPs. *Am J Hum Genet* 2004;75: 190–203.
31. Reich DE, Cargill M, Bolk S, Ireland J, Sabeti PC, Richter DJ, Lavery T, Kouyoumjian R, Farhadian SF, Ward R, Lander ES. Linkage disequilibrium in the human genome. *Nature* 2001;411:199–204.
32. Sachidanandam R, Weissman D, Schmidt SC, Kakol JM, Stein LD, Marth G, Sherry S, Mullikin JC, Mortimore BJ, Willey DL, Hunt SE, Cole CG, Coggill PC, Rice CM, Ning Z, Rogers J, Bentley DR, Kwok PY, Mardis ER, Yeh RT, Schultz B, Cook L, Davenport R, Dante M, Fulton L, Hillier L, Waterston RH, McPherson JD, Gilman B, Schaffner S, Van Etten WJ, Reich D, Higgins J, Daly MJ, Blumenstiel B, Baldwin J, Stange-Thomann N, Zody MC, Linton L, Lander ES, Altshuler D. A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature* 2001;409: 928–933.

33. Weiss KM, Terwilliger JD. How many diseases does it take to map a gene with SNPs? *Nat Genet* 2000;26:151–157.
34. Roberts L. Human genome research. SNP mappers confront reality and find it daunting. *Science* 2000;287:1898–1899.
35. Kruglyak L. Prospects for whole-genome linkage disequilibrium mapping of common disease genes. *Nat Genet* 1999;22:139–144.
36. Collins A, Lonjou C, Morton NE. Genetic epidemiology of single-nucleotide polymorphisms. *Proc Natl Acad Sci USA* 1999;96:15173–15177.
37. Griffin JL, Shockcor JP. Metabolic profiles of cancer cells. *Nat Rev Cancer* 2004;4:551–561.
38. Hieter P, Boguski M. Functional genomics: it's all how you read it. *Science* 1997;278:601–602.
39. Fierz W. Challenge of personalized health care: to what extent is medicine already individualized and what are the future trends? *Med Sci Monit* 2004;10:RA111–RA123.
40. Bordet R, Gautier S, Le Louet H, Dupuis B, Caron J. Analysis of the direct cost of adverse drug reactions in hospitalised patients. *Eur J Clin Pharmacol* 2001;56:935–941.
41. Suh DC, Woodall BS, Shin SK, Hermes-De Santis ER. Clinical and economic impact of adverse drug reactions in hospitalized patients. *Ann Pharmacother* 2000;34:1373–1379.
42. du Souich P. In human therapy, is the drug–drug interaction or the adverse drug reaction the issue? *Can J Clin Pharmacol* 2001;8:153–161.
43. Ginsburg GS, McCarthy JJ. Personalized medicine: revolutionizing drug discovery and patient care. *Trends Biotechnol* 2001;19:491–496.
44. Collins FS, Green ED, Guttmacher AE, Guyer MS. A vision for the future of genomics research. *Nature* 2003;422:835–847.
45. Ross JS, Ginsburg GS. The integration of molecular diagnostics with therapeutics. Implications for drug development and pathology practice. *Am J Clin Pathol* 2003;119:26–36.
46. Meyer JM, Ginsburg GS. The path to personalized medicine. *Curr Opin Chem Biol* 2002;6:434–438.
47. Senn S. Individual response to treatment: is it a valid assumption? *BMJ* 2004;329:966–968.
48. Jain KK. From molecular diagnostics to personalized medicine. The IBC Workshop, London, UK, 1st May 2002. *Expert Rev Mol Diagn* 2002;2:299–301.
49. Evans WE, Relling MV. Moving towards individualized medicine with pharmacogenomics. *Nature* 2004;429:464–468.
50. Group BDW. Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. *Clin Pharmacol Ther* 2001;69:89–95.
51. Lewin DA, Weiner MP. Molecular biomarkers in drug development. *Drug Discov Today* 2004;9:976–983.
52. Lesko LJ. *Guidance for Industry: Pharmacogenomic Data Submissions*. Washington, DC: US Department of Health and Human Services, Food and Drug Administration, CDER, CBER, CDRH; 2003. http://www.fda.gov/cder/guidance/5900dfp.pdf.
53. Koch WH. Technology platforms for pharmacogenomic diagnostic assays. *Nat Rev Drug Discov* 2004;3:749–761.
54. Singleton P, Sainsbury D. *Dictionary of Microbiology and Molecular Biology*. Hoboken, NJ: Wiley-Interscience;1994.
55. Anthony JC. The promise of psychiatric enviromics. *Br J Psychiatry Suppl* 2001;40:s8–s11.

56. Bezold C, Halpern JA, Eng JL. *2020 Visions: Health Care Information Standards and Technologies*. Rockville, MD: The United States Pharmacopoeial Convention, Inc; 1993.
57. Romkes M, Buch SC. Genotyping technologies: application to biotransformation enzyme genetic polymorphism screening. *Methods Mol Biol* 2005;291:399–414.
58. Kwok P-Y, Chen X. Detection of single nucleotide polymorphisms. *Curr Issues Mol Biol* 2003;5:43–60.
59. Daly AK. Development of analytical technology in pharmacogenetic research. *Naunyn Schmiedebergs Arch Pharmacol* 2004;369:133–140.
60. Twyman RM. SNP discovery and typing technologies for pharmacogenomics. *Curr Top Med Chem* 2004;4:1423–1431.
61. Shi MM. Enabling large-scale pharmacogenetic studies by high-throughput mutation detection and genotyping technologies. *Clin Chem* 2001;47:164–172.
62. Cargill M, Altshuler D, Ireland J, Sklar P, Ardlie K, Patil N, Lane CR, Lim EP, Kalayanaraman N, Nemesh J, Ziaugra J, Friedland L, Rolfe A, Warrington J, Lipshutz RJ, Daley GQ, Lander ES. Characterization of single-nucleotide polymorphisms in coding regions of human genes. *Nat Genet* 1999;22:231–238.
63. Halushka MK, Fan JB, Bentley K, Hsie L, Shen N, Weder A, Cooper R, Lipshutz RJ, Chakravarti A. Patterns of single-nucleotide polymorphisms in candidate genes for blood-pressure homeostasis. *Nat Genet* 1999;22:239–247.
64. Carlson CS, Newman TL, Nickerson DA. SNPing in the human genome. *Curr Opin Chem Biol* 2001;5:78–85.
65. Kirk BW, Feinsod M, Favis R, Kliman RM, Barany F. Single nucleotide polymorphism seeking long term association with complex disease. *Nucleic Acids Res* 2002;30: 3295–3311.
66. Syvanen AC. Accessing genetic variation: genotyping single nucleotide polymorphisms. *Nat Rev Genet* 2001;2:930–942.
67. Kwok P-Y. High-throughput genotyping assay approaches. *Pharmacogenomics* 2000;1:95–100.
68. Mullis KB, Faloona FA. Specific synthesis of DNA *in vitro* via a polymerase-catalyzed chain reaction. *Methods Enzymol* 1987;155:335–350.
69. Saiki RK, Walsh PS, Levenson CH, Erlich HA. Genetic analysis of amplified DNA with immobilized sequence-specific oligonucleotide probes. *Proc Natl Acad Sci USA* 1989;86:6230–6234.
70. Mei R, Galipeau PC, Prass C, Berno A, Ghandour G, Patil N, Wolff RK, Chee MS, Reid BJ, Lockhart DJ. Genome-wide detection of allelic imbalance using human SNPs and high-density DNA arrays. *Genome Res* 2000;10:11126–11137.
71. Hacia JG, Sun B, Hunt N, Edgemon K, Mosbrook D, Robbins C, Fodor SP, Tagle DA, Collins FS. Strategies for mutational analysis of the large multiexon ATM gene using high-density oligonucleotide arrays. *Genome Res* 1998;8:1245–1258.
72. Emahazion T, Feuk L, Jobs M, Sawyer SL, Fredman D, St Clair D, Prince JA, Brookes AJ. SNP association studies in Alzheimer's disease highlight problems for complex disease analysis. *Trends Genet* 2001;17:407–413.
73. Howell WM, Jobs M, Gyllensten U, Brookes AJ. Dynamic allele-specific hybridization. A new method for scoring single nucleotide polymorphisms. *Nat Biotechnol* 1999;17:87–88.
74. Tyagi S, Bratu DP, Kramer FR. Multicolor molecular beacons for allele discrimination. *Nat Biotechnol* 1998;16:49–53.

75. Kostrikis S, Tyagi S, Mhlanga MM, Ho DD, Kramer FR. Spectral genotyping of human alleles. *Science* 1998;279:1228–1229.
76. Livak KJ. Allelic discrimination using fluorogenic probes and the 5′-nuclease assay. *Genet Anal* 1999;14:143–149.
77. Latif S, Bauer-Sardina I, Ranade K, Livak K, Kwok P-Y. Fluorescence polarization in homogeneous nucleic acid analysis II: 5′-nuclease assay. *Genome Res* 2001;11:436–440.
78. Hsu TM, Law SM, Duan S, Neri B, Kwok P-Y. Genotyping single nucleotide polymorphisms by the Invader assay with dual-color fluorescence polarization detection. *Clin Chem* 2001;47:1373–1377.
79. Kaiser MW, Lyamicheva N, Ma W, Miller C, Neri B, Fors L, Lyamichev VI. A comparison of eubacterial and archaeal structure-specific 5′-exonucleases. *J Biol Chem* 1999;274: 21387–21394.
80. Kwiatkowksi RW, Lyamichev VI, de Arruda M, Neri B. Clinical, genetic, and pharmacogenetic applications of the Invader assay. *Mol Diagn* 1999;4:353–364.
81. Lyamichev VI, Kaiser MW, Lyamicheva NE, Vologodskii AV, Hall JG, Ma WP, Allawi HT, Neri B. Experimental and theoretical analysis of the invasive signal amplification reaction. *Biochemistry* 2000;39:9523–9532.
82. Hall JG, Eis PS, Law SM, Reynaldo LP, Prudent JR, Marshall DJ, Allawi HT, Mast AL, Dahlberg JE, Kwiatkowski RW, de Arruda M, Neri BP, Lyamichev VI. Sensitive detection of DNA polymorphisms by the serial invasive signal amplification reaction. *Proc Natl Acad Sci USA* 2000;97:8272–8277.
83. Sosnowski RG, Tu E, Butler WF, O'Connell JP, Heller MJ. Rapid determination of single base mismaatch mutations in DNA hybrids by electric field control. *Proc Nat Acad Sci USA* 1997;94:1119–1123.
84. Edman CF, Raymond DE, Wu DJ, Tu E, Sosnowski RG, Butler WF, Nerenberg M, Heller MJ. Electric field directed nucleic acid hybridization by electric fleld control. *Nucleic Acids Res* 1997;25:4907–4914.
85. Cornell BA, Braach-Maksvytis VL, King LG, Osman PD, Raguse B, Wieczorek L, Pace RJ. A biosensor that uses ion channel switches. *Nature* 1997;387:580–583.
86. Wang J, Cai X, Rivas G, Shiraishi H, Dontha N. Nucleic-acid immobilization, recognition and detection at chronopotentiometric DNAchips. *Biosens Bioelectron* 1997;12: 587–599.
87. Heller MJ, Forster AH, Tu E. Active microeletronic chip devices which utilize controlled electrophoretic fields for multiplex DNA hybridization and other genomic applications. *Electrophoresis* 2000;21:157–164.
88. Radtkey R, Feng L, Muralhidar M, Duhon M, Canter D, DiPierro D, Fallon S, Tu E, McElfresh K, Nerenberg M, Sosnowski R. Rapid, high fidelity analysis of simple sequence repeats on an electronically active DNA microchip. *Nucleic Acids Res* 2000;28:E17.
89. Westin L, Xu X, Miller C, Wang L, Edman CF, Nerenberg M. Anchored multiplex amplification on a microelectronic chip array. *Nat Biotechnol* 2000;18:199–204.
90. Todd AV, Fuery CJ, Impey HL, Applegate TL, Haughton MA. DzyNA-PCR; use of DNAzymes to detect and quantify nucleic acid sequences in a real-time fluorescent format. *Clin Chem* 2000;46:625–630.
91. Myakishev MV, Khripin Y, Hu S, Hamer DH. High-throughput SNP genotyping by allele-specific PCR with universal energy-transfer-labeled primers. *Genome Res* 2001;11:163–169.
92. Germer S, Higuchi R. Single-tube genotyping without oligonucleotide probes. *Genome Res* 1999;9:72–78.

93. Beaudet L, Bedard J, Breton B, Mercuri RJ, Budarf ML. Homogeneous assays for single-nucleotide polymorphism typing using AlphaScreen. *Genome Res* 2001;11:600–608.
94. Nazarenko IA, Bhatnagar SK, Hohman RJ. A closed tube format for amplification and detection of DNA based on energy transfer. *Nucleic Acids Res* 1997;25:2516–2521.
95. Whitcombe D, Theaker J, Guy SP, Brown T, Little S. Detection of PCR products using self-probing amplicons and fluorescence. *Nat Biotechnol* 1999;17:804–807.
96. Pastinen T, Raitio M, Lindroos K, Tainola P, Peltonen L, Syvanen AC. A system for specific, high-throughput genotyping by allele-specific primer extension on microarrays. *Genome Res* 2000;10:1031–1042.
97. Tonisson N, Kurg A, Kaasik K, Lohmussaar E, Metspalu A. Unravelling genetic data by arrayed primer extension. *Clin Chem Lab Med* 2000;38:165–170.
98. Shumaker JM, Metspalu A, Caskey CT. Mutation detection by solid phase primer extension. *Hum Mutat* 1996;7:346–354.
99. Pastinen T, Kurg A, Metspalu A, Peltonen L, Syvanen AC. Minisequencing: a specific tool for DNA analysis and diagnostics on oligonucleotide arrays. *Genome Res* 1997;7:606–614.
100. Dubiley S, Kirillov E, Mirzabekov A. Polymorphism analysis and gene detection by minisequencing on an array of gel-immobilized primers. *Nucleic Acids Res* 1999;27:E19.
101. Hsu TM, Chen X, Duan S, Miller R, Kwok P-Y. A universal SNP genotyping assay with fluorescence polarization detection. *BioTechniques* 2001;31:560–570.
102. Chen X, Kwok P-Y. Template-directed dye-terminator incorporation (TDI) assay: a homogeneous DNA diagnostic method based on fluorescence energy transfer. *Nucleic Acids Res* 1997;25:347–353.
103. Chen X, Zehnbauer B, Gnirke A, Kwok P-Y. Fluorescence energy transfer detection as a homogeneous DNA diagnostic method. *Proc Nat Acad Sci USA* 1997;94: 10756–10761.
104. Chen X, Levine L, Kwok P-Y. Fluorescence polarization in homogeneous nucleic acid analysis. *Genome Res* 1999;9:492–498.
105. Buetow KH, Edmonson M, MacDonald R, Clifford R, Yip P, Kelley J, Little DP, Strausberg R, Koester H, Cantor CR, Braun A. High-throughput development and characterization of a genome wide collection of gene-based single nucleotide polymorphism markers by chip-based matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Proc Nat Acad Sci USA* 1996;98:581–584.
106. Ross P, Hall L, Smirnov I, Haff L. High level multiplex genotyping by MALDI-TOF mass spectrometry. *Nature Biotechnol* 1998;16:1347–1351.
107. Sauer S, Lechner D, Berlin K, Lehrach H, Escary JL, Fox N, Gut IG. A novel procedure for efficient genotyping of single nucleotide polymorphisms. *Nucleic Acids Res* 2000;28:E13.
108. Braun A, Little DP, Koster H. Detecting CFTR gene mutations by using primer oligo base extension and mass spectrometry. *Clin Chem* 1997;43:1151–1158.
109. Nordstrom T, Nourizad K, Ronaghi M, Nyren P. Method enabling pyrosequencing on double-stranded DNA. *Anal Biochem* 2000;282:186–193.
110. Ahmadian A, Gharizadeh B, Gustafsson AC. Single-nucleotide polymorphism analysis by pyrosequencing. *Anal Biochem* 2000;280:103–110.
111. Ronaghi M. Pyrosequencing sheds light on DNA sequencing. *Genome Res* 2001;11:3–11.

112. Alderborn A, Kristofferson A, Hammerling U. Determination of single-nucleotide polymorphisms by real-time pyrophosphate DNA sequencing. *Genome Res* 2000;10: 1249–1258.

113. Nyren P, Pettersson B, Uhlen M. Solid phase DNA minisequencing by an enzymatic luminometric inorganic pyrophosphate detection assay. *Anal Biochem* 1993;208: 171–175.

114. Fan JB, Chen X, Halushka MK, Berno A, Huang X, Ryder T, Lipshutz RJ, Lockhart DJ, Chakravarti A. Parallel genotyping of human SNPs using generic high-density oligonucleotide tag arrays. *Genome Res* 2000;10:853–860.

115. Chen J, Iannone MA, Li MS, Taylor JD, Rivers P, Nelsen AJ, Slentz-Kesler KA, Roses A, Weiner MP. A microsphere-based assay for multiplexed single nucleotide polymorphism analysis using single base chain extension. *Genome Res* 2000;10:549–557.

116. Cai H, White PS, Torney D, Deshpande A, Wang Z, Keller RA, Marrone B, Nolan JP. Flow cytometry-based minisequencing: a new platform for high-throughput single-nucleotide polymorphism scoring. *Genomics* 2000;66:135–143.

117. Hirschhorn JN, Sklar P, Lindblad-Toh K, Lim YM, Ruiz-Gutierrez M, Bolk S, Langhorst B, Schaffner S, Winchester E, Lander ES. SBE-TAGS: an array-based method for efficient single-nucleotide polymorphism genotyping. *Proc Natl Acad Sci USA* 2000; 97:12164–12169.

118. Nickerson DA, Kaiser R, Lappin S, Stewart J, Hood L, Landegren U. Automated DNA diagnostics using an ELISA-based oligonucleotide ligation assay. *Proc Natl Acad Sci USA* 1990;87:8923–8927.

119. Grossman PD, Bloch W, Brinson E, Chang CC, Eggerding FA, Fung S, Iovannisci DM, Woo S, Winn-Deen ES. High-density multiplex detection of nucleic acid sequences: oligonucleotide ligation assay and sequence-coded separation. *Nucleic Acids Res* 1994;22:4527–4534.

120. Samiotaki M, Kwiatkowski M, Parik J, Landegren U. Dual-color detection of DNA sequence variants by ligase-mediated analysis. *Genomics* 1994;20:238–242.

121. Baner J, Nilsson M, Mendel-Hartvig M, Landegren U. Signal amplification of padlock probes by rolling circle replication. *Nucleic Acids Res* 1998;26:5073–5078.

122. Lizardi PM, Huang X, Zhu Z, Bray-Ward P, Thomas DC, Ward DC. Mutation detection and single-molecule counting using isothermal rolling-cycle amplification. *Nat Genet* 1998;19:225–232.

123. Nilsson M, Malmgren H, Samiotaki M, Kwiatkowski M, Chowdhary BP, Landegren U. Padlock probes: circularizing oligonucleotides for localized DNA detection. *Science* 1994;265:2085–2088.

124. Zhang DY, Brandwein M, Hsuih TC, Li H. Amplification of target-specific, ligation-dependent circular probe. *Gene* 1998;211:277–285.

125. Chen X, Livak KJ, Kwok PY. A homogeneous, ligase-mediated DNA diagnostic test. *Genome Res* 1998;8:549–556.

126. Gerry NP, Witowski NE, Day J, Hammer RP, Barany G, Barany F. Universal DNA microarray method for multiplex detection of low abundance point mutations. *J Mol Biol* 1999;292:251–262.

127. Wallace RB, Shaffer J, Murphy RF, Bonner J, Hirose T, Itakura K. Hybridization of synthetic oligodeoxyribonucleotides to phi chi 174 DNA: the effect of single base pair mismatch. *Nucleic Acids Res* 1979;6:3543–3557.

128. Conner BJ, Reyes AA, Morin C, Itakura K, Teplitz RL, Wallace RB. Detection of sickle cell beta S-globin allele by hybridization with synthetic oligonucleotides. *Proc Natl Acad Sci USA* 1983;80:278–282.

129. Bonnet G, Tyagi S, Libchaber A, Kramer FR. Thermodynamic basis of the enhanced specificity of structured DNA probes. *Proc Natl Acad Sci USA* 1999;96:6171–6176.
130. Tyagi S, Kramer FR. Molecular beacons: probes that fluoresce upon hybridization. *Nat Biotechnol* 1996;14:303–308.
131. Kuimelis RG, Livak KJ, Mullah B, Andrus A. Structural analogues of TaqMan probes for real-time quantitative PCR. *Nucleic Acids Symp Ser* 1997;255–256.
132. Duggan DJ, Bittner M, Chen Y, Meltzer P, Trent JM. Expression profiling using cDNA microarrays. *Nat Genet* 1999;21:10–14.
133. Maughan NJ, Lewis FA, Smith V. An introduction to arrays. *J Pathol* 2001;195:3–6.
134. Saiki RK, Chang CA, Levenson CH, Warren TC, Boehm CD, Kazazian HH Jr, Erlich HA. Diagnosis of sickle cell anemia and beta-thalassemia with enzymatically amplified DNA and nonradioactive allele-specific oligonucleotide probes. *N Engl J Med* 1988; 319:537–541.
135. Blohm DH, Guiseppi-Elie A. New developments in microarray technology. *Curr Opin Biotechnol* 2001;12:41–47.
136. Prince JA, Feuk L, Howell WM, Jobs M, Emahazion T, Blennow K, Brookes AJ. Robust and accurate single nucleotide polymorphism genotyping by dynamic allele-specific hybridization (DASH): design criteria and assay validation. *Genome Res* 2001; 11:152–162.
137. Ross PL, Lee K, Belgrader P. Discrimination of single-nucleotide polymorphisms in human DNA using peptide nucleic acid probes detected by MALDI-TOF mass spectrometry. *Anal Chem* 1997;69:4197–4202.
138. Griffin TJ, Tang W, Smith LM. Genetic analysis by peptide nucleic acid affinity MALDI-TOF mass spectrometry. *Nat Biotechnol* 1997;15:1368–1372.
139. Orum H. Purification of nucleic acids by hybridization to affinity tagged PNA probes. *Curr Issues Mol Biol* 1999;1:105–110.
140. Orum H, Jakobsen MH, Koch T, Vuust J, Borre MB. Detection of the factor V Leiden mutation by direct allele-specific hybridization of PCR amplicons to photoimmobilized locked nucleic acids. *Clin Chem* 1999;45:1898–1905.
141. Newton CR, Graham A, Heptinstall LE, Powell SJ, Summers C, Kalsheker N, Smith JC, Markham AF. Analysis of any point mutation in DNA. The amplification refractory mutation system (ARMS). *Nucleic Acids Res* 1989;17:2503–2516.
142. Day DJ, Speiser PW, White PC, Barany F. Detection of steroid 21-hydroxylase alleles using gene-specific PCR and a multiplexed ligation detection reaction. *Genomics* 1995;29:152–162.
143. Khanna M, Park P, Zirvi M, Cao W, Picon A, Day J, Paty P, Barany F. Multiplex PCR/LDR for detection of K-ras mutations in primary colon tumors. *Oncogene* 1999;18:27–38.
144. Broude NE, Woodward K, Cavallo R, Cantor CR, Englert D. DNA microarrays with stem-loop DNA probes: preparation and applications. *Nucleic Acids Res* 2001;29:E92.
145. Faruqi AF, Hosono S, Driscoll MD, Dean FB, Alsmadi O, Bandaru R, Kumar G, Grimwade B, Zong Q, Sun Z, Du Y, Kingsmore S, Knott T, Lasken RS. High-throughput genotyping of single nucleotide polymorphisms with rolling circle amplification. *BMC Genomics* 2001;2:4.
146. Nilsson M, Krejci K, Koch J, Kwiatkowski M, Gustavsson P, Landegren U. Padlock probes reveal single-nucleotide differences, parent of origin and *in situ* distribution of centromeric sequences in human chromosomes 13 and 21. *Nat Genet* 1997;16:252–255.
147. Fire A, Xu SQ. Rolling replication of short DNA circles. *Proc Natl Acad Sci USA* 1995;92:4641–4645.

148. Iannone MA, Taylor JD, Chen J, Li MS, Rivers P, Slentz-Kesler KA, Weiner MP. Multiplexed single nucleotide polymorphism genotyping by oligonucleotide ligation and flow cytometry. *Cytometry* 2000;39:131–140.

149. Lyamichev V, Mast AL, Hall JG, Prudent JR, Kaiser MW, Takova T, Kwiatkowski RW, Sander TJ, de Arruda M, Arco DA, Neri BP, Brow MA. Polymorphism identification and quantitative detection of genomic DNA by invasive cleavage of oligonucleotide probes. *Nat Biotechnol* 1999;17:292–296.

150. Mein CA, Barratt BJ, Dunn MG, Siegmund T, Smith AN, Esposito L, Nutland S, Stevens HE, Wilson AJ, Phillips MS, Jarvis N, Law S, de Arruda M, Todd JA. Evaluation of single nucleotide polymorphism typing with invader on PCR amplicons and its automation. *Genome Res* 2000;10:330–343.

151. Wilkins Stevens P, Hall JG, Lyamichev V, Neri BP, Lu M, Wang L, Smith LM, Kelso DM. Analysis of single nucleotide polymorphisms with solid phase invasive cleavage reactions. *Nucleic Acids Res* 2001;29:E77.

152. Lockhart DJ, Winzeler EA. Genomics, gene expression and DNA arrays. *Nature* 2000;405:827–836.

153. Green CD, Simons JF, Taillon BE, Lewin DA. Open systems: panoramic views of gene expression. *J Immunol Methods* 2001;250:67–79.

154. Mockler TC, Ecker JR. Applications of DNA tiling arrays for whole-genome analysis. *Genomics* 2005;85:1–15.

155. Ross JS, Schenkein DP, Kashala O, Linette GP, Stec J, Symmans WF, Pusztai L, Hortobagyi GN. Pharmacogenomics. *Adv Anat Pathol* 2004;11:211–220.

156. Shi L, Tong W, Goodsaid F, Frueh FW, Fang H, Han T, Fuscoe JC, Casciano DA. QA/QC: challenges and pitfalls facing the microarray community and regulatory agencies. *Expert Rev Mol Diagn* 2004;4:761–777.

157. Liang P, Pardee AB. Differential display of eukaryotic messenger RNA by means of the polymerase chain reaction. *Science* 1992;257:967–971.

158. Welsh J, Chada K, Dalal SS, Cheng R, Ralph D, McClelland M. Arbitrarily primed PCR fingerprinting of RNA. *Nucleic Acids Res* 1992;20:4965–4970.

159. Bertelsen AH, Velculescu VE. High-throughput gene expression analysis using SAGE. *Drug Discov Today* 1998;3:152–159.

160. Velculescu VE. Serial analysis of gene expression. *Science* 1995;270:484–487.

161. Hubank M, Schatz DG. Identifying differences in mRNA expression by representational difference analysis of cDNA. *Nucleic Acids Res* 1994;22:5640–5648.

162. Shimkets RA, Lowe DG, Tai JT, Sehl P, Jin H, Yang R, Predki PF, Rothberg BE, Murtha MT, Roth ME, Shenoy SG, Windemuth A, Simpson JW, Simons JF, Daley MP, Gold SA, McKenna MP, Hillan K, Went GT, Rothberg JM. Gene expression analysis by transcript profiling coupled to a gene database query. *Nat Biotechnol* 1999;17:798–803.

163. Prashar Y, Weissman SM. READS: a method for display of 3′-end fragments of restriction enzyme-digested cDNAs for analysis of differential gene expression. *Methods Enzymol* 1999;303:258–272.

164. Sutcliffe JG, Foye PE, Erlander MG, Hilbush BS, Bodzin LJ, Durham JT, Hasel KW. TOGA: an automated parsing technology for analyzing expression of nearly all genes. *Proc Natl Acad Sci USA* 2000;97:1976–1981.

165. Spinella DG, Bernardino AK, Redding AC, Koutz P, Wei Y, Pratt EK, Myers KK, Chappell G, Gerken S, McConnell SJ. Tandem arrayed ligation of expressed sequence tags (TALEST): a new method for generating global gene expression profiles. *Nucleic Acids Res* 1999;27:E22.

166. Pandey A, Mann M. Proteomics to study genes and genomes. *Nature* 2000;405: 837–846.
167. Hunter TC, Andon NL, Koller A, Yates JR, Haynes PA. The functional proteomics toolbox: methods and applications. *J Chromatogr B Analyt Technol Biomed Life Sci* 2002;782:165–181.
168. Zhu H, Bilgin M, Snyder M. Proteomics. *Annu Rev Biochem* 2003;72:783–812.
169. Tonge R, Shaw J, Middleton B, Rowlinson R, Rayner S, Young J, Pognan F, Hawkins E, Currie I, Davison M. Validation and development of fluorescence two-dimensional differential gel electrophoresis proteomics technology. *Proteomics* 2001;1:377–396.
170. Unlu M, Morgan ME, Minden JS. Difference gel electrophoresis: a single gel method for detecting changes in protein extracts. *Electrophoresis* 1997;18:2071–2077.
171. Templin MF, Stoll D, Schrenk M, Traub PC, Vohringer CF, Joos TO. Protein microarray technology. *Trends Biotechnol* 2002;20:160–166.
172. Zhu H, Snyder M. Protein arrays and microarrays. *Curr Opin Chem Biol* 2001;5:40–45.
173. Zhu H, Snyder M. "Omic" approaches for unraveling signaling networks. *Curr Opin Cell Biol* 2002;14:173–179.
174. Seetharaman S, Zivarts M, Sudarsan N, Breaker RR. Immobilized RNA switches for the analysis of complex chemical and biological mixtures. *Nat Biotechnol* 2001;19:336–341.
175. Lueking A, Horn M, Eickhoff H, Bussow K, Lehrach H, Walter G. Protein microarrays for gene expression and antibody screening. *Anal Biochem* 1999;270:103–111.
176. Wiese R, Belosludtsev Y, Powdrill T, Thompson P, Hogan M. Simultaneous multianalyte ELISA performed on a microarray platform. *Clin Chem* 2001;47:1451–1457.
177. Moody MD, Van Arsdell SW, Murphy KP, Orencole SF, Burns C. Array-based ELISAs for high-throughput analysis of human cytokines. *Biotechniques* 2001;31: 186–190,192–194.
178. Huang RP, Huang R, Fan Y, Lin Y. Simultaneous detection of multiple cytokines from conditioned media and patient's sera by an antibody-based protein array system. *Anal Biochem* 2001;294:55–62.
179. Sreekumar A, Nyati MK, Varambally S, Barrette TR, Ghosh D, Lawrence TS, Chinnaiyan AM. Profiling of cancer cells using protein microarrays: discovery of novel radiation-regulated proteins. *Cancer Res* 2001;61:7585–7593.
180. Robinson WH, DiGennaro C, Hueber W, Haab BB, Kamachi M, Dean EJ, Fournel S, Fong D, Genovese MC, de Vegvar HE, Skriner K, Hirschberg DL, Morris RI, Muller S, Pruijn GJ, van Venrooij WJ, Smolen JS, Brown PO, Steinman L, Utz PJ. Autoantigen microarrays for multiplex characterization of autoantibody responses. *Nat Med* 2002;8:295–301.
181. Hiller R, Laffer S, Harwanegg C, Huber M, Schmidt WM, Twardosz A, Barletta B, Becker WM, Blaser K, Breiteneder H, Chapman M, Crameri R, Duchene M, Ferreira F, Fiebig H, Hoffmann-Sommergruber K, King TP, Kleber-Janke T, Kurup VP, Lehrer SB, Lidholm J, Muller U, Pini C, Reese G, Scheiner O, Scheynius A, Shen HD, Spitzauer S, Suck R, Swoboda I, Thomas W, Tinghino R, Van Hage-Hamsten M, Virtanen T, Kraft D, Muller MW, Valenta R. Microarrayed allergen molecules: diagnostic gatekeepers for allergy treatment. *FASEB J* 2002;16:414–416.
182. Joos TO, Schrenk M, Hopfl P, Kroger K, Chowdhury U, Stoll D, Schorner D, Durr M, Herick K, Rupp S, Sohn K, Hammerle H. A microarray enzyme-linked immunosorbent assay for autoimmune diagnostics. *Electrophoresis* 2000;21:2641–2650.
183. Davies H, Lomas L, Austen B. Profiling of amyloid beta peptide variants using SELDI Protein Chip arrays. *Biotechniques* 1999;27:1258–1261.

184. Nelson RW. The use of bioreactive probes in protein characterization. *Mass Spectrom Rev* 1997;16:353–376.

185. Xiao Z, Prieto D, Conrads TP, Veenstra TD, Issaq HJ. Proteomic patterns: their potential for disease diagnosis. *Mol Cell Endocrinol* 2005;230:95–106.

186. Semmes OJ, Feng Z, Adam BL, Banez LL, Bigbee WL, Campos D, Cazares LH, Chan DW, Grizzle WE, Izbicka E, Kagan J, Malik G, McLerran D, Moul JW, Partin A, Prasanna P, Rosenzweig J, Sokoll LJ, Srivastava S, Thompson I, Welsh MJ, White N, Winget M, Yasui Y, Zhang Z, Zhu L. Evaluation of serum protein profiling by surface-enhanced laser desorption/ionization time-of-flight mass spectrometry for the detection of prostate cancer: I. Assessment of platform reproducibility. *Clin Chem* 2005;51:102–112.

187. Gygi SP, Rist B, Gerber SA, Turecek F, Gelb MH, Aebersold R. Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat Biotechnol* 1999;17:994–999.

188. Han DK, Eng J, Zhou H, Aebersold R. Quantitative profiling of differentiation-induced microsomal proteins using isotope-coded affinity tags and mass spectrometry. *Nat Biotechnol* 2001;19:946–951.

189. Michaud GA, Snyder M. Proteomic approaches for the global analysis of proteins. *Biotechniques* 2002;33:1308–1316.

190. Smolka MB, Zhou H, Purkayastha S, Aebersold R. Optimization of the isotope-coded affinity tag-labeling procedure for quantitative proteome analysis. *Anal Biochem* 2001;297:25–31.

191. Uetz P, Giot L, Cagney G, Mansfield TA, Judson RS, Knight JR, Lockshon D, Narayan V, Srinivasan M, Pochart P, Qureshi-Emili A, Li Y, Godwin B, Conover D, Kalbfleisch T, Vijayadamodar G, Yang M, Johnston M, Fields S, Rothberg JM. A comprehensive analysis of protein–protein interactions in *Saccharomyces cerevisiae*. *Nature* 2000;403: 623–627.

192. Vidal M, Endoh H. Prospects for drug screening using the reverse two-hybrid system. *Trends Biotechnol* 1999;17:374–381.

193. Nicholson JK, Wilson ID. Opinion: understanding "global" systems biology: metabonomics and the continuum of metabolism. *Nat Rev Drug Discov* 2003;2:668–676.

194. Fiehn O. Metabolomics—the link between genotypes and phenotypes. *Plant Mol Biol* 2002;48:155–171.

195. Nicholson JK, Connelly J, Lindon JC, Holmes E. Metabonomics: a platform for studying drug toxicity and gene function. *Nat Rev Drug Discov* 2002;1:153–161.

196. Kell DB. Metabolomics and systems biology: making sense of the soup. *Curr Opin Microbiol* 2004;7:296–307.

197. Fiehn O, Kloska S, Altmann T. Integrated studies on plant biology using multiparallel techniques. *Curr Opin Biotechnol* 2001;12:82–86.

198. Cornish-Bowden A, Cardenas ML. From genome to cellular phenotype—a role for metabolic flux analysis? *Nat Biotechnol* 2000;18:267–268.

199. Saghatelian A, Cravatt BF. Discovery metabolite profiling—forging functional connections between the proteome and metabolome. *Life Sci* 2005;77:1759–1766.

200. Trethewey RN. Metabolite profiling as an aid to metabolic engineering in plants. *Curr Opin Plant Biol* 2004;7:196–201.

201. Saghatelian A, Cravatt BF. Global strategies to integrate the proteome and metabolome. *Curr Opin Chem Biol* 2005;9:62–68.

202. Fernie AR, Trethewey RN, Krotzky AJ, Willmitzer L. Metabolite profiling: from diagnostics to systems biology. *Nat Rev Mol Cell Biol* 2004;5:763–769.

203. Bino RJ, Hall RD, Fiehn O, Kopka J, Saito K, Draper J, Nikolau BJ, Mendes P, Roessner-Tunali U, Beale MH, Trethewey RN, Lange BM, Wurtele ES, Sumner LW. Potential of metabolomics as a functional genomics tool. *Trends Plant Sci* 2004;9:418–425.
204. Jenkins H, Hardy N, Beckmann M, Draper J, Smith AR, Taylor J, Fiehn O, Goodacre R, Bino RJ, Hall R, Kopka J, Lane GA, Lange BM, Liu JR, Mendes P, Nikolau BJ, Oliver SG, Paton NW, Rhee S, Roessner-Tunali U, Saito K, Smedsgaard J, Sumner LW, Wang T, Walsh S, Wurtele ES, Kell DB. A proposed framework for the description of plant metabolomics experiments and their results. *Nat Biotechnol* 2004;22:1601–1606.
205. Raamsdonk LM, Teusink B, Broadhurst D, Zhang N, Hayes A, Walsh MC, Berden JA, Brindle KM, Kell DB, Rowland JJ, Westerhoff HV, van Dam K, Oliver SG. A functional genomics strategy that uses metabolome data to reveal the phenotype of silent mutations. *Nat Biotechnol* 2001;19:45–50.
206. Watkins SM, German JB. Toward the implementation of metabolomic assessments of human health and nutrition. *Curr Opin Biotechnol* 2002;13:512–516.
207. Shockcor JP, Holmes E. Metabonomic applications in toxicity screening and disease diagnosis. *Curr Top Med Chem* 2002;2:35–51.
208. Tweeddale H, Notley-McRobb L, Ferenci T. Effect of slow growth on metabolism of *Escherichia coli*, as revealed by global metabolite pool ("metabolome") analysis. *J Bacteriol* 1998;180:5109–5116.
209. Oliver SG, Winson MK, Kell DB, Baganz F. Systematic functional analysis of the yeast genome. *Trends Biotechnol* 1998;16:373–378.
210. Bochner BR, Gadzinski P, Panomitros E. Phenotype microarrays for high-throughput phenotypic testing and assay of gene function. *Genome Res* 2001;11:1246–1255.
211. Hanlon EB, Manoharan R, Koo TW, Shafer KE, Motz JT, Fitzmaurice M, Kramer JR, Itzkan I, Dasari RR, Feld MS. Prospects for *in vivo* Raman spectroscopy. *Phys Med Biol* 2000;45:R1–59.
212. Rees JL. The genetics of sun sensitivity in humans. *Am J Hum Genet* 2004;75:739–751.
213. Smith MW, Dean M, Carrington M, Winkler C, Huttley GA, Lomb DA, Goedert JJ, O'Brien TR, Jacobson LP, Kaslow R, Buchbinder S, Vittinghoff E, Vlahov D, Hoots K, Hilgartner MW, O'Brien SJ. Contrasting genetic influence of CCR2 and CCR5 variants on HIV-1 infection and disease progression. Hemophilia Growth and Development Study (HGDS), Multicenter AIDS Cohort Study (MACS), Multicenter Hemophilia Cohort Study (MHCS), San Francisco City Cohort (SFCC), ALIVE Study. *Science* 1997;277:959–965.
214. Chen J, Giovannucci EL, Hunter DJ. MTHFR polymorphism, methyl-replete diets and the risk of colorectal carcinoma and adenoma among U.S. men And women: an example of gene–environment interactions in colorectal tumorigenesis. *J Nutr* 1999;129: 560S–564S.
215. Chen J, Stampfer MJ, Hough HL, Garcia-Closas M, Willett WC, Hennekens CH, Kelsey KT, Hunter DJ. A prospective study of *N*-acetyltransferase genotype, red meat intake, and risk of colorectal cancer. *Cancer Res* 1998;58:3307–3311.
216. Bloemenkamp KW, Rosendaal FR, Helmerhorst FM, Buller HR, Vandenbroucke JP. Enhancement by factor V Leiden mutation of risk of deep-vein thrombosis associated with oral contraceptives containing a third-generation progestagen. *Lancet* 1995;346: 1593–1596.
217. Chan AT, Tranah GJ, Giovannucci EL, Hunter DJ, Fuchs CS. Genetic variants in the UGT1A6 enzyme, aspirin use, and the risk of colorectal adenoma. *J Natl Cancer Inst* 2005;97:457–460.
218. Lehtimaki T, Moilanen T, Porkka K, Akerblom HK, Ronnemaa T, Rasanen L, Viikari J, Ehnholm C, Nikkari T. Association between serum lipids and apolipoprotein E phenotype

is influenced by diet in a population-based sample of free-living children and young adults: the Cardiovascular Risk in Young Finns Study. *J Lipid Res* 1995;36:653–661.

219. Hines LM, Stampfer MJ, Ma J, Gaziano JM, Ridker PM, Hankinson SE, Sacks F, Rimm EB, Hunter DJ. Genetic variation in alcohol dehydrogenase and the beneficial effect of moderate alcohol consumption on myocardial infarction. *N Engl J Med* 2001;344: 549–555.
220. Memisoglu A, Hu FB, Hankinson SE, Manson JE, De Vivo I, Willett WC, Hunter DJ. Interaction between a peroxisome proliferator-activated receptor gamma gene polymorphism and dietary fat intake in relation to body mass. *Hum Mol Genet* 2003;12:2923–2929.
221. Maier LA. Genetic and exposure risks for chronic beryllium disease. *Clin Chest Med* 2002;23:827–839.
222. Weinshilboum R. Thiopurine pharmacogenetics: clinical and molecular studies of thiopurine methyltransferase. *Drug Metab Dispos* 2001;29:601–605.
223. Israel E, Drazen JM, Liggett SB, Boushey HA, Cherniack RM, Chinchilli VM, Cooper DM, Fahy JV, Fish JE, Ford JG, Kraft M, Kunselman S, Lazarus SC, Lemanske RF, Martin RJ, McLean DE, Peters SP, Silverman EK, Sorkness CA, Szefler SJ, Weiss ST, Yandava CN. The effect of polymorphisms of the beta(2)-adrenergic receptor on the response to regular use of albuterol in asthma. *Am J Respir Crit Care Med* 2000;162:75–80.
224. Weatherall DJ. Phenotype–genotype relationships in monogenic disease: lessons from the thalassaemias. *Nat Rev Genet* 2001;2:245–255.
225. Huisman TH, Carver MFH., Baysai E. *A Syllabus of Thalassemia Mutations*. Augusta, GA: The Sickle Cell Anemia Foundation; 1997.
226. Weatherall DJ, Clegg JB. *The Thalassemia Syndromes*. Oxford, UK: Blackwell Science; 2001.
227. Whittaker M, Britten JJ, Dawson PJG. Comparison of a commercially available assay system with two reference methods for the determination of cholinesterase variants. *Clin Chem* 1983;29:1746–1751.
228. Weinshilbom RM, Raymond FA. Inheritance of low erythrocyte catechol-O-methyltransferase activity in man. *Am J Hum Genet* 1977;29:125–135.
229. Kouri RE, McKinney CE, Slomiany DJ, Snodgrass DR, Wray NP, McLemore TL. Positive correlation between high aryl hydrocarbon hydroxylase activity and primary lung cancer as analysed in cryopreserved lymphocytes. *Cancer Res* 1982;42:5030–5037.
230. Fuhr U, Rost K. Simple and reliable CYP1A2 phenotyping by the paraxanthine/caffeine ration in plasma and saliva. *Pharmacogenetics* 1994;4:109–116.
231. Campbell ME, Spielberg SP, Kalow W. A urinary metabolite ratio that reflects systemic caffeine clearance. *Clin Pharmacol Ther* 1987;42:157–165.
232. Cholerton S, Idle ME. Comparison of a novel thin-layer chromatographic-flourescence detection method with a spectrofluorometric method for the determination of 7-hydroxycoumarin in human urine. *J Chromatogr* 1992;575:325–330.
233. Veronese ME, Miners JO, Randles D, Gregov D, Birkett DJ. Validation of the tolbutamide metabolic ratio for population screening with the use of sulfaphenazole to produce model phenotypic poor metabolizers. *Clin Pharmacol Ther* 1990;47:403–411.
234. Wedlund PJ, Aslanian WS, McAllister CB, Wilkinson GR, Branch RA. Mephenytoin hydroxylation deficiency in Caucasians: frequency of a new oxidative drug metabolism polymorphism. *Clin Pharmacol Ther* 1984;36:773–780.
235. Chang M, Dahl ML, Tybring G, Gotharson E, Bertilsson L. Use of omeprazole as a probe drug for CYP2C19 phenotype in Swedish Caucasians: comparison with *S*-

methylphenytoin hydroxylation phenotype and CYP2C19 genotype. *Pharmacogenetics* 1995;5:358–363.

236. Idle JR, Mahgoub A, Angelo MM, Dring LG, Lancaster R, Smith RL. The metabolsim of [$^{14}$C] debrisoquine in man. *Br J Clin Pharmacol* 1979;7:257–266.
237. Eichelbaum M, Spannbrucker N, Steincke B, Dengler HJ. Defective N-oxidation of sparteine in man: a new pharmacogenetic defect. *Eur J Clin Pharmacol* 1979;17:153–155.
238. Jacqz-Aigrain E, Menard Y, Popon M, Mathieu H. Dextromethorphan phenotypes determined by high-performance liquid chromatography and fluorescence detection. *J Chromatogr* 1989;495:361–363.
239. Lennard MS. Quantitative-analysis of metoprolol and 3 of its metabolites in urine and liver-microsomes by high-performance liquid-chromatography. *J Chromatogr* 1985;342: 199–205.
240. Kim RB, O'Shea D, Wilkinson GR. Interindividual variability of chlorzoxazone 6-hydroxylation in men and women and its relationship to CYP2E1 genetic polymorphisms. *Clin Pharmacol Ther* 1995;57:645–655.
241. Raucy JL, Schultz ED, Wester MR, Arora S, Johnson DE, Omdahl JL, Carpenter SP. Human lymphocyte cytochrome P450 2E1, a putative marker for alcohol-mediated changes in hepatic chlorozoxazone activity. *Drug Metab Dispos* 1997;25:1429–1435.
242. Thummel KE, Shen DD, Podoll TD, Kunze KL, Trager WF, Hartwell PS, Raisys VA, Marsh CL, McVicar JP, Barr DM, Perkins JD, Carithers RL. Use of midazolam as a human cytochrome-P450 3A pobe. I. *In-vitro*, *in-vivo* correlations in liver-transplant patients. *J Pharmacol Exp Ther* 1994;271:549–556.
243. Zhiri A, Mayer HA, Michaux V. 6-Beta-hydroxycortisol in serum and urine as determined by enzyme-immunoassay on microtitre plates. *Clin Chem* 1986;32:2094–2097.
244. Watkins PB, Hamilton TA, Annesley TM, Ellis CN, Kolars JC, Voorhees JJ. The erythromycin breath test as a predictor of cyclosporine blood-levels. *Clin Pharmacol Ther* 1990;48:120–129.
245. Alwaiz M, Ayesh R, Mitchell SC, Idle ME, Smith RL. Trimethylaminuria—the detection of carriers using a trimethylamine load test. *J Inherit Metab Dis* 1989;12:80–85.
246. Sieidegard J, Pero RW. The hereditary transmission of high glutathione transferase-activity towards *trans*-stilbene oxide in human mononuclear leukocytes. *Hum Genet* 1985;69:66–68.
247. Hallier E, Langhof T, Dannaappel D, Leutbecher M, Schroder K, Goergens HW, Muller A, Bolt HM. Polymorphism of glutathione conjugation of methyl-bromide, ethylene-oxide and dichloromethane in human blood—influence on the induction of sister chromatid exchanges (SCE) in lymphocytes. *Arch Toxicol* 1993;67:173–178.
248. Cribb AE, Isbrucker R, Levatte T, Tsui B, Gillespie CT. Acetylator phenotyping: the urinary caffeine metabolite ratio in slow acetylators correlates with a marker of systemic NAT1 activity. *Pharmacogenetics* 1994;4:166–170.
249. Grant DM, Tang BK, Kalow W. Polymorphic N-acetylation of a caffeine metabolite. *Clin Pharmacol Ther* 1983;33:355–359.
250. Playfer JR, Eze LC, Bullen MF, Evans DAP. Genetic polymorphism and inter-ethnic variability of plasma paraozonase activity. *J Med Genet* 1976;13:337–342.
251. Price RA, Cox NJ, Spielman RS, Van Loon JA, Maidak BL, Weinshilbom RM. Inheritance of human platelet thermolabile phenol sulphotransferase (TLPST) activity. *Genet Epidemiol* 1988;5:1–15.

252. Price RA, Spielman RS, Lucena AL, Van Loon JA, Maidak BL, Weinshilbom RM. Genetic polymorphism for human platelet thermostable phenol sulphotransferase (TSPST) activity. *Genetics* 1989;122:905–914.
253. Weinshilbom RM, Sladek SL. Mercaptopurine pharmacogenetics: monogenic inheritance of erythrocyte thiopurine methyltransferase activity. *Am J Hum Genet* 1980;32:651–662.
254. Greiner B, Eichelbaum M, Fritz P, Kreichgauer HP, VonRichter O, Zundler J, Kroemer HK. The role of intestinal P-glycoprotein in the interaction of digoxin and rifampin. *J Clin Invest* 1999;104:147–153.
255. Thomas DC. Genetic epidemiology with a capital "E." *Genet Epidemiol* 2000;19:289–300.
256. Botto LD, Khoury MJ. Commentary: facing the challenge of gene–environment interaction: the two-by-four table and beyond. *Am J Epidemiol* 2001;153:1016–1020.
257. Khoury MJ, Adams MJ Jr, Flanders WD. An epidemiologic approach to ecogenetics. *Am J Hum Genet* 1988;42:89–95.
258. Ottman R. An epidemiologic approach to gene–environment interaction. *Genet Epidemiol* 1990;7:177–185.
259. Rothman KJ, Greenland S. *Modern Epidemiology*. Philadelphia: Lippincott-Raven; 1998.
260. Rebbeck TR, Spitz M, Wu X. Assessing the function of genetic variants in candidate gene association studies. *Nat Rev Genet* 2004;5:589–597.
261. Ge H, Walhout AJ, Vidal M. Integrating "omic" information: a bridge between genomics and systems biology. *Trends Genet* 2003;19:551–560.
262. Hartwell LH, Hopfield JJ, Leibler S, Murray AW. From molecular to modular cell biology. *Nature* 1999;402:C47–C52.

# GENOMICS

Dimitri Semizarov and Eric A. G. Blomme
*Global Pharmaceutical Research and Development, Abbott Laboratories, Abbott Park, Illinois*

**Contents**

1 Introduction
2 Gene Expression Microarrays
  2.1 Technology
  2.2 Applications of Microarrays in Preclinical Development
3 Comparative Genomic Hybridization
  3.1 Technology
  3.2 Applications of CGH in Preclinical Development
4 Gene Silencing
  4.1 Antisense Oligonucleotides and Ribozymes
  4.2 Short Interfering RNA
5 Conclusion
  References

## 1 INTRODUCTION

In the past five to ten years a growing number of new drugs have been discovered using a target-based approach [1], implying a significant paradigm shift in the pharmaceutical industry. This shift has been facilitated by the decoding of the complete sequence of the human genome [2–4]. Today, a significant portion of the research and development (R&D) effort in the pharmaceutical industry is focused on drug target identification and validation. This paradigm change also requires that new

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

methodologies be used in preclinical development. The preclinical development process for targeted agents involves genomic analysis of preclinical model systems to elucidate the mechanism of target inhibition, delineate off-target effects, and identify biomarkers.

Genomic technologies may potentially play an important role at all stages of drug discovery. For example, target identification in cancer often involves large scale genomic screens for genes or their products that alter cell proliferation and survival. Identification of genes frequently amplified at the chromosomal level may reveal a target for therapeutic intervention, whose inhibition by small molecules or antibodies may deprive the cancer cell of its proliferation/survival advantage. Because gene amplification often leads to overexpression [5–7], the search for genes overexpressed in cancer cells is a logical and common strategy for target identification.

Genomic technologies are now frequently utilized in lead selection and compound optimization. They are routinely used to profile candidate compounds in order to identify the pathways activated and thus delineate the on-target and off-target effects. Large databases are being created to correlate the chemical structures of compounds and their toxicogenomic profiles. An important component of preclinical development is biomarker discovery. The increasing importance of biomarkers is closely connected with the exponential growth in R&D costs currently experienced by the pharmaceutical industry [8]. An early discovery of an efficacy biomarker could substantially decrease development costs and cut the development time, thus allowing an earlier market entry and hence an improved patent life cycle [9]. Therefore, biomarker discovery programs are now incorporated early into the discovery process, typically at the lead optimization or preclinical testing stages. So-called patient stratification biomarkers, that is, markers correlated with the disease type or response to a drug candidate, are particularly valuable in oncology development, as cancer represents a heterogeneous genetic disease. In one of the recent successful examples of patient stratification strategies, the response to a drug called Herceptin was found to correlate with the amplification of the *HER2* gene [10–12]. These biomarkers offer the possibility of rationally selecting patients for clinical trials and are now being pursued early in the discovery process. For example, cell culture screens for target inhibition are often accompanied by detailed genomic analysis of the cell lines to identify genetic markers correlating with the response in the assay.

Several genomic technologies are used in biomarker discovery. Gene expression and CGH microarrays are used to determine the molecular profiles that predict the outcome in various cancers and correlate with the response to the drug (for examples see Refs. 13–21). Microarrays can also be utilized to determine the molecular signatures of the response to the agent, leading to discovery of efficacy biomarkers [22].

In the following sections, we describe the current state of the aforementioned microarray technologies, as well as several methods for manipulating gene activity, and review the applications of these tools in preclinical drug development. The objective of this chapter is to introduce to the reader the most common genomic technologies, to describe the relevant protocols, and to briefly summarize their applications in preclinical drug development. Chapter Toxicogenomics in Preclinical Development comprehensively reviews the applications of DNA microarrays in the field of toxicology.

## 2 GENE EXPRESSION MICROARRAYS

### 2.1 Technology

We define some of the microarray terms used in this chapter.

*Probe.* A DNA fragment attached to the microchip and used to detect transcripts in the test sample.

*Reverse Transcriptase.* An enzyme capable of synthesizing DNA using RNA as a template.

*DNA Polymerase.* An enzyme capable of synthesizing DNA on a DNA template.

*cDNA.* DNA synthesized off an mRNA template.

*cDNA.* RNA synthesized in an *in vitro* transcription reaction using cDNA as a template.

***Standard Microarray Protocol*** Generally, gene expression microarrays represent microchips containing thousands of DNA probes, which are used to analyze the abundance of multiple transcripts in a sample. Based on the type of probe, microarrays can be classified into oligonucleotide and cDNA arrays. While cDNA microarrays may potentially offer higher sensitivity of mRNA detection, difficulties in manufacturing and deposition of cloned and purified long DNA sequences have limited the use of cDNA arrays largely to academic laboratories. In this chapter, we focus on the more commonly used oligonucleotide microarrays.

Figure 1 presents an outline of a typical microarray experiment. Total RNA is isolated from the test sample using one of the known techniques [23, 24] and is used as a template to synthesize cDNA. An oligo(dT) primer with an attached T7 sequence is used as a primer for reverse transcription. An enzyme called reverse transcriptase catalyzes the synthesis of cDNA using the input RNA as a template. The resulting cDNA is then subjected to one round of DNA replication to generate double-stranded DNA. This reaction is catalyzed by a DNA polymerase. The resulting double-stranded DNA then serves as a template for T7 RNA polymerase, which recognizes the T7 sequences in the cDNA [25]. The *in vitro* transcription is performed in the presence of biotinylated rNTPs to label the cRNA. The cRNA is fragmented and hybridized to the array.

After array hybridization, the array is washed to remove unbound molecules, stained using streptavidin-phycoerythrin and a bioinylated anti-streptavidin antibody, and scanned in a fluorescent scanner in order to quantify the signal for all the probes. After the acquisition of the image by the scanner, a specialized program overlays a grid on the array to identify the spots and generates a table of signal intensities. A different program then processes the signal intensities for individual probes to generate the intensities for each individual gene, determine the background, and perform normalization. The signal intensity for each gene serves as a measure of the abundance of the corresponding transcript in the initial sample.

The quantity of the test RNA sample is an important factor in microarray analysis. The RNA polymerase synthesizes multiple copies of cRNA from each cDNA molecule, and the target preparation protocol results in an amplification of the original sample. For RNA quantities > 1 μg, the above-described protocol typically

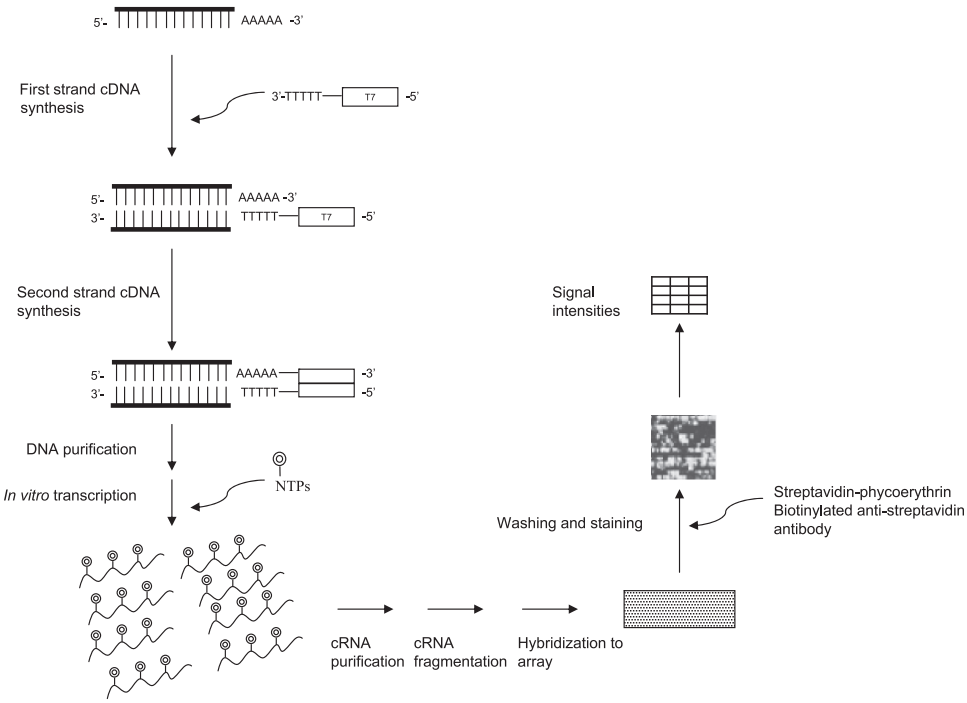


**FIGURE 1** Gene expression profiling using DNA microarrays (single-round RNA amplification protocol). The total RNA sample is amplified to generate biotin-labeled cRNA, which is fragmented and hybridized to a microarray.

produces enough cRNA for at least one or two array hybridizations. Another important consideration is the integrity of the sample. If the total RNA is degraded, the reverse transcriptase will not be able to synthesize sufficiently long cDNAs and the cRNA products will not hybridize to all the probes for the transcript. Although 3′-bias is an important consideration in microarray probe design, many probes on a microarray are removed by hundreds of nucleotides from the 3′-terminus of the transcript. Probes that are distant from the 3′-terminus will not hybridize if the cRNA products are of insufficient length. Similarly, suboptimal functioning of the reverse transcriptase may lead to shortened cDNAs and cRNAs and hence will result in an underestimation of the abundance of the transcripts. As microarray applications expand, more different sample types will need to be analyzed. Next, we consider two special situations with regard to the sample type.

***Microarray Analysis of Archived and Small Samples*** The standard expression microarray protocol requires high quality intact RNA. However, as microarrays became an important tool in biomarker research, the researchers started looking for ways to apply the microarray technology to retrospectively analyze archived human tissue samples. Freezing of a sample immediately after surgical resection typically preserves RNA. Therefore, such samples can be analyzed with a standard microarray protocol. It is critical that the sample be frozen immediately, because even quick manipulation of tissue results in changes in gene expression [26, 27] that

may be mistaken for true characteristics of the sample. However, the most difficult challenge is presented by formalin-fixed paraffin-embedded (FFPE) samples, as formalin fixation results in irreversible modification and degradation of RNA [28, 29]. It is noteworthy that formalin fixation results in a wide spectrum of RNA modifications, including cross-linking, addition of monomethylol residues to the nucleic bases, and adenine dimerization [28]. Until recently, such samples were deemed unsuitable for microarray analysis. In 2003, a specialized microarray for analysis of FFPE samples was developed [30]. It contains mostly probe sets that are directed against the three hundred 3′-terminal nucleotides of the transcripts (instead of the 600 nucleotide limit set for regular microarrays). The increased 3′-bias is intended to facilitate binding of shortened cRNAs synthesized off truncated cDNAs. However, a solution remains to be found for analysis of highly modified RNA, as RNA with modified nucleic bases has a limited capacity to produce cDNA in the reverse transcription reaction. The chip is designed for use with a reagent system [31], which enables RNA isolation from FFPE tissues as well as its amplification and labeling.

Significant progress has also been made in analysis of small samples. The progress in this area was fueled by the introduction of a tissue dissection technology called laser capture microdissection [26, 32]. The technique involves placing a transparent film over a tissue section and selectively adhering the cells of interest to the film with a fixed-position, short-duration, focused pulse from an infrared laser (Fig. 2). During the procedure, the tissue is visualized microscopically. The film with the

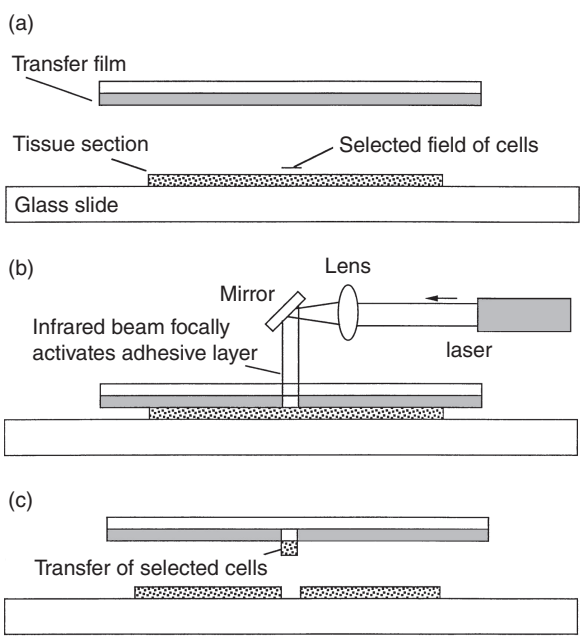


**FIGURE 2** [32] Laser capture microdissection protocol. (a) A tissue section is mounted on a microscope slide and covered with transparent film. Cells of interest are selected visually under a microscope. (b) A laser beam focused on the cells of interest is activated, causing the film to adhere to the selected cells. (c) The film is removed together with the attached cells. At this point, the cells of interest can be lysed and further processed.

procured tissue is then removed from the section and used to isolate DNA or RNA [32]. As laser capture microdissection of tissue samples became common, an urgent need arose for a microarray protocol suitable for analysis of samples ranging from 100 to 1000 cells. The problem was solved by introducing an additional round of RNA amplification. In a two-round amplification protocol (Fig. 3), the first round is performed with regular rNTPs, while the second round uses labeled rNTPs as in the single-round amplification protocol described in the previous subsection. Today, protocols involving laser capture microdissection and RNA isolation from single cells followed by gene expression analysis have become routine. They made possible the analysis of pure tumor cells and comparison with adjacent normal tissue [33–35].

***Microarray Data Analysis*** Most microarray experiments involve either a comparison between a treatment and the baseline or a comparison between the test sample and a reference. Therefore, the first level of data analysis almost inevitably involves building gene expression ratios, that is, calculating the ratios between the intensity values for the same gene from two different chips. A *t*-test is typically used to determine the significance of the difference between the control and the test values for each gene. The data can then be filtered to remove insignificant changes. Methods based on conventional *t*-tests provide the probability ($p$) that a difference in gene expression occurred by chance. It is common to set up a significance threshold at $p$-value $\leq 0.01$. Although $p = 0.01$ is a reasonably stringent cut-off for experiments designed to evaluate small numbers of genes, a microarray experiment measuring the expression of 20,000 genes (such as an experiment using Affymetrix U133A arrays) would identify 200 genes by chance.

To reduce the number of false positives, significance analysis of microarrays [36] can be used. This method identifies genes with statistically significant changes in expression by assimilating a set of gene-specific *t*-tests. Each gene is assigned a score on the basis of its change in gene expression relative to the standard deviation of repeated measurements for that gene. Genes with scores greater than a threshold are deemed potentially significant. The percentage of such genes identified by chance is the false discovery rate (FDR). To estimate the FDR, nonsense genes are identified by analyzing permutations of the measurements. The threshold can be adjusted to identify smaller or larger sets of genes, and FDRs are calculated for each set. Other false discovery analysis methods have recently been introduced, some of which include analysis of false negatives [37, 38].

To improve the robustness of analysis, multiple replicates of the same sample are typically run. The commonly accepted minimum is two replicates; however, use of triplicates minimizes the false-positive rate [39]. A microarray experiment consists of multiple steps, and each step represents a potential source of variation. The variation of the measured gene expression data can be categorized into two generic sources: biological and technical variations. The biological variation in gene expression measured comes from different animals or different cell lines or tissues. It reflects the variability in gene expression between the different biological samples used in the experiment. Biological variation can be assessed only by using independent biological replicates. If all biological samples are pooled, the biological variation is minimized, but the potentially useful information on the variability in gene expression between different animals or cells is lost. The technical variation accounts

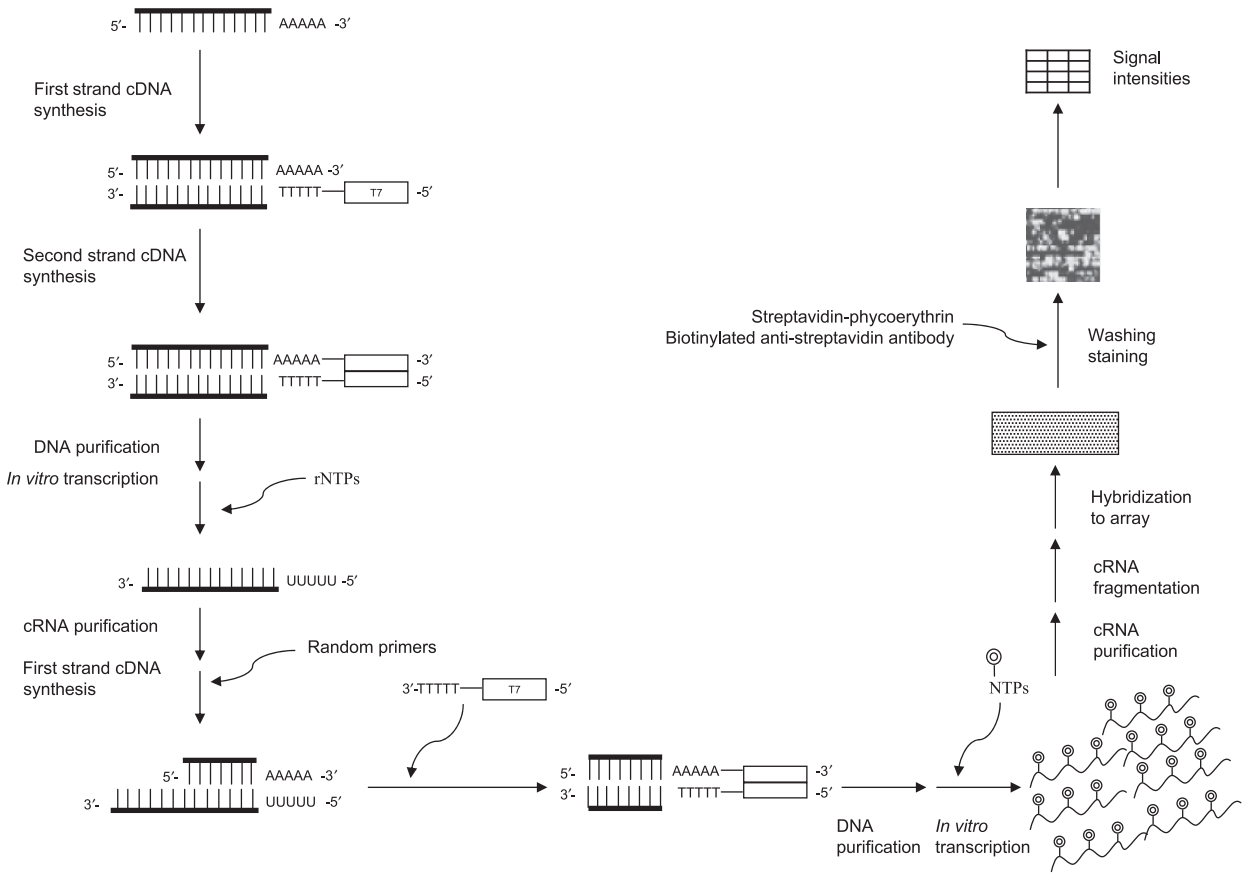


**FIGURE 3** Gene expression profiling of small samples using DNA microarrays. Unlike the standard protocol outlined in Fig. 1, this procedure involves two rounds of RNA amplification. The first round is performed in the presence of regular rNTPs, while the second round involves biotin-modified rNTPs to label the cRNA product.

for the variation associated with the use of microarray techniques unrelated to the biological samples. The biological, technical, and residual variations are mutually independent. The variation in a measured intensity is the sum of these two variations.

The contributions of the technical and biological variabilities to the overall variability have been studied extensively. It has been established that the biological variation is the main component of the variation between microarray experiments [40, 41]. Therefore, biological replicates (multiple plates of cells, multiple animals, etc.) rather than technical replicates (multiple arrays run for the same sample) should be used whenever possible.

Once the data are filtered with respect to the statistical significance, an additional filter is usually set up to remove genes with a small fold change, which are less likely to be biologically relevant. As the robustness of microarrays improved with time, the fold change threshold was lowered; it is now commonly set at 1.5 or 2.

Simple lists of genes regulated as a result of a biological process provide limited information. As applications of expression microarrays widened and the numbers of genes analyzed increased, the analysis methods have become more and more complex.

One of the most common tasks in microarray data analysis is identification of common patterns of gene regulation in a population of samples. An example of such a task would be identification of genes coinduced in a series of treatments or discovery of genes associated with a particular biological characteristic of the samples (disease category, tissue type, etc.). Problems of this type are commonly solved by two-dimensional clustering, a statistical procedure whereby samples (each represented by one or more microarrays) are aggregated into clusters based on the similarity of their expression "signatures," while the genes are simultaneously clustered based on the similarity of their expression levels across the samples. The rationale behind clustering samples according to their expression profiles is simple: samples with similar gene expression "signatures" are more likely to have common biological characteristics. Similarly, genes coregulated in a series of samples are more likely to be part of a common biological pathway activated in the samples under consideration. Thus, two-dimensional clustering may provide very useful information on the degree of relatedness between samples and reveal the genes potentially relevant to the classification. Clustering results can be conveniently visualized using a gene expression matrix, or a heatmap, in which each column represents an experiment and each row represents a gene (Fig. 4). Each element of the heatmap is colored based on the expression level, thus providing a convenient visual representation of the gene expression patterns across all the experiments. One of the most notable applications of clustering is in cancer classification, which was pioneered in the late 1990s (for examples see Refs. 18 and 42–44).

If clustering is done without any a priori introduced sample classification, it is referred to as unsupervised clustering. Because of its unbiased nature, unsupervised clustering is often used to identify patterns in previously unclassified complex datasets. Several unsupervised clustering algorithms are used for microarray data analysis (reviewed in Refs. 45 and 46). Hierarchical clustering is the most common algorithm. It uses an agglomerative approach, whereby expression profiles are successively joined to form groups based on similarity between them, thus forming a hierarchical tree, or dendrogram [47]. The latter presents a convenient visualization

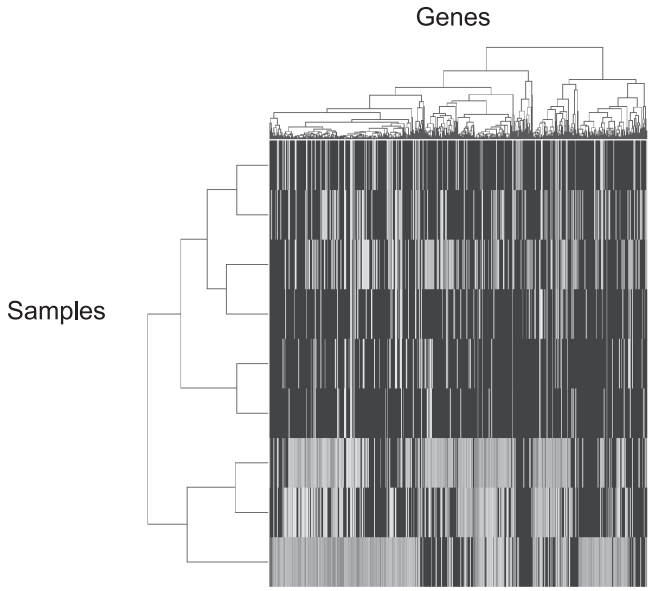


**FIGURE 4** An example of a heatmap obtained by hierarchical two-dimensional clustering of nine samples. Each row represents a sample and each column represents a gene. (Although not visible here, a red color is used for upregulation and a blue color is used for downregulation of genes, with black reserved for unaffected genes.) The dendrogram on the left illustrates the degree of relatedness between the expression profiles of the samples and the dendrogram on the top reflects the similarity of the expression levels for each gene across the samples.

option and is often presented together with a heatmap (Fig. 4). An alternative algorithm is *k*-means clustering, a divisive approach based on partitioning the dataset into a predefined number ($k$) of clusters [48]. Obviously, it requires some a priori knowledge of the biology of the dataset so that the number of clusters could be preset. When the researcher can specify in advance not only the number of clusters but also the relationships between them, self-organizing maps (SOMs) can be used, which organize the clusters into a "map" where similar clusters are close to each other [49].

Unsupervised algorithms can find novel patterns in datasets but are not designed to classify data according to known classes. On the contrary, supervised clustering approaches, such as support vector machines (SVMs) [50], take known classes and create rules for assigning genes or experiments into these classes. The user initially runs microarrays for a training set with known class labels and enters the gene expression profiles together with the classification information into the algorithm. This "trains" the algorithm or teaches it to associate certain gene expression patterns with the predefined sample class labels. The next step is to profile samples from a new set of samples, the test set, and input the gene expression data into the algorithm. The latter will then classify the samples using the knowledge on class–expression pattern associations learned from the training set. SVMs have been used to identify genes with similar expression patterns, but their most powerful application is in classification of samples. They have been used extensively in cancer classifica-

tion and in some cases proved to be more reliable than the traditional diagnostic methods [42, 51]. Examples of supervised clustering relevant to drug discovery are discussed in Section 2.2.

The gene expression pattern of a cell reflects its phenotype and may provide information on the intracellular signaling pathways functioning in the cell. A comparison of the gene expression profiles for diseased tissue and the adjacent normal tissue may provide insight into the mechanisms and pathways driving the disease. In *in vitro* experiments, studying the gene induction patterns caused by a treatment may help identify the pathways activated or repressed by the treatment. A prerequisite for all these applications is the ability of the researcher to map gene expression profiles to signaling pathways, that is, identify the associations between the affected genes and the known pathways for the entire signature. There are several programs that allow association of gene expression patterns with predefined biological classes [52–60]. They use one of the existing gene classification systems, such as Gene Ontology [61, 62], Biocarta [63], or KEGG [64], to determine the enrichment of an expression signature in a certain motif, such as "cell cycle control" or "DNA biosynthesis." Gene Ontology is the most commonly used annotation system, which classifies a significant fraction of the genome (~15,000 genes) according to their involvement in a biological process or molecular function or their cellular localization. It is built hierarchically and involves a parent–child relationship between its terms. Programs such as MappFinder allow the researcher to identify the GO terms that show correlated gene expression changes in a microarray experiment. The affected GO terms can then be rank-ordered based on the $Z$-score, a statistic that reflects the number of genes in the term meeting the criteria for fold change in the microarray experiment [52]. MappFinder was one of the first GO-based programs designed for analysis of gene expression data. It has since been used to study the effects of various factors on intracellular pathways *in vitro* [65, 66] and *in viv*o [67]. Combined use siRNA-mediated gene silencing and MappFinder analysis of expression signatures has been suggested as an approach to pathway profiling [65]. Other pathway analysis programs have been developed in the past several years, which allow convenient visualization of the pathway analysis results [55–60]. Many of them use manually curated pathways instead of Gene Ontology, which permits greater focus and lower redundancy, especially when studying specific disease-related pathways.

Given the large amounts of gene expression data accumulated in the literature, integrative analysis of multiple datasets related to the same disease represents a very attractive idea. The precedent for such analysis was established when Rhodes et al. [68] performed so-called meta-analysis of four different gene expression datasets for prostate cancer. The authors identified a molecular signature common to the datasets, thus generating a robust signature of the disease. The signature was then mapped to KEGG pathways [64] to reveal a common biological motif, activation of polyamine biosynthesis [68]. Other studies identified common gene signatures in different breast and lung cancer datasets [69–71]. The existence of common motifs in datasets from different laboratories despite the well-publicized problem of interplatform variability presents strong evidence in favor of microarrays as tools for identification of drug targets and biomarkers.

As the amount of information derived from microarrays continues to increase, new and more complex data analysis procedures will emerge that will facilitate the current and future applications of the technology.

## 2.2 Applications of Microarrays in Preclinical Development

The applications of gene expression microarrays in drug discovery are summarized in Fig. 5. Chronologically, the first application of gene expression microarrays in drug discovery was in the area of identification of therapeutic targets.

***Target Identification and Validation*** One of the most obvious and logical strategies for target identification is a genome-wide scan for genes or their protein products overexpressed in the diseased tissue relative to its normal counterpart. Gene expression microarrays are ideally suited for this purpose because they provide wide genome coverage and permit efficient screening of hundreds of samples using a standardized reproducible protocol. To date, they have been most widely used to identify drug targets in cancer.

The development of sophisticated data analysis methods, such as two-dimensional clustering, has permitted cancer classification on the genomic basis and identification of groups of genes overexpressed in subsets of target cancers. Hematological cancers were historically the first group of cancers to be subjected to microarray classification, because of the better availability of homogeneous tumor cell populations. In a pioneering study of 72 acute leukemia samples, Golub et al. [42] identified a gene expression signature that reliably determined the disease type. The signature correctly assigned the samples into the acute myeloid leukemia (AML) or acute lymphoid leukemia (ALL) categories and further determined the subtype in the ALL category (B cell and T cell). In another study, Yeoh et al. [72] profiled 360 ALL blasts and identified gene expression patterns associated with the 6 known clinical subgroups of disease. While these and other works produced robust gene expression-based disease classifiers, no members of these classifiers appeared suitable as therapeutic targets. A recent exemplary microarray study of the mixed-lineage leukemia (MLL) [73, 74] went one step further. Armstrong et al. [73] not only identified a distinct gene expression profile that distinguishes leukemias with MLL translocations, but also identified a single gene that is consistently overexpressed in this leukemia type and validated it as a therapeutic target [74]. Mixed-lineage leukemia has traditionally been defined as a subtype harboring specific chromosomal translocations, so-called MLL translocations. The first study established a globally distinct nature of this disease by identifying its characteristic gene expression profile [73]. The authors then analyzed the MLL signature and identified

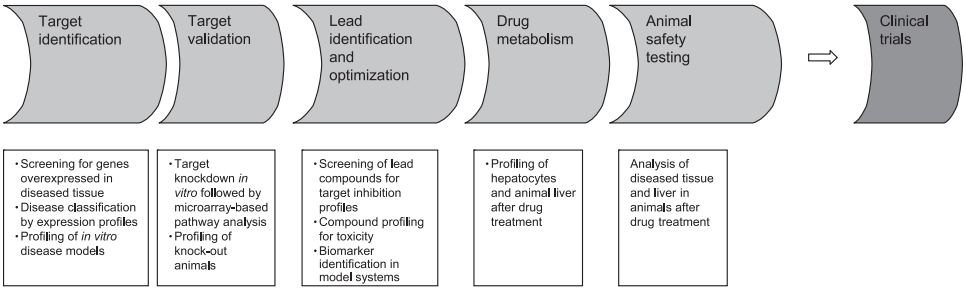


**FIGURE 5** Applications of expression microarrays at different stages of drug discovery and development.

one gene consistently overexpressed in most MLLs compared to other acute leukemias [74]. This gene codes for a receptor tyrosine kinase, FLT3, which plays an important role in hematopoietic development. The authors then determined that many MLLs also carry mutations in the FLT3 gene, which result in constitutive activation of the kinase. An inhibitor of FLT3 was shown to differentially kill cells carrying MLL translocations *in vitro* and inhibit MLL development in a mouse model, thus validating FLT3 as a therapeutic target for MLL. This study truly demonstrates the power of gene expression microarrays in drug target discovery.

Microarrays have also been used for target identification is solid tumors. Expression profiling of medulloblastoma, a highly invasive tumor of the cerebellum, revealed expression signatures of the metastatic and nonmetastatic subtypes of the disease [75]. Several genes, including PDGFRA and members of the downstream RAS/mitogen-activated protein kinase pathway, were found to be upregulated in the metastatic tumors. The study thus yielded several potential therapeutic targets. Dhanasekaran et al. [76] have profiled over 50 normal and neoplastic prostate samples and three common prostate cell lines. Characteristic expression signatures were identified for localized prostate cancer, metastatic hormone-refractory prostate cancer, and benign prostatic hyperplasia. Two genes were consistently overexpressed in prostate cancer samples, hepsin (a transmembrane serine protease) and PIM1 (a serine/threonine kinase). The expression of the protein products of these genes was examined on a tissue microarray containing 738 tissues and found to be elevated in prostate cancer. An independent microarray screen of 11 malignant and 4 normal prostate samples also revealed hepsin overexpression in prostate cancer and implicated hepsin as a promising drug target [77]. Other studies later confirmed the conclusions on hepsin as a promising therapeutic target [78].

It is important to note here that target identification by microarrays should not be limited to direct identification of genes overexpressed in diseased tissue. Most genes in the disease signature encode proteins that are not druggable. Moreover, overexpression of certain genes in a disease signature only represents the bottom part of the signaling pathway modulating the disease process. Recent bioinformatics developments described in the preceding subsection permitted mapping of gene expression signatures to signaling pathways. Thus, an alternative approach to target identification may involve identification of pathways activated in diseased tissue and a search for druggable targets within the pathways. Subsequent target validation steps may involve modulation of the target activity *in vitro* followed by microarray profiling to determine whether the target-related signature is affected.

All the aforementioned studies have been performed in patient samples. In many cases patient samples are not available and therefore target discovery has to be performed in preclinical model systems, such as cultured cells or animal models of disease. However, these systems may produce a lot of noise, because cultured cells accumulate a lot of secondary genetic changes (such as gene amplifications/deletions), which are not relevant to the initial disease-originating events. Therefore, microarray-based target discovery in cultured cells requires inventive experimental design and careful validation. In an example from the oncology area, Sarang et al. [79] performed a rescue screen in neuroblastoma cells for approximately 900 known therapeutic agents. They found that 26 of these agents are capable of rescuing the cells from oxidant stress. The compounds were profiled by microarrays to identify the common gene expression signature. One of the genes in the signature codes for

a secreted peptide called galanin. In a series of validation experiments, galanin was found to reverse cell death cased by oxidant stress. It was thus concluded that galanin receptor may represent a therapeutic target. Indirect identification of therapeutic targets has been performed by analyzing pathways activated by expression of known oncogenes [80, 81].

Although microarrays can reliably associate gene overexpression with disease, they cannot tell whether the overexpression is the cause or an effect of the disease. In most cases, microarray screens generate long lists of overexpressed genes and thus necessitate complex hypothesis-driven follow-up experiments aimed at selecting the genes that have a causative role in the disease process. Target validation is often performed in model systems using a loss-of-function or gain-of-function approach. Microarrays can also be used at this stage to assess the global effects of target knock-down or target overexpression. An example of such application is a study by Cho et al. [82], who knocked down the protein kinase RIα gene with an antisense oligonucleotide in cultured cells and subjected the cells to microarray analysis. It was found that suppression of the protein kinase RIα gene causes coordinated changes in gene expression that can be mapped to cell growth, differentiation, and activation pathways. The genes that compose the proliferation–transformation signature were downregulated, whereas those that defined the differentiation signature were upregulated in antisense-treated cancer cells and tumors, but not in host livers, thus validating the mechanism of protein kinase RIα. Short interfering RNA (siRNA) has recently emerged as a promising tool in target validation [83, 84]. Gene silencing with siRNA followed by microarray experiments and systematic pathway analysis may prove to be a powerful strategy in functional genomics. Early studies established the feasibility of this approach [65, 85]. One should expect that this approach will be actively used in target validation in the near future.

***Compound Characterization*** More recently, DNA microarrays have become an important tool in compound selection and optimization. Since microarrays are capable of generating a genome-wide view of the physiological state of the cell, they are well suited for characterization of compounds and elucidating their mechanism. A global approach to the problem would be to create a database of expression signatures associated with compound treatments in a relevant model system (Fig. 6). If such database is sufficiently populated with compounds with known mechanisms, one could envision its use as a look-up table: a series of novel compounds can be profiled by microarrays and their expression signatures can be plugged into the table. One of the available clustering algorithms would then be used to determine which of the known compounds have the most similar signatures to the test compounds.

Although more effort is needed to compile a comprehensive activity/expression signature database, numerous studies have used microarray profiling to determine the mechanism of known and novel agents. In an early example of microarray-based compound profiling, Glaser et al. [86] profiled three known histone deacetylase (HDAC) inhibitors in two cell lines to generate an HDAC inhibition signature. The gene expression signatures of the three active HDAC inhibitors were generally similar to each other and differed significantly from the signatures for inactive analogues, suggesting that the expression signatures are mechanism based. A core signature of 13 genes was identified that was common to all the HDAC inhibitors and

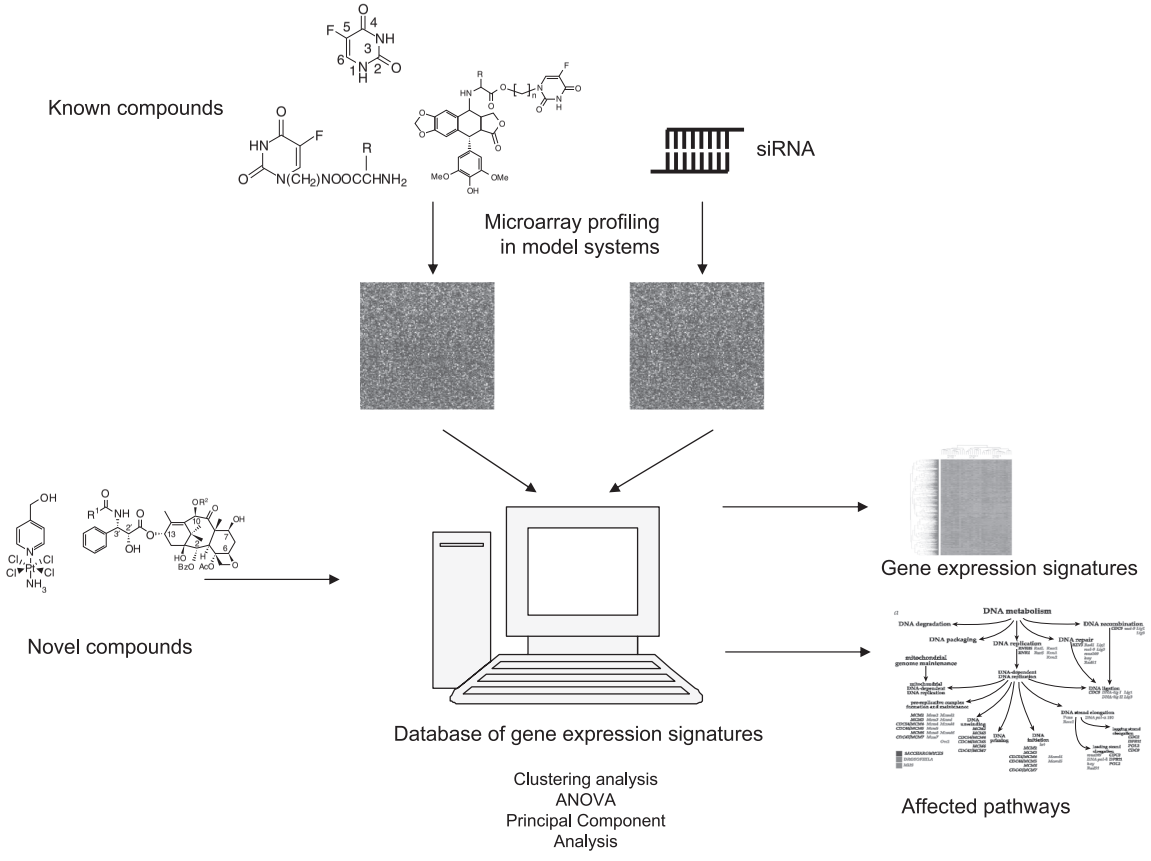


**FIGURE 6** Creation of a genomics database for compound selection and optimization. Gene expression signatures of target inhibition and elimination are obtained using known inhibitors and siRNA, respectively, for multiple targets. Novel compounds synthesized to inhibit the target are profiled using the same microarray and their gene expression signatures are utilized to identify the affected pathways and targets.

all the cell lines used. The inhibitors were structurally different, implying that the signatures reflect the HDAC inhibition mechanism rather than off-target effects of the compounds.

An inhibitor of cyclooxygenase 1 (COX1) and cyclooxygenase 2 (COX2), sulindac sulfide, has been profiled in colorectal carcinoma cells, which express the COX1 enzyme but little COX2 [87]. A group of 11 genes was identified; their expression was further analyzed in another colon cancer cell line, HCT116, which expresses very low levels of both enzymes. The drug did not affect the expression of these 11 genes in HCT116, suggesting that their induction is COX dependent. This study is important because COX inhibitors have been shown to induce growth arrest and apoptosis of colon cancer cells. Elucidation of the mechanism of this potential anticancer effect could result in either selection of one of the known COX inhibitors for preclinical studies or optimization of the structure and creation of new COX inhibitors with potential anticancer activity. More compound profiling studies are analyzed in a very comprehensive review by Clarke et al. [88].

In an example from a different therapeutic area, Gunther et al. [22] have profiled multiple classes of antidepressants, antipsychotics, and opioid receptor agonists in primary human neurons. The gene expression patterns obtained for these groups of drugs were then used to construct statistical models capable of predicting drug efficacy. Several supervised classification algorithms were shown to reliably predict the functional group of each of the drugs based on their expression signatures.

A similar approach can be taken with respect to toxicity if the appropriate model system is used (hepatocytes, liver of model animals, etc.). This application of DNA microarrays will be comprehensively covered in Chapter Toxicogenomics in Preclinical Development. The toxic mechanism of a new compound can be determined from the database by association with the toxicity signatures of known compounds. In an early proof-of-concept study, Waring et al. [89] treated rats with 15 different known hepatotoxins, which are known to cause a variety of hepatocellular injuries including necrosis, DNA damage, cirrhosis, hypertrophy, and hepatic carcinoma. Gene expression signatures of the livers of treated rats were clustered and compared to the histopathology findings and clinical chemistry values. The results show strong correlation between the histopathology, clinical chemistry, and gene expression profiles induced by the agents. Other studies also demonstrated the feasibility of predicting the toxicity of compounds using microarray gene expression signatures [90–94]. A recent publication describes the creation of a comprehensive chemogenomics database [95]. The authors have profiled approximately 600 known compounds, including over 400 FDA-approved drugs in up to seven different tissues of rats. It was demonstrated that gene expression signatures, when used in conjunction with the database, can be used to predict the toxicity of a compound.

***Biomarker Discovery*** Biomarker identification is a major application for DNA microarrays in preclinical development, which often parallels and aids compound optimization. A biomarker is "a characteristic that is objectively measured and evaluated as an indicator of normal biologic processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention" [96]. An early identification of a useful biomarker could result in substantial savings in the development process. Therefore, a substantial amount of effort has been devoted to biomarker discovery at the preclinical development stage, using model systems, such as cultured cells and

animal models. Studies similar to the ones described in the previous subsection may yield useful pharmacodynamic biomarkers related to the mechanism. However, if the goal is to identify patient stratification biomarkers—that is, biomarkers that can predict a patient's response to the drug—one needs to screen basal levels of expression and then relate them to the sensitivity or resistance of the system to the drug. Early attempts to correlate gene expression signatures with drug sensitivity have been done in cultured cell lines [97–99]. Gene expression microarrays have been used to generate basal expression profiles of the 60 cell lines used by the NCI for cancer drug discovery screens (NCI-60 panel) [97]. The gene expression profiles were then correlated with the sensitivity of the cells to several common drugs, such as 5-fluorouracil and L-asparaginase [98]. The authors then correlated the expression of certain genes to the mechanisms of drug resistance. Gene expression signatures were used to predict the chemosensitivity of the cell lines. Staunton et al. [99] developed an algorithm to predict the sensitivity of the NCI-60 cell lines to 232 known agents. All 232 compounds were profiled in the 60 cell lines and the data were divided into a training set and a test set. The training set was used to develop classifiers and the test set was used to test the accuracy of the classifiers. The study yielded accurate classifiers ($p \leq 0.05$) and thus proved the feasibility of chemosensitivity prediction by microarrays.

Correlations between drug sensitivity and expression signatures have also been determined in more complex model systems, such as animal cancer models [100, 101]. Zembutsu et al. [100] profiled 85 cancer xenografts derived from nine human organs. The xenografts, implanted into nude mice, were examined for sensitivity to nine anticancer drugs (5-fluorouracil, 3-[(4-amino-2-methyl-5-pyrimidinyl)methyl]-1-(2-chloroethyl)-1-nitrosourea hydrochloride, adriamycin, cyclophosphamide, cisplatin, mitomycin C, methotrexate, vincristine, and vinblastine). The authors identified gene expression signatures, which correlated with drug sensitivity of the xenografts, and established an algorithm to calculate the drug sensitivity score based on the gene expression pattern.

Studies of this type have established the feasibility of predicting response to the drug based on the expression profile. However, it remains to be proved that correlations established in model systems will reproduce in patient samples. To date, only a limited number of studies correlated drug response with the basal expression profile in human samples [20, 21, 102–104]. Cell culture and xenograft-based models have a number of limitations. Most importantly, cultured cells undergo multiple rounds of selection, acquire additional genetic abnormalities, and therefore may poorly represent the original tumor, both in terms of their genomic patterns and their drug sensitivity profile. Future research will show whether gene expression profiles obtained in preclinical model systems can be used to predict patient response in clinical trials.

## 3 COMPARATIVE GENOMIC HYBRIDIZATION

### 3.1 Technology

Chromosomal aberrations are detrimental events associated with a number of developmental diseases, such as Down syndrome. Amplifications and deletions of chromosomal regions occurring in somatic cells are believed to be one of the main

factors leading to cancer. Although fluorescent *in situ* hybridization (FISH; for a recent review, see Ref. 105) has been effectively applied to analyze known genetic aberrations for decades, until recently there was no method for detecting gene copy number alterations on a whole-genome scale. Comparative genomic hybridization (CGH), a technique that enables genome-wide analysis of chromosomal aberrations, was first described by Kallioniemi and colleagues in 1992 [106]. The method involves hybridization of the test DNA (sometimes mixed with reference DNA) to a complete representation of the genome attached to a solid support. Originally, CGH was performed on metaphase chromosome spreads, but in the past five to seven years, microarray-based CGH has become dominant (reviewed in Refs. 107 and 108). An overview of the array CGH procedure is shown in Fig. 7.

The original array CGH protocols have employed a two-color hybridization scheme, whereby the test DNA is labeled with a red fluorescent dye, while the reference normal DNA is labeled green. Although genomic DNA can be labeled and hybridized directly, many CGH protocols involve a PCR-based amplification step. Once the DNA is labeled, the test and the reference samples are mixed and hybridized to the array. Cot-1 DNA is typically added to suppress the hybridization of

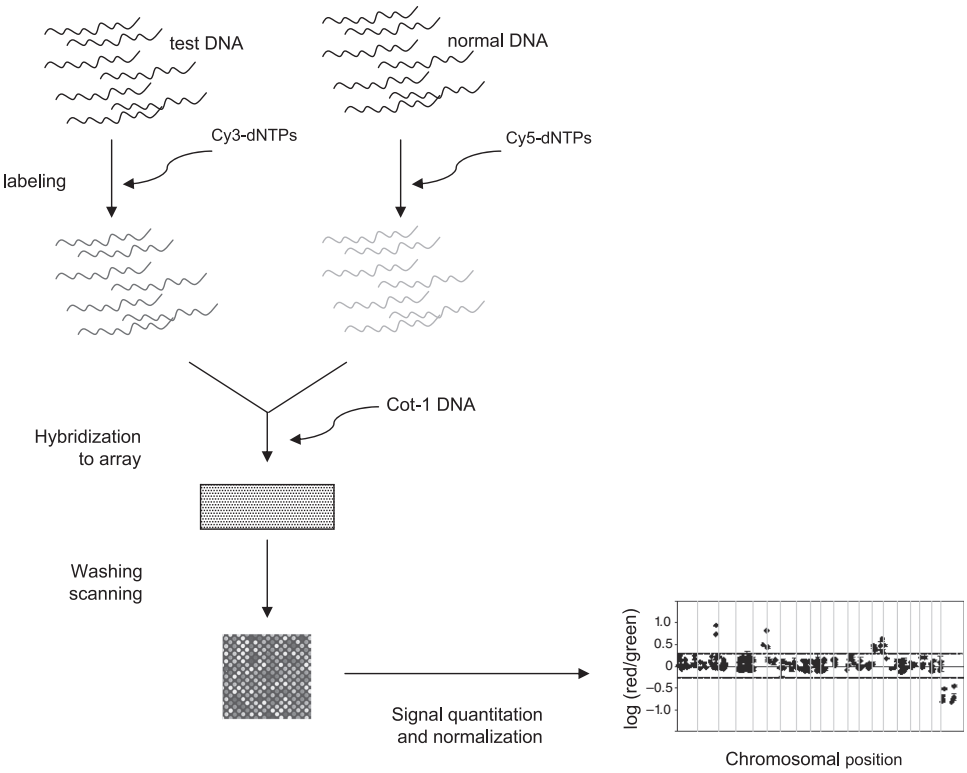


**FIGURE 7** Two-color procedure for comparative genomic hybridization. The test gDNA is labeled red and the reference normal gDNA is labeled green. The gDNA samples are mixed and hybridized to a CGH array. Cot-1 DNA is added to eliminate the signal from repetitive sequences. After hybridization, the array is washed and scanned to generate signal intensities for all regions of interest.

repetitive sequences. After hybridization, the array is washed and scanned to generate the fluorescent intensity values for each probe on the array. The data are then normalized and presented as ratios of test/normal (usually on a log scale). An example of the CGH output for one chromosome in one sample is shown in Fig. 7. Ratios between the test and reference samples for multiple positions on a chromosome provide information on the copy number for each region measured. The copy number profile of a sample typically consists of a series of plateaus corresponding to regions with a constant copy number, flanked by abrupt transitions. An important limitation on the use of CGH is that it can measure changes in copy number, but it cannot detect balanced chromosomal translocations or changes in ploidy.

The value of CGH arrays increases with an improvement in genome coverage, resolution, and reproducibility. Several types of array platforms are currently used for CGH. Historically, the genome was represented on CGH arrays as a collection of bacterial artificial chromosomes (BACs) (e.g., see Refs. 109 and 110). Thousands of BACs were propagated and used as templates to generate PCR products. The PCR products were purified and deposited on a microarray. BAC arrays provide sufficiently high sensitivity to detect single copy amplifications and deletions [109]. The main drawback of the earlier BAC arrays was low resolution. Spotting 2460 BAC clones in triplicate provided an average resolution of 1.4 MB across the genome [109]. However, a high density BAC array was recently developed that contains approximately 30,000 clones arranged in a tiling fashion and covering the entire genome [109]. The array provided a significantly higher resolution and made possible detecting amplifications as small as 300 kb [109].

Arrays containing cDNAs have been used extensively for CGH [4, 6, 111, 112]. The advantages of cDNA arrays include higher reproducibility, easier manufacturing, and better representation of the genome [108]. However, multiple probes are required to detect small copy number changes, and more sample needs to be used (several micrograms), due to the lower sensitivity of the array [108].

A significant breakthrough was achieved in 2004 when two oligonucleotide-based platforms were developed for CGH [113, 114]. One of them was a microarray containing over 21,000 60-mer probes synthesized *in situ* by an ink jet technology [113]. The array provided a significant improvement in resolution and was shown to reliably detect single-copy losses, homozygous deletions, and various types of amplifications. It used the two-color protocol outlined in Fig. 7, with the addition of a PCR step to amplify the test and control DNAs. The other platform represented a high density microarray originally designed for detection of single nucleotide polymorphisms (SNPs) [114]. The array covered over 10,000 SNPs distributed across the genome. Each SNP was interrogated by multiple 25 mers synthesized *in situ* by a photolithographic method. Unlike the two-color CGH protocol described in Fig. 7, the SNP array protocol involves labeling DNA by incorporation of biotinylated dNTPs (Fig. 8). After array hybridization, the array is stained using streptavidin-phycoerythrin and a biotinylated anti-streptavidin antibody. Signal intensities from individual SNP measurements are smoothed across a user-defined smoothing window using a specialized algorithm. The resulting values are then compared to a preloaded reference dataset for normal DNA to produce an estimate of the copy number in the experimental sample. The array was used to evaluate chromosomal aberrations on a genome-wide scale in a number of cancer cells. It reliably detected

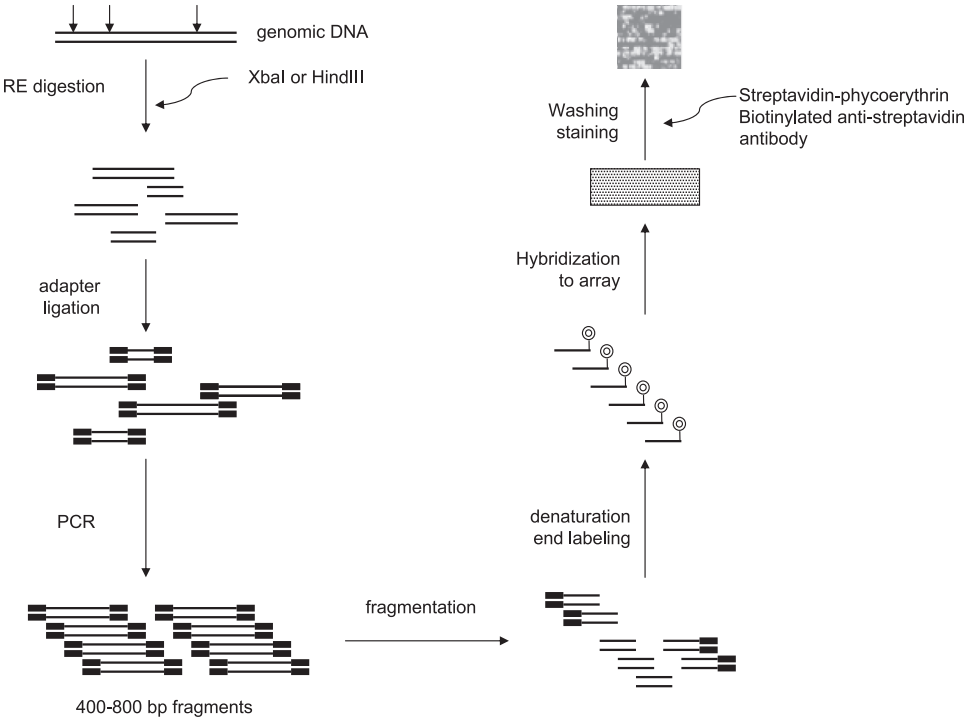


**FIGURE 8** Gene copy number analysis using SNP genotyping microarrays. The genomic DNA sample is digested with a restriction endonuclease for complexity reduction. Adapters are ligated to the restriction fragments and the fragments are amplified by PCR. The PCR products are denatured, labeled, and hybridized to a microarray. The array is washed and stained to produce signal intensities for each probe. The data for multiple SNPs are smoothed and compared to an internal control to generate an estimate of the copy number for the chromosomal region.

chromosomal amplifications as well as homozygous and hemizygous deletions simultaneously with LOH detection [114]. The arrays produced results generally comparable with those obtained on BAC and cDNA arrays, but the authors reported a substantially lower noise level and a much higher resolution, averaging approximately 300 kb [114].

The next generation of SNP microarray has an increased SNP coverage (approximately 114,000 SNPs), which corresponds to a resolution of <100 kb. The software used for data analysis smoothes the signals for all the SNPs and compares the data with the internal reference dataset for over 100 individuals. The output of the software is the absolute copy number for each SNP position on the chip [115]. The new microarray has been validated in a large study of 101 lung carcinoma samples (tumors and cell lines) [116]. The increased resolution of the array enabled identification of several small amplifications and homozygous deletions that had not previously been detected by other CGH protocols.

One possible limitation on the use of high density oligonucleotide arrays for CGH is the requirement for high purity of the target tissue in the sample. The SNP array protocol uses a PCR amplification step, which may produce nonlinear ampli-

fication effects and thus diminish the ability of the procedure to detect deletions. In particular, Zhao et al. [114] reported a drop in the accuracy of scoring homozygous deletions with decreasing purity of tumor DNA. For example, the deletion of a small region in a cell line was detectable only with ≥90% purity of tumor DNA. Deletion of another region was only detected at 100% purity of tumor DNA. The future will show whether this limitation can be overcome in subsequent generations of oligonucleotide-based CGH arrays, but it should be considered today when making the platform decision.

Most existing CGH protocols require at least 500 ng of input DNA for the labeling reaction, which is equivalent to approximately 50,000–100,000 cells. This often presents a constraint as many clinical specimens are small. Numerous DNA amplification methods have been suggested for CGH on small samples [117–123]. The type of sample is also an important consideration in array CGH. The easiest type of sample to work with is cell culture, because the isolation of high quality DNA is routine and the cell population is homogeneous. Analysis of frozen tissue samples presents more difficulties because of the potential sample heterogeneity. For example, tumor samples often contain significant amounts of normal tissue, and this dilutes the signal obtained for aberrations in the tumor. Profiling of archived FFPE samples by CGH presents the greatest challenges, because of the poor quality of DNA isolated from such samples. Fixation protocols used in hospitals typically result in a number of known alterations in DNA, including degradation, cross-linking, and modification or loss of nucleic bases [124–126]. The average fragment size of DNA decreases with increasing fixation time [127]. The concentration of formalin used for tissue fixation and the age of the sample also affect the quality of the genomic DNA preparation [128].

Archived tissue samples represent an invaluable resource for genetic analysis because the existence of large banks of FFPE tissues with clinical annotation makes possible retrospective analysis of correlation between the genomic profile of the disease and the outcome or response to treatment. This goal undoubtedly justifies the amount of effort devoted to the optimization of FFPE CGH protocols. Additionally, the task of genomic analysis would be significantly facilitated if the fixation protocols used by hospitals were standardized, thus eliminating the variation in the DNA quality. Obviously, a protocol minimizing DNA degradation would be preferred.

### 3.2 Applications of CGH in Preclinical Development

Unlike gene expression microarrays, CGH has not yet become a mainstream tool in drug discovery, mostly because it was not commercially available until recently. However, we believe that CGH has a great potential to become a dominant tool in the next decade, particularly in cancer. Indeed, CGH detects alterations at the DNA level, which are believed to be fundamental to the disease. Chromosomal alterations are the main genetic feature of cancer; they often play a causative role in tumorigenesis by altering intracellular signaling and gene expression.

In preclinical development of oncology compounds, one frequently has to face decisions on selecting the appropriate *in vitro* and *in vivo* model systems. The genetic heterogeneity of cancer necessitates stratification of patients on the basis of the predominant genetic aberrations. The susceptibility to novel targeted agents

correlates with the presence and the amount of the target, which may in turn be determined by the presence of a genetic aberration. Therefore, the genomic profile of the preclinical model system needs to reflect that of the target patient population. For example, Herceptin, an approved agent for the treatment of breast cancer, was developed using cell lines with a HER2 amplification as a model system [129, 130].

Herceptin acts by binding to the HER2 receptor and causing its internalization and the blockage of signal transduction [10–12, 131]. The *HER2* gene is frequently amplified in breast cancer; its amplification status is determined by FISH [132]. When a monoclonal antibody was developed against the extracellular domain of HER2, it was first tested on cultured breast cancer cells [129]. Of six mammary carcinoma cell lines tested, only the lines with HER2 amplification (SK-BR3, MBA-MB-175, and MDA-MD-361) were sensitive to the antibody. The established correlation between HER2 amplification and sensitivity to Herceptin was later used in the clinical development of the drug [10, 11].

In most cases, however, *in vitro* and *in vivo* model systems for compound screening are used without consideration for their genomic profile. Most commonly, oncology drug candidates are screened using cultured cancer cells and rodent xenografts that comprise the same cell lines grown subcutaneously in immunocompromised mice. One issue with using tumor cell lines is that the cells have been cultured on plastic for many generations, have acquired additional chromosomal aberrations, and therefore are not representative of the original tumor. In most cases, the main (and the only) selection criterion for cells to be used as model systems is their tissue origin. Such indiscriminate selection of model systems is one of the factors behind the current low success rate in the clinic [133, 134].

Most transformed cell lines used in preclinical development possess multiple secondary chromosomal aberrations that have been acquired during their propagation on plastic. These aberrations are not reflective of the genomic profile of the tumor of origin, but they may affect the sensitivity of cells to drugs. We have recently screened 23 small cell lung carcinoma cell lines previously used for compound screening *in vitro* and detected numerous large chromosomal amplifications and deletions that have not been detected in the target tumor [135]. In other studies, concurrent CGH screening of primary tumors and cultured cell lines of the same origin revealed gene copy number changes in the cell lines that were not present in the tumor samples [116].

In the near future, the adoption of whole-genome CGH screens should enable rational selection of model systems based on their genomic profiles. The first step in the selection process should be a comprehensive analysis of the genetic heterogeneity of the target population (Fig. 9). Model systems would then be selected whose genomic profiles reflect those found in the patient population (e.g., primary cell lines established from tumors). If the selected model systems vary in their sensitivity to the lead compounds, their genomic profiles need to be correlated with the sensitivity/resistance status. One of the established statistical procedures (e.g., analysis of variance, or ANOVA) can be used for this correlation. Once genomic features are identified that correlate with the sensitivity to the lead compounds, they can be used to stratify patients. If the correlation is confirmed in the early phases of clinical development, then enrollment of patients into Phase III should be based on the presence of this key genomic feature. Patient stratification based

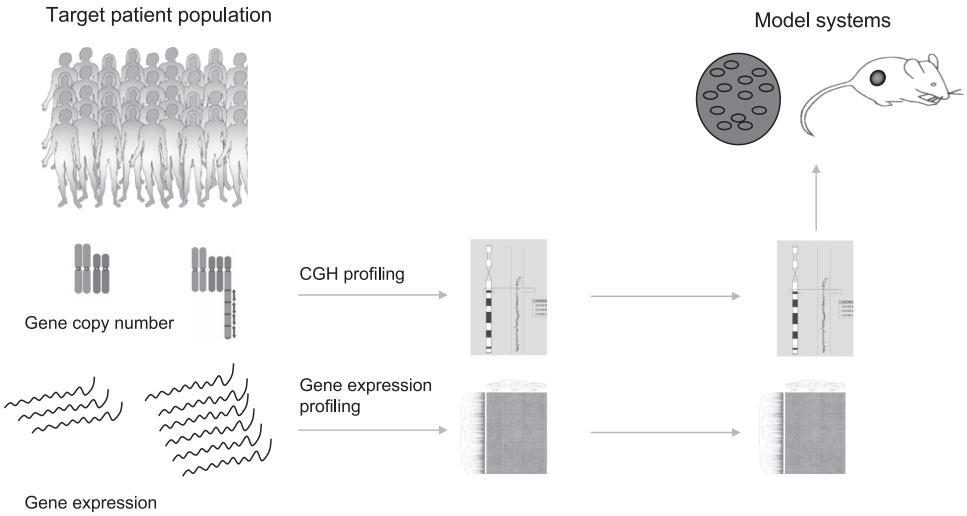


**FIGURE 9** Genomics-based selection of model systems for preclinical drug development in oncology. The genetic heterogeneity of the patient population is examined by comprehensive genomic profiling with CGH and expression microarrays. A preclinical model is selected based on the similarity of its genomic profile to that of the target patient population, in particular, with respect to the key genomic aberrations that affect the sensitivity to the drugs under development. Testing of the compounds in the preclinical models is accompanied by detailed genomic analysis of the models to identify the key genomic features that correlate with the sensitivity to the compounds. The presence of these genomic features will then serve as a criterion for selecting patients for clinical trials.

on genomic biomarkers would improve the response rate in clinical trials and decrease the duration of the trial and its cost.

A number of recent studies have proved the feasibility of comprehensive profiling of cancers for gene copy number. For several types of cancer, correlation has been established between the genetic profile and the outcome or susceptibility to existing treatments. For example, neuroblastoma, a childhood tumor derived from neural crest cells, displays remarkable genetic heterogeneity, with several well-known recurrent aberrations, such as MYCN amplification, 17q gains, and 1p losses [136, 137]. The presence of these aberrations correlates with the outcome [136, 137], raising a possibility of patient stratification and creation of targeted therapies. This suggests that any neuroblastoma drug development program would benefit from a stratification scheme, whereby the model system would be screened for genetic abnormalities and the response to the agent would be correlated with the genomic profile.

A number of studies have been performed to identify genes amplified in specific tumors. Cheng et al. [138] have profiled ovarian and breast cancers using array CGH and identified a recurrent amplification on chromosome 1q22. The amplification was centered on the gene coding for a small GTPase called RAB25. The aberration was associated with decreased survival in both types of cancer. RAB25 was then shown to increase anchorage-independent cell proliferation and suppress apoptosis and anoikis. The authors concluded that RAB25 represents an attractive therapeutic

target. Chibon et al. [139] used BAC-based CGH arrays to characterize the genomic profiles of malignant fibrous histiocytoma, an aggressive tumor, which shows no distinct line of differentiation. The 6q23 band was found to be frequently amplified in this type of cancer. The authors characterized the genes residing in the amplified region and identified ASK1 (MAP3K5) as a candidate target for therapeutic intervention. Tonon et al. [140] profiled 44 lung adenocarcinomas and small cell carcinomas and determined the most frequent aberrations in these tumors. Two genomic regions, 8p12 and 20q11, were found to be frequently amplified in both types of cancer. Several genes residing in these regions were identified, and their overexpression was confirmed by quantitative RT-PCR. Ehlers et al. [141] used both CGH and gene expression microarrays to study uveal melanoma and determined that the gain of chromosome 8 correlates most strongly with the expression of DDEF1, a gene located at 8q24. It was shown that overexpression of DDEF1 results in increased cell motility. The authors concluded that DDEF1 represents an attractive therapeutic target.

In the past few years, CGH has been used to discover patient stratification biomarkers for several types of cancer. Paris et al. [142] profiled archived tumors from 64 prostate cancer patients, 32 of whom had recurred postoperatively. The authors identified a loss at 8p23.2 that was associated with advanced disease, but most importantly, they discovered a chromosomal gain at 11q13.1 that was predictive of postoperative recurrence, independent of the stage of the disease. One gene (MEN1) coding for a nuclear protein menin was mapped to the amplified region, and its expression was correlated with disease recurrence, thus establishing the gene as a biomarker of the aggressive recurrent disease. In another study, 35 gastric carcinomas were profiled on ~2400-element BAC CGH arrays, and the patterns of chromosomal aberrations were correlated with the clinical history of the patients [16]. Hierarchical clustering of the CGH profiles revealed three predominant groups. Membership in these groups correlated with lymph node status and survival. Patients from cluster 3 had a significantly better prognosis than patients from clusters 1 and 2. Although no commonly amplified genes were identified, this study clearly demonstrates the power of genome-wide CGH in patient stratification and biomarker discovery.

A study by Wreesmann et al. [143] dealt with papillary thyroid cancer (PTC). Two distinct variants of this disease exist, the more aggressive tall-cell thyroid cancer (TCV) and conventional thyroid cancer (cPTC). A panel of 25 TCV and 45 cPTCs was profiled by CGH and gene expression microarrays to identify genomic features that would distinguish between the two subtypes of the disease. Significant differences were identified in the patterns of chromosomal gains and losses. One gene, MUC1, was of particular interest because it was both amplified and overexpressed in TCV. Its overexpression was confirmed by immunohistochemistry on independent TCV samples. Multivariate analysis showed a significant correlation between MUC1 expression level and the outcome of treatment, establishing the gene as a prognostic marker in PTC.

Diffuse large B-cell lymphoma is a highly heterogeneous disease that displays significant diversity with respect to clinical presentation and outcome. Different subtypes of the disease require different therapeutic approaches. A recent report [144] described CGH profiling of 224 diffuse large B-cell lymphomas previously examined by expression microarrays. The authors reported a high degree of correlation between the gene copy number and the expression data and identified a chromosomal region (3p11-p12)

that provided prognostic information that is statistically independent of the previously built gene expression-based model.

The first steps have been taken toward developing a diagnostic tool based on a CGH array. Schwaenen et al. [145] reported a novel CGH chip that can detect recurrent chromosomal abnormalities in B-cell chronic lymphocytic leukemia (B-CLL). The chip contained a total of 644 DNA elements and covered all the known regions frequently altered in B-CLL, as well as some other B-cell neoplasms. The array was tested and validated in 106 primary B-CLL tumor samples.

Although the bulk of the CGH work has been of an exploratory nature, one can expect that comprehensive CGH profiling of preclinical development models will become routine in the near future. Once the recurrent aberrations are identified for the tumor type of interest, a focused array similar to that described earlier can be developed to routinely interrogate preclinical model systems and stratify patients in clinical trials according to their genomic profile. Thus, the use of CGH in preclinical development will facilitate development of agents targeted to specific patient populations.

## 4 GENE SILENCING

We define some of the gene silencing terms used in this chapter.

*Antisense Oligonucletide*. A short DNA fragment complementary to the target mRNA sequence.

*RnaseH*. An enzyme that specifically binds to and degrades RNA–DNA duplexes.

*Dicer*. An RNase III-type endonuclease that processes long double-stranded RNA into siRNA.

*RISC*. An RNA–protein complex containing an siRNA molecule and effector proteins that unwind the siRNA and catalyze the cleavage of the target mRNA.

*Transfection*. Transfer of nucleic acids into the cell.

*Electroporation*. Transfection of cells with nucleic acids by using electric current to increase the permeability of the cells.

High throughput alteration of gene activity has recently emerged as a powerful tool in drug discovery. Gene knock-down screens are often performed in cell culture to identify therapeutic targets, and gene overexpression is a common tool in target validation. In the area of preclinical development, manipulation of the target gene's activity is used in compound optimization and selection. In this section we describe gene silencing, a set of techniques frequently used in drug development to identify therapeutic targets and optimize candidate compounds.

### 4.1 Antisense Oligonucleotides and Ribozymes

Antisense oligonucleotides were historically the first tool for targeted gene silencing. Their initial applications were focused on inhibiting the replication of viruses

in cell culture [146]. Typically, antisense oligonucleotides ~20 nucleotides in length are used. To increase stability and prevent enzymatic degradation, antisense oligonucleotides are usually modified, for example, by introducing phosphorothioate groups [147]. Other common types of modification include morpholino [148] and methoxyethyl [149] groups and peptide nucleic acids (PNAs) [150].

The mechanism of gene silencing by antisense oligonucleotides is illustrated in Fig. 10. They act by hybridizing to mRNA and causing RNaseH-mediated degradation of the transcript or creating a steric hindrance for the enzymes that catalyze translation and splicing [151].

In more than 20 years since the first use, antisense oligonucleotides have produced notable successes, including an approved therapeutic agent (Vitravene™ by Isis Pharmaceuticals), but their applications were limited by several drawbacks of the technology. Most importantly, antisense oligonucleotides have revealed significant nonspecific effects [82, 147, 152, 153]. In particular, phosphothioate oligonucleotides have been shown to bind various proteins and thus cause off-target effects [153]. Some oligonucleotides also induce expression of interferons through binding to Toll-like receptors [154, 155]. The potency of antisense oligonucleotides is determined by many factors, including the type of modification and accessibility of the target. Overall, they are less potent as the more recently discovered gene silencing tools [156–159].

Ribozymes also hybridize to mRNA through Watson–Crick base pairing and catalyze its degradation through hydrolysis of the phosphodiester bonds (reviewed in Ref. 160). The most commonly used type is the hammerhead ribozyme. The catalytic sequence of a ribozyme is flanked by sequences complementary to the target transcript, a feature that provides target specificity. The accessibility of the target sequence is an important factor, which determines the potency of silencing [161].

### 4.2 Short Interfering RNA

***Mechanism of Gene Silencing*** A breakthrough in the field of functional genomics was achieved with the discovery of RNA interference in the late 1990s. In 1998, Fire et al. [162] showed that double-stranded RNA causes specific gene silencing when injected into *Caenorhabditis elegans*. Several years later it was found that in *Drosophila melanogaster* embryo extracts, long double-stranded RNA is cleaved into

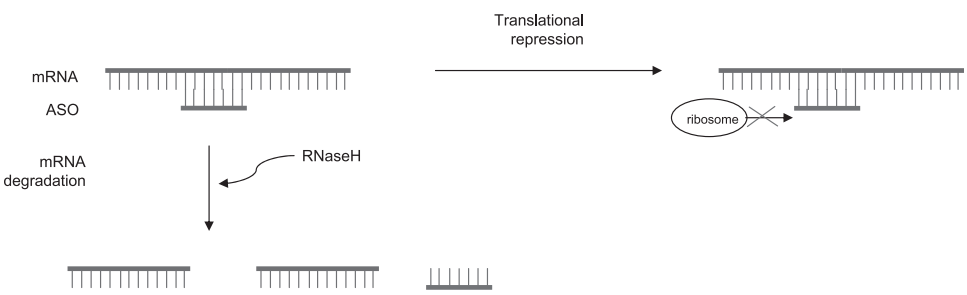


**FIGURE 10** Mechanism of gene silencing by antisense oligonucleotides. Binding of an antisense oligonucleotide to the target mRNA results in target degradation by RNaseH and translational repression due to steric hindrances.

fragments containing approximately 22 base pairs by an enzyme complex called Dicer [163]. This observation prompted the first use of synthetic RNA fragments to induce RNA interference. Elbashir et al. [83, 84] demonstrated that chemically synthesized 21 and 22 nucleotide RNA duplexes cause degradation of homologous mRNA when introduced into cultured cells. These duplexes were named short interfering RNAs (siRNAs). The mechanism of gene silencing by siRNA is illustrated in Fig. 11 (for a recent review, see Ref. 164). Long dsRNA is processed in the cell by the Dicer. The enzyme cleaves long dsRNA strands into fragments 21–28 nucleotides long, which contain two-nucleotide overhangs. The siRNA duplexes are then incorporated into an enzymatic complex called RNA-induced silencing complex (RISC). Within the silencing complex, the siRNA duplex is unwound in an ATP-dependent manner. Once the duplex is unwound, the antisense strand guides the

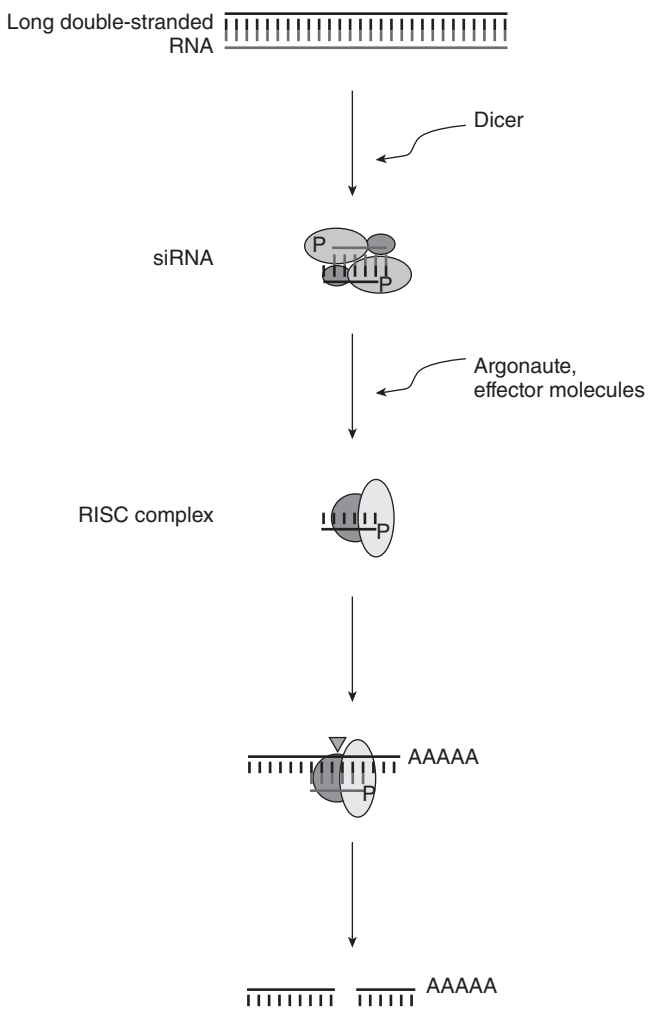


**FIGURE 11** Mechanism of gene silencing by siRNA. Long double-stranded RNA is cleaved by the Dicer complex. The resulting siRNA binds to the RISC complex and is guided to the target site to degrade the target mRNA.

RISC complex to the target site on the mRNA. The latter then causes endonucleolytic cleavage of the target mRNA sequence [164, 165].

In a typical *in vitro* siRNA experiment, the researcher transfects cells with chemically synthesized siRNA duplexes (typically 22 nucleotides long with two-nucleotide overhangs on both ends) and 24–48 hours later measures the effects of the siRNA on the target levels by either QPCR or Western blot. Alternatively, a vector coding for a so-called short hairpin RNA (shRNA) can be transiently or stably transfected into cells to produce a more lasting silencing effect [166]. Several reports have addressed the issue of differential siRNA potency [167, 168]. Duplexes that target different regions of the same gene vary significantly in their potency. The potency of siRNA is likely to be determined by the primary and secondary structures of the target site [164]. More research is necessary to clearly define the rules for designing potent siRNAs.

Another important consideration in gene silencing is the specificity of the silencing tool. In the case of siRNA, the low concentrations used and the enzymatic nature of its mechanism are factors that diminish the potential for nonspecific effects. There are several types of nonspecific effects that could be displayed by siRNAs, including (1) degradation of mRNAs other than the target due to cross-hybridization, (2) binding to cellular proteins in a sequence-specific manner (aptamer effect), (3) translational silencing through the miRNA effect, and (4) induction of the "dsRNA response" that is nonspecific with respect to the siRNA sequence. Several recent studies have demonstrated that transfection of siRNA into the cell does not cause global nonspecific effects on gene expression [169, 170]. However, other studies have described nonspecific effects caused by siRNA, which depended on the concentration, cell type, and the type of siRNA [171, 172]. Overall, the combination of high potency, relatively higher specificity, and ease of finding accessible target sites have quickly made siRNA the preferred gene silencing tool.

***Use of siRNA for Target Identification*** One of the major applications of siRNA is in drug discovery. The initial application was in target identification and validation. A typical siRNA-based target identification screen involves high throughput transfection of a large siRNA library into cultured cells and observation of the resulting phenotypes, most often by an established assay. Such screens are most common in cancer drug discovery. Prerequisites for a successful screen include a representative cell culture model, a robust transfection protocol, and a reliable quantitative assay for the desired phenotype. In cancer drug discovery, the desired phenotype is typically cell death, apoptosis, or inhibition of cell proliferation. Other assays are currently being adapted to high throughput siRNA screens, including cell migration, invasion, and colony formation assays. Cells that are hard to transfect, for example, primary cells, may be subjected to electroporation. Electroporation has been shown to be amenable to high throughput screening [173]. While the results of a comprehensive cancer drug discovery screen are yet to be published, several focused studies have been reported. Berns et al. [174] used a set of retroviral vectors targeting approximately 8000 different human genes for suppression. This RNAi library was used in human cells to identify five new modulators of the p53-dependent proliferation arrest. Suppression of these genes confers resistance to both p53-dependent and p19ARF-dependent proliferation arrest and abolishes a DNA-damage-induced G1 cell-cycle arrest.

To facilitate high throughput screening with siRNA libraries, an siRNA microarray has recently been developed [175]. The siRNAs are deposited onto a microchip, which is then used for reverse transfection of mammalian cells, thus facilitating high throughput screening of hundreds of genes in multiple cell types. The technique has shown significant promise for drug discovery, but it has several limitations. In particular, only adherent cells can be used in the screen and the transfection conditions need to be optimized for each cell type.

***Use of siRNA for Target Validation*** Short interfering RNA has recently been shown to be an extremely useful tool for target validation. It has been applied to validate targets identified in DNA microarray experiments [176, 177]. Indeed, DNA microarrays identify genes overexpressed in the diseased tissue, but they do not prove the functional involvement of the genes into the disease process. Meanwhile, if silencing these genes with siRNA reverses the disease phenotype, the target candidate is typically taken to the next stage of validation. In a recent study from the oncology area, Williams et al. [176] used DNA microarrays to profile colon tumors from 20 patients alongside with normal colon tissue samples. Over 500 genes were found to be consistently overexpressed in one-third of the cancer samples, and 13 of them were confirmed by quantitative real-time PCR. To identify genes that play an important role in colorectal carcinogenesis, siRNA was used to disrupt expression of several of the overexpressed genes in a colorectal carcinoma cell line, HCT116. Silencing of one of these genes (survivin) severely reduced tumor growth both *in vitro* and in an *in vivo* xenograft model, thus validating the target in colorectal cancer.

In a recent study, Li and colleagues [178] used siRNA to validate a target *in vivo*. They used inducible RNA interference in mice to silence hypoxia-inducible factor-1α (HIF-1α) in established xenograft tumors. It was shown that HIF-1α inhibition results in transient tumor stasis or tumor regression. A differential requirement of HIF-1α for tumor growth was also observed among different tumor types. Examination of tumors resistant to HIF-1α inhibition suggested that the resistance might result from a less hypoxic tumor environment and that the level of HIF-1α expression in tumors may be a useful marker for predicting tumor response to HIF-1α inhibition. This study demonstrated the versatility of inducible RNAi as a tool for evaluating cancer targets *in vivo*.

Overall, siRNA and DNA microarrays are highly complementary technologies, when used successively in the target discovery process. First, DNA microarrays can be used to identify genes overexpressed in a diseased tissue. Then siRNA can be used to silence these genes in a disease model system, while the phenotype is observed. If silencing of a gene affects the disease phenotype, then the gene may indeed have functional involvement in the disease process. This would warrant further investigation, primarily in an *in vivo* system.

***Use of siRNA for Compound Characterization*** The application of siRNA in compound selection and optimization is rapidly gaining momentum. Several possible schemes have been developed for compound selection that involve combined use of siRNA-based gene silencing and DNA microarrays. siRNA against the chosen target can be used to generate gene expression signatures associated with target knock-down. Candidate compounds can then be screened in the same system and

their microarray signatures can be compared with that of the siRNA (Fig. 6). Theoretically, if the compounds inhibit the target, the overlap of the compound and the siRNA signatures should represent the target inhibition signature, while the signatures unique to the compounds represent their off-target effects. In our experience, however, an siRNA against a gene and an inhibitor of the gene product do not always produce similar signatures. This could be explained by the fact that silencing of a gene and inhibiting its product are fundamentally different events. Indeed, inhibitors are often directed against one specific site of a protein, leaving the possibility that the other sites may still maintain their function or interact with other proteins. Meanwhile, silencing the gene results in the elimination of the entire protein. The time course of the two events is almost certainly different as well: 24 hours after the addition of the inhibitor one should not expect to see the same effects as 24 hours after the siRNA transfection. Additionally, either silencing or inhibition may be incomplete, with the residual activity responsible for the differences in the gene expression effects between the silencing agent and the inhibitor. Nonspecific effects of both the inhibitor and the siRNA may also contribute to the divergence. While nonspecific effects are expected from the compound, the siRNA is meant to serve as the gold standard of specificity in this type of experiment. One approach to selecting only the specific effects of siRNA is to profile several siRNAs against the target and then select the overlap of their signatures as the target silencing profile.

Overall, manipulation of gene activity has become a routine tool in drug discovery. Today, it is hard to imagine a target identification or validation strategy without a step that would involve manipulation of the target activity. The approach is becoming more common with each new technique developed, as the techniques provide higher potency and specificity.

## 5 CONCLUSION

In the years following the publication of the human genome sequence, genomic technologies have begun to revolutionize drug discovery and development. Drug development has undergone a major paradigm change, whose essence is in shifting from the trial-and-error approach to an innovative, hypothesis-driven, and systematic strategy based on target selection and validation, followed by selection and optimization of a compound that would modulate the activity of the target with minimal side effects.

Today, the initial stages of the drug development process, target identification and validation, almost always involve analysis of gene content or expression or manipulation of gene activity. Genomic technologies are also beginning to be used in preclinical and clinical development, following innovative proof-of-concept studies published in the past few years. As we have demonstrated in this chapter, the use of genomics in preclinical development may help bridge the gap between the preclinical and clinical stages of drug development by helping select preclinical models more relevant to the target disease. Thus, genomics may solve one of the most significant problems in today's drug development, the high failure rate in clinical trials of drugs that have been selected based on their activity in preclinical systems. One of the implications of this upcoming change for pharmaceutical

research and development organizations is the necessity of organizational changes to facilitate close collaboration between discovery and development. Just as the promise of the biomarker research compelled the managers of many pharmaceutical companies to create cross-functional translational biology teams, the new strategies for genomics-based preclinical development may necessitate formation of groups dedicated to the genomic analysis of target patient population and the appropriate selection of preclinical model systems. It is not always easy to quantify the benefits of genomics in the drug development process. The list of purely genomics-based drugs remains short. However, the success of the new drug development paradigm is dependent on the genomic data and therefore there is no alternative to continuing and expanding the use of genomic technologies in drug discovery and development.

## REFERENCES


1. Orth AP, et al. The promise of genomics to identify novel therapeutic targets. *Expert Opin Ther Targets* 2004;8(6):587–596.
2. Lander ES, et al. Initial sequencing and analysis of the human genome. *Nature* 2001;409(6822):860–921.
3. Venter JC, et al. The sequence of the human genome. *Science* 2001;291(5507): 1304–1351.
4. Pollack JR, et al. Genome-wide analysis of DNA copy-number changes using cDNA microarrays. *Nat Genet* 1999;23(1):41–46.
5. Linn SC, et al. Gene expression patterns and gene copy number changes in dermatofibrosarcoma protuberans. *Am J Pathol* 2003;163(6):2383–2395.
6. Pollack JR, et al. Microarray analysis reveals a major direct role of DNA copy number alteration in the transcriptional program of human breast tumors. *Proc Natl Acad Sci USA* 2002;99(20):12963–12968.
7. Heidenblad M, et al. Microarray analyses reveal strong influence of DNA copy number alterations on the transcriptional patterns in pancreatic cancer: implications for the interpretation of genomic amplifications. *Oncogene* 2005;24(10):1794–1801.
8. Lewin DA, Weiner MP. Molecular biomarkers in drug development. *Drug Discov Today* 2004;9(22):976–983.
9. Frank R, Hargreaves R. Clinical biomarkers in drug discovery and development. *Nat Rev Drug Discov* 2003;2(7):566–580.
10. Cobleigh, MA, et al. Multinational study of the efficacy and safety of humanized anti-HER2 monoclonal antibody in women who have HER2-overexpressing metastatic breast cancer that has progressed after chemotherapy for metastatic disease. *J Clin Oncol* 1999;17(9):2639–2648.
11. Vogel CL, et al. Efficacy and safety of trastuzumab as a single agent in first-line treatment of HER2-overexpressing metastatic breast cancer. *J Clin Oncol* 2002;20(3):719–726.
12. Carter P, et al. Humanization of an anti-p185HER2 antibody for human cancer therapy. *Proc Natl Acad Sci USA* 1992;89(10):4285–4289.
13. Singh D, et al. Gene expression correlates of clinical prostate cancer behavior. *Cancer Cell* 2002;1(2):203–209.

14. Shipp MA, et al. Diffuse large B-cell lymphoma outcome prediction by gene-expression profiling and supervised machine learning. *Nat Med* 2002;8(1):68–74.
15. Weiss MM, et al. Genomic alterations in primary gastric adenocarcinomas correlate with clinicopathological characteristics and survival. *Cell Oncol* 2004;26(5–6):307–317.
16. Weiss MM, et al. Genomic profiling of gastric cancer predicts lymph node status and survival. *Oncogene* 2003;22(12):1872–1879.
17. Pomeroy SL, et al. Prediction of central nervous system embryonal tumour outcome based on gene expression. *Nature* 2002;415(6870):436–442.
18. vant Veer LJ, et al. Gene expression profiling predicts clinical outcome of breast cancer. *Nature* 2002;415(6871):530–536.
19. Khan J, et al. Classification and diagnostic prediction of cancers using gene expression profiling and artificial neural networks. *Nat Med* 2001;7(6):673–679.
20. Bohen SP, et al. Variation in gene expression patterns in follicular lymphoma and the response to rituximab. *Proc Natl Acad Sci USA* 2003;100(4):1926–1930.
21. Takata R, et al. Predicting response to methotrexate, vinblastine, doxorubicin, and cisplatin neoadjuvant chemotherapy for bladder cancers through genome-wide gene expression profiling. *Clin Cancer Res* 2005;11(7):2625–2636.
22. Gunther EC, et al. Prediction of clinical drug efficacy by classification of drug-induced genomic expression profiles *in vitro*. *Proc Natl Acad Sci USA* 2003;100(16):9608–9613.
23. Chomczynski P, Sacchi N. Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Anal Biochem* 1987;162(1):156–159.
24. Chomczynski P. A reagent for the single-step simultaneous isolation of RNA, DNA and proteins from cell and tissue samples. *Biotechniques* 1993;15(3):532–534, 536–537.
25. Van Gelder RN, et al. Amplified RNA synthesized from limited quantities of heterogeneous cDNA. *Proc Natl Acad Sci USA* 1990;87(5):1663–1667.
26. Perou CM, et al. Molecular portraits of human breast tumours. *Nature* 2000;406(6797):747–752.
27. Huang J, et al. Effects of ischemia on gene expression. *J Surg Res* 2001;99(2):222–227.
28. Masuda N, et al. Analysis of chemical modification of RNA from formalin-fixed samples and optimization of molecular biology applications for such samples. *Nucleic Acids Res* 1999;27(22):4436–4443.
29. Cronin M, et al. Measurement of gene expression in archival paraffin-embedded tissues: development and performance of a 92-gene reverse transcriptase–polymerase chain reaction assay. *Am J Pathol* 2004;164(1):35–42.
30. Erlander MG, et al. Global gene expression profiling from formalin-fixed paraffin-embedded (FFPE) tissues: implications for retrospective and prospective clinical studies. *Proc Am Soc Clin Oncol* 2003;22:125.
31. http://www.arctur.com/research_portal/products/paradise_main.htm.
32. Emmert-Buck MR, et al. Laser capture microdissection. *Science* 1996;274(5289): 998–1001.
33. Ma XJ, et al. A two-gene expression ratio predicts clinical outcome in breast cancer patients treated with tamoxifen. *Cancer Cell* 2004;5(6):607–616.
34. Fuller AP, et al. Laser capture microdissection and advanced molecular analysis of human breast cancer. *J Mammary Gland Biol Neoplasia* 2003;8(3):335–345.
35. Ma XJ, et al. Gene expression profiles of human breast cancer progression. *Proc Natl Acad Sci USA* 2003;100(10):5974–5979.
36. Tusher VG, Tibshirani R, Chu G. Significance analysis of microarrays applied to the ionizing radiation response. *Proc Natl Acad Sci USA* 2001;98(9):5116–5121.

37. Tsai CA, Hsueh HM, Chen JJ. Estimation of false discovery rates in multiple testing: application to gene microarray data. *Biometrics* 2003;59(4):1071–1081.
38. Delongchamp RR, et al. Multiple-testing strategy for analyzing cDNA array data on gene expression. *Biometrics* 2004;60(3):774–782.
39. Lee ML, et al. Importance of replication in microarray gene expression studies: statistical methods and evidence from repetitive cDNA hybridizations. *Proc Natl Acad Sci USA* 2000;97(18):9834–9839.
40. Novak JP, Sladek R, Hudson TJ. Characterization of variability in large-scale gene expression data: implications for study design. *Genomics* 2002;79(1):104–113.
41. Chen JJ, et al. Analysis of variance components in gene expression data. *Bioinformatics* 2004;20(9):1436–1446.
42. Golub TR, et al. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science* 1999;286(5439):531–537.
43. Ramaswamy S, et al. Multiclass cancer diagnosis using tumor gene expression signatures. *Proc Natl Acad Sci USA* 2001;98(26):15149–15154.
44. Bittner M, et al. Molecular classification of cutaneous malignant melanoma by gene expression profiling. *Nature* 2000;406(6795):536–540.
45. Quackenbush J. Computational analysis of microarray data. *Nat Rev Genet* 2001;2(6):418–427.
46. Raychaudhuri S, et al. Basic microarray analysis: grouping and feature reduction. *Trends Biotechnol* 2001;19(5):189–193.
47. Eisen MB, Brown PO. DNA arrays for analysis of gene expression. *Methods Enzymol* 1999;303:179–205.
48. Tavazoie S, et al. Systematic determination of genetic network architecture. *Nat Genet* 1999;22(3):281–285.
49. Toronen P, et al. Analysis of gene expression data using self-organizing maps. *FEBS Lett* 1999;451(2):142–146.
50. Brown MP, et al. Knowledge-based analysis of microarray gene expression data by using support vector machines. *Proc Natl Acad Sci USA* 2000;97(1):262–267.
51. Nutt CL, et al. Gene expression-based classification of malignant gliomas correlates better with survival than histological classification. *Cancer Res* 2003;63(7):1602–1607.
52. Doniger SW, et al. MAPPFinder: using Gene Ontology and GenMAPP to create a global gene-expression profile from microarray data. *Genome Biol* 2003;4(1):R7.
53. Volinia S, et al. GOAL: automated Gene Ontology analysis of expression profiles. *Nucleic Acids Res* 2004;32(Web Server issue):W492–W49.
54. http://david.niaid.nih.gov/david/.
55. Vaquerizas JM, et al. GEPAS, an experiment-oriented pipeline for the analysis of microarray gene expression data. *Nucleic Acids Res* 2005;33(Web Server issue):W616–W620.
56. Al-Shahrour F, et al. BABELOMICS: a suite of web tools for functional annotation and analysis of groups of genes in high-throughput experiments. *Nucleic Acids Res* 2005;33(Web Server issue):W460–W464.
57. Al-Shahrour F, Diaz-Uriarte R, Dopazo J. FatiGO: a web tool for finding significant associations of Gene Ontology terms with groups of genes. *Bioinformatics* 2004;20(4):578–580.
58. Mlecnik B, et al. PathwayExplorer: web service for visualizing high-throughput expression data on biological pathways. *Nucleic Acids Res* 2005;33(Web Server issue):W633–W637.
59. www.ingenuity.com.

60. www.genego.com.
61. Harris MA, et al. The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Res* 2004;32(Database issue):D258–D261.
62. Ashburner M, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* 2000;25(1):25–29.
63. www.biocarta.com.
64. Kanehisa M, et al. The KEGG resource for deciphering the genome. *Nucleic Acids Res* 2004;32(Database issue):D277–D280.
65. Semizarov D, Kroeger P, Fesik S. siRNA-mediated gene silencing: a global genome view. *Nucleic Acids Res* 2004;32(13):3836–3845.
66. Zucchini C, et al. Identification of candidate genes involved in the reversal of malignant phenotype of osteosarcoma cells transfected with the liver/bone/kidney alkaline phosphatase gene. *Bone* 2004;34(4):672–679.
67. Meyer MH, Dulde E, Meyer RA Jr. The genomic response of the mouse kidney to low-phosphate diet is altered in X-linked hypophosphatemia. *Physiol Genomics* 2004;18(1):4–11.
68. Rhodes DR, et al. Meta-analysis of microarrays: interstudy validation of gene expression profiles reveals pathway dysregulation in prostate cancer. *Cancer Res* 2002;62(15): 4427–4433.
69. Jiang H, et al. Joint analysis of two microarray gene-expression data sets to select lung adenocarcinoma marker genes. *BMC Bioinformatics* 2004;5(1):81.
70. Sorlie T, et al. Repeated observation of breast tumor subtypes in independent gene expression data sets. *Proc Natl Acad Sci USA* 2003;100(14):8418–8423.
71. Shen R, Ghosh D, Chinnaiyan AM. Prognostic meta-signature of breast cancer developed by two-stage mixture modeling of microarray data. *BMC Genomics* 2004;5(1):94.
72. Yeoh EJ, et al. Classification, subtype discovery, and prediction of outcome in pediatric acute lymphoblastic leukemia by gene expression profiling. *Cancer Cell* 2002; 1(2):133–143.
73. Armstrong SA, et al. MLL translocations specify a distinct gene expression profile that distinguishes a unique leukemia. *Nat Genet* 2002;30(1):41–47.
74. Armstrong SA, et al. Inhibition of FLT3 in MLL. Validation of a therapeutic target identified by gene expression based classification. *Cancer Cell* 2003;3(2):173–183.
75. MacDonald TJ, et al. Expression profiling of medulloblastoma: PDGFRA and the RAS/MAPK pathway as therapeutic targets for metastatic disease. *Nat Genet* 2001;29(2):143–152.
76. Dhanasekaran SM, et al. Delineation of prognostic biomarkers in prostate cancer. *Nature* 2001;412(6849):822–826.
77. Magee JA, et al. Expression profiling reveals hepsin overexpression in prostate cancer. *Cancer Res* 2001;61(15):5692–5696.
78. Fromont G, et al. Differential expression of 37 selected genes in hormone-refractory prostate cancer using quantitative taqman real-time RT-PCR. *Int J Cancer* 2005;114(2):174–181.
79. Sarang SS, et al. Discovery of molecular mechanisms of neuroprotection using cell-based bioassays and oligonucleotide arrays. *Physiol Genomics* 2002;11(2):45–52.
80. Huang E, et al. Gene expression phenotypic models that predict the activity of oncogenic pathways. *Nat Genet* 2003;34(2):226–230.
81. Svaren J, et al. EGR1 target genes in prostate carcinoma cells identified by microarray analysis. *J Biol Chem* 2000;275(49):38524–38531.

82. Cho YS, et al. Antisense DNAs as multisite genomic modulators identified by DNA microarray. *Proc Natl Acad Sci USA* 2001;98(17):9819–9823.
83. Elbashir SM, et al. Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. *Nature* 2001;411(6836):494–498.
84. Elbashir SM, Lendeckel W, Tuschl T. RNA interference is mediated by 21- and 22-nucleotide RNAs. *Genes Dev* 2001;15(2):188–200.
85. Ziegelbauer J, Wei J, Tjian R. Myc-interacting protein 1 target gene profile: a link to microtubules, extracellular signal-regulated kinase, and cell growth. *Proc Natl Acad Sci USA* 2004;101(2):458–463.
86. Glaser KB, et al. Gene expression profiling of multiple histone deacetylase (HDAC) inhibitors: defining a common gene set produced by HDAC inhibition in T24 and MDA carcinoma cell lines. *Mol Cancer Ther* 2003;2(2):151–163.
87. Bottone FG Jr et al. Gene modulation by the cyclooxygenase inhibitor, sulindac sulfide, in human colorectal carcinoma cells: possible link to apoptosis. *J Biol Chem* 2003;278(28):25790–25801.
88. Clarke PA, Te Poele R, Workman P. Gene expression microarray technologies in the development of new therapeutic agents. *Eur J Cancer* 2004;40(17):2560–2591.
89. Waring JF, et al. Clustering of hepatotoxins based on mechanism of toxicity using gene expression profiles. *Toxicol Appl Pharmacol* 2001;175(1):28–42.
90. Burczynski ME, et al. Toxicogenomics-based discrimination of toxic mechanism in HepG2 human hepatoma cells. *Toxicol Sci* 2000;58(2):399–415.
91. Thomas RS, et al. Identification of toxicologically predictive gene sets using cDNA microarrays. *Mol Pharmacol* 2001;60(6):1189–1194.
92. Thomas RS, et al. Developing toxicologically predictive gene sets using cDNA microarrays and Bayesian classification. *Methods Enzymol* 2002;357:198–205.
93. Hamadeh HK, et al. Prediction of compound signature using high density gene expression profiling. *Toxicol Sci* 2002;67(2):232–240.
94. Waring JF, et al. Microarray analysis of hepatotoxins *in vitro* reveals a correlation between gene expression profiles and mechanisms of toxicity. *Toxicol Lett* 2001;120(1–3):359–368.
95. Ganter B, et al. Development of a large-scale chemogenomics database to improve drug candidate selection and to understand mechanisms of chemical toxicity and action. *J Biotechnol* 2005;119(3):219–244.
96. Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. *Clin Pharmacol Ther* 2001;69(3):89–95.
97. Ross DT, et al. Systematic variation in gene expression patterns in human cancer cell lines. *Nat Genet* 2000;24(3):227–235.
98. Scherf U, et al. A gene expression database for the molecular pharmacology of cancer. *Nat Genet* 2000;24(3):236–244.
99. Staunton JE, et al. Chemosensitivity prediction by transcriptional profiling. *Proc Natl Acad Sci USA* 2001;98(19):10787–10792.
100. Zembutsu H, et al. Genome-wide cDNA microarray screening to correlate gene expression profiles with sensitivity of 85 human cancer xenografts to anticancer drugs. *Cancer Res* 2002;62(2):518–527.
101. Zembutsu H, et al. Gene-expression profiles of human tumor xenografts in nude mice treated orally with the EGFR tyrosine kinase inhibitor ZD1839. *Int J Oncol* 2003;23(1):29–39.

102. Oshita F, et al. Genomic-wide cDNA microarray screening to correlate gene expression profile with chemoresistance in patients with advanced lung cancer. *J Exp Ther Oncol* 2004;4(2):155–160.
103. Chen X, et al. Variation in gene expression patterns in human gastric cancers. *Mol Biol Cell* 2003;14(8):3208–3215.
104. Leung SY, et al. Phospholipase A2 group IIA expression in gastric adenocarcinoma is associated with prolonged survival and less frequent metastasis. *Proc Natl Acad Sci USA* 2002;99(25):16203–16208.
105. Levsky JM, Singer RH. Fluorescence *in situ* hybridization: past, present and future. *J Cell Sci* 2003;116(Pt 14):2833–2838.
106. Kallioniemi A, et al. Comparative genomic hybridization for molecular cytogenetic analysis of solid tumors. *Science* 1992;258(5083):818–821.
107. Pinkel D, Albertson DG. Array comparative genomic hybridization and its applications in cancer. *Nat Genet* 2005;37(Suppl):S11–S17.
108. Albertson DG, Pinkel D. Genomic microarrays in human genetic disease and cancer. *Hum Mol Genet* 2003;12(Spec No 2):R145–R152.
109. Snijders AM, et al. Assembly of microarrays for genome-wide measurement of DNA copy number. *Nat Genet* 2001;29(3):263–264.
110. Pinkel D, et al. High resolution analysis of DNA copy number variation using comparative genomic hybridization to microarrays. *Nat Genet* 1998;20(2):207–211.
111. Clark J, et al. Genome-wide screening for complete genetic loss in prostate cancer by comparative hybridization onto cDNA microarrays. *Oncogene* 2003;22(8):1247–1252.
112. Hyman E, et al. Impact of DNA amplification on gene expression patterns in breast cancer. *Cancer Res* 2002;62(21):6240–6245.
113. Barrett MT, et al. Comparative genomic hybridization using oligonucleotide microarrays and total genomic DNA. *Proc Natl Acad Sci USA* 2004;101(51):17765–17770.
114. Zhao X, et al. An integrated view of copy number and allelic alterations in the cancer genome using single nucleotide polymorphism arrays. *Cancer Res* 2004;64(9): 3060–3071.
115. www.affymetrix.com.
116. Zhao X, et al. Homozygous deletions and chromosome amplifications in human lung carcinomas revealed by single nucleotide polymorphism array analysis. *Cancer Res* 2005;65(13):5561–5570.
117. Hosono S, et al. Unbiased whole-genome amplification directly from clinical samples. *Genome Res* 2003;13(5):954–964.
118. Devries S, et al. Array-based comparative genomic hybridization from formalin-fixed, paraffin-embedded breast tumors. *J Mol Diagn* 2005;7(1):65–71.
119. Lage JM, et al. Whole genome analysis of genetic alterations in small DNA samples using hyperbranched strand displacement amplification and array-CGH. *Genome Res* 2003;13(2):294–307.
120. Daigo Y, et al. Degenerate oligonucleotide primed-polymerase chain reaction-based array comparative genomic hybridization for extensive amplicon profiling of breast cancers: a new approach for the molecular analysis of paraffin-embedded cancer tissue. *Am J Pathol* 2001;158(5):1623–1631.
121. Wang G, et al. Balanced-PCR amplification allows unbiased identification of genomic copy changes in minute cell and tissue samples. *Nucleic Acids Res* 2004;32(9):E76.
122. Tanabe C, et al. Evaluation of a whole-genome amplification method based on adaptor-ligation PCR of randomly sheared genomic DNA. *Genes Chromosomes Cancer* 2003;38(2):168–176.

123. Guillaud-Bataille M, et al. Detecting single DNA copy number variations in complex genomes using one nanogram of starting DNA and BAC-array CGH. *Nucleic Acids Res* 2004;32(13):E112.
124. Giannella C, et al. Comparison of formalin, ethanol, and Histochoice fixation on the PCR amplification from paraffin-embedded breast cancer tissue. *Eur J Clin Chem Clin Biochem* 1997;35(8):633–635.
125. Koshiba M, et al. The effect of formalin fixation on DNA and the extraction of high-molecular-weight DNA from fixed and embedded tissues. *Pathol Res Pract* 1993;189(1):66–72.
126. Tbakhi A, et al. The effect of fixation on detection of B-cell clonality by polymerase chain reaction. *Mod Pathol* 1999;12(3):272–278.
127. Ghazvini S, et al. Comparative genomic hybridization analysis of archival formalin-fixed paraffin-embedded uveal melanomas. *Cancer Genet Cytogenet* 1996;90(2):95–101.
128. Semizarov D, unpublished observations.
129. Hudziak RM, et al. p185HER2 monoclonal antibody has antiproliferative effects *in vitro* and sensitizes human breast tumor cells to tumor necrosis factor. *Mol Cell Biol* 1989;9(3):1165–1172.
130. Lupu R, et al. Direct interaction of a ligand for the erbB2 oncogene product with the EGF receptor and p185erbB2. *Science* 1990;249(4976):1552–1555.
131. Harris M. Monoclonal antibodies as therapeutic agents for cancer. *Lancet Oncol* 2004;5(5):292–302.
132. Hicks DG, Tubbs RR. Assessment of the *HER2* status in breast cancer by fluorescence *in situ* hybridization: a technical review with interpretive guidelines. *Hum Pathol* 2005;36(3):250–261.
133. Kola I, Landis J. Can the pharmaceutical industry reduce attrition rates? *Nat Rev Drug Discov* 2004;3(8):711–715.
134. Kamb A. What's wrong with our cancer models? *Nat Rev Drug Discov* 2005;4(2):161–165.
135. Semizarov D, unpublished observations.
136. Maris JM. The biologic basis for neuroblastoma heterogeneity and risk stratification. *Curr Opin Pediatr* 2005;17(1):7–13.
137. Mosse YP, et al. High-resolution detection and mapping of genomic DNA alterations in neuroblastoma. *Genes Chromosomes Cancer* 2005;43(4):390–403.
138. Cheng KW, et al. The RAB25 small GTPase determines aggressiveness of ovarian and breast cancers. *Nat Med* 2004;10(11):1251–1256.
139. Chibon F, et al. ASK1 (MAP3K5) as a potential therapeutic target in malignant fibrous histiocytomas with 12q14-q15 and 6q23 amplifications. *Genes Chromosomes Cancer* 2004;40(1):32–37.
140. Tonon G, et al. High-resolution genomic profiles of human lung cancer. *Proc Natl Acad Sci USA* 2005;102(27):9625–9630.
141. Ehlers JP, et al. DDEF1 is located in an amplified region of chromosome 8q and is overexpressed in uveal melanoma. *Clin Cancer Res* 2005;11(10):3609–3613.
142. Paris PL, et al. Whole genome scanning identifies genotypes associated with recurrence and metastasis in prostate tumors. *Hum Mol Genet* 2004;13(13):1303–1313.
143. Wreesmann VB, et al. Genome-wide profiling of papillary thyroid cancer identifies MUC1 as an independent prognostic marker. *Cancer Res* 2004;64(11):3780–3789.
144. Bea S, et al. Diffuse large B-cell lymphoma subgroups have distinct genetic profiles that influence tumor biology and improve gene expression-based survival prediction. *Blood* 2005;106(9):3183–3190.

145. Schwaenen C, et al. Automated array-based genomic profiling in chronic lymphocytic leukemia: development of a clinical tool and discovery of recurrent genomic alterations. *Proc Natl Acad Sci USA* 2004;101(4):1039–1044.

146. Zamecnik PC, Stephenson ML. Inhibition of Rous sarcoma virus replication and cell transformation by a specific oligodeoxynucleotide. *Proc Natl Acad Sci USA* 1978;75(1):280–284.

147. Stein CA. Does antisense exist? *Nat Med* 1995;1(11):1119–1121.

148. Summerton J, Weller D. Morpholino antisense oligomers: design, preparation, and properties. *Antisense Nucleic Acid Drug Dev* 1997;7(3):187–195.

149. Dorn G, et al. Specific inhibition of the rat ligand-gated ion channel P2X3 function via methoxyethoxy-modified phosphorothioated antisense oligonucleotides. *Antisense Nucleic Acid Drug Dev* 2001;11(3):165–174.

150. Egholm M, et al. PNA hybridizes to complementary oligonucleotides obeying the Watson–Crick hydrogen–bonding rules. *Nature* 1993;365(6446):566–568.

151. Dias N, Stein CA. Antisense oligonucleotides: basic concepts and mechanisms. *Mol Cancer Ther* 2002;1(5):347–355.

152. Fisher AA, et al. Evaluating the specificity of antisense oligonucleotide conjugates. A DNA array analysis. *J Biol Chem* 2002;277(25):22980–22984.

153. Lebedeva I, Stein CA. Antisense oligonucleotides: promise and reality. *Annu Rev Pharmacol Toxicol* 2001;41:403–419.

154. Roman M, et al. Immunostimulatory DNA sequences function as T helper-1-promoting adjuvants. *Nat Med* 1997;3(8):849–854.

155. Hafner M, et al. Antimetastatic effect of CpG DNA mediated by type I IFN. *Cancer Res* 2001;61(14):5523–5528.

156. Kretschmer-Kazemi Far R, Sczakiel G. The activity of siRNA in mammalian cells is related to structural target accessibility: a comparison with antisense oligonucleotides. *Nucleic Acids Res* 2003;31(15):4417–4424.

157. Grunweller A, et al. Comparison of different antisense strategies in mammalian cells using locked nucleic acids, 2′-O-methyl RNA, phosphorothioates and small interfering RNA. *Nucleic Acids Res* 2003;31(12):3185–3193.

158. Miyagishi M, Hayashi M, Taira K. Comparison of the suppressive effects of antisense oligonucleotides and siRNAs directed against the same targets in mammalian cells. *Antisense Nucleic Acid Drug Dev* 2003;13(1):1–7.

159. Xu Y, et al. Effective small interfering RNAs and phosphorothioate antisense DNAs have different preferences for target sites in the luciferase mRNAs. *Biochem Biophys Res Commun* 2003;306(3):712–717.

160. Doudna JA, Cech TR. The chemical repertoire of natural ribozymes. *Nature* 2002;418(6894):222–228.

161. Kuwabara T, Warashina M, Taira K. Cleavage of an inaccessible site by the maxizyme with two independent binding arms: an alternative approach to the recruitment of RNA helicases. *J Biochem (Tokyo)* 2002;132(1):149–155.

162. Fire A, et al. Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. *Nature* 1998;391(6669):806–811.

163. Zamore PD, et al. RNAi: double-stranded RNA directs the ATP-dependent cleavage of mRNA at 21 to 23 nucleotide intervals. *Cell* 2000;101(1):25–33.

164. Dorsett Y, Tuschl T. siRNAs: applications in functional genomics and potential as therapeutics. *Nat Rev Drug Discov* 2004;3(4):318–329.

165. Dykxhoorn DM, Novina CD, Sharp PA. Killing the messenger: short RNAs that silence gene expression. *Nat Rev Mol Cell Biol* 2003;4(6):457–467.
166. Lee NS, et al. Expression of small interfering RNAs targeted against HIV-1 rev transcripts in human cells. *Nat Biotechnol* 2002;20(5):500–505.
167. Vickers TA, et al. Efficient reduction of target RNAs by small interfering RNA and RNase H-dependent antisense agents. A comparative analysis. *J Biol Chem* 2003;278(9):7108–7118.
168. Holen T, et al. Positional effects of short interfering RNAs targeting the human coagulation trigger tissue factor. *Nucleic Acids Res* 2002;30(8):1757–1766.
169. Semizarov D, et al. Specificity of short interfering RNA determined through gene expression signatures. *Proc Natl Acad Sci USA* 2003;100(11):6347–6352.
170. Chi JT, et al. Genomewide view of gene silencing by small interfering RNAs. *Proc Natl Acad Sci USA* 2003;100(11):6343–6346.
171. Jackson AL, et al. Expression profiling reveals off-target gene regulation by RNAi. *Nat Biotechnol* 2003;21(6):635–637.
172. Persengiev SP, Zhu X, Green MR. Nonspecific, concentration-dependent stimulation and repression of mammalian gene expression by small interfering RNAs (siRNAs). *Rna* 2004;10(1):12–18.
173. Walters DK, Jelinek DF. The effectiveness of double-stranded short inhibitory RNAs (siRNAs) may depend on the method of transfection. *Antisense Nucleic Acid Drug Dev* 2002;12(6):411–418.
174. Berns K, et al. A large-scale RNAi screen in human cells identifies new components of the p53 pathway. *Nature* 2004;428(6981):431–437.
175. Wheeler DB, Carpenter AE, Sabatini DM. Cell microarrays and RNA interference chip away at gene function. *Nat Genet* 2005;37(Suppl):S25–S30.
176. Williams NS, et al. Identification and validation of genes involved in the pathogenesis of colorectal cancer using cDNA microarrays and RNA interference. *Clin Cancer Res* 2003;9(3):931–946.
177. Zhou A, et al. Identification of NF-kappa B-regulated genes induced by TNFalpha utilizing expression profiling and RNA interference. *Oncogene* 2003;22(13):2054–2064.
178. Li L, et al. Evaluating hypoxia-inducible factor-1alpha as a cancer therapeutic target via inducible RNA interference in vivo. *Cancer Res* 2005;65(16):7249–7258.

# PROTEOMICS

Juan Casado and J. Ignacio Casal
*Spanish National Cancer Center (CNIO), Madrid, Spain*

## Contents

1 Introduction
2 Protein Extraction and Purification
  2.1 Disruption of Tissues and Solubilization of Proteins
  2.2 Depletion of Abundant Proteins
3 Two-Dimensional Polyacrylamide Gel Electrophoresis
  3.1 Protein Sample Preparation
  3.2 First Dimension: Isoelectrofocusing
  3.3 Second Dimension: Polyacrylamide Gel Electrophoresis—Choice of Gel Size, Concentration, and Protein Molecular Weight Range
4 Gel Staining and Image Processing
  4.1 Coomassie Brilliant Blue and Silver Staining
  4.2 Fluorescent Dyes
  4.3 Differential In-Gel Electrophoresis Staining
  4.4 Software-Based Image Analysis for Quantification and Comparison of Protein Patterns
  4.5 Drawbacks and Limitations of 2D-PAGE
5 In-Gel and In-Solution Endoprotease Digestion of Proteins
  5.1 Reduction and Alkylation of Proteins
  5.2 Trypsin Digestion
  5.3 Alternative Proteases
  5.4 In-Gel Digestion of Proteins
  5.5 In-Solution Digestion of Proteins
6 Principles of Mass Spectrometry: Ionization Sources and Different Mass Analyzers
  6.1 Matrix Assisted Laser Desorption Ionization–Time of Flight (MALDI-TOF): Spotting on MALDI Plates and Use of Different Matrices
  6.2 Acquisition of Peptide Mass Fingerprints (PMFs) for Protein Identification
  6.3 Time of Flight–Time of Flight (TOF-TOF) Analysis of Precursor Ions

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

7 Electrospray Ionization of Proteins and Tandem Mass Spectrometry
7.1 Basic Concepts of Electrospray-Based Mass Analyzers
7.2 Mass Determination of Intact Proteins
7.3 Peptide Bond Fragmentation Using Collision Energy
7.4 Electron Transfer Dissociation
8 Multidimensional Liquid Chromatography Coupled to Mass Spectrometers
8.1 Multidimensional LC-ESI-Q-TRAP
8.2 LC-MALDI-TOF(-TOF)
9 Database Search and Use of Bioinformatic Tools for Protein Identification
9.1 Search Engines Against Nucleotide and Amino Acid Databases
9.2 Use of Customized Databases for Tagged Recombinant Proteins and *De Novo* Sequencing
10 Identification of Posttranslational Modification of Proteins
11 Protein Quantitation
12 Protein Arrays
12.1 Abundance-Based Microarrays
12.2 Function-Based Protein Microarrays
13 Links of Interests
14 Conclusion
References

## 1 INTRODUCTION

The term *proteomics* was first coined in the mid-1990s [1] and has become a key research area during the last decade. It is dedicated to study the biological implication of proteins in disease, pathologies, or alterations of living organisms. The *proteome* was first defined as the protein complement of the genome [1]. The proteome is dynamic and its composition changes as a consequence of alterations in an effort to adapt to environmental changes and keep the homeostasis of living systems. There are at least four main questions that proteomic studies aim to find a response for: (1) How many proteins change as a consequence of an alteration? (2) How much do these proteins change? (3) Did the changes observed involve the regulation of any biochemical pathways? (4) What is the biological relevance or can we take advantage of that information to solve the alteration? To answer these questions, a number of techniques are routinely used to comprehensively and systematically separate, visualize, identify, and quantify the plethora of proteins coexisting in tissues and cells.

Two-dimensional polyacrylamide gel electrophoresis (2D-PAGE), high performance liquid chromatography (HPLC), and mass spectrometry (MS) are the most important tools in proteomics. The combination of all these techniques allows the identification and quantification of up to thousands of proteins, the identification of proteins differentially expressed in normal versus altered organisms/tissues, and the characterization of posttranslational modification of proteins.

## 2 PROTEIN EXTRACTION AND PURIFICATION

One of the most important steps in a proteomic study is the selection of the biological material to work with. Whole proteome analysis of any organism, cell, organelle, biological fluid, or a single protein can be considered. Therefore, due to the great diversity of protein samples used, the optimal procedure must be determined empirically for each sample. Here, we provide some general guidelines.

### 2.1 Disruption of Tissues and Solubilization of Proteins

The proteins of interest must be completely soluble. A number of chemicals are frequently found in literature for protein solubilization, the most frequently used being 8 M urea or 6 M guanidine hydrochloride.

Proteomic experiments can also focus on the identification and characterization of a subset of proteins of interest, such as membrane proteins [2, 3] or glycoproteins. A wide number of chemicals have been described to be useful for the extraction of membrane proteins including treatment with alkaline pH, *N*-acetyl trimethylammonium, and different concentrations of detergents, such as Triton X-100, Triton X-114 [4], *N*-lauroylsarcosine, *N*-octylglucoside, and sodium dodecyl sulfate (SDS) among others. The enrichment of glycoproteins is achieved by two different methods, using hydrazide chemistry or lectin binding.

### 2.2 Depletion of Abundant Proteins

Some of the proteins with meaningful biological information can be present at low concentrations in the experimental samples. One example is the identification of proteins present in human plasma, containing six abundant proteins (albumin, IgG, IgA, haptoglobin, transferrin, and alpha-1 anti-trypsin) that can be depleted from the sample by selective immunodepletion of the proteins using chromatographic columns containing antibodies against everyone of the six proteins [5]. The selective removal of the most abundant "housekeeping" proteins favors the enrichment of proteins present at a lower abundance in the sample that can be visualized by 2D-PAGE and identified with mass spectrometry.

## 3 TWO-DIMENSIONAL POLYACRYLAMIDE GEL ELECTROPHORESIS

Two-dimensional polyacrylamide gel electrophoresis (2D-PAGE) allows the separation of hundreds (up to thousands) of proteins in a single experiment. The result is a gel matrix where proteins appear as single spots that can be visualized after a protein staining procéss. After digitalization of the gel to obtain an image with the help of transmission light scanners, the subsequent images can be processed for spot analysis. It can be used to find differential expression profiles between normal versus altered protein patterns and to quantify these changes by densitometric measurement of the intensity of the protein spots observed in the gel. The quality and biological significance of the results are compromised by the protein sample preparation

and the reproducibility of the experiments. The main steps involved in 2D-PAGE analysis and some of the potential problems are described next.

### 3.1 Protein Sample Preparation

Proteins can be found soluble in the cytosol, embedded in cell membranes, and bound to nucleic acids or to other proteins. For 2D-PAGE, it is necessary that the pool of proteins submitted be completely soluble. Different treatments can be used to solubilize different types of proteins. The choice of cell/tissue disruption method, composition of lysis buffer, and the use of detergents directly affect the solubilization of proteins.

Since proteases are normally present in biological samples and proteolysis may cause undesired results in the 2D gels, cocktails of protease inhibitors should be added during the protein extraction steps. Some of the protease inhibitors frequently used include 4-(2-aminoethyl) benzenesulfonyl fluoride (AEBSF), ethylenediaminetetraacetic acid (EDTA), bestatin, E-64, leupeptin, aprotinin, and phenylmethylsulfonyl fluoride (PMSF), among others.

If only a subset of proteins from the biological sample is of interest (i.e., subcellular proteins), prefractionation methods need to be employed [6]. This is normally achieved by differential solubilization or ultracentrifugation in sucrose gradients. Once the proteins of interest have been extracted, it is common to include some steps to remove contaminants. Dialysis for salt removal, lyophilization, precipitation of proteins with trichloroacetic acid–cold acetone [7], or DNA removal with the use of endonucleases are among the current procedures.

***Protocol A: Protein Cleaning and Preparation for Isoelectrofocusing*** Sample volumes corresponding to 100 μg of protein, previously quantified [8], are pelleted by centrifugation at 13,000 rpm after addition of 6% trichloroacetic acid in chilled acetone and incubation at 4 °C for 30 min [7]. The addition of 0.2% deoxycholate usually facilitates the precipitation of proteins. The precipitation can be repeated several times if there is an excess of salts in the sample of interest. After removal of the supernatant, the protein pellet is dried under a nitrogen stream at room temperature. The appearance of several spots corresponding to the same protein is frequently observed in 2D gels. This can be attributed most times to the oxidation state of thiol groups in the protein of interest. This problem is solved by oxidizing the thiol groups, for instance, adding chemicals, such as DeStreak (Amersham Biosciences), prior to or during the rehydration of immobilized pH gradient (IPG) strips.

### 3.2 First Dimension: Isoelectrofocusing

Prior to the isoelectrofocusing (IEF), dry and clean precipitates need to be solubilized in a rehydration solution (typically 100–300 μL) containing 8 M urea, 2% 3-[(3-cholamidopropyl)dimethylamonio]-1-propanesulfonate (CHAPS), 2 mg/mL dithiothreitol (DTT), 0.0125 g/mL iodoacetamide (IAM), a trace of bromophenol blue, and carrier ampholytes IPG buffer (commercially available depending on the pH range of the IPG strip). Proteins are embedded in the IPG strip at low voltages (50 V per IPG strip) and are ready for IEF. It consists of different steps: a first step

of 500 V for 1 h, a second step of 1000 V for 1 h. The final step requires 8,000 V for 4 h, until a total voltage of about 40,000 (volt × hour)V · h. After completion of the IEF, each strip is deposited in a vial with 5 mL of equilibrating solution and left with agitation for 15 min.

### 3.3 Second Dimension: Polyacrylamide Gel Electrophoresis—Choice of Gel Size, Concentration, and Protein Molecular Weight Range

After the IEF, the IPG strip must be equilibrated by covering it with a solution containing 50 mM tris(hydroxymethyl)aminomethane (Tris-base) pH 8.8, 6 M urea, 30% glycerol, 2% SDS, 0.01 g bromophenol blue, and 10 mg/mL DTT. Proteins are separated according to their molecular weight under denaturing conditions in the presence of SDS [9].

***Protocol B: Standard SDS-PAGE Gel Solution for 100 mL***

| Solution for 12.5% Acrylmide-Bisacrylamide Gels | Volume |
|---|---|
| Stock solution 40% acrylamide/bisacrylamide (37.5:1, 2.6% C) | 41.7 mL |
| 1.5 M Tris-HCl pH 8.8 | 25 mL |
| 10% Sodium dodecyl sulfate (SDS) | 1 mL |
| Water | 31.8 mL |
| 10% Ammonium persulfate | 0.5 mL |
| *N,N,N′,N′*-tetramethylethylenediamine (TEMED) | 33 μL |

Once the gel is polymerized, the IPG strips, containing the proteins focused at their isoelectric point (p*I*) and equilibrated, are placed on top of the gel and the proteins are separated according to their molecular weight in a chamber containing running buffer (25 mM Tris-HCl, 192 mM glycine, and 0.1% SDS).

## 4 GEL STAINING AND IMAGE PROCESSING

The staining of 2D gels is a crucial step in proteomic analysis. It allows the visualization of protein spots resolved on the 2D gel and enables further analysis, including the identification of differential proteins among samples, densitometric quantification, and recovery of proteins of interest.

### 4.1 Coomassie Brilliant Blue and Silver Staining

An improved method for Coomassie Brilliant Blue (CBB) stained gel is the colloidal Coomassie Brilliant Blue [10]. Silver staining is used to visualize proteins that are present at a low concentration in the gel (2 ng/spot) (Fig. 1a).

***Protocol C: Colloidal Coomassie Brilliant Blue G-250 Staining*** Fix the gels with 40% methanol and 10% acetic acid for 30 min. Prepare a stock solution of colloidal Coomassie Brilliant Blue G-250 staining as follows: mix 4 g of Brilliant Blue G-205 with 230 mL of 85% phosphoric acid and stir. Add 640 mL of saturated ammonium sulfate solution and stir. Add distilled water to 4 L. Prepare 100 mL of staining

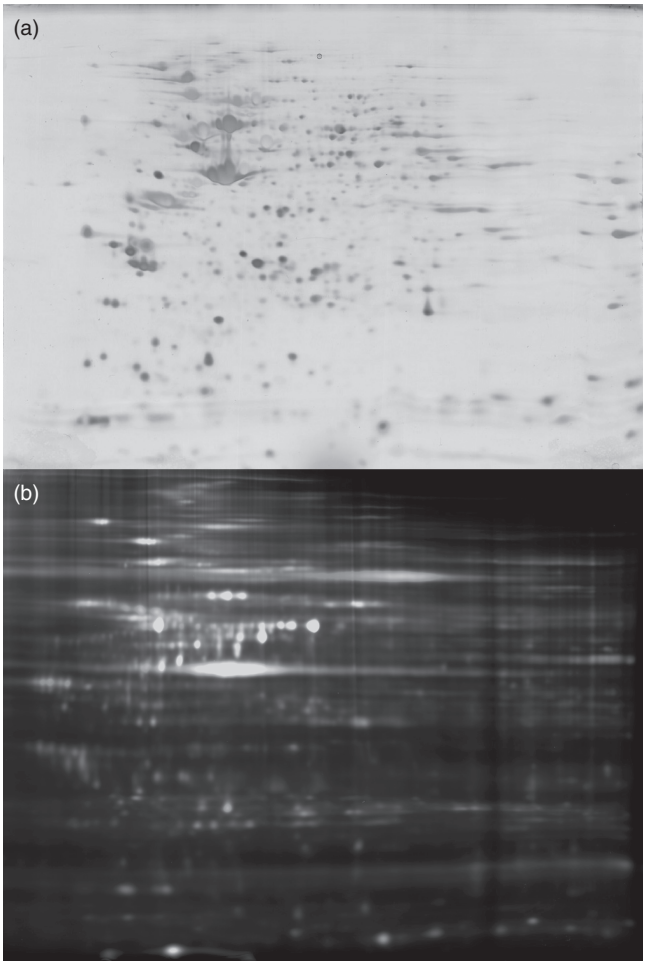


**FIGURE 1** Representative images corresponding to (a) silver stained 2D gels (14 × 16 cm) and (b) DIGE stained corresponding to the overlapping Cy3 + Cy5 images. Protein spots appear all over the gel matrix separated according to their isoelectric point and molecular weight. After staining, the spots of interest (differential spots) can be recovered and further processed for protein identification and characterization using mass spectrometry and database searches.

solution by adding 20 mL methanol to 80 mL of the stock solution. The staining solution should be prepared immediately before use, as the colloid only lasts for a few hours. Stain gels overnight (if maximum sensitivity is required) using a rocking platform to avoid colloidal stain sticking to the surface of the gel. Finally, destain the gels with 3–5 changes of 25% methanol until the spots appear clearly defined in the gel. The use of rocking platforms during the staining is crucial, because the staining of gels without agitation can cause the appearance of noise and background during software-based image densitometry of the gels. The duration and temperature of staining and destaining steps are also crucial in order to obtain reproducible results.

***Protocol D: Silver Staining/Destaining of Proteins in Polyacrylamide Gels*** Cover the gel with fixing solution (50% methanol, 10% acetic acid), remove the fixing solution, cover gel with 5% methanol for 15 min, and wash 3 times with distilled water. Cover the gel with sensitizing solution 0.2 g/L of $Na_2S_2O_3$ freshly prepared for 2 min and wash with distilled water. Cover gel with 0.2 g/L of silver nitrate ($AgNO_3$) for 25 min and wash 3 times with distilled water. Cover gel with developing solution containing 3 g/100 mL $Na_2CO_3$, 50 μL/100 mL of 37% HCOH, and 2 mL of sensitizing solution. Allow staining until the protein spots are visible over the background of the gel (10 min maximum). Remove the developing solution and stop with 14 g/L $Na_2$-ethylenediaminetetraacetic acid (EDTA) for 10 min. Finally, wash with distilled water.

Once stained, the gels can be destained by covering the gels with 0.2 g potassium ferricyanide ($K_3Fe(CN)_6$) in 100 mL of sensitizing solution, leaving a yellow background in the gel. Restaining is also possible starting the protocol again by adding the silver nitrate solution (adapted from Ref. 11).

### 4.2 Fluorescent Dyes

Silver staining, while sensitive, exhibits a low dynamic range. Handling may significantly change the results and the background of the gels frequently interferes with further characterization of silver-stained proteins using mass spectrometry. Fluorescent staining is superior in sensitivity and linear response as compared to Coomassie Brilliant Blue and silver staining, respectively [12]. A range of fluorescent dyes are commercially available for different purposes, including Sypro$^{(r)}$ Ruby, Deep purple, 5-hexadecanoylamino-fluorescein [13]. Sypro Ruby is a sensitive fluorescent stain for detecting proteins separated by polyacrylamide gel electrophoresis. This fluorescent stain does not interfere with subsequent analysis of proteins by mass spectrometry and is quantitative over three orders of magnitude. Apart from its high sensitivity, Sypro Ruby will not stain nucleic acids [14].

Other specific applications of fluorescent dyes for the detection of subsets of proteins have also been developed; such as the case of ProQ diamond for phosphoproteins.

***Protocol E: Sypro Ruby Staining of Proteins in Polyacrylamide Gels*** It is highly recommended to bind the gel to one of the glass plates prior to running the electrophoresis as follows: prepare a Bind-Silane working solution by mixing 8 mL of ethanol, 200 μL of acetic acid, 10 μL of Bind-Silane, and 1.8 mL of distilled water. Cover the inner surface of one of the glass plates used to polymerize the acrylamide gel and dry to room temperature for 1–2 h. Perform the assembly and polymerization of the gels as usual. The gels will stay attached to the glass during electrophoresis, staining procedures, scanning, and storage.

Incubate the gel in fixing solution (30% methanol, 7.5% acetic acid) for at least 2 h. Remove fixing solution and cover the gel with Sypro Ruby stain. Incubate 3 h or overnight with gentle shaking protected from direct light. Wash extensively with distilled water (at least 4 changes of 20 min). Finally, scan the gels and select the spots of interest. Sypro Ruby protein gel stain can be imaged using laser-based imaging systems equipped with 450, 473, 488, or 532 nm laser lines. For further details

on Sypro Ruby staining you may access http://probes.invitrogen.com/lit/bioprobes31/section1.html.

### 4.3 Differential In-Gel Electrophoresis Staining

An outstanding application of fluorescent dyes to proteomics using 2D gels is the differential in-gel electrophoresis (DIGE) technology, which is based on the use of three CyDye™ fluorochroms, Cy2, Cy3, and Cy5, for protein labeling (Fig. 1b). It allows one to label up to three samples that can be separated in a 2D gel in order to search for differential protein profiles, thus avoiding experimental variations from gel to gel and increasing the accuracy of the results. A pooled internal standard, containing equal amounts of protein from each sample, should also be created as a reference for normalization in these experiments.

Fluorescent two-dimensional DIGE constitutes the proteomic equivalent to gene expression analysis by DNA microarrays. Even minor differences of protein expression can be detected across multiple samples simultaneously with statistical confidence by using the DeCyder software [15, 16]. The comparison of spot intensities using the DIGE approach and the DeCyder software is more objective than the conventional approach based on the comparison of the brightness of gel images obtained by conventional staining. The quantification of abundance changes are obtained over a linear dynamic range of almost four orders of magnitude for the Cy dyes [15]. This novel technology is being applied to the analysis of differential protein expression in different neoplasia for the search of cancer markers [17–19] or to study drug resistance [20].

***Protocol F: DIGE Protein Labeling*** Reconstitute each vial of commercial CyDye in high quality *N,N*-dimethylformamide (DMF) to a final concentration of 400 pmol/μL. After complete resuspension, each dye gives a visible color, Cy2—yellow, Cy3—red, and Cy5—blue. Add the proper amount of dye to the different protein (up to three) extracts. Normally, 400 pmol of dye is used to label 50 μg of protein, although higher concentration of protein can be efficiently labeled by adding proportional amounts of each dye. Let dyes react with the sample during 30 min at room temperature in the dark and quench the labeling reaction by adding 10 mM lysine. Combine the labeled samples into a single microfuge tube and mix. Centrifuge for 10 min at 13,000 rpm to pellet any particles not dissolved. Add to each labeled sample and pooled samples 100–300 μL (depending on drystrip length) of rehydration solution, containing 8 M urea, 2% CHAPS, 2 mg/mL DTT, 0.0125 g/mL iodoacetamide, a trace of bromphenol blue, and carrier ampholytes IPG buffer (commercially available depending on the pH range of the IPG strip). The following steps of the 2D electrophoresis are performed as described in Section 3.2.

### 4.4 Software-Based Image Analysis for Quantification and Comparison of Protein Patterns

Software-based analysis of conventional (non-DIGE) 2D gels includes the detection of spots (individual proteins) present inside the gel matrix. Although the reproducibility from gel to gel is relatively high, some degree of experimental variability can be observed from gel to gel (even if the same protein mixture is analyzed). For this

reason, triplicates (at least) of 2D gels should be run for every individual sample in order to compare the pattern of spots that appear in each of the gels. In this way, the software is able to measure the average "volume" of each spot, measured as the area (in dots per square area of the gel surface) by the intensity of the same spot (the value of intensity assigned to each spot) assigned as a percentage of the total intensity of the surface of the 2D gel. The scanner-densitometer used has been calibrated by assigning an algorithm able to transform the transmission light that reaches the detectors through the gel [21]. Therefore, the software can reveal the differences in protein profiles and measure changes in the intensity of spots from each sample.

### 4.5 Drawbacks and Limitations of 2D-PAGE

2D-PAGE gels can resolve a limited range of proteins, in terms of molecular weight and isoelectric point. In general, high range (above 150 kDa) or low range (below 12 kDa) molecular weight proteins are difficult to detect with this technique. The same limitations apply to hydrophobic proteins (e.g., membrane proteins), which normally migrate very poorly in these gels. Finally, it should be taken into account that only relatively abundant proteins can be visualized.

Commercial p*I* strips including immobilized ampholites range from 3 (most acidic) to 11 (most basic). Although these p*I* values cover a wide range of the proteins present in living organism, it is not possible to include proteins with higher or lower p*I* points in a 2D gel.

The size of the gel needs to be considered. Larger 2D gels allow a better resolution of proteins that could appear as overlapped spots in small 2D gels, but they require longer p*I* strips and considerable consumption of acrylamide and other reagents (especially when replicates of the experiments are necessary), and the handling for staining and scanning is a hurdle.

On the other hand, conventional staining approaches have a limited dynamic range. The use of sensible staining using fluorescent dyes requires the use of equipment such as laser scanners and robotized spot-pickers to pick up the spots of interest for further analysis. This instrumentation is expensive and not available in every lab.

## 5 IN-GEL AND IN-SOLUTION ENDOPROTEASE DIGESTION OF PROTEINS

In most proteomic studies, the start-up material used for high throughput protein identification is a proteolytic mixture of the proteins. Here we describe some of the current methods for protein digestion.

### 5.1 Reduction and Alkylation of Proteins

Reduction with 10 mM DTT is necessary to break disulfide bonds and to unfold the protein. Alkylation with 50 mM iodoacetamide is an efficient reaction that yields carboxyamidomethylated residues and keeps the protein unfolded. This allows the endoprotease to access all the cleavage sites efficiently.

### 5.2 Trypsin Digestion

Trypsin is the most frequently used endoprotease for proteomic analyses for several reasons. It possesses high cleavage specificity after Lys or Arg residues (provided Pro is not the following amino acid in the protein sequence). The frequency of appearance of Lys and Arg residues inside proteins is relatively high compared to other amino acids. Modified trypsins that reduce the autoproteolysis and increase the efficiency of the reaction are readily available in the market. Moreover, trypsin cleaves after basic residues, which favors the protonation of the peptides under low acidic conditions (which is crucial for mass spectrometry in positive ionization mode). The amount of trypsin added to samples must be proportional to the amount of protein, normally 1:20 or even 1:50 (w/w) trypsin:protein ratio is a suitable ratio.

### 5.3 Alternative Proteases

In some cases, tryptic digestion is not sufficient to characterize the protein of interest because the resulting peptides are too small or too large for their separation with C18 reverse phase columns and subsequent analysis with mass spectrometry. In these cases, the combination of two (or more) endopeptidase digestions with mass spectrometric analysis of the resulting peptides enables the comprehensive identification or proteins present in complex mixtures. It is advisable to use combinations of proteases with different specificities, such as trypsin combined with chymotrypsin, Asp-N, or Glu-C.

### 5.4 In-Gel Digestion of Proteins

This method is based on the behavior of gel plugs to act as a "sponge," shrinking when acetonitrile is added or swelling when aqueous buffer or other reagents are added to the piece of gel [22].

***Protocol G: In-Gel Digestion of Proteins*** Cover the gel plug containing the protein of interest with acetonitrile for 10 min; remove the acetonitrile. Cover the gel with 10 mM DTT in 50 mM ammonium bicarbonate and store at 56 °C for 30 min; remove the DTT. Cool to room temperature. Alkylate the sample by covering the gel with 50 mM iodoacetamide in 50 mM ammonium bicarbonate and keep at room temperature in the dark for 15 min; remove the iodoacetamide. Add trypsin at a 1:20 ratio and store at 37 °C for 8 hours. Stop the endoprotease reaction by adding 1 μL of concentrated acetic acid. Remove the supernatant and place it in a new clean tube. Cover the plug with acetonitrile for 10 min and recover the supernatant. Mix both supernatants and lyophilize the peptide mixture. The peptide mixture will be ready for mass spectrometry after resuspension in appropriate buffer.

### 5.5 In-Solution Digestion of Proteins

Proteins in solution in different buffers can be lyophilized to dryness and redissolved in 50 mM ammonium bicarbonate containing relatively high concentrations of urea (4–8 M) to help the solubilization of the proteins at basic pH, necessary for tryptic or chymotryptic endoprotease digestion. When 8 M urea is used to solubilize

proteins, the concentration of urea should be diluted to 2 M prior to endoprotease digestion in order to maintain the protease activity.

***Protocol H: In-Solution Digestion of Proteins*** Lyophilize the sample to dryness and redissolve in a small volume (10 μL of 50 mM ammonium bicarbonate). Vortex and/or bath sonicate to ensure the solubilization of your protein(s) of interest. Add 10 μL of DTT in 50 mM ammonium bicarbonate and store at 56 °C for 30 min. Cool to room temperature. Alkylate by adding 10 μL of iodoacetamide in ammonium bicarbonate and keep at room temperature in the dark for 15 min. Dilute in one volume of 50 mM ammonium bicarbonate to reduce the urea concentration to 2 M. Add trypsin at a 1:20 ratio and store at 37 °C for 8 h. Your peptide mixture will be ready for mass spectrometry analysis after resuspension in the appropriate buffer.

## 6 PRINCIPLES OF MASS SPECTROMETRY: IONIZATION SOURCES AND DIFFERENT MASS ANALYZERS

Mass spectrometry has become a key area in proteomic studies [23]. Several sources of soft ionization for peptides and proteins have been developed during the last decades to allow high throughput analysis of peptides and routine identification and characterization of proteins (or their proteolytically derived peptides) present at very low concentrations in biological samples (even attomoles / nanograms of protein). Moreover, tandem mass spectrometry (MS/MS) approaches enable the confident identification of mixtures of proteins based on amino acid sequence information. Mass spectrometers are composed of three parts: the ionization source, the mass analyzer, and the detector. Several mass analyzers can be found in mass spectrometers, including triple quadrupoles (Q), time-of-flight (TOF) analyzers, ion traps (IT), Fourier cyclotron resonance (FCR), and combinations such as quadrupole time-of-flight (Q-TOF). All these analyzers are able to accurately measure the mass of compounds (proteins or peptides) that enter the analyzer in gas phase as a function of their charge state (mass-to-charge ($m/z$) ratio).

### 6.1 Matrix Assisted Laser Desorption Ionization–Time of Flight (MALDI-TOF): Spotting on MALDI Plates and Use of Different Matrices

The MALDI technique [24] has primarily been used in conjunction with TOF analyzers for molecular mass determination. For MALDI, peptides are mixed with a matrix that transfers the energy of a laser beam to the peptides embedded in the crystallized matrix. The resulting spectra, characterized by singly charged precursor ions, present high energy collision fragments, which are readily interpretable.

***Protocol I: Dried Droplet Method for Spotting of Samples on MALDI Plates*** After tryptic digestion and lyophilization, the resulting peptides are resuspended in 5 μL water with 0.1% trifluoroacetic acid. It is advisable to introduce desalting steps by using Zip-tips® containing C18 beads (commercially available). The solution containing the peptides is mixed with 5 μL of MALDI-TOF matrix. Normally, 1% α-cyano-4-hydroxy-cinnamic acid dissolved in 90% acetonitrile and 10% trifluoroacetic

acid is used as matrix for peptide mass fingerprinting of proteins. Normally, 1 μL of the mixture (peptides and matrix) is spotted on the MALDI plate and allowed to dry for 5 min before MALDI analysis.

### 6.2 Acquisition of Peptide Mass Fingerprints (PMFs) for Protein Identification

Peptide mass fingerprinting is the most common strategy for identifying proteins proceeding from SDS-PAGE separations, either 1D or 2D. The proteins are usually in-gel digested with trypsin or other endoproteases (e.g., Lys C or Glu C). The resulting proteolytic peptides are extracted from the gel piece and analyzed by MALDI-TOF MS [25]. The peptide masses contained in the resulting spectra are the peptide mass fingerprints. This mass profile is matched against the theoretical masses obtained from the *in silico* digestion of all proteins contained in the database. This interrogation is made with algorithms such as MASCOT (http://www.matrixscience.co.uk), or Profound (http://prowl.rockefeller.edu/profound_bin) among others. The choice of the algorithm may influence the final identification. A combination of several search algorithms is advisable [26]. The results of the matching are ranked according to the number of peptide masses matching their sequence within a given mass error tolerance. Successful protein identification requires several factors: (1) mass accuracy of the instrument, (2) ratio of assigned versus nonassigned peaks in the spectrum, and (3) quality and size of the database used [27]. Common databases used for the analysis are SwissProt or NCBI. The whole process is now automated and hundreds of samples can be analyzed per day. As a general rule, when MASCOT is used for identification, protein identification is considered as significant when the MASCOT score is higher than 66 and the sequence coverage is higher than 15%.

### 6.3 Time of Flight–Time of Flight (TOF-TOF) Analysis of Precursor Ions

TOF-TOF instruments combine the high throughput of the PMF analysis with the increased confidence given by peptide fragmentation that extends protein identification coverage. This approach is particularly useful when complex mixtures of peptides are analyzed. These instruments have the ability to acquire MALDI-TOF mass spectra followed by high energy collision-induced dissociation (CID) and to perform subsequent MS/MS analysis of fragment ions for more definitive protein identification. The peptide sequence information gained in MS/MS experiments leads to higher confidence in the database search results and identification of more proteins. The acquisition queue is automated in such a way that the five or ten most prominent ions of every spectrum are subjected to further MS/MS fragmentation. The accuracy of these instruments has improved considerably in the last few years; 2.5 parts per million (ppm) are now easily achieved. This accuracy is very useful for the study of proteomes from unknown genomes and the *de novo* sequencing approaches.

Although MALDI instruments are very useful for classical gel-based proteomics and protein identification, they are not the best option for the analysis of complex proteomes, since they lack multidimensionality. On the other hand, they are not

particularly appropriate for the study of posttranslational modifications and more specifically phosphoproteome analysis. For these cases, other mass spectrometry techniques based on electrospray ionization coupled to liquid chromatography are more efficient.

## 7 ELECTROSPRAY IONIZATION OF PROTEINS AND TANDEM MASS SPECTROMETRY

In electrospray ionization, proteins present in aqueous buffers are desolvated into a gas phase when flowing through a narrow capillary and are subjected to high voltage [28]. Contrary to other ionization techniques, such as fast atom bombardment (FAB), electrospray ionization (ESI) is more efficient and robust for the ionization of large organic molecules, including peptides. Therefore, it has become widely used in proteomics. Basically, the mechanism underlying ESI is that the application of voltage to a capillary (made of metal or silica), at atmospheric pressure and containing proteins or peptides in solution, produces charged tiny droplets that form a spray. As droplets evaporate, charged peptides pass from the aqueous phase to the gas phase.

### 7.1 Basic Concepts of Electrospray-Based Mass Analyzers

The basic principle of electrospray-based mass spectrometers is that ions, produced in an ESI source, can be isolated as a function of voltage applied to opposite electrodes and radiofrequency. By controlling these variables, ions can be stored, fragmented, or forced to hit a detector in order to measure their mass-to-charge ratio (*m/z*). Once peptides are charged and in the gas phase, they enter the mass spectrometer with the aid of rotary vacuum pumps. Basic amino acid residues become protonated when solved in low acidic buffers. Therefore, proteins and peptides in low acidic buffers become multiply charged and can be analyzed as positive ions.

ESI sources are usually coupled to ion traps, triple quadrupoles, and Q-TOF instruments. Ion traps and quadrupoles are able to isolate ions of interest and perform tandem mass spectrometry (MS/MS or $MS^n$) at high speed and resolution rates. In order to choose what kind of mass analyzer is more appropriate for a certain experimental approach, three issues need to be considered: mass resolution, scan speed, and accuracy of mass measurement. Ion traps are composed of a three-dimensional (3D) chamber (3D traps) or linear ion traps (i.e., LTQ, Thermo Finnegan), where ions can be stored, isolated, and fragmented in order to achieve structural information on those. Ion traps are capable of performing multiple stages of precursor ion fragmentation ($MS^n$, where $n$ = number of sequential fragmentation) in a single experiment. Linear ion traps are characterized by higher ion capacity and scan rates, compared to 3D ion traps. Triple quadrupoles can be used to detect, isolate, and fragment specific precursor ions that produce a selected product ion of interest, such as a neutral loss.

During the last few years, ultrahigh resolution equipment has been developed, including Orbitrap or Fourier transform (FT) ion cyclotron resonance technologies, which are able to reach up to sub-ppm mass accuracy.

### 7.2 Mass Determination of Intact Proteins

A direct measurement of the molecular weight of proteins can be performed with ESI-equipped mass spectrometers. In this approach, proteins of interest are resuspended in low acidic buffers (0.1% acetic or formic acid) and placed inside small capillaries with metallic tips to facilitate the appearance of a spray with the aid of voltage. Protein ions spraying from the ESI source are normally detected as a mixture of different charge states, which is called a charge envelope. The mass determination is performed by deconvoluting the set of mass-to-charge ratios, corresponding to consecutive charged species, detected for a specific protein, with specific algorithms such as MagTran (free software available at http://www.geocities.com/SiliconValley/Hills/2679/magtran.html) or Maxent software (http://www.cs.princeton.edu/~schapire/maxent/). Mass determination of intact proteins can routinely be performed on purified proteins in solution. Protein purifications using chromatographic procedures usually contain certain amounts of salt or other impurities that can dramatically affect the mass measurement of the protein of interest. Desalting steps prior to mass determination are carried out by dialysis or by reverse phase chromatography with hydrophobic columns (C2, C4, or C8 depending on the size of the protein).

Using Fourier transform ion cyclotron resonance (FTICR), the mass determination of whole intact proteins ranging from 10,000 to 100,000 Da may be provided with an accuracy near 1 ppm. Protein masses of up to 229 kDa have also been reported [29]. In some cases, this data can lead to unambiguous protein identification. However, the mass of the intact protein is compromised by posttranslational modifications or partial proteolysis of the purified protein [30]. In that case, ions corresponding to the intact protein are subsequently fragmented in the mass spectrometer, yielding the molecular masses of both the protein and the fragment ions. If a sufficient number of informative fragment ions are observed, this analysis can provide a complete description of the primary structure of the protein and reveal all of its primary structure and its modifications.

Many proteomic studies, including the ability to fragment and analyze large intact proteins for protein characterization, protein–protein interaction, and chemical cross-link location inside proteins, are called *top–down* approaches. A second widely used alternative is the *bottom–up* approach, in which complex mixtures of peptides are analyzed by mass spectrometry in two stages. In the first stage, the masses of peptides are determined; in a second stage, peptides are fragmented to produce information on the sequence of the peptides to reconstruct the primary structure of the original protein.

### 7.3 Peptide Bond Fragmentation Using Collision Energy

Peptide structure determination using mass spectrometry is based on ionization of peptides and subsequent fragmentation of the peptide amide bonds after collision-induced dissociation (CID) by controlled collision with an inert gas (such as He or Ar). The collision of ionized peptide molecules against the inert gas inside a collision chamber is sufficient to overcome the energetic threshold of the amide bonds. After fragmentation, the series of fragment ions generated from the N and C termini, respectively, are scanned, which leads to the elucidation of the primary structure

backbone of the peptide. The series of fragment ions derived from the N terminus are designated as *a, b*, or *c* ions, whereas the ions derived from the C terminus are *x, y*, or *z* ions. This is a nomenclature widely accepted for the fragmentation of peptides, first proposed in 1984 and modified four years later [31].

### 7.4 Electron Transfer Dissociation

Two highly effective methods for fragmenting large peptides and proteins, electron capture dissociation (ECD) and electron transfer dissociation (ETD) [32, 33], have emerged as promising alternatives for protein and peptide characterization, in both, top–down and bottom–up approaches. Electron transfer dissociation is an alternative to CID fragmentation that fragments the backbone of peptides with the aid of anthracene anions. Anthracene is introduced in the collision chamber of ion traps and transfers electrons to multiply protonated peptides, leading to their eventual fragmentation. ETD fragmentation technologies are still in development and require the improvement of new software bioinformatic tools to aid in the interpretation of ETD-derived MS/MS spectra. Although quite recent, ETD seems particularly well suited for characterization of posttranslational modifications, particularly phosphorylation. Therefore, it is foreseeable that ETD may become an indispensable tool for posttranslational modification and protein sequence analysis in the near future.

## 8 MULTIDIMENSIONAL LIQUID CHROMATOGRAPHY COUPLED TO MASS SPECTROMETERS

### 8.1 Multidimensional LC-ESI-Q-TRAP

Here we focus on the analysis of complex mixtures of peptides by coupling an ESI source at the end of a separation step using high performance liquid chromatography (HPLC). Although ESI was originally designed to be performed at a few microliters per minute flow rate (micro HPLC), separation techniques have rapidly evolved to run at nano-flow rates below 100 nL per minute using nanocolumns with diameters ranging from 1.5 to 3 μm. The major advantage of using nano-LC flow rates is the significant increase in the sensitivity of the identification.

Peptide mixtures can routinely be separated using reverse phase (RP) chromatographic columns packed with, normally, C18. Peptides elute as independent chromatographic peaks as long as an increasing concentration of acetonitrile is pumped through the reverse phase column. As soon as the peptides elute from the RP column and reach the ESI source, ions are produced and can readily be scanned and fragmented. If the mixture of peptides of interest is more complex, a prefractionation can be achieved by coupling a strong cation exchange (SCX) column before the RP. This is called two-dimensional liquid chromatography (2D-LC) or multidimensional protein identification (MudPit) [34]. In this approach, after injecting the peptide mixture, increasing concentrations of salt (normally from 0.05 mM up to 2 M NaCl or ammonium formate) are pumped first through the SCX column. The eluting peptides from each salt concentration go through a second separation through the RP column. Chromatographic runs from 30 min up to 6

hours can be performed, depending on the complexity of the peptide mixture. Modern HPLC instruments allow one to configure this experimental approach online, which permits a comprehensive identification of proteins derived from tryptic digestions with a high degree of automation. Nevertheless, if high concentrations of peptides need to be analyzed, the first fractionation using SCX may be performed offline using capillary or even preparative columns and salt gradients, instead of salt plugs.

The recent development of C18 beads smaller than 2 μm, together with ultrahigh pressure liquid chromatrography (UPLC) (Waters Corp.) has led to better separation and resolution during chromatographic runs and has been shown to be very advantageous for both peptide and metabolite screening.

### 8.2 LC- MALDI-TOF(-TOF)

The same principle of "divide and conquer" can be applied for MALDI-TOF analysis. Complex mixtures of peptides can be separated through RP columns and directly spotted as independent spots onto the surface of a MALDI plate, where an appropriate matrix is added for subsequent analysis. This process can be automated using specific robots such as Probot (LC-Packings). Contrary to what happens during peptide mass fingerprint analysis, where all the peptides derived from a unique protein are spotted in a single position, in this new approach the peptides belonging to the same protein are usually spotted in consecutive positions of the MALDI plate. Hence, a correction in the acquisition and search algorithm needs to be introduced in order to take into account this possibility. TOF instruments are characterized by a high sensitivity and accuracy to measure the *m/z* ratios of peptides quickly and reliably, but when coupled to peptide fractionation using HPLC, these capabilities significantly improve.

## 9 DATABASE SEARCH AND USE OF BIOINFORMATIC TOOLS FOR PROTEIN IDENTIFICATION

### 9.1 Search Engines Against Nucleotide and Amino Acid Databases

The analysis of even a single protein tryptic digest yields a vast amount of fragmentation data after analysis with a mass spectrometer. Manual sequencing of peptides and subsequent identification of the original protein would be a major constraint. Several search engines have been developed to help in this task, including MS-Fit, MASCOT, OMSSA, PHENYX, SEQUEST, and X! Tandem, among others. Basically, search engines compare the experimental mass-to-charge ratios measured by mass spectrometers with *in silico* mass-to-charge ratios derived from the fragmentation of amino acid sequence databases. Nucleotide databases can also be used for protein identification as long as protein sequence can be derived from nucleotides.

Most proteomic studies are therefore biased toward the identification of peptides and proteins contained in databases (either as nucleic acid or as amino acid sequences). Hence, most "shotgun" protein identification experiments, aimed at the successful identification of as many proteins present in the sample as possible, are

constrained to organisms whose genomes have been sequenced, or at least partially sequenced.

Annotated databases, including functional information of proteins as well as their modifications, are rapidly improving and new nucleotide databases from several organisms are also emerging. The Human Genome Project and other genome sequencing projects are turning out in rapid succession the complete genome sequences of specific species and thus, in principle, the amino acid sequence of every protein potentially encoded [35].

### 9.2 Use of Customized Databases for Tagged Recombinant Proteins and *De Novo* Sequencing

Mass spectrometry is also useful to characterize the quality and sequence content of recombinant proteins that contain peptide tags (His-tag, GST-tag, etc.) for their purification after expression in *E. coli* or other systems. In these cases, where the exact amino acid sequence of the recombinant protein is known, customized databases can be prepared in order to identify all the proteolytic peptides of the protein of interest. The goal is to detect the possible occurrence of deletions or partial degradation from the N or C terminus or the loss of the tags.

An alternative to protein identification using search algorithms is the identification of unknown proteins, from unknown genomes, by *de novo* sequencing. There are two things that need to be considered in *de novo* sequencing experiments. First, better accuracy of the measured mass leads to better sequence information from the fragmented peptide. Second, appropriate software (i.e., PEAKS, http://www.bioinfor.com) needs to be used to gather all the information detected after fragmentation of the peptides of interest.

## 10 IDENTIFICATION OF POSTTRANSLATIONAL MODIFICATION OF PROTEINS

Many critical events involved in cellular responses (control of enzymatic activity, protein–protein interaction, cellular localization, signal transduction) are mediated by changes in posttranslational protein modifications rather than transcriptional changes [36]. More than 300 modifications have been reported on proteins and peptides both in prokaryotic and eukaryotic cells (http://www.abrf.org/index.cfm/dm.home). Currently, mass spectrometry is considered as the most appropriate approach for the identification and characterization of posttranslational modification of proteins [37]. It can unambiguously identify the occurrence of posttranslational modifications at the residue level.

Almost 2% of the human genome encodes protein kinases [38], and it is estimated that protein phosphorylation can affect up to one-third of all proteins [39], due to the relative abundance of phosphorylated residues near 1% pTyr, 12% pThr, and 87% pSer [40]. Phosphorylation analysis of proteins is made by cleavage into their constituent peptides using endopeptidase digestion, generally trypsin, separation, and analysis of the resulting peptides by LC-MS/MS. The high sensitivity and the ability to isolate and fragment peptides of ion traps and triple quadrupoles serves as a quick and reliable method for the detection of phosphorylations [41–43].

The use of nano-HPLC coupled to mass spectrometers with high ion capacity and fast scan rates enhances the detection of phosphorylated peptides and enables the determination of the phosphorylation at the residue level [44, 45].

In order to detect phosphorylation sites on protein digests, data-dependent mass analysis is frequently used to trigger $MS^3$ scans, where specific neutral losses (98.0, 49.0, 32.7 Da) are observed. This is called data-dependent neutral loss (DDNL) analysis. Nevertheless, phosphopeptide analysis with mass spectrometry has to face the inherent problem of low concentrations inside the peptide mixture [46]. Therefore, several enrichment strategies have been developed to favor their analysis using mass spectrometry. These methods include the use of immobilized metal ion ($Fe^{3+}$ or $Ga^{3+}$) affinity chromatography (IMAC), zirconium dioxide [47], or titanium dioxide [45, 48]. Phosphopeptides can also be collected in a few fractions by strong cation exchange (SCX) chromatography using a salt gradient (e.g., 1 mM to 1 M NaCl or ammonium formate). Another strategy for tyrosine kinases is the use of immunological methods based on anti-phosphotyrosine antibodies, which are sufficiently specific for the enrichment of phosphotyrosine-containing peptides by affinity or immunoprecipitation.

As an example of the identification of phosphorylation using mass spectrometry, Fig. 2 shows the identification of a chymotryptic peptide from Cdc25 protein, involved in cell cycle regulation in *Xenopus laevis*, after HPLC separation using a reverse phase column online with linear ion trap mass spectrometry. The fragmentation spectrum shows an $MS^2$ event of the triply charged precursor ion $[M+3H]^{3+}$ $m/z$ = 843.4 that corresponds to the singly phosphorylated peptide, SVSNKENEGEL-FKpSPNCKPVAL. The peptide sequence and the occurrence of the phosphorylated residue can be derived from the singly and doubly charged *y* and *b* ion series. A diagnostic prominent neutral loss as phosphoric acid (−98 Da) is also observed. The expected ion masses are listed above the sequence.

On the other hand, acetylation and methylation patterns of chromatin proteins (most times on the N terminus and/or lysine residues of histones) has become an increasingly important aspect of epigenetics and cancer biology [49, 50].

## 11 PROTEIN QUANTITATION

Until recently, the quantitative study of the pool of proteins expressed in biological samples using differential display was restricted to 2D gel analyses. Protein quantitation using the 2D gel-based approach consists of the measurement of the volume (area in pixels by intensity of the spots) of spots identified as a single protein in different gels. A significant statistical variation of the amount of protein measured in gels corresponding to drug treatments or specific biological alterations of interest with respect to a control can therefore be expressed in terms of quantitative variation. An alternative non-gel-based approach has been the use of isotope coded affinity tags (ICAT) for the quantitative study of protein expression at the proteome level [51]. Recently, an improved approach, analogous to ICAT, has been developed. It is called iTRAQ™ (Applied Biosystems) and consists of a range of up to eight different amine-specific, stable isotope tags that label all peptides in up to eight different biological samples, enabling simultaneous identification and quantitation of the whole set of proteins.

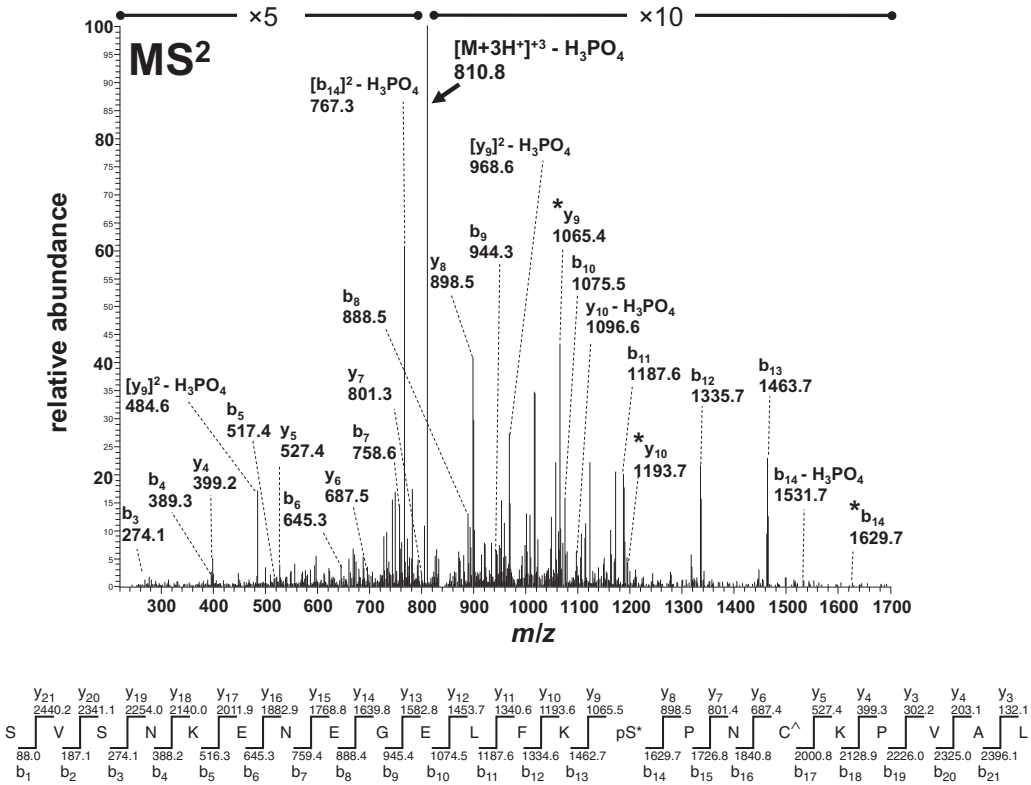


**FIGURE 2** Identification of a chymotryptic peptide by RP-HPLC online with linear ion trap mass spectrometry. The fragmentation spectrum shows an $MS^2$ event of the triply charged precursor ion $[M + 3H]^{3+}$ $m/z = 843.4$ that corresponds to the singly phosphorylated peptide SVSNKENEGELFKpS*PNC^KPVAL, where p shows the presence of a phosphorylated Ser and ^ indicates carboxyamidomethylated Cys. The peptide sequence and the occurrence of the phosphorylated residue can be derived from the singly and doubly charged *y* and *b* ion series. The loss of the a phosphoric acid molecule in the $MS^2$ event yielded a prominent neutral loss (black arrow) with *m/z* ratio 810.8. The phosphorylated serine (pS) residue is indicated and the diagnostic ions bearing phosphorylations are indicated with an asterisk (*). The peptide sequence and the assignment of the fragmentation series are also indicated. The expected ion masses have been listed above the sequence. Loss of phosphate from the ions is indicated by $H_3PO_4$.

A more recent and powerful alternative is the stable isotope labeling with amino acids in cell culture (SILAC) [52–54], which can be used mainly in cell cultures because it is based on the metabolic incorporation of labeled amino acids containing heavy isotopes (i.e., $^2H$ instead of $^1H$, $^{13}C$ instead of $^{12}C$, or $^{15}N$ instead of $^{14}N$) in living cells. After at least five cell population turnovers, complete labeling of all the proteins is achieved. The cell cultures to be compared are mixed and the changes in the expression of proteins versus the control can be quantified because the peptides containing the different isotopes coelute simultaneously from the C18 column, leading to a mass shift that is observed in the mass spectrometer. A major limitation

of the iTRAQ and SILAC approaches is the cost of the reagents necessary to perform these experiments.

A more economical approach for isotopic peptide labeling and quantification is based on the enzymatic $^{16}O/^{18}O$ labeling. This methodology is based on the trypsin catalyzed oxygen exchange at the free terminus from a peptide [55, 56]. Proteins are digested initially with normal water and they are further incubated either with $H_2\ ^{16}O$ or with $H_2\ ^{18}O$ for the differential labeling.

## 12 PROTEIN ARRAYS

Protein microarrays represent an innovative and versatile approach to study protein abundance, function, and protein–protein interaction at an unprecedented scale. The molecular characteristics of proteins make construction of protein arrays much more challenging than building DNA microarrays. The major obstacle to the development of this field is the content of the microarrays. In contrast to the PCR technique for nucleic acids that can generate large quantities of virtually any DNA, there is no counterpart for protein amplification. Protein expression or production systems are difficult to automate and always unpredictable. Many proteins are unstable, which complicates microarray shelf life. Finally, in contrast to the simple hybridization procedures for nucleic acids, proteins have shown a wide range of chemistries and specificities that complicate substantially any interaction study at the multiplex level.

Two major approaches have been followed for the production of protein microarrays: abundance-based microarrays and function-based microarrays.

### 12.1 Abundance-Based Microarrays

There are two types of abundance-based microarrays: (1) antibody or antibody-like microarrays and (2) reverse phase protein microarrays. Antibody microarrays are the most popular option given the stability, selectivity, and high affinity of the antibodies as well as the high number of antibodies commercially available. Therefore, this option has been the fastest growing in the field. In any case, some precautions should be taken into account: not all the antibodies are actually functional in the microarray format, and even monospecific antibodies can show cross-reaction on the microarray when confronted with complex protein mixtures. The most practical approach for the use of capture microarrays is the direct labeling of the analytes with one or two fluorochroms. This approach can be used for the analysis of differential expression of proteins in different conditions (e.g., for the analysis of tumor tissues versus normal tissues in cancer). This approach has recently been reported by a number of authors in prostate cancer [57] and breast cancer [58].

Reverse phase protein microarrays rely entirely on the existence of analyte-specific reagents [59]. This is particularly important as the concentration of analytes in the spots is locally high and may facilitate spurious interactions. Other authors [60] used reverse phase microarrays to classify different types of cancer using the NCI-60 cell lines. The cell lines were spotted as lysates and then probed with antibodies. More than 200 antibodies were tested to finally select 52 that gave a reliable signal. Also, some interference problems may appear from the locally high concentration of analytes. This might be alleviated by prefractionation of the samples before spotting.

### 12.2 Function-Based Protein Microarrays

Function-based microarrays consist of panels of proteins spotted at defined positions. They are used to study the biochemical properties and activities of the target proteins spotted on the microarray. Thus, protein microarrays are mainly used to examine protein interactions with other proteins or other molecules [61]. There are also other arrays used for studying the enzyme activity and substrate specificity. Proteins have to be printed in the right conditions to maintain conformation, integrity, and activity. In some cases, the integrity of the protein requires it to be printed as a complex of several proteins to remain active and functional. The major challenge about this type of microarray continues to be the production of the proteins. It is time consuming and costly to produce proteins of good purity and yield.

Generally, these arrays are made by using three different approaches: (1) chemical linkage [62], (2) peptide fusion tag [61], and (3) self-assembling protein arrays [63]. All three methods have advantages and disadvantages [64]. Alternatively, it may be simpler to study protein domains rather than the full length proteins, because small domains are usually simple to express and purify [65]. Self-assembling microarrays, also called nucleic acid programmable protein arrays (NAPPAs), are a new and interesting alternative based on the printing of cDNAs on the microarray surface and expressed *in situ* using a mammalian cell-free expression system. The nascent molecule, labeled with a tag, is captured by an adjacent antibody against the tag printed on the chip surface. This approach is very promising because it excludes the necessity of expressing, purifying, and storing the proteins. Also, it opens new avenues to prepare custom-made protein microarrays.

There is great enthusiasm about this new area of protein microarrays. Still, major efforts need to be made in different areas of production before these tools become commonly used in laboratories, as happens with DNA chips. Few articles have reported original findings of proteomic interest and most of the communications to date have been based more on technology optimization than in solving real problems. Therefore, the utility of protein microarrays remains to be fully determined.

## 13 LINKS OF INTERESTS

**Metabolic Pathways**

http://expasy.org/tools/pathways/
http://www.genome.ad.jp/kegg/kegg2.html

**Protocols Online**

http://www.protocol-online.org/
www.thermo.com/eThermo/CMA/PDFs/Articles/articlesFile_21631.pdf
www.bio-rad.com/cmc_upload/Literature/13023/4006173B-wMS.pdf
http://probes.invitrogen.com/lit/bioprobes31/section1.html
www.fluorotechnics.com/content/documents/ap_dp_comp_sypro.pdf
http://www5.amershambiosciences.com/APTRIX/upp00919.nsf/Content/Proteomics+DIGE

**Reference Books for Proteomics**

http://www.proteinsandproteomics.org/

**Handbook of 2D-PAGE**

http://www6.amershambiosciences.com/aptrix/upp01077.nsf/Content/Products?OpenDocument&parentid=575392&moduleid=164456&zone=Proteomics

**2D Gels and Gel Repository**

http://expasy.org/ch2d/
www.apczech.cz/pdf/DF_DeStreak.pdf

**HPLC and Mass Spectrometry Resources**

http://www.ionsource.com/
http://www.lcgcmag.com/LC%2FHPLC

**Nucleic Acid and Protein Databases**

http://www.ebi.ac.uk/FTP/
http://www.uniprot.o
http://www.tigr.org/

**Bioinformatic Tools**

http://expasy.org/tools/
http://ncrr.pnl.gov/software/
http://www.ionsource.com/links/programs.htm
http://www.geocities.com/SiliconValley/Hills/2679/magtran.html
http://www.cs.princeton.edu/~schapire/maxent/

**Search Engines Used in Mass Spectrometry**

MS-Fit: (http:/prospector.ucsf.edu) University of California (San Francisco, EEUU)
MASCOT: (www.matrixscience.com) Imperial Cancer Research Fund (London, UK)
OMSSA: http://pubchem.ncbi.nlm.nih.gov/omssa/
PHENYX: http://www.phenyx-ms.com/
SEQUEST: http://fields.scripps.edu/sequest/
X! Tandem: http://www.thegpm.org/TANDEM/index.html

**Posttranslational Modifications**

http://expasy.org/tools/findmod/findmod_masses.html
http://www.abrf.org/index.cfm/dm.home

**Proteomic Forums and Survey Studies**

http://www.abrf.org/index.cfm/group.show/Proteomics.34.htm
http://www.swissproteomicsociety.org/digest/

## 14 CONCLUSION

The measurement of changes in gene expression as a response to the appearance of an alteration/disease is not sufficient in many cases. It is estimated that only 50% of the measured increments of mRNA are directly correlated to an increase in the amount of the corresponding expressed protein, as observed in human [66], yeast [67, 68], bacteria [69], and other cell lines [70]. Therefore, changes detected at the protein level using proteomic techniques could be more indicative of the existence of alterations and modifications.

The classical approach to study proteins and biological processes, which was based on purification to homogeneity followed by biochemical assays of the specific activity of the purified proteins, has been replaced today by a tremendous increase in the application of proteomic technologies to the study and characterization en masse of proteins and complete proteomes. Although the analysis of full proteomes remains a formidable task, the new generation of multidimensional chromatographic steps coupled to the unprecedented resolution of the new mass spectrometers allows optimism for the future. Technological advances have translated into major improvements in the three major branches of the current proteomics: two-dimensional electrophoresis, mass spectrometry, and protein/antibody microarrays. Increased reproducibility, better statistical analysis, and improvements in mass accuracy, resolving power, and accuracy of quantifications is giving proteomics a major impulse within the life sciences.

## REFERENCES

1. Wilkins MR, Sanchez JC, Gooley AA, Appel RD, Humphery SI, Hochstrasser DF, Williams KL. Progress with proteome projects: why all proteins expressed by a genome should be identified and how to do it. *Biotechnol Genet Eng Rev* 1996;13:19–50.
2. Wu CC, MacCoss MJ, Howell KE, Yates JR. A method for the comprehensive proteomic analysis of membrane proteins. *Nat Biotechnol* 2003;21:532–538.
3. Alfonso P, Dolado I, Swat A, Núñez A, Cuadrado A, Nebreda A, Casal JI. Membrane protein alterations induced by oncogenic H-ras transformation of p38 MAPK-deficient fibroblasts. *Proteomics* 2006;6(Supp 1):S262–S271.
4. Bordier C. Phase separation of integral membrane proteins in Triton X-114 solution. *J Biol Chem* 1981;256:1604–1607.
5. States DJ, Omenn GS, Blackwell TW, Fermin D, Eng J, Speicher DW, Hanash SM. Challenges in deriving high-confidence protein identifications from data gathered by a HUPO plasma proteome collaborative study. *Nat Biotechnol* 2006;24:333–338.
6. Foster LJ, de Hoog CL, Zhang YL, Zhang Y, Xie XH, Mootha VK, Mann M. A mammalian organelle map by protein correlation profiling. *Cell* 2006;125:187–199.

7. Bensadoun A, Weinstein D. Assay of proteins in the presence of interfering materials. *Anal Biochem* 1975;70:241–250.
8. Bradford M. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal Biochem* 1976;72: 248–254.
9. Laemmli UK. Cleavage of structural proteins during assembly of head of bacteriophage-T4. *Nature* 1970;227:680–685.
10. Neuhoff V, Arold N, Taube D, Ehrhardt W. Improved staining of proteins in polyacrylamide gels including isoelectric focusing gels with clear background at nanogram sensitivity using Coomassie Brilliant Blue G-250 and R-250. *Electrophoresis* 1988;9:255–262.
11. Blum H, Beier H, Gross HJ. Improved silver staining of plant proteins, RNA and DNA in polyacrylamide gels. *Electrophoresis* 1987;8:93–99.
12. Lanne B, Panfilov O. Protein staining influences the quality of mass spectra obtained by peptide mass fingerprinting after separation on 2-D gels. A comparison of staining with Coomassie brilliant blue and SYPRO Ruby. *J Proteome Res* 2005;4:175–179.
13. Chevalier F, Rofidal V, Vanova P, Bergoin A, Rossignol M. Proteomic capacity of recent fluorescent dyes for protein staining. *Phytochemistry* 2004;65:1499–1506.
14. White IR, Pickford R, Wood J, Skehel JM, Gangadharan B, Cutler P. A statistical comparison of silver and SYPRO Ruby staining for proteomic analysis. *Electrophoresis* 2007;25:3048–3054.
15. Tonge R, Shaw J, Middleton B, Rowlinson R, Rayner S, Young J, Pognan F, Hawkins E, Currie I, Davison M. Validation and development of fluorescence two-dimensional differential gel electrophoresis proteomics technology. *Proteomics* 2001;1:377–396.
16. Alban A, David SO, Bjorkesten L, Andersson C, Sloge E, Lewis S, Currie I. A novel experimental design for comparative two-dimensional gel analysis: two-dimensional difference gel electrophoresis incorporating a pooled internal standard. *Proteomics* 2003;3:527–533.
17. Gharbi S, Gaffney P, Yang A, Zvelebil MJ, Cramer R, Waterfield MD, Timms JF. Evaluation of two-dimensional differential gel electrophoresis for proteomic expression analysis of a model breast cancer cell system. *Mol Cell Proteomics* 2002;1:91–98.
18. Friedman DB, Hill S, Keller JW, Merchant NB, Levy SE, Coffey RJ, Caprioli RM. Proteome analysis of human colon cancer by two-dimensional difference gel electrophoresis and mass spectrometry. *Proteomics* 2004;4:793–811.
19. Alfonso P, Núñez A, Madoz-Gurpide J, Lombardia L, Sánchez L, Casal JI. Proteomic expression analysis of colorectal cancer by two dimensional gel electrophoresis. *Proteomics* 2005;5:2602–2611.
20. González-Santiago L, Alfonso P, Suárez Y, Núñez A, García-Fernández LF, Alvarez E, Muñoz A, Casal JI. Proteomic analysis of the resistance to Aplidin in human cancer cells. *J Proteome Res* 2007;6:1286–1294.
21. Rauman B, Cheung A, Marten M. Quantitative comparison and evaluation of two commercially available, two-dimensional electrophoresis image analysis software packges, Z3 and Melanie. *Electrophoresis* 2002;23:2194–2202.
22. Shevchenko A, Wilm M, Vorm O, Mann M. Mass spectrometric sequencing of proteins from silver stained polyacrylamide gels. *Anal Chem* 1996;68:850–858.
23. Domon B, Aebersold R. Mass spectrometry and protein analysis. *Science* 2006;212–217.
24. Karas M, Hillenkamp F. Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. *Anal Chem* 1988;60:2299–2301.
25. Mann M, Hojrup P, Roepstorff P. Use of mass spectrometric molecular weight information to identify proteins in sequence databases. *Biol Mass Spectrom* 1993;22:338–345.

26. Chamrad DC, Korting G, Stuhler K, Meyer HE, Klose J, Bluggel M. Evaluation of algorithms for protein identification from sequence databases using mass spectrometry data. *Proteomics* 2004;4:619–668.
27. Baldwin MA. Protein identification by mass spectrometry: issues to be considered. *Mol Cell Proteomics* 2004;3:1–9.
28. Fenn JB, Mann M, Meng CK, et al. Electrospray ionisation for mass spectrometry of large biomolecules. *Science* 1989;246:64–71.
29. Han XM, Jin M, Breuker K, McLafferty FW. Extending top–down mass spectrometry to proteins with masses greater than 200 kilodaltons. *Science* 2006;314:109–112.
30. Hayter JR, Robertson DHL, Gaskell SJ, Beynon RJ. Proteome analysis of intact proteins in complex mixtures. *Mol Cell Proteomics* 2003;2:85–95.
31. Biemann K. Contributions of mass spectrometry to peptide and protein structure. *Biomed Environ Mass Spectrom* 1988;16:99–111.
32. Syka JEP, Coon JJ, Schroeder MJ, Shabanowitz J, Hunt DF. Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. *Proc Natl Acad Sci USA* 2004;101:9528–9533.
33. Coon JJ, Ueberheide B, Syka JEP, Dryhurst DD, Ausio J, Shabanowitz J, Hunt DF. Protein identification using sequential ion/ion reactions and tandem mass spectrometry. *Proc Natl Acad Sci USA* 2005;102:9463–9468.
34. Washburn MP, Wolters D, Yates JR. Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat Biotechnol* 2001;19:242–247.
35. Peri S, Navarro JD, Amanchy R, Kristiansen TZ, Jonnalagadda CK, Surendranath V, Niranjan V, Muthusamy B, Gandhi TKB, Gronborg M, Ibarrola N, Deshpande N, Shanker K, Shivashankar HN, Rashmi BP, Ramya MA, Zhao ZX, Chandrika KN, Padma N, Harsha HC, Yatish AJ, Kavitha MP, Menezes M, Choudhury DR, Suresh S, Ghosh N, Saravana R, Chandran S, Krishna S, Joy M, Anand SK, Madavan V, Joseph A, Wong GW, Schiemann WP, Constantinescu SN, Huang LL, Khosravi-Far R, Steen H, Tewari M, Ghaffari S, Blobe GC, Dang CV, Garcia JGN, Pevsner J, Jensen ON, Roepstorff P, Deshpande KS, Chinnaiyan AM, Hamosh A, Chakravarti A, Pandey A. Development of human protein reference database as an initial platform for approaching systems biology in humans. *Genome Res* 2003;13:2363–2371.
36. Olsen JV, Blagoev B, Gnad F, Macek B, Kumar C, Mortensen P, Mann M. Global, *in vivo*, and site-specific phosphorylation dynamics in signaling networks. *Cell* 2006;127: 635–648.
37. Mann M, Jensen ON. Proteomic analysis of post-translational modifications. *Nat Biotechnol* 2003;21:255–261.
38. Manning G, Whyte DB, Martinez R, Hunter T, Sudarsanam S. The protein kinase complement of the human genome. *Science* 2002;298:1912–1934.
39. Cohen P. The role of protein phosphorylation in human health and disease. *Eur J Biochem* 2001;268:5001–5010.
40. Hunter T, Sefton BM. Transforming gene product of Rous sarcoma virus phosphorylates tyrosine. *Proc Natl Acad Sci USA Biol Sci* 1980;77:1311–1315.
41. Mann M, Hendrickson RC, Pandey A. Analysis of proteins and proteomes by mass spectrometry. *Annu Rev Biochem* 2001;70:437–473.
42. Meng FY, Forbes AJ, Miller LM, Kelleher NL. Detection and localization of protein modifications by high resolution tandem mass spectrometry. *Mass Spectrom Rev* 2005;24:126–134.
43. Villar M, Ortega-Perez I, Were F, Cano E, Redondo JM, Vazquez J. Systematic characterization of phosphorylation sites in NFATc2 by linear ion trap mass spectrometry. *Proteomics* 2006;6(Suppl 1):S16–S27.

44. Beausoleil SA, Jedrychowski M, Schwartz D, Elias JE, Villen J, Li JX, Cohn MA, Cantley LC, Gygi SP. Large-scale characterization of HeLa cell nuclear phosphoproteins. *Proc Natl Acad Sci USA* 2004;101:12130–12135.
45. Larsen MR, Thingholm TE, Jensen ON, Roepstorff P, Jorgensen TJD. Highly selective enrichment of phosphorylated peptides from peptide mixtures using titanium dioxide microcolumns. *Mol Cell Proteomics* 2005;4:873–886.
46. Steen H, Jebanathirajah JA, Rush J, Morrice N, Kirschner MW. Phosphorylation analysis by mass spectrometry—myths, facts, and the consequences for qualitative and quantitative measurements. *Mol Cell Proteomics* 2006;5:172–181.
47. Kweon HK, Hakansson K. Selective zirconium dioxide-based enrichment of phosphorylated peptides for mass spectrometric analysis. *Anal Chem* 2006;78:1743–1749.
48. Pinkse MWH, Uitto PM, Hilhorst MJ, Ooms B, Heck AJR. Selective isolation at the femtomole level of phosphopeptides from proteolytic digests using 2D-nanoLC-ESI-MS/MS and titanium oxide precolumns. *Anal Chem* 2004;76:3935–3943.
49. Fraga MF, Ballestar E, Villar-Garea A, Boix-Chornet M, Espada J, Schotta G, Bonaldi T, Haydon C, Ropero S, Petrie K, Iyer NG, Perez-Rosado A, Calvo E, Lopez JA, Cano A, Calasanz MJ, Colomer D, Piris MA, Ahn N, Imhof A, Caldas C, Jenuwein T, Esteller M. Loss of acetylation at Lys16 and trimethylation at Lys20 of histone H4 is a common hallmark of human cancer. *Nat Genet* 2005;37:391–400.
50. Boix-Chornet M, Fraga MF, Villar-Garea A, Caballero R, Espada J, Nunez A, Casado J, Largo C, Casal JI, Cigudosa JC, Franco L, Esteller M, Ballestar E. Release of hypoacetylated and trimethylated histone H4 is an epigenetic marker of early apoptosis. *J Biol Chem* 2006;281:13540–13547.
51. Shiio Y, Aebersold R. Quantitative proteome analysis using isotope-coded affinity tags and mass spectrometry. *Nat Protocols* 2006;1:139–145.
52. Ong SE, Blagoev B, Kratchmarova I, Kristensen DB, Steen H, Pandey A, Mann M. Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol Cell Proteomics* 2002;1:376–386.
53. Gruhler A, Olsen JV, Mohammed S, Mortensen P, Faergeman NJ, Mann M, Jensen ON. Quantitative phosphoproteomics applied to the yeast pheromone signaling pathway. *Mol Cell Proteomics* 2005;4;310–327.
54. Mann M. Functional and quantitative proteomics using SILAC. *Nat Rev Mol Cell Biol* 2006;7:952–958.
55. Yao X, Freas A, Ramirez J, Demirev PA, Fenselau C. Proteolytic $^{18}O$ labeling for comparative proteomics: model studies with two serotypes of adenovirus. *Anal Chem* 2001;73:2836–2842.
56. Yao X, Afonso C, Fenselau C. Dissection of proteolytic $^{18}O$ labeling: endoprotease-catalyzed $^{16}O$-to-$^{18}O$ exchange of truncated peptide substrates. *J Proteome Res* 2003;2:147–152.
57. Miller JC, Zhou H, Kwekel J, Cavallo R, Burke J, Butler EB, Teh BS, Haab BB. Antibody microarray profiling of human prostate cancer sera: antibody screening and identification of potential biomarkers. *Proteomics* 2003;3:56–63.
58. Celis JE, Moreira JM, Cabezon T, Gromov P, Friis E, Rank F, Gromova I. Identification of extracellular and intracellular signaling components of the mammary adipose tissue and its interstitial fluid in high risk breast cancer patients: toward dissecting the molecular circuitry of epithelial–adipocyte stromal cell interactions. *Mol Cell Proteomics* 2005;4: 492–522.
59. Speer R, Wulfkuhle JD, Liotta LA, Petricoin EF3. Reverse-phase protein microarrays for tissue-based analysis. *Curr Opin Mol Ther* 2005;7:240–245.

60. Nishizuka S, Charboneau L, Young L, Major S, Reinhold WC, Waltham M, Kouros-Mehr H, Bussey KJ, Lee JK, Espina V, Munson PJ, Petricoin E3, Liotta LA, Weinstein JN. Proteomic profiling of the NCI-60 cancer cell lines using new high-density reverse-phase lysate microarrays. *Proc Natl Acad Sci USA* 2003;100:14229–14234.
61. Zhu H, Bilgin M, Bangham R, Hall D, Casamayor A, Bertone P, Lan N, Jansen R, Bidlingmaier S, Houfek T, Mitchell T, Miller P, Dean RA, Gerstein M, Snyder M. Global analysis of protein activities using proteome chips. *Science* 2001;293:2101–2105.
62. MacBeath G, Schreiber SL. Printing proteins as microarrays for high-throughput function determination. *Science* 2000;289:1760–1763.
63. Ramachandran N, Hainsworth E, Bhullar B, Eisenstein S, Rosen B, Lau AY, Walter JC, LaBaer J. Self-assembling protein microarrays. *Science* 2004;305:86–90.
64. LaBaer J, Ramachandran N. Protein microarrays as tools for functional proteomics. *Curr Opin Chem Biol* 2005;9:14–19.
65. Espejo A, Cote J, Bednarek A, Richard S, Bedford MT. A protein-domain microarray identifies novel protein–protein interactions. *Biochem J* 2002;367:697–702.
66. Varambally S, Yu JJ, Laxman B, Rhodes DR, Mehra R, Tomlins SA, Shah RB, Chandran U, Monzon FA, Becich MJ, Wei JT, Pienta KJ, Ghosh D, Rubin MA, Chinnaiyan AM. Integrative genomic and proteomic analysis of prostate cancer reveals signatures of metastatic progression. *Cancer Cell* 2005;8:393–406.
67. Griffin TJ, Gygi SP, Ideker T, Rist B, Eng J, Hood L, Aebersold R. Complementary profiling of gene expression at the transcriptome and proteome levels in *Saccharomyces cerevisiae*. *Mol Cell Proteomics* 2002;1:323–333.
68. Washburn MP, Koller A, Oshiro G, Ulaszek RR, Plouffe D, Deciu C, Winzeler E, Yates JR. Protein pathway and complex clustering of correlated mRNA and protein expression analyses in *Saccharomyces cerevisiae*. *Proc Natl Acad Sci USA* 2003;100:3107–3112.
69. Baliga NS, Pan M, Goo YA, Yi EC, Goodlett DR, Dimitrov K, Shannon P, Aebersold R, Ng WV, Hood L. Coordinate regulation of energy transduction modules in *Halobacterium* sp analyzed by a global systems approach. *Proc Natl Acad Sci USA* 2002;99: 14913–14918.
70. Tian Q, Stepaniants SB, Mao M, Weng L, Feetham MC, Doyle MJ, Yi EC, Dai HY, Thorsson V, Eng J, Goodlett D, Berger JP, Gunter B, Linseley PS, Stoughton RB, Aebersold R, Collins SJ, Hanlon WA, Hood LE. Integrated genomic and proteomic analyses of gene expression in mammalian cells. *Mol Cell Proteomics* 2004;3:960–969.

# TOXICOGENOMICS IN PRECLINICAL DEVELOPMENT

Eric A. G. Blomme, Dimitri Semizarov, and Jeffrey F. Waring
*Global Pharmaceutical Research and Development, Abbott Laboratories, Abbott Park, Illinois*

**Contents**

1 Introduction
- 1.1 Toxicogenomics and Other Emerging Technologies in Perspective
- 1.2 Definitions and Basics of Toxicogenomics
- 1.3 History of Toxicogenomics
- 1.4 Applications of Toxicogenomics

2 Practical and Logistic Aspects of Toxicogenomics
- 2.1 Technical Considerations
- 2.2 Species Considerations
- 2.3 Toxicogenomics in *In Vitro* and *In Vivo* Studies

3 Toxicogenomic Reference Databases
- 3.1 The Need for Toxicogenomic Reference Databases
- 3.2 Design and Development of Toxicogenomic Reference Databases
- 3.3 Existing Toxicogenomic Databases

4 Toxicogenomics in Drug Discovery
- 4.1 Predictive Toxicology
- 4.2 Development of Predictive Gene Expression Signatures
- 4.3 Case Examples
- 4.4 Predicting Species-Specific Toxicity

5 *In Vitro* Toxicogenomics
- 5.1 Objectives of *In Vitro* Toxicogenomics
- 5.2 Proof-of-Concept Using Primary Rat Hepatocytes
- 5.3 Use of Gene Expression Profiling to Assess Genotoxicity
- 5.4 Current and Future Use of *In Vitro* Toxicogenomics

6 Toxicogenomics in Mechanistic Toxicology
- 6.1 Objectives of Mechanistic Toxicology
- 6.2 Case Examples of Mechanistic Toxicology

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

7 Toxicogenomics and Target-Related Toxicity
8 Toxicogenomics and Idiosyncratic Toxicity
8.1 Definition of Idiosyncratic Toxicity
8.2 Preclinical Models of Idiosyncratic Toxicity
8.3 Case Example: Idiosyncratic Hepatotoxicity
9 Toxicogenomics in Regulatory Submissions
9.1 Overview of the FDA Pharmacogenomics Guidance
9.2 Future Impact of Toxicogenomic Data in Regulatory Decision Making
10 Conclusion
References

## 1 INTRODUCTION

### 1.1 Toxicogenomics and Other Emerging Technologies in Perspective

The cost of drug discovery and development has risen exponentially in the last decades. According to the Pharmaceutical Research and Manufacturers of America (PhRMA), the industry's trade group, pharmaceutical companies spent $33 billion on R&D in 2003, a threefold rise since 1990 and nearly 30-fold since 1977 [1]. This increase in investment has so far failed to deliver a surge of new medicines, and this is reflected by a concerning low productivity of pharmaceutical R&D [1, 2]. The cost of developing a new chemical entity (NCE) ranges from $800 million (U.S.) to $1.1 billion [1, 3]. These rising R&D costs are not sustainable and the lack of productivity of pharmaceutical R&D units has to be addressed. There are several factors underlying this change in the drug development economics. First, pharmaceutical companies are now tackling diseases of greater complexity than in the past and the industry's interest in the development of blockbusters requires running longer and more expensive clinical trials [1]. Second, the requirements for approval are notably higher than in the past because of the enhanced standard of care for most diseases and more stringent regulations intended to improve drug quality and safety [4].

One striking aspect of the drug discovery and development process is the high failure rate of compounds with an estimated 99.9% of compounds eliminated from the discovery and development pipeline [1]. Obviously, the vast majority of these compounds are eliminated very early in the process because of suboptimal pharmacological, physicochemical, pharmacokinetic, or toxicologic properties. Nevertheless, failure rates in the subsequent, more costly stages of development are substantially high with the vast majority of clinical attrition occurring in Phases IIb and III [4]. Eliminating unsuccessful drugs earlier than in full development is definitely a prerequisite for a decrease in the overall R&D costs. Indeed, the recently implemented laboratory technologies (such as combinatorial chemistry, genomics tools, or high throughput screening) driving discovery efforts are resulting in a constantly increasing number of novel compounds being synthesized and a similarly rising number of therapeutically interesting targets. Therefore, approaches that allow for an earlier, multidirectional characterization of compounds are needed to face these major challenges.

The major causes of attrition in the clinic in 2000 were lack of efficacy and safety, both accounting for approximately 30% of failures, respectively [4]. This is in contrast to what was occurring in the late 1980s when poor pharmacokinetic properties were the main reason for termination (around 40%), while lack of clinical safety already accounted for 30% of failures [5]. This remarkable improvement in the pharmacokinetic properties of advanced compounds was mostly the result of a significant effort by the pharmaceutical industry to develop preclinical tools to better predict the pharmacokinetic properties of experimental compounds. While the industry has successfully addressed the failures related to pharmacokinetics, it has not significantly improved its ability to better characterize early the toxicologic properties of compounds. In fact, one may argue that with an increased number of compounds to be evaluated, less toxicologic characterization has been possible, and often compounds are selected for animal testing without sufficient data regarding their toxicologic potential. Furthermore, traditional toxicologic evaluation through *in vivo* studies typically creates a bottleneck in the R&D process because of its length and cost, and due to the requirement for significant amounts of compound. Approaches designed to characterize the toxicologic profile of compounds earlier would allow discovery scientists to select the molecules with an optimal, or at least adequate, toxicological profile for these costly studies. To be cost effective and applicable in the drug discovery setting, these approaches must use small amounts of compounds (typically an amount that would not require scale-up chemistry), have acceptable accuracy (the level of acceptability being dependent on the stage of testing) and reproducibility, and have an appropriate throughput [6]. Various technologies, including the "omics" technologies, potentially meet these criteria and are addressed in this textbook in various chapters.

## 1.2 Definition and Basics of Toxicogenomics

In this chapter, we use the term *toxicogenomics* to refer to the use of gene expression analysis in the field of toxicology. The sequencing of several whole genomes has led to the development of methodologies that make it feasible to monitor in several animal species (humans, mice, rats) the expression levels of large numbers of genes expressed in a specific tissue at a certain time, an activity referred to as *gene expression profiling* or simply expression profiling. Among these laboratory tools reviewed in Chapter 22, molecular toxicologists have mostly used *DNA microarrays* to generate transcription profiles from tissues collected from *in vivo* studies or cells derived from *in vitro* experiments. Consequently, this chapter heavily emphasizes the use of microarrays in toxicology studies. It is noteworthy, however, that other technologies are available or may become available in the future for gene expression profiling. In particular, for specific applications, it is generally agreed that more cost-effective platforms with a higher throughput will be needed for toxicogenomics to realize its full potential in drug discovery and development [7].

Toxicogenomics is based on the relatively simple assumption that toxicants acting through a similar mechanism of action will generate similar gene expression profiles or at least affect similar pathways, leading to common gene expression changes (Fig. 1). In other words, these gene expression changes (either upregulation or downregulation) induced in common by toxicants with similar toxicologic properties could represent an easy and sensitive endpoint to identify and classify toxicants.

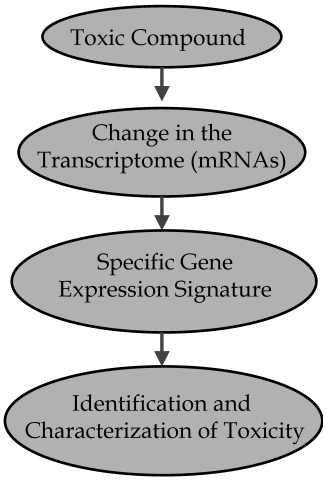


**FIGURE 1** Principle of toxicogenomics. Toxic compounds induce changes in the cellular transcriptome, including changes in the expression of particular gene sets that correlate with the mechanism of toxicity. These gene sets are typically referred to as signatures. These signatures can be used for the identification and characterization at the molecular level of toxic changes.

These common gene expression changes are typically referred to as *fingerprints* or *signatures* of toxicity. Proof-of-principle for the signature concept has already been nicely demonstrated in various types of cancer, as mentioned in Chapter 22. Using gene expression analysis, investigators have identified new classes of hematological malignancies or predicted prognosis in lung cancer and breast cancer [8]. Furthermore, because microarrays allow for the global evaluation of the cell transcriptome, the assumption is also that the identified gene expression changes will allow toxicologists to understand better the molecular mechanisms whereby toxicants injure cells. Again, this is clearly supported by the numerous mechanistic observations made in various diseases using microarrays [8]. These observations are typically a stepping stone for the articulation of new hypotheses on the mechanism of action, which can then be evaluated and confirmed by subsequent, appropriately designed experiments.

The use of most of the early microarray platforms was associated with reproducibility and accuracy issues. Since then, the technology has rapidly improved and several recent studies have demonstrated the reproducibility and accuracy of gene expression data [9–11]. This rapid improvement has led to a change of practices. For instance, it was initially recommended to confirm specific gene expression changes observed with microarrays with techniques that were considered more accurate, such as real-time reverse transcription-PCR (RT-PCR) [12]. In our laboratory, we do use RT-PCR to validate microarray data, but this confirmation step is now only occasional. An example demonstrating the good correlation between a microarray platform and RT-PCR is illustrated in Fig. 2. For gene expression profiling to reliably identify and characterize toxicity, gene expression data must be sufficiently reproducible following exposure to chemicals. The consistency and reproducibility were evaluated by different industrial, governmental, and academic

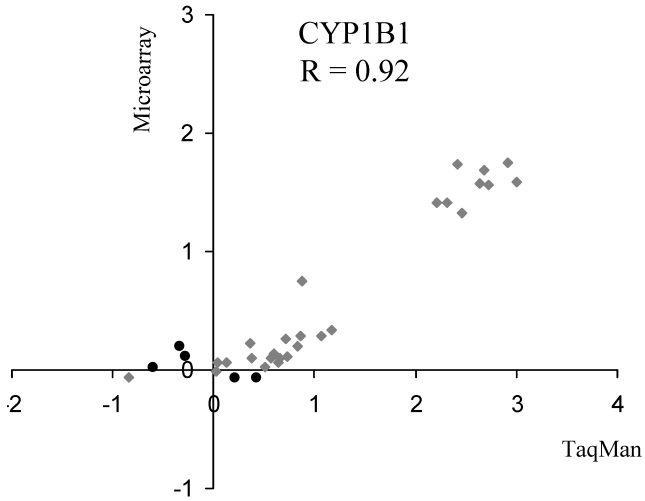


**FIGURE 2** Correlation between expression levels of cytochrome P450, 1b1 (CYP1B1) measured with Affymetrix microarrays and RT-PCR (TaqMan). The log10 values of the fold changes determined by Affymetrix microarrays (*Y* axis) are plotted against those determined by RT-PCR (*X* axis). Included are the gene expression levels from 32 hepatocyte treatments measured by both platforms. The values represented by squares indicate genes that exhibit the same directionality in terms of gene expression changes using both technologies. Circles indicate genes that exhibit different directionality between the two platforms.

laboratories using various commercial and customized microarray platforms and shown to be sufficient for assessing toxic reactions in tissues such as liver, kidney, and heart [9, 13, 14].

Any experimental *in vivo* study is associated with some degree of interindividual variability. This is the reason behind the use of treatment groups of appropriate size in toxicology studies. Developing an understanding of the interindividual variability in gene expression data was critical to fully understanding the optimal number of animals per group to be used in toxicogenomic studies. Clearly, gene expression data are also subject to interindividual variability inherent to any *in vivo* model. In our experience, gene expression data generated following short *in vivo* exposures (typically less than 24 hours) are typically quite variable. However, gene expression profiles from tissues exposed for longer periods of time (1–5 days) are usually less variable than other endpoints typically used in toxicological assessment, such as serum chemistry, hematology, or histopathology. Figure 3 illustrates this concept using liver as an example. This low variability has clear implications and advantages, as it allows the toxicologist to use as few as 2–3 rats/group to generate interpretable and reliable data.

The question of species extrapolation has been evaluated and studied by toxicologists for decades. In most instances, especially in the case of pharmaceuticals, the objective of the toxicologist is to identify hazards and assess their risks for humans. A particular interest with toxicogenomics is to understand if this technology can lead to an improved prediction of toxicologic reactions in humans, and in particular whether gene expression profiling allows toxicologists to reliably deter-

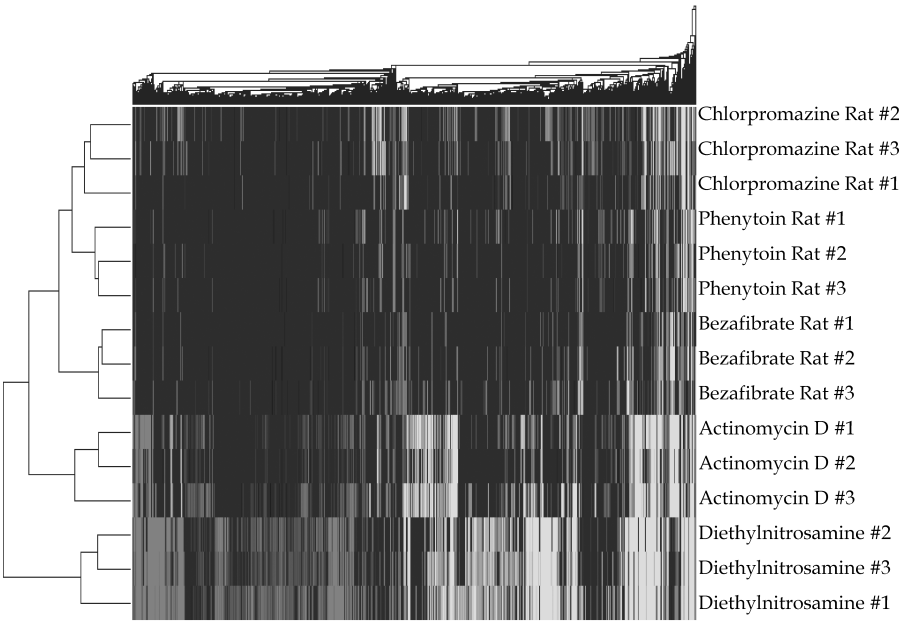


**FIGURE 3** Heatmap with hierarchical clustering illustrating gene expression changes in the liver of rats treated for 3 days with various hepatotoxicants at toxic doses. Genes shown include genes that were up- or downregulated by at least two fold with a $p < 0.01$. Despite significant variability in response observed with clinical pathology and histopathology, there is limited interindividual variability in gene expression profiles.

mine which of the changes occurring in traditional toxicologic preclinical species would not be relevant to humans. Several studies have shown that, in general, toxicogenomics improves the robustness of cross-species extrapolation [15]. Indeed, despite the differences in genomes, many responses to toxicants are evolutionary conserved. Therefore, since analysis of gene expression changes following the exposure to a toxicant improves the understanding of the mechanism of toxicity and of the major cellular subsystems affected, it becomes easier to assess the relevance of a specific toxic change or to predict how a specific species would react to a compound. However, it is important to point out that most of these studies have focused on a specific class of toxicants or a particular mechanism of toxicity. Furthermore, *in vitro* studies have mostly been used to evaluate how humans may react to a specific compound and these studies make it difficult to fully comprehend a cause-and-effect relationship. Consequently, it is prudent to state that while a toxicogenomic analysis in the context of a particular toxicologic study can and will ultimately improve the overall risk assessment to humans, it is likely that exceptions will remain.

### 1.3 History of Toxicogenomics

Soon after the microarray technology was first invented in 1995, toxicologists and molecular biologists rapidly realized its potential to generate large amounts of valu-

able molecular data, which could help in understanding the pathogenesis of various toxic changes [16, 17]. This technology has rapidly been embraced by the pharmaceutical industry as a potentially useful tool to identify safer drugs in a faster, more cost-effective manner [18]. Almost all major pharmaceutical companies now have dedicated groups applying gene expression analysis in toxicology. Academic and governmental institutions have also aggressively adopted this technology, and the regulatory community has clearly identified toxicogenomics as an important part for the success of its critical path initiative [19]. Growing interest in this field is best exemplified by the vast numbers of workshops, committees, and consortia created to address various technical issues, to improve the science base of the community, to allow contributors to keep pace with these new emerging technologies, and to establish guidelines for determining how genomic data should be submitted to regulatory agencies. An exponentially growing number of studies are being published on this topic. Some studies have used microarrays to identify the mechanism of toxicity of pharmaceutical agents or standard toxicants [20–22]. Other studies have evaluated the predictive power of toxicogenomics by identifying potential toxic liabilities before the development of other manifestations of toxicity, such as clinical chemistry or histopathology [23–25]. Finally, several investigators have used this technology in an attempt to identify new markers of toxicity [26]. The majority of published studies have evaluated reference and tool compounds that induce well-characterized, general, or specific toxicities. These compounds either have been on the market for many years or have never been developed as pharmaceutical agents. However, in the last few years, several pharmaceutical companies have clearly moved beyond this proof-of-concept stage and have applied this technology to drug development programs to address critical, development-limiting toxicologic issues, as indicated by recent submissions of gene expression profiling datasets to regulatory agencies as well as by various published studies.

### 1.4 Applications of Toxicogenomics

It would clearly be beyond the scope of this chapter to provide an in-depth review of all published studies using toxicogenomics. Rather, in this chapter, we focus on potential applications of toxicogenomics using selected examples as illustrations. Regrettably, numerous very valuable published studies have not been referenced here because of space limitations, and therefore, we encourage the reader interested in a specific application to proceed to a more elaborate literature search.

In the drug discovery and development process, toxicogenomics has been or can be applied at different stages to address different issues (Fig. 4). Its applications include the following:

- Prediction and characterization of the toxic properties of experimental compounds from short-term *in vivo* studies or *in vitro* systems. This activity typically takes place in the discovery setting at the lead identification, lead optimization, and candidate selection stages.
- Elucidation of the mechanism of toxic changes. These types of studies are more often conducted on candidate compounds that do induce unexpected changes of unknown mechanism in repeat-dose toxicology studies. The objectives are to develop an understanding of the relevance of these changes for humans or

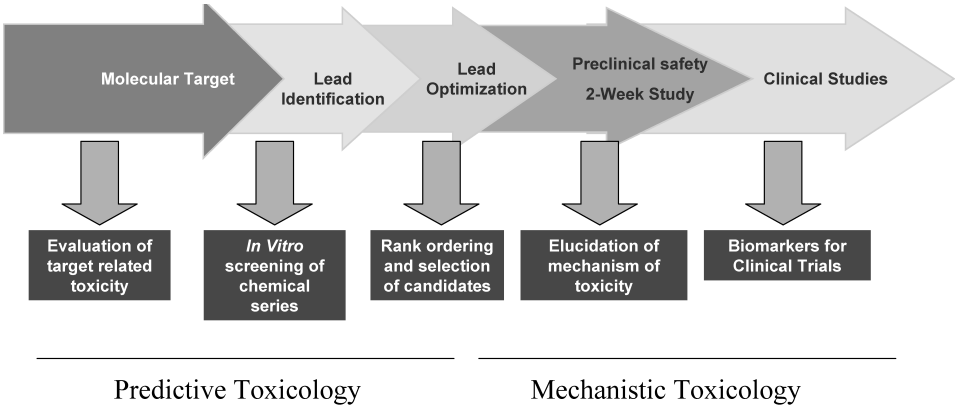


**FIGURE 4** Use of toxicogenomics in drug discovery and development. In drug discovery and development, toxicogenomics can be applied at different stages. An understanding of potential target-related toxicity with gene expression profiling can help to prioritize for the most drugable targets. Toxicogenomic studies using short-term *in vivo* studies or *in vitro* assays (predictive toxicology) are useful for the early characterization of the toxic properties of compounds at the lead identification/lead optimization stages. Toxicogenomics can also be used to elucidate the mechanism of toxicity associated with compounds. These mechanistic studies typically take place after candidate selection but are also useful to establish appropriate counterscreens for backup compound selection. Finally, gene expression profiling studies can help identify novel biomarkers for the identification and monitoring of toxicologic changes, in both preclinical and clinical studies.

to generate mechanistic data for the establishment of appropriately designed counterscreens for selecting backup compounds.

- Characterization of new therapeutic targets and the proactive identification of potential toxicity issues associated with their modulation (also referred to as toxicity related to primary pharmacology or on-target toxicity). The purpose is to investigate further the biology of the target as early as the target selection stage. Given the rising number of targets that pharmaceutical companies are working with, learning what makes a good or bad target at the gene expression level with respect to toxicology can lead to better target prioritization in the discovery pipeline.
- Identification of selective gene expression signatures that could be used as sensitive biomarkers. There is a definite need for additional biomarkers that could help in the identification and monitoring of specific toxicologic changes, in both preclinical and clinical studies. In the last few years, very few additional sensitive and specific correlates of particular toxic changes have been made available to the toxicology community. Whole-genome analysis represents an ideal approach for the identification of genes or signaling pathways whose deregulation leads to specific toxic effects.
- Molecular characterization of compounds associated with idiosyncratic reactions in the clinic. Idiosyncratic toxic events are low-incidence toxic events of unknown mechanism that occur in humans, usually in a non-dose-dependent manner and that were not observed in preclinical species. This type of toxicity

leads to costly failures of drugs in clinical trials or market withdrawal. Approaches that would allow for understanding the mechanism of idiosyncratic adverse events could have an enormous impact on the productivity of the pharmaceutical industry. They could rescue useful drugs through identification of populations sensitive to the toxic effects or through early prediction of idiosyncratic adverse events.

## 2 PRACTICAL AND LOGISTIC ASPECTS OF TOXICOGENOMICS

### 2.1 Technical Considerations

At the time when this chapter is written, gene expression profiling is still a rapidly evolving discipline for which efforts have so far essentially focused on addressing key technical issues and on validating approaches through proof-of-concept studies. Because of the immaturity of the microarray technology, there are many different technical issues, including external or internal control selection, probe design, platform variation and comparison, scanner performance characteristics, data normalization, or analysis software packages. These technical issues are being aggressively addressed by various academic, governmental, and industry groups, often in the form of consortia or collaborations [9, 10, 14, 27, 28]. While technical issues are critical to the reliability and correct interpretation of microarray data, most of them are closely related to the instrumentation being used and are consequently rapidly evolving and becoming obsolete. For this reason, in this chapter, we only address technical issues that will have some relevance for a significant period of time and that need to be understood by molecular toxicologists. For other issues, the reader is referred to Chapter 22 or other existing publications [10,12, 27–30].

### 2.2 Species Considerations

The vast majority of toxicogenomic studies are conducted using rat tissues or cell lines of rat or human origin. The major reason is that the rat is the most commonly used small laboratory animal species for toxicology testing in the pharmaceutical industry. Another reason is the incomplete genome annotation for other species (dog or monkey) and the lack of historical gene expression data for these species. Nevertheless, gene expression studies have been conducted in large animal species (dog and monkey) and several investigators have experimental microarray platforms available for these large species [31, 32].

### 2.3 Toxicogenomics in *In Vitro* and *In Vivo* Studies

***Tissue Considerations*** Toxicogenomic studies have been performed using cell cultures and tissues from *in vivo* studies. Cell cultures are very homogeneous, and gene expression changes induced by toxicants can be reproduced reliably in a laboratory under similar experimental conditions. The use of tissues can be more challenging. Several tissues, such as liver or heart, are relatively homogeneous in their phenotypes and transcriptomes, and robust, consistent changes in gene expression can reliably be detected, as long as a consistent tissue collection protocol is followed.

Other tissues, such as brain or testis, are more heterogeneous and complex, and their gene expression profiles are therefore less consistent. For instance, the brain is composed of various cell types (neurons, glia) with immense phenotypic and transcriptional diversity [33, 34]. In addition, depending on the region of the brain, marked differences in functions and transcriptomes are present between cells of the same origin. This complexity explains the limited number of attempts to apply toxicogenomics to these tissues. Technologies, such as laser capture microdissection (LCM) and RNA amplification protocols, have greatly enhanced the ability to perform expression analysis on single cell populations [35]. However, these technologies also require a more extensive investment of time and labor with an overall reduced capacity. In addition, the analysis of gene expression changes in single cell populations may limit one's ability to fully understand major interactions between cells that may play a significant role in the pathogenesis of a toxicologic change. These limitations will be illustrated in our section covering testicular toxicity (Section 6.2). Finally, it should be reemphasized that an accurate interpretation of gene expression changes can only succeed if tissue collection protocols are sufficiently consistent. Even in the case of rather homogeneous tissues like heart or kidney, it is critical to collect samples for gene expression analysis in a consistent manner [36]. Likewise, relatively homogeneous tissues like kidney contain several compartments that are clearly different in structure, function, and transcriptome. Toxicants may induce changes in only specific compartments or cellular subpopulation. Failure to cover all compartments of a tissue would limit one's ability to detect toxicant-induced gene expression changes. In our laboratory, we have established collection protocols for all tissues that we routinely evaluate. For example, we collect kidneys in a manner such that appropriate and consistent proportions of cortex and medulla are included for RNA extraction. Inappropriate collection procedures would lead to the identification of gene expression changes related to the collection procedure rather than to the toxicity being evaluated.

Samples for gene expression analysis should be collected immediately after sacrifice and flash frozen in liquid nitrogen or preserved in an appropriate RNA stabilization solution. They can then be stored for a prolonged period of time at –80 °C without significant RNA degradation. Inappropriate collection procedures or storage conditions will result in RNA degradation, as typically revealed by an overall poor RNA quality after RNA extraction procedures (Fig. 5). It is therefore recommended to always evaluate RNA quality after extraction to ensure that the quality is sufficient to justify the costly step of hybridization to microarrays, but also to ensure that interpretation of the microarray data is feasible.

***Hybridization Design in Toxicogenomic Studies*** Three common hybridization designs are used in experiments using two-color or two-channel microarrays. These designs do not pertain to experiments using one-color microarrays. These designs are referred to as *direct, reference*, and *loop* [37, 38]. In the *direct design*, the test article-treated samples are hybridized against their appropriate control samples. This allows for the identification of differentially expressed genes at a specific time. In the *reference design*, which is the most commonly used design within the biological community, all study samples are hybridized against a common reference sample, and this can be useful in the case of a study containing one control group for several treatment groups. This design is also well suited for the characterization of the

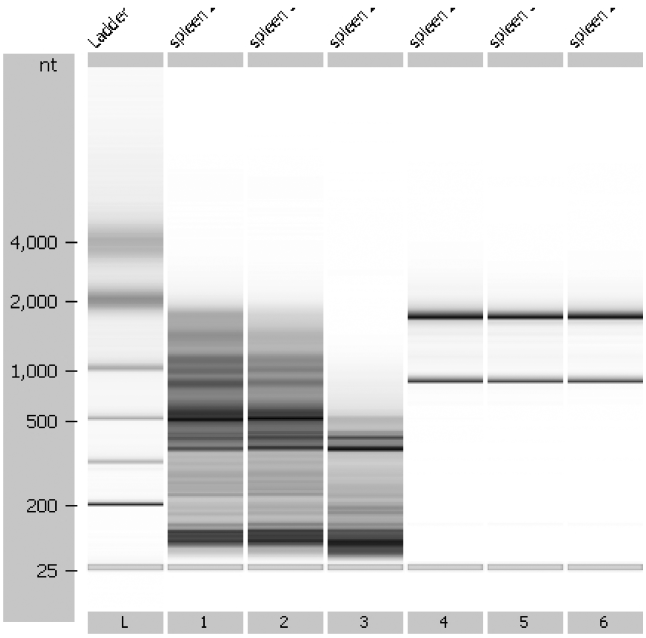


**FIGURE 5** Evaluation of RNA quality after extraction using an Agilent Bioanalyzer. Rat spleens were collected after sacrifice and flash frozen. For samples 1–3, the collection procedure was inappropriate, resulting in RNA degradation as evidenced by the multiple bands and smear observed on the Bioanalyzer-generated electrophoretic image. These samples should not be used for hybridization to microarrays. Note the sharp contrast with samples 4 and 5, where RNA quality is optimal for microarray experiments.

magnitude of gene expression changes and temporal relationships. The *loop design* is seldom used, although it has been shown to attain a higher precision [30, 38]. It entails sequential hybridization of all study samples against one another; it may offer advantages for time course experiments where one needs to understand expression changes over time.

***Experimental Design in Toxicogenomic Studies*** The design of *in vivo* and *in vitro* toxicogenomic studies is determined by the questions to be answered or the issues to be addressed. Specific design will be discussed as different applications are reviewed. However, several considerations are relevant to the current discussion.

*Duration of Dosing* In the case of a study to address the mode of action of a specific toxicologic change, a time-course experiment may be very appropriate to identify gene expression changes linked to the development of the toxicity. Time points are selected based on the particular toxicologic change to be investigated. An important aspect to remember is that gene expression changes are transient and therefore timing is of critical importance. Changes that are more relevant to generate a mechanistic understanding of toxicologic effects are typically those occurring before the effects are fully developed. Therefore, it is usually more insightful to evaluate gene expression during the development of a tissue change, rather than when the change

has already occurred and is fully established. In the case of a study used to rank-order several potential lead molecules for candidate selection, the study needs to be designed with a consideration for compound availability, reference database, performance characteristics of the predictive signatures of interest, and so on. For instance, in our laboratory, an *in vivo* toxicogenomic hepatic reference database has been generated based on 3-day repeat-dose studies in rats. Consequently, we typically assess our compounds for hepatotoxicity in 3-day repeat-dose studies. In several companies, toxicogenomic analysis is now integrated in most 2–4 week rat repeat-dose toxicology studies. The objective is usually to be proactive in case unexpected toxicologic changes occur in the studies or to enhance knowledge and expertise in gene expression analysis. In addition, these types of practice are useful to promote a better acceptance of these new technologies in a company.

*Dose* In a mechanistic *in vivo* study, the dose will be selected based on the best chance to consistently reproduce the toxicologic change of interest. In contrast, the method used to develop the predictive signatures will dictate the dose to be used in a predictive study. Validation of predictive signatures should address to some extent their predictive power and accuracy for various dose levels. For instance, in our laboratory, our *in vivo* databases and signatures have been developed based on a low and a high dose. The low dose corresponds to an estimated pharmacological dose (a dose resulting in an exposure similar to that achieved in the $ED_{50}$ in the most appropriate animal model), while the high dose corresponds to a maximal tolerated dose (MTD) for a 3-day study (defined by the highest dose before rats exhibit clinical signs of toxicity, lack of body weight gain, or a significant decrease in food consumption). It is clear that in the vast majority of cases, because these studies are typically conducted early in a project, only limited information is available for selecting the doses. Consequently, dose setting can be a challenging task and requires extensive communication among all project stakeholders. An adequate dose selection is, however, crucial for proper decision making and, in our experience, is central to a successful prediction of toxicologic changes.

*Samples* When conducting an *in vivo* study, we strongly recommend that all major tissues, including blood, be collected for concurrent histopathologic and clinical pathologic examinations. The latter are relatively inexpensive compared to the current cost of microarrays, can be performed quickly in a discovery setting, and provide critical information that helps interpret gene expression data in the context of changes in the homeostasis of the tissue being analyzed and in the context of the overall status of the animal. For instance, if clinical observations and analysis of the clinical pathology and histopathology changes suggest that an animal is moribund, gene expression changes may be more reflective of the overall poor condition of the animal than of a specific mechanism of toxicity. In our laboratory, for predictive studies, we typically select or prioritize the tissues to be evaluated with microarrays based on a prior histopathologic and clinical pathologic evaluation. For instance, if a test article demonstrates dose-limiting hepatotoxicity as evidenced by hepatocellular necrosis and elevations of serum transaminases, it would be moot and redundant to evaluate the liver by gene expression analysis. However, if the project team considers it important to understand the pathogenesis of this hepatoxicity, microarray analysis of the liver may be warranted.

## 3 TOXICOGENOMIC REFERENCE DATABASES

### 3.1 The Need for Toxicogenomic Reference Databases

Gene expression changes, when viewed in isolation, can often lead to more questions than answers [39]. Furthermore, although changes in expression of individual genes may be of importance, it is usually more appropriate to examine gene expression changes by looking at pathways being regulated. The analysis of pathways increases the confidence that a change in expression of specific genes has biological implications, but increases the accuracy of the overall interpretation. Abundant historical data are not yet available and general experience is still insufficient for the meaningful interpretation of thousands of simultaneous gene expression changes, which may often appear disconnected. Only after a specific toxicologic change has been observed consistently can its real toxicologic significance be understood. The same holds true for gene expression changes. A recent study nicely illustrated this aspect. In this study, over 300 liver microarray experiments covering three different classes of compounds (genotoxic carcinogens, nongenotoxic carcinogens, and noncarcinogens) were clustered across 72 putative oncogenes [40]. The three classes of compounds were strikingly interdispersed within the cluster, indicating that upregulation of oncogene expression was not a surrogate marker for carcinogenesis, as both noncarcinogens and carcinogens were upregulating the expression of these selected genes. This study also reiterates the danger of focusing on the change in expression of single or a limited set of genes in a microarray experiment.

Gene expression changes induced by toxicants typically reflect a large number of complex pharmacological, physiological, and biochemical processes [18, 41]. To generate a plausible mechanistic hypothesis for the pathogenesis of a toxicologic change, the gene expression changes related to the toxicity need to be identified and separated from those that are adaptative, beneficial, or unrelated to the development of the toxicologic change. This requires an appropriate study design, including the evaluation of multiple time points, but mostly the access to reference data. The use of reference compounds may clarify or confirm which gene expression changes are related to a specific lesion or how the lesion develops. However, contextual information from large, established reference databases is optimal to properly interpret gene expression data by correlating unique gene changes to those associated with treatment with a large repository of compounds or with specific toxicologic mechanisms [39]. The large number of compounds, tissues, corroborative toxicologic and pathologic changes, and gene expression data in these reference databases allow one to strengthen statistical inferences [42]. The concept of databases in toxicology is not novel, and as an illustration, one can think of databases for serum chemistry, hematology, pathology, or carcinogenesis.

### 3.2 Design and Development of Toxicogenomic Reference Databases

The ideal toxicogenomic database contains gene expression profiles induced in various tissues following the treatment of the reference species (most often rats) with a variety of reference toxicants (known pharmaceutical agents, prototypical toxicants) and control compounds, at multiple doses and time points [23, 43–47].

The reference compounds profiled in the database should reflect a variety of toxic mechanisms and represent different structure–activity relationships [42]. The use of multiple doses of the reference compounds (e.g., an efficacious dose and a maximum tolerated dose) is extremely useful to distinguish a pharmacological effect from a toxicological effect. In some situations, time-course data can be very useful to identify gene expression changes linked to a time-dependent toxic response and can increase the chances of observing a true toxic response. The number of animals required for each time point or dose is also an important consideration. Biological replicates are useful to establish both the biological and technical variability. As mentioned earlier, when rats are exposed for a sufficient period of time to toxicants, the interindividual variability of gene expression changes is relatively small, so that 3 animals per group and per time point are usually considered sufficient for the generation of meaningful gene expression profiles.

In addition to gene expression profiles, a useful database also contains sufficient technical and biological information. The realization by the scientific community that gene expression data can only be correctly understood and put in a perspective if the critical amount of associated information is available had led to the publication in 2001 by the Microarray Gene Expression Data Society of the minimum information about a microarray experiment (MIAME) guidelines for the reporting and publication of microarray experiments [48]. These guidelines have become a reference for the submission of gene expression data to scientific journals and are used as standards by some public and commercial databases.

Experimental inconsistencies can lead to some confusion, as not all datasets relate to one another. In particular, the platform used to generate the reference profiles can limit the value of some experimental datasets. While some marked improvements have been achieved in the ability to extrapolate data from one platform to another and in establishing more consistent gene nomenclature, different platforms still generate slightly different datasets or may be evaluating different genes [49]. Furthermore, even when the same platform is used, gene expression data have been shown to vary across laboratories because of differences in protocols or instrumentations [37]. These issues are the focus of several major initiatives involving academic, industry, and government laboratories with the objectives of identifying, validating, and implementing standards that could be used for gene expression analysis [9, 10, 14]. Overall, however, several confounding factors still exist that currently limit the comparison of gene expression data from one laboratory to another. These experimental aspects should be carefully evaluated when selecting the database to work with.

The nature of the species or strain used can also be a limitation. Ideally, one should compare gene expression profiles generated in a particular species with those generated in the same species or even strain. Practically, this may not always be feasible because of a lack of reference data in some species or strains. Therefore, certain circumstances require extrapolating across species. There have been significant improvements in the annotations of the genome of the major preclinical species currently used in toxicology [15]. However, the mapping of orthologous genes should still be viewed as approximate and needs considerable additional effort to be optimized. Furthermore, not all species react similarly to a specific toxicant and this difference in response also limits the ability to extrapolate from one species to another. Different rat strains may also have different responses to some toxicity and

the difference in response at the transcriptome level has also been addressed. Overall, the transcriptome response following exposure to toxicants has been shown to be usually very similar across different rat strains [50].

### 3.3 Existing Toxicogenomic Databases

Several publicly available, commercial or proprietary databases exist for the analysis of gene expression datasets [28]. Although not all of these repositories are specific to toxicology, they can still represent a useful source of data for the analysis of gene expression. Several recently developed databases are mostly focused on toxicology. For instance, the National Institute of Environmental Health Sciences has recently established the National Center for Toxicogenomics to create a reference knowledge database (Chemical Effects in Biological Systems or CEBS) that would ultimately allow scientists to understand mechanisms of toxicity through the use of gene expression analysis as well as proteomics and metabolite profiling [51, 52]. Companies such as GeneLogic (Gaithersburg, MD) or Iconix (Mountain View, CA) have created large toxicogenomic reference databases containing gene expression profiles induced by prototypical reference compounds with corroborating toxicologic and pathologic endpoints [42, 53]. A list of selected public databases with a brief description of their general attributes is presented in Table 1.

## 4 TOXICOGENOMICS IN DRUG DISCOVERY

### 4.1 Predictive Toxicology

In this chapter, the term *predictive toxicology* is used to refer to the use of short-term assays that allow the toxicologist to predict with sufficient accuracy toxic changes that would occur after longer exposure and to extrapolate toxic reactions from preclinical species to humans [54, 55]. Predictive toxicology assays can be in the form of *in vitro* assays or short-term *in vivo* studies. Because most attempts to use toxicogenomics in predictive toxicology have so far focused on the use of short-term *in vivo* studies, we limit our discussion to this approach. The use of cell cultures is covered later in this chapter.

Large-scale expression analysis is an extremely sensitive approach to detect deregulated genes and signaling pathways that contribute to toxic changes. The main assumption of toxicogenomics is that following exposure to toxicants at relevant doses, transcriptional changes occur before the development of a toxic phenotype as assessed by traditional endpoints, such as clinical observations, hispathologic examination, or clinical pathology measurements. In our experience, this assumption is accurate for the vast majority of toxicities with few exceptions. For this reason, toxicogenomics offers the unique opportunity to reliably identify compounds with toxic liabilities early in the drug discovery process, and to thus significantly improve the productivity of drug discovery [18, 54, 55]. As mentioned earlier, the development of more sensitive and predictive technologies that would allow for the characterization of toxicology early is critical for the selection of molecules with an optimal toxicologic profile. Pharmaceutical R&D units and private toxicogenomics companies have consequently invested significant resources in the development of

**TABLE 1 Some Public Databases**

| Database | Attributes | URL |
|---|---|---|
| Gene Expression Omnibus (GEO) | World's largest public repository<br>Adherence to MIAME guidelines<br>Toxicology data available<br>Exploration, analysis, and visualization tools | http://www.ncbi.nlm.nih.gov/projects/geo/ |
| ArrayExpress | Large public repository<br>Adherence to MIAME guidelines<br>Toxicology data available in Tox-MIAMExpress<br>Expression data from normal human and mouse tissues | http://www.ebi.ac.uk/arrayexpress/ |
| Chemical Effects in Biological Systems (CEBS) | Evolving public toxicogenomics repository from the National Institute of Environmental Sciences (NIEHS) National Center for Toxicogenomics (NCT)<br>Designed to house data from complex studies having multiple data steams (genetic, proteomic, metabonomic data)<br>Exploration and analysis tools | http://cebs.niehs.nih.gov/ |
| Environment, Drugs, and Gene Expression (EDGE) | Public toxicogenomics repository<br>Standardized experimental conditions including standardized microarray platform<br>Mostly focused on mouse liver microarray data<br>Useful bioinformatics tools (clustering, BLAST searching, rank analysis, classification tools) | http://edge.oncology.wisc.edu/ |
| Symatlas | Product of the Genomics Institute of the Novartis Research Foundation<br>Expression data from a large panel of normal human and mouse tissues or cell culture models | http://symatlas.gnf.org/SymAtlas/ |
| DbZach System | Toxicogenomic database allowing data mining and full knowledge-based understanding of toxicologic mechanisms<br>Contains correlating clinical chemistry parameters and histopatholigc data | http://dbzach.fst.msu.edu/ |

predictive gene expression-based assays. However, relatively few of these efforts have been formerly published, such that the majority of the information used here is based on personal experience.

Predictive toxicology should ideally be applied at the lead selection and lead optimization stages, concurrently with the other assays used to assess drug-like

properties of molecules (pharmacologic and physicochemical properties, ADME pharmacokinetic characterization). Predictive toxicogenomics is not yet amenable to such an early stage because of its relative immaturity, low throughput, and significant cost. In addition, while most available data suggest that expression analysis will allow for the use of gene expression-based assays early in the discovery process, not enough data related to the accuracy of this approach are available to fully assess the value added by this technology. Nevertheless, several companies are using gene expression analysis to rank-order or prioritize compounds using short-term rat studies based on their toxic potential in specific tissues, in particular, liver.

### 4.2 Development of Predictive Gene Expression Signatures

As pointed out earlier, changes in expression of single genes are typically not sufficient to predict or identify existing toxic changes. Rather, changes in the expression of a gene set are more likely to correlate with toxicity. Consequently, the first step in predicting toxic changes entails the development of gene expression signatures that strongly correlate with toxic changes. Albeit simple in concept, this task has proved to be substantially more difficult and more resource intensive than initially anticipated. Not all recently developed genomics approaches have demonstrated superiority compared to traditional methods or have been able to adequately validate signatures based on external samples (forward validation process). Indeed, sophisticated statistical tools and biostatistical expertise are needed to address the complexity of the various changes in the transcriptome and to develop predictive toxicogenomic signatures [18, 53, 56]. Some of the most commonly used statistical methods are briefly reviewed next.

Toxicants typically induce hundreds to thousands of gene expression changes, a large number difficult to manage for most statistical methods. Therefore, the first step in developing predictive signatures is to reduce the number of parameters, focusing on the ones relevant to the classification model. Two major approaches are traditionally used for dimensionality reduction. The most commonly used approach is to rank genes with respect to differences in expression between experimental groups using parametric or nonparametric statistical tests, such as standard or permutation t- or F-test, Wilcoxon statistics, or significance analysis of microarrays (SAMs) [43, 44, 57, 58]. The genes that are differentially expressed at a specified significance level or a fixed number of top ranking genes are then selected for inclusion in the prediction model. The second approach for dimension reduction is to use noise reduction methods, such as principal component analysis (PCA) [59]. The PCA method reduces a large set of genes into several components, where each new component is a weighted linear combination of all genes. These components are rank-ordered according to the amount of variance, and the first component represents the greatest variability among the samples. By selecting the first $n$ components as most informative, the dimension of gene expression data is vastly reduced.

Once data have been reduced and informative genes have been selected, predictive models can be developed with the objective of classifying compounds as toxic or nontoxic in the relevant tissue. This step requires a training set consisting of a repository of gene expression profiles encompassing both toxic and nontoxic compounds, and the use of computational algorithms that will allow for the accurate

classification of unknown samples. These computational methods require significant numbers of gene expression profiles to generate useful predictive models. Several computational algorithms have been used for the analysis of microarray-generated gene expression data. They include logistic regression, linear discriminant analysis (LDA), naive Bayesian classifiers, artificial neural networks (ANNs), and support vector machines (SVMs). Both logistic regression and LDA use statistical inference to weigh the contributions of each signature gene expression value in sample prediction and typically require tens to hundreds of datasets, depending on the variability of the gene expression data. In situations where there is a clear difference between groups, these methods can be quite robust [43, 53]. Naive Bayesian classification is a popular approach for classification. In one study, gene expression profiles were generated from the liver of mice treated with 12 compounds, representing 5 well-characterized classes of hepatotoxicants (peroxisome proliferators, aryl hydrocarbon receptor agonists, noncoplanar polychlorinated biphenyls, inflammatory agents, and hypoxia-inducing agents) [45]. Using a naive Bayesian classification, a predictive signature consisting of 12 genes accurately classified all samples into their chemical groups. Artificial neural networks (ANNs) are analogous in concept to a biological nervous system. They are composed of a number of highly interconnected processing elements called neurons or nodes tied together with weighted connections. An iterative learning process begins by feeding input data into the network, which calculates the predicted output based on predetermined weights. Comparison of the predicted output and the targeted output leads to adjustment of the weights of the connections, and a new output is calculated. This process is reiterated until the network output closely matches the targeted output. ANNs have the advantage of learning complex patterns and of learning from new information; they have gained increasing popularity for the classification of gene expression profiles in many disciplines. In our laboratory, as illustrated later, the ANN approach has proved to be extremely powerful for the generation of predictive signatures of hepatotoxicity in rats. Support vector machines (SVMs) are a relatively new type of learning algorithm with robust performance with respect to sparse and noisy data [60]. SVMs operate by finding an optimal demarcation that most distantly separates positive and negative samples. Unlike other classification algorithms, SVMs perform very well with a large number of data and this feature makes SVMs especially attractive for the classification of gene expression data, which usually have a large number of gene expression endpoints and a limited number of samples [61, 62]. One can easily be confused with the choice and complexity of these various classification algorithms. The ultimate criterion to select an optimal prediction model is the prediction accuracy, which is estimated using a testing set that is different from the training set during a validation step. Although most studies evaluate the robustness of predictive models using various validation approaches, no exhaustive survey of the various prediction methods is available in the literature. For predictive toxicogenomics, the choice of methods is likely dependent on the experimental design. For instance, an LDA approach is an easy alternative and performs very well if there is a clear difference between the classes of toxicants and/or the sample size is relatively large. However, in the case of a more complex system, such as general hepatotoxicity, with a limited number of gene expression profiles, a machine learning approach, like SVMs or ANNs, could be more appropriate [18, 62].

### 4.3 Case Examples

***Prediction of Hepatotoxicity*** Hepatotoxicity has been the toxicity of choice for most toxicogenomic studies. Indeed, the liver is a common target organ for many toxicants and has been extensively studied. This has promoted the generation of comprehensive gene expression databases that facilitate interpretation of liver-derived gene expression profiles. Additionally, the liver is a rather homogeneous tissue, as opposed to tissues such as intestines or brain, for instance. The liver is composed mostly of hepatocytes sharing similar biochemical functions that translate into relatively uniform gene expression profiles. This homogeneity makes identification and interpretation of gene expression changes easier. For these reasons, a wealth of published and proprietary gene expression information is available for the liver, and hepatotoxicity can now be relatively well predicted and understood with gene expression profiling.

As a proof-of-concept, our laboratory has developed, a few years ago, a quantitative approach to predict hepatotoxicity based on gene expression profiles [18]. We first constructed an internal database containing microarray-generated liver gene expression profiles from 3-day rat toxicology studies using over 50 hepatototoxicants and nonhepatotoxicants. All compounds were administered to 3 male rats/group at a high dose (a dose expected to induce hepatotoxicity after 1 week of treatment) and at a lower, nonhepatotoxic dose. A set of marker genes was identified that distinguished the hepatotoxicants from the nonhepatotoxicants using ANOVA analysis. Using an artificial neural network algorithm coupled with principal component analysis for dimensionality reduction, a quantitative model was established to classify compounds according to a composite toxicity score (Fig. 6). A forward validation step was conducted using additional compounds that were not part of the original database. The neural network algorithm could successfully classify these compounds based on their potential to cause hepatotoxicity with a high degree of sensitivity and specificity. This predictive hepatotoxicity assay is now routinely used to prioritize compounds using exploratory 3-day repeat-dose rat studies for various projects.

***Prediction of Nephrotoxicity*** A recent publication illustrates how microarray-generated gene expression profiles from the kidneys of rats treated for short periods of time with various nephrotoxicants and nonnephrotoxicants can be used to predict toxic changes in longer term studies [25]. In this study, using a large commercial reference database, a predictive gene expression signature of renal tubular toxicity was developed and shown to predict with good accuracy renal tubular changes that would typically occur after longer exposure to the toxicants. These gene expression profiles were derived from the kidneys of rats treated for 5 days, a time point where no obvious toxic changes were evident, as evidenced by the lack of histopathologic observations and the lack of changes in serum chemistry parameters (most notably serum creatinine and blood urea nitrogen). In a step to confirm that tubular injury would ultimately occur in longer term studies, the authors dosed rats up to 28 days with the 15 nephrotoxicants used in their positive class, as a phenotypic anchor to the predictive signature. In addition, the signature was validated using compounds naive and not structurally related to the training set. These test compounds included

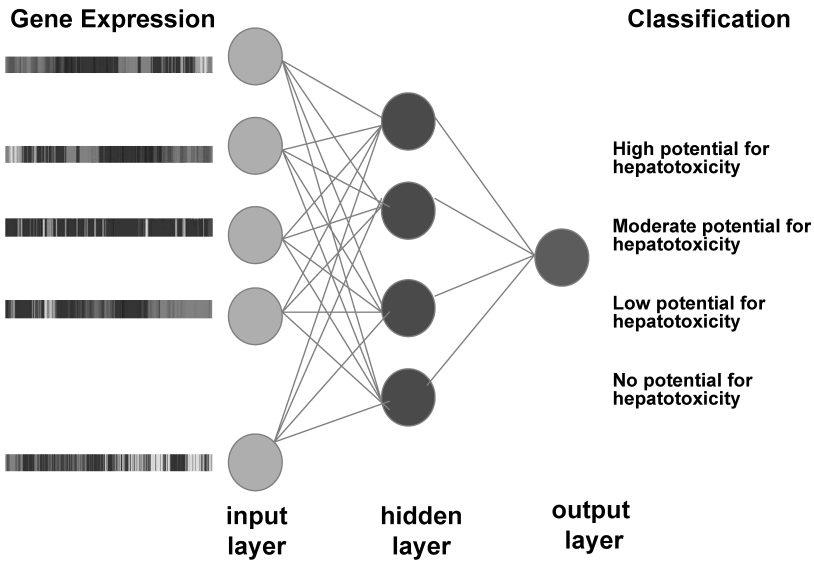


**FIGURE 6** Schematic representation of the ANN-based approach used for the development of a predictive hepatotoxicity toxicogenomic assay. An internal rat liver gene expression database was first constructed. Reference gene expression profiles were then reduced in dimensionality using PCA. The filtered gene expression profiles were fed into an ANN, which processes the information from one layer to the next using multiple weighting factors and transfer functions. The output of the ANN is compared with the ideal toxicity classification and the model is readjusted. This learning process is repeated until the model is able to make an accurate prediction. Using microarray-generated gene expression profiles from male Sprague–Dawley rats treated for 3 days with experimental compounds, the neural network algorithm classifies the compounds based on their potential to cause hepatotoxicity in rats on a four-category scale ranging from no potential to high potential. (From Ref. 18, with permission from Elsevier.)

9 nephrotoxicants and 12 nonnephrotoxicants. The signature correctly predicted the future presence or absence of renal tubular injury in 76% of the compound treatments.

Another study used a similar, yet slightly different, approach to study time- and dose-dependent gene expression changes associated with proximal tubular injury in the rat [63]. In this study, Sprague–Dawley rats were treated with a wide variety of classical nephrotoxicants and renal gene expression profiles were evaluated 1, 3, and 7 days after initiation of dosing. The gene expression profiles nicely clustered based on the severity and nature of the pathologic changes and were consistent with tubular degeneration, regeneration, and necrosis. In addition, using an SVM-based approach and a training set of 120 gene expression profiles, a predictive classifier was developed that was able to predict the type of pathology of a testing set composed of 28 gene expression profiles with 100% selectivity and 82% sensitivity.

These two studies are good illustrations of the validity of the toxicogenomic approach for predictive toxicology but are also a good indication of the scale involved in the generation of predictive signatures of sufficient accuracy to be useful.

Both studies required the use of significant numbers of prototypical toxicants for both the training and testing sets, multiple time points, 3 or 4 rats/treatment group/time point, and sophisticated computational algorithms. In other words, such studies are associated with a significant investment of resources, and it is likely that continued validation of these early signatures by a combined effort of the broad scientific community will be required to improve accuracy, but also to better estimate their predictivity and suitability in early toxicology testing.

***Prediction of Carcinogenicity*** Assessment of the carcinogenic potential of compounds in the rodent bioassay is expensive and lengthy and cannot be performed until late in a program. In addition, the endpoint of the rodent bioassay consists essentially of macroscopically and microscopically detectable tumors, which may spontaneously occur in naive, aging animals and whose incidence varies tremendously among animals, thereby requiring a significant number of animals to obtain sufficient statistical power [64]. Toxicogenomics applied in a predictive mode would allow for a profound reduction in the duration of dosing, thereby reducing the amount of compound required and permitting an earlier assessment of the compound. Furthermore, a response at the transcriptome level more homogeneous than the development of tumors could be expected, thereby allowing for a significant reduction in the number of animals to be dosed. It is obvious that any predictive signature of a carcinogenic effect would not dispense with the requirement to conduct the regulatory carcinogenicity studies. However, such predictive signatures would clearly be valuable when applied early to make more informed decisions on compounds.

Using a 5-day repeat-dose toxicity study in rats, Kramer et al. [65] evaluated whether gene expression profiling could help identify candidate molecular markers that may predict hepatic carcinogenicity induced by either nongenotoxic or genotoxic compounds. They hypothesized that there might exist a small number of critical genes whose expression may be predictive of the early events of carcinogenesis initiated by multiple mechanisms. Their study included three dose levels of a number of well-characterized compounds, including five nongenotoxic carcinogens, one genotoxic carcinogen, one carcinogen that may act via genotoxicity, a mitogen, and a noncarcinogenic hepatotoxicant. Analysis of the hepatic gene expression profiles resulted in the identification of several genes (including CYP-R and TSC-22, which were upregulated and downregulated, respectively) whose expression correlated well with the estimated carcinogenic potential. CYP-R catalyzes the transfer of electrons from NADPH to heme oxygenase, cytochrome-$b_5$, and a variety of cytochromes P450. As correctly stated by the authors, the upregulation of this gene may simply represent a surrogate marker for cytochrome P450 induction, since many nongenotoxic carcinogens are also cytochrome inducers. Alternatively, CYP-R induction may reflect a role for oxidative stress in rodent hepatocarcinogenesis. TSC-22 belongs to a subfamily of leucine zipper transcription factors and had been previously shown to be regulated by various treatments, such as TGF-β, phorbol ester, progesterone, or vesnarionone. Although the exact role of TSC-22 is uncertain, its repression may reflect regulation mediated directly by the test articles or an adaptive response to these compounds. In either case, its regulation supports an altered balance between proliferation and apoptosis, consistent with what would be expected with carcinogens.

A second study used gene expression profiling to evaluate whether known genotoxic carcinogens would induce a common set of genes belonging to defined biological pathways and whether these genes could be used as a predictive signature for hepatic genotoxic carcinogens [44]. This study used potent carcinogens (such as dimethylnitrosamine or 2-nitrofluorene) dosed daily for up to 14 days in rats. Gene expression analysis of livers indicated that the following biological pathways were mostly deregulated: DNA damage response, specific detoxification response, proliferation and survival, and structural changes. This common pattern of deregulation is consistent with what would be expected in the early events of tumorigenesis and could be predictive of later tumor development. The same investigators followed up on these encouraging preliminary results with a study incorporating nongenotoxic carcinogens [66]. In contrast to what was seen with the genotoxic compounds, nongenotoxic carcinogens impacted distinct cellular pathways/response (including oxidative DNA or protein damage, cell cycle progression, tissue regeneration) that were consistent with compound-specific mechanisms and the two-stage model of carcinogenesis.

Several important lessons can be learned from these pioneering studies. First, neither a single gene nor a single pathway will be sufficient to predict and discriminate the two classes of carcinogens. Second, a predictive gene expression signature of relatively good accuracy can likely be generated once a sufficient repository of gene expression profiles from a larger variety of carcinogens at different doses and different time points become available. Indeed, for a project of that scale, it is likely that a collaborative effort from the scientific community will be necessary for the refinement and validation of any predictive signature of carcinogenicity.

### 4.4 Predicting Species-Specific Toxicity

Toxicologic changes occurring in preclinical species are not necessarily relevant to humans because of species differences in cell biology, physiology, or responses to changes induced by compounds [67]. A classic example of a toxicity with no relevance to humans involves the peroxisome proliferators, such as the fibrate class of cholesterol-lowering drugs, that activate the peroxisome proliferator-activated receptor-α (PPAR-α). Upon chronic administration, these compounds cause hepatomegaly and eventually hepatic neoplasms in rats [68, 69]. There are marked species differences in the response to peroxisome proliferators, with mice and rats being highly responsive in contrast to humans. This differential species response correlates directly with the number of hepatic PPAR-α; PPAR-α is expressed in human liver at only 5–10% of rodent liver levels. Consequently, humans are at minimal or no risk to develop hepatic tumors following chronic exposure to peroxisome proliferators. Toxicogenomics has furthered the understanding of the molecular mechanisms associated with the various effects of several peroxisome proliferators [23, 70, 71]. This allows one to more specifically demonstrate the mechanisms of action by which certain compounds lead to rodent hepatomegaly and hepatic carcinogenesis, thereby improving overall risk assessment. The value of toxicogenomics to understand species-specific responses is also illustrated with the case of cyclosporine-induced nephrotoxicity [72, 73]. In the kidneys of cyclosporin A-treated rats, a marked downregulation of calbindin-D28kDa, a calcium binding protein, correlates with and causes the accumulation of calcium in tubules and ultimately renal tubular

calcification in this species. In contrast, cyclosporine does not regulate calbindin-D28kDa expression in the kidneys of dogs and monkeys, two species resistant to cyclosporine-mediated renal toxicity.

## 5 *IN VITRO* TOXICOGENOMICS

### 5.1 Objectives of *In Vitro* Toxicogenomics

Compounds are ultimately assessed in animal toxicology studies. Not surprisingly, the vast majority of published toxicogenomic studies to date have been using tissue from animals dosed *in vivo*. *In vivo* studies require large amounts of compound and consequently do not allow for an early characterization of the toxicologic profiles of compounds. Moreover, the number of compounds that can be analyzed in animal studies is limited, in part because of the cost and practicalities of these studies. Thus, *in vitro* systems may significantly improve the throughput and increase the value of toxicogenomics in drug discovery by allowing for an early toxicological characterization of compounds. In addition, gene expression profiling using *in vitro* systems may identify biomarkers of toxicity in the form of gene sets that could be transferred and investigated in preclinical or clinical studies to monitor possible toxic reactions. Finally, gene expression studies in human cells, such as primary human hepatocytes, may, in some cases, be more relevant to the clinical situation or allow for a better understanding of the relevance of toxic changes and for a better assessment of safety risks for humans.

The selection of an appropriate cell system should be guided by the questions to be addressed. If one desires to identify the mechanism of toxicity of a compound, using a cell system that most closely mimics the target organ, such as a primary cell system, would be preferable. However, if one wishes to identify markers of general toxicity, then the cell type may not be as important. In fact, it is likely that identifying general markers of toxicity (such as DNA damage, apoptosis, or oxidative stress) would be feasible using cell lines of various origins, such as HeLa or Jurkat cells.

The major limitation of *in vitro* systems is their inability to recapitulate the overall complexity of the living organism, which limits their potential in detecting lesions associated with multicellular interactions. Most *in vitro* systems are also short term and therefore inadequate to detect chronic effects [74]. Understanding the predictive value of *in vitro* systems has remained a major challenge for toxicologists for decades and *in vitro* toxicogenomics falls into the same predicament. Several studies have addressed the relationship between *in vitro* and *in vivo* toxicogenomic results. These studies have demonstrated that different mechanisms of toxicity can be identified using gene expression profiles generated from *in vitro* systems and that consequently the concept of predictive signatures was also relevant to *in vitro* systems [75, 76].

Although gene expression profiling using *in vitro* systems can distinguish compounds with different mechanisms of toxicity, signatures of satisfactory accuracy and cost-effective gene expression platforms with adequate throughput are necessary for its implementation in a discovery setting. For an *in vitro* toxicogenomic assay to have practical applications in drug discovery, gene expression signatures need to be

generated and validated for several relevant toxicologic endpoints. In addition, in order to increase the throughput of gene expression profiling (e.g., adapting it to a 96- or 384-well format), it would be advantageous to reduce the number of genes being monitored. Ideally, one would want to rapidly evaluate compounds for several toxicologic endpoints in a simultaneous fashion in a limited number of wells.

The selection of an appropriate dose represents a critical issue. At this point, there is no clear consensus of what represents an ideal dose for *in vitro* toxicogenomic assessment, and it is likely that the dose selection will depend on the cell type. In our experience, development of robust predictive signatures and characterization of the toxicologic profiles of compounds require the use of relatively high doses, sufficient to cause cytotoxicity. For instance, in our primary rat hepatocyte model, we typically characterize compounds at concentrations sufficient to cause death of 20% of cells. Failure to reach these cytotoxic concentrations will result in an insensitive assay of limited value.

### 5.2 Proof-of-Concept Using Primary Rat Hepatocytes

Most published *in vitro* toxicogenomic studies have evaluated rat liver cells for several reasons. First, liver is a common target organ of toxicity. Second, rat liver cells are most commonly used for *in vitro* toxicologic studies. Third, the use of hepatocytes offers the opportunity to assess the toxicity associated with certain metabolites without prior metabolic activation. Finally, since most *in vivo* studies have focused on liver, liver-derived cells can be used to correlate *in vitro* and *in vivo* data. An *in vitro* system for hepatotoxicity could consist of isolated perfused livers, liver slices, isolated hepatocytes, or liver cell lines. Isolated perfused livers and liver slices maintain intact cellular interactions and spatial arrangements and allow for long-term studies [74, 77]. These models are also the most appropriate for studying toxic effects on the biliary system, because they contain phenotypically and functionally intact biliary epithelial cells. However, isolated livers and liver slices are resource intensive, low-throughput systems. Cell lines are readily available, cost effective, and generally yield reproducible results over time. However, liver cell lines are quite different from liver or primary hepatocytes in terms of function and phenotype. Gene expression profiles were compared for rat livers, rat liver slices, primary rat hepatocytes cultured on collagen monolayer or collagen sandwich, and two rat liver cell lines (BRL3A and NRL clone 9 cells) [20]. Liver slices were the most similar to intact rat livers, followed by primary hepatocytes in culture. In contrast, the two rat liver cell lines showed little correlation to intact rat livers. In particular, the cell lines expressed very low or undetectable levels of phase I metabolizing enzymes, both at the RNA and protein levels [20]. These results are consistent with data generated in our laboratory (Fig. 7).

Isolated hepatocytes are not identical but are sufficiently close to intact livers in terms of gene expression analysis (Fig. 8) [20, 78]. In addition, they maintain the enzyme architecture and metabolizing capabilities of intact liver in short-term cultures [79, 80]. However, isolated hepatocytes, especially of human origin, can be very difficult and expensive to obtain. Furthermore, in the case of human hepatocytes, lifestyle differences of the donors, such as smoking or drinking habits, medications, or general health, lead to substantial interindividual variability in gene expression profiles. However, this variability is overall not a major concern, since preliminary

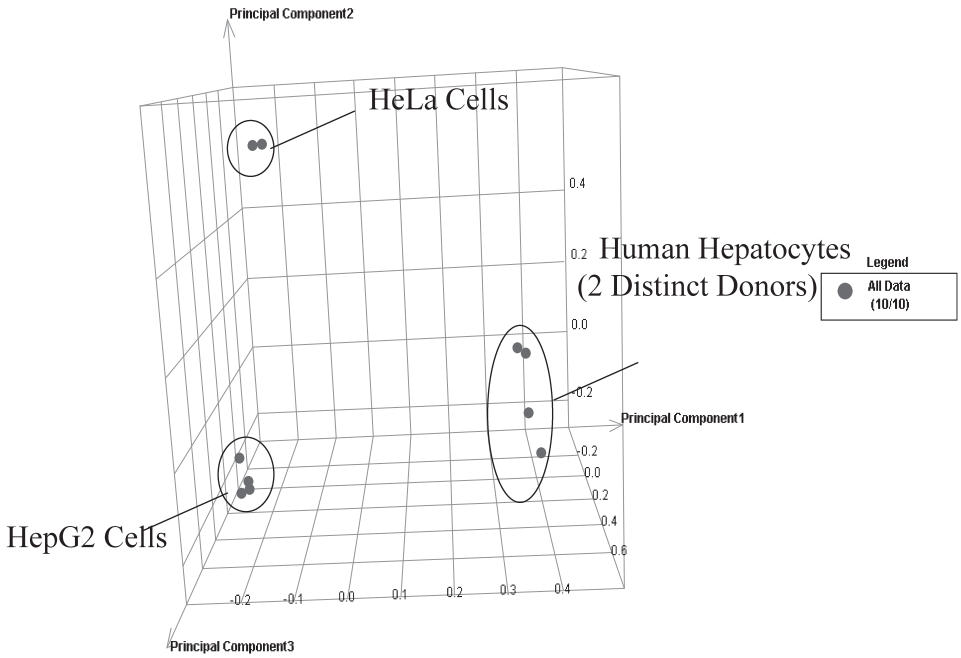


**FIGURE 7** Principal component analysis of Affymetrix HG-U133A microarray-generated gene expression profiles from primary human hepatocytes, HeLa cells, and HepG2 cells. This analysis illustrates the major differences at the transcriptome level between primary human hepatocytes and liver cell lines, such as HepG2 cells, which are almost as dissimilar to human hepatocytes as HeLa cells, which are derived from a cervical carcinoma.

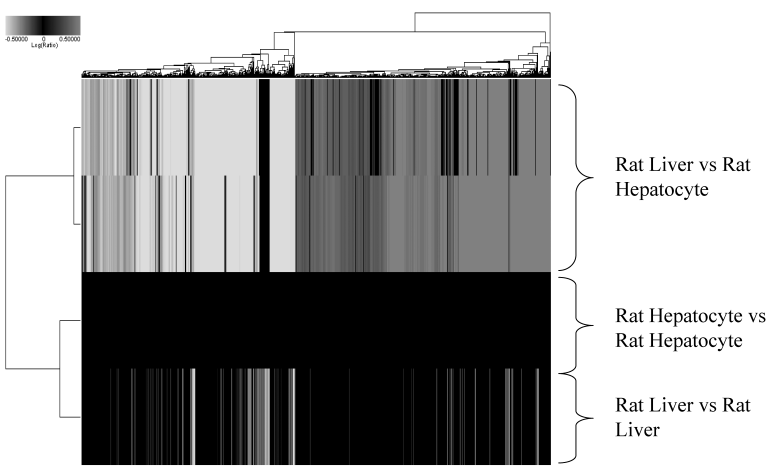


**FIGURE 8** Heatmap illustrating the differences in gene expression profiles between rat liver and primary rat hepatocytes. Gene expression profiles were generated using Affymetrix rat RAE230A microarrays. Genes that are at least showing a two fold difference in expression levels with a $p$ value less than 0.01 between the two systems are shown.

studies indicate a surprisingly low interindividual variability in response to high exposure to toxicants for the most robustly regulated genes [81].

In our laboratory, we routinely use primary rat hepatocytes cultured on collagen to characterize compounds at the gene expression level. In this model, we have profiled a large number of compounds, thereby generating an internal database that has allowed us to develop predictive signatures for several toxicologic endpoints (Fig. 9). For instance, our laboratory has reported results from a study profiling at the gene expression levels 15 well-characterized hepatotoxicants in primary rat hepatocytes [75]. Compounds with similar mechanisms of toxicity, such as the aromatic hydrocarbon (Ah) receptor ligands Aroclor 1254 and 3MC, resulted in similar expression profiles and, using unsupervised hierarchical clustering, could clearly be distinguished from other agents such as carbon tetrachloride and allyl alcohol. This study also demonstrated a significant correlation between the genes regulated *in vivo* and *in vitro* for some toxicants, such as the Ah receptor ligands. Similar results have been confirmed, reproduced, or expanded in studies conducted by others. For instance, in a study using primary rat hepatocytes exposed to 11 different hepatotoxicants and a low-density array platform containing only 59 genes, compounds could correctly be classified into different mechanistic hepatotoxic classes [82]. Other studies have also used liver-derived cell lines with some success. Three separate studies used a human hepatoma cell line (HepG2 cells) to demonstrate that transcriptional analysis differentiates compounds and that *in vitro* toxicogenomics can be used to further the understanding of toxic mechanisms [21, 83, 84].

### 5.3 Use of Gene Expression Profiling to Assess Genotoxicity

Toxicogenomics has also been applied in other cell systems. In particular, gene expression analysis has been evaluated as a potential tool to gain a better under-

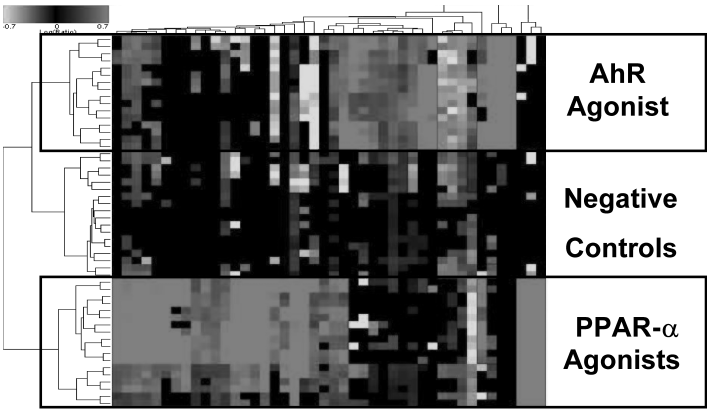


**FIGURE 9** Heatmap of gene expression changes following treatment of primary rat hepatocytes with 16 aryl hydrocarbon receptor (AhR) agonists, 18 negative control compounds, and 18 peroxisome proliferator-activated receptor-α (PPAR-α) agonists for 48 hours. The genes shown were selected by linear discriminant analysis. Using a small internal database of *in vitro* gene expression profiles in primary rat hepatocytes, small gene sets can be identified using various statistical algorithms (in this case linear discriminant analysis) to classify compounds according to their toxic properties for specific toxicologic endpoints.

standing of genotoxic mechanisms. The current *in vitro* genotoxicity assays using mammalian cells (mammalian mutation and/or chromosomal damage assays) provide a limited insight into genotoxic mechanisms [22, 85]. Not surprisingly, the majority of positive genotoxicity findings for marketed drugs with negative carcinogenicity data have been observed in these *in vitro* mammalian assays, demonstrating their low specificity and the need to develop approaches enabling mechanism-based risk assessment [86]. More specifically, the differentiation of DNA-reactive versus DNA-nonreactive mechanisms of genotoxicity would facilitate risk assessment of positive findings in the *in vitro* mammalian cell-based assays. Several studies have addressed the question of whether gene expression profiling of *in vitro* systems would allow a better risk assessment of genotoxicants [85, 87]. Data from these studies demonstrated differences in gene expression profiles between DNA damaging and non-DNA-damaging compounds. In various cell types (p53-deficient mouse lymphoma cells L5178Y/TK$^{+/-}$, TK6 cells), DNA-damaging compounds regulated genes involved in cell cycle regulation, DNA repair, apoptosis, and cellular signaling that were distinct from those regulated by non-DNA-damaging agents [22]. This suggests that, although toxicogenomics will not replace the current standard genotoxicity assays for hazard identification, it can serve as a useful follow-up experimental approach to evaluate compounds with positive findings in these standard assays.

### 5.4 Current and Future Use of *In Vitro* Toxicogenomics

A tremendous amount of work remains to be done for *in vitro* toxicogenomics to become a routine tool for toxicologic characterization of compounds in discovery. However, this area is moving at an extremely rapid pace, and several companies are already using gene expression-based *in vitro* assays for compound characterization and prioritization. For instance, our laboratory has identified predictive signatures for specific toxicologic endpoints in the rat hepatocyte model. Some of these signatures, as well as the methods used to generate them, have already been published [18, 75, 88]. One can envision that efforts in this field will soon clarify the strengths and limitations of the *in vitro* toxicogenomic approach and thus establish the role of *in vitro* toxicogenomic assays in drug discovery.

## 6 TOXICOGENOMICS IN MECHANISTIC TOXICOLOGY

### 6.1 Objectives of Mechanistic Toxicology

The impact of gene expression profiling in drug discovery and development has so far been mostly evident when used to elucidate the mechanism of a specific toxicity. Toxic changes are commonly identified in preclinical studies and obviously not all toxic changes are worth investigating. The decision regarding whether a specific toxicologic change needs to be mechanistically understood is based on multiple factors. These factors include, for instance, the nature of the toxicologic change, the exposures at which the change occurs, the species affected, the availability of good backup compounds, or the stage of the program [18]. For instance, there is clear value in trying to understand the mechanism of tumorigenesis for test article-related

tumors detected in a lifetime bioassay. At this stage of a program, clinical trials are well underway and an enormous amount of resources have already been invested. In contrast, in an early program, if several backup compounds with a different chemistry but similar physicochemical and pharmacological properties are available, it may not be worth investigating the mechanism of a specific toxicologic change occurring with a single lead compound. However, if backup compounds are not yet available, understanding the molecular basis of toxicologic changes may be useful to properly select backup compounds without this toxicologic liability through early structure–toxicity relationship studies during lead optimization. Indeed, when the mechanism of toxicity is understood, appropriate counterscreens can rationally be developed that allow for the selection of backup compounds unlikely to induce the same toxic change. Finally, there is a lack of sensitive and specific biomarkers for some toxic changes. Because gene expression profiling provides a global view of the transcriptional effects induced by a compound, it may identify biomarkers that can subsequently be used in preclinical and conceptually in clinical studies to monitor toxicity.

### 6.2 Case Examples of Mechanistic Toxicology

***Hepatotoxicity*** Early studies have shown that changes in expression of small gene sets can reliably discriminate compounds with distinct mechanism of toxicity in the liver. These gene sets can be used to mechanistically classify compounds and assign a compound with an unknown toxicologic mechanism into predefined classes based on mechanism [23, 70]. This allows one to identify toxicity using predictive gene expression signatures as discussed before and also to establish mechanistic hypotheses by comparing expression profiles with those present in a database. For instance, our laboratory has investigated the hepatic effects of A-277249, a thienopyridine inhibitor of NF-κB-mediated expression of cellular adhesion molecules [89]. This compound induced hepatic changes in rats in a repeat-dose toxicity study, including increased liver weights, changes in serum chemistry (elevations of serum transaminases, alkaline phosphatase, and gamma glutamyl transferase), and histopathologic changes (hypertrophy and hyperplasia of hepatocytes and biliary epithelial cells). To investigate the mechanism of this hepatotoxicity, a 3-day repeat-dose rat toxicity study was conducted. Livers were collected and gene expression profiles were generated using microarrays. The compound was observed to induce extensive changes in gene expression. Using a proprietary gene expression database of known hepatotoxicants, agglomerative hierarchical cluster analysis demonstrated that A-277249 had a gene expression profile similar to Aroclor 1254 and 3-methylcholanthrene (3MC), two well-characterized activators of the aryl hydrocarbon nuclear receptor (AhR), indicating that A-277249 hepatic changes were, at least in part, mediated by the AhR either directly or through effects on NF-κB [89]. The AhR is a nuclear receptor that mediates responses to various toxicants, such as the halogenated aromatic toxicants [90]. In this particular case, the chemical class was abandoned. But, in a case where backup compounds would have been available, one could have used these data to set up an appropriate counterscreen to rapidly evaluate the backup compounds for this toxic mechanism. In particular, one could have evaluated whether these compounds induce an upregulation of CYP1A1 in primary rat hepatocytes, since ligands of the AhR are known to induce CYP1A1.

***Intestinal Toxicity and Notch Signaling*** Gene expression profiling has rarely been applied to study toxicity of the gastrointestinal tract. The complexity and heterogeneity of this tissue makes it very difficult to investigate toxic changes at the level of the transcriptome. However, two recent studies demonstrated that toxicogenomics can be used to elucidate the toxic mechanisms and identify markers of toxicity for this tissue [26]. In the last few years, functional γ-secretase inhibitors (FGSIs) have been developed as potential therapeutic agents for Alzheimer disease [26, 91]. FGSIs can block the cleavage of several transmembrane proteins, including the cell fate regulator Notch-1, which plays an important role in the differentiation of the immune system and gastrointestinal tract. Rats treated with several FGSIs develop a unique gastrointestinal toxicity, characterized by an increase in gastrointestinal weight, distended stomach and small and large intestines, and a mucoid enteropathy related to goblet cell hyperplasia [26, 91]. Microarray analysis of the duodenum or ileum of FGSIs-treated rats identified changes in the expression of several genes, and these changes confirmed that perturbation in Notch signaling was the mechanism for this characteristic enteropathy. These gene expression studies went further and also identified that the gene encoding the serine protease adipsin was significantly upregulated following treatment with FGSIs. The investigators followed up on this interesting finding and demonstrated elevated levels of the adipsin protein in gastrointestinal contents and feces of FGSIs-treated rats, as well as increased numbers of ileal enterocytes expressing adipsin by immunohistochemistry. Based on these data, both laboratories concluded that adipsin may be potentially exploited as a specific, sensitive, and noninvasive biomarker of FGSIs-induced gastrointestinal toxicity.

***Testicular Toxicity*** Early toxicant-induced testicular changes are typically subtle in early stages without striking morphologic changes, which may easily be missed unless more sophisticated techniques, such as tubular staging, are used [92]. However, in longer term *in vivo* studies (4 weeks or longer), changes are typically more pronounced and advanced. Approaches that would allow for an earlier detection of testicular changes would be beneficial in preclinical safety assessment. Current well-established correlating biomarkers (such as serum FSH or semen analysis) are not sensitive enough to allow for an early detection of toxic changes both in preclinical studies and in clinical trials. Recently, a project sponsored by the ILSI Health and Environmental Sciences Institute (HESI) evaluated the suitability and limitations of additional biomarkers, such as Inhibin B, to detect modest testicular dysfunction in rats [93]. While the complete results of this study have not been communicated at this point, preliminary data have indicated that plasma inhibin B is not a useful marker. Finally, little is known in general about the mechanism of toxicity for testicular toxicants. This lack of understanding relates probably to a lower degree of interest by toxicologists in general in contrast to tissues like liver, but also to the complexity of the tissue. The testis is a tissue composed of several different cell types with striking differences in functions and morphology, which can all be targets for toxicants. Furthermore, these different cell types closely interact with each other, modulating their respective functions and status. Consequently, a toxicant targeting a specific cell type will ultimately affect, by a secondary mechanism, the status of another cell type. For instance, a toxicant that affects the function of Sertoli cells, the cell supporting the growth, differentiation, and release of germ cells, will

ultimately lead to effects on germ cells, such as failure of sperm release or germ cell depletion. This complex cellular interdependence has limited the use of *in vitro* studies to investigate the mechanism of testicular toxicity.

Several elegant studies have demonstrated that gene expression profiling can elucidate the molecular basis of testicular toxicity [94–96]. For instance, gene expression changes in the testis were evaluated following exposure of mice to bromochloroacetic acid, a known testicular toxicant. Using a custom nylon DNA array, numerous changes in gene expression were detected in genes with known functions in fertility, such as Hsp70–2 and SP22, as well as genes encoding proteins involved in cell communication, adhesion, and signaling, supporting the hypothesis that the toxicologic effect was the result of disruption of cellular interactions between Sertoli cells and spermatids [94, 97]. Our laboratory has used DNA microarrays to investigate the testicular toxicity of another halogenated acetic acid, dibromoacetic acid (DBAA). Oral treatment of rats with DBAA at high doses (250 mg/kg/day) induces specific early morphologic changes in the testis, characterized by failed spermiation or failure of release by Sertoli cells of mature step 19 spermatids (Fig. 10). While this morphologic change strongly suggests that the Sertoli cell is the target cell for DBAA, our results using whole testes indicated that DBAA induced a small but consistent downregulation of cytochrome P450c17α (CYP17), an enzyme essential for the production of testosterone by gonads (Fig. 10). These results led us to hypothesize that DBAA may induce, at least in part, its toxicity through an effect on testicular testosterone production. In fact, we were able to show that following treatment of rats with DBAA for as little as 4 days, testicular testosterone contents were significantly decreased, indicating that the decrease in CYP17 expression likely has biological implications. This specific study illustrates two important points. First, because gene expression analysis represents a new and

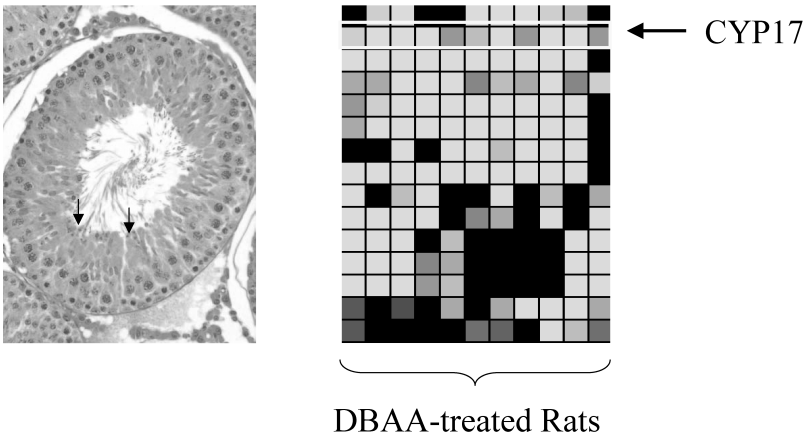


**FIGURE 10** Histologic and gene expression changes induced in rat testis by dibromoacetic acid (DBAA). Oral treatment of rats with DBAA at high doses (250 mg/kg/day for 4 days) leads to failed spermiation, characterized by the failure of release by Sertoli cells of mature step 19 spermatids (arrows). This histologic change correlates with a significant downregulation of cytochrome P450c17a (CYP17) mRNA in the 11 rats evaluated.

global approach to study the molecular mechanism of various pathologies, one should be extremely cautious not to focus exclusively on earlier hypotheses and not to rely on old dogma. In fact, microarray analysis offers the unique opportunity to generate global molecular data that should be used to generate new, unbiased hypotheses. Once generated, these hypotheses should be addressed using appropriately designed studies. Second, although laser-capture microdissection represents a useful technique to study gene expression in complex tissues by focusing on single cell populations, it should also be used with the understanding that critical cell–cell interactions may govern changes in gene expression. In the present case, should we have focused on Sertoli cells only, we probably would have been unable to implicate the decrease in testosterone production as a significant part of the DBAA-induced testicular toxicity.

## 7 TOXICOGENOMICS AND TARGET-RELATED TOXICITY

The recent advances in genomics have spurred a proliferation of novel potential therapeutic targets. The traditional target validation procedure is designed to demonstrate that modulation (inhibition or activation) of target activity in relevant disease models can lead to a therapeutic benefit. The target validation step frequently does not include an assessment of the effects of target modulation on normal cellular, organ, or body function. Yet, developing a good understanding of the potential safety liabilities associated with a target is clearly a critical phase of target drugability assessment. Most of the genomics-derived novel targets play major roles in normal cellular function, and consequently modulation of their activity can lead to toxic changes or what is often referred to as target-related toxicity, mechanism-based toxicity, or on-target toxicity.

Gene expression profiling can provide invaluable information about the biology of these novel targets and help proactively identify target-related safety liabilities. Ideally, this assessment should be conducted at the earliest stages of the discovery process, namely, the target identification/validation stage. With the proliferation of novel attractive targets and the limited resources of discovery units to focus on novel programs, this step becomes especially critical, as it allows a discovery organization to prioritize novel targets based not only on expected therapeutic benefits but also on potential safety liabilities. This ultimately leads to a greater focus on programs most likely to ultimately succeed.

Part of this step involves the evaluation of target expression in normal tissues from both preclinical species and human beings, with the development of quantitative expression tissue maps [54, 98–100]. These tissue maps allow one to proactively identify tissues more likely to be affected by toxic changes. The availability of complete tissue repositories or banks is necessary to generate these tissue maps. Target expression can be evaluated at the level of mRNA expression, protein expression, and enzymatic activity in the case of an enzyme or using assays such as receptor binding assays. Evaluation of mRNA levels is clearly the easiest way to generate expression data, but one should remember that, for a large proportion of proteins, mRNA levels do not necessarily correlate well with protein or activity levels. Therefore, it is usually recommended to confirm or complement mRNA data with appropriate secondary assays. Furthermore, total tissue expression may not

always reflect the critical role played by some targets, especially in cases where a target is expressed in a small compartment of a complex tissue. These situations can be more effectively addressed with either *in situ* hybridization or immunohistochemistry.

The generation of expression tissue maps, albeit useful, is usually insufficient to evaluate potential safety liabilities associated with a target. In particular, these novel targets are frequently identified based on overexpression in diseased samples. Expression in normal tissues does not necessarily translate into nondrugability or on-target toxicity. The difference in expression levels may result in safety margins sufficient enough for development. Determination of a safety window, however, is not a trivial task early in a program. Ideally, it should be determined using appropriate tool compounds (i.e., compounds with adequate pharmacologic activity against and specificity for the target). These tool compounds do not need to have candidate properties as long as sufficient systemic exposure can be achieved for a short repeat-dose study. To control for toxicity related to a chemical class, one should use several tool compounds from different chemical classes and, if available, inactive compounds with close structural similarities (such as inactive enantiomers). In our laboratory, we typically use a 3-day repeat-dose study in male rats using carefully selected tool compounds for target assessment. Doses are selected to result in exposures sufficient to achieve efficacy (as determined by concurrent studies in relevant preclinical efficacy models), but also in higher exposures, so target-related safety margins can be determined. These high exposures are obviously dependent on the characteristics of the tool compounds available (pharmaceutical properties, ADME/PK, and toxicity) and are often limited early in a program. Tissues for gene expression profiling are then selected and prioritized based on the biology of the target (partly generated from the expression tissue maps, but also from the literature), as well as prior evaluation of clinical pathology and histopathology changes. Gene expression changes are evaluated in the context of available reference databases and of our battery of gene expression signatures.

Other approaches have been proposed to proactively evaluate on-target toxicity. These approaches use the same tools as for target validation, such as antibodies, genetically engineered mice, or technologies to modulate mRNA expression levels, such as antisense oligonucleotides, ribozymes, or siRNA [101–103]. In particular, siRNA has gained popularity in the last few years. We have explored the use of siRNA to understand on-target and off-target effects of compounds *in vitro* with mixed results. Clearly, issues such as potency and selectivity of siRNA need to be better understood for this approach to be useful in toxicogenomics. Furthermore, gene silencing approaches induce varying degrees of mRNA and protein downregulation, which may not mimic the pharmacologic inhibition necessary to achieve efficacy. Recent advances in viral and nonviral delivery methods have led to the use of RNA interference for *in vivo* functional genomic studies with successful gene downregulation [104, 105]. Because of the costs of siRNA, these *in vivo* studies are typically conducted in mice, a species for which robust toxicogenomic databases are not available. Furthermore, the required delivery techniques also induce mild adverse effects that complicate the interpretation of gene expression profiles. Our current experience with *in vivo* RNAi studies is too limited to provide a reliable assessment of the potential use of gene silencing to better predict target-related toxicity. However, conceptually, this approach appears promising.

## 8 TOXICOGENOMICS AND IDIOSYNCRATIC TOXICITY

### 8.1 Definition of Idiosyncratic Toxicity

Idiosyncratic toxicity refers to a toxicity not related to the pharmacology of the drug that occurs unexpectedly in a small proportion of treated patients, often in a non-dose-dependent manner [106]. This toxic effect is typically not detected in animals and consequently cannot be predicted during preclinical testing or the early phases of clinical trials. This could be due to host specificity of the toxic reaction or to the insufficient size of preclinical animal studies. Examples of drugs that result in idiosyncratic toxicity include troglitazone (withdrawn in 2000 due to liver toxicity), bromfenac (withdrawn in 1998 due to liver toxicity), fenfluramine (withdrawn in 1997 because of heart valve disease), and cerivastatin (withdrawn in 2001 due to rhabdomyolysis).

Several mechanisms have been proposed to explain the development of idiosyncratic drug toxicity. These include the formation of reactive metabolites in certain individuals due to the presence of polymorphisms in drug-metabolizing enzymes, the development of immune-mediated responses to the drug or one of its metabolites, a synergistic effect of concurrent low-level inflammatory reactions, and changes in mitochondrial function and integrity [107–111]. While substantial experimental evidence exists that these mechanisms occur for certain drugs, there is currently no reliable way to proactively identify compounds that may lead to idiosyncratic responses in a large patient population. Because gene expression profiling allows for a global view of responses occurring simultaneously in cells or tissues, it represents a new approach to study the mechanisms of idiosyncratic reactions and to potentially identify predictive signatures of idiosyncratic toxicity. We illustrate this concept using idiosyncratic hepatotoxicity.

### 8.2 Preclinical Models of Idiosyncratic Toxicity

Several models have been used to study the mechanism of idiosyncratic toxicity. Only a few models are covered here. The brown Norway rat has been commonly used as an animal model of idiosyncratic toxicity, especially for the study of suspected immune-mediated idiosyncratic reactions. For instance, following exposure to nevirapine or D-pencillamine, brown Norway rats develop toxic changes similar to those seen in humans [112, 113]. Another model is a two-hit model for idiosyncratic toxicity, where compounds with potential idiosyncratic liabilities require another underlying factor (such as alcohol intake or concurrent infection) to cause toxicity. This has led to the development of a lipopolysaccharide (LPS)-enhanced toxicity rat model, where rats are coadministered a single dose of compound and LPS. Using this model, several compounds (such as chlorpromazine, ranitidine, and trovafloxacin) known to cause idiosyncratic toxicity in humans have been shown to induce similar changes in rats [114–116]. This LPS-enhanced toxicity rat model has also been investigated at the gene expression level to generate novel hypotheses regarding the exact molecular cascade associated with the toxicity [116, 117].

Idiosyncratic toxicity has also been studied in *in vitro* models. In particular, an assumption is that these types of toxicity are host specific and that, consequently, only the use of human cells will lead to an accurate insight into the molecular and

biochemical events occurring during toxicity. Using this approach, for instance, troglitazone was shown to lead to decreases in cellular ATP and mitochondrial membrane potential in HepG2 cells, and that, consequently, mitochondrial dysfunction may be, at least in part, the cause of the idiosyncratic toxicity induced by troglitazone [118]. Since the liver is often the target organ of idiosyncratic drug reactions, the latter are commonly studied using cultured human hepatocytes or liver-derived cell lines. Although cell lines such as HepG2 can be useful, recent advances in culturing of human hepatocytes have facilitated their use in toxicogenomic studies.

### 8.3 Case Example: Idiosyncratic Hepatotoxicity

As mentioned earlier, the liver is a common target organ of idiosyncratic toxicity and the focus of most research. In particular, the quinolone trovafloxacin has been a tool compound often used in our laboratory. Quinolones are antibacterial agents that act by inhibiting bacterial DNA gyrase and DNA topoisomerase IV [119]. As a class, they are generally well tolerated, except for trovafloxacin [120]. Before its regulatory approval in 1997, trovafloxacin had been tested in over 7000 patients and had not caused any hepatic failures or deaths. Over 2 million people have since received trovafloxacin and 150 cases of liver toxicity have been reported, including 14 cases of acute liver failure. Four patients required liver transplants and an additional five patients died [121]. Because of this hepatotoxicity, severe restrictions were placed on the use of trovafloxacin. The drug can now only be administered in life-threatening situations. The mechanism underlying this adverse effect has not yet been determined.

Our objective was, using human hepatocytes and gene expression profiling, to determine the molecular mechanism of the idiosyncratic toxicity induced by trovafloxacin. Human hepatocytes from four different donors were treated with six quinolone agents (trovafloxacin, levofloxacin, grepafloxacin, gatifloxacin, ciprofloxacin, and clinafloxacin). Using gene expression profiling, trovafloxacin could clearly be distinguished from the other quinolones; the treatment with trovafloxacin resulted in far more gene expression changes than the other compounds [81]. Many of these gene expression changes involved crucial biological pathways that may be involved in the mechanism underlying trovafloxacin-induced hepatotoxicity. In particular, trovafloxacin regulated a number of mitochondrial genes that were not regulated by the other quinolones [81]. In parallel, male Sprague–Dawley rats were treated with levofloxacin (600 mg/kg/day) and trovafloxacin (200 mg/kg/day) for 7 days. Consistent with other measures of toxicity (serum chemistry, histopathology), microarray analysis of the rat livers failed to identify unique gene expression changes induced by trovafloxacin.

Troglitazone is also commonly investigated as a reference compound for idiosyncratic hepatotoxicity. Troglitazone is a thiazolidinedione PPAR-γ agonist that was developed for the treatment of type II diabetes. Troglitazone induced idiosyncratic hepatotoxicity in a small percentage of patients and was removed from the market [106]. In a study similar to ours, human hepatocytes were treated with three thiazolidinedione compounds (troglitazone, rosiglitazone, and pioglitazone). Troglitazone also resulted in a large number of gene expression changes that were not observed with the two other thiazolidinedione compounds [122].

Albeit preliminary, these studies demonstrate two important points. First, the combined use of human hepatocytes and gene expression profiling allowed for the distinction of compounds associated with idiosyncratic hepatotoxicity (trovafloxacin, troglitazone) from compounds of the same chemical classes and not associated hepatotoxicity, suggesting that this *in vitro* system may be appropriate for studying idiosyncratic hepatotoxicity, and potentially for the early identification of safety liabilities. Second, rat repeat-dose studies did not differentiate trovafloxacin from another quinolone known not to induce liver failure in humans, indicating that, at least in the case of the quinolones, the traditional rat model may not be suited to detect potential idiosyncratic toxic liabilities even when coupled with a toxicogenomic evaluation.

## 9 TOXICOGENOMICS IN REGULATORY SUBMISSIONS

### 9.1 Overview of the FDA Pharmacogenomics Guidance

Because of the potential of gene expression profiling to improve the safety assessment of new chemical entities, the FDA issued a guidance in March 2005 for the regulatory submission of pharmacogenomic data (http://www.fda.gov/cder/guidance). This guidance reflects the effort of the FDA to promote the use of genomic technologies in drug development and is also designed to enhance the agency's knowledge of these emerging technologies. In finalizing these guidelines, the FDA has openly cooperated with the various stakeholders and has organized appropriate forums to focus on the major issues and principles that the document should cover. The pharmaceutical industry has welcomed this guidance, as it represents an important stepping stone toward the development of genomics-based drugs and the use of genomics-based safety data. In addition, this guidance provided reassurance to companies that early-stage toxicogenomic experiments would not bring negative regulatory consequences, an important aspect for the wider acceptance of this new technology in the relatively conservative environment of drug safety evaluation.

The guidance clarifies the FDA's policy on the use of pharmacogenomic data in the drug application review process and covers the application of genomics concepts and technologies to nonclinical, clinical pharmacology, and clinical studies. It provides guidelines to sponsors on pharmacogenomic data submission requirements, the format and procedure for data submission, and how the data will be used in regulatory decision making. In general terms, gene expression data for which submission is required include data used for decision making within a specific trial; data used to support scientific arguments about mechanism of action, dose selection, safety, or effectiveness; data that will support registration or labeling language; and data generated on previously validated biomarkers. This guidance demonstrates that the FDA is open to and expects the submission of gene expression profiling data that were generated to support scientific contentions related to toxicity.

The guidance defines pharmacogenomic tests as follows: "An assay intended to study interindividual variations in whole-genome or candidate gene, single-nucleotide polymorphism (SNP) maps, haplotype markers, or alterations in gene expression or inactivation that may be correlated with pharmacological and therapeutic response. In some cases, the pattern or profile of change is the relevant

biomarker, rather than changes in individual markers." This implies that gene expression datasets could ultimately be recognized as validated biomarkers. The guidance also defines "valid biomarkers" and distinguishes between "known valid biomarkers" and "probable valid biomarkers." Valid biomarkers are measured in an analytical test system with well-established performance characteristics, and an established scientific framework or body of evidence exists to understand the significance of the test results. For a known valid biomarker, a widespread agreement exists in the medical or scientific community about the significance of the results. In contrast, for a probable valid biomarker, there is no widespread agreement, but only a scientific framework or body of evidence sufficient to elucidate the significance of the test results. An example would be a biomarker developed by a sponsor and not available for public scientific scrutiny or for independent verification. This distinction also reemphasizes the enormous amount of work and improvement that will be needed in the future to make gene expression data suitable for regulatory decision making. This includes obviously an improved scientific framework for data interpretation through the use of larger, more complete reference databases, but also improved quality control of laboratory procedures, a better understanding of the comparability of different platforms, and some better-defined processes to validate biomarkers.

The FDA recognizes that, currently, most gene expression profiling data are exploratory and would therefore not be required for submission. However, to be prepared to appropriately evaluate future submissions, FDA scientists need to develop an understanding of a variety of relevant scientific issues. The Voluntary Genomic Data Submission (VGDS) provides the material necessary to develop this understanding and is reviewed by a cross-center Interdisciplinary Pharmacogenomic Review Group (IPRG). For more information, the reader is referred to the FDA web site, which reviews the frequently asked questions regarding VGDS (http://www.fda.gov/cder/genomics/FAQ.htm). All VGDS data are protected from disclosure either outside the FDA or to review divisions, are routed directly to the IPRG, and stored on a secured, separate server. These data are not distributed outside the IPRG without the prior agreement of the sponsor and are not to be used for regulatory decision making. The concept of VGDS has in general been well received by the pharmaceutical industry. Voluntary submissions allow sponsors to familiarize FDA scientists with genomic data and their interpretation and, at the same time, to learn about the regulatory decision-making process and expectations involving genomic data. Ultimately, this could prevent delays in future submissions containing required genomic data. So far, many formal submissions have occurred.

### 9.2 Future Impact of Toxicogenomic Data in Regulatory Decision Making

The development of a guidance demonstrates that the FDA is expecting genomic data to become an integral part of the risk assessment of pharmaceuticals. It is, however, difficult at this point to objectively predict the role that a rapidly evolving technology will have in regulatory decision making. When the concept of toxicogenomics first emerged, expectations were high and to some extent unrealistic. While the technology and the development of appropriate analytical tools have considerably improved the ability to generate and interpret large sets of data, it still should

not be considered a mature approach to evaluate toxicology. Furthermore, it is unlikely that toxicogenomics will replace most traditional toxicology studies that are currently part of a regulatory package. Rather, toxicogenomics will more likely complement and increase the value of these studies by providing an improved understanding of the relevance of preclinical toxicologic changes to humans. Toxicogenomics also represents a largely needed novel approach to identify additional biomarkers of safety that could potentially be used to improve monitoring of adverse events in the clinics, resulting in safer clinical trials and potentially earlier identification of outliers with increased sensitivity to particular adverse events. However, it is still unclear how genomics-based biomarkers will be validated to become an integral part of regulatory decision making. The validation of these biomarkers is complex and clearly context specific.

## 10 CONCLUSION

To address the high failure rate due to toxicity, the preclinical toxicologist in the pharmaceutical industry needs to more accurately identify a hazard earlier and provide an improved risk assessment of compounds in discovery and development. In contrast to what was available in the past, toxicogenomics requires a broader and slightly different expertise, often only achieved by multidisciplinary teams composed of toxicologists, molecular biologists, bioinformaticians, and biostatisticians, among others. The formation of productive teams composed of people with strikingly different scientific backgrounds, diverse expertise, and sometimes conflicting interest can be a significant challenge. Properly generated and curated reference databases are also needed to fully exploit the potential benefits of toxicogenomics. Furthermore, major improvements are needed for the microarray technology to become cost effective and meet performance characteristics amenable to its routine implementation in preclinical risk assessment. All these requirements can only be achieved through a substantial investment in human resources and hardware. Most major pharmaceutical companies have committed to significant investments, but so far toxicogenomics still has not been fully integrated in many organizations. This may reflect the current shortage of experts with a pragmatic vision of the future use of this technology, as well as the traditionally conservative nature of toxicology departments in the pharmaceutical industry. Nevertheless, if the trend seen in the last few years continues, it is realistic to predict that molecular toxicology and toxicogenomics will play a growing strategic role in the risk assessment of new chemical entities in the pharmaceutical industry.

## REFERENCES

1. Service RF. Surviving the blockbuster syndrome. *Science* 2004;303:1796–1799.
2. Grabowski HG, Vernon JM. Returns to R&D on new drug introductions in the 1980s. *J Health Econ* 1994;13:383–406.
3. Rawlings MD. Cutting the cost of drug development? *Nat Rev Drug Discov* 2004;3:360–364.

4. Kola I, Landis J. Can the pharmaceutical industry reduce attrition rates? *Nat Rev Drug Discov* 2004;3:711–715.
5. Prentis RA, Lis Y, Walker SR. Pharmaceutical innovation by the seven UK-owned pharmaceutical companies (1964–1985). *Br J Clin Pharmacol* 1988;25:387–396.
6. Ulrich RG, Friend SH. Toxicogenomics and drug discovery: will new technologies help us produce better drugs? *Nat Rev* 2002;1:84–88.
7. MacNeil JS. Genomics goes downstream. *Genome Technol* 2005;54:24–30.
8. Segal E, Friedman N, Kaminski N, Regev A, Koller D. From signatures to models: understanding cancer using microarrays. *Nat Genet* 2005;37(Suppl):S38–S45.
9. Chu TM, Deng S, Wolfinger R, Paules RS, Hamadeh HK. Cross-site comparison of gene expression data reveals high similarity. *Environ Health Perspect* 2004;112:449–455.
10. Shi L, Tong W, Fang H, Scherf U, Han J, Puri RK, et al. Cross-platform comparability of microarray technology: intra-platform consistency and appropriate data analysis procedures are essential. *BMC Bioinformatics* 2005;6(Suppl 2):S12.
11. Yauk CL, Berndt ML, Williams A, Douglas GR. Comprehensive comparison of six microarray technologies. *Nucleic Acids Res* 2004;32:e124.
12. Chuaqui RF, Bonner RF, Best CJ, Gillespie JW, Flaig MJ, Hewitt SM, et al. Post-analysis follow-up and validation of microarray experiments. *Nat Genet* 2002;32(Suppl): 509–514.
13. Baker VA, Harries HM, Waring JF, Duggan CM, Ni HA, Jolly RA, et al. Clofibrate-induced gene expression changes in rat liver: a cross-laboratory analysis using membrane cDNA arrays. *Environ Health Perspect* 2004;112:428–438.
14. Waring JF, Ulrich RG, Flint N, Morfitt D, Kalkuhl A, Staedtler F, et al. Interlaboratory evaluation of rat hepatic gene expression changes induced by methapyrilene. *Environ Health Perspect* 2004;112:439–448.
15. Thomas RS, Rank DR, Penn SG, Zastrow GM, Hayes KR, Hu T, et al. Applications of genomics to toxicology research. *Environ Health Perspect* 2002;110:919–923.
16. Schena M, Shalon D, Davis RW, Brown PO. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* 1995;270:467–470.
17. Nuwaysir EF, Bittner M, Trent J, Barrett JC, Afshari CA. Microarrays and toxicology: the advent of toxicogenomics. *Mol Carcinog* 1999;24:153–159.
18. Yang Y, Blomme EA, Waring JF. Toxicogenomics in drug discovery: from preclinical studies to clinical trials. *Chem Biol Interact* 2004;150:71–85.
19. *Innovation or Stagnation? Challenge and Opportunity on the Critical Path to New Medical Products*. Washington, DC: US Department of Health and Human Services, FDA; 2004.
20. Boess F, Kamber M, Romer S, Gasser R, Muller D, Albertini S, et al. Gene expression in two hepatic cell lines, cultured primary hepatocytes, and liver slices compared to the *in vivo* liver gene expression in rats: possible implications for toxicogenomics use of *in vitro* systems. *Toxicol Sci* 2003;73:386–402.
21. Hong Y, Muller UR, Lai F. Discriminating two classes of toxicants through expression analysis of HepG2 cells with DNA arrays. *Toxicol In Vitro* 2003;17:85–92.
22. Newton RK, Aardema M, Aubrecht J. The utility of DNA microarrays for characterizing genotoxicity. *Environ Health Perspect* 2004;112:420–422.
23. Hamadeh HK, Bushel PR, Jayadev S, Martin K, DiSorbo O, Sieber S, et al. Gene expression analysis reveals chemical-specific profiles. *Toxicol Sci* 2002;67:219–231.
24. Lee J, Richburg JH, Shipp EB, Meistrich ML, Boekelheide K. The Fas system, a regulator of testicular germ cell apoptosis, is differentially up-regulated in Sertoli cell versus germ cell injury of the testis. *Endocrinology* 1999;140:852–858.

25. Fielden MR, Eynon BP, Natsoulis G, Jarnagin K, Banas D, Kolaja KL. A gene expression signature that predicts the future onset of drug-induced renal tubular toxicity. *Toxicol Pathol* 2005;33:675–683.
26. Searfoss GH, Jordan WH, Calligaro DO, Galbreath EJ, Schirtzinger LM, Berridge BR, et al. Adipsin, a biomarker of gastrointestinal toxicity mediated by a functional gamma-secretase inhibitor. *J Biol Chem* 2003;278:46107–46116.
27. Shi L, Tong W, Goodsaid F, Frueh FW, Fang H, Han T, et al. QA/QC: challenges and pitfalls facing the microarray community and regulatory agencies. *Expert Rev Mol Diagn* 2004;4:761–777.
28. Mattes WB, Pettit SD, Sansone SA, Bushel PR, Waters MD. Database development in toxicogenomics: issues and efforts. *Environ Health Perspect* 2004;112:495–505.
29. Shi L, Tong W, Su Z, Han T, Han J, Puri RK, et al. Microarray scanner calibration curves: characteristics and implications. *BMC Bioinformatics* 2005;6(Suppl 2):S11.
30. Churchill GA. Fundamentals of experimental design for cDNA microarrays. *Nat Genet* 2002;32(Suppl):490–495.
31. Higgins MA, Berridge BR, Mills BJ, Schultze AE, Gao H, Searfoss GH, et al. Gene expression analysis of the acute phase response using a canine microarray. *Toxicol Sci* 2003;74:470–484.
32. Sugai T, Kawamura M, Iritani S, Araki K, Makifuchi T, Imai C, et al. Prefrontal abnormality of schizophrenia revealed by DNA microarray: impact on glial and neurotrophic gene expression. *Ann NY Acad Sci* 2004;1025:84–91.
33. Mirnics K, Pevsner J. Progress in the use of microarray technology to study the neurobiology of disease. *Nat Neurosci* 2004;7:434–439.
34. Galvin JE, Ginsberg SD. Expression profiling and pharmacotherapeutic development in the central nervous system. *Alzheimer Dis Assoc Disord* 2004;18:264–269.
35. Todd R, Lingen MW, Kuo WP. Gene expression profiling using laser capture microdissection. *Expert Rev Mol Diagn* 2002;2:497–507.
36. Irwin RD, Boorman GA, Cunningham ML, Heinloth AN, Malarkey DE, Paules RS. Application of toxicogenomics to toxicology: basic concepts in the analysis of microarray data. *Toxicol Pathol* 2004;32:72–83.
37. Hayes KR, Bradfield CA. Advances in toxicogenomics. *Chem Res Toxicol* 2005;18:403–414.
38. Vinciotti V, Khanin R, D'Alimonte D, Liu X, Cattini N, Hotchkiss G, et al. An experimental evaluation of a loop versus a reference design for two-channel microarrays. *Bioinformatics* 2005;21:492–501.
39. Kolaja K, Fielden M. The impact of toxicogenomics on preclinical development: from promises to realized value to regulatory implications. *Preclinica* 2004;2:122–129.
40. Fielden MR, Pearson C, Brennan R, Kolaja KL. Preclinical drug safety analysis by chemogenomic profiling in the liver. *Am J Pharmacogenomics* 2005;5:161–171.
41. Guerreiro N, Staedtler F, Grenet O, Kehren J, Chibout SD. Toxicogenomics in drug development. *Toxicol Pathol* 2003;31:471–479.
42. Ganter B, Tugendreich S, Pearson CI, Ayanoglu E, Baumhueter S, Bostian KA, et al. Development of a large-scale chemogenomics database to improve drug candidate selection and to understand mechanisms of chemical toxicity and action. *J Biotechnol* 2005;119:219–244.
43. Bushel PR, Hamadeh HK, Bennett L, Green J, Ableson A, Misener S, et al. Computational selection of distinct class- and subclass-specific gene expression signatures. *J Biomed Inform* 2002;35:160–170.

44. Ellinger-Ziegelbauer H, Stuart B, Wahle B, Bomann W, Ahr HJ. Characteristic expression profiles induced by genotoxic carcinogens in rat liver. *Toxicol Sci* 2004;77:19–34.
45. Thomas RS, Rank DR, Penn SG, Zastrow GM, Hayes KR, Pande K, et al. Identification of toxicologically predictive gene sets using cDNA microarrays. *Mol Pharmacol* 2001;60:1189–1194.
46. Waring JF, Jolly RA, Ciurlionis R, Lum PY, Praestgaard JT, Morfitt DC, et al. Clustering of hepatotoxins based on mechanism of toxicity using gene expression profiles. *Toxicol Appl Pharmacol* 2001;175:28–42.
47. Waring JF, Cavet G, Jolly RA, McDowell J, Dai H, Ciurlionis R, et al. Development of a DNA microarray for toxicology based on hepatotoxin-regulated sequences. *Environ Health Perspect* 2003;111:863–870.
48. Brazma A, Hingamp P, Quackenbush J, Sherlock G, Spellman P, Stoeckert C, et al. Minimum information about a microarray experiment (MIAME)—toward standards for microarray data. *Nat Genet* 2001;29:365–371.
49. Mattes WB. Annotation and cross-indexing of array elements on multiple platforms. *Environ Health Perspect* 2004;112:506–510.
50. Luhe A, Suter L, Ruepp S, Singer T, Weiser T, Albertini S. Toxicogenomics in the pharmaceutical industry: hollow promises or real benefit? *Mutat Res* 2005;575:102–115.
51. Tennant RW. The national center for toxicogenomics: using new technologies to inform mechanistic toxicology. *Environ Health Perspect* 2002;110:8–10.
52. Waters MD, Fostel JM. Toxicogenomics and systems toxicology: aims and prospects. *Nat Rev Genet* 2004;5:936–948.
53. Castle AL, Carver MP, Mendrick DL. Toxicogenomics: a new revolution in drug safety. *Drug Discov Today* 2002;7:728–736.
54. Searfoss GH, Ryan TP, Jolly RA. The role of transcriptome analysis in pre-clinical toxicology. *Curr Mol Med* 2005;5:53–64.
55. Suter L, Babiss LE, Wheeldon EB. Toxicogenomics in predictive toxicology in drug development. *Chem Biol* 2004;11:161–171.
56. Natsoulis G, El Ghaoui L, Lanckriet GR, Tolley AM, Leroy F, Dunlea S, et al. Classification of a large microarray data set: algorithm comparison and analysis of drug signatures. *Genome Res* 2005;15:724–736.
57. Golub TR, Slonim DK, Gaasenbeek JR, Caligiuri MA, et al. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science* 1999;286:531–537.
58. Tusher VG, Tibshirani R, Chu G. Significance analysis of microarrays applied to the ionizing radiation response. *Proc Natl Acad Sci USA* 2001;98:5116–5121.
59. Khan J, Wei JS, Ringner M, Saal LH, Westermann F, et al. Classification and diagnostic prediction of cancers using gene expression profiling and artificial neural networks. *Nat Med* 2001;7:673–679.
60. Cristianini N, Shawe-Taylor J. *An Introduction to Support Vector Machines*. Cambridge UK: Cambridge University Press; 2000.
61. Furey TS, Cristianini N, Duffy N, Bednarski DW, Schummer M, Haussler D. Support vector machine classification and validation of cancer tissue samples using microarray expression data. *Bioinformatics* 2000;16:906–914.
62. Steiner G, Suter L, Boess F, Gasser R, de Vera MC, Albertini S, et al. Discriminating different classes of toxicants by transcript profiling. *Environ Health Perspect* 2004;112:1236–1248.

63. Thukral SK, Nordone PJ, Hu R, Sullivan L, Galambos E, Fitzpatrick VD, et al. Prediction of nephrotoxicant action and identification of candidate toxicity-related biomarkers. *Toxicol Pathol* 2005;33:343–355.
64. Greim H, Gelbke HP, Reuter U, Thielmann HW, Edler L. Evaluation of historical control data in carcinogenicity studies. *Hum Exp Toxicol* 2003;22:541–549.
65. Kramer JA, Curtiss SW, Kolaja KL, Alden CL, Blomme EA, Curtiss WC, et al. Acute molecular markers of rodent hepatic carcinogenesis identified by transcription profiling. *Chem Res Toxicol* 2004;17:463–470.
66. Ellinger-Ziegelbauer H, Stuart B, Wahle B, Bomann W, Ahr HJ. Comparison of the expression profiles induced by genotoxic and nongenotoxic carcinogens in rat liver. *Mutat Res* 2005;575:61–84.
67. Waring JF, Ulrich RG. The impact of genomics-based technologies on drug safety evaluation. *Annu Rev Pharmacol Toxicol* 2000;40:335–352.
68. Cattley RC. Peroxisome proliferators and receptor-mediated hepatic carcinogenesis. *Toxicol Pathol* 2004;32(Suppl 2):6–11.
69. Holden PR, Tugwood JD. Peroxisome proliferator-activated receptor alpha: role in rodent liver cancer and species differences. *J Mol Endocrinol* 1999;22:1–8.
70. Hamadeh HK, Bushel PR, Jayadev S, DiSorbo O, Bennett L, Li L, et al. Prediction of compound signature using high density gene expression profiling. *Toxicol Sci* 2002; 67:232–240.
71. Kramer JA, Blomme EA, Bunch RT, Davila JC, Jackson CJ, Jones PF, et al. Transcription profiling distinguishes dose-dependent effects in the livers of rats treated with clofibrate. *Toxicol Pathol* 2003;31:417–431.
72. Badr MZ, Belinsky SA, Kauffman FC, Thurman RG. Mechanism of hepatotoxicity to periportal regions of the liver lobule due to allyl alcohol: role of oxygen and lipid peroxidation. *J Pharmacol Exp Ther* 1986;238:1138–1142.
73. Butterworth KR, Carpanini FM, Dunnington D, Grasso P, Pelling D. The production of periportal necrosis by allyl alcohol in the rat. *Br J Pharmacol* 1978;63:353P–354P.
74. Amin K, Ip C, Jimenez L, Tyson C, Behrsing H. *In vitro* detection of differential and cell-specific hepatobiliary toxicity induced by geldanamycin and 17-allylaminogeldanamycin using dog liver slices. *Toxicol Sci* 2005;87:442–450.
75. Waring JF, Ciurlionis R, Jolly RA, Heindel M, Ulrich RG. Microarray analysis of hepatotoxins *in vitro* reveals a correlation between gene expression profiles and mechanisms of toxicity. *Toxicol Lett* 2001;120:359–368.
76. Burczynski ME, McMillian M, Cirvo J, Li L, Parker JB, Dunn RT, et al. Toxicogenomics-based discrimination of toxic mechanism in HepG2 human hepatoma cells. *Toxicol Sci* 2000;58:399–415.
77. Gomez-Lechon MJ, Ponsoda X, Bort R, Castell JV. The use of cultured hepatocytes to investigate the metabolism of drugs and mechanisms of drug hepatotoxicity. *Altern Lab Anim* 2001;29:225–231.
78. Waring JF, Ciurlionis R, Jolly RA, Heindel M, Gagne G, Fagerland JA, et al. Isolated human hepatocytes in culture display markedly different gene expression patterns depending on attachment status. *Toxicol In Vitro* 2003;17:693–701.
79. Li AP, Reith MK, Rasmussen A, Gorski JC, Hall SD, Xu L, et al. Primary human hepatocytes as a tool for the evaluation of structure–activity relationship in cytochrome P450 induction potential of xenobiotics: evaluation of rifampin, rifapentine and rifabutin. *Chem Biol Interact* 1997;107:17–30.
80. Ulrich RG, Bacon JA, Cramer CT, Peng GW, Petrella DK, Stryd RP, et al. Cultured hepatocytes as investigational models for hepatic toxicity: practical applications in drug discovery and development. *Toxicol Lett* 1995;82–83:107–115.

81. Liguori MJ, Anderson LM, Bukofzer S, McKim J, Pregenzer JF, Retief J, et al. Microarray analysis in human hepatocytes suggests a mechanism for hepatotoxicity induced by trovafloxacin. *Hepatology* 2005;41:177–186.
82. de Longueville F, Surry D, Meneses-Lorente G, Bertholet V, Talbot V, Evrard S, et al. Gene expression profiling of drug metabolism and toxicology markers using a low-density DNA microarray. *Biochem Pharmacol* 2002;64:137–149.
83. Harries HM, Fletcher ST, Duggan CM, Baker VA. The use of genomics technology to investigate gene expression changes in cultured human cells. *Toxicol In Vitro* 2001;15:399–405.
84. Morgan KT, Ni H, Brown HR, Yoon L, Crosby LM, et al. Application of cDNA microarray technology to *in vitro* toxicology and the selection of genes for a real-time RT-PCR-based screen for oxidative stress in Hep-G2 cells. *Toxicol Pathol* 2002;30:435–451.
85. Aubrecht J, Caba E. Gene expression profile analysis: an emerging approach to investigate mechanisms of genotoxicity. *Pharmacogenomics* 2005;6:419–428.
86. Snyder RD, Green JW. A review of the genotoxicity of marketed pharmaceuticals. *Mutat Res* 2001;488:151–169.
87. Dickinson DA, Warnes GR, Quievryn G, Messer J, Zhitkovich A, Rubitski E, et al. Differentiation of DNA reactive and non-reactive genotoxic mechanisms using gene expression profile analysis. *Mutat Res* 2004;549:29–41.
88. Yang Y, Abel S, Ciurlionis R, Waring JF. Development of gene expression-based *in vitro* assays for the efficient toxicity characterization of compounds. *Pharmacogenomics* 2006;7:177–186.
89. Waring JF, Gum R, Morfitt D, Jolly RA, Ciurlionis R, Heindel M, et al. Identifying toxic mechanisms using DNA microarrays: evidence that an experimental inhibitor of cell adhesion molecule expression signals through the aryl hydrocarbon nuclear receptor. *Toxicology* 2002;181–182:537–550.
90. Denison MS, Nagy SR. Activation of the aryl hydrocarbon receptor by structurally diverse exogenous and endogenous chemicals. *Annu Rev Pharmacol Toxicol* 2003;43:309–334.
91. Milano J, McKay J, Dagenais C, Foster-Brown L, Pognan F, Gadient R, et al. Modulation of notch processing by gamma-secretase inhibitors causes intestinal goblet cell metaplasia and induction of genes known to specify gut secretory lineage differentiation. *Toxicol Sci* 2004;82:341–358.
92. Creasy DM. Evaluation of testicular toxicity in safety evaluation studies: the appropriate use of spermatogenic staging. *Toxicol Pathol* 1997;25:119–131.
93. Stewart J. Inhibin B as a potential biomarker of testicular toxicity. *Toxicologist* 2005; 74:6.
94. Richburg JH, Johnson KJ, Schoenfeld HA, Meistrich ML, Dix DJ. Defining the cellular and molecular mechanisms of toxicant action in the testis. *Toxicol Lett* 2002; 135:167–183.
95. Adachi T, Ono Y, Koh KB, Takashima K, Tainaka H, Matsuno Y, et al. Long-term alteration of gene expression without morphological change in testis after neonatal exposure to genistein in mice: toxicogenomic analysis using cDNA microarray. *Food Chem Toxicol* 2004;42:445–452.
96. Adachi T, Koh KB, Tainaka H, Matsuno Y, Ono Y, Sakurai K, et al. Toxicogenomic difference between diethylstilbestrol and 17beta-estradiol in mouse testicular gene expression by neonatal exposure. *Mol Reprod Dev* 2004;67:19–25.
97. Rockett JC, Christopher LJ, Brian GJ, Krawetz SA, Hughes MR, Hee KK, et al. Development of a 950-gene DNA array for examining gene expression patterns in mouse testis. *Genome Biol* 2001;2:14.1–14.9.

98. Zappa F, Ward T, Pedrinis E, Butler J, McGown A. NAD(P)H: quinone oxidoreductase 1 expression in kidney podocytes. *J Histochem Cytochem* 2003;51:297–302.
99. Su AI, Wiltshire T, Batalov S, Lapp H, Ching KA, Block D, et al. A gene atlas of the mouse and human protein-encoding transcriptomes. *Proc Natl Acad Sci USA* 2004;101:6062–6067.
100. Su AI, Cooke MP, Ching KA, Hakak Y, Walker JR, Wiltshire T, et al. Large-scale analysis of the human and mouse transcriptomes. *Proc Natl Acad Sci USA* 2002;99:4465–4470.
101. Honore P, Kage K, Mikusa J, Watt AT, Johnston JF, Wyatt JR, et al. Analgesic profile of intrathecal P2X(3) antisense oligonucleotide treatment in chronic inflammatory and neuropathic pain states in rats. *Pain* 2002;99:11–19.
102. Semizarov D, Frost L, Sarthy A, Kroeger P, Halbert DN, Fesik SW. Specificity of short interfering RNA determined through gene expression signatures. *Proc Natl Acad Sci USA* 2003;100:6347–6352.
103. Zambrowicz BP, Turner CA, Sands AT. Predicting drug efficacy: knockouts model pipeline drugs of the pharmaceutical industry. *Curr Opin Pharmacol* 2003;3:563–570.
104. Lu PY, Xie F, Woodle MC. *In vivo* application of RNA interference: from functional genomics to therapeutics. *Adv Genet* 2005;54:117–142.
105. Li L, Lin X, Staver M, Shoemaker A, Semizarov D, Fesik SW, et al. Evaluating hypoxia-inducible factor-1alpha as a cancer therapeutic target via inducible RNA interference *in vivo*. *Cancer Res* 2005;65:7249–7258.
106. Waring JF, Anderson MG. Idiosyncratic toxicity: mechanistic insights gained from analysis of prior compounds. Curr *Opin Drug Discov Dev* 2005;8:59–65.
107. Williams DP, Park BK. Idiosyncratic toxicity: the role of toxicophores and bioactivation. *Drug Discov Today* 2003;8:1044–1050.
108. Knowles SR, Uetrecht J, Shear NH. Idiosyncratic drug reactions: the reactive metabolite syndrome. *Lancet* 2000;356:1587–1591.
109. Uetrecht J. Prediction of a new drug's potential to cause idiosyncratic reactions. *Curr Opin Drug Discov Dev* 2001;4:55–59.
110. Roth RA, Luyendyk JA, Maddox JF, Ganey PE. Inflammation and drug idiosyncrasy—is there a connection? *J Exp Ther* 2003;307:1–8.
111. Park BK, Kitteringham NR, Powell H, Pirmohamed M. Advances in molecular toxicology—towards understanding idiosyncratic drug toxicity. *Toxicology* 2000;153:39–60.
112. Shenton JM, Teranishi M, Abu-Asab MS, Yager JA, Uetrecht JP. Characterization of a potential animal model of an idiosyncratic drug reaction: nevirapine-induced skin rash in the rat. *Chem Res Toxicol* 2003;16:1078–1089.
113. Masson MJ, Uetrecht JP. Tolerance induced by low dose D-penicillamine in the brown Norway rat model of drug-induced autoimmunity is immune-mediated. *Chem Res Toxicol* 2004;17:82–94.
114. Luyendyk JP, Maddox JF, Cosma GN, Ganey PE, Cockerell GL, Roth RA. Ranitidine treatment during a modest inflammatory response precipitates idiosyncrasy-like liver injury in rats. *J Pharmacol Exp Ther* 2003;307:9–16.
115. Buchweitz JP, Ganey PE, Bursian SJ, Roth RA. Underlying endotoxemia augments toxic responses to chlorpromazine: is there a relationship to drug idiosyncrasy? *J Pharmacol Exp Ther* 2002;300:460–467.
116. Waring JF, Liguori MJ, Luyendyk JP, Maddox JF, Ganey PE, Stachlewitz RF, et al. Microarray analysis of LPS potentiation of trovafloxacin-induced liver injury in rats suggests a role for proinflammatory chemokines and neutrophils. *J Pharmacol Exp Ther* 2006;316:1080–1087.

117. Luyendyk JP, Mattes WB, Burgoon LD, Zacharewski TR, Maddox JF, Cosma GN, et al. Gene expression analysis points to hemostasis in livers of rats cotreated with lipopolysaccharide and ranitidine. *Toxicol Sci* 2004;80:203–213.
118. Tirmenstein MA, Hu CX, Gales TL, Maleeff BE, Narayanan PK, Kurali E, et al. Effects of troglitazone on HepG2 viability and mitochondrial function. *Toxicol Sci* 2002;69:131–138.
119. Drlica K, Zhao X. DNA gyrase, topoisomerase IV and the ;4-quinolones. *Microbiol Mol Biol Rev* 1997;61:377–392.
120. Ball P, Mandell L, Niki Y, Tillotson G. Comparative tolerability of the newer fluoroquinolone antibacterials. *Drug Safety* 1999;21:407–421.
121. Bertino J, Fish D. The safety profile of the fluoroquinolones. *Clin Ther* 2000;22:798–817.
122. Kier LD, Neft R, Tang L, Suizu R, Cook T, Onsurez K, et al. Applications of microarrays with toxicologically-relevant genes (tox genes) for the evaluation of chemical toxicants in Sprague Dawley rats *in vivo* and human hepatocytes *in vitro*. *Mutat Res* 2004;549: 101–113.

# TOXICOPROTEOMICS: PRECLINICAL STUDIES

B. Alex Merrick and Maribel E. Bruno

*National Center for Toxicogenomics, National Institute of Environmental Health Sciences, Research Triangle Park, North Carolina*

**Contents**

1 Introduction
2 The Field of Toxicoproteomics
3 Disciplines in Proteomics for Toxicoproteomic Research
4 Proteomic Platforms for Toxicoproteomic Studies
  4.1 Two-Dimensional Gel Electrophoresis and Mass Spectrometry
  4.2 Multidimensional, Quantitative LC-MS/MS: MuDPIT, ICAT, iTRAQ, and SILAC
  4.3 Retentate Chromatography–Mass Spectrometry: SELDI
  4.4 Protein Capture Arrays: Antibody Arrays
5 Toxicoproteomic Studies in Liver Injury
  5.1 Acetaminophen
  5.2 Carbon Tetrachloride
  5.3 Bromobenzene
  5.4 Wyeth 14643
  5.5 Hydrazine
  5.6 Thioacetamide
6 New Blood Biomarkers in Liver Injury
7 Toxicoproteomic Studies in Kidney Injury
  7.1 Cyclosporin A
  7.2 Puromycin and Gentamicin
  7.3 4-Aminophenol, D-Serine, and Cisplatin
  7.4 Dichlorovinyl-L-Cysteine
8 Toxicoproteomics of Adverse Effects and Toxicants: Citation Analysis

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

9 Future Developments in Toxicoproteomics
9.1 Toxicoproteomic Studies in Other Organs and Tissues
9.2 Expectations and Reality
9.3 Future Trends and Tools in Toxicoproteomics
10 Conclusion
Acknowledgments
References

## 1 INTRODUCTION

The primary purposes of this chapter are to define the field of toxicoproteomics, describe its development among "omics" technologies, and demonstrate its utility in experimental toxicological research, preclinical testing, and emerging applications to medicine. Subdisciplines within the field of proteomics deployed in preclinical research are described as Tier I and Tier II proteomic analysis. Tier I analysis involves global protein mapping and protein profiling for differential expression. Tier II proteomic analysis includes global methods for description of function, structure, interactions, and posttranslational modification of proteins. Preclinical toxicoproteomic studies with model liver and kidney toxicants are critically reviewed to assess their contributions toward understanding pathophysiology and in biomarker discovery. Toxicoproteomic studies in other organs and tissues are briefly discussed as well. A literature citation analysis matching proteomic and toxicity publications reveals that most toxicoproteomic research involves Tier I proteomic analysis and that biomarker development is a primary focus of many studies. The final section discusses five key developments in toxicoproteomics, encompassing new tools and research focus areas for the field.

## 2 THE FIELD OF TOXICOPROTEOMICS

Toxicoproteomics is the application of global protein measurement technologies to toxicology research. Chemical exposure frequently modifies proteins directly, produces changes in protein expression, and dysregulates critical biological pathways or processes leading to toxicity. Primary aims in toxicoproteomics are the discovery of key modified proteins, the determination of affected pathways, and the development of biomarkers for eventual prediction of toxicity. A growing base of pharmacogenomic knowledge should also strengthen the field of preclinical assessment and predictive toxicology [1–4].

The wide application of proteomics has generated great interest and enthusiasm in many established scientific disciplines in basic biology and medicine, including the field of toxicology. There are well over 10,000 publications relating to some aspect or application of proteomics in the biosciences. However, the number of published proteomic studies is quite limited in reporting primary data for drug-mediated adverse reactions, biochemical toxicities, or undesirable phenotypes [5–7].

Compared to toxicogenomics or those technologies using DNA microarrays to study the effects of toxic substances, preclinical toxicoproteomics is indeed a smaller enterprise.

Fundamental reasons for the emerging state of preclinical toxicoproteomics compared to other omics fields relate to DNA technologies' ability to exploit the knowledge of whole genome sequence and apply it in mass parallel formats. DNA slides spotted with synthetic oligonucleotides can essentially hybridize and measure expression of every gene in human and many preclinical species. DNA microarrays are the most common transcriptomic format compared to other powerful gene expression methods like SAGE (serial analysis of gene expression) and MPSS (massively parallel signature sequencing). DNA microarray platforms provide high data output and rapid sample throughput in a way that proteomics cannot yet match in providing a "protein for every gene." Thus, as a hybrid "technofield" of proteomics and toxicology, toxicoproteomics is still becoming established as a specialized subject area. For the moment, toxicoproteomics has been placed under the aegis of toxicogenomics [8] or mass spectrometry [9].

The capabilities and critical issues for toxicogenomics and toxicoproteomics in pharmaceutical data submission for nonclinical safety testing to regulatory agencies have recently been reviewed [10]. Some of the unresolved issues raised for regulatory omic data submission include proper data quality standards, disparate platforms and data formats, meaningful data validation, relevance to traditional toxicological endpoints, potential added-value to established biochemical and molecular methods, animal-to-human extrapolation, mechanism of action, impact on the NOAEL, early versus adaptive, nonpharmacologic responses, limitations of bioinformatics algorithms, tools and available databases, and criteria for how genomic and proteomic data would influence regulatory decisions.

Some reviews in the field have closely scrutinized toxicoproteomics and metabonomics for their contributions in biomarker discovery beyond traditional clinical chemistry and histopathology indicators. A review of thirteen toxicoproteomic and metabonomic studies with various nephrotoxic agents examined various endpoints for their respective abilities to determine specificity and sensitivity of renal toxicity. The review concluded that proteomic (and metabonomic) data compared very poorly with traditional methods of blood and urine chemistries and histopathology without significant improvements [11]. However, it is the potential for discovery and new insights into pathobiology and therapeutics that fuels interest in omics technologies. The path to discovery is unfortunately seldom linear or formulaic. For example, progress in cystic kidney disorders leveraged prior proteomic mapping studies [12] to show MSK1 mutations underlie ciliary dysfunction in embryonic development [13]. Apparently, normal ciliary function is critical in renal tubule development and its dysfunction produces several forms of polycystic kidney disease. Clearly, the discovery potential of proteomics is a major driver in the push for better biomarkers of injury and disease.

Despite the discovery potential of toxicoproteomics, the strategies for conducting proteomic analysis and using such data in drug development, preclinical safety, and regulatory submission are far from standardized. The complexity of protein expression, multiple technology platforms, and emerging technical standards are major challenges for continued growth of the field. Researchers are also finding that no one platform is best suited for toxicoproteomic research.

One of the major goals for toxicoproteomics is to translate identified protein changes into new biomarkers and signatures of chemical toxicity that can provide much greater definition than current indicators [7]. The imprecise meaning of the term "biomarker" accounts for its wide variation in use (and misuse) in scientific and regulatory communities [14, 15]. At a biochemical and molecular level, biomarkers can be narrowed down to "singular biological measures with reproducible evidence of a clear association with health, disease, adverse effect or toxicity." This is a necessarily limited definition for quantitative biochemical or molecular measures. Historical and more current examples of biomarkers are the detection of a single *protein* such as C-reactive protein in cardiovascular disease [16], an *enzyme activity* like alanine aminotransferase activity in liver injury [17], *gene transcription* products such as Her2/neu [18] in breast malignancies, gene mutations/polymorphisms like slow acetylators that affect xenobiotic metabolism [19], or small molecules/metabolites such as serum glucose, insulin, and urinary ketone bodies in pathologic or drug-induced diabetes.

Many subcategories of biomarkers are in popular usage as well, including biologic, surrogate, prognostic, diagnostic, and bridging biomarkers [15, 20]. Importantly, a major development of the large datasets derived from omic technologies is the possibility of greater molecular topography compared to a singular biomarker. One of the major tenets of toxicoproteomics and other omic analyses is that specific patterns of protein changes can comprise a consistent "signature" of toxicity [21] or "combinatorial biomarker" [22] that is robust enough to be observed in spite of variations in biology, experimental design, or technology platforms. This is a critical assumption—first, because there is great potential for including nonspecific or indirect protein changes in such a signature, and second, because of the inherent challenges in establishing a causal linkage of multiple protein changes to a toxic or adverse phenotype. Discovery and definition of such toxicity signatures are at an early stage of progress in the field of toxicoproteomics [6, 21].

## 3 DISCIPLINES IN PROTEOMICS FOR TOXICOPROTEOMIC RESEARCH

Proteomics in a global protein analysis mode generally links separation and identification technologies to create a protein profile or differential protein display. Although the focus of proteomics has been grouped in various ways, Fig. 1 shows representative subdisciplines of proteomics that provide a means to categorize toxicoproteomic research.

Four factors often shape the manner in which researchers pursue their activities in toxicoproteomics that include (1) the complex nature of proteins, (2) that portion of the proteome targeted for study, (3) the integrative relationship of toxicoproteomic studies with other omic technologies, and (4) the driving forces behind specific toxicoproteomic projects.

Each of these four factors should be considered. First, a primary objective in proteomics is the isolation and identification of individual proteins from complex biological matrices. In toxicoproteomic analysis, the first tier of proteomic analysis is to determine individual *protein identities* (fingerprint, AA sequence), their relative (or absolute) *quantities*, and their *spatial location* within cell(s), tissues, and biofluids

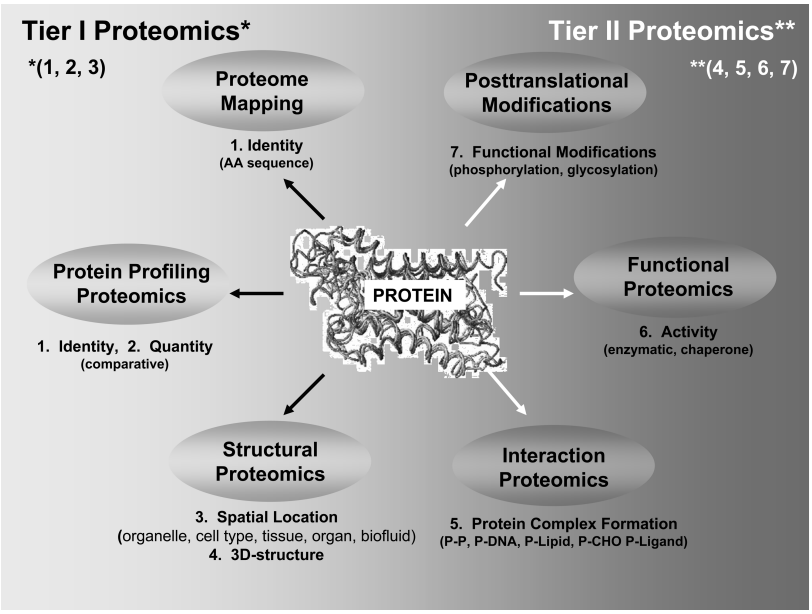


**FIGURE 1** Disciplines of proteomics for application to preclinical drug assessment. Major areas of proteomics attempt to describe some portion of seven protein attributes (1–7) in a global manner. Tier I proteomic analysis (1–3) involves identification and quantitation of all proteins from a defined space for protein mapping or profiling. Tier II proteomic analysis (4–7) involves determination of three-dimensional structures, protein complexes, functions, or posttranslational modifications of proteins in a global manner.

of interest. The second tier of analysis globally screens for protein *functions, protein interactions*, three-dimensional *structure*, and specific *posttranslational modifications*. These first and second tiers of proteomic analysis encompass the seven intrinsic attributes of proteins that factor into toxicoproteomic analysis [23] as shown in Fig. 1. Proteomic platforms vary greatly in their respective abilities to deliver data on all protein attributes simultaneously during one analysis.

Proteome mapping is the most descriptive of proteomic inquiries and usually focuses on identifying all proteins in the sample or at a cellular location at hand. Profiling experiments necessarily require quantitation (relative to control, or absolute) to be comparative among samples. Implicit in protein mapping and profiling are considerations about the spatial "origin" of the sample. Often, sample origins are the same in profiling experiments for comparability; for example, serum samples are most comparable to serum, livers to livers, and so on. Structural proteomics is usually defined as high-throughput determination of protein structures in three-dimensional space and is often determined by X-ray crystallography and NMR spectroscopy. This definition has been expanded in Fig. 1 to include spatial location of proteins within the organism rather than continue to divide proteomic fields by specific levels of protein organization that might range from subcellular, cellular, organ, tissue, or organism to species proteomics.

A second factor to consider in toxicoproteomics is that the proteomes of most cells, tissues, and organs are so vast that, unlike whole genome queries, proteomes

cannot be completely analyzed by existing proteomic platforms. By default, toxicoproteomic studies most often analyze only a portion of the proteome contained in typical biological samples. A frequent strategy to broaden protein coverage is to take steps prior to analysis to reduce sample complexity (analyze a portion of the proteome or "subproteome") by such procedures as subcellular fractionation, affinity or adsorptive chromatography, or electrophoretic separation.

Third, toxicoproteomic analysis may be conducted as an independent activity or alternately as a component of a large, formalized gene expression project for which the study design, type of experimental subjects, and the availability or amount of biological specimens may greatly impact sample preparation procedures and proteomic platform selection [24].

Fourth, the forces and individuals driving toxicoproteomic studies such as drug discovery, biophysical and chemical analyses, safety assessment, drug efficacy, ADME properties, and clinical trials will greatly influence the study design, analysis, and, importantly, biological interpretation of toxicoproteomic data.

## 4 PROTEOMIC PLATFORMS FOR TOXICOPROTEOMIC STUDIES

The complexity of a "proteome" contained in a biological sample presents numerous challenges for comprehensively describing the seven attributes of protein expression during any single proteomic analysis [7]. The primary aims of proteomic analysis are to (1) achieve maximal coverage of the proteome (i.e., Tier I analysis) in each sample, (2) complete analysis at high throughput, (3) produce an accurate quantitative protein measurement, (4) deliver data and interpretable results in a timely period, and (5) use discovery-oriented, open platforms.

All proteomic platforms typically share two common capabilities: a means of global *separation* and a technology for *identification* of proteins. Identification usually means assignment to international gene (i.e., NCBI) or protein (i.e., Uniprot or SwissProt) identification numbers. The following proteomic platforms represent a brief description of the principal technologies used for separating and identifying proteins during toxicoproteomic studies.

### 4.1 Two-Dimensional Gel Electrophoresis and Mass Spectrometry

Two-dimensional (2D) gel electrophoresis systems have been combined with mass spectrometry in an established and adaptable platform since 1975 and it is the most commonly used proteomic platform to separate and comparatively quantitate protein samples [25]. Current state-of-the-art 2D gels use immobilized pH gradient (IPG) gels first to separate proteins by charge and then subsequently to resolve by mass using SDS gel electrophoresis for effective separation of complex protein samples. Proteins are sufficiently separated to homogeneity by 2D gels to permit mass spectrometry (MS) identification. Typical IPG gels at 18–24 cm fitted with similarly sized SDS-PAGE gels can separate between 2000 and 3000 proteins. Each spot represents not a unique protein (i.e., gene product) but often posttranslationally modified forms of the same protein. Fluorescent staining is often the most sensitive means of protein detection (nanogram to microgram range). After electronic alignment (registration) of stained proteins in 2D gels by image analysis software, intensities of identical protein spots are compared among treatment groups and a

ratio (fold change) is calculated for each protein using specialized software. In DIGE 2D-MS (*di*fferential *g*el *e*lectrophoresis two-dimensional gel mass spectrometry), protein samples to be compared are labeled with Cy2-, Cy3-, or Cy5-based linkers. Labeled samples are mixed together and electrophoresed on the same gel. This procedure minimizes image analysis errors from trying to electronically register different gels, since each dye (sample) is read at a different wavelength on the same gel [26]. Up to three or four samples can be run on the same 2D gel.

The combination of 2D gel separation of proteins with mass spectrometry provides a ready means of protein identification after protein excision, enzymatic digestion, and MS analysis. The 2D-MS platform forms a versatile and discovery-oriented standardized approach for use in toxicoproteomic studies [27]. A downside to this platform is the limited coverage of a proteome that can be realized on 2D gels by even the most sensitive fluorescent stains.

One-dimensional gel-based proteomic platforms—1D gel LC-MS/MS—may also be extremely effective for protein separation and identification using SDS-PAGE only (i.e., mass separation) with specially preprocessed samples such as immunodepleted plasma [28] or cell secretomes [29]. Such preprocessing sufficiently reduces the original protein complexity to allow small amounts of sample protein (μg) or serum (μL) to be resolved to near protein homogeneity in stained protein bands. Bands are enzymatically digested to obtain diagnostic peptides for protein identification after amino acid sequencing by LC-MS/MS.

### 4.2 Multidimensional, Quantitative LC-MS/MS: MuDPIT, ICAT, iTRAQ, and SILAC

Multidimensional liquid chromatography (LC-LC) is used to separate protein digests (ng–μg peptides) by charge (strong anion exchange) and hydrophobicity (C18) immediately prior to entry into a tandem mass spectrometer (MS/MS) for protein identification [30]. A premier representative of LC-MS/MS proteomics is the multidimensional protein identification technology or "MuDPIT" platform. This approach has also been called "shotgun proteomics" since entire protein lysates are trypsin digested into thousands of peptide fragments without the need for any fractionation or processing prior to LC-LC separation and MS/MS identification. Advantages of this newer platform are the potential for detection and identification of low abundance proteins that may not be observed in gel-based protein separations. However, the MuDPIT platform is only semiquantitative. The platform is very effective in proteomic mapping and discovery studies and should find great utility in toxicoproteomics.

Other variations on the LC-MS/MS approach, closely linking LC separation to tandem MS instruments, have incorporated isotopic labeling strategies for protein quantitation and in-depth proteomic profiling of samples. Examples of such platforms are ICAT (isotope coded affinity tags), iTRAQ (isobaric tag for relative and absolute quantitation), and SILAC (stable isotope labeling with amino acids in cell culture). These methods use "light" and "heavy" forms of isotopes in linkers that bind to functional groups of proteins (i.e., cysteines or amino groups) in lysates. SILAC and iTRAQ are particularly effective for metabolic incorporation of "light" and "heavy" forms of amino acids (i.e., $^{1}H:^{2}H$, $^{12}C:^{13}C$, $^{14}N:^{15}N$) into cellular proteins during cell culture incubations. Although sample throughput is slow and analysis time is lengthy, the protein coverage has been greatly expanded with the develop-

ment of these new multidimensional proteomic platforms. Careful sample, dose, and time selection to a few samples appears to be a successful strategy in achieving the most value from multidimensional, quantitative LC-MS/MS platforms.

### 4.3 Retentate Chromatography–Mass Spectrometry: SELDI

Retentate chromatography–mass spectrometry (RC-MS) is a high-throughput proteomic platform that creates a laser-based mass spectrum from a chemically absorptive surface. The principle of this approach is the adsorptive retention (pg–ng protein) of a subset of sample proteins on a thin, chromatographic support (i.e., hydrophobic, normal phase, weak cation exchange, strong anion exchange, or immobilized metal affinity supports). The adsorptive surfaces are placed on thin metal chips, which can be inserted into a specially modified MALDI-type mass spectrometer. The laser rapidly desorbs proteins from each sample on a metal chip to create a mass spectrum profile.

RC-MS can be performed on any protein sample but thus far this platform has found the greatest utility in the analysis of serum and plasma for disease biomarker discovery [6, 31]. The lead commercial platform of the RC-MS proteomic platform is the SELDI-TOF-MS instrument or, *s*urface-*e*nhanced *l*aser *d*esorption *i*onization *t*ime-*o*f-*f*light [32]. Analysis of samples is relatively rapid (100/day). Only a few microliters of biofluid sample are necessary and hundreds of samples can be screened in a few days. Downsides are that only a fraction of the proteome is analyzed (what adsorbs to the particular chemical surface), there are sample reproducibility issues, and protein identification of peaks is not readily achieved without additional analysis [33]. However, the RC-MS approach fits many problem areas as a proteomic discovery tool for defining drug or chemical exposure when rapid screening is needed for hundreds or thousands of preclinical or clinical samples.

### 4.4 Protein Capture Arrays: Antibody Arrays

Protein capture arrays (any mass parallel array of proteins, peptides, capture ligands, or adsorptive surfaces for protein analysis) represent a promising new proteomic tool that closely emulates the design for parallel analysis of DNA microarray technology [34]. Many different types of capture molecules can be arrayed (recombinant proteins, aptamers, peptides, drug libraries) but the most prevalent are antibody arrays that directly separate proteins from each other by affinity binding to specific protein targets. Generally, commercial antibody array platforms have widely varying sensitivities (pg–μg peptides) that fall into three classes based on targeted proteins: cytokine/chemokine arrays, cellular function protein arrays, and cell signaling arrays. However, antibody arrays are not presently available for any given cell type, biofluid, or species. This platform provides a rapid screen for limited sets of proteins that may fit some applications in toxicoproteomics.

## 5 TOXICOPROTEOMIC STUDIES IN LIVER INJURY

The liver is the primary site of xenobiotic metabolism and is a major organ for biotransformation and elimination of pharmaceuticals from the body [35]. Many of the

initial studies in toxicoproteomics have been performed to gain insight into drug-induced liver injury and to test the capabilities of proteomic analysis with well studied drugs and chemicals in rodent models of toxicity. While rodent models of liver toxicity may predominantly show one phenotype such as necrosis, hepatitis, cholestasis, steatosis, fibrosis, cirrhosis, or malignancy, many different cellular and molecular processes are ongoing [36].

The removal from the marketplace of several widely prescribed drugs due to hepatotoxicity has attracted considerable attention and there are active efforts to better understand and identify hepatic risk prior to drug approval. Factors affecting susceptibility to drug-induced injury include age, sex, drug–drug interactions, and genetic polymorphism in metabolic pathways involved in activation or disposition of therapeutic drugs [37]. Reactive intermediates produced during metabolism can be toxic or some compounds may dysregulate critical biochemical pathways or functions of the liver [35].

Acute hepatic necrosis is a reproducible phenotype characteristic of many types of drugs and model compounds in preclinical species. The above considerations provide a rationale for exploring and testing the capabilities of emerging omics technologies like proteomics and transcriptomics upon acute hepatic injury. Which agents might be worthwhile for toxicoproteomic studies? A recent toxicogenomic study for classifying hepatotoxicants evaluated a representative list of 25 well-known model compounds or substances showing hepatotoxicity during testing [38]. The aim of this preclinical research report was to determine if biological samples from rats treated with various compounds could be classified based on gene expression profiles. Such model agents causing acute hepatonecrotic injury included acetaminophen, carbon tetrachloride, bromobenzene, hydrazine, and others. Hepatic gene expression profiles were analyzed using a supervised learning method (support vector machines—SVMs) to generate classification rules. The SVM method was combined with recursive feature elimination to improve classification performance. The goal was to identify a compact subset of probe sets (transcripts) with potential use as biomarkers. DNA microarray data have been generated for each substance in this study [38]. Their list of representative hepatotoxic agents for preclinical testing served as a basis for examining the literature for corresponding toxicoproteomic studies.

Table 1 summarizes available primary data from toxicoproteomic studies. Generally, these studies have been conducted on representative, model liver and kidney damaging agents relevant to preclinical assessment of toxicity. The agent, proteomic analysis platform, tissue or preparation, and brief results for each study are summarized in the table. Liver toxicants are addressed first.

### 5.1 Acetaminophen

Acetaminophen has been one of the most commonly tested agents for inducing hepatic injury in toxicoproteomic studies of the liver. It produces centrilobular hepatic necrosis in most preclinical species. Acute hepatocellular injury from acetaminophen (*N*-**a**cetyl-**p**ara-**a**mino**p**henol—APAP) exposure is primarily initiated by CYP2E1 bioactivation to form reactive intermediates such as *N*-acetyl-*p*-benzoquinone imine (NAPQI) that deplete glutathione and then bind to critical cellular

**TABLE 1 Toxicoproteomic Analysis of Liver and Kidney Toxicants**[a]

| Reference | Chemical | Platform | Tissue | Results |
|---|---|---|---|---|
| 46 | APAP, AMAP | 2D-MS | Mouse liver | 35 proteins identified; altered proteins are known APAP adducts |
| 47 | APAP | 2D DIGE MS | Mouse liver | Optimization study for 2D DIGE separation; several proteins altered with APAP |
| 48 | APAP | 2D DIGE MS | Mouse liver | DNA array/proteomics; ↓ Hsps; protein changes in 15 min |
| 49 | APAP, others | NEPHGE 2D-MS | Rat liver, HepG2 | 113, 194 proteins identified in rat liver, HepG2 cells; catalase, carbamoylphosphate synthetase-1, aldo-keto-reductase, altered |
| 53 | APAP | ICAT | Mouse hepatocytes | Optimization study for ICAT analysis |
| 50 | APAP, $CCl_4$, amiodarone | 2D-MS | Rat hepatocytes | 31 proteins identified of 113 proteins altered by APAP |
| 52 | APAP | ICAT | Mouse liver | 1632 proteins identified with 247 proteins diffentially expressed; ↑ hepatoprotective proteins in SJL resistant strain; ↑ loss mitochondrial proteins from C57B1/6 susceptible strain |
| 74 | APAP, ANIT, Wy14643, Pb | 2D DIGE MS | Rat liver, serum | Liver (124), serum (101) proteomics (proteins identified); five serum biomarkers for liver toxicity—PNP, MDH, GcBP, PON1, RBP |
| 51 | APAP, TCN, amiodarone | 2D-MS | Rat hepatocytes | For APAP, ↑15 and ↓25 proteins identified; ↓GPx; ↑PRx1,2 |
| 128 | APAP | 2D-MS | Rat liver, serum | Cluster analysis of transcriptomic, proteomic, and clinical chemistry data |
| 129 | $CCl_4$ | 2D-MS | Rat liver stellate cells | 150 proteins identified; ↑ calcyclin, calgizzarin, galectin-1 |
| 55 | $CCl_4$ | 2D-MS | Rat liver | 30 proteins identified; proliferation and apoptosis proteins altered |
| 57 | BB | 2D-MS | Rat liver | DNA array/proteomics; identified proteins infer degradation, oxidative stress from toxicity |
| 29 | Aflatoxin B1 | 1D gel LC-MS/MS | Rat hepatocyte | Rat secretome of 200 proteins identified; ↓α1-antitrypsin and 2-macroglobulin secretion with aflatoxin |
| 60 | Wy14643, rosiglitizone | 2D-MS | ob/ob Mouse liver | ↑ FA oxidation, lipogenesis in ob/ob with both PPAR activators; gluconeogenesis, glycolysis, AA metabolism affected with both |
| 59 | Wy14643, oxazepam | 2D-MS | Mouse liver | DNA array/proteomics; subcellular fractions; protein identification unique to each chemical |

| | | | | |
|---|---|---|---|---|
| 130 | Hydrazine | 2D-MS | Liver mitochondria | Detection of carbonylated proteins after hypoxia |
| 63 | Hydrazine | 2D DIGE MS | Rat liver | Lipid, $Ca^{2+}$, thyroid, stress pathways activated by toxicity |
| 62 | Hydrazine | 2D DIGE MS | Rat liver | ↑10, ↓10 proteins identified in lipid, $Ca^{2+}$, thyroid, stress pathways |
| 67 | Thioacetamide | 2D-MS | Rat liver | STAP protein, stellate cells related to TAA-cirrhosis model |
| 66 | Thioacetamide | SELDI, MALDI-TOF/TOF | Rat liver | His-rich glycoprotein in serum related to TAA-cirrhosis model |
| 65 | Thioacetamide | 2D-MS | Rat liver | Liver cirrhosis model found ↓ FA β-oxidation, branched chain AA, and methionine breakdown; ↑oxidative stress, lipid peroxidation pathways |
| 78 | Cyclosporin A | 2D | Rat liver, kidney | Multiple 2D gel spots changed with cyclosporine in kidney (19 spots) and liver (29 spots) |
| 79 | Cyclosporin A | 2D AA sequencing | Rat kidney | Specific ↓ of calbindin in kidney by cyclosporine; identified by AA sequencing; ELISA validation |
| 82 | Cyclosporin A | 2D gel, 35S-labeling | Murine T cells | Cyclosporin A induces >100 proteins not found in resting or activated murine T-cells; proteins, like CSTAD, identified in later study [83] |
| 90 | Puromycin | 2D-MS | Rat kidney | Proteomics/metabolomics to study renal glomerular toxicity |
| 131 | PbAc | 2D-MS | Rat kidney, medulla, cortex | Subacute Pb altered 76 cortex proteins and 13 medullar proteins; ↓ calbindin, calcineurin, arginino-succinate synthetase; ↑ GSTM1 |
| 88 | Gentamicin | 2D-MS | Rat urine | 20 proteins identified; mitochondrial dysfunction in renal cortex |
| 95 | 4-Aminophenol, D-serine | 2D-MS | Rat plasma | FAH identified as kidney damage biomarker |
| 94 | 4-Aminophenol, D-serine, cis-Pt | 2D-MS | Rat plasma | T-kininogen, inter-α-inhibitor H4P, complement c3 identified as kidney damage biomarkers |
| 108 | DCVC | 2D DIGE MS | Porcine renal LLC-PK1 line | ↑14, ↓9 proteins with DCVC; Hsp27 phosphorylation isoform found as marker of prosurvival by maintaining cell adhesion |

[a]Summaries of liver and kidney toxicoproteomic studies are intended to briefly overview study details and results. Only abbreviations that are helpful in interpreting the summary notes are included. For further explanation, please refer to the citation. Abbreviations are listed in order of appearance in the table. APAP, acetaminophen; AMAP, 3′-isomer of acetaminophen; 2D-MS, 2D gel mass spectrometry; 2D DIGE MS, 2D gel differential gel expression (Cy3,5 dyes) mass spectrometry; ↑, ↓, increase, decrease; Hsps, heat shock proteins; NEPHGE, nonequilibrium pH gel electrophoresis; ICAT, isotope coded affinity tags; $CCl_4$, carbon tetrachloride; ANIT, α-naphthylisothiocyanate; Wy14643, Wyeth 14643 compound; Pb, phenobarbital; PNP, purine nucleotide phosphorylase; MDH, malic dehydrogenase; GcBP, vitamin D binding protein; PON1, paraoxonase; RBP, retinol binding protein; TCN, tetracycline; GPx, glutathione peroxidase; PRx1,2, peroxiredoxin 1 and 2; BB, bromobenzene; LC-MS/MS, liquid chromatography tandem mass spectrometry; PPAR, peroxisome proliferator-activated receptor; $Ca^{2+}$, calcium; STAP, stellate cell activation associated protein; TAA, thioacetamide; SELDI, surface enhanced laser dissociated ionization; MALDI TOF/TOF, MALDI-based tandem MS; His, histidine; FA, fatty acid; AA, amino acid; CSTAD, cyclosporin A-conditional, T cell activation-dependent gene; PbAc, lead acetate; GSTM1, GSH transferase M1; FAH, fumarylacetoacetate hydrolase; cis-Pt, cisplatin; H4P, inter-α inhibitor H4P heavy chain; DCVC, dichlorovinyl-L-cysteine.

macromolecules [35]. Mitochondria are thought to be primary targets in acetaminophen toxicity with particular attention on the mitochondrial permeability transition [39]. It is worth noting that mitochondrial dysfunction underlies the pathogenesis of several toxicities in preclinical species especially in liver, skeletal, and cardiac muscle, and the CNS [40]. Evidence has also been accumulating for the contribution of nonparenchymal cells such as Kupffer cells, NK cells, neutrophils, and endothelial cells that secrete cytokines and chemokines during acetaminophen-induced liver injury [41–45].

Some of the earliest toxicoproteomic studies using 2D-MS platforms were conducted using standard 2D-MS [46] analysis as well as the 2D DIGE MS platform alone [47] or in combination with DNA microarrays [48]. Proteomic analysis of livers from these studies in mice identified altered proteins that are known targets for adduct formation such as mitochondrial proteins, HSPs, and other structural and intermediary metabolism proteins. A different type of 2D gel separation using a nonequilibrium approach to charge separation of proteins (NEPHGE) found 100–200 differentially expressed proteins in rat liver and HepG2 cells, especially in enzymes involved in intermediary metabolism [49].

Studies using rat hepatocytes exposed to acetaminophen and analyzed by 2D-MS have found it helpful to concurrently evaluate other cytotoxic pharmaceutical agents such as tetracycline, amiodarone, and carbon tetrachloride [50, 51]. These studies found alterations in several metabolic enzymes and identified glutathione peroxidase and peroxiredoxins 1 and 2 (PRX1, PRX2), which serve as cellular responsive antioxidative enzymes during toxicant exposure.

One of the first LC-MS/MS studies using ICAT technology that involved acetaminophen toxicity in mouse liver was published in 2005 [52], which was preceded by an earlier optimization study for ICAT in mouse hepatocytes [53]. This study combined the more comprehensive ICAT analysis procedure with an adept choice of resistant (SJL) and susceptible (C57B1/6) mouse strains to investigate potential susceptibility factors (proteins and pathways) in acetaminophen toxicity [52]. Inherent differences in liver homogenate protein expression levels between resistant SJL and susceptible C57B1/6 mice were found by comparison of hepatic proteomics after vehicle (saline) treatment at 6 h. Of the 1236 protein identified, 121 were differentially expressed between the two mouse strains. After 6 h of 300 mg/kg acetaminophen by intraperitoneal treatment, 1632 proteins were identified from which 247 were different between the two strains and 161 proteins were more abundant in the SJL strain. Some of these naturally more abundant proteins (in the absence of toxicant) may have protective roles against toxicity including two- to fourfold increases in lactoferrin, galectin-1, tripeptidyl-peptidase II, proteasomal subunit β-Type1, and DnaJ homolog A1. Upon administration of acetaminophen, comparative expression showed that SJL mice expressed from 3- to 10-fold higher levels of ubiquitin-like 2 (SUMO1) activating enzyme E1B, complement c5, COX1, peroxiredoxin 1, Grp170, Hsp70, GSTμ-2, and regucalcin. In addition to antioxidant enzyme functions, many of these upregulated proteins may have a reparative role in degrading denatured and damaged proteins, cell proliferation and regeneration, and cellular stress response. A selective loss of several mitochondrial proteins from susceptible C57B1/6 mice suggested this organelle is particularly vulnerable in acetaminophen-induced hepatic injury.

### 5.2 Carbon Tetrachloride

Carbon tetrachloride produces acute centrilobular hepatic necrosis but has frequently been used in a repeated exposure regimen over several weeks to produce an animal model of liver fibrosis [54]. Activation of hepatic stellate cells from a quiescent vitamin A-storing cell to a myofibroblast-like cell is a key event in excessive accumulation of fibril-forming extracellular matrix proteins and development of liver fibrosis. Proteomic analysis was performed on cellular and secreted proteins of normal and activated rat hepatic stellate cells either *in vitro* or *in vivo* after carbon tetrachloride for 8 weeks. Of the 43 altered proteins identified, 27 showed similar changes *in vivo* and *in vitro* including upregulation of calcyclin, calgizzarin, and galectin-1 as well as downregulation of liver carboxylesterase 10. These changes were confirmed in fibrotic liver tissues. A compendium of 150 stellate cellular and secreted proteins were identified.

Another carbon tetrachloride fibrosis study conducted a 2D-MS proteomic analysis on liver tissues from rats exposed to carbon tetrachloride for a period of 4–10 weeks [55]. During this exposure period, collagen deposition and hydroxyproline content of fibrotic livers increased continuously. Differentially expressed proteins from proteomic analysis were categorized as proliferation-related proteins/enzymes (proliferating cell nuclear antigen p120, p40, and cyclin F ubiquitin-conjugating enzyme 7—UBC7), and apoptosis-related proteins, mainly caspase-12, which was absent in the control rats. These researchers found that proliferation- and apoptosis-related proteins are dynamically expressed during different stages of rat liver fibrosis induced by carbon tetrachloride.

### 5.3 Bromobenzene

Bromobenzene is another well-characterized model liver toxicant whose metabolism, reactive intermediates, protein adducts, and liver toxicity phenotype (centrilobular necrosis) have been well characterized [56]. A transcriptomic and proteomic comparison of bromobenzene conducted after 24 h exposure to a single dose of bromobenzene showed alterations in transcripts and genes involved in drug metabolism, oxidative stress, sulfhydryl metabolism, and acute phase response [57]. Of the 1124 proteins resolved from liver homogenates, 24 proteins were differentially expressed and identified as intermediary or drug metabolism enzymes.

### 5.4 Wyeth 14643

The peroxisome proliferator-activated receptors (PPARs) are ligand-activated transcription factors that modulate lipid and glucose homeostasis [58]. Wyeth 14643 is a hepatic metabolic enzyme inducer and acts as a potent agonist of peroxisome proliferator-activated receptor alpha (PPAR-$\alpha$), a member of the nuclear hormone receptor superfamily and a key transcriptional regulator of many genes involved in free fatty acid oxidation systems in liver. Global gene and protein expression changes were compared by cDNA microarray of mouse liver and 2D-MS of mouse liver subcellular fractions from B6C3F1 mice treated from 0.5 to 6 months with oxazepam and the peroxisome proliferator, Wyeth 14643 [59]. Each compound produces hepa-

tocellular cancer after a 2 year bioassay of dietary exposure. The hypothesis was that each compound would produce cancer by different biochemical pathways and that transcript and protein changes measured prior to tumor formation (up to 6 months) would provide mechanistic insights into carcinogenesis. After 6 months, only 36 transcripts were altered after oxazepam compared to 220 transcripts with the Wyeth compound. Notable genes upregulated in the signature profile for oxazepam were Cyp2b20, Gadd45β, TNFα-induced protein 2, and Igfbp5. Upregulated genes with Wyeth compound were cyclin D1, PCNA, Igfbp5, Gadd45β, and CideA. Altered expression of over 100 proteins by proteomic analysis showed upregulation of the cancer biomarker α-fetoprotein in cytosol and cell cycle-controlled p38-2G4 protein in microsomes during both treatments. Both transcriptomic and proteomic analyses were deemed complementary in distinguishing between two chemical carcinogens that appear to proceed through different mechanisms and eventually lead to liver cancer as the common phenotype.

Insights into the therapeutic action of PPAR-α and PPAR-γ agonists, WY14643 and rosiglitazone, respectively, were reported in proteomic analysis of the ob/ob animal model of obesity disease [60]. Hepatic protein expression profiles were developed by 2D-MS analysis of lean and obese (ob/ob) mice, and obese mice treated with WY14643 or rosiglitazone. Livers from obese mice displayed higher levels of enzymes involved in fatty acid oxidation and lipogenesis compared to lean mice and these differences were further amplified by treatment with both PPAR activators. WY14643 normalized the expression levels of several enzymes involved in glycolysis, gluconeogenesis, and amino acid metabolism in the obese mice to the levels of lean mice. Rosiglitazone only partially normalized levels of enzymes involved in amino acid metabolism. This study used an established mouse model of obesity disease to map metabolic pathways and discriminate between PPAR-α and PPAR-γ agonist effects by proteomic analysis.

### 5.5 Hydrazine

Hydrazine is a model, cross-species hepatotoxicant used as an industrial reagent and found as a drug metabolite of the structurally related pharmaceutical isoniazid (antituberculosis drug) and the antihypertensive agent hydralazine. Hydrazine typically causes initial steatosis, macrovesicular degeneration, followed by marked hepatic necrosis. Transcriptomic studies suggest hydrazine initiates a process whereby the production and intracellular transport of hepatic lipids is favored over the removal of fatty acids and their metabolites [61].

Proteomic studies using 2D DIGE MS on the hepatotoxic effects of hydrazine were conducted in rats from 48 to 168 h [62, 63]. In one study, 2D gel patterns from liver were analyzed by principal component analysis (PCA) and partial least squares regression. PCA plots described the variation in protein expression related to dose and time. Regression analysis was used to select 10 upregulated proteins and 10 downregulated proteins that were identified by mass spectrometry. Hydrazine treatment altered proteins in lipid metabolism, $Ca^{2+}$ homeostasis, thyroid hormone pathways, and stress response. In a second study, low density cDNA microarrays and 2D DIGE MS proteomics of liver tissue and metabonomic analysis of serum were performed from hydrazine-treated rats at 48–168 h [62]. Their findings supported known

effects of hydrazine toxicity and provided potential biomarkers of hydrazine-induced toxicity.

### 5.6 Thioacetamide

Thioacetamide is metabolically activated in liver to produce thioacetamide-*S*,*S*-dioxide as a reactive intermediate that binds to liver macromolecules to initiate centrilobular necrosis [64]. Repeated administration of thioacetamide is an established technique for generating rat models of liver fibrosis and cirrhosis depending on dose and length of administration (weeks). A 2D-MS proteomic approach was used to profile liver protein changes in rats receiving thioacetamide for 3, 6, and 10 weeks to induce hepatic cirrhosis [65]. Expression of 59 protein spots altered by thioacetamide can be identified, including three novel, unannotated proteins. Downregulation of enzymes was noted in pathways such as fatty acid β-oxidation, branched chain amino acids, and methionine breakdown, which may relate to succinyl-CoA depletion and affect heme and iron metabolism. Increased levels were found for enzymes responding to oxidative stress and lipid peroxidation such as GSH transferases. Finally, these proteomic data were integrated into a proposed overview model for thioacetamide-induced liver cirrhosis affecting succinyl-CoA and cytochrome P450 production combined with iron release and hydrogen peroxide generation.

In another model of thioacetamide-induced liver cirrhosis in rats, researchers probed for potential serum biomarkers using the SELDI proteomic approach [66]. A weak cation exchange surface was used to analyze serum by SELDI MS from control (normal) rats, thioacetamide-induced liver cirrhosis rats, and rats with bile duct ligation-induced liver fibrosis. A consistently downregulated 3495 Da protein in cirrhosis samples was one of the selected significant biomarkers. This 3495 Da protein was purified on-chip and was trypsin digested on-chip for MS/MS identification of a histidine-rich glycoprotein. The new protein was proposed as a new preclinical biomarker for the rat cirrhosis model that might eventually prove useful for early clinical detection of liver cirrhosis and classification of liver diseases.

An innovative study involving stellate cell activation by 8 week treatment with thioacetamide utilized a proteomic approach that led to the discovery of a novel protein named STAP for "stellate cell activation-associated protein" [67]. Quiescent and thioacetamide-activated stellate cells were analyzed by 2D-MS (ESI-MS/MS) to identify 43 proteins altered during the activation process including upregulation of collagen-$\alpha_1$ (I and III), γ-actin, neural cell adhesion molecule (N-CAM), calcyclin, calgizzarin, and galectin-1. In particular, STAP was highly increased both in activated stellate cells and in fibrotic liver tissues induced by thioacetamide treatment. These researchers cloned the STAP gene and found it was a cytoplasmic protein, expressed only in stellate cells, with molecular weight of 21,496 Da and a 40% amino acid sequence homology to myoglobin. Biochemical characterization showed STAP is a heme protein exhibiting peroxidase activity toward hydrogen peroxide and linoleic acid hydroperoxide. These results indicate that STAP is a novel endogenous peroxidase catabolizing hydrogen peroxide and lipid hydroperoxides, both of which have been reported to trigger stellate cell activation and consequently promote progression of liver fibrosis. STAP was postulated to play a role as an antifibrotic scavenger of peroxides in the liver.

## 6 NEW BLOOD BIOMARKERS IN LIVER INJURY

Blood is one of the most accessible and informative biofluids for specific organ pathology in preclinical studies. Biomarkers that can be assayed in biological fluids from preclinical species may hold relevance for human subjects [17]. A comprehensive mapping of soluble human blood elements of the plasma proteome is currently underway for improved understanding of disease and toxicity by the Human Proteome Oganization (HUPO) [68]. Results from an international survey of soluble human blood proteins by chromatographic and electrophoretic separation have revealed several thousand resolvable proteins for which mass spectrometry has provided evidence for over 1000 unique protein identifications [68, 69]. Researchers are also mapping the mouse [70] and rat [71] serum and plasma proteomes for use in preclinical and experimental studies. An excellent review has been published for 2D gel mapping of rat serum and rat tissue proteomic studies [71].

The sensitivity of 2D gel proteomic approaches to detect and measure alterations in the mouse or rat plasma proteomes has only recently been tested by various labs. Researchers examined changes in the mouse plasma proteome, focused upon inflammation after cutaneous burn injury with superimposed *Pseudomonas aeruginosa* infection [72]. Upregulations of inter α-trypsin inhibitor heavy chain 4 and hemopexin were detected along with other mouse acute-phase proteins, including haptoglobin and serum amyloid A. In another inflammation study, reference maps of the mouse serum proteome were generated by 2D-MS from control animals and from mice injected with lipopolysaccharide (LPS) to induce systemic inflammation, and from mice transgenic for human apolipoproteins A-I and A-II [73]. The greatest changes were noted for haptoglobin and hemopexin.

Finally, a comparative plasma proteome analysis has been reported in which investigators used 1D gel LC-MS/MS analysis upon a few microliters of plasma from lymphoma-bearing SJL mice experiencing systemic inflammation [28]. After removal of albumin and Igs from plasma, these researchers identified a total of 1079 nonredundant mouse plasma proteins; more than 480 in normal and 790 in RcsX tumor-bearing SJL mouse plasma. Of these, only 191 proteins were found in common. Many of the upregulated proteins were identified as acute-phase proteins but several unique proteins, including haptoglobin, proteosome subunits, fetuin-B, 14–3-3ζ, and MAGE-B4 antigen, were found only in the tumor-bearing mouse plasma due to secretion or shedding by membrane vesicles, or externalized due to cell death. These results are very encouraging for the effectiveness of a proteomic approach for protein identification from small sample amounts, and for comparative proteomics in animal models of drug-induced toxicity or disease.

The application of serum or plasma protein maps in toxicoproteomics such as serum profiling of liver injury has just begun to take shape. A recent study reported identification of serum proteins altered in rats treated with four liver-targeted compounds including acetaminophen, ANIT (1-naphthylisothiocyanate), phenobarbital, and Wyeth 14643 at early, fulminant, and recovery periods of effect [74]. Nineteen serum proteins were identified as significantly altered from the four studies and, among them, five serum proteins were of special interest as serum markers for early hepatic toxicity or functional alterations in rats, including vitamin D-binding protein (Gc binding protein), purine nucleotide phosphorylase (PNP), malic dehydrogenase (MDH), paraoxonase (PON1), and retinol-binding protein (RBP). Some of these

proteins may serve as early predictive markers of hepatotoxicity for new drug candidates or may be more sensitive than other conventional methods.

The soluble portion of blood, serum or plasma, is regarded as a complex biofluid tissue. While many organs contribute various proteins as blood solutes, the liver is by far the most productive member of all organs and tissues. It has occurred to some researchers to study the secreted proteome of hepatocytes since liver parenchyma are often primary targets of drug-induced toxicity, and they also secrete many plasma proteins, which can be measured in preclinical species. Secreted proteins were separated and identified from primary rat hepatocytes using a collagen gel sandwich system. Proteomic analysis was conducted by a 1D gel LC-MS/MS procedure. More than 200 secreted proteins were identified that included more than 50 plasma proteins, several structural extracellular matrix proteins, and many proteins involved in liver regeneration. Secretion of two proteins, $\alpha_1$-antitrypsin and $\alpha_2$-macroglobulin, was greatly reduced in aflatoxin B1 exposed hepatocytes. This study provides evidence that proteomic analysis of medium from hepatocyte sandwich culture might represent a new *in vitro* model and general approach for future discoveries of secreted biomarkers in drug-induced chemical toxicity.

## 7 TOXICOPROTEOMIC STUDIES IN KIDNEY INJURY

Kidney is a primary organ for preclinical assessment in pharmaceutical development since its metabolic and excretory functions often render it susceptible to drug-induced toxicity [75]. The kidney is a major organ for filtration, reabsorption, and secretion to maintain homeostasis of water-soluble salts and small molecules. The organ also has a considerable capacity for biotransformation of drugs and xenobiotics. Specific physiological characteristics of the kidney are localized to specific cell types (i.e., vascular endothelial and smooth muscle cells, mesangial cells, interstitial cells, podocytes, proximal and distal tubular epithelial cells), with each demonstrating selective susceptibility to toxicity. Renal damage can be due to several different mechanisms affecting different segments of the nephron, renal microvasculature, or interstitium. The nature of renal injury may be acute and recoverable. However, other drugs with repeated exposure can produce chronic renal changes that may lead to end-stage renal failure. The abilities to perform kidney transplants and other organ replacements have saved many lives because of immunosuppressive drug treatment to prevent organ rejection. However, immunosuppressive drugs also run the risk of renal toxicity over time. New nephrotoxic markers amenable for multiple preclinical models and high throughput screening are major goals for toxicoproteomic and toxicogenomic technologies [75, 76].

### 7.1 Cyclosporin A

Some of the groundbreaking studies that initiated the field of toxicoproteomics took place in the mid-1990s and involved investigating the side effects of the immunosuppressant drug cyclosporin A. Cyclosporin A is a calcineurin inhibitor that has been a mainstay for immunosuppressive therapy following solid-organ transplantation. Cyclosporin A blocks immune responses by inhibiting the calcineurin-dependent dephosphorylation of the nuclear factor of activated T cells (NFAT). However, a

dose-dependent nephrotoxicity occurs with high incidence that is characterized by nonhistologic functional deficits or functional decline, with calcium loss in urine (hypercalciuria), vascular-interstitial lesions, and calcification of renal tubules [77].

Initial 2D gel studies were conducted in rat liver and kidney samples that showed changes in 48 proteins in these tissues in rats treated with cyclosporin A in which an unidentified protein present only in the kidney was uniquely downregulated [78]. A subsequent 2D gel study of kidney homogenates identified a decrease in the 28 kDa kidney protein as calbindin-D using protein microsequencing. Importantly, this same study validated a time-dependent decrease in calbindin expression by ELISA for up to 28 days of cyclosporin A treatment [79]. These toxicoproteomic studies published a decade ago represented a significant advance in understanding an important part of cyclosporine-induced pathophysiology in kidney.

More recently, the contribution of calbindin-D28k has been clarified by the generation of genetically modified mice. Cyclosporin A-induced hypercalciuria represents two pathophysiological processes: a downregulation of calbindin-D28k with subsequent impaired renal calcium reabsorption and a cyclosporin A-induced high turnover bone disease [80]. In addition, there is evidence that one biochemical mechanism underlying cyclosporin A and other calcineurin inhibitors may be a drug-induced mitochondrial dysfunction [81].

The effects of cyclosporin A on gene upregulation were advanced by a 2D gel proteomic analysis of newly synthesized $^{35}$S-methionine-labeled proteins in murine T cells activated in the absence or presence of cyclosporin A [82]. Remarkably, these investigators found more than 100 proteins not present in resting or activated T cells that could be induced by cyclosporin A exposure. It is important to emphasize that the discovery nature of this proteomic study was capitalized upon (same researchers) in identification of the corresponding genes under the same treatment conditions using a transcript enrichment technique called "representational difference analyses" [83]. Among the upregulated transcripts, a new gene was found, named *CSTAD*, for "CSA-conditional, T cell activation-dependent" gene. *CSTAD* encodes two proteins of 104 and 141 amino acids that are localized in mitochondria [83]. *CSTAD* upregulation is observed in mice after cyclosporin A treatment, suggesting that upregulation of *CSTAD* and perhaps many other genes may be newly implicated in cyclosporin A toxicity. Thus, toxicoproteomics has played an important role in furthering the understanding of the critical proteins and biological pathways in cyclosporin A toxicity that should lead to better biomarkers for this important class of pharmaceuticals.

### 7.2 Puromycin and Gentamicin

The regionally specific structure and function of the kidney renders specialized areas to be more susceptible to toxicity from exposure to certain pharmaceutical agents. For example, puromycin aminonucleoside is an antibiotic that causes glomerular podocyte necrosis, nephrosis, and proteinuria in rodent models [84]. Gentamicin is an aminoglycoside antibiotic that accumulates in proximal tubular epithelia and inhibits cell lysosomal function, producing phospholipidosis and tubular degeneration [85]. Some studies have begun to proteomically characterize specific regions such as the medulla and cortex [86] or subcellular structures of kidney cell types such as the nucleus [87]. In one study, the nephrotoxic effects of

gentamicin upon protein expression were studied in rat kidney. Results revealed the identities of more than 20 proteins involved in the citric acid cycle, gluconeogenesis, fatty acid synthesis, and transport or cellular stress responses [88]. The authors believed that impairment of energy production and mitochondrial dysfunction were involved in gentamicin-induced nephrotoxicity.

Another approach to studying nephrotoxicity is by proteomic characterization of urine. Proteomic mapping of rat urine proteins studied by 2D-MS resolved 350 protein spots from which 111 protein components were identified including transporters, transport regulators, chaperones, enzymes, signaling proteins, cytoskeletal proteins, pheromone-binding proteins, receptors, and novel gene products [89]. One toxicoproteomic study examined urinary protein expression profiles to gain insight into puromycin-induced kidney toxicity [90]. Nephropathy and proteinuria caused by puromycin aminonucleoside in rats was studied by metabonomics and a 2D-MS proteomic analysis of urinary proteins from 8 to 672 h after dosing. Prior to exposure, major urinary protein (MUP), $\alpha_2$-microglobulin, and glial fibrillary acid protein isoforms were the major urinary proteins found in addition to many other unidentified low mass urinary proteins. Following puromycin treatment, a gradual increase in higher mass proteins was observed on 2D gels, particularly albumin, at 32 h after dosing. By 120 h, albumin, transthyretin, and Gc-binding protein were identified as major urinary proteins from puromycin-induced kidney damage. After 672 h, the urinary protein profile in 2D gels had largely returned to normal. Many of these plasma-derived proteins appearing in the urine over 0–672 h following puromycin administration were consistent with loss of glomerular integrity and major leakage of plasma protein in urine. This study suggests that urinary proteomics, in conjunction with these other techniques, has the potential to provide significantly more mechanistic information than is readily provided by traditional clinical chemistries and may be a productive means for biomarker discovery of nephrotoxic agents in preclinical species.

### 7.3 4-Aminophenol, D-Serine, and Cisplatin

The nephrotoxin 4-aminophenol (4-AP) produces severe necrosis of the pars recta of the proximal tubule in the rat that is thought to occur through formation of the toxic metabolite 1,4-benzoquinoneimine [91]. D-Serine, an enantiomer of L-serine, is another model nephrotoxicant that selectively damages the pars recta of proximal tubules in the kidney, which may involve formation of toxic oxidative metabolites [92]. The chemotherapeutic agent cisplatin is also a model nephrotoxicant and targets different portions of the kidney. It is metabolized to cytotoxic intermediates in proximal tubular epithelial cells and induces necrosis in distal tubules and collecting ducts along with causing mild glomerular toxicity [93]. Proteomic profiling using 2D-MS was used to investigate plasma protein changes in rats treated with 4-aminophenol, D-serine, and cisplatin compared to saline controls [94]. Nontoxic isomers, L-serine, and transplatin were also studied. Many plasma proteins were found that displayed dose- and time-dependent response to toxicants. Several isoforms of T-kininogen protein were identified as increasing in plasma at early time points and returning to baseline levels after 3 weeks with each nephrotoxicant but not with nontoxic compounds. In addition, inter-$\alpha$ inhibitor H4P heavy chain was increased in the 4-AP and D-serine studies. A further set of proteins correlating with kidney

damage was found to be a component of the complement cascade and other blood clotting factors, indicating a contribution of the immune system to the observed toxicity. It was proposed that T-kininogen may be required to counteract apoptosis in proximal tubular cells in order to minimize tissue damage following a toxic insult.

In a related study, plasma samples from 4-aminophenol and D-serine treated rats were profiled by 2D-MS showing dose- and time-dependent effects of various plasma proteins in response to these nephrotoxicants [95]. One toxicity-associated plasma protein was identified as the cellular enzyme fumarylacetoacetate hydrolase (FAH), a key component of the tyrosine metabolism pathway. FAH was elevated in the plasma of animals treated with 4-aminophenol and D-serine at early time points and returned to baseline levels after 3 weeks. The protein was not elevated in the plasma of control animals or those treated with the nontoxic isomer L-serine. The investigators raised the possibility that FAH might serve as a marker of kidney toxicity in preclinical species.

### 7.4 Dichlorovinyl-L-Cysteine

Dichlorovinyl-L-cysteine (DCVC) is a model nephrotoxicant taken up by renal proximal tubular epithelia, where it is bioactivated by renal cysteine conjugate (β-lyase) to form reactive, cytotoxic intermediates [96]. DCVC is a metabolite of trichloroethylene but can be chemically synthesized for use in experimental studies [97].

A proteomic study of DCVC toxicity was conducted in LLC-PK1 porcine renal epithelial cells by 2D DIGE MS to determine early changes in stress-response pathways preceding focal adhesion disorganization linked to the onset of apoptosis. DCVC treatment caused a greater than 1.5-fold up- and downregulation of 14 and 9 proteins, respectively, prior to apoptosis including aconitase and pyruvate dehydrogenase, and those related to stress responses and cytoskeletal reorganization, such as cofilin, Hsp27, and alpha-β-crystallin. Most noticeable was a p*I* shift in Hsp27 from phosphorylation at Ser-82. Only inhibition of p38 with SB203580 reduced Hsp27 phosphorylation, which was associated with accelerated reorganization of focal adhesions, cell detachment, and apoptosis. Inhibition of active JNK localization at focal adhesions did not prevent DCVC-induced phosphorylation of Hsp27. Overexpression of a phosphorylation-defective mutant Hsp27 acted as a dominant negative and accelerated DCVC-induced focal adhesion changes and the onset of apoptosis. Early p38 activation appears to rapidly phosphorylate Hsp27, to maintain cell adhesion, and to suppress renal epithelial cell apoptosis. This toxicoproteomic study combines both protein identification and posttranslational modification to elucidate critical proteins (Hsp27) and protein attributes (phosphorylation) in critical pathways (p38 stress pathway) to gain insight into mechanisms of renal epithelial cell death.

## 8 TOXICOPROTEOMICS OF ADVERSE EFFECTS AND TOXICANTS: CITATION ANALYSIS

The application of proteomics to toxicology is in its early stages. Published toxicoproteomic reports relevant to pharmaceutical research so far have focused on liver,

kidney, and blood, as described in the previous sections of this chapter. However, it is of great interest to further stratify the areas of toxicoproteomic research. Citation analysis is one approach for further discriminating specific research areas in toxicoproteomics. A citation analysis was performed using SciFinder® software (CAS, Columbus, OH) to acquire an overview of how proteomics has been used in toxicology. A search of "proteomics" and "toxicity" terms retrieved a total of 895 documents from Medline and Chemical Abstract services that included primary literature, reviews, commentary, and chemical and patent oriented data. A subgroup of 323 citations of the total is shown in a two-dimensional bar graph in Fig. 2. The bar graph matches the SciFinder preset subcategories in "Toxicology"

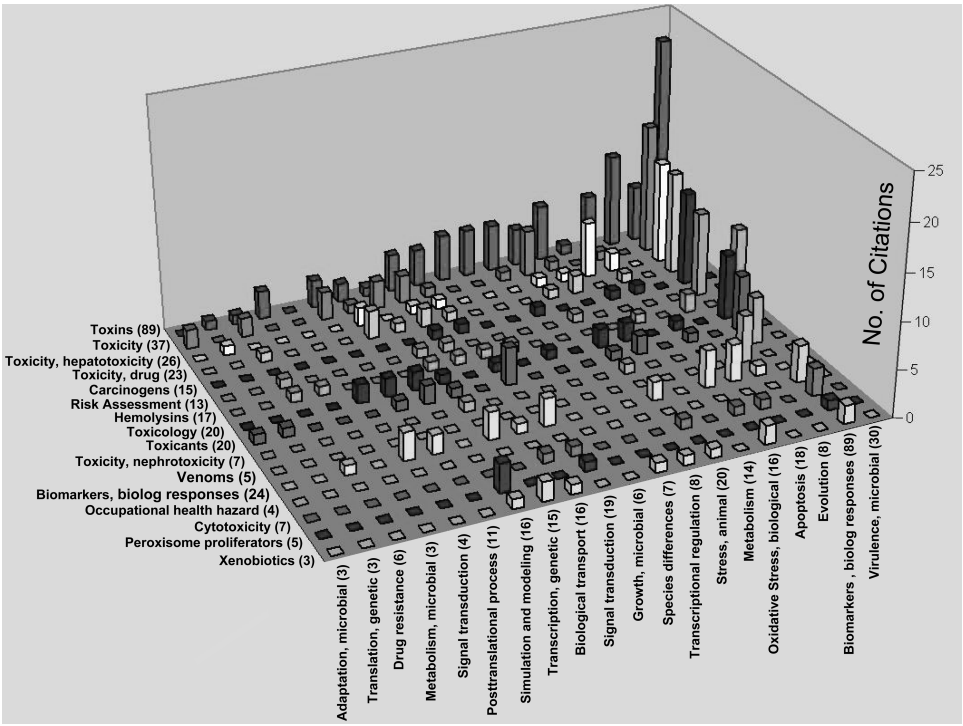


**FIGURE 2** Role of proteomics in preclinical assessment and toxicology research. A 3D histogram displays the frequency and distribution of proteomic research citations in toxicology research as of March 2006. SciFinder® citation analysis matched the terms *proteomics* and *toxicity* and retrieved a total of 895 documents to comprise available toxicoproteomics references. The algorithm searches both Medline and Chemical Abstracts for published journal and conference literature and patents. A subgroup of 323 citations is shown in the bar chart to give more definition to citation analysis. The subgroup of 323 citations is the product of matching the category "Processes and Systems" (its selected subcategories on the *x*-axis) with the category "Toxicology" (its selected subcategory on the *y*-axis). Categories and subcategories are preset divisions of SciFinder software. The *z*-axis shows the number of citations (number in parentheses) for each bar, which represents the product of matching subcategories from the *x*- and *y*-axes. (SciFinder data display output from CAS, a division of the American Chemical Society.) In some subcategories, "biological" is abbreviated as "biolog."

with "Processes and Systems" categories to give more definition to the citation analysis.

The most cumulative citations occurred in the subcategory "Biomarkers, biolog responses," at 89 citations when matched with selected subcategories of toxicology. Similarly, the subcategory "Toxins" had 89 cumulative citations when matched with selected subcategories of "Processes and Systems." Other notable subcategories in "Processes and Systems" with high citations were "Virulence, microbial" (30), "Apoptosis" (18), "Stress, animal" (20), and "Signal transduction" (19). In "Toxicology," notable citations were in subcategories for "Toxicity" (37), "Toxicity, hepatotoxicity" (26), "Toxicity, drug" (23), and "Biomarkers, biolog responses" (24). Clearly, biomarker development is a primary focus in toxicoproteomic studies, particularly in liver.

The interest in preclinical development of biomarkers for drug toxicity may leave one to wonder why there are not more primary data reports in pharmaceutical toxicoproteomics. One reason is that linkages of mass spectrometry to separation technologies have only been available for 6–7 years and are just becoming more widely accessible in the last few years. A second reason is that some excellent toxicoproteomic research in the commercial sector may not be publicly accessible because of proprietary concerns unless compound structures can be coded [98] or only partially divulged [99]. A third reason is that meaningful differential protein expression is not easily achieved in many drug-exposed systems. Often drug targets are specific for receptors, signaling pathways, or portions of the cell that are at low levels of expression and not easily detectable (i.e., membrane receptor proteins). Alternately, there may be an overabundance of noninformative, interfering proteins (i.e., albumin removal from serum) that mask important low level proteins. Reduction of complexity or enrichment can be an extremely useful and necessary strategy for achieving differential protein expression prior to toxicoproteomic analysis. Undoubtedly, the range of citations in toxicoproteomics will widen, particularly as mass spectrometry becomes more pervasive and accessible.

## 9 FUTURE DEVELOPMENTS IN TOXICOPROTEOMICS

### 9.1 Toxicoproteomic Studies in Other Organs and Tissues

The brief historical development of toxicoproteomics has focused primarily on liver and kidney. However, protein profiling studies are being conducted in many other organs and tissues to profile adverse effects of therapeutics. Proteomic approaches are revealing new blood serum and tissue biomarkers in animal models of human neurodegenerative diseases like Parkinson disease, Alzheimer disease, and amyotrophic lateral sclerosis (ALS) [100, 101]. Proteomic studies are being conducted in cardiotoxicity models with doxorubicin [102] and renin–angiotensin models of hypertension [103]. Comparative protein expression studies are systematically examining testicular toxicity in rats with several reproductive toxicants such as cyclophosphamide, sulfasalazine, 2,5-hexanedione, and ethylene glycol monomethyl ether [104]. The effects of formaldehyde on rat lung [105] and protein adduct formation of 1-nitronapthalene metabolites in rat lung [106, 107] are being examined by proteomic techniques to provide insight into pulmonary pathology by these agents.

Thus, protein mapping and profiling studies are exploring a variety of preclinical assessment animal models of toxicity and disease.

## 9.2 Expectations and Reality

The expectations of omics technologies in pharmaceutical development are very high, but the breakthroughs in drug discovery and improvements over traditional measures in preclinical assessment have lagged behind. This situation is understandable since the platforms for proteomics continue to be in dynamic development. Furthermore, applications to toxicology settings are still being explored to match platform sensitivity for differential protein expression with preclinical biological samples. Many of the published toxicoproteomic reports reviewed herein have served as proof-of-principle studies using Tier I proteomic analysis as referred to in Fig. 1. The approach has been to examine a well-characterized toxicant(s) and compare proteomic data output with known toxicological endpoints (i.e., serum and urine chemistries, histopathology). These efforts might be described as the discovery phase of toxicoproteomics, where differential protein expressions are determined in response to compound exposure. However, many of these initial studies have often not been accompanied by any confirmation analysis using ELISA, Western blot, immunohistochemistry, or functional assay (i.e., enzymatic activity). Two other areas show a slow progress in toxicoproteomic research. One is in the follow-up *hypothesis-driven research* that further characterizes discovery findings and establishes causal linkage of toxicant exposure and effect. The other area is in *validation studies* of proposed biomarkers using independent and blinded study samples. However, the full cycle of discovery, focused confirmation analysis, and hypothesis-testing for causality is achievable [67, 82, 83, 108]. Validation studies to determine the general applicability of each biomarker represent a more long-term commitment.

Good study design in toxicoproteomic experiments (or any omic experiment) would be expected to include multiple doses, several time points, positive and negative control compounds, nontoxic chemical isomers, single and multiple dosing, confirmation of results, validation in blinded samples, and many other considerations that are standard practice in good science. Such design considerations have not been evenly applied in many of the reviewed toxicoproteomic studies. These shortcomings should be interpreted to determine where we are with state-of-the-art toxicoproteomics. Long data analysis times in interpreting mass spectra, large data volumes per experiment, statistical analyses, and bioinformatics challenges in deriving biological meaning from complex datasets have made toxicoproteomic studies cumbersome and time consuming. The inability organize, integrate, and communicate data within organizations and across many laboratories [109] underscores the continuing need for international proteomic data standards [110], accessible databases [8], and proteoinformatic tools [111] to most efficiently extract biological meaning from toxicoproteomic experiments [112, 113].

## 9.3 Future Trends and Tools in Toxicoproteomics

Future trends in toxicoproteomic studies will see developments in several areas where special attributes of proteins can be exploited by proteomics in preclinical

assessment. First, further refinements of tandem mass spectrometry with intimately integrated multidimensional separation schemes will continue to dominate proteomic analysis. MS instruments will become more accessible. A driving force is the complementary nature of LC/MALDI and LC/ESI tandem mass spectrometry platforms that provide increased proteome coverage for almost any sample to be analyzed [114].

Second, reduction of sample complexity or any prepurification strategy prior to toxicoproteomic analysis will be the most useful when innovatively applied to appropriate biological samples (e.g., immunodepletion of albumin, Igs in plasma) or research problem areas (e.g., phosphoprotein enrichment in protein signaling).

Third, Tier II proteomics will begin to be applied to four toxicoproteomic problem areas: (1) global and targeted protein phosphorylation [115–117], (a) chemoproteomics [118] using pharmaceuticals or enzyme substrates like ATP [119] as mass capture ligands for proteins, (3) protein–protein interactions via pull-down approaches [120, 121], and (4) the introduction of innovative, functional proteomic screens such as those recently reported for Fas ligand [122].

Fourth, an enormous opportunity for toxicoproteomics lies in exploiting accessible biofluids (i.e., serum/plasma, urine, and cerebrospinal fluid) for biomarker development [123]. These activities in toxicoproteomics might be combined with transcriptomic analysis of blood leukocytes for a parallel approach in biomarker discovery [124]. Furthermore, serum/plasma may be screened for autoantibodies responsible for adverse drug reactions by incubation of serum with protein autoantigens. Autoantibodies to aldolase B have been reported that might be partially responsible for troglitazone-induced liver dysfunction [125].

Fifth, the insightful use of genetically altered animals and cell models will enhance discovery of protein targets and mechanistic insights into adverse drug reactions. The following studies might serve as examples. Proteomic analysis has been used to study important protein differences in leptin deficient ob/ob mice [60] or growth hormone receptor knock-out mice [126]. Molecularly altered cell systems (i.e., siRNA, RNAi, gene mutated) are highly adaptable to toxicoproteomic analysis and discovery such as ATM deficient cell models used to find hnRNP as a p53-coactivator in DNA damage and repair [127]. Use of genetically susceptible animal strains [52] can be highly informative tools in toxicoproteomics such as the use of SJL-resistant and C57B1/6-susceptible mice to acetaminophen toxicity [52].

Finally, continued efforts for integration of proteomics, transcriptomics, and toxicology data to derive mechanistic insight and biomarkers will be a continuing goal to maximize return on the investment in omics technologies [24, 113, 128].

## 10 CONCLUSION

Challenges for toxicoproteomics in preclinical risk assessment are for use as a discovery tool for specific proteins affected by drug and toxicant action, for better understanding of biochemistry and cell biology, and for biomarker development. The discipline of proteome mapping will be a different and more complex enterprise from the high throughput, linear sequencing activities that have been so useful in mapping the human genome. While the mapping and measuring of attributes in any one proteome is a large undertaking, biofluid proteomes such as serum/plasma,

urine, and CSF hold the most immediate promise for preclinical assessment in terms of better biomarkers.

Strides in improving proteomic technologies to map and measure proteomes and subproteomes are being made. However, no one proteomic platform appears ideally suited to quantify the broad range of protein expression in a given tissue, to distinguish the multiple forms and posttranslational modifications of proteins, to address the inadequate annotation of proteomes, or to accomplish integration of proteomic data with transcriptomic and metabolomic data. More than one proteomic platform may likely be needed for complex experimental problem areas. While 2D gels will always be a flexible and adaptable tool in toxicoproteomics, the ascendancy of liquid chromatography-based tandem mass spectrometry (LC-MS/MS) in its many and varied platforms will eclipse most gel-based proteomic platforms. Additionally, toxicoproteomic researchers should make effective use of transcriptomic data that in many cases may likely precede any toxicoproteomic analysis of tissues or blood. In fact, such transcriptomic studies may prove an invaluable guide to inform and guide toxicoproteomic studies of the relevant subproteome for analysis by identifying pathways and subcellular structures affected by toxicity as inferred by changes in the transcriptome. The further challenge lies in conducting the follow-up, hypothesis-driven studies that bring biological meaning to the data contained in lists of altered proteins, transcripts, and metabolites.

Although there are many challenges for toxicoproteomics in preclinical assessment, the opportunities are also close at hand for a greater understanding of toxicant action, the linkage to accompanying dysfunction and pathology, and the development of predictive biomarkers and signatures of toxicity.

## ACKNOWLEDGMENTS

This work was supported by the Intramural Research Program of the NIH, National Institute of Environmental Health Sciences.

## REFERENCES

1. Chapal N, Molina L, Molina F, Laplanche M, Pau B, Petit P. Pharmacoproteomic approach to the study of drug mode of action, toxicity, and resistance: applications in diabetes and cancer. *Fundam Clin Pharmacol* 2004;18:413–422.
2. Leighton JK. Application of emerging technologies in toxicology and safety assessment: regulatory perspectives. *Int J Toxicol* 2005;24:153–155.
3. Ross JS, Symmans WF, Pusztai L, Hortobagyi GN. Pharmacogenomics and clinical biomarkers in drug discovery and development. *Am J Clin Pathol* 2005;124(Suppl): S29–S41.
4. Siest G, Marteau JB, Maumus S, Berrahmoune H, Jeannesson E, Samara A, Batt AM, Visvikis-Siest S. Pharmacogenomics and cardiovascular drugs: need for integrated biological system with phenotypes and proteomic markers. *Eur J Pharmacol* 2005;527:1–22.
5. Bandara LR, Kennedy S. Toxicoproteomics—a new preclinical tool. *Drug Discov Today* 2002;7:411–418.

6. Petricoin EF, Rajapaske V, Herman EH, Arekani AM, Ross S, Johann D, Knapton A, Zhang J, Hitt BA, Conrads TP, Veenstra TD, Liotta LA, Sistare FD. Toxicoproteomics: serum proteomic pattern diagnostics for early detection of drug induced cardiac toxicities and cardioprotection. *Toxicol Pathol* 2004;32(Suppl 1):122–130.
7. Wetmore BA, Merrick BA. Toxicoproteomics: proteomics applied to toxicology and pathology. *Toxicol Pathol* 2004;32:619–642.
8. Waters MD, Fostel JM. Toxicogenomics and systems toxicology: aims and prospects. *Nat Rev Genet* 2004;5:936–948.
9. Turner SM. Stable isotopes, mass spectrometry, and molecular fluxes: applications to toxicology. *J Pharmacol Toxicol Methods* 2006;53:75–85.
10. Kasper P, Oliver G, Lima BS, Singer T, Tweats D. Joint EFPIA/CHMP SWP workshop: the emerging use of omic technologies for regulatory non-clinical safety testing. *Pharmacogenomics* 2005;6:181–184.
11. Gibbs A. Comparison of the specificity and sensitivity of traditional methods for assessment of nephrotoxicity in the rat with metabonomic and proteomic methodologies. *J Appl Toxicol* 2005;25:277–295.
12. Keller LC, Romijn EP, Zamora I, Yates JR, 3rd, Marshall WF. Proteomic analysis of isolated chlamydomonas centrioles reveals orthologs of ciliary-disease genes. *Curr Biol* 2005;15:1090–1098.
13. Kyttala M, Tallila J, Salonen R, Kopra O, Kohlschmidt N, Paavola-Sakki P, Peltonen LKestila M. MKS1, encoding a component of the flagellar apparatus basal body proteome, is mutated in Meckel syndrome. *Nat Genet* 2006;38:155–157.
14. Hackett JL, Gutman SI. Introduction to the Food and Drug Administration (FDA) regulatory process. *J Proteome Res* 2005;4:1110–1113.
15. MacGregor JT. The future of regulatory toxicology: impact of the biotechnology revolution. *Toxicol Sci* 2003;75:236–248.
16. Yeh ET. High-sensitivity C-reactive protein as a risk assessment tool for cardiovascular disease. *Clin Cardiol* 2005;28:408–412.
17. Amacher DE. A toxicologist's guide to biomarkers of hepatic response. *Hum Exp Toxicol* 2002;21:253–262.
18. Ross JS, Fletcher JA, Linette GP, Stec J, Clark E, Ayers M, Symmans WF, Pusztai L, Bloom KJ. The Her-2/neu gene and protein in breast cancer 2003: biomarker and target of therapy. *Oncologist* 2003;8:307–325.
19. Thier R, Bruning T, Roos PH, Rihs HP, Golka K, Ko Y, Bolt HM. Markers of genetic susceptibility in human environmental hygiene and toxicology: the role of selected CYP, NAT and GST genes. *Int J Hyg Environ Health* 2003;206:149–171.
20. Bilello JA. The agony and ecstasy of "OMIC" technologies in drug development. *Curr Mol Med* 2005;5:39–52.
21. Merrick BA, Bruno ME. Genomic and proteomic profiling for biomarkers and signature profiles of toxicity. *Curr Opin Mol Ther* 2004;6:600–607.
22. Koop R. Combinatorial biomarkers: from early toxicology assays to patient population profiling. *Drug Discov Today* 2005;10:781–788.
23. Merrick BA. Introduction to high-throughput protein expression. In Hamadeh HK, Afshari CA, (Eds). *Toxicogenomics: Principles and Applications*. Hoboken, NJ: Wiley; 2004, pp 263–281.
24. Merrick BA, Madenspacher JH. Complementary gene and protein expression studies and integrative approaches in toxicogenomics. *Toxicol Appl Pharmacol* 2005; 207:189–194.

25. Righetti PG, Castagna A, Antonucci F, Piubelli C, Cecconi D, Campostrini N, Antonioli P, Astner H, Hamdan M. Critical survey of quantitative proteomics in two-dimensional electrophoretic approaches. *J Chromatogr A*, 2004;1051:3–17.
26. Freeman WM, Hemby SE. Proteomics for protein expression profiling in neuroscience. *Neurochem Res* 2004;29:1065–1081.
27. Yates JR. Mass spectral analysis in proteomics. *Annu Rev Biophys Biomol Struct* 2004;33:297–316.
28. Bhat VB, Choi MH, Wishnok JS, Tannenbaum SR. Comparative plasma proteome analysis of lymphoma-bearing SJL mice. *J Proteome Res* 2005;4:1814–1825.
29. Farkas D, Bhat VB, Mandapati S, Wishnok JS, Tannenbaum SR. Characterization of the secreted proteome of rat hepatocytes cultured in collagen sandwiches. *Chem Res Toxicol* 2005;18:1132–1139.
30. Macdonald N, Chevalier S, Tonge R, Davison M, Rowlinson R, Young J, Rayner S, Roberts R. Quantitative proteomic analysis of mouse liver response to the peroxisome proliferator diethylhexylphthalate (DEHP) *Arch Toxicol* 2001;75:415–424.
31. Petricoin E, Wulfkuhle J, Espina V, Liotta LA. Clinical proteomics: revolutionizing disease detection and patient tailoring therapy. *J Proteome Res* 2004;3:209–217.
32. Issaq HJ, Conrads TP, Prieto DA, Tirumalai R, Veenstra TD. SELDI-TOF MS for diagnostic proteomics. *Anal Chem* 2003;75:148A–155A.
33. Diamandis EP. Mass spectrometry as a diagnostic and a cancer biomarker discovery tool: opportunities and potential limitations. *Mol Cell Proteomics* 2004;3:367–378.
34. Cutler P. Protein arrays: the current state-of-the-art. *Proteomics* 2003;3:3–18.
35. Park BK, Kitteringham NR, Maggs JL, Pirmohamed M, Williams DP. The role of metabolic activation in drug-induced hepatotoxicity. *Annu Rev Pharmacol Toxicol* 2005;45:177–202.
36. Kaplowitz N. Drug-induced liver injury. *Clin Infect Dis* 2004;38(Suppl 2):S44–S48.
37. Maddrey WC. Drug-induced hepatotoxicity: 2005. *J Clin Gastroenterol* 2005;39:S83–S89.
38. Steiner G, Suter L, Boess F, Gasser R, de Vera MC, Albertini S, Ruepp S. Discriminating different classes of toxicants by transcript profiling. *Environ Health Perspect* 2004;112:1236–1248.
39. Kon K, Kim JS, Jaeschke H, Lemasters JJ. Mitochondrial permeability transition in acetaminophen-induced necrosis and apoptosis of cultured mouse hepatocytes. *Hepatology* 2004;40:1170–1179.
40. Amacher DE. Drug-associated mitochondrial toxicity and its detection. *Curr Med Chem* 2005;12:1829–1839.
41. Ishida Y, Kondo T, Tsuneyama K, Lu P, Takayasu T, Mukaida N. The pathogenic roles of tumor necrosis factor receptor p55 in acetaminophen-induced liver injury in mice. *J Leukoc Biol* 2004;75:59–67.
42. Ito Y, Bethea NW, Abril ER, McCuskey RS. Early hepatic microvascular injury in response to acetaminophen toxicity. *Microcirculation* 2003;10:391–400.
43. James LP, Simpson PM, Farrar HC, Kearns GL, Wasserman GS, Blumer JL, Reed MD, Sullivan JE, Hinson JA. Cytokines and toxicity in acetaminophen overdose. *J Clin Pharmacol* 2005;45:1165–1171.
44. Laskin DL, Laskin JD. Role of macrophages and inflammatory mediators in chemically induced toxicity. *Toxicology* 2001;160:111–118.
45. Liu ZX, Govindarajan S, Kaplowitz N. Innate immune system plays a critical role in determining the progression and severity of acetaminophen hepatotoxicity. *Gastroenterology* 2004;127:1760–1774.

46. Fountoulakis M, Berndt P, Boelsterli UA, Crameri F, Winter M, Albertini S, Suter L. Two-dimensional database of mouse liver proteins: changes in hepatic protein levels following treatment with acetaminophen or its nontoxic regioisomer 3-acetamidophenol. *Electrophoresis* 2000;21:2148–2161.
47. Tonge R, Shaw J, Middleton B, Rowlinson R, Rayner S, Young J, Pognan F, Hawkins E, Currie I, Davison M. Validation and development of fluorescence two-dimensional differential gel electrophoresis proteomics technology. *Proteomics* 2001;1:377–396.
48. Ruepp SU, Tonge RP, Shaw J, Wallis N, Pognan F. Genomics and proteomics analysis of acetaminophen toxicity in mouse liver. *Toxicol Sci* 2002;65:135–150.
49. Thome-Kromer B, Bonk I, Klatt M, Nebrich G, Taufmann M, Bryant S, Wacker U, Kopke A. Toward the identification of liver toxicity markers: a proteome study in human cell culture and rats. *Proteomics* 2003;3:1835–1862.
50. Kikkawa R, Yamamoto T, Fukushima T, Yamada H, Horii I. Investigation of a hepatotoxicity screening system in primary cell cultures—what biomarkers would need to be addressed to estimate toxicity in conventional and new approaches? *J Toxicol Sci* 2005;30:61–72.
51. Yamamoto T, Kikkawa R, Yamada H, Horii I. Identification of oxidative stress-related proteins for predictive screening of hepatotoxicity using a proteomic approach. *J Toxicol Sci* 2005;30:213–227.
52. Welch KD, Wen B, Goodlett DR, Yi EC, Lee H, Reilly TP, Nelson SD, Pohl LR. Proteomic identification of potential susceptibility factors in drug-induced liver disease. *Chem Res Toxicol* 2005;18:924–933.
53. Lee H, Yi EC, Wen B, Reily TP, Pohl L, Nelson S, Aebersold R, Goodlett DR. Optimization of reversed-phase microcapillary liquid chromatography for quantitative proteomics. *J Chromatogr B Analyt Technol Biomed Life Sci* 2004;803:101–110.
54. Weber LW, Boll M, Stampfl A. Hepatotoxicity and mechanism of action of haloalkanes: carbon tetrachloride as a toxicological model. *Crit Rev Toxicol* 2003;33:105–136.
55. Liu Y, Liu P, Liu CH, Hu YY, Xu LM, Mu YP, Du GL. [Proteomic analysis of proliferation and apoptosis in carbon tetrachloride induced rat liver fibrosis]. *Zhonghua Gan Zang Bing Za Zhi* 2005;13:563–566.
56. Nelson SD. Mechanisms of the formation and disposition of reactive metabolites that can cause acute liver injury. *Drug Metab Rev* 1995;27:147–177.
57. Heijne WH, Stierum RH, Slijper M, van Bladeren PJ, Van Ommen B. Toxicogenomics of bromobenzene hepatotoxicity: a combined transcriptomics and proteomics approach. *Biochem Pharmacol* 2003;65:857–875.
58. Staels B, Fruchart JC. Therapeutic roles of peroxisome proliferator-activated receptor agonists. *Diabetes* 2005;54:2460–2470.
59. Iida M, Anna CH, Hartis J, Bruno M, Wetmore B, Dubin JR, Sieber S, Bennett L, Cunningham ML, Paules RS, Tomer KB, Houle CD, Merrick AB, Sills RC, Devereux TR. Changes in global gene and protein expression during early mouse liver carcinogenesis induced by non-genotoxic model carcinogens oxazepam and Wyeth-14,643. *Carcinogenesis* 2003;24:757–770.
60. Edvardsson U, von Lowenhielm HB, Panfilov O, Nystrom AC, Nilsson F, Dahllof B. Hepatic protein expression of lean mice and obese diabetic mice treated with peroxisome proliferator-activated receptor activators. *Proteomics* 2003;3:468–478.
61. Richards VE, Chau B, White MR, McQueen CA. Hepatic gene expression and lipid homeostasis in C57BL/6 mice exposed to hydrazine or acetylhydrazine. *Toxicol Sci* 2004;82:318–332.

62. Kleno TG, Kiehr B, Baunsgaard D, Sidelmann UG. Combination of "omics" data to investigate the mechanism(s) of hydrazine-induced hepatotoxicity in rats and to identify potential biomarkers. *Biomarkers* 2004;9:116–138.
63. Kleno TG, Leonardsen LR, Kjeldal HO, Laursen SM, Jensen ON, Baunsgaard D. Mechanisms of hydrazine toxicity in rat liver investigated by proteomics and multivariate data analysis. *Proteomics* 2004;4:868–880.
64. Chilakapati J, Shankar K, Korrapati MC, Hill RA, Mehendale HM. Saturation toxicokinetics of thioacetamide: role in initiation of liver injury. *Drug Metab Dispos* 2005;33:1877–1885.
65. Low TY, Leow CK, Salto-Tellez M, Chung MC. A proteomic analysis of thioacetamide-induced hepatotoxicity and cirrhosis in rat livers. *Proteomics* 2004;4:3960–3974.
66. Xu XQ, Leow CK, Lu X, Zhang X, Liu JS, Wong WH, Asperger A, Deininger S, Eastwood Leung HC. Molecular classification of liver cirrhosis in a rat model by proteomics and bioinformatics. *Proteomics* 2004;4:3235–3245.
67. Kawada N, Kristensen DB, Asahina K, Nakatani K, Minamiyama Y, Seki S, Yoshizato K. Characterization of a stellate cell activation-associated protein (STAP) with peroxidase activity found in rat hepatic stellate cells. *J Biol Chem* 2001;276:25318–25323.
68. Omenn GS, States DJ, Adamski M, Blackwell TW, Menon R, Hermjakob H, Apweiler R, Haab BB, Simpson RJ, Eddes JS, Kapp EA, Moritz RL, Chan DW, Rai AJ, Admon A, Aebersold R, Eng J, Hancock WS, Hefta SA, Meyer H, Paik YK, Yoo JS, Ping P, Pounds J, Adkins J, Qian X, Wang R, Wasinger V, Wu CY, Zhao X, Zeng R, Archakov A, Tsugita A, Beer I, Pandey A, Pisano M, Andrews P, Tammen H, Speicher DW, Hanash SM. Overview of the HUPO Plasma Proteome Project: results from the pilot phase with 35 collaborating laboratories and multiple analytical groups, generating a core dataset of 3020 proteins and a publicly-available database. *Proteomics* 2005;5:3226–3245.
69. Ping P, Vondriska TM, Creighton CJ, Gandhi TK, Yang Z, Menon R, Kwon MS, Cho SY, Drwal G, Kellmann M, Peri S, Suresh S, Gronborg M, Molina H, Chaerkady R, Rekha B, Shet AS, Gerszten RE, Wu H, Raftery M, Wasinger V, Schulz-Knappe P, Hanash SM, Paik YK, Hancock WS, States DJ, Omenn GS, Pandey A. A functional annotation of subproteomes in human plasma. *Proteomics* 2005;5:3506–3519.
70. Duan X, Yarmush DM, Berthiaume F, Jayaraman A, Yarmush ML. A mouse serum two-dimensional gel map: application to profiling burn injury and infection. *Electrophoresis* 2004;25:3055–3065.
71. Gianazza E, Eberini I, Villa P, Fratelli M, Pinna C, Wait R, Gemeiner M, Miller I. Monitoring the effects of drug treatment in rat models of disease by serum protein analysis. *J Chromatogr B Analyt Technol Biomed Life Sci* 2002;771:107–130.
72. Duan X, Yarmush D, Berthiaume F, Jayaraman A, Yarmush ML. Immunodepletion of albumin for two-dimensional gel detection of new mouse acute-phase protein and other plasma proteins. *Proteomics* 2005;5:3991–4000.
73. Wait R, Chiesa G, Parolini C, Miller I, Begum S, Brambilla D, Galluccio L, Ballerio R, Eberini I, Gianazza E. Reference maps of mouse serum acute-phase proteins: changes with LPS-induced inflammation and apolipoprotein A-I and A-II transgenes. *Proteomics* 2005;5:4245–4253.
74. Amacher DE, Adler R, Herath A, Townsend RR. Use of proteomic methods to identify serum biomarkers associated with rat liver toxicity or hypertrophy. *Clin Chem* 2005;51:1796–1803.
75. Witzmann FA, Li J. Proteomics and nephrotoxicity. *Contrib Nephrol* 2004;141:104–123.
76. Davis JW, Kramer JA. Genomic-based biomarkers of drug-induced nephrotoxicity. *Expert Opin Drug Metab Toxicol* 2006;2:95–101.

77. Mihatsch MJ, Thiel G, Ryffel B. Cyclosporin A: action and side-effects. *Toxicol Lett* 1989;46:125–139.
78. Benito B, Wahl D, Steudel N, Cordier A, Steiner S. Effects of cyclosporine A on the rat liver and kidney protein pattern, and the influence of vitamin E and C coadministration. *Electrophoresis* 1995;16:1273–1283.
79. Steiner S, Aicher L, Raymackers J, Meheus L, Esquer-Blasco R, Anderson NL, Cordier A. Cyclosporine A decreases the protein level of the calcium-binding protein calbindin-D 28 kDa in rat kidney. *Biochem Pharmacol* 1996;51:253–258.
80. Lee CT, Huynh VM, Lai LW, Lien YH. Cyclosporine A-induced hypercalciuria in calbindin-D28k knockout and wild-type mice. *Kidney Int* 2002;62:2055–2061.
81. Serkova N, Christians U. Transplantation: toxicokinetics and mechanisms of toxicity of cyclosporine and macrolides. *Curr Opin Investig Drugs* 2003;4:1287–1296.
82. Mascarell L, Frey JR, Michel F, Lefkovits I, Truffa-Bachi P. Increased protein synthesis after T cell activation in presence of cyclosporin A. *Transplantation* 2000;70:340–348.
83. Mascarell L, Auger R, Alcover A, Ojcius DM, Jungas T, Cadet-Daniel V, Kanellopoulos JM, Truffa-Bachi P. Characterization of a gene encoding two isoforms of a mitochondrial protein up-regulated by cyclosporin A in activated T cells. *J Biol Chem* 2004;279: 10556–10563.
84. Guan N, Ding J, Deng J, Zhang J, Yang J. Key molecular events in puromycin aminonucleoside nephrosis rats. *Pathol Int* 2004;54:703–711.
85. Sundin DP, Meyer C, Dahl R, Geerdes A, Sandoval R, Molitoris BA. Cellular mechanism of aminoglycoside tolerance in long-term gentamicin treatment. *Am J Physiol* 1997;272:C1309–C1318.
86. Witzmann FA, Fultz CD, Grant RA, Wright LS, Kornguth SE, Siegel FL. Differential expression of cytosolic proteins in the rat kidney cortex and medulla: preliminary proteomics. *Electrophoresis* 1998;19:2491–2497.
87. Shakib K, Norman JT, Fine LG, Brown LR, Godovac-Zimmermann J. Proteomics profiling of nuclear proteins for kidney fibroblasts suggests hypoxia, meiosis, and cancer may meet in the nucleus. *Proteomics* 2005;5:2819–2838.
88. Charlwood J, Skehel JM, King N, Camilleri P, Lord P, Bugelski P, Atif U. Proteomic analysis of rat kidney cortex following treatment with gentamicin. *J Proteome Res* 2002;1:73–82.
89. Thongboonkerd V, Klein JB, Arthur JM. Proteomic identification of a large complement of rat urinary proteins. *Nephron Exp Nephrol* 2003;95:e69–e78.
90. Cutler P, Bell DJ, Birrell HC, Connelly JC, Connor SC, Holmes E, Mitchell BC, Monte SY, Neville BA, Pickford R, Polley S, Schneider K, Skehel JM. An integrated proteomic approach to studying glomerular nephrotoxicity. *Electrophoresis* 1999;20:3647–3658.
91. Crowe CA, Yong AC, Calder IC, Ham KN, Tange JD. The nephrotoxicity of *p*-aminophenol. I. The effect on microsomal cytochromes, glutathione and covalent binding in kidney and liver. *Chem Biol Interact* 1979;27:235–243.
92. Kaltenbach JP, Carone FA, Ganote CE. Compounds protective against renal tubular necrosis induced by D-serine and D-2,3-diaminopropionic acid in the rat. *Exp Mol Pathol* 1982;37:225–234.
93. Kuhlmann MK, Horsch E, Burkhardt G, Wagner M, Kohler H. Reduction of cisplatin toxicity in cultured renal tubular cells by the bioflavonoid quercetin. *Arch Toxicol* 1998;72:536–540.
94. Bandara LR, Kelly MD, Lock EA, Kennedy S. A correlation between a proteomic evaluation and conventional measurements in the assessment of renal proximal tubular toxicity. *Toxicol Sci* 2003;73:195–206.

95. Bandara LR, Kelly MD, Lock EA, Kennedy S. A potential biomarker of kidney damage identified by proteomics: preliminary findings. *Biomarkers* 2003;8:272–286.
96. Chen JC, Stevens JL, Trifillis AL, Jones TW. Renal cysteine conjugate beta-lyase-mediated toxicity studied with primary cultures of human proximal tubular cells. *Toxicol Appl Pharmacol* 1990;103:463–473.
97. Lash LH, Qian W, Putt DA, Hueni SE, Elfarra AA, Krause RJ, Parker JC. Renal and hepatic toxicity of trichloroethylene and its glutathione-derived metabolites in rats and mice: sex-, species-, and tissue-dependent differences. *J Pharmacol Exp Ther* 2001;297:155–164.
98. Gao J, Ann Garulacan L, Storm SM, Hefta SA, Opiteck GJ, Lin JH, Moulin F, Dambach DM. Identification of *in vitro* protein biomarkers of idiosyncratic liver toxicity. *Toxicol In Vitro* 2004;18:533–541.
99. Meneses-Lorente G, Guest PC, Lawrence J, Muniappa N, Knowles MR, Skynner HA, Salim K, Cristea I, Mortishire-Smith R, Gaskell SJ, Watt A. A proteomic investigation of drug-induced steatosis in rat liver. *Chem Res Toxicol* 2004;17:605–612.
100. Sheta EA, Appel SH, Goldknopf IL. 2D gel blood serum biomarkers reveal differential clinical proteomics of the neurodegenerative diseases. *Expert Rev Proteomics* 2006;3:45–62.
101. Vercauteren FG, Bergeron JJ, Vandesande F, Arckens L, Quirion R. Proteomic approaches in brain research and neuropharmacology. *Eur J Pharmacol* 2004;500:385–398.
102. Merten KE, Feng W, Zhang L, Pierce W, Cai J, Klein JB, Kang YJ. Modulation of cytochrome C oxidase-va is possibly involved in metallothionein protection from doxorubicin cardiotoxicity. *J Pharmacol Exp Ther* 2005;315:1314–1319.
103. Elased KM, Cool DR, Morris M. Novel mass spectrometric methods for evaluation of plasma angiotensin converting enzyme 1 and renin activity. *Hypertension* 2005; 46:953–959.
104. Yamamoto T, Fukushima T, Kikkawa R, Yamada H, Horii I. Protein expression analysis of rat testes induced testicular toxicity with several reproductive toxicants. *J Toxicol Sci* 2005;30:111–126.
105. Yang YH, Xi ZG, Chao FH, Yang DF. Effects of formaldehyde inhalation on lung of rats. *Biomed Environ Sci* 2005;18:164–168.
106. Wheelock AM, Boland BC, Isbell M, Morin D, Wegesser TC, Plopper CG, Buckpitt AR. *In vivo* effects of ozone exposure on protein adduct formation by 1-nitronaphthalene in rat lung. *Am J Respir Cell Mol Biol* 2005;33:130–137.
107. Wheelock AM, Zhang L, Tran MU, Morin D, Penn S, Buckpitt AR, Plopper CG. Isolation of rodent airway epithelial cell proteins facilitates *in vivo* proteomics studies of lung toxicity. *Am J Physiol Lung Cell Mol Physiol* 2004;286:L399–L410.
108. de Graauw M, TijDens I, Cramer R, Corless S, Timms JF, van De Water B. Heat shock protein 27 is the major differentially phosphorylated protein involved in renal epithelial cellular stress response and controls focal adhesion organization and apoptosis. *J Biol Chem* 2005;280:29885–29898.
109. Rauch A, Bellew M, Eng J, Fitzgibbon M, Holzman T, Hussey P, Igra M, Maclean B, Lin CW, Detter A, Fang R, Faca V, Gafken P, Zhang H, Whitaker J, States D, Hanash S, Paulovich A, McIntosh MW. Computational Proteomics Analysis System (CPAS): an extensible, open-source analytic system for evaluating and publishing proteomic data and high throughput biological experiments. *J Proteome Res* 2006;5:112–121.
110. Orchard S, Hermjakob H, Julian RK Jr, Runte K, Sherman D, Wojcik J, Zhu W, Apweiler R. Common interchange standards for proteomics data: public availability of tools and schema. *Proteomics* 2004;4:490–491.

111. Hamady M, Cheung TH, Resing K, Cios KJ, Knight R. Key challenges in proteomics and proteoinformatics. Progress in proteins. *IEEE Eng Med Biol Mag* 2005;24:34–40.
112. Ekins S, Nikolsky Y, Nikolskaya T. Techniques: application of systems biology to absorption, distribution, metabolism, excretion and toxicity. *Trends Pharmacol Sci* 2005; 26:202–209.
113. Quackenbush J. Extracting meaning from functional genomics experiments. *Toxicol Appl Pharmacol* 2005;207:195–199.
114. Bodnar WM, Blackburn RK, Krise JM, Moseley MA. Exploiting the complementary nature of LC/MALDI/MS/MS and LC/ESI/MS/MS for increased proteome coverage. *J Am Soc Mass Spectrom* 2003;14:971–979.
115. Collins MO, Yu L, Husi H, Blackstock WP, Choudhary JS, Grant SG. Robust enrichment of phosphorylated species in complex mixtures by sequential protein and peptide metal-affinity chromatography and analysis by tandem mass spectrometry. *Sci STKE* 2005;16.
116. Kim SY, Chudapongse N, Lee SM, Levin MC, Oh JT, Park HJ, Ho IK. Proteomic analysis of phosphotyrosyl proteins in the rat brain: effect of butorphanol dependence. *J Neurosci Res* 2004;77:867–877.
117. Wang M, Xiao GG, Li N, Xie Y, Loo JA, Nel AE. Use of a fluorescent phosphoprotein dye to characterize oxidative stress-induced signaling pathway components in macrophage and epithelial cultures exposed to diesel exhaust particle chemicals. *Electrophoresis* 2005;26:2092–2108.
118. Gagna CE, Winokur D, Clark Lambert W. Cell biology, chemogenomics and chemoproteomics. *Cell Biol Int* 2004;28:755–764.
119. Beillard E, Witte ON. Unraveling kinase signaling pathways with chemical genetic and chemical proteomic approaches. *Cell Cycle* 2005;4:434–437.
120. Brandt M, Madsen JC, Bunkenborg J, Jensen ON, Gammeltoft S, Jensen KJ. On-bead chemical synthesis and display of phosphopeptides for affinity pull-down proteomics. *Chembiochem* 2006;7:623–630.
121. Uhlen M, Ponten F. Antibody-based proteomics for human tissue profiling. *Mol Cell Proteomics* 2005;4:384–393.
122. Hauptschein RS, Sloan KE, Torella C, Moezzifard R, Giel-Moloney M, Zehetmeier C, Unger C, Ilag LL, Jay DG. Functional proteomic screen identifies a modulating role for CD44 in death receptor-mediated apoptosis. *Cancer Res* 2005;65:1887–1896.
123. Gao J, Garulacan LA, Storm SM, Opiteck GJ, Dubaquie Y, Hefta SA, Dambach DM, Dongre AR. Biomarker discovery in biological fluids. *Methods* 2005;35:291–302.
124. Merrick BA, Tomer KB. Toxicoproteomics: a parallel approach to identifying biomarkers. *Environ Health Perspect* 2003;111:A578–A579.
125. Maniratanachote R, Shibata A, Kaneko S, Yamamori I, Wakasugi T, Sawazaki T, Katoh K, Tokudome S, Nakajima M, Yokoi T. Detection of autoantibody to aldolase B in sera from patients with troglitazone-induced liver dysfunction. *Toxicology* 2005;216:15–23.
126. Beyea JA, Sawicki G, Olson DM, List E, Kopchick JJ, Harvey S. Growth hormone (GH) receptor knockout mice reveal actions of GH in lung development. *Proteomics* 2006;6:341–348.
127. Moumen A, Masterson P, O'Connor MJ, Jackson SP. hnRNP K: an HDM2 target and transcriptional coactivator of p53 in response to DNA damage. *Cell* 2005; 123:1065–1078.
128. Fostel J, Choi D, Zwickl C, Morrison N, Rashid A, Hasan A, Bao W, Richard A, Tong W, Bushel PR, Brown R, Bruno M, Cunningham ML, Dix D, Eastin W, Frade C, Garcia A, Heinloth A, Irwin R, Madenspacher J, Merrick BA, Papoian T, Paules R, Rocca-Serra

P, Sansone AS, Stevens J, Tomer K, Yang C, Waters M. Chemical effects in biological systems—data dictionary (CEBS-DD): a compendium of terms for the capture and integration of biological study design description, conventional phenotypes, and omics data. *Toxicol Sci* 2005;88:585–601.

129. Kristensen DB, Kawada N, Imamura K, Miyamoto Y, Tateno C, Seki S, Kuroki T, Yoshizato K. Proteome analysis of rat hepatic stellate cells. *Hepatology* 2000;32:268–277.
130. Reinheckel T, Korn S, Mohring S, Augustin W, Halangk W, Schild L. Adaptation of protein carbonyl detection to the requirements of proteome analysis demonstrated for hypoxia/reoxygenation in isolated rat liver mitochondria. *Arch Biochem Biophys* 2000;376:59–65.
131. Witzmann FA, Fultz CD, Grant RA, Wright LS, Kornguth SE, Siegel FL. Regional protein alterations in rat kidneys induced by lead exposure. *Electrophoresis* 1999;20:943–951.

# REGULATORY CONSIDERATIONS

Evan B. Siegel[1] and Duane B. Lakings[2]
[1]*Ground Zero Pharmaceuticals, Inc., Irvine, California*
[2]*Drug Safety Evaluation Consulting, Inc., Elgin, Texas*

**Contents**

1 The Strategic Preclinical Plan: An Integral Element in the FDA Pre-IND Consultation Process
2 Pharmacology
  2.1 Primary and Secondary Pharmacodynamics
  2.2 Safety Pharmacology
3 Pharmacokinetics
  3.1 Methods of Analysis: Bioanalytical Chemistry Method Development and Validation
  3.2 ADME
  3.3 Pharmacokinetic Drug–Drug and Drug–Food Interactions
4 Toxicology
  4.1 Single-Dose Toxicology
  4.2 Repeat-Dose Toxicology
  4.3 Genotoxicity
  4.4 Carcinogenicity
  4.5 Reproductive and Development Toxicity
  4.6 Local Tolerability
  4.7 Immunogenicity
  4.8 Immunotoxicity
  4.9 Other Toxicology Studies
5 Preclinical Conclusions
6 Pre-IND Manufacturing Issues
7 Pre-IND Clinical Issues
  References

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

## 1 THE STRATEGIC PRECLINICAL PLAN: AN INTEGRAL ELEMENT IN THE FDA PRE-IND CONSULTATION PROCESS

The advent of new approaches to small organic molecule (novel chemical entity or NCE) and macromolecule (biological) development, particularly in the biotechnological arena, has provided new opportunities for treatment and prevention of human disease, which has a major impact on patients, the healthcare system, and human productivity. The rapidly increasing cost (both in time and other resources) of new product development has led to unprecedented pressures on sponsors to move their programs from the discovery laboratory through preclinical testing and into human clinical trials.

Sponsors often find it difficult to incorporate formidable intellectual property into the realities of new product development, including the need to perform a logical, efficient, and targeted preclinical development program that effectively characterizes the drug-like attributes of drug candidates and may identify undesirable properties (i.e., demerits) that can prevent successful development. We see many new product failures in Phase III clinical trials, in some cases because poor planning and/or poorly designed preclinical work has been performed prior to the initiation of first-in-human or Phase I clinical trials. Since drug candidates that eventually fail during Phase II or Phase III clinical trials may have already cost sponsors $50–100 million (U.S.) or more, we need to ask ourselves whether a more intensive (not necessarily extensive) preclinical development testing plan could have resulted in early decisions to discontinue drug candidates in late discovery (i.e., discovery lead optimization or developability assessment), preclinical development, or early (Phase I) clinical phases, for which far smaller amounts of precious funding and other resources (personnel time and facilities) will have been spent. We suggest that sponsors, sponsor consultants, contract research organizations (CROs), and advisors undertake preclinical strategic planning for development programs before decisions are made (or not made) that might avoid these later clinical trial phase new product failures.

One approach is to enter into "pre-pre-IND" strategic preclinical discussions with the FDA, before beginning the formal pre-IND preclinical safety (toxicology and pharmacokinetic) studies that often lead to the officially sanctioned pre-IND and IND processes. Informal telephone conversations, submission of small amounts of summary pharmacology, pharmacokinetic, and toxicology data for FDA review, and skeletal proposals for strategic preclinical study plans can be accomplished by agreement with FDA Regulatory Project Managers and Reviewers. The FDA has launched its own initiatives to informally discuss new product development in the early preclinical stage, even before provisional products move into clinical trials via an IND. Sometimes a sponsor assumes that only limited animal pharmacology studies for "proof of principle" and perhaps one or two toxicology studies are needed prior to initiation of human clinical trials. In fact, the preclinical development program may require (usually does require) support from other study types, such as cross-reactivity, safety pharmacology, toxicokinetics, metabolism, genotoxicity and/or immunogenicity, and other development disciplines, such as manufacturing and clinical: in other words, a more sophisticated and detailed preclinical approach. In other situations, the sponsor may be under the misconception that he/she has to embark on a multimillion dollar preclinical development program, with the genera-

tion of reams of data and multiple publications, prior to planning a first-in-human clinical trial. By planning strategically, approaching the FDA early, receiving advice, and putting this advice into practice for a relatively modest expenditure of time and money, the overall development program for a drug candidate may well have a very smooth entry into preclinical development and/or early clinical development.

Since oncology NCEs and biologicals represent a significant percentage of those new products entering preclinical development and clinical trials and are projected to treat a life-threatening human disease (i.e., various types of cancer) for which safe and effective pharmaceutical products are not yet available, sponsors feel a great deal of urgency to move these novel drug candidates quickly into human clinical testing. This chapter uses oncology products as examples that can benefit from strategic preclinical development planning. The considerations included are generally applicable to all new product development programs and we cite regulatory guidances (FDA Guidance for Industry and International Conference on Harmonization (ICH) guidelines) that sponsors can (should) use for preparing a drug development plan and in the design of recommended preclinical studies for the effective characterization of drug candidates.

Several categories of preclinical issues and the associated studies to investigate these issues need to be considered for all NCE and biologicals development programs, including those involving biotechnologically derived active drug substances. A careful review of the specific "flags" for potential clinical safety that might be related to the particular chemical structure, mechanisms of action (if known), or other characteristics of the drug substance will allow the development team to focus on the preclinical development program that will best support a rapid movement of the NCE or biological into human trials. Key preclinical development considerations are discussed in the following sections, which are followed by short sections on preclinical development considerations for the manufacture of drug substance and proposed drug product and for the clinical aspects for submitting an IND for a first-in-human clinical trial.

## 2 PHARMACOLOGY

### 2.1 Primary and Secondary Pharmacodynamics

We recommend that sponsors consider the following when discovery and preclinical pharmacology development studies are designed and conducted to fully evaluate and characterize the pharmacological profiles of active antitumor agents:

- Antitumor animal models evaluated should explore the establishment of a disciplined dose–response relationship as well as the identification of a dose giving maximum biological activity, particularly if a safety margin for significant antitumor activity in animals cannot be established against the no observable adverse effect level (NOAEL) in the same animal species during toxicology studies (discussed later).
- A tissue specificity evaluation to determine if the active pharmaceutical ingredient (API) or drug candidate interacts with (i.e., binds to) specialized tissues such as vascular endothelium from various species, including humans. The

results from this research effort will assist in defining and confirming toxicology animal species that are relevant to humans as the development program unfolds.

- If the drug or biological candidate is expected to cause adverse effects in humans with inflammatory disorders and an appropriate animal model of inflammation is available, we recommend that the provisional product be evaluated in that model.

### 2.2 Safety Pharmacology

The ICH S7a guideline [1] recommends that a standard battery of safety pharmacology (SP) studies be conducted prior to the initiation of clinical trials on a new drug substance. This standard battery normally includes cardiovascular, central nervous system (CNS), and respiratory SP studies to investigate potential undesirable pharmacological effects of a drug substance on physiological functions in relation to exposure in the projected therapeutic range and above. The cardiovascular SP study should be designed to evaluate the potential to cause prolongation of the QT interval (as discussed in the ICH S7b guideline [2]), an area of concern for both the FDA and ICH primarily for NCEs but also for macromolecules.

For drug candidates intended to act in or be delivered via the gastrointestinal (GI) system (i.e., orally administered compounds), additional work should be performed exploring safety for the GI system, including effects on GI motility. Other SP studies that sponsors will need to consider (depending on the characteristics of the drug candidate) include effects on the renal/urinary system for drug candidates and/or their known metabolites cleared from the body by the kidney and on the autonomic nervous system for drug candidates that may have undesirable effects on the autonomic nerves. For drug candidates to be administered as part of a combination therapeutic regimen that contains an oncologic drug known to have adverse effects on the autonomic nervous system (i.e., platinum-containing compounds), SP studies should be designed to evaluate potential undesirable effects to the autonomic nervous system by the drug candidate alone and in combination with the agent known to produce adverse effects in this system.

## 3 PHARMACOKINETICS

We recommend that pharmacokinetic (PK) studies be conducted in the animal species to be evaluated in toxicology studies conducted in compliance with Good Laboratory Practice (GLP) regulations (as described in the Code of Federal Regulations, 21 CFR Part 58) and over the dose range proposed for these toxicology studies. These animal PK studies should be designed to evaluate the proposed clinical route of administration and, if possible, intravenous (IV) dosing to assess the absolute bioavailability (commonly referred to as $F$) of a drug candidate to be administered by a route other than IV. If appropriately designed, toxicokinetic (TK) evaluations during the subchronic and other toxicology studies may be sufficient to describe the PK profile of the NCE or biological in these animal species. This planning can save sponsors both time and funding. The ICH S3a guideline [3] on toxicokinetics can be referenced for further information on this topic. However, if a

drug candidate has limited absorption into systemic circulation after a non-IV route of administration, we recommend that sponsors design and conduct animal PK studies to evaluate the rate and extent of delivery (i.e., the delivery profile) and to obtain information on the delivery characteristics of possible drug candidate formulations. These animal PK studies should be conducted prior to starting the toxicology studies needed for supporting human clinical trials so that drug candidate formulation that provides the desired delivery profile (and the one proposed for use in human trials) can be evaluated in the toxicology studies, thus allowing these animal safety studies to support the proposed first-in-human clinical trial.

### 3.1 Methods of Analysis: Bioanalytical Chemistry Method Development and Validation

Bioanalytical chemistry (BAC) methods for the pharmacologically active drug substance (i.e., parent compound and/or its metabolites, if known) in physiological fluids (e.g., serum/plasma and urine) from each toxicology species (e.g., rat, dog, nonhuman primate) will need to be developed and appropriately validated to provide TK support for toxicology studies and to analyze specimens collected during animal PK evaluations. The developed BAC method, after appropriate intermatrix and interspecies validation, can also be used to support human clinical trials. The *FDA Guidance to Industry: Bioanalytical Method Validation* [4] provides information on the various aspects and parameters (e.g., specificity, sensitivity, linearity, accuracy, precision, and stability) considered necessary to appropriately validate a BAC method.

### 3.2 ADME

In general, nonclinical ADME studies include protein binding, mass balance (including metabolite profiling, isolation, and identification, and metabolite pathway definition and comparison for various species, including humans), single- and possibly multiple-dose tissue distribution in a rodent (usually rats) species, special TK studies to support reproductive and developmental toxicology (reproductive toxicity) studies in a rodent (usually rats) and nonrodent (commonly rabbits) species, evaluation of the potential inhibition or induction of various drug metabolizing enzymes (particularly CYP450 isozymes such as CYP3A4, CYP2C9, CYP2D6), and possibly PK evaluations of metabolites (especially metabolites that have pharmacological or toxicological activity). Many of these ADME studies require an appropriately radiolabeled drug candidate. Where possible, the sponsor should prepare, or have prepared, a carbon-14 radiolabeled drug candidate in order to conduct these nonclinical ADME studies and also to be used in a possible mass balance study in humans (which is considered necessary in order to obtain the physiological samples for comparing the extent of metabolism and the metabolic profiles of a drug candidate in the toxicology animal species and in humans). However, not all these nonclinical ADME studies are necessary for the preclinical (e.g., prior to the first-in-human clinical trial) evaluation of a drug candidate. Sponsors normally conduct *in vitro* metabolism and protein binding studies in the projected toxicology animal species and in humans and mass balance assessments with metabolite profiling in the rodent toxicology animal species to support an IND submission for a

first-in-human clinical trial and then conduct the other animal ADME studies on a drug candidate concurrently with clinical development.

***Absorption*** For small organic molecule and macromolecule drug candidates to be administered intravenously, absorption studies are not necessary. However, for NCEs with limited aqueous solubility and most macromolecules, studies to determine that the drug candidate is compatible with blood components are considered necessary, especially for a protein drug candidate to be administered intravenously.

If the proposed clinical route of administration for the provisional drug product is oral or another non-IV route (e.g., dermal, pulmonary, nasal, buccal), preclinical information on the rate and extent of absorption is usually necessary to show that the absorption profile is linear over the dose range to be evaluated in the toxicology studies, is not substantially different for each of the toxicology species, and is comparable to the absorption profile in humans (when human data are available). We recommend that the sponsor consider conducting an absolute bioavailability, dose proportionality study in each of the toxicology animal species (e.g., rat or other rodent and dog or nonhuman primate) and that the dose range evaluated include the projected pharmacological active dose for that species (if available) and the maximum projected dose level for toxicology assessments.

***Distribution*** Types of nonclinical distribution studies conducted on most drug candidates include protein binding and distribution in blood cells, single- and multiple-dose tissue distribution (TD), and feto-placental transfer and lacteal excretion studies. Preclinical distribution studies considered necessary to support an IND for a first-in-human clinical trial are protein binding and distribution into blood cells. Some sponsors also conduct single-dose TD studies with whole body autoradiography (WBA) prior to initiating first-in-human clinical trials but multiple-dose TD studies (if considered necessary as described in ICH S3b [5]) are conducted concurrently with human clinical studies. If the human disease to be evaluated during clinical trials is gender specific (females only) and/or pregnant or nursing women or women of child-bearing potential are to be studied in the first-in-human clinical trial, feto-placental transfer and lacteal excretion studies may be necessary prior to the initiation of clinical trials.

***Metabolism*** The similarity or difference in the extent of metabolism and possibly in the metabolites formed among the various toxicology animal species and humans should be explored to demonstrate the relevance of the animal model for human exposure. Thus, as part of the strategic preclinical development program planning, the sponsor should consider evaluating the extent of metabolism and *in vitro* metabolism profile of the active drug substance in animals and in humans. These *in vitro* metabolism studies can be conducted in microsomes, which express the various CYP450 isozymes, or hepatocytes, which express all the phase I (oxidation, reduction, hydrolysis) and phase II (conjugation) metabolism systems. If the drug candidate is extensively metabolized in *in vitro* systems from animals and humans, the sponsor may want to develop a metabolite profiling BAC method to determine if the metabolites formed are similar or different for the various species in order to select metabolically-relevant-to-human animal species for toxicology studies.

For a protein drug candidate, *in vitro* or *in vivo* metabolism studies are generally not considered necessary (as discussed in ICH S6 [6]) unless the biological contains not natural amino acids or other structural modifications. However, the potential of a macromolecule drug candidate to induce or inhibit drug metabolism enzymes (i.e., CYP450 isozymes) can be assessed from liver samples collected from the rodent and nonrodent animal species at necropsy during subchronic toxicology studies.

***Excretion*** The rate and routes of excretion for the drug candidate and its metabolites in the toxicology animal species and later in humans are determined during mass balance (MB) studies (using the radiolabeled drug candidate). These animal mass balance studies, especially in the toxicology rodent animal species, are usually completed prior to the submission of an IND for a first-in-human clinical trial. Excreta (urine, bile, feces), and in some cases expired air, collected over various time intervals and blood/plasma at various time points after dosing are collected and counted for total radioactivity or total drug-related material (parent compound and any formed metabolites) to determine the routes of excretion and the extent of excretion by each route of elimination. As development continues, the collected samples from MB studies are profiled for drug candidate metabolites using a developed BAC profiling method (usually HPLC with a radiochemical detector). These collected physiological matrices (or from *in vitro* metabolism systems if the formed metabolites in MB studies and the earlier conducted *in vitro* metabolism evaluations are the same) can also be used for isolating sufficient quantities of each metabolite for identification, which is commonly conducted concurrently with clinical trials.

If urine samples collected and analyzed for total radioactivity during MB studies show little or no renal elimination of drug-related material, additional excretion studies may be necessary to evaluate how the drug candidate and its metabolites are cleared from the body and to obtain the necessary physiological matrices (i.e., bile) for metabolite profiling. If substantial radioactivity is eliminated in the urine, additional studies to evaluate the extent of renal elimination may be necessary and a safety pharmacology study on the renal/urinary system may be warranted.

### 3.3 Pharmacokinetic Drug–Drug and Drug–Food Interactions

Prior to the initiation of clinical trials on drug candidates to be administered concurrently with another active agent (e.g., a novel anticancer agent plus an approved cytotoxic agent), we recommend that animal drug–drug interaction studies be conducted to determine if the PK profiles of the drug candidate and the other anticancer agent are (or are not) changed by the combination regimen. Similarly, information on the optimal time (in relation to meals) for administration of a drug candidate can be evaluated with an appropriately designed drug–food interaction study.

If a particular species (e.g., the nonhuman primate) is a relevant animal model for evaluating the safety and metabolism of a drug candidate, this animal species is also relevant for evaluating potential drug–drug and drug–food interactions and the desired results can be obtained at a substantial reduction in time and cost (compared

to that for human clinical trials). Thus, we recommend that the sponsor consider first conducting possibly needed human drug–drug or drug–food interaction studies in the key (i.e., relevant to humans) animal species. If no or only minimal interaction is found for a particular drug–drug combination or if food and the time of feeding do not adversely affect the extent of absorption or the absorption profile for the drug candidate, evaluations in humans would not be necessary. If substantial drug–drug or drug–food effects are noted in the nonhuman primate or other appropriate animal species, the results can be confirmed with relatively simple clinical trials.

## 4 TOXICOLOGY

In general, toxicology studies conducted prior to a first-in-human clinical trial are to establish a margin of safety for the first dose to be administered to humans and to identify potential tissues and organs of toxicity so that human clinical trials can be appropriately designed. Other animal safety aspects evaluated during preclinical development and recommended for completion prior to the first-in-human clinical trial include safety pharmacology (discussed earlier) and genotoxicity and local tolerance (see Sections 4.3 and 4.6). Preclinical approaches for evaluating the animal toxicology profile of NCEs and biologicals intended for monotherapy and combination/adjunctive therapy are considered next.

### 4.1 Single-Dose Toxicology

Lethal dose ($LD_{50}$) studies are no longer considered necessary (and are not recommended) and have been replaced by acute toxicology evaluations where dosing is continued until a dose level with marked toxicological effect is identified. Acute toxicology studies (as discussed in *FDA Guidance for Industry: Single Dose Acute Toxicity Testing for Pharmaceuticals* [7] and ICH M3 [8]) are conducted in two species, one of which is a nonrodent, and use both the proposed clinical route of administration and intravenous (if possible).

***Monotherapy*** For a drug candidate to be administered as a monotherapy, acute (single dose or the number of doses planned to be administered over a 24-hour period during clinical testing) toxicology studies in two species, one of which is a nonrodent, are generally considered necessary, unless there is ample evidence that a second species is inappropriate and would not be scientifically supportable. The acute studies should be designed to determine a no-observable-adverse-effect-level (NOAEL) and a dose that causes marked toxicity and to determine if the observed acute toxicity is reversible. The results of the acute studies should be available to assist the sponsor in the design and dose selection for subchronic toxicology studies, which should be conducted on the same animal species as completed acute toxicology studies.

***Combination Therapy*** Acute toxicology studies on the provisional experimental drug product administered concurrently with another anticancer therapy may be beneficial in showing that the toxicity profiles of the combination therapy are (or are not) substantially different from the profiles of either agent administered alone.

We recommend that combination therapy acute toxicology studies be conducted to assist in the design of the recommended subchronic combination therapy toxicology evaluations. Even when a drug candidate is to be part of a combination therapy, we recommend that the acute toxicology studies discussed previously for a monotherapy agent be conducted in order to establish the toxicological profile for the drug candidate.

### 4.2 Repeat-Dose Toxicology

***Monotherapy*** For an NCE or macromolecule as a monotherapy and to support single-dose administration to humans, subchronic toxicology studies in two species (one of which is a nonrodent) with a dosing duration of 14 days are recommended (as discussed in ICH M3 [8]). The duration of a proposed multiple-dose clinical trial should not be longer than that for the completed subchronic toxicology studies (i.e., a proposed clinical trial dosing duration of 30 days requires completed subchronic toxicology studies in two species with a duration of dosing of 30 days or longer). The high dose of the drug candidate for subchronic toxicology studies should be sufficient to produce toxic effects and to identify potential tissues and organs of toxicity while the low dose of the drug candidate should produce little or no toxic effect (i.e., NOAEL) in order to establish an acceptable safety margin for the dose levels to be evaluated in the first-in-human clinical trial.

***Combination Therapy*** Prior to initiating clinical trials on a provisional new product to be administered concurrently with another active agent, subchronic toxicology studies in two species (one of which is nonrodent) and using the proposed clinical route of administration, dosing frequency, and duration for each agent are generally necessary. Recommended dose groups for these combination therapy subchronic toxicology studies are vehicle control (no active pharmaceutical agent), active concomitant agent alone (at a dose above the planned clinical dose), and the experimental product at various dose levels with concurrent administration of the known therapy at the dose level evaluated for the active concomitant agent alone. The selection of dose levels should be based on previously conducted dose range-finding studies and/or an appropriately designed pilot study. An FDA (draft) Guidance to Industry [9] on the nonclinical safety evaluations for drug combinations is available. This guidance document provides information on the development recommendations when the combination includes (1) two or more previously marketed drugs, (2) one or more previously marketed drug plus one drug candidate, which could be either an NCE or a macromolecule, and (3) two or more drug candidates. We recommend that sponsors carefully review this draft FDA Guidance to Industry prior to beginning a preclinical development program wherein combination therapy will be studied in humans.

### 4.3 Genotoxicity

All small organic molecule drug substances (NCEs) need to be supported by the standard ICH battery of genotoxicity studies [10, 11], unless specific waiver is given by the FDA Reviewing Division. This standard battery of genotoxicity tests consists of the following:

1. A test for gene mutation in bacteria.
2. An *in vitro* assessment of chromosomal damage using mammalian cells or an *in vitro* mouse lymphoma $tk^{+/-}$ assay.
3. An *in vivo* test for chromosomal damage using rodent hematopoietic cells.

According to *FDA Guidance to Industry: Recommended Approaches to Integration of Genetic Toxicology Study Results* [12], the results from *in vitro* genetic toxicology studies should be available prior to the initiation of Phase I first-in-human clinical trials and the *in vivo* test can be conducted concurrently with clinical trials.

As discussed in ICH S6 [6], most biologicals (e.g., monoclonal antibodies or other proteins) do not require safety support from genotoxicity studies. However, if the macromolecule contains not natural amino acids or other structural modifications, the possible need for genotoxicity studies prior to first-in-human clinical trials should be addressed with the FDA during the pre-IND consultation.

### 4.4 Carcinogenicity

For most NCEs and then only under special circumstances, carcinogenicity studies are not initiated prior to the first-in-human clinical trial and thus are not considered as preclinical studies but are nonclinical studies. However, most small organic molecule development plans will need to include carcinogenicity studies in the rat and in the mouse, each of which are usually initiated after Phase II clinical trials have shown the drug candidate can mediate a human disease or disorder in patients without causing unacceptable toxicity in these patients. ICH S1a [13], S1b [14], and S1c [15, 16] provide guidance on the need for and dose selection for carcinogenicity studies. ICH S1b indicates that one rodent (usually the rat) carcinogenicity study plus one other study (commonly in a transgenic, $p53^{+/-}$ mouse model) that supplements the carcinogenicity study and provides additional information that is not readily available from the long-term assay can be used to support the development of a drug candidate. Thus, we recommend that careful thought be given in the strategic preclinical plan for negotiation with the FDA for conducting one carcinogenicity study plus one additional supportive study. These studies might be conducted during Phase III clinical development, thus deferring the significant cost until the provisional drug product shows enough promise in patients to justify the initiation of these long-term animal studies. A sponsor assumes some risk with this approach, particularly when the clinical development program is able to be truncated (i.e., substantial evidence of efficacy is observed during Phase II clinical trials, which is relatively common for many oncology drug candidates) and a marketing application can be submitted earlier than expected. In this case, a sponsor may be able to submit the completed carcinogenicity study(ies) during the formal NDA review, or as a postapproval, Phase IV commitment.

For macromolecule drug candidates, such as monoclonal antibodies or other proteins with no unusual structural characteristics [6], carcinogenicity studies are generally not necessary. We recommend that during the pre-IND consultation or another early meeting with the FDA, sponsors confirm that carcinogenicity studies for a biological will not be required.

### 4.5 Reproductive and Development Toxicity

Before pregnant or lactating women or women not practicing acceptable methods of birth control can be entered in clinical trials on a provisional drug product, completion of the recommended battery of reproductive and developmental toxicology studies, as delineated in the ICH S5a [17] and S5b [18] guidelines, will normally be necessary. These guidelines indicate that the rat, or other appropriate mammalian species, and a second mammalian species (most commonly the rabbit) should be used for embryotoxicity studies. The timing of inclusion of these studies is dependent on the particular characteristics of the experimental drug substance and any species-specific issues, as well as the speed with which the development program is moving. For most, but not all, NCE and biological drug candidates, these reproductive and developmental toxicology studies are not conducted during preclinical development but are initiated after Phase I clinical trial results are available and are frequently conducted concurrently with early Phase II clinical trials. As noted in Section 3 on pharmacokinetics, a question usually asked when reproductive toxicology studies show no apparent adverse effect to the fetus is whether the fetus was appropriately exposed to the test article and/or its metabolites. To obtain this information, sponsors usually conduct feto-placental transfer studies (using a radiolabeled drug candidate to demonstrate exposure of the fetus to drug-related material and not just the parent compound) in the rodent and nonrodent species used in reproductive toxicology studies. The results of these TK studies should also be available prior to the entry of women of childbearing potential and pregnant women into clinical trials on the provisional drug product.

### 4.6 Local Tolerability

Local tolerability assessments on a provisional experimental drug product can be, and frequently are, obtained during acute, subchronic, and chronic toxicology studies. Separate local tolerance studies in animals may not be necessary. However, sponsors need to ensure that the local tolerance evaluations are generated using the same provisional drug product that will be studied in clinical trials. For some formulation types, particularly for dermal, pulmonary, and nasal formulations, the extent of local exposure, and thus local tolerability, may be determined by the formulation type, the excipients in the formulation, and the rate of release of the drug substance from the formulation. In these cases, the sponsor will need to conduct formulation studies early (i.e., prior to the initiation of toxicology testing) so that the local tolerance or intolerance of provisional drug product can be established. For NCE and biological drug candidates to be administrated by the IV, SC, or IM routes, local tolerance studies are important, particularly when the formulation is not similar to physiological conditions or when the drug candidate has limited aqueous solubility.

### 4.7 Immunogenicity

The potential for a biological, such as a chimeric monoclonal antibody, to cause an untoward immune response [6] in animals can be evaluated using samples collected during the subchronic toxicology studies and thus is commonly determined

during preclinical development. Appropriate assays to determine the ability of the drug substance to cause the stimulation of antibody formation *in vivo* will be necessary.

### 4.8 Immunotoxicity

As noted in ICH S6 [6], the ability of the drug candidate to adversely affect the immune system (i.e., immunotoxicity) should be evaluated early in the drug development process, commonly during preclinical development. If appropriately designed, these results can be obtained during the subchronic toxicology studies in rodent and nonrodent animal species that are conducted prior to the first-in-human clinical trial.

### 4.9 Other Toxicology Studies

Other toxicology studies commonly conducted during drug development include determining the toxicological profile of drug candidate metabolites, drug substance impurities, and novel formulation excipients. Since the need for these study types cannot always be defined during preclinical development, the strategic drug development plan should be updated as necessary. For drug candidate metabolites that occur only in humans (i.e., novel human metabolites) or are at substantially higher levels in humans than in animals, the *FDA Guidance for Industry* (draft): *Safety Testing of Drug Metabolities* [19] is available to assist sponsors.

If a drug candidate has potential application for a human disease or disorder in elderly and/or pediatric patients, we recommend that sponsors conduct at least some toxicology studies in old or young animals, respectively, to ascertain if the observed toxicity profile is age related. If the toxicity profile is age-related, sponsors will need to carefully design clinical trials in elderly and/or pediatric patients to ensure that the patient population is not exposed to toxic levels of the drug candidate.

## 5 PRECLINICAL CONCLUSIONS

In this chapter, we have described and discussed the preclinical drug development studies considered necessary to effectively characterize the pharmacology, pharmacokinetic, and toxicology profiles of drug candidates, both NCEs and macromolecules. We have also provided references to FDA guidances and ICH guidelines that can be used by sponsors to assist in designing these preclinical studies and to determine when these studies need to be conducted in the overall strategic drug development plan. Table 1 provides a summary of the types of preclinical studies that are considered necessary prior to a first-in-human clinical trial on a drug candidate. Since each drug development program is unique, the preclinical studies listed may not be required for all drug candidates and/or additional studies may be necessary to appropriately evaluate whether or not a drug candidate has the necessary attributes, and does not have major demerits, to be a successful clinical drug candidate.

**TABLE 1 Prxeclinical Drug Development Studies Commonly Conducted on NCE and Biological Drug Candidates**

| Study Type | Study Purpose and Goal | Reference |
|---|---|---|
| | *Pharmacology* | |
| Primary pharmacodynamics | *In vitro* and animal pharmacology studies on the mode of action and/or effects of a drug substance in relation to the therapeutic target. | |
| Safety pharmacology | Studies to evaluate the effects of a drug candidate on vital functions, such as the cardiovascular, central nervous, and respiratory systems (standard battery) and renal/urinary, autonomic nervous, gastrointestinal, and other organ systems (supplemental). | ICH S7a and S7b |
| | *Pharmacokinetics* | |
| Absorption | Studies to determine rate and extent of absorption of a drug candidate to be administered by a non-IV route. Studies may be *in vitro* and/or *in vivo* and results may be during toxicokinetic evaluations on a drug candidate. | ICH S3a |
| Distribution | *In vitro* studies to determine the extent of protein binding in animal species and in humans. | |
| Metabolism | *In vitro* studies to determine the extent of metabolism and possible drug metabolizing enzyme systems involved in the metabolism of a drug candidate in toxicology animal species and in humans. | |
| Excretion | Mass balance studies in the rodent and possibly the nonrodent toxicology animal species to determine the routes and extent of elimination. Profiling of drug candidate metabolites may be initiated. | |
| Drug–drug interaction | For drug candidates to be administered concurrently with other pharmaceuticals, animal interaction studies may be desirable to determine if the pharmacokinetic profiles of either agent are affected when coadministered. | |

**TABLE 1** ***Continued***

| Study Type | Study Purpose and Goal | Reference |
|---|---|---|
| | *Toxicology* | |
| Acute toxicology | Single-dose studies conducted in two species, one of which is a nonrodent, and using both the proposed clinical route of administration and intravenous (if possibly). Dosing should continue until evidence of marked toxic effect is observed. | ICH M3 and FDA Guidance to Industry |
| Repeat-dose toxicology | Repeat-dose studies in two species, one of which is not a rodent and using the proposed clinical route of administration and dosing frequency. Duration of dosing should be equal to or greater than that for the proposed clinical trails on the drug candidate. | ICH M3 and S4a |
| Genotoxicity | For NCEs, a standard battery of tests consisting of a test for gene mutation in bacteria, an *in vitro* assessment of chromosomal damage using mammalian cells or an *in vitro* mouse lymphoma $tk^{+/-}$ assay, and an *in vivo* test for chromosomal damage using rodent hematopoietic cells. For macromolecules, genotoxicity studies are usually not necessary. | ICH S2a, S2b, and S6 |
| Carcinogenicity | Not required during preclinical development for NCEs. Not normally required for macromolecules. | ICH S1a, S1b, and S1c |
| Reproductive and developmental toxicology | Not usually required during preclinical development. | ICH S5a and S5b |
| Local tolerance | Tolerability at the site of dosing for both NCEs and biologicals should be conducted using the route of administration proposed for clinical trials and formulation to be evaluated in humans. | ICH M3 and S6 |
| Immunogenicity and immunotoxicity | Studies recommended for macromolecules to evaluate the effects of the drug candidate on the immune system. | ICH S6 |
| Other toxicology studies | Toxicology assessments on drug candidate metabolites and/or impurities and novel formulation excipients. Not usually conducted during preclinical development. | FDA Guidance to Industry |

## 6 PRE-IND MANUFACTURING ISSUES

Prior to submitting an IND for a first-in-human clinical trial on a drug candidate, a sponsor will need to prepare the chemistry, manufacturing, and control (CMC) section (Item 7) of IND Form FDA-1571. This section consists of five subparts:

Subpart a: Manufacturing and controls information on the drug substance or active pharmaceutical ingredient (API)
Subpart b: Manufacturing and controls information on the proposed drug product
Subpart c: Manufacturing and controls information on the placebo to match the proposed drug product
Subpart d: Copies of labeling to be supplied for the clinical trials
Subpart e: Environmental impact analysis statement

CMC information on the API and proposed drug product in an IND should be of sufficient quality and detail to justify and support the use of the API during clinical development; should demonstrate that the sponsor utilized an API of similar quality during preclinical toxicology studies; and that the proposed drug product to be used in clinical trials was evaluated in animal safety studies. If the proposed drug product to be used in clinical trials was not evaluated in toxicology studies, sponsors will need to have conducted "bridging" studies to show that the rate and extent of absorption and the disposition profile for the proposed drug product are not significantly different from those for the formulated API used in the toxicology animal species and possibly in the pharmacology animal models, if the pharmacological response in animals was used in establishing the initial dose for the first-in-human clinical trial.

Information on the API should include the following:

1. The name and address of the facility where the API is manufactured.
2. The manufacturing procedure flowchart and a detailed description of production, including in-process controls.
3. Information regarding proof of structure, which can be obtained by elemental analysis, UV, IR, NMR spectra, and mass spectrum analysis.
4. Physical chemical properties, which include description and appearance, solubility profile, dissociation constant(s) (if any), optical rotation (if necessary), partition coefficient (such as octanol/water for log $P$ determination), UV absorbance, melting point and thermal analysis, X-ray powder diffraction pattern, and qualitative tests.
5. Raw materials, including solvents and water, used in the manufacture of the API.
6. Copies of release specifications or manufacturer's certificates of analysis (COA) for raw materials.
7. Release specifications for the API.
8. Analytical chemistry (AC) methods for raw materials and API release testing.

9. Reference Standard information.
10. Packaging and labeling information.
11. Stability results including methods, procedures, and system suitability for the stability-indicating assay.

For the manufacture of the API, a flowchart of synthesis is recommended and should provide information on the starting materials (with structures) for each step, catalysts and other reagents for each step, reaction conditions for each step, and products (with structures) for each step. Information commonly provided for the description of each step of the API synthesis includes equipment required (including suppliers and specifications), reactions (including amounts and order of addition), solvents, conditions (temperature, pressure, pH, time), tests for completion of reaction, isolation procedure for synthesis product (intermediate or API), purification steps for synthesis product, and yield of synthesis product.

Commonly employed API structure elucidation techniques include elemental analysis, mass spectroscopy, nuclear magnetic resonance, ultraviolet spectrum, infrared spectrum, molecular weight, stereochemistry (if the API has one or more chiral centers), X-ray analysis, and chromatography (usually HPLC for an NCE and SDS-PAGE for a macromolecule).

Sponsors are requested to provide information on the API impurity profile, which requires the development and validation of an impurity profiling AC method. The AC methods should provide chromatographic separation of API from all impurities and/or degradation products and should allow for calculation of the amount of each impurity and/or degradant. At this stage of development, sponsors should initiate a program for the isolation and identification of each impurity and/or degradant in the API. The API used in animal safety studies should have an impurity profile that is similar to the profile for the material to be used in the first-in-human clinical trial. If the clinical trial API material contains a novel impurity or an impurity at higher levels than present in the API material used in toxicology studies, sponsors should consider conducting toxicology studies on the clinical trial API material to ensure that an API impurity or degradant does not cause toxicity that might lead to adverse events or even serious adverse events in humans.

For the proposed drug product or the formulated API that is to be evaluated during the first-in-human clinical trial, sponsors should consider providing the following information:

1. The name and address of the facility where the proposed drug product is manufactured.
2. A detailed description of the manufacturing and packaging procedures.
3. A list and description of all raw materials used in the manufacture of the proposed drug product.
4. Release specifications on all raw materials and/or a Certificate of Analysis (COA) from the manufacturer or producer of the raw materials.
5. Release specifications, including acceptance and rejection criteria for the quantity of API in the proposed drug product, for the final proposed drug product formulation. For solid dosage forms, content uniformity and rate of dissolution or release may also be required.

6. Description of all AC methods and tests employed in the testing and release of raw materials.
7. Details on container and closure systems for storage of the proposed drug product.
8. Proposed drug product stability results, including detailed descriptions of stability-indicating AC methods and procedures.

The description of the manufacturing process for the proposed drug product will most likely not be the same process that will be used for the drug product to be marketed. Sponsors need to include sufficient details on the proposed drug product manufacturing process to demonstrate that the process is under control and will produce similar material during the manufacture of separate batches.

If the first-in-human clinical trial on a drug candidate is to be a placebo-controlled trial, sponsors are requested to provide information on the placebo to match the proposed drug product. This placebo information is to include the following:

1. Name and address of facility where the placebo is manufactured.
2. Detailed description of the manufacturing and packaging procedures for the placebo.
3. List of all raw materials used in manufacturing the placebo.
4. Release specifications on all raw materials and/or a Certificate of Analysis (COA) from the manufacturer or producer of the raw materials used in the placebo.
5. Release specifications for the final placebo formulation, including method to demonstrate the absence of the API. Stability results are not routinely required for the placebo.
6. Description of all AC methods and tests employed in the testing and release of raw materials.
7. Details on container and closure systems.

Copies of the labeling for the proposed drug product and placebo formulations to be used in the clinical trial studies are to be included. The type of clinical trial study to be conducted will generally dictate the information to be included on the label for the clinical trial supplies. In all cases, the statement "Caution: New Drug Limited by US (Federal) Law to Investigation Use Only" has to be placed on all labels for an investigational new drug.

An environmental analysis statement is required to be included in an IND (21 CFR § 312.23 (a) (7) (e)). Most sponsors include a claim for categorical exclusion in an IND since the manufacturing processes for the API and proposed drug product have not yet been fully defined or characterized.

## 7 PRE-IND CLINICAL ISSUES

The clinical development of a drug candidate cannot be initiated until an IND has been submitted to the FDA. This IND contains three major aspects:

1. The results from preclinical animal pharmacology, pharmacokinetic, and toxicology studies conducted on the drug candidate that show the material has an acceptable margin of safety for dosing to humans.
2. The results from CMC efforts to demonstrate that the manufacturing processes for the API and proposed drug product can be controlled and produce material of a sufficient quality for administration to humans.
3. The proposed clinical development program for the drug candidate, including a complete protocol for the first-in-human clinical trial, an Investigator's Brochure to provide summary information on the drug candidate to clinical investigators, and a summary of the proposed clinical development program.

The sponsor, usually the clinical department or group, will need to generate the clinical trial protocol to be included in the IND. ICH E6 [20] provides sponsors with recommendations for conducting clinical trials on a drug candidate in the following areas:

- Institutional Review Board (IRB) and Independent Ethics Committee (IEC)
- Clinical trial investigator
- Sponsor
- Clinical trial protocol and protocol amendment(s)
- Investigator's Brochure (IB)
- Essential documents for the conduct of a clinical trial

In general, the clinical trial protocol is to contain sections that provide background information, objectives and purpose, design, subject selection and withdrawal, subject treatment, assessments of efficacy and safety, statistics, access to source data, quality assurance (QA) and quality control (QC) procedures, ethics, data handling and record keeping, financing and insurance, and publication policy.

The IB usually contains an introduction, general considerations, confidentiality statement, summary on the disease indication, physical and chemical properties of the API and formulation, summaries of completed preclinical pharmacology, pharmacokinetic, and toxicology studies, effects in humans (if any), and guidance for the investigator. For a first-in-human clinical trial, information on effects in humans is not usually available unless the drug candidate has been administered to humans in foreign clinical trials or in investigator-IND clinical studies.

The summary of the proposed clinical development program should contain sufficient detail to explain the clinical testing strategy to determine the safety, pharmacokinetics, and efficacy of the drug candidate in patients with the disease or disorder to be studied and to establish the dosage regimen for treating the disease or disorder without causing unacceptable toxicity.

## REFERENCES

1. ICH S7a Guideline: *Safety Pharmacology Studies for Human Pharmaceuticals*, July 2001.

2. ICH S7b Guideline: *Safety Pharmacology Studies for Assessing the Potential for Delayed Ventricular Repolarization (QT Interval Prolongation) by Human Pharmaceuticals*, released for consultation at Step 2 of the ICH Process, February 2002.
3. ICH S3a Guideline: *Toxicokinetics: The Assessment of Systemic Exposure in Toxicity Studies*, March 1995.
4. *FDA Guidance to Industry: Bioanalytical Method Validation*, May 2001.
5. ICH S3b Guideline: *Pharmacokinetics: Guidance for Repeated Dose Tissue Distribution Studies*, March 1995.
6. ICH S6 Guideline: *Preclinical Safety Evaluation of Biotechnology-Derived Pharmaceuticals*, July 1997.
7. *FDA Guidance for Industry: Single Dose Acute Toxicity Testing for Pharmaceuticals*, August 1996.
8. ICH M3 Guideline: *Nonclinical Safety Studies for the Conduct of Human Clinical Trials for Pharmaceuticals*, July 1997.
9. *FDA Guidance to Industry (draft): Nonclinical Safety Evaluations for Drug Combinations*, January 2005.
10. ICH S2a Guideline: *Guidance on Specific Aspects of Regulatory Genotoxicity Tests for Pharmaceuticals*, July 1995.
11. ICH S2b Guideline: *A Standard Battery for Genotoxicity Testing of Pharmaceuticals*, July 1997.
12. *FDA Guidance to Industry: Recommended Approaches to Integration of Genetic Toxicology Study Results*, November 2004.
13. ICH S1a Guideline: *The Need for Long-Term Rodent Carcinogenicity Studies of Pharmaceuticals*, March 1996.
14. ICH S1b Guideline: *Testing for Carcinogenicity of Pharmaceuticals*, July 1997.
15. ICH S1c Guideline: *Dose Selection for Carcinogenicity Studies of Pharmaceuticals*, March 1995.
16. ICH S1c(R) Guideline: *Addendum to Dose Selection for Carcinogenicity Studies of Pharmaceuticals: Addition of a Limit Dose and Related Notes*, July 1997.
17. ICH S5a Guideline: *Detection of Toxicity to Reproduction for Medicinal Products*, September 1994.
18. ICH S5b Guideline: *Detection of Toxicity to Reproduction for Medicinal Products: Addendum on Toxicity to Male Fertility*, April 1996.
19. *FDA Guidance for Industry (draft): Safety Testing of Drug Metabolites*, June 2005.
20. ICH E6 Guideline: *Guideline for Good Clinical Practice*, May 1996.

# REGULATORY ISSUES IN PRECLINICAL SAFETY STUDIES (U.S. FDA)

Kenneth L. Hastings[1] and William J. Brock[2]

[1]*sanofi-aventis, Bethesda, Maryland*
[2]*Brock Scientific Consulting, LLC, Montgomery Village, Maryland*

**Contents**

1 Introduction
2 Have Sufficient Preclinical (Nonclinical) Toxicology Data Been Submitted?
3 Are the Doses Proposed for Clinical Trials Safe Based on Submitted Data?
4 Is the Length of Exposure Proposed for Clinical Trails Safe?
5 Additional Considerations
6 Conclusion
References

## 1 INTRODUCTION

In drug development, it is important to know and understand the legislation and other "high level" documents (e.g., case law) establishing the regulatory authority of the FDA as well as the specific enabling documents that have been published (e.g., 21 CFR, ICH Guidances, FDA Guidances) and the supporting guidelines [1–10]. However, it is of practical importance to understand the scientific approach that ultimately drives decision making in the drug review divisions that constitute the real world of pharmaceutical development. The pivotal individual in applying pharmacology/toxicology principles and practice to drug development is the FDA reviewer. This individual has the task of interpreting data submitted by the sponsor and making the decisions critical to clinical drug development. Sometimes in trying

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

to understand the drug review process, it is easy to lose sight of the "big picture" and become lost in the minutia. A very useful approach to understanding the review and decision-making process is knowing the fundamental questions needed to be answered by the somewhat bewildering list of published documents. The following list of questions will form the framework for discussing regulatory issues that, when dealt with incorrectly, can derail a drug development program:

- Have sufficient preclinical (nonclinical) toxicology data been submitted?
- Are the doses proposed for clinical trials safe based on submitted data?
- Is the length of exposure proposed for clinical trials safe?

If sponsors keep these questions in mind when designing, conducting, and interpreting preclinical drug safety studies, they will be thinking along the same lines as the pharmacology/toxicology reviewer. Each of these questions is discussed in turn.

## 2 HAVE SUFFICIENT PRECLINICAL (NONCLINICAL) TOXICOLOGY DATA BEEN SUBMITTED?

When attempting to answer this question, two documents are absolutely pivotal: ICH M3 and ICH S6. These documents provide key information on what studies are needed to enable clinical trials. ICH M3 also contains a chart that lists the studies needed to be completed prior to conduct of specific clinical studies (Table 1).

There are several issues that should be considered when interpreting what appears to be a rather simple paradigm. The first is that the term "repeat-dose toxicity studies" is generally interpreted to mean daily oral doses. This is a relatively straightforward matter when considering a new molecular entity (NME) that will be administered in clinical trials by the oral route. However, when a drug will be administered by another route, interpretation becomes more complicated. There is no simple paradigm to which reference can be made, but rather a number of points to consider. The first is whether the drug is an NME or not. If the drug has been previously developed for use by oral administration, it may be sufficient to conduct what are generally referred to as "bridging studies." For example, if a drug was originally developed for oral administration and sufficient oral toxicology data are

**TABLE 1 Duration of Toxicity Studies and Clinical Trials**

| | Duration of Clinical Trials Toxicity Studies | |
|---|---|---|
| Duration of Clinical Trials | Rodents | Nonrodents |
| Single dose | 2 weeks | 2 weeks |
| Up to 2 weeks | 2 weeks | 2 weeks |
| Up to 1 month | 1 month | 1 month |
| Up to 3 months | 3 months | 3 months |
| Up to 6 months | 6 months | 6 months |
| >6 months | 6 months | Chronic: 9 months and longer |

available, it may be possible to support clinical trials with repeat-dose toxicology studies by the new route of exposure (e.g., topical, inhalation) in one species (usually a nonrodent). Route-specific studies may need to be conducted (such as local irritation), and toxicokinetic studies should be incorporated into the nonclinical development program to establish comparative local and systemic exposure. However, the important point is that a complete replication of toxicology studies conducted by the original route of exposure may not be necessary.

An interesting issue that often arises in the nonclinical toxicology evaluation of drugs is the role of acute dose studies. As a general rule, the package of enabling studies should contain acute dose toxicology studies in both rodents and nonrodents. Several points should be considered. The first is the issue of what constitutes an acute dose study. Although occasionally a point of debate, it is generally accepted that an acute dose study is one in which the test article (drug) is administered as a single dose, or as more than one dose, in a single 24 hour period, followed by appropriate clinical and anatomical evaluation. The studies should be conducted over a dose range sufficient to demonstrate adverse effects and should be GLP compliant (21 CFR part 58). The second is that acute studies, as a general rule, are not usually sufficient to support clinical trials. They are essentially "discovery" or *hazard identification* studies intended to determine the toxic potential of the test article. In fact, several studies may need to be conducted (usually in rodents) to determine the doses necessary to produce toxicity. There are examples of drugs for which acute toxicity studies may be adequate to support clinical trials (basically, agents that will be administered as a single dose in very circumscribed clinical settings, such as radioimaging agents or inhalation anesthetics), and there are specific situations recognized in regulatory guidances discussed later, but in general these are not sufficient to initiate clinical trials. Finally, under most circumstances it is not necessary (in fact, it is discouraged) to determine lethal doses. Although acute dose toxicity studies may be useful in evaluating and managing overdose situations in clinical trials, and may eventually be used to write the overdose section of a product label, they are of limited practical use in the design and conduct of clinical trials. This leads to a consideration in designing toxicology studies to enable clinical trials: if appropriate repeat-dose toxicology studies have been conducted, GLP-compliant acute dose studies may not be needed, or even appropriate, especially in nonhuman primates.

As mentioned earlier, there are specific guidance documents that allow for the use of acute dose toxicity studies to enable clinical trials. For the development of imaging agents, for example, it is acceptable to conduct acute dose toxicity studies with significant caveats. First, the studies will need to be GLP compliant and conducted in both rodents and nonrodents, and animals should be monitored for the core battery of safety pharmacology parameters (cardiovascular, central nervous system, and pulmonary function) unless separate single-dose studies are conducted as described in ICH S7A. Second, complete clinical observations, clinical pathology (clinical chemistry and hematology), as well as necropsy and complete histological examinations must be conducted on day 2 (24 hours after test article administration). Third, a recovery group must be included for every dose group with complete clinical and necropsy/histology determinations.

The second example is the *screening IND*. In most respects, screening IND studies should be conducted in a manner identical to those needed to support clini-

cal trials with imaging agents. There are several additional considerations, however. The screening IND is designed to allow a sponsor to submit safety data on several related compounds. Although not explicitly defined, the term "related compounds" is assumed to mean both structurally and pharmacologically related. If the drugs are to be administered by the oral route in clinical trials, acute toxicity studies should be conducted by the intravenous route as well, with complete safety pharmacology, clinical pathology, and pathology determinations. Toxicokinetic determinations also should be included with particular emphasis on oral bioavailability. An ICH-compliant battery of genotoxicity studies should be conducted with each compound. The specific purpose of the screening IND paradigm is to enable single-dose clinical studies in healthy volunteers in order to select the most promising candidate compound. For further discussion of the different types of INDs and the regulatory process, the reader is referred to Mathieu [11].

In 2006, the FDA published a comprehensive guidance entitled *Exploratory IND Studies* [8]. Although often referred to as a single approach, in fact the guidance describes several alternative approaches to first-in-human (FIH) clinical trials, and the types of nonclinical studies needed to support these trials. The general approach is discussed later, but one concept in particular utilizes acute toxicity studies—the *microdose* clinical trial. Essentially, the design of acute dose toxicity studies builds on the screening IND approach described earlier with the notable exception that the safety of a proposed single-dose clinical trial can be determined using one mammalian species. Selection of the species should be justified using comparative *in vitro* pharmacokinetic and pharmacodynamic data (essentially, using animal and human tissues/cells), and the dose selected for use in clinical trials should be no more than 1/100 of the no-observed-adverse-effect level (NOAEL) in the test species, based on relative body surface area and not a mg/kg dose. Safety pharmacology and genotoxicity studies are not needed. As a practical matter, the clinical trials are likely to be conducted using doses in the low microgram level (nanomole level for protein drugs), and the studies are designed primarily to determine (relative) oral bioavailability. Another possible clinical study endpoint is specific receptor binding (discovery pharmacodynamics).

In most instances, a battery of genotoxicity studies should be conducted and submitted for review prior to FIH clinical trials. These are hazard identification tests and should include bacterial mutagenicity (Ames test) and at least one *in vitro* mammalian genotoxicity study (the mouse lymphoma assay and/or a chromosomal aberration assay). Prior to Phase II clinical trials, a complete battery of ICH S2B genotoxicity tests should be completed, which as a practical issue means an *in vivo* rodent micronucleus assay in addition to previously completed studies.

Other studies that are needed according to ICH M3 include determination of local tolerance (if the drug is to be administered by a nonoral route), a battery of ICH S7A safety pharmacology studies, and, prior to Phase III clinical trials, the ICH equivalent of Segment II reproductive toxicology studies. Other studies may also be needed, such as an intravenous toxicology study to enable clinical studies to determine absolute oral bioavailability. Studies to determine potential drug effects on cardiac function (e.g., hERG assay, *in vivo* cardiac function studies), as described in ICH S7B, should be submitted although these are not required. ICH S8 immunotoxicity studies may be needed based on a weight-of-evidence analysis, primarily using results of repeat-dose toxicology studies. Some specific studies, such as assays

to determine the ability of topical drugs to cause allergic contact dermatitis, may be needed.

Chronic nonrodent toxicology studies are likely to be needed if the drug is to be used for more than 3–6 months. As discussed later, there are considerations in deciding the need for chronic toxicology studies, but these apply to a minority of drugs. There are "regional differences" in the definition of chronic use. In the United States, "chronic toxicity studies" are defined as 3 months of repeated exposure to the drug, or intermittent exposure equivalent to 3 months. ICH specifies that a repeat-dose toxicology study in rodents of 6 months duration is adequate. For nonrodents, ICH defines the duration of a chronic repeat-dose toxicology study to be 9 months. However, for an NME that is the first molecule in a class of compounds, it is likely that a 12 month repeat-dose study will be needed. Another situation in which a 12 month study may be expected is "accelerated approval," where a fairly small clinical database is available for evaluation and, especially, where a surrogate marker is used as the efficacy endpoint. On the other hand, a 6 month repeat-dose toxicology study might be deemed sufficient if the drug belongs to a well-known class of drugs. Toxicities are sometimes observed that clearly limit the ability to conduct repeat-dose toxicity studies of greater than 6 months duration at clinically relevant doses. Lifetime rodent carcinogenicity bioassays (mouse and rat) will almost always be needed for drugs that will be used on a chronic basis. These studies are usually conducted late in the development of a drug and occasionally will be performed as part of postmarketing commitments. The FDA does not provide specific guidance on the conduct of most types of nonclinical toxicology studies (with the notable exception of food additives), but in the case of carcinogenicity bioassays, concurrence on the study design should be requested by the sponsor via the protocol assessment review process.

Finally, biologic therapeutics (basically, protein drugs) represent a unique class of therapeutics and nonclinical studies needed to enable clinical trials are considered in ICH S6. It is important to remember that ICH S6 and ICH M3 are complementary documents, and both apply to protein therapeutics. It is the exceptions to ICH M3 that should be noted when planning the development of protein drugs. For example, under most circumstances, rodents are not likely to be appropriate for evaluating the safety of protein drugs. In fact, it is the entire concept of "appropriate animal model" that leads to much of the misunderstanding in the design and interpretation of nonclinical toxicology studies with biologic products. Many of these drugs are immunogenic in animals other than nonhuman primates, and production of neutralizing antibodies in repeat-dose toxicology studies can render findings irrelevant to human safety. If a protein drug is pharmacologically inactive in a species, it is highly likely that safety information will also be uninformative in this model. Unlike small molecular weight drugs, most toxicity observed with protein drugs are likely to be essentially exaggerated pharmacodynamics. There are certainly exceptions, but even under most circumstances these apparently "off-target" adverse effects will be observed only in species fairly closely related to humans.

There are many ways of dealing with the issue of species specificity such as the use of homologous proteins that are pharmacologically active in a standard toxicology species (usually rodents) or the genetic manipulation of mice or rats such that they express the human receptor of interest (drug target) or lack the molecule

(especially useful for evaluating protein antagonists or monoclonal antibodies to endogenous proteins). There are certain studies that are not needed to evaluate the safety of protein therapeutics. For example, genotoxicity studies and hERG assays are usually uninformative. In some cases, there may simply be no relevant animal model to evaluate the safety of a protein therapeutic, and consideration should be given to using *in vitro* data using human cells/tissues. This approach is discouraged by the FDA, but sometimes there will be no justifiable alternative. For example, consider a monoclonal antibody (mAb) directed at a human pathogenic virus. A standard method useful in evaluating potential safety concerns with the mAb is tissue cross-reactivity. Essentially, the mAb is tested with a bank of human (and animal) tissues to detect specific binding to expressed epitopes. If no binding is observed, it is highly likely that nonclincial toxicology studies will not be useful and that the safety of clinical studies cannot be evaluated using standard techniques. In the case of monoclonal antibodies to human epitopes, where no pharmacological effects are observed in nonhuman primates or other standard toxicology species, there are *in vitro* methods that utilize human cells and may be useful. For many (if not most) protein therapeutics, case-by-case study design should be the expected approach.

## 3 ARE THE DOSES PROPOSED FOR CLINICAL TRIALS SAFE BASED ON SUBMITTED DATA?

Extrapolation of doses from animals to humans is a critical issue in evaluating non-clinical toxicology studies. The key document to consult when converting animal doses to human equivalent doses is *FDA Guidance for Industry: Estimating the Maximum Safe Starting Dose in Initial Clinical Trials for Therapeutics in Adult Healthy Volunteers* [6]. The general method presented in this guidance is based on comparative body surface areas and utilizes the standard conversion formula:

$$\text{Human Equivalent Dose} = \text{Animal Dose (mg/kg)} \times (\text{Animal Weight/Human Weight})^{0.33}$$

This method is an inherently conservative approach and is intended to estimate systemic exposure to low molecular weight drugs. For example, if a dose of 30 mg/kg, when administered by the oral route to cynomolgus monkeys, produces an AUC of 50 μg·h/mL, using this formula, the human equivalent dose (HED) would be 10 mg/kg, and if this dose is administered to an adult human volunteer, the measured AUC should also be ~50 μg·h/mL. There are of course many assumptions made in this method and include similar oral bioavailability and metabolism, and the method assumes that the drug has never been administered to humans, so no comparative pharmacokinetic data exist to test the calculation. In addition, the method is meant to apply to FIH doses in healthy volunteers. Therefore, an additional safety factor should be applied to the calculated HED. The default safety factor advocated in the guidance is 10, so the dose calculated in the example given above would be 1 mg/kg. The guidance allows for several considerations in calculating HED, such as steepness of the dose–response curve and nature of adverse effects observed (reversible or not, monitorable or not, etc.), and provides the sponsor the opportunity to present

data supporting an alternative method for estimating relative doses (such as body weight comparison). Although the default dose adjustment is referred to as a safety factor, it is in fact an uncertainty factor. Experience has demonstrated the method to be reliable for most classes of drugs. The exceptions include drugs delivered by routes of administration that do not result in appreciable systemic exposure, drugs that are extensively metabolized (where the object of the calculation is estimating comparative exposure to the parent drug), and protein drugs with a molecular weight ≥100,000 where distribution is primarily confined to the vascular compartment.

An important consideration in calculating a safe starting dose for clinical trials in healthy volunteers based on animal data is that a NOAEL has been determined. The NOAEL is the animal dose that will be used to calculate the FIH dose. In most cases, the FDA will take a conservative approach in identifying a NOAEL. Since most Phase I studies will be conducted in healthy volunteers, essentially no adverse effects should be anticipated when administering FIH doses. An issue frequently debated with sponsors is the simple question of whether a finding in an enabling repeat-dose toxicology study is in fact "adverse." For example, should an increase in blood pressure of 10% compared to controls be considered adverse? Chances are the FDA would want to see a dose at which *no* increase in blood pressure was observed, but for certain drug classes this might be an unrealistic expectation. As discussed later, the pharmacologically active dose (PAD) is a consideration in calculating a safe starting dose. Depending on the intended use of the drug, blood pressure effects may be an important consideration, and if the HED at which the blood pressure effect is observed is higher than the anticipated PAD based on animal pharmacology studies, the PAD should be used to calculate the FIH dose. If the PAD is *greater* than the dose at which elevations in blood pressure are observed, it is likely that the FDA will consider the NOAEL to be the lower dose, and although pharmacokinetic data can be obtained in healthy volunteers, it is likely that pharmacologic activity studies will need to be conducted in patients rather than healthy volunteers.

The concept of NOAEL is important in one clinical trial design allowed under the *Exploratory IND Studies* guidance. The essential approach is that the sponsor would conduct a standard 2-week repeat-dose toxicity study in animals (usually rodents), and a NOAEL would be determined. The sponsor would include a determination of key pharmacokinetic (PK) parameters in this study, most importantly the AUC, $C_{max}$, and $t_{1/2}$. A 2-week repeat-dose oral toxicity study would then be conducted in a second animal species (usually dogs), using only one dose level, calculated to be the NOAEL equivalent based on the first species, and the same PK parameters determined. If the NOAEL in the first species is equivalent to the second species based on relative body surface area, and if PK parameters at the NOAEL are equivalent, then the HED can be calculated and used to establish the FIH dose. If one species appears to be more sensitive based on observed effects (essentially, this would mean that adverse effects were observed in the second species at the NOAEL equivalent in the first species), the sponsor would use the second species to determine the NOAEL in a 2-week repeat-dose study. In either case, the FIH dose would be no greater than 1/50 of the NOAEL in the more sensitive species based on the relative body surface area calculation. The *maximum* clinical dose would be the lowest of the following:

- One-fourth of the NOAEL determined in the first species (usually rodents)
- One-half of the AUC at the NOAEL in the first species, or the AUC in the second species (usually dogs) at the dose equivalent to the NOAEL in the first species (whichever is lower)
- The PAD in humans, or
- The dose at which an adverse clinical effect is observed

Although this seems a complicated scheme, it has the advantage of requiring less test article than would be needed for standard toxicity studies needed to enable FIH clinical trials, and the sponsor could conduct clinical trials of up to 7 days in duration in healthy volunteers to include the determination of PK parameters at steady-state blood levels. Other issues would also need to be addressed, such as genotoxicity and safety pharmacology, but the study design could support more rapid initiation of repeat-dose Phase I clinical trials.

One issue missing from this discussion has been studies to support traditional ascending-dose PK and maximum tolerated dose (MTD) Phase I clinical trials. These still have an important role in drug development. The body surface area dose calculation is relevant to dose setting in these studies. However, some classes of drugs continue to be evaluated in symptomatic patients rather than healthy volunteers, and determination of safe starting dose may follow a different paradigm. For example, when developing cytotoxic chemotherapeutic agents for the treatment of cancer, Phase I studies are typically conducted in patients who are suffering from end-stage disease. In this situation, it has been judged acceptable to use a much more aggressive strategy in determining acceptable clinical doses. A traditional approach is to determine the lethal dose in 10% of animals ($LD_{10}$)—usually mice—and calculate the starting dose based on relative body surface area. Although rodents have been used to determine FIH doses, there is evidence that dogs may be more sensitive. However, there is an ongoing debate over whether increased sensitivity results in more effective dosing strategies. One issue to be considered in assessing the usefulness of a more sensitive species is that nonclinical toxicity studies in these animals may be limited with respect to discovering potential adverse effects. That is, it is at least worthy of consideration that use of a less sensitive species could allow for testing higher doses for longer durations, maximizing the possibility that important, clinically relevant toxicities may be revealed.

## 4 IS THE LENGTH OF EXPOSURE PROPOSED FOR CLINICAL TRIALS SAFE?

This is essentially a cumulative toxicity question. As presented earlier, ICH M3 explicitly states the length of nonclinical toxicity studies needed to support the safety of clinical trials. When designing a drug development program, it is important to consider the timing of nonclinical studies with respect to the duration of clinical trials. For example, the entire amount of time needed to conduct and evaluate findings can be underestimated, which can result in "delayed" clinical development. As a general rule, the importance of this issue is often overestimated. That is, it is much more important to design and conduct enabling toxicology studies correctly—even

if this seems to take more time than might be judged necessary—than to rush studies and commit errors that compromise the reliability of study results. Recruitment of subjects into clinical trials often takes much more time than anticipated; the clinical investigator usually underestimates this time and the numbers of subjects that can be recruited. Carefully conducted nonclinical studies can be extremely valuable, especially with respect to discovering toxicities that require repeated exposure—and often repeated insult—to demonstrate. Although in-life determinations can be useful, clinical chemistry and hematology signs are usually not "leading indicators." When elevations in creatinine are observed, for example, it is likely that extensive kidney pathology has already been produced. Thus, reliance on in-life observations in nonclinical toxicology studies can be a risky clinical trial strategy.

One issue that has arisen repeatedly is the impact of positive findings in genotoxicity studies on the safety of clinical trials. When considering the significance of genotoxicity in length of exposure, a weight-of-evidence approach should be taken [9]. Experience has indicated that often positive genotoxicity findings can be related to product impurities, and it is important to make this determination as soon as possible. If positive genotoxicity findings are due to drug contaminants, this may be a fairly easy situation to address. Another typical situation is a positive finding in an *in vitro* chromosomal aberration assay. This argues for inclusion of *in vivo* micronucleus assays as early as possible in nonclinical development. Many times the situation that needs to be addressed is that the positive findings *in vitro* were produced at high drug concentrations, bacterial mutagenicity assays were negative, and the mouse micronucleus assay is needed to clarify the nature of the hazard.

Carcinogenicity studies are rarely needed to enable clinical trials, but there are exceptions [4]. Clinical development of peroxisome proliferator-activated receptors (PPARs) represents a significant departure from the norm. Current ad hoc policy in the FDA is to allow clinical trials with PPAR agonists for up to 6 months duration provided there are no other toxicities, especially genotoxicity or cardiotoxicity findings, in nonclinical studies. However, in order to proceed with studies of longer duration, it may be necessary to complete carcinogenicity bioassays, thus significantly affecting clinical trials. Although this has not been adopted as formal policy, it remains the practice within the FDA.

## 5 ADDITIONAL CONSIDERATIONS

In addition to the three issues just discussed, there are other questions that the pharmacology/toxicology reviewer will likely want to answer. These may include:

- Are there patient or condition of use restrictions?
- Do studies indicate the need for special monitoring or need for available antidotes?
- Are the demonstrated pharmacological effects relevant to the intended indication?
- Are adverse effects expected at effective doses?
- What is known about related drugs?
- What are the types and sites of toxic effects?

- Has the sponsor established a no observed adverse effect level (NOAEL)?
- What is the apparent therapeutic index?
- Are there irreversible or unmonitorable adverse effects?
- Are there differences in metabolism of the drug, especially between animals used in pivotal toxicology studies and humans?

These "supplemental" questions are no less important, but either apply to a subset of nonclinical studies or have been answered as part of the larger issues dealt with earlier. However, each deserves some comment.

*Are There Patient or Condition of Use Restrictions?* This question usually can be taken as referring to special populations. For example, if a drug is to be studied in clinical trials to prevent transmission of infections from mother to infant (perinatal transmission, e.g., HIV), the sponsor should conduct Segment I and III reproductive toxicology studies earlier than is usually expected [3]. Although these studies are conducted prior to submission of the NDA, these studies would need to be conducted prior to Phase II clinical trials (assuming pregnancy would not be an issue in Phase I studies). Although a relatively simple timing issue for most drugs, this could be a complicated issue with biologics. In fact, the entire issue of reproductive toxicology studies becomes complex when evaluating protein drugs. This is the single case where these types of studies may need to be conducted in nonhuman primates (NHPs). Although great effort is made to avoid these studies (e.g., development of transgenic rodent models expressing the target of interest for use in reproductive toxicology studies), often cynomolgus monkeys or other NHPs may be the only practical choice. A somewhat related topic is studies to enable pediatric drug trials. The FDA has published guidance on the types of juvenile animal studies useful in supporting the safety of clinical trials in infants, children, and adolescents, but the most important point is that these may not always be needed, and consultation with the FDA is encouraged. Length of exposure, age of the clinical trial subjects, and what is already known concerning the adverse effects of the drug are all issues that should be considered in deciding the need for juvenile animal studies [10].

Another issue that should be dealt with under this question is the issue of phototoxicity and photoallergy [5]. The usefulness of nonclinical phototoxicity studies has been debated, and the FDA has published a guidance on this topic, but it is still somewhat unclear under what conditions nonclinical phototoxicity studies may be needed. Although generally considered in the development of topical drug products, this may also be an issue with drugs administered by other routes where systemic exposure results in significant skin deposition. In fact, some of the most important phototoxic drugs (e.g., fluoroquinolones) are given orally. The guidance relies on two parameters to determine the need for nonclinical phototoxicity studies: (1) significant molecular absorption in the UV-VIS spectrum and (2) concentration of the drug and/or metabolite(s) in the skin. Whether these are appropriate criteria can be questioned, however. Recently, the FDA has not requested photoallergy nonclinical studies since interpretation of the results continues to be debated.

*Do Studies Indicate the Need for Special Monitoring or Need for Available Antidotes?* Although often overlooked, risk management is an important issue to be dealt with in toxicology studies. Unfortunately, nonclinical toxicology studies

have not been utilized as well as possible in clinical trial design. There are examples of how these studies may be used. For example, if an association has been established in nonclinical studies between drug AUC and liver toxicity, this should be considered in the design and conduct of clinical trials. If a drug has been shown to cause anaphylactoid reactions, conditions associated with this adverse effect (such as rate of infusion for drugs given intravenously) should be used in the design of clinical trials, and appropriate antidotes (e.g., epinephrine) should be available during (and after) drug administration. There are a number of examples that could be cited, but the important point is to take into consideration toxicity findings when considering the issue of risk management. The effort to discover and apply biomarkers of toxicity is a major driver in the FDA's Critical Path Initiative.

*Are the Demonstrated Pharmacological Effects Relevant to the Intended Indication?* The answer to this question is often put in the "nice to know" category, but the importance of the issue should not be underestimated. An important consideration is that the review team at the FDA develops a perception of how effective a drug is likely to be and will sometimes act accordingly. Consider, for example, a drug intended to prevent organ transplant rejection. Especially in renal transplantation, there are a number of effective drugs currently available. None are without adverse effect problems, but in general these have proved to be "miracle" drugs when used correctly. Thus, in developing a new drug for this indication, it is very important that proof-of-concept pharmacology studies provide compelling efficacy data. One consideration is that usually a candidate drug for an organ transplant indication will likely be added to an ongoing cyclosporine-based immunosuppressive regimen. Nonclinical pharmacology studies should take this into consideration, and studies should be designed to demonstrate whether the candidate drug in any way interferes with the efficacy of cyclosporine. The important consideration here is that sponsors ignore good proof-of-concept studies at their peril.

*Are Adverse Effects Expected at Effective Doses?* This is another risk management question and should be considered as early as possible in the development of a drug. There are many effective drugs on the market that produce headache, nausea, diarrhea, and so on, in a significant portion of the patient population. Many of these adverse effects will only be discovered in clinical trials, but signs of toxicity should be carefully considered in the design and conduct of these trials. Among the signs that can be misinterpreted are neurological effects in animals. Many animals can demonstrate signs of neurotoxicity, such as aberrant behavior in a functional observation battery, that at first seem not predictive of human effects. It is important to remember that animal behavior in response to adverse neurological effects may not, in fact often does not, have a direct correlation in humans. It is entirely conceivable that convulsions in animals may not be associated with similar effects in clinical trials but may signal that the drug does have the potential to produce seemingly unrelated effects in humans, such as depression or anxiety. Once again, sponsors ignore such findings at their peril, especially where obvious adverse effects are observed in animals at target doses for development.

*What Is Known About Related Drugs?* As was pointed out earlier, the FDA has experience with a vast array of drugs. Many have failed in development due to

adverse effects, but findings were never made public in publications. As part of the drug development program, it is incumbent on the sponsor to determine as thoroughly as possible what is known in publicly available sources about possible drug class effects. It is also important that the sponsor address these potential class effects, often by design and conduct of specialized nonclinical studies or incorporation of additional observations in standard toxicology studies. Also, it is important to "listen to the reviewer." All too often suggestions are made to a sponsor that particular attention be given to a specific issue and the advice is ignored. A wise sponsor listens to what is essentially free advice and considers it carefully.

*What Are the Types and Sites of Toxic Effects?* This is essentially a review issue, but as with so many questions in this list there are potentially important points to consider. One is related to the previous question. Simply put, if a drug belongs to a class with known toxicities, and these are not seen in submitted toxicology studies, several questions could be raised by the review team and include:

- Have they been conducted correctly?
- Has a true MTD been achieved?
- Were appropriate observations made and endpoints included?
- Do TK data support that sufficient systemic exposure was achieved?

As you can see, what is a seemingly straightforward issue that should be answered by the results of standard toxicology studies could become a very complicated problem. Another point to consider is whether the types of toxicities are "acceptable" given the intended indication. This can be a very complicated issue and should be addressed. For example, consider a drug that is intended to be administered by the topical route, and dermal toxicity studies have demonstrated relatively minor adverse effects at reasonable multiples of expected clinical doses. However, studies that maximize systemic exposure—that is, the studies were conducted by a parenteral route of administration—demonstrate significant toxicities associated with effective topical doses. These effects should not be ignored. It is frankly impossible to consider every scenario of use and discovery of a hazard should drive appropriate *risk assessment* studies. In this example, it may be adequate to demonstrate that, under conditions of use, significant systemic exposure is so unlikely as to render the toxicity findings irrelevant. A sponsor should never take such issues lightly—it is almost certain that the FDA will not.

*Has the Sponsor Established a No Observed Adverse Effect Level (NOAEL)?* Although discussed extensively earlier, it is included here for emphasis. The sponsor *must* establish a NOAEL in most cases in order to enable FIH clinical trials. There are the usual exceptions (such as chemotherapeutic agents), but these are relatively obvious. A classic mistake made by sponsors is to conduct a nonclinical toxicology study intended to support the safety of a clinical trial in which no adverse effects are observed at one dose and serious toxicities, including death on study, are observed at the next highest dose. Based on the results of the study, the sponsor will claim to have established a NOAEL. Although technically correct, the sponsor has ignored the obvious implications of the study results. That is, either the range of doses was

too great to observe something equivalent to a lowest-observed-adverse-effect level (LOAEL) so that clinically appropriate toxicities were observed (i.e., adverse effects that may be monitored and when observed can be used to determine a reasonable stopping dose in a dose–escalation trial), or the dose–response curve is so steep with respect to serious toxicities that development of the drug is likely to be problematic. Issues such as this can seriously affect a drug development program and should not be taken lightly.

*What Is the Apparent Therapeutic Index?* This question is related to the previous one. All too frequently, sponsors will fail to consider results of nonclinical pharmacology studies in the context of potential safety implications of findings. A common mistake is to conduct a proof-of-concept pharmacology study in an animal model, demonstrate promising effects, and ignore serious toxicities. Enabling nonclinical toxicology studies are conducted, a NOAEL is established, Phase I studies are conducted, and what is discovered is that although a safe dose can be administered to humans, it is nowhere near the projected efficacious dose based on pharmacology studies. This is actually a relatively easy issue to deal with as long as the sponsor conducts good studies and is willing to "live with" the results.

*Are There Irreversible or Unmonitorable Adverse Effects?* This issue can be a "drug killer." There are a number of toxicities for which there are no adequate biomarkers of effect useful in clinical trials (especially prodromal markers). When observed, often the most important determination to make is the dose multiple at which the toxicity is observed. Although sometimes "irreversible or unmonitorable adverse effects" may be toxicologically insignificant, this is not the usual case. Thus, it is very important to establish the apparent therapeutic index. Other parameters that are important to establish include blood levels at which these adverse effects are observed (AUC, $C_{max}$); it is important to determine whether doses or systemic exposures needed to produce the adverse effects change with length of treatment (especially if the doses needed to produce the adverse effect are lower with longer exposures), and whether a clinically useful dose can be found where irreversible/unmonitorable adverse effects are not produced, with an appropriate safety margin. Although this is essentially the same as establishing an apparent therapeutic index, in fact, the toxicity used to determine this might occur at doses lower than those associated with the irreversible/unmonitorable effect and could provide an additional margin of safety.

*Are There Differences in Metabolism of the Drug, Especially Between Animals Used in Pivotal Toxicology Studies and Humans?* This has emerged in recent years as a serious potential problem in drug development. Some authorities maintain that this is essentially an "artifact" of increased sensitivity of analytical methodology (tandem mass spectrometry, etc.). It is important that the sponsor determine as early as possible in the drug development program whether there is/are relatively unique drug metabolite(s) produced by humans that have not been assessed for safety in nonclinical toxicology studies. The FDA has published guidance on this topic [6]. Essentially, the guidance recommends that the sponsor determine, possibly using *in vitro* methods such as isolated human hepatocytes/liver slices/microsomes, whether unique (or disproportionately abundant) metabolites exist compared to species

used in toxicology studies, and be prepared to deal with the issue proactively. This could mean that the sponsor might need to conduct essentially "bridging toxicology" studies with synthesized metabolite. There are other considerations, however, such as relative amount of metabolite. The FDA guidance has set (rather arbitrarily) 10% of administered dose or systemic exposure (both relative to parent drug) as the "concern level" for a unique or "overabundant" metabolite. It is recommended that the sponsor consult the FDA when such situations arise during drug development.

## 6 CONCLUSION

A wise man once said that every new drug requires a new development program. This is very close to the truth. Drug development should be approached as what it, in fact, is: a scientific experiment. There are few if any "givens" in drug development, which is why it is a risky proposition. Seek advice whenever needed, and consider it carefully when offered.

## REFERENCES

1. FDA Guidance website: http://www.fda.gov/cder/guidance/index.htm.
2. ICH Guidance website: http://www.ich.org/cache/compo/276-254-1.html.
3. *Integration of Study Results to Assess Concerns about Human Reproductive and Developmental Toxicities*. Washington, DC: FDA; 2001.
4. *Carcinogenicity Study Protocol Submissions*. Washington, DC: FDA; 2002.
5. *Photosafety Testing*. Washington, DC: FDA; 2003.
6. *Guidance for Industry: Estimating the Maximum Safe Starting Dose in Initial Clinical Trials for Therapeutics in Adult Healthy Volunteers*. Washington, DC: FDA; 2005.
7. *Safety Testing of Drug Metabolites*. Washington, DC: FDA; 2005.
8. *Exploratory IND Studies*. Washington, DC: FDA; 2006.
9. *Recommended Approaches to Integration of Genetic Toxicology Study Results*. Washington, DC: FDA; 2006.
10. *Nonclinical Safety Evaluation of Pediatric Drug Products*. Washington, DC: FDA; 2006.
11. Mathieu M. *New Drug Development: A Regulatory Overview*, 7th ed. Waltham, MA: Parexel; 2005.

# SELECTION AND UTILIZATION OF CROs FOR SAFETY ASSESSMENT

Joanne R. Kopplin and Ward R. Richter
*Druquest International, Inc., Leeds, Alabama*

**Contents**

1 Introduction
2 Historical Development of the Preclinical CRO Post-WWII
3 Role of the CRO in Preclinical Development Today
4 Locating Candidate CROs
5 Telephone Interviews for Initial Screening
6 Site Visit: The Key to Evaluation
7 Evaluation of the Process: Not the Final Report
8 Special Considerations in Selecting a Specialized or University Laboratory
9 Selecting a Single Laboratory Versus Spreading the Work Among Multiple Laboratories
10 Initial Selection of a CRO for a New Program or Single Large Study
11 Tactics for Developing a Relationship with the CRO Staff and Enlisting Them as Members of the Development Team
12 Dealing with Unexpected Compound-Related Changes or Technical Problems
13 Monitoring the Outsourced Work
References

## 1 INTRODUCTION

Outsourcing preclinical development and safety testing to Contract Research Organizations (CROs) has become common in the biopharmaceutical, chemical, medical device, and food additive industries. It is important to understand the culture of the

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

CRO industry and how to locate, evaluate, and work with individual CROs when outsourcing work. In this chapter, we discuss these processes based on personal experience working with and working for preclinical CROs. Our combined experience is derived from outsourcing preclinical work as employees of a major pharmaceutical company, a major chemical company, and a young biotechnology company and as consultants for these industries. We have also held senior management positions in three different CROs during the course of our careers.

## 2 HISTORICAL DEVELOPMENT OF THE PRECLINICAL CRO POST-WWII

Contract Research Organizations (CROs) conducting preclinical testing provide services in pharmacology, pharmacokinetics, analytical chemistry, toxicology, pathology, and associated disciplines. Most were established and/or grew after Word War II as a result of growth in the pharmaceutical industry and because of increasing government regulations. The preclinical contract industry has been concentrated in the United States and the United Kingdom. Smaller laboratories are found in Canada, France, Italy, Japan, and other countries.

Early growth of the preclinical contract industry was fueled by rapid growth of the pharmaceutical industry when construction of new toxicology laboratories and animal facilities did not keep up with the number of new drug candidates selected for development. Surplus work was outsourced but sponsors outsourced only work that was deemed less critical and simpler to conduct because contract laboratories did not have the same level of scientific expertise, facilities, or equipment as the sponsoring companies. The scientific capabilities of the contract laboratories grew in quality and depth over the years and by the 1990s they were generally perceived as the equivalent of that found in the largest pharmaceutical firms. For many specialized technologies they often were superior to the sponsoring companies because of extensive experience in technologies that were used only occasionally by a single company.

Rapid growth of toxicology contract laboratories occurred during the 1960s and the early 1970s. Industrial Biotest (IBT) was an industry leader and was utilized by many major pharmaceutical and chemical companies in the United States, Europe, and Japan. IBT was favored because it turned out studies and reports on time and at low cost and was perceived to provide quality work. By 1975, it was learned that Industrial Biotest fabricated some of the data in its reports and the laboratory was eventually closed with several members of the management team convicted of criminal activity. The criminal behavior of this laboratory was one of the motivators for the establishment of the Good Laboratory Practice Act of 1978. Because so many large companies had relied on this laboratory and because some of the work was invalidated, there was a significant backlash against outsourcing to contract laboratories. Pharmaceutical and chemical companies rapidly built new toxicology laboratories and established policies of conducting all preclinical testing in-house. The amount of outsourcing for preclinical safety testing leveled off or contracted in the 1980s, as a result.

Another unrelated event occurred in 1976 that was to have a major negative financial impact on the preclinical contract laboratory industry when the Toxic

Substances Control Act (TOSCA) was passed, regulating an estimated 75,000 chemicals. Most people in the industry and government expected that TOSCA would mandate extensive toxicology testing of thousands if not tens of thousands of chemicals in industrial use. The contract laboratory industry viewed this as an opportunity and a major new source of revenue and expansion. Many of the laboratories, including the industry leaders, built new facilities, increasing their animal capacity by as much as 30–40%. The work predicted by passage of the Toxic Substances Control Act never materialized due to the manner in which the law was interpreted by regulatory agencies and because of intense lobbying pressure by the chemical industry. By the early 1980s there was a serious surplus of animal rooms in the contract industry and most facilities were not filled. The industry became very competitive and companies cut prices to cost or even below cost to cover overhead. This oversupply of animal and laboratory facilities kept prices and margins depressed until the late 1990s when growth of the pharmaceutical industry caught up with capacity. Prices of studies remained at relatively flat level for 15 years in spite of inflation.

Starting in the 1990s, a series of factors dramatically improved the financial outlook for preclinical toxicology laboratories. (1) The biotechnology industry, considered as a subset of the pharmaceutical industry, matured and grew to spend significant amounts of money on development. (2) The contract laboratories achieved recognition that they were fully as capable as in-house laboratories of sponsors and they no longer carried the stigma of the Industrial Biotest incident. (3) Combinatorial chemistry and new automated early screening technologies increased the number of potential drug candidates in development. (4) Financial administrators in the large pharmaceutical and chemical companies put more and more pressure on the corporations to outsource for economic reasons, because of the lower operating costs in contract laboratories. Purchasing departments became involved and they pressed for preferred provider agreements with one or more CROs.

Today the involvement of CROs is recognized and supported by the FDA. In a speech to the Association of CROs (ACRO) in October 2002, the FDA Commissioner, Dr. Mark McClellan, stated that "CROs hold great potential" to help the process of reducing development time for new drug applications and getting safe, critical products to market more quickly.

## 3 ROLE OF THE CRO IN PRECLINICAL DEVELOPMENT TODAY

The trends just described have filled the empty animal rooms in the major preclinical CROs in the last several years. Pharmaceutical companies have their own laboratories but contract excess or specialized work to CROs. We estimate the annual value of outsourced work (preclinical and clinical) to be in excess of $8 billion (U.S.). Of this $8 billion, an estimated $600–750 million was outsourced for preclinical testing in 2002. It is unclear whether analysts developing these estimates have included the potential outsourcing from the smaller biotechnology companies. Small biotechnology companies without their own internal capabilities exclusively use CROs to conduct their safety testing. Even the largest biotechnology companies have limited animal facilities, usually limited to efficacy testing and pre-GLP safety testing.

Testing procedures conducted by preclinical CROs are standardized by government regulation and all competitors conduct the same process. All of the major players in the marketplace produce a quality report on time, conforming to government regulations with the process being driven by experienced managers and technical staff. Companies differentiate themselves by experience, unique technologies, automation, scientific expertise, and service. The product is the process rather than the end report.

Pharmaceutical companies generally do not feel comfortable with laboratories that appear to be too academic or culturally aligned with practices in government laboratories. These groups are perceived to pay less attention to timeliness and efficiency and to have less understanding of the drug development process. Pharmaceutical companies contract work to laboratories that have familiar facilities and familiar cultural behavior. They contract to laboratories with a known track record or those laboratories that employ experienced staff familiar with the pharmaceutical corporate culture. David Cavalla [1] in *Modern Strategy for Preclinical Pharmaceutical R&D* stated it in this way: "Enduring added value can only come from a strong knowledge base of the core business. In this case this is the technical and scientific complexity of the pharmaceutical development process. CROs who can capture this knowledge through employing professionally experienced leaders, who can then cultivate the scientific culture of pharmaceutical development into their organization, will be the winners of the fight to capture increased market share."

## 4 LOCATING CANDIDATE CROS

For new startup companies it is important to begin the identification and evaluation of CROs early to enroll one or more in assisting development of their business plan. "Before a company even has a drug candidate to test, when it is still devising its business plan and is concerned with matters such as financing, infrastructure, and early-stage discovery, it should begin researching the services available for outsourcing, the economics of pursuing an outsourcing strategy" [2]. For established companies, it is important to establish an ongoing relationship with several CROs to allow them to move rapidly in outsourcing new drug candidate testing.

Colleagues who have worked with several CROs can be the single most useful source of locating companies that are candidates for your outsourcing requirements. They can provide the benefit of their experience in working with specific companies and their positive as well as negative experiences. Frequently, they will have useful contacts on the staffs of the CROs they have worked with. An individual new to the preclinical development business may not have colleagues in the business. Attendance at local and regional biotechnology meetings is useful in making contact with new colleagues on an informal basis.

Consultants in preclinical development, safety assessment, or toxicology work with many CROs and they are usually in the business of assisting in evaluation of CROs and in outsourcing studies. While the major CROs all conduct a broad range of studies and have most technologies available, they are not all equally adept at conducting studies that are not part of the standard core of toxicology studies. Consultants can be very useful in assessing specific capabilities. Interviewing and getting the assistance of one or more consultants for suggestions on CROs helps develop

a useful relationship with the consultant early on in the preclinical development process. If problems develop or unusual findings appear in safety studies, consultants can be invaluable in designing an approach to elucidate the problem and if necessary in working with the FDA to develop a strategy that will satisfy the agency's questions when the data are submitted in an IND. The consultant may not be needed for anything further in the preclinical development program but an established relationship is invaluable, should a consultant's assistance be needed.

Trade shows are a great place to review the capabilities of a large number of CROs, especially those with specialized technologies. The 2005 SOT meeting had 51 exhibitors who classified themselves as CROs. In addition, there were specialty laboratories with a single technology who didn't classify themselves as CROs. It is possible to meet scientists and scientific directors of the companies at shows such as the SOT Exhibition. It is possible to discuss a proposed development program and outsourcing requirements at the trade show and get a sense of the CRO's approach to a relationship with a new client. Develop a file of candidate laboratories for future reference and keep it up to date. We maintain a library of literature from at least 12 companies that we are interested in utilizing. In addition to literature, web sites are useful in getting an overview of a company's capabilities.

## 5 TELEPHONE INTERVIEWS FOR INITIAL SCREENING

Telephone interviews of candidate CROs save time and travel costs. They are useful in developing a short list of potential partners. The telephone will become one of the primary means of communication during the negotiating period and during study design, initiation, and conduct. The telephone interview will provide a sense of how easily the organization deals with you by phone. Determine if it easy to reach the scientific staff, whether you are routed only to marketing staff, and whether your calls are returned in a timely manner.

## 6 SITE VISIT: THE KEY TO EVALUATION

An actual site visit and inspection of the CRO's facilities and staff is the most important single activity in evaluating a new company. Mid-size and large companies considering outsourcing will conduct site visits by their in-house staff, while small virtual companies and new startups may conduct site visits through a consultant. Consultants often have working knowledge of the CRO under consideration and may not need to conduct a new inspection in the detail outlined here. For small companies relying on a consultant, it is still beneficial for a staff member to accompany the consultant on a site visit to meet the staff and begin to develop a working relationship. A site visit provides the opportunity to see work in progress and systems in action. It is the best way to evaluate security and the procedures to protect the confidentiality of data. Breaches of confidentiality can be very subtle and three examples follow. One of the authors (Richter) was alerted to a breach of security in his own CRO by a client. The client arrived at the laboratory early, 15 minutes before the front door was open for business. He walked behind some bushes adjacent to the front door to peer through the window. The desk of the director of

pathology was next to the window and he saw the business cards of two of his colleagues who had visited the week before and he could read pages of a report for his compound. He was very unhappy. One of the authors (Kopplin) inspected a GLP compliant CRO that was affiliated with a university, but operated as an independent agency. They were self-contained except for the ability to print reports and they used the university central reproduction facility in another building. Confidential reports of safety studies were left on a table in the hallway for pickup after completion. They were accessible to the general public walking by. This laboratory was not placed on a list of acceptable CROs. One of the authors (Richter) inspected a large well established CRO and was getting a tour of the facility. They were visiting the test material storage room when the host was paged to a phone call. The author was left alone with access to the stock test material from dozens of companies. The laboratory was not placed on a list of acceptable CROs. These examples provide insight into the type of things that can only be determined by a site visit. A checklist (Table 1) is provided to outline the conduct of an inspection of a new CRO.

## 7 EVALUATION OF THE PROCESS: NOT THE FINAL REPORT

While the checklist covers very specific items, the overall evaluation of the candidate CRO should depend on an evaluation of its culture and an evaluation of the process involved in conducting the work. Unlike the manufacturer of a widget, the process of conducting the research, interpreting it, and summarizing it in a report is critical to its quality and elegance. The widget manufacturer can tolerate mistakes by having a good quality control system that eliminates defective products. That simplistic approach to quality is not applicable to a highly regulated knowledge-based industry. Evaluating the process is fundamental to finding a CRO that meets your needs. All of the major CROs can produce a report on time that is free from error. The quality of the process is what differentiates them. Thus, the culture of the CRO, its systems, and its staff are critical elements in success. When inspecting a CRO, look for signs that the process has broken down, such as piles of work on a study director's desk and floor or rooms full of materials that should have been archived or disposed of. One of the authors (Richter) discovered that a CRO's inability to validate an analytical procedure occurred when a reference standard for another compound that should have been archived was left in the analytical laboratory and used by mistake.

The reliability and quality of the process influence the time to completion of studies and reports. Pharmaceutical companies are under strong pressure to cut the time from drug discovery to market. It is important to determine the CRO's record for on-time reports. Late reports are an indication of failure in the processes or inadequate staffing.

## 8 SPECIAL CONSIDERATIONS IN SELECTING A SPECIALIZED OR UNIVERSITY LABORATORY

Working with university laboratories can require some extra effort. There are times when a specialized procedure performed only in a university setting is needed to

**TABLE 1 Checklist for Site Visit Evaluation**

| |
|---|
| *Meet Key Personnel, Marketing, Management, Scientific and Technical* |
| Marketing Representative |
| Senior Scientific Management |
| Study Director(s) |
| Director of Pathology |
| Director of Clinical Pathology |
| Directors of Analytical and Bioanalytical Chemistry |
| Director of Quality Assurance |
| Director of the Pharmacy |
| Managers and/or supervisors of other specialty services or laboratories |
| *Staff Qualifications* |
| Resumés of key scientific and technical staff |
| Certifications, scientists and technicians |
| Training records, evidence of ongoing training |
| Publication record and presentations at scientific meetings |
| Consultant's and part-time specialist's qualifications |
| Determine whether you would be comfortable working with this staff; go to lunch and/or dinner with some of them; ask to meet key people in their office, view their work environment |
| *Special Things to Observe or Determine During Visit* |
| Determine in-house capabilities. Are any key services outsourced? |
| Assess the morale of the staff |
| Observe how busy the facility is. Are animal rooms mostly full or mostly empty? |
| Find out their on-time record for delivery of reports |
| Establish who will be your primary contact during negotiations and during conduct of work |
| Develop an impression of whether the laboratory management appear to be interested in working with you or whether your visit is treated as a routine |
| Evaluate the degree of automation for data capture and reporting |
| *Physical Plant Areas to Inspect* |
| Reception |
| Conference room |
| Offices |
| Laboratories, chemistry, histology, clinical, specialty |
| Necropsy room(s) |
| Animal rooms |
| Technician work areas, desks and facilities for data review and organization |
| Pharmacy |
| Information Technology |
| Restrooms |
| Cage washing |
| Storage, clean and dirty cages, food, bedding, supplies |

**TABLE 1** ***Continued***

| *What to Inspect Physical Physical Plant For* |
|---|
| Organization of overall facility for efficient work flow; user friendly |
| HVAC, environmental monitoring system, review temperature records |
| Cleanliness of the facility |
| Cleanliness of uniforms |
| Maintenance |
| Organized work flow in laboratories |
| Work flow into and out of animal rooms, clean versus dirty corridors |
| Security and confidentiality. Can you see any data left in open view during a tour? |
| Availability of SOPs in work areas |
| Availability of MSDSs in work areas |
| *Security* |
| Reception and visitor entrance |
| Employee entrance |
| Deliveries and loading dock |
| Data security, hard copy and electronic |
| Computer backup, stored on site or off-site |
| Disaster recovery plan |
| Compound and formulation security |
| Background check on employees |
| Is work left unattended on desks? |
| Pathologists computers on or off while away from desk? |
| End of work day security for active paper files? |
| Backup generator |
| Emergency planning: snow, sleet, or other event |
| *Standard Operating Procedures (SOPs)* |
| Review table of contents for general overview |
| Review key SOPs and those for specialized procedures needed for proposed project(s) |
| Are the SOPs reviewed and updated on a specific schedule? |
| Are the SOPs accessible to the staff in the laboratories and animal rooms? Ask to see the SOPs for a specific procedure during a tour of the laboratories. Is it readily available and does it appear that the staff are following the SOPs? |
| Is the format of SOPs user friendly? |
| Is the process of SOP approval overly bureaucratic? |
| Do you find hand-written notes derived from SOPs in the laboratory—a GLP violation? |
| Evaluate employee health and emergency medical assistance |
| Review animal care program |
| *Other Documentation* |
| FDA and/or EPA inspections and audits, key findings |
| Maintenance records on equipment, view examples for critical equipment |
| Sample contracts and confidentiality agreements |
| Sample protocols |
| Statistical procedures |
| Examples of reports—final, interim, and periodic progress updates |
| Facility certifications, local |
| AAALAC accredited; date of last inspection, inspection report |

evaluate the safety of a drug candidate. Unless the testing procedure has a widespread demand by the pharmaceutical industry, the laboratory is unlikely to be in compliance with the Good Laboratory Practices Act. The requirements of full compliance are of such a magnitude that it is not economically practical for the university laboratory to meet all of the specific details. The FDA and EPA will accept studies from these university laboratories, if the procedures are not available from a fully compliant GLP laboratory and if the laboratory meets the spirit of the GLP requirements. University laboratories generally do not have access to help in setting up the major requirements of GLP compliance. It will be necessary for the sponsoring organization to work with the laboratory to meet minimum standards. An inspection of the facilities and an evaluation of systems is the first step in outlining the deficits for the laboratory director and staff. We have found that the first and most glaring deficiency in many university laboratories is control of the test material and its security. Often the laboratories are used by graduate students who have unlimited access along with their visiting friends in the middle of the night. It is not difficult, but it is necessary, to develop locked and secure cabinets or rooms for the test material, related materials, data records, as well as the appropriate logs of the materials. The extent of installing GLP procedures depends on how critical the laboratory procedure is to determining the ultimate safety of the test article. The sponsor should be prepared to provide QA inspection of the work. At a minimum, the sponsor should be prepared to assist in any or all of the following:

Outlining the SOPs needed
Developing the protocols
Advising and/or assisting in the preparation of SOPs
Preparing forms for data recording, when not automated
Documenting staff performing individual procedures
Determining validation needs for instrumentation and processes
Maintaining CVs and training of staff
Defining raw data
Establishing data security procedures

We have also found that an hour of time spent in reviewing the background and need for GLP standards is productive with an inexperienced academic staff. Most individuals grasp the concept of documentation and chain of evidence for forensic purposes as an analogy and this goes a long way in achieving compliance with the general concept of Good Laboratory Practices.

## 9 SELECTING A SINGLE LABORATORY VERSUS SPREADING THE WORK AMONG MULTIPLE LABORATORIES

Early in the evolution of the CRO industry, many in the sponsor industry worried about the stability of an individual CRO and placed their studies in different laboratories because they felt it reduced the risk. If anything were to go wrong, financially, technically, or logistically in one lab and a study was compromised, the damage was limited. The rest of the development program would be safe.

With the maturation of the CRO industry, the individual companies have become financially stable, they have well established and reliable procedures, and they maintain a staff that is just as stable as that of the sponsoring organizations. With that change, it is usually wiser to place all of the development outsourcing with a single company once a working relationship has been established. The company will have experience with the compound, its formulation, and expected compound effects. Dose setting studies do not transfer well from one lab to another because of subtle differences in animal environment and care. A second lab will almost always insist on repeating dose-ranging studies before undertaking the next step in a development program even though dose ranging was done earlier at another laboratory. They will have learned from experience that they risk running a study that may not meet regulatory guidelines for establishing dose levels. It is critical to place major studies in the same laboratory that did the dose-ranging studies in the same species and strain of animals.

There are distinct financial advantages to placing all of a safety testing program at one laboratory in addition to eliminating the need for several dose-ranging studies. CROs will negotiate favorably on packages of all of the studies needed to reach an IND. Even those sponsors who only outsource one or two studies from a development program will get favorable treatment in pricing if they are repeat clients. However, it may be necessary to use several CROs because the laboratory of choice may not offer all types of studies or may be relatively inexperienced in a specialized study such as reproductive toxicology or genetic toxicology.

## 10 INITIAL SELECTION OF A CRO FOR A NEW PROGRAM OR SINGLE LARGE STUDY

For budget purposes, CROs will provide general prices for specific types of studies. The final price will depend on the specific protocol and variations of the protocol related to the type of compound, mechanism of action, and clinical indication being tested. CROs will also provide templates of protocols for sponsor review and modification or they will modify a protocol to meet the sponsor's specific requirements. Once the protocol is close to the final version, it is time to get specific prices for the work. This may involve submitting a single study for pricing or submitting a whole series of studies for pricing.

When outsourcing a new development package or a single large study such as a carcinogenicity study, it is useful to get prices from at least three different laboratories that have met your criteria during the screening process. The favored laboratory may not be able to start the projected studies on a realistic schedule because it is busy. Laboratories that are busy and fully booked usually do not negotiate on price. By letting the CRO know that you are getting three bids on the program, it will be sure to quote its best price for fear of being underbid by a competitor.

When getting bids from several laboratories and comparing prices, it is important to be certain that all are bidding on the exact same protocol. Rather than asking for the price of a 28 day rat gavage toxicity study, a specific and identical protocol should be submitted to each of the CROs. In that way, prices won't be affected by variations in the several laboratories standard protocols for that type of study. Inform the bidding companies that the protocol may be modified in final negotia-

tions but that the initial protocol will be the basis for selection based on price. It helps to build future relationships, if you agree to share the pricing information after bids are returned. It is not necessary to disclose the bidding laboratory, only the range of prices from them.

Since the laboratories are competitive, bids usually are quite similar. If they are, the choice should be your favorite providing they all meet your time frame for completion of the work. Occasionally, there may be a laboratory that is especially high and out of range. The high price may indicate that the lab is booked and overworked or that it didn't review the protocol carefully, both bad signs. Bids that are significantly lower than the rest may indicate that the laboratory is having difficulty marketing or that it did not review the protocol carefully. If awarded the contract, the lab could be tempted to cut corners to keep costs down. When we were outsourcing studies for our respective employers, we never recommended accepting a price that was significantly below others or that was well below the expected range based on previous experience.

## 11 TACTICS FOR DEVELOPING A RELATIONSHIP WITH THE CRO STAFF AND ENLISTING THEM AS MEMBERS OF THE DEVELOPMENT TEAM

Nothing is more important to the success of outsourcing than developing an effective working relationship with the staff of the CRO. They should be considered part of the development team with just as much invested in the success of the project as the sponsor's staff. They should share the agony of failure and the elation of success. Decades ago, many sponsors felt it necessary to develop an adversarial relationship with a CRO to ensure due diligence in the conduct of the project. Nothing destroys a successful outcome more than a paranoid attitude over the conduct of a study. It is possible to exercise proper quality control over the outsourcing process without making it an adversarial relationship. The staff of the CRO expect that the sponsor will be looking over their shoulders and evaluating their work as it proceeds but this process does not need to be confrontational.

Developing a relationship that makes the technical staff as well as the scientific staff feel part of the team helps ensure success of the project. A good sponsor will present one or more seminars on the drug candidate and the development progress for the CRO staff. Progress reports on the development program keep enthusiasm high for the testing project and encourage careful work on the part of the technical staff. CRO management usually encourages such inclusion of the staff as part of your development team. It helps morale of the staff to have a stake in the outcome of the drug candidate's progress through the development process.

## 12 DEALING WITH UNEXPECTED COMPOUND-RELATED CHANGES OR TECHNICAL PROBLEMS

It is human to immediately assume that errors were made in the conduct of a study when there are unexpected adverse effects. It is in the best interest of the CRO to detect errors if they occur and the staff will assist in resolving the cause of

unexpected results. One of the authors (Richter) worked with a sponsor on an intravenous infusion study that caused the death of all animals at the high dose level with severe damage to the veins. Studies on vein irritation had shown that the compound was not especially irritating. The sponsor immediately accused the laboratory of misdosing the animals, hired several consultants, and spent 3 months reviewing all of the records, interviewing technicians and scientific staff in an attempt to prove that the laboratory had made a serious error. The CRO staff had determined that the problem was related to the formulation within days of the deaths. The compound was only minimally soluble in the formulation and it precipitated out in the veins during infusion, actually obstructing them. It took the sponsor 3 months of lost time before the sponser accepted this interpretation and, working with another consultant, went back to the drawing board and reformulated.

## 13 MONITORING THE OUTSOURCED WORK

Monitoring of the work is a key to a successful outcome. Under GLP regulations, a study monitor is specified by the sponsor and this individual is the primary contact with the CRO's study director during the initiation, conduct, and reporting of the outsourced work. The study monitor should be in contact by email or phone at least weekly to review progress. The contract with the CRO should specify regular letter reports on a periodic basis, weekly or monthly, depending on the complexity of the study.

It is very helpful for the study monitor and/or additional sponsor representatives to visit the CRO laboratory site and review the work during the course of a study. The frequency and depth of these inspections depends on the complexity of the study. Short-term dose-ranging studies usually do not need a physical inspection unless the dosing procedure is unusually difficult. For definitive studies, such as 28 day, 90 day, or longer toxicology studies, it is desirable to have a sponsor representative on-site for at least the first day of dosing and the necropsies. Interim on-site inspections of the data on a monthly or quarterly schedule reduce the amount of time needed in reviewing the final report and in detecting problems. These visits are helpful in nurturing a sense of team work with the CRO staff and lets them know that they are part of the team. When unusual parameters are being measured or when a routine parameter is critical to the mechanism of action of the compound, the sponsor will want to be present for the event. Urinalysis is usually considered to be a very routine analysis in toxicological studies. We worked with a compound that was expected to cause changes in urinary parameters and this was so critical to the outcome of the study that sponsor representatives at the vice presidential level were on site to observe the urine collection and processing in the clinical pathology laboratory.

The best CROs occasionally make mistakes and sometimes these are not obvious to those who observe the work on a daily basis. A sponsor representative may look at the progress of the study with a different viewpoint. One of the authors (Richter) worked as a consultant for a carcinogenicity study at a CRO in which more than half of the female rats in all dose groups, including controls, developed uterine carcinomas. The spontaneous incidence of uterine carcinomas in rats of the strain used was less than 5%. The staff of the CRO were unable to determine the reason for

the increased incidence after detailed review of all records. After the author found that there was light-induced damage to the retinas, he suggested that increased estrogen levels were responsible and that excessive exposure to light could increase estrogen. The CRO staff carefully checked the timers for the animal rooms and found that they were correctly set to go off and on at the protocol specified times. After the author insisted that the timers be checked again, the building engineers revealed that they had disconnected the timers from the lights and although the timers were running, they had no effect on the lights. The rats had been exposed to light 24 hours a day for 2 years! No one had checked to see if the lights actually went off every night. The book by Gralla [3] covers many issues in monitoring toxicology studies.

## REFERENCES

1. Cavalla D. *Modern Strategy for Preclinical Pharmaceutical R&D* 1997;131.
2. Glaser V. *Outsourcing Drug Discov Dev* 2005;25(15):23–26.
3. Gralla EJ. *Scientific Considerations in Monitoring and Evaluating Toxicological Research*. Bristol, DA: Hemisphere Publishing Corporation; 1981.

# AUDITING AND INSPECTING PRECLINICAL RESEARCH AND COMPLIANCE WITH GOOD LABORATORY PRACTICE (GLP)

N.J. Dent
*Country Consultancy Ltd., Copper Beeches, Milton Malsor, United Kingdom*

**Contents**

1 Introduction
2 Objective of the Guidelines
3 Who Does It Affect?
4 Why Have It?
5 How Is It Enforced?
6 What Is GLP?
  6.1 Responsibilities
  6.2 Training and Recording: Confirmation of Suitability
  6.3 Quality Assurance
  6.4 Standard Operating Procedures (SOPs)
  6.5 Study Plans and Reports
  6.6 Data
  6.7 Equipment
  6.8 Computers
  6.9 Test Systems
  6.10 Test Substance
  6.11 Archives
7 How Can Compliance Be Maintained Within a Facility?
8 Pitfalls and Benefits
  References

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

## 1 INTRODUCTION

The Good Laboratory Practice (GLP) Guidelines have been in existence for nonclinical safety studies since 1976 [1–8]. They have progressed through various transitional phases to become guidelines in some countries and regulatory/statutory instruments in others.

The current document is the Organisation for Economic Co-operation and Development (OECD) Principles of GLP [9] and is currently accepted as the industry standard. This was reviewed and published in January 1997 [10, 11] but must be used in conjunction with the appropriate Scientific Guidelines for the scientific side of the study, that is, the OECD Toxicology Guidelines and so on [12]. This sets out to cover all nonclinical safety studies and gives guidance as to how these studies should be conducted in conjunction with the appropriate regulatory toxicology guidelines and, on that basis, when encompassed in the various Directives of the European Union (EU) [13] or in other Memorandums of Understanding, allows data generated under this program to be mutually accepted by other OECD countries. One must not forget, however, the other equally important guidelines and regulations of other countries, such as those of the U.S. Food and Drug Administration (FDA) [14] and the U.S. Environmental Protection Agency (EPA) [15–20], and similar organizations in Japan [21]. All have basically similar rules and, being members of the OECD, data generated to the OECD principles will generally be accepted in the United States and Japan. The EPA regulations used to be quite different and were applied to agrochemical and pesticide products. However, having been revised recently, they have been brought in line with the documents of other agencies.

The guidelines themselves, with the exception of those in countries where they are featured as regulations, are, as stated, guidelines to the conduct of the study and aim to cover compliance with the GLP principles but in no way do they dictate how the science will be performed.

It must be remembered that compliance is monitored by adherence to GLP, whereas the regulatory authority and the receiving authority of the dossier when submitted for the application of a marketing permit or similar document review the science.

## 2 OBJECTIVE OF THE GUIDELINES

The general objective of the guidelines originates in the very early 1970s, when one pharmaceutical company in particular and a contract research organization (CRO) generated data that, when submitted to the FDA, gave them cause for concern in the accuracy of the data presented and, in certain instances, the honesty of the submission.

At that time, a full review of companies and institutions conducting nonclinical safety studies (toxicology) was undertaken by the FDA [22], and in general, the industry was found to be credible. One company (Industrial Biotest) was found to be generating extremely poor quality data—in many instances, in a fraudulent manner. Just prior to this, GD Searle had made a submission to the FDA where some data were apparently missing but subsequently sent, causing the FDA again

to reconsider the total honesty of the pharmaceutical industry. The quote from the FDA was that it had "until that time assumed total honesty and integrity of the data supplied to the FDA for revision; however with recent events this has now given cause for concern."

Industrial Biotest basically set out with the prime objective to commit fraud, and many of the pesticide and agrochemical industries suffered through their lack of attention to detail, inventiveness, and basic dishonesty. Many of the test compounds sent for analysis and testing in animals for safety aspects remained unopened in the test substance store, although the client had received a report saying that the product had been tested and was extremely safe and able to be used. Around 200 studies were eventually rejected by the agencies, causing a great upheaval in the scientific research community and in the marketing of products. This caused the FDA to put together and implement the GLPs [23, 24].

Over the next 10 years, many countries introduced similar good practices guidelines. The EU [25] in general produced its guidelines, and eventually, despite the fact that the world was operating according to similar principles, a standard document was produced by the OECD in the early 1980s and became the industry standard. This was of benefit to the whole industry because it now precluded the fact that every submitting company would have to be inspected by each relevant monitoring authority and, when implemented into several directives and legal statutes [26, 27] within the OECD (particularly in the UK [28]), this allowed data generated by one company to be accepted by several receiving authorities without further inspection.

The objective of the GLPs is to ensure that a standard approach is undertaken, covering traceability and accountability and, while still allowing freedom for the scientists, to impose certain restrictions on the generation of data and the experimental work.

During the period of time when guidelines were being produced and inspections were taking place, several other instances of "dubious data" appeared. Notably, in the United States around the year 2000, a contract research company claimed full GLP compliance for its safety studies but in actual fact was cited by the FDA inspectors as having almost as many problems as Industrial Biotest. Later that year, an analytical contract research organization performing work for the EPA was found to operate a "two notebook system"—producing data as it was received from the machines and then manipulating that data to give better analytical results. Not only was this problem seen in the United States but similar instances were seen in laboratories in Scandinavia; in the United Kingdom a fraudulent submission was made to the regulatory authorities when a report was "doctored" to give better results on the product than had been seen in testing and, a product under license to a company was eventually found to have much of the data made up by the research technician.

The general aspect of nonclinical research could at this time be seen to be the ability to produce extremely poor quality data but in total compliance with GLP!

*One cannot, however, use GLP to prevent any scientific person who is intent on committing fraud in its true sense from performing that act; it, however, has in the past and will in the future continue to enable this practice to be exposed more easily.*

One should not consider that the whole scientific community is setting out to commit fraud or tempting to deceive the regulatory authorities; but, unfortunately,

the whole concept of good laboratory practices was brought about and subsequently reinforced by those several issues and had brought the scientific community into disrepute with the regulatory agencies.

It must be remembered that GLP is merely common sense in a formal environment. The key phrases that are currently seen in a GLP environment include good documentation, good training, maintenance and calibration of all equipment, the archiving and storing of data in a formal and retrievable manner, and the use of high quality, validated equipment and accredited test systems (animals).

This in general is merely good science, and the GLPs have further enhanced this by the addition of an independent Quality Assurance Unit (QAU) and a study director/principal investigator who jointly controls and oversees the project and involves the management in putting together adequate resources and assuming overall responsibility for the study.

This can be seen as good science, with several slight enhancements. The details of these individual subjects are addressed later in this chapter.

## 3 WHO DOES IT AFFECT?

*Any company or institution performing nonclinical safety studies for the submission of data for a new chemical entity; a new biological, immunological, pesticide, veterinary or agrochemical product; or, for that matter, a similar product that will eventually appear in the marketplace and be consumed by the general public must adhere to GLP in the conduct of their nonclinical safety study experimentation.*

Although initially the GLPs were seen to be a bureaucratic restriction on research and development, over the thirty or so years since this practice has been established it is amazing the number of disciplines and companies that persistently apply to the regulatory authorities to enter into their GLP monitoring program while not performing any safety analysis or supporting safety analysis: the food industry, the analytical industry involved in assaying samples from human clinical trials, the microbiology laboratories, central laboratories in hospitals, and central laboratories performing safety analysis for human clinical trials. The list is endless but it merely demonstrates that good laboratory practices have been seen as a quality standard that is internationally accepted and therefore there is a definite bonus in having GLP as one if not the only quality standard within a company.

Within a company, every person from senior management to the junior technician is bound by these GLPs and must exhibit clear understanding and training in these practices.

To ensure that the practices are followed, a regulatory inspection takes place on a 2 year basis in most countries, and the objective of this is to review, as an independent group, how these good practices are being followed. Certification or a guarantee that the company is operating according to these standards is the benchmark standard. This is also addressed later in this chapter.

As we move into the 21st century, it is quite apparent that the industry will shrink as mergers and acquisitions [29–33] take place, and, with this, the emergence of the now familiar CRO will become ever more popular in the conduct of nonclinical safety studies. It is therefore very important that, in this area, the sponsor has the assurance that these facilities are operating not only to the highest standard of

science but also in compliance with GLP and that, as a subcontractor, the data they generate will be equally accepted as if the data were generated by the company itself.

## 4 WHY HAVE IT?

In general terms, for companies conducting nonclinical safety studies [34], it is a regulatory requirement, and without this certificate or certification of compliance, data will generally not be accepted by the receiving/regulatory authorities.

However, one should not embark on the process of obtaining or working to GLP with this sole aim in mind. It should be used as an ongoing improving and quality standard for the laboratory. In fact, in my experience over the past five years, many companies have gone far beyond the requirements of GLP compliance [35], and the overall concept of good scientific design and good science has been superseded by the desire merely to obtain compliance. It is quite often seen that an extremely poor quality scientific study has been conducted in complete compliance with GLP. For example, on several occasions in a laboratory where the refrigerator was operated far from its permitted limits, but where the temperature was diligently recorded, signed, and dated as required by GLP, there was no attempt either to document the excursions outside the accepted range or to rectify the problem. The operative was merely under the impression that as long as temperature in the refrigerator is recorded, this is GLP despite the damage that excursions outside the temperature range may have caused to any investigational product stored in the refrigerator. Over the years, scientists who have worked according to the principles of GLP now readily admit without any prompting that they are unsure how they conducted scientific studies before the advent of these good practices: the ability to reconstruct studies, to work to a standard format across several differing laboratories or countries, and to be able to prove beyond reasonable doubt that these were the values obtained and the results submitted.

Certainly, data with a GLP compliance statement are being accepted more readily by the receiving authority, which has led to fewer repeated studies. This, in turn, is helping to achieve the aim of all scientists in reducing the use of animals.

From a company's point of view, working according to the principles of GLP shows that it has an attitude that is both ethical and moral to the production of scientific data with products that will eventually enter the human food chain or be of benefit to humankind.

## 5 HOW IS IT ENFORCED?

In the OECD countries, for at least 14 years, an Inspectorate has been set up, varying in inspector numbers from several hundred in the United States to one or two in countries not conducting a great deal of scientific nonclinical research. All countries, however, have a regulatory group that, in some instances, also acts as the receiving authority for the review of data and reports to the GLP Monitoring Authority (which visits on a 2 year basis or, in Germany, a 4 year basis, those companies who have claimed compliance and will then be on a rolling program of review [36]).

Unlike its role in many areas of regulatory compliance, it is still the responsibility of the sponsoring company to claim compliance from the Monitoring Authority. This claim is made for a particular company, laboratory, and/or series of tests. From the date of compliance when a letter is written to the Monitoring Authority, data generated from then on are assumed by that company to be in compliance with GLP. This claim in then verified in a visit from the regulatory inspector. The inspection may be performed by one or two persons for 1 to 5 days. At the end of the inspection, an exit meeting is held, and the company is usually given an indication of its performance. Noncompliance points are noted in writing and discussed, and a report is then prepared. In view of the findings, three levels of compliance can be obtained:

1. Sufficient deviations have been seen to question the integrity of the data and, therefore, a complete rejection of the claim of compliance is made, with a revisit necessary.
2. Minor points of compliance have been seen that can be handled in a specified period, in which case, the laboratory is placed under the category of pending compliance.
3. Very minor points of compliance have been seen, which, when addressed in writing by the management in a 1 month period with supportive paperwork and so on, lead to the company being given a Statement of Compliance, a Certificate of Compliance, or an indication that the laboratory is in compliance. It depends on the specific country whether a Certificate of Compliance is given. If a certificate is given, it generally states that on the particular day that the inspection took place, the laboratory was found to be in compliance with the OECD Principles of GLP. Also, the address of the facility is given as well as a listing of areas in which compliance has been confirmed. This could be stated as analytical support facilities, acute toxicology, mutagenicity, or similar designations.

Naturally, the benchmark standard is either the OECD Guidelines or similar standards in Japan or the United States. The Inspectorate carries out inspections against these documents. It could be said that often it is merely a review of the procedure and an opinion of compliance given by the inspector versus the interpretation of the individual conducting the experimental work. To try to overcome this criticism and to ensure that all inspectors work according to a standard format, over the past 4 years, the OECD has instituted a series of mutual joint visits (MJVs) [37]. The process of an MJV is that a company is inspected by its local inspector and that the inspector is accompanied by inspectors from two other countries as observers. At the conclusion of the inspection, the company is given the findings by its local inspector and, outside that meeting, a review of the performance of the inspector with positive and negative points is given by the two observing inspectors. To ensure continuity, one of these three inspectors would then be on the next MJV.

In the past, Memorandums of Understanding (MoUs) [38] have been instituted between certain major countries, such as Japan and the United Kingdom, the United States and Japan or Canada. However, these have generally fallen into nonuse for a variety of reasons, especially in Europe, where it is now, or has been for some time,

not possible for a country to negotiate directly with another country. Brussels, however, being the center of the European Community, has to carry out that discussion with a proposed partner in another country. As such, at the time of this writing, very few MoUs are currently in force.

## 6 WHAT IS GLP?

As noted in Section 1, GLP is a series of guidelines that cover the conduct and data production for nonclinical safety studies.

The OECD covers a series of activities and personnel. Responsibilities, training, quality assurance (QA), Standard operating procedures (SOPs), study plans and study reports, data production and recording, equipment maintenance and calibration, computers and validation, test systems, test substances, and archiving are the primary areas covered by the GLPs. A very brief overview of each of these areas is given hereafter.

### 6.1 Responsibilities

The prime players in a GLP scenario would be the *management*, the *sponsor*, the *study director, the principal investigator*, and the *QA unit.*

In a hierarchical structure, management would be totally responsible for the conduct of the work and for the assurance that resources have been made available and that an active role is played by these people in overseeing the conduct of scientific research.

The sponsor is the company that places a contract with a CRO or requests from within a company that work in another department be undertaken. The sponsor supplies the money and the request for the work.

The study director is prime player and is ultimately responsible for the production of the study plan, the conduct of the study, and the overseeing or production of the final report. Naturally, a large amount of delegation may take place; however, this must always be in writing. The overall responsibility for the conduct of the study, the daily contact with the study staff, the prevention or recording of problems, and the assurance that the study has been conducted in line with the study plan, the GLP, and the scientific guidelines solely belongs to this individual.

The principal investigator is next in the line of responsibility after the study director and is generally the person seen in a field study situation where the crop spraying, for example, may be undertaken at a place remote from the GLP designated site where the study director works. The principal investigator is therefore the person responsible initially for that portion of the work, although the study in total is under the direct control of the study director. It may also be that, within a company, work is subcontracted to the Analytical Department, for example, for the analysis of formulated material. The person responsible for this particular aspect of the scientific work is the principal investigator, who is involved in the study plan and responsible to the study director. Another typical scenario is work conducted in a CRO under the control of the study director, where samples of plasma are taken for toxicokinetics, for example, and these samples are analyzed by the sponsor. The sponsor's analyst, therefore, may well be designated the principal investigator.

The quality assurance unit is an independent group that does not become involved in the conduct of the study but merely reviews the data, experimental work, and documents produced to ensure compliance with the SOPs, the study plans, and GLP. Other activities such as training and assistance in interpreting GLPs may be the responsibility of the QAU. (For more information, see Section 6.3 in this chapter or Section II, 1.4 pp. 17–19 of the OECD GLPs.)

### 6.2 Training and Recording: Confirmation of Suitability

It is the responsibility of management to ensure that training takes place and the responsibility of the study director to assure that the individuals conducting the work are adequately trained and have adequate records. At a minimum, there must be a CV, a training record, and a very clear job description. Specifically, with regard to a study director, there must be explicit details of how the study director position can be met and the responsibilities of that individual in carrying out the relevant duties.

There should be procedures detailing how the training will take place; recording of the training must be made on a regular basis, the records must be stored in archives and regularly updated, and a complete and historical review of a trainee's activities, previous training, and ability to conduct the work according to GLP must be documented. (See the OECD Principles, Section I 2a through 2d.)

### 6.3 Quality Assurance

This function, as already stated, is an independent review. The responsibilities here start with a review of the study plan and continue through the review of the study in the in-life phase, data audits, and the final study report audit.

In addition to these, *system audits* and *process audits* can be undertaken.

There is a total misconception within the industry, particularly in new companies applying good laboratory practice, about the concepts of quality assurance, which is frequently confused with quality control.

The easiest way of describing this would be: *quality control* is a system of regular checks made either 100% or as determined in standard operating procedures by all operatives on data, machines, standards, and so on; *quality assurance*, however, is an independent observation that the system of quality control has been operated.

Therefore, it should be quite clear that quality control is a system operated by all parties and can if necessary encompass elements seen in other accreditation systems; but the overall confirmation that quality control systems have been carried out is the independent review on a random or total basis carried out by the quality assurance unit.

The aim of QA is to assure management that compliance with GLP is maintained throughout the entire study, that the data integrity is maintained, and that compliance with the SOPs and the study plan is adhered to by all experimental study staff.

The study audit is a specific audit of the study in direct relation to the study plan.

A system audit, however, rather than proceeding in a vertical line, takes a horizontal line across all studies and would include such tasks as archiving, training,

SOPs, general computer validation, animal house operation, and management activities. These are but a few areas that would constitute a system audit but they give an idea of the type of activities across studies that would be audited.

Process audits, on the other hand, have specifically been addressed in the revised 1997 GLPs, and these are basically aimed at auditing short-term studies of a repetitive nature, generally undertaken by similar teams of people. A process audit assures that the system is working and that parts of the process are reviewed over a quoted period in the QA SOPs. The aim is to ensure that all critical aspects of this process are reviewed through different studies over a period of time. This does not necessitate QA review of all short-term studies on every occasion, nor does it require the review of such areas as analytical analysis on a batch-by-batch basis or the analysis of hematology or biochemistry samples each time these come up for analysis.

To enable the reader to understand quality assurance, examples of some checklists are included in Table 1. These should be classed as the minimum requirements:

**TABLE 1 General Checklist For Audit/Inspection**

- Introductory meeting to discuss studies (ongoing/completed).
- Review documentation produced for initial meeting.
- Review organizational structure, management awareness, study directors, support personnel, safety aspects relating to animal liberation and to protect the data and client identity.

**Quality Assurance**

Structure and status
Master schedule
Protocol review procedures
Study phase inspection procedures and frequency
Process based inspection
Facility inspection
Report auditing procedures
Reporting systems

**Personnel**

To cover training, policy for training, records, job descriptions, CVs.

**Standard Operating Procedures (SOPs)**

Their production, content, authorization, distribution, staff awareness, review procedure, and locations.

**Animals**

Source and supply
Receipt, acceptance, quarantine
Environmental design and control
Operational systems, clean/dirty regime and so on for stock, supplies, and staff
Cleaning regime
Identification and treatment allocation
Veterinary management
Dose administration

**TABLE 1** ***Continued***

Supplies
Collection of body fluids
Measurement, body weight, diet, consumption, clinical signs, and so on
Waste disposal and incinerator

**Test Substance Control**

Receipt, quarantine, characterization, formulation, accountability, dispatch to toxicology.

**Laboratories**

Clinical biochemistry, methods of analysis, QC, sample receipt, sample treatment preanalysis, storage, raw data, reporting, authorization of results, equipment maintenance and calibration.
Hematology, as for clinical chemistry.
Analytical chemistry, as for clinical chemistry.
ADME labs, for specific studies (e.g., WBA).

**Computing Department**

Validation, maintenance, backup, reporting, security, data handling, data manipulation, programming and staff, CRF Part 11.

**Histopathology**

Autopsy procedures, dissection, preservation, labeling, processing.
Histology, tissue processing, slide production, slide QC, slide reading and recuts, reporting.

**Equipment**

All areas other than those specifically addressed above, logs, service contract, non-routine maintenance, and so on.

**Archives**

Design, security, operation, removal, return.

**Documentation**

To cover protocols, reports, amendments, their production, authorization, issue, review, and distribution. Raw data: what is it and how is it approved?

**Report Audit**

Toxicology reports should be examined to produce an audit trail in the reverse direction to the QA inspections carried out so far (i.e., report traced back through raw data to the animal requisition records).

**Management Briefing on Findings and Recommendations**

I am firmly convinced that the best tool given to a quality assurance auditor is the human body in terms of *eyes, ears*, and *nose*. While checklists can be a very useful memory aid, the best means of carrying out an audit is to look at the work being carried out, observe the operative in complying with the SOPs, and confirm adherence to the study plan or protocol. It must be of paramount importance to all staff and especially to the QA unit that the whole concept of quality assurance involves the study being in compliance with GLP, following the appropriate and up-to-date standard operating procedures, and being in strict adherence to the study plan or protocol. The scientific aspect is not the concern of quality assurance; this is solely the responsibility of the study director, the principal investigator, and the test facility management. Confirmation of satisfactory scientific study completion would be the basis of evaluation by the regulator.

Confirmation of quality assurance activities and compliance with GLP would eventually become the responsibility of the monitoring authority of the country concerned.

QA itself is required to produce SOPs that clearly detail operation, method of selection of critical phases, and studies, and to report its results to the management.

After every audit, a report is produced that is then discussed with the study director and circulated to the management with the overall agreement from the study director relating to the audit findings and their explanation of the resolution. (For in-depth review of QA, see Principles of GLP, Section II 2.1–2.2, p. 20.)

### 6.4 Standard Operating Procedures (SOPs)

These generally have been likened to a complete documented history of the entire aspect of conducting nonclinical safety studies. Any activity needs to be described in one of these documents. There may be a compilation of activities, or they may address single items such as the calibration and use of an electronic balance. They must be produced by the individual most familiar with the task, agreed on by management, and countersigned by a person senior to the author of the document.

Once produced, SOPs must be reviewed on a regular basis (approximately every 2 years), and any changes to these procedures must be made in writing, with the agreement of all parties, and circulated to each owner or user of the SOP. The SOP itself must be filed in the archive and additional copies produced. An SOP management system must be set up, whereby a responsible person knows the whereabouts of all SOPs and can retrieve and replace them with amended or superseded revisions and can make sure that they are reviewed regularly and disposed of when no longer required.

The SOP must appear immediately in the area adjacent to the workplace to be readily available to all persons. Frequently, SOPs are the basis of training, and most companies now have SOP-based training schemes. The content and receipt of the SOP should be acknowledged immediately upon receipt and a training program set up, whereby confirmation of the understanding and the ability to perform the duties stated in the SOP is documented in the appropriate training record.

SOPs should be adequately controlled to prevent unauthorized photocopying, which may lead to a superseded copy not being sent to the proper recipients. Someone making an illegal photocopy would not be on the distribution list and

therefore would not always receive updated versions, with the possibility that an outdated method could be used.

The requirement for archiving historical copies is one of the key attributes of GLP in that traceability can be seen as originating at the archive. The dates and historical record of the SOPs can prove irrevocably that a particular action was the method in use at the time.

SOPs can be paper-based or electronic. The trend toward electronic record-keeping is becoming more common in laboratories. The only requirement made by the inspectorate is that accurate, controlled copies are available on the electronic media and that prevention of copying or unauthorized changing are built into the SOP system. (See GLP Principles Section II, 7 p. 24.)

### 6.5 Study Plans and Reports

Before any study can be undertaken satisfactorily, a study plan must be produced. The study plan is merely an indication of all of the activities that will take place, resources required, time frames, and objectives. The study plan can be likened to a road map that, when given to all the participants, will allow them to start at the beginning and to proceed through the various mazes to the final completion point indicated by the study report. The one golden rule in GLP is one study plan, one study director, and one report.

The study report itself is a mirror image of all the headings in the study plan and serves to confirm that the objectives of the study have been met and that the results and discussions of the data presented give an indication of the outcome of the particular experimental work.

Both the study plan and the report are audited by QA, and each study plan and report are generally determined by the company's format.

Typical headings must be given in both documents. (These can be found in Section II, 8.1–8.3, pp. 25–27 and 9.1–9.2, p. 28, of the OECD Principles of GLP.)

### 6.6 Data

Raw data, or source data, are generally considered the first records made, either electronically in computer-readable form or created the first time that the "pen hits the paper."

These should be original, signed and dated recordings that may be on any type of medium. When results are stored on media such as heat-sensitive paper, the results should be photocopied to insure against deterioration of the medium over time. Electronic data can be regarded as a disk, CD-ROM, or similar medium provided that this material, when reintroduced to the computer and the software, can generate the images stored on the disk or electronic medium in a 100% readable form.

There are many types of electronic media, machines, and source data or raw data within the toxicological environment. Ironically, the most common storage medium for raw data is paper. Paper and its storage partner microfilm have been around for many years, and their stability and reproducibility are well known. Other electronic media, however, do not have the same capability of reproduction known over a long period, and thus most industries and companies prefer paper.

The data should be recorded promptly, legibly, and signed and dated, and any corrections should be made in a format to allow the original record to be seen, the change should be described and justified where applicable, and the change should be signed and dated by the individual making the revision. This procedure, whether on paper or via computer, should have the same standards. With use of the computer, an audit trail is necessary to identify the change and the person making it, along with the reason.

### 6.7 Equipment

As can be imagined, equipment in a toxicological study may be varied, simple, or complex. As such, it is difficult to describe each individual type of equipment in this limited space.

GLP requires that equipment be maintained, calibrated, and generally demonstrated as fit for use.

Equipment such as can be seen in an analytical laboratory supporting toxicology studies, for example, high pressure liquid chromatography (HPLC) and other equally sophisticated equipment such as GC-MS/MS, should have checks of system suitability and installation, operation, and performance qualifications performed *at a minimum*.

Other equipment such as centrifuges and balances should be maintained and calibrated and, with regard to the latter, regular checks should be made with known, standardized, regularly calibrated weights. These should be placed on the balance with a frequency to guarantee that data from the machine are accurate. Even if the balance is an electronic calibrating balance, regular manual check weights should be applied.

Each piece of equipment should have a log book that gives a historical record of its use, breakdown, repair, and service. Generally, it is acceptable that these log books be placed by the equipment generating critical data to be presented in the final report.

All equipment should be clearly identified as to the time that it started producing raw data for experimental use and, when no longer required, the equipment should be removed from the laboratory or suitably labeled "not for GLP use."

The calibration and validation of equipment have been addressed extensively, but equipment that can be shown as "fit for purpose/use" within the GLP environment is generally acceptable to most regulatory inspectors. (For additional information on the GLP requirements for equipment, see Principles, Section II, 4, p. 22.)

### 6.8 Computers

Over the past few years, computers have played a very important role in many aspects of toxicology.

The general trend in the industry and particularly from the Inspectorate is to ensure that the computers are fully validated. Validation, however, means different things to different people. Some companies and their Information Technology Group (ITG) will dismantle the computer and its software components, reconfigure them, test them, and then reinstall them. Others will take a more realistic approach and work on the basis that the computer was brought in for a specific task and is

considered validated provided that task is completed with the aid of the computer in a reproducible and acceptable manner.

However, in the most simplistic form, validation could be covered by *evidence that the computer will perform the task for which it was purchased and, more importantly, continue to perform that task for the foreseeable future*. In other words, as with other equipment, is the computer fit for the purpose?

Several documents have been written from a regulatory standpoint, the most useful being Monograph 10 of the OECD Principles, Application of GLP to Computer Systems [39]. Many books are available and vary in detail and content to cover everything that one would wish to know about computers but were afraid to ask, down to the simple documentation giving the essentials for validation and providing a disk with the SOPs to comply with GLP [40, 41].

The prime concern of the Inspectorate is that the user responsible for performing the validation and producing the report is in control of the equipment and can ensure and prove the integrity of the data when entered into the computer and regenerated in some other form.

It is generally accepted in the industry that acceptance testing is perfectly satisfactory for most computers and assures that the computer, when installed on company premises, will perform the function for which it was purchased. However, each computer must be viewed in the role it will play in the company, and suitable testing must be conducted to ensure that the data and integrity are of the highest quality and that total control over output is maintained.

As with all equipment, computer maintenance and calibration records are of paramount importance.

If in-house software programs are produced, they are tested and validated, and the source code is made available. One of the key elements required in the computer record-keeping is that of change control and password protection and training. (See GLP Principles Section II, 7(b), p. 25.) One very important rule is that the electronic signatures rule [42] must be observed when data are signed off electronically. This FDA requirement became effective 20 August 1997 and covers all data for which signatures are made electronically and requires that the FDA is officially notified.

However, in 2005 the FDA radically reviewed its requirements for electronic signatures and Part 11 compliance and basically required the industry to ensure that any use of computers in any form within a GLP study should follow the predicate rules, which were predominantly explained in the early good laboratory practices, and to confirm that the system is fit for use, adequately controlled, validated, password protected, and in essence completely controlled by the user and company. Much of what was seen to be bureaucratic and totally overenthusiastic approaches by the FDA and their guidance manuals were withdrawn at this time, but Part 11 is still alive and must be strictly adhered to in the situation where data are sent in whatever form to the FDA.

### 6.9 Test Systems

This really is a slightly complex name for what is generally considered the animal subject. Test system, however, has been utilized because, in many instances in toxicology, GLP now applies to such subjects as groundwater, soil, insects (such as

earthworms and honey bees), and microorganisms (such as daphnia) and therefore the use of the word animal is not always applicable.

The main criteria are that the origin of the test system is known with its breeding history, where applicable, that these are purchased from well-known and, if possible, accredited suppliers, and that the quarantine period is observed to ensure that test systems are of high quality and fit for use.

Care, husbandry, intermediate sacrifice if the test system is found to be "in extremis," and humane sacrifice before necropsy are essentials for the test system. Separate housing among species and experimentation is critical, and all aspects of manipulation of the animal from clinical observations, dosing, and special tests such as electrocardiogram (ECG) need to be well documented and outlined in standard operating procedures.

Animal husbandry itself, the animal room, and the animal room diary giving an indication of exactly what occurred in the room and to the animal are essential items of documentation. Unique identification is also of paramount importance with the animals, cages, and the location of the cages.

The industry has very recently turned its attention to the use of individually ventilated cages (IVCs). The plus points for this are that the clean and dirty corridor situation, which has so frequently been vastly expensive in terms of HVAC, is no longer required and that a room with adequate ventilation and temperature control can house the IVC units, which are individually controlling the environment for the animal so that each cage can be considered as an individual animal room. Thus, for short-term toxicology studies, many more studies can be carried out in the same room, which was prohibited by earlier GLPs. Acute studies can therefore be conducted, and no longer is the volume of the room considered an issue by the toxicologist or regulator. New animal facilities can be constructed much quicker and at less cost than involved in the expensive clean and dirty corridor situation and ventilation. Existing animal units can readily be modified with a minimum of disruption to studies and with a minimum of expenditure in terms of reconstruction and HVAC. The animal itself is in a much more controlled environment, but the cautionary tale is that if an organization goes for IVCs, it is essential that the whole system be purchased: this includes not only the station controlling the temperature, humidity, and airflow within the cage, but cage cleaning stations and dosing and handling stations—all of which are filter air controlled to maintain the overall sterility and critical ambience of the experiment. Failure to do this and the purchase of only a partial system will doom the study to failure.

This system has been put into place where specific strains of mice, for example, are required to be used and their delicate nature would not allow for conventional animal housing; feeding studies particularly benefit here as the mass dust and contamination between study groups can be much more tightly controlled within the individual cage environment. The QA unit and regulator, however, must review these facilities in a totally different light with slightly more flexibility.

Compliance with GLP, however, brings particular problems and much of the QA audit and standard operating procedures need revising as do the formal GLP inspections of animal units comprised entirely of IVCs [43].

Full and documented histories of heating and ventilation are required, and, in barrier-maintained rooms, signed and dated records of positive to negative pressures are to be kept. These records should also reference any malfunction and its rectifica-

tion. Furthermore, excursions outside the permitted range must be documented, and the effect on the study and data integrity must be identified and addressed by the study director in the final report. (Additional information is available in the GLP Principles, Section, II, 3.2, p. 21.)

### 6.10 Test Substance

In most instances, the test substance would be a new chemical entity (NCE) or an existing product; a comparator, pharmaceutical, veterinary, or agrochemical product; or even a device. Knowledge of the composition, characterization, stability, and other physiochemical properties is essential. Stability, however, may be determined as the short-term studies progress, with the proviso that the overall stability is known, along with full characterization by the time long-term toxicity studies are carried out. Stability testing may well be carried out in parallel as long as the stability of the active ingredient and formulated product is known sufficiently to allow for control of the dosing to be done within the period of stability known at that time.

One of the key elements of test substance control is accountability. A record of the amount received for toxicity testing should be accurately recorded, and 100% accountability of that product throughout the life of the testing is an essential element of GLP.

Again, formulation of the product is required to be covered in detail, and, in many companies, the elements of GLP are the benchmark standards when dealing with a test substance. Use of the test substance in the animal facility, the maintenance of homogeneity of suspensions, the mixing of the product in feed, and the testing of the product are all essential. This is one particular area in which, within the toxicology testing area, support functions such as analytical studies then come under GLP.

These functions will be required to test formulations and feedstuffs to ensure that the correct amount of the active ingredient is present as determined by the study plan for the various dosing groups. This requires that a validated method be available before any work is carried out, with the ability to analyze samples of the formulated product or plasma samples for toxicokinetics as the study progresses.

It is required that a reserve or retention sample of the active ingredient be retained. This should be retained for as long as it affords reasonable testing and within the expiry period determined by the analytical facility. It is also required that a retention sample be retained for studies that are not considered to be short term. This is one particular area in which revision of the GLPs is sometimes not well thought out. Originally, it had been stated that reserve samples should be taken for studies exceeding 4 weeks. This was subsequently revised to specify "studies that are not considered to be short-term." The glossary in the GLPs defines a short-term study as "a study of short duration with repetitive processes." This, one must admit, does not give a lot of guidance! (See additional information on systems in GLP Section II, 3.3, p. 21 and 6.1–6.2, p. 23.)

### 6.11 Archives

Having addressed all the various aspects of the study, one can see that much documentation, tissues, slides, and wax blocks could well be accumulating. The require-

ment is to store this material for "a period of time." Again, very little guidance is given in the GLP, and one is referred to the national guidelines for the storage of data. However, it is of great importance that this material be maintained in good condition in a retrievable format for at least 15 years or for 2 years past the availability of "the product," whichever is longer.

In 2006 the OECD Principles and GLP Compliance Monitoring Group produced a monograph detailing archiving in a GLP environment. This document gives valuable advice to the problematic subject of archiving [44].

Generally, companies themselves are maintaining that material for far longer, or for 2 years past the availability of the product.

All the material must be retained in a secure location for easy access, under the responsibility of a management-designated archivist and deputy. The security aspect of the archive should preclude damage from outside sources, fire, water, rodents, and so on.

Over the years, it has been seen that vast amounts of data must be retained. Most companies, bearing in mind the fifteen year retention period, tend to carry these documents in their archives for considerably longer periods of time. Many authorities still allow electronic archiving such as scanning to CD/DVD, but the FDA still requires the original paper document to be retained. This has led to a large explosion in contract archiving houses. While these are perfectly satisfactory, it is essential that the company QA unit inspect these for safety, security, long-term financial viability, and most importantly the general concepts that you would expect to see in your own archive. During 2006 the fragile nature of contract archiving was exposed when one major international company having a unit in London caught fire and at the same time a similar unit in The Netherlands also caught fire, both causing major loss of documentation. Either way, whether it is your own archive or the contract archive, this problem is always waiting around the corner.

The entire aspect of archiving is basically one of common sense, and guidance on how to archive these materials can be obtained from government agencies that store personnel records or from libraries. (Additional information on archiving is available in GLPs under Section II, 3.4, p. 22 and 10, p. 29.)

## 7 HOW CAN COMPLIANCE BE MAINTAINED WITHIN A FACILITY?

Compliance, having been granted after an inspection, should be monitored on a daily basis. However, it is my opinion that many companies' standards of compliance relax after the initial certificate has been granted only to find that, 2 years later, for example, an enormous rush 1 month before an announced inspection is required to generate the appropriate documentation and to update the system.

It is suggested that QC reviews be carried out on a regular basis to ensure that points likely to detract from the overall compliance are reviewed regularly and that project meetings be held where QA is invited to give a précis of the regular points seen during audits so that these can be addressed and rationalized.

Training and retraining, along with an awareness of the requirement to comply with GLP, are of immense importance. New equipment, major SOP revisions, and transfer of technicians or scientists among departments are always good signs that

additional training should be carried out. It must be remembered that GLP is teamwork. It is no good considering that there are the scientists and technicians on the one hand and QA on the other. There is no point in hoping that whatever happens and however little QC is carried out, QA will discover all the mistakes in the final report and review. Remember, it is not QA's problem; that department's role is to ensure that compliance has been maintained; QC and data checking are the responsibilities of every member of the staff team.

Improvement targets should be set in line with quality control manuals used in other accreditation systems. It is always a good point to review internally and on a regular basis (1) problems that have been encountered in experimentation, (2) audit findings, (3) ways to improve work by looking at new systems and reviewing SOPs to ensure that these are current, and (4) areas where improvements can be made.

An example can be taken from the accreditation systems [45], in which, in addition to QA audits, departments become involved in self-inspection. Each department can identify a QA representative whose daily responsibility is to review compliance issues, to look at the overall quality policy of the company, and to ensure that, between QA audits, self-inspection is performed and that a departmental review is made of these findings with action points and a time plan identified.

The primary impetus for the maintenance of compliance, however, is the regular external inspection by the Inspectorate. In addition, it is now becoming frequent for independent consultants to be brought in to do preregulatory inspections. Whichever way one views the system, whether through consultation or by assigning a department to perform inspections in one area and to conduct audits in another, the regular review of compliance should be maintained.

When using CROs, the whole aspect of auditing takes on a different light. Here, subcontracting is usually performed because of internal pressures, shortage of space, or lack of in-house expertise.

Dealing with CROs is no different from setting up an in-house GLP system. The CRO should be regarded as an extension of the facility in which the sponsor is conducting its own research.

## 8 PITFALLS AND BENEFITS

In conclusion, it is worthwhile to address the pitfalls and benefits of operating according to the principles of GLP.

A pitfall could be seen as a restriction on the scientist against performing free research. It could also be seen as an intrusion by an independent body looking at why problems occur and at the sorts of problems that occur and carrying out regular reviews with senior management about these problems. Costs will increase because of time pressures and the necessity of involving third-party reviews. The recording of data will now be subject to more QC, more required approvals, extra costs, and, generally, more data presented. Time must be taken to write and review SOPs. This in itself can be a very costly exercise; I know of one company that, having spent more than 6 months writing its SOPs, classed them as capital pieces of equipment and put a value of $20,000 on that volume.

Other companies and personnel may encounter similar pitfalls. The list is not intended to be exhaustive but merely to indicate areas in which additional time, money, and resources will be allocated. However, on the positive side, benefits can be seen immediately.

In talking to many people who have operated under the GLP system for the past 30 years, it is generally heard that the system allows for a better standard of research, less repeated work, the ability to have full accountability and traceability of everything within the experimental phase, and the knowledge that all documentation produced at the end of the study is now safe and secure in the archive and can readily be accessed for regulatory review or inspection.

Fewer studies are being repeated, and, therefore, the immediate benefit is the lowering of subject usage. The fact that data, when generated with a certificate or a compliance statement, will now be accepted by all OECD member countries means that once the study is completed and the regulatory submission made, the time for acceptance by several countries (if submissions are made in a multistate procedure) will be reduced dramatically.

Finally, the initial bureaucratic straitjacket of GLP thrust on the international research community in 1976 has rapidly turned full circle and now is seen as the quality standard to which all companies in all countries want to aspire. From that point of view, all nonclinical safety studies, when conducted according to the principles of GLP and adequately addressing science as well as compliance, can achieve a very high success rate both in the outcome of the science and in the acceptance of data for a regulatory submission.

## REFERENCES


1. US Food and Drug Administration (FDA). Nonclinical Laboratory Studies. Proposed Regulations for Good Laboratory Practice Regulations. Proposed Rule. *Federal Register* 1976;41:51206–51228.
2. US Food and Drug Administration (FDA). Nonclinical Laboratory Studies. Good Laboratory Practice Regulations. Final Rule. *Federal Register* 1978;43:59986–60020.
3. US Food and Drug Administration (FDA). Good Laboratory Practice for Nonclinical Laboratory Studies: Amendment of Good Laboratory Practice Regulations. Amendment. *Federal Register* 1980;45:24865.
4. US Food and Drug Administration (FDA). Good Laboratory Practice Regulations. Proposed Rule. *Federal Register* 1984;49:43530–43537.
5. US Food and Drug Administration (FDA). Good Laboratory Practice Regulations. Final Rule. *Federal Register* 1987;52:33768–33782.
6. US Food and Drug Administration (FDA). Good Laboratory Practice Regulations. Minor Amendments. Final Rule. *Federal Register* 1989;54:15923–15924.
7. US Food and Drug Administration (FDA). Good Laboratory Practice Regulations. Removal of Examples of Methods of Animal Identification. Final Rule. *Federal Register* 1991;56:32088.
8. US Food and Drug Administration (FDA). Good Laboratory Practice Regulations. Technical Amendment. *Federal Register* 1994;59:13200.
9. OECD Series on Principles of Good Laboratory Practice and Compliance Monitoring. ENV/MC/CHEM/9817. As revised in 1997. http://www.oecd.org.ehs/ehsmono.

10. Dent NJ. Comparison between the principles of GLP 1992 as defined in Monograph 45 and the draft revision of the principles of GLP as published by OECD on 6th January 1997. *Qual Assurance J* 1997;2:7–12.
11. Dent NJ. European regulatory compliance issues: good research practices. *J Am Coll Toxicol* 1994;13 (1):79–85.
12. *OECD Guidelines for the Testing of Chemicals*. Paris: OECD; 1993.
13. Directives EEC/87/18, EEC/9, EEC/88/320. *Official J Eur Commun*, 1992.
14. US Food and Drug Administration. *Good Laboratory Practice for Nonclinical Laboratory Studies: Title 21, Part 58, Code of Federal Regulations*. FDA; 1993.
15. US Environmental Protection Agency (EPA). Pesticide Programs: Good Laboratory Practice Standards. Final Rule. *Federal Register* 1993;48:53946–53969.
16. US Environmental Protection Agency (EPA). Toxic Substance Control: Good Laboratory Practice Standards. Final Rule. *Federal Register* 1993;48:53922–53944.
17. US Environmental Protection Agency (EPA). Toxic Substance Control Act (TSCA): Good Laboratory Practice Standards. Final Rule. *Federal Register* 1989;54:34034–34050.
18. US Environmental Protection Agency (EPA). Federal Insecticide, Fungicide and Rodenticide Act (FIFRA): Good Laboratory Practice Standards. Final Rule. *Federal Register* 1989;54:34052–34074.
19. Good Laboratory Practice Standards (FIFRA), EPA. *Federal Register* 1993, Title 40, Part 160, Code of Federal Regulations, Final Rule revised as of July 1.
20. Good Laboratory Practice Standards (TSCA), EPA. *Federal Register* 1993, Title 40, Part 792, Code of Federal Regulations, Final Rule revised as of July 1.
21. Shillam KW. GLP Legislation. In Carson PA, Dent NJ, Eds. *Good Laboratory and Clinical Practices: Techniques for the QA Professional*. GLP Regulations Management Briefings—Post Conference Report. Rockville, MD: US Food and Drug Administration; 1979, pp 16–29.
22. Lepore PD. Overall responsibilities for GLP. In Carson PA, Dent NJ, Eds. *Good Laboratory and Clinical Practices: Techniques for the QA Professional*. Rockville, MD: US Food and Drug Administration; 1979, pp 29–39.
23. Lepore PD. Good Laboratory Practice Regulations: Questions and Answers, Subpart B. Available through FDA FOI, FDA, (HFI-35), Rockville, MD, 1979.
24. Lepore PD. Good Laboratory Practice Regulations: Questions and Answers, Subpart B. Available through FDA FOI, FDA, (HFI-35), Rockville, MD, 1981.
25. Broad RD, Dent NJ. An introduction to Good Laboratory Practice. In Carson PA, Dent NJ, Eds. *Good Laboratory and Clinical Practices: Techniques for the QA Professional*. Rockville, MD: US Food and Drug Administration; 1979, pp 3–16.
26. Commission Directive 1999/11/EC of 8 March 1999. *Official J Eur Commun* 1999; L77/8. 23.03.99.
27. Commission Directive 1999/12/EC of 8 March 1999. *Official J Eur Commun* 1999; L77/8. 23.03.99.
28. Good Laboratory Practice Regulations 1999. Statutory Instrument 1999 No. 3106. http://www.hmso.gov.uk/si/si1999/19993106.htm.
29. Dominated by the urge to merge. *Scrip Mag* 1999;Jan:46.
30. A strategic approach to R&D outsourcing. *Scrip Mag* 1999;July/Aug:29.
31. R&D trends give clues to future. *Scrip Mag* 1999;July/Aug:17.
32. How much of a threat is virtual pharma? *Scrip Mag* 1999;Nov:41.
33. The right treatment for the right patient. *Scrip Mag* 2000;Jan:11.

34. OECD Glossary. OECD Series on Principles of Good Laboratory Practice and Compliance Monitoring. ENV/MC/CHEM/9817. As revised in 1997. http://www.oecd.org.ehs/ehsmono.
35. Dent NJ. Forget compliance and concentrate on science. *Qual Assurance J* 1998;3:103–108.
36. Personal communication from OECD.
37. Personal communication from OECD.
38. Personal communication from OECD.
39. OECD GLP Consensus Document on The Application of the Principles of GLP to Computerised Systems, *Environ Monograph* 1997;10:116. OECD Series on Principles of Good Laboratory Practice and Compliance Monitoring. ENV/MC/CHEM/9817. As revised in 1997. http://www.oecd.org.ehs/ehsmono.
40. *The Red Apple II Symposium*, sponsored by DIA; (in press).
41. Chamberlain R. *Computer Systems Validation for Pharmaceutical and Medical Device Industries*. Buffalo Grove, IL: Interpharm Press; 1994.
42. US Food and Drug Administration. Electronic Records; Electronic Signatures; Final Rule. *Federal Register* 1997;62:13429–13466.
43. Dent NJ. Quality assurance and the use of individually ventilated cages (personal communication). Presentation to the British Association of Research Quality Insurance Symposium, Dublin.
44. Lawrence RG. Archives. In Carson PA, Dent NJ, Eds. *Good Laboratory and Clinical Practices: Techniques for the QA Professional*. Butterworth Heinemann; 1979, pp 155–171.
45. Piton A. Quality assurance: the present and future role of international standardization. In Dent NJ, Ed. *Implementing International Good Practices—GAPs, GCPs, GLPs, GMPs*. Interpharm Press; 1993, pp 251–256 (ISBN 0-935184-44-9).

## Useful Websites and General References

Kracht WR. Implementing the ISO 9000 quality standards in a GMP or GLP environment. In Dent NJ, Ed. *Implementing International Good Practices—GAPs, GCPs, GLPs, GMPs*. Interpharm Press; 1993, pp. 267–277 (ISBN 0-935184-44-9).

Carson PA, Dent NJ. *Good Laboratory, Clinical and Manufacturing Practices*. London: Royal Society of Chemistry; 2007 (ISBN 978-0-85404-834-2).

Dent NJ, ED. *Implementing International Good Practices—GAPs, GCPs, GLPs, GMPs*.

*Good Research Practices—A Practical Guide to the Implementation of the GXPs*.

OECD homepage http://www.oecd.org.

European Quality http://www.euroqual.org.

US Federal Register—Environmental Issues http://www.epa.gov.fedrgstr.

GALPs http://www.epa.gov/doc/irm_galp.

EPA homepage http://www.epa.gov.

International Organization for Standardization http://www.iso.ch.

Good Laboratory Practice on Line http://www.glpguru.com.

Compliance Programme Manual http://www.fda.gov.ora.cpgm.default.htm.

Title 21 Code of Federal Regulations Part 58 http://www.access.gpo.gov/nara/cfr/index.html.

GLP Regulations Questions and Answers http://www.fda.gov/cder/guidance/index.htm.

Website for further information: www.countryconsultancyltd.co.uk.

# DRUG IMPURITIES AND DEGRADANTS AND THEIR SAFETY QUALIFICATION

Robin C. Guy
*Robin Guy Consulting, LLC, Lake Forest, Illinois*

**Contents**

1 Introduction
2 Types and Sources of Impurities
  2.1 Manufacturing Impurities
  2.2 Degradants
  2.3 Leachables and Extractables
3 Regulations
  3.1 International Conference on Harmonisation (ICH) Guidance
  3.2 U.S. FDA (CDER)
4 Technigues
  4.1 General
  4.2 Ultraviolet and Visible (UV/Vis) Spectroscopy
  4.3 Atomic Absorption Spectroscopy
  4.4 Thin Layer Chromatography (TLC)
  4.5 High Pressure Liquid Chromatography (HPLC)
  4.6 Mass Spectrometry (MS)
  4.7 Gas Chromatography (GC) and Combination of GC and MS
  4.8 MS/MS
  4.9 HPLC/Mass Spectrometry (LC/MS or LC/MS/MS)
  4.10 Common Problems
5 Toxicity Assessment
  5.1 Literature-Based Approaches
  5.2 Common Testing Requirements for New Impurities
  5.3 Study Design Based on Exposure, Route, Dosage, and Other Factors
  5.4 Documentation Content and Format

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

6 Bridging Studies
7 Discussion Regarding Testing the Drug with the Impurity Versus Actual Impurity
Acknowledgments
References

## 1 INTRODUCTION

In the United States, since very early times, traveling salespeople touted their latest miracle cures with enough publicity and fanfare to attract many to their circus-like shows. These cures were ointments, elixirs, and creams that would cure everything including obesity, diabetes, tuberculosis, and cancer [1]. Unfortunately, almost all of these products were either toxic or not efficacious. Even some vaccines were not safe, as evidenced in the deaths of thirteen children from St. Louis, Missouri after receiving an inoculation from a diphtheria antitoxin that was contaminated with tetanus [2]. This incident prompted Congress to enact the Biologics Control Act of 1902, a law that required inspections of the manufacturers and sellers of biological products, including testing of these products for purity and strength [3, 4].

In 1906, due to the outcry from the public after publication of Upton Sinclair's novel, *The Jungle*, the original Food and Drug Act was passed by Congress. The law prohibited adulteration and misbranding of food and drugs. Offending products could be seized and condemned and offending persons could be fined and jailed. Drugs had either to abide by standards of purity and quality set forth in the United States Pharmacopoeia (USP) and the National Formulary (composed of committees of physicians and pharmacists; these are now combined) or meet individual standards chosen by their manufacturers and stated on their labels. Adulteration was defined as the removal of valuable constituents, the substitution of ingredients so as to reduce quality, the addition of deleterious ingredients, and the use of spoiled animal and vegetable products. Misbranding was making false or misleading label statements regarding a food or a drug as well as the failure to provide required information in labeling.

Sulfanilamide, a drug used to treat streptococcal infections, had been used safely for years in tablet and powder form. For marketing to children, a company prepared the drug in liquid form after dissolving it in diethylene glycol and adding a raspberry flavor. Unfortunately, the 1906 law did not address testing products for toxicity. Diethylene glycol, a chemical normally used in antifreeze, is a deadly poison. Over 100 people died, most of whom were children. Another public outcry from this incident helped pass the Food, Drug, and Cosmetic Act in June 1938. This new law included cosmetics and medical devices and was instrumental in providing evidence of safety in pharmaceutical products prior to administration in humans. It also authorized factory inspections and gave the agency the ability to use injunctions to enforce the law.

Impurities in drug products have triggered numerous laws and amendments. In 1941 almost 300 people were killed or injured by ingestion of sulfathiazole tablets that were contaminated with phenobarbital. This caused the FDA to revise manu-

facturing and quality control requirements, leading to what would later be called Good Manufacturing Practices (GMPs) [5]. In 1945, batch certification by the FDA became a requirement for penicillin. The Penicillin Amendment requires FDA testing and certification of safety and effectiveness of all penicillin products; later amendments extended this requirement to all antibiotics. In 1983 FDA testing was found to be no longer necessary and was abolished [6].

Through the years, the U.S. Food and Drug Administration (FDA) and other international regulatory agencies, working alone or with industry representatives, have set regulations and guidelines for the handling of impurities in pharmaceuticals. As per the GMPs, it is essential that the purity of the pharmaceuticals is preserved from manufacturing through delivery to the consumer.

Most drugs are not administered as neat chemicals but are mixed with inactive ingredients to form the specific drug product. Many have impurities, due to manufacturing, leachables, or degradants. Impurities are chemicals contained in raw materials or formed during manufacture, storage, or use. Their properties may be different from the desired product with respect to activity, efficacy, and safety. These impurities may be either intentional or unintentional. Intentional impurities may be due to the raw material itself or excipients. Unintentional impurities may occur due to a few determinants, including interactions between the other materials in the drug product, leaching from containers, or degradation due to environmental conditions. Regulations and guidances exist for the safety assessment of impurities. Many focus on different levels of testing, based on the amount of impurity detected. There are many different methods and equipment available for the analysis of impurities. Safety or toxicology studies to test the impurity may be needed, and there are guidances for the selection and design of these studies.

## 2 TYPES AND SOURCES OF IMPURITIES

Impurities may come from a variety of sources. Manufacturing impurities, leachable impurities, and degradants are the most common.

### 2.1 Manufacturing Impurities

Manufacturing impurities can come from raw materials, active ingredients, excipients, process impurities and packaging impurities.

***Raw Material Impurities*** Contamination from raw material sources is important to control. The highest quality starting material is essential to obtain a high quality product. Raw materials with impurities may be cleaned up, and it is important to do so before they have a chance to interact with other materials in the formulation that can lead to additional impurities. Cleaning up a raw material such as a chemical is usually easier than cleaning up natural materials. The latter may contain complex sample matrices that include similar chemical compounds. In addition, the pharmacologically active compounds are usually present in small amounts and therefore may be difficult to separate out from the impurities. Thus, highly sophisticated analytical methodologies must be developed to ensure that the natural drug is uniform and pure.

*Compounds Related to Parent* Pharmaceuticals may contain impurities that are similar in structure to the parent material. These impurities may be more difficult to detect and isolate depending on how similar the structure is as compared to the parent material. A chromatogram may indicate a "shoulder" on the main peak or an eluting peak indicating such an impurity. An example is given in Fig. 1.

*Spurious Compounds* Although each known degradation pathway is considered, and every attempt is made to decrease or obliterate the impurities, pharmaceutical chemistry can still be fickle. The additional unpredicted peak found on the chromatogram or the inexplicable tan flecks in a powder material could be an extractable–leachable, a residual solvent, or an unforeseen degradant. These must be investigated and, depending on the amount of the impurity, must be identified. The active pharmaceutical ingredient (API) from a natural material is an example of how complicated raw materials and impurities can become.

***Active Ingredients*** Impurities arise from the manipulation of an active ingredient in a pharmaceutical. These impurities may be formed due to process or storage conditions.

***Excipients*** An excipient is any inactive ingredient that is intentionally added to therapeutic or diagnostic product, but is not intended to exert a therapeutic effect at the intended dosage. Excipients may act to improve product delivery (e.g., enhance absorption or control release of the drug substance) or allow the product to be formulated (e.g., tablet). Examples of excipients include fillers, extenders, diluents, wetting agents, solvents, emulsifiers, preservatives, flavors, absorption enhancers, sustained-release matrices, and coloring agents [7]. Within the context of this guidance, the term excipient applies to macromolecular substances, such as albumin, or substances such as amino acids and sugars that are used in drug and biological products. It does not, however, apply to process- or product-related impurities (e.g., degradation products, leachates, residual solvents) or extraneous contaminants. Excipients may interact with the active ingredient or container to form impurities or they may be formed due to process or storage conditions.

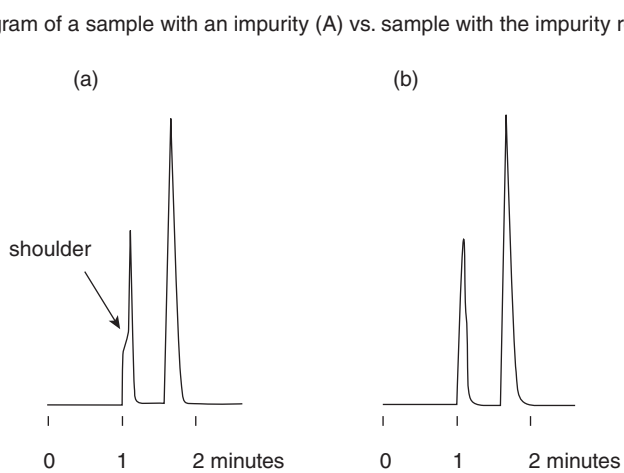


**FIGURE 1** Chromatogram showing eluting peak.

***Process Impurities*** Process impurities, primarily residual solvents, are organic volatile impurities that remain in pharmaceuticals, even after processing, as well as oxygen left in lyophilized vials. Both the United States Pharmacopoeia (USP), the officially recognized drug standard-setting compendium [8], and the International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH), Q3C [9], address commonly used chemical solvents, the latter of which has been adopted by the European and Japanese pharmacopoeias. Other process impurities may be due to by-products of synthetic reactions, chiral or stereoisomeric impurities, contamination, and residual process intermediates or reagents.

Biologics and biotechnology products have similar process impurity issues. The process needs to ensure that there are no residual cellular components in the biologic product. Biotechnology processing needs to avoid causing structural deformities to the protein. In all cases, the process must be scrutinized closely. Checking for impurities at various steps throughout the manufacturing phase may help to pinpoint where the impurities are produced.

***Packaging Impurities*** Packaging impurities can also be an issue in pharmaceuticals. These impurities can occur when chemicals from the packaging of the pharmaceutical are found in the product itself. These are discussed further in Section 2.3.

## 2.2 Degradants

Degradants are formed from the breakdown of the drug substance or product, possibly due to storage conditions, interaction of ingredients from the container, label or closing systems, or treatment of the material.

The ICH has published guidelines for the identification and qualification of degradant-related impurities. These guidelines take into account new and existing drugs and formulations. A forced degradation test is essential to assist in the determination of identifying any degradants and their degradation pathways. Forced degradation studies can be useful in the determination of the chemical and physical stability of crystal forms, the stereochemical stability of the drug substance alone and in the drug product, and for differentiating drug substance–related degradation products in formulations. Forced degradation tests are important to determine if extreme conditions will cause any impurities in the drug product. Forced degradation conditions can include storage under high temperatures, high humidity, extreme pH, freeze–thaw cycles, oxygen exposure, or UV light for specific time intervals. At certain timepoints, samples of the drug product are taken and analyzed for purity. One important consideration is that forced degradation experiments may only produce or release impurities into the drug product that would not generally be present during normal storage and handling conditions. A limitation of forced degradation testing is that a compound may not necessarily degrade under a given stress condition.

There are no specifics in the ICH guidelines for pH levels, temperature, conditions, or light intensity for forced degradation studies. However, for photodegradation, ICH Q1B [10] suggests that the light source should produce combined visible and ultraviolet (UV, 320–400 nm) outputs, and that exposure levels should be justified.

Degradation products if detected in significant amounts should be identified, tested for, and monitored against appropriately established acceptance criteria, according to ICH guidances [11, 12]. Investigation and identification of some specific degradation products may not be necessary if it has been proven that they are not produced under accelerated or long-term storage conditions [10, 13]. In addition, forced degradation may not be necessary when the routes of degradation and the suitability of the analytical procedures can be determined through use of data from stress testing of drug substance, reference materials for process impurities and degradants, data from accelerated and long-term studies on the drug substance, and data from accelerated and long-term studies on the drug product [14]. Additional corroborative information on the specificity of the analytical methods and on the degradation pathways of the drug substance may be available from literature sources.

Degradants will obviously vary for different drug products. Small chemicals may produce degradants that are similar to the parent compound. Biotechnological and biological products may encounter protein variants including truncated fragments, deamidated, oxidized, isomerized, aggregated forms, and mismatched disulfide links [15]. In these studies, there may also be difficulties in working with unstable degradation products.

As previously mentioned, many factors play a role in the formation of degradant products. Raw material handling is important to control early in the production phase. There should be no alteration of the material in any manner by chemicals or processing during manufacture. Changes in products can even lead to mutagenic degradants or degradants with increased immunogenicity, especially in biologics. The packaging used in various stages of the process may produce degradants. In addition, the choice of sterilization techniques may also play a role in producing degradation products, as different techniques may produce different degradants. Instability can lead to degradation over time. These factors would need to be taken into consideration prior to finalizing manufacturing, handling, and storage procedures.

### 2.3 Leachables and Extractables

Leachables and extractables are chemicals that may migrate from packaging or labeling, producing contamination in the drug product. These chemicals may be released throughout different stages of manufacturing, during storage or shipment of the drug components, during contact with processing materials in the plant, and even at the final stage while the drug product is in its primary storage container or in contact with labeling materials.

Extractables are compounds that can be extracted from the container, from the coatings of the container closure system, or from the label during forced degradation conditions including contact with an appropriate solvent or high temperature [16]. Forced degradation experiments may only release impurities into the pharmaceutical that would not generally be present during normal storage and handling conditions.

Leachables are compounds that migrate into the formulation from the label, container components, or coatings of the container and closure system as a result of direct contact with the formulation [17].These are impurities that have formed

during normal storage and handling conditions. Certain types of storage or treatment conditions may also produce leachables more readily, including sterilization procedures.

Sources of leachable and extractable contamination from packaging may include accelerators, adhesives, antioxidants, coatings, components of the container or closure system, excipients, plastic or rubber components, plasticizers, processing aids, and vulcanizing agents. Plasticizers are used to make plastics more flexible. Plasticizers such as bis(2-ethylhexyl)phthalate (BEHP) may be present throughout package manufacturing. Acrylonitrile is a leachable that may be present in manufactured products and is found in a number of pharmaceuticals [16]. Another potential source of leachable or extractable contamination is nitrosamines, which are found in products made of rubber. Rubber and rubber products are used in packaging, in closure components, and in devices, primarily in O-rings. Inhalers are a medical device that is prepared with various plastic, rubber, and stainless steel components. The FDA has a draft guidance for inhalers, which basically states that the levels of degradation products and impurities should be determined by means of stability indicating methods, and that any impurities or degradation products appearing at levels 0.10% or greater should be specified [18]. Many drug products are distributed or administered in packages made of plastic and rubber components; therefore, phthalates and nitrosamines could come into contact with the drug product.

## 3 REGULATIONS

Both the U.S. Food and Drug Administration (FDA) and the ICH offer a plethora of information on dealing with impurities in pharmaceuticals. While the FDA is the regulatory authority in the United States, the ICH is an organization that brings together regulatory authorities and pharmaceutical associations of Europe, Japan, and the United States, in addition to experts from other non-ICH countries to discuss scientific and technical aspects of product registration. The purpose of the ICH is to make recommendations to achieve greater harmonization in technical guidelines and requirements for product registration in order to reduce or obviate the need to duplicate the testing carried out during the research and development of new pharmaceuticals.

It is essential that proper analyses be conducted for impurities. Knowledge gained from these analyses can be used to guide formulation development and improve manufacturing and packaging processes.

There are consequences when impurities are undetected or observed at higher levels than reported. For regulatory submissions for marketing applications, current FDA and ICH guidance recommends inclusion of the results, including chromatograms of stressed samples, demonstration of the stability-indicating nature of the analytical procedures, and the degradation pathways of the drug substance in solution, solid state, and drug product. The structures of significant degradation products and the associated procedures for their isolation and/or characterization also are expected to be included in the application. For the latest updates on these guidances and for more details, check the FDA or ICH websites (www.fda.gov, www.ich.org). The following discussions are summaries of the guidances and include authentic excerpts but are not intended to be all inclusive.

### 3.1 International Conference on Harmonisation (ICH) Guidance

***Q3A*** The ICH *Guidance for Industry, Q3A Impurities in New Drug Substances*, Revision 1, was published February 2003 [11]. It is intended to "provide guidance for registration applications on the content and qualification of impurities in new drug substances produced by chemical syntheses and not previously registered in a region or member state." A new drug substance is not the final marketed product, but the active ingredient used in the marketed product. An active ingredient is intended to furnish pharmacological activity or other direct effect in the diagnosis, cure, mitigation, treatment, or prevention of disease or to affect the structure or any function of the human body. The active ingredient does not include intermediates used in the synthesis of such ingredient. The term includes those components that may undergo chemical change in the manufacture of the drug product and be present in the drug product in a modified form intended to furnish the specified activity or effect (21 CFR 210.3(b)(7) and 21 CFR 314.3(b)). Impurities in new drug substances are addressed from both a chemistry and a safety perspective.

The guidance is not intended to apply to new drug substances used during the clinical research stage of development. Nor does it cover the following types of drug substances: biological/biotechnological, peptide, oligonucleotide, radiopharmaceutical, fermentation products and associated semisynthetic products, herbal products, or crude products of animal or plant origin. The guidance does not cover extraneous contaminants that should not occur in new drug substances and are more appropriately addressed as good manufacturing practice (GMP) issues, polymorphic forms, and enantiomeric impurities.

The guidance divides impurities into three categories: organic impurities (process– and drug-related), inorganic impurities, and residual solvents. These are defined in the guidance, along with their respective rationale for the reporting and control of impurities. The guidance further describes the circumstances in which impurities need to be reported, identified, and qualified.

The rationale for the reporting and control, identification, and qualification of impurities is discussed in the guidance. Organic impurities need to be summarized based on the actual and potential impurities most likely to arise during the synthesis, purification, and storage of a new drug substance. This discussion can be limited to those impurities that might reasonably be expected based on knowledge of the chemical reactions and conditions involved.

In addition, laboratory studies conducted to detect impurities in the new drug substance need to be summarized for the regulatory application. According to the guidance, "this summary should include test results of batches manufactured during the development process and batches from the proposed commercial process, as well as the results of stress testing used to identify potential impurities arising during storage. The impurity profile of the drug substance batches intended for marketing should be compared with those used in development, and any differences discussed" [13].

Studies conducted to characterize the structure of impurities present in a new drug substance at a level greater than the identification threshold (Table 1) should be described and any impurity from any batch or degradation product from stability studies should be identified. If identification of an impurity or degradant is not feasible, a summary of the laboratory studies demonstrating the unsuccessful effort should be included in the application. If an impurity is pharmacologically or

**TABLE 1 Thresholds**

| Maximum Daily Dose[a] | Reporting Threshold[b,c] | Identification Threshold[c] | Qualification Threshold[c] |
|---|---|---|---|
| ≤2 g/day | 0.05% | 0.10% or 1.0 mg/day intake (whichever is lower) | 0.15% or 1.0 mg/day intake (whichever is lower) |
| >2 g/day | 0.03% | 0.05% | 0.05% |

[a]The amount of drug substance administered per day.
[b]Higher reporting thresholds should be scientifically justified.
[c]Lower thresholds can be appropriate if the impurity is unusually toxic.

toxicologically active, identification of the compound should be conducted even if the impurity level is below the identification threshold.

The guidance also states that "qualification is the process of acquiring and evaluating data that establishes the biological safety of an individual impurity or a given impurity profile at the level(s) specified. The applicant should provide a rationale for establishing impurity acceptance criteria that includes safety considerations. The level of any impurity that is present in a new drug substance that has been adequately tested in safety and/or clinical studies would be considered qualified. Impurities that are also significant metabolites present in animal and/or human studies are generally considered qualified. A level of a qualified impurity higher than that present in a new drug substance can also be justified based on an analysis of the actual amount of impurity administered in previous relevant safety studies. If data are unavailable to qualify the proposed acceptance criterion of an impurity, safety studies to obtain such data can be appropriate when the usual qualification thresholds are exceeded."

The "Decision Tree for Identification and Qualification," as revised for Q3B(R) (Fig. 2) describes considerations for the qualification of impurities when thresholds are exceeded. If the level of impurity cannot be decreased to below the threshold, or if adequate data is not available in the scientific literature to justify safety, then additional safety testing should be considered. The studies considered appropriate to qualify an impurity will depend on a number of factors, including the patient population, daily dose, and route and duration of administration. Toxicology studies are discussed briefly later in this chapter and in more detail in other chapters in this volume. Such studies can be conducted on the new drug substance containing the impurities to be controlled, although studies using isolated impurities can sometimes be appropriate.

ICH Q3A states that "safety assessment studies to qualify an impurity should compare the new drug substance containing a representative amount of the new impurity with previously qualified material. Safety assessment studies using a sample of the isolated impurity can also be considered." The latter is especially important to consider for genetic toxicology studies and the importance of testing the isolated impurity is discussed in more detail at the end of this chapter.

Therefore, according to the guidance, if the maximum daily dose of the drug is less than 2 grams per day and the impurity intake is more than 0.15% or 1.0 mg/day, the Qualification Threshold has been reached, meaning safety studies will need to be performed. Lower thresholds can be appropriate if the impurity is unusually toxic. In addition, the impurity will need to be reported and identified. These studies include general and genetic toxicology studies, and possibly other specific toxicology

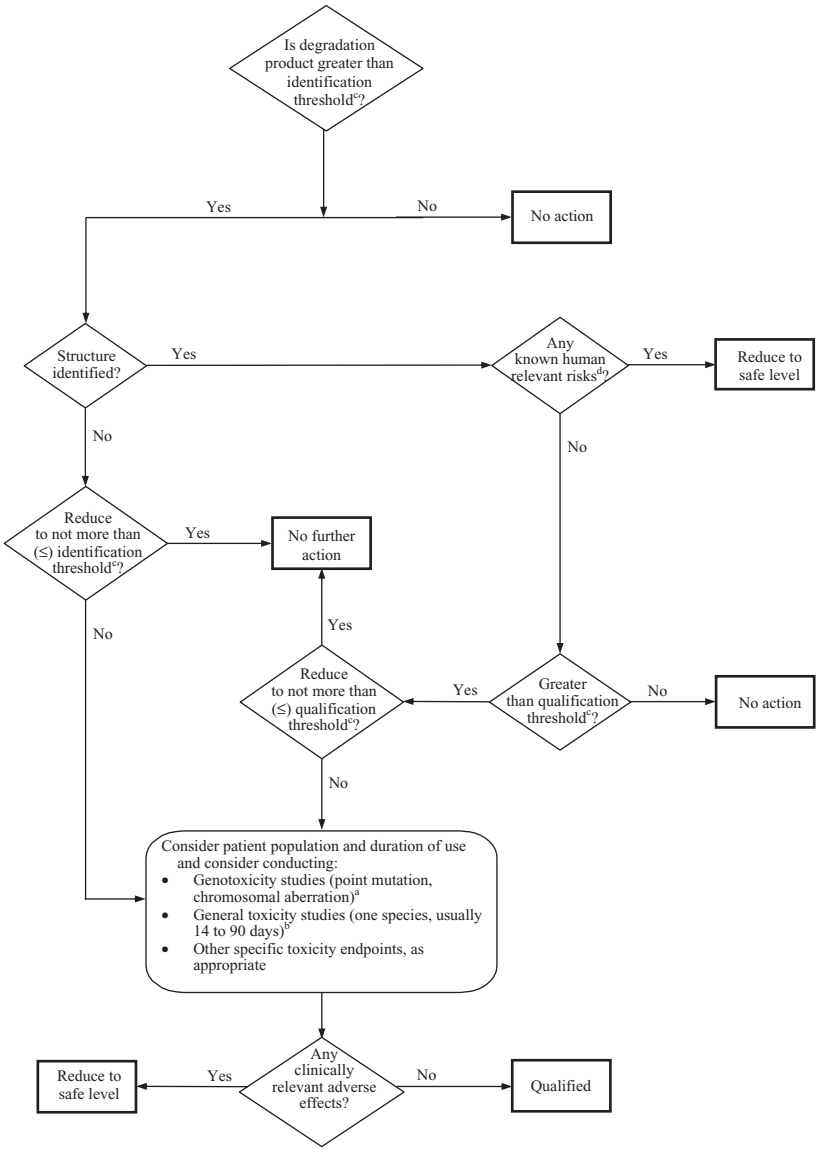


**Notes**

[a] If considered desirable, a minimum screen (e.g., genotoxic potential) should be conducted. A study to detect point mutations and one to detect chromosomal aberrations, both *in vitro*, are considered an appropriate minimum screen.

[b] If general toxicity studies are desirable, one or more studies should be designed to allow comparison of unqualified to qualified material. The study duration should be based on available relevant information and performed in the species most likely to maximize the potential to detect the toxicity of a degradation product. On a case-by-case basis, single-dose studies can be appropriate, especially for single-dose drugs. In general, a minimum duration of 14 days and a maximum duration of 90 days would be considered appropriate.

[c] Lower thresholds can be appropriate if the degradation product is unusually toxic.

[d] For example, do known safety data for this degradation product or its structural class preclude human exposure at the concentration present?

**FIGURE 2** The "Decision Tree for Identification and Qualification," as revised for Q3B(R).

endpoints, as appropriate. It also recommended to have a discussion of specific toxicity testing with the appropriate FDA division.

The genetic toxicology studies can include a minimum screen (a study to detect point mutations and one to detect chromosome aberrations, both *in vitro*). The general toxicology studies should include one or more studies designed to allow comparison of unqualified to qualified material. The study duration should be based on available relevant information and performed in the species most likely to maximize the potential to detect the toxicity of an impurity. On a case-by-case basis, single-dose studies can be appropriate, especially for single-dose drugs. In general, a minimum duration of 14 days and a maximum duration of 90 days would be considered appropriate.

Inorganic impurities are normally detected and quantified using pharmacopoeial or other appropriate procedures. The need for inclusion or exclusion of inorganic impurities in a new drug substance specification should be discussed. Acceptance criteria should be based on pharmacopoeia standards or known safety data. The control of residues of the solvents used in the manufacturing process for a new drug substance should be discussed and presented according to ICH Q3C [9].

A registration application should include documented evidence that the analytical procedures are validated and suitable for the detection and quantification of impurities [19, 20]. Organic impurity levels can be measured by a variety of techniques, including those that compare an analytical response for an impurity to that of an appropriate reference standard or to the response of the new drug substance itself. Differences in the analytical procedures used during development and those proposed for the commercial product should be discussed in the registration application. Analytical results should be provided in an application for all batches of a new drug substance used for clinical, safety, and stability testing, as well as for batches representative of the proposed commercial process. The application should also contain a table that links the specific new drug substance batch to each safety study and each clinical study in which the new drug substance has been used. Any impurity at a level greater than the reporting threshold (Table 1) and total impurities observed in these batches of the new drug substance should be reported with the analytical procedures indicated. Table 2 is an illustration of reporting impurity results for identification and qualification in an application.

The guidance also states that when analytical procedures change, results provided in the application should be linked to the procedure used, with appropriate valida-

**TABLE 2 Reporting Impurity Results in an Application**

| | | Action | |
|---|---|---|---|
| Raw Result (%) | Reported Result (%) | Identification (Threshold 0.10%) | Qualification (Threshold 0.15%) |
| 0.066 | 0.07 | None | None |
| 0.0963 | 0.10 | None | None |
| 0.12 | 0.12[a] | Yes | None[a] |
| 0.1649 | 0.16[a] | Yes | Yes[a] |

[a]After identification, if the response factor is determined to differ significantly from the original assumptions, it may be appropriate to remeasure the actual amount of the impurity present and reevaluate against the qualification threshold.

tion information provided, including representative chromatograms of representative batches. The applicant should ensure that complete impurity profiles (e.g., chromatograms) of individual batches are available, if requested.

The ICH Q3A guidance also states that the specification for a new drug substance should include a list of impurities. Individual impurities with specific acceptance criteria included in the specification for a new drug substance are referred to as specified impurities. Specified impurities can be identified or unidentified. A rationale for the inclusion or exclusion of impurities in a specification should be presented.

"Acceptance criteria should be set no higher than the level that can be justified by safety data and should be consistent with the level achievable by the manufacturing process and the analytical capability. Where there is no safety concern, impurity acceptance criteria should be based on data generated on batches of a new drug substance manufactured by the proposed commercial process, allowing sufficient latitude to deal with normal manufacturing and analytical variation and the stability characteristics of the new drug substance. Although normal manufacturing variations are expected, significant variation in batch-to-batch impurity levels can indicate that the manufacturing process of the new drug substance is not adequately controlled and validated" [21].

***Q3B*** ICH *Guidance for Industry, Q3B(R) Impurities in New Drug Products*, Revision 1, was published November 2003 [12]. It is intended to provide guidance for registration applications on the content and qualification of impurities in new drug products produced from chemically synthesized new drug substances not previously registered in a region or member state. A new drug product is a finished dosage form, for example, a tablet, capsule, or solution, that contains a drug substance, generally, but not necessarily, in association with one or more other ingredients [22]. The Q3B(R) complements the ICH guidance *Q3A guidance Impurities in New Drug Substances*, which should be consulted for basic principles along with ICH *Q3C Impurities: Residual Solvents*, when appropriate.

This guidance addresses only those impurities in new drug products classified as degradation products of the drug substance, or reaction products of the drug substance with an excipient and/or immediate container closure system (collectively referred to as *degradation products*). Generally, impurities present in a new drug substance need not be monitored or specified in new drug product unless they are also degradation products [21]. This guidance does not address impurities arising from excipients present in a new drug product or extracted or leached from the container closure system. This guidance also does not apply to new drug products used during the clinical research stages of development. It also does not cover the same types of products as in 3QA(R): biological/biotechnological, peptides, oligonucleotides, radiopharmaceuticals, fermentation products and associated semisynthetic products, herbal products, and crude products of animal or plant origin. Also excluded from this guidance are extraneous contaminants that should not occur in new drug products and are more appropriately addressed as GMP issues, polymorphic forms, and enantiomeric impurities.

The rationale for the reporting and identification of degradation products is similar to that in Q3A; however, identification thresholds are different (Table 3). The impurity profiles of the batches representative of the proposed commercial process should be compared with the profiles of batches used in development, and any differences should be discussed in the submission.

As expected, qualification of an impurity has similar concerns as Q3A. The main differences are the reporting, identification, and qualification thresholds (Table 3). The thresholds are basically higher than they were in Q3A; however, there are more categories for dosages. If the qualification thresholds given in Table 3 are exceeded and data are unavailable to qualify the proposed acceptance criterion of a degradation product, additional studies to obtain such data may be appropriate (Fig. 2).

The analytical procedures and reporting in 3QB(R) are similar to those in Q3A. Differences between the analytical procedures used during development and those proposed for the commercial product should also be discussed in the registration application. Reported results should be rounded using conventional rules (see Table 4). In Table 4, the doses have been divided into more categories than in 3QA, and, therefore, the action thresholds have changed.

The specification for a new drug product should include a list of degradation products expected to occur during manufacture of the commercial product and under recommended storage conditions. This section, again, is similar to Q3A, with the exception of the threshold levels being different. In addition, a final listing of impurities in the specifications should include the same items as in Q3A.

***Q3C*** ICH *Guidance for Industry, Q3C Impurities: Residual Solvents* was published December 1997 [9]. It is intended to provide guidance for recommending acceptable amounts for residual solvents in pharmaceuticals for the safety of the patient. The guidance recommends use of less toxic solvents and describes levels considered to be toxicologically acceptable for some residual solvents. A complete list of the solvents included in this guidance is provided in a companion document entitled

**TABLE 3 Thresholds for Degradation Products in New Drug Products**

| Maximum Daily Dose[a] | Threshold[b,c] |
|---|---|
| *Reporting Thresholds* | |
| ≤1 g | 0.1% |
| >1 g | 0.05% |
| *Identification Thresholds* | |
| <1 mg | 1.0% or 5 μg TDI, whichever is lower |
| 1–10 mg | 0.5% or 20 μg TDI, whichever is lower |
| >10 mg–2 g | 0.2% or 2 mg TDI, whichever is lower |
| >2 g | 0.10% |
| *Qualification Thresholds* | |
| <10 mg | 1.0% or 50 μg TDI, whichever is lower |
| 10–100 mg | 0.5% or 200 μg TDI, whichever is lower |
| >100 mg–2 g | 0.2% or 3 mg TDI, whichever is lower |
| >2 g | 0.15% |

[a]The amount of drug substance administered per day.
[b]Thresholds for degradation products are expressed either as a percentage of the drug substance or as total daily intake (TDI) of the degradation product. Lower thresholds can be appropriate if the degradation product is unusually toxic.
[c]Higher thresholds should be scientifically justified.

**TABLE 4 Illustration of Reporting Degradation Product Results for Identification and Qualification in an Application**

| Raw Result (%) | Reported Result (%) (Reporting Threshold = 0.1%) | Total Daily Intake (TDI) of the Degradation Product (rounded result in μg) | Action: Identification Threshold 0.2% | Action: Qualification Threshold 200 μg TDI (equivalent to 0.4%) |
|---|---|---|---|---|
| *50 mg Maximum Daily Dose* | | | | |
| 0.04 | Not reported | 20 | None | None |
| 0.2143 | 0.2 | 100 | None | None |
| 0.349 | 0.3[a] | 150 | Yes | None[a] |
| 0.550 | 0.6[a] | 300 | Yes | Yes[a] |

| Raw Result (%) | Reported Result (%) (Reporting Threshold = 0.05%) | Total Daily Intake (TDI) of the Degradation Product (rounded result in mg) | Action: Identification Threshold 2 mg TDI (equivalent to 0.11%) | Action: Qualification Threshold 3 mg TDI (equivalent to 0.16%) |
|---|---|---|---|---|
| *1.9 g Maximum Daily Dose* | | | | |
| 0.049 | Not reported | 1 | None | None |
| 0.079 | 0.08 | 2 | None | None |
| 0.183 | 0.18[a] | 3 | Yes | None[a,b] |
| 0.192 | 0.19[a] | 4 | Yes | Yes[a] |

[a]After identification, if the response factor is determined to differ significantly from the original assumptions, it can be appropriate to remeasure the actual amount of the degradation product present and reevaluate against the qualification threshold (see Attachment 1).
[b]Although the reported result of 0.18% exceeds the calculated threshold value of 0.16%, in this case the action is acceptable since the TDI (when rounded) does not exceed 3 mg. Chromatograms with peaks labeled (or equivalent data if other analytical procedures are used) from representative batches, including chromatograms from analytical procedure validation studies and from long-term and accelerated stability studies, should be provided. The applicant should ensure that complete degradation product profiles (e.g., chromatograms) of individual batches are available, if requested.

*Q3C—Tables and List* [23]. These tables are not included in this chapter but may be found in the ICH or FDA website. The list is not exhaustive, and other solvents may be used and later added to the list.

Residual solvents in pharmaceuticals are defined here as organic volatile chemicals that are used or produced in the manufacture of drug substances or excipients, or in the preparation of drug products. The solvents are not completely removed by practical manufacturing techniques. Appropriate selection of the solvent for the synthesis of drug substance may enhance the yield, or determine characteristics such as crystal form, purity, and solubility. Therefore, the solvent may sometimes be a critical parameter in the synthetic process. This guidance does not address solvents deliberately used as excipients nor does it address solvates. However, the content of solvents in such products should be evaluated and justified.

As there are no therapeutic benefits from residual solvents, all residual solvents should be removed to the extent possible to meet product specifications, good

manufacturing practices, or other quality-based requirements. Drug products should contain no higher levels of residual solvents than can be supported by safety data. Some solvents that are known to cause unacceptable toxicities (carcinogens), such as benzene and carbon tetrachloride (Class 1, see Table 1 in Ref. 23), should be avoided in the production of drug substances, excipients, or drug products unless their use can be strongly justified in a risk–benefit assessment. Some solvents associated with less severe toxicity (nongenotoxic animal carcinogens or possible causative agents of other irreversible toxicity such as neurotoxicity or teratogenicity), such as acetonitrile and chlorobenzene (Class 2, see Table 2 in Ref. 23), should be limited in order to protect patients from potential adverse effects. Ideally, less toxic solvents, such as acetic acid and acetone (Class 3, see Table 3 in Ref. 23), should be used where practical.

Residual solvents in drug substances, excipients, and drug products are within the scope of this guidance. Therefore, testing should be performed for residual solvents when production or purification processes are known to result in the presence of such solvents. It is only necessary to test for solvents that are used or produced in the manufacture or purification of drug substances, excipients, or drug products. Although manufacturers may choose to test the drug product, a cumulative method may be used to calculate the residual solvent levels in the drug product from the levels in the ingredients used to produce the drug product. If the calculation results in a level equal to or below that recommended in this guidance, no testing of the drug product for residual solvents need be considered. If, however, the calculated level is above the recommended level, the drug product should be tested to ascertain whether the formulation process has reduced the relevant solvent level to within the acceptable amount. Drug product should also be tested if a solvent is used during its manufacture.

This guidance does not apply to potential new drug substances, excipients, or drug products used during the clinical research stages of development, nor does it apply to existing marketed drug products.

The guidance applies to all dosage forms and routes of administration. Higher levels of residual solvents may be acceptable in certain cases such as short-term (30 days or less) or topical application. Justification for these levels should be made on a case-by-case basis and discussed with the appropriate FDA division.

The methods for establishing exposure limits in compounds are discussed, as are reporting levels. Some analytical procedures for the determination of exposure typically include chromatographic techniques such as gas chromatography for solvents. Any harmonized procedures for determining levels of residual solvents as described in the pharmacopoeias should be used, if feasible. If only Class 3 solvents are present, a nonspecific method such as loss on drying may be used.

The limits of residual solvents may include a value for the Permitted Daily Exposure (PDE), which is the maximum acceptable intake per day of residual solvent in pharmaceutical products. These limits vary depending on the class.

For solvents to be avoided, solvents in Class 1 should not be employed in the manufacture of drug substances, excipients, and drug products because of their unacceptable toxicity or their deleterious environmental effect. However, if their use is unavoidable in order to produce a drug product with a significant therapeutic advance, then their levels should be restricted as shown in Table 1 of Ref. 23 of the companion document, unless otherwise justified. The solvent 1,1,1-trichloroethane

is included (Table 1 in Ref. 23) because it is an environmental hazard. The stated limit of 1500 ppm is based on a review of the safety data.

For solvents to be limited, solvents in Class 2 (see Table 2 in Ref. 23) should be limited in pharmaceutical products because of their inherent toxicity. PDEs are given to the nearest 0.1 mg/day, and concentrations are given to the nearest 10 ppm. The stated values do not reflect the necessary analytical precision of determination. Precision should be determined as part of the validation of the method.

For solvents with low toxic potential, solvents in Class 3 (see Table 3 in Ref. 23) may be regarded as less toxic and of lower risk to human health. Class 3 includes no solvent known as a human health hazard at levels normally accepted in pharmaceuticals. However, there are no long-term toxicity or carcinogenicity studies for many of the solvents in Class 3. Available data indicate that they are less toxic in acute or short-term studies and negative in genotoxicity studies. It is considered that amounts of these residual solvents of 50 mg/day or less (corresponding to 5000 ppm or 0.5% under Option 1) would be acceptable without justification. Higher amounts may also be acceptable provided they are realistic in relation to manufacturing capability and GMPs.

For solvents for which no adequate toxicological data were found, the solvents listed (see Table 4 in Ref. 23) may also be of interest to manufacturers of excipients, drug substances, or drug products. However, no adequate toxicological data on which to base a PDE were found. Manufacturers should supply justification for residual levels of these solvents in pharmaceutical products.

### 3.2 U.S. FDA (CDER)

U.S. FDA (CDER) *Guidance for Industry, NDAs: Impurities in Drug Substances* was published February 2000 [24]. The guidance refers applicants to ICH *Q3A Impurities in New Drug Substances* when seeking guidance on identification, qualification, and reporting of impurities in drug substances that are not considered new drug substances. Q3A was developed by the ICH to provide guidance on the information that should be provided in a new drug application (NDA) in support of impurities in new drug substances that are produced by chemical syntheses. The FDA believes that the guidance provided there on identification, qualification, and reporting of impurities should also be considered when evaluating impurities in drug substances produced by chemical syntheses that are not considered new drug substances. ICH Q3A defines a new drug substance (also referred to as a new molecular entity or new chemical entity) as a designated therapeutic moiety that has not been previously registered in a region or member state. The definition also states that a new drug substance may be a complex, a simple ester, or a salt of a previously approved drug substance.

This recommendation applies to applicants planning to submit NDAs and supplements for changes in drug substance synthesis or process. It also applies to holders of Type II Drug Master Files (DMFs) that support such applications. Applicants should note that this recommendation would not apply to DMFs cited in an NDA or supplement if the DMF information has been deemed acceptable for that dosage form, route of administration, and daily intake prior to the publication of the final version of this guidance. Examples of NDAs affected by the recommendation include those submitted for new dosage forms of already approved drug products, or drug products containing two or more active moieties that are individually used

in already approved drug products but have not previously been approved or marketed together in a drug product.

This guidance does not apply to applications for biological, biotechnological, peptide, oligonucleotide, radiopharmaceutical, fermentation, and semisynthetic products derived from herbal products or crude products of animal or plant origin, nor does it apply to Abbreviated New Drug Applications (ANDAs). Guidance on drug substances used for ANDA products is available [25].

In February 2005, the FDA published the *Guidance for Industry, ANDAs: Impurities in Drug Substances* [25]. This guidance provides revised recommendations on what chemistry, manufacturing, and controls (CMC) information to include regarding the reporting, identification, and qualification of impurities in drug substances produced by chemical synthesis when submitting original Abbreviated New Drug Applications (ANDAs), Drug Master Files (DMFs) including Type II DMFs, and ANDA supplements for changes in drug substance synthesis or process. The guidance also provides recommendations for establishing acceptance criteria for degradation products (specifically, degradation products of the active ingredient or reaction products of the active ingredient with an excipient(s) and/or immediate container/closure system) in generic drug products. The ANDA document refers to the ICH toxicology test scheme of up to 90 days in two species (rodent and nonrodent). The guidance does not apply to an ANDA or ANDA supplement that has been reviewed prior to the publication of the final guidance.

The submitter is referred to the ICH Q3B(R) guideline [12] that was developed to provide guidance on impurities in drug products for New Drug Applications (NDAs). However, the FDA believes that many of the recommendations provided on impurities in drug products also apply to ANDAs, including:

- Section I, Introduction
- Section II, Rationale for the Reporting and Control of Degradation Products
- Section III, Analytical Procedures
- Section IV, Reporting Degradation Products, Content of Batches
- Attachment 1, Thresholds for Degradation Products

The FDA recommends that the specification for a drug product include a list of degradation products and include in the submission a rationale for the inclusion or exclusion of degradation products in the drug product specification. It is important that the rationale include a discussion of the degradation profiles observed in stability studies and in the degradation profiles observed in the batch(es) under consideration together with a consideration of the degradation profile of the batch(es) manufactured by the proposed commercial process.

The FDA recommends the inclusion of general acceptance criteria of not more than the identification threshold (Fig. 2) for any unspecified degradation product and acceptance criteria for total degradation products. In addition, the drug product specification needs to include, where applicable, the following types of degradation products: each specified identified degradation product, each specified unidentified degradation product, any unspecified degradation product with an acceptance criterion of not more than the figure in the identification threshold in Q3B(R), and total degradation products.

The acceptance criterion for degradants should be set as in Q3B(R). In establishing degradation product acceptance criteria, the first critical consideration is whether

a degradation product is specified in the United States Pharmacopoeia (USP). If there is a monograph in the USP that includes a limit for a specified identified degradation product, the FDA recommends that the acceptance criterion be set no higher than the official compendial limit. If the level of the degradation product is above the level specified in the USP, the FDA recommends qualification. Then, if appropriate qualification has been achieved, an applicant may wish to petition the USP for revision of the degradation product's acceptance criterion.

If the acceptance criterion for a specified degradation product does not exist in the USP and this degradation product can be qualified by comparison to an FDA-approved human drug product, the acceptance criterion should be consistent with the level observed in the approved human drug product. In other circumstances, the acceptance criterion may need to be set lower than the qualified level to ensure drug product quality. For example, if the level of the metabolite impurity is too high, other quality attributes, like potency, could be seriously affected. In this case, the FDA recommends that the degradation product acceptance criterion be set lower than the qualified level.

Recommended qualification thresholds for degradation products based on the maximum daily dose of the drug are provided in ICH Q3B(R). The decision tree (Fig. 2) describes considerations for the qualification of degradation products when the usual qualification threshold recommended in ICH Q3B(R) is exceeded.

## 4 TECHNIQUES

### 4.1 General

For safety and efficacy, it is important to know the purity of the drug substances and drug products. Useful information can often be obtained by carrying out a simple thin layer chromatographic analysis, and the ultraviolet spectrum can also be valuable. It is also possible to measure the absorbance of a solution of the drug and compare the result with tabulated specific absorbance values (the absorbance of a 1% (w/v) solution in a cell of 1 cm path length). For example, the specific absorbances for the drug colchicine in ethanol are 730 and 350 at 243 nm and 425 nm, respectively. Thus, a 10 mg/L solution in ethanol should give absorbance readings of 0.73 and 0.35 at 243 nm and 425 nm, respectively, in a cell of 1 cm path length. However, this procedure, along with other methods, does not take into account the presence of impurities with similar relative molecular masses and specific absorbance values or other impurities. Biologics have their own unique analytical challenges and may need both chemical assays and bioactivity assays (microbial, cellular, metabolic, enzymatic, gene expression). This section discusses chemical assays.

There are many different methods for the detection and quantitation of impurities in materials. Some of these are discussed next. Data on analytical methodology for specific chemicals is readily available on the Web. OSHA has an index of sampling and analytical methods for hundreds of chemicals at http://www.osha.gov/dts/sltc/methods/toc.html.

The National Institute of Standards and Technology (NIST) Chemistry WebBook provides access to data compiled and distributed by NIST under the Standard Reference Data Program and can be found at http://webbook.nist.gov/. The WebBook

contains thermochemical data for over 7000 organic and small inorganic compounds, reaction thermochemistry data for over 8000 reactions, IR spectra for over 16,000 compounds, mass spectra for over 15,000 compounds, UV/Vis spectra for over 1600 compounds, gas chromatography data for over 27,000 compounds, electronic and vibrational spectra for over 5000 compounds, constants of diatomic molecules (spectroscopic data) for over 600 compounds, ion energetics data for over 16,000 compounds, and thermophysical property data for 74 fluids.

***Method Validation*** Method validation is an analytical procedure used to demonstrate that the analytical procedure in question is suitable for its intended purpose and is reproducible. A GMP requirement in 21 CFR 211.165(e) states that the accuracy, sensitivity, specificity, and reproducibility of test methods employed shall be established and documented [26]. The U.S. FDA (CDER) has a draft guidance that discusses analytical procedures and methods validation, chemistry, manufacturing, and controls documentation [27].

This draft guidance provides recommendations to applicants on submitting data for the validation study, including analytical procedures, validation data, and samples to support the documentation of the identity, strength, quality, purity, and potency of drug substances and drug products. The recommendations apply to drug substances and drug products covered in New Drug Applications (NDAs), ANDAs, Biologics License Applications (BLAs), Product License Applications (PLAs), and supplements to these applications. The principles also apply to drug substances and drug products covered in Type II Drug Master Files (DMFs). If a different approach is chosen, the applicant is encouraged to discuss the matter in advance with the center with product jurisdiction to prevent the expenditure of resources on preparing a submission that may later be determined to be unacceptable. Although this guidance does not specifically address the submission of analytical procedures and validation data for raw materials, intermediates, excipients, container closure components, and other materials used in the production of drug substances and drug products, validated analytical procedures should be used to analyze these materials.

The principles of methods validation presented apply to all types of analytical procedures. However, the specific recommendations in this guidance may not be applicable to certain unique analytical procedures for products such as biological, biotechnological, botanical, or radiopharmaceutical drugs. For example, many bioassays are based on animal challenge models, immunogenicity assessments, or other immunoassays that have unique features that should be considered when submitting analytical procedure and methods validation information. For questions on appropriate validation approaches for analytical procedures or submission of information not addressed in this guidance, applicants should consult with the appropriate chemistry review staff at the FDA.

FDA investigators inspect the analytical laboratory testing sites to ensure that the analytical procedures used for release and stability testing comply with GMPs (21 CFR part 211) [26] or good laboratory practices (GLPs; 21 CFR part 58) [28], as appropriate. These assure that proper guidelines and documentation are followed throughout the studies. All analytical procedures are of equal importance from a validation perspective. In general, validated analytical procedures should be used, irrespective of whether they are for in-process, release, acceptance, or stability

testing. Each quantitative analytical procedure should be designed to minimize assay variation. Reference standards must be used. A list of information that should typically be included in a description of an analytical procedure is included in this guidance. Impurities and degradants are addressed in the method validation information and the following should be reported:

- Data to demonstrate the stability of all analytical sample preparations through the time required to complete the analysis.
- Representative calculations using submitted raw data, to show how the impurities in drug substance are calculated.
- Information from stress studies.
- Impurities labeled with their names and location identifiers (e.g., RRT for chromatographic data) for the impurity analytical procedure.
- For drug substances:
  - A discussion of the possible formation and control of polymorphic and enantiomeric substances.
  - Identification and characterization of each organic impurity, as appropriate. This information may not be needed for all products (e.g., botanicals). Other impurities (e.g., inorganics, residual solvents) should be addressed and quantitated.
  - Recommendations on submitting information on impurities is provided in various FDA guidances such as the ICH guidance *Q3A Impurities in New Drug Substances* (January 1996).
  - A list of known impurities, with structure if available, including process impurities, degradants, and possible isomers.
- For drug products:
  - A degradation pathway for the drug substance in the dosage form, where possible.
  - Data demonstrating recovery from the sample matrix as illustrated by the accuracy studies.
  - Data demonstrating that neither the freshly prepared nor the degraded placebo interferes with the quantitation of the active ingredient.

The draft guideline lists examples of common problems that can delay successful validation.

- Failure to provide a sample of a critical impurity, degradation product, internal standard, or novel reagent.
- Failure to submit well-characterized reference standards for noncompendial drugs.
- Failure to provide sufficient detail or use of unacceptable analytical procedures. For example, use of arbitrary arithmetic corrections, failure to provide system suitability tests, and differing content uniformity and assay analytical procedures without showing equivalence factors for defining corrections as required by the current USP Chapter—Uniformity of Dosage Units.
- Failure to submit complete or legible data. For example, failure to label instrument output to indicate sample identity and failure to label the axes.

- Inappropriate shipping procedures. For example, failure to properly label samples, failure to package samples in accordance with product storage conditions, and inadequate shipping forms (e.g., missing customs form for samples from outside the United States).
- Failure to describe proper storage conditions on shipping containers.

Prior to the start of the validation study, conducting a feasibility study will help uncover issues early on so that the main study flows with ease. Feasibility studies address the setup method and system suitability, determine limits of detection (LOD) and quantitation (LOQ), perform forced degradation studies, check placebo for interferences, test related compounds, and check the label and specification claim.

The protocol needs to address specificity, linearity, accuracy, range, LOD/LOQ, precision, and robustness [29].

Specificity is the degree to which the measured response is due to the analyte of interest and not to other substances expected to be present in the sample matrix. For example, degradation products may be formed when the drug substance is exposed to environmental and/or chemical conditions, such as acid hydrolysis (HCl), base hydrolysis (NaOH), oxidation (peroxide), thermal degradation (heat), or photolysis (irradiation).

Linearity is the ability to obtain results that are directly proportional. Standards may be prepared at concentration levels ranging from 50% to 150% of the theoretical range of the analyte in a sample solution concentration prepared according to the method. The upper and lower limits of the linear range need to be reported. A graph of the standard curve and results of linear regression should also be determined. Low-level linearity for impurities needs to be determined, but sometimes impurity reference standards are unavailable. Since impurities are typically determined based on the active ingredient peak, demonstration of linearity at a low level is critical. A low-level linearity curve should include concentrations of 0.05%, 0.1%, 0.5%, 1.0%, 1.5%, and 2.0% of the active compound.

Accuracy is the measure of the total error of a method, including both systematic and random errors. For impurities assays, the accuracy should be assessed on samples (drug substance/drug product) spiked with known amounts of impurities. If impurities or degradation products are not available, it is acceptable to compare results obtained by a second, well-characterized method.

Range is the interval between the upper and lower concentration of analyte for which a suitable level of precision, accuracy, and linearity has been demonstrated.

***Limit of Detection (LOD)*** The LOD is the lowest concentration of analyte that can be detected, but not necessarily quantitated, by the analytical method. It is usually expressed as concentration of analyte generating an instrument response equivalent to three times the noise. The LOQ is the lowest concentration of analyte that can be determined with acceptable accuracy and precision by the analytical method. It is usually expressed as concentration of analyte generating an instrument response equivalent to ten times the noise.

Precision is the measure of how close the data values are to each other for a number of determinations under the same analytical conditions. ICH has defined precision to contain three components: repeatability (six replicate measurements of

a single sample preparation), intermediate precision (ruggedness; second analyst repeats method precision on different instrument, column, day, etc.), and reproducibility (method precision, duplicate measurements of six sample preparations). Precision is expressed as the percent relative standard deviation (%RSD).

Robustness is similar to ruggedness. It is a measure of how the results are affected by variations of different factors internal to the method, as in the mobile phase pH, mobile phase composition, temperature, flow rate, injector/detector temperatures, and oven ramp rates. According to ICH Q2A, the robustness of an analytical procedure is a measure of its capacity to remain unaffected by small but deliberate variations in method parameters and provides an indication of its reliability during normal usage.

***Metabolism of Impurities and Degradants*** Although they are not true impurities in the drug substance or the drug product, metabolites might be produced *in vivo*. The drug substance or product is metabolized in a manner so that the animal becomes exposed to the new metabolite. These metabolites may be detected from bioanalysis samples (whole blood, plasma, serum, urine). Due to exposure *in vivo*, ICH Q3A states that impurities that are also significant metabolites present in animal and/or human studies are generally considered qualified.

### 4.2 Ultraviolet and Visible (UV/Vis) Spectroscopy

A number of the quantitative methods use ultraviolet (UV) (200–400 nm) or visible (400–800 nm) spectrophotometry. Quantitative analysis of solutions is commonly performed by ultraviolet spectroscopy [30]. The major problem encountered with this technique is interference, and some form of sample purification is needed, such as solvent extraction. The spectrophotometer may be of the single-beam or double-beam type. With a single-beam instrument, light passes from the source through a monochromator and then via a sample cell to the detector. With double-beam instruments, light from the monochromator passes through a beam-splitting device and then via separate sample and reference cells to the detector. Double-beam instruments with automated wavelength scanning and a variety of other features are also available.

### 4.3 Atomic Absorption Spectroscopy

Atoms can absorb energy of a specific wavelength that corresponds to the wavelength they would emit when eroded by a high energy source. In atomic absorption, the energy needed for the atomic emission is supplied by a source other than a flame. The flame is actually the cuvette that holds the sample. The decrease in the emission intensity of the source is proportional to the concentration of atoms in the sample. Atomic absorption spectroscopy is important for the analysis of trace metals in biological fluids [31].

### 4.4 Thin Layer Chromatography (TLC)

Thin layer chromatography (TLC) is a technique for determining the composition of a mixture. It involves the movement by capillary action of a liquid phase (usually

an organic solvent) through a thin, uniform layer of stationary phase (usually silica gel, $SiO_2$) held on a rigid or semirigid support, normally a glass, aluminum, or plastic sheet or plate. Compounds are separated by partition between the mobile and stationary phases. It may be used to establish the extent of a reaction, the purity of a compound, or the presence or absence of materials in fractions from column chromatography. TLC is relatively inexpensive and simple to perform and can be a powerful qualitative technique when used together with some form of sample pretreatment, such as solvent extraction. However, some separations can be difficult to reproduce. The interpretation of results can also be very difficult, especially if a number of drugs or metabolites are present.

### 4.5 High Pressure Liquid Chromatography (HPLC)

In general, chromatography is used to separate mixtures of chemicals into individual components. Once separated, the components can be individually evaluated. With chromatography, separation occurs when the sample mixture is introduced (injected) into a mobile phase. In liquid chromatography (LC), the mobile phase is a solvent. The mobile phase transports the mixture through a stationary phase. The stationary phase is a chemical that can selectively draw components from a sample mixture. The stationary phase is usually contained in a tube (column). Columns can be glass or stainless steel of various dimensions.

The mixture of compounds in the mobile phase will interact with the stationary phase. Individual compounds in the mixture tend to interact at various rates. Those that interact the fastest will exit (elute from) the column first. Those that interact slowest will exit the column last. By changing characteristics of the mobile phase and the stationary phase, different mixtures of chemicals can be separated.

HPLC is commonly used for the separation, identification, purification, and quantification of chemical compounds (Fig. 3). Chemical separations can be accomplished since certain compounds have different migration rates given a particular column and mobile phase. Therefore, separation of compounds from each other can occur; the extent or degree of separation is mostly determined by the choice of the stationary phase and the mobile phase. The chromatographer may choose the conditions, such as the proper mobile phase, to allow adequate separation or purification in order to collect or extract the desired compound as it elutes from the stationary phase. Each compound should have a characteristic peak under certain chromatographic conditions. The migration of the compounds and contaminants through the column need to differ enough so that the pure desired compound can be collected or extracted without bringing on any other undesired compound.

Preparative HPLC is the process of isolation and purification of compounds. Analytical HPLC determines information about the sample compound, including identification, quantification, and resolution.

Identification of compounds by HPLC is a critical part of any HPLC assay and is accomplished by researching the literature and by trial and error. After a few steps, a separation assay must be developed. The parameters of this assay should be such that a clean, well-separated peak of the known sample is observed from the chromatograph.

Quantification of compounds by HPLC is the process of determining the unknown concentration of a compound in a known solution. It involves injecting a series of

**FIGURE 3** HPLC. Compliments of Midwest BioResearch, Evanston, IL.

known concentrations of the standard compound solution onto the HPLC for detection. The chromatograph of these known concentrations will give a series of peaks that correlate to the concentration of the compound injected.

### 4.6 Mass Spectrometry (MS)

Mass spectrometry is a powerful analytical tool because it can provide valuable structural information with a high degree of specificity. Compounds enter an electron ionization (mass spectrometer) detector and are bombarded with a stream of electrons causing them to break apart into fragments. These fragments can be large or small pieces of the original molecules and are charged ions with a certain mass. A group of four electromagnets (quadrupole) focuses each of the fragments through a slit and into the detector. The quadrupoles are programmed by the computer to direct only certain fragments through the slit. The rest bounce away. This happens repeatedly for a specific range of fragments. A graph, the mass spectrum, is then produced.

The production of a characteristic mass spectrum or fragmentation pattern acquired for each molecule makes it a definitive and effective tool for identifying unknown impurities or degradation products. This characteristic pattern is a "fingerprint." Comparing the MS "fingerprint" with large mass spectral databases further facilitates identification.

### 4.7 Gas Chromatography (GC) and Combination of GC and MS

As discussed previously, chromatography is used to separate mixtures of chemicals into individual components. The procedure for separation is similar for LC and GC with the exception of the mobile phase. In GC, the mobile phase is an inert gas such as helium.

With the combined GC/MS, the GC separates the components of a mixture. After the individual compounds elute from the GC column, they enter the electron ionization (mass spectrometer) detector. The fragments go through the same procedure as discussed for MS. MS characterizes each of the components individually. A 3D graph is produced, which provides qualitative and quantitative data. By combining the two techniques, an analytical chemist can take an organic solution, inject it into the instrument, separate the individual components, and identify each of them. Furthermore, the researcher can determine the quantities (concentrations) of each of the components.

ICH Q3C states that residual solvents are typically determined using chromatographic techniques such as gas chromatography. Any harmonized procedures for determining levels of residual solvents as described in the pharmacopoeias should be used, if feasible.

### 4.8 MS/MS

The tandem MS method has been used as an alternative to GC/MS. This method uses collision-activated dissociation on a triple quadruple mass spectrometer. This assists in the quick and direct qualitative and semiquantitative analysis. Other benefits of tandem MS include the elimination of most wet chemical and chromatographic separation steps, and the detection of both known and unknown compounds by molecular weight and functional group. A disadvantage is that tandem MS is somewhat less specific than GC/MS in the identification of some isomeric compounds.

### 4.9 HPLC/Mass Spectrometry (LC/MS or LC/MS/MS)

LC/MS data includes molecular weight and structural information that can help identify an impurity or a degradation product.

LC/MS/MS, also known as a triple-quad LC/MS, has become standard in the pharmaceutical analytical laboratory (Fig. 4). For the quantitative analysis of many analytes, LC/MS/MS is a fast, universal, selective, and sensitive tool. The quantitation is typically performed with high selectivity, which practically eliminates matrix components. LC/MS/MS is also a very important technique in the bioanalytical arena for characterizing large molecules such as botanicals, biologics, proteins, and peptides, which are typically present in challenging matrices.

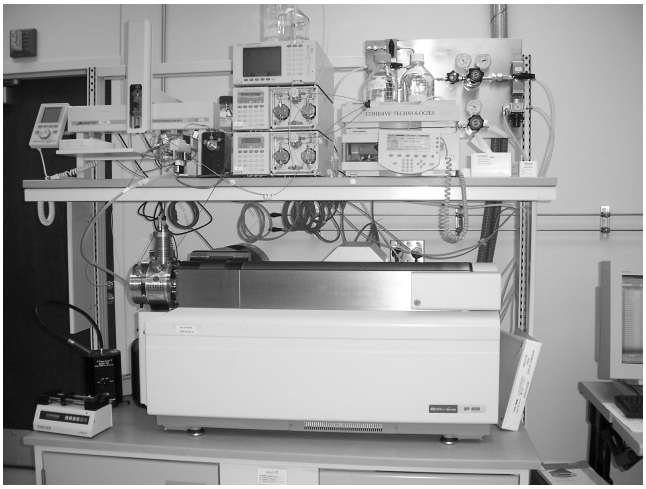

**FIGURE 4** LC/MS/MS. Compliments of Midwest BioResearch, Evanston, IL.

## 4.10 Common Problems

***Meeting GMP Standards for Maintenance and Recordkeeping*** The FDA has promulgated 21CFR Part 211, *Current Good Manufacturing Practice for Finished Pharmaceuticals* [26]. There are other sections for the FDA-issued Good Manufacturing Practices (GMPs) and the manufacturer needs to look closely to ensure that the proper regulations are followed. The regulations in this part (21CFR Part 211) contain the minimum current GMP for preparation of drug products in general for administration to humans or animals: for drug products, in Parts 600 through 680, as they pertain to drugs that are also biological products for human use; and in Part 1271, as they are applicable to drugs that are also human cells, tissues, and cellular and tissue-based products (HCT/Ps) and that are drugs.

Maintenance and recordkeeping are addressed in many sections of 21CFR Part 211. Strict guidelines are in place, such as all equipment and utensils shall be cleaned, maintained, and sanitized at appropriate intervals to prevent malfunctions or contamination that would alter the safety, identity, strength, quality, or purity of the drug product beyond the official or other established requirements.

Virtually all actions need to be documented and stored. All records need to be maintained for specified periods, including any production, control, or distribution records; records for all components, drug product containers, closures, and labeling; records for maintenance; and use logs. Therefore, everything needs to be documented and explained so that the study can be reconstructed, if necessary.

For all impurities, approval of the drug product would be based on the applicant's ability to demonstrate that the product is manufactured to ensure both efficacy and safety for patients. This can only be achieved by paying detailed attention to identity, strength, quality, and purity of the products and their components during the entire manufacturing and distribution process.

***Locating Suitable Standard and Reference Materials*** Occasionally, there may be problems locating suitable standards and reference materials. Some of these may be purchases through a reputable dealer or the USP. Some may be custom made in

a laboratory. But no matter which route is selected, documentation of all procedures and steps is critical to be in compliance with GMPs.

***Difficulty of Separation of Chromatogram Background Noise*** If a chromatograph has a lot of peaks, there is a good possibility that all of the nonactive peaks are due to impurities. It is best to look for a method that puts your peak in a clear area. Depending on the type of equipment used for analyses, changes may be made to an existing method to place the peak in a clear area.

## 5 TOXICITY ASSESSMENT

Once it has been determined that an impurity needs to be qualified, research will need to be conducted to establish safety. The literature from peer-reviewed journals will need to be assessed. Toxicology and safety assessment studies may also need to be performed.

### 5.1 Literature-Based Approaches

Literature searches can be obtained in many different ways. There are many companies who offer subscriptions to their search programs, and Information Services professionals are also available to assist you in your search. For those who prefer to perform their own searches for no charge, many websites will provide you with results that include the citations and possibly abstracts. In many cases, there are fees for full text articles. Below is a brief introduction to some of the free databases on the Web. Links are current as of November 2007.

The U.S. National Library of Medicine (www.nlm.nih.gov) has many links to other databases, including PubMed®, MEDLINE®, TOXNET®, LNM Gateway, LNM Catalog, the Hazardous Substance Data Bank (HSDB) and meeting abstracts. Other resources exist and it is worth a visit to the website.

PubMed/MEDLINE (http://www.ncbi.nlm.nih.gov/entrez/query.fcgi to register) includes over 15 million citations from MEDLINE and other life science journals for biomedical articles back to the 1950s. PubMed includes links to full text articles and other related resources.

TOXNET (http://toxnet.nlm.nih.gov/) contains databases on toxicology, hazardous chemicals, environmental health, and toxic releases. Databases include HSDB, IRIS, CCRIS, GENETOX, TOXLINE, EMIC, DART/ETIC, TRI, and ChemID/*plus*.

LNM Gateway (http://gateway.nlm.nih.gov/gw/Cmd) can be used to search multiple NLM retrieval systems.

LNM Catalog (http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=nlmcatalog) provides access to NLM bibliographic data for journals, books, audiovisuals, computer software, electronic resources, and other materials. Links to the library's holdings in LocatorPlus, NLM's online public access catalog, are also provided.

Local, state, national, and international websites exist to provide information on impurities, chemicals, and regulations. For U.S. FDA information, including guidances and regulations, go to www.fda.gov.

### 5.2 Common Testing Requirements for New Impurities

New impurities, depending on the level present in the drug product, will need to be tested for safety. Studies to be performed are based on ICH and FDA guidances and will vary on a case-by-case basis. A few studies may be needed, or the entire list. Discussions with the appropriate division of the FDA for future study strategy may be prudent. Mammalian studies most likely are done in a rodent and a nonrodent species. Rat and dog, respectively, have previously been recommended, but recent tests have used the minipig in place of the dog. In certain specific cases, nonhuman primates may also be used.

The studies include but are not limited to:

- Pharmacodynamics
- Pharmacokinetics
- Safety pharmacology
- Genetic toxicology battery
- Acute dosing
- Repeated-dose studies
- Carcinogenicity
- Reproductive and teratology studies
- Multigeneration assessments
- Special testing as appropriate (i.e., phototoxicity, dermal, inhalation, sensitization, etc.)

If needed, a repeated-dose general toxicity study in rodents and nonrodents can range from a minimum duration of 14 days to 90 days. The clinical route of administration is recommended; however, other routes (i.e., intravenous, intraperitoneal) may be necessary if exposure is an issue. A minimal genetic toxicology battery, consisting of two *in vitro* assays to detect point mutations and chromosomal aberrations, should be conducted with the test article containing the contaminant, although more precision will be gained if the isolated impurity is tested. There may be a need for carcinogenicity studies if the drug product is administered daily for >3 months. Embryo-fetal development assessment may be necessary when the parent drug is used in a population that includes women of childbearing potential. Other reproductive toxicity studies may be requested on a case-by-case basis depending on results of the general toxicity and embryo-fetal developmental studies.

### 5.3 Study Design Based on Exposure, Route, Dosage, and Other Factors

Details of the studies may also need to be discussed with the FDA or other regulatory agencies. Details such as route, dosage, length of exposure, parameters

measured, species, and age of animals are just a few considerations in the study design. Study protocols developed by the FDA or the Organisation of Economic Cooperation and Development (OECD) have been used.

### 5.4 Documentation Content and Format

Content of regulatory submissions will vary between regulatory bodies and types of submissions. Within the FDA, for example, there are differences between what is needed for approval of pharmaceuticals, biologics, and medical devices. In addition, each regulatory agency has a specific procedure for formatting. Many are moving toward the Common Technical Document (CTD). Virtually all agencies have guidelines posted on their websites to explain what is necessary for submission. More than one guideline for submissions may exist within an agency.

The FDA CDER Guidance page (http://www.fda.gov/cder/guidance/) has a few links to submission information, including information in Chemistry, Drug Safety, Electronic Submissions, ICH (final and draft), and IND (Investigational New Drug).

Providing the regulatory agency with the information that it needs for product approval is both a science and an art, and in many cases, expert assistance should be sought.

## 6 BRIDGING STUDIES

Changes are made many times during development of a drug substance or a drug product. When most of the preclinical safety testing has already been conducted, and changes are made, it may not be necessary to repeat all of the previous studies. However, a change in manufacturing techniques or a change in the impurity profile would necessitate testing to ensure that the materials before and after the change are comparable in terms of quality, safety, and efficacy. Comparability can often be deduced from available data alone but might sometimes need to be supported by comparability bridging studies.

Bridging studies may consist of toxicology studies in an appropriate species, either rodent and/or nonrodent. The nature of the bridging studies must be determined on a case-by-case basis. Examples of bridging studies are:

- An acute study.
- A repeated-dose study of 90 days' duration. Toxicokinetic samples should be analyzed and blood levels compared to levels observed in animals prior to the change.
- Pharmacodynamic studies would also need to be conducted to detect any possible changes.
- Teratology (Segment II) study in the most sensitive species (rat or rabbit, primarily).
- Pre- and postnatal development including maternal functions (Segment III).
- Other studies as needed.

Chapter, Bridging Studies in Pharmaceutical Safety Assessment, has more details of bridging studies.

## 7 DISCUSSION REGARDING TESTING THE DRUG WITH THE IMPURITY VERSUS ACTUAL IMPURITY

What exactly is tested for studies to determine the safety of impurities? As time goes on and the chemists have more experience with the drug substances, they find more efficient methods of producing the substance in hopes that the impurity profile will decrease. Most of the time, especially early on in development, the impurity is already present in the drug substance that is administered to the animals. If the impurity is present in a high enough concentration to provide a relevant dosage to the test systems in repeated-dose studies, the impurity may be considered adequately tested.

ICH Q3A states that the level of any impurity present in a new drug substance that has been adequately tested in safety and/or clinical studies would be considered qualified. ICH Q3B (R) describes considerations for the qualification of degradation products when thresholds are exceeded. If additional safety testing is needed, such studies can be conducted on the new drug product or substance containing the degradation products to be controlled, although studies using isolated degradation products can sometimes be appropriate.

Testing the actual impurity is important in some cases, and it would be prudent for the industry, FDA, and other agencies to hold discussions on this topic. Some impurities are present in rather small amounts. Toxicology tests may not be sensitive enough to characterize their individual toxicities if impurities are tested within a drug substance or product. Testing the toxicity of an impurity present in a drug substance or product in toxicology studies or bridging studies may very well be a waste of time, money, and animals if no toxicity from the impurity is observed due to low levels. However, using the predictive power of several structure–activity relationship (SAR) programs, specific toxicities identified for individual impurities could be assessed.

Testing the impurity at appropriate levels to produce appropriate safety factors for *in vivo* studies would elucidate specific toxicities from the impurity. But more crucial is testing the isolated impurity in genetic toxicology studies. If tested as a component of the drug substance, the exposure of the impurity to the target cells in these studies would be so low that only the most potent mutagens would be detected. To understand the toxicological implications of impurities, testing of the isolated impurity at appropriate concentrations is essential.

## ACKNOWLEDGMENTS

I would like to thank Robert E. Osterberg, Office of New Drugs, Center for Drug Evaluation and Research, U.S. Food and Drug Administration, for his review of the draft manuscript for this chapter. I also thank Mike Schlosser, president of Midwest BioResearch, for allowing me to photograph the equipment.

## REFERENCES


1. Janssen WF. The story of the laws behind the labels, Part I, 1906 Food and Drugs Act. *FDA Consumer*, June 1981. Available at http://vm.cfsan.fda.gov/~lrd/history1.html.
2. US FDA Center for Biologics Evaluation and Research. *Commemorating 100 Years of Biologics Regulation, Science and the Regulation of Biological Products From a Rich History to a Challenging Future*, 25 September 2002. Available at http://www.fda.gov/cber/inside/centscireg.htm.
3. US FDA Office of Women's Health. *FDA Milestones in Women's Health: Looking Back as We Move into the New Millennium*, 2000. Available at http://www.fda.gov/womens/milesbro.html.
4. Roberts R. *Safety Assessment in Pediatrics*. Center for Drug Evaluation and Research, US Food and Drug Administration; July 8, 1999. Available at www.fda.gov/cder/present/dia-699/sa-dia99/index.htm.
5. US FDA Center for Drug Evaluation and Research. *Time Line: Chronology of Drug Regulation in the United States*. Available at www.fda.gov/cder/about/history/time1.htm.
6. US FDA. *Milestones in U.S. Food and Drug Law History, FDA Backgrounder: Current and Useful Information from the Food and Drug Administration*, August 1995. Available at www.fda.gov/opacom/backgrounders/miles.html.
7. US Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research (CDER) and Center for Biologics Evaluation and Research (CBER). *Guidance for Industry, Nonclinical Studies for the Safety Evaluation of Pharmaceutical Excipients*, May 2005.
8. United States Pharmacopoeia (USP), 2005. Available at www.usp.org.
9. ICH. *Guidance for Industry, Q3C Impurities: Residual Solvents*, December 1997. Available at http://www.fda.gov/cder/guidance/index.htm.
10. ICH. *Guidance for Industry, Q1B Photostability Testing of New Drug Substances and Product*, November 1996. Available at http://www.fda.gov/cder/guidance/index.htm.
11. ICH. *Guidance for Industry, Q3A Impurities in New Drug Substances*, February 2003. Available at http://www.fda.gov/cder/guidance/index.htm.
12. ICH. *Guidance for Industry, Q3B(R) Impurities in New Drug Products*, November 2003. Available at http://www.fda.gov/cder/guidance/index.htm.
13. ICH. *Guidance for Industry, Q1A Stability Testing of New Drug Substances and Products*, August 2001. Available at http://www.fda.gov/cder/guidance/index.htm.
14. FDA website, 2005. Available at http://www.fda.gov/cder/guidance/cGMPs/packaging.htm.
15. ICH. *Guidance for Industry, Q6B Specifications: Test Procedures and Acceptance Criteria for Bio-technological/Biological Products*, August 1999. Available at http://www.fda.gov/cder/guidance/index.htm.
16. Osterberg RE. Extractables and leachables in drug products. *Am Coll Toxicol Newslett* 2005;25(1):1020.
17. Osterberg RE. Potential toxicity of extractables and leachables in drug products. *Am Pharm Rev* 2005;8(2):1020–1021.
18. US Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research (CDER). *Guidance for Industry, Metered Dose Inhaler (MDI) and Dry Powder Inhaler (DPI), Drug Products Chemistry, Manufacturing, and Controls Documentation*, October 1998. Available at http://www.fda.gov/cder/guidance/index.htm.

19. ICH. *Guideline for Industry, Q2A Text on Validation of Analytical Procedures*, March 1995. Available at http://www.fda.gov/cder/guidance/index.htm.
20. ICH. *Guidance for Industry, Q2B Validation of Analytical Procedures: Methodology*, November 1996. Available at http://www.fda.gov/cder/guidance/index.htm.
21. ICH. *Guidance for Industry, Q6A Specifications: Test Procedures and Acceptance Criteria for New Drug Substances and New Drug Products: Chemical Substances*, December 2000. Available at http://www.fda.gov/cder/guidance/index.htm.
22. Code of Federal Regulations; Title 21, Volume 5; Revised as of 1 April 2005. 21 CFR 314.3(b); Title 21—Food and Drugs, Chapter I—Food and Drug Administration Department of Health and Human Services, Subchapter D—Drugs for Human Use. Part 314, *Application for FDA Approval to Market a New Drug*, Subpart A—*General Provisions*, Sec. 314.3 *Definitions*. Revised as of 1 April 2005. Available at www.fda.gov.
23. ICH. *Guidance for Industry, Q3C—Tables and List*, November 2003, Revision 1. Available at http://www.fda.gov/cder/guidance/index.htm.
24. US Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research (CDER). *Guidance for Industry, NDAs: Impurities in Drug Substances*, February 2000. Available at http://www.fda.gov/cder/guidance/3622fnl.htm.
25. US Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research (CDER). *Guidance for Industry, ANDAs: Impurities in Drug Substances*, August 2005. Available at http://www.fda.gov/cder/guidance/index.htm.
26. US Department of Health and Human Services, Food and Drug Administration. 21CFR Part 211 *Current Good Manufacturing Practice for Finished Pharmaceuticals*. Revised 1 April 2005. Available at www.fda.gov.
27. US Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research (CDER). *Draft Guidance for Industry, Analytical Procedures and Methods Validation Chemistry, Manufacturing, and Controls Documentation*, August 2000. Available at http://www.fda.gov/cder/guidance/2396dft.htm.
28. US Department of Health and Human Services, Food and Drug Administration. 21CFR Part 58 *Good Laboratory Practice for Nonclinical Laboratory Studies*, Revised 1 April 2005. Available at www.fda.gov.
29. Thornton DJ, Thornton BL. Instrumentation systems: reliability and validity issues. In Ward KM, Lehmann CA, Leiken AM, Eds. *Clinical Laboratory Instrumentation and Automation—Principles, Applications and Selection*. Philadelphia: WB Saunders; 1994, Chap 15.
30. Bender JT. Ultraviolet absorption spectroscopy. In *Principles of Chemical Instrumentation*. Philadelphia: WB Saunders; 1987, Chap 6.
31. Narayanan S. Atomic absorption spectroscopy. In *Principles and Applications of Laboratory Instrumentation*. Chicago: American Society of Clinical Pathologists; 1989.

# BRIDGING STUDIES IN PRECLINICAL PHARMACEUTICAL SAFETY ASSESSMENT

Shayne Cox Gad
*Gad Consulting Services, Cary, North Carolina*

**Contents**

1 Introduction
2 Causes and Solutions
  2.1 Formulation Changes
  2.2 Changes in Manufacturing Processes or Site
  2.3 Alterations in Proportions of Optical Isomers
  2.4 Issues of Test Species Selection
  2.5 Impurity and Degradant Issues
  2.6 Active Metabolite Issues
  2.7 Correlation of Effects Between Nonclinical Species
  2.8 Correlation of Surrogate Markers with Established Effects on Toxicity Indicators
3 Issues of Bridging Study Design
  3.1 Test Compound(s)
  3.2 Basic Bridging Study Design
  3.3 Comparative (or Base) Species Data: Historical or Concurrent
  References

## 1 INTRODUCTION

Preclinical toxicology bridging studies are unique to the safety evaluation of pharmaceuticals [1–3]. While there have been analogues in clinical drug evaluation [4–6], only in the development of new drugs or new drug formulations are we faced with

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

**TABLE 1 Reasons for Performance of Bridging Studies**

| |
|---|
| *Performed to Allow for Changes In:* |
| Formulations (most common) |
| Manufacturing processes |
| Optical isomer proportions |
| Test species selection/issues |
| Impurity or degradant issues |
| Active metabolite issues |
| *Performed to Accommodate:* |
| Correlation of effects between nonclinical species |
| Correlation of surrogate markers with established effect on toxicity markers |

the need to establish relative comparability between different formulations or dosing regimens of an active ingredient, or show the comparability of response of two different species of animals intended to predict toxicity and safety in humans. Such actions are usually attractive compared to having to repeat an entire nonclinical safety evaluation program. Rather, only one or two critical studies (most sensitive points/tests) are performed to establish relationships.

As technology (and the cost of drug development) has advanced, so has the need for carefully considered bridging studies.

Table 1 presents the primary reasons behind the conduct of preclinical bridging studies.

## 2 CAUSES AND SOLUTIONS

An excellent starting place is to consider the reasons for performing bridging studies, as listed in Table 1, as well as the issue behind each.

### 2.1 Formulation Changes

Frequently occurring after initial clinical trials, here bridging studies typically look at the change in pharmacokinetics, to assure that the new formulation components do not increase human tolerance issues.

*Issue*: Addition of new or alteration of combination of formulation components may either cause safety problems not previously seen or alter active ingredient pharmacokinetics in an unfavorable manner.

### 2.2 Changes in Manufacturing Processes or Site

Simple parallel comparative toxicity studies are commonly performed to establish that the active pharmaceutical ingredient (API) and/or clinical product are pharmacokinetically and toxicologically equivalent. These studies are also used if cell lines for producing biotechnology products are changed.

*Issue*: Changes in the process by which an API is manufactured or in the site where manufacture is performed may change the nature of the API in unforeseen

ways; a bridging study serves to confirm equivalent biologic activity (both therapeutic and toxicological).

### 2.3 Alterations in Proportions of Optical Isomers

Assuming that they are not metabolically interconverted, a change in enantiomers in a racemic mixture of API generally calls for a comparative study.

*Issue*: As chemical development proceeds in parallel with other aspects of the development of a new drug, the proportion of optical isomers may change. As this can have widely different biologic effects, bridging studies seek to establish acceptability of any changes (which subquently can lead to changes in API specification).

### 2.4 Issues of Test Species Selection

First-in-Human (FIH) clinical trials reveal that one or both primary (rodent and nonrodent) species used to evaluate or predict safety to this point are not adequate—either in terms of pharmacokinetics or because a metabolite found in humans is not present (or present in inadequate amounts) in the test species. A bridge with another study must then be performed.

*Issue*: Nonclinical safety studies form the principal basis for providing confidence that the clinical development of a potential drug can be pursued with acceptable levels of risks. Change to a new process (rodent or more commonly nonrodent) requires adequate characterization of pharmacodynamics and toxicodynamics before longer term studies are conducted.

### 2.5 Impurity and Degradant Issues

These are perhaps becoming the most common types of bridging studies, although in this context the studies are sometimes called qualification studies. The needs arise either as extended stability studies disclose new or increased levels of a degradant formed as the result of a stability issue, or as impurities are discovered in newly produced lots of either API or clinical formulations.

*Issue*: The detection of an impurity or degradant in a pharmaceutical product presents a potential new and additional risk to patients exposed to the product. The magnitude (and acceptability) of potential risk must be established before use/evaluation clinically.

### 2.6 Active Metabolite Issues

Particularly in accordance with the new International Conference on Harmonization (ICH) and Center for Drug Evaluation and Research (CDER) guidelines, the need for these studies usually arises subsequent to the determination that there is an active metabolite present in humans at significant levels (the guidelines say 10% of the total area under the curve (AUC)), but lower levels have triggered FDA requests. The result may be a variation of the test species case. Selection of test species with the benefit of at least a metabolite profile in microsomes or hepatocytes will reduce but not exclude the possibility of this problem.

*Issue*: Active (or potentially active) metabolites that occur in human beings at levels not previously adequately evaluated in NCS studies may represent a significant risk to patients/subjects and must be evaluated for relevant potential clinical issues before further clinical development is pursued.

### 2.7 Correlation of Effects Between Nonclinical Species

It sometimes comes to pass that findings in one species are not readily apparent in another. This may be due to a wide variety of reasons but special studies "extending" evaluation in the species not showing the effect (such as achieving an equivalent level of systemic exposure) may serve to clarify relevance to humans.

*Issue*: As pointed out in Table 2, sometimes the identified potential toxicities in one of the species utilized in NCS evaluations are vastly different (usually to the worse) from the other species. In this case, unless there is a known basis for the difference that allows identification of which prediction is more relevant to potential clinical safety, such relevance (or basis for difference) needs to be established.

### 2.8 Correlation of Surrogate Markers with Established Effects on Toxicity Indicators

Subsequent to the performance of nonclinical systemic toxicity studies, it can occur that one wishes to demonstrate that surrogate markers may serve to identify an adverse effect in a less invasive manner, or to predict the early stages of effects before they become serious or irreversible. This particularly occurs when the actual toxic effect would be either unacceptable or highly undesirable in patients.

*Issue*: While Table 1 presents the major reasons for performing bridging studies, there are other potential reasons. There are a common set of inconsistencies or

**TABLE 2 Troubleshooting in General Toxicology**

| | |
|---|---|
| Unexpected toxicity compared with prior tests | Change in formulation or batch of test chemical; poor predictivity of dose range-finder studies due to factors such as differences in animal age, supplier, or husbandry. |
| Variation in individual response | Metabolic polymorphism or other genetic factor; social factors in group housing (e.g., nutrition status) |
| Low systemic concentration or area under the curve (AUC) | Poor absorption or poor formulation; isotonicity is important in parenteral formulations; extensive first-pass effect; short half-life. |
| Low toxicity | Low availability; inappropriate route of administration or dose selection. |
| Interspecies differences | Different ADME; different mechanism of effect; species-specific mechanisms such as peroxisome proliferation; enterohepatic recirculation; different expression of or affinity for pharmacological receptors. |
| Different response in males and females | Especially in rodents; due to different activities of metabolism enzymes in liver particularly but also physiological differences such as $\alpha_2$-microglobulin excretion in males. |

problems that arise from many standard safety evaluation programs. Table 2 presents these, as well as the common genesis of their occurrence.

## 3 ISSUES OF BRIDGING STUDY DESIGN

### 3.1 Test Compound(s)

What test material is used in a bridging study is dictated by the issue to resolve because in most cases a portion of the question posed can be phrased as "Is the new drug material equivalent to that previously evaluated?" with equivalence potentially being pharmacokinetic, pharmacodynamic, or toxicodynamic. This being the case, there are two test materials/formulations to be included in a bridging study, with the first being the defined "standard" or material to date. This is often called the reference material.

The second (or "new" or "different") material will then generally be the "standard" with the extreme degree of changed nature (the most impurity or degradant or optical isomer) or isolated metabolite or new formulation.

### 3.2 Basic Bridging Study Design

The most common is the comparative study, in which at least two groups of test animals are given two different doses (low and high) of the reference material, two more are similarly given doses of the changed material (or metabolite), and a fifth group acts as a control, receiving only vehicle or formulations. The test animal species employed is determined by the issue to be resolved.

Test material administration is by the intended clinical route, with the intended frequency of administration (unless, of course, this is the issue to be addressed). Duration should be from 2 to 4 weeks.

### 3.3 Comparative (or Base) Species Data: Historical or Concurrent

While in bridging studies (unlike most other NCS studies), the direct comparison of concern is between the reference and new active material groups, this is not always or purely the case. Sometimes (such as with active metabolites), we are comparing responses or results between new material groups and a control group. While a concurrent control is almost always the most appropriate comparator, considerations of variability of the normal range in a historical population may be critical to understanding the relevance of a change or effect seen with a new material.

## REFERENCES

1. Anderson C, et al. Bridging studies in Asia and the impact of the ICH E5 guideline. *Drug Information J* 2003;37:107(S)–116(S).
2. Arnold DL, Grice HC, Krewski DR. *Handbook of In Vivo Toxicity Testing*. San Diego: Academic Press; 1990.
3. Gad SC. *Drug Safety Evaluation*. Hoboken, NJ: Wiley-Interscience; 2002.

4. Liu JP, Chow SC. Bridging studies in clinical development. *J Biopharm Stat* 2002;12(3):359–367.
5. Naito C. Implementation of the bridging study strategy and extension of the "bridging" concept: current status. *Drug Information J* 2003;37:149(S)–154(S).
6. Naito C. Necessity and requirements of bridging studies and their present status in Japan. *Int J Clin Pharmacol Ther* 2000;38(2):80–86.

# Historical Perspective

ROBERTA S. KING

## 1 CONTROVERSIES SPANNING PAST, PRESENT, AND FUTURE

Two major issues have been debated throughout the history of drug metabolism, and are still disputed to some degree. One is the name itself, the other is the physiological purpose of "drug" metabolism. In the 1800s and early 1900s, the generally agreed purpose of these reactions was reflected in the most widely used name, *detoxication mechanisms*. However, *detoxication* became widely recognized as a misnomer because not all parent compounds were toxic and not all metabolites were less or nontoxic. A better term was not invented until the 1950s when the term "drug metabolism" was coined. While handy, this term was still not entirely valid, and it needed silent agreement that "drug" be not restricted to medicinal compounds (Bachmann and Bickel, 1985–86). Thus, *xenobiotic metabolism* became popular starting in the 1970s, especially in circles studying carcinogens and environmental compounds. Xenobiotic, by definition, included all compounds foreign to the organism, not just medicinal ones. However, even in the early 1900s many examples were already known of metabolism of endogenous compounds, for example, steroids undergoing glucuronidation. These early examples of endogenous substrates were generally dismissed because they typically occurred at much higher concentrations than normally present, so-called "supraphysiological" concentrations. While none of these three terms could be considered ideal, in 1947 R.T. Williams concluded that the field of detoxication included "… all those metabolic processes not specifically covered by the main streams of fat, carbohydrate and protein intermediary metabolism" (Williams, 1947). Williams went on to explain that, "Detoxication is, in fact, the study of the metabolism of organic compounds other than lipids, carbohydrates, proteins and closely related natural compounds, although the lines of demarkation between these two groups is by no means a sharp one …" (Williams, 1947; Bachmann and Bickel, 1985–86). Thereby, the earliest clear description of the field mainly described what it was not, and later terms were not much more precise.

The physiological purpose of these reactions was also widely debated. Of note is that several of the early theories are still considered at least partially valid. The first theories tried to answer the question of how these transformations were related to

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

detoxication. For example, in 1917, Berczeller published the first article trying to answer why conjugation would result in detoxication (Berczeller, 1917; Bachmann and Bickel, 1985–86). His theory, later to be disproved, was that conjugations such as glucuronidation and sulfonation led to a change in the surface tension of dilute aqueous solutions. He related this to *in vivo* conditions by indicating that the conjugates, with less surface activity than the parent drug, would be more easily removed from cellular surfaces while the parent drugs would accumulate at the surfaces to toxic concentrations (Bachmann and Bickel, 1985–86). The next hypothesis, proposed in 1922 by Sherwin (Sherwin, 1922), was based on the idea of "chemical defense" against accumulation of foreign compounds. This hypothesis is still considered valid as one role of drug metabolism. Sherwin proposed that the body needed to completely destroy foreign compounds or excrete them in the urine. Thus, for Sherwin, the purpose of oxidation and reduction reactions was to destroy the foreign molecules. If complete destruction was not possible, then conjugation reactions could make the molecule more aqueous soluble and more easily removed through the urine. In 1925, Schüller (Schüller, 1925) proposed a theory also similar to the modern view of the physiological purpose of drug metabolism. Schüller proposed that conjugation reactions led to an increase in the water solubility of compounds leading to a change in the distribution of the compound in the body (Schüller, 1925). The third theory, which resonates somewhat with today's metabolism scientists, was proposed by Quick in 1932 (Quick, 1932). His view was that conjugations resulted in an increase in acidity, converting a weak acid parent into a strong acid metabolite which could be eliminated more easily. One benefit of this hypothesis was that it included in its rationale the acetylation reaction, which was ignored by Sherwin and Schüller. In his 1947 comprehensive review of metabolism (Williams, 1947), R.T. Williams criticized each of these potential roles of metabolism. Yet rather than proposing his own theory, Williams concluded that "interpretations worthy of the status of theories were lacking at that time" (Bachmann and Bickel, 1985–86).

In 2008, the field is described as an "elaborate defense system against foreign compounds and against the accumulation of potentially toxic endogenous molecules" (Meyer, 2007). However, the same author also adds recognition of the modern "concept of molecular links between xenobiotic metabolism and endogenous pathways of sterol, lipid, bile acid and energy homeostasis." To complicate matters, the current generation of genetic technologies has revealed that all of the enzyme families contributing to drug metabolism include both members with selectivity for endogenous compounds and members with selectivity for exogenous compounds, and that homologues are present throughout diverse species from bacteria to human. Indeed, gene knockout studies have shown that multiple members of the "drug"-metabolizing enzyme families are essential to life or reproductive processes, as well as serving as a means to defend against xenobiotics (Sheets, 2007). Thus, in 2008, the field has no more specific name than the popular term "drug metabolism," yet its physiological function has broadened to include endogenous compounds and endogenous regulatory pathways.

## 2 1800s: DISCOVERY OF MAJOR DRUG METABOLISM PATHWAYS (CONTI AND BICKEL, 1977)

Many current students of metabolism are surprised to learn that drug metabolism experiments were first conducted and published more than 180 years ago, first in dogs in 1824 and next in humans in 1841 (Wöhler, 1824; Ure, 1841). While these experiments

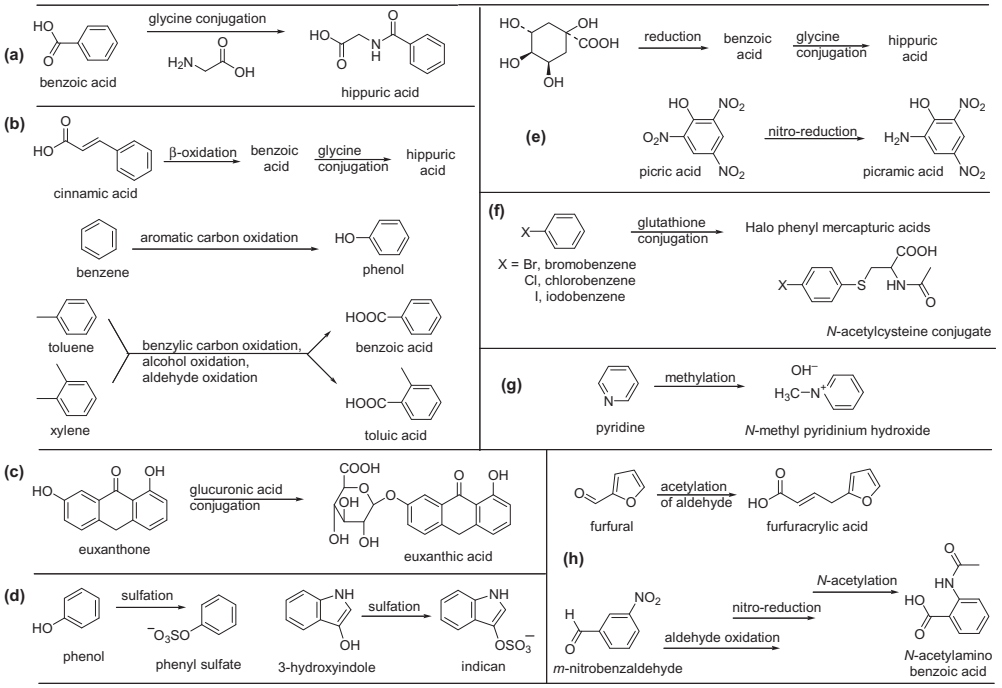


**Figure 1** Examples of first-observed metabolism reactions. **(a)** glycine conjugation; **(b)** β-oxidation, aromatic carbon oxidation, benzylic carbon oxidation, aldehyde oxidation; **(c)** glucuronide conjugation; **(d)** sulfate conjugation; **(e)** reduction; **(f)** glutathione conjugation; **(g)** methylation; **(h)** acetylation.

were quite rudimentary by today's standards, they laid the foundation for all of metabolism by establishing the paradigm that the body could take up exogenous compounds, perform chemical reactions on them, and then remove them via the urine in a chemically altered form. Also noteworthy is that by 1893, one or more examples of each major metabolism pathway had been published (Table 1, Fig. 1). Nearly all of this early work was published in German, often by German scientists. This was a time of significant advances in chemistry and medicine in German laboratories. However, the metabolism studies were limited by the difficulty in purifying and proving structure identification of the proposed metabolites. Advances in organic chemistry and analytical methods often directly led to advances in the metabolism studies.

Glycine conjugation has the honor of being the first-described metabolism reaction. In 1841 and 1842, two scientists independently ingested benzoic acid and observed a compound in the urine "in copious amounts," and "without any apparent unhealthy effects" (Ure, 1841; Keller, 1842; Conti and Bickel, 1977). The excreted compound was similar to benzoic acid, but also contained nitrogen. Identification of the compound as hippuric acid, the glycine conjugate of benzoic acid, was made in 1845 by a French scientist named Dessaignes (Dessaignes, 1845).

The benzoic acid studies soon led to the observation of oxidation as a preliminary step to conjugation, by observing that ingestion of cinnamic acid also caused excretion

**TABLE 1 Early development of the major drug metabolism pathways.**[a]

| Year | Reaction name | Substrate | Intermediate | Product | Original citation (Conti and Bickel, 1977) |
|---|---|---|---|---|---|
| 1824, 1841 | Glycine conjugation | Benzoic acid | None | Hippuric acid | Wöhler (1824), Ure (1841) |
| 1842, | β-oxidation | Cinnamic acid | Benzoic acid | Hippuric acid | Erdmann and Marchand (1842a, b) |
| 1848, | Aldehyde oxidation | Benzaldehyde | None | Benzoic acid | Wöhler and Frerichs (1848) |
| 1867 | Aromatic and benzylic carbon oxidation | Benzene | None | Phenol | Schultzen and Naunyn (1867) |
| | | Toluene | Benzyl alcohol | Benzoic acid (via further oxidation of alcohol to carboxylic acid) | |
| 1844, 1870 | Glucuronide conjugation | Mango leaves | None | Euxanthic acid, Euxanthone and oxidized glucose | Erdmann (1844), Baeyer (1870) |
| 1851, 1876 | Sulfate conjugation | Benzene | Phenol | Conjugated phenol | Staedeler (1851), Baumann (1876) |
| | | Hydroxyindole | | Indole-sulfate | |
| 1863 | Reduction | Quinic acid | Benzoic acid | Hippuric acid | Lautemann (1863) |
| | | Picric acid | | Picramic acid | Karplus (1893) |
| 1879, 1884 | Glutathione conjugation | Bromobenzene | None | Mercapturic acid conjugates | Baumann and Preuss (1879), Jaffe (1879), Baumann (1884) |
| | | Chlorobenzene | | Acetylcysteine conjugates | |
| 1887 | Methylation | Pyridine acetate | Pyridine | *N*-methyl pyridinium hydroxide | His (1887) |
| 1887, 1893 | Acetylation | Furfural | *N*-acetylfurfural | Furfuracrylic acid | Jaffe and Cohn (1887), Cohn (1893) |
| | | *m*-nitrobenzaldehyde | | *N*-acetyl amino benzoic acid | |

[a]Structures of substrates and products are shown in Fig. 1.

of hippuric acid into the urine (Erdmann and Marchand 1842a, b; Wöhler and Frerichs, 1848). We now call this biotransformation of cinnamic acid to benzoic acid, β-oxidation. Wöhler and Frerichs (1848) also discovered aldehyde oxidation when they found that dogs and cats excreted a "considerable amount" of benzoic acid after treatment with benzaldehyde (oil of bitter almonds) (Conti and Bickel, 1977, p. 11).

Nothing of the mechanism of these transformations was yet understood. During this era, the body was considered simply a chemical reaction container. Indeed, the chemists found many transformations "absolutely puzzling" (Conti and Bickel, 1977, p. 8) because, outside the organism, they could only be reproduced only under very harsh conditions, if at all. For example, oxidation of benzene to phenol had never been accomplished when, in 1867, Schultzen and Naunyn published their very clear determination of phenol in the urine of humans and dogs after ingestion of benzene. Even in the twenty-first century, only rather harsh and non-physiological conditions are known to chemically transform benzene to phenol (March, 1992).

In 1844 Erdmann observed that euxanthic acid isolated from urine of cows fed mango leaves could be hydrolyzed to euxanthone, but it took until 1870 to characterize the sugar moiety as an oxidized form of glucose. Similarly, conjugated phenols were observed in 1851, but it took until 1876 to confirm that the conjugate was a sulfate producing sulfuric acid upon hydrolysis (Staedeler, 1851; Baumann, 1876). Baumann is known as the "father of sulfation" and subsequently showed that many other ingested chemicals were also excreted as sulfates in the urine, often after preliminary oxidation to a phenol (Baumann, 1876). Methylation was first described in 1887 after feeding pyridine acetate to dogs and isolating the *N*-methyl product from the urine (His, 1887). Also in 1887, Jaffe and Cohn first observed a product that appeared to be conjugated with acetic acid (Jaffe and Cohn, 1887), although it took until 1893 for Cohn to confirm *N*-acetylation as a major conjugation reaction (Cohn, 1893).

While biochemists, chemists, physiologists, and pharmacologists contributed to these discoveries, the emphasis was on studying the metabolism of "foreign" compounds. More importantly, the emphasis was the fate of the chemical compound, rather than the organism that transformed it. Indeed, Bachmann and Bickel concluded that before circa 1870, "the whole matter was a biochemical curiosity rather than a physiologically meaningful process" (Bachmann and Bickel, 1985–86, p. 213). Evidence that this view began to change comes from Nencki's thesis of 1870, foreseeing that, "… one will on the one hand be able to establish laws allowing predictions on the fate of new compounds, and on the other hand gain increasing insight into the organism as a 'chemical agent'." (Nencki, 1870) Beginning in about 1876 with Baumann's phenyl sulfate (Baumann, 1876), it was often found that the excreted compounds were much less toxic than their parent compounds. And by 1893, enough evidence existed for introduction of the term "detoxication" in a textbook of physiological chemistry (Neumeister, 1893; Bachmann and Bickel, 1985–86, p. 213). Detoxication still stands as one major physiological role of metabolism, but is no longer recognized as the only role.

## 3 1900–1950s: CONFIRMATION OF MAJOR PATHWAYS AND MECHANISTIC STUDIES

As better techniques for identification of compounds were developed, the major reactions were confirmed and put upon somewhat stronger structural foundations; and in two cases the active cofactors for the reactions were elucidated. For example, Lipmann

earned the 1953 Nobel Prize in Medicine, in part, for his 1945 publication of the role of coenzyme A in acetylation of sulfanilamides (Lipmann, 1945). In 1953, Cantoni published evidence for *S*-adenosyl methionine as the active cofactor for methylation reactions (Cantoni, 1953). The biosynthetic source of glycine was elucidated in 1946 (Shemin, 1946), but it was not yet understood how the conjugation reaction occurred under physiological conditions.

Nencki's (1870) prediction of gaining insight into the organism as a chemical agent was also developed in significant ways. In the 1800s, the blood was often considered the localized source of the transformations, and the emphasis was on animal or human ingestion of a compound and isolation of excretion products from the urine. In the early 1900s, techniques such as hepatectomy and perfusion of livers and kidneys of laboratory animals proved the alternative paradigm of organ-based metabolism. One novel approach (Hemingway, Pryde and Williams, 1934) used serially perfused dog organs (such as liver or spleen) in combination with the kidney to demonstrate that the liver was the main site of glucuronic acid conjugation. In 1936, Potter and Elvehjem described an improved alternative to the commonly used tissue slice and tissue mince methods (Potter and Elvehjem, 1936). Their device used a glass test tube as mortar with a motor-driven blown-glass pestle for nondestructive tissue homogenization. Claude added differential centrifugation to the improved homogenization technique, and made possible subcellular localization of metabolism via isolation of individual tissue organelles (Claude, 1940). He also coined the still-popular term "microsome" (Claude, 1943).

Of note is that during the early twentieth century, most metabolism studies were conducted within biochemistry departments of university medical schools. Also during this time, because of (or in spite of) the World Wars, metabolism research expanded across Europe to England and North America, and by 1950 the United States had replaced Germany as the dominant origin of metabolism publications.

Approximately 100 years after the first published discoveries, "Modern" metabolism science was founded by the Welshman R.T. Williams, who in 1947 published the first text devoted to metabolism entitled *Detoxication Mechanisms: The Metabolism of Drugs and Allied Organic Compounds*. This text was expanded in a second edition in 1959 with a modified title, *The Metabolism and Detoxication of Drugs, Toxic Substances, and Other Organic Compounds*. In these texts, Williams brought systematic organization and clarity to what had previously been broad and disconnected research, "so that working hypotheses (could) be advanced" (Williams, 1947, quoted from Caldwell, 2006). It was Williams who proposed that metabolism occurred through reactions representing two distinct phases leading to the still popular terms, "phase I" and "phase II" metabolism reactions. Another revolutionary, B.B. Brodie of the United States National Institutes of Health, first published in 1948 and "led the field into its modern phase" (Bachmann and Bickel, 1985–86, p. 188). As foreseen by Nencki in 1870 (Nencki, 1870), predictive rules for functional group transformations began to emerge, and the enzymology of the reactions began to be elucidated.

## 4 1950s–1980: MODERN DRUG METABOLISM EMERGES, WITH ENZYMATIC BASIS

Because of advances in analytical technologies and biochemical methods, metabolism studies took off starting in about 1950. For example, partition chromatography improved

separations and allowed differentiation of drug versus metabolites. Isotope-tracer methods (mostly 14C and 15N) allowed metabolites of foreign compounds to be detected and quantified at nontoxic doses. Absorption spectrophotometry also improved both quantification and identification of drug versus metabolites. For the first time, the cofactors of the enzymatic reactions were fully elucidated, and the enzymatic biosynthesis pathways of the cofactors were established. For example, the molecular mechanism of glucuronidation utilizing the reactive cofactor, uridine-3′,5′-diphosphate glucuronic acid (UDPGA) was published in 1954 (Dutton and Storey, 1954). Advances in tissue fractionation methods provided the means to prove the enzymatic basis of metabolism. Indeed, Brodie and coworkers published the first review of this emphasis in 1958 entitled, "Enzymatic metabolism of drugs and other foreign compounds" (Brodie, Gillette and La Du, 1958). In the late 1950s and early 1960s, the cytochrome P450 enzyme system was discovered and characterized as the source of microsomal oxidations of several drugs and steroid hormones (Mason, 1957a, b; Klingenberg, 1958; Omura and Sato, 1962; Omura *et al.*, 1965). By the end of this period, the role of cytochrome P450 in drug metabolism was well established and the important modifying factors (inhibition, induction, and polymorphisms) were beginning to be understood (Conney *et al.*, 1980; Netter, 1980; Orrenius, Thor and Jernström, 1980; Ritchie *et al.*, 1980).

## 5 1980–2005: FIELD DRIVEN BY IMPROVED TECHNOLOGIES

Major advances during this 25-year period were driven by improved technologies. The major human cytochrome P450 members were identified, purified, and characterized from human liver tissue starting in 1983 (Wang *et al.*, 1983). Shortly after, development of cloning and heterologous expression techniques led to single isoform preparation with relative ease. Robotics for liquid handling and improvements in fluorescent technologies have driven the development of successful high-throughput screening procedures. Advanced separation systems linked to mass spectrometry or NMR detection systems have driven metabolite structure analysis and quantitation. Advances in viable hepatocyte isolation have improved the ability to characterize inducers of specific enzymes. Other advances include *in silico* modeling of all stages of metabolism including *in vitro* to *in vivo* modeling and computational prediction of human drug metabolism. Improvements in protein crystallization techniques, availability of powerful synchrotron sources, and improvements in 3D-structure generation software have led to the generation of multiple structures of each enzyme family providing an appreciation of structure variability and invariability. More recently, the study of transporters and transport systems has driven a paradigm shift toward the recognition of the important contributions of transport to drug development and drug safety.

## 6 2005+: HIGH TECHNOLOGY

While some would describe the metabolism field as "mature," technological advances still push the field forward in dramatic ways. We are now in an age of high technology in metabolism science, as evidenced by the emphasis in Part II of this text. Indeed, most of the knowledge described in this text would not be possible without the technological advances of the recent 30 years. Much of this modern innovative technology has been

driven by the pharmaceutical industry, including highly advanced analytical instrumentation and high-throughput technologies. It has also been influenced by great advances of the "genomics era," through recognition of the genetic basis of variation in drug metabolism among individuals and populations. The future of the "drug" metabolism field will be driven by further technological advances toward the purposes of elucidating and understanding the impact of metabolic "cross talk": the complex three-way interactions among endogenous compounds and their regulatory pathways, exogenous compounds, and disease states.

## ACKNOWLEDGMENTS

The author acknowledges three English language reviews of the early metabolism literature as major sources: Bachmann and Bickel (1985–86), Caldwell (2006), and Conti and Bickel (1977). All quotations are as provided in the reviews in English translation. Publications published prior to 1900 were not accessed directly and are listed as cited in Conti and Bickel (1977).

## REFERENCES

Bachmann, C., Bickel, M.H. (1985–86). History of drug metabolism: the first half of the 20th century. *Drug Metab. Rev.*, *16* (3), 185–253.

Baeyer, A. (1870). *Ann. Chem. Pharm.*, *155*, 257.

Baumann, E. (1876). Concerning the occurrence of brenzcatechin in the urine. *Pflügers Arch. Physiol.*, *12*, 69.

Baumann, E. (1884). Ueber cystin und cystein. *Z. Physiol. Chem.*, *8*, 299.

Baumann, E., Preusse, C. (1879). *Z. Physiol. Chem.*, *3*, 156.

Berczeller, L. (1917). The excretion of substances foreign to the organism in the urine. *Biochem. Z.*, *84*, 75–79.

Brodie, B.B., Gillette, J.R., La Du, B.N. (1958). Enzymatic metabolism of drugs and other foreign compounds. *Annu. Rev. Biochem.*, *27* (3), 427–454.

Caldwell, J. (2006). Drug metabolism and pharmacogenetics: the British contribution to fields of international significance. *Br. J. Pharmacol.*, *147*, S89–S99.

Cantoni, G.L. (1953). S-Adenosylmethionine; a new intermediate formed enzymically from L-methionine and adenosine triphosphate. *J. Biol. Chem.*, *204*, 403–416.

Claude, A. (1940). Particulate components of normal and tumor cells. *Science*, *91* (2351), 77–78.

Claude, A. (1943). The constitution of protoplasm. *Science*, *97* (2525), 451–456.

Cohn, R. (1893). Concerning the occurrence of acetylated conjugates following the administration of aldehydes. *Z. Physiol. Chem.*, *17*, 274.

Conney, A.H., Buening, M.K., Pantuck, E.J., Pantuck, C.B., Fortner, J.G., Anderson, K.E., Kappas, A. (1980). Regulation of human drug metabolism by dietary factors. *Ciba Found. Symp.*, *76*, 147–167.

Conti, A., Bickel, M.H. (1977). History of drug metabolism: discoveries of the major pathways in the 19th century. *Drug Metab. Rev.*, *6* (1), 1–50.

Dessaignes, V. (1845). *C. R. Acad. Sci.*, *21*, 1224.

Dutton, G.J., Storey, I.D. (1954). Uridine compounds in glucuronic acid metabolism. I. The formation of glucuronides in liver suspensions. *Biochem. J.*, *57* (2), 275–283.

Erdmann, O.L. (1844). *J. Prakt. Chem.*, *33*, 190.

Erdmann, O.L., Marchand, R.F. (1842a). Metabolism of cinnamic acid to hippuric acid in animals. *Ann. Chem. Pharm.*, *44*, 344.

Erdmann, O.L., Marchand, R.F. (1842b). *J. Prakt. Chem.*, *26*, 491.

Hemingway, A., Pryde, J., Williams, R.T. (1934). The biochemistry and physiology of glucuronic acid: the site and mechanism of the formation of conjugated glucuronic acid. *Biochem. J.*, *28* (1), 136–142.

His, W. (1887). On the metabolic products of pyridine. *Arch. Ex. Pathol. Pharmakol. 22*, 253–260.

Jaffe, M. (1879). Ueber die nach einfuhrung von brombenzol und chlorbenzol im organismus entstehenden schwefelhaltigen sauren. *Ber. Deut. Chem. Ges.*, *12*, 1092.

Jaffe, M., Cohn, R. (1887). *Ber. Dtsch. Chem. Ges.*, *20*, 2311.

Karplus, J.P. (1893). *Z. Klin. Med.*, *22*, 210.

Keller, W. (1842). On the conversion of benzoic acid into hippuric acid. *Ann. Chem. Pharm.*, *43*, 108.

Klingenberg, M. (1958). Pigments of rat liver microsomes. *Arch. Biochem. Biophys.*, *75*, 376–386.

Lautemann, E. (1863). Concerning the reduction of quinic acid to benzoic acid and the conversion of the same to hippuric in animals. *Ann. Chem. Pharm.*, *125*, 9.

Lipmann, F. (1945). Acetylation of sulfanilamide by liver homogenates and extracts. *J. Biol. Chem.*, *160*, 173.

March, J. (1992). *Advanced Organic Chemistry*, 4th edn, John Wiley & Sons, Inc., New York, pp. 553–554, 700.

Mason, H.S. (1957a). Mechanisms of oxygen metabolism. *Adv. Enzymol. Relat. Sub. Biochem.*, *19*, 79–233.

Mason, H.S. (1957b). Mechanism of oxygen metabolism. *Science*, *125* (3259), 1185–1188.

Meyer, U.A. (2007). Endo-xenobiotic crosstalk and the regulation of cytochromes P450. *Drug Metab. Rev.*, *39* (2), 639–646.

Nencki, M. (1870). *du Bois-Reymond's Arch. Anat. Physiol.*, 399. As quoted in Conti, A. and Bickel, M.H. (1977). History of drug metabolism: discoveries of the major pathways in the 19th century. *Drug Metab. Rev.*, *6*, 1–50.

Netter, K.J. (1980). Inhibition of oxidative drug metabolism in microsomes. *Pharmacol. Ther.*, *10* (3), 515–535.

Neumeister, R. (1893). *Lehrbuch der physiologischen Chemie mit Berucksichtigung der pathologischen Verhaltnisse*, 2, 346, Gustav Fischer, Jena.

Omura, T., Sato, R. (1962). A new cytochrome in liver microsomes. *J. Biol. Chem.*, *237*, 1375–1376.

Omura, T., Sato, R., Cooper, D.Y., Rosenthal, O., Estabrook, R.W. (1965). Function of cytochrome P-450 of microsomes. *Fed. Proc.*, *24* (5), 1181–1189.

Orrenius, S., Thor, H., Jernström, B. (1980). The influence of inducers on drug-metabolizing enzyme activity and on formation of reactive drug metabolites in the liver. *Ciba Found. Symp.*, *76*, 25–42.

Potter, V.R., Elvehjem, C.A. (1936). The effect of selenium on cellular metabolism. The rate of oxygen uptake by living yeast in the presence of sodium selenite. *Biochem. J.*, *30* (2), 189–196.

Quick, A.J. (1932). The relationship between chemical structure and physiological response. II. The conjugation of hydroxy- and methoxy-benzoic acids. *J. Biol. Chem.*, *97*, 403.

Ritchie, J.C., Sloan, T.P., Idle, J.R., Smith, R.L. (1980). Toxicological implications of polymorphic drug metabolism. *Ciba Found. Symp.*, *76*, 219–244.

Schüller, J. (1925). Detoxication combinations within the body. *Arch. Exp. Pathol. Pharmakol.*, *106*, 265.

Schultzen, O., Naunyn, B. (1867). The behavior of benzene-derived hydrocarbons in the animal organism. *du Bois-Reymond's Arch Anat Physiol.*, 1867, 349.

Sheets, J.J. (2007). Ronald Estabrook's early guidance of a postdoctoral fellow concerning the intricacies of steroid metabolism by cytochromes P450. *Drug Metab. Rev.*, *39* (2–3), 281–283.

Shemin, D. (1946). The biological conversion of L-serine to glycine. *J. Biol. Chem.*, *162*, 297–307.

Sherwin, C.P. (1922). The fate of foreign organic compounds in the animal body. *Physiol. Rev.*, *2*, 238–276.

Staedeler, G. (1851). *Ann. Chem. Pharm.*, *77*, 17.

Ure, A. (1841). On gouty concretions; with a new method of treatment. *Pharm. J. Trans.*, *1*, 24.

Wang, P.P., Beaune, P., Kaminsky, L.S., Dannan, G.A., Kadlubar, F.F., Larrey, D., Guengerich, F.P. (1983). Purification and characterization of six cytochrome P-450 isozymes from human liver microsomes. *Biochemistry*, *22*, 5375–5383.

Williams, R.T. (1947). *Detoxication Mechanisms: The Metabolism of Drugs and Allied Organic Compounds*, Chapman and Hall, London.

Williams, R.T. (1959). *The Metabolism and Detoxication of Drugs, Toxic Substances, and Other Organic Compounds*, 2nd edn, Chapman and Hall, London.

Wöhler, F. (1824). *Tiedemann's Z. Physiol.*, *1*, 142.

Wöhler, F., Frerichs, F.T. (1848). Concerning the modifications which particular organic materials undergo in their transition to the urine. *Ann. Chem. Pharm.*, *65*, 335.

# Factors Affecting Metabolism

ROBERTA S. KING

Multiple pathways generally compete for metabolism of any particular compound. Thus, the relative amounts of each metabolite formed may be somewhat or even quite different from one species to another, one gender to the other, one age group to another, among otherwise apparently similar individuals, and even from one time point to another in a single individual. Furthermore, the proportion of each metabolite formed *in vivo* may be different than the proportion formed *in vitro*. Major influences causing these differences are availability of relevant enzyme systems, availability of relevant cofactors, presence of modulators, and transport properties of the parent and/or metabolites. Thus, in general, *in vitro* metabolism provides examples of the *possible* biotransformations of a particular agent, whereas *in vivo* metabolism is limited by the specific conditions of availability of enzyme, cofactor, modulators, and transport.

Investigation of the differences in metabolism among species, age groups, genders, polymorphic enzymes, and presence of modulators is applicable to both drug development and drug therapy. In development, knowledge of these differences in metabolism helps to correlate pharmacokinetic (PK), toxicology, and structure–effect studies from animal models to humans. During development and therapy, knowledge of these individual differences in metabolism helps to predict and understand differences in drug response and adverse effects. Because metabolism is an important criterion determining the rate of clearance of most drugs, differences in clearance can cause significant difference in individual dose–response and incidence of adverse effects. This is especially important for drugs with a *narrow therapeutic window*, that is, those having a small difference between minimum plasma concentration needed to reach a therapeutic effect and maximum plasma concentration to prevent adverse effects.

The major influences affecting metabolism can be categorized into genetic, environmental, and physiological factors, although the distinction between the categories is confounded by the fact that environmental and physiological factors can modify genetic susceptibilities. *Environmental factors* are thought of as temporal factors, with varying consequences according to their concentration over time. Environmental factors are usually limited to exogenous "small" molecules such as other drugs, compounds

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

naturally present in foods, and compounds otherwise absorbed through our environment. The environmental factors may act directly on an enzyme (as inhibitor or activator) or indirectly on the genetic regulatory system (as suppressor or inducer). *Genetic factors* are coded into one's genome. Genetic factors include variations among species and individuals in the regulatory (non-coded) and/or the transcribed (coded) region of each enzyme system. By definition, genetic factors are inherited (with allowances for natural selection of variants) and, therefore, should not change over one's lifetime. *Physiological factors* include those related to age, disease state, and gender (including hormonal cycling). A fourth category of factors influencing metabolism is currently emerging. This category is the interplay or cross talk among disease and endogenous and/or exogenous regulators of drug metabolism (Meyer, 2007).

Modulators such as enzyme inhibitors and enzyme inducers are considered environmental factors affecting metabolism. As described more fully in Chapter Characterization of Cytochrome P450 Mechanism-Based Inhibition, *enzyme inhibitors* reduce formation of a particular metabolite, allowing buildup of the parent and/or a shifting of biotransformation toward other possible metabolites. Inhibitors bind directly to the enzyme. Three main types of inhibitors are relevant to drug metabolism. *Competitive* (reversible) inhibitors bind directly at the active site and competitively block access of substrate to the active site. *Noncompetitive* (reversible) inhibitors bind outside the active site not blocking access of the substrate, yet decrease the rate of catalysis. *Irreversible* inhibitors bind covalently to the enzyme, and activity can be regenerated only by biosyntheses of more enzyme. *Enzyme activators* are analogous to noncompetitive inhibitors, but have a stimulating effect on enzyme activity. Cytochrome P450 3A4 is especially susceptible to drug–drug interactions via inhibition and activation, although all cytochrome P450s have known inhibitors. As described more fully in Chapter Clinical Drug–Drug Interactions, *enzyme inducers* enhance gene transcription of particular enzyme systems, increasing the amount of that particular enzyme available, thus increasing the proportion of metabolism catalyzed by that particular enzyme. Enzyme inducers bind to one or more transcription factors or promoter elements in the DNA region that regulates expression of the gene. This causes increased transcription of the gene, resulting in more mRNA, and more protein synthesized. The effect of enzyme inducers lasts several days, depending on the degradation rate of the enzyme rather than on clearance of the inducer. *Enzyme suppressors* have the opposite effect from inducers, decreasing transcription and protein expression. Cytochrome P450 3A4 is especially susceptible to drug–drug interactions via the induction/suppression mechanism.

An often overlooked, yet important, factor affecting metabolism is cofactor supply. For each transformation, a specific cofactor is required to supply the source of the reactive transfer moiety. For example, oxidative reactions catalyzed by cytochrome P450 need a ready supply of NAD(P)H. Transferases need a ready supply of the transfer moiety such as uridine-5′-diphosphate glucuronic acid (UDPGA) for glucuronidation or 3′-phosphoadenosine-5′-phosphosulfate (PAPS) for sulfonation. Each cofactor has a limited cellular supply and regeneration system. For example, it is generally accepted that sulfotransferase (SULT) reactions are limited by depletion and slow regeneration of PAPS supply, whereas UDPGA supply for glucuronidation is abundant and readily regenerated. Genetic and environmental factors also modify availability of cofactors via inhibition, activation, induction, or suppression of the enzymes controlling regeneration of each cofactor.

*Physiological factors* affecting drug metabolism include those related to age, disease state, and gender (including hormonal cycling). Age differences are typically observed

in the very young or in the elderly. For example, fetal tissue and newborns lack significant expression of all drug-metabolizing enzymes except CYP3A7 and SULT (Hines, 2007). Other enzyme systems develop over days, weeks, and months. In the elderly, total metabolism is often reduced due to lower hepatic blood flow and diminished activity of nearly all metabolism enzymes. Hepatic insufficiency caused by one or more of a number of diseases also reduces hepatic metabolism. Similarly, renal insufficiency generally reduces renal metabolism. More complicated regulatory mechanisms control the differences in metabolism seen in hormonal cycling and possibly also in endocrine disorders. Evidence is emerging for significant regulatory cross talk among endogenous and exogenous regulators of drug metabolism (Meyer, 2007).

Genetic differences are responsible for many well-known differences in metabolism among different individuals. When a trait has differential expression in >1% of the population, the trait is considered to have a *genetic polymorphism*. These differences may be within coded and/or non-coded regions of the gene for each enzyme member. Of genetic differences affecting drug metabolism, the cytochrome P450 system is considered of greatest importance. Indeed, genetic screens for CYP2D6 and CYP2C19 status (Roche Diagnostics, 2008), and CYP2C9 status (Nanosphere, 2008) have recently been approved by the FDA (US Food and Drug Administration, 2005, 2008a). These tests are currently optional to assist in dosing of, for example, tricyclic antidepressants and antipsychotics (de Leon, Armstrong and Cozza, 2006), β-blockers, antiepileptics, proton pump inhibitors, and warfarin (US Food and Drug Administration, 2008b).

Genetic polymorphisms for three additional drug metabolism systems are currently recognized as clinically relevant: thiopurine methyltransferase (TPMT), *N*-acetyltransferase 2 (NAT2), and UDP-glucuronosyltransferase (UGT1A1). TPMT catalyzes degradation of thiopurine anticancer agents (mercaptopurine, thioguanine). Individuals who are homozygous-deficient for TPMT have little or no enzyme activity and accumulate toxic levels of the corresponding antimetabolites. While only 0.3% of patients are affected (Woodson, Dunnette and Weinshilboum, 1982), avoidance of life-threatening bone marrow suppression requires substantial dose reduction in these patients (Relling *et al.*, 1999). In contrast, up to 50% of patient populations have a NAT2 genetic polymorphism causing "slow" metabolism and increased incidence of adverse effects from isoniazid (Evans, 1989). A common allele for decreased UGT1A1 activity is present in up to 30% of patient populations, and can cause increased adverse effects after irinotecan treatment (Ando *et al.*, 2000). While the adverse effects from the isoniazid and irinotecan treatments are not life-threatening to individuals with low-activity genotypes, lower dosing can impart quality of life benefits while retaining effectiveness of the treatment.

## REFERENCES

Ando, Y., Saka, H., Ando, M., Sawa, T., Muro, K., Ueoka, H., Yokoyama, H., Saitoh, S., Shimokata, K., Hasegawa, Y. (2000). Polymorphisms of UDP-glucuronosyltransferase gene and irinotecan toxicity: a pharmacogenetic analysis. *Cancer Res.*, *60* (24), 6921–6926.

Evans, D.A. (1989). *N*-acetyltransferase. *Pharmacol. Ther.*, *42*, 157–234.

de Leon, J., Armstrong, S.C., Cozza, K.L. (2006). Clinical guidelines for psychiatrists for the use of pharmacogenetic testing for CYP450 2D6 and CYP450 2C19. *Psychosomatics*, *47* (1), 75–85.

Hines, R.N. (2007). Ontogeny of human hepatic cytochromes P450. *J. Biochem. Mol. Toxicol.*, *21* (4), 169–175.

Meyer, U.A. (2007). Endo-xenobiotic crosstalk and the regulation of cytochromes P450. *Drug Metab. Rev.*, *39* (2), 639–646.

Nanosphere (2008). *Verigene Warfarin Metabolism Nucleic Acid Test*, http://www.nanosphere.us/VerigeneWarfarinMetabolismNucleicAcidTest_4472.aspx (accessed 12 February 2008).

Relling, M.V., Hancock, M.L., Rivera, G.K., Sandlund, J.T., Ribeiro, R.J., Krynetski, E.Y., Pui, C.H., Evans, W.E. (1999). Mercaptopurine therapy intolerance and heterozygosity at the thiopurine S-methyltransferase gene locus. *J. Natl. Cancer Inst.*, *91* (23), 2001–2008.

Roche Diagnostics (2008). *Background Information: AmpliChip CYP450*, http://www.roche.com/med_backgr-ampli.htm (updated 21 January 2008, accessed 12 February 2008).

US Food and Drug Administration (2005). US FDA approved the Roche AmpliChip™ CYP450 Array for the rapid genotyping of both *CYP2C19* and *CYP2D6* variants for diagnostic use. *Fed. Regist.*, *70* (46), 11865–11867.

US Food and Drug Administration (2008a). *FDA News, FDA Clears Genetic Lab Test for Warfarin Sensitivity*, http://www.fda.gov/bbs/topics/NEWS/2007/NEW01701.html (release date: 17 September 2007, accessed 12 February 2008).

US Food and Drug Administration (2008b). *Center for Drug Evaluation and Research, Table of Valid Genomic Biomarkers in the Context of Approved Drug Labels*, http://www.fda.gov/cder/genomics/genomic_biomarkers_table.htm (date created: 15 September 2006, updated 29 January 2008, accessed 12 February 2008).

Woodson, L.C., Dunnette, J.H., Weinshilboum, R.M. (1982). Pharmacogenetics of human thiopurine methyltransferase: kidney-erythrocyte correlation and immunotitration studies. *J. Pharmacol. Exp. Ther.*, *222* (1), 174–181.

1

# Biotransformations in Drug Metabolism

ROBERTA S. KING

## 1 DRUG METABOLISM IN DRUG DEVELOPMENT AND DRUG THERAPY

The major function of drug metabolism is to facilitate removal of compounds from the organism, thereby preventing unwanted accumulation of foreign compounds or of potentially toxic levels of endogenous compounds. This physiological function of metabolism is quantified using two pharmacokinetic terms, *bioavailability* and *clearance*. Bioavailability indicates the percentage of administered drug that reaches the systemic circulation, and is indicated by percent absorbed minus pre-systemic metabolism. Clearance indicates the rate of removal of the drug from the systemic circulation, and is indicated by the rates of post-systemic metabolism and excretion. Bioavailability and clearance combine to affect the total drug exposure which is usually quantified by *area under the concentration-versus-time curve* (AUC). These relationships are clarified in Fig. 1.

In addition to clearance, metabolism can affect the spectrum of consequences of drug action including both the desired therapeutic effects and undesired effects. During the early stages of drug development, metabolism-based strategies allow scientists to improve the pharmacokinetic properties of a lead compound by blocking or inserting sites of facile metabolism. During the clinical studies and post-marketing stages, metabolism data can be used to promote safe use of a drug, including dosing adjustments or warnings about drug–drug interactions. The most common fate of metabolism is the formation of *inactive products*—metabolites having no physiological effect (desired or undesired). In the case of inactive products, the rate of clearance would be equal to the loss of parent drug. If the drug is converted to inactive metabolites pre-systemically or too quickly for the desired action, then analogues can be developed with these sites of metabolism blocked, thereby slowing metabolic clearance. A second possible fate of drug metabolism is the formation of *active metabolites*—those having the desired activity to an extent similar to (or greater than) the parent drug. In the case of active metabolites (called "prodrugs" under certain circumstances), clearance of all active

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

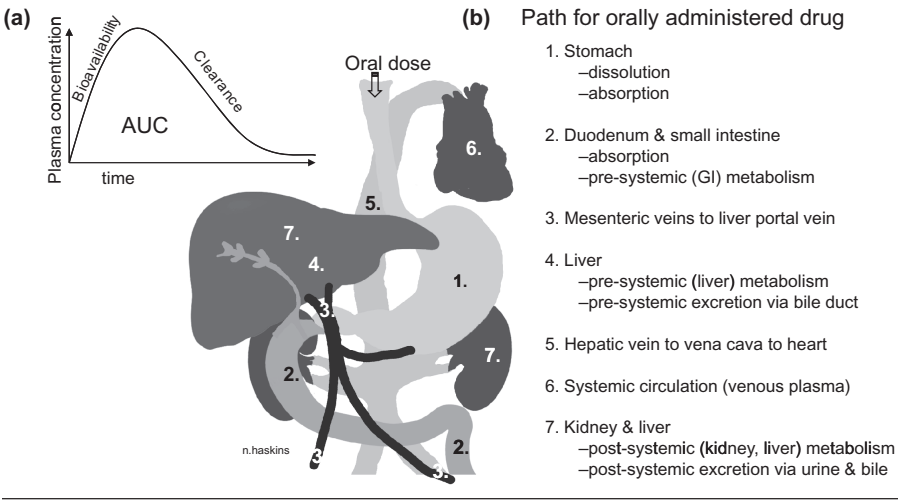


**Figure 1** Role of drug metabolism in bioavailability and clearance.

metabolites needs to be considered. A third possible fate of drug metabolism is the formation of *toxic metabolites*. Toxic metabolites are undesirable, and analogues can be developed with these sites blocked to prevent formation. A fourth possible fate of drug metabolism is the formation of metabolites with physiological action unrelated to the desired mechanism of action. These sites can be blocked, or the metabolite can be considered a lead compound to achieve the unrelated action. For example, an early antihistamine, promethazine, has been successfully used for its anticholinergic (antiemetic) activity. Table 1 summarizes the effective uses of metabolism science in drug development and drug therapy.

The most famous example of the importance of metabolism science to drug therapy and drug development, published in 1990 (Monahan *et al.*, 1990), involved a young woman presenting to the emergency room with a potentially lethal arrhythmia (palpitations, syncope, *torsades de pointes*). After the woman was treated successfully with the standard of care for arrhythmia, the investigation began to shed light on the cause. The patient's drug history showed that 10 days prior to admission, she was prescribed terfenadine (Seldane®—an antihistamine) and cefaclor (Ceclor®—a cephalosporin antibiotic). On the eighth day of terfenadine therapy, the patient began a self-medicated course of ketoconazole (Nizoral®—an azole antifungal) for vaginal candidiasis. She was also taking medroxyprogesterone acetate, although this fact is irrelevant. Existing knowledge led to the prediction that ketoconazole, a known potent cytochrome P450 (CYP) inhibitor, decreased the metabolism-based clearance of terfenadine leading to a high plasma concentration of terfenadine (Fig. 2). At these relatively high concentrations, an unrelated effect of terfenadine was observed—inhibition of potassium channels in the heart leading to impaired control of heart rate. For a time, terfenadine was marketed with strict warnings about coadministration of certain drugs. However, instead of giving up on their investment or being satisfied with a risky prescribing scenario, scientists at Hoechst Marion Roussel (currently Sanofi-Aventis) effectively used

**TABLE 1 Effective uses of metabolism science in drug development and drug therapy.**

| Development | Therapy |
|---|---|
| Find therapeutically active metabolites | Major component controlling duration of action |
| Find toxic or reactive metabolites | Rate of "clearance by metabolism" |
| Alter rate of metabolic clearance | Prodrugs, active metabolites, toxic metabolites |
| Prodrug strategy | Individual variation in response to drug therapy |
| Predict potential drug interactions | Genetic differences |
| Predict individual variation in response | Induction |
| Understand similarity of humans to animals for PK, toxicology, structure–effect studies | Inhibition |
| | Physiological state differences |

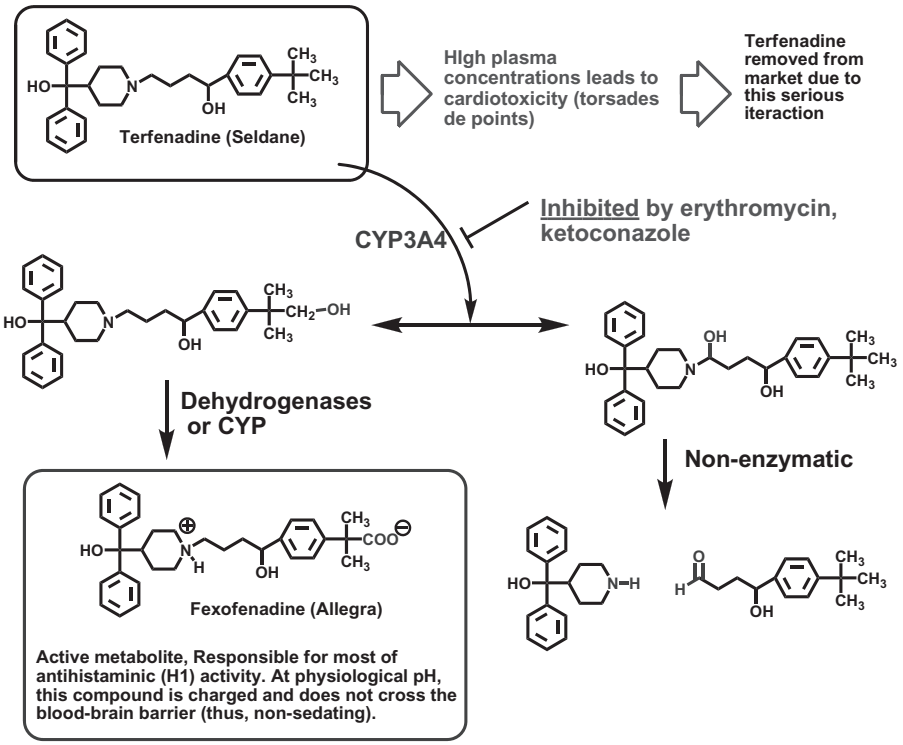


**Figure 2** Example of effective use of metabolism science: fexofenadine (Allegra) as a superior replacement for terfenadine (Seldane).

metabolism science and developed the highly successful fexofenadine (Allegra) as a superior replacement (Fig. 2). This exemplifies an effective case of taking a "lemon" (problematic product) and producing "lemonade" (a highly successful derivative product).

## 2 PREDICTION OF METABOLITE AND RESPONSIBLE ENZYME

As a professor of pharmacy, I often hear the question, "How does one predict exactly what metabolites will be formed?" Well, it is pretty easy to predict the metabolites that *may potentially* form and which enzymes *may be* responsible for their formation. Recently published computational models report 80–90% prediction accuracy for the three major CYP members (Sheridan *et al.*, 2007; Terfloth, Bienfait and Gasteiger, 2007), but it is more difficult to predict in a given individual the exact proportions of each potential metabolite. I also hear the related question, "How does one know whether a particular metabolite will be clinically active or inactive, or whether it will cause an adverse affect?" Again, predictions are somewhat accurate for common functional groups and well-studied groups of drug molecules, yet the predictions are best confirmed experimentally.

A curated and well-maintained source for specific links to original literature for CYP substrates, inhibitors, and inducers is available to the public at http://medicine.iupui.edu/flockhart/table.htm. Two versions of the table are maintained: one relevant for *in vitro* metabolism, and the other relevant for real clinical situations under expected dose ranges. In addition, through the author's work at the University of Rhode Island, a comprehensive publicly available database has been developed which focuses on both phase I and phase II metabolites formed, enzymes responsible, and potential for individual variation. The data represented have been gathered through course projects by students of the College of Pharmacy and currently include many top 100 prescribed drugs (2005 list) and additional drugs used for the treatment of cardiovascular and renal diseases (http://www.uri.edu/pharmacy/departments/bps/faculty/king.html).

Of course, the major goal of the pharmaceutical industry is to make useful medications for human and veterinary use. While humans, agricultural animals, and pets are the major focus of these medications, obvious ethical limitations require much research to be conducted *in silico, in vitro*, or limited to laboratory animal species. Thus, there will always be a need to correlate data from several experimental approaches including comparing *in vitro* studies with *in vivo*, various animal species with humans, among humans of diverse genetic backgrounds, and in the presence of potential modifiers.

As a foundation, one needs to understand two categories of information about metabolism: first, the transformations and, second, the enzymes that catalyze these transformations. If one learns the transformations possible for each type of functional group, then possible metabolites can be predicted for any structure. If one also learns the enzyme systems and cofactors present (or lacking) in each species or tissue type, then the prediction can be narrowed further. Lastly, if one learns the major modifiers of metabolism and which pathways are susceptible to modification, then complete predictions can be made. Neither category is more important than the others, although most texts emphasize only one at the expense of the others, depending on whether they are written from the view of the chemist or of the enzymologist.

## 3 FUNCTIONAL GROUP BIOTRANSFORMATIONS: PHASE I, PHASE II, CATALYSIS

*In vivo*, each compound is typically metabolized through a number of parallel and/or sequential reactions, a so-called *metabolism pathway*, and the relative amounts of each

metabolite formed may be somewhat or even quite different from one situation to another. The total metabolism of any compound is controlled by the total of all pathways available to metabolize the drug and is modified by the specific conditions of availability of enzyme, cofactor, modulators, and transport. As early as 1947, R.T. Williams proposed that metabolism occurred through reactions representing two distinct phases, "phase I" and "phase II" metabolism reactions (Williams, 1947). These terms are still popular, although we must be aware that the order is not exclusive (phase I not always followed by phase II; phase II not always preceded by phase I). *Phase I* reactions are most commonly described as "functionalization" reactions and include oxidations, reductions, and hydrolyses. These reactions (Fig. 3) introduce a new polar functional group to the parent drug (oxidations), modify an existing functional group to be more polar (reductions), or unmask existing polar functional groups (hydrolyses). The most common functional groups exposed or introduced in the phase I reactions are hydroxyl (–OH), amino (–NH–), and carboxylic acid (–COOH).

*Phase II* reactions are most commonly described as *conjugation* reactions, and include glucuronidation, sulfonation, glycine/glutamine conjugation, acetylation, methylation, and glutathione conjugation (Fig. 4). These are substitution-type reactions and they link a new group either to the parent drug or to a phase I metabolite. Some conjugations cause a dramatic increase in the polarity, and thus excretion, of a drug by

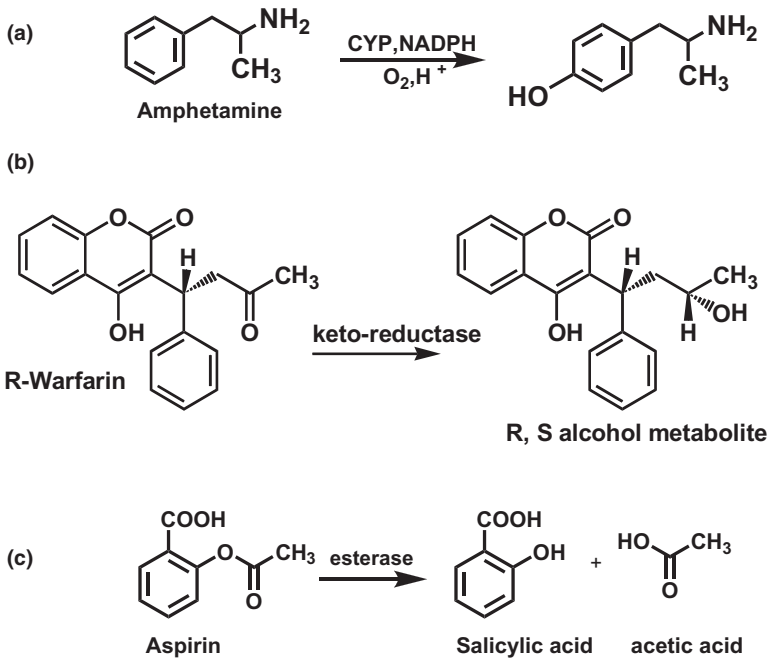


**Figure 3** Examples of phase I reactions. **(a)** Introduce an –OH by oxidation of a carbon atom (aromatic hydroxylation). **(b)** Modify existing group to an –OH by reduction (ketone reduction). **(c)** Expose an existing –OH by hydrolysis (ester hydrolysis).

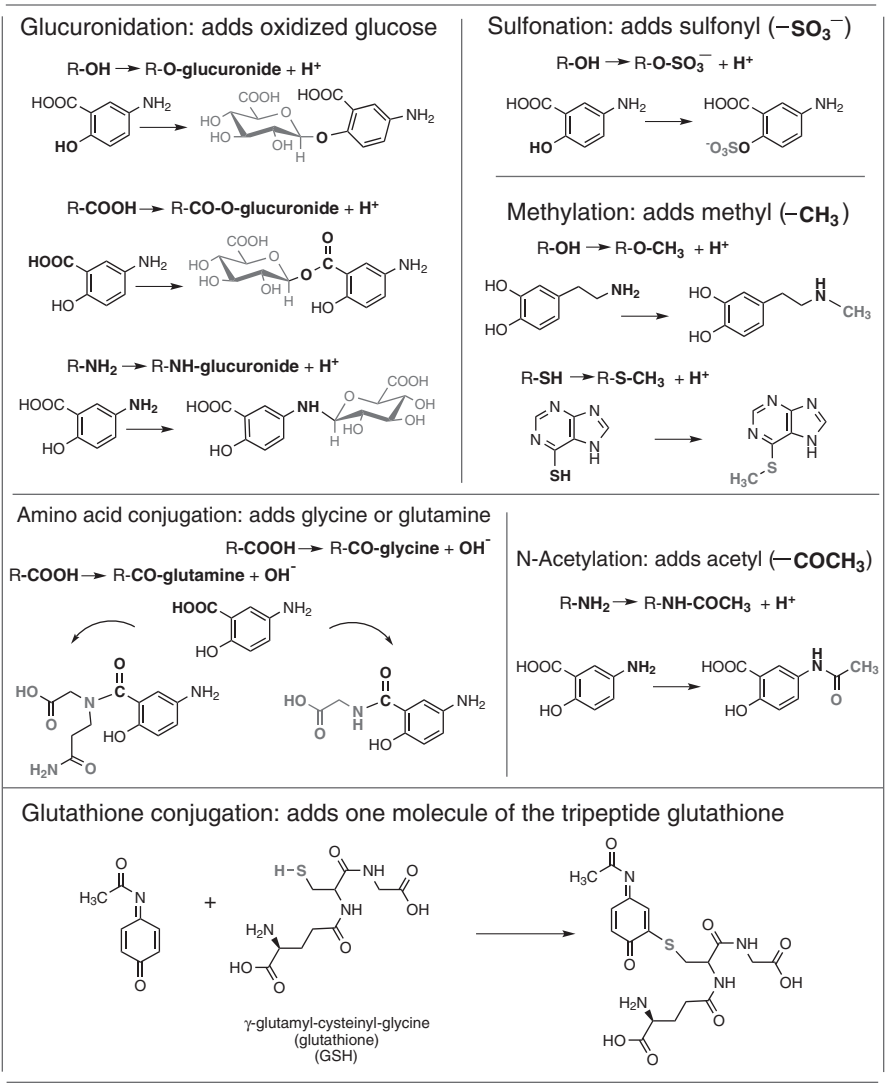


**Figure 4** Examples of phase II reactions.

adding an ionized functional group: sulfonation, glucuronidation, and amino acid conjugation. Other conjugation reactions are just likely to cause termination of therapeutic activity: methylation and acetylation. One type of conjugation reaction protects against reactive metabolites: glutathione conjugation.

Of note is that all of these biotransformations occur under very mild "physiological" conditions: typically neutral pH, approximately 37°C, aqueous solvent, and atmospheric pressure. Only with enzymes as catalysts can these reactions occur under such mild conditions. Every enzyme acts to increase a rate of reaction by a combination of two components, approximation and catalysis. *Approximation* increases the local concen-

tration of the reactants and facilitates the proper three-dimensional orientation of the reactants. This occurs by non-covalent enzyme–ligand interactions such as hydrogen bonds, complementary charge–charge interactions, and attractive lipophilic forces. *Catalysis* serves to lower the activation energy of the reaction. Most enzyme systems are able to stabilize the transition state (and destabilize the reactants) by non-covalent molecular forces such as hydrogen bonding, charge–charge interactions, and dipole interactions. The oxidation/reduction reactions are catalyzed by assisting the electron transport and in the formation of the reactive oxygen moiety. The rate of metabolism by any one enzyme molecule is controlled by the number of product molecules formed per active site per minute, which is called the "turnover rate" of each active site. The rate of metabolism contributed by any one enzyme member (i.e. CYP3A4) is the turnover rate multiplied by the total number of active sites of that member (dependent on CYP3A4 protein expression). And finally, the rate of total metabolism of an administered drug is controlled by the total of all pathways and enzymes available to metabolize the drug.

## 4 OXIDATIONS AND CYTOCHROME P450

Oxidations comprise the bulk of phase I metabolism reactions, and members of the CYP *superfamily* catalyze the majority of these drug oxidation reactions. Of the nearly 60 human CYP members currently identified (http://drnelson.utmem.edu/CytochromeP450.html), only about 10 contribute significantly to metabolism of marketed drugs (Fig. 5). Individual *members* are coded by separate genes, and the members are grouped and named according to amino acid sequence similarity of the gene product. Within each CYP *gene family* (CYP1, CYP2, and CYP3), the members have more than 40% sequence identity. Each family is divided into subfamilies indicated

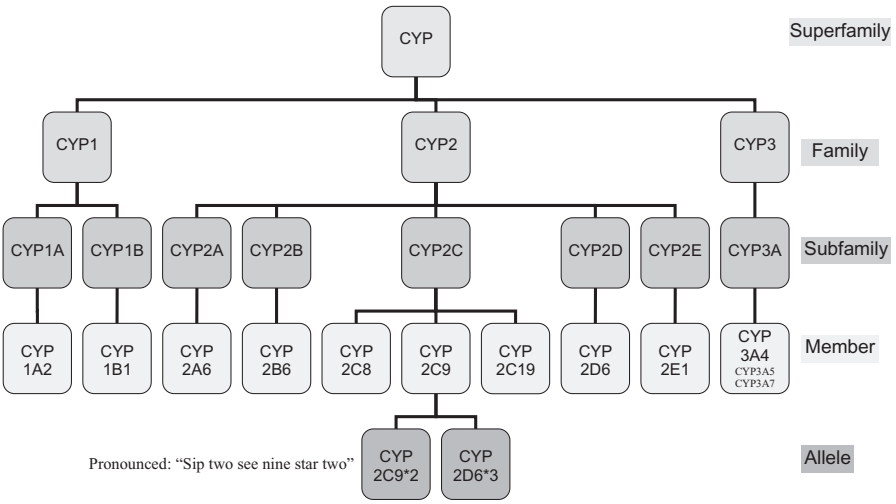


**Figure 5** Relationships and nomenclature of cytochrome P450 members important for human drug metabolism.

**TABLE 2 Major P450 reactions classified according to category and mechanism.**[a]

| Category | Mechanism | Biotransformation |
|---|---|---|
| A. $sp^3$ Carbon hydroxylation | i. Hydrogen atom abstraction<br>ii. Oxygen rebound | 1. Benzylic carbon hydroxylation[b]<br>2. Allylic carbon hydroxylation<br>3. Aliphatic carbon hydroxylation<br>a. end of chain (ω)<br>b. penultimate (ω-1)<br>c. cyclic<br>4. Alcohol/aldehyde carbon oxidation |
| B. $sp^2$ Epoxidation | i. Hydrogen atom abstraction<br>ii. Oxygen rebound | 5. Alkene epoxidation<br>6. Aromatic hydroxylation[b] |
| C. Heteroatom release | i. Nonbonded electron abstraction<br>ii. Hydrogen atom abstraction<br>iii. Oxygen rebound<br>iv. Nonenzymatic rearrangement | 7. *N*-dealkylation[b]/deamination<br>8. *O*-dealkylation[b]<br>9. Oxidative cleavage |
| D. Heteroatom oxygenation | i. Nonbonded electron abstraction<br>ii. Nonbonded electron abstraction<br>iii. Oxygen rebound | 10. Nitrogen oxidation<br>a. Primary amine<br>b. Secondary amine<br>c. Tertiary amine<br>11. Sulfur oxidation<br>a. Thiol (sulfide)<br>b. Sulfoxide |

[a]From Guengerich (2001).
[b]Indicates the four most common cytochrome P450 oxidation reactions in human drug metabolism.

by a capital letter (CYP2C, CYP2D, CYP3A). Each *subfamily* contains members with more than 55% amino acid sequence identity (CYP2C8, CYP2C9, CYP2C19). That brings us to the member level, and remember that members are defined as being coded by separate genes. Members can be further subdivided into allele or variant categories. *Alleles* are responsible for hereditary variation and are defined as alternative versions of a gene resulting from mutation or duplication/deletion events. Each person has two separately inherited alleles that together make up the individual's genotype for that gene. *Genetic polymorphism* is used to describe genes that contain an alternative allele present in >1% of a given population. For example, CYP2C9 includes a wild-type allele (CYP2C9*1) and several alternative alleles (CYP2C9*2, CYP2C9*3). Although not applicable to the CYP superfamily, some enzyme superfamilies include member genes that encode multiple sequence *variants* (commonly by variable splicing). These variants are indicated by appending "_v#" to the member name, such as SULT2B1_v1 and SULT2B1_v2.

There are at least 11 different common oxidation reactions catalyzed by CYP (and many additional less common reactions) (Table 2). How can one enzyme family catalyze so many different reactions? The easiest way to rationalize all these different possible oxidations is by thinking about the CYP catalytic mechanism (Fig. 6). The major reactions can be understood using a single reactive oxygen moiety ($FeO^{3+}$) and only four mechanistic subcategories (Table 2): (A) $sp^3$ carbon hydroxylation, (B) $sp^2$

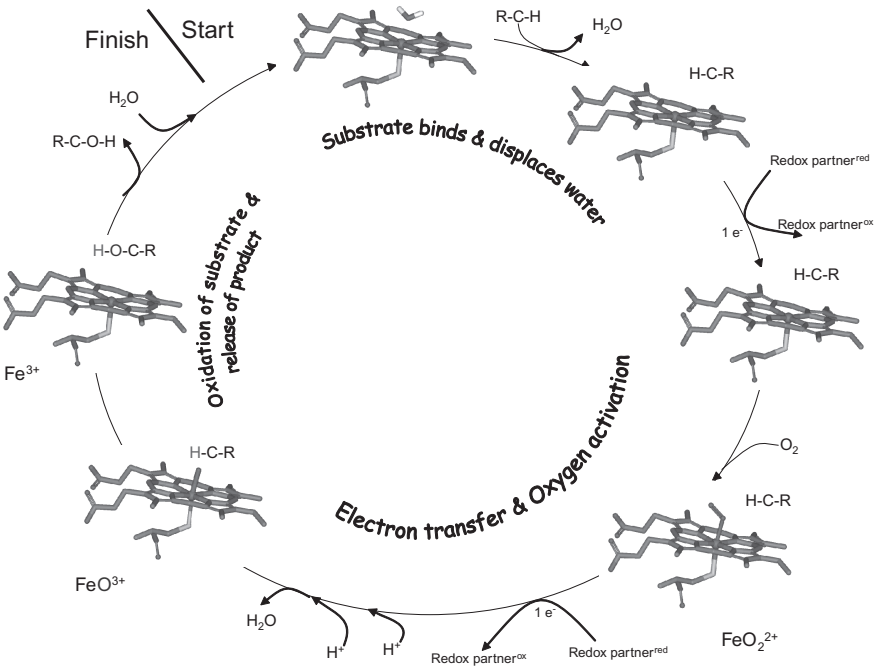


**Figure 6** P450 catalytic mechanism cycle. First, substrate binds by displacing water molecule from heme iron. Second, electron transfer steps and oxygen activation occur. Third, the substrate is oxidized and the product is released.

carbon epoxidation, (C) heteroatom release, and (D) heteroatom oxygenation. This "unified" view of the major P450 reactions was originally described by Guengerich in 1984 (Guengerich and McDonald, 1984; Guengerich, 2001). *Regioselectivity*, which is defined as the atom or atoms oxidized in a molecule when multiple sites are possible, is partially controlled by the ease of hydrogen atom abstraction and the stability of the resulting carbon radical. Mechanism, however, does not singularly control the regioselectivity. Each CYP superfamily member has a slightly different three-dimensional structure of the substrate-binding pocket, each of which exhibits unique structural preferences for substrate-binding affinity and orientation toward the reactive oxygen. Thus, mechanistic and structural features combine to define the substrate selectivity and regioselectivity of each CYP. These features are described in more detail in Section 5 of this chapter.

Because of the combination of mechanistic and structural features controlling CYP oxidations, the probability of forming each potential product is a combination of three criteria including abundance of enzyme isoform catalyst, abundance of each substrate functional group, and relative reactivity of each substrate functional group. Because of its abundance in the liver and GI tract, CYP3A4 catalyzes the largest fraction of human drug metabolism (Fig. 7) (Shimada *et al.*, 1994). Thus, the pool of total formed metabolites is enriched with functional groups oxidized by CYP3A4 within substrate structures preferred by the CYP3A4-binding pocket. CYP2D6, in contrast, only makes up only a

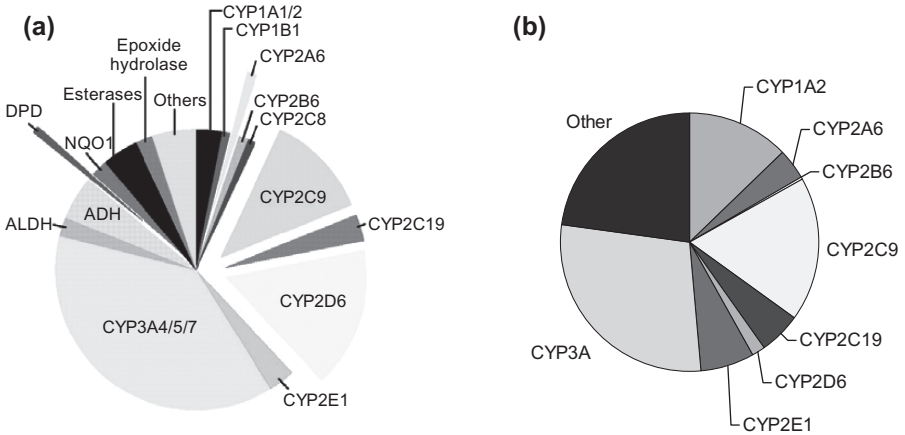


**Figure 7** **(a)** Proportion of all marketed drugs metabolized by each phase I enzyme. Enzyme polymorphisms associated with changes in drug effects are shown offset from the corresponding pie charts (reprinted with permission from Evans and Relling, 1999). **(b)** Average relative proportions of CYPs present in human liver (Shimada *et al.*, 1994; Emoto *et al.*, 2006).

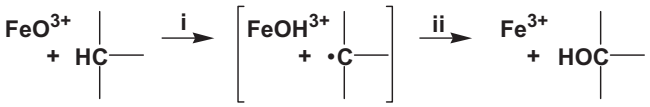


**Figure 8** Carbon hydroxylation under the unified mechanism. (i) H• abstraction from an $sp^3$ carbon; and (ii) oxygen rebound (Guengerich, 2001).

very small percentage of the total human liver P450, yet oxidizes approximately 25% of all drug molecules (Fig. 7) (Shimada *et al.*, 1994; Evans and Relling, 1999). This contradiction occurs because CYP2D6-preferred substrates are highly represented within the pool of all therapeutic drug molecules. Lastly, certain functional groups are more readily oxidized than others due to the stability of the carbon radical or ease of nonbonded electron abstraction. For example, the pool of CYP-oxidized metabolites is enriched in benzylic alcohols due to the relative stability of benzylic carbon radicals. As a result of these combined structural and mechanistic features, four oxidation reactions are most commonly observed in human drug metabolism: (i) aromatic hydroxylation, (ii) *N*-dealkylation, (iii) *O*-dealkylation, and (iv) benzylic carbon hydroxylation.

### 4.1 CYP Carbon Hydroxylation Reactions

$sp^3$ Carbon hydroxylation by CYP is very common and can be used to exemplify the unified mechanism proposed by Guengerich (Guengerich, 2001). This mechanism considers the heme iron to be in the oxyferryl state ($FeO^{3+}$) for the oxidation step (Fig. 6). As shown in Fig. 8, carbon hydroxylation occurs in two steps: (i) H• abstraction

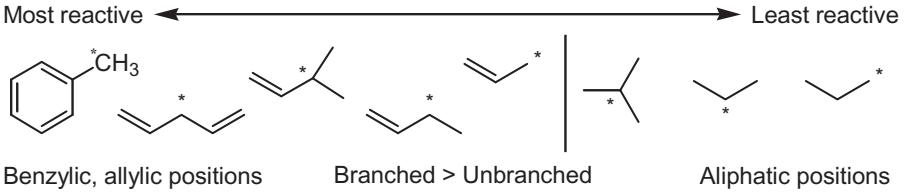


**Figure 9** Order of reactivity of $sp^3$ carbons toward cytochrome P450 oxidation. Asterisk (*) indicates position of carbon radical in each type of functional group.

from an $sp^3$ carbon, and (ii) oxygen rebound. Therefore, the oxidized carbon must have a C–H bond. In the absence of enzyme conformational limitations, the most easily oxidized carbon will be one with the most easily abstracted hydrogen atom, although rearrangements may occur to provide the most stable carbon radical (Fig. 8).

Multiple $sp^3$ carbon atom types are hydroxylated by CYP, including benzylic, allylic, and aliphatic carbons (Table 2). Benzylic positions are the most reactive, followed by allylic, followed by aliphatic (Fig. 9). Indeed, benzylic carbon hydroxylation is one of the four most commonly observed CYP reactions. Within each group, branched carbon positions are more reactive than unbranched ones. While relatively unreactive, the terminal methyls of tertiary-butyl and isopropyl groups are commonly oxidized by CYP because of the statistical probability (9x, 6x) of having any one terminal C–H bond oxidized. As discussed in a later section, steric hindrance around the heme $FeO^{3+}$ limits hydroxylations of sterically hindered carbons, which encourages hydroxylation at the terminal (ω) and penultimate (ω-1) positions of aliphatic chains. Occasionally aliphatic 6-membered rings are oxidized to the least sterically hindered product (usually 4-*trans* product). Examples of benzylic and allylic carbon hydroxylation are shown in Fig. 10. Examples of aliphatic carbon hydroxylation are shown in Fig. 11.

Alcohol and aldehyde oxidation can also be catalyzed by CYP via $sp^3$ carbon hydroxylation (Guengerich, 2001). While these oxidations are often considered products of alcohol dehydrogenase and aldehyde dehydrogenase, respectively, CYP can also contribute to their oxidation. Figure 12 shows the unified mechanism for oxidation of ethanol and acetaldehyde by CYP via carbon hydroxylation. Specifically, ethanol and acetaldehyde are known to be very good substrates for CYP2E1.

### 4.2 CYP Epoxidation Reactions

Oxidation of $sp^2$ carbons results in formation of an epoxide or arene oxide (Fig. 13). In most cases, the epoxides do not survive to excretion, but are hydrolyzed by epoxide hydrolase to a *trans*-diol. Carbamazepine is a well-known exception to this rule: a significant portion of the epoxide is stable enough for renal excretion (Fig. 14). In the case of the arene oxide, the return to aromaticity drives a rearrangement reaction which produces the phenol. The complete conversion to the phenol is called aromatic hydroxylation and is one of the four most common oxidations in drug metabolism (Fig. 13). In general, the order of preference for aromatic hydroxylation position is *para* > *ortho* >> *meta*. When two substituents are present, the least sterically hindered product is generally favored (Fig. 14).

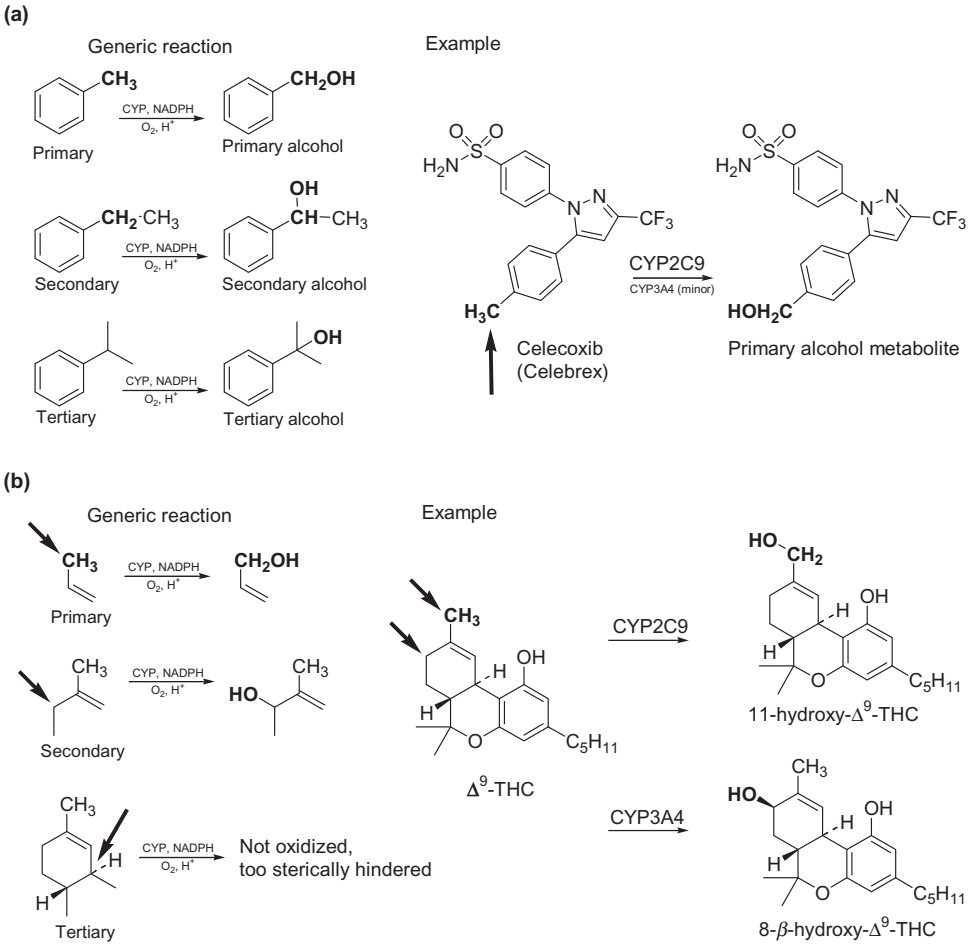


**Figure 10** Examples of carbon hydroxylation: **(a)** benzylic carbon hydroxylation (Paulson *et al.*, 2000; Tang *et al.*, 2000), and **(b)** allylic carbon hydroxylation (Bornheim *et al.*, 1992; Lemberger 1973).

## 4.3 CYP Heteroatom Release Reactions

The remaining two of the four most commonly observed P450 drug oxidations result in the reactions called *N*-dealkylation and *O*-dealkylation. The P450 enzyme catalyzes oxygen insertion on a C–H which is bonded to an amine nitrogen (or ether oxygen). The resulting carbinolamine (or hemiacetal) is not stable and can spontaneously degrade, releasing the aldehyde/ketone and amine (or alcohol). Thus, the bond between the heteroatom and the carbon is broken. As shown in Fig. 15, the reaction on amines can be rationalized using four steps: (i) one-electron transfer, (ii) proton abstraction, (iii) oxygen rebound, and (iv) nonenzymatic rearrangement. In general, the most commonly *N*-dealkylated groups are methyl, ethyl, *n*-propyl, isopropyl, allyl, and

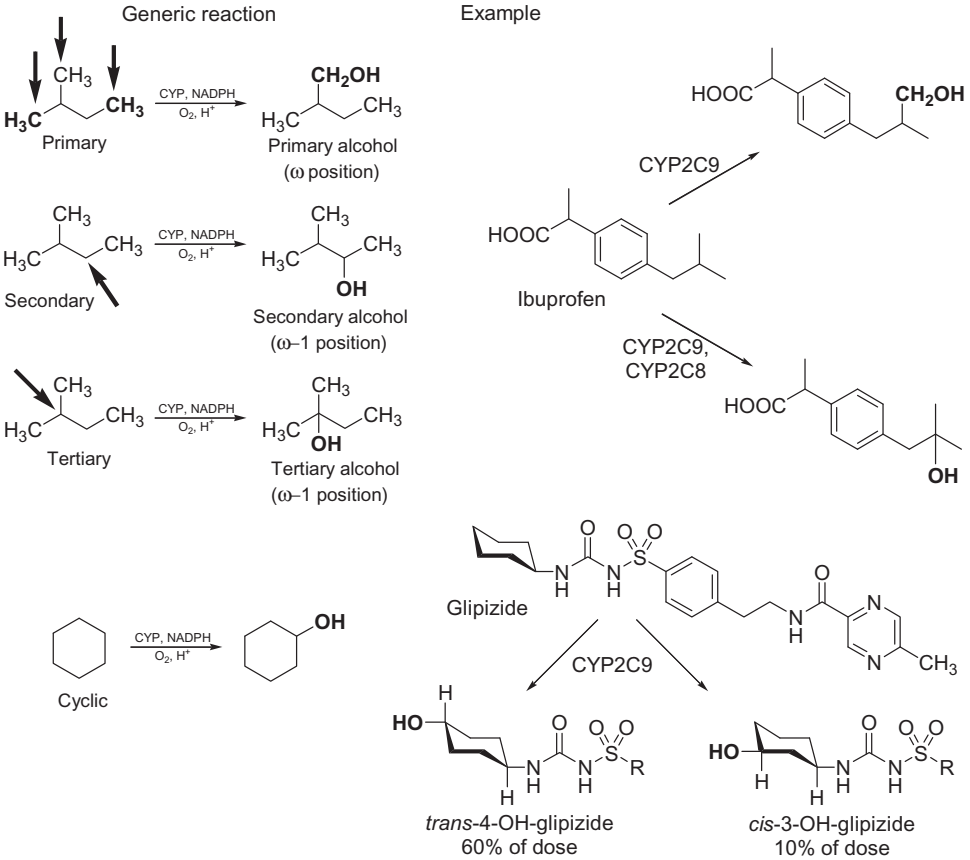


**Figure 11** Examples of aliphatic carbon hydroxylation (Fuccella *et al.*, 1973; Hamman *et al.*, 1997; Rudy *et al.*, 1991).

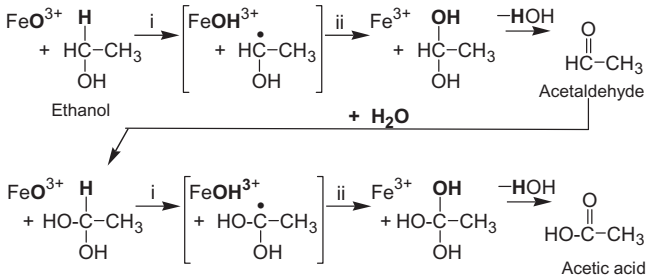


**Figure 12** Mechanism for oxidation of ethanol and hydrated acetaldehyde by CYP2E1 (Guengerich, 2001).

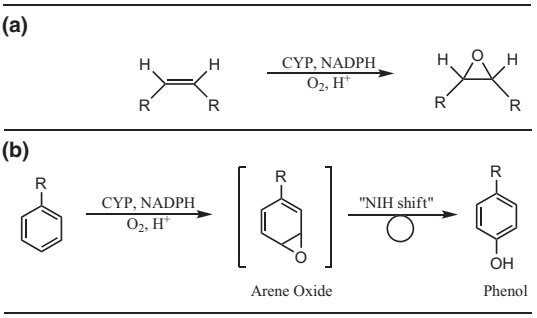


**Figure 13** Oxidation of $sp^2$ systems. **(a)** Alkenes are oxidized to epoxide. **(b)** Arenes (aromatic) are oxidized to arene oxides; rearrangement of the arene oxide results in aromatic hydroxylation.

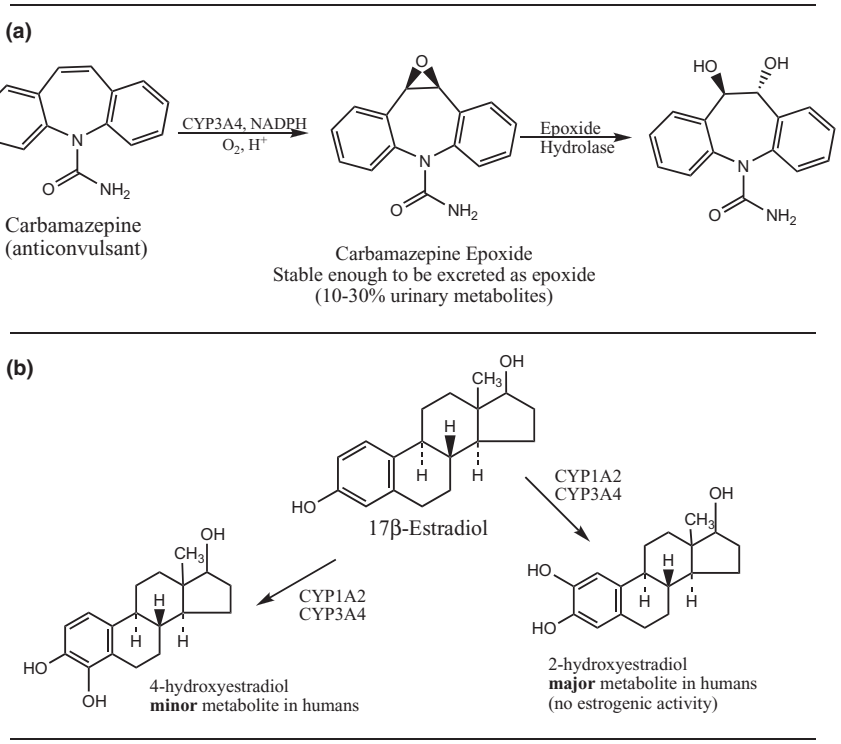


**Figure 14** **(a)** Example of alkene epoxidation and **(b)** aromatic hydroxylation.

benzyl. For amines, tertiary amines are favored for dealkylation over secondary over primary amines. When a primary amine is dealkylated, the reaction is usually called *N*-deamination, rather than *N*-dealkylation, but the mechanism is the same. Examples of *N*-dealkylation/deamination are shown in Fig. 17.

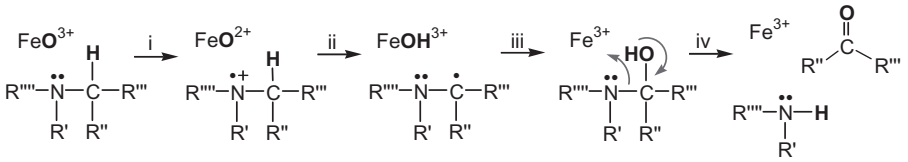


**Figure 15** Unified P450 mechanism for N-dealkylation of amine nitrogen. Mechanistic steps defined as (i) nonbonded electron transfer, (ii) hydrogen atom abstraction, (iii) oxygen rebound, and (iv) non-enzymatic rearrangement (Guengerich, 2001).

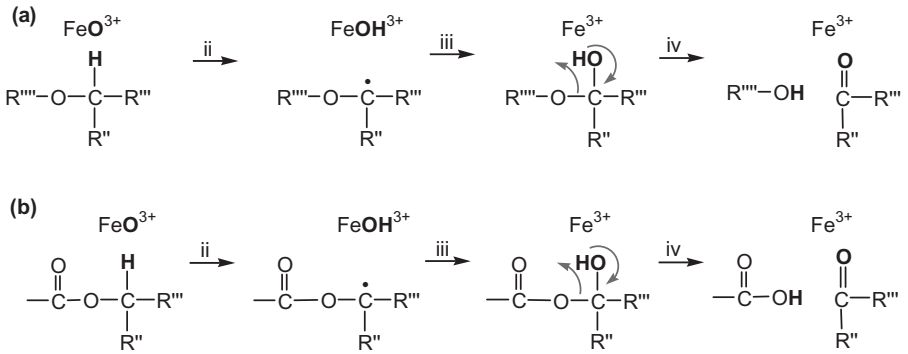


**Figure 16** Unified P450 mechanism for O-dealkylation of ethers **(a)** and oxidative cleavage of esters **(b)**. Mechanistic steps defined as (ii) hydrogen atom abstraction, (iii) oxygen rebound, and (iv) non-enzymatic rearrangement (Guengerich, 2001).

*O*-dealkylations generally occur near terminal positions on the molecule, and methoxy groups are especially susceptible (*O*-demethylation). As shown in Fig. 16, the reaction on ethers can be rationalized similarly to *N*-dealkylation, but starting at step ii: (ii) proton abstraction, (iii) oxygen rebound, and (iv) nonenzymatic rearrangement (Guengerich, 2001). The same mechanism holds for another reaction, oxidative cleavage of esters and amides (Fig. 16). *In vitro*, oxidative cleavage can be distinguished from ester hydrolysis by whether a ketone/aldehyde or an alcohol, respectively, accompanies the carboxylic acid product (Guengerich, 1987; Peng *et al.*, 1995), and by whether the reaction is dependent on presence of CYP and its electron-donating system. Oxidative cleavage also occurs on amides, carbamates, and ureas (Peng *et al.*, 1995). Examples of *O*-dealkylation and oxidative cleavage are shown in Fig. 17.

### 4.4 CYP Heteroatom Oxygenation Reactions

The final category of CYP-catalyzed oxidations involves the direct oxidation of heteroatoms (N, S). Originally these reactions were thought to be exclusively catalyzed by flavin-containing monooxygenase (FMO) enzyme system, but more recently the role of CYPs has been recognized. As shown in Fig. 18, the amine oxidation can be rationalized using three steps: (i) first one-electron transfer, (ii) second one-electron transfer, and (iii) oxygen rebound. Primary amine nitrogens are enzymatically oxidized

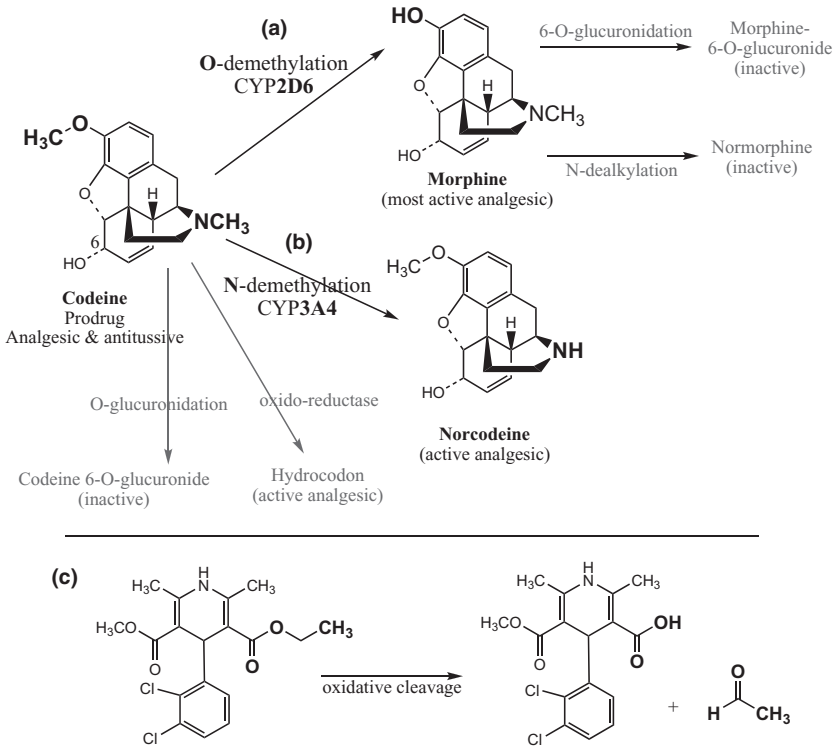


**Figure 17** Examples of N-dealkylation, O-dealkylation, and oxidative cleavage. **(a)** N-dealkylation and O-dealkylation of codeine. **(b)** Oxidative cleavage of ester functional group in felodipine. Amides, carbamates, and ureas are also cleaved through oxidation.

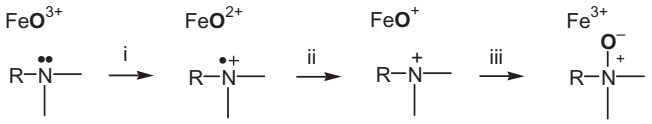


**Figure 18** Unified P450 mechanism for N-oxygenation. Mechanistic steps defined as (i) first nonbonded electron transfer, (ii) second nonbonded electron transfer, and (iii) oxygen rebound (Guengerich, 2001).

to the respective *N*-hydroxylamine, with further spontaneous oxidation to the nitroso and nitro form (Fig. 19a). Secondary amines are enzymatically oxidized to the respective secondary *N*-hydroxylamine, with further spontaneous oxidation to the nitrone form (Fig. 19b). Tertiary amines are enzymatically oxidized to the amine oxide (*N*-oxide) (Fig. 19c). CYPs also catalyze oxidation of sulfur atoms as shown in Fig. 20. Most sulfoxidation reactions are catalyzed by a single isoform, CYP3A4. Indeed, a recent computational regioselectivity model indicated that CYP3A4-catalyzed sulfur

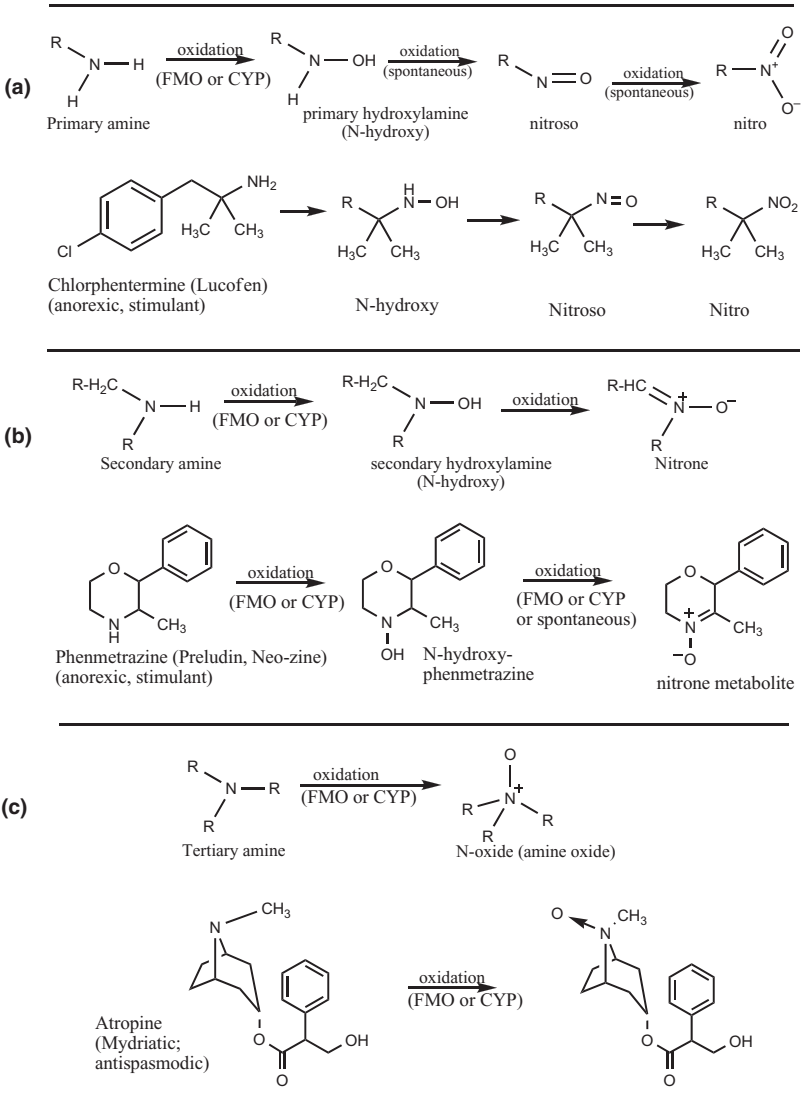


**Figure 19** Examples of nitrogen oxidation: **(a)** primary amine oxidation, **(b)** secondary amine oxidation, and **(c)** tertiary amine oxidation. Each generic reaction is shown first, followed by a specific example.

oxidation is the major reaction for nearly all drug molecules containing an oxidizable sulfur (Sheridan *et al.*, 2007).

CYP also catalyzes less common reactions that are of interest within the field of drug metabolism including oxidation of 1,4-dihydropyridines, oxidative deformylation, oxidation of a double bond with 1,2-migration, and steroid aromatization (Guengerich,

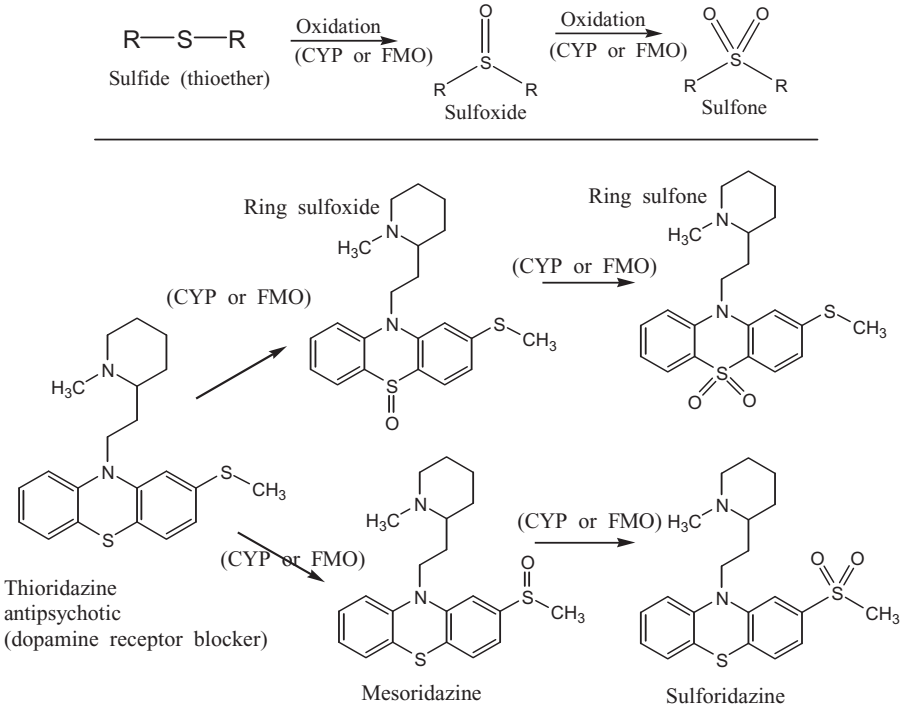


**Figure 20** Examples of sulfur oxidation.

2001). Examples of these reactions relevant to drug metabolism are shown in Fig. 21.

## 5 ENZYMOLOGY AND MODIFIERS OF CYPs

A number of features differentiate the many CYPs, including relative expression levels and 3D protein structure of the substrate-binding cavity and access channel. Relative expression is controlled by tissue localization (GI versus liver versus kidney versus other), age- and gender-related differences, presence of genetic polymorphisms with functional relevance, and susceptibility to induction/suppression. Inhibitors and activators act via direct binding with the enzyme and, thus, their effects are influenced by differences in the 3D protein structure of the substrate-binding cavity and access channel. Substrate selectivity and regioselectivity are also defined by the features of the substrate-binding cavity and access channel.

### 5.1 3D Protein Structure of the Substrate-Binding Cavity and Access Channel

The 3D protein structure of the substrate-binding cavity and access channel is unique for each CYP and, thus, controls the differential binding of ligands. Substrates and

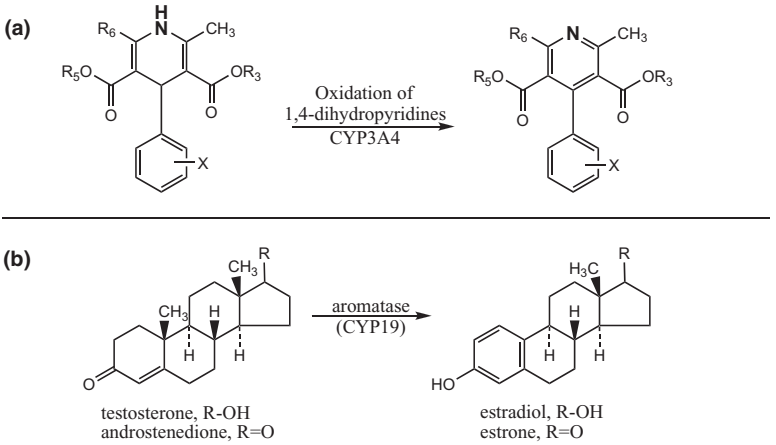


**Figure 21** Other oxidations relevant to drug metabolism. **(a)** Oxidation of 1,4-dihydropyridines, and **(b)** steroid aromatization.

inhibitors must be able to enter into the binding cavity through the available access channel and have a degree of affinity for the binding cavity. Although the ligand and the protein side chains have some conformational flexibility, studies have shown that this flexibility is limited by atomic and bonding interactions. Thus, each CYP member is limited to a range of ligands that can interact with its binding cavity, and these limitations define the isoform-, substrate-, inhibitor-, and regioselectivity of each P450 member.

***5.1.1 Isoform Selectivity for Substrates and Inhibitors*** *Isoform selectivity* refers to the ability of a particular enzyme member to accept a certain ligand into the binding pocket, either as a substrate or as an inhibitor. For example, a potential ligand may be excluded from the binding pocket when it is larger than the access channel or the cavity available for ligand binding. In contrast, a potential ligand may be small enough to fit within the binding cavity but lack complementary structural features necessary for appropriate binding affinity with a particular enzyme isoform. By definition, P450 *substrates* are ligands that can bind with suitable affinity and with an oxidizable functional group within catalytic distance of the activated heme oxygen. Note that some substrates can also inhibit the metabolism of other substrates via competitive binding at the active site. Rational inspection of the pool of substrate structures oxidized by each particular CYP member has allowed vague, but fairly reliable, predictions as to which general classes of ligands will be substrates for each major P450 (Lewis, 2003; Totah and Rettie, 2005). The available X-ray crystal structures and homology models of human P450s support these predictions. A more quantitative approach recently confirmed and extended these predictions for CYP3A4, 2D6, and 2C9 (Terfloth, Bienfait and Gasteiger, 2007). Combining all these sources, the authors have prepared a modified decision tree exemplified in Fig. 22. In brief, CYP1A2 has a relatively flat yet broad binding pocket and accepts mostly planar molecules such as aromatic rings. CYP2E1 and 2A6 have relatively small binding pockets and accept only rather small substrates. In contrast, CYP3A4 and 2C8 have quite large binding pockets,

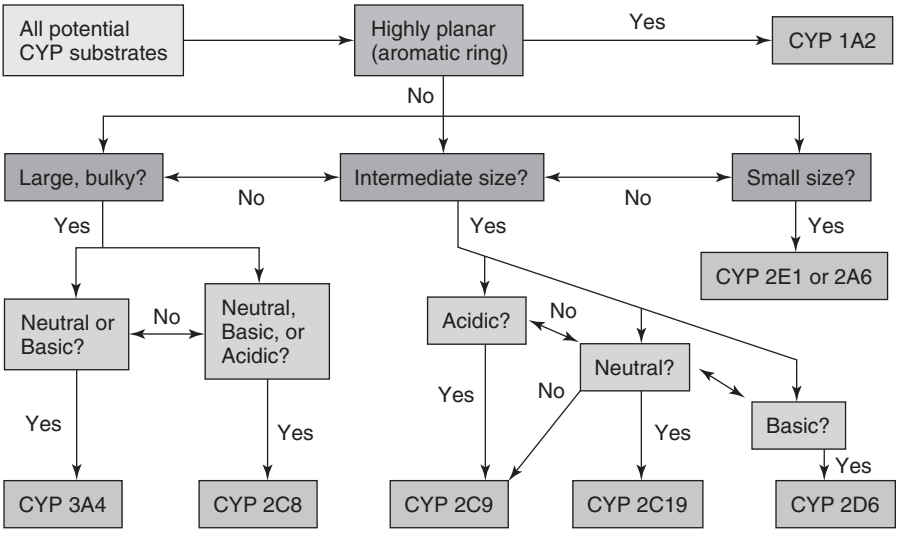


**Figure 22** Decision tree for prediction of dominant cytochrome P450 isoform based on substrate physicochemical features.

accepting up to rather large bulky molecules. CYP2D6 and 2C19 have preference for basic molecules, and CYP2C9 has a strong preference for molecules with one or more acidic functional groups.

These preferences can be rationalized by observation of significant differences in the relative features of the binding cavity for each major human CYP as defined by X-ray crystal structures and homology models. For example, CYP3A4- and 2C8-binding cavities are rather large (calculated to be 1386 and 1438 $Å^3$, respectively). CYP2E1- and 2A6-binding cavities are relatively small (2A6 calculated to be 260 $Å^3$). Other CYP members have intermediate size binding cavities (2D6, 2C9, 2C19), each of which is differentiated through the physicochemical properties of the distinct amino acid side chains lining the access channel and within the binding cavity. For example, CYP2C9 has a strong preference for substrates with acidic functional groups. The structure of CYP2C9 shows a positively charged region at the opening of the access channel, which could attract molecules with acidic functional groups and repel molecules with basic functional groups. While both CYP2C8 and CYP3A4 have rather large binding pockets, visualization of the binding cavities shows that they should each have distinct steric restrictions on ligand binding. In contrast, visualization of CYP2A6-binding cavity rationalizes why this P450 member accepts only rather small ligands.

***5.1.2 Regioselectivity*** The decision tree, however, does not address *regioselectivity*, that is, which atom will be oxidized by a particular CYP member when multiple sites are available. Each CYP member exhibits a unique "metabolite fingerprint" because regioselectivity is influenced not only by relative reactivity of the oxidizable atoms but also by binding affinity and orientation relative to the activated heme oxygen. A recently published computational study confirmed and extended prior

**TABLE 3 Regioselectivity preferences of CYP3A4, CYP2D6, and CYP2C9.**

| |
|---|
| CYP3A4 |
| 1. *N*-dealkylations (*N*-demethylation is most preferred) |
| 2. Sulfur oxidations (whenever available) |
| 3. Allylic and benzylic carbon hydroxylation |
| 4. Aromatic carbon hydroxylation, especially to catechol |
| 5. Oxidizes atoms near end of molecule |
| CYP2D6 |
| 1. Aromatic methoxy (phenol *O*-demethylation) |
| 2. Aromatic methyl oxidation (benzylic carbon oxidation) |
| 3. *O*-demethylation |
| 4. Oxidizes atoms at end or near middle of long molecule |
| CYP2C9 |
| 1. Aromatic methoxy (phenol *O*-demethylation) |
| 2. Aromatic methyl oxidation (benzylic carbon oxidation) |
| 3. Sulfur oxidations |
| 4. *N*-dealkylations |
| 5. Aromatic hydroxylation is not particularly favored, but when it occurs, it results in catechol formation |
| 6. Oxidizes atoms at end or near middle of long molecule |

Sheridan's model predicts these observed preferences (Sheridan *et al.*, 2007).

understanding of the structural parameters associated with the regioselectivity of three major CYP members: CYP3A4, 2C9, and 2C19 (Sheridan *et al.*, 2007). Specifically, all three studied CYP members were found to preferentially oxidize non-sterically hindered atoms, as measured by the magnitude of hydrogen atom exposure. Atoms at the exposed ends of molecules were also found to be preferentially oxidized by all three CYP members. CYP2D6 and 2C9 were found to *allow* oxidation of atoms in the middle of a long molecule; presumably because 2D6 and 2C9 are less sterically hindered in the region proximal to the heme and allow more "folding" of the substrate relative to CYP3A4.

The computational model also found that each CYP member exhibited distinct regioselective preferences which result in quite different metabolic fingerprints (Table 3). For example, Sheridan's model confirmed the long-standing knowledge that the stability of the intermediate carbon radical is a major influence for regioselectivity of CYP3A4. The model also showed that CYP3A4 prefers to oxidize sulfur atoms whenever they are available in a ligand, and to also oxidize aromatic carbons to form phenols and catechols.

## 5.2 Tissue Localization and Relative Expression Levels

***5.2.1 CYP3A4*** The major sites for drug metabolism are the liver and the mucosa of GI tract, with some metabolism occurring in the kidney. While all drug-metabolizing CYP members are expressed to some extent in the liver, CYP3A4 is the only member that is also highly expressed in the mucosa of the GI tract. Thus, CYP3A4 is responsible for much "first-pass" metabolism. Newborn and fetal livers express very little (or no) CYP3A4 and instead express the related member CYP3A7 (Hines, 2007). Some adults express an additional CYP3A family member, CYP3A5, in the liver and GI mucosa,

whereas CYP3A5 appears to be the dominant form in the kidney of most individuals. Because the isoform- and regioselectivity of these related members slightly differ, the relative amounts of 3A4 versus 3A5 versus 3A7 can cause variation in the metabolites formed in each individual. However, at this time, only a few compounds have been characterized for differential metabolism by CYP3A4, 3A5, or 3A7.

In addition to differences in constitutive levels, CYP3A4 expression is widely affected by temporal factors. This is mostly attributable to endogenous and exogenous inducers/suppressors of expression. Exogenous CYP3A4 inducers are well characterized and act mostly through the nuclear transcription factors and response elements. Recently, a few endogenous CYP3A4 expression regulators have been identified such as vitamin D3, growth hormone, and triiodothyronine. Gender, age, and disease-state differences in CYP3A4 expression may also act via endogenous regulators of gene expression.

***5.2.2 CYP2D6*** CYP2D6 constitutes only a small fraction (~2%) of the total enzyme content of typical human liver tissue, yet it contributes to the metabolism of approximately 25% of all marketed drugs. This paradox occurs because CYP2D6 substrates are highly represented among the total pool of drug structures. Because of wide variation in CYP2D6 allele expression due to gene duplication, deletion, or mutation events, individuals may have abolished, decreased, normal, or ultrarapid CYP2D6 activity. The phenotype (CYP2D6 activity) has been separated into four groups: poor metabolizers (PMs), intermediate metabolizers (IMs), extensive metabolizers (EMs), and ultrarapid metabolizers (UMs). The majority (70–90%) of individuals in most studied populations are CYP2D6 EMs (normal activity), 5–30% are IMs (decreased activity), and 0–8% are PMs (abolished activity) (Ingelman-Sundberg *et al.*, 2007). However, gene duplications resulting in much higher prevalence of the ultrarapid phenotype (UM) have been selected for in Northeast Africa (*CYP2D6*2×N* genotype) and Oceania (*CYP2D6*1×N* genotype). A secondary effect modifying CYP2D6 metabolism is that presence of CYP2D6 inhibitors (or competitive substrates) can temporally cause lower 2D6-catalyzed metabolism. Thus, individuals with inherited intermediate expression levels may, in the presence of one or more 2D6 inhibitors, metabolize CYP2D6 substrates more slowly than predicted from the genotype alone. Thus, CYP2D6 is responsible for a large portion of pharmacokinetic variation among individuals. A comprehensive review on CYP2D6 has been recently published (Zanger, Raimundo and Eichelbaum, 2004).

***5.2.3 CYP2C9*** CYP2C9 prefers acidic molecules as substrates and oxidizes such drug categories as nonsteroidal anti-inflammatory drugs, oral antidiabetics, warfarin, antiepileptics, and hypnotics. While CYP2C9 represents ~20% of the average hepatic CYP content, it metabolizes only ~10% of all marketed drugs. CYP2C9 is affected by all three mechanisms of variability (genetic polymorphism, enzyme induction, and enzyme inhibition) and is responsible for individual differences in drug response and side effects. The polymorphic activity of CYP2C9 is due mainly to two common coding variants, *CYP2C9*2* and *CYP2C9*3*. These variant alleles are present in up to 10% of Caucasians, less in Africans, and not at all in Asians (Ingelman-Sundberg *et al.*, 2007). The *CYP2C9*3* allele results in production of protein which has only about 10% of the intrinsic clearance for most substrates versus the wild-type CYP2C9. The effect of the CYP2C9*2 variant varies with substrate, but generally slightly lower catalytically efficiency than the wild-type enzyme.

***5.2.4 CYP2C19*** CYP2C19 represents about 5% of the total P450 content in typical human liver tissue, and metabolizes approximately 3% of all marketed drugs. CYP2C19 is affected by genetic polymorphisms and enzyme inhibition. Individuals carrying two defective *CYP2C19* genes (poor metabolizers) represent up to 5% of Caucasian and African populations, and up to 20% of Asian populations (Ingelman-Sundberg *et al.*, 2007). This poor CYP2C19 metabolizer phenotype has proven beneficial for the treatment of GI disorders with proton pump inhibitors. The poor metabolizers have reduced metabolism of these drugs, with concomitant higher drug plasma levels and increased responsiveness to treatment. The pharmacokinetics of the tricyclic antidepressants amitriptyline and clomipramine, and the selective serotonin reuptake inhibitors (SSRI) sertraline and citalopram, are affected by CYP2C19 genotype/phenotype, but significant clinical effect or risk is as yet unclear.

## REFERENCES


Bornheim, L.M., Lasker, J.M., Raucy, J.L. (1992). Human hepatic microsomal metabolism of delta 1-tetrahydrocannabinol. *Drug Metab. Dispos.*, *20* (2), 241–246.

Emoto, C., Murase, S., Iwasaki, K. (2006). Approach to the prediction of the contribution of major cytochrome P450 enzymes to drug metabolism in the early drug-discovery stage. *Xenobiotica*, *36* (8), 67–683.

Evans, W.E., Relling, M.V. (1999). Pharmacogenomics: translating functional genomics into rational therapeutics. *Science*, *286*, 487–491.

Fuccella, L.M., Tamassia, V., Valzelli, G. (1973). Metabolism and kinetics of the hypoglycemic agent glipizide in man: comparison with glibenclamide. *J. Clin. Pharmacol.*, *13*, 68–75.

Guengerich, F.P. (1987). Oxidative cleavage of carboxylic esters by cytochrome P-450. *J. Biol. Chem.*, *262* (18), 8459–8462.

Guengerich, F.P. (2001). Common and uncommon cytochrome P450 reactions related to metabolism and chemical toxicity. *Chem. Res. Toxicol.*, *14* (6), 611–650.

Guengerich, F.P., McDonald, T.L. (1984). Chemical mechanisms of catalysis by cytochromes P-450: a unified view. *Acc. Chem. Res.*, *17*, 9–16.

Hamman, M.A., Thompson, G.A., Hall, S.D. (1997). Regioselective and stereoselective metabolism of ibuprofen by human cytochrome P450 2C. *Biochem. Pharmacol.*, *54* (1), 33–41.

Hines, R.N. (2007). Ontogeny of human hepatic cytochromes P450. *J. Biochem. Mol. Toxicol.*, *21* (4), 169–175.

Ingelman-Sundberg, M., Sim, S.C., Gomez, A., Rodriguez-Antona, C. (2007). Influence of cytochrome P450 polymorphisms on drug therapies: pharmacogenetic, pharmacoepigenetic and clinical aspects. *Pharmacol. Ther.*, *116*, 496–526.

Lemberger, L. (1973). Tetrahydrocannabinol metabolism in man. *Drug Metab. Dispos.*, *1* (1), 461–468.

Lewis, D.F.V. (2003). Essential requirements for substrate binding affinity and selectivity toward human CYP2 family enzymes. *Arch. Biochem. Biophys.*, *409*, 32–44.

Monahan, B.P., Ferguson, C.L., Killeavy, E.S., Lloyd, B.K., Troy, J., Cantilena, L.R., Jr (1990). Torsades de pointes occurring in association with terfenadine use. *JAMA*, *264* (21), 2788–2790.

Paulson, S.K., Hribar, J.D., Liu, N.W., Hajdu, E., Bible, R.H., Jr, Piergies, A., Karim, A. (2000). Metabolism and excretion of [(14)C]celecoxib in healthy male volunteers. *Drug Metab. Dispos.*, *28* (3), 308–314.

Peng, H.-W., Raner, G.M., Vaz, A.D.N., Coon, M.J. (1995). Oxidative cleavage of esters and amides to carbonyl products by cytochrome P450. *Arch. Biochem. Biophys.*, *318* (2), 333–339.

Rudy, A.C., Knight, P.M., Brater, D.C., Hall, S.D. (1991). Stereoselective metabolism of ibuprofen in humans: administration of R-, S- and racemic ibuprofen. *J. Pharmacol. Exp. Ther.*, *259* (3), 1133–1139.

Rudy, A.C., Knight, P.M., Brater, D.C., Hall, S.D. (1991). Stereoselective metabolism of ibuprofen in humans: administration of R-, S- and racemic ibuprofen. *J. Pharmacol. Exp. Ther.*, *259* (3), 1133–1139.

Sheridan, R.P., Korzekwa, K.R., Torres, R.A., Walker, M.J. (2007). Empirical regioselectivity models for human cytochromes P450 3A4, 2D6, and 2C9. *J. Med. Chem.*, *50* (14), 3173–3184.

Shimada, T., Yamazaki, H., Mimura, M., Inui, Y., Guengerich, F.P. (1994). Interindividual variations in human liver cytochrome P-450 enzymes involved in the oxidation of drugs, carcinogens and toxic chemicals: studies with liver microsomes of 30 Japanese and 30 Caucasians. *J. Pharmacol. Exp. Ther.*, *270* (1), 414–423.

Tang, C., Shou, M., Mei, Q., Rushmore, T.H., Rodrigues, A.D. (2000). Major role of human liver microsomal cytochrome P450 2C9 (CYP2C9) in the oxidative metabolism of celecoxib, a novel cyclooxygenase-II inhibitor. *J. Pharmacol. Exp. Ther.*, *293* (2), 453–459.

Terfloth, L., Bienfait, B., Gasteiger, J. (2007). Ligand-based models for the isoform specificity of cytochrome P450 3A4, 2D6, and 2C9 substrates. *J. Chem. Inf. Model.*, *47*, 1688–1701.

Totah, R.A., Rettie, A.E. (2005). Cytochrome P450 2C8: substrates, inhibitors, pharmacogenetics, and clinical relevance. *Clin. Pharmacol. Ther.*, *77* (5), 341–352.

Williams, R.T. (1947). *Detoxication Mechanisms: The Metabolism of Drugs and Allied Organic Compounds*. Chapman and Hall, London.

Zanger, U.M., Raimundo, S., Eichelbaum, M. (2004). Cytochrome P450 2D6: overview and update on pharmacology, genetics, biochemistry. *Naunyn Schmiedebergs Arch. Pharmacol.*, *369* (1), 23–37.

# *In Vivo* Metabolite Kinetics

ZHENG YANG

## 1 INTRODUCTION

Following the administration of a drug, the pharmacokinetics and disposition of the metabolite(s) that are formed can be described and modeled. Such information is important because any metabolite can be pharmacologically active and contribute to the overall efficacy of the drug of interest. For example, atorvastatin (Lipitor®) produces two active hydroxylated metabolites that are equipotent to the parent drug *in vitro* and contribute to the prolonged inhibition of 3-hydroxy-3-methylglutaryl coenzyme A (HMG-CoA) reductase in an *ex vivo* assay (Christians, Jacobsen and Floren, 1998; Siedlik *et al.*, 1999). Another example is clopidogrel (Plavix®) whose pharmacological activity is due to the production of a reactive thiol metabolite that irreversibly blocks the ADP binding and receptor activation of the platelet $P2Y_{12}$ receptor (Savi *et al.*, 2000; Pereillo *et al.*, 2002). Information on active metabolites and their kinetics in the body is essential to comprehending the pharmacokinetic–pharmacodynamic (PK–PD) relationships of drugs and provides important guidance to their use in the clinic. With recent advancement in bioanalytical techniques, many of pharmacologically active metabolites can be identified and quantified in various biological fluids at the early stages of drug discovery. Understanding the pharmacological activity and pharmacokinetics of metabolites may provide useful insights into the PK–PD disconnect exhibited by a new chemical entity during *in vivo* efficacy testing. Moreover, the structural insights from active metabolites offer an excellent opportunity in the design of novel drug molecules (Fura *et al.*, 2004; Fura, 2006).

From a safety point of view, recent discussions on metabolites in safety testing (Baillie *et al.*, 2002; Hastings *et al.*, 2003; Smith and Obach, 2005, 2006; Davis-Bruno and Atrakchi, 2006) and the FDA Guidance on this topic also highlight the importance of understanding metabolite kinetics. In this context, it is imperative to make a distinction between a major circulating metabolite and a major metabolic pathway, and to understand the factors that determine the exposure of a metabolite in circulation and its appearance in excreta (e.g. urine and bile). In addition, knowledge of how metabolites distribute and accumulate in tissues is valuable to biotransformation-based

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

toxicity (Humphreys and Unger, 2006). Lastly, metabolite concentrations are often monitored as a way to detect alterations in the clearance of a drug in the presence of an inhibitor or inducer (Wilkinson, 2002). Appreciating the relevance of various metabolite-related pharmacokinetic parameters to the change in parent drug disposition is important to assess drug–drug interaction potential of various therapeutic agents. Clearly, in all the aforementioned areas, a thorough understanding of metabolite kinetics as well as factors governing such processes is essential.

In this chapter, the concepts and principles of *in vivo* drug metabolite kinetics will be reviewed and discussed along with examples. The important equations in the chapter are indicated by an asterisk (*). There are several excellent reviews on the subject (Houston, 1982; Pang, 1985). In addition, the book *Clinical Pharmacokinetics: Concepts and Applications* by Rowland and Tozer (1995) is highly recommended for everyone who is interested in exploring this topic further.

## 2 *IN VIVO* METABOLITE KINETIC CONCEPTS AND PRINCIPLES

### 2.1 General Considerations

In this chapter, linear pharmacokinetics is assumed throughout various kinetics models and equation derivations. The metabolite formed is assumed not to convert back to the parent drug; that is, there is no interconversion or reversible metabolism. All doses, drug amounts, and drug concentrations are in molar units in order to avoid conversions between a drug and a metabolite due to a difference in molecular weight.

### 2.2 Formation Clearance versus Elimination Clearance of Metabolite

The kinetics of a metabolite in the systemic circulation after dosing of a parent drug is governed by both its formation and its disposition. As illustrated in Fig. 1, the appearance of the metabolite in circulation is a function of enzyme activities in an eliminating organ (e.g. liver) and permeation of the metabolite formed across cellular membranes. Sequential metabolism and excretion into excreta (e.g. bile and urine) affect the concentration of the metabolite formed in the eliminating organ, thus also impacting on the rate and extent of the metabolite appearance in circulation. Once in the systemic circulation, the disposition of the metabolite in the body is the same as that of any other xenobiotics, which includes distribution and elimination processes.

In a simplified scheme, one could envision the metabolite kinetics using the model as shown in Fig. 2. Accordingly, the rate of change in the plasma and tissue concentration of a metabolite is given by the following equations:

$$V_d(m)\frac{dC(m)}{dt} = CL_f \cdot C - CL(m)\cdot C(m) + [CL_{d,t}(m)\cdot C_t(m) - CL_d(m)\cdot C(m)] \quad (1)$$

$$V_{d,t}(m)\frac{dC_t(m)}{dt} = CL_d(m)\cdot C(m) - CL_{d,t}(m)\cdot C_t(m) \quad (2)$$

where C is the parent drug concentration in plasma; C(m) and $C_t(m)$ are the metabolite concentrations in plasma and tissues, respectively; $V_d(m)$ and $V_{d,t}(m)$ are the volumes

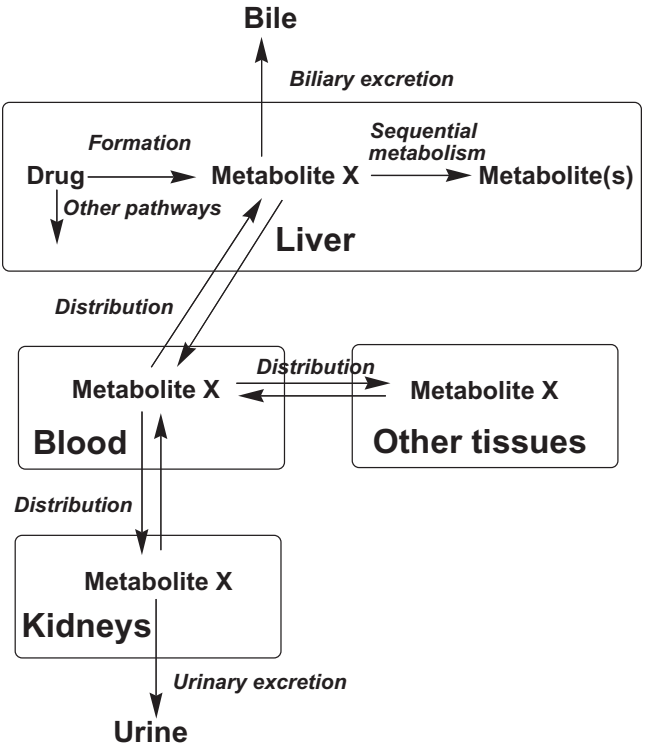


**Figure 1** A schematic representation of the fate of a metabolite (X) in the body following its generation in the liver.

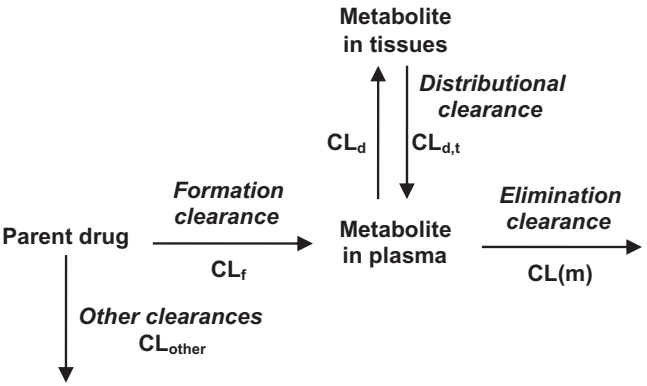


**Figure 2** A simplified model for metabolite kinetics.

of distribution of the metabolite in plasma and tissues, respectively; $CL_f$ is the clearance of the parent drug that forms the metabolite, also known as the formation clearance of the metabolite; CL(m) is the total body clearance of the metabolite, also known as the metabolite elimination clearance; $CL_d(m)$ and $CL_{d,t}(m)$ are the distributional clearances of the metabolite between plasma and tissues. Integrating Equations 1 and 2 from time zero to infinity and appreciating the fact that the metabolite concentration in the body at either time zero or infinity is zero, the following equations are obtained:

$$CL_f \cdot AUC - CL(m) \cdot AUC(m) + [CL_{d,t}(m) \cdot AUC_t(m) - CL_d(m) \cdot AUC(m)] = 0 \quad (3)$$

$$CL_d(m) \cdot AUC(m) - CL_{d,t}(m) \cdot AUC_t(m) = 0 \quad (4)$$

where AUC, AUC(m), and $AUC_t(m)$ are the total areas under the concentration–time curve of the parent drug in plasma, metabolite in plasma, and metabolite in tissues, respectively. Substituting Equation 4 into Equation 3 and rearranging the equation, it follows that:

$$\frac{AUC(m)}{AUC} = \frac{CL_f}{CL(m)} \quad (5)$$

where AUC(m)/AUC is the metabolite-to-parent (M/P) plasma AUC ratio. It is important to note that Equation 5 is valid only if the AUC is a total area from time zero to infinity. By the same token, the following relationship also exists:

$$\frac{C_{ss}(m)}{C_{ss}} = \frac{CL_f}{CL(m)} \quad (6)$$

where $C_{ss}(m)/C_{ss}$ is the M/P steady-state plasma concentration ($C_{ss}$) ratio. Equations 5 and 6 can be expanded by substituting $CL_f$ with $f_m \cdot CL$, where CL is the total body clearance of the parent drug and $f_m$ is the fraction of the total body clearance of the parent drug that forms the metabolite. Accordingly, the M/P AUC or $C_{ss}$ ratio can be represented as follows:

$$\frac{AUC(m)}{AUC} = f_m \cdot \frac{CL}{CL(m)} \quad (7)^*$$

$$\frac{C_{ss}(m)}{C_{ss}} = f_m \cdot \frac{CL}{CL(m)} \quad (8)^*$$

Equations 7 and 8 indicate that the M/P AUC or $C_{ss}$ ratio is determined by both the formation and the elimination clearance of the metabolite. In other words, an increase in the M/P ratio can be due to an increase in the formation clearance, decrease in the elimination clearance, or a combination of both.

### 2.3 Major Circulating Metabolite versus Major Metabolic Pathway

A major circulating metabolite is the metabolite that represents a substantial portion (e.g. ≥10%) of drug-related components in circulation. The exact percentage of the cutoff value for defining a major metabolite in safety testing is a subject of debate at the moment (Baillie *et al.*, 2002; Hastings *et al.*, 2003; FDA Guidance for Industry, 2008). However, the exact cutoff value may not be important; rather, considerations should be given to the structure and potential pharmacology and toxicology of a circulating metabolite prior to further safety assessment (Smith and Obach, 2005, 2006). Furthermore, it has been suggested that absolute abundance (i.e. metabolite concentrations in circulation) is more relevant than relative abundance (i.e. percentage) when considering metabolites in safety testing (Smith and Obach, 2005).

A major metabolic pathway is the route that constitutes a major fraction of biotransformation pathways. Although the same metabolite in some cases can be both a major circulating metabolite and the product of a major metabolic pathway, it is important to make a distinction between these two concepts. To appreciate the difference, please consider the following example as shown in Fig. 3. Drug X is metabolized to two metabolites (M1 and M2). The $CL_f$ and CL(M1) of M1 are 90 and 1000 mL/min, respectively. The $CL_f$ and CL(M2) of M2 are 10 and 5 mL/min, respectively. Based on Equation 5, the M/P AUC ratio of M1 and M2 is calculated to be 0.09 (=90 mL/min/1000 mL/min) and 2 (=10 mL/min/5 mL/min), respectively. The result demonstrates that M2, not M1, is the major circulating metabolite of drug X. However, the fraction of the

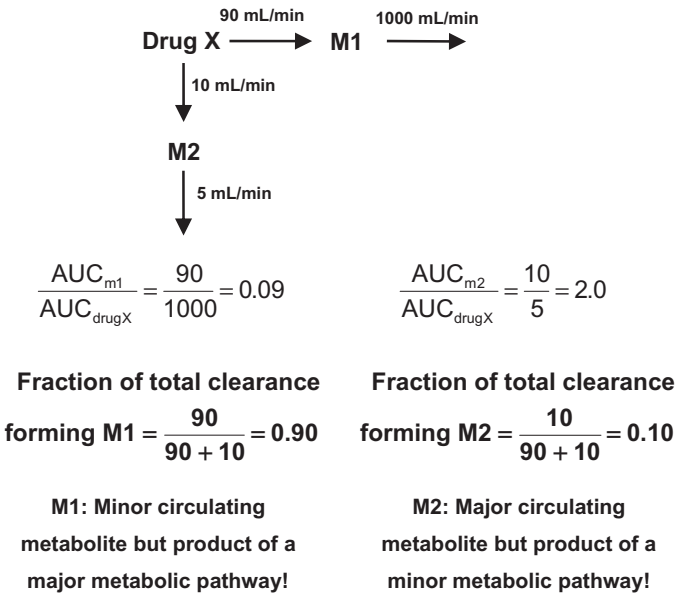


**Figure 3** A hypothetical example that illustrates the conceptual difference between a major circulating metabolite and the product of a major metabolite of biotransformation.

total clearance of drug X forming M1 is 0.90 (=90 mL/min/(90 mL/min + 10 mL/min)), whereas the fraction of drug X forming M2 is only 0.10. This result indicates that although M1 is a minor metabolite in circulation, it represents the product of the major metabolic pathway of drug X. This hypothetical example illustrates the point that a high M/P AUC ratio for a particular metabolite does not necessarily indicate that the pathway is the major biotransformation route of the parent drug. As illustrated in Fig. 1, the fate of a metabolite upon its formation in the liver (i.e. entrance into circulation, biliary secretion, and sequential metabolism) and upon its reaching the systemic circulation would affect the metabolite levels in circulation. Consequently, the M/P AUC or $C_{ss}$ ratio is governed by both the formation and the elimination clearance of the metabolite.

There are many examples in drug discovery and development that contrast the differences between a major circulating metabolite and a product of the major metabolic pathway. For example, muraglitazar underwent glucuronidation as the major elimination pathway in humans (Wang *et al.*, 2006). However, the glucuronide of muraglitazar was only a minor drug-related component in circulation, with an M/P AUC ratio less than 0.05. In another example, the formation of trifluoromethoxy salicylic acid from the substance P receptor antagonist CP-122721 represented only 1.4–3.5% of the dose in human excreta (Colizza, Awad and Kamel, 2007). However, trifluoromethoxy salicylic acid accounted for 29–56% of the total radioactivity in human circulation, whereas the parent drug was only up to 5.6%. The results indicate that trifluoromethoxy salicylic acid is a major circulating metabolite of CP-122721, but its formation pathway is only a minor metabolic pathway in humans. Clearly, understanding these metabolite kinetic concepts is important when considering the safety testing of metabolites observed in human circulation and excreta.

A true determination of the major biotransformation pathway of a drug is to measure the amount (e.g. radioactivity) of a particular metabolite as a fraction of dose in excreta. This can be achieved following the administration of a radiolabeled drug and profiling of excreta (Roffey *et al.*, 2007). The approach, however, can be confounded with sequential metabolism. Therefore, for this approach to be valid, it is important to understand the entire biotransformation pathways.

The pharmacokinetic relationship between a metabolite and a parent drug in excreta can be examined by considering the model scheme shown in Fig. 2. $CL_{other}$ and CL(m) are assumed to be the excretory clearance of the drug and metabolite, respectively. Based on mass balance, the following equations exist:

$$\mathrm{Ae} + \mathrm{Ae(m)} = \mathrm{Dose} \tag{9}$$

$$\mathrm{Ae} = (1 - f_m) \cdot \mathrm{Dose} \tag{10}$$

$$\mathrm{Ae(m)} = f_m \cdot \mathrm{Dose} \tag{11*}$$

$$\frac{\mathrm{Ae(m)}}{\mathrm{Ae}} = \frac{CL_f}{CL_e} \tag{12*}$$

where Ae and Ae(m) are the total amount of the parent drug and the metabolite recovered in excreta, respectively; $CL_e$ is the excretory clearance of the parent drug into the urine or bile, with a sum of the $CL_e$ and $CL_f$ equal to the total body clearance

(CL) of the parent drug; and dose is the dose of the drug administered. Equations 9–12 are valid only if the recovery of the parent drug and the metabolite is complete in excreta. The ratio of Ae(m) and Ae, which equals the ratio of the formation and excretory clearances of the parent drug, is sometimes referred to as the metabolic ratio.

Also, it is clear from Equation 11 that Ae(m), when normalized with dose, equals the $f_m$, a fraction that represents the extent of that particular biotransformation pathway to the total elimination processes. It is worth pointing out that the *extent* ($f_m$) of metabolism cannot be confused with the *rate* ($CL_f$) of metabolism, although a high $CL_f$ generally leads to a high $f_m$. This is because the $f_m$ is dependent on not only the rate of metabolism ($CL_f$) per se but also on the rate of removal from other pathways ($CL_{other}$). In the cases of tolbutamide and warfarin, they have a very low rate of metabolism in humans (i.e. metabolic clearance < 1 mL/min/kg), but they are extensively metabolized, with an $f_m$ approaching 1.0 (Thummel *et al.*, 2006).

### 2.4 Formation Rate-Limited versus Elimination Rate-Limited Metabolite Kinetics

Similar to the M/P AUC or $C_{ss}$ ratio, the concentration–time profile of a metabolite is also governed by the formation and elimination processes. In this case, however, not only does the clearance play a role, but also the volume of distribution of the parent drug and metabolite is important. This is because both clearance and volume, as independent pharmacokinetic parameters, determine the half-life (or elimination rate constant), which in turn governs the shape of concentration–time profiles. The half-life ($t_{1/2}$) as a function of the elimination rate constant ($k_{el}$), volume of distribution, and clearance is shown in the following equations:

$$t_{1/2}(m) = \frac{0.693}{k_{el}(m)} = \frac{0.693 \cdot V_d(m)}{CL(m)} \tag{13}$$

$$t_{1/2} = \frac{0.693}{k_{el}} = \frac{0.693 \cdot V_d}{CL} \tag{14}$$

where $t_{1/2}(m)$ is the half-life of the metabolite when dosed by itself; $t_{1/2}$ is the half-life of the parent drug; $k_{el}(m)$ and $k_{el}$ are the elimination rate constants of the metabolite and parent drug, respectively. Accordingly, when the formation rate constant $k_f$ is much slower than the $k_{el}(m)$ (i.e. the rate-limiting step is the formation of the metabolite), the concentration–time profile of the metabolite is parallel to that of the parent (Fig. 4a). In this case, the apparent $t_{1/2}$ of the metabolite when dosed as a parent drug is the same as that of the parent, but the true half-life ($t_{1/2}(m)$) of the metabolite when dosed by itself is much shorter than that of the parent. This case is known as formation rate-limited metabolite kinetics. In some cases, metabolite kinetics can also be elimination rate-limited. The situation arises when the $k_f$ is much faster than the $k_{el}(m)$ (i.e. the rate-limiting step is the elimination of the metabolite). As a result, the concentration–time profile of the metabolite is no longer parallel to that of the parent, with an elimination $t_{1/2}$ of the metabolite longer than that of the parent (Fig. 4b). In this case, the $t_{1/2}$ of the metabolite when dosed as a parent drug is the same as the $t_{1/2}(m)$ obtained after dosing the metabolite by itself. The situation is much like the flip-flop kinetics

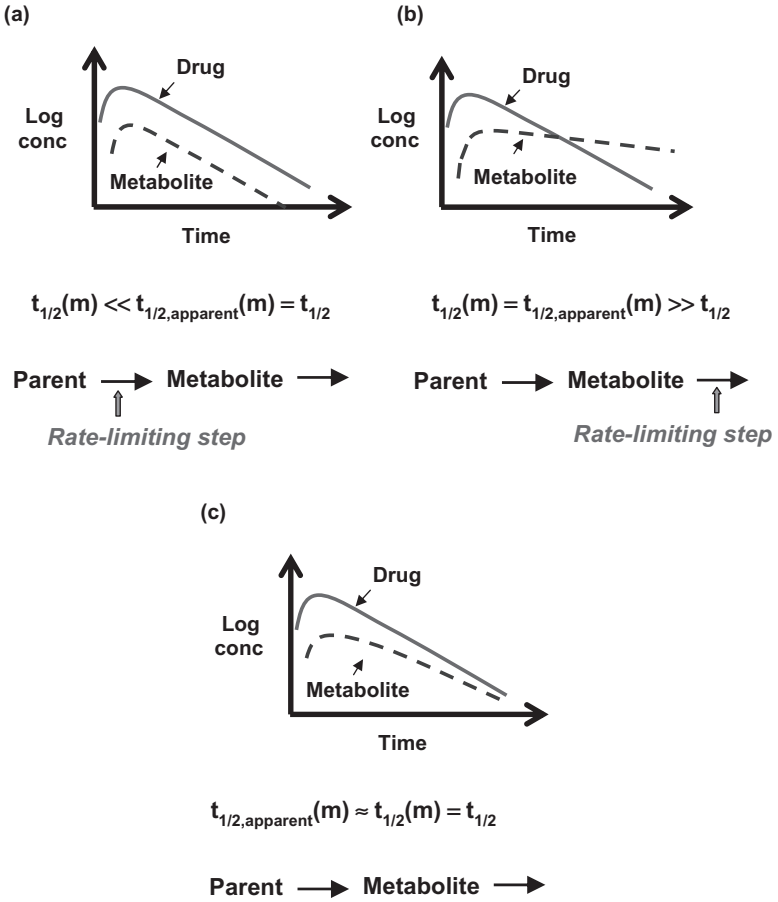


**Figure 4** Effects of the rate-limiting step in metabolite kinetics on the shape of metabolite curves and metabolite half-lives after a single dose of a parent drug, where $t_{1/2}$ is the terminal half-life of the parent drug; $t_{1/2}(m)$ is the terminal half-life of the metabolite after intravenous dosing of the metabolite itself; and $t_{1/2,apparent}(m)$ is the apparent terminal half-life of the metabolite after dosing of the parent drug. **(a)** Formation rate-limited. **(b)** Elimination rate-limited. **(c)** Neither of steps is rate-limiting.

following oral administration, where the apparent oral $t_{1/2}$ is longer than that after intravenous dosing. The third situation is that the $k_f$ is equal to the $k_{el}(m)$. In this case, neither formation nor elimination process is rate-limiting. The concentration–time profile of the metabolite is nearly parallel to that of the parent (Fig. 4c), with an apparent $t_{1/2}$ of the metabolite slightly longer than that of the parent, which is also longer than its true $t_{1/2}(m)$.

Knowledge of whether metabolite kinetics is formation or elimination rate-limited has a direct implication on how a metabolite is cleared from the body, how quickly it reaches steady state, and how much accumulation it may have relative to the parent. This is of great importance in the situations where a metabolite is pharmacologically

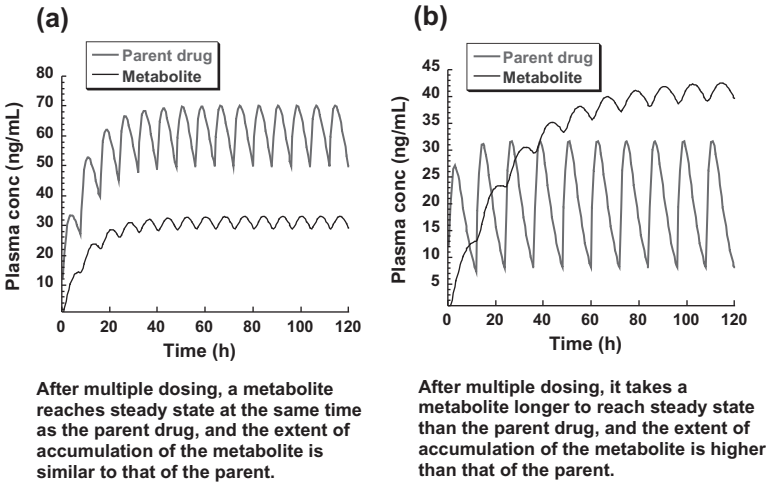


**Figure 5** Effects of the rate-limiting step in metabolite kinetics on the time to reach steady state and extent of accumulation of a metabolite during multiple dosing of a parent drug. **(a)** Formation rate-limited. **(b)** Elimination rate-limited.

active or toxic. For formation rate-limited metabolite kinetics, the metabolite reaches the steady state at the same time as the parent drug with an extent of accumulation similar to that of the parent (Fig. 5a). By the same token, the metabolite is cleared from the body at the same rate as the parent. However, for elimination rate-limited metabolite kinetics, it takes much longer for the metabolite to reach steady state than the parent and the metabolite has a higher accumulation ratio than that of the parent (Fig. 5b). Similarly, the metabolite is cleared from the body at a slower rate than that of the parent. As a result, when a metabolite is pharmacologically active or toxic, it may lead to a disconnect between parent drug concentrations and pharmacological response or toxicity. For example, atorvastatin is an HMG CoA reductase inhibitor that has a mean plasma elimination $t_{1/2}$ of ~14 hours in humans (Williams and Feely, 2002). It undergoes the extensive first-pass metabolism in the liver and produces two active hydroxylated metabolites in humans that have equipotent activity to the parent drug *in vitro* (Christians, Jacobsen and Floren, 1998; Siedlik *et al.*, 1999). Using an *ex vivo* assay that measures the inhibition of HMG-CoA reductase in plasma samples, the plasma $t_{1/2}$ of inhibitory activity was estimated to be 20–45 hours (Siedlik *et al.*, 1999; Williams and Feely, 2002), significantly longer than that of the parent drug. The results indicate that the active metabolites of atorvastatin likely exhibit elimination rate-limited kinetics in plasma.

## 2.5 Pre-systemically Formed versus Systemically Formed Metabolite

Pre-systemically formed metabolites are generally referred to as the metabolites produced during the first pass after oral administration, which generally involves metabolism in the gut wall and liver. Systemically formed metabolites are the metabolites that are produced during the general circulation such as the ones after intravenous dosing or subsequent organ passes after reaching the systemic circulation following oral

administration. Recognizing the anatomical location of the liver in the body, it is unlikely that hepatic metabolism during the first pass is any different from that during subsequent passes or systematic circulation, provided that no saturation in metabolism occurs. On the other hand, the gut wall likely plays a more pronounced role during the first pass than that during subsequent circulation because of its anatomical location and blood flow to that region. In addition, numerous drug-metabolizing enzymes (e.g. cytochrome P450 enzymes, sulfotransferases, and UDP-glucuronosyltransferases) are expressed in enterocytes and have been shown to metabolize drugs such as midazolam, 17α-ethinylestradiol, terbutaline, and isoproterenol (Thummel and Shen, 2002). Therefore, the gut wall, in addition to the liver, is an important organ in generating pre-systemically formed metabolites.

The M/P plasma AUC ratio, based on Equation 7, should be independent of the route of drug administration, provided that the $f_m$ does not change with the dosing route and pharmacokinetics is linear between the routes of administration. However, this may not be the case when a drug is administered orally, where the M/P AUC ratio after oral dosing is often different from that after intravenous dosing. When the M/P AUC ratio after oral dosing is higher than that after intravenous dosing, it indicates that there are additional metabolites formed pre-systemically and these metabolites enter the systemic circulation directly. In this case, it is as if a portion of metabolites is directly dosed and, as a result, the $f_m$ is changed between intravenous and oral routes. One good example is midazolam. The plasma AUC ratio between 1′-hydroxy-midazolam and midazolam in humans after oral administration was ~0.70, whereas the M/P AUC ratio after IV dosing was ~0.12 (Thummel *et al.*, 1996), indicating a substantial amount of 1′-hydroxy-midazolam was produced during the first pass after oral administration and the gut wall could contribute significantly to the metabolite formed pre-systematically. Indeed, in anhepatic patients, the average gut wall extraction ratio after intraduodenal administration was 0.43, significantly higher than that (0.08) after intravenous dosing (Paine *et al.*, 1996). When the M/P AUC ratio after oral dosing is lower than that after intravenous dosing, it suggests that the formation clearance may be saturated during the first-pass metabolism. This is possible as drug concentrations in the gut wall and portal vein during the first pass are generally higher than those after intravenous dosing, which could lead to the saturation of drug-metabolizing enzymes, particularly when a biotransformation reaction is characterized by a low Michaelis–Menten constant (Km) or when the oral dose is high.

In addition to the M/P plasma AUC ratio, pre-systemically formed metabolites could have a significant impact on the shape of the metabolite curve when a metabolite exhibits formation rate-limited kinetics. Using a model scheme originally proposed by Houston (Houston, 1982), the effect of pre-systemically formed metabolites on the shape of the metabolite curve can be examined. When there is no pre-systemically formed metabolite, the metabolite curve is parallel to that of the parent (Fig. 6a), as expected from the formation rate-limited kinetics. When 50% of drug dose is converted to the metabolites pre-systemically, the metabolite curve starts to show a biphasic phenomenon, with metabolites in circulation coming from two sources—pre-systemically and systemically formed (Fig. 6b). When nearly 90% of drug dose is converted to a metabolite pre-systemically during the first-pass metabolism, the metabolite curve exhibits a classic biphasic shape (Fig. 6c). The initial portion of the metabolite curve is largely made up of the pre-systemically formed metabolite, which can peak earlier than that of the parent and have a shorter $t_{1/2}$ than that of the parent depending on the

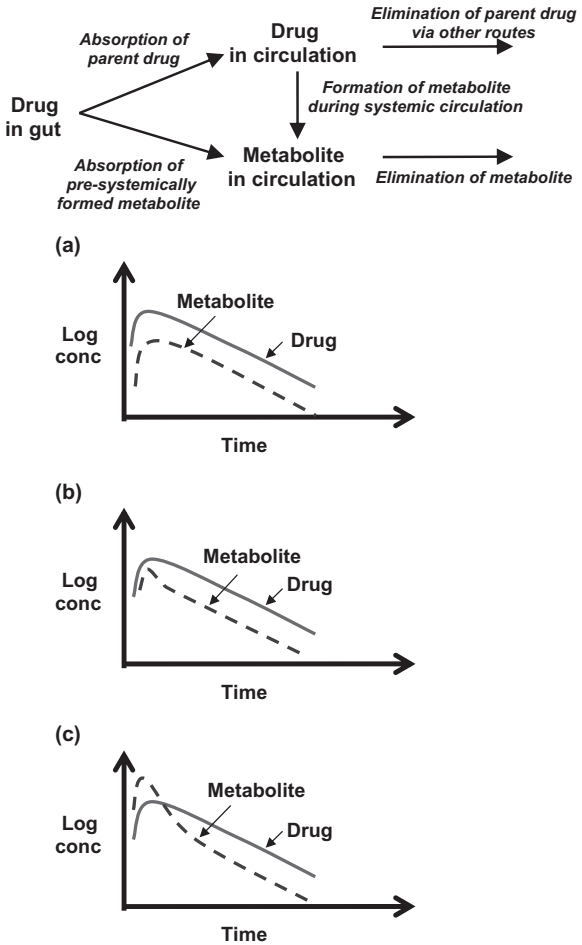


**Figure 6** Effects of pre-systemically formed metabolites on the shape of metabolite curves following a single oral dose of a parent drug. **(a)** No first-pass extraction. **(b)** First-pass extraction is 50% of dose administered. **(c)** First-pass extraction is 90% of dose administered.

sampling duration. But if studied long enough with good assay sensitivity, the metabolite curve will eventually be parallel to that of the parent as expected from the formation rate-limited metabolite kinetics.

One example to illustrate the impact of pre-systemically formed metabolite on the metabolite curve is shown in Fig. 7. In this example, midazolam was administered intravenously and orally, with 1′-hydroxy-midazolam and midazolam measured in plasma (Thummel *et al.*, 1996). After intravenous dosing, the concentration–time curves of 1′-hydroxy-midazolam in both subjects 4 and 8 were parallel to those of the parent drug, indicating formation rate-limited metabolite kinetics. After oral administration,

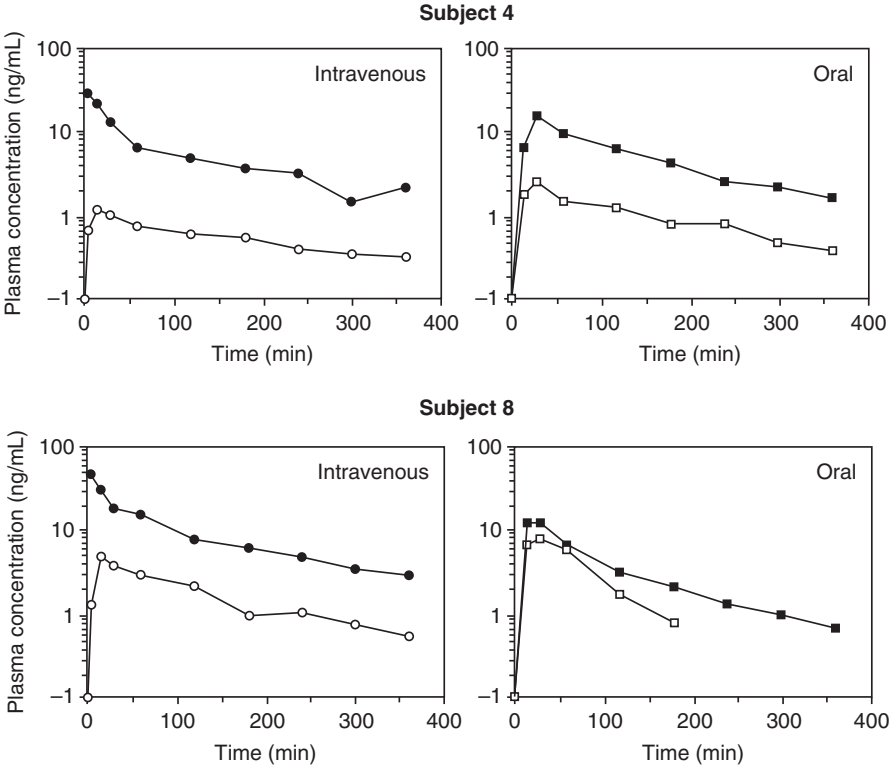


**Figure 7** The plasma concentration–time profiles of midazolam and 1′-hydroxy-midazolam in two human subjects following intravenous (1 mg) and oral (2 mg) doses of midazolam. Solid symbols correspond to midazolam and open symbols indicate 1′-hydroxy-midazolam (from Thummel *et al.*, 1996. Adapted with permission from Macmillan Publishers Ltd.).

however, the metabolite curve behaved differently between the two subjects. The metabolite curve in subject 4 was still parallel to that of the parent, whereas the metabolite curve in subject 8 was no longer parallel to the parent curve, with an apparent $t_{1/2}$ of the metabolite shorter than that of the parent at the time interval studied. The estimated M/P AUC ratio in subject 4 after intravenous and oral administration was 0.13 and 0.43, respectively. In subject 8, the corresponding M/P AUC ratio was 0.19 and 1.2, respectively. Clearly, the fold difference in the M/P AUC ratio between the two routes in subject 8 was bigger than that in subject 4, indicating more metabolite was produced pre-systemically (possibly in the gut wall) in subject 8 than in subject 4. As a result, the metabolite curve in subject 8 after oral administration was affected significantly. Consistent with this assessment is the fact that the oral bioavailability in subject 4 was in good agreement with the predicted hepatic availability, whereas the observed bioavailability in subject 8 was much lower than the one predicted from the hepatic availability, indicating a significant first-pass metabolism in the gut wall.

It is important to note again that, under linear conditions (i.e. no saturation in liver metabolism), hepatic metabolism is the same between the first pass and systemic cir-

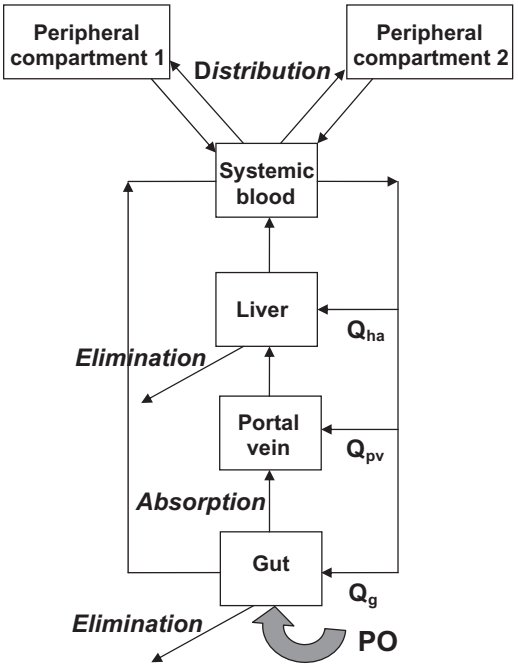


**Figure 8** A representative hybrid physiologically based pharmacokinetic model that is allowed to differentiate pre-systemic and systemic processes following oral administration of a drug. $Q_{pv}$: portal vein blood flow; $Q_{ha}$: hepatic artery blood flow; $Q_g$: gut blood flow.

culation. However, the gut wall may play a more important role in metabolism during the first pass than that during the systemic circulation following oral administration.

One of the challenges in using compartmental models to describe the metabolite kinetics influenced by pre-systemically formed metabolite is the fact that the conventional models cannot differentiate pre-systemic and systemic processes. To overcome this obstacle, one has to rely on physiologically based pharmacokinetic models (Vossen *et al.*, 2007). However, this also presents a challenge as we often do not have sufficient data in tissues to describe tissue partition coefficients. As a result, a semi-physiologically based pharmacokinetic model (or hybrid model) has been developed (Yamano *et al.*, 2001; Ito *et al.*, 2003; Yang *et al.*, unpublished data), where the gut and liver are taken out from the conventional compartments and set up in a physiologically and anatomically meaningful way linking the blood compartment. A representation of this kind of models is shown in Fig. 8. The model is of particular importance to our understanding of metabolite kinetics following oral administration.

### 2.6 *In Situ*-Generated versus Exogenously Dosed Metabolite

An *in situ*-generated metabolite refers to the metabolite formed in an eliminating organ after dosing of the parent drug. An exogenously dosed metabolite is the synthetic

standard of a metabolite administered to the body. This is often performed in order to determine the elimination clearance of the metabolite or assess pharmacological activity or toxicity associated with the metabolite. The exogenously dosed metabolite is also known as the preformed metabolite.

From a pharmacokinetic point of view, an *in situ*-generated metabolite may behave differently from an exogenously dosed metabolite (Pang, 1985; Prueksaritanont, Lin and Baillie, 2006). A number of factors contribute to the differences. First of all, metabolites formed, if not always, are more hydrophilic than the parent drug. As a result, the extent of distribution of an exogenously dosed metabolite into tissues may be different from that of *in situ*-generated metabolites. The differences also reside in the ability of the exogenously dosed metabolite in accessing the influx and efflux drug transporters and drug-metabolizing enzymes in eliminating organs for further metabolism and excretion. For an *in situ*-generated metabolite, its metabolism may be affected in the presence of the parent drug and vice versa, which cannot be reproduced by doing the metabolite exogenously. In addition, for formation rate-limited metabolite kinetics, the kinetics of an *in situ*-generated metabolite in the body is governed by the kinetics of the parent drug. Collectively, these kinetics differences between an *in situ*-generated and an exogenously dosed metabolite may translate to the exposure differences in the tissues that are important to the pharmacology and toxicology of the metabolite. Specific examples of this kind are nicely summarized by Prueksaritanont, Lin and Baillie (2006).

## 2.7 Sequential Metabolism

Many times, a primary metabolite is formed followed by sequential metabolism to produce a secondary metabolite. A typical example is hydroxylation followed by glucuronidation. Using a model similar to the one shown in Fig. 2, with an addition of the secondary metabolite that is eliminated through excretion into the urine or bile (Fig. 9), the following equations can be obtained:

$$\frac{AUC(m1)}{AUC} = f_m \cdot \frac{CL}{CL(m1)} \qquad (15)^*$$

$$\frac{AUC(m2)}{AUC} = f_m \cdot f_m(m1) \cdot \frac{CL}{CL(m2)} \qquad (16)^*$$

where AUC(m1), AUC(m2), and AUC are the AUCs of the primary metabolite, secondary metabolite, and parent drug, respectively; CL(m1), CL(m2), and CL are the elimination clearances of the primary metabolite, secondary metabolite, and parent drug, respectively; $f_m$ is the fraction of the total body clearance of the parent drug forming the primary metabolite; and $f_m(m1)$ is the fraction of the total body clearance of the primary metabolite forming the secondary metabolite. The metabolic ratio in this case equals:

$$\frac{Ae(m1) + Ae(m2)}{Ae} = \frac{CL_f}{CL_e} \qquad (17)^*$$

where Ae, Ae(m1), and Ae(m2) are the amounts of the parent drug, primary metabolite, and secondary metabolite in excreta, respectively. Similarly, the ratio of Ae(m2) to

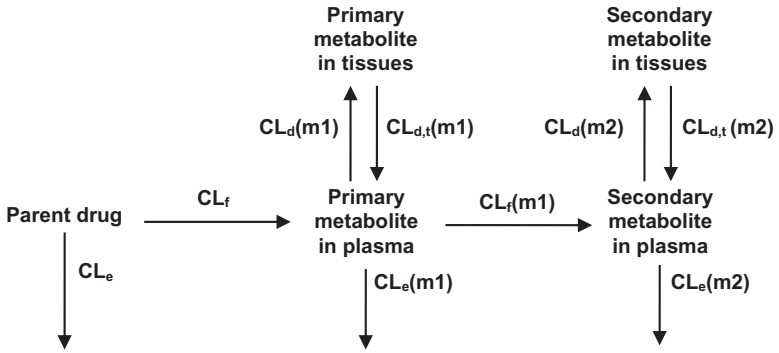


**Figure 9** A kinetic model for sequential metabolism. $CL_e$, $CL_e(m1)$, and $CL_e(m2)$ are excretory clearances of the parent drug, primary metabolite, and secondary metabolite, respectively; $CL_f$ and $CL_f(m1)$ are the formation clearance of the primary and secondary metabolites, respectively; $CL_d(m1)$ and $CL_{d,t}(m1)$ are the distributional clearance of the primary metabolite to and from tissues, respectively; $CL_d(m2)$ and $CL_{d,t}(m2)$ are the distributional clearance of the secondary metabolite to and from tissues, respectively.

Ae(m1) equals the ratio of the formation clearance of the secondary metabolite ($CL_f(m1)$) and the excretory clearance of the primary metabolite ($CL_e(m1)$) and is shown in the following equation:

$$\frac{Ae(m2)}{Ae(m1)} = \frac{CL_f(m1)}{CL_e(m1)} \tag{18*}$$

Once again, Equations 15–18 are valid only if the AUC is the total area of the curve from time zero to infinity and the recovery of the parent drug and metabolites is complete in excreta.

For sequential metabolism that involves a common secondary metabolite produced from two parallel, primary metabolic pathways, the kinetic processes are more complex than the scheme shown in Fig. 9. Interested readers can refer to the paper by Pang (1995) for detailed kinetic treatment.

## 3 EFFECT OF INHIBITION AND INDUCTION ON METABOLITE KINETICS

In a drug–drug interaction study, metabolite concentrations associated with an affected clearance pathway are often measured. Intuitively, one would think that, when the formation clearance is inhibited, that metabolite concentrations and AUC would decrease, and when the formation clearance is induced, the metabolite concentration and AUC would increase. However, this may not be entirely true based on the metabolite kinetic principles. Considering the model scheme shown in Fig. 2, one could rearrange Equation 7 and obtain the following equation:

$$AUC(m) = f_m \cdot \frac{CL \cdot AUC}{CL(m)} \tag{19}$$

Recalling that drug dose equals CL·AUC, the following equation exists:

$$AUC(m) = f_m \cdot \frac{Dose}{CL(m)} \quad (20)^*$$

By the same token, the following equation can be obtained at steady state:

$$C_{ss}(m) = f_m \cdot \frac{Dose/\tau}{CL(m)} \quad (21)^*$$

where $\tau$ is the dosing interval. Equations 20 and 21 indicate that the AUC and $C_{ss}$ of a metabolite are a function of dose, $f_m$, and elimination clearance of the metabolite. When $f_m$ is ~1.0, meaning that the formation of the metabolite is a predominant elimination pathway of the parent drug, then Equations 20 and 21 can be simplified to the following equations:

$$AUC(m) = \frac{Dose}{CL(m)} \quad (22)$$

$$C_{ss}(m) = \frac{Dose/\tau}{CL(m)} \quad (23)$$

In this case, the AUC and $C_{ss}(m)$ of the metabolite would not change regardless of whether its formation clearance is inhibited or induced. Conceptually, this makes sense. When the formation clearance, as a predominant clearance pathway of the parent drug, is inhibited, the parent drug concentration would increase. The rate of formation, however, would not change, as it equals the product of the formation clearance and parent drug concentrations. As a result, AUC and $C_{ss}(m)$ of the metabolite would not change regardless of what happens to its formation clearance. However, the M/P AUC or $C_{ss}$ ratio would change because of the changes in the parent drug concentrations. This is best illustrated through the simulations shown in Fig. 10a. In this case, the formation clearance of the metabolite is the total body clearance of the parent drug (i.e. $f_m = 1$). When the formation clearance is inhibited by 50%, the average parent drug concentrations at steady state increase by twofold as expected. On the one hand, the average metabolite concentration at steady state remains the same. There are many clinical drug–drug interactions that exemplify the case cited. For example, in the presence of rifampicin, the plasma AUC of midazolam after oral administration reduced from 35.8 to 3.7 μg/h/mL (Gorski *et al.*, 2003). On the other hand, the average urinary recovery of 1′-hydroxy-midazolam before and after rifampicin treatment was comparable (74% versus 64%), indicating that the extent of formation of 1′-hydroxy-midazolam, the major elimination pathway of midazolam in humans, is similar in the presence and absence of rifampicin.

When the $f_m$ no longer approaches 1, the AUC and $C_{ss}(m)$ of the metabolite would change when the formation clearance of the metabolite is affected. The extent of change is dependent on the change in the $f_m$. Figure 10b shows a compound that has an $f_m$ of 0.8 for the formation of metabolite A. In this case, when the formation clearance of metabolite A is inhibited by 50%, the average steady-state concentrations of the parent drug are increased by 1.7-fold. Notice that the change is not twofold due

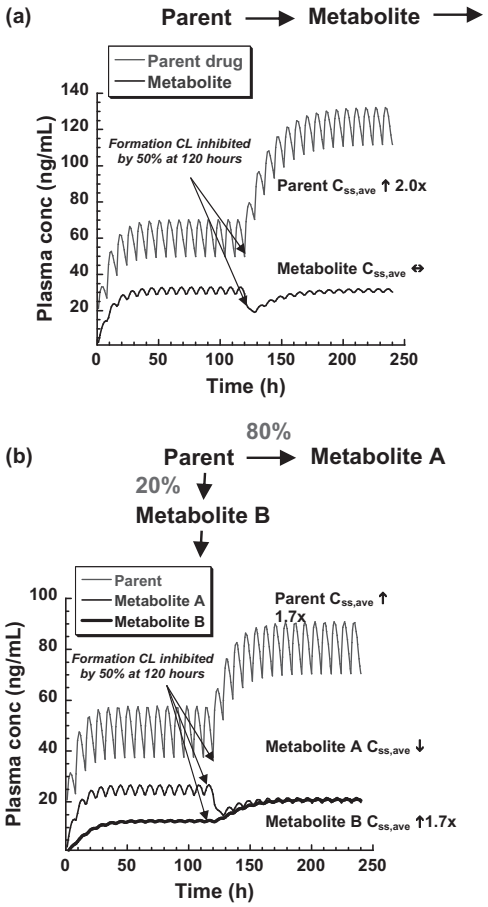


**Figure 10** Inhibition of the formation clearance of a metabolite on metabolite kinetics at steady state during a multiple-dose regimen. **(a)** $f_m = 1$; **(b)** $f_m = 0.8$.

to a shunting effect to another pathway (metabolite B). Accordingly, the average steady-state concentration of metabolite B is also increased by 1.7-fold. The average steady-state concentration of metabolite A, of which the formation clearance is inhibited, was marginally decreased by 17%. Clearly, the existence of additional pathways may significantly attenuate the extent of inhibition or induction on a particular pathway.

A sensitive way to detect the change in an affected formation clearance pathway is to compare the M/P AUC or $C_{ss}$ ratio in the presence and absence of an inhibitor or inducer. Recall from Equation 5 that:

$$\frac{\mathrm{AUC(m)}}{\mathrm{AUC}} = \frac{\mathrm{CL_f}}{\mathrm{CL(m)}}$$

Assuming no change in the CL(m) in the presence of an inhibitor or inducer, Equation 5 becomes:

$$\left\{\frac{AUC(m)}{AUC}\right\}_{affected} = \frac{CL_{f,affected}}{CL(m)} \tag{24}$$

Therefore, the M/P AUC ratio in the presence and absence of an inhibitor or inducer is:

$$\frac{\left\{\frac{AUC(m)}{AUC}\right\}_{affected}}{\left\{\frac{AUC(m)}{AUC}\right\}} = \frac{CL_{f,affected}}{CL_f} \tag{25*}$$

Equation 25 indicates that a change in the M/P AUC ratio directly reflects the change in the metabolite formation clearance.

In addition to the M/P plasma AUC ratio, the metabolic ratio (Ae(m)/Ae) in excreta can be used to assess the alteration in the formation clearance of a metabolite. Recall from Equation 12 that:

$$\frac{Ae(m)}{Ae} = \frac{CL_f}{CL_e}$$

Assuming no change in the $CL_e$ (renal or biliary clearance) of the parent drug in the presence and absence of an inhibitor or inducer, the following equation exists:

$$\frac{\left\{\frac{Ae(m)}{Ae}\right\}_{affected}}{\left\{\frac{Ae(m)}{Ae}\right\}} = \frac{CL_{f,affected}}{CL_f} \tag{26*}$$

Again, like the M/P AUC ratio, a change in the metabolic ratio directly reflects the change in the metabolite formation clearance.

To appreciate these effects, simulations were conducted to examine the effects of $f_m$ and inhibition of the metabolite formation clearance on the metabolite AUC, parent AUC, and M/P AUC (or metabolic) ratio (Fig. 11). It is apparent from the graphs that the higher the inhibition of metabolite formation clearance, the more affected the metabolite AUC. However, the $f_m$ attenuates this effect as discussed earlier. The higher the $f_m$ value, the less affected the metabolite AUC even when the formation clearance is inhibited by ~100%. On the other hand, the higher the $f_m$, the more affected the parent AUC. Lastly, the change in the M/P AUC ratio or metabolic ratio is inversely related to the extent of the inhibition of the metabolite formation clearance and is independent of the $f_m$ as shown in Equations 25 and 26. And this change largely results from the alteration in the parent AUC or Ae. As the $f_m$ approaches unity, the metabolite AUC or Ae(m) will be largely unchanged, assuming that the inhibitor or inducer does not affect the elimination clearance of the metabolite.

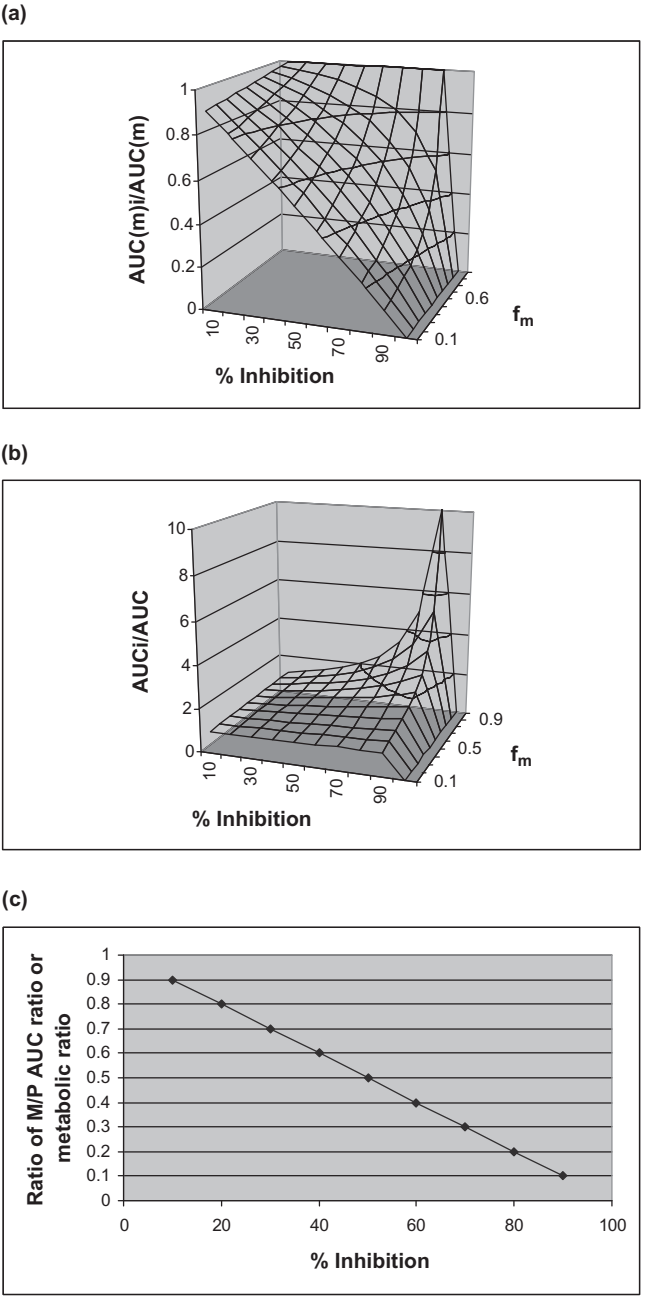


**Figure 11** Sensitivity of various pharmacokinetic parameters to the inhibition of the metabolite formation clearance as a function of $f_m$. **(a)** Metabolite AUC. **(b)** Parent AUC. **(c)** M/P AUC ratio or metabolic ratio.

## 4 DETERMINATION OF FORMATION AND ELIMINATION CLEARANCE OF METABOLITE

The elimination clearance (CL(m)) of a metabolite can be determined by intravenous dosing of a synthetic standard of the metabolite, with a caveat that disposition kinetics may be different between an *in situ*-generated metabolite and an exogenously dosed one (Pang, 1985; Prueksaritanont, Lin and Baillie, 2006). Once the CL(m) is determined, the formation clearance ($CL_f$) of the metabolite can be determined based on the M/P AUC or $C_{ss}$ ratio as shown in Equation 5 or Equation 6.

However, sometimes, a synthetic standard of the metabolite is not available or not administered intravenously to obtain the CL(m). In this case, one could still make a case to estimate the CL(m) and $CL_f$ from dosing the parent drug alone, provided that a number of assumptions are met and metabolite and parent drug data are available in plasma and excreta. In one scenario, one could estimate the CL(m) from the metabolite data in plasma and excreta. Considering the model scheme shown in Fig. 2, it can be shown that:

$$CL(m) = \frac{Ae(m)}{AUC(m)} \qquad (27)^*$$

In this case, the recovery of a metabolite in excreta does not have to be complete, as long as the collection time interval for excreta is the same as that for the AUC of the metabolite. The key assumptions, however, are that the metabolite produced in the body is entirely available to the systemic circulation, as if it was dosed systemically, and there is no sequential metabolism. Conceivably, the CL(m) can be accurately determined from Equation 27 if the metabolite is produced in the liver and/or gut and is eliminated exclusively into the urine unchanged. However, if there is a significant sequential metabolism involved, Equation 27 would underestimate the true CL(m), as Ae(m) does not account for the entire amount eliminated. In addition, if a metabolite is produced in the liver but eliminated significantly via biliary secretion, Equation 27 would overestimate the true CL(m), as the metabolite in the bile may come directly from the liver. Once the CL(m) is determined, the $CL_f$ of the metabolite can be determined from the M/P AUC or $C_{ss}$ ratio using Equation 5 or Equation 6.

In another scenario, if the formation of the metabolite represents a predominant pathway of the metabolic clearance of the parent drug, one could estimate the $CL_f$ of the metabolite by subtracting the total body clearance of the parent drug from its excretory clearance. Alternatively, the $CL_f$ can be calculated from the product of the metabolic ratio and the excretory clearance of the parent drug using Equation 12, with additional assumptions that the recovery of the parent drug and metabolite is complete and there is no sequential metabolism. The excretory clearance of the parent drug can be calculated as the amount of the parent drug in excreta divided by its plasma AUC estimated at the time interval as the excreta. Once the $CL_f$ is determined, the CL(m) of the metabolite can be determined from the M/P AUC or $C_{ss}$ ratio using Equation 5 or Equation 6.

The situation could be further complicated if metabolite information is not available in plasma. In this case, with a number of assumptions, one could still estimate the $CL_f$ of the metabolite but not the CL(m). Recalling that drug dose

equals CL·AUC and $CL_f$ equals $f_m$·CL, the following equation exists based on Equation 11:

$$CL_f = \frac{Ae(m)}{AUC} \quad (28)^*$$

where AUC is the total area under the curve of the parent drug. It is important to note that Equation 28 is valid only if the recovery of the metabolite in excreta is complete and there is no sequential metabolism of the metabolite. If the recovery of the metabolite in excreta is incomplete (e.g. true recovery is 50% of the dose versus observed recovery is 25% over a finite period of time), it may lead to an underestimation of the $CL_f$. In addition, for orally administered drugs, one has to assume that the metabolite is largely formed after the parent drug reaches the systemic circulation (i.e. minimal first-pass metabolism). Violating this assumption leads to an overestimation of the $CL_f$. Alternatively, as discussed earlier, the $CL_f$ can be calculated from the product of the metabolic ratio and the excretory clearance of the parent drug using Equation 12, if both parent and metabolite data are available in excreta. In this case, the assumptions on complete recovery in excreta and no sequential metabolism still apply. In addition, the formation of the metabolite is assumed to represent a predominant metabolic pathway of the parent drug.

For sequential metabolism, the same principles along with the assumptions discussed before can be applied to determine the formation ($CL_f$) and elimination clearance (CL(m1)) of the primary metabolite (model scheme shown in Fig. 9). With plasma and excreta data available for the parent drug, primary metabolite, and secondary metabolite, the $CL_f$(m1), $CL_e$(m1), and CL(m1) of the primary metabolite are calculated as follows:

$$CL_f(m1) = \frac{Ae(m2)}{AUC(m1)} \quad (29)$$

$$CL_e(m1) = \frac{Ae(m1)}{AUC(m1)} \quad (30)$$

$$CL(m1) = \frac{Ae(m1) + Ae(m2)}{AUC(m1)} \quad (31)^*$$

The key assumptions are (i) the primary metabolite produced in the body entirely available in the systemic circulation, (ii) no sequential metabolism of the secondary metabolite, and (iii) complete recovery of metabolites in excreta. As a result, the $CL_f$ calculated from Equation 29 always overestimates the true $CL_f$ as the secondary metabolite can form locally in the liver upon the formation of the primary metabolite. The $CL_e$ can also be overestimated using Equation 30 if biliary secretion is a part of the excretory clearance pathways of the primary metabolite. Once the CL(m1) is known, the $CL_f$ of the primary metabolite can be determined from the M/P AUC or $C_{ss}$ ratio between the primary metabolite and the parent drug using Equation 5 or Equation 6. In the case of lack of the primary metabolite data in plasma, one

could determine the $CL_f$ but not CL(m1) of the primary metabolite using the following equation:

$$CL_f = \frac{Ae(m1) + Ae(m2)}{AUC} \tag{32}$$

In this case, the assumptions of no sequential metabolism of the secondary metabolite and complete recovery of metabolites in excreta also apply. In addition, for orally administered drugs, minimal first-pass metabolism is assumed.

## 5 SUMMARY

*In vivo* drug metabolite kinetics is a complex process in part because it is governed by both formation and elimination of the metabolite(s) of interest. As a result, the M/P AUC or $C_{ss}$ ratio is a function of both the formation and the elimination clearance of the metabolite. The conceptual difference between a major circulating metabolite and the product of a major metabolic pathway needs to be appreciated in the safety testing of metabolites observed in circulation and excreta. Also, the rate of metabolism is not necessarily related to the extent of metabolism, as a low rate of metabolism can still lead to extensive metabolism if no other pathways exist. The rate-limiting step in the formation and elimination of a metabolite is a key determinant to the pharmacokinetic characteristics of the metabolite, including the shape of curve, rate of removal from the body, time to reach steady state, and extent of accumulation. Furthermore, the extent of metabolites formed pre-systemically in the gut wall/lumen following oral administration of a parent drug could have a significant impact on the metabolite kinetics and the shape of the metabolite curve. In addition, a change in the metabolic ratio or M/P plasma AUC ratio directly reflects the change in the formation clearance of a metabolite, whereas an alteration in the metabolite or parent AUC is also a function of the $f_m$. Overall, a clear understanding and appreciation of metabolite kinetic principles is essential to assess the pharmacology and toxicology of metabolites and to guide the better use of therapeutic agents in the clinic.

## ACKNOWLEDGMENTS

The author would like to thank Drs. Punit Marathe and A. David Rodrigues for their critical review of the book chapter and valuable comments and suggestions.

## REFERENCES

Baillie, T.A., Cayen, M.N., Fouda, H., Gerson, R.J., Green, J.D., Grossman, S.J., Klunk, L.J., LeBlanc, B., Perkins, D.G., Shipley, L.A. (2002). Drug metabolites in safety testing. *Toxicol. Appl. Pharmacol.*, *182*, 188–196.

Christians, U., Jacobsen, W., Floren, L.C. (1998). Metabolism and drug interactions of 3-hydroxy-3-methylglutaryl coenzyme A reductase inhibitors in transplant patients: are the statins mechanistically similar? *Pharmacol. Ther.*, *80*, 1–34.

Colizza, K., Awad, M., Kamel, A. (2007). Metabolism, pharmacokinetics, and excretion of the substance P receptor antagonist CP-122,721 in humans: structural characterization of the novel major circulating metabolite 5-trifluoromethoxy salicylic acid by high-performance liquid chromatography-tandem mass spectrometry and NMR spectroscopy. *Drug Metab. Dispos.*, *35*, 884–897.

Davis-Bruno, K.L., Atrakchi, A. (2006). A regulatory perspective on issues and approaches in characterizing human metabolites. *Chem. Res. Toxicol.*, *19*, 1561–1563.

FDA Guidance for Industry: Safety Testing of Drug Metabolite. February 2008. http://www.fda.gov/CDER/GUIDANCE/6897fnl.pdf.

Davis-Bruno, K.L., Atrakchi, A. (2006). A regulatory perspective on issues and approaches in characterizing human metabolites. *Chem. Res. Toxicol.*, *19*, 1561–1563.

Fura, A. (2006). Role of pharmacologically active metabolites in drug discovery and development. *Drug Discov. Today*, *11*, 133–141.

Fura, A., Shu, Y.-Z., Zhu, M., Hanson, R.L., Roongta, V., Humphreys, W.G. (2004). Discovering drugs through biological transformation: role of pharmacologically active metabolites in drug discovery. *J. Med. Chem.*, *47*, 4339–4351.

Gorski, J.C., Vannaprasaht, S., Hamman, M.A., Ambrosius, W.T., Bruce, M.A., Haehner-Daniels, B., Hall, S.D. (2003). The effect of age, sex, and rifampin administration on intestinal and hepatic cytochrome P450 3A activity. *Clin. Pharmacol. Ther.*, *74*, 275–287.

Hastings, K.L., El-Hage, J., Jacobs, A., Leighton, J., Morse, D., Osterberg, R.E. (2003). Letter to the editor on drug metabolites in safety testing. *Toxicol. Appl. Pharmacol.*, *190*, 91–92.

Houston, J.B. (1982). Drug metabolite kinetics. *Pharmacol. Ther.*, *15*, 521–552.

Humphreys, W.G., Unger, S.E. (2006). Safety assessment of drug metabolites: characterization of chemically stable metabolites. *Chem. Res. Toxicol.*, *19*, 1564–1569.

Ito, K., Ogihara, K., Kanamitsu, S., Itoh, T. (2003). Prediction of the in vivo interaction between midazolam and macrolides based on in vitro studies using human liver microsomes. *Drug Metab. Dispos.*, *31*, 945–954.

Paine, M.F., Shen, D.D., Kunze, K.L., Perkins, J.D., Marsh, C.L., McVicar, J.P., Barr, D.M., Gillies, B.S., Thummel, K.E. (1996). First-pass metabolism of midazolam by the human intestine. *Clin. Pharmacol. Ther.*, *60*, 14–24.

Pang, K.S. (1985). A review of metabolite kinetics. *J. Pharmacokinet. Biopharm.*, *13*, 633–662.

Pang, K.S. (1995). Kinetics of sequential metabolism: contribution of parallel, primary metabolic pathways to the formation of a common secondary metabolite. *Drug Metab. Dispos.*, *23*, 166–177.

Pereillo, J.M., Maftouh, M., Andrieu, A., Uzabiaga, M.F., Fedeli, O., Savi, P., Pascal, M., Herbert, J.M., Maffrand, J.P., Picard, C. (2002). Structure and stereochemistry of the active metabolite of clopidogrel. *Drug Metab. Dispos.*, *30*, 1288–1295.

Prueksaritanont, T., Lin, J.H., Baillie, T.A. (2006). Complicating factors in safety testing of drug metabolites: kinetic differences between generated and preformed metabolites. *Toxicol. Appl. Pharmacol.*, *217*, 143–152.

Roffey, S.J., Obach, R.S., Gedge, J.I., Smith, D.A. (2007). What is the objective of the mass balance study? A retrospective analysis of data in animal and human excretion studies employing radiolabeled drugs. *Drug Metab. Rev.*, *39*, 17–43.

Rowland, M., Tozer, T.N. (1995). *Clinical Pharmacokinetics: Concepts and Applications*, 3rd edn, Williams & Wilkins, Baltimore, MD.

Savi, P., Pereillo, J.M., Uzabiaga, M.R., Combalbert, J., Picard, C., Maffrand, J.P., Pascal, M., Herbert, J.M. (2000). Identification and biological activity of the active metabolite of clopidogrel. *Thromb. Haemost.*, *84*, 891–896.

Siedlik, P.H., Olson, S.C., Yang, B.B., Stern, R.H. (1999). Erythromycin coadministration increases plasma atorvastatin concentrations. *J. Clin. Pharmacol.*, *39*, 501–504.

Smith, D.A., Obach, R.S. (2005). Seeing through the mist: abundance versus percentage. Commentary on metabolites in safety testing. *Drug Metab. Dispos.*, *33*, 1409–1417.

Smith, D.A., Obach, R.S. (2006). Metabolites and safety: what are the concerns, and how should we address them? *Chem. Res. Toxicol.*, *19*, 1570–1579.

Thummel, K.E., O'Shea, D., Paine, M.F., Shen, D.D., Kunze, K.L., Perkins, J.D., Wilkinson, G.R. (1996). Oral first-pass elimination of midazolam involves both gastrointestinal and hepatic CYP3A-mediated metabolism. *Clin. Pharmacol. Ther.*, *59*, 491–502.

Thummel, K.E., Shen, D.D. (2002). The role of the gut mucosa in metabolically based drug-drug interactions, in *Drug–Drug Interactions* (ed. A.D. Rodrigues), Marcel Dekker, New York, pp. 359–385.

Thummel, K.E., Shen, D.D., Isoherranen, N., Smith, H.E. (2006). Design and optimization of dosage regimens: pharmacokinetic data, in *Goodman & Gilman's The Pharmacological Basis of Therapeutics*, 11th edn (eds L.L. Brunton, J.S. Lazo, K.L. Parker), McGraw-Hill, New York, pp. 1787–1888.

Vossen, M., Sevestre, M., Niederalt, C., Jang, I.J., Willmann, S., Edginton, A.N. (2007). Dynamically simulating the interaction of midazolam and the CYP3A4 inhibitor itraconazole using individual coupled whole-body physiologically-based pharmacokinetic (WB-PBPK) models. *Theor. Biol. Med. Model.*, *4*, 13. doi:10.1186/1742-4682-4-13.

Wang, L., Zhang, D., Swaminathan, A., Xue, Y., Cheng, P.T., Wu, S., Mosqueda-Garcia, R., Aurang, C., Everett, D.W., Humphreys, W.G. (2006). Glucuronidation as a major metabolic clearance pathway of 14c-labeled muraglitazar in humans: metabolic profiles in subjects with or without bile collection. *Drug Metab. Dispos.*, *34*, 427–439.

Wilkinson, G.R. (2002). In vivo probes for studying induction and inhibition of cytochrome P450 enzymes in humans, in *Drug–Drug Interactions* (ed. A.D. Rodrigues), Marcel Dekker, New York, pp. 439–504.

Williams, D., Feely, J. (2002). Pharmacokinetic-pharmacodynamic drug interactions with HMG-CoA reductase inhibitors. *Clin. Pharmacokinet.*, *41*, 343–370.

Yamano, K., Yamamoto, K., Katashima, M., Kotaki, H., Takedomi, S., Matsuo, H., Ohtani, H., Sawada, Y., Iga, T. (2001). Prediction of midazolam-CYP3A inhibitors interaction in the human liver from in vivo/in vitro absorption, distribution, and metabolism data. *Drug Metab. Dispos.*, *29*, 443–452.

# Pharmacogenetics and Pharmacogenomics

LILIAN G. YENGI

## 1 INTRODUCTION: PHARMACOGENETICS AND PHARMACOGENOMICS

The fields of pharmacogenetics, genomics, and drug transporters have profoundly impacted drug metabolism research by providing plausible mechanisms for interindividual variability in drug response and metabolism-related toxicity. They have provided tools with which to understand enzyme regulation, identify factors that affect drug exposure, the potential for drug–drug interactions, and species differences in drug disposition. Knowledge from these fields is being used to form the scientific basis for designing appropriate clinical studies and data interpretation, leading to the development and use of safer and more efficacious drugs.

Pharmacogenetics is the study of the effects of genetic differences between individuals on interindividual responses to medicines (Ginsburg *et al.*, 2005; Eichelbaum, Ingelman-Sundberg and Evans, 2006). It is an old discipline, which has been invigorated by an increased understanding of molecular biology and development of associated technological tools with which to study these mechanisms (Goldstein, Tate and Sisodiya, 2003). In its simplest form, genetics is the scientific study of heredity, dating back to Mendel, who showed that the inheritance pattern of certain traits in pea plants followed simple statistical rules and described a fundamental unit of heredity, which he called allele. While Mendel used *allele* to refer to what we now know as gene, allele is currently used to denote a specific variant of a particular gene. Pharmacogenetics can therefore be viewed as a more specific and narrow subfield of genetics. Though an old discipline, the term having been around since the 1950s has enjoyed a new lease of life in the last 10–20 years, mainly due to advances in our understanding of the impact of genetics on efficacy and safety of drugs, emergence of global tools with which to study multiple genes at any given time, and the completion of the human genome project.

Pharmacogenomics is a more recent term, coined to define a more holistic or global approach, in which the expression levels, regulation, functions, and interactions of

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

multiple genes are simultaneously studied, and their effects on overall variability in drug response determined (Nebert, 1999). It is sometimes used interchangeably with pharmacogenetics, despite these subtle differences. These fields are continually being integrated in various aspects of the life sciences, including drug discovery and development, with the expectations that they would lead to the development of safer drugs that can be tailored to subsets of patients based on their genetic makeup, the so-called personalized medicines (Wolf, Smith and Smith, 2000; Hedgecoe and Martin, 2003). Genome-wide screening is used to identify single nucleotide polymorphisms (SNPs) that co-segregate with certain diseases with a view to using these SNP regions as potential targets for drug development, and global gene expression analysis is used preclinically and clinically to identify surrogate markers of toxicity (toxicogenomics) and/or efficacy (Searfoss *et al.*, 2003). That individuals differ in their response to drug use is not new and there have always been suggestions that a genetic component could be responsible. However, this could not be directly proven due to lack of appropriate technologies. Nevertheless, differences in plasma levels of antidepressants such as nortriptyline and imiprimine were attributed to genetic factors (Hammer and Sjoqvist, 1967; Alexanderson, Evans and Sjoqvist, 1969) and impact of genetic variations in the toxicity of drugs such as the antitubercular drug Isoniazid, were postulated, and then proven as early as the 1950s (Bonicke and Reif, 1953; Evans, Manley and Mc, 1960). The CYP2D6 polymorphism, which resulted in differences in the metabolism of debrisoquine among different individuals, is a more recent and famous example of the impact of genetic variations on exposure and disposition drugs (Mahgoub *et al.*, 1977; Eichelbaum, 1986). Many functional polymorphisms have now been identified in many gene families, with the variant alleles occurring with varying frequencies among different ethnic populations (Garte *et al.*, 2001; Ingelman-Sundberg, Daly and Nebert, 2007). Those polymorphisms with major implications on the development and clinical applications of major pharmaceutical drugs will be discussed later in this chapter.

## 2 PHARMACOGENOMICS OF DRUG-METABOLIZING ENZYMES (DME)

Interest in gene regulation and its potential impact on the efficacy and safety of new drug candidates is steadily increasing. One of the consequences of transcriptional regulation is clinical drug–drug interactions with coadministered drugs and also the observation that some drugs induce the enzymes involved in their metabolism, hence altering their kinetic properties, including those compounds still in development, where exposures could alter following multiple dosing in safety studies. It is therefore important that the effects of new drug compounds on the expression of drug metabolism genes form an integral part of drug metabolism studies, both preclinically and during clinical development. The discovery and characterization of the cytochrome P450s (CYPs) in the 1950s, their solubilization, and elucidation of the mechanisms underlying their regulation are critical milestones in our understanding of the biotransformation of xenobiotics including pharmaceutical drugs. Although it had been known since their characterization that some P450s were inducible, it was knowledge of the interactions between steroid hormones and their targets within the cell that led to the hypothesis that induction of these microsomal enzymes by ligands was via a mechanism similar to that of steroid hormones. This hypothesis was later supported in subsequent experiments that eventually led to the characterization and identification of the aryl hydrocarbon hydroxylase (Ah) receptor as responsible for CYP1A1

induction (Poland, Glover and Kende, 1976; Guenthner and Nebert, 1977). Several nuclear receptors have now been cloned and the downstream genes they regulate have been identified (Bertilsson *et al.*, 1998; Blumberg *et al.*, 1998; Kliewer *et al.*, 1998). These include pregnane X receptor PXR, with CYP3A4, CYP2C, and UGT as the prototypic downstream genes, constitutive androstane receptor (CAR), which regulates genes in the CYP2B family and UGTs, liver X receptor (LXR), the master regulator of cholesterol homeostasis, farnesoid X receptor (FXR), peroxisome proliferator-activated receptor (PPAR), vitamin D receptor (VDR), glucocorticoid receptor (GR), and hepatocyte nuclear receptor 4 (HNF4). These receptors control a complex network of endogenous pathways and often exhibit cross talk between them in relation to their ligands and downstream regulated genes (Kliewer, Lehmann and Willson, 1999; Repa and Mangelsdorf, 2000; Rushmore and Kong, 2002). It is therefore not surprising that regulation of nuclear receptors by drug candidates has the potential not only to impact targeted pathways, but also affect off-target genes leading to unexpected efficacy and/or safety outcomes. The role of nuclear hormone receptors in drug metabolism and drug–drug interactions will be discussed in detail in later chapters of this book.

## 3 POLYMORPHISMS IN GENES ENCODING DME AND IMPLICATIONS FOR DRUG DISCOVERY AND DEVELOPMENT

Following the concept of two-phase elimination of xenobiotics proposed by R.T. Williams, reactions such as oxidation, reduction, and hydrolysis are considered phase I and activating, while conjugation reactions are considered phase II and detoxifying in nature. Although other phase I enzyme systems exist, the CYPs have emerged as the principal enzymes catalyzing most phase I reactions, while phase II reactions are catalyzed by several gene families including uridine diphosphate glucuronosyltransferase (UGT), glutathione transferases (GSTs), *N*-acetyltransferase (NAT), and sulfotransferases (SULTs). Phase I reactions are often the initial stage in converting often highly lipophilic environmental chemicals, which cannot readily be excreted from the body, to more water-soluble, hydrophilic species (Guengerich, 1991b; Goldstein and Faletto, 1993). Phase II enzymes then catalyze the attachment of highly water-soluble moieties to these intermediate products, thus increasing their hydrophilicity. The net result of phase I and phase II reactions is generally to detoxify and eliminate the compound from the body, although they can also result in activation of some compounds to their ultimate toxic and/or carcinogenic end products (Guengerich, 1991a, b; Goldstein and Faletto, 1993). Research over the years has shown that functional polymorphisms in both phase I and II enzymes are remarkably common and some have been associated with disease susceptibility and prognoses, and interindividual variability in drug response and toxicity. The following section will review some of the functional polymorphisms in DME that impact the safety and/or efficacy of drugs.

### 3.1 P450 Polymorphisms

Although it is now common knowledge that many P450s exhibit genetic polymorphisms with implications for drug efficacy and safety, the first direct association between CYP polymorphisms and alterations in the pharmacokinetics of a drug was not made

until the late 1970s and 1980s. Many of the variations were first identified by the occurrence of adverse reactions in patients or volunteers taking normal doses of drugs later found to be metabolized primarily by polymorphic CYPs. Despite the fact that more than 50 human CYPs have been identified, only genes in the first three families (CYP1 to 3) appear to be responsible for the metabolic transformation of most drugs and xenobiotics. Hence polymorphisms in genes belonging to these families are likely to be relevant in determining the safety and efficacy of most pharmaceuticals. Indeed many of these genes are functionally polymorphic, with different allelic frequencies among different ethnic populations.

CYP1A1 is one of the few major CYPs involved in the metabolism of xenobiotics that appears to have been conserved from an evolutionary perspective. CYP1A1 is constitutively expressed at very low levels predominantly in extrahepatic tissues but is highly induced in response to several endogenous and exogenous ligands. About 20 single nucleotide polymorphisms have been identified at the CYP1A1 locus, with 10 resulting in amino acid substitutions (Daly *et al.*, 1993; Saito *et al.*, 2003; Ingelman-Sundberg, Daly and Nebert, 2007). However, none of them significantly affect protein function and therefore have very little impact on the metabolism of CYP1A1 drug substrates, although some of the allelic variants have been found to be in linkage disequilibrium with certain diseases including various cancers (Taioli *et al.*, 2003; Vineis *et al.*, 2004). This evolutionary conservation in CYP1A1 sequence suggests a critical endogenous role for CYP1A1, although endogenous substrates for CYP1A1 have not been identified. However, CYP1A1 is regulated by the Ah receptor, which has an important role in cell cycle and it has been hypothesized that CYP1A1 might be an important mediator in some cell types (Ingelman-Sundberg, 2001).

CYP1A2 is expressed almost exclusively in the liver and constitutes approximately 12% of total hepatic CYP content. It is involved in the metabolism of various toxic and carcinogenic compounds and several drugs, including caffeine, clozapine (where CYP1A2 has a major role), fluvoxamine, olanzapine, tacrine, theophylline, and acetaminophen (Murayama *et al.*, 2004; Chetty and Murray, 2007). Although CYP1A2-mediated metabolism is often a minor pathway in the metabolism of many drugs, it may become critical in the absence of the major contributing CYP or when the major isoform is saturated (e.g. in CYP2D6 poor metabolizers). Increased abundance of CYP1A2 due to induction may also result in clinically significant interactions. In addition, constitutive expression of CYP1A2 mRNA levels in human liver samples can vary by as much as 15-fold and the protein by up to 60-fold (Shimada *et al.*, 1994; Hammons *et al.*, 2001). Therefore, CYP1A2 polymorphisms have the potential to alter the disposition of many drugs including those in which it only plays a minor role. Approximately 40 single nucleotide variant alleles have been identified at the CYP1A2 locus, five of which result in alterations to the amino acid sequence of the encoded protein (Ingelman-Sundberg, Daly and Nebert, 2007). The 60-fold interindividual variation detected in the activity of CYP1A2 protein may be the phenotypic manifestation of these polymorphisms (Rasmussen *et al.*, 2002; Pucci *et al.*, 2007). A transition polymorphism (C to A) in the first intron of CYP1A2 has been associated with variations in CYP1A2 inducibility in healthy volunteers who are cigarette smokers (Sachse *et al.*, 1999) and has been postulated to be responsible for tardive dyskinesia (TD), a common and potentially irreversible side effect associated with long-term treatment with antipsychotics (Basile *et al.*, 2000). However, studies in a Turkish population found no association between this polymorphism and schizophrenic patients with TD (Boke *et al.*, 2007). A polymorphism (G to A) in the 5′-flanking region of *CYP1A2* in Japanese subjects

has been associated with a significant decrease in CYP1A2-mediated demethylation of caffeine (Nakajima *et al.*, 1999). The implications of this and other CYP1A2 polymorphisms in relation to clinical drug interactions remain to be ascertained.

CYP2D6 was the first human CYP isozyme identified to be polymorphic with clinical drug implication. In the 1970s, it had been known that the metabolism of certain drugs such as debrisoquine and spartein showed interindividual variations as evidenced by different plasma levels of the drugs (Eichelbaum *et al.*, 1979). About a decade later, several scientists were able to show that these differences were due to genetic differences in the enzyme responsible for the metabolism of these drugs. This gene was later identified as CYP2D6, located on the long arm of chromosome 22. Because of this unique place in the history of pharmacogenetics, the CYP2D6 locus has been extensively studied and to date, more than 70 alleles have been identified, with different frequencies occurring in different ethnic populations (Ikenaga *et al.*, 2005; Beverage *et al.*, 2007; Ingelman-Sundberg, Daly and Nebert, 2007). Some of these allelic variants produce nonfunctional proteins, some have reduced function, and others have no effect on phenotype. There are also individuals with multiple alleles resulting in an increase in function. Phenotypically, these are respectively called poor metabolizers (PMs), intermediate metabolizers (IMs), extensive metabolizers (EMs), and ultrarapid metabolizers (UMs). The PM phenotype appears to be more common among Europeans (3–11%), low in African populations, and 1–2% in Asians (Scordo *et al.*, 2004). The UM phenotype is the most common phenotype in Ethiopians, with a reported frequency of up to 39% (Daly *et al.*, 1993; Aklillu *et al.*, 1996). The importance of CYP2D6 polymorphisms in pharmaceutical development is evidenced by the long list of drugs whose efficacy and safety is impacted by CYP2D6 polymorphisms (Meyer *et al.*, 1996; Ingelman-Sundberg, 2001).

The importance of CYP2C8 in the metabolism of several drugs is becoming evident; hence interest in CYP2C8 polymorphisms and impact on drug disposition and clinical drug interactions is increasing. CYP2C8 plays a major role in the metabolism of the thiazolidinediones (glitazones) class of antidiabetic drugs. For instance *in vitro* and *in vivo* studies suggest that pioglitazone, a peroxisome proliferator-activated receptor gamma (PPAR-γ) agonist used in the treatment of type 2 diabetes, is metabolized mainly by CYP2C8, and to a lesser extent by CYP3A4, whereas rosiglitazone is metabolized by CYP2C9 and CYP2C8 (Jaakkola *et al.*, 2006; Scheen, 2007). The antimalarial agent, amodiaquin, is almost exclusively metabolized by CYP2C8, and several variant forms of CYP2C8 with low activity toward amodiaquin clearance have been identified and estimated at 1–4% in African populations (Gil and Gil Berglund, 2007). This obviously raises safety concerns for carriers of these defective alleles, since amodiaquin is poorly tolerated, causing hematologic and hepatic toxicity (Brasseur, 2007). Although a functional allele, CYP2C8*3, coding for the Arg139Lys and Lys399Arg amino acid substitutions has been shown to confer higher *in vivo* metabolic capacity than the wild-type CYP2C8*1 allele, these pharmacokinetic differences were quantitatively moderate and did not impact the overall clearance of the drug (Kirchheiner *et al.*, 2006; Pedersen, Damkier and Brosen, 2006). Nevertheless, developers of drugs that are predominantly metabolized by CYP2C8 should take into consideration the frequency of these polymorphisms in the populations tested, especially if the drug has a narrow safety window.

CYP2D6, CYP2C9, and CYP2C19 polymorphisms are the most clinically relevant and the most widely studied. This is due to the fact that these three isozymes contribute to the metabolism of about 40% of drugs currently on the market (Ingelman-Sundberg,

Oscarson and McLellan, 1999). The CYP2C9 enzyme, the most abundant among human CYP2C isoforms, metabolizes many therapeutically important drugs, including most nonsteroidal anti-inflammatory drugs (NSAIDs), *S*-warfarin, phenytoin, and losartan (Scordo *et al.*, 2004; Zand *et al.*, 2007). The impact of genetic variations at the CYP2C9 and CYP2C19 loci on the metabolism and safety of the anticoagulant, warfarin, and antiepileptic, phenytoin is well documented. About 30 *CYP2C9* alleles have been reported with more than 20 additional SNPs identified (Ingelman-Sundberg, Daly and Nebert, 2007). The CYP2C9*2 variant is the most common detrimental allele of CYP2C9 among Caucasians (8–12%). It is a C430T transition in exon 3, resulting in an Arg144Cys amino acid substitution. The frequency of the CYP2C9*3 allele varies between 3% and 8% among Caucasians and is an Ile359Leu amino acid change due to an A1075C transition in exon 7. The frequency of these two alleles is lower in Asians and black Africans. Patients with CYP2C9*2 and/or CYP2C9*3 variants metabolize warfarin slowly and are more likely to have elevated plasma levels of the drug, potentially leading to adverse events, including hemorrhaging during initiation of warfarin therapy, before dose stabilization (Crespi and Miller, 1997; Gage and Lesko, 2008). Therefore, taking the genetically determined metabolic capacities of CYP2C9 into account during clinical development of new drugs that are substrates for CYP2C9 or during therapy has the potential to improve individual risk/benefit relationships (Rettie and Tai, 2006). Warfarin and drugs with similar risk/benefit parameters in particular will benefit from this information because of the narrow therapeutic index for each patient (Krynetskiy and McDonnell, 2007). Indeed the FDA recently updated the warfarin label, to include information on pharmacogenetic testing and to encourage the use of this information in choosing doses for individuals initiating warfarin therapy (Gage and Lesko, 2008).

CYP2C19 is involved in the metabolism of many drugs including *S*-mephenytoin, diazepam, omeprazole, proguanil, citalopram, *R*-warfarin, and many antidepressants (Herrlin *et al.*, 1998; Zand *et al.*, 2007). More than 20 CYP2C19 alleles have been identified, with some resulting in decreases in the activity of the encoded enzyme. The CYP2C19*2 and CYP2C19*3 alleles have been associated with poor metabolism of CYP2C19 drug substrates, accounting for 50% to 90% of the CYP2C19 poor metabolizer phenotype (Nakamoto *et al.*, 2007; Zand *et al.*, 2007). The frequency of the poor metabolizer phenotype ranges between 2% and 5% in Caucasians, and 18% to 23% in Asians (Ohkubo *et al.*, 2006).

CYP3A4 is the most abundant CYP in human liver and is responsible for the metabolism of more than 50% of the drugs currently on the market (Bertilsson *et al.*, 1998; Siest, Jeannesson and Visvikis-Siest, 2007). Over 30 SNPs have been identified in both the coding and promoter region of CYP3A4, with some variant alleles resulting in variations in CYP3A4 activity, including CYP3A4*4, CYP3A4*5, and CYP3A4*18 (Henningsson *et al.*, 2005; Hu *et al.*, 2005; Ruzilawati, Suhaimi and Gan, 2007). It is interesting that although CYP3A4 catalyzes the largest number of pharmaceutical drugs, the pharmacokinetic properties of none of its drug substrates appear to be notably affected by the genetic mutations that have so far been described (Siest, Jeannesson and Visvikis-Siest, 2007). In addition, the extensive interindividual differences in drug response for drugs metabolized by CYP3A4 do not seem to have any genetic basis. Although the reason for this variation may be multifactorial, it is believed that extensive variation in the levels of CYP3A4 protein in different individuals is likely to play a part. It has been further suggested that polymorphisms in other genes, including those involved in the transcriptional regulation of CYP3A4, may contribute to this

variation (Siest *et al.*, 2004; Siest, Jeannesson and Visvikis-Siest, 2007). Furthermore, expression or polymorphism of CYP3A5, the other CYP3A family gene expressed in adults, has also been implicated in CYP3A4 interindividual variability. However, there have been conflicting reports regarding the importance of CYP3A5 and CYP3A7 (fetal CYP3A) expression/polymorphisms in overall CYP3A-mediated enzyme activities (Fukuda *et al.*, 2004; Jin *et al.*, 2005; Wojnowski and Kamdem, 2006).

### 3.2 Phase II DME Polymorphisms

NATs are a phase II gene family, with two human isoforms, NAT-1 and NAT-2. NATs are responsible for the acetylation conjugation of arylamine drugs such as isoniazid, dapsone, and procainamide (Cascorbi, 2006; Yalin *et al.*, 2007). NAT2 is expressed predominantly in the human liver, whereas NAT1 can be detected in a wide variety of different tissues (Cascorbi, 2006). NATs were among the first family of enzymes reported to exhibit pharmacogenetic variations. The role of NAT in the metabolism of the antitubercular drug isoniazid was identified as far back as the 1950s (Bonicke and Reif, 1953; Weber and Hein, 1985; Sim, Westwood and Fullam, 2007) and *N*-acetylation has since been shown to be the cause of interindividual variation in the metabolism and toxicity of isoniazid (Evans, Manley and Mc, 1960; Weber and Hein, 1985). More than 36 SNPs have been identified at the NAT-2 locus (Hein *et al.*, 2007). Some of these alleles (NAT2*5B, *6A, and *7B) and/or combinations of haplotypes are functional, resulting in the slow, intermediate, or rapid acetylator phenotypes (Pistorius *et al.*, 2007). The slow acetylator phenotype occurs at frequencies of up to 95% in some populations (Relling *et al.*, 1992; Ilett *et al.*, 1993; Lin *et al.*, 1993). The serum half-life of isoniazid varies from 0.9 to 1.8 hours in rapid acetylators and from 2.2 to 4.4 hours in slow acetylators (Peloquin *et al.*, 1997). An association has also been reported between SNPs in NAT2 and adverse events caused by sulfasalazine, a first-line treatment for inflammatory bowel disease and a rheumatoid arthritis drug. A study showed that Chinese patients with inflammatory bowel disease who were slow acetylators, experienced more adverse events (36%) following treatment with sulfasalazine compared with patients with the common (wild-type) allele (Chen *et al.*, 2007; Taniguchi *et al.*, 2007). These studies emphasize the importance of NAT2 polymorphisms in the metabolism and/or toxicity of drugs predominantly eliminated via acetylation.

UGTs are a major phase II detoxication system, responsible for the glucuronidation of many pharmacologically active xenobiotic compounds and endogenous substances in humans and have profound effects on the disposition, metabolism, and elimination of many drugs (Burchell *et al.*, 2000, 2005). UGTs are responsible for converting potentially toxic compounds into water-soluble glucuronides that can be easily excreted from the body (Burchell and Coughtrie, 1989). In humans, they exist as a supergene family comprising two families, UGT1 and UGT2, located on chromosomes 2q37 and 4q13, respectively. Human UGT1 family consists of nine members (A1, A3, A4, A5, A6, A7, A8, A9, and A10) with an identical carboxyl terminus but different amino-terminal domains (28). UGT1A genes have five exons and share complete sequence homology from exons 2 to 5, with alternative splicing resulting in sequence differences in exon 1. The UGT2 family is further divided into two subfamilies (UGT2A and UGT2B) with each gene comprising six exons. UGT2A contains two genes (2A1 and 2A2) while UGT2B contains seven genes (2BIX4, 2BIX7, 2BIX10, 2BIX11, 2BIX15, 2BIX17, and 2BIX28) (Miners, McKinnon and Mackenzie, 2002). This remarkably small number of enzymes is responsible for the conjugation of thousands of xenobiotics and plays a

critical role in the detoxication of endogenous toxins such as bilirubin and hormones such as the thyroid hormones (Burchell, 2003; Yamanaka *et al.*, 2007).

Several polymorphisms have been reported at UGT loci, the functional significance of which is still not known for some. Perhaps the best-known clinical implications of UGT polymorphisms are Gilbert's syndrome and Crigler–Najjar syndromes I and II, which are nonhemolytic unconjugated hyperbilirubinemias, resulting from mutations in or lack of UGT1A1 activity (Tukey and Strassburg, 2000; Paoluzzi *et al.*, 2004). UGT1A1*28 is a functional UGT promoter polymorphism associated with Gilbert's disease (Monaghan *et al.*, 1996; Strassburg and Manns, 2000), which has also been implicated in toxicity associated with irinotecan therapy. Since UGTs are responsible for direct conjugation of parent drugs and also metabolites resulting from phase I biotransformation, functional polymorphisms of UGTs can affect the rate at which a parent drug is cleared or the accumulation of a metabolite to toxic levels. The formation of reactive intermediates such as the acyl glucuronide can be impacted by UGT polymorphisms. Indeed several allelic variants of UGT proteins have been reported as essential determinants of susceptibility to drug-related toxicities. Irinotecan is an anticancer drug approved for use in patients with metastatic colorectal cancer. It is metabolized to an active metabolite, SN-38, which is glucuronidated by hepatic UGTs. Diarrhea is the major irinotecan-related toxicity, believed to be due to direct enteric injury caused by SN-38, which has been shown to accumulate in mouse intestines after intraperitoneal administration (Araki *et al.*, 1993). Furthermore, studies in humans have shown an inverse relationship between SN-38 glucuronidation rates and the severity of diarrhea in patients treated with increasing doses of irinotecan (Gupta *et al.*, 1994), indicating that glucuronidation of SN-38 protects against irinotecan-induced gastrointestinal toxicity. Therefore, the conversion of SN-38 to SN-38 glucuronide by hepatic UGTs is a critical step in the detoxication of this active metabolite that is presumably responsible for the toxicity of irinotecan. It has been shown that UGT1A1 is the major UGT isoform responsible for SN-38 glucuronidation (Nagar and Blanchard, 2006), hence it is reasonable to presume that polymorphisms in UGT1A1 that reduce the activity of the encoded UGT1A1 protein would compromise an individual's ability to deactivate SN-38. Indeed the FDA recently revised the package insert for irinotecan to warn of this association between UGT1A1*28 and potential irinotecan toxicity (Ando *et al.*, 2007).

Other variant alleles of UGT1A1, UTG1A1*6, and UGT1A1*27 found in the coding regions are known to reduce the activity of the enzyme and have also been postulated to impact the ability of variant individuals to metabolize drugs that undergo UGT1A-mediated conjugation (Ando *et al.*, 2007). For instance, UGT1A1*6, together with UGT2B7 and CYP2D6 variant alleles, was associated with a low ability to glucuronidate carvedilol, an adrenergic beta-blocker, strongly affecting the pharmacokinetics of this drug in Japanese subjects (Ohno *et al.*, 2004).

Functional polymorphisms also appear common in the other UGT genes, including UGT1A4, UGT1A6, UGT1A7, UGT1A8, UGT1A9, UGT1A10, and UGT2B7 (Burchell *et al.*, 2000; Guillemette *et al.*, 2000; Tukey and Strassburg, 2000; Saeki *et al.*, 2005; Thibaudeau *et al.*, 2006). Although some of these have been found to encode proteins with reduced activities toward their substrates, the clinical implications of these variations have not yet been definitively defined. For instance, while UGT1A7 and UGT1A9 are also involved in the metabolism of irinotecan and exhibit allelic variations, only UGT1A7 alleles have been shown to influence SN-38 glucuronidation kinetics (Gagne *et al.*, 2002; Jinno *et al.*, 2003b; Lankisch *et al.*, 2005). Two SNPs in exon

1 of UGT1A10 (G177A and C605T) resulting in amino acid alterations have been shown to encode proteins with reduced activities toward UGT substrates in recombinant systems, suggesting that these substitutions can influence the metabolism of drugs that are primarily eliminated by glucuronidation (Saeki *et al.*, 2002; Jinno *et al.*, 2003a).

SULTs are also a major phase II detoxication enzyme system. These cytosolic enzymes evolved from one supergene superfamily and catalyze the sulfation of many xenobiotics including drugs, and endogenous compounds including steroids, bile acids, thyroid hormones, and neurotransmitters (Weinshilboum *et al.*, 1997). Although SULTs catalyze phase II reactions, which are generally considered to render compounds less functional and more soluble, sulfation is also involved in biosynthetic pathways such as steroid biosynthesis and also serves as the terminal step in the bioactivation of many compounds to mutagens (Hobkirk and Glasier, 1993). UGTs and SULTs often share similar substrate requirements and it appears the relative contribution of each system to the disposition of a common substrate is dictated by the subcellular location of the isozymes, kinetic properties, and the availability of obligatory cofactors (Burchell and Coughtrie, 1997).

At least three SULT families (SULT1, SULT2, and SULT4) exist in humans with 10 functional genes, localized on five different chromosomes, identified to date (Nagata and Yamazoe, 2000; Coughtrie, 2002; Gamage *et al.*, 2006). Genetic polymorphisms have been detected in all human SULT genes and although some allelic variants have been associated with disease onset, epidemiology, or prognosis (Wang *et al.*, 2002; Wu *et al.*, 2003; Glatt and Meinl, 2004), definitive correlations between SULT allelic variants and interindividual variations in drug response have not been made. Perhaps this is not surprising, considering that not many drugs are metabolized exclusively by SULTs and also the fact that SULT and UGTs share common substrates; hence, any deficiencies in SULT catalytic activity is likely compensated for by UGT conjugation. Nevertheless, it has been suggested that SULT1A1 polymorphisms may be important in women who are on adjuvant tamoxifen therapy because SULT1A1 catalyzes the sulfation of a variety of phenolic and estrogenic compounds, including 4-hydroxytamoxifen, the active metabolite of tamoxifen. Indeed it has been shown that patients who are homozygous for SULT1A1*2 have significantly reduced platelet sulfotransferase activity and platelet enzyme activity correlates strongly with SULT protein expression (Raftogianis *et al.*, 1997; Nowell *et al.*, 2002).

GSTs are a multifunctional phase II supergene family of enzymes that catalyze many reactions between glutathione (GSH) and lipophilic compounds with electrophilic centers, including those that are cytotoxic and/or genotoxic. The results of GST-mediated conjugation are often stable covalently bound adducts which are usually no longer toxic and can be excreted from the body. For instance, PAHs found in cigarette smoke produce electrophilic substrates for GSTs (Ketterer *et al.*, 1992; Raunio *et al.*, 1995). Aflatoxin $B_1$ ($AFBIX_1$) is a potent hepatocarcinogen in many animals and its carcinogenic activity is derived from its activation to $AFB_1$ exo-8,9-epoxide primarily by CYPs. Several studies have indicated that the extent of GSH conjugation of $AFB_1$ is a major factor in determining risk of $AFB_1$ carcinogenicity in different species (Hayes *et al.*, 1991a, b; Guengerich *et al.*, 1998). Hence, amounts of the CYP enzymes involved in epoxide formation, as well as the levels of the detoxification of AFB1 and the 8,9-epoxide by systems such as GSTs and epoxide hydrolase, determine the toxicity of this compound (Hayes *et al.*, 1992). GSTs form an important component of the cellular response to xenobiotics and endogenous chemicals, and also contribute to physiological

processes such as steroid hormone synthesis and tyrosine catabolism (Litwack, Ketterer and Arias, 1971; Johansson and Mannervik, 2001).

The GSTs have been grouped into seven classes based on sequence homology (Alpha, Mu, Pi, Theta, Zeta, Omega, and Zigma), with about 17 genes identified in humans (Mannervik, Board and Hayes, 2005). Despite the multiplicity of this supergene family, polymorphisms have been identified at almost all GST gene loci and appear to play an important role in the etiology of several malignancies and other diseases (Katoh *et al.*, 1996; Yengi *et al.*, 1996; Strange, Lear and Fryer, 1998). Although several lines of evidence suggest that GST polymorphisms may be important in clinical outcomes of cancer therapy, a clear relationship between GST polymorphisms and drug response has not be established, with data from different studies often conflicting (Anderer *et al.*, 2000; Ambrosone *et al.*, 2001; Mossallam, Hamid and Samra, 2006; Lo and Ali-Osman, 2007). In addition, these associations between clinical outcome and GST polymorphisms do not seem to be supported by pharmacological evidence such as changes in pharmacokinetic properties in patients carrying GST mutations (Innocenti and Ratain, 2002). This could be due to the fact that there is some redundancy in the GST superfamily, with multiple isozymes sharing the same substrate requirements. In addition, GSTs appear to play only a minor role in the metabolic transformation of a few drugs (Eaton and Bammler, 1999; Ates *et al.*, 2004; Kuhne *et al.*, 2008), hence the impact of their polymorphisms is likely to be minimal.

### 3.3 Nuclear Receptors and Transporter Polymorphisms

Genetic variations have been identified at gene loci encoding drug transporters and nuclear receptors responsible for the transcriptional regulation of major DMEs. Interest in deciphering nuclear receptor polymorphisms is driven by suggestions that the high degree of variability in human DME expression and/or activity, may result more from polymorphisms in these regulatory proteins than from direct polymorphisms in the individual enzymes (Harper *et al.*, 2002; Bosch *et al.*, 2006; Martinez-Jimenez *et al.*, 2007). As already mentioned, drugs that are metabolized by CYP3A4 exhibit high interindividual variability in their pharmacokinetic parameters, yet none of the polymorphisms identified at the CYP 3A4 locus can account for this variation, suggesting polymorphisms in PXR protein (Garcia-Martin *et al.*, 2002). In fact PXR has been shown to exhibit alternative splicing, with the PXR-2 splice variant lacking a contiguous stretch of 37 amino acids, with an effect on the basal and/or induced expression of PXR-regulated genes (Dotzlaw *et al.*, 1999; Enoru-Eta *et al.*, 2006). In addition, up to 24 SNPs, occurring in different ethnic populations, have been identified (Hustert *et al.*, 2001; Koyano *et al.*, 2002). Hustert *et al.* (2001) reported six missense mutations that result in variant PXR proteins and studied associations between these PXR variants and CYP3A4 transcription in transiently transfected LS174T cells. Although they reported that three variants affected the basal and/or induced expression of the reporter gene constructs containing two copies of the ER6 motif of the proximal promoter region of CYP3A4, they also suggested that further studies are needed to determine whether this correlation between PXR variants and CYP3A4 promoter activity may be used to predict CYP3A4 activity during drug development or treatment. Although several AhR allelic variants with functional implications and strong effects on AhR phenotypes have been documented in laboratory animals (Poland and Glover, 1990), the few polymorphisms of the human AhR do not appear to have strong and obvious consequences on drug metabolism, although a few have been linked to decreases in

CYP1A activity and induction by 2,3,7,8-tetrachlorodibenzo-p-dioxin (TCDD) (Smart and Daly, 2000; Wong, Okey and Harper, 2001; Harper *et al.*, 2002). Several SNPs have been identified at the CAR locus and two variants have been reported functional in reporter gene assays (Ikeda *et al.*, 2003, 2005). Again the clinical implications of these polymorphisms remain to be established. The clinical importance of genetic polymorphisms in drug transporter genes has also been recognized and extensively reviewed (Kim, 2002; Mizuno *et al.*, 2003; Marzolini, Tirona and Kim, 2004).

## 4 APPLICATIONS OF GENETICS AND GENOMICS IN DRUG DISCOVERY AND DEVELOPMENT

### 4.1 Pharmacogenomics Applications

One of the most attractive possibilities of genomics is the ability to use this technology in the early stages of drug discovery, to predict the safety/toxicity profile of new compounds based on gene expression similarities with previous compounds with similar mechanisms of action. The field toxicogenomics evolved from this premise. The potential benefit is that it would reduce or eliminate the need for costly preclinical animal safety studies (Suter, Babiss and Wheeldon, 2004; Luhe *et al.*, 2005). Thus, by developing databases of the gene expression profiles of compounds with known types of toxicities, the toxic liabilities of new compounds could be determined by querying these databases. Companies such as Gene Logic, in partnership with pharmaceutical companies, developed software and databases, which aimed to apply toxicity-based gene expression assessments to lead selection and optimization. Several pharmaceutical companies have now developed proprietary databases that are applied to rank-order compounds based on toxicogenomics data (Waring *et al.*, 2003; Nie *et al.*, 2006).

Pharmacogenomics, specifically toxicogenomics, is being used mechanistically, to determine the underlying molecular basis of toxicological findings. These data can be particularly useful when several mechanisms can result in a similar pathological end points, or when species- and/or gender-specific differences in toxicity are observed. These data are important when extrapolating data from nonclinical species to humans and have the potential to reduce the attrition rate of compounds in the preclinical stages. The ability to identify which genes are changing in response to treatment with a particular compound, can also be used to determine off-target toxicity (Wang and LeCluyse, 2003; Keshava and Caldwell, 2006). Such an approach was used to determine that thyroid hypertrophy and reductions in thyroid hormone levels following treatment with a nonsteroidal progestin receptor agonist, resulted from induction of UGTs which are known to conjugate thyroid hormones (Yengi *et al.*, 2007).

One of the most successful applications of genomics data is to predict the potential for new chemical entities to cause drug–drug interactions with coadministered drugs that may result in loss of efficacy and ultimately toxicity (Ereshefsky and Dugan, 2000; Rodriguez-Novoa *et al.*, 2006). Interest in gene regulation has increased continuously, with many regulatory agencies requesting information on the ability of new compounds to induce DME. In fact, the FDA draft guidance on *in vitro* and *in vivo* drug interaction studies lends credence to the fact that the agency considers induction data integral to their ability to appropriately evaluate new chemical entities for clinical drug–drug interaction liabilities (FDA, 2006). Preclinical gene regulation studies are performed primarily *ex vivo*, in livers from preclinical safety studies. Increasingly, however, studies

conducted *in vitro* using systems such as primary hepatocytes and cell-based gene reporter assays are yielding more human relevant data, as the need for species to species extrapolation is eliminated (Smith, 2000; Luo *et al.*, 2002; Vermeir *et al.*, 2005; Enoru-Eta *et al.*, 2006; FDA, 2006; Yengi *et al.*, 2007). In addition to drug–drug interaction liabilities, gene regulation can also increase toxic liabilities of compounds by altering rates of metabolism and/or elimination of toxic species. For drugs whose toxicity results from their metabolism to more active and toxic metabolites, such as acetaminophen, induction or down-regulation of genes involved in formation of the metabolite or elimination can greatly affect the safety profile of the compound (Douidar and Ahmed, 1987; DiPetrillo *et al.*, 2002). Genome-based studies can determine early in the discovery process whether metabolism genes are regulated, enabling the design of more appropriate safety and/or clinical studies. Also, since most DMEs are involved in normal biochemical processes such as steroid hormone biosynthesis and elimination, cholesterol homeostasis, and bilirubin metabolism, regulation of DMEs can lead to undesirable end points (You, 2004; Rifkind, 2006). Some studies have shown that drugs associated with idiosyncratic drug reactions (IDRs) also up-regulate some protective genes *in vivo*, suggesting that gene expression profiling could potentially be used to identify compounds with a potential to cause IDRs (Uetrecht, 2003).

Traditionally, gene regulation studies were conducted by determining enzyme activities; however, mRNA data are increasingly being generated either alone or in combination with enzyme activity data, because of the lack of isozyme-specific probe substrates (Smith, 2000; Yengi *et al.*, 2007). In addition, there is increasing evidence that compounds that cause mechanism-based inhibition (MBI) of DMEs can also act as inducers of these enzymes (McConn and Zhao, 2004; Zhou *et al.*, 2005). Occurring post-transcriptionally, mRNA data can help delineate potential toxicological processes taking place at the molecular level that may not be observed at the level of activity because of continuous inactivation of the protein.

### 4.2 Pharmacogenetic Applications

As has been discussed, the efficacy and/or safety of a drug predominantly metabolized by a polymorphic gene can be greatly impacted by the genotype of the individual taking the drug (Evans and Relling, 1999; Shi, Bleavins and de la Iglesia, 2001). Information on the genotypes of participants in clinical trials not only provides data that could be used to explain potential individual differences in response, but also ensures that those with an altered phenotype are properly monitored. This is particularly important with drugs that have a narrow therapeutic window outside of which severe toxicity or therapeutic failure may occur (Linder, Prough and Valdes, 1997). For drugs predominantly metabolized by a polymorphic DME, most companies routinely genotype subjects during clinical trials; however, timing of these studies varies. While some companies genotype clinical samples as a matter of routine, some take a diagnostic approach, only genotyping if the data indicate the genetic variation has an impact on the pharmacokinetics and/or safety profile of the compound. For drugs metabolized by DMEs for which the impact of allelic variants is still unknown, other approaches such as targeted or genome-wide SNP analysis is used to identify SNPs and/or haplotypes that may co-segregate with any observed variations (Swan *et al.*, 2005; Haga and Ginsburg, 2006).

The idea that polymorphisms as yet unidentified either in target or off-target genes may play a causative role in unexpected adverse (idiosyncratic) drug reactions is

increasingly gaining traction. It is currently difficult to accurately predict which drugs will be associated with a significant incidence of IDRs. Formation of reactive metabolites and covalent binding of these intermediates with cellular components such as proteins and DNA has been associated with IDRs and is used as a screening tool to predict IDR liabilities of compounds (Williams *et al.*, 2002; Liebler and Guengerich, 2005). However, since not all reactive metabolites cause IDR and some cause reactions in a species-specific manner, these predictions only provide partial solutions (Gardner *et al.*, 2005; Liebler and Guengerich, 2005). It has been suggested that polymorphisms in DMEs, the proteins to which reactive metabolites covalently bind, receptors, and/or protective genes all have the potential to alter the metabolic clearance of a drug, potentially accounting for individual or species differences in susceptibility to IDR. Use of genomics and genetic technologies to determine gene expression profiles and genotypes therefore has the potential to develop into a more effective screening tool for IDR, as genetically susceptible subsets of the population could be identified (Hosford *et al.*, 2004; Hughes *et al.*, 2004).

The "omics" technologies have been incorporated into traditional drug safety and metabolism studies, to provide information on the efficacy and toxicity of compounds in development and to elucidate the underlying molecular mechanisms responsible. The FDA and other regulatory agencies recognize that in order to reduce the time and costs associated with bringing new drugs to market, drug companies have to be innovative in their approaches, and embrace new technologies and processes that challenge traditional paradigms. To this effect, the Critical Path Initiative was launched by the FDA to encourage collaborations among government regulators, the academic community, and regulated businesses to modernize scientific tools and processes, and harness the potential of bioinformatics to, among others, evaluate and predict the safety, effectiveness, and manufacturability of candidate medical products (FDA, 2004), the ultimate objective being a smarter and more rational approach to the development of safer and more effective drugs.

## REFERENCES

Aklillu, E., Persson, I., Bertilsson, L., Johansson, I., Rodrigues, F., Ingelman-Sundberg, M. (1996). Frequent distribution of ultrarapid metabolizers of debrisoquine in an Ethiopian population carrying duplicated and multiduplicated functional cyp2d6 alleles. *J. Pharmacol Exp. Ther.*, *278*, 441–446.

Alexanderson, B., Evans, D.A., Sjoqvist, F. (1969). Steady-state plasma levels of nortriptyline in twins: influence of genetic factors and drug therapy. *Br. Med. J.*, *4*, 764–768.

Ambrosone, C.B., Sweeney, C., Coles, B.F., Thompson, P.A., McClure, G.Y., Korourian, S., Fares, M.Y., Stone, A., Kadlubar, F.F., Hutchins, L.F. (2001). Polymorphisms in glutathione S-transferases (GSTM1 and GSTT1) and survival after treatment for breast cancer. *Cancer Res.*, *61*, 7130–7135.

Anderer, G., Schrappe, M., Brechlin, A.M., Lauten, M., Muti, P., Welte, K., Stanulla, M. (2000). Polymorphisms within glutathione S-transferase genes and initial response to glucocorticoids in childhood acute lymphoblastic leukaemia. *Pharmacogenetics*, *10*, 715–726.

Ando, Y., Fujita, K., Sasaki, Y., Hasegawa, Y. (2007). UGT1AI*6 and UGT1A1*27 for individualized irinotecan chemotherapy. *Curr. Opin. Mol Ther.*, *9*, 258–262.

Araki, E., Ishikawa, M., Iigo, M., Koide, T., Itabashi, M., Hoshi, A. (1993). Relationship between development of diarrhea and the concentration of SN-38, an active metabolite of CPT-11, in

the intestine and the blood plasma of athymic mice following intraperitoneal administration of CPT-11. *Jpn J. Cancer Res.*, *84*, 697–702.

Ates, N.A., Tursen, U., Tamer, L., Kanik, A., Derici, E., Ercan, B., Atik, U. (2004). Glutathione S-transferase polymorphisms in patients with drug eruption. *Arch. Dermatol. Res.*, *295*, 429–433.

Basile, V.S., Ozdemir, V., Masellis, M., Walker, M.L., Meltzer, H.Y., Lieberman, J.A., Potkin, S.G., Alva, G., Kalow, W., Macciardi, F.M., Kennedy, J.L. (2000). A functional polymorphism of the cytochrome P450 1A2 (CYP1A2) gene: association with tardive dyskinesia in schizophrenia. *Mol. Psychiatry*, *5*, 410–417.

Bertilsson, G., Heidrich, J., Svensson, K., Asman, M., Jendeberg, L., Sydow-Backman, M., Ohlsson, R., Postlind, H., Blomquist, P., Berkenstam, A. (1998). Identification of a human nuclear receptor defines a new signaling pathway for CYP3A induction. *Proc. Natl. Acad. Sci. U.S.A.*, *95*, 12208–12213.

Beverage, J.N., Sissung, T.M., Sion, A.M., Danesi, R., Figg, W.D. (2007). CYP2D6 polymorphisms and the impact on tamoxifen therapy. *J. Pharm. Sci.*, *96*, 2224–2231.

Blumberg, B., Sabbagh, W. Jr, Juguilon, H., Bolado, J., Jr, van Meter, C.M., Ong, E.S., Evans, R.M. (1998). SXR, a novel steroid and xenobiotic-sensing nuclear receptor. *Genes Dev.*, *12*, 3195–3205.

Boke, O., Gunes, S., Kara, N., Aker, S., Sahin, A.R., Basar, Y., Bagci, H. (2007). Association of serotonin 2A receptor and lack of association of CYP1A2 gene polymorphism with tardive dyskinesia in a Turkish population. *DNA Cell Biol.*, *26*, 527–531.

Bonicke, R., Reif, W. (1953). Enzymatic inactivation of isonicotinic acid hydrizide in human and animal organism. *Naunyn Schmiedebergs Arch. Exp. Pathol. Pharmakol.*, *220*, 321–323.

Bosch, T.M., Deenen, M., Pruntel, R., Smits, P.H., Schellens, J.H., Beijnen, J.H., Meijerman, I. (2006). Screening for polymorphisms in the PXR gene in a Dutch population. *Eur. J. Clin. Pharmacol.*, *62*, 395–399.

Brasseur, P. (2007). Tolerance of amodiaquine. *Med. Trop. (Mars)*, *67*, 288–290.

Burchell, B. (2003). Genetic variation of human UDP-glucuronosyltransferase: implications in disease and drug glucuronidation. *Am. J. Pharmacogenomics*, *3*, 37–52.

Burchell, B., Coughtrie, M.W. (1989). UDP-glucuronosyltransferases. *Pharmacol. Ther.*, *43*, 261–289.

Burchell, B., Coughtrie, M.W. (1997). Genetic and environmental factors associated with variation of human xenobiotic glucuronidation and sulfation. *Environ. Health Perspect.*, *105* (Suppl. 4), 739–747.

Burchell, B., Lockley, D.J., Staines, A., Uesawa, Y., Coughtrie, M.W. (2005). Substrate specificity of human hepatic udp-glucuronosyltransferases. *Methods Enzymol.*, *400*, 46–57.

Burchell, B., Soars, M., Monaghan, G., Cassidy, A., Smith, D., Ethell, B. (2000). Drug-mediated toxicity caused by genetic deficiency of UDP-glucuronosyltransferases. *Toxicol. Lett.*, *112*–113, 333–340.

Cascorbi, I. (2006). Genetic basis of toxic reactions to drugs and chemicals. *Toxicol. Lett.*, *162*, 16–28.

Chen, M., Xia, B., Chen, B., Guo, Q., Li, J., Ye, M., Hu, Z. (2007). N-acetyltransferase 2 slow acetylator genotype associated with adverse effects of sulphasalazine in the treatment of inflammatory bowel disease. *Can. J. Gastroenterol.*, *21*, 155–158.

Chetty, M., Murray, M. (2007). CYP-mediated clozapine interactions: how predictable are they? *Curr. Drug Metab.*, *8*, 307–313.

Coughtrie, M.W. (2002). Sulfation through the looking glass—recent advances in sulfotransferase research for the curious. *Pharmacogenomics J.*, *2*, 297–308.

Crespi, C.L., Miller, V.P. (1997). The R144C change in the CYP2C9*2 allele alters interaction of the cytochrome P450 with NADPH:cytochrome P450 oxidoreductase. *Pharmacogenetics*, *7*, 203–210.

Daly, A.K., Cholerton, S., Gregory, W., Idle, J.R. (1993). Metabolic polymorphisms. *Pharmacol. Ther.*, *57*, 129–160.

DiPetrillo, K., Wood, S., Kostrubsky, V., Chatfield, K., Bement, J., Wrighton, S., Jeffery, E., Sinclair, P., Sinclair, J. (2002). Effect of caffeine on acetaminophen hepatotoxicity in cultured hepatocytes treated with ethanol and isopentanol. *Toxicol. Appl. Pharmacol.*, *185*, 91–97.

Dotzlaw, H., Leygue, E., Watson, P., Murphy, L.C. (1999). The human orphan receptor PXR messenger RNA is expressed in both normal and neoplastic breast tissue. *Clin. Cancer Res.*, *5*, 2103–2107.

Douidar, S.M., Ahmed, A.E. (1987). A novel mechanism for the enhancement of acetaminophen hepatotoxicity by phenobarbital. *J. Pharmacol. Exp. Ther.*, *240*, 578–583.

Eaton, D.L., Bammler, T.K. (1999). Concise review of the glutathione S-transferases and their significance to toxicology. *Toxicol. Sci.*, *49*, 156–164.

Eichelbaum, M. (1986). Polymorphic oxidation of debrisoquine and sparteine. *Prog. Clin. Biol. Res.*, *214*, 157–167.

Eichelbaum, M., Ingelman-Sundberg, M., Evans, W.E. (2006). Pharmacogenomics and individualized drug therapy. *Annu. Rev. Med.*, *57*, 119–137.

Eichelbaum, M., Spannbrucker, N., Steincke, B., Dengler, H.J. (1979). Defective N-oxidation of sparteine in man: a new pharmacogenetic defect. *Eur. J. Clin. Pharmacol.*, *16*, 183–187.

Enoru-Eta, J., Yengi, L.G., Kao, J., Scatina, J. (2006). A reporter cellular model for evaluating induction of CYP3A4 by new chemical entities. *Drug Dev. Res.*, *67*, 470–475.

Ereshefsky, L., Dugan, D. (2000). Review of the pharmacokinetics, pharmacogenetics, and drug interaction potential of antidepressants: focus on venlafaxine. *Depress. Anxiety*, *12* (Suppl. 1) 30–44.

Evans, D.A., Manley, K.A., McKusick, K.V. (1960). Genetic control of isoniazid metabolism in man. *Br. Med. J.*, *2*, 485–491.

Evans, W.E., Relling, M.V. (1999). Pharmacogenomics: translating functional genomics into rational therapeutics. *Science*, *286*, 487–491.

FDA (2004). *Critical Path Initiative*, http://www.fda.gov/oc/initiatives/criticalpath/ (accessed 28 October 2008).

FDA (2006). Guidance for Industry. Drug Interaction Studies—Study Design, Data Analysis, and Implications for Dosing and Labeling, Food and Drug Administration, Washington, DC, http://www.fda.gov/cder/guidelines/dft.htm (accessed 28 October 2008).

Fukuda, T., Onishi, S., Fukuen, S., Ikenaga, Y., Ohno, M., Hoshino, K., Matsumoto, K., Maihara, A., Momiyama, K., Ito, T., Fujio, Y., Azuma, J. (2004). CYP3A5 genotype did not impact on nifedipine disposition in healthy volunteers. *Pharmacogenomics J.*, *4*, 34–39.

Gage, B.F., Lesko, L.J. (2008). Pharmacogenetics of warfarin: regulatory, scientific, and clinical issues. *J. Thromb. Thrombolysis*, *25*, 45–51. (Epub 1 October 2007).

Gagne, J.F., Montminy, V., Belanger, P., Journault, K., Gaucher, G., Guillemette, C. (2002). Common human UGT1A polymorphisms and the altered metabolism of irinotecan active metabolite 7-ethyl-10-hydroxycamptothecin (SN-38). *Mol. Pharmacol.*, *62*, 608–617.

Gamage, N., Barnett, A., Hempel, N., Duggleby, R.G., Windmill, K.F., Martin, J.L., McManus, M. E. (2006). Human sulfotransferases and their role in chemical metabolism. *Toxicol. Sci.*, *90*, 5–22.

Garcia-Martin, E., Martinez, C., Pizarro, R.M., Garcia-Gamito, F.J., Gullsten, H., Raunio, H., Agundez, J.A. (2002). CYP3A4 variant alleles in white individuals with low CYP3A4 enzyme activity. *Clin. Pharmacol. Ther.*, *71*, 196–204.

Gardner, I., Popovic, M., Zahid, N., Uetrecht, J.P. (2005). A comparison of the covalent binding of clozapine, procainamide, and vesnarinone to human neutrophils in vitro and rat tissues in vitro and in vivo. *Chem. Res. Toxicol.*, *18*, 1384–1394.

Garte, S., Gaspari, L., Alexandrie, A.K., Ambrosone, C., Autrup, H., Autrup, J.L., Baranova, H., Bathum, L., Benhamou, S., Boffetta, P., Bouchardy, C., Breskvar, K., Brockmoller, J., Cascorbi,

I., Clapper, M.L., Coutelle, C., Daly, A., Dell'Omo, M., Dolzan, V., Dresler, C.M., Fryer, A., Haugen, A., Hein, D.W., Hildesheim, A., Hirvonen, A., Hsieh, L.L., Ingelman-Sundberg, M., Kalina, I., Kang, D., Kihara, M., Kiyohara, C., Kremers, P., Lazarus, P., Le Marchand, L., Lechner, M.C., van Lieshout, E.M., London, S., Manni, J.J., Maugard, C.M., Morita, S., Nazar-Stewart, V., Noda, K., Oda, Y., Parl, F.F., Pastorelli, R., Persson, I., Peters, W.H., Rannug, A., Rebbeck, T., Risch, A., Roelandt, L., Romkes, M., Ryberg, D., Salagovic, J., Schoket, B., Seidegard, J., Shields, P.G., Sim, E., Sinnet, D., Strange, R.C., Stucker, I., Sugimura, H., To-Figueras, J., Vineis, P., Yu, M.C., Taioli, E. (2001). Metabolic gene polymorphism frequencies in control populations. *Cancer Epidemiol. Biomarkers Prev.*, *10*, 1239–1248.

Gil, J.P., Gil Berglund, E. (2007). CYP2C8 and antimalaria drug efficacy. *Pharmacogenomics*, *8*, 187–198.

Ginsburg, G.S., Konstance, R.P., Allsbrook, J.S., Schulman, K.A. (2005). Implications of pharmacogenomics for drug development and clinical practice. *Arch. Intern. Med.*, *165*, 2331–2336.

Glatt, H., Meinl, W. (2004). Pharmacogenetics of soluble sulfotransferases (SULTs). *Naunyn Schmiedebergs Arch. Pharmacol.*, *369*, 55–68.

Goldstein, D.B., Tate, S.K., Sisodiya, S.M. (2003). Pharmacogenetics goes genomic. *Nat. Rev. Genet.*, *4*, 937–947.

Goldstein, J.A., Faletto, M.B. (1993). Advances in mechanisms of activation and deactivation of environmental chemicals. *Environ. Health Perspect.*, *100*, 169–176.

Guengerich, F.P. (1991a). Molecular advances for the cytochrome P-450 superfamily. *Trends Pharmacol. Sci.*, *12*, 281–283.

Guengerich, F.P. (1991b). Reactions and significance of cytochrome P-450 enzymes. *J. Biol. Chem.*, *266*, 10019–10022.

Guengerich, F.P., Johnson, W.W., Shimada, T., Ueng, Y.F., Yamazaki, H., Langouet, S. (1998). Activation and detoxication of aflatoxin B1. *Mutat. Res.*, *402*, 121–128.

Guenthner, T.M., Nebert, D.W. (1977). Cytosolic receptor for aryl hydrocarbon hydroxylase induction by polycyclic aromatic compounds. Evidence for structural and regulatory variants among established cell cultured lines. *J. Biol. Chem.*, *252*, 8981–8989.

Guillemette, C., Ritter, J.K., Auyeung, D.J., Kessler, F.K., Housman, D.E. (2000). Structural heterogeneity at the UDP-glucuronosyltransferase 1 locus: functional consequences of three novel missense mutations in the human UGT1A7 gene. *Pharmacogenetics*, *10*, 629–644.

Gupta, E., Lestingi, T.M., Mick, R., Ramirez, J., Vokes, E.E., Ratain, M.J. (1994). Metabolic fate of irinotecan in humans: correlation of glucuronidation with diarrhea. *Cancer Res.*, *54*, 3723–3725.

Haga, S.B., Ginsburg, G.S. (2006). Prescribing BiDil: is it black and white? *J. Am. Coll. Cardiol.*, *48*, 12–14.

Hammer, W., Sjoqvist, F. (1967). Plasma levels of monomethylated tricyclic antidepressants during treatment with imipramine-like compounds. *Life Sci.*, *6*, 1895–1903.

Hammons, G.J., Yan-Sanders, Y., Jin, B., Blann, E., Kadlubar, F.F., Lyn-Cook, B.D. (2001). Specific site methylation in the 5′-flanking region of CYP1A2 interindividual differences in human livers. *Life Sci.*, *69*, 839–845.

Harper, P.A., Wong, J.Y., Lam, M.S., Okey, A.B. (2002). Polymorphisms in the human AH receptor. *Chem. Biol. Interact.*, *141*, 161–187.

Hayes, J.D., Judah, D.J., McLellan, L.I., Kerr, L.A., Peacock, S.D., Neal, G.E. (1991a). Ethoxyquin-induced resistance to aflatoxin B1 in the rat is associated with the expression of a novel alpha-class glutathione S-transferase subunit, Yc2, which possesses high catalytic activity for aflatoxin B1-8,9-epoxide. *Biochem. J.*, *279* (Pt 2), 385–398.

Hayes, J.D., Judah, D.J., McLellan, L.I., Neal, G.E. (1991b). Contribution of the glutathione S-transferases to the mechanisms of resistance to aflatoxin B1. *Pharmacol. Ther.*, *50*, 443–472.

Hayes, J.D., Judah, D.J., Neal, G.E., Nguyen, T. (1992). Molecular cloning and heterologous expression of a cDNA encoding a mouse glutathione S-transferase Yc subunit possessing high catalytic activity for aflatoxin B1-8,9-epoxide. *Biochem. J.*, *285* (Pt 1), 173–180.

Hedgecoe A., Martin, P. (2003). The drugs don't work: expectations and the shaping of pharmacogenetics. *Soc. Stud. Sci.*, *33*, 327–364.

Hein, D.W., Sim, E., Boukouvala, S., Grant, D.M., Minchin, R.F. (2007). *Consensus Arylamine N-Acetyltransferase (NAT) Gene Nomenclature*, http://www.louisville.edu/medschool/pharmacology/NAT.html (accessed 28 October 2008).

Henningsson, A., Marsh, S., Loos, W.J., Karlsson, M.O., Garsa, A., Mross, K., Mielke, S., Vigano, L., Locatelli, A., Verweij, J., Sparreboom, A., McLeod, H.L. (2005). Association of CYP2C8, CYP3A4, CYP3A5, and ABCB1 polymorphisms with the pharmacokinetics of paclitaxel. *Clin. Cancer Res.*, *11*, 8097–8104.

Herrlin, K., Massele, A.Y., Jande, M., Alm, C., Tybring, G., Abdi, Y.A., Wennerholm, A., Johansson, I., Dahl, M.L., Bertilsson, L., Gustafsson, L.L. (1998). Bantu Tanzanians have a decreased capacity to metabolize omeprazole and mephenytoin in relation to their CYP2C19 genotype. *Clin. Pharmacol. Ther.*, *64*, 391–401.

Hobkirk, R., Glasier, M.A. (1993). Generation of estradiol within the pregnant guinea pig uterine compartment with special reference to the myometrium. *J. Steroid Biochem. Mol. Biol.*, *44*, 291–297.

Hosford, D.A., Lai, E.H., Riley, J.H., Xu, C.F., Danoff, T.M., Roses, A.D. (2004). Pharmacogenetics to predict drug-related adverse events. *Toxicol. Pathol.*, *32* (Suppl. 1), 9–12.

Hu, Y.F., He, J., Chen, G.L., Wang, D., Liu, Z.Q., Zhang, C., Duan, L.F., Zhou, H.H. (2005). CYP3A5*3 and CYP3A4*18 single nucleotide polymorphisms in a Chinese population. *Clin. Chim. Acta*, *353*, 187–192.

Hughes, A.R., Mosteller, M., Bansal, A.T., Davies, K., Haneline, S.A., Lai, E.H., Nangle, K., Scott, T., Spreen, W.R., Warren, L.L., Roses, A.D. (2004). Association of genetic variations in HLA-B region with hypersensitivity to abacavir in some, but not all, populations. *Pharmacogenomics*, *5*, 203–211.

Hustert, E., Zibat, A., Presecan-Siedel, E., Eiselt, R., Mueller, R., Fuss, C., Brehm, I., Brinkmann, U., Eichelbaum, M., Wojnowski, L., Burk, O. (2001). Natural protein variants of pregnane X receptor with altered transactivation activity toward CYP3A4. *Drug Metab. Dispos.*, *29*, 1454–1459.

Ikeda, S., Kurose, K., Jinno, H., Sai, K., Ozawa, S., Hasegawa, R., Komamura, K., Kotake, T., Morishita, H., Kamakura, S., Kitakaze, M., Tomoike, H., Tamura, T., Yamamoto, N., Kunitoh, H., Yamada, Y., Ohe, Y., Shimada, Y., Shirao, K., Kubota, K., Minami, H., Ohtsu, A., Yoshida, T., Saijo, N., Saito, Y., Sawada, J. (2005). Functional analysis of four naturally occurring variants of human constitutive androstane receptor. *Mol. Genet. Metab.*, *86*, 314–319.

Ikeda, S., Kurose, K., Ozawa, S., Sai, K., Hasegawa, R., Komamura, K., Ueno, K., Kamakura, S., Kitakaze, M., Tomoike, H., Nakajima, T., Matsumoto, K., Saito, H., Goto, Y., Kimura, H., Katoh, M., Sugai, K., Minami, N., Shirao, K., Tamura, T., Yamamoto, N., Minami, H., Ohtsu, A., Yoshida, T., Saijo, N., Saito, Y., Sawada, J. (2003). Twenty-six novel single nucleotide polymorphisms and their frequencies of the NR1I3 (CAR) gene in a Japanese population. *Drug Metab. Pharmacokinet.*, *18*, 413–418.

Ikenaga, Y., Fukuda, T., Fukuda, K., Nishida, Y., Naohara, M., Maune, H., Azuma, J. (2005). The frequency of candidate alleles for CYP2D6 genotyping in the Japanese population with an additional respect to the -1584C to G substitution. *Drug Metab. Pharmacokinet.*, *20*, 113–116.

Ilett, K.F., Chiswell, G.M., Spargo, R.M., Platt, E., Minchin, R.F. (1993). Acetylation phenotype and genotype in aboriginal leprosy patients from the north-west region of Western Australia. *Pharmacogenetics*, *3*, 264–269.

Ingelman-Sundberg, M. (2001). Pharmacogenetics: an opportunity for a safer and more efficient pharmacotherapy. *J. Intern. Med.*, *250*, 186–200.

Ingelman-Sundberg, M., Daly, A.K., Nebert, D.W. (2007). *Home Page of the Human Cytochrome P450 (CYP) Allele Nomenclature Committee*, http://www.cypalleles.ki.se/ (accessed 28 October 2008).

Ingelman-Sundberg, M., Oscarson, M., McLellan, R.A. (1999). Polymorphic human cytochrome P450 enzymes: an opportunity for individualized drug treatment. *Trends Pharmacol. Sci.*, *20*, 342–349.

Innocenti, F., Ratain, M.J. (2002). Update on pharmacogenetics in cancer chemotherapy. *Eur. J. Cancer*, *38*, 639–644.

Jaakkola, T., Laitila, J., Neuvonen, P.J., Backman, J.T. (2006). Pioglitazone is metabolised by CYP2C8 and CYP3A4 in vitro: potential for interactions with CYP2C8 inhibitors. *Basic Clin. Pharmacol. Toxicol.*, *99*, 44–51.

Jin, M., Gock, S.B., Jannetto, P.J., Jentzen, J.M., Wong, S.H. (2005). Pharmacogenomics as molecular autopsy for forensic toxicology: genotyping cytochrome P450 3A4*1B and 3A5*3 for 25 fentanyl cases. *J. Anal. Toxicol.*, *29*, 590–598.

Jinno, H., Saeki, M., Tanaka-Kagawa, T., Hanioka, N., Saito, Y., Ozawa, S., Ando, M., Shirao, K., Minami, H., Ohtsu, A., Yoshida, T., Saijo, N., Sawada, J. (2003a). Functional characterization of wild-type and variant (T202I and M59I) human UDP-glucuronosyltransferase 1A10. *Drug Metab. Dispos.*, *31*, 528–532.

Jinno, H., Tanaka-Kagawa, T., Hanioka, N., Saeki, M., Ishida, S., Nishimura, T., Ando, M., Saito, Y., Ozawa, S., Sawada, J. (2003b). Glucuronidation of 7-ethyl-10-hydroxycamptothecin (SN-38), an active metabolite of irinotecan (CPT-11), by human UGT1A1 variants, G71R, P229Q, and Y486D. *Drug Metab. Dispos.*, *31*, 108–113.

Johansson, A.S., Mannervik, B. (2001). Human glutathione transferase A3-3, a highly efficient catalyst of double-bond isomerization in the biosynthetic pathway of steroid hormones. *J. Biol. Chem.*, *276*, 33061–33065.

Katoh, T., Nagata, N., Kuroda, Y., Itoh, H., Kawahara, A., Kuroki, N., Ookuma, R., Bell, D.A. (1996). Glutathione S-transferase M1 (GSTM1) and T1 (GSTT1) genetic polymorphism and susceptibility to gastric and colorectal adenocarcinoma. *Carcinogenesis*, *17*, 1855–1859.

Keshava, N., Caldwell, J.C. (2006). Key issues in the role of peroxisome proliferator-activated receptor agonism and cell signaling in trichloroethylene toxicity. *Environ. Health Perspect.*, *114*, 1464–1470.

Ketterer, B., Harris, J.M., Talaska, G., Meyer, D.J., Pemble, S.E., Taylor, J.B., Lang, N.P., Kadlubar, F.F. (1992). The human glutathione S-transferase supergene family, its polymorphism, and its effects on susceptibility to lung cancer. *Environ. Health Perspect.*, *98*, 87–94.

Kim, R.B. (2002). Pharmacogenetics of CYP enzymes and drug transporters: remarkable recent advances. *Adv. Drug Deliv. Rev.*, *54*, 1241–1242.

Kirchheiner, J., Thomas, S., Bauer, S., Tomalik-Scharte, D., Hering, U., Doroshyenko, O., Jetter, A., Stehle, S., Tsahuridu, M., Meineke, I., Brockmoller, J., Fuhr, U. (2006). Pharmacokinetics and pharmacodynamics of rosiglitazone in relation to CYP2C8 genotype. *Clin. Pharmacol. Ther.*, *80*, 657–667.

Kliewer, S.A., Lehmann, J.M., Willson, T.M. (1999). Orphan nuclear receptors: shifting endocrinology into reverse. *Science*, *284*, 757–760.

Kliewer, S.A., Moore, J.T., Wade, L., Staudinger, J.L., Watson, M.A., Jones, S.A., McKee, D.D., Oliver, B.B., Willson, T.M., Zetterstrom, R.H., Perlmann, T., Lehmann, J.M. (1998). An orphan nuclear receptor activated by pregnanes defines a novel steroid signaling pathway. *Cell*, *92*, 73–82.

Koyano, S., Kurose, K., Ozawa, S., Saeki, M., Nakajima, Y., Hasegawa, R., Komamura, K., Ueno, K., Kamakura, S., Nakajima, T., Saito, H., Kimura, H., Goto, Y., Saitoh, O., Katoh, M., Ohnuma, T., Kawai, M., Sugai, K., Ohtsuki, T., Suzuki, C., Minami, N., Saito, Y., Sawada, J. (2002). Eleven novel single nucleotide polymorphisms in the NR1I2 (PXR) gene, four of which induce non-synonymous amino acid alterations. *Drug Metab. Pharmacokinet.*, *17*, 561–565.

Krynetskiy, E., McDonnell, P. (2007). Building individualized medicine: prevention of adverse reactions to warfarin therapy. *J. Pharmacol. Exp. Ther.*, *322*, 427–434.

Kuhne, A., Sezer, O., Heider, U., Meineke, I., Muhlke, S., Niere, W., Overbeck, T., Hohloch, K., Trumper, L., Brockmoller, J., Kaiser, R. (2008). Population pharmacokinetics of melphalan and glutathione S-transferase polymorphisms in relation to side effects. *Clin. Pharmacol. Ther.*, *83*, 749–759. (Epub 3 October 2007).

Lankisch, T.O., Vogel, A., Eilermann, S., Fiebeler, A., Krone, B., Barut, A., Manns, M.P., Strassburg, C.P. (2005). Identification and characterization of a functional TATA box polymorphism of the UDP glucuronosyltransferase 1A7 gene. *Mol. Pharmacol.*, *67*, 1732–1739.

Liebler, D.C., Guengerich, F.P. (2005). Elucidating mechanisms of drug-induced toxicity. *Nat. Rev. Drug Discov.*, *4*, 410–420.

Lin, H.J., Han, C.Y., Lin, B.K., Hardy, S. (1993). Slow acetylator mutations in the human polymorphic N-acetyltransferase gene in 786 Asians, blacks, Hispanics, and whites: application to metabolic epidemiology. *Am. J. Hum. Genet.*, *52*, 827–834.

Linder, M.W., Prough, R.A., Valdes, R. Jr. (1997). Pharmacogenetics: a laboratory tool for optimizing therapeutic efficiency. *Clin. Chem.*, *43*, 254–266.

Litwack, G., Ketterer, B., Arias, I.M. (1971). Ligandin: a hepatic protein which binds steroids, bilirubin, carcinogens and a number of exogenous organic anions. *Nature*, *234*, 466–467.

Lo, H.W., Ali-Osman, F. (2007). Genetic polymorphism and function of glutathione S-transferases in tumor drug resistance. *Curr. Opin. Pharmacol.*, *7*, 367–374.

Luhe, A., Suter, L., Ruepp, S., Singer, T., Weiser, T., Albertini, S. (2005). Toxicogenomics in the pharmaceutical industry: hollow promises or real benefit? *Mutat. Res.*, *575*, 102–115.

Luo, G., Cunningham, M., Kim, S., Burn, T., Lin, J., Sinz, M., Hamilton, G., Rizzo, C., Jolley, S., Gilbert, D., Downey, A., Mudra, D., Graham, R., Carroll, K., Xie, J., Madan, A., Parkinson, A., Christ, D., Selling, B., LeCluyse, E., Gan, L.S. (2002). CYP3A4 induction by drugs: correlation between a pregnane X receptor reporter gene assay and CYP3A4 expression in human hepatocytes. *Drug Metab. Dispos.*, *30*, 795–804.

McConn, D.J., Zhao, Z. (2004). Integrating in vitro kinetic data from compounds exhibiting induction, reversible inhibition and mechanism-based inactivation: in vitro study design. *Curr. Drug Metab.*, *5*, 141–146.

Mahgoub, A., Idle, J.R., Dring, L.G., Lancaster, R., Smith, R.L. (1977). Polymorphic hydroxylation of Debrisoquine in man. *Lancet*, *2*, 584–586.

Mannervik, B., Board, P.G., Hayes, J.D., Listowsky, I., Pearson, W.R. (2005). Nomenclature for mammalian soluble glutathione transferases. *Methods Enzymol.*, *401*, 1–8.

Martinez-Jimenez, C.P., Jover, R., Donato, M.T., Castell, J.V., Gomez-Lechon, M.J. (2007). Transcriptional regulation and expression of CYP3A4 in hepatocytes. *Curr. Drug Metab.*, *8*, 185–194.

Marzolini, C., Tirona, R.G., Kim, R.B. (2004). Pharmacogenomics of the OATP and OAT families. *Pharmacogenomics*, *5*, 273–282.

Meyer, U.A., Amrein, R., Balant, L.P., Bertilsson, L., Eichelbaum, M., Guentert, T.W., Henauer, S., Jackson, P., Laux, G., Mikkelsen, H., Peck, C., Pollock, B.G., Priest, R., Sjoqvist, F., Delini-Stula, A. (1996). Antidepressants and drug-metabolizing enzymes—expert group report. *Acta Psychiatr. Scand.*, *93*, 71–79.

Miners, J.O., McKinnon, R.A., Mackenzie, P.I. (2002). Genetic polymorphisms of UDP-glucuronosyltransferases and their functional significance. *Toxicology*, *181*–182, 453–456.

Mizuno, N., Niwa, T., Yotsumoto, Y., Sugiyama, Y. (2003). Impact of drug transporter studies on drug discovery and development. *Pharmacol. Rev.*, *55*, 425–461.

Monaghan, G., Ryan, M., Seddon, R., Hume, R., Burchell, B. (1996). Genetic variation in bilirubin UPD-glucuronosyltransferase gene promoter and Gilbert's syndrome. *Lancet*, *347*, 578–581.

Mossallam, G.I., Abdel Hamid, T.M., Samra, M.A. (2006). Glutathione S-transferase GSTM1 and GSTT1 polymorphisms in adult acute myeloid leukemia; its impact on toxicity and response to chemotherapy. *J. Egypt Natl. Canc. Inst.*, *18*, 264–273.

Murayama, N., Soyama, A., Saito, Y., Nakajima, Y., Komamura, K., Ueno, K., Kamakura, S., Kitakaze, M., Kimura, H., Goto, Y., Saitoh, O., Katoh, M., Ohnuma, T., Kawai, M., Sugai, K., Ohtsuki, T., Suzuki, C., Minami, N., Ozawa, S., Sawada, J. (2004). Six novel nonsynonymous CYP1A2 gene polymorphisms: catalytic activities of the naturally occurring variant enzymes. *J. Pharmacol. Exp. Ther.*, *308*, 300–306.

Nagar, S., Blanchard, R.L. (2006). Pharmacogenetics of uridine diphosphoglucuronosyltransferase (UGT) 1A family members and its role in patient response to irinotecan. *Drug Metab. Rev.*, *38*, 393–409.

Nagata, K., Yamazoe, Y. (2000). Pharmacogenetics of sulfotransferase. *Annu. Rev. Pharmacol. Toxicol.*, *40*, 159–176.

Nakajima, M., Yokoi, T., Mizutani, M., Kinoshita, M., Funayama, M., Kamataki, T. (1999). Genetic polymorphism in the 5′-flanking region of human CYP1A2 gene: effect on the CYP1A2 inducibility in humans. *J. Biochem. (Tokyo)*, *125*, 803–808.

Nakamoto, K., Kidd, J.R., Jenison, R.D., Klaassen, C.D., Wan, Y.J., Kidd, K.K., Zhong, X.B. (2007). Genotyping and haplotyping of CYP2C19 functional alleles on thin-film biosensor chips. *Pharmacogenet. Genomics*, *17*, 103–114.

Nebert, D.W. (1999). Pharmacogenetics and pharmacogenomics: why is this relevant to the clinical geneticist? *Clin. Genet.*, *56*, 247–258.

Nie, A.Y., McMillian, M., Parker, J.B., Leone, A., Bryant, S., Yieh, L., Bittner, A., Nelson, J., Carmen, A., Wan, J., Lord, P.G. (2006). Predictive toxicogenomics approaches reveal underlying molecular mechanisms of nongenotoxic carcinogenicity. *Mol. Carcinog.*, *45*, 914–933.

Nowell, S., Sweeney, C., Winters, M., Stone, A., Lang, N.P., Hutchins, L.F., Kadlubar, F.F., Ambrosone, C.B. (2002). Association between sulfotransferase 1A1 genotype and survival of breast cancer patients receiving tamoxifen therapy. *J. Natl. Cancer Inst.*, *94*, 1635–1640.

Ohkubo, Y., Ueta, A., Ando, N., Ito, T., Yamaguchi, S., Mizuno, K., Sumi, S., Maeda, T., Yamazaki, D., Kurono, Y., Fujimoto, S., Togari, H. (2006). Novel mutations in the cytochrome P450 2C19 gene: a pitfall of the PCR-RFLP method for identifying a common mutation. *J. Hum. Genet.*, *51*, 118–123.

Ohno, A., Saito, Y., Hanioka, N., Jinno, H., Saeki, M., Ando, M., Ozawa, S., Sawada, J. (2004). Involvement of human hepatic UGT1A1, UGT2B4, and UGT2B7 in the glucuronidation of carvedilol. *Drug Metab. Dispos.*, *32*, 235–239.

Paoluzzi, L., Singh, A.S., Price, D.K., Danesi, R., Mathijssen, R.H., Verweij, J., Figg, W.D., Sparreboom, A. (2004). Influence of genetic variants in UGT1A1 and UGT1A9 on the in vivo glucuronidation of SN-38. *J. Clin. Pharmacol.*, *44*, 854–860.

Pedersen, R.S., Damkier, P., Brosen, K. (2006). The effects of human CYP2C8 genotype and fluvoxamine on the pharmacokinetics of rosiglitazone in healthy subjects. *Br. J. Clin. Pharmacol.*, *62*, 682–689.

Peloquin, C.A., Jaresko, G.S., Yong, C.L., Keung, A.C., Bulpitt, A.E., Jelliffe, R.W. (1997). Population pharmacokinetic modeling of isoniazid, rifampin, and pyrazinamide. *Antimicrob. Agents Chemother.*, *41*, 2670–2679.

Pistorius, S., Goergens, H., Engel, C., Plaschke, J., Krueger, S., Hoehl, R., Saeger, H.D., Schackert, H.K. (2007). N-Acetyltransferase (NAT) 2 acetylator status and age of tumour onset in patients with sporadic and familial, microsatellite stable (MSS) colorectal cancer. *Int. J. Colorectal Dis.*, *22*, 137–143.

Poland, A., Glover, E. (1990). Characterization and strain distribution pattern of the murine Ah receptor specified by the Ahd and Ahb-3 alleles. *Mol. Pharmacol.*, *38*, 306–312.

Poland, A., Glover, E., Kende, A.S. (1976). Stereospecific, high affinity binding of 2,3,7,8-tetrachlorodibenzo-p-dioxin by hepatic cytosol. Evidence that the binding species is receptor for induction of aryl hydrocarbon hydroxylase. *J. Biol. Chem.*, *251*, 4936–4946.

Pucci, L., Geppetti, A., Maggini, V., Lucchesi, D., Maria Rossi, A., Longo, V. (2007). CYP1A2 F21L and F186L polymorphisms in an Italian population sample. *Drug Metab. Pharmacokinet.*, *22*, 220–222.

Raftogianis, R.B., Wood, T.C., Otterness, D.M., Van Loon, J.A., Weinshilboum, R.M. (1997). Phenol sulfotransferase pharmacogenetics in humans: association of common SULT1A1 alleles with TS PST phenotype. *Biochem. Biophys. Res. Commun.*, *239*, 298–304.

Rasmussen, B.B., Brix, T.H., Kyvik, K.O., Brosen, K. (2002). The interindividual differences in the 3-demthylation of caffeine alias CYP1A2 is determined by both genetic and environmental factors. *Pharmacogenetics*, *12*, 473–478.

Raunio, H., Husgafvel-Pursiainen, K., Anttila, S., Hietanen, E., Hirvonen, A., Pelkonen, O. (1995). Diagnosis of polymorphisms in carcinogen-activating and inactivating enzymes and cancer susceptibility—a review. *Gene*, *159*, 113–121.

Relling, M.V., Lin, J.S., Ayers, G.D., Evans, W.E. (1992). Racial and gender differences in N-acetyltransferase, xanthine oxidase, and CYP1A2 activities. *Clin. Pharmacol. Ther.*, *52*, 643–658.

Repa, J.J., Mangelsdorf, D.J. (2000). The role of orphan nuclear receptors in the regulation of cholesterol homeostasis. *Annu. Rev. Cell Dev. Biol.*, *16*, 459–481.

Rettie, A.E., Tai, G. (2006). The pharmocogenomics of warfarin: closing in on personalized medicine. *Mol. Interv.*, *6*, 223–227.

Rifkind, A.B. (2006). CYP1A in TCDD toxicity and in physiology-with particular reference to CYP dependent arachidonic acid metabolism and other endogenous substrates. *Drug Metab. Rev.*, *38*, 291–335.

Rodriguez-Novoa, S., Barreiro, P., Jimenez-Nacher, I., Soriano, V. (2006). Overview of the pharmacogenetics of HIV therapy. *Pharmacogenomics J.*, *6*, 234–245.

Rushmore, T.H., Kong, A.N. (2002). Pharmacogenomics, regulation and signaling pathways of phase I and II drug metabolizing enzymes. *Curr. Drug Metab.*, *3*, 481–490.

Ruzilawati, A.B., Suhaimi, A.W., Gan, S.H. (2007). Genetic polymorphisms of CYP3A4: CYP3A4*18 allele is found in five healthy Malaysian subjects. *Clin. Chim. Acta*, *383*, 158–162.

Sachse, C., Brockmoller, J., Bauer, S., Roots, I. (1999). Functional significance of a C–>A polymorphism in intron 1 of the cytochrome P450 CYP1A2 gene tested with caffeine. *Br. J. Clin. Pharmacol.*, *47*, 445–449.

Saeki, M., Ozawa, S., Saito, Y., Jinno, H., Hamaguchi, T., Nokihara, H., Shimada, Y., Kunitoh, H., Yamamoto, N., Ohe, Y., Yamada, Y., Shirao, K., Muto, M., Mera, K., Goto, K., Ohmatsu, H., Kubota, K., Niho, S., Kakinuma, R., Minami, H., Ohtsu, A., Yoshida, T., Saijo, N., Sawada, J. (2002). Three novel single nucleotide polymorphisms in UGT1A10. *Drug Metab. Pharmacokinet.*, *17*, 488–490.

Saeki, M., Saito, Y., Jinno, H., Sai, K., Kaniwa, N., Ozawa, S., Komamura, K., Kotake, T., Morishita, H., Kamakura, S., Kitakaze, M., Tomoike, H., Shirao, K., Minami, H., Ohtsu, A., Yoshida, T., Saijo, N., Kamatani, N., Sawada, J. (2005). Genetic polymorphisms of UGT1A6 in a Japanese population. *Drug Metab. Pharmacokinet.*, *20*, 85–90.

Saito, T., Egashira, M., Kiyotani, K., Fujieda, M., Yamazaki, H., Kiyohara, C., Kunitoh, H., Kamataki, T. (2003). Novel nonsynonymous polymorphisms of the CYP1A1 gene in Japanese. *Drug Metab. Pharmacokinet.*, *18*, 218–221.

Scheen, A.J. (2007). Pharmacokinetic interactions with thiazolidinediones. *Clin. Pharmacokinet.*, *46*, 1–12.

Scordo, M.G., Caputi, A.P., D'Arrigo, C., Fava, G., Spina, E. (2004). Allele and genotype frequencies of CYP2C9, CYP2C19 and CYP2D6 in an Italian population. *Pharmacol. Res.*, *50*, 195–200.

Searfoss, G.H., Jordan, W.H., Calligaro, D.O., Galbreath, E.J., Schirtzinger, L.M., Berridge, B.R., Gao, H., Higgins, M.A., May, P.C., Ryan, T.P. (2003). Adipsin, a biomarker of gastrointestinal toxicity mediated by a functional gamma-secretase inhibitor. *J. Biol. Chem.*, *278*, 46107–46116.

Shi, M.M., Bleavins, M.R., de la Iglesia, F.A. (2001). Pharmacogenetic application in drug development and clinical trials. *Drug Metab. Dispos.*, *29*, 591–595.

Shimada, T., Yamazaki, H., Mimura, M., Inui, Y., Guengerich, F.P. (1994). Interindividual variations in human liver cytochrome P-450 enzymes involved in the oxidation of drugs, carcinogens and toxic chemicals: studies with liver microsomes of 30 Japanese and 30 Caucasians. *J. Pharmacol. Exp. Ther.*, *270*, 414–423.

Siest, G., Jeannesson, E., Berrahmoune, H., Maumus, S., Marteau, J.B., Mohr, S., Visvikis, S. (2004). Pharmacogenomics and drug response in cardiovascular disorders. *Pharmacogenomics*, *5*, 779–802.

Siest, G., Jeannesson, E., Visvikis-Siest, S. (2007). Enzymes and pharmacogenetics of cardiovascular drugs. *Clin. Chim. Acta*, *381*, 26–31.

Sim, E., Westwood, I., Fullam, E. (2007). Arylamine N-acetyltransferases. *Expert Opin. Drug Metab. Toxicol.*, *3*, 169–184.

Smart, J., Daly, A.K. (2000). Variation in induced CYP1A1 levels: relationship to CYP1A1, Ah receptor and GSTM1 polymorphisms. *Pharmacogenetics*, *10*, 11–24.

Smith, D.A. (2000). Induction and drug development. *Eur. J. Pharm. Sci.*, *11*, 185–189.

Strange, R.C., Lear, J.T., Fryer, A.A. (1998). Glutathione S-transferase polymorphisms: influence on susceptibility to cancer. *Chem. Biol. Interact.*, *111*–112, 351–364.

Strassburg, C.P., Manns, M.P. (2000). Jaundice, genes and promoters. *J. Hepatol.*, *33*, 476–479.

Suter, L., Babiss, L.E., Wheeldon, E.B. (2004). Toxicogenomics in predictive toxicology in drug development. *Chem. Biol.*, *11*, 161–171.

Swan, G.E., Valdes, A.M., Ring, H.Z., Khroyan, T.V., Jack, L.M., Ton, C.C., Curry, S.J., McAfee, T. (2005). Dopamine receptor DRD2 genotype and smoking cessation outcome following treatment with bupropion SR. *Pharmacogenomics J.*, *5*, 21–29.

Taioli, E., Gaspari, L., Benhamou, S., Boffetta, P., Brockmoller, J., Butkiewicz, D., Cascorbi, I., Clapper, M.L., Dolzan, V., Haugen, A., Hirvonen, A., Husgafvel-Pursiainen, K., Kalina, I., Kremers, P., Le Marchand, L., London, S., Rannug, A., Romkes, M., Schoket, B., Seidegard, J., Strange, R.C., Stucker, I., To-Figueras, J., Garte, S. (2003). Polymorphisms in CYP1A1, GSTM1, GSTT1 and lung cancer below the age of 45 years. *Int. J. Epidemiol.*, *32*, 60–63.

Taniguchi, A., Urano, W., Tanaka, E., Furihata, S., Kamitsuji, S., Inoue, E., Yamanaka, M., Yamanaka, H., Kamatani, N. (2007). Validation of the associations between single nucleotide polymorphisms or haplotypes and responses to disease-modifying antirheumatic drugs in patients with rheumatoid arthritis: a proposal for prospective pharmacogenomic study in clinical practice. *Pharmacogenet. Genomics*, *17*, 383–390.

Thibaudeau, J., Lepine, J., Tojcic, J., Duguay, Y., Pelletier, G., Plante, M., Brisson, J., Tetu, B., Jacob, S., Perusse, L., Belanger, A., Guillemette, C. (2006). Characterization of common UGT1A8, UGT1A9, and UGT2B7 variants with different capacities to inactivate mutagenic 4-hydroxylated metabolites of estradiol and estrone. *Cancer Res.*, *66*, 125–133.

Tukey, R.H., Strassburg, C.P. (2000). Human UDP-glucuronosyltransferases: metabolism, expression, and disease. *Annu. Rev. Pharmacol. Toxicol.*, *40*, 581–616.

Uetrecht, J. (2003). Screening for the potential of a drug candidate to cause idiosyncratic drug reactions. *Drug Discov. Today*, *8*, 832–837.

Vermeir, M., Annaert, P., Mamidi, R.N., Roymans, D., Meuldermans, W., Mannens, G. (2005). Cell-based models to study hepatic drug metabolism and enzyme induction in humans. *Expert Opin. Drug Metab. Toxicol.*, *1*, 75–90.

Vineis, P., Veglia, F., Anttila, S., Benhamou, S., Clapper, M.L., Dolzan, V., Ryberg, D., Hirvonen, A., Kremers, P., Le Marchand, L., Pastorelli, R., Rannug, A., Romkes, M., Schoket, B., Strange, R.C., Garte, S., Taioli, E. (2004). CYP1A1, GSTM1 and GSTT1 polymorphisms and lung cancer: a pooled analysis of gene-gene interactions. *Biomarkers*, *9*, 298–305.

Wang, H., LeCluyse, E.L. (2003). Role of orphan nuclear receptors in the regulation of drug-metabolising enzymes. *Clin. Pharmacokinet.*, *42*, 1331–1357.

Wang, Y., Spitz, M.R., Tsou, A.M., Zhang, K., Makan, N., Wu, X. (2002). Sulfotransferase (SULT) 1A1 polymorphism as a predisposition factor for lung cancer: a case-control analysis. *Lung Cancer*, *35*, 137–142.

Waring, J.F., Cavet, G., Jolly, R.A., McDowell, J., Dai, H., Ciurlionis, R., Zhang, C., Stoughton, R., Lum, P., Ferguson, A., Roberts, C.J., Ulrich, R.G. (2003). Development of a DNA microarray for toxicology based on hepatotoxin-regulated sequences. *EHP Toxicogenomics*, *111*, 53–60.

Weber, W.W., Hein, D.W. (1985). N-acetylation pharmacogenetics. *Pharmacol. Rev.*, *37*, 25–79.

Weinshilboum, R.M., Otterness, D.M., Aksoy, I.A., Wood, T.C., Her, C., Raftogianis, R.B. (1997). Sulfation and sulfotransferases 1: Sulfotransferase molecular biology: cDNAs and genes. *FASEB J.*, *11*, 3–14.

Williams, D.P., Kitteringham, N.R., Naisbitt, D.J., Pirmohamed, M., Smith, D.A., Park, B.K. (2002). Are chemically reactive metabolites responsible for adverse reactions to drugs? *Curr. Drug Metab.*, *3*, 351–366.

Wojnowski, L., Kamdem, L.K. (2006). Clinical implications of CYP3A polymorphisms. *Expert Opin. Drug Metab. Toxicol.*, *2*, 171–182.

Wolf, C.R., Smith, G., Smith, R.L. (2000). Science, medicine, and the future: pharmacogenetics. *BMJ*, *320*, 987–990.

Wong, J.M., Okey, A.B., Harper, P.A. (2001). Human aryl hydrocarbon receptor polymorphisms that result in loss of CYP1A1 induction. *Biochem. Biophys. Res. Commun.*, *288*, 990–996.

Wu, M.T., Wang, Y.T., Ho, C.K., Wu, D.C., Lee, Y.C., Hsu, H.K., Kao, E.L., Lee, J.M. (2003). SULT1A1 polymorphism and esophageal cancer in males. *Int. J. Cancer*, *103*, 101–104.

Yalin, S., Hatungil, R., Tamer, L., Ates, N.A., Dogruer, N., Yildirim, H., Karakas, S., Atik, U. (2007). N-acetyltransferase 2 polymorphism in patients with Diabetes Mellitus. *Cell Biochem. Funct.*, *25*, 407–411.

Yamanaka, H., Nakajima, M., Katoh, M., Yokoi, T. (2007). Glucuronidation of thyroxine in human liver, jejunum, and kidney microsomes. *Drug Metab. Dispos.*, *35*, 1642–1648.

Yengi, L., Inskip, A., Gilford, J., Alldersea, J., Bailey, L., Smith, A., Lear, J.T., Heagerty, A.H., Bowers, B., Hand, P., Hayes, J.D., Jones, P.W., Strange, R.C., Fryer, A.A. (1996). Polymorphism at the glutathione S-transferase locus GSTM3: interactions with cytochrome P450 and glutathione S-transferase genotypes as risk factors for multiple cutaneous basal cell carcinoma. *Cancer Res.*, *56*, 1974–1977.

Yengi, L.G., Xiang, Q., Shen, L., Appavu, C., Kao, J., Scatina, J. (2007). Application of pharmacogenomics in drug discovery and development: correlations between transcriptional modulation and preclinical safety observation. *Drug Metab. Lett.*, *1*, 41–48.

You, L. (2004). Steroid hormone biotransformation and xenobiotic induction of hepatic steroid metabolizing enzymes. *Chem. Biol. Interact.*, *147*, 233–246.

Zand, N., Tajik, N., Moghaddam, A.S., Milanian, I. (2007). Genetic polymorphisms of cytochrome P450 enzymes 2C9 and 2C19 in a healthy Iranian population. *Clin. Exp. Pharmacol. Physiol.*, *34*, 102–105.

Zhou, S., Yung Chan, S., Cher Goh, B., Chan, E., Duan, W., Huang, M., McLeod, H.L. (2005). Mechanism-based inhibition of cytochrome P450 3A4 by therapeutic drugs. *Clin. Pharmacokinet.*, *44*, 279–304.

# Introduction to Drug Transporters

LOUIS LEUNG and ARAM OGANESIAN

## 1 INTRODUCTION

The role of transporter proteins in drug and endogenous substance disposition has increasingly gained recognition over the last decade (Mizuno *et al.*, 2003; Beringer and Slaughter, 2005; Anzai, Kanai and Endou, 2006). For example, the role of transporters in biliary excretion was not recognized as late as 1995 and as a result, species differences in biliary excretion were not readily explainable (Lin, 1995). It has now become clear that transporters are responsible both for the uptake and efflux of drugs and other chemicals in various tissues and may be key determinants of the pharmacokinetic characteristics of a drug. Inhibition or induction of different drug transporters in different tissues can lead to clinical drug–drug interactions that may result in adverse effects or lack of efficacy (Bodo *et al.*, 2003; Van Montfoort *et al.*, 2003; Sai, 2005). Inhibition or induction of transporters that are responsible for endogenous substance disposition (e.g. bile acids) can also result in toxicological consequences (Bodo *et al.*, 2003). This review is not intended to be a comprehensive and in-depth review of all transporter-related topics, but rather focus on the role of drug transporters in drug development. There are many excellent reviews on various topics on transporters (Bodo *et al.*, 2003; Mizuno *et al.*, 2003; Van Montfoort *et al.*, 2003; Beringer and Slaughter, 2005; Sai, 2005; Anzai, Kanai and Endou, 2006; Endres *et al.*, 2006; Zhang *et al.*, 2006; Alrefai and Gill, 2007; Lin, 2007; Urquhart, Tirona and Kim, 2007). The purpose of this review is to provide the readers with fundamental knowledge of drug transporters and their roles in drug development relating to drug absorption, distribution, metabolism, excretion and toxicity, clinical drug–drug interactions, and *in vitro*/*in vivo* methodologies.

## 2 TRANSPORTER CLASSIFICATION, LOCALIZATION, AND FUNCTIONS

Table 1 is a summary of transporter families and the individual genes expressed. Transporters have been classified as primary, secondary, or tertiary active transporters.

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

**TABLE 1 Human drug transporter gene families.**

| Gene family | Gene product (gene symbol) |
|---|---|
| Multidrug resistance protein/P-glycoprotein | MDR1/P-gp (*ABCB1*) |
| Bile salt export pump | BSEP (*ABCB11*) |
| Multidrug resistance-associated protein | MRP1 (*ABCC1*) |
| | MRP2 (*ABCC2*) |
| | MRP3 (*ABCC3*) |
| | MRP4 (*ABCC4*) |
| Breast cancer resistance protein | BCRP (*ABCG2*) |
| Sodium taurocholate cotransporting peptide | NTCP (*SLC10A1*) |
| Oligopeptide transporters | PEPT 1 (*SLC15A1*) |
| | PEPT 2 (*SLC15A2*) |
| Organic anion-transporting polypeptides | OATP-A/OATP1A2 (*SLC21A3*) |
| | OATP-B/OATP2B1 (*SLC21A9*) |
| | OATP-C/OATP1B1 (*SLC21A6*) |
| | OATP8/OATP1B3 (*SLC21A8*) |
| | OATP-D/OATP3A1 (*SLC21A11*) |
| | OATP-E/OATP4A1 (*SLC21A12*) |
| | OATP-F/OATP1C1 (*SLC21A14*) |
| Organic cation transporters | OCT1 (*SLC22A1*) |
| | OCT2 (*SLC22A2*) |
| | OCT3 (*SLC22A3*) |
| Novel organic cation transporters | OCTN1 (*SLC22A4*) |
| | OCTN2 (*SLC22A5*) |
| Organic anion transporters | OAT1 (*SLC22A6*) |
| | OAT2 (*SLC22A7*) |
| | OAT3 (*SLC22A8*) |
| | OAT4 (*SLC22A9*) |

Primary active transporters require ATP-dependent hydrolysis as the first step in catalysis. Examples are ATP-binding cassette transporters such as multidrug resistance protein (MDR), multidrug resistance-associated protein (MRP), and breast cancer resistance protein (BCRP). Secondary or tertiary active transporters are driven by an exchange or cotransport of intracellular and/or extracellular ions with the substrate. Examples include organic anion transporter (OAT), organic anion-transporting polypeptide (OATP), sodium taurocholate cotransporting peptide (NTCP), organic cation transporter (OCT), novel organic cation transporter (OCTN), and oligopeptide transporter (PEPT). The major tissue localization and examples of common substrates and inhibitors of the different transporter families are summarized in Table 2 (see also Fig. 1).

## 2.1 ABC Transporters

***2.1.1 MDR1P-gp (ABCB1)*** Of the known transporters, P-glycoprotein (P-gp) is by far the most characterized and understood efflux transporter. P-gp belongs to the ATP-binding cassette transporter superfamily and is encoded by the MDR1 (multidrug resistance) gene (ABCB1) in humans. It is predominantly expressed in tissues including the luminal surface of intestinal epithelia, the renal proximal tubule, the bile canalicular

**TABLE 2 Major human transporters, tissue localization, substrates, and inhibitors**[a]

| Transporter (gene) | Transporter family | Major tissue localization | Examples of common substrates | Examples of inhibitors |
|---|---|---|---|---|
| | | ABC transporters | | |
| MDR1/P-gp (*ABCB1*) | Multidrug resistance protein/P-glycoprotein | Intestinal enterocytes, liver, kidney, blood–brain barrier, placenta, adrenal, testes, tumor cells | Digoxin, paclitaxel, tacrolimus, atorvastatin | Cyclosporine, terfenadine, verapamil, ketoconazole, quinidine, ritonavir, elacridar (LY-335979), valspodar (PSC-833), GF-120918 |
| BSEP (*ABCB11*) | Bile salt export pump | Liver | Bile salts; paclitaxel, vinblastine, pravastatin | Cyclosporine, troglitazone |
| MRP1 (*ABCC1*) | MRP | Liver, intestine, kidney, brain, testis | Substrates overlap among MRP1, MRP2, and MRP3. | Ketoconazole |
| MRP2 (*ABCC2*) | MRP | Intestine, liver, kidney, brain, placenta | Glutathione, glucuronide, and sulfate conjugates. | Cyclosporine, rifampin, omeprazole |
| MRP3 (*ABCC3*) | MRP | Intestine, pancreas, placenta, adrenal cortex, liver, kidney, prostate | E17βG, indinavir, cisplatin, etoposide, methotrexate, vincristine | |
| MRP4 (*ABCC4*) | MRP | Prostate, lung, adrenals, ovary, testis, pancreas, small intestine | Nucleoside analogues and cyclic nucleotides (cGMP and cAMP). | |
| MRP5 (*ABCC5*) | MRP | Skeletal muscle, heart, and brain | Nucleoside analogues and cyclic nucleotides (cGMP and cAMP). | |
| BCRP (*ABCG2*) | Breast cancer resistance protein | Placenta, intestine, liver, breast, brain, tumor cells | Broad substrate specificity, partly overlapping with P-gp and MRPs.<br>Doxorubicin, daunorubicin, topotecan, rosuvastatin, sulfasalazine | Elacridar (GF120918), gefitinib |

**TABLE 2** ***Continued***

| Transporter (gene) | Transporter family | Major tissue localization | Examples of common substrates | Examples of inhibitors |
|---|---|---|---|---|
| | | Solute carrier transporters | | |
| NTCP (*SLC10A1*) | Sodium taurocholate co-transporting peptide | Liver, pancreas | Bile salts; taurocholate, rosuvastatin | |
| PEPT1 (*SLC15A1*)<br>PEPT2 (*SLC15A2*) | Oligopeptide transporters | Intestine, kidney | Dipeptides, tripeptides, and peptidomimetics; ampicillin, amoxicillin, captopril, valacyclovir | |
| OATP-A/OATP1A2 (*SLC21A3*) | OATP | Brain, liver, testis, prostate | Bile salts, ouabain, *N*-methyl quinidine, certain hormones, and/or hormone conjugates | |
| OATP-B/OATP2B1 (*SLC21A9*) | OATP | Liver, intestine, pancreas, lung, ovary, testes, spleen | Estrone-3-sulfate, DHEAS, pravastatin | |
| OATP-C/OATP1B1 (*SLC21A6*) | OATP | Liver | Estrone-3-sulfate, DHEAS, taurocholate, rifampin, rosuvastatin, methotrexate, pravastatin, thyroxine | Cyclosporine, rifampin |
| OATP8/OATP1B3 (*SLC21A8*) | OATP | Liver | Digoxin, pravastatin, methotrexate, rifampin | |
| OATP-D/OATP3A1 (*SLC21A11*) | OATP | Ubiquitous, strong expression in leukocytes, spleen, tumors | E-3-sulfate, benzylpenicillin | |
| OATP-E/OATP4A1 (*SLC21A12*) | OATP | Ubiquitous, skeletal muscle, tumors | Taurocholate, E-17β-G | |

| | | | | |
|---|---|---|---|---|
| OCT1 (*SLC22A1*) | OCT | Liver | Small organic cations; acyclovir, amantadine, desipramine, ganciclovir, quinidine, quinine, metformin | Disopyramide, midazolam, phenformin, phenoxy-benzamine, quinidine, quinine, Ritonavir, verapamil |
| OCT2 (*SLC22A2*) | OCT | Kidney, brain | Small organic cations; amantadine, cimetidine, memantine | Desipramine, phenoxy-benzamine, quinine |
| OCT3 (*SLC22A3*) | OCT | Skeletal muscle, liver, brain placenta, kidney, heart | Small organic cations; cimetidine | Desipramine, prazosin, phenoxy-benzamine |
| OAT1 (*SLC22A6*) | OAT | Kidney, brain | Organic anions; acyclovir, adefovir, methotrexate, zidovudine | Probenecid, cefadroxil, cefamandole, cefazolin |
| OAT2 (*SLC22A7* | OAT | Liver, kidney | Organic anions; zidovudine, salicylates | |
| OAT3 (*SLC22A8*) | OAT | Kidney, brain | Organic anions; methotrexate, zidovudine, cimetidine, E-3-sulfate | Probenecid, cefadroxil, cefamandole, cefazolin |
| OCTN1 (*SLC22A4*) | Novel organic cation transporters | Kidney, skeletal muscle, prostate, placenta, heart | Small organic cations; quinidine, verapamil | |
| OCTN2 (*SLC22A5*) | Novel organic cation transporters | Kidney, skeletal muscle, prostate, lung, heart, liver | Small organic cations; quinidine, verapamil | |

[a]Adapted and expanded from Xia, Milton and Gan (2007) and Zhang *et al.* (2006).

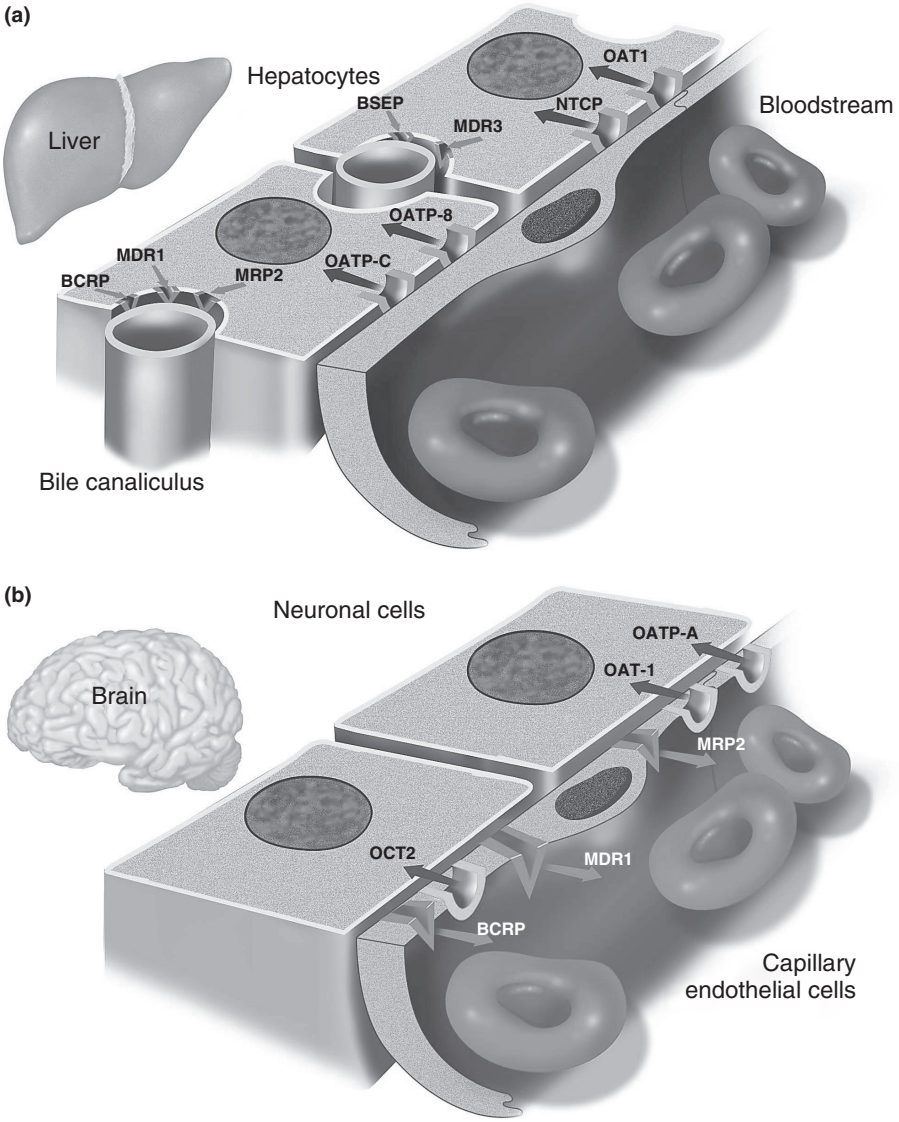


**Figure 1** Human drug transporters expressed in different tissues: **(a)** liver, **(b)** brain, **(c)** kidney, and **(d)** intestine (from Ho and Kim, 2005).

membrane of hepatocytes, and the blood–brain barrier (BBB). P-gp can play an important role in limiting intestinal drug absorption and brain penetration, and in facilitating renal or biliary excretion of drug substrates. P-gp substrates tend to be organic amphipathic molecules with large structural diversities. Drugs that are P-gp substrates represent diverse therapeutic categories; more comprehensive examples of drugs that are P-gp substrates are summarized in Table 3.

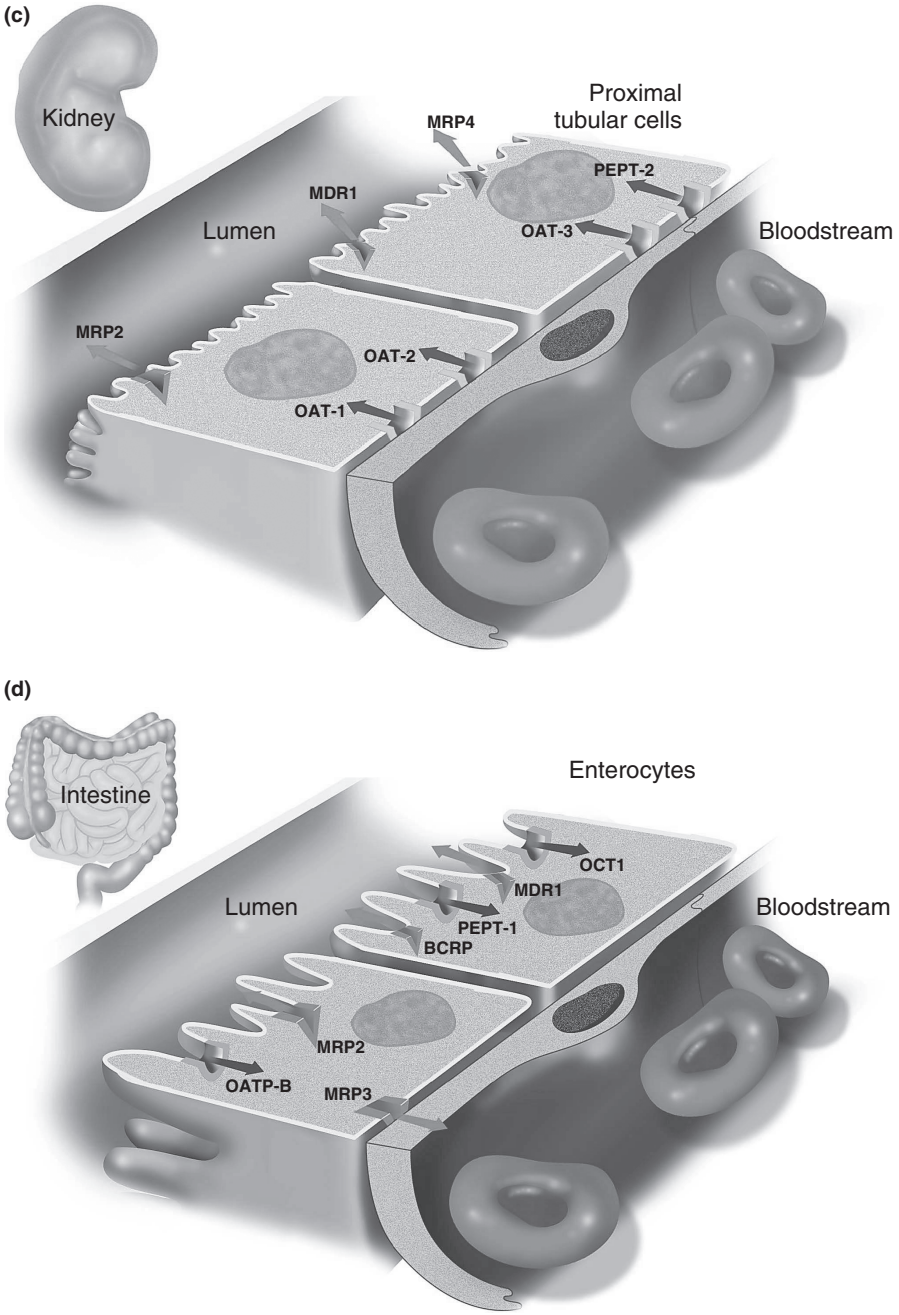


**Figure 1** *Continued*

**TABLE 3 Examples of drugs that are P-gp substrates.**[a]

| | | | |
|---|---|---|---|
| Analgesics | Antiepileptic | Antihypertensives | Oncology |
| Asimadoline | Phenytoin | Losartan | Actinomycin D |
| [d-penicillamine-2,5]-enkephalin (DPDPE) | Antifungals | β-Blockers | Bisantrene |
| Fentanyl | Itraconazole | Celiprolol | Cisplatin |
| Methadone | Ketoconazole | Talinolol | Cytarabine |
| Morphine | Antihistamines | Other cardiovascular | Chlorambucil |
| Antibiotics | Cetirizine | Atorvastatin | Docetaxel |
| Dactinomycin | Fexofenadine | Digoxin | Doxorubicin |
| Erythromycin | Anti-inflamatory | Diltiazem | Daunorubicin |
| Gramicidin D | Methylprednisolone | Quinidine | Fluorouracil |
| Sparfloxacin | Methotrexate | Verapamil | Epirubicin |
| Valinomycin | Corticosteroids | Immunosuppressants | Etoposide |
| Antigout agents | Corticosterone | Cyclosporine A (CsA) | Hydroxyurea |
| Colchicine | Dexamethasone | FK506 | Lapatinib |
| Antidiarrheal | Hydrocortisone | Sirolimus | Mitoxantrone |
| Loperamide | Triamcinolone | Tacrolimus | Octreotide |
| Antiemetics | Diagnostic dyes | HIV protease inhibitors | Paclitaxel |
| Domperidone | Hoechst 33342 | Amprenavir | Tamoxifen |
| Ondansetron | Rhodamine 123 | Indinavir | Teniposide |
| Antidepressant | H2-receptor antagonists | Lopinavir | Topotecan |
| Amitriptyline | Cimetidine | Nelfinavir | Vinblastine |
| Nortriptyline | Ranitidine | Ritonavir | Vincristine |
| | | Saquinavir | Vinorelbine |

[a]Adapted and expanded from Matheny *et al.* (2001) and Endres *et al.* (2006).

***2.1.2 BSEP (ABCB11)*** The bile salt export pump (BSEP) is located on the canalicular membrane of hepatocytes, and mediates the transport of bile acids such as taurocholic acid into the bile. Mutations in BSEP have been associated with cholestatic liver disease (Sai, 2005). BSEP has recently been shown to transport non-bile acid substrates such as pravastatin (Hirano *et al.*, 2005) and vinblastine (Akashi, Tanaka and Takikawa, 2006).

***2.1.3 MRPs (ABCC)*** The multidrug resistance associated proteins (MRPs) have been shown to be associated with drug resistance in cancer cells. The exact physiological role of MRPs is not known. MRPs are involved in the efflux of substrates or drugs from the liver or kidney into the peripheral blood (e.g. MRP1, MRP3, and MRP6), or from the liver, kidney, and small intestines into the bile, urine, or intestinal lumen (MRP2).

MRP1 (ABCC1) is ubiquitous and is present at the basolateral surface of various tissues including the liver, intestine, brain, kidney, and testis. In the liver and kidney, MRP1 is responsible for the secretion of substrates into peripheral blood. MRP1 substrates are glutathione, glucuronide and sulfate conjugates, and organic anions such as methotrexate and estradiol-17β-D-glucuronide.

MRP2 (ABCC2) is by far the most well-studied and important MRP in human drug disposition and toxicity. MRP2 is present at the apical surface of many tissues and include the liver, kidney, brain, and intestines, and is responsible for biliary and renal

excretion and efflux into intestinal lumen of substrates. Substrates for MRP2 overlap with those of MRP1, and include glutathione, glucuronide and sulfate conjugates, and organic anions such as indinavir and cisplatin. A mutation of human MRP2 was shown to lead to hyperbilirubinemia such as the Dubin–Johnson syndrome due to a defect in the biliary excretion of bilirubin glucuronides (Paulusma *et al.*, 1997).

MRP3 (ABCC3) is present at the basolateral surface of various tissues including the liver, kidney, and intestines. Like MRP1, MRP3 is responsible for the secretion of substrates into peripheral blood in the kidneys and liver. Substrates for MRP3 overlap with those of MRP1 and MRP2, and include glutathione, glucuronide and sulfate conjugates, and organic anions such as vincristine and etoposide.

MRP4 (ABCC4) and MRP5 (ABCC5) are present at the apical and basolateral surface, respectively, of many tissues including lungs, intestines, and testes. Substrates for these MRPs are primarily nucleoside analogues and cyclic nucleotides (cGMP and cAMP).

MRP6 (ABCC6) is present at the basolateral surface of the liver and kidneys, and like MRP1 and MRP3, is responsible for the secretion of substrates into peripheral blood in the kidneys and liver. Substrates for MRP6 include cisplatin and daunorubicin.

***2.1.4 BCRP (ABCG2)*** Breast cancer resistance protein (BCRP) has only one ATP-binding cassette and six putative transmembrane domains (Rocchi *et al.*, 2000), suggesting that BCRP is a half-transporter, which may function as a homo- or heterodimer. BCRP is located on the apical surface of the liver, intestines, and breasts, and exhibits broad substrate specificity partly overlapping with P-gp and MRP substrates, such as topotecan and daunorubicin.

## 2.2 Solute Carrier Transporters

***2.2.1 PEPT1 (SLC15A1) and PEPT2 (SLC15A2)*** PEPT1 and PEPT2 are peptide transporters found in the brush border membrane of the intestines and renal cortex, respectively, and are involved in absorption of dipeptides, tripeptides, and peptidomimetic drugs.

***2.2.2 Organic Anion Transporting Polypeptides*** Organic anion-transporting polypeptides (OATPs) are a superfamily of uptake transporters with wide tissue distribution including the liver, gut, and BBB. They are responsible for the uptake from blood or lumen of a variety of amphipathic organic anions such as steroid conjugates, bile acids, and drugs.

*2.2.2.1 OATP1A2 (SLC21A3/SLCO1A2)* OATP1A2 (OATP-A), the first OATP to be cloned (from the liver), is expressed in the BBB and was detected in the brain, lungs, liver, biliary epithelial cells, kidneys, and testes (Van Montfoort *et al.*, 2003). OATP1A2 exhibits a broad substrate specificity including hydrophobic organic anions such as bile acid and bilirubin, neutral compounds such as ouabain, and organic cations such as *N*-methyl quinidine.

*2.2.2.2 OATP1B1 (SLC21A6/SLCO1B1)* OATP1B1 (OATP-C) was the second OATP to be cloned, and is only expressed in the liver responsible for the uptake of

substrates across the basolateral membrane of hepatocytes. OATP1B1 seems to prefer organic anions including bilirubin, steroid hormone conjugates, and the drugs methotrexate and pravastatin.

*2.2.2.3 OATP1B3 (SLC21A8/SLCO1B3)* OATP1B3 (OATP-8) is primarily expressed in the liver, and similar to OATP1B1, transports steroid hormones and organic anions such as pravastatin. In addition to ouabain, OATP1B3 also mediates the hepatic uptake of the uncharged cardiac glycoside digoxin. OATP1B3, together with OATPs 1B1 and 2B1, is primarily responsible for the uptake of sodium-independent bile salt, organic anions, and drugs into the liver.

*2.2.2.4 OATP2B1 (SLC21A9/SLCO2B1)* OATP2B1 (OATP-B) is primarily expressed at the basolateral membrane of hepatocytes, and was also detected in the kidney, small intestine, brain, spleen, placenta, lung, heart, and ovary. OATP2B1 has narrow substrate selectivity when compared with members of the OATP1 family, and transports organic anions including estrone-3-sulfate but not bile salts.

*2.2.2.5 OATP3A1 (SLC21A11/SLCO3A1), OATP4A1 (SLC21A12/SLCO4A1), and OATP1C1 (SLC21A14/SLCO1C1)* OATP3A1 (OATP-D) and OATP4A1 (OATP-E) were cloned from human kidney (Tamai *et al.*, 2000) and are ubiquitous in their tissue distribution. OATP3A1 transports estrone-3-sulfate and benzylpenicillin and OATP4A1 transports taurocholate and estradiol-17β-glucuronide. OATP1C1 (OATP-F) was cloned from human brain (Pizzagalli *et al.*, 2002) and its substrate specificity appears to be narrow.

***2.2.3 Organic Cation Transporters*** The OCT family of transporters is responsible for the uptake of organic cations into the kidney or liver from blood, and represents the first step in renal or biliary excretions (Koepsell, Lips and Volk, 2007).

*2.2.3.1 OCT1 (SLC22A1)* OCT1 is the first member of the OCT family and was cloned from the human liver (Zhang *et al.*, 1997). OCT1 is primarily found on the sinusoidal membrane of hepatocytes and mediates the uptake of quaternary ammonium compounds such as tetraethylammonium (TEA) and *N*-methylquinine, and drugs including quinidine, quinine, and metformin.

*2.2.3.2 OCT2 (SLC22A2)* OCT2 was cloned from human kidney and is primarily expressed on the basolateral membrane of renal proximal tubules. OCT2 is involved in basolateral uptake of organic cations into proximal tubule cells, the first step in the secretion in proximal tubules of cations and drugs such as the weak base cimetidine. OCT2 is also present in neurons in the brain, and may mediate the uptake of cationic drugs.

*2.2.3.3 OCT3 (SLC22A3)* OCT3 is found in various tissues including the liver, kidney, and brain, and shares substrate selectivity as OCT1 and OCT2. Like OCT2, OCT3 is involved in basolateral uptake of organic cations into proximal tubule cells leading to subsequent secretion into proximal tubules.

### *2.2.4 Novel Organic Cation Transporters*

*2.2.4.1 OCTN1 (SLC22A4)* OCTN1 is a polyspecific and sodium-independent cation exchanger found in various tissues including kidney, skeletal muscle, placenta, and heart but not in the liver. OCTN1 transports cations and cationic drugs such as TEA, quinidine, verapamil, and the zwitterion carnitine. Localized in the brush border membrane of the renal proximal tubules, OCTN1 mediates the secretion of various organic cations across the luminal membrane.

*2.2.4.2 OCTN2 (SLC22A5)* OCTN2, also known as carnitine transporter, was also found in various tissues including kidney and liver. Like OCTN1, OCTN2 is involved in the secretion of cations and cationic drugs across the luminal membrane in renal proximal tubule cells. OCTN2 mediates the uptake of carnitine across the luminal membrane, and the cotransport was enhanced in the presence of sodium.

***2.2.5 Organic Anion Transporters*** The OAT family of proteins is involved in the uptake (OAT1, OAT2, or OAT3) of organic anions into the kidney or liver from the blood (OAT1, OAT2, or OAT3), or efflux into the lumen in the kidney (OAT4). OATs contribute collectively to renal or biliary drug elimination.

*2.2.5.1 OAT1 (SLC22A6)* OAT1 was cloned from human kidney (Race *et al.*, 1999) and is primarily expressed at the basolateral membrane of renal proximal tubule cells and was also found in the brain and placenta. OAT1 is involved in the uptake of organic anions across the basolateral membrane of the kidney proximal tubule cells, such as nonsteroidal anti-inflammatory drugs.

*2.2.5.2 OAT2 (SLC22A7)* OAT2 is expressed at the basolateral membrane of the renal proximal tubule cells and also at the sinusoidal membrane of hepatocytes, and is involved in the transport of organic anions such as salicylates and zidovudine.

*2.2.5.3 OAT3 (SLC22A8)* Like OAT2, OAT3 is expressed at the basolateral membrane of the renal proximal tubule cells, and in addition is found in the brain and skeletal muscle. OAT3 mediates uptake of nonsteroidal inflammatory drugs, cimetidine, antiviral drugs, and estrone-3-sulfate.

*2.2.5.4 OAT4 (SLC22A9)* OAT4 is expressed at the brush border membrane of the renal proximal tubule cells, and mediates the secretion across the luminal membrane of the renal proximal tubule cells of organic anions such as estrone-3-sulfate, methotrexate, and zidovudine.

## 3 CLINICAL DRUG–DRUG INTERACTIONS

Since transporters are present on the apical or basolateral membrane of various tissues, drug–drug interactions via inhibition or induction of a particular transporter can occur at the respective sites leading to altered pharmacokinetics or metabolic disposition of the concomitant drugs. Examples of clinical drug–drug interactions involving drug transporters are summarized in Tables 4 and 5. The expected effects of a putative

**TABLE 4 Transporter-mediated drug–drug interactions reported in literature.**[a]

| Transporter | No. of studies reported | No. of showing no effect | No. with effects eliciting >20% change (PK) |
|---|---|---|---|
| P-gp | 251 | 56 | 189 |
| OATP | 24 | 5 | 19 |
| MRP2 | 13 | 3 | 10 |
| MRP1 | 3 | 1 | 2 |
| OAT | 8 | 3 | 5 |
| OAT1 | 2 | 1 | 1 |
| OATP-C (OATP2; OATP1BIX1) | 13 | 1 | 6 |
| OATP-A (OATP1A2) | 3 | 0 | 3 |
| OATP-B (OATP2BIX1) | 1 | 0 | 1 |
| OCT | 5 | 0 | 5 |
| OCT2 | 4 | 1 | 3 |
| BCRP | 1 | 0 | 1 |
| Transporter unidentified | 2 | 0 | 2 |

[a]Survey of University of Washington Drug-Drug Interactions (DDI) database conducted in September 2007.

**TABLE 5 Drug–drug interactions involving P-gp reported in University of Washington DDI database.**

| Inhibitor (precipitant) | Object ("victim") drug |
|---|---|
| (R)-verapamil | Doxorubicin (Wilson *et al.*, 1995), talinolol (Schwarz *et al.*, 1999) |
| Atazanavir | Ritonavir (Boffito *et al.*, 2004, 2006; Von Hentig *et al.*, 2007), saquinavir (Boffito *et al.*, 2004, 2006; Von Hentig *et al.*, 2007) |
| Azithromycin | Fexofenadine (Gupta *et al.*, 2001), ximelagatran (Dorani *et al.*, 2007) |
| Biricodar (VX-710) | Paclitaxel (Rowinsky *et al.*, 1998; Seiden *et al.*, 2002) |
| Carvedilol | Digoxin (Baris *et al.*, 2006) |
| Cefuroxime | Ximelagatran (Dorani *et al.*, 2007) |
| Cetirizine | Pilsicainide (Tsuruoka *et al.*, 2006) |
| Clarithromycin | Colchicine (Rollot *et al.*, 2004), digoxin (Wakasugi *et al.*, 1998; Kurata *et al.*, 2002; Zapater *et al.*, 2002; Rengelshausen *et al.*, 2003; Tanaka *et al.*, 2003) glyburide (glibenclamide) (Lilja *et al.*, 2007) |
| Clotrimazole | Tacrolimus (Vasquez *et al.*, 2005) |
| Cremophor EL | Saquinavir (Martin-Facklam *et al.*, 2002) |
| Cremophor RH40 | Digoxin (Tayrouz *et al.*, 2003) |
| Curcumin | Talinolol (Juan *et al.*, 2007) |
| Cyclosporine | Atorvastatin (Asberg *et al.*, 2001; Lemahieu *et al.*, 2005) cerivastatin (Mück *et al.*, 1999), docetaxel (Malingré *et al.*, 2001a), doxorubicin (Bartlett *et al.*, 1994; Rushing *et al.*, 1994) etoposide (Lum *et al.*, 2000; Lum *et al.*, 1992) idarubicin (Pea *et al.*, 1999; Smeets *et al.*, 2001) irinotecan (Innocenti *et al.*, 2004), paclitaxel (Meerum Terwogt *et al.*, 1999; Britten *et al.*, 2000; Meerum Terwogt *et al.*, 1998) prednisolone (Potter *et al.*, 2004), saquinavir (Brinkman, Huysmans and Burger, 1998), sirolimus (Zimmerman *et al.*, 2003), sitagliptin (Krishna *et al.*, 2007), vincristine (Bertrand *et al.*, 1992) |
| Diltiazem | Methylprednisolone (Booker *et al.*, 2002) |
| Docetaxel | Cyclosporine (Malingré *et al.*, 2001b), vinorelbine (Airoldi *et al.*, 2003) |
| Elacridar (GF120918) | Doxorubicin (Planting *et al.*, 2005), paclitaxel (Malingré *et al.*, 2001c) |

**TABLE 5** ***Continued***

| Inhibitor (precipitant) | Object ("victim") drug |
|---|---|
| Erythromycin | Atorvastatin (Siedlik *et al.*, 1999), fexofenadine (Petri *et al.*, 2006), simvastatin (Kantola, Kivistö and Neuvonen, 1998), talinolol (Schwarz *et al.*, 2000), ximelagatran (Eriksson *et al.*, 2006), digoxin (Eberl *et al.*, 2007) |
| Gemfibrozil | Loperamide (Niemi *et al.*, 2006) |
| Grapefruit juice | Dextromethorphan (Di Marco *et al.*, 2002), erythromycin (Kanazawa, Ohkubo and Sugawara, 2001), fexofenadine (Glaeser *et al.*, 2007), talinolol (Schwarz *et al.*, 2005) |
| Itraconazole | Celiprolol (Lilja *et al.*, 2003), cimetidine (Karyekar *et al.*, 2004), digoxin (Partanen, Jalava and Neuvonen, 1996; Jalava, Partanen and Neuvonen, 1997) fexofenadine (Shon *et al.*, 2005; Shimizu *et al.*, 2006; Shimizu *et al.*, 2006; Uno *et al.*, 2006) loperamide (Niemi *et al.*, 2006), paroxetine (Yasui-Furukori *et al.*, 2007), quinidine (Kaukonen, Olkkola and Neuvonen, 1997) |
| Ketoconazole | Cyclosporine (Akhlaghi *et al.*, 2001), irinotecan (Kehrer *et al.*, 2002), ritonavir (Khaliq *et al.*, 2000), saquinavir (Khaliq *et al.*, 2000), tacrolimus (Floren *et al.*, 1997; Tuteja *et al.*, 2001) |
| Lansoprazole | Erythromycin (Frassetto *et al.*, 2007) |
| Lopinavir and ritonavir | Fexofenadine (Van Heeswijk *et al.*, 2006) |
| Ms209 | Docetaxel (Diéras *et al.*, 2005) |
| Nelfinavir | Azithromycin (Amsden *et al.*, 2000), cyclosporine (Frassetto *et al.*, 2005) |
| Oc144-093 (Ont-093) | Docetaxel (Kuppens *et al.*, 2005), paclitaxel (Chi *et al.*, 2005) |
| Omeprazole | Saquinavir (Winston *et al.*, 2006) |
| Orange juice | Celiprolol (Lilja, Juntti-Patinen and Neuvonen, 2004), dextromethorphan (Di Marco *et al.*, 2002) |
| Paclitaxel | Doxorubicin (Moreira *et al.*, 2001), epirubicin (Fogli *et al.*, 2002) |
| Pilsicainide | Cetirizine (Tsuruoka *et al.*, 2006) |
| Posaconazole | Cyclosporine (Sansone-Parsons *et al.*, 2007), tacrolimus (Sansone-Parsons *et al.*, 2007) |
| Probenecid | Fexofenadine (Yasui-Furukori *et al.*, 2005) |
| Quinidine | Fentanyl (Sadeque *et al.*, 2000; Kharasch *et al.*, 2004) loperamide (Skarke *et al.*, 2003), morphine (Kharasch *et al.*, 2003) |
| Rifampin | Ezetimibe (Oswald *et al.*, 2006) |
| Ritonavir | Buprenorphine (McCance-Katz *et al.*, 2006), digoxin (Penzak *et al.*, 2004; Ding *et al.*, 2004) fexofenadine (Van Heeswijk *et al.*, 2006), Indinavir (Haas *et al.*, 2003), loperamide (Tayrouz *et al.*, 2001), prednisone (Penzak *et al.*, 2005), saquinavir (Buss *et al.*, 2001) |
| Saquinavir | Cyclosporine (Brinkman, Huysmans and Burger, 1998) |
| St. John's wort extract (*Hypericum perforatum*) | Fexofenadine (Wang *et al.*, 2002) |
| Surfactant TPGS | Talinolol (Bogman *et al.*, 2005) |
| Talinolol | Digoxin (Westphal *et al.*, 2000) |
| Tariquidar (XR9576) | Erythromycin (Kurnik, Wood and Wilkinson, 2006) |
| Telithromycin | Digoxin (Nenciu, Laberge and Thirion, 2006) |
| Tenofovir | Saquinavir (Chittick *et al.*, 2006) |
| Valspodar (PSC 833) | Dexamethasone (Kovarik *et al.*, 1998), digoxin (Kovarik *et al.*, 1999), doxorubicin (Sonneveld *et al.*, 1996; Giaccone *et al.*, 1997; Advani *et al.*, 2001; Minami *et al.*, 2001) etoposide (Boote *et al.*, 1996), paclitaxel (Patnaik *et al.*, 2000; Chico *et al.*, 2001) |

**TABLE 5** ***Continued***

| Inhibitor (precipitant) | Object ("victim") drug |
|---|---|
| Verapamil | Doxorubicin (Kerr *et al.*, 1986), fexofenadine (Tannergren *et al.*, 2003; Yasui-Furukori *et al.*, 2005) ranolazine (Lemma *et al.*, 2006), risperidone (Nakagami *et al.*, 2005), simvastatin (Kantola, Kivistö and Neuvonen, 1998) |
| Zosuquidar (LY335979) | Doxorubicin (Sandler *et al.*, 2004), vinorelbine (Lê *et al.*, 2005) |
| **Inducer** | **Object ("victim") drug** |
| Anticonvulsants inducers | Irinotecan (Crews *et al.*, 2002) |
| Antivirals | Paclitaxel (Nannan Panday *et al.*, 1999) |
| Avasimibe | Digoxin (Sahi *et al.*, 2003) |
| Clarithromycin | Colchicine (Rollot *et al.*, 2004) |
| Corticosteroids | Tacrolimus (Anglicheau *et al.*, 2003) |
| Cyclosporine | Cerivastatin (Mück *et al.*, 1999) |
| Docetaxel | Cyclosporine (Malingré *et al.*, 2001b) |
| Efavirenz | Erythromycin (Mouly *et al.*, 2002) |
| Itraconazole | Cimetidine (Karyekar *et al.*, 2004) |
| Ketoconazole | Cyclosporine (Akhlaghi *et al.*, 2001) |
| Lopinavir | Amprenavir (Basso *et al.*, 2002; Kashuba *et al.*, 2005) |
| Lopinavir and ritonavir | Phenytoin (Lim *et al.*, 2004) |
| Mefloquine | Ritonavir (Khaliq *et al.*, 2001) |
| Methylprednisolone | Tacrolimus (Shimada *et al.*, 2002) |
| Nelfinavir or indinavir | Cyclosporine (Frassetto *et al.*, 2005) |
| Phenytoin | Lopinavir (Lim *et al.*, 2004), ritonavir (Lim *et al.*, 2004) |
| Pilsicainide | Cetirizine (Tsuruoka *et al.*, 2006) |
| Posaconazole | Cyclosporine (Sansone-Parsons *et al.*, 2007) |
| Prednisone | Etoposide (Kishi *et al.*, 2004) |
| Red wine | Cyclosporine (Tsunoda *et al.*, 2001) |
| Rifampin | (R)-Talinolol (Zschiesche *et al.*, 2002), (S)-talinolol (Zschiesche *et al.*, 2002), carvedilol (Giessmann *et al.*, 2004), celecoxib (Jayasagar *et al.*, 2003), celiprolol (Lilja, Niemi and Neuvonen, 2004), dicloxacillin (Putnam *et al.*, 2005), digoxin (Greiner *et al.*, 1999; Drescher *et al.*, 2003; Larsen *et al.*, 2007) erythromycin (Paine *et al.*, 2002a), ezetimibe (Oswald *et al.*, 2006), fexofenadine (Hamman *et al.*, 2001), indinavir (Justesen *et al.*, 2004), linezolid (Egle *et al.*, 2005), morphine (Fromm *et al.*, 1997), tacrolimus (Hebert *et al.*, 1999), talinolol (Westphal *et al.*, 2000) |
| Saquinavir | Cyclosporine (Brinkman, Huysmans and Burger, 1998) |
| St. John's wort extract (*Hypericum perforatum*) | (R)-Verapamil (Tannergren *et al.*, 2004), (S)-verapamil (Tannergren *et al.*, 2004), amitriptyline (Johne *et al.*, 2002), cyclosporine (Barone *et al.*, 2000, 2001; Breidenbach *et al.*, 2000; Mai *et al.*, 2000; Ruschitzka *et al.*, 2000; Bauer *et al.*, 2003; Dresser *et al.*, 2003) digoxin (Johne *et al.*, 1999; Dürr *et al.*, 2000) fexofenadine (Xie *et al.*, 2005), simvastatin (Sugimoto *et al.*, 2001), tacrolimus (Mai *et al.*, 2003), talinolol (Schwarz *et al.*, 2007) |
| Tenofovir | Atazanavir (Taburet *et al.*, 2004) |
| Tipranavir | Loperamide (Mukwaya *et al.*, 2005) |

**TABLE 6 The expected effect of a putative inhibitor on the systemic exposure (AUC) of a drug that is a substrate for efflux or uptake transporters in the intestine, kidney, liver, or brain.**

| Tissue | Transporter | Membranelocalization | Drug substrate transport direction | Change in drug substrate AUC by inhibitors |
|---|---|---|---|---|
| Intestine | MDR1 (P-gp) | Brush border | Efflux into lumen | ↑ |
| | MRP2 | Brush border | Efflux into lumen | ↑ |
| | MRP3 | Basolateral | Efflux into blood | ↓ |
| | BCRP | Brush border | Efflux into lumen | ↑ |
| | OATP2B1 | Brush border | Uptake from lumen | ↓ |
| | PEPT1/2 | Brush border | Uptake from lumen | ↓ |
| Kidney | MDR1 (P-gp) | Brush border | Efflux into lumen | ↑ |
| | MRP2 | Brush border | Efflux into lumen | ↑ |
| | MRP4 | Brush border | Efflux into lumen | ↑ |
| | OCTN1/2 | Brush border | Efflux into lumen | ↑ |
| | OAT4 | Brush border | Efflux into lumen | ↑ |
| | OAT1/2/3 | Basolateral | Uptake from blood | ↑ |
| | OCT2 | Basolateral | Uptake from blood | ↑ |
| Liver | MDR1 (P-gp) | Canalicular | Efflux into bile | ↑ |
| | MRP2 | Canalicular | Efflux into bile | ↑ |
| | BSEP | Canalicular | Efflux into bile | ↑ |
| | BCRP | Canalicular | Efflux into bile | ↑ |
| | MRP1/3 | Sinusoidal | Efflux into blood | ↓ |
| | OATP1A2/1B1/1B3/2B1 | Sinusoidal | Uptake from blood | ↑ |
| | OCT1 | Sinusoidal | Uptake from blood | ↑ |
| | OAT2 | Sinusoidal | Uptake from blood | ↑ |
| Brain | MDR1 (P-gp) | Luminal | Efflux into luminal blood | ↑[a] |
| | MRP1/4/5 | Luminal | Efflux into luminal blood | ↑[a] |
| | BCRP | Luminal | Efflux into luminal blood | ↑[a] |
| | OATP1A2 | ND | ND | ND |

[a]Drug concentration in brain.
ND, not determined.

inhibitor on the systemic exposure (area under the curve, AUC) of a drug that is a substrate of efflux or uptake transporters in the intestine, kidney, liver, or brain are summarized in Table 6.

### 3.1 P-Glycoprotein

Since P-gp is present on the apical membrane of various tissues including the intestinal epithelia, renal proximal tubules, hepatocytes, and BBB, drug–drug interactions via inhibition or induction of P-gp can thus occur at the respective sites leading to altered pharmacokinetics or metabolic disposition of the concomitant drugs. Classic examples are P-gp mediated drug–drug interactions involving the prototype P-gp substrate digoxin, and concomitant drugs that are P-gp inhibitors such as quinidine, talinolol,

clarithromycin, itraconazole, and atorvastatin. Since digoxin can be administered intravenously or orally, the effect of P-gp inhibitors on digoxin pharmacokinetics can be attributed to effect on systemic clearance (biliary or renal) or absorption (intestinal). Inhibition of P-gp-mediated biliary or renal clearance, or inhibition of intestinal P-gp-mediated efflux, was thought to result in decreased systemic clearance or increased oral absorption, leading to elevated systemic exposure (AUCs) of digoxin. For drugs with narrow therapeutic windows such as digoxin, anticancer agents, and immunosuppressants, adverse effects resulting from P-gp-related drug interactions are particularly problematic. Although the role of P-gp in limiting brain penetration of drugs has been shown in animal models using mdr1a (–/–) mice (Seelig, Gottschlich and Devant, 1994; Schinkel *et al.*, 1996), there are few examples where clinical drug–drug interactions were demonstrated at the BBB. Using radiolabeled verapamil as the P-gp substrate and cyclosporine A (CsA) as the P-gp inhibitor, the brain : blood AUC ratio of radioactivity in humans was shown to increase by about 88% at 2.8-μM pseudo-steady-state CsA blood concentration (Sasongko *et al.*, 2005). The authors concluded that the extent of P-gp-based drug interactions at the human BBB at therapeutic CsA blood concentrations is not likely to be very pronounced.

### 3.2 BCRP

Like P-gp, drug–drug interactions involving inhibition of BCRP could occur at various organ sites including the liver, kidney, and small intestine due to the expression of BCRP in these organs. GF120918 is a potent inhibitor of BCRP and P-gp, and increases the oral bioavailability and plasma AUC of topotecan by two- to threefold (Kruijtzer *et al.*, 2002). Since topotecan is a substrate for BCRP and is only a weak-to-moderate substrate of P-gp, the effect of GF120918 on the pharmacokinetics of topotecan was thought to be due to inhibition of BCRP-mediated intestinal efflux leading to increased oral drug absorption.

### 3.3 OATP

Organic anion-transporting polypeptides (OATPs) are a superfamily of uptake transporters with wide tissue distribution including the liver, gut, and BBB. They are responsible for the transport of a variety of amphipathic organic anions such as steroid conjugates, bile acids, and drugs. OATP1B1 has been implicated as a factor in the interactions between fruit juices and fexofenadine (Dresser *et al.*, 2002), in which the systemic exposure of orally administered fexofenadine is reduced by approximately 30% when taken with grapefruit, apple, or orange juice. The mechanism appears to be the inhibition of OATP1B1-mediated intestinal absorption of fexofenadine by fruit juices and this has been demonstrated using cell lines expressing OATPs (Cvetkovic *et al.*, 1999). Hepatic drug–drug interactions that might involve OATP1B1 have also been reported. Cyclosporine A, an OATP1B1 inhibitor, increased the plasma AUC of cerivastatin by almost fourfold (Mück *et al.*, 1999). Gemfibrozil, an inhibitor of OATP1B1 and also of MRP2 (Sasaki *et al.*, 2002), increased the plasma AUC of HMG-CoA reductase inhibitors including simvastatin, lovastatin, and cerivastatin by at least twofold (Backman *et al.*, 2000; Kyrklund *et al.*, 2001; Backman *et al.*, 2002). Genetic polymorphisms in OATP1A2 expression has been shown, which appeared to be

ethnicity-dependent (Lee *et al.*, 2005) and may contribute to interindividual variability in drug disposition.

### 3.4 OAT

Since OATs are responsible for the renal excretion of a wide variety of anionic drugs, inhibition of OATs may result in decreased clearance and increased AUCs of drugs, the elimination of which is primarily due to renal clearance. For example, the OAT1 inhibitor probenecid (Hosoyamada *et al.*, 1999) reduced the renal excretion of benzylpenicillin, acyclovir, cepahalosporins, and cidofovir (Roberts *et al.*, 1981; Overbosch *et al.*, 1988; Tsuji *et al.*, 1990; Cundy, 1999), presumably by inhibiting the active tubular secretion of these drugs. Since OATs are also present in the liver and brain, there remains the possibility for drug–drug interactions to occur by reducing hepatic or brain drug uptake via inhibition of OATs.

### 3.5 OCTs

Like OATs, OCTs are responsible for the renal excretion of a wide variety of cationic drugs; inhibition of OCTs therefore may decrease the clearance and increase AUC of drugs where renal drug clearance is primarily responsible for drug elimination. Examples of drug–drug interactions involving OCTs are the increase in AUC and decrease in renal clearance of procainamide and metformin (Somogyi, McLean and Heinzow, 1983; Somogyi *et al.*, 1987) when coadministered with cimetidine, itself a substrate for active renal excretion (Drayer *et al.*, 1982). Metformin, a biguanide antihyperglicemic agent used for the treatment of type 2 diabetes mellitus, was taken up into the liver by OCT1. Genetic variation in OCT1 affects the response to metformin presumably due to variability in liver drug concentration achieved (Reitman and Schadt, 2007; Shu *et al.*, 2007).

### 3.6 MRP and BSEP

Instead of adverse effects that may result from transporter-mediated drug–drug interactions, MRP2 and BSEP have been implicated in drug and endogenous substrate-induced toxicity. Identification of ligands that are substrates or inhibitors of MRP2 or BSEP may help to reduce the possibility of drug-induced toxicity. MRP2 is involved in the biliary excretion of anion drugs and their conjugates at the bile canalicular membrane. It is believed that the toxic effects of methotrexate in the intestines result from the active excretion of methotrexate into the bile by MRP2, with subsequent accumulation in the intestine leading to the observed toxicity (Masuda *et al.*, 1997). The excretion of the reactive glucuronides of diclofenac by MRP2 and its accumulation in bile is also thought to contribute to the toxic effects diclofenac has on bile canalicular membranes (Seitz and Boelsterli, 1998). Inhibition of BSEP has been implicated in drug-induced cholestasis caused by intracellular accumulation of toxic bile salts (taurocholate). Inhibition of BSEP-mediated transport of bile salt is also suggested as a mechanism responsible for cholestasis caused by cyclosporine and estradiol-17β-glucuronides (Stieger *et al.*, 2000). Other examples of transporter involvement in toxicity include troglitazone, a diabetic agent that was withdrawn from the market due to liver toxicity. It is believed the toxicity was via a cholestatic mechanism, since troglitazone and its

sulfate conjugate have been shown to inhibit BSEP-mediated taurocholate transport (Funk *et al.*, 2001).

## 4 POLYMORPHISMS AND REGULATION OF DRUG TRANSPORTERS

Polymorphisms in various drug transporters have been identified including MDR1, MRP2, OATPs, OATs, and OCTs (Kerb, Hoffmeyer and Brinkmann, 2001; Kim *et al.*, 2001; Suzuki and Sugiyama, 2002; Tirona and Kim, 2002; Ieiri, Takane and Otsubo, 2004; Marzolini *et al.*, 2004; Ho and Kim, 2005). Unlike drug metabolism enzymes where polymorphisms have well-documented effects on drug pharmacokinetics and disposition, the clinically relevant effects of drug transporter polymorphisms are not as common and remain to be firmly established. An example of variation of drug response due to drug transporter genetic variation was suggested for OCT1, where reduced function polymorphisms of OCT1 in individuals were thought to affect the response to metformin, a biguanide for the treatment of type 2 diabetes and the hepatic uptake of which is mediated by OCT1 (Shu *et al.*, 2007). Drug transporters like cytochrome P450 (CYP) enzymes are subject to modulations via nuclear receptors. The biology and functions of nuclear receptors are discussed in great detail in separate chapters and will only be briefly described herein. Nuclear receptors that are known to be involved in the regulation of various drug transporters include pregnane X receptor (PXR), constitutive androstane receptor (CAR), glucocorticoid receptor (GR), hepatocyte nuclear factor 4α (HNF4α), farnesoid X receptor (FXR) (Lee *et al.*, 2006), and vitamin D receptor (VDR). Examples of drug transporters that are regulated by nuclear receptors, the nuclear receptors involved, and their ligands are summarized in Table 7.

### 4.1 P-Glycoprotein (P-gp)

P-gp was shown to be regulated by both PXR and CAR. Coadministration of PXR ligands, such as rifampin, induces P-gp protein expression and decreased digoxin plasma

**TABLE 7 Regulation of drug transporters: nuclear receptors and ligands.**[a]

| Target gene | Nuclear receptor | Ligands | Effect on target |
|---|---|---|---|
| *MDR1* | PXR | Flucloxacillin, artemisinin, rifampin, | Up |
| | CAR | hyperforin | Up |
| *MRP2* | PXR | Rifampin, hyperforin, phenobarbital, | Up |
| | CAR | chenodeoxycholic acid | Up |
| | FXR | | Up |
| *MRP4* | CAR | Phenobarbital | Up |
| *BSEP* | FXR | Cholic acid, chenodeoxycholic acid, | Up |
| | | lithocholic acid | Down |
| *BCRP* | PPARγ | Rosiglitazone | Up |
| | | GW9662 | Down |
| *NTCP* | GR | Dexamethasone | Up |
| *OATP8* | FXR | Chenodeoxycholic acid | Up |
| *OATP-A* | PXR | | Up |

[a]Adapted from Urquhart, Tirona and Kim (2007).

levels (Greiner *et al.*, 1999; Geick, Eichelbaum and Burk, 2001). Administration of St. John's wort extract also decreased digoxin plasma levels and increased duodenal P-gp expression (Dürr *et al.*, 2000), presumably via activation of PXR. Antimalarial artemisinin drugs were shown to induce P-gp expression by activation of PXR and CAR (Burk *et al.*, 2005a, b). Certain P-gp substrates such as cyclosporine, amitriptyline, and verapamil were shown to have elevated plasma levels during inflammation (Pheterson *et al.*, 1987; Chen *et al.*, 1994; Mayo *et al.*, 2000). Reduced transporter expression, as a result of decreased expression of PXR in the liver, was thought to be a possible mechanism leading to increased systemic exposure of P-gp substrates during an inflammation response.

### 4.2 BCRP

The BCRP was recently shown to be regulated by the nuclear receptor peroxisome proliferators-activated receptor gamma (PPARγ) in human dendritic cells (Szatmari *et al.*, 2006), as evident by increased BCRP mRNA and protein expression following treatment with the PPARγ agonist rosiglitazone.

### 4.3 MRP2

MRP2 has been shown to be regulated by PXR, CAR, and FXR. Rifampicin and hyperforin (PXR ligands), phenobarbital (a CAR activator), and chenodeoxycholic acid (a FXR ligand) increased the expression of MRP2 (Kast *et al.*, 2002). Drug–drug interactions involving induction of MRP2 though remain to be demonstrated.

### 4.4 BSEP

The BSEP is regulated by FXR. BSEP expression was shown to be up-regulated by the FXR ligand chenodeoxycholic acid and decreased by the FXR antagonist lithocholic acid (Ananthanarayanan *et al.*, 2001; Yu *et al.*, 2002). FXR, together with its target genes, is essential in maintaining normal liver function. For example, individuals affected with progressive familial intrahepatic cholestasis type I, which was caused by mutations in BSEP, were shown to have decreased expression of FXR (Chen *et al.*, 2002; Alvarez *et al.*, 2004).

### 4.5 OATP

OATP1A2 was suggested to be a target gene for PXR in breast carcinoma due to a strong correlation between OATP1A2 and PXR expression (Miki *et al.*, 2006). The FXR ligand chenodeoxycholic acid was shown to up-regulate OATP1B3 in a human hepatoma cell line (Korjamo *et al.*, 2005).

## 5 *IN VITRO* METHODS IN EVALUATION OF DRUG TRANSPORTERS

*In vitro* and *in vivo* methods to assess the involvement of drug transporters in drug disposition and interactions have been the topic of detailed discussions in several recent reviews (Lin, 2007; Xia, Milton and Gan, 2007). Several common *in vitro* approaches to assess transporter involvement are discussed here. A variety of *in vitro*

and *in vivo* models are available to assess the role of different transporters in drug disposition. Among *in vitro* systems, the colon carcinoma cell line (Caco-2) is one of the most widely used models for predicting intestinal drug absorption (Polli *et al.*, 2001; Korjamo *et al.*, 2005). Caco-2 cells express a variety of drug transporters including MDR1/P-gp and MRP2, and can be used to assess drug–drug interactions involving multiple transporters. However, the presence of multiple transporters can also be a limitation because it can be difficult to determine the involvement of a particular transporter, unless there are specific inhibitors and/or substrates available. As a result, Caco-2 cells are used mainly to evaluate P-gp interactions in a monolayer-based format, with selective inhibitors or substrates of P-gp. Various factors such as culture time, passage number, and culture conditions can affect P-gp expression; hence, these conditions need to be optimized and standardized to ensure data reproducibility. Other cell lines include canine kidney cell line (MDCK). Their advantage over Caco-2 cells is shorter culture time and the fact that they can be specifically transfected to overexpress human P-gp, thus increasing sensitivity and specificity (Polli *et al.*, 2001). Sandwich-cultured hepatocytes represent a potentially useful *in vitro model* to study drug transporters (Annaert *et al.*, 2001; Annaert and Brouwer, 2005). Biliary excretion in long-term sandwich-cultured rat hepatocytes has been shown to correlate well with *in vivo* biliary excretion, though evaluation of *in vitro–in vivo* correlation in humans remains difficult. Sandwich-cultured hepatocytes may also allow the study of the biliary excretion of metabolites, provided the relevant metabolic activities are maintained.

Efforts in cloning and transgenic technologies have also been put toward understanding transporter biology and their importance in drug discovery and development. Single-transfected cells that stably or transiently express individual transporters are used to determine whether a compound is a substrate or inhibitor of particular transporters (Sasaki *et al.*, 2004; Lee *et al.*, 2005). Double-transfected cells or monolayers that stably express human or rat uptake and efflux transporters have been developed (Sasaki *et al.*, 2004) to better understand the synergistic role of uptake and efflux transporters under *in vivo* conditions. Although these expression systems have the advantages of allowing the qualitative identification of specific drug transporters involved in a drug's disposition, quantitative prediction of *in vitro–in vivo* correlations remains to be established before these systems can gain widespread applications.

In addition to cell-based systems, canalicular membrane vesicles (CMVs) are used to assess transport of substances and drugs into the bile (Ishizuka *et al.*, 1999; Shilling *et al.*, 2006). CMVs have been shown to provide a good prediction of transporter-mediated biliary drug clearance. CMVs express many of the efflux transporters found on the canalicular membrane of hepatocytes, including P-gp, MRP2, and BSEP. CMVs can be prepared from livers of various species using similar procedures and therefore allow for the simultaneous assessment of species differences or similarities. One limitation of CMVs, however, is variability in quality between different livers (human and nonhuman primates), and strict adherence to protocols to ensure reproducible expression of the various transporters is critical. Sinusoidal membrane vesicles (SMVs) are prepared from the sinusoidal (basolateral) membrane of hepatocytes and are used to evaluate transporters expressed on the basolateral membranes of hepatocytes (Mizuno *et al.*, 2003). They are, however, not as widely used as more validated systems. Scientists from the FDA have published their perspective on using *in vitro* methods to assess whether a drug candidate is a substrate or inhibitor of efflux transporters such as P-gp, MRP2, or BCRP (Zhang *et al.*, 2006). The preferred methods were monolayer methods such as Caco-2 or MDCK cells expressing the transporter of interest, while *in vitro*

methods such as the ATPase assays (Ramachandra *et al.*, 1996; Taguchi *et al.*, 1997) were deemed not as informative.

Like CYP enzymes, the inhibition of drug transporters can result in changes in drug clearance and in drug exposure, leading to either adverse effects or inadequate pharmacological activity (Mizuno *et al.*, 2003; Beringer and Slaughter, 2005). Since P-gp is the most well-understood drug transporter with respect to its role in clinical drug–drug interactions, considerable efforts have been put into identifying and characterizing P-gp substrates or inhibitors during drug discovery and development, using *in vitro* models such as bidirectional transport employing monolayers (e.g. Caco-2 and MDCK cells), uptake/efflux of fluorescent or radiolabeled probes (e.g. Calcein-AM and rhodamine-123), or simulation of ATPase activity using membrane vesicles derived from tissues or cells expressing P-gp (Polli *et al.*, 2001). Bidirectional transport using monolayers expressing P-gp is considered the most reliable method for identifying P-gp substrates or inhibitors, since the method has the advantages of allowing the direct measurement of efflux across the cell barrier and evaluation of P-gp involvement. Similar to studies involving CYPs, criteria can be set up to identify whether a drug is a substrate (e.g. efflux ratios >2) or an inhibitor (e.g. $IC_{50}$ values in inhibiting digoxin efflux) of P-gp. However, unlike CYPs, the extrapolation of *in vitro* P-gp inhibition data to clinical situations has not been well established, in part due to the wide tissue distribution of P-gp (i.e. intestinal lumen, liver, kidney, and BBB) and that relevant inhibitor concentration that needs to be considered (Bjornsson *et al.*, 2003).

Drug transporters have been recognized to play an active role in renal and biliary excretion of many xenobiotics and endogenous substances. Since *in vivo* studies of renal or biliary excretion may not be always readily conducted during the drug discovery phase, and realizing that species differences exist in renal or biliary drug clearance, *in vitro* approaches in assessing renal or biliary clearance have been proposed, and may provide guidance on the significance of these pathways in drug clearance. Together with physicochemical properties such as molecular weight, solubility, and permeability, the identification of transporter involvement (particularly for efflux transporters P-gp and MRP2) greatly enhances our capability to predict compounds that are renally or biliary excreted. Considerable efforts continue to be exerted during the drug discovery phase to identify and select drug candidates that are not (or poor) P-gp substrates in an attempt to improve oral absorption, and for CNS-targeted drugs, to improve brain penetration. The absence of P-gp-mediated efflux, together with high passive permeability, can aid in the selection of CNS drugs (Mahar Doan *et al.*, 2002).

## 6 TRANSPORTERS–DRUG-METABOLIZING ENZYMES INTERPLAY

Due to the overlapping substrate and inhibitor selectivity for CYP3A and P-gp, it has become increasingly recognized that drug–drug interactions may also be mediated by this transporter–enzyme interplay. Wu and Benet (2005) recently proposed a revised Biopharmaceutics Classification System (BCS) that may be useful in predicting overall drug disposition, including routes of elimination and effects of efflux and uptake transporters on oral drug absorption, and the clinical significance of transporter–enzyme interplay on oral bioavailability and drug–drug interactions. Transporter–drug-metabolizing enzyme interplay, primarily involving P-gp-CYP3A4 in the intestine, is thought to limit the oral absorption of certain drugs. Though *in vitro* systems, such as cell monolayers expressing CYP3A4 and P-gp, to assess the role of transporter–enzyme

interplay on drug absorption have been used (Paine *et al.*, 2002a, b; Cummins *et al.*, 2004), the quantitative extrapolation of *in vitro* data to *in vivo* situations remains to be established.

It has been suggested (Wu and Benet, 2005) that drug–drug interactions as a result of transporter–enzyme interplay might be more significant for drugs with low solubility and extensive metabolism (class 2 drugs). An example of a drug where transporter–enzyme interplay may play an important role in its metabolic disposition is sirolimus, a macrolide antibiotic used as an immunosuppressive agent to prevent allograft rejection in kidney transplantation. Sirolimus was metabolized by CYP3A and a majority of the administered dose was excreted into the bile (Leung *et al.*, 2006). Preclinical *in vitro* studies, using models including liver microsomes and Caco-2 cell monolayers expressing CYP3A, indicated that sirolimus was susceptible to macrolide ring opening via hydrolysis/dehydration, and cytochrome P450 CYP3A-mediated oxidation and P-gp-mediated efflux (Paine *et al.*, 2002b). *In vitro* results therefore suggested that sirolimus would be subject to CYP3A- and non-CYP-mediated metabolic clearance, as well as P-gp-mediated biliary excretion. Moreover, its oral bioavailability could be limited by intestinal and hepatic CYP3A-catalyzed first-pass effect and intestinal P-gp-mediated efflux. Inhibition of P-gp in a CYP3A4-transfected Caco-2 cellular system in the apical to basolateral direction decreased the extraction ratio of sirolimus, suggesting that inhibition of intestinal drug efflux can lead to decreased intestinal metabolism of the concomitant drug due to decreased access of the drug to the enzyme by reducing recycling (Paine *et al.*, 2002b; Cummins *et al.*, 2004). In contrast, inhibition of P-gp in the basolateral to apical direction increased the extraction ratio of sirolimus, suggesting that inhibition of hepatic drug efflux can increase the extent of hepatic metabolism of the concomitant drug due to increased access of the drug to the active enzyme. An example of a drug where lack of proper integration of drug–drug interaction data can lead to drastic consequence is mibefradil, a calcium channel blocker used for hypertension, which was withdrawn from the market in 1998 due to serious drug interactions with concomitant medications that are CYP3A or P-gp substrates. Continued research is needed in this area to further demonstrate the clinical significance of transporter–enzyme interplay, and to assess approaches to predict likely clinical outcome from *in vitro* methodologies for drugs that are inhibitors of transporters but not of CYP enzymes.

## 7 OUTLOOK

Despite advances made in the drug transporter field in the last decade, further understanding and improved tools in certain areas are needed. An area of continued importance is understanding and prediction of clinical drug–drug interactions involving concomitant substrates and inhibitors of drug transporters. Due to the ubiquitous nature of drug transporters and possible sites of substrate–inhibitor interaction including absorption (intestinal lumen), distribution (BBB, cell uptake), and elimination (liver, kidney), challenges continue with respect to the usage of the most appropriate *in vivo* and *in vitro* models and parameters in the prediction of clinical outcome. Although parameters such *in vitro* $K_i$ and plasma inhibitor concentrations have been proposed to predict drug transporter-mediated drug–drug interactions (Zhang *et al.*, 2006; Zhang *et al.*, 2008), the appropriateness of the model remains unclear due to the complex nature of drug transporter kinetics that do not necessarily follow simple

Michaelis–Menten kinetics. There remains limited published information on species differences in drug transporters, though differences in transporter activity, such as P-gp, MRP2, OATPs, and OATs, were observed between human and other animal species (Lin and Yamazaki, 2003; Mizuno *et al.*, 2003; Van Montfoort *et al.*, 2003; Shilling *et al.*, 2006). Therefore, due to observed or potential species differences in various drug transporter activities, the use of animal models to extrapolate to human situations should be performed with caution.

The study of the regulation of drug transporters has made tremendous advances in the past decade, and will continue to be an area of active research. Recognition of the role of nuclear receptors in drug transporter expression, and their associated polymorphism, has provided insight into the molecular mechanism for transporter-mediated drug–drug and drug–endogenous substrate interactions and drug toxicity, and provided the opportunity of individualized treatment based on interindividual variability in drug disposition and response. Inflammation-mediated changes in drug disposition have been noted, and might be related to decreased mRNA and protein expressions of drug transporters such as P-gp, MRPs, or OATPs during periods of inflammation (Petrovic, Teng and Piquette-Miller, 2007). Future research will need to provide further understanding of the molecular mechanisms involved in the regulation of drug transporters.

The emerging importance of drug transporters, notably P-gp, in drug disposition and drug–drug interactions has recently prompted scientists from the FDA to publish their scientific perspective on the subject matter (Zhang *et al.*, 2006). It was suggested that our current knowledge allowed the use of *in vitro* models, such as a bidirectional transport assay, to evaluate whether a drug candidate is a substrate or inhibitor of P-gp. Positive outcomes from the *in vitro* studies may warrant confirmation in clinical drug interaction studies employing a P-gp inhibitor (if the drug candidate is a potential P-gp substrate) or a P-gp substrate such as digoxin (if the drug candidate is a potential P-gp inhibitor). Due to current limited knowledge or availability of standardized methods, routine *in vitro* evaluation of P-gp induction or other transporter-based interactions was not recommended at this time. Therefore, additional knowledge and established methods for various transporters should continue to be an active area of research.

## REFERENCES

Advani, R., Fisher, G.A., Lum, B.L., Hausdorff, J., Halsey, J., Litchman, M., Sikic, B.I. (2001). A phase I trial of doxorubicin, paclitaxel, and valspodar (PSC 833), a modulator of multidrug resistance. *Clin. Cancer Res.*, *7*, 1221–1229.

Airoldi, M., Cattel, L., Marchionatti, S., Recalenda, V., Pedani, F., Tagini, V., Bumma, C., Beatrice, F., Succo, G., Maria Gabriele, A. (2003). Docetaxel and vinorelbine in recurrent head and neck cancer: pharmacokinetic and clinical results. *Am. J. Clin. Oncol.*, *26*, 378–381.

Akashi, M., Tanaka, A., Takikawa, H. (2006). Effect of cyclosporin A on the biliary excretion of cholephilic compounds in rats. *Hepatol. Res.*, *34*, 193–198.

Akhlaghi, F., Keogh, A.M., McLachlan, A.J., Kaan, A. (2001). Pharmacokinetics of cyclosporine in heart transplant recipients receiving metabolic inhibitors. *J. Heart. Lung. Transplant.*, *20*, 431–438.

Alrefai, W.A., Gill, R.K. (2007). Bile acid transporters: structure, function, regulation and pathophysiological implications. *Pharm. Res.*, *24*, 1803–1823.

Alvarez, L., Jara, P., Sánchez-Sabaté, E., Hierro, L., Larrauri, J., Díaz, M.C., Camarena, C., De la Vega, A., Frauca, E., López-Collazo, E., Lapunzina, P. (2004). Reduced hepatic expression of farnesoid X receptor in hereditary cholestasis associated to mutation in ATP8B1. *Hum. Mol. Genet.*, *13*, 2451–2460.

Amsden, G.W., Nafziger, A.N., Foulds, G., Cabelus, L.J. (2000). A study of the pharmacokinetics of azithromycin and nelfinavir when coadministered in healthy volunteers. *J. Clin. Pharmacol.*, *40*, 1522–1527.

Ananthanarayanan, M., Balasubramanian, N., Makishima, M., Mangelsdorf, D.J., Suchy, F.J. (2001). Human bile salt export pump promoter is transactivated by the farnesoid X receptor/bile acid receptor. *J. Biol. Chem.*, *276*, 28857–28865.

Anglicheau, D., Flamant, M., Schlageter, M.H., Martinez, F., Cassinat, B., Beaune, P., Legendre, C., Thervet, E. (2003). Pharmacokinetic interaction between corticosteroids and tacrolimus after renal transplantation. *Nephrol. Dial. Transplant.*, *18*, 2409–2414.

Annaert, P.P., Brouwer, K.L. (2005). Assessment of drug interactions in hepatobiliary transport using rhodamine 123 in sandwich-cultured rat hepatocytes. *Drug Metab. Dispos.*, *33*, 388–394.

Annaert, P.P., Turncliff, R.Z., Booth, C.L., Thakker, D.R., Brouwer, K.L. (2001). P-glycoprotein-mediated in vitro biliary excretion in sandwich-cultured rat hepatocytes. *Drug Metab. Dispos.*, *29*, 1277–1283.

Anzai, N., Kanai, Y., Endou, H. (2006). Organic anion transporter family: current knowledge. *J. Pharmacol. Sci.*, *100*, 411–426.

Asberg, A., Hartmann, A., Fjeldså, E., Bergan, S., Holdaas, H. (2001). Bilateral pharmacokinetic interaction between cyclosporine A and atorvastatin in renal transplant recipients. *Am. J. Transplant.*, *1*, 382–386.

Backman, J.T., Kyrklund, C., Kivistö, K.T., Wang, J.S., Neuvonen, P.J. (2000). Plasma concentrations of active simvastatin acid are increased by gemfibrozil. *Clin. Pharmacol. Ther.*, *68*, 122–129.

Backman, J.T., Kyrklund, C., Neuvonen, M., Neuvonen, P.J. (2002). Gemfibrozil greatly increases plasma concentrations of cerivastatin. *Clin. Pharmacol. Ther.*, *72*, 685–691.

Baris, N., Kalkan, S., Güneri, S., Bozdemir, V., Guven, H. (2006). Influence of carvedilol on serum digoxin levels in heart failure: is there any gender difference? *Eur. J. Clin. Pharmacol.*, *62*, 535–538.

Barone, G.W., Gurley, B.J., Ketel, B.L., Lightfoot, M.L., Abul-Ezz, S.R. (2000). Drug interaction between St. John's wort and cyclosporine. *Ann. Pharmacother.*, *34*, 1013–1016.

Barone, G.W., Gurley, B.J., Ketel, B.L., Abul-Ezz, S.R. (2001). Herbal supplements: a potential for drug interactions in transplant recipients. *Transplantation*, *71*, 239–241.

Bartlett, N.L., Lum, B.L., Fisher, G.A., Brophy, N.A., Ehsan, M.N., Halsey, J., Sikic, B.I. (1994). Phase I trial of doxorubicin with cyclosporine as a modulator of multidrug resistance. *J. Clin. Oncol.*, *2*, 835–842.

Basso, S., Solas, C., Quinson, A.M., Ravaux, I., Poizot-Martin, I., Bacconier, J., Durand, A., Lacarelle, B. (2002). Pharmacokinetic interaction between lopinavir/r and amprenavir in salvage therapy. *J. Acquir. Immune. Defic. Syndr.*, *31*, 115–117.

Bauer, S., Störmer, E., Johne, A., Krüger, H., Budde, K., Neumayer, H.H., Roots, I., Mai, I. (2003). Alterations in cyclosporin A pharmacokinetics and metabolism during treatment with St John's wort in renal transplant patients. *Br. J. Clin. Pharmacol.*, *55*, 203–211.

Beringer, P.M., Slaughter, R.L. (2005). Transporters and their impact on drug disposition. *Ann. Pharmacother.*, *39*, 1097–1108.

Bertrand, Y., Capdeville, R., Balduck, N., Philippe, N. (1992). Cyclosporin A used to reverse drug resistance increases vincristine neurotoxicity. *Am. J. Hematol.*, *40*, 158–159.

Bjornsson, T.D., Callaghan, J.T., Einolf, H.J., Fischer, V., Gan, L., Grimm, S., Kao, J., King, S.P., Miwa, G., Ni, L., Kumar, G., McLeod, J., Obach, R.S., Roberts, S., Roe, A., Shah, A., Snikeris,

F., Sullivan, J.T., Tweedie, D., Vega, J.M., Walsh, J., Wrighton, S.A. (2003). The conduct of in vitro and in vivo drug–drug interaction studies: a Pharmaceutical Research and Manufacturers of America (PhRMA) perspective. *Drug Metab. Dispos.*, *31*, 815–832.

Bodo, A., Bakos, E., Szeri, F., Varadi, A., Sarkadi, B. (2003). The role of multidrug transporters in drug availability, metabolism and toxicity. *Toxicol. Lett.*, *140*–141, 133–143.

Boffito, M., Kurowski, M., Kruse, G., Hill, A., Benzie, A.A., Nelson, M.R., Moyle, G.J., Gazzard, B.G., Pozniak, A.L. (2004). Atazanavir enhances saquinavir hard-gel concentrations in a ritonavir-boosted once-daily regimen. *AIDS*, *18*, 1291–1297.

Boffito, M., Maitland, D., Dickinson, L., Back, D., Hill, A., Fletcher, C., Moyle, G., Nelson, M., Gazzard, B., Pozniak, A. (2006). Pharmacokinetics of saquinavir hard-gel/ritonavir and atazanavir when combined once daily in HIV Type 1-infected individuals administered different atazanavir doses. *AIDS Res. Hum. Retroviruses*, *22*, 749–756.

Bogman, K., Zysset, Y., Degen, L., Hopfgartner, G., Gutmann, H., Alsenz, J., Drewe, J. (2005). P-glycoprotein and surfactants: effect on intestinal talinolol absorption. *Clin. Pharmacol. Ther.*, *77*, 24–32.

Booker, B.M., Magee, M.H., Blum, R.A., Lates, C.D., Jusko, W.J. (2002). Pharmacokinetic and pharmacodynamic interactions between diltiazem and methylprednisolone in healthy volunteers. *Clin. Pharmacol. Ther.*, *72*, 370–382.

Boote, D.J., Dennis, I.F., Twentyman, P.R., Osborne, R.J., Laburte, C., Hensel, S., Smyth, J.F., Brampton, M.H., Bleehen, N.M. (1996). Phase I study of etoposide with SDZ PSC 833 as a modulator of multidrug resistance in patients with cancer. *J. Clin. Oncol.*, *14*, 610–618.

Breidenbach, T., Kliem, V., Burg, M., Radermacher, J., Hoffmann, M.W., Klempnauer, J. (2000). Profound drop of cyclosporin A whole blood trough levels caused by St. John's wort (Hypericum perforatum). *Transplantation*, *69*, 2229–2230.

Brinkman, K., Huysmans, F., Burger, D.M. (1998). Pharmacokinetic interaction between saquinavir and cyclosporine. *Ann. Intern. Med.*, *129*, 914–915.

Britten, C.D., Baker, S.D., Denis, L.J., Johnson, T., Drengler, R., Siu, L.L., Duchin, K., Kuhn, J., Rowinsky, E.K. (2000). Oral paclitaxel and concurrent cyclosporin A: targeting clinically relevant systemic exposure to paclitaxel. *Clin. Cancer Res.*, *6*, 3459–3468.

Burk, O., Arnold, K.A., Geick, A., Tegude, H., Eichelbaum, M. (2005a). A role for constitutive androstane receptor in the regulation of human intestinal MDR1 expression. *Biol. Chem.*, *386*, 503–513.

Burk, O., Arnold, K.A., Nussler, A.K., Schaeffeler, E., Efimova, E., Avery, B.A., Avery, M.A., Fromm, M.F., Eichelbaum, M. (2005b). Antimalarial artemisinin drugs induce cytochrome P450 and MDR1 expression by activation of xenosensors pregnane X receptor and constitutive androstane receptor. *Mol. Pharmacol.*, *67*, 1954–1965.

Buss, N., Snell, P., Bock, J., Hsu, A., Jorga, K. (2001). Saquinavir and ritonavir pharmacokinetics following combined ritonavir and saquinavir (soft gelatin capsules) administration. *Br. J. Clin. Pharmacol.*, *52*, 255–264.

Chen, H.L., Chang, P.S., Hsu, H.C., Ni, Y.H., Hsu, H.Y., Lee, J.H., Jeng, Y.M., Shau, W.Y., Chang, M.H. (2002). FIC1 and BSEP defects in Taiwanese patients with chronic intrahepatic cholestasis with low gamma-glutamyltranspeptidase levels. *J. Pediatr.*, *140*, 119–124.

Chen, Y.L., Le Vraux, V., Leneveu, A., Dreyfus, F., Stheneur, A., Florentin, I., De Sousa, M., Giroud, J.P., Flouvat, B., Chauvelot-Moachon, L. (1994). Acute-phase response, interleukin-6, and alteration of cyclosporine pharmacokinetics. *Clin. Pharmacol. Ther.*, *55*, 649–660.

Chi, K.N., Chia, S.K., Dixon, R., Newman, M.J., Wacher, V.J., Sikic, B., Gelmon, K.A. (2005). A phase I pharmacokinetic study of the P-glycoprotein inhibitor, ONT-093, in combination with paclitaxel in patients with advanced cancer. *Invest. New Drugs*, *23*, 311–315.

Chico, I., Kang, M.H., Bergan, R., Abraham, J., Bakke, S., Meadows, B., Rutt, A., Robey, R., Choyke, P., Merino, M., Goldspiel, B., Smith, T., Steinberg, S., Figg, W.D., Fojo, T., Bates, S. (2001). Phase

I study of infusional paclitaxel in combination with the P-glycoprotein antagonist PSC 833. *J. Clin. Oncol.*, 19, 832–842.

Chittick, G.E., Zong, J., Blum, M.R., Sorbel, J.J., Begley, J.A., Adda, N., Kearney, B.P. (2006). Pharmacokinetics of tenofovir disoproxil fumarate and ritonavir-boosted saquinavir mesylate administered alone or in combination at steady state. *Antimicrob. Agents. Chemother.*, *50*, 1304–1310.

Crews, K.R., Stewart, C.F., Jones-Wallace, D., Thompson, S.J., Houghton, P.J., Heideman, R.L., Fouladi, M., Bowers, D.C., Chintagumpala, M.M., Gajjar, A. (2002). Altered irinotecan pharmacokinetics in pediatric high-grade glioma patients receiving enzyme-inducing anticonvulsant therapy. *Clin. Cancer Res.*, *8*, 2202–2209.

Cummins, C.L., Jacobsen, W., Christians, U., Benet, L.Z. (2004). CYP3A4-transfected Caco-2 cells as a tool for understanding biochemical absorption barriers: studies with sirolimus and midazolam. *J. Pharmacol. Exp. Ther.*, *308*, 143–155.

Cundy, K.C. (1999). Clinical pharmacokinetics of the antiviral nucleotide analogues cidofovir and adefovir. *Clin. Pharmacokinet.*, *36*, 127–143.

Cvetkovic, M., Leake, B., Fromm, M.F., Wilkinson, G.R., Kim, R.B. (1999). OATP and P-glycoprotein transporters mediate the cellular uptake and excretion of fexofenadine. *Drug Metab. Dispos.*, *27*, 866–871.

Di Marco, M.P., Edwards, D.J., Wainer, I.W., Ducharme, M.P. (2002). The effect of grapefruit juice and seville orange juice on the pharmacokinetics of dextromethorphan: the role of gut CYP3A and P-glycoprotein. *Life Sci.*, *71*, 1149–1160.

Diéras, V., Bonneterre, J., Laurence, V., Degardin, M., Pierga, J.Y., Bonneterre, M.E., Marreaud, S., Lacombe, D., Fumoleau, P. (2005). Phase I combining a P-glycoprotein inhibitor, MS209, in combination with docetaxel in patients with advanced malignancies. *Clin. Cancer Res.*, *11*, 6256–6260.

Ding, R., Tayrouz, Y., Riedel, K.D., Burhenne, J., Weiss, J., Mikus, G., Haefeli, W.E. (2004). Substantial pharmacokinetic interaction between digoxin and ritonavir in healthy volunteers. *Clin. Pharmacol. Ther.*, *76*, 73–84.

Dorani, H., Schützer, K.M., Sarich, T.C., Wall, U., Logren, U., Ohlsson, L., Eriksson, U.G. (2007). Pharmacokinetics and pharmacodynamics of the oral direct thrombin inhibitor ximelagatran co-administered with different classes of antibiotics in healthy volunteers. *Eur. J. Clin. Pharmacol.*, *63*, 571–581.

Drayer, D.E., Romankiewicz, J., Lorenzo, B., Reidenberg, M.M. (1982). Age and renal clearance of cimetidine. *Clin. Pharmacol. Ther.*, *31*, 45–50.

Drescher, S., Glaeser, H., Mürdter, T., Hitzl, M., Eichelbaum, M., Fromm, M.F. (2003). P-glycoprotein-mediated intestinal and biliary digoxin transport in humans. *Clin. Pharmacol. Ther.*, *73*, 223–231.

Dresser, G.K., Bailey, D.G., Leake, B.F., Schwarz, U.I., Dawson, P.A., Freeman, D.J., Kim, R.B. (2002). Fruit juices inhibit organic anion transporting polypeptide-mediated drug uptake to decrease the oral availability of fexofenadine. *Clin. Pharmacol. Ther.*, *71*, 11–20.

Dresser, G.K., Schwarz, U.I., Wilkinson, G.R., Kim, R.B. (2003). Coordinate induction of both cytochrome P4503A and MDR1 by St John's wort in healthy subjects. *Clin. Pharmacol. Ther.*, *73*, 41–50.

Dürr, D., Stieger, B., Kullak-Ublick, G.A., Rentsch, K.M., Steinert, H.C., Meier, P.J., St, Fattinger, K. (2000). John's Wort induces intestinal P-glycoprotein/MDR1 and intestinal and hepatic CYP3A4. *Clin. Pharmacol. Ther.*, *68*, 598–604.

Eberl, S., Renner, B., Neubert, A., Reisig, M., Bachmakov, I., König, J., Dörje, F., Mürdter, T.E., Ackermann, A., Dormann, H., Gassmann, K.G., Hahn, E.G., Zierhut, S., Brune, K., Fromm, M.F. (2007). Role of p-glycoprotein inhibition for drug interactions : evidence from in vitro and pharmacoepidemiological studies. *Clin. Pharmacokinet.*, *46*, 1039–1049.

Egle, H., Trittler, R., Kümmerer, K., Lemmen, S.W. (2005). Linezolid and rifampin: drug interaction contrary to expectations? *Clin. Pharmacol. Ther.*, *77*, 451–453.

Endres, C.J., Hsiao, P., Chung, F.S., Unadkat, J.D. (2006). The role of transporters in drug interactions. *Eur. J. Pharm. Sci.*, *27*, 501–517.

Eriksson, U.G., Dorani, H., Karlsson, J., Fritsch, H., Hoffmann, K.J., Olsson, L., Sarich, T.C., Wall, U., Schützer, K.M. (2006). Influence of erythromycin on the pharmacokinetics of ximelagatran may involve inhibition of P-glycoprotein-mediated excretion. *Drug Metab. Dispos.*, *34*, 775–782.

Floren, L.C., Bekersky, I., Benet, L.Z., Mekki, Q., Dressler, D., Lee, J.W., Roberts, J.P., Hebert, M.F. (1997). Tacrolimus oral bioavailability doubles with coadministration of ketoconazole. *Clin. Pharmacol. Ther.*, *62*, 41–49.

Fogli, S., Danesi, R., Gennari, A., Donati, S., Conte, P.F., Del Tacca, M. (2002). Gemcitabine, epirubicin and paclitaxel: pharmacokinetic and pharmacodynamic interactions in advanced breast cancer. *Ann. Oncol.*, *13*, 919–927.

Frassetto, L., Baluom, M., Jacobsen, W., Christians, U., Roland, M.E., Stock, P.G., Carlson, L., Benet, L.Z. (2005). Cyclosporine pharmacokinetics and dosing modifications in human immunodeficiency virus-infected liver and kidney transplant recipients. *Transplantation*, *80*, 13–17.

Frassetto, L.A., Poon, S., Tsourounis, C., Valera, C., Benet, L.Z. (2007). Effects of uptake and efflux transporter inhibition on erythromycin breath test results. *Clin. Pharmacol. Ther.*, *81*, 828–832.

Fromm, M.F., Eckhardt, K., Li, S., Schänzle, G., Hofmann, U., Mikus, G., Eichelbaum, M. (1997). Loss of analgesic effect of morphine due to coadministration of rifampin. *Pain*, *72*, 261–267.

Funk, C., Pantze, M., Jehle, L., Ponelle, C., Scheuermann, G., Lazendic, M., Gasser, R. (2001). Troglitazone-induced intrahepatic cholestasis by an interference with the hepatobiliary export of bile acids in male and female rats. Correlation with the gender difference in troglitazone sulfate formation and the inhibition of the canalicular bile salt export pump (Bsep) by troglitazone and troglitazone sulfate. *Toxicology*, *167*, 83–98.

Geick, A., Eichelbaum, M., Burk, O. (2001). Nuclear receptor response elements mediate induction of intestinal MDR1 by rifampin. *J. Biol. Chem.*, *276*, 14581–14587.

Giaccone, G., Linn, S.C., Welink, J., Catimel, G., Stieltjes, H., van der Vijgh, W.J., Eeltink, C., Vermorken, J.B., Pinedo, H.M. (1997). A dose-finding and pharmacokinetic study of reversal of multidrug resistance with SDZ PSC 833 in combination with doxorubicin in patients with solid tumors. *Clin. Cancer Res.*, *3*, 2005–2015.

Giessmann, T., Modess, C., Hecker, U., Zschiesche, M., Dazert, P., Kunert-Keil, C., Warzok, R., Engel, G., Weitschies, W., Cascorbi, I., Kroemer, H.K., Siegmund, W. (2004). CYP2D6 genotype and induction of intestinal drug transporters by rifampin predict presystemic clearance of carvedilol in healthy subjects. *Clin. Pharmacol. Ther.*, *75*, 213–222.

Glaeser, H., Bailey, D.G., Dresser, G.K., Gregor, J.C., Schwarz, U.I., McGrath, J.S., Jolicoeur, E., Lee, W., Leake, B.F., Tirona, R.G., Kim, R.B. (2007). Intestinal drug transporter expression and the impact of grapefruit juice in humans. *Clin. Pharmacol. Ther.*, *81*, 362–370.

Greiner, B., Eichelbaum, M., Fritz, P., Kreichgauer, H.P., von Richter, O., Zundler, J., Kroemer, H.K. (1999). The role of intestinal P-glycoprotein in the interaction of digoxin and rifampin. *J. Clin. Invest.*, *104*, 147–153.

Gupta, S., Banfield, C., Kantesaria, B., Marino, M., Clement, R., Affrime, M., Batra, V. (2001). Pharmacokinetic and safety profile of desloratadine and fexofenadine when coadministered with azithromycin: a randomized, placebo-controlled, parallel-group study. *Clin. Ther.*, *23*, 451–466.

Haas, D.W., Johnson, B., Nicotera, J., Bailey, V.L., Harris, V.L., Bowles, F.B., Raffanti, S., Schranz, J., Finn, T.S., Saah, A.J., Stone, J. (2003). Effects of ritonavir on indinavir pharmacokinetics in cerebrospinal fluid and plasma. *Antimicrob. Agents. Chemother.*, *47*, 2131–2137.

Hamman, M.A., Bruce, M.A., Haehner-Daniels, B.D., Hall, S.D. (2001). The effect of rifampin administration on the disposition of fexofenadine. *Clin. Pharmacol. Ther.*, *69*, 114–121.

Hebert, M.F., Fisher, R.M., Marsh, C.L., Dressler, D., Bekersky, I. (1999). Effects of rifampin on tacrolimus pharmacokinetics in healthy volunteers. *J. Clin. Pharmacol.*, *39*, 91–96.

Hirano, M., Maeda, K., Hayashi, H., Kusuhara, H., Sugiyama, Y. (2005). Bile salt export pump (BSEP/ABCB11) can transport a nonbile acid substrate, pravastatin. *J. Pharmacol. Exp. Ther.*, *314*, 876–882.

Ho, R.H., Kim, R.B. (2005). Transporters and drug therapy: implications for drug disposition and disease. *Clin. Pharmacol. Ther.*, *78*, 260–277.

Hosoyamada, M., Sekine, T., Kanai, Y., Endou, H. (1999). Molecular cloning and functional expression of a multispecific organic anion transporter from human kidney. *Am. J. Physiol.*, *276*, F122–128.

Ieiri, I., Takane, H., Otsubo, K. (2004). The MDR1(ABCB1) gene polymorphism and its clinical implications. *Clin. Pharmacokinet.*, *43*, 553–576.

Innocenti, F., Undevia, S.D., Ramírez, J., Mani, S., Schilsky, R.L., Vogelzang, N.J., Prado, M., Ratain, M.J. (2004). A phase I trial of pharmacologic modulation of irinotecan with cyclosporine and phenobarbital. *Clin. Pharmacol. Ther.*, *76*, 490–502.

Ishizuka, H., Konno, K., Shiina, T., Naganuma, H., Nishimura, K., Ito, K., Suzuki, H., Sugiyama, Y. (1999). Species differences in the transport activity for organic anions across the bile canalicular membrane. *J. Pharmacol. Exp. Ther.*, *290*, 1324–1330.

Jalava, K.M., Partanen, J., Neuvonen, P.J. (1997). Itraconazole decreases renal clearance of digoxin. *Ther. Drug Monit.*, *19*, 609–613.

Jayasagar, G., Krishna Kumar, M., Chandrasekhar, K., Madhusudan Rao, Y. (2003). Influence of rifampicin pretreatment on the pharmacokinetics of celecoxib in healthy male volunteers. *Drug Metab. Drug Interact.*, *19*, 287–295.

Johne, A., Brockmöller, J., Bauer, S., Maurer, A., Langheinrich, M., Roots, I. (1999). Pharmacokinetic interaction of digoxin with an herbal extract from St John's wort (Hypericum perforatum) *Clin. Pharmacol. Ther.*, *66*, 338–345.

Johne, A., Schmider, J., Brockmöller, J., Stadelmann, A.M., Störmer, E., Bauer, S., Scholler, G., Langheinrich, M., Roots, I. (2002). Decreased plasma levels of amitriptyline and its metabolites on comedication with an extract from St. John's wort (Hypericum perforatum). *J. Clin. Psychopharmacol.*, *22*, 46–54.

Juan, H., Terhaag, B., Cong, Z., Bi-Kui, Z., Rong-Hua, Z., Feng, W., Fen-Li, S., Juan, S., Jing, T., Wen-Xing, P. (2007). Unexpected effect of concomitantly administered curcumin on the pharmacokinetics of talinolol in healthy Chinese volunteers. *Eur. J. Clin. Pharmacol.*, *63*, 663–668.

Justesen, U.S., Andersen, A.B., Klitgaard, N.A., Brøsen, K., Gerstoft, J., Pedersen, C. (2004). Pharmacokinetic interaction between rifampin and the combination of indinavir and low-dose ritonavir in HIV-infected patients. *Clin. Infect. Dis.*, *38*, 426–429.

Kanazawa, S., Ohkubo, T., Sugawara, K. (2001). The effects of grapefruit juice on the pharmacokinetics of erythromycin. *Eur. J. Clin. Pharmacol.*, *56*, 799–803.

Kantola, T., Kivistö, K.T., Neuvonen, P.J. (1998). Erythromycin and verapamil considerably increase serum simvastatin and simvastatin acid concentrations. *Clin. Pharmacol. Ther.*, *64*, 177–182.

Karyekar, C.S., Eddington, N.D., Briglia, A., Gubbins, P.O., Dowling, T.C. (2004). Renal interaction between itraconazole and cimetidine. *J. Clin. Pharmacol.*, *44*, 919–927.

Kashuba, A.D., Tierney, C., Downey, G.F., Acosta, E.P., Vergis, E.N., Klingman, K., Mellors, J.W., Eshleman, S.H., Scott, T.R., Collier, A.C. (2005). Combining fosamprenavir with lopinavir/ritonavir substantially reduces amprenavir and lopinavir exposure: ACTG protocol A5143 results. *AIDS*, *19*, 145–512.

Kast, H.R., Goodwin, B., Tarr, P.T., Jones, S.A., Anisfeld, A.M., Stoltz, C.M., Tontonoz, P., Kliewer, S., Willson, T.M., Edwards, P.A. (2002). Regulation of multidrug resistance-associated protein 2 (ABCC2) by the nuclear receptors pregnane X receptor, farnesoid X-activated receptor, and constitutive androstane receptor. *J. Biol. Chem.*, *277*, 2908–2915.

Kaukonen, K.M., Olkkola, K.T., Neuvonen, P.J. (1997). Itraconazole increases plasma concentrations of quinidine. *Clin. Pharmacol. Ther.*, *62*, 510–517.

Kehrer, D.F., Mathijssen, R.H., Verweij, J., de Bruijn, P., Sparreboom, A. (2002). Modulation of irinotecan metabolism by ketoconazole. *J. Clin. Oncol.*, *20*, 3122–3129.

Kerb, R., Hoffmeyer, S., Brinkmann, U. (2001). ABC drug transporters: hereditary polymorphisms and pharmacological impact in MDR1, MRP1 and MRP2. *Pharmacogenetics*, *2*, 51–64.

Kerr, D.J., Graham, J., Cummings, J., Morrison, J.G., Thompson, G.G., Brodie, M.J., Kaye, S.B. (1986). The effect of verapamil on the pharmacokinetics of adriamycin. *Cancer Chemother. Pharmacol.*, *18*, 239–242.

Khaliq, Y., Gallicano, K., Venance, S., Kravcik, S., Cameron, D.W. (2000). Effect of ketoconazole on ritonavir and saquinavir concentrations in plasma and cerebrospinal fluid from patients infected with human immunodeficiency virus. *Clin. Pharmacol. Ther.*, *68*, 637–646.

Khaliq, Y., Gallicano, K., Tisdale, C., Carignan, G., Cooper, C., McCarthy, A. (2001). Pharmacokinetic interaction between mefloquine and ritonavir in healthy volunteers. *Br. J. Clin. Pharmacol.*, *51*, 591–600.

Kharasch, E.D., Hoffer, C., Whittington, D., Sheffels, P. (2003). Role of P-glycoprotein in the intestinal absorption and clinical effects of morphine. *Clin. Pharmacol. Ther.*, *74*, 543–554.

Kharasch, E.D., Hoffer, C., Altuntas, T.G., Whittington, D. (2004). Quinidine as a probe for the role of p-glycoprotein in the intestinal absorption and clinical effects of fentanyl. *J. Clin. Pharmacol.*, *44*, 224–233.

Kim, R.B., Leake, B.F., Choo, E.F., Dresser, G.K., Kubba, S.V., Schwarz, U.I., Taylor, A., Xie, H.G., McKinsey, J., Zhou, S., Lan, L.B., Schuetz, J.D., Schuetz, E.G., Wilkinson, G.R. (2001). Identification of functionally variant MDR1 alleles among European American and African Americans. *Clin. Pharmacol. Ther.*, *70*, 189–199.

Kishi, S., Yang, W., Boureau, B., Morand, S., Das, S., Chen, P., Cook, E.H., Rosner, G.L., Schuetz, E., Pui, C.H., Relling, M.V. (2004). Effects of prednisone and genetic polymorphisms on etoposide disposition in children with acute lymphoblastic leukemia. *Blood*, *103*, 67–72.

Koepsell, H., Lips, K., Volk, C. (2007). Polyspecific organic cation transporters: structure, function, physiological roles, and biopharmaceutical implications. *Pharm. Res.*, *24*, 1227–1251.

Korjamo, T., Honkakoski, P., Toppinen, M.R., Niva, S., Reinisalo, M., Palmgren, J.J., Monkkonen, J. (2005). Absorption properties and P-glycoprotein activity of modified Caco-2 cell lines. *Eur. J. Pharm. Sci.*, *26*, 266–279.

Kovarik, J.M., Purba, H.S., Pongowski, M., Gerbeau, C., Humbert, H., Mueller, E.A. (1998). Pharmacokinetics of dexamethasone and valspodar, a P-glycoprotein (mdr1) modulator: implications for coadministration. *Pharmacotherapy*, *18*, 1230–1236.

Kovarik, J.M., Rigaudy, L., Guerret, M., Gerbeau, C., Rost, K.L. (1999). Longitudinal assessment of a P-glycoprotein-mediated drug interaction of valspodar on digoxin. *Clin. Pharmacol. Ther.*, *66*, 391–400.

Krishna, R., Bergman, A., Larson, P., Cote, J., Lasseter, K., Dilzer, S., Wang, A., Zeng, W., Chen, L., Wagner, J., Herman, G. (2007). Effect of a single cyclosporine dose on the single-dose pharmacokinetics of sitagliptin (MK-0431)., a dipeptidyl peptidase-4 inhibitor, in healthy male subjects. *J. Clin. Pharmacol.*, *47*, 165–174.

Kruijtzer, C.M., Beijnen, J.H., Rosing, H., ten Bokkel Huinink, W.W., Schot, M., Jewell, R.C., Paul, E.M., Schellens, J.H. (2002). Increased oral bioavailability of topotecan in combination with the breast cancer resistance protein and P-glycoprotein inhibitor GF120918. *J. Clin. Oncol.*, *20*, 2943–2950.

Kuppens, I.E., Bosch, T.M., van Maanen, M.J., Rosing, H., Fitzpatrick, A., Beijnen, J.H., Schellens, J.H. (2005). Oral bioavailability of docetaxel in combination with OC144-093 (ONT-093). *Cancer Chemother. Pharmacol.*, *55*, 72–78.

Kurata, Y., Ieiri, I., Kimura, M., Morita, T., Irie, S., Urae, A., Ohdo, S., Ohtani, H., Sawada, Y., Higuchi, S., Otsubo, K. (2002). Role of human MDR1 gene polymorphism in bioavailability and interaction of digoxin, a substrate of P-glycoprotein. *Clin. Pharmacol. Ther.*, *72*, 209–219.

Kurnik, D., Wood, A.J., Wilkinson, G.R. (2006). The erythromycin breath test reflects P-glycoprotein function independently of cytochrome P450 3A activity. *Clin. Pharmacol. Ther.*, *80*, 228–234.

Kyrklund, C., Backman, J.T., Kivistö, K.T., Neuvonen, M., Laitila, J., Neuvonen, P.J. (2001). Plasma concentrations of active lovastatin acid are markedly increased by gemfibrozil but not by bezafibrate. *Clin. Pharmacol. Ther.*, *69*, 340–345.

Larsen, U.L., Hyldahl Olesen, L., Guldborg Nyvold, C., Eriksen, J., Jakobsen, P., Østergaard, M., Autrup, H., Andersen, V. (2007) Human intestinal P-glycoprotein activity estimated by the model substrate digoxin. *Scand. J. Clin. Lab. Invest.*, *67*, 123–134.

Lê, L.H., Moore, M.J., Siu, L.L., Oza, A.M., MacLean, M., Fisher, B., Chaudhary, A., de Alwis, D. P., Slapak, C., Seymour, L. (2005). Phase I study of the multidrug resistance inhibitor zosuquidar administered in combination with vinorelbine in patients with advanced solid tumours. *Cancer Chemother. Pharmacol.*, *56*, 154–160.

Lee, F.Y., Lee, H., Hubbert, M.L., Edwards, P.A., Zhang, Y. (2006). FXR, a multipurpose nuclear receptor. *Trends. Biochem. Sci.*, *31*, 572–580.

Lee, W., Glaeser, H., Smith, L.H., Roberts, R.L., Moeckel, G.W., Gervasini, G., Leake, B.F., Kim, R.B. (2005). Polymorphisms in human organic anion-transporting polypeptide 1A2 (OATP1A2): implications for altered drug disposition and central nervous system drug entry. *J. Biol. Chem.*, *280*, 9610–9617.

Lemahieu, W.P., Hermann, M., Asberg, A., Verbeke, K., Holdaas, H., Vanrenterghem, Y., Maes, B.D. (2005). Combined therapy with atorvastatin and calcineurin inhibitors: no interactions with tacrolimus. *Am. J. Transplant.*, *5*, 2236–2243.

Lemma, G.L., Wang, Z., Hamman, M.A., Zaheer, N.A., Gorski, J.C., Hall, S.D. (2006). The effect of short- and long-term administration of verapamil on the disposition of cytochrome P450 3A and P-glycoprotein substrates. *Clin. Pharmacol. Ther.*, *79*, 218–230.

Leung, L.Y., Lim, H.K., Abell, M.W., Zimmerman, J.J. (2006). Pharmacokinetics and metabolic disposition of sirolimus in healthy male volunteers after a single oral dose. *Ther. Drug Monit.*, *28*, 51–61.

Lilja, J.J., Backman, J.T., Laitila, J., Luurila, H., Neuvonen, P.J. (2003). Itraconazole increases but grapefruit juice greatly decreases plasma concentrations of celiprolol. *Clin. Pharmacol. Ther.*, *73*, 192–198.

Lilja, J.J., Juntti-Patinen, L., Neuvonen, P.J. (2004). Orange juice substantially reduces the bioavailability of the beta-adrenergic-blocking agent celiprolol. *Clin. Pharmacol. Ther.*, *75*, 184–190.

Lilja, J.J., Niemi, M., Neuvonen, P.J. (2004). Rifampicin reduces plasma concentrations of celiprolol. *Eur. J. Clin. Pharmacol.*, *59*, 819–824.

Lilja, J.J., Niemi, M., Fredrikson, H., Neuvonen, P.J. (2007). Effects of clarithromycin and grapefruit juice on the pharmacokinetics of glibenclamide. *Br. J. Clin. Pharmacol.*, *63*, 732–740.

Lim, M.L., Min, S.S., Eron, J.J., Bertz, R.J., Robinson, M., Gaedigk, A., Kashuba, A.D. (2004). Coadministration of lopinavir/ritonavir and phenytoin results in two-way drug interaction through cytochrome P-450 induction. *J. Acquir. Immune. Defic. Syndr.*, *36*, 1034–1040.

Lin, J.H. (1995). Species similarities and differences in pharmacokinetics. *Drug Metab. Dispos.*, *23*, 1008–1021.

Lin, J.H. (2007). Transporter-mediated drug interactions: clinical implications and in vitro assessment. *Expert Opin. Drug Metab. Toxicol.*, *3*, 81–92.

Lin, J.H., Yamazaki, M. (2003). Role of P-glycoprotein in pharmacokinetics: clinical implications. *Clin. Pharmacokinet.*, *42*, 59–98.

Lum, B.L., Kaubisch, S., Fisher, G.A., Brown, B.W., Sikic, B.I. (2000). Effect of high-dose cyclosporine on etoposide pharmacodynamics in a trial to reverse P-glycoprotein (MDR1 gene) mediated drug resistance. *Cancer Chemother. Pharmacol.*, *5*, 305–311.

Lum, B.L., Kaubisch, S., Yahanda, A.M., Adler, K.M., Jew, L., Ehsan, M.N., Brophy, N.A., Halsey, J., Gosland, M.P., Sikic, B.I. (1992). Alteration of etoposide pharmacokinetics and pharmacodynamics by cyclosporine in a phase I trial to modulate multidrug resistance. *J. Clin. Oncol.*, *10*, 1635–1642.

McCance-Katz, E.F., Moody, D.E., Smith, P.F., Morse, G.D., Friedland, G., Pade, P., Baker, J., Alvanzo, A., Jatlow, P., Rainey, P.M. (2006). Interactions between buprenorphine and antiretrovirals. II. The protease inhibitors nelfinavir, lopinavir/ritonavir, and ritonavir. *Clin. Infect. Dis.*, *43* (Suppl 4), S235–S246.

Mahar Doan, K.M., Humphreys, J.E., Webster, L.O., Wring, S.A., Shampine, L.J., Serabjit-Singh, C.J., Adkison, K.K., Polli, J.W. (2002). Passive permeability and P-glycoprotein-mediated efflux differentiate central nervous system (CNS) and non-CNS marketed drugs. *J. Pharmacol. Exp. Ther.*, *303*, 1029–1037.

Mai, I., Krüger, H., Budde, K., Johne, A., Brockmöller, J., Neumayer, H.H., Roots, I. (2000). Hazardous pharmacokinetic interaction of Saint John's wort (Hypericum perforatum) with the immunosuppressant cyclosporin. *Int. J. Clin. Pharmacol. Ther.*, *38*, 500–502.

Mai, I., Störmer, E., Bauer, S., Krüger, H., Budde, K., Roots, I. (2003). Impact of St John's wort treatment on the pharmacokinetics of tacrolimus and mycophenolic acid in renal transplant patients. *Nephrol. Dial. Transplant.*, *18*, 819–822.

Malingré, M.M., Richel, D.J., Beijnen, J.H., Rosing, H., Koopman, F.J., Ten Bokkel Huinink, W.W., Schot, M.E., Schellens, J.H. (2001a). Coadministration of cyclosporine strongly enhances the oral bioavailability of docetaxel. *J. Clin. Oncol.*, *19*, 1160–1166.

Malingré, M.M., Ten Bokkel Huinink, W.W., Mackay, M., Schellens, J.H., Beijnen, J.H. (2001b). Pharmacokinetics of oral cyclosporin A when co-administered to enhance the absorption of orally administered docetaxel. *Eur. J. Clin. Pharmacol.*, *57*, 305–307.

Malingré, M.M., Beijnen, J.H., Rosing, H., Koopman, F.J., Jewell, R.C., Paul, E.M., Ten Bokkel Huinink, W.W., Schellens, J.H. (2001c). Co-administration of GF120918 significantly increases the systemic exposure to oral paclitaxel in cancer patients. *Br. J. Cancer*, *84*, 42–47.

Martin-Facklam, M., Burhenne, J., Ding, R., Fricker, R., Mikus, G., Walter-Sack, I., Haefeli, W.E. (2002). Dose-dependent increase of saquinavir bioavailability by the pharmaceutic aid cremophor EL. *Br. J. Clin. Pharmacol.*, *53*, 576–581.

Marzolini, C., Paus, E., Buclin, T., Kim, R.B. (2004). Polymorphisms in human MDR1 (P-glycoprotein): recent advances and clinical relevance. *Clin. Pharmacol. Ther.*, *75*, 13–33.

Masuda, M., I'Izuka, Y., Yamazaki, M., Nishigaki, R., Kato, Y., Ni'inuma, K., Suzuki, H., Sugiyama, Y. (1997). Methotrexate is excreted into the bile by canalicular multispecific organic anion transporter in rats. *Cancer Res.*, *57*, 3506–3510.

Matheny, C.J., Lamb, M.W., Brouwer, K.R., Pollack, G.M. (2001). Pharmacokinetic and pharmacodynamic implications of P-glycoprotein modulation. *Pharmacotherapy*, *21*, 778–796.

Mayo, P.R., Skeith, K., Russell, A.S., Jamali, F. (2000). Decreased dromotropic response to verapamil despite pronounced increased drug concentration in rheumatoid arthritis. *Br. J. Clin. Pharmacol.*, *50*, 605–613.

Meerum Terwogt, J.M., Beijnen, J.H., ten Bokkel Huinink, W.W., Rosing, H., Schellens, J.H. (1998). Co-administration of cyclosporin enables oral therapy with paclitaxel. *Lancet*, *352* (9124), 285.

Meerum Terwogt, J.M., Malingré, M.M., Beijnen, J.H., ten Bokkel Huinink, W.W., Rosing, H., Koopman, F.J., van Tellingen, O., Swart, M., Schellens, J.H. (1999). Coadministration of oral cyclosporin A enables oral therapy with paclitaxel. *Clin. Cancer Res.*, *5*, 3379–3384.

Miki, Y., Suzuki, T., Kitada, K., Yabuki, N., Shibuya, R., Moriya, T., Ishida, T., Ohuchi, N., Blumberg, B., Sasano, H. (2006). Expression of the steroid and xenobiotic receptor and its possible target gene, organic anion transporting polypeptide-A, in human breast carcinoma. *Cancer Res.*, *66*, 535–542.

Minami, H., Ohtsu, T., Fujii, H., Igarashi, T., Itoh, K., Uchiyama-Kokubu, N., Aizawa, T., Watanabe, T., Uda, Y., Tanigawara, Y., Sasaki, Y. (2001). Phase I study of intravenous PSC-833 and doxorubicin: reversal of multidrug resistance. *Jpn. J. Cancer Res.*, *92*, 220–230.

Mizuno, N., Niwa, T., Yotsumoto, Y., Sugiyama, Y. (2003). Impact of drug transporter studies on drug discovery and development. *Pharmacol. Rev.*, *55*, 425–461.

Moreira, A., Lobato, R., Morais, J., Silva, S., Ribeiro, J., Figueira, A., Vale, D., Sousa, C., Araújo, F., Fernandes, A., Oliveira, J., Passos-Coelho, J.L. (2001). Influence of the interval between the administration of doxorubicin and paclitaxel on the pharmacokinetics of these drugs in patients with locally advanced breast cancer. *Cancer Chemother. Pharmacol.*, 48, 333–337.

Mouly, S., Lown, K.S., Kornhauser, D., Joseph, J.L., Fiske, W.D., Benedek, I.H., Watkins, P.B. (2002). Hepatic but not intestinal CYP3A4 displays dose-dependent induction by efavirenz in humans. *Clin. Pharmacol. Ther.*, *72*, 1–9.

Mück, W., Mai, I., Fritsche, L., Ochmann, K., Rohde, G., Unger, S., Johne, A., Bauer, S., Budde, K., Roots, I., Neumayer, H.H., Kuhlmann, J. (1999). Increase in cerivastatin systemic exposure after single and multiple dosing in cyclosporine-treated kidney transplant recipients. *Clin. Pharmacol. Ther.*, *65*, 251–261.

Mukwaya, G., MacGregor, T., Hoelscher, D., Heming, T., Legg, D., Kavanaugh, K., Johnson, P., Sabo, J.P., McCallister, S. (2005). Interaction of ritonavir-boosted tipranavir with loperamide does not result in loperamide-associated neurologic side effects in healthy volunteers. *Antimicrob. Agents. Chemother.*, *49*, 4903–4910.

Nakagami, T., Yasui-Furukori, N., Saito, M., Tateishi, T., Kaneo, S. (2005). Effect of verapamil on pharmacokinetics and pharmacodynamics of risperidone: in vivo evidence of involvement of P-glycoprotein in risperidone disposition. *Clin. Pharmacol. Ther.*, *78*, 43–51.

Nannan Panday, V.R., Hoetelmans, R.M., van Heeswijk, R.P., Meenhorst, P.L., Inghels, M., Mulder, J.W., Beijnen, J.H. (1999). Paclitaxel in the treatment of human immunodeficiency virus 1-associated Kaposi's sarcoma—drug–drug interactions with protease inhibitors and a nonnucleoside reverse transcriptase inhibitor: a case report study. *Cancer Chemother. Pharmacol.*, *43*, 516–519.

Nenciu, L.M., Laberge, P., Thirion, D.J. (2006). Telithromycin-induced digoxin toxicity and electrocardiographic changes. *Pharmacotherapy*, *26*, 872–876.

Niemi, M., Tornio, A., Pasanen, M.K., Fredrikson, H., Neuvonen, P.J., Backman, J.T. (2006). Eur Itraconazole, gemfibrozil and their combination markedly raise the plasma concentrations of loperamide. *J. Clin. Pharmacol.*, *62*, 463–472.

Oswald, S., Giessmann, T., Luetjohann, D., Wegner, D., Rosskopf, D., Weitschies, W., Siegmund, W. (2006). Disposition and sterol-lowering effect of ezetimibe are influenced by single-dose coadministration of rifampin, an inhibitor of multidrug transport proteins. *Clin. Pharmacol. Ther.*, *80*, 477–485.

Oswald, S., Haenisch, S., Fricke, C., Sudhop, T., Remmler, C., Giessmann, T., Jedlitschky, G., Adam, U., Dazert, E., Warzok, R., Wacke, W., Cascorbi, I., Kroemer, H.K., Weitschies, W., von Bergmann, K., Siegmund, W. (2006). Intestinal expression of P-glycoprotein (ABCB1), multidrug resistance associated protein 2 (ABCC2), and uridine diphosphate-glucuronosyltransferase

1A1 predicts the disposition and modulates the effects of the cholesterol absorption inhibitor ezetimibe in humans. *Clin. Pharmacol. Ther.*, *79*, 206–217.

Overbosch, D., Van Gulpen, C., Hermans, J., Mattie, H. (1988). The effect of probenecid on the renal tubular excretion of benzylpenicillin. *Br. J. Clin. Pharmacol.*, *25*, 51–58.

Paine, M.F., Wagner, D.A., Hoffmaster, K.A., Watkins, P.B. (2002a). Cytochrome P450 3A4 and P-glycoprotein mediate the interaction between an oral erythromycin breath test and rifampin. *Clin. Pharmacol. Ther.*, *72*, 524–535.

Paine, M.F., Leung, L.Y., Lim, H.K., Liao, K., Oganesian, A., Zhang, M.Y., Thummel, K.E., Watkins, P.B. (2002b). Identification of a novel route of extraction of sirolimus in human small intestine: roles of metabolism and secretion. *J. Pharmacol. Exp. Ther.*, *301*, 174–186.

Partanen, J., Jalava, K.M., Neuvonen, P.J. (1996). Itraconazole increases serum digoxin concentration. *Pharmacol. Toxicol.*, *79*, 274–276.

Patnaik, A., Warner, E., Michael, M., Egorin, M.J., Moore, M.J., Siu, L.L., Fracasso, P.M., Rivkin, S., Kerr, I., Litchman, M., Oza, A.M. (2000). Phase I dose-finding and pharmacokinetic study of paclitaxel and carboplatin with oral valspodar in patients with advanced solid tumors. *J. Clin. Oncol.*, *18*, 3677–3689.

Paulusma, C.C., Kool, M., Bosma, P.J., Scheffer, G.L., ter Borg, F., Scheper, R.J., Tytgat, G.N., Borst, P., Baas, F., Oude Elferink, R.P. (1997). A mutation in the human canalicular multispecific organic anion transporter gene causes the Dubin-Johnson syndrome. *Hepatology*, *25*, 1539–1542.

Pea, F., Damiani, D., Michieli, M., Ermacora, A., Baraldo, M., Russo, D., Fanin, R., Baccarani, M., Furlanut, M. (1999). Multidrug resistance modulation in vivo: the effect of cyclosporin A alone or with dexverapamil on idarubicin pharmacokinetics in acute leukemia. *Eur. J. Clin. Pharmacol.*, *55*, 361–368.

Penzak, S.R., Shen, J.M., Alfaro, R.M., Remaley, A.T., Natarajan, V., Falloon, J. (2004). Ritonavir decreases the nonrenal clearance of digoxin in healthy volunteers with known MDR1 genotypes. *Ther. Drug Monit.*, *26*, 322–330.

Penzak, S.R., Formentini, E., Alfaro, R.M., Long, M., Natarajan, V., Kovacs, J. (2005). Prednisolone pharmacokinetics in the presence and absence of ritonavir after oral prednisone administration to healthy volunteers. *J. Acquir. Immune. Defic. Syndr.*, *40*, 573–580.

Petri, N., Borga, O., Nyberg, L., Hedeland, M., Bondesson, U., Lennernas, H. (2006). Effect of erythromycin on the absorption of fexofenadine in the jejunum, ileum and colon determined using local intubation in healthy volunteers. *Int. J. Clin. Pharmacol. Ther.*, *44*, 71–79.

Petrovic, V., Teng, S., Piquette-Miller, M. (2007). Regulation of drug transporters during infection and inflammation. *Mol. Interv.*, *7*, 99–111.

Pheterson, A.D., Miller, L., Fox, C.F., Estroff, T.W., Sweeney, D.R. (1987). Multifocal neurological impairment caused by infection-induced rise in blood lithium and amitriptyline. *Int. J. Psychiatry Med. 1986.*, *16*, 257–262.

Pizzagalli, F., Hagenbuch, B., Stieger, B., Klenk, U., Folkers, G., Meier, P.J. (2002). Identification of a novel human organic anion transporting polypeptide as a high affinity thyroxine transporter. *Mol. Endocrinol.*, *16*, 2283–2296.

Planting, A.S., Sonneveld, P., van der Gaast, A., Sparreboom, A., van der Burg, M.E., Luyten, G. P., de Leeuw, K., de Boer-Dennert, M., Wissel, P.S., Jewell, R.C., Paul, E.M., Purvis, N.B.Jr, , Verweij, J. (2005). A phase I and pharmacologic study of the MDR converter GF120918 in combination with doxorubicin in patients with advanced solid tumors. *Cancer Chemother. Pharmacol.*, *55*, 91–99.

Polli, J.W., Wring, S.A., Humphreys, J.E., Huang, L., Morgan, J.B., Webster, L.O., Serabjit-Singh, C.S. (2001). Rational use of in vitro P-glycoprotein assays in drug discovery. *J. Pharmacol. Exp. Ther.*, *299*, 620–628.

Potter, J.M., McWhinney, B.C., Sampson, L., Hickman, P.E. (2004). Area-under-the-curve monitoring of prednisolone for dose optimization in a stable renal transplant population. *Ther. Drug Monit.*, *26*, 408–414.

Putnam, W.S., Woo, J.M., Huang, Y., Benet, L.Z. (2005). Effect of the MDR1 C3435T variant and P-glycoprotein induction on dicloxacillin pharmacokinetics. *J. Clin. Pharmacol.*, *45*, 411–421.

Race, J.E., Grassl, S.M., Williams, W.J., Holtzman, E.J. (1999). Molecular cloning and characterization of two novel human renal organic anion transporters (hOAT1 and hOAT3) *Biochem. Biophys. Res. Commun.*, *255*, 508–514.

Ramachandra, M., Ambudkar, S.V., Gottesman, M.M., Pastan, I., Hrycyna, C.A. (1996). Functional characterization of a glycine 185-to-valine substitution in human P-glycoprotein by using a vaccinia-based transient expression system. *Mol. Biol. Cell*, *7*, 1485–1498.

Reitman, M.L., Schadt, E.E. (2007). Pharmacogenetics of metformin response: a step in the path toward personalized medicine. *J. Clin. Invest.*, *117*, 1226–1229.

Rengelshausen, J., Göggelmann, C., Burhenne, J., Riedel, K.D., Ludwig, J., Weiss, J., Mikus, G., Walter-Sack, I., Haefeli, W.E. (2003). Contribution of increased oral bioavailability and reduced nonglomerular renal clearance of digoxin to the digoxin-clarithromycin interaction. *Br. J. Clin. Pharmacol.*, *56*, 32–38.

Roberts, D.H., Kendall, M.J., Jack, D.B., Welling, P.G. (1981). Pharmacokinetics of cephradine given intravenously with and without probenecid. *Br. J. Clin. Pharmacol.*, *11*, 561–564.

Rocchi, E., Khodjakov, A., Volk, E.L., Yang, C.H., Litman, T., Bates, S.E., Schneider, E. (2000). The product of the ABC half-transporter gene ABCG2 (BCRP/MXR/ABCP) is expressed in the plasma membrane. *Biochem. Biophys. Res. Commun.*, *271*, 42–46.

Rollot, F., Pajot, O., Chauvelot-Moachon, L., Nazal, E.M., Kélaïdi, C., Blanche, P. (2004). Acute colchicine intoxication during clarithromycin administration. *Ann. Pharmacother.*, *38*, 2074–2077.

Rowinsky, E.K., Smith, L., Wang, Y.M., Chaturvedi, P., Villalona, M., Campbell, E., Aylesworth, C., Eckhardt, S.G., Hammond, L., Kraynak, M., Drengler, R., Stephenson, J. Jr, Harding, M.W., Von Hoff, D.D. (1998). Phase I and pharmacokinetic study of paclitaxel in combination with biricodar, a novel agent that reverses multidrug resistance conferred by overexpression of both MDR1 and MRP. *J. Clin. Oncol.*, *16*, 2964–2976.

Ruschitzka, F., Meier, P.J., Turina, M., Lüscher, T.F., Noll, G. (2000). Acute heart transplant rejection due to Saint John's wort. *Lancet*, *355* (9203), 548–549.

Rushing, D.A., Raber, S.R., Rodvold, K.A., Piscitelli, S.C., Plank, G.S., Tewksbury, D.A. (1994). The effects of cyclosporine on the pharmacokinetics of doxorubicin in patients with small cell lung cancer. *Cancer*, *74*, 834–841.

Sadeque, A.J., Wandel, C., He, H., Shah, S., Wood, A.J. (2000). Increased drug delivery to the brain by P-glycoprotein inhibition. *Clin. Pharmacol. Ther.*, *68*, 231–237.

Sahi, J., Milad, M.A., Zheng, X., Rose, K.A., Wang, H., Stilgenbauer, L., Gilbert, D., Jolley, S., Stern, R.H., LeCluyse, E.L. (2003). Avasimibe induces CYP3A4 and multiple drug resistance protein 1 gene expression through activation of the pregnane X receptor. *J. Pharmacol. Exp. Ther.*, *306*, 1027–1034.

Sai, Y. (2005). Biochemical and molecular pharmacological aspects of transporters as determinants of drug disposition. *Drug Metab. Pharmacokinet.*, *20*, 91–99.

Sandler, A., Gordon, M., De Alwis, D.P., Pouliquen, I., Green, L., Marder, P., Chaudhary, A., Fife, K., Battiato, L., Sweeney, C., Jordan, C., Burgess, M., Slapak, C.A.A. (2004). Phase I trial of a potent P-glycoprotein inhibitor, zosuquidar trihydrochloride (LY335979)., administered intravenously in combination with doxorubicin in patients with advanced malignancy. *Clin. Cancer Res.*, *10*, 3265–3272.

Sansone-Parsons, A., Krishna, G., Martinho, M., Kantesaria, B., Gelone, S., Mant, T.G. (2007). Effect of oral posaconazole on the pharmacokinetics of cyclosporine and tacrolimus. *Pharmacotherapy*, *27*, 825–834.

Sasaki, M., Suzuki, H., Ito, K., Abe, T., Sugiyama, Y. (2002). Transcellular transport of organic anions across a double-transfected Madin-Darby canine kidney II cell monolayer expressing both human organic anion-transporting polypeptide (OATP2/SLC21A6) and multidrug resistance-associated protein 2 (MRP2/ABCC2). *J. Biol. Chem.*, *277*, 6497–6503.

Sasaki, M., Suzuki, H., Aoki, J., Ito, K., Meier, P.J., Sugiyama, Y. (2004). Prediction of in vivo biliary clearance from the in vitro transcellular transport of organic anions across a double-transfected Madin-Darby canine kidney II monolayer expressing both rat organic anion transporting polypeptide 4 and multidrug resistance associated protein 2. *Mol. Pharmacol.*, *66*, 450–459.

Sasongko, L., Link, J.M., Muzi, M., Mankoff, D.A., Yang, X., Collier, A.C., Shoner, S.C., and Unadkat, J.D. (2005). Imaging P-glycoprotein transport activity at the human blood-brain barrier with positron emission tomography. *Clin. Pharmacol. Ther.*, *77*, 503–514.

Schinkel, A.H., Wagenaar, E., Mol, C.A., van Deemter, L. (1996). P-glycoprotein in the blood-brain barrier of mice influences the brain penetration and pharmacological activity of many drugs. *J. Clin. Invest.*, *97*, 2517–2524.

Schwarz, U.I., Gramatté, T., Krappweis, J., Berndt, A., Oertel, R., von Richter, O., Kirch, W. (1999). Unexpected effect of verapamil on oral bioavailability of the beta-blocker talinolol in humans. *Clin. Pharmacol. Ther.*, *65*, 283–290.

Schwarz, U.I., Gramatté, T., Krappweis, J., Oertel, R., Kirch, W. (2000). P-glycoprotein inhibitor erythromycin increases oral bioavailability of talinolol in humans. *Int. J. Clin. Pharmacol. Ther.*, *38*, 161–167.

Schwarz, U.I., Seemann, D., Oertel, R., Miehlke, S., Kuhlisch, E., Fromm, M.F., Kim, R.B., Bailey, D.G., Kirch, W. (2005). Grapefruit juice ingestion significantly reduces talinolol bioavailability. *Clin. Pharmacol. Ther.*, *77*, 291–301.

Schwarz, U.I., Hanso, H., Oertel, R., Miehlke, S., Kuhlisch, E., Glaeser, H., Hitzl, M., Dresser, G. K., Kim, R.B., Kirch, W. (2007). Induction of intestinal P-glycoprotein by St John's wort reduces the oral bioavailability of talinolol. *Clin. Pharmacol. Ther.*, *81*, 669–768.

Seelig, A., Gottschlich, R., Devant, R.M. (1994). A method to determine the ability of drugs to diffuse through the blood-brain barrier. *Proc. Natl. Acad. Sci. U. S. A.*, *91*, 68–72.

Seiden, M.V., Swenerton, K.D., Matulonis, U., Campos, S., Rose, P., Batist, G., Ette, E., Garg, V., Fuller, A., Harding, M.W., Charpentier, D. (2002). A phase II study of the MDR inhibitor biricodar (INCEL, VX-710) and paclitaxel in women with advanced ovarian cancer refractory to paclitaxel therapy. *Gynecol. Oncol.*, *86*, 302–310.

Seitz, S., Boelsterli, U.A. (1998). Diclofenac acyl glucuronide, a major biliary metabolite, is directly involved in small intestinal injury in rats. *Gastroenterology*, *115*, 1476–1482.

Shilling, A.D., Azam, F., Kao, J., Leung, L. (2006). Use of canalicular membrane vesicles (CMVs) from rats, dogs, monkeys and humans to assess drug transport across the canalicular membrane. *J. Pharmacol. Toxicol. Methods*, *53*, 186–197.

Shimada, T., Terada, A., Yokogawa, K., Kaneko, H., Nomura, M., Kaji, K., Kaneko, S., Kobayashi, K., Miyamoto, K. (2002). Lowered blood concentration of tacrolimus and its recovery with changes in expression of CYP3A and P-glycoprotein after high-dose steroid therapy. *Transplantation*, *74*, 1419–1424.

Shimizu, M., Uno, T., Sugawara, K., Tateishi, T. (2006). Effects of itraconazole and diltiazem on the pharmacokinetics of fexofenadine, a substrate of P-glycoprotein. *Br. J. Clin. Pharmacol.*, *61*, 538–544.

Shon, J.H., Yoon, Y.R., Hong, W.S., Nguyen, P.M., Lee, S.S., Choi, Y.G., Cha, I.J., Shin, J.G. (2005). Effect of itraconazole on the pharmacokinetics and pharmacodynamics of fexofenadine in relation to the MDR1 genetic polymorphism. *Clin. Pharmacol. Ther.*, *78*, 191–201.

Shu, Y., Sheardown, S.A., Brown, C., Owen, R.P., Zhang, S., Castro, R.A., Ianculescu, A.G., Yue, L., Burchard, E.G., Brett, C.M., Giacomini, K.M. (2007). Effect of genetic variation in the organic cation transporter 1 (OCT1) on metformin action. *J. Clin. Invest.*, *117*, 1422–1431.

Siedlik, P.H., Olson, S.C., Yang, B.B., Stern, R.H. (1999). Erythromycin coadministration increases plasma atorvastatin concentrations. *J. Clin. Pharmacol.*, *39*, 501–504.

Skarke, C., Jarrar, M., Schmidt, H., Kauert, G., Langer, M., Geisslinger, G., Lötsch, J. (2003). Effects of ABCB1 (multidrug resistance transporter) gene mutations on disposition and central nervous effects of loperamide in healthy volunteers. *Pharmacogenetics*, *13*, 651–660.

Smeets, M., Raymakers, R., Muus, P., Vierwinden, G., Linssen, P., Masereeuw, R., de Witte, T. (2001). Cyclosporin increases cellular idarubicin and idarubicinol concentrations in relapsed or refractory AML mainly due to reduced systemic clearance. *Leukemia*, *15*, 80–88.

Somogyi, A., McLean, A., Heinzow, B. (1983). Cimetidine-procainamide pharmacokinetic interaction in man: evidence of competition for tubular secretion of basic drugs. *Eur. J. Clin. Pharmacol.*, *25*, 339–345.

Somogyi, A., Stockley, C., Keal, J., Rolan, P., Bochner, F. (1987). Reduction of metformin renal tubular secretion by cimetidine in man. *Br. J. Clin. Pharmacol.*, *23*, 545–551.

Sonneveld, P., Marie, J.P., Huisman, C., Vekhoff, A., Schoester, M., Faussat, A.M., van Kapel, J., Groenewegen, A., Charnick, S., Zittoun, R., Löwenberg, B. (1996). Reversal of multidrug resistance by SDZ PSC 833, combined with VAD (vincristine, doxorubicin, dexamethasone) in refractory multiple myeloma. A phase I study. *Leukemia*, *10*, 1741–1750.

Stieger, B., Fattinger, K., Madon, J., Kullak-Ublick, G.A., Meier, P.J. (2000). Drug- and estrogen-induced cholestasis through inhibition of the hepatocellular bile salt export pump (Bsep) of rat liver. *Gastroenterology*, *118*, 422–430.

Sugimoto, K., Ohmori, M., Tsuruoka, S., Nishiki, K., Kawaguchi, A., Harada, K., Arakawa, M., Sakamoto, K., Masada, M., Miyamori, I., Fujimura, A. (2001). Different effects of St John's wort on the pharmacokinetics of simvastatin and pravastatin. *Clin. Pharmacol. Ther.*, *70*, 518–524.

Suzuki, H., Sugiyama, Y. (2002). Single nucleotide polymorphisms in multidrug resistance associated protein 2 (MRP2/ABCC2): its impact on drug disposition. *Adv. Drug Deliv. Rev.*, *54*, 1311–1331.

Szatmari, I., Vámosi, G., Brazda, P., Balint, B.L., Benko, S., Széles, L., Jeney, V., Ozvegy-Laczka, C., Szántó, A., Barta, E., Balla, J., Sarkadi, B., Nagy, L. (2006). Peroxisome proliferator-activated receptor gamma-regulated ABCG2 expression confers cytoprotection to human dendritic cells. *J. Biol. Chem.*, *281* (33), 23812–23823.

Taburet, A.M., Piketty, C., Chazallon, C., Vincent, I., Gérard, L., Calvez, V., Clavel, F., Aboulke, J.P., Girard, P.M. (2004). Interactions between atazanavir-ritonavir and tenofovir in heavily pretreated human immunodeficiency virus-infected patients. *Antimicrob. Agents. Chemother.*, *48*, 2091–2096.

Taguchi, Y., Yoshida, A., Takada, Y., Komano, T., Ueda, K. (1997). Anti-cancer drugs and glutathione stimulate vanadate-induced trapping of nucleotide in multidrug resistance-associated protein (MRP) *FEBS Lett.*, *401*, 11–14.

Tamai, I., Nezu, J., Uchino, H., Sai, Y., Oku, A., Shimane, M., Tsuji, A. (2000). Molecular identification and characterization of novel members of the human organic anion transporter (OATP) family. *Biochem. Biophys. Res. Commun.*, *273*, 251–260.

Tanaka, H., Matsumoto, K., Ueno, K., Kodama, M., Yoneda, K., Katayama, Y., Miyatake, K. (2003). Effect of clarithromycin on steady-state digoxin concentrations. *Ann. Pharmacother.*, *37*, 178–181.

Tannergren, C., Petri, N., Knutson, L., Hedeland, M., Bondesson, U., Lennernäs, H. (2003). Multiple transport mechanisms involved in the intestinal absorption and first-pass extraction of fexofenadine. *Clin. Pharmacol. Ther.*, *74*, 423–436.

Tannergren, C., Engman, H., Knutson, L., Hedeland, M., Bondesson, U., St, Lennernäs, H. (2004). John's wort decreases the bioavailability of R- and S-verapamil through induction of the first-pass metabolism. *Clin. Pharmacol. Ther.*, *75*, 298–309.

Tayrouz, Y., Ding, R., Burhenne, J., Riedel, K.D., Weiss, J., Hoppe-Tichy, T., Haefeli, W.E., Mikus, G. (2003). Pharmacokinetic and pharmaceutic interaction between digoxin and Cremophor RH40. *Clin. Pharmacol. Ther.*, *73*, 397–405.

Tayrouz, Y., Ganssmann, B., Ding, R., Klingmann, A., Aderjan, R., Burhenne, J., Haefeli, W.E., Mikus, G. (2001). Ritonavir increases loperamide plasma concentrations without evidence for P-glycoprotein involvement. *Clin. Pharmacol. Ther.*, *70*, 405–414.

Tirona, R.G., Kim, R.B. (2002). Pharmacogenomics of organic anion-transporting polypeptides (OATP). *Adv. Drug Deliv. Rev.*, *54*, 1343–1352.

Tsuji, A., Terasaki, T., Tamai, I., Takeda, K. (1990). In vivo evidence for carrier-mediated uptake of beta-lactam antibiotics through organic anion transport systems in rat kidney and liver. *J. Pharmacol. Exp. Ther.*, *253*, 315–320.

Tsunoda, S.M., Harris, R.Z., Christians, U., Velez, R.L., Freeman, R.B., Benet, L.Z., Warshaw, A. (2001). Red wine decreases cyclosporine bioavailability. *Clin. Pharmacol. Ther.*, *70*, 462–467.

Tsuruoka, S., Ioka, T., Wakaumi, M., Sakamoto, K., Ookami, H., Fujimura, A. (2006). Severe arrhythmia as a result of the interaction of cetirizine and pilsicainide in a patient with renal insufficiency: first case presentation showing competition for excretion via renal multidrug resistance protein 1 and organic cation transporter 2. *Clin. Pharmacol. Ther.*, *79*, 389–396.

Tuteja, S., Alloway, R.R., Johnson, J.A., Gaber, A.O. (2001). The effect of gut metabolism on tacrolimus bioavailability in renal transplant recipients. *Transplantation*, *71*, 1303–1307.

Uno, T., Shimizu, M., Sugawara, K., Tateishi, T. (2006). Lack of dose-dependent effects of itraconazole on the pharmacokinetic interaction with fexofenadine. *Drug Metab. Dispos.*, *34*, 1875–1879.

Urquhart, B.L., Tirona, R.G., Kim, R.B. (2007). Nuclear receptors and the regulation of drug-metabolizing enzymes and drug transporters: implications for interindividual variability in response to drugs. *J. Clin. Pharmacol.*, *47*, 566–578.

Van Heeswijk, R.P., Bourbeau, M., Campbell, P., Seguin, I., Chauhan, B.M., Foster, B.C., Cameron, D.W. (2006). Time-dependent interaction between lopinavir/ritonavir and fexofenadine. *J. Clin. Pharmacol.*, *46*, 758–767.

Van Montfoort, J.E., Hagenbuch, B., Groothuis, G.M., Koepsell, H., Meier, P.J., Meijer, D.K. (2003). Drug uptake systems in liver and kidney. *Curr. Drug Metab.*, *4*, 185–211.

Vasquez, E.M., Shin, G.P., Sifontis, N., Benedetti, E. (2005). Concomitant clotrimazole therapy more than doubles the relative oral bioavailability of tacrolimus. *Ther. Drug Monit.*, *27*, 587–591.

Von Hentig, N., Müller, A., Rottmann, C., Wolf, T., Lutz, T., Klauke, S., Kurowski, M., Oertel, B., Dauer, B., Harder, S., Staszewski, S. (2007). Pharmacokinetics of saquinavir, atazanavir, and ritonavir in a twice-daily boosted double-protease inhibitor regimen. *Antimicrob. Agents. Chemother.*, *51*, 1431–1439.

Wakasugi, H., Yano, I., Ito, T., Hashida, T., Futami, T., Nohara, R., Sasayama, S., Inui, K. (1998). Effect of clarithromycin on renal excretion of digoxin: interaction with P-glycoprotein. *Clin. Pharmacol. Ther.*, *64*, 123–128.

Wang, Z., Hamman, M.A., Huang, S.M., Lesko, L.J., Hall, S.D. (2002). Effect of St John's wort on the pharmacokinetics of fexofenadine. *Clin. Pharmacol. Ther.*, *71*, 414–420.

Westphal, K., Weinbrenner, A., Giessmann, T., Stuhr, M., Franke, G., Zschiesche, M., Oertel, R., Terhaag, B., Kroemer, H.K., Siegmund, W. (2000). Oral bioavailability of digoxin is enhanced by talinolol: evidence for involvement of intestinal P-glycoprotein. *Clin. Pharmacol. Ther.*, *68*, 6–12.

Westphal, K., Weinbrenner, A., Zschiesche, M., Franke, G., Knoke, M., Oertel, R., Fritz, P., von Richter, O., Warzok, R., Hachenberg, T., Kauffmann, H.M., Schrenk, D., Terhaag, B., Kroemer, H.K., Siegmund, W. (2000). Induction of P-glycoprotein by rifampin increases intestinal secretion of talinolol in human beings: a new type of drug/drug interaction. *Clin. Pharmacol. Ther.* *68*, 345–355.

Wilson, W.H., Jamis-Dow, C., Bryant, G., Balis, F.M., Klecker, R.W., Bates, S.E., Chabner, B.A., Steinberg, S.M., Kohler, D.R., Wittes, R.E. (1995). Phase I and pharmacokinetic study of the multidrug resistance modulator dexverapamil with EPOCH chemotherapy. *J. Clin. Oncol.*, *13*, 1985–1994.

Winston, A., Back, D., Fletcher, C., Robinson, L., Unsworth, J., Tolowinska, I., Schutz, M., Pozniak, A.L., Gazzard, B., Boffito, M. (2006). Effect of omeprazole on the pharmacokinetics of saquinavir-500 mg formulation with ritonavir in healthy male and female volunteers. *AIDS*, *20*, 1401–1406.

Wu, C.Y., Benet, L.Z. (2005). Predicting drug disposition via application of BCS: transport/absorption/elimination interplay and development of a biopharmaceutics drug disposition classification system. *Pharm. Res.*, *22*, 11–23.

Xia, C.Q., Milton, M.N., Gan, L.S. (2007). Evaluation of drug-transporter interactions using in vitro and in vivo models. *Curr. Drug Metabolism.*, *8*, 341–363.

Xie, R., Tan, L.H., Polasek, E.C., Hong, C., Teillol-Foo, M., Gordi, T., Sharma, A., Nickens, D.J., Arakawa, T., Knuth, D.W., Antal, E.J. (2005). CYP3A and P-glycoprotein activity induction with St. John's Wort in healthy volunteers from 6 ethnic populations. *J. Clin. Pharmacol.*, *45*, 352–356.

Yasui-Furukori, N., Uno, T., Sugawara, K., Tateishi, T. (2005). Different effects of three transporting inhibitors, verapamil, cimetidine, and probenecid, on fexofenadine pharmacokinetics. *Clin. Pharmacol. Ther.*, *77*, 17–23.

Yasui-Furukori, N., Saito, M., Niioka, T., Inoue, Y., Sato, Y., Kaneko, S. (2007). Effect of itraconazole on pharmacokinetics of paroxetine: the role of gut transporters. *Ther. Drug Monit.*, *29*, 45–48.

Yu, J., Lo, J.L., Huang, L., Zhao, A., Metzger, E., Adams, A., Meinke, P.T., Wright, S.D., Cui, J. (2002). Lithocholic acid decreases expression of bile salt export pump through farnesoid X receptor antagonist activity. *J. Biol. Chem.*, *277*, 31441–31447.

Zapater, P., Reus, S., Tello, A., Torrús, D., Pérez-Mateo, M., Horga, J.F. (2002). A prospective study of the clarithromycin-digoxin interaction in elderly patients. *J. Antimicrob. Chemother.*, *50*, 601–606.

Zhang, L., Dresser, M.J., Gray, A.T., Yost, S.C., Terashita, S., Giacomini, K.M. (1997). Cloning and functional expression of a human liver organic cation transporter. *Mol. Pharmacol.*, *51*, 913–921.

Zhang, L., Strong, J.M., Qiu, W., Lesko, L.J., Huang, S.M. (2006). Scientific perspectives on drug transporters and their role in drug interactions. *Mol. Pharm.*, *3*, 62–69.

Zhang, L., Zhang, Y.D., Strong, J.M., Reynolds, K.S., Huang, S.M. (2008). A regulatory viewpoint on transporter-based drug interactions. *Xenobiotica*, *38*, 709–724.

Zimmerman, J.J., Harper, D., Getsy, J., Jusko, W.J. (2003). Pharmacokinetic interactions between sirolimus and microemulsion cyclosporine when orally administered jointly and 4 hours apart in healthy volunteers. *J. Clin. Pharmacol.*, *43*, 1168–1176.

Zschiesche, M., Lemma, G.L., Klebingat, K.J., Franke, G., Terhaag, B., Hoffmann, A., Gramatté, T., Kroemer, H.K., Siegmund, W. (2002). Stereoselective disposition of talinolol in man. *J. Pharm. Sci.*, *91*, 303–311.

# Automated Drug Screening for ADMET Properties

MELVIN REICHMAN and HELEN GILL

## 1 INTRODUCTION

The pharmacological basis of therapeutics is comprised of two overarching scientific disciplines: pharmacodynamics and pharmacokinetics. It is often summarily stated: "pharmacodynamics is the study of what a drug does to the body; whereas pharmacokinetics is the study of what the body does to a drug." The latter studies the ***a***bsorption, ***d***istribution, ***m***etabolism, ***e***xcretion, and ***t***oxicology of drugs (ADMET). The survival and growth of the pharmaceutical industry relies on favorable, early clinical trial outcomes. Poor outcomes are commonly due to poor ADMET properties of new chemical entities (NCEs). The difficulty in optimizing ADMET properties of any NCE is universally recognized as an important factor responsible for the increasing cost of bringing a new drug to market. On the other hand, ADMET sciences in drug development can "rescue" drugs that otherwise would fail. While the list of marketed drug withdrawals continues to increase, it appears that significant progress has been made by earlier detection of ADMET issues through automated bioassays amenable to higher throughput screening paradigms. Higher throughput ADMET for drug lead qualification is deployed more commonly in the industry, as well as earlier rather than later preclinical development. There are data suggesting that the paradigm shift of deploying ADMET earlier in the R&D continuum has led to a marked decrease in drug failures. This article traces the development of automation in experimental ADMET to provide an overview of the assays that have been miniaturized to the microtiter plate format and automated on common or specialized laboratory workstations to enable higher throughput ADMET screening.

## 2 BACKGROUND

The term high-throughput screening (HTS) began regularly appearing in the scientific literature in the early 1990s (Harris *et al.*, 1991; Burch, 1993; and see Fig. 1). During

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

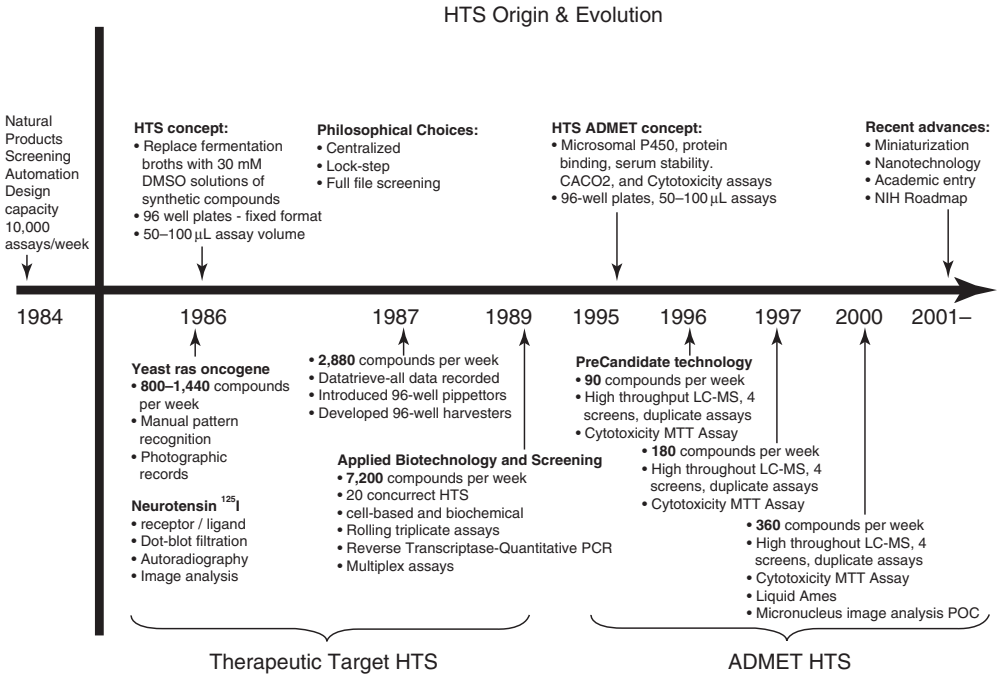


**Figure 1** Chronological sequence of key decisions and developments pertaining to the origin and evolution of HTS to ADMET (Pereira and Williams, 2007).

the 1990s, HTS methodologies became *the* intense focus throughout the entire pharmaceutical industry. In the 1990s, a new era of technology-driven R&D rapidly emerged that focused on clever bioassays and laboratory automation. Prior to this paradigm shift, the focus in pharmaceutical R&D was on better understanding and refining the traditional principles of pharmacology that helped develop first-in-class antimicrobials, cardiovascular, gastrointestinal, psychoactive, and other drugs (see Sneader, 2005).

At the start of the 1990s, animal models were still used for primary screening and lead optimization assays. Potency and efficacy directly relate to bioavailability in animal models; therefore, when a compound showed activity without overt side effects *in vivo*, then that compound would more likely progress on a preclinical development track than a compound that had only been tested *in vitro*. Before the era of HTS, *in vivo* data were complemented by functional assays in smooth muscles and receptor binding assays. The data from this "troika" of assay types, namely, (1) biochemical, (2) *in vitro* functional, and (3) efficacy *in vivo*, were typically generated in parallel. Such data served medicinal chemists extraordinarily well in the 1970s through around 1990, during which time occurred the greatest number of successful first-in-class drug approvals in history (Ng, 2004). The pharmaceutical industry during this period of productivity was viewed as "the perfect sector."

The productivity of pharmaceutical R&D in the 1970–1990s was due to many factors. The most important was that ADMET properties of new leads were considered

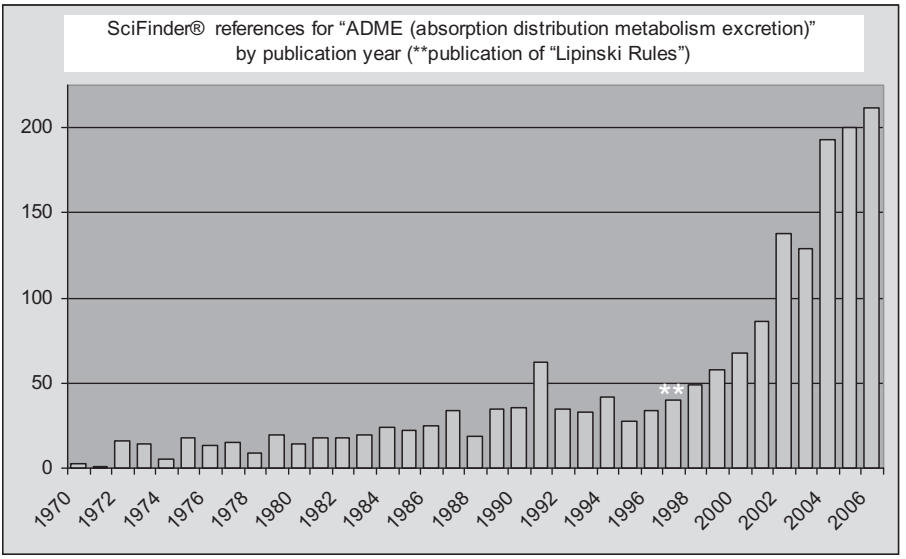


**Figure 2** Impact of the Lipinski Rules. Christopher Lipinski published two papers in 1997–1998 summarizing the properties of "drug-likeness," or guidelines to determine if a chemical compound with a certain pharmacological or biological activity has properties that would make it a likely orally active drug in humans, based mainly on the observation that most medication drugs are relatively small (<500 MW) and lipophilic (but log P < 5) molecules. Figure graciously provided by Wolfgang Sauer of Merck Serono S.A.

at the very start of pharmaceutical R&D continuum. Primary screens were run *in vivo*, and experienced medicinal chemists and pharmacologists knew there were particular requirements for oral bioavailability, such as that compounds should have molecular weights of <500 and log P values <5 (Buxton, 2006). Nonetheless, today's emphasis on more rigorous ADMET profiling earlier in the drug development process emerged almost immediately following Lipinski's seminal publication (see Fig. 2). The "Lipinski Rule of Five" came to form the modern foundation of the field of cheminformatics. Cheminformatics was seen as particularly useful as a *guideline* for what *not* to synthesize.

Medicinal chemists now understand that acquiring data on potency, efficacy, and selectivity of NCEs must be coupled with ADMET for best understanding structure–activity relationships (SARs) and for chemical leads to progress to clinical development (Fig. 3). To generate ADMET data in parallel for dozens of therapeutic targets on the hundreds of compounds each month that are synthesized in a large pharma, manual assay methods are simply not practical. For example, if the target under investigation is a kinase, such as C-MET (Gentile, Trusolino and Comoglio, 2008; Goldstein, Gray and Zarrinkar, 2008), over 100 kinase cross-reactivity assays may be performed on newly synthesized compounds to define on-target versus off-target cross-reactivity profiles. While the problem is less intense in ADMET than for kinase inhibitor profiling, either scale of compound profiling *cannot* be achieved manually. Hence, there has been

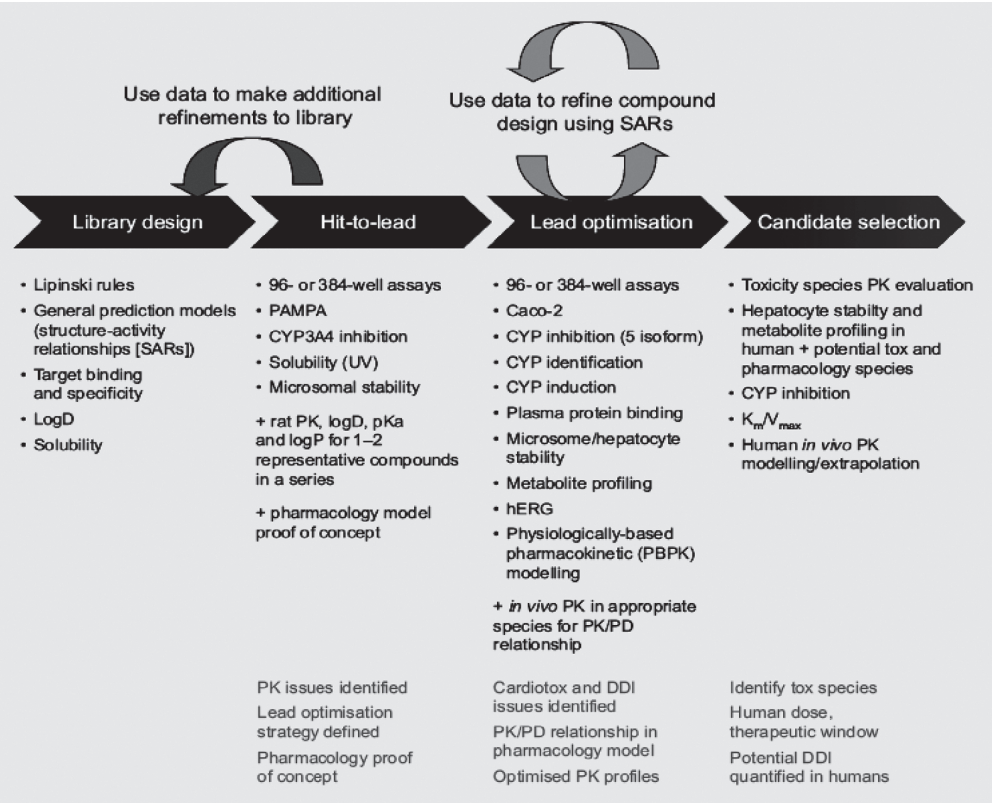


**Figure 3** A graphical representation of the pharmaceutical R&D continuum with respect to ADMET profiling (adapted from McGee, 2005).

intense interest starting in around 2000 in the application of laboratory automation for ADMET assays *in vitro*.

## 3 MODERN LABORATORY AUTOMATION: ORIGINS IN ADMET

Starting in the early 1980s, scientists running *in vitro* screening assays began enjoying the benefits of commercially available laboratory automation. Hamilton Inc. (Switzerland), Tecan Inc. (Switzerland), and Beckman Inc. (USA) developed automated liquid-handling workstations that exist to this day as modernized "upgrades" to the original releases. These workstations could solvate and serially dilute compound dozens or even hundreds of samples, and then add the required reagents to execute bioassays, which were run only manually then. By that time, the microtiter plate was in broad use, which facilitated the development of standardized laboratory workstations with a compact footprint that could generate hundreds of data points per day. The Beckman Biomek 1000 (Fig. 4a) was the first "fully automated" assay workstation and possibly the most popular in the history of pharmaceutical R&D. In fact, given its size, its cost, its simple

(a)

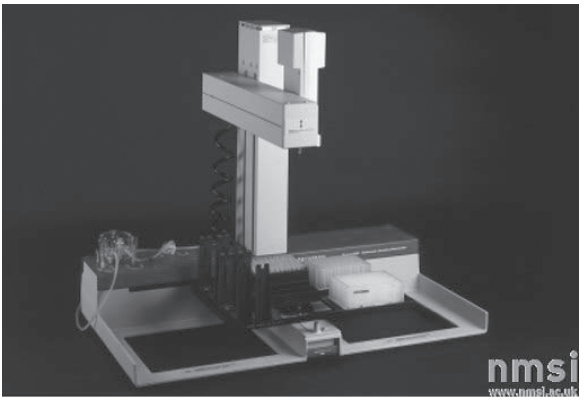


(b)

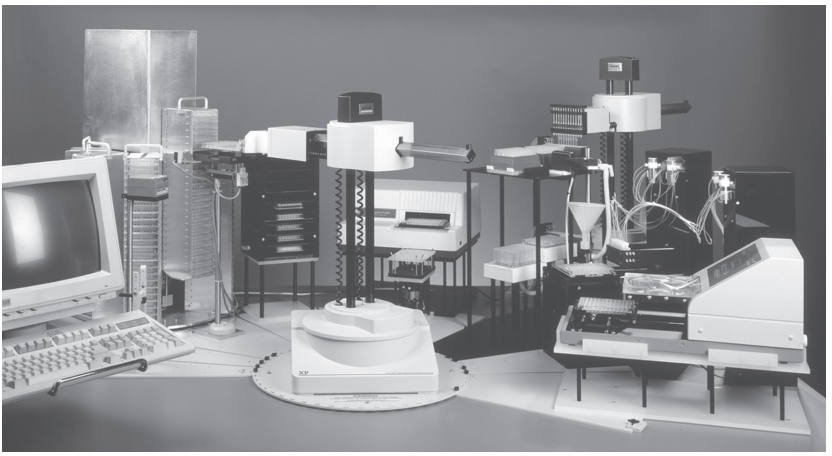

**Figure 4** **(a)** Biomek 1000 robotic workstation, 1991. The Biomek 1000 was the first instrument designed to allow scientists to perform drug screening on one integrated device. Its features included a graphical user interface anyone could use, a full complement of pipetting heads spanning from 1 μL to 1 mL, a plate washer, and a spectrophotometer (http://www.ingenious.org.uk/See/Scienceandtechnology/Biologyandbiotechnology/?target=SeeLarge&ObjectID={4DF77B0C-5B7B-CE96-3A8A-29E944E550D3}&viewby=images). (**b**) The Zymate System. Zymark International Inc. (a company founded by former executives of Waters, the leader in analytical instrumentation in the 1980s) introduced the radial table layout known as "pi" system. Their system was the first designed with "plug and play" modules that could be swapped in and out of the system in an extraordinary flexible manner. The Zymate was the first commercially available, fully automated, "standardized" screening platform that enabled 24/7 laboratory operations. The first adapters of the Zymate were analytical chemists who ran protocols to extract and measure drug and drug metabolites from clinical plasma samples (reproduced with permission of Zymark Corporation).

graphical user interface, ability to perform single- and multiple-channel pipetting, plate washing, and plate reading (with the aid of a unique absorbance tool for visible light), as well as other powerful capabilities (simple tool-changing ability, line-level programming, data logging, etc.), the Biomek 1000, arguably, remains unequaled as an automated bioassay workstation. Furthermore, the Biomek 1000 was sufficiently compact to fit in a standard laminar flow hood for preparing and running cultured cell assays.

Zymark advanced the concept of laboratory robotics to new "enablement" with the release of the first fully automated screening platforms equipped with a robotic arm capable of serving a large work-cell around 6 feet in diameter (Fig. 4b). Frank Zenie (then President and CEO of Waters, Inc., USA) and other senior executives from the Waters Corporation, the leading analytical instrumentation company at that time, founded Zymark Inc. in 1981. The Zymate systems from Zymark could be deployed for automating virtually any common laboratory assays, including solid-phase extraction, liquid chromatography–mass spectrometry (LC-MS) analysis, and other analytical tasks that were the staples of bioavailability assessment in preclinical development. By 1984, Zymark had validated their Zymate systems for the analysis of drugs in biological fluids. Their systems could run unattended around the clock (on good days) and had a significant impact for enabling high-throughput drug metabolism studies at pharmaceutical companies worldwide (see Fig. 4b and Banno and Takahashi, 1991).

## 4 COMPUTATIONAL CHEMISTRY 101: THE LIPINSKI RULES

Physicochemical information is important in drug discovery (Kearns, 2001, see Fig. 3) as follows:

- To rank order compounds for drug-like properties
- For rapid evaluation of the synthetic feasibility of compounds for HTS
- To aid in risk assessment of problems for ADMET, formulation, stability, or process
- For monitoring synthetic optimization
- Help objectively define drug advancement criteria to meet established discovery team goals.

Choosing the right balance of ADMET and pharmacokinetic properties is difficult because the rules that determine success remain, ultimately, not precisely defined. An excellent case in point is represented by the heavily cited "Lipinski Rule of Five" (see Fig. 2, Lipinski *et al.*, 1997; Lipinski, 2000), which originates from the considerable amount of data gathered by the industry on properties that maximize an *oral* drug candidate's probability of surviving development. The "Lipinski Rules" have been received enthusiastically by industry and academia because the rules provide a simple-minded framework (often, there lies genius in simple-minded frameworks) for defining the chemistry space for oral bioavailability, insofar as we understand this limited chemistry space, and recognizing that the "rules" are only around 75% "effective." Such rules have been applied almost universally to the design and selection of compounds for

lead discovery. The Lipinski Rules state that *poor oral bioavailability due to poor absorption or permeation is more likely when*:

- There are more than five H-bond donors expressed as sum of OH and NH.
- The molecular weight is over 500.
- The c log P is over 5.
- There are more than 10 (2 × 5) H-bond acceptors.

There is also danger in applying "rules" on ADMET too early in the R&D continuum; there is danger in applying them too late. Instead, the SR properties can be used for rank ordering or grouping compounds into high, medium, and low categories that alert the project team that there *may* be bioavailability issues with compounds that violate the Lipinski principles.

Most might agree, however, that the most important of the Lipinski and other rules that guide on oral bioavailability of compounds are for MW (<around 500) and c log P (generally <5), the latter of which may be assessed from its measured (i.e. "observed") log P or log D value.

## 5 AUTOMATED METHODS FOR DETERMINING LOG P/LOG D

The partition coefficient is the ratio of concentrations of un-ionized compound between two immiscible solutions (see Wikipedia under "Partition Coefficient"). To measure the partition coefficient of ionizable solutes, the pH of the aqueous phase is adjusted such that the predominant form of the compound is un-ionized. The logarithm of the ratio of the concentrations of the un-ionized solute in the solvents is called log P.

$$\log P_{oct/wat} = \log\left(\frac{[\text{solute}]_{\text{octanol}}}{[\text{solute}]_{\text{water}}^{\text{un-ionized}}}\right)$$

The distribution coefficient is the ratio of the sum of the concentrations of all forms of the compound (ionized plus un-ionized) in each of the two phases. For measurements of distribution coefficient, the pH of the aqueous phase is buffered to a specific value such that the pH is not significantly perturbed by the introduction of the compound. The logarithm of the ratio of the sum of concentrations of the solute's various forms in one solvent, to the sum of the concentrations of its forms in the other solvent is called log D:

$$\log D_{oct/wat} = \log\left(\frac{[\text{solute}]_{\text{octanol}}}{[\text{solute}]_{\text{water}}^{\text{ionized}} + [\text{solute}]_{\text{water}}^{\text{un-ionized}}}\right)$$

The log D is pH-dependent; hence, one must specify the pH at which the log D was measured. Of particular interest is the log D at pH = 4 (the physiological pH of blood serum). For un-ionizable compounds, log P = log D at any pH. Generally, Wikipedia provides an excellent summary of log P and log D, and the reader is referred there for additional hyperlinked references.

"Automation" of log P determination by high performance liquid chromatography (HPLC) inferred from retention times on hydrophobic (C-18) columns was first published in 1978 (Unger, Cook and Hollenberg, 1978). There are provisos, however, that always accompany the majority of expedient (*in vitro*) assay methods. Rapid HPLC methods work best for predicting accurate partition coefficients *within a related structural series of compound*; however, HPLC alone without physical partitioning coefficients of at least some members of the series of compounds under investigation is subject to potential error. Therefore, the classical "shake-flask" method is still used, in combination with HPLC with UV (and MS) detection to quantitatively measure the amount of compound distributed between the organic and aqueous phases. The importance of drug partition coefficients is that they often relate directly to blood–brain barrier (BBB) permeability and predict oral bioavailability to a reasonable degree.

Another approach to automated partition coefficient determination is a hydrophobicity scale based on the migration index from microemulsion electrokinetic chromatography (MEEKC) of anionic solutes. The MEEKC migration index has been automated and commercially available as a laboratory workstation termed the "pKa PRO system" (Fig. 5), which enables parallel, accurate, separation-based log P measurements that are insensitive to impurities and consume minimal amounts of sample. Several published studies suggest that the instrument provides more reliable correlations between log P and the classic octanol : water partitioning method than reverse-phase liquid chromatography.

Basic compounds should have a pKa value less than 10 to minimize charge effects. The use of MEEKC provides a validated method with good correlation to the traditional shake-flask technique (Neubert *et al.*, 1999). The method requires only microgram quantities of compound and is separation-based, allowing for detection and separation of potential impurities. The device is based on parallel, 96-channel microelectrophoresis. Published studies have noted that it is also possible to evaluate the pKa values of compounds in a high-throughput manner using the same instrument platform, by simply exchanging capillary arrays and buffer solutions.

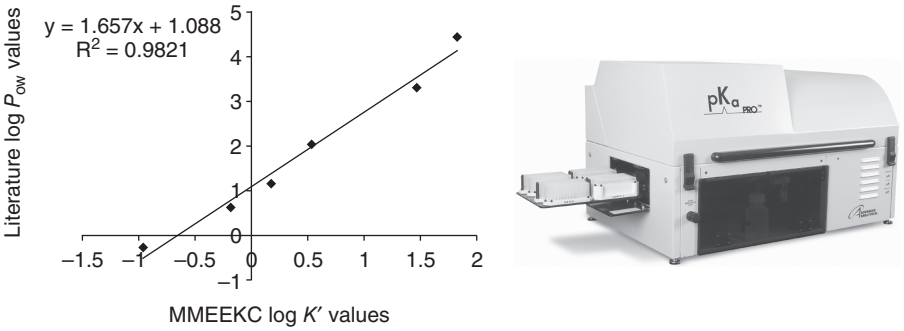


**Figure 5** The pKa PRO system equipped with the log P application package provides automated measurements of neutral and basic compounds over the log P range of –0.5 to 5.0 with a throughput of up to 46 samples/hour (reproduced with permission from Advanced Analytical Technologies, Inc., http://www.aati-us.com/systems/pKa.html).

## 6 DRUG SOLUBILITY

### 6.1 Background

In order for an orally administered compound to act, it must dissolve in the aqueous environment of the gastrointestinal tract and within the cytoplasm when the target is intracellular. The partition coefficient of a compound is directly related to its aqueous solubility—all else equal within a series. The more lipophilic a compound is, the lower is its aqueous solubility. Poor solubility can limit the absorption of compounds from the gastrointestinal tract, resulting in reduced oral bioavailability. It may also necessitate novel formulation strategies and hence increase cost and delays. Moreover, compound solubility can affect other *in vitro* assays. Poor aqueous solubility is the largest physicochemical problem hindering oral drug activity—hence on completion of compound synthesis, the determination of aqueous solubility is a priority.

Solubility data may prove critical in rationalizing disparate results in other bioassays. For example, a compound must be soluble in the assay buffer used to measure activity. Poorly soluble compounds within a series are more likely to have variable solubility in different buffer systems used in common bioassays, which can confound the SARs that are critical for effective lead optimization. In fact, some have shown that determination of the effective concentration of compounds in bioassays reveals marked errors in potency measurements due to marked underestimation of the actual concentration of compound that is in solution in the bioassay, particularly when a single concentration is used to assess activity by extrapolated $IC_{50}$ values (Popa-Burke *et al.*, 2004).

Determination of compound solubility is, therefore, quite important at every stage of the pharmaceutical R&D continuum including:

1. *The earliest stages of discovery.* Poorly soluble compounds can precipitate in the initial screen and lead to "false negatives," or lead to variable data due to minor changes in assay conditions (including room temperature, minute changes in the pH of a buffer, etc.).
2. *Lead optimization stage.* This, along with animal pharmacology, is the most expensive component of preclinical drug development. The greatest variety of bioassays are deployed at this stage to assess cross-reactivity and mechanism of action of nascent drug leads. The standard operating procedures (SOPs) for every bioassay typically call for specific (i.e. nonuniform) buffer recipes, ions, and cofactors—and sometimes even detergents—that surely affect pH, ionic strength, and potential ability to raise (or lower) the critical micelle concentration (CMC) of compounds in aqueous buffers.
3. *Preclinical development.* Compound solubility may be critical to clinical efficacy. One focus for formulation chemists is to optimize compound solubility, which may be accomplished by various physical formulations. For example, progesterone is rather insoluble as an oil and has very poor bioavailability (<1%); however, micronization may increase solubility (>5%), and therefore markedly enhance efficacy and potency (Fitzpatrick and Good, 2000).

### 6.2 Automated Methods Used to Determine Drug Solubility

Compound solubility is usually determined by either kinetic or thermodynamic methods. Kinetic assessments, which start from a solution of dissolved compound, are

now common in an early discovery setting. Often a turbidimetric end point is used for the kinetic solubility measurements. Thermodynamic assessments, where a saturated solution of a compound in equilibrium is prepared from solid, are considered to be the gold standard and take into consideration the crystal lattice of the compound, and therefore, the dissolution rate of the compound. Throughput is generally lower than kinetic measurements and larger quantities of analytical grade of compound are ideally used for these studies. For this reason, thermodynamic assays are usually performed in early development (rather than in the earlier discovery stages) to confirm earlier kinetic solubility results, to rule out potential artifacts, and to generate quality solubility data with crystalline material to support the best selection of potential clinical candidates. The traditional shake-flask method is still frequently used in the drug development phases for measuring thermodynamic solubility. The commercial "pSol Gemini" instrument (pION, Woburn, MA, USA) uses a pH-based approach, which is approved by the FDA to derive a complete solubility–pH profile. This allows experimental validation of prediction calculations for solubility, as well as c log P and c log D. Less than 100 micrograms of compound is required, making the method suitable for early stages of discovery where compounds can be in very limited supply, although throughput is limited to handfuls of compound per day (Avdeef, 2007; see also http://www.pion-inc.com/papers.htm).

A number of automated methods based on kinetic solubility have been developed, which have dramatically increased the throughput of solubility assessment. A brief overview of the methods can be found in Kearns (2001). The MultiScreen Solubility filter plate (Millipore, MA, USA) and Screening Method provides an automation-compatible, high-throughput means to estimate the aqueous solubility of hundreds of compounds per day. Using a single-point calibration, the screening ratio is easily and quickly derived, and compound solubility is readily approximated. Multiple samples, each requiring approximately 200 nanomoles (~100 μg) per result, can be run in parallel. The method allows for the analysis of approximately 45 compounds (duplicate determinations) per plate with the capability of completing four or more plates in a standard 8-hour day (see Fig. 6a and b).

Methods that use turbidimetric detection techniques are now becoming more popular as they bypass the need for HPLC analysis. Turbidimetry measures the onset of precipitation by either absorbance or nephelometry (light scattering) and is often used as the end point for kinetic solubility determination. These techniques are amenable to high-throughput format with rapid data generation and require only small amounts of test compound, usually from readily available dimethyl sulfoxide (DMSO) stock solutions (Bevan and Lloyd, 2000). An additional advantage of the turbidimetric method is that it is easily temperature controlled and incubations can be performed and monitored at physiological temperature.

### 6.3 Summary: Measuring Drug Solubility

Figure 7 depicts a simple approach to automate determination of aqueous drug solubility. The following provisos must be considered in establishing and then validly managing a high-throughput ADMET laboratory for determining drug solubility:

- Use the turbidimetric and equilibrium methods for discovery and development projects, respectively.

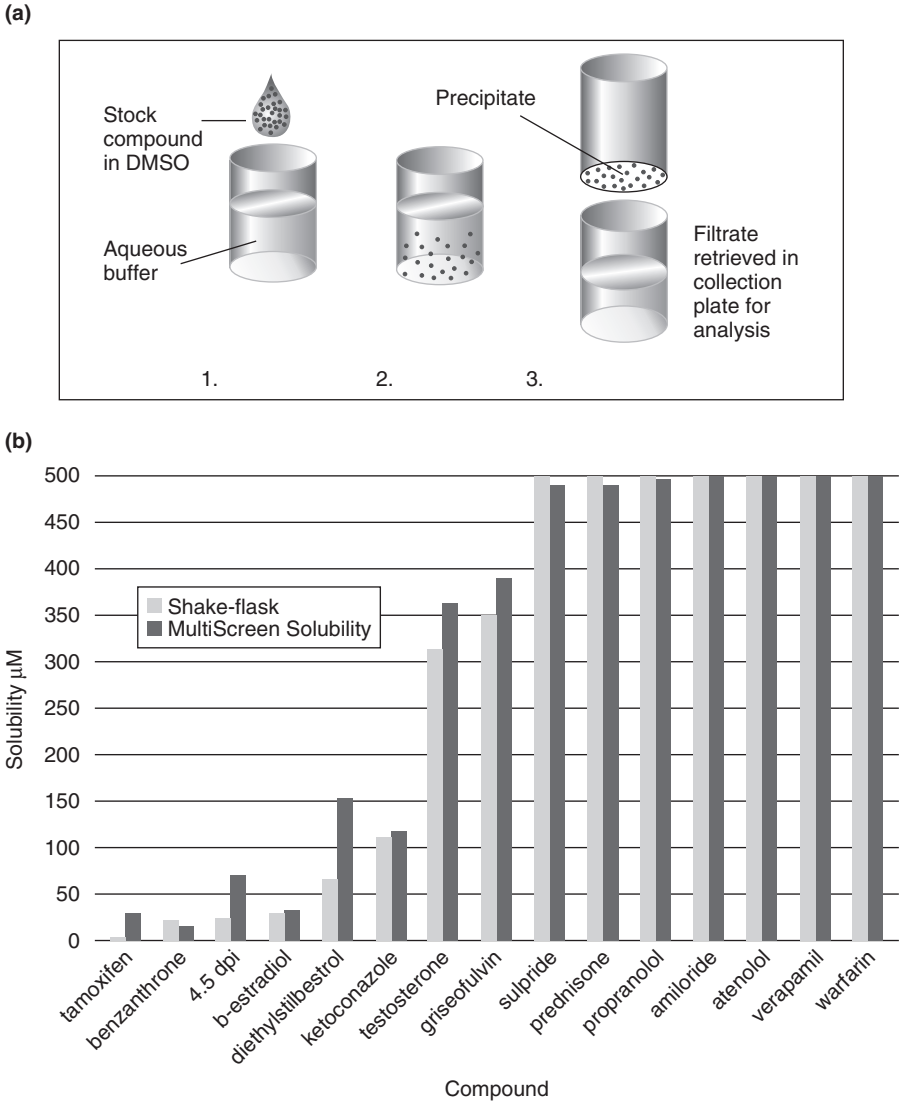


**Figure 6** Graphical depiction of basis for the MultiScreen solubility filter plate method (a) and its comparison with the shake-flask method (b). The filtrate retrieved in the collection plate may be analyzed by HPLC with ELSD (evaporative light scattering detection) (against a standard curve or known concentration of test compound, or by bioassay). (http://www.millipore.com/catalogue/module.do?id=C8875).

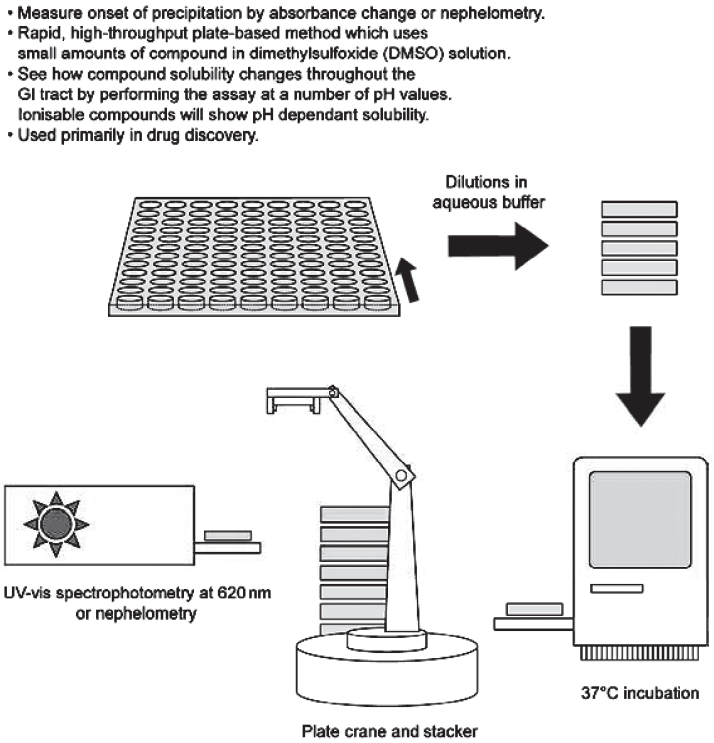


**Figure 7** An example of drug solubility assay automation using a turbidimetric method (from www.cyprotex.com).

- Assess solubility using different pH levels for ionizable groups.
- Assess solubility data first and then use this as a guide to interpreting results from *in vitro* experiments.
- Use surrogate data for ranking purposes rather than as "rejection filters."
- Use solubility data to aid design of new chemical libraries.
- Use solubility data as a guide for formulation strategies and dosage.
- If the therapeutic indication is an area of high unmet medical need (cancer), oral bioavailability would not be an appropriate parameter for early "Go-No-Go" decision making.
- It may be generally shortsighted to reject a compound based solely on solubility data. Route of administration and therapeutic plasma concentrations must also be considered, particularly in areas of desperate unmet medical need, such as oncology and even diabetic complications.
- Optimizing physical form (formulation) can have a significant impact on solubility.

It is notable that both Figures 6b and 8 are in reasonable agreement between automated high-throughput, microtiter-based kinetic methods and classical shake-flask methods. Figure 9 provides data demonstrating that it can be difficult to effectively mix reagents in wells of a microtiter plate uniformly. Others have shown that there can be considerable variability between different lots of the same compound that are simply distributed on multiple microtiter plates, and that the potency of a compound in a given assay can be off by more than threefold because of solubility issues (Popa-Burke *et al.*, 2004). In summary, great care and attention to details are obligatory for successful automation of pharmacokinetics assays run in parallel, and from whose data compounds will either be "killed" or advance. In the field of ADMET we may say, therefore, that "the executioner lies in the details."

## 7 DRUG PERMEABILITY

### 7.1 Why Is Permeability Important?

Drugs with low membrane permeability (i.e. low lipophilicity) are generally absorbed slowly from solution in the stomach and small intestine, despite that they may readily

| | | solubility ($\mu$g/mL) | | | |
|---|---|---|---|---|---|
| no. | compound | from solid by HPLC | from DMSO soln by HPLC | from nephelometry | comment[a] |
| 1 | acetazolamide | >500 | >111 | >111 | ✓ |
| 2 | allopurinol | 337 | >68 | >68 | ✓ |
| 3 | bendroflumethiazide | 12 | 33 | 13 | ✓ |
| 4 | benzocaine | >500 | >83 | >83 | ✓ |
| 5 | benzthiazide | 11 | 20 | 27 | ✓ |
| 6 | betamethasone | 63 | >196 | >196 | ✓ |
| 7 | butamben | 125 | 85 | 100 | ✓ |
| 8 | butylparaben | 139 | 88 | 97 | ✓ |
| 9 | chlorpropamide | >500 | >138 | >138 | ✓ |
| 10 | clofazimine | <1 | 30 | 1 | red soln, insoluble by eye |
| 11 | flurbiprofen | >500 | >122 | >122 | ✓ |
| 12 | gemfibrozil | >500 | >125 | >125 | ✓ |
| 13 | hydrocortisone | 258 | 175 | >181 | ✓ |
| 14 | hydroflumethiazide | 187 | >166 | >166 | ✓ |
| 15 | iodipamide | >500 | >57 | 57 | ✓ |
| 16 | nitrofurazone | 163 | >99 | >99 | ✓ |
| 17 | oxyphenbutazone | >500 | <1 | 40 | ✓with DMSO soln by HPLC result |
| 18 | phenacemide | 182 | >89 | >89 | ✓ |
| 19 | phenylbutazone | 320 | >154 | 77 | |
| 20 | prednisone | 98 | >179 | 89 | ✓ |
| 21 | propylparaben | 273 | 89 | >89 | ✓ |
| 22 | tolazamide | 350 | >174 | 90 | |
| 23 | trimethoprim | >500 | >145 | 145 | ✓ |
| 24 | tyrosine | no result | <1 | 91 | ✓ |
| 25 | hydroquinine | >500 | 138 | 120 | ✓ |
| 26 | morin hydrate | 192 | >151 | nd[b] | red soln |
| 27 | phenyl salicylate | 16 | <1 | >107 | |
| 28 | theobromine | >500 | >90 | 90 | ✓ |
| 29 | 2-hydroxy-3-isopropyl-6-methylbenzoic acid | >500 | >97 | >97 | ✓ |
| 30 | progesterone | 6 | 15 | 20 | ✓ |

[a] ✓ indicates acceptable agreement.

**Figure 8** A comparison of different methods to measure drug solubility. Results of the nephelometric, kinetic, and equilibrium solubility methods (Kearns, 2001).

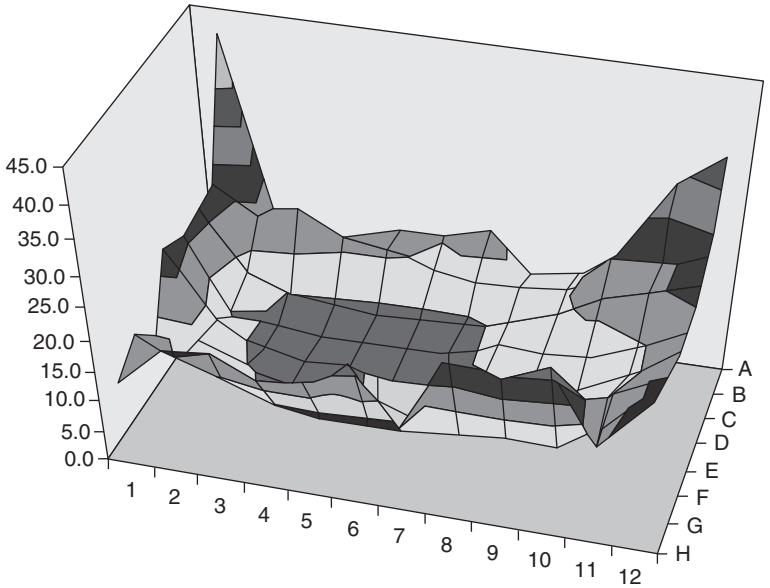


**Figure 9** An example of the importance of small details in the assay significantly influencing results. Stirring efficiency varies considerably from well to well on a 96-well plate when shaken on an orbital shaker. The shaker has room for four plates. The A12-corner faced the center of the shaking tray. The data are instructive of the principle that for successful conversion of classical ADMET assays to high-throughput automated assays, small details are critical (from http://pion-inc.com/images/PermPosterCPH01.pdf).

dissolve in aqueous media. The ability to assess the rate and extent of absorption across the intestinal tract is critical if a small organic molecule is to be developed as an orally available drug; and generally, this is the objective across the entire pharmaceutical industry. Understanding whether a compound is likely to enter the brain and central nervous system (CNS) via the BBB may also be important for any drug, not only psychoactive medicines. If the compound in question is a potential CNS therapeutic, crossing the BBB is a prerequisite, of course. Conversely, BBB penetration might cause unwanted off-target side effects for those drugs not intended for the CNS. Measuring drug permeability through appropriate monolayer model systems is a popular surrogate for measurement of intestinal permeability and oral bioavailability.

### 7.2 What Factors Determine Permeability?

Permeability—or drug absorption—may be simple (as in the case of ethanol, which is permeable through the gastrointestinal tract by passive diffusion) or the result of a complex composite of different physicochemical properties required for optimal oral absorption, as well as certain biological processes that may actively transport—or eliminate—drugs from the bloodstream. There are three pharmacokinetic processes that govern permeability: (a) passive diffusion, (b) active transport via drug transporters (requiring ATP), and (c) facilitated transport (not requiring ATP).

As a general rule, compounds that are very hydrophilic do not cross cell membranes readily; conversely, compounds that are very lipophilic tend to readily cross cell membranes. In general, medicinal chemists tend to focus on optimizing permeability first, because solubility may be improved through formulation at a later stage. For example, a micronized (fine pulverization) drug formulation can increase its rate of dissolution by increasing its surface area by a factor of 10 or greater. Conversely, pegylation can decrease solubility (rate of release) by a factor of 10 or more (Boyd, 2008). *It is not as easy to optimize drug absorption by formulation protocols as it is to enhance (or decrease) solubility*.

### 7.3 Methods Used for Investigating Permeability

Drug permeability cannot be accurately predicted by physicochemical factors alone because there are many drug transport pathways that are driven by protein transporters that recognize their ligands in ways that cannot be simply predicted. Consequently, a number of validated, cell-monolayer systems *in vitro* have been validated, including the human colon adenocarcinoma cell line, Caco-2 and the Madin Darby canine kidney (MDCK) cell line, to help predict intestinal permeability. These assays are commonly employed during early discovery, especially in lead prioritization for the chemotypes discovered by HTS, as well as for lead optimization to guide the medicinal chemist for enhancing intrinsic potency and efficacy (pharmacodynamics) in parallel with bioavailability (pharmacokinetics). Because of the complexity involved in measuring active transport, most lead optimization efforts focus on passive permeation to better predict the oral bioavailability of drugs. Caco-2 and MDCK (transfected and non-transfected) cell cultures are widely used *in vitro* methods for estimating both passive and active permeability. Although advances have been made in throughput and consistency with the introduction of drug transport plates amenable to automation, there is still a requirement for cell culture and the associated time delay in initiating studies and risks of infection (of cell cultures by environmental microorganisms). This has led to the development and adoption of alternative techniques.

A newer *in vitro* model, known as the parallel artificial membrane permeability assay (PAMPA), has been developed (Figs. 10 and 12). This approach ranks compounds on their passive diffusion rates alone. PAMPA is increasingly used as the first-line permeability screen during lead profiling. Both methods have been adopted for use in assessing the ability of compounds to cross the BBB (Fig. 11).

The functional concept depicted by Fig. 10 has been incorporated into the microtiter plate format by Millipore and termed the MultiScreen PAMPA System (also see Fig. 6), which is an important basis for fully automated methods to assess permeability. PAMPA data may be used to bin compounds into high- and low-permeability categories. Furthermore, close examination of BBB and high-resolution PAMPA examples led to the presumption that PAMPA can be considered a sophisticated method of measuring lipophilicity (as measured by octanol/water partitioning; see Fig. 11).

### 7.4 Summary

Figure 12 depicts automation of the PAMPA assay and Fig. 13 displays some reference compounds that may be used to "bin" compounds based on their permeability and solubility classes. Recently, Galinis-Luciani, Nguyen and Yazdanian (2007) have published a critical assessment of the PAMPA method (see Fig. 11). The authors

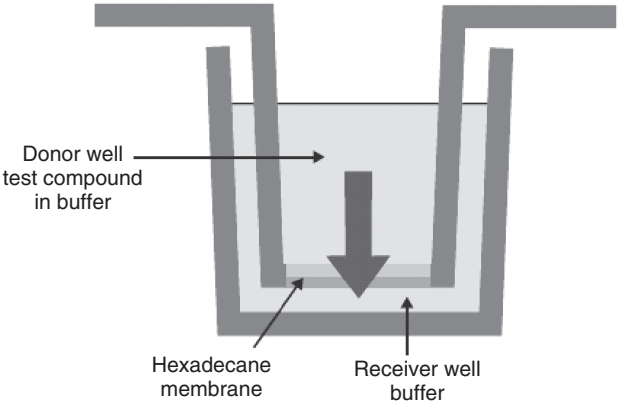


**Figure 10** The basis for the PAMPA assay depicted graphically. Compound is added to the donor well and the amount of compound transferred across an artificial bilayer into the receiver well is measured by HPLC (from www.cyprotex.com).

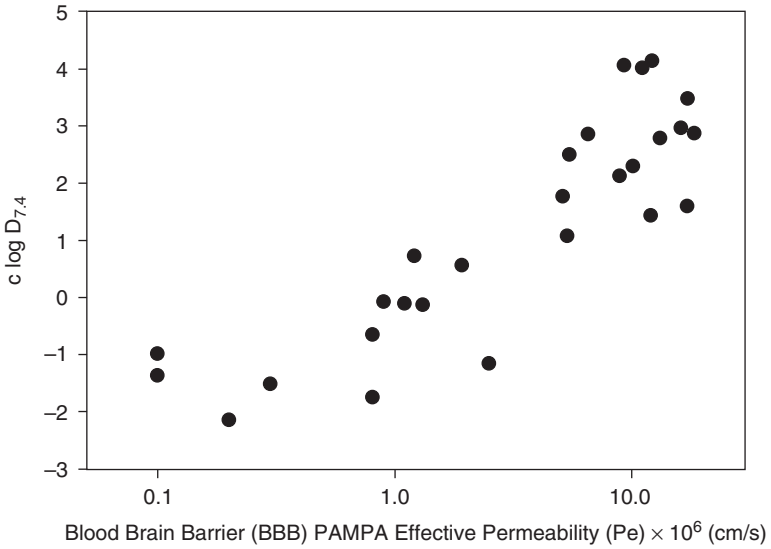


**Figure 11** Comparison of calculated log D7.4 values versus blood–brain barrier PAMPA effective permeability (from Galinis-Luciani, Nguyen and Yazdanian, 2007). On the one hand, Fig. 11 may be interpreted as support for the PAMPA assay as a predictor of blood–brain barrier permeability. On the other hand, the data imply that the PAMPA assay may not yield better information than calculated log D values.

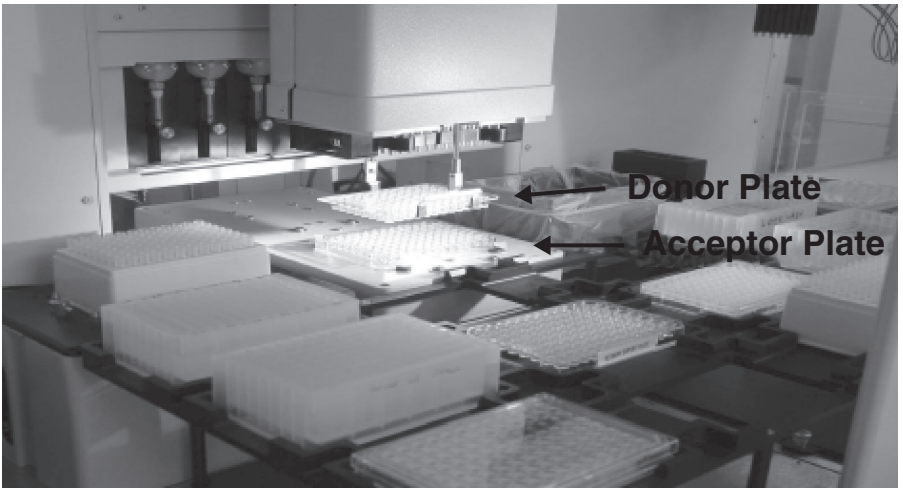


**Figure 12** An example of modern automation of the PAMPA Evolution 96 permeability assay. Compound and plate preparation were accomplished with the Biomek FX Laboratory Automation Workstation. The figure depicts a fully automated turnkey application for accomplishing the PAMPA assay. (Courtesy of Beckman Coulter.)

| Drug | Permeability class | Solubility class |
|---|---|---|
| Carbamazepine | High | Low |
| Ketoprofen | High | Low |
| Verapamil | High | Low |
| Metoprolol | High | High |
| Propranolol | High | High |
| Atenolol | Low | High |
| Ranitidine | Low | High |

**Figure 13** Reference compounds that may be used to "bin" test compounds (new chemical entities) into high-versus-low solubility and permeability classes.

concluded that PAMPA data collected at a single pH permitted compounds to be binned as being predicted to have high and low permeability, but the same outcome could be achieved using calculated log D7.4 values. Their conclusion is that the PAMPA assay has little, if any, advantage over mechanistic permeability assays such as the Caco-2 or MDCK cell models. The cellular models provide additional information about active transport. Upon examination of the balance of quality of data and quantity

of time, the PAMPA assay may be a very limited asset to a drug discovery effort these authors conclude. Others conclude that the success of using PAMPA in drug discovery depends on careful data interpretation, use of optimal assay conditions, implementation and integration strategies, and education of users with respect to expectations and proper application of the data (Avdeef *et al.*, 2007).

## 8 DRUG TRANSPORTERS IN INTESTINAL ABSORPTION

Understanding the role of transporters in pharmacokinetic processes has become a key objective over recent years. Transporters play an important function in renal clearance, biliary clearance, hepatic uptake, intestinal absorption, and BBB permeability, as well as cancer cell resistance. For a more detailed description of transporters and their roles in drug absorption and elimination, see Chapter Introduction to Drug Transporters. Both influx transporters, which facilitate the entry of drugs into cells, and efflux transporters, which limit the entry of drugs or enhance the removal of drugs from the cells, are crucial in determining concentrations of drug in the body and within various cells and tissues (see Fig. 14). As is the case with cytochrome P450 (CYP) enzymes, transporter-mediated drug interactions have been reported in both humans and animals. The substrate specificity of the individual transporters often overlaps and therefore the evidence for transporter interactions is often less conclusive. Many inhibitors and inducers can simultaneously affect both drug transport and CYP enzymes. The lack of specific probe substrates and inhibitors for the individual transporters has hindered our development of simple *in vitro* assays.

Our knowledge is growing of the transporters present in the various tissues of the body. In enterocytes, a number of different transporters exist which are involved in both drug uptake and drug efflux, as illustrated in Fig. 14. Although many transporters have been identified, P-glycoprotein (P-gp) is the most extensively studied and many of the reported transporter-mediated clinical interactions focus around P-gp. These include clinical drug–drug interactions (DDIs) between the P-gp substrate digoxin and other drugs such as erythromycin, quinidine, verapamil, talinolol, and clarithromycin (Doering, 1979; Leahey *et al.*, 1979; Bigger and Leahey 1982; Klein *et al.*, 1982; Belz *et al.*, 1983; Maxwell, Gilmour-White and Hall, 1989; Wakasugi *et al.*, 1998; Verschraagen *et al.*, 1999; Westphal *et al.*, 2000). The tools for evaluating substrates and inhibitors of P-gp are much advanced compared with the other transporters. The most popular *in vitro* methods for investigating P-gp are Caco-2 cells and either the MDCKII cell line or the LLC cell line transfected with the human *MDR1* gene. These cells can be grown as polarized monolayer and the bidirectional transport of the drug through the monolayer can easily be investigated. The Millipore MultiScreen Permeability plates have been successfully used for full automation of transport (permeability) assays.

A number of higher-throughput methods also exist. For example, ATP is the energy source for moving substrates by the ABC transporters; therefore, generation of inorganic phosphate from ATP hydrolysis can easily be measured using a simple colorimetric reaction. This assay is useful for distinguishing between substrates and inhibitors as well as rapidly estimating the kinetics of the drug transporter interaction (Garrigues *et al.*, 2002; Glavinas *et al.*, 2007).

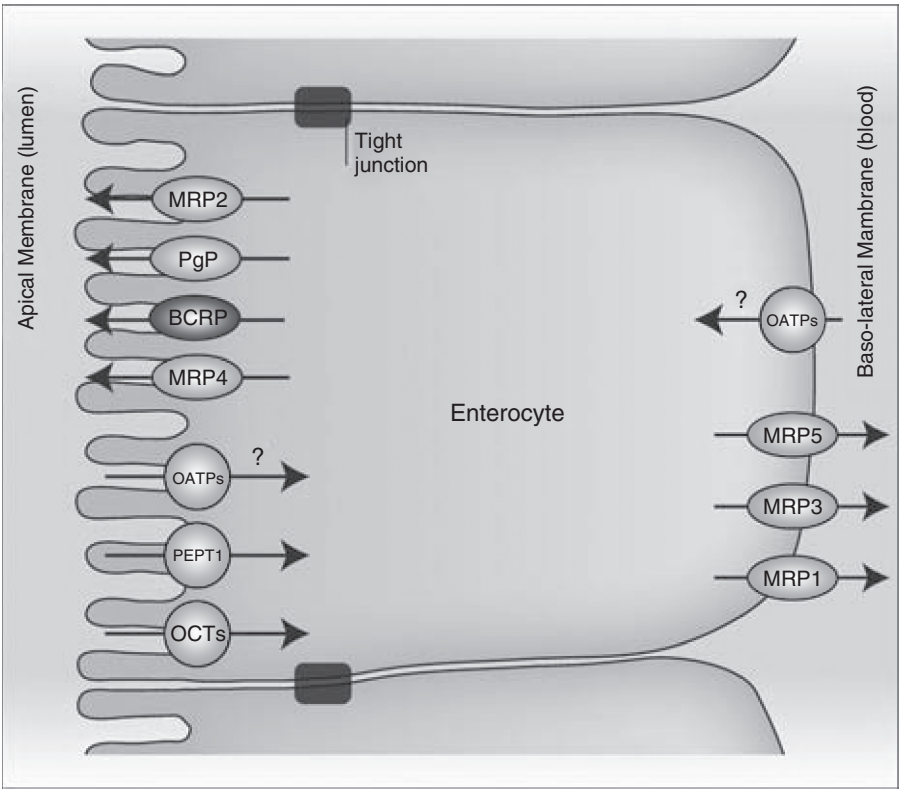


**Figure 14** Localization of efflux and uptake transporters in the enterocyte. It is important to note that there is still dispute over the presence of OATP transporters in the enterocyte and, for this reason, these have been marked with question marks in this figure (with permission from SOLVO Biotechnology, Budapest, Hungary).

## 9 DRUG METABOLISM

### 9.1 Hepatic Intrinsic Clearance

The ideal scenario is that a drug has an optimal half-life for once-a-day dosing with no toxic effects from drug accumulation. As the majority of drug metabolism occurs in the liver, the most popular *in vitro* models employed for evaluating intrinsic clearance in a discovery setting are either human or animal-derived hepatic microsomes, hepatic S9, or hepatocytes. Typically, microsomes are used as a first line screen, and then the most favorable compounds are then assessed in the presence of hepatocytes. The data generated from both these models are used to select compounds for further development.

***9.1.1 Liver Microsomes*** Subcellular fractions such as liver microsomes are one of the most commonly used *in vitro* models of hepatic clearance in drug discovery.

Microsomes are easy to prepare and can be stored for long periods of time. They are easily adaptable to 384-well formats that enable large numbers of compounds to be screened rapidly and cost-effectively. A pool of human donors is typically assessed to reduce the influence of interindividual variability. A wide range of different species is commercially available and therefore the assay is useful in assessing interspecies differences in metabolism.

For metabolic stability studies, the microsomes are incubated with the test compound at 37 °C in the presence of the cofactor, NADPH, which initiates the reaction. Solvent terminates the reaction and precipitates the protein. Following centrifugation, the supernatant is analyzed by HPLC coupled with MS. The substrate concentration is typically set between 1 and 5 μM to ensure adequate analytical sensitivity yet to limit solubility issues or risk of saturation of the enzymes.

In discovery, knowledge of metabolic pathways is limited and, therefore, substrate depletion over a number of time points is now an accepted approach to calculate intrinsic clearance. Due to the low test compound concentration in the final samples, saturation of the response on the mass spectrometer is unlikely and it is generally considered unnecessary to quantify the samples using a standard curve.

Although microsomes contain the full complement of phase I enzymes, cytosolic metabolizing enzymes are absent and phase II metabolism may be overlooked. As a result, microsomes tend to underestimate intrinsic clearance for those compounds that undergo metabolism by these phase II enzymes. One of the most common phase II enzymes, uridine glucuronyl transferase (UGT) is present in microsomes. However, in contrast to CYPs and the flavin-containing monooxygenases, the active site of the UGTs resides in the lumen of the endoplasmic reticulum (ER), and the ER membrane provides a diffusional barrier for substrates, cofactors, and products (Meech and Mackenzie, 1997). In a microsomal incubation, disruption of this barrier is required to remove the latency of UGTs. Traditionally, detergents were added to the incubation; however, the optimal levels of detergent often varied for different substrates which circumvented this as a routine early screen (Lett *et al.*, 1992). More recently, the use of the antibiotic fungal peptide alamethicin has been proposed as a universal substrate-independent *in vitro* method to study glucuronidation activity in human liver microsomes (Fisher *et al.*, 2000). Alamethicin forms well-defined pores in the membrane and allows access to the enzyme without affecting gross membrane structure or general intrinsic enzyme catalytic activity. Supplementation of the microsomal incubations with alamethicin and the cofactors, NADPH and UDGPA, allows both CYP-mediated and UGT-mediated metabolism to be evaluated as an early stage screen. Extending the range of enzymes investigated is likely to improve the prediction of hepatic clearance from microsomal stability data.

The data analysis system at Cyprotex enables issues such as retention time shifts, peaks in blank samples, internal standard fluctuations, atypical profile fluctuations, inconsistencies in the minus NADPH data, and control compound failures to be flagged with warning messages (Fig. 15). These warnings can be automatically integrated into comments on the final spreadsheets. As well as dramatically improving efficiency, this type of system acts as an additional checking mechanism and so reduces the chances of error in the data. Checks performed by scientists are still deemed necessary to maintain the quality of the data. However, as these systems become more sophisticated, it is entirely feasible that checking systems relying entirely on machine interpretation will eventually become more commonplace, generating higher quality and less subjective data than manual human intervention.

| Auto check warning | |
|---|---|
| **Condition failed** | **Warning** |
| FALSE | Interfering peak |
| FALSE | Compound detection problem |
| FALSE | Minus NADPH control low |
| FALSE | Minus NADPH control high |
| FALSE | Low response |
| FALSE | Low internal standard response |
| FALSE | Using all five time points |
| FALSE | Variable profile |
| FALSE | Retention time shifts |
| FALSE | Test compound not detectable |
| TRUE | Control compound out of range |

| | |
|---|---|
| **Full warning message** | Control compound out of range |

| | |
|---|---|
| **Accept** | FALSE |
| **Reject** | TRUE |

**Figure 15** An example of automated checking procedures performed at Cyprotex in the Cloe Screen Microsomal Stability assay to assure QC/QA in automated analytical methods.

***9.1.2 Supplemented S9 Fraction*** S9 fraction consists of both cytosolic and microsomal enzymes. As with microsomes, this model is amenable to HTS, and as such is rapid and cost-effective. Despite containing the full complement of hepatic enzymes, the S9 fraction is not energetically competent and requires cofactor supplementation to enable phase II metabolism. The activity of the S9 tends to be lower than that of microsomes and as such higher protein concentrations are required in the incubations, which can lead to issues with extensive binding to the S9 protein. This model tends to be less popular than human liver microsomes as a first line screen.

***9.1.3 Hepatocytes*** Hepatocytes have obvious advantages over subcellular fractions in that they are whole cell systems that are energetically competent with active transporters. This avoids issues with latency and circumvents the need for cofactor supplementation. Due to these reasons, the predictive power of hepatocytes tends to be considered superior compared with microsomes. With improvements in cryopreservation techniques, practical issues associated with isolation of the hepatocytes are reduced, and intrinsic clearance assessment using cryopreserved hepatocytes has increased noticeably over the past five years (Hewitt *et al.*, 2007). Despite this, the use of hepatocytes as a model for intrinsic clearance still remains a second line screen for many companies, mainly a consequence of increased running costs. Hepatocytes tend to be the system of choice for applications such as metabolite profiling and induction studies which are performed later and are less cost-sensitive.

## 9.2 Approaches for Interpretation of Intrinsic Clearance Data

Often, classification bands are used to categorize compounds into low, medium, or high clearance. These classification bands are calculated assuming an extraction ratio (the

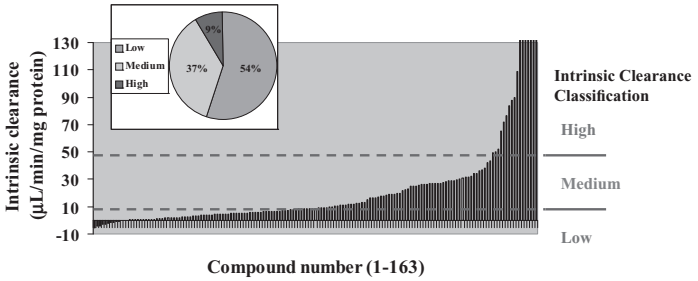


**Figure 16** Classification of microsomal intrinsic clearance for a set of 163 FDA-approved drugs. Drugs were screened in the Cloe Screen Microsomal Stability assay at Cyprotex. From the set of drugs screened, over 90% of the FDA-approved drugs were classified as low or medium intrinsic clearance in the microsomal stability assay. Only 9% of the drugs were rapidly cleared.

**TABLE 1 Classification bands for categorizing compounds into high, medium, and low clearance.**

| | Intrinsic clearance (μL/min/mg protein) | |
|---|---|---|
| Clearance category[a] | Human liver microsomes | Human hepatocytes |
| Low | ≤8.6 μL/min/mg protein | ≤3.5 μL/min/$10^6$ cells |
| Medium | Between 8.6 and 47 μL/min/mg protein | Between 3.5 and 19 μL/min/$10^6$ cells |
| High | ≥47 μL/min/mg protein | ≥19 μL/min/$10^6$ cells |

[a]Assumes:

- Low clearance equivalent to an extraction ratio of less than or equal to 0.3 and high clearance equivalent to an extraction ratio greater than or equal to 0.7 (hepatic clearance = extraction ratio × hepatic blood flow).
- Conversion of hepatic clearance to intrinsic clearance uses the equation: $CL_{int} = CL_H/(fu(1 - E))$ and assumes fu = 1.
- Scaling factors to convert from mL/min/70 kg to μL/min/mg protein assumes 1800 g liver/70 kg man and 40 mg protein/g liver (microsomes) or $99 \times 10^6$ cells/g liver (hepatocytes).

fraction of drug that is eliminated from the blood by an organ) of 0.3 and 0.7 for the low and high boundaries, respectively, as defined in Wilkinson and Shand (1975). Figure 16 suggests that the large majority of FDA-approved drugs (>90%) have low or medium clearance classification in the microsomal stability assay (Table 1).

The caveat with using the classification approach is that only microsomal clearance is taken into account for the prediction of hepatic clearance. Therefore, a common approach is to predict hepatic clearance from the intrinsic clearance by incorporating hepatic blood flow, plasma protein binding, and microsomal binding.

### 9.3 Drug–Drug Interactions

DDIs are a significant safety concern as substantial changes in blood and tissue concentrations of drug or metabolite can occur. These fluctuating exposure levels can alter the safety and efficacy profile of a drug and/or its metabolites, especially for drugs with

a narrow therapeutic range. DDIs remain a major regulatory hurdle, which can lead to early termination of development, refusal of approval, prescribing restrictions, or withdrawal of drugs from the market. Therefore, the potential for drug interactions should be evaluated early. These studies should be of sufficient quality as the outcome ultimately influences the design of clinical trials. Metabolic drug interaction studies focus on the CYP family as these enzymes play a key role in the metabolic clearance of drugs. The availability of human tissue is a valuable tool in the *in vitro* assessment of drug interactions.

***9.3.1 Cytochrome P450 Inhibition*** The two most popular *in vitro* systems for assessment of CYP inhibition are human liver microsomes and recombinant CYP enzymes. These systems investigate the potential of the test compound to inhibit the metabolism of a probe substrate by the main CYP isoforms. The probe substrate selected should be predominantly metabolized by a single CYP enzyme and should have a simple metabolic scheme preferably with no sequential metabolism. Initial evaluation of time and protein linearity as well as enzyme kinetic studies should be performed during the setup of these screening assays. For early discovery screening, an $IC_{50}$ (concentration at which the test compound causes 50% inhibition of the probe substrate metabolism) is deemed sufficient for ranking and selection purposes. $IC_{50}$ studies should be performed using a probe substrate concentration at or below its Michaelis–Menten constant ($K_m$) with a minimum of six inhibitor concentrations and under linear conditions with respect to time and protein. Typically, compounds are categorized into the following classification bands (Krippendorff *et al.*, 2007):

- Potent inhibition $IC_{50} < 1\,\mu M$
- Moderate inhibition $IC_{50}$ between 1 and $10\,\mu M$
- No or weak inhibition $IC_{50} > 10\,\mu M$.

If $IC_{50}$ studies indicate that an investigational drug does not inhibit the main CYP isoforms, then corresponding *in vivo* inhibition-based interaction studies of the investigational drug and concomitant medications eliminated by these pathways are typically not required.

For later stage studies where decisions are made regarding clinical studies and where inhibition is observed in the $IC_{50}$ screen, a more detailed $K_i$ determination is considered necessary. The $K_i$ (or inhibition constant) is the affinity with which the inhibitor binds to the enzyme. $K_i$ determination has several advantages over $IC_{50}$ determination in that it is independent on type and concentration of substrate and incubation conditions and, as such, is more reproducible between laboratories. Data are useful in identifying the type of inhibition (i.e. competitive, noncompetitive, uncompetitive, or mixed) and Eadie Hofstee plots tend to be the preferred method for differentiating between the various types of reversible inhibition.

Probe substrates, which form fluorescent products, have been introduced as an alternative to the traditional probe substrates. These substrates are common in an early discovery setting as fluorescent plate-based analysis is faster than LC-MS analysis and so throughput can be enhanced. However, there are several disadvantages in using fluoroprobes for P450 inhibition. Firstly, most of these substrates are not P450 isoform-specific, and so they have to be screened using individually expressed enzyme. The absence of other P450 enzymes will eliminate competing pathways of metabolism. Secondly, clinical assessment of these substrates is difficult in humans and therefore

understanding their relevance to the *in vivo* situation is limited. Thirdly, there is a poor correlation of inhibitory potential between different fluoroprobes (Stresser *et al.*, 2000). Finally, if the inhibitor or its metabolite fluoresces, then interference in the assay can occur, leading to false results. As a consequence, regulatory authorities have not accepted fluorescent probes, and the majority of companies still currently employ the traditional probe substrate methods, which utilize LC-MS/MS. One successful approach to reduce analytical time in these traditional methods has been to combine supernatants from multiple samples into a single "cassette," following termination of the individual reactions. Coupled with rapid gradient ultra-performance liquid chromatography methods, this enables analysis times to be dramatically reduced and enables HTS of the traditional methods to be a more realistic proposition.

***9.3.2 Mechanism-Based (Irreversible) CYP Inhibition*** Mechanism-based inhibition, also known as "suicide inhibition" or "time-dependent inhibition," is an irreversible interaction where a covalent bond is formed between a reactive metabolite and the active site of the enzyme, destroying the enzyme's activity. The consequences of mechanism-based inhibition are considered to be more serious than reversible inhibition because the inactivated enzyme must be resynthesized before activity is restored. In addition, the irreversible inactivation usually implies the formation of a covalent bond between the metabolite and the enzyme, which can lead to hapten formation and in some cases, can trigger an autoimmune response. The regulatory authorities are now aware that this is a significant problem and are proposing that NCEs are screened for the potential to cause mechanism-based inhibition. Clinically important mechanism-based inhibitors of CYP include antibacterials (e.g. clarithromycin, erythromycin, and isoniazid), anticancer agents (e.g. tamoxifen and irinotecan), anti-HIV agents (e.g. ritonavir and delavirdine), antihypertensives (e.g. dihydralazine, verapamil, and diltiazem), sex steroids and their receptor modulators (e.g. gestodene and raloxifene), and several herbal constituents (e.g. bergamottin and glabridin) (Zhou *et al.*, 2005).

The design of the protocol for mechanism inhibition involves pre-incubating the test compound at a single concentration with NADPH and 10 times the final concentration of human liver microsomes for 30 minutes. If the test compound is a mechanism-based inhibitor, then it will be activated in the presence of the NADPH and bind irreversibly to the active site of the enzyme. Following the pre-incubation stage, an aliquot of the incubation is diluted 10-fold with buffer containing the CYP isoform-specific substrate and then a second incubation occurs. Mechanism-based inhibitors are not dependent on the probe substrate concentration; therefore, it is not necessary to perform the studies at the $K_m$ of the substrate. By diluting the test compound 10-fold for the final incubation with the marker substrate, any potential reversible inhibition is minimized. Several control incubations are included in addition to the test well where the pre-incubation is performed in the absence of NADPH, and absence of test compound. Details of all incubations are illustrated in Fig. 17. The percentage inhibition following pre-incubation is calculated for the test compound using the following equation (the peak area ratio is the peak area of the probe substrate metabolite on LC-MS/MS divided by the peak area of the internal standard on LC-MS/MS):

$$\%\text{Inhibition following pre-incubation} = \left(1 - \left(\frac{\text{Peak area ratio in presence of NADPH}}{\text{Peak area ratio in absence of NADPH}}\right)\right) \times 100$$

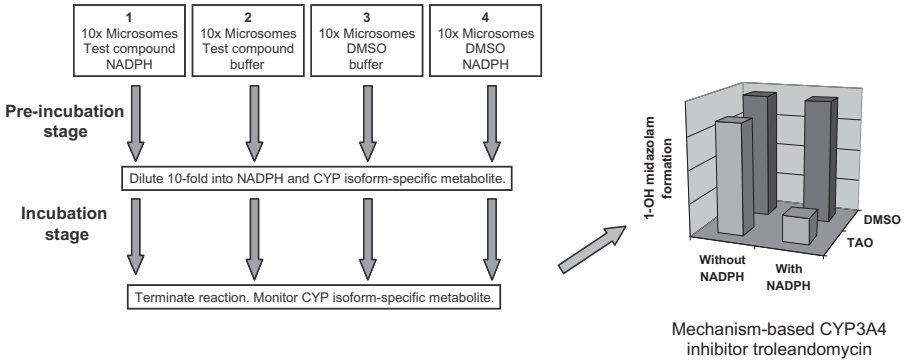


**Figure 17** Overview of mechanism-based inhibition protocol. Pre-incubation 1: This pre-incubation consists of test compound, NADPH, and 10-fold the concentration of microsomes. If the test compound is a mechanism-based inhibitor, then it should be activated and bind irreversibly to the active site. Pre-incubation 2: This pre-incubation consists of test compound, buffer, and 10-fold the concentration of microsomes. As no NADPH is present in the pre-incubation, no metabolic activation of the test compound by CYP should occur in the pre-incubation step; therefore, mechanism-based inhibition should not occur. Pre-incubation 3: This pre-incubation is one of the vehicle control incubations where the test compound is replaced with DMSO, and buffer replaces the NADPH. It consists of DMSO, buffer, and 10-fold the concentration of microsomes. As the test compound and NADPH are absent, no reversible or mechanism-based inhibition of the CYP-specific isoform should occur. Pre-incubation 4: This pre-incubation is one of the vehicle control incubations where DMSO replaces the test compound. It consists of DMSO, NADPH, and 10-fold the concentration of microsomes. As the test compound is absent, no reversible or mechanism-based inhibition of the CYP-specific isoform should occur.

If the compound is identified to be a mechanism-based inhibitor, then more detailed studies can be performed at multiple test compound concentrations and multiple pre-incubation times to calculate the rate of inactivation ($K_{inact}$) and the inhibition constant ($K_I$).

Compared with reversible inhibition, mechanism-based inhibition more frequently causes pharmacokinetic–pharmacodynamic DDIs, as the inactivated enzyme has to be replaced by newly synthesized CYP protein. The resultant drug interactions may lead to adverse drug effects, including some fatal events. For example, when mechanism-based inhibitors of CYP3A4 are coadministered with terfenadine, cisapride, or astemizole, torsades de pointes (a life-threatening ventricular arrhythmia associated with QT prolongation) may occur (Zhou *et al.*, 2005). This method is amenable to HTS and due to the clinical implications of such an interaction, the industry is now addressing this issue at a much earlier stage of drug discovery.

***9.3.3 CYP Reaction Phenotyping*** Drug-metabolizing enzyme identification studies, or reaction phenotyping studies, are a set of experiments that identify the specific enzymes (usually CYP enzymes) responsible for metabolism of a drug. *In vitro* identification of drug-metabolizing CYP enzymes helps predict the potential for *in vivo* DDIs and the impact of polymorphic enzyme activity on drug disposition. Furthermore, if suitable *in vitro* studies at therapeutic concentrations indicate that the main CYP isoforms do not metabolize an investigational drug, then clinical studies to evaluate the

effect of CYP inhibitors/inducers on the elimination of the investigational drug will not be required. The methods for assessing reaction phenotyping include recombinant enzymes, potential of selective P450 inhibitors to inhibit test compound, antibody inhibition studies, and/or correlation analyses with individual donor activity. All these techniques can be adapted to automated plate-based systems.

***9.3.4 CYP Induction*** Induction of CYP can either lead to a reduction in therapeutic effect as a result of increased metabolic clearance, or adverse effects as a consequence of increased formation of an active metabolite. The regulatory authorities currently recommend that hepatocytes be used for the purpose of assessing P450 induction of NCEs. Traditionally, this has been performed by incubating the potential inducers with freshly isolated cultured human hepatocytes prior to quantification of CYP-specific probe substrate metabolism. However, it is becoming more popular to assess mRNA levels and/or protein levels in parallel with activity. This can be helpful when both inhibition and induction are occurring simultaneously. More recently, several studies have also reported that cryopreserved hepatocytes can be used for CYP induction studies (Roymans *et al.*, 2005), and this has become more widely accepted due to the obviously practical advantages.

Due to the fact that the regulatory authorities currently suggest a minimum of three donors be investigated in these studies, it can be expensive to perform these studies, and therefore higher-throughput methods are attractive. One such method that is growing in popularity is the aromatic hydrocarbon receptor (AhR) and pregnane X receptor (PXR) reporter gene assay. However, there is still question as to the *in vivo* relevance of these alternative approaches and the data tend to be used as supporting evidence alongside other more acceptable methods.

***9.3.5 Reactive Metabolites*** Idiosyncratic drug toxicity is a major concern for the pharmaceutical industry, as often the toxicity is not detected until after launch of the drug. A high proportion of drugs that cause idiosyncratic toxicity have the capability of forming reactive metabolites. Walgren, Mitchell and Thompson (2005) found that from a set of 21 drugs that have either been withdrawn from the US market due to hepatotoxicity or have a black box warning for hepatotoxicity, there was evidence for the formation of reactive metabolites in five out of six drugs that were withdrawn, and 8 out of 15 drugs that had black box warnings.

Reactive metabolites are capable of covalently binding to macromolecules by nucleophilic substitution which is thought in some circumstances to instigate toxic effects. The approach by Merck and several other companies has been to impose a panel of screening assays that can detect reactive metabolite formation and the risk can then be minimized by the appropriate structural modification during the lead optimization stage (Evans *et al.*, 2004). The use of chemical trapping agents such as reduced glutathione (Yan *et al.*, 2005; Masubuchi, Makino and Murayama, 2007) or glutathione ethyl ester (Soglia *et al.*, 2004) has become a popular early screen for detecting reactive metabolites. The experiments are usually performed in the presence of liver microsomes and stable adducts formed can be detected by LC-MS/MS. These studies are useful in providing indirect information on the structure of the reactive species. One disadvantage to this screen is that no single small molecule serves as universal surrogate and screening through a panel of trapping agents may be needed.

Covalent binding to proteins is another recognized method for evaluating reactive metabolite formation. As radiolabeled compound is required for these studies, their

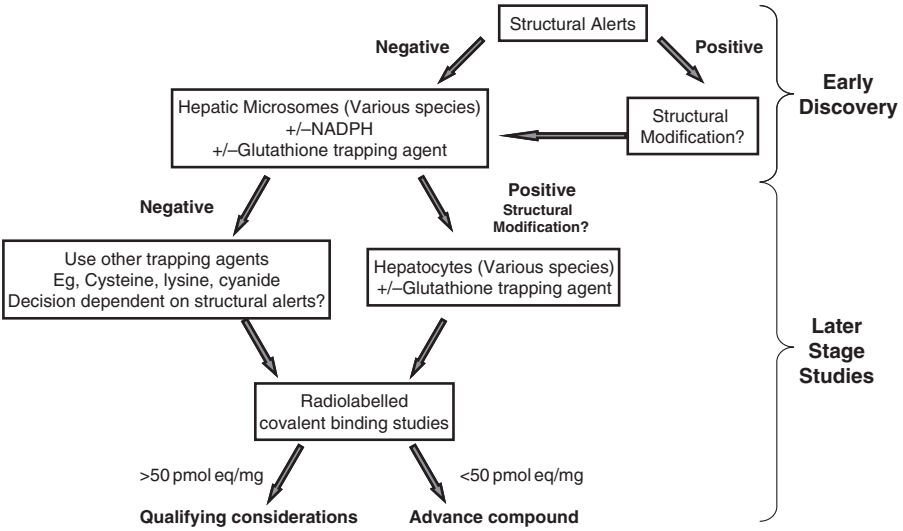


**Figure 18** Suggested screening strategy for reactive metabolite assessment (adapted from Evans *et al.*, 2004).

suitability as early screening tools is restricted. Covalent binding studies can either be performed *in vitro*, by investigating binding to liver microsomes, or *in vivo*, by dosing to rats and measuring adduct formation in the liver or plasma proteins. Merck have proposed a cutoff point for the acceptable level of covalent binding of 50 pmol drug equivalent/mg of protein (Evans *et al.*, 2004) (Fig. 18). Compounds that are lower than this value are advanced further. Compounds that exceed this level have to pass a number of other qualifying considerations before being considered for further development. These considerations include dose, availability of alternative therapies, whether the condition is life threatening, and whether structural modification is possible.

CYP-mediated bioactivation of drugs can lead to intermediates that are so reactive that they bind irreversibly to the active site of the enzyme resulting in loss of enzyme activity. This phenomenon is known as mechanism or time-dependent inhibition, which was discussed earlier in this chapter. Identifying if a compound is a mechanism-based inhibitor is also an important screening tool in the detection of reactive metabolites. Although reactive metabolite formation is an obvious risk factor, not all compounds identified in the screening tools described result in toxicity. Therefore, the approach adopted by Merck seems to be logical. Risks should be minimized at the early lead optimization stage by appropriate structural modification, and qualifying considerations should also be taken into account before deciding to stop further development of a compound.

## 9.4 The Potential Pitfalls of Extrapolating From *In Vitro* to *In Vivo* in Drug Clearance

Extrapolating from *in vitro* to *in vivo* can be problematic and error prone. The design and types of studies should be planned carefully before embarking on drug metabolism screening. Some of the potential pitfalls are described in the next section.

***9.4.1 Nonspecific Binding*** Nonspecific binding reduces the concentration of free drug available to either be metabolized by an enzyme or inhibit an enzyme. Nonspecific binding either to the incubation vessel or to the incubation constituents (e.g. microsomes) can alter the concentration of free drug. There are now several studies where knowledge of the extent of microsomal binding has led to a better understanding of the relationship between *in vitro* metabolism and *in vivo* pharmacokinetics. As well as studies investigating improvements in clearance prediction (Obach, 1997, 1999; Carlile *et al.*, 1999), it has been demonstrated that nonspecific microsomal binding can account for underestimation of inhibitor potency (i.e. overestimation of $IC_{50}$ or $K_i$ values) when dealing with lipophilic basic drugs (Tran *et al.*, 2002; Margolis and Obach, 2003), with the potential implication being an underestimation of risk from DDIs. The risks of not considering nonspecific microsomal binding in estimating inhibitory potency may be greater when dealing with mechanism-based inhibitors, due to the relatively high microsomal concentrations that are typically used during pre-incubation of the inhibitor in these experiments.

***9.4.2 Interindividual Variability*** Interindividual variability can exist both *in vivo* and in the *in vitro* metabolizing systems. The *in vivo* clearance can be influenced dramatically by CYP450 induction (e.g. smoking can induce CYP1A2), polymorphisms in metabolism, gender, age, and physiological conditions such as stress. For *in vitro* studies, pooled microsomes or pooled hepatocytes can reduce the problems associated with interindividual variability. However, this will mask the true extent of variability likely to take place *in vivo*.

***9.4.3 Metabolism in Extrahepatic Tissue*** *In vitro* hepatic metabolizing systems only address clearance in a single organ. In reality, although the liver is the major site of drug metabolism, many other tissues are capable of metabolizing drugs. If the extrahepatic metabolism is significant, the *in vitro* intrinsic clearance is likely to underpredict the *in vivo* clearance.

***9.4.4 Renal and Biliary Clearance*** As well as metabolic clearance, renal and biliary clearances contribute to total *in vivo* clearance. Presently, no routine widely accepted *in vitro* assays exist for biliary or renal clearance, which are appropriate for drug discovery, and as such the reliable prediction of total *in vivo* clearance is a primary unmet need in early ADMET/PK. In the future as a greater understanding of the biology of the processes is gained, it is possible that passive permeability data and transporter data can be used in combination with physiologically based PK modeling techniques to predict these processes.

***9.4.5 Equilibrium between Blood and Hepatocytes*** It is assumed that there is rapid equilibrium of drugs between the blood and hepatocytes. When the intrinsic clearance in hepatocytes is much greater than the efflux clearance from hepatocytes to blood, the *in vivo* intrinsic clearance is rate-limited by the influx process from blood to hepatocytes. Such an incorrect assumption of rapid equilibrium can be one of the reasons why *in vivo* intrinsic clearance is less than *in vitro* intrinsic clearance (Ito *et al.*, 1998).

### 9.5 Transporters in Hepatocytes

The subject of drug metabolism is made even more complex by the role of transporters in the distribution of drug into the hepatocyte and in the biliary excretion of the drug (Fig. 19). Interest in this area is growing rapidly. So far, the majority of studies have focused on P-gp, an efflux transporter present in the gastrointestinal tract, brain, kidney, and liver. More recently, interest has grown in other transporter families such as OATP (organic anion transporting polypeptides) and MRP (multidrug resistance associated proteins). These families of transporters among others play a major role in the efflux and uptake processes in the hepatocyte. A number of transporters also exist on the bile canalicular membrane and are involved in hepatobiliary excretion of drugs.

If a compound is a substrate of any of these transporters, the concentration of compound within the cell and in turn, at the site of the drug-metabolizing enzymes, may be influenced. Furthermore, if the compound is an inhibitor of a transporter, then this could affect the metabolism and elimination of a competing drug resulting in DDIs. Presently, it would appear that the specificity of drug transporters is less well characterized. The impact of this is twofold. Firstly, it may indicate the greater likelihood that multiple transporters may be involved in the membrane transport of a single drug, and so the contribution of the individual transporters to the overall net membrane transport needs to be taken into consideration when extrapolating from systems expressing a single transporter protein. Secondly, the absence of specific probe substrates and selective inhibitors may hinder the identification of the mechanisms of drug transport. Presently, the link between drug transporters and the prediction of pharmacokinetics is limited. By gaining a better insight into the value of *in vitro* transporter assays, it will

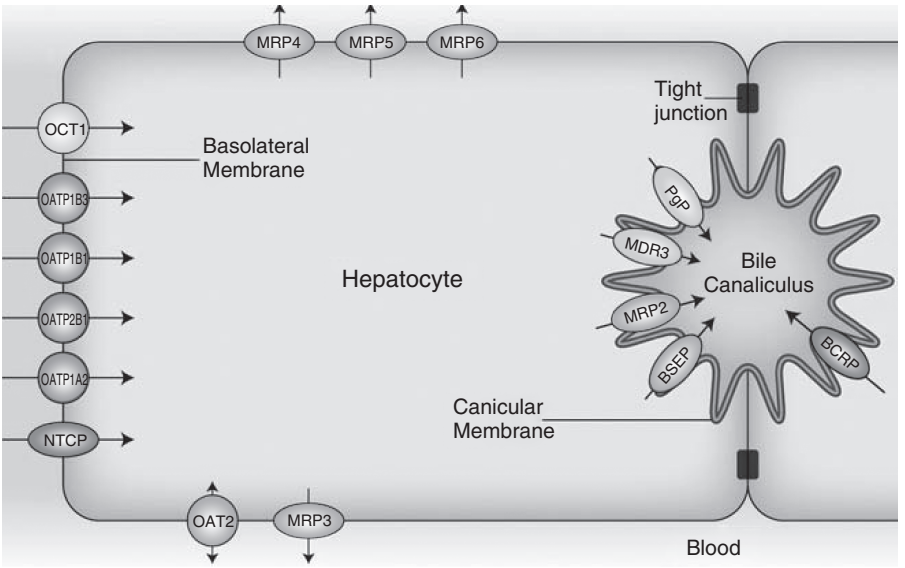


**Figure 19** Localization of efflux and uptake transporters in the hepatocyte (with permission from SOLVO Biotechnology, Budapest, Hungary).

enable *in vitro* screening methods to be more widely accepted and integrated into the panel of routine assays currently being assessed for ADMET. Incorporating the transporter information will enable plasma and tissue levels to be more accurately modeled, leading to an enhancement in the overall prediction of pharmacokinetics.

### 9.6 The Future of Drug Metabolism

As new experimental and analytical technologies become available on the market, both the speed and quantity of data generation will increase. For example, chip technology is now being introduced for recombinant CYP screening (Gilardi *et al.*, 2002; Lee *et al.*, 2005). Advances in the understanding of the complexity of drug metabolism and its interplay with other biological processes such as drug transporters will enable the accuracy of clearance prediction to increase. As this happens, there will be a greater reliance on pharmacokinetic modeling techniques enabling experimental data on individual parameters to be used in combination to predict the pharmacokinetics. More complex experimental models such as bioreactors are also being developed, which are closer to the physiology of the intact organ (Allen, Khetani and Bhatia, 2005; Sivaraman *et al.*, 2005; De Bartolo *et al.*, 2006). These systems can offer flow conditions and long-term culture of human hepatocytes. Recreation of the *in vivo* microarchitecture appears to retain liver-specific function more effectively (Hoffmaster, 2007). However, at present, throughput is restricted with these systems. Yet by combining assessment of metabolism with transporter information and potential toxicity evaluation in a more physiologically relevant system, it may be a smarter way of predicting drug disposition.

## 10 PLASMA PROTEIN BINDING

Distribution refers to the reversible transfer of drug from one location to another within the body. Delivery of drug to tissues by blood, ability to cross tissue membranes, binding within blood and tissues, and partitioning into fat can all affect distribution (Rowland and Tozer, 1995). The propensity of drugs to bind to plasma proteins is one of several key factors in determining distribution. A drug that is extensively and strongly bound to plasma proteins has limited access to sites of action in the cell, and metabolism and elimination may be slower. Plasma protein binding is typically a reversible interaction; however, irreversible or covalent binding can occasionally occur.

Human serum albumin and $\alpha_1$-acid glycoprotein (AAG) are the two main binding proteins in plasma with binding to other plasma proteins usually occurring to a much lesser extent. Generally, acidic drugs have a higher tendency to bind to albumin and basic drugs to AAG. Several factors can influence the unbound fraction in the plasma including drug concentration, binding affinity, and the number of binding sites. At low concentrations (i.e. below the dissociation constant), the fraction unbound is dependent on the concentration of binding sites and the affinity of the drug for the binding site. At high concentrations (i.e. above the dissociation constant), the fraction unbound is dependent on the number of binding sites and the drug concentration.

The levels of the individual proteins in plasma can fluctuate dramatically depending on age, disease, and pregnancy. Therefore, identifying which plasma protein the drug binds to and to what extent can be useful in understanding distribution in certain patient groups.

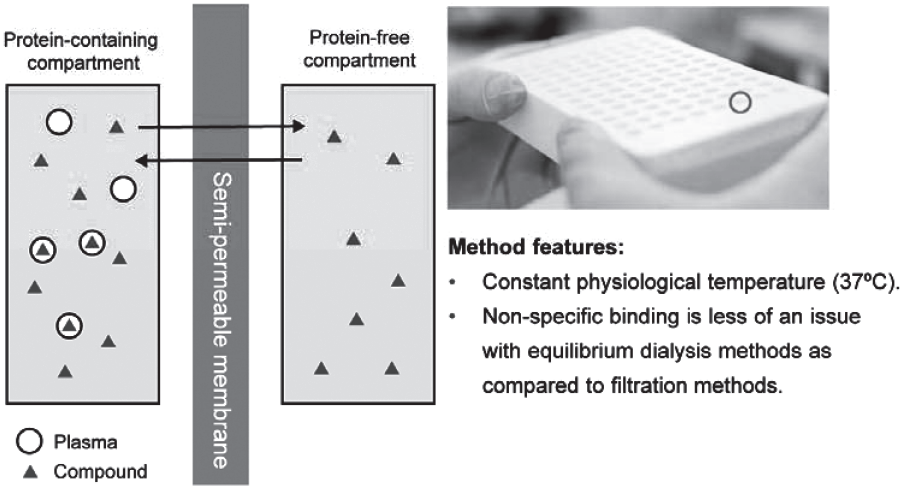


**Figure 20** Equilibrium dialysis is one of the most common methods used for determining plasma protein binding. With the availability of 96-well equilibrium dialysis plates, throughput is now improved dramatically over the traditional Dianorm systems. (From www.cyprotex.com.)

## 10.1 Methods for Determining Protein Binding

***10.1.1 Conventional Methods*** Although several higher-throughput methods (often based on investigating binding to a single plasma protein such as HSA) have been developed, none appear to offer the quality associated with equilibrium dialysis, and as such this remains the most popular method in industry to date (Fig. 20). The apparatus consists of two compartments containing buffer and plasma respectively separated by a semipermeable membrane that allows drug but not protein to pass across. Equilibrium dialysis has many advantages in that it can be used with whole plasma under a temperature-controlled environment. The apparatus is usually made from Teflon that reduces effect of binding. The traditional, and quite laborious, protocol for equilibrium dialysis using Dianorm cells has now largely been replaced with plate-based and automation friendly versions. Coupled with advances in LC-MS/MS, this has improved throughput dramatically.

Ultrafiltration is another popular technique for determining plasma protein binding. It is more rapid than the equilibrium dialysis method and filter plates are now commercially available. However, there are several perceived drawbacks to this method that include issues with binding to the filters and possible changes in binding as the concentration of drug in the plasma increases during centrifugation.

***10.1.2 Alternative Methods*** Advances in chromatographic technology have led to the development of highly automated rapid methods for determining binding to HSA. These methods use HPLC columns with immobilized protein attached. A paper by Valko *et al.* (2003) illustrates that data generated using a column-based method correlate relatively well with literature plasma protein binding measurements, and the method appears to be more accurate for highly bound compounds compared with

ultrafiltration methods. Even compounds with 99.2% and 99.3% can be distinguished and have retention times that differ by several minutes. An earlier paper by Talbert *et al.* (2002) describes how kinetic rate constants and equilibrium constants can be simultaneously determined using these immobilized protein columns by monitoring the retention time and peak width.

HSA immobilized on beads, microparticles, or sensor chips (Sjöholm *et al.*, 1979; Ahmad *et al.*, 2003; Schuhmacher *et al.*, 2004) can also be used to assess protein binding. The TRANSIL technology developed by Sovicell GmbH (formerly Nimbus Biotechnology) consists of either human serum albumin or AAG covalently immobilized on an inert surface. The beads are designed to minimize interactions with drug molecules, and thus reduce any nonspecific binding effects. The protein is randomly oriented on the bead surface, and no loss of function due to the covalent immobilization procedure is observed. The beads are separated from the free fraction at the end of the incubation period by filtration or centrifugation.

## 11 QC/QA CONSIDERATIONS

This review of automation in modern ADMET has provided a brief survey of the variety of assays *in vitro* that are amenable to automation and that are deployed by all modern pharma to assess the pharmacokinetic profiles of nascent drug leads. Some pharma have established highly organized and fully automated ADMET profiling laboratories. There are now contract research organizations (CROs), such as Cyprotex, NovaScreen, PanLabs, CEREP, and others, which have made available panels of semi- and fully automated ADMET assays that are run in parallel from the same stock solutions of compounds. When one set of compounds is submitted to a centralized lab for such cross-profiling, the results from one assay to another are more likely comparable to one another without major confounds that could arise due to, for example, varying compound solubility as a result of the pH varying across the different assays.

Apart from the intrinsic assay issues, there are "extrinsic" issues that should be considered whenever comparing data from different labs in different assays. As shown in Fig. 9, small nuances in SOPs can have large impacts on variables such as mixing efficiency in an assay. Some labs may control better than others the potential "edge effects" due to more rapid evaporation of outer wells of a microtiter plate than those toward the middle of the plate, which may cause precipitation of compounds due to changes in drug concentration, or pH, or ionic strength. For these reasons, results should always be compared with a reference compound, as has been mentioned earlier (as in Fig. 13).

Furthermore, there is always an inherent presumption that the designated structure for a compound powder in a vial is correct, and that the presumed structure is in fact the entity that elicits the primary response in the primary screening assay for determining potency and efficacy (pharmacodynamics), as well as that the same structure mediates the responses from ADMET cross-profiling assays—particularly those that may involve toxicity. This, however, is an assumption until proven, and proving this assumption is not always straightforward.

For example, the standard acceptance criteria for compound purity vary from site to site, and even within a site. Within a site, acceptance criteria for registered (accepted into the corporate compound library) may be as follows; however, the purity targets could vary ±10% between sites:

1. A compound synthesized in-house by discrete organic synthetic methods (i.e. handmade compounds by chemists in-house or by contract research) may have the requirement of being >95% pure as assessed by LC-MS and with structure properly assigned based on NMR.
2. Compounds made by combinatorial synthesis may require only 90% purity by only LC-MS.
3. Compound purchased for HTS may require only 80% purity, and it is not clear whether the analytical data are valid.

All these compound types from the different sources, which may range in age from weeks to years, and which have been stored under various controlled or uncontrolled (ambient) conditions, are found in a "typical" HTS compound library. The "same" compound may exist in both solution and powder forms, and as different stock concentrations and multiple lots, respectively. Different lots of the "same" compound solution may have been freeze-thawed several times and the DMSO solvent may contain varying amounts of water, which not only is known to affect compound solubility (Cheng *et al.*, 2003) but may also affect compound stability.

It should quickly become apparent to the reader that without rigorous compound (lot) management, comparison of data between different assays run at different times may be problematic, irrespective of whether the assays are pharmacodynamic or pharmacokinetic in nature. Figure 15 depicts an approach that may be adopted to assist QC/QA processes using automated rules to mimic checking procedures typically employed by a scientist.

## 12 SUMMARY

There is an overall drive in the pharmaceutical industry to improve efficiency and reduce operating costs. Sophisticated instruments are now widely available for liquid handling and sample analysis, which enable rapid screening of large numbers of compounds. As a consequence, the bottleneck has now shifted to the stage of data and information handling which is traditionally a resource-intensive process. Intelligent workflow techniques play a major role in addressing these challenges. By automating these processes, it allows more efficient use of internal resources, enables knowledge to be captured from individuals and shared, and reduces subjective human intervention. Ultimately this leads to less errors, better consistency, reduced timelines, and lower costs. It appears that a greater reliance on automated *in vitro* methods earlier in the pharmaceutical R&D continuum is having an impact on improving the failure rate of new drugs (see Fig. 21).

Over the next decade, the pharmaceutical industry will continue to radically change its approach to address the lack of efficiency in its current working practice. Implementation of *in silico* techniques, quicker and better screening techniques, and smarter use of experimental data will hopefully move a step forward in achieving this goal. The move toward identifying efficacious, safer drugs with optimal pharmacokinetic parameters at an early stage will continue, enabling efficiencies both in terms of time and money, and ultimately allowing more effective drugs to reach the market quicker.

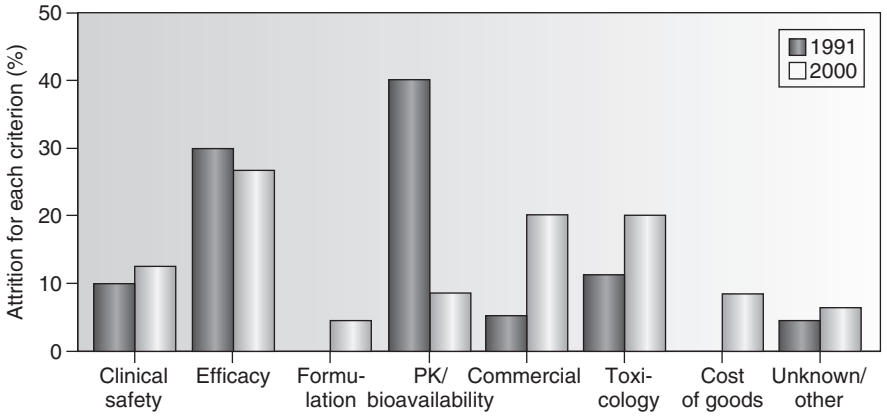


**Figure 21** Drug attrition in drug development in 1991 and 2000 (from Kola and Landis, 2004). The text discusses the reasons for the recent progress in ADMET due to better and earlier filtering from the development queue—both experimentally and computationally—those compounds that display poor bioavailability.

Smarter decision making and interpretation of the data will be required with ADMET properties considered to be whole rather than individual properties. Integrating multiple ADMET properties with physiochemical data to predict whole body pharmacokinetics using physiologically based pharmacokinetic modeling will continue to grow in popularity as models become more accurate (Lüpfert and Reichel, 2005; De Buck *et al.*, 2007).

## REFERENCES

Ahmad, A., Ramakrishnan, A., McLean, M.A., Breau, A.P. (2003). Use of surface plasmon resonance biosensor technology as a possible alternative to detect differences in binding of enantiomeric drug compounds to immobilized albumins. *Biosens. Bioelectron.*, *18* (4), 399–404.

Allen, J.W., Khetani, S.R., Bhatia, S.N. (2005). *In vitro* zonation and toxicity in a hepatocyte bioreactor. *Toxicol. Sci.*, *84* (1), 110–119.

Avdeef, A. (2007). Solubility of sparingly-soluble drugs. *Adv. Drug Deliv. Rev.*, *59* (7), 568–590.

Avdeef, A., Bendels, S., Di, L., Faller, B., Kansy, M., Sugano, K., Yamauchi, Y. (2007). PAMPA–critical factors for better predictions of absorption. *J. Pharm. Sci.*, *96* (11), 2893–2909.

Banno, K., Takahashi, R. (1991). Laboratory automation system for the analysis of drugs in biological fluids. *Anal. Sci.*, *7* (3), 511–513.

Belz, G.G., Doering, W., Munkes, R. *et al.* (1983). Interaction between digoxin and calcium antagonists and antiarrhythmic drugs. *Clin. Pharmacol. Ther.*, *33*, 410–417.

Bevan, C.D., Lloyd, R.S. (2000). A high-throughput screening method for the determination of aqueous drug solubility using laser nephelometry in microtiter plates. *Anal. Chem.*, *72* (8), 1781–1787.

Bigger, J.T., Jr, Leahey, E.B., Jr (1982). Quinidine and digoxin: an important interaction. *Drugs*, *24*, 229–239.

Boyd, B.J. (2008). Past and future evolution in colloidal drug delivery systems. *Expert Opin. Drug Deliv.*, *5* (1), 69–85.

Burch, R.M. (1993). Mass ligand-binding screening strategies for identification of leads for new drug discovery. *NIDA Res. Monogr.*, *134*, 37–45.

Buxton, I. (2006). Pharmacokinetics and pharmacodynamics: The dynamics of drug absorption, distribution, action, and elimination. In Brunton, L. *et al.* (eds.), *Goodman & Gilman's The Pharmacological Basis of Therapeutics*, 11th ed. McGraw-Hill, pp. 1–137.

Buxton, I. (2006). Pharmacokinetics and pharmacodynamics: The dynamics of drug absorption, distribution, action, and elimination In . *NIDA Res. Monogr.*, *134*, 37–45.

Carlile, D.J., Hakooz, N., Bayliss, M.K., Houston, J.B. (1999). Microsomal prediction of *in vivo* clearance of CYP2C9 substrates in humans. *Br. J. Clin. Pharmacol.*, *47*, 625–635.

Cheng, X., Hochlowski, J., Tang, H., Hepp, D., Beckner, C., Kantor, S., Schmitt, R. (2003). Studies on repository compound stability in DMSO under various conditions. *J. Biomol. Screen.*, *8* (3), 292–304.

De Bartolo, L., Salerno, S., Morelli, S., Giorno, L., Rende, M., Memoli, B., Procino, A., Andreucci, V.E., Bader, A., Drioli, E. (2006). Long-term maintenance of human hepatocytes in oxygen-permeable membrane bioreactor. *Biomaterials*, *27* (27), 4794–4803.

De Buck, S.S., Sinha, V.K., Fenu, L.A., Nijsen, M.J., Mackie, C.E., Gilissen, R.A.H.J. (2007). Prediction of human pharmacokinetics using physiologically based modelling: a retrospective analysis of 26 clinically tested drugs. *Drug Metab. Dispos.*, *35* (10), 1766–1780.

Doering, W. (1979). Quinidine-digoxin interaction: pharmacokinetics, underlying mechanism and clinical implications. *N. Engl. J. Med.*, *301*, 400–404.

Evans, D.C., Watt, A.P., Nicoll-Griffith, D.A., Baillie, T.A. (2004). Drug-protein adducts: an industry perspective on minimizing the potential for drug bioactivation in drug discovery and development. *Chem. Res. Toxicol.*, *17* (1), 3–16.

Fisher, M.B., Campanale, K., Ackermann, B.L., VandenBranden, M., Wrighton, S.A. (2000). *In vitro* glucuronidation using human liver microsomes and the pore-forming peptide alamethicin. *Drug Metab. Dispos.*, *28* (5), 560–566.

Fitzpatrick, L.A., Good, A. (2000). Micronized progesterone: clinical indications and comparison with current treatments. *Fertil. Steril.*, *73* (3), 654–754.

Galinis-Luciani, D., Nguyen, L., Yazdanian, M. (2007). Is PAMPA a useful tool for discovery? *J. Pharm. Sci.*, *96* (11), 2886–2892.

Garrigues, A., Nugier, J., Orlowski, S., Ezan, E. (2002). A high-throughput screening microplate test for the interaction of drugs with P-glycoprotein. *Anal. Biochem.*, *305* (1), 106–114.

Gentile, A., Trusolino, L., Comoglio, P.M. (2008). The Met tyrosine kinase receptor in development and cancer. *Cancer Metastasis Rev.*, *27* (1), 85–94.

Gilardi, G., Meharenna, Y.T., Tsotsou, G.E., Sadeghi, S.J., Fairhead, M., Giannini, S. (2002). Molecular Lego: design of molecular assemblies of P450 enzymes for nanobiotechnology. *Biosens. Bioelectron.*, *17* (1–2), 133–145.

Glavinas, H., Kis, E., Pál, A., Kovács, R., Jani, M., Vági, E., Molnár, E., Bánsághi, S., Kele, Z., Janáky, T., Báthori, G., Von Richter, O., Koomen, G.J., Krajcsi, P. (2007). ABCG2 (breast cancer resistance protein/mitoxantrone resistance-associated protein) ATPase assay: a useful tool to detect drug-transporter interactions. *Drug Metab. Dispos.*, *35* (9), 1533–1542.

Goldstein, D.M., Gray, N.S., Zarrinkar, P.P. (2008). High-throughput kinase profiling as a platform for drug discovery. *Nat. Rev. Drug Discov.*, *7* (5), 391–397.

Harris, S.R., Garlick, R.K., Miller, J.J., Jr, Harney, H.N., Monroe, P.J. (1991). Complement C5a receptor assay for high throughput screening. *J. Recept. Res.*, *11* (1–4), 115–128.

Hewitt, N.J., Gómez Lechón, M.J., Houston, J.B., Hallifax, D., Brown, H.S., Maurel, P., Kenna, J. G., Gustaysson, L., Lohmann, C., Skonberg, C., Guillouzo, A., Tuschl, G., Li, A.P., LeCluyse, E. (2007). Primary hepatocytes: current understanding of the regulation of metabolic enzymes

and transporter proteins, and pharmaceutical practice for the use of hepatocytes in metabolism, enzyme induction, transporter, clearance, and hepatotoxicity studies. *Drug Metab. Rev.*, *39*, 159–234.

Hoffmaster, K.A. (2007). *Presentation: AAPS Workshop on Drug Transporters in ADME: from the Bench to the Bedside, March 5–7*.

Ito, K., Iwatsubo, T., Kanamitsu, S., Nakajima, Y., Sugiyama, Y. (1998). Quantitative prediction of *in vivo* drug clearance and drug interactions from *in vitro* data on metabolism together with binding and transport. *Annu. Rev. Pharmacol. Toxicol.*, *38*, 461–499.

Kearns, D.H. (2001). High throughput physicochemical profiling for drug discovery. *J. Pharm. Sci.*, *90* (11), 1838–1858.

Klein, H.O., Lang, R., Weiss, E., Di Segni, E., Libhaber, C., Guerrero, J., Kaplinsky, E. (1982). The influence of verapamil on serum digoxin concentration. *Circulation*, *65*, 998–1003.

Kola, I., Landis, J. (2004). Can the pharmaceutical industry reduce attrition rates? *Nat. Rev. Drug Discov.*, *3* (8), 711–771.

Krippendorff, B.F., Lienau, P., Reichel, A., Huisinga, W. (2007). Optimizing classification of drug-drug interaction potential for CYP450 isoenzyme inhibition assays in early drug discovery. *J. Biomol. Screen.*, *12* (1), 92–99.

Leahey, E.B., Jr, Reiffel, J.A., Heissenbuttel, R.H. *et al.* (1979). Enhanced cardiac effect of digoxin during quinidine treatment. *Arch. Intern. Med.*, *139*, 519–521.

Lee, M.Y., Park, C.B., Dordick, J.S., Clark, D.S. (2005). Metabolizing enzyme toxicology assay chip (MetaChip) for high-throughput microscale toxicity analyses. *Proc. Natl. Acad. Sci. U.S.A.*, *102* (4), 983–987.

Lett, E., Kriszt, W., de Sandro, V., Ducrotov, G., Richert, L. (1992). Optimal detergent activation of rat liver microsomal UDP-glucuronosyl transferases toward morphine and 1-naphthol: contribution to induction and latency studies. *Biochem. Pharmacol.*, *43* (7), 1649–1653.

Lipinski, C.A. (2000). Drug-like properties and the causes of poor solubility and poor permeability. *Pharmacol. Toxicol. Met.*, *44*, 235–249.

Lipinski, C.A., Lombardo, F., Dominy, B.W., Feeney, P.J. (1997). Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug. Del. Rev.*, *23*, 3–25.

Lüpfert, C., Reichel, A. (2005). Development and application of physiologically based pharmacokinetic-modeling tools to support drug discovery. *Chem. Biodivers.*, *2* (11), 1462–1486.

McGee, P. (2005). Tracking drug safety from the ground up. *Drug Discov. Devel.*, 20–28.

Margolis, J.M., Obach, R.S. (2003). Impact of nonspecific binding to microsomes and phospholipid on the inhibition of cytochrome P4502D6: implications for relating *in vitro* inhibition data to *in vivo* drug interactions. *Drug Metab. Dispos.*, *31* (5), 606–611.

Masubuchi, N., Makino, C., Murayama, N. (2007). Prediction of *in vivo* potential for metabolic activation of drugs into chemically reactive intermediate: correlation of *in vitro* and *in vivo* generation of reactive intermediates and *in vitro* glutathione conjugate formation in rats and humans. *Chem. Res. Toxicol.*, *20* (3), 455–464.

Maxwell, D.L., Gilmour-White, S.K., Hall, M.R. (1989). Digoxin toxicity due to interaction of digoxin with erythromycin. *BMJ*, *298*, 572.

Meech, R., Mackenzie, P.I. (1997). Structure and function of uridine diphosphate glucuronosyltransferases. *Clin. Exp. Pharmacol. Physiol.*, *24* (12), 907–915.

Neubert, R.H.H., Schwarz, M.A., Mrestani, Y., Plätzer, M., Raith, K. (1999). Affinity capillary electrophoresis in pharmaceutics. *Pharm. Res.*, *16*, 1663–1673.

Ng, R. (2004). *Drugs: From Discovery to Approval*, John Wiley & Sons, Inc.

Obach, R.S. (1997). Nonspecific binding to microsomes: impact on scale-up of in vitro intrinsic clearance to hepatic clearance as assessed through examination of warfarin, imipramine, and propranolol. *Drug Metab. Dispos.*, *25* (12), 1359–1369.

Obach, R.S. (1999). Prediction of human clearance of twenty-nine drugs from hepatic microsomal intrinsic clearance data: an examination of in vitro half-life approach and nonspecific binding to microsomes. *Drug Metab. Dispos.*, *27* (11), 1350–1359.

Pereira, D.A., Williams, J.A. (2007). Origin and evolution of high throughput screening. *Br. J. Pharmacol.*, *152*, 53–61.

Popa-Burke, I.G., Issakova, O., Arroway, J.D., Bernasconi, P., Chen, M., Coudurier, L., Galasinski, S., Jadhav, A.P., Janzen, W.P., Lagasca, D., Liu, D., Lewis, R.S., Mohney, R.P., Sepetov, N., Sparkman, D.A., Hodge, C.N. (2004). Streamlined system for purifying and quantifying a diverse library of compounds and the effect of compound concentration measurements on the accurate interpretation of biological assay results. *Anal. Chem.*, *76* (24), 7278–7287.

Rowland, M., Tozer, T. (1995). *Clinical Pharmacokinetics, Concepts and Applications*, 3rd edn, Lippincott Williams & Wilkins.

Roymans, D., Annaert, P., Van Houdt, J., Weygers, A., Noukens, J., Sensenhauser, C., Silva, J., Van Looveren, C., Hendrickx, J., Mannens, G., Meuldermans, W. (2005). Expression and induction potential of cytochromes P450 in human cryopreserved hepatocytes. *Drug Metab. Dispos.*, *33* (7), 1004–1016.

Schuhmacher, J., Kohlsdorfer, C., Bühner, K., Brandenburger, T., Kruk, R. (2004). High-throughput determination of the free fraction of drugs strongly bound to plasma proteins. *J. Pharm. Sci.*, *93* (4), 816–830.

Sivaraman, A., Leach, J.K., Townsend, S., Iida, T., Hogan, B.J., Stolz, D.B., Fry, R., Samson, L.D., Tannenbaum, S.R., Griffith, L.G. (2005). A microscale *in vitro* physiological model of the liver: predictive screens for drug metabolism and enzyme induction. *Curr. Drug Metab.*, *6* (6), 569–591.

Sjöholm, I., Ekman, B., Kober, A., Ljungstedt-Påhlman, I., Seiving, B., Sjödin, T. (1979). Binding of drugs to human serum albumin: XI. The specificity of three binding sites as studied with albumin immobilized in microparticles. *Mol. Pharmacol.*, *16*, 767–777.

Sneader, W. (2005). *Drug Discovery: A History*, John Wiley & Sons, Inc.

Soglia, J.R., Harriman, S.P., Zhao, S., Barberia, J., Cole, M.J., Boyd, J.G., Contillio, L.G. (2004). The development of a higher throughput reactive intermediate screening assay incorporating micro-bore liquid chromatography-micro-electrospray ionization-tandem mass spectrometry and glutathione ethyl ester as an in vitro conjugating agent. *J. Pharm. Biomed. Anal.*, *36*, 105–116.

Stresser, D.M., Blanchard, A.P., Turner, S.D., Erve, J.C., Dandeneau, A.A., Miller, V.P., Crespi, C. L. (2000). Substrate dependent modulation of CYP3A4 catalytic activity:analysis of 27 test compounds with four fluorometric substrates. *Drug Metab. Dispos.*, *28*, 1440–1448.

Talbert, A.M., Tranter, G.E., Holmes, E., Francis, P.L. (2002). Determination of drug-plasma protein binding kinetics and equilibria by chromatographic profiling: exemplification of the method using L-tryptophan and albumin. *Anal. Chem.*, *74*, 446–452.

Tran, T.H., Von Moltke, L.L., Venkatakrishnan, K., Granda, B.W., Gibbs, M.A., Obach, R.S., Harmatz, J.S., Greenblatt, D.J. (2002). Microsomal protein concentration modifies the apparent inhibitory potency of CYP3A inhibitors. *Drug Metab. Dispos.*, *30* (12), 1441–1445.

Unger, S.H., Cook, J.R., Hollenberg, J.S. (1978). Simple procedure for determining octanol–aqueous partition, distribution, and ionization coefficients by reversed-phase high-pressure liquid chromatography. *J. Pharm. Sci.*, *67* (10), 1364–1367.

Valko, K., Nunhuck, S., Bevan, C., Abraham, M.H., Reynolds, D.P. (2003). Fast gradient HPLC method to determine compounds binding to human serum albumin. Relationships with octanol/water and immobilized artificial membrane lipophilicity. *J. Pharm. Sci.*, *92* (11), 2236–2248.

Verschraagen, M., Koks, C.H., Schellens, J.H., Beijnen, J.H. (1999). P-glycoprotein system as a determinant of drug interactions: the case of digoxin-verapamil. *Pharmacol. Res.*, *40*, 301–306.

Wakasugi, H., Yano, I., Ito, T., Hashida, T., Futami, T., Nohara, R., Sasayama, S., Inui, K. (1998). Effect of clarithromycin on renal excretion of digoxin: interaction with P-glycoprotein. *Clin. Pharmacol. Ther.*, *64*, 123–128.

Walgren, J.L., Mitchell, M.D., Thompson, D.C. (2005). Role of metabolism in drug-induced idiosyncratic hepatotoxicity. *Crit. Rev. Toxicol.*, *35*, 325–361.

Westphal, K., Weinbrenner, A., Giessmann, T., Stuhr, M., Franke, G., Zschiesche, M., Oertel, R., Terhaag, B., Kroemer, H.K., Siegmund, W. (2000). Oral bioavailability of digoxin is enhanced by talinolol: evidence for involvement of intestinal P-glycoprotein. *Clin. Pharmacol. Ther.*, *68*, 6–12.

Wilkinson, G.R., Shand, D.G. (1975). Commentary: a physiological approach to hepatic drug clearance. *Clin. Pharm. Ther.*, *18*, 377–390.

Yan, Z., Maher, N., Torres, R., Caldwell, G.W., Huebert, N. (2005). Rapid detection and characterization of minor reactive metabolites using stable-isotope trapping in combination with tandem mass spectrometry. *Rapid Commun. Mass Spectrom.*, *19* (22), 3322–3330.

Zhou, S., Yung Chan, S., Cher Goh, B., Chan, E., Duan, W., Huang, M., McLeod, H.L. (2005). Mechanism-based inhibition of cytochrome P450 3A4 by therapeutic drugs. *Clin. Pharmacokinet.*, *44* (3), 279–304.

# Mass Spectrometry

THOMAS R. SHARP

## 1 INTRODUCTION

> "It is a fundamental tenet of chemistry that the structural formula of any compound contains coded within it all that compound's chemical, physical and biological properties." (Katritzky, Karelson and Lobanov, 1997)

If we accept this statement as philosophical truth, the determination of the structure of organic molecules is of paramount interest. Determining the structure will permit determining all other properties of the molecule. Therefore, methods for determining structure are important. Mass spectrometry has become one of the premier methods for determining organic structure, and is generally the first line of experimentation done (if not the only experimentation necessary) for determining the structure of organic molecules. Mass spectrometry, coupled especially with liquid chromatography (normal or reverse phase), is a foundation experimental technique in any structure elucidation effort, and in drug metabolism in particular. The combination of high-resolution chromatography, for handling complex mixtures, and the structural specificity and sensitivity for detecting low-level components of mass spectrometry—both in low relative abundance in a mixture and low absolute amounts of analyte—is unsurpassable for tackling such problems. Sometimes it is sufficient to solve the problem. Always it is necessary to have done the experiments to provide solid foundation for subsequent required work—by isolation, NMR, synthesis, and so on—to finally arrive at the answer.

> "Why learn anything about spectral interpretation when the computer can do all the work? The answer to this question is simple, as most conscientious users quickly realize. The library search often does not provide a realistic answer or (worse) may provide an answer that looks correct but is not. Even software programs that profess to 'interpret' unknown spectra can only provide probable answers. After that, you are left to your own devices." (Smith, 2004)

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

Much of the literature focuses on the instrumentation and techniques for how to generate mass spectral data. There is a substantial body of older literature dealing with interpretation of mass spectra and cataloging of fragmentation patterns. This older body of literature is primarily concerned, however, with electron ionization (EI) mass spectra. Even the more recently published textbooks and treatises, discussing the benefits of the newer soft ionization techniques,[1] do not discuss in much depth how to interpret mass spectra generated by these techniques. The statement that soft ionization techniques generate protonated molecular ions is often all that is seen. Little else is offered. The primary purpose in composing this chapter is to discuss mass spectra generated by the soft ionization methods—electrospray, in the main—and phenomena that often appear in these spectra, and to understand what these phenomena mean.

We will discuss instrumentation briefly here insofar as to provide a reference point for our subsequent discussions of data, and to provide entry into the vast amount of literature on instrumentation and instrumental advancements. Starting with spectral interpretation, we will describe the EI literature for completeness, so that the reader has a connection with this literature. Some of the numerological and interpretational rules are directly transferable from the EI literature to the interpretation of soft ionization spectra. Examining EI-induced fragmentation patterns is instructive to a degree, especially when considering the higher-energy regimes invoked when generating fragmentation of soft ionization-produced molecular ions using tandem mass spectrometry.

## 2 A BRIEF HISTORY

> "… If there is a mortal sin…, it is the teaching of science without connecting to the history of science…. The lack of historical perspective was dramatically demonstrated… by the student who asked if Galileo and Einstein ever met (Klose, 1987)."

Mass spectrometry, as a distinct discipline, is approximately 100 years old. Grayson (2002) has compiled an excellent history covering all aspects of the field of mass spectrometry for the fiftieth anniversary of the annual meetings of the American Society for Mass Spectrometry. We will make a few arbitrarily selected historical comments here. The mass spectrometry community has made significant efforts in the last 20 years to record and preserve the history of the field directly from some of the early workers in the field before that very personal perspective is lost. Much of this accumulated history has been published. Some oral histories with surviving pioneers in the field have been recorded and deposited at the Chemical Heritage Foundation.[2]

Most practitioners of the precise art and subtle science of mass spectrometry acknowledge the field to have originated with the work of J.J. Thomson and associates (Griffiths, 1997), published in 1910–1912, using the parabola mass spectrograph. Seminal

[1] The soft ionization methods, which will be discussed later, most often produce a molecular ion in which a charge-carrying species is attached to the neutral molecule. Typically, an $H^+$ is the attaching species. Many structural classes of compounds, however, show the strong tendency to scavenge and attach monovalent cations, such as $Na^+$.

[2] Chemical Heritage Foundation, 315 Chestnut Street, Philadelphia, Pennsylvania 19106. http://www.chemheritage.org.

discoveries that he and his coworkers made include the fact that the elements could be polyisotopic, by discovering the isotopes of neon. Thomson was awarded the 1906 Nobel Prize in physics (http://nobelprize.org/physics/laureates/1906/index.html) for his work on "investigations on the conduction of electricity by gases," work not directly involved with mass spectrometry, but which helped lay the foundations for it. Thomson's direct involvement in early mass spectrometry ended (Svec, 1985) with the publication of the parabola mass spectrograph work in 1913. Francis Aston (Squires, 1998; Downard, 2007a, b), a student of Thomson's, continued the work, is attributed as coining the term "mass spectrometry," and advanced the field significantly. Aston was awarded the 1922 Nobel Prize in chemistry (http://nobelprize.org/chemistry/laureates/1922/index.html) for his discoveries of a large number of the stable isotopes of the elements using mass spectrometry. Aston and A.J. Demster continued to determine the isotopic composition of the elements until finishing in 1935 with the publication of the isotopic composition of platinum and iridium. Aston is attributed as stating, in a presentation given not long after this 1935 publication, that "mass spectroscopy had served its purpose and would die away as a field for research" (Svec, 1985).

An unfortunate but seminal feat that mass spectrometry accomplished was the isolation of sufficient quantities of the fissile isotope of uranium, $^{235}U$, to produce the first uranium-based fission atomic bombs during World War II. Numerous instrumental and other innovations derived from the efforts in the Manhattan Project. Physical separation of the uranium isotopes was accomplished using a large production array of mass spectrometers. Each individual machine was called a calutron (Nier, 1989; Yergey and Yergey, 1997; Settle, 2002). Gaseous diffusion was used to provide initial enrichment of feedstock (in the form of uranium hexafluoride) to the calutrons. Final enrichment was done mass spectrometrically. Gaseous diffusion did not replace mass spectrometric enrichment until after World War II had ended. The name "calutron" was coined by Ernest Lawrence. The physical principles of magnetic separation were based on Lawrence and other's work building and using cyclotrons for subatomic particle acceleration in laboratories at the University of California—Berkeley (now the Lawrence-Berkeley Laboratories).

The first industrial interest—in particular, chemical applications—of this heretofore physics laboratory oddity started in the petroleum industry in the 1950s (Meyerson, 1986; Quayle, 1987). Application of mass spectrometry to the structure determination of organic compounds, and quantitation of components of mixtures started with the faith that the petroleum industry placed in mass spectrometry during and after World War II. Mass spectrometry seemed an ideal technique to address the detection and determination of gases and volatile components of petroleum distillates from refinery streams. This faith and interest is directly responsible for creation of the first commercially available mass spectrometer systems, produced by Consolidated Engineering Corporation (CEC) in the United States (Meyerson, 1986) and Metropolitan-Vickers (later to become Associated Electrical Industries, AEI) in the United Kingdom (Quayle, 1987). An outpost of mass spectrometry investigation also grew up in Australia, influenced initially by encouragement and a visit from Aston (Morrison, 1991; Downard and de Laeter, 2005).

The importance of mass spectrometry in planetary and interplanetary research to date has been summarized by Neir (1985) and Grayson (2002). Mass spectrometers have been key pieces of instrumentation to ride atmospheric and interplanetary probes. Being well adapted to analysis of gaseous samples, they are ideal for studying the

atmospheres of the earth and other planets, recording information on the chemical composition of those atmospheres. The Viking lander missions to Mars illustrate this application well by the examination of the composition of the Martian atmosphere (McElroy, Yung and Nier, 1976) and the search for signs of life on Mars (Biemann *et al.*, 1976).

Forensic application of mass spectrometry has extended into the arena of professional sports in addition to applications in crime detection and resolution. Recognition of gas chromatography–mass spectrometry (GC-MS) as an important tool in this area was made in 1967. The 1968 Winter Olympics saw the first time in which contestants were monitored for a few stimulants and narcotics (Grayson, 2002). Application to the horse racing industry followed a number of years later. The efforts continue to this day in sports monitoring and horse racing. The drug designers have become more sophisticated, and efforts to discover what they create must keep pace. An example of a relatively recent new performance-enhancing drug was reported by Catlin and associates (Catlin *et al.*, 2004).

Expansion of the utility of mass spectrometry to large and in particular biologically relevant molecules has been a direct result of the desorption ionization techniques—field desorption, fast atom bombardment (FAB), matrix-assisted laser desorption—and the atmospheric pressure ionization (API) techniques—atmospheric pressure chemical ionization (APCI) and electrospray (ESI). The latter are particularly important in achieving the interface between condensed-phase chromatographic techniques—HPLC and supercritical fluid chromatography—and the mass spectrometer. The importance of electrospray ionization, in particular, has been recognized by the 2002 awarding of the Nobel Prize (Fenn, 2003) for chemistry to John Fenn and Koichi Tanaka. Fenn and Tanaka, along with Kurt Wüthrich,[3] were recognized for their development of methods for identification and structure analysis of biological molecules.

## 3 GENERAL BACKGROUND IN MASS SPECTROMETRY

Understanding the powers that various chromatography–mass spectrometry techniques have presupposes that one understands the information that mass spectrometry in its various forms and nuances can provide. We will cursorily survey the field of instrumentation to provide a minimum reference here, with entry points into the more extensive detailed literature. The detailed focus will be on the interpretation of data and experiments.

Discussions of some of the basic mass spectrometric principles in the following will reinforce previous awareness. Should the reader's curiosity stray beyond these bounds, we invite consultation with the literature, and provide entry points to those topics in the following. Recent reasonably comprehensive textbooks on the general topic of mass spectrometry have been published by de Hoffmann and Stroobant (2002), Gross (2004), and Watson and Sparkman (2007). A treatise on the general subject of mass spectrometry, slightly more up-to-date than the textbooks, has been published by

[3] http://nobelprize.org/chemistry/laureates/2002/index.html. Fenn's and Tanaka's work applied electrospray to the mass spectrometric examination of large molecules. Wüthrich's contributions were in application of NMR to this area.

Burinsky (2006). A description of the state of the art on combined chromatography–mass spectrometry (including a discussion on LC-NMR) has been published by Sharp and Marquez (2006). Volume 101 (2001), issue 2 of *Chemical Reviews* covers many of the current active areas of development in mass spectrometry.

There is no one miracle technique which will solve all problems. Specialized techniques address specialized problems. It is one's responsibility as a good scientist to know the breadth of applicability of a technique, what it is good for—and what it is NOT good for—and to use what is needed when it is needed. Do not fall into the trap—if one's favorite tool is a hammer, then everything looks like a nail—of attempting to solve all problems with one's favorite technique.

### 3.1 The Mass Spectrometry Literature

The mass spectrometry research literature is scattered through many journals and publications. *The Journal of the American Chemical Society* and *Analytical Chemistry* have been consistent contributors. In recent years, several journals specializing in fundamentals and applications of mass spectrometry have come and gone. *The International Journal of Mass Spectrometry and Ion Processes* began publication in 1968 and has gone through an evolution in its title, being formerly known as *The International Journal of Mass Spectrometry and Ion Physics*. The title has been further simplified to *The International Journal of Mass Spectrometry. Organic Mass Spectrometry* started in 1968, specializing in organic chemical applications of mass spectrometry. *Biomedical Mass Spectrometry* (later changing to *Biomedical and Environmental Mass Spectrometry*, then to *Biological Mass Spectrometry*) started in 1986 as a companion journal specializing in the biomedical and environmental application areas, as the title implies. These two journals combined in 1995 to become *The Journal of Mass Spectrometry*, currently being published. *Mass Spectrometry Reviews* started in 1982. *Rapid Communications in Mass Spectrometry* started in 1987 as a means of providing (as the title implies) rapid dissemination of new results. *The Journal of the American Society for Mass Spectrometry* started in 1990 as the official archive of the American Society for Mass Spectrometry. *European Journal of Mass Spectrometry* is another specialty journal, published since 1995, containing fundamental and applied work. Specific application areas are also located in the specialty journals for those particular areas.

Two primary mass spectrometry conferences occur on a regular basis. The International Mass Spectrometry Conference is held every three years, and is typically held in a European city (Hamming and Foster, 1972; Quayle, 1987). The proceedings of this conference, in the form of extended abstracts of presented papers, are often published in book form as the series, titled *Advances in Mass Spectrometry*. Proceedings have been published occasionally as a special volume of the *International Journal of Mass Spectrometry and Ion Processes*. The annual conference of the American Society for Mass Spectrometry rotates among numerous cities on the North American continent (Hamming and Foster, 1972). The Society originated as the American Society for Testing and Materials Committee E-14 on mass spectrometry, starting with an annual meeting in 1953. The Society was formally organized in 1969. Extended abstracts of papers presented at this meeting are also collected and distributed, but do not constitute official publications, as they are distributed only to members of the Society and to conference attendees.

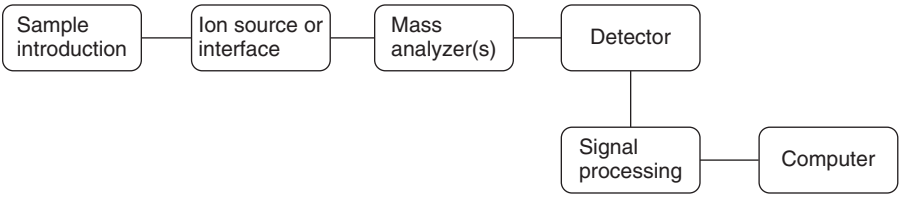


**Figure 1** Modular diagram of a mass spectrometer system (from Burinsky, 2006).

## 4 INSTRUMENTATION

Figure 1 shows a very simplified (and simple-minded) modular diagram of a modern mass spectrometer system. We will make brief commentary on these various parts. Our main focus is to discus what to do with the data after one detects the ions—interpretation!

Mass spectrometry is an inherently gas-phase (vacuum[4]) experiment. The low-pressure gas-phase nature, however, should immediately suggest incompatibilities. Facile sample introduction into an instrument suggests compromising the vacuum chamber. Developing practical, efficient methods and mechanisms for accommodating the transition from samples at atmospheric pressure and in condensed phase (solids or liquids) into a vacuum system—into the mass spectrometer—has been an ongoing effort. Once into the instrument, ion sources and ionization methods convert molecules to charged particles, if they are not already charged.

A molecular ion or fragment ion requires a net electrostatic charge in order to be influenced by and steered by an external magnetic and/or electric field. Once charged particles are produced, a number of different mass filters or mass analyzers are available to achieve sorting of particles—molecular ions, fragments, clusters, and so on—according to their mass-to-charge ratio. This is an opportune time to emphasize that the mass spectrometer does not measure mass, but measures mass-to-charge ratio. Commonly, and until recent times, most data collected by mass spectrometers related to particles that contained a single excess electrostatic charge. The mass-to-charge ratio of a particle reduced to the "mass" of the particle divided by the +1 or –1 electrostatic charge. Although multiply charged ions had been reported in the literature, starting with some of Thomson's original observations (Kiser, 1965), they were uncommon until electrospray was developed for direct observation of macromolecules.

Detection[5] is accomplished most typically by using an electron multiplier, which works essentially the same as a photomultiplier. The ion beam impinges on the first dynode of a multiplier stack. Variations in recent years have replaced the nude electron multiplier with a phosphorescent dynode—the ion beam strikes the dynode and emits a pulse of light, which is then detected and amplified by a photomultiplier. This technique is reminiscent of the original detection mechanism used by Thomson and others, a variation on a phosphorescent television screen picture tube. Microchannel

[4] An entire technological field of vacuum science and techniques exists in order to achieve the necessary vacuum and pumping speed, which we shall ignore at present.

[5] Detectors, like vacuum systems and technology, are another extensive science in and of themselves, which we will simple take for granted here.

plate detectors and other variations on this theme have been introduced, some to implement two-dimensional detection. This property recalls the Mattauch–Herzog classic double focusing mass analyzer geometry, which focused a high-resolution ion beam on a focal plane for, among other things, photographic plate detection of mass spectra.

The "vacuum zone" of a mass spectrometer is classically considered to include the ion source, mass analyzer, and detector components of Fig. 1. Sample introduction mechanisms allow being able to efficiently bridge the gap between atmosphere and vacuum. Direct insertion probe locks for EI and FAB ionization required complex valving systems. Much of the technological effort expended in developing chromatography–mass spectrometry interfaces dealt with the condensed-phase and high-pressure to low-pressure transition, along with disposing of excess solvents and carrier gases. A significant shift occurred with the perfection of the API sources, now used with electrospray and several other methods. The vacuum chamber boundary moved from between the ion source and the sample introduction boxes of Fig. 1 to between the ion source and mass analyzer boxes. Sample introduction, desolvation, and ionization now occur at atmospheric pressure. Introduction of a representative plume of vapors into the mass spectrometer for mass analysis is now done through a small sampling cone strategically placed in the atmospheric ionization chamber.

## 5 INLET SYSTEMS: GETTING SAMPLES IN

### 5.1 Direct Insertion Probe Techniques

Distillation of samples placed on a direct insertion probe, inserted into the mass spectrometer and placed in close proximity to the ion source, was at one time the most common and most frequently used mechanism for conducting mass spectral analysis. The heat generated by the ion source—typically an EI or CI ion source—caused components to volatilize out of the sample. If a pure compound, heating to above the melting point of the compound produced vapors, which were ionized and mass-analyzed. If a mixture, components would distill off the probe according to their melting points and volatilities. It is a crude but effective evaluation of homogeneity of a sample as well as providing identity information about the components. Independently heated probes made the experiment all the more reliable and useful, as rates of heating could be controlled. Samples that would thermally decompose during the distillation could be addressed by using a direct exposure probe—one in which the sample was placed on an electrically heated wire filament, and the filament heated ballistically to volatilize sample very quickly. Running the race between intact volatilization and thermal decomposition often produced valuable structural information. FAB ionization was also used primarily as a direct insertion probe technique. Sample, dispersed in a matrix, was inserted on a probe into proximity with the ion source, and bombarded by a beam of fast atoms (or ions, in the case of liquid secondary ion mass spectrometry or SIMS).

Replacement of a direct insertion probe with the effluent of a gas chromatographic column constitutes the GC-MS experiment, described later. FAB was adapted to serving liquid chromatographic effluents by the flow-FAB experiment, also described later. Once being the standard sample introduction method, instruments capable of conducting direct insertion probe experiments are becoming rare.

### 5.2 Liquid and Gas Septum Inlet

Gases and thermally stable liquids can be introduced into the mass spectrometer by injecting a sample through a septum into an evacuated, heated expansion bulb. The resulting vapor can be leaked into the source of the mass spectrometer, and the spectrum recorded. The caution necessary here is decomposition of the analyte by contact with hot walls of the expansion bulb and inlet tubing. All-glass inlet systems, in which the sample only contacts glass walls, are available, but this does not always solve these kinds of problems. For gases, one must use a gas-tight syringe or other types of transfer device. Introduction of samples by this mechanism presupposes EI or CI is being used. Such inlet systems are also becoming rare in commercial instruments.

### 5.3 Gas Chromatography–Mass Spectrometry

GC-MS is a mature technique. Gohlke and McLafferty (1993) reviewed events in the early history of its development. The first successful experiments using this technique were conducted in late 1955, coupled to a time-of-flight (TOF) mass analyzer. Other analyzers available at that time could not scan fast enough to be able to record an adequate full-scan mass spectrum during the elution of a gas chromatographic peak, even though gas chromatography was being done at that time on packed columns, with very wide chromatographic peaks. Improvements in the scanning speed of magnetic sector analyzers, and the advent of the quadrupole mass filter, supplanted the original TOF analyzers until recent years when TOF technology improved.

Both gas chromatography and mass spectrometry are gas-phase experiments. However, coupling a gas chromatographic effluent to a mass spectrometer requires dealing with the problems associated with making the transition from a high-pressure gas-phase experiment to a low-pressure gas-phase experiment. Two older books on GC-MS by McFadden (1973) and Message (1984) describe the early days in the maturation of the GC-MS technique. The problems of coping with relatively high flows of carrier gas, in the case of packed column chromatography, and of enrichment of the analyte in a large excess of carrier gas have been dealt with straightforwardly.

Several types of carrier gas separators were developed to accommodate packed column gas chromatography—to accommodate the high gas flows and the necessity to enrich analytes. Most assume that helium is used as the carrier gas, and rely on the small molecular size of helium in relation to the larger size of analytes of interest. Typical packed column gas flows accommodated were 30 mL/min (or more) of helium carrier gas, flowing into the separator, and 3 to 5 mL/min flowing on into the mass spectrometer.

Because of the substantially lower total gas flow through a fused silica capillary gas chromatography column, a separator is no longer needed. With flexible fused silica columns, the column end can be inserted directly into the mass spectrometer. Typical carrier gas flows through capillary columns of 1 to 2 mL/min can be readily accommodated by the pumping systems on modern mass spectrometers. The end of the capillary column can be positioned in close proximity to the ion source of the mass spectrometer, if not actually inserted into it. The sensitivity of detection is thus maximized, since this positioning delivers the entire sample directly into the ion source. Separator technology is not entirely obsolete, however. Megabore fused silica columns are often operated with carrier gas flows approaching those of packed columns. The higher flows need to

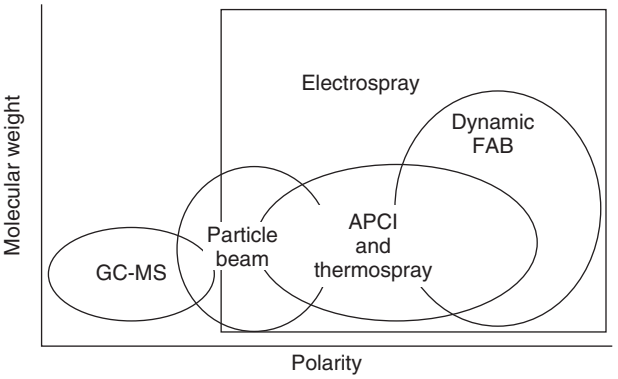


**Figure 2** Applicability of various chromatography–mass spectrometry techniques to the molecular domain.

be processed through a jet separator, with the requisite reduction in total gas flow and analyte enrichment.

Not all ionization methods are available for use in GC-MS experiments. Because the GC experiment is a gas-phase experiment, only those ionization methods that operate in the gas phase are appropriate. Standard EI and CI methods are applicable. Recent years have seen a reintroduction of field ionization coupled to GC-MS experiments. The literature suggests spectra of quality comparable to that generated by conventional EI.

The gas-phase limitation imposed by the very nature of the GC-MS experiment limits its applicability to chemical space. The Venn diagram in Fig. 2 illustrates an approximate mapping of applicability of the combined chromatography–mass spectrometry techniques discussed here. Using only relative molecular mass and polarity as two dimensions that can map chemical space, the figure indicates that the applicability of GC-MS is limited to those relatively small, thermally stable, nonpolar compounds that can be readily volatized (or made to volatilize by derivatization (Knapp, 1979) to pass through a gas chromatographic column.

### 5.4 Liquid Chromatography–Mass Spectrometry (LC-MS)

Combined LC-MS is a rapidly maturing technique. Peter Arpino's cartoon (Fig. 3) captures well the essence of the difficulties associated with coupling the liquid chromatographic experiment with the mass spectrometric experiment. The historical perspective presented in the following is intended to illustrate the difficulties in achieving a useful LC-MS interface, and the variety of devices and the ingenuity of the people attempting, and eventually succeeding, to make it happen. The LC-MS experiment has only become an effective experiment since the early to mid-1990s.

Attempts to couple condensed-phase chromatographic methods with mass spectrometry start with the work reported by Victor Tal'Rose[6] (Baldwin, Burlingame and Nikolaev, 2004) and colleagues in 1968 (Tal'Rose *et al.*, 1968). By referring to

[6] Victor Tal'Rose recently passed away.

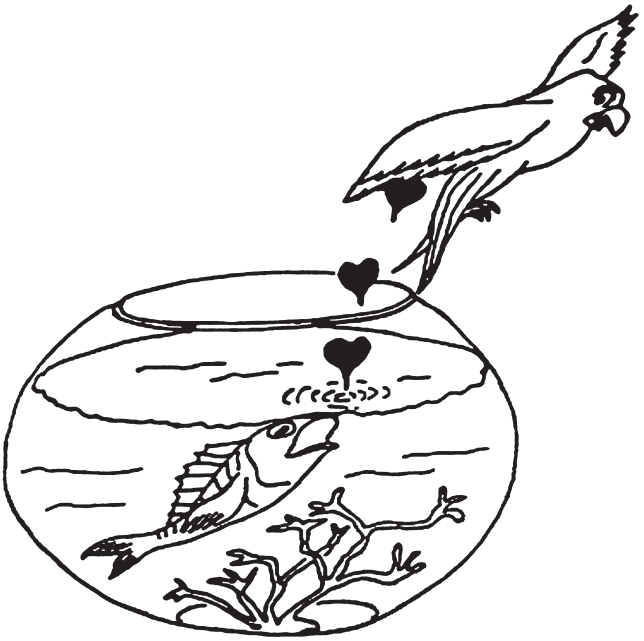

**Figure 3** Combining liquid chromatography with mass spectrometry involves a difficult courtship (from Arpino, 1982).

condensed-phase chromatographic methods here, we primarily imply HPLC. Attempts, with varying degrees of success and acceptance, have been reported to interface other condensed-phase separation methods, such as thin-layer chromatography (Wilson, 1999; Busch, 2004a) and supercritical fluid chromatography (Randall and Wahrhaftig, 1981; Smith and Udseth, 1983). Yergey *et al.* (1990) summarized the progress on development of effective LC-MS interfaces, and compiled a review and comprehensive bibliography of interface development and applications up to 1990. They do not cover electrospray, as the explosion of efforts and applications of this technique only started in this time frame.

Unlike combined GC-MS, LC-MS presents a significant incompatibility between the condensed-phase liquid chromatographic experiment and the gas-phase mass spectrometric experiment. The pumping system of the mass spectrometer must be capable of handling the gas load introduced by the chromatographic experiment. Typical gas loads generated by a gas chromatographic experiment are 3 to 5 mL/min for packed columns (after jet separation!) and 0.5 to 1 mL/min for capillary columns. HPLC mobile phase must be vaporized. Typical mobile phase flow is 1 to 3 mL/min. Hexane, converted to gas phase, produces 180 to 540 mL/min of vapor. Chloroform produces 280 to 840 mL/min. Methanol produces 350 to 1650 mL/min. Water produces 1250 to 3720 mL/min. The design and engineering of an interface to connect a liquid chromatographic experiment with a mass spectrometric experiment must accommodate these large gas flows—approximately 1000× that of the gas chromatographic experiment—and still permit both the chromatographic experiment and the mass spectrometric experiment to work properly.

***5.4.1 Direct Liquid Introduction (DLI)*** DLI is the simplest and most straightforward approach. It was first attempted and reported by Tal'Rose *et al.* (1968). Baldwin and McLafferty (1973) reported their attempts, utilizing CI. Chapman *et al.* (1983) reported their experiences with interfacing to a magnetic sector instrument. A small portion of chromatographic effluent (10 to 20 μL/min) is sprayed into the ion source through a pinhole leak into the mass spectrometer. CI is most frequently used, with the residual solvent molecules in the spray acting as the CI reagent.

***5.4.2 Atmospheric Pressure Chemical Ionization (APCI)*** Rather than leaking a portion of a chromatographic effluent into the vacuum system of the mass spectrometer, the effluent is introduced via a heated nebulizer into an API source. The source is outside the vacuum chamber, operating at or above atmospheric pressure. Ionization occurs in a corona discharge, struck in the plume of effluent sprayed from the heated probe. The solvent vapors provide a source of protonating reagent for the proton transfer CI reactions in the vapor. The corona discharge region is sampled through a sampling orifice by the mass spectrometer. The discharge was first struck by proximity with the ionizing radiation from a $^{63}$Nickel radioactive needle, and later by a needle with an applied high voltage (Carroll *et al.*, 1975). The ionization is soft, as the majority of compounds produce abundant protonated molecular ions.

***5.4.3 Continuous Flow FAB Interfaces*** FAB ionization revolutionized the application of mass spectrometry to large, thermally fragile, nonvolatile molecules. This desorption technique was described by Barber and colleagues in 1981 (Barber *et al.*, 1981; Surman and Vickerman, 1981). Dependence of the technique on a dispersing matrix and its inherent incompatibility with flowing sample introduction into the mass spectrometer prevented its immediate application to chromatographic separations. However, a mobile phase adulterated with 10% (by volume) of glycerol, the classical FAB matrix, was infused through a glass frit, the frit being the FAB target. Low flow chromatographic effluent (circa 2 μL/min) could be analyzed by FAB ionization (Ito *et al.*, 1986). Caprioli *et al.* (Caprioli, Fan and Cottrell, 1986) showed that the frit was not necessary, effluent being introduced directly onto the target through a hole in the center. The fine art of obtaining useful results from either of these approaches involved balancing the rate of evaporation of effluent from the FAB target with replenishment of fluid by chromatographic flow. Adjusting solvent composition, glycerol content in the mobile phase and externally applied heat to the target were some of the operational variables.

***5.4.4 The Moving Belt Interface*** Scott *et al.* (1974) and McFadden, Schwartz and Bradford (1976) first described this mechanical interface in 1974 and 1976. A diagram of a commercialized moving belt interface is shown in Fig. 4. The interface first consisted of a spool of wire, which was unrolled off one spool and onto another. As the wire was wound from spool to spool, the effluent from a liquid chromatographic separation was applied to the wire. As the wire was transported through the ion source of a mass spectrometer, it was heated to desorb analytes that had been applied to it. The desorbed materials were ionized and mass-analyzed in the mass spectrometer. Later, a recycling continuous loop of wire, or a belt of an appropriate material, was used, with added mechanisms for "cleaning" the belt to avoid chromatographic confusion. The moving belt is a versatile interface, offering applicability in EI, CI, and FAB (Stroh *et al.*, 1985) ionization modes. However, it is mechanically the most complex, both in

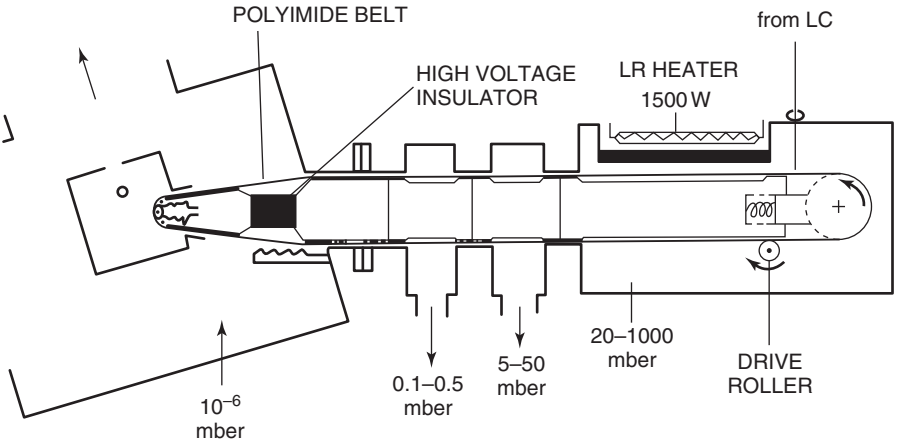


**Figure 4** Schematic diagram of a moving belt interface. Courtesy of V.G. Analytical.

operation and optimization of the interface, and in its installation into a mass spectrometer.

***5.4.5 Thermospray*** The term "thermospray" refers to both a vaporization LC-MS interface device and a soft ionization method. Vestal and coworkers (Blakley, Carmody and Vestal, 1980) developed thermospray, beginning in 1976. Plasmaspray is a variation on the thermospray concept, implemented by V.G. Instruments on their line of magnetic sector and quadrupole instruments. The technique and device were designed as an alternate means of rapidly vaporizing solvent and nonvolatile analytes by heating the tip of a stainless steel tube through which the LC effluent flowed. The resulting spray of vapor and droplets was directed against a heated probe to complete droplet vaporization. An electron beam from a hot filament bombarded this vapor, producing CI, and the vapor was sampled with a sampling orifice into the mass spectrometer. The accidental observation that a filament and electron beam was not needed was made when, unknown to the operators, the filament had burned out and as many ions were being produced and detected without the filament as with the filament. Mass spectra produced by thermospray are soft ionization spectra. The molecular species is typically a protonated molecular ion or an adduct—often an $NH_4^+$ cationated species, because thermospray works best from ammoniated mobile phases. Thoughts on the mechanism(s) involved in this process were reviewed by Vestal (1983). The book by Yergey *et al.* (1990) contains an extensive treatise on this technique. Empirical observations show that ionization is quite efficient when the LC separations are carried out in a polar solvent, which includes ionic solutes such as ammonium acetate. Separations carried out in a nonpolar solvent, without salts, require an assisting electron beam from a filament, if the application works at all.

***5.4.6 MAGIC*** The MAGIC acronym stands for "monodisperse aerosol generation interface for chromatography." This LC-MS interfacing technique was first described by Willoughby and Browner (1984), and evolved into the particle beam interface. Various instrument manufacturers implemented it under a variety of names. It contained no moving parts. Volatile solvents work best. Small portions of split aqueous

flows could also be accommodated. Desolvation was done by evaporation and impinging with a "nebulizing gas." Only sufficient external heat was supplied to counteract the cooling caused by expansion and evaporation of the solvent. Reduction of net flow into the mass spectrometer is done by using the Ryhage jet separator technique (developed for packed column GC-MS). This interface presents the possibility of conducting EI on samples introduced into the mass spectrometer, EI spectra being the "richest" in fragmentation and structural information. EI spectra could be obtained of sufficiently high quality to permit spectral library searching against, for example, the NIST library. The particle beam technique occupies a niche in applications space. Compounds that can produce fragmentation-rich EI mass spectra (e.g. via a probe distillation experiment), but which will not pass well through a gas chromatographic column, work in liquid chromatographic experiments interfaced to the mass spectrometer through a particle beam interface.

Semi-volatile and nonvolatile compounds of interest in environmental sciences (Bellar, Behymer and Budde, 1990), such as the larger polycyclic aromatic hydrocarbons and pesticides (Miles, Doerge and Bajic, 1992), have been effectively addressed. It is theoretically possible to introduce reagent gases into the ion source of the mass spectrometer and do CI on materials introduced through a particle beam interface. Extrel Corporation also developed and commercialized an ion source that included a fast atom source and a target for doing FAB ionization. The mobile phase was adulterated with a low percentage of glycerol. The low percentage, in general, did not interfere significantly with the chromatography. The glycerol acted as a continuously renewing matrix from which to do FAB ionization, the glycerol being sprayed onto the FAB target along with the analyte.

***5.4.7 Electrospray*** The term "electrospray" refers both to the interface design and the ionization mechanism, as do the "thermospray" and "atmospheric pressure chemical ionization" terms. Abbe Jean-Antoine Nollet's 1750 experiments on electrostatic spraying of liquids[7] constitute the first experiments reported with this technique (Fig. 5[8]). Nollet demonstrated a number of ways in which electrostatically charged spraying devices could effectively spray fluids. John Fenn's (Yamashita and Fenn, 1984) more recent elaboration of the electrospray phenomenon has led to a revolution in the application of mass spectrometry, first to biologically relevant large molecules, and subsequently in the application to polar molecules in general.

## 6 IONIZATION METHODS: MAKING IONS

To perform mass spectrometry, one must make ions from neutral molecules. Ionization methods have advanced from the classic EI, through CI (Harrison, 1992), field desorption (Wood, 1982), FAB (Barber *et al.*, 1981; Surman and Vickerman, 1981), and thermospray (Vestal, 1983) to the API techniques currently favored. EI is classic, but is restricted in its applicability to thermally stable, volatilizable compounds. Field

[7] This interesting historical footnote was brought to our attention by Dr. Thomas Covey, MDS Sciex Instrument Corporation.

[8] Information available on the history of chemical instrumentation web site maintained by Prof. Euginii Katz, Department of Chemistry, The Hebrew University of Jerusalem. See http://chem.ch.huji.ac.il/~eugeniik/history/nollet.htm.

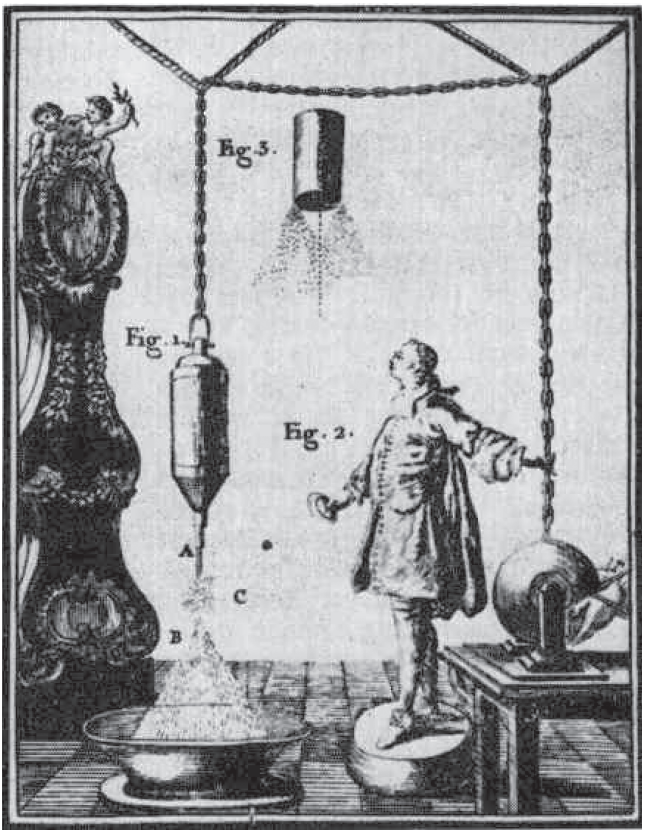


**Figure 5** Abbe Jean-Antoine Nollet's experiments with the electrostatic spraying of liquids.

desorption was always a specialized niche technique applicable to some larger compounds. FAB enjoyed a meteoric rise in use first reported in 1981; but, has all but disappeared now, being replaced by the API techniques: atmospheric pressure chemical ionization (APCI) (Carroll *et al.*, 1975) and electrospray (ESI) (Yamashita and Fenn, 1984). Matrix-assisted laser desorption ionization (MALDI) (Hillenkamp and Karas, 1990) has shown significant utility for characterizing larger proteins, approximately 100 kDa and larger.

The evolution of the ionization techniques has progressed from the hard EI technique, which produces extensive fragmentation, to the soft ionization API techniques, which produce in general only molecular ions.[9] In achieving the goal of being able to

[9] Hardness and softness of ionization refer to the energetics of forming an ion from a neutral molecule, and the extent of spontaneous fragmentation resulting from the ionization process. Hard ionization techniques induce extensive fragmentation and have a high probability of completely destroying the molecular ion. Soft ionization techniques induce minimal or no fragmentation. The ideal soft ionization technique would produce a molecular ion species and nothing else. This is not to imply that substantial amounts of energy need to be deposited in samples in order to induce evaporation, desorption, and/or ionization. See the discussion by Hillenkamp and Karas (1990).

ionize larger, more polar, less thermally stable molecules, and in optimizing for production of molecular ions, much of the structure-indicating fragmentation information has been lost. MS-MS experiments restore structural information content. They select the molecular ion produced by the soft ionization technique in the first mass analyzer of a tandem instrument and activate it by collision with a target gas. The product ion mass spectrum of fragments resulting from collision-induced decomposition (CID) of the parent ion in a later mass analyzer is recorded. The effect of inducing fragmentation by collisional activation can also be achieved through collision with solvent molecules in the API ion source. Various implementations of this procedure are called up-front CID, cone voltage fragmentation, and "poor man's MS-MS." Results obtained this way are comparable to true MS-MS product ion spectra. The strict interpretation of the results in terms of precursor–product ion relationships is lost, but this limitation does not negate the utility of the method when a true tandem instrument is not available.

Discovering and developing new ionization methods is one of the many topics that punctuate the history of mass spectrometry. All of the mass analyzers rely on the ability to use external force fields—magnetic fields, electric fields, radio frequency fields—to steer electrostatically charged particles through space. Methods to convert neutral molecules to charged species while preserving structural information are a perpetual topic of interest and development. Electric discharge in a gas produces ions, but is highly energetic, producing primarily atomic species. It was the ionization method used in Thomson's early parabola mass spectrograph studies that utilized electric discharge, and in the early evaluations of isotopic composition of the elements. EI—a beam of electrons interacting with neutral molecule vapors—is the classical ionization method, producing radical cation molecular ions (when the molecule is stable enough to withstand dislodging of a valence electron) and extensive fragmentation. Development of softer ionization methods is driven by the attempt to produce a molecular ion species without destroying structural integrity of the molecule. CI using a variety of different reagent gases, field ionization (FI), field desorption ionization, FAB, APCI, electrospray (ESI), and MALDI are the more well-known and prominent ionization methods. The order in which they are listed is an approximate order from hardness to softness. This list is also not comprehensive. Vestal (2001) reviews these and a number of other minor specialty ionization methods.

Which method should one use for a particular application? The first factor to consider is the type of experiment to be used. Not all of these methods are compatible with combined chromatography–mass spectrometry experiments. Figure 2 gives a first impression of where, within the universe of chemical compounds, various techniques can be applied. The nature of GC-MS limits use to EI and CI, with some FI, and to compounds that are sufficiently volatile and thermally stable to pass successfully through a gas chromatography column. Attempts have been made to utilize nearly all of the major ionization methods for LC-MS, with varying degrees of success. Figure 2 again tries to capture the circumstances. FAB and MALDI are desorption ionization methods, and present particular engineering difficulties over how to present sample while preserving chromatographic separation. Workable solutions have been developed for FAB. While the general nature of pharmaceutically relevant compounds makes electrospray and APCI nearly ideal for this application area, they do not work well for all compounds. In short, there is no universally applicable ionization method.

### 6.1 Electron Ionization

Electron ionization (also known as electron impact) is the classical ionization method. Vapors of the sample are bombarded with a beam of high-energy electrons (typically at 70 eV), produced by an electrically heated filament. During interaction of an energetic electron from the ionizing beam with an analyte molecule, an outer shell electron is displaced from the neutral molecule, producing a positively charged radical cation. Even with a low-energy transfer efficiency, a large amount of energy is still deposited into the molecule. Substantial fragmentation of the molecule is expected. The timescale of interaction of the ionizing electron with the neutral molecule is very short with respect to any vibrational events within the molecule. As the energy starts to redistribute into the various vibrational modes of the molecule, vibrations increase until a bond is broken and a fragment produced.

Organic mass spectrometry under EI is typically performed at 70 eV. This energy is chosen because the generalized shape of the ionization efficiency curve for most gases and organic molecules plateaus at approximately 60 to 80 eV. Standard libraries of EI spectra are recorded at 70 eV, as the mass spectrum of most compounds by EI at 70 eV is very reproducible from instrument to instrument and from time to time. The EI source is very reliable and easy to operate. Low eV observations, however, can be useful. Less energy is transferred to the molecule during ionization, and less fragmentation occurs. For particularly fragile molecules, low eV observations can often help to identify the molecular ion.

### 6.2 Chemical Ionization

This ionization technique, first christened as chemical ionization, was described by Munson and Field (1966). The first observation of the $CH_5^+$ species in a high-pressure ionization source, however, was reported by Tal'Rose and Lyubimova (1952). Field summarized the early work (Field, 1968) and published a historical retrospective (Field, 1990). A reagent gas is bombarded by electrons to form a reagent ion in the gas phase. The sample of interest is introduced into the mixture of neutral reagent gas and reagent gas ion. Sample molecules are ionized by ion–molecule reaction with the reagent ion, typically by proton transfer, often producing an abundant protonated molecular ion, $[M+H]^+$. The attraction of CI is that compounds, which yield a very low abundance molecular ion, if at all, by EI, will yield molecular mass information, in the form of an $[M+H]^+$ under CI. The most commonly used reagent gases used for CI mass spectrometry are methane, isobutane, and ammonia. Methane was used in the original observations. Isobutane often produced less fragmentation and/or more abundant $[M+H]^+$. Ammonia presents selectivity in ionizing compounds which have proton affinities higher than ammonia, but not ionizing compounds with proton affinities lower than ammonia. A number of exotic reagent gases have been explored for even more selective CI experiments. Negative ion chemical ionization is also a useful variation on this experiment, applicable to highly electronegative molecules. Alex Harrison's compendial book summarizes the fundamental and applied work done to explore this technique (Harrison, 1992).

### 6.3 Field Ionization and Field Desorption Ionization

Field desorption ionization was first described by Beckey (1969), comparing and contrasting to field ionization. Neutral molecules "introduced" into an electric field with a

very large field gradient will ionize. Thermal evaporation of a substance and introduction into the field gradient characterizes field ionization. Field desorption includes evaporation of target analyte from an activated emitter, upon which a fine brush-like structure of micro-needles had been grown. The activated emitter is dipped in a solution of analyte. After introducing into the vacuum chamber of the mass spectrometer, solvent is allowed to evaporate. The emitter is then heated by electric current until the analyte desorbs, producing (when successful) a spectrum of abundant protonated molecular ion. The technique was first demonstrated for monosaccharides, these and larger polysaccharides being notoriously difficult compounds mass spectrometrically. Mechanisms of how this technique works have been investigated (Holland, Soltmann and Sweeley, 1976). Wood (1982) catalogs a list of successful applications of field desorption. The technique, however, tends to be very dependent on the technique of the operator and the skill with which the emitters are prepared (Schulten and Beckey, 1972; Lehmann and Schulten, 1978). For this reason, the technique was not widely used, and has fallen into disfavor, FAB (see further discussion) being the preferred replacement.

### 6.4 Plasma Desorption Mass Spectrometry

Plasma desorption is a particle-induced desorption technique described in 1974 by Torgerson *et al.* (Torgerson, Skowronski and Macfarlane, 1974). A sample is deposited on a foil target and placed in the proximity of a $^{252}$Californium radioactive source in a TOF mass spectrometer. Upon fission, a $^{252}$CF nucleus splits into two high-energy fission fragments, which are ejected in opposite directions. One fragment triggers a clock, while the other bombards the sample foil, causing desorption and ionization of target analyte molecules on the film. The desorbed ions are then accelerated and mass-analyzed in the TOF mass spectrometer. It is a soft ionization technique which produced information on large, nonvolatile molecules. It preceded FAB as one of the few methods at the time for directly observing peptides, proteins, and other biologically important molecules. Reviews describe techniques and successful applications (Macfarlane, 1983; Sundqvist and Macfarlane, 1985; Macfarlane *et al.*, 1994). This technique, while having significant impact during its height, was subsequently replaced by FAB.

### 6.5 Fast Atom Bombardment and Secondary Ion Mass Spectrometry

FAB was first described in 1981 by Barber *et al.* (1981) and by Surman and Vickerman (1981) in a pair of back-to-back published papers. It is a desorption ionization technique, an offshoot of SIMS (Honig, 1985), in which the surface of a liquid matrix containing the target analyte molecule is bombarded with high-energy neutral atoms, rather than with fast ions as in SIMS. Both positive and negative ions are produced by the same process. The yield of negative ions is generally at least an order of magnitude less than that of positive ions. The technique makes large, nonvolatile, and thermally labile molecules accessible for study by mass spectrometry. Field desorption addressed these same classes of molecules, but FAB proved to be a much simpler and more reliable technique.

Sample was dispersed in a drop of matrix on the end of a direct insertion probe. Selection of a matrix compatible with the chemistry of the target analytes became an art form. While glycerol, the originally used matrix, was relatively general purpose,

a number of other widely used matrices, such as *o*-nitrobenzyl alcohol (Meili and Seibl, 1984; Sweetman and Blair, 1988) or a eutectic mixture of dithiothreitol and dithioerythritol (Witten *et al.*, 1986), are prevalent in the literature. FAB mass spectra are mixture spectra, features in the spectrum being from either the target molecule or the matrix. Features derived from commonly used matrices are cataloged by Costello (1991).

### 6.6 Matrix-Assisted Laser Desorption Ionization

Hillenkamp and Karas first described this technique in 1985 (Karas *et al.*, 1987) and reviewed early progress (Hillenkamp and Karas, 1990). Laser desorption of target analytes from surfaces had been exploited extensively, but was restricted to certain classes of compounds, in particular those of no more than approximately 1000 daltons. Matching wavelength of irradiating light from a laser with the absorption maxima of target analytes imposed further technical difficulties. By contrast, dispersing a target analyte (or mixture of analytes) in a matrix of a light-absorbing small molecule matrix, and irradiating with a wavelength at which the matrix absorbed, effected absorption and energy transfer into the target analytes without extensive fragmentation of the analytes and without the necessity of tuning the irradiating wavelength to the absorption maximum of the target. A number of other advantages also presented. The molecular mass boundary of laser desorption could be extended to intact proteins. In the process of examining this application, a number of small aromatic molecules which absorbed in the 270-nm region were examined for utility as matrices (Beavis and Chait, 1989). Attempts to find the magic matrix, choosing or tailoring the matrix to the application, resulted in a significant amount of research activity, similar to searching for that special FAB matrix. Today, MALDI competes effectively with electrospray (see next section) as the ionization method of choice, in particular for large proteins and other macromolecules.

### 6.7 Electrospray

Malcolm Dole (Dole *et al.*, 1968) described his group's attempts to measure the relative molecular masses of oligomers of synthetic polymers based on electrospray ionization in 1968. Fenn's group elaborated on the Dole observations, and built the electrospray ion source shown in Fig. 6 (Yamashita and Fenn, 1984). In essence, liquid sample containing molecules of interest is flowed into the source through a small tube, on which a high electrostatic voltage is placed. The liquid is sprayed from the tube tip, forming a Taylor cone and subsequently a stream of droplets, each electrostatically charged. A countercurrent flow of dry nitrogen gas opposes the spray, causing evaporation of solvent. As the droplets become smaller because of solvent loss, they reach the Rayleigh instability limit, where coulomb charge repulsion overcomes solvent surface tensions and the droplets fragment further. Repetition of this process ultimately produces solvent-free solute ions, which are subsequently swept into the sampling orifice of the mass spectrometer and detected.

The detailed mechanism of electrospray is still an active and lively topic of discussion at current scientific conferences. Numerous research laboratories are actively probing the subject. Volume 35, issue 7 (2000), of the *Journal of Mass Spectrometry* published a series of five papers (Amad *et al.*, 2000; Cole, 2000; Gamero-Castaño and Fernandez de la Mora, 2000; Kebarle, 2000; Van Berkel, 2000) from research groups

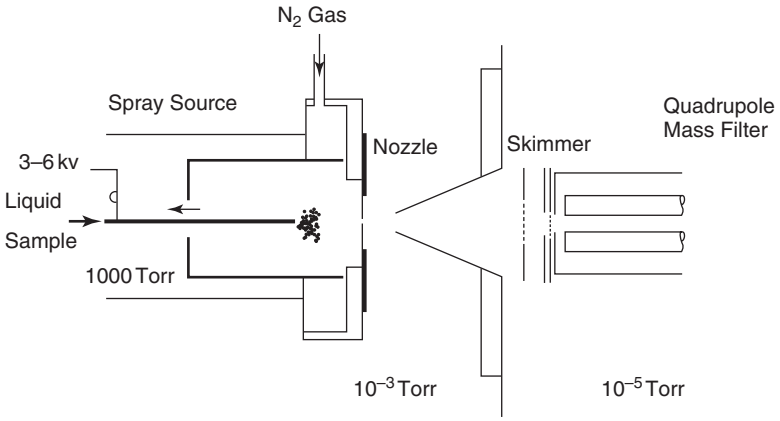


**Figure 6** Electrospray ion source (from Yamashita and Fenn, 1984).

working in this area, followed by a lengthy printed discussion (Fernandes de la Mora *et al.*, 2000) in the subsequent issue.

Two models have been repeatedly discussed in the literature—ion evaporation, attributed to Iribarne and Thomson (1976), and coulomb repulsion, attributed to Dole. Both models rely on the Rayleigh instability of charged droplets to become smaller and smaller. They differ in that coulomb repulsion relies on the continuous repeating of the process until one is left with a single isolated desolvated ion. Ion evaporation invokes the idea that, before a droplet becomes small enough to contain a single solute molecule, the electrostatic field strength at the droplet surface becomes sufficiently intense to eject a surface ion from the droplet into the ambient gas. Kebarle reviewed the data available in 1991 (Blades, Ikonomou and Kebarle, 1991), and updates and interprets (Kebarle, 2000) the summed results to propose that the data indicate that the mechanisms constitute a continuum, and that elements of both interplay, depending upon physical properties such as hydrophobicity or hydrophilicity, most notably the size of the molecule under consideration—small ions are produced by ion evaporation, large ions are produced by coulomb repulsion—and other environmental factors such as ion concentration in solution. Vestal (2001) gives an additional concise summary of the current thoughts on how electrospray works, but more importantly summarizes a number of design variations of electrospray ion sources extant on modern instruments.

### 6.8 Atmospheric Pressure Chemical Ionization

APCI is similar to electrospray in that ionization occurs at atmospheric pressure. The effluent of, for example, an HPLC column is passed through a heated nebulizer and sprayed into the source chamber. Ionization occurs in a corona discharge, struck in the plume of effluent. The solvent vapors provide a source of protonating reagent for the proton transfer CI reactions in the vapor. The corona discharge region of the spray plume is sampled through a sampling orifice by the mass spectrometer. When this experiment was first reported, the discharge was struck by proximity with the ionizing radiation of a radioactive nickel needle (Carroll *et al.*, 1975). The radioactive needle

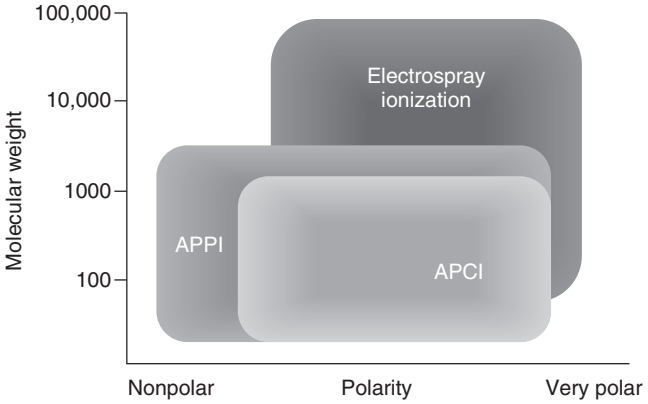


**Figure 7** Applicable range of molecules ionizable by APPI (from Syage, Short and Cai, 2008).

was later replaced by a needle with an applied high voltage. APCI ionization is soft, as the majority of compounds produce abundant protonated molecular ions. It is, however, somewhat harsher than electrospray, and is often used for compounds that do not have readily protonatable sites in the molecule, and require higher energies to make ions. Regarding the "quality" of the spectra, some comparisons have been made to FAB, using extent of fragmentation as a measure of quality.

### 6.9 Atmospheric Pressure Photoionization (APPI)

APPI is a variation on the electrospray and APCI techniques. Instead of a corona discharge or a high-voltage potential to effect ionization, a high-intensity ultraviolet lamp is used to effect photoionization of analytes in a chromatographic effluent stream (Robb, Covey and Bruins, 2000). Photoionization detection combined with gas chromatography has been studied and available since the 1970s. Photoionization, coupled with addition of dopants such as toluene or methanol into the chromatographic effluent to control the photochemistry occurring in the sprayed effluent, has been demonstrated to present advantages of sensitivity and extending the range of applicable types of compounds. Developers claim to be able to address a group of nonpolar analytes which are not accessible by either electrospray or APCI (Fig. 7). Study of the mechanism of APPI (Syage, 2004), where the range of applications was discussed, has been published (Syage, Short and Cai, 2008).

### 6.10 DART and DESI

Two new ionization methods to appear recently are desorption electrospray ionization (DESI) and direct analysis in real time (DART). They are the first in a new area of ionization mechanisms being explored which are collectively referred to as open-air ionization sources. Recall the earlier discussed progression from forming ions in an evacuated chamber to forming ions in an atmospheric pressure chamber. Both DART and DESI form ions in an "open atmosphere" and sample the resulting plume through an entrance cone into the mass spectrometer.

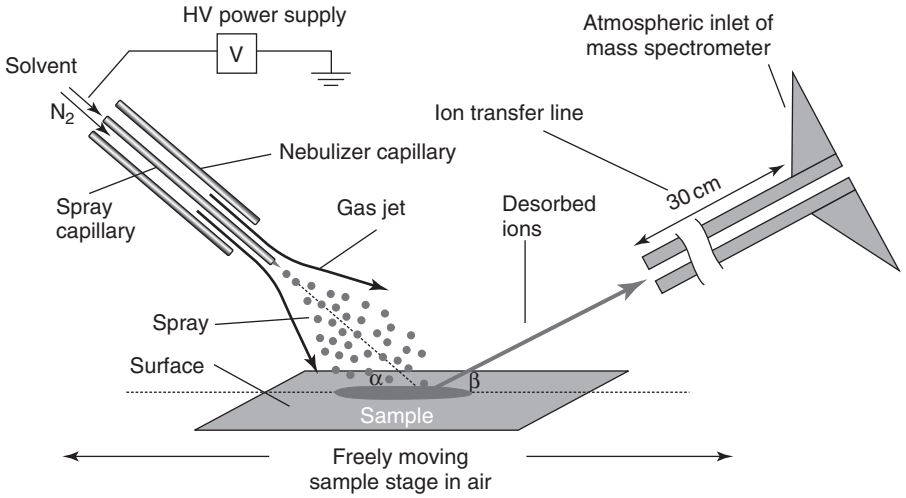


**Figure 8** Desorption electrospray ionization source (from Takáts *et al.*, 2004).

DESI has been described in detail (Takáts *et al.*, 2004, 2005). Figure 8 shows the configuration of the source and sampling arrangement. A spray of solvent, generated by an electrospray spraying device, is impinged on a surface. This beam of impinging particles consists of gas-phase solvent ions, ionic clusters, and charged microdroplets. These charged projectiles act similar to the bombarding particles (ions, atoms, photons) in the desorption ionization methods, chemically sputtering target analyte molecules off the surface. The nearby mass spectrometer entrance orifice introduces a sampling of this collection of sputtered chemical species for mass analysis. The process has been likened to a "power-washing" process.[10] In a very short time, a number of successful applications have developed using this technique, from analyzing thin-layer chromatographic separations to doing two-dimensional imaging of surfaces and of thin sections of biological tissues.

The applicability of DESI to the two-dimensional mapping of a surface resurrects significant interest in the combination of mass spectrometry and thin-layer chromatography. Reports have periodically appeared in the literature on developing devices and techniques for accomplishing this coupling, such as enclosing an entire x-y-translatable stage (such as implied by Fig. 8, and actually used by Takáts *et al.* 2004) within a vacuum chamber and adding a layer of matrix for FAB ionization (Busch, 1995). A relatively recent review (Wilson, 1999) summarizes those efforts. The innovation comes here in that the entire x-y-translatable stage apparatus can be placed in the open air.

The DART ionization source (Cody, Laramée and Durst, 2005) uses an electric discharge to produce a plasma of electrons, ions, and excited states of a gas, typically nitrogen, which is then formed into a high-velocity gas stream. This gas stream can be directed toward a liquid or solid sample, or it can interact with vapor-phase samples. As with the DESI source, the angular and distance orientation of the ionizing gas stream toward the sample surface and the mass spectrometer entrance influences the

[10] The power-washing analogy is attributed to Dr. Gary van Berkel (2007).

quantitative aspects of DART sampling. However, regarding efficient operation of the technique, and for qualitative applications, these orientation parameters are not critical. This noncritical nature makes sample preparation and presentation for study by this technique very easy. Therein lies the attraction, especially when identity of substances is the goal of the experiment and absolute amounts or percentage compositions of mixtures are not necessary results of the analysis.

## 7 MASS ANALYSIS: SORTING BY SIZE

### 7.1 Time-of-Flight Analyzer

The TOF mass analyzer is the simplest type of mass analyzer. Ions in a beam produced by an ion source have a "constant" kinetic energy, but have velocities that differ according to the mass of the particle. Recalling that kinetic energy is expressed as $0.5\,mv^2$, particles of higher mass will have lower velocities and will take longer to travel along a tube of fixed length than lighter particles with higher velocities. By measuring the elapsed time of travel of particles at a detector, a fixed distance from the ion source is correlated to the mass of the particles.

### 7.2 Quadrupole and Triple Quadrupole

The quadrupole mass filter has several properties that make it very useful. It can be scanned very fast over a wide mass range. It has very high transmission efficiency. A high percentage of the ions of a selected mass which are introduced into the filter are transmitted. The ratio of the DC voltage and the radio frequency voltage applied to the quadrupole determines the mass of the particle transmitted. If the DC voltage is turned off, the quadrupole becomes a high-efficiency ion pipe, transmitting all masses. The quadrupole can operate efficiently at higher pressures and much lower voltages than magnetic sector instruments (see further discussion). These several advantages make it a good mass spectrometer to couple to a gas chromatograph—especially a capillary chromatograph which produces very narrow chromatographic peaks. This also presents a significant advantage in coupling to ultrahigh-pressure liquid chromatographs, which also produce very narrow chromatographic peaks. Quadrupoles can be very small, making them ideal for benchtop laboratory instrument applications. The primary disadvantage of this device is that, operated under typical conditions, it is a low-mass resolution device.

Campana has presented an excellent discussion of quadrupolar mass filter theory (Campana, 1980). The quadrupole arrangement is not the only type of mass filter that works according to the basic principles of this device. A number of n-pole mass filters have been described, including a monopole. A hexapole is commonly used in several commercial instruments. In a three-dimensional "quadrupole," the ion trap has been developed.

The triple quadrupole mass analyzer was first assembled by Yost, Enke and Morrison as a device for conducting photoionization experiments. It is a tandem mass spectrometric arrangement where three quadrupole mass filters are connected such that the exiting ion beam of one quadrupole is introduced as the source of ions for the next quadrupole in series. The first quadrupole was used to select a specific precursor ion from a heterogeneous beam of ions and introduce it into the second quadrupole. The second quadrupole was used in its RF-only ion pipe mode to contain and constrain the

selected ions. These selected ions were then irradiated with light in order to induce photochemical reactions. The products were then introduced into the third quadrupole and mass-analyzed to identify the photochemical products. After conducting these experiments, the originators and many others recognized that this configuration could also be used, by conducting collisional activation with a collision gas in the central RF-only quadrupole, to do definitive structural and quantitative experiments. The triple quadrupole configuration is probably now the most common tandem configuration, and perhaps the most common quadrupole configuration in use in modern chemical laboratories.

### 7.3 Ion Trap

Wolfgang Paul shared the 1989 Nobel Prize in physics (http://nobelprize.org/nobel_prizes/physics/laureates/1989/) for his work on the ion trap. The trap is a three-dimensional quadrupole, which can be used in many of the same capacities as the triple quadrupole. Additionally, tandem mass spectrometric experiments with multiple stages of collisional activation and decomposition beyond the single-stage MS-MS experiments of the triple quadrupole can be conducted.

### 7.4 Magnetic Sector

The magnetic sector mass analyzer is the stereotypical mass spectrometer. When one mentions a mass spectrometer, this image is conjured. Single-focusing instruments have been commercially built. Double-focusing magnetic machines combine a magnetic momentum analyzer sector with an energy-focusing electric sector. A number of double-focusing sector geometries have been developed, including the forward and reverse Nier–Johnson designs, the Mattauch–Herzog focal plane design, and numerous others. The Nier–Johnson is the most commonly seen and commonly commercially manufactured geometry. Variations of this geometry have been used in producing multi-sector tandem mass spectrometers. The sector instrument has largely become obsolete in recent years. Its expense, complexity, and difficulty of operation have caused its demise in favor of quadrupole instruments, TOF instruments, and ion traps. Improvements in the reflectron time-of-flight (rTOF) mass analyzer have even challenged the one remaining niche that the sector instruments, and the equally technologically intense Fourier transform mass spectrometers (FT-MS) occupy, namely the high-resolution accurate mass measurement arena.

### 7.5 Ion Cyclotron Resonance Fourier Transform Mass Analyzer

The FT-MS uses the principles of the ion cyclotron resonance (ICR) experiment to make mass measurements. The mass-to-charge ratio of particles undergoing cyclotron resonance in a strong magnetic field correlates with the period of their orbiting in the magnetic field. This frequency domain is transformed into a mass-to-charge domain by use of the mathematical Fourier transform. Many portions of the technology used in FT-MS originated from Fourier transform NMR. A recent review describes this and related high-resolution techniques (Marshall and Hendrickson, 2008). The two strengths FT-MS presents are the potentially extremely high resolution the instrument can generate for making accurate mass measurements, and its tandem mass spectrometry capabilities. The price for this performance is a large, expensive, and relatively complex instrument.

### 7.6 Tandem Mass Spectrometry (MS-MS)

Coupling two or more mass analyzers in series produces a tandem mass spectrometer (an MS-MS) (McLafferty, 1983), the best known of which is the triple quadrupole. A number of tandem sector geometries have been built, some of which have been commercially available (Crow, Tomer and Gross, 1983). The historical first tandem sector instruments were by Futrell and Miller (1966) and by McLafferty (1981). Hybrid geometries, combining a magnetic sector instrument with a quadrupole, an ion trap, or a TOF analyzer, have also been built and characterized and are described in the literature. The ion traps, triple quadrupoles and, more recently, the tandem quadrupole–TOF instruments are the more abundant and user-friendly instruments. The quadrupole and the sector geometries provide the ability to perform tandem-in-space MS-MS experiments. The ion traps and FT-MS instruments permit performing tandem-in-time MS-MS, permitting $MS^n$ (mass spectrometry raised to the nth degree, multiple stages of collisional activation and mass analysis) experiments. Multiple MS-MS stages in space would require a mass analyzer for each MS-MS stage, rapidly making tandem-in-space experiments impractical for an n-value greater than 2.

## 8 DETECTION: SEEING WHAT HAS BEEN GENERATED

### 8.1 Presentation of Mass Spectra

Wide variation is seen in the absolute intensity of signals in a mass spectrum. To compensate for this variation, a mass spectrum of a compound or a mixture is typically presented as a "stick" or histogram plot. The most intense peak in the spectrum is used to normalize the other peaks in the spectrum, is referred to as the base peak, and is assigned to a 100% relative abundance. The intensities of the other peaks in the spectrum are expressed as percentages of the base peak, and plotted on the y-axis of a spectral plot. Rather than presenting an actual peak shape (in reality, approximately Gaussian if one examines the actual raw data), the abundance of an "ion" in a mass spectrum is represented as a single line. The graphical representation generated by a computer data system is often annotated with an "absolute signal intensity" corresponding to the intensity of the base peak so that absolute intensity information is not lost. This number is usually in "counts," a count being the signal intensity interval detected by the analog-to-digital conversion electronics. Different instruments will have different intervals, so that the actual meaning of the count value, in millivolts of signal for example, will have to be traced to the specific instrument. The number is useful in comparing intensities of different spectra.

The x-axis is the customary "mass" scale, more correctly noted as the "m/z" scale. Numerically, the values correspond to daltons (or atomic mass units) of mass. Most often, it is assumed to be an integral scale, peaks being centered on the whole number integers. On mass spectrometers with sufficient resolving power, and with "mass" assignments made with sufficient precision, positioning of peaks at other than integral mass values is correct, and constitutes useful information. Casual practice typically uses whole numbers for the masses of the atoms, while the only atom for which a true integer mass is correct is $^{12}C$, upon which the International Union of Pure and Applied Chemists (IUPAC) atomic mass unit is assigned. While the deviation from integral relative molecular mass is small for molecules with relative molecular masses of approximately 300 daltons or less, and is most often neglected, the significance of nonintegral masses

becomes more important as the molecular mass increases above 300 daltons. Large numbers of hydrogen atoms in the molecule accumulate excess mass. Sixty-five hydrogen atoms will add approximately 0.5 dalton above the integer to the relative molecular mass.

In 1991, Cooks and Rockwood (1991) championed a movement to clarify an ambiguity in referencing the typically presented x-axis in plots of mass spectra. The values along the x-axis have been casually referred to masses, and the axis the mass axis. Fortunately, the axis is generally labeled in print as the "m/z" axis. However, the frequent appearance in recent years of mass spectra, where ions are multiply charged or negatively charged, introduces semantic and algebraic inconsistencies. While most mass spectra concern ions that carry a single positive electrostatic charge, the m/z axis reduces arithmetically to be equivalent to the "mass" axis; no longer so with multiply charged ions, as in proteins. Negative ion spectra are typically presented in the same orientation as positive ions spectra, with the x-axis of the plot starting at zero on the lower left corner and with increasing absolute value of m/z proceeding toward the right, leading to an ambiguity which could lead to misinterpretation. Proposing the Thomson as a unit of measure (comparable to the Dalton), and incorporating the sign (positive or negative) from the charge state of the ion would remedy these ambiguities. Although a reasonable proposal, the suggestion has not yet caught hold.

### 8.2 Data Systems and Signal Processing: Making Sense of It

The ease of processing and displaying mass spectral data that we enjoy today is often taken for granted. Meyerson (1986) provides an enlightening view into the history of using "computer" systems to help with mass spectral data reduction and interpretation. Achieving an effective combination of computing power to control and reduce data from a combined GC-MS system warranted a full paper published in *Analytical Chemistry* in 1970 (Reynolds *et al.*, 1970). In the abstract, the authors claim spectral information to have been "acquired [and] made available to the chemist within minutes in an on-line graphic system." In the next year, Hertz *et al.* (Hertz, Hites and Biemann, 1971) reported success with the progenitor of the library search for identification of compounds by their mass spectra.

### 8.3 The Chromatography–Mass Spectrometry Data Model

In order to compensate for possible separatory shortcomings of the chromatography, a number of data-processing algorithms have been developed to make use of the fact that repetitive scan GC-MS data are really a three-dimensional matrix of data. The x-axis describes the time dependence of the separation. The y-axis describes the mass spectrum of the GC effluent at any "instant" in time, and the z-axis describes the intensity of the signal. Imagine then a cube, with the lower left corner of the front face assigned to be the origin of our data space. Along the bottom edge of the front face is the time axis (the x-axis). The left edge of the front face is the intensity axis (the y-axis), and the axis from front to back is the m/z axis (the z-axis). In general, this model applies to any repetitive scan chromatography–spectroscopy experiment.

The total ion current (TIC) chromatogram is a summation of all the mass intensities for each individual mass spectrum, plotted as a function of time. The TIC chromatogram can be visualized as a projection of all the intensities of the mass spectra onto the front surface of the data cube. (Strictly speaking, the projection of data onto the

front surface of a transparent cube would actually be the base peak intensity (BPI) chromatogram, in which the BPI of each spectrum is plotted as a function of time. Summing the intensities of all peaks in each spectrum, and plotting the sum as a function of time, is the TIC chromatogram.) Slicing the cube in a plane parallel to the front face and at a particular point on the m/z axis would expose a selected ion chromatogram of the time-dependent intensity of that particular m/z value. Slicing the cube orthogonal to the front face at any particular point in time would present on the exposed surface a mass spectrum of the GC effluent at that particular point in time.

### 8.4 Full (Repetitive) Scan Data Collection versus Selected Ion Recording

Full-scan experiments set the mass spectrometer in a cycle of repeatedly scanning a selected range of m/z values. The experiment is done primarily for structure elucidation, and the richness of structural information the fragmentation (induced or spontaneous) provides. It is relatively insensitive. Observing a chromatographic peak relies on sufficient accumulated signal in the mass spectrum of a component to be discernable above background. If background in a given experiment is high, components can be easily missed. Selected ion recording (SIR or SIM) is used for quantitation. The experiment shows greater sensitivity because one maximizes the time that one is looking at the signal that tells one about what one wants to know, and discards everything else. Consider that, in a repetitive scan experiment, a finite amount of time is required for the mass spectrometer to scan across a range of m/z values, typically 1000 daltons. One would envision spending equal amounts of time observing each of those nominal 1000 channels. Only a few (or even one) of those channels contain signals of interest—the molecular ion and/or fragments indicating the compound of interest. The remaining 99.9% of the time, one is looking elsewhere. By performing an SIR experiment, one is watching the information channel(s) containing the signal of interest nearly 100% of the time. One would expect to see sensitivity increases approaching three orders of magnitude. The price, of course, is that the maximum information is being measured on the targeted signal, but all other information about the composition of the original sample is being discarded.

## 9 INTERPRETATION OF MASS SPECTRAL DATA

### 9.1 Mass Spectral Fragmentation Compendia and Libraries

The concept of the organic functional group—that a particular organic functional group will behave in essentially the same way, regardless of other functional groups attached to it—holds true in the mass spectrometer, too. Meyerson (1986) gives insight into the beginnings of the revolutionary concept that mass spectral data correlate to structure of organic compounds.

Compilations of mass spectral fragmentation data for organic compounds are useful sources for help when interpreting mass spectra. Two of the better-known libraries are the NIST/EPA/MSDC mass spectral database, maintained by the United States National Institute of Standards and Technology, and the Wiley-McLafferty database, vended by John Wiley & Sons Publishing Company. These two differ in that the NIST collection contains only one spectrum representing a given compound. The Wiley-McLafferty collection contains a larger number of spectra, but includes multiple

spectra for some compounds, intending to capture information on variability of spectra from observation to observation and instrument to instrument. There are a number of smaller and more specialized collections scattered in the literature. The Thermodynamics Research Center at Texas A&M University still collects compounds and compiles data, which eventually are incorporated into the NIST and Wiley-McLafferty libraries.

In addition to the searchable library compilations, several compendial books on the EI fragmentation behavior of compounds have been published (Budzikiewicz, Djerassi and Williams, 1967; McLafferty and Venkataraghavan, 1982; Porter, 1985). They are dated, but nevertheless effectively capture the collective fragmentation information prior to their publication. Unfortunately, all of these information sources discuss EI spectra. EI fragmentation rules can be of limited assistance in interpreting MS-MS product ion spectra. One small collection of CI information has been published (Harrison, 1992). Fragmentation behavior of molecules under the soft ionization conditions—CI, FAB, API—still remains largely scattered throughout the chemical literature.

Compilation of spectral catalogs for the soft ionization techniques—CI, FAB, API—has not occurred, for at least two reasons. First, the exact nature of a soft ionization spectrum is much more dependent upon instrument and sampling conditions. Second, for many compounds analyzed by the soft ionization techniques, the spectrum consists principally of a molecular ion. A compilation of soft ionization spectra, in its ideal form, would be a compilation of molecular masses. Movements have surfaced several times over recent years to promote a compilation of MS-MS product ion spectra. These too are dependent upon exact operating conditions. An acceptable standard set of conditions has eluded everyone's efforts to make such a project happen.

All of these aspects for extracting structural information from mass spectral data come into play in some way when one is presented with a structure elucidation problem; however, this is not the complete set of tools and techniques. Finding that one obscure but revealing literature reference, flashes of insight in the middle of the night, snippets from one's collective personal and professional experiences, interactions with others practicing the art, are all ill-definable but nevertheless important contributors to this process. One can apply a degree of systematic practice, but structure elucidation from spectral (and other) data is still largely an empirical method, and still requires significant amounts of time simply staring at the data.

## 9.2 Molecular Weight versus Relative Molecular Mass

"Molecular weight" is an incorrect scientific term! The numerical value being discussed, rather, is more correctly known as a relative molecular mass. It is a mass, measuring a quantity of matter, not the influence of an external gravitational field on that matter. It is a relative mass, not an absolute mass measured in grams, because it is expressed relative to the standard atomic mass of 12.0000 atomic mass units (daltons), defined in reference to the $^{12}C$ isotope of carbon. It is a relative molecular mass because it is a sum of the relative atomic masses of the atoms contained within the molecule. The distinction between the incorrect term "atomic weight" and the correct term "relative atomic mass" follows the same logic, only applied to the atomic masses of the elements on the periodic table (Busch, 2004b).

The atomic mass unit is given the name "Dalton." The standard is defined as being 1/12 of the mass of an atom of the most abundant naturally occurring isotope of

carbon, $^{12}C$. The $^{12}C$ standard was adopted by the IUPAC in 1960 (Grayson, 2002). The previous standard had been assigned to be 1/16 of the atomic mass of oxygen. Two scales developed around this ambiguously defined oxygen standard, depending on whether one used the chemically determined atomic mass of oxygen, or the mass of the most abundant stable isotope, $^{16}O$. The chemically determined mass incorporated the natural abundances of the minor stable oxygen isotopes, $^{17}O$ (0.04%) and $^{18}O$ (0.2%). The other was based upon mass measurement specifically of $^{16}O$. The two scales are offset by 0.0044 daltons. Accurate masses reported prior to 1960 are referenced to the oxygen standard. Care should be used when comparing these numbers with more modern measurements. Note should be made to which of the two oxygen standards the older measurements are referenced. The extensive table of accurate mass values, calculated and published in Beynon's book (Beynon, 1960), is based upon the oxygen standard. If using those values, mass differences in that table should be correct, regardless of the original standard. Absolute masses should be corrected for the current definition.

## 9.3 Monoisotopic versus Chemical (Average) Relative Molecular Mass

The relative molecular mass that is measured by the mass spectrometer is numerically different from the chemical (average) molecular mass used to calculate molar concentrations or reaction stoichiometries. For the elements commonly found in organic molecules, the numbers are close and in many contexts interchangeable. However, when one discusses a mass spectrum, and an accurate mass measurement made by a mass spectrometer, the differences are noticeable. For a molecule that contains atoms of polyisotopic elements such as chlorine or bromine, or when the mass of the molecule gets large, the numbers become even more significantly different. One must exercise care when discussing molecular masses and be explicit about which kind of molecular mass number one is discussing. Two examples are calculated in Table 1. The normal hydrocarbon hexatriacontane, $C_{36}H_{74}$, has a nominal relative molecular mass of 506 daltons, an exact monoisotopic molecular mass of 506.5791 daltons, and a chemical

**TABLE 1 Example calculations.**

| Element | No. of atoms | Nominal atomic mass | Mass | Chemical atomic mass | Mass | Accurate atomic mass | Mass |
|---|---|---|---|---|---|---|---|
| Hexatriacontane | | | | | | | |
| C | 36 | 12 | 432 | 12.011 | 432.396 | 12.0000 | 432.0000 |
| H | 74 | 1 | 74 | 1.008 | 74.592 | 1.0078 | 74.5772 |
| r.m.m. | | | 506 | | 506.988 | | 506.5772 |
| Chlorinated compound | | | | | | | |
| C | 14 | 12 | 168 | 12.011 | 168.154 | 12.0000 | 168.0000 |
| H | 12 | 1 | 12 | 1.008 | 12.096 | 1.0078 | 12.0936 |
| N | 1 | 14 | 14 | 14.007 | 14.007 | 14.0031 | 14.0031 |
| O | 1 | 16 | 16 | 15.999 | 15.999 | 15.9949 | 15.9949 |
| F | 1 | 19 | 19 | 18.998 | 18.998 | 18.9984 | 18.9984 |
| $^{35}Cl$ | 2 | 35 | 70 | 35.450 | 70.900 | 34.9689 | 69.9378 |
| r.m.m. | | | 299 | | 300.154 | | 299.0278 |

molecular mass of 506.98 daltons. Note that the difference between the nominal molecular mass and the monoisotopic molecular mass of hexatriacontane derives from the large number of hydrogen atoms in the molecule. Hydrogen is a mass-sufficient element, implying that its monoisotopic accurate mass (1.007825 daltons) is larger than its nominal (integral) atomic mass of 1 dalton. The cumulative excess mass of this large number of hydrogen atoms accounts for the 0.58 dalton above the nominal molecular mass.[11] All three molecular masses for hexatriacontane are correct, when used in their proper contexts, but are substantially different from each other and can generate errors when improperly used.

The second example in Table 1 includes several additional elements. Nitrogen ($^{14}N$) is mass-sufficient, while oxygen ($^{16}O$), fluorine ($^{19}F$), and chlorine ($^{35}Cl$ in the table) are mass-deficient—having monoisotopic accurate masses smaller than their nominal (integral) atomic masses. The mass sufficiencies and deficiencies make small contributions to the discrepancies in the three kinds of molecular masses. The largest contributor, however, is the fact that chlorine is polyisotopic, with two naturally occurring isotopes of significant abundance, differing by 2 daltons in mass. The nominal and accurate mass calculations in Table 1 use only the mass for $^{35}Cl$. The chemical molecular mass calculation uses the periodic table mass, which is a weighted average of the masses and abundances of the $^{35}Cl$ and $^{37}Cl$ isotopes. See the discussion of polyisotopic elements and isotope patterns in the next section.

## 9.4 Polyisotopic Elements and Isotope Patterns

Thomson's discovery and characterization of the isotopes of neon (Griffiths, 1997), and subsequent work by Francis Aston[12] (Griffiths, 1997) (a student of Thomson's) and others, measuring the accurate masses of the isotopes, elaborated the isotopic diversity of the periodic table. Elements that have more than one naturally occurring isotope are called polyisotopic elements. The most familiar polyisotopic element to organic chemists is carbon, with two common stable isotopes—$^{12}C$ (98.9% natural abundance) and $^{13}C$ (1.1% natural abundance)[13] (see Table 2). Next are chlorine and bromine, each with two very abundant isotopes separated by 2 daltons. Although hydrogen, oxygen and nitrogen have well-known naturally occurring stable isotopes, they are of low natural abundance. In practice, hydrogen is treated as monoisotopic in small molecules, where the number of hydrogen atoms is relatively small. Deuterium is only 0.015% naturally abundant. The stable isotopes of oxygen and nitrogen are also of sufficiently low abundance—$^{17}O$, 0.038%; $^{18}O$, 0.2%; $^{15}N$, 0.37%—that, unless a molecule contains significant numbers of these heteroatoms, these elements do not contribute significantly to an isotope pattern. Fluorine and phosphorus are monoisotopic.[14] Sulfur is also a common polyisotopic element appearing in organic molecules. DeBievre and Barnes (1985) report the natural abundances of the isotopes of the elements. Abundances and accurate masses of the isotopes can also be found in the *Handbook*

[11] Sixty-four hydrogen atoms cumulatively contribute 0.5 dalton of additional mass above the integral mass.

[12] Aston was awarded the Nobel Prize in chemistry in 1922 (http://nobelprize.org/chemistry/laureates/1922/index.html).

[13] $^{14}C$ is a naturally occurring isotope, but is a radioactive (unstable) isotope. Its natural abundance is low enough not to be of concern in mass spectrometry, unless one makes special efforts, for example, using accelerator mass spectrometry.

[14] Naturally occurring stable isotopes! We are not considering $^{18}F$ and $^{32}P$ here.

*of Chemistry and Physics* . Masses and abundances of selected isotopes are given in Table 2.

In conventional mass spectra, isotope patterns deriving from the presence of poly-isotopic elements are striking and quite noticeable. An excellent example is the ESI spectrum shown in Fig. 9 for a compound with a $C_{14}H_{12}NOFCl_2$ empirical formula. The m/z 300 $[M+H]^+$ and the m/z 269 neutral loss fragment (resulting from neutral

**TABLE 2 Accurate masses and natural abundances of selected isotopes.**

| Isotope | Accurate mass | Natural abundance |
|---|---|---|
| $^{12}C$ | 12.0000 | 98.9 |
| $^{13}C$ | 13.00336 | 1.1 |
| $^{1}H$ | 1.007825 | 99.99 |
| $^{2}H$ | 2.0140 | 0.01 |
| $^{14}N$ | 14.00307 | 99.6 |
| $^{15}N$ | 15.0001 | 0.4 |
| $^{16}O$ | 15.99491 | 99.8 |
| $^{17}O$ | 16.9991 | 0.04 |
| $^{18}O$ | 17.99916 | 0.2 |
| $^{19}F$ | 18.9984 | 100.0 |
| $^{23}Na$ | 22.9898 | 100.0 |
| $^{31}P$ | 30.9738 | 100.0 |
| $^{32}S$ | 31.9721 | 95.0 |
| $^{34}S$ | 33.9679 | 4.2 |
| $^{35}Cl$ | 34.9689 | 75.8 |
| $^{37}Cl$ | 36.9659 | 24.2 |
| $^{79}Br$ | 78.9184 | 50.7 |
| $^{81}Br$ | 80.9163 | 49.3 |

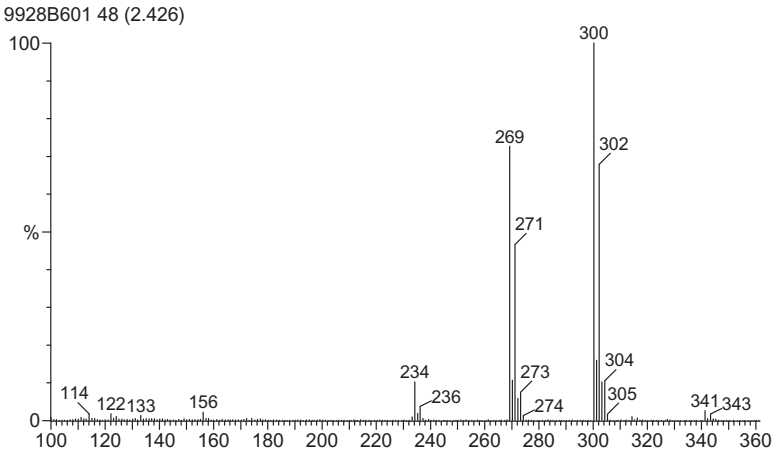


**Figure 9** ESI mass spectrum of a molecule containing two chlorine atoms.

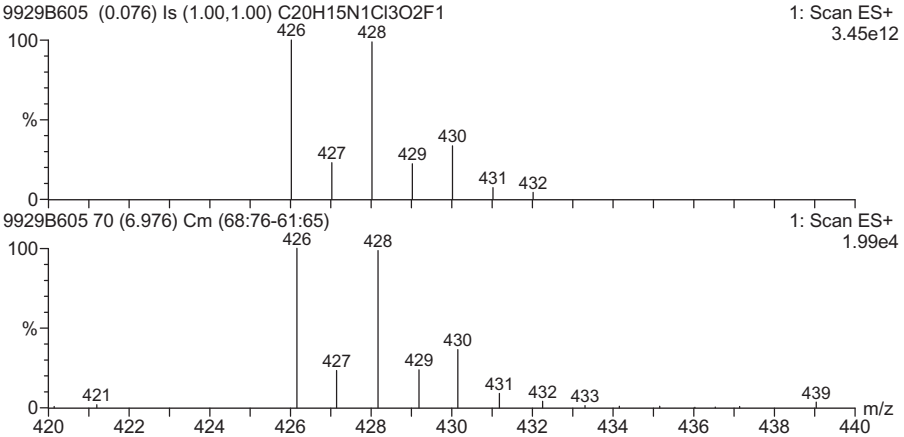


**Figure 10** Comparison of the isotope pattern for the observed $[M+H]^+$ (bottom panel) with the predicted model (top panel) for $C_{20}H_{15}NO_2FCl_3$.

loss of methylamine) clearly show $Cl_2$ isotope patterns which match well with simulations.[15] The m/z 234 fragment results from further loss of one of the chlorine atoms, illustrating a clear $Cl_1$ isotope pattern, and the utility of the isotope patterns in assigning structure.

The close match of simulation (Fig. 10, upper panel) and observation (Fig. 10, lower panel) of the $[M+H]^+$ of another chlorine-containing compound, along with some chemical considerations of the history of the sample, permitted assigning the structure of the unknown compound based solely upon the isotope pattern.

The most extensively polyisotopic element is tin, with 10 stable isotopes ranging in abundance from 0.4% to 33%, over a 12-dalton range. Figure 11 illustrates a dramatic example of an isotope pattern for monooctyl tin ethylhexylthioglycolate (Structure 1). The upper panel of Fig. 11 expands the cationated molecular ion region, $[M+Na]^+$, for this compound, taken from an LC-MS experiment conducted using APCI. Which species is the molecular ion? The answer is "all of them." The empirical formula for all of these species is $C_{38}H_{74}O_6S_3Sn_1Na$. While the chemical relative molecular mass for this compound is correctly calculated from the empirical formula to be 864.9 daltons (based upon periodic table atomic masses), the mass spectrometer reveals these 10 species, all of the same elemental composition but differing by isotopic content, all representing the molecule under discussion. The 38 carbon atoms and the three sulfur atoms contribute to this complex isotope pattern, but the primary contributor is the single tin atom.

Revisiting the question, "Which is the molecular ion species?" and calculating using isotopic masses, the measured values still bear further examination. For the most commonly encountered elements in organic molecules, the lowest mass isotope is the most abundant—for example, carbon, oxygen, sulfur, and chlorine. The calculated molecular mass using lowest-mass-isotope masses in general corresponds to the most abundant

[15] The well-done mass spectrometry data systems all have isotope pattern modeling programs built in. A number of other stand-alone simulation programs are also available. For example, see http://www.alchemistmatt.com/.

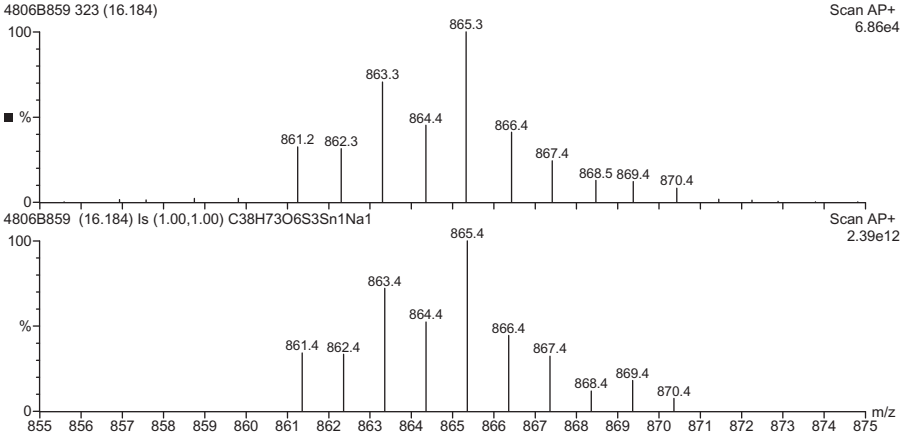


**Figure 11** Observed isotope pattern (upper panel) and predicted model (lower panel) for the $[M+Na]^+$ of dioctyl tin EHTG, taken from a positive ion APCI LC-MS experiment.

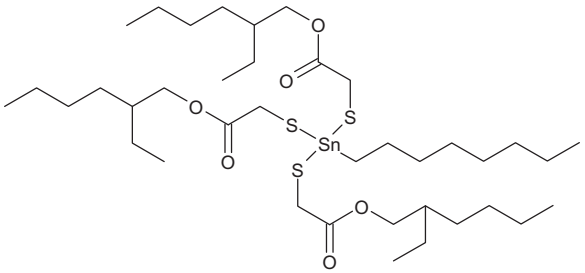


**Structure 1** Monooctyl tin ethylhexyl thioglycolate. $C_{38}H_{74}O_6S_3Sn$.

species observed. Using $^{112}Sn$, the species corresponding to this molecular ion should appear in Fig. 11 at m/z 857. Any signal at m/z 857 in the spectrum is insignificant. $^{112}Sn$ is only 1% abundant. The more appropriately designated molecular ion species here is the one at m/z 865, containing $^{120}Sn$ (the most abundant isotope of tin, at 33% natural abundance). The lower panel of Fig. 11 compares a theoretical prediction of the isotope pattern of this compound with the experimental observation. For complex isotope patterns such as these, one can (and should) simulate the isotope pattern with isotope pattern modeling software, available in the better mass spectrometer data systems, and in several data system-independent programs. When discussing molecular mass results, especially mass spectrometric results, for such polyisotope-containing molecules, one should specify how one is calculating results.

Further, the 865.4-dalton assignment for this species in Fig. 11 and the position at which the peak is plotted, are correct and not a drafting artifact. The only isotope whose mass is truly an integer is $^{12}C$, and only because it is arbitrarily assigned to be 12.0000 daltons. All other isotopic masses are either mass-sufficient or mass-deficient. The

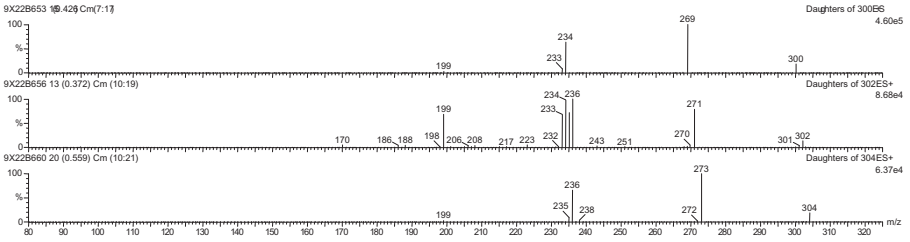


**Figure 12** MS-MS product ion spectra of the m/z 300 (top panel), m/z 302 (middle panel), and m/z 304 (bottom panel) $[M+H]^+$.

relative molecular mass of the m/z 865 species truly is 865.36 daltons. The sufficiency results from the accumulated effect of the 74 hydrogen atoms,[16] offset by the deficiency contributed by all of the other mass-deficient elements (excluding carbon) in the empirical formula.

The consideration of molecular mass is further complicated by the existence of a number of significantly polyisotopic elements. Given this complication, determination (and reporting) of the correct molecular mass is controversial. Polyisotopic elements immediately divide practitioners into two camps. One camp says the molecular mass is that value calculated using the lowest mass isotope of each element in the molecule. The other camp says to calculate using the mass of the most abundant isotope of each element. In most instances, these two rules give rise to the same number. The most abundant isotopes of carbon, hydrogen, oxygen, nitrogen, sulfur, phosphorus, and all of the halogens are indeed the lowest mass isotope. Boron-containing molecules or organometallic compounds containing a polyisotopic transition metal or heavy metal may more appropriately deserve application of the second rule (Sharp *et al.*, 1986; Kowalski, Sharp and Stang, 1987). Applying either of these rules, however, becomes even more complicated when a molecule contains, for example, three or more chlorine atoms, two or more bromine atoms, or more than one polyisotopic metal atom (such as tin).

The interpretation of isotope patterns changes when applied to MS-MS product ion spectra. Consider the product ion spectrum, in the top panel of Fig. 12, of the same m/z 300 $[M+H]^+$ discussed earlier. The experiment was performed on a triple quadrupole instrument by selecting the m/z 300 species with the first analyzer under unit resolution conditions. The m/z 300 species was collisionally activated in the second quadrupole collision cell. The product ion spectrum resulting from CID was recorded in the third quadrupole mass analyzer. Recall that the empirical formula of this molecular ion is $C_{14}H_{13}NOFCl_2$. Why have the very prominent isotope patterns disappeared? Because one has selected the m/z 300 species as the parent, there is no longer any consideration of polyisotopy. This parent is a monoisotopic parent—namely $C_{14}H_{13}NOF^{35}Cl_2$. Methylamine neutral loss produces a single species—m/z 269. Further fragmentation, sequentially losing chlorine atoms, produces single species—m/z 234 and m/z 199—because the only chlorine isotope that can be lost from this selected m/z 300 parent is a $^{35}Cl$ atom.

[16] Sixty-four hydrogen atoms cumulatively contribute 0.5 dalton of additional mass above the integral mass.

Why does an isotope pattern reappear in the product ion spectrum of the m/z 302 parent (Fig. 12, middle panel)? The elemental composition of the parent is now $C_{14}H_{13}NOF^{35}Cl_1{}^{37}Cl_1$. The m/z 271 species shows no polyisotopic signature because its formation results from loss of $CH_3NH_2$ from the parent to give $C_{13}H_8OF^{35}Cl_1{}^{37}Cl_1$—only one possible elemental composition. Loss of the first chlorine atom, however, now permits producing two possible elemental compositions—$C_{13}H_8OF^{35}Cl_1$ and $C_{13}H_8OF^{37}Cl_1$ at m/z 234 and m/z 236. They are of equal abundance because of the equal probability of losing either a $^{35}Cl$ or a $^{37}Cl$ atom in the fragmentation process. The fact that the natural abundances of the chlorine isotopes are approximately 3 to 1 is irrelevant here. Selecting the m/z 302 species as the parent sets the isotope ratio of $^{35}Cl$ to $^{37}Cl$ in this experiment to be 1 to 1. The isotope pattern disappears on losing the second chlorine to give the m/z 199 species, since the source of polyisotopy is now completely gone.

All evidence of isotope patterns in the product ion spectrum of the m/z 304 parent (Fig. 12, bottom panel) disappears because the selected parent is monoisotopic—$C_{14}H_{13}NOF^{37}Cl_2$. Justification for the absence of isotope patterns is identical to that of interpretation of the m/z 300 product ion spectrum. The detailed explanation of this excellent example serves to illustrate the care one needs to exercise in interpreting isotope patterns (or the lack thereof) in MS-MS spectra. At the same time, it demonstrates the structure-indicating power of careful interpretation.

With the recent development of the combined quadrupole TOF tandem instrument geometry, MS-MS product ion spectra are starting to appear in the literature with isotope patterns reflecting the natural abundance of the involved elements. Considering this discussion, how does this occur? Note that our selection of parent ions for CID in the experiments described earlier was done under unit resolution conditions. Only species with a single (although nominal) mass are permitted to pass into the collision cell—thus the argument of above, justifying the selection of a specific isotope content. With a tandem instrument composed of a quadrupole and a TOF mass analyzer, the resolution of the first mass analyzer can be set to substantially less than unit resolution, such that the entire isotope pattern can be selected as the "parent." This approach preserves the isotope pattern information while providing MS-MS product ion information. This new consideration must be added while interpreting isotope patterns in MS-MS spectra—namely to identify on what kind of instrument the experiment is being done, and how the instrument is set up.

## 9.5 High-Resolution Measurements versus Accurate Mass Measurements

The accurate mass measurement technique is being actively rediscovered because of recent improvements in instrumentation. John Beynon (Beynon, 1954, 1960) established the validity of using high-precision mass measurements (to at least four decimal places of precision) to determine elemental compositions of ions with high certainty, capitalizing on the nonintegral nature of the precise atomic masses of the isotopes. Accurate mass measurements were until recently the nearly exclusive province of magnetic sector instruments. Making such measurements presented difficulties, and required careful attention to detail. rTOF instruments and FT-MS instruments have challenged magnetic sector dominance of this area and facilitated a somewhat easier route to making these measurements. Accurate mass measurement became accepted in the synthetic organic chemistry community as an appropriate measure of the "proof" of a structure. Acceptance criteria for including accurate mass measurements in the

suite of data defending structure assignment of a newly synthesized compound are given in the instructions for authors of the synthetic and other journals,[17] even though Clayton Heathcock (Heathcock, 1990) (then editor-in-chief of the *Journal of Organic Chemistry*) acknowledged that combustion analysis, for which accurate mass measurements are often used as a substitute, is really a measure of purity and not of identity. Accurate mass measurements are a measure of identity.

Biemann's discussion (Biemann, 1990) defines the circumstances and limitations very carefully, where one should and should not use these kinds of measurements, and illustrates with examples applied to the elucidation of fragmentation pathways (the least demanding application), evidence supporting the confirmation of structural assignments and the determination of the structure of natural products (the most demanding application). Busch (1994) discusses the utility of an accurate mass measurement in confirming the identification of a phthalate contaminant as an example. Gilliam *et al.* (Gilliam, Landis and Occolowitz, 1983, 1984) report on procedures and practices for making accurate mass measurements in a production type of environment, and the utility the measurements provide in a pharmaceutical discovery setting.

Improvements in FT-MS instrumentation have permitted a re-evaluation of the value of accurate mass measurements to the analysis of very complex mixtures. Alan Marshall's group's work, for example, on vegetable oils (Wu, Rodgers and Marshall, 2004) and on crude and refined petroleum fractions (Marshall and Rodgers, 2004) has resurrected concepts that were developed in the early 1950s when mass spectrometry was being vigorously applied in the petroleum industry. The very high resolving power accessible on a high magnetic field FT-MS instrument emphasizes the potential elemental composition heterogeneity that can exist at a single mass in these complex mixtures, and demonstrates how much structural and compositional information can be extracted from a single mass spectrum. The mass deficiency of a single oxygen or sulfur atom imparts enough difference to distinguish compositions containing these elements from the more abundant, and more readily expected hydrocarbon compositions in a diesel oil sample (Fig. 13). A veritable forest of compositions appears in the negative ion ESI spectrum of olive oil, due to a large number of oxygen atoms present in the compositions (Fig. 14).

A little recognized systematic error in the calculation of accurate masses of, for example, small radical cation molecular ions (as in EI) or protonated molecular ions (as seen in the soft ionization methods) is the fact that the electron has a small but finite mass. The accurate masses of radical cations, in which a valence electron has been removed, of anions that have been created by capture of an electron, and of protonated species produced by soft ionization processes, should take into consideration this small mass difference (Mamer and Lesimple, 2004). For example, there is a small difference between the relative atomic mass of a neutral hydrogen atom and a proton. The accepted accurate mass of $^{1}H^{0}$ is 1.007825 daltons. The accurate mass of $^{1}H^{+}$ is 1.007276 daltons. To be completely correct, expected accurate masses of protonated molecular ions, $[M+H]^{+}$, produced by electrospray should be calculated using the mass of one $H^{+}$, rather than all of neutral hydrogen atoms. Mamer and Lesimple do acknowledge, however, that, for large molecules, the error is of little consequence.

Notwithstanding the efforts of groups attempting to precisely define acceptable practices and requirements (Gross, 1994), confusion still exists on the differences

[17] Please consult, for example, recent instructions for authors in the *Journal of Organic Chemistry* or the *Journal of the American Society for Mass Spectrometry*.

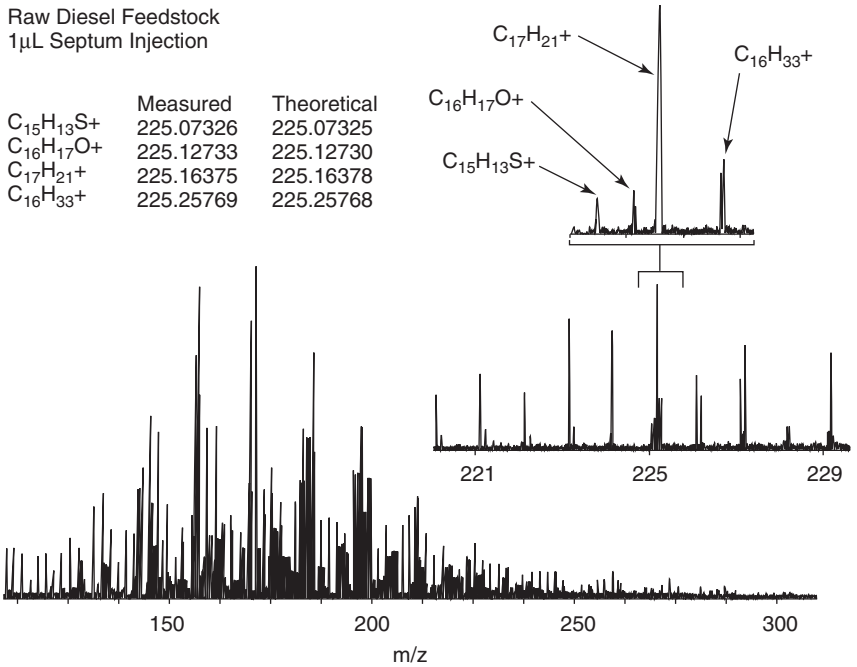


**Figure 13** Example high-resolution experiment on complex mixtures, showing the multiple elemental compositions possible at a single nominal mass. Four elemental compositions at m/z 225 in the positive ion electrospray mass spectrum of a diesel oil feedstock (from Wu, Rodgers and Marshall, 2004).

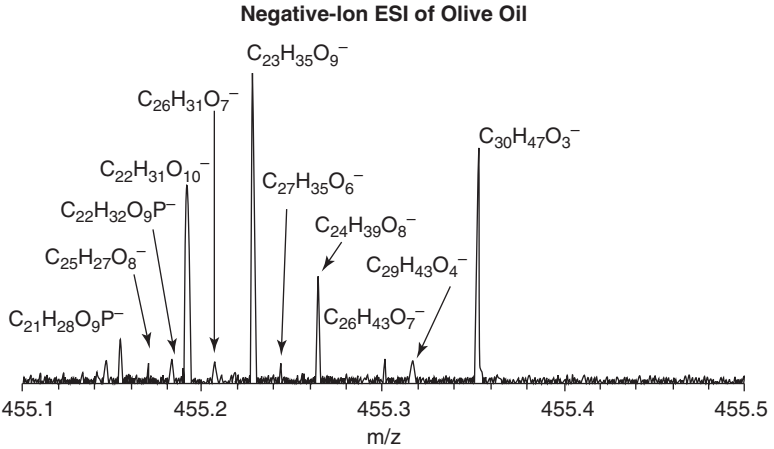


**Figure 14** Example high-resolution experiment on complex mixtures, showing the multiple elemental compositions possible at a single nominal mass. Multiple elemental compositions at m/z 455 in the negative ion electrospray spectrum of olive oil (from Marshall and Rodgers, 2004).

between definition of the terms "high resolution" and "accurate mass" (Kondrat, 1999). High resolution, as in any context, implies the ability to make confident and reliable measurements of small differences. In mass spectrometry, high resolution implies the ability to measure the small differences deriving from different elemental compositions that have the same nominal mass (see the examples in Figs 13 and 14). Measurement can be made at sufficiently high-enough resolution to see the heterogeneity of compositions at a given nominal mass without necessarily being able to assign the precise accurate masses of those compositions. The partial mass spectra shown in Fig. 13 truly are high-resolution spectra. Conversely, accurate mass measurement refers to the ability to confidently assign a precise relative molecular mass (customarily four decimal places or more) to a signal. Accurate mass measurements can be made at low resolution, given a stable, reproducible scan by the instrument, appropriate accurate mass reference standards, and the confidence that one is observing a mass spectral peak composed of a single elemental composition (Tyler, Clayton and Green, 1996). The spectra in Fig. 13 are additionally accurate mass-assigned spectra because of the care taken in making the mass assignments. Indeed, Beynon's original demonstration (Beynon, 1954) of the utility of calculating elemental compositions from accurate mass measurements was done on a single-focusing magnetic sector instrument at a resolving power of only 250, while they were building an instrument intending to achieve a 2500 resolving power to enable them to make better measurements. Note these low resolving powers, in contrast to the 5000 to 15,000 resolving powers used by Gilliam *et al.* (Gilliam, Landis and Occolowitz, 1983, 1984) and the 350,000 resolving power used by Marshall *et al.* (Marshall and Rodgers, 2004) and Wu *et al.* (Wu, Rodgers and Marshall, 2004).

The instrument manufacturers often imply that accurate mass measurements (and therefore elemental composition determinations) are a single measurement solution to many structural problems. This is an oversimplification. A recent report, stemming from a metabolomics application, presents data that high mass measurements, even at sub-1 ppm accuracy, are insufficient to "identify" metabolites (Kind and Fiehn, 2006). Isotopic abundance pattern analysis is a necessary additional constraint to eliminate false elemental composition candidates.

## 9.6 Unimolecular Decompositions versus Granddaughter[18] Ions

The theoretical concept behind MS-MS involves unimolecular decomposition to establish ion-genetic relationships between fragments and their parent structures. The difficulty of interpreting a conventional mass spectrum consisting of more than one component is determining which fragment belongs to which parent. Another problem occurs when a fragment gives rise to another fragment via a sequence of fragmentations. This dilemma is tremendously improved by having a chromatographic separation on the front of the experiment, giving some assurance that one is looking at the mass

[18] Some individuals have taken offense at the usage of the originally coined term "daughter ion" in describing the ion-genetic relationships resulting from a collision-induced decomposition experiment. The more palatable term "product ion" has come into use as an alternative. We have attempted to refer to product ions in our discussions. Notwithstanding, the term "granddaughter" conveys a certain meaning about the skipping-of-generations relationship between products formed by two or more sequential chemical reactions from the original parent. No comparable companion term to "product ion" has yet arisen which conveys this meaning. For this reason, we still use the term "granddaughter ion" here in this limited context.

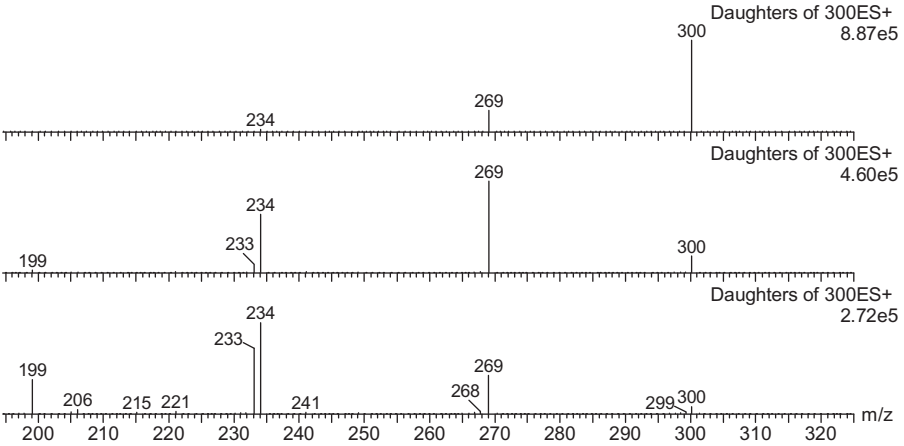


**Figure 15** MS-MS product ion spectra of the m/z 300 $[M+H]^+$ at 10 volts (top panel), 20 volts (middle panel), and 30 volts (bottom panel) interface cone voltages.

spectrum of a single compound, and that all fragments are ultimately derived from one parent. This is not the case when one still has co-eluting components. MS-MS experiments seek to eliminate this uncertainty. These experiments assume an "infinitely thin" collision zone and single collisions of the parent with a collision gas, depositing only sufficient energy to induce one unimolecular decomposition reaction. These assumptions are not sustained when utilizing a physically large quadrupole collision cell. Multiple collisions at high energy are possible. Product ions can also undergo collisional activation before they leave the collision cell. Granddaughter and great-granddaughter ions can be formed and detected in the MS-MS spectrum. How does one tell unimolecular decomposition from multistep decomposition? Fragmentations requiring the breaking of multiple bonds at multiple sites in the molecule can indicate multistep decomposition.

Examination of the collision energy dependence of the product ion spectrum is a functional test. To illustrate, the product ion spectra of the m/z 300 $[M+H]^+$ of a molecule of empirical formula $C_{14}H_{12}NOFCl_2$, at three different collision energies, are shown in Fig. 15. This is the same molecule on which the isotope pattern discussion in Section 9.4 is based. The top panel, at low collision energy, shows significant survival of the $[M+H]^+$ and loss of a 31-dalton species, corresponding to neutral loss of methylamine—a sensible fragmentation, as the original molecule contains an aliphatic *N*-methylamine side chain. Only just appearing is an m/z 234 fragment, corresponding to further loss of one of the chlorine atoms. That the methylamine and chlorine moieties are distant from each other in the original structure implies that these are two separate events, and that the m/z 269 species is losing the chlorine. If chlorine were to be lost directly from the parent ion, an m/z 265 species would be present, and it is not.

Increasing the collision energy, as illustrated in the two lower panels of Fig. 15 further destroys the $[M+H]^+$ and induces more extensive fragmentation of the m/z 269 species. It increases the relative abundance of the first chlorine loss, and induces a further loss of the second chlorine atom, producing the m/z 199 species. The collision energy dependence on the abundance and appearance of the product ions indicates

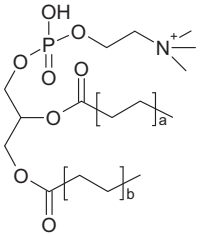


**Structure 2** Lecithin general structure (phosphatidylcholine) for a, b = 7, dipalmitoylphosphatidylcholine. $C_{40}H_{81}NO_8P$; r.m.m. 734.5700.

that these species are the result of a sequence of fragmentations, and helps one correlate back to the original structure of the molecule.

## 9.7 When Is a Molecular Ion Not an $[M+H]^+$?

A molecular ion species should not always be automatically assumed to be a protonated species, one mass unit heavier than the neutral molecular species. Cationic species, such as quaternary ammonium compounds, will appear in the mass spectrum at the mass of the cationic species. A biologically relevant example is the family of phosphatidylcholines (also known as the lecithins), represented by Structure 2. Mass spectra of these compounds, recorded under appropriate conditions, show $M^+$ species, not $[M+H]^+$. The molecular ion species for the dipalmitoyl phosphatidylcholine species, whose empirical formula and relative molecular mass are given with Structure 2, would appear at m/z 734.6. Note also the 0.6 dalton of excess mass above the integral value because of the large number of mass-sufficient hydrogen atoms in this molecule.

## 9.8 Deuterium Exchange Experiments

Substantial amounts of work appear in the literature in recent years on the study of backbone amide hydrogen–deuterium exchange studies of protein folding, based upon the pH dependence of the rates of H-D exchange of amide protons (Bai *et al.*, 1993). The rate minimum for this exchange is between pH 2 and pH 3 at 0 °C. Half-lives for exchange can be greater than 1 hour at these conditions for unprotected amides. This slowness of exchange permits conducting exchange experiments on intact proteins, then cleaving those proteins with acid-active peptidases and determining the regions of the original protein accessible to exchange. These constitute solution state probes of tertiary and higher structure of proteins. Review papers by Englander and colleagues (Krishna *et al.*, 2004; Englander, 2006) describe this important technique.

The utility of deuterium exchange, however, is not limited to large molecules. Identifying and confirming the structures of small molecules can benefit considerably by determining the numbers (and nature) of exchangeable hydrogen atoms in the molecule. Olsen *et al.* (2000) report on the general utility of deuterium exchange LC-MS experiments for structure elucidation and impurity identification. However, one should not forget the importance of the chemical nature of the exchanging hydrogen. Exchange is not instantaneous and universal for exchangeable hydrogen atoms. Such

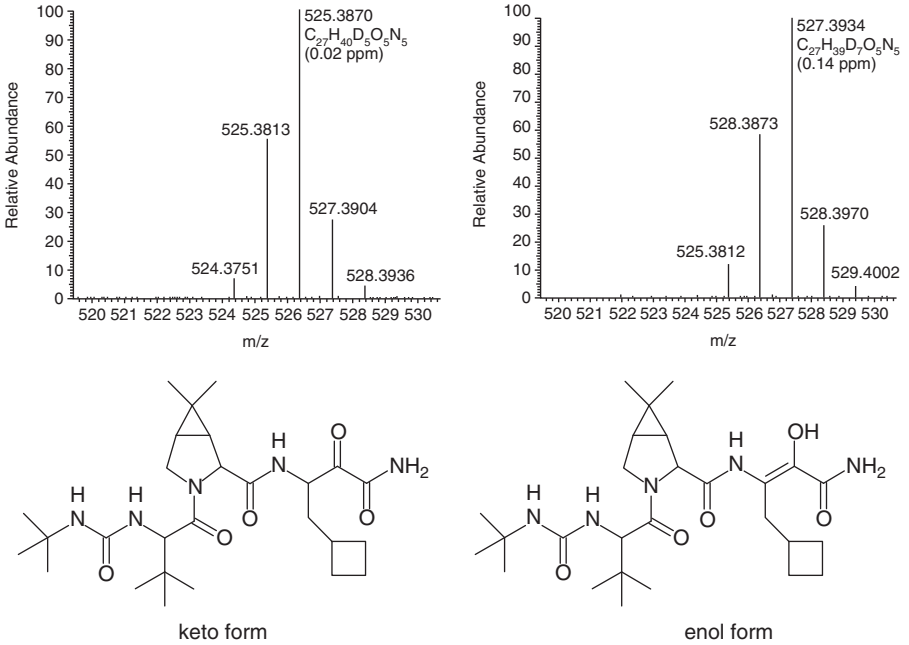


**Figure 16** Differentiation of keto (left structure and mass spectrum) and enol (right structure and mass spectrum) by hydrogen–deuterium exchange (from Chen *et al.*, 2007).

a serendipitous observation in the identification of a process-related impurity (Alsante *et al.*, 2004) in a pharmaceutical drug candidate exemplifies the benefit of an inventory of exchangeable hydrogen atoms in an impurity identification.

An example where deuterium exchange played an important role to imply a keto-enol tautomerization in a pharmaceutical drug substance has been reported (Chen *et al.*, 2007). An impurity exhibiting the same relative molecular mass as the drug substance was observed. Conducting a deuterium exchange LC-MS experiment showed that the drug substance contained five exchangeable hydrogen atoms, while the impurity contained six exchangeable hydrogens, the structural explanation being that the two molecular forms were keto and enol tautomers. The enol tautomer contained an additional hydroxyl group with a readily exchangeable hydrogen (Fig. 16).

The DART ionization source permits performing even more facile hydrogen–deuterium exchange experiments. Investigations into the contamination of pet foods by melamine are reported by Vail *et al.* (Vail, Jones and Sparkman, 2007). The hydrogen–deuterium exchange was conducted simply by placing a source of $D_2O$ vapors in proximity with the ion source and the sample. Exchange was effected in the gas phase between the melamine molecules and the $D_2O$ (see Fig. 17). The six exchangeable hydrogen atoms in the melamine molecule plus the additional deuterium to form the $[M+D]^+$ are accounted for in the spectrum of the exchanged compound. The exchanged spectrum in Fig. 17 points out a concern. Is the deuterium incorporation chemically specific for exchangeable hydrogen atoms or nonspecific? If exchange were to be demonstrated to stop at seven and only seven hydrogens, then exchange is chemically specific. If the molecule could be shown to incorporate more than seven

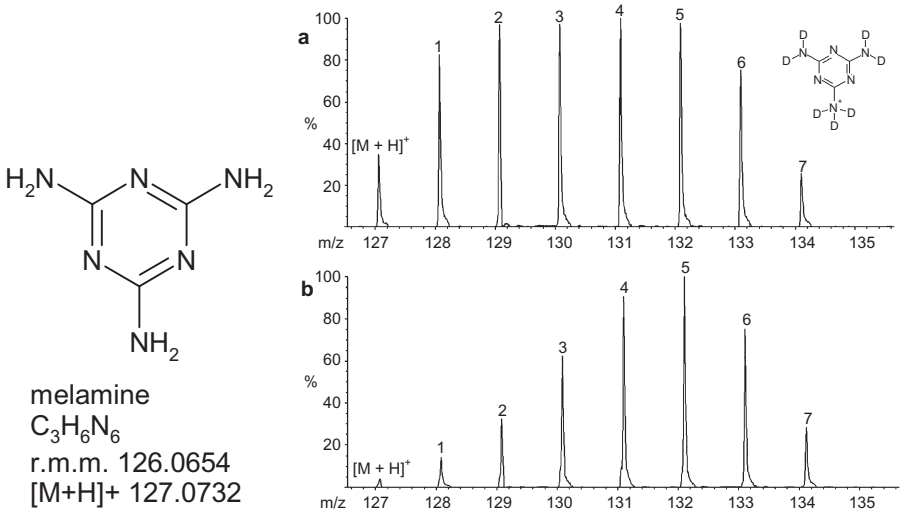


**Figure 17** Hydrogen-deuterium exchange of melamine in a DART ion source. **(a)** Standard melamine in the presence of deuterium oxide. **(b)** Melamine detected in contaminated pet food in the presence of deuterium oxide. From Vail *et al.* 2007

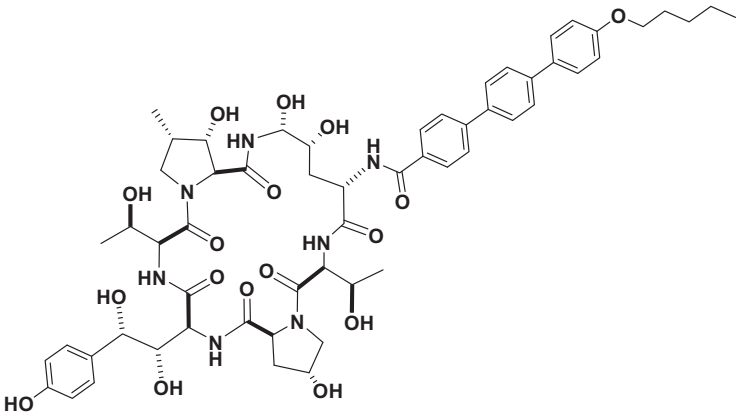

**Structure 3** Anidulafungin. $C_{58}H_{73}N_7O_{17}$; r.m.m. 1139.5063.

deuterium atoms, then exchange would not have the specific chemical implication implied here. While not shown here, communications with the authors and with R. Cody (R.B.Cody,2007,personal communications) indicate that inducing more complete hydrogen–deuterium exchange in melamine and in other compounds used to test this phenomenon indeed shows that exchange is specific and consistent with the chemical nature of the hydrogen atoms involved.

Anidulafungin (Structure 3) (Cappelletty and Eiselstein-McKitrick, 2007) is a semi-synthetic antifungal compound, derived from a natural product. The molecular ion region of the positive ion electrospray mass spectrum of this compound is shown in Fig. 18 (Z. Zhang and R. Morris, 2007, personal communications). The expected monoisotopic $[M+H]^+$ is observed at m/z 1140.9. (The mass calibration in Fig. 18 is

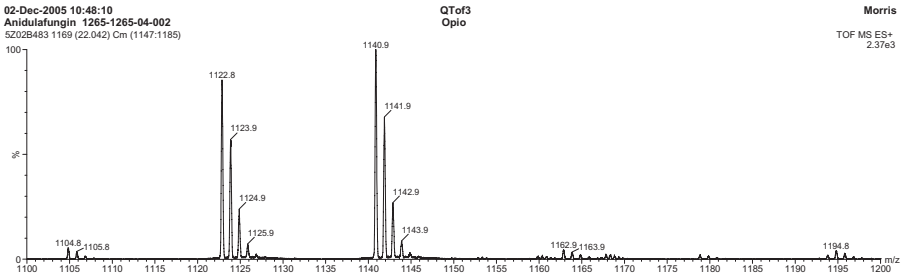


**Figure 18** Positive ion electrospray mass spectrum of the molecular ion region of anidulafungin.

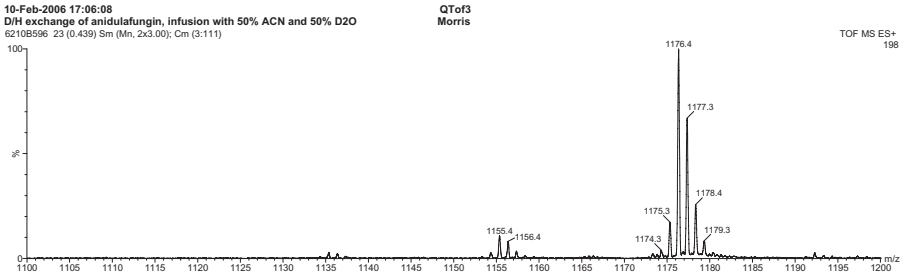


**Figure 19** Deuterium exchanged molecular ion region of anidulafungin.

off by ~0.4 daltons!) A small sodiated molecular ion, $[M+Na]^+$, is observed at m/z 1162.9, resulting from this molecule's scavenging $Na^+$ from the instrument system. The species at m/z 1122.8 and m/z 1104.8 result from mass spectrometrically induced dehydrations of the molecule, of which there are ample opportunities in this molecule. Isotopic abundances are consistent with expectations for this empirical formula.

Evidence supporting the correctness of the anidulafungin structure is available by hydrogen–deuterium exchange experiments. The 14 hydrogen atoms explicitly shown in Structure 3—a phenolic hydrogen, five amide hydrogen atoms, and eight aliphatic alcohol hydrogens—exchange in an aqueous environment. By dissolving anidulafungin in $D_2O$, the mass spectrum of this thoroughly exchanged material is shown in Fig. 19. The m/z 1176.4 species is the $[M+Na]^+$. The isotope pattern suggests that deuterium exchange is not 100%, because of the significant abundances at m/z 1175.3 and m/z 1174.3. However, the mass of the $[M+Na]^+$ has moved up 14 daltons, corresponding to deuterium exchange of the 14 exchangeable hydrogens. The m/z 1155.4 species is the $[M+H]^+$ with exchanged hydrogens. Recall that the reason it has moved up 15 daltons is that the charge carrier, an $H^+$, has also been replaced by a $D^+$ under the exchange conditions—in actual fact, an $[M+D]^+$.

## 9.9 Cluster Ions and Adduct Ions

The quest for an ionization method that guarantees a molecular ion for all compounds has driven the development of softer and softer methods. While approaching this goal,

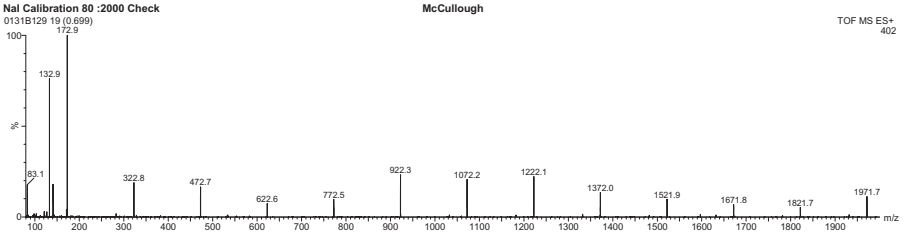


**Figure 20** Sodium iodide cluster spectrum by electrospray.

a wealth of new exceptions to the rule has been observed. Cluster ions and adduct ions form in the matrix (for FAB ionization) or in the spray of the API methods depending on the nature of the compound, the matrix, solvent, or extraneous salts present in the sample or in the instrument. While complicating the interpretation of such results and requiring the appropriate care in interpretation, the presence of these species can be useful (Fig. 20).

The ubiquitous nature of the sodium cation often leads to $Na^+$ cationization and the presence of $[M+Na]^+$ in the mass spectrum in addition to, or in place of, the $[M+H]^+$. Efforts to either remove or add $Na^+$ from the sample result in changes in the relative abundances of the $[M+H]^+$ and $[M+Na]^+$. Addition will often cause replacement of other labile protons with additional sodium atoms. Addition of lithium or potassium can replace the Na, with the appropriate mass changes, to confirm the identity of the Na adducts. Ammonium cationization will often occur from LC-MS mobile phases containing an ammonium buffer salt. Acetonitrile solvent adducts will also often form.

An example of a molecule that exhibits an extreme dependence on solvent conditions is shown in the partial spectra of Fig. 21. The molecule has a 404-dalton relative molecular mass. The material is produced as the sodium salt. Samples of the exact same material have been dissolved in and introduced into the electrospray ion source using three different solvents. The m/z 405 species is the $[M+H]^+$. The m/z 422 species is an $[M+NH_4]^+$. The m/z 427 species is [M+Na]. The m/z 468 species corresponds to an [M+Na+actronitrile]$^+$. Examining any of these three spectra in isolation, one cannot easily tell which is the real $[M+H]^+$. However, once one knows that a particular class of molecule—such as this one—behaves in this manner, one can use the presence of the cluster ions as additional supporting evidence for assigning the molecular mass of the compound. This phenomenon adds uncertainty about assuming the behavior of unknowns. If one can safely assume that an unknown impurity or degradant is sufficiently structurally similar to the parent molecule to behave in the same manner, then the adduct formation helps to confirm molecular mass assignments. Looking for 17 daltons ($NH_4^+$), 22 daltons ($Na^+$), 38 daltons ($K^+$), and 41 daltons (acetonitrile) mass differences, or various combinations of these numbers, will provide some level of comfort in assuming adduction behavior.

HPLC columns and instruments that have been used for numerous projects and been exposed to numerous mobile phases will often be heavily "contaminated" with extraneous buffer and other cations and anions. Some organic molecule structures show a propensity to scavenge such stray ions from the instrument and column. From where does the $Na^+$ in an $[M+Na]^+$ derive if neither the sample nor the mobile phase expressly contains $Na^+$? Stray $Na^+$ or other ions are ubiquitous. In negative ion detec-

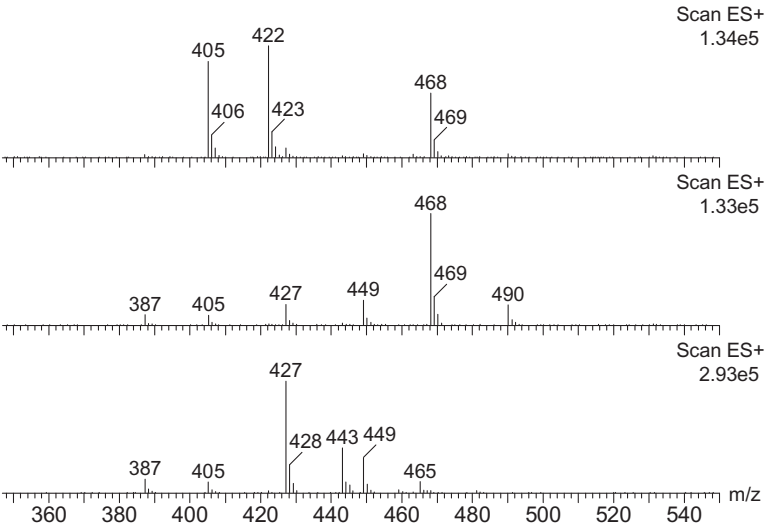


**Figure 21** Formation of different adduct ions under different solvent conditions. Samples of a 404-dalton molecule dissolved in and introduced in 1:1 acetonitrile:20 mM aqueous ammonium acetate (top panel), 1:1 acetonitrile:water (middle panel), and 1:1 methanol:water (bottom panel).

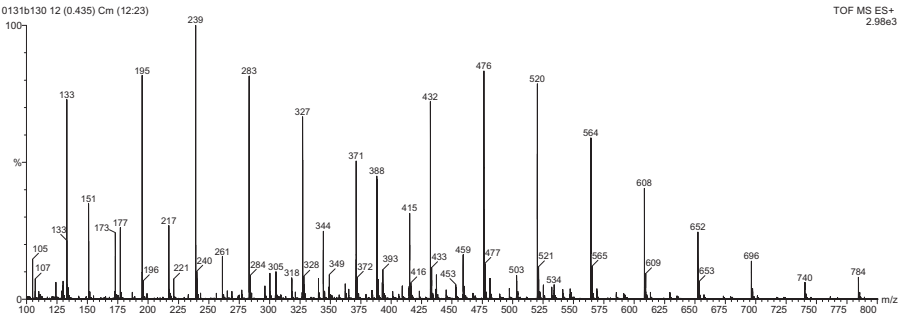


**Figure 22** Positive ion ESI mass spectrum of a poly(ethylene glycol) mixture.

tion mode, the presence of $[M\text{-}H+114]^+$ adducts derives from adduction with trifluoroacetic acid, an often-used mobile phase adulterant to produce low-pH mobile phases. Residual trifluoroacetic acid often abounds in mass spectrometers.

## 9.10 Polymer Patterns

Another striking feature that can be seen in the overall appearance of a mass spectrum is a repeating series of peaks at a regular interval—a "polymer pattern." It is almost as striking as an isotope pattern. Such a pattern is illustrated by the spectrum shown in Fig. 22, an ESI mass spectrum of a mixture of poly(ethylene glycol) molecules—PEG (Structure 4). Various PEG mixtures and similar polymers are frequently used as

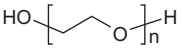


**Structure 4** Poly(ethylene glycol). $(C_2H_4O)_nH_2O$; r.m.m. 44N + 18.

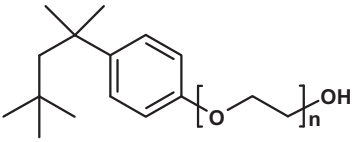


**Structure 5** Octoxynol "Triton X". $C_8H_{17}$ $C_6H_4$ $[C_2H_4O]_nOH$; r.m.m. 206 + $[44]_n$.

mass calibrants and accurate mass reference standards. The spectrum shown is of a mixture of two commercially available PEG preparations of 200- and 400-dalton average molecular masses.[19] The constant interval between the repeating peaks is 44 daltons, the mass of the PEG repeating unit. The m/z 239 peak is the $[M+H]^+$ of a PEG molecule with five repeating units—$PEG_5$. The m/z 261 peak is the $[M+Na]^+$ of $PEG_5$. The m/z 476 species is the $[M+NH_4]^+$ of $PEG_{10}$. The sample solution was prepared by dissolving the PEG in aqueous ammonium acetate, providing a source of ammonium cation. This spectrum illustrates the interesting phenomenon of $Na^+$ cationization, and of preferential $NH_4^+$ cationization by the larger PEG polymers. Recognizing these features, the complicated appearing spectrum can be easily interpreted. One can visually pick out the components of each series, and estimate the average molecular masses of the original PEG starting materials. An obvious PEG polymer pattern was key in identifying a drug-excipient interaction in a formulation containing PEG (Fig. 24). The drug substance, a carboxylic acid, formed a series of esters with the terminal alcohol group of the PEG molecules. The mass spectrum clearly showed this pattern of peaks, separated by 44 daltons, with the masses offset from those of PEG by the mass of the drug substance.

A positive ion electrospray mass spectrum of a sample of octoxynol (1,1,3,3-tetramethyl)butylphenyl (poly)ethylene glycol, Triton X-100, Structure 5) (Octoxynol. Merck Index, 12th edn. Entry # 6858) is shown in Fig. 23. The characteristic polymer pattern is prominent. The peaks in the spectrum are a series of ammoniated molecular ions, $[M+NH_4]^+$. This surfactant molecule is efficient at scavenging ammonium cation from the instrument system.

Table 3 calculates the average relative molecular mass for the triton sample recorded in Fig. 23. The number contribution is the product of the chain length and the abundance of a chain length (simple peak height from the spectrum). The average chain length, 9.5 units, is the ratio of the sum of number contributions divided by the sum of contributions (proportional to the number of molecules in the sample). The average molecular mass of the polymer sample, 644 daltons, can be calculated from the average chain length. Visual inspection of the spectrum, suggesting a center of the distribution between m/z 620 and m/z 664, is consistent with the results of the calculation. The distribution width can be used as a measure of polymer "polydispersity"—in this case, a chain length of 9.5 units ±2.7.

[19] Note here that the term "average molecular mass" applied to a polymer implies yet another nuance in the usage of the term "molecular mass."

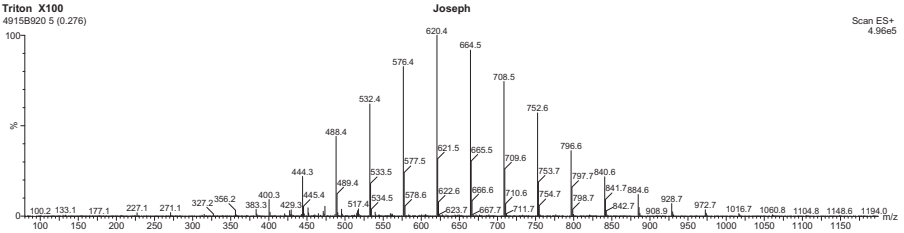


**Figure 23** Electrospray mass spectrum of octoxynol (Triton X).

**TABLE 3 Calculation of average molecular mass and polydispersity for Triton X-100.**

| Length | Mass | Empirical formula | Abundance | Number contribution |
|---|---|---|---|---|
| 0 | 224.2 | $C_{14}H_{26}NO_1$ | | 0 |
| 1 | 268.2 | $C_{16}H_{30}NO_2$ | | 0 |
| 2 | 312.3 | $C_{18}H_{34}NO_3$ | | 0 |
| 3 | 356.3 | $C_{20}H_{38}NO_4$ | 8 | 24 |
| 4 | 400.3 | $C_{22}H_{42}NO_5$ | 18 | 72 |
| 5 | 444.3 | $C_{24}H_{46}NO_6$ | 39 | 195 |
| 6 | 488.4 | $C_{26}H_{50}NO_7$ | 78 | 468 |
| 7 | 532.4 | $C_{28}H_{54}NO_8$ | 109 | 763 |
| 8 | 576.4 | $C_{30}H_{58}NO_9$ | 145 | 1,160 |
| 9 | 620.4 | $C_{32}H_{62}NO_{10}$ | 175 | 1,575 |
| 10 | 664.5 | $C_{34}H_{66}NO_{11}$ | 161 | 1,610 |
| 11 | 708.5 | $C_{36}H_{70}NO_{12}$ | 131 | 1,441 |
| 12 | 752.5 | $C_{38}H_{74}NO_{13}$ | 102 | 1,224 |
| 13 | 796.5 | $C_{40}H_{78}NO_{14}$ | 64 | 832 |
| 14 | 840.6 | $C_{42}H_{82}NO_{15}$ | 38 | 532 |
| 15 | 884.6 | $C_{44}H_{86}NO_{16}$ | 22 | 330 |
| 16 | 928.6 | $C_{46}H_{90}NO_{17}$ | 13 | 208 |
| 17 | 972.6 | $C_{48}H_{94}NO_{18}$ | 7 | 119 |
| 18 | 1,016.7 | $C_{50}H_{98}NO_{19}$ | 3 | 54 |
| 19 | 1,060.7 | $C_{52}H_{102}NO_{20}$ | 1 | 19 |
| 20 | 1,104.7 | $C_{54}H_{106}NO_{21}$ | | 0 |
| | | | $\Sigma = 1{,}114$ | $\Sigma = 10{,}626$ |
| | | | Average length | 9.5 |
| | | | Average mol. mass | 643.7 |

Such average molecular mass determinations are analogous to two figures of merit used in characterizing polymer samples, namely the number-average relative molecular mass, $\overline{M}_n$, and the weight-average relative molecular mass, $\overline{M}_w$.[20] $\overline{M}_n$, indicates the average chain length in a polymer sample, and is classically measured by determination of colligative properties (Yau, Kirkland and Bly, 1979), such as freezing point depression or vapor-phase osmometry. $\overline{M}_w$ correlates with physical properties of larger

[20] Note that we are using the correct terminology here, rather than the more customary and historically used number-average and weight-average molecular weights.

polymers. Literature on determination of $\overline{M}_n$ and $\overline{M}_w$ show that FAB, electrospray, and MALDI as very appropriate techniques for direct determination. Limitations presented by MALDI are discussed by Byrd and McEwen (2000).

The calculation in Table 3, derived from the spectrum in Fig. 23, is directly related to $\overline{M}_n$ and is a much more facile and direct measure of $\overline{M}_n$. $\overline{M}_w$ can also be calculated directly from the data in Table 3, and is 665.6 daltons for this sample. Polydispersity, as defined in polymer chemistry, is the ratio of $\overline{M}_w/\overline{M}_n$, calculated here to be 1.03. Response factors of individual species of a polymeric molecule in the mass spectrometer can vary with the chain length of the molecule (R.E. Borjas, 2007, personal communications). For some types of polymer, lower mass members of the series will, for example, have a higher molar ion yield than will higher mass members of the series. The distribution observed in the mass spectrum will therefore be biased in favor of shorter chain length. Not all polymers, however, exhibit this phenomenon. To a first approximation, however, and for comparative purposes, the measurements are useful and much easier to do than, for example, the classical freezing point depression.

Why examine polymers and polymer patterns in the mass spectrometer here? A number of polymeric materials are used regularly as excipients in pharmaceutical formulations—PEG, various surfactants such as Triton X-100, and others. The opportunity then presents for a drug substance to interact chemically with the polymeric material, which will then appear in analyses of the formulation as an impurity or a degradant. Polymeric materials, which are not UV-active, and not detectable by a typical UV-visible detector in an LC-MS run, will become visible when a UV-active drug substance attaches to a polymer molecule and labels it with a chromophore. Such polymeric materials will often elute in an HPLC chromatogram as broad chromatographic peaks detected by a UV-visible detector, the mass spectrometer TIC, or both. The mass spectrum under the chromatographic peak will show a polymer pattern of a polymeric excipient, perchance offset in mass by attachment of the drug substance molecule or a portion of it. Carboxylic acid drug substances have been observed to esterify with PEG. An example of interaction with cetirizine (the acid portion of Structure 6) with PEG to form a family of esters in a formulation is illustrated in the FAB mass spectrum of these compounds in Fig. 24.

## 9.11 Multiply Charged Ions

ESI-generated multiply charged ions are most frequently discussed in the examination of peptides and intact proteins. The general observation is that peptides need to be on

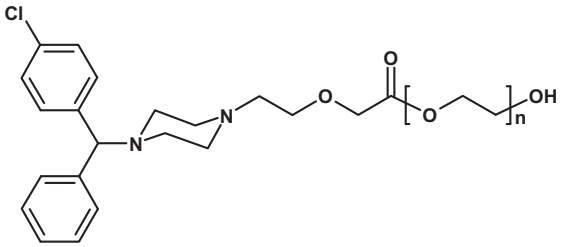


**Structure 6** Cetirizine PEG esters. $C_{21}H_{25}N_2O_3Cl[C_2H_4O]_n$ (monoisotopic); r.m.m. 388 + $[44]_n$.

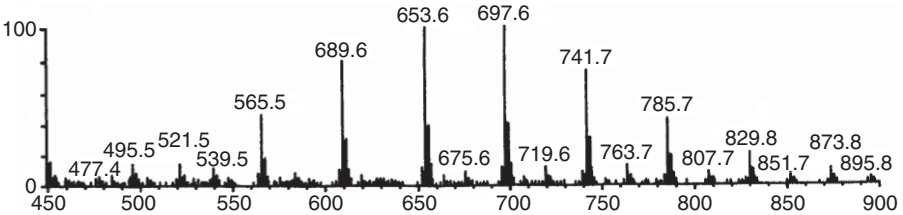


**Figure 24** Fast atom bombardment spectrum of cetirizine poly(ethylene glycol) ester mixture.

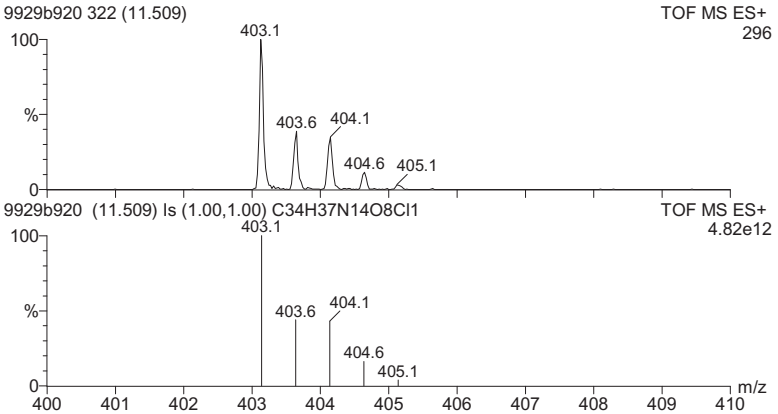


**Figure 25** Comparison of the isotope pattern of the doubly charged molecular ion region (top panel) and the isotope pattern modeled (bottom panel) from the $C_{34}H_{37}N_{14}O_8Cl$ elemental composition of the m/z 403 doubly charged molecular ion.

the order of 1500–2000 daltons or larger in order to form abundant multiply charged ions (i.e. doubly charged ions). Although doubly charged molecular ions have been reported in the EI mass spectra of polycyclic aromatic hydrocarbons (Myerson and Vander Haar, 1962; Sharp *et al.*, 1988), mention of multiply charged ions in the context of classic mass spectra of small molecules is rare. In our work on impurities and degradants of small pharmaceutical molecules, however, we are starting to see an increasing frequency of doubly charged molecular ions in electrospray mass spectra—molecules with relative molecular masses as low as 500 to 600 daltons. The factors that determine whether or not and to what extent a molecule will form doubly charged molecular ions are still under investigation (Gaskell, 1997). How does one tell if one has a doubly charged species? The spacing between the isotope peaks indicates multiple charging.

ESI mass spectra of compounds often contain van der Waals clusters of the molecule. Numerologically, these species appear at two times the relative molecular mass of the molecule plus one. Consider the example spectrum shown in Fig. 25. The m/z 403 species could be assigned to be the $[M+H]^+$ of the compound in question, indicating a relative molecular mass of 402 daltons. It was indeed the most abundant peak in the

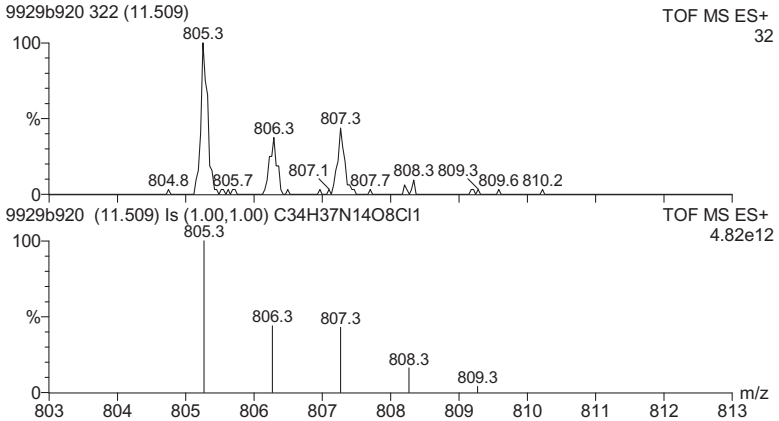


**Figure 26** Comparison of the isotope pattern of the singly charged molecular ion region (top panel) and the isotope pattern modeled (bottom panel) from the $C_{34}H_{37}N_{14}O_8Cl$ elemental composition of the m/z 805 singly charged molecular ion.

spectrum. A proton-bound dimer would correspond to a $[2M+H]^+$ peak at m/z 805, consistent with the expanded spectral portion shown in Fig. 26. The relative abundance of the m/z 805 species is approximately 10% that of the m/z 403 species, consistent with typical relative abundance observations of $[M+H]^+$ and $[2M+H]^+$. The convention would be to assign the $[M+H]^+$ to be the largest peak in the spectrum. The fact that a 402-dalton molecular mass exhibited an illogical mass difference between this and the drug substance molecule under investigation prompted re-examination of the m/z 403 region and the putative assignment of this species as the $[M+H]^+$.

Close examination reveals that the peaks comprising the isotope pattern of the m/z 403 species in Fig. 25 appear at 0.5-dalton intervals. The nature of the instrument on which this spectrum was obtained, and the confidence in mass assignments, indicate that these mass assignments were indeed correct, and that this spectrum represents a doubly charged ion. Recall that the mass spectrometer does not measure mass but mass-to-charge ratio. Typically, for small molecules, ions carry only a single electrostatic charge, and the m/z value reduces to the mass. The interval between isotope peaks in a pattern such as this should be 1 dalton. The 0.5-dalton spacing indicates that the m/z 403 species is really a $[M+2H]^{2+}$. The m/z 805 species is the true $[M+H]^+$, and the molecular mass is really 804 daltons. This reassigned molecular mass and a consideration of the synthetic chemistry of the drug substance being studied were sufficient to propose a reasonable identity of the impurity.

## 9.12 Calculating the Charge State and Chemical Molecular Mass of a Protein

With the increasing application of ESI to the examination of intact proteins (illustrated by the example of Fig. 27), one observes a pattern outwardly resembling a polymer pattern. This feature is not in fact a polymer pattern but a multiply charged ion envelope of a single chemical species. The spectra in Figs. 22 to 24 represent the $[M+H]^+$ of collections of molecules of different sizes—the spectrum in Fig. 27 that of

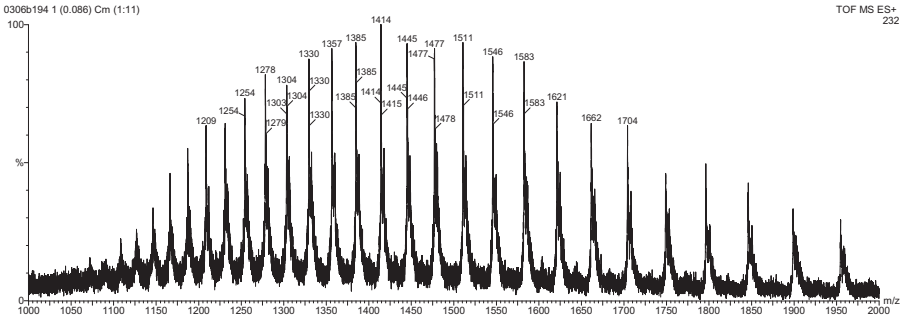


**Figure 27** Positive ion ESI mass spectrum of bovine serum albumin. The average molecular mass observed here for this protein is 66,424 daltons. There are two additional higher mass proteins in this sample.

a single species. The protein used here is bovine serum albumin, a single molecule with a chemical molecular mass calculated from this spectrum of 66,424 daltons. The definitive way to differentiate a polymer pattern from a multiply charged ion series of a protein is by evaluating the spacing of the peaks. A polymer pattern will show peaks spaced at a constant mass interval across the entire series. A multiply charged ion series will show a steadily increasing mass spacing as one traverses the pattern from low mass to high mass.

It is useful to point out some of the differences in interpretation and thought process needed to interpret mass spectra of proteins. The shift from monoisotopic molecular masses in small molecules to chemical molecular masses in peptides and proteins is necessary when the chemical entities exceed 1500 to 2000 daltons. Mass calibration procedures and philosophies of instrument operation change. The concept and application of chemical molecular mass accuracy changes from 0.1 to 0.2 dalton accuracy for nominal monoisotopic measurements to 0.01–0.02% for proteins (acceptable 0.01% accuracy for the bovine serum albumin example used here is ±7 daltons). In short, the mass spectrometry of proteins and other intact biopolymers deserves an entire monograph of its own. We will include some discussions here, however, to introduce the area.

The first useful fact to extract from the spectrum of a protein is the relative molecular mass of the protein. Mass spectrometer data systems from the major manufacturers incorporate programs and algorithms for deconvoluting these complex mass spectra into spectra from which the relative molecular masses can be directly read. Some programs require significant input and intervention in order to give useful results. Others are very automatic. How does one spot-check the results of such programs, rather than relying on blind faith? The procedure for hand calculating chemical molecular masses is discussed elsewhere (de Hoffmann and Stroobant, 2002, pp. 37–39), and will be briefly derived here. Example calculations will illustrate the procedure.

A peak in the mass spectrum will have an assigned "mass" value, m, determined by the mass spectrometer data system. It will have an unknown charge state, q, the number of electrostatic charges carried by it. If we represent the chemical molecular mass of a protein by M, we can relate M, m, and q by Equation 1:

$$M = mq \tag{1}$$

Each peak in the multiply charged envelope will have this same relationship, differing by the numerical values of the assigned mass and the charge state. For two adjacent peaks in the multiply charged envelope, their electrostatic charges, q, will differ by one charge. Expressing relationships for the two peaks as Equations 2 and 3, with the charge relationship as Equation 4:

$$M = m_1 q_1 \tag{2}$$

$$M = m_2 q_2 \tag{3}$$

$$q_2 = q_1 + 1 \tag{4}$$

Setting the expressions for M in terms of $m_1$ and $m_2$ equal, substituting for $q_2$, and rearranging gives Equation 5:

$$m_1/m_2 - 1 = 1/q_1 \tag{5}$$

Note that Equations 4 and 5 are written assuming one is using adjacent peaks in the multiply charged envelope that differ by only one electrostatic charge unit. One could easily use any pair of peaks in the envelope, adjusting the equations for the difference in charge state. For example, for peaks separated by three charge states, Equation 5 changes to $3/q_1$. Using adjacent peaks is easiest. Substituting the appropriate numerical values into Equation 6, one can calculate $q_1$ for the corresponding $m_1$. Once one charge state is known, the rest are assumed, each differing by 1. With the charge states in hand, M of the protein can be calculated from any of the sets of numbers, according to Equation 6. Subtracting q corrects for the charge carriers, namely the $H^+$, that have been attached to the protein. As a check of the result, M should be calculated for different peaks in the spectrum to verify that the charge states are assigned correctly:

$$M = (mq) - q \tag{6}$$

Figure 28 is the positive ion electrospray mass spectrum of insulin. All three prominent species in the spectrum are molecular ions. The pattern is a multiply charged envelope because the differences between the m/z values decrease from high mass to

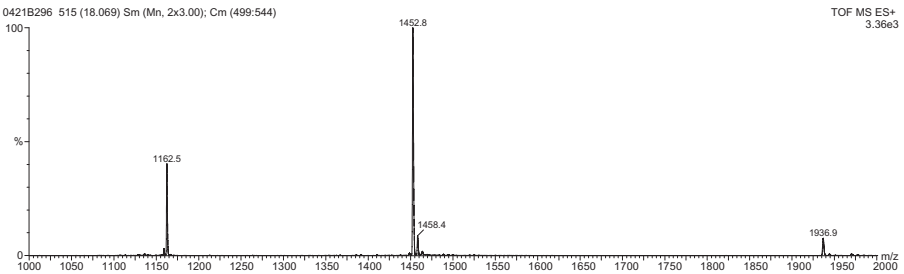


**Figure 28** Electrospray mass spectrum of insulin.

```
                          ┌-S--S┐
A chain   H2N - GIVEQ CCTSI CSLYQ LENYC N - CO2H
                          |             |
                          S             S
                          S            S
                          |             |
B chain   H2N - FVNQH LCGSH LVEAL YLVCG ERGFF YTPKT - CO2H
```

**Structure 7** Amino acid sequence of insulin.

low mass. This feature differentiates this pattern from a polymer pattern (discussed earlier).

The amino acid sequences of the two polypeptide chains of insulin are shown in Structure 7 (using the one letter coding for the amino acids). The disulfide bridges interconnecting the two chains are also shown. The A chain contains 21 amino acids (average molecular mass 2381.7 daltons), with an intramolecular disulfide bridge between cysteine residues $A_6$ and $A_{11}$. The B chain contains 30 amino acid residues (average molecular mass 3430.0 daltons). Two disulfide bridges connect the A and B chains, connecting between cysteine residues $A_7$ and $B_7$ and between $A_{20}$ and $B_{19}$. This composite structure gives a chemical molecular mass of 5807.7 daltons.

We will apply this calculation to the data presented in Fig. 28 for insulin. We first arbitrarily choose the m/z 1452.8 peak as $m_1$, with its corresponding unknown charge state $q_1$. Since the electrostatic charge increases with decreasing m/z, the next higher charge state species, with $q_2 = q_1 + 1$, corresponds to the m/z 1162.5 peak. Substituting and calculating:

$$1/q_1 = (1452.8/1162.5) - 1 \tag{7}$$

$$1/q_1 = 0.2497 \tag{8}$$

$$q_1 = 4.004 \tag{9}$$

Rounding to the nearest integer, $q_1$ is calculated to be 4, and the m/z 1452.8 species therefore corresponds to the +4 charge state. The m/z 1162.5 species corresponds to the +5 charge state. The m/z 1936.9 species corresponds to the +3 charge state. The chemical molecular mass, M, is calculated using the numbers for the +4 charge state as follows:

$$M = (1452.8)(4) - 4 \tag{10}$$

$$M = 5807.2 \text{ daltons} \tag{11}$$

Calculating M from the +5 and +3 charge states gives, respectively, 5807.5 and 5807.7 daltons. One could conduct a statistical evaluation of the results of several combinations of numbers to derive a precision for the molecular mass determination. The data system programs that deconvolute these protein spectra permit such an evaluation.

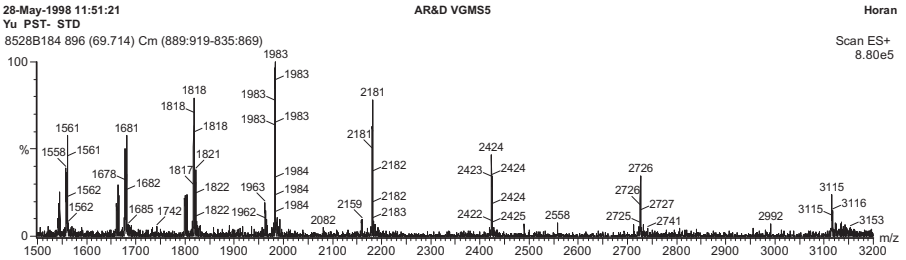


**Figure 29** Multiply charged ESP mass spectrum of pST.

**AFPAMPLSSL FANAVLRAQH LHQLAADTYK EFERAYIPEG QRYSIQNAQA**

**AFCFSETIPA PTGKDEAQQR SDVELLRFSL LLIQSWLGPV QFLSRVFTNS**

**LVFGTSDRVY EKLKDLEEGI QALMRELEDG SPRAGQILKQ TYDKFDTNLR**

**SDDALLKNYG LLSCFKKDLH KAETYLRVMK CRRFVESSCA F**

**Structure 8** Amino acid sequence of porcine somatotropin (pST).

Note the species in Fig. 28 at m/z 1458.4. It differs from the m/z 1452.8 species by 5.6 m/z units. Noting that the m/z 1452.8 species is the +4 charge state of insulin, and assuming that this species is of the same charge state, the difference is $5.6 \times 4 = 22.4$ daltons. This mass increment corresponds to a $Na^+$ cationated molecular ion of insulin, in which one of the $H^+$ charge carriers is replaced by $Na^+$. The molecular mass of insulin from this species calculates to 5807.6 daltons. Note here that resolving power of the mass analyzer becomes important in making mass assignments and in being able to differentiate between species with small mass differences. These species are 22 daltons different in mass, but because they carry 4 electrostatic charges, they are only separated by 5.6 m/z units.

Revisiting the spectrum of bovine serum albumin in Fig. 27, calculations indicate that the m/z 1477 peak corresponds to a +45 charge state, and the chemical molecular mass of bovine serum albumin, from this spectrum, is approximately 66,420 daltons. The 66,424-dalton value cited earlier was generated by a data system deconvolution program, taking into account the entire spectrum and the decimal places of the assigned m/z values not shown in the figure.

Porcine somatotropin (pST) is another protein we will examine. A spectrum of the protein is shown in Fig. 29. Its amino acid sequence is given in Structure 8. An average molecular mass of this protein is calculated, using the m/z 2181 species (charge state +10) in Fig. 29 to be 21,800 daltons. From the sequence, the chemical molecular mass calculates to 21,802 daltons (fully reduced). Note that the state of oxidation of the two known disulfide links in the native protein structure would cause the molecular mass to vary by 4 daltons. The best figure of merit in the literature for accuracy in measuring the average molecular mass of a protein is ±0.01%. This precision corresponds to ±2 daltons for this protein. To achieve this, one must take significant care in making the measurement. A more typical routine figure of merit is ±0.02% (±4 daltons).

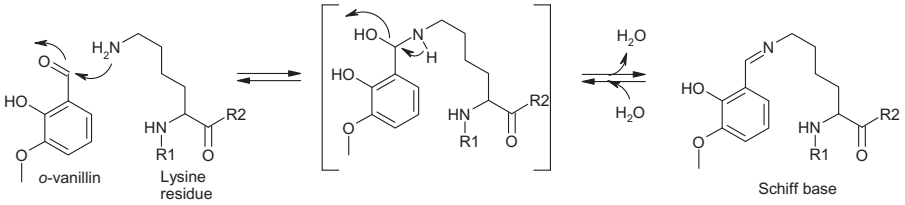


**Scheme 1** Reversible Schiff base formation between *o*-vanillin and a lysine residue.

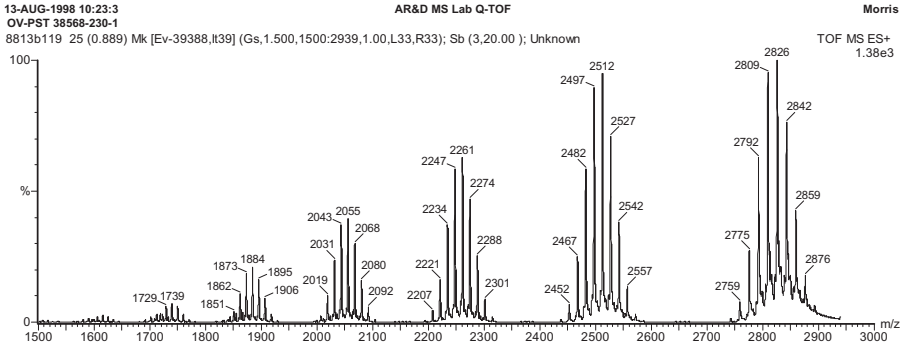


**Figure 30** Multiply charged ESP mass spectrum of ortho-vanillylated porcine somatotropin.

The question arises about whether mass spectrometry can evaluate the state of reduction/oxidation of this protein by the measured molecular mass determination alone. The difference between fully oxidized pST and fully reduced pST is within the range of imprecision of the experimental measurements. The state of oxidation cannot be reliably measured by determining the average molecular mass.

Experimentation aimed at providing stabilization of this protein involved derivatization of the ε-amino groups of lysine residues in the protein by Schiff base formation with aromatic aldehydes, specifically here *o*-vanillin. Reaction according to Scheme 1 introduces an additional 134 daltons for each site on the protein that is derivatized. Assessment of average degree of substitution of pST can be done by direct mass spectrometric observation (Sharp *et al.*, 2006).

The multiply charged mass spectrum in Fig. 30 illustrates a polymer pattern superimposed on top of a multiply charged envelope. Using the functional test of spacing between the peaks, we see that, for each of the distributions, mass spacing between peaks is constant, indicating a polymer pattern. The spacings correspond to 134 daltons adjusted by the charge state of the particular pattern, the mass of an *o*-vanillyl reside coupled as a Schiff base to an ε-amino group. The centers of the distributions, however, decrease in mass interval with decreasing m/z, indicating a multiply charged envelope. Any of these distributions can be used to calculate an average degree of substitution for this preparation, according to similar procedures illustrated for calculating average molecular masses of polymers described earlier.

Most of the mass spectrometer computer data systems—especially those targeted toward making measurements of biological molecules—contain one or more

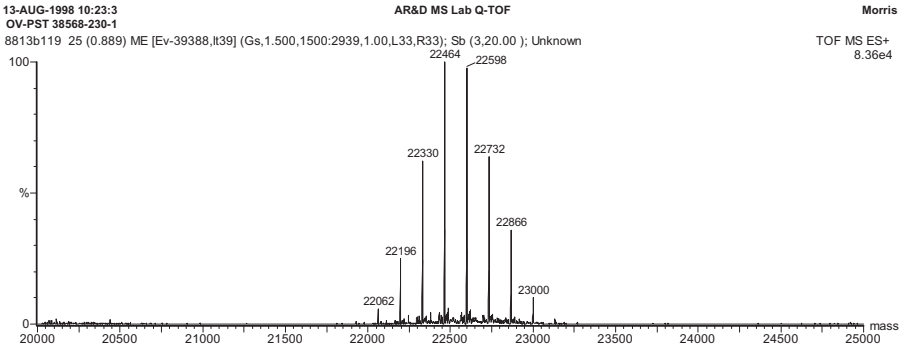


**Figure 31** MaxEnt transformed singly charged mass spectrum of OV-pST. This spectrum is derived from the one shown in Fig. 30.

components for performing the transformation calculation, illustrated before for insulin, to obtain the chemical molecular masses of the macromolecules. If deconvolution of the spectrum in Fig. 30 is performed, the spectrum in Fig. 31 is obtained. This spectrum represents the distribution of derivatized protein molecules in the sample as if they were singly charged species. The annotations at the tops of the peaks are the chemical molecular masses of the species. The peaks are separated by 134 daltons, the mass of the *o*-vanillyl residue. Calculating the average degree of substitution of this protein sample from the transformed spectrum should give the same value as calculation from the untransformed spectrum of Fig. 30. Results are slightly different—from 5.6 for the transformed spectrum to 5.8 for the +8 charge state envelope of Fig. 30. Depending upon the needs of the analysis, these slight differences could be acceptable. The quantitative nature of transformation of multiply charged spectra to singly charged spectra have been discussed (Reinhold and Reinhold, 1992).

## REFERENCES

Alsante, K.M., Boutros, P., Couturier, M.A. *et al.* (2004). Pharmaceutical impurity identification: a case study using a multidisciplinary approach. *Journal of Pharmaceutical Sciences*, *93*, 2296–2309.

Amad, M.H., Cech, N.B., Jackson, G.S., Enke, C.G. (2000). Importance of gas-phase proton affinities in determining the electrospray ionization response for analytes and solvents. *Journal of Mass Spectrometry*, *35*, 784–789.

Arpino, P.J. (1982). On-line liquid chromatography/mass spectrometry? An odd couple. *Trends in Analytical Chemistry*, *1*, 154.

Bai, Y., Milne, J.S., Mayne, L., Englander, S.W. (1993). Primary structure effects on peptide group hydrogen exchange. *Proteins: Structure Function and Genetics*, *17*, 75–86.

Baldwin, M.A., Burlingame, A.L., Nikolaev, E. (2004). Victor L. Talroze: 1922–2004. *Journal of the American Society for Mass Spectrometry*, *15*, 1517–1519.

Baldwin, M.A., McLafferty, F.W. (1973). Liquid chromatography-mass spectrometry interface. I: the direct introduction of liquid solutions into a chemical ionization mass spectrometer. *Organic Mass Spectrometry*, *7*, 1111–1112.

Barber, M., Bordoli, R.S., Sedgwick, R.D., Tyler, A.N. (1981). Fast atom bombardment of solids (F.A.B): a new ion source for mass spectrometry. *Chemical Communications*, 325–326.

Beavis, R.C., Chait, B.T. (1989). Factors affecting the ultraviolet laser desorption of proteins. *Rapid Communications in Mass Spectrometry*, *3*, 233–237.

Beckey, H.D. (1969). Field desorption mass spectrometry: a technique for the study of thermally unstable substances of low volatility. *International Journal of Mass Spectrometry and Ion Physics*, *2*, 500–503.

Bellar, T.A., Behymer, T.D., Budde, W.L. (1990). Investigation of enhanced ion abundances from a carrier process in high-performance liquid chromatography particle beam mass spectrometry. *Journal of the American Society for Mass Spectrometry*, *1*, 92–98.

Beynon, J.H. (1954). Qualitative analysis of organic compounds by mass spectrometry. *Nature*, *174*, 735–737.

Beynon, J.H. (1960). *Mass Spectrometry and its Applications to Organic Chemistry*, Elsevier Publishers, Amsterdam.

Biemann, K. (1990). Utility of exact mass measurements. *Methods in Enzymology*, *193*, 295–305.

Biemann, K., Oro, J., Toulmin, P.III, , Orgel, L.E., Nier, A.O., Anderson, D.M., Simmonds, P.G., Flory, D., Diaz, A.V., Rushneck, D.R., Biller, J.A. (1976). Search for organic and volatile inorganic compounds in two surface samples from the Chryse Planitia region of Mars. *Science*, *194*, 72–76.

Blades, A.T., Ikonomou, M.G., Kebarle, P. (1991). Mechanism of electrospray mass spectrometry. Electrospray as an electrochemical cell. *Analytical Chemistry*, *63*, 2109–2114.

Blakley, C.R., Carmody, J.C., Vestal, M.L. (1980). Liquid chromatograph-mass spectrometer for analysis of nonvolatile samples. *Analytical Chemistry*, *52*, 1636–1641.

Budzikiewicz, H., Djerassi, C., Williams, D.H. (1967). *Mass Spectrometry of Organic Compounds*, Holden-Day, Inc., San Francisco, CA.

Burinsky, D.J. (2006). Mass spectrometry, in *Comprehensive Analytical Chemistry. Volume 47: Modern Instrumental Analysis* (eds S. Ahuja, N. Jespersen), Elsevier, Amsterdam, pp. 319–396.

Busch, K.L. (1994). Using exact mass measurements. *Spectroscopy*, *9* (7), 21–22.

Busch, K.L. (1995). Mass spectrometric detectors for samples separated by planar electrophoresis. *Journal of Chromatography A*, *692*, 275–290.

Busch, K.L. (2004a). Planar separations and mass spectrometric detection. *Journal of Planar Chromatography*, *17*, 398–403.

Busch, K.L. (2004b). Masses in mass spectrometry: balancing the analytical scales. *Spectroscopy*, *19* (11), 32–34.

Byrd, H.C.M., McEwen, C.N. (2000). The limitations of MALDI-TOF mass spectrometry in the analysis of wide polydisperse polymers. *Analytical Chemistry*, *72*, 4568–4576.

Campana, J.E. (1980). Elementary theory of the quadrupole mass filter. *International Journal of Mass Spectrometry and Ion Physics*, *33*, 101–117.

Cappelletty, D., Eiselstein-McKitrick, K. (2007). The echinocandins. *Pharmacotherapy*, *27*, 369–388.

Caprioli, R.M., Fan, T., Cottrell, J.S. (1986). Continuous-flow sample prove for fast atom bombardment mass spectrometry. *Analytical Chemistry*, *58*, 2949–2953.

Carroll, D.I., Dzidic, I., Stillwell, R.N., Haegele, K.D., Horning, E.C. (1975). Atmospheric pressure ionization mass spectrometry: corona discharge ion source for use in liquid chromatograph-mass spectrometer-computer analytical system. *Analytical Chemistry*, *47*, 2369–2373.

Catlin, D.H., Sekera, M.H., Ahrens, B.D., Starcevic, B., Chang, Y.-C., Hatton, C.K. (2004). Tetrahydrogestrinone: discovery, synthesis, and detection in urine. *Rapid Communications in Mass Spectrometry*, *18*, 1245–1249.

Chapman, J.R., Harden, E.H., Evans, S., Moore, L.E. (1983). LC-MS interfacing in sector mass spectrometers. *International Journal of Mass Spectrometry and Ion Physics*, *46*, 201–204.

Chen, G., Khusid, A., Daaro, I., Irish, P., Pramanik, B.N. (2007). Structural identification of trace level enol tautomer impurity by on-line hydrogen/deuterium exchange HR-LC/MS in a LTQ-Orbitrap hybrid mass spectrometer. *Journal of Mass Spectrometry*, *42*, 967–970.

Cody, R.B., Laramée, J.A., Durst, H.D. (2005). Versatile new ion source for the analysis of materials in open air under ambient conditions. *Analytical Chemistry*, *77*, 2297–2302.

Cole, R.B. (2000). Some tenets pertaining to electrospray ionization mass spectrometry. *Journal of Mass Spectrometry*, *35*, 763–772.

Cooks, R.G., Rockwood, A.L. (1991). The "Thomson." A suggested unit for mass spectroscopists. *Rapid Communications in Mass Spectrometry*, *5*, 93.

Costello, C.E. (1991). Appendix 3, Mass spectra of matrix materials. *Methods in Enzymology*, *193*, 875–882.

Crow, F.W., Tomer, K.B., Gross, M.L. (1983). Mass resolution in mass spectrometry-mass spectrometry. *Mass Spectrometry Reviews*, *2*, 47–76.

DeBievre, P., Barnes, I.L. (1985). Table of the isotopic composition of the elements as determined by mass spectrometry. *International Journal of Mass Spectrometry and Ion Processes*, *65*, 211–230.

de Hoffmann, E., Stroobant, V. (2002). *Mass Spectrometry: Principles and Applications*, 2nd edn, John Wiley & Sons, Ltd, Chichester, UK.

Dole, M., Mack, L.L., Hines, R.L., Mobley, R.C., Ferguson, L.D., Alice, M.B. (1968). Molecular beams of macroions. *Journal of Chemical Physics*, *49*, 2240–2249.

Downard, K.M. (2007a). Francis William Aston: the man behind the mass spectrograph. *European Journal of Mass Spectrometry*, *13*, 177–190.

Downard, K.M. (2007b). Cavendish's crocodile and dark horse: the lives of Rutherford and Aston in parallel. *Mass Spectrometry Reviews*, *26*, 713–723.

Downard, K.M., de Laeter, J.R. (2005). A history of mass spectrometry in Australia. *Journal of Mass Spectrometry*, *40*, 1123–1139.

Englander, S.W. (2006). Hydrogen exchange and mass spectrometry: a historical perspective. *Journal of the American Society for Mass Spectrometry*, *17*, 1481–1489.

Fenn, J.B. (2003). Electrospray wings for molecular elephants (Nobel Lecture). *Angewandte Chemie (International)*, *42*, 3871–3894.

Fernandes de la Mora, J. Van Berkel, G.J., Enke, C.G., Cole, R.B., Martinez-Sanchez, M., Fenn, J.B. (2000). Electrochemical processes in electrospray ionization mass spectrometry. *Journal of Mass Spectrometry*, *35*, 939–952..

Field, F.H. (1968). Chemical ionization mass spectrometry. *Accounts of Chemical Research*, *1*, 42–49.

Field, F.H. (1990). The early days of chemical ionization: a reminiscence. *Journal of the American Society for Mass Spectrometry*, *1*, 277–283.

Futrell, J.H., Miller, C.D. (1966). Tandem mass spectrometer for study of ion-molecule reactions. *Review of Scientific Instrumentation*, *37*, 1521–1526.

Gamero-Castaño, M., Fernandez de la Mora, J. (2000). Kinetics of small ion evaporation from the charge and mass distribution of multiply charged clusters in electrospray. *Journal of Mass Spectrometry*, *35*, 790–803.

Gaskell, S.J. (1997). Electrospray: principles and practice. *Journal of Mass Spectrometry*, *32*, 677–688.

Gilliam, J.M., Landis, P.W., Occolowitz, J.L. (1983). Accurate mass measurement in fast atom bombardment mass spectrometry. *Analytical Chemistry*, *55*, 1531–1533.

Gilliam, J.M., Landis, P.W., Occolowitz, J.L. (1984). On-line accurate mass measurement in fast atom bombardment mass spectrometry. *Analytical Chemistry*, *56*, 2285–2288.

Gohlke, R.S., McLafferty, F.W. (1993). Early gas chromatography/mass spectrometry. *Journal of the American Society for Mass Spectrometry*, *4*, 367–371.

Grayson, M.A. (2002). *Measuring Mass: From Positive Rays to Proteins*, Chemical Heritage Press, Philadelphia, p. 149.

Griffiths, I.W. (1997). J.J. Thomson—the centenary of his discovery of the electron and of his invention of mass spectrometry. *Rapid Communications in Mass Spectrometry*, *11*, 2–16.

Gross, J.H. (2004). *Mass Spectrometry: A Textbook*, Springer-Verlag, Berlin, p. 518.

Gross, M.L. (1994). Accurate masses for structure confirmation. *Journal of the American Society for Mass Spectrometry*, *5*, 57.

Hamming, M.C., Foster, N.G. (1972). *Interpretation of Mass Spectra of Organic Compounds*, Academic Press, New York.

Harrison, A.G. (1992). *Chemical Ionization Mass Spectrometry*, 2nd edn, CRC Press, Boca Raton, FL, p. 224.

Heathcock, C.H. (1990). Editorial. *Journal of Organic Chemistry*, *55*, 8A.

Hertz, H.S., Hites, R.A., Biemann, K. (1971). Identification of mass spectra by computer-searching a file of known spectra. *Analytical Chemistry*, *43*, 681–691.

Hillenkamp, F., Karas, M. (1990). *Methods in Enzymology*, *193*, 280–294.

Holland, J.F., Soltmann, B., Sweeley, C.C. (1976). A model for ionization mechanisms in field desorption mass spectrometry. *Biomedical Mass Spectrometry*, *3*, 340–345.

Honig, R.E. (1985). The development of secondary ion mass spectrometry (SIMS): a retrospective. *International Journal of Mass Spectrometry and Ion Processes*, *66*, 31–54.

Iribarne, J.V., Thomson, B.A. (1976). On the evaporation of small ions from charged droplets. *Journal of Chemical Physics*, *64*, 2287–2294.

Ito, Y., Takeuchi, T., Ishii, D., Goto, M., Mizuno, T. (1986). Direct coupling of micro high performance liquid chromatography with fast atom bombardment mass spectrometry. II: application to gradient elution of bile acids. *Journal of Chromatography*, *385*, 201–209.

Karas, M., Bachmann, D., Bahr, U., Hillenkamp, F. (1987). Matrix-assisted ultraviolet laser desorption of non-volatile compounds. *International Journal of Mass Spectrometry and Ion Processes*, *78*, 53–68.

Katritzky, A.R., Karelson, M., Lobanov, V.S. (1997). SPR as a means of predicting and understanding chemical and physical properties in terms of structure. *Pure and Applied Chemistry*, *69*, 245–248.

Kebarle, P. (2000). A brief overview of the present status of the mechanisms involved in electrospray mass spectrometry. *Journal of Mass Spectrometry*, *35*, 804–817.

Kind, T., Fiehn, O. (2006). Metabolomic database annotations via query of elemental compositions: mass accuracy is insufficient even at less than 1 ppm. *BMC Bioinformatics*, *7*, 234–246.

Kiser, R.W. (1965). *Introduction to Mass Spectrometry and Its Applications*, Prentice-Hall, Inc., Englewood Cliffs, NJ.

Klose, R.T. (1987). The Joys of Science. *Newsweek*, 26 October 1987.

Knapp, D.R. (1979). *Handbook of Analytical Derivatization Reactions*, John Wiley & Sons, Inc., New York, p. 741.

Kondrat, R. (1999). High resolution mass spectrometry. *Journal of the American Society for Mass Spectrometry*, *10*, 661.

Kowalski, M.H., Sharp, T.R., Stang, P.J. (1987). *Organic Mass Spectrometry*, *22*, 642–643.

Krishna, M.M.G., Hoang, L., Lin, Y., Englander, S.W. (2004). Hydrogen exchange methods to study protein folding. *Methods*, *34*, 51–64.

Lehmann, W.D., Schulten, H.-R. (1978). Quantitative field desorption mass spectrometry: V-discussion of methodology and examples of applications. *Biomedical Mass Spectrometry*, *5*, 208–214.

McElroy, M.B., Yung, Y.L., Nier, A.O. (1976). Isotopic composition of nitrogen: implications for the past history of Mars' atmosphere. *Science*, *194*, 70–72.

McFadden, W. (1973). *Techniques of Combined Gas Chromatography-Mass Spectrometry: Applications in Organic Analysis*, John Wiley & Sons, Inc., New York.

McFadden, W.H., Schwartz, H.L., Bradford, D.C. (1976). Direct analysis of liquid chromatographic effluents. *Journal of Chromatography*, *122*, 389–396.

Macfarlane, R.D. (1983). Californium-252 plasma desorption mass spectrometry. large molecules, software, and the essence of time. *Analytical Chemistry*, *55*, 1247A–1264A.

Macfarlane, R.D., Hu, Z.H., Song, S., Pittenauer, E., Schmid, E.R., Allmaier, G., Metzger, J.O., Tusznski, W. (1994). $^{252}$Cf-plasma desorption mass spectrometry II—a perspective of new directions. *Biology Mass Spectrometry*, *23*, 117–130.

McLafferty, F.W. (1981). Tandem mass spectrometry. *Science*, *214*, 280–287.

McLafferty, F.W. (1983). *Tandem Mass Spectrometry*, John Wiley & Sons, Inc., New York.

McLafferty, F.W., Venkataraghavan, R. (1982). *Mass Spectral Correlations*, 2nd edn, American Chemical Society Advances in Chemistry Series #40, Washington, DC, p. 124.

Mamer, O.A., Lesimple, A. (2004). Letter to the editor. *Journal of the American Society for Mass Spectrometry*, *15*, 626.

Marshall, A.G., and Hendrickson, C.L. (2008). High resolution mass spectrometers. *Annual Review of Analytical Chemistry*, *1*, 579–599.

Marshall, A.G., Rodgers, R.P. (2004). Petroleomics: the next grand challenge for chemical analysis. *Accounts of Chemical Research*, *37*, 53–59.

Meili, J., Seibl, J. (1984). A new versatile matrix for fast atom bombardment analysis. *Organic Mass Spectrometry*, *19*, 581–582.

Message, G.M. (1984). *Practical Aspects of Gas Chromatography/Mass Spectrometry*, John Wiley & Sons, Inc., New York.

Meyerson, S. (1986). Reminiscences of the early days of mass spectrometry in the petroleum industry. *Organic Mass Spectrometry*, *21*, 197–208.

Miles, C.J., Doerge, D.R., Bajic, S. (1992). Particle beam liquid chromatography-mass spectrometry of national pesticide survey analytes. *Archives of Environmental Contamination and Toxicology*, *22*, 247–251.

Morrison, J.D. (1991). Personal reminiscences of forty years of mass spectrometry in Australia. *Organic Mass Spectrometry*, *26*, 183–194.

Munson, M.S.B., Field, F.H. (1966). Chemical ionization mass spectrometry I: general introduction. *Journal of the American Chemical Society*, *88*, 2621–2630.

Myerson, S., Vander Haar, R.W. (1962). Multiply charged ions in mass spectra. *Journal of Chemical Physics*, *37*, 2458–2462.

Nier, A.O. (1985). Mass spectrometry in planetary research. *International Journal of Mass Spectrometry and Ion Processes 66*, *55*, 73.

Nier, A.O. (1989). Some reminiscences of mass spectrometry and the Manhattan Project. *Journal of Chemical Education*, *66*, 385–388.

Octoxynol. Merck Index, 12th edn. Entry # 6858.

Olsen, M.A., Cummings, P.G., Kennedy-Gabb, S., Wagner, B.M., Nicol, G.R., Munson, B. (2000). The use of deuterium oxide as a mobile phase for structural elucidation by HPLC/UV/ESI/MS. *Analytical Chemistry*, *72*, 5070–5078.

Porter, Q.N. (1985). *Mass Spectrometry of Heterocyclic Compounds*, 2nd edn, John Wiley & Sons, Inc., New York.

Quayle, A. (1987). Recollections of mass spectrometry of the fifties in a UK Petroleum Laboratory. *Organic Mass Spectrometry*, *22*, 569–585.

Randall, L.G., Wahrhaftig, A.L. (1981). Direct coupling of a dense (supercritical) gas chromatograph to a mass spectrometer using a supersonic molecular beam interface. *Review of Scientific Instrumentation*, *52*, 1283–1295.

Reinhold, B.B., Reinhold, V.N. (1992). Electrosprayionization mass spectrometry: deconvolution by an entropy-based algorithm. *Journal of the American Society for Mass Spectrometry*, *3*, 207–215.

Reynolds, W.E., Bacon, V.A., Bridges, J.C., Coburn, T.C., Halpern, B., Lederberg, J., Levinthal, E.C., Steed, E., Tucker, R.B. (1970). A computer operated mass spectrometer system. *Analytical Chemistry*, *42*, 1122–1129.

Robb, D.B., Covey, T.R., Bruins, A.P. (2000). Atmospheric pressure photoionization: an ionization method for liquid chromatography-mass spectrometry. *Analytical Chemistry*, *72*, 3653–3659.

Schulten, H.R., Beckey, H.D. (1972). Field desorption mass spectrometry with high temperature activated emitters. *Organic Mass Spectrometry*, *6*, 885–985.

Scott, R.P.W., Scott, C.G., Munroe, M., Hess, J., Jr (1974). Interface for on-line liquid chromatography-mass spectroscopy analysis. *Journal of Chromatography*, *99*, 395–405.

Settle, F.A. (2002). Analytical chemistry and the Manhattan Project. *Analytical Chemistry*, *74*, 36A–43A.

Sharp, T.R., Marquez, B.L. (2006). Hyphenated methods, in *Comprehensive Analytical Chemistry. Volume47: Modern Instrumental Analysis, Amsterdam* (eds S. Ahuja and N. Jespersen), Elsevier, Amsterdam, pp. 691–754.

Sharp, T.R., Smith, D.E., Heah, P.C., Buhro, W.E., Crocco, G.L., Bodner, G.S., Fernandez, J.M., Gladysz, J.A. (1986). *Advances in Mass Spectrometry*, *10B*, 1359–1360.

Sharp, T.R., Lee, H., Ferguson, A., Marsh, K.N., Harvey, R.G. (1988). 36th Annual Conference on Mass Spectrometry and Allied Topics.

Sharp, T.R., Horan, G.J., Morris, R., Pezzullo, L.H., Stroh, J.G. (2006). Method for determining the average degree of substitution of *o*-vanillin derivatized porcine somatotropin. *Journal of Pharmaceutical and Biomedical Analysis*, *40*, 185–189.

Smith, R.D., Udseth, H.R. (1983). Mass spectrometry with direct supercritical fluid injection. *Analytical Chemistry*, *55*, 2266–2272.

Smith, R.M. (2004). *Understanding Mass Spectra*, 2nd edn, John Wiley & Sons, Inc., New York, Preface, pg xi.

Squires, G. (1998). Francis Aston and the mass spectrograph. *Journal of the Chemical Society, Dalton Transactions*, 3893–3899.

Stroh, J.G., Cook, J.C., Milberg, R.M., Brayton, L., Kihara, T., Huang, Z., Rinehart, K.L.Jr, , Lewis, I.A.S. (1985). On-line liquid chromatography fast atom bombardment mass spectrometry. *Analytical Chemistry*, *57*, 985–991.

Sundqvist, B., Macfarlane, R.D. (1985). $^{252}$Cf-Plasma desorption mass spectrometry. *Mass Spectrometry Reviews*, *4*, 421–460.

Surman, D.J., Vickerman, J.C. (1981). Fast atom bombardment quadrupole mass spectrometry. *Chemical Communications*, 324–325.

Svec, H.J. (1985). Mass spectroscopy—ways and means. A historical prospectus. *International Journal of Mass Spectrometry and Ion Processes*, *66*, 3–29.

Sweetman, B.J., Blair, I.A. (1988). 3-nitrobenzyl alcohol has wide applicability as a matrix for FAB MS. *Biomedical and Environmental Mass Spectrometry*, *17*, 337–340.

Syage, J.A. (2004). Mechanism of $[M^{+}H]+$ formation in photoionization mass spectrometry. *Journal of the American Society for Mass Spectrometry*, *15*, 1521–1533.

Syage, J.A., Short, L.C., Cai, S.-S. (2008). Atmospheric pressure photoionization—the second source for LC-MS? *LCGC North America*, *26* (3), 286–296.

Takáts, Z., Wiseman, J.M., Gologan, B., Cooks, R.G. (2004). Mass spectrometry sampling under ambient conditions with desorption electrospray ionization. *Science*, *306*, 471–473.

Takáts, Z., Wiseman, J.M., Gologan, B., Cooks, R.G. (2005). Ambient mass spectrometry using desorption electrospray ionization (DESI): instrumentation, mechanisms and applications in forensics, chemistry, and biology. *Journal of Mass Spectrometry 40*, 1261–1275.

Tal'Rose, V.L., Lyubimova, A.K. (1952). *Doklady Akadamea Nauk SSSR*, *86*, 909.

Tal'Rose, V.L., Karpov, G.V., Grdetskii, I.G., Skurat, V.E. (1968). Capillary system for the introduction of liquid mixtures into an analytical mass spectrometer. *Russian Journal of Physical Chemistry*, *42*, 1658–1664.

Torgerson, D.F., Skowronski, R.P., Macfarlane, R.D. (1974). New approach to the mass spectrometry of non-volatile compounds. *Biochemical and Biophysical Research Communications*, *60*, 616–621.

Tyler, A.N., Clayton, E., Green, B.N. (1996). Exact mass measurement of polar organic molecules at low resolution using electrospray ionization and a quadrupole mass spectrometer. *Analytical Chemistry*, *68*, 3561–3569.

Vail, T.M., Jones, P.R., Sparkman, O.D. (2007). Rapid and unambiguous identification of melamine in contaminated pet food based on mass spectrometry with four degrees of confirmation. *Journal of Analytical Toxicology*, *31*, 304–312.

Van Berkel, G.J. (2000). Electrolytic deposition of metals on to the high-voltage contact in an electrospray emitter: implications for gas-phase ion formation. *Journal of Mass Spectrometry*, *35*, 773–783.

Vestal, M.L. (1983). Ionization techniques for nonvolatile molecules. *Mass Spectrometry Reviews*, *2*, 447–480.

Vestal, M.L. (2001). Methods of ion generation. *Chemical Reviews*, *101*, 361–375.

Watson, J.T., Sparkman, O.D. (2007). *Introduction to Mass Spectrometry: Instrumentation, Applications and Strategies for Data Interpretation*, 4th ed. John Wiley & Sons, Chichester, U.K.

Willoughby, R.C., Browner, R.F. (1984). Monodisperse aerosol generation interface for combining liquid chromatography with mass spectrometry. *Analytical Chemistry*, *56*, 2625–2631.

Wilson, I.D. (1999). The state of the art in thin-layer chromatography-mass spectrometry: a critical appraisal. *Journal of Chromatography A*, *856*, 429–442.

Wilson, I.D. (1999). The state of the art in thin-layer chromatography–mass spectrometry: a critical appraisal. *Journal of Chromatography A*, *856*, 429–442.

Witten, J.L., Schaffer, M.H., O'Shea, M., Cook, J.C., Hemling, M.E., Rinehart, K.L. (1986). Structure of two cockroach neuropeptides assigned by fast atom bombardment mass spectrometry. *Biochemical and Biophysical Research Communications*, *124*, 350–358.

Wood, G.W. (1982). Field desorption ionization: applications. *Mass Spectrometry Reviews*, *1*, 63–102.

Wu, Z., Rodgers, R.P., Marshall, A.G. (2004). Characterization of vegetable oils: detailed compositional fingerprints derived from electrospray ionization Fourier transform ion cyclotron resonance mass spectrometry. *Journal of Agricultural and Food Chemistry*, *52*, 5322–5328.

Yamashita, M., Fenn, J.B. (1984). Electrospray ion source. Another variation on the free-jet theme. *Journal of Physical Chemistry*, *88*, 4452–4459.

Yau, W.W., Kirkland, J.J., Bly, D.D. (1979). *Modern Size-Exclusion Liquid Chromatography*, John Wiley & Sons, Inc., New York, pp. 5–13.

Yergey, A.L., Edmonds, C.G., Lewis, I.A.S., Vestal, M.L. (1990). *Liquid Chromatography/Mass Spectrometry: Techniques and Applications*, Plenum Publishers, New York, p. 306.

Yergey, A.L., Yergey, A.K. (1997). Preparative scale mass spectrometry: a brief history of the Calutron. *Journal of the American Society for Mass Spectrometry*, *8*, 943–953.

# Approaches to Performing Metabolite Elucidation: One Key to Success in Drug Discovery and Development

ALA F. NASSAR

Absorption, distribution, metabolism, excretion, and toxicology (ADMET) studies are widely used in drug discovery and development to help obtain the optimal balance of properties necessary to convert lead compounds into drugs that are safe and effective for human use. Drug discovery efforts have been aimed at identifying and addressing metabolism issues at the earliest possible stage, by developing and applying innovative liquid chromatography–mass spectrometry (LC-MS)-based techniques and instrumentation, which are both faster and more accurate than prior techniques. Such new approaches are demonstrating considerable potential to improve the overall safety profile of drug candidates throughout the drug discovery and development process. These emerging techniques streamline and accelerate the process by eliminating potentially harmful candidates earlier and improving the safety of new drugs. In the area of drug metabolism, for example, revolutionary changes have been achieved by the combination of LC-MS with innovative instrumentation such as triple quadrupoles, ion traps, Orbitrap, and time-of-flight mass spectrometry. In turn, most ADMET studies have come to rely on LC-MS for the analysis of an ever-increasing workload of potential candidates. This chapter provides a discussion on the importance of LC-MS in supporting metabolic activation testing, metabolite characterization, and radiolabeled-drug testing.

## 1 INTRODUCTION

Mass spectrometry (MS) and nuclear magnetic resonance (NMR) are critical to the success of such ADMET studies. NMR spectroscopic techniques are most often used to confirm and elucidate metabolite identification in drug metabolism studies. Liquid

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

chromatography (LC)-NMR is a good choice for these studies, but MS has advantages over NMR with respect to sensitivity, smaller sample size, and greater speed compared with NMR. LC-MS-NMR has become a commercially available technique and is used in the late discovery stages to confirm and characterize metabolites. LC-MS is an analytical technique that still shows room for development; already, significant improvements have been made in sensitivity and resolution. It is probably the most powerful technique currently available for pharmaceutical analysis, and has significantly accelerated the drug discovery and development process. LC-MS has become the dominant technique for performing almost all of the analyses involved in ADMET studies, and is likely to remain the principal tool for such studies.

The major aim of LC-MS is the application of its analytical power to create straightforward, sensitive, fast, and reliable data. Improved hardware and software for LC-MS have led to greater sensitivity, greater ease-of-use, and improved post-analysis of data (Kubinyi, 1977; Bakke *et al.*, 1995; Lesko *et al.*, 2000; Meyboom, Lindquist and Egberts, 2000; Thompson, 2000; Greene, 2002; Lasser *et al.*, 2002; Taylor *et al.*, 2002; Tiller and Romanyshyn, 2002; Yang *et al.*, 2002; Jemal *et al.*, 2003; Kostiainen *et al.*, 2003; Lin *et al.*, 2003; Nassar, 2003; Nassar and Adam, 2003; Plumb *et al.*, 2003; Roberts, 2003; Kassel, 2004; Nassar and Lopez-Anaya, 2004; Nassar and Talaat, 2004; Balimane *et al.*, 2005; Castro-Perez *et al.*, 2005a; Johnson and Plumb, 2005; Leclercq *et al.*, 2005). Techniques such as electron impact, chemical ionization, atmospheric pressure chemical ionization (APCI), fast atom bombardment, thermospray, gas chromatography–mass spectrometry (GC-MS), and electrospray ionization (ESI) are used in ADMET studies. Metabolite characterization ion trap ($MS^n$—this refers to multistage MS/MS experiments where n is the number of product ion stages (progeny ions)) and quadrupole time-of-flight (QTOF)-MS are widely used; $MS^n$ because it allows the relatively rapid construction of fragmentation maps and a higher degree of specificity than with other methods. QTOF-MS has advantages for metabolite identification over $MS^n$ or triple quadrupole MS, including fast mass spectral acquisition speed with high full-scan sensitivity, enhanced mass resolution, and accurate mass measurement capabilities which allow for the determination of elemental composition. Exact mass measurement is a valuable tool for solving structure elucidation problems by helping to confirm elemental composition, and is also invaluable in eliminating false positives and determining nontrivial metabolites. Accurate mass measurements are routine experiments performed by modern mass spectrometers. The accuracy of a measurement refers to the degree of conformity of a measured quantity to its actual true value. Accurate mass measurements may not be sufficient if mixtures are analyzed or if the isotopic fine structure of compounds needs to be evaluated. In such cases, high resolving power is needed; this can be easily achieved by using Orbitrap or Fourier-transform ion cyclotron resonance (FTICR)-MS. Also, chromatography or comprehensive chromatography (GC-GC, LC-LC), or multidimensional chromatography (LC-GC) can help to enhance the efficiency.

Data-dependent scans on QTOF or $MS^n$ provide a highly useful tool. Increasing sample complexity, sample volume restrictions, and throughput requirements necessitate that the maximum amount of useful information is extracted from a single experiment. Data-directed analysis (DDA) or data-dependent scans enable intelligent MS and MS/MS acquisitions to be performed automatically on multiple co-eluting components. DDA is able to make intelligent decisions about which ions to select for MS/MS using its inherent high resolution, exact mass measurement capability, and full mass range. One shortcoming, however, is that due to limited sensitivity, DDA often

overlooks minor metabolites that could be toxic or active in nature. MS/MS or $MS^n$ fragmentation data contain a tremendous variety of information. Unfortunately, this wealth of information is contained within a mosaic of rearrangements, fragmentations, ion physics, and gas-phase chemistry reactions, which we do not yet fully understand. MS/MS fragmentation data cannot solve all our problems without an expert scientist to analyze and interpret this information, and use it to identify and characterize metabolites. While this analysis does not provide certain, "yes-or-no" answers, it does aid in directing the researcher as to which further studies will be needed. Herein, one demonstrably successful analytical strategy for metabolite characterization and identifying reactive metabolites by LC-MS is presented.

## 2 CRITERIA FOR LC-MS METHODS

The growing realization of the importance of ADMET properties early in the drug discovery process has led to a dramatic increase in the numbers of compounds requiring screening. Errors, as well as false positives and/or false negatives, will always be a risk factor with any screening procedure in the early stages of drug discovery because of the resulting necessity for speed. While discovery scientists seek to improve their turnaround time, it is vital to maintain or even improve the quality of chromatographic resolution of the metabolites produced in screening efforts. The high-throughput methods needed to achieve the required speed would not be possible without the innovative enabling technologies of computing, automation, new sample preparation technologies, and highly sensitive and selective detection systems. To ensure the integrity of the data, the screening procedure should meet criteria such as relevance, effectiveness, speed, robustness, accuracy, and reproducibility (White, 2000; Nassar *et al.*, 2006).

## 3 MATRICES EFFECT

The analysis of compounds in complex biological matrices, such as blood, plasma, bile, urine, and feces samples, is probably the largest application of LC-MS/MS. LC-MS has become the principal technique used in quantitative bioanalysis due to a combination of factors, such as cost, ease, and speed of performing selected reaction monitoring (SRM), and a wide dynamic range and good sensitivity due to favorable signal-to-noise ratio (S/N) (Jemal and Bergum, 1992; Jemal, 2000; Nassar *et al.*, 2001; Nassar, Varshney and Getek, 2001; Zhang and Henion, 2001; Jemal and Ouyang, 2003; Guevremont, 2004; Castro-Perez *et al.*, 2005b; Kapron *et al.*, 2005). It is interesting to note that it was initially thought that the invention of MS/MS would enable the analysis of matrices without separation but suppression effects have prevented this. The presence of endogenous components in the matrix can suppress the analyte response, probably by competition for ESI droplet surface and hence ionization. These effects can cause differences in response between samples in matrices and standards, leading to difficulties in quantitative analysis and compound identification. This emphasizes the importance of the chromatographic step in the analysis, where good separation can reduce or eliminate these effects. Because of the high degree of selectivity routinely provided by SRM,

bioanalytical method development time for quantitative determinations of one or several analytes has been reduced to a few days or less. Although the SRM approach demonstrates excellent sensitivity, selectivity, and efficiency, one drawback is that it fails to produce the qualitative information required to support the recognition and structural elucidation of metabolites that could be present in the samples. Metabolite identification and structure elucidation requires development of a method that is separate and distinct from the quantitation of the known components present, which requires additional time, effort, and expertise. During spectral scanning, the duty cycle of a quadrupole mass spectrometer is such that only a small fraction of the total time is spent monitoring any one ion. To obtain an optimum S/N, a quadrupole analyzer must allow a limited number of selected ions to pass. Because most ions are filtered out, much of the qualitative information is lost. A potential strategy to deal with this situation is the use of LC-TOF-MS to generate data that will simultaneously provide qualitative and quantitative information about drug candidate metabolism and disposition (Jemal and Bergum, 1992; Jemal, 2000; Zhang and Henion, 2001; Castro-Perez *et al*., 2005b).

## 4 METABOLITE CHARACTERIZATION

Metabolite identification is crucial to the drug discovery and development process, as it can be used to investigate the phase I metabolites likely to be formed *in vivo*, the differences between species in drug metabolism, the major circulating metabolites of an administered drug and the phase I and phase II pathways of metabolism, and pharmacologically active or toxic metabolites, and also to help determine the effects of metabolizing enzyme inhibition and/or induction. The ability to produce this information early in the discovery phase is becoming increasingly important as a basis for judging whether a drug candidate merits further development. Table 1 shows a list of a wide variety of phase I and phase II biotransformations, together with the mass changes from the parent drug (modified from Mortishire *et al*., 2005). This table could also be used retroactively to find the potential reactions responsible for certain metabolites.

Metabolite identification enables early detection of potential metabolic liabilities or issues, provides a metabolism perspective to guide synthesis efforts with the aim of either blocking or enhancing metabolism so as to optimize the pharmacokinetic and safety profiles of newly synthesized drug candidates, and assists the prediction of the metabolic pathway(s) of potential drug candidates for development (Taylor *et al*., 2002; Kantharaj *et al*., 2003; Nassar, Kamel and Clarimont, 2004a, b). The author's experience has demonstrated that the best combination for accomplishing rapid and accurate metabolite identification involves a robotic system for sample preparation and *in silico* software to predict and find possible metabolites as well as to predict hypothetical metabolite chemical structure (Obach, 1997, 1999; Greene, 2002). This can be combined with the use of LC-MS to determine exact mass measurements (accurate mass) for sample analysis and LC-MS-NMR, and online hydrogen–deuterium (HD) exchange for further metabolite structure confirmation and elucidation. In the following sections, the relative merits of current and potential strategies for dealing with metabolite characterization in various stages of drug discovery and development are examined and illustrated with examples. Techniques, tools, and approaches are suggested for each of these stages, as summarized in Table 2.

**TABLE 1 Common biotransformation reactions.**

| Metabolic reaction | Monoisotopic mass change |
|---|---|
| Oxidative debromination | –61.9156 |
| tert-Butyl dealkylation | –56.0624 |
| Hydrolysis of nitrate esters | –44.9851 |
| Decarboxylation | –43.9898 |
| Isopropyl dealkylation | –42.0468 |
| Propyl ketone to acid | –40.0675 |
| tert-Butyl to alcohol | –40.0675 |
| Reductive dechlorination | –33.9611 |
| Hydroxymethylene loss | –30.0106 |
| Nitro reduction | –29.9742 |
| Propyl ether to acid | –28.0675 |
| Deethylation | –28.0312 |
| Decarboxylation | –27.9949 |
| Ethyl ketone to acid | –26.0519 |
| Isopropyl to alcohol | –26.0519 |
| Alcohols dehydration | –18.0105 |
| Dehydration of oximes | –18.0105 |
| Reductive defluorination | –17.9906 |
| Oxidative dechlorination | –17.9662 |
| Sulfoxide to thioether | –15.9949 |
| Thioureas to ureas | –15.9772 |
| Ethyl ether to acid | –14.0519 |
| Demethylation | –14.0157 |
| tert-Butyl to acid | –12.0726 |
| Methyl ketone to acid | –12.0363 |
| Ethyl to alcohol | –12.0363 |
| Two sequential desaturation | –4.0314 |
| Hydroxylation and dehydration | –2.0157 |
| Primary alcohols to aldehyde | –2.0157 |
| Secondary alcohols to ketone | –2.0157 |
| Desaturation | –2.0157 |
| 1,4-Dihydropyridines to pyridines | –2.0157 |
| Oxidative defluorination | –1.9957 |
| Oxidative deamination to ketone | –1.0316 |
| Demethylation and methylene to ketone | –0.0365 |
| 2-Ethoxyl to acid | –0.0363 |
| Oxidative deamination to alcohol | 0.984 |
| Isopropyl to acid | 1.943 |
| Demethylation and hydroxylation | 1.9792 |
| Ketone to alcohol | 2.0157 |
| Methylene to ketone | 13.9792 |
| Hydroxylation and desaturation | 13.9792 |
| Alkene to epoxide | 13.9792 |
| (O, N, S) methylation | 14.0157 |
| Ethyl to carboxylic acid | 15.9586 |
| Hydroxylation | 15.9949 |
| Secondary amine to hydroxylamine | 15.9949 |
| Tertiary amine to *N*-oxide | 15.9949 |
| Thioether to sulfoxide, sulfoxide to sulfone | 15.9949 |
| Aromatic ring to arene oxide | 15.9949 |

**TABLE 1** ***Continued***

| Metabolic reaction | Monoisotopic mass change |
|---|---|
| Demethylation and two hydroxylation | 17.9741 |
| Hydration, hydrolysis (internal) | 18.0106 |
| Hydrolysis of aromatic nitriles | 18.0106 |
| Hydroxylation and ketone formation | 29.9741 |
| Quinone formation | 29.9741 |
| Demethylation to carboxylic acid | 29.9742 |
| Hydroxylation and methylation | 30.0105 |
| Two hydroxylation | 31.9898 |
| Thioether to sulfone | 31.9898 |
| Alkenes to dihydrodiol | 34.0054 |
| Acetylation | 42.0106 |
| Three hydroxylation | 47.9847 |
| Aromatic thiols to sulfonic acids | 47.9847 |
| Glycine conjugation | 57.0215 |
| Sulfate conjugation | 79.9568 |
| Hydroxylation and sulfation | 95.9517 |
| Cysteine conjugation | 103.0092 |
| Taurine conjugation | 107.0041 |
| S-cysteine conjugation | 119.0041 |
| Decarboxylation and glucuronidation | 148.0372 |
| *N*-acetylcysteine conjugation | 161.0147 |
| Glucuronide conjugation | 176.0321 |
| Two sulfate conjugation | 191.9035 |
| Hydroxylation + glucuronide | 192.027 |
| GSH conjugation | 289.0732 |
| Desaturation + S-GSH conjugation | 303.0525 |
| S-GSH conjugation | 305.0682 |
| Epoxidation + S-GSH conjugation | 321.0631 |
| Two glucuronide conjugation | 352.0642 |

Below is a general list of tools that can be used to identify and characterize metabolites. It is not necessary to use all of these tools for each analysis; for example, LC-NMR can be saved for use in the late stage of drug discovery.

- *In silico*
- Full-scan LC-MS positive and negative modes
- Multistages MS ($MS^n$) positive and negative modes
- N-rule
- Ring double-bond equivalent (RDBE)
- Isotopic patterns
- Exact mass measurements
- HD experiments
- Stable isotope
- LC-MS-NMR

### 4.1 Use of LC-MS to Identify and Characterize Metabolites

LC-MS is used with various strategies in drug discovery for identifying compounds and/or their metabolites. One area in which the technique is employed is in confirming

**TABLE 2 Techniques that help in the search for, and identification of, metabolites in drug metabolism.**

| Stage of compound | Technique | Comments |
|---|---|---|
| Very early stage of drug discovery | *In silico* as well as *in vitro* screening | Useful during synthesis to help select/eliminate compounds |
| Ranking compounds for drug discovery | *In silico* as well as *in vitro* screening<br>LC-MS and LC-MS/MS techniques | Useful during synthesis efforts to adjust metabolism<br>Helps in identifying of simple/major metabolites, for example, dealkylations, and conjugations such as glucuronide<br>Predicts likely metabolites formed *in vivo* |
| Selected candidate in late drug discovery | *In silico* as well as *in vitro* screening<br>LC-MS and LC-MS/MS techniques<br>Online HD exchange, ion trap, Q-TOF and Orbitrap | Aids in determination of metabolic differences between species<br>Used to identify potential pharmacologically active or pharmacologically toxic metabolites |
| Nominate compound for clinical trials | Same as selected candidate in the late drug discovery. Also, LC-MS-NMR and LC-Radioactive-MS | Used to determine the percentage of metabolite formed *in vitro* or *in vivo*<br>Aids synthesis of metabolites for toxicology testing<br>Aids in comparison of human pathways<br>Aids identification of drug–drug interactions |

Reprinted from Nassar and Lopez-Anaya, 2004, with permission from Thomson Scientific.

the structure of a known compound. A second area of application, and most relevant to this discussion, is the identification of unknown metabolites of drug candidates.

## 5 STRATEGIES FOR IDENTIFYING UNKNOWN METABOLITES

When using full-scan mass spectra to search for metabolites, one should search for the most intense ion with singly charged ions, or doubly charged ions at 0.5 mass-to-charge ratio (m/z), adducts, multiply charged ions, and/or dimers in the full-scan spectrum. According to the nitrogen rule, if a compound contains either no nitrogen atoms or an even number, its molecular ion will be at an even mass number, while an odd number of nitrogen atoms will produce an odd mass number for the molecular ion. This rule, applied to the molecular ion, determines whether the unknown agent has an even or odd number of nitrogen atoms. A check of the isotopic peak of the molecular ions serves to confirm patterns, being aware of interferences from other ions. Lastly, by applying the RDBE rule, one can determine or confirm the total number of rings plus double bonds for a compound containing carbon, hydrogen, nitrogen, oxygen, and elements with the same valences as these. The value of DBE can be calculated according to the following equation:

$$\mathrm{RDBE} = \mathrm{C} - \tfrac{1}{2}[\mathrm{H} + \mathrm{F} + \mathrm{Cl} + \mathrm{Br} + \mathrm{I}] + \tfrac{1}{2}[\mathrm{N} + \mathrm{P}] + 1$$

Next, one would perform accurate mass measurement to confirm/determine possible molecular formulae, and then compare the product ion MS/MS or $MS^n$ spectra of the parent with the metabolites. After having identified molecular ions for possible metabolites, MS/MS or $MS^n$ should be performed, and comparisons made between the product ion MS/MS or $MS^n$ spectra of the parent and metabolites. Given the structure of the parent drug, with its corresponding fragmentation, elucidation of metabolite structure is greatly facilitated. The specific fragment ion that creates a shift in the m/z value leads to identification of the site of the modification on the molecule.

Once drug candidates reach the late drug discovery/candidate selection phase, LC-MS/MS, QTOF, Orbitrap (high-resolution and exact mass measurement), ion trap ($MS^n$), and HD exchange come into prominence in order to determine metabolic differences between species and identify potential pharmacologically active or toxic metabolites (Table 2). Exact mass analysis serves to confirm the elemental composition of metabolites. The exact mass (accurate mass) shift between a drug candidate and its metabolites can be used to predict the elemental composition of those metabolites. Knowledge of the molecular formulae of unknown metabolites is one of the tools that can be used to identify the metabolites and then propose their structure using additional MS data and tools.

### 5.1 Online HD-LC-MS and Derivatization

HD exchange occurs in solution in the presence of exchangeable (labile) hydrogen atoms in a molecule. When this method is combined with LC-MS, it has the advantage of providing an easy estimation of the number of labile hydrogen atoms in such groups as -OH, -SH, -NH, $-NH_2$, and -COOH (Nassar, 2003). This number is useful in comparing metabolite structure with that of the parent drug to determine the presence or absence of these groups. HD exchange experiments are also valuable for structural elucidation and interpretation of MS/MS fragmentation processes (Nassar, 2003). A method for metabolite identification in drug discovery and development utilizing online HD exchange and a tandem QTOF mass spectrometer coupled with LC (LC-QTOF-MS) recently has been developed (Nassar, 2003). This method apparently works very well for the identification of metabolites produced by dehydrogenation, oxidation, glucuronidation, and dealkylation. It provides discrimination between *N*- or *S*-oxide formation and monohydroxylation; also, it becomes easy to identify conjugations such as quaternary amine glucuronide versus primary or secondary glucuronides using this technique. The generic method is simple, easy, fast, sensitive, robust, and reliable, as well as achieving enhanced throughput; these factors facilitate rapid characterization of metabolites *in vitro* or *in vivo*. This time saving, combined with the particular benefits of QTOF or Orbitrap, such as higher resolution, makes this technique a valuable tool for structure elucidation. This method also works well with low-resolution instruments. Derivatization can be useful to stabilize unstable metabolites, improve chromatographic properties for highly polar compounds, and reduce volatility for volatile metabolites. Furthermore, it can be useful for characterizing chirality and site of metabolism. Many methods of derivatization have been reported that can be used to identify most functional groups. These techniques have the limitation of time-consuming sample preparation. Derivatization of metabolites in a sample demands either isolation of a specific metabolite before derivatization, or derivatization of the entire sample, which

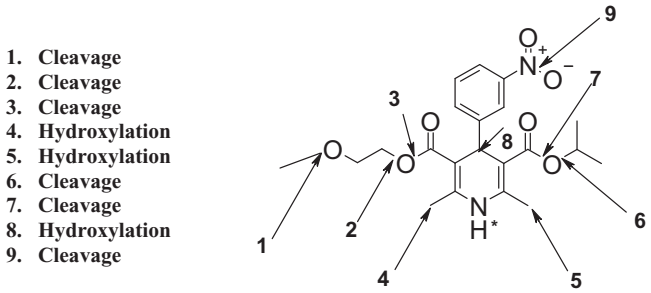


**Figure 1** The chemical structure of nimodipine with the labile H exchange as indicated by the asterisk. The arrows indicate the positions of the possible metabolites as predicted using Pallas software.

**TABLE 3 Exchange of labile hydrogens in nimodipine and metabolites formed *in vitro*.**

| Compound | Labile hydrogen | Mass $[M_H+H]^+$, m/z | Mass $[M_D+D]^+$, m/z | |
|---|---|---|---|---|
| | | | Predicted | Measured |
| Nimodipine | 1 | 419 | 421 | 421 |
| M-1 | 1 | 359 | 361 | 361 |
| M-2 | 2 | 435 | 438 | 438 |
| M-3 | 1 | 403 | 405 | 405 |
| M-4 | 2 | 405 | 408 | 408 |
| M-5 | 0 | 417 | 418 | 418 |

$M_H$ = the molecular weight in $H_2O$; $M_D$ = the molecular weight in $D_2O$.

will create more complex ions in MS. The speed and reduction in preparation time and effort involved with HD-LC-MS, and the fact that LC-MS-NMR has become more widely available and sensitive, have pushed derivatization methods out of favor.

As an example of how this process works, five metabolites of nimodipine formed in human liver microsomes; all were identified and characterized by LC-QTOF (Nassar, 2003). Figure 1 shows the chemical structure of nimodipine with the labile H exchange as indicated by asterisks; the arrows indicate the positions of the possible metabolites as predicted using Pallas software. Table 3 shows the exchange of labile hydrogens in nimodipine and its metabolites formed *in vitro*. The labile hydrogens ranged from none to two, which provides a significant means of identification, particularly with dehydrogenation metabolites, suggesting that the dehydrogenation took place on the pyridine moiety, resulting in a loss of one labile hydrogen. For example, in Fig. 2a, the full-scan mass spectrum of M-3 revealed a protonated molecular ion $[M_H+H]^+$ at m/z 403, 16 amu lower than nimodipine, suggesting that this metabolite was a cleaved product. The product ion spectrum of $[M_H+H]^+$ at m/z 403 showed the fragment ions of m/z 361, 343, 317, and 301. When $H_2O$ was replaced with $D_2O$ in the mobile phase, the full-scan mass spectrum of M-3 revealed a molecular ion $[M_D+D]^+$ at m/z 405 (Fig. 2b), 2 mass units higher than $[M_H+H]^+$ M-3, indicating the presence of one exchangeable hydrogen atom

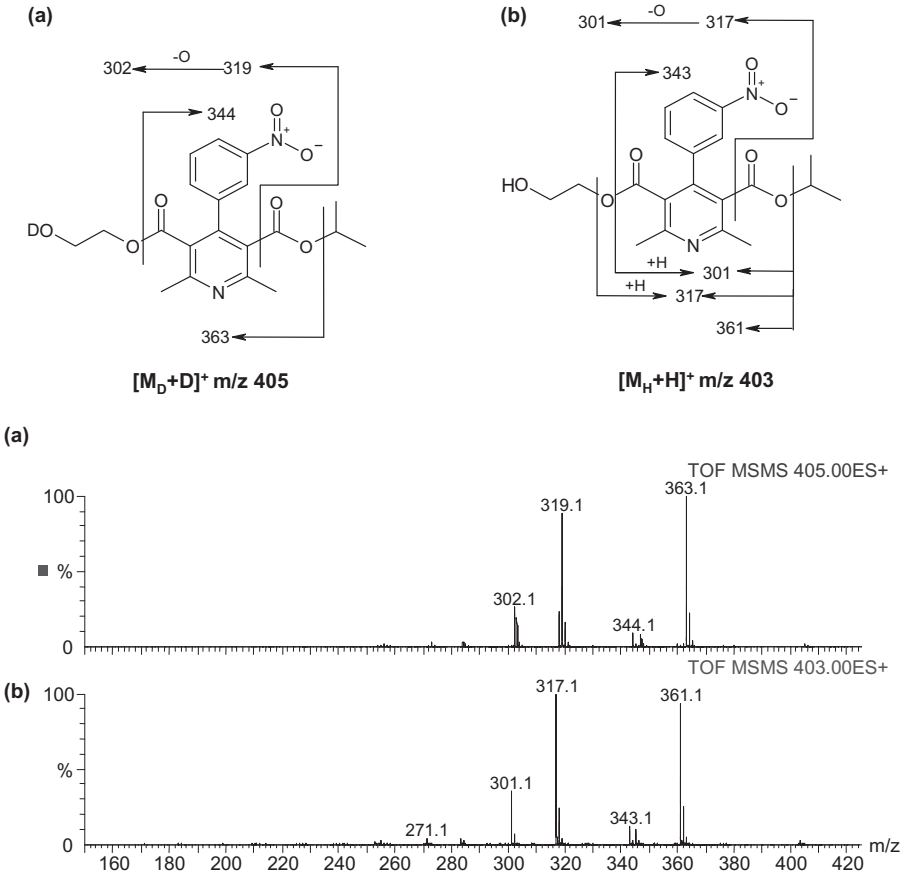


**Figure 2** QTOF MS/MS spectra of metabolite M-3 in **(a)** $D_2O$ and **(b)** $H_2O$. Reprinted from Nassar, 2003, with permission from Preston Publications.

in M-3. The product ion spectrum of m/z 405 showed the fragment ions of m/z 363, 344, 319, and 302. Nimodipine has one exchangeable hydrogen atom, with the cleavage producing another exchangeable hydrogen atom. This eliminates the possibility that M-3 is due to direct cleavage of nimodipine. It is possible that M-3 can be formed from the dehydrogenation and cleavage of nimodipine. The proposed TOF-MS/MS fragmentation for M-3 in $D_2O$ and $H_2O$ is shown in Fig. 2.

The proposed metabolic pathways of nimodipine in hepatic microsomal incubations in $H_2O$ and $D_2O$ are shown in Figs 3 and 4. Nimodipine is metabolized by means of dehydrogenation, demethylation of the methoxy group, cleavage of the ester groups by hydrolysis or oxidation, and hydroxylation of methyl groups. HD exchange provided significant information for all five metabolite identifications. One advantage of the HD exchange method is that, with LC-MS/MS, it offers an easy estimation of the number of labile hydrogen atoms in such groups as -SH, -OH, -NH, $-NH_2$, and -COOH. This number is useful in comparing metabolite structure with that of the parent drug to determine the presence or absence of these groups. HD exchange experiments have

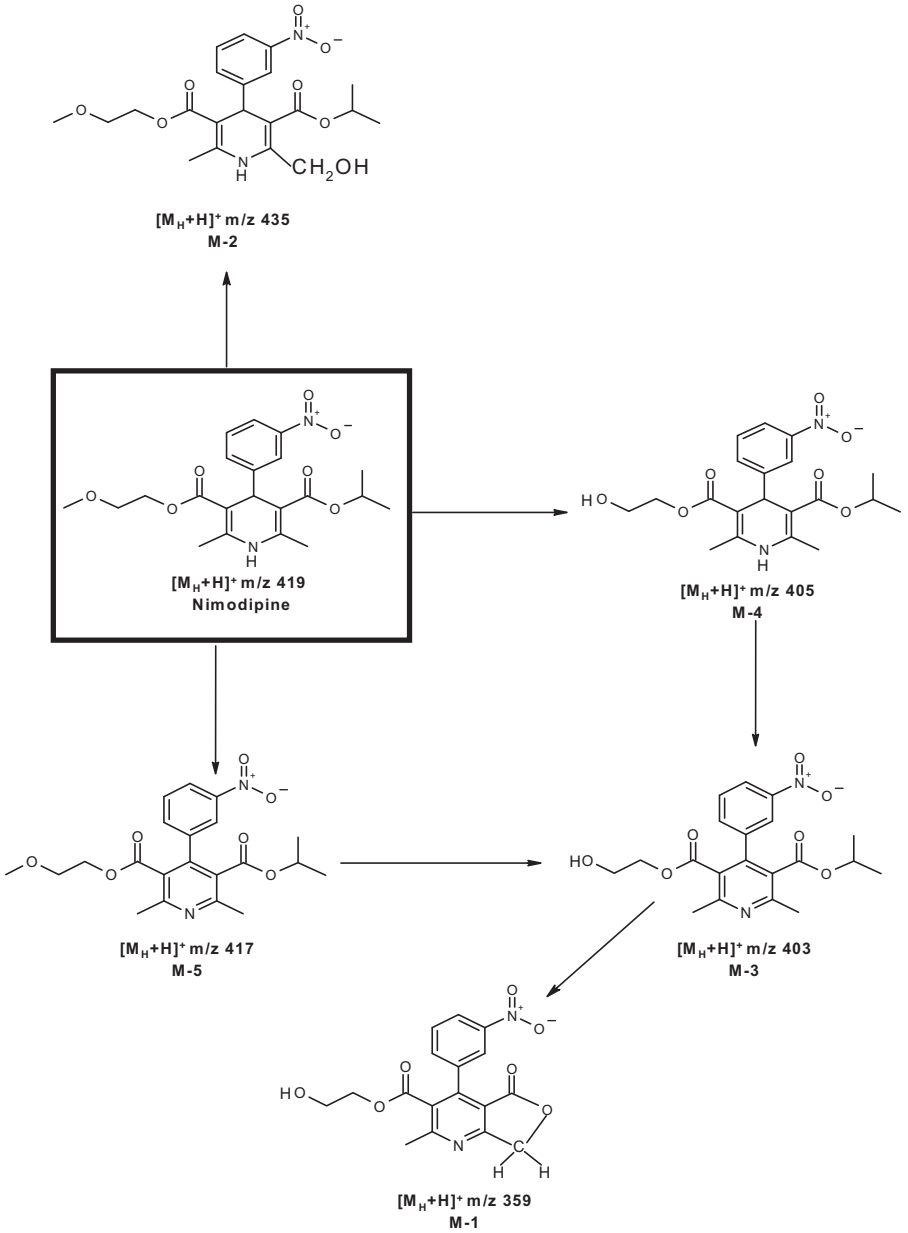


**Figure 3** Proposed metabolic pathways of nimodipine in hepatic microsomal incubations in $H_2O$. Reprinted from Nassar, 2003, with permission from Preston Publications.

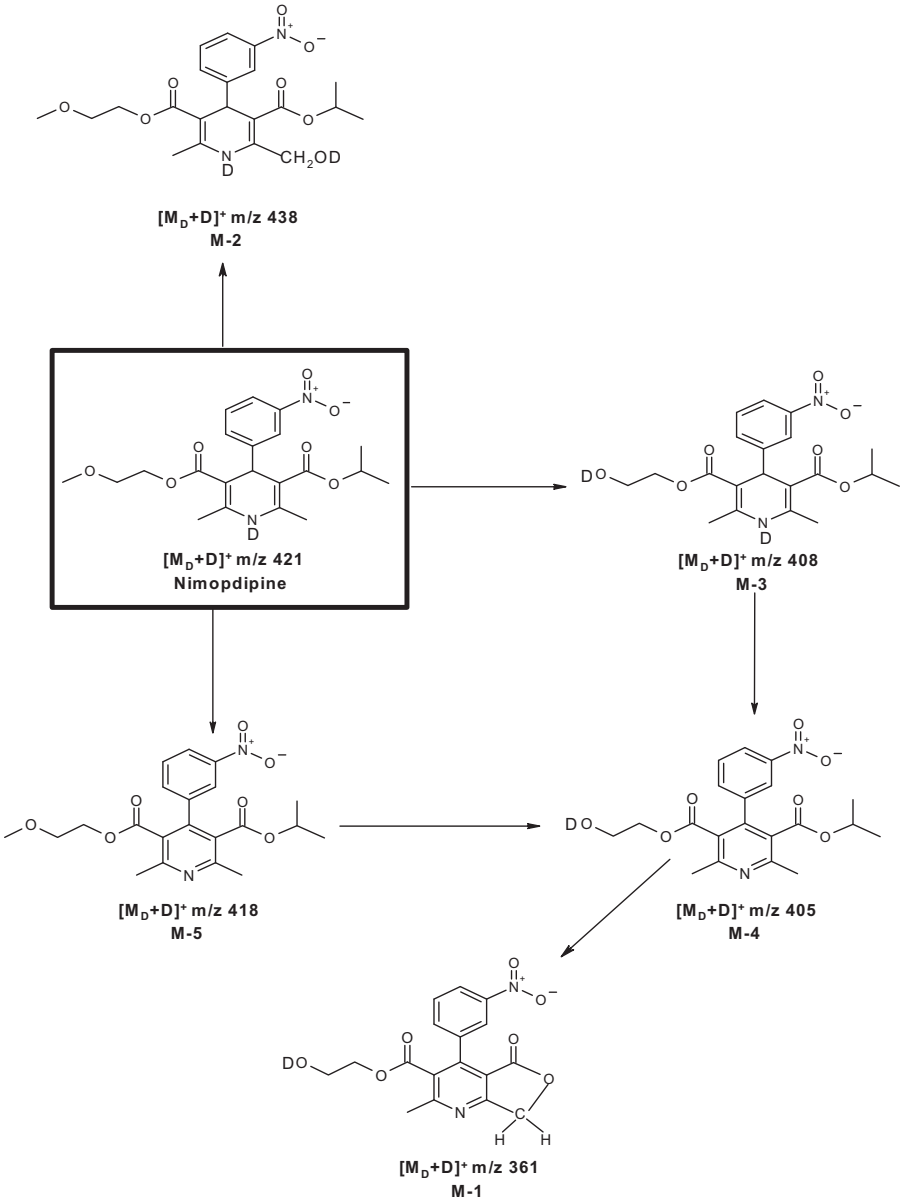


**Figure 4** Proposed metabolic pathways of nimodipine in hepatic microsomal incubations in $D_2O$. Reprinted from Nassar, 2003, with permission from Preston Publications.

facilitated structural elucidation and interpretation of fragmentation processes as well. These results show that this method should be particularly desirable for identification of metabolites produced by dehydrogenation, oxidation, and dealkylation.

### 5.2 LC-MS-NMR

For determination of the structure of metabolites, NMR spectroscopy has become one of the most commonly used techniques. The sensitivity of NMR has increased significantly with the introduction of modern high-field-strength (500, 600, and 800 MHz) NMR spectrometers. For the purposes of drug metabolism, the most commonly accessible nuclei are $^{1}H$, $^{19}F$, $^{13}C$, $^{31}P$, and $^{15}N$. When using NMR, sample concentrations must be between 1 μg and 10 mg, with analysis times between 5 minutes and 1 hour. Several operational modes are available for LC-MS-NMR such as isocratic or gradient elution, continuous flow, stop flow, time-sliced stop flow, peak collection into capillary loops for post-chromatographic analysis, and automatic detection of chromatographic peaks with triggered NMR acquisition. Given considerations for sample concentration and chromatographic resolution, any of these techniques is available.

The online HD-LC-MS method has the ability to rapidly provide characterization of metabolites, such as those formed by dealkylation, those from *in vitro* or *in vivo* samples, and can distinguish between *N*- or *S*-oxide versus monohydroxyl and quaternary amine versus primary or secondary glucuronide. Despite these advantages, NMR is still necessary for identifying the regiochemistry of aromatic oxidation, determining the site of aliphatic oxidation where fragmentation pathways are unavailable or inconclusive, and locating groups such as OH, epoxide, and sulfate by comparing NMR spectra of the parent with the metabolite. When seeking to identify unknown metabolites, NMR and MS are clearly complementary methods, and confirmation of a definite metabolite structure often requires data from both methods. There are strong benefits from the high quality and dual nature of information gained from a single run when using LC-MS-NMR; as the instrumentation and techniques involved undergo continued evolution, efficiency and sensitivity in terms of the chromatographic properties are bringing LC-MS-NMR well on the way to becoming a mainstream approach.

### 5.3 Stable Isotope Labeling

To improve the reliability of metabolite identification by full-scan MS, stable isotopic labeling ($^{2}H$, $^{13}C$, $^{15}N$, $^{18}O$, $^{34}S$) has been used. Also, the use of stable isotope-labeled drugs allows safe experiments in humans, which provides great advantages over the use of radiolabels. In this type of study, a mixture of known amounts of labeled and non-labeled drug is used for the metabolic experiments and analyzed by LC-MS. The identification of a metabolite should have the following criteria: (i) two peaks with identical shapes and retention times must be recorded in the ion chromatograms; (ii) the mass difference of the two peaks must be the same as the mass difference between the labeled and the unlabeled parent drug; and (iii) the relative abundance ratio of the peaks must be the same as the concentration ratio of the labeled and the unlabeled parent drug. Labeling may not be necessary in those cases where the compound includes one or more Cl or Br atoms, which will show abundant "mass + 2" isotopes with known abundance ratios in the mass spectra. Also, the metabolite standards must be synthesized, and in some cases chemical synthesis of metabolites may be difficult. Alternatively, enzymatic synthesis of metabolites has been found to be an

easy and rapid way to produce a few milligrams of desired metabolites for structure confirmation.

## 6 "ALL-IN-ONE" RADIOACTIVITY DETECTOR, STOP FLOW, AND DYNAMIC FLOW FOR METABOLITE IDENTIFICATION

Many of the studies of drug metabolism such as absorption, bioavailability, distribution, biotransformation, excretion, metabolite identification, and other pharmacokinetic research use radiolabeled drugs, with the radioactive isotopes $^{14}C$ or tritium ($^{3}H$) most commonly used for the labeling of a given drug. Radioactive labeling is a good fit with HPLC separation, as it allows high-resolution, quantitative detection of unknown metabolites and real-time monitoring by connecting the HPLC-radioactivity detector outlet to MS. The use of these detector interfaces aids the generation of data for structural elucidation of metabolites and biotransformation pathways for an administered drug. The detection sensitivity of both radioactivity and MS is the key to successfully identify and characterize metabolites (Nassar, Bjorge and Lee, 2003; Nassar *et al.*, 2004).

Regulatory policy dictates that exposure to administered radioactivity be held as low as possible in most studies, which demands much greater sensitivity of the radioactivity detector to be able to detect metabolites. There has been great progress in the emerging science of detecting trace amounts of radiolabeled or non-radiolabeled drugs and their metabolites. The availability of these technologies should have a dramatic impact on drug discovery and development for metabolite profiling studies. It has been reported that a microplate scintillation counter combined with capillary LC can be used to enhance sensitivity by eluent fractionation and subsequent offline counting. The limitations with this method are that the sample must be completely dry before counting, that any volatile compounds are likely to be lost, and that there is the potential for apolar compounds to adsorb on the surface of the plate. In addition, this technique has the limitations of time-consuming sample preparation, high analysis costs, and the inability to elucidate metabolite structure.

A new development for metabolite identification in drug discovery and development is a detection method which couples online stop flow and dynamic flow with a radioactivity detector and mass spectrometer. Figure 5 shows the hardware schematic diagram of the ν.ARC system. This system is simple and requires no custom-made

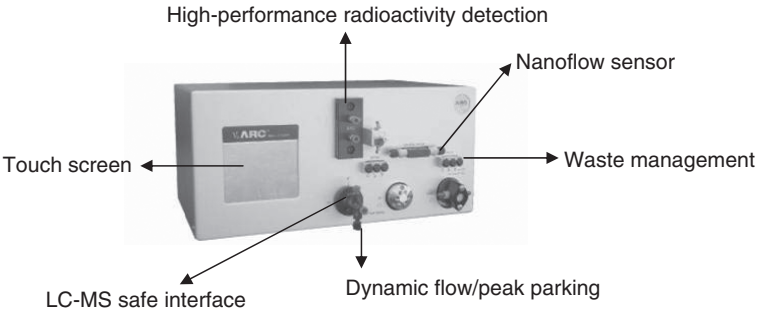


**Figure 5** Hardware schematic diagram of the LC-ARC system. Courtesy of Aim Research Company, Hochessin, DE.

hardware. The ν.ARC system uses ARC's specially designed cells, and is operated under the ARC Data System which controls the entire radio-LC system, including LC and the radioactivity detector. When interfaced with LC-MS, the ν.ARC system enhances detection of low-level radioactive peaks while increasing the flexibility and productivity of MS testing. This system has the major benefit of enhancing the sensitivity of radioisotope measurement for metabolite identification in drug metabolism studies. Another advantage to this system is the easy interface with the mass spectrometer, which allows acquisition of mass spectrometric data online. This system dramatically improves the sensitivity for $^{14}C$ peaks by up to 10-fold over conventional flow-through detection methods, and eliminates the need for a fraction collector and time-consuming sample preparation. In addition, the system produces accurate column recovery, quantification of low-level radioactivity, and consistently high resolution throughout the run. These factors give the combination of radioactivity detector, stop flow, and dynamic flow and mass spectrometry great potential as a powerful tool for improving the sensitivity of radioisotope measurement in metabolite identification studies, which also highlights the impressive progress that has been made in the technology of radioisotope counting in drug metabolism.

When online dynamic flow coupled with a radioactivity detector and mass spectrometer is used, the total run time remains similar to conventional radio-LC. A technique has been developed to collect the peaks online and then infuse them to the MS, which allows acquisition of a higher order of multistage fragmentation for both major and minor metabolites. The multistage MS fragmentation pattern obtained for the metabolites enabled determination of the sites of metabolism.

### 6.1 Online Peak Collection and Multistage Mass Experiments

It is important to note that this system can also be used to collect the peaks of interest online (peak parking), then infuse them for extended periods of time at flow rates as low as 1 μL/min while maintaining the column pressure. The peaks of interest can be triggered by radioactivity signals, UV signals, or specified retention time, allowing analysts to use direct infusion with a low flow rate, with only one run and a small sample size. The peak parking feature allows sustained analytical signal input (infusion fashion), which allows any or all of the following experiments to be done as desired: optimizing mass spectrometric condition; tuning for any individual metabolite, tuning for both positive and negative polarities, and optimizing collision energy. Also, it allows the operator to conduct multistage mass experiments, automatically acquire ion mass data, perform $MS^n$, compare isotopic mass spectra, and perform neutral loss tests. This method provides the capability of identifying the structures of unknown metabolites or impurities, again requiring a limited sample amount and a single run. This is a significant improvement over an offline fraction collector, which may lose volatile compounds during the fraction collection process. Because this method retains these compounds for analysis, it greatly expands the ability to characterize and identify metabolites of a given compound, which in turn is of significant benefit to analysts.

For purposes of evaluating this system, *in vitro* human liver microsomal (HLM) incubations were performed with only [$^{14}C$]-dextromethorphan (Fig. 6), a semisynthetic narcotic cough-suppressing ingredient in a variety of over-the-counter cold and cough medications. The online separation and identification of dextromethorphan metabolites did not require intensive sample preparation, concentration or fraction collection. Following incubation of [$^{14}C$]-dextromethorphan with human liver microsomes for

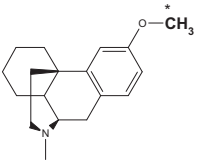


**Figure 6** Chemical structure of [$^{14}$C]dextromethorphan, the test compound used for this study (* = $^{14}$C).

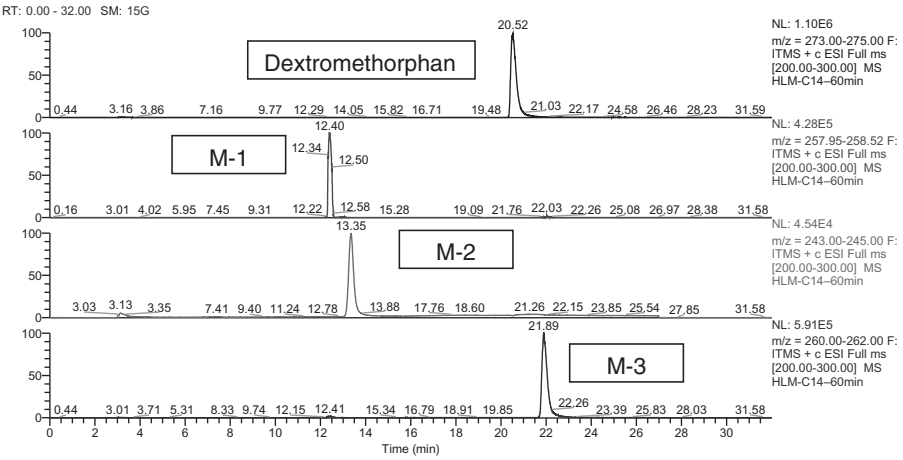


**Figure 7** HPLC-MS chromatograms of [$^{14}$C]dextromethorphan following incubation with pooled human liver microsomes for 60 minutes in the presence of NADPH showing M-1, M-2, M-3, and non-metabolized dextromethorphan. Reprinted from Nassar and Lee, 2007, with permission from Preston Publications.

60 minutes in the presence of nicotinamide adenine dinucleotide phosphate (NADPH), three metabolites were detected—M-1, M-2, and M-3—along with the parent drug. The retention times of dextromethorphan and these metabolites were between 12 and 23 minutes with excellent separation efficiency. Metabolites M-1, M-2, and M-3 have retention times of 12.4, 13.3, and 22.9 minutes, respectively. Figure 7 shows the LC-MS chromatograms of dextromethorphan following incubation with HLM for 60 minutes in the presence of NADPH. The LC-radio chromatogram shows that M-1 and M-2 did not have radioactive peaks, while M-3 and the parent dextromethorphan did. The full-scan mass spectra for M-1, M-2, and M-3 revealed protonated molecular ions $[M+H]^+$ at m/z 258, 244, and 260, suggesting that M-1 lost one methyl that contained C14, M-2 lost two methyl groups (one of which contained C14), and M-3 lost one methyl group.

The structures of dextromethorphan and metabolites were elucidated by LTQ (linear trap quadrupole)-MS/MS analysis. These peaks were collected online and then infused at flow rate 10 μL/min; $MS^2$, $MS^3$, and $MS^4$ were performed with sufficient time to examine the MS data and decide on the next fragment ions to be used for metabolite characterizations. Separate runs were performed for $MS^2$, $MS^3$, and $MS^4$, each without collecting the peaks; the results were similar. Figure 8 represents LC-MS spectra of

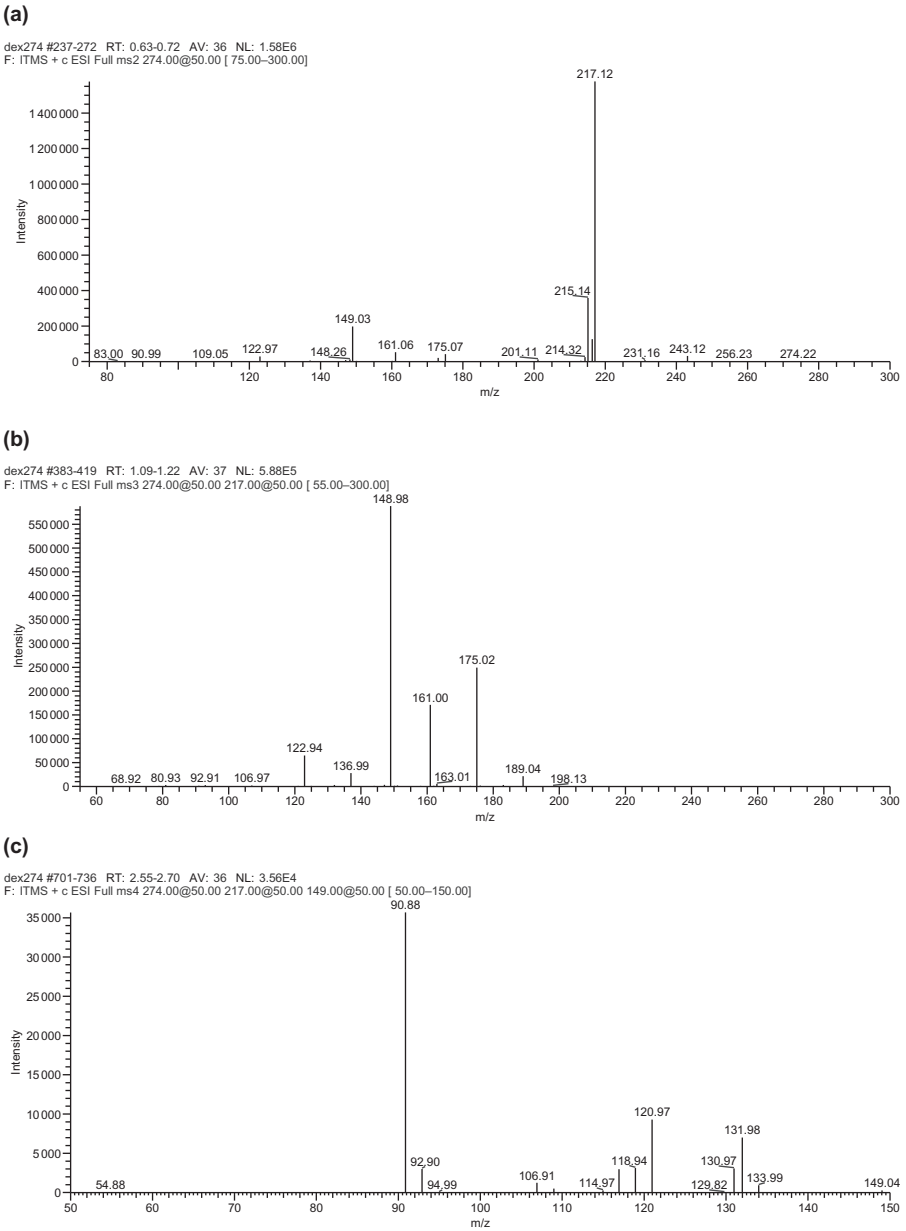


**Figure 8** Representative LC-MS spectra of $[M+H]^+$ m/z 274, M-1: **(a)** $MS^2$, **(b)** $MS^3$, and **(c)** $MS^4$ following incubation of $[^{14}C]$dextromethorphan with human liver microsomes. Reprinted from Nassar and Lee, 2007, with permission from Preston Publications.

non-metabolized [$^{14}$C]dextromethorphan [M+H]$^+$ at m/z 274: (a) MS$^2$, (b) MS$^3$, and (c) MS$^4$. The fragment ions are m/z 217, 215, 201, 175, 149, 123, 121, and 91. Comparison of the fragmentation ions for M-1 with the parent drug fragmentation suggests that *N*-dealkylation took place on the dextromethorphan molecule. Figure 9 shows proposed metabolic pathways of dextromethorphan in hepatic microsomal incubations. Mass spectrometric analysis showed the presence of dextromethorphan metabolites formed by *N*- and *O*-dealkylation, correlating with previously published results.

One of the major benefits of this method is that it is up to 10 times more sensitive in detecting $^{14}$C peaks than commercially available flow-through radioactivity detectors. This method enhances the resolution of radiochromatograms and is able to measure volatile metabolites. Another advantage to this system is the easy interface with the mass spectrometer, which allows acquisition of mass spectrometric data online. The method gives accurate column recovery and quantification of low-level radioactivity and high resolution throughout the run. An important safety benefit is that by using this method, injection size has been reduced, thereby decreasing potential exposure to

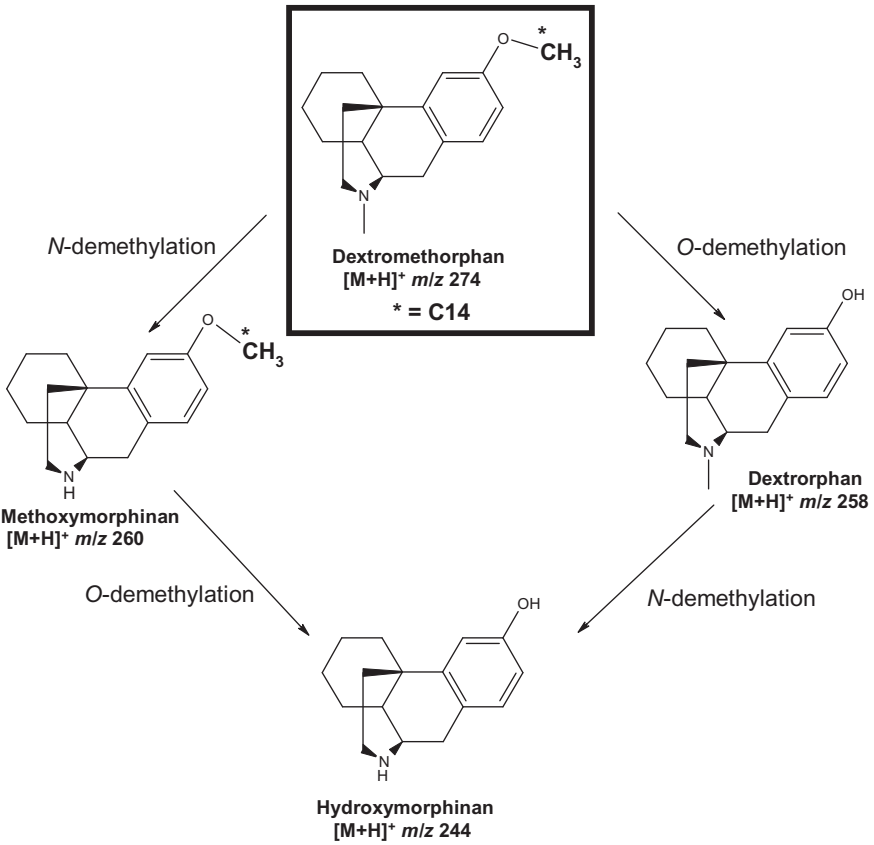


**Figure 9** Proposed metabolic pathways of dextromethorphan in human hepatic microsomal incubations based on LC-ARC-MS/MS data. Reprinted from Nassar and Lee, 2007, with permission from Preston Publications.

radioactivity and reducing the amount of radioactive wastes. This smaller injection size, as well as the capability to perform many tests on a single run, also reduces both time and expenses significantly. Furthermore, it is easier because it reduces manual operations. This study showed that impressive progress that has been made in the technology of radioisotope counting and metabolite characterization in drug metabolism using LC-ARC. The overall results suggest that the combination of LC-ARC dynamic flow with radioactivity detection and mass spectrometry has great potential as a powerful tool for radioisotope measurement in metabolite identification studies during drug discovery and development.

### 6.2 Metabolic Activation Studies by Mass Spectrometry

Adverse drug reactions can be classified into type A (predictable) and type B (unpredictable or idiosyncratic) reactions. Type A reactions are dose-dependent and predictable based on the pharmacology of the drug; it is possible to reverse the reactions by reducing the dosage or, if necessary, discontinuing the drug altogether. Type B reactions cannot be predicted on the basis of the pharmacology of the drug. Type B reactions include anaphylaxis, blood dyscrasias, hepatoxicity, and skin reactions, and may result from a combination of genetic predisposition and environmental factors. Such unacceptable consequences can result from the occurrence of idiosyncratic drug reactions during late-stage clinical trials or even after a drug has been released. It is this uncertainty that has driven researchers to develop improved means for predicting the potential for a drug to cause such reactions. Currently, it appears that idiosyncratic drug reactions may be due to reactive metabolites. A typical cause of toxicity is the formation of electrophilic reactive metabolites which bind to nucleophilic groups present in vital cellular proteins and/or nucleic acids. Significant evidence from previously published reports indicates that chemically reactive metabolites may be responsible for serious forms of toxicity, including cellular necrosis, mutagenesis, carcinogenesis, teratogenesis, hypersensitivity, and blood disorders.

## 7 STRATEGIES TO SCREEN FOR REACTIVE METABOLITES

LC-MS may be used to screen for reactive electrophiles generated from *in vitro* experiments in several tissues (e.g. liver microsomes) following reaction with glutathione (GSH) (Nassar and Lopez-Anaya, 2004; Nassar, Kamel and Clarimont, 2004a). There are several types of reactive metabolites (epoxides, arene oxides, quinones, quinone imines, quinone methides, iminoquinone methides, nitroso derivatives, nitrenium ions, nitro reduction products, nitro radicals, iminium ions, Michael acceptors, *S*-oxides, aliphatic aldehydes, or hydrolysis/acetylation products) that can be trapped in the presence of either GSH or an equimolar mixture of *N*-acetyl-cysteine and *N*-acetyl-lysine. This method is relatively simple and rapid, and has been implemented for high-throughput screening to identify reactive metabolites generated by bioactivation with phase I and phase II enzymes. Since thousands of compounds must be screened in a typical pharmaceutical industry setting, it is almost impossible, and highly expensive, to have all of them radiolabeled. Using radiolabeled reagents such as GSH and potassium cyanide is easier and more cost-effective for obtaining an accurate measurement of adduct formation. The reactive intermediates may be formed in other subcellular fractions such as liver S9, mitochondria, and cytosol (e.g. valproic acid forms reactive

metabolites in subcellular mitochondria), and thus, screening for reactive intermediates in these fractions will minimize false negatives. As an example of such a screening procedure, the author compared *in vitro* cellular and subcellular models for identifying drug candidates with the potential to produce reactive metabolites. For purposes of evaluating these models, six known compounds currently in the clinic but with dose limitations due to hepatotoxicity were selected. Labetalol, acetaminophen, niacin, iproniazid, 8-hydroxyquinoline, and isoniazid were incubated separately with human liver mitochondria, S9, microsomes, cytosol, and hepatocytes. LC-QTOF-MS methods were used to elucidate metabolic profiles of selected compounds. The phase II metabolites were in good agreement with previously published results, although without supplementation of reagents in subcellular fractions, phase II metabolites were not detected. Hepatocytes also formed GSH conjugates but the relative abundance was lower, indicating that the best model to identify potential reactive metabolites may involve the subcellular fractions. The results from incubation of acetaminophen and labetalol, respectively, in subcellular fractions are described elsewhere. The results demonstrated that for acetaminophen, when reactive intermediates are produced, they can be trapped by GSH; while in the case of labetalol, no reactive intermediates were produced and subsequently no GSH adducts were detected.

## 8 SUMMARY

There is no doubt as to the value of LC-MS in drug metabolism, as it has proved to be a sensitive and specific technique, which gives it important roles in ADMET studies. This chapter has demonstrated that LC-MS can provide a rapid, accurate, and relatively easy-to-use approach for metabolism studies without compromising the quality of the assays. Although LC-MS has improved the efficiency of metabolite analysis, it is still a daunting challenge to identify all of the metabolites generated by a particular drug candidate. Despite the high resolution MS provides, neutral-loss and precursor ion scans may miss some metabolites with unexpected fragmentation. Also, the current MS techniques may not provide the sensitivity required to detect trace quantities of metabolites in complex biological matrices. It is the opinion of the author that future implementation plans will apply in the following areas:

1. Modern MS instrumentations can greatly assist the analyst in providing high-quality data in a more rapid and time-efficient manner. Techniques employing QTOF, Orbitrap, micro-, and capillary separation techniques with nanospray MS are emerging and gaining wider use. As the sensitivity and resolution of MS continue to improve, the technique will further enhance metabolite characterization. For example, another recent innovation, the hybrid two-dimensional quadrupole ion trap–Fourier transform ion cyclotron resonance mass spectrometer, offers good potential for metabolite identification with high sensitivity and resolution.
2. Good progress has been made and will continue in the automation of sample preparation and data handling.
3. Advances in column technology and LC-MS have made chiral separation routine; more assays now have direct coupling of various chiral columns with an atmospheric pressure ion source without compromising the detector sensitivity or the LC resolving power.

4. It is not always possible to determine the site of metabolic reaction in a drug molecule with MS. LC-MS-NMR provides unambiguous structural characterization of metabolites. Although the sensitivity of NMR is currently not sufficient for the analysis of metabolites in trace quantities, continuous improvements are being made.
5. Online HD-LC-MS has become more widely used, because it is much faster and requires less preparation than NMR for metabolite confirmation, and further progress is anticipated.
6. An important consideration is that any screening process for idiosyncratic drug reactions will produce false results, both positive and negative, and a goal must be to reduce or eliminate these, which will in turn lead to improved drug safety. More studies are required to improve our knowledge of, and hence the techniques used for, screening with the use of radiolabeled or stable isotope reagents, which would give an accurate measurement of the amount of reactive intermediate. Reactive intermediate formation should be investigated in subcellular fractions, which will aid in reducing false results.

## REFERENCES

Bakke, O.M., Manocchia, M., de Abajo, F., Kaitin, K.I., Lasagna, L. (1995). Drug safety discontinuations in the United Kingdom, the United States and Spain from 1974 through 1993: a regulatory perspective. *Clin. Pharmacol. Ther.*, *58* (1), 108–117.

Balimane, P.V., Pace, E., Chong, S., Zhu, M., Jemal, M., Pelt, CK (2005). A novel high-throughput automated chip-based nanoelectrospray tandem mass spectrometric method for PAMPA sample analysis. *J. Pharm. Biomed. Anal.*, *39* (1–2), 8–16.

Castro-Perez, J., Plumb, R., Granger, J.H., Beattie, I., Joncour, K., Wright, A. (2005a). Increasing throughput and information content for *in vitro* drug metabolism experiments using ultra-performance liquid chromatography coupled to a quadrupole time-of-flight mass spectrometer. *Rapid Commun. Mass Spectrom.*, *19* (6), 843–848.

Castro-Perez, J., Plumb, R., Liang, L., Yang, E. (2005b). A high-throughput liquid chromatography/tandem mass spectrometry method for screening glutathione conjugates using exact mass neutral loss acquisition. *Rapid Commun. Mass Spectrom.*, *19* (6), 798–804.

Greene, N. (2002). Computer systems for the prediction of toxicity: an update. *Adv. Drug Deliv. Rev.*, *54* (3), 417–431.

Guevremont, R. (2004). High-field asymmetric waveform ion mobility spectrometry: a new tool for mass spectrometry. *J. Chromatogr. A*, *1058* (1–2), 3–19.

Jemal, M. (2000). High-throughput quantitative bioanalysis by LC/MS/MS. *Biomed. Chromatogr.*, *14* (6), 422–429.

Jemal, M., Bergum, J. (1992). Effect of the amount of internal standard on the precision of an analytical method. *J. Clin. Pharmacol.*, *32* (7), 676–677.

Jemal, M., Ouyang, Z. (2003). Enhanced resolution triple-quadrupole mass spectrometry for fast quantitative bioanalysis using liquid chromatography/tandem mass spectrometry: investigations of parameters that affect ruggedness. *Rapid Commun. Mass Spectrom.*, *17* (1), 24–38.

Jemal, M., Ouyang, Z., Zhao, W., Zhu, M., Wu, W.W. (2003). A strategy for metabolite identification using triple-quadrupole mass spectrometry with enhanced resolution and accurate mass capability. *Rapid Commun. Mass Spectrom.*, *17* (24), 2732–2740.

Johnson, K.A., Plumb, R. (2005). Investigating the human metabolism of acetaminophen using UPLC and exact mass oa-TOF MS. *J. Pharm. Biomed. Anal.*, *39* (3–4), 805–810.

Kantharaj, E., Tuytelaars, A., Proost, P.E., Ongel, Z., Van Assouw, H.P., Gilissen, R.A. (2003). Simultaneous measurement of drug metabolic stability and identification of metabolites using ion-trap mass spectrometry. *Rapid Commun. Mass Spectrom.*, *17* (23), 2661–2668.

Kapron, J.T., Jemal, M., Duncan, G., Kolakowski, B., Purves, R. (2005). Removal of metabolite interference during liquid chromatography/tandem mass spectrometry using high-field asymmetric waveform ion mobility spectrometry. *Rapid Commun. Mass Spectrom.*, *19* (14), 1979–1983.

Kassel, D.B. (2004). Applications of high-throughput ADME in drug discovery. *Curr. Opin. Chem. Biol.*, *8* (3), 339–345.

Kostiainen, R., Kotiaho, T., Kuuranne, T., Auriola, S. (2003). Liquid chromatography/atmospheric pressure ionization-mass spectrometry in drug metabolism studies. *J. Mass Spectrom.*, *38* (4), 357–372.

Kubinyi, H. (1977). Quantitative structure-activity relationships. 7. The bilinear model, a new model for nonlinear dependence of biological activity on hydrophobic character. *J. Med. Chem.*, *20* (5), 625–629.

Lasser, K.E., Allen, P.D., Woolhandler, S.J., Himmelstein, D.U., Wolfe, S.M., Bor, D.H. (2002). Timing of new black box warnings and withdrawals for prescription medications. *J. Am. Med. Assoc.*, *287* (17), 2215–2220.

Leclercq, L., Delatour, C., Hoes, I., Brunelle, F., Labrique, X., Castro-Perez, J. (2005). Use of a five-channel multiplexed electrospray quadrupole time-of-flight hybrid mass spectrometer for metabolite identification. *Rapid Commun. Mass Spectrom.*, *19* (12), 1611–1618.

Lesko, L.J., Rowland, M., Peck, C.C., Blaschke, T.F. (2000). Optimizing the science of drug development: opportunities for better candidate selection and accelerated evaluation in humans. *Pharm. Res.*, *17* (11), 1335–1344.

Lin, J., Sahakian, D.C., de Morais, S.M., Xu, J.J., Polzer, R.J., Winter, S.M. (2003). The role of absorption, distribution, metabolism, excretion and toxicity in drug discovery. *Curr. Top. Med. Chem.*, *3* (10), 1125–1154.

Meyboom, R.H., Lindquist, M., Egberts, A.C. (2000). An ABC of drug-related problems. *Drug Saf.*, *22* (6), 415–423.

Mortishire, R.J., O'Connor, D., Castro-Perez, J.M., Kirby, J. (2005). Accelerated throughput metabolic route screening in early drug discovery using high-resolution liquid chromatography/quadrupole time-of-flight mass spectrometry and automated data analysis. *Rapid Commun. Mass Spectrom.*, *19*, 2659–2670.

Nassar, A.E. (2003). Online hydrogen-deuterium exchange and a tandem quadrupole time-of-flight mass spectrometer coupled with liquid chromatography for metabolite identification in drug metabolism. *J. Chromatogr. Sci.*, *41* (8), 398–404.

Nassar, A.E., Adam, P. (2003). Metabolite characterization in drug discovery utilizing robotic liquid-handling, quadrupole time-of-flight mass spectrometry and *in silico* prediction. *Curr. Drug Metab.*, *4* (4), 259–271.

Nassar, A.E., Lee, D.Y. (2007). Novel approach to performing metabolite identification in drug metabolism. *J. Chromatogr. Sci.*, *45* (3), 113–119.

Nassar, A.E., Lopez-Anaya, A. (2004). Strategies for dealing with reactive intermediates in drug discovery and development. *Curr. Opin. Drug Discov. Devel.*, *7* (1), 126–136.

Nassar, A.E., Bjorge, S.M., Lee, D.Y. (2003). On-line liquid chromatography-accurate radioisotope counting coupled with a radioactivity detector and mass spectrometer for metabolite identification in drug discovery and development. *Anal. Chem.*, *75* (4), 785–790.

Nassar, A.E., Kamel, A.M., Clarimont, C. (2004a). Improving the decision-making process in structural modification of drug candidates: reducing toxicity. *Drug Discov. Today*, *9* (24), 1055–1064.

Nassar, A.E., Kamel, A.M., Clarimont, C. (2004b). Improving the decision-making process in the structural modification of drug candidates: enhancing metabolic stability. *Drug Discov. Today*, *9* (23), 1020–1028.

Nassar, A.E., Talaat, R. (2004). Strategies for dealing with metabolite elucidation in drug discovery and development. *Drug Discov. Today*, *9* (7), 317–327.

Nassar, A.E., Parmentier, Y., Martinet, M., Lee, D.Y. (2004). Liquid chromatography-accurate radioisotope counting and microplate scintillation counter technologies in drug metabolism studies. *J. Chromatogr. Sci.*, *42* (7), 348–353.

Nassar, A.E., Varshney, N., Getek, T., Cheng, L. (2001). Quantitative analysis of hydrocortisone in human urine using a high-performance liquid chromatographic-tandem mass spectrometric-atmospheric-pressure chemical ionization method. *J. Chromatogr. Sci.*, *39* (2), 59–64.

Nassar, A.E., Talaat, R.E., Kamel, A.M. (2006). The impact of recent innovations in the use of liquid chromatography-mass spectrometry in support of drug metabolism studies: are we all the way there yet? *Curr. Opin. Drug Discov. Devel.*, *9* (1), 61–74.

Obach, R.S. (1997). Nonspecific binding to microsomes: impact on scale-up of *in vitro* intrinsic clearance to hepatic clearance as assessed through examination of warfarin, imipramine, and propranolol. *Drug Metab. Dispos.*, *25* (12), 1359–1369.

Obach, R.S. (1999). Prediction of human clearance of twenty-nine drugs from hepatic microsomal intrinsic clearance data: an examination of *in vitro* half-life approach and nonspecific binding to microsomes. *Drug Metab. Dispos.*, *27* (11), 1350–1359.

Plumb, R.S., Stumpf, C.L., Granger, J.H., Castro-Perez, J., Haselden, J.N., Dear, G.J. (2003). Use of liquid chromatography/time-of-flight mass spectrometry and multivariate statistical analysis shows promise for the detection of drug metabolites in biological fluids. *Rapid Commun. Mass Spectrom.*, *17* (23), 2632–2638.

Roberts, S.A. (2003). Drug metabolism and pharmacokinetics in drug discovery. *Curr. Opin. Drug Discov. Devel.*, *6* (1), 66–80.

Taylor, E.W., Jia, W., Bush, M., Dollinger, G.D. (2002). Accelerating the drug optimization process: identification, structure elucidation, and quantification of *in vivo* metabolites using stable isotopes with LC/$MS^n$ and the chemiluminescent nitrogen detector. *Anal. Chem.*, *74* (13), 3232–3238.

Thompson, T.N. (2000). Early ADME in support of drug discovery: the role of metabolic stability studies. *Curr. Drug Metab.*, *1* (3), 215–241.

Tiller, P.R., Romanyshyn, L.A. (2002). Liquid chromatography/tandem mass spectrometric quantification with metabolite screening as a strategy to enhance the early drug discovery process. *Rapid Commun. Mass Spectrom.*, *16* (12), 1225–1231.

White, R.E. (2000). High-throughput screening in drug metabolism and pharmacokinetic support of drug discovery. *Ann. Rev. Pharmacol. Toxicol.*, *40*, 133–157.

Yang, L., Amad, M., Winnik, W.M., Schoen, A.E., Schweingruber, H., Mylchreest, I., Rudewicz, P.J. (2002). Investigation of an enhanced resolution triple quadrupole mass spectrometer for high-throughput liquid chromatography/tandem mass spectrometry assays. *Rapid Commun. Mass Spectrom.*, *16* (21), 2060–2066.

Zhang, H., Henion, J. (2001). Comparison between liquid chromatography-time of-flight mass spectrometry and selected reaction monitoring liquid chromatography-mass spectrometry for quantitative determination of idoxifene in human plasma. *J. Chromatogr. B Biomed. Sci. Appl.*, *757* (1), 151–159.

# Structural Modifications of Drug Candidates: How Useful Are They in Improving Metabolic Stability of New Drugs? Part I: Enhancing Metabolic Stability

ALA F. NASSAR

The rule of three, which attempts to define the activity–exposure–toxicity triangle, presents the single most difficult challenge in the design and advancement of drug candidates to the development stage. Absorption, distribution, metabolism, and excretion (ADME) studies are widely used in drug discovery to optimize this balance of properties necessary to convert lead compounds into drugs that are both safe and yet effective for human patients. Metabolite characterization has become one of the main drivers of the drug discovery process, helping to optimize ADME properties and increase the success rate for drugs. The study of structural modifications produces valuable information which, throughout the process of drug discovery, keeps on giving direction and focus in the effort to improve and balance the relationships between activity, exposure, and toxicity for drug candidates. In this chapter, the strategies for the decision-making process in structural modification of drug candidates are addressed to improve metabolic stability. Several examples are included to show how metabolic stability has influenced and guided drug design.

## 1 BACKGROUND

During the process of metabolism, the molecular structure of a drug is changed from one that is absorbed (lipophilic, or capable of crossing the membrane lipid core of membranes) to one that can be readily eliminated from the body (incapable of crossing

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

the lipid core of membranes, or hydrophilic). If lipophilic drugs are not metabolized, they will remain in the body for longer than intended, and their cumulative biological effects will eventually cause harm. Thus, the formation of water-soluble metabolites not only enhances drug elimination but also leads to compounds that are generally pharmacologically inactive and relatively nontoxic. Phase I reactions include oxidative, reductive, and hydrolytic biotransformations. These reactions introduce a polar functional group (e.g. OH, COOH, $NH_2$, and SH) into the drug molecule and usually result in only a small increase in hydrophilicity. This can occur through direct introduction of the functional group (e.g. aromatic and aliphatic hydroxylation) or by modifying existing functionalities (e.g. reduction of ketone and aldehydes to alcohols; oxidation of alcohols to acids; hydrolysis of ester and amides to yield COOH, $NH_2$, and OH groups; reduction of azo and nitro compounds to give $NH_2$ moieties; oxidative *N*-, *O*-, and *S*-dealkylation to give $NH_2$, OH, and SH groups). Phase II reactions form water-soluble conjugated products by reaction with polar and ionizable endogenous compounds such as glucuronic acid, sulfate, glycine, and other amino acids to the functional groups of phase I metabolites. Most phase II biotransformation reactions result in a large increase in drug hydrophilicity, thus they greatly promote the excretion of foreign chemicals. Conjugated metabolites are readily excreted in the urine or bile depending on the molecular weight and are generally pharmacologically inactive and nontoxic. Other phase II pathways, such as methylation and acetylation, serve to terminate biological activity, while glutathione conjugation serves to protect the body against chemically reactive compounds. The opposite may sometimes be true: drugs can form metabolically activated or reactive electrophilic species in a process known as bioactivation. Drugs that already have existing functional groups, such as OH, COOH, and NH, are often directly conjugated by reactions with phase II enzymes.

## 2 INTRODUCTION

There is no doubt as to the value of predicting the ADME/TOX of compounds as early as possible in the drug discovery process. Pharmaceutical companies are coming under increasing pressure to more precisely prove long-term safety, and to provide more data on the safety of their products. Failure to successfully address these concerns causes companies to withdraw drugs from the market, not only those that present human health consequences but also negative economic and public relations impact on the pharmaceutical industry. Because of human health safety concerns, a number of drugs have been withdrawn from the market. Failures in the drug discovery process account for a substantial part of the cost of drug development because many compounds are discarded after substantial investment owing to unforeseen toxicological effects.

The drug discovery process in the pharmaceutical industry has undergone a dramatic change in the last decade due to advances in molecular biology, high-throughput pharmacological screens, and combinatorial synthesis (Kubinyi *et al.*, 1977; Bakke *et al.*, 1995; Pirmohamed *et al.*, 1998; Richard *et al.*, 1998; Lesko *et al.*, 2000; Meyboom *et al.*, 2000; Greene *et al.*, 2002; Lasser *et al.*, 2002; Taylor *et al.*, 2002; Tiller *et al.*, 2002; Uetrecht *et al.*, 2002; Kantharaj *et al.*, 2003; Kostiainen *et al.*, 2003; Nassar *et al.*, 2003a, b, 2004a, b; Plumb *et al.*, 2003; Watt, 2003). The high cost, long development time, and high failure rate in bringing drugs to market have been important factors behind this change. The major reasons for this low success rate include poor pharmacological activity, low bioavailability, or high toxicity. Thus, the design of drugs with optimal potency, pharmacokinetic properties, and less toxicity is challenging given the opposing requirements for

absorption and metabolism. During this process, high metabolic liability needs to be avoided which usually leads to poor bioavailability and high clearance as well as the formation of active or toxic metabolites, which also have an impact on the pharmacological and toxicological outcomes. One approach to this dilemma would be to thoroughly explore the structure requirements for *in vitro* pharmacological activity through combinatorial synthesis and identify the key structural features that are essential for activity for each chemical series, and then attempt to improve the metabolism, PK and toxicity properties through structure modifications to the regions of the molecule that have little or no impact on the desired activity. Thus, one can maintain an optimal level of potency while introducing structural features that will improve the metabolism and toxicity characteristics.

## 3 SIGNIFICANCE OF METABOLITE CHARACTERIZATION AND STRUCTURE MODIFICATION

Optimization of the PK properties of candidates is one of the important considerations during the drug discovery process. The rate and sites of metabolism of new chemical entities by drug-metabolizing enzymes are amenable to modulation by appropriate structural changes. Until recently, this optimization has been achieved largely by empirical methods and trial and error. Now, metabolite identification enables early identification of potential metabolic liabilities or issues, provides a metabolism perspective to guide synthesis efforts with the aim of either blocking or enhancing metabolism so as to optimize the pharmacokinetic and safety profiles of newly synthesized drug candidates, and assists the prediction of the metabolic pathways of potential drug candidates. This is crucial to the drug discovery process, as it can be used to investigate the phase I metabolites likely to be formed *in vivo*, differences between species in drug metabolism, the major circulating metabolites of an administered drug, phase I and phase II pathways of metabolism, pharmacologically active or toxic metabolites, and to help determine the effects of metabolizing enzyme inhibition and/or induction. The ability to produce this information early in the discovery phase has become increasingly important as a basis for judging whether a drug candidate merits further development.

Herein, the strategies for the decision-making process in structural modification of drug candidates are discussed to enhance the metabolic stability (Part I). Also, some of the chemical substructures that form reactive metabolites and are involved in toxicities in humans are discussed. Several examples from literature are discussed to show how metabolic stability studies have influenced drug design strategies, and helped to facilitate improvements.

## 4 PART I: ENHANCE METABOLIC STABILITY

### 4.1 *In Vitro* Studies

Various *in vitro* methods are available to evaluate metabolic stability. Among the most popular and widely utilized systems are liver microsomes. Microsomes retain activity of key enzymes involved in drug metabolism which reside in the smooth endoplasmic reticulum, such as cytochrome P450s (CYPs), flavin monooxygenases, and glucuronosyltransferases. Isolated hepatocytes offer a valuable whole cell *in vitro* model, and retain a broader spectrum of enzymatic activities, including not only reticular systems,

but cytosolic and mitochondrial enzymes as well. Because of a rapid loss of hepatocyte-specific functions, it has been possible to generate useful data only with short-term hepatocyte incubations or cultures. However, significant strides have been made toward enhancing the viability of both cryopreserved and cultured hepatocytes. Liver slices, like hepatocytes, retain a wide array of enzyme activities, and are used to assess metabolic stability. Furthermore, both hepatocytes and liver slices are capable of assessing enzyme induction *in vitro*. The choice of which system to employ in a drug discovery screening program is dependent on numerous factors such as information about a particular chemical series and availability of tissue. For example, if a pharmacological probe has been shown to induce CYP activity, it may be preferable to use liver slices rather than microsomes for metabolic screening in support of this program. Cross-comparison of metabolic turnover rates in various tissues from various species can be quite beneficial. In addition to providing information on potential rates and routes of metabolism, interspecies comparisons may help in choosing species to be used in efficacy and toxicology models. The use of *in vivo* methods to assess metabolism has advantages and disadvantages. On the plus side, data are being generated within a living species, allowing a full early PK profile for the compound to be assessed. However, the methods are generally of low throughput, are time-consuming, and can suffer from marked species differences. In this respect, *in vitro* studies can be used in conjunction with *in vivo* experiments in order to select the animal model with metabolism most similar to human.

## 5 METABOLIC STABILITY AND INTRINSIC METABOLIC CLEARANCE

Using the concept of intrinsic metabolic clearance ($CL_{int}$), it has been demonstrated that *in vitro* metabolism rates for a selected set of model substrates correlated well with hepatic extraction ratios determined from isolated perfused rat livers (Rane *et al.*, 1977). Thus, when measured under appropriate conditions, the intrinsic hepatic metabolic clearance ($CL_{int}$) of human and preclinical species can be predicted by measuring enzyme kinetics parameters: Km (Michaelis–Menten constant) and $V_{max}$ (maximum velocity) ($CL_{int} = V_{max} / Km$) (Segel *et al.*, 1993). However, it is time-consuming to measure Km and $V_{max}$. In drug discovery, one can use a single substrate concentration ($\ll$Km) to determine the parent drug disappearance half-life ($t_{1/2}$). The intrinsic hepatic metabolic clearance ($CL_{int}$) can be calculated based on scaling factors from microsomes as well as hepatocytes (Obach *et al.*, 1999a). Then, the hepatic metabolic clearance ($CL_h$) and hepatic extraction ratio ($E_h$) can be calculated (Obach *et al.*, 1999b). Integration of metabolic stability studies with the other high-throughput pharmacology screens and other ADME screens such as inhibition screening is essential. Solving a metabolic stability problem may not necessarily lead to a compound with an overall improvement in activity or even PK properties. For example, compounds with improved metabolic stability may be developed only to find out that absorption becomes a problem. Reduction in metabolic clearance may be accompanied by an increase in renal or biliary clearance of the parent drug. It is also possible that improvement of *in vitro* intrinsic clearance ($CL_{int}$) may result from self-inhibition of one or more drug-metabolizing enzymes. This apparent improvement is only realized during the *in vitro* screening process whereas *in vivo*, the compound may exhibit saturable metabolism, nonlinear PK, or drug–drug interactions. Thus, it is advisable that *in vitro* metabolic stability data be integrated with inhibition screening.

## 6 ADVANTAGES OF ENHANCING METABOLIC STABILITY

Several advantages to be gained from enhancing metabolic stability have been reported (Ariens *et al.*, 1982):

- Increased bioavailability and longer half-life, which in turn should allow lower and less frequent dosing, thus promoting better patient compliance.
- Better congruence between dose and plasma concentration, thus reducing or even eliminating the need for expensive therapeutic monitoring.
- Reduction in metabolic turnover rates from different species which, in turn, may permit better extrapolation of animal data to humans.
- Lower patient-to-patient and intra-patient variability in drug levels, since this is largely based on differences in drug metabolic capacity.
- Diminishing the number and significance of active metabolites and thus lessening the need for further studies on drug metabolites in both animals and man.

## 7 STRATEGIES TO ENHANCE METABOLIC STABILITY

To enhance microsomal stability, *in vitro* metabolism studies can be performed to confirm formation of metabolites as well as to provide quantitative analysis of major metabolites. After identifying moieties that contribute to activity (the pharmacophore) and other moieties necessary for activity, several modifications can be used to enhance metabolic stability. In general, metabolism can be reduced by incorporation of stable functions (blocking groups) at metabolically vulnerable sites. Substrate structure–activity relationships of metabolizing enzymes have to be accommodated within the structure–activity relationships of the actual pharmacological target. However, when it is appropriate to enhance metabolic stability in the molecular design, the following strategies have been used (Testa *et al.*, 1990; Humphrey and Smith, 1992):

- Deactivating aromatic rings toward oxidation by substituting them with strongly electron-withdrawing groups (e.g. $CF_3$, $SO_2NH_2$, $SO_3$–).
- Introducing an *N*-t-butyl group to prevent *N*-dealkylation.
- Replacing a labile ester linkage with an amide group.
- Constraining the molecule in a conformation which is unfavorable to the metabolic pathway, more generally, protecting the labile moiety by steric shielding.
- The phenolic function has consistently been shown to be rapidly glucuronidated. Thus, avoidance of this moiety in a sterically unhindered position is advised in any compound intended for oral use.
- Avoidance of other conjugation reactions as primary clearance pathways, would also be advised in the design stage in any drug destined for oral usage.
- Sometimes the best strategy is to anticipate a likely route of metabolism and prepare the expected metabolite if it has adequate intrinsic activity. For example, often *N*-oxides are just as active as the parent amine, but will not undergo further *N*-oxidation.

Examples from literature for these strategies to enhance metabolic stability in the molecular design are summarized in Table 1 (Genin *et al.*, 1996; Bouska *et al.*, 1997;

**TABLE 1 Enhancement of metabolic stability through structural modification.**

| Category | Approach/ strategy | Enzyme, pathway | Lead compound |
|---|---|---|---|
| I. Lipophilicity and metabolic stability | Reduce log P and log D | NA | 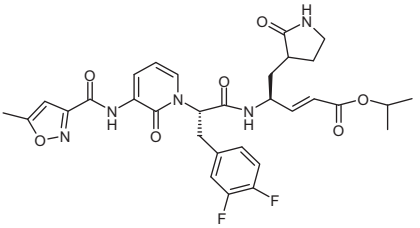 |
| | | NA | EC50 = 0.078 mM, ClogP = 2.07, C7 h = 0.012 μM |
| | Introduce isosteric atoms or polar functional group | NA | 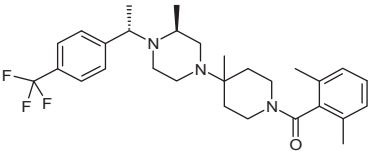 |
| | | | Ki = 1 nM, $AUC_{0\text{-}6h}$ = 922 ng/mL h |
| | | NA | |
| | | NA | |
| | | NA | |
| II. Metabolically labile groups | Remove or block the vulnerable site of metabolism | Benzylic oxidation | 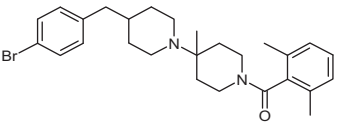 |
| | | | Ki = 66 nM, $AUC_{0\text{-}6h}$ = 40 ng/mL h |

| Optimized compound | Experimental model | Therapeutic Class | Reference |
|---|---|---|---|
| 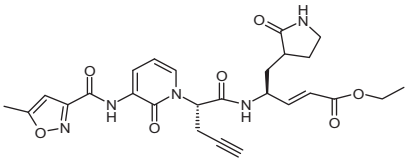 EC50 = 0.058 μM, ClogP = 0.18, C7h = 0.057 μM | Orally dosed monkey | 3C protease inhibitor | Dragovich *et al.* (2003) |
| 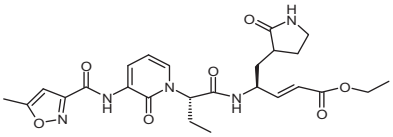 EC50 = 0.047 μM, ClogP = 0.66, C7h = 0.896 μM | Orally dosed monkey | 3C protease inhibitor | Dragovich *et al.* (2003) |
| 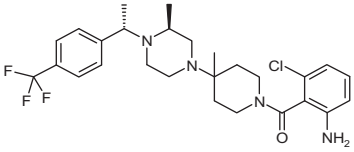 Ki = 5 nM, $AUC_{0\text{-}6h}$ = 1872 ng/mL h | Orally dosed rat | CCR5 antagonists | Tagat *et al.* (2001) |
| 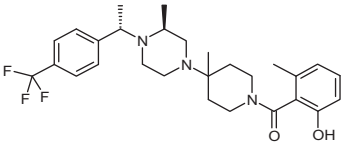 Ki = 5 nM, $AUC_{0\text{-}6h}$ = 2543 ng/mL h | Orally dosed rat | CCR5 antagonists | Tagat *et al.* (2001) |
| 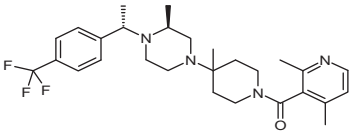 Ki = 2.3 nM, $AUC_{0\text{-}6h}$ = 3,905 ng/mL h | Orally dosed rat | CCR5 antagonists | Tagat *et al.* (2001) |
| 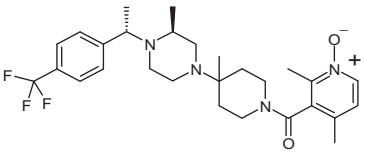 Ki = 7 nM, $AUC_{0\text{-}6h}$ = 9243 ng/mL h | Orally dosed rat | CCR5 antagonists | Tagat *et al.* (2001) |
| 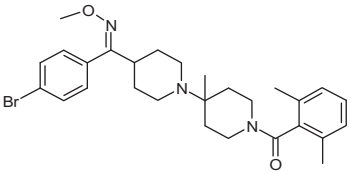 Ki = 2 nM, $AUC_{0\text{-}6h}$ = 1400 ng/mL h | Orally dosed rat | CCR5 antagonists | Palani *et al.* (2002) |

**TABLE1** ***Continued***

| Category | Approach/ strategy | Enzyme, pathway | Lead compound |
|---|---|---|---|
| | | Allylic oxidation | 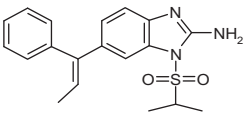 $IC_{50}$ = 0.06 μg/mL, $C_{max}$ = 14–140 ng/mL |
| | | Phenyl oxidation | 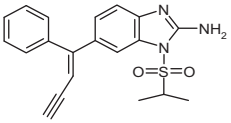 $IC_{50}$ = 0.02 μg/mL, %F = 9 |
| | | *N*-oxidation | 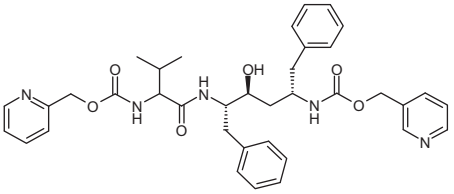 AUC = 1.98 μg·h/mL, %F = 26 |
| | | *N*-demethylation | 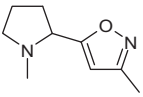 $t_{1/2}$ = 3 h, %F = 1.2 |
| | | *N*-dealkylation | 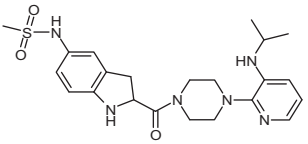 $t_{1/2}$ = 10.8 min |
| | | Ester hydrolysis | 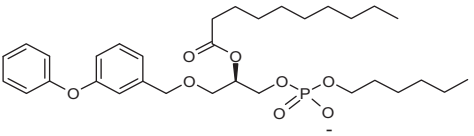 $t_{1/2}$ = 33 min, $C_{max}$ = 465 ng/mL, %F = 4 |

| Optimized compound | Experimental model | Therapeutic Class | Reference |
|---|---|---|---|
| 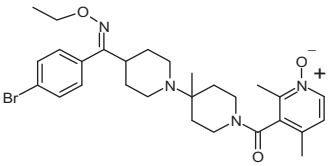 Ki = 2.1 nM, $AUC_{0\text{-}6h}$ = 6500 ng/mL h | Orally dosed rat | CCR5 antagonists | Palani *et al.* (2002) |
| 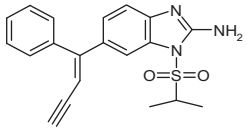 $IC_{50}$ = 0.02 μg/mL, $C_{max}$ = 70–300 ng/mL | Orally dosed monkey | Vinylacetylene antivirals | Victor *et al.* (1997) |
| 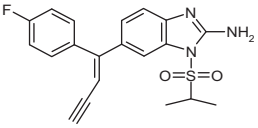 $IC_{50}$ = 0.04 μg/mL, %F = 23 | Orally dosed monkey | Vinylacetylene antivirals | Victor *et al.* (1997) |
| 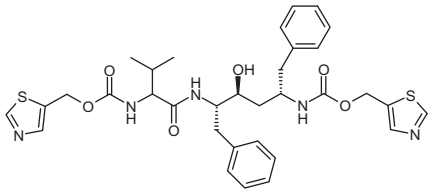 AUC = 4.24 μg·h/mL, %F = 47 | Orally dosed rat | HIV protease inhibitors | Kempf *et al.* (1998) |
| 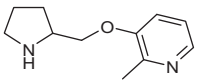 $t_{1/2}$ = 24 h, %F = 61.5 | Dog liver slices | NAChR ligands | Lin *et al.* (1997) |
| 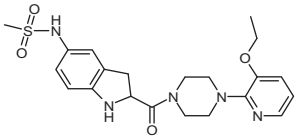 $t_{1/2}$ = 46.8 min | Rat liver microsomes | BHAP reverse transcriptase inhibitors | Genin *et al.* (1996) |
| 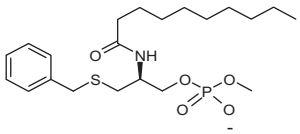 $t_{1/2}$ = 39 min, $C_{max}$ = 3261 ng/mL, %F = 90 | Rat blood and plasma, liver microsomes and homogenate | Phospholipase A inhibitors | Blanchard *et al.* (1998) |

**TABLE 1** ***Continued***

| Category | Approach/ strategy | Enzyme, pathway | Lead compound |
|---|---|---|---|
| | | Amide hydrolysis | 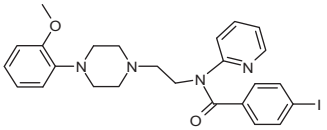 Ki = 0.2 nM, 40% and >60% degradation in human cytosole and microsomes, respectively |
| | | Glucuronidation (effect of linker) | 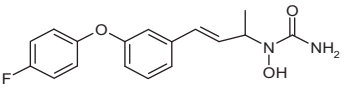 UDPGA rate (nmol/min/mg protein) = 0.19, $t_{1/2}$ = 4.7 h |
| | | Glucuronidation (effect of template) | 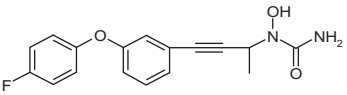 UDPGA rate (nmol/min/mg protein) = 0.05, $t_{1/2}$ = 5.5 h |
| | | Glucuronidation (effect of stereochemistry) | 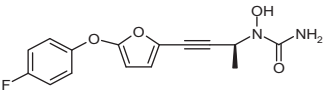 UDPGA rate (nmol/min/mg protein) = 0.02, $t_{1/2}$ = 7.7 h |

*Note*: NA = not applicable; $C_{max}$ = maximum plasma concentration; $K_i$ = inhibition constant; F = bioavailability. Reprinted from Nassar *et al*. 2004, with permission from Elsevier.

Lin *et al.*, 1997; Victor *et al.*, 1997; Blanchard *et al.*, 1998; Kempf *et al.*, 1998; Zhuang *et al.*, 1998; Tagat *et al.*, 2001; Palani *et al.*, 2002; Dragovich *et al.*, 2003). One strategy for improving the metabolic stability of a compound is to reduce the overall lipophilicity (log P, log D) of the structure. This is due to the fact that the metabolizing enzymes generally have a lipophilic binding site and hence more readily accept lipophilic molecules. As exemplified in a Pfizer study of the human rhinovirus 3C protease inhibitor (Dragovich *et al.*, 2003), the lead compound, which exhibited poor oral bioavailability *in vivo* in the monkey, was subjected to PK optimization with the benzyl group being identified as a position for reducing lipophilicity without compromising activity. Introduction of substituents with reduced lipophilicity led to the propargyl analog 2 and the ethyl analog 3, both of which had reduced ClogPs when compared with the parent. These compounds demonstrated improved oral exposure in the monkey, highlighting a simple yet effective way to improve the metabolic profile of a compound. This method is not always successful however, as lipophilic groups are usually involved in binding to the biological target.

An alternative approach toward lowering the lipophilicity of a lead is to introduce isosteric atoms or functional groups into the molecule that impart increased polarity. This could be in the form of a heteroatom (e.g. a nitrogen introduced into a benzene

| Optimized compound | Experimental model | Therapeutic Class | Reference |
|---|---|---|---|
| 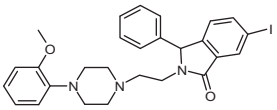 Ki = 0.069 nM, 10% and <5% degradation in human liver cytosole and microsomes, respectively | Human liver cytosole and microsomes | $5HT_{1A}$ receptor ligand | Zhuang *et al.* (1998) |
| 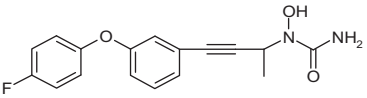 UDPGA rate (nmol/min/mg protein) = 0.05, $t_{1/2}$ = 5.5 h | Monkey liver microsomes and plasma | 5-LO inhibitors | Bouska *et al.* (1997) |
| 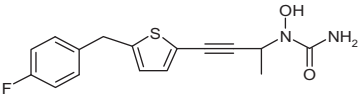 UDPGA rate (nmol/min/mg protein) = 0.012, $t_{1/2}$ = 14.5 h | Monkey liver microsomes and plasma | 5-LO inhibitors | Bouska *et al.* (1997) |
| 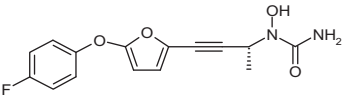 UDPGA rate (nmol/min/mg protein) = 0.01, $t_{1/2}$ = 8.7 h | Monkey liver microsomes and plasma | 5-LO inhibitors | Bouska *et al.* (1997) |

ring to form the more polar pyridine) or instigating functional group transformations toward more polar functionalities (e.g. conversion of a ketone to the corresponding carboxamide). This strategy was utilized in a Schering-Plough program to enhance the pharmacokinetics of a CCR5 antagonist lead in their HIV-1 inhibitor program (Tagat *et al.*, 2001). The benzamide of the lead compound was identified as a position for reducing lipophilicity, and changes were made to alter the substitution pattern of the ring. The more polar anthranilamide and salicylamide gave significant improvements in the rat oral blood levels, with the area under the curve (AUC) increasing substantially. However, the best results were achieved by introducing a heteroatom into the phenyl ring, with the more polar nicotinamide showing satisfactory oral blood levels in the rat. During this study, the authors noticed a single metabolite of nicotinamide rapidly forming in the plasma samples—they concluded this was due to *N*-oxidation. Thus, the pyridine *N*-oxide was prepared and found to be an excellent compound both in terms of potency and oral bioavailability in rat, dog, and monkey.

A much more elegant approach toward improving metabolic stability is to remove or block the vulnerable site of metabolism. For example, identification of sites potentially labile toward oxidation (e.g. benzylic or allylic positions) can be addressed by blocking the site of metabolism (introduction of a halogen onto the carbon atom

involved, replacing the benzylic $CH_2$ with an isostere such as O). Using this approach, workers at Schering-Plough conducted a metabolism-driven optimization of a second-generation lead from the aforementioned CCR5 antagonist program (Palani *et al.*, 2002). The lead compound exhibited good potency against the receptor, but showed very poor oral bioavailability in the rat. The benzylic position was identified as a metabolically labile site, and a variety of isosteres were introduced including O, CHOH, and C O among others. Eventually, the methoxime was identified as a potent, metabolically stable derivative and formed the next-generation lead compound.

The work of Victor *et al.* provides another example of blocking sites of potential oxidative metabolism to improve metabolic stability (Victor *et al.*, 1997). Replacing a methyl group vulnerable to allylic oxidation with acetylene resulted in up to fivefold improvement in blood levels of monkeys. A further improvement of about twofold in oral bioavailabilty in the monkey was achieved when an acetylenyl moiety replaced the methyl and the *p*-fluoro substitution was made on the phenyl ring. It may be that the electron-withdrawing character of fluorine deactivates the aromatic ring toward metabolic oxidation. Or perhaps the somewhat lipophilic character of fluorine coupled with its ability to hydrogen bond provides a unique mixture of physical properties which assist in the absorption process.

Kempf and coworkers employed the principle of bioisosterism in their campaign to synthesize the potent HIV protease inhibitor ritonovir (Kempf *et al.*, 1998). Pyridyl *N*-oxidation was a limiting route of metabolism in one of their early lead compounds. The authors embarked on a systematic search for other heterocyclic groups which would decrease the rate of metabolism while at the same time maintaining adequate solubility to ensure absorption as well as inhibitory activity. Ultimately, pyridyl groups were replaced with metabolism-resistant thiazole groups resulting in analog (ritonovir), which exhibited a nearly twofold improvement in bioavailability while maintaining potent inhibition of HIV protease.

Lin and colleagues (Lin *et al.*, 1997) observed up to about a 50-fold improvement in potency at the nicotinic acetylcholine receptor when the *N*-desmethyl analog was used instead of *N*-methyl analog, which was vulnerable to *N*-demethylation.

Genin and colleagues (Blanchard *et al.*, 1998) realized a significant improvement of about fourfold when they replaced a metabolically vulnerable 3-isopropylamino substituent on the lead compound with an ethoxy substituent. While one might reason that this substituent would also be rapidly metabolized, in this case it proved to be stable, at least when compared with the lead compound.

Functional groups can be open to other types of metabolism other than oxidation; it is well documented that amidases and esterases are present in the liver and can hydrolyze amides and esters, respectively. Additionally, phase II metabolism where polar functionality (e.g. glucuronides and sulfates) is imparted onto a molecule in order to make it more water-soluble and hence easier to clear, can be targeted as a method for improving overall metabolic stability.

Replacing a labile ester linkage with an amide group is known to impart added stability to esterase activity. In Blanchard's study (Blanchard *et al.*, 1998), this can be illustrated by observing the 22-fold improvement in bioavailability in amide when compared with its corresponding ester.

During a metabolism-driven optimization of a series of new arylpiperazine benzamido derivatives as potential ligands for 5-HT1A receptors, Zhuang *et al.* (1998) attributed the low brain uptake in the human subject to rapid metabolism (amid hydrolysis), whereas in other species (rats and monkeys), the metabolic pathway (cleavage

of the amid bond by liver microsomes) appears to be much slower. The authors reported that the cyclized amide derivatives provided better than twofold improvement in amide stability compared with the open-chain lead compound.

An extensive body of work on *N*-hydroxyurea inhibitors of 5-lipogenase related to zileuton provides an excellent opportunity to examine structure–activity relationships with respect to conjugation with glucuronic acid (Bouska *et al.*, 1997). It was discovered early on in this investigation that glucuronidation of the *N*-hydroxyl moiety was an activity-limiting step, so it became of pivotal importance to this program to impart metabolic stability toward this conjugation. The *N*-hydroxyurea compounds were divided into three areas for structural modifications: the template, the linking group, and the pharmacophore. The *N*-hydroxyurea was identified in zileuton as an optimal pharmacophore for potency and selectivity; therefore, it was not modified in the search for metabolically stable compounds. The template and link components were the major focus of medicinal chemistry efforts to optimize *in vivo* duration in the cynomolgus monkey. This group systematically modified first the linkage group joining the template with the *N*-hydroxyurea pharmacophore and then the benzthiophene template. Each compound was screened for stability to glucuronidation of the hydroxyurea. Of the linkage moieties, compounds with acetylene linkage groups were generally found to have lower uridine diphosphate glucuronic acid (UDPGA) rates than any of the links tested. This lower UDPGA rate resulted in the longest *in vivo* duration in monkey. Apparently, the rigid conformation required by the acetylenic group disrupts binding in the active site of UDPGA and diminishes conjugation.

Many variations of the template group were also tested in the UDPGA assay. Compounds containing the benzthiophene template found in zileuton demonstrated more rapid rates of glucuronidation. Conversely, biaryl templates such as phenoxyphenyl, phenoxyfuran, and phenoxythiophene demonstrated reduced rates of glucuronidation. These different metabolic stability results may be attributed to changes in lipophilicity and/or remote electronic effects.

Finally, the importance of the stereochemistry of the carbon center immediately adjacent to the hydroxyurea was investigated. *N*-hydroxyurea compounds containing a methyl branch on the linking group had already shown a longer *in vivo* duration than their methylene analogs, which encouraged further work with this methyl branch series. The addition of the methyl group introduced a chiral center and resulted in a compound with separable *R*- and *S*-enantiomers. Approximately twofold difference in the glucuronidation rate, depending on the configuration of the adjacent carbon center, was reported.

## 8 CONCLUSIONS

We have discussed strategies and tools to enhance metabolic stability to enable researchers in the drug companies to design drugs that are safer and more robust for humans. Structural information on metabolites helps in enhancing as well as streamlining the process of developing new drug candidates, which in turn has great value in several important aspects of drug discovery and development. By improving our ability to identify both helpful and harmful metabolites, suggestions for structural modifications will optimize the likelihood that other compounds in the series are more successful. *In silico* and *in vitro* techniques are available to screen compounds for key ADME characteristics, which, when applied within a rational strategy, can make a

major contribution to the design and selection of successful drug candidates. Structural modifications to solve a metabolic stability problem may not necessarily lead to a compound with an overall improvement in PK properties. Solving metabolic stability problems at one site could result in the increase in the rate of metabolism at another site, a phenomenon known as metabolic switching. Further, reduction in hepatic clearance may lead to increased renal or biliary clearance of a parent drug or inhibition of one or more drug-metabolizing enzymes. Therefore, it is advisable that *in vitro* metabolic stability data be integrated with other ADME screening.

## REFERENCES


Ariens, E.J. *et al.* (1982). *Strategy in Drug Research*, Elsevier Scientific Publishing Company, Amsterdam, pp. 165–178.

Bakke, O.M. *et al.* (1995). Drug safety discontinuations in the United Kingdom, the United States and Spain from 1974 through 1993: a regulatory perspective. *Clin. Pharmacol. Ther.*, *58* (1), 108–117.

Blanchard, S.G. *et al.* (1998). Discovery of bioavailable inhibitors of secretory phospholipase A2. *Pharm. Biotechnol.*, *11*, 445–463.

Bouska, J.J. *et al.* (1997). Improving the in vivo duration of 5-lipoxygenase inhibitors: application of an in vitro glucuronosyltransferase assay. *Drug Metab. Dispos.* *25* (9), 1032–1038.

Dragovich, P. *et al.* (2003). Structure-based design, synthesis, and biological evaluation of irreversible human rhinovirus 3C protease inhibitors. 8. Pharmacological optimization of orally bioavailable 2-pyridone-containing peptidomimetics. *J. Med. Chem.*, *46* (21), 4572–4585.

Genin, M. *et al.* (1996). Synthesis and bioactivity of novel Bis(heteroaryl)piperazine reverse transcriptase inhibitors: structure-activity relationships and increased metabolic stability of novel substituted pyridine analogs. *J. Med. Chem.*, *39* (26), 5267–5275.

Greene, N. *et al.* (2002). Computer systems for the prediction of toxicity: an update. *Adv. Drug Deliv. Rev.*, *54* (3), 417–431.

Humphrey, M.J., Smith, D.A. (1992). Role of metabolism and pharmacokinetic studies in the discovery of new drugs—present and future perspectives. *Xenobiotica*, *22* (7), 743–755.

Kantharaj, E. *et al.* (2003). Simultaneous measurement of drug metabolic stability and identification of metabolites using ion-trap mass spectrometry. *Rapid Commun. Mass Spectrom.*, *17* (23), 2661–2668.

Kempf, D. *et al.* (1998). Discovery of ritonavir, a potent inhibitor of HIV protease with high oral bioavailability and clinical efficacy. *J. Med. Chem.*, *41* (4), 602–617.

Kostiainen, R. *et al.* (2003). Liquid chromatography/atmospheric pressure ionization-mass spectrometry in drug metabolism studies. *J. Mass Spectrom.*, *38* (4), 357–372.

Kubinyi, H. *et al.* (1977). Quantitative structure-activity relationships. 7. The bilinear model, a new model for nonlinear dependence of biological activity on hydrophobic character. *J. Med. Chem.*, *20*, 625–629.

Lasser, K.E. *et al.* (2002). Timing of new black box warnings and withdrawals for prescription medications. *J. Am. Med. Assoc.*, *287* (17), 2215–2220.

Lesko, L.L. *et al.* (2000). Optimizing the science of drug development: opportunities for better candidate selection and accelerated evaluation in humans. *Pharm. Res.*, *14*, 1335–1343.

Lin, N.H. *et al.* (1997). Structure-activity studies on 2-methyl-3-(2(S)-pyrrolidinylmethoxy) pyridine (ABT-089): an orally bioavailable 3-pyridyl ether nicotinic acetylcholine receptor ligand with cognition-enhancing properties. *J. Med. Chem.*, *40* (3), 385–390.

Meyboom, R.H.B. *et al.* (2000). An ABC of drug-related problems. *Drug Saf.*, *22* (6), 415–423.

Nassar, A.-E.F. *et al.* (2003a). Metabolite characterization in drug discovery utilizing robotic liquid-handling, quadrupole time-of-flight mass spectrometry and *in-silico* prediction. *Curr. Drug Metab.*, *4*, 259–271.

Nassar, A.-E.F. *et al.* (2003b). Online hydrogen-deuterium exchange and a tandem quadrupole time-of-flight mass spectrometer coupled with liquid chromatography for metabolite identification in drug metabolism. *J. Chromatog. Sci.*, *41*, 398–404.

Nassar, A.F. *et al.* (2004a). Strategies for dealing with reactive intermediates in drug discovery and development. *Curr. Opin. Drug Discov. Devel.*, *7*, 126–136.

Nassar, A.-E.F. *et al.* (2004b). Strategies for dealing with metabolite elucidation in drug discovery and development. *Drug Discov. Today*, *9*, 317–327.

Nassar, A.E., Kamel, A.M., Clarimont, C. (2004). Improving the decision-making process in the structural modification of drug candidates: enhancing metabolic stability, *Drug Discov. Today*, *9*, (23), 1020–1028.

Obach, R.S. *et al.* (1999a). Prediction of human clearance of twenty-nine drugs from hepatic microsomal intrinsic clearance data: an examination of in vitro half-life approach and nonspecific binding to microsomes. *Drug Metab. Dispos.*, *27* (11), 1350–1359.

Obach, R.S. *et al.* (1999b). Nonspecific binding to microsomes: impact on scale-up of in vitro intrinsic clearance to hepatic clearance as assessed through examination of warfarin, imipramine, and propranolol. *Drug Metab. Dispos.*, *25* (12), 1359–1369.

Palani, A. *et al.* (2002). Synthesis, SAR, and biological evaluation of oximino-piperidino-piperidine amides. 1. Orally bioavailable CCR5 receptor antagonists with potent anti-HIV activity. *J. Med. Chem.*, *45* (14), 3143–3160.

Pirmohamed, M. *et al.* (1998). Adverse drug reactions. *Br. Med. J.*, *316* (7140), 1295–1298.

Plumb, R.S. *et al.* (2003). Use of liquid chromatography/time-of-flight mass spectrometry and multivariate statistical analysis shows promise for the detection of drug metabolites in biological fluids. *Rapid Commun. Mass Spectrom.*, *17* (23), 2632–2638.

Rane, A. *et al.* (1977). Prediction of hepatic extraction ratio from in vitro measurement of intrinsic clearance. *J. Pharmacol. Exp. Ther.*, *200* (2), 420–424.

Richard, A.M. *et al.* (1998). Structure-based methods for predicting mutagenicity and carcinogenicity: are we there yet?. *Mutat. Res.*, *400* (1–2), 493–507.

Segel, I.H. *et al.* (1993). Chapter two: kinetics of unireactant enzymes, in *Enzyme Kinetics*, John Wiley & Sons, Inc., New York, pp. 18–89.

Tagat, J.R. *et al.* (2001). Piperazine-based CCR5 antagonists as HIV-1 inhibitors. II. Discovery of 1-[(2,4-dimethyl-3-pyridinyl)carbonyl]-4- methyl-4-[3(S)-methyl-4-[1(S)-[4-(trifluoromethyl)p henyl]ethyl]-1-piperazinyl]- piperidine N1-oxide (Sch-350634), an orally bioavailable, potent CCR5 antagonist. *J. Med. Chem.*, *44* (21), 3343–3346.

Taylor, E.W. *et al.* (2002). Accelerating the drug optimization process: identification, structure elucidation, and quantification of in vivo metabolites using stable isotopes with LC/MSn and the chemiluminescent nitrogen detector. *Anal. Chem.*, *74* (13), 3232–3238.

Testa, B. *et al.* (1990). Drug metabolism and pharmacokinetics: implications for drug design. *Acta Pharm. Jugosl. 40* (3), 315–350.

Tiller, P.R. *et al.* (2002). Liquid chromatography/tandem mass spectrometric quantification with metabolite screening as a strategy to enhance the early drug discovery process. *Rapid Commun. Mass Spectrom.*, *16*, 1225–1231.

Uetrecht, J.P. *et al.* (2002). Screening for the potential of a drug candidate to cause idiosyncratic drug reactions. *Drug Discov. Today*, *5*, 832–837.

Victor, F. *et al.* (1997). Synthesis, antiviral activity, and biological properties of vinylacetylene analogs of enviroxime. *J. Med. Chem.*, *40* (10), 1511–1518.

Watt, A.P. (2003). Metabolite identification in drug discovery. *Curr. Opin. Drug Discov. Devel.*, *6* (1), 57–65.

Zhuang, Z.P. *et al.* (1998). Isoindol-1-one analogues of 4-(2′-methoxyphenyl)-1-[2′-[N-(2″-pyridyl)-p-iodobenzamido]ethyl]pipera zine (p-MPPI) as 5-HT1A receptor ligands. *J. Med. Chem.*, *41* (2), 157–166.

# Structural Modifications of Drug Candidates: How Useful Are They in Improving PK Parameters of New Drugs? Part II: Drug Design Strategies

ALA F. NASSAR

## 1 INTRODUCTION

One key to successful drug design and development is the process of finding the right combination of diverse properties such as activity, toxicity, exposure, and so on. It is very important to first determine, and then optimize, the exposure–activity–toxicity relationships for drug candidates, and thus their suitability for advancement to development. Therefore, the concern of the drug metabolism scientist is to optimize plasma half-life, drug/metabolic clearance, metabolic stability, and the ratio of metabolic to renal clearance. The effort to achieve these positives must be balanced by the need to minimize or eliminate potential difficulties or dangers such as gut/hepatic-first-pass metabolism, inhibition/induction of drug-metabolizing enzymes by metabolites, biologically active metabolites, metabolism by polymorphically expressed drug-metabolizing enzymes, and formation of reactive metabolites. The optimal result will be a safer drug that undergoes predictable metabolic inactivation or even undergoes no metabolism. Among the approaches available to the drug design team, as they seek to meet these goals, are active metabolites, prodrugs, and hard and soft drugs; these will be discussed with some examples from recent case studies.

## 2 ACTIVE METABOLITES

Some drugs produce metabolites that are pharmacologically active; in some cases, these metabolites surpass their parent drugs as candidates for development. Active metabolites are often indicated by an elevated level of pharmacological effect, which is revealed through pharmacokinetic tests on the parent drug candidate. These metabolites often are subject to phase II metabolism and thus have better safety profiles. Table 1 contains

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

**TABLE 1 Metabolism-driven optimization of PK properties by structure modification.**

| Category | Pre-optimized compound |
|---|---|
| Active metabolites | 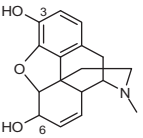 Morphine AUC 0–12 = 614 ng·h/mL |
| | 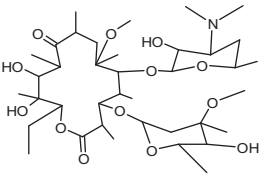 Clarithromycin $t_{1/2}$ = 3.7 h day 1, $t_{1/2}$ = 5.1 h day 7 |
| Prodrug | 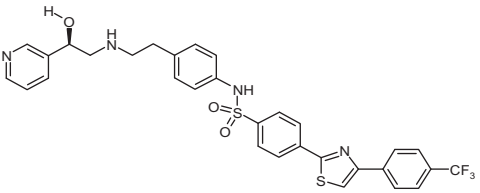 %F = 17 (rats), %F = 4 (monkeys) |
| | 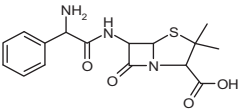 Ampicillin, mean peak serum level = 3.7 μg/mL, %F = 47–49 |
| First-pass effect | 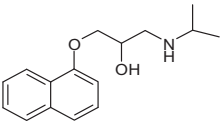 Propranolol AUC 0–6 = 132 ng/mL·h |
| | 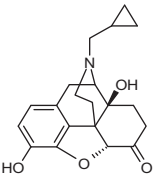 Naltrexone %F = 1.1 |

| Optimized compound | Optimized compound | Reference |
|---|---|---|
| 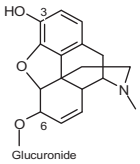 <br>Morphine-6-glucuronide, AUC 0–12 = 1672 ng·h/mL | | Koren *et al.* (2004) |
| 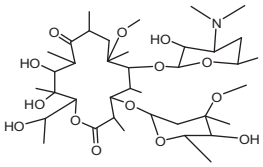 <br>14-hydroxy-clarithromycin $t_{1/2}$ = 6.2 h day 1, $t_{1/2}$ = 10.1 h day 7 | | Lode *et al.* (2002) |
| 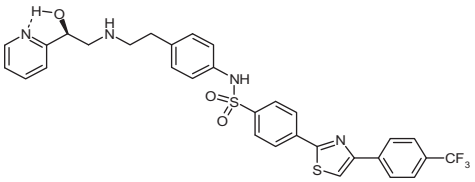 <br>%F = 30 (rats), %F = 23 (monkeys) | | Stearns *et al.* (2002) |
| 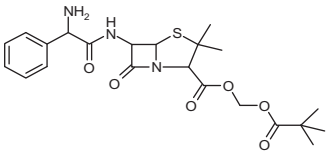 <br>Pivampicillin, mean, peak serum level = 7.1 µg/mL, %F = 82–89 | 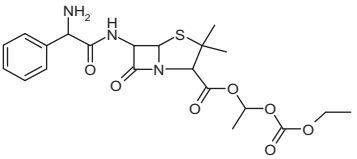 <br>Bacampicillin, mean, peak serum level = 8.3 µg/mL, %F = 87–95 | Sjövall, Magni and Bergan (1978) |
| 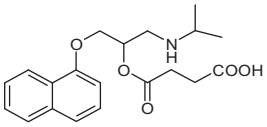 <br>Hemisuccinate ester of propranolol AUC 0–6 = 1075 ng/mL·h | | Hasegawa *et al.* (1978) |
| 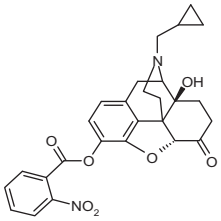 <br>Anthranilate %F = 49.2 | 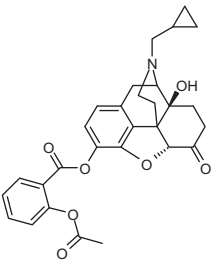 <br>Acetylsalicylate %F = 31.0 | Hussain *et al.* (1987) |

**TABLE 1** ***Continued***

| Category | Pre-optimized compound |
|---|---|
| | 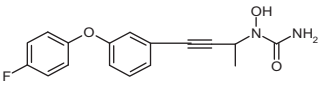 5-LO inhibitors $t_{1/2}$ = 5.5 h |
| Half-life | 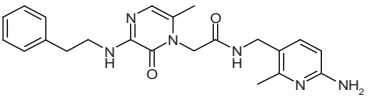 Thrombin $t_{1/2}$ = 3.8 h |
| | 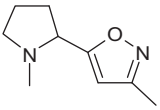 5 pyrrolidine ABT-418 $t_{1/2}$ = 0.2 h |
| PO and IV | 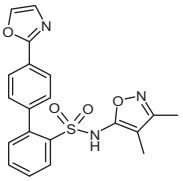 BMS 187308, AUC 0–inf. (μM·h) = 343 (iv) and 166 (po) at 100 μmol/kg |

the examples of 14-hydroxy-clarithromycin and morphine 6-glucuronide; they are more potent and less toxic than the parent drugs clarithromycin and morphine, respectively (Dabernat *et al.*, 1991; Bergeron, Bernier and L'Ecuyer, 1992; Burkhardt *et al.*, 2002; Kopecky *et al.*, 2004). Phase I metabolism of clarithromycin produces 14-hydroxy-clarithromycin. Clarithromycin is an important antibacterial drug in the treatment of community-acquired respiratory tract infections. In a double-blind, randomized, two-period crossover study, the pharmacokinetics properties of clarithromycin were determined after single and multiple doses in 12 healthy male volunteers. On day 1, the half-life was 3.7 hours for clarithromycin and 6.2 hours for its metabolite 14-hydroxy-clarithromycin. On day 7, the half-life was longer than on day 1 (5.1 hours for clarithromycin and 10.1 hours for 14-hydroxy-clarithromycin). Also, the *in vitro* activity of clarithromycin and its main metabolite 14-hydroxy clarithromycin against *Haemophilus influenzae* was evaluated. The 14-hydroxy metabolite was more active than the parent compound against *Haemophilus influenzae*. Microdilution broth minimum inhibitory concentrations (MICs) and minimum bactericidal concentrations (MBCs) of both clarithromycin and 14-hydroxy-clarithromycin were determined. For clarithromycin, the MIC50 was 4 mg/L and the MIC90 was 8 mg/L, while the hydroxy metabolite was two- to fourfold more active with an MIC50 and MIC90 of 2 mg/L. As a second example of working with an active metabolite, morphine 6-glucuronide is formed by phase II metabolism of morphine. Morphine 6-glucuronide is more potent than morphine itself without side effects. Morphine-6-glucuronide is a major metabolite of morphine with potent analgesic actions. The PK of morphine and its active metabolite

| Optimized compound | Optimized compound | Reference |
|---|---|---|
| 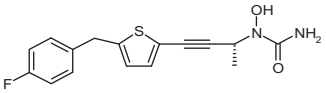 $t_{1/2}$ = 16 h | | Bouska *et al.* (1997) |
| 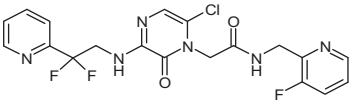 $t_{1/2}$ = 6.6 h | 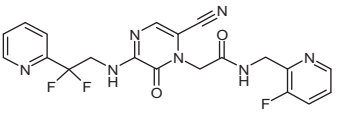 $t_{1/2}$ = 9.7 h | Burgey *et al.* (2003) |
| 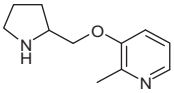 ABT-089 $t_{1/2}$ = 1.6 h | | Lin *et al.* (1997) |
| 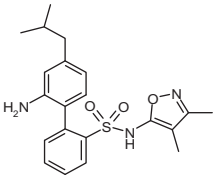 BMS 193884, AUC 0–inf. (μM·h) = 840 (iv) at 100 μmol/kg | | Humphreys *et al.* (2003) |

morphine-6-glucuronide were studied in humans, giving $AUC_{0-12h}$ results of 614.3 and 1672.3 ng·h/mL for morphine and morphine 6-glucuronide, respectively. Thus, it becomes clear that morphine 6-glucuronide is a highly potent metabolite, and there is strong evidence that it is important to the overall effect of morphine.

## 3 PRODRUGS

The prodrug approach has been widely used in drug design. Prodrugs were discovered when it was demonstrated that the antibacterial agent protosil was active *in vivo* only when it was metabolized to the actual drug sulfanilamide. Prodrugs most commonly use either oxidative or reductive activation; for example, protosil is activated by reduction of its azo linkage to the amine sulfa drug. Another example of a prodrug is the antipyretic agent phenacetin, which provides its activity after conversion to acetaminophen. Carbamazepine is an anticonvulsant drug that is the metabolic precursor of the active agent carbamazepine 10,11-oxide. Minoxidil is a potent vasodilator that also induces hypertrichosis of facial and body hair (Buhl *et al.*, 1990). Research data show that sulfation is a critical step for the hair growth effects of minoxidil and that it is the sulfated metabolite that directly stimulates hair follicles.

As an example of the effort to improve PK properties through structure modification, the oral bioavailability of a 3-pyridyl thiazole benzenesulfonamide adrenergic receptor agonist (Table 1) was investigated in rats, dogs, and monkeys (Stearns *et al.*, 2002). It was shown that the limited bioavailability of this compound was due to

poor oral absorption in rats and monkeys (%F = 17 and 4 for rats and monkeys, respectively). A slight modification to the structure of this compound resulted in a significant improvement to its oral bioavailability. The linkage to the pyridine moiety was changed from the 3- to the 2-position so that the pyridyl-nitrogen atom was positioned to the hydrogen bond with the ethanolamine hydroxyl group; this minimized intermolecular interactions that may limit the oral absorption of this compound class. Bioavailability in rats and monkeys improved to %F = 30 and 23 for rats and monkeys, respectively.

The PK evaluation of various structural modifications of ampicillin is another example of how the prodrug strategy can improve PK properties (see Table 1). Pivampicillin, talampicillin, and bacampicillin are prodrugs of ampicillin, each of which is produced from the esterification of the polar carboxylate group to form lipophilic, enzymatically labile esters (Sjövall, Magni and Bergan, 1978; Ehrnebo, Nilsson and Boreus, 1979; Ensink *et al.*, 1996). These prodrugs result in nearly total absorption, whereas that of ampicillin is <50%. A randomized crossover study on 11 healthy volunteers to determine the pharmacokinetics of bacampicillin, ampicillin, and pivampicillin showed that the relative bioavailability of bacampicillin (%F = 87–95) and pivampicillin (%F = 82–89) was comparable, whereas ampicillin was less than two-thirds that of the others (%F = 47–49). Additionally, the mean of the individual peak concentrations in serum was 8.3 μg/mL for bacampicillin, 7.1 μg/mL for pivampicillin, and 3.7 μg/mL for ampicillin.

## 4 HARD AND SOFT DRUGS

Hard drugs have been defined as drugs that are biologically active and non-metabolizable *in vivo*. Soft drugs are defined as those that produce predictable and controllable *in vivo* metabolism to form nontoxic products after they have achieved their therapeutic role. For example, quaternary ammonium compounds, such as benzalkonium chloride, are hard antibacterial agents. They are toxic to humans and animals, and their chemical stability (non-metabolizable) makes them unsuitable for general environmental sanitation. Soft analogs of such hard antibacterial agents have been developed, and are less toxic. Although the soft analogs have been shown to possess antibacterial activity in *in vitro* studies, a problem for researchers is that it is likely that their *in vivo* activity will be hampered by their chemical instability. When considering toxicology and pharmacology, a safer drug is one that undergoes predictable metabolic inactivation or even undergoes no metabolism (Schuster *et al.*, 1999; Fromigue, Lagneaux and Body, 2000; Bos *et al.*, 2002).

### 4.1 Hard Drugs

The hard drug approach is very attractive to drug design teams because the problem of toxicity due to reactive or active metabolites is eliminated, and the pharmacokinetics also are simplified because the drugs are excreted primarily through either the bile or kidney. When excretion of a drug occurs mainly through the kidney, differences in its elimination between humans and animals will depend mainly on the renal function, which is a strong predictor of PK profiles. Examples of hard drugs include biophosphonates and certain ACE inhibitors. Because they are not metabolized in animals or humans, and are only eliminated through renal excretion, biophosphonates are very safe with no significant systemic toxicity. Similarly, ACE inhibitors undergo very

limited metabolism and are exclusively excreted by the kidney. It is worth noting that biophosphonates and carboxyalkyldipeptide ACE inhibitors were not intentionally designed as hard drugs; rather, their hard qualities are the result of structural improvement.

### 4.2 Soft Drugs

Soft drugs are those that produce predictable *in vivo* metabolism which forms nontoxic products after achieving their therapeutic role. The soft drug approach also aims to produce safer drugs with an increased therapeutic index by integrating metabolism considerations into the drug design process (Bodor and Kaminski, 1980; Bodor, 1984; Bodor and Buchwald, 2000, 2004; Huang *et al.*, 2003). Because most oxidative reactions of drugs are produced by hepatic cytochrome P450 enzyme systems that are often affected by age, sex, disease, and environmental factors, biotransformation becomes more complex and PK is variable. P450 oxidative reactions have the potential to form reactive intermediates and active metabolites which can produce toxicity. The distribution and clearance of such metabolites may be different from those of the parent drug. The design of soft drugs aims to avoid oxidative metabolism as much as possible and to use hydrolytic enzymes to achieve predictability and control of drug metabolism. There are a number of drugs currently on the market, such as esmolol, remifentanil, or loteprednol etabonate, which are products of the soft drug design approach. Other promising drug candidates are currently under investigation in a variety of fields including possible soft antimicrobials, anticholinergics, corticosteroids, beta-blockers, analgetics, ACE inhibitors, antiarrhythmics, and others. An example of this concept in drug design is the isosteric soft analog of the hard antifungal drug cetylpyridinium chloride. While cetylpyridinium is quite effective against both gram-positive and gram-negative bacteria, it is also quite toxic ($LD_{50}$ = 108 mg/kg). The soft analog has a metabolically soft spot (the ester group) built into the structure to replace the ethylene group for detoxification as well as 3-methyl imidazole derivative to replace the pyridine moiety. The soft nature of the new antimicrobial is low toxicity with $LD_{50}$ = 4110 mg/kg.

## 5 PK PARAMETERS

To increase the chances of success for a drug's development, it is essential that a drug candidate has good bioavailability and a desirable half-life, while avoiding problems such as poor absorption and extensive first-pass metabolism. Therefore, an accurate measurement of the pharmacokinetic parameters and a good understanding of the factors that affect the pharmacokinetics will guide drug design. In the following sections, the strategies for structural modification of drug candidates are discussed to improve PK properties such as first-pass effect, half-life, and intravenous versus oral absorption.

### 5.1 First-Pass Effect/Prodrug

Table 1 shows three examples of how the first-pass effect can be minimized through structure modification. The first example, propranolol, shows the potential usefulness of the prodrug approach when a highly metabolized drug is protected from first-pass

elimination in order to improve bioavailability. Oral dosage of propranolol produces a low bioavailability and a wide variation from patient to patient when compared with intravenous administration; this difference is attributed to first-pass elimination of the drug. Hemisuccinate ester of propranolol was selected as a potential prodrug with the hypothesis that propranolol hemisuccinate ester administration would avoid glucuronide formation during absorption and subsequently be released in the blood by hydrolysis. Following 80 mg po of propranolol hydrochloride and an equivalent dose of its hemisuccinate ester derivative, the $AUC_{0-6h}$ was 1075 and 132 ng/mL·h for hemisuccinate ester and propranolol, respectively (Garceau, Davis and Hasegawa, 1978; Shameem, Imai and Otagiri, 1993; Imai *et al.*, 2003).

A similar approach was successfully employed to enhance the bioavailability of naltrexone [17-(cyclopropylmethyl)-4,5 alpha-epoxy-3,14-dihydroxymorphinan-6-one]. Naltrexone is used in the treatment of opioid addiction and absorbs well from the gastrointestinal tract. However, its systemic bioavailability is very poor, as it undergoes extensive first-pass metabolism when administered orally. Employing the prodrug strategy, a number of prodrug esters on the 3-hydroxyl group were prepared: the anthranilate, acetylsalicylate, benzoate, and pivalate. The oral bioavailability of these prodrugs was determined in dogs. The anthranilate and acetylsalicylate compounds exhibited the greatest enhancement of naltrexone bioavailability (45 and 28 times greater than naltrexone, respectively). No correlation was found between the rates of plasma hydrolysis and bioavailability. Naltrexone-3-acetylsalicylate hydrolyzed in human and dog plasma, with a fast deacetylation step to naltrexone salicylate followed by a slower hydrolysis step to naltrexone (Hussain *et al.*, 1987; Reuning *et al.*, 1989).

An extensive body of work on *N*-hydroxyurea inhibitors of 5-lipoxygenase related to zileuton provides an excellent opportunity to examine structure-activity relationships (SARs) with respect to conjugation with glucuronic acid (Bouska *et al.*, 1997). In zileuton, the *N*-hydroxyurea moiety was identified as an optimal pharmacophore for potency and selectivity; therefore, it was not modified in the search for metabolically stable compounds. Medicinal chemistry efforts to optimize *in vivo* duration in the cynomolgus monkey were focused on the template and link components. In general, compounds with acetylene linker groups were found to have lower uridine 5-diphosphoglucuronic acid (UDPGA) rates than any of the links tested. Biaryl templates such as phenoxyphenyl, phenoxyfuran, and phenylthiophene also demonstrated reduced rates of glucuronidation and longer *in vivo* duration in monkeys. Finally, the importance of the stereochemistry of the carbon center immediately adjacent to *N*-hydroxyurea was investigated. It had already been shown that *N*-hydroxyurea compounds containing a methyl branch on the linking group had longer *in vivo* durations than their methylene analogs, which led to additional research on this methyl branch series. By adding the methyl group, a chiral center was created, and resulted in a compound with separable *R*- and *S*-enantiomers. The glucuronidation rate showed an approximately twofold difference, depending on the configuration of the adjacent carbon center. The result was the identification of a clinical candidate, ABT-761, which demonstrated the longest *in vivo* duration ($t_{1/2}$ = 16 hours) compared with that of zileuton ($t_{1/2}$ = 0.4 hours).

### 5.2 Half-life

Most drugs are administered as a fixed dose given at regular intervals to achieve therapeutic objectives. The plasma half-life is a measure of the duration of drug action. Thus,

the half-life of drugs in plasma is one of the most important factors that determine the selection of a dosage regimen. Shorter half-life means more frequent dosing, which tends to reduce the likelihood that the patient will maintain the regimen. The two factors affecting half-life are volume of distribution and elimination clearance. An increase in the first or a reduction of the second will improve the drug's half-life. Modification of the chemical structure is one viable approach to slowing a drug's clearance.

A metabolism-based approach to the optimization of the 3-(2-phenethylamino)-6-methylpyrazinone acetamide template for thrombin is an example of these efforts, which resulted in the identification of several potent thrombin inhibitors with high levels of oral bioavailability and long plasma half-lives. Thrombin is vital to hemostasis because it mediates conversion of fibrinogen to fibrin and activates platelets. By developing the aminopyrazinone acetamide thrombin-inhibitor template, an efficacious, orally bioavailable pre-optimized compound with plasma half-life of 3.8 hours was identified. This compound was found to metabolize through oxidation at the benzylic positions and phase II conjugation of the amino group. These structural areas of metabolism represented major means of clearance of the drug. Their elimination enabled generation of compounds with improved pharmacokinetics. Both the chloro ($t_{1/2}$, 6.6 hours) and cyano ($t_{1/2}$, 9.7 hours) modifications imparted significantly improved half-lives versus the pre-optimized compound (Burgey *et al.*, 2003).

As a second example, structural modification of ABT-418 was required to achieve further improvements in the margin of safety and to identify compounds with favorable oral bioavailability. ABT-418, an analog of (*S*)-nicotine in which the pyridine ring is replaced by the 3-methyl-5-isoxazole moiety, has been shown to possess cognitive-enhancing and anxiolytic-like activities in animal models with an improved safety profile compared with that of nicotine (Lin *et al.*, 1997). One shortcoming of ABT-418 was its very poor bioavailability (%F = 1.2), with a short plasma half-life ($t_{1/2}$ = 0.21 hour). Research on structural modification led to the identification of ABT-089, 2-methyl-3-(2(*S*)-pyrrolidinylmethoxy)pyridine, with a vastly improved oral bioavailability (%F = 61.5) with $t_{1/2}$ = 1.6 hours.

### 5.3 Oral Absorption and Intravenous Dose

The oral bioavailability of a drug is defined as the fraction of an oral dose of the drug that reaches the systemic circulation. An important consideration for orally administered drugs is that the entire blood supply of the upper gastrointestinal tract passes through the liver before reaching the systematic circulation. This means that the drug may be metabolized by the gut wall and/or the liver during the first passage of drug absorption. A drug with high metabolic clearance is always subject to an extensive first-pass effect, thus reducing its bioavailability. By contrast, administering a drug intravenously ensures that all of it enters the blood. Rapid injection will promptly achieve elevated concentrations of the drug in the blood; infusion at a controlled rate produces a constant concentration which can be maintained for any desired length of time. The lipophilicity of a drug is important to its membrane permeability as well as metabolic activity. Higher lipophilicity for a drug generally results in higher permeability and greater clearance, and thereby higher first-pass metabolism.

Bristol-Myers Squibb Pharmaceutical Research Group (Humphrey and Smith, 1992; Humphreys *et al.*, 2003) present a good example of working with structure–metabolism relationships to produce favorable pharmacokinetic properties. They identified an

initial lead (2′-amino-*N*-(3,4-dimethyl-5-isoxazolyl)-40-(2-methylpropyl)[1,10-biphenyl]-2-sulfonamide) for endothelin (ET) receptor antagonists. However, testing in preclinical animal species and human *in vitro* systems revealed the compound was extensively metabolized due to oxidative biotransformation. The site of metabolism of the candidate was determined and structure–activity and structure–metabolism studies were aimed at optimizing this structural class by finding more metabolically stable analogs that maintained potency. This resulted in the identification of an analog (*N*-(3,4-dimethyl-5-isoxazolyl)-40-(2-oxazolyl)[1,10-biphenyl]-2-sulfonamide) with improved *in vitro* properties; further studies revealed that the new compound had improved pharmacokinetic properties (Table 1).

## 6 PK ANALYSIS

Pharmacokinetic analysis is important to drug discovery as it provides insight into how new drug candidates may be absorbed and excreted by the human body (Lesko *et al.*, 2000; Bapiro *et al.*, 2001; Nassar and Adams, 2003; Nassar, Kamel and Clarimont, 2004a, b). Pharmacokinetic studies in early drug discovery involve dosing an animal with the compound of interest and measuring the drug concentration in biological fluids (e.g. plasma) as a function of time. This yields valuable information such as the time required to reach peak concentration and the half-life of the drug. Other parameters such as volume of distribution, clearance, and bioavailability can also be determined from these experiments. This important information is valuable to the decision-making process because it gives medicinal chemists targets for optimization. With many candidates to screen, pharmacokinetic studies are routine but demand a significant amount of time and resources. Sample preparation is a highly important aspect of these experiments, and robotic-based automation is commonly used. The biological fluid is then analyzed by liquid chromatography–mass spectrometry (LC-MS) directly after online or off-line solid-phase extraction, liquid–liquid extraction, or protein precipitation. By reducing the number of pharmacokinetic samples through pooling at certain time points or across all time points, and by the use of an abbreviated standard curve, the analysis is streamlined. The new technology of ultra-performance liquid chromatography (UPLC) can offer significant advantages in resolution, speed, and sensitivity for analytical determinations, which are all important factors for the demands of high-throughput screening in drug discovery. UPLC is more beneficial when coupled with mass spectrometers capable of high-speed acquisitions. This approach allows attainment of more information about a lead candidate while maintaining a rapid analytical turnaround time. Chromatographic separations with analysis times of around 15 seconds can be envisioned; however, such separations will require high-speed MS systems. Also, by using micro and capillary separation techniques with nanospray MS, significant additional sensitivity can be gained.

## 7 CONCLUSIONS

The clinical success of a drug candidate depends upon desirable ADME/TOX properties. An accurate measurement of the pharmacokinetic parameters and a good

understanding of the factors that affect the pharmacokinetics will help to guide drug design. By seeking a balance based upon the rule of three (exposure–activity–toxicity), and taking pharmaceutical research beyond the initial candidate or parent drug, drug metabolism scientists have opened up a whole new front in the effort to develop safer and more efficacious drugs. New approaches, evolving technologies, and the wisdom gained from examination of previous difficulties and failures have aided the drug development team to focus their efforts. It is the opinion of the author that improvement in the drug discovery process will be facilitated most by learning how and when to use each of these tools and approaches to design drugs to achieve the maximum benefit, and by increasing cooperation and dialogue among the disciplines involved in the entire process. The author believes that the following suggestions should be considered when developing a new drug:

1. Considering the risks and costs of new drug development, it is critical to eliminate high-risk compounds as early as possible. Fast and reliable *in silico* screens are needed to filter out problematic molecules at the earliest stages of discovery.
2. The metabolism of new drugs should be studied *in vitro* before the initiation of any clinical studies. Early information on *in vitro* metabolic processes in humans, such as the identification of the enzymes responsible for drug metabolism and sources of potential enzyme polymorphism, can be useful in the design of clinical studies, particularly those that examine drug–drug interactions. High metabolic ability usually leads to poor bioavailability and high clearance, and formation of active or toxic metabolites will have an impact on the pharmacological and toxicological outcomes.
3. Appropriate structural modifications to the drug candidate help to optimize the metabolic ability, drug–drug interactions with coadministered drugs due to inhibition and/or induction of drug metabolism pathways, and the rate and sites of metabolism of new chemical entities by drug-metabolizing enzymes.
4. Good *in vitro* activity does not correlate to good *in vivo* activity unless a drug has a good bioavailability and a desirable duration of action. These qualities are essential to successful development.
5. To avoid exclusion of good compounds, the selection of animal species and the experimental design of studies are important in providing a reliable prediction of drug absorption and elimination in humans. A good compound could be excluded on the basis of results from an inappropriate animal species or poor experimental design.
6. Drug candidates should have little or none of the following: gut/hepatic-first-pass metabolism, inhibition/induction of drug-metabolizing enzymes, biologically active metabolites, metabolism by polymorphically expressed drug-metabolizing enzymes, and formation of reactive metabolites. Also, it is important to have the most desirable plasma half-life and ratio of metabolic to renal clearance.
7. One of the most important keys to successful drug design and development is finding the right combination of multiple properties such as activity, toxicity, and exposure. To design good drug candidates, the rule of three should be applied regardless of the therapeutic index.

## REFERENCES


Bapiro, T.E., Egnell, A.C., Hasler, J.A., Masimirembwa, C.M. (2001). Application of higher throughput screening (HTS) inhibition assays to evaluate the interaction of antiparasitic drugs with cytochrome P450s. *Drug Metab. Dispos.*, *29*, 30–35.

Bergeron, M.G., Bernier, M., L'Ecuyer, J. (1992). In vitro activity of clarithromycin and its 14-hydroxy-metabolite against 203 strains of *Haemophilus influenzae*. *Infection*, *20*, 164–167.

Bodor, N. (1984). Soft drugs: principles and methods for the design of safe drugs. *Med. Res. Rev.*, *4*, 449–469.

Bodor, N., Buchwald, P. (2000). Soft drug design: general principles and recent applications. *Med. Res. Rev.*, *20*, 58–101.

Bodor, N., Buchwald, P. (2004). Designing safer (soft) drugs by avoiding the formation of toxic and oxidative metabolites. *Mol. Biotechnol.*, *26*, 123–132.

Bodor, N., Kaminski, J.J. (1980). Soft drugs. 2. Soft alkylating compounds as potential antitumor agents. *J. Med. Chem.*, *23*, 566–569.

Bos, H., Henning, R.H., De Boer, E., Tiebosch, A.T., De Jong, P.E., De Zeeuw, D., Navis, G. (2002). Addition of AT1 blocker fails to overcome resistance to ACE inhibition in adriamycin nephrosis. *Kidney Int.*, *61*, 473–480.

Bouska, J.J., Bell, R.L., Goodfellow, C.L., Stewart, A.O., Brooks, C.D., Carter, G.W. (1997). Improving the in vivo duration of 5-lipoxygenase inhibitors: application of an in vitro glucuronosyltransferase assay. *Drug Metab. Dispos.*, *25*, 1032–1038.

Buhl, A.E., Waldon, D.J., Baker, C.A., Johnson, G.A. (1990). Minoxidil sulfate is the active metabolite that stimulates hair follicles. *J. Invest. Dermatol.*, *95*, 553–557.

Burgey, C.S., Robinson, K., Lyle, T.A., Sanderson, P.E.J., Lewis, S.D., Lucas, B.J., Krueger, J.A., Singh, R., Miller-Stein, C., White, R.B., Wong, B., Lyle, E.A., Williams, P.D., Coburn, C.A., Dorsey, B.D., Barrow, J.C., Stranieri, M.T.; Holahan, M.A., Sitko, G.R., Cook, J.J., McMasters, D.R., McDonough, C.M., Sanders, W.M., Wallace, A.A., Clayton, F.C., Bohn, D., Leonard, Y. M., Detwiler, T.J., Lynch, J.J., Yan, Y., Chen, Z., Kuo, L., Gardell, S.J., Shafer, J.A., Vacca, J.P. (2003). Metabolism-directed optimization of 3-aminopyrazinone acetamide thrombin inhibitors. Development of an orally bioavailable series containing P1 and P3 pyridines. *J. Med. Chem.*, *46*, 461–473.

Burkhardt, O., Borner, K., Stass, H., Beyer, G., Allewelt, M., Nord, C., Lode, H. (2002). Single- and multiple-dose pharmacokinetics of oral moxifloxacin and clarithromycin, and concentrations in serum, saliva and faeces. *Scand. J. Infect. Dis.*, *34*, 898–903.

Dabernat, H., Delmas, C., Seguy, M., Fourtillan, J.B., Girault, J., Lareng, M.B. (1991). The activity of clarithromycin and its 14-hydroxy metabolite against Haemophilus influenzae, determined by in-vitro and serum bactericidal tests. *J. Antimicrob. Chemother.*, *27*, 19–30.

Ehrnebo, M., Nilsson, S.O., Boreus, L.O. (1979). Pharmacokinetics of ampicillin and its prodrugs bacampicillin and pivampicillin in man. *J. Pharmacokinet. Biopharm.*, *7*, 429–451.

Ensink, J.M., Vulto, A.G., van Miert, A.S., Tukker, J.J., Winkel, M.B., Fluitman, M.A. (1996). Oral bioavailability and in vitro stability of pivampicillin, bacampicillin, talampicillin, and ampicillin in horses. *Am. J. Vet. Res.*, *57*, 1021–1024.

Fromigue, O., Lagneaux, L., Body, J.J. (2000). Bisphosphonates induce breast cancer cell death in vitro. *J. Bone Miner. Res.*, *15*, 2211–2221.

Garceau, Y., Davis, I., Hasegawa, J. (1978). Plasma propranolol levels in beagle dogs after administration of propranolol hemisuccinate ester. *J. Pharm. Sci.*, *67*, 1360–1363.

Huang, F., Browne, C.E., Wu, W.M., Juhasz, A., Ji, F., Bodor, N. (2003). Design, pharmacokinetic, and pharmacodynamic evaluation of a new class of soft anticholinergics. *Pharm. Res.*, *20*, 1681–1689.

Humphrey, M.J., Smith, D.A. (1992). Role of metabolism and pharmacokinetic studies in the discovery of new drugs—present and future perspectives. *Xenobiotica*, *22*, 743–755.

Humphreys, W.G., Obermeier, M.T., Barrish, J.C., Chong, S., Marino, A.M., Murugesan, N., Wang-Iverson, D., Morrison, R.A. (2003). Application of structure-metabolism relationships in the identification of a selective endothelin A antagonist, BMS-193884, with favourable pharmacokinetic properties. *Xenobiotica*, *33*, 1109–1123.

Hussain, M.A., Koval, C.A., Myers, M.J., Shami, E.G., Shefter, E. (1987). Improvement of the oral bioavailability of naltrexone in dogs: a prodrug approach. *J. Pharm. Sci.*, *76*, 356–358.

Imai, T., Yoshigae, Y., Hosokawa, M., Chiba, K., Otagiri, M. (2003). Evidence for the involvement of a pulmonary first-pass effect via carboxylesterase in the disposition of a propranolol ester derivative after intravenous administration. *J. Pharmacol. Exp. Ther.*, *307*, 1234–1242.

Kopecky, E.A., Jacobson, S., Joshi, P., Koren, G. (2004). Systemic exposure to morphine and the risk of acute chest syndrome in sickle cell disease. *Clin. Pharmacol. Ther.*, *75*, 140–146.

Lesko, L.J., Rowland, M., Peck, C.C., Blaschke, T.F. (2000). Optimizing the science of drug development: opportunities for better candidate selection and accelerated evaluation in humans. *Pharm. Res.*, *17*, 1335–1343.

Lin, N.H., Gunn, D.E., Ryther, K.B., Garvey, D.S., Donnelly-Roberts, D.L., Decker, M.W., Brioni, J.D., Buckley, M.J., Rodrigues, A.D., Marsh, K.G., Anderson, D.J., Buccafusco, J.J., Prendergast, M.A., Sullivan, J.P., Williams, M., Arneric, S. P., Holladay, M.W. (1997). Structure-activity studies on 2-methyl-3-(2(S)-pyrrolidinylmethoxy) pyridine (ABT-089): an orally bioavailable 3-pyridyl ether nicotinic acetylcholine receptor ligand with cognition-enhancing properties. *J. Med. Chem.*, *31*, 385–390.

Nassar, A.E., Adams, P.E. (2003). Metabolite characterization in drug discovery utilizing robotic liquid-handling, quadrupole time-of-flight mass spectrometry and *in silico* prediction. *Curr. Drug Metab.*, *4*, 259–271.

Nassar, A.E., Kamel, A.M., Clarimont, C. (2004a). Improving the decision-making process in structural modification of drug candidates: reducing toxicity. *Drug Discov. Today*, *24*, 1055–1064.

Nassar, A.E., Kamel, A.M., Clarimont, C. (2004b). Improving the decision-making process in the structural modification of drug candidates: enhancing metabolic stability. *Drug Discov. Today*, *23*, 1020–1028.

Reuning, R.H., Ashcraft, S.B., Wiley, J.N., Morrison, B.E. (1989). Disposition and pharmacokinetics of naltrexone after intravenous and oral administration in rhesus monkeys. *Drug Metab. Dispos.*, *17*, 583–589.

Schuster, C., Reinhart, W.H., Hartmann, K., Kuhn, M. (1999). Angioedema induced by ACE inhibitors and angiotensin II-receptor antagonists: analysis of 98 cases. *Schweiz. Med. Wochenschr.*, *129*, 362–369.

Shameem, M., Imai, T., Otagiri, M. (1993). An in-vitro and in-vivo correlative approach to the evaluation of ester prodrugs to improve oral delivery of propranolol. *J. Pharm. Pharmacol*, *45*, 246–452.

Sjövall, J., Magni, L., Bergan, T. (1978). Pharmacokinetics of bacampicillin compared with those of ampicillin, pivampicillin, and amoxycillin. *Antimicrob. Agents Chemother.*, *13*, 90–96.

Stearns, R.A., Miller, R.R., Tang, W., Kwei, G.Y., Tang, F.S., Mathvink, R.J., Naylor, E.M., Chitty, D., Colandrea, V.J., Weber, A.E., Colletti, A.E., Strauss, J.R., Keohane, C.A., Feeney, W.P., Iliff, S.A., Chiu, S.H. (2002). The pharmacokinetics of a thiazole benzenesulfonamide beta 3-adrenergic receptor agonist and its analogs in rats, dogs, and monkeys: improving oral bioavailability. *Drug Metab. Dispos.*, *30*, 771–777.

# Minimizing the Potential for Drug Bioactivation of Drug Candidates to Success in Clinical Development

ALA F. NASSAR

## 1 INTRODUCTION

Toxicity problems, especially those that may only occur under unusual or idiosyncratic conditions during the late stages of drug development, are one of the most devastating surprises for pharmaceutical companies. Variations in human drug-metabolizing enzymes can produce subtle evidence of potential toxicity, or none at all, during preclinical safety studies. Such problems are also unlikely to show up in all but the largest clinical trials, but if the side effects are serious, can result in product withdrawal. There are indications that some substructures found in drugs can form reactive metabolites that are involved in toxicities in humans. These substructures include arylacetic and arylpropionic acids, aryl hydroxamic acids, oximes, anilines, anilides, hydrazines, hydrazides, hydantoins, quinones, quinone methides, nitroaromatics, heteroaromatics, halogenated hydrocarbons, some halogenated aromatics, chemical groups that can be oxidized to acroleins, and medium-chain fatty acids. Reactive metabolites are unstable, and are intermediates to more stable metabolites. Table 1 shows several examples of drugs that undergo metabolic activation and cause adverse reaction in humans, which have been withdrawn from the market or restricted in use with toxicity warnings. Clearly, a drug candidate that does, or may, metabolize to such substructures would increase the risk of failure or withdrawal.

### 1.1 Importance of Reactive Intermediates in Drug Discovery and Development

One key to success in clinical development is to minimize reactive intermediates in drug metabolism. There is a growing consensus that idiosyncratic drug reactions (IDRs) have enormous consequences for patients and the pharmaceutical industry. It is estimated that IDRs account for ~5% of all hospital admissions and occur in 10% to 20%

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

**TABLE 1 Examples of chemical structures activating to produce toxic metabolites.**

| Chemical class | Bio-transformation | Toxic metabolite | Compound | | Biological effects | Ref |
|---|---|---|---|---|---|---|
| | | | Name | Clinical use | | |
| Quinone | Oxidation | Quinone-type | Tacrine | Alzheimer's disease | Hepatic toxicity | Madden *et al.* (1993) |
| | | | Troglitazone | Treat Type II diabetes | Hepatic toxicity | Kassahun *et al.* (2001; Tettey *et al.* 2001) |
| | | | Minocycline | Antibiotics | Hepatic toxicity<br>Lupus-like syndrome | Shapiro *et al.* (1997) |
| | | | Acetaminophen | Analgesic agent | Hepatic toxicity | Isaacson *et al.* (1999) |
| | | | Aminosalicylic acid | Inflammatory bowel disease | Lupus-like syndrome<br>Pancreatic toxicity<br>Hepatic toxicity<br>Renal toxicity | Liu *et al.* (1995) |
| | | | Amodiaquine | Treat malaria | Hepatic toxicity<br>Agranulocytosis | Ruscoe *et al.* (1995) |
| | | | Phenytoin | Anticonvulsant | Drug-induced hypersensitivity<br>Teratogenicity | Munns *et al.* (1997) |
| | | | Carbamazepine | Anticonvulsant | Teratogenicity | Ju *et al.* (1999) |
| | | | Vesnarinone | Phosphodiesterase inhibitor | Agranulocytosis | Uetrecht *et al.* (1994) |
| | | | Prinomide | Antiinflammatory | Agranulocytosis | Parrish *et al.* (1997) |
| | | | Estrogens | NSAID | Breast cancer<br>Uterine cancer | Mohsin *et al.* (2004) |
| | | | Tamoxifen | NSAID | Endometrial cancer | Fan *et al.* (2001) |
| | | | Fluperlapine | Antipsychotic agent | Agranulocytosis | Mann *et al.* (1987) |

| | | | | | | |
|---|---|---|---|---|---|---|
| Aryl nitro | Reduction | Nitroso | Tolcapone | Parkinson's disease | Liver toxicity | Borges *et al.* (2000) |
| | | | Chloramphenicol | Antibiotic | Aplastic anemia<br>Bone marrow toxicity | Yunis *et al.* (1980) |
| | | | Dantrolene | Muscle relaxant | Liver toxicity | Utili *et al.* (1977) |
| | | | Nimesulide | COX 2 inhibitors | Liver toxicity | Merlani *et al.* (2001) |
| Nitrogen-containing aromatic | Oxidation | Nitrenium ion<br>Free radical | Clozapine | Antipsychotic agent | Agranulocytosis<br>Liver toxicity<br>Myocarditis | Idanpaan-Heikkila *et al.* (1977; Haack *et al.* 2003) |
| | | | Aminopyrine | Painkiller | Agranulocytosis<br>CNS toxicity | Uetrecht *et al.* (1995) |
| | | | Dipyrone | Painkiller | Agranulocytosis | Sabbaga *et al.* (1993) |
| Aryl amines | Oxidation to hydroxylamine | Nitroso | Sulfamethoxazole | Antibacterial agent | Hepatotoxicity<br>Agranulocytosis<br>Lupus-like syndrome<br>Skin rashes | Cribb *et al.* (1992) |
| | | | Dapsone | Antiparasitic | Agranulocytosis<br>Flu-like syndrome<br>Hemolytic anemia<br>Methemoglobinemia | Coleman *et al.* (1994) |
| | | | Procainamide | Cardiac antiarrhythmic | Lupus-erythematosis<br>Agranulocytosis<br>Fever | Woosley *et al.* (1978) |
| | | | Nomifensine | Antidepressant | Hemolytic anemia<br>Allergic reactions | Salama *et al.* (1983) |
| | | | Sulfasalazine | Ulcerative colitis | Abnormal liver function<br>Decreased blood counts<br>Allergic reactions | Senturk *et al.* (1997) |
| | | | Aminoglutethimide | Breast cancer | Skin rashes<br>Fever<br>Agranulocytosis<br>Thrombocytopenia<br>Liver toxicity | Stuart-Harris *et al.* (1984) |

**TABLE 1** ***Continued***

| Chemical class | Bio-transformation | Toxic metabolite | Compound | | Biological effects | Ref |
|---|---|---|---|---|---|---|
| | | | Name | Clinical use | | |
| Michael Acceptors | Hydrolysis<br>Oxidation | Aldehyde<br>Co-A conjugate | Felbamate | Anticonvulsant | Aplastic anemia<br>Liver toxicity | Thompson *et al.* (1996; Dieckhaus *et al.* 2002) |
| | | | Terbinafine | Antifungal agent | Bone marrow toxicity<br>Liver toxicity<br>Skin rashes | Gupta *et al.* (1998) |
| | | | Valproic acid | Anticonvulsant | Liver toxicity | Grillo *et al.* (2001) |
| | | | Mianserin | Antidepressant | Agranulocytosis | Inman *et al.* (1988) |
| | | | Leflunomide | Inflammatory arthritis | Liver toxicity<br>Agranulocytosis | Jardine (2002) |
| Carboxylic acids | Glucuronidation | Acyl glucuronides | Diclofenac | NSAID | Liver toxicity<br>Agranulocytosis | Ware *et al.* (1998) |
| | | | Zomepirac | NSAID | Liver toxicity | Bailey *et al.* (1996) |
| | | | Ibufenac | NSAID | Liver toxicity | Prescott *et al.* (1986) |
| | | | Bromfenac | NSAID | Liver toxicity | (Skjodt *et al.*, 1999; Moses *et al.*, 1999) |
| | | | Benoxaprofen | NSAID | Liver toxicity | Halsey *et al.* (1982) |
| | | | Indomethacine | NSAID | Bone marrow toxicity | Godessart *et al.* (1999) |

Reprinted from Nassar, Kamel, and Clairmont, 2004, with permission from Elsevier.

of hospital inpatients (Bakke *et al.*, 1995; Pirmohamed *et al.*, 1998; Richard *et al.*, 1998; Meyboom *et al.*, 2000; Greene *et al.*, 2002; Lasser *et al.*, 2002; Tiller *et al.*, 2002; Uetrecht *et al.*, 2002). Of the new prescription drugs approved in the United States during the period between 1975 and 2000, 10.2% acquired a new "black box" warning or had to be withdrawn from the market because of IDRs (Nassar *et al.*, 2004b). Idiosyncratic drug toxicity is generally believed to be a phenomenon that cannot be readily evaluated experimentally because it is a rare event (usually <1 in 5000 cases), making clinical trials impractical since an extremely large patient population would be required. The current data suggest that IDRs are human-specific events that may not be detectable with experimental animals currently used. IDRs may have a complex pathology and can lead to a large number of symptoms, ranging from nonspecific rashes to specific organ damage, such as agranulocytosis or cholestasis. Immunological reactions can affect several organs including the liver, kidney, or lung, or they may have systemic effects. Hepatotoxicity is the most common type of idiosyncratic reaction. The time to onset of liver reactions varies from several days to almost one year, suggesting that the rare cases of liver injury may be caused by a metabolic idiosyncrasy (Pirmohamed *et al.*, 1998). Immune system responses may account for between 3% and 25% of all IDRs. For example, the antimalarial drug amodiaquine produces life-threatening agranulocytosis and hepatotoxicity in approximately 1 in 2000 patients during prophylactic administration. The unexpected occurrence of IDRs during late clinical trials has led to severe restrictions on use or failure to launch, and in the case of launched drugs, even withdrawal from the market. During clinical development, toxicity of drug candidates accounts for a significant portion of attrition (~20%) (Uetrecht *et al.*, 2002). Some examples of drugs withdrawn from the market due to IDRs are troglitazone, practolol, benoxaprofen, ticrynafen, zomepirac, and nomifensine. While all pharmaceutical companies consider IDRs to be a significant issue and are making efforts to predict them, the elusive nature and mechanisms of IDRs hinder the development of a clear and universal approach (Mehta *et al.*, 1982; Mattocks *et al.*, 1990; Li *et al.*, 2002; Matzinger *et al.*, 2002; Nassar *et al.*, 2004b).

This chapter discusses the recent efforts to overcome or assess the problem of reactive metabolites in drug discovery and development, which might help to evaluate the safety profile of new drug candidates for IDRs during the preclinical phase. Recently developed tools for identifying and determining the potential toxicity of reactive metabolites are discussed. The use of these tools is evaluated at several stages of drug discovery and development, along with their potential for improving drug safety. Also, some of the chemical substructures that form reactive metabolites and are involved in toxicities in humans are discussed.

## 2 IDIOSYNCRATIC DRUG TOXICITY AND MOLECULAR MECHANISMS

### 2.1 Adverse Drug Reactions

From a clinical perspective, adverse drug reactions (ADRs) can be classified as type A (predictable) and type B (unpredictable or idiosyncratic) reactions. Type A reactions, which account for ~80% of all ADRs, are dose-dependent and predictable based on the pharmacology of the drug. Type B, or IDRs, cannot be predicted on the basis of the pharmacology of the drug and lack simple dose dependency. Type B reactions, which include anaphylaxis, blood dyscrasias, hepatotoxicity, and skin reactions, may

result from a combination of genetic predisposition and environmental factors, and are reviewed elsewhere (Uetrecht *et al.*, 2002).

## 2.2 Types of Reactive Metabolites

There is significant evidence to indicate that some substructures found in drugs can form reactive metabolites that are involved in toxicities in humans. These substructures include arylacetic and arylpropionic acids, anilines, anilides, hydrazines, hydantoins, quinones, quinone methides, nitroaromatics, heteroaromatics, halogenated hydrocarbons, some halogenated aromatics, chemical groups that can be oxidized to acroleins, and medium-chain fatty acids. The thiazolidinedione ring, a relatively new substructure, can form reactive metabolites that may cause hepatotoxicity (Pirmohamed *et al.*, 1998). Reactive metabolites are unstable, and are intermediates to more stable metabolites. The different types of reactive metabolites and bioactivation have been thoroughly reviewed elsewhere (Pirmohamed *et al.*, 1998). For the purpose of this review, it is sufficient to categorize the three major types of metabolites as electrophiles, polarized double bonds, and free radicals. Previously published reports have provided significant evidence that chemically reactive metabolites may be responsible for serious forms of toxicity, including cellular necrosis, mutagenesis, carcinogenesis, teratogenesis, hypersensitivity, and blood disorders (Fig. 1).

## 2.3 Hypotheses for Idiosyncratic Reactions

Several hypotheses have been proposed to explain potential mechanisms for IDRs (Kassahun *et al.*, 1993; Tang *et al.*, 1996, 1997; Chen *et al.*, 2001; Burkhart *et al.*, 2002;

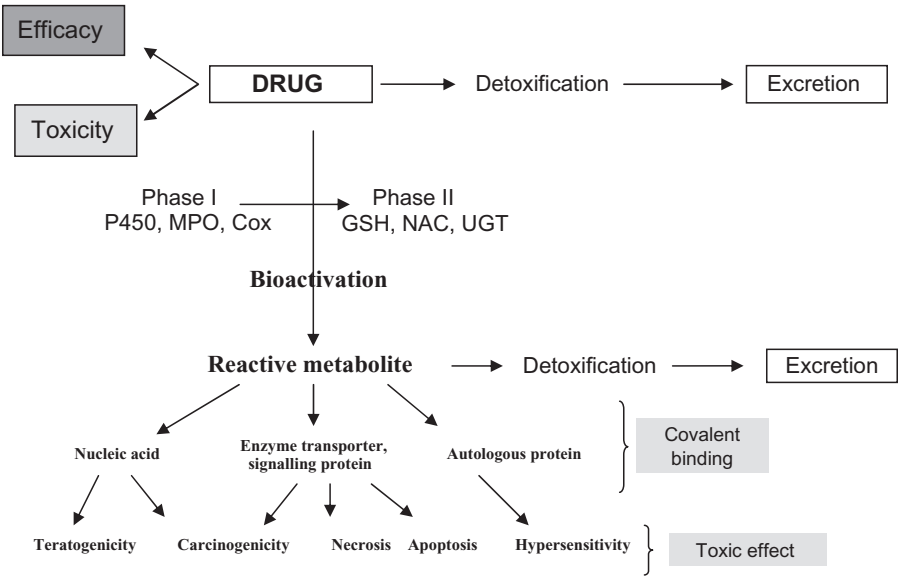


**Figure 1** Reactive intermediate theory and idiosyncratic reactions/toxic effects. Reprinted from Nassar and Lopez-Anaya, 2004, with permission from Thomson Reuters (Scientific) Limited. Copyright 2004, © The Thomson Corporation.

Judson *et al.*, 2003a, b; Steghens *et al.*, 2003; Evans, 2004; Soglia *et al.*, 2004). These include the hapten, the P-I, and the danger hypotheses. Alternatively, the theory of tolerance toward an immune reaction may explain why the majority of individuals are not susceptible to IDRs.

The hapten hypothesis proposes that modification of an endogenous protein by a reactive metabolite or directly by a reactive parent drug generates a "foreign" protein that, in some cases, leads to an immune-mediated adverse reaction. Immune-mediated mechanisms have been proposed for several drugs including halothane, phenytoins, sulfonamides, and tienilic acid. Failure to down-regulate harmful immune responses due to a "foreign" protein may cause IDRs in susceptible individuals as described in the danger hypothesis (Chen *et al.*, 2001).

The P-I hypothesis provides a new explanation for the occurrence of IDRs in the absence of a reactive intermediate. It is proposed that T cells do not necessarily bind to protein, but may reversibly bind to major histocompatibility complex/T cell receptors (MHC/TCRs). There is a lack of evidence linking such direct interactions with immune system responses (Soglia *et al.*, 2004).

The danger hypothesis proposes that, rather than the "foreignness" of a reactive metabolite triggering the immune system, cell damage produced by reactive metabolites results in a "danger signal." If the reactive metabolites are not toxic enough to cause cell injury, or if the target cells are resistant to the stress induced by reactive metabolites, protein adducts will have a lower probability of escaping the target cells and inducing a specific immune response. This could be one of the reasons why not all drugs that form reactive metabolites cause IDRs, and not all people are susceptible to these reactions (Evans, 2004).

Currently, the risk assessment of a given toxic effect in humans is usually based on a safety margin of the compound, which is the ratio of the no observable adverse effect level (NOAEL) in the most sensitive species and the expected therapeutic dose in man. The NOAEL is typically derived from a rodent and non-rodent species. Clinical and pathological measurements are the gold standard for identification of organ toxicity in animals. The nonclinical studies are conducted in healthy animals, and the risk assessment is extrapolated to healthy and health-compromised patients. This traditional approach to risk assessment has performed reasonably well in predicting most ADRs; however, it has poor predictability for IDRs (Judson *et al.*, 2003a). Therefore, strategies need to be developed or adapted from current drug safety assessment methods to consider reactive intermediates in the overall assessment of the potential risk profile during drug discovery and development.

***2.3.1 Considerations*** To illustrate how to reduce or eliminate reactive metabolites, consider drug-induced adverse reactions, particularly IDRs (referred to as type B reactions). Such reactions are a major issue because, given current techniques and approaches, they often go undetected until late in the process, or even after the drug has been released onto the market, and the consequences are disastrous for all concerned. IDRs mediated through a reactive metabolite may be associated with several mechanisms (Mehta *et al.*, 1982; Mattocks *et al.*, 1990; Li *et al.*, 2002; Matzinger *et al.*, 2002). Currently there is no general approach or "one-size-fits-all" screen that addresses idiosyncratic reactions, because of the difficulties in understanding the mechanisms of these reactions and in accurately predicting clinical results. Some of the strategies under consideration to improve our understanding of these mechanisms and to develop safer drugs include:

1. Avoidance or flagging of chemical functional groups that are known to cause toxicity during drug design (e.g. aromatic and hydroxy amines, phenols, epoxides, acyl halides, acyl glucuronides, thiopenes, furans, fatty acid-like compounds, hydroxylated metabolites, and quinines).
2. Development of suitable *in vitro* and *in vivo* systems to elucidate the role of short-lived, potentially toxic metabolites in the pathogenesis of idiosyncratic toxicity.
3. Identification of chemical functional groups that are associated with low or no toxicity, and development of more metabolically stable drugs to potentially avoid metabolic interactions.

IDRs mediated through a reactive metabolite may be associated with several mechanisms, and it is unlikely that one test or screen will accurately predict all such mechanisms. Although there is no evidence that safer drugs will be identified by addressing the following questions in the drug discovery/development stages, it is considered that drugs without these potential liabilities may have a better safety profile:

1. Does the candidate have the potential to form reactive metabolites based on chemical "structure-alerts"?
2. Does the compound form reactive metabolites in liver, blood, or skin tissue or cells?
3. Is the binding of the drug >50 pmol/mg of microsomal protein?
4. Does the candidate form reactive intermediates that are able to "travel" and react covalently with other tissues?
5. What proteins are affected and what are the effects of the modified proteins?
6. Which genes are affected? Could these affected genes generate a potential IDR?
7. What cellular functions are affected by the reactive intermediate?
8. Bioaccumulation in liver?
9. Glutathione depletion?
10. Drug–drug interactions?
11. Ames and/or micronucleus positive responses?

It remains to be determined whether the answers to these questions will allow the pharmaceutical industry to reduce or eliminate IDRs. Since the major organs involved in IDRs are the liver, skin, and bone marrow, preclinical approaches should consider these target organs to study the effects of reactive intermediates. Thus, a sound strategy should evaluate three major elements: reactive intermediate characterization, covalent binding, and biological impact and/or function.

## 3 KEY TOOLS AND STRATEGIES TO IMPROVE DRUG SAFETY

The exact mechanism(s) for IDRs are unknown, and because many drugs form reactive metabolites, screening processes may eliminate potentially good candidates (false negatives). Current safety screening methods with animal models do not accurately predict IDRs, according to the Multinational Pharmaceutical Company Survey and the outcome

**TABLE 2 Selected tools to screen for reactive metabolites, covalent binding and biological effects.**

| Stage | Tool | Comment |
|---|---|---|
| Early drug discovery | *In silico* methods/drug design/data mining. | Prior to drug synthesis. |
| Available compounds for screening | Microsomal/CYP time-dependent inactivation.<br>Reactive intermediate screens in human tissues. | *In vitro* studies in cells, subcellular fractions or purified enzymes. |
| Late drug discovery/ candidate selection | Covalent binding.<br>Oxidative stress.<br>Hapten characterization.<br>Reactive intermediate comparison (human versus preclinical species). | Cellular/subcellular/ enzymatic or preliminary *in vivo* studies.<br>Is the binding >50 pmol/mg microsomal protein? |
| Preclinical development | Animal models/biomarkers.<br>Proteonomics/metabolomics/genomics.<br>Reactive intermediate exposure in animal models.<br>Prediction of species differences in exposure of reactive intermediates. | Available animal models have limited prediction rates of IDRs, but new tools may improve predictability. |
| Clinical | Biomarkers/proteonomics/genomics/ metabolomics.<br>Reactive intermediate exposure when considered relevant. | Hapten identification.<br>mRNA.<br>Immunotoxicity.<br>Reactive metabolites (in blood, skin and liver). |

Reprinted from Nassar and Lopez-Anaya, 2004, with permission from Thomson Reuters (Scientific) Limited. Copyright 2004, © The Thomson Corporation.

of the International Life Sciences Institute Workshop (Judson *et al.*, 2003a). Many tools are in the exploratory stage, encompassing computational approaches and experimental assays to predict the formation of reactive metabolites *in vivo* and their role in IDRs (Nassar *et al.*, 2004b). For example, during *in vitro* testing, if significant protein binding is found, an attempt is made to discover the chemical basis for the binding so that it can be designed out of the structure. Another potential technique is to trap reactive intermediates by incubating them in human liver microsomes. To evaluate the biological effects of reactive intermediates initially, methods for cytotoxicity testing including tests for cell viability, membrane integrity, protein synthesis, DNA synthesis, glutathione (GSH) level, apoptosis, free radical production, lipid peroxidation, enzyme inhibition, and other enzymatic activities due to oxidative stress are being developed for early drug discovery. Results suggest that these types of screens can be used to differentiate known toxicants from relatively safe drugs. Table 2 presents some of the current tools with the advantages and disadvantages summarized and related to the problems facing those in drug discovery and development. The study also attempts to synthesize these tools and strategies into an alternative strategy.

### 3.1 *In Silico* Screens as Filtering Process

A promising new trend in early drug discovery is the collaboration between medicinal and computational chemists and drug metabolism scientists seeking to identify and eliminate potential toxic chemical groups. Structure–activity relationships are well

known as a means to identify several toxic end points. One way to improve screening for such potentially toxic compounds would be to increase the use of *in silico* testing methods at earlier stages in drug discovery (Richard *et al.*, 1998; Greene *et al.*, 2002; Judson *et al.*, 2003a, b). Several programs such as TOPKAT, CASE/MULTI-CASE, DEREK, HazardExpert, and OncoLogic are commercially available for the prediction of mutagenicity and carcinogenicity based on chemical structure. These available commercial systems for mutagenicity and/or carcinogenicity prediction differ in their specifics, yet most fall into two major categories. One is automated approaches that rely on the use of statistics for extracting correlations between structure and activity. The others are knowledge-based expert systems that rely on a set of programmed rules distilled from available knowledge and human expert judgment. The advantages of *in silico* techniques include time and money savings, reduced use of laboratory animals, and the ability to rapidly screen large numbers of structures even before synthesis occurs. This high capacity improves prioritization in early discovery for toxicology testing and highlighting toxophores for easy identification. Because it may be difficult to avoid some of these functionalities in the design of new compounds, in some cases their presence should be considered a "structural alert" for the drug metabolism scientist. In turn, studies can be conducted at an earlier stage to determine whether the compound in question undergoes metabolism at the site to generate a potentially toxic intermediate. However, *in silico* methods still cannot completely replace conventional toxicity testing. These systems are not designed/validated to screen for potential IDRs but could be useful as a first step in eliminating compounds with other potential toxic effects.

### 3.2 *In Vitro* Assays

Most *in vitro* assays are amenable to high-throughput screening (HTS), and their use is preferred in early drug discovery. To try to identify reactive metabolites *in vitro*, human tissue, cells, or purified/recombinant enzymes have been used under similar conditions to those in typical drug metabolism studies. Since liver, blood, and skin are the most common targets for the majority of IDRs, these systems have been considered. The human liver and blood systems are the most metabolically active in forming reactive intermediates, while the skin has a lower metabolic activity. Although these *in vitro* assays can identify reactive metabolites, it has been difficult to extrapolate the results to *in vivo* settings. Many drugs have reactive intermediates and *in vitro* assays may result in a significant number of false negatives for IDRs. Metabolic models of cytotoxicity and methods to assess cell death have been reviewed elsewhere (Li *et al.*, 2002; Matzinger *et al.*, 2002). These methods of testing cell viability can be useful to evaluate potential toxic mechanisms. For hepatocellular necrosis, hepatocyte culture and crude cytotoxicity or cell death end points (i.e. enzyme leakage and dye exclusion) can be considered toxicologically relevant. Hepatocyte incubation/cultures have received the most attention (Tiller *et al.*, 2002), while fresh hepatocytes are of limited use. Cryopreserved human hepatocytes are commercially available and technological improvements in preservation will facilitate studies to further understand the mechanisms of IDRs.

### 3.3 Reactive Intermediates: Phase I and Phase II

Phase I and/or phase II enzymes can be involved in drug bioactivation. Several phase I enzymes are recognized as important in forming reactive intermediates such

as cytochrome P450s (CYPs), myeloperoxidases (MPOs), cyclooxygenases (COXs), aldehyde oxidase, and flavin-containing monooxygenases (FMOs) (Pritsos *et al.*, 1985; Kassahun *et al.*, 1993; Liu *et al.*, 1994; Munns *et al.*, 1997; Timmins *et al.*, 1999; Cuttle *et al.*, 2000). In theory, any enzyme is capable of forming reactive intermediates, and the involvement of CYPs has received most attention. Metabolism by FMOs has a lower potential for the formation of reactive metabolites (except for certain *S* or *N* groups) because substrates do not bind directly to the enzyme and electrons are not transferred, avoiding formation of free radicals. In general, phase II metabolism, involving enzymes such as glucuronyl transferases, sulfatase, and GSH transferases, are less frequently recognized as being involved in reactive intermediate formation, despite being responsible for the formation of many reactive intermediates. The study of acyl glucuronides has attracted much interest, as their reactivity is strongly related to covalent binding and potential toxicity.

***3.3.1 Trapping Reactive Intermediates*** Reactive electrophile screenings generated from *in vitro* experiments in several tissues (i.e. liver microsomes) have been designed to react with GSH and allow for subsequent analysis of the GSH adducts by liquid chromatography–mass spectrometry (LC-MS) (Chen *et al.*, 2001). Several types of reactive metabolites (epoxide, arene oxide, quinone, quinone imine, quinone methide, iminoquinone methide, nitroso, nitrenium ion, nitro reduction, nitro radical, iminium ion, free radical, *S*-oxidation, Michael acceptor, *S*-oxide, aliphatic aldehyde, or hydrolysis/acetylation) can be trapped in the presence of either GSH or an equimolar mixture of *N*-acetyl-cysteine and *N*-acetyl-lysine. This method is relatively simple and rapid, and has been implemented for HTS to identify reactive metabolites generated by bioactivation with phase I and phase II enzymes (Kassahun *et al.*, 1993; Tang *et al.*, 1996, 1997; Chen *et al.*, 2001; Soglia *et al.*, 2004). A preliminary evaluation of this method was conducted using 20 commercially available compounds with known toxicological profiles in liver microsomes at substrate concentrations of 10 μM (Chen *et al.*, 2001). The results indicated that the method is unlikely to produce false-negative response because relatively safe compounds did not generate GSH conjugates, while 8 of the 10 compounds that are known to generate reactive metabolites resulted in positive responses. The two compounds that are known to generate reactive metabolites, but did not produce positive responses in the current assay were valproic acid and phenytoin. The reason for this was that the formation of the valproic acid reactive metabolite (2,4-dieneVPA) requires P450-oxidation and β-oxidation catalyzed by a mitochondrial coenzyme A-dependent process rather than microsomal enzymes (Kassahun *et al.*, 1993; Tang *et al.*, 1996, 1997; Soglia *et al.*, 2004). The reactive metabolite of phenytoin may be a free radical instead of an epoxide which cannot form a stable GSH conjugate (Liu *et al.*, 1994; Cuttle *et al.*, 2000). This method was further evaluated using 43 compounds as positive controls and 16 compounds as negative controls. The results indicated that 40 of 43 compounds tested gave positive results as expected from the literature. Those producing false negatives were felbamate, trimethoprim, and sulfamethoxazole. The potential reason for this was a relatively low *in vitro* concentration used in the assay (100 μM). On the other hand, all 16 of the compounds used as negative controls were found to be negative as expected, indicating that this method may be useful in determining the potential for a compound to form reactive intermediates. Although feasible for HTS, the strategy has limited applications to trap the most stable GSH adducts. Those unstable GSH adducts can spontaneously regenerate GSH and the reactive metabolite. A highly reactive metabolite can react with an active site

residue of the enzyme that forms it and thus may not be trapped by GSH. If the reactive metabolite is a free radical, it will most likely abstract a hydrogen atom from GSH rather than react with it. Alternative trapping reagents may be used for more reactive intermediates (Timmins *et al.*, 1999).

Since thousands of compounds must be screened in a typical pharmaceutical industry setting, it is almost impossible, and very expensive, to have all of them radiolabeled. Using radiolabeled reagents such as GSH and KCN is easier and more cost-effective to obtain an accurate measurement of adduct formation (Nassar *et al.*, 2004a). The reactive intermediates may be formed in other subcellular fractions such as liver S9, mitochondria, and cytosol, and thus, screening for reactive intermediates in these fractions will minimize false negatives. As an example, valproic acid forms reactive metabolites in subcellular mitochondria, as explained earlier. We have compared *in vitro* cellular and subcellular models for identifying drug candidates with the potential to produce reactive metabolites. For purposes of evaluating these models, six known compounds currently in the clinic with dose limitations due to hepatotoxicity were selected. Labetalol, acetaminophen, niacin, iproniazid, 8-hydroxyquinoline, and isoniazid were incubated separately with human liver mitochondria, S9, microsomes, cytosol, and hepatocytes. Liquid chromatography hybrid quadrupole time of flight mass spectrometry (LC-QTOF-MS) methods were used to elucidate metabolic profiles of selected compounds. The phase II metabolites were in good agreement with previously published results. Without supplementation of reagents in subcellular fractions, phase II metabolites were not detected. Hepatocytes also formed GSH conjugates but the relative abundance was lower, indicating that the best model to identify potential reactive metabolites may involve the subcelluar fractions (Nassar *et al.*, 2004a). Using the results for acetaminophen, for example, when reactive intermediates are produced, they can be trapped by GSH, while in the case of labetalol, no reactive intermediates were produced and subsequently no GSH adducts were detected. Studies should be done to improve our knowledge of, and hence techniques used for, screening with the use of radiolabeled reagents, which would give an accurate measurement of the amount of reactive intermediates (Nassar *et al.*, 2004a). Also, we should investigate reactive intermediate formation in subcellular fractions, which will aid in reducing false results (Nassar *et al.*, 2004a).

***3.3.2 Trapping Free Radicals*** Techniques for trapping free radicals are well established. Free radicals resulting from reduction and oxidation can be trapped with spin-trapping reagents (Pritsos *et al.*, 1985; Timmins *et al.*, 1999; Sanders *et al.*, 2000). For example, a method for trapping free radicals was used on hydrazine analogs and their derivatives which are responsible for hemolytic and hepatotoxic events, presumably via the alkyl or aryl free radicals that oxidize essential cysteinyl residues in proteins or covalently react with biomacromolecules (Gorrod, Whittesea and Lam, 1991). Another example is the bioactivation of phenytoin to a free radical species, proposed to be mediated by COX-1, which can be trapped with the spin trapping agent α-phenyl-*N*-t-butylnitrone (PNB) *in vitro* in embryo cell cultures (Liu *et al.*, 1994; Munns *et al.*, 1997; Timmins *et al.*, 1999; Cuttle *et al.*, 2000). The techniques to trap free radicals need further evaluation to develop high-throughput methodology and establish their role in IDRs.

***3.3.3 Trapping Iminium Ion*** Trapping iminium ions with cyanide is also a well-known technique (Gorrod, Whittesea and Lam, 1991). The technique can be implemented as a screen for the detection of reactive iminium intermediates by trapping

them with radiolabeled cyanide. With (*S*)-nicotine as the reference compound, the extent of radiolabeled cyanide incorporated into the test compounds can be quantified. If compounds have higher cyanide incorporation than (*S*)-nicotine, it is considered indicative of the formation of iminium intermediates. However, the iminium ions can also form GSH adducts depending on their relative stability. For example, U-89843 incubated in liver microsomes formed *N*-acetylcysteinyl and GSH adducts in NADPH-supplemented rat liver microsomes supporting a bioactivation pathway potentially involving an iminium intermediate (Zhao *et al.*, 1996a, b). The reactive metabolites of both DMP 406 and mianserin reacted with a range of nucleophiles, but in many cases the reaction was reversible. The best nucleophile for trapping these reactive metabolites was cyanide (Gorrod, Whittesea and Lam, 1991). These results suggest the potential need for a combination of trapping methods to evaluate the formation of reactive intermediates.

Taken together, the results suggest the need for a combination of trapping methods to evaluate the formation of reactive intermediates.

## 4 PEROXIDASES

Peroxidases are present in polymorphonuclear granulocyte (PMN) and play a significant role in the bioactivation of many drugs in humans. Several *in vitro* systems have been used to try to mimic bioactivation processes *in vivo* (i.e. myeloperoxidase [MPO]/$H_2O_2/Cl_2$, HOCl, chloroperoxidase, and horseradish peroxidase [HRP]/$H_2O_2$). A preliminary evaluation of the nucleophile-activated system was conducted to assess its ability to predict drug-induced agranulocytosis (Richard *et al.*, 1998). In this study, DMP-406 (a clozapine analog), which caused agranulocytosis in dogs, was used as the testing agent, with clozapine and mianserin as positive controls, and olanzapine as a negative control. Clozapine is thought to cause agranulocytosis via a reactive nitrenium ion metabolite produced by neutrophils. It has been demonstrated that the products of clozapine oxidation by HRP/$H_2O_2$ induced apoptosis in neutrophils at therapeutic concentrations (Colburn, 2003). The major reactive intermediate of DMP-406 is an imine ion in activated neutrophils and is similar to the reactive intermediate responsible for mianserin-induced agranulocytosis. DMP-406 did not increase apoptosis at concentrations <50 μM, while mianserin increased apoptosis at a concentration 10 μM above its therapeutic concentration, indicating that this assay lacks predictability (resulting in false negatives) for both drugs. Olanzapine increased apoptosis at the same concentration as clozapine (1 μM), but this concentration is above the therapeutic concentration of olanzapine, consistent with *in vivo* results. There was no increase in apoptosis with any drug in the absence of HRP/$H_2O_2$, which forms the reactive intermediates of these drugs. These preliminary results indicate that this assay may be unable to reliably predict the ability of different types of drugs to cause agranulocytosis and that different drugs may induce agranulocytosis by different mechanisms. Drugs that form nitrenium ion intermediates might be better predicted by this system.

## 5 ACYL GLUCURONIDATION AND *S*-ACYL-COA THIOESTERS

Phase II biotransformation is generally considered to be a bioinactivation pathway, but some drugs become bioactivated during this phase (Madden *et al.*, 1993; Takeyama *et al.*, 1993; Liu *et al.*, 1995; Shapiro *et al.*, 1997; Cziraky *et al.*, 1998; Isaacson *et al.*, 1999; Kassahun *et al.*, 2001; Tettey *et al.*, 2001; Nassar *et al.*, 2004a; Schonbeck *et al.*, 2004).

Since acyl glucuronide metabolites of some carboxylic acids might cause fatal IDRs in some patients (Takeyama *et al.*, 1993), much attention has been focused on the formation of reactive intermediates through glucuronidation and the methods to screen these conjugates to select for safer drugs (Takeyama *et al.*, 1993; Cziraky *et al.*, 1998; Schonbeck *et al.*, 2004). Some acyl glucuronides are reactive intermediates, which bind covalently to protein by mechanisms that may or may not result in the cleavage of the glucuronic acid moiety. Carboxylic acids can be metabolized by alternative pathways: (i) by glucuronide conjugation to form *O*-acyl glucuronides and (ii) by CoA conjugation to form *S*-acyl-CoA thioesters. Both pathways result in the formation of potentially reactive, electrophilic intermediates, due to the reactivity of the carbonyl carbon. The electrophilic carbon center can react with nucleophilic targets on macromolecules, forming covalent adducts with biomacromolecules, for example, proteins. Three major factors have been suggested to determine the reactivity of acyl glucuronides *in vivo*. The first is the relative stability of a given acyl glucuronide in aqueous buffer at pH 7.4/37°C. Some compounds with short, first-order half-lives (<0.5 hour) result in a relatively high concentration of drug–protein adducts in human tissues and are associated with a high risk for causing IDRs in patients (Takeyama *et al.*, 1993; Liu *et al.*, 1995; Ruscoe *et al.*, 1995). In contrast, acyl glucuronides with longer half-lives (>5 hours) have rarely been reported to cause IDRs. Second is the degree of substitution at the α-carboxyl carbon of the aglycone (an increase in degree of substitution leads to a decrease in acyl glucuronide reactivity). The third factor is the overall drug exposure (including the up-concentration of the acyl glucuronides) in particular tissue compartments. A recent example of apparent success using this screening strategy is the assessment of telmisartan 1-*O*-acyl glucuronide, the principal metabolite of telmisartan in humans. Telmisartan 1-*O*-acyl glucuronide exhibited a long, first-order degradation half-life of 26 hours compared with the short half-life of diclofenac (0.5 hour), suggesting telmisartan has a low potential for covalent binding (Cziraky *et al.*, 1998). The current literature indicates a low incidence/absence of ADRs for telmisartan, supporting the potential use of this screening approach.

*S*-acyl-CoA thioesters have received significant attention as reactive intermediates involved in IDRs (Takeyama *et al.*, 1993). They can be formed with carboxylic acid drugs, and the resultant thioesters can be much more reactive than the acyl glucuronides due to the nature of the thioester bond. Similarly to acyl glucuronides, the reactivity of synthetic *S*-acyl-CoA thioesters at pH 7.4/37°C may be an important indicator of their potential to covalently bind to biomolecules. The covalent binding of seven structurally different carboxylic acid drugs for GSH was compared at pH 7.4/37°C. The results indicated that hydrolysis rates might be a good predictor for GSH conjugation, and similarly to acyl glucuronides, substitution at the α-carbon to the thioester bond affects reactivity (ISSX 2003). An HTS assay could be easily implemented to assess reactivity or stability of *S*-acyl-CoA thioesters, but additional studies are necessary to further evaluate the role of this metabolic pathway in IDRs.

## 6 COVALENT BINDING

Several methods are available to detect or quantify covalent binding of drugs and their metabolites to macromolecules, including radiochemical and immunological methods, and protein analysis (LC-MS, matrix-assisted laser desorption ionization time-of-flight (MALDI-TOF), and quadrupole time-of-flight (Q-TOF)). A large number of studies

indicate that there is a good correlation between drug-induced toxicities and protein covalent binding (Takeyama *et al.*, 1993; Liu *et al.*, 1995; Ruscoe *et al.*, 1995). The alkylation of certain proteins has been suggested to lead to formation of neoantigens that may trigger immunomediated hepatotoxicity. Immunochemical techniques have been used to identify the protein adducts of hepatotoxic compounds. In an effort to understand the significance of *in vitro* covalent binding, studies have been carried out in animals, where GSH adduct and/or protein adducts were detected in biological fluids (i.e. in samples of bile) (Munns *et al.*, 1997; Ju *et al.*, 1999). Subsequent experiments with radiolabeled compound can reveal more detail on the propensity of the reactive intermediate to covalently bind cellular proteins *in vitro* and *in vivo*. From a quantitative viewpoint, it has been suggested that as a rule of thumb, compounds with potential toxic profiles have a binding affinity of 50 pmol/mg microsomal protein; however, in order to quantify the covalent binding, radiolabeled drug is required (ISSX 2002, 2003). An example of the use of this approach in drug discovery has been presented recently using tritiated analogs to rank compounds with lower covalent binding in microsomal protein (Munns *et al.*, 1997). Another example of this approach has been discussed using four radiolabeled compounds including imipramine and diclophenac, which exhibited >50 pmol/mg suggesting potential risk for documented IDRs. However, additional studies may be necessary to further evaluate the clinical relevance of the 50 pmol/mg value since only liver microsome proteins have been considered. There are many other enzymes that can also form reactive metabolites in other cellular fractions (cytosol or mitochondria), blood, lung, and so on.

The question remains as to what is an acceptable level of covalent binding to proteins/macromolecules in the liver (or other organs) for a drug candidate to be taken into development. Another important consideration is that covalent binding might be an important detoxification pathway for some reactive intermediates, and it should therefore not always be perceived as a negative attribute for a drug in development. Without a sound understanding of the potential implications/mechanisms of covalent binding, a safe and effective drug could be eliminated from further development.

***6.1.1 Reactive Oxygen Species (ROS)*** It has been proposed that some IDRs may be the result of oxidative stress (increases in the intracellular levels of ROS) (Uetrecht *et al.*, 1989; Takeyama *et al.*, 1993; Masubuchi *et al.*, 2000, 2002). Generally, ROS are generated as by-products from cellular metabolism, primarily in the mitochondria. When the cellular production of ROS exceeds the antioxidant capacity of the cell, cellular macromolecules such as lipids, proteins, and DNA can be damaged. To prevent oxidative stress under normal physiological conditions, these free radicals are neutralized by an elaborate antioxidant defense system consisting of enzymes (e.g. catalase, superoxide dismutase, and GSH peroxidase), and numerous nonenzymatic antioxidants (e.g. vitamins A, E and C, GSH, ubiquinone, and flavonoids) (Nikulina *et al.*, 2000; Yin *et al.*, 2001; Miyamoto *et al.*, 2003). GSH-associated metabolism is a major mechanism for cellular protection against agents or reactive intermediates. It has been reported that several mechanisms could contribute to cell death associated with oxidative stress, which can be summarized in three steps (O'Donovan *et al.*, 2000). The first stage is the recognition of stress by sensitive protein(s) (e.g. the depletion of GSH), mitochondrial damage, inactivation of critical cellular functions, activation of transcription factors, defense gene expression, protein expression, protein function, and release of pro-inflammatory cytokinase. Stage 2 is the subsequent activation of cellular defenses through phase II metabolism enzymes (e.g. uridine

diphosphate-glucuronyltransferase (UDP-GT)), GSH-related enzymes (e.g. glutathione S-transferase (GST)), heat shock proteins (e.g. Hsp72), antioxidants, and cell cycle inhibitors. At stage 3, the cells of tolerant individuals are able to dynamically protect themselves from continued stress, but the cells of susceptible individuals may not be able to do so, leading to premature apoptotic death. Several strategies of antioxidative defense include transition metals inactivated by chelating proteins (e.g. ferritin), and ROS reduced enzymatically (e.g. by the glutathione peroxidase) or nonenzymatically by antioxidants (e.g. by vitamin E, vitamin C, and glutathione).

## 7 MECHANISTIC STUDIES

Screening for cytotoxicity, covalent binding, and oxidative stress may help to characterize reactive intermediates. However, a greater understanding of the role of the reactive metabolite will allow for better decisions to be made in the final assessment of quality candidates. The apparent lack of predictability of IDRs by animal models may be due to (i) differences in drug metabolism (and, therefore, reactive metabolites being formed) between species, (ii) species differences in immune-mediated IDRs, and (iii) predisposing factors in humans that do not usually exist in animal models (e.g. alcohol, coadministration of drugs, or disease). Some of the following mechanistic studies have furthered our understanding of the safety profile of a drug candidate with reactive intermediates.

### 7.1 Covalent Binding to Albumin/Plasma Proteins

In general, covalently modified proteins can be repaired or degraded; if not, they may impair important cellular functions, which could be directly pathogenic (Parrish *et al.*, 1997). Several acyl glucuronides have been demonstrated to covalently bind albumin/plasma proteins *in vitro* and *in vivo* (human volunteers). Studies of covalent binding to albumin can be performed with or without radiolabeled compounds. To facilitate these studies, radiolabeled material is preferred and can be useful to identify plasma proteins. Synthetic standards of intermediate metabolites (i.e. acyl glucuronides) are required to assess covalent binding to these proteins. Although non-radiolabeled material can be used to identify albumin covalent binding, quantitation via this approach is difficult. The purpose of these *in vitro* or *in vivo* studies is to characterize the disposition of reactive intermediates in plasma; these studies cannot answer the question of whether plasma proteins can function as protectants of reactive intermediate toxicity or are pathogenic after modification. For some acyl glucuronides, current, indirect evidence suggests a link between covalent plasma binding of tissues and potential for IDRs.

### 7.2 Time-Dependent Inhibition

CYP time-dependent inhibition studies can potentially identify reactive intermediates when they are covalently bound to metabolizing enzymes, as this interaction decreases the activity of the latter. CYP time-dependent inhibition assays are widely used and implemented in a high-throughput manner during early drug discovery as part of the candidate selection criteria. However, these data may be used only as preliminary information since some reactive intermediates do not exhibit time dependency. For

example, acetaminophen (APAP) is not a CYP inhibitor but its reactive intermediate *N*-acetyl-4-benzoquinone imine (NAPQI) is formed by CYP2E1. The stability of the reactive intermediate may be an important factor that determines the effect on time dependency. A highly reactive intermediate may react with the metabolizing enzyme, rendering it unable to distribute to other cell compartments or tissues, while a reactive intermediate with moderate reactivity may diffuse into additional cell compartments. For example, APAP reactive metabolite appears to distribute from the endothelium reticulum into the mitochondria where it reacts with critical proteins. On the other hand, tienilic acid induces hepatotoxicity, and patients form anti-LKM2 (liver kidney microsomes type 2) antibodies in the serum which recognizes CYP2C9. Tienilic acid covalently binds to CYP2C9, the major CYP metabolizing enzyme, and inhibits its activity in a time-dependent manner (Mohsin *et al.*, 2004). The time inhibition assay may record reactive intermediates with moderate reactivity as false negatives, and may be limited to the activity of the enzyme that is being measured.

### 7.3 Antioxidants/Trapping Reagents

*In vivo* and *in vitro* studies have been conducted with supplementation/depletion of antioxidants to elucidate toxicity mechanisms. The results of these studies indicate that antioxidants can prevent covalent binding of several types of reactive intermediates to biomolecules and they can also prevent oxidative stress generated by non-covalent interactions. For example, addition of antioxidants such as GSH or ascorbic acid resulted in significant prevention of CYP1A2-dependent cytotoxicity and protein-reactive metabolite formation of tacrine. In another *in vitro* example during early discovery, the amount of reactive intermediate covalently bound to microsomal protein was reduced on addition of GSH to the microsomal incubation, as determined using radiolabeled compounds (Munns *et al.*, 1997). Another interesting approach is the use of the spin-trap agent PNB *in vivo*. This trapping agent was used with thalidomide, which initiates embryonic DNA oxidation and teratogenicity in rabbits, both of which were abolished by pre-treatment with PNB.

Several potential biomarkers of oxidative stress can be monitored to evaluate the potential toxic effects of drugs (Uetrecht *et al.*, 1994; Masubuchi *et al.*, 2000). Drugs that undergo redox cycling or form free radicals can generate toxic effects through oxidative stress without forming covalent adducts with biomolecules. This suggests that certain groups of compounds can be screened out in the drug design. The overall data in this area of research suggest that more study is required to select the appropriate biomarker(s) of oxidative stress that can aid in the assessment of the safety profile of a new potential drug candidate. The potential advantage of an oxidative stress biomarker is that it can be monitored during *in vitro* and/or *in vivo* studies.

***7.3.1 Biological Markers (Biomarkers)*** Biomarkers can be measured and quantified, providing useful information for a wide range of clinical and preclinical uses (Colburn, 2003; Ilyin, Belkowski and Plata-Salaman, 2004; Lesko and Atkinson, 2001). Some potential examples of biological parameters which can be measured include: concentration of specific enzyme(s) and/or specific hormones; specific gene phenotype distribution in a population; presence of biological substances which are useful as indicators for health and physiology related assessments such as disease risk, psychiatric disorders, environmental exposure and its effects; disease diagnosis; metabolic processes; substance abuse; pregnancy; cell line development; and epidemiologic studies.

These and other parameters can be used to identify a toxic effect in an individual organism and can be used to extrapolate between species. Biomarkers can serve to confirm diagnoses, monitor treatment effects or disease progression, and predict clinical results.

Biomarkers are clearly indicated as having important roles in drug development for a number of situations, such as their ability to provide a rational basis for selection of lead compounds, as a help in determining the ability to work toward qualification and use of a biomarker as a surrogate end point. Changes in a biomarker can provide useful indicators for pathophysiology, which in turn are important in identifying a suitable therapeutic target. For example, the association of elevated serum cholesterol levels with an increased incidence of coronary heart disease provides an underlying rationale for developing drugs that lower cholesterol by inhibiting 3-hydroxy-3-methylglutaryl coenzyme A reductase. Thus, total cholesterol is a good example of a clinical biomarker that has been qualified for use as a surrogate end point (Schonbeck *et al.*, 2004; Cziraky *et al.*, 1998).

Biomarkers are also important to the preclinical assessment of the potential benefits and harmful effects of a new drug candidate. Screening tests in animals using biomarkers, such as reduction of blood pressure, provide important demonstration that a compound is likely to produce the intended therapeutic activity in patients. By measuring blood levels during adverse events, such as seizures, in animal toxicology studies may help guide the design of dose escalation studies in humans and serve as a surrogate for preventing or reducing the likelihood of similar adverse events in humans. Biomarkers for potential toxicity play an equally important role. For example, a drug found to prolong the QT interval in animals may warn of potential cardiovascular risk in subsequent clinical studies. Also, biomarkers could be glutathione conjugates and/or glutathionylated or oxidized proteins detectable both in vitro and in vivo. Biomarkers of oxidative stress involving gene/protein expression at least require a whole cell system. It is highly unlikely that a single compound/gene/protein/function will be an effective biomarker, however combinations of these may prove more successful. Several potential biomarkers of oxidative stress can be monitored to evaluate the potential toxic effects of drugs. Drugs that undergo redox cycling or form free radicals can generate toxic effects through oxidative stress without forming covalent adducts with biomolecules. This suggests that certain groups of compounds can be screened out in the drug design phase. The potential advantage of an oxidative stress biomarker is that it can be monitored during in vitro and/or in vivo studies. Pharmacokinetic-pharmacodynamic studies using biomarkers may be particularly useful; for example, one such study showed good correlation between the hypotensive effects of an antiarrhythmic drug in dogs and humans. One shortcoming is that most biomarkers, used singly, are unlikely to capture all the effects of a drug, and thereby fulfill the most stringent criterion for a surrogate end point, although by using several of them in combination, it is more likely, and desirable, to produce evidence which is consistent enough to point in a particular direction.

## 8 PRECLINICAL DEVELOPMENT

The main goal in preclinical development is the extrapolation of efficacy and safety data from animal and *in vitro* models to humans (Fan *et al.*, 2001). This approach should be considered when a metabolic pathway is suspected to play a major role in this

extrapolation. One example of the usefulness of this technique is illustrated by tamoxifen, which results in the development of hepatic tumors in rats, although these tumors were not detected in mice in the rodent carcinogenicity assay (Fan *et al.*, 2001). The proposed metabolic pathway responsible for tumorigenesis involves sequential bioactivation of tamoxifen via α-hydroxylation and *O*-sulfonation, and the resultant reactive metabolite reacts with DNA. *In vitro* bridging studies with hepatic subcellular fractions, which formed part of the human risk assessment, demonstrated that rat microsomes were 3-fold more active than human microsomes in forming α-hydroxytamoxifen, and 5-fold more active than human hepatic cytosol with regard to *O*-sulfonation (bioactivation). In contrast, the rate of *O*-glucuronidation (bioinactivation) in human hepatic microsomes was at least 100-fold greater than for rat hepatic microsomes. The dose of tamoxifen required to induce tumors in rats was 40 mg/kg, which is ~150-fold greater than the therapeutic dose in man (0.3 mg/kg). This data suggests a safety margin of 150,000 for risk of tumors, which is consistent with the clinical experience of the drug to date, and is consistent with the mouse data. This example, along with others, illustrates the need for a better understanding of the connection between exposure of reactive intermediates and their reactivity in the assessment of a safety profile. However, a reliable animal model to predict IDRs is needed to be able to justify the use of metabolic profiles in the safety profile assessment. Currently, species differences in metabolic pathways may be considered in the attempt to bridge safety data between animals and humans. Additional knowledge of biomarkers may also help improve the development of this approach.

## 9 CLINICAL DEVELOPMENT: STRATEGY

Drugs dosed at ≤10 mg/day appear to have a low incidence of IDRs, while low potent drugs (β-lactam antibiotics, sulfonamides, phenobarbital, phenytoin, carbamazepine, tricyclic antidepressants, and non-steroidal anti-inflammatory drugs administered at unusually high doses) administered at doses of >10 mg/day may exhibit a higher incidence of IDRs. It has been suggested that more potent drugs should be developed to reduce the risk of IDRs. For example, olanzapine forms a reactive nitrenium ion similar to that formed by clozapine, which is considered to be responsible for clozapine-induced agranulocytosis. Olanzapine, however, is not associated with a significant incidence of agranulocytosis. A critical difference between the two drugs is that of the dose delivered: clozapine is administered at a dose of several hundred mg/day while olanzapine is administered at a maximum dose of 10 mg/day. Another example is troglitazone, which, when administered at 200 to 600 mg/day, resulted in serious hepatic injury and was withdrawn from the market by the FDA in 2000. Rosiglitazone, an analog of troglitazone, is administered at 4 to 8 mg/day and is not hepatoxic (Mann *et al.*, 1987). The *in vitro* induction of CYP enzymes by rosiglitazone and troglitazone suggests that other thiazolidinediones may have the potential to cause clinically significant drug interactions if administered at sufficiently high doses (Borges *et al.*, 2000).

In an effort to characterize the spectrum of IDRs more accurately, the following issues may be considered during early-to-late clinical phases:

1. Study of high-risk patients to identify pharmacokinetic and pharmacodynamic factors that influence susceptibility to drug toxicity.

2. Development of computer-based schemes to monitor for adverse reactions and adverse events in primary and secondary care.
3. Encouragement to report ADRs to regulatory agencies.
4. Identification of risk factors for different types of drug toxicity using pharmacoepidemiological approaches.
5. Identification of multigenic predisposing factors to permit the prediction of individual susceptibility.

## 10 CONCLUSION AND FUTURE POSSIBILITIES

During the past decade, there has been an enormous increase in our understanding of how cells and organisms respond to the generation of metabolites, which are chemically reactive. An important consideration is that any screening process will produce false positive and false negative results, and our goal must be to reduce or eliminate these, which in turn will improve drug safety.

There are two major inter-related points of emphasis in the effort to understand and manage IDRs. Firstly, research should focus on the process or mechanism by which IDRs occur. Unfortunately, due to their relative rarity and unpredictability, these have proven difficult to determine with any certainty. Secondly, further studies are required to improve our knowledge in and hence techniques used for screening, and this will be aided by continued improvement in cooperation and dialogue between pharmaceutical companies and academia. For example, greater advances could be gained if researchers involved in this field shared data to correlate human toxicity with animal toxicity or functional assays. Future screens could focus on biomarkers for oxidative stress. Since there is a general inconsistency in the correlation of toxicity with covalent binding, further validation of *in vitro* covalent binding with regard to *in vivo* toxicity is required. In our opinion, future implementation plans should focus on: (i) correlating covalent binding in different *in vitro* systems (animal models and human); (ii) defining biomarkers for oxidative stress; (iii) correlating covalent binding *in vitro* with findings in animals; (iv) continued improvement of databases of genomics/proteomics; and (v) extrapolating data to humans. It is worth stating that drug safety should be further investigated by pharmaceutical companies to conduct post-marketing studies, which are required by the Food and Drug Administration when a safety question arises during the pre-approval period.

## REFERENCES

Bailey, M.J. *et al.* (1996). Chemical and immunochemical comparison of protein adduct formation of four carboxylate drugs in rat liver and plasma. *Chem. Res. Toxicol.*, *9*, 659–666.

Bakke, O.M. *et al.* (1995). Drug safety discontinuations in the United Kingdom, the United States and Spain from 1974 through 1993: a regulatory perspective. *Clin. Pharmacol. Ther.*, *58* (1), 108–117.

Borges, N. *et al.* (2000). Tolcapone-related liver dysfunction: implications for use in Parkinson's disease therapy. *Drug Saf.*, *26* (11), 743–747.

Burkhart, C. *et al.* (2002). Non-covalent presentation of sulfamethoxazole to human CD4+ T cells is independent of distinct human leucocyte antigen-bound peptides. *Clin. Exp. Allergy*, *32* (11), 1635–1643.

Chen, W.G. *et al.* (2001). Reactive metabolite screen for reducing candidate attrition in drug discovery, in *Biological Reactive Intermediates VI. Chemical and Biological Mechanisms in Susceptibility to and Prevention of Environmental Diseases* (eds P.M. Dansetter, R. Snyder, M. Delaforge, G.G. Gibson, H. Geim, D.J. Jollow, T.J. Monks and I.G. Sipes), Kluwer Academic/ Plenum Press, New York, pp. 521–524.

Colburn, W.A. (2003). Biomarkers in drug discovery and development: from target identification through drug marketing. *J. Clin. Pharmacol.*, *43* (4), 329–341.

Coleman, M.D. *et al.* (1994). Reduction of dapsone hydroxylamine to dapsone during methaemoglobin formation in human erythrocytes in vitro. IV: Implications for the development of agranulocytosis. *Biochem. Pharmacol.*, *48* (7), 1349–1354.

Cribb, A.E. *et al.* (1992). Sulfamethoxazole is metabolized to the hydroxylamine in humans. *Clin. Pharmacol. Ther.*, *51*, 522–526.

Cuttle, L. *et al.* (2000). Phenytoin metabolism by human cytochrome P450: involvement of P450 3A and 2C forms in secondary metabolism and drug-protein adduct formation. *Drug Metab. Dispos.*, *28* (8), 945–950.

Cziraky, M. *et al.* (1998). Clinical positioning of HMG-CoA reductase inhibitors in lipid management protocols. *Pharmacoeconomics*, *14* (Suppl. 3), 29–38.

Dieckhaus, C.M. *et al.* (2002). Mechanisms of idiosyncratic drug reactions: the case of felbamate. *Chem. Biol. Interact.*, *142* (1–2), 99–117.

Evans, D.C. (2004). Drug-protein adducts: an industry perspective on minimizing the potential for drug bioactivation in drug discovery and development. *Chem. Res. Toxicol.*, *17* (1), 3–16.

Fan, P.W. *et al.* (2001). Bioactivation of tamoxifen to metabolite E quinone methide: reaction with glutathione and DNA. *Drug Metab. Dispos.*, *29*, 891–896.

Godessart, N. *et al.* (1999). Role of COX-2 inhibition on the formation and healing of gastric ulcers induced by indomethacin in the rat. *Adv. Exp. Med. Biol.*, *469*, 157–163.

Gorrod, J.W., Whittesea, C.M.C., Lam, S.P. (1991). Trapping of reactive intermediates by incorporation of 14C-sodium cyanide during microsomal oxidation. *Adv. Exp. Med. Biol.*, *283*, 657–664.

Greene, N. *et al.* (2002). Computer systems for the prediction of toxicity: an update. *Adv. Drug Deliv. Rev.*, *54* (3), 417–431.

Grillo, M.P. *et al.* (2001). Effect of alpha-fluorination of valproic acid on valproyl-S-acyl-CoA formation in vivo in rats. *Drug Metab. Dispos.*, *29*, 1210–1215.

Gupta, A.K. *et al.* (1998). Severe neutropenia associated with oral terbinafine therapy. *J. Am. Acad. Dermatol.*, *38*, 765–767.

Haack, M.J. *et al.* (2003). Toxic rise of clozapine plasma concentrations in relation to inflammation. *Eur. Neuropsychopharmacol.*, *13* (5), 381–385.

Halsey, J.P. *et al.* (1982). Benoxaprofen: side-effect profile in 300 patients. *BMJ*, *284*, 1365–1368.

Idanpaan-Heikkila, J. *et al.* (1977). Agranulocytosis during treatment with chlozapine. *Eur. J. Clin. Pharmacol.*, *11*, 193–198.

Ilyin, S.E., Belkowski, S.M., Plata-Salaman, C.R. (2004). Biomarker discovery and validation: technologies and integrative approaches. *Trends Biotechnol.*, *22* (8), 411–416.

Inman, W.H. *et al.* (1988). Blood disorders and suicide in patients taking mianserin or amitriptyline. *Lancet*, *2* (8602), 90–92.

Isaacson, J. *et al.* (1999). Index of suspicion. Case 3. Acetaminophen overdose. *Pediatr. Rev.*, *20*, 309–310, 312–313.

Jardine, D.L. (2002). Hodgkin's disease following methotrexate therapy for rheumatoid arthritis. *N. Z. Med. J.*, *115* (1156), 293–294.

Ju, C. *et al.* (1999). Detection of 2-hydroxyiminostilbene in the urine of patients taking carbamazepine and its oxidation to a reactive iminoquinone intermediate. *J. Pharmacol. Exp. Ther.*, *288*, 51–56.

Judson, P.N. *et al.* (2003a). A comprehensive approach to argumentation. *J. Chem. Inf. Comput. Sci.*, *43*, 1356–1363.

Judson, P.N. *et al.* (2003b). Using argumentation for absolute reasoning about the potential toxicity of chemicals. *J. Chem. Inf. Comput. Sci.*, *43*, 1364–1370.

Kassahun, K. *et al.* (1993). In vivo formation of the thiol conjugates of reactive metabolites of 4-ene VPA and its analog 4-pentenoic acid. *Drug Metab. Dispos.*, *21* (6), 1098–1106.

Kassahun, K. *et al.* (2001). Studies on the metabolism of troglitazone to reactive intermediates in vitro and in vivo. Evidence for novel biotransformation pathways involving quinone methide formation and thiazolidinedione ring scission. *Chem. Res. Toxicol.*, *14*, 62–70.

Lasser, K.E. *et al.* (2002). Timing of new black box warnings and withdrawals for prescription medications. *JAMA*, *287* (17), 2215–2220.

Lesko, L.J., Atkinson, A.J., Jr (2001). Use of biomarkers and surrogate endpoints in drug development and regulatory decision making: criteria, validation, strategies. *Annu. Rev. Pharmacol. Toxicol.*, *41*, 347–366.

Li, A.P. *et al.* (2002). A review of the common properties of drugs with idiosyncratic hepatotoxicity and the "multiple determinant hypothesis" for the manifestation of idiosyncratic drug toxicity. *Chem. Biol. Interact.*, *142* (1–2), 7–23.

Liu, L. *et al.* (1994). In vivo phenytoin-initiated oxidative damage to proteins and lipids in murine maternal hepatic and embryonic tissue organelles: potential molecular targets of chemical teratogenesis. *Toxicol. Appl. Pharmacol.*, *125* (2), 247–255.

Liu, Z.C. *et al.* (1995). Oxidation of 5-aminosalicylic acid by hypochlorous acid to a reactive iminoquinone. Possible role in the treatment of inflammatory bowel diseases. *Drug Metab. Dispos.*, *23*, 246–250.

Madden, S. *et al.* (1993). An investigation into the formation of stable, protein-reactive and cytotoxic metabolites from tacrine in vitro. Studies with human and rat liver microsomes. *Biochem. Pharmacol.*, *46*, 13–20.

Mann, K. *et al.* (1987). Differential effects of a new dibenzo-epine neuroleptic compared with haloperidol. Results of an open and crossover study. *Pharmacopsychiatry*, *20*, 155–159.

Masubuchi, Y. *et al.* (2000). Possible mechanism of hepatocyte injury induced by diphenylamine and its structurally related nonsteroidal anti-inflammatory drugs. *J. Pharmacol. Exp. Ther.*, *292* (3), 982–987.

Masubuchi, Y. *et al.* (2002). Role of mitochondrial permeability transition in diclofenac-induced hepatocyte injury in rats. *Hepatology*, *35* (3), 544–551.

Mattocks, A.R. *et al.* (1990). Trapping of short-lived electrophilic metabolites of pyrrolizidine alkaloids escaping from perfused rat liver. *Toxicol. Lett.*, *54* (1), 93–99.

Matzinger, P. *et al.* (2002). The danger model: a renewed sense of self. *Science*, *296* (5566), 301–305.

Mehta, J.R. *et al.* (1982). Trapping of DNA-reactive metabolites of therapeutic or carcinogenic agents by carbon-14-labeled synthetic polynucleotides. *Cancer Res.*, *42* (8), 2996–2999.

Merlani, G. *et al.* (2001). Fatal hepatoxicity secondary to nimesulide. *Eur. J. Clin. Pharmacol.*, *57*, 321–326.

Meyboom, R.H.B. *et al.* (2000). An ABC of drug-related problems. *Drug Saf.*, *22* (6), 415–423.

Miyamoto, Y. *et al.* (2003). Oxidative stress caused by inactivation of glutathione peroxidase and adaptive responses. *Biol. Chem.*, *384* (4), 567–574.

Mohsin, S.K. *et al.* (2004). Progesterone receptor by immunohistochemistry and clinical outcome in breast cancer: a validation study. *Mod. Pathol.*, *17*, 1545–1554.

Moses, P.L. *et al.* (1999). Severe hepatotoxicity associated with bromfenac sodium. *Am. J. Gastroenterol.*, *94* (5), 1393–1396.

Munns, A.J. *et al.* (1997). Bioactivation of phenytoin by human cytochrome P450: characterization of the mechanism and targets of covalent adduct formation. *Chem. Res. Toxicol.*, *10* (9), 1049–1058.

Nassar, A.E. *et al.* (2004a). Detecting and minimizing reactive intermediates in R&D. *Curr. Drug Discov.*, July, 20–25.

Nassar, A.F. *et al.* (2004b). Strategies for dealing with reactive intermediates in drug discovery and development. *Curr. Opin. Drug Discov. Dev.*, *7*, 126–136.

Nassar, A.E. Kamel, A.M., Clarimont, C. (2004). Improving the decision-making process in structural modification of drug candidates: reducing toxicity. *Drug Discov. Today*, *9* (24), 1055–1064.

Nassar, A.E., Lopez-Anaya, A. (2004). Stagegies for dealing with reactive intermediates in drug discovery and development. *Curr. Opin. Drug Discov. Devel.*, *7* (1), 126–136.

Nikulina, M.A. *et al.* (2000). Glutathione depletion inhibits IL-1 beta-stimulated nitric oxide production by reducing inducible nitric oxide synthase gene expression. *Cytokine*, *12* (9), 1391–1394.

O'Donovan, D.J. *et al.* (2000). Mitochondrial glutathione and oxidative stress: implications for pulmonary oxygen toxicity in premature infants. *Mol. Genet. Metab.*, *71* (1–2), 352–358.

Parrish, D.D. *et al.* (1997). Activation of CGS 12094 (prinomide metabolite) to 1,4-benzoquinone by myeloperoxidase: implications for human idiosyncratic agranulocytosis. *Fundam. Appl. Toxicol.*, *35*, 197–204.

Pirmohamed, M. *et al.* (1998). Adverse drug reactions. *BMJ*, *316* (7140), 1295–1298.

Prescott, L.F. *et al.* (1986). Effects of non-narcotic analgesics on the liver. *Drugs*, *32* (4), 129–147.

Pritsos, C.A., Constantinides, P.P., Tritton, T.R., Heimbrook, D.C., Sartorelli, A.C. (1985). Use of high-performance liquid chromatography to detect hydroxyl and superoxide radicals generated from mitomycin C. *Anal. Biochem.*, *150* (2), 294–299.

Richard, A.M. *et al.* (1998). Structure-based methods for predicting mutagenicity and carcinogenicity: are we there yet? *Mutat. Res.*, *400* (1–2), 493–507.

Ruscoe, J.E. *et al.* (1995). The effect of chemical substitution on the metabolic activation, metabolic detoxication, and pharmacological activity of amodiaquine in the mouse. *J. Pharmacol. Exp. Ther.*, *273*, 393–404.

Sabbaga, J. *et al.* (1993). Acute agranulocytosis after prolonged high-dose usage of intravenous dipyrone—a different mechanism of dipyrone toxicity? *Ann. Hematol.*, *66* (3), 153–155.

Salama, A. *et al.* (1983). The role of metabolite-specific antibodies in nomifensine-dependent immune hemolytic anemia. *N. Engl. J. Med.*, *313*, 469–474.

Sanders, S.P., Bassett, D.J., Harrison, S.J., Pearse, D., Zweier, J.L., Becker, P.M. (2000). Measurements of free radicals in isolated, ischemic lungs and lung mitochondria. *Lung*, *178* (2), 105–118.

Schonbeck, U. *et al.* (2004). Inflammation, immunity, and HMG-CoA reductase inhibitors: statins as antiinflammatory agents? *Circulation*, *109* (21 Suppl. 1), II18–II26.

Senturk, T. *et al.* (1997). Seizures and hepatotoxicity following sulphasalazine administration. *Rheumatol. Int.*, *17* (2), 75–77.

Shapiro, L.E. *et al.* (1997). Comparative safety of tetracycline, minocycline, and doxycycline. *Arch. Dermatol.*, *133*, 1224–1230.

Skjodt, N.M. *et al.* (1999). Clinical pharmacokinetics and pharmacodynamics of bromfenac. *Clin. Pharmacokinet.*, *36* (6), 399–408.

Soglia, J.R. *et al.* (2004). The development of a higher throughput reactive intermediate screening assay incorporating micro-bore liquid chromatography-micro-electrospray ionization-tandem mass spectrometry and glutathione ethyl ester as an in vitro conjugating agent. *J. Pharm. Biomed. Anal.*, *36* (1), 105–116.

Steghens, J.P. *et al.* (2003). Fast liquid chromatography-mass spectrometry glutathione measurement in whole blood: micromolar GSSG is a sample preparation artifact. *J. Chromatogr. B Analyt. Technol. Biomed. Life Sci.*, *798* (2), 343–349.

Stuart-Harris, R.C. *et al.* (1984). Aminoglutethimide in the treatment of advanced breast cancer. *Cancer Treat. Rep.*, *11*, 189–204.

Takeyama, N. *et al.* (1993). Oxidative damage to mitochondria is mediated by the Ca(2+)-dependent inner-membrane permeability transition. *Biochem. J.*, *294* (Pt 3), 719–725.

Tang, W. *et al.* (1996). Characterization of thiol-conjugated metabolites of 2-propylpent-4-enoic acid (4-ene VPA), a toxic metabolite of valproic acid, by electrospray tandem mass spectrometry. *J. Mass Spectrom.*, *8*, 926–936.

Tang, W. *et al.* (1997). A comparative investigation of 2-propyl-4-pentenoic acid (4-ene VPA) and its alpha-fluorinated analogue: phase II metabolism and pharmacokinetics. *Drug Metab. Dispos.*, *25* (2), 219–227.

Tettey, J.N. *et al.* (2001). Enzyme-induction dependent bioactivation of troglitazone and troglitazone quinone in vivo. *Chem. Res. Toxicol.*, *14* (8), 965–974.

Thompson, C.D. *et al.* (1996). Synthesis and in vitro reactivity of 3-carbamoyl-2-phenylpropionaldehyde and 2-phenylpropenal: putative reactive metabolites of felbamate. *Chem. Res. Toxicol.*, *9*, 1225–1229.

Tiller, P.R. *et al.* (2002). Liquid chromatography/tandem mass spectrometric quantification with metabolite screening as a strategy to enhance the early drug discovery process. *Rapid Commun. Mass Spectrom.*, *16*, 1225–1231.

Timmins, G.S., Liu, K.J., Bechara, E.J., Kotake, Y., Swartz, H.M. (1999). Trapping of free radicals with direct in vivo EPR detection: a comparison of 5,5-dimethyl-1-pyrroline-N-oxide and 5-diethoxyphosphoryl-5-methyl-1-pyrroline-N-oxide as spin traps for HO* and SO4. *Free Radic. Biol. Med.*, *27* (3–4), 329–333.

Uetrecht, J.P. *et al.* (1989). Idiosyncratic drug reactions: possible role of reactive metabolites generated by leukocytes. *Pharm. Res.*, *6* (4), 265–273.

Uetrecht, J.P. *et al.* (1994). Metabolism of vesnarinone by activated neutrophils: implications for vesnarinone-induced agranulocytosis. *J. Pharmacol. Exp. Ther.*, *270*, 865–872.

Uetrecht, J.P. *et al.* (1995). Oxidation of aminopyrine by hypochlorite to a reactive dication: possible implications for aminopyrine-induced agranulocytosis. *Chem. Res. Toxicol.*, *8* (2), 226–233.

Uetrecht, J.P. *et al.* (2002). Screening for the potential of a drug candidate to cause idiosyncratic drug reactions. *Drug Discov. Today*, *5*, 832–837.

Utili, R. *et al.* (1977). Dantrolene-associated hepatic injury. Incidence and character. *J. Gastroenterol.*, *72*, 610–616.

Ware, J.A. *et al.* (1998). Immunochemical detection and identification of protein adducts of diclofenac in the small intestine of rats: possible role in allergic reactions. *Chem. Res. Toxicol.*, *1998* (11), 164–171.

Woosley, R.L. *et al.* (1978). Effect of acetylator phenotype on the rate at which procainamide induces antinuclear antibodies and the lupus syndrome. *N. Engl. J. Med.*, *298*, 1157–1159.

Yin, J.H. *et al.* (2001). Inducible nitric oxide synthase neutralizes carbamoylating potential of 1,3-bis(2-chloroethyl)-1-nitrosourea in c6 glioma cells. *J. Pharmacol. Exp. Ther.*, *297* (1), 308–315.

Yunis, A.A. *et al.* (1980). Chloramphenicol toxicity: pathogenetic mechanisms and the role of the p-NO2 in aplastic anemia. *Clin. Toxicol.*, *17* (3), 359–373.

Zhao, Z. *et al.* (1996a). Bioactivation of 6,7-dimethyl-2,4-di-1-pyrrolidinyl-7H-pyrrolo[2,3-d] pyrimidine (U-89843) to reactive intermediates that bind covalently to macromolecules and produce genotoxicity. *Chem. Res. Toxicol.*, *9* (8), 1230–1239.

Zhao, Z. *et al.* (1996b). In vitro and in vivo biotransformation of 6,7-dimethyl-2,4-di-1-pyrrolidinyl-7H-pyrrolo[2,3-D]pyrimidine (U-89843) in the rat. *Drug Metab. Dispos.*, *24* (2), 187–198.

# Screening for Reactive Metabolites Using Genotoxicity Arrays and Enzyme/DNA Biocolloids

JAMES F. RUSLING, ELI G. HVASTKOVS, and JOHN B. SCHENKMAN

## 1 INTRODUCTION

Therapeutic drugs and chemicals used in our bodies or in our environment must be guaranteed safe to the people who are exposed to them. Extensive procedures for screening and predicting toxicity have been developed in the pharmaceutical industry. Nevertheless, drugs that are toxic to some subset of the population are not always identified by these procedures. About 30% of drug development failures are linked to toxicity issues, and unfortunately some of these do not come to light until clinical trials or even after the drug is introduced to the market. In addition, drug costs correlate with drug development failures (Caldwell and Yan, 2006). For these reasons, predicting drug toxicity at the earliest stages of development has become a critical goal (Nasser, Kamel and Clarimont, 2004).

A wide variety of strategies have been proposed for early toxicity prediction, including *in silico* methods along with a range of *in vitro* and *in vivo* biological approaches (Nasser, Kamel and Clarimont, 2004; Mayne, Ku and Kennedy, 2006; Kramer, Sagartz and Morris, 2007). Established methods use microsomes, cell cultures, or animal models and tend toward utilization of biochemical end points that are the result of complex responses to the drug (Kramer, Sagartz and Morris, 2007). These methods are typically combined into a panel of methodologies that in many cases provide a reasonably good prediction of human *in vivo* toxicity (Mayne, Ku and Kennedy, 2006). Nonetheless, unpredicted or idiosyncratic drug toxicity can result from interindividual variations in human biochemistry and resulting drug behavior in specific individuals that may be impossible to predict from batteries of toxicity tests and sometimes even from human clinical trials limited to subsets of the population that will eventually use the drug.

Certainly existing toxicity testing and prediction methods are important, viable, and useful. However, there is an unfilled niche for simple, cheap, high-throughput,

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

biochemically based screening assays that can be arranged into biosensor or biosensor array formats that are currently emerging (Stoll *et al.*, 2004; Rusling, Hvastkovs and Schenkman, 2007). Inexpensive toxicity devices of this sort could be used at very early stages of drug development and contribute important information about toxic drug candidates to aid in screening decisions. While the development of such sensing devices is in its infancy, this approach has the potential to help decrease drug development expenditures and, ultimately, costs of drugs to the public. In this chapter, we discuss fabrication and measurement principles for such devices, as well as recent examples of such approaches.

Reactions of DNA with drug molecules or their enzyme-generated metabolites can produce covalently linked nucleobase adducts which may initiate cancer (Jacoby, 1980; Friedberg, 2003; Scharer, 2003). These adducts most often occur on guanines and adenines in DNA, and serve as good biomarkers for cancer risk (Warren and Shields, 1997; Phillips *et al.*, 2000). They are also convenient biomarkers for detecting reactive metabolites and predicting drug toxicity (Tarun and Rusling, 2005a; Rusling, Hvastkovs and Schenkman, 2007). The reactive metabolites also modify proteins and other biomolecules, but DNA adducts are readily measured by a number of modern bioanalytical techniques including liquid chromatography–mass spectrometry (LC-MS). "Bioactivation" is the term used to denote generation of reactive metabolites by cytochrome P450s (cyt P450) and other metabolic enzymes. The process of forming reactive metabolites and causing DNA damage is usually called genotoxicity. Many substrates yield DNA-reactive metabolites, including styrene, benzo[*a*]pyrene (B[*a*]P), nitrosamines, napthylamines, and tamoxifen and other chemotherapeutic agents (Bond, 1989; Cavalieri *et al.*, 1990; Pauwels *et al.*, 1996; Umemoto *et al.*, 2001; Wang *et al.*, 2001). These reactive intermediates can also damage proteins and other biomolecules. In general, there is much more published information about the chemistry of reactive intermediates from nondrug molecules such as pollutants, since information for drug candidates that never came to market may be considered proprietary.

Cyt P450s are metabolic iron-heme enzymes (P—Fe) that catalyze the transfer of oxygen atoms to organic substrates (Schenkman *et al.*, 1993; Lippard and Berg, 1994; Ortiz de Montellano, 2005). Figure 1 depicts an accepted mechanistic model for cyt P450 enzyme catalysis that can lead to bioactivation. The resting state of the catalytic iron heme has water bound to the distal side of P—Fe(III) **(1).** This water freely exchanges with its environment and does not participate in oxygen transfer reactions (Lippard and Berg, 1994). This form of the enzyme binds the substrate RH and eliminates the distal water **(2).** Substrate is not bound directly to the iron heme of P—Fe(III), but sits above it in a hydrophobic pocket within the protein. Next, **2** is reduced by one electron to the P—Fe(II) state by an NADPH-dependent reductase to give **3**. P—Fe(II) **3** then binds dioxygen at the distal site to form a ferrous-dioxygen or ferric superoxy complex **(4)** (Guengerich, Bell and Okazaki, 1995; Guengerich, 2001; Ortiz de Montellano and De Voss, 2002). This P—Fe(II)—$O_2$ complex is converted to a P—Fe(III)—OOH complex via a one-electron reduction to give **5**, then protonated to **6**. The electron to generate **5** comes from an NADPH-dependent reductase in most cases. P—Fe(III)—OOH **(6)** may also be generated by addition of hydrogen peroxide or organic peroxides in a reversible process called the peroxide shunt. Protonation of **6** and dioxygen bond cleavage leads to formation of the highly reactive heme-iron(IV)-oxo radical cation {(P•+)Fe(IV)-oxo, **7**}, which transfers oxygen to the bound substrate to generate the product (ROH). The product dissociates from the enzyme and the distal site is reoccupied by water for another catalytic cycle. In addition, exposure of ferrous form **3** to

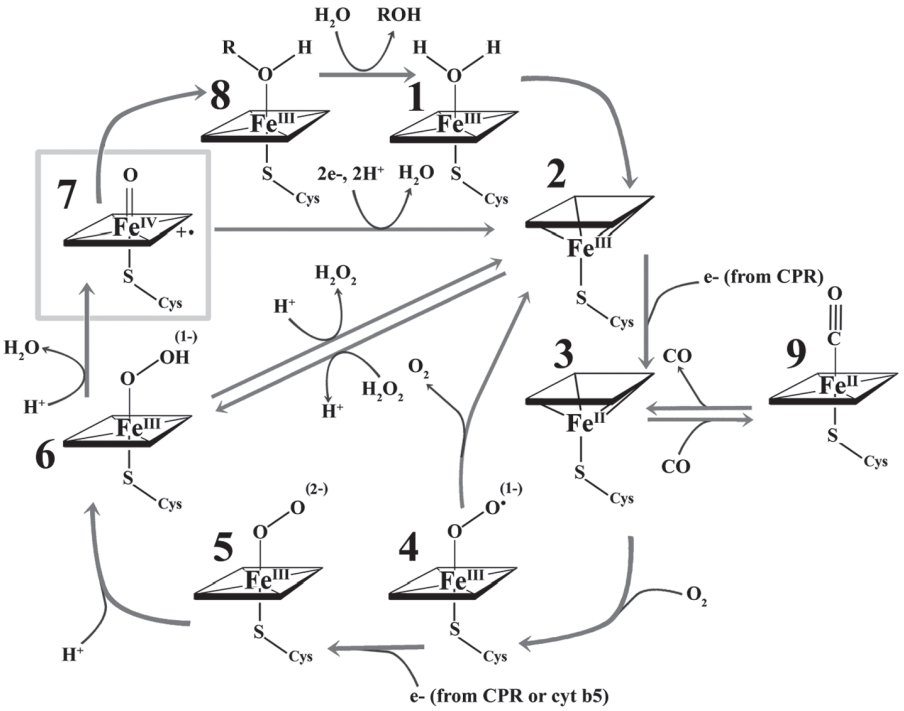


**Figure 1** Pathway for cyt P450-catalyzed metabolic reactions.

carbon monoxide produces the P—Fe(II)—CO complex **9** that absorbs light at ~450 nm and gives cyt P450s their characteristic name.

A simple example of bioactivation involves the conversion of styrene to styrene oxide by cyt P450 enzymes. In this case, styrene oxide is the only enzyme-generated metabolite. We use this example to show how the natural bioactivation process can be mimicked to produce and detect DNA adducts.

DNA damage can be used as an end point for formation of the reactive metabolite styrene oxide. In Fig. 2, we see that if styrene is converted to styrene oxide by a cyt P450 enzyme in the presence of DNA, DNA will be damaged, mainly by forming covalent adducts with guanine bases (Tarun and Rusling, 2005a; Rusling, Hvastkovs and Schenkman, 2007). Further, we show how thin films of metabolic enzymes and DNA grown on arrays or colloidal particles can facilitate detection of these damage events, rapidly and nonspecifically by novel array technology, or more specifically by capillary LC-MS/MS (capLC-MS).

In the next section, we describe how these concepts can be utilized to make electrochemiluminescent arrays for drug screening, specifically addressing formation of reactive intermediates, genotoxicity, and enzyme inhibition. In Section 3, we show how the same concepts can be used for the rapid, sensitive identification and quantitation of major reactive intermediates and their nucleobase adducts by capLC-MS. In Section 4, we survey alternative emerging technologies including optical protein microarrays that can be applied to drug toxicity screening.

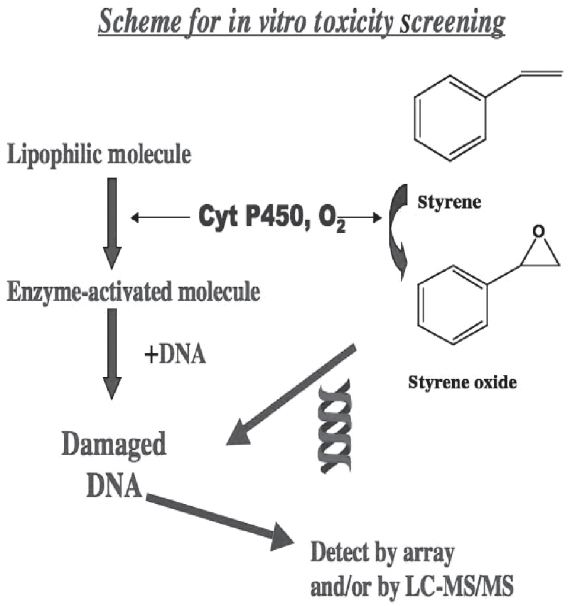


**Figure 2** Illustration of scheme used for toxicity screening from the formation of reactive intermediates using DNA damage as a measured end point.

## 2 ELECTROCHEMICAL AND ELECTROCHEMILUMINESCENT ARRAYS

A major research effort in our laboratory has been directed toward developing biosensor arrays for genotoxicity screening (Rusling, 2004, 2005; Rusling *et al.*, 2007; Rusling, Hvastkovs and Schenkman, 2007). These arrays all utilize a versatile layer-by-layer film fabrication technique (Lvov, 2000, 2001; Ariga, Hill and Ji, 2007; Zhang, Chen and Zhang, 2007) to grow 20- to 40-nm-thick films containing metabolic enzymes and DNA (Fig. 3). These films facilitate the two-step process illustrated in Fig. 2. That is, the enzyme reaction produces metabolites that may or may not damage DNA, and the array measurement step detects the relative degree of DNA damage.

The enzyme reaction is the first step of array operation. It was established long ago that metabolic specificity is a property of the terminal oxidase, cyt P450 (Rahimtula *et al.*, 1978; Werringloer, Kawano, Estabrook, 1980; White, Sligar, Coon, 1980; Adams and Adams, 1992). Thus, cyt P450 enzymes in the films can be activated by small concentrations (≤1 mM) of hydrogen peroxide, in the reverse of the well-known hydrogen peroxide shunt (see Fig. 1), to give identical products as when the natural oxidoreductase system is used. A few reports documented cases of altered product distributions when cyt P450s were activated using peroxides as opposed to the natural electron donor/reductase cycle (Bichara *et al.*, 1996; Kupfer *et al.*, 2001; Ortiz de Montellano, 2005). However, in all cases examined in our laboratory, hydrogen peroxide activation gives the same metabolites as natural cyt P450 activation by NADPH and cyt P450 reductase. Thus, peroxide activation is appropriate and convenient for use in these sensors. In addition, we shall see that these arrays can also be made using liver micro-

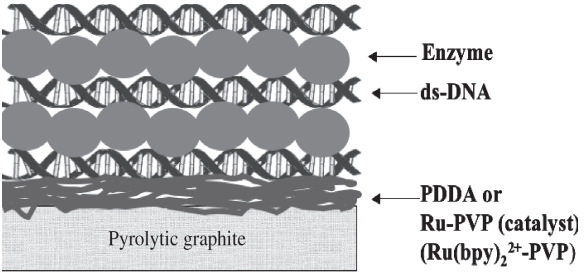


**Figure 3** Idealized representation of enzyme/DNA films that can be made on any solid surface. The enzyme reaction produces metabolites in close proximity to DNA. In arrays, resulting damage to DNA from the metabolites is detected by voltammetry or electrochemiluminescence.

$$\mathrm{ClRu^{II}\text{-}PVP} \leftrightarrow \mathrm{ClRu^{III}\text{-}PVP} + e^- \text{(at electrode)} \quad (1)$$

$$\mathrm{ClRu^{III}\text{-}PVP} + \mathrm{G} \rightarrow \mathrm{ClRu^{II}\text{-}PVP} + \mathrm{G}^{\bullet} + \mathrm{H}^+ \quad (2)$$

**Scheme 1** Catalytic electrochemical pathway for detection of DNA damage using ClRu-PVP.

somes as the source of metabolic enzymes, and activated by NADPH and cyt P450 reductase as in living systems.

The enzymes in the "nanoreactor spots" on the arrays synthesize reactive metabolites near large concentrations of DNA in each film spot. The rate of DNA damage from metabolite–nucleobase adduct formation is a measure of relative genotoxicity, and can be detected by voltammetric (Wang *et al.*, 2005) or electrochemiluminescent (Hvastkovs *et al.*, 2007) measurement steps.

Arrays for genotoxicity screening arose from of our research on single electrode biosensors that established the optimized parameters and operating conditions (Rusling, 2004, 2005; Rusling, Hvastkovs and Schenkman, 2007). The enzyme reaction is the first step of sensor operation. We ascertained that pH ~5.5 was optimum for DNA damage (Rusling, 2004, 2005), and that enzyme/polyion films 20–40 nm thick eliminate performance limitations from mass transport of reactant entering the film and product leaving (Munge *et al.*, 2003). The second step in the assay is detection of DNA damage, and we found that electrochemical detection with very good S/N can be done with square wave voltammetry (SWV) for DNA oxidation using a soluble electrochemical catalyst, $\mathrm{Ru(bpy)_3^{2+}}$ (Zhou *et al.*, 2003; Wang *et al.*, 2005). Alternatively, a catalytic Ru-polyvinylpyridine polymer [$\mathrm{Ru(bpy)_2^{2+}}$Cl-PVP or ClRu-PVP) can be incorporated within the film for "reagentless" sensing (Mugweru, Yang and Rusling, 2004). The signal in the measurement step after the enzyme reaction results predominantly from catalytic electrochemical oxidation of guanines in DNA (Scheme 1). Here, the catalyst active sites have better access to guanines in the partly unfolded, damaged DNA than in intact ds-DNA. Improved access of catalyst to the guanines provides faster rates of the reaction (Eq. 2) in damaged DNA than in undamaged ds-DNA and consequently larger signals (Rusling, 2004, 2005; Rusling *et al.*, 2007).

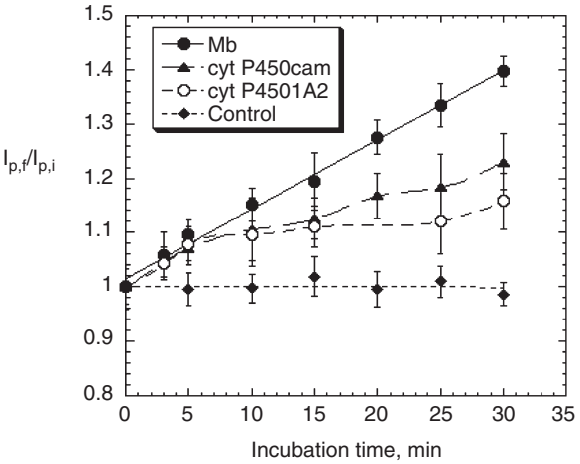


**Figure 4** Influence of incubation time with 50-μM benzo[*a*]pyrene and 1-mM $H_2O_2$ on the final/initial peak current ratios from SWV of PDDA/DNA/(enzyme/DNA)$_2$ films in an eight-electrode array. Control is PDDA/DNA/(Mb/DNA)$_2$ film in 50-μM benzo[*a*]pyrene alone (four replicates for Mb films, three each for cyt P450 films). Amounts of proteins in the films in nmol cm$^{-2}$ were 0.26 for Mb, 0.060 for cyt P450cam, and 0.054 for cyt P540 1A2. Adapted with permission from Wang *et al.* (2005). Copyright 2005 American Chemical Society.

Prototype genotoxicity screening sensors in the format described combined enzyme bioactivation leading to potential DNA damage, mimicking events in the human liver. Direct measurement of formation rates of altered nucleobases by LC-MS from similar films that had been reacted with damage agents and hydrolyzed confirmed that the sensors actually detect DNA damage (Tarun and Rusling, 2005a). Sensor and LC-MS results also correlated well with animal genotoxicity estimated by TDL$_o$, that is, the lowest dose producing carcinogenicity, in mice and rats (Tarun and Rusling, 2005a; Yang, Wang and Rusling, 2005).

Inclusion of several cyt P450s in electrochemical arrays facilitates analysis of the relative ability of the enzymes to generate genotoxic intermediates. Our first such array had eight electrodes and included cyt P450$_{cam}$, human cyt P450 1A2, and myoglobin (Mb) (Wang *et al.*, 2005). Mb is a good control in these arrays as it catalyzes some of the reactions of cyt P450s, but usually much less efficiently. In an array with cyt P450s, it should be the least active enzyme or have no activity at all. B[*a*]P was used as a test substrate for these arrays because of its well-understood metabolism resulting in several DNA-reactive metabolites from cyt P450-catalyzed oxidations (Rogan *et al.*, 1993; Todorovic *et al.*, 1997; Neilson, 1998). DNA damage was detected by increases in SWV peak ratios using soluble $Ru(bpy)_3^{2+}$ as the catalyst. Figure 4 shows the influence of enzyme incubation time on final/initial sensor peak ratio from such an array. Initial slopes of these plots divided by the amount of each enzyme in the array spots give relative turnover rates in {(nmol enzyme)$^{-1}$ min$^{-1}$)}, which were 3.0 for cyt P450$_{cam}$, 3.5 for cyt P450 1A2, and 0.9 for Mb. Cyt P450$_{cam}$ and cyt P450 1A2 showed threefold higher activity for bioactivation of B[*a*]P for DNA damage than the control enzyme Mb.

The ability to detect metabolite-based DNA damage simultaneously for several enzymes represents a promising approach to identify and characterize genotoxicity

pathways. However, the electrochemical arrays described require individual, electronically addressed electrodes and reproducible microelectronic chip manufacture. While this is certainly possible, an instrumentally more convenient approach is based on optical detection of electrochemiluminescence (ECL) from damaged DNA using a similar catalytic detection scheme. This approach requires no special electronics, and detection can been done with a simple charge-coupled device (CCD) camera.

Using ECL, application of a voltage to a polymer in a simple array can produce light. In the absence of DNA, when oxidation catalyst $Ru(bpy)_3^{2+}$ is oxidized to $Ru(bpy)_3^{3+}$, ECL has been generated by using a sacrificial reductant, often tripropylamine or oxalate. This produces a photoexcited $[Ru(bpy)_3^{2+}]^*$ by a pathway involving reaction of a radical form of the reductant with electrochemically generated $Ru(bpy)_3^{3+}$ (Ege, Becker and Bard, 1984; Kenten *et al.*, 1991; Xu *et al.*, 1994). Alternatively, $Ru(bpy)_3^{+}$ is formed by reduction of $Ru(bpy)_3^{2+}$ by the radical, followed by combination of $Ru^{I}$ and $Ru^{III}$ complexes to give $[Ru(bpy)_3^{2+}]^*$ (Rubinstein and Bard, 1980), which emits light at 610 nm. $Ru(bpy)_3^{2+}$-labeled DNA using a sacrificial reductant to produce ECL has been developed as the basis of a very sensitive method to detect oligonucleotide hybridization (Leland and Powell, 1990; Blackburn *et al.*, 1991; Xu and Bard, 1995).

The polymer $[Ru(bpy)_2(PVP)_{10}]^{2+}$ (Ru-PVP, PVP = polyvinylpyridine) shown in Fig. 5 can be used for "reagentless" DNA sensors since it can be electrochemically activated to react with DNA in films and generate ECL directly (Dennany, Forster and Rusling, 2003). That is, DNA itself is the sacrificial reductant, removing the need for an external sacrificial reductant in DNA detection. ECL can be measured from a single sensor by positioning an optical fiber under an electrode coated with a Ru-PVP/enzyme/DNA film, and directing the light to a photomultiplier tube. This type of sensor allows the simultaneous recording of voltammetric and ECL signals (Dennany, Forster and Rusling, 2003; So *et al.*, 2007). While these sensors are excellent for exploratory studies, high-throughput analysis requires an array format.

The chemistry of the detection process is summarized in Scheme 2. ECL is generated from guanine moieties present on DNA strands in the films (Dennany, Forster and Rusling, 2003). Electrochemical oxidation generates $Ru^{III}$-PVP, and guanine radicals are initially formed (Eq. 3) by catalytic oxidation of guanines (Eq. 4). These guanine radicals can react with the metallopolymer (Eq. 5) to produce electronically excited $Ru^{II}$* sites in the film (Eq. 6). Alternatively, $Ru^{I}$-PVP can be formed from reaction of $Ru^{II}$-PVP with $G^{\bullet}$, and can then react with $Ru^{III}$-PVP to produce the excited $Ru^{II}$* state to give ECL. As in purely electrochemical detection of DNA damage, catalyst active sites gain better access to guanines in partly unfolded, damaged DNA than in intact ds-DNA, thus increasing the rate of the catalytic process and consequently the output of ECL light.

ECL is well suited to arrays without individually addressable electrodes. An ECL array can be spotted directly on a conductive plate, for example, a single pyrolytic graphite (PG) block electrode. We have developed convenient, prototype high-throughput toxicity screening arrays based on multiple spots of DNA, enzyme, and the Ru-PVP polymer (Fig. 5). Up to 50 individual spots, each an LbL film containing DNA, enzyme, and Ru-PVP, can be manually micropipetted onto the 1 × 1 in. array plate. We are also evaluating automated spotting devices that can be used to deposit hundreds of spots on similar arrays.

As in the single probe sensors, the enzyme reaction is run first on this array, which then is washed and placed into an electrochemical cell housed in a dark box with a CCD camera (Fig. 6) (Hvastkovs *et al.*, 2007). Upon application of 1.25 V versus

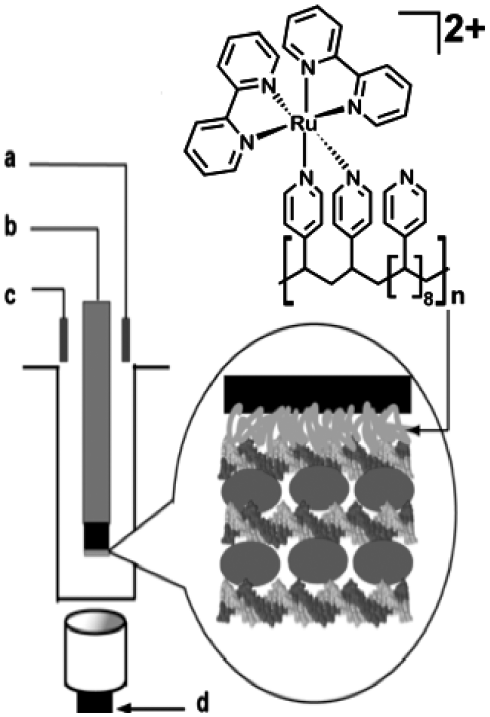


**Figure 5** Sensor system for simultaneous ECL and voltammetry detection. The reference electrode **(a)**, working electrode **(b),** and Pt counter electrode **(c)** are located in a glass cell with an extended cylindrical glass base. A fiber optic cable **(d)** is positioned on the outside of the cell directly under the sensor surface and leads to a photomultiplier tube. Ru-PVP (structure shown), DNA (double helix), and enzyme (ovals) layers forming the active sensor film are shown on the right. Only one layer of RuPVP is shown on the sensor, although several can be used to enhance S/N.

$$Ru^{II}\text{-PVP} \leftrightarrow Ru^{III}\text{-PVP} + e^{-}\,(\text{at electrode}) \quad (3)$$

$$Ru^{III}\text{-PVP} + G \rightarrow Ru^{II}\text{-PVP} + G^{\bullet} + H^{+} \quad (4)$$

$$G^{\bullet} + Ru^{III}\text{-PVP} \rightarrow G_{2ox} + Ru^{II*}\text{-PVP} \quad (5)$$

$$Ru^{II*}\text{-PVP} \rightarrow Ru^{II}\text{-PVP} + h\nu\,(610\ \text{nm}) \quad (6)$$

**Scheme 2** Electrocatalytic oxidation of guanine employing ECL generating Ru-PVP polymer.

SCE, the electrochemical oxidation of Ru-PVP initiates the measurement chemistry in Scheme 2 and generates light from each spot, which is measured with the CCD camera over an integration time of ~20 seconds. As in all catalytic DNA detection schemes, larger signals are obtained from damaged DNA because of better accessibility of the guanines to the catalyst as the ds-DNA unravels.

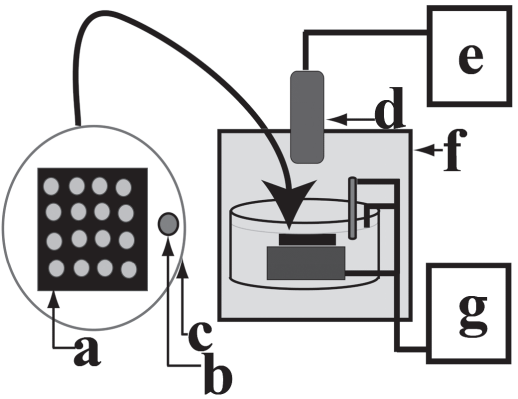


**Figure 6** ECL arrays for toxicity screening: **(a)** array, **(b)** reference electrode, **(c)** counter electrode, **(d)** CCD camera, **(e)** computer, **(f)** dark box housing, and **(g)** potentiostat to apply voltage for ECL generation. Adapted with permission from Hvastkovs *et al.* (2007). Copyright 2007 American Chemical Society.

ECL arrays can be configured to measure the time course of reactions catalyzed by a single enzyme, or can contain a collection of enzymes for simultaneous comparison of reactive intermediate formation kinetics. Results can be reorganized and presented in a number of ways by computer software. Figure 7 shows an ECL array with 49 RuPVP/DNA/enzyme spots designed to study rates of DNA damage for the bioactivation of B[*a*]P by the human enzyme cyt P450 1B1. Individual array spots were subjected to 0.5-mM $H_2O_2$ and 100-μM B[*a*]P for various times marked on the figure. We can see that light intensity increased with the time of the enzyme reaction. Control spots were exposed to B[*a*]P alone or $H_2O_2$ alone, and showed little change in ECL. These controls are necessary to confirm that $H_2O_2$ activates the enzymes but does not damage DNA. Figure 7b illustrates a useful way of expressing the array data, as a plot of the final/initial ECL ratio.

Figure 8 illustrates oxidation of B[*a*]P with five enzymes in a single array (Hvastkovs *et al.*, 2007). The data have been rearranged by computer software so that representative spots for a single enzyme lie in a specific row. Relative rates of DNA damage were estimated simultaneously in this way for these five enzymes in ~1 minute of enzyme reaction time and 20 seconds of array development time. The slopes of the linear ECL increases correlate with DNA damage rates, and are compared in the graph on the right of Fig. 8. Bioactivation producing DNA damage taken as the initial slopes of these graphs was in the order cyt P4501B1 > cyt P4501A2 > cyt $P450_{cam}$ > cyt P4502E1 > Mb, the same as the order of relative metabolic activity of these enzymes toward B[*a*]P (Hvastkovs *et al.*, 2007). The slope of the linear ECL increase estimates the relative activity of different enzymes to produce reactive metabolites.

We have also developed and tested ECL arrays for *N*-nitrosamines (Krishnan *et al.*, 2007). Nitrosamines have been reported to be carcinogens in more than 30 species (Hecht, 1998). Cyt P450s, in particular CYP2E1, bioactivate *N*-nitrosamines to reactive intermediates, mainly by α-hydroxylation. *N*-Nitrosopyrrolidine (NPYR) (Preussmann and Stewart, 1984) was used as the test compound. Adducts form from reaction of DNA with NPYR metabolites (Scheme 3) (Wang *et al.*, 2001).

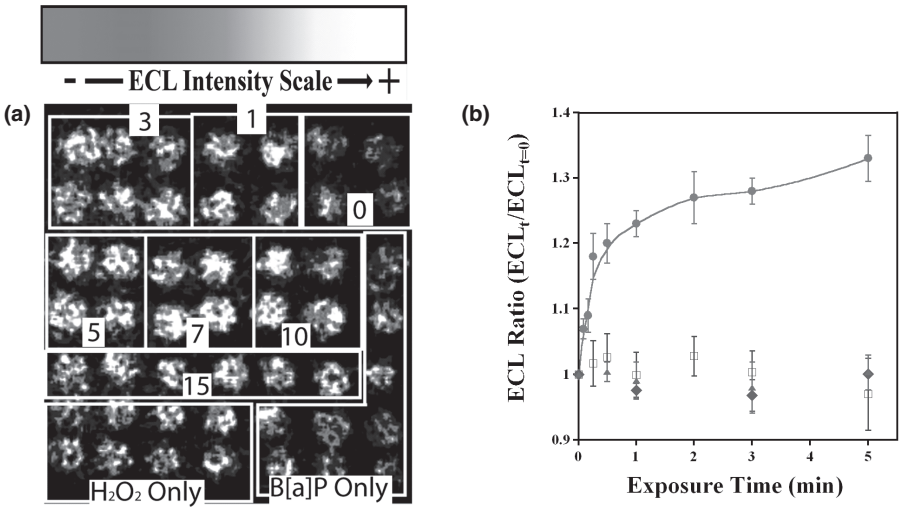


**Figure 7** **(a)** CCD image of ECL array with 49 individual RuPVP/DNA/enzyme spots all containing cyt P450 1B1. Boxes denote spots that were exposed to 0.5-mM $H_2O_2$ + 100-μM B[*a*]P for the labeled times (min). Controls (bottom) were exposed to B[*a*]P or $H_2O_2$ only for increasing amounts of time from 1 to 7 minutes as viewed from right to left (not marked for clarity). **(b)** ECL ratio plot demonstrating the increase in ECL intensity versus time of enzyme reaction. Controls show ECL response versus time exposed to B[*a*]P only (squares), $H_2O_2$ only (diamonds), and 0.5-mM $H_2O_2$ + 100-μM B[*a*]P + 30 μM of inhibitor αNF (triangles). Adapted with permission from Hvastkovs *et al.* (2007). Copyright 2007 American Chemical Society.

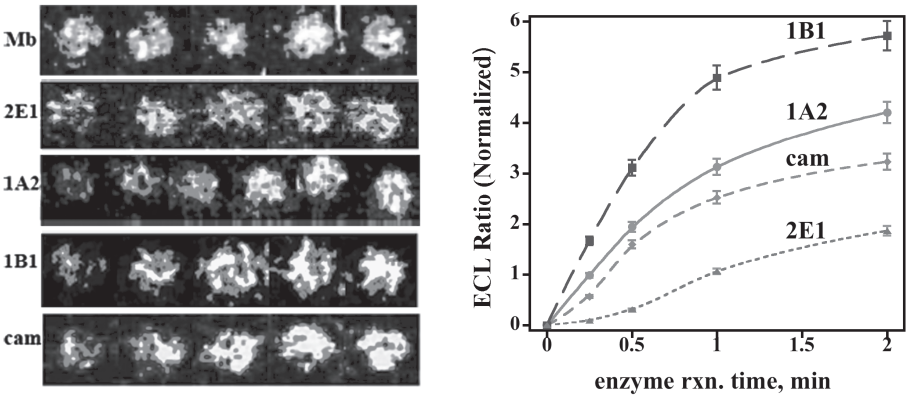


**Figure 8** ECL array results for enzyme reactions with 100-μM benzo[*a*]pyrene + 0.5-mM $H_2O_2$ **(a)** reconstructed images for reaction times of 0, 1, 3, 5, and 7 minutes for cyt P450 enzymes and myoglobin on the same array. CCD images are recolorized with the brighter spots showing more DNA damage. **(b)** ECL initial/final ratios normalized for the amount of enzyme in each spot. Adapted with permission from Hvastkovs *et al.* (2007). Copyright 2007 American Chemical Society.

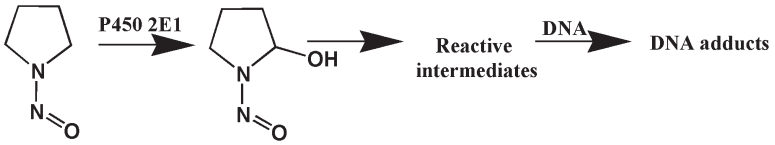


**Scheme 3** Reaction pathway for the cyt P450 mediated bioactivation of NPYR.

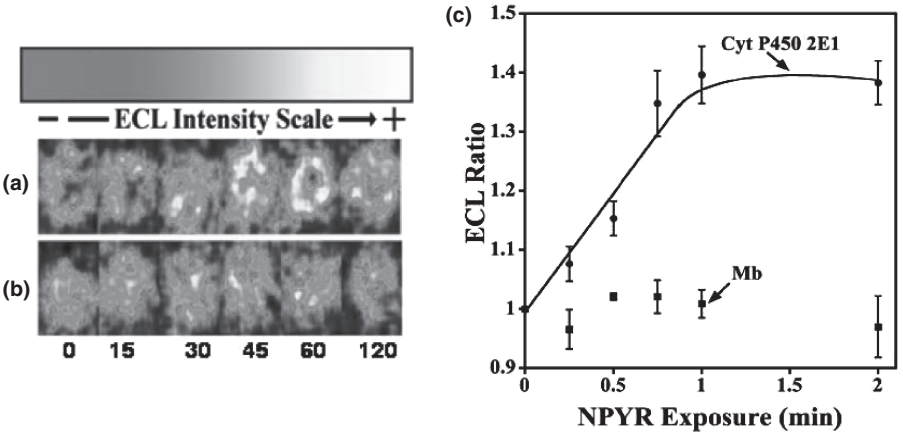


**Figure 9** ECL array data after exposure to NPYR: **(a)** digitally reconstructed image demonstrating CCD-captured ECL emitted from RuPVP/DNA/cyt P450 2E1 array and **(b)** RuPVP/DNA/Mb array exposed to 150-μM NPYR + 1-mM $H_2O_2$ for the denoted amounts of time (s). **(c)** Ratio plot demonstrating the ECL signal increase from films containing cyt P450 2E1 or Mb. Adapted with permission from Krishnan *et al.* (2007). Copyright 2007 Royal Society of Chemistry.

Reconstructed ECL arrays (Krishnan *et al.*, 2007) show the ECL response when the DNA/cyt P450 2E1 and DNA/Mb films are exposed to increasing times of damage solution (Fig. 9a,b). Each spot contains enzyme, DNA, and RuPVP. The increase in light intensity with time of the Cyt P450 2E1 reaction indicates increased DNA damage via generation of a reactive intermediate. No change in ECL occurred for RuPVP/DNA/Mb spots, showing that Mb does not bioactivate NPYR. Figure 9c shows the ECL data in the ratio plot format, and is another way to present the result that cyt P450 2E1 bioactivates NPYR, but Mb does not.

The future vision for this ECL toxicity arrays includes screening for the formation of reactive intermediates (Krishnan *et al.*, 2007; Rusling, Hvastkovs and Schenkman, 2007; So *et al.*, 2007), identifying the enzymes producing them (Wang *et al.*, 2005; Hvastkovs *et al.*, 2007), and applications in drug–drug interactions involving enzyme inhibition (Mugweru and Rusling, 2006). For this type of array, we will need also to utilize phase II and multienzyme metabolic reaction in the arrays. This development is currently under way in our laboratories.

We can also use liver microsomes as enzyme sources in the arrays. Figure 10 shows a digitally reconstructed image of an array in which each spot contains DNA, rat liver

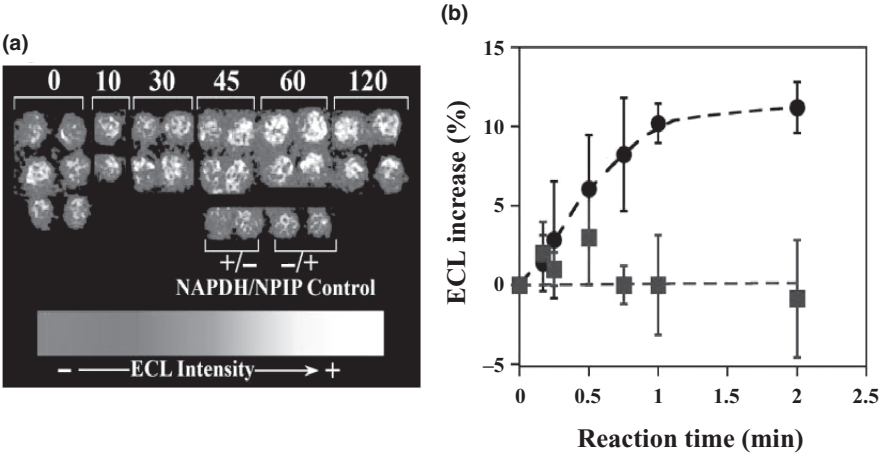


**Figure 10** Data from ECL arrays using rat liver microsomes (RLMs) as the enzyme source. **(a)** Reconstructed image showing the results of an array exposed to NADPH and 1-mM NPIP for increasing times (in seconds at top). Spot had film architecture $(DNA/RuPVP)_2DNA(RuPVP/RLM/DNA)_3$. Controls (bottom) were exposed to only NPIP in buffer or NADPH alone. **(b)** ECL percentage change versus reaction time (●) demonstrating from array runs. Controls (■) correspond to spots incubated in NPIP (no NADPH) under the same conditions.

microsomes (RLMs), and RuPVP. The spots were exposed to an NADPH-generating system containing 1-mM *N*-nitrosopiperidine (NPIP) for increasing amounts of time at 37 °C. NPIP is a known genotoxic nitrosamine compound that induces esophageal and liver tumors after exposure (Young-Scaime *et al.*, 1995). The use of liver microsomes allows the cyt P450 enzymes in the array to be activated by the natural *in vivo* route (Scheme 1) through electron transfer from NADPH via with cyt P450 reductase (CPR) (Ortiz de Montellano, 2005). NPIP is metabolically activated by the cyt P450 enzymes present in the RLM and diffuses through the films to react with nucleophilic DNA. Figure 10 shows the increase in ECL intensity with reaction time. Figure 10 also shows control experiments where spots were incubated with only NPIP in buffer (no NADPH) or the NADPH system alone. Controls show negligible ECL increases. The relative turnover for NPIP was found by dividing the initial slope of the ECL plot by the amount of RLM present in the film estimated by a quartz crystal microbalance (QCM). For this example, turnover rate was $660\,min^{-1}$ (mg RLM)$^{-1}$.

## 3 DNA/ENZYME BIOCOLLOIDS FOR LC-MS TOXICITY SCREENING

Progress in techniques like electrospray ionization (ESI) and matrix-assisted laser desorption ionization (MALDI) has made MS unprecedented in popularity for bioanalysis. ESI has facilitated coupling of capillary liquid chromatography (capLC) to MS (Vanhoutte *et al.*, 1997; Appruzzese and Vouros, 1999; Gangl, Turesky and Vouros, 2001), minimizing the need to isolate and purify DNA adduct samples and facilitating small sample size and ultrahigh sensitivity in detection of nucleobase adducts (Tarun

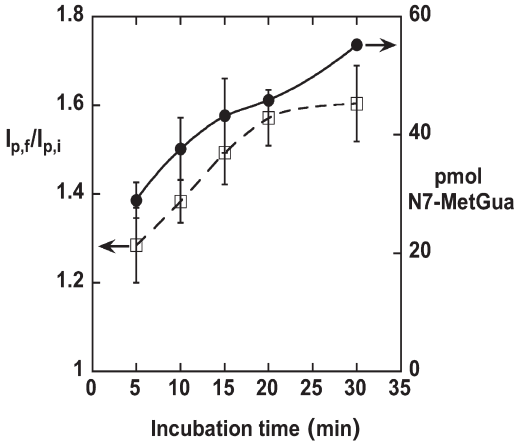


**Figure 11** Comparisons of electrochemical toxicity sensor with nucleobase adduct formation rate. Peak current ratio $I_{p,f}/I_{p,i}$ (final/initial) for sensors consisting of $(PDDA/DNA)_2$ films on graphite electrode-based (□), and pmol *N*7-methylguanine found by LC-MS (•) of hydrolyzed DNA from similar films on carbon cloth. Both assays after incubation of DNA films with 2-mM MMS. Adapted with permission from Tarun and Rusling (2005b). Copyright 2005 American Chemical Society.

and Rusling, 2005a). LC-MS methods are among the most specific and versatile methods to detect DNA adducts, and can provide chemical structures and rates of formation of specific nucleobase adducts.

In the early days of developing toxicity sensors, we used LC-MS to measure DNA adducts to confirm that the sensors actually detect DNA damage (Zhou *et al.*, 2003; Tarun and Rusling, 2005b). We measured rates of altered nucleobases by LC-MS using DNA films on carbon cloth that had been reacted with damage agents and hydrolyzed (Yang, Wang and Rusling, 2005; Tarun and Rusling, 2006). In an early example, response of electrochemical toxicity sensors gave good correlations with the rate of formation of *N*-7-methylguanine found by LC-MS after incubation with methyl methane sulfonate (MMS) (Fig. 11) and epoxides. Such correlations provided direct evidence that the slopes of sensor response versus enzyme reaction time measure relative rates of DNA damage. Sensor and LC-MS results also correlated well with animal genotoxicity estimated by carcinogenicity index $TDL_o$ in rodents (Tarun and Rusling, 2005b, 2006).

Preliminary results showed that films of enzymes and DNA similar to those used in the arrays could provide nucleobase adducts for LC-MS detection after running the enzyme reaction. These early films for LC-MS were made on carbon cloth, but we later found that increased reaction rates resulting in more sensitive adduct detection could be effected when the films were made on 500-nm-diameter silica colloids. We shall see in the succeeding discussion that these enzyme/DNA biocolloids combined with neutral hydrolysis of the DNA, and capLC-MS with online sample preconcentration provides a very rapid and sensitive method for measuring and identifying reactive metabolites that may be involved in genotoxicity. Together, ECL arrays and capLC-MS analyses provide complementary tools for toxicity studies. The ECL arrays are rapid, inexpensive, high-throughput screening tools, while the capLC-MS approach is a bit more

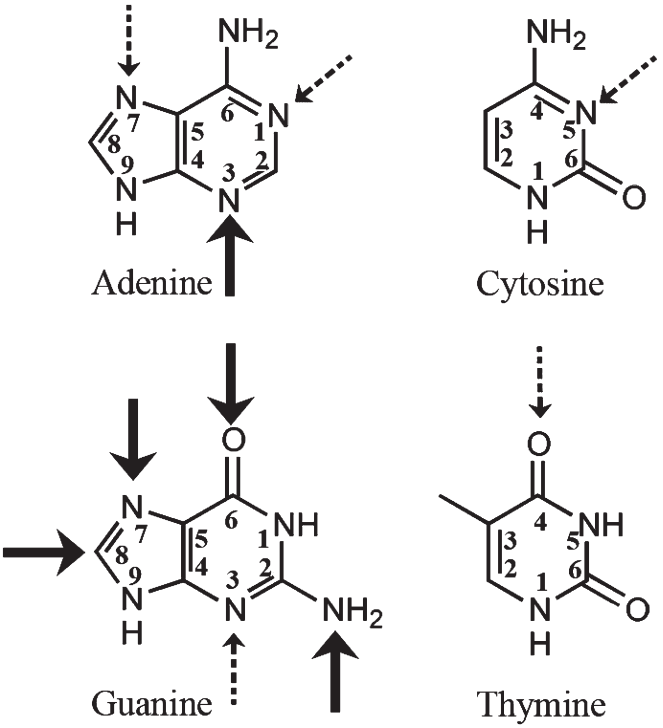


**Scheme 4** Sites of DNA alkylation subject to attach by reactive drug metabolites. Arrows show major (bold, solid arrow) and minor (broken arrow) alkylation sites.

time-consuming and costly, but provides a wealth of information on reactive intermediates, their nucleobase adducts, and the rates of formation of these products.

Before we describe specific DNA/enzyme biocolloids, we digress to summarize basic DNA chemistry necessary for understanding their operation. Many DNA adducts are formed by alkylating agents (Marnett and Burcham, 1993), and nitrogen and oxygen atoms on nucleobases are active sites (Scheme 4). Reactive metabolites are often good alkylating agents or electrophiles and react at one of these nucleophilic sites on the bases (Scheme 5). In practice, guanine is the most reactive base. An example of an *N*7-guanine adduct is shown in Scheme 6.

Scheme 5 also illustrates the useful sample preparation method of neutral thermal hydrolysis, which selectively ejects damaged DNA nucleobases. Alkylation at *N*7-guanine and *N*3-adenine makes the *N*-glycosidic bonds thermally labile for G and A derivatized at these positions (Zamenhof and Arikawa, 1966; Zoltewicz and Clark, 1972). Thus, simple heating of damaged DNA produces a sample enriched in *N*7-guanine- and *N*3-adenine adducts, with a much smaller fraction of native nucleobases than would be obtained by acid or enzymatic hydrolysis (Jacoby, 1980; Tarun and Rusling, 2005b). For this reason, neutral thermal hydrolysis is an excellent sample preparation method for LC-MS.

The use of nanoparticles for biocatalysis was pioneered by Lvov and Caruso (Caruso *et al.*, 1998; Caruso and Möhwald, 1999; Caruso and Schüler, 2000; Lvov and Caruso,

**Scheme 5** Mechanism of *N*7-alkylation of guanine and subsequent depurination of the damaged nucleobase that can be effected by neutral thermal hydrolysis. R = electrophilic reactive metabolite.

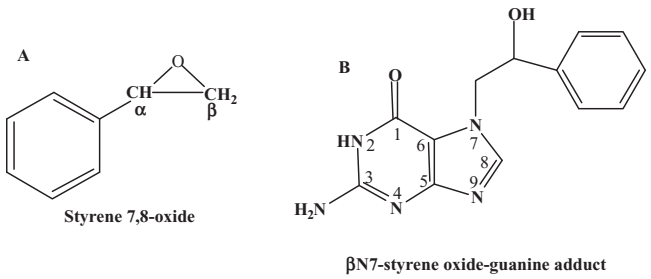


**Scheme 6** An example of an *N*7-guanine adduct formed from reaction of the styrene metabolite styrene 7,8-oxide with guanine moieties in DNA.

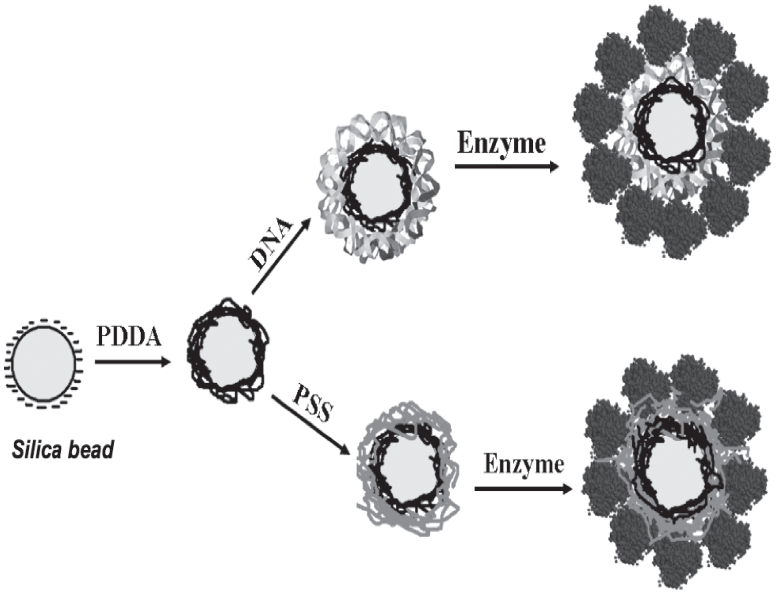


**Scheme 7** Conceptual illustration of DNA/enzyme film formation on silica microbeads. Layer-by-layer electrostatic assembly is employed to immobilize positively charged PDDA on the negatively charged microbeads, followed by sequential adsorption of negatively charged poly(styrene sulfonate) (PSS) or DNA, and a positively charged enzyme layer.

2001; Fang *et al.*, 2002; Shutava *et al.*, 2004). They used alternate electrostatic layer-by-layer (LbL) film assembly (Lvov, 2000, 2001; Ariga, Hill and Ji, 2007; Zhang, Chen and Zhang, 2007), the same method used for our sensor arrays, to form active enzyme films of glucose oxidase, horseradish peroxidase, and urease on spherical nanoparticles. Enzyme and DNA/enzyme *biocolloids* for reactive intermediate screening are illustrated in Scheme 7. Their tiny size offers a very large active surface area and the ability to employ tiny solution volumes to conserve enzyme and obtain more product per unit time for subsequent determination of major and minor products of both the enzyme and DNA reactions.

To demonstrate the biocolloid approach, we show results for styrene oxide and 4-(methylnitrosoamino)-1-(3-pyridyl)-butanone (NNK) oxidized using human cyt P450 2E1 (CYP 2E1) (Bajrami *et al.* , submitted). Cyt P450 2E1 is a prime catalyst in the metabolism of styrene to styrene oxide (Scheme 6) *in vivo* (Tanaka, Terada and Misawa, 2000). NNK is an *N*-nitroso procarcinogen in tobacco smoke that has been implicated in cancer risk (Scheme 8) (Hecht, 2003; Wang *et al.*, 2003). Styrene and NNK were oxidized with cyt P450 2E1 biocolloids activated by 1 mM $H_2O_2$. Styrene forms styrene oxide exclusively from the enzyme reaction, while metabolites of NNK undergo hydrolysis to 4-hydroxy-1-(3-pyridyl)-butanone (HPB) (Wang *et al.*, 2003) (Scheme 8). Figure 12 shows styrene oxide and HPB production as detected by GC for styrene oxide and LC-MS for HPB. The rate of production of both metabolites was initially similar, but over 5 minutes the rate of styrene oxide production was larger than that of HPB.

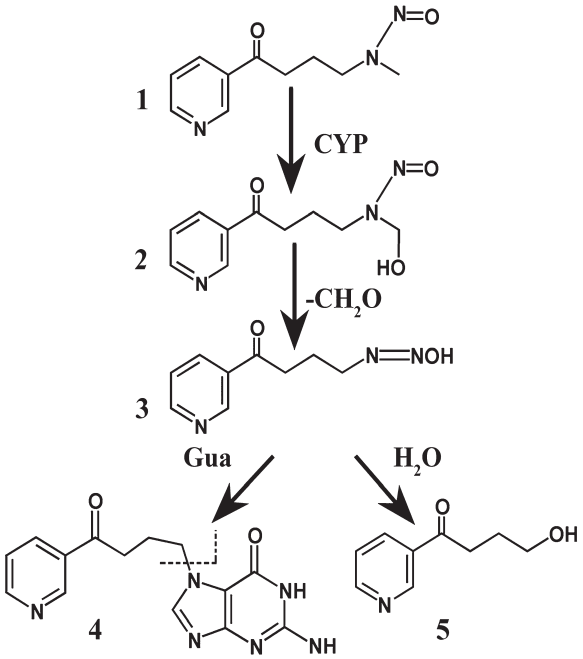


**Scheme 8** Pathway for cyt P450 (CYP)-catalyzed bioactivation of NNK **(1)** (Wang *et al.*, 2003), which undergoes α-hydroxylation to unstable intermediate **(2)** that loses formaldehyde to form 4-oxo-4-(3-pyridyl)-1-butanediazohydroxide **(3)**. Reaction of **3** with guanine produces several adducts including pyridyloxobutylation at *N*7 (**4**, 7-[4-oxo-4-(3-pyridyl)but-1-yl]Gua). Hydrolysis of **3** produces 4-hydroxy-1-(3-pyridyl)-1-butanone (HPB, **5**). The dashed line in **4** represents fragmentation in MS.

Figure 13 shows data for DNA adducts obtained from NNK metabolites made using DNA/cyt P450 2E1 biocolloids. Figure 13a is the capLC/MS-MS chromatogram and 13b is the SRM spectrum showing the product m/z peaks after reaction and neutral thermal hydrolysis. The m/z 299 → 152 is consistent with the previously reported pyridyloxobutylation of guanine via attachment at the most likely *N*7 location in Scheme 8 (Wang *et al.*, 2003), which shows an *N*7-pyridyloxobutylated guanine that would produce m/z 152 (Gua) upon fragmentation.

The multiple peaks in Fig. 13a are probably due to several positional isomers (Wang *et al.*, 2003), since monitoring at this m/z also detects adducts with linkages at positions other than *N*7 of guanine. The m/z 299 peaks for the hydrolysate are consistent with the formation of pyridyloxobutyl-guanine from the enzyme-catalyzed oxidation of NNK.

In this case, HPB is not the DNA-reactive metabolite. The precursor to HPB is a short-lived species that reacts with DNA (**3**, Scheme 8). Figure 13c shows that the integrated area of this peak increases linearly similar to HPB in Fig. 12. Thus, rate of formation of the specific metabolite produced directly mirrors the rate of formation of the major DNA adducts in the films.

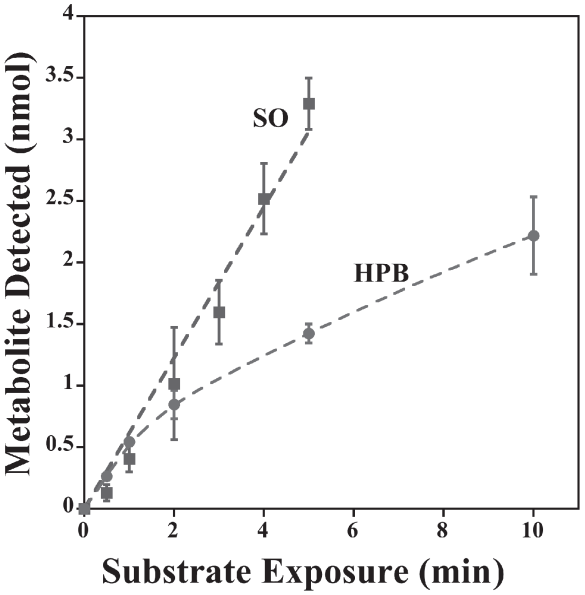


**Figure 12** Influence of reaction time on amount of styrene oxide (SO, ■) from 1% styrene and HPB (●) from 100-mM NNK found from reaction with PSS/cyt P450 2E1 colloids activated by 1-mM $H_2O_2$ in pH 5.5 buffer. Adapted with permission from Bajrami *et al.* (2008). Copyright 2007 American Chemical Society.

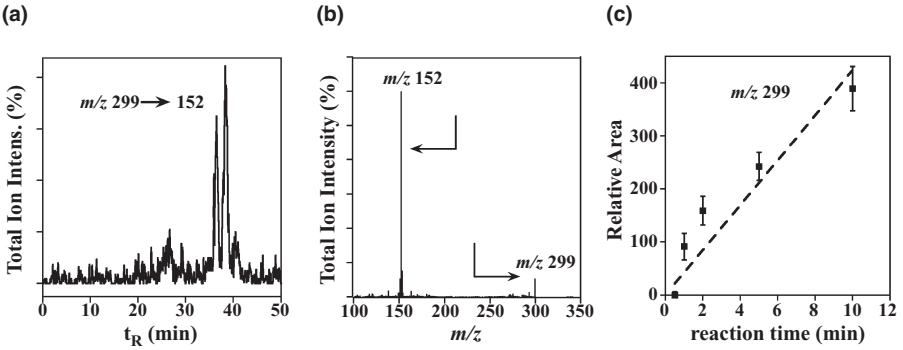


**Figure 13** CapLC/MS-MS results for NNK reaction catalyzed by DNA/cyt P450 2E1 microbeads in pH 5.5 buffer. **(a)** Chromatogram after a 15-minute reaction. **(b)** Mass spectrum analyzed by SRM (m/z 299 → 152) of same sample as in panel a. © Influence of reaction time on total integrated area of the peaks in panel a at $t_R = 37$. Adapted with permission from Bajrami *et al.* (2008). Copyright 2007 American Chemical Society.

These results show that DNA/enzyme biocolloids can be used to measure formation rates and confirm structures of both metabolites and DNA adducts. This technique can be used for substrates that do not produce DNA adducts as well as those that do, and can thus be used in other chemical and drug discovery studies such as enzyme inhibition. Finally, we have recently made biocolloids that employ liver microsomes

and microsomes genetically engineered to contain a single cyt P450 enzyme and its reductase, so that tedious purification of individual enzymes can be avoided. These microsome biocolloids contain cyt P450 reductase and can be activated with NADPH, exactly as in conventional biocatalysis with microsomes (Krishnan *et al.*, in preparation).

## 4 ALTERNATIVE ARRAYS AND OTHER NOVEL APPROACHES

The distillation of vast numbers of potential lead drug candidates involves initial screening of prospective hit molecules for possible toxicity that might be missed in later *in vivo* tests. Such *in vitro* screens include those for genotoxicity, hERG channel block, drug–drug interactions, and metabolite-mediated toxicity. In addition, these assays often include a standard cytotoxicity (cell destruction) assay (Kramer, Sagartz and Morris, 2007). These assays must be able to detect potential toxicity from thousands of potential candidates in a high-throughput and cost-efficient manner, which is not always realized. In addition, as significant monetary losses in pharmaceutical development occur upon late stage discovery of toxicity, preliminary *in vitro* screens must selectively and sensitively identify toxic candidates while keeping those deemed acceptable in the development pipeline (Nicholson *et al.*, 2002).

### 4.1 Traditional Cell-Based Toxicity Detection

Typically, initial *in vitro* toxicity assays are designed to monitor degrees of cytotoxicity arising from exposure to a compound of interest. Cytotoxicity is typically induced via some form of DNA damage, or genotoxicity (Kramer, Sagartz and Morris, 2007). The bacterial reverse mutation mutagenicity assay (Ames test) (Ames, Mcann and Yamasaki, 1975), the SOS/*umu* test (Oda *et al.*, 1985; Liebler and Guengerich, 2005), the micronucleus test (MNT) (Galloway *et al.*, 1994), the chromosomal aberration (CA) test, the mouse lymphoma assay (MLA) (Kramer, Sagartz and Morris, 2007), and the Comet assay (Singh *et al.*, 1998; Sasaki *et al.*, 2000) for chromosome breakage are currently employed as indirect genotoxicity screens for potential drug candidates and environmental contaminants. The most well known is the Ames test that assays the ability of a compound to induce frame or base pair shifts that facilitate the growth of a bacterial strain in the absence of certain amino acids (Ames, Mcann and Yamasaki, 1975). These assays predominantly measure cell growth, protein expression, or, in the case of the Comet assay, the electrophoresis gel tail length (Singh *et al.*, 1998), to gauge relative genotoxicity. The desirable quality of these tests is that predefined thresholds provide *yes*/*no* answers to facilitate developmental decisions. In cases generating a "yes" toxicity answer, the tested compound may be shelved or redesigned unless a certain amount of genotoxicity is tolerated, for instance in cancer drug development (Kramer, Sagartz and Morris, 2007). Despite their ubiquitous and required use, generally these tests have low specificity providing negative output for certain classes of compounds, such as aneugens (aneuploidy inducing), or give false positives upon exposure to non-genotoxic compounds (Hastwell *et al.*, 2006).

High-throughput modifications to these mutagenicity and clastogenicity (chromosome breakage) tests have been reported, typically employing microtiter plate formats (Kramer, Sagartz and Morris, 2007). For instance, the use of such a platform imparts a bit of high-throughput nature to the electrophoresis-based Comet (Kiskinis, Suter and Hartmann, 2002) and bacterial mutagenesis Ames II assays (Flukiger-Isler *et al.*, 2004).

Laser scanning cytometry allowed for more rapid analysis of stained erythrocytes in the MNT assay (Styles *et al.*, 2001). A similar detection approach was employed using bone marrow cells in an *in vitro* approach to the MNT assay (Shuga *et al.*, 2007). However, newer genotoxicity screens designed to improve upon the traditional Ames assays have yet to attract significant mainstream industrial attention due to high throughput or specificity limitations (Van Gompel *et al.*, 2005).

Employing a fusion plasmid coupling genes for green fluorescent protein (GFP) to growth arrest and DNA damage (GADD) protein, Hastwell *et al.* demonstrated higher specificity and comparable sensitivity in determining a host of genotoxic agents compared with the Ames test (Hastwell *et al.*, 2006). In addition, the GADD-GFP assay was able to determine the action of aneugens, which the Ames assay cannot. These benefits were hypothetically due to the eukaryotic nature of the assay and the importance of the p53-binding motif designed into the GADD protein. Due to the removal of DNA repair mechanisms in the bacterial strains for Ames assays, some relatively benign xenobiotics produced characteristic false-positive outputs. The GADD-GFP system accounted for these false positives presumably due the use of repair-oriented p53 competent cells in the system (Hastwell *et al.*, 2006).

### 4.2 Metabolite Screening

Another important aspect for *in vitro* toxicity screening is the consideration that many compounds administered *in vivo* are toxic upon metabolism (bioactivation) (Oda *et al.*, 1985; Kramer, Sagartz and Morris, 2007). A common drawback that exists in the typical mutagenicity/toxicity screens is that addition of exogenous liver homogenate (S9 fraction) is necessary to account for reactive metabolite genotoxicity. Indeed, throughput in the aforementioned GPDD-GFP test was hampered by this modification (Hastwell *et al.*, 2006). Although the addition of a liver fraction introduces a variety of enzymes involved in general metabolism, no information is garnered regarding enzyme specificity or mechanism. In an attempt to ascertain such information in toxicity screening, purified or microsomal preparations of cyt P450 enzymes in conjunction with cell lines have been incorporated. Arrays of cyt P450 3A4 and 2D6 sol-gel immobilized in contacting a MCF7 breast cancer cell monolayer were used to determine metabolite efficacy in promoting cytotoxicity. Scheme 9 demonstrates the metabolite enzyme toxicity analysis chip (MetaChip) technology, which involves suspending cyt P450 enzymes in a sol-gel mixture and clamping the sol-gel to a monolayer of cells. Cytoxan and Tegafur were metabolized by the immobilized cyt P450s, the respective reactive metabolites diffused into the proximal cancer cells, and cell viability was measured by fluorescence after cell staining (Lee *et al.*, 2005b).

Overexpression of cyt P450 3A4 in supersomes and liver cells has also been exploited to screen for metabolite toxicity. The toxicity of several drugs with known adverse drug reactions (ADRs) toward the HepG2 cell line was reported with cell viability measured using MTT colorimetric and ATP chemiluminescent assays. Ketoconozole inhibition elucidated the action of the particular P450 isoform in the metabolism of these drugs (Vignati *et al.*, 2005). Idiosyncratic drug reactions (IDRs) were studied employing a proteomics approach where a mixture of drugs causing hepatotoxicity *in vivo* induced intersecting extracellular protein biomarkers from the cells (Gao *et al.*, 2004). Drug–protein interactions can be studied by hepatocyte or microsomal incubation of a scrutinized compound and radioactive or LC-MS analysis of adducted proteins (Evans *et al.*, 2004). Further metabolite analysis can be done employing LC-MS methods after

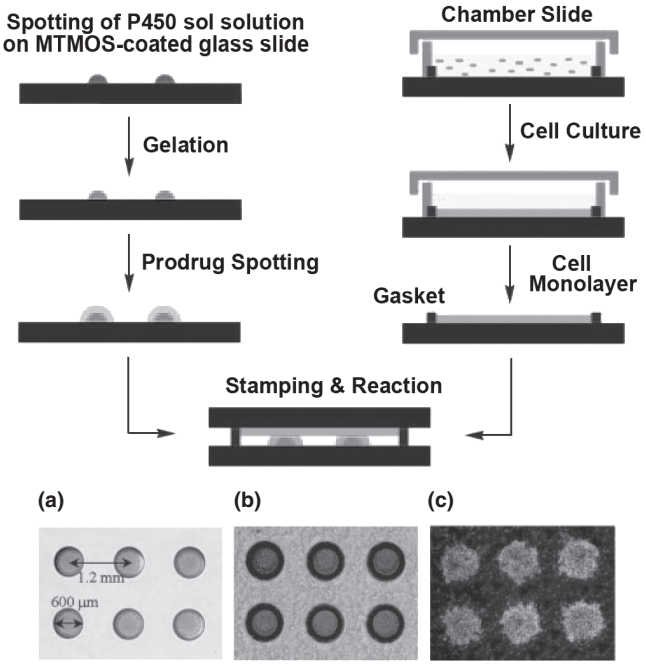


**Scheme 9** Schematic of the MetaChip platform and microscopic photographs of sol-gel spots. Shown are 30-nL P450 sol-gel spots (**a**), 30-nL sol-gel spots with 60 nL of prodrug solution after being stamped by MCF7 cell monolayer **(b)**, and MCF7 cell monolayer after removal from sol-gel array and staining **(c)**. Adapted with permission from Lee *et al.* (2005b). Copyright 2005 National Academy of Sciences.

adduction or "trapping" with nucleophilic compounds, such as glutathione (Liebler and Guengerich, 2005; Park *et al.*, 2005; Caldwell and Yan, 2006).

Beyond detection of toxicity at the cellular level, several *in vitro* tests have been designed to predict histological (tissue) level toxicity. Inter-organ interactions were modeled using six different organ and tissue cell types placed in the same well of a specially designed microtiter plate connected with a substrate containing common fluid (Li, Bode and Sakai, 2004). *In vitro* modeling of the blood–brain barrier (BBB) has been accomplished employing brain endothelial and glial cells attached in multi-well plates fitted with a porous membrane modeling the capillary barrier. Transfer between the two wells was monitored and can be used to estimate pharmaceutical transfer between capillaries and the brain (Cecchelli *et al.*, 1999). Embryonic stem cells have been used as a screen for potential terotogenic compounds. The embryonic stem cell test (EST) has the ability to accurately determine strong embryotoxic materials. However, this assay needs to be refined to include maternal metabolism enzymes to preclude false positives (McNeish, 2004).

High-content screening (HCS) is a method that measures several cytotoxicity parameters simultaneously. HepG2 hepatocyte cells in 96-well plates were incubated for three days and exposed to several fluorescent dyes to analyze calcium content, mitochondrial membrane potential, DNA content, and cell number. This system was

able to predict the hepatotoxicity of 90% of varying degree liver toxins. Of those, 98% are known to cause some form of *in vivo* toxicity demonstrating the predictive ability of the HCS system (O'Brien *et al.*, 2006).

In order to more accurately model the action of cancerous cell growths, breast cancer cells were encapsulated in alginate poly-L-lysine alginate (ALA) microcapsules and exposed to varying chemotherapeutic drugs. Compared with monolayer culture assays, the studied drugs exhibited altered toxicity profiles in the 3-D microcapsule assay, which demonstrated the complex nature of actual tumor cells and elucidated some serious drawbacks to monolayer cell-based assay formats (Zhang *et al.*, 2005).

### 4.3 -Omics Technologies

Genome sequence alone does not predict disease phenotypes, giving rise to a significant number of approaches designed to analyze gene expression to elucidate such phenotypic outcomes. These fields are termed the -omics fields: genomics (genetic complement), transcriptomics (gene expression), proteomics (cell signaling and protein synthesis), metabolomics (cellular metabolite flux), and metabonomics (endogenous small molecule flux) (Nicholson *et al.*, 2002).

The use of DNA microarrays in transcriptomics or toxicogenomics promises to be a powerful tool in the elucidation of DNA damage mechanism via drug and environmental toxin exposure. In such experiments, selected animals or cell lines are exposed to a compound of interest and after sufficient time, RNA is isolated from the exposed tissue or cells. The RNA is amplified, fluorescently labeled, and allowed to hybridize to array-immobilized DNA probes (Butte, 2002). Additionally, the RNA can undergo reverse-transcription producing the original DNA sequence which can then be analyzed. The overall goal is to determine the regulation of a set of genes due to physiological changes from drug exposure. From a genotoxic standpoint, damage induced by a drug would elicit a damage response from a host of genes that will be detected in the microarray (Ulrich and Friend, 2002). In this fashion, a selection of direct genotoxic agents was discerned from indirect genotoxic agents (i.e. mode of action not directly on DNA itself) based on differences in gene expression profiles after exposure to mouse lymphoma cells (Hu *et al.*, 2004). Gene expression analysis of rat liver cells exposed to a series of marketed thiazolidinedione diabetes drugs was performed to elucidate the IDR mechanism that caused recall of one of the drugs. This method identified several expressed genes involved in DNA damage response, cyt P450 induction, and cell regulatory cycles demonstrating that the IDR response was multifaceted in nature. Also, the DNA expression procedure was able to more sensitively determine IDRs than *in vivo* tests (Vansant *et al.*, 2006).

Microarrays have been employed to measure the damaging effects of electron dosing at 3 eV. Fluorescent DNA microarrays demonstrated that 300 electrons per dT 25-mer oligomer was sufficient to preclude hybridization (Solomun and Sturm, 2007). Overall, despite the promise to elucidate genotoxic mechanism, initial microarray studies have shown less sensitivity than traditional genotoxic assays, such as the MNT assay (Newton, Aardema and Aubrecht, 2004). In addition, questions exist about reproducibility and interpretation, but standardization of protocols can improve interlaboratory findings (Irizarry *et al.*, 2005; Owens, 2005).

Proteomic microarrays have been developed to study translation products in a cellular context. From a toxicity standpoint, the interest lies in the charting of expression levels, emergence of new protein biomarkers, alteration of translated products, or

post-translation modifications (Stoll *et al.*, 2004). For instance, microarrays of capture antibodies have been employed to ascertain the changes in protein expression in cancer cells upon radiation treatment (Sreekumar *et al.*, 2001). Sandwich immunoassay arrays have been developed to quantify several different cellular proteins at femtomolar levels (Schweitzer *et al.*, 2000). Limitations in capture antibody development and high-throughput issues will need to be improved before proteomic microarray applications become more widespread (Stoll *et al.*, 2004).

Metabonomics attempts to study the time flux of relative concentrations of endogenous small molecule components of biofluids and tissues (Lindon *et al.*, 2003). In this fashion, a complete end-point analysis due to genetic expression, protein translation, or metabolite exposure of these samples can be discerned and multiple biomarkers of drug exposure can be elucidated. The benefits of this application are numerous and include a global profiling tool at relatively low per sample expense compared with DNA microarray analysis (Griffin and Bollard, 2004). Magic angle resonance (MAR) $^{1}$H NMR allows for histological screening on solid tissue samples after drug exposure. This technique requires little sample workup and does not destroy the sample so that it can be used in multiple analyses (Nicholson *et al.*, 2002; Lindon *et al.*, 2003). Coupled with LC-MS, vast amounts of metabolite and other proteomic biomarker data due to drug exposure can be discerned. An example of metabonomic toxicology was presented by Nicholson *et al.* with the rapid determination of fatty acid metabolism impairment by a failed pharmaceutical candidate. The metabonomic profile of the drug was consistent with several lines of *in vivo* rodent and primate experiments (Mortishire-Smith *et al.*, 2004). Despite high promise, drawbacks of this approach are the throughput and sensitivity of the NMR, the complexity of the data analysis, and initial start-up expenses in acquiring equipment with sufficient frequency resolution. Overall, due to the expense and high-throughput limitations, the -omics fields will likely be used extensively in detailed studies of further developed drug candidates to elucidate findings from preliminary *in vitro* genotoxicity screens (Newton, Aardema and Aubrecht, 2004).

### 4.4 Emerging Toxicity Detection Technologies

In addition to some of the aforementioned cell-based cytotoxicity assays, other novel means of detecting xenobiotic toxicity have also been reported. Cell morphology changes eliciting a transduction signal alteration have shown promise. Scheme 10 shows the experimental setup described by Schwartz *et al.* using scattered light intensity changes to monitor real-time cellular viability. Rat hepatocytes were immobilized on a porous silica photonic crystal and upon exposure to cytotoxic agents cadmium chloride or acetaminophen, the intensity of scattered light at 640 nm increased demonstrating the alteration in the cellular structure due to the xenobiotic action (Schwartz *et al.*, 2006).

In a similar vein, microelectrode-containing microtiter plates have been employed to measure the impedance change between the electrode and the ambient upon exposure to cytotoxic agents. Cells adhered to the electrode surface were exposed to various cytotoxic agents that induce morphological changes. Cytotoxicity affects cell viability, cell number, and electrode adherence producing a resistance change from electrode to the surrounding buffer that correlates with rates of xenobiotic-induced cytotoxicity. The results obtained in this fashion charted well with standard colorimetric detection (Xiao and Luong, 2005; Xing *et al.*, 2005; Zhu *et al.*, 2006).

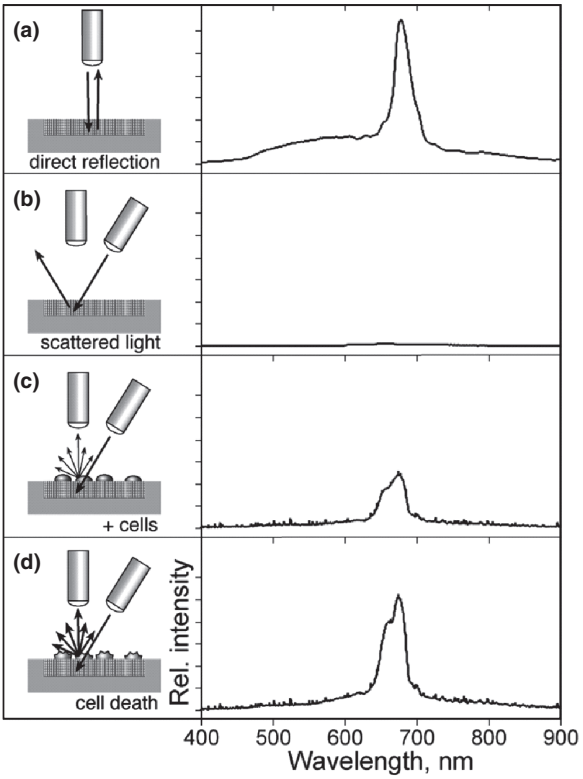


**Scheme 10** Optical design used to monitor biological events on a polystyrene-filled porous Si photonic crystal, with representative spectra. **(a)** To measure the specular reflection from the photonic crystal, illumination and observation share the same optical path, along the axis normal to the chip surface, leading to a sharp reflectivity peak, as shown on the right of the diagram. **(b)** To measure scattered (diffuse) reflectivity from the photonic crystal, the light source is incident from an off-normal position. The specular reflection from the photonic crystal is no longer observed. **(c)** Placing cells on the surface of the chip introduces scattering centers that direct some of the light into the detection optics. **(d)** Changes in cell morphology alter scattering efficiency, which are detected as changes in the intensity of the spectral peak. Adapted with permission from Schwartz *et al.* (2006). Copyright 2006 American Chemical Society.

Similar electronic microsensors have been employed to monitor certain areas of toxicity upon exposure to exogenous agents. Metabolic oxygen consumption changes and extracellular acidities of tumor cells exposed to drugs were measured concurrently using a multi-parametric electronic sensor. Differences in measured parameters allowed discernment of the mechanism of drug action on the cells (Montrescu *et al.*, 2005). Electrochemical measurement of $K^+$-induced dopamine release from a microarray of cells demonstrated that exposure to nomifensine, a dopamine transporter inhibitor, can provide a more sensitive toxicity screen than standard staining and counting viable cells (Cui *et al.*, 2006). Also using a microelectrode array format, the decrease in action potentials of cardiac cells was used to estimate the toxicity of pesticides (Natarajan *et al.*, 2006).

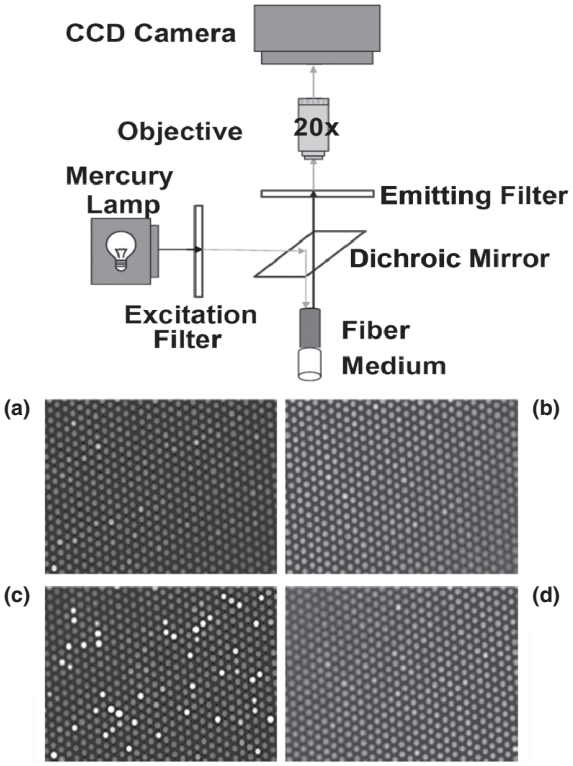


**Scheme 11** Single-cell array biosensor platform setup and fluorescent images from a small portion of two imaging fiber-based arrays. The left panel shows an array where cells were incubated with medium containing 5 μg mL$^{-1}$ mitomycin C. Fluorescent signals were acquired at **(a)** t = 0 minute and **(b)** t = 90 minutes. The right panel shows the control array where cells were incubated with only medium and fluorescent signals were taken at **(c)** t = 0 minute and **(d)** t = 90 minutes. The bright spots indicate a fluorescence increase, and the cells exposed to MMC were much brighter than the control. Adapted with permission from Kuang, Biran and Walt (2004). Copyright 2004 American Chemical Society.

Other optical detection strategies to monitor toxicity have also been described. Surface plasmon resonance (SPR) has been used to monitor adsorption processes modeling serum-binding behavior of drugs on surfaces (Cooper, 2002). Alveolar epithelial cells anchored to a fiber optic were employed to assay the cytotoxicity of inhalation hazards employing fiber evanescent wave spectroscopy (FEWS). Wavelengths in the mid-IR range were monitored and compared with a MTT colorimetric assay to ascertain damage by various airborne toxic agents (Riley *et al.*, 2006).

Fiber optic arrays to measure genotoxicity were developed by Walt *et al.* employing *Escherichia coli* carrying a *recA:gfp* fusion reporter plasmid. A single bacterium was loaded into a properly sized microwell on the tip of each fiber optic (shown in Scheme 11). Bundling of the individual fibers created a high-density biosensor array. Upon exposure to genotoxic medium, fluorescence from expressed GFP was captured

by a distally located CCD camera (Kuang, Biran and Walt, 2004). Likewise, compatible *recA:gfp* and *katG:lux* plasmids inserted into *E. coli* were described to detect and distinguish genotoxic from oxidative cell damage based on increases in fluorescence and bioluminescence, respectively (Mitchell and Gu, 2004a, b). Extending this methodology, *E. coli* containing several different promoters fused with *lux* genes were placed in well plates or constructed chip arrays. Toxicity was assessed from bioluminescence increases detected by a CCD camera and was dependent on reporter gene and employed damage agent (Lee *et al.*, 2005a).

Improvements to the "lab on a chip" concept for biological monitoring continue to be reported. Several unique microfluidic and cell trapping approaches have been described with potential toxicity detection applications (Dittrich and Manz, 2006). *In vitro* protein expression in manufactured oil and water droplets ("biocontainers") was detected fluorescently using a microfluidic approach. Genetic and protein expression components were mixed simultaneously with the oil and water. Expression occurred along a perpendicular reaction strip after mixing, and detection at certain positions along the strip allowed for time-dependent monitoring of protein expression (Dittrich, Jahnz and Schwille, 2005). Also, improvements in such microfluidic devices allow for single-cell immobilization and analysis of the cellular actions upon exposure to an exogenous medium (Dittrich and Manz, 2006). A novel concept in the microfluidics area involves simultaneously monitoring of multiple cell types using a four-compartment flow cell microdevice to mimic *in vivo* metabolism processes. As a test, naphthalene was metabolized in the liver compartment by cyt P450-containing hepatocyte cells followed by flow dispersion into the other compartments of the cell. Depletion of glutathione in the lung compartment denoted reactive metabolite formation (Viravaidya, Sin and Shuler, 2004).

Overall, numerous clever approaches have been explored to measure toxicity of new drug candidates and environmental xenobiotics with varying optical and electrochemical transduction strategies. In order to gain wider acceptance, these platforms must be able to assess chemical toxicity from a wide range of mechanisms in a cost-efficient manner. They must be able to provide more sensitive information, and even chemical structural information, if they are to eventually be used as replacements or in conjunction with traditional toxicity assays. Based on the continual financial driving force to discover toxic pharmaceutical candidates at early stages, we envision that more novel *in vitro* toxicity assays will be developed and employed as initial toxicity screens followed by a more thorough analysis using *-omics* analyses.

## 5 SUMMARY AND FUTURE OUTLOOK

In this chapter, we described new array-based methodology utilizing enzyme/DNA films to screen for reactive drug metabolites based on a DNA-damage end point measured by ECL. Similar enzyme/DNA films on nanoparticles coupled with LC-MS to obtain structure, and rate information for formation of nucleobase adducts and reactive metabolites was also described. Relative rates of DNA damage from both of these biochemical assays can be used to predict relative genotoxicity and the tendency to form reactive metabolites for new drugs. These methods are complementary since the ECL arrays provide rapid screening for reactive metabolites, while the LC-MS approach provides valuable structural and rate of formation information about metabolites and possible nucleobase adducts. Both approaches are adaptable to enzyme inhibition

studies. These technologies are currently being developed in higher-throughput versions taking advantage of modern robotics and biomolecule printing devices that handle sub-nanoliter volumes.

This chapter also surveyed recent developments utilizing other types of array and nanoscience-based technologies for drug toxicity assessment. The future appears bright for incorporation of some of these technologies in high-throughput drug toxicity screening. Many of these approaches could be used alongside or in place of certain conventional toxicity tests.

We feel that the new approaches described herein will prove to be valuable in drug toxicity testing along with other reliable tests and *in silico* assessments. We agree with many pharmaceutical scientists that the future lies in developing predictive batteries of tests upon which to base reliable informed decisions about toxicity at early stages in the development life of a drug (Nasser, Kamel and Clarimont, 2004; Caldwell and Yan, 2006; Mayne, Ku and Kennedy, 2006; Kramer, Sagartz and Morris, 2007). Indeed, this is already happening. We feel that it would be advantageous to evaluate newer, faster, more biochemical-based approaches to the toxicity testing toolbox on an experimental basis, and that perhaps drug companies should develop small groups of bioanalytical scientists charged with this specific task. After a period of evaluation, the new technologies could be adopted if they prove valuable, or discarded if they do not. The hope and dream is that such an approach would establish reliable new high-throughput tests in the toxicity toolbox that would eventually lower the number of drug failures late in the development process, and thus lower drug development costs, as well as drug costs to patients.

## REFERENCES

Adams, C., Adams, P.A. (1992). A comparative study of pH/activity profiles for the anaerobic $H_2O_2$ and alkyl hydroperoxide supported N demethylation of N methylaniline catalyzed by alkaline haematin and microsomal cytochrome P-450. *J. Inorg. Biochem.*, *45*, 47–52.

Ames, B.N., Mcann, J., Yamasaki, E. (1975). Methods for detecting carcinogens and mutagens with the Salmonella/mammalian-microsome mutagenicity test. *Mutat. Res.*, *31*, 347–364.

Appruzzese, W.A., Vouros, P. (1999). Analysis of DNA adducts by capillary methods coupled to mass spectrometry: a perspective. *J. Chromatogr. B*, *794*, 97–108.

Ariga, K., Hill, J.P., Ji, Q. (2007). Layer by layer assembly as a versatile bottom-up nanofabrication technique for exploratory research and realistic application. *Phys. Chem. Chem. Phys.*, *9*, 2319–2340.

Bajrami, B., Hvastkovs, E.G., Jensen, G., Schenkman, J.B., Rusling, J.F. (2008). Enzyme-DNA biocolloids for DNA adduct and reactive metabolite detection by chromatography-mass spectrometry. *Anal. Chem.*, *80*, (4) 922–932.

Bichara, N., Ching, M.S., Blake, C.L., Ghabrial, H., Smallwood, R.A. (1996). Propranolol hydroxylation and N-desisopropylation by cytochrome P4502D6: studies using the yeast-expressed enzyme and NADPH/O2 and cumene hydroperoxide-supported reactions. *Drug Metab. Dispos.*, *24*, 112–118.

Blackburn, G.F., Shah, H.P., Kenten, J.H., Leland, J., Kamin, R.A., Link, J., Petermann, J., Powell, M.J., Shah, A., Talley, D.B., Tyagi, S.K., Wilkins, E., Wu, T.-G., Massey, R.J. (1991). Electrochemiluminescence detection for development of immunoassays and DNA probe assays for clinical diagnostics. *Clin. Chem.*, *37*, 1534–1539.

Bond, J.A. (1989). Review of the toxicology of styrene. *CRC Crit. Rev. Toxicol.*, *19*, 227–249.

Butte, A. (2002). The use and analysis of microarray data. *Nat. Rev. Drug Discov.*, *1*, 951–960.

Caldwell, G.W., Yan, L. (2006). Screening for reactive intermediates and toxicity assessment in drug discovery. *Curr. Opin. Drug Discov. Devel.*, *9*, 47–50.

Caruso, F., Lichtenfeld, H., Giersig, M., Möhwald, H. (1998). Electrostatic self-assembly of silica nanoparticle-polyelectrolyte multilayers on polystyrene latex particles. *J. Am. Chem. Soc.*, *120*, 8523–8524.

Caruso, F., Möhwald, H. (1999). Protein multilayer formation on colloids through a stepwise self-assembly technique. *J. Am. Chem. Soc.*, *121*, 6039–6046.

Caruso, F., Schüler, C. (2000). Enzyme multilayers on colloid particles: assembly, stability, and enzymatic activity. *Langmuir*, *16*, 9595–9603.

Cavalieri, E.L., Rogan, E.G., Devaneshan, P.D., Cremonesi, P., Cerny, R.L., Gross, M.L., Bodell, W.J. (1990). Binding of benzo[a]pyrene to DNA by cytochrome P 450 catalyzed one-electron oxidation in rat liver microsomes and nuclei. *Biochemistry*, *29*, 4820–4827.

Cecchelli, R., Dehouck, B., Fenart, L., Buee-Scherrer, V., Duhem, C., Lundquist, S., Rentfel, M., Torpier, G., Dehouck, M.P. (1999). In vitro model for evaluating drug transport across the blood-brain barrier. *Adv. Drug Deliv. Rev.*, *36*, 165–178.

Cooper, M.A. (2002). Optical biosensors in drug discovery. *Nat. Rev. Drug Discov.*, *1*, 515–528.

Cui, H., Ye, J., Chen, Y., Chong, S., Sheu, F. (2006). Microelectrode array biochip: tool for in vitro drug screening based on the detection of a drug effect on dopamine release from PC12 cells. *Anal. Chem.*, *78*, 6347–6355.

Dennany, L., Forster, R.J., Rusling, J.F. (2003). Simultaneous direct electrochemiluminescence and catalytic voltammetry detection of DNA in ultrathin films. *J. Am. Chem. Soc.*, *125*, 5213–5218.

Dittrich, P.S., Jahnz, M., Schwille, P. (2005). A new embedded process for compartmentalized cell-free protein expression and on-line detection in microfluidic devices. *Chembiochem*, *6*, 811–814.

Dittrich, P.S., Manz, A. (2006). Lab on a chip: microfluidics in drug discovery. *Nat. Rev. Drug Dev.*, *5*, 210–218.

Ege, D., Becker, W.G., Bard, A.J. (1984). Electrogenerated chemiluminescent determination of tris(2,2′-bipyridine)ruthenium ion (Ru(bpy)32+) at low levels. *Anal. Chem.*, *56*, 2413–2417.

Evans, D.C., Watt, A.P., Nicoll-Griffith, D.A., Baillie, T.A. (2004). Drug-protein adducts: an industry perspective on minimizing the potential for drug bioactivation in drug discovery and development. *Chem. Res. Toxicol.*, *17*, 3–16.

Fang, M., Grant, P.S., McShane, M.J., Sukhorukov, G.B., Golub, V.O., Lvov, Y. (2002). Magnetic bio/nanoreactor with multilayer shells of glucose oxidase and inorganic nanoparticles. *Langmuir*, *18*, 6336–6344.

Flukiger-Isler, S., Baumeister, M., Braun, K., Gervais, V., Hasler-Nguyen, N., Reimann, R., Van Gompel, J., Wunderlich, H.-G., Engelhardt, G. (2004). Assessment of the performance of the Ames II assay: a collaborative study with 19 coded compounds. *Mutat. Res.*, *558*, 181–197.

Friedberg, E.C. (2003). DNA damage and repair. *Nature*, *421*, 436–440.

Galloway, S.M., Aardema, M.J., Ishidate, M., Jr, Ivett, J.L., Kirkland, D.J., Morita, T., Mosesso, P., Sofuni, T. (1994). Report from working group on in vitro tests for chromosomal aberrations. *Mutat. Res.*, *312*, 241–261.

Gangl, E.T., Turesky, R.J., Vouros, P. (2001). Detection of in vivo formed DNA adducts at the part-per-billion level by capillary liquid chromatography/microelectrospray mass spectrometry. *Anal. Chem.*, *73*, 2397–2404.

Gao, J., Garulacan, L., Storm, S., Hefta, S., Opiteck, G., Lin, J., Moulin, F., Dambach, D. (2004). Identification of in vitro protein biomarkers of idiosyncratic liver toxicity. *Toxicol. In Vitro*, *18*, 533–541.

Griffin, J.L., Bollard, M.E. (2004). Metabonomics: its potential as a tool in toxicology for safety assessment and data integration. *Curr. Drug Metab.*, *5*, 389–398.

Guengerich, F.P. (2001). Common and uncommon cytochrome P450 reactions related to metabolism and chemical toxicity. *Chem. Res. Toxicol.*, *14*, 611–650.

Guengerich, F.P., Bell, L.C., Okazaki, O. (1995). Interpretations of cytochrome P450 mechanisms from kinetic studies. *Biochemie*, *77*, 573–580.

Hastwell, P.W., Chai, L.-L., Roberts, K.J., Webster, T.W., Harvey, J.S., Rees, R.W., Walmsley, R.M. (2006). High specificity and high sensitivity genotoxicity assessment in a human cell line: validation of the GreenScreen HC GADD45α-GFP genotoxicity assay. *Mutat. Res.*, *607*, 160–175.

Hecht, S.S. (1998). Biochemistry, biology, and carcinogenicity of tobacco specific N-nitrosamines. *Chem. Res. Toxicol.*, *11*, 559–603.

Hecht, S.S. (2003). Tobacco carcinogens, their biomarkers and tobacco-induced cancer. *Nat. Rev. Cancer*, *10*, 733–744.

Hu, T., Gibson, D.P., Carr, G.J., Torontali, S.M., Tiesman, J.P., Chaney, J.G., Aardema, M.J. (2004). Identification of a gene expression profile that discriminates indirect-acting genotoxins from direct-acting genotoxins. *Mutat. Res.*, *549*, 5–27.

Hvastkovs, E.G., So, M., Krishnan, S., Bajrami, B., Tarun, M., Jansson, I., Schenkman, J.B., Rusling, J.F. (2007). Electrochemiluminescent arrays for cytochrome P450-activated genotoxicity screening. DNA damage from benzo[a]pyrene metabolites. *Anal. Chem.*, *79*, 1897–1906.

Irizarry, R.A., Warren, D., Spencer, F., Kim, I.F., Biswal, S., Frank, B.C., Gabrielson, E., Garcia, J.G.N., Geoghegan, J., Germino, G., Griffin, C., Hilmer, S.C., Hoffman, E., Jedlicka, A.E., Kawasaki, E., Martinez-Murillo, F., Morsburger, L., Lee, H., Petersen, D., Quackenbush, J., Scott, A., Wilson, M., Yang, Y., Ye, Q.S., Yu, W. (2005). Multiple-laboratory comparison of microarray platforms. *Nat. Methods*, *2*, 345–349.

Jacoby, W.B. (ed.) (1980). *Enzymatic Basis of Detoxification*, Vols I and II, Academic, New York.

Kenten, J.H., Casedei, J., Link, J., Lupold, S., Willey, J., Powell, M., Rees, A., Massey, R.J. (1991). Rapid electrochemiluminescence assays of polymerase chain reaction products. *Clin. Chem.*, *37*, 1626.

Kiskinis, E., Suter, W., Hartmann, A. (2002). High throughput Comet assay using 96-well plates. *Mutagenesis*, *17*, 37–43.

Kramer, J.A., Sagartz, J.E., Morris, D.L. (2007). The application of discovery toxicology and pathology towards the design of safer pharmaceutical lead candidates. *Nat. Rev. Drug Discov.*, *6*, 636–649.

Krishnan, S., Hvastkovs, E.G., Bajrami, B., Choudhary, D., Schenkman, J.B., Rusling, J.F. (2008). Synergistic metabolic toxicity screening using microsome/DNA electrochemiluminescent arrays and nanoreactors. *Anal. Chem.*, *80*, (14) 5279–5285.

Krishnan, S., Hvastkovs, E.G., Bajrami, B., Jansson, I., Schenkman, J.B., Rusling, J.F. (2007). Genotoxicity screening for N-nitroso compounds. Electrochemical and electrochemiluminescent detection of human enzyme-generated DNA damage from N-nitrosopyrrolidine. *Chem. Commun.*, 1713–1715.

Kuang, L., Biran, I., Walt, D. (2004). Living bacterial cell array for genotoxin monitoring. *Anal. Chem.*, *76*, 2902–2909.

Kupfer, R., Liu, S.Y., Allentoff, A.J., Thompson, J.A. (2001). Comparisons of hydroperoxide isomerase and monooxygenase activities of cytochrome P450 for conversions of allylic hydroperoxides and alcohols to epoxyalcohols and diols: probing substrate reorientation in the active site. *Biochemistry*, *40*, 11490–11501.

Lee, J., Mitchell, R., Kim, B., Cullen, D., Gu, M. (2005a). A cell array biosensor for environmental toxicity analysis. *Biosens. Bioelectron.*, *21*, 500–507.

Lee, M., Park, C., Dordick, J., Clark, D. (2005b). Metabolizing enzyme toxicology assay chip (MetaChip) for high throughput microscale toxicity analyses. *Proc. Natl. Acad. Sci. U.S.A.*, *102*, 983–987.

Leland, J.K., Powell, M.J. (1990). Electrogenerated chemiluminescence: an oxidative-reduction type ECL reaction sequence using tripropyl amine. *J. Electrochem. Soc.*, *137*, 3127–3131.

Li, A., Bode, C., Sakai, Y. (2004). A novel in vitro system, the integrated discrete multiple organ cell culture (IdMOC) system, for the evaluation of human drug toxicity: comparative cytotoxicity of tamoxifen towards normal human cells from five major organs and MCF-7 adenocarcinoma breast cells. *Chem. Biol. Interact.*, *150*, 129–136.

Liebler, D.C., Guengerich, F.P. (2005). Elucidating the mechanisms of drug-induced toxicity. *Nat. Rev. Drug Dev.*, *4*, 410–420.

Lindon, J.C., Nicholson, J.K., Holmes, E., Antti, H., Bollard, M.E., Keun, H., Beckonert, O., Ebbels, T.M., Reily, M.D., Robertson, D.R., Stevens, G.J., Luke, P., Breau, A.P., Cantor, G.H., Bible, R.H., Niederhauser, U., Senn, H., Schlotterbeck, G., Sideelmann, U.G., Laursen, S.M., Tymiak, A., Car, B.D., Lehman-McKeeman, L., Colet, J.-M., Loukaci, A., Thomas, C. (2003). Contemporary issues in toxicology: the role of metabonomics in toxicology and its evaluation by the COMET project. *Toxicol. Appl. Pharm.*, *187*, 137–146.

Lippard, S.J., Berg, J.M. (1994). *Principles of Bioorganic Chemistry*, University Science Books, Mill Valley, CA.

Lvov, Y. (2000). *Protein Architecture: Interfacing Molecular Assemblies and Immobilization Biotechnology* (eds Y. Lvov and H. Möhwald), Marcel Dekker, New York, pp. 125–167.

Lvov, Y. (2001). *Handbook of Surfaces and Interfaces of Materials*, Vol. *3*. Nanostructured Materials, Micelles and Colloids (ed. R.W. Nalwa), Academic Press, San Diego, pp. 170–189.

Lvov, Y., Caruso, F. (2001). Biocolloids with ordered urease multilayer shells as enzymatic reactors. *Anal. Chem.*, *73*, 4212–4217.

McNeish, J. (2004). Embryonic stem cells in drug discovery. *Nat. Rev. Drug Dev.*, *3*, 70–80.

Marnett, L.J., Burcham, P.C. (1993). Endogenous DNA adducts: potential and paradox. *Chem. Res. Toxicol.*, *6*, 771–785.

Mayne, J.T., Ku, W.W., Kennedy, S.P. (2006). Informed toxicity assessment in drug discovery: systems-based toxicology. *Curr. Opin. Drug Discov. Devel.*, *9*, 75–83.

Mitchell, R., Gu, M. (2004a). An Escherichia coli biosensor capable of detecting both genotoxic and oxidative damage. *Appl. Microbiol. Biotechnol.*, *64*, 46–52.

Mitchell, R., Gu, M. (2004b). Construction and characterization of novel dual stress-responsive bacterial biosensors. *Biosens. Bioelectron.*, *19*, 977–985.

Montrescu, E.R., Otto, A.M., Brischwein, M., Zahler, S., Wolf, B. (2005). Dynamic analysis of metabolite effects of chloroacetaldehyde and cytochalasin B on tumor cells using bioelectronic sensor chips. *J. Cancer Res. Clin. Oncol.*, *131*, 683–691.

Mortishire-Smith, R.J., Skiles, G.L., Lawrence, J.W., Spence, S., Nicholls, A.W., Johnson, B.A., Nicholson, J.K. (2004). Use of metabonomics to identify impaired fatty acid metabolism as the mechanism of drug induced toxicity. *Chem. Res. Toxicol.*, *17*, 165–173.

Mugweru, A., Rusling, J.F. (2006). Studies of DNA damage inhibition by dietary antioxidants using metallopolyion/DNA sensors. *Electroanalysis*, *18*, 327–332.

Mugweru, A., Yang, J., Rusling, J.F. (2004). Comparison of hemoglobin and myoglobin for in-situ metabolite generation in chemical toxicity sensors using a metallopolymer catalyst for DNA damage detection. *Electroanalysis*, *16*, 1132–1138.

Munge, B., Estavillo, C., Schenkman, J.B., Rusling, J.F. (2003). Optimizing electrochemical and peroxide-driven oxidation of styrene with ultrathin polyion films containing cytochrome P450cam and myoglobin. *Chembiochem*, *4*, 82–89.

Nasser, A.E.F., Kamel, A.M., Clarimont, C. (2004). Improving the decision making process in structural modification of drug candidates: reducing toxicity. *Drug Discov. Today*, *9*, 1055–1064.

Natarajan, A., Molnar, P., Sieverdes, K., Jamshidi, A., Hickman, J. (2006). Microelectrode array readings of cardiac action potentials as a high throughput method to evaluate pesticide toxicity. *Toxicol. In Vitro*, *20*, 375–381.

Neilson, A.H. (ed.) (1998). *PAHs and Related Compounds*, Springer, Berlin.

Newton, R.K., Aardema, M., Aubrecht, J. (2004). The utility of DNA microarrays for characterizing genotoxicity. *Environ. Health Pers.*, *112*, 420–422.

Nicholson, J.K., Connelly, J., Lindon, J.C., Holmes, E. (2002). Metabonomics: a platform for studying drug toxicity and gene function. *Nat. Rev. Drug Discov.*, *1*, 156–161.

O'Brien, P., Irwin, W., Diaz, D., Cofield, H., Krejsa, C., Slaughter, M., Gao, B., Kaludercic, N., Angeline, A., Bernardi, P., Brain, P., Hougham, C. (2006). High concordance of drug-induced human hepatotoxicity with in vitro cytotoxicity measured in a novel cell based model using high content screening. *Arch. Toxicol.*, *80*, 580–604.

Oda, Y., Nakamura, S., Oki, I., Kato, T., Shinagawa, H. (1985). Evaluation of a new system (umu-test) for the detection of environmental mutagens and carcinogens. *Mutat. Res.*, *147*, 212–229.

Ortiz de Montellano, P.R. (2005). *Cytochrome P-450: Structure, Mechanism, and Biochemistry*, 3rd edn, Kluwer Academic/Plenum, New York.

Ortiz de Montellano, P.R., De Voss, J.J. (2002). Oxidizing species in the mechanism of cytochrome P450. *Nat. Prod. Rep.*, *19*, 477–493.

Owens, J. (2005). Do microarrays match up? *Nat. Rev. Drug Discov.*, *4*, 459.

Park, K., Williams, D.P., Naisbitt, D.J., Kitteringham, N.R., Pirmohamed, M. (2005). Investigation of toxic metabolites during drug development. *Toxicol. Appl. Pharmacol.*, *207*, S425–S434.

Pauwels, W., Vodiceka, P., Servi, M., Plna, K., Veulemans, H., Hemminki, K. (1996). Adduct formation on DNA and haemoglobin in mice intraperitoneally administered with styrene. *Carcinogenisis*, *17*, 2673–2680.

Phillips, D.H., Farmer, P.B., Beland, F.A., Nath, R.G., Poirier, M.C., Reddy, M.V., Turtletaub, K.W. (2000). Methods of DNA adduct determination and their application to testing compounds for genotoxicity. *Environ. Mol. Mutagen.*, *35*, 222–233.

Preussmann, R., Stewart, B.W. (1984). N-Nitroso carcinogens, in *Chemical Carcinogenesis* (ed. C.E. Searle), ACS monograph 182, American Chemical Society, Washington, DC, pp. 643–828.

Rahimtula, A.D., O'Brien, P.J., Seifried, H.E., Jerina, D.M. (1978). The mechanism of action of cytochrome P-450. Occurrence of the "NIH shift" during hydroperoxide-dependent aromatic hydroxylations. *Eur. J. Biochem.*, *89*, 133–141.

Riley, M.R., Lucas, P., Le Coq, D., Juncker, C., Boesewetter, D.E., Collier, J.L., DeRosa, D.M., Katterman, M.E., Boussard-Pledel, C., Bureau, B. (2006). Lung cell fiber evanescent wave spectroscopic biosensing of inhalation health hazards. *Biotechnol. Bioeng.*, *95*, 599–612.

Rogan, E.G., Devanesan, P.D., Ramakrishna, N.V.S., Higgenbotham, S., Padvavathi, N.S., Chapman, K., Cavalieri, E.L., Jeong, H., Jankowiak, R., Small, G.J. (1993). Identification and quantitation of benzo[a]pyrene-DNA adducts formed in mouse skin. *Chem. Res. Toxicol.*, *6*, 356–363.

Rubinstein, I., Bard, A.J. (1980). Polymer films on electrodes. 4. Nafion-coated electrodes and electrogenerated chemiluminescence of surface-attached tris(2,2′-bipyridine)ruthenium(2+). *J. Am. Chem. Soc.*, *102*, 6642–6644.

Rusling, J.F. (2004). Sensors for toxicity of chemicals and oxidative stress based on electrochemical catalytic DNA oxidation. *Biosens. Bioelectron.*, *20*, 1022–1028.

Rusling, J.F. (2005). *Electrochemistry of Nucleic Acids and Proteins* (eds E. Palecek, F. Scheller and J. Wang), Elsevier, pp. 433–450.

Rusling, J.F., Hvastkovs, E.G., Hull, D.O., Schenkman, J.B. (2007). Biochemical applications of ultrathin films of enzymes, polyions and DNA (feature article). *Chem. Commun.*, DOI: 10.1039/b709121b.

Rusling, J.F., Hvastkovs, E.G., Schenkman, J.B. (2007). Toxicity screening using biosensors that measure DNA damage. *Curr. Opin. Drug Discov. Devel.*, *10*, 67–73.

Sasaki, Y.F., Sekihashi, K., Izumiyama, F., Nishidate, E., Saga, A., Ishida, K., Truda, S. (2000). The Comet assay with multiple mouse organs: comparison of Comet assay results and carcinogenicity with 208 chemicals selected from the IARC monographs and U.S. NTP Carcinogenicity Database. *Crit. Rev. Toxicol.*, *30*, 629–799.

Scharer, O.D. (2003). Chemistry and biology of DNA repair. *Angew. Chem. Int. Ed. Engl.*, *42*, 2946–2974.

Schenkman, J.B., Greim, H. (eds) (1993). *Cytochrome P450*, Springer-Verlag, Berlin.

Schwartz, M.P., Derfus, A.M., Alvarez, S.D., Bhatia, S.N., Sailor, M.J. (2006). The smart petri dish: a nanostructured photonic crystal for real-time monitoring of living cells. *Langmuir*, *22*, 7084–7090.

Schweitzer, B., Wiltshire, S., Lambert, J., O'Malley, S., Kukanskis, K., Zhu, Z., Kingsmore, S.F., Lizardi, P.M., Ward, D.C. (2000). Immunoassays with rolling circle DNA amplification: a versatile platform for ultrasensitive antigen determination. *Proc. Natl. Acad. Sci. U.S.A.*, *97*, 10113–10119.

Shuga, J., Zhang, J., Samson, L.D., Lodish, H.F., Griffith, L.G. (2007). In vitro erythropoiesis from bone marrow-derived progenitors provides a physiological assay for toxic and mutagenic compounds. *Proc. Natl. Acad. Sci. U.S.A.*, *104*, 8737–8742.

Shutava, T., Zheng, Z., John, V., Lvov, Y. (2004). Microcapsule modification with peroxidase-catalyzed phenol polymerization. *Biomacromolecules*, *5*, 914–921.

Singh, N.P., McCoy, M.T., Tice, R.R., Schneider, E.L. (1998). A simple technique for quantitation of low levels of DNA damage in individual cells. *Exp. Cell. Res.*, *175*, 184–191.

So, M., Hvastkovs, E.G., Schenkman, J.B., Rusling, J.F. (2007). Electrochemiluminescent/voltammetric toxicity screening sensor using enzyme-generated DNA damage. *Biosens. Bioelectron.*, *23* (4), 492–498.

Solomun, T., Sturm, H. (2007). Bringing electrons and microarray technology together. *J. Phys. Chem. B*, *111*, 10636–10638.

Sreekumar, A., Nyati, M.K., Varambally, S., Barrette, T.R., Ghosh, D., Lawrence, T.S., Chinnaiyan, A.M. (2001). Profiling of cancer cells using protein microarrays: discovery of novel radiation-regulated proteins. *Cancer Res.*, *61*, 7585–7593.

Stoll, D., Bachmann, J., Templin, M.F., Joos, T.O. (2004). Microarray technology: an increasing variety of screening tools for proteomic research. *Drug Discov. Today: Targets*, *3*, 24–31.

Styles, J.A., Clark, H., Festing, M.F.W., Rew, D.A. (2001). Automation of mouse micronucleus genotoxicity assay by laser scanning cytometry. *Cytometry*, *44*, 153–155.

Tanaka, E., Terada, M., Misawa, S. (2000). Cytochrome P450 2E1: its clinical and toxicological role. *J. Clin. Pharm. Ther.*, *3*, 165–175.

Tarun, M., Rusling, J.F. (2005a). Measuring DNA nucleobase adducts using neutral hydrolysis and liquid chromatography-mass spectrometry. *Crit. Rev. Eukaryot. Gene Expr.*, *15*, 295–315.

Tarun, M., Rusling, J.F. (2005b). Quantitative measurement of DNA adducts using neutral hydrolysis and LC-MS. Validation of genotoxicity sensors. *Anal. Chem.*, *77*, 2056–2062.

Tarun, M., Rusling, J.F. (2006). Genotoxicity screening using biocatalyst-DNA films and capillary LC-MS/MS. *Anal. Chem.*, *78*, 624–627.

Todorovic, R., Ariese, F., Devenesan, P., Jankowiak, R., Small, G.J., Rogan, E.G., Cavalieri, E.L. (1997). Determination of benzo[a]pyrene- and 7,12-dimethylbenz[a]anthracene-DNA adducts formed in rat mammary glands. *Chem. Res. Toxicol.*, *10*, 941–947.

Ulrich, R., Friend, S.H. (2002). Toxicogenomics and drug discovery: will new technologies help us produce better drugs. *Nat. Rev. Drug Discov.*, *1*, 84–88.

Umemoto, A., Komaki, K., Monden, Y., Suwa, M., Kanno, Y., Kitagawa, M., Suzuki, M., Lin, C.-X., Ueyama, Y., Momen, M.A., Ravindernath, A., Shibutani, S. (2001). Identification and quantification of tamoxifen-DNA adducts in the liver of rats and mice. *Chem. Res. Toxicol.*, *14*, 1006–1013.

Van Gompel, J., Woestenborghs, F., Beerens, D., Mackie, C., Cahill, P.A., Knight, A.W., Billinton, N., Tweats, D.J., Walmsley, R.M. (2005). An assessment of the utility of the yeast GreenScreen assay in pharmaceutical screening. *Mutagenesis*, *20*, 449–454.

Vanhoutte, K., Dongen, W., Hoes, I., Lemiere, F., Esmans, E.L., Van Onckelen, H., Van den Eckhout, E., Soest, R.E.J., Hudson, A.J. (1997). Development of a nanoscale liquid chromatography/electrospray mass spectrometry methodology for the detection and quantification of DNA adducts. *Anal. Chem.*, *69*, 3161–3168.

Vansant, G., Pezzoli, P., Saiz, R., Birch, A., Duffy, C., Ferre, F., Monforte, J. (2006). Gene expression analysis of troglitazone reveals its impact on multiple pathways in cell culture: a case for in vitro platforms combined with gene expression analysis for early (idiosyncratic) toxicity screening. *Int. J. Toxicol.*, *25*, 85–94.

Vignati, L., Turlizzi, E., Monaci, S., de Grossi, P., Kanter, P., Monshouwe, M. (2005). An in vitro approach to detect metabolite toxicity due to CYP3A4-dependent bioactivation of xenobiotics. *Toxicology*, *216*, 154–167.

Viravaidya, K., Sin, A., Shuler, M. (2004). Development of a microscale cell culture analog to probe naphthalene toxicity. *Biotechnol. Prog.*, *20*, 316–323.

Wang, B., Jansson, I., Schenkman, J.B., Rusling, J.F. (2005). Evaluating enzymes that generate genotoxic benzo[a]pyrene metabolites using sensor arrays. *Anal. Chem.*, *77*, 1361–1367.

Wang, M., Cheng, G., Sturla, S.J., Shi, Y., McIntee, E.J., Villalta, P.W., Upadhyaya, P., Hecht, S.S. (2003). Identification of adducts formed by pyridyloxobutylation of deoxyguanosine and DNA by 4-(acetoxymethylnitrosamino)-1-(3-pyridyl)-1-butanone, a chemically activated form of tobacco specific carcinogens. *Chem. Res. Toxicol.*, *16*, 616–626.

Wang, M., McIntee, E.J., Shi, Y., Cheng, G., Upadhyaya, P., Villalta, P.W., Hecht, S.S. (2001). Reactions of acetoxy-N-nitrosopyrrolidine with deoxyguanosine and DNA. *Chem. Res. Toxicol.*, *14*, 1435–1445.

Warren, A.J., Shields, P.G. (1997). Molecular epidemiology: carcinogen-DNA adducts and genetic susceptibility. *Proc. Soc. Exp. Biol. Med.*, *216*, 172–180.

Werringloer, J., Kawano, S., Estabrook, R.W. (1980). Comparison of the interaction of cumene hydroperoxide and hydrogen peroxide with liver microsomal cytochromes, in *Microsomes, Drug Oxidations, and Chemical Carcinogenesis* (eds M.C. Coon, A.H. Conney, R.W. Estabrook and H.V. Gelboin), Academic Press, New York, pp. 403–406.

White, R.E., Sligar, S.G., Coon, M.W. (1980). Evidence for a homolytic mechanism of peroxide oxygen-oxygen bond cleavage during substrate hydroxylation by cytochrome P-450. *J. Biol. Chem.*, *255*, 11108–11111.

Xiao, C., Luong, H.T. (2005). Assessment of cytotoxicity by emerging impedance spectroscopy. *Toxicol. Appl. Pharm.*, *206*, 102–112.

Xing, J.Z., Zhu, L., Jackson, J., Gabos, S., Sun, X.-J., Wang, X.-b., Xu, X. (2005). Dynamic monitoring of cytotoxicity on microelectrode sensors. *Chem. Res. Toxicol.*, *18*, 154–161.

Xu, X.-H., Bard, A.J. (1995). Immobilization and hybridization of DNA on an aluminum(III) alkanebisphosphonate thin film with electrogenerated chemiluminescent detection. *J. Am. Chem. Soc.*, *117*, 2627–2631.

Xu, X.-H., Yang, H.C., Mallouk, T.E., Bard, A.J. (1994). Immobilization of DNA on an aluminum(III) alkanebisphosphonate thin film with electrogenerated chemiluminescent detection. *J. Am. Chem. Soc.*, *116*, 8386–8387.

Yang, J., Wang, B., Rusling, J.F. (2005). Genotoxicity sensor response correlated with DNA nucleobase damage rates. *Mol. Biosyst.*, *1*, 251–259.

Young-Scaime, R., Wang, M., Chung, F.-L., Hecht, S.S. (1995). Reactions of acetoxy-N-nitrosopyrrolidine and acetoxy-N-nitrosopiperidine with deoxyguanosine: formation of N2-tetrahydrofuranyl and N2-tetrahydropyranyl adducts. *Chem. Res. Toxicol.*, *8*, 607–616.

Zamenhof, S., Arikawa, S. (1966). Depurination and alkylation of DNAs of different base compositions. *Mol. Pharmacol.*, *2*, 570–573.

Zhang, X., Chen, H., Zhang, H. (2007). Layer-by-layer assembly: from conventional to unconventional methods. *Chem. Commun.*, 1395–1405.

Zhang, X., Wang, W., Yu, W., Xie, Y., Zhang, X., Zhang, Y., Ma, X. (2005). Development of an in vitro multicellular tumor spheroid model using microencapsulation and its application in anticancer drug screening and testing. *Biotechnol. Prog.*, *21*, 1289–1296.

Zhou, L., Yang, J., Estavillo, C., Stuart, J.D., Schenkman, J.B., Rusling, J.F. (2003). Toxicity screening by electrochemical detection of DNA damage by metabolites generated *in-situ* in ultrathin DNA-enzyme films. *J. Am. Chem. Soc.*, *125*, 1431–1436.

Zhu, J., Wang, X., Yu, W., Abassi, Y.A. (2006). Dynamic and label-free monitoring of natural killer cell cytotoxic activity using electronic cell sensor arrays. *J. Immunol. Methods*, *309*, 25–33.

Zoltewicz, J.A., Clark, D.F. (1972). Kinetics and mechanism of the hydrolysis of guanosine and 7-methylguanosine nucleosides in perchloric acid. *J. Org. Chem.*, *37*, 1193–1197.

# Enzyme Inhibition

PAUL F. HOLLENBERG

## 1 INTRODUCTION

Inhibition of the catalytic activities of the drug-metabolizing enzymes has received considerable attention from clinical pharmacologists and pharmaceutical companies in the past 20 years, because of the fact that several widely used drugs or drugs in the late stages of development were shown to cause life-threatening adverse affects when given to patients taking other drugs due to their ability to inhibit the metabolism of those drugs. The inhibition of the metabolism of one drug by another drug taken concurrently by the patient has now been shown to be a very important cause of adverse drug reactions (ADRs). The potential for drugs to interact with one another by competing for metabolism by the same drug-metabolizing enzyme continues to be a significant health concern and, if anything, will probably become of greater concern in the coming years. ADRs are suggested to be responsible for up to 10% of all hospital admissions (Kohler *et al.*, 2000). In addition, drug–drug interactions are estimated to be implicated in up to 20% of all ADRs (Levy *et al.*, 1980). The probability that ADRs associated with drug–drug interaction increase as the number of drugs administered concomitantly increases is generally accepted as true, and the occurrence of ADRs is believed to increase very rapidly as the number of drugs given simultaneously to a patient increases. When the activity of a given drug-metabolizing enzyme is modified by inhibition, this modification may not only change the pharmacokinetics of the drug whose metabolism is inhibited by leading to an increase in the levels of the parent drug in the patient but, where more than one enzyme may be responsible for the metabolism of a drug and the different enzymes may result in the formation of different products, it may result in a significant change in the profile of the metabolites formed. If any of the different products formed have significantly enhanced or decreased biological and/or toxicological properties, this may lead to significantly altered biological or toxicological activity. Since many of the drug-metabolizing enzymes may be viewed as being relatively nonspecific (i.e. very promiscuous) in their substrate specificity, it is not surprising that two or more drugs can compete for metabolism by the same enzyme resulting in inhibition of the metabolism of one or both of the drugs. This can lead to enhanced metabolism

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

of one of the drugs by a second enzyme that shows specificity for forming another product from the drugs whose metabolism is inhibited. In the situation where two or more drugs are competing for metabolism by the same enzyme, the effective $K_I$ (the inhibitor concentration that results in half maximal inhibition) for the inhibition of the metabolism of one drug by the other may be comparable to the $K_M$ (the substrate concentration that gives half maximal velocity) for its metabolism. If the peak plasma levels in normal clinical usage approximate the $K_M$ (or $K_I$) for the first drug, this can result in significant inhibition of the metabolism of the second drug. In addition to inhibition, the induction of the drug-metabolizing enzymes also is an important factor in many of the drug interactions documented in the literature (Lin and Lu, 1998). This is discussed in a subsequent chapter.

In addition to the potential for drug–drug interactions leading to ADRs, the pharmaceutical industry is also very interested in enzyme inhibition as a basis for developing drugs that target specific enzymes of interest. For example, some of the cytochrome P450s as well as monoamine oxidase are targets for drug development because the products of their normal reactions may be of importance under certain conditions. A good example is the yeast P450 51 (lanosterol protein α-demethylase), which is a target for many common fungicides currently on the market. Aromatase (P450 19A1), which oxidizes the androgens androstandione and testosterone to form the estrogens estrone and 17β-estradiol, respectively, is also a target for inhibition in cancer therapy.

The development of technologies that can be used to study the interactions of drugs with individual human P450 enzymes *in vitro* by using either the purified recombinant P450 enzymes in reconstituted systems or enzyme selective substrates in human microsomal preparations or freshly isolated primary hepatocytes, all of which are now readily available commercially, has led to major breakthroughs in studying drug–drug interactions *in vitro*. The development of these *in vitro* technologies has made evaluation of drug–drug interactions an integral part of the drug development process such that *in vitro* drug–drug interaction studies are now routinely performed by pharmaceutical companies. However, the *in vitro–in vivo* correlations for drug–drug interactions have been much more problematic. Details of the development of these methodologies are covered in a subsequent chapter.

Competition between two or more different drugs for metabolism by the same enzyme resulting in the inhibition of the metabolism of one or more of the competing drugs may result in unexpected elevations in the plasma concentration(s) that can ultimately result in a variety of adverse effects. Depending on the final plasma concentrations of the inhibited drugs, this can ultimately result in a variety of minor as well as serious adverse effects and, in some cases, has resulted in fatalities. The onset of inhibition in patients is usually rapid following a single administration of the inhibitory compound. However, in some cases, as will be discussed later, the inhibition is not observed immediately.

Several very prominent drugs have been withdrawn from the market and some drugs in late-stage development have ultimately been stopped due to their potential for ADRs resulting from inhibition of the drug-metabolizing enzymes. Drug–drug interactions responsible for the limited utility of a given drug or even its withdrawal from the market generally result in significant financial losses to the pharmaceutical companies involved. Therefore, pharmaceutical companies are continuing to develop and utilize a variety of new approaches for the prediction of possible drug–drug interactions resulting from inhibition of the drug-metabolizing enzymes. Our knowledge of the structures, catalytic mechanisms, and regulation of the drug-metabolizing enzymes has increased greatly during the last decade due to a number of major advances in

molecular biology and biochemical approaches to the study of enzymes. Although the mechanistic aspects of the inhibition of the drug-metabolizing enzymes *in vitro* are much better understood than previously, the accurate prediction of the occurrence and consequences of drug–drug interactions due to inhibition continue to be an important unsolved challenge. This chapter will focus on approaches used to investigate the inhibition of the drug-metabolizing enzymes. The theoretical underpinnings for understanding the various mechanisms by which inhibition may occur as well as experimental design and the interpretation of inhibition data will be discussed. Particular attention is paid to the cytochrome P450 enzymes since many of the clinically relevant pharmacokinetic drug interactions are in one way or another related to the P450s. In addition, the principles used in the study of the P450s and for understanding the mechanisms of their inhibition are applicable, in general, to all of the other drug-metabolizing enzymes. Thus, although the focus will be on P450s, the general principles described here are applicable to all drug-metabolizing enzymes.

## 2 MECHANISMS OF ENZYME INHIBITION

The inhibitors of the drug-metabolizing enzymes can be divided into three general categories that differ in their mechanisms. They are (i) compounds that bind reversibly in the active site or an effector site of the enzyme, (ii) compounds that form quasi-irreversible complexes with a prosthetic group of the enzyme in question, and (iii) compounds that bind irreversibly to the prosthetic group of the protein or to the apoprotein itself (Guengerich, 1999). The most common cause of drug–drug interactions is generally believed to be reversible inhibition. Reversible inhibition includes competitive, noncompetitive, uncompetitive, product inhibition, inhibition by transition-state analogs, and inhibition by slow, tight-binding inhibitors. Reversible inhibition may occur rapidly following a single dose of the inhibitory drug. It is transient and, following elimination of the inhibitor from the body, the normal metabolic functions of the enzymes will be regained. As opposed to the quasi-irreversible and irreversible inhibitors that exhibit both dose-dependent and time-dependent inhibition of catalytic activity, reversible inhibitors generally only exhibit dose-dependent inhibition of activity.

## 3 COMPETITIVE INHIBITION

Competitive inhibition occurs when the inhibitor binds reversibly to the enzyme and prevents the binding of the substrate to the catalytically active site of the enzyme. This is generally believed to result from the fact that the inhibitor shares some degree of structural similarity with the drug whose metabolism is being inhibited. The inhibitor itself may or may not be metabolized by the enzymes that it inhibits. Depending on the substrate specificity of the drug-metabolizing enzyme in question, some of which exhibit relatively little apparent specificity, the structural similarities between the substrate whose metabolism is inhibited and the competitive inhibitor may or may not be apparent. Competition of the inhibitor with the drug whose metabolism is inhibited for occupancy of the active site of the drug-metabolizing enzyme, may involve simple competition for binding to a lipophilic or hydrophilic domain in the active site, or it may involve the formation of hydrogen bonds or ionic bonds with specific amino acid residues in the active site (Ortiz de Montellano and Correia, 2005). In some cases, a competitive inhibitor may bind reversibly to a site other than the active site and block

$$E + S \underset{k_{-1}}{\overset{k_1}{\rightleftarrows}} ES + I \underset{k_{-2}}{\overset{k_2}{\rightleftarrows}} ES^* \underset{k_{-3}}{\overset{k_3}{\rightleftarrows}} EP \underset{k_{-4}}{\overset{k_4}{\rightleftarrows}} E + P$$

$$E + I \underset{k_{-1'}}{\overset{k_{1'}}{\rightleftarrows}} EI$$

**Scheme 1**

$$v = \frac{V_{max}\,[S]}{K_M\,(1 + [I]/K_I) + [S]}$$

**Scheme 2**

substrate binding to the active site. Competitive inhibition is commonly observed when two different substrates of the same drug-metabolizing enzyme are present in the *in vitro* system under study or are given at the same time to the patient. The classic enzyme kinetic concept for this process is depicted in Scheme 1.

The intermediate depicted by ES* may or may not be present in any appreciable concentration. The kinetic constants $k_1$, $k_{1'}$, and $k_{-4}$ may be lumped together and termed "$k_{on}$" and $k_{-1}$, $k_{-1'}$, and $k_4$ may be lumped together as the "$k_{off}$." For classic competitive inhibition, the steady-state apparent $K_M$ value for the reaction of the substrate under study will increase by the factor $(1+[I]/K_I)$ in the presence of the inhibitor while the $V_{max}$ does not change. The inhibition constant, $K_I$, for a given compound can be estimated from steady-state kinetic studies with increasing concentrations of the competitive inhibitor since the apparent $K_M$ value and the $K_I$ are related by the expression in Scheme 2, where the apparent $K_M$ is equal to $K_M\,(1+[I]/K_I)$.

The most commonly used method for characterizing inhibition data is to systematically vary the concentrations of both the substrate and the inhibitor. Initial estimates for the kinetic parameters can be obtained from data for control studies run in the absence of inhibitor and from a double-reciprocal plot of the inhibition data. These analyses provide preliminary estimates of $V_{max}$ (the maximal velocity at saturating substrate concentrations), $K_M$, and $K_I$. Once an estimate of the $K_M$ for the reaction is known and competitive inhibition is assumed initially, the $K_I$ can be determined by varying the concentration of the inhibitor at a single-substrate concentration. Data obtained by varying the concentration of the inhibitor at several different single-substrate concentrations can be used to provide information regarding the $K_M$ and the $K_I$ as well as the type of inhibition. When both substrate and inhibitor concentrations are varied, the data can be fit to equations for not only competitive inhibition but also noncompetitive, uncompetitive, and mixed type and the fits can be compared. In general, the most prevalent type of inhibition is competitive with the partial mixed type of inhibition being the second most prevalent.

Competitive inhibition can be identified by a common intersection point on the ordinate in either an Eadie–Hofstee (V versus V/S) or Lineweaver–Burk (1/V versus 1/S) plot. An alternative method involves using the Dixon plot, where 1/V is plotted versus the concentration of the substrate, S (Bronson *et al.*, 1995). Since the weighting of the data points is a problem for all of the linear transforms of kinetic data

(Cornish-Bowden, 1979), various nonlinear regression programs have been developed and are readily available for analysis of data obtained from kinetic studies on various drug-metabolizing enzymes (EZ-Fit, GraFit, KaleidaGraph, SigmaPlot, etc.).

In principle, for studies involving reversible inhibition, the $K_I$ is equivalent to the binding constant for the inhibitor to the enzyme ($K_D$), whereas the $K_M$ usually is not equivalent to the binding constant but is composed of a variety of kinetic terms. However, when the inhibitor is also a substrate for the enzyme it inhibits and the competitive inhibition is due to the fact that the two substrates are competing with each other for metabolism, then the $K_I$ is not equal to the actual binding constant ($K_D$) of the inhibitor.

The *in vitro* assays are most reliable when used with either purified enzymes in the appropriate reconstituted system or with microsomal or cytosolic preparations, in which there is only one enzyme responsible for metabolizing the probe substrate to the product that is measured, and the enzyme under study is the only one responsible for binding the compound being investigated as the inhibitor. Since the concentrations of marker substrates and inhibitors added to the growth medium of cells in culture do not necessarily reflect the intracellular concentrations experienced by the drug-metabolizing enzymes, studies aimed at determination of the type of inhibition are difficult to perform with freshly isolated hepatocytes, and the interpretation of data obtained from these studies with respect to the calculation of meaningful kinetic constants and the type of inhibition is suspect.

When using microsomal preparations or purified reconstituted enzyme systems *in vitro*, competitive inhibition is observed relatively frequently in these types of studies. It is generally presumed that this is also the case *in vivo* and approaches have been developed for characterizing inhibitory interactions *in vivo* by analyzing various pharmacokinetic parameters (Black *et al.*, 1996). Since many of the most important drug-metabolizing enzymes are relatively nonspecific and have a variety of different drugs as substrates, competition of several drugs for metabolism by a specific drug-metabolizing enzyme is a common occurrence leading to drug–drug interactions in patients who are simultaneously administered a number of different drugs.

## 4 NONCOMPETITIVE INHIBITION

Some drugs can inhibit the drug-metabolizing enzymes even when they are not metabolized by the enzyme that they affect. A simple noncompetitive inhibitor binds to a site on the enzyme that is distinct from that of the substrate and thus, it is not metabolized by that enzyme and does not alter substrate binding and vice versa. The substrate and the inhibitor bind reversibly and independently. Although the bound inhibitor has the effect of apparently decreasing the amount of enzyme present (decrease in $V_{max}$ with no change in $K_M$), it does not inactivate the enzyme. In this case, the enzyme–substrate–inhibitor complex is formed; however, it is unable to function catalytically. This type of inhibition, although discussed routinely in any presentation of enzyme kinetics, is relatively uncommon and is not routinely encountered in studies of the drug-metabolizing enzymes. This type of behavior might be observed in studies with microsomes or expressed enzymes when some of the enzyme involved in metabolism is inactivated during the course of the study, for example, by the formation of reactive intermediates which inactivate the protein or by the presence of a tight-binding inhibitor, and can be differentiated from these possibilities by approaches described later.

## 5 UNCOMPETITIVE INHIBITION

Similar to noncompetitive inhibition, uncompetitive inhibition is readily defined but is rarely observed in drug metabolism studies done *in vitro*. In uncompetitive inhibition, the inhibitor does not bind to the free enzyme but, in fact, only binds to the enzyme that already has substrate bound and prevents subsequent metabolism. This favors the equilibrium shown in Scheme 1 to form the ES complex and in this case leads to the formation of a nonproductive enzyme–substrate–inhibitor complex. Therefore, both the $V_{max}$ and the $K_M$ decrease proportionately so that the ratio of the $V_{max}$ to $K_M$ remains constant. In a Lineweaver–Burk plot of the data, parallel lines would be observed for plots of 1/V versus 1/S at different concentrations of the inhibitor.

The easiest way to estimate the inhibition constants for these classical mechanisms involves determining the initial rates of metabolism at various substrate concentrations in the absence and in the presence of a single concentration of the inhibitor under study. From each of the double-reciprocal plots, the values $V_{max}$ and $K_M$ are determined graphically or by linear regression. Subsequently, the calculations may be performed using the relationship between $K_M$ in the presence of the inhibitor and $K_I$ from Scheme 2 (competitive inhibition), Scheme 3 (noncompetitive inhibition), and Scheme 4 (uncompetitive inhibition).

## 6 PRODUCT INHIBITION

Due to the relative lack of specificity of many of the drug-metabolizing enzymes and their ability to metabolize a substrate on several different parts of the parent molecule to give different metabolites, it is possible that a product of an initial reaction catalyzed by a drug-metabolizing enzyme may bind to the enzyme again in the same or a different orientation and inhibit the initial reaction for the metabolism of another substrate molecule. Thus, the product may have physical characteristics that are not significantly different from that of the initial substrate and it could thereby bind competitively and inhibit metabolism of the parent compound either by serving as a substrate for a second reaction or by just binding in the substrate binding site. For example, cytochrome P450 2E1 hydroxylates benzene to give the phenol and then hydroxylates it once again to form the hydroquinone (in this case, the phenol competes with the benzene for metabolism and vice versa). It is possible in some cases that the first product may have a greater affinity for binding to the active site than the parent drug.

$$v = \frac{V_{max}\,[S]}{[S]\,(1 + [I]/K_I) + K_M}$$

**Scheme 3**

$$v = \frac{V_{max}\,[S]}{([S] + K_M)\,(1 + [I]/K_I)}$$

**Scheme 4**

## 7 TRANSITION-STATE ANALOGS

Transition-state analogs resemble the transition state for the enzymatic reaction catalyzed by the enzyme. That is to say, they resemble the transient complex formed within the catalytic cycle that has the maximum free energy. These compounds are tight-binding noncovalently bound inactivators in which the $k_{on}$ rate is rapid and the $k_{off}$ is slow. The inactivation of the enzyme occurs rapidly and, under typical assay conditions, there is no time dependence. In principle, the enzyme activity can be restored by removal of the inhibitor by a variety of approaches including gel filtration, centrifugal separation, or dialysis. In this case, the kinetics are characterized by a decrease in the $V_{max}$ but no alteration in the $K_M$. Transition-state inhibitors have been developed to target a variety of different clinically important enzymes since the transition state tends to be uniquely characteristic of one kind of enzyme reaction, whereas substrates may be shared by two or more different enzymes. So far, transition-state analogs have been of relatively minor importance in the study or the targeting of drug-metabolizing enzymes.

## 8 SLOW, TIGHT-BINDING INHIBITORS

These compounds are characterized by the fact that they exhibit relatively slow $k_{on}$ rates but have even slower $k_{off}$ rates. Thus, they have very high affinities for the target enzyme. In these cases, the binding may be noncovalent or possibly even covalent, but due to the slowness of the binding, the decrease in the activity of the enzyme may be time-dependent and can easily be mistaken for mechanism-based inactivation. However, in this case, gel filtration, centrifugal separation, or dialysis may be used to restore enzyme activity, even though the process may be slow. Although rapid formation of a complex between the inhibitor and the enzyme may occur initially, the tight binding of the inhibitor to the enzyme resulting in inhibition may require a conformational change of the enzyme as a result of the initial binding, a displacement of a water molecule in the enzyme active site, a change in the protonation state of an amino acid residue in the active site, or reversible formation of a covalent bond (Silverman, 1995).

## 9 MECHANISM-BASED INACTIVATORS

Specific covalent binding of reactive intermediates formed by the drug-metabolizing enzymes within the active sites of the enzymes that initially form them is oftentimes referred to as “mechanism-based,” “catalysis-dependent,” or “suicide” inactivation. A broad definition of mechanism-based or suicide inactivators would include any compound that is activated via the normal catalytic mechanism of the enzyme that it subsequently inactivates. However, we will restrict the definition to that of Silverman (1995), who defines it as “an unreactive compound the structure of which resembles that of either the substrate or the product of the target enzyme and that undergoes a catalytic transformation by the enzyme to a species that, before release from the active site, inactivates the enzyme.” Key to this concept is the requirement that the inactivation occurs without the release of the reaction product from the enzyme active site. This definition rules out other types of inactivators such as affinity labels, transition-state analogs, and slow, tight-binding inhibitors.

Although mechanism-based inactivation is generally thought to be a relatively unusual occurrence for enzyme-catalyzed reactions, it has been observed more frequently than might be expected with the cytochrome P450s. Presumably, this results from the increased reactivity of many of the oxygenated intermediates formed by these enzymes during the course of the oxygenation reactions. Thus, mechanism-based inactivation has been encountered with P450s as well as other drug-metabolizing enzymes with a variety of different drugs. The utility of mechanism-based inactivators in the design of new drugs has attracted significant interest recently since, in principle, these compounds could be designed so that they would only inhibit the target enzyme. They are also of significant interest because of their potential for use in elucidating mechanisms catalyzed by the drug-metabolizing enzymes. In addition, a great deal of effort has been put into the development of mechanism-based inactivators as diagnostic inhibitors of the various drug-metabolizing enzymes in order to identify which enzymes may be involved in catalyzing a particular reaction. A primary focus of many of these studies has been aimed at identifying which form(s) of the various cytochrome P450s are involved in catalyzing a particular reaction in microsomal preparations. Following catalytic formation of the reactive intermediate, these compounds result in the formation of covalent bonds to amino acid residues or prosthetic groups in the enzyme active site that cannot be broken to regenerate the catalytically active enzyme. Although the onset of inhibition by these compounds either *in vivo* or *in vitro* may occur more slowly than that observed with reversible or quasi-irreversible inhibitors, the final effect of the mechanism-based inactivators is generally much more profound than observed with other inhibitors since the inhibition of drug metabolism can only be reversed *in vivo* by the synthesis of new, catalytically active enzymes. Due to the formation of a covalent bond, gel filtration, dialysis, or separation by centrifugation is not effective in regenerating the catalytic activity.

The relevant kinetic scheme for mechanism-based inactivation is shown in Scheme 5. In this scheme, MBI is the mechanism-based inactivator, E-MBI is the initial Michaelis–Menten enzyme–substrate complex, E-MBI* is the reactive intermediate complexed to the enzyme, E-MBI** is the inactivated enzyme, and P is the product that leaves the enzyme. The derivation and the kinetic implications of the various kinetic constants have been described previously by Silverman (1995). Parameters that are routinely determined to describe mechanism-based inactivators include $k_{inact}$, $K_I$, and the partition ratio. The $k_{inact}$ and the $K_I$ are defined in a manner analogous to the k and $K_M$ in standard enzyme-catalyzed reactions. The partition ratio, a concept introduced by Walsh (Wang and Walsh, 1978), is determined from the ratio of $k_3/k_4$ and is a measure of the number of latent inactivator molecules metabolized and released as product for each inactivator molecule metabolized to give a form that reacts with the active enzyme to inactivate it. It can be thought of as the number of cycles of metabolism that can be traversed by the enzyme, on average, before it is inactivated by the mechanism-based inactivator. The partition ratio can also be considered to be a measure of the efficiency of the mechanism-based inactivator and is dependent on a number of factors including

$$\begin{array}{ccccccc} & k_1 & & k_2 & & k_4 & \\ \text{E + MBI} & \rightleftarrows & \text{E} \cdot \text{MBI} & \rightarrow & \text{E} \cdot \text{MBI*} & \rightarrow & \text{EMBI**} \\ & k_{-1} & & & \downarrow k_3 & & \\ & & & & \text{E + P} & & \end{array}$$

**Scheme 5**

the reactivity of the reactive intermediate, the proximity of the appropriate target molecule(s) (nucleophile, electrophile, or radical) in the active site to the site of formation of the reactive intermediate, the rate of diffusion of the reactive intermediate from the active site of the enzyme, and the presence in the active site of the enzyme of other molecules which may react with the reactive intermediate and render it harmless (e.g. water, or low-molecular-weight nucleophiles or electrophiles). It is not dependent on the concentration of the protein or inactivator. Partition ratios ranging from almost zero to several thousand have been reported for various mechanism-based inactivators. The most efficient mechanism-based inactivator would have a partition ratio of zero in that every reactive intermediate formed would inactivate the enzyme. In this case, in a standard assay where metabolite formation is measured, the conclusion would be that the inactivator is not a substrate for the enzyme since there would be no measurable product formation.

As pointed out by Ortiz de Montellano and Correia (2005), mechanism-based inactivators may exhibit much better specificity for their target enzymes than reversible inhibitors since (i) the initial step(s) for the binding of the inhibitor to the specific enzyme must satisfy all of the constraints imposed on reversible inhibitors; (ii) the inhibitor must also be acceptable as a substrate for the enzyme due to the requirement that it must undergo catalytic activation to form a reactive species; and (iii) the reactive species formed as a result of metabolism by the target enzyme results in an irreversible modification of that enzyme which permanently removes it from the pool of active enzymes.

A variety of different criteria are routinely used to assess whether a substrate is a mechanism-based inactivator of an enzyme (Silverman, 1995; Kent, Jushchyshyn and Hollenberg, 2001; Ortiz de Montellano and Correia, 2005). The criteria include:

1. The loss of enzyme activity must exhibit first-order kinetics. Thus, a plot of the logarithm of the percent enzyme activity remaining versus time should give a straight line. The slope of the straight line gives the $k_{app}$, the rate constant for the inactivation reaction at the concentration of the inactivator used.
2. The inactivation should exhibit saturation kinetics with respect to the concentration of the inactivator. The plot of $k_{app}$ versus [I] should be hyperbolic and a linear transformation of the data (e.g. a plot of $1/k_{app}$ versus 1/[I]) can be used to calculate $K_I$, the concentration of the inactivator which gives the half-maximal rate of inactivation and the $k_{inact}$, the rate of inactivation at saturating concentrations of inactivator. The half-life, $t_{1/2}$, can be calculated from the relationship $t_{1/2} = 0.693/k_{inact}$.
3. The inactivation requires that all of the typical cofactors be present and that metabolism is occurring. It should be possible to demonstrate metabolism of the inactivator to a reactive intermediate which can then lead to the formation of product or to inactivation.
4. In the presence of one of its typical substrates, the enzyme should be protected from inactivation and thus show a slower rate of inactivation.
5. The inactivation should be irreversible. It should not be possible to show a regain of activity upon dialysis, gel filtration, or centrifugation since the inhibitor should be covalently bound to the apoprotein or to a prosthetic group of the enzyme.
6. Following inactivation, there should be a 1 : 1 stoichiometry of inactivator bound to enzyme molecule inactivated.

7. The presence of exogenous "scavenger" nucleophiles should not cause a decrease in the rate of inactivation or the stoichiometry of the reaction.
8. The inactivation should not exhibit any lag time. Some substrates of drug-metabolizing enzymes are converted to reactive intermediates that may leave the enzyme active site and form covalent adducts with proteins as well as with DNA or RNA. Therefore, it is necessary to differentiate between specific inactivation by mechanism-based inactivators and nonspecific inactivation which may result from extensive modification of the protein by reactive intermediates released from the active site which bind nonspecifically to the protein and eventually will lead to inactivation. Oftentimes, careful analysis of the time course for the reaction is required in order to differentiate between the two, since the latter case often exhibits some lag in the inactivation as the level of modification of the enzyme must increase to a certain point before significant inactivation is observed.

Under optimal conditions, all of these criteria will be readily observable for mechanism-based inactivator of a drug-metabolizing enzyme. However, not all of these tests may be applicable in every situation.

Experiments with mechanism-based inactivators do not need to be restricted to purified enzymes. Initial studies may be done with relatively crude enzyme preparations (e.g. cytosolic fractions or microsomes) using probe substrates demonstrated to be specific for the individual enzymes under investigation in order to monitor inactivation. Radioactive labeling studies can also be performed with such preparations followed by SDS-PAGE (sodium dodecyl sulfate-polyacrylamide gel electrophoresis) or HPLC (high performance liquid chromatography) separation of the enzymes of interest to examine the specificity and stoichiometry of the labeling.

## 10 INHIBITORS THAT ARE METABOLIZED TO REACTIVE PRODUCTS THAT COVALENTLY ATTACH TO THE ENZYME

This group of inhibitors may oftentimes be grouped with the mechanism-based inactivators previously discussed. Some of the same criteria may apply, such as saturation kinetics with respect to the concentration of the inactivator, the requirement for the presence of all of the typical cofactors and that metabolism is occurring, protection from inactivation by normal substrates, and the irreversible nature of the inactivation. However, the stoichiometry for the inactivation may be much greater than one, the presence of exogenous "scavengers" such as glutathione may affect both the rate of inactivation and the stoichiometry and the inactivation reaction may exhibit a lag time. Thus, the kinetics may not show first-order loss of activity with respect to time. However, it is important to realize that deviations from pseudo-first-order kinetics may not be easy to detect.

## 11 SUBSTRATE INHIBITION

Substrate inhibition kinetics are observed when a substrate causes a decrease in the rate of product formation as its concentration increases above a certain level. Although the exact mechanism for this type of inhibition with drug-metabolizing enzymes is not well understood, it has been observed in some cases, particularly with the cytochrome

P450s (Shou *et al.*, 2001). In classical enzymology, this behavior is oftentimes seen with enzymes that function using a ping-pong mechanism. For cytochrome P450s, this has been explained by proposing a two-site model (Shou *et al.*, 2001). In this model, the substrate can bind to the active site to yield a productive complex and substrate can also bind to an inhibitory site which is assumed to be nonproductive. When both sites are occupied with substrates, the turnover leading to product formation is no longer as rapid as that seen with the initial binding of the substrate to the productive site. In this model, the $K_S$ (the binding constant) for substrate binding to the catalytic site is generally smaller than the $K_I$ for the binding of the substrate to the inhibitory site, indicating that the substrate has a binding affinity to the catalytically active site greater than that for binding to the inhibitory site.

## 12 PARTIAL INHIBITION

Partial inhibition is observed when the enzyme is saturated with inhibitor but the inhibition is not complete. This phenomenon has been primarily observed in P450-catalyzed reactions and corresponds to a mixed-type inhibition. The cause of the partial inhibition is the formation of a Michaelis–Menten enzyme–substrate complex which also has the inhibitor bound and yet it is still catalytically active. The kinetics of this reaction can best be understood in terms of the substrate binding in its preferred binding site in the active site, and in terms of the inhibitor binding either in the active site, to a site where it does not necessarily compete with the substrate, or outside of the active site, in some way affecting the catalytic activity. Partial inhibition may result either from steric hindrance that hampers optimal binding of the substrate to its binding site in the active site or from a conformational change of the enzyme.

## 13 INHIBITION OF CYTOCHROME P450 ENZYMES

As discussed already, enzyme inhibition is a very important issue in drug discovery and development as well as in the clinic. The inhibitory effects of one drug on the disposition, efficacy, or toxicity of another drug have major implications for the continued development and use of that entity. In many cases, drug–drug interactions can result in alterations in the pharmacokinetics of one or both agents. Although enzyme inhibition can affect all of the various drug-metabolizing enzymes through a variety of mechanisms, the most common drug interactions generally occur through the cytochrome P450 system since they play such an important role in the metabolism of many drugs. Therefore, the rest of this chapter will focus on various aspects of the inhibition of these enzymes and we will now address some things that are more specific to the P450s; however, in many cases, the topics discussed are applicable to other drug-metabolizing enzymes although maybe not as widely as those areas discussed previously.

Generally, the metabolism of a drug by a cytochrome P450 represents the rate-limiting step in its metabolism; therefore, inhibition of the P450s is recognized by the FDA and other regulatory agencies as an important cause of drug–drug interactions (US Food and Drug Administration, Center for Drug Evaluation and Research, 1997). The catalytic cycle for drug metabolism by the cytochrome P450s provides for a number

of potential points at which inhibition of metabolism may occur. The primary steps in the catalytic cycle for cytochrome P450-catalyzed reactions are (i) binding of the substrate; (ii) one-electron reduction of the ferric ($Fe^{+3}$) enzyme to the ferrous ($Fe^{+2}$) form by the NADPH-cytochrome P450 reductase; (iii) binding of the molecular oxygen by the ferrous ($Fe^{+2}$) iron; (iv) transfer of the second electron from the reductase to the ferrous-oxy-substrate complex with the subsequent release of one molecule of water and the formation of an activated oxygen intermediate; (v) insertion of the activated oxygen into the substrate to form the oxygenated product; and (vi) release of the oxygenated product from the enzyme leading to the regeneration of the native ferric ($Fe^{+3}$) form of the enzyme that can now bind another molecule of drug and undergo another catalytic cycle (White and Coon, 1980). Three of the preceding steps: (i) binding of the substrate, (ii) binding of the molecular oxygen to the ferrous ($Fe^{+2}$) enzyme, and (iii) the catalytic step in which the activated oxygen is transferred from the heme iron to the enzyme-bound substrate, appear to be particularly susceptible to inhibition (Ortiz de Montellano and Correia, 2005). Another potential target for inhibition of cytochrome P450-catalyzed drug metabolism is the transfer of electrons from the P450 reductase to the cytochrome P450, which occurs at two points in the catalytic cycle (after the addition of the substrate and after the binding of molecular oxygen to the ferrous P450). In general, inhibitors of these two steps interfere with the transfer of electrons from the reductase to the P450 by accepting electrons directly from the reductase and therefore they do not cause inhibition by interaction directly with the P450 protein. Thus, they are relatively nonspecific with respect to the cytochrome P450 forms that they inhibit and have the potential to inhibit all P450-catalyzed drug metabolism (Hollenberg, 2002).

## 14 REVERSIBLE INHIBITORS

The cytochrome P450s can be inhibited by all of the mechanisms already described in this chapter. The inhibitors of the P450s can be divided into three general categories that differ in their mechanisms: (i) inhibitors that bind reversibly to the enzyme; (ii) inhibitors that form quasi-irreversible complexes with either the ferric ($Fe^{+3}$) or ferrous ($Fe^{+2}$) iron of the heme prosthetic group; and (iii) inhibitors that bind irreversibly to the apoprotein or to the prosthetic heme, or that cause covalent binding of the prosthetic heme or a degradation product of the heme to the apoprotein. In general, reversible inhibitors interfere with the catalytic cycle prior to the formation of the activated oxygen intermediate and they may be competitive, noncompetitive, uncompetitive, product, or transition-state inhibitors. Those inhibitors that act during or subsequent to the formation of the activated oxygen intermediate are generally either quasi-irreversible or irreversible inhibitors.

As already indicated, some reversible inhibitors of the P450s may act by coordinating with the iron in the prosthetic heme atom. Coordination of a strong ligand to the ferric ($Fe^{+3}$) heme iron shifts the iron from the high- to the low-spin form giving rise to the "type II" difference spectrum (Schenkman, Sligar and Cinti, 1981). This change in the spin state occurs concomitantly with a change in the redox potential of the P450 that makes the reduction by the P450 reductase more difficult (Guengerich, 1983). Conversely, binding of a substrate to the active site of the P450 often shifts the heme iron to a more high-spin form causing a change in the redox potential which facilitates reduction. Thus, the inhibition of P450s by ligands to the heme iron which cause a shift

in the spin state to a more low-spin form is a result not only of the occupation of the sixth coordination site of the heme iron by the ligand, but also the change in the reduction potential of the iron.

Inhibitors that bind to both the iron in the heme and to a hydrophobic binding site in the P450 active site are generally much more effective inhibitors than those that utilize only one of those interactions (Ortiz de Montellano and Correia, 2005). A number of nitrogen-containing aliphatic and aromatic compounds exhibit these properties and are relatively potent reversible inhibitors. Some of the most widely used compounds in this class are derivatives of imidazole and pyridine.

## 15 QUASI-IRREVERSIBLE INHIBITORS

Another category of inhibitors includes those compounds that inhibit by forming quasi-irreversible complexes with the iron atom in the heme prosthetic group. In general, these compounds require initial activation by the enzyme resulting in the formation of transient intermediates that can then coordinate very tightly to the prosthetic heme in the P450 active site leading to inhibition. Although these linkages are not covalent, the coordination is generally so tight that the inhibitory complexes can only be broken down to release the native catalytically active enzyme under special experimental conditions. Several different classes of compounds have been shown to be capable of forming these metabolic intermediates that coordinate so effectively with the P450 that they are converted to a catalytically nonfunctional state. These classes include compounds containing a dioxymethylene function and nitrogen-containing compounds including acyl hydrazines, 1,1-disubstituted hydrazines, and a variety of alkyl amines that can be oxidized to nitroso compounds by the P450s. Aryl and alkyl methylenedioxy compounds are oxidized by the P450s to yield species that can coordinate very tightly to the prosthetic heme iron, leading to the formation of metabolite intermediate complexes (MI complexes) that are extremely stable. The requirement for an initial oxidation by the P450 has been unequivocally demonstrated (Franklin, 1971). The resulting ferrous MI complex is extremely stable as demonstrated by the fact that it can be isolated intact from rats treated with isosafrole. The ferric MI complex is much less stable than the ferrous complex and incubation with various lipophilic molecules can result in displacement of the inhibitor from the active site, leading to regeneration of the native catalytically active enzyme (Elcombe *et al.*, 1975). The chemical nature of the side-chain substituent on the methylenedioxy compound plays an important role in determining the stability of the MI complex. Both the lipophilicity and the size of the alkyl side chain are important since alkyl chains of one to three carbon atoms lead to the formation of relatively unstable complexes while side chains containing larger alkyl groups result in the formation of stable MI complexes (Murray, Hetnarski and Wilkinson, 1985).

A second class of compounds that can form MI complexes with the P450 heme iron is the relatively large class composed of aromatic and alkyl amines. Included in this class are a number of clinically important antibiotics such as troleandomycin and erythromycin (Mansuy *et al.*, 1977). Oxidation of these amines results in the formation of intermediates that can coordinate very tightly to the ferrous heme resulting in optical spectra exhibiting an absorbance maxima in the range of 445–455 nm. Following oxidation, primary amines are capable of forming these types of MI complexes although secondary or tertiary amines are not able to do this. However, secondary and tertiary

amines may also form MI complexes following conversion to primary amines by *N*-dealkylation. The primary amines initially undergo hydroxylation followed by a two-electron oxidation of the resulting hydroxylamine to give the *C*-nitroso group, which currently is thought to be responsible for chelation with the iron to form the MI complex (Mansuy *et al.*, 1977). Although the formation of MI complexes might be considered to fall into the category of mechanism-based inactivation, a better description would probably be transformation of the substrates by the enzyme(s) to slow, tight-binding inhibitors since the linkages are not covalent and oxidation by an oxidant such as $Fe(CN)_6^{3-}$ will oxidize the ferrous complex and release the ligand, thereby regenerating the native ferric enzyme which is now fully competent to carry out metabolism.

## 16 MECHANISM-BASED INACTIVATORS

Over the past decade, considerable interest has developed with respect to mechanism-based inactivators of the cytochrome P450s. Three general classes of mechanism-based inactivators of the cytochrome P450s have been identified. These classes consist of compounds that (i) bind covalently to the P450 apoprotein, (ii) bind irreversibly to the prosthetic heme group, and (iii) cause covalent binding of the heme prosthetic group or its degradation product to the apoprotein (Osawa and Pohl, 1987). These modifications, when they occur *in vivo*, appear to greatly accelerate the degradation of the inactivated protein. It is now clear that many mechanism-based inactivators may inactivate a given P450 by more than one mechanism and may inactivate different forms of P450s by different mechanisms. Thus, the predominant way in which a mechanism-based inactivator inactivates a P450 may be determined by a number of factors including the identity of the form of P450 being studied. So far, the factors that determine how a specific mechanism-based inactivator will modify a particular form of P450 are not well understood.

A wide variety of different types of compounds have been found to be excellent mechanism-based inactivators of the cytochrome P450s. These include (i) acetylenes and terminal aryl and alkyl olefins such as 2-ethynylnaphthalene, 5-phenyl-1-pentyne, 9-ethynylphenanthrene, 10-dodecynoic acid, 17α-ethynylestradiole, gestodene, mifepristone (RU486), and secobarbital; (ii) a variety of different organosulfur compounds such as carbon disulfide, disulfiram, cimetidine, thiophenes, mercaptosteroids, isothiocyanates, thioureas, parathion, dialkylsulfides, and diethyldithiocarbamate; (iii) furanocoumarins such as 8-methoxypsoralen, bergamottin, and L-754,394 (a Merck compound synthesized as a potential inhibitor of the HIV protease); (iv) 1-aminobenzotriazole and its *N*-aralkylated derivatives; and (v) halogenated compounds such as chloramphenicol and various *N*-monosubstituted dichloroacetamides (Kent, Jushchyshyn and Hollenberg, 2001; Ortiz de Montellano and Correia, 2005).

The marked lack of substrate specificity of many of the P450s involved in drug metabolism, as evidenced by the differences in the sizes and the chemical properties of their substrates as well as the number of different substrates that they can metabolize, speaks to the apparent flexibility of the active sites of these enzymes. This apparent promiscuity of the P450s offers the possibility of designing a multitude of mechanism-based inactivators having a variety of different potentially reactive moieties. Thus, it would seem to be a relatively straightforward exercise to design mechanism-based inactivators that will inhibit a desired P450. However, the problem in designing

mechanism-based inactivators for P450s lies not in designing one that will inactivate a P450 but in designing one that will be specific for a single form of P450.

The development of reversible, quasi-irreversible, and irreversible inhibitors of the P450s involved in the metabolism of drugs, other xenobiotics, and endogenous substrates such as steroid hormones, and our understanding of the mechanisms of action of these various types of inhibitors have increased remarkably over the past decade and have provided important insights that can be used to develop highly selective isozyme-specific inhibitors of these P450s. The development of highly selective inhibitors of P450s is of great importance not only for studies aimed at probing the three-dimensional structures of the proteins, their mechanisms of action, and the biological roles of the various P450s, but also because of their potential as modulators of cytochrome P450 activity that can be used as therapeutic agents in a fashion similar to the way that inhibitors of the P450s involved in steroid metabolism have been used to treat endocrine disorders and also as anticancer agents. In addition, since different P450s play major roles in the metabolic activation and detoxication of a variety of different chemical carcinogens as well as other toxins, the development of inhibitors that can be used to selectively inactivate these enzymes is of great interest. These types of inhibitors have the potential to shift the balance between various pathways for the metabolism of a chemical carcinogen or toxin such that those pathways involved in metabolic activation will be minimized, whereas those pathways leading to detoxication will be enhanced. Thus, these agents would be of great value in improving the quality of human life. In addition, a better understanding of inhibition mechanisms and selectivity may lead to more efficient screening protocols for drug–drug interactions in the pharmaceutical industry.

## REFERENCES

Black, D.J., Kunze, K.L., Wienkers, L.C., Gidal, B.E., Seaton, T.L., McDonnell, N.D., Evans, J.S., Bauwens, J.E., Trager, W.F. (1996). Warfarin-fluconazole II. A metabolically based drug interaction: in vivo studies. *Drug Metab. Dispos.*, *24*, 422.

Bronson, D.D., Daniels, D.M., Dixon, J.T., Redick, C.C., Haaland, P.D. (1995). Virtual kinetics: using statistical experimental design for rapid analysis of enzyme inhibitor mechanisms. *Biochem. Pharmacol.*, *50*, 823.

Cornish-Bowden, A. (1979). *Fundamentals of Enzyme Kinetics*, Butterworths, London.

Elcombe, C.R., Bridges, J.W., Gray, T.J.B., Nimmo-Smith, R.H., Netter, K.J. (1975). Studies on the interaction of safrole with rat hepatic microsomes. *Biochem. Pharmacol.*, *24*, 1427.

Franklin, M.R. (1971). The enzymic formation of a methylenedioxyphenyl derivative exhibiting an isocyanide-like spectrum with reduced cytochrome P-450 in hepatic microsomes. *Xenobiotica*, *1*, 581.

Guengerich, F.P. (1983). Oxidation-reduction properties of rat liver cytochromes P450 and NADPH-cytochrome P450 reductase related to catalysis in reconstituted systems. *Biochemistry*, *22*, 2811.

Guengerich, F.P. (1999). Inhibition of drug metabolizing enzymes: molecular and biochemical aspects, in *Handbook of Drug Metabolism* (ed. T.F. Woolf), Marcel Dekker, New York, p. 203.

He, K., Iyer, K., Hayes, R.N., Sing, M.W., Woolf, T.F., Hollenberg, P.F. (1998). Inactivation of P450 3A4 by bergamottin, a component of grapefruit juice. *Chem. Res. Toxicol.*, *11*, 252.

Hollenberg, P.F. (2002). Characteristics and common properties of inhibitors, inducers and activators of CYP enzymes. *Drug Metab. Rev.*, *34*, 17.

Kent, U.M., Jushchyshyn, M.I., Hollenberg, P.F. (2001). Mechanism-based inactivators as probes of cytochrome p450 structure and function. *Current Drug Metab.*, *2*, 215.

Kohler, G.I., Bode-Boger, S.M., Busse, R., Hoopman, M., Welete, T., Boger, R. (2000). Drug-drug interactions in medical patients: effects of in-hospital treatment and relation to multiple drug use. *Int. J. Clin. Pharmacol. Ther.*, *38*, 504.

Levy, M., Kewitz, H., Altwein, W., Hillebrand, J., Eliakim, M. (1980). Hospital admissions due to adverse drug reactions: a comparative study from Jerusalem and Berlin. *Eur. J. Clin. Pharmacol.*, *17*, 25.

Lin, J.H., Lu, A.Y.H. (1998). Inhibition and induction of cytochrome P450 and the clinical implications. *Clin. Pharmacokinet.*, *35*, 361.

Mansuy, D., Gans, P., Chottard, J.-C., Bartoli, J.-F. (1977). Nitrosoalkanes as Fe(II) ligands in the 455-nm-absorbing cytochrome P-450 complexes formed from nitroalkanes in reducing conditions. *Eur. J. Biochem.*, *76*, 607.

Murray, M., Hetnarski, K., Wilkinson, C.F. (1985). Selective inhibitory interactions of alkoxymethylene-dioxybenzenes towards mono-oxygenase activity in rat hepatic microsomes. *Xenobiotica*, *15*, 369.

Ortiz de Montellano P.R. and Correia, M.A. (2005). Inhibition of cytochrome P450 enzymes, in *Cytochrome P450. Structure, Mechanism and Structure*, 3rd edn (ed. P.R. Ortiz de Montellano), Kluwer Academic/Plenum Press, New York, p. 247.

Osawa, Y., Pohl, L.R. (1987). Covalent binding of the prosthetic heme to protein: a potential mechanism for the suicide inactivation or activation of hemoproteins. *Chem. Res. Toxicol.*, *2*, 131.

Schenkman, J.B., Sligar, S.G., Cinti, D.L. (1981). Substrate interactions with cytochrome P-450. *Pharmacol. Ther.*, *12*, 43.

Shou, M., Lin, Y., Lu, P., Tang, C., Mei, Q., Cui, D., Tang, W., Ngui, J.S., Lin, C.C., Singh, R., Wong, B.K., Yergey, J.A., Lin, J.H., Pearson, P.G., Baillie, T.A., Rodrigues, A.D., Rushmore, T.H. (2001). Enzyme kinetics of cytochrome P450-mediated reactions. *Current Drug Metab.*, *2*, 17.

Silverman, R.B. (1995). Mechanism-based enzyme inactivators. *Methods Enzymol.*, *249*, 240.

US Food and Drug Administration, Center for Drug Evaluation and Research (CDER) (1997). *Guidance for Industry: Drug Metabolism/Drug Interaction Studies in the Drug Development Process: Studies In Vitro*, www.fda.gov/cder/guidance/clin3.pdf (accessed April 1997).

Wang, E., Walsh, C. (1978). Suicide substrates for the alanine racemase of *Escherichia coli* B. *Biochemistry*, *17*, 1313.

White, R.E., Coon, M.J. (1980). Oxygen activation by cytochrome P450. *Annu. Rev. Biochem.*, *49*, 315.

# Evaluating and Predicting Human Cytochrome P450 Enzyme Induction

MICHAEL SINZ, SEAN KIM, STEPHEN FERGUSON, and EDWARD LECLUYSE

## 1 INTRODUCTION

Cytochrome P450 (CYP) isoforms are the predominate enzymes involved in the biotransformation and elimination of xenobiotics. Since their identification and characterization in 1962, they have become one of the most published enzyme systems in pharmaceutical research (Omura and Sato, 1962). Given their prevalence for metabolizing and eliminating drugs, interactions or events that manipulate the concentration or function of CYP enzymes have great impact on drug disposition, toxicity, and pharmacology. The most common alterations to drug-metabolizing enzyme function are due to (i) genetic polymorphisms, (ii) enzyme inhibition, and (iii) enzyme induction. The focus of this chapter will be on CYP enzyme induction, including methods to evaluate, predict, and quantitate drug interactions due to enzyme induction. Enzyme induction is the action of creating more enzymes than is normally present in a biological system and can be manifested by increased gene transcription or decreased protein or mRNA degradation. As early as 1954, the first report of enzyme induction appeared in a manuscript by Brown *et al.*, who described the enzyme-inducing effects of various food diets when given to rodents and the resulting increased enzymatic effects of *N*-demethylation in liver homogenates (Brown, Miller and Miller, 1954). This was followed some time later by the first review on microsomal enzyme induction in 1967 (Conney, 1967). Interestingly, some of the original nomenclature used to describe CYP enzymes was based on enzyme-inducing agents employed to increase CYP-isoform levels making them easier to characterize. For example, BNF-B was used to identify a CYP1A enzyme and PB-4/5 for a CYP2B enzyme, after treatment with the inducing agents β-naphthoflavone and phenobarbital (PB), respectively.

The increase in enzyme activity caused by a drug is reflected in an increased hepatic clearance of drugs metabolized by the induced enzyme. Various terms are used to describe the drug that causes enzyme induction (perpetrator or inducer) and the drug

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

affected by the induced enzyme (victim or substrate). Pharmacokinetically, the drugs affected by enzyme induction generally demonstrate reduced AUC (area under the curve), $C_{max}$ (maximum concentration), and half-life as a reflection of increased clearance. For example, coadministration of the antidiabetic drug rosiglitazone (a CYP2C8 substrate) with the inducing agent rifampicin leads to 65%, 31%, and 62% decreases in AUC, $C_{max}$, and half-life of rosiglitazone, respectively (Park *et al.*, 2004). A more dramatic example is illustrated by the interaction between midazolam (a CYP3A4 substrate) and St. John's wort (an herbal antidepressant and CYP3A4 inducer), where the AUC and $C_{max}$ decreased by 79% and 65%, respectively. However, the half-life of midazolam decreased only by 9%, which is anticipated for a high-clearance drug such as midazolam (Mueller *et al.*, 2006). These pharmacokinetic changes have significant pharmacological consequences because the increased metabolism (elimination) reduces the duration and pharmacological effect of a coadministered drug. For example, the induction of CYP3A4 results in reduced ethinylestradiol levels that can lead to unexpected pregnancies and reduced cyclosporine concentrations that can lead to organ transplant rejection (both drugs are predominately metabolized by CYP3A4). Drug–drug interactions are known to occur via enzyme induction for several CYP enzymes, most notably CYP3A4. In general, the decrease in AUC observed ranges from 15% to 98% depending on the potency of the inducing agent as well as the fraction metabolized and overall elimination of a drug (victim) via the induced pathway.

In addition to the aforementioned scenarios where the perpetrator drug causes the induction and a second victim drug is affected, there exists a situation where a single drug acts as both perpetrator and victim, referred to as "autoinduction." Carbamazepine is an anticonvulsant that causes CYP3A4 enzyme induction in epileptic patients and this induction increases the clearance of carbamazepine itself because carbamazepine is also a CYP3A4 substrate (Bertilsson *et al.*, 1980). The antimalarial drug artemisinin also exhibits this effect by inducing CYP2B6 which is significantly involved in artemisinin elimination (Simonsson *et al.*, 2003). Autoinduction is a phenomenon that can occur anytime a drug induces an enzyme that is also predominately involved in its own metabolic clearance. Therefore, when understanding the overall liability of an enzyme inducer, a complete understanding of the major mechanisms of elimination and the enzymes involved in elimination is important to assess potential autoinduction.

Drug–drug interactions are occurring with increasing frequency due to the rise in multiple prescriptions taken by individual patients. Recently it was estimated that 32 million Americans take three or more prescriptions daily (American Heart Association, www.americanheart.org). Hence, the potential for drug–drug interactions is also increasing and the need to develop compounds lacking the ability to induce drug-metabolizing enzymes has become a significant concern while developing new drugs. Most pharmaceutical companies have implemented several layers of screening and characterization to assess the potential liability of enzyme induction. In early drug discovery stages, large numbers of compounds are screened by means of high-throughput assays to either eliminate compounds with enzyme induction potential or reduce the liability by selecting compounds with lower potencies for the inductive effect. During preclinical development, characterization in more complex human systems helps to further define the induction potential of a compound. Ultimately, the effect or lack of effect of an enzyme inducer is evaluated in human volunteers/patients with well-characterized CYP probe substrates (as indicated in Table 1) where the actual magnitude of pharmacokinetic change is measured.

**TABLE 1 *In vivo* probe substrates and enzyme inducers used in clinical drug interaction studies (oral administration).[a]**

| CYP | *In vivo* probe substrates | Inducers |
|---|---|---|
| CYP1A | Theophylline, caffeine | Smokers versus nonsmokers |
| CYP2B6 | Efavirenz | Rifampicin |
| CYP2C8 | Rosiglitazone, repaglinide | Rifampicin |
| CYP2C9 | Warfarin, tolbutamide | Rifampicin |
| CYP2C19 | Omeprazole, esoprazole, lansoprazole, pantoprazole | Rifampicin |
| CYP2D6 | Desipramine, dexamethasone, atomoxetine | None identified |
| CYP2E1 | Chlorzoxazone | Ethanol |
| CYP3A4 | Midazolam | Rifampicin, carbamazepine |

[a]From FDA (2006).

## 2 RECEPTOR-BASED *IN VITRO* BIOASSAYS

For many drug-metabolizing enzymes, transcriptional activation and subsequent induction are mediated by nuclear hormone receptors that function as transcription factors, such as CAR (constitutive androstane receptor), PXR (pregnane X receptor or SXR-steroid X receptor), and PPAR (peroxisome proliferator-activated receptor) (Honkakoski, Sueyoshi and Negishi, 2003; Wang and LeCluyse, 2003; Qatanani *et al.*, 2005; Tirona and Kim, 2005). One notable exception is AhR (aromatic hydrocarbon receptor) which regulates the expression of CYP1A enzymes (Mandal, 2005). AhR is not a member of the nuclear receptor (NR) family, but belongs to the basic helix-loop-helix PER-ARNT-SIM (PAS) transcription factor family.

The identification of receptors that regulate CYP/transporter expression has led to the development of several high-throughput *in vitro* methods such as receptor binding and transactivation assays which assess the interaction between receptors and potential ligands (Vignati *et al.*, 2004; Zhu *et al.*, 2004). These assays are widely employed throughout drug discovery and development to identify potential inducers of CYP and transporter enzymes (Stanley *et al.*, 2006). In practice, these assays are utilized, in a parallel or sequential manner, along with lower-throughput but more definitive cell-based assays. As described in the preceding chapter, the human pregnane X receptor (hPXR, SXR, NR1I2) is a nuclear hormone receptor/transcription factor principally responsible for the induction of CYP3A4, as well as the expression of CYP2B6, CYP2C8/9, Pgp, MRP-2, and OATP-2 (Kliewer, Goodwin and Willson, 2002). PXR-derived assays are the most common high-throughput assays due to the simplicity of ligand-based activation and the importance of target genes in drug interactions. As such, the following discussion will mainly focus on PXR-based bioassays.

Ligand-binding assays generally consist of expressed receptors incubated with test compounds and a tight-binding radiolabeled ligand. Competition (displacement) of the radiolabeled ligand with the test compound is measured and an $IC_{50}$ (inhibitor concentration that causes 50% inhibition) is determined. Cell-based PXR transactivation-reporter assays employ (i) an expression vector containing some form of the CYP3A4 5′-promoter region coupled to a reporter gene (typically luciferase) and (ii) an hPXR expression vector. Due to increased expression of luciferase in response to ligand–receptor interaction, the induction response can be measured as an increased

production of luminescent product and an $EC_{50}$ (concentration at which 50% efficacy is found) value determined from the concentration–response curve. Typically these assays are augmented by co-transfecting an expression vector for human PXR to enhance the magnitude of induction response because most immortalized cell lines generally do not significantly express PXR (Goodwin, Hodgson and Liddle, 1999; El-Sankary *et al.*, 2001). It has been shown that the choice of host cell has an impact on the magnitude of induction response. For example, rifampicin treatment of HepG2 cells induces luciferase activity, while the same treatment in COS7 or NIH3T3 cell lines results in minimal responses (Goodwin, Hodgson and Liddle, 1999). In addition, it does not appear that a maximal PXR response requires a cell of hepatic origin. For example, the African Green Monkey kidney cell line, CV-1, is widely used in PXR reporter assays. However, because the ligand–PXR interaction is a complex process requiring multiple co-regulators, the most appropriate host cell lines presumably have all the regulatory partners necessary to mediate an induction response.

In general, there is a good correlation between $EC_{50}$ and $IC_{50}$ values in PXR transactivation and binding assays, respectively (Zhu *et al.*, 2004). However, there are examples where significant binding to the PXR receptor does not lead to transactivation or enzyme induction. These unique situations can be caused by PXR antagonists or ligands which bind, but do not elicit the appropriate displacement of corepressors or recruitment of co-activators. Such is the example of docetaxel and paclitaxel, two structurally similar anticancer agents, in which both drugs bind to PXR; however, only paclitaxel binding results in significant gene activation and enzyme induction (Harmsen *et al.*, 2007). Ectesinascidin 743 (ET-743) is an experimental anticancer drug that can also inhibit PXR transactivation at nanomolar concentrations via PXR antagonism (Synold, Dussault and Forman, 2001). With the increasing number of compounds being tested simultaneously in PXR receptor binding and reporter assays, more PXR antagonists (characterized by a strong PXR binding but no or weak PXR transactivation) are expected to be identified. However, it is not yet clear what the clinical impact that PXR antagonists will have on the expression of PXR target genes in humans.

The change in hPXR transactivation-reporter activities has also been shown to correlate with the changes in CYP enzyme activities measured from primary human hepatocytes, which are widely accepted as the standard *in vitro* assessment of enzyme induction (Luo *et al.*, 2002). Presently, the PXR transactivation assay is the preferred method to screen new chemical entities over the PXR binding assay due to the improved correlation to human drug–drug interactions, fewer false positives, and no radiolabeled reagents. While many reporter assays utilize transiently transfected systems, the variation in transfection efficiency and the resulting induction responses from one transfection to another is not ideal in an industrial paradigm of high-throughput screening. By introducing an antibiotic resistance gene as a selection marker, one can establish a stably transfected cell such as the DPX-2 cell line to minimize run-to-run variations (Lemaire, de Sousa and Rahmani, 2004; Trubetskoy *et al.*, 2005). Another approach is to cryopreserve a single large batch of transfected cells and thus ensure a more consistent induction response from one experiment to another. The bulk cryopreservation method does not appear to affect the functionality of the cells as the induction responses ($EC_{50}$ and maximum response) observed between cryopreserved and fresh cells were found to be similar (Zhu *et al.*, 2007).

While PXR-mediated gene transcription is primarily initiated by direct ligand binding, several alternative mechanisms of gene induction have been identified. CAR

is constitutively activated and predominately sequestered in the cytoplasm by accessory proteins, such as HSP90 and CAR cytoplasmic retention protein (CCRP), in order to regulate transcriptional activity. Translocation of CAR to the nucleus can be initiated by direct agonist binding to the receptor but also through a partially elucidated ligand-independent mechanism involving kinases. Phenobarbital is an example of a drug that does not bind to CAR yet causes nuclear translocation and transcriptional activation of a target gene, CYP2B6 (Qatanani *et al.*, 2005). Similarly, the inactive AhR resides in the cytoplasm and can be activated upon ligand binding or via protein tyrosine kinases. Omeprazole-mediated induction of CYP1A has been shown to proceed through this kinase-mediated pathway although this premise has come under some debate (Hu *et al.*, 2007). Induction of CYP2E1 is notably different from other CYP enzymes in that no receptor is known to be involved. CYP2E1 inducers, such as ethanol, increase the stability of mRNA and/or protein and thus result in increased enzyme activity due to diminished degradation pathways (Gonzalez, 2007).

The development of CAR-based assays has proven to be more difficult than PXR assays. CAR transactivation assays frequently encounter a high basal CAR activity and CAR often spontaneously translocates to the nucleus when overexpressed in cell lines (Chang and Waxman, 2006). These experimental challenges hamper the identification of CAR activators and only a few compounds have been identified to date. Also due to dual mechanisms of CAR activation, direct and indirect, screening for CAR modulators (potential CYP2B6 inducers) involves a nuclear translocation assay to complement a transactivation assay. The CAR nuclear translocation assays are frequently conducted in primary hepatocytes due to the spontaneous translocation of CAR in cell lines as described earlier. The translocation of hCAR from cytoplasm to nucleus can be monitored and measured either by confocal microscopy using an expressed CAR protein tagged with fluorescent proteins such as GFP or EYFP or by measuring the ratio of CAR proteins between cytoplasm and nucleus using Western blot analysis (Faucette *et al.*, 2007). Recently, a human splice variant of CAR was identified and termed "CAR3." A subsequent analysis showed that CAR3 has ~80% lower basal activity compared with wild-type CAR (CAR1) when expressed in COS-1 cells (Auerbach *et al.*, 2005; Faucette *et al.*, 2007). The CAR3-based transactivation assay can more easily identify potential CAR activators by lowering the basal activity that can otherwise mask a potential ligand-activated transactivation response.

It should be noted that the ligand binding domains (LBDs) of many nuclear receptors are different between various animal species and human (Wang and LeCluyse, 2003). For example, rodent and human PXR only share ~76% LBD sequence homology. Therefore, *in vitro* bioassays utilizing animal PXR are generally not predictive of the potential induction effect in humans. Nonetheless, enzyme induction in animals can complicate preclinical safety assessments in which these animal species are utilized (Worboys and Carlile, 2001). Frequently, enzyme induction (especially autoinduction) is found in multiple-dose animal pharmacokinetic or toxicokinetic studies, and the use of animal *in vitro* assays can aid in the investigation of induction mechanisms as well as species differences.

While nuclear hormone receptor-based assays provide reliable high-throughput screening opportunities, one disadvantage is that each assay only predicts one particular receptor-mediated process at a time. In some cases, multiple nuclear hormone receptors can work in concert and provide a cumulative effect on target genes (Wang and LeCluyse, 2003). This nuclear receptor cross talk with target genes is best evaluated in more complex cell-based models to be described in subsequent sections.

## 3 IMMORTALIZED HEPATOCYTES

Primary human hepatocytes are commonly used in drug metabolism-related studies as well as in the assessment of hepatotoxicants (Soars *et al.*, 2007). However, their limited supply, rapid decline in expression of key drug-metabolizing enzymes, and significant donor-to-donor variation have fostered research toward the generation of human hepatocyte-like cells that can provide a continuous supply of cells while maintaining stable expression of necessary enzymes and transporters. There are currently three different approaches being taken to generate cell lines with these characteristics: refinement of hepatocarcinoma-derived cells into an appropriate phenotype, immortalization of human primary hepatocytes, and differentiation of stem cells into hepatocyte-like cells (to be discussed in Section 14).

Immortalized cells are defined as cells that can grow and divide indefinitely under optimal culture conditions (Hahn, 2002). Immortalization of cells can occur naturally as exemplified by cells of tumor origin; however, a considerable effort is being made to convert primary cells into non-tumorigenic immortalized cells (Sinz and Kim, 2006). There are many different approaches used to immortalize primary hepatocytes but the most common methods of immortalization are (i) overexpression of SV40 large T antigen (viral oncogene) or (ii) expression of telomerase reverse transcriptase (TERT). While viral transformation is relatively simpler, the cells produced through this method tend to be genetically unstable and readily lose their phenotypic characteristics. Combined, both SV40 T antigen and TERT-mediated transformations are predicted to produce genetically more stable cells (Cascio, 2001). In all types of cells, the goal is to either find or produce a cell line that maintains the phenotypic characteristics of the parental tissue (hepatocytes) and thus provide an unlimited and stable supply of cells for ADMET (absorption, distribution, metabolism, elimination, toxicity) research. The goal of this section is to review the most widely studied immortalized cells used in the assessment of CYP induction as well as the methodologies that are employed.

## 4 HepG2 AND HepG2-DERIVED CELL LINES

In the absence of a suitable alternative to primary hepatocytes, several lines of hepatocarcinoma cells have been used in the study of drug metabolism. Among them, the HepG2 cell line (established in 1979 from a hepatoma tissue) is the most well characterized and frequently used. While retaining some liver-specific functions, HepG2 cells have low expression and activity of CYP enzymes, as well as phase II enzymes (Vermeir *et al.*, 2005). Mostly fetal isoforms of CYPs, such as CYP1A1 and CYP3A7, are expressed in high levels and the expression of other phase I and II enzymes are much lower than those in primary hepatocytes (Wilkening, Stahl and Bader, 2003). The same observation was made in recent comprehensive reports on the expression analysis of phase I and II enzymes in HepG2 cells (Westerink and Schoonen, 2007a, b). Compared with the expression levels in cryopreserved human hepatocytes, HepG2 cells express much lower levels of all major CYPs (CYP1A1, -1A2, -2A6, -2B6, -2C8, 2C9, -2C19, -2D6, -2E1, and -3A4) as well as phase II enzymes (UGT1A1 and -1A6). However, the transcript levels of several isozymes of sulfotransferases (SULTs), glutathione-*S*-transferases (GSTs), *N*-acetyltransferases (NATs), as well as epoxide hydrolase (EPHX1) were similar to those in primary human hepatocytes in these studies.

In these studies, the transcript levels of several CYPs including CYP3A4 and -2B6 as well as phase II enzymes such as EPHX1, SULTs, and GTSs were increased in response to model inducers. The most significant induction was observed with CYP1A1/2 RNA expression and activity by known AhR agonists (BNF, 3MC, TCDD, indirubin, and indigo). This observation is consistent with previous reports of CYP1A1/2 induction by other xenobiotics including but not limited to polyaromatic hydrocarbons, omeprazole, and quinine (Bapiro *et al.*, 2002; Iwanari *et al.*, 2002; Baliharova *et al.*, 2003). The induction of CYP3A4 in HepG2 cells was also reported with *o, p′*-DDT (main isomer of DDT) which induced both activity and mRNA levels of CYP3A4 via an activation of PXR (Medina-Diaz and Elizondo, 2005; Medina-Diaz *et al.*, 2007). Others reported the induction of CYP3A4 by PB and dexamethasone (Hewitt and Hewitt, 2004) as well as induction of UGT1A1 by the dietary flavonoid, chrysin (Walle *et al.*, 2000; Smith *et al.*, 2005). However, due to the low basal expression of enzymes and minimal induction responses, HepG2 cells are not considered an appropriate *in vitro* model to study induction of CYPs.

Recently, HepG2 cells have been recognized for their heterogeneity in genetic characteristics in that several different cell subtypes can coexist in a cell population. Hewitt *et al.* reported that even the source of the cells can have an impact on basal activities in HepG2 cells (Hewitt and Hewitt, 2004). By enriching only the type of cells that have better induction responses through means of altered nutritional requirements, several HepG2 subtypes have been established. Amphioxus ACTIVTox cells are a highly selected subclone of HepG2 cells developed by Amphioxus Cell Technologies (Stem Cell Innovations, Houston, TX). The ACTIVTox cells have been shown to be useful in assessing the induction of both CYP1A and CYP3A4 as well as toxicity of new chemical entities (www.stemcellinnovations.com). The WGA cell line is also a highly differentiated subclone of HepG2 cells enriched by culturing the cells in a defined medium that lacks arginine (Rencurel *et al.*, 2005). Unlike the parental HepG2 cells or other immortalized cells, the WGA cells were shown to be responsive to PB-induced CYP2B6 induction, which holds promise as a cell line to study CAR-mediated CYP regulatory mechanisms.

## 5 HBG-BC2

The human hepatoma cell line, HBG, was isolated from the resected tumor of a male patient (Le Jossic *et al.*, 1996). When seeded at confluency, these cells undergo a differentiation process, resulting in expression of liver-specific functions, as well as drug-metabolizing enzymes. These differentiated cells are known as BC2 cells (O'Connor *et al.*, 2005). Gomez-Lechon *et al.* performed an extensive analysis of the biotransformation properties of BC2 cells by measuring basal and inducible expression of phase I and II enzymes (Gomez-Lechon *et al.*, 2001). The analysis revealed low but measurable activities of CYP1A1/2, -2A6, -2B6, -2C9, and -3A4. Upon treatment of the cells with appropriate inducers, the activity of CYP1A1/2 was greatly increased while no or marginal induction was observed for CYP2B6. Even though the authors observed induction of CYP3A4 in response to 100-μM dexamethasone, no induction of CYP3A4 activity (testosterone 6β-hydroxylase) was found with either rifampicin or phenobarbital suggesting that BC2 cells may not be a good model to assess induction potential of suspected CYP3A inducers. Thus, recent studies using BC2 cells have focused more on their use in generating metabolites or assessing toxicities mediated by biotransformation (Fabre *et al.*, 2003; O'Connor *et al.*, 2005).

## 6 HEPATOCARCINOMA-DERIVED CELL LINE: HepaRG

The HepaRG cell line was isolated from a resected liver tumor of a female patient suffering from hepatitis C-related hepatocellular carcinoma (HCC) (Gripon *et al.*, 2002). When the cells are cultured at confluency for several weeks in the presence of 2% DMSO and hydrocortisone, a highly differentiated hepatocyte-like cell line develops. Subsequent examination revealed that these cells express several liver-specific markers such as albumin, transferrin, CYP2E1, CYP3A4, aldoase B and GSTα, and retain a susceptibility to hepatitis B virus infection which was previously validated only in primary hepatocytes. A comprehensive expression analysis performed by Aninat *et al.* (2006) showed that RNA expressions of major nuclear hormone receptors (PXR, AhR, and PPARα) were similar between primary human hepatocyte and differentiated HepaRG cells (aided by the culture of cells in DMSO). Interestingly, the expression of CAR was high compared with that of HepG2 cells when HepaRG cells were cultured in the presence of DMSO, and was 20–30% of that in primary human hepatocytes. Previously, none of the immortalized cell lines have reported substantial expression of CAR and thus the HepaRG cell line may be of interest in studying CAR-mediated induction. In this study, some of CYP enzymes also required the addition of DMSO (2%) to exhibit similar expression levels as in primary hepatocytes. For example, mRNA levels of CYP2B6, -2E1, and -3A4 were markedly increased in the presence of DMSO while CYP1A2, -2C9, -2D6, and other liver-specific markers were relatively unchanged. The enzyme activities of CYP1A, -2C9, -2E1, and -3A4 were also highly enhanced by the addition of DMSO. However, in the presence of DMSO, HepaRG cells were refractory to induction by several prototypical inducers and the investigators postulated that the DMSO treatment induced the expression of CYPs to the maximum level that can be achieved. This hypothesis was supported by the observation that prototypical inducers increased enzyme activities of CYP3A4, -2C9, and -1A by approximately 10-, 2-, and 63-fold, respectively, in HepaRG cells grown without DMSO. Le Vee *et al.* also examined the expression of hepatic transporters in HepaRG cells (Le Vee *et al.*, 2006). The transcript levels of OCT-1, OATP-C, NTCP, OATP-8, BSEP, and MRP-2 were approximately 10–55% of those found in primary human hepatocytes, while the expressions of OATP-B, MDR-1, and MRP-3 were higher in HepaRG than in primary human hepatocytes. The measurement of uptake transporter activities showed that OCT-1, OATP/OAT2, and NTCP activities were detectable but lower than in primary human hepatocytes. However, the activities of efflux transporters such as MDR-1 (Pgp) and MRP were similar or even higher than those in primary human hepatocytes. The treatment of HepaRG cells with known inducers of transporters such as rifampicin, phenobarbital, and chenodeoxycholic acid significantly increased the mRNA expression of MDR-1, MRP2, and BSEP, respectively. Overall, the HepaRG cell line could represent a reliable surrogate *in vitro* model to primary human hepatocytes as an unlimited source of hepatocyte-like cells while retaining high expression as well as inducibility of drug-metabolizing enzymes and transporters (Guillouzo *et al.*, 2007).

## 7 CELLS OF INTESTINAL ORIGIN: LS180 AND Caco-2

The PXR is highly expressed in the liver, as well as in the intestine, and is known to regulate the expression of intestinal CYP3A4 and MDR-1. The induction of CYP3A4 and MDR-1 in the intestine can negatively impact the absorption of orally delivered

drugs and thus a suitable cell line of intestinal origin has been sought after to study the regulation of these genes. LS180 and Caco-2 cells are human colon carcinoma cell lines routinely used and well characterized in studying intestinal absorption of drugs and regulation of CYP3A4 and MDR-1. In LS180 cells, CYP3A4 and MDR-1 have been shown to be responsive to induction stimuli by reserpine, rifampicin, PB, and verapamil (Schuetz, Beck and Schuetz, 1996). The rank order for inducers of MDR-1 was similar to that of CYP3A4 inducers in this study, suggesting a parallel regulation of these genes in LS180 cells. While inducers of MDR-1 and CYP3A4 demonstrate similar changes in LS180 cells, similar responses were not observed for Caco-2 cells (Li *et al.*, 2003). MDR-1 induction was often prominent even in the absence of CYP3A4 induction which could be due to a lower expression of PXR in Caco-2 cells (Li *et al.*, 2003; Brandin *et al.*, 2007).

In addition, induction of CYP1A2 was also reported when LS180 cells were exposed to a prototypical CYP1A inducer, omeprazole (Brandin *et al.*, 2007). A variety of herbal and food supplements were shown to induce the expression of CYP1A2, -3A4, and MDR-1 in this study suggesting herbal/food–drug interactions can be studied in the LS180 cell line. Interestingly, vitamin supplements appear to play an important role in the regulation of drug-metabolizing enzymes and transporter in these intestinal cells. Tocotrienols, major components of vitamin E, induced expression of MDR-1 and UGT1A1 but not CYP3A4 in LS180 cells (Zhou *et al.*, 2004). This was in contrast to primary hepatocytes where only the induction of CYP3A4 by tocotrienols was observed. Subsequently it was shown that the expression of nuclear receptor corepressor (NCoR) which negatively regulates PXR activity was higher in LS180 cells than in primary human hepatocytes. By attenuating the expression of NCoR, the ability of tocotrienols to induce CYP3A4 in LS180 cells was restored. The addition of supra-physiological concentrations of 1α, 25-dihydroxyvitamin D3 (1α, 25-$(OH)_2$-$D_3$) increased the otherwise low expression of CYP3A4 in Caco-2 and TC7 cells (Schmiedlin-Ren *et al.*, 1997; Engman *et al.*, 2001). Further induction of CYP3A4 by known PXR ligands was not demonstrated in 1α, 25-$(OH)_2$-$D_3$-treated Caco-2 cells, suggesting that PXR is not involved in vitamin $D_3$-mediated CYP3A4 induction (Schmiedlin-Ren *et al.*, 2001). Overall, LS180 cells appear to be a suitable model to study the regulation of CYP3A4 and MDR-1 while Caco-2 cells are better suited for drug absorption studies (Aiba *et al.*, 2005).

## 8 IMMORTALIZED HUMAN HEPATOCYTE CELL LINES: Fa2N-4

Developed by researchers at Multicell Technologies, the Fa2N-4 cell line is widely used for the assessment of CYP induction. This cell line is derived from primary hepatocytes (12-year-old female donor) through a clonal selection process following immortalization using the SV40 large T antigen. While most immortalized human hepatocyte cell lines have lost basal expression as well as inducibility of CYPs, Fa2N-4 cells retain normal hepatocellular morphology as well as expression and inducibility of CYPs and transporters. Using known inducers of CYP enzymes, Mills *et al.* reported concentration-dependent increases in both transcript levels and enzyme activities of CYP3A4, -2C9, and -1A2 (Mills *et al.*, 2004). These changes were comparable to the induction responses observed in primary human hepatocytes. In addition, UGT1A and MDR-1 were also induced by rifampicin treatment in this cell line. However, it is generally believed that Fa2N-4 cells have no or very low expression of CAR and thus induction of CYP2B6 cannot be evaluated in this cell line (Hariparsad *et al.*, 2008).

The procedure for assessing enzyme induction in Fa2N-4 cells is well defined. While it is similar to the protocol using primary hepatocytes, there are some subtle differences. Typically, Fa2N-4 cells are propagated on a collagen substratum in a proprietary medium. Following trypsinization, cells are plated in a wide variety of formats (6-, 12-, 24-, and 96-well). After one to two days of adaptation, the cells are treated once daily for two to three consecutive days with test articles and control compounds (negative/positive). Whereas the cells are immortalized, they proliferate throughout this three- to five-day process. Hence, some consistency must be maintained to control the cell number/density from experiment to experiment to decrease variability of the induction response. The induction experiment can start either by subculturing actively proliferating cells or by thawing and plating cryopreserved Fa2N-4 cells. The latter is more widely used as it eliminates the need for maintaining cells in culture and also possible negative effect of multiple passages on the overall induction response.

Fa2N4 have several attributes that make them amendable to the high-throughput environment of drug discovery. Fa2N-4 cells appear to maintain their induction response even in higher-density well formats, such as the 96-well plate format. Ripp *et al.* have developed a high-throughput induction assay using 96-well plates and produced detailed concentration–response curves for known inducers of CYP3A4 (Ripp *et al.*, 2006). In addition, Youdim *et al.* reported the use of 96-well plate incubations of Fa2N-4 cells with a cocktail of probe substrates (Youdim *et al.*, 2007).

## 9 OTHER IMMORTALIZED HEPATOCYTE CELL LINES

The expansion of immortalized hepatocyte cell lines as well as development of more sophisticated immortalization techniques is expected to produce proliferating yet functional hepatocytes that can be used in many areas of ADMET research. While the Fa2N-4 cell line is commonly used, there are several other immortalized hepatocyte cell lines currently under evaluation. NKNT-3 is a reversibly immortalized human hepatocyte cell line that was generated by retroviral transfer of the SV40 large T antigen that can be subsequently excised by Cre/Lox site-specific recombination (Kobayashi *et al.*, 2000). NKNT-3 cells express CYP2C9 and CYP3A4, which can be further enhanced by coculturing NKNT-3 cells with another immortalized human hepatocyte stellate cell line, TWNT-1 (Watanabe *et al.*, 2003). YOCK-13 is also a reversibly immortalized, insulin-secreting human hepatocyte cell line generated using human TERT as the immortalizing gene. YOCK-13 cells express markers of hepatocyte differentiation that include albumin, asialoglycoprotein receptor, bilirubin-uridine diphosphate glucronosyltransferase, CYP-associated enzyme 3A4, glutamine synthetase, and GST-π (Okitsu *et al.*, 2004). OSUM-29 is an immortalized human fetal hepatocyte cell line which has demonstrated expression of CYP1A1, CYP1A2, and CYP3A4 when cultured in a 3-D radial-flow bioreactor (Akiyama *et al.*, 2004). Finally, the HepZ cell line was immortalized using antisense RNA against Rb and P53 genes, which control cell-cycle regulation and have been shown to exhibit PB-inducible CYP activities (Werner *et al.*, 1999).

## 10 PRIMARY HUMAN HEPATOCYTE CULTURE SYSTEMS

The use of cultured primary human hepatocytes has become the "gold standard" *in vitro* system for assessing the potential for chemicals to induce human CYP expression (LeCluyse *et al.*, 2005). Enzyme induction data from *in vitro* experiments with

primary human hepatocytes are known to correlate well with clinical observations when pharmacologically relevant concentrations of test article are used (LeCluyse *et al.*, 2000). When hepatocytes are cultured using appropriate conditions that facilitate liver-like cell morphology and expression of liver-specific proteins (e.g. albumin), CYP enzymes are effectively induced *in vitro* analogous to the *in vivo* situation in terms of the magnitude and specificity of CYP induction (LeCluyse *et al.*, 2000; Runge *et al.*, 2000). Several reviews on the application of primary human hepatocytes in assessing and predicting human drug interactions due to enzyme induction have been recently published (Dickins, 2004; Hewitt, de Kanter and LeCluyse, 2007; Soars *et al.*, 2007).

The process of inducing CYP expression is a complex series of responses (e.g. receptor binding, receptor translocation, receptor activation, recruitment of co-activators and corepressors, receptor dimerization, receptor binding to target DNA, chromatin reorganization, stabilization of mRNA, and stabilization of protein). As a result, a metabolically competent system capable of metabolizing drugs in a liver-like manner that contains human gene targets, human receptors, human co-activators, human protein stabilization machinery, and human enzymes with analogous properties to human liver (e.g. mechanism-based inhibition) is optimal to most effectively model the human inducibility by xenobiotics and their hepatic metabolites. Primary human hepatocyte culture systems have been shown to effectively model these responses and are recognized by the FDA as an effective tool for assessing xenobiotic induction potential (FDA, 2006).

Species differences in induction response can occur at multiple levels of the process. Most importantly, many of the species differences in induction response arise due to unique interactions between xenobiotics and nuclear receptors (transcription factors), making prediction of human induction responses difficult (Wang and LeCluyse, 2003). For example, rifampicin is a potent and effective inducer of human CYP3A4 but not rat CYP3A (Silva, Day and Nicoll-Griffith, 1999; LeCluyse *et al.*, 2000; Zhu *et al.*, 2000). This is effectively modeled in cultures of rat and human hepatocytes and when performing *in vivo* induction studies in rats and humans. Similar species differences are also known for omeprazole where the drug has been shown to be a more effective inducer of human CYP1A2 than of rat CYP1A (Shih *et al.*, 1999; LeCluyse *et al.*, 2000). Also, the $EC_{50}$ of omeprazole induction of human CYP1A2 is much larger than the exposure in patients treated with omeprazole for gastric ulcers; thus, it rarely causes significant clinical drug interactions due to induction because the concentration of omeprazole required to induce CYP1A2 *in vitro* generally exceeds the concentrations achieved *in vivo* (Daujat *et al.*, 1992; Dilger, Zheng and Klotz, 1999; LeCluyse *et al.*, 2000). Omeprazole is a drug that illustrates the need to assess enzyme induction in an appropriate species and at a therapeutic drug concentration in order to obtain the correct induction response/prediction. Phenobarbital and other anticonvulsants (e.g. phenytoin) are known to induce human CYP2B, CYP2C, and CYP3A enzymes. In human patients, phenytoin has been reported to induce the metabolism of multiple drugs including cyclophosphamide, thiotepa, and tacrolimus (de Jonge *et al.*, 2005; Wada *et al.*, 2007). However, clear species differences exist for the induction of CYPs by anticonvulsants. For example, induction of the murine CYP2Cs by phenobarbital and phenytoin has been reported to be mediated through CAR, rather than PXR, as the major determinant for murine Cyp2c29 and Cyp2c37 induction (Jackson *et al.*, 2004, 2006). This is a clear distinction from the human CYP2Cs which appear to be regulated by both PXR and CAR (Chen *et al.*, 2003, 2004; Ferguson *et al.*, 2005).

Comparison of immortalized liver cell systems with primary human hepatocyte cultures reveals differences in their respective abilities to proliferate. This distinction may be important because the cell-signaling pathways that cause a liver cell to become immortalized may influence the pathways involved in the regulation of CYP expression. As previously discussed, the expression level of CYPs and transcription factors are much lower in immortalized cell systems compared with primary hepatocytes (Hewitt, de Kanter and LeCluyse, 2007). Therefore, to comprehensively assess the induction (and/or suppression) of multiple pharmacologically important genes, an immortalized cell system may be limited.

The CAR is thought to be essential for phenobarbital-type induction and appears to play a role in the regulation of CYP2B6, CYP2C9, CYP2C8, and UGT1A1. Therefore, assessing induction potential of new chemical entities in a CAR responsive model system is important to evaluate the full potential of new chemical entities to induce CYP expression. Primary human hepatocyte systems and immortalized cell systems have both been shown to effectively model AhR and PXR receptor pathways as discussed earlier. However, CAR's unique constitutive activity coupled with its spontaneous translocation to nuclei in immortalized cell systems makes it difficult to model in immortalized cell lines (Kawamoto *et al.*, 1999). Primary human hepatocyte culture systems effectively model CAR activation via both translocation and ligand activation, and presently are the most effective model system to assess human CAR activation (Wang *et al.*, 2004; Faucette *et al.*, 2007).

Immortalized cell systems are often deficient in terms of baseline enzymatic activity but can model induced expression and enzymatic activity of certain pathways and enzymatic target genes (see Section 3). However, due to the poor metabolism of compounds in these systems, it is difficult to evaluate the effects of metabolites on these induction pathways. Primary human hepatocyte systems are known to be metabolically active and are useful in studying the metabolic stability of chemicals (Jacobson *et al.*, 2007). This, coupled with their demonstrated ability to effectively model multiple induction pathways and other pharmacologically important events such as mechanism-based inhibition, highlights the utility of this model system.

A limitation of primary hepatocyte culture systems has been tissue availability. While fresh human hepatocytes are the standard for evaluating *in vitro* induction of CYP enzymes, attachable cryopreserved hepatocytes can also be used and data using cryopreserved cells are accepted by the FDA. The drug-metabolizing enzymes remain inducible after cryopreservation, and due to the significant variation in activities of drug-metabolizing enzymes between individual human livers, cryopreserved cells generate results essentially indistinguishable from freshly isolated cells (Schehrer, Regan and Westendorf, 2000; Hewitt *et al.*, 2007). mRNA, protein expression, and activities of CYP1A2, 2B6, 2C9, 2E1, and 3A4 in cryopreserved hepatocytes have been shown to be inducible by many common inducers (Roymans *et al.*, 2005), as well as the activity of multiple uridine diphosphate-glucuronosyl-transferases, carboxylesterases, and sulfotransferases (Reinach *et al.*, 1999; Nishimura *et al.*, 2004). The advantage of cryopreserved cells is that experiments can be routinely conducted, as the experiments are not dependent on the sporadic availability of fresh hepatocytes.

Most research currently uses a "sandwich" culture model, where isolated hepatocytes are plated on simple collagen (or other extracellular matrix) and sandwiched between the collagen basement membrane and an extracellular matrix overlay such as Matrigel (LeCluyse *et al.*, 2000, 2005; Hamilton *et al.*, 2001). This environment appears optimal for maintenance of hepatocyte cultures in liver-like cell morphology and over

multiple days in culture often forms liver-like bile canaliculi (Bi, Kazolias and Duignan, 2006; Marion, Leslie and Brouwer, 2007). This culture system affords the opportunity to maintain mature hepatocytes in a liver-like environment capable of assessing the cumulative effects of metabolism, induction, inhibition (time-dependent), and metabolite effects over multiple days, thus replicating the more steady-state-like effects on CYP activity and gene expression.

Data should be collected from at least three individual donors, treated with the test compound, vehicle, and positive controls (known inducers, Table 2) for two to three days. Drug treatment begins 48 hours after cell plating once the cells have had an opportunity to reform cell–cell contacts. Typically, two to three days of drug treatment are necessary for RNA expression and three days for enzyme activity assessments. Each hepatocyte donor preparation is deemed acceptable if the recommended positive control activity demonstrates a greater than twofold increase in enzyme activity of the probe substrate. A minimum of three test compound concentrations, based on the expected human plasma drug levels are suggested, one of which should be an order of magnitude greater than this concentration. In the absence of knowledge of human plasma levels, concentrations ranging over at least two orders of magnitude should be studied.

**TABLE 2 Common CYP inducers and CYP probe substrates used in *in vitro* studies.**

| CYP Enzyme | Inducer | Substrate/probe |
|---|---|---|
| 1A2 | **Omeprazole**, **β-naphthoflavone, 3-methylcholantrhrene**, lansoprazole | **Phenacetin**, theophylline, caffeine, acetaminophen, tacrine |
| 2A6 | **Dexamethasone**, pyrazole | **Coumarin, nicotine**, butadiene |
| 2B6 | **Phenobarbital**, phenytoin | **Efavirenz, buproprion**, propofol, *S*-mephenytoin |
| 2C8 | **Rifampicin**, phenobarbital | **Taxol**, repaglinide, rosiglitazone, amodiaquine |
| 2C9 | **Rifampicin**, phenobarbital | ***S*-warfarin**, **tolbutamide, diclofenac**, phenytoin, flurbiprofen |
| 2C19 | **Rifampicin** | ***S*-mephenytoin**, fluoxetine, omeprazole, esoprazole, lansoprazole, pantoprazole, citalopram, diazepam, hexobarbital, imipramine, proguanil, propanolol |
| 2D6 | None identified | **Dextromethorphan, bufuralol**, codeine, ethylmorphine, desipramine, atomoxetine, nicotine, debrisoquine |
| 2E1 | Ethanol, isoniazid, acetone | **Chlorzoxazone**, aniline, *p*-nitrophenol, lauric acid, acetaminophen, caffeine, dapsone, enflurane, theophylline |
| 3A4/3A5 | **Rifampicin**, rifabutin, Rifapentine, sulfinpyrazone, carbamazepine, dexamethasone, phenytoin, troleandomycin, troglitazone, taxol | **Midazolam, testosterone**, dextromethorphan, triazolam, terfenadine, buspirone, felodipine, lovastatin, eletriptan, sildenafil, simvastatin, triazolam, acetaminophen, carbamezapine, cyclosporin, digitoxin, diazepam, erythromycin, fluoxetine, nifedipine, quinidine, saquinavir, cortisol, terfenadine, verapamil |

Drugs highlighted in bold are preferred inducers and substrates.
Compiled from Tucker, Houston and Huang (2001) and FDA (2006).

The draft FDA guidance on drug interaction studies recommends analysis of catalytic activity of the major drug-metabolizing enzymes CYP1A2 and CYP3A4 in freshly isolated or attachable cryopreserved hepatocyte cultures. Since co-induction of CYP2C, CYP2B, and P-gp occurs with CYP3A, negative results *in vitro* with CYP3A may also eliminate the need to address interactions eliminated by these CYPs and P-gp. However, when utilizing data on only CYP1A1/2 and CYP3A4, it may be possible to underpredict clinical induction. This can be due to cross talk between nuclear receptors, for example, PXR and CAR (Faucette *et al.*, 2006). When evaluating only CYP1A and CYP3A and not CYP2B6, induction via CAR could be misrepresented. Enzymatic analysis can be conducted in microsomes prepared from hepatocytes or *in situ*, using media-based assays and CYP-specific probe substrates, where the metabolite formed can be attributed to one major CYP enzyme. The enzyme activity measurement is thought to be more reflective of actual enzyme changes (pharmacokinetic changes) that occur in patients. Additional methods to quantitate *in vitro* enzyme induction are CYP protein concentration (by Western immunoblotting) and mRNA expression (by reverse transcriptase–polymerase chain reaction, RT-PCR). RNA expression is particularly useful when both enzyme induction and inhibition are occurring with the same test article, such as with troleandomycin and DPC681 (Luo *et al.*, 2003; McGinnity *et al.*, 2006). An advantage of primary human hepatocyte culture systems is their ability to model metabolism, inhibition, and induction (Luo *et al.*, 2002). We have examined the effects of rifampicin and phenobarbital (prototypical PXR and CAR activators), ketoconazole (a potent competitive CYP3A4 inhibitor), and several metabolism-dependent inhibitors including erythromycin, troleandomycin, and ritonavir on CYP2B6 and CYP3A4 mRNA, protein content, and enzymatic activity over multiple time points. Generally, clear distinctions were observed between early time points (where inhibition would be predicted to be prominent) and longer time points where contributions from both inhibition and induction would be operative. These studies highlight the power of simultaneously probing induction and inhibition in a single, metabolically competent cell system. Whereas measurement of enzyme activities from incubations is considered the most reliable measure of induction potential, most often, both RNA expression and enzyme activity are measured to better understand the overall effects of a test article.

A relatively new area of research has recently emerged in the area of down-regulation of CYP expression (also called suppression). Suppression is generally thought to be the opposite of induction; however, the mechanisms controlling induction of CYPs may not be the most prominent factors for basal CYP expression in various tissues. For example, in human liver, factors such as GR and HNF4α have been shown to be important for basal expression, but appear to have supportive roles for classical CYP induction (Pascussi *et al.*, 2001). Primary human hepatocyte culture systems have also been shown to model CYP suppression. The most intuitive example is that of interleukins on CYP gene expression (Abdel-Razzak *et al.*, 1995; Chen *et al.*, 1995a, b). It has been known historically that inflammation results in decreased metabolic capacity in humans, and these studies identified specific inflammatory factors. Additional research is needed to probe the mechanisms by which this occurs at the transcriptional level. Another recent example of CYP suppression was with colchicine where Dvorak and coworkers demonstrated suppression of several CYPs including CYP2C9 and CYP3A4 (Dvorak *et al.*, 2003). The mechanism of this suppression appears to be down-regulation of the glucocorticoid receptor which is important for the basal expression of CYPs on multiple levels including down-regulation of nuclear receptor expression

and the possible impact on chromatin histone deacetylation (Dvorak *et al.*, 2003). These examples highlight the need to understand chemical interactions that may affect the basal, as well as the induced levels of CYPs, and primary human hepatocyte culture systems have the advantage of modeling these processes. Additional modifications to culture systems such as coculture (e.g. with Kupfer cells) will provide even better models of the human liver and the effects of chemicals to perturb inflammatory pathways in the liver.

Numerous studies have been reported using primary hepatocyte culture systems to assess induction of a variety of gene targets from both phase I and II enzymes, transporters, and genes involved in endogenous processes, such as gluconeogenesis (Raucy *et al.*, 2002, 2004; Raucy, 2003; Kodama *et al.*, 2004). In summary, primary cultures of human hepatocytes are an effective tool to assess the potential for human induction by drugs and drug candidates due to their more liver-like origin, cell morphology, metabolic competence, gene expression profile, and cell-signaling pathways. Technologies continue to emerge to better utilize these systems to probe human hepatic metabolism and drug interactions with new promising drugs.

## 11 INTERPRETATION AND QUANTITATIVE PREDICTION OF ENZYME INDUCTION

The common end points for assessment of enzyme induction are RNA expression (measured by real-time RT-PCR), protein level (measured by ELISA or Western blotting), and enzyme activity (measured by quantitating probe drug turnover or metabolite formation). Enzyme activity is the actual measure of increased enzyme capacity to turn over a substrate (which is the basis of a drug–drug interaction) and most accepted by regulatory agencies. Table 2 lists the most common *in vitro* positive control inducers and probe substrates for each CYP enzyme. However, quantitation of RNA expression and protein level often provide a more direct measurement of actual gene activation or increased enzyme concentration when the test article is also a potent enzyme inhibitor. Complications arise when interpreting enzyme activity measurements when a compound both induces and inhibits an enzyme. For example, ritonavir and troleandomycin are both potent inhibitors and inducers of CYP3A4 and result in increased CYP3A4 gene transcription, but generally exhibit normal or reduced enzyme activity (Luo *et al.*, 2002). Also, DPC681 is an inducer and metabolism-based inhibitor of CYP3A4 and an *in vitro* enzyme activity assay revealed no change in enzyme activity (Luo *et al.*, 2003). Additional *in vitro* studies in human hepatocytes indicated increased CYP3A4 protein concentrations and clinical studies revealed autoinduction when patients were dosed with DPC681 for several days. These examples illustrate the complicated nature of predicting drug interactions in patients when competing enzyme inhibition and induction occur *in vitro* and *in vivo*. In most cases, measuring RNA expression and enzyme activity is advised to take advantage of both methods (transcription and enzyme activity). Western blotting analysis is used less frequently in the industrial setting due to additional time and resources necessary to perform these analyses.

While *in vitro* receptor- or cell-based assays yield important qualitative assessments (mild, moderate, strong inducer) of enzyme induction, a quantitative assessment of how the induction will impact the exposure of a "victim" drug and resulting drug–drug interactions remains elusive. The data from *in vitro* experiments can take several forms,

such as fold increase over a vehicle-treated control incubation, percent increase of a positive control, or $EC_{50}$ value. A recently published draft FDA guidance for industry on drug interactions proposed two numerical end points to be used in the prediction: percent increase of a positive control and $EC_{50}$ (FDA, 2006). The guidance indicates that a drug that produces a change (enzyme activity) equal to or greater than 40% of the positive control is a potential inducer and requires further *in vivo* evaluation. As for the $EC_{50}$, the guidance suggests that this parameter can be used to compare the potency of different compounds. While simple rank ordering of $EC_{50}$ values and comparisons to known CYP3A4 inducers, such as rifampicin, can provide useful information on potency, more accurate predictions of drug interactions incorporate clinically relevant drug concentrations, such as total or free drug concentrations (portal vein, $C_{max}$, or steady state concentration), after a therapeutic dose of drug (Ripp *et al.*, 2006; Sinz *et al.*, 2006; Hewitt, de Kanter and LeCluyse, 2007). Percent activation at a therapeutic drug concentration ($C_{max}$) and rank ordering ($EC_{50}$) have been employed when evaluating PXR transactivation data (Sinz *et al.*, 2006). Using this approach, it was shown that compounds causing >40% transactivation at a therapeutic drug concentration are predicted to "likely" have drug interactions, compounds with percent transactivation between 15% and 40% "may" elicit drug interactions, and those with percent transactivation <15% are "not" anticipated to cause drug interactions (Sinz *et al.*, 2006). For example, hyperforin has an $EC_{50} = 0.04\,\mu M$ in the hPXR transactivation assay, an approximate efficacious plasma concentration of $0.3\,\mu M$, and a percent transactivation of 80% at the efficacious concentration in the hPXR transactivation assay. Pioglitazone has an $EC_{50} = 31\,\mu M$ in the hPXR transactivation assay, an approximate efficacious plasma concentration of $3.8\,\mu M$, and a percent transactivation of only 10% at the efficacious concentration. Contrasting these parameters between the two drugs, one would predict that hyperforin will cause CYP3A4 drug interactions while pioglitazone will not, similar to results observed in actual clinical drug interaction studies (Sinz *et al.*, 2006).

Combining values and obtaining a ratio of $E_{max}/EC_{50}$ ($E_{max}$ is maximum effect) can be predictive in assessing the likelihood of a drug interaction much like the approach used for drug inhibition assessments ($I/K_i$). More complex calculations include multiple parameters, such as $EC_{50}$, $E_{max}$, and drug concentration, which have been used to assess induction effects for multiple induction models (transactivation, immortalized hepatocytes, and primary hepatocytes). The fundamental equation used in these models is the same and can be described as follows:

$$E = \frac{C \times E_{max}}{C + EC_{50}}$$

where E is the inductive effect and C is the drug concentration. However, there is considerable discussion of which drug concentration to use in these calculations. Kato *et al.* showed that the steady-state concentration of unbound drug is useful in predicting clinical induction from human hepatocyte assays (Kato *et al.*, 2005) and Ripp *et al.* demonstrated that, using the Fa2N-4 cell line, unbound therapeutic plasma concentration of drug ($C_{max}$(fu)) can be used to determine a relative induction score (RIS) (Ripp *et al.*, 2006):

$$RIS = \frac{E_{max} \times C_{max}(fu)}{EC_{50} + C_{max}(fu)}$$

More recently, Hewitt *et al.* proposed that the NOEL (no observed effect level) be added to the equation to improve the predictive power (Hewitt, de Kanter and LeCluyse, 2007). The following equation describes a model that combines these parameters (where NOEL is defined as "the highest concentration at which no induction response is observed"):

$$\text{Induction risk factor} = \frac{E_{max} \times C_{max}(fu)}{EC_{50} + C_{max}(fu)} \times \frac{NOEL}{C_{max}(fu)}$$

Some difficulties exist when using cell-based systems, such as effects of protein binding, partial agonists, cytotoxicity, and insolubility of test compounds. While the effect of protein binding on enzyme induction in *in vitro* or *in vivo* systems is not clearly defined, the use of protein binding (i.e. ($C_{max}$(fu)) in some situations appears to be more predictive. However, the use of total drug concentration represents a more conservative approach which defines a maximal induction effect. Certainly more studies on the *in vitro* and *in vivo* relationship between protein binding and intracellular receptor-mediated effects will be required in the future.

Another caveat in applying these quantitative *in vitro* approaches, such as hPXR transactivation or cell-based induction assays, are concentration-response plots that reveal partial agonists or bell-shaped curves which make $EC_{50}$ and $E_{max}$ determinations difficult. In these situations, it may be difficult to establish a concentration–response relationship at concentrations above a certain threshold. In the future, a better understanding of molecular mechanisms of induction as well as advancement in tools to study the interaction will enable more sophisticated quantitative prediction of clinical enzyme induction. Cytotoxicity assessments should be conducted with all cell-based systems to ensure that a negative response is not due to cell death. Often this can be as simple as visualizing the cells under a microscope (cell morphology) or more sophisticated measures of cell membrane permeability or cell death. An understanding of drug solubility in cell culture media can also help avoid false-negative results when compounds are highly insoluble. Both cytotoxicity and insolubility often display bell-shaped response curves when inducers are evaluated, making it difficult to assess the full induction potential of such compounds.

Finally, the aforementioned methods work well for rank ordering of these compounds and predicting the magnitude of a drug interaction for compounds that are highly metabolized and metabolized exclusively by the individual CYP enzyme being induced (e.g. midazolam). However, for victim drugs whose elimination includes other pathways (i.e. renal or biliary) and/or whose metabolism is through multiple CYP enzymes, these methods can overestimate the change in exposure of a victim drug (Rowland and Tozer, 1988; Zhang *et al.*, 2007). Hence, a full understanding of elimination pathways (metabolic and non-metabolic) of victim drugs is necessary to predict exposure changes due to drug–drug interactions *in vivo*.

## 12 CLINICAL DRUG INTERACTION STUDIES

When a drug is predicted, based on *in vitro* data, to induce CYP enzymes in humans, then a clinical drug–drug interaction study is necessary to assess the magnitude of pharmacokinetic changes resulting from the induction. The purpose of these drug interaction studies is to determine if an interaction is significant enough to warrant

discontinuation, dosage adjustment, or therapeutic drug monitoring of the drug itself or a coadministered drug. In some instances, enzyme induction is not anticipated and a drug–drug interaction study may still be conducted in order to include labeling information indicating "no potential for drug interactions due to enzyme induction" (i.e. safety or marketing advantages). The majority of drug interaction clinical trials are related to CYP3A4, although other enzymes such as CYP2C9 or CYP1A2, as well as transporters such as MDR1, have been studied (Malmstrom *et al.*, 1998; Sahi *et al.*, 2003; Shadle *et al.*, 2004). Probe substrates recommended by the FDA for various CYP enzymes are listed in Table 1 (FDA, 2006). The general design of a CYP3A4 drug–drug interaction study includes 12–20 healthy volunteers (or sometimes patients) including males and females between the ages of 18 and 60. Patients are often required to abstain from medications, herbal supplements, or other dietary components, such as grapefruit juice, which may affect CYP3A4 enzyme levels during the study. Women who are included in such studies are often of non-childbearing age or required to use alternative methods of birth control other than oral contraceptives which may be affected during the trial.

In the case of CYP3A4 enzyme induction studies, midazolam either po or IV is used as a probe substrate to indirectly assess CYP3A4 enzyme changes by measuring pharmacokinetic changes of midazolam before and after drug treatment. Oral doses of midazolam capture the CYP3A4 induction effects from both the liver and intestine while an IV dose of midazolam measures the liver effect alone (Ma *et al.*, 2006a, b). Oral doses of midazolam range between 2 and 5 mg and IV doses are generally ~2 mg, while plasma sample collection periods for po and IV are typically 0.5–8 hours and 0.25–6 hours, respectively. Midazolam is dosed just prior to drug treatment and once again following several days of continuous drug treatment. The dose and duration of drug administration (perpetrator-inducer) between the two doses of midazolam is based on the expected use for drug treatment. For both perpetrator and victim, the dosages should be employed that maximize the likelihood of finding an interaction; hence, the highest approved or predicted dose of each drug should be used in such studies. Typical treatment durations are approximately seven days, as illustrated in Fig. 1, during which several samples are taken to measure perpetrator drug levels. Also, a third or multiple midazolam doses may be given several days after the last dose of drug in order to assess the de-induction phase and determine when patients have restabilized to basal CYP3A4 levels. Depending on the half-life of the compound, the time

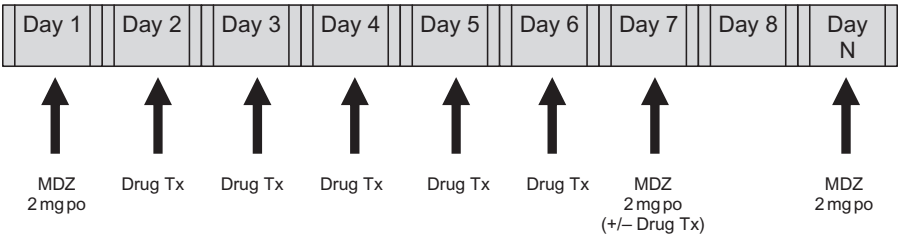


**Figure 1** A typical CYP3A4 clinical drug–drug interaction design. The probe drug, midazolam (MDZ), is dosed before and after several days of drug treatment (Drug Tx). MDZ may also be administered sometime after the last drug treatment (Day N) to assess the de-induction phase.

needed to return to normal CYP3A4 levels may vary. For example, the CYP3A4 de-induction phase of pleconaril was studied for a length of 34 days due to the 180-hour half-life of the drug, and it was determined that enzyme levels returned to normal sometime between days 6 and 13 after the last dose of pleconaril (Ma *et al.*, 2006a). Most often, a drug with a reasonable half-life will allow enzyme levels to return to pretreatment levels within three to four days.

Typical pharmacokinetic parameters that are compared between the two doses of midazolam include AUC and $C_{max}$. The least squares geometric mean ratios with 90% confidence intervals for midazolam AUC and $C_{max}$ are calculated and a drug interaction is established if these values fall outside of the predefined and generally accepted bioequivalence range of 0.80 and 1.25. Relevant *in vitro* or *in vivo* drug interaction information describing the drug's inducing effects on substrates is presented under Drug–Drug Interactions of the Clinical Pharmacology section of the label. Often, additional information describing the interaction and steps necessary to avoid interactions will be presented in the Precautions, Warnings/Precautions, Dosage and Administration, and/or Contraindication sections as necessary (FDA, 1999).

## 13 INDUSTRIAL AND REGULATORY PERSPECTIVES

In an industrial setting, there is a balance of resources dedicated toward first screening compounds for the potential to be enzyme inducers and second, greater characterization of positive hits from screening assays. Most often, the earlier screening assays are high throughput and have the ability to accommodate large numbers of compounds in a short period of time. These characteristics, coupled with a low cost per compound, allow such assays to be placed in early drug discovery. Most often, these high-throughput assays are PXR binding or transactivation assays in 96- or 384-well formats. Although excellent screening and sometimes predictive of human induction potential, assays based on receptors and transcription factors are considered supportive evidence for a compound's induction potential by the FDA as they do not directly reflect the actual enzyme activity (FDA, 2006; Sinz *et al.*, 2006).

Gaining in popularity are the immortalized hepatocyte models which afford many of the advantages of nuclear receptor-based assays (availability and speed) and often incorporate more than one receptor-mediated process in a single system (e.g. Fa2N-4 cell line incorporates CYP1A and 3A induction processes). In many ways, these are similar to primary human hepatocyte experiments and accepted by the FDA as models to assess enzyme induction. However, caution should be noted when using immortalized hepatocyte models, in that they do not incorporate all potential nuclear hormone/transcription factor-mediated CYP induction pathways. Even when CYP1A and CYP3A4 are shown to properly induce with known positive controls, this implies that AhR- and PXR-specific pathways are functioning properly, but not necessarily all possible pathways. For example, CYP3A4 enzyme induction, which can be mediated by both PXR and CAR, may demonstrate an attenuated CYP3A4 induction response in cell-based models that only incorporate PXR and not both PXR and CAR (e.g. Fa2N-4) (Hariparsad *et al.*, 2008). Nevertheless, immortalized hepatocytes are an acceptable model to screen and/or characterize the enzyme induction potential of drug candidates. For the most part, immortalized hepatocytes can be used in conjunction with primary hepatocytes to better characterize the enzyme induction potential but not in lieu of primary human hepatocytes.

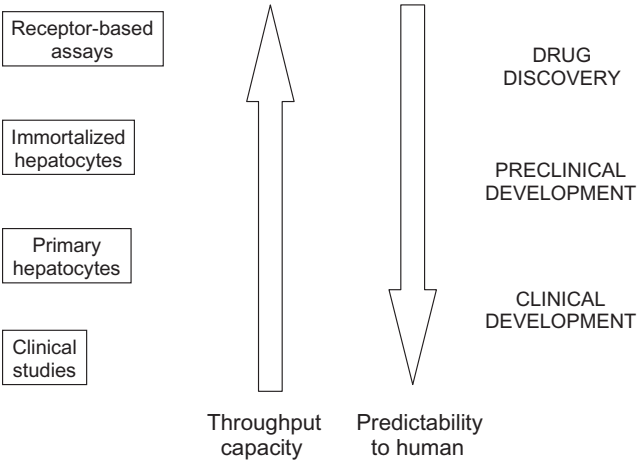


**Figure 2** Illustration of *in vitro* models and clinical studies used to predict or assess the enzyme induction potential of a drug candidate.

Figure 2 describes the variety of assays, throughput capacity, predictability to human drug–drug interactions, and approximate time frame that each assay may be employed during drug discovery/development. Although Fig. 2 incorporates all possible *in vitro* human model systems, not all companies will employ the entire suite of assays. In the very least, nearly all companies will eventually incorporate primary human hepatocytes as an induction model to assess PXR- and AhR-mediated enzyme induction (CYP3A4/2C/2B6 and CYP1A, respectively), as this is the standard procedure to assess CYP induction prior to clinical drug–drug interaction studies.

As described here and in the previous chapter, there exist significant species differences in ligand affinities for various nuclear hormone receptors or transcription factors. Hence, most animal models do not properly recapitulate the effects observed in humans, and *in vivo* animal models (mice, rats, dogs) are generally not applicable in predicting enzyme induction potential in humans. Enzyme induction in animals is, however, problematic when assessing the toxicological properties of a new drug candidate. Autoinduction in animals can affect drug exposure to the extent that high drug exposures necessary to assess toxicity cannot be achieved. This poses a serious problem when attempting to achieve adequate exposure–safety margins in the absence of increasing drug exposures with increasing doses of drug.

## 14 FUTURE MODELS

There are multiple new models of enzyme induction being developed with varying degrees of success and validation. All appear promising; however, additional evaluation and a greater understanding of how best to employ these models are still necessary.

*In silico* simulations and structure–activity relationships (SARs) are useful tools to help eliminate potentially problematic chemotypes early in drug development. Virtual screening can be conducted on large numbers of compounds, for example, chemical

library screening. In addition, these models may be useful in defining structural features or structural modifications necessary to eliminate or attenuate the affinity of a ligand to PXR. Successful induction models generally combine nuclear receptor crystal structures with *in silico* structural biology (Handschin, Podvinec and Meyer, 2003). However, merely docking the compound into the ligand-binding domain may not be sufficient in capturing a complete picture of transcriptional events. For example, the large ligand-binding domain of hPXR accommodates ligands of all sizes. Larger ligands can alter the protein structure after binding and smaller molecules have been shown to bind at several sites within the LBD (Handschin, Podvinec and Meyer, 2003). Another complicating factor in ligand-binding pharmacophore models are adjacent sites on the nuclear receptor (allosteric sites) which can be involved in conformational changes that affect ligand or DNA binding. For example, Ekins *et al.* reported that agonists and antagonists bind to distinct regions on PXR and that the effect of ketoconazole on CYP3A transcription was mediated through binding to the AF-2 region of PXR and not the LBD (Ekins *et al.*, 2007; Wang *et al.*, 2007).

"Stem cells" are defined as "omnipotent progenitor cells that retain the ability to replicate indefinitely *in vitro* as undifferentiated cells and under appropriate conditions, differentiate into specialized cell types, such as hepatocytes, neurons and cardiomyocytes" (Pouton and Haynes, 2005; Vermeir *et al.*, 2005). Stem cells can be derived from the inner cell mass of a blastocyst (embryonic stem cell, ESC) or can be obtained from a variety of tissues, such as bone marrow, cord blood, liver, and adipose tissue (adult stem cell, ASC) (McNeish, 2004; Pouton and Haynes, 2007).

Petersen *et al.* demonstrated that bone marrow stem cells can be a potential source of hepatic oval cells, which can further differentiate into two types of liver epithelia cells: ductular cells and hepatocytes (Petersen *et al.*, 1999). Subsequently, it was found that human and murine hematopoietic stem cells in bone marrow can differentiate into hepatocyte-like cells (Alison *et al.*, 2000; Lagasse *et al.*, 2000), opening the possibility of generating functional hepatocytes from non-hepatic adult stem cells. In general, human stem cells express low levels of CYPs but higher levels of GST (Shao *et al.*, 2007). Under the presence of fibroblast growth factor (FGF) and hepatocyte growth factor (HGF), they differentiate into functional hepatocyte-like cells that express high levels of PB-inducible CYP activity and some selected liver-specific functions, such as secretion of urea and albumin and LDL uptake (Schwartz *et al.*, 2002; Lee *et al.*, 2004).

Embryonic stem cells provide another potential source of hepatocytes. The most comprehensive analysis of expression and induction of drug-metabolizing enzymes in human embryonic stem cells was reported by Ek *et al.* (2007). Hepatocyte-like cells derived from hESC under the influence of bFGF expressed CYP1A2, CYP3A4/7, and low levels of CYP1A1 and CYP2C proteins. When treated with a cocktail of known CYP inducers, the protein expressions of CYP1A2 and CYP3A4 were also increased, suggesting the inducibility of these genes is retained. However, the activities of CYP1A1 and CYP3A4 were lower than those in primary hepatocytes (3% and 6%, respectively). An examination of CYP expression in hepatocytes derived from murine ESCs also demonstrated expression of Cyp2b10, Cyp2c29, Cyp2d9, Cyp3a11, and Cyp7a1 as well as PB-inducible CYP activities (Tsutsui *et al.*, 2006). While stem cells are not widely used in ADMET research today (Sinz and Kim, 2006), advancements in stem cell technology and characterization will greatly expand their application in drug metabolism and safety assessment.

Humanized nuclear receptor transgenic mice and knockout mice can be useful in assessing *in vivo* enzyme induction or mechanisms of enzyme induction as described

in the previous chapter. A newer humanized model, known as the chimeric mouse, involves transplanting human hepatocytes into a mouse liver, thus providing a useful *in vivo* model capable of assessing enzyme regulation by drug candidates. Humanized chimeric mice were developed by injecting human hepatocytes into SCID mice (80% of the hepatocytes in the liver were of human origin) (Tateno *et al.*, 2004). CYP1A1/2 and CYP3A4 mRNA, protein content, and enzyme activity have been shown to be induced in chimeric mice treated with known human inducers and in hepatocytes isolated from humanized chimeric mice (Katoh *et al.*, 2005a, b; Nishimura *et al.*, 2005). These *in vivo* models (transgenic or chimeric) provide an advantage over previous *in vitro* models, as the systems are dynamic, with drug absorption, distribution, metabolism, and elimination occurring naturally (albeit a humanized mouse). Drug candidates can be administered over several days at doses/exposures equivalent to those expected in humans to enable improved assessments of human induction potential.

Recently, there is an increased effort to validate nonhuman primates (rhesus or cynomolgus monkey) as an animal model for CYP3A induction. The high sequence homology (~96%) between human and monkey PXR is presumed to result in similar induction profiles (Prueksaritanont *et al.*, 2006; Nishimura *et al.*, 2007). CYP1A, -2B, and -3A in the cynomolgus monkey were found to respond to known human inducers, such as β-naphthoflavone, phenobarbital, and dexamethasone, respectively, and CYP3A64 responded to rifampicin induction in the rhesus monkey (Bullock *et al.*, 1995; Prueksaritanont *et al.*, 2006). A more recent *in vitro* report by Nishimura *et al.* indicated that cynomolgus monkey hepatocytes responded to CYP1A and -3A inducers; however, only CYP1A1 responded in the appropriate manner to omeprazole treatment and CYP1A2 did not (Nishimura *et al.*, 2007). This indicates some unknowns still exist due either to the enzyme inducer, omeprazole, or the mechanism(s) of CYP1A2 induction between monkeys and humans. However, if successfully validated, monkeys could provide a useful *in vivo* model to assess human induction at therapeutically relevant concentrations, thus providing a more quantitative prediction of enzyme induction.

## 15 CONCLUSIONS

One of the biggest challenges in predicting human drug–drug interactions from *in vitro* experiments is the lack of understanding between *in vitro* drug concentrations and drug concentration at the site of action in the liver. This is complicated by several factors such as protein binding, nonspecific binding *in vitro*, the use of total versus free (unbound) efficacious plasma concentrations, or the use of systemic versus portal vein drug concentrations, and compounds with extremely low or high liver-to-plasma ratios. The two most common approaches in all human *in vitro* models are (i) total efficacious systemic $C_{max}$ or $C_{ss}$ concentration or (ii) unbound drug concentration at the portal vein (Ripp *et al.*, 2006; Sinz *et al.*, 2006; Hewitt, de Kanter and LeCluyse, 2007). The former approach does not account for protein binding effects on drug disposition, yet is considered a conservative estimate that uses a high (total) drug concentration after the first-pass effect (metabolism/elimination). The latter approach does account for protein binding on drug disposition and reflects the maximum drug absorbed and presented to the liver.

Equally challenging is the extrapolation from *in vitro* induction studies when the compound being evaluated is both an inhibitor and inducer of a CYP enzyme. If the compound is an inhibitor, enzyme induction can be overlooked when only enzyme

activity is assessed. Due to the common nature of CYP inhibition, conducting both activity and mRNA analysis is encouraged. Even if a complete understanding of the separate *in vitro* inhibition and induction potentials of a compound is assessed, the combined effects *in vivo* are difficult to predict (e.g. DPC681, Luo *et al.*, 2003).

In summary, the last decade has seen a significant progression forward in our understanding and the predictability of enzyme induction, both *in vitro* and *in vivo*. The use of nuclear hormone receptor and hepatocyte-based cellular assays, combined with our improved understanding of the mechanisms leading to enzyme and transporter induction, has allowed us to predict and understand enzyme induction to a much greater degree than in the past.

## ACKNOWLEDGMENTS

The authors wish to thank Drs Jasminder Sahi, Kenneth Santone, and Gillian Wallace for their contributions and comments.

## REFERENCES

Abdel-Razzak, Z., Corcos, L., Fautrel, A., Guillouzo, A. (1995). Interleukin-1 beta antagonizes phenobarbital induction of several major cytochromes P450 in adult rat hepatocytes in primary culture. *FEBS Lett.*, *366* (2–3), 159–164.

Aiba, T., Susa, M., Fukumori, S., Hashimoto, Y. (2005). The effects of culture conditions on CYP3A4 and MDR1 mRNA induction by 1alpha,25-dihydroxyvitamin D(3) in human intestinal cell lines, Caco-2 and LS180. *Drug Metab. Pharmacokinet.*, *20* (4), 268–274.

Akiyama, I., Tomiyama, K., Sakaguchi, M. *et al.* (2004). Expression of CYP3A4 by an immortalized human hepatocyte line in a three-dimensional culture using a radial-flow bioreactor. *Int. J. Mol. Med.*, *14* (4), 663–668.

Alison, M.R., Poulsom, R., Jeffery, R. *et al.* (2000). Hepatocytes from non-hepatic adult stem cells. *Nature*, *406* (6793), 257.

Aninat, C., Piton, A., Glaise, D. *et al.* (2006). Expression of cytochromes P450, conjugating enzymes and nuclear receptors in human hepatoma HepaRG cells. *Drug Metab. Dispos.*, *34* (1), 75–83.

Auerbach, S.S., Stoner, M.A., Su, S., Omiecinski, CJ. (2005). Retinoid X receptor-alpha-dependent transactivation by a naturally occurring structural variant of human constitutive androstane receptor (NR1I3). *Mol. Pharmacol.*, *68* (5), 1239–1253.

Baliharova, V., Skalova, L., Maas, R.F., De Vrieze, G., Bull, S., Fink-Gremmels, J. (2003). The effects of benzimidazole anthelmintics on P4501A in rat hepatocytes and HepG2 cells. *Res. Vet. Sci.*, *75* (1), 61–69.

Bapiro, T.E., Andersson, T.B., Otter, C., Hasler, J.A., Masimirembwa, C.M. (2002). Cytochrome P450 1A1/2 induction by antiparasitic drugs: dose-dependent increase in ethoxyresorufin O-deethylase activity and mRNA caused by quinine, primaquine and albendazole in HepG2 cells. *Eur. J. Clin. Pharmacol.*, *58* (8), 537–542.

Bertilsson, L., Hojer, B., Tybring, G., Osterloh, J., Rane, A. (1980). Autoinduction of carbamazepine metabolism in children examined by a stable isotope technique. *Clin. Pharmacol. Ther.*, *27* (1), 83–88.

Bi, Y.A., Kazolias, D., Duignan, D.B. (2006). Use of cryopreserved human hepatocytes in sandwich culture to measure hepatobiliary transport. *Drug Metab. Dispos.*, *34* (9), 1658–1665.

Brandin, H., Viitanen, E., Myrberg, O., Arvidsson, A.K. (2007). Effects of herbal medicinal products and food supplements on induction of CYP1A2, CYP3A4 and MDR1 in the human colon carcinoma cell line LS180. *Phytother. Res.*, *21* (3), 239–244.

Brown, R.R., Miller, J.A., Miller, E.C. (1954). The metabolism of methylated aminoazo dyes. IV. Dietary factors enhancing demethylation in vitro. *J. Biol. Chem.*, *209* (1), 211–222.

Bullock, P., Pearce, R., Draper, A. *et al.* (1995). Induction of liver microsomal cytochrome P450 in cynomolgus monkeys. *Drug. Metab. Dispos.*, *23* (7), 736–748.

Cascio, S.M. (2001). Novel strategies for immortalization of human hepatocytes. *Artif. Organs*, *25* (7), 529–538.

Chang, T.K., Waxman, D.J. (2006). Synthetic drugs and natural products as modulators of constitutive androstane receptor (CAR) and pregnane X receptor (PXR). *Drug Metab. Rev.*, *38* (1–2), 51–73.

Chen, J., Nikolova-Karakashian, M., Merrill, A.H., Jr, Morgan, E.T. (1995a). Regulation of cytochrome P450 2C11 (CYP2C11) gene expression by interleukin-1, sphingomyelin hydrolysis, and ceramides in rat hepatocytes. *J. Biol. Chem.*, *270* (42), 25233–25238.

Chen, J.Q., Strom, A., Gustafsson, J.A., Morgan, E.T. (1995b). Suppression of the constitutive expression of cytochrome P-450 2C11 by cytokines and interferons in primary cultures of rat hepatocytes: comparison with induction of acute-phase genes and demonstration that CYP2C11 promoter sequences are involved in the suppressive response to interleukins 1 and 6. *Mol. Pharmacol.*, *47* (5), 940–947.

Chen, Y., Ferguson, S.S., Negishi, M., Goldstein, J.A. (2003). Identification of constitutive androstane receptor and glucocorticoid receptor binding sites in the CYP2C19 promoter. *Mol. Pharmacol.*, *64* (2), 316–324.

Chen, Y., Ferguson, S.S., Negishi, M., Goldstein, J.A. (2004). Induction of human CYP2C9 by rifampicin, hyperforin, and phenobarbital is mediated by the pregnane X receptor. *J. Pharmacol. Exp. Ther.*, *308* (2), 495–501.

Conney, A.H. (1967). Pharmacological implications of microsomal enzyme induction. *Pharmacol. Rev.*, *19* (3), 317–366.

Daujat, M., Peryt, B., Lesca, P., Fourtanier, G., Domergue, J., Maurel, P. (1992). Omeprazole, an inducer of human CYP1A1 and 1A2, is not a ligand for the Ah receptor. *Biochem. Biophys. Res. Commun.*, *188* (2), 820–825.

Dickins, M. (2004). Induction of cytochromes P450. *Curr. Top. Med. Chem.*, *4* (16), 1745–1766.

Dilger, K., Zheng, Z., Klotz, U. (1999). Lack of drug interaction between omeprazole, lansoprazole, pantoprazole and theophylline. *Br. J. Clin. Pharmacol.*, *48* (3), 438–444.

Dvorak, Z., Modriansky, M., Pichard-Garcia, L. *et al.* (2003). Colchicine down-regulates cytochrome P450 2B6, 2C8, 2C9, and 3A4 in human hepatocytes by affecting their glucocorticoid receptor-mediated regulation. *Mol. Pharmacol.*, *64* (1), 160–169.

Ek, M., Soderdahl, T., Kuppers-Munther, B. *et al.* (2007). Expression of drug metabolizing enzymes in hepatocyte-like cells derived from human embryonic stem cells. *Biochem. Pharmacol.*, *74* (3), 496–503.

Ekins, S., Chang, C., Mani, S. *et al.* (2007). Human pregnane x receptor antagonists and agonists define molecular requirements for different binding sites. *Mol. Pharmacol.*, *72* (3), 592–603.

El-Sankary, W., Gibson, G.G., Ayrton, A., Plant, N. (2001). Use of a reporter gene assay to predict and rank the potency and efficacy of CYP3A4 inducers. *Drug Metab. Dispos.*, *29* (11), 1499–1504.

Engman, H.A., Lennernas, H., Taipalensuu, J., Otter, C., Leidvik, B., Artursson, P. (2001). CYP3A4, CYP3A5, and MDR1 in human small and large intestinal cell lines suitable for drug transport studies. *J. Pharm. Sci.*, *90* (11), 1736–1751.

Fabre, N., Arrivet, E., Trancard, J. *et al.* (2003). A new hepatoma cell line for toxicity testing at repeated doses. *Cell Biol. Toxicol.*, *19* (2), 71–82.

Faucette, S.R., Sueyoshi, T., Smith, C.M., Negishi, M., LeCluyse, E.L., Wang, H. (2006). Differential regulation of hepatic CYP2B6 and CYP3A4 genes by constitutive androstane receptor but not pregnane X receptor. *J. Pharmacol. Exp. Ther.*, *317* (3), 1200–1209.

Faucette, S.R., Zhang, T.C., Moore, R. *et al.* (2007). Relative activation of human pregnane X receptor versus constitutive androstane receptor defines distinct classes of CYP2B6 and CYP3A4 inducers. *J. Pharmacol. Exp. Ther.*, *320* (1), 72–80.

FDA (1999). Guidance for industry in vivo drug metabolism/drug interaction studies—study, design, data analysis, and recommendations for dosing and labelling.

FDA (2006). FDA Draft Guideline: drug interaction studies—study design, data analysis, and implications for dosing and labeling.

Ferguson, S.S., Chen, Y., LeCluyse, E.L., Negishi, M., Goldstein, JA. (2005). Human CYP2C8 is transcriptionally regulated by the nuclear receptors constitutive androstane receptor, pregnane X receptor, glucocorticoid receptor, and hepatic nuclear factor 4alpha. *Mol. Pharmacol.*, *68* (3), 747–757.

Gomez-Lechon, M.J., Donato, T., Jover, R. *et al.* (2001). Expression and induction of a large set of drug-metabolizing enzymes by the highly differentiated human hepatoma cell line BC2. *Eur. J. Biochem.*, *268* (5), 1448–1459.

Gonzalez, F.J. (2007). The 2006 Bernard B. Brodie Award Lecture. Cyp2e1. *Drug Metab. Dispos.*, *35* (1), 1–8.

Goodwin, B., Hodgson, E., Liddle, C. (1999). The orphan human pregnane X receptor mediates the transcriptional activation of CYP3A4 by rifampicin through a distal enhancer module. *Mol. Pharmacol.*, *56* (6), 1329–1339.

Gripon, P., Rumin, S., Urban, S. *et al.* (2002). Infection of a human hepatoma cell line by hepatitis B virus. *Proc. Natl. Acad. Sci. U.S.A.*, *99* (24), 15655–15660.

Guillouzo, A., Corlu, A., Aninat, C., Glaise, D., Morel, F., Guguen-Guillouzo, C. (2007). The human hepatoma HepaRG cells: a highly differentiated model for studies of liver metabolism and toxicity of xenobiotics. *Chem Biol. Interact.*, *168* (1), 66–73.

Hahn, W.C. (2002). Immortalization and transformation of human cells. *Mol. Cells*, *13* (3), 351–361.

Hamilton, G.A., Jolley, S.L., Gilbert, D., Coon, D.J., Barros, S., LeCluyse, E.L. (2001). Regulation of cell morphology and cytochrome P450 expression in human hepatocytes by extracellular matrix and cell-cell interactions. *Cell Tissue Res.*, *306* (1), 85–99.

Handschin, C., Podvinec, M., Meyer, U.A. (2003). In silico approaches, and in vitro and in vivo experiments to predict induction of drug metabolism. *Drug News Perspect.*, *16* (7), 423–434.

Hariparsad, N., Carr, B., Evers, R., Chu, X. (2008). Comparison of immortalized Fa2N-4 cells and human hepatocytes as in vitro models for cytochrome P450 induction. *Drug Metab. Dispos.*, *36* (6), 1046–1055.

Harmsen, S., Meijerman, I., Beijnen, J.H., Schellens, J.H. (2007). The role of nuclear receptors in pharmacokinetic drug-drug interactions in oncology. *Cancer Treat. Rev.*, *33* (4), 369–380.

Hewitt, N.J., Hewitt, P. (2004). Phase I and II enzyme characterization of two sources of HepG2 cell lines. *Xenobiotica*, *34* (3), 243–256.

Hewitt, N.J., de Kanter, R., LeCluyse, E. (2007). Induction of drug metabolizing enzymes: a survey of in vitro methodologies and interpretations used in the pharmaceutical industry—do they comply with FDA recommendations? *Chem. Biol. Interact.*, *168* (1), 51–65.

Hewitt, N.J., Lechon, M.J., Houston, J.B. *et al.* (2007). Primary hepatocytes: current understanding of the regulation of metabolic enzymes and transporter proteins, and pharmaceutical practice for the use of hepatocytes in metabolism, enzyme induction, transporter, clearance, and hepatotoxicity studies. *Drug Metab. Rev.*, *39* (1), 159–234.

Honkakoski, P., Sueyoshi, T., Negishi, M. (2003). Drug-activated nuclear receptors CAR and PXR. *Ann. Med.*, *35* (3), 172–182.

Hu, W., Sorrentino, C., Denison, M.S., Kolaja, K., Fielden, M.R. (2007). Induction of cyp1a1 is a nonspecific biomarker of aryl hydrocarbon receptor activation: results of large scale screening of pharmaceuticals and toxicants in vivo and in vitro. *Mol. Pharmacol.*, *71* (6), 1475–1486.

Iwanari, M., Nakajima, M., Kizu, R., Hayakawa, K., Yokoi, T. (2002). Induction of CYP1A1, CYP1A2, and CYP1B1 mRNAs by nitropolycyclic aromatic hydrocarbons in various human tissue-derived cells: chemical-, cytochrome P450 isoform-, and cell-specific differences. *Arch. Toxicol.*, *76* (5–6), 287–298.

Jackson, J.P., Ferguson, S.S., Moore, R., Negishi, M., Goldstein, J.A. (2004). The constitutive active/androstane receptor regulates phenytoin induction of Cyp2c29. *Mol. Pharmacol.*, *65* (6), 1397–1404.

Jackson, J.P., Ferguson, S.S., Negishi, M., Goldstein, J.A. (2006). Phenytoin induction of the cyp2c37 gene is mediated by the constitutive androstane receptor. *Drug Metab. Dispos.*, *34* (12), 2003–2010.

Jacobson, L., Middleton, B., Holmgren, J. *et al.* (2007). An optimized automated assay for determination of metabolic stability using hepatocytes: assay validation, variance component analysis, and in vivo relevance. *Assay Drug Dev. Technol.*, *5* (3), 403–415.

de Jonge, M.E., Huitema, A.D., van Dam, S.M., Beijnen, J.H., Rodenhuis, S. (2005). Significant induction of cyclophosphamide and thiotepa metabolism by phenytoin. *Cancer Chemother. Pharmacol.*, *55* (5), 507–510.

Kato, M., Chiba, K., Horikawa, M., Sugiyama, Y. (2005). The quantitative prediction of in vivo enzyme-induction caused by drug exposure from in vitro information on human hepatocytes. *Drug Metab. Pharmacokinet.*, *20* (4), 236–243.

Katoh, M., Matsui, T., Nakajima, M. *et al.* (2005a). In vivo induction of human cytochrome P450 enzymes expressed in chimeric mice with humanized liver. *Drug Metab. Dispos.*, *33* (6), 754–763.

Katoh, M., Watanabe, M., Tabata, T. *et al.* (2005b). In vivo induction of human cytochrome P450 3A4 by rifabutin in chimeric mice with humanized liver. *Xenobiotica*, *35* (9), 863–875.

Kawamoto, T., Sueyoshi, T., Zelko, I., Moore, R., Washburn, K., Negishi, M. (1999). Phenobarbital-responsive nuclear translocation of the receptor CAR in induction of the CYP2B gene. *Mol. Cell. Biol.*, *19* (9), 6318–6322.

Kliewer, S.A., Goodwin, B., Willson, T.M. (2002). The nuclear pregnane X receptor: a key regulator of xenobiotic metabolism. *Endocr. Rev.*, *23* (5), 687–702.

Kobayashi, N., Fujiwara, T., Westerman, K.A. *et al.* (2000). Prevention of acute liver failure in rats with reversibly immortalized human hepatocytes. *Science*, *287* (5456), 1258–1262.

Kodama, S., Koike, C., Negishi, M., Yamamoto, Y. (2004). Nuclear receptors CAR and PXR cross talk with FOXO1 to regulate genes that encode drug-metabolizing and gluconeogenic enzymes. *Mol. Cell. Biol.*, *24* (18), 7931–7940.

Lagasse, E., Connors, H., Al-Dhalimy, M. *et al.* (2000). Purified hematopoietic stem cells can differentiate into hepatocytes in vivo. *Nat. Med.*, *6* (11), 1229–1234.

Le Jossic, C., Glaise, D., Corcos, L. *et al.* (1996). trans-Acting factors, detoxication enzymes and hepatitis B virus replication in a novel set of human hepatoma cell lines. *Eur. J. Biochem.*, *238* (2), 400–409.

Le Vee, M., Jigorel, E., Glaise, D., Gripon, P., Guguen-Guillouzo, C., Fardel, O. (2006). Functional expression of sinusoidal and canalicular hepatic drug transporters in the differentiated human hepatoma HepaRG cell line. *Eur. J. Pharm. Sci.*, *28* (1–2), 109–117.

LeCluyse, E.L., Alexandre, E., Hamilton, G.A. *et al.* (2005). Isolation and culture of primary human hepatocytes. *Methods Mol. Biol.*, *290*, 207–229.

LeCluyse, E.L., Madan, A., Hamilton, G., Carroll, K., DeHaan, R., Parkinson, A. (2000). Expression and regulation of cytochrome P450 enzymes in primary cultures of human hepatocytes. *J. Biochem. Mol. Toxicol.*, *14* (4), 177–188.

Lee, K.D., Kuo, T.K., Whang-Peng, J. *et al.* (2004). In vitro hepatic differentiation of human mesenchymal stem cells. *Hepatology*, *40* (6), 1275–1284.

Lemaire, G., de Sousa, G., Rahmani, R.A. (2004). PXR reporter gene assay in a stable cell culture system: CYP3A4 and CYP2B6 induction by pesticides. *Biochem. Pharmacol.*, *68* (12), 2347–2358.

Li, Q., Sai, Y., Kato, Y., Tamai, I., Tsuji, A. (2003). Influence of drugs and nutrients on transporter gene expression levels in Caco-2 and LS180 intestinal epithelial cell lines. *Pharm. Res.*, *20* (8), 1119–1124.

Luo, G., Cunningham, M., Kim, S. *et al.* (2002). CYP3A4 induction by drugs: correlation between a pregnane X receptor reporter gene assay and CYP3A4 expression in human hepatocytes. *Drug Metab. Dispos.*, *30* (7), 795–804.

Luo, G., Lin, J., Fiske, W.D. *et al.* (2003). Concurrent induction and mechanism-based inactivation of CYP3A4 by an L-valinamide derivative. *Drug Metab. Dispos.*, *31* (9), 1170–1175.

Ma, J.D., Nafziger, A.N., Rhodes, G., Liu, S., Bertino, J.S., Jr (2006a). Duration of pleconaril effect on cytochrome P450 3A activity in healthy adults using the oral biomarker midazolam. *Drug Metab. Dispos.*, *34* (5), 783–785.

Ma, J.D., Nafziger, A.N., Rhodes, G., Liu, S., Gartung, A.M., Bertino, J.S., Jr (2006b). The effect of oral pleconaril on hepatic cytochrome P450 3A activity in healthy adults using intravenous midazolam as a probe. *J. Clin. Pharmacol.*, *46* (1), 103–108.

McGinnity, D.F., Berry, A.J., Kenny, J.R., Grime, K., Riley, R.J. (2006). Evaluation of time-dependent cytochrome P450 inhibition using cultured human hepatocytes. *Drug Metab. Dispos.*, *34* (8), 1291–1300.

McNeish, J. (2004). Embryonic stem cells in drug discovery. *Nat. Rev. Drug Discov.*, *3* (1), 70–80.

Malmstrom, K., Schwartz, J., Reiss, T.F. *et al.* (1998). Effect of montelukast on single-dose theophylline pharmacokinetics. *Am. J. Ther.*, *5* (3), 189–195.

Mandal, P.K. (2005). Dioxin: a review of its environmental effects and its aryl hydrocarbon receptor biology. *J. Comp. Physiol. [B]*, *175* (4), 221–230.

Marion, T.L., Leslie, E.M., Brouwer, K.L. (2007). Use of sandwich-cultured hepatocytes to evaluate impaired bile acid transport as a mechanism of drug-induced hepatotoxicity. *Mol. Pharm.*, *4* (6), 911–918.

Medina-Diaz, I.M., Arteaga-Illan, G., de Leon, M.B. *et al.* (2007). Pregnane X receptor-dependent induction of the CYP3A4 gene by o,p′-1,1,1,-trichloro-2,2-bis (p-chlorophenyl)ethane. *Drug Metab. Dispos.*, *35* (1), 95–102.

Medina-Diaz, I.M., Elizondo, G. (2005). Transcriptional induction of CYP3A4 by o,p′-DDT in HepG2 cells. *Toxicol. Lett.*, *157* (1), 41–47.

Mills, J.B., Rose, K.A., Sadagopan, N., Sahi, J., de Morais, S.M. (2004). Induction of drug metabolism enzymes and MDR1 using a novel human hepatocyte cell line. *J. Pharmacol. Exp. Ther.*, *309* (1), 303–309.

Mueller, S.C., Majcher-Peszynska, J., Uehleke, B. *et al.* (2006). The extent of induction of CYP3A by St. John's wort varies among products and is linked to hyperforin dose. *Eur J. Clin. Pharmacol.*, *62* (1), 29–36.

Nishimura, M., Imai, T., Morioka, Y., Kuribayashi, S., Kamataki, T., Naito, S. (2004). Effects of NO-1886 (Ibrolipim), a lipoprotein lipase-promoting agent, on gene induction of cytochrome P450s, carboxylesterases, and sulfotransferases in primary cultures of human hepatocytes. *Drug Metab. Pharmacokinet.*, *19* (6), 422–429.

Nishimura, M., Koeda, A., Suganuma, Y. *et al.* (2007). Comparison of inducibility of CYP1A and CYP3A mRNAs by prototypical inducers in primary cultures of human, cynomolgus monkey, and rat hepatocytes. *Drug Metab. Pharmacokinet.*, *22* (3), 178–186.

Nishimura, M., Yokoi, T., Tateno, C. *et al.* (2005). Induction of human CYP1A2 and CYP3A4 in primary culture of hepatocytes from chimeric mice with humanized liver. *Drug Metab. Pharmacokinet.*, *20* (2), 121–126.

O'Connor, J.E., Martinez, A., Castell, J.V., Gomez-Lechon, M.J. (2005). Multiparametric characterization by flow cytometry of flow-sorted subpopulations of a human hepatoma cell line useful for drug research. *Cytometry A*, *63* (1), 48–58.

Okitsu, T., Kobayashi, N., Jun, H.S. *et al.* (2004). Transplantation of reversibly immortalized insulin-secreting human hepatocytes controls diabetes in pancreatectomized pigs. *Diabetes*, *53* (1), 105–112.

Omura, T., Sato, R. (1962). A new cytochrome in liver microsomes. *J. Biol. Chem.*, *237*, 1375–1376.

Park, J.Y., Kim, K.A., Kang, M.H., Kim, S.L., Shin, J.G. (2004). Effect of rifampin on the pharmacokinetics of rosiglitazone in healthy subjects. *Clin. Pharmacol. Ther.*, *75* (3), 157–162.

Pascussi, J.M., Drocourt, L., Gerbal-Chaloin, S., Fabre, J.M., Maurel, P., Vilarem, M.J. (2001). Dual effect of dexamethasone on CYP3A4 gene expression in human hepatocytes. Sequential role of glucocorticoid receptor and pregnane X receptor. *Eur. J. Biochem.*, *268* (24), 6346–6358.

Petersen, B.E., Bowen, W.C., Patrene, K.D. *et al.* (1999). Bone marrow as a potential source of hepatic oval cells. *Science*, *284* (5417), 1168–1170.

Pouton, C.W., Haynes, J.M. (2005). Pharmaceutical aspplications of embryonic stem cells. *Adv. Drug Deliv. Rev.*, *57* (13), 1918–1934.

Pouton, C.W., Haynes, J.M. (2007). Embryonic stem cells as a source of models for drug discovery. *Nat. Rev. Drug Discov.*, *6* (8), 605–616.

Prueksaritanont, T., Kuo, Y., Tang, C. *et al.* (2006). In vitro and in vivo CYP3A64 induction and inhibition studies in rhesus monkeys: a preclinical approach for CYP3A-mediated drug interaction studies. *Drug Metab. Dispos.*, *34* (9), 1546–1555.

Qatanani, M., Moore, DD. (2005). CAR, the continuously advancing receptor, in drug metabolism and disease. *Curr. Drug Metab.*, *6* (4), 329–339.

Raucy, J.L. (2003). Regulation of CYP3A4 expression in human hepatocytes by pharmaceuticals and natural products. *Drug Metab. Dispos.*, *31* (5), 533–539.

Raucy, J.L., Lasker, J., Ozaki, K., Zoleta, V. (2004). Regulation of CYP2E1 by ethanol and palmitic acid and CYP4A11 by clofibrate in primary cultures of human hepatocytes. *Toxicol. Sci.*, *79* (2), 233–241.

Raucy, J.L., Mueller, L., Duan, K., Allen, S.W., Strom, S., Lasker, J.M. (2002). Expression and induction of CYP2C P450 enzymes in primary cultures of human hepatocytes. *J. Pharmacol. Exp. Ther.*, *302* (2), 475–482.

Reinach, B., de Sousa, G., Dostert, P., Ings, R., Gugenheim, J., Rahmani, R. (1999). Comparative effects of rifabutin and rifampicin on cytochromes P450 and UDP-glucuronosyl-transferases expression in fresh and cryopreserved human hepatocytes. *Chem. Biol. Interact.*, *121* (1), 37–48.

Rencurel, F., Stenhouse, A., Hawley, S.A. *et al.* (2005). AMP-activated protein kinase mediates phenobarbital induction of CYP2B gene expression in hepatocytes and a newly derived human hepatoma cell line. *J. Biol. Chem.*, *280* (6), 4367–4373.

Ripp, S.L., Mills, J.B., Fahmi, O.A. *et al.* (2006). Use of immortalized human hepatocytes to predict the magnitude of clinical drug-drug interactions caused by CYP3A4 induction. *Drug Metab. Dispos.*, *34* (10), 1742–1748.

Rowland, M., Tozer, T.N. (1988). *Clinical Pharmacokinetics: Concepts and Applications*, Lea & Febiger, Philadelphia.

Roymans, D., Annaert, P., Van Houdt, J. *et al.* (2005). Expression and induction potential of cytochromes P450 in human cryopreserved hepatocytes. *Drug Metab. Dispos.*, *33* (7), 1004–1016.

Runge, D., Michalopoulos, G.K., Strom, S.C., Runge, D.M. (2000). Recent advances in human hepatocyte culture systems. *Biochem. Biophys. Res. Commun.*, *274* (1), 1–3.

Sahi, J., Milad, M.A., Zheng, X. *et al.* (2003). Avasimibe induces CYP3A4 and multiple drug resistance protein 1 gene expression through activation of the pregnane X receptor. *J. Pharmacol. Exp. Ther.*, *306* (3), 1027–1034.

Schehrer, L., Regan, J.D., Westendorf, J. (2000). UDS induction by an array of standard carcinogens in human and rodent hepatocytes: effect of cryopreservation. *Toxicology*, *147* (3), 177–191.

Schmiedlin-Ren, P., Thummel, K.E., Fisher, J.M., Paine, M.F., Lown, K.S., Watkins, P.B. (1997). Expression of enzymatically active CYP3A4 by Caco-2 cells grown on extracellular matrix-coated permeable supports in the presence of 1alpha,25-dihydroxyvitamin D3. *Mol. Pharmacol.*, *51* (5), 741–754.

Schmiedlin-Ren, P., Thummel, K.E., Fisher, J.M., Paine, M.F., Watkins, P.B. (2001). Induction of CYP3A4 by 1 alpha,25-dihydroxyvitamin D3 is human cell line-specific and is unlikely to involve pregnane X receptor. *Drug Metab. Dispos.*, *29* (11), 1446–1453.

Schuetz, E.G., Beck, W.T., Schuetz, J.D. (1996). Modulators and substrates of P-glycoprotein and cytochrome P4503A coordinately up-regulate these proteins in human colon carcinoma cells. *Mol. Pharmacol.*, *49* (2), 311–318.

Schwartz, R.E., Reyes, M., Koodie, L. *et al.* (2002). Multipotent adult progenitor cells from bone marrow differentiate into functional hepatocyte-like cells. *J. Clin. Invest.*, *109* (10), 1291–1302.

Shadle, C.R., Lee, Y., Majumdar, A.K. *et al.* (2004). Evaluation of potential inductive effects of aprepitant on cytochrome P450 3A4 and 2C9 activity. *J. Clin. Pharmacol.*, *44* (3), 215–223.

Shao, J., Stapleton, P.L., Lin, Y.S., Gallagher, E.P. (2007). Cytochrome p450 and glutathione S-transferase mRNA expression in human fetal liver hematopoietic stem cells. *Drug Metab. Dispos.*, *35* (1), 168–175.

Shih, H., Pickwell, G.V., Guenette, D.K., Bilir, B., Quattrochi, L.C. (1999). Species differences in hepatocyte induction of CYP1A1 and CYP1A2 by omeprazole. *Hum. Exp. Toxicol.*, *18* (2), 95–105.

Silva, J.M., Day, S.H., Nicoll-Griffith, D.A. (1999). Induction of cytochrome-P450 in cryopreserved rat and human hepatocytes. *Chem. Biol. Interact.*, *121* (1), 49–63.

Simonsson, U.S., Jansson, B., Hai, T.N., Huong, D.X., Tybring, G., Ashton, M. (2003). Artemisinin autoinduction is caused by involvement of cytochrome P450 2B6 but not 2C9. *Clin. Pharmacol. Ther.*, *74* (1), 32–43.

Sinz, M.W., Kim, S. (2006). Stem cells, immortalized cells and primary cells in ADMET assays. *Drug Discov. Today Technol.*, *3* (1), 79–85.

Sinz, M.W., Kim, S., Zhu, Z. *et al.* (2006). Evaluation of 170 xenobiotics as transactivators of human pregnane X receptor (hPXR) and correlation to known CYP3A4 drug interactions. *Curr. Drug Metab.*, *7* (4), 375–388.

Smith, C.M., Graham, R.A., Krol, W.L. *et al.* (2005). Differential UGT1A1 induction by chrysin in primary human hepatocytes and HepG2 Cells. *J. Pharmacol. Exp. Ther.*, *315* (3), 1256–1264.

Soars, M.G., McGinnity, D.F., Grime, K., Riley, R.J. (2007). The pivotal role of hepatocytes in drug discovery. *Chem. Biol. Interact.*, *168* (1), 2–15.

Stanley, L.A., Horsburgh, B.C., Ross, J., Scheer, N., Wolf, CR. (2006). PXR and CAR nuclear receptors which play a pivotal role in drug disposition and chemical toxicity. *Drug Metab. Rev.*, *38* (3), 515–597.

Synold, T.W., Dussault, I., Forman, B.M. (2001). The orphan nuclear receptor SXR coordinately regulates drug metabolism and efflux. *Nat. Med.*, *7* (5), 584–590.

Tateno, C., Yoshizane, Y., Saito, N. *et al.* (2004). Near completely humanized liver in mice shows human-type metabolic responses to drugs. *Am. J. Pathol.*, *165* (3), 901–912.

Tirona, R.G., Kim, R.B. (2005). Nuclear receptors and drug disposition gene regulation. *J. Pharm. Sci.*, *94* (6), 1169–1186.

Trubetskoy, O., Marks, B., Zielinski, T., Yueh, M.F., Raucy, J. (2005). A simultaneous assessment of CYP3A4 metabolism and induction in the DPX-2 cell line. *Aaps. J.*, *7* (1), E6–13.

Tsutsui, M., Ogawa, S., Inada, Y. *et al.* (2006). Characterization of cytochrome P450 expression in murine embryonic stem cell-derived hepatic tissue system. *Drug Metab. Dispos.*, *34* (4), 696–701.

Tucker, G.T., Houston, J.B., Huang, S.M. (2001). Optimizing drug development: strategies to assess drug metabolism/transporter interaction potential—toward a consensus. *Pharm. Res.*, *18* (8), 1071–1080.

Vermeir, M., Annaert, P., Mamidi, R.N., Roymans, D., Meuldermans, W., Mannens, G. (2005). Cell-based models to study hepatic drug metabolism and enzyme induction in humans. *Expert Opin. Drug Metab. Toxicol.*, *1* (1), 75–90.

Vignati, L.A., Bogni, A., Grossi, P., Monshouwer, M. (2004). A human and mouse pregnane X receptor reporter gene assay in combination with cytotoxicity measurements as a tool to evaluate species-specific CYP3A induction. *Toxicology*, *199* (1), 23–33.

Wada, K., Takada, M., Ueda, T. *et al.* (2007). Drug interactions between tacrolimus and phenytoin in Japanese heart transplant recipients: 2 case reports. *Int. J. Clin. Pharmacol. Ther.*, *45* (9), 524–528.

Walle, T., Otake, Y., Galijatovic, A., Ritter, J.K., Walle, U.K. (2000). Induction of UDP-glucuronosyltransferase UGT1A1 by the flavonoid chrysin in the human hepatoma cell line hep G2. *Drug Metab. Dispos.*, *28* (9), 1077–1082.

Wang, H., Faucette, S., Moore, R., Sueyoshi, T., Negishi, M., LeCluyse, E. (2004). Human constitutive androstane receptor mediates induction of CYP2B6 gene expression by phenytoin. *J. Biol. Chem.*, *279* (28), 29295–29301.

Wang, H., Huang, H., Li, H. *et al.* (2007). Activated pregnenolone X-receptor is a target for ketoconazole and its analogs. *Clin. Cancer Res.*, *13* (8), 2488–2495.

Wang, H., LeCluyse, E.L. (2003). Role of orphan nuclear receptors in the regulation of drug-metabolising enzymes. *Clin. Pharmacokinet.*, *42* (15), 1331–1357.

Watanabe, T., Shibata, N., Westerman, K.A. *et al.* (2003). Establishment of immortalized human hepatic stellate scavenger cells to develop bioartificial livers. *Transplantation*, *75* (11), 1873–1880.

Werner, A., Duvar, S., Muthing, J. *et al.* (1999). Cultivation and characterization of a new immortalized human hepatocyte cell line, HepZ, for use in an artificial liver support system. *Ann. N. Y. Acad. Sci.*, *875*, 364–368.

Westerink, W.M., Schoonen, W.G. (2007a). Phase II enzyme levels in HepG2 cells and cryopreserved primary human hepatocytes and their induction in HepG2 cells. *Toxicol. In Vitro*, *21* (8), 1592–1602.

Westerink, W.M., Schoonen, W.G. (2007b). Cytochrome P450 enzyme levels in HepG2 cells and cryopreserved primary human hepatocytes and their induction in HepG2 cells. *Toxicol. In Vitro*, *21* (8), 1581–1591.

Wilkening, S., Stahl, F., Bader, A. (2003). Comparison of primary human hepatocytes and hepatoma cell line Hepg2 with regard to their biotransformation properties. *Drug Metab. Dispos.*, *31* (8), 1035–1042.

Worboys, P.D., Carlile, D.J. (2001). Implications and consequences of enzyme induction on preclinical and clinical drug development. *Xenobiotica*, *31* (8–9), 539–556.

Youdim, K.A., Tyman, C.A., Jones, B.C., Hyland, R. (2007). Induction of cytochrome P450: assessment in an immortalized human hepatocyte cell line (Fa2N4) using a novel higher throughput cocktail assay. *Drug Metab. Dispos.*, *35* (2), 275–282.

Zhang, H., Davis, C.D., Sinz, M.W., Rodrigues, A.D. (2007). Cytochrome P450 reaction-phenotyping: an industrial perspective. *Expert Opin. Drug Metab. Toxicol.*, *3* (5), 667–687.

Zhou, C., Tabb, M.M., Sadatrafiei, A., Grun, F., Blumberg, B. (2004). Tocotrienols activate the steroid and xenobiotic receptor, SXR, and selectively regulate expression of its target genes. *Drug Metab. Dispos.*, *32* (10), 1075–1082.

Zhu, W., Song, L., Zhang, H., Matoney, L., LeCluyse, E., Yan, B. (2000). Dexamethasone differentially regulates expression of carboxylesterase genes in humans and rats. *Drug Metab. Dispos.*, *28* (2), 186–191.

Zhu, Z., Kim, S., Chen, T. *et al.* (2004). Correlation of high-throughput pregnane X receptor (PXR) transactivation and binding assays. *J. Biomol. Screen*, *9* (6), 533–540.

Zhu, Z., Puglisi, J., Connors, D. *et al.* (2007). Use of cryopreserved transiently transfected cells in high-throughput pregnane X receptor transactivation assay. *J. Biomol. Screen*, *12* (2), 248–254.

# An Introduction to Metabolic Reaction Phenotyping

CARL D. DAVIS and A. DAVID RODRIGUES

## 1 INTRODUCTION

The role of drug-metabolizing enzymes (DMEs) and metabolic clearance in the overall determination of a drug's oral bioavailability and pharmacokinetics (PK) is well established and fully incorporated into the process of discovery, design, and development of new and better therapeutic agents. Given the very many different types, classes, and members of enzyme families known to metabolize drugs and other xenobiotics, the task of optimizing metabolic stability can appear somewhat daunting. Fortunately, certain enzymes feature more predominantly than others, and drug optimization often begins and ends with them.

In this chapter we will describe a few of the major enzyme families, the experimental and kinetic methods used to identify their role in the overall metabolic clearance of a drug (so-called "reaction phenotyping" or "enzyme mapping") and illustrate some concerns related to the contribution and interindividual variation of some notable enzymes. An attempt has been made to provide a useful and comprehensive outline of what is requisite for drug discovery, but an introductory text can only go so far. So for greater depth of analysis and discussion, the reader is referred to the many reviews published on bioanalysis, enzyme kinetics, substrates and biotransformation, drug–drug interactions (DDIs), adverse drug reactions, and *in silico* methods and pharmacogenetics, a few of which are listed at the end of this chapter.

## 2 DRUG METABOLISM

The membrane-bound hemoproteins that belong to the superfamily of cytochrome P450s (CYPs), by virtue of expression, number, diversity, and substrate promiscuity, dominate most other enzymes in their contribution to drug clearance. Members of three CYP subfamilies in particular (CYP1, CYP2, and CYP3) are responsible for the

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

metabolic clearance of most drugs and xenobiotics (Rendic and DiCarlo, 1997; Guengerich, 2003; Williams *et al.*, 2004). The flavin-containing monooxygenases (FMOs), which tend to mediate *N*- and *S*-oxidation reactions, and the uridine 5′-diphosphate glucuronosyltransferases (UGTs) that catalyze drug conjugation with uridine 5′-diphosphate glucuronic acid (UDPGA), are two other important classes of enzymes that are also found co-localized with CYPs in the endoplasmic reticulum. In addition to these membrane-bound forms, there are significant families of cytosolic DMEs, such as *N*-acetyl transferases (NATs), sulfotransferases (SULTs), glutathione transferases (GSTs), alcohol dehydrogenases (ADHs), and aldehyde dehydrogenases (ALDHs). Almost all biotransformation and metabolic clearance pathways encountered in drug metabolism involve one or more of these DMEs—most of them involve CYPs, UGTs, and/or FMOs—and although extrahepatic metabolism can be very important, especially on a case-by-case basis, the highest levels of many important enzymes are found in the liver and this organ dominates the metabolic clearance of most drugs.

DMEs have evolved in part to regulate some endogenous reactions (including steroid synthesis and bilirubin clearance) and also, if not primarily, to protect living organisms from exposure to environmental toxins (Guengerich, 1989). Orally ingested xenobiotics such as drugs, environmental chemicals, toxins, and so on, follow a common fate. If soluble and permeable, the drug in question is absorbed from the intestine and may undergo metabolism in the enterocytes forming the microvilli of the duodenum and proximal jejunum. The fraction that is absorbed and not metabolized in the gut subsequently passes into the hepatic portal vein where it enters the highly perfused liver, where, in many cases, it undergoes metabolism by DMEs such as CYPs. This combination of absorption and metabolism (gut and liver) is known as the "first-pass effect" and determines the oral bioavailability of a drug (i.e. the amount of the dose that reaches the systemic circulation and is available for effect on target tissues).

In general, the function of metabolic clearance is to convert a lipophilic molecule into a more polar and more aqueous soluble metabolite that can then be eliminated in the bile and/or urine. The liver possesses the highest levels of the major and most commonly involved DMEs, and subject to dose and saturation kinetics, it is usually the single most important organ of metabolic clearance. The liver thus serves, on one hand, as the *de facto* organ of detoxification, thus protecting us from an accidental or deliberate exposure to poisons and toxins; on the other hand, the liver acts as an effective barrier to achieving efficacious exposure from the oral administration of reasonable dosages and practicable dose regimens. Balancing these properties often determines the safety margins of a drug—achieving exposure for efficacy while not causing significant adverse reactions.

Most other organs and tissues in humans and other mammals are known to express DMEs that can play a significant or even a major role in metabolic clearance. For instance, the lungs receive 100% of the cardiac output, and the nasal and pulmonary epithelia express DMEs that can serve to protect against systemic exposure as well as inhaled toxins, including solvents and volatile agents such as naphthalene (Buckpitt *et al.*, 2002; Castell, Donato and Gómez-Lechón, 2005). Keratinocytes, epithelial cells, and the major cell type in the epidermis, are also known to express DMEs that can protect against permeable agents absorbed via skin contact (Swanson, 2004). The liver and extrahepatic DMEs can lead to the formation of more reactive and/or more toxic metabolites that can cause toxicity locally or *ex situ*, which can become a concern in drug development and ultimately determine the manifestation and prevalence of any ensuing toxicity in humans. For a more detailed discussion on this aspect of drug

metabolism, the reader is referred to Chapter 21, which deals with the specific roles and complexities involved in detoxification and bioactivation reactions that can ultimately define the safety profile of a drug.

The expression, localization, and substrate selectivity of the DMEs are key factors essential to understanding and predicting tissue-specific effects (be they benign or malignant) and individual susceptibility to adverse drug reactions. The focus of this chapter will deal specifically with how one (i) measures the rate of metabolism, (ii) identifies the enzymes involved, (iii) extrapolates from *in vitro* to *in vivo* clearance, and (iv) considers metabolic clearance when predicting the potential for DDIs that affect a drug. In addition, examples are provided in order to illustrate why reaction phenotyping is important in the risk evaluation and safety profile of a new drug. In drug discovery, these studies usually provide *a priori* predictions of human metabolism without any human clinical data for the potential drug candidate. Therefore, it is often necessary to draw parallels between results obtained with the test drug and those from drugs in the clinic possessing similar characteristics.

## 3 OBJECTIVES OF REACTION PHENOTYPING IN DRUG DISCOVERY

Rarely, if ever, does a compound offer a perfect profile of efficacy, safety, and ease of development. Therefore, the primary goal in drug discovery is to find the compound that offers the best balance of properties with an acceptable risk–benefit ratio and thus minimize, rather than negate, the chance of failure in development. Before advancing a drug into clinical trials, it is desirable to know (i) the intrinsic clearance of a drug and from that the predicted *in vivo* clearance in humans and (ii) the enzymes involved in its metabolism (reaction phenotype). These are useful since knowing its *in vitro* clearance in human and animal assays helps one to use *in vivo* animal data to help predict a drug's human pharmacokinetics (including half-life and oral bioavailability). At the same time, knowing the enzyme(s) involved in the biotransformation reaction(s) allows one to predict the consequences of individual variability in metabolism and the risks associated with coadministration of drugs that alter the function of such enzyme(s).

These days mention is often made of a drug acting as a "perpetrator" of DDIs, with the drug itself causing offense by acting as an inhibitor or inducer. In reaction phenotyping, arguably one is tasked with the identification of the risks associated with the drug as a potential "victim" of DDIs. As will be discussed, the higher the metabolic clearance, and the fewer the number of enzymes involved in that clearance, the greater is the chance of a significant and potentially dangerous interaction occurring. Such an interaction manifests itself as an increase in the plasma $C_{max}$ (maximal plasma concentration) or AUC (area under the plasma concentration–time curve) of the victim drug with a twofold or greater increase usually considered significant.

## 4 METABOLIC STABILITY SCREENING

The simplest way to minimize the risks associated with drug metabolism is to find a drug that is not cleared by metabolism (assuming that a prolonged half-life is not inherently a concern). Today's pharmaceutical R&D efforts incorporate metabolic screening

in the drug discovery process, with the goal of finding and developing drugs without significant metabolism-mediated clearance (Parkinson, 1996; Rodrigues, 1999; Venkatakrishnan, Von Moltke and Greenblatt, 2001; Bjornsson *et al.*, 2003; Williams *et al.*, 2003a; Zhang *et al.*, 2007). This does not guarantee good or even sufficient oral bioavailability—solubility and permeability are two other, often critical, factors that must be considered—however, and all else being equal, a drug with greater metabolic stability will have higher oral exposure and lower interindividual variability. The key advantage to screening metabolic stability *in vitro* is that it allows one to evaluate many more compounds than is possible by *in vivo* screening in preclinical animal models. Hepatic DMEs most frequently involved in drug metabolism, *in vitro* stability, and clearance assays and screens are designed accordingly. For example, liver microsomes (prepared from liver tissue) and hepatocytes (fresh and/or cryopreserved) are most often used to measure and compare metabolic stability. With liver and hepatocytes available from human donors and preclinical animal models, *in vitro* assays in conjunction with allometric-scaling methods can help to bridge *in vitro* and *in vivo* studies to support predictions of PK and metabolism in humans, as well as identify biotransformation reactions that appear specific to humans and/or animals.

Liver microsomes provide a rich source of CYPs, FMOs, and UGTs, are available from human and various animal species, relatively cheap to prepare, simple to use, and can be stored frozen for relatively long periods. Fresh hepatocytes, in contrast, require more handling than liver microsomes yet offer the significant advantage of possessing a full complement of microsomal, cytosolic, and mitochondrial hepatic DMEs. Progress has been made to improve cryopreservation and thawing methods, which has increased the ease of handling of hepatocytes (Gómez-Lechón *et al.*, 2006). Thawing can result in the loss of metabolic activity; therefore, cryopreserved cells tend not to be used to make a definitive prediction of *in vivo* clearance with freshly isolated hepatocytes used instead. Cryopreserved cells are useful, however, as a preliminary screen of metabolic stability and/or as a bioreactor to generate metabolites for identification. Hepatocytes also allow one to look directly at hepatotoxicity associated with the drug and/or its metabolite(s) and, potentially, evaluate directly the role of a specific CYP in that toxicity (Mennes *et al.*, 1994; Li *et al.*, 1999; Hewitt *et al.*, 2002).

Liver microsomes and hepatocytes require different buffers for optimal activity of DMEs, and there are a few other differences; however, in general the *in vitro* methodologies for predicting metabolic clearance in these systems are similar. Often, the rate of loss of test compound over a single predetermined time period is used to rank order metabolic stability. This is very useful in identifying the worst performing compounds, yet it does lack precision and accuracy in predicting properties such as the *in vivo* clearance of a drug. Other methods are needed to scale properties from *in vitro* to *in vivo* and are employed as a compound advances through the screening process.

Two matters must be considered when assessing results from these kinds of screens: (i) the rate of metabolism in liver microsomes does not *per se* indicate which enzyme or enzymes are involved in the reaction; and (ii) measuring the rate of loss of compound does not specify what metabolite or metabolites are formed.

## 5 DRUG-METABOLIZING ENZYMES

The role of enzymes in drug metabolism has long been known and the last 40 years in particular have produced a wealth of data describing and elaborating on many of these

enzymes and provided a mechanistic insight into the pharmacokinetics observed in animals and patients. There are very many DMEs expressed in humans and animals with considerable variety in their substrate affinity and expression between tissues and between species. Fortunately, a few enzymes are encountered more frequently than others and we will focus on three such enzyme families: the CYPs, the FMOs, and the UGTs. Understanding how these work, and knowing how to measure their activity and predict their contribution to metabolic clearance, will be relevant to most small molecule drugs.

### 5.1 Cytochrome P450s (CYPs)

The CYPs [E.C. 1.14.14.1] arguably are the single most important family of enzymes involved in regulating exposure to pharmacologically active, toxic, or potentially toxic xenobiotics. These enzymes possess a common ancestor dating back over a billion years and exhibit highly conserved regions in their amino acid sequence across different species including humans. The CYPs are membrane-bound hemoproteins that are found in the smooth endoplasmic reticulum and can be prepared as microsomal fractions for *in vitro* use. CYPs are known to have a structurally diverse array of substrates that usually undergo metabolism via oxidation. The CYP-catalyzed monooxygenase reaction is the most commonly encountered and can be described simply as:

$$RH + O_2 + 2H^+ + 2e^{-1} \rightarrow ROH + H_2O$$

Substrate oxidation proceeds via redox cycling between the ferric to ferrous states of the heme protoporphyrin group in the active site of the enzyme, with the NAD(P)H oxidoreductase and cytochrome b5 accessory flavoproteins involved in the electron transfer from nicotinamide adenine dinucleotide phosphate (NADPH) to the heme and the hemo-substrate complex, respectively (Fig. 1). Ultimately, this redox process catalyzes the incorporation of oxygen from molecular oxygen producing water and a metabolite that is usually more polar, more water-soluble, and more easily eliminated in the bile or urine.

In most cases, CYP-mediated metabolic clearance is an elimination process and thus serves as a potential mechanism of detoxification via removal of a potentially hazardous xenobiotic. There are instances, however, in which the transition intermediates formed in the CYP catalytic cycle can prove to be highly reactive. If they react with the heme (diltiazem, tamoxifen or troleandomycin, and CYP3A4) or apoprotein (dihydralazine and CYP1A2; tamoxifen and CYP2B6; tienilic acid and CYP2C9), mechanism-based inactivation can occur that results in irreversible inactivation of the CYP (covalent binding to the heme will completely inhibit the CYP reaction, whereas alkylation of the amino acid can cause a loss of activity but not necessarily complete inactivation).

Mechanism-based inactivation can exacerbate and prolong the risk of DDIs and toxicity from coadministered drugs with effects overcome only once a new enzyme has been synthesized to replace that inactivated. Or if the reactive species is sufficiently stable to escape the active site and covalently bind with nucleophilic macromolecules such as proteins, RNA, or DNA, it may culminate directly in adverse drug reactions and toxicity. (For a comprehensive review of mechanisms and methods related to mechanism-based inactivation, see Venkatakrishnan and Obach, 2007 and references contained therein.)

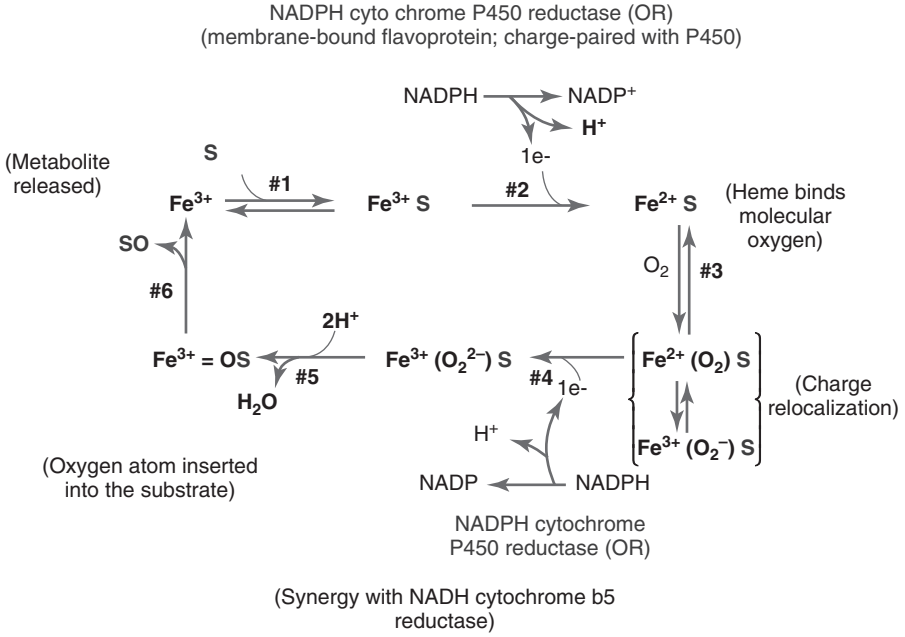


**Figure 1** Cytochrome P450 (CYP)-catalyzed oxidation and redox cycling. Binding of the substrate to the low-spin ferric heme iron of the P450 (step #1) initiates the transfer of an electron from the cytochrome P450 oxidoreductase (CYPOR) accessory protein to form a substrate complex with a ferrous heme iron (step #2). This can bind with molecular oxygen (step #3), followed by charge relocalization presumed to form a more stable ferric complex that undergoes a second reduction step with CYPOR, with or without involvement from the cytochrome b5 accessory protein (step #4), to form water and an $(Fe\text{-}O)^{3+}$ complex or $O_2^-$ superoxide anion, with subsequent incorporation of an oxygen atom into the substrate (step #5). In the last step, the oxidized product (metabolite) is released returning the active site to its original low-spin ferric heme (step #6).

CYPs are divided into subfamilies based on their shared sequence in protein homology, with >40% sequence similarity designating the family and >55% shared sequence homology designating the subfamily (the family is represented by an Arabic numeral, the subfamily by a capital letter, and then another Arabic numeral designates the individual gene member, e.g. CYP3A4). Over 50 different CYPs have been identified in humans, with CYP1A2, 2A6, 2B6, 2C8, 2C9, 2C19, 2D6, 2E1, 2J2, 3A4, 3A5, 4F2, 4F3a, 4F3b, and 4F12 responsible for the metabolic clearance of most drugs. In fact, historically CYP1A2, 2C8, 2C9, 2C19, 2D6, and 3A4 are the most important CYPs in this regard and are the focus of most reaction-phenotyping efforts, at least in the early stages. (The other enzymes are often assessed via a process of elimination.)

The same nomenclature is used for CYPs from human and from other species. Several CYPs are well conserved across species and share sufficient identity that they have the same name (CYP1A1, 1A2, 1B1, and 2E1); however, although these CYPs are generally similar in sequence and function, it is important to note that they are not identical to the human forms. Even a small difference in amino acid sequence can affect substrate affinity and the ability of the enzyme to metabolize a substrate and/or the

regiospecificity in the site of the molecule metabolized. (For a complete and updated repository of the classified CYPs, see Home Page of the Human Cytochrome P450 (CYP) Allele Nomenclature Committee at http://www.cypalleles.ki.se/.)

As a family, the CYPs unite to catalyze the oxidation and metabolic clearance of a highly diverse collection of lipophilic substrates that are otherwise not readily eliminated in urine or bile. This lack of specialization distinguishes the CYPs from more efficient enzymes like carbonic anhydrase and provides a safeguard against the tremendous diversity and risk present in nature's chemistry. Individual CYPs, however, tend to be more selective in their substrate characteristics and can be distinguished somewhat by their preferred substrate properties, as shown in Table 1. CYP1A1 is often involved in the metabolism of aryl hydrocarbons and although not constitutively expressed in liver tissue, its mRNA is highly inducible with translation to the protein also reported (Drahushuk *et al.*, 1998). CYP1A1 and CYP1A2 show an overlap in their substrate affinity; however, CYP1A2 preferentially metabolizes polycyclic aromatic hydrocarbons and their amine analogues. Both CYP1A1 and 1A2 are involved in the metabolism of procarcinogens to carcinogens; however, CYP1A2 has been shown to be the primary enzyme involved in the metabolism of drugs such as phenacetin and theophylline. CYP2A6 shows a preference for smaller, neutral, less lipophilic molecules such as ketones or nitrosamines. CYP2C9 and CYP2C19 are able to metabolize acidic molecules, with CYP2C9 preferring acidic drugs that have a reactive site six to eight carbon lengths removed from an H-bonding residue (e.g. diclofenac). In contrast CYP2D6, constitutive in liver, is better able to metabolize basic drugs, usually with the reactive site six to eight carbons removed from a basic nitrogen atom (e.g. fluoxetine, nortriptyline, dextromethorphan, and bufuralol). CYP2E1 is highly inducible and often effective in metabolizing small molecules (e.g. ethanol, acetaminophen, halothane, and para-nitrophenol), and CYP3A4 has an active site big enough to accommodate multiple substrates including relatively large and/or very lipophilic compounds (e.g.

**TABLE 1 Human cytochrome P450s and their substrate selectivity.**

| CYP | Substrate | Example substrates |
|---|---|---|
| CYP1A1, 1A2 | Planar lipophilic; neutral or basic molecules with log P 1.39–6.35 (CYP1A1); log P 0.08–3.61 (CYP1A2) | Phenacetin, dapsone, tacrine, ramelteon |
| CYP2C9 | Medium-sized acidic molecules with 1–2 hydrogen bond acceptors; log P 0.89–5.18 | Diclofenac, warfarin, tolbutamide |
| CYP2C19 | Medium-sized basic molecules with 2–3 hydrogen bond donor/acceptor atoms; log P 1.49–4.42 | (*S*)-mephenytoin, omeprazole |
| CYP2D6 | Medium-sized basic molecules with a protonatable nitrogen 5–7 Å from the site of metabolism; log P 0.75–5.04 | Fluoxetine, dextromethorphan, nortriptyline, bufuralol |
| CYP2E1 | Small, neutral molecules (structurally diverse); log P –1.35–3.63 | Para-nitrophenol, ethanol, chlorzoxazone |
| CYP3A4 | Large, lipophilic molecules (structurally diverse); log P 0.97–7.54 | Testosterone, cyclosporin A, nifedipine, midazolam, simvastatin |

Adapted from Lewis and Dickins (2002) and Lewis, Jacobs and Dickins (2004).

simvastatin, testosterone, cyclosporine A, and midazolam), a combination of properties that has led to its involvement in a majority of CYP-mediated drug reactions and significant consequences that result from DDIs (Guengerich *et al.*, 2002; Lewis and Dickins, 2002).

Substrate and protein conformation, intra- and intermolecular H-bonding, and electronic and steric constraints all influence the propensity for metabolism and make it difficult to model or predict *in silico* how a substrate will behave *a priori*. Despite this complexity, advances have been made with *in silico* models, especially with regard to modeling CYP inhibition governed by binding properties and thus simpler than predicting the site or rate of metabolism (Stouch *et al.*, 2003; O'Brien and de Groot, 2005). Nonetheless, when beginning reaction-phenotyping studies, it is always worth considering the structure and physicochemical properties of the molecule and how this relates to the preferences known for the CYPs described.

A fundamental concern that drives the need to perform reaction-phenotyping studies is that if a new drug is metabolized to a significant degree (>40% of metabolic clearance), then it is at risk of displaying markedly different exposure when coadministered with an inhibitor of that same enzyme. For instance, terfenadine (the original prodrug of fexofenadine or Allegra) is a classic example of a high-metabolic-clearance drug in humans. Terfenadine was shown to undergo extensive first-pass metabolism, with its clearance mediated predominantly by CYP3A4 (Garteiz *et al.*, 1982; Yun, Okerholm and Guengerich, 1993). Unfortunately, terfenadine can also bind to the hERG (human *ether-a-go-go*-related gene) potassium channel receptor and cause QT prolongation and cardiovascular toxicity including potentially fatal arrhythmias (Woosley *et al.*, 1993). When terfenadine was dosed with CYP3A4 inhibitors such as ketoconazole, itraconazole, and grapefruit juice, there was an increase in its exposure and that of its metabolites. This was associated with an increase in prolongation of QT effects, linked to fatal cardiac arrhythmias in patients (Honig *et al.*, 1993a, b, 1996). As a consequence of such poor clinical control, terfenadine was soon withdrawn from the market and now serves as a worst-case example of what can go wrong insofar as a narrow therapeutic index and too specific (single CYP) reaction phenotype is concerned.

***5.1.1 Human CYP Polymorphisms and Individual Variability*** Humans show significant interindividual variability in their total and relative expression of DMEs such as CYPs, UGTs, FMOs, SULTs, GSTs, and NATs. This variability extends to the relative exposure seen with drugs and their metabolites in the patient population. By knowing what enzymes are involved in the metabolism of a drug, it is possible to make a prediction of the likely variability in its exposure.

Variability in DMEs can involve several factors, some of which are easier to characterize than others and some are compound- or drug-specific. One source of interindividual variability in humans stems from enzyme induction that can be elicited by environmental influences, such as pollutants or lifestyle and dietary factors such as cigarette smoking and eating cooked meat, which are both known to induce CYP1A1, 1A2, among others; the popular herbal supplements like St. John's wort (*Hypericum perforatum* L.) are also known to induce CYP3A (Markowitz, 2003). Variability in drug metabolism and exposure can also stem from powerful CYP inhibitors found in natural products, for example, furanocoumarins related to and including bergamottin in grapefruit juice (Girennavar *et al.* 2006).

Another important contributor to individual variation in CYP activity is genetics, with heterozygous and homozygous mutants in CYP expression and/or function often

resulting in polymorphisms. A polymorphism is defined as an allelic variant that is expressed with ≥1% frequency in the human population. CYP polymorphic phenotypes usually describe bimodal or multimodal distribution in the pharmacokinetics of a drug given at an equivalent dose to healthy humans. Based on their phenotype, patients can be classified as poor metabolizers (PMs), intermediate metabolizers (IMs), extensive metabolizers (EMs), and in the instance of CYP2D6, ultrarapid metabolizers (UMs). At the genetic level, there are numerous allelic variants leading to various mutations that can affect the ability of that individual to metabolize a particular substrate or class of substrates. This is evident with ethnic differences seen in the frequency and expression of CYP polymorphisms. For instance, CYP2D6 slow metabolizers comprise ~7% of Caucasian and only ~0.8% of Asian populations, whereas CYP2C19 slow metabolizers comprise ~2% of Caucasians and ~19% of Asians. This is enough to merit a distinctive shift in focus when evaluating the safety and development of CYP2D6 or CYP2C19 drugs in these different patient populations and markets. CYP3A4 is highly variable within and between humans yet is generally not considered to be polymorphic in terms of its phenotype (i.e. it is present and functionally active in most people). In contrast, its closest family member, CYP3A5, is found in only 25% of Caucasians and only 75% of Africans and people of African descent (Kuehl *et al.* 2001; Mouly, 2005). An overlapping substrate affinity of CYP3A4 and CYP3A5 likely contributes significantly to the interindividual variability in drug exposure. Since CYP3A5 is known to be expressed in the intestine and other extrahepatic tissues, its contribution to metabolic clearance could be underestimated if only human liver microsomal systems are used to predict metabolic clearance.

In Table 2 are listed the frequencies of a few known polymorphisms in Caucasians versus Asians reported for CYP2D6, CYP2C19, and also CYP2A6, 2C9, and 3A4. It should be noted that these represent only a very few number of the genetic polymorphisms already identified for DMEs. Numerous mutant forms have been identified that have greater, lesser, or even no discernible effect on CYP activity, including the possibility of multiple gene copies that can lead to markedly higher CYP protein expression. One example of which is the ultrarapid CYP2D6 metabolizers found at relatively high frequencies in some African populations (~29%) and Southern Europeans (~10%). In contrast to null polymorphisms and slow metabolizer phenotype, a CYP2D6 substrate will show much higher clearance and thus much lower exposure of drug in patients possessing multiple copies of the CYP2D6 gene and potentially much higher exposure to metabolite(s) formed from that drug, for example, the antidepressant nortriptylline (Dalén *et al.*, 1998).

Although there exist some very useful technologies that facilitate genotyping of patients (e.g. the Roche Diagnostics P450 AmpliChip), there can be considerable complexity involved in translating a particular genotype to a predicted phenotype (Jain, 2005; Li *et al.*, 2006). Advances are expected that ultimately could allow individualized dose regimens to become economically feasible. Looking to the future, this will provide more therapeutic opportunities for R&D with both new drugs and also the opportunity to revisit the use of older drugs that were withdrawn due to a poor safety profile in the average population. Meanwhile, the reaction phenotype of a drug when coupled with its predicted or known safety margins does allow one to define the risk involved in administering to various populations. For instance, a new drug eliminated primarily by a polymorphic metabolic clearance with a narrow safety margin is unlikely ever to be developed beyond testing in humans under strictly controlled clinical conditions.

**TABLE 2 Common polymorphisms of human cytochrome P450s (CYP) in Asian and Caucasian populations.**

| CYP | | Poor metabolizer frequency (%) | Major variant | Allele frequency (%) | Mutation effect | References |
|---|---|---|---|---|---|---|
| CYP2A6 | Asian | 2–4 | 2A6*4 | 11–18 | Gene deletion; nonenzyme | Ariyoshi *et al.* (2000) |
| | Caucasian | ~0 | 2A6*2 | 4 | Gene deletion | Kwon *et al.* (2001); Oscarson *et al.* (1999a, b) |
| CYP2C9 | Asian | 0.04 | 2C9*3 | 2.1 | Stop codon; inactive | Takanashi *et al.* (2000) |
| | Caucasian | 1 | 2C9*2 | 8–13 | Reduced affinity | Aithal *et al.* (2000) |
| | | | 2C9*3 | 7–9 | Altered substrate specificity | |
| CYP2C19 | Asian | 19 | 2C19*2 | 32 | Splicing defect | Kaneko *et al.* (1999a, b) |
| | | | 2C19*3 | 6–10 | Stop codon; inactive | |
| | Caucasian | 2 | 2C19*2 | 13 | Inactive | Griese *et al.* (2001) |
| CYP2D6 | Asian | 0.8 | 2D6*5 | 6.2 | Gene deletion; no enzyme | Hiroi, Imaoka and Funae (1998) |
| | | | 2D6*10 | 33–50 | Unstable (intermediate metabolizer) | Ramamoorthy, Tyndale and Sellers (2001); Shimada *et al.* (2001) |
| | Caucasian | 7 | 2D6*4 | 12–21 | Inactive | Sachse *et al.* (1997) |
| CYP3A4 | Asian | 1 | 3A4*6 | 1 | Frameshift | Hsieh *et al.* (2001) |
| | Caucasian | ~0 | 0 | 0 | — | Ozdemir *et al.* (2000) |

Adapted from Mizutani, 2003, with references therein.

### 5.2 Flavin-Containing Monooxygenases

FMOs [E.C. 1.14.13.8] are found in species from bacteria to mammals. They are identified by several classes: external monooxygenases are the most common form and depend on NADPH or nicotinamide adenine dinucleotide (NADH) to reduce the flavin (flavin mononucleotide, FMN, or flavin adenine dinucleotide, FAD) to the reactive form that can bind molecular oxygen; in contrast, the rarely encountered internal monooxygenases (e.g. lactate monooxygenase) can use the substrate itself to reduce the flavin. In humans, genes for ~57 CYPs have been reported (plus approximately five pseudogenes), whereas to date, only five functional human FMOs are known (FMO1-5) with six pseudogenes. Of these enzymes, FMO1 has the broadest substrate specificity; FMO2 is the narrowest because of its restricted active site; FMO3 has the highest functional activity and also a high expression; FMO4 shows the lowest expression—with its low activity this is considered insignificant with respect to drug metabolism (with possible exceptions); and FMO5 has significant expression, albeit with relatively low activity and very narrow substrate selectivity.

FMOs show distinctive and potentially characteristic differences in their relative expression in human tissues: FMO1 is the major FMO expressed in fetal liver and adult human kidney and intestine; FMO2 is the major FMO expressed in human lung, albeit inactive in Caucasians, Asians, and the majority of African-Americans (attributed to the FMO2*2 allele associated with expression of a nonfunctional truncated protein encoded by a nonsense mutation g.23238C > T (Q472X)); and FMO3 is regarded generally as the FMO most highly expressed in the liver (Dolphin *et al.*, 1998; Cashman 2000; Whetstine, 2000; Krueger and Williams 2005). Some studies have, however, shown that FMO5 can be expressed at higher levels in human liver (3.5 to 34 pmol/mg microsomal protein for FMO5 versus 12.5 to 117 pmol/mg microsomal protein for FMO3) (Overby *et al.*, 1997). This could reflect how well the tissues were handled in the preparation of liver microsomes (Cashman and Zhang 2006). In general, FMO3 and FMO5 are the forms most likely to be involved in FMO-mediated drug metabolism.

With liver FMO3 measured at immunoquantified levels on average ~60% of those observed for CYP3A4 (Overby, Carver and Philpot, 1997) and considering its relatively broad substrate selectivity, FMO3 has become the most significant form involved in FMO-mediated drug metabolism. When microsome preparations are properly handled, FMO activity can demonstrate a $k_{cat}$ at 30–60 per minute, which is considerably greater than that typically seen with CYPs (e.g. $k_{cat}$ of 1–20 per minute) (Krueger and Williams, 2005). It is thus conceivable that even with CYP3A4 substrates, FMO3 may constitute the primary metabolic clearance pathway and potentially mitigate CYP3A4 DDIs. With FMOs and CYPs are both NADPH-dependent enzymes, and thus both active in the liver microsomal assays used typically for metabolic stability studies. The two families of enzymes can even compete to form the same metabolite(s) as reaction products (albeit with FMOs being more specific than CYPs and tending to favor *S*- and *N*-oxidation reactions). This is a consideration when performing reaction-phenotyping studies and evaluating CYP versus FMO reactions.

One important characteristic of FMOs is that they are thermally unstable when compared with CYPs, which is useful for reaction-phenotyping studies. Conversely, however, if liver microsomes are not properly handled, FMO activity can be lost *in vitro*, leading to a prediction that underestimates its contribution to the total metabolic clearance.

***5.2.1 Human FMO Polymorphisms and Individual Variability*** No inducers have been found for FMOs and DDIs appear limited in effect since few drugs are metabolized

primarily by FMOs (itopride is an example of a drug metabolized primarily by FMOs and not by CYPs) (Mushiroda *et al.*, 2000). In general, the FMOs appear to behave as high-capacity enzymes with a low affinity for substrates relative to CYPs. To date, 34, 57, 40, 30, and 40 genomic allelic variants have been found for FMO1, 2, 3, 4, and 5, respectively. In addition to various wild-type forms are variants that are inactive or show decreased or functional activity (Cashman, 2004). The most well-known polymorphism is associated with FMO3, and Trimethylaminuria (TMAU). FMO3 exclusively catalyzes the *N*-oxidation and clearance of trimethylamine (TMA), a product of choline metabolism (Cashman *et al.*, 1997; Lang *et al.*, 1998). A rare subpopulation of humans (0.1% to 1% incidence in Caucasians) is deficient in FMO3 and displays a slow metabolizer phenotype status in their clearance of TMA. As TMA accumulates in excreta such as breath, sweat, and urine, it produces a characteristic "fish odor" syndrome (Arseculeratne *et al.*, 2007).

Polymorphisms in FMO2 are also known, with a 1414C allele expressed in 13% of African-Americans coding for a full-length active form of FMO2, rather than the 1414T allele that has a premature stop codon producing a truncated and inactive form of FMO2 in human lung microsomes from Caucasians and Asians (Dolphin *et al.*, 1998; Whetstine, 2000). In contrast, 26% of African-Americans and 4.5% of Hispanics express an active form of FMO2 in their lung (Cashman, 2000). Since the polymorphic and active form of FMO2 can metabolize the formation of reactive and toxic sulfenic acid metabolites of thioureas, it is plausible that individual and ethnic differences exist in the susceptibility and incidence of pulmonary toxicity (Henderson *et al.*, 2004).

In kidney microsomes, the relative expression of detected FMOs was found to be FMO1 > FMO5 > FMO3. Although an FMO1 polymorphism has not been identified, levels of this flavoprotein appeared to be greater in kidney microsomes from African-Americans than from Caucasians (Krause, Lash and Elfarra, 2003).

### 5.3 Uridine 5′-Diphosphate Glucuronosytransferases

The UGTs [E.C. 2.4.1.17] are a third class of membrane-bound enzymes expressed in the liver and intestine and often play an important role in the first-pass effect. Like CYPs and FMOs, UGTs also utilize a nucleotide cofactor, in this case, UDPGA. Unlike CYPs and FMOs, UGTs effect metabolism not by oxygenation of the substrate but by *N*- and *O*-conjugation of the substrate by group transfer of glucuronic acid (supplied by UDPGA). This typically generates a polar glucuronide conjugate that is amenable to biliary and/or renal elimination. With phenols and carboxylic acids as the typical substrates for UGTs, relative to CYPs there are few drugs that are directly eliminated by this class of enzymes. However, many oxidized products are formed from CYP-mediated reactions that can serve as substrates for conjugation and UGTs are often involved in regulating the exposure of such metabolites.

In addition to their role in drug clearance, the UGT-mediated formation of acyl glucuronides and ether conjugates has been linked to reactivity and toxicity (e.g. diclofenac) (Grillo *et al.*, 2003a, b). This can be a concern in the development of a drug and careful consideration should be given to evaluating the risks involved. This is discussed in detail in Chapter 21.

***5.3.1 Human UGT Polymorphisms and Individual Variability*** UGTs are classified by primary amino acid sequence into families and subfamilies. In humans, to date,

two gene families have been identified with a growing number of members that are currently grouped into three subfamilies (UGT1A, UGT2A, and UGT2B). Table 3 shows the major UGTs and their reported expression in liver, intestine, kidney, and/or nasal/pulmonary epithelia.

Gene mutations in the coding region and/or promoter of the UGT1A1 gene lead to decreased activity including its ability to conjugate the endogenous substrate bilirubin. This results in an observed polymorphism of UGT1A1 activity in humans, with interindividual variation in the levels of unconjugated bilirubin. A profound (90%) reduction in bilirubin clearance is seen in patients suffering from the rare Crigler–Najjar syndrome, whereas a moderate (40%) decrease is reported on average for patients suffering from the more commonplace Gilbert syndrome, which is seen in 10% of the US population and associated with the UGT1A1*28 homozygous form (Tukey and Strassburg, 2000). Inhibition of UGT1A1 can present symptomatically like Crigler–

**TABLE 3 Tissue distribution and substrates of major human uridine 5′-diphosphate glucuronyltransferases (UGTs).**

| UGT | Liver | Intestine | Kidney | Nasal/pulmonary | Substrates |
|---|---|---|---|---|---|
| 1A1 | √ | √ | | | Bilirubin; 3-OH-estradiol (major) |
| 1A3 | √ | √ | | | Hexafluoro-1α, 25-dihydroxyvitamin D3; naproxen; 3-OH-estradiol |
| 1A4 | √ | √ | | | Trifluoroperazine; amitriptyline; cyclobenzaprine; lamotrigine; olanzapine |
| 1A6 | √ | √ | | | Serotonin; 1-naphthol |
| 1A7 | | √ | | | 7-ethyl-10-hydroxycamptothecin (SN38) |
| 1A8 | | √ | | | Propofol; 3-OH-estradiol |
| 1A9 | √ | | √ | | Propofol (major) |
| 1A10 | | √ | | | 3-OH-estradiol |
| 2A1 | | | | √ | Scopoletin |
| 2B4 | √ | √ | | | Hyodeoxycholic acid (bile acid) |
| 2B7 | √ | √ | √ | | Zidovudine; morphine-6-glucuronidation; clofibric acid; diclofenac; naloxone; valproic acid; denopamine (substrate affinity encompasses UGT2B4) |
| 2B10 | √ | √ | | | Nicotine, cotinine |
| 2B11 | √ | | | | 12-HETE; 15-HETE; 13-HODE |
| 2B15 | √ | √ | | | *S*-Oxazepam |
| 2B17 | √ | | | | Testosterone; alizarin (fastest rate), also eugenol, scopoletin, galagin |
| 2B28 | √ | | | | None identified |

12-HETE: 12-hydroxyeicosatetraenoic acid; 15-HETE: 15-hydroxyeicosatetraenoic acid; 13-HODE: 13-hydroxyoctadecadienoic acid (also substrates for UGT2B10) (Fournel-Gigleux *et al.*, 1989; Jedlitschky *et al.*, 1999; King *et al.*, 2000; Tukey and Strassburg, 2000; Fisher *et al.*, 2001; Lévesque *et al.*, 2001; Gagné *et al.*, 2002; Turgeon *et al.*, 2003a, b; Williams *et al.*, 2004; Court, 2005; Kaji and Kume, 2005; Jakobsson *et al.*, 2006; Miners *et al.*, 2006; Barre *et al.*, 2007; Chen *et al.*, 2007; Kaivosaari *et al.*, 2007).

Najjar or Gilbert syndrome, and in UGT1A1-deficient individuals, could well present as profound hyperbilirubinemia. Irinotecan, an ester camptothecin derivative used to treat metastatic colon cancer, is metabolized by carboxyesterases to form SN-38, a carboxylic acid and active and toxic form of the drug. SN-38 is subsequently cleared by UGT1A1-mediated conjugation and examples exist that show increased incidence and severity of neutropenia and diarrhea in UGT1A1-deficient patients (Toffoli and Cecchin, 2004; Toffoli *et al.*, 2006).

## 6 ASSAY OPTIMIZATION FOR METABOLIC CLEARANCE AND REACTION-PHENOTYPING STUDIES

The first step in reaction phenotyping is to optimize metabolic clearance and predict total metabolic clearance *in vivo*. In general, this is done by measuring the rate of change of a therapeutically relevant substrate concentration determined over an appropriate time course. To detail all mechanisms of metabolic clearance and assay protocols is beyond the scope of this chapter. Therefore, the present discussion will focus on the most commonly encountered DME systems, specifically (i) how to optimize CYP, FMO, and/or UGT activity in liver microsomes and (ii) the tools available to identify the specific enzyme or enzymes involved and the importance of integrating the data. Beyond the necessity for routine and standardized metabolic stability screening, drug metabolism is best approached on a case-by-case basis and it is important to note that the major enzymes involved in the metabolism of a new drug may not be the commonplace DMEs described herein. In addition to obvious structural characteristics—aromatic amines tend to be substrates for *N*-acetyltransferases, monoamines substrates for monoamine oxidases, epoxides substrates of epoxide hydrolases, and so on—*in vivo* excretion balance and biotransformation studies are extremely helpful in identifying metabolic reactions that are not covered by standardized assays. Often the clearest indication of an "atypical" metabolic clearance mechanism is seen when a compound is found to be metabolically stable in liver microsomes and yet is rapidly and/or extensively metabolized *in vivo* (i.e. relatively little unchanged parent compound is measured in excreta). In such circumstances, one must be prepared to take a different approach and consider using other assay conditions, subfractions such as cytosol or mitochondria, and non-hepatic tissues such as gut, skin, nasal mucosa, kidney microsomes, with experimental protocols adapted accordingly. Sequential pathways of metabolism can be encountered that involve several of these enzymes: microsomal, cytosolic, hepatic, as well as extrahepatic (e.g. carbamazepine). In these cases, radiolabeled and stable-label substrates are very useful in understanding the mechanisms involved.

### 6.1 Optimizing CYP Activity in Human Liver Microsomes

Livers procured from animals or human organ donors can be perfused to isolate hepatocytes or homogenized and centrifuged to isolate liver mitochondrial, S9, microsomal, and/or cytosol subfractions. Liver microsomes are the cheapest and easiest system to use and the most commonly used enzyme system for *in vitro* drug metabolism studies. The rate and extent of turnover of test compound are measured in incubations supplemented with an excess of NADPH as a cofactor (directly or as part of a regenerating

system). Buffers like potassium phosphate or Tris are used to control the pH, with reactions normally performed at pH 7.4 and 37 °C to simulate physiological conditions. Magnesium chloride is often included in the incubation medium to modify the ionic strength and sometimes this can affect the turnover of substrate. Test compounds usually are added in solution, with the solvent often a mix of dimethylsulfoxide (DMSO) and acetonitrile or methanol and aqueous buffer. It is important to keep the total volume of organic solvent as low as possible since DMSO, acetonitrile, and methanol are each known to differentially inhibit CYP activity (Chauret, Gauthier and Nicoll-Griffith, 1998; Hickman *et al.*, 1998). If not taken into account, the quantification and prediction of a particular CYP's role in the drug's metabolism can be incorrectly estimated.

Human liver microsomes (HLMs) are usually prepared from a pool of donors (typically from N = 12 to 20), with male and female donors included. This helps account for the known heterogeneity in relative expression and activity that exists in the human population. If the individuals are chosen correctly, the metabolic clearance in the HLM pool will represent an "average" but fully functioning human liver (i.e. one without any null polymorphisms and with basal activity of the inducible CYP forms). Animals tend to be more homogeneous in their enzyme expression and liver microsomes are often prepared from the liver of a single untreated animal (or a pooled group if animals have been pretreated with an inducer or potential inducer). Variability does exist even in animals, and may also need to be taken into account; for instance, CYP2D15 polymorphism affects the metabolism and pharmacokinetics of celecoxib in dogs (Paulson *et al.*, 1999). More commonly encountered are gender-specific differences in CYP expression; for example, CYP2C11 is a male-specific enzyme in the rat (Czerniak, 2001). These animal polymorphisms and animal-specific DMEs may show up in the safety and tolerability studies conducted in support of a regulatory filing submitted prior to human clinical trials. Such findings may necessitate additional mechanistic studies to show their clinical relevance.

### 6.2 *In Vitro* Tools for Identifying Cytochrome P450s

Of the human DMEs, the CYPs are the most mature and advanced in terms of reaction-phenotyping tools. In fact, *in vitro* procedures have become increasingly standardized. However, it is important to remember that the different *in vitro* data sets have to be consistent in order to enable an integrated *in vitro* CYP reaction phenotype (Fig. 2), sound decision making, and successful *in vitro–in vivo* correlations (Zhang *et al.*, 2007).

***6.2.1 Selective Inhibitors*** Possibly the most useful and direct way of identifying what CYP metabolizes a drug is to co-incubate it in the absence and presence of a known selective inhibitor of that enzyme. In this instance, the activity is measured in the absence (solvent alone) and presence of the inhibitor. The "% inhibition" or fractional loss of activity (versus solvent alone) represents the fraction (or %) of the metabolism catalyzed by the enzyme in question:

$$\% \text{ Inhibition} = \frac{\text{vo} - \text{vi}}{\text{vo}} * 100$$

$$\text{or fractional loss of activity for each CYPi} = \frac{\text{vo} - \text{vi}}{\text{vo}}$$

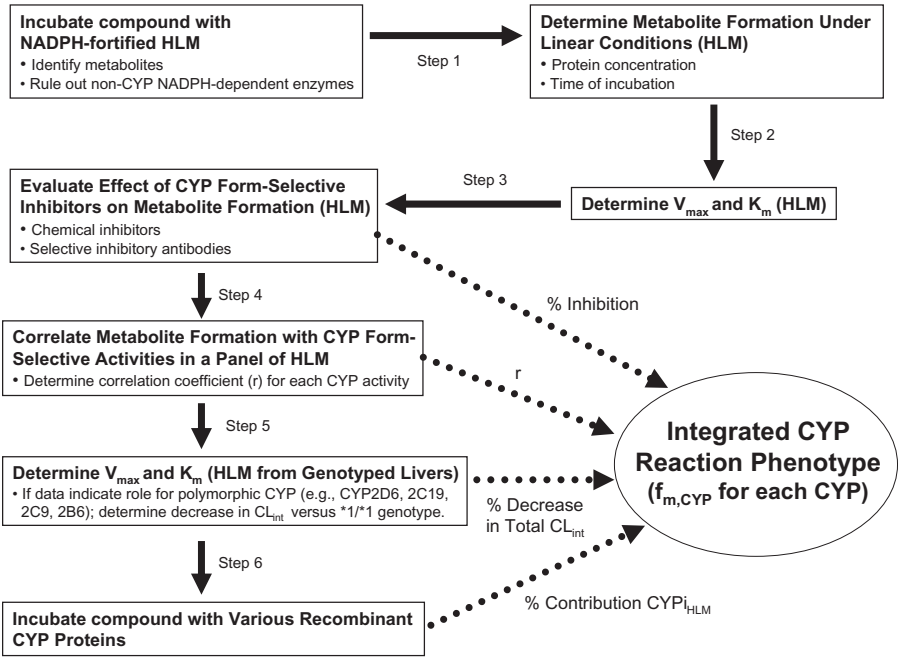


**Figure 2** Illustration of a possible *in vitro* cytochrome P450 (CYP) reaction-phenotyping strategy. Adapted from Zhang *et al.* (2007), where $V_{max}$ is the maximal rate of reaction; $K_m$ is the Michaelis constant; HLM is human liver microsomes; $f_m$ is the fraction metabolized; CYP is cytochrome P450s, $CL_{int}$ is intrinsic clearance; and r is the correlation coefficient.

where vo is the disappearance rate constant or velocity of reaction and vi is the disappearance rate constant in the presence of an inhibitor; CYPi represents a specific CYP or CYP-mediated reaction.

The fraction metabolized ($f_m$) is an important factor in understanding the consequences of a coadministered inhibitor:

$$\text{Total Clearance} = \text{Total Clearance} * (f_{\text{non-metabolic}} + f_m)$$
$$1 = f_{\text{non-metabolic}} + f_m$$

where $f_{non\text{-}metabolic}$ is the fraction cleared by mechanisms other than metabolic clearance (e.g. renal and/or biliary elimination); $f_m$ is the fraction metabolized—measured as the fraction of dose recovered as metabolites in excreta.

In a situation where all of the clearance is governed by metabolism, then $f_m = 1$. In reaction phenotyping, this concept is extended further to define $f_{m.CYP}$, which is the fraction metabolized by CYPs, and further still to $f_{m.CYPi}$, which describes the fraction metabolized by a specific CYP (the sum of these individual reactions should account for the total $f_{m.CYP}$). For example, if $f_{m.CYP3A4} = 1$, then the entire CYP-mediated clearance is by CYP3A4. If in this case $f_m$ also accounts for most of the *in vivo* clearance ($f_m \sim 1$), then CYP3A4 determines the *in vivo* clearance and coadministered CYP3A4-selective inhibitors of this enzyme could have a great effect on total exposure, especially

relative to a drug with a low $f_m$ or one with an equivalent $f_m$ that results from multiple CYP-mediated reactions, where each $f_{m.CYPi}$ is <0.5. Identifying the fraction metabolized by a specific CYP is the goal of reaction phenotyping. The product of $f_m$ and $f_{m.CYP}$ determines the overall contribution of that CYP to clearance:

$$\text{Clearance}_{CYPi} = \text{Total Clearance} * f_m * f_{m.CYPi}$$

Knowing this for a drug candidate allows one to identify some of the risks associated with DDIs, polymorphisms, and interindividual variability in the target population.

Given that CYPs are distinguished in part by their promiscuity, there are relatively few drugs and chemicals that are uniquely selective. A group of tool compounds has been identified as most appropriate for the purpose of *in vitro* reaction phenotyping. Table 4 lists the substrates and inhibitors identified by their selectivity for the major CYPs and that come recommended in the FDA draft guidelines for *in vitro* characterization of drug candidates (FDA, 2006). In the table are listed the preferred choice of substrate and inhibitor and acceptable alternative compounds. (The FDA guidelines are subject to review and a change to the preferred list of substrates and inhibitors is very possible. It is advised, if practicable, to perform studies using the most current recommendations available.) It should be noted that with CYP3A reactions, in particular, it is common to use multiple selective yet structurally dissimilar probe substrates or inhibitors to account for some of the more complex kinetics associated with this enzyme (Kenworthy *et al.*, 1999). The same also appears to be true for CYP2C9, with a panel of CYP2C9 inhibitors differentiating in their potency toward inhibition of (*S*)-flurbiprofen, (*S*)-warfarin, phenytoin, tolbutamide, and diclofenac (Kumar *et al.*, 2006). The heteroactivation of CYP2C9 substrates by dapsone also supports the kinetic similarity with CYP3A4 and suggests a "two-site" binding model is appropriate for CYP2C9 (Hutzler, Hauer and Tracy, 2001). This is further supported by the discovery of a second binding site for warfarin, based on studies of a crystal structure of CYP2C9, and the apparent capacity of a relatively large active site (~470 $Å^3$) that can accommodate the simultaneous binding of multiple ligands (Williams *et al.*, 2003b).

If a single CYP metabolizes a drug, then coadministered inhibitors of that CYP should demonstrate a profound effect on the rate and extent of metabolism. This is most easily discerned by measuring the rate of formation of a specific metabolite and, for substrates with shorter half-lives, depletion-kinetic approaches also work quite well. Table 5 lists the most common CYP-selective chemical inhibitors used and the recommended concentrations for reaction phenotyping (based on the literature reported maximal inhibition of various CYP-specific biotransformation reactions). For a substrate metabolized by multiple CYPs, a single selective inhibitor will affect only one component of the turnover and several inhibitors are needed to characterize the compound (the individual effects are summed to account for the total clearance). In reality, the more CYPs that are involved in the metabolic clearance of a compound, the harder it becomes to identify the contribution of each enzyme to the total clearance. Fortunately, this is also the preferred profile of a drug, and the same issue applies *in vivo*—consequently, CYP DDIs become unlikely to profoundly alter its clearance or AUC and render it a so-called "victim."

***6.2.2 CYP Inhibitory Antibodies*** CYP inhibitory antibodies are an alternative to chemical inhibitors and yield similar data (e.g. % inhibition for each CYP form). Unlike

**TABLE 4 Cytochrome P450 (CYP) inhibitors, substrates, and inducers (United States FDA recommended *in vitro* tools).**

| | Inhibitor | | Substrate | | Inducer | |
|---|---|---|---|---|---|---|
| CYP | Preferred | Acceptable | Preferred | Acceptable | Preferred | Acceptable |
| 1A2 | Furafylline | α-Naphthoflavone | Phenacetin-O-deethylation | 7-Ethoxyresorufin-O-deethylation | Omeprazole | Lansoprazole |
| | | | | Theophylline-*N*-demethylation | β-Naphthoflavone | |
| | | | | Caffeine-3-*N*-demethylation | 3-Methylcholanthrene | |
| | | | | Tacrine 1-hydroxylation | | |
| 2A6 | Tranylcypromine | Pilocarpine | Coumarin-7-hydroxylation | | Dexamethasone | Pyrazole |
| | Methoxsalen | Tryptamine | Nicotine C-oxidation | | | |
| 2B6 | | 3-Isopropenyl-3-methyl diamantane | Efavirenz hydroxylase | Propofol hydroxylation | Phenobarbital | Phenytoin |
| | | 2-Isopropenyl-2-methyl adamantane | Bupropion-hydroxylation | *S*-mephenytoin-*N*-demethylation | | |
| | | Sertraline, phencyclidine | | | | |
| | | Triethylenethiophosphoramide (thioTEPA) | | | | |
| | | Clopidogrel, ticlopidine | | | | |
| 2C8 | Montelukast | Trimethoprim, gemfibrozil | Taxol 6-hydroxylation | Amodiaquine *N*-deethylation | Rifampin | Phenobarbital |
| | Quercetin | Rosiglitazone, pioglitazone | | Rosiglitazone para-hydroxylation | | |
| 2C9 | Sulfaphenazole | Fluconazole | Tolbutamide methyl-hydroxylation | Flurbiprofen 4′-hydroxylation | Rifampin | Phenobarbital |
| | | Fluvoxamine | *S*-warfarin 7-hydroxylation | Phenytoin-4-hydroxylation | | |
| | | Fluoxetine | Diclofenac 4′-hydroxylation | | | |
| 2C19 | | Ticlopidine | *S*-mephenytoin 4′-hydroxylation | Omeprazole 5-hydroxylation | Rifampin | |
| | | Nootkatone | | Fluoxetine O-dealkylation | | |
| 2D6 | Quinidine | | (±)-bufuralol 1′-hydroxylation | Debrisoquine 4-hydroxylation | None identified | |
| | | | Dextromethorphan O-demethylation | | | |
| 3A4/5 | Ketoconazole | Azamulin | Midazolam 1′-hydroxylation | Erythromycin *N*-demethylation | Rifampin | Phenobarbital |
| | Itraconazole | Troleandomycin | Testosterone 6 β-hydroxylation | Dextromethorphan *N*-demethylation | | Phenytoin, taxol |
| | | Verapamil | | Triazolam 4-hydroxylation | | Rifapentine |
| | | | | Terfenadine C-hydroxylation | | Troglitazone |
| | | | | | | Dexamethasone |

Adapted from FDA (2006).

**TABLE 5 Common chemical inhibitors and recommended concentrations for cytochrome P450 (CYP) reaction phenotyping *in vitro*.**

| CYP form(s) | Inhibitor | Inhibitor concentrations (μM) |
|---|---|---|
| CYP1A2 | Furafylline | 10–30[a] |
| | α-Naphthoflavone | 1 |
| CYP2A6 | Methoxsalen | 1 |
| CYP2B6 | ThioTEPA | 50 |
| CYP2C8 | Montelukast | 0.1[b] |
| CYP2C9 | Sulfaphenazole | 10 |
| CYP2C19 | Benzylnirvanol | 1 |
| CYP2D6 | Quinidine | <2 |
| CYP2D6 and CYP3A4/5 | | 10 |
| CYP2E1 | Diethyldithiocarbamate | 50 |
| CYP3A4/5 and others | Ketoconazole | 1 and 10 |
| CYP3A4/5 | Troleandomycin | 50[a] |

[a]Inhibition potency increases with preincubation in the presence of NADPH (time-dependent inhibitors).
[b]Conditions may be subject to the concentration of human liver microsomal protein used in the assay.
Data from Muralidharan *et al.* (1991); Kunze and Trager (1993); Chang, Gonzalez and Waxman (1994); Baldwin *et al.* (1995); Bourrié *et al.* (1996); Eagling, Tjia and Back (1998); Yamazaki and Shimada (1998); Zhang *et al.* (2001); Suzuki *et al.* (2002); Cai *et al.* (2004); Turpeinen *et al.* (2004); Walsky *et al.* (2005).

chemical inhibitors that tend to work competitively, inhibitory antibodies work non-competitively by modulating the conformation of the CYP hemoprotein complex. This can simplify the kinetic optimization of substrate inhibition otherwise based on $K_m$ and $K_i$ properties of substrate and inhibitor, respectively. At the National Institutes of Health (NIH), Gelboin and Krausz have isolated and characterized specific CYP inhibitory monoclonal antibodies for the major human DMEs and shown ascites and purified MAbs to selectively inhibit up to 80–90% of human liver microsomal activity (Gelboin and Krausz, 2006). They even identified an MAb selective for CYP2C9*2 allele that does not significantly inhibit the CYP2C9*1 wild-type form, despite just one amino acid difference in their sequence (this level of specificity should be taken into account, especially when using liver microsomes from individual donors). MAbs are potentially most valuable when the available chemical inhibitors lack the desired selectivity; with CYP2A6 and 2B6, especially, inhibitory MAbs may be the only way to provide a reasonably definitive assessment of their contribution to total clearance.

***6.2.3 Correlation Analysis*** A more labor-intensive but occasionally useful approach to reaction phenotyping is to measure the rate of metabolism and/or intrinsic clearance in a panel of HLMs prepared from individual donors (usually a minimum of $N \geq 12$ is recommended). This exploits the relationship between rate of metabolism and total amount of active enzyme present in the liver microsomes (see Section 8). The observed interindividual variability of the new drug is then correlated with the metabolism of known CYP-specific substrates using the same human liver donors. For metabolic clearance or specific metabolic pathways governed by a single CYP, a positive correlation should exist between the archetypal probe compound and the new drug. Benetton *et al.* (2007), Ghosal *et al.* (2007), Lee *et al.* (2007a, b, 2008), Obach, Margolis and Logman (2007), and Yoon *et al.* (2007) offer recent examples of correlation analyses coupled with other reaction-phenotyping methods to identify the major enzymes involved in the metabolic clearance of selegiline, vicroviroc, eupatilin,

mirodenafil, CP-122,721, and benidipine, respectively. The substrates listed in Table 4 are commercially available and considered to have the most favorable specificity for a particular CYP; these are best suited for most correlation analyses.

A significant disadvantage to this method is the possibility of covariates and the difficulty in discerning multiple pathways of metabolic clearance, especially difficult if *in vitro* metabolism is relatively slow. As such, correlation analysis is not used as a sole means of identifying the reaction phenotype but as another tool to help identify those CYPs without ideal selective inhibitors. It can also help to illustrate quite clearly the potential for interindividual variability in metabolic clearance and possible effect on safety margins for a drug.

***6.2.4 Recombinant Enzymes*** Very selective tools are made for drug metabolism studies by transfecting a human cDNA for a particular CYP into a vector and its transfer into a rapidly growing and easily cultured cell line, such as lymphoblastoid cells, *Escherichia coli*, or insect cells. The culture is then grown to produce a batch of cells from which microsomes are made that possess enzyme activity enriched with a single human CYP or other enzyme (CYP, FMO, and UGT recombinant enzymes are commercially available). To secure functional CYP activity, these systems require co-expression with NADPH oxidoreductase and, for some substrates, co-expression also with cytochrome b5. A panel of recombinant CYP enzymes (e.g. CYP1A2, 2A6, 2B6, 2C8, 2C9, 2C19, 2E1, 3A4, and 3A5, or a subset thereof) is incubated with the test compound, using buffer and conditions similar to those described for liver microsomal assays. By measuring the turnover of compound in the individual CYP microsomes, the potential of each enzyme to metabolize the substrate can be quite easily determined and provides useful data to help predict its phenotype in humans. Recombinant and reconstituted enzyme systems are especially useful in showing the potential role in drug metabolism for a known polymorphic enzyme (e.g. CYP2C19, CYP2D6, and UGT1A1). This can help devise a strategy to evaluate and define any related issues. Enzyme kinetics and *in vitro* to *in vivo* extrapolation from liver microsomal or recombinant enzyme data are covered in a later section.

### 6.3 Optimizing FMO Activity in Human Liver Microsomes

Conditions that are optimal for liver microsomal CYP activity usually enable FMO reactions to take place (although the true optimum for the flavoproteins is at a higher non-physiological pH of 9–10). FMOs, especially in the absence of NADPH, are less thermally stable than CYPs and livers need to be handled carefully and microsomes prepared under controlled and proper conditions to maintain full activity. Without this, there is a risk of underestimating the contribution of FMOs to metabolic clearance. This thermal instability also provides a useful tool to aid in the reaction phenotyping of gross FMO activity in HLMs by substituting as a nonselective *in vitro* inhibitor (Cashman, 2005).

Reaction phenotyping of FMO activity tends to follow on from CYP studies, that is, if a significant contribution of NADPH-dependent metabolic clearance is not accounted for by CYPs, then evaluating a contribution from FMO-mediated clearance becomes an obvious next step. This is done by using a short (e.g. five minutes) preincubation of HLM at 45 °C in the absence and presence of saturating levels of NADPH. Without NADPH, the FMOs are thermally inactivated while CYP activity remains viable, while NADPH provides a protective effect and acts as a control (it is recom-

mended that one also compare activity with the same HLM without the heat treatment, i.e. the CYP and FMO are both constitutively active). Another approach to distinguish FMO versus CYP involvement is to use 1-benzylimidazole as a nonselective CYP inhibitor (Ziegler, 1988; Grothusen *et al.*, 1996).

If these experiments demonstrate FMO to be a major contributor to metabolic clearance *in vitro*, then the next step is to identify the specific enzyme form(s) involved. The reported characteristics for different FMOs allow one to take a rational approach in identifying the specific FMO with a process of elimination based on anecdotal evidence. Although FMO5 has the highest expression in the human liver, it tends to have much lower activity and much narrower substrate selectivity than FMO3. Despite possible exceptions in which FMO5 cannot be discounted, in most cases the activity measured in HLM will reflect primarily FMO3-mediated metabolism. As with CYPs, recombinant FMO enzymes are available that allow one to easily demonstrate functional activity and possibly selectivity. If turnover is seen with FMO3 and FMO5, then, if necessary, a correlation analysis can be performed using a panel of HLMs from a range of donors characterized for FMO3 and/or FMO5 activity. Since most drugs metabolized by human FMOs are substrates of FMO3, reaction phenotyping usually starts with substrates relatively selective for this form (used either to correlate with interindividual variability in rate of metabolism and/or as potential competitive inhibitors of the test drug). TMA, (*S*)-nicotine, cimetidine, clozapine, and ranitidine are typical substrates for FMO3 and lend themselves to such correlation analyses (TMA: Lang *et al.*, 1998; (*S*)-nicotine: Cashman *et al.*, 1992; cimetidine: Cashman *et al.*, 1993; clozapine: Fang, 2000; ranitidine: Kang *et al.*, 2000). As mentioned previously, covariates are always a risk when it comes to correlation analyses and the potential contribution from other FMOs, and also from CYPs, should be considered before concluding a major role for FMO in metabolic clearance.

Substrates and metabolic reactions that appear to be relatively selective for the individual FMOs are shown in Table 6. Reaction phenotyping is made difficult by a paucity of truly selective probe compounds and an even worse dearth of FMO form-selective inhibitors. One alternative approach is to perform a correlation analysis of activity using tissues known to have a relatively greater expression of a particular FMO (Table 7). For instance, kidney microsomes can be used as an enriched source of FMO1 and FMO2, lung microsomes from a panel of African-Americans with the 1414C allele provide a source of active FMO2, and microsomes from the small intestine are enriched with FMO5 activity (Cashman and Zhang, 2006). Obtaining these tissues is very

**TABLE 6 Known substrates for the major human flavin-containing monooxygenases (FMOs).**

| FMO | Substrates |
|---|---|
| FMO1 | Xanomeline, *S*-oxidation of the *S*-methyl metabolite of disulfiram[a] |
| FMO2 | Thioureas, thioethers |
| FMO3 | Trimethylamine, *N*′-oxidation of (*S*)-nicotine, clozapine-*N*-oxidation, cimetidine-*S*-oxide formation, ranitidine-*N*-oxidation |
| FMO4 | Unstable form with no selective substrate identified |
| FMO5 | Thioethers with proximal carboxylic acid[b] |

[a]Cashman and Zhang (2006).
[b]Ohmi *et al.* (2003).
References contained within Cashman (2000)—unless indicated otherwise.

**TABLE 7 Relative expression of flavin-containing monooxygenases (FMOs) in various adult human tissues.**

| | Copies per ng RNA | | | | |
|---|---|---|---|---|---|
| Adult tissue | FMO1 | FMO2 | FMO3 | FMO4 | FMO5 |
| Liver | 96 | 987 | 23,087 | 4,882 | 26,540 |
| Kidney | 6,198 | 4,683 | 531 | 2,510 | 1,628 |
| Small Intestine | 523 | 929 | 74 | 403 | 2,586 |
| Lung | 596 | 115,896 | 2,224 | 738 | 2,274 |
| Brain | 3.1 | 141 | 11 | 20 | 57 |

Data normalized to HPRT (hypoxanthine phosphoribosyl transferase) and expressed as copies per ng RNA (adapted from Cashman and Zhang, 2006).

difficult and expensive and care must be taken when handling the tissues to conserve the FMO activity. Except in the most unusual circumstances, it is unlikely that such reductive measures will be taken in drug discovery.

### 6.4 Optimizing UGT Activity in Human Liver Microsomes

The *in vitro* to *in vivo* extrapolation of UGT activity is notoriously difficult to do and optimizing *in vitro* liver microsomal conditions for metabolism is normally done on a compound-specific basis. This does not lend itself readily to quantitative screening, at least not beyond a "yes" or "no" conclusion about whether or not a compound is a substrate for UGT enzymes. Usually the bias of UGT phenotyping is to identify possible substrates that may act competitively with others and result in DDIs (UGT1A1 and UGT2B7 being the main concern). With this class of enzymes, often it is necessary to correlate *in vitro* studies in liver microsomes or hepatocytes with *in vivo* data, and use that to decide on its value insofar as discriminating compounds for the structure-activity relationship (SAR) and oral exposure. With UGTs expressed on the luminal side of the endoplasmic reticulum, liver microsomes need to be "activated" to obtain optimal activity. The pore-forming agent alamethicin has become commonly used for this purpose (Fulceri *et al.*, 1994; Fisher *et al.*, 2000). Other methods include using detergents like Brij58 or mechanical techniques like sonication to disrupt the membranes and enable the substrates to access the UGTs (Vanstapel and Blanckaert, 1988; Shepherd *et al.*, 1989; Coughtrie *et al.*, 1991; Visser *et al.*, 1993; Soars *et al.*, 2001; Soars, Ring and Wrighton, 2003). UGTs utilize UDPGA as a requisite cofactor, and usually this is supplemented *in vitro* at high concentrations (10 mM). With UDPGA and NADPH both quite soluble in aqueous buffers, they can be added together allowing both UGT and CYP reactions to be measured simultaneously.

UGT-mediated clearance and metabolism are best predicted using hepatocytes rather than liver microsomes (Miners *et al.*, 2006). The advantage to hepatocytes is that they possess the full complement of native liver DMEs, which includes CYPs and UGTs. They also provide a more "natural" environment than liver microsomes, and appear not to have a significant issue with the release of UGT inhibitory unsaturated fatty acids that can affect the liver microsomal metabolism of lamotrigine, zidovudine, and 4-methylumbelliferone (Rowland *et al.*, 2006, 2007; Uchaipichat *et al.*, 2006b). Hepatocytes do suffer the disadvantage of being harder to obtain and use than liver microsomes, with cell viability and yield important additional factors to consider. The use of cryopreserved hepatocytes rather than freshly isolated cells helps to circumvent some of these problems.

There are very few UGT-selective inhibitors identified to date, nor are there many UGT isozyme-specific probes available for correlation analysis. Therefore, reaction phenotyping of UGTs is presently much harder to do than for CYPs. Probably the most useful tool available is the recombinant UGT enzymes. Unfortunately, the UGTs are not very well characterized with respect to their absolute abundance in the liver or other tissues, and scaling from recombinant or other *in vitro* data to human *in vivo* metabolic clearance is not yet possible. Correlation analysis with substrates and reactions that are relatively selective is probably the next best approach to take in reaction phenotyping a new drug. Kaji and Kume (2005), Tachibana *et al.* (2005), Yamanaka *et al.* (2005), Chen *et al.* (2007), Katoh *et al.* (2007), Mano, Usui and Kamimura (2007a), and Yu *et al.* (2007) provide a few examples of correlation analyses as used in the reaction phenotyping of afloqualone, trans-3′-hydroxycotinine, fluoroquinolones (levofloxacin, grepafloxacin, moxifloxacin, and sitafloxacin), racemic flurbiprofen, tranilast, and mitiglinide, respectively.

Table 3 lists the major human UGTs and substrates with reactions known to be mediated by those enzymes and their known expression in several major tissues. Correlation analysis can be done with these substrates; however, since many are metabolized by multiple UGTs and other enzymes (e.g. CYPs), several such studies may be needed using several different reference compounds. Although absolute human levels of UGTs are not yet available, there are tissue-specific differences in the relative expression of UGTs that may help to pinpoint a role for certain UGTs in the conjugative clearance of a new drug (e.g. UGT1A8 in small intestine and UGT1A9 in kidney; Table 3).

The current lack of good *in vitro* tools, coupled with the complex and compound-specific kinetics that are often encountered with glucuronidation reactions, makes it difficult to identify with confidence the contribution of a single UGT to the total clearance of a drug. At this stage, often the goal in drug discovery is to (i) identify any potential risk of UGT1A1 inhibition and issues associated with hyperbilirubinemia and (ii) select a compound that is metabolized by multiple isoforms (UGT and CYP). More work is needed to support quantitative UGT reaction phenotyping, and one should proceed cautiously if developing a UGT substrate with a relatively narrow therapeutic index. This is true also if a potentially reactive acyl glucuronide metabolite is formed, for example, the acylglucuronide metabolite of diclofenac and its putative role in immunotoxicity (Ebner *et al.*, 1999; Grillo *et al.*, 2003a; Naisbitt *et al.*, 2007).

## 7 *IN VITRO* TO *IN VIVO* EXTRAPOLATION OF METABOLIC CLEARANCE

### 7.1 *In Vitro* Intrinsic Clearance

By characterizing a drug's properties *in vitro* and *in vivo* using animal systems and available *in vitro* human systems, it becomes possible to predict *a priori* the *in vivo* properties of a drug in humans. Clearance (CL) and volume of distribution (Vd) are key properties in deciding the exposure and duration of a drug *in vivo* with half-life derived from both ($t_{1/2}$ = Vd/CL). Therefore, in addition to identifying the enzymes involved in the metabolism of a drug, one also needs to predict the total metabolic clearance and contribution of individual metabolic clearance reactions to the overall clearance. This is done simply using a rate or metabolic half-life of metabolism to calculate the intrinsic clearance of a drug that in turn is scaled to predict its *in vivo* clearance in the animal or human. In general, if the *in vitro* assays correctly predict the

*in vivo* clearance measured in animal models, then there is some confidence that the same will hold true for the prediction in humans.

Simple liver microsomal assays are very useful for differentiating compounds with metabolic clearance issues. As discussed previously, metabolic stability screens are often used as a first tier assay, with a single fixed protein concentration used and loss of test compound measured at a single time point relative to the starting concentration (e.g. $C_{10\,\text{minutes}}$ relative to $C_0$). These assays differentiate compounds with rapid turnover from those with no measurable turnover and thus easily identify those compounds with metabolic stability issues. However, in reaction phenotyping and *in vitro* to *in vivo* extrapolation, intrinsic clearance is a more useful measure and accuracy of prediction becomes extremely important. Intrinsic clearance or $CL_{int}$ can be defined as the proportionality constant between the concentration of substrate at the enzyme ($C_E$) and its rate of metabolism in the absence of physiological constraints:

$$\text{Rate}(v) = CL_{int} \times C_E$$

For a single substrate-binding reaction, the chemical reaction can be described as:

$$\mathbf{E} + \mathbf{S} \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} \mathbf{ES} \leftrightarrow \mathbf{EP} \underset{k_{-2}}{\overset{k_2}{\rightleftharpoons}} \mathbf{P} + \mathbf{E}$$

Further simplification is possible if this reaction demonstrates Briggs–Haldane kinetics: (i) a quick rate of substrate, S, binding to the enzyme, E; (ii) a relatively slow reaction step in the formation of EP; (iii) a rapid rate of release of product, P (the latter defined by $k_2$); and (iv) the reaction is irreversible. These are reasonable assumptions for most reactions and allow the chemical reaction scheme to be reduced to a pseudo-first-order reaction:

$$\mathbf{E} + \mathbf{S} \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} \mathbf{ES} \xrightarrow{k_2} \mathbf{P} + \mathbf{E}$$

where $k_2 = k_{cat}$, catalytic rate constant or turnover number.

The enzyme kinetics involved in a typical bimolecular reaction can be daunting but fortunately can also be greatly simplified—subject to various assumptions and the correct use of appropriate assay conditions—while still providing a reasonably close approximation. For most CYP-mediated reactions, it is assumed that the substrate and enzyme are in equilibrium and in quasi-steady-state conditions, with the enzyme-substrate binding complex (ES) forming and dissociating at an equal rate (i.e. $d[ES]/dt \sim 0$), which is not true initially but becomes a valid assumption upon mixing of substrate and enzyme.

When these assumptions are true, the kinetics for the reaction can be simply described using the classical Michaelis–Menten equation:

$$v = \frac{V_{max} * [S]}{K_m + [S]}$$

where v = rate or $-dS/dt$; $V_{max}$ = the maximal rate of reaction; [S] = substrate concentration and $K_m$ = Michaelis constant, the concentration at which the rate of reaction is half of its maximum.

If a substrate is metabolized by multiple pathways, then an additive approach is used to calculate the kinetics of metabolism:

$$v = \frac{V_{max1} * [S]}{K_{m1} + [S]} + \frac{V_{max2} * [S]}{K_{m2} + [S]} + \frac{V_{maxn} * [S]}{K_{mn} + [S]}$$

Michaelis–Menten kinetics are usually determined by nonlinear regression analysis of substrate concentration ([S]) versus the rate of reaction (v), which in the classical case results in a hyperbolic curve (Fig. 3a). The data can also be transformed to make linear plots. The Eadie–Hoftsee transformation plots v/[S] against v, the slope of which provides the $K_m$ and the intercept on the y-axis provides $V_{max}$. This transformation is especially useful in identifying multiple pathways in the metabolism of a compound. Figure 3b illustrates the Eadie–Hofstee transformation of a hypothetical drug that is metabolized by a single enzyme (or multiple enzymes with identical properties), with a straight-line plot generated. Figure 3c illustrates an Eadie–Hofstee transformation of a hypothetical drug that is now metabolized by two distinct enzymes, with a biphasic profile generated as a result of the low- and high-affinity enzymes.

More complicated kinetic models account for homo- and hetero-cooperativity, where multiple molecules of the same or different substrate, respectively, bind to the same molecule or complex of enzyme and modulate the catalytic function of the enzyme. This can result in sigmoidal kinetics, often represented in a simplified form using the Hill equation:

$$v = \frac{V_{max} * [S]^{\gamma}}{S_{50}^{\gamma} + [S]^{\gamma}}$$

where $\gamma$ is the Hill coefficient: if $\gamma = 1$, then the reaction obeys classical Michaelis–Menten kinetics.

These equations all possess the mathematical form that represents the hyperbolic nature of CYP-mediated reactions. This means that the rate of metabolism or turnover increases as the substrate concentration increases, with the rate eventually limited by the amount of enzyme present and its innate catalytic capacity (known as the enzyme's $k_{cat}$ or turnover number). This leads to an observed $V_{max}$ in the reaction ($V_{max} = k_{cat}$*amount of active enzyme present, Ai). By classical Michaelis–Menten kinetics, the $V_{max}$ occurs at twice the $K_m$ of that drug for that enzyme, and when $[S] << K_m$, $v/[S] = V_{max}/K_m$, which is the apparent intrinsic clearance. Therefore, by measuring the apparent *in vitro* $V_{max}$ and $K_m$ for each metabolic reaction involved, one obtains a definitive kinetic analysis for the drug—in addition to being able to calculate the intrinsic clearance, which helps to predict the overall metabolic clearance, one can use the $V_{max}$ and $K_m$ to predict nonlinearity in the drug's pharmacokinetics and predict the effect of enzyme induction or inhibition on drug clearance.

The typical and most definitive method used to measure $K_m$ and $V_{max}$ is to incubate a range of substrate concentrates and measure the rate of metabolism at each concentration under linear kinetic conditions. This means that substrate depletion is minimized during the course of incubation, with total turnover ideally <8% (practically <20% is often used because of the constraints imposed by analytical methods). In practice, linear conditions are identified for enzyme content and time for the lowest concentration of substrate and then used to measure rate of metabolism of a range of

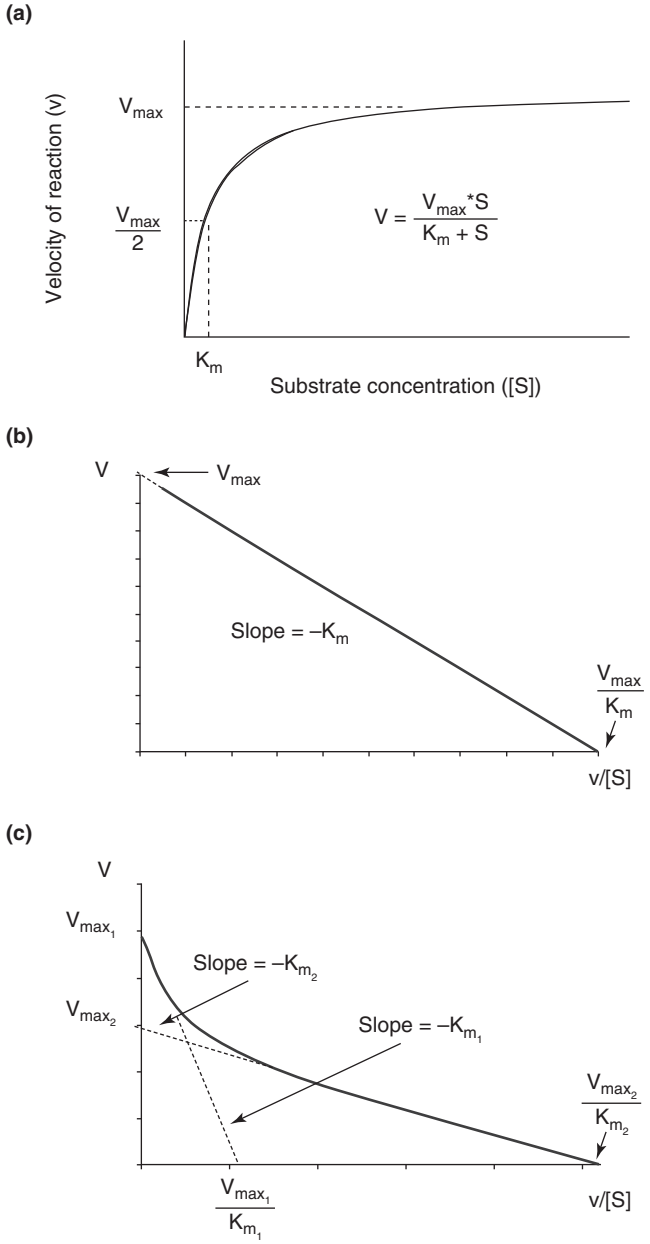


**Figure 3** Michaelis–Menten kinetics and Eadie–Hofstee transformations: **(a)** Michaelis–Menten kinetics—single enzyme; **(b)** Eadie–Hofstee transformation—single-enzyme reaction; **(c)** Eadie–Hofstee transformation—two-enzyme reaction; v = reaction rate or velocity; $V_{max}$ = maximal rate of reaction; [S] = substrate concentration; $K_m$ = Michaelis constant.

substrate concentrations (e.g. 1 to 1000 μM, subject to the sensitivity limits of the assay). A typical profile is shown in Fig. 3a. The $K_m$ and $V_{max}$ are calculated using nonlinear regression analysis of the concentration versus rate data (various commercial software are available for this purpose).

Measuring the rate of formation of metabolites requires the use of either radiolabeled substrate, with resolution achieved by liquid chromatography and/or by quantification relative to calibration curves prepared from reference standards. Radiolabel and metabolite standards are not always available, especially in the early drug discovery stage. In which case, substrate-depletion kinetics are normally used to estimate intrinsic clearance. This is achieved easily enough by measuring the half-life of substrate turnover in the liver microsomal incubation and applying some further simplification to the equations involved:

$$V = \frac{-d[S]}{dt} = \frac{V_{max} * [S]}{K_m + [S]}$$

This equation transposes to:

$$V_{max} * dt = \frac{-d[S] * (K_m + [S])}{[S]}$$

Over one half-life ($t_{1/2}$), $[S] = 0.5*[S]_0$ and the equation can be reduced to:

$$\frac{V_{max} * t_{1/2}}{K_m} = Ln2 + \frac{0.5 * [S]_0}{K_m}$$

For this to be valid, the $[S]_0 << K_m$, in which case:

$$\frac{0.5 * [S]_0}{K_m} \ll Ln2$$

Usually a low concentration of substrate is used (typically ~1 μM) and assumed to be below the $K_m$ with the experimental data inspected to confirm the assumption. If valid, the equation reduces further to:

$$\frac{V_{max} * t_{1/2}}{K_m} = Ln2$$

which is easily rearranged to provide an apparent intrinsic clearance ($CL'_{int}$):

$$\frac{V_{max}}{K_m} = \frac{Ln2}{t_{1/2}} = CL'_{int}$$

where $t_{1/2}$ for substrate depletion is obtained by linear regression of the log percentage $[S]_0$ remaining v time to obtain a value for the slope (k, or elimination rate constant) (Fig. 4), since:

$$t_{1/2} = -Ln2/k$$

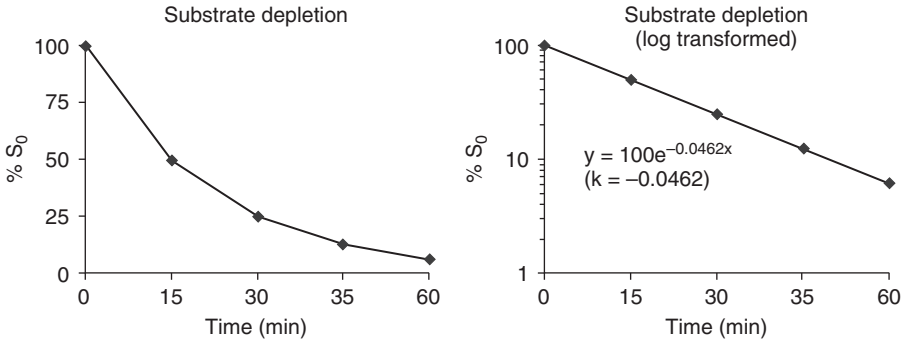


**Figure 4** Illustration of a substrate-depletion plot for $t_{1/2}$ measurement and calculation of intrinsic clearance. $S_0$ = initial substrate concentration.

To properly balance the units the full equation assumed is:

$$CL'_{int} = \frac{Ln2}{t_{1/2}(min)} * \frac{\text{volume incubation (mL)}}{\text{microsomal protein (mg)}}$$

Units for $CL'_{int}$ are usually expressed as mL/min/mg microsomal protein.

The substrate-depletion approach is convenient and useful in discovery, yet provides only a gross measure of metabolic clearance and represents the summed processes of all enzymes involved at the substrate concentration used. The advantage to Michaelis–Menten kinetics is that the role and potential for saturation of individual metabolic pathways can be determined and predictions made across a range of relevant drug concentrations and scenarios.

### 7.2 Scaling *In Vitro* Intrinsic Clearance to Predicted *In Vivo* Intrinsic Clearance

Liver microsomal data are scaled by applying an empirical factor to adjust CYP activity in liver microsomes to an equivalent of gram liver weight. Houston, at the University of Manchester, UK, reported that a numeric factor of "×45" gave a product of liver microsomal activity that improved the prediction of *in vivo* metabolic clearance, that is, 45 mg of liver microsomal protein contains the equivalent functional CYP activity of 1 g of liver (also referred to as mg microsomal protein per gram liver or MPPGL) (Houston, 1994). A similar concept is applied to hepatocytes and termed "hepatocellularity," which assumes approximately 3 million cells as equivalent to 1 mg of liver microsomal protein or approximately 120 million cells is equivalent in functional CYP activity to 1 g of liver. Recently, Barter *et al.* performed a meta-analysis that showed 95% confidence intervals for the geometric mean of 27–32 MPPGL and 74–131 million cells per g liver for hepatocellularity (Barter *et al.*, 2007).

One more factor to consider with respect to scaling of $CL'_{int}$ is liver weight. Values for liver weight are available for most species including mouse, rat, monkey, dog, and human (2, 10, 160, 320, and 1500 g, respectively) (Davies and Morris, 1993). Clearance is often represented per kg of body weight rather than per animal (i.e. per liver) and

for conversion, the generalized body weights assumed for mouse, rat, monkey, dog, and human are usually 0.025, 0.25, 5, 10, and 70 kg, respectively (Davies and Morris, 1993).

Putting all of this together, if Michaelis–Menten kinetics are measured for each metabolic pathway, one can scale *in vitro* $CL_{int}$ to an *in vivo* hepatic intrinsic clearance ($CL_{H,int}$) using:

$$CL_{H,int} = \frac{V_{max}}{K_m} * 45\,(\text{MPPGL}) * \frac{1500\,\text{g human liver}}{70\,\text{kg human body weight}}$$

If a substrate-depletion approach is taken, the apparent intrinsic clearance from the sum of all metabolic processes active at the substrate concentration tested will scale using:

$$CL_{H,int} = \frac{\text{Ln}2}{t_{1/2}(\text{min})} * \frac{\text{volume incubation (mL)}}{\text{microsomal protein (mg)}} * 45\,(\text{MPPGL}) * \frac{1500\,\text{g human liver}}{70\,\text{kg human body weight}}$$

Liver weight and body weight are substituted accordingly to similarly scale animal liver microsomal metabolic clearance data and correlate with the measured *in vivo* clearance. (If a positive correlation is not observed between *in vitro* and *in vivo* animal data, then additional mechanisms may be involved, e.g. non-CYP, non-microsomal, and/or extrahepatic metabolism.)

## 8 *IN VITRO* TO *IN VIVO* EXTRAPOLATION AND PHYSIOLOGICAL MODELS OF HEPATIC CLEARANCE

Intrinsic clearance represents the capacity of a particular enzyme to metabolize a specific substrate in the absence of physiological limitations. The amount of circulating drug presented to the liver enzymes and cleared from the blood depends on the rate of hepatic blood flow ($Q_H$). There are several models of hepatic blood flow available, with each assuming that hepatic clearance ($CL_H$) is perfusion rate-limited with only unbound drug entering the hepatocyte and available for metabolism and/or excretion (Wilkinson and Shand, 1975).

Possibly the simplest and most commonly used approach is the venous equilibrium or "well-stirred" model. This model assumes a rapidly distributed and uniform concentration of drug in the whole liver (i.e. no concentration gradient across the liver or cells):

$$CL_H = \frac{Q_H * (f_{ub} * CL_{H,int})}{Q_H + (f_{ub} * CL_{H,int})}$$

where $f_{ub}$ = fraction of drug unbound in blood. This can be reduced further to provide bioavailability ($F_H$):

$$F_H = \frac{Q_H}{Q_H + (f_{ub} * CL_{H,int})}$$

Other models that can be used include the sinusoidal or "parallel tube" model that assumes a concentration gradient across the liver:

$$CL_H = Q_H - Q_H * \exp(-f_{ub} * CL_{H,int} / Q_H)$$

There is also a dispersion blood flow model that accommodates a radial distribution of drug from blood through the liver:

$$CL_H = Q_H * \frac{1 - 4a}{(1+a)^2 * \exp[(a-1)/2Dn] - (1-a)^2 * \exp[-(a+1)/2Dn]}$$

where a = $(1 = 4Rn*Dn)^{1/2}$; Dn = 0.17 (dispersion number); Rn = $f_{ub}*CL_{H,int}/Q_H$ (efficiency number).

High and low hepatic extraction drugs show markedly different characteristics with respect to hepatic blood flow. Using the well-stirred model as an example and assuming no significant protein binding (i.e. $f_{ub} = 1$):

$$CL_H = \frac{Q_H * CL_{H,int}}{Q_H + CL_{H,int}}$$

With a high-clearance drug, the *in vitro* clearance greatly exceeds the flow of blood through the liver (i.e. $CL_{H,int} >> Q_H$), in which case the equation reduces to:

$$CL_H = Q_H$$

Consequently high-clearance drugs are hepatic blood flow limited, which means that increased expression of the enzymes involved in metabolism and/or an increase in the free fraction from protein displacement will alter pharmacokinetics and pharmacodynamics less than changes in hemodynamics such as blood flow.

In contrast, if the hepatic blood flow greatly exceeds the metabolic clearance (i.e. $Q_H >> CL_{int}$), then the equation reduces to:

$$CL_H = CL_{H,int}$$

CYP inducers, for instance, can profoundly increase intrinsic clearance and thus increase hepatic clearance and the dose required to achieve exposure for efficacy—if induction is sufficient to turn clearance from low to high, then hepatic blood flow will become the rate-limiting factor. Likewise, coadministered CYP inhibitors can alter the phenotype of a drug effectively converting high clearance to a low clearance, thus profoundly affecting the drug's AUC and effects *in vivo*. Inducers and inhibitors can thus change the phenotype in a patient resulting in a "phenocopy" that mimics a rapid metabolizer or slow metabolizer, respectively.

To predict the effect of metabolic clearance on oral bioavailability:

$$F = Fa * (1 - CL_{non\text{-}renal} / Q_H)$$

where F = bioavailability; Fa = fraction of dose absorbed; $CL_{non\text{-}renal}$ = any pathway not mediated by renal clearance, for example, metabolic, biliary, or chemical instability.

If metabolic clearance governs elimination, then this equation becomes:

$$F = Fa * (1 - CL_H / Q_H)$$

For a more detailed review of these methods, enzyme kinetics and the assumptions made, and examples of how they can be applied to the prediction of *in vivo* clearance, see Segel (1975), Wilkinson and Shand (1975), Houston (1994), Obach *et al.* (1997), and Obach (1999).

### 8.1 cDNA-Expressed Human Recombinant Enzymes

cDNA-transfected recombinant enzyme sources are another useful tool for drug reaction phenotyping. These are made by transfecting a lymphoblastoid, bacterial, or insect cell source with negligible basal CYP or other DME activity, with a vector containing cDNA for a particular human enzyme (recombinant CYPs, FMOs, and UGTs are all commercially available). The cDNA for human cytochrome P450 oxidoreductase (CYPOR) is also included as an essential redox partner to support CYP activity, with or without the cDNA for human cytochrome b5 (cyt b5). The transfected cells are grown in culture and used to prepare microsomes enriched with the specific recombinant enzyme, and used to investigate the role of a specific enzyme in the potential clearance and biotransformation of a drug.

The kinetics associated with CYP-mediated reactions are quite complex and not all substrates necessarily behave the same way. CYPOR and cyt b5 are accessory proteins and electron donors involved in the CYP catalytic cycle and should be included when using artificial constructs such as recombinant or reconstituted enzyme systems. CYPOR is essential for all CYP-mediated biotransformation reactions (see Fig. 1). Cyt b5 in contrast is more substrate-specific in its role and although usually involved in stimulating CYP-mediated reactions (Venkatakrishnan *et al.*, 2000), cyt b5 can also inhibit CYP-catalyzed reactions (Gorsky and Coon, 1986). It was also shown that apoprotein of cyt b5 was as effective as holoprotein in stimulating the reactions of several CYPs, with a role postulated for divalent cations in influencing the ability of substrates and cyt b5 to bind with CYP (Yamazaki *et al.*, 1995, 1996). Occasionally for these reasons (e.g. optimization of metabolic stability), assays will benefit from further evaluation of buffers, salts, and effect of redox partner concentration on drug biotransformation and turnover, and whether this improves the accuracy of prediction of *in vivo* clearance from *in vitro* data.

The same kinetic analyses that are done with liver microsomes can also be done with recombinant enzymes, for example, $K_m$ and $V_{max}$ determinations performed and/or a substrate-depletion kinetic approach as discussed previously. Theoretically, the recombinant enzyme should have the same $K_m$ as that of the native liver microsomal form, assuming no significant differences in protein binding and free fraction of test substrate. However, to use recombinant enzyme systems for *in vitro* to *in vivo* extrapolation of metabolic clearance, one needs to account for the relative expression of those enzymes in the human liver and other tissues of concern. For reference, Table 8 lists the range and mean values reported for most of the major CYPs measured in microsomes prepared from human liver, small intestine, and skin. There is a significant degree of variability evident in the values reported by the laboratories referenced, which presumably reflects the variability in the donors themselves, as well as differences in the methods used to prepare tissue microsomes and quantify the expression of holoprotein.

**TABLE 8 Immunoquantified expression of different cytochrome P450 (CYP) forms in the human liver, intestine, and skin microsomal fractions.**

| | Liver[a] | | | | Intestine[b] | | Skin[c] | |
|---|---|---|---|---|---|---|---|---|
| | Range (pmol/mg-protein) | | Range R.A. (%) | | | | | |
| CYP | Min | Max | Min | Max | Range (pmol/mg-protein) | Mean R.A. (%) | Mean (pmol/mg-protein) | R.A. (%) |
| CYP1A1 | | | | | 3.6–7.7 | 7.4 | 3.6 | 16 |
| CYP1A2 | 19 | 67 | 7.5 | 13 | ND | ND | | |
| CYP2A6 | 14 | 68 | 5.5 | 13 | BLD | BLD | | |
| CYP1B1 | | | | | | | 2.0 | 9 |
| CYP2B6 | 1 | 45 | 0.4 | 8.4 | BLD | BLD | 0.035 | 0.2 |
| CYP2C8 | 12 | 64 | 4.5 | 12 | ND | ND | | |
| CYP2C9 | 50 | 96 | 20 | 18 | 2.9–27 | 11 | | |
| CYP2C19 | 8 | 20 | 3.1 | 3.7 | <0.6–3.9 | 1.3 | | |
| **Total CYP2C** | **60** | **64** | **24** | **12** | | | | |
| CYP2D6 | 5 | 11 | 2 | 2.1 | <0.2–3.1 | 0.66 | | |
| CYP2E1 | 22 | 52 | 8.6 | 9.8 | BLD | BLD | 11 | 49 |
| CYP2J2 | | | | | <0.2–3.1 | 1.2 | | |
| CYP3A4 | 37 | 108 | 15 | 20 | 8.8–150 | 57 (variable) | | |
| CYP3A5 | 1 | 117 | 0.4 | 22 | 4.9–25 | 21 (variable) | 5.6 | 25 |
| **Total CYP3A** | **96** | **262** | **38** | **49** | **Highly variable** | | | |
| **Total pmol CYP/mg** | **255** | **534** | | | | | | |

[a]The ranges shown are based on the mean values reported, based on data from Shimada *et al.* (1994), Rodrigues (1999), Snawder and Lipscomb (2000), Venkatakrishnan *et al.* (2000), Rowland-Yeo *et al.* (2003), and Paine *et al.* (2006).
[b]Paine *et al.* (2006).
[c]Bergström *et al.* (2007).
R.A. = relative abundance; BLD = below limits of detection; ND = not detected.

These values can be used to scale recombinant enzyme activity using the methods described as follows. The recombinant enzymes are incubated at a known concentration of CYP which provides a rate of reaction (typically pmoles metabolized/min/pmol CYPi). Michaelis–Menten kinetics can be measured for each individual enzyme (CYPi) and biotransformation pathway, and summed to give a total rate of metabolism at a specific concentration of substrate:

$$V_{total} = \sum_{i=1} \frac{V_{maxi} * S}{K_{mi} + S}$$

Using rates or apparent intrinsic clearance, the nominal absolute expression per mg of human microsomal protein can be factored to provide a microsomal equivalent that, if all the CYPs are functionally active and expressed at levels corresponding to those published (Table 8), will sum to give a total intrinsic clearance that is theoretically equivalent to liver microsomal intrinsic clearance:

$$V_{HLM} = \sum_{i=1}^{n} A_i * V_{rCYPi}$$

where $V_{HLM}$ = predicted rate of metabolism in HLM; $V_{rCYPi}$ = rate of metabolism in CYPi; $A_i$ = abundance of CYPi.

***8.1.1 Relative Abundance or Total Normalized Rate (TNR)*** The simplest approach to scaling rate of metabolism in recombinant human enzymes to an *in vivo* equivalent is to assume that the rate of metabolism by each molecule of recombinant CYP incubated can be scaled by its expression in a human liver (as listed in Table 8). This relative abundance (Nakajima *et al.*, 1999; Soars *et al.*, 2003) or total normalized rate (TNR) (Rodrigues, 1999) assumes in effect that the recombinant enzyme is identical in activity to the native enzyme. If correct, the relative contribution of a specific enzyme to the overall metabolism in HLMs can be calculated quite easily:

$$\%\ \text{Contribution CYPi}_{HLM} = \frac{Ai * V_{rCYPi}}{V_{HLM}} * 100$$

This can also be expressed as the fraction metabolized by a specific CYP ($f_{m.CYP}$):

$$f_{m.CYP} = \frac{Ai * V_{rCYPi}}{V_{HLM}}$$

which assumes that all of the metabolism is governed by CYPs expressed in the human liver microsomes.

***8.1.2 Relative Activity Factors (RAFs)*** In reality, the $k_{cat}$ of the recombinant enzyme can differ markedly from that of the native enzyme form. This was recognized early on in the development and use of recombinant enzymes for drug metabolism reactions and a relative activity factor (RAF) was proposed to make allowance for this disconnect (Crespi, 1995).

The RAF is determined for each individual enzyme by measuring the kinetics (e.g. $V_{max}$, intrinsic clearance) of a prototypic substrate, that is, one known to be selective for the enzyme of interest. The substrate, for example midazolam, is incubated with HLM and, in this case, also with the recombinant human CYP3A4 (rCYP3A4). Under equivalent conditions for CYP content and unbound concentrations of substrate, and assuming no difference in substrate affinity (e.g. $K_m$), rCYP3A4 should demonstrate the same substrate-dependent rate of metabolism as CYP3A4 in HLMs. If there is a difference between them in relative activity, it is presumed to result from a difference between the recombinant and native forms in the turnover number, necessitating a correction factor for relative activity (the "RAF"):

$$RAF_{CYPi} = \frac{\text{Rate of index reaction CYPi in HLM}}{\text{Rate of index reaction in rCYPi}}$$

This assumes the $K_m$ of midazolam for rCYP3A4 is the same as liver CYP3A4, and that no correction for free fraction is needed.

It is useful when using a specific batch of recombinant enzymes and liver microsomes for reaction phenotyping to assume that any difference in the turnover kinetics of a prototypic substrate will apply equally to all other substrates of that CYP. Consequently, one need only measure the kinetics of a suitable reference compound once per batch of CYPi to obtain a RAF value that can be applied to scale and extrapolate the metabolic clearance of any other drug or compound that is incubated in the same batch of rCYPi.

When the relative abundances of the CYPs involved are applied, the sum of individual reactions in recombinant enzymes should equal that measured in HLMs:

$$V_{HLM} = \sum_{i=1}^{n} Ai * (V_{CYPi} * RAF_{CYPi})$$

When substrates are relatively stable (i.e. showing slow rates of turnover), Michaelis–Menten kinetics of the rate of formation of metabolite(s) is usually the best and possibly only way to get an accurate prediction of clearance (as opposed to measuring the loss of parent substrate). However, if the compound is rapidly and extensively metabolized, then intrinsic clearance can be determined instead using substrate-depletion kinetics ($S < K_m$).

The RAF adjusted rate of reaction by a specific CYP in HLMs is calculated as follows:

$$V_{HLM.CYPi} = A_i * (V_{CYPi} * RAF_{CYPi})$$

The contribution of that CYP to the overall rate of metabolism in HLMs is then calculated:

$$\%\ \text{Contribution of CYPi} = \frac{V_{HLM.CYPi}}{V_{HLM}} * 100$$

The relationship between relative abundance, RAF, and the fraction metabolized by a specific CYP can then be represented by:

$$f_{m.CYPi} = \frac{Ai * (V_{CYPi} * RAF_{CYPi})_i}{V_{HLM}}$$

In practicality, the relative abundance or TNR approach is very useful, especially in early drug discovery and in the preliminary stages of reaction phenotyping. For single high-clearance pathways, it is often all that is needed, often corroborated easily by CYP-selective inhibition studies. However, if multiple enzymes show significant turnover of test substrate, and when differentiation and accuracy of prediction assumes greater significance (e.g. a potential CYP polymorphic clearance and a relatively narrow safety margin), the more laborious RAF method becomes the preferred and definitive approach.

To illustrate how these approaches work, Fig. 5 shows four hypothetical scenarios that can be observed *in vitro* with recombinant CYPs. In the first example (Fig. 5a), with "compound A," all of the rCYPs show the same extent and rate of turnover. This compound will therefore scale by the relative abundances of those enzymes in human tissues (assuming no kinetic correction is needed for $k_{cat}$ and/or $K_m$). Using the relative abundance of CYPs reported in the liver (Shimada *et al.*, 1994; Lasker *et al.*, 1998), the small intestine (Paine *et al.*, 2006), and skin (Bergström *et al.*, 2007), this hypothetical compound is predicted to have the reaction phenotype shown in the bar plots. In the liver, CYP3A is predicted to be the major CYP, with significant contributions to hepatic metabolism expected from 2C9, 1A2, and 2E1. Although CYP3A4 is again predicted to be the major contributor to metabolism in the intestine, the differential expression of CYPs in this tissue predicts a significant role for CYP3A5, 2C9 and 1A1, and liver and intestinal CYP DDIs should be limited in effect. In contrast, in the skin, CYP2E1 is predicted to be the major CYP, with contributions from CYP3A5, 1A1, and 1B1 to the metabolism of compound A in keratinocytes.

In the second example (Fig. 5b), which is the most commonly encountered scenario, especially for large lipophilic molecules, the greatest turnover of "compound B" was measured in CYP3A4. In this particular example, compound B is simulated to have 20% turnover in the CYPs. With compound B, CYP3A4 is predicted to dominate its metabolic clearance in liver and intestine; however, with negligible CYP3A4 expression in the keratinocyte, its reaction phenotype in the skin is predicted similar to that for compound A. A drug with the characteristics of compound B is likely to be a potential victim of CYP3A4 DDIs.

In the third example (Fig. 5c), "compound C" is most rapidly metabolized by the recombinant CYP2D6 enzyme. In this example, the other CYP enzymes again show significant turnover and the contribution of CYP2D6 depends on its relative expression. Since CYP2D6 protein is expressed in liver microsomes at only about 2% of the total CYP content, whereas CYP3A4 is more like 30% of the total, only a minor role for CYP2D6 is predicted in the metabolic clearance of this hypothetical drug.

In the fourth simulated example (Fig. 5d), recombinant human CYP1A1 most efficiently metabolized "compound D." CYP1A1 is not constitutively expressed in human liver; therefore, the hepatic clearance cannot be scaled accurately to predict *in vivo* clearance. CYP1A1 in the gut and skin is likely to have a major role in extrahepatic metabolism, which could prove significant, for instance, if reactive or toxic and/or active metabolite(s) are formed. With CYP1A1 known to be highly inducible in the liver, the lung and other tissues, and assuming that metabolic clearance accounts for most, if not all, of the total clearance, it is reasonable to expect considerable interindividual variability in a human population dosed with a compound possessing a phenotype such as this. If drug safety could be affected by such a pronounced interindividual variability, then additional studies may be warranted.

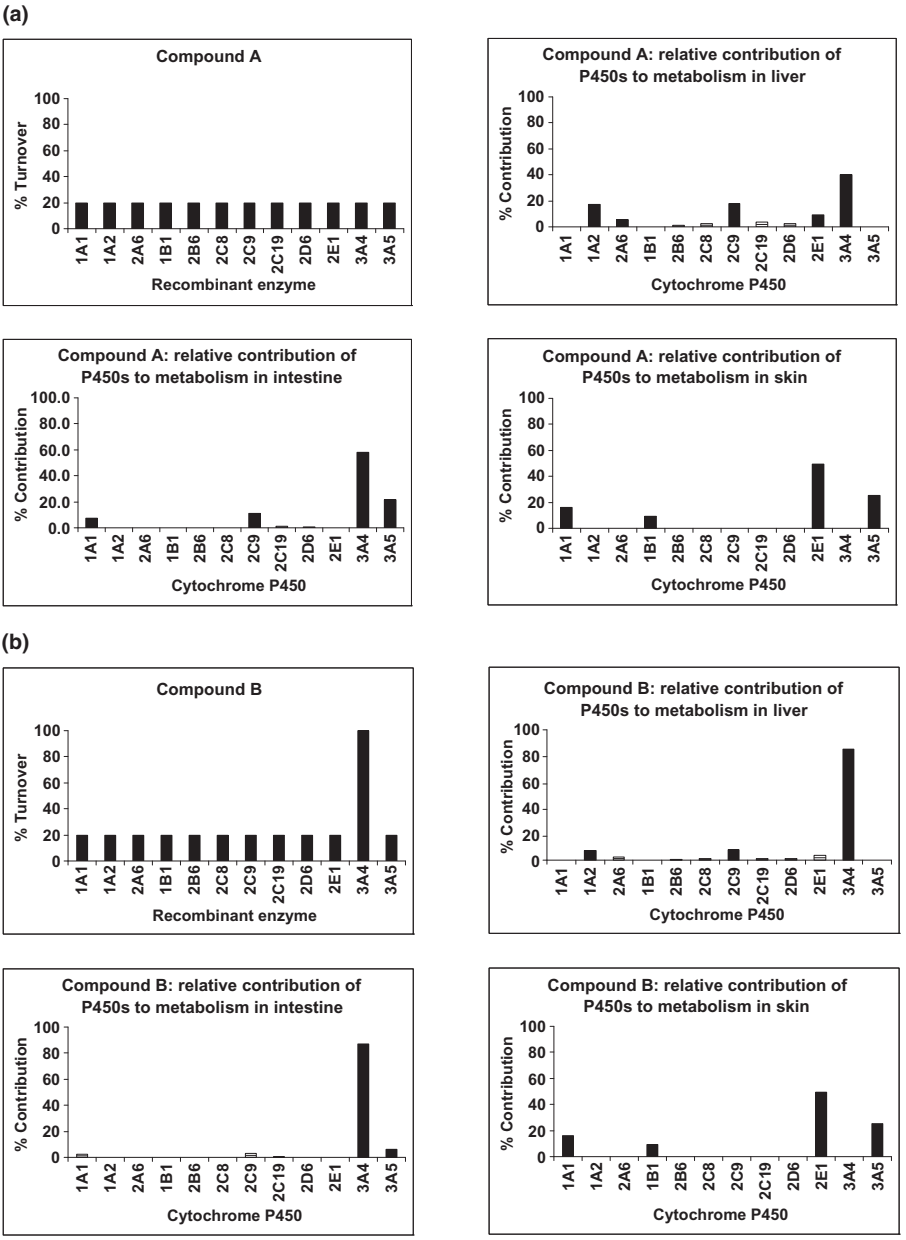


**Figure 5** **(a)** Scaling of human recombinant cytochrome P450 (CYP) turnover using relative abundance in the liver, intestine and skin. Hypothetical representation: a theoretical compound that is metabolized to an equal extent in human recombinant enzymes. The relative expression of those enzymes varies between liver, intestine, and skin and affects the relative contribution of each enzyme to the total metabolism in that tissue. **(b)** Scaling of human recombinant CYP turnover using relative abundance in the liver, intestine and skin. Hypothetical representation: a theoretical compound that is metabolized rapidly and extensively in human recombinant CYP3A4 and to an equal extent in the other human recombinant enzymes. The relative expression of those enzymes varies between liver, intestine, and skin and affects the relative contribution of each enzyme to the total metabolism in that tissue.

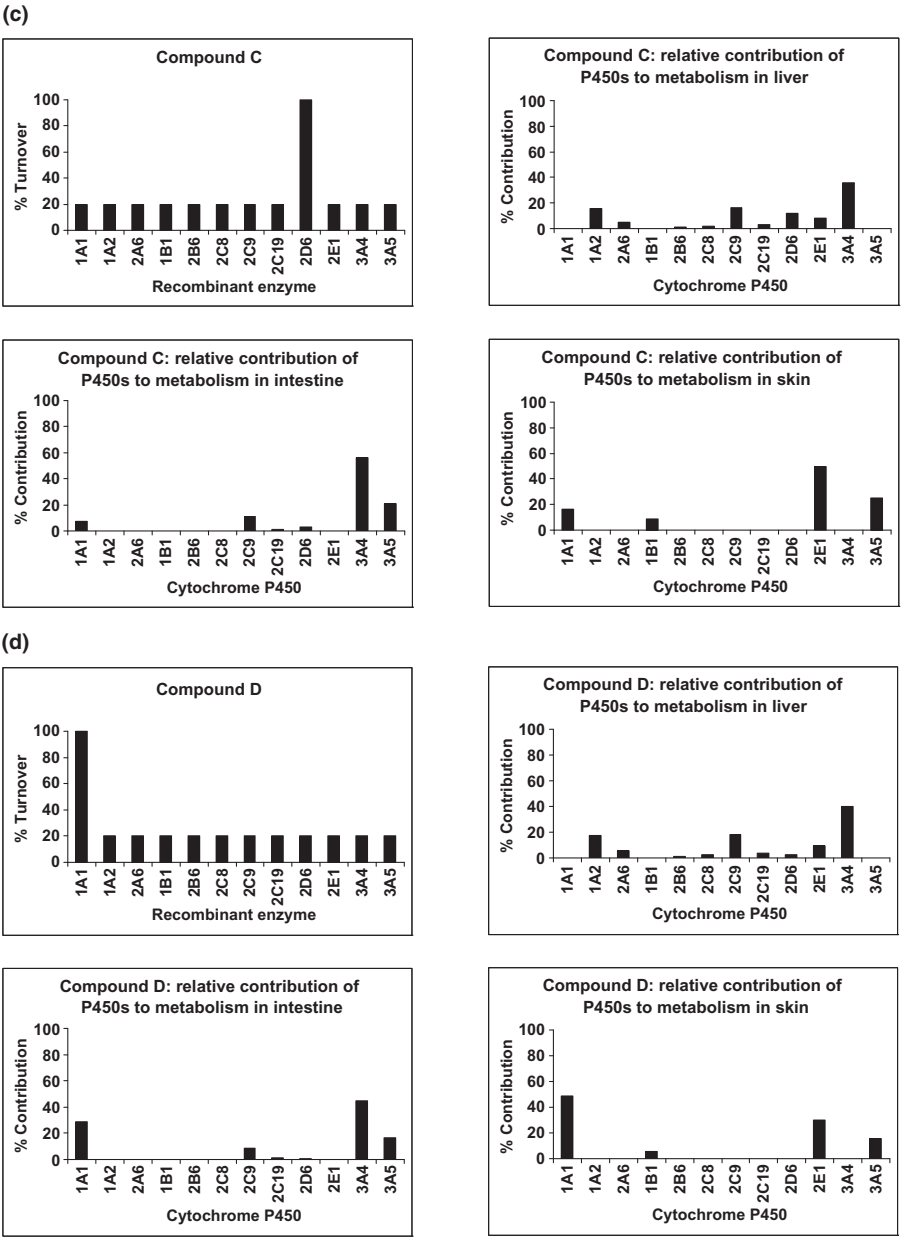


**Figure 5** **(c)** Scaling of human recombinant CYP turnover using relative abundance in the liver, intestine, and skin. Hypothetical representation: a theoretical compound that is metabolized rapidly and extensively in human recombinant CYP2D6 and to an equal extent in the other human recombinant enzymes. The relative expression of those enzymes varies between the liver, intestine, and skin and affects the relative contribution of each enzyme to the total metabolism in that tissue. **(d)** Scaling of human recombinant CYP turnover using relative abundance in the liver, intestine, and skin. Hypothetical representation: a theoretical compound that is metabolized rapidly and extensively in human recombinant CYP1A1 and to an equal extent in the other human recombinant enzymes. The relative expression of those enzymes varies between the liver, intestine, and skin and affects the relative contribution of each enzyme to the total metabolism in that tissue.

The example of "compound D" also illustrates some issues that apply to DMEs—including hepatic, extrahepatic, CYP, and non-CYP enzymes (e.g. UGTs)—that are not as well defined in their relative expression and thus not so readily scaled to predict their role in the *in vivo* metabolism. It is clear that relatively minor enzymes can scale quite differently between different human tissues. For most orally administered drugs, this is unlikely to be relevant. However, if different routes of administration are part of the drug's life cycle management, then tissue-specific metabolism could well become a concern, especially if these enzymes mediate the formation of CYP-specific metabolites and/or demonstrate tissue-specific toxicity.

These hypothetical scenarios assume that the recombinant and native forms are essentially identical with the same intrinsic clearance *in vitro*. If that is not the case, a RAF is needed to correct the predicted rate. Figure 6 uses the example of compound C from Fig. 5c to illustrate the potential effect of RAFs on the predicted relative contribution of CYP2D6 to the total hepatic metabolic clearance. In Fig. 6a, compound C showed 20% turnover in each of the recombinant CYPs except for CYP2D6 where it is 100%; however, in this case, an index substrate in a separate theoretical incubation showed that the rate of metabolism of CYP2D6 in liver microsomes was much greater than that with an equivalent concentration of the same batch of recombinant CYP2D6. In this case, a TNR approach will be wrong and a correction factor must be used. In this example, a RAF of ×10 was applied and CYP2D6 is now predicted to be the major contributor to metabolic clearance in the liver. In this scenario, CYP2D6 polymorphisms could become a major issue in the development of this compound as a drug. In Fig. 6b, the same compound is used to illustrate the opposite scenario in which the index reaction for CYP2D6 showed hypothetically much higher rates of metabolism in the same batch of recombinant CYP2D6. A correction factor is now needed to avoid overestimating the contribution of this enzyme to the total metabolic clearance. Applying a RAF of 0.1 for the recombinant enzyme results in CYP2D6 predicted to have no significant role in the metabolic clearance of compound C.

These are relatively simple but realistic illustrations that serve to demonstrate the potential complexity involved in reaction phenotyping. If the kinetics are not performed or applied correctly, there is a real possibility in making a wrong prediction. Rather than rely on any single method, instead it is recommended that multiple assays be used for reaction phenotyping, and HLM CYP inhibition studies and/or correlation analysis should be used to corroborate recombinant CYP assays, or vice versa, or help identify any anomalies. These hypothetical examples also serve to illustrate that differences may exist locally in the metabolism of a drug, which could prove particularly significant if the route of administration is not oral (e.g. transdermal, intranasal, and inhalation). This can prove especially significant if metabolites are formed that are specific to the enzymes expressed locally to the site of administration.

In Table 9 are shown the results of an integrated reaction phenotype for etoricoxib, ABT-761, zileuton, naproxen, clarithromycin, and celecoxib (see Zhang *et al.*, 2007 and references therein). In each case, a specific reaction was monitored using a panel of HLM from at least 10 donors, HLM incubated with and without selective CYP inhibitors, and results from incubations with recombinant human CYPs scaled using a total normalized rate to determine a relative contribution of each to the total hepatic clearance. Knowing the relative fraction metabolized ($f_m$) and the contribution from each CYP ($f_{m.CYP}$) to the metabolism of that drug, one can predict the potential effect of a coadministered inhibitor on its relative exposure.

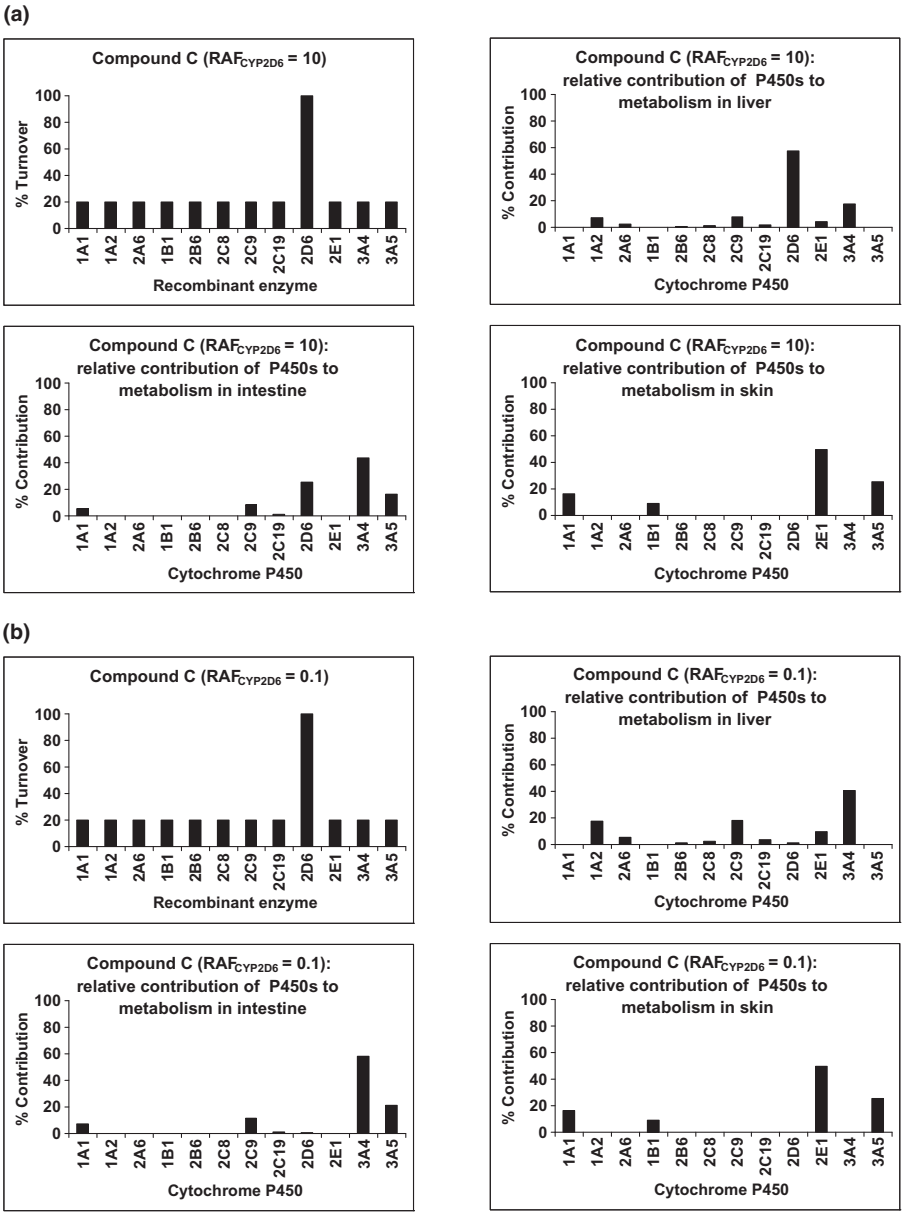


**Figure 6** **(a)** Effect of RAFs on the scaling of human recombinant cytochrome P450 (CYP) turnover. Hypothetical representation: a theoretical compound that is metabolized rapidly and extensively in human recombinant CYP2D6 and to an equal extent in the other human recombinant enzymes. A RAF correction is needed since the relative activity of CYP2D6 was greater in the native form than in the recombinant form. The relative expression of those enzymes varies between the liver, intestine, and skin and affects the relative contribution of each enzyme to the total metabolism in that tissue. **(b)** Effect of RAFs on the scaling of human recombinant CYP turnover. Hypothetical representation: a theoretical compound that is metabolized rapidly and extensively in human recombinant CYP2D6 and to an equal extent in the other human recombinant enzymes. A RAF correction is needed since the relative activity of CYP2D6 was lower in the native form than in the recombinant form. The relative expression of those enzymes varies between the liver, intestine, and skin and affects the relative contribution of each enzyme to the total metabolism in that tissue.

**TABLE 9 Examples of an integrated *in vitro* cytochrome P450 (CYP) reaction phenotype: major CYP(s) shown for each reaction.**

| Compound | Reaction | CYP form(s) | Human liver microsomes: Correlation coefficient ($r^2$)[a] | Human liver microsomes: % Inhibition[b] | % Total normalized rate (rCYPs)[c] |
|---|---|---|---|---|---|
| Etoricoxib | Hydroxylation | 3A4 | 0.64 | 65 | 58 |
| | | 1A2 | <0.40 | 5.0 | 2.0 |
| | | 2C9 | <0.40 | 10 | 6.0 |
| | | 2D6 | <0.40 | 10 | 26 |
| ABT-761 | Hydroxylation | 3A4 | 0.67 | 55 | 66 |
| | | 2C9 | 0.40 | 20 | 19 |
| Zileuton | Sulfoxidation | 3A4 | 0.67 | 60 | 95 |
| | Hydroxylation | 1A2 | 0.58 | 45 | 57 |
| Naproxen | *O*-demethylation | 2C9 | 0.67 | 50 | 56 |
| | | 1A2 | 0.46 | 55 | 33 |
| Clarithromycin | *N*-demethylation | 3A4 | 0.81 | >95 | 100 |
| | Hydroxylation | 3A4 | 0.85 | >95 | 100 |
| Celecoxib | Hydroxylation | 2C9 | 0.92 | 80 | 86 |
| | | 3A4 | 0.55 | 20 | 6.0 |

[a]Reaction rates were measured in a panel of human liver microsomes from at least 10 different organ donors. Rates were correlated with various CYP form-selective activities and a correlation matrix setup.
[b]Reaction rates were measured in the presence of human liver microsomes co-incubated with solvent or CYP form-selective inhibitor. Inhibition expressed as % loss of activity in the presence of solvent alone.
[c]Reaction rate (turnover) was determined after incubation with a panel of individual expressed (recombinant) CYP proteins (rCYPs). Turnover for each CYP form was then normalized as described in Rodrigues (1999).
Adapted from Zhang *et al.*, 2007 and references therein.

## 8.2 Significance of Metabolic Clearance and Reaction Phenotype

There are numerous examples in which drug metabolism is implicated in toxicity through either impaired clearance in the case of polymorphic mechanisms or increased exposure in the case of reactive or toxic metabolites. Knowing the enzymes involved can help understand the mechanism of toxicity, which can optimize the evaluation of a prospective backup compound, and identify individual susceptibility to such toxicity, which can optimize the design of a clinical study with screening and monitoring of the patient population. For instance, drugs metabolized primarily by CYP2D6, CYP2C19, and CYP2C9 are likely to be subject to significant interindividual variability in their exposure and effects as a result of known genetic polymorphisms that affect the relative expression of these enzymes in humans. Nonetheless, the drugs can still be very useful, for example, fluoxetine (CYP2D6), omeprazole (CYP2C19), and warfarin (CYP2C9); however, close attention must be paid to the safety margins and the feasibility of using genotyped/phenotyped patients in the clinical evaluation and even the selection criteria used to enroll patients into a clinical study (e.g. the CYP2C19 slow metabolizer phenotype is much more prevalent in Asians than in Caucasians) (Goldstein *et al.*, 1997; Shimoda *et al.*, 1999; Yokono *et al.*, 2001; Mizutani, 2003; Niu, Luo and Hao 2004; Bertilsson, 2007).

In general, the consequences of higher metabolic clearance are tied to the potential effect of coadministered inhibitors on the total exposure of the new drug. Figure 7

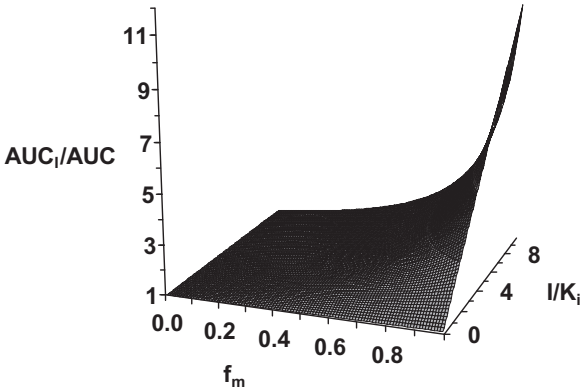


**Figure 7** Simulation of fraction metabolized, potency of inhibition, and predicted effect on exposure: $f_m$ = fraction metabolized; AUC = area under the curve (with and without inhibitor); I = concentration of inhibitor; and $K_i$ = inhibitor constant.

illustrates the relationship between fraction metabolized ($f_m$) and the effect of inhibition ($I/K_i$) on the total exposure of the drug (i.e. the ratio of area under the curve, with and without the inhibitor—$AUC_I/AUC$). Drugs that are not cleared significantly by metabolism (e.g. $f_m < 0.3$) will not suffer much when coadministered with an inhibitor of that enzyme, regardless of how potent the inhibitor is or the concentration at which it is given or attains. With a low-metabolic-clearance drug, its AUC will be very similar, if not identical, with or without the inhibitor (i.e. $AUC_I/AUC = 1$). In contrast, a drug that is cleared entirely by metabolism ($f_m = 1$) will be markedly affected by inhibition of its clearance, with a several-fold increase in AUC possible.

The relationship between the fraction of drug metabolized and the exposure and potency of a coadministered inhibitor and the predicted effect on the drug AUC of its exposure can be calculated using the following formula:

$$\frac{AUC_i}{AUC} = \frac{1}{\dfrac{f_m}{1+[I]/K_i} + (1 - f_m)}$$

where $f_m$ = fraction metabolized; I = concentration of inhibitor, for example plasma; $K_i$ = inhibition constant; AUCi = area under the curve with the inhibitor; AUC = area under the curve without the inhibitor.

To be even more specific, the contribution of a particular CYP to the total metabolic clearance can be indicated by $f_m.f_{mCYP}$, where $f_{mCYP}$ is the fraction metabolized by each CYP form, which substitutes into the equation to give:

$$\frac{AUC_i}{AUC} = \frac{1}{\dfrac{f_m \cdot f_{mCYP}}{1+[I]/K_i} + (1 - f_m \cdot f_{mCYP})}$$

If there is only one enzyme involved, then $f_m = f_m.f_{mCYP}$; however, if more enzymes are involved, then selective inhibition is only going to block part of the clearance pathway

and the total effect on AUC will be less. Consequently, a drug with a narrow therapeutic index that is metabolically cleared by several CYPs is considered, all else being equal, safer than a drug cleared principally by only one CYP. It is possible to determine $f_m$ after the completion of a radiolabeled study, at which point the dose has been recovered and excreta profiled for parent drug, phase I (oxidative) and phase II (conjugates) metabolism (Zhang *et al.*, 2007).

The effect of inhibitors is compounded further for drugs that are metabolized significantly by enzymes in both the intestine and liver (e.g. CYP3A4 and 3A5). In addition to its effect on hepatic clearance, the inhibitor will now significantly affect the amount of drug available for absorption from the intestine. The subsequent change in AUC can be predicted using the following equation:

$$\frac{AUC_i}{AUC} = \frac{F_{Gi}}{F_G} * \frac{1}{\frac{f_m}{1+[I]/K_i} + (1 - f_m)}$$

where $F_{Gi}$ = fraction unmetabolized available from the gut in the presence of inhibitor; $F_G$ = fraction unmetabolized and available from the gut without an inhibitor.

In Fig. 8 are simulations for $F_{Gi}/F_G$ ranging from 1 (i.e. no effect on gut metabolism) to 10 (i.e. a 10-fold increase in the amount of drug absorbed). The effect on AUC can be exponential if the inhibitor also knocks out the hepatic clearance and $f_m$ is high (e.g. $f_m > 0.7$).

By knowing the enzymes involved in metabolic clearance and their relative contribution, one can simulate the potential effect of inhibition on exposure and safety. The relative expression of CYPs differs from one tissue to another and this can be another major consideration when predicting oral bioavailability and/or the effects of DDIs. For example, a drug that has a high metabolic clearance and is selective for CYP3A4/5 is very likely to demonstrate a DDI given the relative expression of CYP3A in the gut and the liver (e.g. the CYP3A inhibitors itraconazole produced a greater than 10-fold increase observed in the $C_{max}$ and AUC of simvastatin, a known selective CYP3A4 substrate with high first-pass metabolism) (Neuvonen, Kantola and Kivistö, 1998).

The importance of this factor is also illustrated in examples of polymorphic expression that results in reduced $f_m$*$f_{mCYP}$, which can lower the oral clearance and alter the susceptibility to interactions from CYP-selective inhibitors. For instance, the CYP2C9*3 allelic variant can result in a lower metabolic clearance of CYP2C9 substrates such as flurbiprofen (Kumar *et al.*, 2006). This is demonstrated *in vivo*, with a recent study reporting a gene-dependent effect on the apparent oral clearance of flurbiprofen in healthy volunteers. Specifically, the CYP2C9 inhibitor, fluvoxamine (400 mg oral dose), showed only a 1.1- and 1.4-fold change in the AUC of flurbiprofen (50 mg oral dose) in the two subjects identified as having the CYP2C9*3/*3 genotype. In comparison, a 2.4-fold change in the median AUC of flurbiprofen was measured in subjects with the CYP2C9*1*3 genotype (n = 8), and a threefold change in subjects with the CYP2C9*1*1 wild-type form (n = 11). Although the subject numbers are admittedly small for the CYP2C9*3*3 group, the authors concluded that, by analysis of fractional clearances, the degree of drug interaction observed was determined by genotype-related differences in the fraction metabolized by CYP2C9 (Kumar *et al.*, 2008).

Further illustration of the role of fraction metabolized in the clearance and AUC is provided by the significance of DDIs elicited via induction of clearance. Induction of

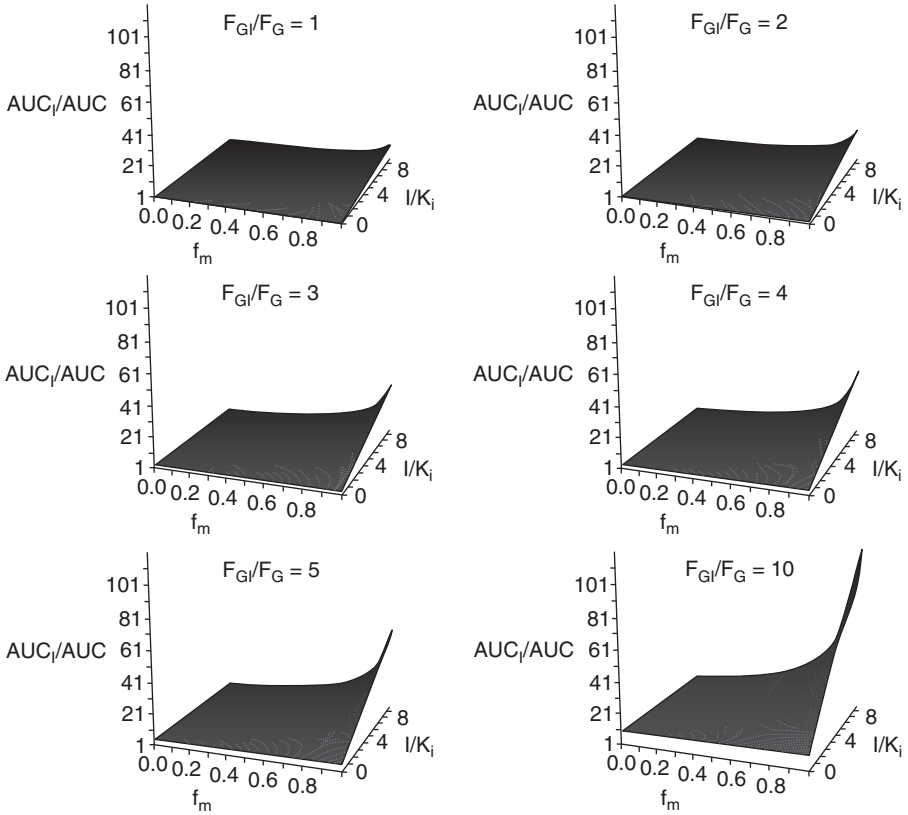


**Figure 8** Simulation of gut first pass, fraction metabolized, potency of inhibition, and predicted effect on exposure: $f_m$ = fraction metabolized; AUC = area under the curve (with and without inhibitor); I = concentration of inhibitor; and $K_i$ = inhibitor constant; $F_G$ = fraction unmetabolized and available from the gut (with and without inhibitor).

CYPs, for instance, effectively can lead to an increase in the $f_m$ to the point that metabolic clearance can become the major contributor to the total clearance of a drug. In such circumstances, drugs can become markedly susceptible to the effects of coadministered inhibitors with the result that profound changes in exposure become possible.

## 8.3 Advanced Reaction Phenotyping

The methods described earlier can be applied to each individual metabolic pathway involved and can be used to predict the relative contribution of each enzyme in the formation of a specific metabolite from a particular concentration of drug. The antidepressant mirtazapine was studied by Störmer *et al.* (2000) and Störmer, von Moltke and Greenblatt (2000), with the 8-hydroxylation, *N*-desmethylation, and *N*-oxidation pathways evaluated. Table 10 summarizes the results of these studies and shows how the contribution of each CYP differs subject to the metabolite formed. In these studies, Störmer *et al.* showed that the formation of the 8-hydroxy metabolite was mediated by CYP2D6, 1A2, 3A4, and 2C9, with a pronounced trend to move away from CYP2D6

**TABLE 10 Examples of *in vitro* cytochrome P450 (CYP) reaction phenotyping of the major biotransformation pathways involved in the metabolism of mirtazapine.**

| | Mirtazapine | | |
|---|---|---|---|
| | % Predicted contribution | | |
| Mirtazapine 8-hydroxylation | 2.5 μM | 25 μM | 250 μM |
| CYP1A2 | 30 | 50 | 55 |
| CYP2C8 | — | — | — |
| CYP2C9 | — | — | 10 |
| CYP2D6 | 65 | 40 | 15 |
| CYP3A4 | 5 | 10 | 20 |
| Mirtazapine *N*-desmethylation | | | |
| CYP1A2 | 45 | 20 | 10 |
| CYP2C8 | <5 | 5 | 15 |
| CYP2C9 | — | — | — |
| CYP2D6 | — | — | — |
| CYP3A4 | 50 | 65 | 75 |
| Mirtazapine *N*-oxide | | | |
| CYP1A2 | 85 | 45 | 15 |
| CYP2C8 | — | — | — |
| CYP2C9 | — | — | — |
| CYP2D6 | — | — | — |
| CYP3A4 | 15 | 55 | 85 |
| Biotransformation pathways | | | |
| Mirtazapine 8-hydroxylation | 55 | 45 | 30 |
| Mirtazapine *N*-desmethylation | 45 | 50 | 55 |
| Mirtazapine *N*-oxidation | <5 | 5 | 15 |

Adapted from Störmer *et al.* (2000) and Störmer, von Moltke and Greenblatt (2000).

at higher concentrations *in vitro*. The formation of the *N*-desmethyl metabolite of mirtazapine was mediated by CYP3A4, 1A2, and 2C8, with CYP3A4 becoming dominant at higher concentrations *in vitro*. The third major metabolite measured, the *N*-oxide, was mediated by CYP3A4 and 1A2, with a move again toward CYP3A4 dominating at higher concentrations *in vitro*. These studies also demonstrated how the relative formation of the metabolites can change, subject to the concentration presented to the metabolizing enzymes; in this case, *N*-desmethylation of mirtazapine became the major pathway at 250 μM. This serves as an excellent example of how reaction-phenotyping methods can be applied to predict *in vivo* biotransformation and kinetics, and illustrates how the biotransformation of a drug depends on the concentration at which it is assessed.

It is recognized that although *in vitro* studies are very useful, especially in the discovery and early development phases, they lack the dynamic environment and complexity offered by real organs and tissues. If, for example, a novel drug candidate has demonstrated a relatively narrow safety margin in preclinical studies, then it follows that variability in its exposure becomes a significant factor in how well the drug will be tolerated and/or efficacious in patients. However, if the exposure can be reasonably well controlled, and the drug has the potential to address a significant unmet medical need, then reaction phenotyping *in vivo* can be useful in helping to predict or define

the likely variability in exposure and risks to the target or general population. For instance, the oral bioavailability of a drug in humans can be measured with and without coadministration of inhibitors such as ketoconazole, itraconazole, or grapefruit juice, which allows one to determine under controlled clinical conditions the real potential for CYP3A4-mediated interactions. Another approach that can be taken is to genotype patients for known variants in CYPs, such as CYP2C9 or CYP2D6, to identify any genetic predisposition toward a high or low clearance of the drug.

Reaction phenotyping extends beyond the DMEs and tissues discussed in this chapter. When adverse effects are seen in preclinical studies, knowing the enzymes involved and their relative contribution to the observed toxicity can prove insightful and help to optimize a chemical series and find a safer molecule. Occasionally, especially when a specific metabolite or fragment is involved, it may even be possible to use reaction-phenotyping studies to define the relevance and risk to patients of a species- or gender-specific toxicity observed in animal models. (Chapter 21 addresses these concepts and application in more detail.)

## 9 SUMMARY

Optimizing the metabolic stability of a new chemotype in drug discovery can improve oral bioavailability and help to identify a drug candidate that will achieve efficacious exposure in the target patient population. In drug discovery, the goal is to identify the compound with the best balance of properties and develop the one that offers the least risk going into clinical trials. By identifying the specific enzymes involved in a drug's metabolic clearance, one can identify the risk of polymorphic variability and the risk of high-magnitude DDIs.

The methods discussed can be applied systematically in drug discovery from a higher-throughput early stage, in which the goal is to eliminate the compounds with the highest possible risk in development, through more predictive methods that enable forecasting of clinical outcomes. Irrespective of the stage of development, however, one has to leverage existing tools (e.g. selective inhibitors and recombinant enzymes) and acquire high-quality *in vitro* CYP reaction phenotype data. Such data are essential when attempting to obtain a fully integrated reaction phenotype, enable decision making, submit data for regulatory review prior to first-in-man, and when prioritizing clinical drug interactions studies, human radiolabel studies, or PK studies with CYP genotyped (phenotyped) subjects.

At the present time, the reaction-phenotyping "tool kit" for other DMEs (e.g. SULTs, UGTs, and FMOs) is poorly developed and additional progress is needed. In most cases, reaction phenotype has been rudimentary and based largely on turnover with recombinant enzyme preparations (Schrag *et al.*, 2004; Bowalgaha *et al.*, 2005; Court, 2005; Picard *et al.*, 2005; Uchaipichat *et al.*, 2006a; Mano, Usui and Kamimura, 2007a, b). As a result, there is a need for selective substrates, immuno-reagents, and chemical inhibitors, so that reaction-phenotyping studies can be performed with native microsomes (e.g. FMOs and UGTs) and cytosol (e.g. SULTs). With the advent of newer approaches (e.g. small interfering RNA or "siRNA"), it may be possible to reaction-phenotype multiple DMEs using primary hepatocytes or other cell-based systems (Chen *et al.*, 2006; Yu, 2007).

Many other (non-CYP) enzymes are expressed polymorphically and are also present in the gut and liver. Therefore, the interplay between these enzymes and CYPs and the

potential consequences for inter-subject variability, drug interactions, and oral bioavailability should always be considered (Nishimura and Naito, 2006).

## ACKNOWLEDGMENT

The authors wish to thank Dr. David M. Stresser of BD Biosciences and Dr. Yang Xu of Amgen for their helpful critique of the manuscript.

## REFERENCES

Aithal, G.P., Day, C.P., Leathart, J.B., Daly, A.K. (2000). Relationship of polymorphism in cyp2c9 to genetic susceptibility to diclofenac-induced hepatitis. *Pharmacogenetics*, *10*, 511–518.

Ariyoshi, N., Takahashi, Y., Miyamoto, M., Umetsu, Y., Daigo, S., Tateishi, T., Kobayashi, S., Mizorogi, Y., Loriot, M.A., Stücker, I., Beaune, P., Kinoshita, M., Kamataki, T. (2000). Structural characterization of a new variant of the CYP2A6 gene (CYP2A6*1B) apparently diagnosed as heterozygotes of CYP2A6*1A and CYP2A6*4C. *Pharmacogenetics*, *10*, 687–693.

Arseculeratne, G., Wong, A.K., Goudie, D.R., Ferguson, J. (2007). Trimethylaminuria (fish-odor syndrome): a case report. *Arch. Dermatol.*, *143*, 81–84.

Baldwin, S.J., Bloomer, J.C., Smith, G.J., Ayrton, A.D., Clarke, S.E., Chenery, R.J. (1995). Ketoconazole and sulphaphenazole as the respective selective inhibitors of P4503A and 2C9. *Xenobiotica*, *25*, 261–270.

Barre, L., Fournel-Gigleux, S., Finel, M., Netter, P., Magdalou, J., Ouzzine, M. (2007). Substrate specificity of the human UDP-glucuronosyltransferase UGT2B4 and UGT2B7. Identification of a critical aromatic amino acid residue at position 33. *FEBS J.*, *274*, 1256–1264.

Barter, Z.E., Bayliss, M.K., Beaune, P.H., Boobis, A.R., Carlile, D.J., Edwards, R.J., Houston, J.B., Lake, B.G., Lipscomb, J.C., Pelkonen, O.R., Tucker, G.T., Rostami-Hodjegan, A. (2007). Scaling factors for the extrapolation of in vivo metabolic drug clearance from in vitro data: reaching a consensus on values of human microsomal protein and hepatocellularity per gram of liver. *Curr. Drug. Metab.*, *8*, 33–45.

Benetton, S.A., Fang, C., Yang, Y.O., Alok, R., Year, M., Lin, C.C., Yeh, L.T. (2007). P450 phenotyping of the metabolism of selegiline to desmethylselegiline and methamphetamine. *Drug Metab. Pharmacokinet.*, *22*, 78–87.

Bergström, M.A., Ott, H., Carlsson, A., Neis, M., Zwadlo-Klarwasser, G., Jonsson, C.A., Merk, H.F., Karlberg, A.T., Baron, J.M. (2007). A skin-like cytochrome P450 cocktail activates prohaptens to contact allergenic metabolites. *J. Invest. Dermatol.*, *127*, 1145–1153.

Bertilsson, L. (2007). Metabolism of antidepressant and neuroleptic drugs by cytochrome p450s: clinical and interethnic aspects. *Clin. Pharmacol. Ther.*, *82*, 606–609.

Bjornsson, T.D., Callaghan, J.T., Einolf, H.J., Fischer, V., Gan, L., Grimm, S., Kao, J., King, S.P., Miwa, G., Ni, L., Kumar, G., McLeod, J., Obach, R.S., Roberts, S., Roe, A., Shah, A., Snikeris, F., Sullivan, J.T., Tweedie, D., Vega, J.M., Walsh, J., Wrighton S.A., Pharmaceutical Research and Manufacturers of America (PhRMA), Drug Metabolism/Clinical Pharmacology Technical Working Group, FDA Center for Drug Evaluation and Research (CDER) (2003). The conduct of in vitro and in vivo drug-drug interaction studies: a Pharmaceutical Research and Manufacturers of America (PhRMA) perspective. *Drug Metab. Dispos.*, *31*, 815–832.

Bourrié, M., Meunier, V., Berger, Y., Fabre, G. (1996). Cytochrome P450 isoform inhibitors as a tool for the investigation of metabolic reactions catalyzed by human liver microsomes. *J. Pharmacol. Exp. Ther.*, *277*, 321–332.

Bowalgaha, K., Elliot, D.J., Mackenzie, P.I., Knights, K.M., Swedmark, S., Miners, J.O. (2005). S-Naproxen and desmethylnaproxen glucuronidation by human liver microsomes and recombi-

nant human UDP-glucuronosyltransferases (UGT): role of UGT2B7 in the elimination of naproxen. *Br. J. Clin. Pharmacol.*, *60*, 423–433.

Buckpitt, A., Boland, B., Isbell, M., Morin, D., Shultz, M., Baldwin, R., Chan, K., Karlsson, A., Lin, C., Taff, A., West, J., Fanucchi, M., Van Winkle, L., Plopper, C. (2002). Naphthalene-induced respiratory tract toxicity: metabolic mechanisms of toxicity. *Drug Metab. Rev.*, *34*, 791–820.

Cai, X., Wang, R.W., Edom, R.W., Evans, D.C., Shou, M., Rodrigues, A.D., Liu, W., Dean, D.C., Baillie, T.A. (2004). Validation of (-)-N-3-benzyl-phenobarbital as a selective inhibitor of CYP2C19 in human liver microsomes. *Drug Metab. Dispos.*, *32*, 584–586.

Cashman, J.R. (2000). Human flavin-containing monooxygenase: substrate specificity and role in drug metabolism. *Curr. Drug Metab.*, *1*, 181–191.

Cashman, J.R. (2004). The implications of polymorphisms in mammalian flavin-containing monooxygenases in drug discovery and development. *Drug Discov. Today*, *9*, 574–581.

Cashman, J.R. (2005). Some distinctions between flavin-containing and cytochrome P450 monooxygenases. *Biochem. Biophys. Res. Commun.*, *338*, 599–604.

Cashman, J.R., Bi, Y.A., Lin, J., Youil, R., Knight, M., Forrest, S., Treacy, E. (1997). Human flavin-containing monooxygenase form 3: cDNA expression of the enzymes containing amino acid substitutions observed in individuals with trimethylaminuria. *Chem. Res. Toxicol.*, *10*, 837–841.

Cashman, J.R., Park, S.B., Yang, Z.C., Washington, C.B., Gomez, D.Y., Giacomini, K.M., Brett, C.M. (1993). Chemical, enzymatic, and human enantioselective S-oxygenation of cimetidine. *Drug Metab. Dispos.*, *21*, 587–597.

Cashman, J.R., Park, S.B., Yang, Z.C., Wrighton, S.A., Jacob, P., III, Benowitz, N.L. (1992). Metabolism of nicotine by human liver microsomes: stereoselective formation of trans-nicotine N′-oxide. *Chem. Res. Toxicol.*, *5*, 639–646.

Cashman, J.R., Zhang, J. (2006). Human flavin-containing monooxygenases. *Annu Rev. Pharmacol. Toxicol.*, *46*, 65–100.

Castell, J.V., Donato, M.T., Gómez-Lechón, M.J. (2005). Metabolism and bioactivation of toxicants in the lung. The in vitro cellular approach. *Exp. Toxicol. Pathol.*, *57*, 189–204.

Chang, T.K., Gonzalez, F.J., Waxman, D.J. (1994). Evaluation of triacetyloleandomycin, alpha-naphthoflavone and diethyldithiocarbamate as selective chemical probes for inhibition of human cytochromes P450. *Arch. Biochem. Biophys.*, *311*, 437–442.

Chauret, N., Gauthier, A., Nicoll-Griffith, D.A. (1998). Effect of common organic solvents on in vitro cytochrome P450-mediated metabolic activities in human liver microsomes. *Drug Metab. Dispos.*, *26*, 1–4.

Chen, G., Blevins-Primeau, A.S., Dellinger, R.W., Muscat, J.E., Lazarus, P. (2007). Glucuronidation of nicotine and cotinine by UGT2B10: loss of function by the UGT2B10 Codon 67 (Asp > Tyr) polymorphism. *Cancer Res.*, *67*, 9024–9029

Chen, J., Yang, X., Huang, M., Hu, Z., He, M., Duan, W., Chan, E., Sheu, F., Chen, X., Zhou, S. (2006). Small interfering RNA-mediated silencing of cytochrome P450 3A4 gene. *Drug Metab. Dispos.*, *34*, 1650–1657.

Coughtrie, M.W., Blair, J.N., Hume, R., Burchell, A. (1991). Improved preparation of hepatic microsomes for in vitro diagnosis of inherited disorders of the glucose-6-phosphatase system. *Clin. Chem.*, *37*, 739–742.

Court, M.H. (2005). Isoform-selective probe substrates for in vitro studies of human UDP-glucuronosyltransferases. *Methods Enzymol.*, *400*, 104–116.

Crespi, C.L. (1995). Xenobiotic-metabolizing human cells as tools for pharmacological and toxicological research. *Adv. Drug Res.*, *26*, 179–235.

Czerniak, R. (2001). Gender-based differences in pharmacokinetics in laboratory animal models. *Int. J. Toxicol.*, *20*, 161–163.

Dalén, P., Dahl, M.L., Bernal Ruiz, M.L., Nordin, J., Bertilsson, L. (1998). 10-Hydroxylation of nortriptyline in white persons with 0, 1, 2, 3, and 13 functional CYP2D6 genes. *Clin. Pharmacol. Ther.*, *63*, 444–452.

Davies, B., Morris, T. (1993). Physiological parameters in laboratory animals and humans. *Pharm. Res.*, *10*, 1093–1095.

Dolphin C.T., Beckett D.J., Janmohamed A., Cullingford T.E., Smith R.L., Shephard E.A., Phillips, I.R. (1998). The flavin-containing monooxygenase 2 gene (FMO2) of humans, but not of other primates, encodes a truncated, nonfunctional protein. *J. Biol. Chem.*, *273*, 30599–30607.

Drahushuk, A.T., McGarrigle, B.P., Larsen, K.E., Stegeman, J.J., Olson, J.R. (1998). Detection of CYP1A1 protein in human liver and induction by TCDD in precision-cut liver slices incubated in dynamic organ culture. *Carcinogenesis*, *19*, 1361–1368.

Eagling, V.A., Tjia, J.F., Back, D.J. (1998). Differential selectivity of cytochrome P450 inhibitors against probe substrates in human and rat liver microsomes. *Br. J. Clin. Pharmacol.*, *45*, 107–114.

Ebner, T., Heinzel, G., Prox, A., Beschke, K., Wachsmuth, H. (1999). Disposition and chemical stability of telmisartan 1-O-acylglucuronide. *Drug Metab. Dispos.*, *27*, 1143–1149.

Fang, J. (2000). Metabolism of clozapine by rat brain: the role of flavin-containing monooxygenase (FMO) and cytochrome P450 enzymes. *Eur. J. Drug Metab. Pharmacokinet.*, *25*, 109–114.

FDA (2006). *FDA Guidelines for In Vitro Characterization of Drug candidates*, http://www.fda.gov/cder/guidance/6695dft.htm (accessed 31 October 2008).

Fisher, M.B., Campanale, K., Ackermann, B.L., VandenBranden, M., Wrighton, S.A. (2000). In vitro glucuronidation using human liver microsomes and the pore-forming peptide alamethicin. *Drug Metab. Dispos.*, *28*, 560–566.

Fisher, M.B., Paine, M.F., Strelevitz, T.J., Wrighton, S.A. (2001). The role of hepatic and extrahepatic UDP-glucuronosyltransferases in human drug metabolism. *Drug Metab. Rev.*, *33*, 273–297.

Fournel-Gigleux, S., Jackson, M.R., Wooster, R., Burchell, B. (1989). Expression of a human liver cDNA encoding a UDP-glucuronosyltransferase catalysing the glucuronidation of hyodeoxycholic acid in cell culture. *FEBS Lett.*, *243*, 119–122.

Fulceri, R., Bánhegyi, G., Gamberucci, A., Giunti, R., Mandl, J., Benedetti, A. (1994). Evidence for the intraluminal positioning of p-nitrophenol UDP-glucuronosyltransferase activity in rat liver microsomal vesicles. *Arch. Biochem. Biophys.*, *309*, 43–46.

Gagné, J.F., Montminy, V., Belanger, P., Journault, K., Gaucher, G., Guillemette, C. (2002). Common human UGT1A polymorphisms and the altered metabolism of irinotecan active metabolite 7-ethyl-10-hydroxycamptothecin (SN-38). *Mol. Pharmacol.*, *62*, 608–617.

Garteiz, D.A., Hook, R.H., Walker, B.J., Okerholm, R.A. (1982). Pharmacokinetics and biotransformation studies of terfenadine in man. *Arzneimittelforschung*, *32*, 1185–1190.

Gelboin, H.V., Krausz, K. (2006). Monoclonal antibodies and multifunctional cytochrome P450: drug metabolism as paradigm. *J. Clin. Pharmacol.*, *46*, 353–372.

Ghosal, A., Ramanathan, R., Yuan, Y., Hapangama, N., Chowdhury, S.K., Kishnani, N.S., Alton, K.B. (2007). Identification of human liver cytochrome P450 enzymes involved in biotransformation of vicriviroc, a CCR5 receptor antagonist. *Drug Metab. Dispos.*, *35*, 2186–2195.

Girennavar, B., Poulose, S.M., Jayaprakasha, G.K., Bhat, N.G., Patil, B.S. (2006). Furocoumarins from grapefruit juice and their effect on human CYP 3A4 and CYP 1B1 isoenzymes. *Bioorg. Med. Chem.*, *14*, 2606–2612.

Goldstein, J.A., Ishizaki, T., Chiba, K., de Morais, S.M., Bell, D., Krahn, P.M., Evans, D.A. (1997). Frequencies of the defective CYP2C19 alleles responsible for the mephenytoin poor metabolizer phenotype in various Oriental, Caucasian, Saudi Arabian and American black populations. *Pharmacogenetics*, *7*, 59–64.

Gómez-Lechón, M.J., Lahoz, A., Jiménez, N., Vicente Castell, J., Donato, M.T. (2006). Cryopreservation of rat, dog and human hepatocytes: influence of preculture and cryoprotectants on recovery, cytochrome P450 activities and induction upon thawing. *Xenobiotica*, *36*, 457–472.

Gorsky, L.D., Coon, M.J. (1986). Effects of conditions for reconstitution with cytochrome b5 on the formation of products in cytochrome P-450-catalyzed reactions. *Drug Metab. Dispos.*, *14*, 89–96.

Griese, E.U., Ilett, K.F., Kitteringham, N.R., Eichelbaum, M., Powell, H., Spargo, R.M., LeSouef, P.N., Musk, A.W., Minchin, R.F. (2001). Allele and genotype frequencies of polymorphic cytochromes P4502D6, 2C19 and 2E1 in aborigines from western Australia. *Pharmacogenetics*, *11*, 69–76.

Grillo, M.P., Hua, F., Knutson, C.G., Ware, J.A., Li, C. (2003a). Mechanistic studies on the bioactivation of diclofenac: identification of diclofenac-S-acyl-glutathione in vitro in incubations with rat and human hepatocytes. *Chem. Res. Toxicol.*, *16*, 1410–1417.

Grillo, M.P., Knutson, C.G., Sanders, P.E., Waldon, D.J., Hua, F., Ware, J.A. (2003b). Studies on the chemical reactivity of diclofenac acyl glucuronide with glutathione: identification of diclofenac-S-acyl-glutathione in rat bile. *Drug Metab. Dispos.*, *31*, 1327–1336.

Grothusen, A., Hardt, J., Bräutigam, L., Lang, D., Böcker, R. (1996). A convenient method to discriminate between cytochrome P450 enzymes and flavin-containing monooxygenases in human liver microsomes. *Arch. Toxicol.*, *71*, 64–71.

Guengerich, F.P. (1989). Characterization of human microsomal cytochrome P-450 enzymes. *Annu. Rev. Pharmacol. Toxicol.*, *29*, 241–264.

Guengerich, F.P. (2003). Cytochromes P450, drugs, and diseases. *Mol. Interv.*, *3*, 194–204.

Guengerich, F.P., Miller, G.P., Hanna, I.H., Martin, M.V., Léger, S., Black, C., Chauret, N., Silva, J.M., Trimble, L.A., Yergey, J.A., Nicoll-Griffith, D.A. (2002). Diversity in the oxidation of substrates by cytochrome P450 2D6: lack of an obligatory role of aspartate 301-substrate electrostatic bonding. *Biochemistry*, *41*, 11025–11034.

Henderson, M.C., Krueger, S.K., Stevens, J.F., Williams, D.E. (2004). Human flavin-containing monooxygenase form 2 S-oxygenation: sulfenic acid formation from thioureas and oxidation of glutathione. *Chem. Res. Toxicol.*, *17*, 633–640.

Hewitt, N.J., Lloyd, S., Hayden, M., Butler, R., Sakai, Y., Springer, R., Fackett, A., Li, A.P. (2002). Correlation between troglitazone cytotoxicity and drug metabolic enzyme activities in cryopreserved human hepatocytes. *Chem. Biol. Interact.*, *142*, 73–82.

Hickman, D., Wang, J.P., Wang, Y., Unadkat, J.D. (1998). Evaluation of the selectivity of In vitro probes and suitability of organic solvents for the measurement of human cytochrome P450 monooxygenase activities. *Drug Metab. Dispos.*, *26*, 207–215.

Hiroi, T., Imaoka, S., Funae, Y. (1998). Dopamine formation from tyramine by CYP2D6. *Biochem. Biophys. Res. Commun.*, *249*, 838–843.

Honig, P.K., Wortham, D.C., Lazarev, A., Cantilena, L.R. (1996). Grapefruit juice alters the systemic bioavailability and cardiac repolarization of terfenadine in poor metabolizers of terfenadine. *J. Clin. Pharmacol.*, *36*, 345–351.

Honig, P.K., Wortham, D.C., Hull, R., Zamani, K., Smith, J.E., Cantilena, L.R. (1993a). Itraconazole affects single-dose terfenadine pharmacokinetics and cardiac repolarization pharmacodynamics. *J. Clin. Pharmacol.*, *33*, 1201–1206.

Honig, P.K., Wortham, D.C., Zamani, K., Conner, D.P., Mullin, J.C., Cantilena, L.R. (1993b). Terfenadine-ketoconazole interaction. Pharmacokinetic and electrocardiographic consequences. *JAMA*, *269*, 1513–1518.

Houston, J.B. (1994). Utility of in vitro drug metabolism data in predicting in vivo metabolic clearance. *Biochem. Pharmacol.*, *47*, 1469–1479.

Hsieh, K.P., Lin, Y.Y., Cheng, C.L., Lai, M.L., Lin, M.S., Siest, J.P., Huang, J.D. (2001). Novel mutations of CYP3A4 in Chinese. *Drug Metab. Dispos.*, *29*, 268–273.

Hutzler, J.M., Hauer, M.J., Tracy, T.S. (2001). Dapsone activation of CYP2C9-mediated metabolism: evidence for activation of multiple substrates and a two-site model. *Drug Metab. Dispos.*, *29*, 1029–1034.

Jain, K.K. (2005). Applications of AmpliChip CYP450. *Mol. Diagn.*, *9*, 119–127.

Jakobsson, J., Ekström, L., Inotsume, N., Garle, M., Lorentzon, M., Ohlsson, C., Roh, H.K., Carlström, K., Rane, A. (2006). Large differences in testosterone excretion in Korean and Swedish men are strongly associated with a UDP-glucuronosyl transferase 2B17 polymorphism. *J. Clin. Endocrinol. Metab.*, *91*, 687–693.

Jedlitschky, G., Cassidy, A.J., Sales, M., Pratt, N., Burchell, B. (1999). Cloning and characterization of a novel human olfactory UDP-glucuronosyltransferase. *Biochem. J.*, *340*, 837–843.

Kaivosaari, S., Toivonen, P., Hesse, L.M., Koskinen, M., Court, M.H., Finel, M. (2007). Nicotine glucuronidation and the human UDP-glucuronosyltransferase UGT2B10. *Mol. Pharmacol.*, *72*, 761–768.

Kaji, H., Kume, T. (2005). Regioselective glucuronidation of denopamine: marked species differences and identification of human udp-glucuronosyltransferase isoform. *Drug Metab. Dispos.*, *33*, 403–412.

Kaneko, A., Bergqvist, Y., Taleo, G., Kobayakawa, T., Ishizaki, T., Björkman, A. (1999a). Proguanil disposition and toxicity in malaria patients from Vanuatu with high frequencies of CYP2C19 mutations. *Pharmacogenetics*, *9*, 317–326.

Kaneko, A., Lum, J.K., Yaviong, L., Takahashi, N., Ishizaki, T., Bertilsson, L., Kobayakawa, T., Björkman, A. (1999b). High and variable frequencies of CYP2C19 mutations: medical consequences of poor drug metabolism in Vanuatu and other Pacific islands. *Pharmacogenetics*, *9*, 581–590.

Kang, J.H., Chung, W.G., Lee, K.H., Park, C.S., Kang, J.S., Shin, I.C., Roh, H.K., Dong, M.S., Baek, H.M., Cha, Y.N. (2000). Phenotypes of flavin-containing monooxygenase activity determined by ranitidine N-oxidation are positively correlated with genotypes of linked FM03 gene mutations in a Korean population. *Pharmacogenetics*, *10*, 67–78.

Katoh, M., Matsui, T., Yokoi, T. (2007). Glucuronidation of antiallergic drug, Tranilast: identification of human UDP-glucuronosyltransferase isoforms and effect of its phase I metabolite. *Drug Metab. Dispos.*, *35*, 583–589.

Kenworthy, K.E., Bloomer, J.C., Clarke, S.E., Houston, J.B. (1999). CYP3A4 drug interactions: correlation of 10 in vitro probe substrates. *Br. J. Clin. Pharmacol.*, *48*, 716–727.

King, C.D., Rios, G.R., Green, M.D., Tephly, T.R. (2000). UDP-glucuronosyltransferases. *Curr. Drug Metab.*, *1*, 143–161.

Krause, R.J., Lash, L.H., Elfarra, A.A. (2003). Human kidney flavin-containing monooxygenases and their potential roles in cysteine s-conjugate metabolism and nephrotoxicity. *J. Pharmacol. Exp. Ther.*, *304*, 185–191.

Krueger, S.K., Williams, D.E. (2005). Mammalian flavin-containing monooxygenases: structure/function, genetic polymorphisms and role in drug metabolism. *Pharmacol. Ther.*, *106*, 357–387.

Kuehl, P., Zhang, J., Lin, Y., Lamba, J., Assem, M., Schuetz, J., Watkins, P.B., Daly, A., Wrighton, S.A., Hall, S.D., Maurel, P., Relling, M., Brimer, C., Yasuda, K., Venkataramanan, R., Strom, S., Thummel, K., Boguski, M.S., Schuetz, E. (2001). Sequence diversity in CYP3A promoters and characterization of the genetic basis of polymorphic CYP3A5 expression. *Nat. Genet.*, *27*, 383–391.

Kumar, V., Brundage, R.C., Oetting, W.S., Leppik, I.E., Tracy, T.S. (2008). Differential genotype dependent inhibition of CYP2C9 in humans. *Drug Metab Dispos*, *36*, 1242–1248. (Epub ahead of print).

Kumar, V., Wahlstrom, J.L., Rock, D.A., Warren, C.J., Gorman, L.A., Tracy, T.S. (2006). CYP2C9 inhibition: impact of probe selection and pharmacogenetics on in vitro inhibition profiles. *Drug Metab. Dispos.*, *34*, 1966–1975.

Kunze, K.L., Trager, W.F. (1993). Isoform-selective mechanism-based inhibition of human cytochrome P450 1A2 by furafylline. *Chem. Res. Toxicol.*, *6*, 649–656.

Kwon, J.T., Nakajima, M., Chai, S., Yom, Y.K., Kim, H.K., Yamazaki, H., Sohn, D.R., Yamamoto, T., Kuroiwa, Y., Yokoi, T. (2001). Nicotine metabolism and CYP2A6 allele frequencies in Koreans. *Pharmacogenetics*, *11*, 317–323.

Lang, D.H., Yeung, C.K., Peter, R.M., Ibarra, C., Gasser, R., Itagaki, K., Philpot, R.M., Rettie, A.E. (1998). Isoform specificity of trimethylamine N-oxygenation by human flavin-containing monooxygenase (FMO) and P450 enzymes: selective catalysis by FMO3. *Biochem. Pharmacol.*, *56*, 1005–1012.

Lasker, J.M., Wester, M.R., Aramsombatdee, E., Raucy, J.L. (1998). Characterization of CYP2C19 and CYP2C9 from human liver: respective roles in microsomal tolbutamide, S-mephenytoin, and omeprazole hydroxylations. *Arch. Biochem. Biophys.*, *353*, 16–28.

Lee, H.S., Ji, H.Y., Park, E.J., Kim, S.Y. (2007a). In vitro metabolism of eupatilin by multiple cytochrome P450 and UDP-glucuronosyltransferase enzymes. *Xenobiotica*, *37*, 803–817.

Lee, S., Hwang, H.J., Kim, J.M., Chung, C.S., Kim, J.H. (2007b). CYP2C19 polymorphism in Korean patients on warfarin therapy. *Arch. Pharm. Res.*, *30*, 344–349.

Lee, H.S., Park, E.J., Ji, H.Y., Kim, S.Y., Im, G.J., Lee, S.M., Jang, I.J. (2008). Identification of cytochrome P450 enzymes responsible for N-dealkylation of a new oral erectogenic, mirodenafil. *Xenobiotica*, *38*, 21–33.

Lewis, D.F., Dickins, M. (2002). Substrate SARs in human P450s. *Drug Discov. Today*, *7*, 918–925.

Lewis, D.F., Jacobs, M.N., Dickins, M. (2004). Compound lipophilicity for substrate binding to human P450s in drug metabolism. *Drug Discov. Today*, *9*, 530–537.

Li, A.P., Lu, C., Brent, J.A., Pham, C., Fackett, A., Ruegg, C.E., Silber, P.M. (1999). Cryopreserved human hepatocytes: characterization of drug-metabolizing enzyme activities and applications in higher throughput screening assays for hepatotoxicity, metabolic stability, and drug-drug interaction potential. *Chem. Biol. Interact.*, *121*, 17–35.

Li, L., Pan, R.M., Porter, T.D., Jensen, N.S., Silber, P., Russo, G., Tine, J.A., Heim, J., Ring, B., Wedlund, P.J. (2006). New cytochrome P450 2D6*56 allele identified by genotype/phenotype analysis of cryopreserved human hepatocytes. *Drug Metab. Dispos.*, *34*, 1411–1416.

Lévesque, E., Turgeon, D., Carrier, J.S., Montminy, V., Beaulieu, M., Bélanger, A. (2001). Isolation and characterization of the UGT2B28 cDNA encoding a novel human steroid conjugating UDP-glucuronosyltransferase. *Biochemistry*, *40*, 3869–3881.

Mano, Y., Usui, T., Kamimura, H. (2007a). Predominant contribution of UDP-glucuronosyltransferase 2B7 in the glucuronidation of racemic flurbiprofen in the human liver. *Drug Metab. Dispos.*, *35*, 1182–1187.

Mano, Y., Usui, T., Kamimura, H. (2007b). The UDP-Glucuronosyltransferase 2B7 isozyme is responsible for gemfibrozil glucuronidation in the human liver. *Drug Metab. Dispos.*, *35*, 2040–2044.

Markowitz, J.S., Donovan, J.L., DeVane, C.L., Taylor, R.M., Ruan, Y., Wang, J.S., Chavin, K.D. (2003). Effect of St John's wort on drug metabolism by induction of cytochrome P450 3A4 enzyme. *JAMA*, *290*, 1500–1504.

Mennes, W.C., van Holsteijn, C.W., van Iersel, A.A., Yap, S.H., Noordhoek, J., Blaauboer, B.J. (1994). Interindividual variation in biotransformation and cytotoxicity of bromobenzene as determined in primary hepatocyte cultures derived from monkey and human liver. *Hum. Exp. Toxicol.*, *13*, 415–421.

Miners, J.O., Knights, K.M., Houston, J.B., Mackenzie, P.I. (2006). In vitro-in vivo correlation for drugs and other compounds eliminated by glucuronidation in humans: pitfalls and promises. *Biochem. Pharmacol.*, *71*, 1531–1539.

Mizutani, T. (2003). PM frequencies of major CYPs in Asians and. *Caucasians. Drug Metab. Rev.*, *35*, 99–106.

Mouly, S.J., Matheny, C., Paine, M.F., Smith, G., Lamba, J., Lamba, V., Pusek, S.N., Schuetz, E.G., Stewart, P.W., Watkins, P.B. (2005). Variation in oral clearance of saquinavir is predicted by CYP3A5*1 genotype but not by enterocyte content of cytochrome P450 3A5. *Clin. Pharmacol. Ther.*, *78*, 605–618.

Muralidharan, G., Hawes, E.M., McKay, G., Korchinski, E.D., Midha, K.K. (1991). Quinidine but not quinine inhibits in man the oxidative metabolic routes of methoxyphenamine which involve debrisoquine 4-hydroxylase. *Eur. J. Clin. Pharmacol.*, *41*, 471–474.

Mushiroda, T., Douya, R., Takahara, E., Nagata, O. (2000). The involvement of flavin-containing monooxygenase but not CYP3A4 in metabolism of itopride hydrochloride, a gastroprokinetic agent: comparison with cisapride and mosapride citrate. *Drug Metab. Dispos.*, *28*, 1231–1237.

Naisbitt, D.J., Sanderson, L.S., Meng, X., Stachulski, A.V., Clarke, S.E., Park, B.K. (2007). Investigation of the immunogenicity of diclofenac and diclofenac metabolites. *Toxicol. Lett.*, *168*, 45–50.

Nakajima, M., Nakamura, S., Tokudome, S., Shimada, N., Yamazaki, H., Yokoi, T. (1999). Azelastine N-demethylation by cytochrome P-450 (CYP)3A4, CYP2D6, and CYP1A2 in human liver microsomes: evaluation of approach to predict the contribution of multiple CYPs. *Drug Metab. Dispos.*, *27*, 1381–1391.

Neuvonen, P.J., Kantola, T., Kivistö, K.T. (1998). Simvastatin but not pravastatin is very susceptible to interaction with the CYP3A4 inhibitor itraconazole. *Clin. Pharmacol. Ther.*, *63*, 332–341.

Nishimura, M., Naito, S. (2006). Tissue-specific mRNA expression profiles of human phase I metabolizing enzymes except for cytochrome P450 and phase II metabolizing enzymes. *Drug Metab. Pharmacokinet.*, *21*, 357–374.

Niu, C.Y., Luo, J.Y., Hao, Z.M. (2004). Genetic polymorphism analysis of cytochrome P4502C19 in Chinese Uigur and Han populations. *Chin. J. Dig. Dis.*, *5*, 76–80.

O'Brien, S.E., de Groot, M.J. (2005). Greater than the sum of its parts: combining models for useful ADMET prediction. *J. Med. Chem.*, *48*, 1287–1291.

Obach, R.S. (1999). Prediction of human clearance of twenty-nine drugs from hepatic microsomal intrinsic clearance data: an examination of in vitro half-life approach and nonspecific binding to microsomes. *Drug Metab. Dispos.*, *27*, 1350–1359.

Obach, R.S., Baxter, J.G., Liston, T.E., Silber, B.M., Jones, B.C., MacIntyre, F., Rance, D.J., Wastall, P. (1997). The prediction of human pharmacokinetic parameters from preclinical and in vitro metabolism data. *J. Pharmacol. Exp. Ther.*, *283*, 46–58.

Obach, R.S., Margolis, J.M., Logman, M.J. (2007). In vitro metabolism of CP-122,721 ((2S,3S)-2-phenyl-3-[(5-trifluoromethoxy-2-methoxy)benzylamino]piperidine), a non-peptide antagonist of the substance P receptor. *Drug Metab. Pharmacokinet.*, *22*, 336–349.

Ohmi, N., Yoshida, H., Endo, H., Hasegawa, M., Akimoto, M., Higuchi, S. (2003). S-oxidation of S-methyl-esonarimod by flavin-containing monooxygenases in human liver microsomes. *Xenobiotica*, *33*, 1221–1231.

Oscarson, M., McLellan, R.A., Gullstén, H., Agúndez, J.A., Benítez, J., Rautio, A., Raunio, H., Pelkonen, O., Ingelman-Sundberg, M. (1999a). Identification and characterisation of novel polymorphisms in the CYP2A locus: implications for nicotine metabolism. *FEBS Lett.*, *460*, 321–327.

Oscarson, M., McLellan, R.A., Gullstén, H., Yue, Q.Y., Lang, M.A., Bernal, M.L., Sinues, B., Hirvonen, A., Raunio, H., Pelkonen, O., Ingelman-Sundberg, M. (1999b). Characterisation and PCR-based detection of a CYP2A6 gene deletion found at a high frequency in a Chinese population. *FEBS Lett.*, *448*, 105–110.

Overby, L.H., Carver, G.C., Philpot, R.M. (1997). Quantitation and kinetic properties of hepatic microsomal and recombinant flavin-containing monooxygenases 3 and 5 from humans. *Chem. Biol. Interact.*, *106*, 29–45.

Ozdemir, V., Kalow, W., Tang, B.K., Paterson, A.D., Walker, S.E., Endrenyi, L., Kashuba, A.D. (2000). Evaluation of the genetic component of variability in CYP3A4 activity: a repeated drug administration method. *Pharmacogenetics*, *10*, 373–388.

Paine, M.F., Hart, H.L., Ludington, S.S., Haining, R.L., Rettie, A.E., Zeldin, D.C. (2006). The human intestinal cytochrome P450 "pie". *Drug Metab. Dispos.*, *34*, 880–886.

Parkinson, A. (1996). An overview of current cytochrome P450 technology for assessing the safety and efficacy of new materials. *Toxicol. Pathol.*, *24*, 48–57.

Paulson, S.K., Engel, L., Reitz, B., Bolten, S., Burton, E.G., Maziasz, T.J., Yan, B., Schoenhard, G.L. (1999). Evidence for polymorphism in the canine metabolism of the cyclooxygenase 2 inhibitor, celecoxib. *Drug Metab. Dispos.*, *27*, 1133–1142.

Picard, N., Ratanasavanh, D., Prémaud, A., Le Meur, Y., Marquet, P. (2005). Identification of the UDP-glucuronosyltransferase isoforms involved in mycophenolic acid phase II metabolism. *Drug Metab. Dispos.*, *33* (1), 139–146.

Ramamoorthy, Y., Tyndale, R.F., Sellers, E.M. (2001). Cytochrome P450 2D6.1 and cytochrome P450 2D6.10 differ in catalytic activity for multiple substrates. *Pharmacogenetics*, *11*, 477–487.

Rendic, S., Di Carlo, F.J. (1997). Human cytochrome P450 enzymes: a status report summarizing their reactions, substrates, inducers, and inhibitors. *Drug Metab. Rev.*, *29*, 413–580.

Rodrigues, A.D. (1999). Integrated cytochrome P450 reaction phenotyping: attempting to bridge the gap between cDNA-expressed cytochromes P450 and native human liver microsomes. *Biochem. Pharmacol.*, *57*, 465–480.

Rowland, A., Elliot, D.J., Williams, J.A., Mackenzie, P.I., Dickinson, R.G., Miners, J.O. (2006). In vitro characterization of lamotrigine N2-glucuronidation and the lamotrigine-valproic acid interaction. *Drug. Metab. Dispos.*, *34*, 1055–1062.

Rowland, A., Gaganis, P., Elliot, D.J., Mackenzie, P.I., Knights, K.M., Miners, J.O. (2007). Binding of inhibitory fatty acids is responsible for the enhancement of UDP-glucuronosyltransferase 2B7 activity by albumin: implications for in vitro-in vivo extrapolation. *J. Pharmacol. Exp. Ther.*, *321*, 137–147.

Rowland-Yeo, K., Rostami-Hodjegan, A., Tucker, G.T., British Pharmacological Society Winter Meeting University of London (2003). Abundance of cytochrome P450 in human liver: a meta-analysis. *Br. J. Clin. Pharmacol.*, *57* (5), 687.

Sachse, C., Brockmöller, J., Bauer, S., Roots, I. (1997). Cytochrome P450 2D6 variants in a Caucasian population: allele frequencies and phenotypic consequences. *Am. J. Hum. Genet.*, *60*, 284–295.

Schrag, M.L., Cui, D., Rushmore, T.H., Shou, M., Ma, B., Rodrigues, A.D. (2004). Sulfotransferase 1E1 is a low km isoform mediating the 3-O-sulfation of ethinyl estradiol. *Drug Metab. Dispos.*, *32*, 1299–1303.

Segel, I. (1975). Kinetics of unireactant enzymes, chapter 2 in *Enzyme Kinetics. Behavior Analysis of Rapid Equilibrium and Steady-State Systems* (ed. I. Segel), John Wiley & Sons, Inc., New York, pp. 54–55.

Shepherd, S.R., Baird, S.J., Hallinan, T., Burchell, B. (1989). An investigation of the transverse topology of bilirubin UDP-glucuronosyltransferase in rat hepatic endoplasmic reticulum. *Biochem. J.*, *259*, 617–620.

Shimada, T., Tsumura, F., Yamazaki, H., Guengerich, F.P., Inoue, K. (2001). Characterization of (+/–)-bufuralol hydroxylation activities in liver microsomes of Japanese and Caucasian subjects genotyped for CYP2D6. *Pharmacogenetics*, *11*, 143–156.

Shimada, T., Yamazaki, H., Mimura, M., Inui, Y., Guengerich, F.P. (1994). Interindividual variations in human liver cytochrome P-450 enzymes involved in the oxidation of drugs, carcinogens and toxic chemicals: studies with liver microsomes of 30 Japanese and 30 Caucasians. *J. Pharmacol. Exp. Ther.*, *270*, 414–423.

Shimoda, K., Jerling, M., Böttiger, Y., Yasuda, S., Morita, S., Bertilsson, L. (1999). Pronounced differences in the disposition of clomipramine between Japanese and Swedish patients. *J Clin. Psychopharmacol.*, *19*, 393–400.

Snawder, J.E., Lipscomb, J.C. (2000). Interindividual variance of cytochrome P450 forms in human hepatic microsomes: correlation of individual forms with xenobiotic metabolism and implications in risk assessment. *Regul. Toxicol. Pharmacol.*, *32*, 200–209.

Soars, M.G., Gelboin, H.V., Krausz, K.W., Riley, R.J. (2003). A comparison of relative abundance, activity factor and inhibitory monoclonal antibody approaches in the characterization of human CYP enzymology. *Br. J. Clin. Pharmacol.*, *55*, 175–181.

Soars, M.G., Riley, R.J., Findlay, K.A., Coffey, M.J., Burchell, B. (2001). Evidence for significant differences in microsomal drug glucuronidation by canine and human liver and kidney. *Drug Metab. Dispos.*, *29*, 121–126.

Soars, M.G., Ring, B.J., Wrighton, S.A. (2003). The effect of incubation conditions on the enzyme kinetics of udp-glucuronosyltransferases. *Drug Metab. Dispos.*, *31*, 762–767.

Störmer, E., von Moltke, L.L., Greenblatt, D.J. (2000). Scaling drug biotransformation data from cDNA-expressed cytochrome P-450 to human liver: a comparison of relative activity factors and human liver abundance in studies of mirtazapine metabolism. *J. Pharmacol. Exp. Ther.*, *295*, 793–801.

Störmer, E., von Moltke, L.L., Shader, R.I., Greenblatt, D.J. (2000). Metabolism of the antidepressant mirtazapine in vitro: contribution of cytochromes P-450 1A2, 2D6, and 3A4. *Drug Metab. Dispos.*, *28*, 1168–1175.

Stouch, T.R., Kenyon, J.R., Johnson, S.R., Chen, X.Q., Doweyko, A., Li, Y. (2003). In silico ADME/Tox: why models fail. *J. Comput. Aided. Mol. Des.*, *17*, 83–92.

Suzuki, H., Kneller, M.B., Haining, R.L., Trager, W.F., Rettie, A.E. (2002). (+)-N-3-Benzyl-nirvanol and (−)-N-3-benzyl-phenobarbital: new potent and selective in vitro inhibitors of CYP2C19. *Drug Metab. Dispos.*, *30*, 235–239.

Swanson, H.I. (2004). Cytochrome P450 expression in human keratinocytes: an aryl hydrocarbon receptor perspective. *Chem. Biol. Interact.*, *149*, 69–79.

Tachibana, M., Tanaka, M., Masubuchi, Y., Horie, T. (2005). Acyl glucuronidation of fluroquinolone antibiotics by the UDP-glucuronosyltransferase 1A subfamily in human liver miscrosomes. *Drug Metab. Dispos.*, *33*, 803–811.

Takanashi, K., Tainaka, H., Kobayashi, K., Yasumori, T., Hosakawa, M., Chiba, K. (2000). CYP2C9 Ile359 and Leu359 variants: enzyme kinetic study with seven substrates. *Pharmacogenetics*, *10*, 95–104.

Toffoli, G., Cecchin, E. (2004). Uridine diphosphoglucuronosyl transferase and methylenetetrahydrofolate reductase polymorphisms as genomic predictors of toxicity and response to irinotecan-, antifolate- and fluoropyrimidine-based chemotherapy. *J. Chemother.*, *4*, 31–35.

Toffoli, G., Cecchin, E., Corona, G., Russo, A., Buonadonna, A., D'Andrea, M., Pasetto, L.M., Pessa, S., Errante, D., De Pangher, V., Giusto, M., Medici, M., Gaion, F., Sandri, P., Galligioni, E., Bonura, S., Boccalon, M., Biason, P., Frustaci, S. (2006). The role of UGT1A1*28 polymorphism in the pharmacodynamics and pharmacokinetics of irinotecan in patients with metastatic colorectal cancer. *J. Clin. Oncol.*, *24*, 3061–3068.

Tukey, R.H., Strassburg, C.P. (2000). Human UDP-glucuronosyltransferases: metabolism, expression, and disease. *Annu. Rev. Pharmacol. Toxicol.*, *40*, 581–616.

Turgeon, D., Carrier, J.S., Chouinard, S., Bélanger, A. (2003a). Glucuronidation activity of the UGT2B17 enzyme toward xenobiotics. *Drug Metab. Dispos.*, *31*, 670–676.

Turgeon, D., Chouinard, S., Belanger, P., Picard, S., Labbe, J.F., Borgeat, P., Belanger, A. (2003b). Glucuronidation of arachidonic and linoleic acid metabolites by human UDP-glucuronosyltransferases. *J. Lipid. Res.*, *44*, 1182–1191.

Turpeinen, M., Nieminen, R., Juntunen, T., Taavitsainen, P., Raunio, H., Pelkonen, O. (2004). Selective inhibition of CYP2B6-catalyzed bupropion hydroxylation in human liver microsomes in vitro. *Drug Metab. Dispos.*, *32*, 626–631.

Uchaipichat, V., Mackenzie, P.I., Elliot, D.J., Miners, J.O. (2006a). Selectivity of substrate (trifluoperazine) and inhibitor (amitriptyline, androsterone, canrenoic acid, hecogenin, phenylbutazone, quinidine, quinine, and sulfinpyrazone) "probes" for human udp-glucuronosyltransferases. *Drug Metab. Dispos.*, *34*, 449–456.

Uchaipichat, V., Winner, L.K., Mackenzie, P.I., Elliot, D.J., Williams, J.A., Miners, J.O. (2006b). Quantitative prediction of in vivo inhibitory interactions involving glucuronidated drugs from in vitro data: the effect of fluconazole on zidovudine glucuronidation. *Br. J. Clin. Pharmacol.*, *61*, 427–439.

Vanstapel, F., Blanckaert, N. (1988). Topology and regulation of bilirubin UDP-glucuronyltransferase in sealed native microsomes from rat liver. *Arch. Biochem. Biophys.*, *263*, 216–225.

Venkatakrishnan, K., Obach, R.S. (2007). Drug-drug interactions via mechanism-based cytochrome P450 inactivation: points to consider for risk assessment from in vitro data and clinical pharmacologic evaluation. *Curr. Drug Metab.*, *8*, 449–462.

Venkatakrishnan, K., von Moltke, L.L., Court, M.H., Harmatz, J.S., Crespi, C.L., Greenblatt, D.J. (2000). Comparison between cytochrome P450 (CYP) content and relative activity approaches to scaling from cDNA-expressed CYPs to human liver microsomes: ratios of accessory proteins as sources of discrepancies between the approaches. *Drug Metab. Dispos.*, *28*, 1493–1504.

Venkatakrishnan, K., von Moltke, L.L., Greenblatt, D.J. (2001). Human drug metabolism and the cytochromes P450: application and relevance of in vitro models. *J. Clin. Pharmacol.*, *41*, 1149–1179.

Visser, T.J., Kaptein, E., van Toor, H., van Raaij, J.A., van den Berg, K.J., Joe, C.T., van Engelen, J.G., Brouwer, A. (1993). Glucuronidation of thyroid hormone in rat liver: effects of in vivo treatment with microsomal enzyme inducers and in vitro assay conditions. *Endocrinology*, *133*, 2177–2186.

Walsky, R.L., Obach, R.S., Gaman, E.A., Gleeson, J.P., Proctor, W.R. (2005). Selective inhibition of human cytochrome P4502C8 by montelukast. *Drug Metab. Dispos.*, *33*, 413–418.

Whetstine, J.R., Yueh, M.F., McCarver, D.G., Williams, D.E., Park, C.S., Kang, J.H., Cha, Y.N., Dolphin, C.T., Shephard, E.A., Phillips, I.R., Hines, R.N. (2000). Ethnic differences in human flavin-containing monooxygenase 2 (FMO2) polymorphisms: detection of expressed protein in African-Americans. *Toxicol. Appl. Pharmacol.*, *168*, 216–224.

Wilkinson, G.R., Shand, D.G. (1975). Commentary: a physiological approach to hepatic drug clearance. *Clin. Pharmacol. Ther.*, *18*, 377–390.

Williams, J.A., Hurst, S.I., Bauman, J., Jones, B.C., Hyland, R., Gibbs, J.P., Obach, R.S., Ball, S.E. (2003a). Reaction phenotyping in drug discovery: moving forward with confidence? *Curr. Drug. Metab.*, *4*, 527–534.

Williams, P.A., Cosme, J., Ward, A., Angove, H.C., Matak Vinkovi , D., Jhoti, H. (2003b). Crystal structure of human cytochrome P450 2C9 with bound warfarin. *Nature, 424* (*6947*), 464–468.

Williams, J.A., Hyland, R., Jones, B.C., Smith, D.A., Hurst, S., Goosen, T.C., Peterkin, V., Koup, J.R., Ball, S.E. (2004). Drug-drug interactions for UDP-glucuronosyltransferase substrates: a pharmacokinetic explanation for typically observed low exposure (AUCi/AUC) ratios. *Drug Metab. Dispos.*, *32*, 1201–1208.

Woosley, R.L., Chen, Y., Freiman, J.P., Gillis, R.A. (1993). Mechanism of the cardiotoxic actions of terfenadine. *JAMA*, *269*, 1532–1536.

Yamanaka, H., Nakajima, M., Katoh, A., Tamura, O., Ishibashi, H., Yokoi, T. (2005). Trans-3′-hydroxycotinine O- and N-glucuronidations in human liver microsomes. *Drug Metab. Dispos.*, *33*, 23–30.

Yamazaki, H., Johnson, W.W., Ueng, Y.F., Shimada, T., Guengerich, F.P. (1996). Lack of electron transfer from cytochrome b5 in stimulation of catalytic activities of cytochrome P450 3A4. Characterization of a reconstituted cytochrome P450 3A4/NADPH-cytochrome P450 reductase system and studies with apo-cytochrome b5. *J. Biol. Chem.*, *271*, 27438–27444.

Yamazaki, H., Shimada, T. (1998). Comparative studies of in vitro inhibition of cytochrome P450 3A4-dependent testosterone 6beta-hydroxylation by roxithromycin and its metabolites, troleandomycin, and erythromycin. *Drug. Metab. Dispos.*, *26*, 1053–1057.

Yamazaki, H., Ueng, Y.F., Shimada, T., Guengerich, F.P. (1995). Roles of divalent metal ions in oxidations catalyzed by recombinant cytochrome P450 3A4 and replacement of NADPH–cytochrome P450 reductase with other flavoproteins, ferredoxin, and oxygen surrogates. *Biochemistry*, *34*, 8380–8389.

Yokono, A., Morita, S., Someya, T., Hirokane, G., Okawa, M., Shimoda, K. (2001). The effect of CYP2C19 and CYP2D6 genotypes on the metabolism of clomipramine in Japanese psychiatric patients. *J. Clin. Psychopharmacol.*, *21*, 549–555.

Yoon, Y.J., Kim, K.B., Kim, H., Seo, K.A., Kim, H.S., Cha, I.J., Kim, E.Y., Liu, K.H., Shin, J.G. (2007). Characterization of benidipine and its enantiomers' metabolism by human liver cytochrome P450 enzymes. *Drug Metab. Dispos.*, *35*, 1518–1524.

Yu, A. (2007). Small interfering RNA in drug metabolism and transport. *Curr. Drug Metab.*, *8*, 700–708.

Yu, L., Lu, S., Lin, Y., Zeng, S. (2007). Carboxyl-glucuronidation of mitiglinide by human UDP-glucuronosyltransferases. *Biochem. Pharmacol.*, *73*, 1842–1851.

Yun, C.H., Okerholm, R.A., Guengerich, F.P. (1993). Oxidation of the antihistaminic drug terfenadine in human liver microsomes. Role of cytochrome P-450 3A(4) in N-dealkylation and C-hydroxylation. *Drug Metab. Dispos.*, *21*, 403–409.

Zhang, H., Davis, C.D., Sinz, M.W., Rodrigues, A.D. (2007). Cytochrome P450 reaction-phenotyping: an industrial perspective. *Expert Opin. Drug Metab. Toxicol.*, *3*, 667–687.

Zhang, W., Kilicarslan, T., Tyndale, R.F., Sellers, E.M. (2001). Evaluation of methoxsalen, tranylcypromine, and tryptamine as specific and selective CYP2A6 inhibitors in vitro. *Drug Metab. Dispos.*, *29*, 897–902.

Ziegler, D.M. (1988). Flavin-containing monooxygenases: catalytic mechanism and substrate specificities. *Drug Metab. Rev.*, *19*, 1–32.

## FURTHER READING

Bjornsson, T.D., Callaghan, J.T., Einolf, H.J., Fischer, V., Gan, L., Grimm, S., Kao, J., King, S.P., Miwa, G., Ni, L., Kumar G., McLeod, J., Obach, S.R., Roberts, S., Roe, A., Shah, A., Snikeris, F., Sullivan, J.T., Tweedie, D., Vega, J.M., Walsh, J., Wrighton, S.A., Pharmaceutical Research and Manufacturers of America Drug Metabolism/Clinical Pharmacology Technical Working Groups. (2003). The conduct of in vitro and in vivo drug-drug interaction studies: a PhRMA perspective. *J. Clin. Pharmacol.*, *43*, 443–469. Review.

Emoto, C., Iwasaki, K. (2007). Approach to predict the contribution of cytochrome P450 enzymes to drug metabolism in the early drug-discovery stage: the effect of the expression of cytochrome b(5) with recombinant P450 enzymes. *Xenobiotica*, *37*, 986–999.

Emoto, C., Murase, S., Iwasaki, K. (2006). Approach to the prediction of the contribution of major cytochrome P450 enzymes to drug metabolism in the early drug-discovery stage. *Xenobiotica*, *36*, 671–683.

Ito, K., Hallifax, D., Obach R.S., Houston J.B. (2005). Impact of parallel pathways of drug elimination and multiple cytochrome P450 involvement on drug-drug interactions: CYP2D6 paradigm. *Drug Metab. Dispos.*, *33*, 837–844.

Ito, K., Iwatsubo, T., Kanamitsu, S., Ueda, K., Suzuki, H., Sugiyama, Y. (1998). Prediction of pharmacokinetic alterations caused by drug-drug interactions: metabolic interaction in the liver. *Pharmacol. Rev.*, *50*, 387–412.

Lu, A.Y., Wang, R.W., Lin, J.H. (2003). Cytochrome P450 in vitro reaction phenotyping: a re-evaluation of approaches used for P450 isoform identification. *Drug Metab. Dispos.*, *31*, 345–350. Review.

Madan, A., Usuki, E., Burton, L., Ogilvie, B., Parkinson, A. (2002). In vitro approaches for studying the inhibition of drug-metabolizing enzymes and identifying the drug-metabolizing enzymes responsible for the metabolism of drugs, in *Drug-Drug Interactions* (ed. A.D. Rodrigues), Marcel Dekker, New York, pp. 217–294.

Obach, R.S., Walsky, R.L., Venkatakrishnan, K., Gaman, E.A., Houston, J.B., Tremaine, L.M. (2006). The utility of in vitro cytochrome P450 inhibition data in the prediction of drug-drug interactions. *J. Pharmacol. Exp. Ther.*, *316*, 336–348.

Proctor, N.J., Tucker, G.T., Rostami-Hodjegan, A. (2004). Predicting drug clearance from recombinantly expressed CYPs: intersystem extrapolation factors. *Xenobiotica*, *34*, 151–178.

Rodrigues, A.D. (1999). Applications of heterologous expressed and purified human drug-metabolizing enzymes: an industrial perspective, in *Handbook of Drug Metabolism* (ed. T.F. Woolf), Marcel Dekker, New York, pp. 279–320.

Rodrigues, A.D. (2005). Impact of CYP2C9 genotype on pharmacokinetics: are all cyclooxygenase inhibitors the same? *Drug Metab. Dispos. 33*, 1567–1575.

Rodrigues, A.D., Rushmore, T. (2002). Cytochrome P450 pharmacogenetics in drug development: in vitro studies and clinical consequences. *Curr. Drug Metab.*, *3*, 289–309.

Shou, M., Lu, T., Krausz, K.W., Sai, Y., Yang, T., Korzekwa, K.R., Gonzalez, F.J., Gelboin H.V. (2000). Use of inhibitory monoclonal antibodies to assess the contribution of cytochromes P450 to human drug metabolism. *Eur. J. Pharmacol.*, *394*, 199–209.

Tang, W., Wang, R.W., Lu, A.Y. (2005). Utility of recombinant cytochrome p450 enzymes: a drug metabolism perspective. *Curr. Drug Metab.*, *6*, 503–517. Review.

Thummel, K.E., Shen, D.D. (2002). The role of the gut mucosa in metabolically based drug-drug interactions, in *Drug-Drug Interactions* (ed. A.D. Rodrigues), Marcel Dekker, New York, pp. 359–385.

Venkatakrishnan, K., von Moltke, L.L., Greenblatt, D.J. (2001). Application of the relative activity factor approach in scaling from heterologously expressed cytochromes p450 to human liver microsomes: studies on amitriptyline as a model substrate. *J. Pharmacol. Exp. Ther.*, *297*, 326–337.

# Nuclear Receptor-Mediated Gene Regulation in Drug Metabolism

HONGBING WANG and WEN XIE

## 1 INTRODUCTION

### 1.1 General Features of Nuclear Receptors

The mammalian nuclear receptor (NR) superfamily contains nearly 50 structurally related members, including both receptors for which ligands are known and "orphan receptors" for which there are, as yet, no known ligands (Laudet *et al.*, 1992; Giguere, 1999). Regulation of gene expression at the transcriptional level by NRs has an important role in both cellular development and the body's defensive systems, including drug metabolism and clearance (Kastner and Chambon, 1995; Mangelsdorf *et al.*, 1995; Honkakoski and Negishi, 2000; Moore *et al.*, 2000a). Members of the nuclear receptor family share several structural features. Two of the modulator domains are common structural motifs shared by most of the nuclear receptors (Fig. 1a). The highly conserved DNA-binding domain (DBD) that links the receptor to specific promoter regions of its target genes is termed hormone response elements (HREs) or xenobiotic response elements (XREs). As the most conserved domain of NRs, DBD can recognize the response elements that contain one or two consensus core half-sites related to the hexamer ACAACA (steroid receptors) or AGGTCA (estrogen receptors and other nuclear receptors) (Cairns *et al.*, 1991; Lee *et al.*, 1993). The HREs or XREs usually contain two NR half-sites of the consensus hexameric sequences, organized as inverted, everted, or direct repeats with a 3- to 6-base-pair spacing (Fig. 1b). The carboxy (C)-terminal portion of the receptor contains a less conserved ligand-binding domain (LBD), which interacts directly with hormones and xenobiotics. Remarkable variations in the shape and size of LBDs provide the structural basis for fitting a wide range of ligands with diverse chemical structures. It appears that the receptors that bind with high affinity to a specific ligand, such as estrogen receptor (ER) and thyroid hormone receptor (TR), have relatively small ligand-binding pockets, whereas the xenobiotic NRs, such as pregnane X receptor (PXR) and the peroxisome proliferator-activated receptors (PPARs), contain much larger and more flexible ligand-binding cavities that

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

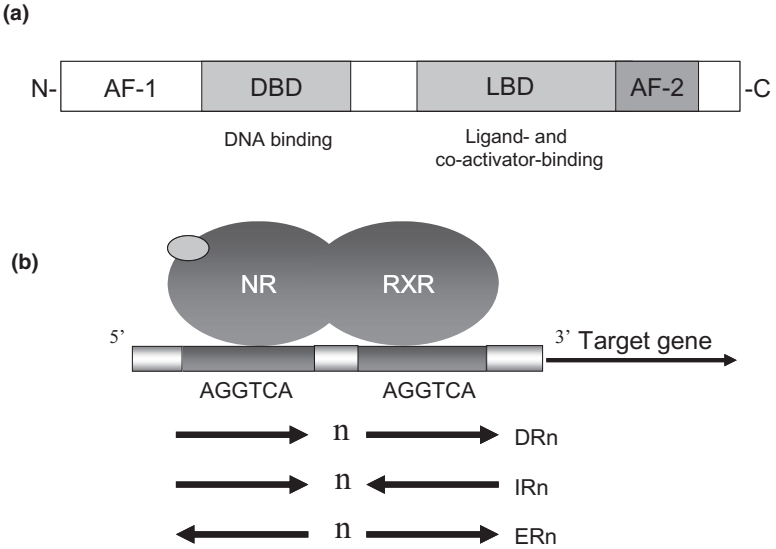


**Figure 1** Structure and functional domains of nuclear receptors. **(a)** Basal structure of nuclear receptors. AF-1 and DBD are at the amino (*N*)-terminal, and AF-2 and LBD are at the carboxy (*C*)-terminal. **(b)** Nuclear receptors can recognize and bind to AGGTCA half-sites spaced by several base pairs as direct repeats separate by n base pairs (DRn), inverted repeats separated by n base pairs (IRn), or everted repeats separated by n base pairs (ERn).

can accommodate the apparently more promiscuous ligands (Nolte *et al.*, 1998; Watkins *et al.*, 2001). Ligand binding induces significant conformational changes in the folding of the LBD, and leads to the recruitment or replacement of transcriptional cofactors, such as co-activators (steroid receptor co-activators, SRCs) and corepressors (nuclear corepressor, NCoR; silencing mediator for retinoid and thyroid hormone receptors, SMRT) (Freedman, 1999; Lanz *et al.*, 1999).

### 1.2 Nuclear Receptors and Drug-Metabolizing Enzymes

An important requirement for physiological homeostasis is the detoxification and removal of endogenous hormones and xenobiotic compounds with biological activities. Much of the detoxification and elimination is carried out by the phase I cytochrome P450 (CYP) enzymes, phase II conjugation enzymes, and "phase III" drug transporters. The products of phase I metabolism are generally more polar and readily excreted than the parent compounds. The phase I enzyme-mediated functionalization also often enables xenobiotics to become better substrates for the phase II conjugating enzymes. The phase II metabolism involves conjugation of hydrophilic moieties, such as sugar, glutathione, and sulfate, to further increase polarity and water solubility of xenobiotics, therefore promoting their excretion and elimination. In addition to phase I and II enzymes, equally important is a group of transporter proteins expressed in various tissues, such as the liver, intestine, brain, and kidney, which modulate the absorption, distribution, and excretion of many drugs.

As our understanding of nuclear receptors grows, it has become evident that nuclear receptors play important roles in the transcriptional regulation of these drug-metabolizing enzymes (DMEs) (Honkakoski and Negishi, 2000; Synold, Dussault and Forman, 2001; Wang and LeCluyse, 2003). The induction profiles of the major DMEs are remarkably linked to the activation of several orphan nuclear receptors. For instance, compounds such as rifampicin (RIF) and phenobarbital (PB) are PXR activators and inducers of CYP3A, CYP2B, UGT1A1, CYP2C9, and multidrug resistance protein 1(MDR1), while 1,4-bis[2-(3,5-dichloropyridyloxy)]benzene (TCPOBOP) and 6-(4-chlorophenyl:imidazo[2,1-b]thiazole-5carbaldehyde O-(3,4-dichlorobenzyl)oxime (CITCO) induce CYP2B, CYP3A, and UGT1A1 through the activation of constitutive androstane receptor (CAR) (Kliewer *et al.*, 1998; Geick, Eichelbaum and Burk, 2001; Ferguson *et al.*, 2002; Maglich *et al.*, 2003). This association of orphan nuclear receptors with particular DMEs not only demonstrates DME specificity, but also defines species-specific differences in the DME induction responses. Therefore, orphan nuclear receptor binding and activation assays have been utilized as useful tools for screening drug candidates *in vitro* for their potentials to cause drug–drug interactions (DDIs) via enzyme induction.

In the following section, we focus on recent advances in our knowledge and understanding of the regulation of DMEs by nuclear receptors. Although it has become increasingly evident that not only PXR and CAR but many other NRs may also be important in regulating the expression of drug disposition genes either directly, or by modulating the activities of PXR and CAR, specific emphasis has been placed on PXR and CAR because of their broad and overlapping xenobiotic specificity and their central role in regulating the expression of multiple DMEs and drug transporters. In addition, the role of a non-NR transcriptional factor, aryl hydrocarbon receptor (AhR), in xenobiotic regulation will also be briefly discussed.

## 2 PREGNANE X RECEPTOR

### 2.1 Cloning and Initial Characterization of PXR

In 1998, several groups published the cloning and initial characterization of the orphan nuclear receptor PXR. Kliewer and colleagues identified mouse PXR on the basis of its sequence homology with other NRs. It was named PXR based on its activation by various natural and synthetic pregnanes (Kliewer *et al.*, 1998). The cloning of human PXR (hPXR) was first reported by Blumberg and colleagues, and this group (led by Dr. Ronald M. Evans at the Salk Institute) initially named this receptor "steroid and xenobiotic receptor (SXR)" based on its activation by steroids and xenobiotics (Blumberg *et al.*, 1998). hPXR was also cloned by Bertilsson and colleagues and was alternately referred to as the pregnane-activated receptor (PAR) (Bertilsson *et al.*, 1998).

PXR shares a common NR modular structure with a conserved *N*-terminal DBD and a *C*-terminal LBD. The DBD contains two zinc-finger motifs, which mediate interaction with specific DNA sequences known as HREs. The LBD, in addition to determining ligand-binding specificity, contains a ligand-inducible transactivation function (AF2) and a motif that directs binding to the common heterodimerization partner retinoid X receptor (RXR). Although DBDs of the mammalian PXRs are highly conserved, sharing more than 95% amino acid identity, the LBDs of PXRs are much more divergent across species than those of other NRs. For example, the human

and rat PXR share only 76% amino acid identity in their LBD, whereas most human and rodent NR orthologs share more than 90% amino acid identity (Maglich *et al.*, 2001). The LBD sequence divergence of PXR between species is believed to be responsible for the species-specific response to ligands/drugs.

PXR is highly expressed in the liver and small intestine, where most DMEs are also highly expressed and induced. In rodents, lower levels of PXR mRNA have also been detected in the kidney, stomach, lung, uterus, ovary, and placenta. PXR can be activated by xenobiotics (including many clinical drugs) and endobiotics (such as bile acids and some hormones, including glucocorticoids, anti-glucocorticoids, and estrogens), which underlie the significance of PXR in both xenobiotic and endobiotic responses. PXR orthologs also exhibit interesting and striking species-dependent ligand specificity. For example, the anti-glucocorticoid pregnenolone-16α-carbonitrile (PCN) is an effective activator of mouse and rat PXRs, but has little effect on human PXR. In contrast, the antibiotic RIF can activate the human and rabbit PXRs, but not the mouse and rat PXRs. Detailed elaboration on the PXR ligand species specificity and its biological significance will be discussed later.

Like many other orphan nuclear receptors, activation of PXR by ligand binding leads to formation of PXR/RXR heterodimer and transcriptional activation of target genes by binding to specific PXR response elements (PXREs) in target gene promoters. The classical PXRE is DR-3 which is composed of two copies of direct repeat of the consensus AG(G/T)TCA NR binding motif and separated by three nucleotides. Other types of PXREs, including DR-4, DR-5, ER-6 (everted repeat spaced by 6-bp), ER-8, and IR-0 (inverted repeat without a spacing nucleotide), have also been shown to be bound by the PXR-RXR heterodimers (Kliewer, Goodwin and Willson, 2002; Sonoda *et al.*, 2003; Saini *et al.*, 2004; Xie *et al.*, 2004).

### 2.2 PXR in Phase I CYP Enzyme Regulation

The oxidative CYP enzymes catalyze the metabolic conversion of xenobiotics to polar derivatives that are more readily eliminated. Among the 78 mouse and 65 human CYP enzymes itemized within the current build of the UniGene sequence data bank, a significant number of these enzymes are known to be induced by various environmental pollutants, pharmaceuticals, or steroid metabolites. A more current list of CYPs is available at http://drnelson.utmem.edu/CytochromeP450.html. Among CYP enzymes, the CYP3A isozymes are of particular medical significance since they are involved in the metabolism of more than 50% of clinical drugs as well as nutraceuticals and herbal medicines (Maurel *et al.*, 1996). In 1998, PXR was isolated as a candidate xenobiotic receptor (xenosensor) postulated to regulate CYP3A gene expression (Bertilsson *et al.*, 1998; Blumberg *et al.*, 1998; Kliewer *et al.*, 1998; Lehmann *et al.*, 1998). PXR regulates the expression of both human and rodent CYP3A genes by directly binding to the XREs ER-6 and DR-3 localized in the promoter regions of the human and rodent CYP3A genes, respectively. Both "loss-of-function" gene knockout and "gain-of-function" transgenic mouse studies have provided convincing genetic and pharmacological evidence to support the role of PXR in CYP3A-mediated xenobiotic responses. Targeted disruption of the mouse PXR locus abolished the CYP3A xenobiotic response to prototypic inducers such as PCN and dexamethasone (Xie *et al.*, 2000; Staudinger *et al.*, 2001). In contrast, hepatic expression of an activated form of hPXR in transgenic mice resulted in sustained induction of CYP3A enzymes and enhanced protection against xenobiotic toxicants, such as zoxazolamine and tribromoethanol (Xie *et al.*, 2000).

In addition to CYP3A isozyme, PXR has also been shown to regulate other CYP enzymes, including CYP2B6 (Honkakoski *et al.*, 1998; Goodwin *et al.*, 2001; Dussault *et al.*, 2003; Wang and Negishi, 2003), CYP2B9 (Dvorak *et al.*, 2003), CYP2C8 (Ferguson *et al.*, 2005), and CYP2C9 (Dvorak *et al.*, 2003).

### 2.3 PXR in Phase II Enzyme Regulation

The phase II conjugating enzymes include broad specificity transferases, such as UDP-glucuronosyltransferase (UGT), glutathione *S*-transferase (GST), and sulfotransferases (SULT) (Hayes and Pulford, 1995; Salinas and Wong, 1999; Nagata and Yamazoe, 2000; Tukey and Strassburg, 2000). The phase II conjugation has a dual role in drug metabolism. First, conjugation reactions not only terminate the ability of electrophiles to react with DNA and proteins, they also prevent nucleophiles from interacting with receptor proteins. At the same time, conjugation increases the water solubility of the compounds, which in turn, promotes renal and biliary excretion (Sheweita, 2000). Therefore, the phase II reactions play a critical role in detoxifying both exogenous and endogenous chemicals.

***2.3.1 PXR in UGT Regulation*** Glucuronidation, a major metabolic pathway for many endo- and xenobiotics, is catalyzed by enzymes belonging to the family of membrane-bound UGTs. Using UDP-glucuronic acid as a sugar donor, UGTs catalyze the transfer of glucuronic acid to a variety of substrates and thus convert small lipophilic molecules to water-soluble glucuronides. UGT-mediated glucuronidation functions as the principle means to eliminate steroid, heme metabolites, environmental toxin, and drugs from the body (Mackenzie *et al.*, 1997, 2003; Radominska-Pandya *et al.*, 1999; Tukey and Strassburg, 2000). Among UGTs, UGT1A1 is one of the most characterized UGT isoform. UGT1A1 is the critical enzyme responsible for the detoxification of bilirubin. Deficiency in the expression and/or activity of UGT1A1 in patients can lead to severe accumulation of unconjugated bilirubin, medically termed "hyperbilirubinemia," a hallmark of the Crigler–Najjar (CN) syndrome (Tukey and Strassburg, 2000).

Our own study showed that UGT1A1 is under the positive control of PXR. UGT1A1 mRNA and protein expression was up-regulated in transgenic mice that express the activated hPXR (VP-hPXR) and in RIF-treated "humanized" hPXR transgenic mice (Xie *et al.*, 2003). The microsomal glucuronidation activity toward β-estradiol (a UGT1A1 substrate), thyroid hormones, corticosterone, and xenobiotics (such as 4-nitrophenol and 4-OH-PhIP), was also increased in the VP-hPXR transgenic mice. Chen and colleagues showed that PCN, a rodent-specific PXR agonist, could induce UGT expression and increase UGT enzymatic activity in wild-type mice and this induction was abolished in PXR-null mice (Chen, Staudinger and Klaassen, 2003). The identification of a DR3-like PXR responsive element (GGTTCATAAAGGGTA) in the human UGT1A1 promoter further established UGT1A1 as a direct target gene of PXR. In addition to UGT1A1, the expression of UGT1A9, but not UGT1A2 and 2B5, was also increased by PXR activation, suggesting that the UGT induction by PXR is isoform-specific (Chen, Staudinger and Klaassen, 2003).

In addition to PXR, the expression of UGT1A1 and several other UGT isoforms have also been reported to be regulated by several other nuclear receptors, including CAR (Sugatani *et al.*, 2001; Huang *et al.*, 2003; Xie *et al.*, 2003) and PPARα (Barbier *et al.*, 2003; Ikeda *et al.*, 2005).

***2.3.2 PXR in SULT Regulation*** The cytosolic sulfotransferases (SULTs) are another important family of phase II enzymes that catalyze the conjugation of nucleophilic compounds. SULTs catalyze the transfer of a sulfonyl group from a sulfate donor 3′-phosphoadenosine 5′-phosphosulfate (PAPS) to hydroxyl or amino groups of acceptor molecules, forming sulfate or sulfamate conjugates. SULTs are specific to sulfonate small lipophilic molecules, such as steroids, bioamines, and therapeutic drugs. It was believed that, although SULTs and UGTs have similar substrate spectrum, SULTs have higher affinities but lower turnover rates compared with UGTs (Morris and Pang, 1987). Thus, UGT activities are likely to play a more significant role when substrates are abundant due to the higher turnover rate of UGTs, whereas SULTs may play a leading role in cases where concentrations of the substrate are lower. SULT-mediated sulfonation is known to play a significant role in the homeostasis of sex hormones that are present in low concentrations in plasma (Qian *et al.*, 2001).

The expression of rodent hepatic Sult2a9, also called Sult2a1 or dehydroepiandrosterone sulfotransferase (DHEA SULT or STD), is subjected to transcriptional regulation by PXR and an IR-0 element located in the 5′-flanking region of rodent Sult2a gene is required for its activation by PXR (Sonoda *et al.*, 2003). The expression of Sult2a9/2a1/STD has also been shown to be regulated by FXR (Song *et al.*, 2001), CAR (Saini *et al.*, 2004), and LXR (Uppal *et al.*, 2007). Interestingly, all four receptors share the same IR-0 response element to regulate the expression of this SULT isoform. The hierarchy and relative contribution of individual nuclear receptors in Sult2a regulation remain to be determined.

***2.3.3 PXR in GST Regulation*** GSTs are soluble homo- and heterodimeric enzymes that use reduced glutathione in conjugation and reduction reactions. GSTs play an important role in the conjugation of electrified chemicals. Since chemical carcinogens are often highly nucleophilic, they also represent potential substrates for GST. For this reason, GSTs are believed to play an important role to protect cells from genotoxic compounds. The cytosolic GST isozymes of rodents and humans can be grouped into several classes, such as Alpha, Mu, Pi, Theta, Omega, and Zeta, based on their amino acid sequences, immunological properties, and substrate specificities (Falkner *et al.*, 2001; Townsend and Tew, 2003). The Alpha, Mu, and Pi classes are the most abundantly expressed GSTs.

GST regulation by PXR was first hinted by several studies of general profiling of gene expression (Maglich *et al.*, 2002; Rosenfeld *et al.*, 2003). It was also reported that GSTA2 expression was induced by PXR and an IR-6 response element was responsible for this transcriptional regulation (Falkner *et al.*, 2001). A more systemic and comparative analysis of GST regulation by PXR was reported in transgenic mice that bear the expression of activated PXR in the liver and intestine (Gong *et al.*, 2006). In this study, it was shown that the expressions of GST Alpha, Pi, and Mu classes are all under the control of PXR. Interestingly, PXR-mediated GST regulation exhibited clear isoform-, tissue-, and gender-specificity: (i) GST Mu was the only isoform that was up-regulated by PXR in both liver and intestine in both sexes; (ii) GST Alpha was induced in the small intestine, but not in the liver; and (iii) PXR had an opposite effect on hepatic GST Pi expression, inducing this class in females, but suppressing it in males.

Paradoxically, although the overall GST expression and activity were increased, activation of PXR sensitized the response to the oxidative xenotoxicant paraquat *in vivo* and in cultured cancer cells. Moreover, heightened paraquat sensitivity in transgenic mice was female-specific. Whether the PXR-mediated, gender-specific GST regu-

lation accounts for the intact paraquat sensitivity in transgenic males remains to be determined. Nevertheless, the regulation of GSTs by PXR suggests that this regulatory pathway may be relevant to carcinogenesis by sensitizing normal and cancerous tissues to oxidative cellular damage (Gong *et al.*, 2006).

### 2.4 PXR in Drug Transporter Regulation

In addition to regulating phase I CYP enzymes and phase II conjugating enzymes, PXR also participates in the regulation of drug transporters. Drug transporters are responsible for uptake and efflux of endogenous and exogenous chemicals, including many clinical prescribed drugs, thus the expression and activity of transporters affect drug efflux and clearance (Klaassen and Slitt, 2005). Organic anion-transporting polypeptides (OATPS) are a family of major uptake drug transporters in the liver. OATPs are localized to the basolateral membrane of hepatocytes and transport compounds, including organic anions, organic cations, and neutral compounds, from blood into hepatocytes. The OATP family includes 9 human, 13 rat, and 15 mouse members (Hagenbuch and Meier, 2003). PXR ligand PCN administration induced Oatp2 mRNA expression in the livers of wild-type mice, but not PXR-null mice (Staudinger *et al.*, 2003). A DR-3 type PXR response element has been identified in the rat Oatp2 promoter (Guo *et al.*, 2002).

Multidrug resistance-associated protein (Mrps) are efflux transporters for structurally diverse amphipathic chemicals and organic anions. The liver-enriched Mrp2 is localized to the canalicular membrane of hepatocytes and is responsible for the hepatobiliary excretion of amphipathic anions (Kast *et al.*, 2002; Kruh and Belinsky, 2003). Natural mutations in MRP2 cause Dubin–Johnson syndrome/hyperbilirubinemia II, a disorder characterized by impaired transfer of anionic conjugates into the bile. Induction of Mrp2 by PXR ligand PCN and dexamethasone was observed in primary hepatocytes isolated from the livers of wild-type mice, but not from PXR-null mice, implicating a role of PXR in the regulation of mouse Oatp2 (Kast *et al.*, 2002). An ER-8 type of NR response element was identified in the promoter of the rat Mrp2 gene that confers the induction of Mrp2 by PXR (Kast *et al.*, 2002). Mrp3 was also reported to be induced by PXR (Teng, Jekerle and Piquette-Miller, 2003). In the intestine, PXR has been shown to stimulate the expression of Mdr1, which encodes an ATP-dependent efflux pump that transports a wide variety of xenobiotics, including many widely used prescription drugs. In humans, activation of PXR in the intestine may decrease intestinal drug absorption by increasing the expression of *Mdr1*. A DR-4 type PXR response element was identified in *Mdr1* gene promoter (Geick, Eichelbaum and Burk, 2001).

### 2.5 Implications of PXR-Mediated Gene Regulation in Drug Metabolism

The implications of PXR-mediated gene regulation in drug metabolism and drug interaction have been recognized since the initial cloning of this receptor. Our body encounters numerous xenobiotic chemicals, including prescription drugs, over-the-counter medications, and herbal medicines. Many of them, especially when accumulated in excess, may exert toxic effects through various mechanisms. The process of drug metabolism is known to largely depend upon a concerted action of phase I and II enzymes, as well as drug transporters. Expressed mainly in the liver and intestine, these enzymes and transporters are capable of recognizing an amazing diversity of xenobiotics to promote their clearance.

The regulation of DMEs by PXR is involved in clinical DDIs, in which one drug accelerates the metabolism of a second medicine and may change or cause adverse results. Because CYP enzymes can recognize a large spectrum of pharmaceutical substrates, a CYP gene-inducing drug is potentially capable of affecting the metabolism and clearance of any co-consumed drugs. As mentioned earlier, the identification of RIF as a potent hPXR agonist has provided an explanation why this antibiotic drug is prone to DDIs. In another example, St. John's wort (SJW), a popular herbal remedy for mild depression, has been reported to trigger severe adverse interactions with several clinical drugs, such as oral contraceptives, the HIV protease inhibitor indinavir, and the immunosuppressant cyclosporin. Such DDIs are likely results of activation of PXR and consequent induction of CYP3A by SJW and the subsequent increased metabolism and/or decreased bioavailability of co-metabolized drugs. In the case of birth control pills, the use of SJW enhances drug clearance, increasing contraceptive failure and thus the birth of "miracle babies." The identification of SJW as a PXR agonist offered a plausible explanation for the propensity of SJW to cause DDIs (Moore *et al.*, 2000b).

Several traditional Chinese medicines (TCMs) have also been implicated in DDIs. TCMs are essential components of alternative medicines. One clinical concern of herbal product use is the effect of herbal products on the metabolism of coadministered drugs. We showed that two TCM herbs Wu Wei Zi (*Schisandra chinensis Baill*) and Gan Cao (*Glycyrrhiza uralensis Fisch*), can activate PXR and induce the expression of several DMEs and transporters, including CYP3A and 2C9 and the multidrug resistance-associated protein 2 (MRP2) in reporter gene assay and in primary hepatocyte cultures (Mu *et al.*, 2006). The anticoagulant warfarin is known to be metabolized by CYP2C9 in humans (Goldstein, 2001). As expected, administration of Wu Wei Zi and Gan Cao extracts in rats resulted in an increased metabolism of coadministered warfarin, reinforcing concerns involving the safe use of herbal medicines and other nutraceuticals to avoid PXR-mediated DDIs (Mu *et al.*, 2006).

Having known the potential of PXR-activating agents in causing DDIs, it is also important to emphasize that PXR activation alone may not be sufficient to predict the propensity of DDIs. Sinz and colleagues recently published the evaluation of 170 xenobiotics in an hPXR transactivation assay and compared these results with known clinical DDIs. Of the 170 xenobiotics tested, 54% of them demonstrated some level of hPXR transactivation. However, by taking into consideration cell culture conditions (solubility, cytotoxicity, appropriate drug concentration in media), as well as *in vivo* pharmacokinetics (therapeutic plasma concentration or $C_{max}$, distribution, route of administration, dosing regimen, liver exposure, potential to inhibit CYP3A4), the risk potential of CYP3A4 enzyme induction for most compounds reduced dramatically. By employing this overall interpretation strategy, the final percentage of compounds predicted to significantly induce CYP3A4 reduced to 5%, all of which are known to cause DDIs in the clinic (Sinz *et al.*, 2006).

## 2.6 Endobiotic Function of PXR

Even though PXR was initially identified as a "xenobiotic receptor," emerging evidence has pointed to an equally important role of PXR as an "endobiotic receptor" that responds to a wide array of endogenous chemicals (endobiotics), such as bile acids and their intermediates, as well as certain steroid hormones. Another mechanism for PXR's involvement in endobiotic homeostasis is that many endobiotics, such as bile acids,

bilirubin, and adrenal steroids, are also substrates of PXR target enzymes and transporters. Activation of PXR by endogenous or xenobiotic ligands has implications in several important physiological and pathological conditions. For this reason, there have been extensive discussions on whether or not PXR can be explored as a therapeutic target (Xie *et al.*, 2004; Gong *et al.*, 2005).

***2.6.1 PXR in Bile Acid Detoxification and Cholestasis*** One family of endogenous PXR ligands identified shortly after the cloning of PXR is bile acid. Bile acids are catabolic end products of cholesterol metabolism. They are physiologically important in the formation of bile and solubilizing biliary lipids and promoting their absorption. However, excessive bile acids are potentially toxic. For example, the secondary bile acid lithocholic acid (LCA) has been shown to cause cholestasis in experimental animals and has long been suspected of doing the same in humans. As an average human releases 600 mL of bile a day, the potential for disrupting bile flow (cholestasis) and the resultant accumulation of toxic by-products is significant. Therefore, excess bile acid should be efficiently eliminated to avoid the toxic effect.

PXR has been demonstrated to acts as an LCA sensor and plays an essential role in detoxification of cholestatic bile acids (Staudinger *et al.*, 2001; Xie *et al.*, 2001). Studies in different animal models showed that activation of PXR protected against severe liver damage induced by LCA. Pretreatment of wild-type mice, but not of the PXR-null mice, with PCN reduced the toxic effects of LCA. Moreover, genetic activation of PXR by expressing the activated PXR in the liver of transgenic mice was sufficient to confer resistance to the hepatotoxicity of LCA. The cholestatic preventive effect of PXR was initially reasoned to be due to the activation of CYP3A, an important CYP enzyme responsible for bile acid hydroxylation (Staudinger *et al.*, 2001; Xie *et al.*, 2001). Subsequent identification of SULT2A, a bile acid detoxifying hydroxysteroid sulfotransferase, as a PXR target gene suggested that additional PXR target genes may have also contributed to the phenotype (Sonoda *et al.*, 2002). Several follow-up studies, including those using mice with individual or combined loss of PXR and CAR, have suggested that PXR-responsive bile acid transporter regulation may also play a role in preventing cholestasis (Zhang *et al.*, 2004; Stedman *et al.*, 2005; Uppal *et al.*, 2005). Since activation of PXR was sufficient to prevent cholestasis, it has been suggested that PXR agonists may prove useful in the treatment of human cholestatic liver disease, a notion that has been supported by several clinical observations. Both RIF and SJW have been empirically used to treat cholestatic liver diseases (Kliewer, Goodwin and Willson, 2002). The relief from cholestasis-associated pruritus and amelioration of cholestasis by RIF was associated with increased 6α-hydroxylation of bile acids, which in turn facilitates glucuronidation by the UGTs at the 6α-hydroxy position. Both RIF and SJW are potent agonists of hPXR and both CYP3A and UGT are PXR target genes, suggesting the anti-cholestatic effects of RIF and SJW are mediated by PXR.

***2.6.2 PXR in Bilirubin Detoxification and Clearance*** Bilirubin is the catabolic by-product of heme proteins, such as β-globin and CYP enzymes. Accumulation of bilirubin in the blood is potentially hepatotoxic and neurotoxic. For example, an insufficiency in expression of UGT1A1, a key enzyme for the conjugation of bilirubin, in the Crigler–Najjar syndrome and Gilbert's diseases results in severe hyperbilirubinemia. Deficiency of MRP2, a drug transporter responsible for the hepatic excretion of conjugated bilirubin, leads to Dubin–Johnson syndrome, characterized by the

accumulation of glucuronidated bilirubin. PXR has been shown to induce the expression of multiple key components in the clearance pathway, including UGT1A1, OATP2, GSTA1 and 2, and MRP2. OATP2 facilitates bilirubin uptake from blood into hepatocytes (Klaassen and Slitt, 2005). GSTA1 and 2 reduce bilirubin back efflux from hepatocytes into blood. MRP2 promotes the canalicular efflux of conjugated bilirubin. Consistent with the pattern of gene regulation, activation of PXR in transgenic mice has been shown to prevent experimental hyperbilirubinemia (Xie *et al.*, 2003).

***2.6.3 PXR in Adrenal Steroid Homeostasis and Drug–Hormone Interactions*** PXR plays an important endobiotic role in adrenal steroid homeostasis. Our recent study showed that genetic (by using the VP-hPXR transgene) and pharmacological (by using the hPXR ligand rifampicin) activation of hPXR in mice markedly increased plasma concentrations of corticosterone and aldosterone, the respective primary glucocorticoid and mineralocorticoid in rodents. The increased levels of corticosterone and aldosterone were associated with activation of adrenal steroidogenic enzymes, including *CYP11a1, CYP11b1, CYP11b2*, and *3β-Hsd*. The PXR-activating transgenic mice also exhibited hypertrophy of the adrenal cortex, loss of glucocorticoid circadian rhythm, and lack of glucocorticoid responses to psychogenic stress (Zhai *et al.*, 2007).

Interestingly, the VP-hPXR transgenic mice had normal pituitary secretion of adrenocorticortropic hormone (ACTH) and the corticosterone suppressing effect of dexamethasone (DEX) was intact, suggesting a functional hypothalamus–pituitary–adrenal (HPA) axis despite a severe disruption of adrenal steroid homeostasis. The ACTH-independent hypercortisolism in the PXR-activating transgenic mice is reminiscent of the pseudo-Cushing's syndrome in patients, the clinical hallmark of which is the normal DEX suppression despite a high circulating level of glucocorticoid. Pseudo-Cushing's syndrome is most seen in alcoholic, depressed, or obese subjects. It is of interest to know whether or not these susceptible patients are associated with increased expression and/or activity of PXR. The glucocorticoid effect appeared to be PXR-specific, as the activation of CAR in transgenic mice had little effect. We propose that PXR is a potential endocrine-disrupting factor that may have broad implications in steroid homeostasis and drug–hormone interactions (Zhai *et al.*, 2007).

***2.6.4 PXR in Lipid Metabolism*** PXR has also recently been shown to play an endobiotic role by impacting lipid homeostasis (Zhou *et al.*, 2006). Expression of an activated PXR in the livers of transgenic mice resulted in an increased hepatic deposit of triglycerides. This PXR-mediated lipid accumulation was independent of the activation of the lipogenic transcriptional factor sterol regulatory element-binding protein 1c (SREBP-1c) and its primary lipogenic target enzymes, including fatty acid synthase (FAS) and acetyl CoA carboxylase 1 (ACC-1). Instead, the lipid accumulation in transgenic mice was associated with an increased expression of the free fatty acid transporter CD36 and several accessory lipogenic enzymes, such as stearoyl CoA desaturase-1 (SCD-1) and long chain free fatty acid elongase (FAE). Studies using transgenic and knockout mice showed that PXR is both necessary and sufficient for *Cd36* activation. Promoter analyses revealed a DR-3 type of PXR response element in the mouse *Cd36* gene promoter, establishing *Cd36* as a direct transcriptional target of PXR. The hepatic lipid accumulation and *Cd36* induction was also seen in the hPXR "humanized" mice treated with the hPXR agonist rifampicin. The activation of PXR was also associated with an inhibition of pro-β-oxidative genes, such as peroxisome proliferators-activated receptor α (PPARα) and thiolase, and an up-regulation of PPARγ, a positive regulator

of CD36. The cross-regulation of CD36 by PXR and PPARγ suggests that this fatty acid transporter may function as a common target of orphan nuclear receptors in their regulation of lipid homeostasis.

## 2.7 Species Specificity of PXR and the Creation of "Humanized" Mice

***2.7.1 Challenges for Rodents as Drug Metabolism Models*** Primary hepatocytes are a common *in vitro* system to evaluate metabolism, toxicity, and enzyme induction (Tucker *et al.*, 2001; Weaver, 2001; Weaver *et al.*, 2001). Human hepatocytes are considered the most relevant system to evaluate or predict human metabolism or effects of a new drug. A significant disadvantage of this system is the lack of routine availability of good quality human liver tissues or cells. Other model systems that may provide more consistent access include immortalized hepatocytes or humanized animal models (Xie *et al.*, 2000; Mills *et al.*, 2004). Over the years, it has been perceived that rodent models have limited utility in predicting drug-related human effects due to significant species differences in DMEs, transporters, and nuclear hormone receptors. For example, PCN is an effective CYP3A inducer in rodents but not in humans, and rifampicin induces CYP3A in humans but not in rats (Kocarek *et al.*, 1995). These findings have been attributed to the species differences in the effect of several drugs on CYP3A expression mediated by PXR (Kocarek *et al.*, 1995; Jones *et al.*, 2000). These differences across species demand the development of humanized animal models to evaluate the potential effect of a chemical in humans using an animal model.

***2.7.2 Species Specificity of the Rodent and Human PXR*** Both hPXR/SXR and mPXR are highly expressed in the liver and small intestine and share many functional properties, in particular, the regulation of *CYP3A* genes. However, as discussed earlier, these two orthologs are pharmacologically distinct in that strong activators of one receptor are often poor activators of the other. This species-specific ligand profile is reflected by the sequence divergence in the LBDs of the mouse and human receptors. The crystal structure of the hPXR LBD has been solved (Watkins *et al.*, 2001). The hydrophobic ligand-binding cavity of hPXR is composed of a large, smooth surface containing only a small number of polar residues, suggesting that it is not necessary for activators to conform to a restricted orientation. Based on site-directed mutational analysis, the position and nature of these polar residues were found to be crucial for establishing the precise pharmacological activation profile of hPXR (Watkins *et al.*, 2001). Indeed, conversion of four amino acids of mouse PXR that correspond to the hPXR-specific activator SR12813-interacting residues in hPXR produces a hybrid mouse–human PXR that was no longer activated by PCN and was only weakly activated by SR12813 in reporter assays (Watkins *et al.*, 2001). The structural and pharmacological differences between hPXR and mPXR and that of other species might reflect the difference in the diets of rodents and primates and the evolutionary need to respond to a different set of ingested nutrients and xenobiotics.

***2.7.3 Creation and Characterization of the hPXR "Humanized" Mice*** Based on the fact that the species origin of the receptor is the determining factor for the ligand specificity between species, we have created transgenic mouse models to determine whether the human receptor is sufficient to establish a human response profile (Xie *et al.*, 2000) (Fig. 2). First, hPXR transgenic mice were generated by expressing the hPXR in the mouse liver. The liver-specific expression of the transgene was

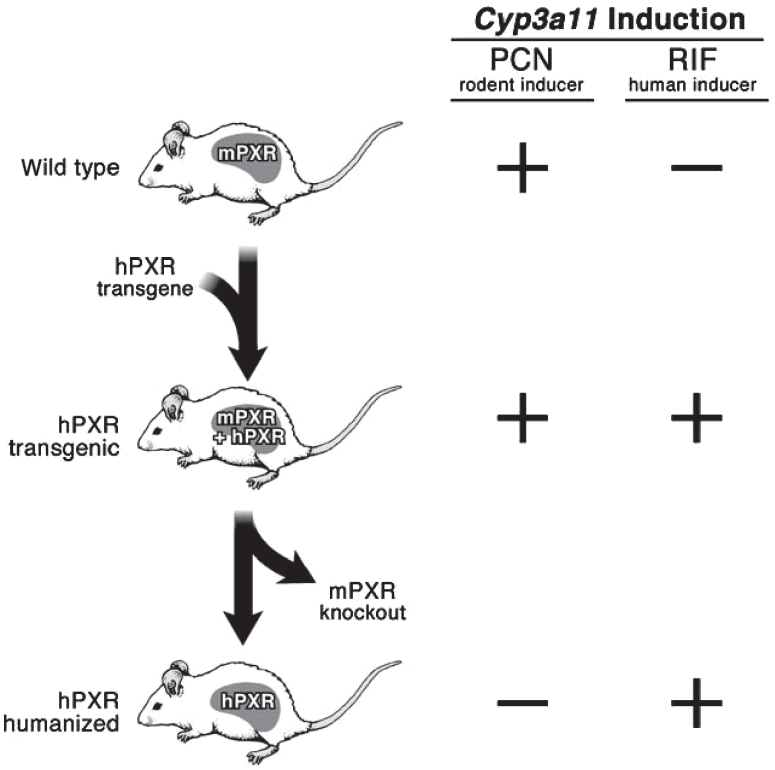


**Figure 2** Schematic representation of the creation of hPXR "humanized" mice. The humanization was achieved in the liver only when the liver-specific albumin promoter was used to direct the transgene expression, or in both the liver and intestine when the fatty acid-binding protein promoter was used. PCN, pregnenolone-16α-carbonitrile; RIF, rifampicin. "+" and "–" mean induction and lack of induction, respectively.

accomplished by using the mouse albumin promoter. Because the resulting mice harbor both mPXR and hPXR in their livers, the transgenic mice exhibited a chimeric or combined CYP3A response to both the rodent-specific inducer PCN and the human-specific inducer rifampicin. These results imply that mice only expressing hPXR could be fully humanized for the xenobiotic response. These animals were created by breeding the hPXR transgene into the mPXR knockout background. In contrast to the null mice that are devoid of CYP3A induction by steroids, replacement of mPXR with transgenic hPXR restores xenobiotic regulation with a humanized response profile. These mice readily responded to human inducers, such as rifampicin, in the equivalent range of the standard oral dosing regimen in humans (300–600 mg per 70 kg man) and exhibited similar pharmacokinetics of CYP3A regulation (Kolars *et al.*, 1992; Xie *et al.*, 2000). A "fully" human profile of CYP3A inducibility is obtained in the mPXR null/hPXR transgenic mice. Therefore, these experiments provide compelling evidence that PXR functions as a species-specific xenosensor mediating the adaptive hepatic response. This is also one of the rare examples in which replacing a single transcriptional regulator enables conversion of species-specific gene regulation.

The original "humanized" mice bear the expression hPXR exclusively in the liver (Xie *et al.*, 2000). Since both the DMEs and xenobiotic receptors are also highly expressed in the intestinal tracts, it is conceivable that mouse models with the humanized receptors expressed in both the liver and intestine would represent a more complete humanized mouse model. The liver and intestine dual humanization has been achieved by using the fatty acid-binding protein (FABP) gene promoter that targets the expression of hPXR transgene to both the liver and intestine (Zhou *et al.*, 2006). An alternative strategy is to "knock-in" hPXR in the mouse locus. This would not only direct expression of hPXR in both liver and intestine, but also normalize expression levels and tissue patterns to the endogenous gene.

***2.7.4 Significance of Humanized Mice in Drug Metabolism Studies and Drug Development*** The creation of mouse models with humanized xenobiotic response may aid pharmaceutical development by predicting potential DDIs (Moore and Kliewer, 2000; Kliewer, Goodwin and Willson, 2002; Sonoda *et al.*, 2003; Xie *et al.*, 2004). For decades, rodent models have been standard components in the assessment of potential toxicity for the development of candidate human drugs. However, their reliability as predictors of the human xenobiotic response is limited due to the species specificity of the xenobiotic response. To date, there has been no reliable system outside of humans to directly and quantitatively assess the DDIs. Primary cultures of human hepatocytes are valuable. However, since the hepatocytes are from individual patients, the utility of human hepatocytes is compromised by interindividual variability, limited tissue resources, and high cost. The humanized mice exhibited a "humanized" hepatic xenobiotic response profile, readily responding to the human-specific inducer RIF in a concentration range equivalent to the standard oral dosing regimen in humans (Xie *et al.*, 2000). The creation of these mice represents a major step toward generating a humanized rodent toxicological model that is continuously renewable and completely standardized. In addition, a PXR-mediated and mechanism-based transfection and reporter gene system has also been shown to be an effective *in vitro* approach to screen for drugs that may be precocious hPXR activators. While the *in vitro* screen is fast, the availability of hPXR "humanized" mice offers a unique screening tool to evaluate DDIs *in vivo*. The humanized mouse models represent important steps in the development of safer human drugs.

Another benefit of the hPXR humanized mice is its utility as a pharmacological model to dissect the function of PXR *in vivo*. The PXR ligand effect on gene expression and/or pathophysiological outcome may require long-term drug treatment. Although PCN is a potent rodent PXR agonist, the wild-type mice cannot tolerate a long-term treatment of PCN due to its toxicity. In this case, the use of humanized mice become necessary since mice can tolerate chronic treatment of RIF, the hPXR agonist. We have successfully used the humanized mice in our recent studies of PXR effect on lipogenesis and adrenal steroid homeostasis (Zhou *et al.*, 2006; Zhai *et al.*, 2007).

## 3 CONSTITUTIVE ANDROSTANE/ACTIVATED RECEPTOR (CAR)

### 3.1 Identification of CAR as the Regulator of CYP2B Genes

Originally termed as MB67 or CARα, the orphan nuclear receptor CAR (NR1I3) was isolated by screening a human liver library with degenerate oligonucleotide probes in

1994 (Baes *et al.*, 1994). The name CAR was initially defined as constitutively activated receptor, because it forms a heterodimer with RXR that binds to retinoic acid response elements (RAREs) and transactivates target genes in the absence of ligands in transfection assays (Baes *et al.*, 1994; Choi *et al.*, 1997). Using human (h) CAR cDNA as probe, the mouse counterpart was isolated subsequently, and was designated CARß (Choi *et al.*, 1997). Recently, two androstane metabolites, androstanol (5α-androstan-3α-ol) and androstenol (5α-androstan-16-en-3α-ol), were found to be endogenous antagonists of mouse CAR, giving an alternative name of CAR as constitutive androstane receptor. As discussed in the following contents, a breakthrough linkage of CAR with CYP2B genes by Negishi and colleagues has triggered a wealth of subsequent studies focusing on the role of CAR on multiple DME and transporter gene expression (Sueyoshi *et al.*, 1999; Xie and Evans, 2001; Ferguson *et al.*, 2002; Wang and Negishi, 2003).

Xenobiotic induction of CYP2B genes in different species has been realized for a long time. Through a series of elegant studies, a specific sequence termed PB-responsive element (PBRE) or PB-responsive unit (PBRU), or later delineated as PB-responsive enhancer module (PBREM), has been identified as the key sequences that are responsible for xenobiotic induction of CYP2B genes (Trottier *et al.*, 1995; Park, Li and Kemper, 1996; Honkakoski and Negishi, 1997; Honkakoski *et al.*, 1998). However, substantial progress in our understanding of PBREM activation by PB was obtained upon identification of CAR as the predominant regulator of PBREM activation. Using cell-based transfection assays, a series of known nuclear receptors, including RXR, CAR, liver X receptor (LXR), TR, hepatocyte nuclear factor (HNF4), and chicken ovalbumin upstream promoter-transcription factor (COUP-TF), were screened for their ability to transactivate mouse PBREM in HepG2 and HEK293 cells (Honkakoski *et al.*, 1998). The results revealed that only liver-enriched mouse (m) CAR was able to activate PBREM-driven reporter gene expression. In HepG2 cells transfected with mCAR, co-transfected PBREM was activated in the absence of PB, consistent with constitutive activation of retinoic acid responses by CAR in the absence of retinoic acid (Honkakoski and Negishi, 1998; Sueyoshi *et al.*, 1999). Co-transfection of RXR with mCAR in HepG2 cells resulted in synergistic augmentation of PBREM reporter activity, indicating that CAR activates PBREM by forming a heterodimer with RXR. Electrophoretic mobility shift assays (EMSAs) confirmed that CAR bound to the NR1 and NR2 motifs of the PBREM in the presence but not in the absence of RXR (Honkakoski *et al.*, 1998). The protein that bound to the PBREM in liver of PB-treated mice was purified using NR1-affinity chromatography and identified as mCAR using binding assays and immunoblotting analysis (Honkakoski *et al.*, 1998). Subsequently, a group of structurally diverse CYP2B inducers was shown to activate the PBREM reporter gene via mCAR. A HepG2 stable cell line designated as g2car-3, which constitutively expresses mCAR, activates transfected PBREM as well as the endogenous *CYP2B6* gene (Sueyoshi *et al.*, 1999; Sugatani *et al.*, 2001). In addition, treatment of g2car-3 cells with mCAR activators such as TCPOBOP and PB could effectively reactivate PBREM activity and *CYP2B6* gene expression initially repressed by androstanol pretreatment. Definite evidence establishing the role of CAR in the regulation of PB induction of the *CYP2B* genes was obtained from experiments using CAR-knockout mice (Wei *et al.*, 2000; Ueda *et al.*, 2002). Two independent mouse lines with targeted ablation of the CAR gene were generated by replacing a segment of the CAR gene that includes part of the DBD, with the coding region of b-galactosidase. As expected, CAR expression was absent in CAR–/– mice, while no overt phenotype was

observed in either male or female mice (Wei *et al.*, 2000). Interestingly, induction of CYP2b10 mRNA by PB and TCPOBOP was abolished completely in the CAR–/– homozygotes.

### 3.2 CAR in Other DME Regulation

***3.2.1 CAR in CYP2Cs Regulation*** The cytochrome P450 subfamily CYP2C enzyme is responsible for the metabolism of approximately 20% of therapeutic drugs and many endogenous compounds in humans (Goldstein, 2001). Induction of human CYP2Cs can result in drug tolerance as well as DDIs. Recently, a series of studies by Goldstein and coworkers demonstrated that xenobiotic induction of CYP2Cs is collaboratively mediated by several NRs, including CAR, PXR, GR, and HNF4α (Ferguson *et al.*, 2002; Chen *et al.*, 2003, 2004). In primary cultured human hepatocytes, CYP2C9 is inducible by xenobiotics including PB, RIF, and dexamethasone (Gerbal-Chaloin *et al.*, 2001; Rae *et al.*, 2001; Raucy *et al.*, 2002). A role for CAR in CYP2C9 regulation was suggested by evidence that both constitutive and drug inducible CYP2C9 mRNA expression was elevated in HepG2 cell lines stably transfected with mCAR or hCAR, in the absence or presence of TCPOBOP or PB, respectively (Ferguson *et al.*, 2002). By analyzing the CYP2C9 promoter region, two CAR-responsive elements, a DR5 located from –2900 to –2841 bp, and a DR4 located between –1822 and –1783 bp, were discovered independently (Ferguson *et al.*, 2002; Gerbal-Chaloin *et al.*, 2002). Gel shift assays demonstrated that both DRs bind CAR/RXR heterodimers. Furthermore, transfection experiments in HepG2 cells revealed that constitutively activated CAR increased reporter gene expression in vectors containing either of the DR4 or DR5 promoter elements (Ferguson *et al.*, 2002).

CYP2C8 is the most strongly inducible member of the CYP2C subfamily in human hepatocyte cultures. Examination of the CYP2C8 promoter region revealed that a distal PXR/CAR-binding site located at –8806 bp of the transcriptional start site, confers inducibility of CYP2C8 via the PXR agonist rifampicin and the CAR agonist CITCO (Ferguson *et al.*, 2005). Further analysis of the CYP2C8 basal promoter region revealed several putative binding motifs, including TATA-box, HNF3, CCAAT enhancer-binding protein, HNF4α, and GATA-binding protein, indicating many other potential nuclear factors may be involved in controlling the transcription of CYP2C8. Another important member of the human CYP2C subfamily, CYP2C19, metabolizes a number of clinically important drugs such as omeprazole, diazepam, propranolol, and *S*-mephenytoin. Chen *et al.* (2003) reported that transcriptional regulation of CYP19 gene expression is mediated by CAR, PXR, and GR through the consensus binding sites (CAR-RE; –1891/–1876 bp) and (GRE; –1750/–1736 bp), respectively. In addition, a functional CAR-RE has been localized and characterized in the promoter region of mouse cyp2c37 gene (Jackson *et al.*, 2006). Intriguingly, this CAR-RE only responds to CAR activation but not to PXR. Overall, these data suggest that CAR has a promiscuous role in the regulation of multiple CYP2C gene expression.

***3.2.2 CAR Regulation of UGT1A1 Gene*** For decades, PB has been used in the treatment of hyperbilirubinemia since this compound induces the expression of UGT1A1 enzyme which involves the glucuronidation of elevated bilirubin (Kopecky, Schwarz and Schwenzel, 1975; Schambach and Menzel, 1975; Ishii *et al.*, 1994). The molecular mechanism of PB induction of UGT1A1 remained elusive until a 290-bp distal enhancer module, located from –3483 to –3194 bp of the UGT1A1 promoter, was

identified as CAR response element by Sugatani *et al.* (2001). Consisting of three putative nuclear receptor palindromes, this enhancer module can be activated in transfection assays by human and mouse CAR in HepG2 cells, and mouse primary hepatocytes. Site-directed mutagenesis of this module abrogated the response. Furthermore, in a separate study, Huang *et al.* (2003) demonstrated that CAR activation increases hepatic expression of several relevant genes involved in bilirubin clearance, including UGT1A1, MRP2, SLC21A6, GSTA1, and GSTA2, in wild-type but not in CAR-null mice. A similar activation profile was observed also in a line of transgenic mice expressing hCAR but not mCAR in the liver. These results suggest that both human and mouse CAR can respond to elevated levels of bilirubin by regulating simultaneously the expression of UGT1A1, MRP2, SLC21A6, GSTA1, and GSTA2. It is of note that within this 290-bp distal enhancer module in the promoter of UGT1A1, several other xenobiotic response elements have also been recognized subsequently, including a PXRE, two GREs, and an AhR-specific XRE (Xie *et al.*, 2003; Yueh *et al.*, 2003; Kuno, Togawa and Mizutani, 2007).

***3.2.3 CAR Regulation of Drug Transporters*** Drug transporters play critical roles in drug absorption and clearance. Several efflux transporters, as well as members of the OATP uptake transporter family, have been identified as transcriptional target genes of the xenobiotic nuclear receptors PXR, CAR, PPAR, and FXR (Dussault *et al.*, 2001; Geick, Eichelbaum and Burk, 2001; Synold, Dussault and Forman, 2001; Guo *et al.*, 2002; Guo, Johnson and Klaassen, 2002; Kast *et al.*, 2002). Multidrug resistance-associated protein 2 (MRP2 or ABCC2) is involved in the transport of organic anions, bile salts, glutathione, and xenobiotics, such as the anticancer drugs cisplatin, anthacyclines, vinca alkaloids, and methotrexate (Konig *et al.*, 1999). Treatment with the PXR-selective activator PCN and the mCAR-specific ligand TCPOBOP resulted in MRP2 induction at the mRNA level. By analyzing the 5′-flanking region of the MRP2 gene, Kast *et al.* isolated an unusual 26-bp sequence, located at 440-bp upstream of the MRP2 transcription initiation site, that contained an everted repeat of AGTTCA spaced by eight nucleotides (ER-8) (Kast *et al.*, 2002). This ER-8 motif was shown to bind both CAR and PXR, and was activated by CAR and PXR ligands in cell-based reporter assays. These findings demonstrated the versatility in the binding abilities of PXR and CAR, and provided the first evidence that CAR and PXR can recognize response element (ER-8) in addition to the previously identified ER-6 element. Interestingly, PB induction of MRP2 was greater in PXR-null mice compared with wild-type animals. Given the fact that PB induction of CYP2B and CYP3A expression was unaffected in PXR-null mice but abolished in CAR-null mice, it is reasonable to speculate that mPXR may interact negatively with PB and that shared PXR/CAR-target genes may be induced by PB more efficaciously in mice lacking PXR due to reduced ligand-binding competition.

Conflicting data have been presented in the literature regarding the contribution of CAR in the regulation of MRP3, another ATP-dependent efflux transporter. Recently, Staudinger *et al.* showed that selective activation of mCAR by TCPOBOP mediates the inducible expression of Oatp2 and Mrp3 (Staudinger *et al.*, 2003). These results were similar to those obtained in PXR-null mice, where PB induction of both Oatp2 and Mrp3 was enhanced. Another report also demonstrated by real-time RT-PCR that both CAR and PXR play critical roles in regulating mouse Mrp3 gene expression (Maglich *et al.*, 2002). In contrast, hCAR appears not to be involved in the regulation of human MRP3 gene expression. Both hCAR- and mCAR-expressing

HepG2 cells have been used to elucidate the role of CAR in regulating induction of human CYP2B6 and UGT1A1 by PB-like inducers (Kawamoto *et al.*, 1999; Sueyoshi *et al.*, 1999; Sugatani *et al.*, 2001). Because of low endogenous CAR expression in HepG2 cells, neither CYP2B6 nor UGT1A1 is inducible by PB. However, higher endogenous CYP2B6 and UGT1A1 expression was observed in HepG2 cells stably transfected with mCAR due to the constitutive activation of CAR in these cells (Sueyoshi *et al.*, 1999). Different from the prototypical CAR target genes, such as CYP2B6 and UGT1A1, human *MRP3* gene expression was induced to a similar extent by PB, but not TCPOBOP in HepG2 and g2car-3 cells, suggesting that PB induction of human MRP3 is mediated by transcription factors other than CAR (Kiuchi *et al.*, 1998; Xiong *et al.*, 2002). Furthermore, sequence analysis of the 5′-flanking region of human MRP3 did not reveal any consensus binding sites for either CAR or PXR, although binding sites were identified for other transcription factors such as Sp1, AP1, AP2, AP3, N-myc, CCAAT/enhancer-binding protein, hepatic nuclear factor-5, and nuclear factor-kB (Fromm *et al.*, 1999; Takada, Suzuki and Sugiyama, 2000). Further study is required to determine if these factors are important in the regulation of *MRP3* gene expression by PB and other compounds. In addition, discrepancies between the results from *in vivo* and *in vitro* studies of the regulation of other transporters should be taken into consideration. For instance, induction of Mdr1 by PCN was observed *in vitro* but not *in vivo* (Salphati and Benet, 1998; Jones *et al.*, 2000). Mrp2 was induced by PB in primary cultured rat hepatocytes but not in rats *in vivo* (Ogawa *et al.*, 2000). These reported *in vitro* and *in vivo* differences suggest that some regulatory mechanisms implicated by *in vitro* methods may not be physiologically relevant under *in vivo* conditions.

### 3.3 Mechanism of CAR Activation

Consistent with its designated name, CAR is constitutively activated in all the immortalized cell lines without xenobiotic stimulation, making the investigation of CAR activation much more challenging. In contrast, CAR activation in primary cultured hepatocytes and intact liver *in vivo* is inducer-dependent, and CAR is localized in the cytoplasm of liver cells and translocated into the nucleus only after exposure to xenobiotic inducers (Kawamoto *et al.*, 1999; Sueyoshi *et al.*, 1999). One of the essential features that are critical for regulating xenobiotic-induced CAR activation involves the nuclear accumulation of this receptor. A number of studies showed that PB and TCPOBOP treatments were associated with decreased CAR expression in cytoplasmic fraction compared with nuclear extracts; this finding was substantiated by the increased binding of a CAR/RXR heterodimer to NR1 in liver nuclear extracts (Kawamoto *et al.*, 1999). These results indicate that nuclear translocation of CAR might be the first activation step in response to PB-type inducers. As a result, the mechanism by which CAR is retained in the cytoplasm has now become a major focus of research interest. Inhibition of nuclear translocation by antagonists such as androstanes might be a possible mechanism for explaining cytoplasmic retention of CAR in primary cells and intact liver. However, physiological serum concentrations of androstane metabolites are several magnitudes lower than concentrations required for repression of CAR activation of the PBREM and induction of *CYP2B* genes in HepG2 cells (Dufort *et al.*, 2001). Moreover, treatment of mouse primary hepatocytes with high concentrations of androstenol could not inhibit PB-driven nuclear accumulation of CAR (Kawamoto *et al.*, 1999). Nuclear translocation is common among steroid receptors such as GR,

progesterone receptor (PR), vitamin D receptor (VDR), and PPAR (Picard and Yamamoto, 1987; Guiochon-Mantel *et al.*, 1991; Walker, Htun and Hager, 1999), which are dissociated from their cytoplasmic complexes with chaperone proteins such as heat shock proteins, immunophilins, or p23 protein. Upon direct binding to their agonists, these receptors are released from their complexes and translocate into the nucleus (Defranco *et al.*, 1995; Pratt, Silverstein and Galigniana, 1999). However, it seems that direct ligand binding is not necessary for a drug to stimulate CAR nuclear translocation. Although PB induces nuclear translocation of both mouse and human CAR, it does not bind to either in various *in vitro* assays (Moore *et al.*, 2000a; Tzameli *et al.*, 2000). Similarly, TCPOBOP treatment resulted in mouse and human CAR accumulation in the nucleus even though hCAR could neither bind nor be activated by TCPOBOP (Tzameli *et al.*, 2000). Advances in our understanding of the CAR translocation mechanism include the discovery that okadaic acid (OA), an inhibitor of protein phosphatase 2A, was able to inhibit PB- and TCPOBOP-driven CAR translocation and CYP2b10 induction in mouse primary hepatocytes (Sidhu and Omiecinski, 1997; Honkakoski and Negishi, 1998; Kawamoto *et al.*, 1999). Nonetheless, subsequent studies have demonstrated that translocation of CAR is necessary but not sufficient to activate this receptor. For instance, calcium/calmodulin-dependent kinase (CaMK) inhibitors, KN-93 and KN-62 could efficiently inhibit PB-mediated CYP2B induction without affecting CAR nuclear accumulation (Negishi, 2000). Activation of nuclear receptors is generally associated with co-regulators such as SRCs. Recently, we observed that the known hCAR deactivator clotrimazole could efficiently translocate EYFP-hCAR into nucleus of mouse hepatocytes *in vivo*, but dissociates SRC-1 from the nuclear localized CAR/RXR heterodimer (unpublished data). Taken together, these observations suggest that tight control of an otherwise constitutively active receptor is achieved by cytosolic sequestration *in vivo* and primary hepatocytes through complicated yet unknown mechanisms.

### 3.4 Identification of CAR Activator

One of the major difficulties for investigating drug-induced CAR activation *in vitro* is the high constitutive activation of CAR in immortalized cell lines, due to the spontaneous accumulation of the receptor in the nucleus in the absence of inducer (Wang and Negishi, 2003). Because of difficulties in evaluation of hCAR activation, the contribution of this receptor to DDIs has remained ambiguous. Mouse CAR-specific target genes and their distinct physiological roles were characterized by utilizing TCPOBOP as a specific mCAR agonist. However, previously identified hCAR ligands such as clotrimazole and 5b-pregnane-3, 20-dione were shown also to activate hPXR. Later, CITCO was identified as a selective hCAR agonist (Maglich *et al.*, 2003). This compound was associated with an $EC_{50}$ of 25 nM for hCAR activation of a CYP3A4-XREM reporter construct in transfected CV-1 cells (Maglich *et al.*, 2003). Furthermore, green fluorescent protein (GFP)-CAR fusion protein was accumulated in the nucleus of rat primary hepatocytes after CITCO treatment. CITCO was also used to delineate hCAR target genes in primary cultures of human hepatocyte. The induction of eight DMEs was investigated in three separate sets of human hepatocytes treated with CITCO or hPXR ligand RIF. As expected, CYP2B6 mRNA displayed the most potent induction by 100 nM of CITCO. In contrast, CYP3A4 mRNA was more robustly induced by rifampicin than by CITCO. Using TCPOBOP in mice and CITCO in humans, specific CAR target genes were also compared between rodents and humans. Results demon-

strated that, in general, human and mouse CAR share a number of similar target genes, including mouse Cyp2b10 and human CYP2B6, the transporter genes mouse Mdr1 and human MDR1, and the phase II sulfotransferase genes mouse SULTN and human SULT1A1. However, to date, there is only limited publication addressing CITCO as a specific hCAR activator; more critical investigation is required to evaluate the specificity of hCAR activation, particularly in *in vitro* cell-based transfection assays by using different cell lines.

Recently, several groups have identified alternative splicing variants of wild-type hCAR with altered functional activities (Auerbach *et al.*, 2003; Arnold, Eichelbaum and Burk, 2004; Jinno *et al.*, 2004; Ikeda *et al.*, 2005). One of these variants, hCAR3, exhibited significantly lower basal activity than wild-type hCAR and was activated extensively by the known hCAR activator CITCO in a cell-based reporter gene assay (Auerbach *et al.*, 2005). In a recent study, we have evaluated the activation profile of hCAR3 more extensively by testing known hCAR activators as well as 16 other CYP3A4 and/or CYP2B inducers (Faucette *et al.*, 2007). Notably, all the known hCAR activators, including the direct activator CITCO, artemisinin and the indirect activators PB and PHN, enhanced CYP2B6 reporter gene expression via hCAR3 in HepG2 cells, suggesting hCAR3 could be activated through both ligand-dependent and -independent mechanisms. Overall, in the absence of other suitable *in vitro* models and based on similar chemical sensitivities to wild-type hCAR, hCAR3 cell-based reporter gene assays may provide a useful tool for *in vitro* screening of hCAR activation in a relatively high-throughput format.

### 3.5 Species Differences in Activation of CAR

Although significant progress on the molecular mechanisms underlying CAR regulation of DME gene expression has been achieved in rodents, information pertaining to hCAR regulation of human DMEs is relatively limited. Two apparent issues that remain to be resolved are whether hCAR is regulated in the same fashion as the rodent CAR and whether CAR is the predominate regulator of CYP2B6 in human liver. At this point, the conclusions regarding the role of CAR in the regulation of CYP2B or other genes are based largely on rodent CAR, primarily mCAR (Hani *et al.*, 1996; Choi *et al.*, 1997). Although hCAR exhibits some common characteristics with its rodent counterparts, such as undergoing nuclear translocation after PB treatment and binding to the PBREM, there are distinct differences between rodent and hCAR. For example, TCPOBOP, the most potent mCAR ligand identified to date, cannot bind or activate either rat or hCAR. All known mCAR inhibitors, such as androstenol, progesterone, and androgens, and CaMK inhibitors do not inhibit hCAR activation. Overall, current evidence suggests that there are clear species-specific differences in CYP2B induction and CAR activation, which severely hamper the extrapolation of animal data to humans. For these reasons, regulation of hCAR by drugs and other xenobiotics has become a more complex and urgent issue that has yet to be solved.

Generation of transgenic PXR-null and CAR-null mice, especially the humanized PXR and CAR mice model, has made it possible to specifically address the effects of these receptors on DME regulation *in vivo*. In CAR-null mice, the strong induction of CYP2b10 gene expression by PB and TCPOBOP was totally absent (Wei *et al.*, 2000). Similarly, in obese Zucker rats, which express extremely low levels of CAR but have normal levels of PXR expression, PB only moderately induces both CYP2B and

CYP3A, whereas CYP3A was significantly induced by the PXR-specific activator PCN (Zelko and Negishi, 2000). Given that xenobiotic induction of DMEs shows striking species specificity due to the species-dependent biochemical properties of nuclear receptors, a CAR humanized mouse line has also been generated (Zhang *et al.*, 2002). This line specifically expresses hCAR in the mCAR-null mice liver driven by the potent albumin promoter. Using this model, opposite responses of mCAR and hCAR to a widely used antiemetic, meclizine, have been established (Huang *et al.*, 2004). Although meclizine treatment augmented the expression of CAR target genes in wild-type mice, both hCAR transactivation and the PB-induced expression of these CAR target genes were inhibited by meclizine treatment in primary hepatocytes isolated from the liver of humanized CAR mice. Furthermore, meclizine decreased the susceptibility of APAP-induced hepatotoxicity in humanized CAR mice (Huang *et al.*, 2004).

### 3.6 Cross talk between CAR, PXR, and Other Nuclear Receptors

Although PXR and CAR have been implicated as the primary regulator of CYP3A and CYP2B, respectively, accumulated evidence has suggested cross talk between these two NRs through recognizing and binding to each other's target gene response elements. Significant overlap exists in the inducers of CYP2B and CYP3A genes in rodent and human primary hepatocytes, such as clotrimazole, RU486, RIF, phenytoin, and PB. Moreover, both receptors can bind to several of the same compounds, including androstanol, clotrimazole, estradiol, and TCPOBOP with different affinities. The current model of cross talk between PXR and CAR is derived primarily from rodent data and commonly depicts symmetrical cross-regulation of CYP2B and CYP3A induction by PXR and CAR, respectively. However, because of documented species differences in nuclear receptor signaling pathways, it is speculative to extrapolate the extents of cross-regulation of *CYP2B* and *CYP3A* genes by rodent PXR and CAR to the corresponding human receptors. The selectivities of hPXR and hCAR for their cross-regulatory genes remain unclear because of the lack of comparisons of CYP2B6 and CYP3A4 promoter binding and functional activation between hPXR and hCAR, as well as the previous unavailability of selective hCAR activators. Recently, we have evaluated the symmetry of hPXR and hCAR cross talk by comparing the selectivities of these receptors for CYP2B6 and CYP3A4 (Faucette *et al.*, 2006). Human hepatocyte studies revealed nonselective induction of both CYP2B6 and CYP3A4 by hPXR activation but marked preferential induction of CYP2B6 by selective hCAR activation. Gel shift assays demonstrated that hPXR exhibited strong and relatively equal binding to all functional response elements in both *CYP2B6* and *CYP3A4* genes, whereas hCAR displayed significantly weak binding to the CYP3A4 proximal ER-6 motif. In cell-based transfection assays, hCAR displayed greater activation of CYP2B6 reporter gene expression compared with CYP3A4 with constructs containing both proximal and distal regulatory elements. Furthermore, in agreement with binding observations, transfection assays using promoter constructs containing repeats of CYP2B6 DR4 and CYP3A4 ER-6 motifs revealed an even greater difference in reporter activation by hCAR. In contrast, hPXR activation resulted in less discernible differences between CYP2B6 and CYP3A4 reporter gene expression. These results lead to a novel model of asymmetrical cross-regulation of these genes that differs from the presumed rodent model (Fig. 3). The greater selectivity of hCAR activation resulted in preferred induction of CYP2B6 over CYP3A4, whereas the lower selectivity of hPXR activation conferred efficacious induction of both target genes.

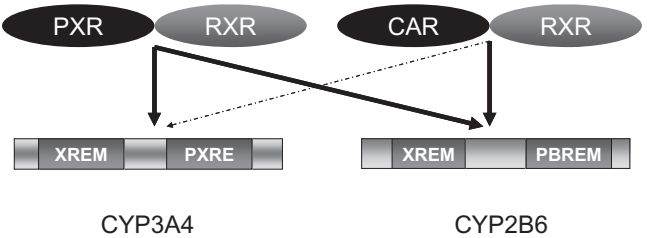


**Figure 3** Asymmetrical model of cross talk between PXR and CAR in the regulation of *CYP2B6* and *CYP3A4* genes. Different from previously presumed rodent model, human PXR and CAR demonstrate an asymmetrical cross-regulation of CYP2B6 and CYP3A4 genes. This model is based on results generated from Faucette *et al.* (2006).

In addition to regulating phase I enzymes, CAR and PXR also cooperate to regulate the expression of phase II enzymes and drug transporters. PXR and CAR have been shown to induce the expression of UGT1A1, OATP2, GSTA1 and 2, all of which are important for bilirubin detoxification. The sharing of bilirubin detoxifying target genes may have accounted for a similar resistance to experimental hyperbilirubinemia in mice with PXR or CAR genetically or pharmacologically activated (Sugatani *et al.*, 2001; Huang *et al.*, 2003; Xie *et al.*, 2003). In another example of receptor cross talk, MRP2, a transporter mediating the efflux of several conjugated compounds across the apical membrane of the hepatocytes into the bile canaliculi, is regulated by FXR, PXR, and CAR (Kast *et al.*, 2002). An ER-8 (everted repeat separated by eight nucleotides) type of NR response element was identified in the promoter of the rat *Mrp2* gene. This element is capable of binding to and conferring PXR, CAR, and FXR responsiveness on a heterologous thymidine kinase promoter (Kast *et al.*, 2002). This is reminiscent of the co-regulation of the rodent Sult2a9/2a1 by FXR, PXR, CAR, and LXR, in which an IR-0 was the shared response element (Song *et al.*, 2001; Sonoda *et al.*, 2002; Saini *et al.*, 2004; Uppal *et al.*, 2007).

Another interesting example of functional interplay between xenobiotic receptors was revealed in the characterization of PXR-null mice in bilirubin clearance (Uppal *et al.*, 2005). Having known that activation of PXR is sufficient to detoxify bilirubin, it was surprising to find that the PXR-null mice also showed increased bilirubin clearance. We proposed that, when both PXR and CAR are present, the ligand-free PXR suppresses the constitutive activity of CAR, maintaining a basal capacity of bilirubin detoxification. The increased bilirubin clearance in PXR-null mice was likely the result of CAR derepression as a consequence of loss of PXR. This notion was supported by the observation that the pattern of enzyme and transporter regulation in the PXR-null mice was remarkably similar to that of transgenic mice expressing activated CAR (Uppal *et al.*, 2005).

In a more recent example, we showed that PXR, LXR, and PPARγ cooperate to regulate the free fatty acid transporter CD36 and to promote fatty acid-mediated hepatic lipogenesis. In this case, three distinct NR response elements were found for each of the three receptors and they are clustered in 500-bp sequences in the mouse *Cd36* gene promoter (J. Zhou and W. Xie, unpublished results). Consistent with the co-regulation of *Cd36*, activation of both PXR and LXR had an additive effect to induce hepatic triglyceride content (J. Zhou and W. Xie, unpublished results).

## 4 ARYL HYDROCARBON RECEPTOR (AhR)

It is noteworthy that transcriptional factors other than NRs also play critical roles in the control of xenobiotic-induced DME expression. One such non-NR transcriptional factor is the AhR, a member of the basic helix-loop-helix PER-ARNT-SIM (PAS) transcription factor family (Burbach, Poland and Bradfield, 1992; Safe, 2001). Similar to several NR family members, AhR expressed predominantly in the cytoplasm as a complex with Hsp90 and several other chaperones without the stimulation of xenobiotic ligands. Once it bonds to its ligands such as TCDD and 3-methylcholanthrene (3-MC), the receptor translocates into the nucleus and forms a heterodimer with its partner, AhR nuclear translocator (ARNT) (Safe, 2001). The newly formed AhR/ARNT heterodimer will bind to the XRE that is located in the promoter region of its target genes.

Like PXR and CAR, activation of AhR has been reported to simultaneously induce both phase I and II enzymes (Rushmore and Pickett, 1990; Ma, 2001) The mammalian CYP1A1, CYP1A2, and CYP1B1 genes are AR target genes identified initially (Hoffman *et al.*, 1991; Gu, Hogenesch and Bradfield, 2000). More recently, it was found that AhR is also implicated in the regulation of UGTs, including UGT1A1 and 1A6 (Emi, Ikushiro and Iyanagi, 1996; Munzel *et al.*, 1998; Klinge *et al.*, 1999; Shelby *et al.*, 2003; Yueh *et al.*, 2003). AhR-binding XREs have been identified in the human and rodent UGT1A1 gene promoter (Yueh *et al.*, 2003). An XRE was reported for the rat *Ugt1a6* gene and found to be both necessary and sufficient to mediate *Ugt1a6* transactivation by AhR (Shelby *et al.*, 2003). Interestingly, the PXR-, CAR-, and AhR-responsive elements are closely localized in the UGT1A1 gene promoter (Zhou, Zhang and Xie, 2005). It remains to be determined whether there are coordination or cross talk between AhR and CAR/PXR in xenobiotic regulation. The existing literature does support cross talk between AhR and nuclear receptor signaling. For example, AhR has been shown to interact with a number of NRs, such as the estrogen receptor α (ERα), COUP-TFI, and estrogen-related receptor α (ERRα). Moreover, the AhR/Arnt heterodimers can utilize certain NR response elements such as the naturally occurring estrogen response elements (Klinge *et al.*, 1999).

## 5 CLOSING REMARKS AND PERSPECTIVES

In the past decade, significant progress has been made in our understanding of the comprehensive role of NRs in the regulation of DMEs and drug transporter system. It has become clear that PXR and CAR can function as master xenosensors to regulate the expression of majority of the DMEs and transporters. Another important concept that has emerged from recent studies is that the interactions between NRs and their target sequences are multifaceted and much more complex than initially estimated. Many of the NRs share marked sequence homology in their DBD, suggesting that there are significant overlaps in their ability to bind to common response elements in a diverse set of target genes. This cross talk between NRs is believed to be the molecular basis for the fail-safe xenobiotic regulatory networks that facilitate host protection.

It also appears that the NR-controlled xenobiotic regulation is a double-edged sword. One of the remaining outstanding challenges is whether the biological action of PXR and CAR makes them drug targets for the treatment of human diseases, such as hyperbilirubinemia, obesity, as well as bile acid-associated cholestasis and colon

cancers. Because of the importance of the DMEs in the biotransformation of marketed and future drugs, DDIs involving the induction and inhibition of CYPs are of great clinical importance. However, the potential DDI alone should not exclude NRs as therapeutic target, and this notion is supported by the facts that many of the xenobiotic receptor ligands, including RIF and PB, have been successfully used as clinical drugs.

Our understanding of the NR biology has further expanded our ability to predict the potential pharmacological and toxicological properties in the early stage of drug development. A series of NR-dependent assays has been developed and well evaluated, including *in vitro* cell-based reporter experiments in a high-throughput manner, *in vitro* ligand-binding assays, and transgenic animal models, especially the "humanized" PXR and CAR mice. It is anticipated that these novel methods may lead to a more rational and molecular-based approach to developing drugs with enhanced therapeutic efficacy and improved safety profiles.

## ACKNOWLEDGMENTS

The original results from our labs that are described in this chapter were generated with the support of NIH grants ES012479, ES014626, CA107011 (to W.X.) and DK061652 (to H.W.).

## REFERENCES

Arnold, K.A., Eichelbaum, M., Burk, O. (2004). Alternative splicing affects the function and tissue-specific expression of the human constitutive androstane receptor. *Nucl. Recept.*, *2* (1), 1.

Auerbach, S.S. *et al.* (2003). Alternatively spliced isoforms of the human constitutive androstane receptor. *Nucleic Acids Res.*, *31* (12), 3194–3207.

Auerbach, S.S. *et al.* (2005). Retinoid X receptor-alpha-dependent transactivation by a naturally occurring structural variant of human constitutive androstane receptor (NR1I3). *Mol. Pharmacol.*, *68* (5), 1239–1253.

Baes, M. *et al.* (1994). A new orphan member of the nuclear hormone receptor superfamily that interacts with a subset of retinoic acid response elements. *Mol. Cell Biol.*, *14* (3), 1544–1551.

Barbier, O. *et al.* (2003). The UDP-glucuronosyltransferase 1A9 enzyme is a peroxisome proliferator-activated receptor alpha and gamma target gene. *J. Biol. Chem.*, *278* (16), 13975–13983.

Bertilsson, G. *et al.* (1998). Identification of a human nuclear receptor defines a new signaling pathway for CYP3A induction. *Proc. Natl. Acad. Sci. U.S.A.*, *95* (21), 12208–12213.

Blumberg, B. *et al.* (1998). SXR, a novel steroid and xenobiotic-sensing nuclear receptor. *Genes Dev.*, *12* (20), 3195–3205.

Burbach, K.M., Poland, A., Bradfield, C.A. (1992). Cloning of the Ah-receptor cDNA reveals a distinctive ligand-activated transcription factor. *Proc. Natl. Acad. Sci. U.S.A.*, *89* (17), 8185–8189.

Cairns, W. *et al.* (1991). Assembly of a glucocorticoid receptor complex prior to DNA binding enhances its specific interaction with a glucocorticoid response element. *J. Biol. Chem.*, *266* (17), 11221–11226.

Chen, C., Staudinger, J.L., Klaassen, C.D. (2003). Nuclear receptor, pregnane X receptor, is required for induction of UDP-glucuronosyltranferases in mouse liver by pregnenolone-16 alpha-carbonitrile. *Drug Metab. Dispos.*, *31* (7), 908–915.

Chen, Y. *et al.* (2003). Identification of constitutive androstane receptor and glucocorticoid receptor binding sites in the CYP2C19 promoter. *Mol. Pharmacol.*, *64* (2), 316–324.

Chen, Y. *et al.* (2004). Induction of human CYP2C9 by rifampicin, hyperforin, and phenobarbital is mediated by the pregnane X receptor. *J. Pharmacol. Exp. Ther.*, *308* (2), 495–501.

Choi, H.S. *et al.* (1997). Differential transactivation by two isoforms of the orphan nuclear hormone receptor CAR. *J. Biol. Chem.*, *272* (38), 23565–23571.

Defranco, D.B. *et al.* (1995). Nucleocytoplasmic shuttling of steroid receptors. *Vitam. Horm.*, *51*, 315–338.

Dufort, I. *et al.* (2001). Comparative biosynthetic pathway of androstenol and androgens. *J. Steroid Biochem. Mol. Biol.*, *77* (4–5), 223–227.

Dussault, I. *et al.* (2001). Peptide mimetic HIV protease inhibitors are ligands for the orphan receptor SXR. *J. Biol. Chem.*, *276* (36), 33309–33312.

Dussault, I. *et al.* (2003). Identification of an endogenous ligand that activates pregnane X receptor-mediated sterol clearance. *Proc. Natl. Acad. Sci. U.S.A.*, *100* (3), 833–838.

Dvorak, Z. *et al.* (2003). Colchicine down-regulates cytochrome P450 2B6, 2C8, 2C9, and 3A4 in human hepatocytes by affecting their glucocorticoid receptor-mediated regulation. *Mol. Pharmacol.*, *64* (1), 160–169.

Emi, Y., Ikushiro, S., Iyanagi, T. (1996). Xenobiotic responsive element-mediated transcriptional activation in the UDP-glucuronosyltransferase family 1 gene complex. *J. Biol. Chem.*, *271* (7), 3952–3958.

Falkner, K.C. *et al.* (2001). Regulation of the rat glutathione S-transferase A2 gene by glucocorticoids: involvement of both the glucocorticoid and pregnane X receptors. *Mol. Pharmacol.*, *60* (3), 611–619.

Faucette, S.R. *et al.* (2006). Differential regulation of hepatic CYP2B6 and CYP3A4 genes by constitutive androstane receptor but not pregnane X receptor. *J. Pharmacol. Exp. Ther.*, *317* (3), 1200–1209.

Faucette, S.R. *et al.* (2007). Relative activation of human pregnane X receptor versus constitutive androstane receptor defines distinct classes of CYP2B6 and CYP3A4 inducers. *J. Pharmacol. Exp. Ther.*, *320* (1), 72–80.

Ferguson, S.S. *et al.* (2002). Regulation of human CYP2C9 by the constitutive androstane receptor: discovery of a new distal binding site. *Mol. Pharmacol.*, *62* (3), 737–746.

Ferguson, S.S. *et al.* (2005). Human CYP2C8 is transcriptionally regulated by the nuclear receptors constitutive androstane receptor, pregnane X receptor, glucocorticoid receptor, and hepatic nuclear factor 4alpha. *Mol. Pharmacol.*, *68* (3), 747–757.

Freedman, L.P. (1999). Increasing the complexity of coactivation in nuclear receptor signaling. *Cell*, *97* (1), 5–8.

Fromm, M.F. *et al.* (1999). Human MRP3 transporter: identification of the 5′-flanking region, genomic organization and alternative splice variants. *Biochim. Biophys. Acta*, *1415* (2), 369–374.

Geick, A., Eichelbaum, M., Burk, O. (2001). Nuclear receptor response elements mediate induction of intestinal MDR1 by rifampin. *J. Biol. Chem.*, *276* (18), 14581–14587.

Gerbal-Chaloin, S. *et al.* (2001). Induction of CYP2C genes in human hepatocytes in primary culture. *Drug Metab. Dispos.*, *29* (3), 242–251.

Gerbal-Chaloin, S. *et al.* (2002). Transcriptional regulation of CYP2C9 gene. Role of glucocorticoid receptor and constitutive androstane receptor. *J. Biol. Chem.*, *277* (1), 209–217.

Giguere, V. (1999). Orphan nuclear receptors: from gene to function. *Endocr. Rev.*, *20* (5), 689–725.

Goldstein, J.A. (2001). Clinical relevance of genetic polymorphisms in the human CYP2C subfamily. *Br. J. Clin. Pharmacol.*, *52* (4), 349–355.

Gong, H. *et al.* (2005). Animal models of xenobiotic receptors in drug metabolism and diseases. *Methods Enzymol.*, *400*, 598–618.

Gong, H. *et al.* (2006). Orphan nuclear receptor pregnane X receptor sensitizes oxidative stress responses in transgenic mice and cancerous cells. *Mol. Endocrinol.*, *20* (2), 279–290.

Goodwin, B. *et al.* (2001). Regulation of the human CYP2B6 gene by the nuclear pregnane X receptor. *Mol. Pharmacol.*, *60* (3), 427–431.

Gu, Y.Z., Hogenesch, J.B., Bradfield, C.A. (2000). The PAS superfamily: sensors of environmental and developmental signals. *Annu. Rev. Pharmacol. Toxicol.*, *40*, 519–561.

Guiochon-Mantel, A. *et al.* (1991). Nucleocytoplasmic shuttling of the progesterone receptor. *EMBO J.*, *10* (12), 3851–3859.

Guo, G.L. *et al.* (2002). Induction of rat organic anion transporting polypeptide 2 by pregnenolone-16alpha-carbonitrile is via interaction with pregnane X receptor. *Mol. Pharmacol.*, *61* (4), 832–839.

Guo, G.L., Johnson, D.R., Klaassen, C.D. (2002). Postnatal expression and induction by pregnenolone-16alpha-carbonitrile of the organic anion-transporting polypeptide 2 in rat liver. *Drug Metab. Dispos.*, *30* (3), 283–288.

Hagenbuch, B., Meier, P.J. (2003). The superfamily of organic anion transporting polypeptides. *Biochim. Biophys. Acta*, *1609* (1), 1–18.

Hani, E-H. *et al.* (1996). Indication for genetic linkage of the phosphoenolpyruvate carboxykinase (PCK1) gene region on chromosome 20q to non-insulin-dependent diabetes mellitus. *Diabetes Metab.*, *22* (6), 451–454.

Hayes, J.D., Pulford, D.J. (1995). The glutathione S-transferase supergene family: regulation of GST and the contribution of the isoenzymes to cancer chemoprotection and drug resistance. *Crit. Rev. Biochem. Mol. Biol.*, *30* (6), 445–600.

Hoffman, E.C. *et al.* (1991). Cloning of a factor required for activity of the Ah (dioxin) receptor. *Science*, *252* (5008), 954–958.

Honkakoski, P. *et al.* (1998). The nuclear orphan receptor CAR-retinoid X receptor heterodimer activates the phenobarbital-responsive enhancer module of the CYP2B gene. *Mol. Cell Biol.*, *18* (10), 5652–5658.

Honkakoski, P., Negishi, M. (1997). Characterization of a phenobarbital-responsive enhancer module in mouse P450 Cyp2b10 gene. *J. Biol. Chem.*, *272* (23), 14943–14949.

Honkakoski, P., Negishi, M. (1998). Protein serine/threonine phosphatase inhibitors suppress phenobarbital-induced Cyp2b10 gene transcription in mouse primary hepatocytes. *Biochem. J.*, *330* (Pt 2), 889–895.

Honkakoski, P., Negishi, M. (2000). Regulation of cytochrome P450 (CYP) genes by nuclear receptors. *Biochem. J.*, *347* (Pt 2), 321–337.

Huang, W. *et al.* (2003). Induction of bilirubin clearance by the constitutive androstane receptor (CAR). *Proc. Natl. Acad. Sci. U.S.A.*, *100* (7), 4156–4161.

Huang, W. *et al.* (2004). Meclizine is an agonist ligand for mouse constitutive androstane receptor (CAR) and an inverse agonist for human CAR. *Mol. Endocrinol.*, *18* (10), 2402–2408.

Ikeda, S. *et al.* (2005). Functional analysis of four naturally occurring variants of human constitutive androstane receptor. *Mol. Genet. Metab.*, *86* (1–2), 314–319.

Ishii, Y. *et al.* (1994). Purification of a phenobarbital-inducible morphine UDP-glucuronyltransferase isoform, absent from Gunn rat liver. *Arch. Biochem. Biophys.*, *315* (2), 345–351.

Jackson, J.P. *et al.* (2006). Phenytoin induction of the cyp2c37 gene is mediated by the constitutive androstane receptor. *Drug Metab. Dispos.*, *34* (12), 2003–2010.

Jinno, H. *et al.* (2004). Identification of novel alternative splice variants of human constitutive androstane receptor and characterization of their expression in the liver. *Mol. Pharmacol.*, *65* (3), 496–502.

Jones, S.A. *et al.* (2000). The pregnane X receptor: a promiscuous xenobiotic receptor that has diverged during evolution. *Mol. Endocrinol.*, *14* (1), 27–39.

Kast, H.R. *et al.* (2002). Regulation of multidrug resistance-associated protein 2 (ABCC2) by the nuclear receptors pregnane X receptor, farnesoid X-activated receptor, and constitutive androstane receptor. *J. Biol. Chem.*, *277* (4), 2908–2915.

Kastner, P.M.M., Chambon, P. (1995). Nonsteroid nuclear receptors: what are genetic studies telling us about their role in real life? *Cell*, *83* (6), 859–869.

Kawamoto, T. *et al.* (1999). Phenobarbital-responsive nuclear translocation of the receptor CAR in induction of the CYP2B gene. *Mol. Cell Biol.*, *19* (9), 6318–6322.

Kiuchi, Y. *et al.* (1998). cDNA cloning and inducible expression of human multidrug resistance associated protein 3 (MRP3). *FEBS Lett.*, *433* (1–2), 149–152.

Klaassen, C.D., Slitt, A.L. (2005). Regulation of hepatic transporters by xenobiotic receptors. *Curr. Drug Metab.*, *6* (4), 309–328.

Kliewer, S.A. *et al.* (1998). An orphan nuclear receptor activated by pregnanes defines a novel steroid signaling pathway. *Cell*, *92* (1), 73–82.

Kliewer, S.A., Goodwin, B., Willson, T.M. (2002). The nuclear pregnane X receptor: a key regulator of xenobiotic metabolism. *Endocr. Rev.*, *23* (5), 687–702.

Klinge, C.M. *et al.* (1999). The aryl hydrocarbon receptor (AHR)/AHR nuclear translocator (ARNT) heterodimer interacts with naturally occurring estrogen response elements. *Mol. Cell Endocrinol.*, *157* (1–2), 105–119.

Kocarek, T.A. *et al.* (1995). Comparative analysis of cytochrome P4503A induction in primary cultures of rat, rabbit, and human hepatocytes. *Drug Metab. Dispos.*, *23* (3), 415–421.

Kolars, J.C. *et al.* (1992). Identification of rifampin-inducible P450IIIA4 (CYP3A4) in human small bowel enterocytes. *J. Clin. Invest.*, *90* (5), 1871–1878.

Konig, J. *et al.* (1999). Conjugate export pumps of the multidrug resistance protein (MRP) family: localization, substrate specificity, and MRP2-mediated drug resistance. *Biochim. Biophys. Acta*, *1461* (2), 377–394.

Kopecky, P., Schwarz, I., Schwenzel, W. (1975). Antepartal phenobarbital therapy for the improvement of fetal bilirubin conjugation. *Arch. Gynakol.*, *219* (1–4), 455–457.

Kruh, G.D., Belinsky, M.G. (2003). The MRP family of drug efflux pumps. *Oncogene*, *22* (47), 7537–7552.

Kuno, T., Togawa, H., Mizutani, T. (2007). Induction of human UGT1A1 by a complex of dexamethasone-GR dependent on proximal site and independent of PBREM. *Mol. Biol. Rep.*, *35* (3), 361–367.

Lanz, R.B. *et al.* (1999). A steroid receptor coactivator, SRA, functions as an RNA and is present in an SRC-1 complex. *Cell*, *97* (1), 17–27.

Laudet, V. *et al.* (1992). Evolution of the nuclear receptor gene superfamily. *EMBO J.*, *11* (3), 1003–1013.

Lee, M.S. *et al.* (1993). Structure of the retinoid X receptor alpha DNA binding domain: a helix required for homodimeric DNA binding, *Science*, *260* (5111), 1117–1121.

Lehmann, J.M. *et al.* (1998). The human orphan nuclear receptor PXR is activated by compounds that regulate CYP3A4 gene expression and cause drug interactions. *J. Clin. Invest.*, *102* (5), 1016–1023.

Ma, Q. (2001). Induction of CYP1A1. The AhR/DRE paradigm: transcription, receptor regulation, and expanding biological roles. *Curr. Drug Metab.*, *2* (2), 149–164.

Mackenzie, P.I. *et al.* (1997). The UDP glycosyltransferase gene superfamily: recommended nomenclature update based on evolutionary divergence. *Pharmacogenetics*, *7* (4), 255–269.

Mackenzie, P.I. *et al.* (2003). Regulation of UDP glucuronosyltransferase genes. *Curr. Drug Metab.*, *4* (3), 249–257.

Maglich, J.M. *et al.* (2001). Comparison of complete nuclear receptor sets from the human, Caenorhabditis elegans and Drosophila genomes. *Genome. Biol.*, *2* (8), RESEARCH0029.1–7.

Maglich, J.M. *et al.* (2002). Nuclear pregnane x receptor and constitutive androstane receptor regulate overlapping but distinct sets of genes involved in xenobiotic detoxification. *Mol. Pharmacol.*, *62* (3), 638–646.

Maglich, J.M. *et al.* (2003). Identification of a novel human constitutive androstane receptor (CAR) agonist and its use in the identification of CAR target genes. *J. Biol. Chem.*, *278* (19), 17277–17283.

Mangelsdorf, D.J. *et al.* (1995). The nuclear receptor superfamily: the second decade. *Cell*, *83* (6), 835–839.

Maurel, D. *et al.* (1996). Effects of acute tilt from orthostatic to head-down antiorthostatic restraint and of sustained restraint on the intra-cerebroventricular pressure in rats. *Brain Res.*, *736* (1–2), 165–173.

Mills, J.B. *et al.* (2004). Induction of drug metabolism enzymes and MDR1 using a novel human hepatocyte cell line. *J. Pharmacol. Exp. Ther.*, *309* (1), 303–309.

Moore, J.T., Kliewer, S.A. (2000). Use of the nuclear receptor PXR to predict drug interactions. *Toxicology*, *153* (1–3), 1–10.

Moore, L.B. *et al.* (2000a). Orphan nuclear receptors constitutive androstane receptor and pregnane X receptor share xenobiotic and steroid ligands. *J. Biol. Chem.*, *275* (20), 15122–15127.

Moore, L.B. *et al.* (2000b). St. John's wort induces hepatic drug metabolism through activation of the pregnane X receptor. *Proc. Natl. Acad. Sci. U.S.A.*, *97* (13), 7500–7502.

Morris, M.E., Pang, K.S. (1987). Competition between two enzymes for substrate removal in liver: modulating effects due to substrate recruitment of hepatocyte activity. *J. Pharmacokinet. Biopharm.*, *15* (5), 473–496.

Mu, Y. *et al.* (2006). Traditional Chinese medicines Wu Wei Zi (Schisandra chinensis Baill) and Gan Cao (Glycyrrhiza uralensis Fisch) activate pregnane X receptor and increase warfarin clearance in rats. *J. Pharmacol. Exp. Ther.*, *316* (3), 1369–1377.

Munzel, P.A. *et al.* (1998). Aryl hydrocarbon receptor-inducible or constitutive expression of human UDP glucuronosyltransferase UGT1A6. *Arch. Biochem. Biophys.*, *350* (1), 72–78.

Nagata, K., Yamazoe, Y. (2000). Pharmacogenetics of sulfotransferase. *Annu. Rev. Pharmacol. Toxicol.*, *40*, 159–176.

Negishi, M. (2000). Nuclear receptor CAR as a phenobarbital induction signal of CYP2B gene [abstract]. *FASEB J.*, *14*, 1306.

Nolte, R.T. *et al.* (1998). Ligand binding and co-activator assembly of the peroxisome proliferator-activated receptor-gamma. *Nature*, *395* (6698), 137–143.

Ogawa, K. *et al.* (2000). Characterization of inducible nature of MRP3 in rat liver. *Am. J. Physiol. Gastrointest. Liver Physiol.*, *278* (3), G438–446.

Park, Y., Li, H., Kemper, B. (1996). Phenobarbital induction mediated by a distal CYP2B2 sequence in rat liver transiently transfected in situ. *J. Biol. Chem.*, *271* (39), 23725–23728.

Picard, D., Yamamoto, K.R. (1987). Two signals mediate hormone-dependent nuclear localization of the glucocorticoid receptor. *Embo J.*, *6* (11), 3333–3340.

Pratt, W.B., Silverstein, A.M., Galigniana, M.D. (1999). A model for the cytoplasmic trafficking of signalling proteins involving the hsp90-binding immunophilins and p50cdc37. *Cell Signal.*, *11* (12), 839–851.

Qian, Y.M. *et al.* (2001). Targeted disruption of the mouse estrogen sulfotransferase gene reveals a role of estrogen metabolism in intracrine and paracrine estrogen regulation. *Endocrinology*, *142* (12), 5342–5350.

Radominska-Pandya, A. *et al.* (1999). Structural and functional studies of UDP-glucuronosyltransferases. *Drug Metab. Rev.*, *31* (4), 817–899.

Rae, J.M. *et al.* (2001). Rifampin is a selective, pleiotropic inducer of drug metabolism genes in human hepatocytes: studies with cDNA and oligonucleotide expression arrays. *J. Pharmacol. Exp. Ther.*, *299* (3), 849–857.

Raucy, J.L. *et al.* (2002). Expression and induction of CYP2C P450 enzymes in primary cultures of human hepatocytes. *J. Pharmacol. Exp. Ther.*, *302* (2), 475–482.

Rosenfeld, J.M. *et al.* (2003). Genetic profiling defines the xenobiotic gene network controlled by the nuclear receptor pregnane X receptor. *Mol. Endocrinol.*, *17* (7), 1268–1282.

Rushmore, T.H., Pickett, C.B. (1990). Transcriptional regulation of the rat glutathione S-transferase Ya subunit gene. Characterization of a xenobiotic-responsive element controlling inducible expression by phenolic antioxidants. *J. Biol. Chem.*, *265* (24), 14648–14653.

Safe, S. (2001). Molecular biology of the Ah receptor and its role in carcinogenesis. *Toxicol. Lett.*, *120* (1–3), 1–7.

Saini, S.P. *et al.* (2004). A novel constitutive androstane receptor-mediated and CYP3A-independent pathway of bile acid detoxification. *Mol. Pharmacol.*, *65* (2), 292–300.

Salinas, A.E., Wong, M.G. (1999). Glutathione S-transferases—a review. *Curr. Med. Chem.*, *6* (4), 279–309.

Salphati, L., Benet, L.Z. (1998). Modulation of P-glycoprotein expression by cytochrome P450 3A inducers in male and female rat livers. *Biochem. Pharmacol.*, *55* (4), 387–395.

Schambach, K., Menzel, K. (1975). Clinical experiences using phenobarbital in the prevention of hyperbilirubinemia in mature newborn infants in a controlled study. *Z. Arztl Fortbild (Jena)*, *69* (10), 535–537.

Shelby, M.K. *et al.* (2003). Tissue mRNA expression of the rat UDP-glucuronosyltransferase gene family. *Drug Metab. Dispos.*, *31* (3), 326–333.

Sheweita, S.A. (2000). Drug-metabolizing enzymes: mechanisms and functions. *Curr. Drug Metab.*, *1* (2), 107–132.

Sidhu, J.S., Omiecinski, C.J. (1997). An okadaic acid-sensitive pathway involved in the phenobarbital-mediated induction of CYP2B gene expression in primary rat hepatocyte cultures. *J. Pharmacol. Exp. Ther.*, *282* (2), 1122–1129.

Sinz, M. *et al.* (2006). Evaluation of 170 xenobiotics as transactivators of human pregnane X receptor (hPXR) and correlation to known CYP3A4 drug interactions. *Curr. Drug Metab.*, *7* (4), 375–388.

Song, C.S. *et al.* (2001). Dehydroepiandrosterone sulfotransferase gene induction by bile acid activated farnesoid X receptor. *J. Biol. Chem.*, *276* (45), 42549–42556.

Sonoda, J. *et al.* (2002). Regulation of a xenobiotic sulfonation cascade by nuclear pregnane X receptor (PXR). *Proc. Natl. Acad. Sci. U.S.A.*, *99* (21), 13801–13806.

Sonoda, J. *et al.* (2003). A nuclear receptor-mediated xenobiotic response and its implication in drug metabolism and host protection. *Curr. Drug Metab.*, *4* (1), 59–72.

Staudinger, J.L. *et al.* (2001). The nuclear receptor PXR is a lithocholic acid sensor that protects against liver toxicity. *Proc. Natl. Acad. Sci. U.S.A.*, *98* (6), 3369–3374.

Staudinger, J.L. *et al.* (2003). Regulation of drug transporter gene expression by nuclear receptors. *Drug Metab. Dispos.*, *31* (5), 523–527.

Stedman, C.A. *et al.* (2005). Nuclear receptors constitutive androstane receptor and pregnane X receptor ameliorate cholestatic liver injury. *Proc. Natl. Acad. Sci. U.S.A.*, *102* (6), 2063–2068.

Sueyoshi, T. *et al.* (1999). The repressed nuclear receptor CAR responds to phenobarbital in activating the human CYP2B6 gene. *J. Biol. Chem.*, *274* (10), 6043–6046.

Sugatani, J. *et al.* (2001). The phenobarbital response enhancer module in the human bilirubin UDP- glucuronosyltransferase UGT1A1 gene and regulation by the nuclear receptor CAR. *Hepatology*, *33* (5), 1232–1238.

Synold, T.W., Dussault, I., Forman, B.M. (2001). The orphan nuclear receptor SXR coordinately regulates drug metabolism and efflux. *Nat. Med.*, *7* (5), 584–590.

Takada, T., Suzuki, H., Sugiyama, Y. (2000). Characterization of 5′-flanking region of human MRP3. *Biochem. Biophys. Res. Commun.*, *270* (3), 728–732.

Teng, S., Jekerle, V., Piquette-Miller, M. (2003). Induction of ABCC3 (MRP3) by pregnane X receptor activators. *Drug Metab. Dispos.*, *31* (11), 1296–1299.

Townsend, D.M., Tew, K.D. (2003). The role of glutathione-S-transferase in anti-cancer drug resistance. *Oncogene*, *22* (47), 7369–7375.

Trottier, E. *et al.* (1995). Localization of a phenobarbital-responsive element (PBRE) in the 5′-flanking region of the rat CYP2B2 gene. *Gene*, *158* (2), 263–268.

Tucker, C.M. *et al.* (2001). Self-regulation predictors of medication adherence among ethnically different pediatric patients with renal transplants. *J. Pediatr. Psychol.*, *26* (8), 455–464.

Tukey, R.H., Strassburg, C.P. (2000). Human UDP-glucuronosyltransferases metabolism, expression, and disease. *Annu. Rev. Pharmacol. Toxicol.*, *40*, 581–616.

Tzameli, I. *et al.* (2000). The xenobiotic compound 1,4-bis[2-(3,5-dichloropyridyloxy)]benzene is an agonist ligand for the nuclear receptor CAR. *Mol. Cell Biol.*, *20* (9), 2951–2958.

Ueda, A. *et al.* (2002). Diverse roles of the nuclear orphan receptor CAR in regulating hepatic genes in response to phenobarbital. *Mol. Pharmacol.*, *61* (1), 1–6.

Uppal, H. *et al.* (2005). Combined loss of orphan receptors PXR and CAR heightens sensitivity to toxic bile acids in mice. *Hepatology*, *41* (1), 168–176.

Uppal, H. *et al.* (2007). Activation of LXRs prevents bile acid toxicity and cholestasis in female mice. *Hepatology*, *45* (2), 422–432.

Walker, D., Htun, H., Hager, G.L. (1999). Using inducible vectors to study intracellular trafficking of GFP- tagged steroid/nuclear receptors in living cells. *Methods*, *19* (3), 386–393.

Wang, H., LeCluyse, E.L. (2003). Role of orphan nuclear receptors in the regulation of drug-metabolising enzymes. *Clin. Pharmacokinet.*, *42* (15), 1331–1357.

Wang, H., Negishi, M. (2003). Transcriptional regulation of cytochrome p450 2B genes by nuclear receptors. *Curr. Drug Metab.*, *4* (6), 515–525.

Watkins, R.E. *et al.* (2001). The human nuclear xenobiotic receptor PXR: structural determinants of directed promiscuity. *Science*, *292* (5525), 2329–2333.

Weaver, R.J. (2001). Assessment of drug-drug interactions: concepts and approaches. *Xenobiotica*, *31* (8–9), 499–538.

Weaver, S.A. *et al.* (2001). Regulatory role of phosphatidylinositol 3-kinase on TNF-alpha-induced cyclooxygenase 2 expression in colonic epithelial cells. *Gastroenterology*, *120* (5), 1117–1127.

Wei, P. *et al.* (2000). The nuclear receptor CAR mediates specific xenobiotic induction of drug metabolism. *Nature*, *407* (6806), 920–923.

Xie, W., Evans, R.M. (2001). Orphan nuclear receptors: the exotics of xenobiotics. *J. Biol. Chem.*, *276* (41), 37739–37742.

Xie, W. *et al.* (2000). Humanized xenobiotic response in mice expressing nuclear receptor SXR. *Nature*, *406* (6794), 435–439.

Xie, W. *et al.* (2001). An essential role for nuclear receptors SXR/PXR in detoxification of cholestatic bile acids. *Proc. Natl. Acad. Sci. U.S.A.*, *98* (6), 3375–3380.

Xie, W. *et al.* (2003). Control of steroid, heme, and carcinogen metabolism by nuclear pregnane X receptor and constitutive androstane receptor. *Proc. Natl. Acad. Sci. U.S.A.*, *100* (7), 4150–4155.

Xie, W. *et al.* (2004). Orphan nuclear receptor-mediated xenobiotic regulation in drug metabolism. *Drug Discov. Today*, *9* (10), 442–449.

Xiong, H. *et al.* (2002). Role of constitutive androstane receptor in the in vivo induction of Mrp3 and CYP2B1/2 by phenobarbital. *Drug Metab. Dispos.*, *30* (8), 918–923.

Yueh, M.F. *et al.* (2003). Involvement of the xenobiotic response element (XRE) in Ah receptor-mediated induction of human UDP-glucuronosyltransferase 1A1. *J. Biol. Chem.*, *278* (17), 15001–15006.

Zelko, I., Negishi, M. (2000). Phenobarbital-elicited activation of nuclear receptor CAR in induction of cytochrome P450 genes. *Biochem. Biophys. Res. Commun.*, *277* (1), 1–6.

Zhai, Y. *et al.* (2007). Activation of pregnane X receptor disrupts glucocorticoid and mineralocorticoid homeostasis. *Mol. Endocrinol.*, *21* (1), 138–147.

Zhang, J. *et al.* (2002). Modulation of acetaminophen-induced hepatotoxicity by the xenobiotic receptor CAR. *Science*, *298* (5592), 422–424.

Zhang, J. *et al.* (2004). The constitutive androstane receptor and pregnane X receptor function coordinately to prevent bile acid-induced hepatotoxicity. *J. Biol. Chem.*, *279* (47), 49517–49522.

Zhou, J. *et al.* (2006). A novel pregnane X receptor-mediated and sterol regulatory element-binding protein-independent lipogenic pathway. *J. Biol. Chem.*, *281* (21), 15013–15020.

Zhou, J., Zhang, J., Xie, W. (2005). Xenobiotic nuclear receptor-mediated regulation of UDP-glucuronosyl-transferases. *Curr. Drug Metab.*, *6* (4), 289–298.

# Characterization of Cytochrome P450 Mechanism-Based Inhibition

DAN ROCK and LARRY C. WIENKERS

## 1 INTRODUCTION

Oxidative metabolism mediated by the superfamily of hepatic enzymes known as the cytochrome P450s represents a significant clearance pathway for many drugs prescribed today (Wienkers and Heath, 2005). Of the possible hepatic cytochrome P450 enzymes identified to date, P450 1A2, P450 2C8, P450 2C9, P450 2C19, P450 2D6, and P450 3A4/5 are responsible for greater than 80% of the known oxidative drug metabolism reactions (Fig. 1). In light of the small number of P450 enzymes responsible for the metabolism of a large number of drugs, it is not surprising that a significant portion of drugs possess overlapping P450 enzyme specificities (Guengerich, 1997). A consequence of this disproportionate ratio between the numbers of enzymes and substrates is that for many multidrug therapies, it is common for a patient to receive two or more drugs that interact with the same P450 enzyme. Under these polytherapeutic conditions, a patient is significantly at risk for incurring a P450-based, drug–drug interaction (DDI) (Black *et al.*, 1996; Friedman *et al.*, 1999). There are many clinically important mechanism-based P450 inhibitors that span several therapeutic areas including *antiarrhythmics* [e.g. amiodarone (Polasek *et al.*, 2004)], *antibacterials* [e.g. clarithromycin (Ito *et al.*, 2003) and troleandomycin (Yamazaki and Shimada, 1998)], *antidepressants* [e.g. fluoxetine (Murray and Murray, 2003) and paroxetine (Hara *et al.*, 2005)], *anti-HIV agents* [e.g. ritonavir (von Moltke *et al.*, 2000) and delavirdine (Voorman *et al.*, 1998)], *antihypertensives* [e.g. diltiazem (Jones *et al.*, 1999), and verapamil (Wang, Jones and Hall, 2005)], *nonsteroidal anti-inflammatory* drugs (NSAIDs) [suprofen (O'Donnell *et al.*, 2003) and zileuton (Lu *et al.*, 2003)], *steroids/receptor modulators* [e.g. gestodene (Guengerich, 1990), raloxifene (Chen *et al.*, 2002; Baer, Wienkers and Rock, 2007)], and methylenedioxymethamphetamine (MDMA) (Van *et al.*, 2007), and *oncology drugs* [e.g. tamoxifen (Zhao *et al.*, 2002) and irinotecan (Hanioka *et al.*, 2002)]. In addition, P450 mechanism-based inhibitors (MBIs) can be found in environmental sources such as *illicit drugs* [e.g. phecyclidine (PCP) (Jushchyshyn *et al.*, 2005)] and *various dietary*

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

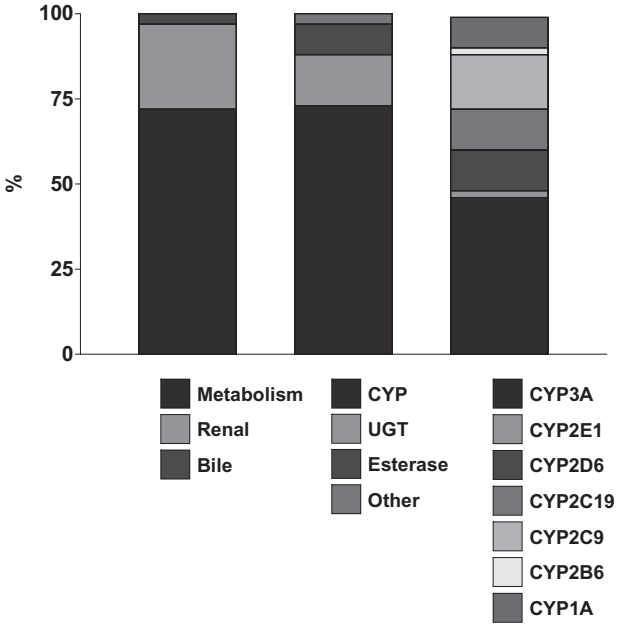


**Figure 1** A graphical breakdown of the various pathways that contribute to the clearance of the top 200 most prescribed drugs in 2002.

*constituents* [e.g. bergamottin (He *et al.*, 1998), resveratrol (Chan and Delucchi, 2000), and 8-methoxypsoralen (Koenigs and Trager, 1998a)].

The magnitude of a clinically observed DDI is primarily dependent upon two features. The number of relevant clearance pathways associated with the victim drug's clearance. Restated, the magnitude of change in the systemic concentration of a drug (area under the curve, AUC) is dependent upon the number of elimination pathways associated with clearance (Bertz and Granneman, 1997; Michalets, 1998). For example, for drugs that are principally cleared by P450 3A enzymes, coadministration of potent MBIs of P450 3A4 can yield large changes in drug exposure (Thummel and Wilkinson, 1998). Secondly, how well the increase in victim drug concentration is tolerated by the system. For compounds with a narrow therapeutic index, an increase in drug exposure, through enzyme inhibition, may precipitate untoward side effects, including some fatal events (Fig. 2). A particularly tragic example of this situation occurred, when potent P450 3A4 inhibitors were coadministered with terfenadine (a P450 3A4 substrate) (Rodrigues *et al.*, 1995), where the subsequent increase in terfenadine exposure precipitated *torsades de pointes* (a life-threatening ventricular arrhythmia associated with QT prolongation) in many patients (Monahan *et al.*, 1990; Honig *et al.*, 1992). Because of this example and others similar to it, global regulatory agencies now require a complete understanding of the potential for drug inhibition of P450 enzymes for every new molecular entity (Davit *et al.*, 1999; Bjornsson *et al.*, 2003).

As described in Chapter Enzyme Inhibition, inhibition of cytochrome P450 enzymes can be broadly classified into two categories: reversible (e.g. competitive and noncompetitive) inhibition and mechanism-based (e.g. quasi-irreversible inhibition and irreversible) inhibition (Fig. 3). Interestingly, when compared with reversible inhibitors of P450 enzymes, MBIs of P450 enzymes are more frequently associated with pharmacokinetic-based

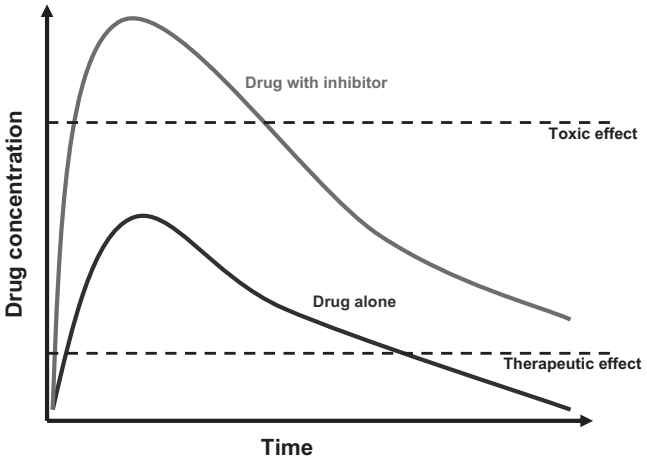


**Figure 2** A simple cartoon depicting an imaginary time–concentration profile that describes the relationship between drug exposure and toxicity.

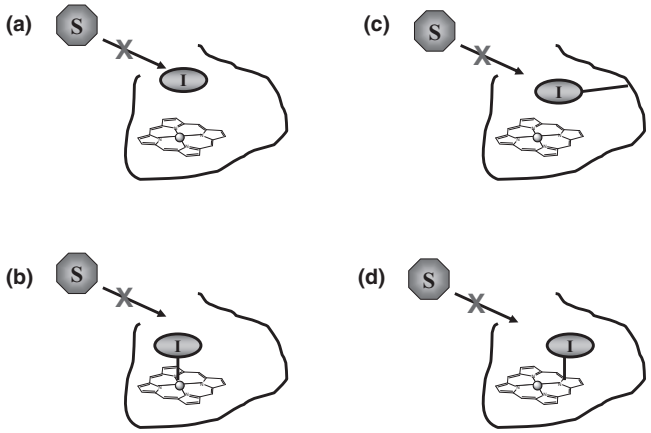


**Figure 3** A simple schematic describing the possible mechanisms by which an inhibitor (I) affects cytochrome P450 substrates (S) metabolism: **(a)** reversible inhibition, **(b)** quasi-irreversible inhibition, **(c)** irreversible inhibition (apoprotein adduction), and **(d)** irreversible inhibition (heme adduction).

DDIs (Kanamitsu *et al.*, 2000; Ghanbari *et al.*, 2006). This observation is primarily a consequence that the inactivated P450 enzyme has to be replaced by newly synthesized P450 protein before the system is capable of metabolizing the victim drug (Bartkowski *et al.*, 1989). However, predicting the clinical significance of a DDI involving P450 enzyme inactivation is difficult, as the outcomes of the interactions are highly dependent upon a number of factors including (i) the particular P450 enzyme involved (i.e. the zero-order synthesis rate of new or replacement enzyme), (ii) the mechanism of P450 enzyme inactivation (i.e. the inhibitors inhibitory constant ($K_I$), $k_{inact}$, and partition ratio), and (iii) the disposition of the inhibited or victim drug (i.e. the number of pathways associated with drug elimination) (Zhou *et al.*, 2005; Kalgutkar, Obach and Maurer, 2007; Riley, Grime and Weaver, 2007).

Cytochrome P450-mediated oxidative metabolism of xenobiotics has been described for over 50 years (Axelrod, 1955; Brodie, Gillette and La Du, 1958), where the enzymatic basis for these reactions was first linked to a unique carbon monoxide-binding pigment found in rat liver microsomes (Klingenberg, 1958). Omura and Sato identified additional properties of the microsomal system in 1962 and coined the name P450 (first used with the hyphen, "P-450") for "pigment 450" because of the absorbance maximum at 450 nm when carbon monoxide was bound to the enzyme in the reduced state (Omura and Sato, 1962). This observation was rapidly followed by experimental evidence, which began to accumulate, for the presence of more than one form of microsomal cytochrome P450, followed by extensive efforts by a number of laboratories to purify P450 enzymes and catalog the oxidative reactions associated with the enzymes (Lu and Coon, 1968). To accommodate the expanding number and diversity of P450 enzymes, a systematic classification based upon protein sequence was developed (Nebert *et al.*, 1987). This system allowed P450 enzymes to be grouped into families, subfamilies, and single enzymes. In this system, the protein sequences within a given gene family are at least 40% identical (e.g. P450 2), the sequences within a given subfamily are >55% identical (e.g. P450 2D and P450 2C), with the last number designating the particular P450 enzyme (e.g. P450 2D6) (Nelson *et al.*, 1996; Nelson, 1999). As stated earlier, in humans, realistically the most important P450s from the point of view of drug metabolism are P450 1A2, P450 2C9, P450 2C19, P450 2D6, and P450 3A4/5 (Wienkers and Heath, 2005).

Prior to developing an in-depth discussion on cytochrome P450 MBIs, it may be worth reviewing the general concepts associated with enzyme catalysis and classical mechanism-based inhibition. In simple terms, in order for a chemical reaction to occur, reactant molecules must possess sufficient energy (activation energy) to cross a potential energy barrier (Fersht, 1974). In this instance, the lower the potential energy barrier to reaction, the more reactants have sufficient energy and in turn the faster the reaction will occur. The high-energy state of the reactants is called the transition state. To illustrate the concept, consider a bond-breaking reaction in which the transition state may be one where the reacting bond, although not completely broken, is vibrating at a high enough frequency that the transition state intermediate is as likely to split apart as to reform (Wolfenden and Snider, 2001). As depicted in Fig. 4, the forming reactants or products result in the loss of energy from the transition state. All enzyme reactions (even reactions that precipitate enzyme destruction) function by forming a transition state, with the reactants, of lower free energy than would be found in the uncatalyzed reaction.

There are several means by which enzymes are able to decrease activation energy. The most important of these involves the enzyme initially binding the substrate, in a fashion often described as "lock and key" orientation, which allows the molecule to exist in close proximity to the catalytic groups found within the active enzyme complex (Koshland and Neet, 1968). When the substrate is bound to the enzyme, the enzyme changes conformation and forces the substrate into a strained or distorted structure that mimics the transition state. In this conformation, the substrate is forced into a reactive state, due to the loss of the substrate's translational and rotational entropy, toward the total activation energy.

The traditional view of "lock and key" does not adequately describe the relationship between substrate and enzyme for cytochrome P450 oxidative reactions. As mentioned earlier, the number of chemotypes metabolized by P450 enzymes is quite large while the number of enzymes that carry out these reactions is exceptionally small. This is due

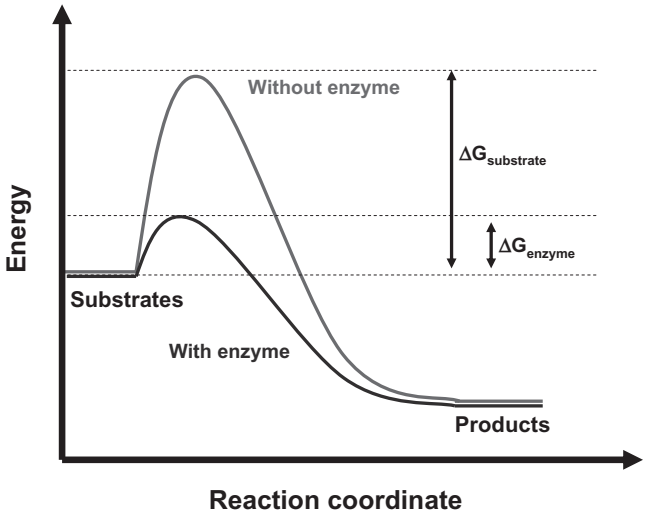


**Figure 4** A schematic diagram showing the free energy profile of the course of an enzyme-catalyzed reaction. The catalyzed reaction pathway (blue line) goes through the transition state, with standard free energy of activation $\Delta G_{enzyme}$, whereas the uncatalyzed reaction (red line) goes through the transition state with standard free energy of activation $\Delta G_{substrate}$. It should be noted that the energetics of catalysis only relate the lowering of the standard free energy of activation from $\Delta G_{substrate}$ to $\Delta G_{enzyme}$ and does not effect the overall free energy change (i.e. the difference between the initial and final states).

to two unique attributes associated with the P450 enzymes. First, the P450s all possess large lipophilic active sites, which are able to accommodate large molecules such as cyclosporine (Fabre *et al.*, 1988). In fact, taken as a whole, P450 enzymes appear to take on a similar structural fold, while in some cases having less than 20% sequence identity and recognizing structurally different substrates (Hasemann *et al.*, 1995). Second, the nature of the P450 oxidative reaction toward nonactivated hydrocarbons is extraordinary efficient. To place the P450 reaction in perspective, oxidations of hydrocarbons of this nature would typically require extremely high temperature and would be nonspecific. At the center of P450 catalysis is the iron protoporphyrin IX (heme) with a thiolate of the conserved cysteine residue as the fifth ligand (White and Coon, 1982).

The sequence of events that make up the P450 catalytic cycle is shown in Fig. 5. The simplified P450 cycle begins with heme iron in the ferric state. In step (i), the substrate (R-H) binds to the enzyme, somewhere near the distal region of the heme group and disrupts the water lattice within the enzyme active site (Sligar, 1976). The loss of water elicits a change in the heme iron spin state (from low-spin to high-spin) (Poulos, 1996). Step (ii) involves the transfer of an electron from NADPH via the accessory flavoprotein NADPH-P450 reductase, with the electron flow going from the reductase prosthetic group FAD to FMN to the P450 enzyme (Vermilion and Coon, 1978). The enzyme, now in the ferrous state can bind oxygen (iii) (Estabrook *et al.*, 1971). The resulting iron–oxygen complex is unstable and can generate ferric iron and superoxide anion, or a second electron may enter the system in step (iv). This second electron may come from NADPH-P450 reductase or, in some cases, from cytochrome $b_5$ (Hildebrandt and Estabrook, 1971). In step (v), a proton is added and the O–O bond

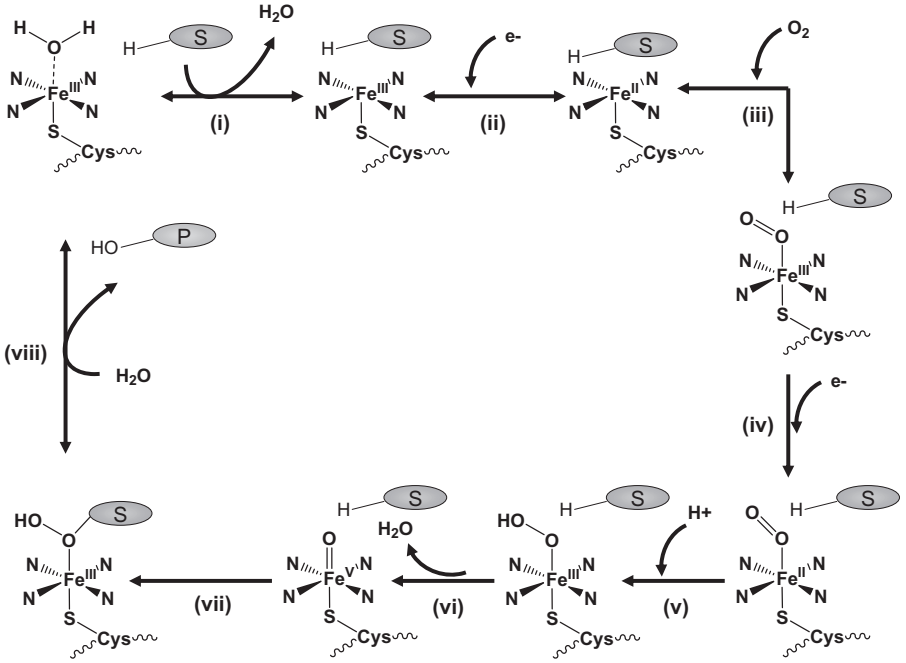


**Figure 5** The catalytic cycle for cytochrome P450 enzyme oxidation reactions.

is then cleaved in step (vi) (Karuzina and Archakov, 1994), yielding $H_2O$ and an entity shown in the scheme as FeO3+. In step (vii), the existing electron-deficient heme complex abstracts either a hydrogen atom or an electron from the bound substrate (Groves and McClusky, 1978). The subsequent intermediate collapses and the metabolite (R-OH) dissociates from the enzyme active site (viii).

Unlike many of the enzymes described earlier in this chapter, the rate-limiting step in cytochrome P450 enzyme catalysis is the activation of molecular oxygen leading to a highly reactive species (Trager, 1982). Based upon the results of several mechanistic studies, it appears that the electronic characteristics of the iron and oxygen species dictate the rates of substrate oxidation for cytochrome P450 enzymes (Fig. 6). As a consequence, the tendency for oxidation of a certain functional group generally follows the relative stability of the radicals that are formed (e.g. *N*-dealkylation > *O*-dealkylation > 2° carbon oxidation > 1° carbon oxidation) (Jones, Shou and Korzekwa, 1996).

While given that these two attributes (reactivity of heme and promiscuous active site) of the P450 enzymes suggest an almost limitless possibility of drug metabolism reactions, the oxygenation of substrates by these enzymes is not completely indiscriminate. For example, the nature of binding and/or oxidation of many P450 substrates can be both region- and stereochemically specific. In the case of warfarin, the anticoagulant medication, it exists as a racemic mixture. Both enantiomers share affinities toward P450 2C9 of less than 10 μM (~4 μM for (*S*)-warfarin and 8 μM for (*R*)-warfarin) (Kunze *et al.*, 1991). However, while both stereoisomers share relative equal affinities toward P450 2C9, only (*S*)-warfarin is metabolized by the enzyme (Rettie *et al.*, 1992). In a second example, quinidine is a pharmaceutical agent that acts as a class I antiarrhyth-

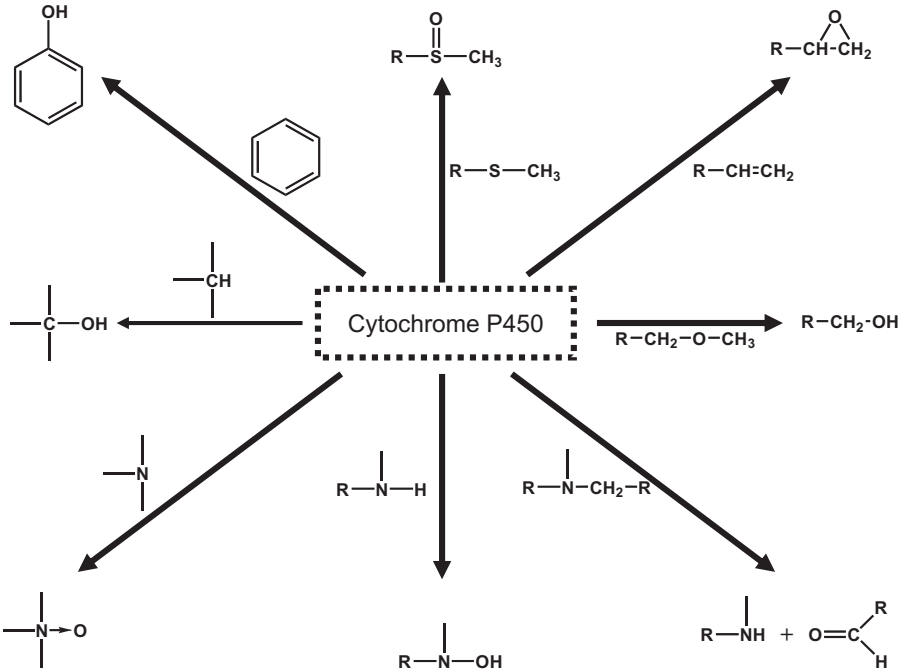


**Figure 6** A brief listing of common P450 oxidation reactions.

mics agent and also a stereoisomer of quinine. Both isomers bind to and inhibit P450 2D6, albeit with markedly different affinities (quinidine has submicromolar $K_i$ values while quinine has a $K_i$ of about 50 μM). Besides possessing radically different affinities for the same enzyme, the isomers also bind completely different within the P450 2D6 active site; quinidine binding orientation yields a type II binding spectra while quinine generates a type I (Hutzler, Walker and Wienkers, 2003).

In simple terms, an enzyme inhibitor is a molecule that is able to prevent a substrate from entering the enzyme's active site and/or preclude the enzyme from catalyzing its reaction. As mentioned earlier, inhibitor binding is either reversible or irreversible. Reversible inhibitors bind with the enzyme through a variety of non-covalent interactions and based upon these binding motifs, elicit different types of inhibition, depending on whether the inhibitor binds the enzyme, the enzyme–substrate complex, or both. In contrast, an irreversible (mechanism-based) inhibitor is an enzyme inhibitor that contains a latent functional group that is chemically unreactive in starting material, but can be activated to a highly reactive intermediate by the enzyme through some part of the normal catalytic mechanism that is effected by the enzyme. The activated species then binds irreversibly to the enzyme and thereby inactivates it.

Given that the reactive groups present within the enzyme's active site are nucleophiles (such as hydroxyl or sulfhydryl groups), irreversible inhibition via covalent addition of reactive species to the enzymes requires that the reactive species needs to be some form of electrophilic moiety. For many enzymes, the formation of a reactive species (using the catalytic mechanism of the enzyme) falls into a few distinct categories, such as the generation of Michael acceptors, haloalkyl derivatives, and rearrangements leading to an acyl-enzyme intermediate (e.g. proteases) (Silverman, 1988).

Initially, characterization of MBIs followed an intuitive kinetic and chemical criteria (Wang and Walsh, 1978). Briefly an MBI was required to be a substrate, where physical binding to the enzyme active site precedes catalysis. The interaction between substrate and enzyme should reflect a binding equilibrium followed by first-order chemical process leading to inactivation. To this end, as inhibitor concentration is increased, a greater fraction of total enzyme will be occupied with inhibitor and that saturation of enzyme with inhibitor will be consistent with the kinetic specificity of the initial binding step. In addition, if covalent modification of the enzyme is a mechanism-based process which proceeds via adduct formation to a key amino acid residue within the enzyme active site, the chemical stoichiometry of modification should be unity (1:1 adduct/enzyme ratio) (Wang and Walsh, 1978).

As the understanding and laboratory experience evolved in this area, researchers expanded and refined the criteria applied to assess whether a substrate is a mechanism-based inactivator (Wang and Walsh, 1978; Walsh, 1984; Kent, Juschyshyn and Hollenberg, 2001). Briefly, the criteria include the following:

1. The loss of enzyme activity must exhibit time dependence.
2. Enzyme inactivation should exhibit saturation kinetics with respect to the concentration of the inhibitor.
3. The inactivation occurs in a catalytically competent system (e.g. necessary cofactors are present and that metabolism is occurring).
4. The enzyme should be protected from inactivation upon co-incubation with a competitive substrate/inhibitor.
5. Lack of suppression of inactivation by reactive intermediate scavengers.
6. The inactivation should be irreversible and activity should not return upon dialysis or gel filtration.
7. Following inactivation, it should be possible to demonstrate a 1:1 stoichiometry of inactivator to enzyme molecule inactivated.

A generalized kinetic scheme for MBI can be simply described (Fig. 7) as a competition or partition between the catalytic turnover of the inhibitor and enzyme inactivation (Walsh, 1984). Through the integration of the MBI's partition ratio, a mechanistic insight into the catalytic branch points associated with the enzyme reaction as well as

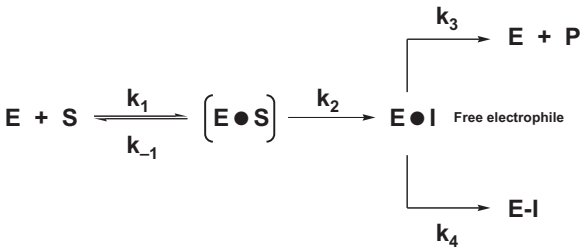


**Figure 7** Reaction scheme of a mechanism-based enzyme inactivator. Rate constants $k_1$ and $k_{-1}$ represent a reversible binding event between the enzyme (E) and substrate (S) to form the enzyme–substrate complex (ES), and $k_2$ represents catalysis of substrate to a reactive intermediate (I). $k_3$ represents the rate constant defining the release of the reactive intermediate represented as product (P), whereas $k_4$ represents inactivation of E by I forming the EI complex. $k_3/k_4$ = partition ratio = amount of formed P/inactivation, EI.

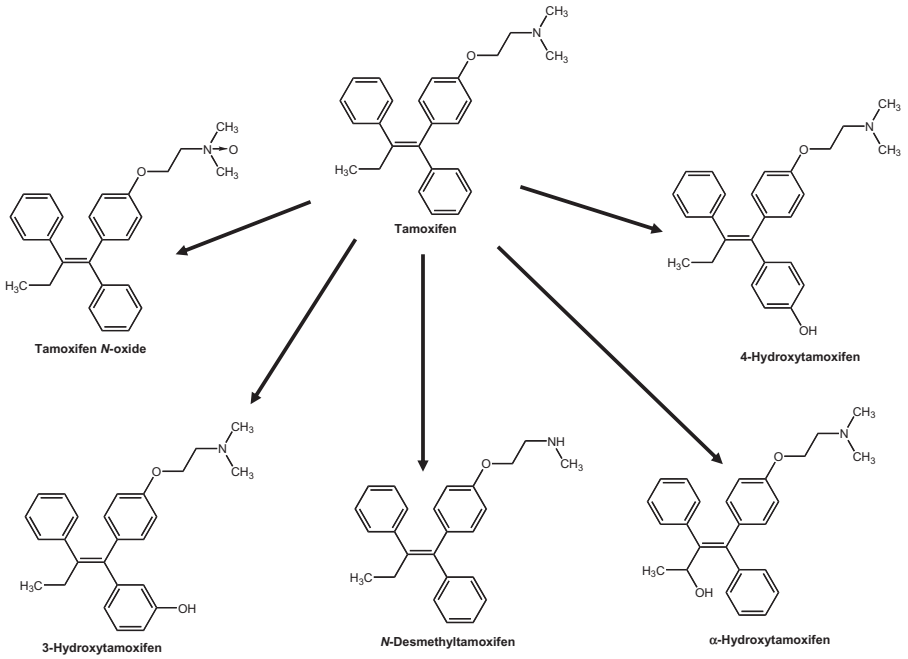


**Figure 8** Multiple P450 3A4 oxidation reactions toward a single substrate, tamoxifen.

the chemical nature of the reactive species in the enzyme active site can be gleaned. In this case, the lower the partition ratio (e.g. the closer to 1), the more efficient the inhibitor is at destroying the enzyme.

For cytochrome P450 MBIs, while metabolic intermediates are quite reactive, low partition ratios associated with MBIs are uncommon. This anomaly between the chemical reactivity of the activated species and partition ratio is explained by the fact that there are normally multiple metabolic pathways associated with many P450 substrates. For many traditional MBIs, the mechanism of enzyme inactivation takes advantage of a single catalytic step between substrate and enzyme. As stated earlier, cytochrome P450 enzymes have large active sites which allow for multiple substrate orientations and thus are susceptible to multiple oxidative reaction. An example of the P450 enzymes' ability to accommodate multiple substrate orientations and carry out a variety of energetically diverse oxidative reactions is seen with the P450 3A4 metabolism of tamoxifen (Fig. 8). In this example, P450 3A4 is able to metabolize tamoxifen to yield five products through four different oxidative reactions (Desta *et al.*, 2004). Adding to the complexity surrounding P450 mechanism-based inhibition is the fact that in some cases, the underlying mechanism of enzyme inactivation from one P450 enzyme does not extrapolate to a different P450 enzyme. For example, subtle sequence differences between P450 3A4 and P4503A5 (e.g. these enzymes share greater than 95% amino acid homology) result in profound differences in susceptibility to mechanism-based inhibition by several inhibitors including verapamil (Wang, Jones and Hall, 2005), tamoxifen (Zhao *et al.*, 2002), and erythromycin (McConn *et al.*, 2004).

As previously described, there are three general classes of mechanism-based inactivators of the cytochrome P450s. They are (1) inhibitors that bind covalently to the apoprotein; (2) inhibitors that interact in a pseudo-irreversible manner with heme iron;

and (3) inhibitors that cause destruction of the prosthetic heme group, oftentimes leading to heme-derived products that covalently modify the apoprotein (Fig. 3). The classes of cytochrome P450 mechanism-based inhibition can be further refined into six common features which serve as latent chemical groups associated with enzyme activation. These include substituted imidazoles, furan rings, thiophenes, alkylamines, methylenedioxyphenyl groups, and acetylenes. The following is a list of the types of P450 mechanism-based inhibition and the chemistry associated with these latent chemical moieties which leads to enzyme inactivation.

## 2 INHIBITORS THAT UPON ACTIVATION, BIND COVALENTLY TO THE P450 APOPROTEIN

### 2.1 Substituted Imidazoles

Putative bioactivation of this latent group requires formation of an imidazomethide, as the electrophilic species (Fig. 9) (Kunze and Trager, 1993). The mechanism(s) for the generation of this reactive species involves initial P450-mediated electron abstraction leading to the formation of a radical intermediate that can undergo oxygen rebound to form the hydroxyl metabolite. Alternatively, the radical may proceed through a second electron or hydrogen atom abstraction to form a reactive imidazomethide species, which can adduct to nucleophilic sites with the P450 enzyme active site or react with water (Hutzler *et al.*, 2004).

### 2.2 Furans

Furans are electron-rich aromatic groups that are readily oxidized by P450 enzymes to form electrophilic species which have been linked to drug-related toxicities (Kouzi, McMurtry and Nelson, 1994) as well as serve as MBIs (Sahali-Sahly *et al.*, 1996). The current hypothesis regarding P450-mediated bioactivation of furan rings (Fig. 10) involves epoxide formation that could either be deactivated via hydrolysis to yield the diol (Khojasteh-Bakht *et al.*, 1999) or react irreversibly with the P450 enzyme, either through direct adduction with the epoxide or through an intramolecular rearrangement of the epoxide to form an α,β-unsaturated carbonyl which serves as a Michael acceptor for protein adduction (Chen, Hecht and Peterson, 1997).

### 2.3 Thiophenes

Currently, the cytochrome P450-mediated bioactivation of thiophene-containing drugs is thought to proceed via two possible mechanisms (Fig. 11). The first mechanism proceeds via P450 oxidation of the thiophene sulfur atom leading to formation of a reactive thiophene-*S*-oxide intermediate. The thiophene-*S*-oxide is susceptible to a Michael-type addition with nucleophilic amino acids within the P450 active site (Lopez-Garcia, Dansette and Mansuy, 1994). The second mechanism stems from a P450-mediated oxidation of the thiophene ring to form a reactive thiophene epoxide metabolite which may react directly with the P450 enzyme (Dansette, Bertho and Mansuy, 2005). Alternatively, the epoxide in a mechanism analogous to furan bioactivation generates a cis-2-butene-1, 4-dialdehyde reactive intermediate. In this instance, the thioketo-α,β-unsaturated aldehyde serves as the electrophilic intermediate that covalently binds to P450 enzyme (O'Donnell *et al.*, 2003).

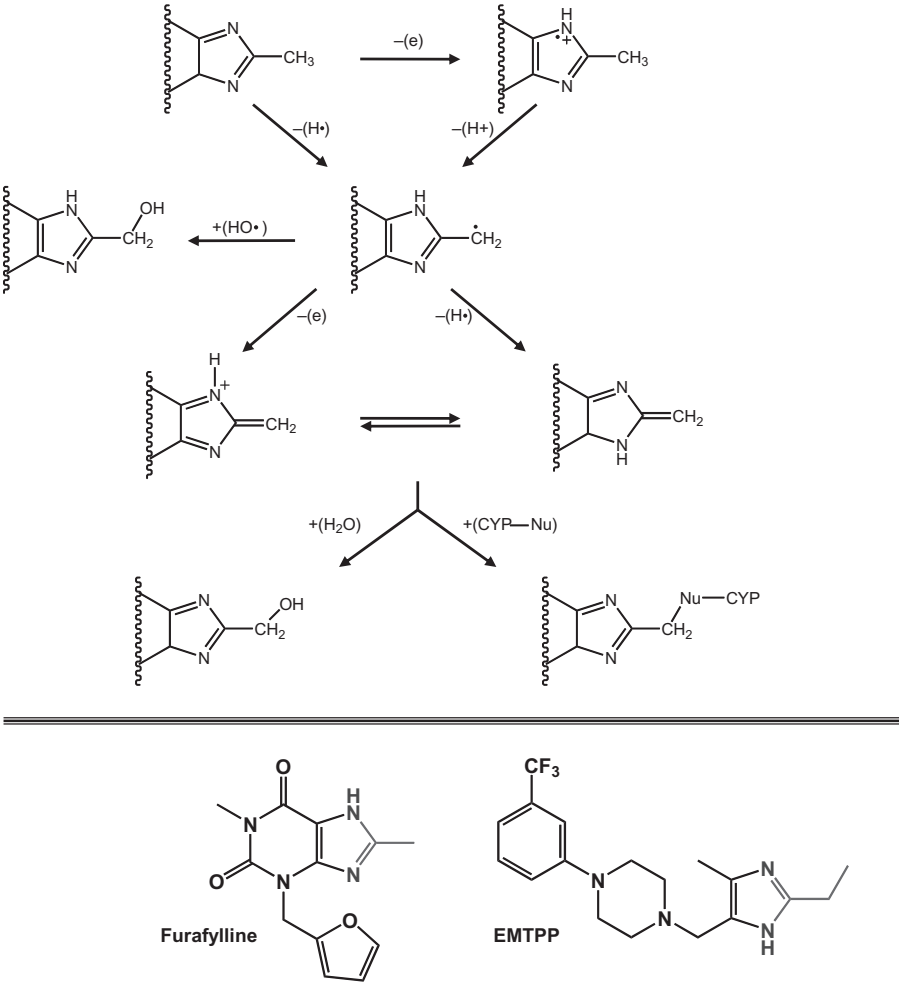


**Figure 9** Mechanistic scheme for P450 enzyme-mediated bioactivation and subsequent enzyme inactivation of a substituted imidazole group. Structure of known P450 MBIs which possess a substituted imidazole as a latent chemical moiety: furafylline (Kunze and Trager, 1993) and 1-[(2-ethyl-4-methyl-1H-imidazol-5-yl)methyl]-4-[4-(trifluoromethyl)-2-pyridinyl]piperazine (EMTPP) (Hutzler *et al.*, 2004).

## 3 INHIBITORS THAT INTERACT IN A PSEUDO-IRREVERSIBLE MANNER WITH HEME IRON

### 3.1 Alkylamines

There are a relatively large number of drugs that contain primary, secondary, or tertiary alkylamines which are also known to be MBIs of cytochrome P450 enzymes (Yamazaki and Shimada, 1998; Jones *et al.*, 1999). The mechanism of inactivation for these molecules proceeds through a pseudo-irreversible interaction with the heme iron described

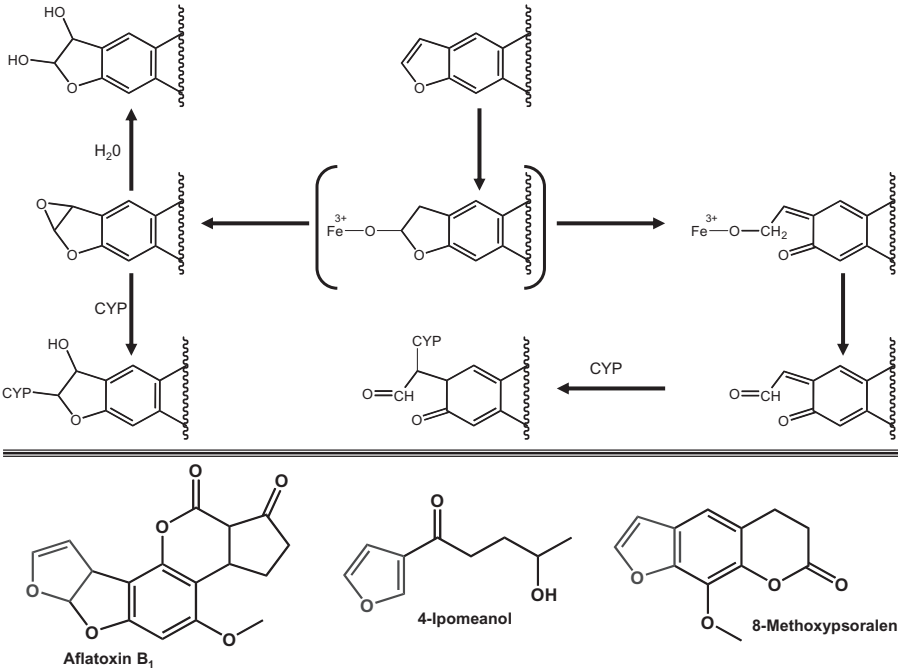


**Figure 10** Mechanistic scheme for P450 enzyme-mediated bioactivation and subsequent enzyme inactivation of a furan group. Structures of known P450 MBIs which possess a furan as a latent chemical moiety: L-754,394 (Sahali-Sahly *et al.*, 1996), 8-methoxypsoralen (Koenigs and Trager, 1998a), and 4-ipomeanol (Alvarez-Diez and Zheng, 2004).

as a metabolite intermediate complex (MIC) (Pershing and Franklin, 1982; Ernest, Hall and Jones, 2005). The salient reactive species in this reaction is generally regarded to be a transient nitrosoalkane (Bensoussan, Delaforge and Mansuy, 1995). To generate the proposed reactive species requires a primary or secondary amine to serve as a precursor to a nitroso species. This involves an initial P450-catalyzed dealkylation step, followed by hydroxylation of the nitrogen to form a secondary hydroxylamine. The subsequent hydroxylamine is then further oxidized to yield a nitrone, which can undergo hydrolysis or be converted to the reactive nitroso species through two-electron oxidation process (Bensoussan, Delaforge and Mansuy, 1995). The schematic for MIC formation and examples of MBIs, which react through this pathway, are provided in Fig. 12.

### 3.2 Methylenedioxyphenyls

The P450 mechanism for bioactivation of this latent chemical moiety is postulated to proceed via proton abstraction of the dioxymethylene bridge which encounters a metabolic branch point of undergoing either typical oxidative demethylation to generate the catechol and formic acid, or intramolecular rearrangement to yield a carbine that is able to coordinate with the heme iron (Fig. 13). Formation of a carbene can arise by abstraction of a hydrogen atom from the methylenic carbon, or by elimination of water from a hydroxymethylene intermediate (Lin *et al.*, 1996; Hutzler *et al.*, 2006). The

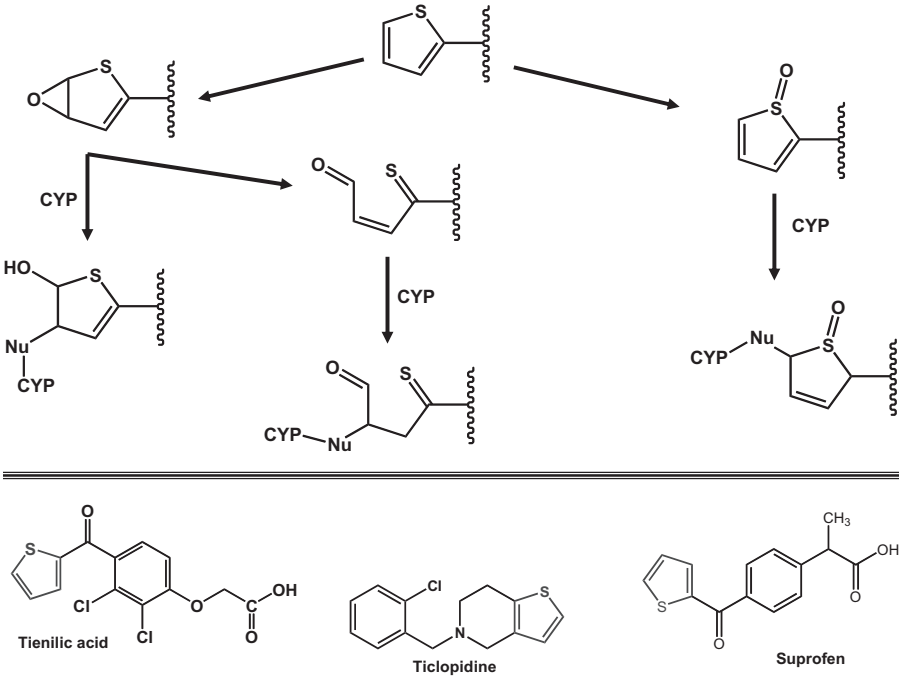


**Figure 11** Mechanistic scheme for P450 enzyme-mediated bioactivation and subsequent enzyme inactivation of a thiophene group. Structure of known P450 MBIs which possess a thiophene as a latent chemical moiety: tienilic acid (Beaune *et al.*, 1987), ticlopidine (Ha-Duong *et al.*, 2001), and suprofen (O'Donnell *et al.*, 2003).

resulting dioxymethylene MI complex with P450 enzymes appears to be stable when the heme is in both the ferric or ferrous state. However, MI complexes with ferrous iron are thought to be more stable than those with ferric iron, due to the potential for back-bonding of electrons from the ferrous iron to the carbine carbon (Dickins *et al.*, 1979).

## 4 INHIBITORS THAT CAUSE DESTRUCTION OF THE PROSTHETIC HEME GROUP, OFTENTIMES LEADING TO HEME-DERIVED PRODUCTS THAT COVALENTLY MODIFY THE APOPROTEIN

### 4.1 Acetylenes

Many compounds containing an acetylenic group have been shown to be mechanism-based inactivators of P450 enzymes (Foroozesh *et al.*, 1997). For this functional moiety, two mechanisms for the inactivation of P450 have been described in detail (Chan, Sui and Ortiz de Montellano, 1993). The first involves the transfer of oxygen from the P450-activated oxygen intermediate to the internal carbon of the acetylene which generates a reactive intermediate that leads to heme alkylation and destruction of the heme chromophore. Alternatively, the transfer of oxygen from the activated oxygen intermediate of the P450 to the terminal carbon of the acetylene would result in an

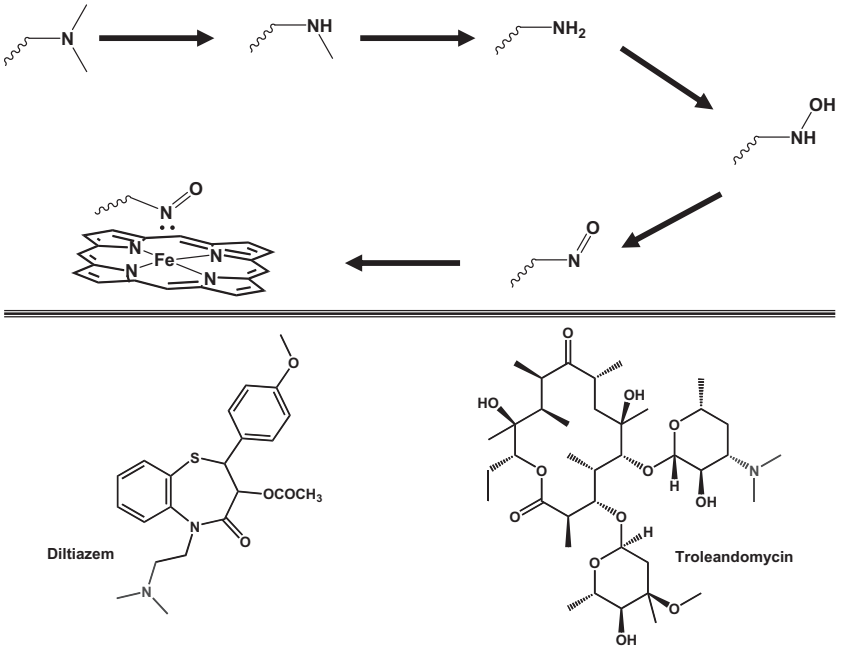


**Figure 12** Mechanistic scheme for P450 enzyme-mediated bioactivation and subsequent enzyme inactivation of alkylamines. Structure of known P450 MBIs which possess an alkylamine as a latent chemical moiety: diltiazem (Jones *et al.*, 1999) and troleandomycin (Yamazaki and Shimada, 1998).

intermediate that would rearrange via a 1,2-shift of the terminal hydrogen to the vicinal carbon to generate a ketene. The reactive ketene species produced by this rearrangement can then be hydrolyzed to produce the carboxylic acid product, or it can acylate nucleophilic residues within the P450 active site to inactivate the protein (Fig. 14).

The objectives of the first portion of this chapter were to serve as a primer to highlight the uniqueness of cytochrome P450 reactions compared with many other enzyme systems, and to develop the relationship between structure of the inhibitor and catalytic function of the enzyme as it relates to P450 mechanism-based inhibition. The remainder of this chapter will focus on expanding these concepts as it relates to conducting studies to characterize the nature and inhibitory potential of mechanism-based inactivators which lead to irreversible destruction of the P450 enzymes.

Characterizing the mechanisms by which a new chemical entity (NCE) inhibits P450s is critical for the development of safe therapeutics as the magnitude of a drug interaction will typically be increased for an MBI (Fig. 15). In all cases, the precursor to P450 MBI results from bioactivation to generate a reactive chemical intermediate and therefore most MBIs can be linked to the generation of reactive metabolites (Evans *et al.*, 2004; Masubuchi, Makino and Murayama, 2007). As a result, MBIs increase the risk to patient safety through three basic pathways: (i) producing unexpectedly high drug levels of the MBI itself or an alternate drug cleared through the P450 affected by the MBI (Fig. 2); (ii) eliciting immune-induced hepatitis potentially as a result of the direct alkylation event; and (iii) oxidative stress through redox cycling and

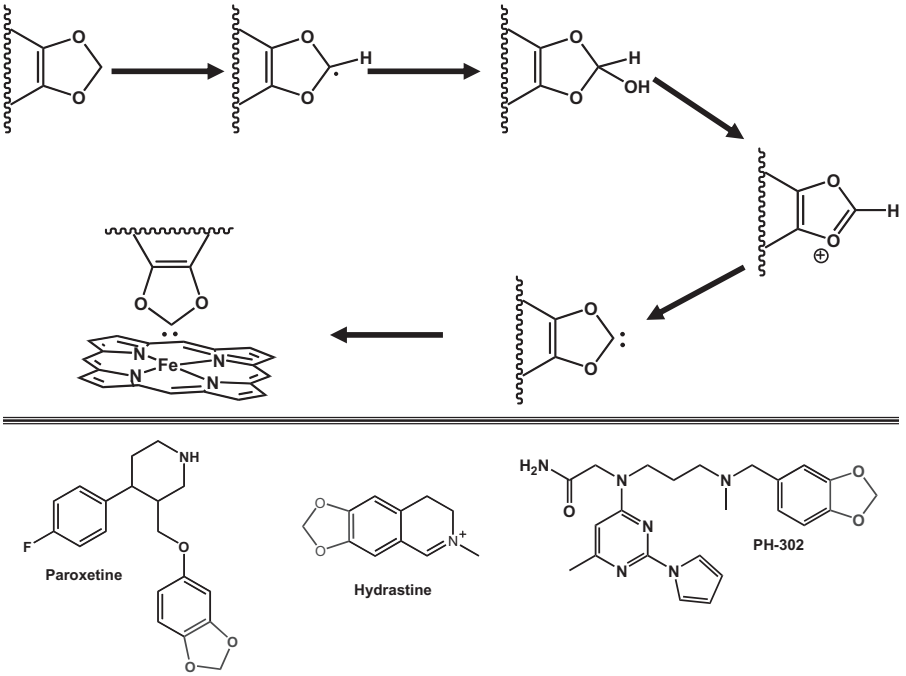


**Figure 13** Mechanistic scheme for P450 enzyme-mediated bioactivation and subsequent enzyme inactivation of methylenedioxyphenyl groups. Structure of known P450 MBIs which possess a methylenedioxyphenyl as a latent chemical moiety: paroxetine (Hara *et al.*, 2005) and PH-302 (Hutzler *et al.*, 2006).

depletion of endogenous glutathione. Regardless of the mode of action, the presence of MBI increases the potential safety risk associated with an NCE.

The most favorable scenario is to eliminate the structural feature(s) responsible for irreversible P450 inhibition, thereby reducing the potential for P450 inhibition and minimizing the risk of additional adverse drug reactions. The selective modification of a chemical series to eliminate MBI requires high-resolution structural information in order to pinpoint the mechanism by which the reactive intermediate forms. However, the inability to remove bioactivation may not preclude an NCE from development but will partially depend upon the severity and progression of the disease the NCE is aimed to treat (Kalgutkar and Soglia, 2005). In addition, a major focus in progressing with an MBI is to provide a risk assessment of the potential effects of the MBI by predicting the expected magnitude of drug interactions. This requires the functional enzymology to be defined including the inactivation parameters of MBI and how they are expected to translate *in vivo* based upon the expected clearance mechanisms, dose, and frequency of administration.

The balance between structural and functional characterization tools for interrogating MBI depends upon when MBI for a particular compound was discovered. For example, if the liability is revealed during lead optimization in discovery, a predominantly structural approach could guide chemistry efforts toward design strategies aimed at the removal or reduction in the extent of MBI. However, if the MBI cannot

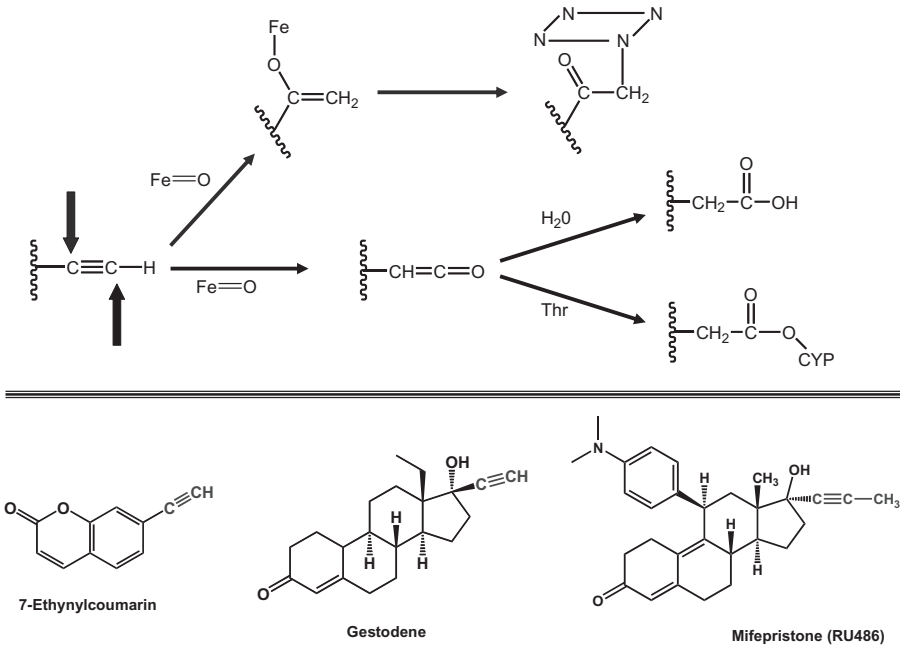


**Figure 14** Mechanistic scheme for P450 enzyme-mediated bioactivation and subsequent enzyme inactivation of an alkyne group. Structure of known P450 MBIs which possess an alkyne group as a latent chemical moiety: gestodene (Guengerich, 1990), mifepristone (RU486) (He, Woolf and Hollenberg, 1999), and 7-ethynylcoumarin (Regal *et al.*, 2000).

be removed or in the unfortunate circumstance that MBI was discovered in preclinical development, a functional approach will dominate research efforts to define the potential impact of the MBI on patient population required to support continuation of the clinical candidate. Both structural and functional characterization is required not only for proper assessment of safety, but also in order to avoid incorporation of similar structural motifs prone to MBI into future chemical templates.

Given the complexity of P450 MBI characterization, a plethora of experimental protocols have evolved with the specific aim of identifying irreversible inhibition. When possible, the primary focus should be on removing MBI from a chemical series which requires the experimental procedures to focus on several key structural and functional end points (Fig. 16). This schematic provides an experimental guidance for the assessment of MBI from which the initial results provide the directive for additional experiments and end points for the continued development of an NCE with MBI liabilities. Therefore, Fig. 16 will serve as an outline for the remainder of the chapter and will cover in detail structural tools used to identify the mechanistic details of MBI and subsequently the functional tools used to determine the kinetics of MBI. A specific focus is given on the different assays used through the progression of a developing NCE, from early removal of the liability in discovery to risk assessment in later stages of development, through the generation of *in vitro* functional data which are used to extrapolate the expected *in vivo* pharmacokinetic impact of MBI on a developing clinical candidate.

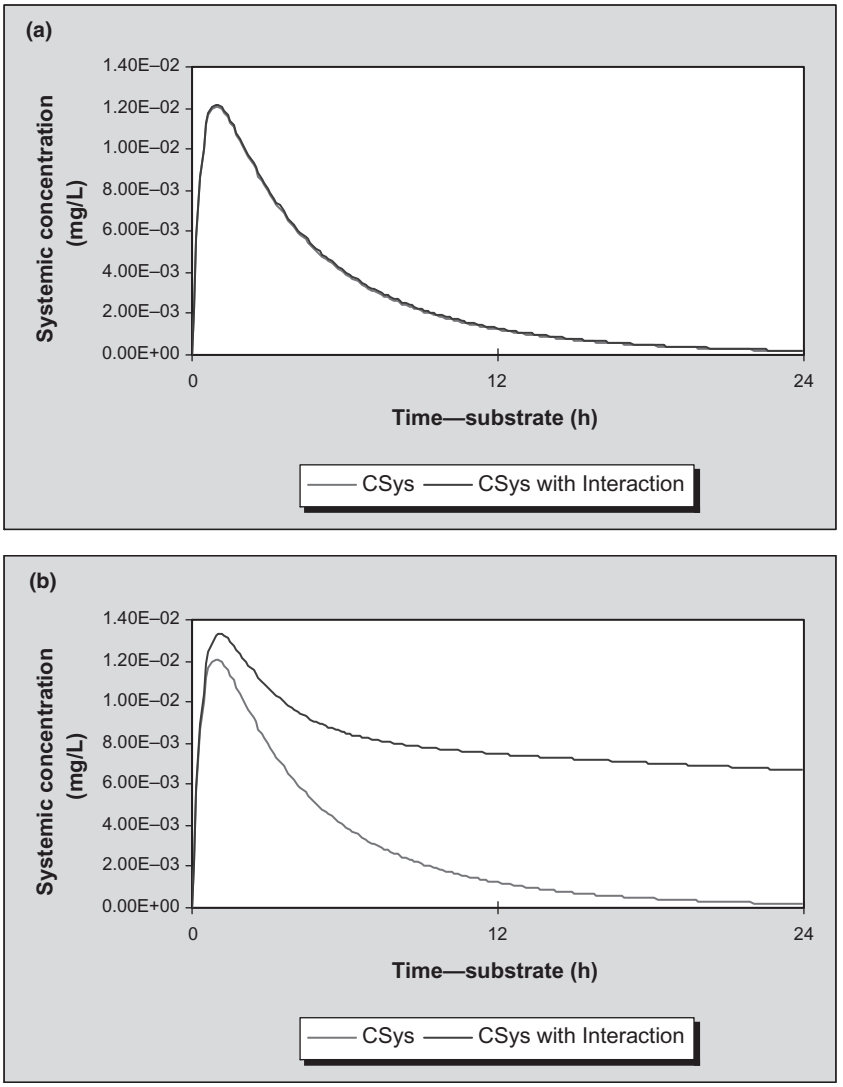


**Figure 15** Simulation of single oral dose pharmacokinetic model of midazolam (5 mg) coadministered with erythromycin (500 mg) using Simcyp (Sheffield, UK). **(a)** Inactivation kinetics were omitted in this simulation resulting in no observed drug interaction. **(b)** $K_I$ and $k_{inact}$ were included in the prediction, highlighting the potential for drug interaction on the basis of erythromycin inactivation of CYP3A4.

Prior to discussing experimental procedures used to study MBI, it is imperative to consider the interactions leading to irreversible P450 inhibition. This provides a conceptual understanding of the MBI process and helps visualize the experimental design requirements necessary to elucidate and remove MBI from a chemical scaffold. The foremost precursor to MBI is that a substrate contains a functional group capable of

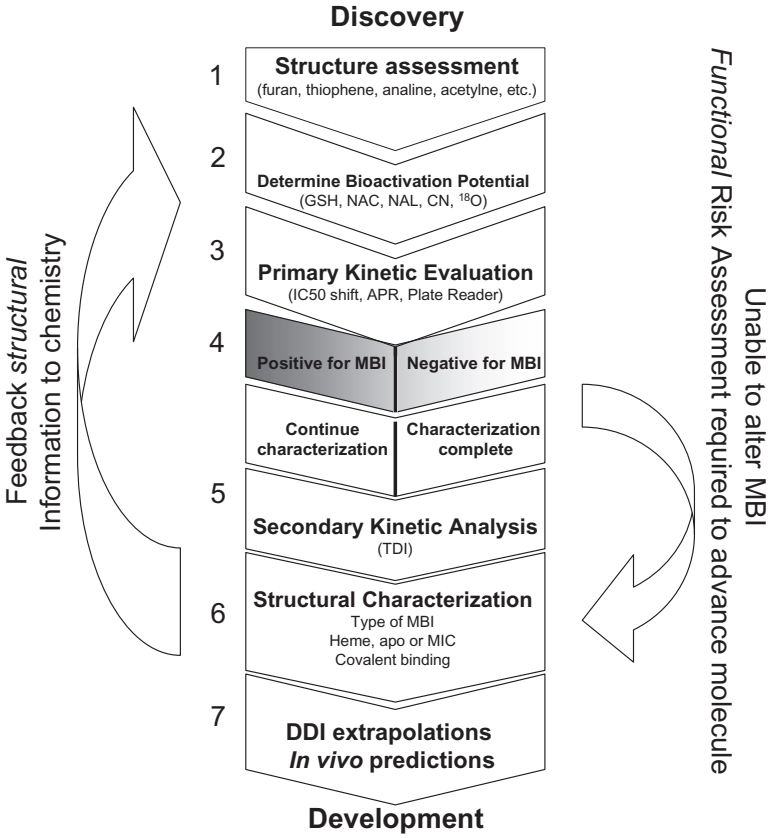


**Figure 16** A schematic of the MBI testing funnel describing the use of structural screening-based assays in discovery which are highlighted by steps 1–6. Structural learnings are then fed back into the chemical design of new compounds until MBI is eliminated. If MBI cannot be attenuated, the program will attempt to manage the risk in development based upon steps 7–8. This schematic will serve as the outline for the remainder of the chapter.

bioactivation as discussed with the seven structural motifs highlighted in Figs 9–14. For P450 MBI is conceptually interesting because the catalytic step is activation of molecular oxygen. This implicitly means substrate oxidation is a by-product of P450 catalysis and consequently substrates do not typically possess a single enzyme–substrate conformation. The repercussions of this trait are that oxidation reactions can occur at multiple positions within a single molecule, favoring the most facile oxidative process unless the substrate is stabilized in a particular binding conformation (Fig. 17). Flurbiprofen metabolism by P450 2C9 can be used to illustrate this point (Fig. 18). The benzyl alcohol of flurbiprofen is the most facile oxidation site; however, the electrostatic interactions of carboxylic acid in flurbiprofen with Arg108 of P450 2C9 position flurbiprofen such that the only 4′-hydroxylation is observed (Fig. 18) (Wester *et al.*, 2004). This particular example shows how P450–substrate interactions may impact the site of metabolism.

| Radical | $\delta\Delta H_f$ Kcal/mole | Reaction type |
|---|---|---|
| $H_2N-\dot{C}H_2$ | 17.3 | *N*-dealkylation |
| 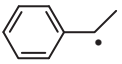 | 19.6 | Benzylic hydroxylation |
| $H_3C-O-\dot{C}H-CH_3$ | 26.6 | *O*-dealkylation |
| $H_3C-\dot{C}H-CH_3$ | 27.7 | Aliphatic hydroxylation |
| | 28.6 | Aliphatic hydroxylation |
| $H_3C-CH_2-\dot{C}H_2$ | 33.0 | ω-Hydroxylation |

Increasing occurrence of metabolism

**Figure 17** The energy differences between different P450-catalyzed reactions based upon hydrogen atom abstraction. The differences in energy have the potential to influence metabolite regioselectivity in P450-mediated metabolism (Jones, Shou and Korzekwa, 1996; Higgins *et al.*, 2001).

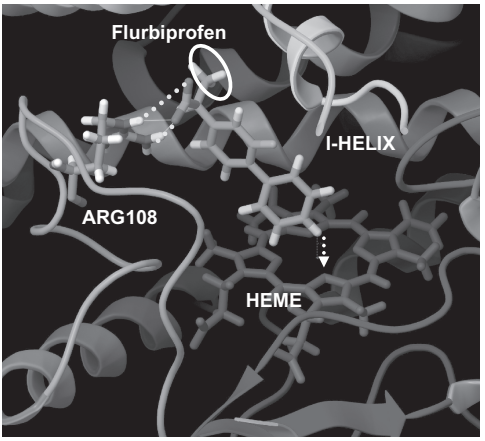


**Figure 18** Crystal structure of P450 2C9 with flurbiprofen bound (1R90.pdb) highlighting the electrostatic interaction of the carboxylic acid of flurbiprofen with the quanidine moiety of side chain from Arg108. The electrostatic interaction is depicted in diagonal dashed line. The site of metabolism is indicated with the white arrow and is positioned directly above the heme, whereas the benzylic site is circled (white) (Wester *et al.*, 2004).

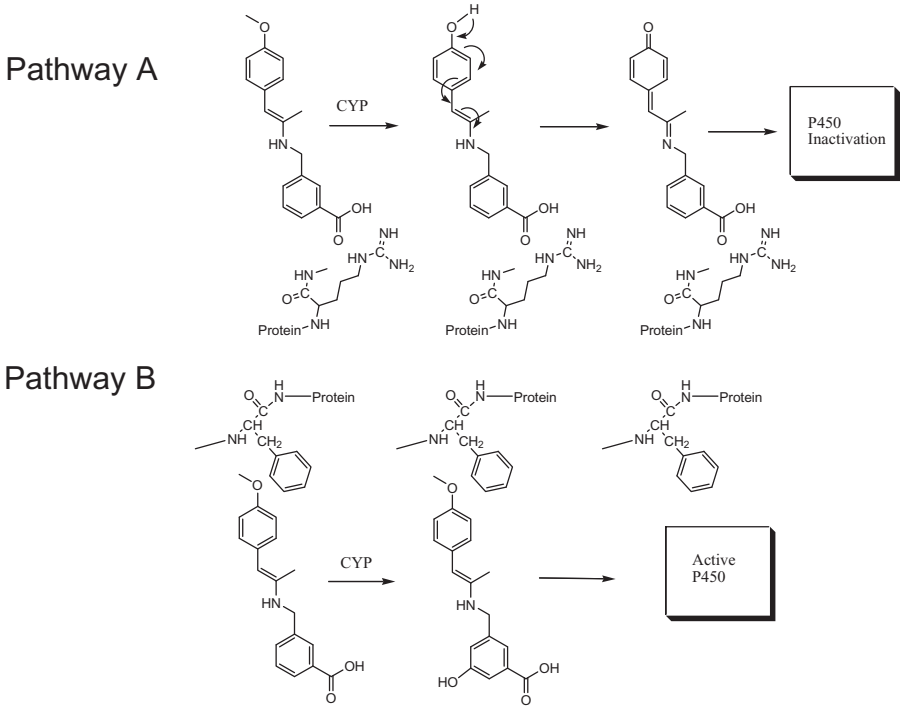


**Figure 19** Hypothetical drug molecule metabolized by two different P450 enzymes. In pathway A, the molecule is positioned such that *O*-dealkylation occurs and leads to the generation of reactive quinone methide metabolite, whereas the metabolism in pathway B leads to formation of benign metabolite because of different metabolite regioselectivity.

As such, specific P450–substrate interactions may dictate whether or not bioactivation occurs. Take the compound in Fig. 19 for example, which has the ability to form a reactive quinone methide upon *O*-dealkylation in pathway A. The compound may be prone to such a reaction if the P450 responsible for metabolism interacts with compound's carboxylic acid group and positions the methoxy group for metabolism. Alternatively, if the compound is metabolized by a different P450 where pi stacking and hydrogen bonding interactions predominate, which favorably orient the same compound for metabolism away from *O*-dealkylation, then the bioactivation step may not occur resulting in the formation of alternate metabolites void of MBI (Fig. 19, pathway B). Therefore, P450–substrate interaction may direct the site of oxidation and influence to the potential propensity for bioactivation significantly contributing to whether or not MBI occurs.

Finally, the active site architecture dictates the ability for MBI. Each P450 has a unique composition of active site amino acids and thus has different nucleophiles capable of trapping reactive metabolites. Structural information offered from P450 crystal structures can provide crucial information regarding the propensity for MBI. To date, there are now structures for most of the key drug-metabolizing enzymes including P450 1A2 (2HI4.pdb) (Sansen *et al.*, 2007), P450 2D6 (2F9Q.pdb)

**Figure 20** Comparison of P450 active sites **(a)** P450 3A4 (1TQN.pdb) and **(b)** P450 2C9 (1R9O.pdb). Representative active site volumes are depicted by the central shaded area. Heme is beneath the shaded area and cysteins are highlighted in white.

(Rowland *et al.*, 2006), P450 2C9 (1R9O.pdb) (Wester *et al.*, 2004), P4502C8 (1PQ2.pdb) (Schoch *et al.*, 2004), P4502A6 (1Z10.pdb) (Yano *et al.*, 2005), and P450 3A4 (1TQN.pdb) (Yano *et al.*, 2004). These three-dimensional representations provide visual coordinates to corroborate experimental findings, including the site of alkylation providing valuable insight to the relevance of adducts with respect to MBI. For example, the selective alkylation of Cys239 was observed with raloxifene (Baer, Wienkers and Rock, 2007). These findings are supported by the crystal structure which shows Cys239 distal to the heme in the enzyme active site (Fig. 20a). Experimental evidence of covalent adducts with P450 2C9 have illustrated the potential role of Ser365 in trapping the glutathione adducting compound, tienilic acid (Melet *et al.*, 2003) suggesting the absence of solvent-accessible cysteines in P450 2C9 which is confirmed upon analysis of its three-dimensional structure in Fig. 20b. Combining structural information of adducts with crystallography data will undoubtedly improve the speed at which future structural information regarding P450 MBI is obtained.

Furanocoumarins commonly found in natural foods such as celery, parsnips, grapefruit juice, and figs have been shown to undergo bioactivation (Fig. 21). Metabolism of the furan moiety leads to bioactivation and is the penultimate step prior to irreversible inhibition of several P450s (Koenigs and Trager, 1998a, b; von Weymarn *et al.*, 2005). The table in Fig. 21 lists three different P450 enzymes affected by the furanocoumarin, 8-methoxypsoralen (8-MOP). Of particular interest from these data is the $K_I$, or the rate of inactivation, at half maximal inhibitor concentration which appears similar for two separate P450 enzymes, P450 2B1 and P450 2A6. This suggests that the affinity and the forward commitment to irreversible inhibition combine in such a way that both enzymes are affected similarly with respect to the bioactivation and inactivation of 8-MOP. However, the $k_{inact}$, or rate at which maximal inactivation occurs, is six times faster with P450 2B1, suggesting either a more reactive residue toward the activated furan or a more proximal orientation of the reactive metabolite with a key active site nucleophile that promotes rapid inactivation versus release of product in the form of a

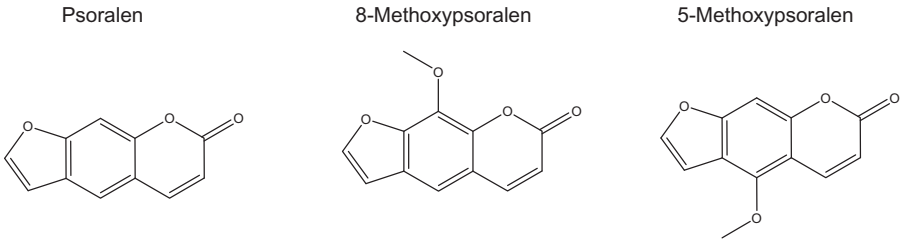


| P450 | $K_I$ (μM) | $K_{inact}$ (/min) | Partition ratio |
|---|---|---|---|
| 1A2 | 1 | NA | NA |
| 2A6 | 1.9 | 2.0 | 11.2 |
| 2A6* | 12 | 0.3 | 25.5 |
| 2A13 | 0.1 | NA | NA |
| 2B1 | 2.9 | 0.34 | 1.3 |
| 2B1* 5mop | 33 | 0.13 | 14.9 |

**Figure 21** The structures of psoralen, 8-methoxypsoralen, and 5-methoxypsoralen are illustrated with the inactivation rates for 8-methoxypsoralen listed for several different P450 enzymes. $K_I$ values were measured to determine the inactivation at half the concentration required to achieve maximal inactivation and $k_{inact}$, or the maximal inactivation, was also determined. The partition ratio was included when measured. The inactivation parameters were not detected for P450 1A2 (Tantcheva-Poor *et al.*, 2001) and the inactivation rate constants for P4502A13 could not be determined due to the high-affinity complex between the P4502A13 and 8-methoxypsoralen (von Weymarn *et al.*, 2005). The close analog 5-methoxypsoralen is shown for comparison.

reactive metabolite. This is consistent with the partition ratios measured for the two enzymes, in which P450 2B1 reacts nearly one to one compared with P450 2A6 which undergoes nearly 10 turnover events to every one inactivation. Furthermore, based upon the proposed mechanism of irreversible inhibition, it may be plausible that one enzyme favors one reactive metabolite pathway versus the other. In combination with different regioselectivity between the two systems, the active site nucleophile may also be unique and could contribute to the different inactivation efficiencies.

5-Methoxypsoralen was also studied with P450 2A6 and P450 2B1, and in both cases showed a decrease in the efficiency of inactivation across all kinetic parameters, $K_I$, $k_{inact}$, and partition ratio (Fig. 21). Interestingly, the 5- and 8-psoralen derivatives were not substrates for P450 2B1 and therefore showed no irreversible inhibition. This provides a nice example of modulation of MBI without drastically changing the chemical scaffold (furan still present). Furthermore, the chemical properties were reasonably conserved with a modest 15-Da mass shift and 0.2 difference in CLogP from the methoxy derivatives. Alternatively in P450 2A6, the introduction of the hydroxyl derivatives led to a reduction in irreversible inhibition the loss of MBI in P450 2B1, indicating the likeliness of a specific enzyme–substrate interaction with 8-MOP which does not exist with the hydroxyl derivative. This type of interaction may prevent bioactivation of the furan and eliminates the potential for MBI in P450 2B1 (Koenigs and Trager, 1998b).

As highlighted with 8-MOP, the chemical structure provides the first indication for the potential of P450 bioactivation and thus the possibility for P450 MBI to occur (Fig. 10 and 21). In many instances, these structures are easily spotted (Figs 9–14). However, because of the robust oxidative chemistry associated with P450 reactions, many novel reaction schemes continue to be discovered (Isin and Guengerich, 2007). For example, the small-molecule ligand for gamma subtype peroxisome proliferator acceptor receptor for the regulation of glucose levels in type II diabetes, troglitazone, was withdrawn from the market after reports of liver failure (Watkins and Whitcomb, 1998; Graham *et al.*, 2003). While the mechanism of hepatoxicity has yet to be clearly defined, significant efforts have focused on the metabolism of troglitazone (Masubuchi, 2006). The structure contains a thiazolidinedione ring which undergoes oxidation at the sulfur, resulting in a ring opening of the thiazolidinedione ring which is then susceptible to nucleophilic attack (Fig. 22) (Kassahun *et al.*, 2001). In addition, the chromane ring structure also forms a GSH adduct after a P450-mediated hydrogen atom abstraction of the phenolic hydroxyl group to produce the ortho-quinone methide which is reactive toward GSH (Fig. 22) (Yamazaki *et al.*, 1999; Kassahun *et al.*, 2001). While it is not believed that the troglitazone adducts function independently to cause the observed hepatoxicity, they may present a contributing factor.

Furafylline serves as an interesting example where the furan structure readily stands out as a potential source of bioactivation and causative ingredient for MBI. In fact it has been highlighted as the site responsible for MBI in a review of P450 bioactivation (Fontana, Dansette and Poli, 2005). However, investigation into the mechanism of P450 1A2 inhibition by furafylline strongly supports the fact that inactivation occurs as a result of oxidative attack at the 8-methyl position of furafylline. A twofold reduction in activation results from the selective incorporation of deuterium replacing the C8 protons supporting the formation of a methide as the precursor to enzyme inactivation (Fig. 9) (Kunze and Trager, 1993). In addition, the use of a furafylline analog where the furan is replaced by a cyclohexyl derivative still yields P450 1A2 inactivation (Racha, Rettie and Kunze, 1998). Similar intermediates have been implicated as the precursor to P450 inactivation from 3-methylindole and 1-[(2-ethyl-4-methyl-1H-imidazol-5-yl)methyl]-4-[4-(trifluoromethyl)-2-pyridinyl]piperazine (Skiles and Yost, 1996; Hutzler *et al.*, 2004).

Aristolochic acid is a known human nephrotoxicant in humans (Bieler *et al.*, 1997). Similar to the activation of furafylline by P450 1A2, aristolochic acid bioactivation also occurs with P450 1A2 but does so in a reductive manner to produce the reactive nitreunium species. The chemical structure possesses a methylene dioxy derivative, known to produce metabolic intermediate complexes. The reduction of the 6-nitro group produces a cyclic nitreunium ion as determined from DNA adducts and appears to be the predominant bioactivation pathway (Fig. 23) (Chae *et al.*, 1999).

The ability to misjudge the safety of a chemical series has driven the development of multiple strategies to facilitate early indicators of MBI. In discovery, the focus is on providing filter points for the high volume of compounds that are screened in early-stage projects; thus, throughput is an important factor and will be a consistent theme within the early screening steps 1, 2, and 3 found in Fig. 16.

The most widely accepted surrogate marker for MBI is the presence of glutathionyl adducts which are generated from hepatic tissue fractions supplemented with the NADPH cofactor plus the addition of glutathione (GSH) (exogenous nucleophile), in order to mimic *in vivo* metabolism with the ability to trap any reactive metabolites during the incubation. In addition to GSH, several additional surrogate nucleophiles can be used and are depicted in Fig. 24. Most commonly, GSH is co-incubated with

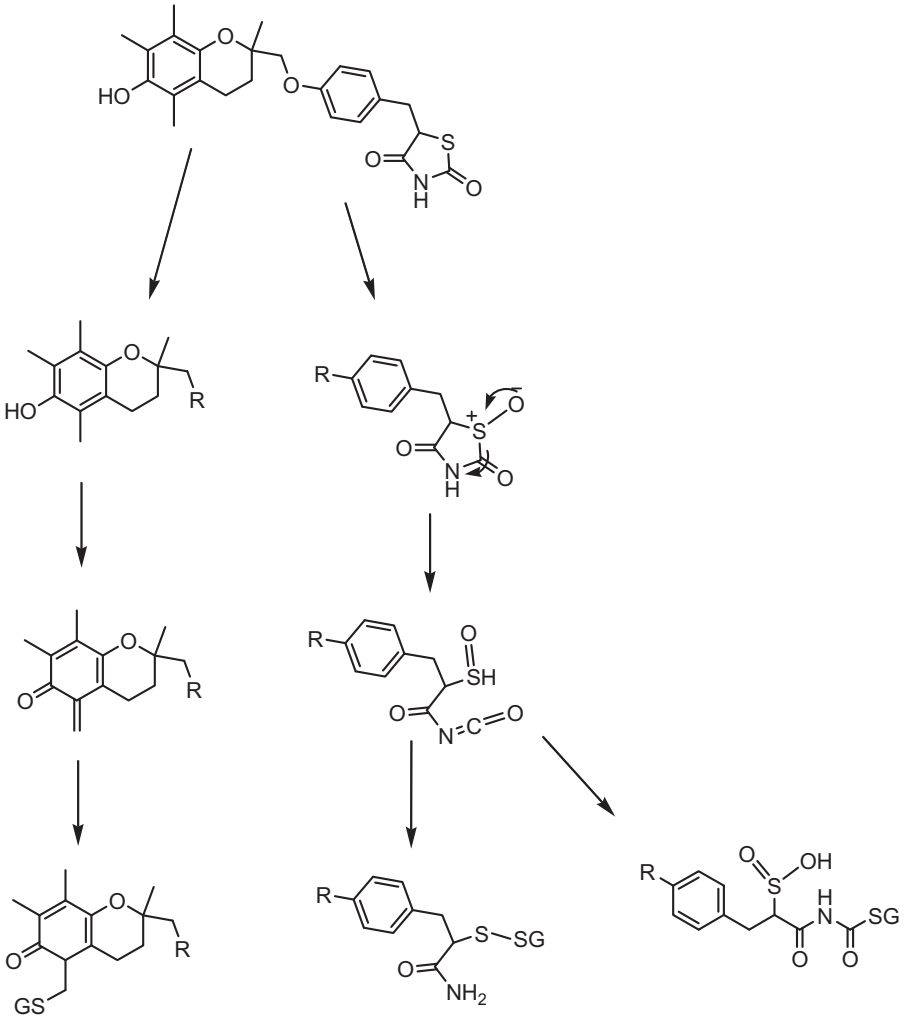


**Figure 22** Mechanistic scheme for P450 enzyme-mediated bioactivation of troglitazone to an *O*-quinone methide and isothiocyanate generated from the ring opening of thializaodine ring (Kassahun *et al.*, 2001; He *et al.*, 2004; Alvarez-Sanchez *et al.*, 2006).

drug and hepatic tissue upon which, drug that undergoes bioactivation to a reactive intermediate can react with the nucleophilic sulfur of cysteine in glutathione. The assay is amenable to liquid chromatography–mass spectrometry/mass spectrometry (LC-MS/MS) and therefore can be designed for high throughput. For example, one of the first high-throughput versions of GSH screening took advantage of the neutral loss of pyroglutamate (129 Da) upon collision-induced dissociation (CID) fragmentation of the GSH conjugate (Baillie and Davis, 1993). This facilitates the identification of GSH-related material without prior knowledge of the parent molecular weight. Unfortunately, not all GSH conjugates produce this fragment ion upon CID which readily confounds attempts to identify GSH conjugates based upon a single LC-MS

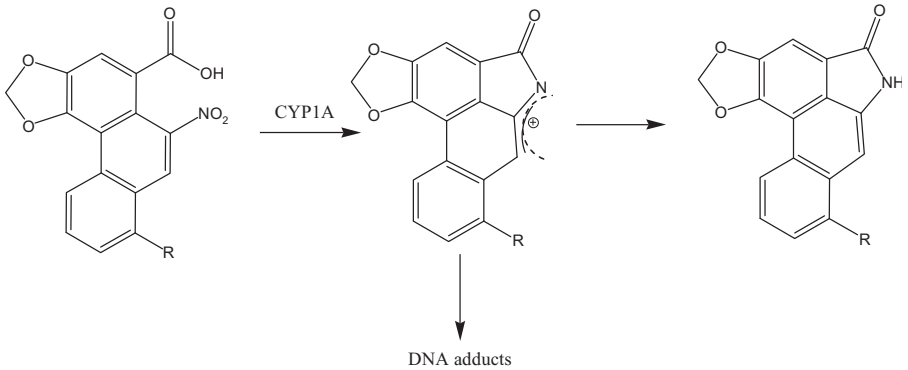


**Figure 23** Mechanistic scheme for P450 enzyme-mediated bioactivation of aristolochic acid (Stiborova *et al.*, 2001). The reductive mechanism is consistent with the previous reports on the P450-mediated reduction of nitropyrenes (Chae *et al.*, 1999).

| Trapping agent | Physical properties | Scan type | Diagnostic fragment | Limitation (s) |
|---|---|---|---|---|
| Glutathione (GSH) 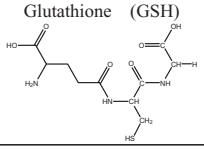 | 307 Da<br>pKa<br>ClogP | Neutral loss (NL)<br>Precursor (PC) | 129 Da<br>m/z -272 Da | Low sensitivity<br>High background interference<br>Irreproducible fragmentation |
| *N*-acetyl lysine (NAL) 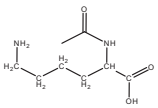 | 307 Da<br>pKa<br>ClogP | | | Reacts with hard electrophiles<br>Reactivity may not represent that of endogenous proteins |
| *N*-acetyl cysteine (NAC) 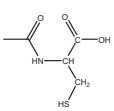 | 307 Da<br>pKa<br>ClogP | NL | 129 Da | Reacts with soft electrophiles<br>Reactivity may not represent that of endogenous proteins |
| Cyanide (CN) | 307 Da<br>pKa<br>ClogP | NL | 27 Da | Reacts with hard electrophiles<br>Reverisible binding |

**Figure 24** GSH and other nucleophiles used in trapping electrophilic intermediates from *in vitro* incubations.

scan (Fig. 24). Instead, different chemically generated conjugates such as aliphatic and benzyl thioether conjugates are known to undergo elimination of 307 Da and produce m/z 308 product ions consistent with the $GSH_2^+$ (Grillo *et al.*, 2003). Alternatively, thioesters can readily lose glutamic acid producing a fragment less than 147 Da from the parent protonated species (Baillie and Davis, 1993; Grillo *et al.*, 2003). The use of negative ion precursor has proven to be less sensitive to the chemical structure of the conjugate itself and serves as a robust method for producing GSH-related

fragments (Dieckhaus *et al.*, 2005). As a result of the GSH dominant fragment patterns generated from the negative ion scan, the method serves as a good initial survey scan for the presence of GSH conjugated drug material with a low incidence of false negatives. Under these conditions, the CID fragmentation generally produces fragments related only to the GSH portion of the adduct and thus, additional MS/MS experiments are typically needed to explore the chemical nature of adduct to support rational chemical design strategies based upon structural information. The recent advent of the linear ion trap mass spectrometer continues to provide additional scanning functionality including enhanced product ion (EPI) scans which have been employed to maximize the information obtained during LC-MS runs used to screen for glutathione conjugates (Zheng *et al.*, 2007).

An important trait to bear in mind with respect to GSH conjugates is not all reactive intermediates are susceptible to attack from the soft electrophile as in the case of GSH. For example, compounds metabolized to hard electrophilic intermediates like iminium ions more readily react with hard nucleophiles like cyanide (Fig. 10) (Nguyen, Gruenke and Castagnoli, 1976). These electrophiles are capable of forming DNA adducts which are a known pathway for carcinogenicity (Stark, Harris and Juchau, 1989). Therefore, the use of additional nucleophiles with alternate reactive properties may be required. Cyanide (CN) or *N*-acetyl lysine represent two nucleophilic surrogates whose reactivity should be considered in order to rule out reactive metabolite formation (Fig. 24) (Argoti *et al.*, 2005; Druckova and Marnett, 2006). Furthermore, screening for cyanide adducts can screen relatively high throughput similar to those efforts described for GSH by monitoring the neutral loss of 27 Da (Argoti *et al.*, 2005).

A major drawback to screening with surrogate electrophiles lies in the assumption that the isolated conjugates relate to P450 inactivation (Baer, Wienkers and Rock, 2007). For example, the formation of multiple glutathione adducts confounds clear interpretation of the adduct responsible for P450 inactivation. Furthermore, many times glutathione offers no protection from MBI, thus increasing the uncertainty that the observed glutathione adducts represent the structural species leading to P450 inactivation. For example, the presence of glutathione in an incubation with P450 2C9 and tienilic acid reduces the partition ratio and stoichiometry of P450 2C9 inactivation, but the adduct itself may not represent the reactive species ultimately responsible for P450 2C9 inactivation (Koenigs *et al.*, 1999). Despite this discrepancy, it has become apparent with the quantification of GSH adducts that the amount of conjugated drug appears highly correlated to covalent binding results in many cases (Masubuchi, Makino and Murayama, 2007). Overall, screening with surrogate electrophiles provides an early complement to covalent binding studies where radiolabeled compound is required (Evans *et al.*, 2004). In a subsequent section of this chapter, the use of mass spectrometry of intact protein is described to elucidate drug-based adducts as a complement to GSH screening procedures.

The third step in the screening progression for MBI (Fig. 16) is to investigate the potential impact of the bioactivation pathway on P450 kinetics. In many cases, a negative finding from this assay provides a decision point at which MBI is no longer considered a safety concern for a particular compound. In some cases, GSH screening is omitted in favor of rapid kinetic screens focused on elucidating the potential for MBI. This choice of tactic may be especially poignant for P450 inhibitors that have multiple modes or mixed modes of inhibition such as 17α-ethynylestradiol (Lin and Hollenberg, 2007). This is feasible partly due to advances in laboratory automation (Lim *et al.*, 2005;

DeWitte and Robins, 2006), which makes the decision to move to a more informative kinetic assay (and in some cases, may replace the GSH LC-MS/MS approach outright) a cost-effective decision. The choice may reflect the stage of a project and if a team is still focused on improving absorption, distribution, metabolism, and excretion (ADME) properties with structure–activity relationships (SARs), then the use of GSH screen should provide mechanistic details necessary to eliminate drug bioactivation. However, in most cases, the inclusion of both screens is recommended as it heightens the awareness of the project team to the potential for MBI, showcases the susceptibility of P450s to generate reactive chemistries leading to complex mechanisms of P450 inhibition, and highlights any potential for other adverse drug reactions.

A straightforward method to assess the impact of bioactivation is performed with a pre-incubation step of the inhibitor in the presence of NADPH. In theory, the pre-incubation step provides a straightforward method to identify if pre-incubating alters the inhibitor through metabolism to increase the potency of inhibition of the compound of interest (Obach, Walsky and Venkatakrishnan, 2007). Two assay formats utilize a pre-incubation to assess the time and concentration effect of the inhibitor on P450 inhibition: (i) the $IC_{50}$ shift assay and (ii) the apparent partition ratio screen (APR) (Fig. 25). Both formats rely upon a pre-incubation step which provides the necessary metabolism to assess whether or not inhibitor potencies increase upon metabolism (Obach, Walsky and Venkatakrishnan, 2007). A schematic of the experimental setup is illustrated in Fig. 26. Based upon the subtle difference in setup, it is clear that the left shift from a pre-incubation in the $IC_{50}$ shift experiment does *not* distinguish the potential contribution of a reversible inhibitor that can occur from a metabolite formed during the initial incubation step.

To minimize the potential for inhibitory metabolites generated during the pre-incubation from contributing to the change in observed inhibition, large dilutions from the primary incubation are utilized to reduce reversible inhibition from inhibitory metabolites (minimum of one-tenth). Moreover, success of this experiment is predicated on the ability of the incubation system (inhibitor concentration, time, and enzyme source) to metabolize the compound. Therefore, an important control is to measure

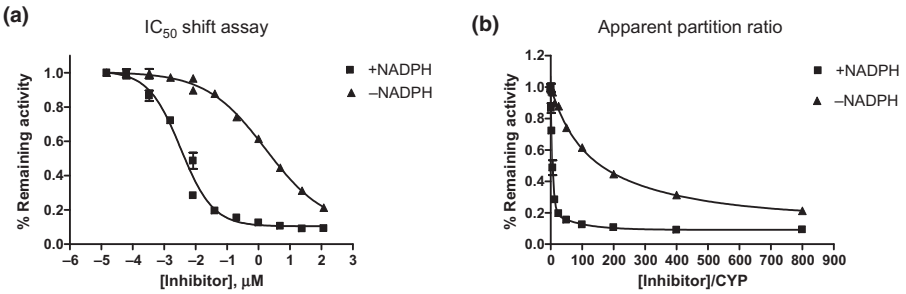


**Figure 25** Data transformed from the experimental design in Figure 26 can be plotted one of two ways depending upon the choice of control. For the $IC_{50}$ shift experiment **(a)**, the control consists of the inhibitor without NADPH. This allows the formation of $IC_{50}$ that can be compared with the pre-incubation with NADPH. The diagnostic sign of TDI is related to a left shift in the $IC_{50}$ or increase in potency when the inhibitor is pre-incubated with NADPH. Alternatively, the apparent partition ratio **(b)** control utilizes NADPH but no inhibitor. The only other difference is the data on the abscissa is inhibitor concentration in the case of the $IC_{50}$ shift compared with inhibitor concentration versus enzyme concentration in the apparent partition ratio experiment.

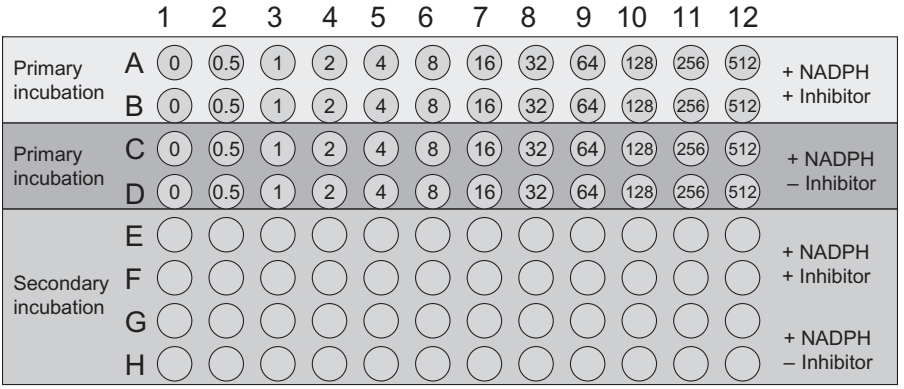


**Figure 26** 96-well plate pre-incubation format for determination of $IC_{50}$ shift or apparent partition ratio. Primary incubations with varying concentrations of suspected MBI are conducted for a finite time period with excess human liver microsomes (0.5 or 1 mg/mL) depending upon probe substrate reaction velocities and dilution scheme of the secondary incubation. The secondary incubations contain reaction buffer plus NADPH and the appropriate probes substrate for the P450 under investigation at four to five times the $K_m$ and are initiated with one-tenth to one-hundredth of the primary incubation volume.

the percent turnover of the inhibitor during the pre-incubation step. This provides confidence that the pre-incubation conditions are sufficient to generate the turnover required to properly investigate the potential for MBI. In the absence of sufficient turnover, it may be necessary to increase the pre-incubation time, change protein concentration, tissue fraction (S-9, microsomes, or hepatocytes), or any combination thereof.

One difference exists in the graphical representation of the data between the $IC_{50}$ shift and the APR formats which provides a possible advantage to representing data as an APR. The x-axis of the $IC_{50}$ shift is represented as the log concentration of inhibitor tested in the experiment whereas for the APR, the x-axis is represented in linear fashion by concentration of the inhibitor relative to P450 levels used in the incubation (Fig. 25). The net effect of this subtle experimental difference is that it provides a clear distinction from the control incubation especially as the concentrations of inhibitor increase, since the altered trajectory of NADPH supported pre-incubation in both the $IC_{50}$ shift and the APR reactions are highly dependent upon inhibitor concentration ($k_{inact}$) for the manifestation of MBI. Furthermore, representing the data as APR provides a direct quantitative measure between different inhibitors by providing an efficiency value for inactivation which is useful for SAR purposes.

Up to this point in Figure 16, the focus has been on Tier 1 or screening-based assays which would often be implemented in the early discovery stage of a project. If the results from these assays point to MBI, then the next step is to define the MBI potency. That is, if the inactivation rate is 1% of the natural degradation rate of the P450, then inactivation will not likely increase the drug interaction beyond that described by reversible inhibition. Therefore, investing in the technical and laborious structural investigations required to define the mechanism(s) behind MBI is likely not warranted.

In order to contextualize the significance of MBI, one must define properties of inactivation including the maximal possible rate of enzyme inactivation at a saturating

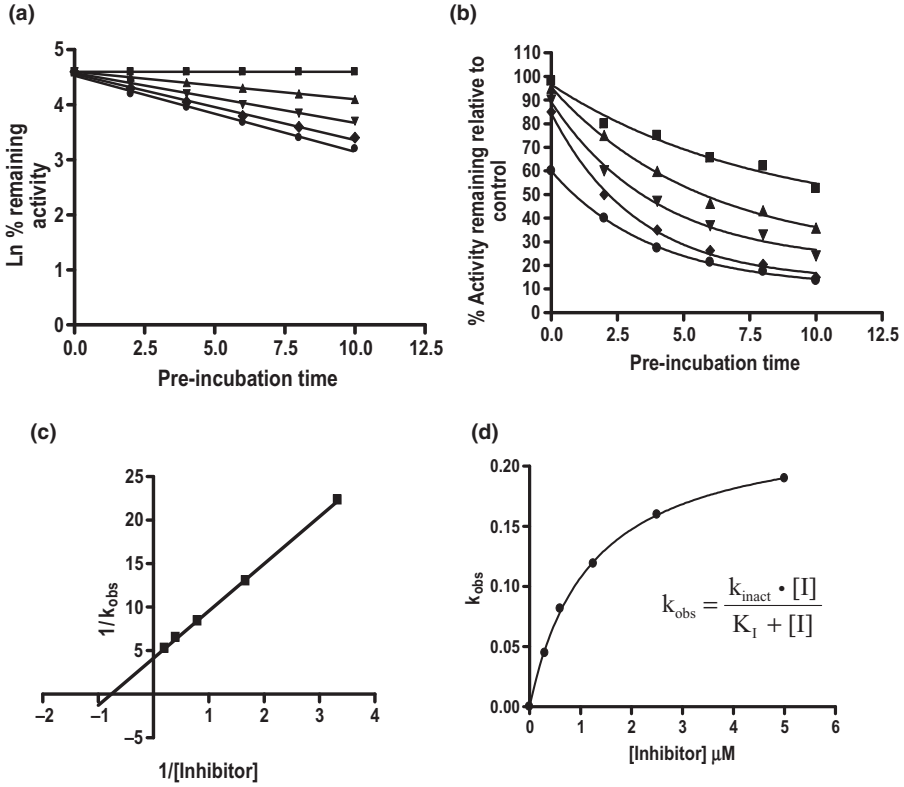


**Figure 27** Example of kinetic plots used for visualizing time-dependent inhibition data. **(a)** Log-linear plot representing the remaining enzyme activity after various pre-incubation times with an MBI, **(b)** linear-linear plot representing the remaining enzyme activity after various pre-incubation times with an MBI, **(c)** linear plot (Kitz Wilson) represented by $1/k_{obs}$ and 1/[inhibitor], and **(d)** non-linear regression of $k_{obs}$ versus inhibitor concentration.

amount of MBI, $k_{inact}$, and the concentration at which half of the maximal inactivation is achieved which is defined as the $K_I$. In doing so, the rate of inactivation can be directly compared with what is anticipated as the natural rate of P450 degradation. This assay commonly referred to as the time-dependent inhibition (TDI) assay includes multiple pre-incubation times and various concentrations of the inhibitor to yield individual inactivation rates for each concentration, which are combined and fit using nonlinear regression to determine the $K_I$ and $k_{inact}$ (Fig. 27) (Mayhew, Jones and Hall, 2000).

The assay requires two steps, starting with a primary incubation to activate the inhibitor (suspected MBI). From the primary incubation, aliquots are transferred into a secondary incubation (10- to 100-fold) containing a marker substrate selective for the particular P450 isoform under investigation (Fig. 28). Several variables must be considered to ensure accurate rates of inactivation are measured: (i) that the competitive inhibition at the secondary stage of the incubation is minimized by sufficiently diluting the primary to secondary incubations; (ii) an appropriate number of inhibitor concentrations are used; (iii) protein concentration is optimized to minimize nonspecific binding yet enough to produce sufficient turnover in the primary and secondary

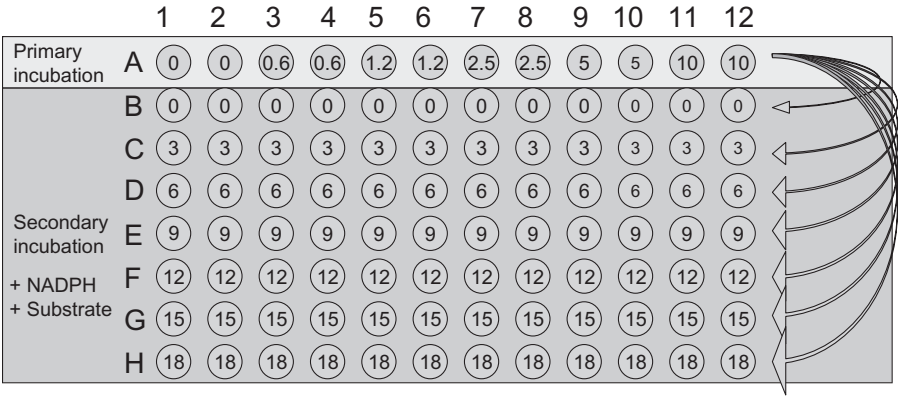


**Figure 28** 96-well plate pre-incubation format for determination of time-dependent inhibition. Primary incubations containing varying concentrations of suspected MBI are conducted in 200 μL final volume with excess human liver microsomes (0.5 or 1 mg/mL) depending upon probe substrate reaction velocities and dilution scheme of the secondary incubation. The secondary incubations contain reaction buffer plus NADPH and the appropriate probes substrate for the P450 under investigation at ~5× the $K_m$ and are initiated with one-tenth to one-hundredth of the primary incubation volume. Transfers are made in time intervals consistent with the expected potency of the TDI and are made easily with 12-channel pipette or automation. Each subsequent row corresponds to unique time interval for each concentration under investigation.

incubations; and (iv) the secondary incubations are carried out under linear conditions with saturating amounts ($V_{max}$ conditions) of the appropriate probe substrates. Careful attention to these design features helps reduce bias in the assay format and provides the necessary end points to accurately quantify the effect of time and metabolism on the inhibitor in question (Yang *et al.*, 2005).

To ascertain the appropriateness of the experimental data, the kinetic results should be examined with several different plots to determine if further kinetic analysis is required. First, each inhibitor concentration and its time course should be plotted as the percent remaining relative to control on the ordinate axis versus the pre-incubation time on the abscissa. These data can be presented in a semilog plot or if the changes are small with a linear plot to facilitate visualization of any aberrant data points (Fig. 27a,b). Furthermore, this plot provides a semiquantitative view to the extent of competitive inhibition observed in the secondary incubations. For example, the highest concentration inhibitor sample in Fig. 27b indicates substantial competitive inhibition occurred at the highest inhibitor concentration. This is diagnosed by the first time point (no pre-incubation) at which the remaining activity should be close to 100%, but in these examples, is only 60% of control. Fitting inactivation data in the presence of significant competitive inhibition will reduce slope of the fit and lead to underestimation of the inactivation rate, confounding the accurate determination of MBI rates. Next, the data should be transformed to a linear plot (Kitz–Wilson) by plotting $1/k_{obs}$ on the ordinate versus 1/inhibitor on the abscissa (Kitz and Wilson, 1962) (Fig. 27c). The intercept of the ordinate axis defines $1/k_{inact}$ which can be used to approximate $k_{inact}$. The slope of the line estimates the $k_{inact}/K_I$. The range in data should be such that a single point does not dominate the linear regression. Secondly, any deviation from linearity may indicate limitations of inhibitor solubility or a mixed mechanism of inhibition. The $K_I$ and $k_{inact}$ values should be derived from nonlinear regression (Fig. 27d) (Maurer and Fung, 2000). At the end of a TDI experiment, there are at least four graphs

that can be used to interrogate the integrity of the data from the TDI experiment. Once this has been completed, the significance of MBI in addition to reversible inhibition can be determined.

Once MBI has been estimated to lead to a significant interaction, it is paramount to structurally characterize the mechanism(s) responsible for irreversible inhibition to define and potentially eliminate the responsible chemical substructure. Three distinct mechanisms of MBI exist for P450 inactivation: (i) heme iron coordination through an MI complex, (ii) heme alkylation, and (iii) apoprotein alkylation. Prior to initiating studies aimed at defining the mechanism of inactivation, it is useful to try and define which P450s are affected by MBI through phenotyping efforts. Phenotyping is presented in detail in Chapter An Introduction to Metabolic Reaction Phenotyping. Defining the metabolic pathways helps define what P450s need to be studied and the experimental tools available to study the structural aspects of MBI. Specifically, it enables the use of recombinant P450 enzymes and while these systems have been shown to produce unique kinetics properties in certain situations, their simplified matrix enables the use of a wider range of applications aiding in the structural characterization and mechanistic details regarding the nature of P450 enzyme inactivation.

Heme iron coordination is a result of P450-mediated metabolite which subsequently coordinates as the sixth axial ligand to the heme iron. This is commonly referred to as MIC formation or MI complex and leads to P450 inactivation. The formation and stability of the MIC is directly related to the chemical substructure of the metabolite and is commonly referred to as a quasi-irreversible inhibitor. Two substructures are prone to MIC formation: the tertiary amine and the methylene dioxy species. The most useful diagnostic for MIC formation is performed by measuring the increase in absorbance at 455 nm over time with a UV-Vis spectrophotometer (Franklin, 1995). In these studies, the enzyme, substrate, and NADPH are combined into a cuvette and pre-incubated at 37 °C. The absorbance at 455 nm is measured kinetically to determine the presence of MIC formation (Fig. 29) (Hutzler *et al.*, 2006). The confirmation of MIC formation as the sole contributor to P450 inactivation can be assessed by the ability to reverse the inhibition upon the addition of potassium ferricyanide (Sharma, Roberts and Hollenberg, 1996).

Direct elucidation of the chemical species responsible for MIC formation represents the most challenging out of the three forms of inactivation because the complex is not physically isolable. Instead, chemistry support is required to develop structural analogs to indirectly show the lack of MIC in the absence of the proposed culprit substructure. For example, for a methylene dioxy to form an MIC, initial hydrogen atom abstraction must occur which can be blocked by introducing a $CF_2$ at the bridge of the two oxygen atoms, or instead the methyl group can be replaced with an ethyl derivative eliminating the ability of the carbene to form after hydrogen atom abstraction (Fig. 13). Alternatively, if the tertiary amine is the precursor to MIC formation, a fluorinated derivative or alpha substitution can be used to deter *N*-dealkylation at the site of MIC formation. If the substructure causing the MIC is also required for therapeutic activity, a more subtle change in structure may need to be explored. For example, with the aid of P450 crystal structures, docking tools may be used to infer enzyme substrate interactions that contribute to the particular substrate regioselectivity. Specifically targeting these interactions may provide a different P450 regioselectivity that reduces MIC formation while maintaining the biological potency.

Heme alkylation is the second mechanism responsible for P450 inactivation and occurs via the bioactivation of a chemical species, which in turn alkylates the heme substructure itself and irreversibly inactivates the affected P450. Substructures prone

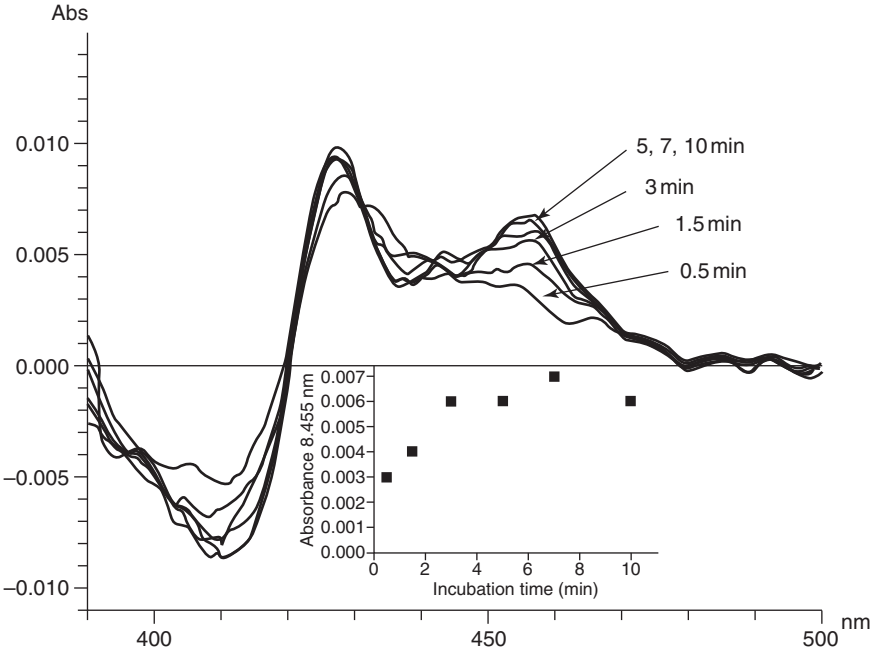


**Figure 29** Kinetic formation of MIC by monitoring absorbance at 455 nm (Franklin, 1995).

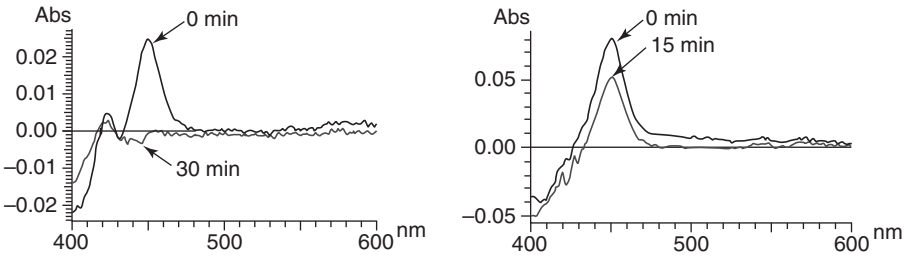


**Figure 30** Measurement of carbon monoxide binding pre- and post-incubation with MBI to qualitatively determine the extent of heme-related adduct formation (Hutzler *et al.*, 2004).

to heme adduct formation include, but are not limited to, furans, aldehydes, and alkynes. Again, the UV-Vis spectrophotometer has been classically used to differentiate a heme adduct from an apoprotein adduct. The saturation of CO in a reduced solution of P450 affords a characteristic absorbance at 450 nm which is used to determine P450 enzyme levels from a complex solution (Omura and Sato, 1964). Heme alkylation prevents the binding of CO to the heme iron and therefore has been used to measure the P450 concentration pre- and post-incubation, whereby the loss in CO binding at the end of incubations represents the amount of alkylated heme. However, experimental data have shown the disappearance of CO binding may not always be linked to heme alkylation (Fig. 30). For example, CO spectral binding revealed ~60% loss of CO binding upon incubation with an MBI tienilic acid, which has been shown to form a P450 apoprotein adduct, indicating the decreased CO binding can occur through alternate

mechanism(s) (Koenigs *et al.*, 1999). In addition, raloxifene has been shown to exclusively alkylate the apoprotein of P450 3A4 and yet after incubation, significant CO binding is lost (Baer, Wienkers and Rock, 2007). Finally, the photoaffinity probe, lapachenole, of which alkylates proteins via Michael addition to selectively label cysteine residues also produced a significant loss of CO binding (Wen *et al.*, 2005). Therefore, CO binding should not be used as the sole indicator of heme alkylation.

Heme alkylated products can also be measured directly using LC-UV-MS. For example, 1-ABT forms a covalent adduct with the heme of P450. The 1-ABT heme adduct was measured directly using mass spectrometry (Ortiz de Montellano and Mathews, 1981). This, combined with NMR data, was used to give the exact details of the adduct and its structure, which provided a basis for the mechanism for 1-ABT heme alkylation (Fig. 31). The advantage of this technique is that information regarding inactivation based on heme adduct formation is determined directly and therefore, the structure can be characterized affording mechanistic hypothesis for its formation leading to the ability to modify the chemistry responsible for inactivation. An alternative method that does not provide mechanistic detail but is not confounded by

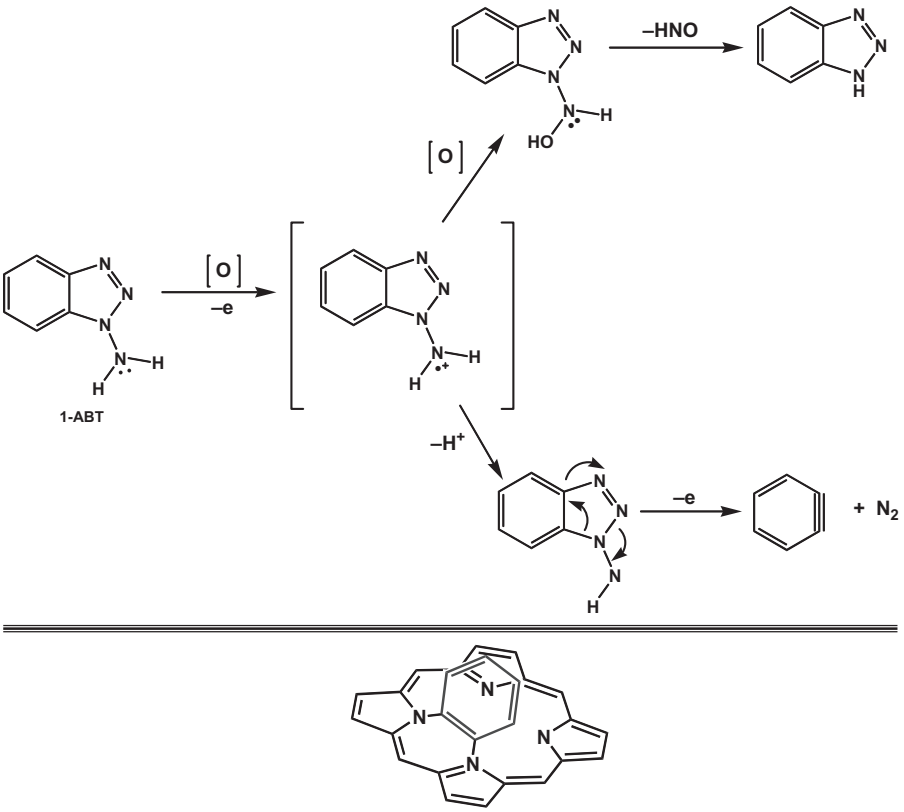


**Figure 31** Mechanistic scheme for P450 enzyme-mediated bioactivation and subsequent enzyme inactivation of 1-aminobenzotriazole (ABT). Structure of ABT rearranges to benzyne upon metabolism and alkylated the heme. Other alkynes have shown similar reactivity toward the heme.

apoprotein adduct formation as the CO measurement is the direct quantification of heme itself pre- and post-incubation. This requires the use of recombinant expressed P450 enzymes without cytochrome $b_5$ so as to measure the disappearance of heme directly in order to implicate heme alkylation or destruction as a mechanism for inactivation.

Apoprotein adducts represent the third mechanism of inactivation and have classically been identified based upon the use of radiolabeled ligand in combination with gel electrophoresis, where P450 binding was implicated through the migration of radioactivity near 55 kDa. More recently, mass spectrometry has been used to characterize adducts of intact proteins which favorably combines the level of detail accustomed to mass spectral data, in addition to providing information regarding covalent binding stoichiometry and confirming the adduct is irreversible in nature. Figure 32 illustrates the use of intact protein mass spectrometry to identify the raloxifene P450 3A4 adduct producing mass shift of 472 Da compared with control P450 3A4 protein without

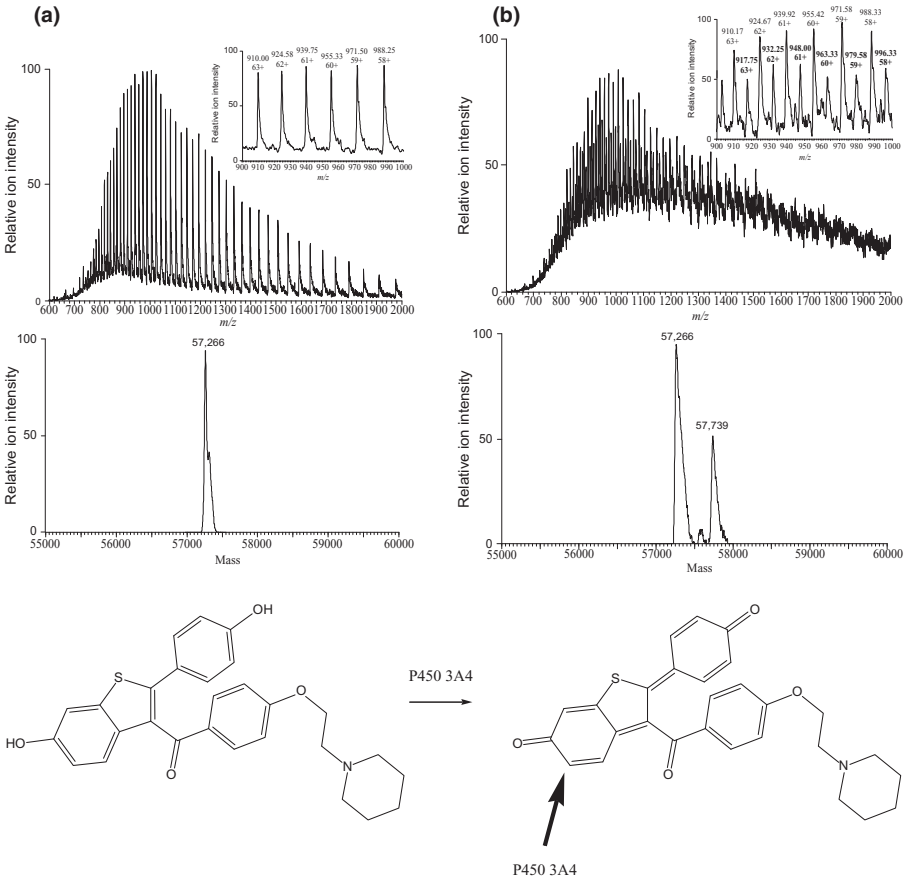


**Figure 32** LC-MS spectra and deconvoluted spectra for P450 3A4 apoprotein. Spectra are shown for unmodified P450 3A4 **(a)**; P450 3A4 following the reconstitution and incubation with raloxifene **(b)** Raloxifene biotransformation to the reactive diquinone methide is shown highlighting one of the nucleophilic sites generated (Baer, Wienkers and Rock, 2007).

raloxifene present. These findings confer the diquinone methide is responsible for P450 3A4 inactivation as opposed to the ortho-quinone reactive metabolite which would have had a mass increase of 488 Da (Baer, Wienkers and Rock, 2007). Presumably, this information could be used to generate raloxifene analogs incapable of forming the diquinone methide preempting MBI formation in a similar fashion to what has been achieved with the close analog, arzoxifene (Liu *et al.*, 2005).

Another interesting example is provided from 1-[(2-ethyl-4-methyl-1H-imidazol-5-yl)methyl]-4-[4-(trifluoromethyl)-2-pyridinyl]piperazine (EMTPP), which was shown to form an apoprotein adduct to P450 2D6 (Fig. 33). The mass difference of the P450 adduct from the unlabeled P450 was 353 Da which was consistent with the parent compound. Interestingly, metabolism data including NMR and CID mass spectrometry were used to support the mechanism of adduct formation that likely resulted from multiple oxidations to form dehydrated methide of EMTPP similar to what has been proposed for 3-methylindole (Nocerini *et al.*, 1985). This example also nicely illustrates the potential limitation of intact protein mass spectrometry and how it must still be combined with additional experimental data in many cases to support the nature of the inactivating species, based upon the mass spectral resolution limitations and the difficulty in differentiating 2 Da with ~60,000 Da proteins. However, with the continued advances in mass spectrometry instrumentation such as the LTQ-Orbitrap and other hybrid mass spectrometry equipment, the use of this technique is rapidly spreading and may, in the near future, have the resolution to differentiate adducts by as little as 2 Da.

The structural limitations of intact protein mass spectrometry can be greatly improved with subsequent proteolytic digest of the adducted protein. This additional step also has the benefit of potentially providing greater detail with respect to the exact nature of the protein adduct, including the ability to provide the site of protein modification and the potential for CID spectral data to be obtained, which increases the potential structural knowledge about the chemistry behind the irreversible binding. The ability to elucidate covalently bound drug adducts to P450s is laden with technical challenges largely associated with difficulty in isolating the membranous P450s and obtaining sufficient sequence coverage from digests necessary to detail the location and nature of MBI adduct (Lightning *et al.*, 2000). To assist in this process, several techniques have been employed to locate nucleophilic residues including photoaffinity labeling (Gartner *et al.*, 2005; Wen *et al.*, 2006), alkylation with xenobiotic linked to biotin (Liu *et al.*, 2005), and direct alkylation of purified P450 with reactive electrophiles such as iodoacetamide (Baer, Wienkers and Rock, 2007). In combination with the nucleophilicity of active site amino acids, these tools provide a focal point for future efforts when trying to identify drug adducted P450s.

Photoaffinity labels have been used to probe enzyme structure function of cytosolic and transmembrane proteins (Wang, Bauman and Colman, 1998; Tessier *et al.*, 2005). The use of photoaffinity probes in P450 labeling has also been explored. Initial efforts were performed with 1-(4-azidophenyl)imidazole and P450 cam (Swanson and Dus, 1979). More recently, lapachenole was selected to investigate the alkylation of P450 3A4 (Fig. 34a) (Gartner *et al.*, 2005). The photo-activation of lapachenole produced two distinct peptide adducts. Two cysteine residues were identified, Cys98 and Cys468. Cys98 is in close proximity to the B-B'loop region located near the substrate recognition site 1 (SRS-1) (Roussel, Khan and Halpert, 2000). Moreover, the region near Cys98 and the B-C loop is postulated to be the access point to the heme iron and small changes in this region impact active site volumes (Williams *et al.*, 2004). Based

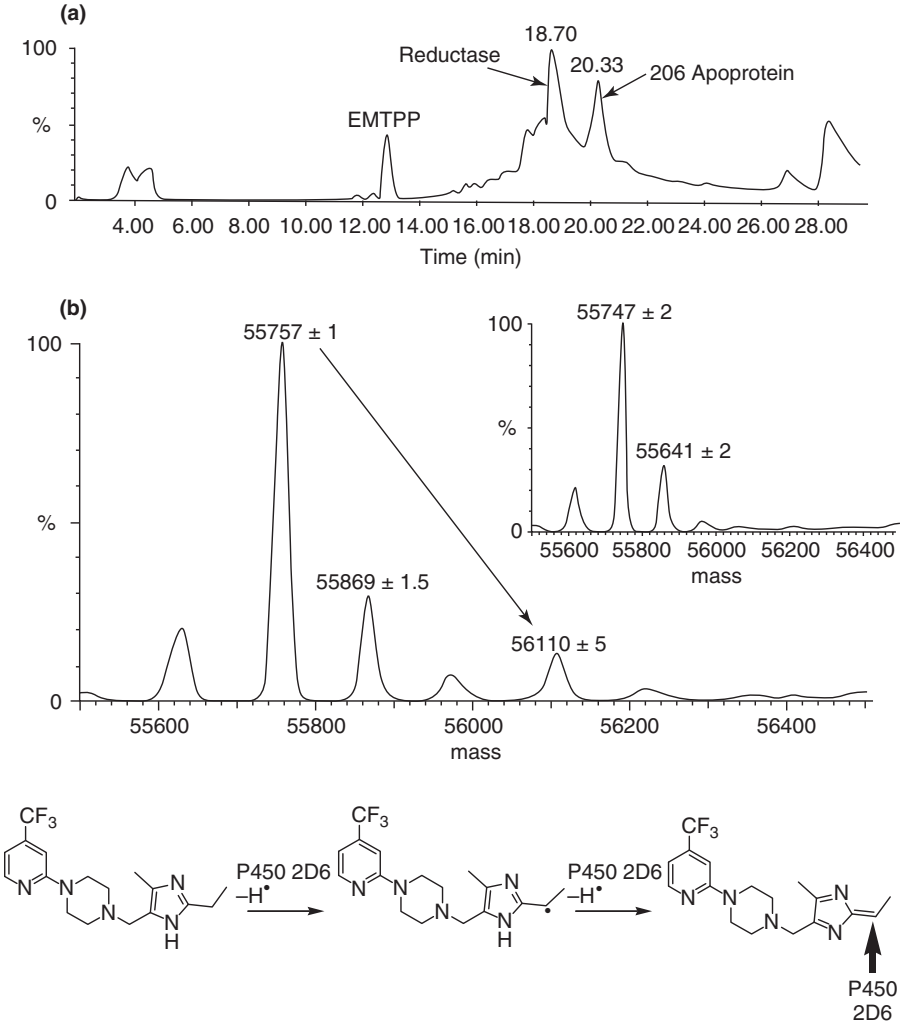


**Figure 33** **(a)** Chromatogram illustrating separation of reductase and apoprotein from an incubation with recombinant P450 2D6. **(b)** Deconvoluted mass spectrum of EMTPP-inactivated P450 2D6 illustrates an EMTPP adduct (56,110 ± 5 Da) consistent with a mass 353 Da. Inset shows the mass of the −NADPH control incubations. Proposed inactivation species of EMTPP by P450 2D6 involving the formation of a radical intermediate that undergoes additional hydrogen atom abstraction from the imidazole nitrogen to form a reactive imidazo-methide-like species (Hutzler *et al.*, 2004).

upon the crystal structure, the second adducted species, Cys468, appears to be a solvent-accessible cysteine on the exterior of the protein.

Alternatively, the use of biotin-linked electrophiles holds promise for use in identifying key enzyme electrophiles. This technique was extended by linking biotin with raloxifene (Fig. 34b) and subsequently used to illustrate binding of the activated raloxifene moiety to Cys47 of GST P1-1 (Liu *et al.*, 2005). When employed against a

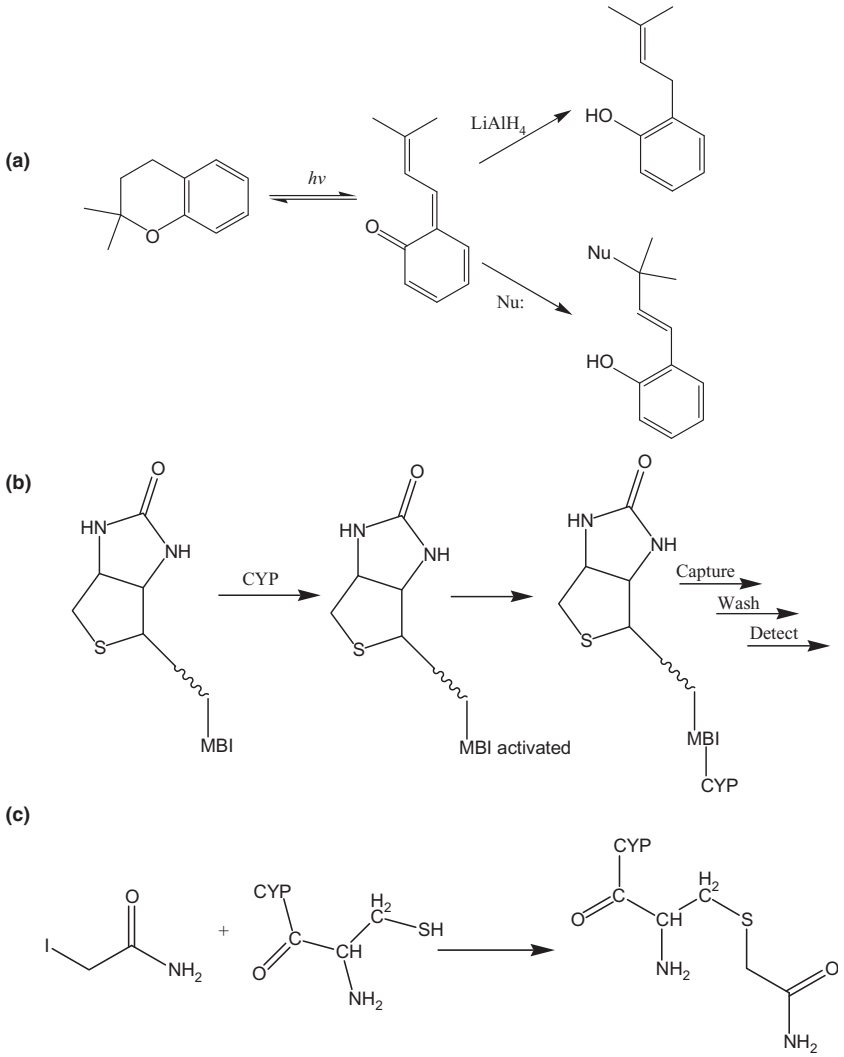


**Figure 34** Different mechanistic approaches to locating P450 nucleophiles including **(a)** photoaffinity labels (Gartner *et al.*, 2005), **(b)** biotin-linked electrophiles (Liu *et al.*, 2005), and **(c)** direct alkylation with reactive electrophile (Baer, Wienkers and Rock, 2007).

dexamethasone-treated preparation of rat liver microsomes, five adducts were identified, none of which corresponded to P450s. The capture and isolation examples provide proof of concept that the technology is capable of isolating nucleophilic amino acids from a complex mixture, and if employed against a simplified *in vitro* system akin to the GST example, key active site residues may be more readily identified from P450s.

Direct alkylation with reactive electrophile iodoacetic acid (IA) and pyrenyliodoacetic acid (PIA) offers a simple approach to locate solvent-accessible and reactive cysteine residues (Fig. 34c). Interestingly, in P450 3A4, the use of IA or PIA produced single stoichiometric reactions when run for a finite time (Baer, Wienkers and Rock, 2007).

Upon digestion with proteinase K, the alkylated residue was identified by LC-MS/MS as Cys239. Upon analysis of the crystal structure of P450 3A4 (1TQN.pdb), Cys239 resides above the B-C loop and may serve as a perfect hook for electrophilic products as they egress the active site of P450 3A4. Interestingly, the label of P450 3A4 by photoaffinity label lapachenole and that of IA produced distinct adducts. It is possible that IA and lapachenole exert structural changes to the P450 exposing unique nucleophiles. Alternatively, the reactivity of Cys239 and Cys98 may chemically favor the two different electrophilic intermediates. Either way, future efforts to identify MBI of P450 3A4 should focus on locating peptides containing cysteine residues 98 and 239.

The use of different digesting reagents has been explored to provide insight to MBI in P450. Cyanogen bromide chemically digests proteins to produce large peptide fragments and has been used to locate the regional location of adduct formation (Lightning *et al.*, 2000). Lysyl endopeptidase also produces rather large fragments of the resultant protein as protein is cleaved at the *C*-terminus of lysine residues. Both of these digestion procedures are very amenable to time of flight (TOF) technology given their large molecular weights and have the advantage of producing fewer fragments of the protein to search. Trypsin digests cleave proteins at the *C*-terminus of lysine and arginine residues. The fragments produced from trypsin are very suitable for LC-MS/MS using a variety of mass spectrometry platforms given their tendency to carry multiple charges ($[M+2H]^{+2}$). Recently, proteinase K has been used to take advantage of the indiscriminant protein cleavage sites to produce small peptide fragments. When separating a proteinase K digested sample on reverse-phase column, the bulk of the small peptides elute quickly, whereas drug adducted peptide creates increased hydrophobic character creating separation from un-adducted peptides (Fig. 35). In addition, the use of proteinase K should improve ionization efficiencies by reduced hydrophobic aggregation and less ion suppression from the reduced number of co-eluting peptides. The tools described for identifying active site nucleophiles provide a focal point for future studies aimed at locating drug-based peptide adducts and over time should lead to improved ability to generate high-quality structural data in a more timely fashion.

Radiolabel is used as a quantitative and qualitative tool to rapidly assess irreversible binding of protein, locate peptides that have radiolabel drug incorporated, and define binding stoichiometry (Fig. 36). An additional advantage of using radiolabel is that *in vivo* and *in vitro* experiments can be used to better understand the reactivity of the bioactivated mechanism (Evans *et al.*, 2004; Baillie, 2006; Williams *et al.*, 2007). Covalent binding results are typically reported as pmol drug bound per milligram of protein. Based upon a meta-analysis, Evans *et al.* issued a perspective highlighting 50 pmol of drug/mg of protein as a reasonable cutoff for continued development of a molecule that possesses covalent binding (Evans *et al.*, 2004). For MBI, the measured quantity of covalent binding may be compared with the relative quantities of P450 enzymes expected from a preparation of human liver microsomes. For example, the amount of P450 in a particular enzyme preparation can be measured (Alterman *et al.*, 2005). The combination of enzyme abundance with covalent binding and MBI may provide additional knowledge regarding the predicted exposure to reactive metabolites. For example, Fig. 37 shows the average P450 abundance data compiled from the Caucasian physiological database from within Simcyp V7.01 (www.simcyp.com, Sheffield, UK). Provided that the bioactivation mechanism produces an MBI with a partition ratio close to 1 and the P450 pathway is known, the quantities of P450 shown in Fig. 37 serve

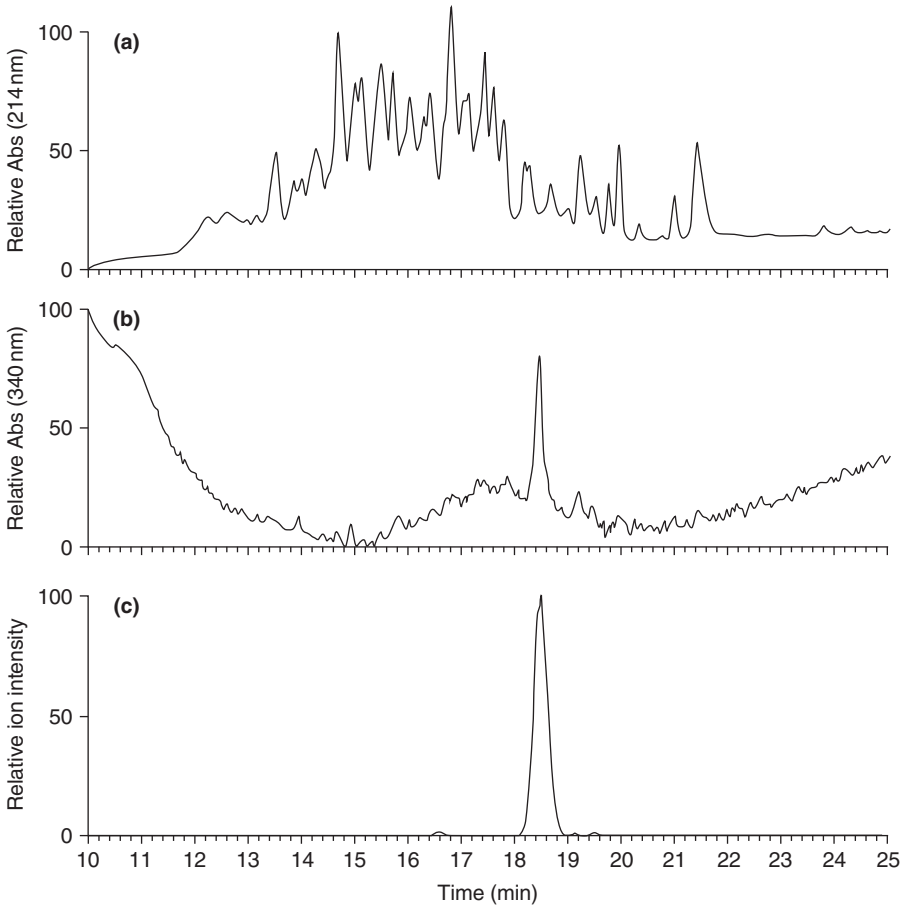


**Figure 35** Chromatographic separation of peptides following proteinase K digestion of raloxifene-adducted P450 3A4: **(a)** UV 214 nm trace; **(b)** UV 340 nm trace; and **(c)** mass spectral trace of adduct (Baer, Wienkers and Rock, 2007).

as a benchmark for the expected amount of radiolabel incorporation into human liver microsome. For example, L-754,394 has a low partition ratio between 1 and 4, is metabolized by P450 3A4, and labeled approximately 120 pmol equivalents per mg of protein when incubated with human liver microsomes (Sahali-Sahly *et al.*, 1996; Lightning *et al.*, 2000). Based upon P450 levels, this is well within the expected range of P450 3A levels from a human liver microsome preparation and is consistent with a low partition ratio where less of the bioactivated drug would be expected to react with extracellular protein.

Alternatively, a series of dihydrobenzoxathiin selective estrogen receptors have been shown to undergo bioactivation via P450 3A4 (Zhang *et al.*, 2005). Two primary analogs were studied to reveal that compound I showed 1106 pmol equivalents per milligram versus compound II which yielded 461 pmol equivalents per milligram of protein (Fig. 38). Based upon the structures in Fig. 38, it is clear that a subtle

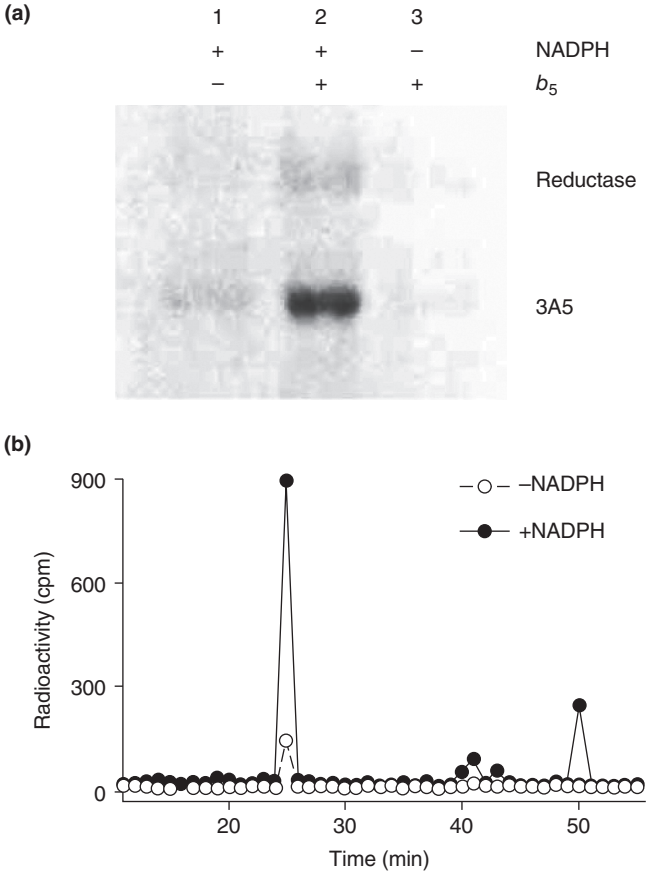


**Figure 36** Covalent binding of EE to P450 3A5. **(a)** SDS-PAGE separation of 3A5 apoprotein and reductase in the reconstituted system incubated with [$^3$H]EE with NADPH but no $b_5$ (Wienkers and Heath, 2005), with NADPH and $b_5$ (Guengerich, 1997), and with $b_5$ but no NADPH (Friedman *et al.*, 1999). **(b)** HPLC separation and analysis by liquid scintillation counting of the fractions after the [$^3$H]EE-inactivated P450 3A5 was digested by Lys C (Lin and Hollenberg, 2007).

difference can have a dramatic effect on the level of covalent binding. More importantly, this example shows despite being P450 3A4-mediated, significant higher levels of protein adduction are observed from covalent binding studies. Unfortunately, these investigations did not include a partition ratio but based upon the covalent binding and P450 3A4 levels, the ratio may be expected to be estimated >10. This example illustrates that the disproportionate covalent binding levels relative to P450 levels should indicate an increase in extracellular microsomal protein and should be reflected by large partition ratios. Of course, heme-based adduct or small peptide fragments (<1000 Da) may not be effectively captured by the typical covalent binding filters which have a pore size of 0.7–1.0 μM. Therefore, MBI which occurs by MIC or heme alkylation may produce significantly lower binding values compared with apoprotein adducts.

| CYP | Enzyme abundance (pmol/mg protein) | %CV |
|---|---|---|
| 1A2 | 52 | 67 |
| 2A6 | 20.1 | 173 |
| 2B6 | 11 | 143 |
| 2C8 | 24 | 81 |
| 2C9 | 73 | 54 |
| 2C18 | 1 | 106 |
| 2C9 | 14 | 106 |
| 2D6 | 8 | 61 |
| 2E1 | 61 | 61 |
| 3A | 210 | 163 |

**Figure 37** Average enzyme abundances taken from SimP450 V7.01 illustrating the various levels of P450 from the human liver. Values are represented as picomoles per milligram of protein with coefficient of variance (%CV) reported (Sheffield, UK).

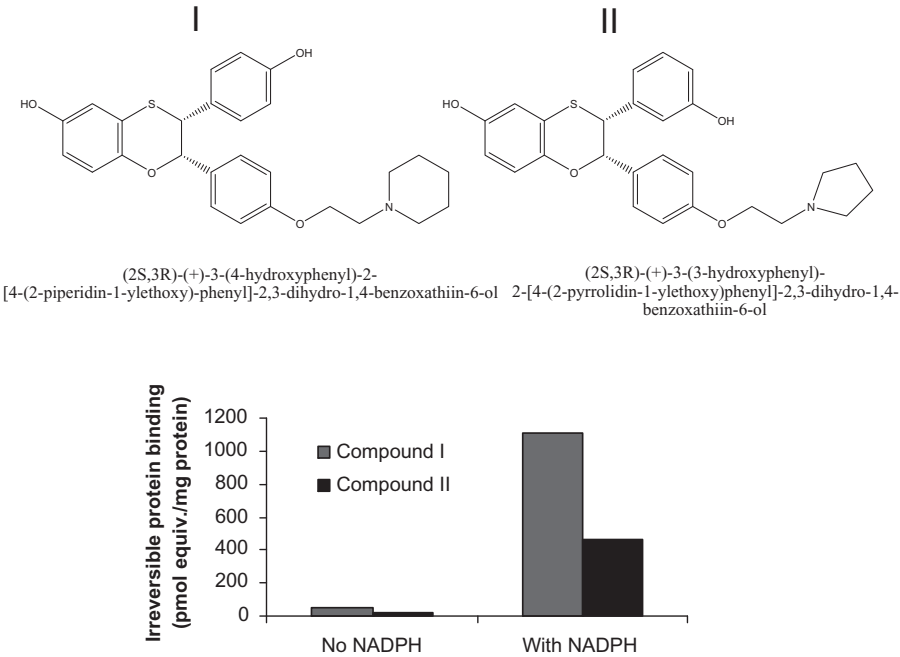


**Figure 38** Structural comparison of two compounds aimed as selective estrogen modulators were studied. Based upon the data from compound I, compound II was developed partially due to the reduction observed in covalent binding (Zhang *et al.*, 2005).

The integration of the covalent binding data with the described experiments (Fig. 16) provides the potential for an integrated picture of the structural and functional information as it relates to P450 MBI for a particular compound. However, despite the low levels of binding that may be observed, even low levels of protein adducts may be sufficient to induce an immune response detrimental to patients. Therefore, results from covalent binding experiments are critical to context against additional risk factors related to the intended therapy including frequency, dose, and disease state the intended therapeutic is aimed to treat. A thorough review and discussion of the risk/benefit factors related to covalent binding are provided elsewhere (Baillie, 2006; Doss and Baillie, 2006).

Implementing the structural and functional tools discussed should help to develop the most appropriate clinical candidate. If MBI still persists, it is imperative that the potential impact on the metabolism of other medications is quantitatively assessed in order to define the clinical limitations of the therapeutic agent including the ability to coadminister other medications, define at-risk patient population (e.g. age or pharmacogenetics), expected exposure of the MBI, and duration of treatment. This requires the combination of much of the data already accumulated from Fig. 16 including the $K_I$ and $k_{inact}$ with the endogenous enzyme degradation rate in order to describe the extent of drug interaction that could result based upon irreversible inhibition kinetics (Fig. 27). Additional information regarding the expected metabolism and disposition of the substrate responsible for producing irreversible inhibitor will typically serve to increase the predictive power of the anticipated changes in AUC, including an understanding of the overall metabolic pathway (existence of multiple inactivation pathways) and potential for gut metabolism may be incorporated in an attempt to refine the *in vivo* extrapolation. Furthermore, the fraction of total hepatic metabolism expected for a victim drug through the affected P450 may be considered for improved predictability of the magnitude of the drug interaction (Wang, Jones and Hall, 2004). These additional properties will be individually discussed.

For most purposes, an AUC change of greater than twofold can define a drug interaction (Kalgutkar, Obach and Maurer, 2007). The extrapolation of *in vitro* MBI data has been described and makes use of the general equation described in Fig. 39 (Eq. 1). This equation is the same for reversible inhibition with the exception of the $K_I/k_{inact}$ addition in the denominator of the equation, which for reversible inhibition reduces to 1 since there is no additional enzyme inactivation. The impact of this difference is illustrated in Fig. 15a, where the difference in midazolam AUC based upon coadministration of erythromycin is calculated without inactivation parameters compared with Fig. 15b, where the inactivation parameters are included in prediction.

Another important factor to consider is the $f_{m(P450)}$ which takes into account the extent to which the MBI-affected P450 contributes to the overall metabolism of the victim drug ($AUC_{po}$). Without accounting for the fraction metabolized, there may be a tendency to overpredict the clinically observed change in AUC. For P450 3A substrates, considering the extent of gut metabolism is an important factor in predicting the magnitude of an anticipated drug interaction. Based upon the inactivation kinetics, the percent of inhibition of gut metabolism results in an exponential relationship; therefore, when gut metabolism is inhibited substantially (>60%), the effect on AUC becomes more pronounced. When this is expected, Equation 1 from Fig. 39 takes on a new form represented by Equation 2.

Understanding the metabolite species responsible for P450 MBI may improve the accuracy of drug interaction predictions. This is especially poignant for MI complex

Expected *in vivo* interaction

1 $$\frac{AUCi}{AUC} = \frac{1}{1 + \frac{k_{inact} \cdot [I]}{K_I + [I]}}$$

2 $$\frac{AUCi}{AUC} = \frac{1}{1 + \left( \frac{f_{m\,(CYP)}}{\left( \frac{[I]}{K_I} \cdot \frac{k_{inact}}{k_{deg}} \right)} \right) + \left(1 - f_{m\,(CYP)}\right)}$$

3 $$\frac{AUCi}{AUC} = \frac{1}{F_g + \left( 1 - F_g \cdot \frac{1}{\left( \frac{[I]}{([I] + KI)_I} \cdot \frac{k_{inact}}{k_{deg}} \right)} \right) + \left(1 - f_{m\,(CYP)}\right)} \cdot \frac{1}{\left( \frac{f_{m\,(CYP)}}{\left( \frac{[I]}{K_I} \cdot \frac{k_{inact}}{k_{deg}} \right)} \right) + \left(1 - f_{m\,(CYP)}\right)}$$

**Figure 39** Kinetic equations used to extrapolate *in vitro* MBI into clinical predicted effect on AUC (1) general equation taking into account $K_I$, $k_{inact}$, $f_{m(P450)}$ and (2) same equation taking into consideration the gut metabolism by P4503A (Kalgutkar, Obach and Maurer, 2007).

formation from tertiary alkyl amine drugs, where multiple precursors to MBI may form. For example, verapamil exerted an unexpected drug interaction based upon the low plasma concentrations in combination with the initial assessment of P450 inhibition which was estimated to be low (Ma, Prueksaritanont and Lin, 2000). However, after further investigation, the *N*-dealkylation of verapamil occurs at two different sites producing two metabolites capable of MIC formation with P4503A (Wang, Jones and Hall, 2005). Furthermore, verapamil is administered as a racemic mixture and yields a mixture of metabolites from which the *S*-enantiomer of norverapamil is predicted to have the most pronounced effect with an $K_{inact}/K_I$ of 164 min/nmol and offers an explanation for the observed interactions with various drugs (Wang, Jones and Hall, 2005).

Choosing the inhibitor concentration is another variable that can impact the final estimate of a drug interaction. The most accurate inhibitor concentration would be that available to the liver hepatocyte but is unmeasurable (Obach *et al.*, 2006). Alternatively, inhibitor concentrations that could be used in addition to total systemic inhibitor concentrations are free systemic or total hepatic inlet concentrations represented again either as free or total (Fig. 40). The hepatic inlet concentrations can be estimated based upon the fraction of the inhibitor which is absorbed and unchanged in the intestine in combination with the absorption rate constant, projected dose, and human hepatic blood flow (Kanamitsu, Ito and Sugiyama, 2000). However, the absorption rate constant and fraction absorbed unchanged are challenging values to obtain and potentially represent a source of error. Bearing patient safety in mind, the most conservative approach is to use total inhibitor concentration. In many cases, using all four inhibitor concentrations for the extrapolation may provide an upper and lower bounds of the potential clinical manifestation of the drug interaction (Fig. 40).

| Inhibiting drug | $IC_{50}$ (μM) | | | | Affected drug | Prediction based on most sensitive $IC_{50}$ | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Testosterone | Midazolam | Felodipine | fu | | [I]= Csys | [I]= fu*Csys | [I]= Chep | [I]= fu*Chep | Actual drug interaction |
| Buspirone | 70 | 59 | >300 | 0.05 | Alprazolam | 1 | 1 | 1.04 | 1 | 1.08 |
| Cyclosporin | 4.5 | 1.4 | 300 | 0.07 | Simvastatin | 2.52 | 1.54 | 4.83 | 1.74 | 2.55 |
| Diltiazem | 60 | 83 | >300 | 0.22 | Buspirone | 1.4 | 1.39 | 1.76 | 1.48 | 5.33 |
| Disulfiram | 6.9 | 3.6 | 7.9 | 0.04 | Midazolam | 1.74 | 1.74 | 13.8 | 2.7 | 0.94 |
| Erythromycin | 220 | 18 | 16 | 0.16 | Buspirone | 4.64 | 3.42 | 12.2 | 4.95 | 5.91 |
| Fluconazole | 17 | 6.8 | 23 | 0.89 | Midazolam | 7.64 | 7.15 | 10.5 | 9.86 | 3.6 |
| Fluoxetine | 16 | 36 | 88 | 0.05 | Alprazolam | 1.04 | 1.01 | 1.16 | 1.01 | 1.32 |
| Fluvoxamine | 28 | 43 | 62 | 0.23 | Buspirone | 2.41 | 2.38 | 3.4 | 2.61 | 2.4 |
| Itraconazole | 0.01 | 0.045 | 0.021 | 0.002 | Buspirone | 59.5 | 5.56 | 67.3 | 14.3 | 18.2 |
| Ketoconazole | 0.012 | 0.011 | 0.019 | 0.01 | Midazolam | 24.5 | 8.5 | 25 | 18.2 | 17 |
| Mibefradil | 0.33 | 0.19 | 0.68 | 0.005 | Midazolam | 10.3 | 1.82 | 20 | 2.16 | 8.86 |
| Nefazodone | 1.4 | 0.9 | 8.6 | 0.009 | Midazolam | 10.3 | 1.87 | 19.6 | 2.42 | 5.44 |
| Nelfinavir | 0.65 | 0.17 | 0.18 | 0.015 | Simvastatin | 19.3 | 3.54 | 21.4 | 12.7 | 6.07 |
| Paroxetine | 23 | 23 | 17 | 0.05 | Alprazolam | 1.02 | 1.01 | 1.12 | 1.01 | 0.99 |
| Ritonavir | 0.0088 | 0.0042 | 0.0037 | 0.015 | Triazolam | 22.2 | 17 | 22.3 | 20.3 | 20.3 |
| Saquinavir | 2 | 0.82 | 0.51 | 0.02 | Midazolam | 16.9 | 2.58 | 24.4 | 10.9 | 5.17 |
| Sertraline | 8.6 | 16 | 18 | 0.02 | Alprazolam | 1.05 | 1.01 | 1.52 | 1.02 | 1 |
| Simvastatin | 2 | 0.77 | 2.6 | 0.06 | Midazolam | 1.65 | 1.49 | 2.32 | 1.53 | 1.13 |
| Troleandomycin | 9.3 | 2.6 | 12 | 0.15 | Midazolam | 3.72 | 2.05 | 14.8 | 5.54 | 4.1 |
| Verapamil | 23 | 23 | 50 | 0.1 | Midazolam | 1.45 | 1.4 | 2.62 | 1.52 | 2.91 |

**Figure 40** A table of *in vitro* $IC_{50}$ generated against three different P4503A probe substrates with numerous chemical inhibitors. The subsequent extrapolated DDI expected based upon different inhibitor concentrations which can then be compared with the actual clinically observed interaction in the last column.

The successful extrapolation of drug interaction with an irreversible inhibitor is also reliant upon an accurate rate of *de novo* P450 biosynthesis and degradation. Experimentally measured values for such a parameter in humans are not obtainable. Furthermore, the degradation rates are known to fluctuate based upon environmental factors and are also unique for individual P450s. To date, the best expected outcome is an estimate of the degradation value based upon existing experimental data. Degradation values for different P450 enzyme have been estimated using numerous techniques including (i) modeling the time course of P450 to return to baseline post-induction, (ii) turnover in primary human hepatocytes using pulse-chase methodology, (iii) Caco-2 cell line to estimate $k_{deg}$ of P4503A, and (iv) the degradation of rat P450 3A. Clinical pharmacokinetic data for P450 1A2 (Faber and Fuhr, 2004), P450 2D6 (Venkatakrishnan and Obach, 2005), and intestinal P450 3A (Greenblatt *et al.*, 2003) based on the time course of de-induction following smoking cessation, recovery following paroxetine inactivation, and recovery following inactivation by grapefruit juice, respectively, have been utilized to determine degradation rate constants. The resulting $k_{deg}$ estimates (per minute) were 0.000296, 0.000226, and 0.000481 for P450 1A2, P450 2D6, and intestinal P450 3A, respectively. For hepatic P450 3A, an initial estimate of 0.000321/min was used based on the kinetics of de-induction of the oral clearance of verapamil (Fromm *et al.*, 1996), general clinical pharmacologic understanding of the kinetics of induction and de-induction of P450 3A (Lin *et al.*, 2001), and additionally supported by *in vitro* estimates of P450 3A turnover in primary human hepatocytes using pulse-chase methodology (Pichard *et al.*, 1992).

The integration of most of these kinetic factors has been successful within the simulation software provided by Simcyp. The program provides a platform to integrate the aforementioned concepts in order to interrogate a compound's ability to impact the AUC of other drugs based upon different dosing regimens, inactivation rates, fraction metabolized, and gut metabolism to successfully extrapolate the DDI potential of a new clinical candidate. In addition, Simcyp has developed the complexity necessary to investigate induction combined with reversible or irreversible P450 inhibition.

The potential for pharmacogenetics to impact MBI has only recently been investigated (McConn *et al.*, 2004; Pinto *et al.*, 2005; Op den Buijsch *et al.*, 2007). Hall *et al.* have shown numerous examples where P450 3A4 and P4503A5 display different kinetics with respect to inactivation of each P450, predominantly with MIC-forming compounds including diltiazem and verapamil. Differences in metabolism and generation of reactive metabolites could also contribute to different levels of exposure to reactive metabolites. Furthermore, with the advances in genotyping, it may become more apparent that some patients are prone to SNP (single nucleotide polymorphism) clusters. A cluster of SNPs in drug-metabolizing enzymes could exacerbate MBI and the potential for other adverse drug reactions. A recent retrospective study examined individuals linked to the ingestion of cooking oil contaminated with polychlorinated biphenyls (PCBs) and polychlorinated dibenzofurans (PCDFs) in 1979. Subjects genotyped to have the combination of the P4501A1-MspI mutant and GSTM1-null genotype were associated with increased risk of chloracne from the exposure of PCBs and PCDFs (odds ratio 2.8, 95% confidence interval 1.1–7.6) (Tsai *et al.*, 2006). A similar study found higher levels of BPDE-DNA adducts in individuals with the combined P4501A1(1/*2 or *2A/*2A)-GSTM1*0/*0 genotype suggest that these genotype combinations are at increased risk for contracting lung cancer when exposed to PAH (Goth-Goldstein *et al.*, 2000). As illustrated in the latter example, benzo[*a*]pyrene metabolism and moreover the generation of its DNA reactive metabolite were

correlated to SNPs, and increased risk of cancer illustrates how the merging of experimental technologies and bioinformatics can provide exciting new opportunities to explore SNPs and disease states as they relate to xenobiotic metabolism.

Given the poor predictability of preclinical *in vivo* models for toxicity, *in vitro* systems continue to provide important experimental support for safe therapeutics. The integration of new *in vitro* test systems for toxicity (sandwich-cultured hepatocytes) may be able to provide a more global assessment of a compound's safety profile by combining transporters, phase I and phase II metabolism into a single functioning assay. This may expose new mechanism of complex metabolism-based inhibition, for example, the possibility of phase II conjugates capable of eliciting P450 inhibition. Only recently have other glucuronides been examined for their potential to elicit an adverse effect on drug-metabolizing enzymes as in the case of the gemfibrazil glucuronide which functions as an MBI against P450 2C8 (Ogilvie *et al.*, 2006). Interestingly, the unique bioactivation of troglitazone was postulated to have the potential to elicit a multiple drug binding event if disulfide formation occurred from the isocyanate intermediate followed by subsequent bioactivation of the chromane ring to the quinone; this could inactivate other systems such as transporters (Kassahun *et al.*, 2001). The combination of technology and biology should provide methods and insight to the structural and functional aspects of MBI including a better understanding of idiosyncratic drug reactions.

## REFERENCES

Alterman, M.A., Kornilayev, B., Duzhak, T., Yakovlev, D. (2005). Quantitative analysis of cytochrome p450 isozymes by means of unique isozyme-specific tryptic peptides: a proteomic approach. *Drug Metabolism and Disposition*, *33*, 1399–1407.

Alvarez-Diez, T.M., Zheng, J. (2004). Mechanism-based inactivation of cytochrome P450 3A4 by 4-ipomeanol. *Chemical Research in Toxicology*, *17*, 150–157.

Alvarez-Sanchez, R., Montavon, F., Hartung, T., Pahler, A. (2006). Thiazolidinedione bioactivation: a comparison of the bioactivation potentials of troglitazone, rosiglitazone, and pioglitazone using stable isotope-labeled analogues and liquid chromatography tandem mass spectrometry. *Chemical Research in Toxicology*, *19*, 1106–1116.

Argoti, D., Liang, L., Conteh, A., Chen, L., Bershas, D., Yu, C.P., Vouros, P., Yang, E. (2005). Cyanide trapping of iminium ion reactive intermediates followed by detection and structure identification using liquid chromatography-tandem mass spectrometry (LC-MS/MS). *Chemical Research in Toxicology*, *18*, 1537–1544.

Axelrod, J. (1955). The enzymatic conversion of codeine to morphine. *The Journal of Pharmacology and Experimental Therapeutics*, *115*, 259–267.

Baer, B.R., Wienkers, L.C., Rock, D.A. (2007). Time-dependent inactivation of P450 3A4 by raloxifene: identification of Cys239 as the site of apoprotein alkylation. *Chemical Research in Toxicology*, *20*, 954–964.

Baillie, T.A. (2006). Future of toxicology-metabolic activation and drug design: challenges and opportunities in chemical toxicology. *Chemical Research in Toxicology*, *19*, 889–893.

Baillie, T.A., Davis, M.R. (1993). Mass spectrometry in the analysis of glutathione conjugates. *Biological Mass Spectrometry*, *22*, 319–325.

Bartkowski, R.R., Goldberg, M.E., Larijani, G.E., Boerner, T. (1989). Inhibition of alfentanil metabolism by erythromycin. *Clinical Pharmacology and Therapeutics*, *46*, 99–102.

Beaune, P., Dansette, P.M., Mansuy, D., Kiffel, L., Finck, M., Amar, C., Leroux, J.P., Homberg, J.C. (1987). Human anti-endoplasmic reticulum autoantibodies appearing in a drug-induced hepa-

titis are directed against a human liver cytochrome P-450 that hydroxylates the drug. *Proceedings of the National Academy of Sciences of the United States of America*, *84*, 551–555.

Bensoussan, C., Delaforge, M., Mansuy, D. (1995). Particular ability of cytochromes P450 3A to form inhibitory P450-iron-metabolite complexes upon metabolic oxidation of aminodrugs. *Biochemical Pharmacology*, *49*, 591–602.

Bertz, R.J., Granneman, G.R. (1997). Use of in vitro and in vivo data to estimate the likelihood of metabolic pharmacokinetic interactions. *Clinical Pharmacokinetics*, *32*, 210–258.

Bieler, C.A., Stiborova, M., Wiessler, M., Cosyns, J.P., van Ypersele de Strihou, C., Schmeiser, H.H. (1997). 32P-post-labelling analysis of DNA adducts formed by aristolochic acid in tissues from patients with Chinese herbs nephropathy. *Carcinogenesis*, *18*, 1063–1067.

Bjornsson, T.D., Callaghan, J.T., Einolf, H.J., Fischer, V., Gan, L., Grimm, S., Kao, J., King, S.P., Miwa, G., Ni, L., Kumar, G., McLeod, J., Obach, R.S., Roberts, S., Roe, A., Shah, A., Snikeris, F., Sullivan, J.T., Tweedie, D., Vega, J.M., Walsh, J., Wrighton, S.A. (2003). The conduct of in vitro and in vivo drug-drug interaction studies: a Pharmaceutical Research and Manufacturers of America (PhRMA) perspective. *Drug Metabolism and Disposition: The Biological Fate of Chemicals*, *31*, 815–832.

Black, D.J., Kunze, K.L., Wienkers, L.C., Gidal, B.E., Seaton, T.L., McDonnell, N.D., Evans, J.S., Bauwens, J.E., Trager, W.F. (1996). Warfarin-fluconazole. II. A metabolically based drug interaction: in vivo studies. *Drug Metabolism and Disposition: The Biological Fate of Chemicals*, *24*, 422–428.

Brodie, B.B., Gillette, J.R., La Du, B.N. (1958). Enzymatic metabolism of drugs and other foreign compounds. *Annual Review of Biochemistry*, *27*, 427–454.

Chae, Y.H., Thomas, T., Guengerich, F.P., Fu, P.P., El-Bayoumy, K. (1999). Comparative metabolism of 1-, 2-, and 4-nitropyrene by human hepatic and pulmonary microsomes. *Cancer Research*, *59*, 1473–1480.

Chan, W.K., Delucchi, A.B. (2000). Resveratrol, a red wine constituent, is a mechanism-based inactivator of cytochrome P450 3A4. *Life Sciences*, *67*, 3103–3112.

Chan, W.K., Sui, Z. and Ortiz de Montellano, P.R. (1993). Determinants of protein modification versus heme alkylation: inactivation of cytochrome P450 1A1 by 1-ethynylpyrene and phenylacetylene. *Chemical Research in Toxicology*, *6*, 38–45.

Chen, L.J., Hecht, S.S., Peterson, L.A. (1997). Characterization of amino acid and glutathione adducts of cis-2-butene-1,4-dial, a reactive metabolite of furan. *Chemical Research in Toxicology*, *10*, 866–874.

Chen, Q., Ngui, J.S., Doss, G.A., Wang, R.W., Cai, X., DiNinno, F.P., Blizzard, T.A., Hammond, M.L., Stearns, R.A., Evans, D.C., Baillie, T.A., Tang, W. (2002). Cytochrome P450 3A4-mediated bioactivation of raloxifene: irreversible enzyme inhibition and thiol adduct formation. *Chemical Research in Toxicology*, *15*, 907–914.

Dansette, P.M., Bertho, G., Mansuy, D. (2005). First evidence that cytochrome P450 may catalyze both S-oxidation and epoxidation of thiophene derivatives. *Biochemical and biophysical Research Communications*, *338*, 450–455.

Davit, B., Reynolds, K., Yuan, R., Ajayi, F., Conner, D., Fadiran, E., Gillespie, B., Sahajwalla, C., Huang, S.M., Lesko, L.J. (1999). FDA evaluations using in vitro metabolism to predict and interpret in vivo metabolic drug-drug interactions: impact on labeling. *Journal of Clinical Pharmacology*, *39*, 899–910.

Desta, Z., Ward, B.A., Soukhova, N.V., Flockhart, D.A. (2004). Comprehensive evaluation of tamoxifen sequential biotransformation by the human cytochrome P450 system in vitro: prominent roles for CYP3A and CYP2D6. *The Journal of Pharmacology and Experimental Therapeutics*, *310*, 1062–1075.

DeWitte, R.S., Robins, R.H. (2006). A hierarchical screening methodology for physicochemical/ADME/Tox profiling. *Expert Opinion on Drug Metabolism & Toxicology*, *2*, 805–817.

Dickins, M., Elcombe, C.R., Moloney, S.J., Netter, K.J., Bridges, J.W. (1979). Further studies on the dissociation of the isosafrole metabolite-cytochrome P-450 complex. *Biochemical Pharmacology*, *28*, 231–238.

Dieckhaus, C.M., Fernandez-Metzler, C.L., King, R., Krolikowski, P.H., Baillie, T.A. (2005). Negative ion tandem mass spectrometry for the detection of glutathione conjugates. *Chemical Research in Toxicology*, *18*, 630–638.

Doss, G.A., Baillie, T.A. (2006). Addressing metabolic activation as an integral component of drug design. *Drug Metabolism Reviews*, *38*, 641–649.

Druckova, A., Marnett, L.J. (2006). Characterization of the amino acid adducts of the enedial derivative of teucrin A. *Chemical Research in Toxicology*, *19*, 1330–1340.

Ernest, C.S., II, Hall, S.D., Jones, D.R. (2005). Mechanism-based inactivation of CYP3A by HIV protease inhibitors. *The Journal of Pharmacology and Experimental Therapeutics*, *312*, 583–591.

Estabrook, R.W., Hildebrandt, A.G., Baron, J., Netter, K.J., Leibman, K. (1971). A new spectral intermediate associated with cytochrome P-450 function in liver microsomes. *Biochemical and Biophysical Research Communications*, *42*, 132–139.

Evans, D.C., Watt, A.P., Nicoll-Griffith, D.A., Baillie, T.A. (2004). Drug-protein adducts: an industry perspective on minimizing the potential for drug bioactivation in drug discovery and development. *Chemical Research in Toxicology*, *17*, 3–16.

Faber, M.S., Fuhr, U. (2004). Time response of cytochrome P450 1A2 activity on cessation of heavy smoking. *Clinical Pharmacology and Therapeutics*, *76*, 178–184.

Fabre, I., Fabre, G., Maurel, P., Bertault-Peres, P., Cano, J.P. (1988). Metabolism of cyclosporin A. III. Interaction of the macrolide antibiotic, erythromycin, using rabbit hepatocytes and microsomal fractions. *Drug Metabolism and Disposition: The Biological Fate of Chemicals*, *16*, 296–301.

Fersht, A.R. (1974). Catalysis, binding and enzyme-substrate complementarity. *Proceedings of the Royal Society of London. Series B, Containing Papers of a Biological Character*, *187*, 397–407.

Fontana, E., Dansette, P.M., Poli, S.M. (2005). Cytochrome p450 enzymes mechanism based inhibitors: common sub-structures and reactivity. *Current Drug Metabolism*, *6*, 413–454.

Foroozesh, M., Primrose, G., Guo, Z., Bell, L.C., Alworth, W.L., Guengerich, F.P. (1997). Aryl acetylenes as mechanism-based inhibitors of cytochrome P450-dependent monooxygenase enzymes. *Chemical Research in Toxicology*, *10*, 91–102.

Franklin, M.R. (1995). Enhanced rates of cytochrome P450 metabolic-intermediate complex formation from nonmacrolide amines in rifampicin-treated rabbit liver microsomes. *Drug Metabolism and Disposition*, *23*, 1379–1382.

Friedman, M.A., Woodcock, J., Lumpkin, M.M., Shuren, J.E., Hass, A.E., Thompson, L.J. (1999). The safety of newly approved medicines: do recent market removals mean there is a problem? *JAMA*, *281*, 1728–1734.

Fromm, M.F., Busse, D., Kroemer, H.K., Eichelbaum, M. (1996). Differential induction of prehepatic and hepatic metabolism of verapamil by rifampin. *Hepatology*, *24*, 796–801.

Gartner, C.A., Wen, B., Wan, J., Becker, R.S., Jones, G., 2nd, Gygi, S.P., Nelson, S.D. (2005). Photochromic agents as tools for protein structure study: lapachenole is a photoaffinity ligand of cytochrome P450 3A4. *Biochemistry*, *44*, 1846–1855.

Ghanbari, F., Rowland-Yeo, K., Bloomer, J.C., Clarke, S.E., Lennard, M.S., Tucker, G.T., Rostami-Hodjegan, A. (2006). A critical evaluation of the experimental design of studies of mechanism based enzyme inhibition, with implications for in vitro-in vivo extrapolation. *Current Drug Metabolism*, *7*, 315–334.

Goth-Goldstein, R., Stampfer, M.R., Erdmann, C.A., Russell, M. (2000). Interindividual variation in CYP1A1 expression in breast tissue and the role of genetic polymorphism. *Carcinogenesis*, *21*, 2119–2122.

Graham, D.J., Green, L., Senior, J.R., Nourjah, P. (2003). Troglitazone-induced liver failure: a case study. *American Journal of Medicine*, *114*, 299–306.

Greenblatt, D.J., von Moltke, L.L., Harmatz, J.S., Chen, G., Weemhoff, J.L., Jen, C., Kelley, C.J., LeDuc, B.W., Zinny, M.A. (2003). Time course of recovery of cytochrome p450 3A function after single doses of grapefruit juice. *Clinical Pharmacology and Therapeutics*, *74*, 121–129.

Grillo, M.P., Hua, F., Knutson, C.G., Ware, J.A., Li, C. (2003). Mechanistic studies on the bioactivation of diclofenac: identification of diclofenac-S-acyl-glutathione in vitro in incubations with rat and human hepatocytes. *Chemical Research in Toxicology*, *16*, 1410–1417.

Groves, J.T., McClusky, G.A. (1978). Aliphatic hydroxylation by highly purified liver microsomal cytochrome P-450. Evidence for a carbon radical intermediate. *Biochemical and Biophysical Research Communications*, *81*, 154–160.

Guengerich, F.P. (1990). Mechanism-based inactivation of human liver microsomal cytochrome P-450 IIIA4 by gestodene. *Chemical Research in Toxicology*, *3*, 363–371.

Guengerich, F.P. (1997). Role of cytochrome P450 enzymes in drug-drug interactions. *Advances in Pharmacology (San Diego, California)*, *43*, 7–35.

Ha-Duong, N.T., Dijols, S., Macherey, A.C., Goldstein, J.A., Dansette, P.M., Mansuy, D. (2001). Ticlopidine as a selective mechanism-based inhibitor of human cytochrome P450 2C19. *Biochemistry*, *40*, 12112–12122.

Hanioka, N., Ozawa, S., Jinno, H., Tanaka-Kagawa, T., Nishimura, T., Ando, M., Sawada, J.-I. (2002). Interaction of irinotecan (CPT-11) and its active metabolite 7-ethyl-10-hydroxycamptothecin (SN-38) with human cytochrome P450 enzymes. *Drug Metabolism and Disposition: The Biological Fate of Chemicals*, *30*, 391–396.

Hara, Y., Nakajima, M., Miyamoto, K.I., Yokoi, T. (2005). Inhibitory effects of psychotropic drugs on mexiletine metabolism in human liver microsomes: prediction of in vivo drug interactions. *Xenobiotica; the Fate of Foreign Compounds in Biological Systems*, *35*, 549–560.

Hasemann, C.A., Kurumbail, R.G., Boddupalli, S.S., Peterson, J.A., Deisenhofer, J. (1995). Structure and function of cytochromes P450: a comparative analysis of three crystal structures. *Structure*, *3*, 41–62.

He, K., Iyer, K.R., Hayes, R.N., Sinz, M.W., Woolf, T.F., Hollenberg, P.F. (1998). Inactivation of cytochrome P450 3A4 by bergamottin, a component of grapefruit juice. *Chemical Research in Toxicology*, *11*, 252–259.

He, K., Woolf, T.F., Hollenberg, P.F. (1999). Mechanism-based inactivation of cytochrome P-450-3A4 by mifepristone (RU486). *The Journal of Pharmacology and Experimental Therapeutics*, *288*, 791–797.

He, K., Talaat, R.E., Pool, W.F., Reily, M.D., Reed, J.E., Bridges, A.J., Woolf, T.F. (2004). Metabolic activation of troglitazone: identification of a reactive metabolite and mechanisms involved. *Drug Metabolism and Disposition*, *32*, 639–646.

Higgins, L., Korzekwa, K.R., Rao, S., Shou, M., Jones, J.P. (2001). An assessment of the reaction energetics for cytochrome P450-mediated reactions. *Archives of Biochemistry and Biophysics*, *385*, 220–230.

Hildebrandt, A., Estabrook, R.W. (1971). Evidence for the participation of cytochrome b 5 in hepatic microsomal mixed-function oxidation reactions. *Archives of Biochemistry and Biophysics*, *143*, 66–79.

Honig, P.K., Woosley, R.L., Zamani, K., Conner, D.P., Cantilena, L.R., Jr (1992). Changes in the pharmacokinetics and electrocardiographic pharmacodynamics of terfenadine with concomitant administration of erythromycin. *Clinical Pharmacology and Therapeutics*, *52*, 231–238.

Hutzler, J.M., Melton, R.J., Rumsey, J.M., Schnute, M.E., Locuson, C.W., Wienkers, L.C. (2006). Inhibition of cytochrome P450 3A4 by a pyrimidineimidazole: evidence for complex heme interactions. *Chemical Research in Toxicology*, *19*, 1650–1659.

Hutzler, J.M., Steenwyk, R.C., Smith, E.B., Walker, G.S., Wienkers, L.C. (2004). Mechanism-based inactivation of cytochrome P450 2D6 by 1-[(2-ethyl-4-methyl-1H-imidazol-5-yl)methyl]- 4-[4-

(trifluoromethyl)-2-pyridinyl]piperazine: kinetic characterization and evidence for apoprotein adduction. *Chemical Research in Toxicology*, *17*, 174–184.

Hutzler, J.M., Walker, G.S., Wienkers, L.C. (2003). Inhibition of cytochrome P450 2D6: structure-activity studies using a series of quinidine and quinine analogues. *Chemical Research in Toxicology*, *16*, 450–459.

Isin, E.M., Guengerich, F.P. (2007). Complex reactions catalyzed by cytochrome P450 enzymes. *Biochimica et Biophysica Acta*, *1770*, 314–329.

Ito, K., Ogihara, K., Kanamitsu, S., Itoh, T. (2003). Prediction of the in vivo interaction between midazolam and macrolides based on in vitro studies using human liver microsomes. *Drug Metabolism and Disposition: The Biological Fate of Chemicals*, *31*, 945–954.

Jones, D.R., Gorski, J.C., Hamman, M.A., Mayhew, B.S., Rider, S., Hall, S.D. (1999). Diltiazem inhibition of cytochrome P-450 3A activity is due to metabolite intermediate complex formation. *The Journal of Pharmacology and Experimental Therapeutics*, *290*, 1116–1125.

Jones, J.P., Shou, M., Korzekwa, K.R. (1996). Predicting the regioselectivity and stereoselectivity of cytochrome P450-mediated reactions: structural models for bioactivation reactions. *Advances in Experimental Medicine and Biology*, *387*, 355–360.

Jushchyshyn, M.I., Hutzler, J.M., Schrag, M.L., Wienkers, L.C. (2005). Catalytic turnover of pyrene by CYP3A4: evidence that cytochrome b5 directly induces positive cooperativity. *Archives of Biochemistry and Biophysics*, *438*, 21–28.

Kalgutkar, A.S., Obach, R.S., Maurer, T.S. (2007). Mechanism-based inactivation of cytochrome P450 enzymes: chemical mechanisms, structure-activity relationships and relationship to clinical drug-drug interactions and idiosyncratic adverse drug reactions. *Current Drug Metabolism*, *8*, 407–447.

Kalgutkar, A.S., Soglia, J.R. (2005). Minimising the potential for metabolic activation in drug discovery. *Expert Opinion on Drug Metabolism & Toxicology*, *1*, 91–142.

Kanamitsu, S., Ito, K., Green, C.E., Tyson, C.A., Shimada, N., Sugiyama, Y. (2000). Prediction of in vivo interaction between triazolam and erythromycin based on in vitro studies using human liver microsomes and recombinant human CYP3A4. *Pharmaceutical Research*, *17*, 419–426.

Kanamitsu, S., Ito, K., Sugiyama, Y. (2000). Quantitative prediction of in vivo drug-drug interactions from in vitro data based on physiological pharmacokinetics: use of maximum unbound concentration of inhibitor at the inlet to the liver. *Pharmaceutical Research*, *17*, 336–343.

Karuzina, I.I., Archakov, A.I. (1994). The oxidative inactivation of cytochrome P450 in monooxygenase reactions. *Free Radical Biology & Medicine*, *16*, 73–97.

Kassahun, K., Pearson, P.G., Tang, W., McIntosh, I., Leung, K., Elmore, C., Dean, D., Wang, R., Doss, G., Baillie, T.A. (2001). Studies on the metabolism of troglitazone to reactive intermediates in vitro and in vivo. Evidence for novel biotransformation pathways involving quinone methide formation and thiazolidinedione ring scission. *Chemical Research in Toxicology*, *14*, 62–70.

Kent, U.M., Juschyshyn, M.I., Hollenberg, P.F. (2001). Mechanism-based inactivators as probes of cytochrome P450 structure and function. *Current Drug Metabolism*, *2*, 215–243.

Khojasteh-Bakht, S.C., Chen, W., Koenigs, L.L., Peter, R.M., Nelson, S.D. (1999). Metabolism of (R)-(+)-pulegone and (R)-(+)-menthofuran by human liver cytochrome P-450s: evidence for formation of a furan epoxide. *Drug Metabolism and Disposition: The Biological Fate of Chemicals*, *27*, 574–580.

Kitz, R., Wilson, I.B. (1962). Esters of methanesulfonic acid as irreversible inhibitors of acetylcholinesterase. *Journal of Biological Chemistry*, *237*, 3245–3249.

Klingenberg, M. (1958). Pigments of rat liver microsomes. *Archives Biochemistry and Biophysics*, *75*, 376–386.

Koenigs, L.L., Trager, W.F. (1998a). Mechanism-based inactivation of P450 2A6 by furanocoumarins. *Biochemistry*, *37*, 10047–10061.

Koenigs, L.L., Trager, W.F. (1998b). Mechanism-based inactivation of cytochrome P450 2B1 by 8-methoxypsoralen and several other furanocoumarins. *Biochemistry*, *37*, 13184–13193.

Koenigs, L.L., Peter, R.M., Hunter, A.P., Haining, R.L., Rettie, A.E., Friedberg, T., Pritchard, M.P., Shou, M., Rushmore, T.H., Trager, W.F. (1999). Electrospray ionization mass spectrometric analysis of intact cytochrome P450: identification of tienilic acid adducts to P450 2C9. *Biochemistry*, *38*, 2312–2319.

Koshland, D.E., Jr, Neet, K.E. (1968). The catalytic and regulatory properties of enzymes. *Annual Review of Biochemistry*, *37*, 359–410.

Kouzi, S.A., McMurtry, R.J., Nelson, S.D. (1994). Hepatotoxicity of germander (Teucrium chamaedrys L.) and one of its constituent neoclerodane diterpenes teucrin A in the mouse. *Chemical Research in Toxicology*, *7*, 850–856.

Kunze, K.L., Eddy, A.C., Gibaldi, M., Trager, W.F. (1991). Metabolic enantiomeric interactions: the inhibition of human (S)-warfarin-7-hydroxylase by (R)-warfarin. *Chirality*, *3*, 24–29.

Kunze, K.L., Trager, W.F. (1993). Isoform-selective mechanism-based inhibition of human cytochrome P450 1A2 by furafylline. *Chemical Research in Toxicology*, *6*, 649–656.

Lightning, L.K., Jones, J.P., Friedberg, T., Pritchard, M.P., Shou, M., Rushmore, T.H., Trager, W.F. (2000). Mechanism-based inactivation of cytochrome P450 3A4 by L-754,394. *Biochemistry*, *39*, 4276–4287.

Lim, H.K., Duczak, N., Jr, Brougham, L., Elliot, M., Patel, K., Chan, K. (2005). Automated screening with confirmation of mechanism-based inactivation of CYP3A4, CYP2C9, CYP2C19, CYP2D6, and CYP1A2 in pooled human liver microsomes. *Drug Metabolism and Disposition*, *33*, 1211–1219.

Lin, H.L., Hollenberg, P.F. (2007). The inactivation of cytochrome P450 3A5 by 17alpha-ethynylestradiol is cytochrome b5-dependent: metabolic activation of the ethynyl moiety leads to the formation of glutathione conjugates, a heme adduct, and covalent binding to the apoprotein. *Journal of Pharmacology and Experimental Therapeutics*, *321*, 276–287.

Lin, L.Y., Fujimoto, M., Distefano, E.W., Schmitz, D.A., Jayasinghe, A., Cho, A.K. (1996). Selective mechanism-based inactivation of rat CYP2D by 4-allyloxymethamphetamine. *The Journal of Pharmacology and Experimental Therapeutics*, *277*, 595–603.

Lin, Y.S., Lockwood, G.F., Graham, M.A., Brian, W.R., Loi, C.M., Dobrinska, M.R., Shen, D.D., Watkins, P.B., Wilkinson, G.R., Kharasch, E.D., Thummel, K.E. (2001). In-vivo phenotyping for CYP3A by a single-point determination of midazolam plasma concentration. *Pharmacogenetics*, *11*, 781–791.

Liu, H., Liu, J., van Breemen, R.B., Thatcher, G.R., Bolton, J.L. (2005). Bioactivation of the selective estrogen receptor modulator desmethylated arzoxifene to quinoids: 4'-fluoro substitution prevents quinoid formation. *Chemical Research in Toxicology*, *18*, 162–173.

Liu, J., Li, Q., Yang, X., van Breemen, R.B., Bolton, J.L., Thatcher, G.R. (2005). Analysis of protein covalent modification by xenobiotics using a covert oxidatively activated tag: raloxifene proof-of-principle study. *Chemical Research in Toxicology*, *18*, 1485–1496.

Lopez-Garcia, M.P., Dansette, P.M., Mansuy, D. (1994). Thiophene derivatives as new mechanism-based inhibitors of cytochromes P-450: inactivation of yeast-expressed human liver cytochrome P-450 2C9 by tienilic acid. *Biochemistry*, *33*, 166–175.

Lu, A.Y., Coon, M.J. (1968). Role of hemoprotein P-450 in fatty acid omega-hydroxylation in a soluble enzyme system from liver microsomes. *The Journal of Biological Chemistry*, *243*, 1331–1332.

Lu, P., Schrag, M.L., Slaughter, D.E., Raab, C.E., Shou, M., Rodrigues, A.D. (2003). Mechanism-based inhibition of human liver microsomal cytochrome P450 1A2 by zileuton, a 5-lipoxygenase inhibitor. *Drug Metabolism and Disposition: The Biological Fate of Chemicals*, *31*, 1352–1360.

Ma, B., Prueksaritanont, T., Lin, J.H. (2000). Drug interactions with calcium channel blockers: possible involvement of metabolite-intermediate complexation with CYP3A. *Drug Metabolism and Disposition*, *28*, 125–130.

Masubuchi, N., Makino, C., Murayama, N. (2007). Prediction of in vivo potential for metabolic activation of drugs into chemically reactive intermediate: correlation of in vitro and in vivo generation of reactive intermediates and in vitro glutathione conjugate formation in rats and humans. *Chemical Research in Toxicology*, *20*, 455–464.

Masubuchi, Y. (2006). Metabolic and non-metabolic factors determining troglitazone hepatotoxicity: a review. *Drug Metabolism and Pharmacokinetics*, *21*, 347–356.

Maurer, T., Fung, H.L. (2000). Comparison of methods for analyzing kinetic data from mechanism-based enzyme inactivation: application to nitric oxide synthase. *AAPS PharmSci*, *2*, E8.

Mayhew, B.S., Jones, D.R., Hall, S.D. (2000). An in vitro model for predicting in vivo inhibition of cytochrome P450 3A4 by metabolic intermediate complex formation. *Drug Metabolism and Disposition*, *28*, 1031–1037.

McConn, D.J., II, Lin, Y.S., Allen, K., Kunze, K.L., Thummel, K.E. (2004). Differences in the inhibition of cytochromes P450 3A4 and 3A5 by metabolite-inhibitor complex-forming drugs. *Drug Metabolism and Disposition: The Biological Fate of Chemicals*, *32*, 1083–1091.

Melet, A., Assrir, N., Jean, P., Pilar Lopez-Garcia, M., Marques-Soares, C., Jaouen, M., Dansette, P.M., Sari, M.A., Mansuy, D. (2003). Substrate selectivity of human cytochrome P450 2C9: importance of residues 476, 365, and 114 in recognition of diclofenac and sulfaphenazole and in mechanism-based inactivation by tienilic acid. *Archives of Biochemistry and Biophysics*, *409*, 80–91.

Michalets, E.L. (1998). Update: clinically significant cytochrome P-450 drug interactions. *Pharmacotherapy*, *18*, 84–112.

von Moltke, L.L., Durol, A.L., Duan, S.X., Greenblatt, D.J. (2000). Potent mechanism-based inhibition of human CYP3A in vitro by amprenavir and ritonavir: comparison with ketoconazole. *European Journal Clinical Pharmacology*, *56*, 259–261.

Monahan, B.P., Ferguson, C.L., Killeavy, E.S., Lloyd, B.K., Troy, J., Cantilena, L.R., Jr (1990). Torsades de pointes occurring in association with terfenadine use. *JAMA*, *264*, 2788–2790.

Murray, M., Murray, K. (2003). Mechanism-based inhibition of CYP activities in rat liver by fluoxetine and structurally similar alkylamines. *Xenobiotica; the Fate of Foreign compounds in Biological Systems*, *33*, 973–987.

Nebert, D.W., Adesnik, M., Coon, M.J., Estabrook, R.W., Gonzalez, F.J., Guengerich, F.P., Gunsalus, I.C., Johnson, E.F., Kemper, B., Levin, W. *et al.* (1987). The P450 gene superfamily: recommended nomenclature. *DNA (Mary Ann Liebert. Inc.)*, *6*, 1–11.

Nelson, D.R. (1999). Cytochrome P450 and the individuality of species. *Archives of Biochemistry and Biophysics*, *369*, 1–10.

Nelson, D.R., Koymans, L., Kamataki, T., Stegeman, J.J., Feyereisen, R., Waxman, D.J., Waterman, M.R., Gotoh, O., Coon, M.J., Estabrook, R.W., Gunsalus, I.C., Nebert, D.W. (1996). P450 superfamily: update on new sequences, gene mapping, accession numbers and nomenclature. *Pharmacogenetics*, *6*, 1–42.

Nguyen, T.L., Gruenke, L.D., Castagnoli, N., Jr (1976). Metabolic N-demethylation of nicotine. Trapping of a reactive iminium species with cyanide ion. *Journal of Medicinal Chemistry*, *19*, 1168–1169.

Nocerini, M.R., Yost, G.S., Carlson, J.R., Liberato, D.J., Breeze, R.G. (1985). Structure of the glutathione adduct of activated 3-methylindole indicates that an imine methide is the electrophilic intermediate. *Drug Metabolism and Disposition*, *13*, 690–694.

O'Donnell, J.P., Dalvie, D.K., Kalgutkar, A.S., Obach, R.S. (2003). Mechanism-based inactivation of human recombinant P450 2C9 by the nonsteroidal anti-inflammatory drug suprofen. *Drug Metabolism and Disposition: The Biological Fate of Chemicals*, *31*, 1369–1377.

Obach, R.S., Walsky, R.L., Venkatakrishnan, K. (2007). Mechanism-based inactivation of human cytochrome p450 enzymes and the prediction of drug-drug interactions. *Drug Metabolism and Disposition*, *35*, 246–255.

Obach, R.S., Walsky, R.L., Venkatakrishnan, K., Gaman, E.A., Houston, J.B., Tremaine, L.M. (2006). The utility of in vitro cytochrome P450 inhibition data in the prediction of drug-drug interactions. *Journal of Pharmacology and Experimental Therapeutics*, *316*, 336–348.

Ogilvie, B.W., Zhang, D., Li, W., Rodrigues, A.D., Gipson, A.E., Holsapple, J., Toren, P., Parkinson, A. (2006). Glucuronidation converts gemfibrozil to a potent, metabolism-dependent inhibitor of CYP2C8: implications for drug-drug interactions. *Drug Metabolism and Disposition*, *34*, 191–197.

Omura, T., Sato, R. (1962). A new cytochrome in liver microsomes. *The Journal of Biological Chemistry*, *237*, 1375–1376.

Omura, T., Sato, R. (1964). The carbon monoxide-binding pigment of liver microsomes. I. Evidence for its hemoprotein nature. *Journal of Biological Chemistry*, *239*, 2370–2378.

Op den Buijsch, R.A., Christiaans, M.H., Stolk, L.M., de Vries, J.E., Cheung, C.Y., Undre, N.A., van Hooff, J.P., van Dieijen-Visser, M.P., Bekers, O. (2007). Tacrolimus pharmacokinetics and pharmacogenetics: influence of adenosine triphosphate-binding cassette B1 (ABCB1) and cytochrome (CYP) 3A polymorphisms. *Fundamental & Clinical Pharmacology*, *21*, 427–435.

Ortiz de Montellano, P.R., Mathews, J.M. (1981). Autocatalytic alkylation of the cytochrome P-450 prosthetic haem group by 1-aminobenzotriazole. Isolation of an NN-bridged benzyne-protoporphyrin IX adduct. *The Biochemical Journal*, *195*, 761–764.

Pershing, L.K., Franklin, M.R. (1982). Cytochrome P-450 metabolic-intermediate complex formation and induction by macrolide antibiotics; a new class of agents. *Xenobiotica; the Fate of Foreign Compounds in Biological Systems*, *12*, 687–699.

Pichard, L., Fabre, I., Daujat, M., Domergue, J., Joyeux, H., Maurel, P. (1992). Effect of corticosteroids on the expression of cytochromes P450 and on cyclosporin A oxidase activity in primary cultures of human hepatocytes. *Molecular Pharmacology*, *41*, 1047–1055.

Pinto, A.G., Wang, Y.H., Chalasani, N., Skaar, T., Kolwankar, D., Gorski, J.C., Liangpunsakul, S., Hamman, M.A., Arefayene, M., Hall, S.D. (2005). Inhibition of human intestinal wall metabolism by macrolide antibiotics: effect of clarithromycin on cytochrome P450 3A4/5 activity and expression. *Clinical Pharmacology and Therapeutics*, *77*, 178–188.

Polasek, T.M., Elliot, D.J., Lewis, B.C., Miners, J.O. (2004). Mechanism-based inactivation of human cytochrome P4502C8 by drugs in vitro. *The Journal of Pharmacology and Experimental Therapeutics*, *311*, 996–1007.

Poulos, T.L. (1996). Ligands and electrons and haem proteins. *Nature Structural Biology*, *3*, 401–403.

Racha, J.K., Rettie, A.E., Kunze, K.L. (1998). Mechanism-based inactivation of human cytochrome P450 1A2 by furafylline: detection of a 1:1 adduct to protein and evidence for the formation of a novel imidazomethide intermediate. *Biochemistry*, *37*, 7407–7419.

Regal, K.A., Schrag, M.L., Kent, U.M., Wienkers, L.C., Hollenberg, P.F. (2000). Mechanism-based inactivation of cytochrome P450 2B1 by 7-ethynylcoumarin: verification of apo-P450 adduction by electrospray ion trap mass spectrometry. *Chemical Research in Toxicology*, *13*, 262–270.

Rettie, A.E., Korzekwa, K.R., Kunze, K.L., Lawrence, R.F., Eddy, A.C., Aoyama, T., Gelboin, H.V., Gonzalez, F.J., Trager, W.F. (1992). Hydroxylation of warfarin by human cDNA-expressed cytochrome P-450: a role for P-4502C9 in the etiology of (S)-warfarin-drug interactions. *Chemical Research in Toxicology*, *5*, 54–59.

Riley, R.J., Grime, K., Weaver, R. (2007). Time-dependent CYP inhibition. *Expert Opinion on Drug Metabolism & Toxicology*, *3*, 51–66.

Rodrigues, A.D., Mulford, D.J., Lee, R.D., Surber, B.W., Kukulka, M.J., Ferrero, J.L., Thomas, S.B., Shet, M.S., Estabrook, R.W. (1995). In vitro metabolism of terfenadine by a purified recombinant fusion protein containing cytochrome P4503A4 and NADPH-P450 reductase. Comparison to human liver microsomes and precision-cut liver tissue slices. *Drug Metabolism and Disposition: The Biological Fate of Chemicals*, *23*, 765–775.,

Roussel, F., Khan, K.K., Halpert, J.R. (2000). The importance of SRS-1 residues in catalytic specificity of human cytochrome P450 3A4. *Archives of Biochemistry and Biophysics*, *374*, 269–278.

Rowland, P., Blaney, F.E., Smyth, M.G., Jones, J.J., Leydon, V.R., Oxbrow, A.K., Lewis, C.J., Tennant, M.G., Modi, S., Eggleston, D.S., Chenery, R.J., Bridges, A.M. (2006). Crystal structure of human cytochrome P450 2D6. *Journal of Biological Chemistry*, *281*, 7614–7622.

Sahali-Sahly, Y., Balani, S.K., Lin, J.H., Baillie, T.A. (1996). In vitro studies on the metabolic activation of the furanopyridine L-754,394, a highly potent and selective mechanism-based inhibitor of cytochrome P450 3A4. *Chemical Research in Toxicology*, *9*, 1007–1012.

Sansen, S., Yano, J.K., Reynald, R.L., Schoch, G.A., Griffin, K.J., Stout, C.D., Johnson, E.F. (2007). Adaptations for the oxidation of polycyclic aromatic hydrocarbons exhibited by the structure of human P450 1A2. *Journal of Biological Chemistry*, *282*, 14348–14355.

Schoch, G.A., Yano, J.K., Wester, M.R., Griffin, K.J., Stout, C.D., Johnson, E.F. (2004). Structure of human microsomal cytochrome P450 2C8. Evidence for a peripheral fatty acid binding site. *Journal of Biological Chemistry*, *279*, 9497–9503.

Sharma, U., Roberts, E.S., Hollenberg, P.F. (1996). Formation of a metabolic intermediate complex of cytochrome P4502B1 by clorgyline. *Drug Metabolism and Disposition*, *24*, 1247–1253.

Silverman, R.B. (1988). The potential use of mechanism-based enzyme inactivators in medicine. *Journal of Enzyme Inhibition*, *2*, 73–90.

Skiles, G.L., Yost, G.S. (1996). Mechanistic studies on the cytochrome P450-catalyzed dehydrogenation of 3-methylindole. *Chemical Research in Toxicology*, *9*, 291–297.

Sligar, S.G. (1976). Coupling of spin, substrate, and redox equilibria in cytochrome P450. *Biochemistry*, *15*, 5399–5406.

Stark, K.L., Harris, C., Juchau, M.R. (1989). Influence of electrophilic character and glutathione depletion on chemical dysmorphogenesis in cultured rat embryos. *Biochemical Pharmacology*, *38*, 2685–2692.

Stiborova, M., Frei, E., Wiessler, M., Schmeiser, H.H. (2001). Human enzymes involved in the metabolic activation of carcinogenic aristolochic acids: evidence for reductive activation by cytochromes P450 1A1 and 1A2. *Chemical Research in Toxicology*, *14*, 1128–1137.

Swanson, R.A., Dus, K.M. (1979). Specific covalent labeling of cytochrome P-450CAM with 1-(4-azidophenyl)imidazole, an inhibitor-derived photoaffinity probe for P-450 heme proteins. *Journal of Biological Chemistry*, *254*, 7238–7246.

Tantcheva-Poor, I., Servera-Llaneras, M., Scharffetter-Kochanek, K., Fuhr, U. (2001). Liver cytochrome P450 CYP1A2 is markedly inhibited by systemic but not by bath PUVA in dermatological patients. *Br. J. Dermatol.*, *144*, 1127–1132.

Tessier, S., Boivin, S., Aubin, J., Lampron, P., Detheux, M., Fournier, A. (2005). Transmembrane domain V of the endothelin-A receptor is a binding domain of ETA-selective TTA-386-derived photoprobes. *Biochemistry*, *44*, 7844–7854.

Thummel, K.E., Wilkinson, G.R. (1998). In vitro and in vivo drug interactions involving human CYP3A. *Annual Review Pharmacology and Toxicology*, *38*, 389–430.

Trager, W.F. (1982). The postenzymatic chemistry of activated oxygen. *Drug Metabolism Reviews*, *13*, 51–69.

Tsai, P.C., Huang, W., Lee, Y.C., Chan, S.H., Guo, Y.L. (2006). Genetic polymorphisms in CYP1A1 and GSTM1 predispose humans to PCBs/PCDFs-induced skin lesions. *Chemosphere*, *63*, 1410–1418.

Van, L.M., Swales, J., Hammond, C., Wilson, C., Hargreaves, J.A., Rostami-Hodjegan, A. (2007). Kinetics of the time-dependent inactivation of CYP2D6 in cryopreserved human hepatocytes by methylenedioxymethamphetamine (MDMA). *European Journal of Pharmaceutical Sciences*, *31*, 53–61.

Venkatakrishnan, K., Obach, R.S. (2005). In vitro-in vivo extrapolation of CYP2D6 inactivation by paroxetine: prediction of nonstationary pharmacokinetics and drug interaction magnitude. *Drug Metabolism and Disposition*, *33*, 845–852.

Vermilion, J.L., Coon, M.J. (1978). Identification of the high and low potential flavins of liver microsomal NADPH-cytochrome P450 reductase. *The Journal of Biological Chemistry*, *253*, 8812–8819.

Voorman, R.L., Maio, S.M., Payne, N.A., Zhao, Z., Koeplinger, K.A., Wang, X. (1998). Microsomal metabolism of delavirdine: evidence for mechanism-based inactivation of human cytochrome P450 3A. *The Journal of Pharmacology and Experimental Therapeutics*, *287*, 381–388.

Walsh, C.T. (1984). Suicide substrates, mechanism-based enzyme inactivators: recent developments. *Annual Review of Biochemistry*, *53*, 493–535.

Wang, E., Walsh, C. (1978). Suicide substrates for the alanine racemase of Escherichia coli B. *Biochemistry*, *17*, 1313–1321.

Wang, J., Bauman, S., Colman, R.F. (1998). Photoaffinity labeling of rat liver glutathione S-transferase, 4-4, by glutathionyl S-[4-(succinimidyl)-benzophenone]. *Biochemistry*, *37*, 15671–15679.

Wang, Y.H., Jones, D.R., Hall, S.D. (2004). Prediction of cytochrome P450 3A inhibition by verapamil enantiomers and their metabolites. *Drug Metabolism and Disposition*, *32*, 259–266.

Wang, Y.H., Jones, D.R., Hall, S.D. (2005). Differential mechanism-based inhibition of CYP3A4 and CYP3A5 by verapamil. *Drug Metabolism and Disposition: The Biological Fate of Chemicals*, *33*, 664–671.

Watkins, P.B., Whitcomb, R.W. (1998). Hepatic dysfunction associated with troglitazone. *N. Engl. J. Med.*, *338*, 916–917.

Wen, B., Doneanu, C.E., Gartner, C.A., Roberts, A.G., Atkins, W.M., Nelson, S.D. (2005). Fluorescent photoaffinity labeling of cytochrome P450 3A4 by lapachenole: identification of modification sites by mass spectrometry. *Biochemistry*, *44*, 1833–1845.

Wen, B., Lampe, J.N., Roberts, A.G., Atkins, W.M., David Rodrigues, A., Nelson, S.D. (2006). Cysteine 98 in CYP3A4 contributes to conformational integrity required for P450 interaction with CYP reductase. *Archives of Biochemistry and Biophysics*, *454*, 42–54.

Wester, M.R., Yano, J.K., Schoch, G.A., Yang, C., Griffin, K.J., Stout, C.D., Johnson, E.F. (2004). The structure of human cytochrome P450 2C9 complexed with flurbiprofen at 2.0-A resolution. *Journal of Biological Chemistry*, *279*, 35630–35637.

von Weymarn, L.B., Zhang, Q.Y., Ding, X., Hollenberg, P.F. (2005). Effects of 8-methoxypsoralen on cytochrome P450 2A13. *Carcinogenesis*, *26*, 621–629.

White, R.E., Coon, M.J. (1982). Heme ligand replacement reactions of cytochrome P-450. Characterization of the bonding atom of the axial ligand trans to thiolate as oxygen. *The Journal of Biological Chemistry*, *257*, 3073–3083.

Wienkers, L.C., Heath, T.G. (2005). Predicting in vivo drug interactions from in vitro drug discovery data. *Nature Reviews*, *4*, 825–833.

Williams, D.P., Antoine, D.J., Butler, P.J., Jones, R., Randle, L., Payne, A., Howard, M., Gardner, I., Blagg, J., Park, B.K. (2007). The metabolism and toxicity of furosemide in the Wistar Rat and CD-1 mouse: a chemical and biochemical definition of the toxicophore. *Journal of Pharmacology and Experimental Therapeutics*, *322*, 1208–1220.

Williams, P.A., Cosme, J., Vinkovic, D.M., Ward, A., Angove, H.C., Day, P.J., Vonrhein, C., Tickle, I.J., Jhoti, H. (2004). Crystal structures of human cytochrome P450 3A4 bound to metyrapone and progesterone. *Science*, *305*, 683–686.

Wolfenden, R., Snider, M.J. (2001). The depth of chemical time and the power of enzymes as catalysts. *Accounts of Chemical Research*, *34*, 938–945.

Yamazaki, H., Shibata, A., Suzuki, M., Nakajima, M., Shimada, N., Guengerich, F.P., Yokoi, T. (1999). Oxidation of troglitazone to a quinone-type metabolite catalyzed by cytochrome P-450 2C8 and P-450 3A4 in human liver microsomes. *Drug Metabolism and Disposition*, *27*, 1260–1266.

Yamazaki, H., Shimada, T. (1998). Comparative studies of in vitro inhibition of cytochrome P450 3A4-dependent testosterone 6beta-hydroxylation by roxithromycin and its metabolites, troleandomycin, and erythromycin. *Drug Metabolism and Disposition: The Biological Fate of Chemicals*, *26*, 1053–1057.

Yang, J., Jamei, M., Yeo, K.R., Tucker, G.T., Rostami-Hodjegan, A. (2005). Kinetic values for mechanism-based enzyme inhibition: assessing the bias introduced by the conventional experimental protocol. *European Journal of Pharmaceutical Sciences*, *26*, 334–340.

Yano, J.K., Hsu, M.H., Griffin, K.J., Stout, C.D., Johnson, E.F. (2005). Structures of human microsomal cytochrome P450 2A6 complexed with coumarin and methoxsalen. *Nature Structural & Molecular Biology*, *12*, 822–823.

Yano, J.K., Wester, M.R., Schoch, G.A., Griffin, K.J., Stout, C.D., Johnson, E.F. (2004). The structure of human microsomal cytochrome P450 3A4 determined by X-ray crystallography to 2.05-A resolution. *Journal of Biological Chemistry*, *279*, 38091–38094.

Zhang, Z., Chen, Q., Li, Y., Doss, G.A., Dean, B.J., Ngui, J.S., Silva Elipe, M., Kim, S., Wu, J.Y., Dininno, F., Hammond, M.L., Stearns, R.A., Evans, D.C., Baillie, T.A., Tang, W. (2005). In vitro bioactivation of dihydrobenzoxathiin selective estrogen receptor modulators by cytochrome P450 3A4 in human liver microsomes: formation of reactive iminium and quinone type metabolites. *Chemical Research in Toxicology*, *18*, 675–685.

Zhao, X.J., Jones, D.R., Wang, Y.H., Grimm, S.W., Hall, S.D. (2002). Reversible and irreversible inhibition of CYP3A enzymes by tamoxifen and metabolites. *Xenobiotica; the Fate of Foreign Compounds in Biological Systems*, *32*, 863–878.

Zheng, J., Ma, L., Xin, B., Olah, T., Humphreys, W.G., Zhu, M. (2007). Screening and identification of GSH-trapped reactive metabolites using hybrid triple quadruple linear ion trap mass spectrometry. *Chemical Research in Toxicology*, *20*, 757–766.

Zhou, S., Yung Chan, S., Cher Goh, B., Chan, E., Duan, W., Huang, M., McLeod, H.L. (2005). Mechanism-based inhibition of cytochrome P450 3A4 by therapeutic drugs. *Clinical Pharmacokinetics*, *44*, 279–304.

# Clinical Drug–Drug Interactions

SUJAL V. DESHMUKH

## 1 INTRODUCTION

In the current health-care system, patients are typically on multiple medications for the treatment of one or more medical conditions. It is thus plausible that drugs may interact with each other in the human body. When these *in vivo* interactions result in a significant change in the pharmacokinetic or pharmacodynamic properties of at least one of the drugs, it is considered a clinical drug–drug interaction (DDI). Pharmacokinetic interactions are often multifactoral; these involve alteration of drug metabolism, drug transporter functions, percent of drug bound to plasma proteins, and drug excretion. Pharmacokinetic interactions that lead to adverse reaction or lack of efficacy are extremely detrimental to the patients and constitute as a major concern to the health-care providers. Pharmacodynamic interactions are those where the drug levels at the site of action are unaffected but its effects are altered; these may be synergistic, additive, or antagonistic in nature. Given the scope of this book, the following chapter will focus on the pharmacokinetic-based DDIs.

Recognizing and remembering pharmacokinetic-based DDIs are a daily challenge for many clinicians. It is thus important that drugmakers identify potential interactions that may occur when a drug is coadministered to humans with other medications. Several *in vitro* studies conducted at the drug discovery/development stages can alert against the possible risk for drug interaction and can help designing appropriate clinical studies to fully address the potentials of pharmacokinetic-based DDI. Early clinical studies directed toward studying potential DDIs are an integral and critical component of the drug discovery/development processes. Although the occurrence of a DDI does not preclude the use of a drug, it alerts the medical practitioner against the potential for DDI when used in combination with other drugs.

The factors that influence DDIs can be categorized as patient-related and drug-related. The patient-related factors are age, sex, disease condition, diet/nutrition, genetics, and environment. The drug-related factors include dose, dosing frequency, pharmacokinetic parameters like half-life, volume of distribution, oral availability, route of elimination, and interaction mechanism (Herman, 1999) (Fig. 1).

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

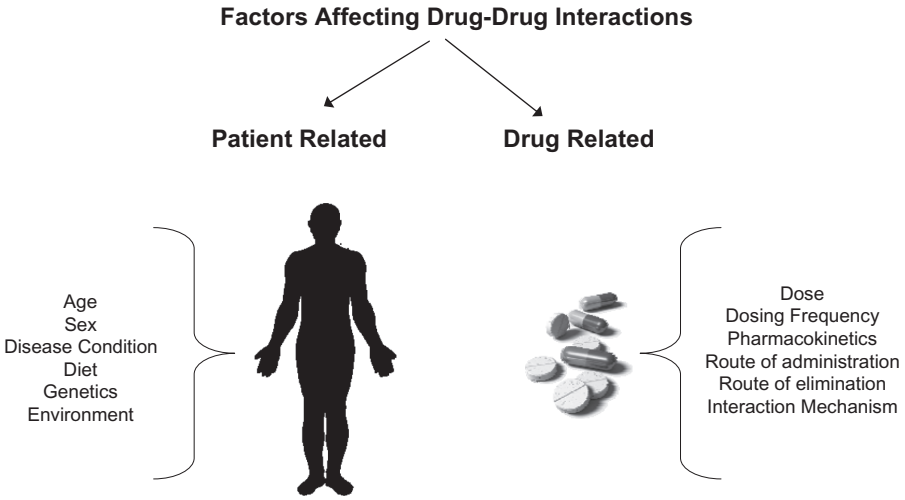


**Figure 1** Patient and drug-related factors affecting DDIs.

When a DDI occurs, the drug causing the interaction is commonly referred to as the perpetrator and the drug that undergoes a change is the victim drug. Typically, a perpetrator is a potent inhibitor or inducer of drug-metabolizing enzymes and/or drug transporters, whereas a victim is an enzyme or transporter substrate.

Most commonly observed clinical DDIs are due to changes in the metabolism of the drugs. At the center of these DDIs are the phase I enzymes, predominantly the cytochrome P450s (CYPs). Among the various CYPs, CYP3A4 is responsible for metabolism of approximately 50% of the marketed drugs (Williams *et al.*, 2004). Drugs that cause inhibition or induction of CYP3A4 are most likely to cause DDIs. Additionally, drugs that are solely dependent on CYP3A4 for their elimination have a likelihood of being significantly affected by CYP3A4 inducers or inhibitors. Table 1 summarizes the marketed drugs as substrates, inhibitors, or inducers of CYPs. The second most common mechanism of DDIs is inhibition of drug transporters such as multidrug resistance transporter (MDR1)/P-glycoprotein (P-gp), organic anion-transporting polypeptide (OATP), breast cancer-resistant protein (BCRP), and organic anion transporter (OAT), which are found in the gut, liver, kidney, and brain (Ayrton and Morgan, 2001; Ho and Kim, 2005; Marchetti *et al.*, 2007). Table 2 provides a summary of various transporter-mediated DDIs.

The liver is a major organ for drug metabolism and thus DDIs predominantly occur in this organ. Furthermore, the hepatobiliary system has several transporters responsible for excreting drugs into the bile (Shitara, Sato and Sugiyama, 2005). Several drugs can compete for these transporters for biliary excretion, leading to DDIs. For orally administered drugs, the intestinal drug-metabolizing enzymes and transporters are sites for DDIs. Another major organ where DDIs commonly occur is the kidney (Fig. 2).

Although DDIs are generally associated with undesirable outcomes and in some cases lead to detrimental adverse events, there may be some benefits of DDIs as well. The modulation of drug transporters or enzymes is a strategy that has been explored to increase drug exposure. Inhibition of P-gp can be used to increase oral bioavailability and tumor levels of some anticancer drugs; for example, valspodar in combination with paclitaxel is being tested in clinical trails (Chico *et al.*, 2001).

**TABLE 1 Substrates, inhibitors and inducers of CYPs.**

| CYP isoform | Substrates | Inhibitors | Inducers |
|---|---|---|---|
| CYP1A2 | Caffeine, clozapine, imipramine, phenacetin, propranolol, *R*-warfarin, theophylline | Cimetidine, ciprofloxacin, fluvoxamine, furafylline, grepafloxacin | Barbiturates, omeprazole, phenytoin |
| CYP2C9 | Diclofenac, dofetilide, fluvastatin, ibuprofen, mefanamic acid, naproxen, phenytoin, piroxicam, *S*-warfarin, tobutamide | Fluconazole, fluoxetine, fluvoxamine, ritonavir | Barbiturates, rifampicin |
| CYP2C19 | Clomipramine, diazepam, hexobarbital, imipramine, mephobarbital, omeprazole, phenytoin, propranolol, *S*-mephenytoin | Fluoxetine, fluvoxamine, omeprazole | |
| CYP2D6 | Amitriptyline, amphetamine, clomipramine, codeine, desipramine, dextromethorphran, dihydrocodeine, diphenylhydramine, fluoxetine, haloperidol, hydrocodone, imipramine, metoprolol, nortriptyline, ondensatron, oxycodone, paroxetine, penbutolol, perphenazine, propranolol, thioridazine, timolol, trimiramine | Fluoxetine, haloperidol, paroxetine, quinidine, ritonavir, sertraline, thioridazine | |
| CYP3A4 | Amiodarone, amitriptyline, alprazolam, astemizole, carbamazepine, cyclosporine, cisapride, clindamycin, clomipramine, clonazepam, dapsone, dexamethasone, dextromethorphan, diazepam, diltiazem, erythromycin, felodipine, hydrocortisone, imipramine, indinavir, lidocaine, lovastatin, midazolam, nefazoldone, nelifinavir, nevirapine, nifedapine, nimodipine, propafenone, quinidine, *R*-warfarin, ritonavir, saquinavir, sertraline, simvastatin, tamoxifen, terfenadine, triazolam, verapamil, zolpidem | Cimetidine, clarithromycin, erythromycin, fluvoxamine, itraconazole, ketoconazole, miconazole, nefazodone, nelfinavir, remacemide, ritonavir | Barbiturates, carbamazepine, dexamethasone, phenytoin, rifabutin, rifampicin |

From FDA (2006) and Indiana University School of Medicine (2008).

## 2 MECHANISMS OF DDIs

In order to find ways to mitigate against DDIs, it is important to understand the mechanism(s) by which these interactions occur. Drugs may interact with each other by one or more of the following mechanisms.

### 2.1 Interactions Affecting Oral Absorption

For orally administered drugs, if one of the drugs has an effect on the gut absorption, it may result in alteration of drug levels. The interaction may cause changes in either the rate of absorption or the amount of drug absorbed. The former affect the maximum drug concentration ($C_{max}$) and time to reach maximum drug concentration ($t_{max}$) but may not have an effect on the total drug absorbed, while the latter may affect both the

**TABLE 2 Transporter-mediated DDIs.**

| Transporter | Victim | Perpetrator | Mechanism | Clinical outcome | Reference |
|---|---|---|---|---|---|
| MDR1 | Digoxin | Quinidine | Inhibition of MDR1 | Decreased renal clearance, increase in plasma levels | Doering (1979) |
| MDR1 | Digoxin | Ritonavir | Inhibition of MDR1 | Increased plasma levels and terminal half-life | Phillips, Rachlis and Ito (2003) |
| MDR1 | Paclitaxel | Cyclosporin A | Inhibition of MDR1 | Increased bioavailability | Meerum Terwogt *et al.* (1998) |
| MDR1 | Talinolol | Rifampin | Induction of MDR1 | Lowered plasma AUC | Westphal *et al.* (2000) |
| MDR1 | Fexofenadine | Rifampin | Induction of MDR1 | Decreased plasma levels | Hamman *et al.* (2001) |
| BCRP | Methotrexate | Omeprazole | Inhibition of BCRP | Increased plasma AUC and low clearance | Reid *et al.* (1993); Breedveld *et al.* (2004) |
| OATP C | Rosuvastatin | Gemfibrozil | Inhibition of OATP | Increased $C_{max}$ and AUC | Schneck *et al.* (2004) |
| OAT | Cephalosporin (cefadroxil) | Probenecid | Inhibition of OAT | Decreased renal clearance, increase in $C_{max}$ and half-life | Somogyi (1996) |
| OAT | ACE inhibitors (enalapril, lisinopril) | Probenecid | Inhibition of OAT | Decreased renal clearance, prolonged half-life | Somogyi (1996) |
| OAT, OCT, OATP | Procainamide, levofloxacin, dofetilide | Cimetidine | Inhibition of OAT, OCT, OATP | Decreased renal clearance, increased AUC | Christian, Meredith and Speeg (1984) |
| OAT3 | Methotrexate | NSAIDs | Inhibition of OAT3 | Decreased renal clearance | Takeda *et al.* (2002) |
| BSEP | Digoxin | Cyclosporine | Inhibition of BSEP | Causes cholestasis, digoxin toxicity | Dorian *et al.* (1988), Stieger *et al.* (2000) |

$C_{max}$ and the exposure (area under curve, AUC). Most oral DDIs that alter absorption result in decreased drug levels. Examples of DDI due to effects on oral absorption are presented in Table 3.

***2.1.1 Changes in Gastrointestinal pH*** Absorption of drug from the gastrointestinal tract is dependent on the pH of the gut, pKa of the drug, and the drugs' physiochemical properties. Changes in pH due to drugs such as antacids, or proton pump inhibitors can affect systemic levels of drugs whose absorption is pH-dependent. Omeprazole, a proton pump inhibitor, which causes a reduction in gastric acid secretion, increases the exposure of saquinavir (Winston *et al.*, 2006).

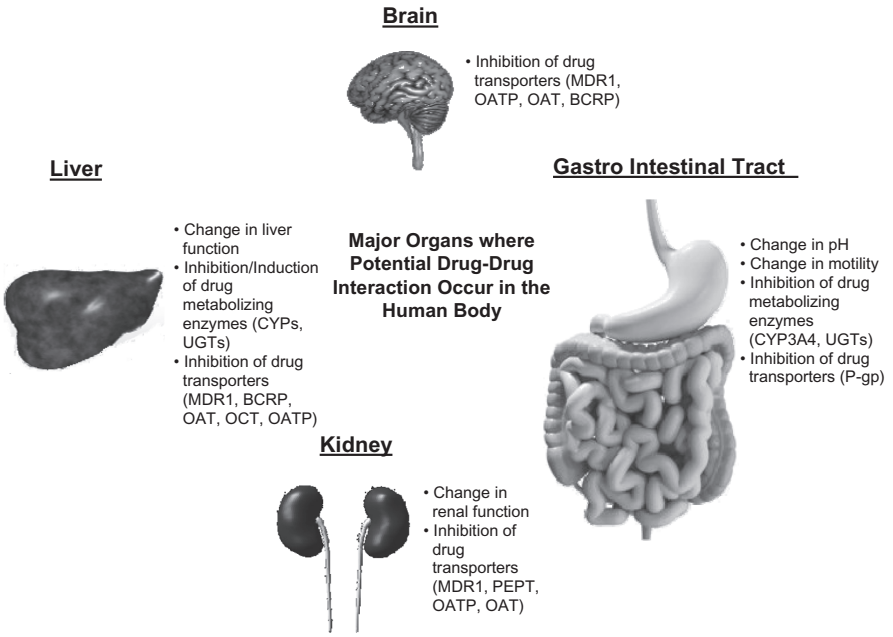


**Figure 2** Major DDI sites in the human body.

***2.1.2 Change in Gastric Motility*** Gastric emptying time is an important factor that affects drug absorption; thus, drugs that alter gastric emptying will also affect drugs that are predominantly absorbed from the upper small intestine. Metoclopramide is a classical example in this category since it is a gastrointestinal prokinetic drug that increases the rate of gastric emptying, causing a reduction in the $t_{max}$ and an increase in the $C_{max}$ of several drugs (Johnson *et al.*, 1984; Greiff and Rowbotham, 1994).

***2.1.3 Complex Formation/Chelation*** Drugs that can potentially form a complex with other drugs can affect absorption. Activated charcoal is used to treat drug overdose. Charcoal can adsorb drugs and prevent them from being absorbed. Tetracycline antibiotics form complexes with metals such as calcium, aluminum, and bismuth, thus resulting in a reduction in their oral absorption (Tatro, 1972).

***2.1.4 Malabsorption*** Certain drugs may cause a malabsorption syndrome that results in decreased absorption from the gastrointestinal tract. Neomycin can cause this syndrome and is known to decrease the absorption of penicillin V and digoxin (Rothfeld and Osborne, 1963).

***2.1.5 Effect on Gut Metabolizing Enzymes*** Metabolism of drugs by gut enzymes is a significant component of first-pass effect for some orally absorbed drugs. CYP3A4 has been identified as one of the major enzymes in human enterocytes (Zhang *et al.*, 1999). Coadministration of potent CYP3A4 inhibitors can increase bioavailability of drugs which undergo substantial CYP3A4-mediated first-pass metabolism in the gut.

**TABLE 3 Clinical DDIs affecting oral absorption.**

| Mechanism | Victim | Perpetrator | Effect | Clinical outcome | Reference |
|---|---|---|---|---|---|
| Change in gastric pH | Indinavir | Didanosine | Impaired indinavir absorption | Low plasma indinavir levels leading to failure of treatment | Bristol-Myers Squibb Company (2001) |
| | Saquinavir | Omeprazole | Increased oral exposure | No toxicity observed | Winston *et al.* (2006) |
| | Atazanavir | Omeprazole/ famotidine | Impaired atazanavir absorption | Decrease in plasma levels not significantly affecting treatment | Bristol-Myers Squibb Company (2008) |
| Complex formation/ chelation | Quinolone antibiotic (fluoroquinolones) | Antacids | Reduction in quinolone oral exposure | Reduced antimicrobial activity | Polk (1989) |
| | Tetracyclines | Antacids (containing Al3+, Ca2+, Mg2+, Bi2+, Zn2+, Fe2+) | Reduced antibiotic exposure | Reduction in antibiotic activity | Tatro (1972) |
| | Penicillamine | Antacids | Reduced absorption of penicillamine | Decreased therapeutic effects of penicillamine | Osman *et al.* (1983) |
| Change in gut motility | Digoxin | Metoclopramide | Reduced absorption | No effect of digoxin | Johnson *et al.* (1984) |

It is, however, difficult to demonstrate this mechanism in the clinic since CYP-mediated gut metabolism cannot be isolated from metabolism that occurs in the liver.

***2.1.6 Effect on Drug Transporters in the Gut*** Enterocytes are known to harbor drug transporters such as MDR1, multidrug resistance-associated protein (MRP)1/2/3, BCRP, and OATP2, which actively uptake or efflux drugs (Ho and Kim, 2005). Potent inhibitors of these transporters can thus affect oral bioavailability of drugs which are substrates for these transporters.

### 2.2 Interaction Affecting Hepatic Metabolism

The liver is a major drug-metabolizing organ responsible for the elimination of drugs. Several drug-metabolizing enzymes of the liver transform drugs into more polar entities that can be readily excreted from the body. Alteration in the levels of these enzymes or their function will change the metabolic profile of drugs and thus will affect

their blood levels. In addition to the liver's role as a drug-metabolizing organ, it also excretes drugs via the biliary system.

***2.2.1 Enzyme Inhibition*** Inhibition of drug-metabolizing enzymes is one of the most common mechanisms observed in clinical DDIs. Drugs that are potent inhibitor of drug-metabolizing enzymes cause severe DDIs. CYPs are responsible for the metabolism of approximately 70% of the drugs on the market, of this, CYP3A4 is involved in the metabolism of about 50% of the drugs followed by CYP2C9 and CYP2D6, each metabolizing about 20% drugs (Williams *et al.*, 2004). Several *in vitro* approaches can be used to demonstrate the inhibitory effect of a drug and based on this information, predictions can be made regarding the potential for clinical DDIs. There are two distinct mechanisms by which drugs inhibit CYP enzymes: (i) reversible and (ii) irreversible (time-dependent/mechanism-based).

Reversible inhibitors compete with the victim drug at the site of the enzyme and prevent it from being metabolized. There are several drugs in the clinic that are reversible inhibitors of CYPs. Drugs such as ketoconazole, quinidine, omeprazole, and methadone are some of the examples of reversible CYP inhibitors (Table 1). Interactions due to reversible inhibitors are clinically manageable.

On the other hand, irreversible/time-dependent inhibition has emerged as a greater concern in the clinic. These inhibitors cause inactivation of the enzyme due to formation of reactive species that irreversibly bind to the enzyme(s). The effect of these inhibitors persists long after the inhibitor is completely eliminated from the body. The rate of enzyme synthesis determines when the CYP activity is restored. Examples of irreversible inhibitors known to cause clinical DDIs include clarithromycin, fluoxetine, and ritonavir (Zhou *et al.*, 2004).

***2.2.2 Enzyme Induction*** Certain drugs can increase levels of drug-metabolizing enzymes by causing an induction in the expression of the gene associated with the respective enzyme. Several drugs have been shown to bind to response elements in the promoter region of genes and increase the expression of one or more enzymes depending on the transcription factor involved. Some of the nuclear receptors involved in expression of CYP enzymes are aryl hydrocarbon receptor (AHR), constitutive androstane receptor (CAR), and pregnane X receptor (PXR) (Lin, 2006). The extent to which enzyme induction occurs is generally dependent on the dose of the drug. DDIs arising from enzyme induction take a few days to a few weeks to develop fully. Since the effects originate at the molecular level, they also persist for a similar time after the drug is discontinued. Interaction resulting from enzyme induction can be managed by increasing the dose of the victim drug, but this has to be closely monitored so as to avoid toxicity due to overdose. Rifampin is a classical example of an enzyme inducer of hepatic and intestinal CYPs, and is often used as an *in vitro* and *in vivo* probe to investigate DDIs (Finch *et al.*, 2002).

***2.2.3 Alteration of Hepatic Blood Flow*** Hepatic clearance is dependent on hepatic blood flow and is an important factor for drugs with a high hepatic extraction ratio. Changes in hepatic blood flow will have an effect on the hepatic clearance. Cimetidine is known to decreases hepatic blood flow, it has been observed that it increases the bioavailability of propranolol via its effect on blood flow (Baciewicz and Baciewicz, 1989). Another example is that of verapamil and metoprolol which have been shown to affect hepatic blood flow and thus cause DDI (Bauer *et al.*, 2000).

***2.2.4 Interactions Affecting Hepatobiliary Excretion*** Drugs are excreted in the bile either unchanged or as conjugated metabolites. Transporters located in the hepatocytes around the bile canaliculus can actively excrete drugs into the bile. Transporters identified so far include BCRP, MDR1/3, and MRP (Shitara, Sato and Sugiyama, 2005). Drugs can inhibit such transporters and affect elimination of drugs that are excreted by these transporters, causing an increase in the levels of the drug or its metabolite(s).

## 2.3 Interactions Due to Changes in Renal Excretion

Several drugs are excreted unchanged through the kidney and eliminated from the body along with urine. Some drugs are eliminated from the glomerular membrane via passive diffusion and filtered into the lumen of the tubules. In the tubules, active energy transport is responsible for excretion of some drugs. Active and passive transport in tubule cells can also play a role in the reabsorption of drugs back into the blood. The interplay between drug excretion and reabsorption determines the fate of a drug in the body. If any one of these processes is affected, it will directly affect drug exposure.

***2.3.1 Changes in Urine pH*** Similar to oral drug absorption, reabsorption of drug from the urine depends on the ionization of the drugs. The non-ionized form of the drug undergoes passive diffusion back into the blood. This diffusion process is dependent on the pKa of the drug and the pH of the urine. In a highly alkaline pH, weakly acidic drugs exist predominantly in the ionized form and will thus be poorly absorbed from the urine. Conversely, weakly basic drugs are readily eliminated in the urine at acidic pH. Altering the pH of urine is a commonly used practice in treatment of overdose of drugs mainly eliminated via kidney, as it can facilitate the elimination of drugs from the body.

***2.3.2 Changes in Active Tubule Excretion*** Active transport/secretion across the tubules is an important mechanism for drug elimination (Kosoglou and Vlasses, 1989; Shitara, Sato and Sugiyama, 2005). Transporters identified in the renal tubule include OAT1/3/4, MDR1, and MRP2 (Ho and Kim, 2005). Drugs that use the same transporters for their excretion will compete with each other. These interactions can affect their elimination from the body. A classical example is the interaction between penicillin and probenecid. The elimination of penicillin is affected by probenecid as they both compete for the same transporter, resulting in decreased elimination of penicillin (Barza and Weinstein, 1976). Thus, probenecid is used in conjunction to reduce the dose of penicillin.

***2.3.3 Changes in Renal Blood Flow*** Factors that regulate renal blood flow such as prostaglandins can be affected by drugs. These changes will cause altered renal blood flow, resulting in changes in the rate at which drugs are eliminated from the body. Indomethacin inhibits the synthesis of prostaglandins and reduces renal blood flow. As a result, elimination of drugs such as lithium is reduced, resulting in increased blood levels of lithium (Weinblatt, 1989).

## 2.4 Interactions Affecting Plasma Protein Binding

In blood, drugs bind to plasma proteins and exist in a state of equilibrium between the bound and unbound form. The unbound drug is responsible for the pharmacological

activity; hence, if levels of unbound drug increase, there may be enhanced therapeutic or toxicological effects. Drugs can bind to several plasma proteins such as alpha-1-glycoprotein, albumin, lipoprotein, and globulins. If two drugs compete to bind to the same protein, one drug could displace the other drug, increasing its free fraction. For tightly bound drugs, small changes in the binding cause a large change in the unbound fraction. For example, a drug that is 99.9% bound (i.e. 0.1% unbound) is displaced by another drug such that it is now 99% bound. The unbound fraction increases to 1%, that is, a change of 10-fold. However, this change in unbound fraction affects drugs with low distribution volume where the drugs remain in the plasma compartment rather than distributing into the tissue. In 1960s, it was reported that plasma protein binding may be responsible for sulfaphenazole-induced hypoglycemia in tolbutamide-treated patients (Christensen, Hansen and Kristensen, 1963) and potentiation of anti-coagulants by pyrazole compounds (Fox, 1964). However, the scientific rationale for the involvement of plasma protein binding has been challenged. Several reports have presented theoretical arguments about the clinical relevance of plasma protein binding on the exposure of patients to a drug (Sellers, 1979; Rolan, 1994; Benet and Hoener, 2002). Clinical effects attributed to plasma protein binding are not common and this phenomenon is not as profound as it was previously thought.

## 3 AVOIDING DDI

There are several different approaches that the medical practitioners can use to avoid DDIs. They involve use of resources that list DDIs, basic knowledge of pharmacology and pharmacokinetics, clinical judgment, and problem-solving skills. Information on clinical DDIs can be obtained from well-established drug interaction studies or new clinical data as they emerge. The Food and Drug Administration (FDA) published its first guidance on DDIs in 1997 (FDA, 1997), followed by a second guidance in 1999 (FDA, 1999). These included best practices for *in vitro* and *in vivo* drug interaction studies, including study design, dosing, statistical analysis, and guidance for appropriate labeling. *In vivo* and *in vitro* studies are mainly performed for metabolism and transporter-based DDIs. Investigational drugs are studied for their potential to inhibit or induce CYP enzymes or P-gp, or to know if drugs are substrates for CYPs or P-gp or any other transporter. Research in pharmaceutical industry has focused on screening compounds for possible DDI liability based on the limited *in vitro–in vivo* correlation models currently available. Models that can accurately predict clinical outcomes are currently lacking. Additional research targeted toward refining these models to improve clinical predictability is an impending need. In summary, understanding the mechanism of DDIs and the interplay of various factors affecting these interactions can assist in managing the risk associated with these DDIs so as to minimize undesirable clinical outcomes.

## REFERENCES

Ayrton, A., Morgan, P. (2001). Role of transport proteins in drug absorption, distribution and excretion. *Xenobiotica*, *31*, 469–497.

Baciewicz, A.M., Baciewicz, F.A., Jr (1989). Effect of cimetidine and ranitidine on cardiovascular drugs. *Am. Heart J.*, *118*, 144–154.

Barza, M., Weinstein, L. (1976). Pharmacokinetics of the penicillins in man. *Clin. Pharmacokinet.*, *1*, 297–308.

Bauer, L.A., Horn, J.R., Maxon, M.S., Easterling, T.R., Shen, D.D., Strandness, D.E., Jr (2000). Effect of metoprolol and verapamil administered separately and concurrently after single doses on liver blood flow and drug disposition. *J. Clin. Pharmacol.*, *40*, 533–543.

Benet, L.Z., Hoener, B.A. (2002). Changes in plasma protein binding have little clinical relevance. *Clin. Pharmacol. Ther.*, *71*, 115–121.

Breedveld, P., Zelcer, N., Pluim, D., Sonmezer, O., Tibben, M.M., Beijnen, J.H., Schinkel, A.H., van Tellingen, O., Borst, P., Schellens, J.H. (2004). Mechanism of the pharmacokinetic interaction between methotrexate and benzimidazoles: potential role for breast cancer resistance protein in clinical drug-drug interactions. *Cancer Res.*, *64*, 5804–5811.

Bristol-Myers Squibb Company (2001). *Didanosine Information*, http://www.fda.gov/cder/foi/label/2001/20155s25lbl.pdf.

Bristol-Myers Squibb Company (2008). *Highlights of Prescribing Information for Atazanavir Sulfate*, http://packageinserts.bms.com/pi/pi_reyataz.pdf.

Chico, I., Kang, M.H., Bergan, R., Abraham, J., Bakke, S., Meadows, B., Rutt, A., Robey, R., Choyke, P., Merino, M., Goldspiel, B., Smith, T., Steinberg, S., Figg, W.D., Fojo, T., Bates, S. (2001). Phase I study of infusional paclitaxel in combination with the P-glycoprotein antagonist PSC 833. *J. Clin. Oncol.*, 19, 832–842.

Christensen, L.K., Hansen, J.M., Kristensen, M. (1963). Sulphaphenazole-induced hypoglycaemic attacks in tolbutamide-treated diabetics. *Lancet*, *2*, 1298–1301.

Christian, C.D., Jr, Meredith, C.G., Speeg, K.V., Jr (1984). Cimetidine inhibits renal procainamide clearance. *Clin. Pharmacol. Ther.*, *36*, 221–227.

Doering, W. (1979). Quinidine-digoxin interaction: pharmacokinetics, underlying mechanism and clinical implications. *N. Engl. J. Med.*, *301*, 400–404.

Dorian, P., Strauss, M., Cardella, C., David, T., East, S., Ogilvie, R. (1988). Digoxincyclosporine interaction: severe digitalis toxicity after cyclosporine treatment. *Clin. Invest. Med.*, *11*, 108–112.

FDA (1997). *Guidance for Industry: Drug Metabolism/Drug Interactions in the Drug Development Process: Studies In Vitro*, http://www.fda.gov/Cder/guidance/clin3.pdf (accessed April 1997).

FDA (1999). *Guidance for Industry: In Vivo Metabolism/Drug Interactions: Study Design, Data Analysis and Recommendation for Dosing and Labeling*, http://www.fda.gov/Cder/guidance/2635fnl.pdf (accessed December 1999).

FDA (2006). *Table of Substrate, Inhibitors and Inducers*, http://www.fda.gov/cder/drug/drugInteractions/tableSubstrates.htm.

Finch, C.K., Chrisman, C.R., Baciewicz, A.M., Self, T.H. (2002). Rifampin and rifabutin drug interactions: an update. *Arch. Intern. Med.*, *162*, 985–992.

Fox, S.L. (1964). Potentiation of anticoagulants caused by pyrazole compounds. *JAMA*, *188*, 320–321.

Greiff, J.M., Rowbotham, D. (1994). Pharmacokinetic drug interactions with gastrointestinal motility modifying agents. *Clin. Pharmacokinet.*, *27*, 447–461.

Hamman, M.A., Bruce, M.A., Haehner-Daniels, B.D., Hall, S.D. (2001). The effect of rifampin administration on the disposition of fexofenadine. *Clin. Pharmacol. Ther.*, *69*, 114–121.

Herman, R.J. (1999). Drug interactions and the statins. *CMAJ*, *161*, 1281–1286.

Ho, R.H., Kim, R.B. (2005). Transporters and drug therapy: implications for drug disposition and disease. *Clin. Pharmacol. Ther.*, *78*, 260–277.

Indiana University School of Medicine (2008). *Cytochrome P450 Drug Interaction Table*, http://medicine.iupui.edu/flockhart/clinlist.htm.

Johnson, B.F., Bustrack, J.A., Urbach, D.R., Hull, J.H., Marwaha, R. (1984). Effect of metoclopramide on digoxin absorption from tablets and capsules. *Clin. Pharmacol. Ther.*, 36, 724–730.

Kosoglou, T., Vlasses, P.H. (1989). Drug interactions involving renal transport mechanisms: an overview. *DICP*, *23*, 116–122.

Lin, J.H. (2006). CYP induction-mediated drug interactions: in vitro assessment and clinical implications. *Pharm. Res.*, *23*, 1089–1116.

Marchetti, S., Mazzanti, R., Beijnen, J.H., Schellens, J.H. (2007). Concise review: clinical relevance of drug–drug and herb–drug interactions mediated by the ABC transporter ABCB1 (MDR1, P-glycoprotein). *Oncologist*, *12*, 927–941.

Meerum Terwogt, J.M., Beijnen, J.H., ten Bokkel Huinink, W.W., Rosing, H., Schellens, J.H. (1998). Co-administration of cyclosporin enables oral therapy with paclitaxel. *Lancet*, *352*, 285.

Osman, M.A., Patel, R.B., Schuna, A., Sundstrom, W.R., Welling, P.G. (1983). Reduction in oral penicillamine absorption by food, antacid, and ferrous sulfate. *Clin. Pharmacol. Ther.*, *33*, 465–470.

Phillips, E.J., Rachlis, A.R., Ito, S. (2003). Digoxin toxicity and ritonavir: a drug interaction mediated through p-glycoprotein? *AIDS*, *17*, 1577–1578.

Polk, R.E. (1989). Drug-drug interactions with ciprofloxacin and other fluoroquinolones. *Am. J. Med.*, *87*, 76S–81S.

Reid, T., Yuen, A., Catolico, M., Carlson, R.W. (1993). Impact of omeprazole on the plasma clearance of methotrexate. *Cancer Chemother. Pharmacol.*, *33*, 82–84.

Rolan, P.E. (1994). Plasma protein binding displacement interactions—why are they still regarded as clinically important? *Br. J. Clin. Pharmacol.*, *37*, 125–128.

Rothfeld, B., Osborne, D. (1963). Malabsorption syndrome produced by neomycin. *Am. J. Dig. Dis.*, *8*, 763–768.

Schneck, D.W., Birmingham, B.K., Zalikowski, J.A., Mitchell, P.D., Wang, Y., Martin, P.D., Lasseter, K.C., Brown, C.D., Windass, A.S., Raza, A. (2004). The effect of gemfibrozil on the pharmacokinetics of rosuvastatin. *Clin. Pharmacol. Ther.*, *75*, 455–463.

Sellers, E.M. (1979). Plasma protein displacement interactions are rarely of clinical significance. *Pharmacology*, *18*, 225–227.

Shitara, Y., Sato, H., Sugiyama, Y. (2005). Evaluation of drug-drug interaction in the hepatobiliary and renal transport of drugs. *Annu. Rev. Pharmacol. Toxicol.*, *45*, 689–723.

Somogyi, A. (1996). Renal transport of drugs: specificity and molecular mechanisms. *Clin. Exp. Pharmacol. Physiol.*, *23*, 986–989.

Stieger, B., Fattinger, K., Madon, J., Kullak-Ublick, G.A., Meier, P.J. (2000). Drug- and estrogen-induced cholestasis through inhibition of the hepatocellular bile salt export pump (Bsep) of rat liver. *Gastroenterology*, *118*, 422–430.

Takeda, M., Khamdang, S., Narikawa, S., Kimura, H., Hosoyamada, M., Cha, S.H., Sekine, T., Endou, H. (2002). Characterization of methotrexate transport and its drug interactions with human organic anion transporters. *J. Pharmacol. Exp. Ther.*, *302*, 666–671.

Tatro, D.S. (1972). Tetracycline-antacid interactions. *JAMA*, *220*, 586.

Weinblatt, M.E. (1989). Drug interactions with non steroidal anti-inflammatory drugs (NSAIDs). *Scand. J. Rheumatol. Suppl.*, *83*, 7–10.

Westphal, K., Weinbrenner, A., Zschiesche, M., Franke, G., Knoke, M., Oertel, R., Fritz, P., von Richter, O., Warzok, R., Hachenberg, T., Kauffmann, H.M., Schrenk, D., Terhaag, B., Kroemer, H.K., Siegmund, W. (2000). Induction of P-glycoprotein by rifampin increases intestinal secretion of talinolol in human beings: a new type of drug/drug interaction. *Clin. Pharmacol. Ther.*, *68*, 345–355.

Williams, J.A., Hyland, R., Jones, B.C., Smith, D.A., Hurst, S., Goosen, T.C., Peterkin, V., Koup, J. R., Ball, S.E. (2004). Drug-drug interactions for UDP-glucuronosyltransferase substrates: a

pharmacokinetic explanation for typically observed low exposure (AUCi/AUC) ratios. *Drug Metab. Dispos.*, *32*, 1201–1208.

Winston, A., Back, D., Fletcher, C., Robinson, L., Unsworth, J., Tolowinska, I., Schutz, M., Pozniak, A.L., Gazzard, B., Boffito, M. (2006). Effect of omeprazole on the pharmacokinetics of saquinavir-500 mg formulation with ritonavir in healthy male and female volunteers. *AIDS*, *20*, 1401–1406.

Zhang, Q.Y., Dunbar, D., Ostrowska, A., Zeisloft, S., Yang, J., Kaminsky, L.S. (1999). Characterization of human small intestinal cytochromes P-450. *Drug Metab. Dispos.*, *27*, 804–809.

Zhou, S., Chan, E., Lim, L.Y., Boelsterli, U.A., Li, S.C., Wang, J., Zhang, Q., Huang, M., Xu, A. (2004). Therapeutic drugs that behave as mechanism-based inhibitors of cytochrome P450 3A4. *Curr. Drug Metab.*, *5*, 415–442.

## FURTHER READING

Stockley, I.H. (1999). *Drug Interactions: A Source Book of Adverse Interactions, Their Mechanisms, Clinical Importance and Management*, 5th edn, Pharmaceutical Press, London.

Goodman, L.S. (2005). *Goodman & Gilman's the Pharmacological Basis of Therapeutics*, 11th edn, McGraw-Hill, New York.

# Pharmaceutical Excipients in Drug–Drug Interaction

CHUAN CHEN and ABU J.M. SADEQUE

## 1 INTRODUCTION

Drug formulations such as tablets, capsules, or injectable solutions consist of active pharmaceutical ingredients and other "inactive" ingredients that are added for many purposes. For example, surfactants are often used in formulations to solubilize water-insoluble drugs, and methylcellulose may be used to prepare drug suspensions or added to tablets as a disintegrating agent or binder. During the development of a given drug formulation, it is well understood that the interactions between the drug and excipients need to be carefully studied to ensure that the drug is stable in the formulation and that the solubility, as well as the dissolution properties in the case of orally administered drugs, is satisfactory (Kalasz and Antal, 2006).

Not all of the excipients in approved drug formulations are "inactive;" some are known to alter drug metabolism and/or transporter activities. Therefore, in the drug formulation development process, one also needs to consider whether the chosen excipients in drug formulations can alter systemic exposure of the drug itself or other potentially coadministered therapeutic agents due to interactions with drug-metabolizing enzymes and/or transporters, especially under situations where the excipients are not approved by the regulatory agencies, or if the amounts of the excipients used are higher than that used for previously approved drug formulations.

Interestingly, among the many formulation components found in marketed drug products (see FDA web site: http://www.fda.gov/cder/drug/iig/default.htm), most of the excipients that interact with drug biotransformation processes and/or transporters belong to a group known as solubilizing excipients, that is, additives used to solubilize drugs in oral and injectable dosage forms. Solubilizing excipients include *water-soluble organic solvents* such as polyethylene glycol 300 (PEG300), polyethylene glycol 400 (PEG400), ethanol, propylene glycol, glycerin, *N*-methyl-2-pyrrolidone, dimethylacetamide, and dimethylsulfoxide (DMSO); *nonionic surfactants* such as Cremophor EL, Cremophor RH40, Cremophor EH60, D-α-tocopherol polyethylene glycol

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

1000 succinate (TPGS1000), polysorbate 20 (Tween 20), polysorbate 80 (Tween 80), Solutol HS15, sorbitan monooleate, Poloxamer 407, Labrafil M-1944CS, Labrafil M-2125CS, Labrasol, Gellucire 44/14, Softigen 767, and mono- and di-fatty acid esters of PEG300, 400, or 1750; *water-insoluble lipids* such as castor oil, corn oil, cottonseed oil, olive oil, peanut oil, peppermint oil, safflower oil, sesame oil, soybean oil, hydrogenated vegetable oils, hydrogenated soybean oil, and medium-chain triglycerides of coconut oil and palm seed oil; *organic liquids/semisolids* such as beeswax, d-α-tocopherol, oleic acid, medium-chain mono- and diglycerides; various *complexing agents* such as α-cyclodextrin, β-cyclodextrin, hydroxypropyl-β-cyclodextrin, and sulfobutylether-β-cyclodextrin; and *phospholipids* such as hydrogenated soy phosphatidylcholine, distearoylphosphatidylglycerol, L-α-dimyristoylphosphatidylcholine, and L-α-dimyristoylphosphatidylglycerol (Strickley, 2004).

In this chapter, the literature on the drug–drug interaction potential of pharmaceutical excipients is reviewed, with a focus on solubilizing excipients.

## 2 INTERACTION OF PHARMACEUTICAL EXCIPIENTS WITH CYTOCHROME P450s (CYPs)

The majority of drugs are metabolized by one or more of the five major cytochrome P450 (CYP) isoforms, namely, CYP1A2, CYP2A6, CYP2C9, CYP2C19, CYP2D6, and CYP3A4, of which CYP3A4 is predominant. A drug can interact with CYP-mediated biotransformation of itself or other drugs via either inhibition of CYP activities or induction of CYP expression. *In vitro* studies aiming to assess the potential of CYP-mediated drug–drug interactions are routinely conducted during the drug discovery and development phases. Sometimes, clinical drug–drug interaction studies are necessary to fully address the possibility of CYP-mediated drug–drug interactions. Compared with drugs, less attention has been paid to the potential of CYP-mediated interactions on drug metabolism due to the excipients present in drug formulations. CYP isoforms are involved in the metabolism of some excipients. For example, vitamin E is metabolized via side-chain degradation initiated by CYP-catalyzed ω-hydroxylation followed by β-oxidation (Brigelius-Flohe, 2003). Therefore, some excipients are likely to interact with CYP enzymes. There are a few examples in which excipients interact with CYPs in *in vitro* studies and/or animal studies. However, conclusive findings from clinical studies that link excipients to CYP-mediated drug–drug interaction are still lacking. The interactions of different excipients with the CYP isoforms responsible for drug metabolism are summarized in Table 1.

### 2.1 CYP Inhibition

Mountfield *et al.* (2000) investigated the CYP inhibition potentials of 16 commonly used excipients in recombinant human CYP3A4, human liver microsomes, and intestinal microsomes from dogs and monkeys. Their data indicated that CYP3A4 activity was inhibited by *N*-methylpyrrolidone, dimethylacetamide, propylene glycol, PEG400, soybean oil, miglyol, glycocholic acid/lecithin, Tween 20, oleic acid, and a mixture of taurocholic acid/lecithin. However, the $IC_{50}/K_i$ values were in the high μM to mM range, indicating the inhibition is not likely of significance from a practical drug–drug interaction point of view.

Solutol HS15 is used in intravenous and oral formulations in various countries. It was found to inhibit recombinant human CYP3A activity ($IC_{50}$ value of 5 μM) (Bittner *et al.*, 2003a). Moreover, Solutol HS15 significantly reduced the metabolism of the CYP3A4 substrates colchicine and midazolam in rat hepatocytes at concentrations below its critical micellar concentration of 0.021% (Gonzalez *et al.*, 2004). Intravenous coadministration of Solutol HS15 and colchicine to rats resulted in a significant reduction in colchicine clearance as well as an increase in $C_{max}$ values (Bittner *et al.*, 2003b). However, intravenous injection of Solutol HS15 had little effect on the pharmacokinetics of coadministered midazolam in rats (Bittner *et al.*, 2003a). In addition to Solutol HS15, 0.03% (w/v) of Cremophor EL and Tween 80 decreased the intrinsic clearance of midazolam in rat hepatocytes by 30% and 25%, respectively, without causing overt cytotoxicity (Gonzalez *et al.*, 2004). Cremophor RH40 inhibited CYP3A-mediated nifedipine oxidation in human liver microsomes ($IC_{50}$ = 0.3%, v/v) (Wandel, Kim and Stein, 2003).

Cyclodextrins are cyclic oligosaccharides with lipophilic inner cavities that are capable of forming non-covalent inclusion complexes with a large variety of molecules (Strickley, 2004). Although only a few approved drug formulations contain cyclodextrins, there is a growing interest in using cyclodextrins as a solubilizer in drug formulations (Challa *et al.*, 2005). Ishikawa, Yoshii and Furuta (2005) reported that hydroxypropyl-β-cyclodextrin inhibited recombinant CYP2C19 and CYP3A4 activities, while methyl-β-cyclodextrin inhibited recombinant CYP1A2, 2D6, 2C9, and 3A4. However, the $IC_{50}$ values were in the mM range, indicating that the inhibition may not be potent enough to cause any clinically meaningful impact on CYP-mediated drug metabolism (Table 1).

Propylene glycol and DMSO are classified as water-soluble organic solvents (Strickley, 2004). Propylene glycol is used in some approved oral and injectable formulations and is also commonly used as a dosing vehicle in animal studies. DMSO is used in only some oral formulations. In a mouse model of acetaminophen-induced liver injury, the dosing vehicle propylene glycol was found to attenuate hepatotoxicity (Snawder *et al.*, 1993). The underlying mechanism was thought to be due to inhibition of CYP2e1 (Snawder *et al.*, 1993; Thomsen *et al.*, 1995), which is the major pathway for acetaminophen bioactivation in rodents. DMSO also diminished acetaminophen-induced liver damage in mice. This protective effect of DMSO was correlated with the inhibition of CYP-mediated aminopyrine-*N*-demethylation (Park *et al.*, 1988).

In summary, although some excipients inhibit CYP activities, the inhibition potency appears to be not high enough to be clinically relevant. However, in preclinical PK and toxicity studies, such excipients can potentially change the clearance of drug candidates by inhibiting CYP activities, because the injection volume in these studies can be up to 1% of body weight. So far, the underlying mechanisms for CYP inhibition by the aforementioned excipients have not been fully investigated.

### 2.2 CYP Induction

CYP3A4 and CYP4F2 are involved in vitamin E metabolism (Kikuta *et al.*, 1999; Parker, Sontag and Swanson, 2000). CYP3A induction is mainly mediated by the nuclear pregnane X receptor (PXR). PXR is activated by a large number of structurally unrelated xenobiotics. Interestingly, different forms of vitamin E were reported to activate a PXR-driven reporter gene in HepG2 cells (Landes *et al.*, 2003). The rank order of the potency of PXR activation was α- and γ-tocotrienol > δ-, α-, γ-tocopherol.

**TABLE 1 Summary of literature findings: effects of excipients on function of CYPs and transporters.**

| Excipient category | Excipient | CYP-related effects | Transporter-related effects |
|---|---|---|---|
| Water-soluble organic solvents | Dimethylacetamide | Inhibits recombinant CYP3A4 ($IC_{50}$: 29.1 mM) (Mountfield *et al.*, 2000) | |
| | DMSO | Inhibits Cyp2e1 (Park *et al.*, 1988) | |
| | Propylene glycol | Inhibits Cyp2e1 (Snawder *et al.*, 1993; Thomsen *et al.*, 1995)<br>Inhibits recombinant CYP3A4 ($IC_{50}$: >50 mM) (Mountfield *et al.*, 2000) | |
| | PEG400 | Inhibits recombinant CYP3A4 ($IC_{50}$: 16.5 mM) (Mountfield *et al.*, 2000) | |
| Nonionic surfactants | Cremophor RH40 | Inhibits CYP3A4 in liver microsomes ($IC_{50}$: 0.3%, v/v) (Wandel, Kim and Stein, 2003) | Causes 20% increase in systemic exposure of the P-gp substrate digoxin in human (Tayrouz *et al.*, 2003)<br>Inhibits P-gp in Caco-2 cell (Wandel, Kim and Stein, 2003) |
| | Cremophor EL | Reduces metabolism of midazolam in rat hepatocytes and liver microsomes (Gonzalez *et al.*, 2004) | Inhibits monocarboxylic acid transporter in Caco-2 cells (Rege, Kao and Polli, 2002)<br>Inhibits P-gp in rat everted gut sac study (Cornaire *et al.*, 2004)<br>Inhibits BCRP in cell-based assay (Yamagata *et al.*, 2007b)<br>No effect on MRP2 function (Bogman *et al.*, 2003) |
| | Labrasol | | Inhibits P-gp in rat everted gut sac study (Cornaire *et al.*, 2004) |
| | Pluronic P85 | | Inhibits Bcrp function in animal study and in cell-based assay (Yamagata *et al.*, 2007a, b) |
| | Solutol HS15 | Inhibits recombinant CYP3A4 ($IC_{50}$: 5 μM) (Bittner *et al.*, 2003a)<br>Reduces metabolism of midazolam in rat hepatocytes and liver microsomes (Gonzalez *et al.*, 2004) | Inhibits P-gp in *in vitro* assays (Coon *et al.*, 1991; Cornaire *et al.*, 2004) |

| | | | |
|---|---|---|---|
| | TPGS1000 | | Potent P-gp inhibitor, causing clinically relevant interaction with drugs (Chang, Benet and Hebert, 1996; Yu *et al.*, 1999; Bogman *et al.*, 2005)<br>No effect on MRP2 function (Bogman *et al.*, 2003) |
| | Tween 20 | Inhibits recombinant CYP3A4 ($IC_{50}$: 3.8 μM) (Mountfield *et al.*, 2000) | Inhibits digoxin transport in rat everted gut sac study (Cornaire *et al.*, 2004)<br>Inhibits Bcrp function in animal study and in cell-based assay (Yamagata *et al.*, 2007a, b) |
| | Tween 80 | | Inhibits peptide transporter in Caco-2 assay (Rege, Kao and Polli, 2002)<br>Inhibits digoxin transport in rat everted gut sac study (Cornaire *et al.*, 2004)<br>No effect on MRP2 function (Bogman *et al.*, 2003) |
| Organic liquids/semisolids | Oleic acid | Inhibits recombinant CYP3A4 ($IC_{50}$: 2.0 mM) (Mountfield *et al.*, 2000) | |
| | Vitamin E | Induces CYP3A via PXR activation (Landes *et al.*, 2003; Kluth *et al.*, 2005) | |
| Complexing agents | β-Cyclodextrin | | No effect on absorption of rhodamine 123 in rat jejunal loops (Oda *et al.*, 2004) |
| | HP-β-CD | Inhibits recombinant CYP2C19 ($IC_{50}$: 6.81 mM) and CYP3A4 ($IC_{50}$: 2.68 mM) (Ishikawa, Yoshii and Furuta, 2005) | |
| | Methyl-β-cyclodextrin | Inhibits recombinant CYP2C19 ($IC_{50}$: 6.45 mM) and CYP3A4 ($IC_{50}$: 0.69 mM) (Ishikawa, Yoshii and Furuta, 2005) | Inhibits P-gp in Caco-2 cells (Arima *et al.*, 2001) |
| Water-insoluble lipids | Soybean oil | Inhibits recombinant CYP3A4 ($IC_{50}$: >50 mM) (Mountfield *et al.*, 2000) | |

Furthermore, CYP3a11 mRNA was induced by α-tocopherol but not by γ-tocotrienol in mice (Kluth *et al.*, 2005). These findings indicate that at least some forms of vitamin E may increase their own metabolism and may also impact the metabolism of other CYP substrates.

## 3 INTERACTION OF PHARMACEUTICAL EXCIPIENTS WITH TRANSPORTERS

Transporter-based drug–drug interactions have been increasingly documented, leading to the inclusion of transporter studies in the recent FDA guidance on drug interaction studies (FDA Draft Guidance for Industry on Drug Interaction Studies, 2006). Some solubilizing excipients such as TPGS1000 have been shown to cause clinically relevant transporter-based drug interactions (Bogman *et al.*, 2005). Therefore, both formulation and drug metabolism/PK scientists involved in drug development need to be aware of the potential interactions between excipients and membrane-associated transporters when developing a drug formulation. Literature findings on the effects of excipients on transporter functions are summarized in Table 1.

### 3.1 P-glycoprotein (P-gp)

P-gp is a 170-kDa plasma membrane-associated transporter belonging to the superfamily of ATP-binding cassette transporters. It was originally discovered in multidrug-resistant tumor cells as a drug efflux transporter. This transporter was later found to be highly expressed in normal tissues. P-gp is localized on the columnar epithelial cells of the intestine, the canalicular membrane of hepatocytes in the liver, the brush-border domain of the proximal tubular cells in the kidneys, and the luminal surface of the blood–brain barrier. The anatomical localization of P-gp suggests that this transporter functions to protect the body against toxic xenobiotics by excreting these compounds into the intestinal lumen, bile, and urine, and by limiting their access to the brain (Ambudkar *et al.*, 2003). P-gp transports many structurally diverse drugs such as digoxin, vinca alkaloids, and HIV protease inhibitors. A number of drugs such as verapamil (Pedersen *et al.*, 1981; Ito *et al.*, 1993), clarithromycin (Wakasugi *et al.*, 1998), and quinidine (Sadeque *et al.*, 2000) are potent P-gp inhibitors. Coon *et al.* (1991) reported that Solutol HS15 renders multidrug-resistant tumor cells more susceptible to cytotoxic agents by inhibiting their efflux transport. Thereafter, a plethora of *in vitro* and *in vivo* studies demonstrated that many other solubilizing excipients inhibit P-gp function to different extents. For example, in a rat everted gut sac study, various excipients such as TPGS1000, Cremophor EL, Labrasol, Solutol HS15, Tween 20, Tween 80, Softigen 767, Miglyol, sucrose monolaurate, Imwitor742, and Acconon E inhibited the transport of digoxin across the intestinal mucosa (Cornaire *et al.*, 2004). Similarly, TPGS, Cremophor EL, Cremophor RH40, Tween 80, PEG300, Pluronic PE8100 and PE6100 inhibited P-gp function in several *in vitro* studies using cell-based transport assays (Hugger *et al.*, 2002; Rege, Kao and Polli, 2002; Bogman *et al.*, 2003). 2,6-D-*O*-methyl-β-cyclodextrin, a hydrophilic complexing agent, was found to decrease the efflux of two P-gp substrates, tacrolimus and rhodamine 123, across Caco-2 monolayers (Arima *et al.*, 2001), whereas β-cyclodextrin (1%) has little effect on the absorption of rhodamine 123 from *in situ* jejunal loops from rats (Oda *et al.*, 2004).

TPGS1000 is a water-soluble derivative of vitamin E comprised of a hydrophilic polar (water-soluble) head and a lipophilic (lipid-soluble) alkyl tail. Sokol *et al.* (1991)

reported that treatment with TPGS1000 (25 mg/kg, daily) enhanced absorption of the highly lipophilic drug cyclosporine in liver transplant patients. While the authors originally suggested that the improved oral bioavailability was due to the solubility enhancing effect of TPGS1000, Chang, Benet and Hebert (1996) later demonstrated that enhanced permeability due to P-gp inhibition also played a role. TPGS1000 (12%) is one of the components in the Agenerase oral formulation of the poorly water-soluble HIV protease inhibitor amprenavir from Glaxo Wellcome (Yu *et al.*, 1999). Amprenavir in capsule or tablet formulations is barely orally available, but formulations containing 20% and 50% TPGS1000 increase its oral bioavailability in beagle dogs at 25 mg/kg to 70% and 80%, respectively (Yu *et al.*, 1999). TPGS1000 was identified as a potent inhibitor of an active efflux system in Caco-2 cell monolayers at concentrations much lower than its critical micelle concentration (0.2 mg/mL) (Yu *et al.*, 1999). Paclitaxel, an anticancer drug, is typically administered intravenously due to its poor solubility and poor permeability. A recent study showed that TPGS1000 enhanced oral paclitaxel absorption by improving both parameters (Varma and Panchagnula, 2005). In these three examples, TPGS1000 enhances the oral bioavailability of poorly water-soluble drugs that are P-gp substrates not only by solubility-enhancing micelle formation, but also by increasing the permeability via inhibiting efflux transporter function.

Bogman *et al.* (2005) reported that the $AUC_{0\text{-}\infty}$ and $C_{max}$ of the P-gp substrate talinolol were increased by 39% and 100%, respectively, in healthy human volunteers perfused intraduodenally with talinolol and 0.04% TPGS1000, whereas another solubilizing excipient, Poloxamer 188 (0.8%), had no effect on talinolol pharmacokinetics. Additionally, TPGS1000, but not Poloxamer 188, inhibited the P-gp-mediated talinolol transport in Caco-2 cells (Bogman *et al.*, 2005). These findings further support that TPGS1000 can cause clinically relevant P-gp-mediated drug interactions.

TPGS analogs with different PEG chain length, and/or a modified hydrophobic core were found to exhibit differential effects on rhodamine123 dye transport in Caco-2 monolayers (Collnot *et al.*, 2006). Moreover, the inhibitory potency of these TPGS analogs on P-gp-mediated efflux in Caco-2 cells highly correlated with that of P-gp ATPase activity but not with changes in membrane fluidity determined by electron spin resonance (Collnot *et al.*, 2007). These *in vitro* studies shed important light on the mechanism(s) through which TPGS modulates P-gp function, supporting the notion that efflux pump energy source depletion is a major factor in the inhibitory mechanism of TPGS.

In a clinical study, a high dose of oral Cremophor RH40 (600 mg three times daily) in combination with a single oral dose of 0.5 mg digoxin in a hard gelatin capsule resulted in an increase in the systemic exposure of digoxin by approximately 22% (Tayrouz *et al.*, 2003). This finding is in agreement with *in vitro* observation that Cremophor RH40 inhibited P-gp function in Caco-2 cells ($IC_{50}$ = 0.03%, v/v) (Wandel, Kim and Stein, 2003). Although this *in vivo* and *in vitro* correlation suggests that P-gp inhibition led to the pharmacokinetic interaction between Cremophor RH40 and digoxin, the influence of Cremophor on the pharmacokinetics of P-gp substrates is not straightforward. For example, although Cremophor EL in the intravenous formulation can result in greater than dose-proportional increases in paclitaxel exposure in cancer patients (Sparreboom *et al.*, 1996), a study by the same authors in mice suggests that P-gp inhibition may not be the mechanism for the observed decrease in clearance. At comparable blood levels achieved in the human study, tissue levels of Cremophor EL were below the limit of detection, and presumably too low to inhibit excretion from organs such as the liver and/or kidney (Sparreboom *et al.*, 1996). Further investigation

by the same research group showed that addition of Cremophor EL at a clinically achievable concentration (0.5%, v/v) to human blood decreased the whole blood to plasma concentration ratio of paclitaxel from 1.07 to 0.69 (Sparreboom *et al.*, 1999). These authors concluded that Cremophor EL formed micelles in blood that entrap compounds such as paclitaxel, resulting in an altered cellular distribution and a concomitant increase in plasma concentration. Furthermore, Cremophor EL in an oral paclitaxel formulation was found to reduce absorption of paclitaxel from the gut, presumably by micellar entrapment of the drug (Bardelmeijer *et al.*, 2002).

### 3.2 Breast Cancer Resistance Protein (BCRP)

BCRP also belongs to the ATP-binding cassette family of drug efflux transporters. The tissue distribution of BCRP is similar to that of P-gp. BCRP transports a wide variety of compounds including anticancer drugs, food carcinogens, antibiotics, and conjugated metabolites. Therefore, BCRP, like P-gp, plays an important role in the absorption, distribution, and elimination of many drugs (Krishnamurthy and Schuetz, 2006). Yamagata *et al.* (2007a) reported that Pluronic P85 and Tween 20 given orally along with topotecan increased the AUC of topotecan by approximately twofold in wild-type mice but not in Bcrp knockout mice. Pluronic P85 and Tween 20 did not affect the AUC of topotecan after intravenous administration. These findings suggest that Pluronic P85 and Tween 20 can improve the oral bioavailability of BCRP substrates by inhibiting BCRP function in the gut. The same research group also reported that Cremophor EL, Tween 20, Span 20, Pluronic P85, and Brij30 enhanced the uptake of the BCRP substrate mitoxantrone into BCRP-expressing cells (Yamagata *et al.*, 2007b). A recent study showed that in BCRP-expressing MDCKII cells, cholesterol depletion with methyl-β-cyclodextrin decreased BCRP activity by 40%, whereas cholesterol repletion with methyl-β-cyclodextrin/cholesterol-inclusion complexes restored BCRP function (Storch *et al.*, 2007). These findings suggest that cyclodextrin affects BCRP function by altering the cholesterol composition of cellular membranes. Whether the BCRP inhibitory effects of these excipients can alter the pharmacokinetics of drugs in a clinical situation is still unknown.

### 3.3 Other Transporters

In contrast to P-gp and BCRP, there are only a few studies examining the effects of excipients on the function of other transporters. For example, Rege, Kao and Polli (2002) reported that Cremophor EL inhibited the monocarboxylic acid transporter function as measured by benzoic acid permeability, and Tween 80 inhibited the permeability of glycyl sarcosin, a substrate of peptide transporter. Bogman *et al.* (2003) showed that TPGS1000, Pluronic PE8100, Cremophor EL, Pluronic PE6100, and Tween 80 had little effects on multidrug resistance-associated protein 2 (MRP2)-mediated efflux in MRP2-overexpressing MDCK cells.

## 4 SUMMARY

The reasons to investigate the potential of drug candidates to cause enzyme- and transporter-mediated drug–drug interaction are well understood. Similarly, during formulation development, formulation and PK scientists should be equipped with

information regarding the potential of selected excipients to cause transporter- and/or enzyme-based drug interaction. Such information is important, because (i) the amount of excipients in dosage formulations often far exceeds that of the active ingredients and (ii) some solubilizing excipients (e.g. TPGS1000) have been convincingly shown to affect the clinical PK of drugs by inhibiting P-gp function. The P-gp inhibition properties of excipients have been applied to increase the oral bioavailability of drugs that are P-gp substrates as exemplified by the development of the TPGS1000-containing Agenerase oral formulation of the HIV protease inhibitor amprenavir (Yu *et al.*, 1999). For many clinical indications and therapeutic areas, however, a dose formulation containing a large amount of a P-gp-inhibiting excipient such as TPGS1000 may result in undesirable product labeling and should be carefully considered. Other transporters, such as BCRP have been increasingly recognized as important factors that contribute to the absorption, distribution, and elimination of drugs. More basic and clinical research is needed in order to understand whether the functions of these transporters can be affected by excipients and whether the effects are clinically relevant. Despite many *in vitro* and *in vivo* studies showing that a number of excipients affect CYP activity and expression, clinically relevant data are still lacking to the best of our knowledge. Therefore, enzyme-based drug interaction involving excipients appears to occur predominantly in preclinical studies where the amount of excipients used in the dose formulations is much higher than that in the formulations intended for clinical usage.

## REFERENCES


Ambudkar, S.V., Kimchi-Sarfaty, C., Sauna, Z.E., Gottesman, M.M. (2003). P-glycoprotein: from genomics to mechanism. *Oncogene*, *22*, 7468–7485.

Arima, H., Yunomae, K., Hirayama, F., Uekama, K. (2001). Contribution of P-glycoprotein to the enhancing effects of dimethyl-beta-cyclodextrin on oral bioavailability of tacrolimus. *J. Pharmacol. Exp. Ther.*, *297*, 547–555.

Bardelmeijer, H.A., Ouwehand, M., Malingre, M.M., Schellens, J.H., Beijnen, J.H., van Tellingen, O. (2002). Entrapment by Cremophor EL decreases the absorption of paclitaxel from the gut. *Cancer Chemother. Pharmacol.*, *49*, 119–125.

Bittner, B., Gonzalez, R.C., Isel, H., Flament, C. (2003a). Impact of Solutol HS 15 on the pharmacokinetic behavior of midazolam upon intravenous administration to male Wistar rats. *Eur. J. Pharm. Biopharm.*, *56*, 143–146.

Bittner, B., Gonzalez, R.C., Walter, I., Kapps, M., Huwyler, J. (2003b). Impact of Solutol HS 15 on the pharmacokinetic behavior of colchicine upon intravenous administration to male Wistar rats. *Biopharm. Drug Dispos.*, *24*, 173–181.

Bogman, K., Erne-Brand, F., Alsenz, J., Drewe, J. (2003). The role of surfactants in the reversal of active transport mediated by multidrug resistance proteins. *J. Pharm. Sci.*, *92*, 1250–1261.

Bogman, K., Zysset, Y., Degen, L., Hopfgartner, G., Gutmann, H., Alsenz, J., Drewe, J. (2005). P-glycoprotein and surfactants: effect on intestinal talinolol absorption. *Clin. Pharmacol. Ther.*, *77*, 24–32.

Brigelius-Flohe, R. (2003). Vitamin E and drug Metabolism. *Biochem. Biophys. Res. Comm.*, *305*, 737–740.

Challa, R., Ahuja, A., Ali, J., Khar, R.K. (2005). Cyclodextrins in drug delivery: an updated review. *AAPS Pharm. Sci. Tech.*, *6*, E329–E357.

Chang, T., Benet, L.Z., Hebert, M.F. (1996). The effect of water-soluble vitamin E on cyclosporine pharmacokinetics in healthy volunteers. *Clin. Pharmacol. Ther.*, *59*, 297–303.

Collnot, E.M., Baldes, C., Wempe, M.F., Hyatt, J., Navarro, L., Edgar, K.J., Schaefer, U.F., Lehr, C.M. (2006). Influence of vitamin E TPGS poly(ethylene glycol) chain length on apical efflux transporters in Caco-2 cell monolayers. *J. Control. Release*, *111*, 35–40.

Collnot, E.M., Baldes, C.B., Wempe, M.F., Kappl, R., Hutermann, J., Hyatt, J.A., Edgar, K.J., Schaefer, U.F., Lehr, C.M. (2007). Mechanism of inhibition of P-glycoprotein mediated efflux by vitamin E TPGS: influence on ATPase activity and membrane fluidity. *Mol. Pharm.*, *4*, 465–474.

Coon, J.S., Knudson, W., Clodfelter, K., Lu, B., Weinstein, R.S. (1991). Solutol HS15, nontoxic polyoxyethylene esters of 12-hydroxystearic acid, reverses multidrug resistance. *Cancer Res.*, *51*, 897–902.

Cornaire, G., Woodley, J., Hermann, P., Cloarec, A., Arellano, C., Houin, G. (2004). Impact of excipients on the absorption of P-glycoprotein substrates *in vitro* and *in vivo*. *Inter. J. Pharm.*, *278*, 119–131.

FDA Draft Guidance for Industry on Drug Interaction Studies. (2006). www.fda.gov/cder/guidance/6695dft.htm.

Gonzalez, R.C., Huwyler, J., Boess, F., Walter, I., Bittner, B. (2004). In vitro investigation on the impact of the surface-active excipients Cremophor EL, Tween 80 and Solutol HS 15 on the metabolism of midazolam. *Biopharm. Drug Dispos.*, *25*, 37–49.

Hugger, E.D., Novak, B.L., Burton, P.S., Audus, K.L., Borchardt, R.T. (2002). A comparison of commonly used polyethoxylated pharmaceutical excipients on their ability to inhibit P-glycoprotein activity in vitro. *J. Pharm. Sci.*, *91*, 1991–2002.

Ishikawa, M., Yoshii, H., Furuta, T. (2005). Interaction of modified cyclodextrins with cytochrome P-450. *Biosci. Biotechnol. Biochem.*, *69*, 246–248.

Ito, S., Woodland, C., Harper, P.A., Koren, G. (1993). The mechanism of the verapamil-digoxin interaction in renal tubular cells (LLC-PK1). *Life Sci.*, *53*, 399–403.

Kalasz, H., Antal, I. (2006). Drug excipients. *Curr. Med. Chem.*, *13*, 2535–2563.

Kikuta, Y., Miyauchi, Y., Kusunose, E., Kusunose, M. (1999). Expression and molecular cloning of human liver leukotriene B4 omega-hydroxylase (CYP4F2) gene. *DNA Cell Biol.*, *18*, 723–730.

Kluth, D., Landes, N., Pfluger, P., Muller-Schmehl, K., Weiss, K., Bumke-Vogt, C., Ristow, M., Brigelius-Flohe, R. (2005). Modulation of CYP3a11 mRNA expression by a-tocopherol but not α-tocotrienol in mice. *Free Radic. Biol. Med.*, *38*, 507–514.

Krishnamurthy, P., Schuetz, J.D. (2006). Role of ABCG2/BCRP in biology and medicine. *Annu. Rev. Pharmacol. Toxicol.*, *46*, 381–410.

Landes, N., Pfluger, P., Kluth, D., Birringer, M., Ruhl, R., Bol, G.F., Glatt, H.R., Brigelius-Flohe, R. (2003). Vitamin E activates gene expression *via* the pregnane X receptor. *Biochem. Pharmacol.*, *65*, 269–273.

Mountfield, R.J., Senepin, S., Schleimer, M., Walter, I., Bittner, B. (2000). Potential inhibitory effects of formulation ingredients on intestinal Cytochrome P450. *Int. J. Pharm.*, *211*, 89–92.

Oda, M., Saitoh, H., Kobayashi, M., Aungst, B.J. (2004). Beta-cyclodextrin as a suitable solubilizing agent for *in situ* absorption study of poorly water-soluble drugs. *Int. J. Pharm.*, *280*, 95–102.

Park, Y., Smith, R.D., Combs, A.B., Kehrer, J.P. (1988). Prevention of acetaminophen-induced hepatotoxicity by dimethyl sulfoxide. *Toxicology*, *52*, 165–175.

Parker, R.S., Sontag, T.J., Swanson, J.E. (2000). Cytochrome P4503A-dependent metabolism of tocopherols and inhibition by sesamin. *Biochem. Biophys. Res. Commun.*, *277*, 531–534.

Pedersen, K.E., Dorph-Pedersen, A., Hvidt, S., Klitgaard, N.A., Nielsen-Kudsk, F. (1981). Digoxin-verapamil interaction. *Clin. Pharmacol. Ther.*, *30*, 311–316.

Rege, B.D., Kao, J.P., Polli, J.E. (2002). Effects of nonionic surfactants on membrane transporters in Caco-2 cell monolayers. *Eur. J. Pharm. Sci.*, *16*, 237–246.

Sadeque, A.J., Wandel, C., He, H., Shah, S., Wood, A.J. (2000). Increased drug delivery to the brain by P-glycoprotein inhibition. *Clin. Pharmacol. Ther.*, *68*, 231–237.

Snawder, J.E., Benson, R.W., Leaky, J.E., Roberts, D.W. (1993). The effects of propylene glycol on the P450-dependent metabolism of acetaminophen and other chemicals in subcellular fractions of mouse liver. *Life Sci.*, *52*, 183–189.

Sokol, R.J., Johnson, K.E., Karrer, F.M., Narkewicz, M.R., Smith, D., Kam, I. (1991). Improvement of cyclosporine absorption in children after liver transplantation by means of water-soluble vitamin E. *Lancet*, *338*, 212–214.

Sparreboom, A., van Tellingen, O., Nooijen, W.J., Beijnen, J.H. (1996). Nonlinear pharmacokinetics of paclitaxel in mice results from the pharmaceutical vehicle Cremophor EL. *Cancer Res.*, *56*, 2112–2115.

Sparreboom, A., van Zuylen, L., Brouwer, E., Loos, W.J., de Bruijn, P., Gelderblom, H., Pillay, M., Nooter, K., Stoter, G., Verweij, J. (1999). Cremophor EL-mediated alteration of paclitaxel distribution in human blood: clinical pharmacokinetics implications. *Cancer Res.*, *59*, 1454–1457.

Storch, C.H., Ehehalt, R., Haefeli, W.E., Weiss, J. (2007). Localization of the human breast cancer resistance protein (BCRP/ABCG2) in lipid rafts/caveolae and modulation of its activity by cholesterol *in vitro*. *J. Pharmacol. Exp. Ther.*, *323*, 257–264.

Strickley, R.G. (2004). Solubilizing excipients in oral and injectable formulations. *Pharm. Res.*, *21*, 201–230.

Tayrouz, Y., Ding, R., Burhenne, J., Riedel, K., Weiss, J., Hoppe-Tichy, T., Haefeli, W.E., Mikus, G. (2003). Pharmacokinetics and pharmaceutic interaction between digoxin and Cremophor RH40. *Clin. Pharmacol. Ther.*, *73*, 397–405.

Thomsen, M.S., Loft, S., Roberts, D.W., Poulsen, H.E. (1995). Cytochrome P450 2E1 inhibition by propylene glycol prevents acetaminophen hepatotoxicity in mice without Cytochrome P450 1A2 inhibition. *Pharmacol. Toxicol.*, *76*, 395–399.

Varma, M.V., Panchagnula, R. (2005). Enhanced oral paclitaxel absorption with vitamin E-TPGS: effect on solubility and permeability *in vitro, in situ* and *in vivo*. *Eur. J. Pharm. Sci.*, *25*, 445–453.

Wakasugi, H., Yano, I., Ito, T., Hashida, T., Futami, T., Nohara, R., Sasayama, S., Inui, K. (1998). Effect of clarithromycin on renal excretion of digoxin: interaction with P-glycoprotein. *Clin. Pharmacol. Ther.*, *64*, 123–128.

Wandel, C., Kim, R.B., Stein, C.M. (2003). "Inactive" excipients such as Cremophor can affect in vivo drug disposition. *Clin. Pharm. Ther.*, *73*, 394–396.

Yamagata, T., Kusuhara, H., Morishita, M., Takayama, K., Benameur, H., Sugiyama, Y. (2007a). Improvement of the oral drug absorption of topotecan through the inhibition of intestinal xenobiotics efflux transporter, breast cancer resistance protein, by excipients. *Drug Metab. Dispos.*, *35*, 1142–1148.

Yamagata, T., Kusuhara, H., Morishita, M., Takayama, K., Benameur, H., Sugiyama, Y. (2007b). Effect of excipients on breast cancer resistance protein substrate uptake activity. *J. Control. Release*, *124*, 1–5.

Yu, L., Bridgers, A., Polli, J., Vickers, A., Long, S., Roy, A., Winnike, R., Coffin, M. (1999). Vitamin E-TPGS increases absorption flux of an HIV protease inhibitor by enhancing its solubility and permeability. *Pharm. Res.*, *16*, 1812–1817.

# The Role of Drug Metabolism in Toxicity

CARL D. DAVIS and UMESH M. HANUMEGOWDA

## 1 INTRODUCTION

The term "toxicity" in general denotes any adverse event that results from exposure to an agent, which includes a wide variety of substances: natural or man-made, physical, chemical, biological or radiological, and endogenous or exogenous. Toxicity can result from exaggerated on-target pharmacology, unrelated off-target activity, or biotransformation to a metabolite that is reactive and/or more potent in its on- or off-target activity. Evolutionary pressures have equipped species, from bacteria to humans, with a complement of enzymes that help to regulate and limit the exposure and/or accumulation of xenobiotic agents, such as those encountered in the diet, the environment, accidental or deliberate exposure to poisons, or from administration of therapeutic agents. Usually these drug-metabolizing enzymes (DMEs) facilitate the elimination of xenobiotics from the body and serve principally as a detoxification process. However, in some cases, the same enzymes can be involved in the bioactivation of a xenobiotic and/or its metabolite(s) to facilitate adverse reactions and toxicity. In addition to the toxic agent itself, several important factors also contribute to or even determine the interspecies and interindividual variability and susceptibility to toxicity. These include differences in the expression of metabolizing enzymes, population differences or polymorphisms in the enzymes expressed, physiological state, underlying disease, age, and interactions associated with concurrently administered drugs.

There are many textbooks and reviews on both toxicity and the role of biotransformation; however, very few articles have reviewed the specific role of DMEs and susceptibility to toxicity. This chapter is intended for the practicing pharmaceutical scientist and provides real-world examples potentially encountered when evaluating and developing drug candidates. We describe the drug toxicity that is mediated predominantly through the formation of reactive metabolites. To extend the scope beyond the specific case to the general, we also summarize some key characteristics of the common enzymes encountered in metabolic clearance of drugs and highlight elements to consider, such as differences in tissue expression between species and/or gender. Examples are offered

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

to illustrate how DMEs are involved in the bioactivation of drugs to form reactive metabolites. Included are many of the well-studied drugs that provide molecular mechanisms of toxicity and examples that describe toxicity by target organ. This is followed by a brief overview of experimental methods that can be used to evaluate drug metabolism and toxicity. Lastly, a brief discussion is offered to describe the pharmaceutical industry's progress to date, with some case examples to illustrate how drug metabolism and toxicity of one drug can influence the development and success of subsequent drugs.

## 2 DRUG-METABOLIZING ENZYMES

Drug toxicity and the factors that govern it are quite complex. In drug discovery, toxicity can depend on factors that may be species-specific, thus forcing toxicologists to determine the relevance of toxicity in animals to that in humans. The balance of risk versus benefit is always weighed prior to exposing humans to a new chemical entity (NCE), with the pharmaceutical industry and regulatory agencies favoring safety over uncertainty. Fortunately, by combining an understanding of drug metabolism and toxicology, some common adverse effects in humans—with the possible exception of idiosyncratic drug reactions—can be predicted and to some extent prevented. This allows the safety and efficacy of NCEs to be assessed in clinical trials with minimal risk to the humans tested.

The efficacy and toxicity of a drug are both dependent on its overall exposure (including at the target site) and its intrinsic potency at the target. Adverse effects of an otherwise well-tolerated drug can be seen after its accidental or deliberate overdose (e.g. acetaminophen-induced liver toxicity). There are several mechanisms that help to regulate exposure to drugs: biliary, renal, and metabolic clearances are three such processes by which well-absorbed drugs can be eliminated. The net result is that drug accumulation is avoided and toxicity limited. These same mechanisms, when operating effectively, also make it difficult to achieve oral bioavailability or sustain an exposure that is adequate for a therapeutic effect. Therefore, the objective in discovery and optimization of drug candidates is to achieve adequate exposure for efficacy while staying below concentrations that potentially result in toxicity. Pragmatically, the emphasis therefore is not to eliminate the potential for toxicity per se, but rather to find drug candidates that offer the widest separation in exposure between therapeutic and adverse effects.

Biliary and renal clearance can be indirectly associated to the etiology of certain types of toxicity. The bile duct canaliculi (liver), nephron/proximal tubules (kidney), or the bladder, in particular, can all be exposed to relatively high local concentrations of drug and/or metabolites associated with their elimination. If biliary or renal toxicity are a direct result of unchanged drug levels, then, generally speaking, these adverse effects would simply be predicted to occur in a human population and thus provide evidence against the subsequent development of that compound as a drug. In contrast, toxicity associated with metabolic clearance is more complex (the net result of potential contributions from numerous and diverse enzymes and transporters that show tremendous differences between tissues and species) and usually much harder to extrapolate from animals to humans. With an NCE, one needs a good understanding of the enzymes involved in its metabolism and the functional activity and reactivity of its metabolite(s) to predict any consequences in humans.

DMEs are characterized by broad substrate specificity and tend not to be named by a specific reaction. Instead they have been identified and grouped into families based on functional activity and their shared homology in the amino acid sequence of the functional protein. Table 1 lists many of the major families of human enzymes involved in metabolic clearance and often implicated in adverse drug reactions. Most DMEs are expressed ubiquitously in tissues, usually with the liver expressing the highest activities (there are notable exceptions, a few of which are indicated in Table 1). With metabolism-related toxicity most often observed in the liver, hepatocytes and liver subfractions have become central to drug lead identification and optimization. Adverse effects and toxicity occurring specifically in other cells, tissues, and/or organ systems may require a tailored screening paradigm.

### 2.1 Cytochrome P450 Enzymes

The cytochrome P450s (CYPs) (EC 1.14.14.1) arguably are the single most important family of enzymes involved in regulating exposure to pharmacologically active, toxic, or potentially toxic xenobiotics. These enzymes exhibit highly conserved regions in their amino acid sequence across different species including humans. The CYPs are membrane-bound hemoproteins found in the smooth endoplasmic reticulum and can be prepared as microsomal fractions for use *in vitro*. Individual CYPs show some specialization with almost distinctive substrate selectivity that partly reflects roles in endogenous metabolism. As a family, CYPs catalyze oxidation and reduction reactions of a highly diverse collection of substrates providing metabolic clearance for compounds not readily eliminated in urine or bile. In general, CYP-mediated oxidation is the more relevant reaction for generating reactive or toxic metabolites.

CYP-mediated oxidation proceeds via redox cycling between the ferric to ferrous states of the heme protoporphyrin group in the active site of the enzyme, with the NAD(P)H oxidoreductase and cytochrome b5 accessory flavoproteins involved in the electron transfer from nicotinamide adenine dinucleotide phosphate (NADPH) to the heme and the hemo-substrate complex. Ultimately, this process catalyzes the

**TABLE 1 Families and major distribution of common human drug-metabolizing enzymes.**

| Enzyme family | Significant family members | Expression |
|---|---|---|
| Cytochrome P450s (membrane-bound; microsomal and mitochondrial) | CYP1A1, 1A2, 2A6, 2A13, 2B6, 2C8, 2C9, 2C18, 2C19, 2D6, 2E1, 2F1, 2J2, 3A4, 3A5, 3A7; 4A11, 4F2, 4F3a, 4F3b, and 4F12 | Skin, brain, kidney, heart, lung, nasal mucosa, intestine, with major expression of most commonly encountered forms in the liver |
| FMO (microsomal) | FMO1, FMO2, FMO3, FMO4, FMO5 | Diverse tissues. Major expression FMO1—kidney; FMO2—lung (inactive Caucasians, Asians), FMO3, 4 and 5—all liver |
| Uridine glucuronosyl-S-Transferases (microsomal) | UGT1A1, 1A3, 1A4, 1A6, 1A7, 1A8, 1A9, 1A10, 2B4, 2B7, 2B10, 2B11, 2B15, 2B17 | Kidney, intestine, lung, with major expression in the liver |
| *N*-acetyl transferases (cytosolic) | NAT1, NAT2 | Liver, intestine, and others |

**TABLE 1** *Continued*

| Enzyme family | Significant family members | Expression |
|---|---|---|
| Monoamine/diamine oxidase (mitochondrial) | MAO-A, MAO-B; DAO (histaminase) | Outer mitochondrial membrane with ubiquitous expression: MAO-A > MAO-B in lung, duodenal mucosa, placenta, adult brain; MAO-B > MAO-A in myocardium, platelets, lymphocytes, fetus brain; diamine oxidase highest in the intestinal mucosa |
| Alcohol dehydrogenases (cytosolic) | ADH classes 1–5. ADH1A, ADH1B, ADH1C; | ADH1A, -1B, -1C (liver); ADH4—stomach |
| Aldehyde dehydrogenases (cytosolic. microsomal, mitochondrial) | ALDH 10 families, 13 subfamilies: ALDH1A1, 1A2, 1A3, 3A1, 3A2, 4A1, 5A1, 8A1, 9A1, 1L1 (many others) | Most tissues |
| Aldehyde oxidase (cytosolic) | AO | Liver, kidney, lung, brain |
| Sulfotransferases (cytosolic) | SULT1A1, 1A2, 1A3, 1A4, 1B1, 1C2, 1C4, 1E1, 2A1, 2B1a, 2B1b, 4A1 | Ubiquitous, with highest expression usually in liver, intestine, and placenta |
| Esterases (cytosolic and microsomal) | CE1, CE2, AChE, PON1, PON2, PON3 | Ubiquitous expression as a family with tissue-specific expression characteristic of some enzymes (e.g. AChE in brain, muscle, red blood cells; PON1, PON3 in liver and blood; PON2 brain, liver, kidney, testes but not blood) |
| Epoxide hydrolases (cytosolic and microsomal) | mEH (HYL1), cEH (HYL2) | Mainly liver with a relatively low expression in other tissues, for example, lung, brain, blood, and ovary |
| Glutathione-*S*-transferases (cytosolic) | GSTA1-A4, GSTM1-M5, GSTP,Zeta-,Omega-,Sigma- | GSTA1-1, A2-2, A4-4, M1-1, M3-3, M4-4, K1-1,O1-1, T1-1, and Z1-1 are all expressed in the liver; GSTA1-1, A2-2, A4-4, M1-1, M3-3, K1-1,O101, P1-1, T1-1, and Z1-1 are expressed in the small intestine; and GSTP1-1 and T1-1 are found in erythrocytes |
| Peroxidases (cytosolic) | MPO, EPO, catalase; COX-1 (PHS-1), COX-2 (PHS-2) | Neutrophils, eosinophils, platelets, intestine, uterus, mammary gland, thyroid; COX-1 is expressed ubiquitously (housekeeping gene); COX-2 is an inducible form |

CYPs, cytochrome P450s; FMO, flavin-containing monooxygenase; UGT, uridine diphosphate glucuronosytransferases; NAT, *N*-acetyl transferase; MAO, monoamine oxidase; DAO, diamine oxidase; ADH, alcohol dehydrogenase; ALDH, aldehyde dehydrogenase; AO, aldehyde oxidase; SULT, sulfotransferase; CE, carboxylesterase; AChE, acetylcholinesterase; PON, paraoxonase; mEH, microsomal epoxide hydrolase; cEH, cytosolic epoxide hydrolase; GST, glutathione *S*-transferase; MPO, myeloperoxidase; EPO, eosinophil peroxidase; COX, cylcooxygenase; PHS, prostaglandin H synthase.

incorporation of oxygen from molecular oxygen producing water and a metabolite. This reaction usually results in a more polar molecule that can be eliminated directly in bile or urine, and/or subjected to further metabolism such as conjugation with sulfate or glucuronide as mediated by the sulfotransferases (SULTs) or uridine diphosphate glucuronosyltransferases (UGTs), respectively. Occasionally a less polar metabolite can be formed, for example, the CYP-mediated formation of an *N*-oxide metabolite of clozapine in human liver microsomes (Pirmohamed *et al.*, 1995). Peroxidase-like reactions can also occur in the CYP catalytic cycle, with the ferrous-$O_2$ complex decomposing to form a superoxide anion or hydrogen peroxide that potentially can react with drug and/or its metabolites.

In some instances, the transition intermediates formed in the CYP catalytic cycle can prove to be highly reactive and bind with the heme (diltiazem, tamoxifen or troleandomycin and CYP3A4) or apoprotein (dihydralazine and CYP1A2; tamoxifen and CYP2B6; tienilic acid and CYP2C9). Mechanism-based inactivation then occurs that results in partial or total irreversible inactivation of the CYP enzyme. Mechanism-based inactivation can exacerbate the risk of drug–drug interactions and toxicity from coadministered drugs by increasing their exposure for a prolonged period (i.e. until new enzyme has been synthesized to replace that which is inactivated). If the reactive species formed in the CYP is able to escape the active site and bind covalently with nucleophilic macromolecules such as proteins, RNA, or DNA, this can culminate in adverse drug reactions and toxicity (Kalgutkar, Obach, and Maurer, 2007; Kalgutkar and Soglia, 2005). The chemistry of these reactions has been well studied and structural alerts identified (Table 2). Based on such anecdotal experiences, medicinal chemistry often is guided by blocking the potential site of metabolism that could result in a problematic reactive intermediate or avoiding certain functional groups entirely. Table 2 lists the most common structural alerts encountered in medicinal chemistry.

CYPs are divided into subfamilies based on their shared sequence in protein homology, with >40% sequence similarity designating the family and >55% shared sequence homology designating the subfamily. Over 50 different CYPs have been identified in humans, with the CYP1A1, 1A2, 2A6, 2A13, 2B6, 2C8, 2C9, 2C18, 2C19, 2D6, 2E1, 2F1, 2J2, 3A4, 3A5, 3A7; 4A11, 4F2, 4F3a, 4F3b, and 4F12 forms responsible for the metabolic clearance of most drugs (CYP3A4, 2C9, 2D6, 2C19, 1A2, and 2C8 being most often encountered). Although CYPs are quite well conserved across species, it is important to note that they are not identical. Even a small difference in amino acid sequence can affect substrate affinity and the capacity to metabolize a substrate. This is exemplified by allelic variants of CYP2C9 in humans where, for instance, individuals possessing the CYP2C9*3 allele (an A1075→C transition in exon 7 changes the amino acid sequence from $Ile^{359}$→$Leu^{359}$) result in a form of CYP2C9 that demonstrates a lower clearance of several substrates, including (*S*)-warfarin (Takanashi *et al.*, 2000; Scordo *et al.*, 2002; Xie *et al.*, 2002). With even greater differences existing between orthologues in humans and other species, caution must be exercised in extrapolating from preclinical animals models and when interpreting the relevance of animal toxicity or lack thereof to humans.

### 2.2 Flavin-Containing Monooxygenases

Flavin-containing monooxygenases (EC 1.14.13.8) (FMOs) are another class of membrane-bound enzymes found in the endoplasmic reticulum. In common with CYPs, FMOs also mediate NADPH-dependent reactions that lead to oxidation of xenobiotics,

**TABLE 2 Fuctional groups and drugs associated with reactive metabolite formation.**

| Functional Group | Example |
|---|---|
| Aniline (e.g. procainamide) | 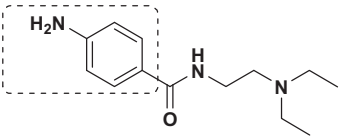 |
| Thiophene (e.g. tienilic acid) | 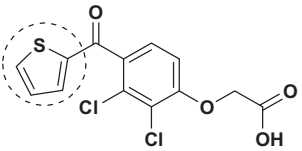 |
| Hydrazines (e.g. isoniazid) | 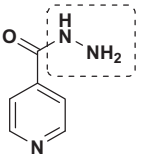 |
| Nitroarene (e.g. choramphenicol) | 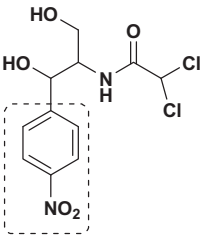 |
| Quinones (e.g. diethylstilbestrol) | 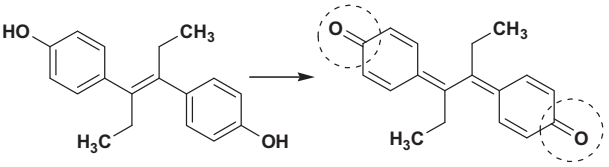 |

with FMOs favoring more specific *N*- and *S*-oxidation reactions. FMOs are otherwise quite distinct from CYPs: they are flavoproteins that contain a single flavin adenine dinucleotide (FAD) have a redox cycle that does not depend on a transition state metal for binding of molecular oxygen and generally a narrower substrate selectivity that is limited to soft nucleophiles, and, preferably, uncharged or single positive charged compounds with an electron-rich center (usually nitrogen or sulfur, but phosphorous and selenium can also provide the electrons) (Ziegler, 1993). The so-called "external monooxygenases" are the most common form of FMO and depend on NADPH or nicotinamide adenine dinucleotide (NADH) as external factors to reduce the flavin (flavin mononucleotide, FMN, or FAD) to the reactive form that binds molecular oxygen. In contrast, the rarely encountered "internal monooxygenases" can use the substrate itself

| Functional Group | Example |
|---|---|
| Thiols (e.g. methimazole) | 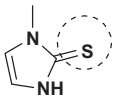 |
| Benzylamine (e.g. terbinafine) | 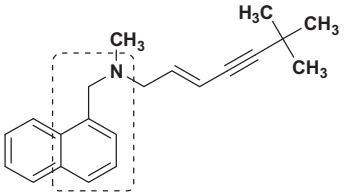 |
| Thiazole (e.g. sudoxicam) | 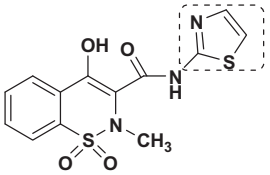 |
| Carboxylic acid (e.g. diclofenac) | 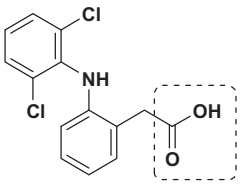 |
| Epoxide (e.g. phenytoin) | 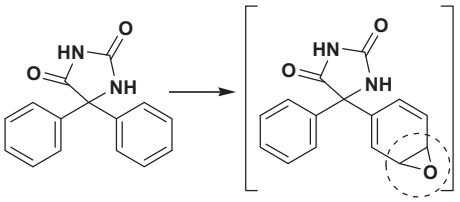 |

to reduce the flavin (lactate monooxygenase is one example of this type of FMO) (van Berkel *et al.*, 2006).

In humans, five functional FMOs are known and six pseudogenes, which are considerably fewer than the ~57 CYPs, are reported to date. The FMO family is expressed at the highest levels in the liver. The individual FMOs, however, show distinctive differences in their relative expression across tissues. FMO1 is predominantly expressed in fetal liver and adult human kidney and intestine. In kidney microsomes, FMO1 is expressed usually at the highest levels (FMO1 > FMO5 > FMO3) and FMO1 is seemingly greater in kidney microsomes prepared from African-Americans than from Caucasians (Krause, Lash and Elfarra, 2003). FMO2 is inactive in Caucasians and Asians, while an active full-length form of FMO2 is expressed polymorphically in the lung of

26% of African-Americans and 4.5% of Hispanics (Cashman, 2000; Whetstine *et al.*, 2000). In contrast, FMO3 and FMO5 are expressed at the highest levels in the liver and are the forms most likely to be involved in FMO-mediated drug metabolism (Cashman, 2000; Whetstine *et al.*, 2000; Krueger and Williams, 2005). The relevance of these findings is that if tissue-specific toxicity is indicated with a known FMO substrate, then tissue-specific differences in the expression of FMOs could underlie the mechanism in the detoxication or bioactivation of the drug. The ability to predict these findings in humans depends on correlating animal *in vitro* models with *in vivo* toxicity in those species, and finding similarities between those mechanisms and the biochemical toxicology indicated in human *in vitro* models.

Far fewer drugs are metabolized by FMOs than by CYPs and very few drugs are cleared primarily by FMOs—the exceptions include compounds like xanomeline, itopride, benzydamine, and olopatiadine, with each drug cleared by formation of its *N*-oxide metabolite catalyzed by the human FMO1 and FMO3 forms (Ring *et al.*, 1999; Lang and Rettie, 2000; Störmer *et al.*, 2000; Mushiroda *et al.*, 2001; Kajita *et al.* 2002). Consequently, the role of FMOs in toxicity is limited when compared with CYPs or other DMEs; however, there are notable examples that exemplify their importance on a case-by-case basis. For instance, monoamine oxidases (MAOs) catalyze the oxidation of the neurotoxin MPTP (1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine) to form the highly toxic MPDP+ (1-methyl-4-phenyl-2,3-dihydropyridine) and MPP+ (1-methyl-4-phenylpyridine) metabolites that block mitochondrial transport, destroy the nigrastriatal dopaminergic pathways, and elicit a rapid onset of Parkinsonism (Davis *et al.*, 1979; Langston *et al.*, 1983; Williams and Ramsden, 2005). CYP2D6, CYP1A2 and, to some extent, CYP3A4 have been shown to form the *N*-desmethyl metabolite (4-phenyl-1,2,3,6-tetrahydropyridine, PTP) and the *N*-oxide metabolite (1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine *N*-oxide) that serve to inactivate MPTP (Weissman *et al.*, 1985; Coleman *et al.*, 1996). FMO in the brain has also been shown to inactivate MPTP via formation of the *N*-oxide metabolite (Weissman *et al.*, 1985; Mushiroda *et al.*, 2001). It is possible that FMO-mediated clearance, in addition to supporting CYPs in protecting against neurotoxicity and hepatotoxicity from MPTP exposure, could potentially play an even more significant role in the inactivation of MPTP in CYP2D6 poor metabolizers. FMOs are also involved in the hepatotoxicity associated with *N*-deacetylketoconazole and the sulfoxidation of thioureas to form sulfenic acids (the latter subsequent to the formation of the reactive sulfinic acid, ethionamide) (Rodriguez and Miranda, 2000; Rodriguez and Buckholz, 2003; Henderson *et al.*, 2004). These reactions lead to redox cycling and depletion of glutathione and NADPH cofactors that affect the metabolism of the culpable agents and potentially that of other xenobiotics (Henderson *et al.*, 2004). FMOs also oxidize secondary and primary amines to form hydroxylated metabolites that can form a complex with hemoproteins in rat tissues, with inactivation of CYPs and ensuing drug–drug interactions and concomitant issues (Jönsson and Lindeke, 1992).

There are few drug–drug interactions associated with FMOs since no inducers have been found for FMOs and few drugs are metabolized exclusively by FMOs. Variability in FMO activity tends to be associated with polymorphic expression, with the best known example being FMO3. This form exclusively catalyzes the *N*-oxidation and clearance of trimethylamine (TMA), a product of choline (Cashman *et al.*, 1997; Lang *et al.*, 1998). In Caucasians, about 0.1% to 1% of the population is deficient in FMO3 and exhibits a "slow" phenotype for clearance of TMA. In these individuals, TMA is secreted in urine, sweat, and breath giving rise to the "fish odor" syndrome. The next most significant polymorphism is linked to FMO2 and the polymorphic expression of

an active form in human lung. If a polymorphic form of FMO is significantly involved in the elimination and/or bioactivation of a drug, then it is plausible that certain individuals will have an inherently greater predisposition toward toxicity, either via the accumulation of toxic drug or formation of higher levels of reactive metabolites.

### 2.3 Uridine Diphosphate Glucuronosyltransferases

The UGTs are another class of membrane-bound enzymes that are expressed in the liver and intestine and often play an important role in the first-pass effect. UGTs catalyze the transfer of glucuronic acid from a nucletoide cofactor, uridine diphosphate glucuronic acid (UDPGA), to specific functionalities (-$NH_2$, -OH, and -COOH) on substrates. This typically generates a more polar metabolite that is amenable to biliary and/or renal elimination. Since UGTs are located on the luminal side of the endoplasmic reticulum, microsomal preparations require "activation" by pore-forming agents such as the antibacterial alamethacin, the detergent Brij58, or by mechanical disruption (e.g. sonication). The use of alamethacin is preferred since it is nonspecific and activates all UGTs.

UGTs, like CYPs, are classified by primary amino acid sequence into families and subfamilies. In humans, two gene families have been identified with a growing number of members that are currently grouped into three subfamilies (UGT1A, UGT2A, and UGT2B). Unlike CYPs, selective inhibitors (chemical or antibody) or substrates for a specific UGT are lacking, which makes it very difficult to identify the exact UGTs involved and their contribution to the total metabolic clearance. This apparent lack of selectivity of UGT substrates does, however, prove advantageous in clinical practice, with a relatively low incidence of UGT-related drug–drug interactions reported in humans (Williams *et al.*, 2004).

Both UGTs and CYPs are co-localized in the endoplasmic reticulum and evidence exists for protein interactions between CYPs and UGTs that can affect glucuronidation—for instance, the ratio of morphine-3-glucuronide to morphine-6-glucuronide is affected by CYP3A4 content (Takeda *et al.*, 2005). Complexities like this, coupled with the atypical kinetics observed with UGT reactions, issues with optimizing protein and phospholipid content, inhibition by fatty acids and the need for activation, and other conditions for metabolic turnover, make it difficult to accurately extrapolate from *in vitro* systems to *in vivo* models (Miners *et al.*, 2006). This is extremely unfortunate because, in addition to substrates cleared directly via UGT-mediated conjugation (e.g. the 1,4-benzodiazepine anxiolytic, (*S*)-oxazepam and the HIV integrase anti-AIDS drug, raltegravir), many metabolites that are formed by CYPs (e.g. phenols and carboxylic acids) are themselves substrates for UGT-mediated glucuronidation (Abernethy *et al.*, 1983; Kassahun *et al.*, 2007). What this means is that even though UGTs may not be directly responsible for the clearance and/or therapeutic dose of a drug, they are often involved in the overall metabolism. This becomes especially important should the metabolites be implicated in adverse reactions, for example, diclofenac acyl-glucuronide and its potential to react with proteins.

Although UGT2B7 is more often implicated in drug clearance than any other UGT enzyme, the most frequent issue concerning UGTs involves the potential for inhibition of UGT1A1 and effects on the clearance of endogenous substrates such as estradiol, estriol, and the blood breakdown product, bilirubin (Ritter, Crawford and Owens, 1991; Court, 2005). UGT1A1 inhibition can slow the clearance of bilirubin causing an accumulation of unconjugated bilirubin, which leads to hyperbilirubinemia and toxicity. People with genetic polymorphisms of UGT1A1 expression exhibit increased levels of

unconjugated bilirubin. A profound (90%) reduction in bilirubin clearance is seen in patients with the rare Crigler–Najjar syndrome, whereas a moderate (40%) decrease is reported on average for patients with the more commonplace Gilbert's syndrome, which is seen in 10% of the US population (associated with the UGT1A1*28 homozygous form) (Tukey and Strassburg, 2000). Drug-induced inhibition of UGT1A1 in humans can present symptomatically like Crigler–Najjar or Gilbert's syndrome and profound hyperbilirubinemia is observed in UGT1A1-deficient individuals. An example of where UGT1A1 genotype or phenotype has implications on human safety and drug tolerability is irinotecan, an ester camptothecin derivative used to treat metastatic colon cancer. Irinotecan is first metabolized by carboxyesterases to form SN-38, a carboxylic acid and the active chemotherapeutic agent that is subsequently cleared by glucuronidation mediated by UGT1A1. UGT1A1-deficient patients show increased incidence and severity of neutropenia and diarrhea likely stemming from prolonged exposure to SN-38 (Toffoli and Cecchin, 2004; Toffoli *et al.*, 2006).

Acyl glucuronides are an interesting class of metabolite, formed by glucuronidation of carboxylic acids, because of their (i) propensity to form covalent adducts with proteins and (ii) potential to hydrolyze and regenerate the carboxylic acid. Intramolecular chemical rearrangement of the biosynthetic 1-*O*-β-acyl glucuronide can form 2-, 3-, and 4-*O*-β-isomers via migration of the acyl group. These acyl glucuronide β-isomers can subsequently anomerize via ring opening and formation of planar aldehyde intermediates followed by recyclization to the 2-*O*-, 3-*O*-, and 4-*O*-α acyl glucuronide isomers. Reactivity can occur via transacylation reactions with the initial acyl glucuronide or via glycation of target proteins such as albumin. The proposed reaction scheme is shown in Fig. 1 (Garlick and Mazer, 1983; Smith, McDonagh and Benet, 1986; Smith, Benet and McDonagh, 1990). Adverse drug reactions associated with acyl glucuronides are seen with many nonsteroidal anti-inflammatory drugs (NSAIDs), including ibuprofen, diclofenac, naproxen, and aspirin (Sanford-Driscoll and Knodel, 1986). In fact, Bailey and Dickinson reported that carboxylic acids account for 25% of drug withdrawals associated with severe toxicity, including zomepirac, ibufenac, and benoxaprofen (Bailey and Dickinson, 2003).

The propensity for acyl glucuronides to cause toxicity and/or immune responses necessitates diligent analysis of reactivity *in vitro*. Recently, lysine-phenylalanine dipeptide was shown to act as a possible surrogate for proteins and offers a useful *in vitro* tool for screening acyl glucuronides for reactivity (Wang *et al.*, 2004). In terms of structure–activity relationships and optimization of medicinal chemistry, the reactivity can be circumvented by substituting a potentially more stable group for the carboxylic acid moiety (e.g. an amide) and/or developing a drug that can be administered at low doses and/or cleared by multiple mechanisms other than UGT.

Another complexity associated with acyl glucuronides is the tendency of 1-*O*-β-acyl glucuronides to be amenable to hydrolysis, more so than the transacylated isomers. Serum albumin, hydroxyl anion, and esterases can hydrolyze acyl glucuronides back to the drug and this requires additional precautions to be taken in bioanalysis and increases the cost of sample analysis and potential inaccuracies. Moreover, β-glucuronidases in the intestine can hydrolyze acyl glucuronides to regenerate drug or metabolite, which if reabsorbed results in the phenomenon of enterohepatic recirculation (EHR), such as seen with valproic acid in the rat (Pollack and Brouwer, 1991; Slattum, Cato and Brouwer, 1995). EHR can significantly prolong the residence time of a drug and add to difficulties in predicting exposure in humans from preclinical animal data. Lastly, consideration also must be made toward pH-dependent hydrolysis

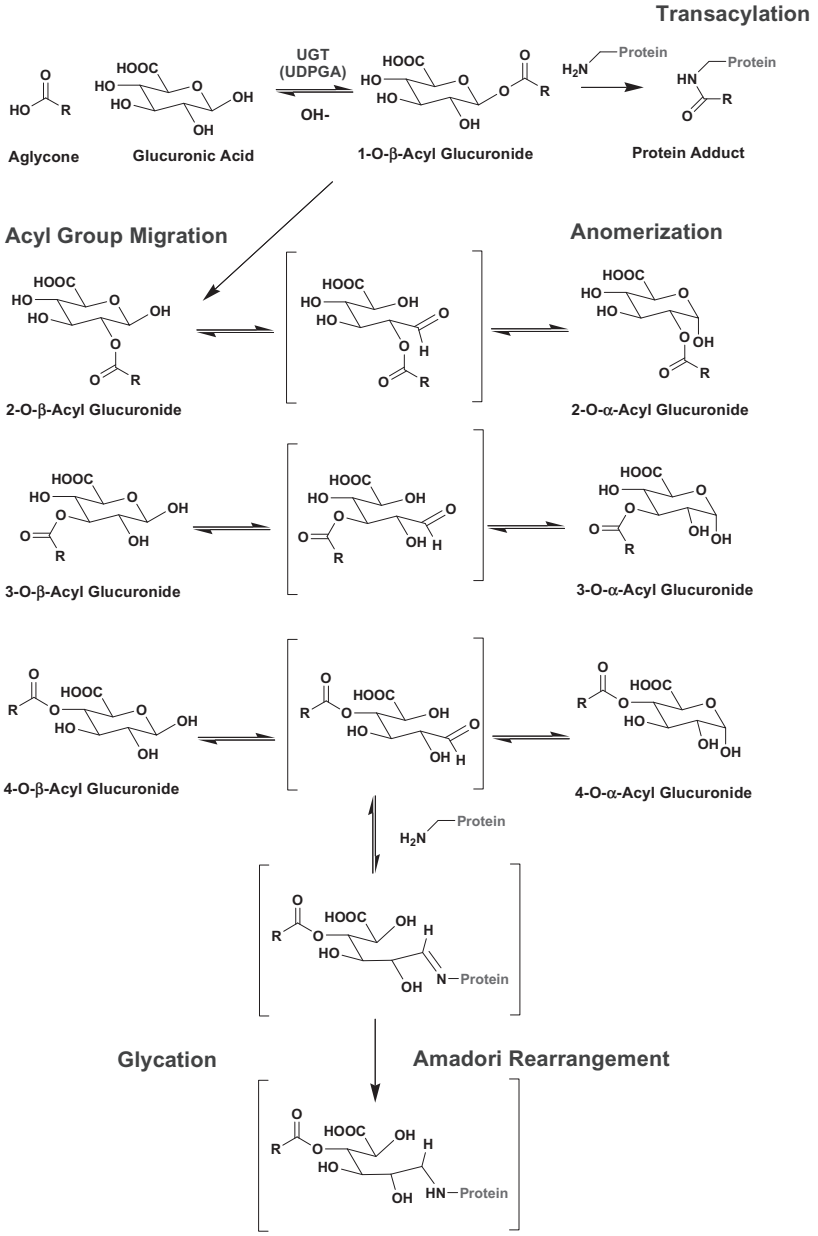


**Figure 1** Reactivity scheme for acyl glucuronides. Acyl glucuronide can react with target nucleophiles directly via transacylation or by glycation via acyl group migration, with formation of an intermediate aldehyde and Amadori rearrangement.

and its effect on anomers. Blood and excreta samples may need to be acidified to allow an accurate quantitation of drug, acyl glucuronide metabolite, and its isomers, which adds bioanalytical cost and increases variability and uncertainty of the results.

### 2.4 Arylamine *N*-Acetyl Transferases

Arylamine *N*-acetyl transferases (EC 2.3.1.5) (NATs) are expressed in humans in two forms, NAT1 and NAT2, each the product of a distinct gene locus. Unlike the DMEs discussed thus far, NATs are soluble, cytosolic enzymes (*in vitro* assays need to be optimized accordingly). NAT1 and NAT2 share 87% identity in their nucleotide sequence and 81% homology in their amino acid sequence. At least 26 allelic variants of NAT1 and 36 variants of NAT2 have been identified to date (Butcher *et al.*, 2002). NAT1 is expressed ubiquitously in humans, with its highest levels of mRNA found in the urothelium and colon epithelial cells, whereas NAT2 is found mainly in the liver and intestine (Debiec-Rychter, Land and King, 1999). With the exception of *N*-acetylation of cysteine conjugates to form mercapturic acids, NAT1 and NAT2 metabolize predominantly aromatic amines. Both forms use acetyl CoA as the cofactor and have different but overlapping substrate specificities. For instance, *p*-aminobenzoic acid and sulfamethazine are relatively selective probe substrates of NAT1 and NAT2, respectively.

Understanding polymorphisms and interindividual differences in expression of both human NAT forms is an important part of the safe evaluation of drugs that serve as their substrates. Although NAT1 was initially considered to be a monomorphic enzyme—albeit one with a wide range in expression and activity between patients—it is now known to have a bimodal polymorphic distribution in humans, with approximately 8% of individuals found to be slow acetylators (Vatsis and Weber, 1993). The slow phenotype is attributed to NAT1*14 and NAT1*17 alleles and lower expression/reduced active form, with NAT1*14 the more frequently occurring (up to 3.8% in Caucasians). A higher occurrence of the NAT1*14 allele was seen in a Lebanese population, where up to 50% of the population is a slow acetylator (Dhaini and Levy, 2000).

When compared with NAT1, the NAT2 form is more distinctive in its polymorphism. The observed bimodal distribution in the plasma exposure of isoniazid—now associated with NAT2—is one of the earliest reported drug phenotypes (Peters, Miller and Brown, 1965). This bimodal distribution is very well illustrated in a relatively recent study of isoniazid plasma exposure in genetically unrelated German subjects (Weinshilboum and Wang, 2004). NAT2 displays an interesting distribution across different ethnicities with a slow acetylator phenotype of 5% in Canadian Eskimos, 10–20% in Asians, 40–70% in Europeans/Caucasians, 80% in Egyptians, and 90% in Moroccans (Evans, 1989). Drugs such as isoniazid, sulfamethazine, sulfadiazine, diaminodiphenyl sulfone, hydralazine, phenelzine, sulfapyridine, sulfamethoxypyridazine, and dapsone are metabolically cleared by NAT2 and should be administered with some regard to the target population to avoid overexposure and minimize the incidence of adverse effects.

NAT1 and NAT2 both catalyze the acetylation of arylamines or hydralazine, or catalyze the *O*-acetylation of a hydroxylamine. This usually serves as a detoxification mechanism in limiting exposure of xenobiotics, illustrated by the greater incidence of bladder cancer, isoniazid-induced neuropathy, and sulfamethoxazole- or dapsone-induced Stevens–Johnson syndrome and quicker onset of positive antinuclear antibody (ANA) syndrome following procainamide therapy in NAT2 slow acetylators. However,

NATs are also implicated directly in the bioactivation of procarcinogens, for example, biotransformation of aromatic and heterocyclic amines (e.g. 4-aminobiphenyl, 2-naphthylamine, benzidine, and cooked meat-derived heterocyclic amines) and also an increased risk of amonafide-induced myelotoxicity in fast acetylators (Innocenti, Iyer and Ratain, 2001).

Bioactivation of aromatic amines is an example where the metabolite of one enzyme reaction can often serve as the substrate for another as part of the mechanism of drug toxicity. Two mechanisms are proposed by which aryl amines can be bioactivated by NATs to an ultimate electrophilic aryl nitrenium ion, each requiring a sequential role for CYP and NAT. The first mechanism involves CYP-mediated *N*-hydroxylation to form a hydroxylamine with subsequent NAT-mediated *O*-acetylation to form *N*-acetoxyarylamine. The second mechanism proposed is initiated by direct *N*-acetylation to form an *N*-acetyl arylamine that undergoes CYP-mediated *N*-hydroxylation to form an *N*-acetohydroxamic acid, followed by *N,O*-transacetylation to form the *N*-acetoxyarylamine. In each case, heterolytic cleavage of the N–O bond of the common *N*-acetoxyarylamine product leads to aryl nitrenium formation and reactivity with target nucleophiles such as DNA and proteins—such adducts thus have potential to cause toxicity in humans and animals (Fig. 2). Today, there is a plenitude of evidence marking this functionality as a clear structural alert. Therefore, even though steric and electronic factors can block the reactive sites of metabolism and low-dose and acute administration regimens may obviate the risk of bioactivation, unembedded aryl amines are usually avoided in drug discovery and development.

NATs provide an interesting example of a distinct species difference in drug metabolism. The liver cytosol from domestic dogs and other canids (e.g. wolf, jackal, and fox) does not show any expression of NATs and this species appears to be NAT-deficient (Trepanier *et al.*, 1997). Consequently, the same idiosyncratic sulfonamide toxicities seen in human can be observed in dogs (Cribb *et al.*, 1996). In the event that an NAT substrate is developed as a drug, the dog is worth consideration as a preclinical model for the evaluation of *N*-acetylation and its role in pharmacokinetics (PK), pharmacodynamics (PD), and/or toxicity (Lavergne *et al.*, 2006).

## 2.5 Sulfotransferases

SULTs (EC 2.8.2) are another class of cytosolic enzymes that catalyze metabolic clearance via conjugation of xenobiotic and endogenous substrates. There is an overlap between SULTs and UGTs in the substrates they metabolize, with SULTs mediating the sulfation of hydroxy or amine functional group present in compounds such as catecholamines, neurotransmitters, steroids, and thyroid hormones. There are also

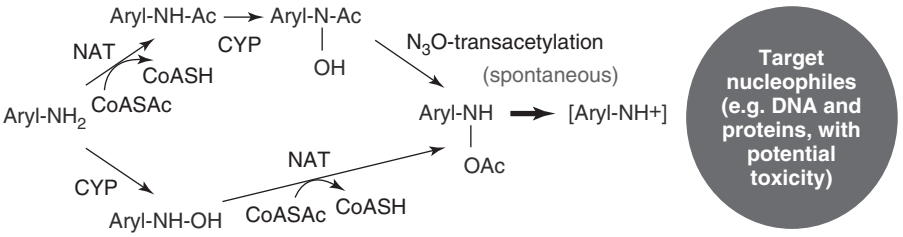


**Figure 2** Reactivity scheme for *N*-acetyl transferases.

membrane-bound SULTs that are found in the Golgi apparatus: these tend not to be significant with respect to the metabolism of typical small-molecular-weight drugs; however, they are involved in the sulfonation and regulation of activity associated with proteins, peptides, lipids, and glycosaminoglycans (Baeuerle and Huttner, 1987; Huttner, 1988).

SULTs are found in many species including plants, birds, and mammals. To date, there are four families of human SULTs: SULT1 has four subfamilies and eight members (SULT1A1, 1A2, 1A3, 1A4, 1B1, 1C2, 1C4, 1E1); SULT2 has two subfamilies and three members (2A1 and 2B1a, 2B1b, which result from alternate slicing of the SULT2B1 gene); SULT4 has one subfamily and one member (SULT4A1); and SULT6 has one subfamily and one known member (SULT6B1). (Families share ≥45% amino acid sequence identity; subfamilies share ≥60% amino acid sequence identity.) Recently, 217 polymorphisms were found in the sequences of 118 DNA samples, including 64 polymorphisms in the 8785-bp coding regions of all 12 SULT genes. Of these, 39 were nonsynonymous polymorphisms that alter the functional activity of the expressed enzyme, with evidence for lowered expression of immunoreactive protein rather than expression of mutant forms with altered enzyme kinetics (e.g. SULT1E1*2 showed 13% expression of the wild-type protein and 7% of its activity, SULT1C2*5 showed no expression, whereas SULT1A3/4*5 showed 120% of wild-type protein expression with 110% of its activity) (Hildebrandt *et al.*, 2007).

SULTs are expressed in many tissues with some differentiation seen between isozymes: SULT1A1 is expressed in adult liver and platelets; SULT1A3 is expressed in placenta, fetal liver, and adult platelets; and SULT 4A1 has been found only in the brain (Coughtrie, 2002). An interesting difference between SULTs and enzymes such as CYPs or UGTs is a relatively high expression of SULTs in the fetus, which points to a role in embryogenesis and/or development—a potential consideration should embryotoxicity be indicated in safety studies.

SULTs are not commonly involved as the principal clearance mechanism yet they do play a significant role in the overall biotransformation and excretion balance of a drug. With regard to direct clearance, 17α-ethinyl estradiol (EE), troglitazone, and acetaminophen are three drugs that involve SULTs in their *O*-sulfation (they also show a possible role for SULTs in their toxicity); cisapride and DPC-423 are examples of drugs that undergo *N*-sulfonation reactions (Back *et al.*, 1982; Slattery *et al.*, 1987; Meuldermans *et al.*, 1988; Kawai *et al.*, 1997; Mutlib *et al.*, 2002). It should be noted that sulfonation does not necessarily inactivate a drug—morphine-6-sulfate is an example of an active sulfonate conjugate (Zuckerman *et al.*, 1999).

PAPS (3′-phosphoadenosine-5′-phosphosulfate) serves as the typical donor of the sulfuryl or sulfonate group ($SO_3^-$). In most tissues, only 4–80 μM of PAPS is found, which is low compared with the 200 μM levels reported for UDPGA and even lower in comparison with the 2–10 mM typical of GSH. Drugs such as acetaminophen that are given at high doses can quickly deplete the sulfonation pathway and dose-dependent kinetics are seen in its biotransformation. This phenomenon was shown in rats dosed with acetaminophen. In humans, however, increased plasma levels of acetaminophen were not seen and induction of the glucuronidation pathway seemingly compensates for the depletion of PAPS (Galinsky and Levy, 1981; Gelotte *et al.*, 2007).

SULTs are able to bioactivate procarcinogens and genotoxic agents including *N*-hydroxyacetyl aminofluorene and 1-(α-hydroxyethyl) pyrene. The latter provides an example of enantioselective metabolism and bioactivation, with 160-fold ratio between the *S*- and *R*-enantiomers for mutagenicity in *Salmonella typhimurium* TA1538 express-

ing human SULT1E1 (Glatt, 2000). Sulfuric acid esters of benzylic and allylic alcohols, in particular, can be susceptible to bioactivation as a result of heterolytic cleavage of the electron-withdrawing sulfate group and formation of an electrophilic cation that is stabilized by mesomerism. This same mechanism is also implicated in the activation of hydroxylamines and hydroxamic acids, which suggests a possible interplay between SULTs and NATs in the metabolism-mediated toxicity associated with aryl amines (Glatt, 1997; Glatt *et al.*, 1998a, b).

SULTs can play a key role in the PK/PD of a new drug and should not be overlooked when assessing drug metabolism and predicting its effects in humans. For instance, hydroxytamoxifen is a potent active metabolite of tamoxifen inactivated by SULT1A1-mediated sulfonation. The variability in efficacy and apparent resistance to tamoxifen in breast cancer patients is attributed, in part, to interindividual differences in functional activity associated with sulfonate conjugation (Hayes and Pulford, 1995). In contrast, the hair-growth stimulant minoxidil is an *N*-oxide prodrug activated by SULTs specifically, to form a sulfonate conjugate that is more hydrophobic and functionally active than minoxidil as a result of an intramolecular H-bond (Falany and Kerl, 1990).

### 2.6 Peroxidases

Peroxidases (EC 1.11.1) involved in drug metabolism are typically heme-containing DMEs that have a characteristic ability to catalyze one-electron oxidations of substrates (e.g. aromatic amines) that produce free radicals with potential to propagate and terminate reactions. Peroxidases are classified into two types based on the mechanism of inactivation elicited by the common substrate phenylhydrazine. Phenylhydrazine forms phenyl radicals that can characteristically bind to the heme-iron of peroxidases such as CYPs, prostaglandin H synthase, catalase, and hemoglobin *inter alia*, or characteristically bind instead to the δ-meso heme edge and protein in the case of peroxidase such as myeloperoxidase (MPO), eosinophil peroxidase (EPO), and lactoperoxidase (LPO) (Tafazoli and O'Brien, 2005).

MPO is a lysosomal peroxidase found at high levels in the primary granules of neutrophils (2–5% of the dry weight) and monocytes, and is also found at significant levels in tissue macrophages (e.g. Kupffer cells in the liver). MPO is distinguished from the other peroxidases in its unique ability to oxidize chlorine to hypochlorous acid, a potent oxidizing agent with a key role in phagocytosis and function as an antimicrobial agent. Known substrates of MPO/HOCl include procainamide, clozapine, chlorpromazine, lamotrigine, dapsone, vesnarinone, sulfonamides, 3-hydroxy carbamazepine, 4-hydroxy phenytoin, and 7-hydroxy fluperlapine (Lai *et al.*, 2000; O'Brien, 2000; Lu and Uetrecht, 2007). MPO is released extracellularly into phagocytic vacuoles upon activation of neutrophils in an immune response (*in vitro* phorbol myristate acetate is often used to activate neutrophils). Hydrogen peroxide, formed by NADPH oxidase and superoxide dismutase (SOD), is also released and acts with chloride ion as cofactors in the MPO-mediated formation of hypochlorous acid. Hypochlorous acid itself is a powerful two-electron oxidant that can nonenzymatically oxidize many drugs and be directly responsible for tissue damage (Klebanoff, 1999). Hypothetically, this can elicit a "danger signal" and trigger idiosyncratic drug toxicity in susceptible individuals.

Infiltration of tissues by neutrophils or eosinophils can lead to increased local MPO activity and toxicity linked with hypochlorous acid and/or reactive oxygen species (ROS) formation, including oxygen radicals (e.g. indomethacin-induced kidney

toxicity) (Basivireddy *et al.*, 2004). For instance, a 50- to 100-fold increase in MPO activity was reported in the liver following infiltration of neutrophils triggered by the injury sustained by ischemia/reperfusion (Kato *et al.*, 2000). Pulmonary insults, such as tobacco smoke, cause neutrophil infiltration into the lung (Schmekel *et al.*, 1990). Benzo[*a*]pyrene is one of several toxins found in cigarette smoke; its 7,8-diol metabolite can be bioactivated by MPO to form the very reactive and carcinogenic diol-epoxide metabolite (Mallet, Mosebrook and Trush, 1991). MPO release may play a direct role in the bioactivation and toxicity of procarcinogens.

A $^{-}$463G→A polymorphism in the promoter region of the MPO gene is associated with decreased transcriptional activity and could result in lower expression and improved risk in lung cancer (Feyler *et al.*, 2002). Unfortunately, identifying a significant role in cancer for low-penetrance genes has proven to be notoriously difficult and counterevidence is available with no association found in the $^{-}$463G→A polymorphism based on data from 988 cases of lung cancer in Caucasians and 1128 control subjects (Xu *et al.*, 2002). The evidence does, however, suggest that measuring MPO activity in target tissues of toxicity could provide a useful marker of oxidative stress and a possible biomarker of bioactivation.

Activated neutrophils, MPO-$H_2O_2$-Cl and HOCl are useful tools for generating and identifying potential metabolites and often HOCl alone can be used as a facile method to form metabolites. DMP 406 (a novel antipsychotic) and mianserin (a tetracyclic antidepressant) are two examples for which HOCl alone does not replicate the metabolism seen with activated neutrophils and *in vivo* their oxidation may proceed instead by sequential MPO-mediated single-electron mechanisms (Iverson, Zahid and Uetrecht, 2002).

There are peroxidases other than MPO to consider in drug metabolism and mechanistic studies of biochemical toxicology. For instance, EPO is involved in the formation of hypobromous acid and associated with an increase in hepatotoxicity seen with acetaminophen and high bromide plasma levels. EPO activity is not limited to the blood and the presence of eosinophils in the rat intestine accounts for most of the peroxidase activity seen in that tissue. Just as an inflammatory response can increase MPO activity in the liver, infiltration by eosinophils in the liver can also occur (e.g. triggered by parasite infections) and effect a profound increase in EPO activity (De, De and Banerjee, 1986).

Catalases, another family of peroxidases distinctive in being able to disproportionate hydrogen peroxide to oxygen and water, play a regulatory role in control of hydrogen peroxide levels formed by MPO and EPO and so on. Catalases also have a direct role in drug metabolism; they are able to oxidize low-molecular-weight compounds such as alcohols (e.g. methanol and ethanol) and hydrazines (e.g. phenylhydrazine, phenelzine, and isoniazid) when concentrations of hydrogen peroxide are low (Artemchik, Kurchenko and Metelitsa, 1985).

Peroxidases are not selective in their substrate affinity: ellipticine, for example, is a topoisomerase inhibitor that forms covalent DNA adducts implicated in the cytotoxicity and anticancer effects. These adducts are formed by oxidation reactions that involve multiple peroxidases including human CYP1A1, 1A2 and 3A4, MPO, human COX-2, human renal microsomes incubated in the presence of arachidonic acid (COX activity), bovine lactate peroxidase, ovine COX-1, and horseradish peroxidase (Stiborová *et al.*, 2004, 2007). With regard to drug therapy, tumor and tissue differences in the expression of peroxidases could offer an opportunity to exploit local bioactivation (e.g. tumor-specific cytotoxicity) and possibly predict the potential susceptibility of humans to cancer or therapy.

### 2.7 Prostaglandin H Synthases/Cyclooxygenases

Prostaglandin H synthase (PHS), also known as cyclooxygenase (COX), is named for its role in the biosynthesis of prostanoids from arachidonic acid. PHS catalyzes the oxidation of arachidonic acid to form $PGG_2$ and then a peroxidase-like reduction to $PGH_2$. Subsequent synthases lead to the formation of various prostaglandins and thromboxanes that feature in cell regulation and differentiation. In humans, there are two forms of PHS, each encoded by a single distinct gene: PHS-1 (COX-1) is found on the luminal surface of the endoplasmic reticulum and PHS-2 (COX-2) is found in the endoplasmic reticulum as well as in the inner and outer membranes of the nuclear envelope. They share approximately 60% of their amino acid sequence identity and have overlapping substrate affinities.

PHS-1 is considered to be a housekeeping gene with well-controlled basal levels expressed in most cells. In contrast, PHS-2 has significant basal levels in select cells only (e.g. brain hippocampus), yet upon inflammation, it is highly inducible in several cell types, a mechanism believed involved in the development of inflammatory responses such as pain and fever. Although NSAIDs work by inhibiting PHS-2, the NSAIDs (ibuprofen and sulindac) and peroxisome proliferators also cause its induction, which could influence the metabolism of coadministered drugs. The regulation of PHS is complex and reviewed in depth by Vogel (2000).

PHS is implicated in the neurodegeneration associated with the amphetamine derivatives, MDMA ("ecstasy") and methamphetamine ("speed"). One hypothesis involves their bioactivation by CYP2D6 in the brain to a catechol intermediate that subsequently forms a quinone that engages in redox cycling (Tucker *et al.*, 1994; Lin *et al.*, 1997). A second hypothesis proposes that PHS in the brain functions as a hydroperoxidase, metabolizing methylenedioxy amphetamine derivatives such as MDMA and forming free radicals and ROS that elicit tissue damage and irreversible loss of monoaminergic nerve terminals and long-term neurodegeneration (Jeng *et al.*, 2006). In addition to these drugs, various other substrates can serve as electron-donating redox partners for PHS and in the process are metabolized to radical cations, for example, arylamines, indoles, estrogens, and heterocyclic amines (e.g. methylenedioxy amphetamine derivatives). Phenytoin, Phenobarbital, and *N*-desmethylated metabolites of mephenytoin and trimethadione are all teratogenic via embryonic PHS bioactivation (Parman, Chen, and Wells, 1998).

### 2.8 Glutathione Transferases

Glutathione transferases, also known as glutathione *S*-transferases (EC 2.5.1.18) (GSTs), are dimeric proteins that are primarily cytosolic but also have a mitochondrial form and a membrane-bound form known as "membrane-associated proteins in eicosanoid and glutathione metabolism" (Jakobsson *et al.*, 1999). These enzymes catalyze the conjugation of glutathione (γ-glutamyl-cysteinyl-glycine; GSH) with reactive soft electrophiles such as epoxides, haloalkanes, alkenes, and aromatic nitro groups. GSTs with ≥60% shared amino acid sequence identity are grouped into the same class and those with <30% shared identity are assigned their own class. Subfamilies of GSTs exist and share up to 90% sequence identity. The sequence at the *N*-terminus includes an important component of the active site and is also taken into consideration when classifying the GSTs. The Alpha-, Mu-, Pi-, and Theta-class of cytosolic GSTs are encoded by distinctive genes. (Zeta-, Omega-, Sigma-, and Kappa-GST forms also exist and are

found in mitochondria and peroxisomes.) The Alpha-, Mu-, and Pi- forms are most similar in terms of structure, substrate affinity, and shared effects of inhibitors and are the classes most frequently involved in drug metabolism reactions (Sheehan *et al.*, 2001). The human forms identified to date include GSTA1-A4, GSTM1-M5, and GSTP. Various allelic forms exist for the GSTs including null allele polymorphisms linked with GSTM1-1 and GSTT1-1. In the case of GSTM1*0, no GSTM1-1 protein is expressed in approximately 50% of the Caucasian and Asian populations and for GSTT1*0, no GSTT1-1 protein is expressed in 10% of Mexican-Americans and up to 60% of East Asians. These individuals are potentially more susceptible to toxicity from substrates otherwise detoxified by these GST enzymes (Nakajima *et al.*, 1995; To-Figueras *et al.*, 1996; Bolt and Thier 2006).

There are some differences known in the tissue expression of UGTs, which could prove useful in understanding the mechanism of toxicity. GSTA1-1, A2-2, A4-4, M1-1, M3-3, M4-4, K1-1,O1-1, T1-1, and Z1-1 are all expressed in the liver; GSTA1-1, A2-2, A4-4, M1-1, M3-3, K1-1,O101, P1-1, T1-1, and Z1-1 are expressed in the small intestine and GSTP1-1 and T1-1 are found in erythrocytes (Mannervik *et al.*, 2005). Based on its known selectivity, the GSH conjugation of methyl halides in erythrocytes is a useful peripheral marker of GSTT-1 phenotype and genotype (Pemble *et al.*, 1994).

Most substrates of GSTs are also able to undergo a nonenzymatic reaction with GSH, which needs to be accounted for when predicting the role and contribution of the GST-mediated reaction to the overall metabolic clearance of a new drug. In drug discovery, screens are often used to identify early on potential issues with a new drug or chemotype. One such assay uses liver microsomal incubations performed in the presence of GSH as a trapping agent for the formation of reactive electrophiles. Although it is rather difficult to predict the consequences of forming GSH conjugates *in vitro* (vis-à-vis, many drugs form GSH adducts yet are generally considered to be safe), since this is principally a detoxification mechanism the presence of adducts is taken as evidence of formation of a reactive species and the key assumption made in medicinal chemistry is that when it comes to GSH adducts, less is better and none is ideal. GSH conjugates usually are excreted in the bile and/or urine either intact or as cysteine-glycine, cysteine, *N*-acetyl cysteine (NAC), and mercapturic acid metabolites. Therefore, when evaluating the *in vivo* biotransformation of a drug in human or animal models, the presence of any one of these metabolites can serve as a marker of GSH conjugation and potential formation of reactive drug-related species. In drug discovery, usually these studies are supported by *in vitro* liver microsomal covalent-binding experiments performed in the absence and presence of GSH. Since GST expression and activity can be induced (e.g. by phenobarbital, 3-methylcholanthrene, and β-naphthoflavone), pretreated *in vitro* or *in vivo* systems can provide additional insight into the mechanism of toxicity associated with the formation and detoxification of reactive electrophiles.

With GSH measured at up to 10 mM in cells, including hepatocytes, GSTs serve as a high-capacity detoxification mechanism that works well except in the event of significant drug overdose, such as acetaminophen toxicity and mitochondrial GSH depletion. The risk of toxicity is potentiated when acetaminophen is combined with ethanol use and alcoholics are more susceptible to acetaminophen toxicity than nonalcoholics (Seeff *et al.*, 1986; Zimmerman and Maddrey, 1995). The toxicity of acetaminophen is very well studied and known to result from the interplay of several factors. Ethanol is especially effective at potentiating adverse reactions since it has been shown to affect both the import of GSH from the cytosol into the mitochondria and also induces

CYP2E1 activity, with the latter responsible for an increase in the formation of the reactive electrophile *N*-acetyl-*p*-benzoquinoneimine (NAPQI) metabolite of acetominophen that, when unchecked by conjugation with GSH, reacts with mitochondrial proteins leading to toxicity (Zhao and Slattery, 2002).

Depletion of GSH is associated with toxicity from pyrrolizidine alkaloids, clozapine and amodiaquine (Williams *et al.*, 1997; Naisbitt *et al.*, 1998; Fu *et al.*, 2004). Although not itself a drug, acrolein is an end product of lipid peroxidation and a potentially toxic fragment of several drugs, including cyclophosphamide. Acrolein is potently toxic to the mitochondria, affecting respiratory function and cell function; however, it is also a substrate of GSTP1-1 (Hayes, Flanagan and Jowsey, 2005). As an added complexity, when GSH levels are low, acrolein was shown to inactivate GSTP1 which could have consequences for other coadministered substrates of GSTs (van Iersel *et al.*, 1997).

The glutathione redox cycle represents a protective mechanism helping to regulate the levels of hydrogen peroxide formed in inflammatory response. Glutathione peroxidase catalyzes the oxidation of GSH by $H_2O_2$ to form a glutathione disulfide (GSSG). Glutathione reductase can then reduce GSSG back to GSH to begin a new cycle (Fig. 3). This redox cycle can be disrupted in several ways: for example, GSH depletion by reactive electrophiles, as described previously; inhibition of the synthesis of GSH (e.g. buthionine sulfoxime); and/or inhibition of the reduction of GSSG (e.g. 1,3-*bis*-(chloroethyl)-1-nitrosourea) (Harlan *et al.*, 1984). Interindividual variability in these processes, bioactivation mechanisms, and intrinsic susceptibility to drugs and/or toxins could factor into the expression of adverse effects including idiosyncratic drug reactions.

Induction of GSTs, GSTP1-1 in particular (characterized by significant basal expression in most tissues except hepatocytes), is linked to an increased resistance to chemotherapeutic agents (Mannervik *et al.*, 1987; Hamada *et al.*, 1994). This provides an opportunity in drug discovery to exploit GSTs in the bioactivation of chemotherapeutic prodrugs, such as described for the thiopurine prodrugs, cis-6-(2-acetylvinylthio)purine (cAVTP) and trans-6-(2-acetylvinylthio)guanine (tAVTG), which become conjugated with GSH prior to formation of the cytotoxic metabolites, 6-mercaptopurine (6-MP) or 6-thioguanine (6-TG), respectively, and release of *cis*-, *trans*-, and *bis*-glutathione-

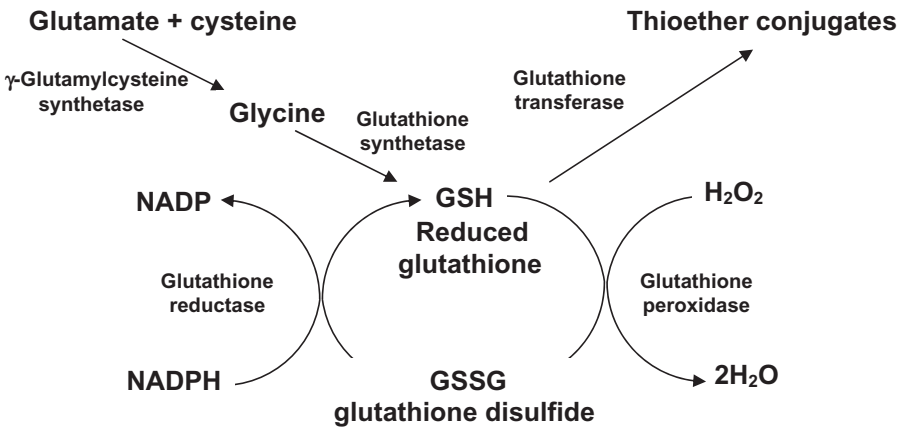


**Figure 3** Glutathione redox cycle (adapted from Harlan *et al.*, 1984).

butenone (Eklund *et al.*, 2007). A similar reaction was reported with azathioprine, with 6-MP released and associated with cytotoxic effects and adverse reactions in individuals with high GST activity (Eklund *et al.*, 2006). Because GSTs are involved in the biosynthesis of endogenous molecules (e.g. steroids, leukotrienes, and prostaglandins) and help regulate cell signaling and expression of genes trans-activated by peroxisome proliferators, control of GST activity is a potential target for pharmacology. Perturbation of these processes by GSH depletion, GST inhibition, or GST induction could lead to toxicity following exposure to a GST substrate.

### 2.9 Other Drug-Metabolizing Enzymes

CYPs, FMOs, UGTs, NATs, SULTs, and peroxidases such as MPO are all part of a detoxification defense system that has evolved to protect humans and other species from accumulation of xenobiotics and endobiotics. Most of the time these enzymes effectively limit exposure of organs to xenobiotics, including drugs. Yet examples exist for all these enzymes being involved directly in the development of drug toxicity. GSTs generally represent essentially a detoxification pathway catalyzing the conjugation of GSH with many of the reactive electrophilic species formed in reactions mediated by the aforementioned enzymes. In contrast, toxicity associated with GSTs mostly concerns their deficiency (e.g. null polymorphism of GSTM1-1) or depletion of the GSH cofactor, rather than any specific metabolite that is formed via GST-mediated catalysis.

The aforementioned families of DMEs are those most commonly encountered in adverse reactions and can offer a significant challenge in understanding toxicity and the risk associated with a drug and, although not discussed in any detail, the relevance of animal models to human toxicity. Many other enzymes (e.g. epoxide hydrolases and monoamine oxidases) are known to be involved in drug metabolism; this adds to the difficulty faced by toxicologists and medicinal chemists. Fortunately, many of these enzymes are quite selective in their substrates. Therefore, the structure of the drug and/or its metabolite(s) provides an initial, simple indication of whether these enzymes participate in biotransformation, elimination, and/or toxicity.

For instance, epoxides represent an electrophilic species commonly formed from drugs and environmental agents, and often are subject to nucleophilic attack from GSH and, in parallel, hydrolysis catalyzed by the epoxide hydrolases (microsomal, mEH or HYL1; and cytosolic, cEH or HYL2) (Arand *et al.*, 2005). Epoxide hydrolases can react with epoxides to form polar dihydrodiol metabolites that tend to be unreactive and readily excreted. Epoxide hydrolases therefore usually represent an effective detoxification pathway, with the exception of epoxides of polycyclic aromatic compounds (e.g. benzo[*a*]pyrene) that become highly reactive and carcinogenic diol epoxides. Other examples are the mitochondrial monoamine oxidases (MAO-A, MAO-B) that can metabolize primary, secondary, and tertiary amines (e.g. sumatriptan). The MAOs are also able to form reactive aldehydes and metabolize the inert tertiary amine MPTP to the reactive $MPP^+$ metabolite that destroys the brain's dopaminergic pathways and induces Parkinson's disease in humans (Davis *et al.*, 1979; Langston *et al.*, 1983).

Other classes of DMEs include alcohol dehydrogenase (ADH) and aldehyde dehydrogenase (ALDH) (cytosolic enzymes that metabolize ethanol in a coupled reaction to form acetic acid), aldehyde oxidase (AO), xanthine oxidase (XO), carboxyesterases, superoxide dismutase, DT-diaphorase, NAD(P)H oxidase, carbonyl reductase, aldose reductase, and aldehyde reductase. The metabolite of one enzyme can become the substrate for another and their interplay can ultimately determine the individual risk

associated with reactive metabolites from drugs. For example, carbonyl reductase can reduce a carbonyl group to an alcohol that can theoretically undergo an oxidation reaction catalyzed by ADH to form an aldehyde. This aldehyde can then either be further oxidized by ALDH to a polar carboxylic acid metabolite (which is excreted directly and/or conjugated with glucuronic acid by UGTs), or the aldehyde is oxidized by AO to a species that could be more or less reactive than its precursors, including the parent drug. Any of these steps could determine the individual risk of drug-related toxicity.

## 3 CLASSIFICATION OF TOXICITY

Drug-related toxicity or adverse drug reactions can be categorized in general terms such as based on target organ (e.g. hepatotoxicity) or based on mechanism of action (e.g. oxidative stress). Adverse drug reactions can be categorized specifically as related to (i) the exaggerated pharmacology of the compound (on-target effects)—for example, hypotension from beta-blockers; (ii) a consequence of off-target pharmacology—for example, QT prolongation by terfenadine; (iii) immunological and hypersensitivity reactions—for example, halothane hepatitis; (iv) formation of reactive metabolites—for example, agranulocytosis with clozapine; and (v) idiosyncratic reactions—for example, hepatotoxicity with carbamazepine. The general scheme of toxicity is depicted in Fig. 4.

Adverse effects can be anticipated in the first two categories and are generally quick to onset. Such effects if reversible are reasonably well tolerated. In such cases, subjects

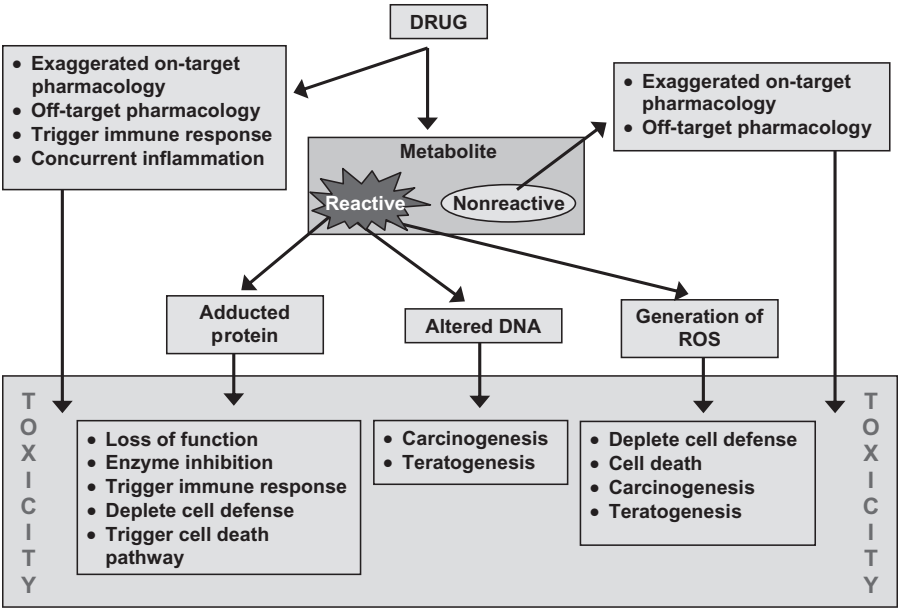


**Figure 4** General scheme of toxicity. Toxicity of a drug and its metabolite, stable or reactive, is complicated and in parallel could precipitate various responses that result in toxicity.

can be monitored closely during clinical trials (e.g. ECG recording and blood-clotting times), with the exposure and/or doses titrated to maximize safety and tolerability. In contrast, those cases that fall into the other three categories of toxicity are more complicated and harder to evaluate and often require a good understanding of the biotransformation of the drug and the enzymes involved in its clearance. It is important to note that drugs can exhibit toxicities that fall into more than one category—their fate ultimately depends on the risk–benefit analysis in the target patient populations.

## 4 MOLECULAR MECHANISMS OF TOXICITY

In this section, we provide a general overview of the molecular mechanisms of toxicity that range from loss of function of cellular macromolecules, oxidative stress, balance between survival and cell death pathways, to antibody-driven and inflammation-influenced toxicities. It should be noted that at any given time with any given compound, there can be more than one mechanism active in precipitating toxicity.

Under normal physiological state, cellular functions such as energy metabolism, survival, and cell death signals are in homeostasis. Dysregulation is usually due to the deterioration in function of critical cellular macromolecules resulting in cell death. Such events can be triggered by the drugs themselves and/or by formation of reactive metabolites that bind with essential cellular macromolecules, effectively compromising cell survival and initiating oxidative stress and/or events that culminate in cell death (apoptosis, necrosis) or an immune response.

Some drugs (e.g. acetaminophen, adriamycin, nilutamide, and bleomycin) can cause toxicity by formation of oxyradicals that deplete cellular defenses. With drugs and reactive metabolites that cause toxicity directly, the number of protein adducts formed and the related incidence of toxicity increases proportionally with time and/or increasing dose. A few drugs are known to generate reactive intermediates that are not directly toxic when they bind with macromolecules but in doing so, form haptens that activate the immune system and cause potentially serious or fatal immune reactions. Lastly, there are a few drugs whose toxicity is termed "idiosyncratic," which means they do not display an obvious dose–response in animals or humans nor can their toxicity be predicted based on known pharmacology of the drug. Unfortunately this kind of toxicity is observed usually in late-stage development or post-marketing of a drug, only after a much bigger patient population has been exposed to the drug. When idiosyncratic drug reactions (IDRs) are observed drug withdrawals are likely to follow, unless an appropriate biomarker or predisposition to the adverse reactions is identified and can be monitored safely and cost-effectively in patients.

### 4.1 Loss of Function of Cellular Macromolecules

Many drugs (e.g. carbamazepine, diclofenac, and valproic acid) form reactive metabolites; however, not all reactive metabolites are quantitatively or qualitatively the same in risk. In fact, only a few drugs cause reactions that are significant enough to manifest as clinical symptoms. The working hypothesis for reactive intermediate-driven toxicity is that the reactive entity binds (usually by covalent bond formation) to critical cellular macromolecules, compromising the physiological function and eventually resulting in cell death and tissue damage. Given the broad range of macromolecules to which the reactive entity can theoretically bind, toxicity is seldom triggered by a single cascade

of events and instead is usually a consequence of multiple parallel events (reviewed in Park, Pirmohamed, and Kitteringham, 1995).

Generally, the enzyme that mediates formation of the reactive metabolite is the target of covalent modification and subsequent inactivation. Such a phenomenon is seen as a time-dependent loss of enzyme function *in vitro* and is an indicator of potential clinical drug–drug interactions. A few examples include inactivation of CYP1A2 by dihydralazine (Masubuchi and Horie, 1999), CYP2B6 by tamoxifen (Sridar *et al.*, 2002) and CYP2C9 by tienilic acid (Melet *et al.*, 2003), CYP2C11 by diclofenac (Masubuchi *et al.*, 2001), and CYP3A4 by diltiazem (Jones *et al.*, 1999).

Other proteins, usually in the vicinity of the reactive intermediate generating system, can become a target for covalent modification. A good example is that of cisplatin, a commonly used chemotherapeutic agent, which produces a selective destruction of renal proximal tubule cells after bioactivation (cysteine conjugate followed by generation of a reactive platinum-thiol compound by cysteine *S*-conjugate beta-lyase). This reactive platinum-thiol binds to proteins in the mitochondria (e.g. alpha-ketoglutarate dehydrogenase and aconitase) that are in close proximity to cysteine *S*-conjugate beta-lyase (Zhang *et al.*, 2006). Similarly, methapyrilene, an antihistamine associated with hepatotoxicity and liver tumors in rats, is metabolized by CYPs to a reactive intermediate that covalently modifies mitochondrial proteins (Anderson *et al.*, 1992; Ratra *et al.*, 1998).

Sometimes a reactive metabolite can react with proteins in several cellular compartments. The classic example is the bioactivation of acetaminophen. Acetaminophen hepatotoxicity is precipitated by hepatic glutathione depletion by the reactive intermediate NAPQI which binds to and modifies many cellular proteins in most intracellular compartments of hepatocytes. Examples of enzymes inhibited by a toxic dose of acetaminophen include glyceraldehyde-3-phosphate dehydrogenase and Ca2+/Mg2+ ATPase, thus affecting ATP production and Ca homeostasis and initiating a cascade of events that ends in hepatocellular necrosis (Dietze *et al.*, 1997; Park *et al.*, 2005) Similarly, zomiperac, an NSAID, is metabolized to a chemically reactive acyl glucuronide that covalently modifies essential proteins like dipeptidyl peptidase IV and tubulin in liver in rats (Bailey *et al.*, 1998; Wang *et al.*, 2001).

The chemical stability of the drug-related reactive species is an important determinant in how far removed from the site of formation its protein adducts can be seen. There are many drugs whose reactive intermediates target not only the macromolecules in the cells that formed them, but also proteins in distant tissues. For example, NSAID acyl glucuronide metabolites are formed by UGTs in the liver, yet are stable enough to be secreted in the bile to form adducts with intestinal proteins (Atchison *et al.*, 2000; Aithal *et al.*, 2004). The reactive acyl glucuronide of diclofenac is implicated in intestinal injury in rats and through formation of protein adducts (Seitz and Boelsterli, 1998).

### 4.2 Oxidative Stress

Cellular stress and damage induced by ROS is a well-studied phenomenon and numerous compounds (e.g. acetaminophen, adriamycin, nilutamide, and bleomycin) induce cell death by this mechanism, either directly or via their reactive metabolites. The ROS (e.g. superoxide radical, hydrogen peroxide, and the highly reactive hydroxyl radical) initiate toxicity through oxidative damage of DNA, proteins, and lipids. Phospholipids, for instance, are highly unsaturated fatty acids and an integral part of the biological membranes; peroxidation of these lipids thus results in cell damage.

Because ROS are by-products of many enzyme mechanisms (e.g. CYPs and MPO), cells are equipped with defense mechanisms to counter the toxic effects of oxyradicals. These defenses include enzymes such as superoxide dismutase, catalase, glutathione peroxidase and stores of alpha-tocopherol, and reduced GSH and ascorbate. Toxicity usually results when these defenses are overwhelmed. In alcoholic liver disease, for example, the increased expression of CYP2E1 in hepatocytes leads to the generation of ROS and the subsequent depletion of GSH, which thus increases the vulnerability of cells (Jones *et al.*, 2002). Many anticancer drugs are known to cause redox cycling and thereby generate oxygen radicals that elicit cellular damage. For example, anthracycline-induced cardiotoxicity is attributed to the free radicals formed during redox cycling (Doroshow, Locker and Myers, 1980).

Hepatocytes are equipped with the ability to sense and respond to oxidative stress. Nuclear factor-erythroid 2-related factor 2 (Nrf2) is a redox-sensitive transcription factor that is normally kept inactive by Kelch-like (ECH)-associated protein 1 (Keap1), an adaptor of the proteasome (Jaiswal, 2004). Under oxidative stress, Nrf-2 ubiquitination is perturbed leading to saturation of Keap1 and newly synthesized Nrf2 translocates to the nucleus where it initiates the antioxidant response element-mediated gene transcription (Okawa *et al.*, 2006; Tong *et al.*, 2006). Nrf-2-responsive genes are enzymes involved in the metabolism of reactive intermediates (e.g. epoxide hydrolase) and phase II conjugation (e.g. GSTs) (Copple *et al.*, 2008). In the case of acetaminophen, for example, hepatic nuclear translocation of Nrf-2 was seen at nontoxic doses and before overt hepatotoxicity in mice (Goldring *et al.*, 2004). In fact, Nrf2-null mice are more susceptible to acetaminophen toxicity and hepatocyte-specific Keap1-null mice are highly resistant to doses of acetaminophen that are lethal to wild-type mice (Chan, Han and Kan, 2001; Okawa *et al.*, 2006). These suggest the critical roles of Nrf2 and Keap1 in protecting against oxidative stress. Similarly, in the case of anticonvulsant drugs felbamate, carbamazepine, phenobarbital, and phenytoin that are associated with varying degrees of effects on liver, pronounced effects on oxidative stress-reactive metabolite-responsive genes (aflatoxin B1 aldehyde reductase, epoxide hydrolase 1 and 2, DT-diaphorase, GST Ya subunit, glucose-6-phosphatase catalytic subunit) were seen in the liver of rats (Leone *et al.*, 2007).

Markers of oxidative stress that occur during the course of toxicity are demonstrated at the organ level. For example, valproic acid is an antiepileptic agent that is associated with cases of hepatotoxicity and hepatic failure in humans. When given to rats, valproic acid increased plasma and liver levels of 15-F(2t)-isoprostane, a marker of oxidative stress that precedes the onset of liver necrosis (Tong *et al.*, 2005). Similarly, a high level of plasma total peroxide concentration, indicative of oxidative stress, is seen in HIV-infected patients on protease inhibitor therapy. Oxidative stress also is implicated in the premature development of atherosclerosis and coronary heart disease in patients on antiretroviral therapy (Masiá *et al.*, 2007).

Oxidative damage to proteins and DNA is also associated with carcinogenesis and teratogenesis. For example, use of diethylstilbestrol during pregnancy was associated with adenocarcinoma of the vagina in daughters at puberty; ROS generation in embryonic tissue is one of the many hypotheses of carcinogenesis associated with diethylstilbestrol (Kodama *et al.*, 1993). Teratogenicity of thalidomide is also linked to ROS generation in embryonic tissue. Substantial DNA oxidation is seen in rabbits following thalidomide; however, pretreatment with a free radical-trapping agent can prevent the teratogenicity, supporting the hypothesis that free radicals are involved (Wells *et al.*, 1997; Parman, Wiley and Wells, 1999). Inhibition of gene expression related to DNA methylation during oxidative stress has been linked to altered function of protein kinases and

transcription factors. Impaired signal transduction is one of the many factors implicated in teratogenesis and carcinogenesis (Fratelli *et al.*, 2005; López-Neblina and Toledo-Pereyra, 2006). In addition to initiating cellular damage, ROS are also implicated in the activation of alternative pathways that can contribute to adverse effects. Tamoxifen, for example, is an agent used in the treatment of breast cancer with an increased risk of thrombosis in humans. This has been attributed to reactive metabolites of tamoxifen, with *in vitro* evidence that these metabolites can increase superoxide release by platelets, thereby increasing platelet aggregation (Vitseva *et al.*, 2005).

### 4.3 Altered Balance of Cell Survival and Death Pathways

Pertubation of cellular homeostasis is a general sequel to changes initiated by functional compromise of critical cellular proteins or cellular defense systems. This ultimately results in cell death, modes of which are commonly described as apoptosis or necrosis. Numerous examples exist of toxic chemicals that can alter cellular homeostasis by affecting the balance between the cell survival and cell death pathways. A few drugs will be discussed as examples.

Apoptosis—or programmed cell death—involves activation of "death pathways," one component of which is p53, a critical death effector protein. Disulfiram, used to treat alcoholism, can induce apoptosis *in vitro*. The reactive metabolites of disulfiram that carbamoylate and inhibit aldehyde dehydrogenase also induce p53-dependent apoptosis in HepG2 cells (Liu *et al.*, 1998). Similarly, cisplatin-induced apoptosis in a dorsal root ganglion hybrid cell line is by accumulation of p53, with protection supplied by pre-incubation with antioxidants (Park *et al.*, 2000).

A portion of acetaminophen toxicity occurs via the initiation of cell death pathways, as demonstrated by some protection against the early apoptotic processes (and the later necrotic processes) conferred by neutralization of Fas ligand and tumor necrosis factor (TNF) (Blazka *et al.*, 1996; Zhang *et al.*, 2000). Neuronal loss in HIV dementia after treatment with the nucleoside reverse transcriptase inhibitor, 2′, 3′-dideoxycytidine, was linked to induction of apoptosis in neuronal cells. In gerbils, the drug reduces expression of the anti-apoptotic protein Bcl-2 and increases expression of pro-apoptotic protein caspase-3 in brain mitochondria (Opii *et al.*, 2007).

Troglitazone, a peroxisome proliferator-activated receptor gamma agonist in the treatment of type 2 diabetes, was removed from the market due to concerns over hepatotoxicity. In rat hepatocytes and HepG2 cells, troglitazone induced apoptosis (Toyoda *et al.*, 2001; Yamamoto *et al.*, 2001). Pretreatment of HepG2 cells with a c-Jun *N*-terminal protein kinase inhibitor blocked the troglitazone-induced change in pro-apoptotic protein levels with a reduction in apoptotic cell death (which suggests a role exists for signaling pathways in troglitazone-induced apoptosis) (Bae and Song, 2003).

Toxic epidermal necrolysis is a type of reactive metabolite-triggered immune-mediated cutaneous adverse reaction characterized by epidermal detachment. The incidence of such a reaction is rare but has been described for many drugs (e.g. sulfonamides, lamotrigine, carbamazepine, and nevirapine). Mechanistic studies have revealed apoptosis of keratinocytes as the critical event leading to necrosis, a classic histological feature of this condition. Unlike keratinocytes in normal, healthy skin, cells in this condition express high levels of Fas ligand at the cell surface (Viard-Leveugle *et al.*, 2003; French, 2006). Fas-mediated keratinocyte apoptosis can be inhibited *in vitro* by monoclonal antibodies to Fas and *in vivo* by immunoglobulins containing anti-Fas antibodies suggesting a prominent role for apoptosis in this condition (Tan *et al.*, 2005; French, Trent, and Kerdel, 2006).

### 4.4 Immune-Mediated

In contrast to the drugs that cause direct toxicity, including many of those described earlier, there are compounds that are able to form protein conjugates that can initiate toxicity via activation of the immune system. Well-known examples are cases of hepatitis following administration of halothane, tienilic acid, dihydralazine, or diclofenac. The incidences of such adverse effects are rare and often unpredictable, hence their designation as idiosyncratic adverse drug reactions.

The reasons for this type of reaction are far from understood and several hypotheses are proposed. The first is the "hapten hypothesis," which suggests that antibodies raised against drug/metabolite–protein (typically to the enzyme that produced the metabolite or proteins in the vicinity)—adducts seen by the immune system as "neoantigens"—subsequently recognize the modified or even the native proteins. These autoantibodies lead to cytolytic toxicity with further or repeated exposure to the same or similar drug that generated the neoantigens (Padovan *et al.*, 1997). A second hypothesis is the "danger hypothesis," which argues that the reactive metabolite causes damage to the tissue and this leads to activation of antigen-presenting cells in an immune response (Matzinger, 1994). A third hypothesis is "pharmacological interaction" which involves a complex formed between the reactive metabolite and the major histocompatibility complex class 1 (MHC-1) leading to a cell-mediated immune response (Schnyder *et al.*, 1997). Given an absence of overwhelming evidence in favor of just one hypothesis, it seems reasonable to assume that any combination of one or all and possibly other mechanisms, as yet unknown, could well be involved in the pathogenesis of immune-mediated drug toxicity.

Many drugs that initiate an immune response are demonstrated to follow the hapten or neoantigen hypothesis. Target proteins are generally localized within the cell. For example, halothane, a volatile anesthetic, is metabolized primarily by CYP2E1, generating a reactive trifluoroacetyl halide metabolite that acetylates various proteins including CYP2E1 (Spracklin *et al.*, 1997). Consequently, many patients suffering with halothane hepatitis have autoantibodies to CYP2E1 (Bourdi *et al.*, 1996). Halothane produces adducts that are almost exclusively found in the liver microsomal subfraction. These modified proteins are accessible to the immune system by exposure on the plasma membrane. Similarly dihydralazine, an antihypertensive drug associated with hepatitis, can trigger an immune response producing antibodies against CYP1A2 (Beaune *et al.*, 1993). Tienilic acid, a diuretic associated with hepatitis is metabolized by CYP2C9 to a reactive intermediate that in turn covalently modifies the same enzyme and eventually leads to the generation of antibodies to CYP2C9 (Lecoeur *et al.*, 1994). These mechanisms are not just limited to reactive metabolites formed via CYP-mediated reactions; theoretically, any reactive metabolite that can form a protein adduct is potentially able to initiate an immune response. For instance, diclofenac-induced hepatitis is also in part immune-mediated. In the rat, both CYPs and UGTs are involved in forming reactive intermediates of diclofenac, with a majority of protein adducts detected on the plasma membrane and localized within the bile canalicular membrane (Hargus *et al.*, 1994). A similar pattern of covalent binding is seen with other NSAIDs such as diflunisal, zomepirac, sulindac, and ibuprofen (Bailey and Dickinson, 1996; Wade, Kenna and Caldwell, 1997). Metabolic activation of abacavir to an aldehyde intermediate is thought to be involved in covalent binding to proteins leading to hypersensitive reactions in a small percent of patients (Walsh, Reese and Thurmond, 2002).

Haptenation in itself is not always sufficient to provoke an immune response and requires co-stimulatory signals to initiate activation of antigen-presenting cells and a subsequent immune response (which is evidence in support of the danger hypothesis). For example, the cutaneous adverse reactions seen with dapsone and sulfamethoxazole are attributed to an up-regulation of danger signals by their reactive arylhydroxylamine metabolites. In human epidermal keratinocytes, arylhydroxylamine metabolites up-regulate stress signals such as pro-inflammatory cytokines and heat shock protein 70—this can subsequently activate dendritic cells and initiate an immune reaction with the skin (Khan *et al.*, 2007).

The hapten hypothesis, and its basis in reactive metabolite formation as a trigger for immune response, was challenged when an immune response was observed for drugs such as lidocaine, lamotrigine, carbamazepine, and sulfamethoxazole but not their metabolites. The pharmacological interaction hypothesis explains this phenomenon by stating that, in this case, the drugs bind directly and reversibly to an immune receptor (e.g. the major histocompatibility complex or the T cell receptor), which then stimulates the cells in a manner similar to a pharmacological activation of the receptors (Pichler *et al.*, 2006; Roujeau, 2006).

### 4.5 Influence of Inflammatory Mediators and Other Cells on Toxicity

The pathophysiological role of inflammatory mediators in the manifestation of toxicity is well known and generally considered to be an integral part of the pathology. In the context of idiosyncratic reactions, however, substantial evidence exists that shows that animals with a low-grade inflammation develop a predisposition to drug toxicity from low dosages that otherwise are not toxic. Best known examples are acetaminophen-, ranitidine-, and chlorpromazine-associated hepatotoxicity. Experimentally, toxicities of these compounds have been shown to be precipitated by an ongoing low-grade inflammation initiated by treatment with lipopolysaccharide (LPS) (Ganey and Roth, 2001). It should be noted that these are examples of idiosyncratic toxicity and not reactive metabolite-driven adverse reactions.

Non-parenchymal cells of the liver, such as sinusoidal endothelial cells, stellate cells, and Kupffer cells, along with infiltrating neutrophils, can influence the toxicities of some drugs. Kupffer cells are the resident macrophages in the liver and during hepatic inflammation are activated to release pro-inflammatory cytokines (e.g. TNFalpha), in addition to a barrage of destructive mediators (e.g. ROS, reactive nitrogen species, and proteases) that could influence the toxicity of other compounds exposed at the same time. In the case of acetaminophen toxicity, for example, evidence that Kupffer cells play a direct role in acetaminophen toxicity is the protection seen in mice depleted of their Kupffer cells by dichloromethylene diphosphonate (Goldin *et al.*, 1996). Inflammatory mediators can also activate other cells, such as sinusoidal cells, with the subsequent expression of adhesion molecules and stellate cells causing a contraction of the sinusoids, which leads to the recruitment of neutrophils and facilitates their transmigration (Ganey *et al.*, 1994; Kharbanda *et al.*, 2004).

In addition to their hypothesized role in idiosyncratic toxicity, inflammatory mediators play a role in the precipitation of toxicity, as seen with compounds such as halothane and acetaminophen. In Balb/c mice, halothane increased the hepatic messenger levels of the inflammatory mediators TNF-alpha, IL-1beta, IL-6, and IL-8; this is in addition to recruiting neutrophils to the liver. These events suggest a role for pro-inflammatory mediators in the halothane-initiated hepatotoxicity observed in this

strain of mice (You *et al.*, 2006). Similarly, an increase in macrophage migration inhibitory factor, a pro-inflammatory signal, was seen in guinea pigs with hepatotoxicity from halothane or acetaminophen. Migration inhibitory factor null mice were significantly less susceptible to acetaminophen toxicity, suggesting a role for this inflammatory mediator in the predisposition toward toxicity (Bourdi *et al.*, 2002). In addition to their role in inflammation, cytokines such as interleukins, TNF-alpha, and interferons can down-regulate CYP enzymes and thus affect the extent of bioactivation and/or clearance of drugs (Bertini *et al.*, 1988; Renton and Knickle, 1990; Abdel-Razzak *et al.*, 1993).

## 5 ORGAN SYSTEMS TOXICOLOGY

In this section, we describe the toxicity and mechanisms of several well-known drugs whose adverse reactions are mediated primarily by the formation of reactive intermediates. These drugs demonstrate distinctive specificity toward target organs and the examples are grouped accordingly.

### 5.1 Hepatotoxicity

Liver is the primary organ of metabolism of most compounds and the primary target for drug toxicity. The majority of compounds that form reactive metabolites are hepatotoxic. Apart from the innate reactivity of the reactive intermediate, its local concentration and the proximity of proteins, such as the enzyme that helped form it, are other factors that predispose toward injury of the hepatocyte. Since biliary clearance of endogenous and exogenous substrates is a key function of the liver, bile may therefore contain a relatively high concentration of reactive metabolites that can damage the cholangiocyte and, in severe cases, be seen as vanishing bile duct syndrome. The hepatotoxicities of selected drugs (e.g. acetaminophen, diclofenac, carbamazepine, valproic acid, and troglitazone), pathological features of their toxicities (e.g. hepatic necrosis, microvascular steatosis, and cholestasis), and the importance of risk–benefit analysis in the fate of drugs (e.g. bromfenac and troglitazone) are discussed in the following section. A comprehensive but certainly not exhaustive list of drugs associated with organ-specific toxicity is provided in Table 3. This serves to illustrate some of the issues and complexities that face scientists working in drug discovery and development.

Acetaminophen is a classic example of a drug that causes hepatotoxicity, which can present as acute liver failure and accounts for as much as 39% of the liver failure cases reported in tertiary care centers in the United States (Ostapowicz *et al.*, 2002). Acetaminophen hepatotoxicity is dose-dependent and is precipitated by loss of hepatic defenses. The pathology is characterized by hepatic necrosis, predominantly in the centrilobular region and is mediated by its reactive intermediate NAPQI. Within the therapeutic dose range, acetaminophen is detoxified by glucuronidation and sulfation; however, a portion of the drug is also metabolized primarily by CYP3A4 to the reactive NAPQI, which is immediately quenched by mitochondrial glutathione (Davis *et al.*, 1974). At toxic doses of the drug, CYP2E1, which is also ethanol inducible, plays a major role in the generation of NAPQI (CYP1A2 and 2A6 are also involved in the generation of NAPQI *in vitro*) (Patten *et al.*, 1993; Chen *et al.*, 1998). In the event of an overdose of acetaminophen, and/or situations in which glutathione is depleted, NAPQI is not quenched and instead binds covalently to critical cellular proteins.

**TABLE 3 Drugs, putative reactive intermediates, implicated enzymes, and associated toxicity.**

| Drug | Putative reactive intermediate | Enzyme | Adverse reaction | References |
|---|---|---|---|---|
| Abacavir | Aldehyde | Alcohol dehydrogenase | Hypersensitivity | Walsh, Reese and Thurmond, 2002 |
| Acetaminophen | Quinoneimine | CYP3A4, 2E1, 1A2, 2A6 | Hepatotoxicity, nephrotoxicity | Chen *et al.*, 1998; Patten *et al.*, 1993; Davis *et al.*, 1974; Hart *et al.*, 1994 |
| Adriamycin | Quinones, ROS | Redox cycling | Cardiotoxicity, neurotoxicity | Berthiaume and Wallace, 2007; Joshi *et al.*, 2005 |
| Amodiaquine | Quinoneimine | Peroxidases | Hepatotoxicity, agranulocytosis | Tingle *et al.*, 1995; Maggs *et al.*, 1988 |
| Alclofenac | Epoxide | CYP | Hepatotoxicity, skin rash | Hort, 1975; Slack *et al.*, 1981 |
| Aminoglutethimide | Hydroxylamine | MPO | Agranulocytosis | Coleman, Khalaf and Nicholls, 2003; Siraki *et al.*, 2007 |
| Benoxaprofen | Acyl glucuronide? | UGT | Hepatotoxicity, photosensitivity | Halsey and Cardoe, 1982 |
| Bromfenac | Acyl glucuronide? | UGT | Hepatotoxicity | Fontana *et al.*, 1999 |
| Carbamazepine | Arene oxide, quinoneimine | CYP2D, 1A2 | Hepatotoxicity | Masubuchi *et al.*, 2001 |
| Chloramphenicol | Nitroso | | Aplastic anemia | Yunis *et al.*, 1980 |
| Cisplatin | ROS, platinum-thiol | Redox cycling; GGT | Neurotoxicity, nephrotoxicity, ototoxicity | Townsend *et al.*, 2003; Smyth *et al.*, 1997 |
| Clozapine | Nitrenium | MPO, CYP | Agranulocytosi, hepatotoxicity, myocarditis | Uetrecht, Zahid and Whitfield, 1994; Pirmohamed *et al.*, 1995; Killian *et al.*, 1999 |
| Cyclophosphamide | Acrolein | CYP2B1, 2C6, 2C11 | Bladder toxicity, ovarian toxicity | Stillwell and Benson, 1988; Chang and Waxman, 1993; Clarke and Waxman, 1989; Plowchalk and Mattison, 1991 |
| Diethylstilbestrol | Quinones | CYP | Carcinogenesis | Cavalieri, Rogan and Chakravarti, 2002; Chae *et al.*, 1998 |
| Dapsone | Nitrenium, hydroxylamine | CYP3A4, MPO | Methemoglobinemia, agranulocytosis, cutaneous reactions | Grossman and Jollow, 1988; Uetrecht *et al.*, 1988; Vyas *et al.*, 2005 |
| Daunorubicin | Quinones, ROS | Redox cycling; xanthine oxidase? | Cardiotoxicity | Gustafson, Swanson and Pritsos, 1991 |
| Dihydralazine | Diazene or diazonium | CYP1A2 | Hepatitis | Bourdi *et al.*, 1992; Masubuchi and Horie, 1999 |

**TABLE 3** *Continued*

| Drug | Putative reactive intermediate | Enzyme | Adverse reaction | References |
|---|---|---|---|---|
| Diclofenac | Quinoneimine, arene oxide, acyl glucuronide | CYP3A4, 2C9, 2C11, MPO, UGT | Hepatotoxicity | Tang *et al.*, 1999; Masubuchi *et al.*, 2001; Poon *et al.*, 2001 |
| Erythromycin | Nitrosoalkane | CYP3A4 | DDI | Periti *et al.*, 1992 |
| Fluperlapine | Quinoneimine | MPO | Agranulocytosis | Lai *et al.*, 2000 |
| Flutamide | Electrophilic intermediate? | CYP3A4, 1A2 | Hepatotoxicity | Berson *et al.*, 1993 |
| Felbamate | 2-Phenylpropenal | CYP | Aplastic anemia, hepatotoxicity | O'Neil *et al.*, 1996; Kaufman *et al.*, 1997; Popovic *et al.*, 2004 |
| Haloperidol | Pyridinium | CYP3A4 | Parkinsonism, tardive dyskinesia | Usuki, Van der Schyf and Castagnoli, 1998 |
| Halothane | Trifluoroacetyl chloride | CYP2E1 | Hepatotoxicity | Spracklin *et al.*, 1997 |
| Hydralazine | Diazene or diazonium | CYP?, MPO | Autoimmunological syndrome | Hofstra. Matassa and Uetrecht, 1991 |
| Ibufenac | Acyl glucuronide | UGT | Hepatotoxicity | Castillo and Smith, 1995 |
| Irinotecan | SN-38 | Esterases | Intestinal toxicity | Takasuna *et al.*, 1998; Wadkins *et al.*, 2004 |
| Indomethacin | Quinoneimine | Amidase, MPO, CYP | Aplastic anemia, agranulocytosis | Ju and Uetrecht, 1998 |
| Isoniazid | *N*-Acetylhydrazine | Amidase, CYP1A1, 1A2, 2E1 | Hepatotoxicity | Sarich *et al.*, 1999 |
| Lamotrigine | Arene oxide | CYP | Skin rashes; toxic epidermal necrolysis | Maggs *et al.*, 2000 |
| Metronidazole | 5-Aminoimidazole, hydroxylamine | CYP | Genotoxicity | Perez-Reyes, Kalyanaraman and Mason, 1980; Kedderis, Argenbright and Miwa, 1989 |
| Methimazole | Sulfenic acid, sulfinic acid | CYP, FMO | Agranulocytosis, hepatotoxicity | Decker and Doerge, 1992; Woeber, 2002; Oh *et al.*, 2007 |
| Mianserin | Iminium | CYP2D | Skin rashes, hepatotoxicity, blood dyscrasias | Masubuchi, Konishi and Horie, 1999; Iverson, Zahid and Uetrecht, 2002 |
| Nilutamide | Nitroanion, ROS | CYP450 reductase, redox cycling | Pulmonary toxicity, hepatotoxicity | Berger *et al.*, 1992; Fau *et al.*, 1992 |
| Nitrofurantoin | Nitroanion, ROS | CYP450 reductase, redox cycling | Pulmonary toxicity, hepatotoxicity | Wang *et al.*, 2008 |

| | | | | |
|---|---|---|---|---|
| Procainamide | Nitrenium | CYP2D6 | Lupus erythematosus; agranulocytosis | Lessard *et al.*, 1999 |
| Paroxetine | Catechol | CYP2D6 | DDI | Bertelsen *et al.*, 2003 |
| Phenytoin | Unstable nitrogen-centered radical | COX, peroxidases | Teratogenesis | Wells *et al.*, 1997; Parman, Wiley and Wells, 1999 |
| Ritonavir | Ring scission | CYP3A4 | Enzyme inactivation | Kempf *et al.*, 1997 |
| Raloxifene | Arene epoxide; quinone | CYP3A4 | Enzyme inactivation | Yukinaga *et al.*, 2007 |
| Sudoxicam | Ring scission | | Hepatotoxicity | Hobbs and Twomey, 1977 |
| Suprofen | Epoxide, acyl glucuronide | CYP2C9 | Nephrotoxicity | O'Donnell *et al.*, 2003; Smith and Liu, 1993 |
| Sulfamethoxazole | Hydroxylamine | CYP2C9 | Hypersensitivity | Miller and Trepanier, 2002 |
| Terbinafine | Aldehyde metabolite? | CYP | Hepatotoxicity | Mallat *et al.*, 1997; Iverson and Uetrecht, 2001 |
| Tacrine | Quinone methide | CYP1A2 | Hepatotoxicity | Madden *et al.*, 1993; Spaldin *et al.*, 1995 |
| Thalidomide | Unstable nitrogen-centered radical | COX | Teratogenesis | Wells *et al.*, 1997; Parman, Wiley and Wells, 1999 |
| Tolbutamide | *n*-Butylisocyanate | Degradation? | Teratogenesis | Guan *et al.*, 1999 |
| Tolcapone | Quinoneimine | CYP2E1, CYP1A2 | Hepatotoxicity | Smith *et al.*, 2003 |
| Troglitazone | Quinone, quinone methide | CYP3A4 | Hepatotoxicity | Kassahun *et al.*, 2001; Yamamoto *et al.*, 2002 |
| Tamoxifen | Quinone methide | CYP | Carcinogenesis | Fan and Bolton, 2001 |
| Trimethoprim | Quinone methide | CYP | Hepatotoxicity | Lai, Zahid and Uetrecht, 1999 |
| Ticlopidine | | CYP2C19, 2B6 | Agranulocytosis, aplastic anemia, thrombocytopenia | Liu and Uetrecht, 2000 |
| Tienilic acid | Tienilic acid-*S*-oxide | CYP2C9 | Hepatitis | Melet *et al.*, 2003 |
| Thiabendazole | Thioformamide | CYP | Nephrotoxicity | Mizutani, Yoshida and Kawazoe, 1993 |
| Valproic acid | 4-ene | CYP2C9 | Hepatotoxicity | Stephens and Levy, 1992; Kiang *et al.*, 2006 |
| Vesnarinone | Imidoiminium ion | CYP3A4, 2E1 | Agranulocytosis | Uetrecht, Zahid and Whitfield, 1994; Frye *et al.*, 1999 |
| Zomepirac | Acyl glucuronide? | UGT? | Allergic reactions | Bailey and Dickinson, 1999; Olsen *et al.*, 2005 |

Possible mechanisms of hepatocellular necrosis include modification of sulfhydryl groups, oxidative stress, lipid peroxidation, and arylation of essential proteins. In rodents, more than 30 hepatic enzymes are shown to be covalently modified by NAPQI with many of them functionally compromised as a result (Park *et al.*, 2005).

NSAIDs as a class are associated with some degree of hepatotoxicity; however, due to the structural diversity of NSAIDs, the toxicity presented is not uniform. For example, a higher incidence of liver toxicity is reported for nimesulide compared with diclofenac or ibuprofen (Traversa *et al.*, 2003). Diclofenac is associated with hepatotoxicity with a low incidence rate of 6 to 18 cases/100,000 person-years (Walker, 1997). The pattern of toxicity is predominantly hepatocellular and to a lesser extent cholestatic (Banks *et al.*, 1995). Histologically, acute lobular hepatitis and in severe cases bridging necrosis and fibrous expansion of portal tracts are observed (Chitturi and George, 2002). Both immunological and nonimmunological mechanisms are proposed for diclofenac hepatotoxicity. UGT-mediated glucuronic acid conjugation of diclofenac results in an acyl glucuronide that forms protein adducts either directly or after acyl migration, forming a reactive aldehyde. The mechanisms of acyl glucuronidation and reactivity are shown in Fig. 1. The reactive acyl glucuronide of diclofenac is shown to form covalent adducts with hepatic and circulating proteins. A similar pattern of covalent binding was reported for zomepirac, diflunisal, sulindac, and ibuprofen (Bailey and Dickinson, 1996; Wade, Kenna and Caldwell, 1997). Adverse reactions with diclofenac are not limited to formation of a reactive acyl glucuronide since diclofenac also undergoes CYP3A4 and CYP2C9 oxidation to metabolites such as 5-hydroxydiclofenac that are processed further by oxidation to form a reactive iminoquinone. CYPs and MPO are implicated in the formation of microsomal protein adducts and postulated to play a role in hepatotoxicity (Hargus *et al.*, 1994; Miyamoto, Zahid and Uetrecht, 1997; Tang *et al.*, 1999; Poon *et al.*, 2001). *In vitro*, an arene oxide intermediate generated during the process of diclofenac hydroxylation is shown to inactivate CYP2C11 (Masubuchi, Ose, and Horie, 2001). Furthermore, covalent modification of proteins on the cell surface is possibly involved in an immune-mediated hepatitis associated with diclofenac therapy.

Carbamazepine is an anticonvulsant linked to a variety of understood and idiosyncratic toxicities, including skin rash, hepatotoxicity, and hematotoxicity, with adverse reactions reported in 33–50% of patients (reviewed in Kalgutkar *et al.*, 2005). Carbamazepine is metabolized to 2-hydroxycarbamazepine and subsequently to 2-hydroxyiminostilbene and finally to a putative reactive iminoquinone, primarily via CYP3A4 metabolism (Pearce, Uetrecht and Leeder, 2005). Time-dependent inactivation of CYP2D in rat liver microsomes and CYP1A2 in human liver microsomes was observed which, although not conclusive, does implicate these enzymes as directly involved in the formation of reactive metabolites of carbamazepine (Masubuchi *et al.*, 2001).

Valproic acid is another anticonvulsant associated with a large incidence of hepatotoxicity in patients. An asymptomatic dose-dependent increase in serum markers of hepatotoxicity was seen in about 44% of patients on valproic acid therapy and non-dose-dependent liver failure in rare instances (Parra, Iriarte and Pierre-Louis, 1996). Reactive intermediates of valproic acid, such as the 4-ene and 2,4-diene, inhibit fatty acid metabolism by alkylation of enzymes such as 3-ketoacyl-coenzyme A thiolase. Hepatotoxicity is therefore linked to inhibition of mitochondrial beta-oxidation by its metabolites, as demonstrated *in vitro* and *in vivo*. Consequently, liver injury is characterized by microvesicular steatosis and hepatocellular necrosis (Kassahun, Farrell and Abbott, 1991; Stephens and Levy, 1992; Kossak *et al.*, 1993).

Isoniazid (antituberculosis agent) causes hepatotoxicity, usually seen as massive necrosis or chronic hepatitis and is related to dose and duration of exposure. Mild elevations of plasma liver enzyme activities have been reported in 20% of the patients (Garibaldi *et al.*, 1972). The mechanism of isoniazid hepatotoxicity is attributed to its metabolism to an intermediate acetylhydrazine, whose bioactivation in turn leads to the formation of reactive acetylating species that can covalently modify hepatic proteins (Timbrell *et al.*, 1980).

Troglitazone (Rezulin), a peroxisome proliferator-activated receptor gamma agonist, effective in the treatment of type 2 diabetes, was associated with cases of severe hepatotoxicity and liver failure that led to its withdrawal from the US market. Troglitazone-associated liver injury, in some instances continued despite discontinuation of troglitazone treatment (Menon, Angulo and Lindor, 2001). Histologically massive necrosis with post-collapse scarring and bile duct proliferation were seen (Gitlin *et al.*, 1998; Vella, de Groen and Dinneen, 1998). When introduced, it was considered a novel drug with a unique mechanism of action that benefited certain patients. The withdrawal was therefore not immediate but was subsequent to the introduction to the market of rosiglitazone and pioglitazone, considered safer drugs in the same class (Lee, 2003). Troglitazone is metabolized by CYP3A4 to several metabolites, one of which is a reactive *O*-quinone methide derivative, which suggests a role for bioactivation (Kassahun *et al.*, 2001; Yamamoto *et al.*, 2002). The actual mechanism of hepatotoxicity of troglitazone is, however, unclear, with several factors apparently involved—including bioactivation, mitochondrial dysfunction by troglitazone itself, and inhibition of bile salt export pump by troglitazone sulfate (Masubuchi, 2006).

Lastly, bromfenac, an NSAID approved as a short-term analgesic for orthopedic pain for use for periods of 10 days or less, is another example that illustrates the importance of risk–benefit analysis. Bromfenac was withdrawn from the market within a year of release following the many reports of severe liver injury (all in patients taking the drug for more than 30 days) (Fontana *et al.*, 1999; Lee, 2003). Although the NSAID class has toxicity issues (to a varying degree) possibly due to reactive intermediates such as acyl glucuronides, the precise mechanism by which bromfenac causes hepatotoxicity is not clear.

### 5.2 Hematotoxicity

Blood cells and hematopoietic tissue is, besides the liver, the organ system most exposed to chemicals. Neutropenia and agranulocytosis (a common manifestation of toxicity to neutrophils/its precursors), methemoglobinemia and anemia (effects reflecting toxicity to erythrocytes/its precursors), and thrombocytopenia (toxicity to platelets/its precursors) are adverse reactions often reported for drugs that are known to form reactive intermediates. Reactive intermediates that are formed in the liver are to some extent responsible for toxicity to the blood cells; however, blood cells such as neutrophils can also metabolize drugs that escape first-pass metabolism and it is this mechanism that contributes most to the majority of examples of agranulocytosis described in the literature (Uetrecht, 1992).

As described earlier, MPO is the major oxidizing enzyme in neutrophils, with hypochlorous acid (HOCl) the major oxidant produced by activated neutrophils. The MPO–HOCl system is involved in generation of reactive intermediates for the vast number of compounds that cause agranulocytosis (Uetrecht, 1992). For example, clozapine, an

atypical antipsychotic drug is associated with neutropenia and agranulocytosis in human. The toxicity of clozapine is attributed to a reactive nitrenium ion metabolite formed by MPO in neutrophils (Williams *et al.*, 1997). Similarly, fluperlapine, another antipsychotic structurally related to clozapine, is also associated with agranulocytosis. The major metabolite of fluperlapine, 7-hydroxyfluperlapine, is converted to a reactive iminoquinone by neutrophil MPO and is the likely proximate cause of toxicity (Lai *et al.*, 2000). Carbamazepine-induced agranulocytosis is also linked to reactive metabolites, in this case the metabolite 9-acridine carboxaldehyde, which is generated by MPO in neutrophils and monocytes (Furst and Uetrecht, 1995). Bioactivation of amodiaquine, tebuquine, cycloquine, and pyronaridine to respective chemically reactive quinoneimine metabolites occurs in the presence of horseradish peroxidase and hydrogen peroxide, suggesting a mechanism for the agranulocytosis seen with some of these drugs (Naisbitt *et al.*, 1998). Diclofenac is also associated with bone marrow toxicity, with a reactive iminoquinone formed by HOCl or by activated neutrophil oxidation of an intermediate 5-hydroxydiclofenac (Miyamoto, Zahid and Uetrecht, 1997). Similarly, indomethacin also demonstrates agranulocytosis associated with a reactive iminoquinone formed by activated neutrophils, in this case via oxidation of its major metabolite desmethyldeschlorobenzoylindomethacin (Ju and Uetrecht, 1998). Trimethoprim, an antibacterial drug used synergistically with sulfonamides, is associated with agranulocytosis—a reactive pyrimidine iminoquinone formed in activated neutrophils may be responsible (Lai, Zahid and Uetrecht, 1999). Similarly, agranulocytosis induced by vesnarinone, an agent used in the treatment of congestive heart failure, is linked to a reactive quinoneimine formed by neutrophils or neutrophil precursors in bone marrow (Uetrecht, Zahid and Whitfield, 1994). Activated neutrophils are also shown to metabolize phenytoin to reactive intermediates that bind irreversibly to macromolecules in neutrophils (Mays *et al.*, 1995).

There are some examples that link reactive metabolite formation to toxicity to erythrocytes or its precursors leading to methemoglobinemia and anemia. Ticlopidine, an ADP-receptor antagonist used in the prevention of atherothrombosis, is associated with incidences of agranulocytosis and aplastic anemia, with a putative thiophene-*S*-chloride formed by HOCl or activated neutrophils implicated (Liu and Uetrecht, 2000). Dapsone, a drug used to treat leprosy, is associated with methemoglobinemia and agranulocytosis, with CYP3A4 and/or MPO forming a hydroxylamine metabolite that is toxic to erythrocytes and leukocytes (Zuidema, Hilbers-Modderman and Merkus, 1986; Grossman and Jollow, 1988; Uetrecht *et al.*, 1988). Nomifensine is an antidepressant drug that was withdrawn from the market due to allergic reactions including hemolytic anemia. The arylhydroxylamine metabolite and nitroso metabolites of nomifensine are thought to be responsible for its toxicity (Salama and Mueller-Eckhardt, 1985; Lindberg and Syvälahti, 1986). The arylhydroxylamine metabolite of phenacetin (*N*-hydroxyphenetidine) through direct actions in the erythrocytes is responsible for phenacetin-induced hemolytic anemia and methemoglobinemia (Jensen and Jollow, 1991). Similarly, the arylhydroxylamine metabolite, 6-methoxy-8-hydroxylaminoquinoline of the antimalarial drug primaquine, is toxic to erythrocytes (Bolchoz *et al.*, 2001). Aplastic anemia associated with the use of the antibiotic chloramphenicol is believed to be due to its nitroso metabolite that is formed in the bone marrow (Yunis *et al.*, 1980). Lastly, aplastic anemia reported with felbamate toxicity is thought to be immune-mediated and associated with its reactive metabolite, 2-phenylpropenal, a potent immunogen (Popovic *et al.*, 2004).

Drug-induced thrombocytopenia can result from either a decrease in platelet production (e.g. from bone marrow toxicity) or destruction of platelets (e.g. from

an immune response). Antibodies that are drug-dependent, hapten-dependent, and platelet-reactive are involved in drug-induced thrombocytopenia (van den Bemt, Meyboom and Egberts, 2004). Quinine and sulfonamides are examples of drugs that induce immune thrombocytopenia. Aminoglutethimide is an aromatase inhibitor used in the treatment of breast cancer. Thrombocytopenia is one of the adverse reactions seen with this drug and possibly due to its hydroxylamine metabolite (Stuart-Harris and Smith, 1984; Dalrymple and Nicholls, 1988).

### 5.3 Gastrointestinal Toxicity

Gastrointestinal toxicity, primarily of the intestines and as a consequence of reactive intermediates, is described for many compounds. Intestinal toxicity occurs more frequently from reactive metabolites formed in the liver and then secreted in bile than from direct bioactivation within enterocytes (Treinen-Moslen and Kanz, 2006). This is linked directly to the significant difference that exists between enterocytes and hepatocytes in their metabolic capacity at the organ level. It should be recognized that orally ingested drugs can have very high local concentrations in the intestine and toxicity can occur directly as a result of toxicodynamics related to the drug itself.

Carboxylic acid-containing NSAIDs may be extensively metabolized by UGTs in the liver to produce reactive acyl glucuronides that are secreted in the bile and are implicated in NSAID enteropathy. One example is diclofenac, which was shown to cause ulceration significantly in the small intestine of rats given bile from diclofenac-treated rats compared with rats that received diclofenac alone. Further, hepatocanalicular conjugate export pump deficient (TR-) rats were refractory to intestinal injury by diclofenac and therefore intestinal injury of diclofenac therapy was linked to reactive acyl glucuronide metabolites of diclofenac secreted in the bile (Seitz and Boelsterli, 1998). Similarly, the reactive acyl glucuronide metabolite of the immunosuppressant drug mycophenolic acid (MPA) is thought to contribute to intestinal injury. Although the reactive acyl glucuronide is formed to a lesser extent than the more stable ether glucuronide metabolites, significant levels are still reported in MPA-treated patients (Shipkova *et al.*, 2002). Furthermore, the reactive metabolite of MPA is also formed in human intestinal microsomes, suggesting a possible role for *in situ* formation of this metabolite in intestinal injury (Shipkova *et al.*, 2001).

Irinotecan is a semisynthetic derivative of camptothecin used in cancer chemotherapy and another example of a drug associated with intestinal injury. As described earlier, irinotecan and its more potent metabolite, SN-38, cause irreversible DNA damage and cell death and concomitant anticancer activity. Probably the same mechanisms are also involved in the intestinal toxicity. Irinotecan administered intravenously is bioactivated by carboxylesterase in the liver to SN-38, which in turn undergoes glucuronidation to an acyl glucuronide. Glucuronidated SN-38, unconjugated SN-38, and unchanged irinotecan are secreted in the bile into the intestine, where more SN-38 is formed by two processes: (i) intestinal carboxylesterase metabolism of irinotecan and (ii) deconjugation of glucuronidated SN-38 by bacterial beta-glucuronidase (Takasuna *et al.*, 1998; Wadkins *et al.*, 2004). SN-38 is also delivered to intestine by direct secretion across intestinal exporters (Itagaki *et al.*, 2005). Collectively, it seems that exposure of the intestines to SN-38 is most likely the cause of intestinal toxicity associated with irinotecan therapy, with interindividual variation in the expression of the enzymes involved potentially linked to the relative susceptibility of patients to adverse drug reactions.

### 5.4 Kidney and Bladder Toxicity

The kidneys are similar to the intestines in that they are not only exposed to a wide array of metabolites generated in the liver and excreted (in this case in the urine), but are also capable of metabolizing drugs *in situ*. Many DMEs including CYPs (e.g. CYP1A1, 2D6, and 3A4), prostaglandin synthetase, GST, and epoxide hydrolase are expressed in kidneys, although their distribution varies markedly between different cell types (Korashy, Elbekai and El-Kadi, 2004). In addition to effects related directly to the unchanged drug itself, renal toxicity can therefore occur either as a consequence of metabolism in the kidneys (especially in cells of the proximal renal tubules), or as a secondary event from exposure to drug metabolites formed in the liver.

An example that shows a role for liver metabolism in renal toxicity is methoxyflurane, a general anesthetic that undergoes extensive CYP-mediated metabolism in the liver, producing fluoride ions that cause renal toxicity by inhibiting chloride transport in the ascending loop of Henle (Mazze, Trudell and Cousins, 1971; Roman *et al.*, 1977). Another example is cyclophosphamide, which causes hemorrhagic cystitis and increases the risk of bladder cancer. Cyclophosphamide is metabolized in the liver by CYPs to a metabolite that undergoes spontaneous decomposition to yield phosphoramide mustard and acrolein (Sladek, 1988). Acrolein is a toxic fragment excreted in the urine and responsible for bladder injury following prolonged exposure to high concentrations in urine (Stillwell and Benson, 1988).

There are a few drugs that illustrate how *in situ* metabolism in the renal cells can cause toxicity. Efavirenz, a widely prescribed non-nucleoside reverse transcriptase inhibitor for the treatment of HIV infection, produces renal tubular epithelial cell necrosis in rats but not in humans. Detailed studies revealed that rats produce a unique glutathione conjugate that is further processed by the renal gamma-glutamyltranspeptidase (GGT) to yield reactive metabolites and renal toxicity. Pretreatment with acivicin, a potent and selective inhibitor of GGT, prevented or reduced the nephrotoxicity of efavirenz in rats, which supports the hypothesis that renal processing of glutathione conjugate is responsible for the adverse drug reactions (Mutlib *et al.*, 2000). Cisplatin is another example of a drug that undergoes bioactivation in kidneys and treatment selectively destroys renal proximal tubule cells. The glutathione conjugate of cisplatin is further processed in kidney proximal tubular cells by enzymes, GGT, aminopeptidase, and cysteine *S*-conjugate beta-lyase, sequentially to a cysteinyl-glycine-conjugate, a cysteine *S*-conjugate and finally to a reactive platinum-thiol compound that is likely responsible for the kidney-specific toxicity (Townsend *et al.*, 2003). Pretreatment with acivicin or aminooxyacetic acid (an inhibitor of cysteine *S*-conjugate beta-lyase) protected mice from cisplatin nephrotoxicity (Townsend and Hanigan, 2002). We should dispel the notion that toxicity is always as simplistic as a reactive metabolite functionally inactivating a single essential target molecule—clearly other parallel processes can be involved. In the case of cisplatin, for instance, the selective deposition of platinum in mitochondria in kidneys, as seen in mice, along with a selective inactivation of alpha-ketoglutarate dehydrogenase enzyme complex, as demonstrated in LLC-PK1 cells, could also contribute to renal injury from cisplatin therapy (Zhang *et al.*, 2006). Acetaminophen is also associated with nephrotoxicity. Bioactivation of acetaminophen to the reactive intermediate NAPQI by renal CYP 2E1 and/or processing of glutathione conjugates of acetaminophen by renal GGT is implicated in its toxicity (Hart *et al.*, 1994).

### 5.5 Pulmonary Toxicity

The lung consists of various cell types, of which type II alveolar epithelial cells, pulmonary alveolar macrophages, capillary endothelial cells, and non-ciliated bronchiolar epithelial cells (Clara cells) are especially important as they express many enzymes of drug metabolism. Esterases, acid phosphatases, CYPs, and conjugating enzymes are all found in type II alveolar epithelial cells. Levels of CYPs are substantially higher (≤10×) in Clara cells than in other pulmonary cell types (reviewed in Gram, 1997).

Nitrofurantoin, an anti-infective used to treat urinary tract infections, with chronic use is associated with pneumotoxicity, as evidenced by cough, dyspnea, infiltrates in lung, and pulmonary fibrosis (Witten, 1989). It displays covalent binding *in vitro*, to a greater extent under anaerobic conditions than under the aerobic conditions necessary for CYP reactions. When nitrofurantoin is incubated with lung microsomes and NADPH under aerobic conditions, large amounts of superoxide are formed. Recombinant cytochrome P450 reductase was found to support redox cycling of nitrofurantoin and to generate ROS. The evidence collectively points to the formation of ROS (e.g. superoxide, $H_2O_2$, and hydroxyl radical) as the cause of pneumotoxicity from nitrofurantoin, rather than a result of protein covalent binding from putative reactive intermediates (Martin, 1983; Wang *et al.*, 2008). Nilutamide, another nitroaromatic drug (used in the treatment of prostatic carcinoma), also causes pulmonary interstitial fibrosis by ROS formation (Seigneur *et al.*, 1988; Berger *et al.*, 1992). *In vitro*, nilutamide is reduced by rat lung microsomes NADPH-cytochrome reductase into a nitroanion free radical which under aerobic conditions further undergoes redox cycling with the generation of ROS (Berger *et al.*, 1992).

Bleomycin, an antineoplastic agent used in the treatment of lymphomas and squamous cell tumors in humans, is limited by a dose-dependent pneumonitis often progressing to interstitial pulmonary fibrosis. The cytotoxicity of bleomycin is thought to be due to single- and double-strand cleavage of DNA by the bleomycin-Fe(II)-oxygen complex (Scheulen *et al.*, 1981). *In vivo*, bleomycin is inactivated by bleomycin hydrolase, whose level of activity is inversely correlated with the susceptibility of tissue or species to bleomycin toxicity. For example, mice lack bleomycin hydrolase activity in lung and hence are susceptible to toxicity, in contrast to rabbits that have high levels of bleomycin hydrolase activity and are relatively resistant to lung damage. The lung and skin, target sites for bleomycin toxicity, are low in bleomycin hydrolase activity compared with resistant tissues such as the liver and kidney (Lazo and Humphreys, 1983; Sebti, DeLeon and Lazo, 1987).

### 5.6 Neurotoxicity

Neurons are capable of metabolism and hence susceptible to toxicity from drugs that are bioactivated. It should be noted that such an occurrence is relatively rare given the predominance of hepatic metabolism, as well as the limited access to the central nervous system imposed by the blood–brain barrier when functioning under normal physiological conditions. Haloperidol, a neuroleptic agent, is associated with adverse reactions that include Parkinsonism and tardive dyskinesia. The intermediate pyridinium metabolite, 4-(4-chlorophenyl)-1-(4-fluorophenyl)-4-oxobutylpyridinium [HPP(+)], is formed by CYP3A4 from haloperidol and can be detected in the urine and brain of humans on haloperidol therapy. HPP+ is a potent inhibitor of complex I of the

mitochondrial respiratory chain and causes ATP depletion with subsequent neurotoxicity (Subramanyam *et al.*, 1991; Eyles *et al.*, 1997; Avent, DeVoss and Gillam, 2006).

Additionally, ROS-mediated toxicities are also described for neurons. For example, peripheral neuropathy is seen in some patients on isoniazid therapy that is attributed primarily to the perturbation of vitamin B6 metabolism; however, some of the intermediates of isoniazid metabolism are also thought to contribute to neuropathy via formation of ROS (Sanfeliu, Wright and Kim, 1999). Similarly, induction of ROS via redox cycling of quinone-containing anthracycline adriamycin is implicated in somnolence or chemobrain (loss of memory or executive functions) in patients on chemotherapy—since ROS-mediated protein oxidation and lipid peroxidation is seen in the brain of treated mice (Joshi *et al.*, 2005). Cisplatin is associated with sensory neuropathy in patients on chemotherapy. The neuropathy induced by cisplatin can be explained by apoptosis of neurons in the dorsal root ganglion induced by oxygen radicals, as seen by accumulation of p53 and Fas/Fas-L proteins; since antioxidants can effectively protect the neurons from cisplatin-induced neuropathy, ROS are implicated in the pathogenesis (Smyth *et al.*, 1997; Gill and Windebank, 1998).

### 5.7 Reproductive Organ Toxicity

Spironolactone is a diuretic associated with inhibition of steroidogenesis in testes. Its metabolite, 7-alpha-thiospironolactone, is converted to a reactive intermediate that results in destruction of testicular CYPs. SU-10603, an inhibitor of 17-alpha-hydroxylase, prevented degradation of testicular CYP by 7-alpha-thiospironolactone, which indicates a prominent role for 17-alpha hydroxylase in the formation of a reactive intermediate (Menard *et al.*, 1979; Kossor *et al.*, 1992).

Cyclophosphamide is an ovarian toxicant that is shown to target dormant primordial ovarian follicles in animals and induce apoptosis of human granulosa cell line (Plowchalk and Mattison, 1991; Tsai-Turton *et al.*, 2007). Phosphoramide mustard, a metabolite of cyclophosphamide, is likely involved in this toxicity and is toxic *in vitro* to mouse ovary cultures as well as *in vivo* in mice. Acrolein, another reactive fragment from cyclophosphamide metabolism, was not found to cause ovarian toxicity in mice (Plowchalk and Mattison, 1991; Desmeules and Devine, 2006).

### 5.8 Adverse Drug Reactions in Skin

Cutaneous adverse drug reactions are seen with many compounds and present symptomatically as a broad spectrum, ranging from minor rash to epidermal necrosis. Stevens–Johnson syndrome and toxic epidermal necrolysis are two extreme conditions of adverse drug reactions, both characterized morphologically by apoptosis and the rapid onset of keratinocyte cell death. Use of therapeutic drugs was reported in over 95% of cases with toxic epidermal necrolysis and in about 50% of cases with Stevens–Johnson syndrome (French, 2006). More than 100 drugs, including sulfonamides, quinolones, cephalosporins, and tetracyclines are associated with these conditions (French, 2006). Generally, these reactions are considered to be immune-mediated and initiated by the modification of proteins by reactive intermediates of drugs. Keratinocyte apoptosis mediated by Fas is the pathological feature of this condition. Intravenous immunoglobins containing anti-Fas antibodies has been used with some success in treating toxic epidermal necrolysis (Tan *et al.*, 2005; French, Trent, and Kerdel, 2006). Reactions with sulfonamides are attributed to its reactive metabolites and the resultant oxidative

stress in cells. Intracellular protein adducts were seen in human dermal fibroblasts incubated with sulfamethoxazole, for instance, which suggests a mechanistic role for reactive intermediates in the observed cutaneous toxicity (Bhaiya *et al.*, 2006). Similarly, dapsone is associated with cutaneous drug reactions via its bioactivation to a reactive arylhydroxylamine metabolite in human epidermal keratinocytes (Vyas *et al.*, 2005).

The relative expression of CYPs in the skin is notably distinct from that in the liver, with inducible forms such as CYP1A1, 1B1, and 2E1 much more dominant even at basal levels (Bergström *et al.*, 2007). The metabolic reaction phenotype of a drug, especially when it involves the generally regarded "lesser" hepatic enzyme forms, should always be considered whenever adverse reactions are encountered and marked by a distinctive tissue specificity.

### 5.9 Other Organ Systems

Any organ or cell system can be a target for toxicity (usually involving the pharmacology of the drug), although the majority of cases in drug development will involve the aforementioned tissues and organs. A few examples that describe metabolism-related adverse drug reactions and toxicity in other organs now follow.

Ototoxicity is one of the limiting side effects of cisplatin and is characterized by bilateral high-frequency hearing loss in patients (Anniko and Sobin, 1986). Oxidative stress is implicated in cisplatin ototoxicity. In the rat, chochlear damage induced by cisplatin is associated with the generation of ROS and probably nitric oxide, as inferred from up-regulation of an inducible form of nitric oxide synthase in damaged cochlear tissue (Kopke *et al.*, 1997; Li, Liu and Frenz, 2006). Protection from ototoxicity is seen in rats given cytoprotective antioxidants, like D- or L-methionine, which further supports the hypothesis that oxidative stress is responsible for cisplatin ototoxicity (Li *et al.*, 2001).

In addition to its hematotoxicity, clozapine is also associated with myocarditis. This toxicity is likely a result of a reactive nitrenium ion metabolite, such as formed by microsomes prepared from cardiac tissue (Iverson, Zahid and Uetrecht, 2002). Anthracycline quinine-containing anticancer agents adriamycin and daunorubicin (anthracyclines) exert their therapeutic effect by generating ROS via quinone-hydroxyquinone redox cycling. Toxicity is therefore dependent on the tissue's ability to defend against ROS. For example, cardiotoxicity seen with anthracycline therapy is attributed to an inability of the myocytes to defend against the redox cycling (Doroshow, Locker and Myers, 1980; Berthiaume and Wallace, 2007).

Spironolactone can inhibit steroidogenesis in adrenal cortex leading to toxicity (Menard *et al.*, 1979). The mechanism involves a series of enzymes with *in situ* metabolism of spironolactone in the adrenal gland. Enzymes in adrenal microsomes deacetylate spironolactone to form 7-alpha-thiospironolactone, which is subsequently metabolized by 17-alpha-hydroxylase to form a reactive entity that covalently modifies and inhibits CYP and 17-alpha hydroxylase and their functional role in steroidogenesis (Colby *et al.*, 1991; Kossor *et al.*, 1991).

## 6 CARCINOGENESIS

Carcinogenesis is a widely studied and specialized toxicity. Significant time and resources are expended during the drug development process to assess the potential of drug

candidates to cause cancer. The mechanisms of bioactivation and DNA adduct formation are well studied for carcinogens such as aflatoxin. Fortunately, relatively few examples exist of drugs used in therapy that demonstrate carcinogenesis mediated through metabolic activation. Screening for this toxicity is a requisite part of drug development and the Ames assay with and without liver S9 for bioactivation helps to eliminate many drugs that would otherwise pose a risk to humans.

Tamoxifen is a nonsteroidal antiestrogen used in the treatment of breast cancer. Tamoxifen is clastogenic in lymphoblastoid cell line expressing CYPs and hepatocarcinogenic in animal models, including rats (Dragan *et al.*, 1994; Styles *et al.*, 1994). However, even though tamoxifen is usually given as long-term therapy, there is no conclusive evidence of hepatocarcinogenesis in women. The apparent species difference seen in the toxicity of tamoxifen appears to be explained by differences in how animals and humans biotransform and eliminate the drug and its metabolites. Tamoxifen is metabolized to an alpha-hydroxylated intermediate that is subsequently either conjugated by glucuronidation, or converted to a genotoxic sulfate metabolite that collapses to a reactive carbocation that forms DNA adducts (Martin *et al.*, 1995; Park *et al.*, 2005). *In vitro*, rat liver enzymes favor the sulfonation reaction, thereby bioactivating tamoxifen, whereas human liver enzymes favor the glucuronidation pathway and elimination (Boocock *et al.*, 1999, 2000). The difference in relative rates of hepatic bioactivation and bioinactivation of tamoxifen in rats and humans seems to explain the difference in susceptibility of rodents to a genotoxic insult. This provides a good example of how species can differ in their susceptibility to toxicity and, if evidence can be provided that the mechanisms involved are specific to an animal species and not present in humans, it is possible that a drug candidate can progress in clinical trials and gain marketing approval. Although there is not enough evidence for hepatocarcinogenesis, there is an association with long-term tamoxifen therapy and an increased incidence of endometrial cancer in patients (Colacurci *et al.*, 2000). The precise mechanism of carcinogenesis in the endometrium and the relative susceptibility is unknown, but with metabolic activation seen in explant cultures of human endometrium it is speculated that bioactivation is involved, similar to that involving liver CYPs (Sharma, Shubert, Sharma *et al.*, 2003).

Estrogens are known to cause cancer (e.g. breast cancer). Natural estrogens (estrone and estradiol) have been shown to form quinone metabolites that can react with DNA and initiate carcinogenesis. It should, however, be noted that estrogen receptor-mediated tumor promotion is a better understood mechanism of carcinogenesis. Diethylstilbestrol, a synthetic nonsteroidal estrogen, was used in the United States until the 1970s to prevent threatened miscarriage. It worked by stimulating placental synthesis of progesterone and estrogen. Maternal use of diethylstilbestrol prior to gestational age of 18 weeks was associated later with incidences (<0.1%) of vaginal clear cell adenocarcinoma of female offspring at puberty and non-adenocarcinoma reproductive tract problems in an estimated 95% of female offspring (Herbst, Ulfelder and Poskanzer, 1971; Melnick *et al.* 1987). A genotoxic epoxide intermediate of diethylstilbestrol formed by fetal CYPs during gestation is thought to be responsible for the subsequent carcinogenesis at puberty (Balling *et al.*, 1985).

The anticonvulsant phenytoin (diphenylhydantoin) is genotoxic *in vitro*. Its major nontoxic metabolite, 5-(p-hydroxyphenyl)-5-phenylhydantoin (HPPH), is conjugated by UGTs and eliminated; however, it is also bioactivated by peroxidases such as prostaglandin H synthase to form a reactive intermediate that initiates DNA oxidation and genotoxicity (e.g. micronuclei formation in rat skin fibroblasts) (Kim *et al.*, 1997). An

individual deficiency in UGT activity (e.g. via depletion of UDPGA and/or polymorphism) therefore plays an important role in a patient's susceptibility to phenytoin genotoxicity (Kim *et al.*, 1997).

Nitrofurazone is an antimicrobial drug used in veterinary medicine and a carcinogen in experimental animals and also a suspected human carcinogen (IARC, 1990). The carcinogenesis of nitrofurazone possibly results from the initiation of redox cycling during metabolic activation, with generation of ROS and resultant oxidative damage to the DNA (Hiraku *et al.*, 2004). Hydroxylamine derivatives can also form DNA adducts that may play a role in carcinogenesis (Fan, Schut and Cytotoxicity, 1995).

## 7 TERATOGENESIS

Teratogenesis is another widely studied and specialized toxicity, with a major effort and investment made in the drug development process to assess the potential of drug candidates to cause birth defects. The expressions of DMEs in the fetus and embryo are notably different to that seen in adult humans and a developmental role is suggested (e.g. NAT1 expression in the fetus is implicated in neurulation, with deficiencies associated with cleft palate formation). There is a marked ontogeny that can factor significantly in bioactivation and the mechanism involved in adverse drug reactions. CYP3A7, for instance, is expressed as the major CYP form in the human fetus, but within days of birth it is virtually absent in the liver of neonates. The liver of human fetus or embryo has the capacity to metabolize compounds, and other fetal extrahepatic tissues such as adrenal gland also express CYP enzymes. The human placenta also expresses CYPs and contributes to the metabolism of xenobiotics, albeit to a minor extent (Hakkola *et al.*, 1998). Collectively, the feto-placental unit contributes to the bioactivation of several known teratogens. Although there are many compounds—including drugs—that are known to be teratogenic, relatively few undergo bioactivation to embryotoxic reactive intermediates. Carbamazepine, phenytoin, and tolbutamide are presented as examples of such teratogens.

Carbamazepine when used during pregnancy is associated with fetal malformations such as craniofacial defects, spina bifida, and fingernail hypoplasia (Rosa, 1991). The teratogenicity of carbamazepine is associated with the metabolite carbamazepine-10,11-epoxide and also the oxidation of the epoxide or carbamazepine at positions on the aromatic ring that can form putative reactive arene oxide or quinone intermediates (Bennett *et al.*, 1996). Coadministration of stiripentol, a broad spectrum CYP inhibitor, protects mice fetus from carbamazepine effects and provides evidence that CYP bioactivation is required for teratogenesis (Finnell *et al.*, 1995).

Phenytoin is teratogenic in animals and human. The teratogenicity of phenytoin is associated with its bioactivation to embryotoxic reactive intermediates. As discussed previously, the major nontoxic metabolite of phenytoin, HPPH, is conjugated by UGTs and eliminated, thus offering protection from alternative pathways including bioactivation. HPPH is also bioactivated by peroxidases (e.g. PHS) to form a reactive intermediate that is genotoxic and implicated in teratological initiation (Kim *et al.*, 1997). UGTs play a role in protecting against the teratogenic effects of phenytoin.

Thalidomide is now used as an immunomodulatory agent but when originally introduced, it was as a sedative/hypnotic (used as a sleeping aid and to ameliorate morning sickness in pregnant women). It was tragic that thalidomide use was linked to numerous

cases of birth defects—estimated to be as high as 7000 worldwide (Lenz, 1966). Thalidomide is metabolized by PHS to a teratogenic intermediate as demonstrated by significant reduction in birth defects when rabbits were pretreated with acetylsalicylic acid, an inhibitor of PHS (Arlen and Wells, 1996). One of the proposed molecular mechanisms of thalidomide teratogenicity is the oxidative damage to cellular macromolecules in the embryo. Thalidomide causes a substantial increase in embryonic DNA oxidation in rabbits but not in mice, which do not exhibit thalidomide-induced teratogenicity. When rabbits are pretreated with a free radical spin-trapping agent, alpha-phenyl-*N*-tert-butylnitrone, prior to thalidomide administration, the embryonic DNA oxidation is blocked and teratogenicity is not observed (Wells *et al.*, 2005).

As discussed earlier, glutathione is present in cells usually at relatively high concentrations and serves to protect tissues from oxidative stress. This nucleophile must be continuously replenished by reduction of oxidized glutathione by glutathione reductase. If glutathione homeostasis is compromised, it can lead to an increased risk of teratogenicity (Hales and Brown, 1991). Tolbutamide, for instance, is a hypoglycemic agent used in the treatment of non-insulin-dependent diabetes mellitus and another drug known to be teratogenic in animals and humans. Tolbutamide teratogenicity is attributed to the reactive metabolites *n*-butyl isocyanate and *S*-(*n*-butylcarbamoyl) glutathione, which can carbamoylate and inhibit glutathione reductase (Guan *et al.*, 1999).

## 8 ABROGATION/MITIGATION OF BIOACTIVATION—CASE EXAMPLES

Drug evaluation is conducted on a case-by-case basis, which can prove costly and time-consuming. Fortunately, historical evidence has led to the compilation of "structural alerts" that are used to focus and optimize chemistry efforts in drug discovery and maximize the opportunity for success in development (Table 2). One strategy adopted in the drug discovery stage is to avoid these features entirely, while another is to evaluate their safety specifically and with full awareness of anecdotal issues related to such structural motifs. The latter strategy is supported by the many examples that exist of drugs that required only relatively minor structural changes to improve their safety profile and yet retain the desired pharmacological properties. This is not necessarily an easy task.

In this section, clozapine and olanzapine, captopril and enalapril, tolcapone and entacapone, and ritonavir and indinavir are discussed as examples (Table 4).

Functional groups such as thiophenes can be oxidized to form reactive species that have a propensity to react with nucleophilic macromolecules. In the case of tienilic acid, its metabolite is sufficiently reactive to inactivate the CYP2C9 enzyme involved in its formation (Melet *et al.*, 2003). Cases involving nitroaromatics, hydrazines, and anilines have been implicated in toxicity, including mutagenicity, carcinogenicity, and hepatotoxicity and issues with drugs like isoniazid and procainamide (Freeman *et al.*, 1981; Blanco *et al.*, 1998; Bomhard and Herbold, 2005; Maeda *et al.*, 2007). Molecules with quinone or quinoid functional groups can conjugate in a manner that leads to covalent binding with nucleophiles, as exemplified by acetaminophen and troglitazone, both drugs possessing marked hepatotoxicity at toxic dosages (He *et al.*, 2004; Shin *et al.*, 2007). The hydrazine-based drugs, hydralazine and isoniazid, are associated with autoimmune disease and hepatotoxicity, respectively (Hari, Raza and Clayton, 1998; Sarich *et al.*, 1999). Drugs containing benzylamine (such as terbinafine), a metabolic soft spot, can undergo a variety of biotransformations, forming aldehydes, carboxylic acids,

hydroxylamines, and GSH conjugates (Mallat *et al.*, 1997). Thiol-containing drugs such as methamizole, penicillamine, and captopril cause toxicity by covalently modifying proteins by reacting with cysteinyl-disulfide linkages or by oxidation to reactive sulfenic acid (Jaffe, 1986; Decker and Doerge, 1992; Piepho, 2000). The thiazole ring can undergo oxidative scission to form reactive intermediates, as seen with sudoxicam and associated with its hepatotoxicity (Hobbs and Twomey, 1977). Carboxylic acid-containing drugs are predisposed to form acyl glucuronides, which are potentially reactive and toxic metabolites and known products of carboxylic acid NSAIDs (e.g. diclofenac and zomepirac) (Bailey and Dickinson, 1996; Seitz and Boelsterli, 1998). Unsaturated carbon–carbon double and triple bonds are electron-rich targets and subject to oxidation and formation of epoxides or oxirene metabolites, respectively—these are associated with subsequent reactions including toxicity and mechanism-dependent CYP inhibition. For instance, 17alpha-ethynylestradiol is metabolized by CYP3A4, a process that results in the inactivation of this enzyme by reactive acetylenic intermediate (Guengerich, 1990).

The clinical utility of clozapine is limited by its known propensity to cause agranulocytosis, hepatotoxicity, and cardiotoxicity. The reactive nitrenium ion of clozapine is thought to be responsible for its toxicity. Olanzapine is a close structural analog of clozapine, but associated with much lower incidences of toxicities than clozapine. Although olanzapine can be metabolized to a reactive nitrenium ion, the significantly lower clinical dose of olanzapine compared with clozapine provides a lower overall systemic exposure to the nitrenium ion, which presumably explains its low incidence of agranulocytosis (Gardner *et al.*, 1998; Uetrecht, 2001). In addition, olanzapine is also metabolized to an *N*-glucuronide, which likely circumvents the bioactivation pathway and contributes to its overall improved safety relative to clozapine (Kalgutkar *et al.*, 2005). Similarly, captopril, an angiotensin-converting enzyme (ACE) inhibitor, is associated with bone marrow toxicity and skin rashes, linked to the thiol moiety. Enalapril, a subsequently developed ACE inhibitor, lacks the thiol goup and is not associated with as many of the toxicities seen with captopril (Piepho, 2000; Kalgutkar *et al.*, 2005).

Tolcapone, a catechol-*O*-methyltransferase inhibitor used in the treatment of Parkinson's disease, is associated with a relatively high incidence of liver toxicity. Reduction of the nitro group in tolcapone and subsequent generation of the reactive quinoneimine is implicated in its hepatotoxicity (Smith *et al.*, 2003). In contrast, entacapone is structurally related to tolcapone but is not associated with hepatotoxicity. The metabolism of entacapone is governed primarily by glucuronidation and not reduction of the nitro group, which results in an overall low exposure to its reactive quinoneimine (Lautala *et al.*, 2000; Smith *et al.*, 2003). In addition to metabolic activation, direct mitochondrial toxicity of tolcapone but not entacapone is also implicated in tolcapone hepatotoxicity, based on studies with isolated mitochondria (Boelsterli *et al.*, 2006). Ritonavir, an HIV protease inhibitor, is a known mechanism-based inactivator of CYP3A4 through oxidative thiazole ring opening to form reactive intermediates. The HIV protease inhibitors indinavir and saquinavir lack the thiazole ring and do not inactivate CYP3A4 (Kempf *et al.*, 1997; Koudriakova *et al.*, 1998).

These examples illustrate the benefit of understanding the mechanism of toxicity, the biotransformation pathways, and the reaction phenotype involved in the adverse drug reactions. Relatively small structural changes that minimize or block the formation of reactive intermediates can profoundly affect the safety of a drug, especially when those changes are coupled with increased potency at the target, reduced metabolic clearance, and overall lower dosages.

**TABLE 4 Structural modification and mitigation of toxicity.**

| Drug with relatively greater liability due to reactive intermediate generation | Drug with mitigated effect |
|---|---|
| Clozapine 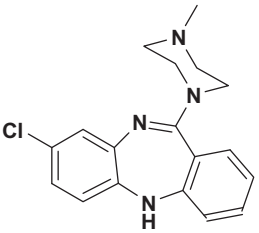 | Olanzapine 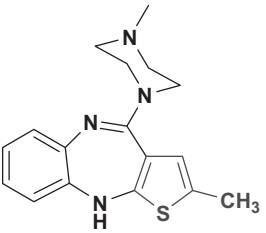 |
| Captopril 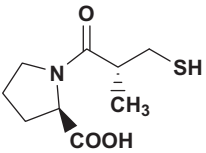 | Enalapril 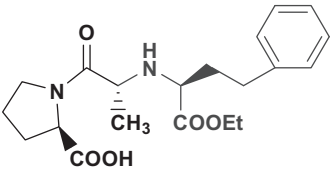 |
| Tolcapone 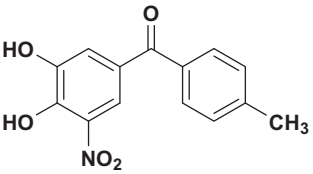 | Entacapone 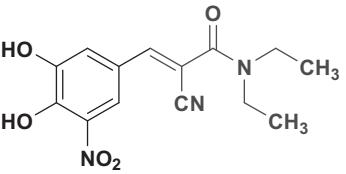 |
| Ritonavir 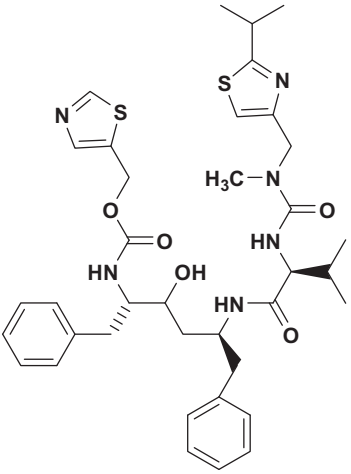 | Indinavir 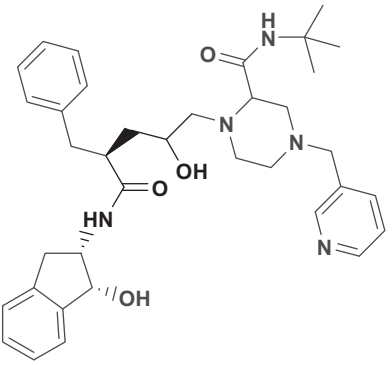 |

## 9 EXPERIMENTAL METHODS FOR SCREENING

The early detection of reactive metabolites and prediction of an associated potential liability in drug development is useful in deciding the fate of a novel drug candidate and/or optimizing the structure-activity relationship (SAR) of the chemotype. Part of the repertoire available to medicinal chemistry includes reactivity screens that can be performed *in silico, in vitro*, or *in vivo*.

A number of software are available commercially that can help predict metabolites, potential for mutagenicity and carcinogenicity (Caldwell and Yan, 2006). Systems such as METEOR and MetabolExpert exist for prediction of metabolites and DEREK and HazardExpert exist to help identify potential genotoxicants. These programs rely on a set of rules derived from the available metabolism and toxicological knowledge (e.g. structural alerts). There are also systems like TOPKAT and MCASE that use toxicity data from known drugs to help determine the potential genotoxicity (CompuDrug International, Inc.; Lhasa Limited; Accelrys; Multicase Inc.). *In silico* predictions are very attractive and easy to perform; however, such predictions are not always accurate and should be used to provide alerts for a chemotype and subsequent testing in relevant assays (e.g. miniAmes).

There are several methods available to identify and quantify reactive metabolites *in vitro*. Time-dependent CYP inhibition is one commonly used *in vitro* method. Another is the qualitative or quantitative determination of adduct formation in liver microsomes. This is usually performed using either a co-incubated soft nucleophile (e.g. glutathione, *N*-acetyl-cysteine, or *N*-acetyl-lysine) to trap soft electrophiles (e.g. epoxides, arene oxides, nitrenium, and/or quinoneimines), or a hard nucleophile (e.g. potassium cyanide and sodium cyanide) to trap hard electrophiles. If a radiolabeled drug candidate is available, then its covalent binding to protein can be measured *in vitro* using liver microsomes from human or any other species of interest (usually done with and without NADPH and with and without a co-incubated nucleophile). Based on liver microsomal covalent binding, Evans *et al.* (2004) concluded that a binding activity of >50 pmol compound per mg of total liver protein poses an increased risk of reactive metabolite formation *in vivo* and associated adverse drug reactions. Non-liver microsomal *in vitro* assays also exist, including MPO for reactive metabolite generation; fluorometric probes such as 2′,7′-dichlorodihydrofluorescein for ROS detection. There are also cell-based assays to evaluate the functional consequence of reactive metabolite formation, and primary hepatocytes or immortalized cell lines expressing CYPs are routinely used in drug discovery to evaluate the hepatotoxic potential. Lastly, the Ames test in *S. typhimurium* and *E. coli* using Arochlor-induced rat liver S9 enzyme fraction, is a well-established *in vitro* assay to evaluate the potential for mutagenicity.

The *in vivo* determination of glutathione adducts, as well as qualitative and quantitative determination of covalently bound proteins, either by a radiolabeled compound or immunological methods, can be useful—albeit relatively labor- and resource-intensive. There are also *in vivo* clastogenicity assays, such as the comet assay, that are an important part of the drug development process. These *in vivo* assays are probably best used to show the mitigation of an *in vitro* liability (e.g. by drug clearance) and/or to identify extrahepatic and non-CYP-mediated toxicity that may not be indicated by routine *in vitro* screens and assays.

When toxicity indicates a role for DMEs in the pathogenesis, reaction phenotyping and drug–drug interaction studies can prove informative in evaluating and potentially even predicting the susceptibility of specific patient populations to adverse drug

reactions (e.g. CYP3A4 interactions and terfenadine cardiac arrhythmias; CYP2E1 induction and acetaminophen hepatotoxicity; UGT1A1 polymorphism and irnitoecan gastrointestinal toxicity).

## 10 SUMMARY

A reasonable prediction of a drug candidate's safety, safety or exposure multiples, and/or safe starting dose for clinical trials can be made if the exposure limits and mechanism of action are known for its pharmacodynamic and toxicodynamic effects. The fate and effects of a drug and its metabolites can be measured *in vivo* using animal models as surrogates for human anatomy and physiology, and also *in vitro* using tissue available from animals and human donors; interspecies scaling and extrapolation factors are then applied to optimize the prediction of the drug's behavior in humans. Normally, preclinical animal models provide the no observed effect level (NOEL) and/or the no observed adverse effect level (NOAEL), which are combined with the target efficacious exposure to predict safety/exposure multiples in humans and drive a safe and viable evaluation of drug candidates in human volunteers and patients.

None of these *a priori* predictions is guaranteed and the estimates serve instead to provide a conservative and ethically acceptable safe starting dose. Human clinical trials are typically conducted with the drug candidate formulated at a starting dose that is a significantly low fraction of the preclinical NOEL or NOAEL. Evaluating the potential toxicity and predicting the margins of safety in humans helps to mitigate the risk in human clinical trials and has become *sine qua non* in the modern drug discovery and development process. Achieving that objective is not always easy and is often iterative. In many cases, one or more backup drug candidates will be required to address observed issues encountered with lead compounds.

While potential adverse drug reactions can be predicted for the majority of drug candidates that advance through the optimization process, such predictions cannot be made as easily when compounds form reactive metabolites. Drugs such as acetaminophen, diclofenac, and ibuprofen are known to produce reactive intermediates and are associated with occasional toxicities (overdose, idiosyncratic), and yet these and many others like them occupy a prominent position in the market place being invaluable in their therapeutic potential and use. If the benefits outweigh the risks, then a properly characterized and monitored drug candidate with potential for human toxicity can still become a successful therapeutic. A risk–benefit analysis, therefore, is particularly critical in deciding the fate of a drug or a drug candidate that forms a reactive intermediate. This analysis should consider a wide variety of factors that include the dose, the therapeutic area, target population, urgency of medical need, duration of treatment, availability of safer alternatives, and post-marketing surveillance. This is often a dynamic process within and between the industry and regulatory agencies.

A low predicted human dose (e.g. 10 mg) is a major factor in mitigating the risk associated with a reactive metabolite (Uetrecht, 2001). For example, 17alpha-ethynylestradiol is metabolized to reactive intermediates that can inactivate CYP3A4 but it is not associated with idiosyncratic toxicity in humans, probably because of its very low therapeutic dose (<0.035 mg). The hepatotoxicity associated with troglitazone (200–400 mg) but not pioglitazone (<10 mg) is very likely because of the difference in therapeutic dose.

Pragmatically, a certain level of risk must be accepted when the benefits are known to outweigh the risk, such as in the treatment of unmet medical conditions or serious medical conditions. For example, felbamate is an antiepileptic drug that shortly after its introduction was reported to be associated with multiple cases of aplastic anemia, hepatotoxicity, and deaths. Felbamate is, however, an effective therapy in refractory patients and instead of withdrawing the drug entirely, its use is limited to only that population of patients. Amodiaquine is associated with cases of life-threatening agranulocytosis and hepatotoxicity, yet as a malarial therapy it is effective against chloroquine-resistant isolates of *Plasmodium falciparum*. Its clinical use is thus limited to the treatment of acute malaria and not for prophylaxis.

A careful monitoring program can also help manage the risk of useful drugs, as in the case of bosentan, a drug used in treatment of pulmonary hypertension. Bosentan's approval as a drug included a risk management program, limited distribution, and monthly monitoring for hepatotoxicity (Temple, 2006). Measures such as these are clearly expensive but do allow potentially dangerous drugs to be used optimally in those patients who will benefit the most. Thalidomide, because of teratogenicity, is now restricted in its use and prescribed for immunomodulation only when the prescribing physicians and patients are enrolled in the System for Thalidomide Education and Prescribing Safety (STEPS). This is an oversight program that helps to ensure that pregnancy does not occur during therapy.

The availability of safer alternatives is an important factor that decides the fate of drugs. For example, bromfenac was withdrawn from the market because of its higher incidence of severe hepatotoxicity relative to the already existing alternatives (e.g. diclofenac) and because it offered no significant advantage over the others. When troglitazone was introduced to the market, it was unique in its mechanism of action and appropriately labeled for its risk; however, troglitazone was soon withdrawn after the introduction of rosiglitazone, which was without any reports of hepatotoxicity and deemed to be a safer drug in the same class.

There are numerous factors to consider in finding and developing a novel drug that are not limited to drug metabolism, although often this is key to understanding the potential fate and risks predicted for that drug in the patient population. There are also many things that are unknown about toxicity and drug metabolism, especially as it relates to differences between individuals in their susceptibility to adverse effects. And in many cases, the exact role of a given enzyme in the mechanism of toxicity is still poorly defined, if defined at all.

Some considerations of safety have become standardized, such as the evaluation of carcinogenesis; however, other factors encountered on a more singular basis often can become an issue and a potential showstopper to the success of a drug. Given that very few, if any, drugs have identical properties and issues, to properly understand the risk–benefit ratio of an experimental drug candidate it is prudent to approach its characterization on an individual level. This can be a daunting goal in the fast-paced, ever-changing and always competitive landscape of modern pharmaceutical research and development. Fortunately, the structure of a molecule can offer many clues to its fate. When this is combined with knowledge of the properties the experimental drug shares with its established ilk, a comprehensive understanding of the therapeutic target and the various mechanisms of toxicities, and also taking into consideration the characteristics of the enzymes commonly involved in metabolism and adverse drug reactions, a modern safety assessment can be tailored somewhat specifically to each new drug to help accelerate its development.

## ACKNOWLEDGMENT

The authors acknowledge the many thoughtful comments and suggestions from Dr. Stephen P. Adams.

## REFERENCES

Abdel-Razzak, Z., Loyer, P., Fautrel, A., Gautier, J.C., Corcos, L., Turlin, B., Beaune, P., Guillouzo, A. (1993). Cytokines down-regulate expression of major cytochrome p-450 enzymes in adult human hepatocytes in primary culture. *Mol. Pharmacol.*, *44* (4), 707–715.

Abernethy, D.R., Greenblatt, D.J., Divoll, M., Shader, R.I. (1983). Enhanced glucuronide conjugation of drugs in obesity: studies of lorazepam, oxazepam and acetaminophen. *J. Lab. Clin. Med.*, *101* (6), 873–880.

Aithal, G.P., Ramsay, L., Daly, A.K., Sonchit, N., Leathart, J.B., Alexander, G., Kenna, J.G., Caldwell, J., Day, C.P. (2004). Hepatic adducts, circulating antibodies and cytokine polymorphisms in patients with diclofenac hepatotoxicity. *Hepatology*, *39* (5), 1430–1440.

Anderson, N.L., Copple, D.C., Bendele, R.A., Probst, G.S., Richardson, F.C. (1992). Covalent protein modifications and gene expression changes in rodent liver following administration of methapyrilene: a study using two-dimensional electrophoresis. *Fundam. Appl. Toxicol.*, *18* (4), 570–580.

Anniko, M., Sobin, A. (1986). Cisplatin: evaluation of its ototoxic potential. *Am. J. Otolaryngol.*, *7* (4), 276–293.

Arand, M., Cronin, A., Adamska, M., Oesch, F. (2005). Epoxide hydrolases: structure, function, mechanism and assay. *Methods Enzymol.*, *400*, 569–588. Review.

Arlen, R.R., Wells, P.G. (1996). Inhibition of thalidomide teratogenicity by acetylsalicylic acid: evidence for prostaglandin H synthase-catalyzed bioactivation of thalidomide to a teratogenic reactive intermediate. *J. Pharmacol. Exp. Ther.*, *277* (3), 1649–1658.

Artemchik, V.D., Kurchenko, V.P., Metelitsa, D.I. (1985). Peroxidase activity of catalase with respect to aromatic amines. *Biokhimiia*, *50* (5), 826–832.

Atchison, C.R., West, A.B., Balakumaran, A., Hargus, S.J., Pohl, L.R., Daiker, D.H., Aronson, J.F., Hoffmann, W.E., Shipp, B.K., Treinen-Moslen, M. (2000). Drug enterocyte adducts: possible causal factor for diclofenac enteropathy in rats. *Gastroenterology*, *119* (6), 1537–1547.

Avent, K.M., DeVoss, J.J., Gillam, E.M. (2006). Cytochrome P450-mediated metabolism of haloperidol and reduced haloperidol to pyridinium metabolites. *Chem. Res. Toxicol.*, *19* (7), 914–920.

Back, D.J., Breckenridge, A.M., MacIver, M., Orme, M., Purba, H.S., Rowe, P.H., Taylor, I. (1982). The gut wall metabolism of ethinyloestradiol and its contribution to the pre-systemic metabolism of ethinyloestradiol in humans. *Br. J. Clin. Pharmacol.*, *13* (3), 325–330.

Bae, M.A., Song, B.J. (2003). Critical role of c-Jun N-terminal protein kinase activation in troglitazone-induced apoptosis of human HepG2 hepatoma cells. *Mol. Pharmacol.*, *63* (2), 401–408.

Baeuerle, P.A., Huttner, W.B. (1987). Tyrosine sulfation is a trans-Golgi-specific protein modification. *J. Cell Biol.*, *105* (6 Pt 1), 2655–2664.

Bailey, M.J., Dickinson, R.G. (1996). Chemical and immunochemical comparison of protein adduct formation of four carboxylate drugs in rat liver and plasma. *Chem. Res. Toxicol.*, *9* (3), 659–666.

Bailey, M.J., Dickinson, R.G. (1999). Limitations of hepatocytes and liver homogenates in modelling *in vivo* formation of acyl glucuronide-derived drug-protein adducts. *J. Pharmacol. Toxicol. Methods*, *41* (1), 27–32.

Bailey, M.J., Dickinson, R.G. (2003). Acyl glucuronide reactivity in perspective: biological consequences. *Chem. Biol. Interact.*, *145* (2), 117–137. Review.

Bailey, M.J., Worrall, S., de Jersey, J., Dickinson, R.G. (1998). Zomepirac acyl glucuronide covalently modifies tubulin *in vitro* and *in vivo* and inhibits its assembly in an *in vitro* system. *Chem. Biol. Interact.*, *115* (2), 153–166.

Balling, R., Haaf, H., Maydl, R., Metzler, M., Beier, H.M. (1985). Oxidative and conjugative metabolism of diethylstilbestrol by rabbit preimplantation embryos. *Dev. Biol.*, *109* (2), 370–374.

Banks, A.T., Zimmerman, H.J., Ishak, K.G., Harter, J.G. (1995). Diclofenac-associated hepatotoxicity: analysis of 180 cases reported to the Food and Drug Administration as adverse reactions. *Hepatology*, *22* (3), 820–827.

Basivireddy, J., Jacob, M., Pulimood, A.B., Balasubramanian, K.A. (2004). Indomethacin-induced renal damage: role of oxygen free radicals. *Biochem. Pharmacol.*, *67* (3), 587–599.

Beaune, P., Bourdi, M., Belloc, C., Gautier, J.C., Guengerich, F.P., Valadon, P. (1993). Immunotoxicology and expression of human cytochrome P450 in microorganisms. *Toxicology*, *82* (1–3), 53–60.

van den Bemt, P.M., Meyboom, R.H., Egberts, A.C. (2004). Drug-induced immune thrombocytopenia. *Drug Saf.*, *27* (15), 1243–1252.

Bennett, G.D., Amore, B.M., Finnell, R.H., Wlodarczyk, B., Kalhorn, T.F., Skiles, G.L., Nelson, S.D., Slattery, J.T. (1996). Teratogenicity of carbamazepine-10, 11-epoxide and oxcarbazepine in the SWV mouse. *J. Pharmacol. Exp. Ther.*, *279* (3), 1237–1242.

Berger, V., Berson, A., Wolf, C., Chachaty, C., Fau, D., Fromenty, B., Pessayre, D. (1992). Generation of free radicals during the reductive metabolism of nilutamide by lung microsomes: possible role in the development of lung lesions in patients treated with this anti-androgen. *Biochem. Pharmacol.*, *43* (3), 654–657.

Bergström, M.A., Ott, H., Carlsson, A., Neis, M., Zwadlo-Klarwasser, G., Jonsson, C.A., Merk, H.F., Karlberg, A.T., Baron, J.M. (2007). A skin-like cytochrome P450 cocktail activates prohaptens to contact allergenic metabolites. *J. Invest. Dermatol.*, *127* (5), 1145–1153.

Berson, A., Wolf, C., Chachaty, C., Fisch, C., Fau, D., Eugene, D., Loeper, J., Gauthier, J.C., Beaune, P., Pompon, D. *et al.* (1993). Metabolic activation of the nitroaromatic antiandrogen flutamide by rat and human cytochromes P-450, including forms belonging to the 3A and 1A subfamilies. *J. Pharmacol. Exp. Ther.*, *265* (1), 366–372.

Bertelsen, K.M., Venkatakrishnan, K., Von Moltke, L.L., Obach, R.S., Greenblatt, D.J. (2003). Apparent mechanism-based inhibition of human CYP2D6 in vitro by paroxetine: comparison with fluoxetine and quinidine. *Drug Metab. Dispos.*, *31* (3), 289–293.

Berthiaume, J.M., Wallace, K.B. (2007). Adriamycin-induced oxidative mitochondrial cardiotoxicity. *Cell Biol. Toxicol.*, *23* (1), 15–25. Review.

Bertini, R., Bianchi, M., Villa, P., Ghezzi, P. (1988). Depression of liver drug metabolism and increase in plasma fibrinogen by interleukin 1 and tumor necrosis factor: a comparison with lymphotoxin and interferon. *Int. J. Immunopharmacol.*, *10* (5), 525–530.

Bhaiya, P., Roychowdhury, S., Vyas, P.M., Doll, M.A., Hein, D.W., Svensson, C.K. (2006). Bioactivation, protein haptenation and toxicity of sulfamethoxazole and dapsone in normal human dermal fibroblasts. *Toxicol. Appl. Pharmacol.*, *215* (2), 158–167.

Blanco, M., Martínez, A., Urios, A., Herrera, G., O'Connor, J.E. (1998). Detection of oxidative mutagenesis by isoniazid and other hydrazine derivatives in Escherichia coli WP2 tester strain IC203, deficient in OxyR: strong protective effects of rat liver S9. *Mutat. Res.*, *417* (1), 39–46.

Blazka, M.E., Elwell, M.R., Holladay, S.D., Wilson, R.E., Luster, M. (1996). Histopathology of acetaminophen-induced liver changes: role of interleukin 1 alpha and tumor necrosis factor alpha. *Toxicol. Pathol.*, *24* (2), 181–189.

Boelsterli, U.A., Ho, H.K., Zhou, S., Leow, K.Y. (2006). Bioactivation and hepatotoxicity of nitroaromatic drugs. *Curr. Drug Metab.*, *7* (7), 715–727.

Bolchoz, L.J., Budinsky, R.A., McMillan, D.C., Jollow, D.J. (2001). Primaquine-induced hemolytic anemia: formation and hemotoxicity of the arylhydroxylamine metabolite 6-methoxy-8-hydroxylaminoquinoline. *J. Pharmacol. Exp. Ther.*, *297* (2), 509–515.

Bolt, H.M., Thier, R. (2006). Relevance of the deletion polymorphisms of the glutathione S-transferases GSTT1 and GSTM1 in pharmacology and toxicology. *Curr. Drug Metab.*, *7* (6), 613–628. Review.

Bomhard, E.M., Herbold, B.A. (2005). Genotoxic activities of aniline and its metabolites and their relationship to the carcinogenicity of aniline in the spleen of rats. *Crit. Rev. Toxicol.*, *35* (10), 783–835.

Boocock, D.J., Maggs, J.L., Brown, K., White, I.N., Park, B.K. (2000). Major inter-species differences in the rates of O-sulphonation and O-glucuronylation of alpha-hydroxytamoxifen *in vitro*: a metabolic disparity protecting human liver from the formation of tamoxifen-DNA adducts. *Carcinogenesis*, *21* (10), 1851–1858.

Boocock, D.J., Maggs, J.L., White, I.N., Park, B.K. (1999). Alpha-hydroxytamoxifen, a genotoxic metabolite of tamoxifen in the rat: identification and quantification *in vivo* and *in vitro*. *Carcinogenesis*, *20* (1), 153–160.

Bourdi, M., Gautier, J.C., Mircheva, J., Larrey, D., Guillouzo, A., Andre, C., Belloc, C., Beaune, P.H. (1992). Anti-liver microsomes autoantibodies and dihydralazine-induced hepatitis: specificity of autoantibodies and inductive capacity of the drug. *Mol. Pharmacol.*, *42* (2), 280–285.

Bourdi, M., Chen, W., Peter, R.M., Martin, J.L., Buters, J.T., Nelson, S.D., Pohl, L.R. (1996). Human cytochrome P450 2E1 is a major autoantigen associated with halothane hepatitis. *Chem. Res. Toxicol.*, *9* (7), 1159–1166.

Bourdi, M., Reilly, T.P., Elkahloun, A.G., George, J.W., Pohl, L.R. (2002). Macrophage migration inhibitory factor in drug-induced liver injury: a role in susceptibility and stress responsiveness. *Biochem. Biophys. Res. Commun.*, *294* (2), 225–230.

Butcher, N.J., Boukouvala, S., Sim, E., Minchin, R.F. (2002). Pharmacogenetics of the arylamine N-acetyltransferases. *Pharmacogenomics J.*, *2* (1), 30–42. Review.

Caldwell, G.W., Yan, Z. (2006). Screening for reactive intermediates and toxicity assessment in drug discovery. *Curr. Opin. Drug Discov. Devel.*, *9* (1), 47–60. Review.

Cashman, J.R. (2000). Human flavin-containing monooxygenase: substrate specificity and role in drug metabolism. *Curr. Drug Metab.*, *1*, 181–191.

Cashman, J.R., Bi, Y.A., Lin, J., Youil, R., Knight, M., Forrest, S., Treacy, E. (1997). Human flavin-containing monooxygenase form 3: cDNA expression of the enzymes containing amino acid substitutions observed in individuals with trimethylaminuria. *Chem. Res. Toxicol.*, *10* (8), 837–841.

Castillo, M., Smith, P.C. (1995). Disposition and reactivity of ibuprofen and ibufenac acyl glucuronides *in vivo* in the rhesus monkey and *in vitro* with human serum albumin. *Drug Metab. Dispos.*, *23* (5), 566–572.

Cavalieri, E.L., Rogan, E.G., Chakravarti, D. (2002). Initiation of cancer and other diseases by catechol ortho-quinones: a unifying mechanism. *Cell Mol. Life Sci.*, *59* (4), 665–681. Review.

Chae, K., Lindzey, J., McLachlan, J.A., Korach, K.S. (1998). Estrogen-dependent gene regulation by an oxidative metabolite of diethylstilbestrol, diethylstilbestrol-4′,4″-quinone. *Steroids*, *63* (3), 149–157.

Chan, K., Han, X.D., Kan, Y.W. (2001). An important function of Nrf2 in combating oxidative stress: detoxification of acetaminophen. *Proc. Natl. Acad. Sci. U.S.A.*, *98* (8), 4611–4616.

Chang, T.K., Waxman, D.J. (1993). Cyclophosphamide modulates rat hepatic cytochrome P450 2C11 and steroid 5 alpha-reductase activity and messenger RNA levels through the combined action of acrolein and phosphoramide mustard. *Cancer Res.*, *53* (11), 2490–2497. Erratum in: *Cancer Res* (1993) *53* (16):3846.

Chen, W., Koenigs, L.L., Thompson, S.J., Peter, R.M., Rettie, A.E., Trager, W.F., Nelson, S.D. (1998). Oxidation of acetaminophen to its toxic quinoneimine and nontoxic catechol metabolites by baculovirus-expressed and purified human cytochromes P450 2E1 and 2A6. *Chem. Res. Toxicol.*, *11* (4), 295–301.

Chitturi, S., George, J. (2002). Hepatotoxicity of commonly used drugs: nonsteroidal anti-inflammatory drugs, antihypertensives, antidiabetic agents, anticonvulsants, lipid-lowering agents, psychotropic drugs. *Semin. Liver Dis.*, *22* (2), 169–183. Review.

Clarke, L., Waxman, D.J. (1989). Oxidative metabolism of cyclophosphamide: identification of the hepatic monooxygenase catalysts of drug activation. *Cancer Res.*, *49* (9), 2344–2350.

Colacurci, N., De Seta, L., De Franciscis, P., Mele, D., Fortunato, N., Cassese, S. (2000). Tamoxifen effects on endometrium. *Panminerva Med.*, *42*, 45–47.

Colby, H.D., O'Donnell, J.P., Flowers, N.L., Kossor, D.C., Johnson, P.B., Levitt, M. (1991). Relationship between covalent binding to microsomal protein and the destruction of adrenal cytochrome P-450 by spironolactone. *Toxicology*, *67* (2), 143–154.

Coleman, T., Ellis, S.W., Martin, I.J., Lennard, M.S., Tucker, G.T. (1996). 1-Methyl-4-phenyl-1,2,3,6-tetrahydropyridine (MPTP) is N-demethylated by cytochromes P450 2D6, 1A2 and 3A4—implications for susceptibility to Parkinson's disease. *J. Pharmacol. Exp. Ther.*, *277* (2), 685–690.

Coleman, M.D., Khalaf, L.F., Nicholls, P.J. (2003). Aminoglutethimide-induced leucopenia in a mouse model: effects of metabolic and structural determinates. *Environ. Toxicol. Pharmacol.*, *15*, 27–32.

Copple, I.M., Goldring, C.E., Kitteringham, N.R., Park, B.K. (2008). The Nrf2-Keap1 defence pathway: role in protection against drug-induced toxicity. *Toxicology*, *246* (1), 24–33.

Coughtrie, M.W. (2002). Sulfation through the looking glass—recent advances in sulfotransferase research for the curious. *Pharmacogenomics J.*, *2* (5), 297–308. Review.

Court, M.H. (2005). Isoform-selective probe substrates for *in vitro* studies of human UDP-glucur onosyltransferases. *Methods Enzymol.*, *400*, 104–116.

Cribb, A.E., Lee, B.L., Trepanier, L.A., Spielberg, S.P. (1996). Adverse reactions to sulphonamide and sulphonamide-trimethoprim antimicrobials: clinical syndromes and pathogenesis. *Adverse Drug React. Toxicol. Rev.*, *15* (1), 9–50.

Dalrymple, P.D., Nicholls, P.J. (1988). Metabolism profiles and excretion of 14C-aminoglutethimide in several animal species and man. *Xenobiotica*, *18* (1), 75–81.

Davis, D.C., Potter, W.Z., Jollow, D.J., Mitchell, J.R. (1974). Species differences in hepatic glutathione depletion, covalent binding and hepatic necrosis after acetaminophen. *Life Sci.*, *14* (11), 2099–2109.

Davis, G.C., Williams, A.C., Markey, S.P., Ebert, M.H., Caine, E.D., Reichert, C.M., Kopin, I.J. (1979). Chronic Parkinsonism secondary to intravenous injection of meperidine analogues. *Psychiatry Res.*, *1* (3), 249–254.

De, S.K., De, M., Banerjee, R.K. (1986). Localization and origin of the intestinal peroxidase—effect of adrenal glucocorticoids. *J. Steroid Biochem.*, *24* (2), 629–635.

Debiec-Rychter, M., Land, S.J., King, C.M. (1999). Histological localization of acetyltransferases in human tissue. *Cancer Lett.*, *143* (2), 99–102.

Decker, C.J., Doerge, D.R. (1992). Covalent binding of 14C- and 35S-labeled thiocarbamides in rat hepatic microsomes. *Biochem. Pharmacol.*, *43* (4), 881–888.

Desmeules, P., Devine, P.J. (2006). Characterizing the ovotoxicity of cyclophosphamide metabolites on cultured mouse ovaries. *Toxicol. Sci.*, *90* (2), 500–509.

Dhaini, H.R., Levy, GN. (2000). Arylamine N-acetyltransferase 1 (NAT1) genotypes in a Lebanese population. *Pharmacogenetics*, *10* (1), 79–83.

Dietze, E.C., Schäfer, A., Omichinski, J.G., Nelson, S.D. (1997). Inactivation of glyceraldehyde-3-phosphate dehydrogenase by a reactive metabolite of acetaminophen and mass spectral characterization of an arylated active site peptide. *Chem. Res. Toxicol.*, *10* (10), 1097–1103.

Doroshow, J.H., Locker, G.Y., Myers, C.E. (1980). Enzymatic defenses of the mouse heart against reactive oxygen metabolites: alterations produced by doxorubicin. *J. Clin. Invest.*, *65* (1), 128–135.

Dragan, Y.P., Fahey, S., Street, K., Vaughan, J., Jordan, V.C., Pitot, H.C. (1994). Studies of tamoxifen as a promoter of hepatocarcinogenesis in female Fischer F344 rats. *Breast. Cancer Res. Treat*, *31* (1), 11–25.

Eklund, B.I., Moberg, M., Bergquist, J., Mannervik, B. (2006). Divergent activities of human glutathione transferases in the bioactivation of azathioprine. *Mol. Pharmacol.*, *70* (2), 747–754.

Eklund, B.I., Gunnarsdottir, S., Elfarra, A.A., Mannervik, B. (2007). Human glutathione transferases catalyzing the bioactivation of anticancer thiopurine prodrugs. *Biochem. Pharmacol.*, *73* (11), 1829–1841.

Evans, D.A. (1989). N-acetyltransferase. *Pharmacol. Ther.*, *42* (2), 157–234.

Evans, D.C., Watt, A.P., Nicoll-Griffith, D.A., Baillie, T.A. (2004). Drug-protein adducts: an industry perspective on minimizing the potential for drug bioactivation in drug discovery and development. *Chem. Res. Toxicol.*, *17* (1), 3–16. Erratum in: *Chem Res Toxicol.* (2005) *18* (11):1777.

Eyles, D.W., Avent, K.M., Stedman, T.J., Pond, S.M. (1997). Two pyridinium metabolites of haloperidol are present in the brain of patients at post-mortem. *Life Sci.*, *60* (8), 529–534.

Falany, C.N., Kerl, E.A. (1990). Sulfation of minoxidil by human liver phenol sulfotransferase. *Biochem. Pharmacol.*, *40* (5), 1027–1032.

Fan, L., Schut, H.A., Snyderwine, E.G. (1995). Cytotoxicity DNA adduct formation and DNA repair induced by 2-hydroxyamino-3-methylimidazo[4,5-f]quinoline and 2-hydroxyamino-1-methyl-6-phenylimidazo[4,5-b]pyridine in cultured human mammary epithelial cells. *Carcinogenesis*, *16* (4), 775–779.

Fan, P.W., Bolton, J.L. (2001). Bioactivation of tamoxifen to metabolite E quinone methide: reaction with glutathione and DNA. *Drug Metab. Dispos.*, *29* (6), 891–896.

Fau, D., Berson, A., Eugene, D., Fromenty, B., Fisch, C., Pessayre, D. (1992). Mechanism for the hepatotoxicity of the antiandrogen, nilutamide. Evidence suggesting that redox cycling of this nitroaromatic drug leads to oxidative stress in isolated hepatocytes. *J. Pharmacol. Exp. Ther.*, *263* (1), 69–77.

Feyler, A., Voho, A., Bouchardy, C., Kuokkanen, K., Dayer, P., Hirvonen, A., Benhamou, S. (2002). Point myeloperoxidase $^{-463}$G → a polymorphism and lung cancer risk. *Cancer Epidemiol. Biomarkers Prev.*, *11* (12), 1550–1554.

Finnell, R.H., Bennett, G.D., Slattery, J.T., Amore, B.M., Bajpai, M., Levy, R.H. (1995). Effect of treatment with phenobarbital and stiripentol on carbamazepine-induced teratogenicity and reactive metabolite formation. *Teratology*, *52* (6), 324–332.

Fontana, R.J., McCashland, T.M., Benner, K.G., Appelman, H.D., Gunartanam, N.T., Wisecarver, J.L., Rabkin, J.M., Lee, W.M. (1999). Acute liver failure associated with prolonged use of bromfenac leading to liver transplantation. The Acute Liver Failure Study Group. *Liver Transpl. Surg.*, *5* (6), 480–484.

Fratelli, M., Goodwin, L.O., Ørom, U.A., Lombardi, S., Tonelli, R., Mengozzi, M., Ghezzi, P. (2005). Gene expression profiling reveals a signaling role of glutathione in redox regulation. *Proc. Natl. Acad. Sci. U.S.A.*, *102* (39), 13998–14003.

Freeman, R.W., Uetrecht, J.P., Woosley, R.L., Oates, J.A., Harbison, R.D. (1981). Covalent binding of procainamide *in vitro* and *in vivo* to hepatic protein in mice. *Drug Metab. Dispos.*, *9* (3), 188–192.

French, L.E. (2006). Toxic epidermal necrolysis and Stevens-Johnson syndrome: our current understanding. *Allergol. Int.*, *55* (1), 9–16.

French, L.E., Trent, J.T., Kerdel, F.A. (2006). Use of intravenous immunoglobulin in toxic epidermal necrolysis and Stevens-Johnson syndrome: our current understanding. *Int. Immunopharmacol.*, *6* (4), 543–549. Review.

Frye, R.F., Tammara, B., Cowart, T.D., Bramer, S.L. (1999). Effect of disulfiram-mediated CYP2E1 inhibition on the disposition of vesnarinone. *J. Clin. Pharmacol.*, *39* (11), 1177–1183.

Fu, P.P., Xia, Q., Lin, G., Chou, M.W. (2004). Pyrrolizidine alkaloids—genotoxicity, metabolism enzymes, metabolic activation and mechanisms. *Drug Metab. Rev.*, *36* (1), 1–55. Review.

Furst, S.M., Uetrecht, J.P. (1995). The effect of carbamazepine and its reactive metabolite, 9-acridine carboxaldehyde, on immune cell function *in vitro*. *Int. J. Immunopharmacol.*, *17* (5), 445–452.

Galinsky, R.E., Levy, G. (1981). Dose- and time-dependent elimination of acetaminophen in rats: pharmacokinetic implications of cosubstrate depletion. *J. Pharmacol. Exp. Ther.*, *219* (1), 14–20.

Ganey, P.E., Bailie, M.B., VanCise, S., Colligan, M.E., Madhukar, B.V., Robinson, J.P., Roth, R.A. (1994). Activated neutrophils from rat injured isolated hepatocytes. *Lab. Invest.*, *70* (1), 53–60.

Ganey, P.E., Roth, R.A. (2001). Concurrent inflammation as a determinant of susceptibility to toxicity from xenobiotic agents. *Toxicology*, *169* (3), 195–208. Review.

Gardner, I., Leeder, J.S., Chin, T., Zahid, N., Uetrecht, J.P. (1998). A comparison of the covalent binding of clozapine and olanzapine to human neutrophils *in vitro* and *in vivo*. *Mol. Pharmacol.*, *53* (6), 999–1008.

Garibaldi, R.A., Drusin, R.E., Ferebee, S.H., Gregg, MB. (1972). Isoniazid-associated hepatitis. Report of an outbreak. *Am. Rev. Respir. Dis.*, *106* (3), 357–365.

Garlick, R.L., Mazer, J.S. (1983). The principal site of nonenzymatic glycosylation of human serum albumin *in vivo*. *J. Biol. Chem.*, *258* (10), 6142–6146.

Gelotte, C.K., Auiler, J.F., Lynch, J.M., Temple, A.R., Slattery, J.T. (2007). Disposition of acetaminophen at 4, 6 and 8 g/day for 3 days in healthy young adults. *Clin. Pharmacol. Ther.*, *81* (6), 840–848.

Gill, J.S., Windebank, A.J. (1998). Cisplatin-induced apoptosis in rat dorsal root ganglion neurons is associated with attempted entry into the cell cycle. *J. Clin. Invest.*, *101* (12), 2842–2850.

Gitlin, N., Julie, N.L., Spurr, C.L., Lim, K.N., Juarbe, H.M. (1998). Two cases of severe clinical and histologic hepatotoxicity associated with troglitazone. *Ann. Intern. Med.*, *129* (1), 36–38.

Glatt, H. (1997). Sulfation and sulfotransferases 4: bioactivation of mutagens via sulfation. *FASEB. J.*, *11* (5), 314–321. Review.

Glatt, H. (2000). Sulfotransferases in the bioactivation of xenobiotics. *Chem. Biol. Interact.*, *129* (1–2), 141–170. Review.

Glatt, H., Bartsch, I., Christoph, S., Coughtrie, M.W., Falany, C.N., Hagen, M., Landsiedel, R., Pabel, U., Phillips, D.H., Seidel, A., Yamazoe, Y. (1998a). Sulfotransferase-mediated activation of mutagens studied using heterologous expression systems. *Chem. Biol. Interact.*, *109* (1–3), 195–219.

Glatt, H., Davis, W., Meinl, W., Hermersdörfer, H., Venitt, S., Phillips, D.H. (1998b). Rat, but not human, sulfotransferase activates a tamoxifen metabolite to produce DNA adducts and gene mutations in bacteria and mammalian cells in culture. *Carcinogenesis*, *19* (10), 1709–1713.

Goldin, R.D., Ratnayaka, I.D., Breach, C.S., Brown, I.N., Wickramasinghe, S.N. (1996). Role of macrophages in acetaminophen (paracetamol)-induced hepatotoxicity. *J. Pathol.*, *179* (4), 432–435.

Goldring, C.E., Kitteringham, N.R., Elsby, R., Randle, L.E., Clement, Y.N., Williams, D.P., McMahon, M., Hayes, J.D., Itoh, K., Yamamoto, M., Park, B.K. (2004). Activation of hepatic Nrf2 *in vivo* by acetaminophen in CD-1 mice. *Hepatology*, *39* (5), 1267–1276.

Gram, T.E. (1997). Chemically reactive intermediates and pulmonary xenobiotic toxicity. *Pharmacol. Rev.*, *49* (4), 297–341.

Grossman, S.J., Jollow, D.J. (1988). Role of dapsone hydroxylamine in dapsone-induced hemolytic anemia. *J. Pharmacol. Exp. Ther.*, *244* (1), 118–125.

Guan, X., Davis, M.R., Tang, C., Jochheim, C.M., Jin, L., Baillie, T.A. (1999). Identification of S-(n-butylcarbamoyl)glutathione, a reactive carbamoylating metabolite of tolbutamide in the rat and evaluation of its inhibitory effects on glutathione reductase *in vitro*. *Chem. Res. Toxicol.*, *12* (12), 1138–1143.

Guengerich, F.P. (1990). Mechanism-based inactivation of human liver microsomal cytochrome P-450 IIIA4 by gestodene. *Chem. Res. Toxicol.*, *3* (4), 363–371.

Gustafson, D.L., Swanson, J.D., Pritsos, C.A. (1991). Role of xanthine oxidase in the potentiation of doxorubicin-induced cardiotoxicity by mitomycin C. *Cancer Commun.*, *3* (9), 299–304.

Hakkola, J., Pelkonen, O., Pasanen, M., Raunio, H. (1998). Xenobiotic-metabolizing cytochrome P450 enzymes in the human feto-placental unit: role in intrauterine toxicity. *Crit. Rev. Toxicol.*, *28* (1), 35–72. Review.

Hales, B.F., Brown, H. (1991). The effect of *in vivo* glutathione depletion with buthionine sulfoximine on rat embryo development. *Teratology*, *44* (3), 251–257.

Halsey, J.P., Cardoe, N. (1982). Benoxaprofen side-effect profile in 300 patients. *Br. Med. J. (Clin. Res. Ed.)*, *284* (6326), 1365–1368.

Hamada, S., Kamada, M., Furumoto, H., Hirao, T., Aono, T. (1994). Expression of glutathione S-transferase-pi in human ovarian cancer as an indicator of resistance to chemotherapy. *Gynecol. Oncol.*, *52* (3), 313–319.

Hargus, S.J., Amouzedeh, H.R., Pumford, N.R., Myers, T.G., McCoy, S.C., Pohl, L.R. (1994). Metabolic activation and immunochemical localization of liver protein adducts of the nonsteroidal anti-inflammatory drug diclofenac. *Chem. Res. Toxicol.*, *7* (4), 575–582.

Hari, C.K., Raza, S.A., Clayton, M.I. (1998). Hydralazine-induced lupus and vocal fold paralysis. *J. Laryngol. Otol.*, *112* (9), 875–877.

Harlan, J.M., Levine, J.D., Callahan, K.S., Schwartz, B.R., Harker, L.A. (1984). Glutathione redox cycle protects cultured endothelial cells against lysis by extracellularly generated hydrogen peroxide. *J. Clin. Invest.*, *73* (3), 706–713.

Hart, S.G., Beierschmitt, W.P., Wyand, D.S., Khairallah, E.A., Cohen, S.D. (1994). Acetaminophen nephrotoxicity in CD-1 mice. I. Evidence of a role for in situ activation in selective covalent binding and toxicity. *Toxicol. Appl. Pharmacol.*, *126* (2), 267–275.

Hayes, J.D., Flanagan, J.U., Jowsey, I.R. (2005). Glutathione transferases. *Annu. Rev. Pharmacol. Toxicol.*, *45*, 51–88. Review.

Hayes, J.D., Pulford, D.J. (1995). The glutathione S-transferase supergene family: regulation of GST and the contribution of the isoenzymes to cancer chemoprotection and drug resistance. *Crit. Rev. Biochem. Mol. Biol.*, *30* (6), 445–600. Review.

He, K., Talaat, R.E., Pool, W.F., Reily, M.D., Reed, J.E., Bridges, A.J., Woolf, T.F. (2004). Metabolic activation of troglitazone: identification of a reactive metabolite and mechanisms involved. *Drug Metab. Dispos.*, *32* (6), 639–646.

Hearse, D.J., Weber, WW. (1973). Multiple N-acetyltransferases and drug metabolism. Tissue distribution, characterization and significance of mammalian N-acetyltransferase. *Biochem. J.*, *132* (3), 519–526.

Henderson, M.C., Krueger, S.K., Stevens, J.F., Williams, D.E. (2004). Human flavin-containing monooxygenase form 2 S-oxygenation: sulfenic acid formation from thioureas and oxidation of glutathione. *Chem. Res. Toxicol.*, *17*, 633–640.

Herbst, A.L., Ulfelder, H., Poskanzer, D.C. (1971). Adenocarcinoma of the vagina. Association of maternal stilbestrol therapy with tumor appearance in young women. *N. Engl. J. Med.*, *284* (15), 878–881.

Hildebrandt, M.A., Carrington, D.P., Thomae, B.A., Eckloff, B.W., Schaid, D.J., Yee, V.C., Weinshilboum, R.M., Wieben, E.D. (2007). Genetic diversity and function in the human cytosolic sulfotransferases. *Pharmacogenomics J.*, *7* (2), 133–143.

Hiraku, Y., Sekine, A., Nabeshi, H., Midorikawa, K., Murata, M., Kumagai, Y., Kawanishi, S. (2004). Mechanism of carcinogenesis induced by a veterinary antimicrobial drug, nitrofurazone, via oxidative DNA damage and cell proliferation. *Cancer Lett.*, *215* (2), 141–150.

Hobbs, D.C., Twomey, T.M. (1977). Metabolism of sudoxicam by the rat, dog and monkey. *Drug Metab. Dispos.*, *5* (1), 75–81.

Hofstra, A.H., Matassa, L.C., Uetrecht, J.P. (1991). Metabolism of hydralazine by activated leukocytes: implications for hydralazine induced lupus. *J. Rheumatol.*, *18* (11), 1673–1680.

Hort, J.F. (1975). Adverse reactions to alclofenac. *Curr. Med. Res. Opin.*, *3* (5), 333–337.

Huttner, W.B. (1988). Tyrosine sulfation and the secretory pathway. *Annu. Rev. Physiol.*, *50*, 363–376. Review.

IARC Working Group (1990). Nitrofural (nitrofurazone), in *IARC Monographs on the Evaluation of the Carcinogenic Risks of Chemicals to Humans*, *50*, IARC, Lyon, pp. 195–209.

van Iersel, M.L., Ploemen, J.P., Lo Bello, M., Federici, G., van Bladeren, P.J. (1997). Interactions of alpha, beta-unsaturated aldehydes and ketones with human glutathione S-transferase P1-1. *Chem. Biol. Interact.*, *108* (1-2), 67–78.

Innocenti, F., Iyer, L., Ratain, M.J. (2001). Pharmacogenetics of anticancer agents: lessons from amonafide and irinotecan. *Drug Metab. Dispos.*, *29* (4 Pt 2), 596–600.

Itagaki, S., Sumi, Y., Shimamoto, S., Itoh, T., Hirano, T., Takemoto, I., Iseki, K. (2005). Secretory transport of irinotecan metabolite SN-38 across isolated intestinal tissue. *Cancer Chemother. Pharmacol.*, *55* (5), 502–506.

Iverson, S.L., Uetrecht, J.P. (2001). Identification of a reactive metabolite of terbinafine: insights into terbinafine-induced hepatotoxicity. *Chem. Res. Toxicol.*, *14* (2), 175–181.

Iverson, S., Zahid, N., Uetrecht, J.P. (2002). Predicting drug-induced agranulocytosis: characterizing neutrophil-generated metabolites of a model compound, DMP 406 and assessing the relevance of an *in vitro* apoptosis assay for identifying drugs that may cause agranulocytosis. *Chem. Biol. Interact.*, *142* (1–2), 175–199.

Jaffe, I.A. (1986). Adverse effects profile of sulfhydryl compounds in man. *Am. J. Med.*, *80* (3), 471–476.

Jaiswal, A.K. (2004). Nrf2 signaling in coordinated activation of antioxidant gene expression. *Free Radic. Biol. Med.*, *36* (10), 1199–1207. Review.

Jakobsson, P.J., Morgenstern, R., Mancini, J., Ford-Hutchinson, A., Persson, B. (1999). Common structural features of MAPEG—a widespread superfamily of membrane associated proteins with highly divergent functions in eicosanoid and glutathione metabolism. *Protein Sci.*, *8* (3), 689–692.

Jeng, W., Ramkissoon, A., Parman, T., Wells, P.G. (2006). Prostaglandin H synthase-catalyzed bioactivation of amphetamines to free radical intermediates that cause CNS regional DNA oxidation and nerve terminal degeneration. *FASEB J.*, *20* (6), 638–650.

Jensen, C.B., Jollow, D.J. (1991). The role of N-hydroxyphenetidine in phenacetin-induced hemolytic anemia. *Toxicol. Appl. Pharmacol.*, *111* (1), 1–12.

Jones, D.R., Gorski, J.C., Hamman, M.A., Mayhew, B.S., Rider, S., Hall, S.D. (1999). Diltiazem inhibition of cytochrome P-450 3A activity is due to metabolite intermediate complex formation. *J. Pharmacol. Exp. Ther.*, *290* (3), 1116–1125.

Jones, B.E., Liu, H., Lo, C.R., Koop, D.R., Czaja, M.J. (2002). Cytochrome P450 2E1 expression induces hepatocyte resistance to cell death from oxidative stress. *Antioxid. Redox. Signal.*, *4* (5), 701–709.

Jönsson, K.H., Lindeke, B. (1992). Cytochrome P-455 nm complex formation in the metabolism of phenylalkylamines. XII. Enantioselectivity and temperature dependence in microsomes and reconstituted cytochrome P-450 systems from rat liver. *Chirality*, *4* (8), 469–477.

Joshi, G., Sultana, R., Tangpong, J., Cole, M.P., St Clair, D.K., Vore, M., Estus, S., Butterfield, D.A. (2005). Free radical mediated oxidative stress and toxic side effects in brain induced by the anti cancer drug adriamycin: insight into chemobrain. *Free Radic. Res.*, *39* (11), 1147–1154.

Ju, C., Uetrecht, J.P. (1998). Oxidation of a metabolite of indomethacin (Desmethyldeschlorobenzoylindomethacin) to reactive intermediates by activated neutrophils, hypochlorous acid and the myeloperoxidase system. *Drug Metab. Dispos.*, *26* (7), 676–680.

Kajita, J., Inano, K., Fuse, E., Kuwabara, T., Kobayashi, H. (2002). Effects of olopatadine, a new antiallergic agent, on human liver microsomal cytochrome P450 activities. *Drug Metab. Dispos.*, *30* (12), 1504–1511.

Kalgutkar, A.S., Gardner, I., Obach, R.S., Shaffer, C.L., Callegari, E., Henne, K.R., Mutlib, A.E., Dalvie, D.K., Lee, J.S., Nakai, Y., O'Donnell, J.P., Boer, J., Harriman, S.P. (2005). A comprehensive listing of bioactivation pathways of organic functional groups. *Curr. Drug Metab.*, *6* (3), 161–225.

Kalgutkar, A.S., Obach, R.S., Maurer, T.S. (2007). Mechanism-based inactivation of cytochrome P450 enzymes: chemical mechanisms, structure-activity relationships and relationship to clinical drug-drug interactions and idiosyncratic adverse drug reactions. *Curr. Drug Metab.*, *8* (5), 407–447.

Kalgutkar, A.S., Soglia, J.R. (2005). Minimising the potential for metabolic activation in drug discovery. *Expert Opin. Drug Metab. Toxicol.*, *1* (1), 91–142.

Kassahun, K., Farrell, K., Abbott, F. (1991). Identification and characterization of the glutathione and N-acetylcysteine conjugates of (E)-2-propyl-2,4-pentadienoic acid, a toxic metabolite of valproic acid, in rats and humans. *Drug Metab. Dispos.*, *19* (2), 525–535.

Kassahun, K., McIntosh, I., Cui, D., Hreniuk, D., Merschman, S., Lasseter, K., Azrolan, N., Iwamoto, M., Wagner, J.A., Wenning, L.A. (2007). Metabolism and disposition in humans of raltegravir (MK-0518), an anti-AIDS drug targeting the human immunodeficiency virus 1 integrase enzyme. *Drug Metab. Dispos.*, *35* (9), 1657–1663.

Kassahun, K., Pearson, P.G., Tang, W., McIntosh, I., Leung, K., Elmore, C., Dean, D., Wang, R., Doss, G., Baillie, T.A. (2001). Studies on the metabolism of troglitazone to reactive intermediates *in vitro* and *in vivo*. Evidence for novel biotransformation pathways involving quinone methide formation and thiazolidinedione ring scission. *Chem. Res. Toxicol.*, *14* (1), 62–70.

Kato, A., Yoshidome, H., Edwards, M.J., Lentsch, A.B. (2000). Regulation of liver inflammatory injury by signal transducer and activator of transcription-6. *Am. J. Pathol.*, *157* (1), 297–302.

Kaufman, D.W., Kelly, J.P., Anderson, T., Harmon, D.C., Shapiro, S. (1997). Evaluation of case reports of aplastic anemia among patients treated with felbamate. *Epilepsia*, *38* (12), 1265–1269. Review.

Kawai, K., Kawasaki-Tokui, Y., Odaka, T., Tsuruta, F., Kazui, M., Iwabuchi, H., Nakamura, T., Kinoshita, T., Ikeda, T., Yoshioka, T., Komai, T., Nakamura, K. (1997). Disposition and metabolism of the new oral antidiabetic drug troglitazone in rats, mice and dogs. *Arzneimittelforschung*, *47* (4), 356–368.

Kedderis, G.L., Argenbright, L.S., Miwa, G.T. (1989). Covalent interaction of 5-nitroimidazoles with DNA and protein *in vitro*: mechanism of reductive activation. *Chem. Res. Toxicol.*, *2* (3), 146–149.

Kempf, D.J., Marsh, K.C., Kumar, G., Rodrigues, A.D., Denissen, J.F., McDonald, E., Kukulka, M.J., Hsu, A., Granneman, G.R., Baroldi, P.A., Sun, E., Pizzuti, D., Plattner, J.J., Norbeck, D.W., Leonard, J.M. (1997). Pharmacokinetic enhancement of inhibitors of the human immunodeficiency virus protease by coadministration with ritonavir. *Antimicrob. Agents Chemother.*, *41* (3), 654–660.

Khan, F.D., Vyas, P.M., Gaspari, A.A., Svensson, C.K. (2007). Effect of arylhydroxylamine metabolites of sulfamethoxazole and dapsone on stress signal expression in human keratinocytes. *J. Pharmacol. Exp. Ther.*, *323* (3), 771–777.

Kharbanda, K.K., Rogers, D.D., II, Wyatt, T.A., Sorrell, M.F., Tuma, D.J. (2004). Transforming growth factor-beta induces contraction of activated hepatic stellate cells. *J. Hepatol.*, *41* (1), 60–66.

Kiang, T.K., Ho, P.C., Anari, M.R., Tong, V., Abbott, F.S., Chang, T.K. (2006). Contribution of CYP2C9, CYP2A6 and CYP2B6 to valproic acid metabolism in hepatic microsomes from individuals with the CYP2C9*1/*1 genotype. *Toxicol. Sci.*, *94* (2), 261–271.

Killian, J.G., Kerr, K., Lawrence, C., Celermajer, D.S. (1999). Myocarditis and cardiomyopathy associated with clozapine. *Lancet*, *354* (9193), 1841–1845.

Kim, P.M., Winn, L.M., Parman, T., Wells, P.G. (1997). UDP-glucuronosyltransferase-mediated protection against *in vitro* DNA oxidation and micronucleus formation initiated by phenytoin and its embryotoxic metabolite 5-(p-hydroxyphenyl)-5-phenylhydantoin. *J. Pharmacol. Exp. Ther.*, *280* (1), 200–209.

Klebanoff, S.J. (1999). Myeloperoxidase. *Proc. Assoc. Am. Physicians*, *111* (5), 383–389. Review.

Kodama, M., Inoue, F., Kaneko, M., Oda, T., Sato, Y. (1993). Formation of free radicals and active oxygen species from diethylstilbestrol and its derivatives. *Anticancer Res.*, *13* (4), 1209–1213.

Kopke, R.D., Liu, W., Gabaizadeh, R., Jacono, A., Feghali, J., Spray, D., Garcia, P., Steinman, H., Malgrange, B., Ruben, R.J., Rybak, L., Van de Water, T.R. (1997). Use of organotypic cultures of Corti's organ to study the protective effects of antioxidant molecules on cisplatin-induced damage of auditory hair cells. *Am. J. Otol.*, *18* (5), 559–571.

Korashy, H.M., Elbekai, R.H., El-Kadi, A.O. (2004). Effects of renal diseases on the regulation and expression of renal and hepatic drug-metabolizing enzymes: a review. *Xenobiotica*, *34* (1), 1–29.

Kossak, B.D., Schmidt-Sommerfeld, E., Schoeller, D.A., Rinaldo, P., Penn, D., Tonsgard, J.H. (1993). Impaired fatty acid oxidation in children on valproic acid and the effect of L-carnitine. *Neurology*, *43* (11), 2362–2368.

Kossor, D.C., Kominami, S., Takemori, S., Colby, H.D. (1991). Role of the steroid 17 alpha-hydroxylase in spironolactone-mediated destruction of adrenal cytochrome P-450. *Mol. Pharmacol.*, *40* (2), 321–325.

Kossor, D.C., Kominami, S., Takemori, S., Colby, H.D. (1992). Destruction of testicular cytochrome P-450 by 7 alpha-thiospironolactone is catalyzed by the 17 alpha-hydroxylase. *J. Steroid Biochem. Mol. Biol.*, *42* (3–4), 421–424.

Koudriakova, T., Iatsimirskaia, E., Utkin, I., Gangl, E., Vouros, P., Storozhuk, E., Orza, D., Marinina, J., Gerber, N. (1998). Metabolism of the human immunodeficiency virus protease inhibitors indinavir and ritonavir by human intestinal microsomes and expressed cytochrome P4503A4/3A5: mechanism-based inactivation of cytochrome P4503A by ritonavir. *Drug Metab. Dispos.*, *26* (6), 552–561.

Krause, R.J., Lash, L.H., Elfarra, A.A. (2003). Human kidney flavin-containing monooxygenases and their potential roles in cysteine s-conjugate metabolism and nephrotoxicity. *J. Pharmacol. Exp. Ther.*, *304*, 185–191.

Krueger, S.K., Williams, D.E. (2005). Mammalian flavin-containing monooxygenases: structure/function, genetic polymorphisms and role in drug metabolism. *Pharmacol. Ther.*, *106*, 357–387.

Lai, W.G., Gardner, I., Zahid, N., Uetrecht, J.P. (2000). Bioactivation and covalent binding of hydroxyfluperlapine in human neutrophils: implications for fluperlapine-induced agranulocytosis. *Drug Metab. Dispos.*, *28* (3), 255–263.

Lai, W.G., Zahid, N., Uetrecht, J.P. (1999). Metabolism of trimethoprim to a reactive iminoquinone methide by activated human neutrophils and hepatic microsomes. *J. Pharmacol. Exp. Ther.*, *291* (1), 292–299.

Lang, D.H., Rettie, A.E. (2000). *In vitro* evaluation of potential *in vivo* probes for human flavin-containing monooxygenase (FMO): metabolism of benzydamine and caffeine by FMO and P450 isoforms. *Br. J. Clin. Pharmacol.*, *50* (4), 311–314.

Lang, D.H., Yeung, C.K., Peter, R.M., Ibarra, C., Gasser, R., Itagaki, K., Philpot, R.M., Rettie, A.E. (1998). Isoform specificity of trimethylamine N-oxygenation by human flavin-containing monooxygenase (FMO) and P450 enzymes: selective catalysis by FMO3. *Biochem. Pharmacol.*, *56*, 1005–1012.

Langston, J.W., Ballard, P., Tetrud, J.W., Irwin, I. (1983). Chronic Parkinsonism in humans due to a product of meperidine-analog synthesis. *Science*, *219* (4587), 979–980.

Lautala, P., Ethell, B.T., Taskinen, J., Burchell, B. (2000). The specificity of glucuronidation of entacapone and tolcapone by recombinant human UDP-glucuronosyltransferases. *Drug Metab. Dispos.*, *28* (11), 1385–1389.

Lavergne, S.N., Danhof, R.S., Volkman, E.M., Trepanier, L.A. (2006). Association of drug-serum protein adducts and anti-drug antibodies in dogs with sulphonamide hypersensitivity: a naturally occurring model of idiosyncratic drug toxicity. *Clin. Exp. Allergy*, *36* (7), 907–915.

Lazo, J.S., Humphreys, C.J. (1983). Lack of metabolism as the biochemical basis of bleomycin-induced pulmonary toxicity. *Proc. Natl. Acad. Sci. U.S.A.*, *80* (10), 3064–3068.

Lecoeur, S., Bonierbale, E., Challine, D., Gautier, J.C., Valadon, P., Dansette, P.M., Catinot, R., Ballet, F., Mansuy, D., Beaune, P.H. (1994). Specificity of *in vitro* covalent binding of tienilic acid metabolites to human liver microsomes in relationship to the type of hepatotoxicity: comparison with two directly hepatotoxic drugs. *Chem. Res. Toxicol.*, *7* (3), 434–442.

Lee, W.M. (2003). Drug-induced hepatotoxicity. *N. Engl. J. Med.*, *349* (5), 474–485.

Lenz, W. (1966). Malformations caused by drugs in pregnancy. *Am. J. Dis. Child.*, *112* (2), 99–106.

Leone, A.M., Kao, L.M., McMillian, M.K., Nie, A.Y., Parker, J.B., Kelley, M.F., Usuki, E., Parkinson, A., Lord, P.G., Johnson, M.D. (2007). Evaluation of felbamate and other antiepileptic drug toxicity potential based on hepatic protein covalent binding and gene expression. *Chem. Res. Toxicol.*, *20* (4), 600–608.

Lessard, E., Hamelin, B.A., Labbé, L., O'Hara, G., Bélanger, P.M., Turgeon, J. (1999). Involvement of CYP2D6 activity in the N-oxidation of procainamide in man. *Pharmacogenetics*, *9* (6), 683–696.

Li, G., Frenz, D.A., Brahmblatt, S., Feghali, J.G., Ruben, R.J., Berggren, D., Arezzo, J., Van De Water, T.R. (2001). Round window membrane delivery of L-methionine provides protection from cisplatin ototoxicity without compromising chemotherapeutic efficacy. *Neurotoxicology*, *22* (2), 163–176.

Li, G., Liu, W., Frenz, D. (2006). Cisplatin ototoxicity to the rat inner ear: a role for HMG1 and iNOS. *Neurotoxicology*, *27* (1), 22–30.

Lin, L.Y., Di Stefano, E.W., Schmitz, D.A., Hsu, L., Ellis, S.W., Lennard, M.S., Tucker, G.T., Cho, A.K. (1997). Oxidation of methamphetamine and methylenedioxymethamphetamine by CYP2D6. *Drug Metab. Dispos.*, *25* (9), 1059–1064.

Lindberg, R.L., Syvälahti, E.K. (1986). Metabolism of nomifensine after oral and intravenous administration. *Clin. Pharmacol. Ther.*, *39* (4), 378–383.

Liu, G.Y., Frank, N., Bartsch, H., Lin, J.K. (1998). Induction of apoptosis by thiuramdisulfides, the reactive metabolites of dithiocarbamates, through coordinative modulation of NFkappaB, c-fos/c-jun and p53 proteins. *Mol. Carcinog.*, *22* (4), 235–246.

Liu, Z.C., Uetrecht, J.P. (2000). Metabolism of ticlopidine by activated neutrophils: implications for ticlopidine-induced agranulocytosis. *Drug Metab. Dispos.*, *28* (7), 726–730.

López-Neblina, F., Toledo-Pereyra, L.H. (2006). Phosphoregulation of signal transduction pathways in ischemia and reperfusion. *J. Surg. Res.*, *134* (2), 292–299.

Lu, W., Uetrecht, J.P. (2007). Possible bioactivation pathways of lamotrigine. *Drug Metab. Dispos.*, *35* (7), 1050–1056.

Madden, S., Woolf, T.F., Pool, W.F., Park, B.K. (1993). An investigation into the formation of stable, protein-reactive and cytotoxic metabolites from tacrine in vitro. Studies with human and rat liver microsomes. *Biochem. Pharmacol.*, *46* (1), 13–20.

Maeda, T., Nakamura, R., Kadokami, K., Ogawa, H.I. (2007). Relationship between mutagenicity and reactivity or biodegradability for nitroaromatic compounds. *Environ. Toxicol. Chem.*, *26* (2), 237–241.

Maggs, J.L., Naisbitt, D.J., Tettey, J.N., Pirmohamed, M., Park, B.K. (2000). Metabolism of lamotrigine to a reactive arene oxide intermediate. *Chem. Res. Toxicol.*, *13* (11), 1075–1081.

Maggs, J.L., Tingle, M.D., Kitteringham, N.R., Park, B.K. (1988). Drug-protein conjugates–XIV. Mechanisms of formation of protein-arylating intermediates from amodiaquine, a myelotoxin and hepatotoxin in man. *Biochem. Pharmacol.*, *37* (2), 303–311.

Mallat, A., Zafrani, E.S., Metreau, J.M., Dhumeaux, D. (1997). Terbinafine-induced prolonged cholestasis with reduction of interlobular bile ducts. *Dig. Dis. Sci.*, *42* (7), 1486–1488.

Mallet, W.G., Mosebrook, D.R., Trush, M.A. (1991). Activation of (+-)-trans-7,8-dihydroxy-7,8-dihydrobenzo[a]pyrene to diolepoxides by human polymorphonuclear leukocytes or myeloperoxidase. *Carcinogenesis*, *12* (3), 521–524.

Mannervik, B., Board, P.G., Hayes, J.D., Listowsky, I., Pearson, W.R. (2005). Nomenclature for mammalian soluble glutathione transferases. *Methods Enzymol.*, *401*, 1–8.

Mannervik, B., Castro, V.M., Danielson, U.H., Tahir, M.K., Hansson, J., Ringborg, U. (1987). Expression of class Pi glutathione transferase in human malignant melanoma cells. *Carcinogenesis*, *8* (12), 1929–1932.

Martin, E.A., Rich, K.J., White, I.N., Woods, K.L., Powles, T.J., Smith, L.L. (1995). 32P-postlabelled DNA adducts in liver obtained from women treated with tamoxifen. *Carcinogenesis*, *16* (7), 1651–1654.

Martin, W.J., II (1983). Nitrofurantoin: evidence for the oxidant injury of lung parenchymal cells. *Am. Rev. Respir. Dis.*, *127* (4), 482–486.

Masiá, M., Padilla, S., Bernal, E., Almenar, M.V., Molina, J., Hernández, I., Graells, M.L., Gutiérrez, F. (2007). Influence of antiretroviral therapy on oxidative stress and cardiovascular risk: a prospective cross-sectional study in HIV-infected patients. *Clin. Ther.*, *29* (7), 1448–1455.

Masubuchi, Y. (2006). Metabolic and non-metabolic factors determining troglitazone hepatotoxicity: a review. *Drug Metab. Pharmacokinet.*, *21* (5), 347–356. Review.

Masubuchi, Y., Horie, T. (1999). Mechanism-based inactivation of cytochrome P450s 1A2 and 3A4 by dihydralazine in human liver microsomes. *Chem. Res. Toxicol.*, *12* (10), 1028–1032.

Masubuchi, Y., Konishi, M., Horie, T. (1999). Imipramine- and mianserin-induced acute cell injury in primary cultured rat hepatocytes: implication of different cytochrome P450 enzymes. *Arch. Toxicol.*, *73* (3), 147–151.

Masubuchi, Y., Nakano, T., Ose, A., Horie, T. (2001). Differential selectivity in carbamazepine-induced inactivation of cytochrome P450 enzymes in rat and human liver. *Arch. Toxicol.*, *75* (9), 538–543.

Masubuchi, Y., Ose, A., Horie, T. (2001). Mechanism-based inactivation of CYP2C11 by diclofenac. *Drug Metab. Dispos.*, *29* (9), 1190–1195.

Matzinger, P. (1994). Tolerance, danger and the extended family. *Annu. Rev. Immunol.*, *12*, 991–1045.

Mays, D.C., Pawluk, L.J., Apseloff, G., Davis, W.B., She, Z.W., Sagone, A.L., Gerber, N. (1995). Metabolism of phenytoin and covalent binding of reactive intermediates in activated human neutrophils. *Biochem. Pharmacol.*, *50* (3), 367–380.

Mazze, R.I., Trudell, J.R., Cousins, M.J. (1971). Methoxyflurane metabolism and renal dysfunction: clinical correlation in man. *Anesthesiology*, *35* (3), 247–252.

Melet, A., Assrir, N., Jean, P., Pilar Lopez-Garcia, M., Marques-Soares, C., Jaouen, M., Dansette, P.M., Sari, M.A., Mansuy, D. (2003). Substrate selectivity of human cytochrome P450 2C9: importance of residues 476, 365 and 114 in recognition of diclofenac and sulfaphenazole and in mechanism-based inactivation by tienilic acid. *Arch. Biochem. Biophys.*, *409* (1), 80–91.

Melnick, S., Cole, P., Anderson, D., Herbst, A. (1987). Rates and risks of diethylstilbestrol-related clear-cell adenocarcinoma of the vagina and cervix. An update. *N. Engl. J. Med.*, *316* (9), 514–516.

Menard, R.H., Guenthner, T.M., Kon, H., Gillette, J.R. (1979). Studies on the destruction of adrenal and testicular cytochrome P-450 by spironolactone. Requirement for the 7alpha-thio group and evidence for the loss of the heme and apoproteins of cytochrome P-450. *J. Biol. Chem.*, *254* (5), 1726–1733.

Menon, K.V.N., Angulo, P., Lindor, K.D. (2001). Severe cholestatic hepatitis from troglitazone in a patient with nonalcoholic steatohepatitis and diabetes mellitus. *Am. J. Gastroenterol.*, *96* (5), 1631–1634.

Meuldermans, W., Van Peer, A., Hendrickx, J., Lauwers, W., Swysen, E., Bockx, M., Woestenborghs, R., Heykants, J. (1988). Excretion and biotransformation of cisapride in dogs and humans after oral administration. *Drug Metab. Dispos.*, *16* (3), 403–409.

Miller, J.L., Trepanier, L.A. (2002). Inhibition by atovaquone of CYP2C9-mediated sulphamethoxazole hydroxylamine formation. *Eur. J. Clin. Pharmacol.*, *58* (1), 69–72.

Miners, J.O., Knights, K.M., Houston, J.B., Mackenzie, P.I. (2006). *In vitro-in vivo* correlation for drugs and other compounds eliminated by glucuronidation in humans: pitfalls and promises. *Biochem. Pharmacol.*, *71*, 1531–1539.

Miyamoto, G., Zahid, N., Uetrecht, J.P. (1997). Oxidation of diclofenac to reactive intermediates by neutrophils, myeloperoxidase and hypochlorous acid. *Chem. Res. Toxicol.*, *10* (4), 414–419.

Mizutani, T., Yoshida, K., Kawazoe, S. (1993). Possible role of thioformamide as a proximate toxicant in the nephrotoxicity of thiabendazole and related thiazoles in glutathione-depleted mice: structure-toxicity and metabolic studies. *Chem. Res. Toxicol.*, *6* (2), 174–179.

Mushiroda, T., Ariyoshi, N., Yokoi, T., Takahara, E., Nagata, O., Kato, H., Kamataki, T. (2001). Accumulation of the 1-methyl-4-phenylpyridinium ion in suncus (Suncus murinus) brain: implication for flavin-containing monooxygenase activity in brain microvessels. *Chem. Res. Toxicol.*, *14* (2), 228–232.

Mutlib, A.E., Gerson, R.J., Meunier, P.C., Haley, P.J., Chen, H., Gan, L.S., Davies, M.H., Gemzik, B., Christ, D.D., Krahn, D.F., Markwalder, J.A., Seitz, S.P., Robertson, R.T., Miwa, G.T. (2000). The species-dependent metabolism of efavirenz produces a nephrotoxic glutathione conjugate in rats. *Toxicol. Appl. Pharmacol.*, *169* (1), 102–113.

Mutlib, A.E., Shockcor, J., Chen, S.Y., Espina, R.J., Pinto, D.J., Orwat, M.J., Prakash, S.R., Gan, L.S. (2002). Disposition of 1-[3-(aminomethyl)phenyl]-N-[3-fluoro-2′- (methylsulfonyl)-[1,1′-biphenyl]-4-yl]-3-(trifluoromethyl)- 1H-pyrazole-5-carboxamide (DPC 423) by novel metabolic pathways. Characterization of unusual metabolites by liquid chromatography/mass spectrometry and NMR. *Chem. Res. Toxicol.*, *15* (1), 48–62.

Naisbitt, D.J., Williams, D.P., O'Neill, P.M., Maggs, J.L., Willock, D.J., Pirmohamed, M., Park, B.K. (1998). Metabolism-dependent neutrophil cytotoxicity of amodiaquine: a comparison with pyronaridine and related antimalarial drugs. *Chem. Res. Toxicol.*, *11* (12), 1586–1595.

Nakajima, T., Elovaara, E., Anttila, S., Hirvonen, A., Camus, A.M., Hayes, J.D., Ketterer, B., Vainio, H. (1995). Expression and polymorphism of glutathione S-transferase in human lungs: risk factors in smoking-related lung cancer. *Carcinogenesis*, *16* (4), 707–711.

O'Brien, P.J. (2000). Peroxidases. *Chem. Biol. Interact.*, *129* (1–2), 113–139. Review.

O'Donnell, J.P., Dalvie, D.K., Kalgutkar, A.S., Obach, R.S. (2003). Mechanism-based inactivation of human recombinant P450 2C9 by the nonsteroidal anti-inflammatory drug suprofen. *Drug Metab. Dispos.*, *31* (11), 1369–1377.

O'Neil, M.G., Perdun, C.S., Wilson, M.B., McGown, S.T., Patel, S. (1996). Felbamate-associated fatal acute hepatic necrosis. *Neurology*, *46* (5), 1457–1459.

Oh, E.J., Chae, H.J., Park, Y.J., Park, J.W., Han, K. (2007). Agranulocytosis, plasmacytosis and thrombocytosis due to methimazole-induced bone marrow toxicity. *Am. J. Hematol.*, *82* (6), 500.

Okawa, H., Motohashi, H., Kobayashi, A., Aburatani, H., Kensler, T.W., Yamamoto, M. (2006). Hepatocyte-specific deletion of the keap1 gene activates Nrf2 and confers potent resistance against acute drug toxicity. *Biochem. Biophys. Res. Commun.*, *339* (1), 79–88.

Olsen, J., Li, C., Bjørnsdottir, I., Sidenius, U., Hansen, S.H., Benet, L.Z. (2005). *In vitro* and *in vivo* studies on acyl-coenzyme A-dependent bioactivation of zomepirac in rats. *Chem. Res. Toxicol.*, *18* (11), 1729–1736.

Opii, W.O., Sultana, R., Abdul, H.M., Ansari, M.A., Nath, A., Butterfield, D.A. (2007). Oxidative stress and toxicity induced by the nucleoside reverse transcriptase inhibitor (NRTI)—2′,3′-dideoxycytidine (ddC): relevance to HIV-dementia. *Exp. Neurol.*, *204* (1), 29–38.

Ostapowicz, G., Fontana, R.J., Schiødt, F.V., Larson, A., Davern, T.J., Han, S.H., McCashland, T.M., Shakil, A.O., Hay, J.E., Hynan, L., Crippin, J.S., Blei, A.T., Samuel, G., Reisch, J., Lee, WM, U.S. Acute Liver Failure Study Group (2002). Results of a prospective study of acute liver failure at 17 tertiary care centers in the United States. *Ann. Intern. Med.*, *137* (12), 947–954. Summary for patients in: Ann Intern Med. 2002 Dec 17;137(12):I24.

Padovan, E., Bauer, T., Tongio, M.M., Kalbacher, H., Weltzien, H.U. (1997). Penicilloyl peptides are recognized as T cell antigenic determinants in penicillin allergy. *Eur. J. Immunol.*, *27* (6), 1303–1307.

Park, B.K., Pirmohamed, M., Kitteringham, N.R. (1995). The role of cytochrome P450 enzymes in hepatic and extrahepatic human drug toxicity. *Pharmacol. Ther.*, *68* (3), 385–424.

Park, S.A., Choi, K.S., Bang, J.H., Huh, K., Kim, S.U. (2000). Cisplatin-induced apoptotic cell death in mouse hybrid neurons is blocked by antioxidants through suppression of cisplatin-mediated accumulation of p53 but not of Fas/Fas ligand. *J. Neurochem.*, *75* (3), 946–953.

Park, B.K., Kitteringham, N.R., Maggs, J.L., Pirmohamed, M., Williams, D.P. (2005). The role of metabolic activation in drug-induced hepatotoxicity. *Annu. Rev. Pharmacol. Toxicol.*, *45*, 177–202. Review.

Parman, T., Chen, G., Wells, P.G. (1998). Free radical intermediates of phenytoin and related teratogens. Prostaglandin H synthase-catalyzed bioactivation, electron paramagnetic resonance spectrometry, and photochemical product analysis. *J. Biol. Chem.*, *273* (39), 25079–25088.

Parman, T., Wiley, M.J., Wells, P.G. (1999). Free radical-mediated oxidative DNA damage in the mechanism of thalidomide teratogenicity. *Nat. Med.*, *5* (5), 582–585.

Parra, J., Iriarte, J., Pierre-Louis, S.J. (1996). Valproate toxicity. *Neurology*, *47* (6), 1608.

Patten, C.J., Thomas, P.E., Guy, R.L., Lee, M., Gonzalez, F.J., Guengerich, F.P., Yang, C.S. (1993). Cytochrome P450 enzymes involved in acetaminophen activation by rat and human liver microsomes and their kinetics. *Chem. Res. Toxicol.*, *6* (4), 511–518.

Pearce, R.E., Uetrecht, J.P., Leeder, J.S. (2005). Pathways of carbamazepine bioactivation *in vitro*: II. The role of human cytochrome P450 enzymes in the formation of 2-hydroxyiminostilbene. *Drug Metab. Dispos.*, *33* (12), 1819–1826.

Pemble, S., Schroeder, K.R., Spencer, S.R., Meyer, D.J., Hallier, E., Bolt, H.M., Ketterer, B., Taylor, J.B. (1994). Human glutathione S-transferase theta (GSTT1): cDNA cloning and the characterization of a genetic polymorphism. *Biochem. J.*, *300* (Pt 1), 271–276.

Pemble, S.E., Wardle, A.F., Taylor, J.B. (1996). Glutathione S-transferase class Kappa: characterization by the cloning of rat mitochondrial GST and identification of a human homologue. *Biochem. J.*, *319* (Pt 3), 749–754.

Perez-Reyes, E., Kalyanaraman, B., Mason, R.P. (1980). The reductive metabolism of metronidazole and ronidazole by aerobic liver microsomes. *Mol. Pharmacol.*, *17* (2), 239–244.

Periti, P., Mazzei, T., Mini, E., Novelli, A. (1992). Pharmacokinetic drug interactions of macrolides. *Clin. Pharmacokinet.*, *23* (2), 106–131. Review. Erratum in: *Clin Pharmacokinet*, (1993) *24* (1):70.

Peters, J.H., Miller, K.S., Brown, P. (1965). Studies on the metabolic basis for the genetically determined capacities for isoniazid inactivation in man. *J. Pharmacol. Exp. Ther.*, *150* (2), 298–304.

Pichler, W.J., Beeler, A., Keller, M., Lerch, M., Posadas, S., Schmid, D., Spanou, Z., Zawodniak, A., Gerber, B. (2006). Pharmacological interaction of drugs with immune receptors: the p-i concept. *Allergol. Int.*, *55* (1), 17–25. Review.

Piepho, R.W. (2000). Overview of the angiotensin-converting-enzyme inhibitors. *Am. J. Health Syst. Pharm.*, *57*, S3–7.Suppl 1.

Pirmohamed, M., Williams, D., Madden, S., Templeton, E., Park, B.K. (1995). Metabolism and bioactivation of clozapine by human liver *in vitro*. *J. Pharmacol. Exp. Ther.*, *272* (3), 984–990.

Plowchalk, D.R., Mattison, D.R. (1991). Phosphoramide mustard is responsible for the ovarian toxicity of cyclophosphamide. *Toxicol. Appl. Pharmacol.*, *107* (3), 472–481.

Pollack, G.M., Brouwer, K.L. (1991). Physiologic and metabolic influences on enterohepatic recirculation: simulations based upon the disposition of valproic acid in the rat. *J. Pharmacokinet. Biopharm.*, *19* (2), 189–225.

Poon, G.K., Chen, Q., Teffera, Y., Ngui, J.S., Griffin, P.R., Braun, M.P., Doss, G.A., Freeden, C., Stearns, R.A., Evans, D.C., Baillie, T.A., Tang, W. (2001). Bioactivation of diclofenac via benzoquinone imine intermediates-identification of urinary mercapturic acid derivatives in rats and humans. *Drug Metab. Dispos.*, *29* (12), 1608–1613.

Popović, M., Nierkens, S., Pieters, R., Uetrecht, J. (2004). Investigating the role of 2-phenylpropenal in felbamate-induced idiosyncratic drug reactions. *Chem. Res. Toxicol.*, *17* (12), 1568–1576.

Ratra, G.S., Morgan, W.A., Mullervy, J., Powell, C.J., Wright, M.C. (1998). Methapyrilene hepatotoxicity is associated with oxidative stress, mitochondrial disfunction and is prevented by the Ca2+ channel blocker verapamil. *Toxicology*, *130* (2–3), 79–93.

Renton, K.W., Knickle, L.C. (1990). Regulation of hepatic cytochrome P-450 during infectious disease. *Can. J. Physiol. Pharmacol.*, *68* (6), 777–781. Review.

Ring, B.J., Wrighton, S.A., Aldridge, S.L., Hansen, K., Haehner, B., Shipley, LA. (1999). Flavin-containing monooxygenase-mediated N-oxidation of the M(1)-muscarinic agonist xanomeline. *Drug Metab. Dispos.*, *27* (10), 1099–1103.

Ritter, J.K., Crawford, J.M., Owens, I.S. (1991). Cloning of two human liver bilirubin UDP-glucuronosyltransferase cDNAs with expression in COS-1 cells. *J. Biol. Chem.*, *266* (2), 1043–1047.

Rodriguez, R.J., Buckholz, C.J. (2003). Hepatotoxicity of ketoconazole in Sprague-Dawley rats: glutathione depletion, flavin-containing monooxygenases-mediated bioactivation and hepatic covalent binding. *Xenobiotica*, *33* (4), 429–441.

Rodriguez, R.J., Miranda, C.L. (2000). Isoform specificity of N-deacetyl ketoconazole by human and rabbit flavin-containing monooxygenases. *Drug Metab. Dispos.*, *28* (9), 1083–1086.

Roman, R.J., Carter, J.R., North, W.C., Kauker, M.L. (1977). Renal tubular site of action of fluoride in Fischer 344 rats. *Anesthesiology*, *46* (4), 260–264.

Rosa, F.W. (1991). Spina bifida in infants of women treated with carbamazepine during pregnancy. *N. Engl. J. Med.*, *324* (10), 674–677. Review.

Roujeau, J.C. (2006). Immune mechanisms in drug allergy. *Allergol. Int.*, *55* (1), 27–33. Review.

Salama, A., Mueller-Eckhardt, C. (1985). The role of metabolite-specific antibodies in nomifensine-dependent immune hemolytic anemia. *N. Engl. J. Med.*, *313* (8), 469–474.

Sanfeliu, C., Wright, J.M., Kim, S.U. (1999). Neurotoxicity of isoniazid and its metabolites in cultures of mouse dorsal root ganglion neurons and hybrid neuronal cell line. *Neurotoxicology*, *20* (6), 935–944.

Sanford-Driscoll, M., Knodel, L.C. (1986). Induction of hemolytic anemia by nonsteroidal antiinflammatory drugs. *Drug Intell. Clin. Pharm.*, *20* (12), 925–934. Review.

Sarich, T.C., Adams, S.P., Petricca, G., Wright, J.M. (1999). Inhibition of isoniazid-induced hepatotoxicity in rabbits by pretreatment with an amidase inhibitor. *J. Pharmacol. Exp. Ther.*, *289* (2), 695–702.

Scheulen, M.E., Kappus, H., Thyssen, D., Schmidt, C.G. (1981). Redox cycling of Fe(III)-bleomycin by NADPH-cytochrome P450 reductase. *Biochem. Pharmacol.*, *30* (24), 3385–3388.

Schmekel, B., Karlsson, S.E., Linden, M., Sundström, C., Tegner, H., Venge, P. (1990). Myeloperoxidase in human lung lavage. I. A marker of local neutrophil activity. *Inflammation*, *14* (4), 447–454.

Schnyder, B., Mauri-Hellweg, D., Zanni, M., Bettens, F., Pichler, W.J. (1997). Direct MHC-dependent presentation of the drug sulfamethoxazole to human alphabeta T cell clones. *J. Clin. Invest.*, *100* (1), 136–141.

Scordo, M.G., Pengo, V., Spina, E., Dahl, M.L., Gusella, M., Padrini, R. (2002). Influence of CYP2C9 and CYP2C19 genetic polymorphisms on warfarin maintenance: dose and metabolic clearance. *Clin. Pharmacol. Ther.*, *72* (6), 702–710.

Sebti, S.M., DeLeon, J.C., Lazo, J.S. (1987). Purification, characterization and amino acid composition of rabbit pulmonary bleomycin hydrolase. *Biochemistry*, *26* (14), 4213–4219.

Seeff, L.B., Cuccherini, B.A., Zimmerman, H.J., Adler, E., Benjamin, S.B. (1986). Acetaminophen hepatotoxicity in alcoholics. A therapeutic misadventure. *Ann. Intern. Med.*, *104* (3), 399–404. Review.

Seigneur, J., Trechot, P.F., Hubert, J., Lamy, P. (1988). Pulmonary complications of hormone treatment in prostate carcinoma. *Chest*, *93* (5), 1106.

Seitz, S., Boelsterli, U.A. (1998). Diclofenac acyl glucuronide, a major biliary metabolite, is directly involved in small intestinal injury in rats. *Gastroenterology*, *115* (6), 1476–1482.

Sharma, M., Shubert, D.E., Sharma, M., Lewis, J., McGarrigle, B.P., Bofinger, D.P., Olson, J.R. (2003). Biotransformation of tamoxifen in a human endometrial explant culture model. *Chem. Biol. Interact.*, *146* (3), 237–249.

Sheehan, D., Meade, G., Foley, V.M., Dowd, C.A. (2001). Structure, function and evolution of glutathione transferases: implications for classification of non-mammalian members of an ancient enzyme superfamily. *Biochem. J.*, *360* (Pt 1), 1–16. Review.

Shin, N.Y., Liu, Q., Stamer, S.L., Liebler, D.C. (2007). Protein targets of reactive electrophiles in human liver microsomes. *Chem. Res. Toxicol.*, *20* (6), 859–867.

Shipkova, M., Armstrong, V.W., Weber, L., Niedmann, P.D., Wieland, E., Haley, J., Tönshoff, B., Oellerich, M. (2002). German Study Group on Mycophenolate Mofetil Therapy in Pediatric Renal Transplant Recipients. Pharmacokinetics and protein adduct formation of the pharmacologically active acyl glucuronide metabolite of mycophenolic acid in pediatric renal transplant recipients. *Ther. Drug Monit.*, *24* (3), 390–399.

Shipkova, M., Strassburg, C.P., Braun, F., Streit, F., Gröne, H.J., Armstrong, V.W., Tukey, R.H., Oellerich, M., Wieland, E. (2001). Glucuronide and glucoside conjugation of mycophenolic acid by human liver, kidney and intestinal microsomes. *Br. J. Pharmacol.*, *132* (5), 1027–1034.

Siraki, A.G., Bonini, M.G., Jiang, J., Ehrenshaft, M., Mason, R.P. (2007). Aminoglutethimide-induced protein free radical formation on myeloperoxidase: a potential mechanism of agranulocytosis. *Chem. Res. Toxicol.*, *20* (7), 1038–1045.

Slack, J.A., Ford-Hutchinson, A.W., Richold, M., Choi, B.C. (1981). Some biochemical and pharmacological properties of an epoxide metabolite of alclofenac. *Chem. Biol. Interact.*, *34* (1), 95–107.

Sladek, N.E. (1988). Metabolism of oxazaphosphorines. *Pharmacol. Ther.*, *37* (3), 301–355. Review.

Slattery, J.T., Wilson, J.M., Kalhorn, T.F., Nelson, S.D. (1987). Dose-dependent pharmacokinetics of acetaminophen: evidence of glutathione depletion in humans. *Clin. Pharmacol. Ther.*, *41* (4), 413–418.

Slattum, P.W., Cato, A.E., III, Pollack, G.M., Brouwer, K.L. (1995). Age-dependent intestinal hydrolysis of valproate glucuronide in rat. *Xenobiotica*, *25* (3), 229–237.

Smith, P.C., Liu, J.H. (1993). Covalent binding of suprofen acyl glucuronide to albumin in vitro. *Xenobiotica*, *23* (4), 337–348.

Smith, P.C., Benet, L.Z., McDonagh, A.F. (1990). Covalent binding of zomepirac glucuronide to proteins: evidence for a Schiff base mechanism. *Drug Metab. Dispos.*, *18* (5), 639–644.

Smith, P.C., McDonagh, A.F., Benet, L.Z. (1986). Irreversible binding of zomepirac to plasma protein *in vitro* and *in vivo*. *J. Clin. Invest.*, *77* (3), 934–939.

Smith, K.S., Smith, P.L., Heady, T.N., Trugman, J.M., Harman, W.D., Macdonald, T.L. (2003). *In vitro* metabolism of tolcapone to reactive intermediates: relevance to tolcapone liver toxicity. *Chem. Res. Toxicol.*, *16* (2), 123–128.

Smyth, J.F., Bowman, A., Perren, T., Wilkinson, P., Prescott, R.J., Quinn, K.J., Tedeschi, M. (1997). Glutathione reduces the toxicity and improves quality of life of women diagnosed with ovarian cancer treated with cisplatin: results of a double-blind, randomized trial. *Ann. Oncol.*, *8* (6), 569–573.

Spaldin, V., Madden, S., Adams, D.A., Edwards, R.J., Davies, D.S., Park, B.K. (1995). Determination of human hepatic cytochrome P4501A2 activity in vitro use of tacrine as an isoenzyme-specific probe. *Drug Metab. Dispos.*, *23* (9), 929–934.

Spracklin, D.K., Hankins, D.C., Fisher, J.M., Thummel, K.E., Kharasch, E.D. (1997). Cytochrome P450 2E1 is the principal catalyst of human oxidative halothane metabolism *in vitro*. *J. Pharmacol. Exp. Ther.*, *281* (1), 400–411.

Sridar, C., Kent, U.M., Notley, L.M., Gillam, E.M., Hollenberg, P.F. (2002). Effect of tamoxifen on the enzymatic activity of human cytochrome CYP2B6. *J. Pharmacol. Exp. Ther.*, *301* (3), 945–952.

Stephens, J.R., Levy, R.H. (1992). Valproate hepatotoxicity syndrome: hypotheses of pathogenesis. *Pharm. Weekbl Sci.*, *14* (3A), 118–121. Review.

Stiborová, M., Poljaková, J., Ryslavá, H., Dracínský, M., Eckschlager, T., Frei, E. (2007). Mammalian peroxidases activate anticancer drug ellipticine to intermediates forming deoxyguanosine adducts in DNA identical to those found *in vivo* and generated from 12-hydroxyellipticine and 13-hydroxyellipticine. *Int. J. Cancer*, *120* (2), 243–251.

Stiborová, M., Sejbal, J., Borek-Dohalská, L., Aimová, D., Poljaková, J., Forsterová, K., Rupertová, M., Wiesner, J., Hudecek, J., Wiessler, M., Frei, E. (2004). The anticancer drug ellipticine forms covalent DNA adducts, mediated by human cytochromes P450, through metabolism to 13-hydroxyellipticine and ellipticine N2-oxide. *Cancer Res.*, *64* (22), 8374–8380.

Stillwell, T.J., Benson R.C., Jr (1988). Cyclophosphamide-induced hemorrhagic cystitis. A review of 100 patients. *Cancer*, *61* (3), 451–457.

Störmer, E., Roots, I., Brockmöller, J. (2000). Benzydamine N-oxidation as an index reaction reflecting FMO activity in human liver microsomes and impact of FMO3 polymorphisms on enzyme activity. *Br. J. Clin. Pharmacol.*, *50* (6), 553–561.

Stuart-Harris, R.C., Smith, I.E. (1984). Aminoglutethimide in the treatment of advanced breast cancer. *Cancer Treat. Rev.*, *11* (3), 189–204.

Styles, J.A., Davies, A., Lim, C.K., De Matteis, F., Stanley, L.A., White, I.N., Yuan, Z.X., Smith, L.L. (1994). Genotoxicity of tamoxifen, tamoxifen epoxide and toremifene in human lymphoblastoid cells containing human cytochrome P450s. *Carcinogenesis*, *15* (1), 5–9.

Subramanyam, B., Pond, S.M., Eyles, D.W., Whiteford, H.A., Fouda, H.G., Castagnoli N., Jr (1991). Identification of potentially neurotoxic pyridinium metabolite in the urine of schizophrenic patients treated with haloperidol. *Biochem. Biophys. Res. Commun.*, *181* (2), 573–578.

Tafazoli, S., O'Brien, P.J. (2005). Peroxidases: a role in the metabolism and side effects of drugs. *Drug Discov. Today*, *10* (9), 617–625. Review.

Takanashi, K., Tainaka, H., Kobayashi, K., Yasumori, T., Hosakawa, M., Chiba, K. (2000). CYP2C9 Ile359 and Leu359 variants: enzyme kinetic study with seven substrates. *Pharmacogenetics*, *10* (2), 95–104.

Takasuna, K., Hagiwara, T., Hirohashi, M., Kato, M., Nomura, M., Nagai, E., Yokoi, T., Kamataki, T. (1998). Inhibition of intestinal microflora beta-glucuronidase modifies the distribution of the active metabolite of the antitumor agent, irinotecan hydrochloride (CPT-11) in rats. *Cancer Chemother. Pharmacol.*, *42* (4), 280–286.

Takeda, S., Ishii, Y., Iwanaga, M., Mackenzie, P.I., Nagata, K., Yamazoe, Y., Oguri, K., Yamada, H. (2005). Modulation of UDP-glucuronosyltransferase function by cytochrome P450: evidence for the alteration of UGT2B7-catalyzed glucuronidation of morphine by CYP3A4. *Mol. Pharmacol.*, *67* (3), 665–672.

Tan, A.W., Thong, B.Y., Yip, L.W., Chng, H.H., Ng, S.K. (2005). High-dose intravenous immunoglobulins in the treatment of toxic epidermal necrolysis: an Asian series. *J. Dermatol.*, *32* (1), 1–6.

Tang, W., Stearns, R.A., Wang, R.W., Chiu, S.H., Baillie, T.A. (1999). Roles of human hepatic cytochrome P450s 2C9 and 3A4 in the metabolic activation of diclofenac. *Chem. Res. Toxicol.*, *12* (2), 192–199.

Temple, R. (2006). Hy's law: predicting serious hepatotoxicity. *Pharmacoepidemiol. Drug Saf.*, *15* (4), 241–243.

Timbrell, J.A., Mitchell, J.R., Snodgrass, W.R., Nelson, S.D. (1980). Isoniazid hepatoxicity: the relationship between covalent binding and metabolism *in vivo*. *J. Pharmacol. Exp. Ther.*, *213* (2), 364–369.

Tingle, M.D., Jewell, H., Maggs, J.L., O'Neill, P.M., Park, B.K. (1995). The bioactivation of amodiaquine by human polymorphonuclear leucocytes in vitro: chemical mechanisms and the effects of fluorine substitution. *Biochem. Pharmacol.*, *50* (7), 1113–1119.

Toffoli, G., Cecchin, E. (2004). Uridine diphosphoglucuronosyl transferase and methylenetetrahydrofolate reductase polymorphisms as genomic predictors of toxicity and response to irinotecan-, antifolate- and fluoropyrimidine-based chemotherapy. *J. Chemother.*, *4*, 31–35.

Toffoli, G., Cecchin, E., Corona, G., Russo, A., Buonadonna, A., Pasetto, L.M., D'Andrea, M., Pessa, S., Errante, D., De Pangher, V., Giusto, M., Medici, M., Gaion, F., Sandri, P., Galligioni, E., Bonura, S., Boccalon, M., Biason, P., Frustaci, S. (2006). The role of UGT1A1*28 polymorphism in the pharmacodynamics and pharmacokinetics of irinotecan in patients with metastatic colorectal cancer. *J. Clin. Oncol.*, *24*, 3061–3068.

To-Figueras, J., Gene, M., Gomez-Catalan, J., Galan, C., Firvida, J., Fuentes, M., Rodamilans, M., Huguet, E., Estape, J., Corbella, J. (1996). Glutathione-S-Transferase M1 and codon 72 p53 polymorphisms in a northwestern Mediterranean population and their relation to lung cancer susceptibility. *Cancer Epidemiol Biomarkers Prev.*, *5* (5), 337–42.

Tong, K.I., Kobayashi, A., Katsuoka, F., Yamamoto, M. (2006). Two-site substrate recognition model for the Keap1-Nrf2 system: a hinge and latch mechanism. *Biol. Chem.*, *387* (10–11), 1311–1320. Review.

Tong, V., Teng, X.W., Karagiozov, S., Chang, T.K., Abbott, F.S. (2005). Valproic acid glucuronidation is associated with increases in 15-F2t-isoprostane in rats. *Free Radic. Biol. Med.*, *1*; 38 (11), 1471–1483.

Townsend, D.M., Deng, M., Zhang, L., Lapus, M.G., Hanigan, M.H. (2003). Metabolism of cisplatin to a nephrotoxin in proximal tubule cells. *J. Am. Soc. Nephrol.*, *14* (1), 1–10.

Townsend, D.M., Hanigan, M.H. (2002). Inhibition of gamma-glutamyl transpeptidase or cysteine S-conjugate beta-lyase activity blocks the nephrotoxicity of cisplatin in mice. *J. Pharmacol. Exp. Ther.*, *300* (1), 142–148.

Toyoda, Y., Tsuchida, A., Iwami, E., Miwa, I. (2001). Toxic effect of troglitazone on cultured rat hepatocytes. *Life Sci.*, *68* (16), 1867–1876.

Traversa, G., Bianchi, C., Da Cas, R., Abraha, I., Menniti-Ippolito, F., Venegoni, M. (2003). Cohort study of hepatotoxicity associated with nimesulide and other non-steroidal anti-inflammatory drugs. *BMJ*, *327* (7405), 18–22.

Treinen-Moslen, M., Kanz, M.F. (2006). Intestinal tract injury by drugs: importance of metabolite delivery by yellowbile road. *Pharmacol. Ther.*, *112* (3), 649–667. Review.

Trepanier, L.A., Ray, K., Winand, N.J., Spielberg, S.P., Cribb, A.E. (1997). Cytosolic arylamine N-acetyltransferase (NAT) deficiency in the dog and other canids due to an absence of NAT genes. *Biochem. Pharmacol.*, *54* (1), 73–80.

Tsai-Turton, M., Luong, B.T., Tan, Y., Luderer, U. (2007). Cyclophosphamide-induced apoptosis in COV434 human granulosa cells involves oxidative stress and glutathione depletion. *Toxicol. Sci.*, *98* (1), 216–230.

Tucker, G.T., Lennard, M.S., Ellis, S.W., Woods, H.F., Cho, A.K., Lin, L.Y., Hiratsuka, A., Schmitz, D.A., Chu, T.Y. (1994). The demethylenation of methylenedioxymethamphetamine ("ecstasy") by debrisoquine hydroxylase (CYP2D6). *Biochem. Pharmacol.*, *47* (7), 1151–1156.

Tukey, R.H., Strassburg, C.P. (2000). Human UDP-glucuronosyltransferases: metabolism, expression and disease. *Annu. Rev. Pharmacol. Toxicol.*, *40*, 581–616.

Uetrecht, J. (2001). Prediction of a new drug's potential to cause idiosyncratic reactions. *Curr. Opin. Drug Discov. Devel.*, *4* (1), 55–59. Review.

Uetrecht, J., Zahid, N., Shear, N.H., Biggar, W.D. (1988). Metabolism of dapsone to a hydroxylamine by human neutrophils and mononuclear cells. *J. Pharmacol. Exp. Ther.*, *245* (1), 274–279.

Uetrecht, J.P. (1992). The role of leukocyte-generated reactive metabolites in the pathogenesis of idiosyncratic drug reactions. *Drug Metab. Rev.*, *24* (3), 299–366. Review.

Uetrecht, J.P. (1999). New concepts in immunology relevant to idiosyncratic drug reactions: the "danger hypothesis" and innate immune system. *Chem. Res. Toxicol.*, *12* (5), 387–395. Review.

Uetrecht, J.P., Zahid, N., Whitfield, D. (1994). Metabolism of vesnarinone by activated neutrophils: implications for vesnarinone-induced agranulocytosis. *J. Pharmacol. Exp. Ther.*, *270* (3), 865–872.

Usuki, E., Van der Schyf, C.J., Castagnoli N., Jr (1998). Metabolism of haloperidol and its tetrahydropyridine dehydration product HPTP. *Drug Metab. Rev.*, *30* (4), 809–826.

van Berkel, W.J., Kamerbeek, N.M., Fraaije, M.W. (2006). Flavoprotein monooxygenases, a diverse class of oxidative biocatalysts. *J. Biotechnol.*, *124* (4), 670–689.

Vatsis, K.P., Weber, W.W. (1993). Structural heterogeneity of Caucasian N-acetyltransferase at the NAT1 gene locus. *Arch. Biochem. Biophys.*, *301* (1), 71–76.

Vella, A., de Groen, P.C., Dinneen, S.F. (1998). Fatal hepatotoxicity associated with troglitazone. *Ann. Intern. Med.*, *129* (12), 1080.

Viard-Leveugle, I., Bullani, R.R., Meda, P., Micheau, O., Limat, A., Saurat, J.H., Tschopp, J., French, L.E. (2003). Intracellular localization of keratinocyte Fas ligand explains lack of cytolytic activity under physiological conditions. *J. Biol. Chem.*, *278* (18), 16183–16188.

Vitseva, O., Flockhart, D.A., Jin, Y., Varghese, S., Freedman, J.E. (2005). The effects of tamoxifen and its metabolites on platelet function and release of reactive oxygen intermediates. *J. Pharmacol. Exp. Ther.*, *312* (3), 1144–1150.

Vogel, C. (2000). Prostaglandin H synthases and their importance in chemical toxicity. *Curr. Drug Metab.*, *1* (4), 391–404. Review.

Vyas, P.M., Roychowdhury, S., Woster, P.M., Svensson, C.K. (2005). Reactive oxygen species generation and its role in the differential cytotoxicity of the arylhydroxylamine metabolites of sulfamethoxazole and dapsone in normal human epidermal keratinocytes. *Biochem. Pharmacol.*, *70* (2), 275–286.

Wade, L.T., Kenna, J.G., Caldwell, J. (1997). Immunochemical identification of mouse hepatic protein adducts derived from the nonsteroidal anti-inflammatory drugs diclofenac, sulindac and ibuprofen. *Chem. Res. Toxicol.*, *10* (5), 546–555.

Wadkins, R.M., Hyatt, J.L., Yoon, K.J., Morton, C.L., Lee, R.E., Damodaran, K., Beroza, P., Danks, M.K., Potter, P.M. (2004). Discovery of novel selective inhibitors of human intestinal carboxylesterase for the amelioration of irinotecan-induced diarrhea: synthesis, quantitative structure-activity relationship analysis and biological activity. *Mol. Pharmacol.*, *65* (6), 1336–1343.

Walker, A.M. (1997). Quantitative studies of the risk of serious hepatic injury in persons using nonsteroidal antiinflammatory drugs. *Arthritis. Rheum.*, *40* (2), 201–208.

Walsh, J.S., Reese, M.J., Thurmond, L.M. (2002). The metabolic activation of abacavir by human liver cytosol and expressed human alcohol dehydrogenase isozymes. *Chem. Biol. Interact.*, *142* (1–2), 135–154.

Wang, J., Davis, M., Li, F., Azam, F., Scatina, J., Talaat, R. (2004). A novel approach for predicting acyl glucuronide reactivity via Schiff base formation: development of rapidly formed peptide adducts for LC/MS/MS measurements. *Chem. Res. Toxicol.*, *17* (9), 1206–1216.

Wang, M., Gorrell, M.D., McGaughan, G.W., Dickinson, R.G. (2001). Dipeptidyl peptidase IV is a target for covalent adduct formation with the acyl glucuronide metabolite of the anti-inflammatory drug zomepirac. *Life Sci.*, *68* (7), 785–797.

Wang, Y., Gray, J.P., Mishin, V., Heck, D.E., Laskin, D.L., Laskin, J.D. (2008). Role of cytochrome P450 reductase in nitrofurantoin-induced redox cycling and cytotoxicity. *Free Radic. Biol. Med.*, *44* (6), 1169–1179.

Weinshilboum, R., Wang, L. (2004). Pharmacogenomics: bench to bedside. *Nat. Rev. Drug Discov.*, *3* (9), 739–748. Review.

Weissman, J., Trevor, A., Chiba, K., Peterson, L.A., Caldera, P., Castagnoli, N.Jr, , Baillie, T. (1985). Metabolism of the nigrostriatal toxin 1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine by liver homogenate fractions. *J. Med. Chem.*, *28* (8), 997–1001.

Wells, P.G., Bhuller, Y., Chen, C.S., Jeng, W., Kasapinovic, S., Kennedy, J.C., Kim, P.M., Laposa, R.R., McCallum, G.P., Nicol, C.J., Parman, T., Wiley, M.J., Wong, A.W. (2005). Molecular and biochemical mechanisms in teratogenesis involving reactive oxygen species. *Toxicol. Appl. Pharmacol.*, *207* (Suppl. 2), 354–366.

Wells, P.G., Kim, P.M., Laposa, R.R., Nicol, C.J., Parman, T., Winn, L.M. (1997). Oxidative damage in chemical teratogenesis. *Mutat. Res.*, *396* (1–2), 65–78. Review.

Whetstine, J.R., Yueh, M.F., McCarver, D.G., Williams, D.E., Park, C.S., Kang, J.H., Cha, Y.N., Dolphin, C.T., Shephard, E.A., Phillips, I.R., Hines, R.N. (2000). Ethnic differences in human flavin-containing monooxygenase 2 (FMO2) polymorphisms: detection of expressed protein in African-Americans. *Toxicol. Appl. Pharmacol.*, *168*, 216–224.

Williams, A.C., Ramsden, D.B. (2005). Autotoxicity, methylation and a road to the prevention of Parkinson's disease. *J. Clin. Neurosci.*, *12* (1), 6–11. Review.

Williams, D.P., Pirmohamed, M., Naisbitt, D.J., Maggs, J.L., Park, B.K. (1997). Neutrophil cytotoxicity of the chemically reactive metabolite(s) of clozapine: possible role in agranulocytosis. *J. Pharmacol. Exp. Ther.*, *283* (3), 1375–1382.

Williams, J.A., Hyland, R., Jones, B.C., Smith, D.A., Hurst, S., Goosen, T.C., Peterkin, V., Koup, J.R., Ball, S.E. (2004). Drug-drug interactions for UDP-glucuronosyltransferase substrates: a pharmacokinetic explanation for typically observed low exposure (AUCi/AUC) ratios. *Drug Metab. Dispos.*, *32* (11), 1201–1208.

Witten, C.M. (1989). Pulmonary toxicity of nitrofurantoin. *Arch. Phys. Med. Rehabil.*, *70* (1), 55–57.

Woeber, KA. (2002). Methimazole-induced hepatotoxicity. *Endocr. Pract.*, *8* (3), 222–224. Review.

Xie, H.G., Prasad, H.C., Kim, R.B., Stein, C.M. (2002). CYP2C9 allelic variants: ethnic distribution and functional significance. *Adv. Drug Deliv. Rev.*, *54* (10), 1257–1270.

Xu, L.L., Liu, G., Miller, D.P., Zhou, W., Lynch, T.J., Wain, J.C., Su, L., Christiani, D.C. (2002). Counterpoint: the myeloperoxidase $^{-463}$G → a polymorphism does not decrease lung cancer susceptibility in Caucasians. *Cancer Epidemiol. Biomarkers Prev.*, *11* (12), 1555–1559.

Yamamoto, Y., Nakajima, M., Yamazaki, H., Yokoi, T. (2001). Cytotoxicity and apoptosis produced by troglitazone in human hepatoma cells. *Life Sci.*, *70* (4), 471–482.

Yamamoto, Y., Yamazaki, H., Ikeda, T., Watanabe, T., Iwabuchi, H., Nakajima, M., Yokoi, T. (2002). Formation of a novel quinone epoxide metabolite of troglitazone with cytotoxicity to HepG2 cells. *Drug Metab. Dispos.*, *30* (2), 155–160.

You, Q., Cheng, L., Reilly, T.P., Wegmann, D., Ju, C. (2006). Role of neutrophils in a mouse model of halothane-induced liver injury. *Hepatology*, *44* (6), 1421–1431.

Yukinaga, H., Takami, T., Shioyama, S.H., Tozuka, Z., Masumoto, H., Okazaki, O., Sudo, K. (2007). Identification of cytochrome P450 3A4 modification site with reactive metabolite using linear ion trap-Fourier transform mass spectrometry. *Chem. Res. Toxicol.*, *20* (10), 1373–1378.

Yunis, A.A., Miller, A.M., Salem, Z., Corbett, M.D., Arimura, GK. (1980). Nitroso-chloramphenicol possible mediator in chloramphenicol-induced aplastic anemia. *J. Lab. Clin. Med.*, *96* (1), 36–46.

Zhang, H., Cook, J., Nickel, J., Yu, R., Stecker, K., Myers, K., Dean, N.M. (2000). Reduction of liver Fas expression by an antisense oligonucleotide protects mice from fulminant hepatitis. *Nat. Biotechnol.*, *18* (8), 862–867.

Zhang, L., Cooper, A.J., Krasnikov, B.F., Xu, H., Bubber, P., Pinto, J.T., Gibson, G.E., Hanigan, M.H. (2006). Cisplatin-induced toxicity is associated with platinum deposition in mouse kidney mitochondria *in vivo* and with selective inactivation of the alpha-ketoglutarate dehydrogenase complex in LLC-PK1 cells. *Biochemistry*, *45* (29), 8959–8971.

Zhao, P., Slattery, J.T. (2002). Effects of ethanol dose and ethanol withdrawal on rat liver mitochondrial glutathione: implication of potentiated acetaminophen toxicity in alcoholics. *Drug Metab. Dispos.*, *30* (12), 1413–1417.

Ziegler, D.M. (1993). Recent studies on the structure and function of multisubstrate flavin-containing monooxygenases. *Annu. Rev. Pharmacol. Toxicol.*, *33*, 179–199.

Zimmerman, H.J., Maddrey, W.C. (1995). Acetaminophen (paracetamol) hepatotoxicity with regular intake of alcohol: analysis of instances of therapeutic misadventure. *Hepatology*, *22* (3), 767–773. Erratum in Hepatology 1995 Dec, 22 (6), 1898.

Zuckerman, A., Bolan, E., de Paulis, T., Schmidt, D., Spector, S., Pasternak, G.W. (1999). Pharmacological characterization of morphine-6-sulfate and codeine-6-sulfate. *Brain Res.*, *842* (1), 1–5.

Zuidema, J., Hilbers-Modderman, E.S., Merkus, F.W. (1986). Clinical pharmacokinetics of dapsone. *Clin. Pharmacokinet.*, *11* (4), 299–315.

# Allergic Reactions to Drugs

MARK P. GRILLO

> Now this is not the end. It is not even the beginning of the end. But it is, perhaps, the end of the beginning.[1]

## 1 INTRODUCTION

There are many pharmaceutical scientists today who say that one of the most difficult obstacles to overcome in drug discovery is the prediction, and therefore the knowledgeable prevention, of idiosyncratic allergic reactions to drugs. The immune system is a highly complex and developed system and is our natural defense in fighting varied pathogenic organisms. The more we have learned about the immune system over the last two decades, due in large part to molecular advancements in immunology, the more it has revealed for researchers to come closer to understanding the mechanism(s) of allergic reactions to drugs (Uetrecht, 2008). Unfortunately, the immune system is susceptible to devastating diseases. Such diseases can occur when the immune system does not differentiate between host cells/macromolecules and foreign substances which leads to autoimmune-type disorders like multiple sclerosis, systemic lupus, rheumatoid arthritis, type I diabetes, and inflammatory bowel disease. In this chapter, some of the proposed chemical mechanisms by which example drugs cause allergic reactions will be discussed. Allergic reactions to drugs can lead to potentially life-threatening reactions such as asthma, anaphylaxis, dermatitis, hepatitis, hemolytic anemia, lupus, nephritis, Stevens–Johnson syndrome, toxic epidermal necrolysis, urticaria, and vasculitis (Dansette *et al.*, 1998; Ewan, 1998; Zimmerman, 1999; Fung *et al.*, 2001). These drug-induced immune-based diseases can create an enormous burden to affected patients and to the respective pharmaceutical companies involved in their manufacturing. There

[1] A quotation from Winston Churchill—*Lord Mayor's Luncheon, Mansion House following the victory at El Alameinin North Africa, London, 10 November 1942* (http://www.winstonchurchill.org/).

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

were 121 drugs withdrawn from the worldwide market between 1960 and 1999, where hepatic (26.2%), hematologic (10.5%), dermatologic (6.3%), and anaphylaxis (3.3%) were among the top six toxicological reasons for their withdrawal (Fung *et al.*, 2001). Such toxicities can deprive patients of organ function that sometimes results in fatalities. A primary reason why drugs fail today in drug development is due to animal toxicity and human adverse events (Guengerich and MacDonald, 2007). A functional understanding of the association between the use of some drugs and the unpredictable formation of drug-induced hypersensitivity reactions is of major importance in the pharmaceutical industry today (Baillie, 2006; Guengerich and MacDonald, 2007; Kumar *et al.*, 2008). The medical and clinical implications are clear. Immune-based drug-induced allergic diseases may also present scientific opportunities. Studying them could answer critical questions about the nature of the immune system and yield increasingly valuable information as it relates to the discovery of safer drugs. In this chapter, an examination of a set of drug-induced allergic reactions will be reviewed, each mediated by different mechanisms that are driven by the inherently different chemistry and biochemistry of the drug. Examples of drug-induced allergic reactions reviewed in the present chapter include penicillin, a β-lactam antibiotic; felbamate, an anticonvulsant drug; halothane, an inhalation anesthetic; tienilic acid, a diuretic drug; sulfamethoxazole, a sulfonamide-containing antibiotic drug; and benoxaprofen, a carboxylic acid-containing nonsteroidal anti-inflammatory drug (NSAID).

## 2 IMMUNE SYSTEM: A BRIEF OVERVIEW

Both innate (non-memory) and adaptive (specificity and memory) immunity lines of defense against foreign antigens are involved in the formation of allergic reactions (Goldsby *et al.*, 2003). An allergic reaction can be divided into two primary stages, namely sensitization and elicitation. The sensitization stage occurs when a foreign substance, for example a drug–protein adduct or a pathogenic organism, is dealt with by a mechanism known as antigen processing, sometimes referred to as the "gateway" to the immune response. Here the antigen becomes engulfed and then degraded by antigen-presenting cells (macrophages), resulting in peptide fragments (approximately 8–15 amino acids) of the foreign substance (or microorganism) being presented on the surface of antigen-presenting cells, and specifically on the major histocompatibility complex (MHC) for T lymphocyte recognition (Kalish, 1995; Park, Pirmohamed and Kitteringham, 1998). Antigen-specific helper and killer T cells (T lymphocytes from the thymus) recognize, and then become activated by, these antigen-presenting cells. The activated helper and killer T cells then replicate, including the production specialized memory T cells, which are cells that respond to subsequent exposure to a foreign substance or infection. Importantly, killer T cells (which destroy cells containing the foreign antigen, i.e. cellular-mediated immunity) require helper T cells for activation. Helper T cells also function in the activation of B cells which are important in the elicitation stage of an allergic reaction. The elicitation (inflammation) stage begins with immature B cells (lymphocytes that produce antibodies) found in bone marrow. Once B cells become mature, which occurs in secondary lymphoid organs, they interact with, and become further activated by, free specific antigens in the blood. B cells are important because they function in the development of antibody-mediated immunity (i.e. humoral immunity). Just as in the case of killer T cell activation, helper T cells also activate B cells, leading to B cell replication and subsequent production of antibodies that bind

to specific antigens. Then, either a complement system destroys the specific antigen, or phagocytic cells engulf the antigen prior to their degradation. Memory B cells, also formed upon B cell activation by helper T cells, are very important because they can respond to subsequent exposures of an antigen, resulting in an immune response sometimes associated with immune-mediated drug toxicity. The present chapter is confined primarily to the metabolism of drugs as it applies to the elicitation of allergic reactions; therefore, the reader is directed to a more in-depth review of the cellular basis of hypersensitivity reactions (Park, Pirmohamed and Kitteringham, 1998; Janeway *et al.*, 2005) for further reading.

## 3 DRUG METABOLISM AND THE HAPTEN HYPOTHESIS

A principal working hypothesis still used today since its inception in 1935 is the hapten hypothesis, where a drug or reactive drug metabolite (see Chapter Chemical Mechanisms in Toxicology) becomes covalently bound to protein, is recognized by the immune system as foreign, and leads to an allergic-type reaction (Landsteiner and Jacobs, 1935; Park, Coleman and Kitteringham, 1987; Uetrecht, 2008) (Fig. 1). For example, there are four stages of allergic hepatitis mediated by reactive metabolites formed in the liver. The first stage is metabolic activation of a xenobiotic to a reactive metabolite leading to the formation of a hepatocellular protein adduct. The second stage is subcellular

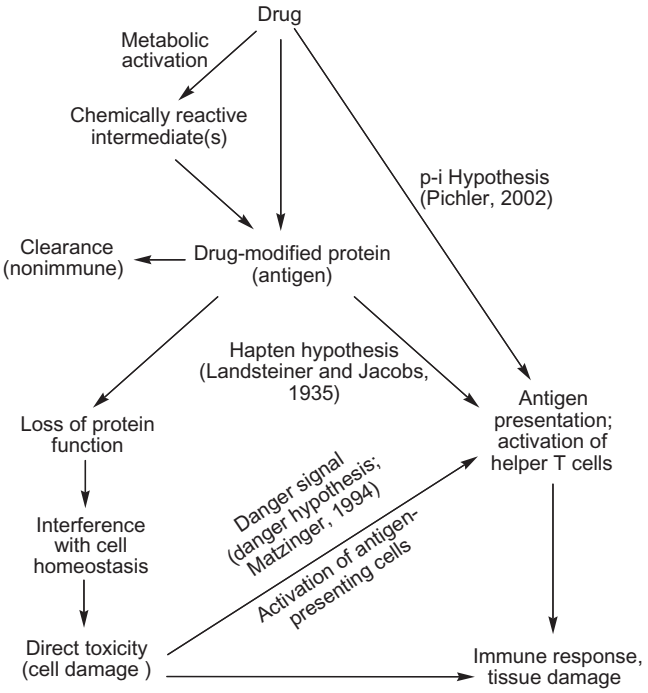


**Figure 1** Scheme for the metabolism of drugs to chemically reactive intermediates leading to direct toxicity or immune-mediated toxicity (adapted from Pirmohamed, Madden and Park, 1996 and Dansette *et al.*, 1998).

trafficking and then to cell surface expression of protein adduct fragments. Then, in the third stage, antigen presentation and immune recognition promote an antibody response and cellular sensitization. Finally, in the fourth stage, an immune attack on adducted hepatocytes leads to immune-mediated liver injury.

Other more recent hypotheses for the cause of drug-induced allergic reactions include the danger hypothesis and the p-i (pharmacological interaction) hypothesis. The danger hypothesis follows a scheme where a reactive drug metabolite causing cellular damage induces an endogenous signal, referred to as the "danger" signal, which leads to the activation of antigen-presenting cells as a major factor in an immune response (Matzinger, 1994; Uetrecht, 1999). An example of a drug proposed to mediate allergic reactions in this manner is the NSAID diclofenac (see Section 4.6). The p-i hypothesis proposes that immune system T cells recognize the parent drug as foreign, rather than drug–protein adducts, leading to an immune reaction (Pichler, 2002). The anticancer drug oxaliplatin provides an example where the parent molecule is believed to have a direct effect on the immune system leading to an allergic reaction (Maindrault-Goebel *et al.*, 2005). Another example is for the anticonvulsant drug lamotrigine, where it was shown that T cells from lamotrigine-hypersensitive patients recognize the drug in the absence of drug metabolism, covalent binding to protein, or antigen processing (Naisbitt *et al.*, 2003).

Allergic-based toxicities that might result from the haptenization of proteins can lead to a triggering of the immune system in hypersensitive individuals (Ju and Uetrecht, 2002; Uetrecht, 2003; Kaplowitz, 2005). These toxicities, which are sometimes fatal, can show an extremely low incidence, ranging from 1 in 1000 to as low as 1 in 100,000 patients (Zimmerman, 1999; Uetrecht, 2008). This is a very serious concern in the pharmaceutical industry, since there have been several products withdrawn from the market because of immune-based toxicities (Fung *et al.*, 2001; Walgren, Mitchell and Thompson, 2005; Guengerich and MacDonald, 2007). It is of especially strong concern because there are almost no experimental models for the human immune system (Shenton, Chen and Uetrecht, 2004; van Wijk and Nierkens, 2006).

## 4 ALLERGIC REACTIONS TO DRUGS (EXAMPLES)

### 4.1 Penicillin

Penicillin (benzylpenicillin; Fig. 2) is a β-lactam antibiotic that was discovered serendipitously in 1928 by the Scottish biologist Sir Alexander Fleming (Fleming, 1929). The mechanism of action of penicillin occurs through inhibition of the growth of gram-positive bacteria by interrupting the biosynthesis of cell wall peptidoglycan proteins that are important for membrane shape and integrity (Zubay, 1988). Of course, the selective toxicity of penicillin for bacteria is due to the fact that animal cells do not possess a comparable structure. Penicillin is a structural analog to that of the D-alanyl-D-alanine *C*-terminal end of peptidoglycan proteins that are normally cross-linked by transpeptidase enzymes. Because of the affinity of penicillin for this transpeptidase enzyme, and due to the high chemical reactivity of the strained 4-membered β-lactam ring, penicillin covalently binds to an active site serine, via transacylation, thereby irreversibly inactivating the enzyme (Wise and Park, 1965; Yocum *et al.*, 1979).

The first treatment and cure of streptococcal infection by penicillin occurred in 1942 in New Haven, Connecticut (Saxon, 1999; Markel, 2004). Since then, penicillin has

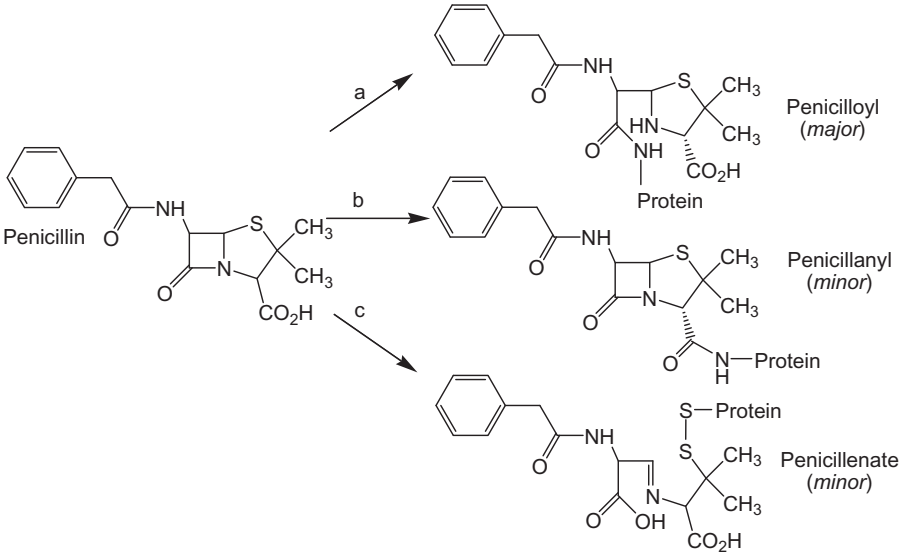


**Figure 2** Chemical mechanisms for the covalent binding of penicillin to protein by **(a)** reaction of a protein-lysinyl amine with the carbonyl carbon of the β-lactam moiety; **(b)** reaction of a protein-lysinyl amine with a reactive acylating metabolite of penicillin; and **(c)** reaction of a protein-cysteinyl sulfhydryl with the thiazolidine ring sulfhydryl group (adapted from Gruchalla, 2003).

become the most widely used antibiotic worldwide for gram-positive bacterial infections (Wright, 1999).

Allergic reactions to penicillin are the most common adverse side effect of the drug and also represent the most common cause of drug hypersensitivity reactions in humans (Mandell and Petri, 1996; Neugut, Ghatak and Miller, 2001; Solensky, 2003). An allergic reaction to penicillin, where an IgE-mediated response occurs, has long been considered a prototype for drug-mediated allergic reactions. Allergic reactions to penicillin can lead to mild side effects such as skin rashes, but also to a potentially life-threatening, and sometimes fatal, anaphylactic shock.

The mechanism of penicillin allergic reaction in patients is believed to be due to the inherent chemical reactivity of the drug with nucleophiles such as lysine groups on proteins that react with the carbonyl carbon of the β-lactam ring, leading to the formation of protein adducts. Such adducts then function as antigens and trigger the immune system response (Fig. 2) (Levine, 1966; Josephson, 2004). The ε-lysyl–amide-linked penicillin adduct is referred to as the penicilloyl–protein conjugate and is designated the major antigenic determinant with respect to its quantitative formation, since approximately 95% of penicillin molecules are bound to protein in this manner (Levine, 1966; Gruchalla, 2003). Other penicillin–protein adducts, and therefore putative penicillin antigens, have been identified as penicillenate– and penicillanyl–protein adducts (Levine and Redmond, 1969; Baldo, 1999). Penicillenate–protein adducts are proposed to be formed by a mechanism where penicillin first undergoes isomerization to penicillenic acid, a thiazolidine ring opened structure bearing a free sulfhydryl group, and secondly reacts with protein sulfhydryls to form stable drug–protein mixed disulfide

linkages (Parker *et al.*, 1962). The penicillenate–protein adduct is known to be inherently more immunogenic than penicilloyl–protein adduct (Park, Pirmohamed and Kitteringham, 1998). The third known antigenic determinant is the penicillanyl–protein adduct, which is formed when penicillin becomes adducted to ε-lysyl amine residues on proteins via the carbonyl carbon of the carboxylic acid moiety of penicillin. The mechanism of formation of the penicillanyl–protein adduct, to my knowledge, has not been shown, but presumably occurs by reaction of an acyl-linked metabolite of penicillin with protein lysine groups forming the penicillanyl–protein amide linkage. Both the penicillenate– and penicillanyl–protein adducts are known as minor antigenic determinants, together comprising approximately 5% of total penicillin-related protein adducts. In summary, penicillin-mediated hypersensitivity reactions go together with the hapten hypothesis, in which the inherent chemical reactivity and instability of the β-lactam moiety of penicillin allows it to function as the hapten (Uetrecht, 2007). The mechanism of the idiosyncratic nature of the IgE response that some patients have to penicillin is not yet known.

## 4.2 Felbamate

Felbamate (2-phenyl-1,3-propanediol dicarbamate; Fig. 3) is a broad-spectrum antiepileptic drug that was approved by the Food and Drug Administration and introduced into the market in 1993, where it was the first new anticonvulsant in 20 years (Wagner, 1994). This was an exciting medication for the treatment of epilepsy because it was effective in patients refractive to other types of anticonvulsant drugs, and also because

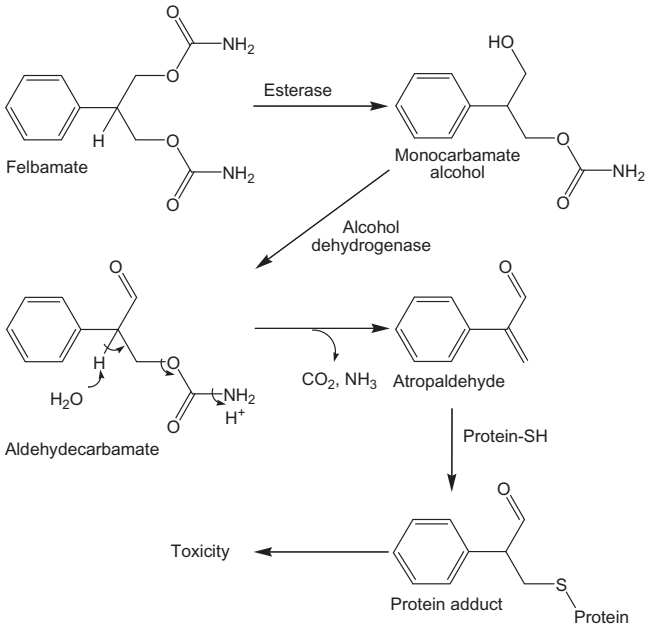


**Figure 3** Proposed chemical mechanism for the metabolic activation of felbamate to the reactive intermediate, atropaldehyde, which reacts covalently with protein leading to toxicity (Dieckhaus *et al.*, 2002).

it was non-sedating. Unfortunately, the use of felbamate was severely curtailed when black box warnings were given after only one year on the market. It was found that out of the 120,000 patients that had received the drug, 33 cases of aplastic anemia and 18 cases of hepatic failure occurred (Pellock, 1999). Since 1994, the Food and Drug Administration has restricted felbamate for compassionate use only. The mechanism of hepatotoxicity is proposed to be mediated by reactive metabolites of the drug (Dieckhaus *et al.*, 2002; Kapetanovic *et al.*, 2002). Bioactivation of felbamate has been shown to occur through a mechanism where it is first metabolized by an esterase to provide a monocarbamate alcohol intermediate. This alcohol metabolite is further oxidized by alcohol dehydrogenase to the corresponding aldehyde carbamate intermediate, which decomposes by the loss of $CO_2$ and ammonia to the reactive metabolite, atropaldehyde. Atropaldehyde is then proposed to react with protein cysteinyl sulfhydryls, in a Michael acceptor-type fashion, resulting in the formation of covalent protein adducts that are immunogenic (Dieckhaus *et al.*, 2000) (Fig. 3).

In a recent toxicogenomic study, felbamate was evaluated along with other anticonvulsant drugs for potential toxicity based on *in vitro* and *in vivo* covalent binding to protein and corresponding gene expression. In that study, felbamate was shown to produce robust effects on oxidative stress/reactive metabolite reporter genes in rat liver. However, the drug was relatively weak in terms of covalent binding to protein in model *in vivo* and *in vitro* systems. For instance, in incubations with rat liver microsomes (100 μM for 2 hours), felbamate-derived covalent binding to protein was measured to be 80 pmol equivalents/mg protein versus 523 pmol equivalents/mg protein detected for acetaminophen in the same study (Leone *et al.*, 2007). These toxicogenomic studies showed that gene expression results from rat liver (as indicators of oxidative stress/reactive metabolite formation) can be very useful for screening candidate drugs for potential idiosyncratic drug toxicity (McMillian *et al.*, 2004).

Fluorofelbamate, a fluorinated analog of felbamate (where the hydrogen atom attached at the benzylic carbon is replaced by a fluorine atom), was designed to block the breakdown to the atropaldehyde reactive metabolite (Hovinga, 2002; Roecklein *et al.*, 2007). The carbon–fluorine bond is chemically inert, such that the corresponding aldehyde carbamate metabolite of fluorofelbamate does not lose hydrofluoric acid leading to the formation of atropaldehyde. Fluorofelbamate, which retains the broad-spectrum activity of felbamate, is a much more potent anticonvulsant than felbamate, shows no evidence of glutathione conjugation *in vivo*, and is currently a phase I drug candidate (Bialer, 2006; Roecklein *et al.*, 2007).

### 4.3 Halothane

Halothane (2-bromo-2-chloro-1,1,1-trifluoroethane) is a volatile inhalation anesthetic used for general anesthesia that has been shown to lead to cases of hepatotoxicity resembling an allergic reaction (Fig. 4) (Davies, 1973; Satoh *et al.*, 1985; Kenna *et al.*, 1988). Halothane is metabolized to reactive intermediates (Fig. 4) that bind covalently to hepatic proteins (van Dyke and Gandolfi, 1974; Gandolfi *et al.*, 1980). In some cases, halothane causes a fulminant hepatitis that is associated with an immune response to halothane-derived protein adducts (Pohl *et al.*, 1988). Increases in plasma transaminases are detected in less than 20% of patients, indicating a mild hepatotoxicity which is proposed to be due to reductive metabolism of halothane by cytochrome P450 (P450 2D6 and P450 3A4), leading to free radical formation causing lipid peroxidation and covalent binding to protein (Fig. 4) (Spracklin, Thummel and Kharasch,

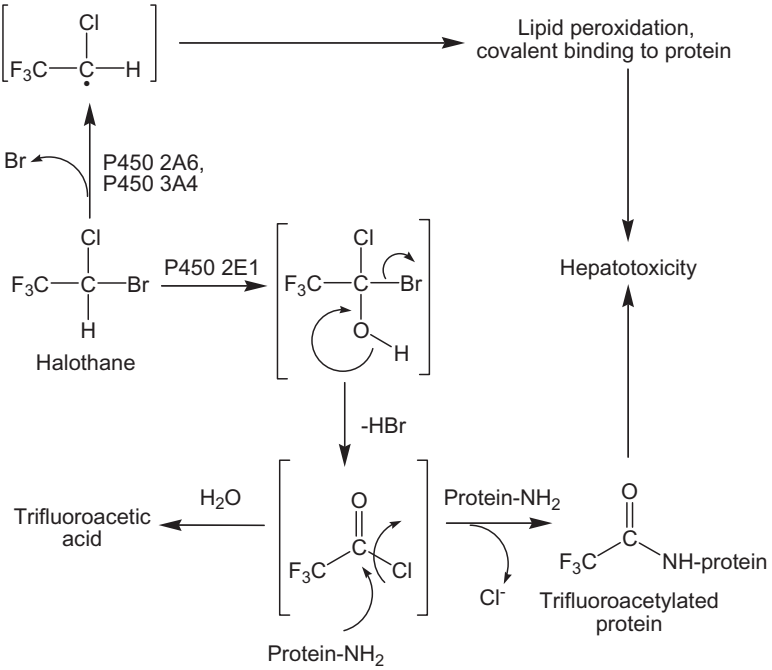


**Figure 4** Chemical mechanism for the P450-mediated bioactivation of halothane, leading to the formation of a reactive free radical intermediate and a trifluoroacetyl chloride acylating species that covalently binds to protein nucleophiles (Spracklin, Thummel and Kharasch, 1996).

1996). Fulminant hepatic necrosis, also referred to as "halothane hepatitis," is rare (1:3500 patients after multiple exposures to halothane) but can be fatal (National Halothane Study, 1966). Halothane hepatitis is proposed to be immunologically mediated by the reactive trifluoroacetylchloride metabolite hapten which is formed from P450 2E1-mediated oxidative metabolism. Chemically reactive trifluoroacetyl chloride reacts, in a transacylation fashion, with lysine residues on hepatic proteins leading to the formation of antigenic trifluoroacetylated-protein adducts. These antigens are proposed to initiate humoral and cellular immune responses that lead to hepatitis. Possible immunogens causing halothane hepatitis include protein disulfide isomerase, carboxylesterase, Erp72 heat shock protein, immunoglobulin heavy chain-binding protein/glucose-regulated protein 78 (BiP/GRP78), Erp99/GRP94 chaperone, and P450 2E1 (Martin, Reed, and Pohl, 1993; Pumford *et al.*, 1993; Bourdi *et al.*, 1996). Sevoflurane (2,2,2-trifluoro-1-[trifluoromethyl]ethyl fluoromethyl ether), which does not form reactive metabolites that lead to trifluoroacetylated-protein adducts and has a much less risk of hepatotoxicity, is used instead of halothane for inhalation anesthesia today (Stoelting, 1999).

### 4.4 Tienilic Acid

Tienilic acid (Fig. 5) is a potent diuretic that was introduced into the market in 1979. Several reports of immune-mediated hepatotoxicity, which was sometimes fatal,

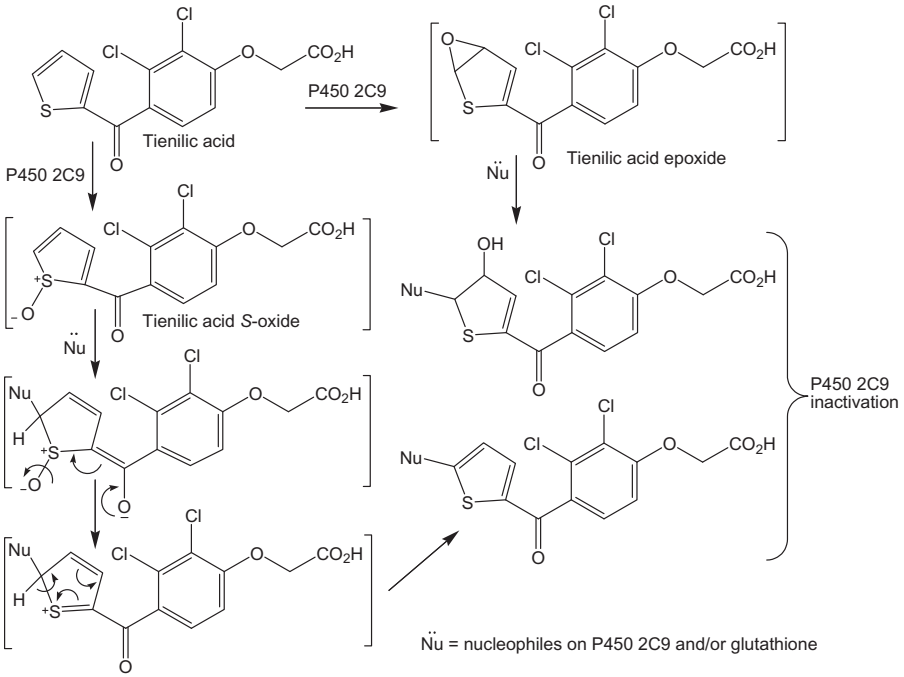


**Figure 5** Proposed chemical mechanism for the metabolic activation of tienilic acid to epoxide- and *S*-oxide-type reactive intermediates that undergo covalently binding reactions with P450 2C9 apoprotein (Koenigs *et al.*, 1999).

occurred and led to its rapid removal from the market in 1980 (Barclay, 1980; Zimmerman *et al.*, 1984; Dansette *et al.*, 1998; Fung *et al.*, 2001). Tienilic acid undergoes P450 2C9-mediated metabolic activation to epoxide- and *S*-oxide-type reactive intermediates (Fig. 5) that inactivate the P450 2C9 enzyme (López-Garcia, *et al.*, 1994), covalently bind to protein (Lecoeur *et al.*, 1994; Koenigs *et al.*, 1999), and react with glutathione (Fig. 5) (Valadon *et al.*, 1996). Anti-P450 2C9 antibodies have been detected in patients with tienilic-induced hepatotoxicity, which suggests that metabolic activation of the drug is related to the mechanism of hepatotoxicity (Beaune *et al.*, 1987; Neuberger and Williams, 1989; Mitchell *et al.*, 1990; Beaune *et al.*, 1994; Robin *et al.*, 1996).

Tienilic acid also contains a carboxylic acid structural alert (Chapter Chemical Mechanisms in Toxicology). As discussed later, carboxylic acid-containing drugs can be metabolized to reactive acyl-linked metabolites that are capable of undergoing covalent binding to protein. Tienilic acid is a phenoxyacetic acid derivative, and these types of carboxylic acids have been shown to lead to highly reactive acyl-coenzyme A (acyl-CoA) thioester-linked metabolites (Li, Grillo and Benet, 2003). However, the corresponding routes of metabolism for tienilic acid have not yet been investigated.

### 4.5 Sulfamethoxazole

Sulfonamide-containing antibiotic drugs are derivatives of *p*-aminobenzenesulfonamide (sulfanilamide; Mandell and Petri, 1996). These drugs inhibit the

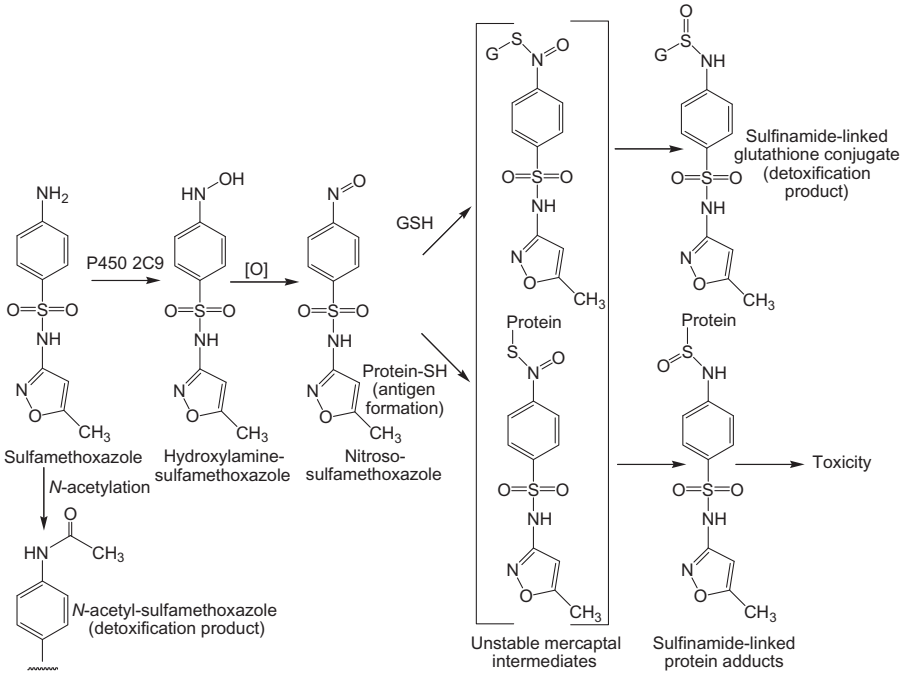


**Figure 6** Chemical mechanism for the P450-mediated bioactivation of sulfamethoxazole to nitroso-sulfamethoxazole followed by subsequent reactions with glutathione and protein nucleophiles (Pirmohamed, Madden and Park, 1996).

biosynthesis of folic acid in bacteria by competitively inhibiting dihydropteroate synthase, the enzyme that converts *p*-aminobenzoic acid into folic acid, a necessary cofactor (Mandell and Petri, 1996). The chemical structure of sulfonamides contains a primary aromatic amine that can be metabolized by oxidation to *N*-hydroxylation metabolites. These oxidized metabolites then can be oxidized further to chemically reactive nitrosoamine products, which are known to react with protein nucleophiles forming drug–protein adducts that serve as antigens for hypersensitivity reactions (Fig. 6). Many sulfonamide-type drugs cause untoward allergic reactions such as Stevens–Johnson syndrome and toxic epidermal necrolysis (Meekins, Sullivan and Gruchalla, 1994; Mandell and Petri, 1996; French, 2006).

The antibiotic drug sulfamethoxazole (4-amino-*N*-[5-methylisoxazol-3-yl]-benzenesulfonamide) is known to cause idiosyncratic hypersensitivity reactions which resemble immunopathologic reactions. These allergic reactions occur in 2% to 4% of normal individuals, but in as much as 60% in AIDS patients (Gruchalla and Sullivan, 1991). The allergic reaction is a classic idiosyncratic drug hypersensitivity that is rarely experimentally reproduced in animals. The toxicity has a delayed onset of 7–14 days after starting therapy. The clinical symptoms are fever, severe cutaneous skin reactions (Stevens–Johnson syndrome and toxic epidermal necrolysis), and multi-organ toxicity. The toxicity is rare, unpredictable, and potentially fatal. Sulfamethoxazole undergoes oxidative metabolism in dendritic cells by cyclooxygenase enzymes, in lymphocytes by myeloperoxidase enzymes, and in the liver (Cribb, Spielberg and Griffin, 1995) and skin (Reilly *et al.*, 2000) by cytochrome P450 2C9 yielding nitroso metabolites that react with

glutathione, or with proteins forming adducts leading to immune-mediated hypersensitivity reactions (Fig. 6). Direct evidence of an immune-mediated response to sulfamethoxazole comes from studies showing that IgE antibodies against cellular proteins or drug are present in the serum of hypersensitive patients having immediate-type allergic reaction to sulfamethoxazole (Carrington *et al.*, 1987; Meekins, Sullivan and Gruchalla, 1994). In addition, sulfamethoxazole has been shown to activate T cells in patients with previous sulfamethoxazole-induced hypersensitivity reactions (Pichler, 2002; Sanderson, Naisbitt and Park, 2006).

### 4.6 Carboxylic Acid-Containing Drugs

Increasing evidence suggests that many carboxylic acid-containing drugs are metabolized to chemically reactive intermediates that form covalent adducts with proteins (Faed, 1984; Boelsterli, 2002; Li, Benet and Grillo 2002a, b). Such adducts have been proposed to act as immune complexes leading to immunotoxic reactions (Spahn-Langguth and Benet, 1992; Zimmerman, 1994). Of the 121 drugs withdrawn from the worldwide market for hepatotoxic, anaphylactic, dermatologic, gastrointestinal, and nephrotoxic reactions that were sometimes fatal, 16 were carboxylic acid-containing drugs (Fung *et al.*, 2001), some of which are shown in Table 1. Many of these carboxylic acid-containing drugs are metabolized to two prospective reactive metabolites (Fig. 7), namely acyl glucuronides (Stogniew and Fenselau, 1982) and acyl-CoA thioesters (Caldwell, 1984; Boelsterli, 2002). Acyl glucuronides, which can be major metabolites of carboxylic acid-containing drugs, have been shown to covalently bind to protein by two alternative mechanisms. The first mechanism occurs when, because of the electrophilic nature of the carbonyl carbon of the 1-*O*-acyl linkage, the acyl glucuronide reacts with protein nucleophiles (e.g. lysinyl-$NH_2$, serine-OH, and cysteinyl-SH) in a nucleophilic displacement-type reaction leading to the transacylation of cellular proteins (Fig. 8). The second mechanism occurs when the 1-*O*-acyl glucuronide undergoes acyl migration, where the aglycone migrates from hydroxyl group to hydroxyl group around the glucuronic acid ring (Hasegawa, Smith and Benet, 1982). These acyl migration isomers then undergo ring-chain tautomerism leading to an open-chain aldehyde that is able to form Schiff-base adducts with protein-lysinyl residues. These Schiff-base-linked adducts are stabilized by undergoing an Amadori rearrangement that yields stable 1-amino-2-keto-linked acyl glucuronide protein adducts (Smith, Benet and McDonagh, 1990; Ding *et al.*, 1993) (Fig. 9). It is important to mention here that some of the carboxylic acid-containing drugs shown in Table 1 may also be metabolized by cytochrome P450 to chemically reactive metabolites. For example, zomepirac, in addition to being metabolized to acyl glucuronide and acyl-CoA reactive metabolites, has been shown recently to be bioactivated on the pyrrole structural moiety leading to a reactive epoxide metabolite (Chapter Chemical Mechanisms in Toxicology, Section 5.9 of Chemical Mechanisms in Toxicology, Fig. 16 of Chemical Mechanisms in Toxicology) (Chen *et al.*, 2006). Another example is for the drug alclofenac, which contains an alkene group that has been shown to form a reactive epoxide metabolite as well (Brown and Ford-Hutchinson, 1982). A final example is the NSAID bromfenac which was on the market in the United States for about nine months when there were many cases of liver damage and sudden death, before being pulled from the market in 1998 (Fontana *et al.*, 1999). The chemical structure of bromfenac contains at least two other structural alerts, namely the parachlorophenyl group and the aniline moiety, in terms of the potential for the formation of chemically reactive metabolites that have been associated with toxicological problems and liabilities (Nelson, 1995).

**TABLE 1 Partial list of carboxylic-acid containing drugs withdrawn from clinical use and/or also associated with allergic reactions.**

| Withdrawn | Known to cause allergic reactions |
|---|---|
| 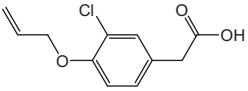 Alclofenac | 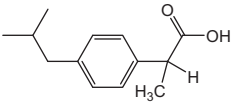 Ibuprofen |
| 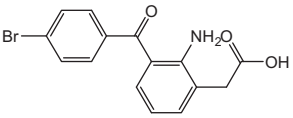 Bromfenac | 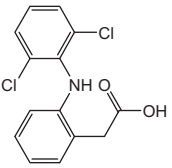 Diclofenac |
| 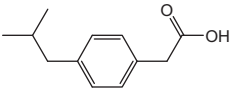 Ibufenac | 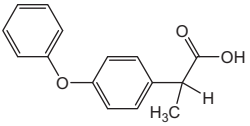 Fenoprofen |
| 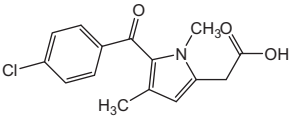 Zomepirac | 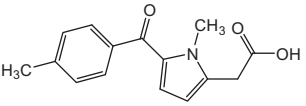 Tolmetin |
| 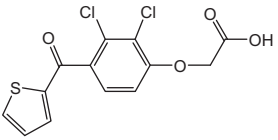 Tienilic acid | 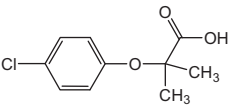 Clofibric acid |
| 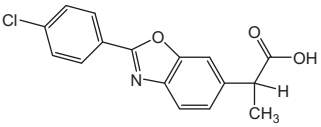 Benoxaprofen | 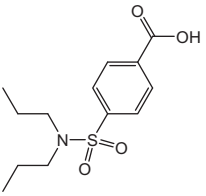 Probenecid |
| 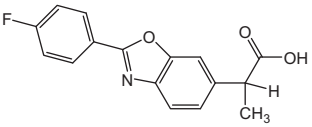 Flunoxaprofen | 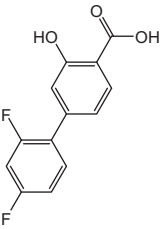 Diflunisal |
| 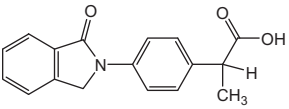 Indoprofen | 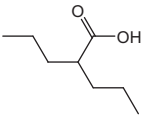 Valproic acid |

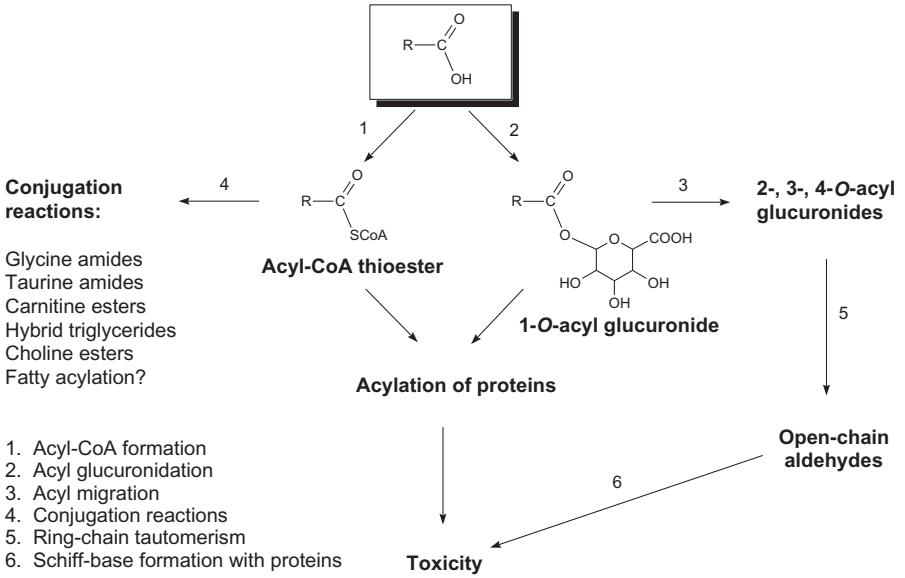


**Figure 7** Proposed scheme for the metabolic activation of carboxylic acid-containing drugs to reactive, and hence potentially toxic, intermediates, and to nontoxic metabolites.

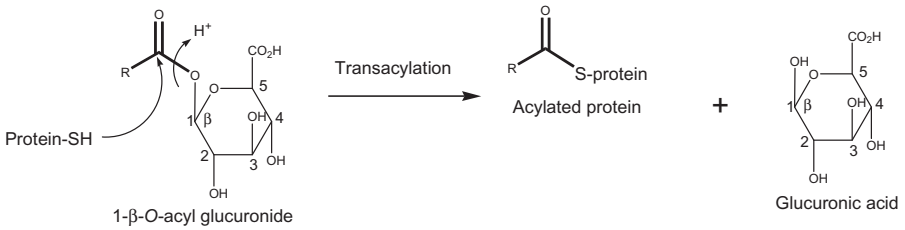


**Figure 8** Mechanism for transacylation of protein nucleophiles by 1-β-*O*-acyl-linked glucuronides.

Thioester-linked acyl-CoA intermediary metabolites of carboxylic acid-containing drugs possess an electrophilic carbonyl carbon at the thioester linkage (Huxtable, 1986). Because of their large size and amphipathic nature, xenobiotic acyl-CoA thioesters do not diffuse out of the cell, but are substrates for further metabolism such as amino acid conjugation, carnitine ester formation, and hybrid triglyceride formation, all of which occur by transacylation-type reactions, albeit enzyme-mediated, which require a reactive thioester carbonyl carbon. Xenobiotic acyl-CoA thioesters are reactive species that have been shown to transacylate endogenous nucleophiles. For example, salicylic acid acyl-CoA thioester is able to react with glycine nonenzymatically *in vitro*, leading to the formation of a salicyluric glycine amide conjugate (Tishler and Goldman, 1970). Other studies have shown that endogenous long-chain acyl-CoA derivatives such as palmitoyl-CoA, arachidonyl-CoA, and retinoyl-CoA react in an "autoacylation" fashion with the cysteinyl-sulfhydryls of proteins and peptides (Yamashita *et al.*, 1995; Duncan and Gilman, 1996; Kubo *et al.*, 2005). Recent examples

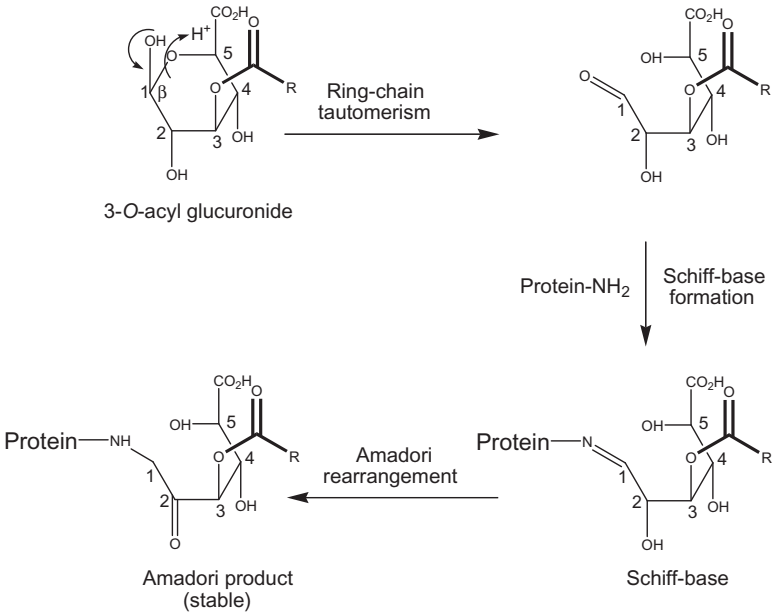


**Figure 9** Mechanism for stable adduct formation with lysine residues on proteins by acyl glucuronide migration isomers.

of reactive acyl-CoA thioesters include those formed from the endogenous bile acid cholic acid (Mano *et al.*, 2001), the lipid-lowering drug clofibric acid (Grillo and Benet, 2002), the NSAIDs tolmetin (Olsen *et al.*, 2005) and zomepirac (Olsen *et al.*, 2007), and nafenopin (Sallustio *et al.*, 2000). The prediction of acyl glucuronides (Benet *et al.*, 1993; Bolze *et al.*, 2002) and acyl-CoA thioester (Sidenius *et al.*, 2004) metabolites for their ability to covalently bind to protein based on their chemical structures has been shown. In general, the chemical reactivity of acyl glucuronides and acyl-CoA thioester derivatives with biological nucleophiles, such as glutathione and human serum albumin, decreases with increasing substitution at the carbon adjacent to the carbonyl carbon of the carboxylic acid moiety.

An extensively studied example of a drug known to cause idiosyncratic toxicities is diclofenac (Boelsterli, 2003; Tang, 2003). Diclofenac (2-[2-(2,6-dichlorophenyl)-aminophenyl]ethanoic acid; see Chapter 23, Fig. 23.21) is an NSAID used for the treatment of osteoarthritis, rheumatoid arthritis, ankylosing spondylitis, and acute muscle pain (Small, 1989). Adverse reactions to diclofenac include enteropathy, which is common during prolonged use of the drug, acute and chronic hepatitis (15% of patients have increased plasma transaminase levels), hemolytic anemia (autoantibodies have been detected), and fatal anaphylaxis which is related to an IgE- or IgG-mediated immune response. One molecular mechanism that is proposed to contribute to diclofenac-induced hepatitis includes the recognition of plasma membrane drug–protein adducts by the immune system, leading to an immunotoxic response. Another mechanism is when the cytotoxic reaction leads to the release of reactive metabolite–covalently bound protein adducts and or immune "danger" signals (Uetrecht, 1999) (Fig. 1). The cytotoxicity is proposed to be caused by mitochondrial damage by the drug, reactive diclofenac metabolites, or by oxidative stress mediated by free radical-

type metabolites (Masubuchi, Yamada and Horie, 2000; Galati *et al.*, 2002). Reactive quinone imine metabolites formed by P450, for example, can lead to the formation of a P450 2C11 adduct with diclofenac (Tang *et al.*, 1999; Masubuchi, Ose and Horie, 2001). In addition, diclofenac acyl glucuronide reacts with protein, forming adducts with dipeptidyl peptidase IV (110 kDa) as well as with unidentified plasma membrane proteins that weigh 140 and 200 kDa (Kretz-Rommel and Boelsterli, 1994; Hargus *et al.*, 1995). Results from one study showed the formation of an IgM autoantibody to the acyl glucuronide conjugate of the 4′-hydroxydiclofenac occurring in a patient dosed with diclofenac (Bougie *et al.*, 1997). The glucuronide metabolite was shown also to promote agglutination of normal red blood cells which was proposed to potentially lead to immune hemolytic anemia.

Diclofenac-induced small intestinal injury has been proposed to occur by enterocyte cytotoxicity mediated by the reactive acyl glucuronide of diclofenac, or the acyl glucuronide of one of its oxidative metabolites (Seitz and Boelsterli, 1998). In another study, diclofenac was administered to rats by gastric gavage, and immunoblot analysis of small intestine homogenates and isolated enterocyte subcellular fractions with drug-specific antiserum revealed 142-, 130-, 110-, and 55-kDa protein adducts mediated by diclofenac (Ware *et al.*, 1998). The 142- and 130-kDa protein adducts of diclofenac were identified as aminopeptidase N (CD13) and sucrase-isomaltase, respectively. These authors proposed that oral tolerance to intestinal drug–protein adducts can be formed in gut-associated lymphoid tissue, leading to the down-regulation of drug-induced allergic reactions in many individuals that do not have gut toxicity to diclofenac.

An example of an NSAID carboxylic acid-containing drug removed from the market due to toxicity is benoxaprofen (Table 1). Benoxaprofen (2-[2-(4-chlorophenyl)-1,3-benzoxazol-5-yl]propanoic acid; Fig. 10) had been shown to cause a rare hepatotoxicity which led to its withdrawal from the market in 1982. The proposal was that benoxaprofen-induced toxicity was related to the formation of covalent adducts to proteins via a reactive acyl glucuronide metabolite (Fig. 10). In covalent binding studies in rat, benoxaprofen was shown to become irreversibly bound to plasma and liver proteins. An immunochemical approach was used to detect hepatic protein targets and showed adduct masses of 110 and 70 kDa as the major liver protein targets modified by benoxaprofen (Dong, Liu and Smith, 2005). In other studies, tandem mass spectrometry was used to identify the structures and specific covalent binding sites for protein adducts formed upon the reaction of benoxaprofen acyl glucuronide and human serum albumin *in vitro* (Qiu, Burlingame and Benet, 1998). Results showed that benoxaprofen glucuronide forms covalent adducts with protein nucleophiles on human serum albumin by both nucleophilic displacement (transacylation) and Schiff-base formation with acyl migration isomers with ε-amino-lysyl residues, where Lys-159 was determined to be the major covalent binding site. Similar investigations with the acyl glucuronide metabolite of the NSAID tolmetin (Table 1) showed covalent binding to human serum albumin by both transacylation- and Schiff-base-type mechanisms; however, in this case Lys-199 was the major site of protein adduct formation (Ding *et al.*, 1995). Although the acyl-CoA thioester intermediary metabolite of benoxaprofen has not been identified, its formation has been shown indirectly by the detection of the corresponding taurine conjugate (Mohri, Okada and Benet, 2005). Therefore, benoxaprofen-*S*-acyl-CoA may also contribute to covalent binding to hepatic proteins *in vivo* that leads to an immune response and subsequent liver injury (Boelsterli, 2002). For benoxaprofen, the benzoxazole substructure might be metabolized to a reactive intermediate that could explain some of the covalent binding to hepatic protein; however,

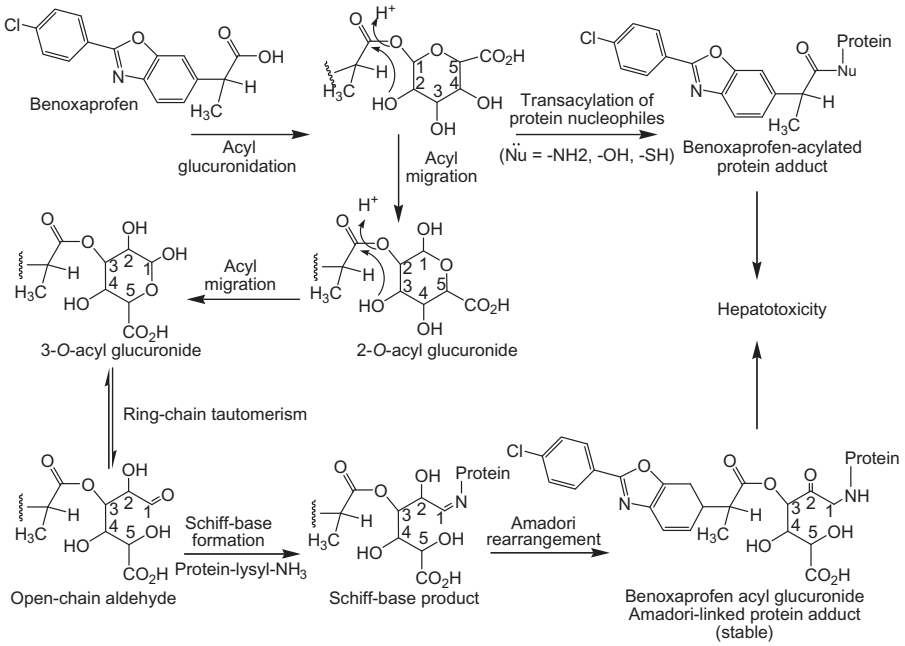


**Figure 10** Proposed chemical mechanism for the metabolic activation of benoxaprofen to reactive acyl glucuronide metabolites that undergo covalent binding to protein (Qiu, Burlingame and Benet, 1998; Dong, Liu and Smith, 2005).

no work has been done with this agent in this regard because the drug was not made readily available for study.

In summary, many carboxylic acid-containing drugs (Table 1), including the discontinued drugs ibufenac (Castillo and Smith, 1995), zomepirac (Smith, McDonagh and Benet, 1986), and suprofen (Smith and Liu, 1995), and currently used drugs such as diclofenac (Boelsterli, 2003), diflunisal (McKinnon and Dickinson, 1989), valproic acid (Williams *et al.*, 1992), tolmetin (Zia-Amirhosseini *et al.*, 1995), clofibric acid (Sallustio *et al.*, 1991), and salicylic acid (Dickinson, Baker and King, 1994) form reactive acyl-linked metabolites that lead to covalent binding to protein and may be responsible for mediating hypersensitivity reactions related to their use (Fung *et al.*, 2001).

## 5 CONCLUSIONS

Pharmaceutical and academic scientists have been trying to put together a picture of the phenomenon of drug-induced allergic reactions in terms of what has happened in the pharmaceutical industry over the past 40 or so years. Things got started with allergic reactions to drugs like sulfonamides and penicillins, but which are now used to understand allergic reactions related to the hapten hypothesis. Pharmaceutical scientists are continuously searching for ways to design drugs that do not cause allergic reactions and to be able to identify those that do at an early stage in drug discovery in order to have significant impact on candidate drug selection. As pharmaceutical scientists from

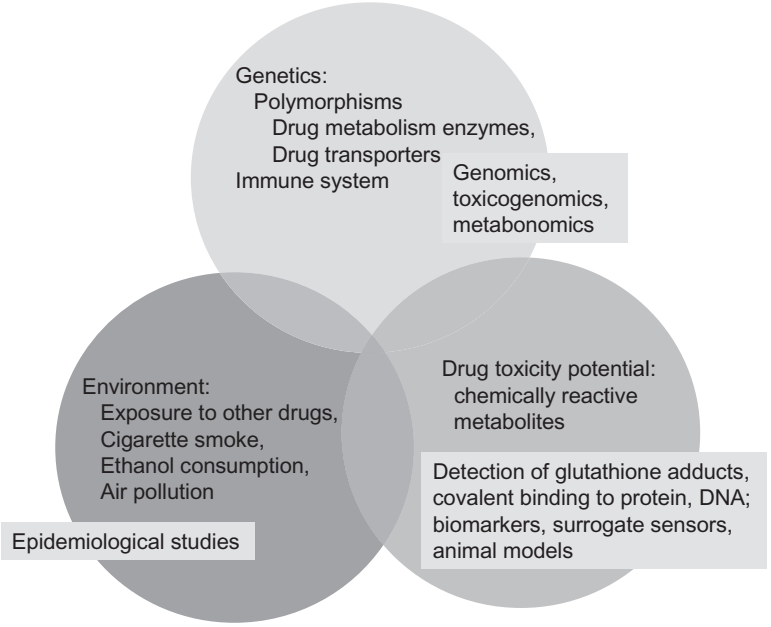


**Figure 11** Risk of allergic reactions to drugs (circles) and corresponding areas of ongoing research (squares).

drug metabolism, biology, toxicology, and immunology disciplines dig deeper and deeper, they have realized that there is a lot more work to do. Such research includes the generation of predictive animal models of drug hypersensitivity reactions, the application of genomics and toxicogenomics related to drug allergic reactions, the increased study of the influence of environmental factors (such as other drugs, toxins, or disease) on allergic diseases, and on the role of reactive drug metabolites and covalent binding to protein on subsequent immunotoxicity (Fig. 11).

In the future, animal models should become very important for toxicological research into drug allergic reactions, because these models are essentially the only way to study the currently available hypotheses (Dansette *et al.*, 1998; Uetrecht, 2006). However, useful animal models for allergic reactions are very few because allergic reactions to drugs are not predictable in animals either (Shenton, Chen and Uetrecht, 2004). Some examples of drugs that have been studied in animal models for drug-induced immune reactions are penicillamine, halothane, isoniazid, and sulfamethoxazole. Penicillamine, a drug sometimes used for the treatment of rheumatoid arthritis, can cause autoimmune diseases such as lupus-like syndrome and skin rash (Stein *et al.*, 1980) and can also induce autoimmunity in Brown Norway rats (Mason and Uetrecht, 2004). Halothane has been shown to demonstrate a cellular immune response in halothane-exposed guinea pigs (Furst *et al.*, 1997). Isoniazid is a drug known to cause a high incidence of hypersensitivity reactions leading to liver injury (Zimmerman, 1993). An animal model for isoniazid has been reported to be the rabbit, where treatment with isoniazid led to elevation in transaminase enzyme levels and the formation of liver necrosis in some test animals (Sarich *et al.*, 1995). Sulfonamide-type antimicrobials, such as sulfamethoxazole, have been shown to lead to drug hypersensitivity

reactions in dogs, especially in larger dogs such as Doberman pinschers (Lavergne *et al.*, 2006). Increased numbers of animal models for drug hypersensitivity reactions are necessary in order to speed up the progress toward understanding the mechanisms of drug-induced allergies (Uetrecht, 2006). An increased understanding of how and why allergic reactions to drugs occur should assist in improving decision-making processes for the selection of safer drug candidates. Hopefully, in the future, this greater and greater understanding will lead to fewer and fewer drugs being developed that might someday cause allergic reactions in patients.

## ACKNOWLEDGMENTS

I would like to thank Dr. Lixia Jin and Judy Huang, B.A. (Department of Pharmacokinetics and Drug Metabolism, Amgen Inc.), Dr. Kathila Rajapaksa (Toxicology, Amgen Inc.), Dr. Joseph A. Ware (Late Stage Pharmacokinetics and Pharmacodynamics, Genentech, South San Francisco), and Dr. Christian Skonberg (Department of Pharmaceutics and Analytical Chemistry, University of Copenhagen, Denmark) for critically reviewing this chapter. I also thank the constructive criticism and helpful suggestions of an anonymous reviewer, which has been included in the references.

## REFERENCES

Baillie, T.A. (2006). Future of toxicology-metabolic activation and drug design: challenges and opportunities in chemical toxicology. *Chem. Res. Toxicol.*, *19*, 889–893.

Baldo, B. (1999). Penicillins and cephalosporins as allergens-structural aspects of recognition and cross-reactions. *Clin. Exp. Allergy.*, *29*, 744–749.

Barclay, W.R. (1980). Ticrynafen's withdrawal from the market. *JAMA*, *243*, 771.

Beaune, P., Dansette, P.M., Mansuy, D., Kiffel, L., Finck, M., Amar, C., Leroux, J.P., Homber, J.C. (1987). Human anti-endoplasmic reticulum autoantibodies appearing in a drug-induced hepatitis are directed against a human liver cytochrome P-450 that hydroxylates the drug. *Proc. Natl. Acad. Sci. U.S.A.*, *84*, 551–555.

Beaune, P., Pessayre, D., Dansette, P.M., Mansuy, D., Manns, M. (1994). Autoantibodies against cytochromes P450: role in human diseases. *Adv. Pharmacol.*, *30*, 199–245.

Benet, L.Z., Spahn-Langguth, H., Iwakawa, S., Volland, C., Mizuma, T., Mayer, S., Mutschler, E., Lin, E.T. (1993). Predictability of the covalent binding of acidic drugs in man. *Life Sci.*, *53*, 141–146.

Bialer, M. (2006). New antiepileptic drugs that are second generation to existing antiepileptic drugs. *Expert Opin. Investig. Drugs*, *15*, 637–647.

Boelsterli, U.A. (2002). Xenobiotic acyl glucuronides and acyl CoA thioesters as protein-reactive metabolites with the potential to cause idiosyncratic drug reactions. *Curr. Drug Metab.*, *3*, 439–450.

Boelsterli, U.A. (2003). Diclofenac-induced liver injury: a paradigm of idiosyncratic drug toxicity. *Toxicol. Appl. Pharmacol.*, *192*, 307–322.

Bolze, S., Bromet, N., Gay-Feutry, C., Massiere, F., Boulieu, R., Hulot, T. (2002). Development of an in vitro screening model for the biosynthesis of acyl glucuronide metabolites and the assessment of their reactivity toward human serum albumin. *Drug Metab. Dispos.*, *30*, 404–413.

Bougie, D., Johnson, S.T., Weitekamp, L.A., Aster, R.H. (1997). Sensitivity to a metabolite of diclofenac as a cause of acute immune hemolytic anemia. *Blood*, *90*, 407–413.

Bourdi, M., Chen, W., Peter, R.M., Martin, J.L., Buters, J.T., Nelson, S.D., Pohl, L.R. (1996). Human cytochrome P450 2E1 is a major autoantigen associated with halothane hepatitis. *Chem. Res. Toxicol.*, *9*, 1159–1166.

Brown, L.M., Ford-Hutchinson, A.W. (1982). The destruction of cytochrome P-450 by alclofenac: possible involvement of an epoxide metabolite. *Biochem. Pharmacol.*, *31*, 195–199.

Caldwell, J. (1984). Xenobiotic acyl-coenzyme A: critical intermediates in the biochemical pharmacology and toxicology of carboxylic acids. *Biochem. Soc. Trans.*, *12*, 9–11.

Carrington, D., Earl, H., Sullivan, T. (1987). Studies of human IgE to a sulfonamide determinant. *J. Allergy. Clin. Immunol.*, *79*, 442–447.

Castillo, M., Smith, P.C. (1995). Disposition and reactivity of ibuprofen and ibufenac acyl glucuronides in vivo in rhesus monkey and in vitro with human serum albumin. *Drug Metab. Dispos.*, *23*, 566–572.

Chen, Q., Doss, G.A., Tung, E.C., Liu, W., Tang, Y.S., Braun, M.P., Didolka, V., Strauss, J.R., Wang, R.W., Stearns, R.A., Evans, D.C., Baillie, T.A., and Tang, W. (2006). Evidence for the bioactivation of zomepirac and tolmetin by an oxidative pathway: identification of glutathione adducts in vitro in human liver microsomes and in vivo in rats. *Drug Metab. Dispos.*, *34*, 145–151.

Cribb, A.E., Spielberg, S.P., Griffin, G.P. (1995). N4-hydroxylation of sulfamethoxazole by cytochrome P450 of the cytochrome P4502C subfamily and reduction of sulfamethoxazole hydroxylamine in human and rat hepatic microsomes. *Drug Metab. Dispos.*, *23*, 406–414.

Dansette, P.M., Bonierbale, E., Minoletti, C., Beaune, P.H., Pessayre, D., Mansuy, D. (1998). Drug-induced immunotoxicity. *Eur. J. Drug Metab. Pharmacokinet.*, *23*, 443–451.

Davies, G.E. (1973). Halothane hepatotoxicity: is halothane hepatotoxicity an allergic reaction to halothane. *Proc. Roy. Soc. Med.*, *66*, 1–2.

Dickinson, R.G., Baker, P.V., King, A.R. (1994). Studies on the reactivity of acyl glucuronides—VII. Salicyl acyl glucuronide reactivity in vitro and covalent binding of salicylic acid to plasma protein of humans taking aspirin. *Biochem. Pharmacol.*, *47*, 469–476.

Dieckhaus, C.M., Miller, T.A., Sofia, R.D., Macdonald, T.L. (2000). A mechanistic approach to understanding species differences in felbamate bioactivation: relevance to drug-induced idiosyncratic reactions. *Drug Metab. Dispos.*, *28*, 814–822.

Dieckhaus, C.M., Thompson, C.D., Roller, S.G., Macdonald, T.L. (2002). Mechanisms of idiosyncratic drug reactions: the case of felbamate. *Chem. Biol. Interact.*, *142*, 99–117.

Ding, A., Ojingwa, J.C., McDonagh, A.F., Burlingame, A.L., Benet, L.Z. (1993). Evidence for covalent binding of acyl glucuronides to serum albumin via an imine mechanism as revealed by tandem mass spectrometry. *Proc. Natl. Acad. Sci. U.S.A.*, *90*, 3797–3801.

Ding, A., Zia-Amirhosseini, P., McDonagh, A.F., Burlingame, A.L., Benet, L.Z. (1995). Reactivity of tolmetin glucuronide with human serum albumin. Identification of binding sites and mechanisms of reaction by tandem mass spectrometry. *Drug Metab. Dispos.*, *23*, 369–376.

Dong, J.Q., Liu, J., Smith, P.C. (2005). Role of benoxaprofen and flunoxaprofen acyl glucuronides in covalent binding to rat plasma and liver proteins in vivo. *Biochem. Pharmacol.*, *70*, 937–948.

Duncan, J.A., Gilman, A.G. (1996). Autoacylation of G protein alpha subunits. *J. Biol. Chem.*, *271*, 23594–23600.

van Dyke, R.A., Gandolfi, A.J. (1974). Studies on irreversible binding of radioactivity from (14C)halothane to rat hepatic microsomal lipids and protein. *Drug Metab. Dispos.*, *2*, 469–476.

Ewan, P.W. (1998). ABC of allergies: anaphylaxis. *BMJ*, *316*, 1442–1445.

Faed, E.M. (1984). Properties of acyl glucuronides: implications for studies of the pharmacokinetics and metabolism of acidic drugs. *Drug Metab. Rev.*, *15*, 1213–1249.

Fleming, A. (1929). On the antibacterial action of cultures of a penicillium, with special reference to their use in the isolation of *B. influenzae*. *Br. J. Pathol.*, *10*, 226–236.

Fontana, R.J., McCashland, T.M., Benner, K.G., Appelman, H.D., Gunartanam, N.T., Wisecarver, J.L., Rabkin, J.M., Lee, W.M. (1999). Acute liver failure associated with prolonged use of bromfenac leading to liver transplantation. The Acute Liver Failure Study Group. *Liver Transpl. Surg.*, *5*, 480–484.

French, L.E. (2006). Toxic epidermal necrolysis and Stevens Johnson syndrome: our current understanding. *Allergol. Int.*, *55*, 9–16.

Fung, M., Thornton, A., Mybeck, K., Wu, J.H., Hornbuckle, K., Muniz, E. (2001). Evaluation of the characteristics of safety withdrawal of prescription drugs from world wide pharmaceutical markets—1960 to 1999. *Drug Inform. J.*, *35*, 293–317.

Furst, S.M., Luedke, D., Gaw, H.H., Reich, R., Gandolfi, A.J. (1997). Demonstration of a cellular immune response in halothane-exposed guinea pigs. *Toxicol. Appl. Pharmacol.*, *143*, 245–255.

Galati, G., Tafazoli, S., Sabzevari, O., Chan, T.S., O'Brien, P.J. (2002). Idiosyncratic NSAID drug induced oxidative stress. *Chem. Biol. Interact.*, *142*, 25–41.

Gandolfi, A.J., White, R.D., Sipes, I.G., Pohl, L.R. (1980). Bioactivation and covalent binding of halothane in vitro: studies with [3H]- and [14C]halothane. *J. Pharmacol. Exp. Ther.*, *214*, 721–725.

Goldsby, R.A., Kindt, T.K., Osborne, B.A., Kubym, J. (2003). *Immunology*, 5th edn, W.H. Freeman and Company, New York.

Grillo, M.P., Benet, L.Z. (2002). Studies on the reactivity of clofibryl-S-acyl-CoA thioester with glutathione in vitro. *Drug Metab. Dispos.*, *30*, 55–62.

Gruchalla, R.S. (2003). 10. Drug allergy. *J. Allergy Clin. Immunol.*, *111*, S548–559.

Gruchalla, R.S., Sullivan, T.J. (1991). Detection of human IgE to sulfamethoxazole by skin testing with sulfamethoxazoyl-poly-L-tyrosine. *J. Allergy Clin. Immunol.*, *88*, 784–792.

Guengerich, F.P., MacDonald, J.S. (2007). Applying mechanisms of chemical toxicity to predict drug safety. *Chem. Res. Toxicol.*, *20*, 344–369.

Hargus, S.J., Martin, B.M., George, J.W., Pohl, L.R. (1995). Covalent modification of rat liver dipeptidyl peptidase IV (CD26) by the nonsteroidal anti-inflammatory drug diclofenac. *Chem. Res. Toxicol.*, *8*, 993–996.

Hasegawa, J., Smith, P.C., Benet, L.Z. (1982). Apparent intramolecular acyl migration of zomepirac glucuronide. *Drug Metab. Dispos.*, *10*, 469–473.

Hovinga, C.A. (2002). Novel anticonvulsant medications in development. *Expert Opin. Investig. Drugs*, *11*, 1387–1406.

Huxtable, R. (1986). Thiols, disulfides and thioesters, in *Biochemistry of Sulfur* (ed. E. Frieden), Plenum Press, New York, 230–245.

Janeway, C., Travers, P., Walport, M., Shlomchik, M. (2005). *Immunobiology*, 6th edn, Garland Science. New York.

Josephson, A.S. (2004). Penicillin allergy: a public health perspective. *J. Allergy Clin. Immunol.*, *113*, 605–606.

Ju, C., Uetrecht, J.P. (2002). Mechanism of idiosyncratic drug reactions: reactive metabolite formation, protein binding and the regulation of the immune system. *Curr. Drug Metab.*, *3*, 367–377.

Kalish, R.S. (1995). Antigen processing: the gateway to the immune response. *J. Am. Acad. Dermatol.*, *31*, 640–652.

Kapetanovic, I.M., Torchin, C.D., Strong, J.M., Yonekawa, W.D., Lu, C., Li, A.P., Dieckhaus, C.M., Santos, W.L., Macdonald, T.L., Sofia, R.D., Kupferberg, H.J. (2002). Reactivity of atropaldehyde, a felbamate metabolite in human liver tissue in vitro. *Chem. Biol. Interact*, *142*, 119–134.

Kaplowitz, N. (2005). Idiosyncratic drug hepatotoxicity. *Nat. Rev. Drug Discov.*, *4*, 489–499.

Kenna, J.G., Satoh, H., Christ, D.D., Pohl, L.R. (1988). Metabolic basis for a drug hypersensitivity: antibodies in sera from patients with halothane hepatitis recognize liver neoantigens that

contain the trifluoroacetyl group derived from halothane. *J. Pharmacol. Exp. Ther.*, *245*, 1103–1109.

Koenigs, L.L., Peter, R.M., Hunter, A.P., Haining, R.L., Rettie, A.E., Friedberg, T., Pritchard, M.P., Shou, M., Rushmore, T.H., Trager, W.F. (1999). Electrospray ionization mass spectrometric analysis of intact cytochrome P450: identification of tienilic acid adducts to P450 2C9. *Biochemistry*, *38*, 2312–2319.

Kretz-Rommel, A., Boelsterli, U.A. (1994). Mechanism of covalent adduct formation of diclofenac to rat hepatic microsomal proteins. Retention of the glucuronic acid moiety in the adduct. *Drug Metab. Dispos.*, *22*, 956–961.

Kubo, Y., Wada, M., Ohba, T., Takahashi, N. (2005). Formation of retinoylated proteins from retinoyl-CoA in rat tissues. *J. Biochem. (Tokyo)*, *138*, 493–500.

Kumar, S., Kassahun, K., Tschirret-Guth, R.A., Mitra, K., Baillie, T.A. (2008). Minimizing metabolic activation during pharmaceutical lead optimization: progress, knowledge gaps and future directions. *Curr. Opin. Drug Discov. Devel.*, *11*, 43–52.

Landsteiner, K., Jacobs, J. (1935). Studies on the sensitization of animals with simple chemical compounds. *J. Exp. Med.*, *61*, 643–656.

Lavergne, S.N., Danhof, R.S., Volkman, E.M., Trepanier, L.A. (2006). Association of drug-serum protein adducts and anti-drug antibodies in dogs with sulfonamide hypersensitivity: a naturally occurring model of idiosyncratic drug toxicity. *Clin. Exp. Allergy*, *36*, 907–915.

Lecoeur, S., Bonierbale, E., Challine, D., Gautier, J.C., Valadon, P., Dansette, P.M., Catinot, R., Ballet, F., Mansuy, D., Beaune, P.H. (1994). Specificity of in vitro covalent binding of tienilic acid metabolites to human liver microsomes in relationship to the type of hepatotoxicity: comparison with two directly hepatotoxic drugs. *Chem. Res. Toxicol.*, *7*, 434–442.

Leone, A.M., Kao, L.M., McMillian, M.K., Nie, A.Y., Parker, J.B., Kelley, M.F., Usuki, E., Parkinson, A., Lord, P.G., Johnson, M.D. (2007). Evaluation of felbamate and other antiepileptic drug toxicity potential based on hepatic protein covalent binding and gene expression. *Chem. Res. Toxicol.*, *20*, 600–608.

Levine, B. (1966). Immunologic mechanisms of penicillin allergy. A haptenic model system for the study of allergic disease in man. *N. Engl. J. Med.*, *275*, 1115–1125.

Levine, B., Redmond, A. (1969). Minor haptenic determinant-specific regains of penicillin hypersensitivity in man. *Int. Arch. Allergy Appl. Immunol.*, *35*, 445–455.

Li, C., Benet, L.Z., Grillo, M.P. (2002a). Studies on the chemical reactivity of 2-phenylpropionic acid 1-O-acyl glucuronide and S-acyl-CoA thioester metabolites. *Chem. Res. Toxicol.*, *15*, 1309–1317.

Li, C., Benet, L.Z., Grillo, M.P. (2002b). Enantioselective covalent binding of 2-phenylpropionic Acid to protein in vitro in rat hepatocytes. *Chem. Res. Toxicol.*, *15*, 1480–1487.

Li, C., Grillo, M.P., Benet, L.Z. (2003). In vitro studies on the chemical reactivity of 2,4-dichlorophenoxyacetyl-S-acyl-CoA thioester. *Toxicol. Appl. Pharmacol.*, *187*, 101–109.

López-Garcia, M.P., Dansette, P.M., Mansuy, D. (1994). Thiophene derivatives as new mechanism-based inhibitors of cytochromes P-450: inactivation of yeast-expressed human liver cytochrome P-450 2C9 by tienilic acid. *Biochemistry*, *33*, 166–175.

McKinnon, G.E., Dickinson, R.G. (1989). Covalent binding of diflunisal and probenecid to plasma protein in humans: persistence of the adducts in the circulation. *Res. Commun. Chem. Pathol. Pharmacol.*, *66*, 339–354.

McMillian, M., Nie, A.Y., Parker, J.B., Leone, A., Bryant, S., Kemmerer, M., Herlich, J., Liu, Y., Yieh, L., Bittner, A., Liu, X., Wan, J., Johnson, M.D. (2004). A gene expression signature for oxidant stress/reactive metabolites in rat liver. *Biochem. Pharmacol.*, *68*, 2249–2261.

Maindrault-Goebel, F., André, T., Tournigand, C., Louvet, C., Perez-Staub, N., Zeghib, N., de Gramont, A. (2005). Allergic-type reactions to oxaliplatin: retrospective analysis of 42 patients. *Eur. J. Cancer*, *41*, 2262–2267.

Mandell, G.L., Petri, W.A. (1996). Antimicrobial agents, in *Goodman and Gilman's the Pharmacological Basis of Therapeutics*, 9th edn (eds J.G. Hardman, A.G. Gilman and L.E. Limbard), McGraw-Hill, New York, pp. 1073–1101.

Mano, N., Uchida, M., Okuyama, H., Sasaki, I., Ikegawa, S., Goto, J. (2001). Simultaneous detection of cholyl adenylate and coenzyme A thioester utilizing liquid chromatography/electrospray ionization mass spectrometry. *Anal. Sci.*, *17*, 1037–1042.

Markel, H. (2004). Shaping the mold, from lab glitch to lifesaver. The New York Times (Science Times), April 20, 2004, p. D6.

Martin, J.L., Reed, G.F., Pohl, L.R. (1993). Association of anti-58 kDa endoplasmic reticulum antibodies with halothane hepatitis. *Biochem. Pharmacol.*, *46*, 1247–1250.

Mason, M.J., Uetrecht, J.P. (2004). Tolerance induced by low dose D-penicillamine in the Brown Norway rat model of drug-induced autoimmunity is immune-mediated. *Chem. Res. Toxicol.*, *17*, 82–94.

Masubuchi, Y., Ose, A., Horie, T. (2001). Mechanism-based inactivation of CYP2C11 by diclofenac. *Drug Metab. Dispos.*, *29*, 1190–1195.

Masubuchi, Y., Yamada, S., Horie, T. (2000). Possible mechanism of hepatocyte injury induced by diphenylamine and its structurally related nonsteroidal anti-inflammatory drugs. *J. Pharmacol. Exp. Ther.*, *292*, 982–987.

Matzinger, P. (1994). Tolerance, danger and the extended family. *Annu. Rev. Immunol.*, *12*, 991–1045.

Meekins, C.V., Sullivan, T.J., Gruchalla, R.S. (1994). Immunochemical analysis of sulfonamide drug allergy: identification of sulfamethoxazole-substituted human serum proteins. *J. Allergy Clin. Immunol.*, *94*, 1017–1024.

Mitchell, J.A., Gillam, E.M., Stanley, L.A., Sim, E. (1990). Immunotoxic side-effects of drug therapy. *Drug Saf.*, *5*, 168–178.

Mohri, K., Okada, K., Benet, L.Z. (2005). Stereoselective taurine conjugation of (R)-benoxaprofen enantiomer in rats: in vivo and in vitro studies using rat hepatic mitochondria and microsomes. *Pharm. Res.*, *22*, 79–85.

Naisbitt, D.J., Farrell, J., Wong, G., Depta, J.P., Dodd, C.C., Hopkins, J.E., Gibney, C.A., Chadwick, D.W., Pichler, W.J., Pirmohamed, M., Park, B.K. (2003). Characterization of drug-specific T cells in lamotrigine hypersensitivity. *J. Allergy Clin. Immunol.*, *111*, 1393–1403.

National Halothane Study (1966). Summary of the National Halothane Study. Possible association between halothane anesthesia and post-operative hepatic necrosis. *JAMA*, *197*, 775–788.

Nelson, S.D. (1995). Mechanisms of the formation and disposition of reactive metabolites that can cause acute liver injury. *Drug Metab. Rev.*, *27*, 147–177.

Neuberger, J., Williams, R. (1989). Immune mechanisms in tienilic acid associated hepatotoxicity. *Gut*, *30*, 515–519.

Neugut, A., Ghatak, A., Miller, R. (2001). Anaphylaxis in the United States: an investigation into its epidemiology. *Arch. Intern. Med.*, *161*, 15–21.

Olsen, J., Li, C.C., Bjørnsdottir, I., Sidenius, U., Benet, L.Z., Hansen, S.H., Benet, L.Z. (2005). In vitro and in vivo studies on acyl-coenzyme A-dependent bioactivation of zomepirac in rats. *Chem. Res. Toxicol.*, *18*, 1729–1736.

Olsen, J., Li, C., Skonberg, C., Bjørnsdottir, I., Sidenius, U., Benet, L.Z., Hansen, S.H. (2007). Studies on the metabolism of tolmetin to the chemically reactive acyl-coenzyme A thioester intermediate in rats. *Drug Metab. Dispos.*, *35*, 758–764.

Park, B.K., Coleman, J.W., Kitteringham, N.R. (1987). Drug disposition and drug hypersensitivity. *Biochem. Pharmacol.*, *36*, 581–590.

Park, B.K., Pirmohamed, M., Kitteringham, N.R. (1998). Role of drug disposition in drug hypersensitivity: a chemical, molecular, and clinical perspective. *Chem. Res. Toxicol.*, *11*, 969–988.

Parker, C.W., Deweck, A.L., Kern, M., Eisen, H.N. (1962). The preparation and some properties of penicillenic acid derivatives relevant to penicillin hypersensitivity. *J. Exp. Med.*, *115*, 803–819.

Pellock, J.M. (1999). Felbamate. *Epilepsia*, *40*, 57–62.

Pichler, W.J. (2002). Pharmacological interaction of drugs with antigen-specific immune receptors: the p-i concept. *Curr. Opin. Allergy Clin. Immunol.*, *2*, 301–305.

Pirmohamed, M., Madden, S., Park, K. (1996). Idiosyncratic drug reactions: metabolic bioactivation as a pathogenic mechanism. *Clin. Pharmacokinet.*, *31*, 215–230.

Pohl, L.R., Satoh, H., Christ, D.D., Kenna, J.G. (1988). The immunologic and metabolic basis of drug hypersensitivities. *Annu. Rev. Pharmacol. Toxicol.*, *28*, 367–387.

Pumford, N.R., Martin, B.M., Thomassen, D., Burris, J.A., Kenna, J.G., Martin, J.L., Pohl, L.R. (1993). Serum antibodies from halothane hepatitis patients react with the rat endoplasmic reticulum protein ERp72. *Chem. Res. Toxicol.*, *6*, 609–615.

Qiu, Y., Burlingame, A.L., Benet, L.Z. (1998). Mechanisms for covalent binding of benoxaprofen glucuronide to human serum albumin: studies by tandem mass spectrometry. *Drug Metab. Dispos.*, *26*, 246–256.

Reilly, T.P., Lash, L.H., Doll, M.A., Hein, D.W., Woster, P.M., Svensson, C.K. (2000). A role for bioactivation and covalent binding within epidermal keratinocytes in sulfonamide-induced cutaneous drug reactions. *J. Invest. Dermatol.*, *114*, 1164–1173.

Robin, M.A., Maratrat, M., Le Roy, M., Le Breton, F.P., Bonierbale, E., Dansette, P., Ballet, F., Mansuy, D., Pessayre, D. (1996). Antigenic targets in tienilic acid hepatitis. Both cytochrome P450 2C11 and 2C11-tienilic acid adducts are transported to the plasma membrane of rat hepatocytes and recognized by human sera. *J. Clin. Invest.*, *98*, 1471–1480.

Roecklein, B.A., Sacks, H.J., Mortko, H., Stables, J. (2007). Fluorofelbamate. *Neurotherapeutics*, *4*, 97–101.

Sallustio, B.C., Knights, K.M., Roberts, B.J., Zacest, R. (1991). In vivo covalent binding of clofibric acid to human plasma proteins and rat liver proteins. *Biochem. Pharmacol.*, *42*, 1421–1425.

Sallustio, B.C., Nunthasomboon, S., Drogemuller, C.J., Knights, K.M. (2000). In vitro covalent binding of nafenopin-CoA to human liver proteins. *Toxicol. Appl. Pharmacol.*, *163*, 176–182.

Sanderson, J.P., Naisbitt, D.J., Park, B.K. (2006). Role of bioactivation in drug-induced hypersensitivity reactions. *AAPS J.*, *8*, 55–64.

Sarich, T.C., Zhou, T., Adams, S.P., Bain, A.I., Wall, R.A., Wright, J.M. (1995). A model of isoniazid-induced hepatotoxicity in rabbits. *J. Pharmacol. Toxicol. Methods*, *34*, 109–116.

Satoh, H., Fukuda, Y., Anderson, D.K., Ferrans, V.J., Gillette, J.R., Pohl, L.R. (1985). Immunological studies on the mechanisms of halothane induced hepatotoxicity: immunohistochemical evidence of trifluoroacetylated hepatocytes. *J. Pharmacol. Exp. Ther.*, *223*, 857–862.

Saxon, W. (1999). Anne Miller, 90, first patient who was saved by penicillin. The New York Times, June 9.

Seitz, S., Boelsterli, U.A. (1998). Diclofenac acyl glucuronide, a major biliary metabolite, is directly involved in small intestinal injury in rats. *Gastroenterology*, *115*, 1476–1482.

Shenton, J.M., Chen, J., Uetrecht, J.P. (2004). Animal models of idiosyncratic drug reactions. *Chem. Biol. Interact.*, *150*, 53–70.

Sidenius, U., Skonberg, C., Olsen, J., Hansen, S.H. (2004). In vitro reactivity of carboxylic acid-CoA thioesters with glutathione. *Chem. Res. Toxicol.*, *17*, 75–81.

Small, R.E. (1989). Diclofenac sodium. *Clin. Pharm.*, *8*, 545–558.

Smith, P.C., Benet, L.Z., McDonagh, A.F. (1990). Covalent binding of zomepirac glucuronide to proteins: evidence for a Schiff base mechanism. *Drug Metab. Dispos.*, *18*, 639–644.

Smith, P.C., Liu, J.H. (1995). Covalent binding of suprofen to renal tissue of rat correlates with excretion of its acyl glucuronide. *Xenobiotica*, *25*, 531–540.

Smith, P.C., McDonagh, A.F., Benet, L.Z. (1986). Irreversible binding of zomepirac to plasma protein in vitro and in vivo. *J. Clin. Invest.*, *77*, 934–939.

Solensky, R. (2003). Hypersensitivity reactions to beta-lactam antibiotics. *Clin. Rev. Allergy Immunol.*, *24*, 201–220.

Spahn-Langguth, H., Benet, L.Z. (1992). Acyl glucuronides revisited: is the glucuronidation process a toxification as well as a detoxification mechanism? *Drug Metab. Rev.*, *24*, 5–47.

Spracklin, D.K., Thummel, K.E., Kharasch, E.D. (1996). Human reductive halothane metabolism in vitro is catalyzed by cytochrome P450 2A6 and 3A4. *Drug Metab. Dispos.*, *24*, 976–983.

Stein, H.B., Patterson, A.C., Offer, R.C., Atkins, C.J., Teufel, A., Robinson, H.S. (1980). Adverse effects of D-penicillamine in rheumatoid arthritis. *Ann. Intern. Med.*, *92*, 24–29.

Stoelting, R.K. (1999). Inhaled anesthetics, in *Pharmacology and Physiology in Anesthetic Practice*, Lippincott-Raven Publishers, pp. 36–76.

Stogniew, M., Fenselau, C. (1982). Electrophilic reactions of acyl-linked glucuronides. *Drug Metab. Dispos.*, *10*, 609–613.

Tang, W. (2003). The metabolism of diclofenac—enzymology and toxicology perspectives. *Curr. Drug Metab.*, *4*, 319–329.

Tang, W., Stearns, R.A., Bandiera, S.M., Zhang, Y., Raab, C., Braun, M.P., Dean, D.C., Pang, J., Leung, K.H., Doss, G.A., Strauss, J.R., Kwei, G.Y., Rushmore, T.H., Chiu, S.H., Baillie, T.A. (1999). Studies on cytochrome P-450-mediated bioactivation of diclofenac in rats and in human hepatocytes: identification of glutathione conjugated metabolites. *Drug Metab. Dispos.*, *27*, 365–372.

Tishler, S.L., Goldman, P. (1970). Properties and reactions of salicyl-coenzyme A. *Biochem. Pharmacol.*, *19*, 143–150.

Uetrecht, J. (2006). Role of animal models in the study of drug-induced hypersensitivity reactions. *AAPS J.*, *7*, 914–921.

Uetrecht, J.P. (1999). New concepts in immunology relevant to idiosyncratic drug reactions: the "danger hypothesis" and innate immune system. *Chem. Res. Toxicol.*, *12*, 387–395.

Uetrecht, J.P. (2003). Screening for the potential of a drug candidate to cause idiosyncratic drug reactions. *Drug Discov. Today*, *8*, 832–837.

Uetrecht, J.P. (2008). Idiosyncratic drug reactions: past, present, and future. *Chem. Res. Toxicol*, *21*, 84–92.

Valadon, P., Dansette, P.M., Girault, J.P., Amar, C., Mansuy, D. (1996). Thiophene sulfoxides as reactive metabolites: formation upon microsomal oxidation of a 3-aroylthiophene and fate in the presence of nucleophiles in vitro and in vivo. *Chem. Res. Toxicol.*, *9*, 1403–1413.

Wagner, M.L. (1994). Felbamate: a new antiepileptic drug. *Am. J. Hosp. Pharm.*, *51*, 1657–1666.

Walgren, J.L., Mitchell, M.D., Thompson, D.C. (2005). Role of metabolism in drug-induced idiosyncratic hepatotoxicity. *Crit. Rev. Toxicol.*, *35*, 325–361.

Ware, J.A., Graf, M.L., Martin, B.M., Lustberg, L.R., Pohl, L.R. (1998). Immunochemical detection and identification of protein adducts of diclofenac in the small intestine of rats: possible role in allergic reactions. *Chem. Res. Toxicol.*, *11*, 164–171.

van Wijk, F., Nierkens, S. (2006). Assessment of drug-induced immunotoxicity in animal models. *Drug Discov. Today*, *3*, 103–109.

Williams, A.M., Worrall, S., de Jersey, J., and Dickinson, R.G. (1992). Studies on the reactivity of acyl glucuronides—III. Glucuronide-derived adducts of valproic acid and plasma protein and anti-adduct antibodies in humans. *Biochem. Pharmacol. 43*, 745–755.

Wise, E.M., Jr, Park, J.T. (1965). Penicillin: its basic site of action as an inhibitor of a peptide cross-linking reaction in cell wall mucopeptide synthesis. *Proc. Natl. Acad. Sci. U.S.A.*, *54*, 5–81.

Wright, A.J. (1999). Penicillins. *Mayo Clin. Proc.*, *74*, 290–307.

Yamashita, A., Watanabe, M., Tonegawa, T., Sugiura, T., Waku, K. (1995). Acyl-CoA binding and acylation of UDP-glucuronosyltransferase isoforms of rat liver: their effect on enzyme activity. *Biochem J.*, *312*, 301–308.

Yocum, R.R., Waxman, D.J., Rasmussen, J.R., Stromincser, J.L. (1979). Mechanism of penicillin action: penicillin and substrate bind covalently to the same active site serine in two bacterial D-alanine carboxypeptidases. *Biochemistry*, *76*, 2730–2734.

Zia-Amirhosseini, P., Ding, A., Burlingame, A.L., McDonagh, A.F., Benet, L.Z. (1995). Synthesis and mass-spectrometric characterization of human serum albumins modified by covalent binding of two non-steroidal anti-inflammatory drugs: tolmetin and zomepirac. *Biochem. J.*, *311*, 431–435.

Zimmerman, H. (1999). *Hepatotoxicity: The Adverse Effects of Drugs and Other Chemicals on The Liver*, Lippincott Williams & Wilkins, Philadelphia, PA.

Zimmerman, H.J. (1993). Hepatotoxicity. *Disease-A-Month*, *39*, 673–788.

Zimmerman, H.J. (1994). Hepatic in jury associated with nonsteroidal anti-inflammatory drugs, in *Nonsteroidal Antiinflammatory Drugs. Mechanisms and Clinical Use* (eds A.J. Lewis and D.E. Furst), Marcel Dekker Inc, New York, pp. 171–194.

Zimmerman, H.J., Lewis, J.H., Ishak, K.G., Maddrey, W.C. (1984). Ticrynafen-associated hepatic injury: analysis of 340 cases. *Hepatology*, *4*, 315–323.

Zubay, G.L. (1988). Carbohydrates, in *Biochemistry*, Macmillan Publishing Co., New York, pp. 663–667.

# Chemical Mechanisms in Toxicology

MARK P. GRILLO

> We pray that every field of science may contribute in bringing happiness—not disaster—to human beings.[1]

## 1 INTRODUCTION

This chapter deals with the mechanisms involved in the biotransformation of drugs to chemically reactive metabolites, current experimental techniques used to identify reactive metabolites, and the potential consequences of reactive metabolite formation as it relates to the discovery of nontoxic drugs. Many pharmaceutical researchers in the fields of drug metabolism and toxicology currently perceive chemically reactive metabolites as an unwanted feature of any drug or drug candidate (Baillie, 2008; Kumar *et al.*, 2008). Therefore, an imperative goal in drug discovery is to eliminate, or to substantially decrease, the metabolic activation liability of drug candidates leading to the increased probability of safer drugs being successfully developed (Baillie and Kassahun, 2001; Kalgutkar and Soglia, 2005; Park *et al.*, 2005b; Baillie, 2006; Doss and Baillie, 2006; Williams, 2006). Drug metabolism scientists today are trained in taking closer and closer looks at the metabolism of new chemical entities sometimes having greater and greater complexity. One of the most important objectives in the pharmaceutical industry is to be able to make knowledgeable judgments about hidden risks so that new drugs that could be potentially toxic to patients are never released (Evans *et al.*, 2004; Nassar and Lopez-Anaya, 2004; Evans and Baillie, 2005).

The history of the subject discussed in this chapter may have started more than 200 years ago with a report published by the English surgeon Percivall Pott (1714–1788; Pott, 1775). Pott is credited with recognizing one of the first known industrial-linked diseases when he described the association between working as a chimney sweep and

[1] A quotation from Kenichi Fukui. He was corecipient of the Nobel Prize in Chemistry in 1981 for investigations into the mechanisms of chemical reactions.

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

the development of scrotal carcinoma (Waldron, 1983; Greene, 1990; Ottoboni, 1991). Pott was impressed by the idea that coal soot contained some substance that was causing the cancer, thereby linking this occupational-linked chemical exposure with formation of the unusual tumor (Melicow, 1975). This was the dawn of a whole new era of the way that we now think about substance-induced carcinogenesis. During that time in the rest of Europe, scrotal cancer was not occurring at the same rate, presumably because of the reduced carcinogenicity of soot produced from wood-burning furnaces (IARC, 1985), and also because of the practice of wearing protective clothing (Davis, 2007).

Progress in understanding carcinogenesis sort of sat there for a very long period of time until the early 1900s, when it began to be critically analyzed again by a research group in Japan studying the cancer-causing property of coal tar. A scientific breakthrough in understanding carcinogenesis occurred when, for the first time, scientists were able to experimentally induce cancer in laboratory animals (Yamagiwa and Ichikawa, 1918). In their experiments, Yamagiwa and Ichikawa showed that after applying coal tar to the ear skin of rabbits, tumor formation would occur at the site of exposure.

The next major development was when the isolation and identification of the carcinogen benzo[*a*]pyrene from coal tar occurred (Cook, Hewett and Hieger, 1933; Berenblum and Schoental, 1947; Kennaway, 1955) (Fig. 1). It was shown that when radiolabeled benzo[*a*]pyrene was applied on mouse skin (Pereira, Burns and Albert, 1979), added to mouse embryo cell cultures (Duncan, Brookes and Dipple, 1969), or incubated in the presence of rat liver microsomes, NADPH, and calf thymus nucleic acids (Gelboin, 1969), electrophilic metabolites were formed that became covalently bound to DNA. From these observations, the initiation of carcinogenesis by polycyclic aromatic hydrocarbons was proposed to involve covalent binding of benzo[*a*]pyrene to DNA macromolecules (Brookes and Lawley, 1964). This covalent binding phenomenon was considered very peculiar at the time, because benzo[*a*]pyrene was known to

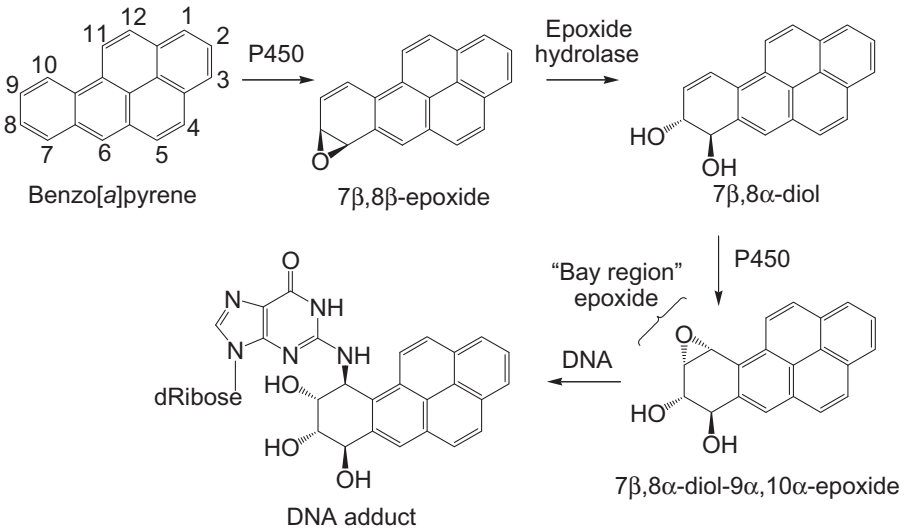


**Figure 1** Metabolic activation of benzo[*a*]pyrene leading to covalent binding to DNA.

**TABLE 1 Hard versus soft nucleophiles and electrophiles.**

| Hard nucleophiles | Intermediate | Soft nucleophiles |
|---|---|---|
| Oxygen in purines and pyrimidines of nucleic acids; phosphate oxygen in DNA; $OH^-$; $H_2O$; ROH; $NH_3$ | Nitrogen in primary and secondary amino groups of proteins; nitrogen in amino groups in purine bases in DNA; $CN^-$ | Sulfur in thiols (e.g. cysteine residues in proteins and peptides like glutathione) |
| **Hard electrophiles** | **Intermediate** | **Soft electrophiles** |
| Alkyl carbonium ions | Aryl carbonium ions; aryl halides; benzylic carbonium ions; nitrenium ions; carbon of epoxides | Carbon in polarized double bonds (e.g. α,β-unsaturated ketones, quinones, and quinone imines) |

Adapted from Carlson (1990) and Gregus and Klaassen (2001).

be inherently unreactive, and the mechanism whereby it became irreversibly bound to DNA was not understood. But we now know that the bioactivation of benzo[*a*]pyrene follows a mechanism where it undergoes metabolism by cytochrome P450 to a chemically reactive, electrophilic, bay region diol epoxide that binds covalently to DNA (Sims *et al.*, 1974) (Fig. 1).

Observations from collected mechanistic studies were combined into a working hypothesis during the 1960s and 1970s from the innovative research and major insights of Miller and Miller (Miller and Miller, 1974, 1979, 1981; Heidelberger, 1975). The hypothesis that was developed stated that for some compounds to become carcinogenic, they first must be metabolically activated by membrane-bound cytochrome P450-dependent monooxygenases to chemically reactive species (Miller and Miller, 1974). Predominantly, it was viewed that chemically reactive hard electrophiles (those species that have a concentrated charge such as a carbocation, a nitrenium ion, or an iminium ion) react covalently with hard nucleophiles such as those located in nucleic acids (Carlson, 1990; Gregus and Klaassen, 2001) (Table 1). This mechanistic proposal, which became the framework by which we now study reactive drug metabolites, is considered the beginning of the field of xenobiotic bioactivation and led to a transformation of thinking about chemical carcinogenesis (Mitchell *et al.*, 1976; Nelson and Pearson, 1990).

These approaches continued to be advanced at the National Institute of Health by research from Brodie, Gillette, Mitchell, and Nelson, and their colleagues which concentrated on the mechanistic toxicology of drugs (Jollow *et al.*, 1973; Gillette and Pohl, 1977; Nelson *et al.*, 1978; Gillette, 1981; Baillie, 2006). They showed that drug-induced hepatotoxicity could be associated with the metabolic activation of therapeutic agents to chemically reactive soft electrophiles. Soft electrophiles are often uncharged and less electronegative than hard, highly polarized electrophiles. Soft electrophiles react with soft nucleophiles, such as the amino acid side chains of proteins (e.g. the sulfur atom of thiols found in cysteinyl residues of proteins and glutathione) (Table 1). In general, soft electrophiles react with soft nucleophiles, whereas hard electrophiles react more readily with hard nucleophiles (Carlson, 1990).

A focus of this chapter is on the implication of metabolic activation in the discovery of drugs. This is a topic of substantial interest in the pharmaceutical industry due to the major concern for the potential of some drugs to form reactive metabolites that

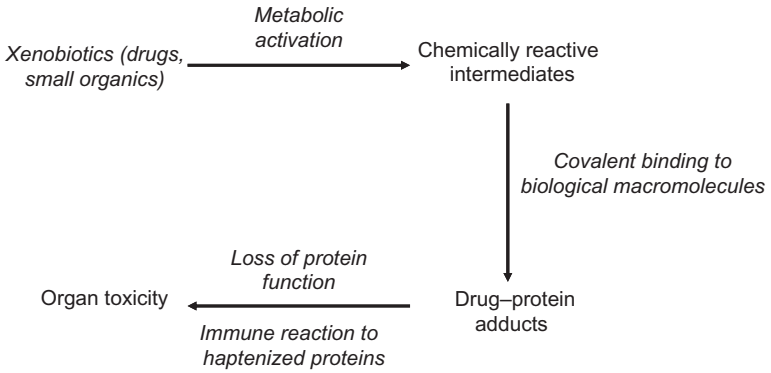


**Figure 2** Metabolic activation of xenobiotics to chemically reactive and potentially toxic products.

**TABLE 2 Drugs or chemicals known to cause idiosyncratic hepatotoxicity.**

| Acetaminophen[a] | Diclofenac | Isoxanine[b] | Tacrine[a] |
|---|---|---|---|
| Alpidem[b] | Dihydralazine[b] | Ketoconazole | Tamoxifen[a] |
| Amineptine[b] | Disulfiram[a] | Leflunomide[a] | Terbinafine[a] |
| Amodiaquine | Ebrotidine[b] | Methyldopa[a] | Tienilic acid[b] |
| Benoxaprofen[b] | Ethanol | Nefazadone[b] | Tilbroquinol[b] |
| Bromefenac[b] | Felbamate[a] | Niperotidien[b] | Troglitazone[b] |
| Carbamazepine[a] | Halothane[a] | Phenytoin | Trovofloxacin[b] |
| Cinchophen[b] | Ibufenac[b] | Pirprofen[b] | Valproic acid[a] |
| Clozapine[a] | Iproniazid[b] | Rifampin[a] | Zileuton[a] |

[a]Drugs with black box warnings.
[b]Withdrawn drugs.
Adapted from Guengerich and MacDonald (2007) and Walgren, Mitchell and Thompson (2005).

can react with macromolecules in the body causing immune-based idiosyncratic toxicities (Uetrecht, 2003; Kalgutkar and Soglia, 2005; Baillie, 2006; Williams, 2006) (Fig. 2). These toxicities sometimes are caused by the haptenization of protein that leads to an untoward "triggering" event in the immune system of hypersensitive individuals (Ju and Uetrecht, 2002; Uetrecht, 2003; Kaplowitz, 2005). These severe, and occasionally fatal, immunologic-based toxic reactions resulting from this abnormality show a very low incidence of 1 in 1000 or even 1 in 100,000 patients (Zimmerman, 1999; Calvey, 2005; Pirmohamed, 2006). This is a tremendous concern in the pharmaceutical industry, since there have been several drugs that have been withdrawn from the market for this reason (Walgren, Mitchell and Thompson, 2005; Guengerich and MacDonald, 2007) (Table 2). It is especially of frustrating concern because there are almost no animal models available today to be able to predict toxic interactions of drugs with the human immune system (Shenton, Chen and Uetrecht, 2004; van Wijk and Nierkens, 2006).

After almost four decades of research showing strong indirect evidence that many types of reactive intermediates may be cytotoxic, pharmaceutical scientists are still trying to understand the significance of chemically reactive metabolites in drug toxicity (Kalgutkar and Soglia, 2005; Baillie, 2006). Over the last 10 to 20 years, there have been substantial improvements in understanding how reactive metabolites are formed and

how they behave chemically. These improvements occurred because the disciplines of drug metabolism and analytical chemistry became increasingly capable at characterizing the chemical structures of reactive intermediates (Baillie, 2006; Liebler, 2006). These advancements are primarily due to considerable discoveries in liquid chromatography–tandem mass spectrometry (LC-MS/MS) techniques, such as the advent of the electrospray ionization interface, accurate mass measurement to enhance metabolite structural identification, and recent improvements in data acquisition and processing for accelerating metabolite identification. However, "Advances in the field of biochemical toxicology have not kept pace with those in drug metabolism and bioanalytical chemistry, such that we are unable to predict, *a priori*, which reactive intermediates may be toxic to the host cell *-vs* those that will be relatively benign" (Baillie, 2006). One accomplishable path forward used by drug metabolism scientists today has been to try to "minimize" the formation of reactive metabolites before taking drug discovery candidates into development, but unfortunately, today, still not being able to predict which drug candidate(s) may potentially cause idiosyncratic toxicity in patients (Evans and Baillie, 2005; Kumar *et al.*, 2008). This pathway forward currently involves two approaches that require a team of experts from drug metabolism, medicinal chemistry, and toxicology working closely together within a drug discovery program (Evans *et al.*, 2004; Mutlib *et al.*, 2005). The first approach currently used by many pharmaceutical companies involves screening techniques to find glutathione adducts formed when candidate drugs are incubated with liver microsomal preparations from a range of species usually including rat, dog, monkey, and human (Dieckhaus *et al.*, 2005; Kalgutkar and Soglia, 2005; Kalgutkar *et al.*, 2005).

## 2 GLUTATHIONE ADDUCTS

*In vitro* experiments in liver microsomes fortified with NADPH (the cofactor necessary for P450-mediated oxidation) are conducted in the presence of nucleophilic trapping agents such as glutathione (a soft nucleophile that reacts with soft electrophiles) or cyanide anion (a hard nucleophile that can react with hard electrophiles), in order to trap chemically reactive species that may have been formed in these incubations. Glutathione reacts with, and thereby captures, electrophilic metabolites by both nucleophilic substitution and nucleophilic addition reactions as depicted in Fig. 3.

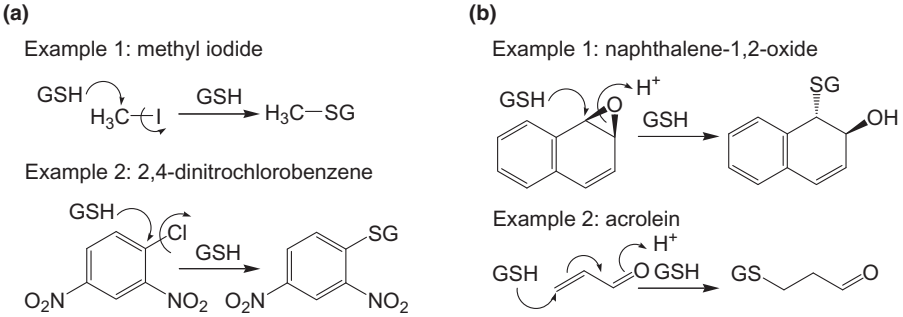


**Figure 3** Chemically reactive compounds react with GSH by one of two general mechanisms: **(a)** nucleophilic substitution-type and **(b)** nucleophilic addition-type reactions.

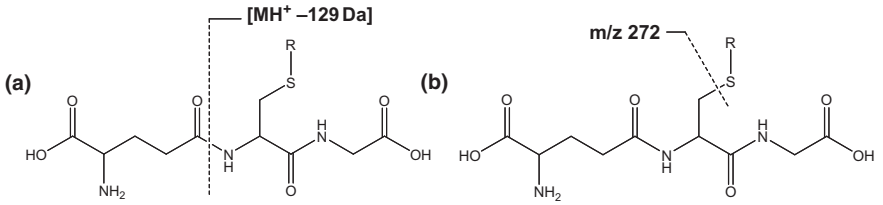


**Figure 4** LC-MS/MS screening for glutathione adducts present in biological extracts by **(a)** constant neutral loss scanning (–129 Da in positive ion scan mode) and **(b)** precursor ion scanning (m/z 272 in negative ion scan mode).

If a conjugate is detected, then the focus is to elucidate the adduct structure by LC-MS/MS characterization, and, when sufficient amount of adduct has been isolated, by nuclear magnetic resonance (NMR) analysis to confirm the atomic site of adduct linkage. Once the site(s) of metabolic activation of a compound is (are) determined, then medicinal chemists can modify the drug candidate (e.g. by fluorine atom substitution) to block, or at least hinder, bioactivation. LC-MS/MS methods are now routinely used to examine biological extracts for the presence of glutathione adducts produced *in vitro* or *in vivo* using specialized scanning techniques. The LC-MS/MS detection techniques employed most often for glutathione conjugate detection are traditional constant neutral loss scanning for 129 Da, which is based on the loss of the elements of pyroglutamic acid from the glutathione adduct protonated molecular ion upon collision-induced dissociation (Fig. 4a) (Baillie and Davis, 1993), and by negative ion tandem LC-MS/MS for precursors of the product ion at m/z 272, which is the fragment ion consisting of the elements of glutathione minus $H_2S$ (Fig. 4b) (Dieckhaus *et al.*, 2005). Therefore, new chemical entities that form glutathione adducts can, in most cases, be detected by these two LC-MS/MS screening methodologies which work independent of compound structure, can be detected without exact knowledge of the adduct structure, and are amenable to high-throughput screening (Baillie and Davis, 1993; Kalgutkar and Soglia, 2005).

The capturing of chemically reactive intermediates with glutathione, and subsequent identification and structural characterization by LC-MS/MS and NMR, is an advantageous methodology because it does not require often expensive and time-consuming synthesis of radioactive drug candidates (Zheng *et al.*, 2007). Other useful strategies include the use of an equal mixture of stable isotope-labeled (e.g. $^{13}C$, $^{15}N$) glutathione and unlabeled glutathione to trap reactive metabolites which assists in adduct detection by LC/MS analysis as an "isotope doublet" (Yan and Caldwell, 2004; Mutlib *et al.*, 2005), as well as the use of derivatized glutathione for semiquantitative determination of glutathione adduct levels (Gan *et al.*, 2005; Soglia *et al.*, 2006). However, these methods employing glutathione stable isotope labeled or derivatized analogs are not suitable for the trapping and identification of hard electrophilic intermediates, where a mixture of $KCN/K^{13}C^{15}N$ should be used to capture, for example, reactive carbocation or iminium ion intermediates (Evans *et al.*, 2004; Argoti *et al.*, 2005).

For successful NMR analysis of purified conjugates, up to milligram amounts of adduct may be necessary for structural characterization. Therefore, *in vivo* studies in preclinical species can be performed in bile duct cannulated animals, where glutathione or mercapturic acid adducts can be isolated by HPLC purification from bile or urine, respectively (Fig. 5).

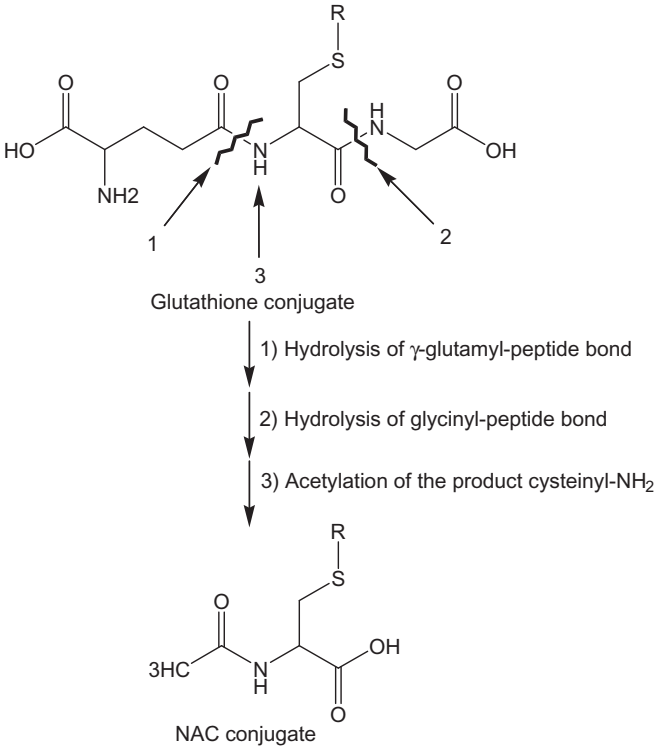


**Figure 5** Enzyme-mediated degradation of glutathione conjugates in the kidney leading to the corresponding mercapturic acid conjugate (*N*-acetylcysteine [NAC] conjugate).

This first approach analysis is performed to provide an initial mechanistic understanding of the route of metabolic activation and is of express interest early in the discovery program when changes can be made to a drug structure due to enhanced program resources. These issues regarding chemically reactive metabolite are inherently important in a drug discovery program and require rapid answers. With the knowledge of reactive metabolite structure, medicinal chemists can, in most cases, modify lead structures to block, or at least substantially decrease, the potential for metabolic activation. Glutathione adduct detection is performed as an iterative process in the early stages of drug discovery and lead candidate optimization. In summary, reactive metabolite-trapping experiments with glutathione, *N*-acetylcysteine (NAC), or cyanide anion can provide stable adducts for structural elucidation to assist in the mechanistic understanding of metabolic activation of candidate drugs (Evans *et al.*, 2004).

Mercapturic acid adducts excreted in urine can also be used as biomarkers for exposure to chemically reactive metabolites formed *in vivo* in preclinical species and in humans (van Welieet *et al.*, 1992; Scholz *et al.*, 2005). Rapid and sensitive detection of mercapturic acid conjugates of unknown structure can be analyzed by negative ion mode LC-MS/MS constant neutral loss scanning for 129 Da from urine extracts, or from extracts obtained from liver microsomal incubations fortified with NADPH and NAC (Scholz *et al.*, 2005) (Fig. 6).

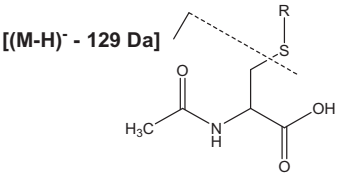


**Figure 6** LC-MS/MS screening for mercapturic acid adducts present in biological extracts by constant neutral loss scanning (–129 Da in negative ion scan mode).

## 3 COVALENT BINDING

Drug candidate optimization during late-stage discovery often leads to only a few qualified compounds remaining that retain desirable drug properties such as high target potency, excellent receptor selectivity, and appropriate animal pharmacokinetics. A key decision to be made at this stage is in the selection of the best compound to take forward into development. Making this decision can be assisted by evaluating drug candidate covalent binding potential, especially if glutathione conjugates were detected during preliminary *in vitro* and *in vivo* metabolic activation studies. Once a compound is introduced into drug development, there is major economic expenditure made in making bulk drug to conduct preclinical toxicity studies. Therefore, the obvious need is to perform experiments on metabolic activation rapidly so that medicinal chemists can make the necessary structural changes such that bioactivation properties of the compound are eliminated, of course without dramatically affecting potency, receptor selectivity, and preclinical species pharmacokinetics (Doss and Baillie, 2006; Guengerich and MacDonald, 2007; Kumar *et al.*, 2008). This second experimental approach, which has been embraced by many pharmaceutical companies, is to measure covalent binding of radioactivity from radiolabeled [$^{14}$C-, $^{3}$H-, or $^{35}$S-labeled (if applicable)] drug candidates formed *in vitro* in microsomal liver incubations (Evans *et al.*, 2004; Kalgutkar and Soglia, 2005). The utmost impediment for these covalent binding studies is the availability of radiolabeled drug candidates. However, once radiolabeled derivatives are made available, liver microsomal incubations are performed in the presence and absence of NADPH (to provide a negative control) and are compared directly with results from a positive control compound studied in the same experiment. Covalent binding of radiolabeled compounds to liver microsomal proteins is measured under standardized conditions (Day *et al.*, 2005). Such studies are performed in microsomal preparations cross-species, including human. If covalent binding to protein is detected *in vitro*, then *in vivo* covalent binding to liver protein can also be measured to evaluate *in vivo* exposure to reactive metabolites, although usually only studied in rats. The highest levels of covalent binding to protein are usually detected in the liver due to the high concentration of P450 enzymes in hepatic tissue (Waring and Anderson, 2005). The focal point of this second approach is to assess the exposure to reactive metabolites, but it does not foretell toxicity or potential toxicity of the drug candidate. To date, it is not known what extent of covalent binding to protein correlates with the risk of a drug candidate being able to cause an adverse idiosyncratic drug reaction (Evans *et al.*, 2004; Baillie, 2006; Uetrecht, 2006).

If covalent binding to protein is detected, then the goal is to block exposure to these reactive intermediates by making strategic changes to the chemical structure of the candidate molecule to hinder bioactivation pathway(s) (Evans *et al.*, 2004; Doss and

Baillie, 2006). What covalent binding amount is considered acceptable? This important question has been prudently and carefully addressed by the Preclinical Drug Metabolism department at Merck Research Laboratories (Evans *et al.*, 2004; Evans and Baillie, 2005; Doss and Baillie, 2006). They came up with a target covalent binding number of 50 pmol equivalents of drug residue/mg protein/h of incubation. It was proposed that this value be used as a target covalent binding level that discovery programs would want their candidate drugs to be below. At Merck Research Laboratories, they stress that this target value should not be considered a threshold value (Evans *et al.*, 2004). The number of 50 pmol equivalents of the drug residue/mg protein/h was not arbitrarily derived, but came from a thorough literature search for the levels of covalent binding to liver proteins in animals dosed with known hepatotoxins, for example bromobenzene (Monks, Hinson and Gillette, 1982), isoniazid (Nelson *et al.*, 1978), and acetaminophen (Matthews *et al.*, 1997), under conditions where these drugs induced hepatotoxicity (Evans *et al.*, 2004). When the values of covalent binding to protein for these compounds were measured, the levels were as high as 1000 to 2000 pmol equivalents/mg liver protein. Therefore, the covalent binding target adopted by Merck Research Laboratories is about 5% as much as that caused by many of these model hepatotoxic compounds. In addition, an experimental reason for this value is that the number of 50 pmol equivalents of the drug residue/mg protein/h is about 10-fold higher than the limits of detection of standard liquid scintillation counters used today (Evans *et al.*, 2004). Covalent binding studies can be very helpful because they aid drug discovery scientists in screening out drugs candidates that form reactive intermediates from going forward into late drug discovery or early drug development.

Having obtained results on covalent binding for a drug candidate, a drug discovery team can put this data into perspective using a set of qualifying considerations as proposed by Merck Research Laboratories (Evans *et al.*, 2004; Doss and Baillie, 2006). The qualifying considerations reached by Evans *et al.* (2004) state that if a drug is designed for the treatment of a disabling or life-threatening disease, if there is an unmet medical need, or if the duration of the use of the drug is acute rather than chronic, then the drug discovery program might decide to take the drug forward into development. In addition, if the drug is designed for a novel therapeutic target and to be used only for proof-of-concept purposes, then the compound might still be considered appropriate for progression into target validation-type studies in humans. When pursuing to develop drugs for pediatric indications, the compounds would probably have to be as "issue-free" as possible (i.e. no detectable covalent binding or metabolic activation potential). Another important aspect is related to the clinical dose. There have been very few drugs taken off the market for toxicity reasons when the daily dose was less than 10 mg (Uetrecht, 2001; Kalgutkar *et al.*, 2002). Therefore, if candidate drugs have high potency and excellent pharmacokinetic properties, such that these drugs could be given to patients at low daily doses, then a project team might decide to advance such compounds into development, even when faced with covalent binding issues.

## 4 STRUCTURAL ALERTS

Hypotheses dealing with direct (intrinsic) or immunologic toxicities mediated by chemically reactive species were increasingly being developed and tested during the 1990s (Tirmenstein and Nelson, 1989; Nelson and Pearson, 1990; Rashed, Myers and Nelson, 1990; Roberts, Price and Jollow, 1990; Myers *et al.*, 1995; Nelson, 1995). Since then, there

has been a merging of the chemical, biochemical, and toxicologic information concerning metabolic activation which has afforded hypotheses to address the roles of chemically reactive intermediates as mediators in drug-induced toxicity. This work has resulted in the generation of a list of "structural alerts" which are structural moieties within larger drug structures that seem to be problematic for drugs based on a historical perspective (Nelson, 1995; Kalgutkar *et al.*, 2005; Guengerich and MacDonald, 2007). These structural alerts include, for example, aromatics, 3-methylindoles, hydrazines and hydrazides, carboxylic acids (especially arylacetic and arylpropionic acids), thiophenes, furans, pyrroles, anilines and anilides, moieties that form quinones and quinone imines, halogenated hydrocarbons and some halogenated aromatics, nitroaromatics, and moieties that form reactive α,β-unsaturated ketone-like structures, thiazolidinediones, alkenes, alkynes, and finally thioureas (Table 3).

As shown in Table 3, xenobiotics can be biotransformed to chemically reactive intermediates by a wide variety of mechanisms leading to products that can covalently bind to macromolecules such as nucleic acids and proteins. Covalent binding in some cases corresponds to toxicity (Nelson and Pearson, 1990; Cohen *et al.*, 1997). Toxic xenobiotics can be categorized into two basic groups: intrinsic toxins and idiosyncratic toxins (Pumford and Halmes, 1997). Intrinsic toxins are substances that have a very high potential to cause organ damage at increased doses, but will not be toxic if used appropriately. These compounds can produce a predictable, dose-dependent, and acute toxicity. In other words, intrinsic toxicity is demonstrated by the concentration of a substance that produces a toxic effect in an organism. One example of an intensely studied intrinsic toxin is acetaminophen, a commonly used over-the-counter medication.

Idiosyncratic toxins are xenobiotics that lead to unpredictable and unexpected toxic drug reactions. Idiosyncratic reactions to drugs can be defined as reactions that do not occur in most patients at any dose, they do not relate to the pharmacological activity of the drug, and the onset of reaction usually takes more than one week after initial exposure to the xenobiotic (Ju and Utrecht, 2002). These types of toxic reactions are commonly referred to as hypersensitivity, or type-B, reactions (Uetrecht, 1999). A well-studied example of a drug known to cause idiosyncratic toxicity is the nonsteroidal anti-inflammatory drug (NSAID) diclofenac (Boelsterli, 2003; Tang, 2003).

## 5 EXAMPLES OF METABOLIC ACTIVATION

### 5.1 Aromatics

Naphthalene (Fig. 7) is an aromatic hydrocarbon that is made up of two fused benzene rings and can be found in cigarette smoke (Schmeltz *et al.*, 1978), petroleum-based fuels (Clark *et al.*, 1982), and in mothballs. Naphthalene is a lung toxicant that undergoes P450 2F-, 1A1- and 1A2-mediated bioactivation forming a chemically reactive epoxide, namely naphthalene-1,2-oxide. This epoxide reacts with glutathione, leading to the formation of three separate thioether-linked glutathione adducts (Buonarati, Jones and Buckpitt, 1990; Shultz *et al.*, 1999). Radiolabeled [$^{14}$C]naphthalene is commercially available and is often used as a positive control in standardized covalent binding studies (Day *et al.*, 2005). *In vitro* covalent binding studies performed under standardized conditions show that [$^{14}$C]naphthalene (10 μM) covalently binds to microsomal protein (1 mg/mL) in the presence of NADPH (1 mM) at ~900 and ~1200 pmol equivalents/mg

protein/h for rat liver microsomal and human liver microsomal incubations, respectively (Day *et al.*, 2005).

### 5.2 Thiophenes

Suprofen (Fig. 8), an NSAID, was used in the United States for about three years before several cases of nephrotoxicity occurred prior to the drug being removed from the market (Hart, Ward and Lifschitz, 1987; Fung *et al.*, 2001). Suprofen contains the thiophene structural alert and is metabolically activated to thiophene *S*-oxide, thiophene epoxide, and γ-thioketo-α,β-unsaturated aldehyde reactive intermediates. The drug also causes a mechanism-based inactivation of P450 2C9, where the electrophilic γ-thioketo-α,β-unsaturated aldehyde intermediate was considered the most likely species to covalently bind to, and inactivate, the enzyme (O'Donnell, *et al.*, 2003) (Fig. 8). Another thiophene-containing drug that forms chemically reactive metabolites is diuretic drug tienilic acid, where it undergoes P450 2C9-mediated metabolic activation to epoxide- and *S*-oxide-type reactive intermediates (see Chapter Allergic Reactions to Drugs; Fig. 5 of Allergic Reactions to Drugs).

### 5.3 Furans

Furosemide (Fig. 9) is an example of a drug that contains the furan-type structural alert (Kalgutkar *et al.*, 2005; Williams, 2006) (Fig. 9). A major use of the drug is for pulmonary edema, and is widely and safely used for this purpose (Jackson, 1996). Furosemide is metabolically activated by P450 leading to the formation of chemically reactive epoxide and α,β-unsaturated dicarbonyl intermediates. Both of these reactive species lead to covalent binding to protein (Mitchell, Snodgrass, Gillette, 1976; Wirth, Bettis and Nelson, 1976; Kalgutkar *et al.*, 2005), and to the production of glutathione adducts formed *in vitro* in microsomal incubations and *in vivo* in rats (Williams *et al.*, 2007). Other published examples of furan metabolic activation include the novel 5-lipoxygenase inhibitor, L-739,010 (Zhang *et al.*, 1996) and the potent HIV-1 encoded furano-pyridine-type protease inhibitor, L-754,394 (Sahali-Sahly *et al.*, 1996). Although not used as drugs, other furan-containing molecules known to undergo bioactivation are the furanocoumarin, bergamottin (Kent *et al.*, 2006), the dihydrofuran, aflatoxin (Degen and Neumann, 1978), and the lung-toxic furanoterpenoid, 4-ipomeanol (Boyd, 1976).

### 5.4 Thiazolidinediones

Troglitazone (Fig. 10) is an oral antidiabetic drug that was used effectively to treat type-2 diabetes and which contains the thiazolidinedione structural alert. However, the drug was removed from the market in March 2000, when it was shown to cause a severe hepatotoxicity that resulted in some fatalities (Isley, 2003). Cytochrome P450 mediates the opening of the thiazolidinedione ring system with the production of isothiocyanate- and sulfenic acid-type electrophilic moieties formed in the same product (Kassahun *et al.*, 2001). Troglitazone is also metabolized by P450 on the chromane moiety, giving rise to a quinone methide-type reactive metabolite (Table 3) (Kassahun *et al.*, 2001). Therefore, more than one pathway is involved in the metabolic activation of troglitazone, but whether reactive intermediates are involved in troglitazone-induced liver failure is not yet understood. There are other thiazolidinedione-containing antidiabetic drugs on the market, including pioglitazone (therapeutic dose 15–30 mg/day) and

**TABLE 3 Examples of chemicals containing structural alerts and their corresponding reactive metabolites.**

| Structural alert | Example chemical | Proposed reactive metabolite |
|---|---|---|
| Hydrazines | 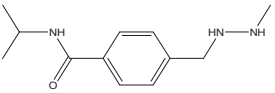 Procarbazine | 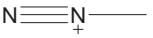 Methyl diazonium |
| Hydrazides | 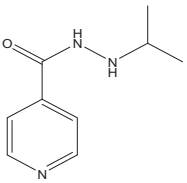 Iproniazid | 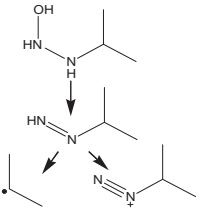 Radical and diazonium ion |
| Arylacetic acids | 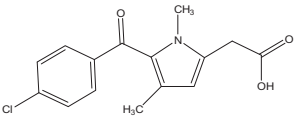 Zomepirac | 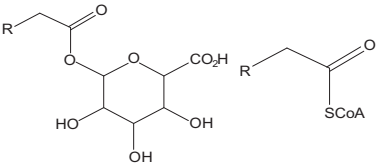 Acyl glucuronide and acyl-CoA |
| Arylpropionic acids | 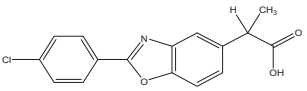 Benoxaprofen | 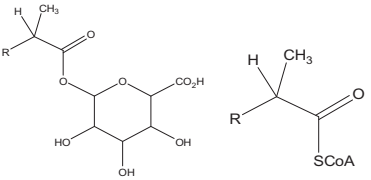 Acyl glucuronide and acyl-CoA |
| Thiophene | 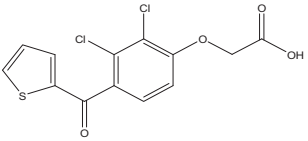 Tienilic acid | 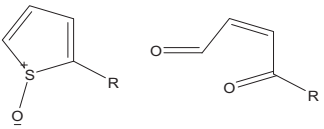 Thioohene-*S*-oxide and a,β-unsaturated dicarbonyl |
| Furans | 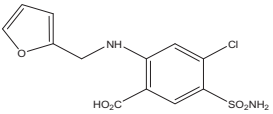 Furosemide | 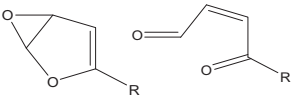 Epoxide and a,β-unsaturated dicarbonyl |

| Structural alert(s) | Example chemical | Proposed reactive metabolite(s) |
|---|---|---|
| Halogenated hydrocarbons | 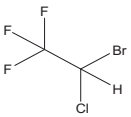 Halothane | 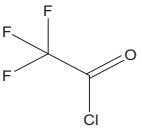 Acyl halide |
| Halogenated aromatics | 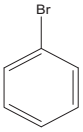 Bromobenzene | 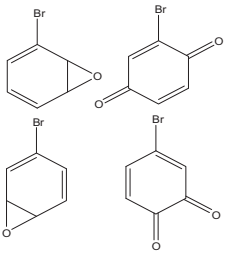 Epoxides, quinones |
| Nitro-aromatics | 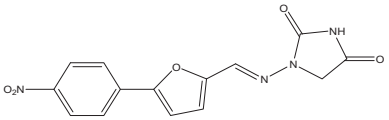 Dantrolene | 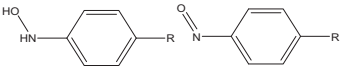 Hydroxylamine, nitroso |
| Thioureas | 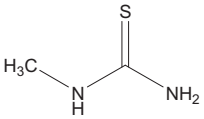 1-Methylthiourea | 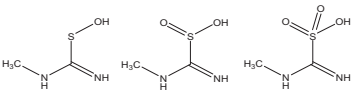 Sulfenic, sulfinic, and sulfonic acids |
| Moieties that form reactive a,β-unsaturated ketone-like structures | 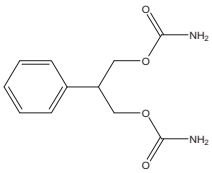 Felbamate | 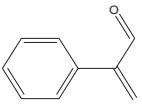 Michael acceptor |
| Thiazolidinediones | 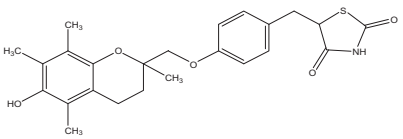 Troglitazone | 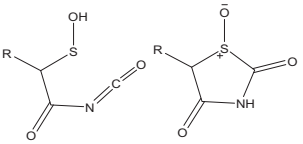 Isocyanate, *S*-oxide |

**TABLE 3** *Contiuned*

| Structural alert | Example chemical | Proposed reactive metabolite |
|---|---|---|
| Anilines | 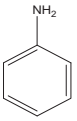 Aniline | 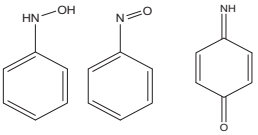 Hydroxylamines, nitroso and quinone imine |
| Anilides | 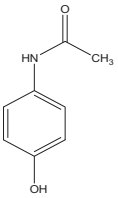 Acetaminophen | 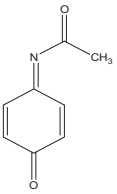 Quinone imine |
| Phenols | 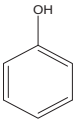 Phenol | 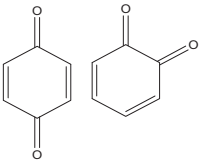 *p*-Quinone, *o*-quinone |
| Moieties that form a reactive quinone imine | 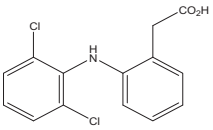 Diclofenac | 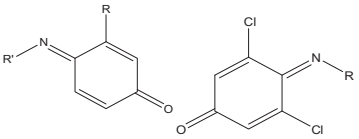 Quinone imines |
| Alkenes | 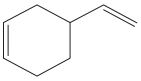 4-Vinylcyclohexene | Epoxide |

Adapted from Kalgutkar *et al.* (2005) and Guengerich and MacDonald (2007).

rosiglitazone (therapeutic dose 4–8 mg/day), both of which have been shown to cause hepatotoxicity. However, the incidence rate of hepatotoxicity is very low compared with troglitazone (therapeutic dose of 600 mg/day), potentially due to the comparatively low doses used for these new-generation thiazolidinediones (Scheen, 2001a, b).

## 5.5 Compounds That Form Reactive Quinones

***5.5.1 Halogenated Aromatics*** The organic solvent bromobenzene (Fig. 11) is a halogenated aromatic molecule that undergoes P450-mediated bioactivation to the reactive intermediates, namely bromobenzene epoxide and bromobenzene *p*-quinone,

| Structural alert(s) | Example chemical | Proposed reactive metabolite(s) |
|---|---|---|
| Pyrroles | 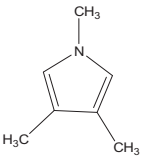 1,3,4-Trimethylpyrrole | 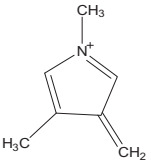 Imine methide |
| Moieties that form a reactive quinone methide | 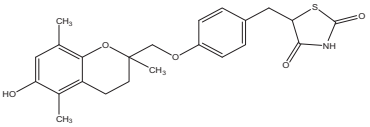 Troglitazone | 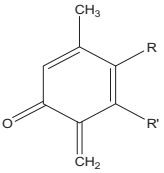 Quinone methide |
| Aromatics | 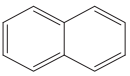 Napthalene | 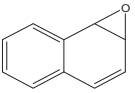 Naphthalene oxide |
| Indoles | 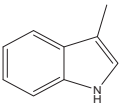 3-Methylindole | 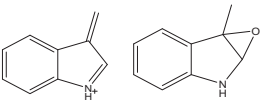 Imine methide, indole epoxide |
| Alkynes | 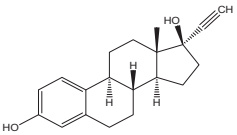 17a-Ethynylestradiol | 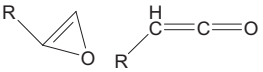 Oxirene and ketene |

both of which are able to arylate tissue macromolecules *in vivo* (Bambal and Hanzlik, 1995), arylate proteins *in vitro* (Narasimhan *et al.*, 1988), and react with glutathione forming the corresponding adducts as depicted in Fig. 11 (Fakjian and Buckpitt, 1984). Bromobenzene has been shown to cause hepatotoxicity and nephrotoxicity which is proposed to be mediated by the formation of these chemically reactive metabolites (Reid *et al.*, 1971; Gregus and Klaassen, 2001).

***5.5.2 Phenols*** Raloxifene (Fig. 12) is a benzothiophene-containing selective estrogen receptor modulator used for the treatment of osteoporosis in postmenopausal women (Lufkin, Wong and Deal, 2001). Raloxifene contains both aromatic- and

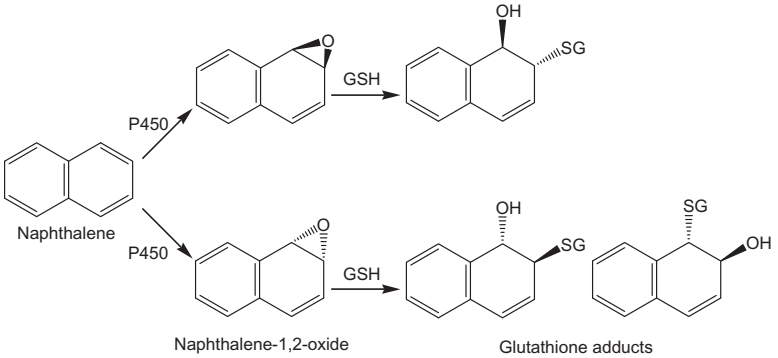


**Figure 7** Scheme for the metabolic activation of naphthalene and formation of corresponding glutathione adducts.

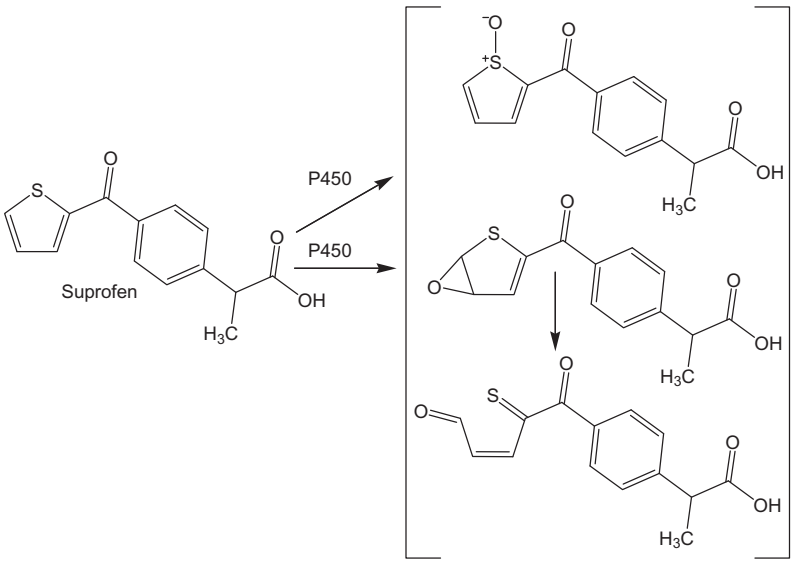


**Figure 8** Scheme for the metabolic activation of suprofen to electrophilic thiophene *S*-oxide-, a thiophene epoxide-, and an γ-thioketo-α,β-unsaturated aldehyde-type intermediates.

phenol-type structural alerts. The drug was recently shown to undergo metabolic activation in incubations with human liver microsomes, leading to the proposed formation of arene oxide(s) and a quinone-type reactive metabolites that react with glutathione, but also lead to the time-dependent inactivation of P450 3A4 (Chen *et al.*, 2002) (Fig. 12). Mass spectrometric analysis was used to show that raloxifene becomes covalently bound to intact P450 3A4 apoprotein (Baer, Wienkers and Rock, 2007). Further mass spectrometric studies on the adducted apoprotein provided strong evidence (mass increase by 471 Da) indicating that raloxifene-mediated P450 3A4 inactivation occurred

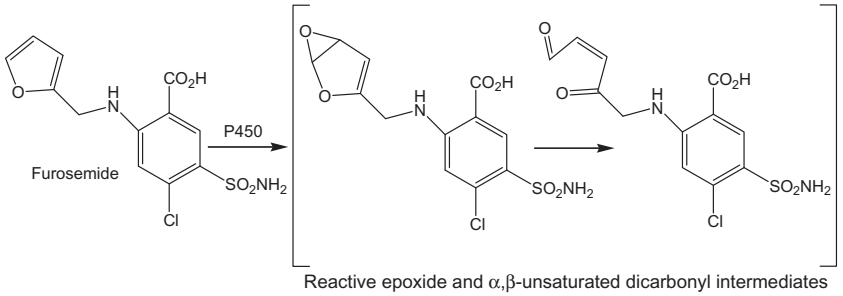


**Figure 9** Scheme for the metabolic activation of furosemide to reactive furan epoxide and α,β-unsaturated dicarbonyl intermediates.

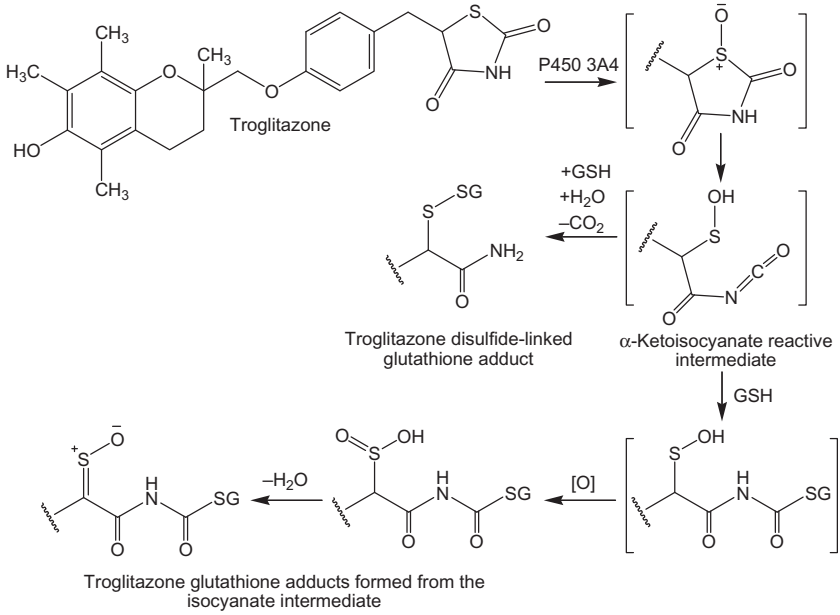


**Figure 10** Scheme for the metabolic activation of the thiazolidinedione moiety of troglitazone leading to the formation of glutathione adducts (adapted from Kassahun *et al.*, 2001).

via bioactivation to the extended quinone reactive metabolite (Fig. 12) with subsequent arylation of the Cys239 active-site residue.

### 5.6 3-Methylindoles

Zafirlukast, a leukotriene receptor antagonist used for the treatment of asthma, contains a 3-methylindole-like structural alert (Fig. 13). The use of zafirlukast has been associated with an idiosyncratic hepatotoxicity that is consistent with a hypersensitivity reaction to the drug (Reinus *et al.*, 2000). Recently, it was shown that zafirlukast is metabolically activated by P4503A4 in human liver microsomes to a highly reactive electrophilic α,β-unsaturated iminium methide intermediate that forms adducts with

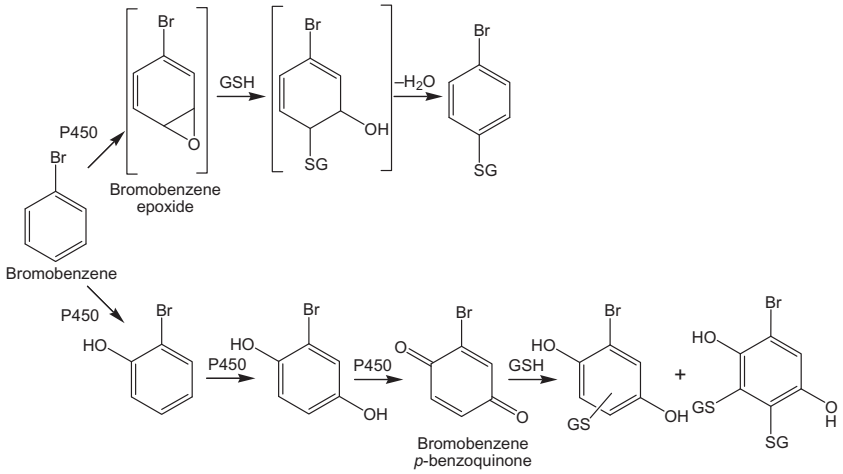


**Figure 11** Scheme for the metabolic activation of bromobenzene and formation of the corresponding glutathione adducts (adapted from Kalgutkar *et al.*, 2005).

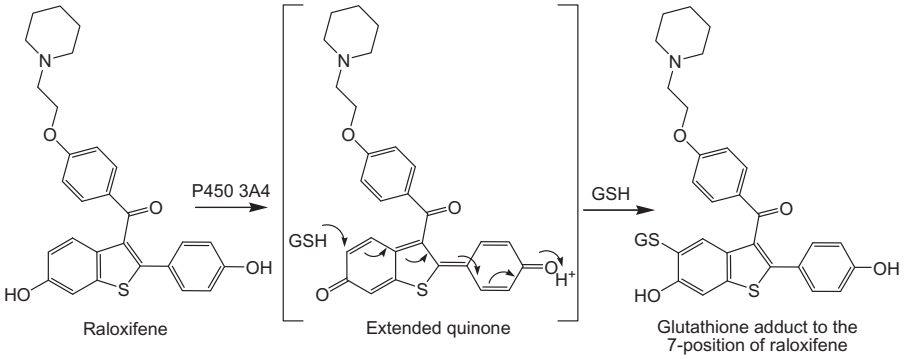


**Figure 12** Proposed scheme for the metabolic activation of the raloxifene to the chemically reactive extended quinone intermediate followed by reaction with glutathione to form the most abundant 7-glutathionyl-raloxifene adduct.

glutathione, is proposed to mediate detected time-dependent inactivation of P450 3A4, and is proposed to be responsible for zafirlukast-induced liver injury (Fig. 13) (Kassahun *et al.*, 2005). The bioactivation of zafirlukast follows a scheme very consistent with the mechanism of P450-mediated metabolic activation of the lung toxicant 3-methylindole (Yan *et al.*, 2007) (Table 3).

## 5.7 Drugs Metabolized to α,β-Unsaturated Ketones

Valproic acid (VPA, Fig. 14) is an anticonvulsant drug that has been shown to induce a rare, but sometimes fatal, hepatotoxicity characterized by microvesicular steatosis (Zimmerman and Ishak, 1982). VPA is metabolized by P450 to an unsaturated metabolite, namely 4-ene-VPA (Rettie *et al.*, 1987), which undergoes further metabolism by the enzymes of fatty acid β-oxidation, leading to the formation of the reactive

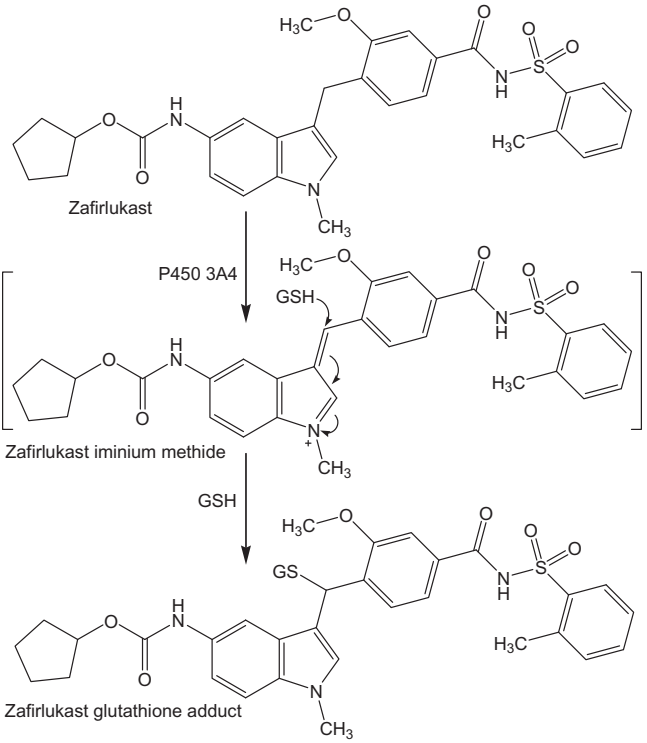


**Figure 13** Scheme for the metabolic activation of zafirlukast to an electrophilic α,β-unsaturated iminium methide intermediate and formation of the corresponding glutathione adduct.

intermediates, (E)-2,4-diene-VPA-CoA and 3-keto-4-ene-VPA-CoA. These intermediates can react with glutathione (Kassahun and Abbott, 1993), but also are proposed to inhibit fatty acid metabolism by covalent binding to, and thereby irreversibly inhibiting, fatty acid β-oxidation enzymes (Baillie, 1988; Porubek *et al.*, 1991) (Fig. 14). Studies in rat have shown that α-fluorination of 4-ene-VPA resulted in a non-hepatotoxic derivative (Tang *et al.*, 1995). Further mechanistic studies comparing VPA with its α-fluorinated derivative (F-VPA) for their abilities to form acyl-CoA thioester derivatives *in vivo* in rat liver showed that F-VPA did not form the corresponding acyl-CoA metabolite (Grillo *et al.*, 2001). From these observations, and from related studies showing the lack of toxicity due to α-fluoro-substitution, it was proposed that metabolism of VPA by acyl-CoA (see Section 5.8 and Section 4.6 of Allergic Reactions to Drugs in Chapter Allergic Reactions to Drugs) formation may mediate, at least in part, the idiosyncratic hepatotoxicity of the drug.

## 5.8 Carboxylic Acids

Zomepirac (5-[*p*-chlorobenzoyl]-1,4-dimethylpyrrole-2-acetic acid, Fig. 15) is a carboxylic acid-containing NSAID that was withdrawn from clinical use in March 1983 due to severe, and sometimes fatal, hepatotoxic reactions (Kiani and Kushner, 1983; Levy and Vasilomanolakis, 1984). Zomepirac is metabolized primarily to an unstable and

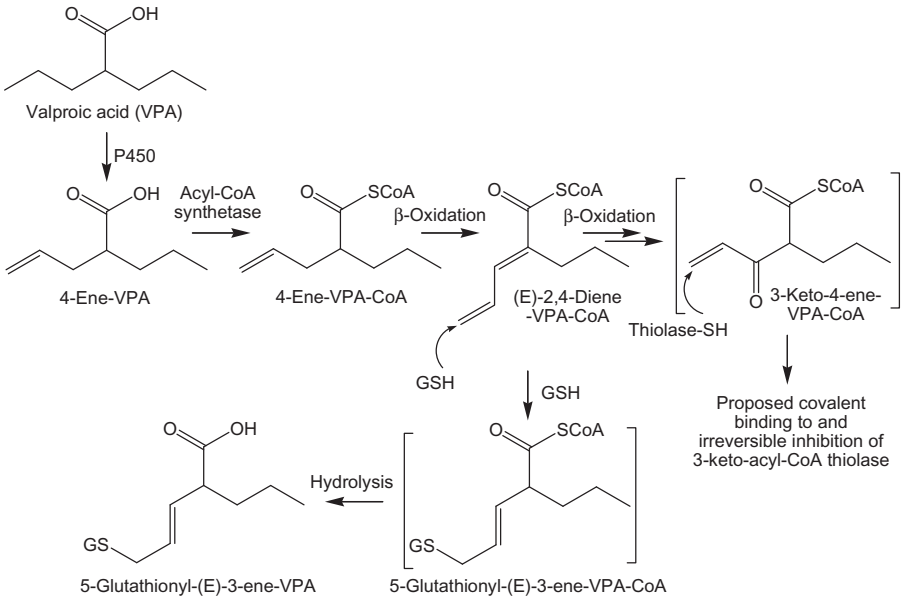


**Figure 14** Proposed scheme for the concerted metabolic activation of valproic acid by P450 and fatty acyl-CoA β-oxidation enzymes.

chemically reactive acyl glucuronide, namely zomepirac-1-*O*-acyl glucuronide (Smith, Benet and McDonagh, 1990), which has been implicated in the toxicity of the drug. Zomepirac-1-*O*-acyl glucuronide is proposed to covalently bind to protein by two different mechanisms, including transacylation-type reactions with protein nucleophiles by the 1-*O*-acyl glucuronide isomer, and through a glycation mechanism involving the reaction of protein amino-groups with open-chain aldehyde forms of the acyl migration isomers of the acyl glucuronide (Fig. 15) (Spahn-Langguth and Benet, 1992). Drug–protein adducts resulting from these reactions then are proposed to be recognized by the immune system as foreign, thereby eliciting an untoward immune-mediated hepatotoxicity (Zia-Amirhosseini, Spahn-Langguth and Benet, 1994). Experiments with zomepirac have shown that the drug becomes covalently bound to protein *in vivo* in human plasma (Smith, Benet and McDonagh, 1990) and *in vivo* in rat liver (Wang and Dickinson, 2000). Recently, zomepirac was shown to form another reactive acylating species, namely zomepirac-*S*-acyl-CoA (Olsen *et al.*, 2005), and that the acyl-CoA metabolite may be more important than the corresponding acyl glucuronide in contributing to the acylation of GSH and protein nucleophiles in liver tissue (Grillo and Hua, 2003).

## 5.9 Pyrroles

Zomepirac (Fig. 16) also contains a pyrrole-type structural alert that has been shown recently to undergo metabolic activation by P450 (Chen *et al.*, 2006). In incubations with human liver microsomes, and *in vivo* in rat, zomepirac is metabolized by P450 forming a proposed reactive epoxide intermediate that reacts with glutathione forming 5-[4′-chlorobenzoyl]-1,4-dimethyl-3-glutathionyl-pyrroleacetic acid. In these studies,

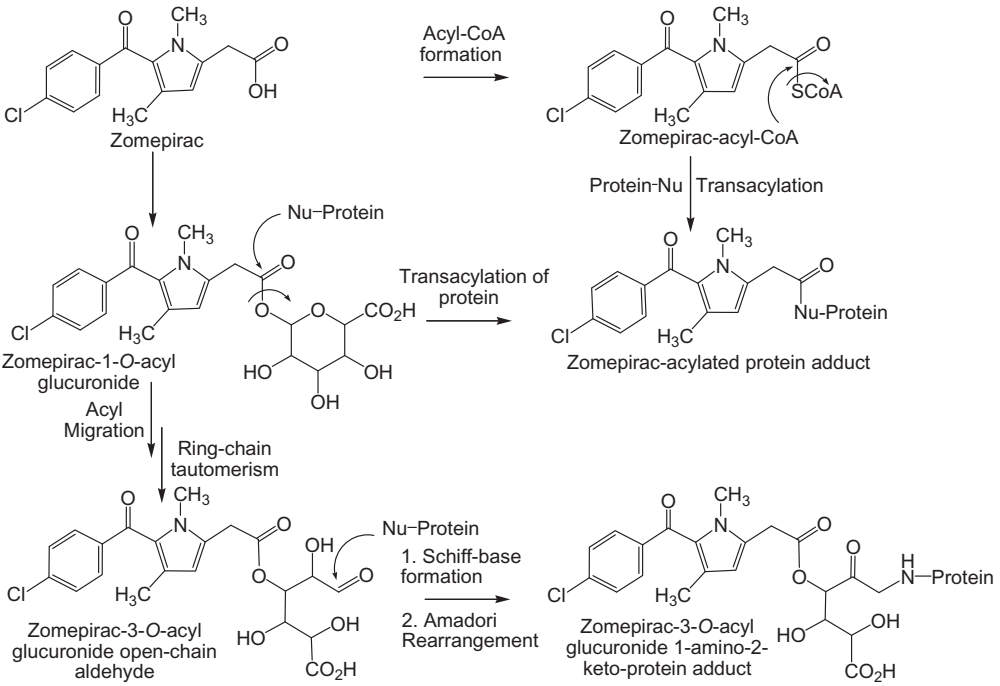


**Figure 15** Proposed scheme for the metabolic activation of the zomepirac by acyl glucuronidation and acyl-CoA formation.

the corresponding glutathione adduct was also shown to be formed for the drug tolmetin, a structurally similar pyrrole-containing NSAID. The authors from these studies proposed that, in addition to the formation of reactive acyl glucuronides, P450-dependent metabolism of zomepirac and tolmetin leads to the production of a reactive epoxide-type intermediate that may form covalent adducts with protein, and thereby contribute to potential immune-mediated hepatotoxic reactions.

## 5.10 Hydrazines

Procarbazine, *N*-isopropyl-α-(2-methylhydrazine)-*p*-toluamide hydrochloride (Fig. 17), is a hydrazine-containing drug that is an effective anticancer agent used for the treatment of several types of cancer including Hodgkin's disease (Calabresi and Chabner, 1996). Procarbazine undergoes oxidative dealkylation, via the formation of an azo-intermediate (spontaneously or catalyzed by P450), which ultimately leads to the production of a methyl diazonium reactive species which methylates, and subsequently inactivates, DNA (Horstman, Meadows and Yost, 1987; Swaffar *et al.*, 1992) (Fig. 17).

## 5.11 Hydrazides

Isoniazid (Fig. 18) is an antituberculosis drug that is well known to be able to cause drug-induced liver toxicity and that contains a hydrazide structural alert (Zimmerman,

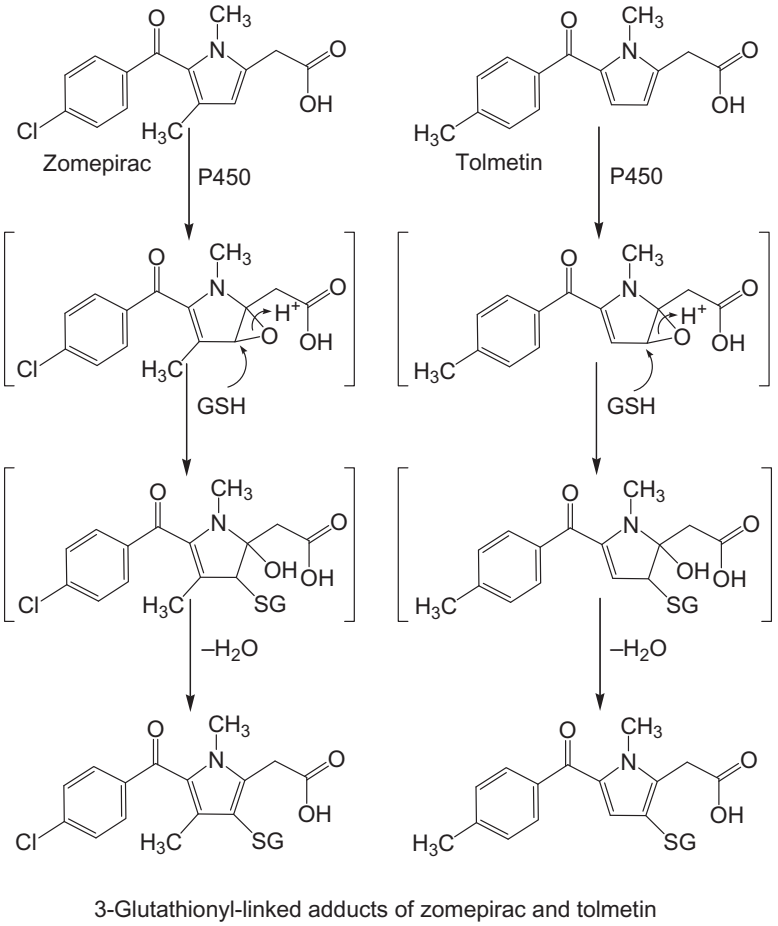


**Figure 16** Proposed scheme for the metabolic activation of the zomepirac and tolmetin by P450-mediated epoxidation of the pyrrole moiety (adapted from Chen *et al.*, 2006).

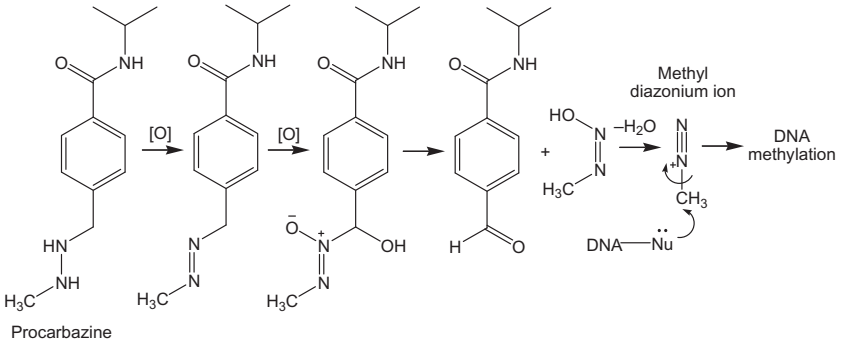


**Figure 17** Proposed scheme for the metabolic activation of procarbazine leading to the generation of a reactive methyl diazonium ion intermediate that methylates DNA nucleophiles.

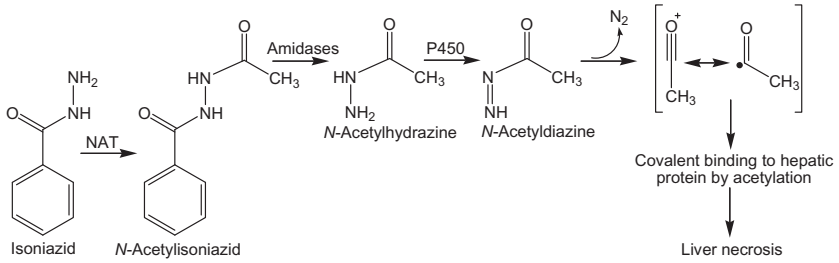


**Figure 18** Proposed mechanism in the metabolic activation of isoniazid (adapted from Timbrell *et al.*, 1980 and Kalgutkar *et al.*, 2005).

1993). The proposed mechanism of bioactivation of isoniazid involves the initial acetylation of the primary amine moiety by *N*-acetyltransferase enzymes leading to the formation of *N*-acetylisoniazid. This product then is hydrolyzed by amidase-mediated cleavage to give *N*-acetylhydrazine, which serves as a substrate for P450. The P450-oxidized product, namely *N*-acetyldiazine, is unstable and decomposes releasing nitrogen gas and forming both electrophilic and free radical-type intermediates that acetylate hepatic protein and which are proposed to induce liver necrosis (Timbrell *et al.*, 1980; Kalgutkar *et al.*, 2005).

## 5.12 Drugs That Form a Reactive Quinone Imine

***5.12.1 Acetaminophen*** Acetaminophen (*N*-acetyl-*p*-aminophenol, APAP; Fig. 19) is a commonly used over-the-counter analgesic and antipyretic medication that, when consumed in an overdose fashion, has been shown to produce serious and sometimes fatal hepatic necrosis (Nelson, 1990; Pumford and Halmes, 1997). APAP is a very safe drug when taken at therapeutic doses (~15 mg/kg), but can cause elevation in aminotransferases in healthy adults receiving 4 g (~57 mg/kg) daily (Watkins *et al.*, 2006). Centrilobular hepatic necrosis occurs at doses of 150 mg/kg and higher (>10 g). At a dose of 350 mg/kg, APAP can induce hepatic failure. There are ~9000 reported cases/year in the United States alone, and approximately 60 deaths annually (Holubek *et al.*, 2006). Many of these overdoses are due to suicide attempts, unintentional overdosing in infants, and from adverse drug interactions in alcoholics. Early on in the testing of APAP toxicity, Jollow *et al.* (1973) detected covalent binding of radioactivity to protein in hepatocytes in the centrilobular location of the liver, where liver cells became necrotic post-APAP overdose. The proposed mechanism of bioactivation of APAP is shown in Fig. 19.

The metabolic activation of APAP, which contains an anilide-type structural alert, occurs by cytochrome P450 2E1-, 2D6-, 3A4-, and 1A2-mediated biotransformation, leading to the formation of a reactive *N*-acetyl-*p*-benzoquinone imine (NAPQI) intermediate that can spontaneously, or catalytically (via glutathione *S*-transferase π), lead to the formation of a glutathione-thioether-linked conjugate detoxification product, namely 3-(glutathione-*S*-yl)APAP (Dahlin *et al.*, 1984; Nelson, 1990; Dong *et al.*, 2000) (Fig. 19). This same adduct can then be degraded to the corresponding mercapturic acid conjugate in the kidney, followed by excretion into urine (Wilson *et al.*, 1982). In an overdose situation, the detoxification of APAP by glucuronidation and/or sulfation becomes saturated such that increased amounts of drug are converted to the reactive

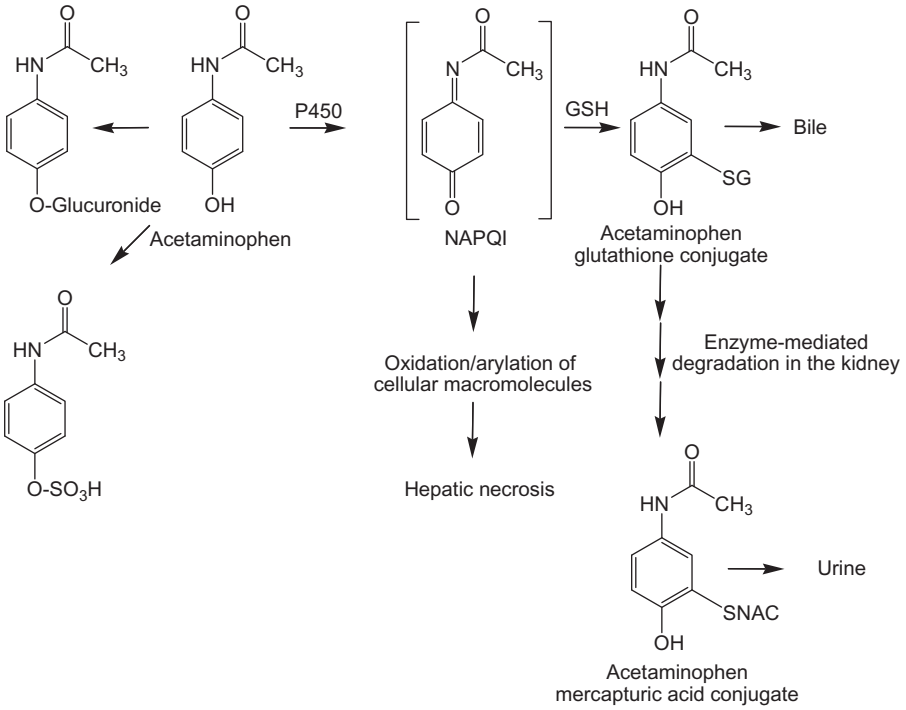


**Figure 19** Proposed mechanism in the metabolic activation of acetaminophen.

NAPQI electrophile. In addition, at overdose levels of the drug, stores of GSH can be depleted such that the amount of NAPQI formed can exhaust hepatocyte defenses through covalent binding to, and destruction of, critical proteins (Park *et al.*, 2005a). An important critical protein and part of this cellular defense is the Keap1:Nrf2 signaling pathway (Goldring *et al.*, 2004; Okawa *et al.*, 2006). Keap1 and Nrf2 form a heterodimer in cytosol of cells. Keap1 is a small protein that is rich in cysteinyl sulfhydryl groups. These free sulfhydryls are targets for alkylation, arylation, or oxidation by many electrophiles such as NAPQI. The mechanism follows on such that when reactive electrophiles bind covalently to, or oxidize, cysteine thiols of Keap1, the heterodimer dissociates, modified Keap1 is ubiquinated and degraded, and the Nrf2 moiety translocates to the nucleus where it binds to the antioxidant response element (ARE) which triggers the transcription of those genes that are cytoprotective such as glutathione *S*-transferase and glutathione peroxidase. (Fig. 20) (Lee *et al.*, 2005). Therefore, Keap1:Nrf2 is proposed to be a protective signaling pathway for reactive electrophiles and oxidizing reagents, and it is believed to be activated by APAP through metabolic activation of the drug to NAPQI (Goldring *et al.*, 2004).

Comparative mechanistic studies have been performed with 3-hydroxyacetanalide (AMAP), which is a simple positional isomer of APAP (Tirmenstein and Nelson, 1989; Nelson and Pearson, 1990; Rashed, Myers and Nelson, 1990; Roberts, Price and Jollow, 1990; Myers *et al.*, 1995). Under conditions where APAP causes hepatotoxicity, AMAP was shown to undergo metabolic activation and subsequent covalently binding to liver tissue, without eliciting hepatotoxicity in test animals. This was, and still is, a confusing

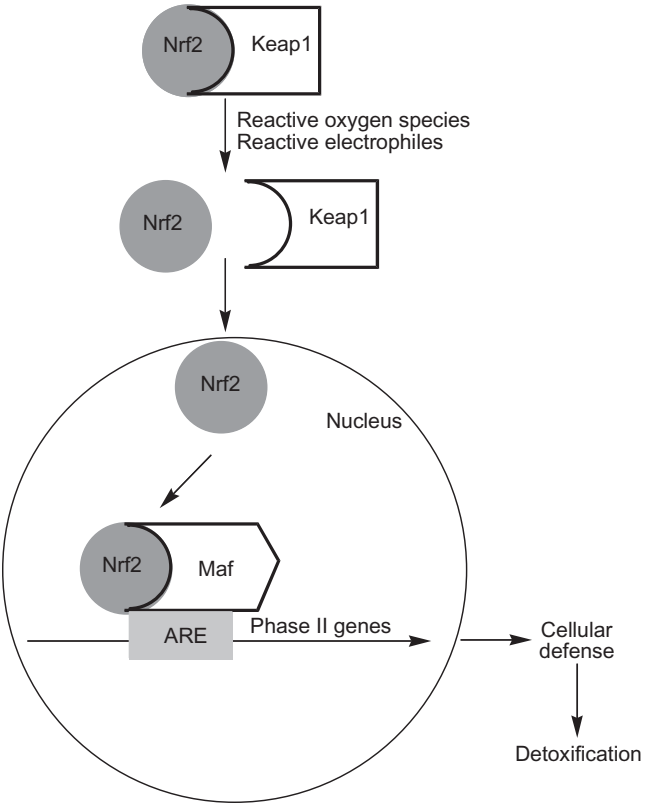


**Figure 20** Proposed mechanism of Nrf2 activation by reactive oxygen species and reactive electrophilic species (adapted from Lee *et al.*, 2005).

result since both APAP and AMAP are converted to very similar reactive electrophiles. APAP is converted to NAPQI, and AMAP is converted to a *p*-benzoquinone intermediate. Both of these reactive species have been shown to arylate cysteine residues on liver proteins. The reactive metabolites of AMAP formed in the liver covalently bind to protein at approximately the same level as acetaminophen. However, it was found that the relative distribution of covalent binding to protein in subcellular locations was different (Tirmenstein and Nelson, 1989). The AMAP regioisomer, which again elicits no toxicity, was shown to covalently bind to cytoplasmic proteins and a few endoplasmic reticulum proteins, many of which were found to be the same proteins are as determined for APAP. The major difference was in the covalent binding to mitochondria, where it was shown that APAP binds extensively to many critical proteins (Tirmenstein and Nelson, 1989). By contrast, the reactive metabolites of AMAP were not found to arylate these same critical proteins of this important organelle. It was proposed that the electrophilic metabolites of AMAP are more reactive than NAPQI formed from acetaminophen, because they bind closer to where they are formed by P450s located in the endoplasmic reticulum. These *p*-benzoquinone AMAP metabolites may be too reactive to escape reactions with protein and other nucleophiles in the endoplasmic reticulum or cytosol and therefore cannot reach, covalently bind to, and

subsequently injure critical mitochondrial proteins. Using two-dimensional gel electrophoresis and mass spectrometric methods, more than 20 APAP-labeled hepatic protein adducts were identified from [$^{14}$C]APAP-treated mice (Qiu, Benet and Burlingame, 1998). The challenge continues to be able to identify the protein targets that are critical for the onset of hepatotoxicity caused by APAP (Nelson, 1995; Qiu, Benet and Burlingame, 2001; Welch *et al.*, 2005).

***5.12.2 Diclofenac*** Diclofenac (2-[2-(2,6-dichlorophenyl)aminophenyl]ethanoic acid; Fig. 21), an NSAID, is metabolically activated by cytochrome P450 to chemically reactive species that react with endogenous nucleophiles such as glutathione and protein, and that have been proposed to play a role in the rare, but sometimes severe, hepatotoxicity associated with use of the drug (Banks *et al.*, 1995; Boelsterli, 2003; Tang, 2003) (Fig. 21). Diclofenac is used for the treatment of osteoarthritis, rheumatoid arthritis, ankylosing spondylitis, and for acute muscle pain (Small, 1989). Adverse reactions to diclofenac include enteropathy (common with long-term use), acute and chronic hepatitis (15% of patients have increased plasma transaminase levels), hemolytic anemia, and fatal anaphylaxis (i.e. IgE- or IgG-mediated immune response); however, no animal models are available to predict these toxicities (Boelsterli, 2003; Shenton, Chen and Uetrecht, 2004).

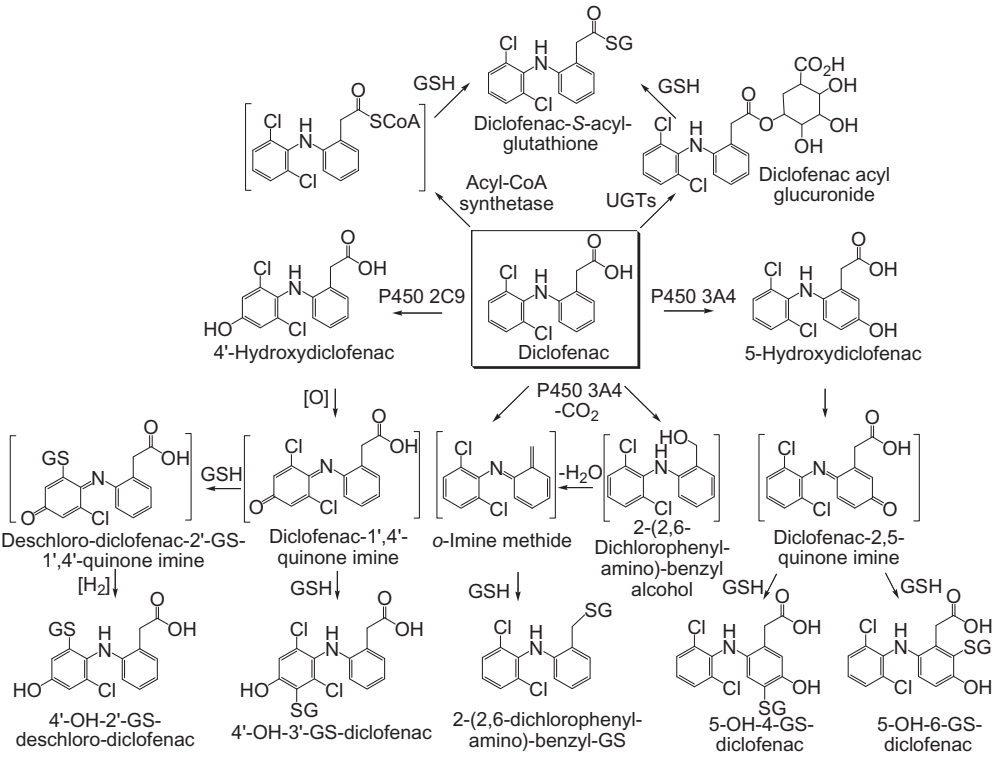


**Figure 21** Proposed mechanisms in the metabolic activation of diclofenac.

Diclofenac is metabolized by cytochrome P450s 2C9 and 3A4 by oxidation of the aromatic ring to 4′- and 5-hydroxylated products, respectively. Both of these hydroxylated metabolites can undergo further P450-mediated oxidation to electrophilic quinone imine metabolites that react, in a Michael acceptor fashion, with glutathione forming derivatives easily detected by LC-MS/MS analysis (Tang *et al.*, 1999). An additional glutathione conjugate of diclofenac was identified as 4′-hydroxy-2′-glutathione-deschloro-diclofenac, which is formed from the reaction of diclofenac-1′,4′-quinone imine, the product of 4′-hydroxydiclofenac oxidation, with glutathione via an *ipso* substitution at the carbon attached to the chlorine atom (Yu, Chen and Deninno, 2005). The formation of this glutathione adduct was shown to be mediated by P450 2C9 bioactivation of diclofenac (Yan *et al.*, 2007) (Fig. 21). Yet another glutathione was recently detected *in vitro* in rat and human hepatocyte incubations, and rat and human liver microsomes, as well as *in vivo* as detected in bile from diclofenac-treated rats (Grillo *et al.*, 2008; Teffera *et al.*, 2008). This adduct, namely 2-(2,6-dichlorophenylamino)-benzyl-*S*-thioether glutathione is proposed to be formed from P450-mediated oxidative decarboxylation of diclofenac leading to an intermediate 2-(2,6-dichlorophenylamino)-benzyl-carbon centered radical, which could either recombine with hydroxyl radical forming 2-(2,6-dichlorophenylamino)benzyl alcohol and undergo dehydration leading to the *o*-imine methide, or undergo further oxidation leading directly to the *o*-imine methide reactive intermediate. Reaction of the *o*-imine methide species at the benzylic position with glutathione then forms the stable conjugate. Therefore, diclofenac undergoes varied types of P450-mediated bioactivation mechanisms leading to reactive metabolite formation and hence potential toxicity. Diclofenac, in standardized covalent binding assays in human liver microsomes, becomes irreversibly bound to protein at a level of ~300 pmol equivalents/mg protein/h (Evans *et al.*, 2004). The reactive metabolite responsible for the bulk of the covalent binding to protein *in vitro* in incubations with human liver microsome remains to be determined.

In addition to metabolism by P450 enzymes, diclofenac is metabolized by acyl glucuronidation to diclofenac-1-*O*-acyl glucuronide (Fig. 21) (Boelsterli, 2003), an unstable and chemically reactive metabolite (Grillo *et al.*, 2003a) that has also been implicated as playing a role in diclofenac-induced hepatotoxicity (Kretz-Rommel and Boelsterli, 1993; Boelsterli, 2003). Acyl glucuronide metabolites of acidic drugs are proposed to bind covalently to protein by two different mechanisms (Spahn-Langguth and Benet, 1992) (see Section 5.8). These include transacylation-type reactions with protein nucleophiles by the 1-*O*-acyl glucuronide isomer, and by a glycation mechanism that involves the reaction of open-chain aldehyde forms of the acyl migration glucuronide isomers with protein amino groups (Ding *et al.*, 1993; Zia-Amirhosseini, Spahn-Langguth and Benet, 1994). Then it is proposed that the drug–protein adducts are recognized by the immune system as foreign, resulting in an immune response leading to the associated idiosyncratic hepatotoxicity (Spahn-Langguth and Benet, 1992). Diclofenac acyl glucuronide was shown to react irreversibly with the canalicular protein dipeptidyl peptidase IV, and which corresponded to a decrease in functional activity of this protein in diclofenac-treated rats (Hargus *et al.*, 1995).

The *S*-acyl-glutathione thioester-linked metabolite, namely diclofenac-*S*-acyl-glutathione, has been detected *in vivo* in rats and *in vitro* in rat and human hepatocyte incubations (Fig. 21) (Grillo *et al.*, 2003a, b). It was shown that inhibition of diclofenac acyl glucuronide production in incubations with rat hepatocytes had no effect on diclofenac-*S*-acyl-glutathione formation. Therefore, another potential reactive

metabolite, namely diclofenac-*S*-acyl-CoA thioester, was proposed. Acyl-CoA metabolites of carboxylic acid-containing drugs are being increasingly investigated as potentially toxicologically important reactive metabolites of carboxylic acid-containing drugs (Li, Benet and Grillo, 2002; Sidenius *et al.*, 2004; Olsen *et al.*, 2007).

In summary, chemically reactive metabolites of diclofenac have been proposed to mediate the idiosyncratic hepatotoxicity associated with the use of the drug (Boelsterli, 2003; Tang, 2003), but which one if any is important in this respect remains to be answered. One recent study showed that glucuronidation-inhibited rat hepatocytes were more susceptible to diclofenac-induced cytotoxicity than vehicle-treated cells (Siraki, Chevaldina and O'Brien, 2005). Metabolic activation studies in rat hepatocytes showed that diclofenac acyl glucuronide was the most important mediator of covalent binding to protein. However, cytotoxicity was attributed primarily to P450-mediated oxidative products (Kretz-Rommel and Boelsterli, 1993).

A proteomic profiling study was conducted for diclofenac to examine the effects of the drug on proteins excreted in rat bile (Jones *et al.*, 2003). Six proteins were detected in bile from diclofenac-treated rats that were not present in untreated animals. These proteins included dipeptidyl peptidase IV, dihydropyrimidinase, apoptosis-inducing factor, α-1-antitrypsin precursor, vitamin D-binding protein prepeptide, and 60S ribosomal protein L4. The most important change observed was the increased presence of dipeptidyl peptidase IV, which is an integral membrane protein having a homodimeric mass of 110 kDa (Hargus *et al.*, 1995). However, no data were collected that confirmed the presence of diclofenac-derived protein adducts excreted in bile in the same study. To date, convincing evidence for the causative hepatotoxic reactive metabolite(s) of diclofenac remains elusive (Tang, 2003).

## 5.13 Alkenes

4-Vinylcyclohexene (Fig. 22) is a chemical that contains the alkene structural alert and is used in the manufacture of flame-retardants, insecticides, and plasticizers. Exposure of mice to 4-vinylcyclohexene results in a loss of small preantral (primordial and primary) follicles in the mouse ovary (Rajapaksa *et al.*, 2007). 4-Vinylcyclohexene-

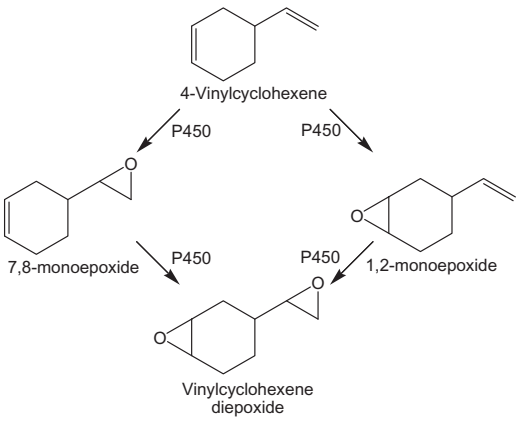


**Figure 22** Proposed mechanisms in the metabolic activation of the ovotoxin 4-vinylcyclohexene (adapted from Rajapaksa *et al.*, 2007).

induced ovarian toxicity has been attributed to bioactivation of the compound to a chemically reactive diepoxide metabolite (Fig. 22). Therefore, this occupational chemical is metabolically activated to either a 1,2- or 7,8-monoepoxide, and subsequently to the ovotoxic metabolite vinylcyclohexene diepoxide via cytochrome P450 enzymes. Results from mechanistic structure–activity studies suggest that bioactivation to the diepoxide is necessary for vinylcyclohexene-induced ovarian toxicity (Rajapaksa *et al.*, 2007).

## 5.14 Alkynes

17α-Ethynylestradiol is an estrogenic oral contraceptive that contains an alkyne structural alert moiety (Fig. 23). 17α-Ethynylestradiol is known to inactivate cytochrome P450 in a mechanism-based fashion. Bioactivation is proposed to occur via P450-mediated oxidation to 17α-oxirene-related reactive intermediates (oxirene and ketene species; Fig. 23) that react with and form heme and apoprotein adducts, resulting in the inactivation of P450 3A5. In addition, the compound was shown to form two glutathione adducts in incubations with the reconstituted cytochrome P450 3A5 system (Lin and Hollenberg, 2007).

## 5.15 Anilines

Dapsone (diamino-diphenyl sulphone; Fig. 24) is an antimicrobial drug used for the treatment of leprosy (Mandell and Petri, 1996). Use of dapsone has been associated

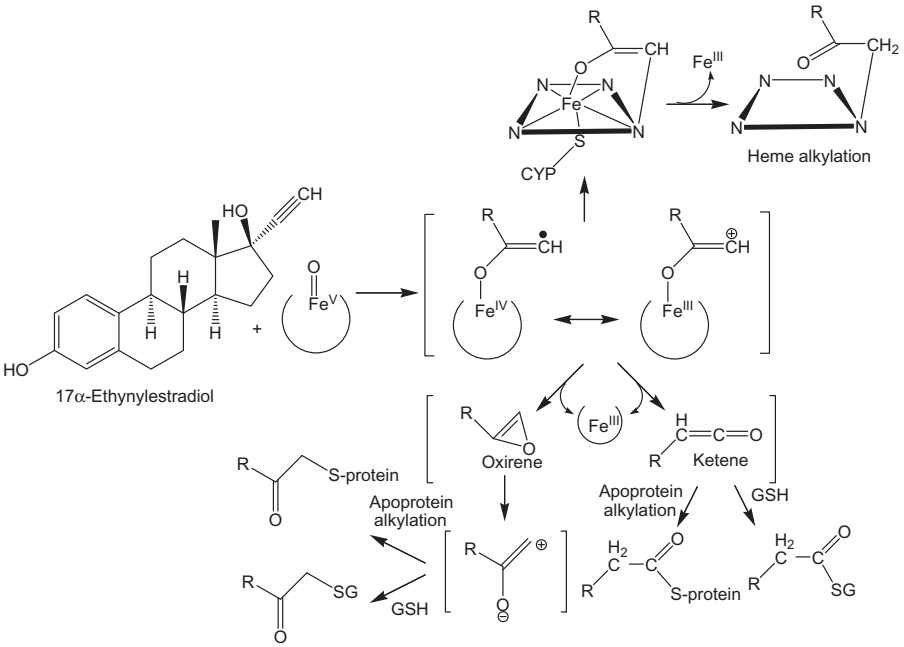


**Figure 23** Proposed mechanisms in the metabolic activation of the estrogenic oral contraceptive 17α-ethynylestradiol to reactive oxirene- and ketene-type intermediates (adapted from Kalgutkar *et al.*, 2005).

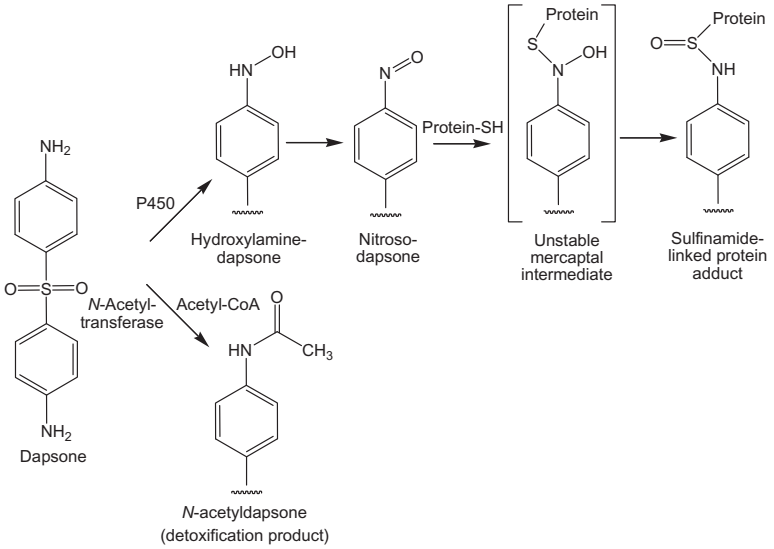


**Figure 24** Proposed mechanism in the metabolic activation of dapsone to a reactive aromatic nitroso-type intermediate (adapted from Vyas *et al.*, 2006).

with the formation of allergic reactions sometimes leading to hepatitis (Joseph, 1985). The chemical structure of dapsone contains the aniline structural alert which is metabolized by P450-mediated oxidation to an *N*-hydroxylated product. This metabolite then can be further oxidized to a chemically reactive nitroso product, which is known to react with protein nucleophiles forming drug–protein adducts that can serve as antigens that trigger hypersensitivity reactions (Vyas *et al.*, 2006) (Fig. 24). The *N*-hydroxylated product might also serve as a substrate for sulfation leading to the formation of a reactive hydroxylamine-sulfate conjugate (see Section 5.16). Another route of dapsone metabolism occurs by acetylation of the aromatic amine moiety which is mediated by *N*-acetyltransferase (utilizing the cofactor acetyl-CoA) leading to the formation of a nontoxic metabolite, namely *N*-acetyldapsone. It has been proposed that the relative extents of these two routes of metabolism occurring in a patient undergoing dapsone therapy could determine the onset of an adverse allergic reaction (Bluhm *et al.*, 1999).

## 5.16 Nitroaromatics

The nitroaromatic moiety is another structural alert that has been associated with the induction of liver toxicity by some drugs. One proposed pathway of bioactivation of the aromatic nitro group is through reductive metabolism generating nitroanion and hydronitroxide radicals which can lead to DNA damage (Hewick, 1982) (Fig. 25). Another bioactivation mechanism occurs by additional one-electron reduction of the nitroanion radical to a nitroso-benzene intermediate which reacts with protein nucleophiles, leading to the formation of sulfinamide-linked protein adducts (Fig. 25). Additional one-electron reduction of the nitroso-benzene intermediate leads to a hydroxyl-amine intermediate which can be further metabolized to a reactive hydroxylamine-sulfate conjugate that is able to react with and damage DNA

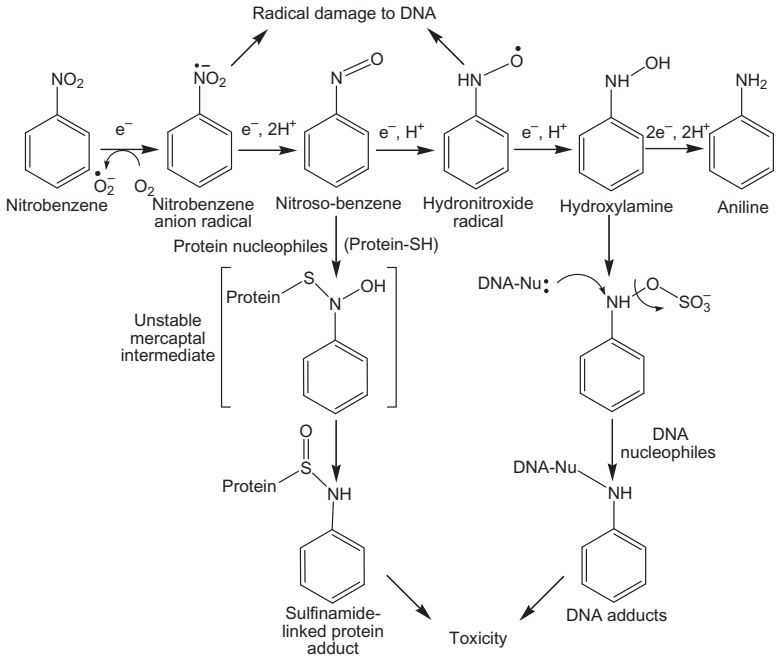


**Figure 25** Proposed mechanism in the metabolic activation of nitrobenzene to reactive aromatic hydroxylamine- and nitroso-type intermediates (adapted from Hewick, 1982).

(Kalgutkar *et al.*, 2005). Nitroaromatic-containing drugs known to be associated with toxicity, including rare cases of idiosyncratic liver injury, are tolcapone, nimesulide, nilutamide, flutamide, and nitrofurantoin (Boelsterli *et al.*, 2006). The reduction, and hence bioactivation, of the nitro group can be catalyzed P450 enzymes, xanthine and aldehyde oxidases, and quinone reductases (Kedderis, Argenbright and Miwa, 1989; Boelsterli *et al.*, 2006).

## 6 CONCLUSIONS

Pharmaceutical scientists working in drug metabolism and toxicology today consider reactive metabolites of drugs or drug candidates an important potential toxicological issue (Kalgutkar and Soglia, 2005; Park *et al.*, 2005b; Baillie, 2006; Doss and Baillie, 2006). Many pharmaceutical companies are dealing with this concern by evaluating data harvested from varied types of metabolic activation studies. Experimental approaches (from glutathione trapping of reactive intermediates to covalent binding of drug-related radioactivity to protein) to assess drug candidates for their ability to form reactive intermediates, both *in vitro* and *in vivo*, are performed routinely by many pharmaceutical companies today (Kalgutkar and Soglia, 2005; Mutlib *et al.*, 2005; Baillie, 2006). Skilled in ferreting out glutathione adducts, and probable structures of reactive intermediate, as well as in modifying drug candidates accordingly, drug metabolism scientists know full well that no matter how capable we have become at describing and finding reactive metabolites, we still struggle to understand why an individual

comes down with a toxic reaction to a certain drug. The causes of toxicity are still difficult to decipher, even when using up-to-date technologies. Nonetheless, it is hoped that by having closer interactions between the fields of drug metabolism, medicinal chemistry, biology, and toxicology that it will play a key role in eliminating potentially toxic compounds from entering development. Importantly, mechanistic metabolic activation studies should be performed prior to the selection of drug candidates due to the tremendous economic investment in further stages of development (Guengerich and MacDonald, 2007; Kumar *et al.*, 2008). It is an ongoing goal in the pharmaceutical industry to decrease the metabolic activation properties of drug candidates in the hope that this may increase the probability of compounds being successfully developed as nontoxic drugs (Baillie, 2006). With the recognition of metabolic activation occurring for a candidate drug, pharmaceutical companies will have to make knowledgeable judgments about potential risks to patients so that toxic drugs are never released onto the market.

In terms of emerging trends, there will be a substantial application of mass spectrometry-based proteomic techniques to identify critical targets of reactive metabolites (Qiu, Benet and Burlingame, 1998; Liebler, 2002; Liebler, 2006). The important challenge then is to be able to use these new techniques to assist in providing convincing evidence that modification of pivotal proteins may have toxic consequences. Coincident advancement of toxicogenomic methodologies, in terms of observing both covalent binding data in juxtaposition with gene expression data, should be very promising in determining the toxic significance of covalent binding (Guengerich and MacDonald, 2007; Leone *et al.*, 2007).

Representative indicators of cellular electrophilic exposure related to the Keap1/Nrf2 signaling system, which is important for triggering cellular defense mechanisms (Enomoto *et al.*, 2001; Dinkova-Kostova, Holtzclaw and Kensler, 2005; Hong *et al.*, 2005), are increasingly being studied. Keap1/Nrf2 could be used as a surrogate marker of exposure to electrophilic intermediates or to oxidative stress that occurs within the cell (Hong *et al.*, 2005). It is envisioned that the accumulation of Nrf2 as a marker for the evaluation of potential toxicity could be developed in order to assess compounds for their potential to trigger a warning signal (Baillie, 2006).

The development of computer simulation techniques to anticipate chemical structures susceptible to bioactivation is becoming more effective (Roberts, 2001; Gombar, Silver and Zhao, 2003). Structural alerts are functional groups of molecules that pharmaceutical chemists attempt not to use because these moieties have a background of forming chemically reactive, potentially toxic intermediates. Hopefully, in the near future, drug metabolism scientists will be able to predict with *in silico* tools, together with knowledge of the tertiary structures of the various enzymes and the electronic characteristics of substrates, the tendency of a molecule to be metabolized by certain enzymes (Korzekwa and Jones, 1993) to chemically reactive species.

Finally, there is a critical demand for increased research on the mechanisms of immune-mediated toxicities at the molecular level (Baillie, 2006; Liebler, 2006; Guengerich and MacDonald, 2007; Kumar *et al.*, 2008; Uetrecht, 2008). The maturation of consilient approaches toward the use of new technologies which are finding immediate applications includes metabonomics (Nicholson *et al.*, 2002; Mortishire-Smith *et al.*, 2004), proteomics (Jones *et al.*, 2003; Bradshaw and Burlingame, 2005; Welch *et al.*, 2005), and toxicogenomics (Martin *et al.*, 2006; Leone *et al.*, 2007), each being brought together to aid in someday understanding the implications of metabolic activation in drug-mediated toxicities (Baillie, 2006, 2008; Liebler, 2006; Tang, 2007).

## ACKNOWLEDGMENTS

I would like to thank Dr. Christian Skonberg (Department of Pharmaceutics and Analytical Chemistry, University of Copenhagen, Denmark), Dr. Kathila Rajapaksa (Toxicology, Amgen Inc.), and Judy Huang, B.A. (Pharmacokinetics and Drug Metabolism, Amgen Inc.) for critically reviewing this chapter. I also thank the constructive criticism and helpful suggestions of two anonymous reviewers, which have been included in the references.

## REFERENCES

Argoti, D., Liang, L., Conteh, A., Chen, L., Bershas, D., Yu, C.P., Vouros, P., Yang, E. (2005). Cyanide trapping of iminium ion reactive intermediates followed by detection and structure identification using liquid chromatography-tandem mass spectrometry (LC-MS/MS). *Chem. Res. Toxicol.*, *18*, 1537–1544.

Baer, B.R., Wienkers, L.C., Rock, D.A. (2007). Time-dependent inactivation of P450 3A4 by raloxifene: identification of Cys239 as the site of apoprotein alkylation. *Chem. Res. Toxicol.*, *20*, 954–964.

Baillie, T.A. (1988). Metabolic activation of valproic acid and drug-mediated hepatotoxicity. Role of the terminal olefin, 2-n-propyl-4-pentenoic acid. *Chem. Res. Toxicol.*, *1*, 195–199.

Baillie, T.A. (2006). Future of toxicology-metabolic activation and drug design: challenges and opportunities in chemical toxicology. *Chem. Res. Toxicol.*, *19*, 889–893.

Baillie, T.A. (2008). Metabolism and toxicity of drugs. Two decades of progress in industrial drug metabolism. *Chem. Res. Toxicol.*, *21*, 129–137.

Baillie, T.A., Davis, M.R. (1993). Mass spectrometry in the analysis of glutathione conjugates. *Biol. Mass Spectrom.*, *22*, 319–325.

Baillie, T.A., Kassahun, K. (2001). Biological reactive intermediates in drug discovery and development: a perspective from the pharmaceutical industry. *Adv. Exp. Med. Biol.*, *500*, 45–51.

Bambal, R.B., Hanzlik, R.P. (1995). Bromobenzene 3,4-oxide alkylates histidine and lysine side chains of rat liver proteins in vivo. *Chem. Res. Toxicol.*, *8*, 729–735.

Banks, A.T., Zimmerman, H.J., Ishak, K.G., Harter, J.G. (1995). Diclofenac-associated hepatotoxicity: analysis of 180 cases reported to the Food and Drug Administration as adverse reactions. *Hepatology*, *22*, 820–827.

Berenblum, I., Schoental, R. (1947). Carcinogenic constituents of coal tar. *Br. J. Cancer*, *1*, 157–165.

Bluhm, R.E., Adedoyin, A., McCarver, D.G., Branch, R.A. (1999). Development of dapsone toxicity in patients with inflammatory dermatoses: activity of acetylation and hydroxylation of dapsone as risk factors. *Clin. Pharmacol. Ther.*, *65*, 598–605.

Boelsterli, U.A. (2003). Diclofenac-induced liver injury: a paradigm of idiosyncratic drug toxicity. *Toxicol. Appl. Pharmacol.*, *192*, 307–322.

Boelsterli, U.A., Ho, H.K., Zhou, S., Leow, K.Y. (2006). Bioactivation and hepatotoxicity of nitroaromatic drugs. *Curr. Drug Metab.*, *7*, 715–727.

Boyd, M.R. (1976). Role of metabolic activation in the pathogenesis of chemically induced pulmonary disease: mechanism of action of the lung-toxic furan, 4-ipomeanol. *Environ. Health Perspect.*, *16*, 127–138.

Bradshaw, R.A., Burlingame, A.L. (2005). From proteins to proteomics. *IUBMB Life*, *57*, 267–272.

Brookes, P., Lawley, P.D. (1964). Evidence for the binding of polynuclear aromatic hydrocarbons to the nucleic acids of mouse skin: relation between carcinogenic power of hydrocarbons and their binding to DNA. *Nature (Lond.)*, *202*, 781–784.

Buonarati, M., Jones, A.D., Buckpitt, A. (1990). In vivo metabolism of isomeric naphthalene oxide glutathione conjugates. *Drug Metab. Dispos.*, *18*, 183–189.

Calabresi, P., Chabner, B.A. (1996). Chemotherapy of nepoplastic diseases, in *Goodman and Gilman's The Pharmacological Basis of Therapeutics*, 9th edn (eds M.J. Wonsiewicz and P. McCurdy), McGraw-Hill, New York, pp. 1225–1287.

Calvey, N. (2005). Adverse drug reactions. *Anaesth. Int. Care Med.*, *6*, pp. 245–249.

Carlson, R.M. (1990). Assessment of the propensity for covalent binding of electrophiles to biological substrates. *Environ. Health Perspect.*, *87*, 227–232.

Chen, Q., Doss, G.A., Tung, E.C., Liu, W., Tang, Y.S., Braun, M.P., Didolka, V., Strauss, J.R., Wang, R.W., Stearns, R.A., Evans, D.C., Baillie, T.A., Tang, W. (2006). Evidence for the bioactivation of zomepirac and tolmetin by an oxidative pathway: identification of glutathione adducts in vitro in human liver microsomes and in vivo in rats. *Drug Metab. Dispos.*, *34*, 145–151.

Chen, Q., Ngui, J.S., Doss, G.A., Wang, R.W., Cai, X., DiNinno, F.P., Blizzard, T.A., Hammond, M.L., Stearns, R.A., Evans, D.C., Baillie, T.A., Tang, W. (2002). Cytochrome P450 3A4-mediated bioactivation of raloxifene: irreversible enzyme inhibition and thiol adduct formation. *Chem. Res. Toxicol.*, *15*, 907–914.

Clark, C.R., Henderson, T.R., Royer, R.E., Brooks, A.L., McClellan, R.O., Marshall, W.F., Naman, T.M. (1982). Mutagenicity of diesel exhaust particle extracts: influence of fuel composition in two diesel engines. *Fundam. Appl. Toxicol.*, *2*, 38–43.

Cohen, S.D., Pumford, N.R., Khairallah, E.A., Boekelheide, K., Pohl, L.R., Amouzadeh, H.R., Hinson, J.A. (1997). Selective protein covalent binding and target organ toxicity. *Toxicol. Appl. Pharmacol.*, *143*, 1–12.

Cook, J.W., Hewett, C.R., Hieger, I. (1933). The isolation of cancer-producing hydrocarbon from coal tar. *J. Chem. Soc. (Lond.)*, *1*, 395–405.

Dahlin, D.C., Miwa, G.T., Lu, A.Y., Nelson, S.D. (1984). N-acetyl-p-benzoquinone imine: a cytochrome P-450-mediated oxidation product of acetaminophen. *Proc. Natl. Acad. Sci. U.S.A.*, *81*, 1327–1331.

Davis, D. (2007). *The Secret History of The War on Cancer*, Basic Books, New York, p. 27.

Day, S.H., Mao, A., White, R., Schulz-Utermoehl, T., Miller, R., Beconi, M.G. (2005). A semi-automated method for measuring the potential for protein covalent binding in drug discovery. *J. Pharmacol. Toxicol. Methods*, *52*, 278–285.

Degen, G.H., Neumann, H.G. (1978). The major metabolite of aflatoxin B1 in the rat is a glutathione conjugate. *Chem. Biol. Interact.*, *22*, 239–255.

Dieckhaus, C.M., Fernández-Metzler, C.L., King, R., Krolikowski, P.H., Baillie, T.A. (2005). Negative ion tandem mass spectrometry for the detection of glutathione conjugates. *Chem. Res. Toxicol.*, *18*, 630–638.

Ding, A., Ojingwa, J.C., McDonagh, A.F., Burlingame, A.L., Benet, L.Z. (1993), Evidence for covalent binding of acyl glucuronides to serum albumin via an imine mechanism as revealed by tandem mass spectrometry. *Proc. Natl. Acad. Sci. U.S.A.*, *90*, 3797–3801.

Dinkova-Kostova, A.T., Holtzclaw, W.D., Kensler, T.W. (2005). The role of Keap1 in cellular protective responses. *Chem. Res. Toxicol.*, *18*, 1779–1791.

Dong, H., Haining, R.L., Thummel, K.E., Rettie, A.E., Nelson, S.D. (2000). Involvement of human cytochrome P450 2D6 in the bioactivation of acetaminophen. *Drug Metab. Dispos.*, *28*, 1397–1400.

Doss, G.A., Baillie, T.A. (2006). Addressing metabolic activation as an integral component of drug design. *Drug Metab. Rev.*, *38*, 641–649.

Duncan, M., Brookes, P., Dipple, A. (1969). Metabolism and binding to cellular macromolecules of a series of hydrocarbons by mouse embryo cells in culture. *Int. J. Cancer*, *4*, 818–819.

Enomoto, A., Itoh, K., Nagayoshi, E., Haruta, J., Kimura, T., O'Connor, T., Harada, T., Yamamoto, M. (2001). High sensitivity of Nrf2 knockout mice to acetaminophen hepatotoxicity associated with decreased expression of ARE-regulated drug metabolizing enzymes and antioxidant genes. *Toxicol. Sci.*, *59*, 169–177.

Evans, D.C., Baillie, T.A. (2005). Minimizing the potential for metabolic activation as an integral part of drug design. *Curr. Opin. Drug Discov. Devel.*, *8*, 44–50.

Evans, D.C., Watt, A.P., Nicoll-Griffith, D.A., Baillie, T.A. (2004). Drug-protein adducts: an industry perspective on minimizing the potential for drug bioactivation in drug discovery and development. *Chem. Res. Toxicol.*, *17*, 3–16.

Fakjian, N., Buckpitt, A.R. (1984). Metabolism of bromobenzene to glutathione adducts in lung slices from mice treated with pneumotoxicants. *Biochem. Pharmacol.*, *33*, 1479–1486.

Fung, M., Thornton, A., Mybeck, K., Wu, J.H., Hornbuckle, K., Muniz, E. (2001). Evaluation of the characteristics of safety withdrawal of prescription drugs from world wide pharmaceutical markets-1960 to 1999. *Drug Inform. J.*, *35*, 293–317.

Gan, J., Harper, T.W., Hsueh, M.M., Qu, Q., Humphreys, W.G. (2005). Dansyl glutathione as a trapping agent for the quantitative estimation and identification of reactive metabolites. *Chem. Res. Toxicol.*, *18*, 896–903.

Gelboin, H.V.A. (1969). Microsomal-dependent binding of benzo(a)pyrene to DNA. *Cancer Res.*, *29*, 1272–1276.

Gillette, J.R. (1981). An integrated approach to the study of chemically reactive metabolites of acetaminophen. *Arch. Intern. Med.*, *23*, 375–379.

Gillette, J.R., Pohl, L.R. (1977). A prospective on covalent binding and toxicity. *J. Toxicol. Environ. Health*, *2*, 849–871.

Goldring, C.E., Kitteringham, N.R., Elsby, R., Randle, L.E., Clement, Y.N., Williams, D.P., McMahon, M., Hayes, J.D., Itoh, K., Yamamoto, M., Park, B.K. (2004). Activation of hepatic Nrf2 in vivo by acetaminophen in CD-1 mice. *Hepatology*, *39*, 1267–1276.

Gombar, V.K., Silver, I.S., Zhao, Z. (2003). Role of ADME characteristics in drug discovery and their in silico evaluation: in silico screening of chemicals for their metabolic stability. *Curr. Top. Med. Chem.*, *3*, 1205–1225.

Greene, F.L. (1990). Environmental carcinogens–Percivall Pott revisited. *J. S. C. Med. Assoc.*, *86*, 328.

Gregus, Z., Klaassen, C.D. (2001). Mechanisms of toxicity, in *Casarett and Doull's Toxicology (2001) The Basic Science of Poisons*, 6th edn (ed. C.D. Klaassen), The McGraw-Hill Company, Inc., pp. 35–82.

Grillo, M.P., Chiellini, G., Tonelli, M., Benet, L.Z. (2001). Effect of alpha-fluorination of valproic acid on valproyl-S-acyl-CoA formation in vivo in rats. *Drug Metab. Dispos.*, *29*, 1210–1215.

Grillo, M.P., Hua, F. (2003). Identification of zomepirac-S-acyl-glutathione in vitro in incubations with rat hepatocytes and in vivo in rat bile. *Drug Metab. Dispos.*, *31*, 1429–1436.

Grillo, M.P., Knutson, C.G., Sanders, P.E., Waldon, D.J., Hua, F., Ware, J.A. (2003a). Studies on the chemical reactivity of diclofenac acyl glucuronide with glutathione: identification of diclofenac-S-acyl-glutathione in rat bile. *Drug Metab. Dispos.*, *31*, 1327–1336.

Grillo, M.P., Hua, F., Knutson, C.G., Ware, J.A., Li, C. (2003b). Mechanistic studies on the bioactivation of diclofenac: identification of diclofenac-S-acyl-glutathione in vitro in incubations with rat and human hepatocytes. *Chem. Res. Toxicol.*, *16*, 1410–1417.

Grillo, M.P., Ma, J., Teffera, Y., Waldon, D.J. (2008). A novel bioactivation pathway for diclofenac initiated by P450-mediated oxidative decarboxylation. *Drug Metab. Dispos.*, *36*, 1740–1744.

Guengerich, F.P., MacDonald, J.S. (2007). Applying mechanisms of chemical toxicity to predict drug safety. *Chem. Res. Toxicol.*, *20*, 344–369.

Hargus, S.J., Martin, B.M., George, J.W., Pohl, L.R. (1995). Covalent modification of rat liver dipeptidyl peptidase IV (CD26) by the nonsteroidal anti-inflammatory drug diclofenac. *Chem. Res. Toxicol.*, *8*, 993–996.

Hart, D., Ward, M., Lifschitz, M.D. (1987). Suprofen-related nephrotoxicity. A distinct clinical syndrome. *Ann. Intern. Med.*, *106*, 235–238.

Heidelberger, C. (1975). Chemical carcinogenesis. *Ann. Rev. Biochem.*, *44*, 79–121.

Hewick, D.S. (1982). Reductive metabolism of nitrogen-containing functional groups, in *Metabolic Basis of Detoxication: Metabolism of Functional Groups* (eds W.B. Jakoby, J.R. Bend and J. Caldwell), Academic Press, Inc., New York, pp. 151–170.

Holubek, W.J, Kalman, S., and Hoffman, R.S. (2006). Acetaminophen-induced acute liver failure: results of a United States multicenter, prospective study. *Hepatology*, *43*, 880–882.

Hong, F., Sekhar, K.R., Freeman, M.L., Liebler, D.C. (2005). Specific patterns of electrophile adduction trigger Keap1 ubiquitination and Nrf2 activation. *J. Biol. Chem.*, *280*, 31768–31775.

Horstman, M.G., Meadows, G.G., Yost, G.S. (1987). Separate mechanisms for procarbazine spermatotoxicity and anticancer activity. *Cancer Res.*, *47*, 1547–1550.

IARC (1985). Polynuclear aromatic 4. bitumens, coal tars and derived products, shale oils and soots. IARC Monographs on the Evaluation of Carcinogenic Risk of Chemicals to Humans, International Agency for Research on Cancer, pp. 271.

Isley, W.L. (2003). Hepatotoxicity of thiazolidinediones. *Expert Opin. Drug Saf.*, *2*, 581–586.

Jackson, E.K. (1996). Diuretics, in *Goodman and Gilman's. The Pharmacological Basis of Therapeutics*, 9th edn (eds M.J. Wonsiewicz and P. McCurdy), McGraw-Hill, New York, pp. 685–713.

Jollow, D.J., Mitchell, J.R., Potter, W.Z., Davis, D.C., Gillette, J.R., Brodie, B.B. (1973). Acetaminophen-induced hepatic necrosis. II. Role of covalent binding in vivo. *J. Pharmacol. Exp. Ther.*, *187*, 195–202.

Jones, J.A., Kaphalia, L., Treinen-Moslen, M., Liebler, D.C. (2003). Proteomic characterization of metabolites, protein adducts, and biliary proteins in rats exposed to 1,1-dichloroethylene or diclofenac. *Chem. Res. Toxicol.*, *16*, 1306–1317.

Joseph, M.S. (1985). Hypersensitivity reaction to dapsone. *Lepr. Rev.*, *56*, 315–320.

Ju, C., Uetrecht, J.P. (2002). Mechanism of idiosyncratic drug reactions: reactive metabolite formation, protein binding and the regulation of the immune system. *Curr. Drug Metab.*, *3*, 367–377.

Kalgutkar, A.S., Dalvie, D.K., O'Donnell, J.P., Taylor, T.J., Sahakian, D.C. (2002). On the diversity of oxidative bioactivation reactions on nitrogen-containing xenobiotics. *Curr. Drug Metab.*, *3*, 379–424.

Kalgutkar, A.S., Gardner, I., Obach, R.S., Shaffer, C.L., Callegari, E., Henne, K.R., Mutlib, A.E., Dalvie, D.K., Lee, J.S., Nakai, Y., O'Donnell, J.P., Boer, J., Harriman, S.P. (2005). A comprehensive listing of bioactivation pathways of organic functional groups. *Curr. Drug Metab.*, *6*, 161–225.

Kalgutkar, A.S., Soglia, J.R. (2005). Minimising the potential for metabolic activation in drug discovery. *Expert Opin. Drug Metab. Toxicol.*, *1*, 91–142.

Kaplowitz, N. (2005). Idiosyncratic drug hepatotoxicity. *Nat. Rev. Drug Discov.*, *4*, 489–499.

Kassahun, K., Abbott, F. (1993). In vivo formation of the thiol conjugates of reactive metabolites of 4-ene VPA and its analog 4-pentenoic acid. *Drug Metab. Dispos.*, *21*, 1098–1106.

Kassahun, K., Pearson, P.G., Tang, W., McIntosh, I., Leung, K., Elmore, C., Dean, D., Wang, R., Doss, G., Baillie, T.A. (2001). Studies on the metabolism of troglitazone to reactive intermedi-

ates in vitro and in vivo. Evidence for novel biotransformation pathways involving quinone methide formation and thiazolidinedione ring scission. *Chem. Res. Toxicol.*, *14*, 62–70.

Kassahun, K., Skordos, K., McIntosh, I., Slaughter, D., Doss, G.A., Baillie, T.A., Yost, G.S. (2005). Zafirlukast metabolism by cytochrome P450 3A4 produces an electrophilic alpha,beta-unsaturated iminium species that results in the selective mechanism-based inactivation of the enzyme. *Chem. Res. Toxicol.*, *9*, 1427–1437.

Kedderis, G.L., Argenbright, L.S., Miwa, G.T. (1989). Covalent interaction of 5-nitroimidazoles with DNA and protein in vitro: mechanism of reductive activation. *Chem. Res. Toxicol.*, *2*, 146–149.

Kennaway, E. (1955). The identification of a carcinogenic compound in coal tar. *Br. Med. J.*, *2*, 749–752.

Kent, U.M., Lin, H.L., Noon, K.R., Harris, D.L., Hollenberg, P.F. (2006). Metabolism of bergamottin by cytochromes P450 2B6 and 3A5. *J. Pharmacol. Exp. Ther.*, *318*, 992–1005.

Kiani, R., Kushner, M. (1983). Zomepirac-induced serum sickness. A report of two cases. *JAMA*, *249*, 2812–2813.

Korzekwa, K.R., Jones, J.P. (1993). Predicting the cytochrome P450 mediated metabolism of xenobiotics. *Pharmacogenetics*, *3*, 1–18.

Kretz-Rommel, A., Boelsterli, U.A. (1993). Diclofenac covalent protein binding is dependent on acyl glucuronide formation and is inversely related to P450-mediated acute cell injury in cultured rat hepatocytes. *Toxicol. Appl. Pharmacol.*, *120*, 155–161.

Kumar, S., Kassahun, K., Tschirret-Guth, R.A., Mitra, K., Baillie, T.A. (2008). Minimizing metabolic activation during pharmaceutical lead optimization: progress, knowledge gaps and future directions. *Curr. Opin. Drug Discov. Devel.*, *11*, 43–52.

Lee, J.M., Li, J., Johnson, D.A., Stein, T.D., Kraft, A.D., Calkins, M.J., Jakel, R.J., Johnson, J.A. (2005). Nrf2, a multi-organ protector? *FASEB J.*, *19*, 1061–1066.

Leone, A.M., Kao, L.M., McMillian, M.K., Nie, A.Y., Parker, J.B., Kelley, M.F., Usuki, E., Parkinson, A., Lord, P.G., Johnson, M.D. (2007). Evaluation of felbamate and other antiepileptic drug toxicity potential based on hepatic protein covalent binding and gene expression. *Chem. Res. Toxicol.*, *20*, 600–608.

Levy, D.B., Vasilomanolakis, E.C. (1984). Anaphylactic reactions due to zomepirac. *Drug Intell. Clin. Pharm.*, *18*, 983–984.

Li, C., Benet, L.Z., Grillo, M.P. (2002). Enantioselective covalent binding of 2-phenylpropionic acid to protein in vitro in rat hepatocytes. *Chem. Res. Toxicol.*, *15*, 1480–1487.

Liebler, D.C. (2002). Proteomic approaches to characterize protein modifications: new tools to study the effects of environmental exposures. *Environ. Health Perspect.*, *110* (Suppl. 1), 3–9.

Liebler, D.C. (2006). The poisons within: application of toxicity mechanisms to fundamental disease processes. *Chem. Res. Toxicol.*, *19*, 610–613.

Lin, H.L., Hollenberg, P.F. (2007). The inactivation of cytochrome P450 3A5 by 17alpha-ethynylestradiol is cytochrome b5-dependent: metabolic activation of the ethynyl moiety leads to the formation of glutathione conjugates, a heme adduct, and covalent binding to the apoprotein. *J. Pharmacol. Exp. Ther.*, *321*, 276–287.

Lufkin, E.G., Wong, M., Deal, C. (2001). The role of selective estrogen receptor modulators in the prevention and treatment of osteoporosis. *Rheum. Dis. Clin. North. Am.*, *27*, 163–185.

Mandell, G.L., Petri, W.A. (1996). Antimicrobial agents, in *Goodman and Gilman's The Pharmacological Basis of Therapeutics*, 9th edn (eds J.G. Hardman, A.G. Gilman and L.E. Limbard), McGraw-Hill, New York, pp. 1155–1174.

Martin, R., Rose, D., Yu, K., Barros, S. (2006). Toxicogenomics strategies for predicting drug toxicity. *Pharmacogenomics*, *7*, 1003–1016.

Matthews, A.M., Hinson, J.A., Roberts, D.W., Pumford, N.R. (1997). Comparison of covalent binding of acetaminophen and the regioisomer 3′-hydroxyacetanilide to mouse liver protein. *Toxicol. Lett.*, *90*, 77–82.

Melicow, M.M. (1975). Percivall Pott (1713–1788): 200th anniversary of first report of occupation-induced cancer scrotum in chimney sweepers (1775). *Urology*, *6*, 745–749.

Miller, E.C., Miller, J.A. (1974). Biochemical mechanisms of chemical carcinogenesis, in *The Molecular Biology of Cancer* Academic Press, New York, pp. 377–402.

Miller, E.C., Miller, J.A. (1979). Milestones in chemical carcinogenesis. *Oncology*, *6*, 445–456.

Miller, E.C., Miller, J.A. (1981). Reactive metabolites as key intermediates in pharmacologic and toxicologic responses: examples from chemical carcinogenesis. *Adv. Exp. Med. Biol.*, *136*, 1–21.

Mitchell, J.R., Nelson, W.L., Potter, W.Z., Sasame, H.A., Jollow, D.J. (1976). Metabolic activation of furosemide to a chemically reactive, hepatotoxic metabolite. *J. Pharmacol. Exp. Ther.*, *199*, 41–52.

Mitchell, J.R., Snodgrass, W.R., Gillette, J.R. (1976). The role of biotransformation in chemical-induced liver injury. *Environ. Health Perspect.*, *15*, 27–38.

Monks, T.J., Hinson, J.A., Gillette, J.R. (1982). Bromobenzene and p-bromophenol toxicity and covalent binding in vivo. *Life Sci.*, *30*, 841–848.

Mortishire-Smith, R.J., Skiles, G.L., Lawrence, J.W., Spence, S., Nicholls, A.W., Johnson, B.A., Nicholson, J.K. (2004). Use of metabonomics to identify impaired fatty acid metabolism as the mechanism of a drug-induced toxicity. *Chem. Res. Toxicol.*, *17*, 165–173.

Mutlib, A., Lam, W., Atherton, J., Chen, H., Galatsis, P., Stolle, W. (2005). Application of stable isotope labeled glutathione and rapid scanning mass spectrometers in detecting and characterizing reactive metabolites. *Rapid Commun. Mass Spectrom.*, *19*, 3482–3492.

Myers, T.G., Dietz, E.C., Anderson, N.L., Khairallah, E.A., Cohen, S.D., Nelson, S.D. (1995). A comparative study of mouse liver proteins arylated by reactive metabolites of acetaminophen and its nonhepatotoxic regioisomer, 3′-hydroxyacetanilide. *Chem. Res. Toxicol*, *8*, 403–413.

Narasimhan, N., Weller, P.E., Buben, J.A., Wiley, R.A., Hanzlik, R.P. (1988). Microsomal metabolism and covalent binding of [3H/14C]-bromobenzene. Evidence for quinones as reactive metabolites. *Xenobiotica*, *18*, 491–499.

Nassar, A.E., Lopez-Anaya, A. (2004). Strategies for dealing with reactive intermediates in drug discovery and development. *Curr. Opin. Drug Discov. Devel.*, *7*, 126–136.

Nelson, S.D. (1990). Molecular mechanisms of the hepatotoxicity caused by acetaminophen. *Semin. Liver Dis.*, *10*, 267–278.

Nelson, S.D. (1995). Mechanisms of the formation and disposition of reactive metabolites that can cause acute liver injury. *Drug Metab. Rev.*, *27*, 147–177.

Nelson, S.D., Pearson, P.G. (1990). Covalent and noncovalent interactions in acute lethal cell injury caused by chemicals. *Ann. Rev. Pharmacol. Toxicol.*, *30*, 169–195.

Nelson, S.D., Mitchell, J.R., Snodgrass, W.R., Timbrell, J.A. (1978). Hepatotoxicity and metabolism of iproniazid and isopropylhydrazine. *J. Pharmacol. Exp. Ther.*, *206*, 574–585.

Nicholson, J.K., Connelly, J., Lindon, J.C., Holmes, E. (2002). Metabonomics: a platform for studying drug toxicity and gene function. *Nat. Rev. Drug Discov.*, *1*, 153–161.

O'Donnell, J.P., Dalvie, D.K., Kalgutkar, A.S., Obach, R.S. (2003). Mechanism-based inactivation of human recombinant P450 2C9 by the nonsteroidal anti-inflammatory drug suprofen. *Drug Metab. Dispos.*, *31*, 1369–1377.

O'Neill, P.J., Yorgey, K.A., Renzi, N.L., Williams, R.L., Benet, L.Z. (1982). Disposition of zomepirac sodium in man. *J. Clin. Pharmacol.*, *22*, 470–476.

Okawa, H., Motohashi, H., Kobayashi, A., Aburatani, H., Kensler, T.W., Yamamoto, M. (2006). Hepatocyte-specific deletion of the keap1 gene activates Nrf2 and confers potent resistance against acute drug toxicity. *Biochem. Biophys. Res. Commun.*, *339*, 79–88.

Olsen, J., Li, C.C., Bjørnsdottir, I., Sidenius, U., Benet, L.Z., Hansen, S.H., Benet, L.Z. (2005). In vitro and in vivo studies on acyl-coenzyme A-dependent bioactivation of zomepirac in rats. *Chem. Res. Toxicol.*, *18*, 1729–1736.

Olsen, J., Li, C., Skonberg, C., Bjørnsdottir, I., Sidenius, U., Benet, L.Z., Hansen, S.H. (2007). Studies on the metabolism of tolmetin to the chemically reactive acyl-coenzyme A thioester intermediate in rats. *Drug Metab. Dispos.*, *35*, 758–764.

Ottoboni, M.A. (1991). *The Dose Makes The Poison: A Plain-Language Guide to Toxicology*, 2nd edn, Van Nostrand Reinhold, New York, p. 96.

Park, B.K., Kitteringham, N.R., Maggs, J.L., Pirmohamed, M., Williams, D.P. (2005a). The role of metabolic activation in drug-induced hepatotoxicity. *Ann. Rev. Pharmacol. Toxicol.*, *45*, 177–202.

Park, K., Williams, D.P., Naisbit, D.J., Kitteringham, N.R., Pirmohamed, M. (2005b). Investigation of toxic metabolites during drug development. *Toxicol. Appl. Pharmacol.*, *207*, 425–434.

Pereira, M.A., Burns, F.J., Albert, R.E. (1979). Dose response for benzo(a)pyrene adducts in mouse epidermal DNA. *Cancer Res.*, *39*, 2556–2559.

Pirmohamed, M. (2006). Genetic factors in the predisposition to drug-induced hypersensitivity reactions. *AAPS J*, *8*, 20–26.

Porubek, D.J., Grillo, M.P., Olsen, R.K., Baillie, T.A. (1991). Toxic metabolites of valproic acid: inhibition of rat liver acetoacetyl-CoA thiolase by 2-*n*-propyl-$^4$-pentenoic acid ($\Delta^4$-VPA) and related branched-chain carboxylic acids, in *Idiosyncratic Reactions to Valproate Clinical Risk Patterns and Mechanisms of Toxicity* (ed. R.H. Levy, J.K. Penry), Raven Press, Ltd., New York, pp. 53–58.

Pott, P. (1775). *Cancer Scroti, Chiurgical Observations. The Chiurgical Works of Percival Pott*, Hawes, Clark, and Collings, London.

Pumford, N.R., Halmes, N.C. (1997). Protein targets of xenobiotic reactive intermediates. *Annu. Rev. Pharmacol. Toxicol.*, *37*, 91–117.

Qiu, Y., Benet, L.Z., Burlingame, A.L. (1998). Identification of the hepatic protein targets of reactive metabolites of acetaminophen in vivo in mice using two-dimensional gel electrophoresis and mass spectrometry. *J. Biol. Chem.*, *273*, 17940–17953.

Qiu, Y., Benet, L.Z., Burlingame, A.L. (2001). Identification of hepatic protein targets of the reactive metabolites of the non-hepatotoxic regioisomer of acetaminophen, 3′-hydroxyacetanilide, in the mouse in vivo using two-dimensional gel electrophoresis and mass spectrometry. *Adv. Exp. Med. Biol.*, *500*, 663–673.

Rajapaksa, K.S., Cannady, E.A., Sipes, I.G., Hoyer, P.B. (2007). Involvement of CYP 2E1 enzyme in ovotoxicity caused by 4-vinylcyclohexene and its metabolites. *Toxicol. Appl. Pharmacol.*, *221*, 215–221.

Rashed, M.S., Myers, T.G., Nelson, S.D. (1990). Hepatic protein arylation, glutathione depletion, and metabolite profiles of acetaminophen and a non-hepatotoxic regioisomer, 3′-hydroxyacetanilide, in the mouse. *Drug Metab. Dispos.*, *18*, 765–770.

Reid, W.D., Christie, B., Krishna, G., Mitchell, J.R., Moskowitz, J., Brodie, B.B. (1971). Bromobenzene metabolism and hepatic necrosis. *Pharmacology*, *6*, 41–55.

Reinus, J.F., Persky, S., Burkiewicz, J.S., Quan, D., Bass, N.M., Davern, T.J. (2000). Severe liver injury after treatment with the leukotriene receptor antagonist zafirlukast. *Ann. Intern. Med.*, *133*, 964–968.

Rettie, A.E., Rettenmeier, A.W., Howald, W.N., Baillie, T.A. (1987). Cytochrome P-450-catalyzed formation of delta 4-VPA, a toxic metabolite of valproic acid. *Science*, *235*, 890–893.

Roberts, S.A. (2001). High-throughput screening approaches for investigating drug metabolism and pharmacokinetics. *Xenobiotica*, *31*, 557–589.

Roberts, S.A., Price, V.F., Jollow, D.J. (1990). Acetaminophen structure-toxicity studies: in vivo covalent binding of a nonhepatotoxic analog, 3-hydroxyacetanilide. *Toxicol. Appl. Pharmacol.*, *105*, 195–208.

Sahali-Sahly, Y., Balani, S.K., Lin, J.H., Baillie, T.A. (1996). *In vitro* studies on the metabolic activation of the furanopyridine L-754,394, a highly potent and selective mechanism-based Inhibitor of cytochrome P450 3A4. *Chem. Res. Toxicol.*, *9*, 1007–1012.

Scheen, A.J. (2001a). Hepatotoxicity with thiazolidinediones: is it a class effect? *Drug Saf.*, *24*, 873–888.

Scheen, A.J. (2001b). Thiazolidinediones and liver toxicity. *Diabetes Metab*, *27*, 305–333.

Schmeltz, I., Tosk, J., Hilfrich, J., Jirota, N., Hoffman, D., Wynder, E. (1978). Bioassays of naphthalene for co-carcinogenic activity. Relation to tobacco carcinogenesis, in *Polynuclear Aromatic Hydrocarbons* (eds P. Jones, R. Freudenthal), Raven Press, New York, Vol. 3, pp. 47–60.

Scholz, K., Dekant, W., Völkel, W., Pähler, A. (2005). Rapid detection and identification of N-acetyl-L-cysteine thioethers using constant neutral loss and theoretical multiple reaction monitoring combined with enhanced product-ion scans on a linear ion trap mass spectrometer. *J. Am. Soc. Mass Spectrom.*, *16*, 1976–1984.

Shenton, J.M., Chen, J., Uetrecht, J.P. (2004). Animal models of idiosyncratic drug reactions. *Chem. Biol. Interact.*, *150*, 53–70.

Shultz, M.A, Choudary, P.V, and Buckpitt A.R. (1999). Role of murine cytochrome P-450 2F2 in metabolic activation of naphthalene and metabolism of other xenobiotics. *J. Pharmacol. Exp. Ther.*, *290*, 281–288.

Sidenius, U., Skonberg, C., Olsen, J., Hansen, S.H. (2004). In vitro reactivity of carboxylic acid-CoA thioesters with glutathione. *Chem. Res. Toxicol.*, *17*, 75–81.

Sims, P., Grover, P.L., Swaisland, A., Pal, K., Hewer, A. (1974). Metabolic activation of benzo(a)pyrene proceeds by a diol-epoxide. *Nature*, *22*, 326–328.

Siraki, A.G., Chevaldina, T., O'Brien, P.J. (2005). Application of quantitative structure-toxicity relationships for acute NSAID cytotoxicity in rat hepatocytes. *Chem. Biol. Interact.*, *151*, 177–191.

Small, R.E. (1989). Diclofenac sodium. *Clin. Pharm.*, *8*, 545–558.

Smith, P.C., Benet, L.Z., McDonagh, A.F. (1990). Covalent binding of zomepirac glucuronide to proteins: evidence for a Schiff base mechanism. *Drug Metab. Dispos.*, *18*, 639–644.

Soglia, J.R., Contillo, L.G., Kalgutkar, A.S., Zhao, S., Hop, C.E., Boyd, J.G., Cole, M.J. (2006). A semiquantitative method for the determination of reactive metabolite conjugate levels in vitro utilizing liquid chromatography-tandem mass spectrometry and novel quaternary ammonium glutathione analogues. *Chem. Res. Toxicol.*, *19*, 480–490.

Spahn-Langguth, H., Benet, L.Z. (1992). Acyl glucuronides revisited: is the glucuronidation process a toxification as well as a detoxification mechanism? *Drug Metab. Rev.*, *24*, 5–47.

Swaffar, D.S., Pomerantz, S.C., Harker, W.G., Yost, G.S. (1992). Non-enzymatic activation of procarbazine to active cytotoxic species. *Oncol. Res.*, *4*, 49–58.

Tang, W. (2003). The metabolism of diclofenac—enzymology and toxicology perspectives. *Curr. Drug Metab.*, *4*, 319–329.

Tang, W. (2007). Drug metabolite profiling and elucidation of drug-induced hepatotoxicity. *Expert Opin. Drug Metab. Toxicol.*, *3*, 407–420.

Tang, W., Borel, A.G., Fujimiya, T., Abbott, F.S. (1995). Fluorinated analogues as mechanistic probes in valproic acid hepatotoxicity: hepatic microvesicular steatosis and glutathione status. *Chem. Res. Toxicol*, *8*, 671–782.

Tang, W., Stearns, R.A., Bandiera, S.M., Zhang, Y., Raab, C., Braun, M.P., Dean, D.C., Pang, J., Leung, K.H., Doss, G.A., Strauss, J.R., Kwei, G.Y., Rushmore, T.H., Chiu, S.H., Baillie, T.A. (1999). Studies on cytochrome P-450-mediated bioactivation of diclofenac in rats and in human hepatocytes: identification of glutathione conjugated metabolites. *Drug Metab. Dispos.*, *27*, 365–372.

Teffera, Y., Waldon, D.J., Colletti, A.E., Albrecht, B.K., Zhao, Z. (2008). Identification of a novel glutathione conjugate of diclofenac by LTQ-orbitrap. *Drug Metab. Lett.*, *2*, 35–40.

Timbrell, J.A., Mitchell, J.R., Snodgrass, W.R., Nelson, S.D. (1980). Isoniazid hepatotoxicity: the relationship between covalent binding and metabolism in vivo. *J. Pharmacol Exp Ther.*, *213*, 364–369.

Tirmenstein, M.A., Nelson, S.D. (1989). Subcellular binding and effects on calcium homeostasis produced by acetaminophen and a nonhepatotoxic regioisomer, 3′-hydroxyacetanilide, in mouse liver. *J. Biol. Chem.*, *264*, 9814–9819.

Uetrecht, J. (2001). Prediction of a new drug's potential to cause idiosyncratic reactions. *Curr. Opin. Drug Discov. Devel.*, *4*, 55–59.

Uetrecht, J. (2006). Evaluation of which reactive metabolites, if any, is responsible for a specific idiosyncratic reaction. *Drug Metab. Rev.*, *38*, 745–753.

Uetrecht, J.P. (1999). New concepts in immunology relevant to idiosyncratic drug reactions: the "danger hypothesis" and innate immune system. *Chem. Res. Toxicol.*, *12*, 387–395.

Uetrecht, J.P. (2003). Screening for the potential of a drug candidate to cause idiosyncratic drug reactions. *Drug Discov. Today.*, *8*, 832–837.

Uetrecht, J.P. (2008). Idiosyncratic drug reactions: past, present and future. *Chem. Res. Toxicol.*, *21*, 84–92.

Vyas, P.M., Roychowdhury, S., Khan, F.D., Prisinzano, T.E., Lamba, J., Schuetz, E.G., Blaisdell, J., Goldstein, J.A., Munson, K.L., Hines, R.N., Svensson, C.K. (2006). Enzyme-mediated protein haptenation of dapsone and sulfamethoxazole in human keratinocytes: I. expression and role of cytochromes P450. *J. Pharmacol. Exp. Ther.*, *319*, 488–496.

Waldron, H.A. (1983). A brief history of scrotal cancer. *Br. J. Ind. Med.*, *40*, 390–401.

Walgren, J.L., Mitchell, M.D., Thompson, D.C. (2005). Role of metabolism in drug-induced idiosyncratic hepatotoxicity. *Crit. Rev. Toxicol.*, *35*, 325–361.

Wang, M., Dickinson, R.G. (2000). Bile duct ligation promotes covalent drug-protein adduct formation in plasma but not in liver of rats given zomepirac. *Life Sci.*, *68*, 525–537.

Waring, J.F., Anderson, M.G. (2005). Idiosyncratic toxicity: mechanistic insights gained from analysis of prior compounds. *Curr. Opin. Drug Discov. Devel.*, *8*, 59–65.

Watkins, P.B., Kaplowitz, N., Slattery, J.T., Colonese, C.R., Colucci, S.V., Stewart, P.W., Harris, S.C. (2006). Aminotransferase elevations in healthy adults receiving 4 grams of acetaminophen daily: a randomized controlled trial. *JAMA.*, *296*, 87–93.

Welch, K.D., Wen, B., Goodlett, D.R., Yi, E.C., Lee, H., Reilly, T.P., Nelson, S.D., Pohl, L.R. (2005). Proteomic identification of potential susceptibility factors in drug-induced liver disease. *Chem. Res. Toxicol.*, *18*, 924–933.

van Welie, R.T., van Dijck, R.G., Vermeulen, N.P., van Sitter, N.J. (1992). Mercapturic acids, protein adducts, and DNA adducts as biomarkers of electrophilic chemicals. *Crit. Rev. Toxicol.*, *22*, 271–306.

van Wijk, F., Nierkens, S. (2006). Assessment of drug-induced immunotoxicity in animal models. *Drug Discov. Today*, *3*, 103–109.

Williams, D.P. (2006). Toxicophores: investigations in drug safety. *Toxicology*, *226*, 1–11.

Williams, D.P., Antoine, D.J., Butler, P.J., Jones, R., Randle, L., Payne, A., Howard, M., Gardner, I., Blagg, J., Park, B.K. (2007). The metabolism and toxicity of furosemide in the Wistar rat and CD-1 mouse: a chemical and biochemical definition of the toxicophore. *J. Pharmacol. Exp. Ther.*, *322*, 1208–1220.

Wilson, J.M., Slattery, J.T., Forte, A.J., Nelson, S.D. (1982). Analysis of acetaminophen metabolites in urine by high-performance liquid chromatography with UV and amperometric detection. *J. Chromatogr.*, *227*, 453–462.

Wirth, P.J., Bettis, C.J., Nelson, W.L. (1976). Microsomal metabolism of furosemide evidence for the nature of the reactive intermediate involved in covalent binding. *Mol. Pharmacol.*, *12*, 759–768.

Yamagiwa, K., Ichikawa, K. (1918). Experimental study of the pathogenesis of carcinoma. *J. Cancer Res*, *3*, 1–13.

Yan, Z., Caldwell, G.W. (2004). Stable-isotope trapping and high-throughput screenings of reactive metabolites using the isotope MS signature. *Anal. Chem.*, *76*, 6835–6847.

Yan, Z., Easterwood, L.M., Maher, N., Torres, R., Huebert, N., Yost, G.S. (2007). Metabolism and bioactivation of 3-methylindole by human liver microsomes. *Chem. Res. Toxicol.*, *20*, 140–148.

Yu, L.J., Chen, Y., Deninno, M.P., O'Connell, T.N., Hop, C.E. (2005). Identification of a novel glutathione adduct of diclofenac, 4′-hydroxy-2′-glutathion-deschloro-diclofenac, upon incubation with human liver microsomes. *Drug Metab. Dispos.*, *33*, 484–488.

Zhang, K.E., Naue, J.E., Arison, B., Vyas, K.P. (1996). Microsomal metabolism of the 5-lipoxygenase inhibitor L-739,010: evidence for furan bioactivation. *Chem. Res. Toxicol.*, *9*, 547–554.

Zheng, J., Ma, L., Xin, B., Olah, T., Humphreys, W.G., Zhu, M. (2007). Screening and identification of GSH-trapped reactive metabolites using hybrid triple quadruple linear ion trap mass spectrometry. *Chem. Res. Toxicol.*, *20*, 757–766.

Zia-Amirhosseini, P., Spahn-Langguth, H., Benet, L.Z. (1994). Bioactivation by glucuronide-conjugate formation. *Adv. Pharmacol.*, *27*, 385–397.

Zimmerman, H. (1999). *Hepatotoxicity: The Adverse Effects of Drugs and Other Chemicals on the Liver*, Lippincott Williams & Wilkins, Philadelphia, PA.

Zimmerman, H., Ishak, K.G. (1982). Valproate-induced hepatic injury: analysis of 23 fatal cases. *Hepatology*, *2*, 591–597.

Zimmerman, H.J. (1993). Hepatotoxicity. *Disease-a-Month*, *39*, 673–788.

# Mechanisms of Reproductive Toxicity

AILEEN F. KEATING and PATRICIA B. HOYER

## 1 INTRODUCTION

Today, women are entering the workplace in record numbers. Many are postponing childbearing to establish a career. For this reason and because people are exposed to ever-increasing concentrations of environmental chemicals, there is an increased awareness of the impact of environmental chemicals on reproductive health. The female must perform two distinct reproductive functions: development and support of female germ cells and maintenance of the fetus until it can survive in the outside world. The reproductive functions of the male include production of the male germ cell (spermatogenesis) and male steroid hormones (steroidogenesis). Therefore, reproductive toxicology involves detecting and understanding potentially detrimental environmental influences on reproductive success in females and males.

This field has been developing in response to observations linking clustered effects in humans to specific types of exposures. There have been several examples of how xenobiotics in the form of pharmaceuticals can impact reproductive function in humans. One is the effect of the sedative thalidomide, prescribed to women in the 1950s for morning sickness during early pregnancy (Seegmiller, 1997). A greatly increased incidence of children born with developmental organ and limb malformations was traced to *in utero* exposure to thalidomide. This drug was reported to have been responsible for 8000 malformed children over a two-year period. Whereas thalidomide is an example involving developmental limb defects, reproductive effects in humans have been seen with the synthetic estrogen, diethylstilbestrol (DES). From the 1940s to 1960s, DES was widely prescribed for women with high-risk pregnancies. In 1971, Herbst, Ulfelder and Poskanzer (1971) observed an increased incidence of rare vaginal clear cell adenocarcinoma in their daughters who had been exposed *in utero* to DES. Subsequent research has demonstrated that prenatal exposure to DES can also cause fertility defects, teratogenesis, and neoplasia throughout the male and female reproductive tracts (Hendry *et al.*, 1999).

The following chapter will outline what is currently known about how xenobiotics can impact reproductive function in females and males. The information is based

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

largely on results from animal studies. Furthermore, how this information in some cases has been translated into effects in humans will also be discussed. To facilitate an understanding of how these chemicals can compromise reproductive function, an overview of reproductive physiology in females and males will first be presented.

## 2 OVERVIEW—FEMALE REPRODUCTION

The major functions of the female gonad, the ovary, are production of the female germ cell, the oocyte, and production of female sex steroid hormones. The ovary of the mature mammalian female contains a heterogeneous mixture of structures that undergo dynamic changes during the estrous/menstrual cycle (Fig. 1). Specifically, the ovary contains two endocrine glands, the follicle and the corpus luteum. The follicle is responsible for gametogenesis (oogenesis) and production of the hormone 17β-estradiol (steroidogenesis). The other endocrine system, the corpus luteum, is derived from follicular tissue following ovulation and is present in the ovary of the nonpregnant female only during the latter part of the estrous/menstrual cycle, but is maintained if pregnancy occurs. The corpus luteum produces the steroid hormone progesterone.

17β-Estradiol and progesterone comprise the major ovarian steroids. 17β-Estradiol is responsible for follicular maturation and hyperplasia of the endometrium and uterine

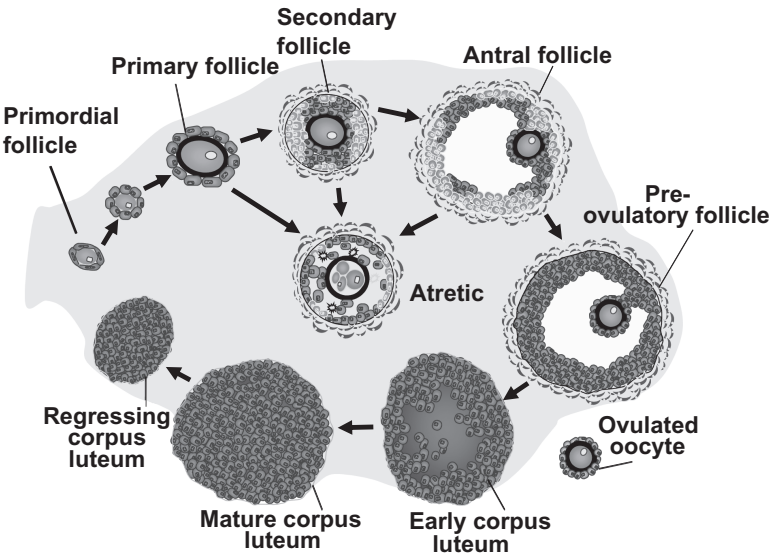


**Figure 1** Development of ovarian follicles. Primordial follicles are signaled to grow and develop toward ovulation. Following ovulation, the remaining tissue differentiates to become a corpus luteum, which undergoes regression if a pregnancy has not been established. The majority (>99%) of follicles die by the natural process of atresia (adapted from Hoyer and Devine, 2002).

vasculature during the follicular phase of the cycle. Additionally, 17β-estradiol provides both negative and positive feedbacks on the hypothalamus and pituitary for regulation of secretion of the gonadotropins, luteinizing hormone (LH) and follicle-stimulating hormone (FSH). Progesterone facilitates implantation by preparing the uterus to accommodate a blastocyst, and provides maintenance of pregnancy by inhibiting uterine contractions and endometrial sloughing. Additionally, progesterone provides negative feedback on the hypothalamus and pituitary for inhibition of the release of gonadotropins.

LH regulates androgen production in the follicle and progesterone production in the corpus luteum, whereas, FSH regulates follicle maturation and estrogen production. Tonic secretion of LH and FSH from the anterior pituitary is affected by the inhibitory actions of 17β-estradiol and progesterone in a classical loop of negative feedback. Because both steroids participate in the inhibition of gonadotropin release, the absence of 17β-estradiol, progesterone, or both, causes basal LH and FSH to increase. Additionally, a protein hormone, inhibin, produced by granulosa cells in developing follicles contributes to the selective inhibition of FSH release. In contrast to tonic LH secretion, which is controlled by a negative feedback mechanism, an LH surge that triggers ovulation is produced by a neuroendocrine reflex arc of positive feedback, which is stimulated by increasing, marked elevations in circulating 17β-estradiol levels produced by the largest developing preovulatory (Graafian) follicle. As a result, when levels become sufficiently elevated, 17β-estradiol shifts from its inhibitory role in regulating LH and FSH release and initiates an LH surge which triggers ovulation. In contrast to 17β-estradiol, progesterone inhibits the LH surge. This ensures that estrogen output during early pregnancy does not stimulate ovulation.

As mentioned earlier, one of the primary functions of the ovary is to provide mature oocytes for successful reproduction. Development and maturation of oocytes occurs within ovarian follicles. Immediately following formation during fetal development, follicles enter the most immature stage of development, termed "primordial." Successful ovulation requires appropriate follicular development, during which the follicle has passed through a number of distinct developmental stages (Hirshfield, 1991). Primordial follicles provide the pool for recruitment of developing follicles; therefore, they are the fundamental reproductive unit within the ovary.

### 2.1 Prenatal Follicle Formation

During fetal development of the ovary, primordial germ cells (oogonia) are formed and become oocytes when they stop dividing, and become arrested at the diplotene stage (prophase) of the first meiotic division. The oocyte does not commence meiosis again unless triggered to ovulate, should that occur. As a result, the lifetime supply of oocytes is set at the time of birth. Around the time of birth, individual oocytes within the ovary become surrounded by a single layer of flattened somatic cells (pre-granulosa cells) and a basement membrane to form primordial follicles (Hirshfield, 1991). Association of the granulosa cells with the oocyte is critical at all subsequent times for maintenance of viability and follicle development (Buccione, Schroeder and Eppig, 1990).

### 2.2 Follicular Development in Adults

Puberty, the time at which sexual reproduction becomes possible, generally occurs between the ages of 9 and 16 in humans. Puberty is identified in women as the first

menstruation, which usually occurs prior to the first ovulation. From birth and throughout the prepubertal period, waves of follicular development in the ovary occur; however, all follicles become atretic. Thus, puberty marks the time at which oocytes can be recruited and developed to ovulation (reviewed in Devine and Hoyer, 2005; Hoyer, 2005).

In humans, 1 to 2 preovulatory follicles develop approximately every 28 days, whereas in rats, 6 to 12 follicles develop every 4 to 5 days (Richards, 1980). Primordial follicles form the pool from which these preovulatory follicles develop. Throughout the reproductive life span, the total number of primordial follicles that ovulate is small compared with the total population. Instead, the vast majority of follicles are lost to attrition in various early stages of development by a process called atresia. The exact determinant for selection of a follicle for ovulation is not understood, but is believed to be under intra-ovarian control (Hirshfield, 1991; Devine and Hoyer, 2005).

The first sign of oocyte growth in primordial follicles is alteration of the surrounding squamous (flattened) granulosa cells into cuboidal shaped cells, followed by initiation of proliferation of these cells (Hirshfield, 1991). Once a follicle makes the transition from primordial to primary, other structural changes occur such as development of the zona pellucida, a protective glycoprotein matrix (Richards, 1980; Hirshfield, 1991). At this stage, another layer of specialized somatic cells, designated theca interna cells, begins to proliferate outside the basement membrane enclosing the oocyte and granulosa cell layer. Theca cells express receptors for LH which stimulates production of androstenedione, the substrate for aromatization to 17β-estradiol by neighboring granulosa cells (Richards, 1980).

As follicles continue to develop, the layers of granulosa cells surrounding the oocyte increase rapidly to become large pre-antral, growing follicles, with diameters reaching 250 μm. The somatic cells acquire receptors for FSH to enhance follicle growth, and develop steroidogenic capacity for synthesis of 17β-estradiol.

The number of follicles that reach the final stage of development is quite small compared with those that began development from the primordial pool. As the follicle develops beyond the pre-antral stage, it acquires a fluid-filled cavity, called the antrum, formed by separations within the granulosa cell layer. The antral follicle continues to grow, and at its most mature stage prior to ovulation is known as a preovulatory (Graafian) follicle. During the final phase of development, a preovulatory follicle becomes more sensitive to the gonadotropins FSH and LH than smaller antral and pre-antral follicles (Richards, 1980). Prior to ovulation, granulosa cells begin to express receptors for LH in readiness for receiving a signal for ovulation, and progression of the oocyte through to the second meiotic division. Following ovulation, the cells remaining, which formed the structure of the follicle, infiltrate and differentiate (luteinize) to form a solid gland, the corpus luteum (Fig. 1). The second meiotic division is only completed if fertilization of the oocyte occurs (reviewed in Devine and Hoyer, 2005 and Hoyer, 2005).

### 2.3 Follicular Atresia

The number of oocytes present in ovaries is dynamic and varies with age, with a peak in the total number of oocytes occurring during embryonic development. In humans, that number (about 7 million) occurs at five months' gestation; at birth the number has dropped to 2 million; 250,000 to 400,000 at puberty; and no viable follicles remain at menopause (Mattison and Schulman, 1980; Hirshfield, 1991). During the lifetime of a

woman, ovulation accounts for only 400 to 600 oocytes; the others have been lost at various stages of development by the process of atresia which occurs via programmed cell death, apoptosis (Tilly *et al.*, 1991). Therefore, atresia is the natural fate of the vast majority of ovarian follicles (>99%), since only a select few will ever be ovulated (Hirshfield, 1991).

### 2.4 Menopause

Depletion of functional primordial follicles from the ovary in women is the underlying cause of ovarian failure (menopause), because this dormant follicle pool represents the cohort for recruitment of all developing follicles. Thus, absence of primordial follicles ultimately leads to the complete loss of follicles of all sizes. The average age of menopause in the United States is 51, and this is a direct consequence of depletion of the follicular reserve. As a result, estrogen-producing (granulosa) cells in preovulatory follicles also become depleted. Menopause is preceded by a period of increasingly irregular cycles (the perimenopausal period). This progressive failure in ovarian function is accompanied by a gradual increase in circulating levels of FSH (and eventually LH), and a decline in circulating 17β-estradiol concentrations. Besides the loss of fertility, menopause has been associated with a variety of health problems in women, such as osteoporosis and increased risk of cardiovascular disease (Mattison *et al.*, 1989). It is thought that the absence of 17β-estradiol is the underlying cause of most of the clinical symptoms associated with menopause. Recently, this loss of 17β-estradiol has been implicated in other health risks such as depression (Sowers and LaPietra, 1995), colon cancer (Fernandez *et al.*, 1998), Alzheimer's disease (Waring *et al.*, 1999), and macular degeneration (Smith, Mitchell and Wang, 1997). Thus, identifying and protecting against environmental or occupational exposures that may accelerate the onset of menopause in women becomes even more critical (Devine and Hoyer, 2005; Hoyer, 2005).

## 3 OVERVIEW—MALE REPRODUCTION

The purpose of the male reproductive system is to produce sperm and deliver them to the oocyte for fertilization. The male gonad is the testis. The two major functions of the testis are production of the male germ cell, sperm (spermatogenesis), and production of male sex steroid hormones (steroidogenesis). These two functions are compartmentalized within the testis. The testes in man are paired organs located in the scrotum, outside the abdominal cavity. This location is essential for spermatogenesis because this process is temperature-sensitive and cannot occur at normal body temperature. Each testis is divided into a number of pyramidal septa which are separated by a connective tissue capsule. Each septum contains a large number of tubules, the seminiferous tubules, which are convoluted such that they occupy a minimum amount of space within the septum. The seminiferous tubules are avascular and are the site of spermatogenesis and sperm transport. The interstitial compartment contains the cells of Leydig, lymphatics, blood vessels, and various connective tissue elements, and occupies the space between the seminiferous tubules. The interstitium is the site of steroidogenesis (in Leydig cells) and is highly vascularized. Therefore, testicular compartmentalization affords the ability of mature sperm to be selectively transported outside the body, and the steroid hormones to have localized effects on spermatogenesis, but also to communicate with distant sites within the body via the circulation (Herbert, Suprakar and Roy, 1995).

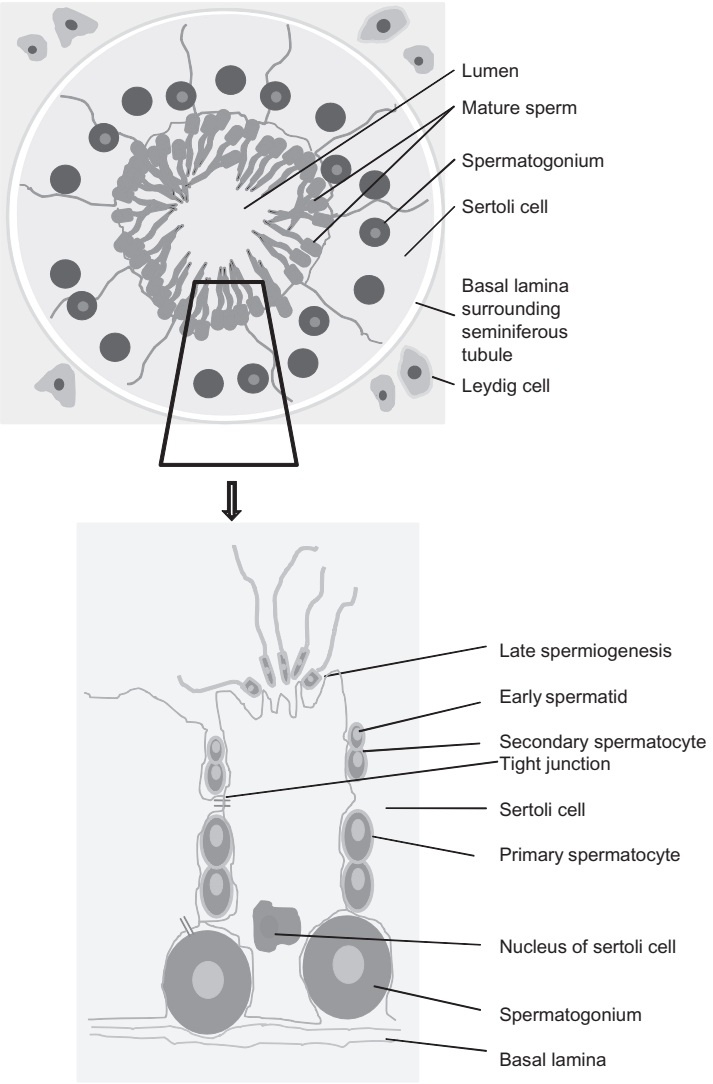


**Figure 2** **(a)** Cross section of a seminiferous tubule (adapted from Martini and Bartholomew, 1997). **(b)** Spermatogenesis occurring in the Sertoli cell (adapted from Sherwood, 1997).

Approximately 80% of testicular mass consists of seminiferous tubules (Fig. 2a). Examination of these tubules in cross section reveals anatomical progression that parallels the temporal sequence of spermatogenesis. Sertoli cells which line the basement membrane of the tubules provide crucial support for spermatogenesis. Also lining the basement membrane are undifferentiated spermatogonia with each clone of differentiating daughter cells migrating toward the tubular lumen as it moves progressively through the stages of development. A major function of Sertoli cells is to maintain tight junctions between themselves and thereby form the blood–testis barrier. This barrier

separates the basal and adluminal compartments within the seminiferous tubules, and provides avascular compartmentalization for the tubular lumen. The barrier prevents movement of proteins, charged organic molecules, and ions from interstitial fluid into seminiferous tubules. Consequently, the composition of intratubular fluid differs markedly from interstitial fluid, and this composition may be critical for later stages of spermatogenesis. The blood–testis barrier also provides important immunological protection, preventing the formation of antibodies against the highly differentiated spermatozoa (Constanzo, 1998).

### 3.1 Spermatogenesis

Spermatogenesis involves the complex process by which relatively undifferentiated germ cells containing a diploid complement of 46 chromosomes are turned into extremely specialized sperm in which the chromosome number is reduced by half (haploid) through meiotic division. This halving of the diploid chromosome number occurs prior to the complete differentiation of spermatozoa. Meiosis also occurs in the female ovum. Therefore, subsequent pairing of male and female chromosomes at the time of fertilization has the effect of giving each individual a unique genotype. Spermatogenesis (formation of spermatozoa from spermatogonia) in humans requires 64 days; however, the time between initiations of two successive spermatogenic generations is approximately 16 days. The rate of spermatogenesis is hormone-independent and quite consistent within a species; however, the yield is hormone-dependent and varies from one individual to another (Sherwood, 1997).

Spermatogenesis can be divided into three major stages: (i) mitotic proliferation, (ii) meiosis, and (iii) packaging (also called spermiogenesis). It begins with mitosis of a spermatogonium to form two daughter cells. One of these cells then progresses to form mature sperm, while the other is retained at the basement membrane as a ready source for maintaining the germ cell line. The mitotic capabilities of spermatogonia are demonstrated by the number of spermatogonia in the testes at birth ($0.6 \times 10^6$), compared with the number present in the testes at puberty ($1.2 \times 10^6$). In fact, a normal adult male produces $100–200 \times 10^6$ sperm per day (Fig. 2b). After undergoing a variable number of mitotic divisions, the daughter cell becomes a primary spermatocyte and enters a resting phase during which chromosomes are duplicated in preparation for meiosis. In man, the daughter cell divides once to become a secondary spermatocyte, and again to become a spermatid. These two meiotic divisions are without DNA replication; therefore, four haploid (23 chromosomes) spermatids are produced from one primary spermatocyte. In addition to the reduction in chromosome number, meiosis allows some chromosomes to exchange segments of DNA, thus mixing their genetic information. After meiosis, spermatids resemble normal epithelial cells. Production of mature spermatozoa requires extensive remodeling of both nucleus and cytoplasm. This packaging (spermiogenesis) involves (i) condensation of chromatin into a tight inert packet, (ii) formation of the acrosome by aggregation of Golgi-produced enzymatic vesicles, (iii) growth of the tail out of one centriole, and (iv) extrusion of most of the cytoplasm (Sherwood, 1997).

### 3.2 Somatic Cells

The hypothalamic hormone gonadotropin-releasing hormone (GnRH) stimulates secretion of FSH and LH from gonadotrophes in the anterior pituitary. The primary

testicular target for FSH is Sertoli cells, located within the seminiferous tubules, which perform several tasks essential for normal spermatogenesis. Sertoli cells are located in an avascular compartment; however, since they provide formation of the blood–testis barrier, the basal face communicates with the vascularized interstitium. The major functions of the Sertoli cells are to (i) maintain the blood–testis barrier, (ii) clean up the damaged cells by phagocytosis, (iii) provide nourishment for developing sperm, (iv) maintain appropriate ionic composition of the seminal fluid, (v) serve as a target cell for hormonal regulation of spermatogenesis, (vi) produce estrogens from androgens, and (vii) synthesize and secrete specific proteins such as inhibin. Inhibin is a protein that feeds back to selectively inhibit secretion of FSH, but not LH, from the anterior pituitary. This provides a specific form of regulation that can preferentially alter secretion of gonadotropins on an individual basis (Sherwood, 1997).

The somatic cells of the testes provide a crucial function to male reproduction. Leydig cells located in the vascularized interstitial compartment are the primary target of LH and the site of synthesis and secretion of the male sex steroid hormones, the androgens (testosterone and androstenedione). LH is also required for maintenance of testicular tissue. In males deprived of adequate LH levels (such as those on anabolic steroids which provide negative feedback to the pituitary), reversible testicular atrophy can occur.

Testosterone is the major male sex steroid produced in the testes, and is lipophilic. Therefore, once released into the blood, it circulates largely (97%) bound to serum-binding proteins. Once secreted, testosterone can arrive at distant sites and be further metabolized. Testosterone can be converted to two active metabolites, as well as an inactive metabolite, androsterone 17-keto-steroid. In sex accessory tissues, testosterone is converted to dihydrotestosterone (DHT) by the enzyme 5α-reductase. DHT is known to be more effective than testosterone in producing fetal development of the prostrate, penis, and scrotum, and in promoting sperm maturation and transport in the adult. Conversion of testosterone by the enzyme aromatase to another active metabolite, 17β-estradiol occurs in Sertoli cells and peripheral tissues; therefore, in addition to its important physiological role, testosterone can serve as a precursor for two other active steroid hormones (Constanzo, 1998). Testosterone also provides negative feedback on LH and FSH release at the level of the hypothalamus and pituitary (Rhoads and Pflanzer, 1996).

In addition to its absolute requirement for spermatogenesis, testosterone provides a variety of reproductive and nonreproductive effects. Testosterone promotes differentiation of the reproductive tract in the developing male fetus, maturation of the reproductive tract at puberty, and maintenance of the reproductive tract in adults. At puberty, testosterone is responsible for development of secondary sex characteristics in the male and development of the libido. Nonreproductive effects of testosterone include anabolic effects on muscle mass. Although the physiological significance of androstenedione is not fully known, it may be required for growth of the secondary sex structures (Sherwood, 1997).

## 4 SITES OF TARGETING AND OUTCOMES

In evaluating reproductive effects in toxicological studies, it is important to consider whether gender differences can be seen. In some cases, females may be susceptible

and males not, or vice versa. That is, chemicals that produce a selective reduction in gonadotropin secretion from the pituitary may affect 17β-estradiol versus testosterone production in a selective manner. As a result, ovary-dependent end points could be impaired, while testes-dependent end points may remain unaffected. Additionally, the impact may be reversible in one gender, and irreversible in the other. For example, in the case of chemicals that destroy germ cells in the early stages of meiotic division, extensive damage may cause irreversible ovarian failure in females, and reversible reductions in spermatogenesis in males (4-vinylcyclohexene diepoxide, VCD). This is because within the ovary the oocyte is arrested in prophase of the first meiotic division, and once destroyed no more can be formed, whereas within the testis, spermatogonia (the most immature form of sperm) can divide by mitosis to maintain a continuously renewable source of germ cells. Because of these possibilities, it is important to investigate the effects of reproductive toxicants in both genders and determine the targeted site(s) in order to evaluate and predict different levels of risk between women and men (Hoyer, 2001). There is currently no general understanding as to what underlies a specific chemical causing reproductive toxicity. In some cases, a chemical may be known to have widespread effects that are not selective for only reproductive tissues, such as polycyclic aromatic hydrocarbons (PAHs). Whereas in other cases, a chemical may selectively target the gonads as with VCD (Richards, 1980; Hoyer and Devine, 2002). Certain aspects of reproductive physiology may render reproductive systems more susceptible to toxicity, such as germ cell-specific growth factors or signaling pathways. Additionally, target tissue-specific expression of xenobiotic-metabolizing enzymes may contribute to localized bioactivation.

### 4.1 Female Effects

Exposure to environmental or occupational chemicals can disrupt female reproductive function (Mattison, 1985). These chemicals can have their effects at a variety of sites in the reproductive axis, including the hypothalamus, pituitary, ovary, or reproductive tract (Mattison and Schulman, 1980; Thomas, 1993; Hoyer and Sipes, 1996; Hoyer, 1997). Disruption of any of these sites can ultimately manifest as a disruption of ovarian function, resulting in infertility; however, certain classes of compounds have been identified that directly affect the ovary. A number of studies have shown that exposure to direct ovarian toxicants often leads to destruction of oocytes, premature ovarian failure, and development of neoplasms (Krarup, 1967, 1969, 1970; Beamer and Tennent, 1986; Tennent and Beamer, 1986; Maronpot, 1987; Melnick *et al.*, 1990; Hoyer and Sipes, 1996). How these effects are produced is generally not well understood, but may be due to one of several possible mechanisms. Oocyte destruction can result from a direct effect of the toxic chemical. Conversely, toxicity can occur via one of several routes. The ovary contains enzymes responsible for biotransformation and detoxification of many xenobiotics. Both the rat and mouse ovary express epoxide hydrolase, glutathione-*S*-transferases, and cytochrome P450s, which metabolize known ovarian toxicants (Mukhtar, Philpot and Bend, 1978; Mattison and Thorgeirsson, 1979; Heinrichs and Juchau, 1980; Bengtsson and Rydstrom, 1983; Bengtsson, Reinholt and Rydstrom, 1992; Flaws *et al.*, 1994; Cannady *et al.*, 2002; Rajapaksa, Sipes and Hoyer, 2007). Therefore, biotransformation of a chemical within the ovary might occur within selected compartments and thereby provide specific exposure of certain classes of oocytes to toxicity. Additionally, since oocytes at all stages of follicular development are surrounded by granulosa cells, these mechanisms might be indirect and involve

effects within the granulosa cell, resulting in loss of its ability to maintain viability of the oocyte (Buccione, Schroeder and Eppig, 1990).

For chemicals that destroy oocytes, the stage of development at which the follicle is lost determines the impact on reproduction. Chemicals that selectively damage large growing or antral follicles only temporarily interrupt reproductive function because these follicles can be replaced by recruitment from the pool of primordial follicles (Hoyer and Sipes, 1996). Thus, these chemicals produce a readily reversible form of infertility that manifests itself relatively soon after exposure (Generoso, Stout and Huff, 1971; Mattison and Schulman, 1980; Jarrell *et al.*, 1991; Davis, Maronpot and Heindel, 1994). Conversely, chemicals that destroy oocytes contained in primordial and primary follicles can lead to permanent infertility and premature ovarian failure (early menopause in humans) because once a primordial follicle is destroyed, it cannot be replaced. Selective destruction of oocytes contained in primordial follicles may have a delayed effect on reproduction until such a time that recruitment for the number of growing and antral follicles can no longer be supported (Generoso, Stout and Huff, 1971; Mattison and Schulman, 1980; Hooser *et al.*, 1994; Mayer *et al.*, 2002). Ultimately, as a result of loss of negative feedback on the hypothalamus and pituitary normally imposed by the ovary (steroids and inhibin), circulating levels of FSH increase (Jarrell *et al.*, 1991; Hooser *et al.*, 1994; Mayer *et al.*, 2002), followed eventually by increases in circulating levels of LH. The increase in release of these hormones from the pituitary will not rise until a sufficient loss of ovarian cyclicity has resulted (several months after disrupted cyclicity is observed).

The level and duration of exposure to an environmental toxicant can determine its effects on reproduction. Individuals are only rarely exposed acutely to high levels of xenobiotic chemicals. These exposures can usually be readily identified. However, the possible effects of chronic, low-dose exposures are of particular concern, because they are more likely to be environmental and are more difficult to identify. These types of exposures may cause reproductive or fertility problems that go unrecognized for years due to the potential for cumulative damage. For example, cigarette smoking in women appears to accelerate the onset of menopause. This is probably due to exposure to PAHs contained in cigarette smoke (Mattison and Schulman, 1980; Mattison *et al.*, 1989; Devine and Hoyer, 2005). Therefore, ongoing selective PAH-induced damage of small pre-antral follicles may not initially cause detectable effects that signal the onset of premature ovarian failure (early menopause). In addition to reproductive effects caused by environmental exposures in fertile women, the age at which exposure occurs can impact the outcome. For example, exposure during childhood might cause sterility by chemical-induced destruction of germ cells. Furthermore, exposures *in utero* may cause improper development of ovarian follicles or permanent alterations in the reproductive tract (Fig. 3) (Hoyer and Sipes, 1996; Hoyer and Devine, 2002). This might result from greater sensitivity of developing fetal organs to xenobiotics or increased metabolic contributions from the placenta. Finally, the increased risk of ovarian cancer in postmenopausal women might be potentiated by a lifetime of exposure to environmental chemicals and/or by increased sensitivity of the senescent ovary (Hoyer, 1997).

### 4.2 Male Effects

Paternal exposure to toxic agents is of concern due to the potential for congenital abnormalities, carcinogenesis, and developmental delays in offspring. Effects of xeno-

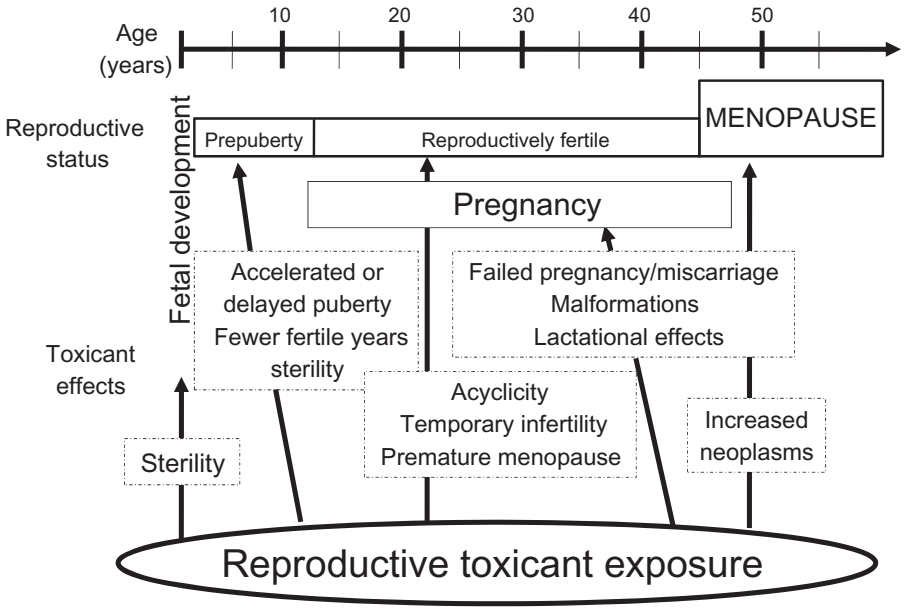


**Figure 3** Potential age-related effects of reproductive toxicants in females. The impact(s) of reproductive toxicants is partially dependent upon the reproductive status of the exposed individual. In most cases, direct ovarian toxicity can lead to premature ovarian failure and infertility (menopause).

biotic exposure on the male reproductive system are also dependent on the site of exposure. Some of the potential direct effects include reduced sperm concentrations and production of compromised spermatozoa. Indirect effects include impaired hormone production of testosterone and the gonadotropins. Xenobiotic exposure to developing germ cells in the male reproductive tract can result in temporary or permanent infertility. Metabolism of chemicals can cause production of highly reactive free radicals which can damage the germinal epithelium due to the lack of catalase in these cells (Buhrley and Ellis, 1973; Docampo, 1987). Examples of xenobiotics whose metabolism results in free radical production include nitrofurans, adriamycin, and bleiomycins (Docampo, 1987). Another anti-spermatic agent is gossypol, isolated from cotton plants, which produces free radicals and also interacts with membrane proteins or the cellular lipid bilayer (Sundaram and Witorsch, 1995). Additionally, xenobiotic exposures can damage the germinal epithelium by initiating apoptosis.

Toxic agents which target developing germ cells (PAHs, cisplatin) can inhibit mitosis (spermatogonia) or meiosis (spermatocytes) and lead to azoospermia or oligospermia. A common side effect due to chemotherapeutic treatment for cancer in males is fertility impairment. Mechanisms by which this occurs include inhibition of DNA synthesis, direct DNA damage, or changes to the kinetics of DNA division. Exposure of germ cells to toxicants can result in irreversible infertility if spermatogonia in the germinal epithelium are targeted (Sundaram and Witorsch, 1995).

Direct targeting of Sertoli cells can damage germ cells by an indirect mechanism (Vinclozolin, VCD, di-(2-ethylhexyl)phthalate; DEHP). Exposure of male rodents to Sertoli cell toxicants results in a rapid reduction in fertility which is prolonged and

sometimes permanent. This toxicity is thought to be due to the dependency of spermatogenesis on the Sertoli cell (Chapin *et al.*, 1988; Foster, 1988).

The major function of the Leydig cell is the production of testosterone, the main circulating androgen in males. This production is regulated by pituitary-derived LH, which is regulated by GnRH from the hypothalamus. Testosterone is important for the reproductive health of the adult male, and plays roles in both sexual behavior and maintenance of spermatogenesis. Thus, chemicals that impair Leydig cell function (2,3,7,8-tetrachlorodibenzo-*p*-dioxin; TCDD, cadmium, lead, mercury, methoxychlor) can have widespread consequences in the male (Sundaram and Witorsch, 1995). Such impairment could result from exposure to chemicals that interfere with GnRH or LH production or by direct targeting of Leydig cells.

The blood–testes barrier is another potential target for xenobiotic-induced effects. Breakdown of the blood–testes barrier by toxic agents can be caused by disruption of tight junctions that form between Sertoli cells (Huang *et al.*, 1988; Weber *et al.*, 1988; Kopf-Maier, 1992).

## 5 EFFECTS OF SPECIFIC CHEMICALS

A number of studies have demonstrated reproductive effects of a wide variety of classes of chemicals in females and males. For many, the results have suggested a mechanism of action (Table 1).

### 5.1 Chemotherapeutics

Roughly 25 females per 1000 will develop some type of cancer before the age of 35 (Statistics Canada), and approximately 1 in 1000 adults are survivors of childhood cancers (Revel and Laufer, 2002). Concerns over side effects of cancer therapy have increased as survival rates of cancer patients improve (currently ~56% overall survival rate; Byrne, 1990). Thus, the toxic effects of chemotherapeutic drugs in women cancer survivors have become important issues. Since the beginning of the use of antineoplastic agents to treat a variety of diseases and malignancies, the ability of these agents to produce reproductive failure has been reported. These effects have been described in patients treated with cyclophosphamide (CPA), nitrogen mustard, chlorambucil, cisplatin, or vinblastine (Miller, Williams and Leissring, 1971; Sobrinho, Levine and DeConti, 1971; Warne *et al.*, 1973; Koyama *et al.*, 1977; Chapman, 1983; Dnistrian *et al.*, 1983; Damewood and Grochow, 1986). In some cases, combined treatment with CPA and methotrexate and/or 5-fluorouracil produced the same effect; however, the individual contribution to ovarian failure by the latter compounds alone was not clearly identified (Koyama *et al.*, 1977; Dnistrian *et al.*, 1983). Effects of chemotherapeutic exposures on reproductive function are primarily a concern for those under the age of 40 who may wish to have children. Children exposed to chemotherapy prior to puberty are less likely to become permanently infertile than adults (Blumenfeld *et al.*, 2002); however, the Childhood Cancer Survivor Study reported that 6.3% of girls will undergo acute ovarian failure, and 8% will experience premature menopause (Garcia, 2007). Retention of ovarian function following chemotherapy does not mean ovarian damage did not occur, and premature menopause is still likely. Furthermore, undergoing early reproductive failure also increases the risks for developing other diseases. Studies on male fertility after cancer treatment have shown that chemotherapeutic agents are

**TABLE 1 Summary of targeted sites and predicted effects of specific chemicals on reproductive function in females and males.**

| Chemical | Small pre-antral follicles | Growing follicles | Spermatogenesis | Hormone levels | |
|---|---|---|---|---|---|
| | | | | Male | Female |
| Cisplatin | – | – | – | + | – |
| CPA | + | + | – | – | – |
| BaP | + | +/– | + | – | + |
| DMBA | + | + | + | – | + |
| 3-MC | + | – | – | – | – |
| 1,3-BD | + | + | + | + | – |
| VCH/VCD | + | – | + | – | – |
| 1-BP | – | + | + | – | – |
| 2-BP | + | + | + | – | – |
| DBP | – | – | + | – | – |
| MEHP | – | + | + | – | + |
| DEPH | – | + | + | – | +/– |
| Cadmium | – | + | + | + | + |
| Lead | – | – | + | + | + |
| Mercury | + | + | + | + | + |
| TCDD | – | – | + | + | + |
| PCBs | + | + | + | – | – |
| BPA | + | – | + | – | – |
| DDT/DDE | – | – | + | – | – |
| Methoxychlor | – | + | – | + | + |
| Vinclozolin | – | – | + | – | – |
| Psoralens | – | – | + | – | + |
| Atrazine | – | – | + | – | + |

+, effect observed; –, no effect observed; +/–, possible effect.
CPA, cyclophosphamide; BaP, benzo[*a*]pyrene; DMBA, 9,10-dimethylbenz[*a*]anthracene; 3-MC, 3-methylcholanthrene; 1,3-BD, 1,3-butadiene; VCH, 4-vinylcyclohexene; VCD, 4-vinylcyclohexene diepoxide; 1-BP, 1-bromopropane; 2, BP-2-bromopropane; DBP, di-butylphthalate; MEHP, mono-(2-ethylhexyl)phthalate; DEHP, di-(2-ethylhexyl)phthalate; TCDD, 2,3,7,8-tetrachlorodibenzo-*p*-dioxin; PCB, polychlorinated biphenyl; BPA, bisphenol A; DDT, dichlorodiphenyltrichloroethane; DDE, dichlorodiphenyldichloroethene.

particularly harmful to male germ cells. A study of testicular cancer survivors found that there was a 30% reduction in fertility after chemotherapeutic treatment (Huyghe *et al.*, 2004). Susceptibility to treatment is dependent on both drug and dose of treatment (Howell and Shalet, 2002a, b). In addition, mutagenic changes in spermatozoa have also been observed; however, they may not persist after treatment (Thomas *et al.*, 2004).

These observations in humans have motivated a variety of studies with CPA in rodents to better elucidate the mechanism(s) by which chemotherapeutic agents cause reproductive toxicity (Miller and Cole, 1970; Botta, Hawkins and Weikel, 1974; Jarrell *et al.*, 1987; Ataya *et al.*, 1995). Plowchalk and Mattison (1992) observed a time- and dose-dependent relationship between CPA and ovarian toxicity by looking at changes in ovarian structure and function in C57BL/6N mice. Loss of primordial follicles was essentially complete in three days and the estimated ED50 (concentration that produced 50% follicle loss) was 122 mg/kg body weight. From these results, it appears that premature ovarian failure in women following chemotherapy is likely via destruction of primordial follicles. Phosphoramide mustard has been determined to be the antineoplastic and ovotoxic form of this chemical (Plowchalk and Mattison, 1992). Mice

were dosed with chemicals capable of forming specific metabolites of CPA. Only chemicals that released phosphoramide mustard induced ovarian toxicity. The greater potency of phosphoramide mustard-producing chemicals compared with CPA in mice was attributed to a bypassing of detoxification steps, allowing more toxic metabolite to reach the ovary (Plowchalk and Mattison, 1992).

In addition to the widespread destruction of primordial follicles by CPA, growing follicles have also been reported to be affected. Ataya *et al.* (1995) described specific losses of follicles >30 μm in rhesus monkeys given multiple doses of CPA. Also, in rats, CPA caused destruction of antral follicles at doses that did not affect primordial follicles (Jarrell *et al.*, 1987). In contrast, in mice under conditions that completely destroyed primordial follicles, only partial destruction of antral follicles was observed (Plowchalk and Mattison, 1992). This demonstrated the greater sensitivity of primordial germ cells to this compound in mice as compared with rats. However, in both studies in rats and mice, lower ovarian weight, reduced follicular and luteal volume, and lower circulating 17β-estradiol levels were most highly associated with the loss of antral follicles (Jarrell *et al.*, 1987; Plowchalk and Mattison, 1992). The effect on antral follicle numbers in both species was reversible. This provided a demonstration that unlike irreversible effects resulting from ovotoxicity that impacts the primordial follicle pool, damage to larger follicles can have a temporary impact on cyclicity.

Cisplatin is a chemotherapeutic agent used in the treatment of testicular cancers. This therapy can have long-lasting effects on spermatogenesis (Petersen *et al.*, 1994). In mice, cisplatin at low doses has been shown to be selectively toxic for spermatogonia, while at high doses all cell types were affected (Meistrich *et al.*, 1982; Boekelheide, 2005). Thus, even low-dose exposure might cause sterility in males.

### 5.2 Polycyclic Aromatic Hydrocarbons

PAHs are widespread in the environment from various combustion processes (including automobile exhaust). Additionally, cigarette smoke contains high levels of PAHs. In women, a positive connection between smoking and early menopause has been established (Jick and Porter, 1977). In animal studies, three PAHs have been shown to be carcinogenic and ovotoxic; benzo[*a*]pyrene (BaP), 3-methylcholanthrene (3-MC), and 7,12-dimethylbenz[*a*]anthracene (DMBA) (Mattison and Thorgeirsson, 1978; Vahakangas, Rajaniemi and Pelkonen, 1985). Initially, it was reported that DMBA depletes oocytes and produces ovarian tumors in mice (Krarup, 1967, 1969, 1970). Subsequently, these effects have also been reported in a number of studies for BaP and 3-MC. These three compounds have been shown to destroy oocytes in small follicles in Sprague-Dawley (SD) rats and in B6 and D2 mice within 14 days following a single dose (Mattison and Thorgeirsson, 1979). The order of potency for oocyte destruction by PAH was determined to be DMBA > 3-MC > BaP. This order of potency was similar to the relative toxicities determined for these compounds in experiments exposing mice and rats to multiple doses at lower levels of these compounds (Borman *et al.*, 2000).

A direct relationship between the dose of PAHs and destruction of primordial follicles has been shown in the mouse ovary (Mattison and Thorgeirsson, 1979). Interestingly, significant oocyte destruction was demonstrated following a single high dose of BaP (100 mg/kg), whereas the same level of oocyte loss was observed with a low dose (10 mg/kg) given daily for 10 days (Mattison and Nightingale, 1980). This observation was supported in another study in mice in which, relative to a single high dose of DMBA, there was approximately 250 times greater ovotoxicity from repeated low-dose

exposure (Borman *et al.*, 2000). Taken together, these observations provide support for a cumulative effect of chronic exposures to low doses. In addition to pre-antral follicles, an apparent effect of DMBA and BaP on antral follicles has also been shown in mice (Mattison, 1980). Beside effects in adults, *in utero* exposure to PAHs during fetal development can also affect reproductive potential of offspring. Daily oral exposure (10 mg/kg) of pregnant mice between 7 and 16 days of gestation with BaP caused severely compromised fertility of female offspring (MacKenzie and Angevine, 1981).

In male rats, treatment of seminiferous tubule segments with BaP or DMBA resulted in inhibition of DNA replication during spermatogenesis, with DMBA being 10-fold more potent than BaP (Georgellis, Parvinen and Rydström, 1989). Additionally, high concentrations of BaP and DMBA inhibited progression of spermatocytes through meiotic division (Georgellis *et al.*, 1990; Schrader and Kanitz, 1994).

PAHs must become bioactivated to arene oxides to become ovotoxic. Inhibition of PAH metabolism with α-naphthoflavone prevented oocyte destruction observed in mice (Mattison, 1983; Shiromizu and Mattison, 1984). Expression of enzymes involved in the bioactivation of PAHs (i.e. aryl hydrocarbon hydroxylase and epoxide hydrolase) has been identified in ovaries of mice and rats (Mattison and Thorgeirsson, 1977, 1979; Mattison and Nightingale, 1980; Bengtsson *et al.*, 1983; Bengtsson, Hamberger and Rydstrom, 1988; Bengtsson and Mattison, 1989; Rajapaksa, Sipes and Hoyer, 2007), monkeys (Bengtsson and Mattison, 1989), and humans (Bengtsson, Hamberger and Rydstrom, 1988). Therefore, localized ovarian metabolism of these compounds may be involved in induction of their ovotoxic effects. In males, BaP metabolites have also been detected in the testes (Smith *et al.*, 2007). Overall, it is the combined processes of bioactivation, distribution, and detoxification both throughout the body and locally in the reproductive tract that determine susceptibility to these compounds.

### 5.3 Occupational Chemicals

In general, compounds known to contain epoxide moieties (or which are capable of bioactivation by epoxidation) have been shown to affect reproductive function in laboratory animals. These compounds include butadiene and its derivatives and 4-vinylcyclohexene (VCH) and its respective derivatives (reviewed in National Toxicology Program, 1986, 1989 and Hoyer and Sipes, 1996). Many epoxidated compounds have been associated with increased mutagenicity in *in vitro* bacterial assays. The ability of epoxides to produce DNA adducts and induce sister chromatin exchanges has also demonstrated effects at the molecular level (James, Fouts and Bend, 1976; DeRaat, 1978; Citti *et al.*, 1984). The role of enzymatic epoxidation in bioactivation of parent compounds to these toxic metabolites has been implicated in a variety of studies (reviewed in Hoyer and Sipes, 1996). Because of the ability of these compounds to become epoxidated, they have the potential to be reproductive toxicants and carcinogenic agents.

***5.3.1 Butadiene*** 1,3-Butadiene (BD) and the related olefins, isoprene and styrene, are released during the manufacture of synthetic rubber and thermoplastic resins, and the occupational exposure of US employees has been estimated to be 3700 to 1,000,000 people annually (IARC 1994). These chemicals have also been reported in cigarette smoke and automobile exhaust. Chronic animal inhalation studies have shown that carcinogenesis caused by BD is higher in mice than in rats. At low doses, female mice exposed daily to BD by inhalation for as long as two years exhibited ovarian atrophy,

granulosa cell hyperplasia, and benign and malignant cell tumors (Melnick *et al.*, 1990). Furthermore, ovarian effects of BD appear at lower concentrations than are required to produce effects in other tissues.

In one study, the metabolite of BD, 1,3-butadiene monoepoxide (1.43 mmole/kg), depleted small follicles by 98% and growing follicles by 87% in female B6C3F$_1$ mice dosed daily for 30 days (Doerr *et al.*, 1995). Whereas, the diepoxide of 1,3-butadiene at a much lower dose (0.14 mmole/kg) depleted small follicles by 85% and growing follicles by 63%. The results of this study support that the diepoxide formed in the metabolism of BD is more potent than the monoepoxide at inducing follicle loss (Doerr *et al.*, 1995).

Male rats exposed to 0, 200, 1000, or 8000 ppm of 1,3-BD by inhalation for five days showed concentration-related increases in the frequency of abnormal sperm morphology at five weeks after exposure (Morrissey *et al.*, 1990). Another study measured the effect of diepoxybutane, a metabolite of butadiene on reproductive function. Mice were given a single i.p. dose at varying concentrations and tissues were collected 7, 14, 21, 28, and 35 days after treatment. At 7, 21, and 28 days after dosing, there were dose-dependent reductions in round and elongated spermatids; however, this was not detected at 35 days after dosing, indicating transient damage to differentiating spermatogonia (Spano *et al.*, 1996).

***5.3.2 4-Vinylcyclohexene*** The dimerization of BD forms VCH. The VCH family of compounds are occupational chemicals released at low concentrations during the manufacture of rubber tires, plasticizers, and pesticides (IARC 1994). VCH and its diepoxide metabolite VCD have been shown in mice and rats to (i) produce extensive destruction of small pre-antral follicles (Smith, Mattison and Sipes, 1990), (ii) cause premature ovarian failure (Hooser *et al.*, 1994; Mayer *et al.*, 2002), (iii) increase the risk for development of ovarian tumors (Hoyer and Sipes, 1996), and (iv) affect normal ovarian development of female offspring exposed *in utero* (Hoyer and Sipes, 1996; Devine and Hoyer, 2005). Because no significant effects on other tissues have been reported in studies with this class of chemicals, the damage they produce appears to be highly selective, and does not involve widespread toxicity. Ovarian damage caused by VCH and its related epoxide metabolites has been demonstrated by a variety of exposure routes, including dermal (National Toxicology Program, 1989), oral (Grizzle *et al.*, 1994), inhalation (Bevan *et al.*, 1996), and i.p. injection (Smith, Mattison and Sipes, 1990). It is therefore important to consider the potential risks for human exposure.

The role of biotransformation in VCH-induced ovarian toxicity has been well established. Following 30 days of daily dosing of female B6C3F$_1$ mice and Fischer 344 rats with VCH, primordial and primary follicles were reduced in ovaries of mice, but not of rats (Smith, Mattison and Sipes, 1990). The greater sensitivity of mice compared with rats was partly due to different capabilities for bioactivation of VCH. Studies have shown that the diepoxidation of VCH to VCD represents bioactivation, and it has been concluded that the species variability in susceptibility to VCH is, in part, due to differences in the capacity of mice and rats to form VCD. However, the increased sensitivity of rats versus mice to the ovotoxicity of VCD has also suggested that detoxification of VCD may play a role in species specificity to VCH-induced ovotoxicity. It was shown that the rat had greater capacity for conversion of VCD to its inactive tetrol, as compared with the mouse, and that only rats possessed detectable ovarian VCD-hydrolytic enzymatic activity (Keller *et al.*, 1997). Therefore, the greater susceptibility of mice

versus rats relates to both enhanced bioactivation to and reduced detoxification of the ovotoxic epoxides.

Ovarian follicle destruction by VCD is selective for primordial and primary follicles. Mechanistic studies in rats have determined that VCD causes ovotoxicity by accelerating the natural process of atresia (apoptosis) and this requires repeated exposures (Springer *et al.*, 1996; Borman *et al.*, 1999). Pro-apoptotic signaling events in the Bcl-2 and mitogen-activated protein kinase families have been shown to be selectively activated in fractions of small pre-antral follicles (targets for VCD; Hu *et al.*, 2002a, b). Thus, because VCD causes a physiological form of ovotoxicity, follicle loss is "silent" and mimics normal cellular atresia. Therefore, damage caused by VCD would go unnoticed in exposed individuals that are affected. As a result, chronic exposure in women to low levels of this chemical may represent a risk for early menopause (Hoyer and Sipes, 2007).

In males, VCD targets testicular germ cells; however, unlike in females this effect is reversible. Male mice were dosed daily with 320 mg/kg/day i.p. and tissues collected after 5, 10, 15, 20, 25, or 30 days of dosing (Hooser *et al.*, 1995). Two groups were also dosed daily for 30 days and allowed to recover for an additional 30 or 60 days. Decreases in testicular weight were observed by day 5 and this continued to day 30. The decrease in weight corresponded to necrosis of germ cells. After day 5 of VCD treatment, there was a loss of spermatogonia, and after 30 days of dosing, seminiferous tubules were devoid of germ cells. In mice that were allowed to recover for 30 days, however, there was complete repopulation of seminiferous tubules, and epididymal spermatogonia were present 60 days after dosing (Hooser *et al.*, 1995). This finding supported VCD targeting of cells undergoing meiotic division since the stem cell population of spermatogonia remained viable and could support renewal of spermatogenesis.

***5.3.3 2-Bromopropane*** Due to their detrimental effects on the ozone layer, chlorofluorocarbons were banned and 1- and 2-bromopropane (BP) were proposed as substitute industrial propellants and cleaning solvents. These chemicals were found to have adverse reproductive effects on both male and female workers in a Korean factory (Park *et al.*, 1997). Subsequent animal studies have determined in female rats that 2-BP causes destruction of ovarian follicles in all stages of development, substantiating the cause of amenorrhea and increased hormone levels in the female workers (Yu *et al.*, 1999). For 1-BP, reproductive studies of long-term inhalation exposures (400 to 800 ppm) suggest that growing and antral follicles, and not primordial follicles are the target of this chemical (Yamada *et al.*, 2003). 2-BP (500, 1000 mg/kg, i.p., once every 2 to 3 days for 15 to 17 days) was found to prolong estrous cycles in rats, decrease ovulations, and alter preovulatory follicle morphology (Sekiguchi *et al.*, 2001). Direct effects of 2-BP on the oocyte might also occur, as suggested by increased micronuclei and decreased cell numbers observed in preimplantation mouse embryos following exposures to 2-BP (Ishikawa, Tian and Yamauchi, 2001).

In male rats, spermatogonia have been identified as targets of 1- and 2-BP, thus providing an explanation for reduced sperm production detected in the male workers (Omura *et al.*, 1999). Exposure of male rats to 1-BP reduced epididymal sperm count and motility and increased percentages of abnormal sperm (Ichihara *et al.*, 2000). Further treatment of male rats with 2-BP over a time course resulted in reductions in number and increases in degenerating spermatogonia, oligospermia, and testicular atrophy (Son *et al.*, 1999). A transplacental effect of 2-BP exposure has also been shown (Kang *et al.*, 2002). There was a decrease in the pup survival due to 2-BP exposure to

the dam. In addition, atrophy of the uterus and testes were present in the offspring (Kang *et al.*, 2002).

### 5.4 Phthalates

One class of chemicals receiving much recent attention in reproductive toxicology is the phthalates. Phthalates are the most abundant class of synthetic chemical in the environment. Detrimental reproductive effects have been reported in both human epidemiological and rodent studies (Kavlock *et al.*, 2002). These chemicals, which are diesters of *o*-phthalic acid, are utilized in the plastics industry to enhance the flexibility of polyvinyl chloride products. They are included in such products as cosmetics, lubricants, plastic tubing, medical devices, vinyl upholstery, surgical gloves, toys, solvents, and pesticides. Phthalates are of particular concern because they can leach out of plastics into air, water, or food (Kavlock *et al.*, 2002; Calafat and McKee, 2006).

DEHP is widely used in the production of many polyvinyl chloride-based plastics, including medical and food packages. Annual production of DEHP is 1–4 million tons, making it the most commercially important phthalate ester plasticizer (Mylchreest *et al.*, 2000). In a reference human population, multiple phthalate metabolites were detected in the urine of approximately 75% of those tested (Blount *et al.*, 2000), demonstrating that humans in the general population are exposed to significant levels. Daily exposure can vary greatly among individuals (3–30 μg/kg body weight/day), but is estimated to be 2 mg/day for DEPH alone (Fay, Donohue and DeRosa, 1999; Kavlock *et al.*, 2002). Additionally, workers in both the phthalate and PVC industries are estimated to have exposures of 143 and 286 μg/kg body weight/workday, respectively (Kavlock *et al.*, 2002). Correlations were identified between urinary concentrations of phthalates in humans and pregnancy complications (Tabacova, Little and Balabaeva, 1999) or rates of unsuccessful pregnancies (Aldyreva *et al.*, 1975). Additionally, decreased rates of pregnancy, increased rates of miscarriage, and anovulation have been associated with occupational exposure of Russian women to phthalates.

DEHP and its monoester metabolite (MEHP) are likely the most widely studied of the phthalate esters. Repeated oral exposure of female SD rats to DEHP caused delayed ovulation, reduced granulosa cell size in antral follicles, decreased circulating 17β-estradiol, progesterone and decreased LH levels, and increased FSH (Davis, Maronpot and Heindel, 1994). Mechanistic analyses have linked effects of phthalates to peroxisome proliferator-activated receptors (PPARs; reviewed in Lovekamp-Swan and Davis, 2003). It is known that multiple phthalates can induce hepatic 17β-hydroxysteroid dehydrogenase (17β-HSD IV), which metabolizes estradiol to estrone (Fan, Cattley and Corton, 1998). Further evidence demonstrates that this up-regulation of 17β HSD occurs through the PPAR-α receptor. In addition to PPAR-α, MEHP can also activate PPAR-γ, which suppresses expression of aromatase in granulosa cells of antral follicles (Lovekamp and Davis, 2001; Lovekamp-Swan, Jetten and Davis, 2003). From the responses to DEHP in rats, it can be concluded that these reproductive effects resulted from a specific targeting of large antral follicles, because of suppressed granulosa cell 17β-estradiol production. An *in vitro* study supported this conclusion because granulosa cells collected from diethylstilbestrol (DES)-primed female rats produced lower 17β-estradiol as increasing MEHP was added to the medium (Davis, Maronpot and Heindel, 1994). By evaluating the effect of FSH and dibutyryl-cAMP, it was concluded that inhibition of steroidogenesis was via inhibition of aromatase activity.

In males, DEHP caused a decrease in Sertoli cell proliferation *in vivo* (Heindel and Chapin, 1989). This chemical has been shown to inhibit FSH signal transduction in cultured rat Sertoli cells, and this may provide a mechanism for the *in vivo* observation. Furthermore, *in vitro* experiments have demonstrated that the effect results from inhibition of FSH-stimulated cAMP production. Based on these observations, the phthalate target is likely to be located in the Sertoli cell membrane. Therefore, DEHP and MEHP appear to have analogous targets within the testis (Sertoli cells) and ovary (granulosa cells). Male rat pups were dosed with increasing concentrations of the diester, di-*n*-pentylphthalate (DPP) (0, 0.25, 1, or 2 g/kg), and tissues collected weekly over a 10-week period. In the high-dose group, there was reduced fertility, and epididymal sperm count from three weeks of treatment, persisting until seven weeks after treatment. Testicular lesions were also reported in the high-dose (2 g/kg) animals (Lindström *et al.*, 1988; Sundaram and Witorsch, 1995).

Another phthalate, di-butylphthalate (DBP), is a component of adhesives, paper coatings, printing inks, aerosols, nail polishes, and hairsprays. In Western industrialized countries, annual production of DBP is 10,000–50,000 tons (Mylchreest *et al.*, 2000). Animal studies investigating transplacental exposure to DBP have demonstrated effects on sexual differentiation in male offspring. The effects include missing epididymides and vasa deferentia, effects on seminal vesicles, induction of hypospadias, decreased anogenital distance, retained thoracic nipples, and cryptorchidism (Mylchreest *et al.*, 2000). DBP exposure to male rats *in utero* has been shown to cause increases in the transcription factor Oct4 from embryonic days 15.5 to 17.5. Additionally, decreases in testicular germ cell number and testes descent were seen (Ferrara *et al.*, 2006). Although transplacental effects of DBP on reproductive tract development have been widely reported in male offspring, analogous effects in female offspring have not been observed. During fetal development, androgenic influences are critical in sexual differentiation in males, but not in females (Hoyer, 2001). Thus, this is likely due to the fact that male effects are thought to result from the anti-androgenic activity of DBP.

In summary, comparing male and female reproductive responses to phthalates provides interesting insight into gender differences and similarities. In rats, DEHP and MEHP are both male and female reproductive toxicants with opposing effects on FSH. Conversely, DBP and its active metabolite, monobutyl phthalate, produce developmental effects mainly in males. Thus, in males and females, within the class of phthalates, different compounds display similar as well as different effects.

## 5.5 Metals

Animal studies have provided evidence that several metals found in the environment have the potential to affect ovarian function either directly or indirectly via hypothalamic-pituitary targeting (Hoyer, 2005). There is evidence for testicular and ovarian accumulation of these metals and in some cases, they have been detected in seminal and follicular fluid. While there is little concrete evidence for adverse reproductive effects in humans caused by metals, future studies are required to determine whether environmental exposures to metals actually pose a risk to reproductive function.

***5.5.1 Cadmium*** Cadmium is one of the most extensively reported metals with respect to effects on reproductive function (Hoyer, 2005). Cadmium exposure in the environment is widespread and occurs via food, modern industrial processes, waste

disposal, and terrestrial and aquatic ecosystems (Nath *et al.*, 1984; Piasek and Laskey, 1994). One particular source is cigarette smoke (Zenzes *et al.*, 1995; Piasek *et al.*, 2001; Younglai *et al.*, 2002) and cadmium concentrations have been demonstrated to be higher in follicular fluid of smokers versus nonsmokers (Zenzes *et al.*, 1995). It was suggested that this accumulation might compromise the quality of oocytes, becoming a risk factor for fertility. However, the findings from two studies in which fertility was evaluated in women with high versus low levels of cadmium in follicular fluid did not support this conclusion (Drbohlav *et al.*, 1998; Younglai *et al.*, 2002). Cadmium has also been identified in seminal fluid of smokers and positive associations were observed between increased cadmium levels and reduction in testes size, along with increases in 17β-estradiol, FSH, and testosterone (Jurasovic *et al.*, 2004).

Ovarian accumulation of cadmium in animals can cause direct adverse effects. In mice, atretic follicles and degenerating corpora lutea were detected following exposure to cadmium (Godowicz and Pawlus, 1985). Cadmium exposure has also been reported to impact ovarian steroidogenesis. Acute exposure of rats to cadmium (up to 5 mg/kg) resulted in decreases in circulating progesterone and 17β-estradiol, and this was cycle stage-dependent (Piasek *et al.*, 1996). Support that the effects of cadmium on steroidogenesis result from direct ovarian targeting is provided by *in vitro* studies. Whole ovarian cell cultures from rats demonstrated particular susceptibility to *in vitro* cadmium exposure when tissues were collected during proestrous or early pregnancy. During those stages, progesterone and testosterone (Piasek and Laskey, 1999; Piasek *et al.*, 2002) as well as 17β-estradiol (Piasek and Laskey, 1994) decreased when cadmium was added to the medium.

Cadmium in male rats at two levels of exposure showed irreversible impairment of the germinal epithelium with toxic effects on spermatogenesis at high doses (Xu *et al.*, 2001). Additionally, daily oral dosing of male sheep with cadmium chloride (3 mg/kg) resulted in reductions in semen density and number of spermatozoa (Lymberopoulos *et al.*, 2000).

***5.5.2 Lead*** Exposure to lead is common as it is also widespread in the environment. The greatest source of exposure is from commercial products such as paint, printing material, and acid batteries (Dearth *et al.*, 2002). High levels of exposure can also occur in industrial settings, such as mining and refining plants (Taupeau *et al.*, 2001). A major concern related to lead exposure is its impact on fertility in women, fetal development, and prepubertal growth. Thus, lead exposure and toxicity are a leading environmental health concern for children, as well as women of childbearing age (Dearth *et al.*, 2002; Taupeau *et al.*, 2003; Hoyer, 2005).

One human study provided evidence for direct ovarian impairment of steroid production. Expression of aromatase and the estrogen receptor-beta mRNA and protein were significantly reduced by *in vitro* lead acetate exposure of granulosa cells collected from women undergoing *in vitro* fertilization (IVF; Taupeau *et al.*, 2003). Exposure to lead in animals has been shown to have a transplacental effect on the female F1 offspring. *In utero* exposure to lead chloride in SD rats resulted in decreased ovarian conversion of progesterone to androstenedione in prepubertal and pubertal females (Wiebe, Barr and Buckingham, 1988). It was concluded that prenatal exposure to lead significantly altered subsequent ovarian steroid production and gonadotropin binding. *In utero* exposure to lead acetate also resulted in delayed timing of puberty in female rat offspring, which was associated with suppressed serum levels of LH and 17β-estradiol (Dearth *et al.*, 2002). Another study showed that prepubertal exposure of

female rats to lead acetate resulted in delayed vaginal opening and disrupted estrous cyclicity (Ronis *et al.*, 1996).

In males, rat Leydig cells cultured with lead acetate showed reduced levels of testosterone and progesterone, along with reductions in steroidogenic enzymes (Thoreux-Manlay *et al.*, 1995). Epidemiological studies of men with high body blood lead levels have shown effects on sperm morphology (Telisman *et al.*, 2007). Another study of males exposed to high levels of occupational lead found structural damage to sperm (Naha and Chowdhury, 2006). Moreover, high seminal lead concentrations in sub-fertile males were found to be associated with decreased sperm motility and concentration (Pant *et al.*, 2003).

***5.5.3 Mercury*** Agriculture, consumption of fossil fuels, and industry are the main sources of mercury pollution in the environment. These are largely in the form of metallic mercury as inorganic or organic compounds. Additionally, ingestion of organomercurials from eating fish can be absorbed via the gastrointestinal tract, and eating fish from contaminated waters is a current concern (Hoyer, 2005). All chemical forms of mercury given to animals have been shown to cause female reproductive problems including infertility, disturbances in the estrous cycle, and inhibition of ovulation—all potentially arising from ovarian targeting (Schuurs, 1999). Mercuric chloride has been shown to accumulate in the female hamster pituitary (Lamperti and Printz, 1974); thus, exposure to mercury may cause hormonal effects by targeting a hypothalamic-pituitary site.

Low-dose mercuric chloride exposure in male mice exhibited reduced epididymal sperm number, dissociation of spermatozoa from the basement membrane, and necrosis. There were additional effects on Leydig cells and epididymis (Orisakwe *et al.*, 2001). A study of mercuric chloride treatment on bull sperm samples showed dose-dependent increases in DNA damage and decreases in cellular viability (Arabi, 2005). Furthermore, methylmercuric chloride exposure to male rats showed reductions in testes size and testosterone concentrations, and increases in testicular cell apoptosis (Homma-Takeda *et al.*, 2001).

## 5.6 Endocrine Disruptors

Endocrine disruptors that display estrogenic/antiestrogenic effects have been widely studied in their ability to target developmental and uterine sites of action, although intracellular mechanisms involved in these actions remain poorly understood. The effects of TCDD on sexual development and fertility have been studied in depth (Safe and Krishnan, 1995). There is much evidence for a hypothalamic-pituitary site of action, although this compound also has been shown to cause direct ovotoxicity (Li, Johnson and Rozman, 1995). The effect of a single oral dose of TCDD on numbers of oocytes ovulated and estrous cyclicity was observed in female rats (Li, Johnson and Rozman, 1995). Exposure to TCDD prolonged the diestrous stage and reduced the time in proestrus and estrus. Additionally, there was a reduction in the number of oocytes ovulated in treated rats. Thus, these findings provide strong evidence that TCDD impairs ovulation (Li, Johnson and Rozman, 1995). The data from this study were consistent with reduced ovulation via a hypothalamic-pituitary effect. However, ovulation was also reduced in hypophysectomized animals, suggesting a direct ovarian effect as well. The mechanism by which TCDD directly targets the ovary might be due to inhibition of 17β-estradiol production in granulosa cells, as suggested by *in vitro* studies (Moran

*et al.*, 1997). Further investigation demonstrated that TCDD produced these effects via interactions with the EGF linked-mitotic signaling pathway involving mitogen-activated protein kinase and protein kinase A. TCDD also affects male reproduction, and has been shown to cause reduced levels of plasma testosterone, reduced testis and epididymis weights, reduced sperm production, and alteration of the number of GnRH receptors in the pituitary gland (Bookstaff, Moore and Peterson, 1990; Bookstaff *et al.*, 1990; Mably, Moore and Peterson, 1992; Mably *et al.*, 1992). Additionally, a recent study showed that male rat pups exposed *in utero* to TCDD had increased testicular CYP1A1 and reduced CYP17 mRNA levels (Taketoh *et al.*, 2007).

The polychlorinated biphenyl (PCB) compound 3,4,5,5-tetrachlorobiphenyl (TCB) has been shown to be teratogenic in the mouse and embryolethal in the rat, as well as to cause transplacental ovarian toxicity in the mouse (Ronnback, 1991). Follicles at all stages of development were reduced 40% to 50% in female offspring at 28 days of age when mice were exposed *in utero* on day 13 of gestation. Interestingly, during a five-month period of testing, this extent of follicular damage did not adversely affect reproductive function in these offspring. Exposure of male rats to another PCB, 2,2′,3,4′,5′,6-hexachlorobiphenyl (PCB 132) has been shown to cause reductions in sperm count and motility (Hsu *et al.*, 2007).

Another widely studied endocrine disruptor has been bisphenol A (BPA; Papaconstantinou *et al.*, 2000). BPA is frequently used in the manufacture of polycarbonate plastics, epoxy resins, dental sealants, and as a stabilizing agent in plastics such as polyvinyl chloride. Possible exposure of humans to BPA has been reported due to its leaching from laboratory flasks, baby-feeding bottles, epoxy resins used for lining food cans, several types of dental sealants, and plastic waste samples. BPA has been detected in human follicular fluid in 1–2 ng/mL concentrations (Ikezuki *et al.*, 2002). *In vitro* animal studies have determined that BPA can bind the estrogen receptor ($\alpha$ and $\beta$), and induce estrogen-dependent gene expression/responses. Estrogenicity associated with BPA has been demonstrated in a number of *in vitro* and *in vivo* assays. *In vitro* assay end points include proliferation of MCF-7 human breast cancer cells, and the activation of estrogen response element (ERE)-driven reporter gene constructs. *In vivo* effects include the induction of vaginal cornification, growth and differentiation of the mammary gland, decrease in serum cholesterol levels, and increases in prolactin levels, uterine vascular permeability, and c-fos mRNA levels in the uterus and vagina (Papaconstantinou *et al.*, 2000). In mice, BPA exposure caused disruptions in puberty onset, regularity of estrous cyclicity (Markey *et al.*, 2002), and development of polycystic ovaries (Kato *et al.*, 2003). It has also been hypothesized that BPA may interfere with FSH-regulated progesterone synthesis in granulosa cells (Mlynarcikova, Fickova and Scsukova, 2005). *In utero* exposure of mice to BPA resulted in disturbances to recombination events during oogenesis (Susiarjo *et al.*, 2007).

Male mice exposed to BPA *in utero* or postnatally were shown to have reduced testes weights and lower testicular catalase activity (Kabuto, Amakawa and Shishibori, 2004); however, another study which exposed rats to BPA neonatally showed no effect on testicular development (Kato *et al.*, 2006). In spite of compelling animal data, to date, there have been no convincing demonstrations of human effects caused by BPA exposure.

## 5.7 Agricultural Compounds

***5.7.1 Pesticides*** A particular area of concern has focused on the effects of endocrine-disrupting chemical exposure on fish and wild life (Peterson *et al.*,

1997; Thomas, 1999). Feminization of male birds, alligators, and fish and the production of the estrogen-induced yolk precursor, vitellogenin, in male freshwater fish have been reported after exposure to xenobiotic estrogens, such as *o,p′*-dichlorodiphenyltrichloroethane (DDT), kepone, pulp kraft mill effluent, and sewage containing nonylphenols. DDT was used in the 1940s to control wartime typhus and agricultural pests (De Jager *et al.*, 2006). Juvenile male alligators exposed to dicofol and DDT in Florida had depressed plasma testosterone concentrations, poorly organized testes, and micropenises (Guillette *et al.*, 1995). Injection of fertilized gull eggs with DDT caused male embryos to develop ovarian tissue and oviducts (De Jager *et al.*, 2006). DDT pollution in California has been associated with feminization of male gulls. Dall's purpoises have elevated blubber concentrations of dichlorodiphenyldichloroethene (DDE) that are correlated with depressed plasma testosterone concentrations. An epidemiological study of men in Mexico showed that DDT exposure was associated with increased sperm tail defects and decreased sperm motility (De Jager *et al.*, 2006).

Methoxychlor (MXC) is an organochlorine pesticide that is presently used on agricultural crops as a replacement for DDT (Murono and Derk, 2005). MXC is believed to be metabolized by the liver to 2,2-bis(p-hydroxyphenyl)-1,1,1-trichloroethane (HPTE), which in some tissues is a more active metabolite than MXC (Murono, Derk and Akgul, 2006). MXC has been shown to cause reproductive effects in female as well as in male rodents. Early studies reported that MXC induces ovarian atrophy in mice (Eroschenko, Abuel-Atta and Grober, 1995) and decreases steroidogenesis in rat ovarian cells (Bal, 1984). More recently, *in vitro* studies have shown that HPTE in cultured rat granulosa cells inhibits progesterone and 17β-estradiol production by interfering with FSH signaling (Zachow and Uzumcu, 2006). Mechanistic investigations into ovarian effects of MXC on isolated mouse antral follicles have indicated that MXC accelerates atresia via increased apoptosis involving the Bcl-2 proto-oncogene family (Borgeest *et al.*, 2002; Miller *et al.*, 2005). In determining pathways with which MXC might interact, incubation of antral follicles with 17β-estradiol protected against MXC or HPTE-induced atresia (Miller, Gupta and Flaws, 2006). This finding was further supported in a study in which atresia was enhanced in antral follicles from MXC-exposed ERβ-overexpressing mice compared with their wild-type counterparts (Tomic *et al.*, 2006). Thus, it appears that ER-mediated pathways may mediate the ovarian toxicity of MXC. MXC has also been shown to exert its effects via increases in oxidative stress-induced mitochondrial damage (Gupta *et al.*, 2006a, b). Whether the mechanisms involving Bcl-2, ERβ and oxidative stress pathways interact in MXC-induced atresia is of interest.

In males, an early study reported that exposure to MXC caused testicular atrophy in rats (Tullner and Edgcomb, 1962). More information related to this effect has been provided by the demonstration that repeated exposure of rats to MXC (p.o.) decreased seminal vesicle weights, and circulating testosterone and dehydroepiandrosterone levels (Eroschenko, Abuel-Atta and Grober, 1995). Furthermore, in incubations of Leydig cells cultured from those animals, testosterone production and P450 cholesterol side-chain cleavage activity were reduced in cells from MXC-treated rats. In one *in vitro* study, HPTE inhibited basal and hCG-stimulated testosterone production in cultured fetal rat Leydig cells (Murono and Derk, 2005). In another *in vitro* study, HPTE inhibited LH-stimulated testosterone production in cultured Leydig cells from immature rats and the estrogen receptor antagonist ICI 182,780 reversed this inhibition (Akingbemi *et al.*, 2000). Long-term effects in male offspring were seen following

transplacental exposure to MXC (Amstislavsky *et al.*, 2006). At six months of age, testosterone levels were reduced, seminal vesicle weight was increased, and sexual arousal was compromised in males exposed *in utero* to MXC. Thus, MXC and its metabolite HPTE cause reproductive effects in females as well as in males, and these effects relate to compromised gonadal steroidogenesis. Interestingly, these effects appear to be mediated by the estrogen receptor in both genders; thus, MXC can be considered an endocrine disruptor.

***5.7.2 Fungicides*** Vinclozolin is a dicarboximide fungicide used to treat fruits, vegetables, ornamental plants, and turfgrass. Administration of vinclozolin to pregnant rats at 100 or 200 mg/kg/day demasculinizes and feminizes male offspring (Gray *et al.*, 1999). This effect is thought to be the result of vinclozolin interacting with the androgen receptor, and in that manner acting as an endocrine disruptor. Earlier studies reported the lowest observed effect level (LOEL) for vinclozolin ranged from 50 to 111 g/kg/day (Van Ravenzwaay, 1992). However, in a study in which even lower doses of vinclozolin were given to pregnant rats (3.125 and 6.25 mg/kg/day), subtle alterations in the differentiation of the external genitalia, ventral prostate, and nipple tissue in male offspring were observed. This is an order of magnitude below previously reported developmental no observed effect levels (NOELs, 22–77 mg/kg/day; Amstislavsky *et al.*, 2006). Thus, it has been proposed that the anti-androgenic effects of vinclozolin can be produced at very low levels, and studies that do not include androgen-sensitive end points might yield NOELs that are at least an order of magnitude too high. Vinclozolin also has an *in utero* effect on male offspring, with reductions in numbers of spermatids and epididymal sperm, sperm motility, and increased spermatogenic cell apoptosis (Anway *et al.*, 2006).

***5.7.3 Psoralens*** The psoralens (or linear furanocoumarins) are photoactivated (Cole, 1970; Murray, Mendez and Brown, 1982; Trumble *et al.*, 1991) naturally occurring plant biosynthetic metabolites that are found in several common fruits and vegetables (Beier, 1990; Berenbaum, 1991; Diawara and Trumble, 1997). Synthetic 5-methoxypsoralen (bergapten) and 5- and 8-methoxypsoralen (5-MOP; 8-MOP; xanthotoxin or methoxsalen) are widely used in the treatment of several skin disorders including psoriasis, conditions of skin depigmentation (such as leprosy, vitiligio, and leucoderma), mycosis fungoides, polymorphous dermatitis, and eczema (Musajo and Rodighiero, 1962; Scott, Pathak and Mohm, 1976; Van Scott, 1976; Pathak and Fitzpatrick, 1992; Chadwick *et al.*, 1994; Ehrsson *et al.*, 1994). However, the psoralens have been linked to a higher incidence of skin cancer (Musajo and Rodighiero, 1962; Scott, Pathak and Mohm, 1976; Van Scott, 1976; Stern *et al.*, 1979; Grekin and Epstein, 1981) and other disorders such as sister chromatid exchange, gene mutation, and chromosomal aberrations in humans (Calzavara-Pinton *et al.*, 1994). Studies on the effect of psoralens on reproductive function have reported a dose-dependent effect of 5-MOP and 8-MOP in the presence of UVA light on birth rate in female rats. The number of offspring was reduced along with a reduction in uterine weight and serum estradiol levels (Diawara *et al.*, 1999). This was determined to be due to reduced ovulations resulting from targeting of granulosa cells. Additionally, male rats treated with 5-MOP or 8-MOP in the absence of UVA light showed a decrease in pituitary gland weight and number of sperm, with an increase in relative testicular weights and serum testosterone. Subsequent matings in this study resulted in a reduced number of pregnancies in females (Diawara *et al.*, 2001).

***5.7.4 Atrazine*** Atrazine is a chlorotriazine and was introduced in the 1950s as a broad-spectrum herbicide and is widely used worldwide (Eldridge, Wetzel and Tyrey, 1999; Cooper *et al.*, 2007). Total annual use of atrazine is approximately 167 kg of active ingredient. In female SD rats, atrazine caused mammary tumors which were attributed to a hormonal influence due to reproductive aging (Wetzel *et al.*, 1994). Hysterectomy prevented mammary tumor development due to atrazine exposure. It is thought that atrazine prevents the LH surge associated with ovulation, thus supporting a hypothalamic/pituitary site of atrazine action. In addition, atrazine exposure leads to significant increases in serum levels of estrone and 17β-estradiol in male and ovariectomized female rats (Cooper *et al.*, 2007). Male rats exposed to atrazine *in utero* had lower body weights, experienced delayed preputial separation, and increased prostate inflammation compared with controls (Rayner *et al.*, 2007). Additionally, atrazine exposure resulted in delayed puberty onset in both male and female rats (Stoker *et al.*, 2002; Laws *et al.*, 2003).

## 6 TRANSLATIONAL APPLICATIONS

A major goal of toxicological research is to enhance the ability to predict and evaluate potential risk of the environment to human health. However, controlled toxicological studies are usually carried out in laboratory rodents because of the economy of using them in large numbers, along with the fact that controlled, manipulative studies in humans are rarely feasible. A reasonable criticism of this approach is the validity of extrapolating responses in rodents to humans. This is especially true of reproductive physiology, in which species variation is the hallmark. Thus, the ideal goal is to coordinate human information related to exposure outcomes with that obtained from animal studies.

Assessment of reproductive effects in humans usually begins with a recognizable outcome in women (acyclicity, delayed time to conception, early menopause), pregnancy outcome (spontaneous abortion) or in men (reduced sperm production, impotence, infertility). Generally, identification of reproductive effects in humans progresses slowly because it relies on population-based and clinical observations (MacKenzie and Angevine, 1981; Whelan, 1997). There are, however, several situations in which reproductive effects in humans have been linked to exposures before animal studies have predicted the outcomes. For example, a number of infertility cases were discovered in men working in a California pesticide factory (Whorton *et al.*, 1977). The suspected cause was exposure to 1,2-dibromo-3-chloropropane (DBCP). Upon evaluation of these men, major effects including azoospermia or oligospermia were accompanied by increased circulating levels of FSH and LH. Follow-up testing in rats has demonstrated that DBCP is directly toxic to sperm (Kluwe *et al.*, 1983). Another example is the case of 2-BP which was discovered to have toxic effects on the reproductive systems of both males and females after occupational exposure of workers in South Korea. Evaluation of workers exposed in factories to this industrial propellant and cleaning solvent uncovered a high incidence of bone marrow effects as well as secondary amenorrhea accompanied by increased circulating FSH and LH levels and hot flashes in 16 of 25 exposed women. Additionally, six of eight exposed men displayed azoospermia or oligospermia, and reduced sperm motility. Based on these observations, animal studies were designed and have since shown that 2-BP causes destruction of ovarian follicles and spermatogonia (Omura *et al.*, 1999; Yu *et al.*, 1999).

Epidemiological studies conducted over the last four decades have demonstrated a relationship between smoking and impaired fertility. Cigarette smoke is a well-known reproductive toxicant. One study reported that rates of pregnancy were reduced to 75% in light smokers and 57% in heavy smokers when compared with nonsmokers. Furthermore, smokers required one year longer to conceive than did nonsmokers (Baird and Wilcox, 1985). Along with the impact on fertility, there were also effects of cigarette smoking on pregnancy and the fetus. Prenatal exposure to cigarette smoke has been associated with retarded intrauterine growth and premature deliveries (Mattison *et al.*, 1989). Additionally, conception in women whose mothers smoked while pregnant was significantly reduced when compared with women whose mothers had not (Weinberg, Wilcox and Baird, 1989). Decreased follicular levels of 17β-estradiol were observed in smoking women when compared with nonsmokers, and cultured human granulosa cells were reported to secrete decreased amounts of 17β-estradiol in the presence of an extract of cigarette smoke (VanVoorhis *et al.*, 1992). In a study of females undergoing IVF, increased levels of the PAH, BaP, were found in serum and follicular fluid in those who were smokers, and this was associated with lower rates of implantation and successful pregnancies compared with control nonsmoking women (Neal *et al.*, 2007). Thus, these effects may relate to the infertility associated with cigarette smoking. When the ovaries of women are depleted of primordial follicles, the result is menopause. Because the ovary contains a finite number of oocytes at birth, human exposure to environmental chemicals that destroy primordial follicles can cause early menopause. Female smokers have also been reported to experience a one- to four-year-earlier age for the onset of menopause (Fig. 4) (Jick and Porter, 1977). Therefore, in women, cigarette smoking may impair fertility as well as accelerate the onset of menopause.

Cigarette smoke is a complex mixture of alkaloids (nicotine), PAHs, nitroso compounds, aromatic amines, and protein pyrosylates, many of which are carcinogenic (Stedman, 1968). Several ovotoxic PAHs are found in cigarette smoke, including DMBA, 3-MC, and BaP. Because these PAHs are known to damage primordial and primary follicles in mice and rats, direct oocyte destruction caused by one or more of these chemicals is the most logical mechanism by which cigarette smoke induces early menopause (Mattison *et al.*, 1989). Of additional concern is the finding in animal studies

**Figure 4** Effect of smoking on age of menopause. Toxicants in cigarette smoke appear to destroy small oocytes. No major effects on fertility are noted during the reproductive life span, but shortening of reproductive life span occurs (adapted from Jick and Porter, 1977).

that exposure of mice *in utero* to cigarette smoke resulted in a smaller number of ovarian primordial follicles in female offspring (Vahakangas, Rajaniemi and Pelkonen, 1985). Secondary follicles in laboratory animals are also susceptible to PAHs (Mattison, 1980; Borman *et al.*, 2000). Thus, in women smokers, prior to early menopause, damage to larger follicles (secondary through antral) caused by exposure to PAHs could also cause problems with fertility. Another component of cigarette smoke, nicotine, acting on the central nervous system, may affect the secretion of hormones involved in regulation of ovarian function and in that way may be responsible for impaired fertility in female smokers (Jick and Porter, 1977). In addition to ovarian effects, cigarette smoke may also induce liver-metabolizing enzymes which may accelerate metabolism of steroid hormones thereby compromising gonadal function (Jick and Porter, 1977; Yeh and Barbieri, 1989). Information obtained from animal studies may be useful for planning epidemiological investigations about potential reproductive effects resulting from exposure of women to ovotoxic chemicals in cigarette smoke.

## 7 CONCLUSIONS AND FUTURE DIRECTIONS

Once the reproductive effects of a chemical have been identified and characterized, it becomes of interest to elucidate mechanisms by which they occur. It is important, therefore, to determine cellular targets and intracellular pathways that are involved. In the case of some chemicals, mechanistic insight has been gained; however, this is generally limited in humans and more research is required to further develop the field of reproductive toxicology. The application of widely available molecular technologies that have been emerging over the past decade can help achieve this aim. Three of these hold particular promise for the field: DNA microarray, laser capture microdissection (LCM), and proteomic analyses tools. DNA microarray involves examining changes that occur in gene expression during xenobiotic exposure. Profiling DNA microarrays can be used to evaluate changes in expression of genes that occur due to a specific chemical exposure. A second use of this method is in single nucleotide polymorphism (SNP) profiling. This area will be useful for determining individual susceptibility to reproductive toxicants. This application also allows genetic screening for mutations, which may impair reproductive capacity as well as enhance susceptibility to cancer. A third use of DNA microarrays involves identifying effects of chemical exposures on regulatory regions of genes that are important in cell signaling using chromatin immunoprecipitation assay (ChIP-on-chip). LCM is a tool that allows specific regions of heterogeneous tissues to be isolated for evaluation of changes in a specific cell population. Proteomics consists of an array of techniques that allows proteins to be separated from complex mixtures, and analyzed for post-transcriptional/translational modifications caused by chemical exposures. In addition, transgenic animals that either over- or underexpress genes of importance in xenobiotic metabolism are proving to play vital roles in determining pathways that are involved in effects on reproductive toxicology.

In conclusion, environmental exposure to xenobiotics that have the potential to impact human reproduction has increased. Research in this field in the next decade is likely to see an increase in mechanistic studies that will identify cellular and molecular events involved in producing effects caused by a wide variety of environmental chemicals using the emerging technologies that have been discussed. Gaining improved mechanistic insight into the effects of reproductive toxicants in the laboratory

will provide a greater understanding for translating these effects into more accurate assessments of human risk.

## REFERENCES


Akingbemi, B.T., Ge, R.S., Klinefelter, G.R., Gunsalus, G.L., Hardy, M.P. (2000). A metabolite of methoxychlor, 2,2-bis(p-hydroxyphenyl)-1,1,1-trichloroethane, reduces testosterone biosynthesis in rat leydig cells through suppression of steady-state messenger ribonucleic acid levels of the cholesterol side-chain cleavage enzyme. *Biol. Reprod.*, *62*, 571–578.

Aldyreva, M.V., Klimova, T.S., Iziumova, A.S., Timofeevskaia, L.A. (1975). The effect of phthalate plasticizers on the generative function. *Gig. Tr. Prof. Zabol.*, *12*, 25–29.

Amstislavsky, S.Y., Amstislavskaya, T.G., Amstislavsky, V.S., Tibeikina, M.A., Osipov, K.V., Eroschenko, V.P. (2006). Reproductive abnormalities in adult male mice following preimplantation exposures to estradiol or pesticide methoxychlor. *Reprod. Toxicol.*, *21*, 154–150.

Anway, M.D., Memon, M.A., Uzumcu, M., Skinner, M.K. (2006). Transgenerational effect of the endocrine disruptor vinclozolin on male spermatogenesis. *J. Androl.*, *27*, 868–879.

Arabi, M. (2005). Bull spermatozoa under mercury stress. *Reprod. Domest. Anim.*, *40*, 454–459.

Ataya, K., Rao, L.V., Lawrence, E., Kimmel, R. (1995). Luteinizing hormone-releasing hormone agonist inhibits cyclophosphamide-induced ovarian follicular depletion in rhesus monkeys. *Biol. Reprod.*, *52*, 365–372.

Baird, D.D., Wilcox, A.J. (1985). Cigarette smoking associated with delayed conception. *JAMA*, *253*, 2979–2983.

Bal, H.S. (1984). Effect of methoxychlor on reproductive systems of the rat. *Proc. Soc. Exp. Biol. Med.*, *176*, 187–196.

Beamer, W.G., Tennent, B.J. (1986). Gonadotropin uptake in genetic and irradiation models of ovarian tumorigenesis. *Biol. Reprod.*, *34*, 761–770.

Beier, R.C. (1990). Natural pesticides and bioactive components in food. *Rev. Environ. Contam. Toxicol.*, *113*, 47–137.

Bengtsson, M., Hamberger, L., Rydstrom, J. (1988). Metabolism of 7,12-dimethylbenz[a]anthracene by different types of cells in the human ovary. *Xenobiotica*, *18*, 1255–1270.

Bengtsson, M., Mattison, D.R. (1989). Gonadotropin-dependent metabolism of 7,12-dimethylbenz(a)anthracene in the ovary of rhesus monkey. *Biochem. Pharmacol.*, *38*, 1869–1872.

Bengtsson, M., Montelius, J., Mankovitz, L., Rydstrom, J. (1983). Metabolism of polycyclic aromatic hydrocarbons in the rat ovary: comparison with metabolism in adrenal and liver tissues. *Biochem. Pharmacol.*, *32*, 129–136.

Bengtsson, M., Reinholt, F.P., Rydstrom, J. (1992). Cellular localization and hormonal regulation of 7,12-dimethylbenzanthracene mono-oxygenase activity in the rat ovary. *Toxicology*, *71*, 203–222.

Bengtsson, M., Rydstrom, J. (1983). Regulation of carcinogen metabolism in the rat ovary by the estrous cycle and gonadotropin. *Science*, *219*, 1437–1438.

Berenbaum, M.R. (1991). Coumarins, in *Herbivores: Their Interactions with Secondary Plant Metabolites, the Chemical Participants*, Vol. 1 (eds G.A. Rosenthal and M.R. Berenbaum), Academic Press, New York, pp. 221–249.

Bevan, C., Stadler, J.C., Elliot, G.S., Frame, S.R., Baldwin, J.K., Leung, H.W., Moran, E., Panepinto, A.S. (1996). Subchronic toxicity of 4-vinylcyclohexene in rats and mice by inhalation exposure. *Fundam. Appl. Toxicol.*, *32*, 1–10.

Blount, B.C., Silva, M.J., Caudill, S.P., Needham, L.L., Pirkle, J.L., Sampson, E.J., Lucier, G.W., Jackson, R.J., Brock, J.W. (2000). Levels of seven urinary phthalate metabolites in a human reference population. *Environ. Health Perspect.*, *108*, 979–982.

Blumenfeld, Z., Dann, E., Avivi, I., Epelbaum, R., Rowe, J.M. (2002). Fertility after treatment for Hodgkins disease. *Ann. Oncol.*, *13*, 138–147.

Boekelheide, K. (2005). Mechanisms of toxic damage to spermatogenesis. *J. Natl. Cancer Inst. Monogr.*, *34*, 6–8.

Bookstaff, R.C., Kamel, F., Moore, R.W., Bjerke, D.L., Peterson, R.E. (1990). Altered regulation of pituitary gonadotropin-releasing hormone (GnRH) receptor number and pituitary responsiveness to GnRH in 2,3,7,8-tetrachlorodibenzo-p-dioxin-treated male rats. *Toxicol. Appl. Pharmacol.*, *105* (1), 78–92.

Bookstaff, R.C., Moore, R.W., Peterson, R.E. (1990). 2,3,7,8-tetrachlorodibenzo-p-dioxin increases the potency of androgens and estrogens as feedback inhibitors of luteinizing hormone secretion in male rats. *Toxicol. Appl. Pharmacol.*, *104* (2), 212–224.

Borgeest, C., Symonds, D., Mayer, L.P., Hoyer, P.B., Flaws, J.A. (2002). Methoxychlor may cause ovarian follicular atresia and proliferation of the ovarian epithelium in the mouse. *Toxicol. Sci.*, *68*, 473–478.

Borman, S.M., Christian, P.J., Sipes, I.G., Hoyer, P.B. (2000). Ovotoxicity in female Fischer rats and B6 mice induced by low-dose exposure to three polycyclic aromatic hydrocarbons: comparison through calculation of an ovotoxic index. *Toxicol. Appl. Pharmacol.*, *167*, 191–198.

Borman, S.M., VanDePol, B.J., Kao, S., Thompson, K.E., Sipes, I.G., Hoyer, P.B. (1999). A single dose of the ovotoxicant 4-vinylcyclohexene diepoxide is protective in rat primary ovarian follicles. *Toxicol. Appl. Pharmacol.*, *158*, 244–252.

Botta, J.A., Jr, Hawkins, H.C., Weikel, J.H., Jr (1974). Effects of cyclophosphamide on fertility and general reproductive performance in rats. *Toxicol. Appl. Pharmacol.*, *27*, 602–611.

Buccione, R., Schroeder, A.C., Eppig, J.J. (1990). Interactions between somatic cells and germ cells throughout mammalian oogenesis. *Biol. Reprod.*, *43*, 543–547.

Buhrley, L.E., Ellis, L.C. (1973). Catalase activity in rat testicular preparations: distribution and changes in activity induced by hypophysectomy and x-irradiation. *Fertil. Steril.*, *24*, 956–961.

Byrne, J. (1990). Fertility and pregnancy after malignancy. *Semin. Perinatol.*, *14*, 423–429.

Calafat, A.M., McKee, R.H. (2006). Integrating biomonitoring exposure data into the risk assessment process: phthalates [diethyl phthalate and di(2-ethylhexyl) phthalate] as a case study. *Environ. Health Perspect.*, *114*, 1783–1789.

Calzavara-Pinton, P., Carlino, A., Manfedi, E., Semeraro, F., Zane, C., Panfilis, G. (1994). Ocular side-effects of PUVA-treated patients refusing eye sun protection. *Acta Derm. Venereol. Suppl. (Stockh)*, *186*, 164–165.

Cannady, E.A., Dyer, C.A., Christian, P.J., Sipes, I.G., Hoyer, P.B. (2002). Expression and activity of microsomal epoxide hydrolase in follicles isolated from mouse ovaries. *Toxicol. Sci.*, *68*, 24–31.

Chadwick, C.A., Potten, C.S., Cohen, A.J., Young, A.R. (1994). The time of onset and duration of 5-methoxypsoralen photochemoprotection from UVR-induced DNA damage in human skin. *Br. J. Dermatol.*, *131*, 483–494.

Chapin, R.E., Gray, T.J., Phelps, J.L., Dutton, S.L. (1988). The effects of mono-(2-ethylhexyl)-phthalate on rat Sertoli cell-enriched primary cultures. *Toxicol. Appl. Pharmacol.*, *92*, 467–479.

Chapman, R.M. (1983). Gonadal injury resulting from chemotherapy. *Am. J. Ind. Med.*, *4*, 149–161.

Citti, L., Gervasi, P.G., Turchi, G., Bellucci, G., Bianchini, R. (1984). The reaction of 3,4-epoxy-1-butene with deoxyguanosine and DNA in vitro: synthesis and characterization of the main adducts. *Carcinogenesis*, *5*, 47–52.

Cole, R.S. (1970). Light-induced cross-linking of DNA in the presence of a furocoumarin (psoralen). Studies with phase lambda, Escherichia coli and mouse leukemia cells. *Biochem. Biophys. Acta*, *217*, 30–39.

Constanzo, L.S. (1998). Reproductive physiology, in *Physiology* (eds W.R. Schmitt and A.M. Shaw), W.B. Saunders Co., Philadelphia, PA, pp. 393–411.

Cooper, R.L., Laws, S.C., Das, P.C., Narotsky, M.G., Goldman, J.M., Lee Tyrey, E., Stoker, T.E. (2007). Atrazine and reproductive function: mode and mechanism of action studies. *Birth Defects Res. B Dev. Reprod. Toxicol.*, *80*, 98–112.

Damewood, M.D., Grochow, L.B. (1986). Prospects for fertility after chemotherapy or radiation for neoplastic disease. *Fertil. Steril.*, *45*, 443–459.

Davis, B.J., Maronpot, R.R., Heindel, J.J. (1994). Di-(2-ethylhexyl)phthalate suppresses estradiol and ovulation in cycling rats. *Toxicol. Appl. Pharmacol.*, *128*, 216–223.

De Jager, C., Farias, P., Barraza-Villarreal, A., Avila, M.H., Ayotte, P., Dewailly, E., Dombrowski, C., Rousseau, F., Sanchez, V.D., Bailey, J.L. (2006). Reduced seminal parameters breakthroughs in andrology associated with environmental DDT exposure and p,p′-DDE concentrations in men in Chiapas, Mexico: a cross-sectional study. *J. Androl.*, *27* (1), 16–27.

Dearth, R.K., Hiney, J.K., Srivastave, V., Burdick, S.B., Bratton, G.R., Dees, W.L. (2002). Effects of lead (Pb) exposure during gestation and lactation on female prepubertal development in the rat. *Reprod. Toxicol.*, *16*, 343–352.

DeRaat, W.K. (1978). Induction of sister chromatid exchanges by styrene and its presumed metabolite styrene oxide in the presence of rat liver homogenate. *Chem. Biol. Interact.*, *20*, 163–170.

Devine, P.J., Hoyer, P.B. (2005). Ovotoxic environmental chemicals: indirect endocrine disruptors, in *Endocrine Disruptors: Effects on Male and Female Reproductive System* (ed. R. Nog), CRC Press, Boca Raton, FL, pp. 67–100.

Diawara, M.M., Chavez, K.J., Hoyer, P.B., Williams, D.E., Dorsch, J., Kulkosky, P., Franklin, M.R. (1999). A novel group of ovarian toxicants: the psoralens. *J. Biochem. Mol. Toxicol.*, *13* (3/4), 195–203.

Diawara, M.M., Chavez, K.J., Simpleman, D., Williams, D.E., Franklin, M.R., Hoyer, P.B. (2001). The psoralens adversely affect reproductive function in male wistar rats. *Reprod. Toxicol.*, *15*, 137–144.

Diawara, M.M., Trumble, J.T. (1997). Linear furanocoumarins, in *Handbook on Plant and Fungal Toxicants* (ed. J.P.F. D'Mello), CRC Press, Boca Raton, FL, pp. 175–189.

Dnistrian, A.M., Schwartz, M.K., Fracchia, A.A., Kaufman, R.J., Hakes, T.B., Currie, V.E. (1983). Endocrine consequences of CMF adjuvant therapy in premenopausal and postmenopausal breast cancer patients. *Cancer*, *51*, 803–807.

Docampo, R. (1987). Male contraception and chemotherapy for trypanosomiasis and leishmaniasis: an analysis of biochemical similarities and considerations for future research, in *Cellular and Molecular Events in Spermiogenesis*, Cambridge University Press, Oaxtepec, Mexico.

Doerr, J.K., Hooser, S.B., Smith, B.J., Sipes, I.G. (1995). Ovarian toxicity of 4-vinylcyclohexene and related olefins in B6C3F1 mice: role of diepoxides. *Chem. Res. Toxicol.*, *8*, 963–969.

Drbohlav, P., Bencko, V., Masata, J., Bendl, J., Rezacova, J., Zouhar, T., Cerny, V., Halkova, E. (1998). Detection of cadmium and zinc in the blood and follicular fluid of women in the IVF and ET program. *Ceska Gynekol.*, *63*, 292–300.

Ehrsson, H., Wallin, I., Ros, A.M., Eksborg, S., Berg, M. (1994). Food-induced increase in bioavailability of 5-methoxypsoralen. *Eur. J. Clin. Pharmacol.*, *46*, 375–377.

Eldridge, J.C., Wetzel, L.T., Tyrey, L. (1999). Estrous cycle patterns of Sprague-Dawley rats during acute and chronic atrazine administration. *Reprod. Toxicol.*, *13*, 491–199.

Eroschenko, V.P., Abuel-Atta, A.A., Grober, M.S. (1995). Neonatal exposures to technical methoxychlor alters ovaries in adult mice. *Reprod. Toxicol.*, *9*, 379–387.

Fan, L.Q., Cattley, R.C., Corton, J.C. (1998). Tissue-specific induction of 17 beta-hydroxysteroid dehydrogenase type IV by peroxisome proliferator-activated receptor alpha. *J. Endocrinol.*, *158*, 237–246.

Fay, M., Donohue, J.M., DeRosa, C. (1999). ATSDR evaluation of health effects of chemicals. VI. Di(2-ethylhexyl)phthalate. Agency for toxic substances and disease registry. *Toxicol. Ind. Health*, *15*, 651–746.

Fernandez, E., La Vecchia, C., Braga, C., Talamini, R., Negri, E., Parazzini, F., Franceschi, S. (1998). Hormone replacement therapy and risk of colon and rectal cancer. *Cancer Epidemiol. Biomarkers Prev.*, *7*, 329–333.

Ferrara, D., Hallmark, N., Scott, H., Brown, R., McKinnell, C., Mahood, I.K., Sharpe, R.M. (2006). Acute and long-term effects of in utero exposure of rats to di(n-butyl) phthalate on testicular germ cell development and proliferation. *Endocrinology*, *147* (11), 5352–5362.

Flaws, J.A., Salyers, K.L., Sipes, I.G., Hoyer, P.B. (1994). Reduced ability of rat preantral ovarian follicles to metabolize 4-vinyl-1-cyclohexene diepoxide in vitro. *Toxicol. Appl. Pharmacol.*, *126*, 286–294.

Foster, P.M. (1988). M-dinitrobenzene: studies on its toxicity to the testicular sertoli cell. *Arch. Toxicol.* Suppl., *13*, 3–17.

Garcia, C.R. (2007). Fertility risk in pediatric and adolescent cancers. XVIth Ovarian Workshop. Ovarian Differentiation, development, function and persistence. San Antonio, TX, Biosymposia, Inc.

Generoso, W.M., Stout, S.K., Huff, S.W. (1971). Effects of alkylating chemicals on reproductive capacity of adult female mice. *Mutat. Res.*, *13*, 172–184.

Georgellis, A., Parvinen, M., Rydström, J. (1989). Inhibition of stage-specific DNA synthesis in rat spermatogenic cells by polycyclic aromatic hydrocarbons. *Chem. Biol. Interact.*, *72*, 79–92.

Georgellis, A., Toppari, J., Veromaa, T., Rydstrom, J., Parvinen, M. (1990). Inhibition of meiotic divisions of rat spermatocytes in vitro by polycyclic aromatic hydrocarbons. *Mutat. Res.*, *231*, 125–135.

Godowicz, P., Pawlus, M. (1985). Effect of cadmium chloride on the ovulation and structure of ovary in the inbred KP and CBA mice strains. *Folia Histochem.*, *23*, 209–215.

Gray, L.E., Jr, Ostby, J., Monosson, E., Kelce, W.R. (1999). Environmental antiandrogens: low doses of the fungicide vinclozolin alter sexual differentiation of the male rat. *Toxicol. Ind. Health*, *15*, 48–64.

Grekin, D.A., Epstein, J.H. (1981). Psoralens, UVA (PUVA) and photocarcinogenesis. *Photochem. Photobiol.*, *33*, 957–960.

Grizzle, T.B., George, J.D., Fail, P.A., Seely, J.C., Heindel, J.J. (1994). Reproductive effects of 4-vinylcyclohexene in Swiss mice assessed by a continuous breeding protocol. *Fundam. Appl. Toxicol.*, *22*, 122–129.

Guillette, L.J., Jr, Pickford, D.B., Crain, D.A., Rooney, A.A., Percival, H.F. (1995). Reduction in penis size and plasma testosterone concentrations in juvenile alligators living in a contaminated environment. *Gen. Comp. Endocrinol.*, *101*, 32–42.

Gupta, R.K., Miller, K.P., Babus, J.K., Flaws, J.A. (2006a). Methoxychlor inhibits growth and induces atresia of antral follicles through an oxidative stress pathway. *Toxicol. Sci.*, *93*, 382–389.

Gupta, R.K., Schuh, R.A., Fiskum, G., Flaws, J.A. (2006b). Methoxychlor causes mitochondrial dysfunction and oxidative damage in the mouse ovary. *Toxicol. Appl. Pharmcol.*, *216*, 436–445.

Heindel, J.J., Chapin, R.E. (1989). Inhibition of FSH-stimulated cAMP accumulation by mono(2-ethylhexyl) phthalate in primary rat sertoli cell cultures. *Toxicol. Appl. Pharmacol.*, *97*, 377–385.

Heinrichs, W.L., Juchau, M.R. (1980). *Extrahepatic Metabolism of Drugs and Other Foreign Compounds*, Vol. 1, SP Medical and Scientific Books, New York, p. 319.

Hendry, W.J., DeBrot, B.L., Zheng, X.L., Branham, W.S., Sheehan, D.M. (1999). Differential activity of diethylstilbestrol versus estradiol as neonatal endocrine disrupters in the female hamster (Mesocricetus auratus) reproductive tract. *Biol. Reprod.*, *61*, 91–100.

Herbert, D.C., Suprakar, P.C., Roy, A.K. (1995). Male reproduction, in *Reproductive Toxicology* (ed. R.J. Witorsch), Raven Press, New York, pp. 3–22.

Herbst, A.L., Ulfelder, H., Poskanzer, D.C. (1971). Adenocarcinoma of the vagina. Association of maternal stilbestrol therapy with tumor appearance in young women. *N. Engl. J. Med.*, *284*, 878–881.

Hirshfield, A.N. (1991). Development of follicles in the mammalian ovary. *Int. Rev. Cytol.*, *124*, 43–101.

Homma-Takeda, S., Kugenuma, Y., Iwamuro, T., Kumagai, Y., Shimojo, N. (2001). Impairment of spermatogenesis in rats by methylmercury: involvement of stage- and cell-specific germ cell apoptosis. *Toxicology*, *169*, 25–35.

Hooser, S.B., DeMerell, D.G., Douds, D.A., Hoyer, P.B., Sipes, I.G. (1995). Testicular germ cell toxicity caused by vinylcyclohexene diepoxide in mice. *Reprod. Toxicol.*, *9* (4), 359–367.

Hooser, S.B., Douds, D.A., Hoyer, P.B., Sipes, I.G. (1994). Long-term ovarian and hormonal alterations due to the ovotoxin, 4-vinylcyclohexene. *Reprod. Toxicol.*, *8*, 315–323.

Howell, S.J., Shalet, S.M. (2002a). Testicular function following chemotherapy. *Hum. Reprod. Update*, *7*, 363–369.

Howell, S.J., Shalet, S.M. (2002b). Effect of cancer therapy on pituitary-testicular axis. *Int. J. Androl.*, *25*, 269–276.

Hoyer, P.B. (1997). Female reproductive toxicology: introduction and overview, in *Comprehensive Toxicology: Reproductive and Endocrine Toxicology* (eds K. Boekelheide, R. Chapin, P.B. Hoyer, C. Harris, I.G. Sipes, C.A. McQueen and A.J. Gandolfi), Elsevier, p. 10.

Hoyer, P.B. (2001). Reproductive toxicology: current and future directions. *Biochem. Pharmacol.*, *62*, 1557–1564.

Hoyer, P.B. (2005). Impact of metals on ovarian function, in *Metals, Fertility and Reproductive Toxicity*, CRC Press, Boca Raton, FL, pp. 155–173.

Hoyer, P.B., Devine, P.J. (2002). Endocrine toxicology: the female reproductive system, in *Handbook of Toxicology* (eds M.J. Derelanko and M.A. Hollinger), CRC Press, pp. 573–595.

Hoyer, P.B., Sipes, I.G. (1996). Assessment of follicle destruction in chemical-induced ovarian toxicity. *Ann. Rev. Pharmacol. Toxicol.*, *36*, 307–331.

Hoyer, P.B., Sipes, I.G. (2007). Development of an animal model for ovotoxicity using 4-vinylcyclohexene diepoxide: a case study. *Birth Defects Res. Part B Dev. Reprod. Toxicol.*, *80*, 113–125.

Hsu, P.C., Pan, M.H., Li, L.A., Chen, C.J., Tsai, S.S., Guo, Y.L. (2007). Exposure in utero to 2,2′,3,3′,4,6′-hexachlorobiphenyl (PCB 132) impairs sperm function and alters testicular apoptosis-related gene expression in rat offspring. *Toxicol. Appl. Pharmacol.*, *221*, 68–75.

Hu, X., Christian, P.J., Sipes, I.G., Hoyer, P.B. (2002a). Expression and redistribution of cellular Bad, Bax, and Bcl-X(L) protein is associated with VCD-induced ovotoxicity in rats. *Biol. Reprod.*, *65*, 1489–1495.

Hu, X., Flaws, J.A., Sipes, I.G., Hoyer, P.B. (2002b). Activation of mitogen-activated protein kinases and AP-1 transcription factor in ovotoxicity induced by 4-vinyl cyclohexene diepoxide in rats. *Biol. Reprod.*, *67*, 718–724.

Huang, H.F., Yang, C.S., Meyenhofer, M., Gould, S., Boccabella, A.V. (1988). Disruption of sustentacular (Sertoli) cell tight junctions and regression of spermatogenesis in vitamin-A-deficient rats. *Acta Anat. (Basel)*, *133*, 10–15.

Huyghe, E., Matsuda, T., Daudin, M., Chevreau, C., Bachaud, J.M., Plante, P., Bujan, L., Thonneau, P. (2004). Fertility after testicular cancer treatments: results of a large multicenter study. *Cancer*, *100*, 732–737.

Ichihara, G., Yu, X., Kitoh, J., Asaeda, N., Kumazawa, T., Iwai, H., Shibata, E., Yamada, T., Wang, H., Xie, Z., Maeda, K., Tsukmura, H., Takeuchi, Y. (2000). Reproductive toxicity of 1-bromopropane, a newly introduced alternative to ozone layer depleting solvents, in male rats. *Toxicol. Sci.*, *54*, 416–423.

Ikezuki, Y., Tsutsumi, O., Takai, Y., Kamei, Y., Taketani, Y. (2002). Determination of bisphenol A concentrations in human biological fluids reveals significant early prenatal exposure. *Hum. Reprod.*, *17*, 2839–2841.

International Agency for Research on Cancer (IARC) (1994). 4-Vinylcyclohexene. IARC monographs on the evaluation of carcinogenic risks to humans: some industrial chemicals. Lyon, France. *60*, 347.

Ishikawa, H., Tian, Y., Yamauchi, T. (2001). Induction of micronuclei formation in preimplantation mouse embryos after maternal treatment with 2-bromopropane. *Reprod. Toxicol.*, *15*, 81–85.

James, M.O., Fouts, J.R., Bend, J.R. (1976). Hepatic and extrahepatic metabolism, in vitro, of an epoxide (8-14C-styrene oxide) in the rabbit. *Biochem. Pharmacol.*, *25*, 187–193.

Jarrell, J., Lai, E.V., Barr, R., McMahon, A., Belbeck, L., O'Connell, G. (1987). Ovarian toxicity of cyclophosphamide alone and in combination with ovarian irradiation in the rat. *Cancer Res.*, *47*, 2340–2343.

Jarrell, J.F., Bodo, L., Younglai, E.V., Barr, R.D., O'Connell, G.J. (1991). The short-term reproductive toxicity of cyclophosphamide in the female rat. *Reprod. Toxicol.*, *5*, 481–485.

Jick, H., Porter, J. (1977). Relation between smoking and age of natural menopause. *Lancet*, *1*, 1354–1355.

Jurasovic, J., Cvitkovic, P., Pizent, A., Colak, B., Telisman, S. (2004). Semen quality and reproductive endocrine function with regard to blood cadmium in Croatian male subjects. *Biometals*, *17*, 735–743.

Kabuto, H., Amakawa, M., Shishibori, T. (2004). Exposure to bisphenol A during embryonic/fetal life and infancy increases oxidative injury and causes underdevelopment of the brain and testis in mice. *Life Sci.*, *74*, 2931–2940.

Kang, K.S., Li, G.X., Che, J.H., Lee, Y.S. (2002). Impairment of male rat reproductive function in F1 offspring from dams exposed to 2-bromopropane during gestation and lactation. *Reprod. Toxicol.*, *16*, 151–159.

Kato, H., Furuhashi, T., Tanaka, M., Katsu, Y., Watanabe, H., Ohta, Y., Iguchi, T. (2006). Effects of bisphenol A given neonatally on reproductive function of male rats. *Reprod. Toxicol.*, *22*, 20–29.

Kato, H., Ota, T., Furuhashi, T., Ohta, Y., Iguchi, T. (2003). Changes in reproductive organs of female rats treated with bisphenol A during the neonatal period. *Reprod. Toxicol.*, *17*, 283–288.

Kavlock, R., Boekelheide, K., Chapin, R., Cunningham, M., Faustman, E., Foster, P., Golub, M., Henderson, R., Hinberg, I., Little, R., Seed, J., Shea, K., Tabacova, S., Tyl, R., Williams, P., Zacharewski, T. (2002). NTP center for the evaluation of risks to human reproduction: phthalates expert panel report on the reproductive and developmental toxicity of di(2-ethylhexyl) phthalate. *Reprod. Toxicol.*, *16*, 529–653.

Keller, D.A., Carpenter, S.C., Cagen, S.Z., Reitman, F.A. (1997). In vitro metabolism of 4-vinylcyclohexene in rat and mouse liver, lung, and ovary. *Toxicol. Appl. Pharmacol.*, *144*, 36–44.

Kluwe, W.M., Lamb, J.C., IV, Greenwell, A.E., Harrington, F.W. (1983). 1,2-dibromo-3-chloropropane (DBCP)-induced infertility in male rats mediated by a post-testicular effect. *Toxicol. Appl. Pharmacol.*, *71*, 294–298.

Kopf-Maier, P. (1992). Effects of carboplatin on the testis. A histological study. *Cancer Chemother. Pharmacol.*, *29*, 227–235.

Koyama, H., Wada, T., Nichizawa, Y., Iwanaga, T., Aoki, Y. (1977). Cyclophosphamide-induced ovarian failure and its therapeutic significance in patients with breast cancer. *Cancer*, *39*, 1403–1409.

Krarup, T. (1967). 9,10-Dimethyl-1,2-benzanthracene induced ovarian tumors in mice. *Acta Pathol. Microbiol. Scand.*, *70*, 241–248.

Krarup, T. (1969). Oocyte destruction and ovarian tumorigenesis after direct application of a chemical carcinogen (9,10-dimethyl-benzanthracene) to the mouse ovary. *Int. J. Cancer*, *4*, 61–75.

Krarup, T. (1970). Oocyte survival in the mouse ovary after treatment with 9,10-dimethyl-1,2-benzanthracene. *J. Endocrinol.*, *46*, 483–495.

Lamperti, A.A., Printz, R.H. (1974). Localization, accumulation and toxic effects of mercuric chloride on the reproductive axis of the female hamster. *Biol. Reprod.*, *11*, 180–186.

Laws, S.C., Ferrell, J.M., Stoker, T.E., Cooper, R.L. (2003). Prepubertal development in female Wistar rats following exposure to propazine and atrazine biotransformation by-products, diamino-S-chlorotriazine and hydroxyatrazine. *Toxicol. Sci.*, *76*, 190–200.

Li, X., Johnson, D.C., Rozman, K.K. (1995). Reproductive effects of 2,3,7,8-tetrachlorodibenzo-p-dioxin (TCDD) in female rats: ovulation, hormonal regulation, and possible mechanism(s). *Toxicol. Appl. Pharmacol.*, *133*, 321–327.

Lindström, P., Harris, M., Ross, M., Lamb, J.C., IV, Chapin, R.E. (1988). Comparison of changes in serum androgen binding protein with germinal epithelial damage and infertility induced by di-n-pentyl phthalate. *Fundam. Appl. Toxicol.*, *11*, 528–539.

Lovekamp, T.N., Davis, B.J. (2001). Mono-(2-ethylhexyl) phthalate suppresses aromatase transcript levels and estradiol production in cultured rat granulosa cells. *Toxicol. Appl. Pharmacol.*, *172*, 217–224.

Lovekamp-Swan, T., Davis, B.J. (2003). Mechanisms of phthalate ester toxicity in the female reproductive system. *Environ. Health Perspect.*, *111*, 139–145.

Lovekamp-Swan, T., Jetten, A.M., Davis, B.J. (2003). Dual activation of PPARalpha and PPARgamma by mono-(2-ethylhexyl) phthalate in rat ovarian granulosa cells. *Mol. Cell. Endocrinol.*, *201*, 133.

Lymberopoulos, A.G., Kotsaki-Kovatsi, V.P., Taylor, A., Papaioannou, N., Brikas, P. (2000). Effects of cadmium chloride administration on the macroscopic and microscopic characteristics of ejaculates from Chios ram-lambs. *Theriogenology*, *54*, 1145–1157.

Mably, T.A., Bjerke, D.L., Moore, R.W., Gendron-Fitzpatrick, A., Peterson, R.E. (1992). In utero and lactational exposure of male rats to 2,3,7,8-tetrachlorodibenzo-p-dioxin. 3. Effects on spermatogenesis and reproductive capacity. *Toxicol. Appl. Pharmacol.*, *114* (1), 118–126.

Mably, T.A., Moore, R.W., Peterson, R.E. (1992). In utero and lactational exposure of male rats to 2,3,7,8-tetrachlorodibenzo-p-dioxin. 1. Effects on androgenic status. *Toxicol. Appl. Pharmacol.*, *114* (1), 97–107.

MacKenzie, K.M., Angevine, D.M. (1981). Infertility in mice exposed in utero to benzo[a]pyrene. *Biol. Reprod.*, *24*, 183–191.

Markey, C.M., Rubin, B.S., Soto, A.M., Sonnenschein, C. (2002). Endocrine disruptors: from wingspread to environmental developmental biology. *J. Steroid. Biochem. Mol. Biol.*, *83*, 235–244.

Maronpot, R.R. (1987). Ovarian toxicity and carcinogenicity in eight recent National Toxicological Program studies. *Environ. Health Perspect.*, *73*, 125–130.

Martini, F.H., Bartholomew, E.F. (1997). The reproductive system of the male, in *Essentials of Anatomy and Physiology* (ed. D.K. Brake), Prentice Hall, New Jersey, pp. 516–523.

Mattison, D.R. (1980). Morphology of oocyte and follicle destruction by polycyclic aromatic hydrocarbons in mice. *Toxicol. Appl. Pharmacol.*, *53*, 249–259.

Mattison, D.R. (1983). The mechanisms of action of reproductive toxins. *Am. J. Ind. Med.*, *4*, 65–79.

Mattison, D.R. (1985). Clinical manifestation of ovarian toxicity, in *Reproductive Toxicology* (ed. R.L. Dixon), Raven, New York, pp. 109–130.

Mattison, D.R., Nightingale, M.R. (1980). The biochemical and genetic characteristics of murine aryl ovarian hydrocarbon (benzo[a]pyrene) hydroxylase activity and its relationship to primordial oocyte destruction by polycyclic aromatic hydrocarbons. *Toxicol. Appl. Pharmacol.*, *56*, 399–408.

Mattison, D.R., Schulman, J.D. (1980). How xenobiotic chemicals can destroy oocytes. *Contemp. Obstet. Gynecol.*, *15*, 157.

Mattison, D.R., Thorgeirsson, S.S. (1977). Genetic differences in mouse ovarian metabolism of benzo[a]pyrene and oocyte toxicity. *Biochem. Pharmacol.*, *26*, 909–912.

Mattison, D.R., Thorgeirsson, S.S. (1978). Gonadal aryl hydrocarbon hydroxylase in rats and mice. *Cancer Res.*, *38*, 1368–1373.

Mattison, D.R., Thorgeirsson, S.S. (1979). Ovarian aryl hydrocarbon hydroxylase activity and primordial oocyte toxicity of polycyclic aromatic hydrocarbons in mice. *Cancer Res.*, *39*, 3471–3475.

Mattison, D.R., Plowchalk, B.S., Meadows, M.J., Miller, M.M., Malek, A., London, S. (1989). The effect of smoking on oogenesis, fertilization, and implantation. *Semin. Reprod. Endocrinol.*, *7*, 291.

Mayer, L.P., Pearsall, N.A., Christian, P.J., Devine, P.J., Payne, C.M., McCuskey, M.K., Marion, S.L., Sipes, I.G., Hoyer, P.B. (2002). Long-term effects of ovarian follicular depletion in rats by 4-vinylcyclohexene diepoxide. *Reprod. Toxicol.*, *16*, 775–781.

Meistrich, M.L., Finch, M., da Cunha, M.F., Hacker, U., Au, W.W. (1982). Damaging effects of fourteen chemotherapeutic drugs on mouse testis cells. *Cancer Res.*, *42*, 122–131.

Melnick, R.L., Huff, J., Chou, B.J., Miller, R.A. (1990). Carcinogenicity of 1,3-butadiene in C57BL/6 X C3HF1 mice at low exposure concentrations. *Cancer Res.*, *50*, 6592–6599.

Miller, J.J., Cole, L.J. (1970). Changes in mouse ovaries after prolonged treatment with cyclophosphamide. *Proc. Soc. Exp. Biol. Med.*, *133*, 190–193.

Miller, J.J., Williams, G.F., Leissring, J.C. (1971). Multiple late complications of therapy with cyclophosphamide, including ovarian destruction. *Am. J. Med.*, *50*, 530–535.

Miller, K.P., Gupta, R.K., Flaws, J.A. (2006). Methoxychlor metabolites may cause ovarian toxicity through estrogen-regulated pathways. *Toxicol. Sci.*, *93*, 180–188.

Miller, K.P., Gupta, R.K., Greenfeld, C.R., Babus, J.K., Flaws, J.A. (2005). Methoxychlor directly affects ovarian antral follicle growth and atresia through Bcl-2- and Bax-mediated pathways. *Toxicol. Sci.*, *88*, 213–221.

Mlynarcikova, A., Fickova, M., Scsukova, S. (2005). Ovarian intrafollicular processes as a target for cigarette smoke components and selected environmental reproductive disruptors. *Endocr. Regul.*, *39*, 20–31.

Moran, F.M., Enan, E., Vandevoort, C.A., Stewart, D.R., Conley, A.J., Overstreet, J.W., Lasley, B.L. (1997). 2,3,7,8-Tetrachlorodibenzo-p-dioxin (TCDD) effects on steroidogenesis of human luteinized granulosa cells in vitro. *Biol. Reprod.* Suppl., *56*, 65.

Morrissey, R.E., Schwetz, B.A., Hackett, P.L., Sikov, M.R., Hardin, B.D., McClanahan, B.J., Decker, J.R., Mast, T.J. (1990). Overview of reproductive and developmental toxicity studies of 1,3-butadiene in rodents. *Environ. Health Perspect.*, *86*, 79–84.

Mukhtar, H., Philpot, R.M., Bend, J.R. (1978). The postnatal development of microsomal epoxide hydrase, cytosolic glutathione S-transferase, and mitochondrial and microsomal cytochrome P-450 in adrenals and ovaries of female rats. *Drug Metab. Dispos.*, *6*, 577–583.

Murono, E.P., Derk, R.C. (2005). The reported active metabolite of methoxychlor, 2,2-bis(p-hydroxyphenyl)-1,1,1-trichloroethane, inhibits testosterone formation by cultured Leydig cells from neonatal rats. *Reprod. Toxicol.*, *20*, 503–513.

Murono, E.P., Derk, R.C., Akgul, Y. (2006). In vivo exposure of young adult male rats to methoxychlor reduces serum testosterone levels and ex vivo Leydig cell testosterone formation and cholesterol side-chain cleavage activity. *Reprod. Toxicol.*, *21*, 148–153.

Murray, R.D.H., Mendez, J., Brown, S.A. (1982). *The Natural Coumarins*, John Wiley & Sons, Ltd, Chichester.

Musajo, L., Rodighiero, G. (1962). The skin-photosensitizing furanocoumarins. *Experientia*, *18*, 153–200.

Mylchreest, E., Wallace, D.G., Cattley, R.C., Foster, P.M. (2000). Dose-dependent alterations in androgen-regulated male reproductive development in rats exposed to di(n-butyl) phthalate during late gestation. *Toxicol. Sci.*, *55*, 143–151.

Naha, N., Chowdhury, A.R. (2006). Inorganic lead exposure in battery and paint factory: effect on human sperm structure and functional activity. *J. UOEH*, *28* (2), 157–171.

Nath, R., Prasad, R., Palinal, V.K., Chopra, R.K. (1984). Molecular basis of cadmium toxicity. *Prog. Food Nutr. Sci.*, *8*, 109–163.

National Toxicology Program (1986). Toxicology and carcinogenesis studies of 4-vinylcyclohexene in F344/N rats and B6C3F1 mice. NTP Tech. Rep. No. 303. Research Triangle Park, NC., US Dep. Health Hum. Serv. Public Health Serv. Natl. Inst. Health. Public Inf. Natl. Toxicol. Program.

National Toxicology Program (1989). Toxicology and carcinogenesis studies of 4-vinylcyclohexene diepoxide in F344/N rats and B6C3F1 mice. NTP technical report: 362. Research Triangle Park, NC., US Dep. Health Hum. Serv. Public Health Serv. Natl. Inst. Health. Public Inf. Natl. Toxicol. Program.

Neal, M.S., Zhu, J., Holloway, A.C., Foster, W.G. (2007). Follicle growth is inhibited by benzo-[a]-pyrene at concentrations representative of human exposure, in an isolated rat follicle culture assay. *Hum. Reprod.*, *22*, 961–967.

Omura, M., Romero, Y., Zhao, M., Inoue, N. (1999). Histopathological evidence that spermatogonia are the target cells of 2-bromopropane. *Toxicol. Lett.*, *104*, 19–26.

Orisakwe, O.E., Afonne, O.J., Nwobodo, E., Asomugha, L., Dioka, C.E. (2001). Low-dose mercury induces testicular damage protected by zinc in mice. *Eur. J. Obstet. Gynecol. Reprod. Biol.*, *95*, 92–96.

Pant, N., Upadhyay, G., Pandey, S., Mathur, N., Saxena, D.K., Srivastava, S.P. (2003). Lead and cadmium concentration in the seminal plasma of men in the general population: correlation with semen quality. *Reprod. Toxicol.*, *17*, 447–450.

Papaconstantinou, A.D., Umbreit, T.H., Fisher, B.R., Goering, P.L., Lappas, N.T., Brown, K.M. (2000). Bisphenol A-induced increase in uterine weight and alterations in uterine morphology in ovariectomized B6C3F1 mice: role of the estrogen receptor. *Toxicol. Sci.*, *56*, 332–339.

Park, J., Kim, Y., Park, D., Choi, K., Park, S., Moon, Y. (1997). An outbreak of hemapoietic and reproductive disorders due to solvents containing 2-bromopropane in an electronic factory, South Korea: epidemiological study. *J. Occup. Health*, *39*, 138–143.

Pathak, M.A., Fitzpatrick, T.B. (1992). The evolution of photochemotherapy with psoralens and UVA (PUVA): 200 BC to 1992 AD. *J. Photochem. Photobiol. B*, *14*, 3–33.

Petersen, P.M., Hansen, S.W., Giwercman, A., Rørth, M., Skakkebaek, N.E. (1994). Dose-dependent impairment of testicular function in patients treated with cisplatin-based chemotherapy for germ cell cancer. *Ann. Oncol.*, *5*, 355–358.

Peterson, R.E., Cooke, P.S., Klece, W.R., Gray, L.E. (1997). Environmental endocrine disruptors, in *Comprehensive Toxicology*, Vol. 10 (eds I.G. Sipes, C.A. McQueen and A.G. Gandolfi), Elsevier Science, New York, pp. 181–191.

Piasek, M., Blanusa, M., Kostial, K., Laskey, J.W. (2001). Placental cadmium and progesterone concentrations in cigarette smokers. *Reprod. Toxicol.*, *15*, 673–681.

Piasek, M., Laskey, J.W. (1994). Acute cadmium exposure and ovarian steroidogenesis in cycling and pregnant rats. *Reprod. Toxicol.*, *8*, 495–507.

Piasek, M., Laskey, J.W. (1999). Effects of in vitro cadmium exposure on ovarian steroidogenesis in rats. *J. Appl. Toxicol.*, *19*, 211–217.

Piasek, M., Laskey, J.W., Kostial, K., Blanusa, M. (2002). Assessment of steroid disruption using cultures of whole ovary and/or placenta in rat and in human placental tissue. *Int. Arch. Occup. Environ. Health*, *75*, S36–44.

Piasek, M., Schonwald, N., Blanusa, M., Kostial, K., Laskey, J.W. (1996). Biomarkers of heavy metal reproductive effects and interaction with essential elements in experimental studies on female rats. *Arh. Hig. Rada Toksikol.*, *47*, 245–259.

Plowchalk, D.R., Mattison, D.R. (1992). Reproductive toxicity of cyclophosphamide in the C57GBL/6N mouse. 1. Effects on ovarian structure and function. *Reprod. Toxicol.*, *6*, 411–421.

Rajapaksa, K.S., Sipes, I.G., Hoyer, P.B. (2007). Involvement of microsomal epoxide hydrolase in ovotoxicity caused by 7,12-dimethylbenz[a]anthracene. *Toxicol. Sci.*, *96*, 327–334.

Rayner, J.L., Enoch, R.R., Wolf, D.C., Fenton, S.E. (2007). Atrazine-induced reproductive tract alterations after transplacental and/or lactational exposure in male Long-Evans rats. *Toxicol. Appl. Pharmacol.*, *218*, 238–248.

Revel, A., Laufer, N. (2002). Protecting female fertility from cancer therapy. *Mol. Cell. Endocrinol.*, *187*, 83–91.

Rhoads, R., Pflanzer, R. (1996). Reproductive physiology, in *Human Physiology* (ed. J.L. Alexander), Saunders College Publishing, Orlando, FL, pp. 900–950.

Richards, J.S. (1980). Maturation of ovarian follicles: actions and interactions of pituitary and ovarian hormones on follicular cell differentiation. *Physiol. Rev.*, *60*, 51–89.

Ronis, M.J., Badger, T.M., Shema, S.J., Roberson, P.K., Shaikh, F. (1996). Reproductive toxicity and growth effects in rats exposed to lead at different periods during development. *Toxicol. Appl. Pharmacol.*, *136*, 361–371.

Ronnback, C. (1991). Effect of 3,3′,4,4′-tetrachlorobiphenyl (TCB) on ovaries of foetal mice. *Pharm. Toxicol.*, *69*, 340.

Safe, S., Krishnan, V. (1995). Chlorinated hydrocarbons: estrogens and antiestrogens. *Toxicol. Lett.*, *82*, 731–736.

Schrader, S.M., Kanitz, M.H. (1994). Occupational hazards to male reproduction, in *Occupational Medicine: State of the Art Reviews: Reproductive Hazards*, Vol. 9 (eds E. Gold, M. Schenker and B. Lasley), Hanley & Bilfus, Philadelphia, PA, pp. 405–414.

Schuurs, A.H. (1999). Reproductive toxicity of occupational mercury. A review of the literature. *J. Dent.*, *27*, 249–256.

Scott, B.R., Pathak, M.A., Mohm, G.R. (1976). Molecular and genetic basis of furanocoumarin reactions. *Mutat. Res.*, *39*, 29–74.

Seegmiller, R.E. (1997). Selected examples of developmental toxicants, in *Reproductive and Endocrine Toxicology*, Vol. 10 (eds K. Boekelheide, R.E. Chapin, P.B. Hoyer and C. Harris), Elsevier Science, New York, Chapter 45.

Sekiguchi, S., Asano, G., Suda, M., Honma, T. (2001). Influence of 2-bromopropane on reproductive system-short-term administration of 2-bromopropane inhibits ovulation in F344 rats. *Toxicol. Ind. Health*, *16*, 277–283.

Sherwood, L. (1997). The reproductive system, in *Human Physiology: From Cells to Systems* (ed. P. Lewis), Wadsworth Publishing Co., pp. 700–750.

Shiromizu, K., Mattison, D.R. (1984). The effect of intraovarian injection of benzo(a)pyrene on primordial oocyte number and ovarian aryl hydrocarbon [benzo(a)pyrene] hydroxylase activity. *Toxicol. Appl. Pharmacol.*, *76*, 18–25.

Smith, B.J., Mattison, D.R., Sipes, I.G. (1990). The role of epoxidation in 4-vinylcyclohexene-induced ovarian toxicity. *Toxicol. Appl. Pharmacol.*, *105*, 372–381.

Smith, T.L., Merry, S.T., Harris, D.L., Joe Ford, J., Ike, J., Archibong, A.E., Ramesh, A. (2007). Species-specific testicular and hepatic microsomal metabolism of benzo(a)pyrene, an ubiquitous toxicant and endocrine disruptor. *Toxicol. In Vitro*, *21*, 753–738.

Smith, W., Mitchell, P., Wang, J.J. (1997). Gender, oestrogen, hormone-replacement and age-related macular degeneration: results from the blue mountains eye study. *Aust. N. Z. J. Ophthalmol.*, *25*, S13–S15.

Sobrinho, L.G., Levine, R.A., DeConti, R.C. (1971). Amenhorrhea in patients with Hodgkin's disease treated with antineoplastic agents. *Am. J. Obstet. Gynecol.*, *109*, 135–139.

Son, H.Y., Kim, Y.B., Kang, B.H., Cho, S.W., Ha, C.S., Roh, J.K. (1999). Effects of 2-bromopropane on spermatogenesis in the Sprague-Dawley rat. *Reprod. Toxicol.*, *13* (3), 179–187.

Sowers, M.R., LaPietra, M.T. (1995). Menopause: its epidemiology and potential association with chronic diseases. *Epidemiol. Rev.*, *17*, 287–302.

Spano, M., Bartoleschi, C., Cordelli, E., Leter, G., Segre, L., Mantovani, A., Fazzi, P., Pacchierotti, F. (1996). Flow cytometric and histological assessment of 1,2:3,4-diepoxybutane toxicity on mouse spermatogenesis. *J. Toxicol. Environ. Health*, *47*, 423–441.

Springer, L.N., McAsey, M.E., Flaws, J.A., Tilly, J.L., Sipes, I.G., Hoyer, P.B. (1996). Involvement of apoptosis in 4-vinylcyclohexene diepoxide-induced ovotoxicity in rats. *Toxicol. Appl. Pharmacol.*, *139*, 394–401.

Stedman, R.L. (1968). The chemical composition of tobacco and tobacco smoke. *Chem. Rev.*, *68*, 153–207.

Stern, R.S., Thibodeau, L.A., Kleinerman, R.A., Parrish, J.A., Fitzpatrick, T.B. (1979). Risk of cutaneous carcinoma in patients treated with oral methoxsalen photochemotherapy for psoriasis. *N. Engl. J. Med.*, *300*, 809–813.

Stoker, T.E., Guidici, D.L., Laws, S.C., Cooper, R.L. (2002). The effects of atrazine metabolites on puberty and thyroid function in the male Wistar rat. *Toxicol. Sci.*, *67*, 198–206.

Sundaram, K., Witorsch, R.J. (1995). Toxic effects on the testes, in *Reproductive Toxicology* (ed. R.J. Witorsch), Raven Press, New York, pp. 99–122.

Susiarjo, M., Hassold, T.J., Freeman, E., Hunt, P.A. (2007). Bisphenol A exposure in utero disrupts early oogenesis in the mouse. *PLoS Genet.*, *3* (1), 63–70.

Tabacova, S., Little, R., Balabaeva, L. (1999). Maternal exposure to phthalates and complications of pregnancy. *Epidemiology*, *10*, S127.

Taketoh, J., Mutoh, J., Takeda, T., Ogishima, T., Takeda, S., Ishii, Y., Ishida, T., Yamada, H. (2007). Suppression of fetal testicular cytochrome P450 17 by maternal exposure to 2,3,7,8-tetrachlorodibenzo-p-dioxin: a mechanism involving an initial effect on gonadotropin synthesis in the pituitary. *Life Sci.*, *80*, 1259–1267.

Taupeau, C., Poupon, J., Nome, F., Lefevre, B. (2001). Lead accumulation in the mouse ovary after treatment-induced follicular atresia. *Reprod. Toxicol.*, *15*, 385–391.

Taupeau, C., Poupon, J., Treton, D., Brosse, A., Richard, Y., Machelon, V. (2003). Lead reduces messenger RNA and protein levels of cytochrome P450 aromatase and estrogen receptor beta in human ovarian granulosa cells. *Biol. Reprod.*, *68*, 1982–1988.

Telisman, S., Colak, B., Pizent, A., Jurasovic, J., Cvitkovic, P. (2007). Reproductive toxicity of low-level exposure in men. *Environ. Res.*, *105*, 256–266.

Tennent, B.J., Beamer, W.G. (1986). Ovarian tumors not induced by irradiation and gonadotropins in hypogonadal (hpg) mice. *Biol. Reprod.*, *34*, 751–769.

Thomas, C., Cans, C., Pelletier, R., De Robertis, C., Hazzouri, M., Sele, B., Rousseaux, S., Hennebicq, S. (2004). No long-term increase in sperm aneuploidy rates after anticancer therapy: sperm fluorescence in situ hybridization analysis in 26 patients treated for testicular cancer or lymphoma. *Clin. Cancer Res.*, *10*, 6535–6543.

Thomas, J.A. (1993). Toxic responses of the reproductive system, in *Toxicology* (eds M.O. Andue, J. Doull and C.D. Klassen), McGraw Hill, New York, pp. 484–520.

Thomas, P. (1999). Nontraditional sites of endocrine disruption by chemicals on the hypothalamus-pituitary-gonadal axis: interactions with steroid membrane receptors. monoaminergic pathways, and signal transduction systems, in *Endocrine Disruptors: Effects on Male and Female Reproductive Systems* (ed. R.K. Naz), CRC Press, Boca Raton, FL, pp. 3–38.

Thoreux-Manlay, A., Le Goascogne, C., Segretain, D., Jegou, B., Pinon-Lataillade, G. (1995). Lead affects steroidogenesis in rat Leydig cells in vivo and in vitro. *Toxicology*, *103*, 53–62.

Tilly, J.L., Kowalski, K.I., Johnson, A.L., Hsueh, A.J. (1991). Involvement of apoptosis in ovarian follicular atresia and postovulatory regression. *Endocrinology*, *129*, 2799–2801.

Tomic, D., Frech, M.S., Babus, J.K., Gupta, R.K., Furth, P.A., Koos, R.D., Flaws, J.A. (2006). Methoxychlor induces atresia of antral follicles in ER alpha-overexpressing mice. *Toxicol. Sci.*, *93*, 196–204.

Trumble, J.T., Moar, W.J., Brewer, M.J., Carson, W.G. (1991). Impact of UV radiation on activity of linear furanocoumarins and Bacillus thuringiensis var kurstaki against Spodoptera exigua: implications for tritrophic interactions. *J. Chem. Ecol.*, *17*, 973–987.

Tullner, W.W., Edgcomb, J.H. (1962). Cystic tubular nephropathy and decrease in testicular weight in rats following oral methoxychlor treatment. *J. Pharmacol. Exp. Ther.*, *138*, 126–130.

Vahakangas, K., Rajaniemi, H., Pelkonen, O. (1985). Ovarian toxicity of cigarette smoke during pregnancy in mice. *Toxicol. Lett.*, *25*, 75–80.

van Ravenzwaay, B. (1992). Discussion of prenatal and reproductive toxicity to Reg. No. 83-258 (Vinclozolin). Data submission to US Environmental Protection Agency from BASF, MRIA 425813-02.

van Scott, E.J. (1976). Therapy of psoriasis 1975. *JAMA*, *235*, 197–198.

VanVoorhis, B.J., Syrop, C.H., Hammitt, D.G., Dunn, M.S., Snyder, G.D. (1992). Effects of smoking on ovulation induction for assisted reproduction techniques. *Fertil. Steril.*, *58*, 981–985.

Waring, S.C., Rocca, W.A., Peterson, R.C., O'Brien, P.C., Tangalos, E.G., Kokmen, E. (1999). Postmenopausal estrogen replacement therapy and risk of AD: a population-based study. *Neurology*, *52*, 965–970.

Warne, G.L., Fairley, K.F., Hobbs, J.B., Martin, F.I. (1973). Cyclophosphamide-induced ovarian failure. *N. Engl. J. Med.*, *289*, 1159–1162.

Weber, J.E., Turner, T.T., Tung, K.S., Russell, L.D. (1988). Effects of cytochalasin D on the integrity of the Sertoli cell (blood-testis) barrier. *Am. J. Anat.*, *182*, 130–147.

Weinberg, C.R., Wilcox, A.J., Baird, D.D. (1989). Reduced fecundibility in women with prenatal exposure to cigarette smoking. *Am. J. Epidemiol.*, *129*, 1072–1078.

Wetzel, L.T., Luempert, L.G., III, Breckenridge, C.B., Tisdel, M.O., Stevens, J.T., Thakur, A.K., Extrom, P.J., Eldridge, J.C. (1994). Chronic effects of atrazine on estrus and mammary tumor formation in female Sprague-Dawley and Fischer 344 rats. *J. Toxicol. Environ. Health*, *43*, 169–182.

Whelan, E.A. (1997). Risk assessment studies: epidemiology, in *Comprehensive Toxicology*, Vol. 10 (I.G. Sipes, C.A. McQueen and A.J. Gandolfi), Elsevier Science, New York, pp. 359–366.

Whorton, D., Krauss, R.M., Marshall, S., Milby, T.H. (1977). Infertility in male pesticide workers. *Lancet*, *2*, 1259–1261.

Wiebe, J.P., Barr, K.J., Buckingham, K.D. (1988). Effect of prenatal and neonatal exposure to lead on gonadotropin receptors and steroidogenesis in rat ovaries. *J. Toxicol. Environ. Health*, *24*, 461–476.

Xu, L.C., Wang, S.Y., Yang, X.F., Wang, X.R. (2001). Effects of cadmium on rat sperm motility evaluated with computer assisted sperm analysis. *Biomed. Environ. Sci.*, *14*, 312–317.

Yamada, T., Ichihara, G., Wang, H., Yu, X., Maeda, K., Tsukamura, H., Kamaijima, M., Nakajima, T., Takeuchi, Y. (2003). Exposure to 1-bromopropane causes ovarian dysfunction in rats. *Toxicol. Sci.*, *71*, 96–103.

Yeh, J., Barbieri, R.L. (1989). Twenty-four-hour urinary-free cortisol in premenopausal cigarette smokers and nonsmokers. *Fertil. Steril.*, *52*, 1067–1069.

Younglai, E.V., Foster, W.G., Hughes, E.G., Trim, K., Jarrell, J.F. (2002). Levels of environmental contaminants in human follicular fluid, serum and seminal plasma of couples undergoing in vitro fertilization. *Arch. Environ. Contam. Toxicol.*, *43*, 121–126.

Yu, X., Kamijima, M., Ichihara, G., Li, W., Kitoh, J., Xie, Z., Shibata, E., Hisanaga, N., Takeuchi, Y. (1999). 2-Bromopropane causes ovarian dysfunction by damaging primordial follicles and their oocytes in female rats. *Toxicol. Appl. Pharmacol.*, *159*, 185–193.

Zachow, R., Uzumcu, M. (2006). The methoxychlor metabolite, 2,2-bis-(p-hydroxyphenyl)-1,1,1-trichloroethane, inhibits steroidogenesis in rat ovarian granulosa cells in vitro. *Reprod. Toxicol.*, *22*, 659–665.

Zenzes, M.T., Krishnan, S., Krishnan, B., Zhang, H., Casper, R.F. (1995). Cadmium accumulation in follicular fluid of women in in vitro fertilization-embryo transfer is higher in smokers. *Fertil. Steril.*, *64*, 599–603.

# An Introduction to Toxicogenomics

AARON L. VOLLRATH and CHRISTOPHER A. BRADFIELD

## 1 INTRODUCTION

Toxicogenomics can be viewed as the assessment of chemically induced biological change through a merger of classical toxicology, high-throughput genomics, and bioinformatics. The growth of toxicogenomics is largely due to the complete sequencing of a growing number of genomes, coupled with the development of novel technologies that allow us to detect chemically induced changes in global gene transcription. The product of toxicogenomics is often a high-density data set that expands our understanding of chemical–biological interactions. Through toxicogenomics approaches, the ability to predict the toxicities of a chemical compound and classify the hazard of a large number of chemicals is now possible. This ability is derived from the idea that chemicals with similar toxicological mechanisms will be revealed by their common influence on subsets of the global genomic response. Moreover, the high-throughput nature of toxicogenomics results in a shorter time for chemical testing which in turn allows for a more timely understanding of toxicities that arise from exposure to chemicals in commerce and those from our natural environment. This emerging scientific discipline is poised to have a significant impact on a number of important toxicological questions.

## 2 MICROARRAYS: GENOMICS-BASED METHODS FOR TRANSCRIPTIONAL PROFILING

The methodological domain of toxicogenomics is quite broad and growing rapidly. This basic introduction to the field will focus primarily on the use of microarrays to assess transcriptional events associated with exposure to a toxicant of interest. This is a method commonly known as transcriptional profiling or toxicant profiling. Microarray-based methods to detect whole-genome transcriptional changes or transcriptional profiling provide a means to identify how a chemical affects cells at the level of gene transcription. In this review, we define a transcriptional profile as the set of genes represented on the microarray platform and their associated measured relative

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

transcription values as compared with a control state. Thus, a transcriptional profile can be thought of as a vector of values in which the number of values is dependent on the number of individual probes on the array. In the case where a transcriptional profile is generated on an array of all known open reading frames, this transcriptional profile would represent a global transcriptional profile. While microarray-based methods are not the only means to assess changes in transcriptional events, they are currently one of the more widely used methods.

## 3 CHEMICAL CLASSIFICATION—IDENTIFYING CHEMICALLY INDUCED SIGNATURES AND THEIR POTENTIAL USE IN RISK ASSESSMENT

One of the applications of toxicogenomics is to classify hazardous chemicals based upon their transcriptional profiles. Once profiling is performed on a large number of potentially toxic chemicals, a chemical class can be defined as a set of chemicals that induce transcriptional responses that are highly similar. It follows that once classes of known chemicals are established by the identification of unique subsets of genes, or chemical signatures, derived from transcriptional profiles, chemicals novel to commerce or understudied natural products may then be classified based on their comparison with the set of established classes (Hayes and Bradfield, 2005). If an unknown chemical is then associated with a known class of chemicals based on the utilization of chemical signatures, predictions can be made regarding the effects of the exposure and, in the case of well-characterized classes of chemicals, mechanisms of action can be predicted.

Recently, *The Committee on Applications of Toxicogenomic Technologies to Predict Toxicology and Risk Assessment* from the National Research Council released a report suggesting toxicogenomic methods be integrated by regulatory agencies into the risk assessment process (Committee on Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment, 2007). Prior to the release of this report, a number of studies in the field of toxicogenomics have been conducted based on this premise with promising results providing proof of principle (Thomas *et al.*, 2001b; Bushel *et al.*, 2002; Hamadeh *et al.*, 2002a, b, 2004; Steiner *et al.*, 2004) In fact, some studies report that their use of chemically induced signatures has a greater degree of predictive accuracy relative to current short-term pathology observations or genomic biomarkers. Additionally, the identified chemically induced signatures are superior to estimates based on longer-term sub-chronic studies (Elcombe *et al.*, 2002; Allen *et al.*, 2004; Heinloth *et al.*, 2004; Fielden *et al.*, 2007).

In theory, the identification of chemical signatures and the feasibility of their subsequent use in predicting or classifying the effects of a chemical exposure and risk assessment analyses could be put into practice in the coming years. However, the Committee asserts that "Fully integrating toxicogenomic technologies into predictive toxicology will require a coordinated effort approaching the scale of the Human Genome Project." Given that assessment, there are some limiting factors that prevent the use of transcriptional profiles and associated chemical signatures in these capacities. One limiting factor is the lack of a database containing a large number of transcriptional profiles of chemicals and associated chemical classes from which the chemical signatures are derived. This is largely due to a lack of a well-organized and executed plan to generate the necessary data. The lack of a large-scale approach is due to a number of unresolved issues. Examples of these issues include a lack of consensus on the

utilization of a model system (e.g. organisms such as the mouse and rat), tissues/organs to analyze, and types of experiments to perform. The question of what array platform to utilize is also a prominent issue.

We have organized the rest of this review into the following sections. First, we provide a basic description of microarray technologies consisting of the two most popular approaches to estimate relative transcriptional changes: one-color and two-color arrays. Second, statistical issues associated with microarrays will be presented, specifically identifying and establishing the level of variation within a microarray experiment. Third, we present an established standard protocol used to describe microarray experiments and address the important decisions that influence the choice of microarray-based approaches. Specifically, regarding the latter, we present some basic experimental design methods utilized to obtain the relative changes in transcription. Fourth, we address the types of questions that can be answered utilizing microarrays and the issues related to interpretation and representation of large data sets. In other words, what bioinformatics methods can be utilized to answer questions regarding microarray data? Finally, we will address how transcriptional events elucidated by microarray technologies can be utilized to gain insight into establishing a mechanism of action.

### 3.1 Basics of Microarray Gene Expression Analyses

Currently, it is now feasible to simultaneously monitor gene expression levels on a scale encompassing all the known gene products of toxicological importance of organisms including *Saccharomyces cerevisiae*, *Mus musculus*, *Rattus norvegicus*, *Canis familiari*, and *Homo sapiens*. There are two popular microarray-based approaches, one-color and two-color, utilized to measure gene expression. In both methods, array platforms have been developed and utilized to assay gene expression in this wide range of organisms. Conceptually, both approaches are similar as each requires a step to incorporate a measurable fluorescent label and these methods can be utilized to quantify changes in gene expression in a relative manner.

In one popular one-color-based approach used by Affymetrix (Fig. 1a), total RNA is first isolated and then reverse-transcribed into double-stranded cDNA. A subsequent *in vitro* transcription step is performed and a biotin tag is attached to the antisense cRNAs. After incorporation of the biotin tag, the labeled cRNAs are fragmented into 25- to 200-bp pieces and dispensed on an array so that they can hybridize to their complementary probes on the gene chip. After hybridization, a staining process (i.e. fluorescent labeling) occurs whereby streptavidin-phycoerythrin biotinylated anti-streptavidin antibodies bind to the biotin-tagged cRNAs hybridized to the chip. A laser excites the phycoerythrin bound to the cRNA at each spot producing a signal. The data are derived from a scanner that measures the corresponding fluorescent signal at each spot or DNA probe location on the array. The relative transcriptional differences between the treated and control samples can be determined by taking the ratio of the signals at corresponding spots on each array.

In the two-color approach, two distinct fluorescent dyes, commonly Cy3 (fluorescein/Cyanine 3) and Cy5 (rhodamine/Cyanine 5) are used (Fig. 1b), such that the treated and control samples are labeled with different dyes (e.g. treated sample labeled with Cy3 and a reference sample/control labeled with Cy5). The labeled cDNA or cRNA samples are mixed, and both are allowed to competitively hybridize to DNA probes on a single array. To capture data, this approach requires that every spot on the array

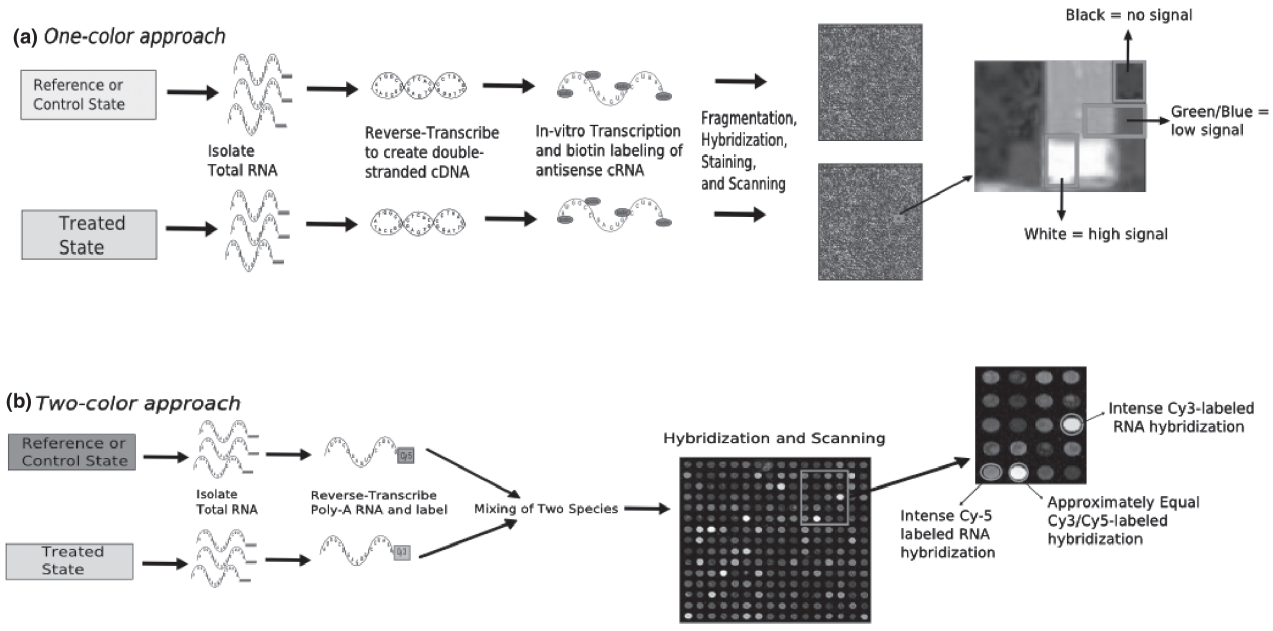


**Figure 1** **(a)** A flowchart depicting the one-color approach used by Affymetrix (*Array images courtesy of Affymetrix*). In this approach, when comparing gene expression between two samples, control and treated, a separate array is required for each sample. For both, total RNA is isolated and reverse-transcribed to create double-stranded cDNA. The double-stranded cDNA serves as a template for the *in vitro* transcription of antisense cRNA. A biotin label is added at this step to facilitate the detection of those cRNAs that bind to complementary probes on the array. After the hybridization step, the array is washed to remove unbound cRNA, stained with streptavidin phycoerythrin conjugate and subsequently scanned with a laser emitting light at 570 nm. The strength of signal is measured by the amount of light emitted. When looking at the array, white, red, and yellow indicate a strong signal at a probe location while green, light blue, and dark blue indicate a weak signal. Black represents no signal at that probe location. These signals are quantified using specialized software to determine the level of expression. Different samples can be compared by taking the ratio of the signals at corresponding probes on each array. **(b)** A flowchart depicting the two-color approach. To obtain the relative expression ratios between an RNA sample from a control/reference state and an RNA sample from a treated state, RNA is first isolated and then the poly-A RNA is reverse-transcribed into cRNA or cDNA while being labeled with a fluor. In this case, the control sample is labeled with Cy5 and the treated sample is labeled with Cy3. After labeling, the labeled species are mixed together and subsequently hybridized to a microarray. During hybridization, the two labeled species bind to the probes and relative differences can be detected by scanning the array with a laser at two different wavelengths. One wavelength excites the Cy5 fluor and the other excites the Cy3 fluor, resulting in the emission of light which is quantified for each by the use of specialized software.

be excited with a laser at two distinct frequencies to generate a signal based on the amount of Cy3- and Cy5-labeled targets that hybridized. Relative differences in transcription can then be calculated based on the raw fluorescence values obtained for each laser-induced signal. This value is usually represented as a log ratio or as a fold change. As an example, assuming no dye bias, if the Cy3 channel was measured at 500 fluorescent units and the Cy5 channel was measure at 1000 fluorescent units, the ratio (Cy5/Cy3) would be 1000/500, representing a twofold change.

### 3.2 Statistical Issues

Variation can affect the results of a microarray experiment. There are two types of variation: biological and technical. Biological variation is variation that mainly occurs prior to RNA extraction. It can be thought of as variation arising because two samples are biologically different. Technical variation occurs as the result of differences that arise during amplification, labeling, and hybridization of the source RNA. Additionally, technical variation is introduced during the measurement of the signals produced from the scanning of the array and in data analysis. Each step in the microarray experiment is a potential source of variation in the final value used to represent the relative difference in transcription (Churchill, 2002).

Biological sources of variation include environment and genetic factors that influence the levels of a transcript and its response to chemical treatment (Oleksiak *et al.*, 2002; Spruill *et al.*, 2002). Even animals on the same genetic background raised in the same environment will exhibit some variation in transcriptional response to treatment with a chemical. Any attempt to assess relative transcriptional differences must take into consideration the underlying biological variation within the experiment being conducted. These estimations are accomplished through the utilization of biological replicates. Examples of such replicates include multiple independent samples for control and treated sources of RNA obtained under the same experimental condition (Fig. 2). A large number of biological replicate comparisons are the ideal, but in cases when the RNA source or cost of microarray analysis is the limiting experimental factor, it is common practice to pool RNA from replicates. However, pooling comes at the cost of reducing the number of independent biological replicates. When large numbers of individual, unique samples obtained under the same experimental conditions are analyzed, it is easier to detect anomalous samples and obtain a better understanding of the inherent biological variation. An example of an anomalous sample would be when one of five samples came from an animal model that has experienced an inflammatory response to a pathogen or wound specific to the individual in question and not the experiment. A discussion of the determination of the required number of biological replicates that should be utilized to obtain statistically significant results is beyond the scope of this work, but has been discussed elsewhere (Churchill, 2002).

The second source of variation, technical variation, is introduced when the RNA is isolated, processed (i.e. reverse-transcribed and fluorescently labeled), and hybridized to the microarray. Utilizing technical replicates allows for the estimation of the amount of technical variation present within the experiment (i.e. running the same sample multiple times). Technical replicates can be RNA from the same source (and control) hybridized multiple times on unique microarrays from the same array platform (i.e. two or more different arrays that have the exact probe set). Ideally, the measurements from the same samples should be able to be replicated with high fidelity from slide to slide if the same protocols and practices are utilized (Fig. 3). In our own experience

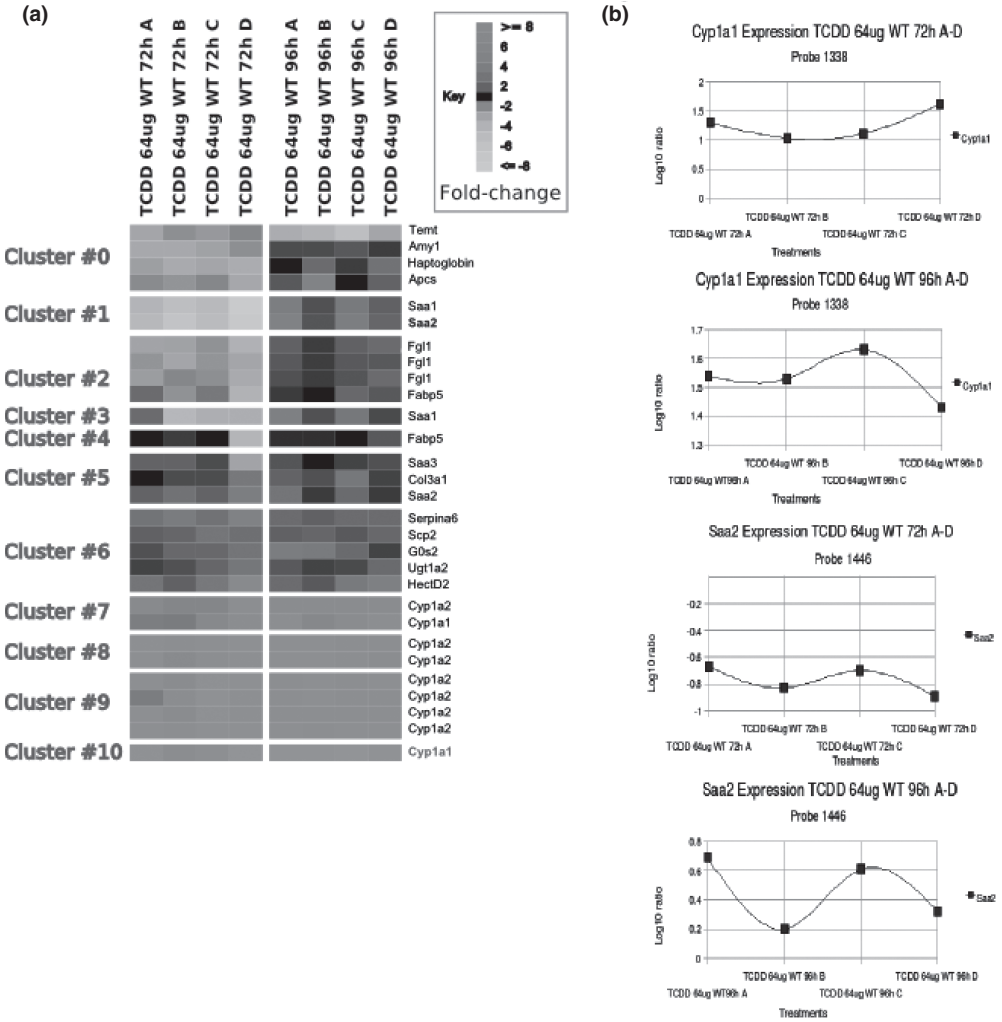


and that of others, technical variance is typically much smaller than biological variance when utilizing commercial arrays (Yauk *et al.*, 2004; Ayroles and Gibson, 2006). Thus, we are advocates of performing biological replicates in the hope that these samples also capture the technical variance.

In the case of the two-color approach, a dye swap can be performed to estimate technical error and account for any dye bias. Dye bias is defined as an intensity difference between samples labeled with different dyes attributable to the dyes instead of a difference in the actual gene expression of the samples (Dobbin *et al.*, 2005). Dye swap is a simple process where the originally Cy5-labeled sample is labeled with Cy3 and the originally Cy3 sample is labeled with Cy5; the samples are then rehybridized to another array. In addition to a single dye-swap approach, if the necessary amount of

**Figure 2** Identifying the biological variation of two sets of treatments. **(a)** This figure represents a k-means (where k = 11, with 11 arbitrarily chosen) clustering analysis of two sets of biological replicates at two different time points with a fold-change threshold of ±3 for genes in at least one of the biological replicate treatments in either set. The first set, TCDD 64 μg WT 72 h A–D (listed at top), composes the left-hand side of the heat map and is separated by a vertical white-space margin from the second set, TCDD 64 μg WT 96 h A–D (listed at top), composing the right-hand side of the heat map. The genes that met the query criteria are listed vertically on the right of the heat map. The heat map is meant to give a qualitative assessment of the biological variation of the biological replicates. **(b)** To obtain a quantitative understanding of biological variation, it is important to look at biological variation at each probe across the replicates. This can be done by utilizing statistical methods. The four charts show the $Log_{10}$ expression values of two genes across the biological replicate sets. These charts represent a qualitative and quantitative assessment of the biological variation of the biological replicates at the two different time points of 72 and 96 hours. The top two charts are for the gene Cyp1a1, labeled in the gene list of the heat map in bold, green text. For the top chart, the average $Log_{10}$ ratio value for Cyp1a1 expression of probe 1338 across the TCDD 64 μg WT 72 h A–D biological replicates is 1.26 and the calculated standard deviation is 0.22. The average $Log_{10}$ ratio value for Cyp1a1 expression of probe 1338 across the TCDD 64 μg WT 96 h A–D biological replicates is 1.53 and the calculated standard deviation is 0.07. The simple statistics suggest that the range of expression for the TCDD 64 μg WT 72 h biological replicates has more biological variation than the TCDD 64 μg WT 96 h biological replicates based on the standard deviation values. The bottom two charts are for the gene Saa2 in the gene list of the heat map in bold, blue text. For the top chart, the average $Log_{10}$ ratio value for Saa2 expression of probe 1446 across the TCDD 64 μg WT 72 h A–D biological replicates is –0.77 and the calculated standard deviation is 0.09. The average $Log_{10}$ ratio value for Saa2 expression of probe 1446 across the TCDD 64 μg WT 96 h A–D biological replicates is 0.46 and the calculated standard deviation is 0.2. The simple statistics suggest that the range of expression for the TCDD 64 μg WT 72 h biological replicates has less biological variation than the TCDD 64 μg WT 96 h biological replicates based on the standard deviation values.

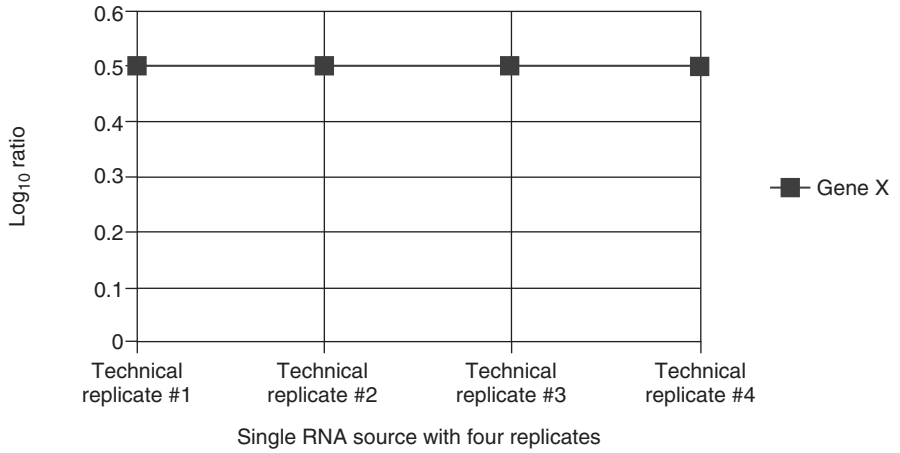


**Figure 3** Ideal set of technical replicates. This figure shows the expression of gene X from a hypothetical set of technical replicates from a single, unique RNA source. Ideally, technical variation would be zero and all of the expression values for gene X across the technical replicates would be the same for this probe and all other probes on the array.

sample is available, multiple dye-swap hybridizations of the same sample/control can be performed.

Finally, the arrangement and number of spots of each probe on the array can also allow for the estimation of technical variation. Placing replicate probe spots sharing the same sequence on the array and randomizing the placement of each probe on the array can address the variance introduced by spot-to-spot printing variation. In practice, assessing variation of probe placement is not always feasible due to the complications of printing probes randomly on an array, but replicate spotting of some or all unique probe sequences is common.

In summary, the final fluorescence value calculated for each probe on the array ideally represents the level of gene transcription. The calculated expression values can be affected by two different sources of variation: biological and technical. It is important to understand both forms of variation occur within the experiment. Biological variation can be estimated by utilizing multiple individual sources of RNA obtained under the same experimental conditions. Technical variation can be assessed by repeating experiments using the same sample sources and can be minimized by experience and use of standardized protocols. In fact, technical variance can be minimized to the extent that cross-platform comparisons can begin to be made due to increased sensitivity allowing for measurement of meaningful biological differences (Bammler *et al.*, 2005; Dobbin *et al.*, 2005; Irizarry *et al.*, 2005). Understanding the biological variation while limiting the technical variation can increase the fidelity of the results obtained. Put simply, technical variation can be minimized by technique, but biological variation is often inherent in the experiment.

## 4 EXPERIMENTAL DESIGNS OF MICROARRAY-BASED TECHNOLOGIES

We have presented a brief introduction to the basic idea of microarrays as a means to identify gene transcription changes on a genomic scale. Though microarrays represent a powerful tool, when toxicogenomics data are published, it is imperative to provide a detailed description of the entire experimental process to ensure the results can be confirmed and reanalyzed by scientific peers. Additionally, proper experimental design and analysis methods must be employed to ensure that those data obtained are the most informative. To that end, the need for standardizing the description of a microarray experiment has been carefully examined (Brazma *et al.*, 2001). In the following sections, we will give a brief overview of some of the basic experimental designs utilized with microarray-based gene experiments, as well as methods for their reporting to the scientific community.

### 4.1 MIAME—A Way to Standardize the Description of Microarray-Based Experiments

After the publication of early microarray studies, it was apparent that there was a need to develop a set of standards to ensure that the necessary technical information was available to interpret or reproduce the reported results. A set of guidelines, called the "Minimum Information About Microarray Experiment" (MIAME) was established as a standard (Brazma *et al.*, 2001). The agreed-upon minimum information consists of six descriptions (Fig. 4). This standard has been accepted by numerous editorial boards and consequently adherence to these guidelines is necessary in order for microarray

**MIAME guidelines**

- Experimental Design—A formal description of the design utilized
- Array Design—A formal description of the array and the probes
- Experimental Samples—A formal description of the samples utilized and the methods used for RNA extraction and labeling
- Hybridizations—A description of the protocol utilized
- Measurements—A formal description of the quantifications utilized as well as images of the actual arrays
- Normalization—A formal description of the types of normalization utilized

**Figure 4** MIAME guidelines.

data to be published by most peer-reviewed journals. For a more detailed analysis, the reader is highly encouraged to refer to Brazma *et al.* (2001) and http://www.mged.org/Workgroups/MIAME/miame.html. An extension of the MIAME guidelines applicable to toxicogenomics, MIAME/Tox, has been designed to address toxicogenomic-specific requirements (http://www.mged.org/Workgroups/rsbi/rsbi.html).

In addition to adhering to the MIAME standards, some peer-reviewed journals in biology require that microarray data must be uploaded into a publicly available database. The purpose of this is to provide transparency and allow other scientists to analyze those data using their own methods. Two of the main databases that accept submission of MIAME-compliant microarray data are Gene Expression Omnibus (GEO) (http://www.ncbi.nlm.nih.gov/geo/) and Array Express (http://www.ebi.ac.uk/microarray-as/aer/). Knowing the requirements for publication of studies utilizing microarray platforms ahead of time will make things easier when submitting those microarray data.

### 4.2 Design of Microarray Gene Expression Experiments

In this section, we will address the design of microarray-based experiments to assess relative changes in gene transcription. One important design choice is identifying the proper control/reference state(s) with which to relate the treated states(s). To that end, the basis of the design is predicated on the question posed and the resources available. To address specific types of experimental questions, "reference," "direct," and "loop" designs can be employed (Fig. 5).

The *reference* design is applicable to the two-color approach and utilizes a reference sample against which the control and experimental samples are compared. A reference sample is a common control RNA source that might come from a historical control sample or a representation of all possible RNAs that might be analyzed in the future. For large-scale experiments, where adequate amounts of reference sample are available, the reference design provides a means of common comparison and historical

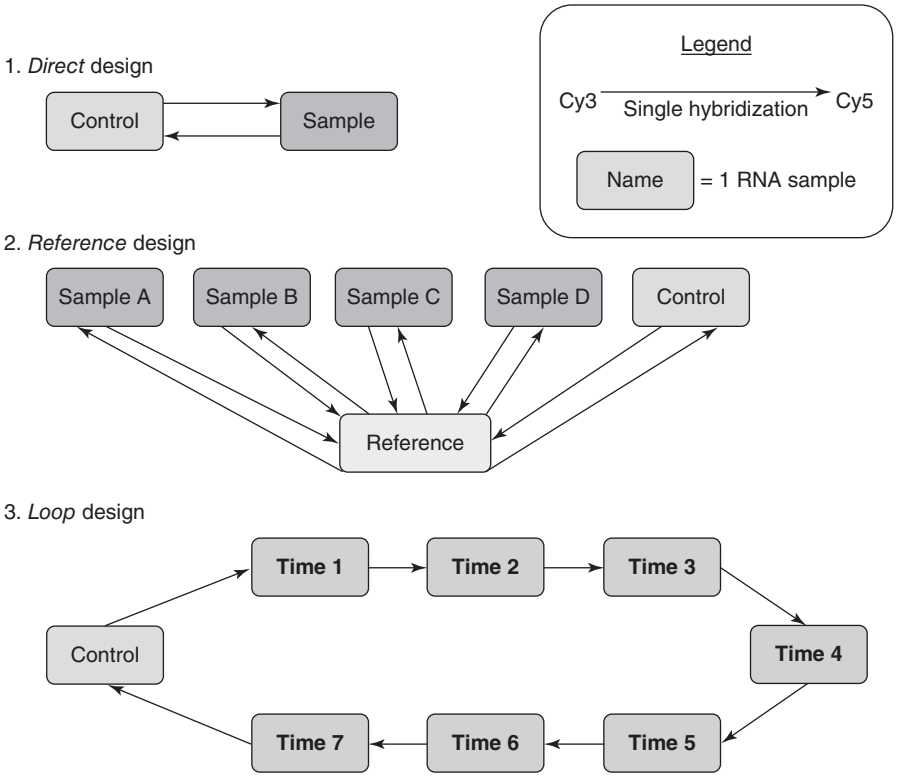


**Figure 5** Three representations of the basic types of microarray design. **1.** The direct design is the most straightforward and utilizes a dye swap to compare the expression of a treated sample directly with a control sample for a total of two microarrays used. **2.** The reference example depicted here is comprised of 10 arrays where four separate treated samples and the control state for this experiment are hybridized against a reference sample. Dye swap is utilized. **3.** The loop design depicts a time series experiment where seven samples and a control state (time = 0) are hybridized in a clockwise loop. This design does not utilize dye swap. Additionally, if one of the arrays were to fail, the loop would be broken affecting the results. None of the representations utilize biological replicates, limiting the ability to identify the level of biological variance.

reflection. From this design, differences in expression can be calculated and compared by subtracting log ratios. As an example, if your reference sample is labeled with the same fluor for all comparisons, different samples can be compared utilizing the following equation:

$$\log_{\mathbf{x}}(\mathbf{Y}/\mathbf{Z}) = \log_{\mathbf{x}}(\mathbf{Y}/\mathbf{R}) - \log_{\mathbf{x}}(\mathbf{Z}/\mathbf{R}) = \log_{\mathbf{x}}(\mathbf{Y}/\mathbf{R})/\log_{\mathbf{x}}(\mathbf{Z}/\mathbf{R}),$$

where **x** is the base of the log, **R** is the fluorescence measurement value of the reference in one channel, and **Y** and **Z** are the fluorescent measurement values in the other channel of the samples whose comparison is being computed. The use of a common reference facilitates the identification of chemical signatures, because the chemically treated samples are all compared with the same reference over the duration of the

study (which could be years). An element of inefficiency inherent to the reference design is that half of the hybridization resources are dedicated to a reference sample.

The *direct* design entails the direct comparison of a treated sample with a control sample, allowing for the identification of relative changes in transcription. When coupled with biological replicates, this is often the most sensitive approach to identify changes in transcription. This is due to the fact that one can compare samples of most interest directly without having a comparison with intermediate reference sample that has the potential to add technical variation. Additionally, subsequent mathematical calculations are not required to obtain the relative differences between two different samples hybridized against a reference on separate arrays.

When the number of conditions being tested is equal to or greater than three, more complex designs can be helpful. The *loop* design involves the sequential hybridization of one sample against another until all the samples have been connected. The loop design is commonly utilized in time-series analyses. In some ways the loop design represents an extension of the basic direct design, because under most circumstances the most interesting samples are compared directly and in sequence. Thus, chronologically proximate samples are usually hybridized against each other to obtain the most temporally sensitive changes in gene transcription. The loop design should incorporate redundancy (i.e. replicates) to ameliorate the effects of a microarray failure that would otherwise break the chain of hybridizations.

In summary, efficiency in sample comparisons is dictated by the length of the path connecting the samples of interest and the number of paths connecting them (Kerr and Churchill, 2001; Yang and Speed, 2002). The direct design offers the shortest path between two samples of interest. Although it is the most simple of the experimental designs, it is limited to the comparison of two conditions. The loop and reference designs require additional data-processing steps to obtain a relative comparison between the different conditions; however, they are amenable to more than two conditions such as in time-series and dose–response studies. The loop design can be more cost-efficient than the reference design (Vinciotti *et al.*, 2005) because in the reference design, 50% of the hybridization resources are directed toward the reference sample. It is important to choose the most efficient design that has the potential to provide the highest degree of statistical significance. However, other factors such as cost and sample resources can affect the choice of experimental design and the level of biological and technical replication. For a more in-depth analysis of these experimental designs and suggestions for when to use them, the reader is directed to the following references: Kerr and Churchill (2001), Yang and Speed (2002), and Hayes and Bradfield (2005).

## 5 THE INTERPRETATION AND ANALYSIS OF DATA GENERATED BY MICROARRAY METHODOLOGIES

One of the major goals of toxicogenomics is to identify and establish global expression profiles resulting from the treatment of a particular chemical or group of chemicals. The common premise of such an experiment is that this profile can be utilized as a means to classify a chemical based on knowledge about related or similar chemicals of known toxicity. We have defined the concept of transcriptional profiles and chemical signatures. We have also highlighted the importance of choosing the correct experimental design to address the particular question being asked. However, simply generating high-density data is not enough. In order to gain biological insight, one must utilize

informatics approaches to analyze and extract meaningful information. Next, we will describe four microarray-based approaches used to analyze microarray data.

## 6 FOUR BASIC MICROARRAY-BASED APPROACHES

Four basic microarray-based approaches deserve special attention: class discovery, class comparison, class prediction, and pathway/network analysis (Quackenbush, 2006). In using these approaches, a basic framework of data analysis takes shape. In general, class discovery approaches are utilized first to look for interesting patterns in the data that may be biologically relevant. Used in conjunction with additional data such as pathology, similarly grouped transcriptional profiles can begin to be assigned to distinct classes. After classes of toxicants are established based on objective criteria, differentially expressed genes (i.e. chemical signatures) can then be identified using class comparison methods to determine how to best distinguish classes in the data. In addition, established classes allow for the methods of class prediction to be utilized to identify the most probable class of unassigned transcriptional profiles. Finally, the incorporation of pathway/network analysis can be utilized for mechanistic analyses.

### 6.1 Class Discovery

Unsupervised methods are utilized during class discovery experiments when one is at the early stages of classification. Unsupervised methods are informatics methods that do not take into consideration *a priori* knowledge about the samples and provide the potential to organize data in an unbiased manner. Common unsupervised approaches used in transcriptional profiling are the clustering approaches. Clustering approaches commonly utilize metrics such as Euclidean distance or correlation coefficients to determine which genes and treatments are most similar to each other. The underlying mathematics of these algorithms is beyond the scope of this work, but the reader is encouraged to consult the following for a more rigorous mathematical description: Eisen *et al.* (1998), Sherlock (2000), and Quackenbush (2001).

There are a large number of different methods to cluster data generated by microarray analyses and this review will concentrate on two of the more common clustering methods: Hierarchical and k-means. Hierarchical clustering is probably the most common means of depicting gene expression profiles (Fig. 6). In this method, genes, as well as the treatments themselves, can be grouped based solely on the underlying changes in gene transcription. In hierarchical clustering, clustered data are represented by a heat map and dendrograms, or hierarchical trees, to represent gene expression changes and similarity relations between the genes and treatments. The most similar treatments and genes can be visually identified as being the most proximate relative to other treatments and genes. The branches within the dendrogram can be assigned a value whose magnitude is utilized to determine relative differences between genes and/or treatments by simply summing the distances along the branches.

k-means clustering allows the researcher to specify the number of distinct groups he or she would like the data to be partitioned into prior to the execution of the clustering algorithm (Fig. 7). Normally, before the portioning process begins, genes are randomly assigned to each cluster to "seed" the algorithm. At that point, the algorithm reassigns genes to different cluster groups such that within each group, the distance between the genes is minimized and between each group the distance between the

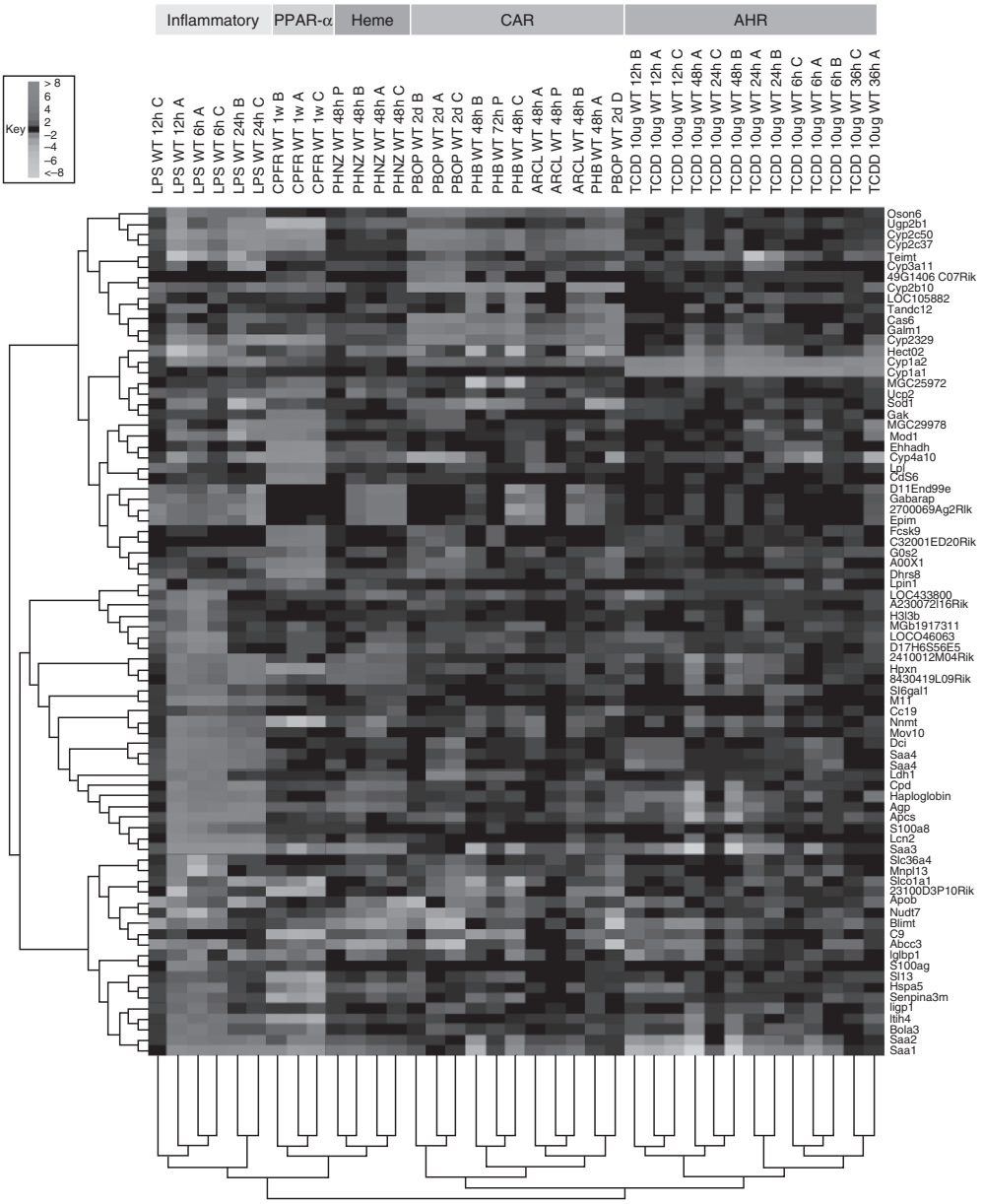


**Figure 6** Hierarchical clustering. This figure represents an example of the class discovery step. Transcription profiles of 37 different C57BL/6 mice treated with either lipopolysaccharide (LPS), a known inflammatory agent; ciprofibrate (CPFR), a known PPAR-α agonist; phenylhydrazine (PHNZ), a known inducer of hemolytic anemia; 1,4-bis(2-(3,5-dichloropyridyloxy))benzene (PBOP), a known CAR agonist; phenobarbital (PHB), a known CAR agonist; Aroclor (ARCL), a known CAR agonist; or 2,3,7,8-tetrachlorodibenzo-*p*-dioxin (TCDD), were clustered using hierarchical clustering. A total of 80 genes for each profile were included based on a magnitude fold change greater than or equal to 3.5 (±3.5). The classes of the transcriptional profiles are listed across the top of the cluster. Though the classes of the transcriptional profiles were known beforehand, by looking at the dendrogram for the treatments at the bottom of the cluster, it is evident that the algorithm was able to accurately cluster the transcriptional profiles based on their classes.

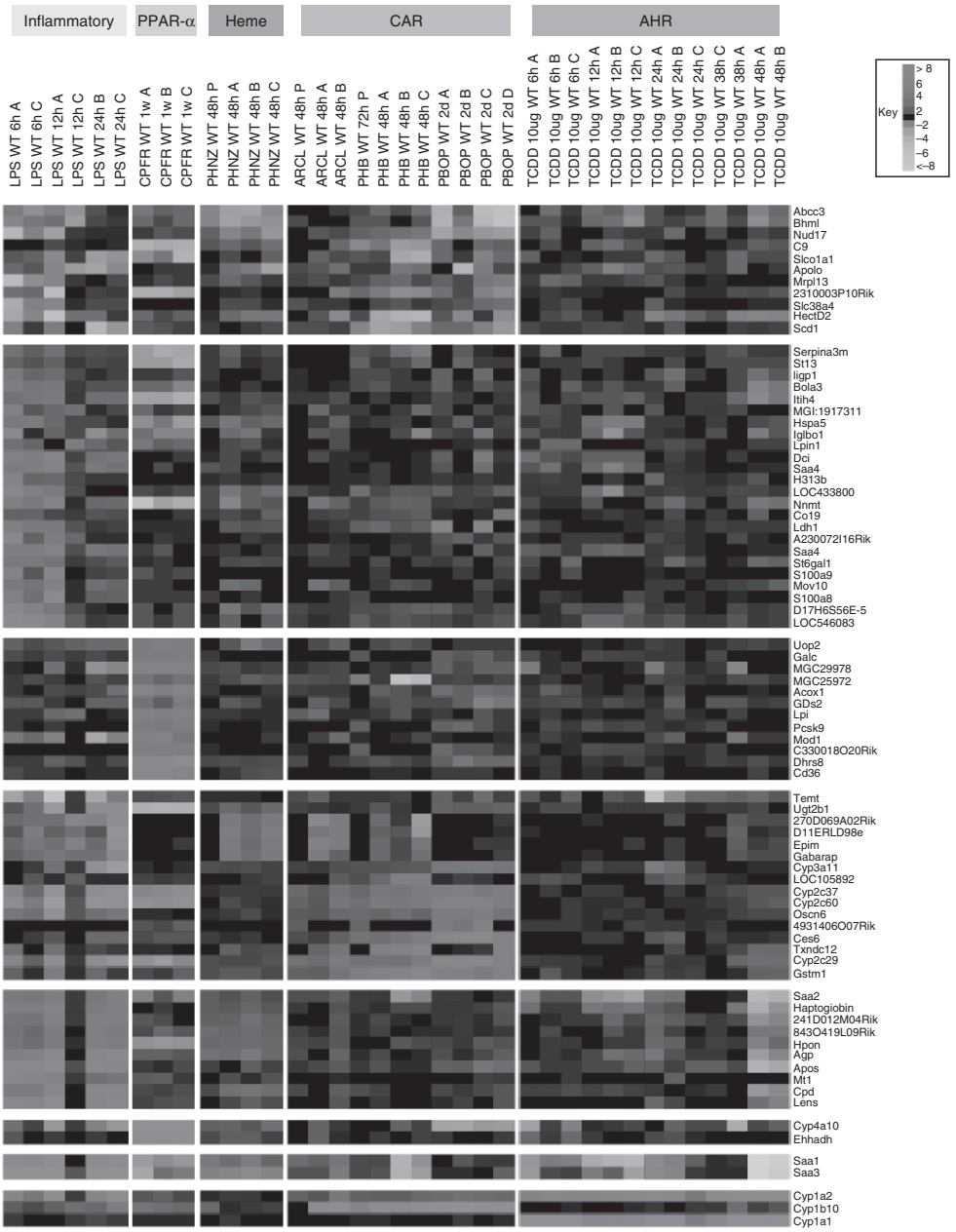


**Figure 7** k-means clustering. This figure represents an example of the class discovery step. Transcription profiles of the same transcriptional profiles from Fig. 5 were clustered using k-means clustering. Again, a total of 80 genes for each profile were included based on a magnitude fold change greater than or equal to 3.5 (±3.5). The classes of the transcriptional profiles are listed across the top of the cluster and were ordered based upon their assigned class. An arbitrary number for k, 8, was chosen to segregate the genes based on their expression levels. This allows for the qualitative identification of distinct clusters of gene expression patterns allowing for discrimination between the classes. Clustering also allows for the qualitative examination of disparities in biological replicates.

gene groups is maximized. The algorithm runs a predetermined amount of time until a particular threshold (e.g. algorithm running time) is reached. The end result is a heat map that consists of a predefined number of groups of genes that have been deemed to be most similar within each group and most disparate between groups.

The class discovery or clustering methods presented are useful in identifying potential classes of chemicals and biologically interesting patterns in gene expression. Additionally, they provide a means of generating hypotheses. However, they are limited in their ability to identify chemically induced signatures due to their unsupervised nature. On the other hand, class comparison methods can be utilized to identify chemically induced signatures.

### 6.2 Class Comparison

Class comparison approaches can be utilized to analyze transcriptional profiles and identify chemically induced signatures based on differences in expression between predefined classes of transcriptional profiles. A goal of toxicogenomics is to utilize transcriptional profiles to identify chemically induced signatures. To that end, statistical methods have been developed to discover chemically induced signatures. Some examples of these methods include statistical analysis of microarrays (SAM) (Tusher *et al.*, 2001), *t*-tests (Baggerly *et al.*, 2001), analysis of variance (Shannon *et al.*, 2003), regression modeling (Thomas *et al.*, 2001a), the Empirical Bayes method (Efron and Tibshirani, 2001), and the mixture model (Pan *et al.*, 2003). When looking to identify chemically induced signatures that can discriminate between different classes, class comparison approaches start with a set of transcriptional profiles that have been assigned to predefined classes based on *a priori* knowledge. Then, statistical methods are able to compare the gene expression profiles of different groups and establish sets of genes (i.e. chemically induced signatures) that can be utilized to discriminate between classes of toxins. These chemically induced signatures provide a basis for the classification of novel chemicals through the utilization of class prediction approaches. These signatures have been referred to as predictive gene sets, because they can subsequently be utilized in class prediction approaches (Thomas *et al.*, 2001b).

Unlike class discovery methods, class comparison requires the intervention of expert knowledge to assign transcriptional profiles to distinct classes. This aspect represents a supervised approach to the analysis of microarray data. In assigning the classes, an inherent bias is entered into the class comparison approach. If the classes of the chemicals are well known and characterized, theoretically the identified chemically induced signatures can be considered to be accurate representations of the chemicals in the assigned class. However, improper assignment of classes to transcriptional profiles can adversely affect the results and interpretation of methods utilizing supervised approaches.

### 6.3 Class Prediction

Class prediction is a supervised approach allowing for the incorporation of *a priori* information into algorithms so that members in a class can be predicted based on transcriptional profiles. In general, class prediction approaches require the utilization of a training set of data and a test set of data. Prior to the actual classification of an unknown transcriptional profile, some class prediction methods incorporate the methods of the class comparison approach to determine chemically induced signatures

(a)

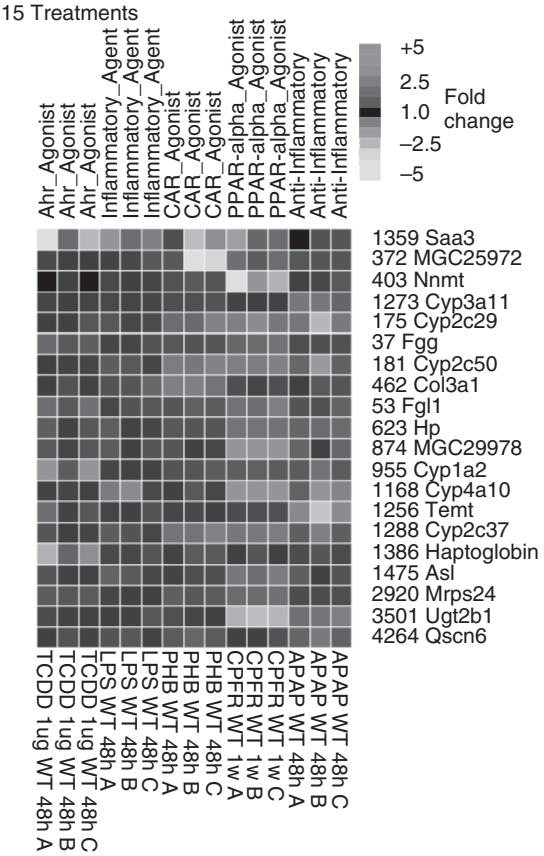


(b)

| Rank | Gene name | Accuracy |
|---|---|---|
| 1 | 1359 Saa3 | 46.67 |
| 2 | 372 MGC25972 | 93.33 |
| 3 | 403 Nnmt | 93.33 |
| 4 | 1273 Cyp3a11 | 96 |
| 5 | 175 Cyp2c29 | 100 |
| 6 | 37 Fgg | 100 |
| 7 | 181 Cyp2c50 | 100 |
| 8 | 462 Col3a1 | 100 |
| 9 | 53 Fg11 | 100 |
| 10 | 623 Hp | 100 |
| 11 | 874 MGC29978 | 100 |
| 12 | 955 Cyp1a2 | 100 |
| 13 | 1168 Cyp4a10 | 100 |
| 14 | 1256 Temt | 100 |
| 15 | 1288 Cyp2c37 | 100 |
| 16 | 1386 Haptoglobin | 100 |
| 17 | 1475 Asl | 100 |
| 18 | 2920 Mrps24 | 100 |
| 19 | 3501 Ugt2b1 | 100 |
| 20 | 4264 Qscn6 | 100 |

**Figure 8** An example of Naive Bayes classification. This figure represents the output of a Naive Bayes-based classification algorithm that first identifies a predefined number of genes constituting a subset of the genes in a transcriptional profile produced by the array platform. In this case, the transcriptional profiles represent a set of gene probes printed on a custom cDNA array. RNA was obtained from murine livers and then competitively hybridized with control RNA. Three different treatments from five different class designations defined by the mechanism of action were selected for the generation of a set of genes that could be used to discriminate between the different classes. The classes and respective treatments are (i) Ahr agonist class consisting of three biological replicate treatments of 1-μg TCDD (dioxin) at 48 hours after dosing; (ii) inflammatory agent class consisting of three biological replicate treatments of 1000-μg lipopolysaccharide (LPS) at 48 hours after exposure; (iii) CAR agonist class consisting of three biological replicates of phenobarbital (PHB) at 48 hours after dosing; (iv) PPAR-alpha agonist class consisting of three biological replicate treatments of ciprofibrate (CPFR) at seven days after dosing; and (v) anti-inflammatory class consisting of three biological replicate treatments of 500 μg/kg acetominophen (APAP) at 48 hours after dosing. **(a)** A heat map representation of the 20 genes selected by information gain. The 20 selected genes are placed along the right-hand side of the heat map in the order of their selection. The numbers preceding the gene names represent the probe number on the array. The classes of treatments are labeled across the top of the heat map and the corresponding treatments are labeled across the bottom. **(b)** A table depicting the order in which the genes were selected by the algorithm based on their ability to classify the treatments

that can be used to discriminate between the unknown transcriptional profiles. The training set and test set of data should represent sets of chemically induced signatures or transcriptional profiles with well-known characteristics and an appropriate class assignment based on expert knowledge and objective criteria. The supervised algorithm then develops a model based on the training set of data and utilizes the generated model to determine the classification of profiles in the test set. In general, the accuracy of the algorithm is based on the number of correctly assigned transcriptional profiles in the test set of data. The applicability of a generated model to future classification of novel transcriptional profiles is dependent on the robustness (i.e. number of transcription profiles, number of unique chemical classes, etc.) of the original transcriptional profiles it was built from. As an example, the broader the application, say for instance in classifying a random chemical drawn from the currently 70,000–80,000 chemicals in commerce for which toxicity is not known, the greater the requirement for a more robust training set to effectively assign a class.

Some examples of class prediction methods utilized to classify chemical profiles include k-nearest neighbors (Theilhaber *et al.*, 2002), weighted voting (Golub *et al.*, 1999), artificial neural networks (Bloom *et al.*, 2004), discriminant analysis (Nguyen and Rocke, 2002; Orr and Scherf, 2002; Antoniadis *et al.*, 2003; Le *et al.*, 2003), classification and regression trees (Boulesteix *et al.*, 2003), Naive Bayes (Thomas *et al.*, 2001b), and support vector machines (Brown *et al.*, 2000; Ramaswamy *et al.*, 2001, 2003).

Proof of principle for utilizing classification methods in toxicogenomics has support from the field of medical oncology (Ross *et al.*, 2000; Scherf *et al.*, 2000; Nielsen *et al.*, 2002; Ramaswamy *et al.*, 2003). Not only have transcriptional profiles been utilized to differentiate tumor types, but studies have correlated transcriptional profiles with clinical outcomes (Dhanasekaran *et al.*, 2001; Sorlie *et al.*, 2001; West *et al.*, 2001; Pomeroy *et al.*, 2002; Singh *et al.*, 2002; van 't Veer *et al.*, 2002; van de Vijver *et al.*, 2002; Williams *et al.*, 2004; Dyrskjot *et al.*, 2005).

To provide a brief example of a class prediction-based algorithm, the output of a Naive Bayes-based classification algorithm used in our laboratory is presented in Fig. 8. The purpose of this class prediction approach is to identify a set of genes that can predict the chemical class based on a model generated from a set of transcriptional profiles and their assigned classes. For an in-depth explanation of this approach and the mathematical basis for this algorithm, the reader is encouraged to refer to Thomas *et al.* (2001b).

**Figure 8** *Continued*
correctly. After the initial selection of genes utilizing information gain, each gene is tested for its ability to classify the treatments with the greatest degree of accuracy using a "leave-one-out cross validation" approach. The first gene is removed from the initial set of 20 and then subsequently tested with all the other genes to see which two provide the greatest predictive accuracy. The second is removed and the three genes that give the best predictive accuracy are selected and the iterative process of removing one gene and adding it to the most predictive list continues until all 20 genes have been placed in order. As can be seen here, five genes (highlighted in red) are enough to give 100% predictive accuracy with the five classes of treatments. Any additional genes (highlighted in gray) added to the list after the first five do not improve or degrade the predictive accuracy. The reader is encouraged to consult the following for further insight: Thomas *et al.* (2001b).

### 6.4 Mechanistic Analysis—Methods Used for Pathway/Network Analysis

In addition to the identification of toxic signatures, toxicogenomics can provide evidence for the identification of the underlying mechanisms of a chemical's toxicity. Utilizing "pathway mapping," toxicogenomic methods can assess the effects of a poison on known toxic signaling pathways and elucidate potential novel mechanisms when a signature cannot be ascribed to any previously identified mechanism. In the context of transcriptional profiling, pathway mapping can be defined as the process of overlaying data derived from transcription studies onto their respective established networks. This process can allow for the generation of hypotheses that are testable by utilizing complementary functional genomics methods to selectively target members of an identified signaling pathway for perturbation.

Class comparison and class prediction approaches can, respectively, identify groups of genes that are differentially expressed and provide a means for classifying unknown transcriptional profiles. However, alone they do not provide utility in terms of identifying the underlying biological mechanisms of response. Pathway/network analysis methods used in combination with transcriptional profiles can help to establish a biological context. Pathway/network analysis methods are in the early stages of development at the moment. The major limiting factor in these analyses is the lack of knowledge of the underlying biology in higher organisms such as the mouse, rat, and human. This factor is being addressed as interactions between biomolecules of all types—proteins, genes, noncoding DNA, and metabolites—are now being elucidated. Much like the genomic projects that have been utilized to catalog every known gene and its regulatory elements, there are a number of projects with the goal of cataloging the functional role of the product(s) of every gene and the role of every metabolite within cells of higher organisms. Some of these projects include those of the Gene Ontology Consortium (GO) (Ashburner, Ball, *et al.*, 2000) and the Kyoto Encyclopedia of Genes and Genomes (KEGG) (Kanehisa, Goto, *et al.*, 2006). With these annotations and interaction data, it is now possible to probe transcriptional profiles for potential biological pathways affected by a toxic exposure. There are tools available for the analysis, visualization, and modeling of these annotation and molecular interaction data. These tools include Cytoscape (Shannon, Markiel, *et al.*, 2003), Osprey (Breitkreutz, Stark, *et al.*, 2003), PathwayAssist (Nikitin, Egorov, *et al.*, 2003), Pathways Database System (Krishnamurthy, Nadeau, *et al.*, 2003), GeneGO (www.genego.com), VisANT (Hu, Mellor, *et al.*, 2005), and BiologicalNetworks (Baitaluk, Sedova, *et al.*, 2006).

At the simplest level, interrogation of transcriptional profiles via integration with these pathway/network data involves using a list of genes of interest and selecting from the interaction data all interactions associated with those genes, and overlaying the gene expression values onto the members of the network. A gene network is a graphical representation of the molecular relationships between genes or gene products. Genes or gene products are represented as nodes, and the biological relationship between two nodes is represented as a line or edge. The nodes can be color-coded based upon the relative expression values assigned to them by microarray analyses.

Another method utilizes already defined signaling pathways and networks and overlays the data from transcriptional profiles (Dahlquist *et al.*, 2002; Doniger *et al.*, 2003). When used in conjunction with metabolic pathways, these methods can elucidate chemically induced perturbations. This method may be more appropriate for identifying dose–response relationships when dealing with well-characterized toxins.

More sophisticated approaches use probabilistic/statistical methods to determine or predict the metabolic pathways, transcriptional networks, or protein interaction networks affected based on the transcriptional profile(s) used as input. Examples of these type of methods include simple Boolean Networks (Akutsu *et al.*, 2000; Savoie *et al.*, 2003; Soinov, 2003), Probabilistic Boolean Networks (Shmulevich *et al.*, 2002a, b; Datta *et al.*, 2004; Hashimoto *et al.*, 2004), and Bayesian Networks (Friedman *et al.*, 2000; Imoto *et al.*, 2003; Savoie *et al.*, 2003; Tamada *et al.*, 2003; Zou and Conzen, 2005). In-depth analysis of these types of models is beyond the scope of this work.

In summary, pathway/network analysis holds great promise. However, the major limitation of these methods is the current overall paucity of interaction data. Currently, these methods are utilized more as a descriptive tool. However, in "simpler" organisms such as *Escherichia coli*, predictive models are being generated (Famili, Mahadevan, *et al.*, 2005; Wiback, Mahadevan, *et al.*, 2004; Palsson, 2006).

## 7 INFORMATICS TOOLS UTILIZED BY THESE INFORMATICS APPROACHES

Most software required to perform these algorithms is commercially available and new algorithms appear in the literature with great frequency. Commercial options can be expensive for a smaller lab, but can offer a good customer support system to answer questions as well as a "turn-key" approach helpful to a smaller lab. An alternative solution is to use an open-source software solution that can be customized to specific user objectives. There are a number of robust statistical packages available in the public domain. The "R" statistical package in conjunction with Bioconductor is a freely available software package project to provide tools for the analysis and comprehension of genomic data. Bioconductor is a community-supported effort and provides fairly detailed documentation of its features. Both commercial and free software options will require some effort on the part of the end user to implement.

## 8 CONCLUSION

Within this review we have focused on how microarray gene expression analyses can be utilized in the context of toxicogenomics. We began by introducing fundamental aspects of the microarray assay for detecting gene transcriptional changes and conceptualized transcriptional profiles and chemically induced signatures. Because there are a number of different microarray platforms available and the scope of their utilization is applicable to a large number of different fields of study, we presented the importance of standardized description of microarray experiments as a way to provide transparency and a means for others to confirm the results. From there we stressed the importance of a well-thought-out experimental design to generate data that can be used to answer the question posed while taking into consideration limiting factors such as cost and resources. Finally, a basic bioinformatic framework was presented to facilitate the process of using transcriptional profiles to identify potential classes of toxic signatures, to identify genes that are differentially expressed between classes of transcriptional profiles, to identify the class of an unknown transcriptional profile based on a model generated by supervised methods, and to utilize transcriptional profiles as a

means of probing known biological pathways or networks. We provided examples of some of the basic algorithms associated with the different methods of this framework.

At the moment, toxicogenomic methods provide an attractive complement to classical toxicology experiments. However, it is thought that at some point, enough data can be generated by these methods to begin to utilize known toxic signatures to streamline and expedite the testing process of the current large number of outstanding uncharacterized chemicals. This could entail the utilization of *in vitro* cell or tissue culture methods to predict the toxic effects of a chemical without the costly utilization of animal models. In general, the dose of a particular toxicant determines whether or not it will have adverse effects. As the sensitivity and fidelity of toxicogenomic methods increase, the ability of these methods to identify subtle changes will enhance our ability to identify thresholds of toxic exposure, as well as make more informed regulatory decisions. Finally, as the functional annotation of gene products progresses and the interactions between biomolecules are elucidated, the ability to use pathway/network analysis methods to identify mechanisms of action for toxins will be enhanced.

## REFERENCES

Akutsu, T., Miyano, S. *et al.* (2000). Algorithms for identifying Boolean networks and related biological networks based on matrix multiplication and fingerprint function. *J. Comput. Biol.*, *7* (3–4), 331–343.

Allen, D.G., Pearse, G. *et al.* (2004). Prediction of rodent carcinogenesis: an evaluation of prechronic liver lesions as forecasters of liver tumors in NTP carcinogenicity studies. *Toxicol. Pathol.*, *32* (4), 393–401.

Antoniadis, A., Lambert-Lacroix, S. *et al.* (2003). Effective dimension reduction methods for tumor classification using gene expression data. *Bioinformatics*, *19* (5), 563–570.

Ashburner, M., Ball, C.A. *et al.* (2000). Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.*, *25* (1): 25–29.

Ayroles, J.F., Gibson, G. (2006). Analysis of variance of microarray data. *Methods Enzymol.*, *411*, 214–233.

Baggerly, K.A., Coombes, K.R. *et al.* (2001). Identifying differentially expressed genes in cDNA microarray experiments. *J. Comput. Biol.*, *8* (6), 639–659.

Baitaluk, M., M. Sedova, *et al.* (2006). BiologicalNetworks: visualization and analysis tool for systems biology. *Nucleic Acids Res.*, *34* (Web Server issue): W466–471.

Bammler, T., Beyer, R.P. *et al.* (2005). Standardizing global gene expression analysis between laboratories and across platforms. *Nat. Methods*, *2* (5), 351–356.

Bloom, G., Yang, I.V. *et al.* (2004). Multi-platform, multi-site, microarray-based human tumor classification. *Am. J. Pathol.*, *164* (1), 9–16.

Boulesteix, A.L., Tutz, G. *et al.* (2003). A CART-based approach to discover emerging patterns in microarray data. *Bioinformatics*, *19* (18), 2465–2472.

Brazma, A., Hingamp, P. *et al.* (2001). Minimum information about a microarray experiment (MIAME)-toward standards for microarray data. *Nat. Genet.*, *29* (4), 365–371.

Breitkreutz, B.J., Stark, C. *et al.* (2003). Osprey: a network visualization system. *Genome Biol.*, *4* (3): R22.

Brown, M.P., Grundy, W.N. *et al.* (2000). Knowledge-based analysis of microarray gene expression data by using support vector machines. *Proc. Natl. Acad. Sci. U.S.A.*, *97* (1), 262–267.

Bushel, P.R., Hamadeh, H.K. *et al.* (2002). Computational selection of distinct class- and subclass-specific gene expression signatures. *J. Biomed. Inform.*, *35* (3), 160–170.

Churchill, G.A. (2002). Fundamentals of experimental design for cDNA microarrays. *Nat. Genet.*, *32* (Suppl.), 490–495.

Committee on Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment, N. R. C. (2007). *Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment*, The National Academies Press.

Dahlquist, K.D., Salomonis, N. *et al.* (2002). GenMAPP, a new tool for viewing and analyzing microarray data on biological pathways. *Nat. Genet.*, *31* (1), 19–20.

Datta, A., Choudhary, A. *et al.* (2004). External control in Markovian genetic regulatory networks: the imperfect information case. *Bioinformatics*, *20* (6), 924–930.

Dhanasekaran, S.M., Barrette, T.R. *et al.* (2001). Delineation of prognostic biomarkers in prostate cancer. *Nature*, *412* (6849), 822–826.

Dobbin, K.K., Kawasaki, E.S. *et al.* (2005). Characterizing dye bias in microarray experiments. *Bioinformatics*, *21* (10), 2430–2437.

Doniger, S.W., Salomonis, N. *et al.* (2003). MAPPFinder: using Gene Ontology and GenMAPP to create a global gene-expression profile from microarray data. *Genome Biol.*, *4* (1), R7.

Dyrskjot, L., Zieger, K. *et al.* (2005). A molecular signature in superficial bladder carcinoma predicts clinical outcome. *Clin. Cancer Res.*, *11* (11), 4029–4036.

Efron, B., Tibshirani, R. (2001). Microarrays, empirical Bayes methods, and false discovery rates. *Genet. Epidemiol.*, *23* (1), 70–86.

Eisen, M.B., Spellman, P.T. *et al.* (1998). Cluster analysis and display of genome-wide expression patterns. *Proc. Natl. Acad. Sci. U.S.A.*, *95* (25), 14863–14868.

Elcombe, C.R., Odum, J. *et al.* (2002). Prediction of rodent nongenotoxic carcinogenesis: evaluation of biochemical and tissue changes in rodents following exposure to nine nongenotoxic NTP carcinogens. *Environ. Health Perspect.*, *110* (4), 363–375.

Famili, I., Mahadevan, R. *et al.* (2005). k-Cone analysis: determining all candidate values for kinetic parameters on a network scale. *Biophys. J.*, *88* (3): 1616–1625.

Fielden, M.R., Brennan, R. *et al.* (2007). A gene expression biomarker provides early prediction and mechanistic assessment of hepatic tumor induction by nongenotoxic chemicals. *Toxicol. Sci.*, *99* (1), 90–100.

Friedman, N., Linial, M. *et al.* (2000). Using Bayesian networks to analyze expression data. *J. Comput. Biol.*, *7* (3–4), 601–620.

Golub, T.R., Slonim, D.K. *et al.* (1999). Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science*, *286* (5439), 531–537.

Hamadeh, H.K., Bushel, P.R. *et al.* (2002a). Prediction of compound signature using high density gene expression profiling. *Toxicol. Sci.*, *67* (2), 232–240.

Hamadeh, H.K., Bushel, P.R. *et al.* (2002b). Gene expression analysis reveals chemical-specific profiles. *Toxicol. Sci.*, *67* (2), 219–231.

Hamadeh, H.K., Jayadev, S. *et al.* (2004). Integration of clinical and gene expression endpoints to explore furan-mediated hepatotoxicity. *Mutat. Res.*, *549* (1–2), 169–183.

Hashimoto, R.F., Kim, S. *et al.* (2004). Growing genetic regulatory networks from seed genes. *Bioinformatics*, *20* (8), 1241–1247.

Hayes, K.R., Bradfield, C.A. (2005). Advances in toxicogenomics. *Chem. Res. Toxicol.*, *18* (3), 403–414.

Heinloth, A.N., Irwin, R.D. *et al.* (2004). Gene expression profiling of rat livers reveals indicators of potential adverse effects. *Toxicol. Sci.*, *80* (1), 193–202.

Hu, Z., Mellor, J. *et al.* (2005). VisANT: data-integrating visual framework for biological networks and modules. *Nucleic Acids Res.*, *33* (Web Server issue): W352–357.

Imoto, S., Kim, S. *et al.* (2003). Bayesian network and nonparametric heteroscedastic regression for nonlinear modeling of genetic network. *J. Bioinform. Comput. Biol.*, *1* (2), 231–252.

Irizarry, R.A., Warren, D. *et al.* (2005). Multiple-laboratory comparison of microarray platforms. *Nat. Methods*, *2* (5), 345–350.

Kanehisa, M., Goto, S. *et al.* (2006). From genomics to chemical genomics: new developments in KEGG. *Nucleic Acids Res.*, *34* (Database issue): D354–357.

Kerr, M.K., Churchill, G.A. (2001). Experimental design for gene expression microarrays. *Biostatistics*, *2* (2), 183–201.

Krishnamurthy, L., Nadeau, J. *et al.* (2003). Pathways database system: an integrated system for biological pathways. *Bioinformatics*, *19* (8): 930–937.

Le, Q.T., Sutphin, P.D. *et al.* (2003). Identification of osteopontin as a prognostic plasma marker for head and neck squamous cell carcinomas. *Clin. Cancer Res.*, *9* (1), 59–67.

Nguyen, D.V., Rocke, D.M. (2002). Multi-class cancer classification via partial least squares with gene expression profiles. *Bioinformatics*, *18* (9), 1216–1226.

Nielsen, T.O., West, R.B. *et al.* (2002). Molecular characterisation of soft tissue tumours: a gene expression study. *Lancet*, *359* (9314), 1301–1307.

Nikitin, A., Egorov, S., Daraselia, N., Mazo, I. (2003). Pathway studio—the analysis and navigation of molecular networks. *Bioinformatics*, *19* (16): 2155–2157.

Oleksiak, M.F., Churchill, G.A. *et al.* (2002). Variation in gene expression within and among natural populations. *Nat. Genet.*, *32* (2), 261–266.

Orr, M.S., Scherf, U. (2002). Large-scale gene expression analysis in molecular target discovery. *Leukemia*, *16* (4), 473–477.

Palsson, B. (2006). *Systems Biology: Properties of Reconstructed Networks*. Cambridge University Press, New York.

Pan, W., Lin, J. *et al.* (2003). A mixture model approach to detecting differentially expressed genes with microarray data. *Funct. Integr. Genomics*, *3* (3), 117–124.

Pomeroy, S.L., Tamayo, P. *et al.* (2002). Prediction of central nervous system embryonal tumour outcome based on gene expression. *Nature*, *415* (6870), 436–442.

Quackenbush, J. (2001). Computational analysis of microarray data. *Nat. Rev. Genet.*, *2* (6), 418–427.

Quackenbush, J. (2006). Computational approaches to analysis of DNA microarray data. *Methods Inf. Med.*, *45* (Suppl. 1), 91–103.

Ramaswamy, S., Ross, K.N. *et al.* (2003). A molecular signature of metastasis in primary solid tumors. *Nat. Genet.*, *33* (1), 49–54.

Ramaswamy, S., Tamayo, P. *et al.* (2001). Multiclass cancer diagnosis using tumor gene expression signatures. *Proc. Natl. Acad. Sci. U.S.A.*, *98* (26), 15149–15154.

Ross, D.T., Scherf, U. *et al.* (2000). Systematic variation in gene expression patterns in human cancer cell lines. *Nat. Genet.*, *24* (3), 227–235.

Savoie, C.J., Aburatani, S. *et al.* (2003). Use of gene networks from full genome microarray libraries to identify functionally relevant drug-affected genes and gene regulation cascades. *DNA Res.*, *10* (1), 19–25.

Scherf, U., Ross, D.T. *et al.* (2000). A gene expression database for the molecular pharmacology of cancer. *Nat. Genet.*, *24* (3), 236–244.

Shannon, P., Markiel, A. *et al.* (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.*, *13* (11), 2498–2504.

Sherlock, G. (2000). Analysis of large-scale gene expression data. *Curr. Opin. Immunol.*, *12* (2), 201–205.

Shmulevich, I., Dougherty, E.R. *et al.* (2002a). Probabilistic Boolean Networks: a rule-based uncertainty model for gene regulatory networks. *Bioinformatics*, *18* (2), 261–274.

Shmulevich, I., Dougherty, E.R. *et al.* (2002b). Gene perturbation and intervention in probabilistic Boolean networks. *Bioinformatics*, *18* (10), 1319–1331.

Singh, D., Febbo, P.G. *et al.* (2002). Gene expression correlates of clinical prostate cancer behavior. *Cancer Cell*, *1* (2), 203–209.

Soinov, L.A. (2003). Supervised classification for gene network reconstruction. *Biochem. Soc. Trans.*, *31* (Pt 6), 1497–1502.

Sorlie, T., Perou, C.M. *et al.* (2001). Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc. Natl. Acad. Sci. U.S.A.*, *98* (19), 10869–10874.

Spruill, S.E., Lu, J. *et al.* (2002). Assessing sources of variability in microarray gene expression data. *Biotechniques*, *33* (4), 916–920, 922–923.

Steiner, G., Suter, L. *et al.* (2004). Discriminating different classes of toxicants by transcript profiling. *Environ. Health Perspect.*, *112* (12), 1236–1248.

Tamada, Y., Kim, S. *et al.* (2003). Estimating gene networks from gene expression data by combining Bayesian network model with promoter element detection. *Bioinformatics*, *19* (Suppl. 2), ii227–ii236.

Theilhaber, J., Connolly, T. *et al.* (2002). Finding genes in the C2C12 osteogenic pathway by k-nearest-neighbor classification of expression data. *Genome Res.*, *12* (1), 165–176.

Thomas, J.G., Olson, J.M. *et al.* (2001a). An efficient and robust statistical modeling approach to discover differentially expressed genes using genomic expression profiles. *Genome Res.*, *11* (7), 1227–1236.

Thomas, R.S., Rank, D.R. *et al.* (2001b). Identification of toxicologically predictive gene sets using cDNA microarrays. *Mol. Pharmacol.*, *60* (6), 1189–1194.

Tusher, V.G., Tibshirani, R. *et al.* (2001). Significance analysis of microarrays applied to the ionizing radiation response. *Proc. Natl. Acad. Sci. U.S.A.*, *98* (9), 5116–5121.

van 't Veer, L.J., Dai, H. *et al.* (2002). Gene expression profiling predicts clinical outcome of breast cancer. *Nature*, *415* (6871), 530–536.

van de Vijver, M.J., He, Y.D. *et al.* (2002). A gene-expression signature as a predictor of survival in breast cancer. *N. Engl. J. Med.*, *347* (25), 1999–2009.

Vinciotti, V., Khanin, R. *et al.* (2005). An experimental evaluation of a loop versus a reference design for two-channel microarrays. *Bioinformatics*, *21* (4), 492–501.

West, M., Blanchette, C. *et al.* (2001). Predicting the clinical status of human breast cancer by using gene expression profiles. *Proc. Natl. Acad. Sci. U.S.A.*, *98* (20), 11462–11467.

Wiback, S.J., Mahadevan, R. *et al.* (2004). Using metabolic flux data to further constrain the metabolic solution space and predict internal flux patterns: the Escherichia coli spectrum. *Biotechnol. Bioeng.*, *86* (3): 317–331.

Williams, R.D., Hing, S.N. *et al.* (2004). Prognostic classification of relapsing favorable histology Wilms tumor using cDNA microarray expression profiling and support vector machines. *Genes Chromosomes Cancer*, *41* (1), 65–79.

Yang, Y.H., Speed, T. (2002). Design issues for cDNA microarray experiments. *Nat. Rev. Genet.*, *3* (8), 579–588.

Yauk, C.L., Berndt, M.L. *et al.* (2004). Comprehensive comparison of six microarray technologies. *Nucleic Acids Res.*, *32* (15), e124.

Zou, M., Conzen, S.D. (2005). A new dynamic Bayesian network (DBN) approach for identifying gene regulatory networks from time course microarray data. *Bioinformatics*, *21* (1), 71–79.

# Role of Bioactivation Reactions in Chemically Induced Nephrotoxicity

LAWRENCE H. LASH

## 1 OVERVIEW OF RENAL STRUCTURE AND FUNCTION: TOXICOLOGICAL IMPLICATIONS

The kidneys are uniquely sensitive to toxicants because they receive and filter a large quantity of blood relative to their weight. On average, the two kidneys receive 25% of the cardiac output, while comprising only about 1% of total body weight. This fact, along with the presence of a myriad of plasma membrane transport proteins, which can result in a high degree of accumulation and concentration of chemicals within the tubular epithelial cells, and drug metabolism enzymes that can result in bioactivation of inert or nontoxic chemicals, contribute to the susceptibility of the kidneys. The major function of the kidneys is to excrete waste products while maintaining total body electrolytes, water, and acid–base balance. Thus, the kidneys are the primary organs responsible for maintaining the constancy of the internal environment. They accomplish this task by three general mechanisms: (i) glomerular filtration, (ii) tubular reabsorption, and (iii) tubular secretion. Although the focus of this chapter is on the role of metabolism in chemically induced nephrotoxicity, it is important to appreciate the role that normal renal physiology plays in the intoxication process.

The mammalian kidney is a complex organ that possesses structural divisions at both the whole organ level and the level of the nephron, which is the basic functional unit of the kidney. A detailed consideration of renal physiology and morphology is beyond the scope of this chapter. However, some key points will be made as they critically impact renal handling of drugs and patterns of chemically induced injury that are observed. Interested readers are referred to a variety of reviews and other publications that discuss renal physiology and nephron structure in greater depth (e.g. Horster, 1978; Jacobson, 1981; Walker and Valtin, 1982; Guder and Ross, 1984; Brenner, 1996; Guder, Wagner and Wirthensohn, 1986; Sands and Verlander, 2005).

At the whole organ level, the kidney is subdivided into cortex and medulla, with the latter being further divided into outer and inner stripes of the outer medulla and inner medulla or papilla. Each of these zones is comprised of distinctive cell types and

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

receives distinctive patterns of blood flow. For example, the renal medulla receives much less blood supply than the cortex, generally contains lower as well as different profiles of activities of drug-metabolizing enzymes and membrane transporters than the cortex, and exhibits a higher degree of anaerobic metabolism than the cortex. Some of the consequences of these differences are that the renal medulla is particularly sensitive to hypoxia, ischemia, or drugs that result in inhibition of energy metabolism (Shanley *et al.*, 1988; Epstein *et al.*, 1989), and that most drugs exhibit regional selectivity in how they are accumulated in and metabolized by the kidneys.

The nephron, which is the major functional unit of the renal epithelium, is comprised of multiple and diverse cell types. Depending on the detail by which one classifies nephron cell types, at least 12 distinct cell populations are found: early and late proximal convoluted tubule, proximal straight tubule, descending and ascending thin limbs of the loop of Henle, medullary and cortical thick ascending limb of the loop of Henle, distal convoluted tubule, connecting tubule, cortical collecting tubule, and medullary and papillary collecting ducts. Each of these cell types possesses characteristic morphology, cellular energetics, transporters, and drug metabolism enzymes. The primary nephron cell type that is of interest for drug metabolism is the proximal tubule. Reasons for this include that this is the first cell population that becomes exposed to either filtered or blood-borne chemicals and that, in contrast to many other cell types in the kidney, such as distal tubules or medullary thick ascending limb cells, proximal tubules possess relatively high activities of plasma membrane transporters and intracellular drug metabolism enzymes.

The following four factors are, therefore, critical determinants of susceptibility to chemically induced nephrotoxicity: (i) high renal blood flow relative to organ weight delivers a disproportionately high amount of blood-borne chemicals to the kidneys; (ii) glomerular filtration, which concentrates chemicals in the tubular fluid; (iii) the presence of a large array of transporters on the basolateral and brush-border plasma membranes, particularly in proximal tubular epithelial cells, for organic anions and cations, amino acids, glucose, and other chemicals; this enables accumulation of many drugs and xenobiotics inside the cell; and (iv) once inside the cell (primarily proximal tubular epithelial cells), accumulated chemicals can be substrates for a wide array of both phase I and phase II drug metabolism enzymes, some of which can lead to bioactivation by formation of reactive intermediates. The last factor, that of bioactivation of accumulated drugs and xenobiotics, is the subject of this chapter.

## 2 INTERORGAN AND INTRA-RENAL BIOACTIVATION PATHWAYS

Taking into account renal blood flow, glomerular filtration, and the presence of several active transport proteins on basolateral and brush-border plasma membranes of proximal tubular epithelial cells, patterns of interorgan and intra-renal pathways for metabolism and transport of drugs and other xenobiotics that highlight all the key functions of the kidneys in general, and the proximal tubular epithelial cells in particular, are shown in Figs 1 and 2. For both drugs that are metabolized by the cytochrome P450 (CYP) system and the glutathione (GSH) conjugation pathway, most of the parent compound is metabolized in the liver. In many cases, the liver efficiently catalyzes efflux of the metabolites into bile or plasma and, through enterohepatic and renal-hepatic circulations, delivers a variety of metabolites to the kidneys. The ability of the liver to translocate chemicals to the circulatory system and that of the kidneys to concentrate

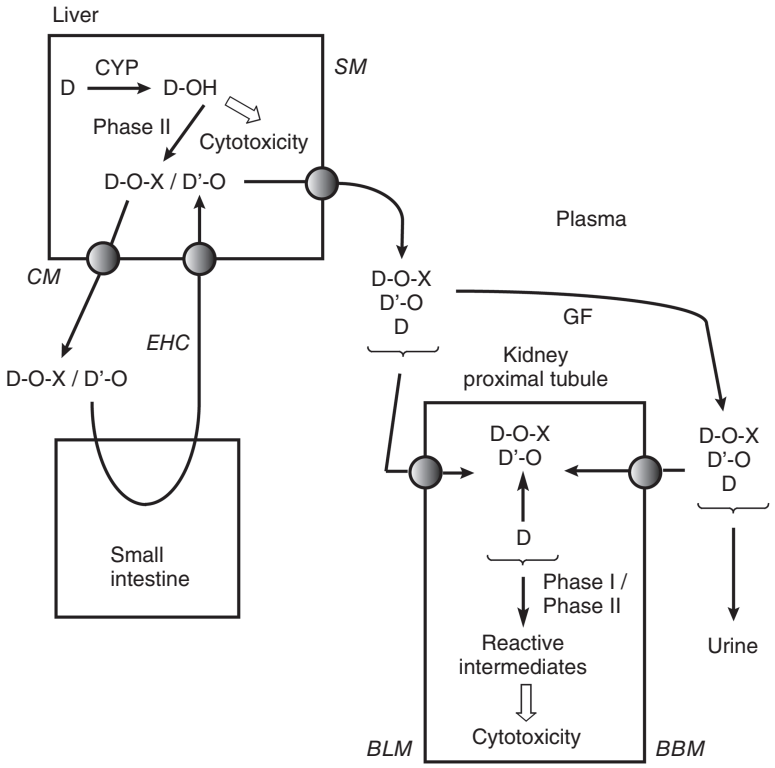


**Figure 1** Scheme of interorgan pathways involving cytochrome P450 (CYP). Metabolism and transport of drugs (D) in the liver, kidneys, and small-intestinal epithelium. Processes depicted in the liver include metabolism of a drug by CYP enzymes to yield the hydroxylated metabolite (D-OH), phase II metabolism to yield either a conjugate (D-O-X) or another species (D′-O). These metabolites can be exported into bile by transport across the canalicular plasma membrane (CM), where it may undergo additional metabolism in the intestines and then be returned to the liver via enterohepatic circulation (EHC). The various metabolites are then exported across the sinusoidal plasma membrane (SM) to be delivered to the kidneys, where the drugs may either undergo glomerular filtration (GF) or be absorbed from the renal periplasmic space by transport across the basolateral plasma membrane (BLM). Filtered drugs and their metabolites may either be excreted in the urine or reabsorbed by transport into the cell across the brush-border plasma membrane (BBM). Inside the proximal tubular cell, the drugs and their various metabolites may undergo various bioactivation reactions to yield reactive intermediates and cause cytotoxicity.

both blood-borne and filtered chemicals, play a major role in delivery of chemicals to bioactivation enzymes in the proximal tubular epithelium.

For drugs that are metabolized by CYP reactions (Fig. 1), the initial reaction occurs predominantly in liver, where it may either exert hepatotoxicity or be metabolized by a phase II conjugation reaction. The conjugate thus formed (glucuronide, sulfate, or GSH conjugate) is readily excreted into bile and eventually gets to the kidneys by enterohepatic and renal-hepatic circulations. The parent compound that is not absorbed and metabolized initially in the liver may be metabolized by renal CYPs, although this

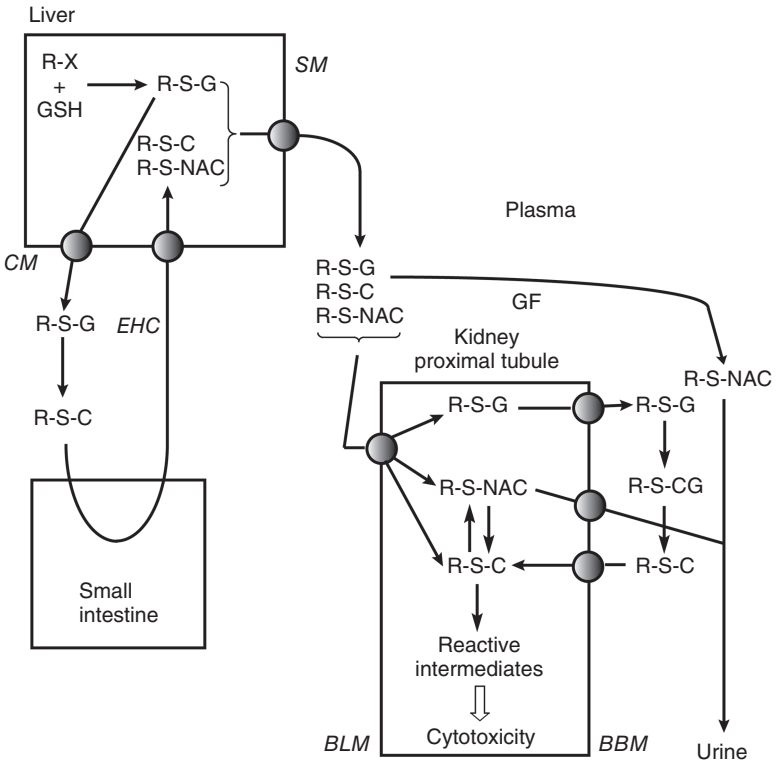


**Figure 2** Scheme of interorgan pathways involving GSH conjugation. Xenobiotics with good leaving groups (R-X) may undergo conjugation with GSH to form the GSH conjugate (R-S-G), which occurs primarily in the liver. The liver readily exports conjugates into bile by transport across the canalicular plasma membrane (CM), where they are largely metabolized to the corresponding cysteine conjugate (R-S-C) and mercapturate (R-S-NAC). Enterohepatic circulation (EHC) delivers these conjugates back to the liver, which then exports them into plasma by transport across the sinusoidal plasma membrane (SM) for delivery to the kidneys. The renal proximal tubular cell is exposed to the various conjugates via either glomerular filtration (GF), which delivers the conjugates to the tubular lumen, or via the renal periplasmic space, from which they can be transported into the cell by carriers on the basolateral plasma membrane (BLM). GSH conjugate in the lumen is subject to degradation, forming the cysteinylglycine conjugate (R-S-CG) and then R-S-C. The latter is transported into the cell by transport across the brush-border plasma membrane (BBM), where they can either be *N*-acetylated to form the mercapturate, or can be a substrate for bioactivation enzymes that generate reactive intermediates that cause cytotoxicity. R-S-NAC can either be secreted into the lumen and excreted in the urine or can be deacetylated in the cell to regenerate R-S-C.

occurs to a much smaller extent than in the liver. Various CYP-derived metabolites may elicit toxicity within the kidney or be excreted into the urine.

For drugs that are metabolized by the GSH conjugation pathway (Fig. 2), all pathways of metabolism, transport, and interorgan translocation eventually lead to the kidneys, regardless of whether the initial reaction occurs in the kidneys or in the liver. If the GSH conjugate is formed in the liver, it is primarily excreted into bile, where it

can be degraded by biliary or subsequently intestinal γ-glutamyltransferase (GGT) and dipeptidase (DP) to the corresponding cysteine conjugate. This cysteine conjugate is ultimately delivered to the kidneys by enterohepatic and renal-hepatic circulations. However, most of the degradation of GSH conjugate to the corresponding cysteine conjugate occurs in the kidneys as this is the predominant site of GGT in the body (Hinchman and Ballatori, 1990). Once the cysteine conjugate is formed in the kidneys, it has several potential fates, which will be discussed in detail in Section 4.

Regardless of whether the key metabolite is derived from CYP activity or GSH conjugation, these interorgan pathways illustrate the close interdependence of metabolism and transport processes in determining target organ specificity. As will be discussed later, similar activities that lead to bioactivation of cysteine conjugates and formation of reactive species in the kidneys are also present in the liver. No hepatotoxicity, however, is observed because of the efficiency of transport processes that mediate excretion of these metabolites from the liver and uptake by the kidneys.

## 3 CYTOCHROME P450 (CYP)-DEPENDENT BIOACTIVATION IN RENAL PROXIMAL TUBULAR CELLS

The basic biochemistry of the CYP monooxygenase system in the kidneys is essentially the same as that in the more studied liver, although important differences exist (Lock and Reed, 1997; Lohr, Willsky and Acara, 1998). First, amounts of CYP enzymes are generally only 5% to 20% of those in the liver. Second, CYP enzymes are not uniformly distributed throughout the cell types of the nephron. Rather, most of the CYP enzymes are found in the proximal tubular epithelium, although some are also found in other cell types, such as the glomerular epithelium or distal tubular cells (Guder and Ross, 1984; Lock and Reed, 1997; Lohr, Willsky and Acara, 1998). Third, the pattern of expression of CYP enzymes differs between the liver and kidney, with the liver expressing a more extensive array of individual enzymes. Finally, another tissue-specific difference is that for many CYP enzymes that are expressed in both liver and kidney, substrate specificity and inducibility differ (Anders, 1980; Jones, Orrenius and Jakobson, 1980; Lock and Reed, 1997; Lohr, Willsky and Acara, 1998; Ronis *et al.*, 1998). This has important implications when considering metabolism of a specific chemical in tissue samples from animals exposed either prior to or simultaneously with a CYP inducer.

When considering CYP-dependent metabolism in the kidneys, species-dependent differences are very prominent (Table 1). This has to be taken into account when extrapolating metabolism data from rodents to humans. For example, the environmental contaminant and "probable human carcinogen" trichloroethylene (TRI) (NTP, 2005) is metabolized by several CYP enzymes, although CYP2E1 is the primary enzyme that catalyzes TRI oxidation (Lash *et al.*, 2000). CYP-dependent metabolism of TRI in rat kidney occurs primarily by CYP2E1 and secondarily by CYP2C11 (Cummings, Parker and Lash, 2001), but is undetectable in human kidney (Cummings and Lash, 2000). Another factor that is important when considering renal CYP metabolism in humans is the existence of interindividual variability due to genetic polymorphisms (Lash, Putt and Cai, 2008).

As in the liver, the kidney contains four major families of CYPs, CYP1–CYP4. Although endogenous substrates exist for several enzymes, particularly for enzymes of the CYP2, CYP3, and CYP4 families, enzymes of the CYP1, CYP2, and CYP3 families are those that are involved in xenobiotic metabolism. A brief summary of the

**TABLE 1 Major cytochrome P450 (CYP) enzymes in rat and human kidney.**

| CYP enzyme | Rats | Humans |
|---|---|---|
| CYP1A1/2 | Present | Absent or low levels |
| CYP1B1/2 | Present | Present |
| CYP2B1/2 | Present | Absent |
| CYP2C11 (CYP2C19) | Present | Absent |
| CYP2D6 | Present | Present at low levels |
| CYP2E1 | Present | Absent |
| CYP3A1/2 (CYP3A4/5) | Present | Present |
| CYP4A2/3 (CYP4A11) | Present | Present |

CYP enzymes present in rats are listed with the human orthologues, when appropriate, listed in parentheses.

properties of enzymes of each CYP family, with an emphasis on species and tissue differences and their role in metabolism of endogenous as opposed to exogenous, or xenobiotic, chemicals will be presented later. In Sections 5 and 6, some examples of drugs and xenobiotics that are activated by these and other pathways will be given.

The CYP1 family consists of CYP1A1/2 and CYP1B1/2, with the former being that which is inducible by 3-methylcholanthrene, 2,3,7,8-tetrachlorodibenzo-*p*-dioxin (TCDD), and β-naphthoflavone. In the rat, CYP1A1 is only detected after exposure to inducers.

The CYP2 family is a highly diverse group of enzymes that metabolize a broad group of substrates, including endogenous compounds such as androgens and estrogens, and xenobiotics such as ethanol, halogenated solvents, and debrisoquine. Enzymes of the CYP2A family are expressed in mouse, but not in rat or human, kidney (Lock and Reed, 1997). CYP2B1/2 are inducible by phenobarbital in rat liver but not in rat kidney. In contrast, clofibrate induces expression of CYP2B1/2 in rat kidney (Cummings *et al.*, 1999) but cannot be detected in human kidney (Lash, Putt and Cai, 2008). Enzymes of the CYP2C family are generally considered to be constitutively expressed and exhibit sex specificity and developmental differences. For drug and xenobiotic metabolism, the major CYP2C enzyme of interest in rats is CYP2C11 (Cummings *et al.*, 1999; Cummings, Parker and Lash, 2001). There is no evidence for expression of the human orthologue CYP2C19 in kidney (Lash, Putt and Cai, 2008). Enzymes of the CYP2D family have not been readily detected in the kidney, although there is evidence that some family members are present in rat (Gonzalez, 1990) and CYP2D6 is present at low levels in human kidney microsomes (Lash, Putt and Cai, 2008). The CYP2E family, the most prominent of which is CYP2E1, metabolizes a large array of small chemicals, including ethanol, acetaminophen, halogenated solvents such as TRI and chloroform, and is inducible by ethanol, acetone, pyridine, pyrazole, and 4-methylpyrazole. Although CYP2E1 is expressed in rat kidney (Ronis *et al.*, 1998; Cummings *et al.*, 1999), its expression has not been detected in human kidney (Amet *et al.*, 1997; Cummings and Lash, 2000; Cummings, Lasker and Lash, 2000; Lash, Putt and Cai, 2008). In the rat and mouse kidney, CYP2E1 expression is under testosterone regulation and males have significantly higher levels than females. Accordingly, male mice are markedly more sensitive to the nephrotoxicity of certain CYP2E1 substrates than are females (Hu, Rhoten and Yang, 1990). This is a clear example of where extrapolation of metabolism and toxicity data from rodents to humans is not possible because of qualitative differences in metabolism.

The CYP3 gene family, of which CYP3A1 and CYP3A2 are the major enzymes in rats and CYP3A4 and CYP3A5 are the major orthologous enzymes in humans, is prominent in the liver, is inducible by glucocorticoids, and is involved in steroid hydroxylation and metabolism of macrolide antibiotics such as cyclosporine. CYP3A1/2 was detected in microsomes from rat kidney cortex homogenates (Schuetz *et al.*, 1992; Bebri *et al.*, 1995; Cummings *et al.*, 1999) and CYP3A4/5 was detected in microsomes from human kidney cortex homogenates (Schuetz *et al.*, 1992; Lash, Putt and Cai, 2008). CYP3A1/2 protein expression, however, was not detected in microsomes from renal proximal tubular cells (Cummings *et al.*, 1999), suggesting that these CYPs may be localized primarily in other regions of the nephron besides the proximal tubule, such as the glomerulus. In human kidney microsomes, CYP3A4/5 were readily detected in some, but not all samples, consistent with the presence of genetic polymorphisms, which is well known for the liver (Lash, Putt and Cai, 2008).

Enzymes of the CYP4A family are primarily involved in metabolism of fatty acids and arachidonic acid, are prominently expressed in the kidneys (Okita *et al.*, 1997; Ito *et al.*, 1998; Stec *et al.*, 2003), and are important in renal physiology, because of the generation of metabolites such as hydroxyeicosatetraenoic acid that influence hemodynamics. CYP4A enzymes are induced by hypolipidemic drugs, such as the fibrates, and other chemicals that cause peroxisome proliferation. These inducing effects, however, are more prominent in the liver than in the kidney and in rodents than in humans (Goldsworthy and Popp, 1987). The major CYP4A enzymes in rat kidney are CYP4A2 and CYP4A3. They hydroxylate fatty acids such as lauric acid and catalyze ω- and ω-1 hydroxylation of arachidonic acid, generating metabolites that, as noted earlier, are important in the regulation of blood pressure and renal hemodynamics. The major CYP4A enzyme in human kidney is CYP4A11, which also metabolizes lauric and arachidonic acids. CYP4A11 is localized exclusively in the S2 and S3 segments of the proximal tubule (Lasker *et al.*, 2000) and its expression in primary cultures of human proximal tubular cells is induced by ethanol and dexamethasone (Cummings, Lasker and Lash, 2000). CYP4A11 is considered the human orthologue of rat CYP4A2.

## 4 FLAVIN-CONTAINING MONOOXYGENASE (FMO)-DEPENDENT BIOACTIVATION IN RENAL PROXIMAL TUBULAR CELLS

FMOs are a multigene family of microsomal enzymes that catalyze oxidation of sulfur-, selenium-, and nitrogen-containing drugs and xenobiotics (Ziegler, 1993; Cashman and Zhang, 2006). Five active isoforms (FMO1–5) have been identified and are present in most mammalian tissues, although species-, sex-, tissue-, and developmental-dependent differences exist in expression of these isoforms (Hines *et al.*, 1994; Dolphin *et al.*, 1998; Ripp *et al.*, 1999; Koukouritaki *et al.*, 2002). Like the CYPs, FMOs are localized in the endoplasmic reticulum (microsomes), catalyze oxidation of a broad range of xenobiotics as well as endogenous compounds, and are most prominently expressed in the liver, although extrahepatic expression occurs.

Primary cultures of human proximal tubular cells express only three isoforms of FMO: FMO1, FMO3, and FMO5 (Krause, Lash and Elfarra, 2003). In contrast to human liver, which expresses primarily FMO3 as well as several other FMO enzymes at significant levels, human kidney expresses primarily FMO1, somewhat lower levels of protein for FMO5, and very low levels of protein for FMO3 (Krause, Lash and Elfarra, 2003). Nishimura and Naito (2006) determined mRNA expression profiles of

FMOs in human kidney, and found that by far the major mRNA expressed was that of FMO1; mRNAs for FMO2, FMO3, FMO4, and FMO5 were detected at 4.0%, 0.09%, 24.7%, and 12.6%, respectively, of the levels found for FMO1. Additionally, FMO1 protein expression levels varied significantly among human kidney samples, suggesting the presence of polymorphisms (Krause, Lash and Elfarra, 2003).

Although cysteine conjugate β-lyase (CCBL) was the first discovered enzymatic activity of the GSH-dependent bioactivation pathway that acts on cysteine conjugates (see Section 5), is the primary catalyst of cysteine conjugate bioactivation in many systems, and is the best and most frequently studied reaction, it became increasingly evident around 1990 that other metabolites could be formed and that additional enzymes besides the CCBL were involved in their formation. Sulfoxides have long been known as metabolites of many cysteine conjugates and are often recovered in the urine of animals treated with cysteine conjugates or their precursors, as stable end products. Elfarra and colleagues identified and subsequently purified an enzymatic activity from rat liver and kidney microsomes that metabolizes several cysteine conjugates, including *S*-(1,2-dichlorovinyl)-L-cysteine (DCVC), to sulfoxides (Sausen and Elfarra, 1990; Sausen, Duescher and Elfarra, 1993). Early evidence from these studies implicating FMOs included inhibition of *S*-oxidation of cysteine conjugates by other FMO substrates (e.g. *N,N*-dimethylaniline, methimazole, and *n*-octylamine), NADPH dependence, restoration of activity of flavin-depleted microsomes by exogenous flavin adenine dinucleotide (FAD) or flavin mononucleotide (FMN), and immunoreactivity of purified *S*-oxidase with antibodies to FMOs.

The role of FMOs in xenobiotic bioactivation in the kidneys is not well characterized, with the exception of its role in the metabolism of certain nephrotoxic cysteine conjugates to reactive sulfoxides (Lash *et al.*, 1994, 2003; Cummings and Lash, 2000; Elfarra and Krause, 2007). As described in Section 5, the cysteine conjugate of TRI, DCVC, is a substrate for FMOs, which produce the reactive DCVC sulfoxide that can then react with GSH and other cellular nucleophiles to produce cytotoxicity (Sausen and Elfarra, 1991). Incubations of human cDNA-expressed FMO1, FMO3, FMO4, and FMO5 with DCVC showed that DCVC sulfoxide was only detected with FMO3 (Krause, Lash and Elfarra, 2003). The observation that FMO1, rather than FMO3, is the major isoform present in rat and human proximal tubular cells raises the question of the importance of this pathway in DCVC bioactivation. In support of a role for FMO in DCVC-induced nephrotoxicity, however, is the observation of significant protection by pretreatment with the FMO substrate methimazole (Lash *et al.*, 1994, 2003; Cummings and Lash, 2000). Moreover, DCVC sulfoxide is a more potent nephrotoxicant *in vivo* and cytotoxicant *in vitro* than DCVC (Lash *et al.*, 1994, 2003). Further studies are needed to resolve this uncertainty about the relative importance of FMO-dependent bioactivation.

## 5 RENAL BIOACTIVATION OF CYSTEINE CONJUGATES: GSH CONJUGATION AS A BIOACTIVATION RATHER THAN A DETOXICATION PATHWAY

### 5.1 GSH Conjugation

GSH conjugation of reactive electrophiles is classically considered the first step in the mercapturic acid pathway, which is a prominent detoxification mechanism for many xenobiotics. With studies beginning in the 1960s, however, it became apparent that several classes of chemicals, chief among them halogenated alkanes and alkenes, are

bioactivated by GSH conjugation. Although there are some reports of toxicity in other target organs, by and large toxicity from xenobiotics that are bioactivated by the GSH conjugation pathway is restricted to the kidneys. Factors responsible for this target-tissue specificity are those that are discussed in Section 2, and include the discrete tissue localization of some of the enzymes of the GSH conjugation pathway and of membrane transporters for efflux and uptake of various metabolic intermediates in the pathway.

Multiple families of glutathione *S*-transferases (GSTs) exist in mammals. GSTs can be found primarily in three subcellular compartments: cytoplasm, mitochondria, and endoplasmic reticulum (or microsomes). The isoforms that are important for xenobiotic metabolism in the kidneys are those found in the cytoplasm and microsomes. All cytoplasmic GSTs are dimeric with subunits of 199 to 244 amino acids in length. Based on amino acid sequence, seven classes of cytoplasmic GSTs exist in mammals, and are designated Alpha, Mu, Pi, Sigma, Theta, Omega, and Zeta. In rodents and humans, GST classes are defined as those enzymes that share >40% amino acid sequence identity. The convention used is to refer to rodent GSTs by Greek letters (i.e. GSTα, μ, π, σ, τ, ϕ, and ζ) and human GSTs by capital Arabic letters (i.e. GSTA, M, P, S, T, O, and Z). At least 16 cytoplasmic GST subunits exist in humans, many of which exist as heterodimers. Consequently, a large variety of isoenzymes can be generated from the various subunits.

In the rat kidney, three families of cytoplasmic GSTs have been studied in some detail and have been found to be expressed: GSTα, GSTπ, and GSTμ. Immunolocalization studies (Rozzel *et al.*, 1993) and Western blot analyses in renal cortical microsomes and isolated renal proximal tubular and distal tubular cells (Cummings, Parker and Lash, 2000) showed that only GSTα is expressed in the proximal tubules whereas GSTπ and GSTμ are expressed in the distal nephron. In contrast to this distribution of GST isoforms, human renal proximal tubular cells express GSTA, GSTP, and GSTM (Cummings, Lasker and Lash, 2000; Lash, Putt and Cai, 2008). Inasmuch as the various GST isoforms have broad and often overlapping substrate specificities, it is difficult to assess the impact of this difference in GST isoform expression pattern on extrapolation of renal metabolism data from rats to humans. Suffice it to say, however, the markedly distinct profiles of GST isoform expression will have an impact on substrate specificities and relative activities.

Besides the more commonly studied families of cytoplasmic GSTs, the Zeta class was recently discovered by a bioinformatics approach using human expressed sequence tag databases (Board *et al.*, 1997, 2001; Board and Anders, 2005). GSTZs are widely distributed in eukaryotes and were shown to be identical to maleylacetoacetate isomerase, which catalyzes the penultimate step in tyrosine catabolism. The substrate specificity pattern of GSTZ differs from the other GST families in lacking significant activity toward the prototypical substrate 1-chloro-2,4-dinitrobenzene. Although GSTZ exhibits some activity with organic peroxides, it plays a critical role in the metabolism of various α-haloacids, including dichloroacetic acid (DCA), which is metabolized to glyoxylic acid. Biotransformation of DCA is important because DCA is used in the clinical management of congenital lactic acidosis, is a common drinking water contaminant, and is a metabolite of the environmental contaminant and probable human carcinogen TRI (Lash *et al.*, 2000). At least four polymorphic variants of GSTZ are known in humans, each exhibiting markedly distinct catalytic activities.

The MAPEG (membrane-associated proteins in eicosanoid and glutathione metabolism) represent a unique family of GSTs that share no sequence identity with either

the cytoplasmic or mitochondrial GSTs. These proteins are found in the endoplasmic reticulum or microsomes and have the distinct property of being involved in the production of eicosanoids. Four subgroups of MAPEG are known, with proteins within each subgroup sharing at least 20% sequence identity. In humans, six MAPEGs have been identified, which fall into subgroups I, II, and IV (Jakobsson, Morgenstern and Mancini, 1999). Two of the proteins are critical for leukotriene biosynthesis, and three are cytoprotective, exhibiting GSH *S*-transferase and peroxidase activities. Evidence suggests that MGST1 (microsomal GSH *S*-transferase 1) functions solely as a detoxication enzyme, whereas human MGST2 and MGST3 can both detoxify foreign compounds and synthesize leukotriene $C_4$. MGSTs have been immunolocalized to several rat tissues, including the kidneys (Otieno *et al.*, 1997). The role of MAPEG in bioactivation of chemicals that ultimately become cysteine conjugate substrates for other bioactivation enzymes, such as the CCBL or FMO, is less clear, although TRI is metabolized to DCVG in the presence of GSH and liver or kidney microsomal fractions from rats or humans (Lash *et al.*, 1995, 1998, 1999).

### 5.2 γ-Glutamyltransferase and Cysteinylglycine Dipeptidase Activities

Whether the GSH conjugate is formed in the liver or in the kidneys, the kidneys are the primary sites for its further processing because of the interorgan translocation pathways described in Section 2. The next two steps in the metabolism of GSH *S*-conjugates involve successive cleavage of the γ-glutamyl isopeptide and cysteinylglycine peptide bonds, to release the corresponding cysteine *S*-conjugate as the product. The first step, catalyzed by γ-glutamyltransferase (GGT), is found on the luminal membrane of several epithelial tissues, although most prominently in the renal proximal tubule (Hinchman and Ballatori, 1990). Hence, although GGT activity is found on extrarenal sites such as the hepatic canalicular plasma membrane and the jejunal brush-border plasma membrane, its extremely high activity on the renal proximal tubular brush-border plasma membrane coupled with the high activity of plasma membrane transporters and the glomerular filtration process, serves to deliver GSH conjugates primarily to the kidneys (cf. Fig. 2). By these mechanisms, most of the whole-body turnover of GSH and GSH conjugates can be accounted for by renal metabolism. The importance of GGT and subsequent DP activity in the turnover of GSH (and by analogy, GSH *S*-conjugates) is highlighted by the profound glutathionuria that occurs when GGT activity is inhibited (Griffith and Meister, 1979).

### 5.3 Cysteine Conjugate β-Lyase Activities

Metabolism to the cysteine conjugate represents a branch point because alternative reactions can lead to either detoxication and excretion or bioactivation and generation of reactive intermediates. Multiple enzymes can bioactivate those cysteine conjugates whose properties make them substrates for these enzymes in preference to the microsomal *N*-acetyltransferase that forms mercapturates. This concept is illustrated in Fig. 3. After processing of the GSH conjugate by the actions of GGT and DP activities, the cysteine conjugate thus formed can be *N*-acetylated by the cysteine conjugate *N*-acetyltransferase (NAT) to yield the mercapturate. Mercapturates are highly polar and are generally readily excreted in urine. However, they may also be acted on by a deacetylase (or aminoacylase) to regenerate the cysteine conjugate. As a consequence of this activity, many *N*-acetyl-L-cysteine-*S*-conjugates of nephrotoxic haloalkenes or haloalkanes

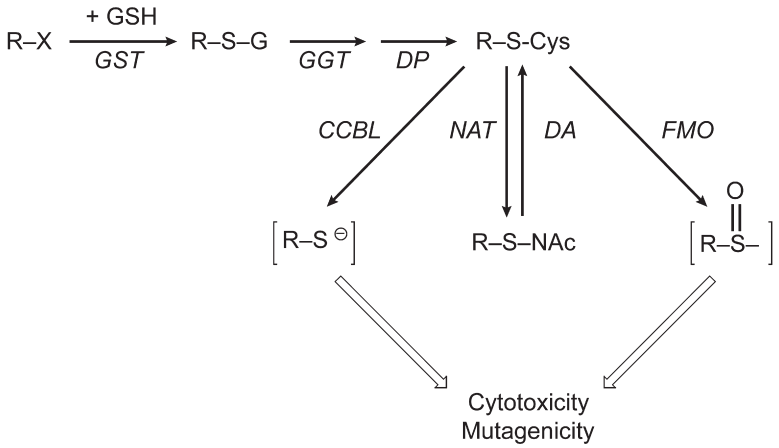


**Figure 3** Bioactivation of GSH conjugates. Drug/xenobiotic (R-X) forms a GSH conjugate (R-S-G) that undergoes degradation by γ-glutamyltransferase (GGT) and dipeptidase (DP) to yield the cysteine conjugate (R-S-Cys). R-S-Cys functions as a branch point in that it is subject to either detoxication by action of the *N*-acetyltransferase (NAT) to form the mercapturate (R-S-NAc) or bioactivation by the action of either the cysteine conjugate β-lyase (CCBL) or flavin-containing monooxygenase (FMO), which generate reactive thiolates and sulfoxides, respectively. Mercapturates may also regenerate the cysteine conjugate by deacetylation, catalyzed by a deacetylase (DA) or aminoacylase.

may be deacetylated and exhibit toxicity (Wolfgang *et al.*, 1989; Zhang and Stevens, 1989; Birner *et al.*, 1998). Alternatively, certain cysteine conjugates can be substrates of either the CCBL or FMO, which generate reactive and chemically unstable metabolites.

One of the most studied activities for bioactivation of cysteine conjugates is the CCBL. The most recently published review by Cooper and Pinto (2006) highlights the multiplicity of enzymes that can catalyze this overall reaction. The various mammalian enzymes that have been identified as possessing CCBL activity in both the cytoplasm and mitochondria are listed in Table 2. The various CCBL activities are all pyridoxal 5′-phosphate (PLP)-containing enzymes and can catalyze either a direct β-elimination reaction to yield a reactive thiolate, or, in some cases, a transamination reaction. The transamination reaction generally yields an unstable α-keto acid that spontaneously rearranges to yield the reactive thiolate. In the kidneys, there are two known CCBL enzymes in the cytoplasm and two in the mitochondria, each capable of catalyzing both β-elimination and transamination reactions. The high-molecular-weight forms in renal cytoplasm and mitochondria have not been extensively characterized, but it has been suggested that at least in the mitochondria, this form is the primary enzyme catalyzing CCBL activity (Abraham, Patel and Cooper, 1995; Abraham, Thomas and Cooper 1995).

One consequence of the ability of the various renal CCBL enzymes to catalyze transamination reactions is that this reaction can compete with the β-elimination reaction, so that maximal CCBL activity requires the presence of an α-keto acid in the reaction mixture (Stevens, Robbins and Byrd, 1986; Elfarra, Lash and Anders, 1987). In the absence of an α-keto acid, therefore, the PLP prosthetic group of the CCBL

**TABLE 2 Mammalian enzymes that catalyze cysteine conjugate β-lyase activity.**

| Enzyme | Tissue | Reaction | Subunit MW |
|---|---|---|---|
| Cytoplasmic enzymes | | | |
| GTK (EC 2.6.1.64)/KAT (EC 2.6.1.7) | Kidney, choroid plexus | BE, TA | 45 kDa; homodimer |
| Kynureninase (EC 3.7.1.3) | Liver | BE | 55 kDa; homodimer |
| cytASAT (EC 2.6.1.1) | Heart | BE | 45 kDa; homodimer |
| ALAT (EC 2.6.1.2) | Heart | BE | 45 kDa; homodimer |
| $BCAT_c$ (EC 2.6.1.42) | Brain | BE | 44 kDa; homodimer |
| High-MW β-lyase | Liver, kidney | BE, TA | 330 kDa |
| Mitochondrial enzymes | | | |
| mtASAT (EC 2.6.1.1) | Liver | BE, TA | 45 kDa; homodimer |
| $BCAT_m$ (EC 2.6.1.42) | Brain | BE | 44 kDa; homodimer |
| GTK (EC 2.6.1.64)/KAT (EC 2.6.1.7) | Brain, kidney | BE, TA | 45 kDa; homodimer |
| High-MW β-lyase | Liver, kidney | BE, TA | 330 kDa |

ALAT, alanine aminotransferase; $BCAT_{c/m}$, branched-chain aminotransferase (cytoplasm/mitochondria); BE, beta-elimination; cyt/mitASAT, cytosolic and mitochondrial aspartate aminotransferase; GTK, glutamine transaminase K; KAT, kynurenine aminotransferase; TA, transamination.

becomes trapped in the pyridoxamine 5′-phosphate (PMP) form and activity is very low or non-detectable.

The most commonly studied nephrotoxic cysteine conjugates are those of halogenated alkanes and alkenes (Fig. 4). Each of the cysteine conjugates has in common the presence of a good leaving group for the CCBL reaction. The haloalkenyl conjugates are all derived from environmental contaminants that are of interest for human health risk assessment because of their prevalence in the environment, their widespread use in industry, and their known carcinogenicity in animals and "probable" carcinogenicity in humans (e.g. Lash, Parker and Scott, 2000; Lash and Parker, 2001; NTP, 2005). The examples shown here are the cysteine *S*-conjugates of trichloroethylene (DCVC), perchloroethylene (TCVC), 1,1,2,2-tetrafluoroethylene (TFEC), and 1-chloro-1,2,2-trifluoroethylene (CTFC). The haloalkenyl cysteine conjugates (DCVC, TCVC), when metabolized by the CCBL, generate reactive thiol-containing species that can cause both cytotoxicity and mutagenicity. In contrast, the haloalkyl cysteine conjugates (TFEC, CTFC) that contain only fluorines or chlorines are cytotoxic but not mutagenic, whereas those containing a bromine atom are also mutagenic (Finkelstein *et al.*, 1994). These other conjugates include several environmentally and clinically important compounds, such as chlorofluorocarbons that have been used as refrigerants (Yin, Jones and Anders, 1995) and degradation products of anesthetic agents such as sevoflurane (Iyer *et al.*, 1998; Altuntas and Kharasch, 2002; Anders, 2005).

Besides these types of halogenated solvent substrates, studies by Hanigan and colleagues (Hanigan *et al.*, 1994, 2001; Townsend and Hanigan, 2002; Townsend *et al.*, 2003a, b; Zhang and Hanigan, 2003) demonstrated that the mono-platinum-monoglutathione conjugate is a substrate for GGT, that the cysteine conjugate of cisplatin is a substrate for CCBL activities, and that nephrotoxicity due to these conjugates in mice and cytotoxicity in various *in vitro* proximal tubular cell models is variously inhibited by inhibitors of GGT and/or the CCBL (see further discussion).

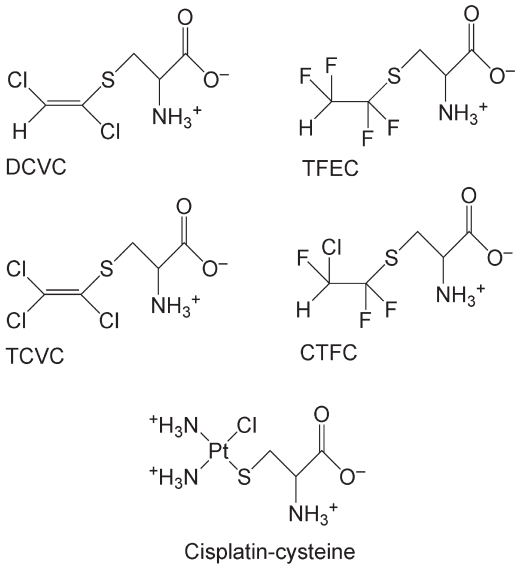


**Figure 4** Structures of representative cysteine conjugates that undergo renal bioactivation. CTFC, *S*-(1-chloro-1,2,2-trifluoroethyl)-L-cysteine; DCVC, *S*-(1,2-dichlorovinyl)-L-cysteine; TFEC, *S*-(1,1,2,2-tetrafluoroethyl)-L-cysteine; TCVC, *S*-(1,2,2-trichlorovinyl)-L-cysteine.

### 5.4 Validation of GSH-Dependent Bioactivation Pathways

The critical importance of the CCBL activity in determining the metabolism and toxicity of various haloalkyl and haloalkenyl cysteine conjugates has been demonstrated in whole animals and in various *in vitro* preparations of renal proximal tubular cells or renal subcellular fractions (Fig. 5). Thus, inhibition of the CCBL activity with aminooxyacetic acid (AOAA) prevents some or all of the toxicity associated with the cysteine conjugate. Addition of appropriate α-keto acids to maximize CCBL activity also enhances cysteine conjugate-induced cytotoxicity. For those cysteine conjugates that are also substrates for FMO, alternative FMO substrates such as methimazole are protective both *in vivo* and *in vitro*, as discussed earlier. The scheme also illustrates how the toxicity of conjugates at earlier steps in the pathway can be modulated by the use of inhibitors or co-substrates. Thus, acivicin, which is an irreversible inhibitor of GGT, prevents the toxicity of GSH conjugates and addition of glycylglycine (GlyGly) or another suitable γ-glutamyl group acceptor, can increase GGT activity and enhance the toxicity of GSH conjugates. Addition of either 1,10-phenanthroline or phenylalanylglycine (PheGly), which is an irreversible or competitive inhibitor of the dipeptidase activity, respectively, protects from the toxicity of both GSH and cysteinylglycine conjugates. Finally, several studies have demonstrated, in both *in vivo* studies and in isolated proximal tubular cell preparations, that probenecid can protect against the cytotoxicity of either GSH or cysteine conjugates, implicating function of the basolateral organic anion transporters in the uptake and accumulation of these conjugates in proximal tubular cells. Hence, modulation of both transport and metabolism can result in alteration of conjugate nephrotoxicity *in vivo* and cytotoxicity *in vitro*.

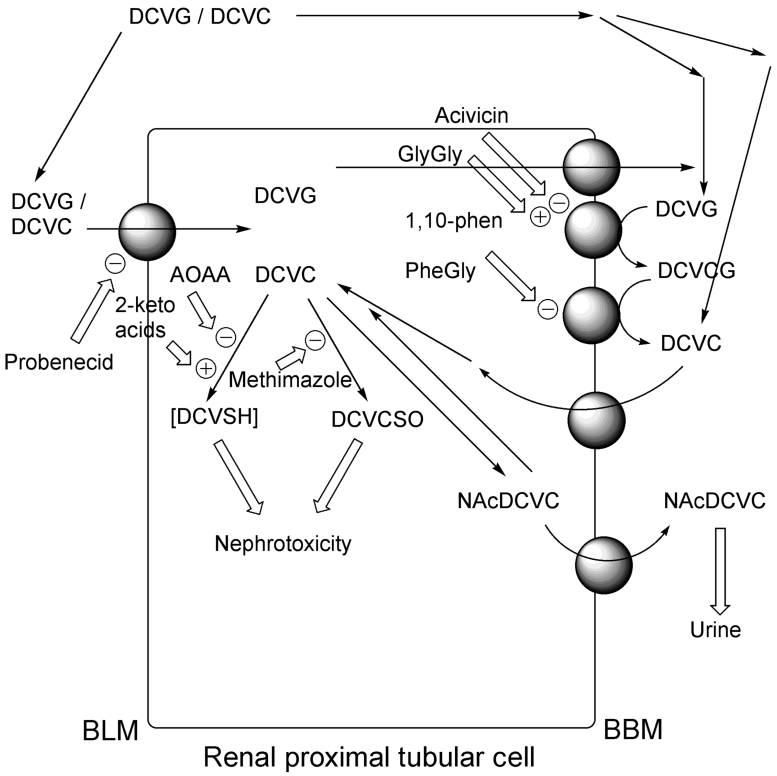


**Figure 5** Validation of role of enzymatic and transport processes in the renal proximal tubule producing nephrotoxicity. Using the GSH-derived conjugates of the environmental contaminant trichloroethylene as an example, transport and metabolism of the various conjugates are illustrated. The GSH conjugate *S*-(1,2-dichlorovinyl)glutathione (DCVG) is either transported across the basolateral plasma membrane (BLM) by probenecid-sensitive organic anion transporters or undergoes glomerular filtration, which delivers it to the lumen for metabolism by γ-glutamyltransferase (GGT) to the cysteinylglycine conjugate *S*-(1,2-dichlorovinyl)-L-cysteinylglycine (DCVCG) and the dipeptidase (DP) to the cysteine conjugate *S*-(1,2-dichlorovinyl)-L-cysteine (DCVC). DCVC enters the renal proximal tubular cell by either transport across the BLM by probenecid-sensitive organic anion or amino acid transporters or transport across the brush-border plasma membrane (BBM). GGT and DP activities can be modulated by acivicin or glycylglycine (GlyGly) and 1,10-phenanthroline (1,10-phen) or phenylalanylglycine (PheGly), respectively. Once inside the proximal tubular cell, DCVC can be *N*-acetylated to form the mercapturate *S*-(1,2-dichlorovinyl)-*N*-acetyl-L-cysteine (NAcDCVC), which can either be excreted in the urine or deacetylated to regenerate DCVC. DCVC can also be a substrate for cysteine conjugate β-lyase (CCBL) or flavin-containing monooxygenase (FMO) activities to generate reactive 1,2-dichlorovinylthiol (DCVSH) or DCVC sulfoxide (DCVCSO), respectively. CCBL activity can be modulated by aminooxyacetic acid (AOAA) or 2-keto acids, and FMO activity can be modulated by co-substrates such as methimazole. Correspondence between inhibition and stimulation of each transport or metabolic step and nephrotoxicity *in vivo* or cytotoxicity *in vitro* validated these pathways. –, inhibition; +, stimulation.

## 6 RENAL BIOACTIVATION OF ACETAMINOPHEN

Although acetaminophen (APAP) is a commonly used analgesic and is considered safe to use under appropriate conditions, it represents a significant drug of concern for public health as it is a frequent cause of poisonings due to overdose. Whereas the liver is typically the most prominent target of toxicity in APAP overdose, the kidneys can also significantly metabolize APAP to both detoxication and bioactivation products (Newton *et al.*, 1982), which can result in analgesic nephropathy (Duggin, 1980). The overall metabolic pathway for APAP, which occurs in both liver and kidneys (Fig. 6), illustrates the presence of several competing pathways for both detoxication and bioactivation. Quantitatively, the predominant pathways are either glucuronidation or sulfation, which lead to highly polar metabolites that are readily excreted. A lower activity pathway is mediated by CYP (predominantly CYP2E1 but also CYP1A2). CYP-dependent metabolism produces a reactive intermediate, *N*-acetyl-*p*-benzoquinoneimine (NAPQI), which readily reacts with nucleophiles such as protein sulfhydryl groups and GSH. GSH conjugation of APAP or its deacetylated congener *p*-aminophenol (PAP) has usually been viewed as a detoxication route, whereby the GSH conjugates are processed to mercapturates, which are readily excreted in the

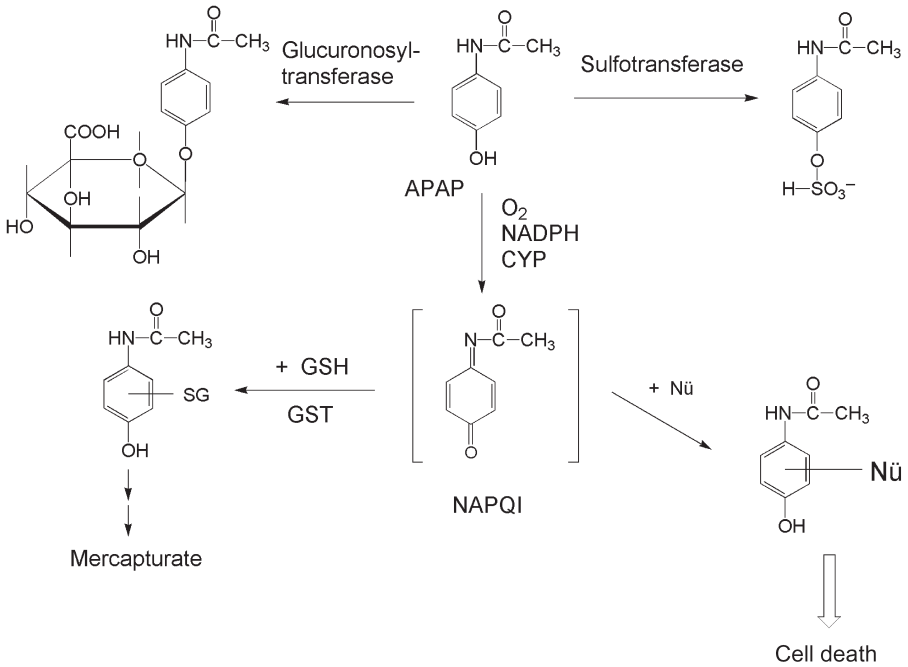


**Figure 6** Renal metabolism of acetaminophen (APAP). Acetaminophen may undergo either glucuronidation, sulfation, or oxidation by cytochrome P450 (CYP). The latter reaction yields a reactive intermediate, *N*-acetyl-*p*-benzoquinoneimine (NAPQI). NAPQI reacts with either GSH, as catalyzed by GSH *S*-transferases (GSTs) to ultimately yield a mercapturate, or other nucleophiles (Nü). The latter reaction may lead to cytotoxicity.

urine. Several studies, however, have found that GSH conjugation of APAP or PAP leads to nephrotoxicity in rats or mice that could be prevented by pre-incubation with acivicin or probenecid (Klos *et al.*, 1992; Emeigh Hart *et al.*, 1996; Trumper, Monasterolo and Elias, 1996). While these findings are similar to those with conjugates of chemicals such as TRI that are ultimately bioactivated by CCBL or FMO, there is no evidence that these latter two bioactivation enzymes are involved. Nonetheless, the findings suggest that GGT-dependent metabolism is required to elicit nephrotoxicity for APAP or PAP.

## 7 SUMMARY AND CONCLUSIONS

The kidneys express a diverse array of enzymes that can catalyze the bioactivation of many drugs and xenobiotics. Unlike the liver, which is comprised predominantly of hepatocytes that express high levels of many phase I and phase II enzymes, the kidneys are comprised of numerous cell types that differ with respect to morphology, physiological functions, and expression of drug metabolism enzymes and membrane transporters, among other factors. Thus, most of the drug metabolism enzymes that are critical for bioactivation are found in the proximal tubular cells. One implication of this selective localization along the nephron is that when one looks to quantify many drug metabolism enzymes in renal homogenates, the apparent activity may be low because of dilution. In other words, the enzyme of interest may only be present in a few discrete cell types and the homogenate will likely include those cell types as well as several others that either do not express the enzyme or express it at low levels. Nonetheless, drug metabolism capability for the kidneys is typically much lower than that for the liver, even when cell type distribution is taken into account.

The kidneys are particularly susceptible to chemically induced toxicity because of their basic physiology and biochemistry. In terms of physiology, the high rate of renal blood flow delivers chemicals to the kidneys where these chemicals can undergo glomerular filtration or absorption from the periplasmic space by uptake across the basolateral membrane. The high flow rate of renal blood flow and glomerular filtration serve to deliver blood-borne chemicals to plasma membrane transporters that mediate their uptake and intracellular accumulation. Once inside the cell (primarily the proximal tubular cell), bioactivation enzymes similar to those found in hepatocytes are present to convert drugs to reactive and toxic species.

This chapter reviewed three major bioactivation pathways in the kidneys: the CYP, FMO, and GSH conjugation pathways. Many of the same CYP enzymes that are found in the liver are also found in the kidneys, although predominantly in the proximal tubules. Even with the same enzymes, some differences exist between the liver and kidney with respect to substrate specificity and inducibility. Significant species differences also exist for the CYPs, making conclusions made in studies with rodents not always readily applicable to humans. The GSH conjugation pathway initially leads to generation of cysteine conjugates by processing of the GSH conjugates by GGT and DP activities on the brush-border membrane of the renal proximal tubule. The cysteine conjugates thus formed serve as a branch point for several types of GST substrates, including halogenated solvents and the chemotherapeutic agent cisplatin, as they can either be metabolized to mercapturates and excreted or bioactivated by CCBL or FMO. The final example, that of APAP, illustrates the interaction and competition between alternate pathways for the given chemical.

## REFERENCES


Abraham, D.G., Patel, P.P., Cooper, A.J.L. (1995). Isolation from rat kidney of a cytosolic high molecular weight cysteine-*S*-conjugate β-lyase with activity toward leukotriene $E_4$. *J. Biol. Chem.*, *270*, 180–188.

Abraham, D.G., Thomas, R.J., Cooper, A.J.L. (1995). Glutamine transaminase K is not a major cysteine *S*-conjugate β-lyase of rat kidney mitochondria: evidence that a high-molecular weight enzyme fulfills this role. *Mol. Pharmacol.*, *48*, 855–860.

Altuntas, T.G., Kharasch, E.D. (2002). Biotransformation of L-cysteine *S*-conjugates and *N*-acetyl-L-cysteine *S*-conjugates of the sevoflurane degradation product fluoromethyl-2,2-difluoro-1-(trifluoromethyl)vinyl ether (Compound A) in human kidney in vitro: interindividual variability in *N*-acetylation, N-deacetylation, and β-lyase-catalyzed metabolism. *Drug Metab. Dispos.*, *30*, 148–154.

Amet, Y., Berthou, F., Fournier, G., Dreano, Y., Bardou, L., Cledes, J., Menez, J.-F. (1997). Cytochrome P-450 4A and 2E1 expression in human kidney microsomes. *Biochem. Pharmacol.*, *53*, 765–771.

Anders, M.W. (1980). Metabolism of drugs by the kidney. *Kidney Int.*, *18*, 636–647.

Anders, M.W. (2005). Formation and toxicity of anesthetic degradation products. *Annu. Rev. Pharmacol. Toxicol.*, *45*, 147–176.

Bebri, K., Boobis, A.R., Davies, D.S., Edward, R.J. (1995). Distribution and induction of CYP3A1 and CY3A2 in rat liver and extrahepatic tissue. *Biochem. Pharmacol.*, *50*, 2047–2056.

Birner, G., Werner, M., Rosner, E., Mehler, C., Dekant, W. (1998). Biotransformation, excretion, and nephrotoxicity of the hexachlorobutadiene metabolite (E)-N-acetyl-S-(1,2,3,4,4-pentachlorobutadienyl)-L-cysteine sulfoxide. *Chem. Res. Toxicol.*, *11*, 750–757.

Board, P.G., Anders, M.W. (2005). Human glutathione transferase-Zeta. *Methods Enzymol.*, *401*, 61–77.

Board, P.G., Baker, R.T., Chelvanayagam, G., Jermiin, L.S. (1997). Zeta, a novel class of glutathione transferases in a range of species from plants to humans. *Biochem. J.*, *328*, 929–935.

Board, P.G., Chelvanayagam, G., Jermiin, L.S., Tetlow, N., Tzeng, H.F., Anders, M.W., Blackburn, A.C. (2001). Identification of novel glutathione transferases and polymorphic variants by expressed sequence tag database analysis. *Drug Metab. Dispos.*, *29*, 544–547.

Brenner, B.M. (1996). *The Kidney*, 5th edn, W.B. Saunders Company, Philadelphia.

Cashman, J.R., Zhang, J. (2006). Human flavin-containing monooxygenases. *Annu. Rev. Pharmacol. Toxicol.*, *46*, 65–100.

Cooper, A.J.L., Pinto, J.T. (2006). Cysteine *S*-conjugate β-lyases. *Amino Acids*, *30*, 1–15.

Cummings, B.S., Lash, L.H. (2000). Metabolism and toxicity of trichloroethylene and S-(1,2-dichlorovinyl)-L-cysteine in freshly isolated human proximal tubular cells. *Toxicol. Sci.*, *53*, 458–466.

Cummings, B.S., Lasker, J.M., Lash, L.H. (2000). Expression of glutathione-dependent enzymes and cytochrome P450s in freshly isolated and primary cultures of proximal tubular cells from human kidney. *J. Pharmacol. Exp. Ther.*, *293*, 677–685.

Cummings, B.S., Parker, J.C., Lash, L.H. (2000). Role of cytochrome P450 and glutathione *S*-transferase α in metabolism and cytotoxicity of trichloroethylene in rat kidney. *Biochem. Pharmacol.*, *59*, 531–543.

Cummings, B.S., Parker, J.C., Lash, L.H. (2001). Cytochrome P450-dependent metabolism of trichloroethylene in rat kidney. *Toxicol. Sci.*, *60*, 11–19.

Cummings, B.S., Zangar, R.C., Novak, R.F., Lash, L.H. (1999). Cellular distribution of cytochromes P-450 in rat kidney. *Drug Metab. Dispos.*, *27*, 542–548.

Dolphin, C.T., Beckett, D.T., Janmohamed, A., Cullingford, T.E., Smith, R.L., Shephard, E.A., Phillips, I.R. (1998). The flavin-containing monooxygenase 2 gene (FMO2) of humans, but

not of other primates, encodes a truncated, nonfunctional protein. *J. Biol. Chem.*, *273*, 30599–30607.

Duggin, G.G. (1980). Mechanisms in the development of analgesic nephropathy. *Kidney Int.*, *18*, 553–561.

Elfarra, A.A., Krause, R.J. (2007). *S*-(1,2,2-Trichlorovinyl)-L-cysteine sulfoxide, a reactive metabolite of S-(1,2,2-trichlorovinyl)-L-cysteine formed in rat liver and kidney microsomes, is a potent nephrotoxicant. *J. Pharmacol. Exp. Ther.*, *321*, 1095–1101.

Elfarra, A.A., Lash, L.H., Anders, M.W. (1987). Alpha-keto acids stimulate rat renal cysteine conjugate β-lyase activity and potentiate the cytotoxicity of *S*-(1,2-dichlorovinyl)-L-cysteine. *Mol. Pharmacol.*, *31*, 208–212.

Emeigh Hart, S.G., Wyand, D.S., Khairallah, E.A., Cohen, S.D. (1996). Acetaminophen nephrotoxicity in the CD-1 mouse. II. Protection by probenecid and AT-125 without diminution of renal covalent binding. *Toxicol. Appl. Pharmacol.*, *136*, 161–169.

Epstein, F.H., Silva, P., Spokes, K., Brezis, M., Rosen, S. (1989). Renal medullary Na-K-ATPase and hypoxic injury in perfused rat kidneys. *Kidney Int.*, *36*, 768–772.

Finkelstein, M.B., Vamvakas, S., Bittner, D., Anders, M.W. (1994). *S*tructure-mutagenicity and structure-cytotoxicity studies on bromine-containing cysteine S-conjugates and related compounds. *Chem. Res. Toxicol.*, *7*, 157–163.

Goldsworthy, T.L., Popp, J.A. (1987). Chlorinated hydrocarbon-induced peroxisomal enzyme activity in relation to species and organ carcinogenicity. *Toxicol. Appl. Pharmacol.*, *88*, 225–233.

Gonzalez, F.J. (1990). Molecular genetics of the P-450 superfamily. *Pharmacol. Ther.*, *45*, 1–38.

Griffith, O.W., Meister, A. (1979). Translocation of intracellular glutathione to membrane-bound γ-glutamyl transpeptidase as a discrete step in the γ-glutamyl cycle: glutathionuria after inhibition of transpeptidase. *Proc. Natl. Acad. Sci. U.S.A.*, *76*, 268–272.

Guder, W.G., Ross, B.D. (1984). Enzyme distribution along the nephron. *Kidney Int.*, *26*, 101–111.

Guder, W.G., Wagner, S., Wirthensohn, G. (1986). Metabolic fuels along the nephron: pathways and intracellular mechanisms of interaction. *Kidney Int.*, *29*, 41–45.

Hanigan, M.L., Gallagher, B.C., Taylor, P.T., Jr, Large, M.K. (1994). Inhibition of gamma-glutamyl transpeptidase activity by acivicin in vivo protects the kidney from cisplatin-induced toxicity. *Cancer Res.*, *54*, 5925–5929.

Hanigan, M.L., Lykissa, E.D., Townsend, D.M., Ou, C., Barrios, R., Lieberman, M.W. (2001). Gamma-glutamyl transpeptidase-deficient mice are resistant to the nephrotoxicity of cisplatin. *Am. J. Pathol.*, *159*, 1889–1894.

Hinchman, C.A., Ballatori, N. (1990). Glutathione-degrading capacities of liver and kidney in different species. *Biochem. Pharmacol.*, *40*, 1131–1135.

Hines, R.N., Cashman, J.R., Philpot, R.M., Williams, D.E., Ziegler, D.M. (1994). The mammalian flavin-containing monooxygenases: molecular characterization and regulation of expression. *Toxicol. Appl. Pharmacol.*, *125*, 1–6.

Horster, M. (1978). Principles of nephron differentiation. *Am. J. Physiol.*, *235*, F387–F393.

Hu, J.J., Rhoten, W.B., Yang, C.S. (1990). Mouse renal cytochrome P450IIE1: immunocytochemical localization, sex-related differences and regulation by testosterone. *Biochem. Pharmacol.*, *40*, 2597–2602.

Ito, O., Alonso-Galicia, M., Hopp, K.A., Roman, R.J. (1998). Localization of cytochrome P-450 4A isoforms along the rat nephron. *Am. J. Physiol.*, *274*, F395–F404.

Iyer, R.A., Frink, E.J., Jr, Ebert, T.J., Anders, M.W. (1998). Cysteine conjugate β-lyase-dependent metabolism of Compound A (2-[fluoromethoxy]-1,1,3,3,3-pentafluoro-1-propene) in human subjects anesthetized with sevoflurane and in rats given Compound A. *Anesthesiology*, *88*, 611–618.

Jacobson, H.R. (1981). Functional segmentation of the mammalian nephron. *Am. J. Physiol.*, *241*, F203–F218.

Jakobsson, P.-J., Morgenstern, R., Mancini, J., Ford-Hutchinson, A., Persson, B. (1999). Common structural features of MAPEG—a widespread superfamily of membrane associated proteins with highly divergent functions in eicosanoid and glutathione metabolism. *Protein Sci.*, *8*, 689–692.

Jones, D.P., Orrenius, S., Jakobson, S.W. (1980). Cytochrome P-450-linked monooxygenase systems in the kidney, in *Extrahepatic Metabolism of Drugs and Other Foreign Compounds* (ed. T.E. Gram), Spectrum Publications, New York, pp. 123–158.

Klos, C., Koob, M., Kramer, C., Dekant, W. (1992). p-Aminophenol nephrotoxicity: biosynthesis of toxic glutathione conjugates. *Toxicol. Appl. Pharmacol.*, *115*, 98–106.

Koukouritaki, S.B., Simpson, P., Yeung, C.K., Rettie, A.E., Hines, R.N. (2002). Human hepatic flavin-containing monooxygenase 1 (FMO1) and 3 (FMO3) developmental expression. *Pediatr. Res.*, *51*, 236–243.

Krause, R.J., Lash, L.H., Elfarra, A.A. (2003). Human kidney flavin-containing monooxygenases and their potential roles in cysteine S-conjugate metabolism and nephrotoxicity. *J. Pharmacol. Exp. Ther.*, *304*, 185–191.

Lash, L.H., Fisher, J.W., Lipscomb, J.C., Parker, J.C. (2000). Metabolism of trichloroethylene. *Environ. Health Perspec.*, *108* (Suppl. 2), 177–200.

Lash, L.H., Lipscomb, J.C., Putt, D.A., Parker, J.C. (1999). Glutathione conjugation of trichloroethylene in human liver and kidney: kinetics and individual variation. *Drug Metab. Dispos.*, *27*, 351–359.

Lash, L.H., Parker, J.C. (2001). Hepatic and renal toxicities associated with perchloroethylene. *Pharmacol. Rev.*, *53*, 177–208.

Lash, L.H., Parker, J.C., Scott, C.S. (2000). Modes of action of trichloroethylene for kidney tumorigenesis. *Environ. Health Perspec.*, *108* (Suppl. 2), 225–240.

Lash, L.H., Putt, D.A., Cai, H. (2008). Drug metabolism enzyme expression and activity in primary cultures of human proximal tubular cells. *Toxicology*, *244*, 56–65.

Lash, L.H., Putt, D.A., Hueni, S.E., Krause, R.J., Elfarra, A.A. (2003). Roles of necrosis, apoptosis, and mitochondrial dysfunction in *S*-(1,2-dichlorovinyl)-L-cysteine sulfoxide-induced cytotoxicity in primary cultures of human renal proximal tubular cells. *J. Pharmacol. Exp. Ther.*, *305*, 1163–1172.

Lash, L.H., Qian, W., Putt, D.A., Jacobs, K., Elfarra, A.A., Krause, R.J., Parker, J.C. (1998). Glutathione conjugation of trichloroethylene in rats and mice: sex-, species-, and tissue-dependent differences. *Drug Metab. Dispos.*, *26*, 12–19.

Lash, L.H., Sausen, P.J., Duescher, R.J., Cooley, A.J., Elfarra, A.A. (1994). Roles of cysteine conjugate β-lyase and *S*-oxidase in nephrotoxicity: studies with *S*-(1,2-dichlorovinyl)-L-cysteine and *S*-(1,2-dichlorovinyl)-L-cysteine sulfoxide. *J. Pharmacol. Exp. Ther.*, *269*, 374–383.

Lash, L.H., Xu, Y., Elfarra, A.A., Duescher, R.J., Parker, J.C. (1995). Glutathione-dependent metabolism of trichloroethylene in isolated liver and kidney cells of rats and its role in mitochondrial and cellular toxicity. *Drug Metab. Dispos.*, *23*, 846–853.

Lasker, J.M., Chen, W.B., Wolf, I., Bloswick, B.P., Wilson, P.D., Powell, P.K. (2000). Formation of 20-hydroxyeicosatetraenoic acid, a vasoactive and natriuretic eicosanoid, in human kidney: role of Cyp4F2 and Cyp4A11. *J. Biol. Chem.*, *275*, 4118–4126.

Lock, E.A., Reed, C.J. (1997). Renal xenobiotic metabolism, in *Comprehensive Series in Toxicology, Vol. 7: Kidney Toxicology* (ed. R.S. Goldstein), Elsevier, Oxford, pp. 77–97.

Lohr, J.W., Willsky, G.R., Acara, A. (1998). Renal drug metabolism. *Pharmacol. Rev.*, *50*, 107–141.

National Toxicology Program (NTP) (2005). *Report on Carcinogens*, 11th edn, U.S. Department of Health and Human Services, Public Health Service, National Toxicology Program, Research Triangle Park, NC.

Newton, J.F., Braselton, W.E., Kuo, C.-H., Kluwe, W.M., Gemborys, M.W., Mudge, G.H., Hook, J.B. (1982). Metabolism of acetaminophen by the isolated perfused kidney. *J. Pharmacol. Exp. Ther.*, *221*, 76–79.

Nishimura, M., Naito, S. (2006). Tissue-specific mRNA expression profiles of human phase I metabolizing enzymes except cytochrome P450 and phase II metabolizing enzymes. *Drug Metab. Pharmacokinet.*, *21*, 357–374.

Okita, J.R., Johnson, S.B., Castle, P.J., Dezellem, S.C., Okita, R.T. (1997). Improved separation and immunodetection of rat cytochrome P450 4A forms in liver and kidney. *Drug Metab. Dispos.*, *25*, 1008–1012.

Otieno, M.A., Baggs, R.B., Hayes, J.D., Anders, M.W. (1997). Immunolocalization of microsomal glutathione *S*-transferases in rat tissues. *Drug Metab. Dispos.*, *25*, 12–20.

Ripp, S.L., Itagak, I.K., Philpot, R.M., Elfarra, A.A. (1999). Species and sex differences in expression of flavin-containing monooxygenase form 3 in liver and kidney microsomes. *Drug Metab. Dispos.*, *27*, 46–52.

Ronis, M.J.J., Huang, J., Longo, V., Tindberg, B., Ingelman-Sundberg, M., Badger, T.M. (1998). Expression and distribution of cytochrome P450 enzymes in male rat kidney: effects of ethanol, acetone and dietary conditions. *Biochem. Pharmacol.*, *55*, 123–129.

Rozzel, B., Hansson, H.-A., Guthenberg, C., Tahir, M.K., Mannervik, B. (1993). Glutathione transferases of classes αμ and π show selective expression in different regions of rat kidney. *Xenobiotica*, *23*, 835–849.

Sands, J.M., Verlander, J.W. (2005). Anatomy and physiology of the kidneys, in *Toxicology of the Kidney*, 3rd edn (eds J.B. Tarloff and L.H. Lash), CRC Press, Boca Raton, Chapter 1.

Sausen, P.J., Duescher, R.J., Elfarra, A.A. (1993). Further characterization and purification of the flavin-dependent *S*-benzyl-L-cysteine *S*-oxidase activities of rat liver and kidney microsomes. *Mol. Pharmacol.*, *43*, 388–396.

Sausen, P.J., Elfarra, A.A. (1990). Cysteine conjugate S-oxidase: characterization of a novel enzymatic activity in rat hepatic and renal microsomes. *J. Biol. Chem.*, *265*, 6139–6145.

Sausen, P.J., Elfarra, A.A. (1991). Reactivity of cysteine S-conjugates sulfoxides: formation of S-[1-chloro-2-(S-glutathionyl)vinyl]-L-cysteine sulfoxide by the reaction of S-(1,2-dichlorovinyl)-L-cysteine sulfoxide with glutathione. *Chem. Res. Toxicol.*, *4*, 655–660.

Schuetz, E.G., Schuetz, J.D., Grogan, W.M., Naray-Fejes-Toth, A., Fejes-Toth, G., Raucy, J., Guzelian, P., Gionela, K., Watlington, C.O. (1992). Expression of cytochrome P450 3A in amphibian, rat, and human kidney. *Arch. Biochem. Biophys.*, *294*, 206–217.

Shanley, P.F., Shapiro, J.I., Chan, L., Burke, T.J., Johnson, G.C. (1988). Acidosis and hypoxic medullary injury in the isolated perfused kidney. *Kidney Int.*, *34*, 791–796.

Stec, D.E., Flasch, A., Roman, R.J., White, J.A. (2003). Distribution of cytochrome P-450 4A and 4F isoforms along the nephron in mice. *Am. J. Physiol.*, *284*, F95–F102.

Stevens, J.L., Robbins, J.D., Byrd, R.A. (1986). A purified cysteine conjugate β-lyase from rat kidney cytosol: requirement for an α-keto acid or an amino acid oxidase for activity and identity with soluble glutamine transaminase K. *J. Biol. Chem.*, *261*, 15529–15537.

Townsend, D.M., Deng, M., Zhang, L., Lapus, M.G., Hanigan, M.H. (2003a). Metabolism of cisplatin to a nephrotoxin in proximal tubule cells. *J. Am. Soc. Nephrol.*, *14*, 1–10.

Townsend, D.M., Hanigan, M.H. (2002). Inhibition of gamma-glutamyl transpeptidase or cysteine conjugate β-lyase activity blocks the nephrotoxicity of cisplatin in mice. *J. Pharmacol. Exp. Ther.*, *300*, 142–148.

Townsend, D.M., Marto, J.A., Deng, M., Macdonald, T.J., Hanigan, M.H. (2003b). High pressure liquid chromatography and mass spectrometry characterization of the nephrotoxic biotransformation products of cisplatin. *Drug Metab. Dispos.*, *31*, 705–713.

Trumper, L., Monasterolo, L.A., Elias, M.M. (1996). Nephrotoxicity of acetaminophen in male Wistar rats: role of hepatically derived metabolites. *J. Pharmacol. Exp. Ther.*, *279*, 548–554.

Walker, L.A., Valtin, H. (1982). Biological importance of nephron heterogeneity. *Annu. Rev. Physiol.*, *44*, 203–219.

Wolfgang, G.H.I., Gandolfi, A.J., Stevens, J.L., Brendel, K. (1989). N-Acetyl S-(1,2-dichlorovinyl)-L-cysteine produces a similar toxicity to S-(1,2-dichlorovinyl)-L-cysteine in rabbit renal slices: differential transport and metabolism. *Toxicol. Appl. Pharmacol.*, *101*, 205–219.

Yin, H., Jones, J.P., Anders, M.W. (1995). Metabolism of 1-fluoro-1,1,2-trichloroethane, 1,2-dichloro-1,1-difluoroethane, and 1,1,1-trifluoro-2-chloroethane. *Chem. Res. Toxicol.*, *8*, 262–268.

Zhang, G., Stevens, J.L. (1989). Transport and activation of S-(1,2-dichlorovinyl)-L-cysteine and N-acetyl-S-(1,2-dichlorovinyl)-L-cysteine in rat kidney proximal tubules. *Toxicol. Appl. Pharmacol.*, *100*, 51–61.

Zhang, L., Hanigan, M.H. (2003). Role of cysteine S-conjugate β-lyase in the metabolism of cisplatin. *J. Pharmacol. Exp. Ther.*, *306*, 988–994.

Ziegler, D.M. (1993). Recent studies on the structure and function of multisubstrate flavin-containing monooxygenases. *Annu. Rev. Pharmacol. Toxicol.*, *33*, 179–199.

# Drug Metabolism in Regulatory Guidances, Clinical Trials, and Product Labeling

MARK N. MILTON

## 1 INTRODUCTION

The role of the drug metabolism scientist has changed markedly over the past 20 years with our enhanced understanding of the enzymology of drug metabolism, the importance of drug metabolism in the efficacy and safety of drugs, and the increasingly interdisciplinary nature of the drug development process. Nowadays, the drug metabolism scientist will need to interact with many different disciplines within a pharmaceutical company. These disciplines include toxicology, pharmacology, formulations, chemistry, clinical, clinical pharmacology, regulatory affairs, marketing, medical affairs, competitive intelligence, and business development. Additionally, they may be responsible for representing the multiple subdisciplines (e.g. bioanalytical, *in vitro* metabolism, *in vivo* metabolism, transporters, and pharmacokinetics) that contribute to the understanding of the absorption, distribution, metabolism, and elimination (ADME) properties of a drug. Therefore, the drug metabolism scientist need not only be an expert in drug metabolism and its related disciplines, but also be a generalist who understands the need of his or her customers, whether they be medicinal chemists, toxicologists, regulatory agencies, or clinicians. They will need to possess a basic understanding of how knowledge of the ADME properties of a drug can be used to efficiently and effectively develop a drug and the regulatory environment that pharmaceutical companies operate within.

The objective of this chapter is to give the reader a high-level overview of how drug metabolism contributes to the risk–benefit analysis of a novel pharmaceutical; the role of drug metabolism in *clinical trials* and the contribution of drug metabolism to the *product label*. Readers should not rely on this chapter as a definitive source of information. They should read the relevant source information for themselves and come to their own conclusions regarding the interpretation of a given regulatory guidance. A

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

more detailed review of the drug metabolism and pharmacokinetics in drug development has been performed by Bonate and Howard (2004) and Rogge and Taft (2005). The role of drug metabolism in the development of agricultural chemicals, veterinary products, and food additives will not be addressed in this chapter. It is difficult, at times, to separate drug metabolism from pharmacokinetics and their contributions to safety and efficacy. Therefore, this chapter will view drug metabolism, pharmacokinetics, and drug transporters as being inexorably linked and will attempt to give drug metabolism scientists an overview of how their discipline relates to, and interacts with, the regulatory environment, clinical trials, and the product label.

Certain parts of this chapter will contain materials that have been lifted verbatim from regulatory guidances. Since the exact wording of regulatory guidances can be of paramount importance, it should be noted that such extracts of the guidances will no longer be in the context of the full guidance and therefore the reader is encouraged to use this chapter as a starting point for understanding the role of drug metabolism in regulatory guidances and to review the source documentation for themselves. As all guidances are open to interpretation, the views expressed here are those of the author and may differ from the views of the reader, colleagues, or the regulatory agencies. It should also be noted that the interpretation of the guidance documents may change with time based on the experience of the regulatory agencies, may vary between countries, and even within a given country's regulatory agency.

Subsections have been included in this chapter for ease of navigation, although it should be noted that it can be difficult to place a given guidance in a given category due to the multidisciplinary nature of many of these guidances and the overlap between nonclinical and clinical development.

## 2 THE REGULATORY ENVIRONMENT

### 2.1 Why Do We Have Drug Regulations?

The regulations for the pharmaceutical industry have tended to be created in reaction to tragedy rather than being created proactively in order to avoid a tragedy such as what occurred with sulfanilamide, TGN1412, thalidomide, terfenadine, the floxacins, FIAU, Phen-fen, Ephedra, encainide/flecainide, propulsid, Redux, and Duract. The subsequently created guidances are an effort in order to avoid further and greater tragedies. Drug metabolism has been implicated to varying extents in these tragedies that have shaped the regulatory environment for pharmaceuticals. Many of the regulatory safeguards in the United States came about as part of the Kefauver–Harris drug amendments to the Food, Drug and Cosmetic Act in 1962. Interestingly, the original purpose of these amendments was to introduce price-cutting but in the wake of the Thalidomide tragedy, the price-cutting elements of the amendments were downplayed and a greater importance on safety was introduced.

### 2.2 The Global Nature of Drug Development

The regulation of the development of pharmaceuticals will be broadly similar throughout the world, although each country or region will have its own specific guidances or areas of interest. This chapter will focus primarily on regulatory submissions to the FDA (United States) and EMEA (European Union), since these regulatory agencies

**TABLE 1 Web sites containing regulatory guidances.**

| Country or region | Web site |
|---|---|
| United States | http://www.fda.gov |
| Canada | http://www.hc-sc.gc.ca |
| EU | http://www.emea.europa.eu |
| Japan | http://www.mhlw.go.jp/english/index.html |
| Australia | http://www.tga.gov.au |
| Global (ICH) | http://www.ich.org |
| Global (OECD) | http://www.oecd.org |

have created the majority of the existing regulatory guidances or guidelines. However, additional guidances have been created in other regions (e.g. Canada and Japan). Efforts have been made to globalize certain guidances under the auspices of the International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH). This is a unique project that brings together the regulatory authorities of Europe, Japan, and the United States, and experts from the pharmaceutical industry in the three regions to discuss scientific and technical aspects of product registration. The purpose of the ICH is to make recommendations on ways to achieve greater harmonization in the interpretation and application of technical guidelines and requirements for product registration in order to reduce or obviate the need to duplicate the testing carried out during the research and development of new medicines. The objective of such harmonization is a more economical use of human, animal, and material resources, and the elimination of unnecessary delay in the global development and availability of new medicines while maintaining safeguards on quality, safety, and efficacy, and regulatory obligations to protect public health. It should be noted that the different countries that form the European Medicines Agency (EMEA) may have different local practices with respect to the implementation of the EMEA guidelines or EU directives. Sources of information on regulatory guidances are summarized in Table 1.

The regulatory process and requirements in Japan can be quite different to the requirements in other parts of the world. An overview of the regulatory process and guidelines in Japan was published by the Japan Pharmaceutical Manufacturers Association in 2007 (Japan Pharmaceutical Manufacturers Association, 2007) and serves as a very good reference source for drug development in Japan. Additionally, a very detailed document was issued in 1995 (Japanese Guidelines for Nonclinical Studies of Drugs Manual 1995) under the editorial supervision of the Pharmaceuticals and Cosmetics Division, Pharmaceutical Affairs Bureau, Japanese Ministry of Health and Welfare (Ministry of Health and Welfare, Pharmaceutical Affairs Bureau, 1995). This document contains guidelines for toxicity studies of drugs, guidelines for general pharmacology studies, and guidelines for nonclinical pharmacokinetic studies. This document is a very good resource for a general description of the design and conduct of nonclinical studies.

### 2.3 The Process of Drug Development

The development of new drugs and the expense of clinical trials to demonstrate the safety and efficacy of innovative drugs is rewarded through granting of a period of

marketing exclusivity that shields the product from competition. The purpose of the regulatory agencies, such as the FDA, is to protect the safety of the patient from people that would want to sell them snake oil (drugs that are non-efficacious and harmful) while providing an avenue for drug developers to demonstrate efficacy, assess toxicity, and secure marketing approval.

Since the development of a new drug is a costly and timely effort and since it is of paramount importance to protect the safety of the patient or volunteer, it is important to build a strong working relationship with the regulatory agencies. For the FDA, this relationship can be initiated prior to the initiation of the first clinical trial by holding a Pre-IND meeting (US FDA, 2000a). The focus of the regulatory agency during its review of an IND application (Investigational New Drug) for a first-in-human (FIH) clinical trial is to ensure that the drug is safe to administer to the proposed population (patients or normal human volunteers) and not to determine whether or not there is a market need for the drug (US FDA, 2007o). At this stage of development, the FDA's focus is not on efficacy. This statement applies to regulatory agencies worldwide, and not just to the FDA. Prior to the submission of the IND, most sponsors will have taken advantage of holding a Pre-IND meeting with the FDA (US FDA, 2005a). The pre-IND meeting can be very valuable in planning a drug development program, especially if sponsors' questions are not fully answered by guidances and other information provided by the FDA. Early interactions with FDA staff can help to prevent clinical hold issues from arising. A pre-IND meeting can also provide sponsors information that will assist them in preparing to submit complete IND applications. For a Pre-IND meeting, the sponsors will submit specific questions that they wish to have answered by the FDA along with a briefing book that will provide sufficient background information for the FDA to provide its response. It is recommended that the sponsor does not ask open-ended questions or ask whether a certain study should be performed or not. The most successful meetings are those in which there is a distinct objective and which proposals are made that the regulatory agency can respond to. This approach will apply to all regulatory agencies (not just the FDA) and to all types of meetings.

In the EU, the process for interacting with the EMEA is, in general, similar to the process for interacting with the FDA. The EMEA has created a guideline for obtaining scientific advice (EMEA, 2007c). However, few sponsors request scientific advice from the EMEA prior to the FIH study due to the length of time and cost of obtaining scientific advice.

At the time of the FDA's review of the sponsor's marketing application (NDA; New Drug Application), the focus will be on safety and efficacy (i.e. the drug must be safe and efficacious). However, the FDA may also choose to take into account the current standard of care for the disease, the benefit that the new drug offers over the current therapies, and projected cost of the therapy when approving the drug. This latter approach has recently been criticized by Bob Essner, the outgoing chief executive of Wyeth. In an article that was published in the *Financial Times*, he expressed concerns that the FDA assessments are exceeding their legal mandate and such reviews now include whether a drug can work better than an existing product, instead of allowing the market to determine what it wants (Bowe, 2007). Concern has been expressed that the FDA has become ultracautious in the past few years which may be in response to high-visibility safety concerns with drugs such as Vioxx. Similarly, regulatory agencies in the EU have become cautious in the wake of the "TeGenero incident" (Wikipedia, 2008a). Recently, the osteoarthritis drug Prexige has been withdrawn from the market in several countries including Canada and Australia amid concerns about the drug

increasing the risk for serious liver-related damage including hepatitis (Health Canada, 2007b; Novartis Australia Pty Limited, 2007).

In anticipation of the expiration of marketing exclusivity and applicable patent protections, potential generic manufacturers begin the scientific and technical process of generic drug development. To receive approval, generic drug applicants must demonstrate that their products are pharmaceutically equivalent and bioequivalent to the reference product. Pharmaceutically equivalent products have the same active ingredient(s) in the same strength in the same dosage form. Bioequivalent products show no significant difference in the rate and extent of absorption of the active ingredient at the site of action. For many drug products, demonstrating pharmaceutical equivalence and bioequivalence is straightforward. Analytical chemistry can identify and quantitate the active ingredient. Comparison of pharmacokinetic parameters is used to evaluate bioequivalence. Bioequivalence based on plasma drug concentration has been identified as the most commonly used and successful biomarker of safety and efficacy. When common approaches to the assessment of bioequivalence and pharmaceutical equivalence are not applicable, as is the case for complex drug products and locally acting drugs, scientific challenges have presented barriers to the development and approval of generic drugs. Examples of where these challenges have limited development include topical and inhaled drug products.

### 2.4 The Nature of Regulatory Guidances

The requirements for the development of a pharmaceutical are primarily managed by guidance (FDA terminology) or guideline (EMEA terminology) documents and to a far lesser extent by legislation. Guidances or guidelines are not legally enforceable documents but instead provide the regulated community with a better understanding of agency positions on specific issues. This is illustrated by a statement that is found on most, if not all, of the FDA guidance documents: "This guidance represents the Food and Drug Administration's (FDA's) current thinking on this topic. It does not create or confer any rights for or on any person and does not operate to bind FDA or the public. You can use an alternative approach if the approach satisfies the requirements of the applicable statutes and regulations. If you want to discuss an alternative approach, contact the FDA staff responsible for implementing this guidance." Alternative statements include "FDA's guidance documents, including this guidance, do not establish legally enforceable responsibilities. Instead, guidances describe the Agency's current thinking on a topic and should be viewed only as recommendations, unless specific regulatory or statutory requirements are cited." The lack of binding nature of the guidances is again emphasized in the FDA's Good Guidance Practice rules (U.S. Environmental Protection Agency, 2008), guidance documents "do not legally bind the public or FDA" and entities "may choose to use an approach other than the one set forth in a guidance document." In addition to guidance documents, the FDA will occasionally issue Q&A documents that provide clarification to a given regulatory guidance document. It should be noted that the Q&A documents do not appear on the FDA's CDER guidances web page. This lack of coordination is unfortunate and will be described further in Section 9, "Procedural Guidances." Guidelines issued by the EMEA (excepting those generated by the ICH process) will contain a statement regarding the legal status of the guideline and indicate whether the guideline should be read in conjunction with other documentation such as directives (laws) or guidelines.

It is important to note that guidance documents are essentially recommendations and not mandates. This point is illustrated by the following statement that is commonly found in FDA guidance documents: "The use of the word should in Agency guidances means that something is suggested or recommended, but not required." However, once agencies commit to a position in a guidance document, they may become inflexible in applying it, and may apply it in such a way as to create binding regulatory obligations. The flexibility in the regulatory guidances can be a double-edged sword. Not only do the guidances allow the sponsors the latitude to propose alternative strategies to the regulatory agency, they also allow for great variability in interpretation between review divisions at the FDA and between reviewers within a given organization. The guidances are not intended to provide a "one size fits all" approach but are intended to be implemented on a "case-by-case basis" that takes into account the nature of the therapeutic and the disease.

Guidances have been used by regulatory agencies in an attempt to nudge sponsors in the "correct" direction and to provide some level of standardization of the materials that are provided to the regulatory agencies for review. The regulatory reviewers have a high workload and have to make very impactful decisions in a very short period of time. These time frames for review of an application for a FIH clinical trial are very short (30 days in the United States and 60 days in the United Kingdom, although the goal for the review in the United Kingdom is 14 days). The reviews are based on the quality and the content of the written documentation that is submitted to the regulatory agencies. The regulatory agencies will not take into account the scientific expertise of the individual(s) involved in the conduct and interpretation of these studies, even if they did know the identities of these individuals. It should be noted that a sponsor could have performed an excellent job of understanding the properties of the drug, only for it not to be approved (either for entry in to clinical trials or for marketing) based on the level and quality of the documentation. The topic of documentation will be addressed in Section 9, "Procedural Guidances."

### 2.5 Other Important Sources of Information

In addition to regulatory guidances, there are other sources of information that will be of great use to the drug metabolism scientists in their role of designing a robust plan for the characterization of the ADME properties of a drug. These sources include the regulatory agency's review of the sponsor's marketing application. These reviews are known as the Summary Basis of Approval (SBA) for drugs approved by the US FDA and the European Public Assessment Report (EPAR) for drugs approved by the Committee for Medicinal Products for Human Use (CHMP) which is part of the EMEA (EMEA, 2008a; US FDA, 2008c). The European Medicines Agency (EMEA) is a decentralized body of the European Union with headquarters in London. Both the SBAs and EPARs will contain a summary of the information used for the regulatory agency's decision to approve the marketing application. There is a greater level of detail in the SBA than in the EPAR, although both documents will not contain commercially confidential information (usually related to chemistry, manufacturing, and controls or CMC). The Division of Oncology Drug Products of the FDA has recently begun to publish an "approval summary" for a newly approved oncologic agent in the peer-reviewed literature. For example, the FDA has published a paper on the approval of Vorinostat for the treatment of advanced primary cutaneous T-cell lymphoma (Mann *et al.*, 2007). Other sources of pertinent information include "withdrawal reports" that

are published by the EMEA which are similar in detail to the EPAR but also containing the reason why the drug's marketing application was not approved (EMEA, 2008f), scientific publications or presentations by regulatory agency personnel, meeting presentations and transcripts from FDA Advisory Committee meetings, approval letters published by the FDA and product labels, and web sites for regional pharmaceutical manufacturer's associations. As can be seen, there is a large amount of information that is available to the drug metabolism scientist.

### 2.6 The Process of Creating a New Regulatory Guidance

When FDA plans to issue a new regulation or revise an existing one, it places an announcement in the Federal Register on the day the public comment period begins. Published every weekday, the Federal Register is available at many public libraries and colleges, and on the FDA web site (www.fda.gov). In the Federal Register, the "notice of proposed rulemaking" describes the planned regulation and provides background on the issue. It also gives the address for submitting written comments and the name of the person to contact for more information. The draft guidance is also made available on the FDA's guidance web page. Also noted is the "comment period," which specifies how long the agency will accept public comments. Usually, the file (or docket) stays open for comments at least 60 days, though some comment periods have been as short as 10 days or as long as nine months. Weekends and holidays are included in the comment period. This comment period offers the drug metabolism scientists in the regulated industry the opportunity to influence the final guidance. All drug metabolism scientists are highly encouraged to keep themselves abreast of the appropriate regulatory guidances (final or draft) as well as participate in providing comments on the draft guidances. A similar review and comment approach is taken by the EMEA. However, the EMEA will often first issue a "concept paper" that outlines what the draft guideline will contain. In addition to publishing draft guidances, the FDA and EMEA will issue lists of guidances that they are planning on issuing in the current year. This list can provide a useful insight into the issues that these regulatory agencies view as being unaddressed. It should be noted that the FDA and EMEA handle draft guidances/guidelines in a very different manner once the guidance/guideline is finalized. The FDA updates its web site with the finalized version of the guidance and removes the draft guidance. Occasionally, the draft guidance can be located by searching the FDA web site, although in most cases it is no longer available to the general public. The EMEA takes a different approach. When the draft guideline is finalized, it will be added to the relevant web page. However, the web page will also contain a link to the draft guidance as well as a summary of the comments that were received on the draft guideline. Therefore, it is possible to readily determine what changed between the draft and final versions of the guideline.

In recent years, both the FDA and EMEA have made available to the public the comments on guidances made by interested parties. These documents can help the reader understand the extent and depth of comments from individuals, pharmaceutical companies, industry groups, and others. The comments on EMEA guidelines are colocated with the final guideline. The EMEA has organized the guidelines for a given topic (e.g. nonclinical) in a comprehensive and easily understandable tabular format (EMEA, 2008h). The comments are tabulated and include the EMEA's responses to the individual comments, noting whether they will be accepted or not. Comments on draft FDA guidances can be found in the docket (e.g. Amgen's comments on the draft guidance

on metabolites in safety testing) (Amgen, 2005) but cannot be easily navigated to and from the main guidance document page (US FDA, 2008g).

From time to time, the pharmaceutical industry may publish a position paper on a topic that describes an approach to a certain element of the drug development process. Such position papers are often created with the hope that the existence of a standardized industry approach may negate the need for the creation of a regulatory guidance or to proactively influence the FDA's thinking on a particular topic. While this approach may be successful in certain circumstances, it may not work in others. For example, in 2002, a proposal was published on "metabolites in safety testing" (MIST) by Baillie *et al.* (2002), which suggested some guidelines regarding when it is necessary to provide greater assessment of the safety of metabolites. This position paper addressed the circumstances under which the toxicity of a metabolite should be evaluated. The cutoff below (in terms of exposure of humans to the metabolite relative to the exposure to total compound-related material) which no testing would be required was far greater than the FDA was comfortable with. This resulted in a response to the original article in terms of a letter to the editor by Hastings *et al.* (2003) and a response by Baillie *et al.* (2003) and a subsequent draft guidance that was issued for review in June 2005 (US FDA, 2005c). It has been speculated that the draft guidance was created in response to the PhRMA white paper and may not have been created if the white paper had not been published. The final guidance was issued in February 2008 and the draft guidance is no longer available on the FDA's Guidance Documents web page. As a note of caution, the web page on the FDA web site that is intended to redirect the reader from the draft guidance to the final guidance does not (as of 12 March 2008) direct the reader to the finalized version of the guidance "Safety Testing of Drug Metabolites" but instead redirects the reader to a different guidance, namely "Clinical Trial Endpoints for the Approval of Cancer Drugs and Biologics" (US FDA, 2007i).

There is an increasingly common phenomenon, namely the "permanently draft" guidance. A draft guidance may be issued for review but never finalized. There are draft guidances issued by the FDA that are over a decade old. This can lead to draft guidances becoming *de facto* finalized guidances. It would be preferable if draft guidances were finalized in a timely manner or moved to a separate section where they are not intermixed with finalized guidances. A table of finalized and draft guidances can be found on the FDA's web site (US FDA, 2008a) and are summarized by type in Table 2. A review of this list on May 1, 2008 revealed that there were 460 documents listed (FDA guidances, ICH guidances, and industry letters). The vast majority (approximately 33%) of these documents were still in draft form.

Twenty-one (~13%) of the draft documents were issued in 2007. The oldest of the draft documents dates from January 1985. The FDA is currently developing additional draft guidances and as the projects under the Critical Path Initiative (US FDA, 2006a) progress, even more guidances are likely to be created. Based on data from 1998 to May 2008, it is predicted that in 2022 there will be approximately 900 FDA guidances of which approximately 50% will be draft. However, given funding and resource constraints, it is possible that the rate of growth of the number of guidances may slow and the percentage of draft guidances may increase.

The FDA routinely publishes a list of guidances under development (US FDA, 2008f). Guidances that CDER (Center for Drug Evaluation and Research) is planning to develop during 2008 that are of interest to the drug metabolism scientist include "Immunogenicity Assessment for Therapeutic Proteins," "PET CGMPs," "Determining Whether Human Research Studies Can Be Conducted Without an IND,"

**TABLE 2 Draft and final FDA guidances and letters by type.**

| | Number of guidances | | Percent of guidances | |
|---|---|---|---|---|
| Type | Final | Draft | Final | Draft |
| Advertising | 3 | 5 | 38 | 63 |
| Biopharmaceutics | 11 | 3 | 79 | 21 |
| Chemistry | 31 | 9 | 78 | 23 |
| Clinical antimicrobial | 5 | 26 | 16 | 84 |
| Clinical medical | 51 | 37 | 58 | 42 |
| Clinical pharmacology | 8 | 3 | 73 | 27 |
| Combination products | 1 | 3 | 25 | 75 |
| Compliance | 17 | 7 | 71 | 29 |
| Current good manufacturing practices | 7 | 5 | 58 | 42 |
| Drug safety | 1 | 1 | 50 | 50 |
| Electronic submissions | 6 | 7 | 46 | 54 |
| Generic drugs | 19 | 4 | 83 | 17 |
| Good review practices | 3 | 0 | 100 | 0 |
| ICH—efficacy | 19 | 4 | 83 | 17 |
| ICH—multidisciplinary | 9 | 2 | 82 | 18 |
| ICH—quality | 23 | 4 | 85 | 15 |
| ICH—safety | 15 | 1 | 94 | 6 |
| INDs | 1 | 0 | 100 | 0 |
| Industry letters | 14 | 0 | 100 | 0 |
| Labelling | 7 | 10 | 41 | 59 |
| OTC | 5 | 6 | 45 | 55 |
| Pharm/Tox | 14 | 4 | 78 | 22 |
| Procedural | 31 | 9 | 78 | 23 |
| Small entity compliance guide | 1 | 0 | 100 | 0 |
| User fee | 7 | 1 | 88 | 13 |
| Total | 309 | 151 | 67 | 33 |

"Content and Formatting of the Clinical Pharmacology Section of a Product Label," "Biotechnology-Derived Pharmaceuticals: Nonclinical Safety Evaluation," "Genotoxic and Carcinogenic Impurities in Drug Substances and Products: Recommended Approaches," "Assessment of Abuse Potential of Drugs," and "Determining Whether Human Research With a Radioactive Drug Can Be Conducted Under a Radioactive Drug Research Committee (RDRC)." Two guidances on the list, "Nonclinical Safety Evaluation of Reformulated Drug Products Intended for Administration by an Alternate Route" and "Diabetes Mellitus: Developing Drugs and Therapeutic Biologics for Treatment and Prevention" were issued in early March 2008 (US FDA, 2008h, 2008i). The ICH web site (ICH, 2008) provides insight into proposed topics (concept papers). Similarly, the EMEA (CHMP) publishes work plans indicating the status of ongoing projects. However, these work plans are somewhat buried on its web site as it is presently structured (EMEA, 2008i).

Recently, there has been an interesting turn of events that has led to the "release" of a draft guidance that did not appear to be on the list of guidances that the FDA is developing. This draft guidance, "Nonclinical Safety Evaluation of Biotechnology-Derived Pharmaceuticals," is dated August 2006 and was not officially released for

review but nevertheless had been widely distributed within the pharmaceutical industry. However, this topic is now on the list of guidances that the FDA intends to develop in 2008 (US FDA, 2008f).

Over the past few years, there seems to have been a level of unofficial collaboration between the FDA and EMEA in terms of which guidances are being developed. Logistically, it makes sense for only one agency to develop guidance on a given topic in order to conserve resources. For example, the EMEA has issued a guideline on the analysis of control samples from toxicology studies. Since sponsors may potentially submit a marketing application in both the European Union and United States, they will follow this guidance even if they are presently only developing the drug within the United States. Similarly, the FDA has issued a draft guidance on the safety testing of drug metabolites. At present, the EMEA does not appear to be planning to create an analogous guideline.

In addition to regulatory guidances, there are other important sources of information that the drug metabolism scientist should pay attention to. For example, the Oncology Division of the FDA has published articles in peer-reviewed journals or as book chapters. These published articles by DeGeorge *et al.* (1997) and Farrell *et al.* (2003) have become *de facto* guidances. These articles are often used by the FDA reviewers to supplement the existing regulatory guidances. Recently, the FDA published an article that provided an update on CYP enzymes, transporters, and the guidance process (Huang *et al.*, 2008). This article provides a good overview of the topic and how the FDA arrived at its current position.

In addition to regulatory guidances, certain aspects of the drug development process have been codified into law. A list of pertinent regulations for the United States can be found on the FDA web site (US FDA, 2008k). The legislation tends to be high level compared with the regulatory guidances and the information contained in the legislation is objective rather than subjective as is the case with the guidances. Examples of US legislation that pertains to drug metabolism in the development of a novel pharmaceutical include Good Laboratory Practice (21 CFR Part 58), Product Labeling (21 CFR Part 201), Bioavailability and Bioequivalence (21 CFR Part 320), and Radioactive Drugs for Certain Research Purposes (21 CFR Part 361.1) (US FDA, 2007a, b, c, d). In the EU, the legislative documents are known as directives (European Commission, 2008). One example of a directive is the Clinical Trials Directive on the approximation of the laws, regulations, and administrative provisions of the Member States relating to the implementation of good clinical practice in the conduct of clinical trials on medicinal products for human use (European Commission, 2001).

### 2.7 The Challenge of Finding the Information That You Require

The drug metabolism scientist can face several challenges when attempting to locate guidances or guidelines that mention drug metabolism. This challenge is not unique to drug metabolism and pharmaceutical scientists because many disciplines may find similar challenges. However, the challenges for the drug metabolism scientist may be greater than for other disciplines since, although drug metabolism is a very important factor in the evaluation of the safety and efficacy of drugs, it has very rarely been the main topic of regulatory guidances or guidelines. The lack of visibility of the contribution of drug metabolism to the risk–benefit analysis for a new pharmaceutical is reflected in the fact that the nonclinical section of the US IND is known as "Pharm/Tox" and the reviewers at the FDA are called Pharm/Tox reviewers.

The guidances on the FDA web site are arranged according to organizational areas such as advertising, chemistry, clinical pharmacology, pharmacology/toxicology, and procedural. There are 22 such categories. The guidelines on the EMEA web site are organized into five categories: Quality (Chemical and Herbal), Biologics, Non-Clinical, Clinical Efficacy and Safety, and Multidisciplinary. The ICH guidances are organized into four categories: Safety (S), Efficacy (E), Multidisciplinary (M), and Quality (Q). In all cases, there presently is no simple way to find all guidances that relate to drug metabolism by searching for keywords or the content of each guidance.

### 2.8 Using Regulatory Guidances

As mentioned previously, it is not mandatory to follow regulatory guidances, although it is most often advisable to do so. Utilizing information in finalized regulatory guidances is relatively straightforward. However, the existence of a large number of draft guidances creates a conundrum for the sponsor. Since the guidance has not been finalized, it may be difficult to use the draft guidance as precedent as it may no longer represent the regulatory agency's current thoughts on a given topic. Interestingly, since there appears to be no obligation for the draft guidances to be finalized within a defined time frame, or to be withdrawn if they have not been finalized, it leaves open the possibility that regulatory guidances could be released as draft guidances without the need to take the extra step to finalize them. Since these draft guidances can be found on the FDA's guidance web page, many sponsors will attempt to comply with them even though they are not final guidances. This may be risky, since the sponsor will not know *a priori* why the guidance has not been finalized. The reasons could range from workload to difficulty in coming to consensus. The draft guidance could contain outdated or inaccurate information, although it should be noted that finalized guidances may also contain outdated information. Another confusing point is that some guidances state that they are "guidance for industry" and others state that they are "guidance for industry and review staff." It is unclear whether the FDA is trying to make a subtle distinction or not.

## 3 BACKGROUND INFORMATION IN REGULATORY GUIDANCES

Several regulatory guidances provide some basic drug metabolism information in order to establish the context for the guidance. Such information is useful to the nonexpert and helps the expert better understand the rationale behind the guidance. Several examples of such background information that pertains to drug metabolism are presented further in the discussion. One such example of the background information that may be found in a guidance document comes from the 1997 FDA guidance, "Drug Metabolism/Drug Interaction Studies in the Drug Development Process: Studies In Vitro" (US FDA, 1997a). This guidance states:

> After entering the body, a drug is eliminated either by excretion or by metabolism to one or more active or inactive metabolites. When elimination occurs primarily by metabolism, the routes of metabolism can significantly affect the drug's safety and efficacy and the directions for use. When elimination occurs via a single metabolic pathway, individual differences in metabolic rates can lead to large differences in drug and metabolite concentrations in the blood and tissue. In some instances, differences exhibit a bimodal distribution

> indicative of a genetic polymorphism (e.g. CYP450 2D6, CYP450 2C19, N-acetyl transferase). When a genetic polymorphism affects an important metabolic route of elimination, large dosing adjustments may be necessary to achieve the safe and effective use of the drug. Pharmacogenetics already has influenced therapeutics. For a drug that is primarily metabolized by CYP450 2D6, approximately 7% of Caucasians will not be able to metabolize the drug, but the percentage for other racial populations is generally far lower. Similar information is known for other pathways, prominently, CYP450 2C19 and N-acetyltransferase. Equally important, if not more so, many enzymatic metabolic routes of elimination, including most of those occurring via the CYP450 enzymes, can be inhibited or induced by concomitant drug treatment. As a result, abrupt changes can occur with a co-administered agent in a single individual. Such interactions can lead to a substantial decrease or increase in the blood and tissue concentrations of a drug or metabolite or cause the accumulation of a toxic substance (e.g. certain antihistamine-antifungal interactions). These types of changes can alter a new drug's safety and efficacy profile in important ways, particularly for a drug with a narrow therapeutic range.
>
> An understanding of metabolic pathways and potential interactions sometimes allows the use of a drug that would produce an unacceptable level of toxicity if blood levels were not predictable. For these reasons, it is important to learn at an early stage of development whether a drug is eliminated primarily by excretion of unchanged drug or by one or more routes of metabolism. If elimination is primarily by metabolism, the principal metabolizing route(s) should be understood. This information will help identify the implications of metabolic differences between and within individuals and the importance of certain drug-drug and other interactions. Having such information also will aid in determining whether the pharmacologic properties of certain metabolites should be explored further.

This background information clearly communicates the FDA's rationale for the development of this guidance.

Similarly, the draft guidance on "Drug Interaction Studies—Study Design, Data Analysis, and Implications for Dosing and Labeling" that was issued in 2006 (US FDA, 2006g) states:

> The desirable and undesirable effects of a drug arising from its concentrations at the sites of action are usually related either to the amount administered (dose) or to the resulting blood concentrations, which are affected by its absorption, distribution, metabolism, and/or excretion. Elimination of a drug or its metabolites occurs either by metabolism, usually by the liver or gut mucosa, or by excretion, usually by the kidneys and liver. In addition, protein therapeutics may be eliminated through a specific interaction with cell surface receptors, followed by internalization and lysosomal degradation within the target cell. Hepatic elimination occurs primarily by the cytochrome P450 family (CYP) of enzymes located in the hepatic endoplasmic reticulum, but may also occur by non-P450 enzyme systems, such as N-acetyl and glucuronosyl transferases. Many factors can alter hepatic and intestinal drug metabolism, including the presence or absence of disease and/or concomitant medications, or even some foods, such as grapefruit juice. While most of these factors are usually relatively stable over time, concomitant medications can alter metabolism abruptly and are of particular concern. The influence of concomitant medications on hepatic and intestinal metabolism becomes more complicated when a drug, including a prodrug, is metabolized to one or more active metabolites. In this case, the safety and efficacy of the drug/prodrug are determined not only by exposure to the parent drug but by exposure to the active metabolites, which in turn is related to their formation, distribution, and elimination. Therefore, adequate assessment of the safety and effectiveness of a drug includes a description of its metabolism and the contribution of metabolism to overall elimination. For this reason, the development of sensitive and specific assays for a drug and its important metabolites is critical to the study of metabolism and drug-drug interactions.

A third guidance, the 1999 FDA guidance on "In Vivo Drug Metabolism/Drug Interaction Studies—Study Design, Data Analysis, and Recommendations for Dosing and Labeling" also makes some general statements and additionally provides some context for the guidance (US FDA, 1999a). The guidance states:

> No distinction is made in this document between the effects of concomitant drugs and/or alterations in metabolism on gastrointestinal absorption and hepatic elimination, although the pharmacokinetic effects of the two may be different. The desirable and undesirable effects of a drug arising from its concentrations at the sites of action are usually related either to the amount administered (dose) or to the resulting blood concentrations, which are affected by its absorption, distribution, metabolism and/or excretion.
>
> Elimination of a drug or its metabolites occurs either by metabolism, usually by the liver, or by excretion, usually by the kidneys and liver. In addition, protein therapeutics may be eliminated via a specific interaction with cell surface receptors, followed by internalization and lysosomal degradation within the target cell. Hepatic elimination occurs primarily by the cytochrome P450 family of enzymes located in the hepatic endoplasmic reticulum but may also occur by non-P450 enzyme systems, such as N-acetyl and glucuronosyl transferases. P450 enzyme systems located in gut mucosa can also significantly affect the amount of drug absorbed into the systemic circulation. Many factors can alter hepatic and intestinal drug metabolism, including the presence or absence of disease and/or concomitant medications. While most of these factors are usually relatively stable over time, concomitant medications can alter metabolic routes of absorption and elimination abruptly and are of particular concern. The influence of concomitant medications on hepatic and intestinal metabolism becomes more complicated when a drug, including a prodrug, is metabolized to one or more active metabolites. In this case, the safety and efficacy of the drug/prodrug are determined not only by exposure to the parent drug but by exposure to the active metabolites, which in turn is related to their formation, distribution, and elimination.

These latter two guidances not only give a background rationale for the guidance, but they also make the reader recognize the fact that not all metabolism occurs in the liver and not all metabolisms are CYP-mediated. These are important points to recognize since the details of this (and other guidances) do give the strong impression that only CYP-mediated hepatic metabolism needs to be taken into account when evaluating the metabolism of a drug. This background information also raises the topic of prodrugs which is a topic that is mentioned in passing in several guidances but for which there is not a specific guidance.

A final example comes from a guidance document in which metabolism is mentioned in passing and is not the main thrust of the guidance. ICHS3b was issued in 1995 and is titled "Pharmacokinetics: Guidance for Repeated Dose Tissue Distribution Studies" (ICH, 1994a). In the introduction to this guidance, it states:

> A comprehensive knowledge of the absorption, distribution, metabolism and elimination of a compound is important for the interpretation of pharmacology and toxicology studies. Tissue distribution studies are essential in providing information on distribution and accumulation of the compound and/or metabolites, especially in relation to potential sites of action; this information may be useful for designing toxicology and pharmacology studies and for interpreting the results of these experiments.

Interestingly, there is no such guidance for single-dose tissue distribution studies nor is there a clear suggestion that such studies are a crucial part of the drug development

process in any regulatory guidance. This topic will be discussed further later on in this chapter.

## 4 NONCLINICAL GUIDANCES AND EARLY-STAGE DRUG DEVELOPMENT

### 4.1 Introduction

Drug metabolism scientists work closely with their nonclinical (toxicology and pharmacology) colleagues to design studies and interpret the data from these studies. The vast majority of nonclinical guidances primarily relate to toxicology, several to drug metabolism and few, if any, to pharmacology. It should be noted that the line between toxicology and pharmacology can become blurred. Safety pharmacology has been defined as "those studies that investigate the potential undesirable effects of a substance on physiological functions in relation to exposure in the therapeutic range and above." These studies are most often performed by toxicologists and are part of the safety assessment of a drug, but will be found in the pharmacology section of a regulatory submission. It is not surprising that there are a large number of guidances related to toxicology since protecting the safety of the patient is of paramount importance and the fact that all drugs have side effects. This latter fact was noted as long ago as the 1500s by Paracelsus, who stated that "All things are poison and nothing is without poison, only the dose permits something not to be poisonous" (Wikipedia, 2008b).

Toxicology guidances cover a range of topics including species selection, duration of studies and types of studies that may be conducted in order to evaluate the potential for a drug to cause a particular toxicity. Drug metabolism and pharmacokinetics are frequently mentioned in such guidances, although often only in passing. The FDA guidance on "Safety Testing of Drug Metabolites" is an exception where the whole premise of the guidance relates to drug metabolism (US FDA, 2008j). The broader topic of preclinical considerations for regulatory submissions has been reviewed by Baldrick (2001) and Larson (2005).

Similarly, drug metabolism scientists provide very similar support to both nonclinical and clinical functions. This subsection of the chapter will focus on those items that are clearly nonclinical in nature (i.e. supporting animal pharmacology and toxicology). Items that clearly relate to support to clinical studies will be addressed later in the chapter. However, there will be some element of redundancy between the two subsections. It is recommended that the nonclinical and clinical subsections be read in concert in order to obtain the fullest understanding of the role of drug metabolism in nonclinical and clinical development.

### 4.2 Good Laboratory Practice

Good Laboratory Practice (GLP) is a concept that drug metabolism scientists should ensure that they are familiar with. Good Laboratory Practice embodies a set of principles that provides a framework within which laboratory studies are planned, performed, monitored, recorded, reported, and archived. These studies are undertaken to generate data by which the hazards and risks to users, consumers and third parties, including the environment, can be assessed for pharmaceuticals, agrochemicals, cosmetics, food and feed additives and contaminants, novel foods, and biocides. The GLPs

help assure regulatory authorities that the data submitted are a true reflection of the results obtained during the study and can therefore be relied upon when making risk/safety assessments. GLP regulations have been implemented globally and are very consistent from country to country, although there may exist some local requirements. Details regarding GLPs can be found in country-specific legislation and on regulatory agencies' web sites (European Parliament, 2004; US FDA, 2007j; OECD, 2007, 2008; U.S. Environmental Protection Agency, 2008).

The GLP regulations became effective in 1979 under 21 CFR Part 58.1 which applies to all nonclinical safety studies intended to support research permits or marketing authorizations for products regulated by the FDA (Dannan, 2003; US FDA, 2007a). The regulations were established after FDA inspected several research laboratories during the mid-1970s, which revealed serious problems with the conduct of safety studies submitted to the agency. The violations included poor record keeping and storage of raw data, lack of proper personnel training and handling of test facilities, and fraud. To apply the fundamental concepts of data quality and integrity, FDA requires that the regulations cover all operations in facilities conducting nonclinical studies and, most importantly, it requires that all facility operations and procedures be strictly documented.

The interpretation of the GLP regulations, particularly related to the subject of which studies should be compliant with GLPs, has engendered much debate. The regulations (in 21 CFR Part 58) state that the GLPs apply to "all non-clinical laboratory studies that support or intend to support applications for research or marketing permits for products regulated by FDA." The FDA defines nonclinical laboratory studies as "in vivo or in vitro experiments in which a test article is studied under laboratory conditions to determine its safety." As can be seen, this definition leaves room for interpretation. GLP compliance will be required for all toxicology studies (except mechanistic or investigative studies) and for the toxicokinetics support (including the bioanalysis) for these studies. Pharmacology (excepting safety pharmacology), drug metabolism, and pharmacokinetic studies do not need to be GLP-compliant since such studies do not directly provide information related to the safety of the test article. However, there will be some circumstances under which some ADME studies should be conducted in a GLP-compliant manner. The first is for the tissue distribution studies that are conducted for a pharmaceutical for the purpose of dosimetry (i.e. the confirmation that the dose of radioactivity that is proposed to be given to humans falls below the maximum allowable limits). Although this is an ADME study, it is clear that it is also a safety study and therefore should be conducted in accordance with GLP regulations. The second circumstance is for veterinary products or agrochemicals. For these products, studies to determine tissue accumulation and depletion should be GLP-compliant since the intent of these studies is to predict the exposure of humans to compound (and/or metabolites) due to ingestion of food that may contain residues of veterinary products or agrochemicals. The FDA has stated that "For drugs and feed additives used in food producing animals, metabolism studies come under the GLPs. In these cases, the studies are intended to define the tissue residues of toxicological concern as well as to estimate tissue depletion. Such studies on other regulated products are usually conducted as part of the pharmacological evaluation and would not be covered. However, metabolism studies on food additives are covered" (US FDA, 2007j). Despite the fact that most drug metabolism studies are not required to be GLP-compliant, they should still be conducted with high standards of quality ensuring reproducibility of data (i.e. "in the spirit of GLP"). Examples of other studies that are outside the scope of

GLP regulations include those utilizing human subjects, which are covered by good manufacturing practice (GMP) and good clinical practice (GCP) guidelines and efficacy studies (covered by GCP guidelines). Although clinical studies fall outside the scope of the GLP regulations, the bioanalytical support to these studies is conducted in compliance with the GLP regulations.

### 4.3 Bioanalytical

The measurement of parent drug and relevant metabolites (however they are defined) is an integral part of understanding the risk–benefit for a drug. This topic is relevant to both nonclinical and clinical studies and is mentioned in a large number of guidances. For brevity, not all guidances that mention the measurement of parent drug and/or metabolites will be covered here. Instead, unique or interesting aspects will be highlighted.

***4.3.1 Bioanalytical Method Validation*** In May 2001, the FDA issued a guidance document, "Bioanalytical Method Validation" that describes some of the criteria that should be applied to the validation of "bioanalytical procedures such as gas chromatography (GC), high-pressure liquid chromatography (LC), combined GC and LC mass spectrometric (MS) procedures such as LC-MS, LC-MS-MS, GC-MS, and GC-MS-MS performed for the quantitative determination of drugs and/or metabolites in biological matrices such as blood, serum, plasma, or urine" (US FDA, 2001b). This guidance also applies to other bioanalytical methods, such as immunological and microbiological procedures, and to other biological matrices, such as tissue and skin samples. The guidance was created after much discussion between the FDA, pharmaceutical companies, and contract research organizations (CROs). This dialog was initiated in 1990 at the first American Association of Pharmaceutical Scientists (AAPS)/Food and Drug Administration (FDA) Bioanalytical Workshop which focused on key issues relevant to bioanalytical methodology and provided a platform for scientific discussions and deliberations. The workshop, and the report, raised awareness of the need for validated bioanalytical methods for the regulatory acceptance of bioequivalence and pharmacokinetic data (Shah *et al.*, 1992). Just over eight years later (in January 1999), the FDA issued a draft guidance on bioanalytical validation based on the workshop report and recent FDA experience. A second workshop was held in January 2000 at which the draft guidance was discussed (Shah *et al.*, 2000). The guidance was finalized in May 2001.

The guidance addresses when a full or partial validation should be performed. The guidance notes that the circumstances that would dictate that a full validation would be required include "when developing and implementing a bioanalytical method for the first time" and "A full validation of the revised assay is important if metabolites are added to an existing assay for quantification." Partial validations are "modifications of already validated bioanalytical methods. Typical bioanalytical method changes that fall into this category include, selectivity demonstration of an analyte in the presence of specific metabolites."

The guidance advises that "the stability of the analyte (drug and/or metabolite) in the matrix during the collection process and the sample storage period should be assessed, preferably prior to sample analysis." This is of particular importance for potentially labile metabolites. The guidance notes that "the stability of analyte in matrix from dosed subjects (or species) should be confirmed." It should be noted that such

potentially labile metabolites could degrade during the sample processing or in the mass spectrometer itself. The conversion of a labile metabolite to the parent drug could potentially lead to the overestimation of the concentrations of the parent drug. The topic of potentially labile metabolites is one that will plague the determination of the metabolite profile in *in vitro* and *in vivo* studies. It should be noted that if metabolites are labile, they may not be detected and identified in the metabolism studies.

The guidance notes that "The accuracy, precision, reproducibility, response function, and selectivity of the method for endogenous substances, metabolites, and known degradation products should be established for the biological matrix. For selectivity, there should be evidence that the substance being quantified is the intended analyte." As mentioned previously, the selectivity of the assays is extremely important. The guidance states that "As with chromatographic methods, microbiological and ligand-binding assays should be shown to be selective for the analyte. Cross-reactivity of metabolites, concomitant medications, or endogenous compounds should be evaluated individually and in combination with the analyte of interest." The guidance includes a couple of interesting and important definitions:

> "Analyte: A specific chemical moiety being measured, which can be intact drug, biomolecule or its derivative, metabolite, and/or degradation product in a biologic matrix."
>
> "Selectivity: The ability of the bioanalytical method to measure and differentiate the analytes in the presence of components that may be expected to be present. These could include metabolites, impurities, degradants, or matrix components."

It is interesting to note that the guidance mentions "degradation product" and "impurities" since such substances are very rarely quantitated routinely, unless there was the specific need to qualify (in a toxicology study) a degradation product or impurity that was present in a given batch of clinical trial material.

It can be speculated that the guidance was written to include the possibility that it may be necessary to quantitate degradation products or impurities on a case-by-case basis or on a more routine basis in the future.

The development of bioanalytical assays for metabolites can be important for several reasons. These reasons include:

- Confirmation that the selected toxicology species have been exposed to "relevant" human metabolites to a sufficient extent (either after the administration of the parent drug or the metabolite itself)
- Confirmation (or refutation) of misdosing of control group animals in toxicology studies
- Develop sensitive assays for the evaluation of drug–drug interactions (DDIs) (change in metabolite concentrations as well as, or instead of, parent concentrations).

One element that is not discussed in great, if any, detail in the regulatory guidances is the selection of the matrix for the bioanalytical assay. The default matrix is plasma which is prepared from whole blood samples. This matrix is selected due to its ease of collection and the assumption that the plasma concentrations of drug and/or metabolites are in equilibrium with the tissue concentrations and therefore the plasma concentrations of the drug and/or metabolite can act as a surrogate for the tissue concentrations. Despite the caveats that can be made to this assumption, the exposure

to the drug in plasma can be correlated with efficacy or toxicity in the majority of cases. Serum is sometimes selected as the matrix instead of plasma, particularly for protein-based therapeutics. A word of caution should be made regarding the use of plasma as the matrix for the quantitation of drug. If the drug partitions into the red blood cells to a marked extent, an erroneous estimate of the plasma concentration of the drug and/or metabolites may be obtained if the blood sample undergoes hemolysis (i.e. the drug and/or metabolites are released from the RBCs after the collection of the blood sample). For drugs that partition heavily into the RBCs, consideration should be given to the use of whole blood as the matrix for the quantitation of drug and/or metabolites.

In May 2006, the third AAPS/FDA Bioanalytical Workshop was held. The report from this workshop, along with several related articles, was published in a themed issue of *The AAPS Journal* (Rocci *et al.*, 2006; Viswanathan *et al.*, 2007). The report (Scharberg, 2007) addressed "Determination of Metabolites During Drug Development." The report from the workshop states: "A draft FDA Guidance for Industry, entitled 'Safety Testing of Drug Metabolites' was issued in June 2005 by the Center for Drug Evaluation and Research (CDER)" (US FDA, 2005c). "There is general support from the pharmaceutical community for the idea that a more extensive characterization of the pharmacokinetics of unique and/or major human metabolites (UMMs) would provide greater insight into the connection between metabolites and toxicological observations. This information would be best generated by the use of rugged, bioanalytical methods applied at appropriate times in drug development. Characterization of UMMs should proceed using a flexible, 'tiered' approach to bioanalytical methods validation. This tiered approach would allow metabolite screening studies to be performed in early drug development using bioanalytical methods with limited validation, with validation criteria increasing as a product moves into clinical trials. A tiered validation approach to metabolite determination would defer bioanalytical resource allocation to later in the drug development timeline when there is a greater likelihood of drug success. As a minimum, the specifics of this tiered validation process should be driven by scientifically appropriate criteria, established a priori." The finalized guidance that was issued in February 2008 does not support the concept of deferring the validation of bioanalytical methods for metabolites until later in the development process (US FDA, 2008j). The topic of "Safety Testing of Drug Metabolites" is discussed in great depth in Section 4.5.

Despite the fact that there were many areas of consensus from the third workshop, consensus could not be achieved on all topics. One such area was "Cross-Validation Required When Using Different Strains or Sexes of a Species." The report states that "No decisive arguments came forward in support of this activity although it was agreed that there could be some differences in biological matrix originating from the different strains or sexes. The general trend of the debate was that validation experiments to address such differences should not be considered the norm and should be performed when there are method-related concerns that can be attributed to a specific strain or sex-related difference." Although not mentioned here, it is possible for a drug to be metabolized differently in males and females which may mean that the original evaluation of the selectivity of the assay may not have been sufficiently comprehensive if the validation was performed for one gender only. An obvious extension of this topic is the effect of age or provider of the animals. However, this topic did not appear to be addressed in this report.

***4.3.2 Analysis of Samples*** The potential for misdosing of control group animals was addressed in an EMEA guideline (Guideline on the Evaluation of Control Samples in Nonclinical Safety Studies: Checking for Contamination with the Test Substance) that was issued in 2005 (EMEA, 2005c). The FDA has not issued guidance on this topic and does not appear to be developing such guidance. There most probably is no need to issue such guidance since sponsors will most likely follow the EMEA guidance in order to be able to file for marketing application in a major market. Hence, this EMEA guideline is *de facto* also an FDA guidance. The guideline provides some clear rationale for the creation of this guideline. "Since 1999 when the first situation was identified, several applications have revealed that samples from control animals contained levels of test substance, or antibodies against the test substance, suggesting that contamination of control animals or samples from control animals with the test substance had occurred. The causes for contamination of controls samples may be multiple and in the majority of the situations the source of contamination has not been identified. The levels of contamination of controls samples were considered relevant/significant in some studies (e.g. carcinogenicity and reproductive toxicity) and led to their invalidation or impacted on a negative opinion for the Marketing Authorisation Application. A survey conducted by the European Federation of Pharmaceutical Industries Association (EFPIA) shows that contamination of controls with different levels of test compound often occurs during toxicology studies, regardless of the route of administration used, the dose levels and duration of treatment. The CPMP/ICH/384/95 Note for Guidance, Toxicokinetics: A Guidance Assessing Systemic Exposure in Toxicology Studies reads in its Note 8 that 'it is often considered unnecessary to assay samples from control groups. Samples from controls may be collected and then assayed if it is deemed that this may help in the interpretation of the toxicity findings, or in the validation of the assay methods.' Significant contamination of controls may lead to the invalidation of studies due to their poor overall quality or to inappropriateness of toxicokinetic data to allow adequate interpretation of non-clinical safety studies and human risk assessment."

The EMEA guideline states that "When significant contamination of control samples with the test compound at a level susceptible to impact on the validity of the study is observed, the sources of contamination should be investigated and identified. The applicant should make appropriate efforts to clarify whether contamination occurred in vivo or ex-vivo. Approaches may include, but are not exclusive to, identification of test substance in control tissues, identification of plasma metabolites in controls, or detection of antibodies against the test substance in case of a biotechnology product. Corrective measures should be taken accordingly." As can be seen, the existence of an assay for the quantitation of metabolites in plasma samples may be extremely useful to help prove that the presence of test article in plasma samples obtained from control group animals was due to either endogenous interference or *ex vivo* contamination. Despite the clear intent of this guideline, there can be great differences in opinion regarding how it should be implemented. The guideline states: "For all pivotal studies that include a toxicokinetic evaluation, control samples should be collected and analysed" but does not define what pivotal is. The term "pivotal" is not used in the ICHM3(R1) guidance. This term could be interpreted as being the longest duration toxicology study that is submitted with the marketing application. However, it is best defined as the GLP toxicology studies that are used to support a given clinical trial, which would imply that samples

should be collected and analyzed from control animals in all GLP-compliant toxicology studies.

The guideline further states that "in non-rodent studies, control samples should be collected and analysed in the same way as treated samples. In rodent studies, control samples should be collected and analysed in at least the general proximity of $T_{max}$ of the test substance." Unfortunately, this will add significant cost (potentially 25% for a traditional study design that includes one control group and three test article groups) to the bioanalytical portion of a non-rodent toxicology study, although the impact on the bioanalytical costs of a rodent toxicology study will be less due to the fewer number of samples that will require analysis. However, there will be a significant increase in the costs of the toxicology portion of a rodent toxicology studies in which samples are collected from control animals. For rodent studies, the toxicokinetics of the drug is most often conducted in "satellite" groups of animals. By using satellite animals, the sponsor is able to avoid the repeated blood draws required to collect the toxicokinetics samples having an impact on certain toxicology tests such as hematology and clinical chemistry assessments. In the past, the satellite groups did not include a control group. However, a control group is now frequently included in the satellite groups which leads to increased costs and increased use of animals. It should be noted that the number of animals in the control satellite group of a rodent study will be less than in a drug-dosed satellite group due to the fact that samples need only be collected in the region of the $T_{max}$. Although the guidance implies that one sample needs to be collected from a control group rodent, the best practice is most probably to collect two samples from the same rat so that the sponsor can distinguish between misdosed animals and samples that have been contaminated post-sampling. It should be noted that for mice, only one blood sample can be collected per mouse in a toxicology study. The ability to be able to distinguish between misdosing and post-sampling contamination or bioanalytical carryover is important. If a control sample is analyzed immediately after a sample containing high levels of drug, it may appear that the control sample contains drug due to sample "carryover" which may be due to small amounts of sample remaining on the injection needle prior to injection of the sample from the control animal. Care should therefore be taken when analyzing samples from control animals to minimize or eliminate sample carryover.

The report from the third AAPS/FDA Bioanalytical Workshop identified several areas of consensus but also illustrated areas in which consensus could not be achieved (Scharberg, 2007). Some of the items of interest to the drug metabolism scientist/pharmacokineticist in this report include the dynamic range of the standard curve and the analysis of incurred samples. The dynamic range for the standard curve was an area of consensus. The report stated that "If a narrow range of analysis values is unanticipated, but observed after start of the sample analysis, it is recommended that the analysis be stopped and either the standard curve narrowed, existing QC concentrations revised, or QC samples at additional concentrations added to the original curve prior to continuing with sample analysis. It is not necessary to reanalyze samples analyzed prior to optimizing the standard curve or QC concentrations." This statement reflects what had been the FDA's current practice over the years prior to the third workshop. The FDA had issued 483s (the form used by the FDA to record observations during an FDA inspection of a facility or study) when bioanalytical methods used to support bioequivalence studies had too great an analytical range. While laboratories like to make calibration ranges as wide as linearity permits, it is clear that the FDA expects calibration ranges to be equal to the actual range of concentrations found in

the study samples and that the low, medium, and high QC samples represent the actual concentration range of samples analyzed within the batch. This means that labs will have to take one of two approaches: validate a series of ranges, each with three levels of QC concentrations; or validate a very wide range but with multiple QC concentrations spread over the entire range. Another implication is that the pharmacokineticist will need to work very closely with the bioanalytical scientist in order to accurately predict the required dynamic range for the calibration curve and/or provide input regarding the extent of dilution that the samples may require.

During inspections by FDA investigators, large inconsistencies between original and repeated results from acceptable runs were found. For example, during validation, inter-day precision and accuracy was less than 6% but original and repeated $C_{max}$ values differed by 30–80%. Such discrepancies are often noted upon repeat analysis of samples that have been identified as requiring repeat analysis by the pharmacokineticist. The analysis of incurred samples is thus addressed:

> There are several situations where the performance of standards and QCs may not adequately mimic that of study samples from dosed subjects (incurred samples). Examples include metabolites converting to the parent species, protein binding differences in patient samples, recovery issues, sample inhomogeneity, and mass spectrometric ionization matrix effects. These factors can affect both the reproducibility and accuracy of the concentration determined in incurred samples. While these effects are often characterized and minimized during method development using QC samples, it is important to ensure that they are under control when the method is applied to the analysis of incurred samples.
>
> A proper evaluation of incurred sample reproducibility and accuracy needs to be performed on each species used for Good Laboratory Practice (GLP) toxicology experiments. It is not necessary for additional incurred sample investigations to be performed in toxicology species once the initial assessment has been performed. Incurred sample evaluations performed using samples from one study would be sufficient for all other studies using that same species. It is generally accepted that the chance of incurred sample variability is greater in humans than in animals, so the following discussion pertains primarily to clinical studies. The final decision as to the extent and nature of the incurred sample testing is left to the analytical investigator, and should be based on an in-depth understanding of the method, the behavior of the drug, metabolites, and any concomitant medications in the matrices of interest. There should be some assessment of both reproducibility and accuracy of the reported concentration. Sufficient data should be generated to demonstrate that the current matrix produces results similar to those previously validated. It is recognized that accuracy of the result generated from incurred samples can be more difficult to assess. It requires evaluation of any additional factors besides reproducibility upon storage, which could perturb the reported concentration. These could include metabolites converted to parent during sample preparation or LC-MS/MS analysis, matrix effects from high concentrations of metabolites, or variable recovery between analyte and internal standard. If a lack of accuracy is not a result of assay performance (ie, analyte instability or interconversion) then the reason for the lack of accuracy should be investigated and its impact on the study assessed. The extent and nature of these experiments is dependent on the specific sample being addressed and should provide sufficient confidence that the concentration being reported is accurate. The results of incurred sample reanalysis studies may be documented in the final bioanalytical or clinical report for the study, and/or as an addendum to the method validation report. In selecting samples to be reassayed, it is encouraged that issues such as concentration, patient population, and special populations (eg, renally impaired) be considered, depending on what is known about the drug, its metabolism, and its clearance. First-in-human, proof-of-concept in patients, special population, and bioequivalence studies are examples of studies that should be considered for incurred-sample

concentration verification. The study sample results obtained for establishing incurred sample reproducibility may be used for comparison purposes, and do not necessarily have to be used in calculating reported sample concentrations.

It will be interesting to see how the analysis of incurred samples will be implemented. Before the analysis of incurred samples becomes a routine part of the bioanalytical support for a program, it may be beneficial for the bioanalytical community to hold a dialog with the pharmacokinetic community, who are the customers for the data generated by the bioanalytical groups.

Although the drug metabolism scientists or pharmacokineticists may not be directly responsible for the bioanalytical support to a project, they should be aware of the FDA's close scrutiny of bioanalytical data since the bioanalytical data are a key element in their understanding of the ADME properties of a drug. As was mentioned earlier, much can be learned from warning letters issued by the FDA regarding the FDA's position on a topic and other company's approaches to drug development. This can be illustrated by the series of warning letters issued by the FDA to MDS Pharma Services in Montreal, Canada. There were several different issues that were identified in the initial warning letter that was issued in April 2004 as a result of the FDA's audit of a bioequivalence study of loratadine in July 2003 (US FDA, 2004e). In December 2004, a second warning letter was issued as a result of the inspection of a larger number of BE studies (US FDA, 2004e). The FDA asked MDS Pharma Services to conduct a five-year retrospective review of all BE studies to ensure that the issues that had been identified in the original warning letter were isolated incidents and not part of a larger pattern. A third warning letter was issued in May 2006 and clearly indicates that the FDA was not pleased with MDS Pharma Services' response to the FDA's original warning letter and the quality of its follow-through on this topic (US FDA, 2006k). In January 2007, the FDA sent a letter to pharmaceutical companies that had used MDS between January 2000 and December 2004 to perform bioanalysis in support of bioequivalence studies (Buehler, 2007). The letter noted that serious questions remain about the validity of bioequivalence data generated by MDS in studies during this time period. The letter stated: "In view of these findings, FDA is informing holders of approved ANDA(s) of these issues and would like to know what steps are being taken by you to assure the accuracy of data submitted in these applications and confirm the validity of MDS's analytical studies that were conducted from January 2000 through December 2004 and subsequently submitted to the FDA. Accordingly, with respect to these studies submitted in your applications, we recommend that within 6 months of the date of this letter you do one of the following, in order of FDA preference:

1. Repeat the bioequivalence studies.
2. Re-assay the samples at a different bioanalytical facility. For this option, the integrity of the original samples must be demonstrated for the frozen storage period.
3. Commission a scientific audit by a qualified independent expert, who is knowledgeable in the area of bioequivalence studies and bioanalytical data, and selected by your company rather than by MDS, to verify the results obtained by MDS.

In addition, because one of the agency's significant findings for the inspected MDS studies was the presence of anomalous results, we are recommending for all of the

above options that the blood/plasma level results obtained in the studies be compared to any published literature or other relevant information that is publicly available."

This clearly places a great burden on the sponsor. At the present time, it is unknown as to whether further deficiencies have been found and whether any approved generics have had to be removed from the market as a consequence of these deficiencies. It is clear that the sponsor should play a more active role in the oversight of the bioanalytical CRO (Scharberg, 2008).

### 4.4 Pharmacokinetics and Toxicokinetics

An early guidance from Europe addresses the topic of "Pharmacokinetics and Metabolic Studies in the Safety Evaluation of New Medicinal Products in Animals" (EMEA, 1994b). This guidance states: "Data on the levels of substance and metabolites in blood, body fluids, organs and in the excreta can be obtained by physical, chemical or biological methods. The investigator should justify the details of the methods used, their validity and reproducibility, including the specificity, precision and accuracy. (The study of the time course of its pharmacodynamic effects may provide useful additional information)." The guidance further states: "When using labeled substances attention must be given to the fact that the measured label in body fluids may not correspond to that of the unmodified substance, but may include labeled metabolites and conjugates." This point is a very important one that is often not appreciated by many pharmaceutical scientists outside of the ADME community and is a key limitation in the interpretation of data from tissue distribution studies. The final point of note is that "Attention should be given to the possibility of isotope exchange with endogenous compounds." In particular, care should be taken when conducting studies with compounds labeled with tritium, since the tritium can exchange with hydrogen atoms from water.

An assessment of the systemic exposure to parent drug and metabolites is a key part of a toxicology study and is important in putting any observed toxicities in the context of the exposure of humans to the parent drug and/or metabolites. This topic is covered in another ICH guidance, "ICHS3A—Toxicokinetics: The assessment of Systemic Exposure in Toxicity Studies" that was issued in 1995 (ICH, 1994b). This guidance addresses the possibility of establishing a maximum dose to be administered in the toxicology study if "toxicokinetic data indicate that absorption of a compound limits exposure to parent compound and/or metabolite(s)." However, it should be recognized that the use of a dose "cap" may not address local (GI) toxicity or the effect of metabolites generated by the GI mucosa or microflora for orally administered drugs. The guidance also advocates the use of metabolism, tissue distribution data, and plasma protein-binding data in the interpretation of the exposure data. For highly protein-bound drugs, the guidance suggests that the exposure to drug be expressed in terms of free (unbound) drug, although it does not address whether concentrations of free drug should be directly measured or can be calculated in other ways as is suggested in other guidances. Exposure to metabolites is addressed thus: "Measurement of metabolite concentrations may be especially important when documentation of exposure to human metabolite(s) is needed in the non-clinical toxicity studies in order to demonstrate adequate toxicity testing of these metabolites. It is recognized that measurement of metabolite(s) as a part of toxicokinetic evaluation serves only to assess exposure and cannot account for possible reactive intermediate metabolites." This latter statement emphasizes the importance of the availability of bioanalytical assays to quantitate metabolites (as appropriate). This guidance encourages the sponsor to perform

"concomitant toxicokinetics" (i.e. describe exposure at the same time that the toxicology study is conducted rather than at a later time). The guidance notes that the most commonly described PK parameters are $T_{max}$, $C_{max}$, and $AUC_{0\text{-}t}$. The guidance further states that half-life or $AUC_{0\text{-infinity}}$ is rarely calculated. This latter point is important since the standard sample collection scheme (five or six samples) is not usually sufficient for the calculation of these PK parameters, especially for drugs that are administered extravascularly (i.e. oral or subcutaneous). The inappropriate calculation of TK parameters (e.g. AUC, half-life, and clearance) is a common pitfall for the inexperienced pharmaceutical scientist.

The guidance also notes that "Increases in exposure may arise unexpectedly as a result of nonlinear kinetics due to saturation of a clearance process. Increasing exposure may also occur during the course of a study for those compounds which have a particularly long plasma half-life." It should be noted that the pharmacokinetics of most drugs are determined at a low dose level (e.g. 10 mg/kg) whereas the toxicokinetics of the drug may be described at doses that are often one to two orders of magnitude greater than the dose used in the PK study. This discrepancy may lead to a species as being viewed as inappropriate for use in toxicology studies for a given compound due to its low bioavailability and hence potentially high compound requirements. However, if the clearance is saturable (e.g. due to saturation of metabolism), it may be feasible to achieve adequate exposure to parent drug in that species. Further points made by the guidance include "Careful attention should also be paid to compounds which achieve high $C_{max}$ values over comparatively short time periods within the dosing interval" and "Conversely, unexpectedly low exposure may occur during a study as a result of auto-induction of metabolizing enzymes." The former point speaks to the appropriateness of selection of the time points at which blood is collected. The latter point pertains primarily to toxicology studies conducted in the rat. Auto-induction of metabolism will, as pointed out in the guidance, lead to low exposure to parent compound but will lead to high exposure to metabolites. This point will be further elucidated later on in this chapter.

### 4.5 Metabolism

One of the most controversial draft guidances that has been issued in the past decade has been the FDA draft guidance "Safety Testing of Drug Metabolites" that was issued for comment in June 2005 (US FDA, 2005c). This guidance was finalized in February 2008 (US FDA, 2008j). The history of the genesis of this guidance has been described previously elsewhere. Since approximately 30 months elapsed between the issuance of the draft and final guidances, the draft guidance was essentially being used as a *de facto* finalized guidance. This section of the chapter will compare the draft and final guidance. Both documents have the same title, are 11 pages in length, and contain a decision tree for determining when to evaluate the safety of a drug metabolite. Despite these similarities, the content of the two versions of the guidance including the content of the decision tree is quite different.

The draft guidance provided some key terms, some of which were retained in the finalized guidance and some of which were revised. The terms used in the draft guidance were:

> "Major metabolite—A metabolite in humans that accounts for plasma levels greater than 10% of the administered dose or systemic exposure, whichever is less.

> Metabolite—A compound derived from the parent compound through Phase I and/or Phase II metabolic pathways.
>
> Pharmacologically active metabolite—A metabolite that has pharmacological activity at the target receptor that is greater than, equal to, or less than the parent compound.
>
> Unique human metabolite—A metabolite produced only in humans."

The finalized guidance retained the terminology for "metabolite" and "pharmacologically active metabolite" but dropped the terminology for "major metabolite" and "unique human metabolite." These dropped terms were replaced with the term "disproportionate drug metabolite."

The terminology used in the final guidance is as follows:

> "Disproportionate drug metabolite—A metabolite present only in humans or present at higher plasma concentrations in humans than in the animals used in nonclinical studies. In general, these metabolites are of interest if they account for plasma levels greater than 10% of parent systemic exposure, measured as area under the curve (AUC) at steady state.
>
> Metabolite—A compound derived from the parent drug through Phase I and/or Phase II metabolic pathways.
>
> Pharmacologically active metabolite—A metabolite that has pharmacological activity at the target receptor. The activity may be greater than, equal to, or less than that of the parent drug."

The move away from the terms "major metabolite" and "unique human metabolite" is important since it emphasizes that metabolites are unlikely to be unique and also downplays the importance of whether a metabolite was major or not and softens the % exposure limit for safety testing by using the term "generally." The finalized guidance states that:

> An active metabolite may bind to the therapeutic target receptors or other receptors, interact with other targets (e.g. enzymes, proteins), and cause unintended effects. This is a particularly important problem when such a metabolite is formed in humans and not in animals, but the occurrence of a metabolite only in humans and not in any animal test species is rare. A more common situation is the formation of a metabolite at disproportionately higher levels in humans than in the animal species used in safety testing of the parent drug. This disproportionality stems from the typical qualitative and/or quantitative differences in metabolic profiles between humans and animals. If at least one animal test species forms this drug metabolite at adequate exposure levels (approximately equal to or greater than human exposure), as determined during toxicology testing of the parent drug, it can be assumed that the metabolite's contribution to the overall toxicity assessment has been established.

The last statement acknowledges the fact that the doses used in animal toxicity studies are far higher than those used in human clinical trials and consequently the safety of a given metabolite may have been assessed in the animal toxicity studies by virtue of the large doses used (and hence large exposures obtained), even if such a metabolite is present in humans on a higher percentage basis than in animals.

The terminology (or similar terminology) described earlier has been used in other guidances, although it is unknown whether the same terminology has the same or different meanings.

The intent and key conclusions of the guidance can be found in the introduction of both the draft and the finalized guidance: "This guidance makes recommendations on when and how to identify, characterize, and evaluate the safety of unique human metabolites and major metabolites of small molecule (nonbiologic) drug products. These metabolites may not be adequately assessed during standard nonclinical studies because they occur only in humans (unique metabolite), or at much higher levels (major metabolite) in humans than in the species used during standard nonclinical toxicology testing. If such metabolites are identified, they should be evaluated as early as possible during the clinical development program."

The draft guidance also states that "This guidance defines major metabolites primarily as those identified in human plasma that account for greater than 10% of drug related material (administered dose or systemic exposure whichever is less) and that were not present at sufficient levels to permit adequate evaluation during standard nonclinical animal studies." This statement does not appear in the finalized guidance. However, an additional statement has been added in the finalized guidance: "This guidance applies to small molecule nonbiologic drug products. This guidance does not apply to some cancer therapies where a risk-benefit assessment is considered. The Food and Drug Administration (FDA) is developing more specific guidance for the safety testing of drug metabolites in cancer therapies." This last statement is interesting, since such a draft guidance is not included in the current (March 2008) list of guidances that are under development for 2008 (US FDA, 2008f). This is not the only guidance that states that a separate guidance will be developed for oncology drugs (US FDA, 2006h). However, no such oncology-specific guidances have been created to date.

The draft guidance also gives a fairly succinct summary of the history of the use of metabolism data in safety evaluation:

> Traditionally, drug metabolites in general have not been routinely evaluated in cross-species safety assessments because their specific contribution to the overall toxicological potential of the parent drug has been unknown. With the availability during the past decade of technologies that can identify, measure, and characterize metabolites, we have gained a better understanding of the role metabolites play in drug safety assessment. Generally, we have used measurements of circulating concentrations of a parent drug in animals as an index of systemic exposure in humans. Quantitative and qualitative differences in metabolite profiles are important when comparing exposure and safety of a drug in a nonclinical species relative to humans during risk assessment. Based on data obtained from in vitro and in vivo metabolism studies, when the metabolic profile of a parent drug is similar qualitatively and quantitatively across species, we can generally assume that potential clinical risks of the parent drug and its metabolites have been adequately characterized during standard nonclinical safety evaluations. However, metabolic profiles and metabolite concentrations can vary across species, and there are cases when clinically relevant metabolites have not been identified or adequately evaluated during nonclinical safety studies. This may be because the metabolite being formed in humans was absent in the animal test species (unique human metabolite) or because the metabolite was present at much higher levels in humans (major metabolite) than in the species used during standard toxicity testing.

The concepts of these statements have been retained in the finalized guidance, although the wording has been changed slightly to read:

> Generally, drug plasma concentration and systemic exposure in the nonclinical studies are compared with systemic exposure in humans to assess the potential risks suggested by nonclinical findings and guide monitoring in clinical trials. This testing paradigm usually is

> sufficient when the metabolic profile in humans is similar to that in at least one of the animal species used in nonclinical studies, but metabolic profiles can vary across species both quantitatively and qualitatively, and there are cases when clinically relevant metabolites have not been identified or adequately evaluated during nonclinical safety studies. This situation can occur if the metabolite is formed only in humans and is absent in the animal test species or if the metabolite is present at disproportionately higher levels in humans than in the animal species used in the standard toxicity testing with the parent drug.
>
> It is not standard practice for drug metabolites to be evaluated separately in a cross-species safety assessment. As a result, their specific contribution to the overall toxicity of the parent drug has often remained unknown. This lack of appreciation of the role of metabolites in drug toxicity may be partly because of the inadequate sensitivity of the analytical methods used to detect and characterize metabolites derived from the parent drug. Technological advances during the past decade have greatly improved the analytical capabilities to detect, identify, and characterize metabolites and allow for better understanding of the role metabolites play in drug safety assessment.

The statement "This lack of appreciation of the role of metabolites in drug toxicity may be partly because of the inadequate sensitivity of the analytical methods used to detect and characterize metabolites derived from the parent drug. Technological advances during the past decade have greatly improved the analytical capabilities to detect, identify, and characterize metabolites and allow for better understanding of the role metabolites play in drug safety assessment" is magnanimous since the approach used to the detection of metabolites has not changed significantly for many decades, although advances in LC-MS/MS have improved our ability to determine the structure of metabolites and to detect metabolites at very low levels. It would be fair to state that the relative lack of evaluation of the toxicity of drug metabolites in the past has been due to a scientific blind spot that has caused many people to assume that the parent drug is responsible for both toxicity and efficacy. Similarly, the pharmaceutical industry has not paid great attention to the contribution of active metabolites to the overall therapeutic effect until recently.

The draft guidance placed a special emphasis on certain types of metabolites: "It is especially important to identify metabolites that may be unique to humans." The guidance also states that "compounds with the following characteristics are of particular concern and may warrant additional investigation:

- Narrow therapeutic indices
- Significant toxicity
- Significantly diverse metabolic profiles between human and nonclinical species
- Irreversible toxicity, or adverse effects not readily monitored in the clinic"

This statement helps the reader to understand some of the sources of the FDA's concerns. However, this section was not included in the finalized guidance. The finalized guidance does provide some statements regarding general considerations for nonclinical study designs. The finalized guidance states: "When designing a nonclinical study for a disproportionate drug metabolite, it is important to consider the following factors:

- Similarity of the metabolite to the parent molecule
- Pharmacological or chemical class

- Solubility
- Stability in stomach pH
- Phase I versus Phase II metabolite
- Relative amounts detected in humans versus the amounts detected in animals"

These are all factors worth considering, but the guidance does not give insight as to how such information would be used in designing the nonclinical study.

The FDA proposed a cutoff for exposure to a metabolite in humans above which additional safety testing may be required. The draft guidance states that:

> Generally, we recommend that metabolites identified in human plasma that account for greater than 10% of drug related material (administered dose or systemic exposure whichever is less) be considered for safety assessment. The rationale for setting the level at greater than 10% for characterization of metabolites reflects consistency with other FDA and EPA regulatory guidances (U.S. Food and Drug Administration 2002; U.S. Environmental Protection Agency 1998) and is supported by actual cases, described below, in which it has been determined that the toxicity of a drug could be attributed to one or more metabolites present at greater than 10% of the administered dose. Of the cases that follow, the last two are examples of a situation when a metabolite present at less than 10% caused toxicity. As a result, depending on the situation, some metabolites present at less than 10% should also be tested.

The draft guidance provides examples (halothane, felbamate, cyclophosphamide, and acetaminophen) to support the selection of 10% as a cutoff. The cutoff is one of the greatest sources of contention between the FDA and the pharmaceutical industry on this draft guidance. The PhRMA white paper proposed a cutoff of 25% for the definition of a major metabolite (Baillie *et al.*, 2002). In response to this proposed cutoff, Hastings *et al.* wrote a letter to the editor (Hastings *et al.*, 2003) that noted that "we want to emphasize that this threshold carries no regulatory authority or implications with respect to the need for safety assessment." They further pointed out that the EPA has established a limit of 5% of administered compound as the threshold for metabolite testing in the toxicological evaluation of pesticides, that ICHQ3A (R2) (Impurities in New Drug Substances) (ICH, 2006a) states that a contaminant/degradant should be "qualified" if present in drug product at >0.1% of the drug substance, and that the FDA Center for Veterinary Medicine has used 10% of total drug residue present in edible tissue as the definition of "major metabolite" in food-producing animals. These points clearly indicate the direction that the FDA was heading and therefore it was not surprising when the draft guidance was issued that the cutoff for the definition of a major metabolite was 10% of the exposure to total compound-related material. As an aside to the letter to the editor created by Hastings *et al.* (2003), it should be noted that ICHQ3A(R2) states that "Excluded from this document are extraneous contaminants that should not occur in new drug substances and are more appropriately addressed as Good Manufacturing Practice (GMP) issues." An "Extraneous Contaminant" is defined as "An impurity arising from any source extraneous to the manufacturing process" (ICH, 2006a).

An alternative approach to the use of a relative abundance was proposed by Smith and Obach (2005). Their proposal was that the absolute abundance of a metabolite in plasma or excreta should be combined with other information regarding the chemical structure of the metabolite. The approach of Smith and Obach carries some merit since

the use of relative abundance could potentially lead to some bizarre situations. For example, duloxetine is extensively metabolized *in vivo* in humans (Lantz *et al.*, 2003). The extent of metabolism is so great that only 2.7% of the plasma exposure to total compound-related material can be attributed to duloxetine. Interestingly, duloxetine itself does not meet either the FDA's or PhRMA's criteria for the definition of a major metabolite. There appear to be four potential major metabolites (using the FDA's criteria) in the plasma of humans dosed with duloxetine, although the exposure to these metabolites in plasma was not reported in this publication.

The finalized guidance adhered to the 10% limit, although the rationale for selecting this limit was changed slightly. The examples that were provided in the draft guideline are not present in the final guidance. The rationale for the selection of the 10% cutoff is now given as being harmonization with EPA guidances. The finalized guidance states that:

> Human metabolites that can raise a safety concern are those formed at greater than 10% of parent drug systemic exposure at steady state (Exposure should be at steady state unless there is some justification for a different measure of exposure. Comparison between human and animal exposure generally is based on area under the curve, but sometimes it may be more appropriate to use $C_{max}$). The choice of a level of greater than 10% for characterization of drug metabolites reflects consistency with FDA and Environmental Protection Agency guidances.

The final guidance provides some caveats regarding the 10% cutoff. It states that:

> Although the drug metabolite of toxicological concern usually is one circulating in plasma at greater than 10% of parent systemic exposure, other metabolites also can elicit safety concern. For example, a drug metabolite representing greater than 10% of urinary excretion relative to the bioavailable dose, or a human fecal metabolite in cases where biliary elimination is the predominant route of excretion in humans, may reflect potential localized renal or bile duct toxicity, respectively. Further characterization in these instances should be addressed on a case-by-case basis with the review division.

The draft guidance also provides some background information on drug metabolism and its role in the inactivation of drugs:

> Drugs entering the body undergo biotransformation via Phase I and Phase II metabolic pathways. Based on the nature of the chemical reactions involved, metabolites formed from Phase I reactions (e.g. oxidation, reduction) are more likely to be pharmacologically active, and require safety evaluation, than Phase II products (e.g. glucuronidation, sulfation). Although conjugated metabolites from Phase II reactions are generally pharmacologically inactive, more water soluble, and readily eliminated from the body, some are toxic. Sulfate and some glucuronide metabolites (e.g. acyl glucuronides of carboxylic acids) may retain pharmacological activity as well as toxicity of the parent drug and may require toxicological evaluation. Demonstration that a metabolite is pharmacologically inactive at the target receptor does not guarantee that it is not toxic, however. If the unique or major metabolites are suspected to contain a reactive functional group, it is important to assess the toxicity potential of these reactive metabolites. Chemically reactive intermediates are rarely detectable due to their short half-life, although stable products (i.e. glutathione conjugates) resulting from such intermediates can provide some indication of exposure to these potentially toxic species.

The finalized guidance does not use the same text, but embodies the same concepts, stating that:

> Metabolites that form chemically reactive intermediates can be difficult to detect and measure because of their short half-lives. However, they can form stable products (e.g. glutathione conjugates) that can be measured and, therefore, may eliminate the need for further evaluation. Phase II conjugation reactions generally render a compound more water soluble and pharmacologically inactive, thereby eliminating the need for further evaluation. However, if the conjugate forms a toxic compound such as acylglucuronide, additional safety assessment may be needed.

The points that are made here serve to illustrate that not all metabolites are inert. It is fairly well appreciated in the pharmaceutical industry that metabolites may be toxic, but the concept that metabolites may retain (or have enhanced) pharmacological activity is not as well understood. This latter point can be illustrated by the cholesterol absorption inhibitor, ezetimibe. After oral administration, ezetimibe is absorbed and extensively conjugated to a pharmacologically active phenolic glucuronide (ezetimibe-glucuronide) (Merck/Schering-Plough Pharmaceuticals, 2008). Ezetimibe is primarily metabolized in the small intestine and liver via glucuronide conjugation (a phase II reaction) with subsequent biliary and renal excretion. Minimal oxidative metabolism (a phase I reaction) has been observed in all species evaluated. In humans, ezetimibe is rapidly metabolized to ezetimibe-glucuronide. Ezetimibe and ezetimibe-glucuronide are the major drug-derived compounds detected in plasma, constituting approximately 10–20% and 80–90% of the total drug in plasma, respectively. Both ezetimibe and ezetimibe-glucuronide are slowly eliminated from plasma with a half-life of approximately 22 hours for both ezetimibe and ezetimibe-glucuronide. Plasma concentration–time profiles exhibit multiple peaks, suggesting enterohepatic recycling. Ezetimibe not only emphasizes the point that metabolites can be pharmacologically active and may be the major pharmacologically active component in the circulation, it also illustrates that not all meaningful metabolisms occur in the liver.

The timing of the generation of an understanding of the metabolism of a drug candidate is also addressed in the draft guidance: "The Agency recommends that—and this guidance encourages—attempts be made to identify as early as possible during the drug development process differences in drug metabolism in animals used in nonclinical safety assessments compared to humans." The guidance further states that "The discovery of unique or major human metabolites late in drug development can cause development delays and could have possible implications for marketing approval. Early identification of unique or major metabolites will allow for timely assessment of potential safety issues." These points are further elaborated as follows:

> The in vitro studies are generally conducted prior to the in vivo studies and provide an initial comparative metabolic profile. Results from these studies can assist in the selection of the appropriate animal species for toxicological assessments, should qualitative interspecies differences in metabolism be detected.
>
> Identifying a major metabolite in animals that does not exist in humans can mean that a toxicity observed in that animal species may not be relevant to humans. Conversely, identifying a human metabolite during clinical development that did not form at appreciable levels in animals would raise safety concerns because it probably was not evaluated in the nonclinical studies due to inadequate exposure. Additionally, when a potentially clinically relevant toxicity is observed during standard nonclinical studies, it is prudent to determine if metabolites contribute to that finding. In such cases, we recommend that the metabolites

> be synthesized and directly administered to the appropriate animal species for further pharmacological/toxicological evaluation. When qualitative and/or quantitative species differences in metabolite profiles are discovered, we also recommend investigation of different routes of administration or use of alternative animal species for safety assessments.

The guidance goes as far as stating that "Discovery of such a metabolite could delay development until the relationship between metabolite exposure and toxicity is understood."

These statements serve to emphasize the potential impact of metabolites on the safety assessment of a drug and the timelines for the development of a drug. The finalized guidance clearly states that the clinical program may be delayed if adequate evaluation of the nonclinical safety of disproportionate drug metabolites is not performed in a timely manner. Implicit in these statements is the fact that it is essential to determine the metabolism of an NCE early on in clinical development rather than at a later stage. The addition of the statement regarding the timing of such safety evaluation studies for drugs that are intended to treat life-threatening diseases is a great addition to the guidance. This enables the sponsor and the FDA to perform a true risk–benefit analysis rather than apply a "one size fits all approach" to the safety assessment of disproportionate drug metabolites.

The finalized guidance provides an overview of the general approaches for assessing metabolite safety. This section was not included in the draft guidance and represents a useful addition. The finalized guidance states:

> A metabolite identified in animals that is not present in humans can mean that a toxicity observed in that animal species, attributed to the metabolite, may not be relevant to humans. Conversely, a drug metabolite identified during clinical development that is not present in animal test species or is present at much lower levels in animals than in humans can suggest the need for further studies in animals to determine the potential toxicity of the metabolite. In such cases, two approaches can be considered to assess the drug metabolite. The first approach is to identify an animal species routinely used in toxicity studies that forms the metabolite at adequate exposure levels (equivalent to or greater than the human exposure), and then investigate the drug's toxicity in that species. The second approach, if a relevant animal species that forms the metabolite cannot be identified, is to synthesize the drug metabolite and directly administer it to the animal for further safety evaluation. In this approach, analytical methods that are capable of identifying and measuring the metabolite in nonclinical toxicity studies should be developed.
>
> The difficulties associated with synthesizing a specific metabolite as well as the inherent complexities that accompany its direct administration are acknowledged. Direct dosing of a metabolite to animals may lead to subsequent metabolism that may not reflect the clinical situation and thus may complicate the toxicity evaluation. Moreover, new and different toxicities may arise from administration of the metabolite that were not observed with the parent drug. However, notwithstanding these possible complications, identification and evaluation of the potential toxicity of the drug metabolite is considered important to ensure clinical safety, and the decision to conduct direct safety testing of a metabolite should be based on a comprehensive evaluation of the data on the parent drug and any information available for the metabolite. Appendix B provides three case examples when drug metabolites were formed at disproportionately higher levels in humans than in test animals used in the nonclinical studies and how the safety evaluation was approached. In Case 1, testing of the drug metabolite was not needed because the metabolite was adequately characterized in nonclinical toxicity studies with the parent drug. However, in Cases 2 and 3, the

drug metabolites had to be tested in toxicity studies by direct administration to the animal. In Case 3, the drug metabolite was pharmacologically inactive at the therapeutic target receptor but showed a unique toxicity not observed with the parent molecule.

The acknowledgment of the difficulty in synthesizing drug metabolites as well as the inherent complexities of the administration of these metabolites is welcomed. Additionally, the creation of examples greatly helps the reader understand the situations under which the FDA may reasonably expect that safety testing of drug metabolites be warranted.

The case studies in the finalized guidance are as follows.

"**Case 1.** From an initial mass balance study, a metabolite represented 1 to 2 percent of total radioactive dose in rat plasma, 5 percent in dogs, and 20 percent in humans (radioactivity of this metabolite in urine and/or feces was minimal). Based on the up-to-20 fold greater exposure in humans than in animals, nonclinical safety testing was recommended. However, the data generated in the general toxicology studies with the parent drug in the rat and dog suggested that the maximum doses tested produced metabolite exposures that represented at least the therapeutic exposure at the maximum recommended human dose. Also, the plasma concentrations of this metabolite measured in the in vivo genetic toxicity study, embryo-fetal development toxicity study, and carcinogenicity studies conducted with the parent drug provided adequate exposure and characterization of the metabolite. Therefore, no additional testing with the metabolite was needed."

This case implies the *in vivo* metabolism data were available relatively early in development since the exposure to the metabolite was addressed in the *in vivo* genetic toxicity study and the embryo-fetal development study. However, it is possible that the sponsor re-conducted these studies in order to show the exposure to the metabolite in these studies. This case also serves to emphasize the estimation, however crude, of the maximum recommended human dose early in the development plan. The case implies, but does not specifically state, that mean exposure data were used.

"**Case 2.** Two primary hydroxylated metabolites, M1 and M2, were shown to undergo further oxidation to form secondary metabolites M3 and M4 using hepatic microsomes and hepatocytes from human, monkey, rat, dog, rabbit, and mouse. This metabolic profile was later confirmed by in vivo data. The results showed the following:

- M1 and M4 were the predominant metabolites in human, monkey, and dog microsomes, whereas rat, mouse, and rabbit formed M2 and M3.
- M4 was formed in humans at 4 fold higher levels than parent drug, but M4 was formed at very low levels in rodents and only represented one-third of the parent exposure in monkey.

**$AUC_{0-24hr}$ at the Maximum Dose**

| | Human (MRHD) | Monkey | Rat |
|---|---|---|---|
| Parent | 1,800 | 15,000 | 12,500 |
| M4 | 7,700 | 5,000 | 135 |

MRHD = Maximum Recommended Human Dose

Severe drug-related and novel target organ toxicities were observed with the parent drug in monkeys but not in rats.

- M4 was pharmacologically inactive at the drug target receptors.

The following additional studies were done with M4:

- Subchronic toxicity study: 3 months in the rat
- Embryo-fetal development study in the rat
- In vitro genotoxicity testing: M4 was positive for point mutation and chromosomal aberration; the parent drug was negative
- Because of the positive genotoxicity, a carcinogenicity study that included M4 was recommended."

This case is interesting, and should have warranted further discussion in order to help the readers. The statement that M4 was not pharmacologically active against drug target receptors seems somewhat inconsistent with the statement that "Severe drug-related and novel target organ toxicities were observed with the parent drug in monkeys but not in rats." It could be concluded from these data that M4 is not toxic based on the observed toxicities in the monkey and apparent lack of toxicity in the rat. However, there may be additional data that are not included in this case study that would support the course of action taken. A more fully fleshed out description of the case study would have been more useful to the reader. It would have also been enlightening to mention whether M4 was generated by Aroclor-induced rat liver S9 or not.

"**Case 3.** M2 is a Phase I oxidative metabolite that formed up to 50 percent of parent drug exposure in humans, 10 percent of parent drug exposure in mice, 15 percent of parent drug exposure in dogs, and only trace amounts in rats. In vitro metabolism studies in these species supported the in vivo findings. Based on structure activity relationship (SAR) analyses, there was no reason to anticipate any difference or exaggeration in toxicity of the metabolite compared to parent molecule. The parent drug showed no significant toxicity or identifiable target organ of toxicity in any of the animal species tested in safety assessment studies. Because disproportionate human exposure was identified, further safety testing was needed. When M2 was tested in a short-term tolerance study in the dog, it produced unexpected and significant cardiotoxicity at all doses and in all of the dogs. M2 was pharmacologically inactive at the therapeutic target receptor."

This case also raises some interesting points. Firstly, the existence of SAR data, as well as the acceptance of such data by the FDA, is uncommon. It implies the existence of a validated hypothesis for the mechanism of toxicity, the existence of a series of analogs of the compound under development as well as the existence of an *in vitro* model that could be used to expeditiously and efficiently generate the SAR data. The observation that there was no significant toxicity observed in the dog with the parent compound could have implications for the validity of the safety assessment program for this compound. A successful program would have identified the potential human toxicities. A study that shows a lack of toxicity could be regarded as a failed study. The observation of an unanticipated toxicity with M2 raises questions regarding the relevance of the study based on the lack of toxicity of the parent drug and the ability of the dog to produce this metabolite. Additionally, the case does not mention whether such toxicities were observed in clinical trials. This case could benefit from a discussion

of how the FDA viewed the data and whether the program was able to progress or not.

Despite these reservations, the inclusion of case studies is a great step forward and case studies should be used, as appropriate, in future guidances.

The identification of metabolites is addressed in relatively great detail compared with other guidances. The draft guidance states:

> In vitro studies using liver slices, microsomes, or hepatocytes from animals and humans to identify the drug metabolic profile are generally conducted before initiation of clinical trials. It is important to also try to determine whether the concomitant use of drugs results in the inhibition or the induction of common metabolic pathways. In vivo metabolic profiles in nonclinical test species are generally available early in drug development, and their results may reveal significant quantitative and/or qualitative differences in metabolism across species. However, a unique metabolite may only be recognized after completion of in vivo metabolic profiling in humans. Therefore, we recommend the in vivo metabolic evaluation in humans be performed as early as feasible.

This statement emphasizes the limitations of *in vitro* studies and strongly advises that *in vivo* human metabolism data be generated as soon as is reasonable.

Once again, the finalized guidance uses different wording but retains the same general content. The finalized guidance states:

> Metabolite concentrations cannot be inferred by measurement of parent drug concentrations. The metabolic profile of the drug should be identified during the drug development process. This identification can be accomplished at different stages of development using in vitro and in vivo methods. In vitro studies can use liver microsomes, liver slices, or hepatocytes from animals and humans and generally should be conducted before initiation of clinical trials. In vivo metabolism study results in nonclinical test species generally should be available early in drug development, and their results will either confirm the results obtained from the in vitro studies or reveal quantitative and/or qualitative differences in metabolism across species. It is the latter situation that may pose a safety concern. Human in vivo metabolism studies usually have been performed relatively later in drug development, but we strongly recommend in vivo metabolic evaluation in humans be performed as early as feasible. Adequacy of exposure to drug metabolites that are present at disproportionately lower levels in animals used in nonclinical studies should be considered on a case-by-case basis. Generally, systemic exposure is assessed by measuring the concentration of the parent drug at steady state, in serum or plasma. However, when measurements cannot be made in plasma of the test species for any reason, verification of adequate exposure can be made in other biological matrices such as urine, feces, or bile. We encourage contacting the FDA early in drug development to discuss these issues.

This statement is not as clear as it could be. It accidentally implies that exposure to a metabolite can be assessed by measuring the concentrations of the parent drug at steady state. It also implies that plasma exposure to a metabolite in humans can be compared with the exposure to that metabolite in the excreta in animals, although it does not provide information as to how such a comparison can be made. Care must be taken when determining that steady state has been achieved in animals (and to a lesser extent in humans) since the achievement of steady state may not readily be determined based on the single-dose pharmacokinetics of the drug. Changes in clearance mechanisms with time due to aging or induction of the clearance mechanisms can confound the prediction of when steady state will have been achieved. It is common practice to

determine the exposure to drug and/or metabolites at the end of the safety assessment study, regardless of whether steady state may have been achieved before that point in time.

The design and conduct of human ADME studies are described in Section 5.7.10. Although this guidance does give some insight into how the metabolism of a drug could be determined, it does not give much guidance on the logistics of determining the metabolism *in vivo*. The use of radiotracers in drug development has been reviewed by Lappin and Tracey (2006). There are several ways in which the *in vivo* metabolism of a drug can be determined. The metabolism and mass balance are best determined using a radioactive tracer (usually $^{14}C$ but occasionally $^{3}H$). In most cases, the metabolism is determined after the administration of a single dose of drug. However, the exposure to metabolites after a single dose may not reflect the exposure after repeat since the exposure to parent drug and metabolites can change with duration of dosing (e.g. auto-induction or auto-inactivation of metabolism) and the age of the animals. Consideration should therefore be given to determining the *in vivo* metabolism after repeat-dose administration. One such design would be to administer unlabeled drug for a period of 7–14 days followed by the administration of a dose containing the radiotracer. It should also be noted that most *in vivo* metabolism studies are conducted at low doses (e.g. 10 mg/kg) since this is the dose range that the drug metabolism scientist or pharmacokineticist usually works at. However, this may not be appropriate since the purpose of identifying the metabolites in animals is to provide information for the safety assessment of the drug. Such studies are conducted at doses that are one or two orders of magnitude higher than the dose at which the PK studies are conducted. The metabolism and the pharmacokinetics of the drug may be markedly different at these higher dose levels and consequently, the *in vivo* metabolic profile may be markedly different. The drug metabolism scientist is strongly encouraged to consider conducting the radiolabeled *in vivo* metabolism studies in animals after repeat-dose administration and at toxicologically relevant doses. The use of the plural in the previous sentence is deliberate since the metabolism should be described at multiple dose levels since the metabolism could be dose-dependent. Once the relevant metabolites are identified, bioanalytical assays should be developed and the exposure to these metabolites determined in the subsequent toxicology studies (as necessary). The characterization of the *in vitro* metabolism of a drug is a task that is increasingly performed during the discovery phase of drug development. These studies tend to be conducted using non-radiolabeled material which can lead to limitations in the interpretation of the data from these studies. It is often not feasible to determine the relative percentages of metabolites and parent drug due to differences in ionization potential in the mass spectrometer. Additionally, the *in vitro* system is a static system and cannot take into account differences in rates of clearance of metabolites and parent drug *in vivo*. Therefore, data generated *in vitro* can be valuable in showing qualitative similarities or differences but should not be used to show quantitative differences. One product of the early investigation of the *in vitro* metabolism of a drug is a graphic that shows the metabolites of the drug and potential pathways and enzymes involved in the metabolism of the drug. This scheme may be very complicated, depending upon the amount of effort that the drug metabolism scientist has expended in identifying metabolites that are present in low amounts. Therefore, the metabolism *in vitro* may appear to be very complex and raise some potential concerns that are not realized once data on the metabolism of the drug *in vivo* are obtained. Another area of caution is the determination of the enzymes involved in the metabolism of the drug. Such a study may report

that 80% of the metabolism is mediated by CYP3A4/5, which may cause some concerns regarding the potential for DDIs or interindividual variability in exposure to the parent drug. However, such data are out of context since it is important to understand whether metabolism is the major route of clearance of the drug and the contribution of hepatic microsomal metabolism to total metabolism of the drug. To date, none of the guidances or publications give much insight into how to determine the extent of metabolism of a drug. Strategies for determining the metabolism of a drug *in vivo* are described later in Section 5.7.10, "Human ADME Studies."

The draft guidance also makes some general statements regarding species differences in metabolism and how exposure to metabolites is assessed:

> In general, systemic exposure to metabolites varies among species, and it is uncommon for humans to form unique metabolites. Therefore, identification of major human metabolites at levels higher than those measured in the test species used for toxicological assessment is of serious concern. For metabolites detected in humans as well as in nonclinical species (although at lower levels in the latter), adequacy of exposure should be considered on a case-by-case basis. Generally, systemic exposure is assessed by measuring the concentration of the compound in serum or plasma. However, when measurements cannot be made in plasma for any one or a number of reasons, measurements can be made in other biological matrices such as urine, feces, or bile. Noncirculating metabolites (i.e. excreted in bile, urine) are sometimes identified before clinical trials, but are not usually monitored. It is quite likely that excreted metabolite levels may be a more appropriate metric in many instances. For example, if Phase II conjugation products of a metabolite are present in the excreta, it can be assumed that systemic exposure to the metabolite has occurred. We recommend consulting the ICH Q3A guidance with regard to the development of analytical methods for measuring metabolites in selected matrices. If the systemic exposure in nonclinical species is equivalent to human exposure when measured in plasma and/or excreta, levels may be considered sufficient and alleviate the need for additional toxicity testing. We encourage contacting the Agency early in drug development to discuss these issues.

The latter part of the quoted statement can be used to illustrate the care that needs to be taken in the review of draft guidances. The guidance states: "We recommend consulting the ICH Q3A guidance with regard to the development of analytical methods for measuring metabolites in selected matrices." It should be noted that the ICHQ3A guidance is titled "Impurities in New Drug Substances" (ICH, 2006a). The title of the guidance is noted as a footnote in the "Safety Testing of Drug Metabolites" draft guidance (US FDA, 2005c). This guidance does not really address the measurement of metabolites in selected matrices. The only statement related to metabolites is the following statement: "Impurities that are also significant metabolites present in animal and/or human studies are generally considered qualified. A level of a qualified impurity higher than that present in a new drug substance can also be justified based on an analysis of the actual amount of impurity administered in previous relevant safety studies." It is unclear whether this was the intended reference or whether the FDA intended to cite a different guidance. It is unfortunate that this "potential error" could actually become an ingrained fact if the draft guidance does not become a final guidance and the draft guidance remains on the FDA web site. The existence of this "potential error" could raise the question as to how many other "potential errors" exist in the large number of draft guidances that exist. This error was corrected in the final guidance.

The guidance is one of the few guidances (draft or final) that encourage the use of *in silico* predictive tools for the prediction of potential toxicity. The guidance makes

that point thus: "It is now possible to identify the molecular structures of metabolites early in drug development. With the availability of computational software designed to predict activity relative to a known structure, the mutagenic, carcinogenic, or teratogenic potential of a drug or a metabolite can be evaluated as soon as a structure is identified. Although structure activity relationship analyses are not considered a substitute for actual testing, we encourage submission of the results from these analyses."

The guidance gives insight into the types of studies that should be conducted in order to evaluate the potential toxicity of a unique or major human metabolite. The guidance provides the following general considerations:

> When designing a nonclinical study for a unique or major metabolite, it is important to consider physicochemical characteristics of the metabolite, including solubility, permeability, extent of absorption, route of administration, and exposure. The indicated patient population, duration of use, and exposures at the therapeutic dose are also important considerations for the risk assessment. Another important consideration is the potential for biotransformation of directly administered human metabolites in animals as well as the presence of impurities in the synthesized metabolites.
>
> It is important to consider combined exposure to parent and pharmacologically active metabolites in safety assessments. A pharmacologically active metabolite can be more, equal, or less active than the parent drug at the target receptor. Similarly, a metabolite may cause toxicity by (1) eliciting exaggerated pharmacological effects via the target receptor, (2) activating receptors different from the parent drug target receptors, or (3) through nonreceptor mediated mechanisms (e.g. physico-chemical)."

The relevance of conducting studies with metabolites has been questioned. For example, if a metabolite is administered orally, there is the possibility that the metabolite will be poorly absorbed and therefore limit the validity of the study. The metabolite could be administered intravenously to avoid such issues, but this approach will still not address all potential concerns. Other concerns include that the metabolite may not readily be taken up into cells and may distribute differently throughout the body and that the metabolite undergoes sequential metabolism to a downstream product that is toxic, leading to differences in tissue-specific toxicity. Additionally, the testing of a reactive metabolite may be an exercise in futility, even if it could be synthesized, due to its likely short half-life and the fact that the safety of the metabolite will be tested out of context (i.e. not at the site of its generation *in vivo* from the parent drug. The challenges of detecting reactive metabolites will be addressed in the section of this chapter that covers drug metabolism in clinical trials and clinical guidances. The complicating factors in safety testing of drug metabolites have recently been addressed by Prueksaritanont, Lin and Baillie (2006). It should also be noted that by administering only the metabolite of interest, the sponsor will not be able to address the potential for additive or synergistic toxicity between the parent drug (or other metabolites) and the administered metabolite.

The statement that the metabolite may cause toxicity through activating receptors different from the parent drug has some large implications. This statement can be taken as implying that the metabolite should be evaluated for potential off-target effects in the same *in vitro* panel that the parent drug was assessed. It may not be sufficient to determine that the metabolite does not have the same intended pharmacology as the parent compound.

Both the draft and final guidances provided a list of studies that may be required for the assessment of the safety of metabolites. The information provided in both the draft and final guidances is essentially the same, although there were subtle differences. The finalized guidance states that "Good laboratory practice guidelines apply to the nonclinical studies with the drug metabolite designed to evaluate safety" (21 CFR Part 58).

The guidance describes the types of studies (general toxicity, genotoxicity, embryo-fetal development, and carcinogenicity studies) that should be conducted in order to assess the safety of the disproportionate drug metabolite. In essence, the studies are similar to those that would be conducted for the parent drug. However, the guidance does make some important statements that pertain particularly to the safety testing of drug metabolites. For example, in the section on General Toxicity Studies, the guidance states: "The toxicity of the drug metabolite should be investigated at multiples of the human exposure or at least at levels comparable to those measured in humans. We also recommend using the parent drug's intended clinical route of administration. However, with justification, other routes can be used to achieve sufficient exposure to the disproportionate metabolite. If the clinical route is oral, it is important to verify the stability of the metabolite in the stomach environment. It is crucial to gather toxicokinetic data from this type of study to ensure adequate exposure." The concern that the administered metabolite may not be stable in the stomach environment is a valid one. If the metabolite is not stable in the GI tract, then the animal may not be exposed systemically to the metabolite. It is unclear whether the concern is related to chemical stability, metabolism by GI microflora, or metabolism by the GI mucosa. It is unclear how the FDA would wish the sponsor to address the stability of the metabolite in the stomach environment. The guidance notes that "Sometimes the conduct of an embryo-fetal development toxicity study in only one species that forms the drug metabolite can be justified." However, the guidance does not address in which species the embryo-fetal development toxicity study should be conducted, but it would seem prudent to conduct the study in rat if at all possible due to the lower amount of chemical required compared with the rabbit and the fact that general toxicity studies would also be conducted in the rat. Finally, the guidance addresses carcinogenicity studies by stating that "Carcinogenicity studies should be conducted on metabolites of drugs that are administered continuously for at least 6 months, or that are used intermittently in the treatment of chronic or recurrent conditions when the carcinogenic potential of the metabolite cannot be adequately evaluated from carcinogenicity studies conducted with the parent drug. A single carcinogenicity study or an alternative bioassay should be conducted." Again, the guidance does not address which species (rat or mouse) should be used for the carcinogenicity assessment of a metabolite. Although a study in the mouse would use less chemical than a study in the rat, additional range-finding studies would need to be conducted in order to establish the doses for the carcinogenicity assessment whereas the general rat toxicity studies could be utilized to establish the doses for a rat carcinogenicity study.

The timing of the safety assessments is an important topic and is addressed in both the draft and final guidances. Both guidances state: "Early identification of unique human or major metabolites can provide clear justification for nonclinical testing in animals, assist in planning and interpreting clinical studies, and prevent delays in drug development." The draft guidance then proceeds to go into more detail, stating:

> Sponsors are encouraged to conduct in vitro studies to identify and characterize unique human or major metabolites early in drug development. If toxicity studies of a human metabolite are warranted, we recommend studies be completed and the study reports be submitted to the Agency before beginning large-scale phase 3 trials. In some cases, it may be appropriate for these nonclinical safety studies with unique human metabolites to be conducted before phase 3 studies; for example, (1) if the metabolite belongs to a chemical class with known toxicity; (2) if the metabolite has positive structural alerts for genotoxicity, carcinogenicity, or reproductive toxicity; or (3) if clinical findings suggest the metabolite or related compounds have indicated special clinical safety concerns, such as QT prolongation.
>
> To optimize and expedite development of drugs for serious or life-threatening diseases that lack an approved effective therapy, the number of nonclinical studies for the unique or major human metabolites may be limited on a case-by-case basis. We recommend sponsors contact the relevant review division to discuss such situations.

The duration of the toxicology studies that are required to support clinical trials of a particular duration or stage of development are covered by ICH guidances, ICHM3(R) and ICHS6. The final guidance is more succinct without the caveats of the draft guidance. It states that:

> If toxicity studies of a drug metabolite are warranted, studies should be completed and study reports provided to the FDA before beginning large-scale clinical trials. To optimize and expedite drug development for serious or life-threatening diseases other than cancer (e.g. ALS, stroke, human immunodeficiency virus), the number and type of nonclinical studies for the drug metabolites can be modified on a case-by-case basis for those drugs with major beneficial therapeutic advances, and for drugs for illnesses that lack an approved effective therapy. Sponsors should contact the appropriate review division to discuss such situations.

The FDA issued, in October 2007, a draft guidance related to the testing of drugs for the treatment of HIV. This guidance "Role of HIV Resistance Testing in Antiretroviral Drug Development" notes the importance of active metabolites by stating that "Any metabolite that exerts inhibitory activity should be delineated and its specific target shown." It is interesting that the draft guidance states "any metabolite" and does not mention "major metabolite" (US FDA, 2007k). Therefore, this draft guidance does not provide any clear guidance on the cutoff level for the characterization of the activity of human metabolites.

The hot nature of this topic can be seen by the relatively large number of publications in the scientific literature over the past two years. These publications have come from industry (Davis-Bruno and Atrakchi, 2006; Fura, 2006; Humphreys and Unger, 2006; Prueksaritanont, Lin and Baillie, 2006; Smith and Obach, 2006; Naito *et al.*, 2007). The publication by Davis-Bruno and Atrakchi (2006) includes two examples where safety testing of metabolites was performed. The SBA and EPAR for Nexavar (sorafenib) provide an illustration of the studies that may be performed to evaluate the safety of a metabolite. M-2 (an *N*-oxide) appears to be the major metabolite of sorafenib in humans. Since the metabolic profile of sorafenib in rats and dogs differed from that in humans, the sponsor conducted a four-week toxicology study with the metabolite M-2 in rats. An Ames test (for genotoxicity) was also performed with M-2 (US FDA, 2005h; EMEA, 2006i).

As can be gathered, the costs of evaluating the safety of metabolites can be very high, regardless of whether the assessment of the toxicity is performed as part of the assessment of the toxicity of the parent drug or as separate studies of the assessment of the toxicity of metabolites. In both situations, there will be the need to synthesize the metabolites for use as standards in the bioanalytical assay, the need to synthesize internal standards for use in the bioanalytical assay, the need to develop and validate methods for the metabolites, the need to re-evaluate the specificity of the existing bioanalytical methodology for parent drug, and the need to analyze the samples from toxicology or other studies. In the situation that the major human metabolite is generated in the animal species for the toxicology evaluation of the parent drug, it will be important to show exposure to these metabolites in the toxicology studies (particularly the chronic or carcinogenicity assessment studies). If identification of major human metabolites (i.e. the conduct of the human ADME study) is conducted late in the drug development process, the situation may arise where the chronic toxicology studies have been conducted but the exposure to the major human metabolites has not been assessed. In this situation, it may be possible to analyze banked samples from the chronic study for the metabolite(s) of interest, although this would be dependent upon the metabolites being stable in the matrix during their long-term storage. More likely, a bridging study would need to be conducted to describe the exposure in those species. Since the exposure to parent drug and metabolites can change with duration of dosing (e.g. auto-induction or auto-inactivation of metabolism), age of the animals, and source of the animals, it is possible that the sponsor may be requested to reconduct the chronic toxicology studies. Additionally, the sponsor may need to consider conducting studies to assess the exposure to metabolites in other studies such as the reproductive toxicology and safety pharmacology studies. It should be noted that the FDA's draft guidance states that the safety testing of metabolites should be conducted prior to large (phase III) clinical trials which would imply that the *in vivo* metabolism of the drug in humans should have been assessed long before that point in time.

In the situation where toxicology studies need to be conducted in animals dosed with the metabolite(s) of interest, there will obviously be additional costs for the synthesis of these metabolites and the conduct of the toxicology studies as well as the bioanalytical costs outlined herein. It should be noted that there may be additional bioanalytical costs if the administered metabolite is not stable by the route that it is administered and can only be quantitated indirectly (e.g. by quantitating a secondary metabolite). Synthesis of metabolites, particularly glucuronide metabolites, is a far from trivial task and may not be achievable in a reasonable time frame. It is speculated that the development of at least one drug has been discontinued due to difficulty in generating the metabolites required for the additional safety evaluation studies that were requested by a regulatory agency.

The scenarios noted here are complicated but it is possible that even more complicated scenarios exist. For example, it may be possible that there are two major metabolites identified in humans, only one of which is present in the toxicology species. This raises the question of what is the appropriate toxicity testing schema. Is it a study in which the unrepresented major human metabolite is administered to animals or is it a study that involves the administration of both the parent drug and the under-represented metabolite? The worst-case scenario may be that both types of study are required.

Given the large impact to budgets and timelines, it can be seen that it may be prudent to obtain an understanding of the *in vivo* metabolism of the drug in humans

at the earliest opportunity. Although such a study may incur short-term costs that could be viewed by some people as being unnecessary, such knowledge could be very cost-effective in the long term. It is also interesting to speculate whether extensively metabolized compounds should be bought into development since they may have the potential to generate "unique" human metabolites.

It should be noted that genotypic differences in drug-metabolizing enzymes (DMEs) can be found between strains of rats and within colonies of dogs (Morita *et al.*, 1998; Paulson *et al.*, 1999; Kamimura, 2006). Such differences may have a profound effect on the metabolism of certain drugs and hence complicate the design and interpretation of animal studies with these drugs. Less understood is the potential for phenotypic "drift" or variation that is not related to any genetic change. Robosky *et al.* (2005, 2006) recently published on this phenomenon. They noted that "stock Sprague-Dawley (Crl: CD(SD)) rats obtained from the Charles River Raleigh facility demonstrated a distinct endogenous urinary metabonomic profile that differed from historical control SD urine spectral profiles obtained over the past several years in our laboratory. In follow-up studies, the origin of the variant phenotype was narrowed down to animals of both sexes that were housed in one specific room (Room 9) in the Raleigh facility. It is likely that the two phenotypes are related to distinct populations of gut flora that particularly impact the metabolism of aromatic molecules." It is possible that differences in GI microflora may cause variability in the metabolism and pharmacokinetics of certain drugs, although this is an under-researched topic (Holmes and Nicholson, 2005; Nicholson, Holmes and Wilson, 2005; Li *et al.*, 2008; Qin, 2008).

## 4.6 Distribution

***4.6.1 Plasma Protein Binding*** The topic of plasma protein binding is an interesting one. There are no guidances for which plasma protein biding is the main topic. However, it is addressed in some guidances (US FDA, 2002a). Plasma protein binding is usually only addressed *in vitro* using plasma obtained from the blood of supposedly healthy individuals. These individuals are usually employees of the laboratory conducting the work and are deemed as being healthy because they are not being actively treated for a disease. Additionally, these volunteers are often asked to fast overnight prior to the blood draw. Plasma is prepared from the blood and the binding of the drug to this plasma is determined. As can be seen, such an evaluation will not give insight into the role of lipoproteins on the binding of the drug and hence the effect of variations in levels of lipoproteins on the pharmacokinetics of a drug. The role of lipoproteins in modifying the biological activity of hydrophobic drugs such as aminoglycosides and cyclosporine A has been reviewed by several authors (Wasan and Cassidy, 1998; Chung and Wasan, 2004; Wasan *et al.*, 2008). Lipoprotein-based formulations of anticancer drugs are being evaluated as potential ways by which to minimize some of the side effects of these drugs (Lacko *et al.*, 2007).

***4.6.2 Tissue Distribution*** The final topic to be discussed in this section is that of tissue distribution. The ICHS3b guidance, "Pharmacokinetics: Guidance for Repeated Dose Tissue Distribution Studies" was issued in 1995 (ICH, 1994a). This guidance states that "A comprehensive knowledge of the absorption, distribution, metabolism and elimination of a compound is important for the interpretation of pharmacology and toxicology studies. Tissue distribution studies are essential in providing information on distribution and accumulation of the compound and/or metabolites, especially in

relation to potential sites of action; this information may be useful for designing toxicology and pharmacology studies and for interpreting the results of these experiments." Interestingly, there is no such guidance for single-dose tissue distribution studies nor is there a clear suggestion that such studies are a crucial part of the drug development process in any regulatory guidance. This guidance describes circumstances under which repeated-dose tissue distribution studies should be considered. These include "When single dose tissue distribution studies suggest that the apparent half-life of the test compound (and/or metabolites) in organs or tissues significantly exceeds the apparent half-life of the elimination phase in plasma and is also more than twice the dosing interval in the toxicity studies, repeated dose tissue distribution studies may be appropriate" and "When steady-state levels of a compound/metabolite in the circulation, determined in repeated dose pharmacokinetic or toxicokinetic studies, are markedly higher than those predicted from single dose kinetic studies, then repeated dose tissue distribution studies should be considered." Despite the fact that this guidance has been in place for over a decade, very few repeat-dose tissue distribution studies are performed, even though the criteria mentioned may have been met for a given drug. The guidance gives input on the study design, "Information from previous pharmacokinetic and toxicokinetic studies should be used in selecting the duration of dosing in repeated dose tissue distribution studies. One week of dosing is normally considered to be a minimum period. A longer duration should be selected when the blood/plasma concentration of the compound and/or its metabolites does not reach steady state. It is normally considered unnecessary to dose for longer than three weeks." The guidance further states that "Consideration should be given to measuring unchanged compound and/or metabolites in organs and tissues in which extensive accumulation occurs or if it is believed that such data may clarify mechanisms of organ toxicity."

Despite the fact that there is no guidance that specifically relates to single-dose tissue distribution studies, such studies are commonly conducted as part of the nonclinical plan. The primary reason for conducting such studies is to determine the exposure of specific organs and organ systems to radioactivity in order to calculate the maximum amount of radioactivity that can be administered in a human ADME study. The human ADME study will be described in greater detail in the next section of this chapter. Occasionally, as mentioned earlier, the toxicologist may request that the drug metabolism scientist perform a tissue distribution study to aid in this prediction. When discussing accumulation or tissue distribution with a toxicologist, the drug metabolism scientist (or pharmacokineticist) should seek to understand the concern of the toxicologist (i.e. if they are concerned about high exposure of a drug in a particular organ) and explain to them the limitations of the techniques used to assess the tissue distribution of a drug.

Tissue distribution studies are relatively straightforward in design, reasonably easy to conduct, but can be very difficult to interpret. These studies are usually conducted in pigmented animals and involve an evaluation of the distribution at a range of time points in a limited number (often N = 1 / time point) of animals and usually only in male animals. The study may be a "tissue excision" study in which the tissues are removed from the body or a "quantitative whole body autoradiography (QWBA)" study. In the former type of study, either unlabeled drug can be administered with quantitation of drug concentrations by techniques such as LC-MS/MS or radiolabeled drug can be administered with the concentration of drug being determined by liquid scintillation counting (LSC) after combustion of the tissues. Tissue distribution studies utilizing unlabeled drug are uncommon, but this study design is appropriate when the

question being posed is "What is the concentration of the drug (or a particular metabolite) in a given tissue?" For the QWBA study, a radiolabeled drug is administered, the animals are frozen, embedded in a matrix, and the radioactivity determined in the tissue section by an imaging technique. A limitation of the QWBA studies is that the equipment used for imaging can accommodate rodents and small monkeys but is inadequate to accommodate large monkeys or dogs, unless the animals are further sectioned and placed in multiple detection cassettes. In the studies that utilize the administration of radioactivity, the concentrations of radioactivity represent the sum of the concentration of parent drug plus the concentration of all metabolites. Therefore, a tissue distribution study may be able to provide information as to which tissues the drug did not distribute to (at least, to a minimum level), but the study will not be able to provide information as to whether the radioactivity was present as parent drugs and/or metabolites. Therefore, it can be easy to assume that the organs that contain high levels of radioactivity will contain high levels of parent drug when, in reality, the radioactivity may represent metabolites and not parent drug. Another pitfall of tissue distribution studies is the assumption that organs that show high levels of radioactivity will show toxicity in animal studies. The resolution of autoradiography studies is usually limited to the tissue level, although it is possible to also use autoradiography to determine the location of drugs within cells (Stumpf, 2005). Recently, noninvasive imaging techniques, such as positron emission tomography (PET) or magnetic resonance spectroscopy (MRS), have become available for *in vivo* drug distribution assessment. These techniques will be able to provide distribution of compound-related material over time within a given animal or person which will provide an advantage over QWBA. However, these techniques have several disadvantages (cost, availability of equipment, limited number of drugs that can be used, and duration of the study) compared with QWBA and suffer from some of the same limitations as QWBA (low spatial resolution and inability to discriminate between bound and unbound drug or parent compound and metabolite) (Brunner and Langer, 2006).

In addition to the caveats regarding the lack of ability to discriminate between parent drug and metabolites, it should be recognized that the presence of parent drug and/or metabolites in a given tissue does not mean that there will be a toxic insult in that organ. A common conclusion drawn from the observation of the retention of radioactivity in a tissue is that the retained radioactivity is due to the covalent binding of the drug or metabolite which would raise a red flag regarding the potential for toxicity. However, it should be noted that the retained radioactivity may represent the incorporation of the drug and/or metabolite into lipids. The incorporation of compounds containing carboxylic acids into lipids has long been known but is often overlooked (Dodds, 1991, 1995). A common finding in tissue distribution studies conducted in pigmented animals is the retention of radioactivity by melanin in the eye. The correlation of binding of drugs or their metabolites to melanin and subsequent toxicity is equivocal, and it is possible that the retained radioactivity represents an animal-specific metabolite and therefore be a false-positive result (Salazar-Bookaman, Wainer and Patil, 1994; Leblanc *et al.*, 1998). Another consequence of the detection of binding of parent drug and/or metabolites to melanin is the need to conduct a photosensitivity study as described previously. Therefore, unless a tissue distribution study is required for dosimetry purposes, tissue distribution studies should be only conducted if there is a very specific question to be answered and then the appropriate study design should be selected in order to avoid the generation of potentially misleading information. For example, if the intent of the study is to determine the concentrations of parent drug

and/or a specific metabolite in a given tissue, an appropriate approach may be to dose the animals with unlabeled drug, remove the tissue (or tissues) of interest, and quantitate the concentrations of the analyte of interest using LC-MS/MS. Variations on this approach include the analysis of tissue sections by mass spectrometric techniques and the use of microdialysis (Joukhadar and Müller, 2005; Rohner, Staab and Stoeckli, 2005; Brunner and Langer, 2006; Chaurand, Corentt and Caprioli, 2006). The analysis of tissue sections will allow for spatial resolution of the distribution of the analyte within a tissue which may be important for correlating exposure to the analyte with a histopathological observation. Microdialysis offers the ability to obtain an estimate of the tissue distribution of a drug across time within a single animal. This latter technique can also be used in humans. It should be noted that both of these techniques have their limitations and are not widely used.

## 4.7 Toxicology

***4.7.1 General Toxicology*** The primary guidance that pertains to toxicology studies is the ICH guidance, ICHM3(R1), on "Non-Clinical Safety Studies for the Conduct of Human Clinical Trials for Pharmaceuticals" (ICH, 2000a). The purpose of this document is to recommend international standards for, and promote harmonization, of the nonclinical safety studies needed to support human clinical trials of a given scope and duration. Although it does not address what metabolism-related studies should be performed, it does make some statements that are pertinent to the drug metabolism statement: "Exposure data in animals should be evaluated prior to human clinical trials. Further information on absorption, distribution, metabolism and excretion in animals should be made available to compare human and animal metabolic pathways. Appropriate information should usually be available by the time the Phase I (Human Pharmacology) studies have been completed." The subject of the timing of the availability of ADME data will be addressed later in this chapter.

A concept paper for the revision of ICH M3 was issued by the ICH in September 2006 (ICH, 2006c). The rationale for the revision is that the approved text of the existing ICH-M3 guideline includes significant regional differences needing to be revisited in the light of accumulated information. In February 2008, the EMEA issued a draft guideline, "Guideline on Repeated Dose Toxicity" (EMEA, 2008c). The executive summary of this draft guideline states that "The purpose of testing toxicity after repeated dosing is to contribute to the development of safe medicinal products that need repeated administration to patients. General principles are provided on substance quality and excipients. The criteria discussed takes into account the choice of animal species, the size of groups and animal husbandry. Dose regimen, duration and route of administration should be selected based on the intended clinical use. Guidance is given on the parameters to be monitored during the in-life phase, and which special studies may be needed in case of a special activity of a certain medicinal product. A list of recommended tissues to be studied histopathologically is attached." This draft guideline complements the ICHM3 guidance by providing more detail regarding the conduct of the studies. It addresses the selection of the species and the safety testing of drug metabolites stating:

> Within the usual spectrum of laboratory animals used for toxicity testing, the species should be chosen based on their similarity to humans with regard to pharmacokinetic profile including biotransformation.

Exposure to the main human metabolite(s) should be ensured. If this can not be achieved in toxicity studies with the parent compound, specific studies with the metabolite(s) should be considered. When the product administered is a prodrug, its conversion to the active substance should be demonstrated in the species under study. Whenever possible, the selected species should be responsive to the primary pharmacodynamic effect of the substance. In certain cases for example when the pharmacodynamic effect by itself will cause toxicity, studies in disease models may be warranted.

***4.7.2 Safety Pharmacology*** Safety pharmacology is an area of toxicology that has been addressed in two separate ICH guidances, "ICHS7A—Safety Pharmacology Studies for Human Pharmaceuticals" and "ICHS7B—The Nonclinical Evaluation of the Potential for Delayed Ventricular Repolarization (QT Interval Prolongation) by Human Pharmaceuticals" that were issued in 2001 and 2005, respectively (ICH, 2000b, 2005a). ICHS7A states that "Generally, any parent compound and its major metabolites that achieve, or are expected to achieve, systemic exposure in humans should be evaluated in safety pharmacology studies. Evaluation of major metabolites is often accomplished through studies of the parent compound in animals. If the major human metabolites are found to be absent or present only at relatively low concentrations in animals, assessment of the effects of such metabolites on safety pharmacology endpoints should be considered. Additionally, if metabolites from humans are known to substantially contribute to the pharmacological actions of the therapeutic agent, it could be important to test such active metabolites. When the in vivo studies on the parent compound have not adequately assessed metabolites, as discussed above, the tests of metabolites can use in vitro systems based on practical considerations." The role of drug metabolism in the selection of the species used in the safety pharmacology testing is addressed thus: "Data from humans (e.g., in vitro metabolism), when available, should also be considered in the test system selection. The time points for the measurements should be based on pharmacodynamic and pharmacokinetic considerations. Justification should be provided for the selection of the particular animal model or test system." The timing of the availability of metabolism data is also addressed: "During early development, sufficient information (e.g., comparative metabolism) may not always be available to rationally select or design the studies in accordance with the points stated above; in such circumstances, a more general approach in safety pharmacology investigations can be applied." The emphasis on the relative lack of data on species differences in metabolism and the emphasis on the use of *in vitro* is a reflection of the fact that safety pharmacology studies should be conducted prior to the FIH study. The guidance also addresses the exposure to metabolites in the following manner: "Regardless of the route of administration, exposure to the parent substance and its major metabolites should be similar to or greater than that achieved in humans when such information is available." These statements appear to set up a circular argument. Since the objective is to conduct the safety pharmacology studies before the phase I study, there will be no information available on the *in vivo* exposure of humans to metabolites of the drug that is being administered. Such information will usually only be available after the human ADME study has been conducted. This then raises the possibility that the sponsor will need to conduct a battery of safety pharmacology studies prior to phase I and potentially need to reconduct selected studies based on the outcome of the human ADME study. At least, the sponsor may be required to conduct a PK study in the species used for the safety pharmacology studies in order to describe the exposure to selected human metabolites

and therefore retrospectively confirm the relevance of the original safety pharmacology studies.

ICHS7B addresses the topic of cardiovascular safety pharmacology studies and addresses the conduct of an integrated risk assessment. The guidance states that the objectives of the studies covered by this guidance are to "(1) identify the potential of a test substance and its metabolites to delay ventricular repolarization, and (2) relate the extent of delayed ventricular repolarization to the concentrations of a test substance and its metabolites." This guidance notes that "Cardiac cells and tissues have limited capacity for drug metabolism; therefore, in vitro studies using the parent substance do not provide information on the effects of metabolites. When in vivo nonclinical or clinical studies reveal QT interval prolongation that is not consistent with data from in vitro studies using the parent substance, testing metabolites in the in vitro test systems should be considered."

The integrated risk assessment can include information from "Pharmacokinetic studies, including plasma levels of parent substance and metabolites (including human data if available); drug interaction studies; and tissue distribution and accumulation studies." Since some of the studies are conducted *in vitro* in systems that have little, if any, drug metabolism capacity, the guidance implies that synthesis and testing of metabolites will need to be performed. The guidance encourages the sponsor to perform S7B nonclinical studies prior to the FIH study which would imply that additional studies may be required later on in development for certain human metabolites. The guidance also suggests that the use of metabolic inducers or inhibitors be considered for studies to determine the relative role of the parent drug and the metabolites in any observed positive results.

***4.7.3 Reproductive Toxicology*** Drug metabolism is mentioned to a limited extent in guidances that pertain to the assessment of reproductive toxicity. The area of reproductive toxicology in the large part evolved from the thalidomide tragedy of the late 1950s and early 1960s (Stephens and Brynner, 2001). It took many decades for the role of drug metabolism in the teratogenicity of thalidomide to be recognized (Arlen and Wells, 1996; Ando, Fuse and Figg, 2002).

The ICHS5(R2) guidance, "Detection of Toxicity to Reproduction for Medicinal Products & Toxicity to Male Fertility" contains some interesting statements that should pertain to each and every guidance: "To employ this concept successfully, flexibility is needed. No guideline can provide sufficient information to cover all possible cases, all persons involved should be willing to discuss and consider variations in test strategy according to the state of the art and ethical standards in human and animal experimentation. Areas where more basic research would be useful for optimization of test designs are male fertility assessment, and kinetic and metabolism in pregnant/lactating animals" (ICH, 2005b). The potential for the need to evaluate the toxicity of certain metabolites is described thus: "For secondary testing or further substance characterization other test systems offer the possibility to study some of the observable developmental processes in detail, for example to reveal specific mechanisms of toxicity, to establish concentration-response relationships, to select 'sensitive periods', or to detect effects of defined metabolites." The concept of body burden is used in this guidance and is defined as being "The total internal dosage of an individual arising from the administration of a substance, comprising parent compound and metabolites, taking distribution and accumulation into account." This definition implies that a tissue distribution study is performed in the pregnant animal.

In 2001, the FDA issued a draft guidance for reviewers on the "Integration of Study Results to Assess Concerns about Human Reproductive and Developmental Toxicities" (US FDA, 2001e). There are several statements of note in this draft guidance. These include "The evaluation should also compare animal and human pharmacodynamic effects, animal and human metabolism and disposition, animal and human pharmacologic and toxic effects, and drug exposures in animal studies in relation to the highest proposed dose in humans." The guidance further states that:

> Drug distribution, elimination, and biotransformation (pathways and metabolites) in the test species and in humans should be compared. Quantitative differences in metabolic/drug distribution profiles between the test species and humans are often seen, and may not have important implications and should not be overemphasized. Reproductive and developmental toxicities induced by compounds whose metabolic and distribution profiles are very similar in animals and humans increases concern for reproductive or developmental toxicity in humans. For compounds with highly dissimilar metabolic or tissue distribution profiles in animals and humans, there is less concern if the toxic effect seen in the test species can be attributed to a metabolite or tissue distribution profile not seen in humans. For any other scenario, concern is unchanged. When there are significant differences in drug distribution or metabolic profiles between several species, yet each test species demonstrates a positive signal for a reproductive or developmental toxicity, the toxicity is assumed to be attributable to the parent drug or a common bio-transformed product and concern is increased.

Finally, the guidance states that:

> Comparison of systemic drug exposure at the NOEL for the reproductive or toxicity class in the test species to that in humans at the maximum recommended dose is a critical determination. This comparison should be based on the most relevant metric (e.g., AUC, $C_{max}$, $C_{min}$, BSA [body surface area] adjusted dose). In general, there is increased concern for reproductive or developmental toxicity in humans for relative exposure ratios (animal: human) that are <10, decreased concern for exposure ratios >25, and no change in concern for ratios between 10 and 25. When applicable, the relative exposure ratio should consider both the parent compound and its metabolites. For example, it is appropriate to combine parent and metabolite when both are pharmacologically active and the activity relates to the reproductive or developmental toxicity.

The EMEA issued a draft guideline (Guideline on Risk Assessment of Medicinal Products on Human Reproduction and Lactation: From Data to Labelling) in 2006 that covers similar topics to the two FDA guidances described earlier (EMEA, 2006c). This draft guideline states that "Information about the excretion into milk of the active substance and/or its metabolites should be available" and "The exposure in pregnant animals measured by plasma concentrations of the compound and/or metabolites should be assessed."

The points in the FDA draft guidance and the EMEA guideline imply that the secretion of drug and/or metabolites should be evaluated in animals and/or humans, and that the metabolism of the drug in animals be determined prior to the determination of the exposure of the pregnant animal to the drug and/or metabolites. In most cases, the toxicology study is performed prior to the determination of the metabolism in the pregnant animal, if the metabolism is described in the pregnant animal. If the metabolism of the drug is determined in the pregnant animal, it is often only done in one species (usually rabbit). It is assumed that the metabolisms in the nonpregnant and

pregnant rat are identical, although that is not always the case. Additionally, distribution of drug and/or metabolites to the fetus is only occasionally performed and is often only performed in one species (it is assumed that the placental transfer is similar in both the rat and rabbit). The timing of when (in the course of the pregnancy) to address the metabolism of the drug and fetal exposure to the drug and/or metabolites is a subject that can engender great debate. It is well known that the metabolic capacity and capabilities of the fetus will change with the stage of pregnancy; that pregnancy can affect the clearance of the drug by the mother and that the placental transfer of the drug and/or metabolites can change with the stage of the pregnancy. Therefore, it can be seen that the description of the metabolism in pregnant animals is a complicated matter that can easily become resource- and time-intensive.

The utilization of animal studies to determine the risk of birth defects and other reproductive effects has been reviewed in depth by Brent (2004).

It should also be noted that none of the aforementioned guidances emphasize the possibility of a drug causing toxicity to the female when the drug has been administered to a male due to excretion of the drug in the semen. This route of excretion has been observed for several drugs such as antiviral drugs, chloroquine, and caffeine but the potential for a drug to be excreted in the semen is very rarely evaluated proactively (Pichini, Zuccaro and Pacifici, 1994; Cao *et al.*, 2007).

In summary, it appears fair to state that, at present, there appear to be gaps in our knowledge regarding how to determine the metabolism of drugs during pregnancy and that this may be an area that may warrant further discussion.

***4.7.4 Organ-Specific Toxicity*** Both the FDA and EMEA have issued draft guidances on hepatoxicity. The EMEA guideline (Non-Clinical Guideline on Drug-Induced Hepatotoxicity) was issued in 2006, with the comment period ending in January 2007. This version of the draft guideline is no longer available. The guideline was re-released for comments in January 2008 and it is anticipated that the guideline will be finalized in 2008–2009 (EMEA, 2008g).

The 2008 draft guideline states that "A drug can cause liver toxicity *via* several mechanisms. For instance, it can be directly acting or indirectly through reactive metabolites. The drug or its metabolites may cause liver toxicity after specific receptor binding, or reactive metabolites can react with hepatic macromolecules leading to direct cytotoxicity." It also notes that important signals related to hepatoxicity may be obtained from many areas, including bioaccumulation, expression of xenobiotic-metabolizing enzymes, and generation of reactive metabolites. The guidance further addresses these topics by stating that "In the absence of other signals of hepatotoxicity, accumulation of the drug in the liver is not a cause of concern. Investigation of metabolic pathways provides clues to the production of (possible reactive) metabolites. At therapeutic dose levels, drugs may affect expression of xenobiotic metabolizing enzymes (induction or inhibition) and potentially increase the concentration of metabolites. In the presence of hepatotoxicity signals, possible reactive metabolites should be identified."

Accumulation is a mass balance effect where input exceeds output. Factors involved in accumulation include selective binding of the drug to tissue molecules, concentration of lipophilic drugs in body fat, absent or slow metabolism of the drug, and slow excretion of the drug. Drug accumulation is a topic that can engender great debate between toxicologists and pharmacokineticists. The toxicologist may inquire about accumulation of drug in order to aid them in determining potential organs of toxicity. The toxicologist

may request that the drug metabolism scientist perform a tissue distribution study to aid in this prediction. When discussing accumulation or tissue distribution with a toxicologist, the drug metabolism scientist (or pharmacokineticist) should seek to understand the concern of the toxicologist (i.e. if they are concerned about high exposure of a drug in a particular organ) and explain to them any limitations of the techniques used to assess the tissue distribution of a drug. Tissue distribution studies and their limitations have been previously discussed in Section 4.6.2.

The EMEA guideline addresses the topic of the cytotoxicity of metabolites by stating that "If hepatotoxic signals occur *in vivo* (Step I) or signals are described in literature, parent compound or (reactive) metabolites if feasible, should be screened in *in vitro* mechanistic tests for potential hepatotoxicity." This may be easier said than done due to the difficulty in generating sufficient amounts of a reactive metabolite for testing in cytotoxicity assays. It should also be noted that the reactive metabolite may be administered to the cell culture medium and be detoxified before it enters the cell. Therefore, testing the cytotoxicity of a reactive metabolite will have many caveats. Several of these caveats will also apply to nonreactive metabolites and have been addressed earlier in the discussion of the safety testing of drug metabolites.

The first version of the draft guideline addressed the role of metabolism in immune-mediated mechanisms of hepatotoxicity. It stated that:

> The mechanism underlying immune-based liver toxicity is often triggered by the formation of reactive metabolite(s) that covalently bind(s) to proteins. However, as this event is considered insufficient to cause pathogenic immune reaction, a second co-stimulatory trigger is considered to be required (e.g. mild hepatic injury, inflammatory conditions, viral or bacterial infection). Immune-based liver damage could also result from an autoimmune response against hepatic antigens, due to loss of tolerance towards those antigens (secondary to cell damage or cross-reaction with the drug/drug-protein adduct), or from bystander damages to a non-antigen bearing hepatocytes. Moreover, a general inflammatory reaction in response to previous liver damage or unspecific activation of resident or accumulated leukocytes could occur. Inflammatory processes produce cytokines that can affect liver damage or repair. For example an activation Kupffer cells would in turn secrete pro inflammatory cytokines like IL-1, TNF alpha or nitric oxide (NO) which may induce apoptosis, stimulate cell growth or initiate inflammatory processes.

However, this mechanism of immune-mediated hepatoxicity is not discussed in the most recent version of the draft guideline. The reason for this change is unclear. The generation of autoantibodies against CYP isoforms has been observed for several drugs including cyclosporine, tacrolimius, hydralazine, and tienilic acid (Lytton *et al.*, 2002; Villeneuve and Pichette, 2004). Anti-CYP2D6 antibodies have been observed to be present in the serum of patients with type II autoimmune hepatitis (Miyakawa *et al.*, 2000). The role of bioactivation in drug-induced hypersensitivity has recently been reviewed by Sanderson, Naisbitt and Park (2006).

In October 2007, the FDA released its own draft guidance on hepatotoxicity, "Drug-Induced Liver Injury: Premarketing Clinical Evaluation" (US FDA, 2007l). This draft guidance, as can be seen from its title, is focused on the clinical aspects of hepatotoxicity. The guidance notes the limitations of animal models and states:

> The drugs that have caused severe DILI in humans have not shown clear hepatotoxicity in animals, generally have not shown dose-related toxicity, and, as noted, generally have caused low rates of severe injury in humans (1 in 5,000 to 10,000 or less). These reactions

thus appear to reflect host factors and individual susceptibility. Consequently, they have been termed *idiosyncratic*, meaning dependent upon the individual person's particular constitution. Whether they are the result of genetic or acquired differences has not yet been established, and to date no genetic, metabolic, or other characteristic has been found to predict severe DILI in an individual.

The timing of the evaluation of hepatoxicity is addressed thus: "Based on our experience, we recommend that the following analyses related to liver injury potential be carried out and included in an NDA or BLA, or included in an investigational new drug application when DILI is suspected and being evaluated."

The role of drug metabolism in hepatotoxicity is recognized by this draft guidance and is described thus: "The metabolism of a drug can have serious consequences for the safety profile of the drug. A drug may be metabolized to a hepatotoxic metabolite (e.g., acetaminophen, halothane, and isoniazid). Most hepatotoxic drugs have been oxidatively metabolized by the CYP450 system. Several in vitro methods are available to detect and quantify binding for a drug or its metabolites to liver proteins, including radiochemical and immunological methods." It should be noted that the guidance does not address exactly how the evaluation of the binding of the drug and/or its metabolites should be performed.

In addition to the aforementioned guidances, there have been several recent publications on the topic of drug-induced hepatoxicity by both industry and the FDA (Senior, 2007; Tang, 2007). The FDA has a web page dedicated to drug-induced liver toxicity (US FDA, 2008b).

***4.7.5 Genotoxicity Testing*** Genotoxicity has been addressed in several regulatory guidances, including the ICHS2A guidance "Specific Aspects of Regulatory Genotoxicity Tests for Pharmaceuticals" that was issued in 1996 (ICH, 1995). The guidance states that, in general, only one gender (male) of rodent (rat or mouse) needs to be evaluated in the *in vivo* bone marrow micronucleus test or other established *in vivo* tests. The guidance further states that "If there is a clear qualitative difference in metabolites between male and female rodents, then both sexes should be used." The clear implication is that the sponsor should have an understanding of the whether there is a gender-related difference in the metabolism of their drug before conducting the *in vivo* study. This study is not required before the phase I clinical studies unless there is a positive result in an *in vitro* genetic toxicity test. In any circumstance, the *in vivo* study should be conducted prior to phase II studies. The determination that there are gender-related differences in metabolism is usually made based on the observation of gender-related differences in exposure to parent drug in general toxicity studies rather than on *in vivo* metabolism studies conducted in both genders of rodent. It should also be noted that, even if there is clear evidence of gender-related difference in metabolism, the sponsor does not always perform the *in vivo* study in both genders.

Metabolism plays a key role in the *in vitro* assessment of the genotoxicity potential of a drug. The standard bacterial reverse mutation assay (Ames assay) is conducted in the presence or absence of a metabolic activation system. The most commonly used metabolic activation system is Aroclor-induced male rat liver S9. Interestingly, the sponsors will very rarely, if ever, conduct studies to describe the metabolism of their drug in this test system. Instead, it is assumed that by using an induced system, a large number of metabolites will have been generated. Very few, if any, Ames assays are performed using human liver S9, although it can be argued that this would be the more

relevant metabolic activation system. The reason for the use of rat liver S9 rather than human liver S9 is that the rat system will have less lot-to-lot variability than the human system and hence give more reproducible results.

The guidance addresses the role of metabolism in understanding negative results in the *in vitro* assays. It states: "For in vitro negative results, special attention should be paid to the following considerations (the examples given are not exhaustive, but are given as an aid to decision-making): Does the structure or known metabolism of the compound indicate that standard techniques for in vitro metabolic activation (e.g., rodent liver S9) may be inadequate? Does the structure or known reactivity of the compound indicate that the use of other test methods/systems may be appropriate?" However, if a sponsor obtains a negative result in the *in vitro* test, they tend to move the program forward rather than spend time in determining whether the result was a false negative or not.

At times, the *in vitro* and *in vivo* tests can produce discordant results. This guidance gives some possible explanations for the discordant results: "They include: (i) An active metabolite produced in vitro may not be produced in vivo, (ii) an active metabolite may be rapidly detoxified in vivo but not in vitro, and (iii) rapid and efficient excretion of a compound may occur in vivo." The guidance appears to clearly emphasize the role of drug metabolism in genotoxicity testing and that the drug metabolism scientist and toxicologist should work closely together in this area.

There are a few additional words of caution to be made regarding drug metabolism in genotoxicity testing. One element of the *in vitro* battery of tests includes the evaluation of the potential for the induction of chromosomal aberrations in cellular systems. It should be noted that these systems may not have a large capacity for drug metabolism. In fact, the metabolism of drugs by these cellular systems is very rarely, if ever, evaluated. Additionally, the stability of the drug in the buffer or media used for these studies is very rarely, if ever, determined. This last caveat would also apply to other *in vitro* studies.

In 2006, the ICH issued a final concept paper for a revision to ICHS2 (ICH, 2006d). This concept paper does not contain any mention of the role of metabolism in the activation of genotoxic chemicals.

***4.7.6 Carcinogenicity Assessment*** Drug metabolism is discussed to a small extent in several guidances issued by the FDA and ICH on the subject of the assessment of carcinogenic potential for a drug (list guidances). The FDA's 2002 guidance on "Carcinogenicity Study Protocol Submissions" states that "Exposure (steady state $AUC_{(0-24)}$) data should be provided for the parent drug and for the major metabolites from clinical trials conducted at the maximum recommended human dose (MRHD) or other appropriate human reference dose if the MRHD exposure data are unavailable" (US FDA, 2002a). It further states that "Where pharmacokinetics differ significantly between genders, data from males and females should be reported separately, as a gender difference can modify the study approach and conclusions. It is our experience that gender differences occur rarely." The guidance also states that "Plasma protein binding data should be provided for the parent drug and the major human metabolites (to the extent feasible) in the rodent test species over the range of concentrations encountered in the dose-rangefinding experiment and in humans at concentrations encountered in clinical trials conducted at the reference dose."

The availability of metabolism data is addressed as follows: "Metabolic profiles should be provided for the drug in humans and in the species employed for assessment

of carcinogenic potential. In cases where in vivo data are unavailable, in vitro data can be used (see the recommendations in ICH S1C)."

The statements in this guidance have some interesting implications. Since carcinogenicity assessment studies take approximately three years from first dose to final report, they are often critical path to the submission of the marketing application. Therefore, the protocol for the carcinogenicity study may need to be submitted to the FDA before the maximum recommended human dose has been determined. In this case, the sponsors would base their calculations on the highest dose that will be evaluated in the phase III clinical trials. Another implication of these statements is that the exposure to the major human metabolites has been determined prior to the submission of the carcinogenicity protocol to the FDA. Therefore, the implicit assumption is that the human ADME has been conducted, any major metabolites identified, bioanalytical methods to quantitate such metabolites have been developed and validated, and the exposure to these metabolites determined at the MRHD prior to the submission of the carcinogenicity protocol. If this assumption were true, then this would mean that the human ADME study should have been conducted relatively early on in the course of the development program. However, the guidance does allow for the provision of *in vitro* data in lieu of *in vivo* data. Caution should be employed in using this approach since the use of such data relies on the assumption that the *in vitro* data are qualitatively and quantitatively similar to the *in vivo* data. If *in vitro* data are used in lieu of *in vivo* data, the sponsor runs the risk of discovering, at a later date, that the *in vivo* data do not support the doses selected for the carcinogenicity assessment study, with the possible need to repeat this very expensive and time-consuming study. The timing of the human ADME study will be covered in the sections on drug metabolism in clinical studies (Sections 4.5 and 5.7.10).

This guidance uses the term "major" metabolite in a manner that is subtly different from other guidances. The guidance states: "For determining the appropriateness of using AUC/limit dose approaches, major metabolites are defined as metabolites that, if excluded from the analysis, significantly change the comparison ratios between species. Data (point estimates as well as individual animal values) should be reported separately for males and females from the same strain as proposed for the bioassay." This definition is different from that used in the guidance "Drug Safety Testing of Metabolites" (US FDA, 2008j). This guidance is one of the few guidances that addresses the potential for strain-related differences in the metabolism of a drug and notes that the metabolism should be determined in the appropriate strain. Although not specifically mentioned in this guidance document, the metabolism of the drug should be assessed in the transgenic mouse if they are being used for carcinogenicity assessment. Simply evaluating the metabolism in the background strain of mouse may be insufficient due to potential differences in the metabolism of the drug in the parental and transgenic strain of mouse (Ariyoshi *et al.*, 2001; Sanders *et al.*, 2001).

The ICH guidance, "Dose Selection for Carcinogenicity Studies of Pharmaceuticals & Limit Dose S1C(R1)," is one of the few guidances to acknowledge that metabolism after a single dose can be different to the metabolism after repeat-dose administration (ICH, 2005c). The guidance states: "Changes in metabolite profile or alterations in metabolising enzyme activities (induction or inhibition) over time, should be understood to allow for appropriate interpretation of studies." Auto-induction of metabolism may markedly change the exposure to metabolites upon repeat-dose phenomenon. Auto-induction of metabolism of a drug is not an uncommon phenomenon and is often observed in rats for compounds that are CYP3A substrates. Occasionally, this

phenomenon is observed in dog, monkey, and human. Therefore, the assessment of the exposure to metabolites in the rat after a single dose may underestimate the true exposure to metabolites and may lead the sponsor to make an incorrect assessment that a given human metabolite is under-represented in the rat, leading to the potential for the need for safety testing of that metabolite or incorrect dose setting for the carcinogenicity assessment studies. The concept of linearity in the metabolism of a drug is also discussed in this guidance. The guidance states that "The mid and low doses selected for the carcinogenicity study should take into account saturation of metabolic and elimination pathways." The implication of this statement is that the metabolism of the drug may need to be assessed at multiple dose levels rather than the one dose level that is traditionally used.

The guidance also provides an appropriate caveat regarding the correlation of plasma and tissue exposure to drug or metabolites by stating: "It is recognised that the doses administered to different species may not correspond to tissue concentrations because of different metabolic and excretory patterns. Comparability of systemic exposure is better assessed by blood concentrations of parent drug and metabolites than by administered dose. The unbound drug in plasma is thought to be the most relevant indirect measure of tissue concentrations of unbound drug. The AUC is considered the most comprehensive pharmacokinetic endpoint since it takes into account the plasma concentration of the compound and residence time in vivo." The guidance does not address how unbound drug concentrations are determined (i.e. directly or indirectly). The standard practice is to determine total drug concentrations and then correct the concentrations for the extent of plasma protein binding. This is a relatively simple task if the drug has concentration-independent plasma protein binding. However, if the plasma protein binding is concentration-dependent, the task can be a very laborious one. Additionally, the relationship between plasma protein binding and concentration should be well established in order to provide an accurate assessment of the exposure to unbound drug. This guidance appears to use the terms "blood" and "plasma" interchangeably and the reader should not necessarily assume that this guidance advocates the quantitation of drug in whole blood samples rather than in plasma samples.

***4.7.7 Combination Drugs*** The subject of drug combinations was addressed in an FDA guidance that was issued in 2006 (US FDA, 2006h). This guidance, "Nonclinical Safety Evaluation of Drug or Biologic Combinations," is not intended to cover fixed dose combination and co-packaged drug products for the treatment of HIV, since these drugs are covered in a separate guidance. Additionally, the guidance states that the Agency is developing a draft guidance specifically addressing oncologic drug combinations. However, this guidance has not yet been issued and does not appear on the list of guidances that the FDA plans to work on in 2007. Drug metabolism is mentioned in this guidance in the context of the possibility of a PK interaction. The guidance states: "A PK interaction can manifest in several ways, some of which can be monitored in vivo and some of which cannot. One drug/biologic product may alter the absorption or excretion of another product, change its distribution into one or more tissues, or change its pattern or rate of metabolism. Drugs may compete for serum protein binding, resulting in an increase in circulating free levels and tissue uptake of one drug. Metabolic interactions have been seen for combinations of two biologics." The guidance further states: "The FDA recommends that sponsors conduct combination PK/ADME studies to assess the potential for a PK interaction between the drugs/biologics. These data are valuable for supporting the safety profile and guiding the

drug/biologic development process. The FDA further recommends that PK/ADME combination studies (e.g., in vitro drug metabolism studies) be conducted early in drug development. The FDA encourages sponsors to evaluate serum protein binding and to monitor plasma concentrations of each drug in the toxicology studies. It may be possible to collect PK data as part of the toxicology studies instead of in a separate study." One of the many challenges in the development of drug combinations is the fact that many of the drugs that are used in the combinations may have been approved for marketing several decades ago and much of the basic ADME data that are contained in present day submissions will not exist. Therefore, a sponsor that is developing a drug combination may be required to determine the metabolism of these older drugs. The development of combination therapies is becoming more common and therefore this guidance will begin to assume an ever-increasing importance. A recent example of an approved combination therapy is Zetia (ezetimibe and simvastatin). Although this drug was approved prior to the FDA's guidance being issued, the SBA for Zetia is informative regarding the approach taken to the development of a combination therapy (US FDA, 2002e). In 2007, the FDA approved Janumet which is a combination of the recently approved DPP-4 inhibitor sitagliptin and the generic drug metformin (US FDA, 2007p).

In approximately the same time frame, the EMEA issued a draft guideline "Guideline on the Non-Clinical Development of Fixed Combinations of Medicinal Products" (EMEA, 2005f). The guideline was finalized in January 2008 with an effective date of August 1, 2008 (EMEA, 2008e). This guideline covers very similar ground to the FDA's guidance.

This guideline does not make mention of metabolism, other than related to the selection of the species to be used in the toxicology studies, saying that "For general toxicity studies usually one species only may be sufficient. The selection of the species used should be scientifically justified based on pharmacodynamic responsiveness to both components, pharmacokinetics, metabolism, target organs and sensitivity to the toxicological aspects." The rationale for combination therapy is described thus: "The rationale for a combination therapy and a fixed combination is often that pharmacological or pharmacokinetic interactions, leading to improved efficacy or safety profiles, compared with the single components. The main aim with the non-clinical studies to support the clinical development of a fixed combination is to characterise potential additive, synergistic, potentiation or antagonistic effects of the compounds when used together and to characterise the pharmacology, pharmacokinetics and toxicology of the combination under development. Additionally, these studies may identify toxicity unique to the combination and previously not seen for either compound administered alone."

The finalized guideline addresses fixed combination of compounds already approved as free combination therapy by stating that "When the fixed combination under development includes compounds for which there is sufficiently documented human experience of their individual and combined use, safety studies in animals are in general not required (CPMP/EWP/240/95)."

The EMEA document cited here is the "Note for guidance on fixed dose combination medicinal products" that was issued in 1996. This document does not appear to be available on the EMEA web site. However, the EMEA did issue a draft guideline in February 2008 (Guideline on Fixed Combination Medicinal Products) that has the same document reference number (CPMP/EWP/240/95) as the 1996 note for guidance

(EMEA, 2008b). However, this draft guideline does not refer to the 1996 note for guidance.

The guideline further states that "If the combination contains compounds from the same classes, as other compounds in well established combinations, for which there is considerable clinical experience and there are no pharmacokinetic interactions identified, further non-clinical studies may not be needed. A justification should in such case be provided. However in some situations certain non-clinical testing may still be warranted for example when the non-clinical data package does not fulfil the recommendations outlined on the draft Note for Guidance on the Non-Clinical Documentation of Medicinal Products for Mixed Marketing Applications (CPMP/SWP/799/95)."

This statement seems to imply that there may be class effects for pharmacokinetic interactions. This would appear to be a novel concept that is not supported by specific guidances or guidelines on drug–drug interactions. Additionally, the guideline does not define what is meant by "class." This terminology could be used to describe compounds with the same mechanism of action (pharmacological class), similar chemical structure (chemical class), or even BCS classification. It should be noted that the document "Draft Note for Guidance on the Non-Clinical Documentation of Medicinal Products for Mixed Marketing Applications (CPMP/SWP/799/95)" is no longer available. However, the finalized guidance does list (in the Reference section) the document "Guidance on the Non-Clinical Documentation of Medicinal Products for Mixed Marketing Applications (CPMP/SWP/799/95)" which is available on the EMEA web site (EMEA, 2005g).

The guideline also addresses fixed combination of approved compounds not approved as free combination therapy by stating that:

> For combination of approved compounds not approved as combination therapy, even though the safety and efficacy of the individual substances have been considered adequately documented for approval, some aspects regarding expected as well as potential unexpected interactions may still need to be addressed (CPMP/EWP/240/95). If pharmacodynamic interactions are the rationale for the combination, such effects should be documented. Relevant non-clinical studies may be part of the proof of principle. Pharmacodynamic data with the combination may also be justified to allow assessment of unexpected/undesirable interactions. Non-clinical (*in vitro* or *in vivo*) or clinical pharmacokinetic interaction studies with the combination should be considered, when appropriate. Provided that the pharmacokinetics of the single components are adequately characterised in animals, including the profile for enzyme induction and inhibition and drug–drug interactions, additional non-clinical documentation on pharmacokinetic interactions is generally not needed (CPMP/EWP/240/95).
>
> If a pharmacokinetic interaction constitutes the rationale for the combination, such interaction should be documented and appropriate non-clinical data may be needed.
>
> For the safety evaluation of the fixed combination, safety pharmacology and toxicity studies should be considered. The need for (combination) studies will depend on the type of the anticipated interactions between the components and on the range of concentrations and exposures covered in the available studies with the single components. If the systemic exposure expected from the use of the fixed combination is not covered adequately by the existing data, additional studies may be warranted.
>
> Additional testing may also be justified if the compounds to be combined target the same organ system(s), or belong to a class of compounds associated with a specific type of toxicity. The toxicological profile of the combination may be studied in non-clinical bridging studies, to support the safe human use and identify potential interactions. In these bridging

> studies, specific endpoints addressing safety concerns based on the characteristics (pharmacology, toxicology) of the individual components may be included in addition to the conventional toxicological evaluation. If sufficient systemic exposure can be obtained, identified relevant safety pharmacology endpoints to be studied may also be included into the (bridging) toxicity studies. In addition, special and/or mechanistic studies may be warranted, for example studies addressing immunotoxicity or dependence, depending on the properties of the individual components in the combination.

The statement "Provided that the pharmacokinetics of the single components are adequately characterised in animals, including the profile for enzyme induction and inhibition and drug-drug interactions, additional non-clinical documentation on pharmacokinetic interactions is generally not needed (CPMP/EWP/240/95)" appears to be confusing for several different reasons. Firstly, it implies that enzyme induction and inhibition and drug interaction studies should be performed in animals and that such studies are predictive of the human situation. Secondly, it cites the document CPMP/EWP/240/95 as the source for this statement. However, this document (Draft Guideline on Fixed Combination Medicinal Products) does not seem to address these topics (EMEA, 2008b).

This draft guideline does address the topic of PD and PK interactions, stating that "The possibility of interactions between the substances should always be considered. The applicant should submit data either to establish that such interactions do not occur or that they are clearly recognized and defined." The guideline further addresses PK studies, stating:

> In general, the applicant must demonstrate that the various substances do not affect each others respective pharmacokinetic patterns. In some cases, however, a pharmacokinetic interaction (i.e. combination with a metabolism inhibitor) constitutes the rationale of the fixed combination.
>
> These interactions should be studied in healthy volunteers but also in patients if the disease modifies the pharmacokinetics of one substance and in high risk subgroups (elderly, patients with renal failure or hepatic impairment).

### *4.7.8 Special Populations*

*4.7.8.1 Anticancer Drugs* As mentioned previously, the FDA does not have a guidance document on the preclinical development of anticancer drugs, although the EMEA does have one on this topic. Instead, the FDA utilizes an article published by DeGeorge *et al.* (1997) as well as a book chapter written by Farrell *et al.* (2003). The use of combinations of drugs for the treatment of cancer is extremely common. The DeGeorge article notes that "When synergistic effects may be anticipated such as when one agent interferes with the metabolism or elimination of the other agent or both cytotoxic agents target the same metabolic pathway or cellular function, pre-clinical testing of the combination is desirable."

Changes in the route of administration or in the formulation of anticancer drugs are often pursued with a goal of improving drug utility (DeGeorge *et al.*, 1997). If a clinical trial is proposed by the oral route for a drug that has already been investigated by intravenous administration, then additional preclinical studies should address whether there is enhanced liver toxicity, direct gastrointestinal toxicity, or altered metabolism (due to microflora in the gastrointestinal tract, the intestinal wall, or a

first-pass effect through the liver). This is one of the few documents or guidances that acknowledges the potential role of the gut microflora in the metabolism of drugs.

The book chapter written by Farrell *et al.* (2003) provides the clearest indication of what preclinical information is recommended for inclusion in the initial IND filing for an oncologic drug. These studies are single-dose toxicity, repeat-dose toxicity and genotoxicity, although the latter may not be required at the IND filing for studies in patients with advanced cancers. What is enlightening is not so much what is included in the list but what is not included. For example, it would appear that the FDA is recommending that pharmacology or ADME data are not required for the initial IND filing for an oncologic drug. However, many such IND submissions contain a great deal of pharmacology and ADME information and these IND submissions look very similar to an IND submission for a non-life-threatening disease. This chapter further states that "Preclinical combination testing may not be necessary provided that each agent's toxicities have been fully characterized, and data do not indicate that the combination use would be unsafe. Concern may increase if one agent interferes with the metabolism or elimination of another, or if both cytotoxic agents target the same metabolic or cellular pathway, or cellular function."

An additional useful source of information is the material (presentations and transcripts) from the FDA Oncologic Drugs Advisory Committee meeting that was held on March 13, 2006 (US FDA CDER, 2006). This meeting discussed preclinical requirements and phase I trial design issues for the development of oncologic products. At this meeting, John Leighton, a Pharm/Tox reviewer from the FDA stated that "Pharmacokinetic studies provide information on absorption, distribution, metabolism and excretion. These studies are strongly encouraged, particularly for drugs with extended expected duration of exposure, for example drugs administered by depot formulation." It is interesting that the word "encouraged" was used rather than "should" or "expected," indicating the somewhat secondary nature of pharmacokinetics and metabolism in the development of novel oncologic agents.

It should be noted that the EU issued "Note for Guidance on the Pre-Clinical Evaluation of Anticancer Medicinal Products" in 1999 (EMEA, 1998). The purpose of this guideline is to "define the preclinical data which are considered obtainable from preclinical studies with respect to pharmacodynamic, pharmacokinetic and toxicological properties of new anticancer drugs and which are considered relevant with respect to Phase I (Human Pharmacology), Phase II (Therapeutic Exploratory) and Phase III (Therapeutic Confirmatory) Clinical Trials and Marketing Applications. Furthermore, the guideline serves the purpose of avoiding unnecessary tests, thus enabling the promptest possible introduction of newly developed anticancer medicinal products into clinical trials without compromising safety." This note for guidance does not contain any specific mention of metabolism or metabolites. However, it does note that the mechanism of resistance could be investigated, "In parallel with the characterization of the mechanism(s) of action, the corresponding profile with respect to possible mechanism(s) of resistance (e.g. overexpression of P-glycoprotein/multidrug resistance protein/glutathione, changes in topoisomerase I and II) can be obtained. Observed resistance could be investigated for its circumvention by resistance modulating agents. Investigation of the possible induction of resistance by long-term exposure of cell lines to the new drug and further characterization of mechanism(s) of resistance are encouraged." Pharmacokinetic/toxicokinetics studies are mentioned, although the extent of such studies are limited. "The evaluation of limited kinetic parameters, for example peak plasma levels and AUC, at doses around the MTD in the animal species used for

preclinical studies may facilitate dose escalation during Phase I studies. Further information on ADME in animals should normally be made available prior to Phase II/III studies."

The topic of nonclinical safety evaluation of oncology drugs has also been taken up by the ICH based on a proposal by PhRMA in October 2006. In May 2007, the ICH issued a concept paper entitled "S9: Preclinical Guideline on Oncology Therapeutic Development" (ICH, 2007). The rationale for the need for an ICH guidance on this topic is described thus:

> The approaches to the preclinical development of oncology drug products have been and continue to be independently discussed and developed in Europe, the USA and Japan. The preclinical approaches are not agreed on across product classes, between small molecules of different molecular mechanisms and across product classes such as biologics and drugs. The available disharmonized guidance has resulted in inefficient use of animal resources, and ineffective drug development in a critical area of human health. The European Medicines Evaluation Agency (EMEA) has official preclinical guidance for the development of cytotoxic cancer treatments. While cytotoxic agents constitute a decreasing class of agents in development, guidance is needed to address the newer, non-cytotoxic (signaling pathways) and context dependent (tumor mutation specific) mechanisms for treatment of cancer now in development. Nor are biologic approaches that include both cytotoxic and non-cytotoxic mechanism addressed in these guidances. The FDA is developing guidances for biologic therapeutics and drugs that will broadly address preclinical development of the various mechanisms of cancer therapy. Further, in the context of FDA's Exploratory IND Guidance, some preclinical approaches to early investigative oncology drug development are briefly outlined, although not limited to this therapeutic application. The MHLW is currently developing preclinical guidance that will address various mechanisms of anticancer therapy, but will not include biologics in its scope. Thus, there is substantial concern that, when these guidances are completed, they will not offer a harmonized approach for preclinical development of drugs needed for the treatment of cancer.
>
> This independent proliferation of guidances focused on preclinical issues in oncology drug development highlights the critical need for guidance on this topic and indicates the inadequacy of the current state. While these current independent approaches have their merits, the dissociated efforts will likely result in disparate recommendations that will adversely impact global development of needed cancer therapies. The result will be both delays in the availability of needed agents and waste of valuable resources, including test animals.
>
> Amplifying this problem, is the fact that most ICH S guidances and M3 either explicitly or implicitly exclude cancer therapies from their recommendations, yet there is no alternative offered for a harmonized product development strategy. As a result, M3 or S6 guidance is often inappropriately relied on (or expected) particularly for non-cytotoxic and biologic agents in development of cancer therapeutics.

These are very strong statements and indicate that this topic is one that will likely advance quickly. The issues to be resolved include:

1. Scope of the guidance: Therapeutic classes to be included; preclinical development approach in relation to product class and how to cover the breadth of the product classes in the oncology guidance; classes to be excluded are therapeutic vaccines, cell and gene therapy.
2. Necessity for long-term preclinical testing to support early clinical trials.
3. Duration, species, type of nonclinical safety studies to support the early and late development phases.

4. Harmonization of approaches to the extent justified for biological and small molecules.

It is interesting to note that the ICH will aggressively develop this guideline. However, the aggressive timeline calls for final guidance to be issued in early 2010.

In 2004, the FDA issued a guidance on "IND Exemptions for Studies of Lawfully Marketed Drug or Biological Products for the Treatment of Cancer" (US FDA, 2004b). This guidance identifies the circumstances under which a sponsor may be able to claim an exemption from the requirements for the submission of an IND. It includes a statement that is very similar to that made in the DeGeorge article: "Because of the danger of synergistic toxicity (i.e., enhanced effects from the combination) occurring with a new drug combination, if there are no data from the literature on its safety, the initial study of a new drug combination should ordinarily be performed under an IND. Synergistic toxicity may be anticipated when one agent interferes with the metabolism or elimination of the other agent; when both agents target the same metabolic pathway or cellular function; or when one agent targets signaling pathways that are reasonably expected to modulate sensitivity to the other agent. If it is determined that synergistic toxicity is likely, animal studies should be considered for determining a safe starting dose for the drug combination in humans" (DeGeorge *et al.*, 1997). This statement is important since it emphasizes the understanding of the enzymology of the metabolism of the drug. However, it should be noted that the enzyme systems involved in the metabolism of many of the drugs that were approved before the early to mid-1990s may not be well characterized.

*4.7.8.2 Diabetes* In February 2008, the FDA issued a draft guidance "Diabetes Mellitus: Developing Drugs and Therapeutic Biologics for Treatment and Prevention" (US FDA, 2008i). Appendix A of this draft guidance addresses preclinical considerations for peroxisome proliferator-activated receptor antagonists. Of interest are the statements "The ICH guidance regarding the duration of chronic toxicity studies in rodents and nonrodents has been adopted and for the nonrodent chronic toxicity study, a 9-month duration generally is appropriate for supporting chronic human use. However, since the no observed adverse effect levels for some of the toxicities associated with PPAR agonists can be adequately defined only after chronic administration, a 1-year study in nonrodents is recommended for drugs in the PPAR class." The peroxisome proliferators have long been known to cause liver enlargement and be non-genotoxic carcinogens. Therefore, it is not surprising to see the hepatic effects discussed in this Appendix: "The cause of any liver enlargement observed should be determined (peroxisome proliferation, mitochondrial proliferation/swelling). Liver tissues should be stained to detect the presence of fatty changes. The sections can be stained with Sudan IV or Oil Red-O. Liver enzyme levels and biochemical markers of peroxisome proliferation (Acyl CoA and CYP 4A) should be analyzed in rodents and nonrodents." The last statement should be clarified in the final guidance since the terminology used is not as accurate as it should be.

*4.7.8.3 Pediatrics* The safety testing of drug products for use in a pediatric population has been addressed in a guidance (Nonclinical Safety Evaluation of Pediatric Drug Products) issued by the FDA in 2006 (US FDA, 2006i).

This guidance notes that there may be differences in drug safety profiles between mature and immature systems, and that metabolism may be involved in these differences. The guidance states:

> Some therapeutics have shown different safety profiles in pediatric and adult patients. Inherent differences between mature and immature systems introduce the possibility of drug toxicity, or resistance to toxicity in immature systems that are not observed in mature systems. Several factors contribute to these potential differences. Postnatal growth and development can affect drug disposition and action. Examples include developmental changes in metabolism (including the maturation rate of Phase I and II enzyme activities), body composition (i.e., water and lipid partitions), receptor expression and function, growth rate, and organ functional capacity. These developmental processes are susceptible to modification or disruption by drugs. Although some age-dependent effects can be largely predicted by knowledge of the changes in drug metabolic pathways during development, others cannot.

The guidance then proceeds to provide some examples of drugs that exhibit differences in toxicity between adult and pediatric patients. These include acetaminophen. "Acute acetaminophen toxicity is a classic example of how maturation can affect the toxicity profile of a drug. Young children are far less susceptible to acute acetaminophen toxicity than adults because children possess a higher rate of glutathione turnover and more active sulfation. Thus, they have a greater capacity to metabolize and detoxify an overdose of acetaminophen when compared to adults."

The selection of the species is also addressed by the guidance:

> The species of the juvenile animal tested should be appropriate for evaluating toxicity endpoints important for the intended pediatric population. Traditionally, rats and dogs have been the rodent and nonrodent species of choice. In some circumstances, however, other species may be more appropriate. For example, when drug metabolism in a particular species differs significantly from humans, an alternative species (e.g., minipigs, pigs, monkeys) may be more appropriate for testing.

The implication of this statement is that the sponsor may need to perform a wider "species screen" for the metabolically relevant species. The guidance does not address whether both the rodent and non-rodent species need to be metabolically relevant or whether only one species needs to be metabolically relevant. The availability of suitable materials (e.g. hepatic microsomes obtained from human infants) may make an *in vitro* comparison difficult. Interestingly, the guidance intimates that the non-rodent species will be the species in which the metabolism will most likely be different from that in humans. The guidance does not address whether such a species screen is performed *in vivo* or *in vitro*. Additionally, the search for alternative species for testing of the potential for toxicity in juveniles may have strategic implications if it is found that the minipig would be a more relevant species for toxicology studies than the dog. If this is the case, an argument could be made that the general non-rodent toxicology studies should be performed in minipigs. The guidance further states that "Assessment of developmental differences in parent drug disposition and profiles of significant metabolites in juvenile animals should be made according to established guidelines (see the ICH guideline for industry S3A Toxicokinetics: Assessment of Systemic Exposure in Toxicity Studies)." Significant metabolite is not defined in this guidance. The guidance makes the point that the dose frequency should "be relevant to the intended clinical

use of the drug. In some cases, however, the use of dosing frequencies similar to those anticipated for clinical administration are not feasible because of technical considerations for the animal models used. Changes in frequency can be made when variables such as metabolic and kinetic differences are considered." This implies that there the sponsor may need to conduct several dose-ranging studies that involve an evaluation of the exposure top both parent and relevant metabolites. Finally, the guidance provides, in Table 7, an overview of the developmental modulation of phase I/II metabolism. This table lists a range of enzyme activities and their expression in human, rat, and rabbit. No information is provided regarding other species such as dogs, minipigs, pigs, or monkeys.

In 2005, the CHMP issued a draft guidance entitled "Guideline on the Need for Non-Clinical Testing in Juvenile Animals on Human Pharmaceuticals for Paediatric Indications" (EMEA, 2005e). This guidance was finalized in January 2008 and came into effect in August 2008 (EMEA, 2008d). The guideline notes that the development of certain functions related to the pharmacokinetic handling of an active substance may take up to several years stating that "The development of the major systems is age dependent, for example:

- Nervous system: Development up to adulthood
- Reproductive system: Development up to adulthood
- Pulmonary system: Development up to two years old
- Immune system: Development up to 12 years old
- Renal system: Development up to one year of age
- Skeletal system: Development up to adulthood
- Organs and/or systems involved in absorption and metabolism of drugs. Development of biotransformation enzymes up to adolescence."

The guideline also addresses pre- and postnatal reproduction studies as follows: "Before performing a juvenile animal toxicity study, it should be considered whether a developmental toxicity issue could be addressed in a modified pre- and postnatal development study in rats. Key factors that need to be examined include, but are not restricted to, the amount of the active substance and/or relevant metabolites excreted via the milk and the resulting plasma exposure of the pups, which organs under development that will be exposed during the pre-weaning period, physical development and histopathological investigations."

The importance of pharmacokinetic data is described:

> Pharmacokinetic data (including in vitro data, if relevant) contributes to the evaluation of the relevance of the animal model for human risk assessment. It is recognised that the collection of blood samples to obtain a full kinetic profile of a test compound under study in juvenile animals might sometimes be difficult. However, sampling at a few time points, using pooled samples if necessary, should be performed to obtain an estimate of basic kinetic characteristics, for example Cmax and AUC. Inclusion of satellite groups of dosed animals, for example excess offspring from a pre-postnatal study, for blood sampling may be considered. The use of methodology allowing population pharmacokinetic/toxicokinetic determinations may also be considered. Toxicokinetics should also be considered to confirm appropriate exposure levels in different treatment groups.

Finally, the guideline states that "under special circumstances, data on absorption, distribution, metabolism and/or excretion in juvenile animals may be valuable to further investigate and for example understand a specific safety concern."

### 4.7.9 Other Topics

*4.7.9.1 Dependence Potential* The topic of dependence potential was addressed by the EMEA in its "Guideline on the Non-Clinical Investigation of the Dependence Potential of Medicinal Products" that was issued in March 2006 (EMEA, 2006g). This guidance states that the impetus for conducting such studies will include early indicators that may come from pharmacokinetic studies (brain penetration, metabolites entering or formed in the brain). However, it should be noted that the brain penetration of parent drug or metabolites, or the metabolism of a drug by the brain is rarely, if ever, determined.

Consistent with the FDA's guidance on Safety Testing of Drug Metabolites (US FDA, 2008j), the guideline states: "Separate studies evaluating the dependence potential of metabolites would be needed when there is reasonable evidence that such metabolites are formed in humans, are suspected to penetrate the human CNS in significant concentrations and are not adequately evaluated in non-clinical studies with the parent."

The guideline addresses the timing of these studies in the following statements:

> Receptor binding studies and safety pharmacology studies regarding the CNS as outlined under ICH S7A should be completed before phase 1 clinical trials. … In general, metabolites do not need to be evaluated for dependence potential prior to human trials. Separate receptor binding studies on metabolites should be initiated as soon as human pharmacokinetic data are available and such studies are considered appropriate. When studies on metabolite receptor affinity and other studies described under the first tier are deemed relevant, data from these studies should become available before large populations are exposed to the investigational medicinal product, generally before Phase III. All relevant data should be available at the time of a Marketing Authorisation Application.

The same topic was addressed by Health Canada in a draft guidance "Clinical Assessment of Abuse Liability for Drugs with Central Nervous System Activity" that was also issued in March 2006 (Health Canada, 2007a).

*4.7.9.2 Photosafety Testing* The 2003 FDA guidance on "photosafety testing" is one of several guidances that points out that toxicity may be due to a metabolite rather than (or as well as) due to the parent drug itself (US FDA, 2003a). In addition, this guidance makes the statement that "Tissue distribution studies of systemically administered drug products, usually included in IND submissions, can be used to assess the extent of partitioning into the skin or eyes. In the absence of partitioning into light-exposed compartments, photoirritation testing is unlikely to be informative and need not be conducted." However, there is no mandate for tissue distribution studies to be included in the IND submission. It would be interesting to conduct a survey of pharmaceutical and biotechnology companies to determine the frequency of inclusion of such studies in their IND submissions and the reason for the conduct of these studies.

*4.7.9.3 Excipients* Excipients were the subject of a guidance entitled "Nonclinical Studies for the Safety Evaluation of Pharmaceutical Excipients" that was issued by the FDA in 2005 (US FDA, 2005d). The intent of this guidance is to outline the process for the safety evaluation of excipients that are not fully qualified by existing safety data with respect to the currently proposed level of exposure, duration of exposure, or route of administration. The basis for this guidance is the observation that not all excipients are inert substances and that some have been shown to be potential toxicants. The Federal Food, Drug, and Cosmetic Act of 1938 (the Act) was enacted after the tragedy of the elixir of sulfanilamide in 1937 in which an untested excipient was responsible for the death of many children who consumed the pharmaceutical. Essentially, the guidance states that you would characterize the ADME properties of the excipient in a similar manner to a drug. The topic of the toxicity of excipients has been extensively reviewed by Weiner and Kotkoskte (2007).

*4.7.9.4 Immunotoxicology* The assessment of the distribution of drug and/or metabolites to a particular tissue is mentioned in the ICHS8 guidance "Immunotoxicity Studies for Human Pharmaceuticals" that was issued in April 2006 (ICH, 2005d). The guidance states that "If the compound and/or its metabolites are retained at high concentrations in cells of the immune system, additional immunotoxicity testing should be considered." However, the assessment of the retention of compound-related material in the cells of the immune system is rarely, if ever, performed unless there is a toxicity signal that such information would be useful in interpreting.

*4.7.9.5 Alternate Routes of Administration* In March 2008, the FDA released a draft guidance entitled "Nonclinical Safety Evaluation of Reformulated Drug Products Intended for Administration by an Alternate Route" (US FDA, 2008h). Metabolism is addressed in a couple of places in the draft guidance. The guidance states that:

> All routes of administration can result in systemic exposure. Therefore, the adequacy of the available systemic toxicity information should be evaluated based on the systemic exposure obtained after administration of a proposed new formulation or of a previous formulation by a new route. Additional toxicity studies might be recommended if the available toxicity information is not sufficient to support the exposure measured with the new formulation or if a significantly different pattern of exposure results from the new formulation. An adequate evaluation of the pharmacokinetics and absorption, distribution, metabolism, and elimination (PK/ADME) of the drug substance is recommended for new formulations. These data and any available human data can be helpful in determining what additional nonclinical toxicity data, if any, are recommended. When comparing the PK/ADME of a new formulation with a previously approved formulation, it is important to examine the shape of the concentration/time curve and not just the total area under curve. For example, alterations in absorption or the dosing frequency can produce significantly different concentration/time profiles that might lead to different toxicological effects. Changes in the vehicle composition or form also can alter the PK of active ingredients. In some cases, PK/ADME for the new formulation might not be available. In these cases, an assumption of 100% bioavailability from the proposed clinical dose might be used to judge the adequacy of available systemic toxicity information."

The guidance describes many different routes of administration. It is only for the ocular route that pharmacokinetics is expressly mentioned "If the active ingredient has

not been used by the ocular route, then toxicity studies in two species with complete eye and systemic evaluation for the appropriate duration should be carried out with the new formulation. In certain cases, studies in one most appropriate species may be adequate. Optimal design of these studies would include the evaluation of ocular and systemic PK."

***4.7.10 Support to First-in-Human Studies*** In March 2006, the "TeGenero incident" occurred in which very serious adverse reactions occurred in a first-in-human clinical trial of TGN412 which is a monoclonal antibody that was being developed as a medicine to treat leukemia and autoimmune diseases such as rheumatoid arthritis (Wikipedia, 2008a). In the clinical trial, six healthy male volunteers experienced severe systemic adverse reactions soon after receiving TGN1412 intravenously. All six volunteers developed a cytokine release syndrome with multi-organ failure and required intensive treatment and supportive measures that were provided by the Intensive Therapy Unit at Northwick Park Hospital (the site of the clinical trial). As a result of this incident, an Expert Scientific Group (ESG), led by Prof. Gordon W. Duff, was set up following this incident. The group received input from a variety of groups and individuals and issued their final report in December 2006 (Expert Group on Phase One Clinical Trials, 2006). The TeGenero incident also led to the Association of the British Pharmaceutical Industry updating its "Guidelines for Phase I Clinical Trials" in 2007 (Association of the British Pharmaceutical Industry, 2007).

As a result of the TeGenero incident, the EMEA issued a controversial guideline in September 2007, entitled "Guideline on Strategies to Identify and Mitigate Risks for First-In Human Clinical Trials with Investigational Medicinal Products" (EMEA, 2007d). Originally, the intent of the guideline was to address potential high-risk biological products. However, the final guidance covers both biologicals and small molecules.

Metabolism is only mentioned in the guideline in terms of the selection of the relevant species for toxicology testing. However, this guidance will be discussed in some depth due to the potential profound impact of the guidance on the nonclinical groups and the potential need for information that may not routinely be generated.

This guideline acknowledges that:

> For many new investigational medicinal products, the non-clinical safety pharmacology and toxicology program provides sufficient safety data for estimating risk prior to first administration in humans. However, for some novel medicinal products this non-clinical safety program might not be sufficiently predictive of serious adverse reactions in man and the non-clinical testing and the design of the first-in-human study requires special consideration. When planning a first-in-human clinical trial, sponsors and investigators should identify the factors of risk and apply risk mitigation strategies accordingly as laid down in this guideline. In addition to the principles expressed in this guideline, some special populations such as paediatrics may deserve specific considerations.

The focus of risk identification and risk mitigation before performing an FIH study is a somewhat new approach. Previously, the approach had been primarily focused on hazard identification with a lesser emphasis on risk identification and risk mitigation.

The risk factors that should be taken into account are described in the following manner: "Predicting the potential severe adverse reactions for the first-in-human use of an investigational medicinal product, involves the identification of the factors of risk. Concerns may be derived from particular knowledge or lack thereof regarding (1) the

mode of action, (2) the nature of the target, and/or (3) the relevance of animal models."

This statement makes the point regarding lack of knowledge and starts to lay the groundwork for the conclusion that new drugs should be regarded as having high risk unless data exist to the contrary. The guidance further states, "The Sponsor should discuss the following criteria for all first-in-human trials in their clinical trial authorisation application. These criteria should be taken into account on a case-by-case basis."

*Mode of action*

While a novel mechanism of action might not necessarily add to the risk per se, consideration should be given to the novelty and extent of knowledge of the supposed mode of action. This includes the nature and intensity (extent, amplification, duration, reversibility) of the effect of the medicinal product on the specific target and non-targets and subsequent mechanisms, if applicable. The type and steepness of the dose response as measured in experimental systems, which may be linear within the dose range of interest, or non-linear (e.g. plateau with a maximum effect, over-proportional increase, U-shaped, bell-shaped), is of importance.

For example, the following modes of action might require special attention:

— A mode of action that involves a target which is connected to multiple signaling pathways (target with pleiotropic effects), e.g. leading to various physiological effects, or targets that are ubiquitously expressed, as often seen in the immune system.

— A biological cascade or cytokine release including those leading to an amplification of an effect that might not be sufficiently controlled by a physiologic feedback mechanism (e.g., in the immune system or blood coagulation system). CD3 or CD28 (super-) agonists might serve as an example.

When analysing risk factors associated with the mode of action, aspects to be considered may include:

— Previous exposure of human to compounds that have related modes of action.

— Evidence from animal models (including transgenic, knock-in, or knock-out animals) for the potential risk of serious, pharmacologically mediated toxicity

— Novelty of the molecular structure of the active substance(s), for example, a new type of engineered structural format, such as those with enhanced receptor interaction as compared to the parent compound.

*Nature of the target*

The target in human should be discussed in detail. Beyond the mode of action, the nature of the target itself might impact on the risk inherent to a first administration to humans, and sponsors should discuss the following aspects, based on the available data:

— the extent of the available knowledge on the structure, tissue distribution (including expression in/on cells of the human immune system), cell specificity, disease specificity, regulation, level of expression, and biological function of the human target including "downstream" effects, and how it might vary between individuals in different populations of healthy subjects and patients.

— If possible, a description of polymorphisms of the target in relevant animal species and humans, and the impact of polymorphisms on the pharmacological effects of the medicinal product.

*Relevance of animal species and models*

The sponsor should compare the available animal species to humans taking into account the target, its structural homology, distribution, signal transduction pathways and the nature of pharmacological effects.

The guidance then refers to the section on "Demonstration of relevance of the animal model." This section of the guideline states that "Qualitative and quantitative differences may exist in biological responses in animals compared to humans. For example, there might be differences in affinity for molecular targets, tissue distribution of the molecular target, cellular consequences of target binding, cellular regulatory mechanisms, metabolic pathways, or compensatory responses to an initial physiological perturbation.

Where there is evidence of species-specificity of action from *in vitro* studies with human cells compared with cells from a test species, the value of the *in vivo* response of the test species may be significantly reduced in terms of predicting the *in vivo* human response. It should be noted that a similar response in human and animal cells *in vitro* is not necessarily a guarantee that the *in vivo* response will be similar.

In practice this means that animal studies with highly species-specific medicinal products may:

- not reproduce the intended pharmacological effect in humans;
- give rise to misinterpretation of pharmacokinetic and pharmacodynamic results;
- not identify relevant toxic effects.

A weight-of-evidence approach should involve integration of information from *in vivo*, *ex vivo* and *in vitro* studies into the decision-making process.

High species-specificity of a medicinal product makes the non-clinical evaluation of the risk to humans much more difficult, but does not imply that there is always an increased risk in first-in-human trials.

The demonstration of relevance of the animal model(s) may include comparison with humans of:

- Target expression, distribution and primary structure. However, a high degree of homology does not necessarily imply comparable effects;
- Pharmacodynamics
  - Binding and occupancy, functional consequences, including cell signalling if relevant.
  - Data on the functionality of additional functional domains in animals, if applicable, e.g. Fc receptor system for monoclonal antibodies.
- Metabolism and other pharmacokinetic aspects
- Cross-reactivity studies using human and animal tissues (e.g. monoclonal antibodies).

The search for a relevant animal model should be documented and justified in detail. Where no relevant species exists, the use of homologous proteins or the use of relevant transgenic animals expressing the human target may be the only choice. The data gained is more informative when the interaction of the product with the target receptor has similar physiological consequences to those expected in humans. The use of *in vitro* human cell systems could provide relevant additional information. The relevance and limitations of

> all models used should be carefully considered and discussed fully in the supporting documentation.

The good news in these lengthy statements is that the drugs will be evaluated on a "case-by-case" basis and that the "weight of evidence" approach will be used. The bad news is that it is clear that a far greater understanding of the biology of the target and the pharmacology of the drug. A far more detailed understanding of the dose or concentration–effect relationship will be required as will be an understanding of the biology of the pharmacological target in humans.

In summary on the topic of species relevance, the guideline states that "Where available animal species/models or surrogates are perceived to be of questionable relevance for thorough investigation of the pharmacological and toxicological effects of the medicinal product, this should be considered as adding to the risk."

The guideline provides further detail on two topics (pharmacodynamics and pharmacokinetics) that are of interest to the drug metabolism scientists. Regarding pharmacodynamics, the guideline states that:

> Pharmacodynamic studies should address the mode of action, and provide knowledge on the biology of the target. These data will help to characterise the pharmacological effects and to identify the most relevant animal models. The primary and secondary pharmacodynamics, should be conducted in *in vitro* animal and human systems and *in vivo* in the animal models. These studies should include target interactions preferably linked to functional response, e.g. receptor binding and occupancy, duration of effect and dose-response. A dose/concentration-response curve of the pharmacological effect(s) should be established with sufficient titration steps in order to increase the likelihood to detect significant pharmacological effects with low doses and to identify active substances with U-shaped or bell-shaped dose response curves. Such significant or even reverse effects have been reported with biological compounds. Since a low dose is to be administered to humans in the first-in-human trial, this is of high importance. Although GLP compliance is not mandatory for pharmacodynamic and pharmacokinetic studies, they should be of high quality and consistent with the principles of GLP.

The guideline also addresses pharmacokinetics. "Standard pharmacokinetic and toxicokinetic data should be available in all species used for safety studies before going into human (ICH S3, S6, M3). Exposures at pharmacodynamic doses in the relevant animal models should be determined especially when pharmacodynamic effects are suspected to contribute to potential safety concerns." If this statement is followed, there will be the need for an increased level of evaluation of the pharmacokinetics of the drug in the efficacy species. Presently, the exposure is usually determined only at one dose level (e.g. the $ED_{50}$ dose) rather than at the multiple dose levels that are implied by the guideline. Additionally, since a theme of the guideline is the understanding of the concentration–effect relationship, there is the implication that the sponsor would have determined whether metabolites contribute meaningfully to the efficacy in the animal model. It also implies that if metabolites do contribute meaningfully to the efficacy, the sponsor will have described the exposure to the parent drug and relevant metabolites in the animal efficacy model.

The estimation of the first dose in humans is addressed by this guideline:

> The estimation of the first dose in human is an important element to safeguard the safety of subjects participating in first-in-human studies. All available information has to be taken

in consideration for the dose selection and this has to be made on a case-by-case basis. Different methods can be used.

In general, the No Observed Adverse Effect Level (NOAEL) determined in non-clinical safety studies performed in the most sensitive and relevant animal species, adjusted with allometric factors (see references "Other guidelines") or on the basis of pharmacokinetics gives the most important information. The relevant dose is then reduced/adjusted by appropriate safety factors according to the particular aspects of the molecule and the design of the clinical trials.

For investigational medicinal products for which factors influencing risk according to section 4.1 have been identified, an additional approach to dose calculation should be taken. Information about pharmacodynamics can give further guidance for dose selection. The 'Minimal Anticipated Biological Effect Level' (MABEL) approach is recommended. The MABEL is the anticipated dose level leading to a minimal biological effect level in humans. When using this approach, potential differences of sensitivity for the mode of action of the investigational medicinal product between humans and animals, need to be taken into consideration e.g. derived from *in-vitro* studies. A safety factor may be applied for the calculation of the first dose in human from MABEL as discussed below.

The calculation of MABEL should utilise all *in vitro* and *in vivo* information available from pharmacokinetic/pharmacodynamic (PK/PD) data such as:

i) target binding and receptor occupancy studies *in vitro* in target cells from human and the relevant animal species;
ii) concentration-response curves *in vitro* in target cells from human and the relevant animal species and dose/exposure-response *in vivo* in the relevant animal species.
iii) exposures at pharmacological doses in the relevant animal species.

Wherever possible, the above data should be integrated in a PK/PD modeling approach for the determination of the MABEL.

In order to further limit the potential for adverse reactions in humans, a safety factor may be applied in the calculation of the first dose in human from the MABEL. This should take into account criteria of risks such as the novelty of the active substance, its biological potency and its mode of action, the degree of species specificity, and the shape of the dose-response curve and the degree of uncertainty in the calculation of the MABEL. The safety factors used should be justified. When the methods of calculation (e.g. NOAEL, MABEL) give different estimations of the first dose in man, the lowest value should be used, unless justified. Other approaches may also be considered in specific situations, e.g. for studies with conventional cytotoxic IMPs in oncology patients.

It is heartening to note that it has been recognized that the approach outlined in this guideline will not be applied across the board and should not inhibit the development of new anticancer drugs for which the FIH study will be conducted in a patient population, that is, in a patient population that clearly has a different risk–benefit than do normal healthy volunteers.

It can be seen from the statements in this guideline that there has been a shift in basic philosophy regarding the safety of new drugs. Previously, the mind-set was that medicinal products *are not* high risk unless we have evidence to the contrary. Now the mind-set seems to be that medicinal products *are* high risk unless we have evidence to the contrary.

Clearly, there are many potential implications of this guideline and many open questions. These implications and open questions occur for many different disciplines include, but are not restricted to the following:

*Pharmacology*

- How many animal models of efficacy should be evaluated?
  - More models in order to better define MABEL?
  - Fewer models due to increased workload?
  - Fewer models but repeated more than once?
- What dosing regimens should be used?
- May need to increase the number of dose levels in animal efficacy studies
- May need increased PK assessment in animal efficacy studies
- May need robust PD markers for both NCEs and biologics
- May need to enhance understanding of variability in, and polymorphisms of, the target
- Need to better understand the pharmacology of the target in humans, including potential sources of variability of the target
- May need to conduct animal efficacy studies "in the spirit of GLP"
- How relevant are our animal models?
  - for example, utility of CIA-induced arthritis model in rat.

*Toxicology*

- Selection of species for toxicology testing
  - How do we address differences in target based on supplier, age, etc.?
  - Need to determine tissue distribution of target for a NCE?
- Need robust PD markers in toxicology studies for both NCEs and biologics
- May need to include limited toxicity assessments in animal efficacy studies
- Design of Tox studies
  - Using repeat-dose Tox studies to support single-dose clinical trials is already conservative
    - Should we use single-dose Tox studies to support single-dose clinical studies?
      - This approach may work in the United States but not in the EU

*PK/ADME*

- Greater need for PK/PD modelers in nonclinical
- Need robust PD markers for both NCEs and biologics
- Will need more studies to define PK/PD than are presently conducted
- Need to ensure relevance of *in vitro* metabolism data in selection of Tox species

*Clinical*

- Selection of starting dose
  - What does "minimal" mean?
    - 10% effect?
    - Need to have robust assays that can distinguish small changes in response
  - All FIH clinical trials may be (in essence) a microdose study

- Each cohort will take longer to perform due to sequential dosing
- Any adverse event that was not observed in animal toxicology studies will likely halt dose escalation
- Slow conduct of clinical trial due to lag time between cohorts due to stopping rules and increased reliance on PK/PD
- Need for real-time (bedside) PD assays?
- Need to develop "rescue" procedures for patients that develop adverse events
- Slow IV infusions will be used instead of bolus administration

*Regulatory*

- Presently, the EU regulatory agencies do not require study reports for review but they may now start to ask for them
- Gaining scientific advice from the EMEA prior to FIH used to be rare but may well become common (equivalent of Pre-IND meeting)
  - Can the EMEA handle an increased level of meetings?
- Will the FDA default to the EU guideline?
- FIH clinical trials may become rare in the EU
- Since some high-risk products are only high risk upon repeat-dose administration, will multiple-dose clinical trials also be placed under similar scrutiny?
- Will need better competitive intelligence regarding other compounds with similar MOAs

*CMC*

- May need to provide a greater range of dose strengths for the FIH study
- May need to provide solution doses to aid flexibility in dose escalation

*Strategic*

- The pharmaceutical industry may choose to no longer develop drugs with:
  - Steep dose/response curves
  - Novel targets
  - Big Pharma may leave this risk to the small biotech companies.
- Fewer FIH studies may be conducted in the EU
- Need for greater sharing of information between companies working on the same MOA?
- Need for greater interaction between pharmacology, toxicology, PK, and clinical groups within a company.

It should be noted that there are some strong similarities between this guideline and the FDA guidance on Exploratory INDs that will be discussed in Section 5.3.

In July 2005, the FDA issued a guidance, "Estimating the Maximum Safe Starting Dose in Initial Clinical Trials for Therapeutics in Adult Healthy Volunteers" (US FDA, 2005e). This guidance outlines a process (algorithm) and vocabulary for deriving the

maximum recommended starting dose (MRSD) for FIH clinical trials of new molecular entities in adult healthy volunteers, and recommends a standardized process by which the MRSD can be selected. The purpose of this process is to ensure the safety of the human volunteers. The guidance states that:

> Although the process outlined in this guidance uses administered doses, observed toxicities, and an algorithmic approach to calculate the MRSD, an alternative approach could be proposed that places primary emphasis on animal pharmacokinetics and modeling rather than dose. In a limited number of cases, animal pharmacokinetic data can be useful in determining initial clinical doses. However, in the majority of investigational new drug applications (INDs), animal data are not available in sufficient detail to construct a scientifically valid, pharmacokinetic model whose aim is to accurately project an MRSD.

The following statement is included as a footnote to the previous paragraph:

> If the parent drug is measured in the plasma at multiple times and is within the range of toxic exposures for two or more animal species, it may be possible to develop a pharmacokinetic model predicting human doses and concentrations and to draw inferences about safe human plasma levels in the absence of prior human data. Although quantitative modeling for this purpose may be straightforward, the following points suggest this approach can present a number of difficulties when estimating a safe starting dose. Generally, at the time of IND initiation, there are a number of unknowns regarding animal toxicity and comparability of human and animal pharmacokinetics and metabolism: (1) human bioavailability and metabolism may differ significantly from that of animals; (2) mechanisms of toxicity may not be known (e.g., toxic accumulation in a peripheral compartment); and/or (3) toxicity may be due to an unidentified metabolite, not the parent drug. Therefore, relying on pharmacokinetic models (based on the parent drug in plasma) to gauge starting doses would require multiple untested assumptions. Modeling can be used with greatest validity to estimate human starting doses in special cases where few underlying assumptions would be necessary. Such cases are exemplified by large molecular weight proteins (e.g., humanized monoclonal antibodies) that are intravenously administered, are removed from circulation by endocytosis rather than metabolism, have immediate and detectable effects on blood cells, and have a volume of distribution limited to the plasma volume. In these cases, allometric, pharmacokinetic, and pharmacodynamic models have been useful in identifying the human mg/kg dose that would be predicted to correlate with safe drug plasma levels in nonhuman primates. Even in these cases, uncertainties (such as differences between human and animal receptor sensitivity or density) have been shown to affect human pharmacologic or toxicologic outcomes, and the use of safety factors as described in this guidance is still warranted.

The guidance goes on to state:

> Toxicity should be avoided at the initial clinical dose. However, doses should be chosen that allow reasonably rapid attainment of the phase 1 trial objectives (e.g., assessment of the therapeutic's tolerability, pharmacodynamic or pharmacokinetic profile). All of the relevant preclinical data, including information on the pharmacologically active dose, the full toxicologic profile of the compound, and the pharmacokinetics (absorption, distribution, metabolism, and excretion) of the therapeutic, should be considered when determining the MRSD. Starting with doses lower than the MRSD is always an option and can be particularly appropriate to meet some clinical trial objectives.

These statements are confusing since they appear to encourage the sponsor to utilize nonclinical metabolism and pharmacokinetic data to establish the starting dose for the

clinical trial, yet also make some strong caveats as to why the use of such data may not be appropriate.

The guidance further states that:

> Initial IND submissions for first-in-human studies by definition lack *in vivo* human data or formal allometric comparison of pharmacokinetics. Measurements of systemic levels or exposure (i.e., AUC or $C_{max}$) cannot be employed for setting a safe starting dose in humans, and it is critical to rely on dose and observed toxic response data from adequate and well-conducted toxicology studies. However, there are cases where nonclinical data on bioavailability, metabolite profile, and plasma drug levels associated with toxicity may influence the choice of the NOAEL. One such case is when saturation of drug absorption occurs at a dose that produces no toxicity. In this instance, the lowest saturating dose, not the highest (nontoxic) dose, should be used for calculating the HED.

The selection of the most appropriate species for calculating the MRSD is addressed thus:

> After the HEDs have been determined from the NOAELs from all toxicology studies relevant to the proposed human trial, the next step is to pick one HED for subsequent derivation of the MRSD. This HED should be chosen from the most appropriate species. In the absence of data on species relevance, a default position is that the most appropriate species for deriving the MRSD for a trial in adult healthy volunteers is the most sensitive species (i.e., the species in which the lowest HED can be identified). Factors that could influence the choice of the most appropriate species rather than the default to the most sensitive species include: (1) differences in the absorption, distribution, metabolism, and excretion (ADME) of the therapeutic between the species, and (2) class experience that may indicate a particular animal model is more predictive of human toxicity. Selection of the most appropriate species for certain biological products (e.g., human proteins) involves consideration of various factors unique to these products. Factors such as whether an animal species expresses relevant receptors or epitopes may affect species selection (refer to ICH guidance for industry S6 Preclinical Safety Evaluation of Biotechnology-Derived Pharmaceuticals for more details). When determining the MRSD for the first dose of a new therapeutic in humans, absorption, distribution, and elimination parameters will not be known for humans. Comparative metabolism data, however, might be available based on in vitro studies. These data are particularly relevant when there are marked differences in both the in vivo metabolite profiles and HEDs in animals."

These statements could be interpreted as implying that data on the ADME of the drug in animals are known at the time that the MRSD is calculated.

The FDA allows for the possibility of decreasing the safety factor used for the selection of the safe starting dose. The guidance states:

> Safety factors of less than 10 may be appropriate under some conditions. The toxicologic testing in these cases should be of the highest caliber in both conduct and design. Most of the time, candidate therapeutics for this approach would be members of a well-characterized class. Within the class, the therapeutics should be administered by the same route, schedule, and duration of administration; should have a similar metabolic profile and bioavailability; and should have similar toxicity profiles across all the species tested including humans.

However, it would appear unlikely that compounds of the same class of drugs would have a similar metabolic profile and bioavailability, even if one could agree what similar meant in this circumstance.

The subject of allometric scaling was raised in the FDA guidance, "Estimating the Maximum Safe Starting Dose in Initial Clinical Trials for Therapeutics in Adult Healthy Volunteers" (US FDA, 2005e). This topic is one for which there are no regulatory guidances and which can generate vigorous debate. Allometric scaling relates a pharmacokinetic parameter, such as clearance or volume of distribution, to total body weight (or other physiological parameter) using a power equation. The use of allometry makes many assumptions and has a mixed record in its ability to predict the pharmacokinetics of a drug in humans. It should be noted that allometry cannot actually be used to predict the pharmacokinetics of a drug specifically in humans since it is simply a means to relate a PK parameter to a physiological parameter such as body weight. In other words, allometry can be used to determine the PK parameters of a drug in a 70-kg animal, but that animal may be a human, a pig, or a small wildebeest. The nonclinical pharmacokineticist is often asked to predict the human pharmacokinetics of a drug before it goes into the clinic, although how such information is used will vary from company to company.

Some companies place great weight on the predicted human pharmacokinetics and expend great effort in trying to refine the calculations, while others use it to a far lesser extent and are comfortable with a broad estimate of the predicted human pharmacokinetics. It should be noted that pharmacokineticists may feel that their prediction of the pharmacokinetics was successful if the observed and the predicted results were within twofold (Rowland, 2007). It has been reported that even with such a wide acceptance range (i.e. the observed clearance is within 50–200% of the predicted clearance), the success rate is <53% (Nagilla and Ward, 2004, 2005a, b; Ward and Smith, 2004a, b; Mahmood, 2005). However, other groups within the company may expect that the predictions should be more accurate. No matter how much effort was expended to predict the human pharmacokinetics, the predicted pharmacokinetics will be of limited, if any, value as soon as the PK parameters are calculated from the first dose cohort in the FIH clinical trial. Despite these caveats, the predicted human PK parameters can be useful in helping to estimate the approximate efficacious dose (based on matching exposures in the animal efficacy model and the predicted human exposure at a given dose).

The predicted dose can be used to provide guidance to the formulation scientists in order for them to determine the appropriate dose strengths for the clinical trial materials (if a capsule or tablet is to be used) and to determine whether the dose will be feasible from a logistical or commercial perspective (e.g. if a dose of 500 mg t.i.d. is predicted, such a dose may have logistical and commercial challenges).

In some cases, the pharmacokineticist may perform a series of simulations in order to predict what the exposure to the parent drug will be at the dose levels that may be used in the FIH clinical trial. This information can be used to help select the doses to be used in the clinical trial, although the starting dose will most likely be based upon the toxicology data as indicated here. Additionally, these data may be used to give an estimate of the highest dose that may be administered in the FIH study if the toxicology data indicate that a dose cap is appropriate. In this scenario, the maximum dose would be re-evaluated on an ongoing basis based on the actual human pharmacokinetic data.

A simulation of the predicted human plasma concentration-versus-time curve could be useful in a couple of situations. Firstly, this information may be useful in the selection of the times at which to collect blood for quantitation of the plasma concentration of parent drug, although the selection of these time points is most commonly driven

by logistical issues rather than by the predicted plasma concentration-versus-time curve. Secondly, and of greater utility, is the use of the simulated plasma concentrations to aid the bioanalytical scientist in the selection of the dynamic range that may be required for the human plasma assay and the potential dilution scheme that may be required for the analysis of the samples. It should be noted that the results of the simulation of the plasma concentration-versus-time curve will be dependent upon several assumptions including the linearity of the pharmacokinetics and the bioavailability of the drug. Usually, dose proportionality (linearity) of the PK parameters is assumed and the simulations are performed for a range of bioavailabilities.

As can be seen, the line between nonclinical and clinical ADME can be very fine, if it indeed exists at all. In some pharmaceutical companies, different groups will be responsible for the nonclinical and clinical work. In other companies, the same group will be responsible for both the nonclinical and clinical work.

The studies required to support FIH clinical trials is being addressed by the ICH as part of the maintenance of ICHM3 as ICHM3(R2). The EMEA is planning to issue a draft guideline in 4Q 2008 based on the 1Q 2006 concept paper, "Guideline on the Development of a CHMP Guideline on the Non-Clinical Requirements to Support Early Phase I Clinical Trials with Pharmaceutical Compounds" (EMEA, 2006d).

## 5 CLINICAL GUIDANCES AND CLINICAL TRIALS

Pharmacokinetics is a key element of clinical studies and is pivotal to understanding the relationship between dose or exposure and effect (either therapeutic or toxic). Like drug metabolism, pharmacokinetics is inherent in many guidances or guidelines and occasionally is the main subject of a guidance document. Since metabolism is one of the factors that shapes the pharmacokinetics of a drug, the drug metabolism scientist should pay attention to guidances that relate to pharmacokinetics as well as drug metabolism. The guidances that contain information about pharmacokinetics are many and varied. These guidances emphasize the importance of pharmacokinetics, the role of protein binding in understanding the PK or PK/PD relationships, the need to develop sensitive and specific assays, and the need to determine the pharmacokinetics of metabolites and/or enantiomers (as appropriate).

### 5.1 General

Many elements of drug metabolism and drug metabolites in clinical trials are covered in the ICH guidance, "E6(R1) Good Clinical Practice: Consolidated Guidance" that was issued in 1996 (ICH, 1996a). Among many other topics, this guidance describes the content of the investigator's brochure (IB). The guidance states that the IB should contain a brief summary (preferably not exceeding two pages), highlighting the significant physical, chemical, pharmaceutical, pharmacological, toxicological, pharmacokinetic, metabolic, and clinical information available that is relevant to the stage of clinical development of the investigational product. It also states that "The results of all relevant nonclinical pharmacology, toxicology, pharmacokinetic, and investigational product metabolism studies should be provided in summary form. This summary should address the methodology used, the results, and a discussion of the relevance of the findings to the investigated therapeutic and the possible unfavorable and unintended effects in humans." Further detail is provided for the different nonclinical disciplines as follows.

*Nonclinical Pharmacology*

A summary of the pharmacological aspects of the investigational product and, where appropriate, its significant metabolites studied in animals should be included. Such a summary should incorporate studies that assess potential therapeutic activity (e.g., efficacy models, receptor binding, and specificity) as well as those that assess safety (e.g., special studies to assess pharmacological actions other than the intended therapeutic effect(s)).

*Pharmacokinetics and Product Metabolism in Animals*

A summary of the pharmacokinetics and biological transformation and disposition of the investigational product in all species studied should be given. The discussion of the findings should address the absorption and the local and systemic bioavailability of the investigational product and its metabolites, and their relationship to the pharmacological and toxicological findings in animal species.

*Pharmacokinetics and Product Metabolism in Humans*

A summary of information on the pharmacokinetics of the investigational product(s) should be presented, including the following, if available: Pharmacokinetics (including metabolism, as appropriate, and absorption, plasma protein binding, distribution, and elimination).

It should be noted that the IB should be a summary of the relevant information and not an extensive overview of the complete knowledge of the ADME properties of the drug. Summarizing the ADME information in a few pages can be a challenge.

Further general points can be found in another ICH guidance, ICHE8, "General Considerations for Clinical Trials" that was issued in 1997 (ICH, 1997a). This guidance states that:

Pharmacokinetic studies are particularly important to assess the clearance of the drug and to anticipate possible accumulation of parent drug or metabolites and potential drug-drug interactions. Some pharmacokinetic studies are commonly conducted in later phases to answer more specialised questions. For many orally administered drugs, especially modified release products, the study of food effects on bioavailability is important. Obtaining pharmacokinetic information in sub-populations such as patients with impaired elimination (renal or hepatic failure), the elderly, children, women and ethnic subgroups should be considered. Drug-drug interaction studies are important for many drugs; these are generally performed in phases beyond Phase I but studies in animals and in vitro studies of metabolism and potential interactions may lead to doing such studies earlier.

The guidance also mentions some special considerations:

A number of special circumstances and populations require consideration on their own when they are part of the development plan.

*Studies of Drug Metabolites*

Major active metabolite(s) should be identified and deserve detailed pharmacokinetic study. Timing of the metabolic assessment studies within the development plan depends on the characteristics of the individual drug.

*Drug-Drug Interactions*

If a potential for drug–drug interaction is suggested by metabolic profile, by the results of non-clinical studies or by information on similar drugs, studies on drug interaction during

> clinical development are highly recommended. For drugs that are frequently co-administered it is usually important that drug-drug interaction studies be performed in non-clinical and, if appropriate in human studies. This is particularly true for drugs that are known to alter the absorption or metabolism of other drugs (see ICH E7), or whose metabolism or excretion can be altered by effects by other drugs.
>
> *Special Populations*
>
> Some groups in the general population may require special study because they have unique risk/benefit considerations that need to be taken into account during drug development, or because they can be anticipated to need modification of use of the dose or schedule of a drug compared to general adult use. Pharmacokinetic studies in patients with renal and hepatic dysfunction are important to assess the impact of potentially altered drug metabolism or excretion. Other ICH documents address such issues for geriatric patients (ICH E7) and patients from different ethnic groups (ICH E5). The need for non-clinical safety studies to support human clinical trials in special populations is addressed in the ICH M3 document.
>
> Excretion of the drug or its metabolites into human milk should be examined where applicable. When nursing mothers are enrolled in clinical studies their babies should be monitored for the effects of the drug.

DDIs and PK in special populations will be addressed in greater depth later in this section. The mention of ICHE7 in the DDI section of the ICHE8 guidance is interesting since the main thrust of the ICHE7 guidance is the clinical evaluation of drugs in the elderly (ICH, 1994c).

The EMEA guideline (EMEA, 2005b), "Guideline on Risk Management Systems for Medicinal Products for Human Use" that was issued in 2005 states that:

> Sometimes, potential risks or unforeseen benefits in special populations might be identified from preapproval clinical trials, but cannot be fully quantified due to small sample sizes or the exclusion of subpopulations of patients from these clinical studies. These populations might include the elderly, children, or patients with renal or hepatic disorder. Children, the elderly, and patients with co-morbid conditions might metabolize drugs differently from patients typically enrolled in clinical trials. Further clinical trials might be used to determine and to quantify the magnitude of the risk (or benefit) in such populations.

## 5.2 Pharmacokinetics and Pharmacodynamics

***5.2.1 General*** One of the earliest EMEA guidelines related to drug metabolism was "Pharmacokinetic Studies in Man Legislative basis Directive 75/318/EEC as amended" which came into force in 1988 (European Economic Community, 1987). This directive contains some basic concepts such as "The percentage and characteristics of binding to serum proteins should be studied using appropriate ex vivo or in vitro methods. Particularly in the case of substances or their active metabolites of which a high percentage is bound to plasma proteins, factors which might alter protein binding and so alter therapeutic response should be studied. Binding to red blood cells and other blood components should also be known." Red blood cells and lipoproteins may be important factors for the disposition of certain drugs (Wasan and Cassidy, 1998; Wasan *et al.*, 2008). However, the contribution of these factors is rarely evaluated. Metabolism is addressed in the following manner in this guidance:

> With a few exceptions substances are to a greater or lesser extent subject to metabolic breakdown within the human body. Pharmacokinetic studies should indicate whether the rate of biotransformation may be substantially modified in case of genetic enzymatic deficiency and whether within the dosage levels normally used, saturation of metabolism may occur, thereby inducing non-linear kinetics. The possibility of enzymatic induction should also be studied if metabolic clearance as a fraction of the systemic clearance is relatively high. If there is an indication that pharmacologically active metabolites (the qualitative activity of which may also occasionally differ from that of the parent substance) are formed, this should be ascertained and, if there is reason to suspect that they contribute to a significant extent to the therapeutic activity and/or adverse reactions in man, they should be examined in suitable animal models or if necessary in appropriate human clinical pharmacological studies. The pharmacokinetic data on such metabolites, the rate of their formation and elimination and their distribution and clearance characteristics should be known.

Excretion is another topic of relevance, although the directive only addresses urinary excretion and not fecal excretion. The reason for this omission is not clear, but it may be as a consequence of the lack of availability of assays to measure drug concentration in fecal samples and the lack of utility of such assays for orally administered drug. In this latter case, the quantitation of drug in the feces cannot provide insight as to whether the drug had been absorbed and eliminated in the bile without metabolism or simply had never been absorbed. The directive does make an interesting statement: "If urinary data are obtained, the urine should be collected until there is no further detectable excretion of parent substance or metabolites within the limits of the method used." What makes this statement interesting is that *a priori* you will not be able to determine how long it will take for the drug to be cleared and for there to be no further detectable excretion of parent substance or metabolites. Additionally, no statement is made as to whether the intent is to measure all metabolites or only major metabolites.

***5.2.2 Exposure–Response*** The contribution of metabolites to the exposure–response relationship is addressed in the 2003 FDA guidance, "Exposure-Response Relationships—Study Design, Data Analysis, and Regulatory Applications" (US FDA, 2003b). This guidance suggests that it is not only important to establish an exposure–response relationship but that it is also important to understand when there is no such response. The guidance states that "For example, it might be reassuring to observe that even patients with increased plasma concentrations (e.g., metabolic outliers or patients on other drugs in a study) do not have increased toxicity in general or with respect to a particular concern (e.g., QT prolongation)."

The concept that metabolism may be different depending upon the patient population is also introduced "In some cases, if there is a change in the mix of parent and active metabolites from one population (e.g., pediatric vs. adult), dosage form (e.g., because of changes in drug input rate), or route of administration, additional exposure-response data with short-term endpoints can support use in the new population, the new product, or new route without further clinical trials." However, the guidance does not address how the change in the mix of parent and active metabolites is determined. The implication is that active metabolites are identified in adults and then the exposure to these metabolites determined in the pediatric population. The role of drug metabolism and drug metabolites in the pediatric population is discussed later in this section.

The guidance also addresses the "Chemical Moieties for Measurement." For active moieties it states:

> To the extent possible, it is important that exposure-response studies include measurement of all active moieties (parent and active metabolites) that contribute significantly to the effects of the drug. This is especially important when the route of administration of a drug is changed, as different routes of administration can result in different proportions of parent compound and metabolites in plasma. Similarly, hepatic or renal impairment or concomitant drugs can alter the relative proportions of a drug and its active metabolites in plasma.

The role of drug metabolism and drug metabolites in renal and hepatic impairment is discussed later in this section. This guidance states that "Blood concentrations can also be helpful when both parent drug and metabolites are active." It should be noted that this guidance seems to use the terms "blood concentrations" and "plasma concentrations" interchangeably, and does not seem to be attempting to imply that drug concentrations should be measured in whole blood samples.

***5.2.3 Bioequivalence*** One of the major topics for regulatory guidances that relate to clinical pharmacology is bioequivalence (BE). Such studies are required for the approval of generic products or when changes are made to formulations during the course of development of the originator product. BE studies are used to "bridge" between data sets and should only be conducted when absolutely necessary since they are highly regulated studies with rigid acceptance criteria. If a BE study fails to show comparability, there may be delays in the development program. At other times, a relative bioavailability study (without rigid acceptance criteria) may be acceptable. For example, if a phase I study is performed with a capsule formulation and the phase IIa study will be performed with a tablet formulation, a BE study would most likely not be required. Instead, a simple comparison of the relative BA of the two formulations (in a study design similar to a BE study but with fewer subjects and less stringent acceptance criteria) would enable the sponsor to determine whether any dosage adjustment is required in the planned doses for the phase IIa study. One such guidance was issued by the FDA in 2003, "Bioavailability and Bioequivalence Studies for Orally Administered Drug Products—General Considerations" (US FDA, 2003c). In the United States, bioequivalence (BE) is defined as the absence of a significant difference in the rate and extent to which an active ingredient or active moiety in pharmaceutical equivalents or pharmaceutical alternatives becomes available at the same molar dose under similar conditions in an appropriately designed study. The FDA's 2003 guidance states that "BA can be generally documented by a systemic exposure profile obtained by measuring drug and/or metabolite concentration in the systemic circulation over time." The guidance further states that:

> BA for orally administered drug products can be documented by developing a systemic exposure profile. A profile can be obtained by measuring the concentration of active ingredients and/or active moieties and, when appropriate, its active metabolites over time in samples collected from the systemic circulation. Systemic exposure patterns reflect both release of the drug substance from the drug product and a series of possible presystemic/systemic actions on the drug substance after its release from the drug product. We recommend that additional comparative studies be performed to understand the relative contribution of these processes to the systemic exposure pattern.

The topic of the quantitation of parent drug or metabolites is also covered:

> The moieties to be measured in biological fluids collected in BA and BE studies are either the active drug ingredient or its active moiety in the administered dosage form (parent

drug) and, when appropriate, its active metabolites (21 CFR 320.24(b)(1)(i)). This guidance recommends the following approaches for BA and BE studies.

For BA studies, we recommend that determination of moieties to be measured in biological fluids take into account both concentration and activity. Concentration refers to the relative quantity of the parent drug or one or more metabolites in a given volume of an accessible biological fluid such as blood or plasma. Activity refers to the relative contribution of the parent drug and its metabolite(s) in the biological fluids to the clinical safety and/or efficacy of the drug. For BA studies, we also recommend that both the parent drug and its major active metabolites be measured, if analytically feasible.

For BE studies, measurement of only the parent drug released from the dosage form, rather than the metabolite, is generally recommended. The rationale for this recommendation is that concentration-time profile of the parent drug is more sensitive to changes in formulation performance than a metabolite, which is more reflective of metabolite formation, distribution, and elimination. The following are exceptions to this general approach.

- Measurement of a metabolite may be preferred when parent drug levels are too low to allow reliable analytical measurement in blood, plasma, or serum for an adequate length of time. We recommend that the metabolite data obtained from these studies be subject to a confidence interval approach for BE demonstration.
- A metabolite may be formed as a result of gut wall or other presystemic metabolism. If the metabolite contributes meaningfully to safety and/or efficacy, we also recommend that the metabolite and the parent drug be measured. When the relative activity of the metabolite is low and does not contribute meaningfully to safety and/or efficacy, it does not have to be measured. We recommend that the parent drug measured in these BE studies be analyzed using a confidence interval approach. The metabolite data can be used to provide supportive evidence of comparable therapeutic outcome.

The guidance also states that "We recommend that under normal circumstances, blood, rather than urine or tissue, be used. In most cases, drug, or metabolites are measured in serum or plasma."

Racemates and enantiomers are discussed in this guidance, "Many drugs are optically active and are usually administered as the racemate. Enantiomers sometimes differ in both their pharmacokinetic and pharmacodynamic properties. Early elucidation of the PK and PD properties of the individual enantiomers can help in designing a dosing regimen and in deciding whether it can be of value to develop one of the pure enantiomers as the final drug product."

The significance of unbound drug and/or active metabolite (protein binding) is also discussed:

Most standard assays of drug concentrations in plasma measure the total concentration, consisting of both bound and unbound drug. Renal or hepatic diseases can alter the binding of drugs to plasma proteins. These changes can influence the understanding of PK and PK-PD relationships. Where feasible, studies to determine the extent of protein binding and to understand whether this binding is or is not concentration-dependent are important, particularly when comparing responses in patient groups that can exhibit different plasma protein binding (e.g., in various stages of hepatic and renal disease). For highly protein bound drugs, PK and PK-PD modeling based on unbound drug concentrations may be more informative, particularly if there is significant variation in binding among patients or in special populations of patients.

This guidance does not give any insight into what would constitute a "significant difference in binding between individuals." Some of the issues and challenges with plasma protein binding have been addressed previously. In addition to this guidance, bioequivalence studies are captured in the CFRs (21 CFR Part 320) (US FDA, 2007c).

Bioequivalence is a topic that is addressed extensively in the FDA's Critical Path Initiative documents, indicating that the pharmaceutical industry will likely receive more guidance on this topic (US FDA, 2008d).

At the present time, the EMEA is in the process of revising its guideline "Note for Guidance on the Investigation of Bioavailability and Bioequivalence CPMP/EWP/QWP/1401/98" that came into force in 2002 and replaced a previous guideline that was issued in 1991 (EMEA, 2001). This guideline includes mention of metabolites in the section "Characteristics to be Investigated." A Q&A document that included a discussion of the use of metabolites in BE testing was issued for this guideline in 2006 (EMEA, 2006j). The comment period on the document "Recommendation on the Need for Revision of ⟨Note for Guidance on the Investigation of Bioavailability and Bioequivalence⟩ CPMP/EWP/QWP/1401/98" ended in August 2007 (EMEA, 2007h). The topics addressed in the Q&A document are addressed in this latter document.

Points of interest in the Q&A and recommendation document include the following.

*When should metabolite data be used to establish bioequivalence?*

"According to the guideline, the only situations where metabolite data can be used to establish bioequivalence are if the concentration of the active substance is too low to be accurately measured in the biological matrix, thus giving rise to significant variability."

Comments: Metabolite data can only be used if the Applicant presents convincing, state-of-the art arguments that measurements of the parent compound are unreliable. Even so, it is important to point out that $C_{max}$ of the metabolite is less sensitive to differences in the rate of absorption than $C_{max}$ of the parent drug. Therefore, when the rate of absorption is considered of clinical importance, bioequivalence should, if possible, be determined for $C_{max}$ of the parent compound, if necessary at a higher dose. Furthermore, when using metabolite data as a substitute for parent drug concentrations, the applicant should present data supporting the view that the parent drug exposure will be reflected by metabolite exposure.

*If metabolites significantly contribute to the net activity of an active substance and the pharmacokinetic system is non-linear.*

Comments: To evaluate the significance of the contribution of metabolites, relative AUCs and non-clinical or clinical pharmacodynamic activities should be compared with those of the parent drug. PK/PD modeling may be useful. If criteria for significant contribution to activity and pharmacokinetic non-linearity are met, then "it is necessary to measure both parent drug and active metabolite plasma concentrations and evaluate them separately." Any discrepancy between the results obtained with the parent compound and the metabolites should be discussed based on relative activities and AUCs. If the discrepancy lies in $C_{max}$, the results of the parent compound should usually prevail. Pooling of the plasma concentrations or pharmacokinetic parameters of the parent drug and its metabolite for calculation of bioequivalence is not acceptable.

*When using metabolite data to establish bioequivalence, may one use the same justification for widening the $C_{max}$ acceptance criteria as in the case of the parent compound?*

Comments: In principle, the same criteria apply as for the parent drug (see Question on widening the acceptance range for $C_{max}$). However, as stated above (see Question regarding when metabolite data can be used), $C_{max}$ of the metabolite is less sensitive to differences in the rate of absorption than $C_{max}$ of the parent drug. Therefore, widening the $C_{max}$ acceptance range when using metabolites instead of the parent compound is generally not accepted. When the metabolite has a major contribution to, or is completely responsible for, the therapeutic effect, and if it can be demonstrated that a widened acceptance range would not lead to any safety or efficacy concerns, which will usually prove more difficult than for the parent compound (see Question on widening the acceptance range for $C_{max}$), then a widened acceptance range for $C_{max}$ of metabolite may be accepted.

It is interesting to note that the original guideline (Note for Guidance on the Investigation of Bioavailability and Bioequivalence CPMP/EWP/QWP/1401/98) (EMEA, 2001) is not located in the Efficacy section of the EMEA guidelines are as the subsequent documents that pertain to bioequivalence, although a hyperlink to the original guideline can be found in the Efficacy guidelines section. Instead, it resides in the Quality section of the guidelines.

The Q&A document also mentions the subject of bioequivalence of highly variable drugs. This topic has also been addressed in a discussion document by Health Canada (2003), at the Clinical Pharmacology Subcommittee of the FDA's Pharmaceutical Science Advisory Committee (US FDA, 2006e)and in recent publications by the FDA (Haidar *et al.*, 2008; Davit *et al.*, 2008). Haidar *et al.* (2008) summarize the challenges for highly variable drugs by stating that "Over the past decade, concerns have been expressed increasingly regarding the difficulty for highly variable drugs and drug products (%CV greater than 30) to meet the standard bioequivalence (BE) criteria using a reasonable number of study subjects. The topic has been discussed on numerous occasions at national and international meetings. Despite the lack of a universally accepted solution for the issue, regulatory agencies generally agree that an adjustment of the traditional BE limits for these drugs or products may be warranted to alleviate the resource burden of studying relatively large numbers of subjects in bioequivalence trials." Davit *et al.* (2008) reviewed 1010 acceptable bioequivalence studies for 180 drugs that were submitted to the FDA between 2003 and 2005. Of these drugs, 57/180 (31%) were highly variable. They concluded that about 60% of the highly variable drugs were highly variable due to drug substance pharmacokinetic characteristics. For about 20% of the highly variable drugs, it appeared that formulation performance contributed to the high variability. For the remaining 20% of the highly variable drugs, it was not possible to identify the factors that contributed to the high variability.

The subject of the effect of food on the pharmacokinetics of a drug has been addressed in the FDA's 2002 guidance, "Food-effect Bioavailability and Fed Bioequivalence Studies" and in the EMEA's 2000 guidance "Note for Guidance on Modified Release Oral and Transdermal Dosage Forms: Section II (Pharmacokinetic and Clinical Evaluation)" (EMEA, 1999; US FDA, 2002b). The FDA guidance notes that "Food can change the BA of a drug and can influence the BE between test and reference products. Food effects on BA can have clinically significant consequences. Food can alter BA by various means, including:

- Delay gastric emptying
- Stimulate bile flow
- Change gastrointestinal (GI) pH

- Increase splanchnic blood flow
- Change luminal metabolism of a drug substance
- Physically or chemically interact with a dosage form or a drug substance"

The FDA's guidance also describes a standard test meal which is a high-fat (approximately 50% of total caloric content of the meal) and high-calorie (approximately 800 to 1000 calories) meal. This test meal should derive approximately 150, 250, and 500–600 calories from protein, carbohydrate, and fat, respectively. An example test meal would be two eggs fried in butter, two strips of bacon, two slices of toast with butter, four ounces of hash brown potatoes, and eight ounces of whole milk. Substitutions in this test meal can be made as long as the meal provides a similar amount of calories from protein, carbohydrate, and fat, and has comparable meal volume and viscosity. This meal represents an "extreme" in terms of diet and will not reflect the standard meal in most places around the world.

Bioequivalence studies can be costly and take a long time to conduct. Dependent upon the physicochemical properties of the drugs, there may be a scientific rationale that such studies may not be required for certain drugs. This concept is embodied in the FDA's guidance, "Waiver of In Vivo Bioavailability and Bioequivalence Studies for Immediate-Release Solid Oral Dosage Forms Based on a Biopharmaceutics Classification System" (US FDA, 2000b).

> The BCS is a scientific framework for classifying drug substances based on their aqueous solubility and intestinal permeability. When combined with the dissolution of the drug product, the BCS takes into account three major factors that govern the rate and extent of drug absorption from IR solid oral dosage forms: dissolution, solubility, and intestinal permeability.
>
> The BCS classifies drug substances as follows:
>
> Class I—high permeability, high solubility
> Class II—high permeability, low solubility
> Class III—low permeability, high solubility
> Class IV—low permeability, low solubility.
>
> In addition, IR solid oral dosage forms are categorized as having rapid or slow dissolution. Within this framework, when certain criteria are met, the BCS can be used as a drug development tool to help sponsors justify requests for biowaivers for highly soluble and highly permeable drug substances (i.e. Class I) in IR solid oral dosage forms that exhibit rapid *in vitro* dissolution using the recommended test methods.

The drug metabolism scientist/pharmacokineticist should play a role in creating the application for a biowaiver by providing information on the permeability of the drug. However, this is not always the case and often the formulation scientist will make the assessment independent of input from ADME specialists. This is an area in which there should be close collaboration between the formulation and ADME groups.

The FDA guidance states:

> The permeability class boundary is based indirectly on the extent of absorption (fraction of dose absorbed, not systemic BA) of a drug substance in humans and directly on measurements of the rate of mass transfer across human intestinal membrane. Alternatively, nonhuman systems capable of predicting the extent of drug absorption in humans can be

used (e.g., in vitro epithelial cell culture methods). In the absence of evidence suggesting instability in the gastrointestinal tract, a drug substance is considered to be *highly permeable* when the extent of absorption in humans is determined to be 90% or more of an administered dose based on a mass balance determination or in comparison to an intravenous reference dose.

The guidance provides extensive details regarding the determination of the drug substance permeability class:

The permeability class of a drug substance can be determined in human subjects using mass balance, absolute BA, or intestinal perfusion approaches. Recommended methods not involving human subjects include in vivo or in situ intestinal perfusion in a suitable animal model (e.g., rats), and/or in vitro permeability methods using excised intestinal tissues, or monolayers of suitable epithelial cells. In many cases, a single method may be sufficient (e.g., when the absolute BA is 90% or more, or when 90% or more of the administered drug is recovered in urine). When a single method fails to conclusively demonstrate a permeability classification, two different methods may be advisable. Chemical structure and/or certain physicochemical attributes of a drug substance (e.g., partition coefficient in suitable systems) can provide useful information about its permeability characteristics. Sponsors may wish to consider use of such information to further support a classification.

*1. Pharmacokinetic Studies in Humans*

a. Mass Balance Studies
Pharmacokinetic mass balance studies using unlabeled, stable isotopes or a radiolabeled drug substance can be used to document the extent of absorption of a drug. Depending on the variability of the studies, a sufficient number of subjects should be enrolled to provide a reliable estimate of extent of absorption. Because this method can provide highly variable estimates of drug absorption for many drugs, other methods described below may be preferable.

b. Absolute Bioavailability Studies
Oral BA determination using intravenous administration as a reference can be used. Depending on the variability of the studies, a sufficient number of subjects should be enrolled in a study to provide a reliable estimate of the extent of absorption. When the absolute BA of a drug is shown to be 90% or more, additional data to document drug stability in the gastrointestinal fluid is not necessary.

*2. Intestinal Permeability Methods*

The following methods can be used to determine the permeability of a drug substance from the gastrointestinal tract: (1) in vivo intestinal perfusion studies in humans; (2) in vivo or in situ intestinal perfusion studies using suitable animal models; (3) in vitro permeation studies using excised human or animal intestinal tissues; or (4) in vitro permeation studies across a monolayer of cultured epithelial cells.

In vivo or in situ animal models and in vitro methods, such as those using cultured monolayers of animal or human epithelial cells, are considered appropriate for passively transported drugs. The observed low permeability of some drug substances in humans could be caused by efflux of drugs via membrane transporters such as P-glycoprotein (P-gp). When the efflux transporters are absent in these models, or their degree of expression is low compared to that in humans, there may be a greater likelihood of misclassification of permeability class for a drug subject to efflux compared to a drug transported passively. Expression of known transporters in selected study systems should be characterized.

Functional expression of efflux systems (e.g., P-gp) can be demonstrated with techniques such as bidirectional transport studies, demonstrating a higher rate of transport in the basolateral-to-apical direction as compared to apical-to-basolateral direction using selected model drugs or chemicals at concentrations that do not saturate the efflux system (e.g., cyclosporin A, vinblastine, rhodamine 123). An acceptance criterion for intestinal efflux that should be present in a test system cannot be set at this time. Instead, this guidance recommends limiting the use of nonhuman permeability test methods for drug substances that are transported by passive mechanisms. Pharmacokinetic studies on dose linearity or proportionality may provide useful information for evaluating the relevance of observed in vitro efflux of a drug. For example, there may be fewer concerns associated with the use of in vitro methods for a drug that has a higher rate of transport in the basolateral-to-apical direction at low drug concentrations but exhibits linear pharmacokinetics in humans.

For application of the BCS, an apparent passive transport mechanism can be assumed when one of the following conditions is satisfied:

- A linear (pharmacokinetic) relationship between the dose (e.g., relevant clinical dose range) and measures of BA (area under the concentration-time curve) of a drug is demonstrated in humans
- Lack of dependence of the measured in vivo or in situ permeability is demonstrated in an animal model on initial drug concentration (e.g., 0.01, 0.1, and 1 times the highest dose strength dissolved in 250 ml) in the perfusion fluid
- Lack of dependence of the measured in vitro permeability on initial drug concentration (e.g., 0.01, 0.1, and 1 times the highest dose strength dissolved in 250 ml) is demonstrated in donor fluid and transport direction (e.g., no statistically significant difference in the rate of transport between the apical-to-basolateral and basolateral-to-apical direction for the drug concentrations selected) using a suitable in vitro cell culture method that has been shown to express known efflux transporters (e.g., P-gp).

To demonstrate suitability of a permeability method intended for application of the BCS, a rank-order relationship between test permeability values and the extent of drug absorption data in human subjects should be established using a sufficient number of model drugs. For in vivo intestinal perfusion studies in humans, six model drugs are recommended. For in vivo or in situ intestinal perfusion studies in animals and for in vitro cell culture methods, twenty model drugs are recommended. Depending on study variability, a sufficient number of subjects, animals, excised tissue samples, or cell monolayers should be used in a study to provide a reliable estimate of drug permeability. This relationship should allow precise differentiation between drug substances of low and high intestinal permeability attributes.

For demonstration of suitability of a method, model drugs should represent a range of low (e.g., <50%), moderate (e.g., 50–89%), and high (≥90%) absorption. Sponsors may select compounds from the list of drugs and/or chemicals provided in Attachment A or they may choose to select other drugs for which there is information available on mechanism of absorption and reliable estimates of the extent of drug absorption in humans.

After demonstrating suitability of a method and maintaining the same study protocol, it is not necessary to retest all selected model drugs for subsequent studies intended to classify a drug substance. Instead, a low and a high permeability model drug should be used as internal standards (i.e., included in the perfusion fluid or donor fluid along with the test drug substance). These two internal standards are in addition to the fluid volume marker (or a zero permeability compound such as PEG 4000) that is included in certain types of perfusion techniques (e.g., closed loop techniques). The choice of internal standards should be based on compatibility with the test drug substance (i.e., they should not exhibit any

significant physical, chemical, or permeation interactions). When it is not feasible to follow this protocol, the permeability of internal standards should be determined in the same subjects, animals, tissues, or monolayers, following evaluation of the test drug substance. The permeability values of the two internal standards should not differ significantly between different tests, including those conducted to demonstrate suitability of the method. At the end of an in situ or in vitro test, the amount of drug in the membrane should be determined.

For a given test method with set conditions, selection of a high permeability internal standard with permeability in close proximity to the low/high permeability class boundary may facilitate classification of a test drug substance. For instance, a test drug substance may be determined to be highly permeable when its permeability value is equal to or greater than that of the selected internal standard with high permeability.

*3. Instability in the Gastrointestinal Tract*

Determining the extent of absorption in humans based on mass balance studies using total radioactivity in urine does not take into consideration the extent of degradation of a drug in the gastrointestinal fluid prior to intestinal membrane permeation. In addition, some methods for determining permeability could be based on loss or clearance of a drug from fluids perfused into the human and/or animal gastrointestinal tract either in vivo or in situ. Documenting the fact that drug loss from the gastrointestinal tract arises from intestinal membrane permeation, rather than a degradation process, will help establish permeability. Stability in the gastrointestinal tract may be documented using gastric and intestinal fluids obtained from human subjects. Drug solutions in these fluids should be incubated at 37°C for a period that is representative of in vivo drug contact with these fluids; for example, 1 hour in gastric fluid and 3 hours in intestinal fluid. Drug concentrations should then be determined using a validated stability-indicating assay method. Significant degradation (>5%) of a drug in this protocol could suggest potential instability. Obtaining gastrointestinal fluids from human subjects requires intubation and may be difficult in some cases. Use of gastrointestinal fluids from suitable animal models and/or simulated fluids such as Gastric and Intestinal Fluids USP can be substituted when properly justified.

In May 2007, the CHMP issued a concept paper (Concept Paper on BCS-Based Biowaiver) for review (EMEA, 2007b). This concept paper describes some concerns that have risen regarding the use of the BCS system and aims to provide clarification leading to a harmonized approach to the use of the BCS-based biowaiver approach.

### 5.3 Phase 0 studies, Microdosing, and the Exploratory IND

In the past decade, a new phase of clinical development has been described. This phase of development is known as phase 0 and phase 0 studies are conducted before the traditional phase I studies are conducted. There is no universal definition of a phase 0 study, but the definition can include:

- Designed to study the PK/PD of a drug
- Studies a limited number of doses
- Studies lower doses than a traditional phase I study
- Utilizes fewer patients compared to a phase I study
- An evaluation of disease biomarkers.

Phase 0 studies are uncommon but can play a key role in the development of a drug. Phase 0 studies may be conducted under an exploratory IND which is a concept that came out of the FDA's "Critical Path Report" that was issued in 2004 and in which the Agency explained that to reduce the time and resources expended on candidate products that are unlikely to succeed, new tools are needed to distinguish earlier in the process those candidates that hold promise from those that do not (US FDA, 2004a). Prior to the issuance of a draft guidance on exploratory IND studies in April 2005 (US FDA, 2005b), the concept was known as a "screening IND." A MAPP (Manual of Policies and Procedures) was issued for screening INDs in 2001 (US FDA, 2001d). Over 350 comments were submitted to the docket. In 2006, the FDA issued the final guidance, "Exploratory IND Studies" that "describes some early phase 1 exploratory approaches that are consistent with regulatory requirements while maintaining needed human subject protection, but that involve fewer resources than is customary, enabling sponsors to move ahead more efficiently with the development of promising candidates" (US FDA, 2006f). Although this guidance does contain some procedural elements, it is located in the Pharm/Tox section of the FDA guidance page. The topic of exploratory INDs has recently been reviewed by Sarapa (2007).

The FDA's guidance contains the following introductory statement:

> This guidance is intended to clarify what preclinical and clinical approaches, as well as chemistry, manufacturing, and controls information, should be considered when planning exploratory studies in humans, including studies of closely related drugs or therapeutic biological products, under an investigational new drug (IND) application (21 CFR 312). Existing regulations allow a great deal of flexibility in the amount of data that needs to be submitted with an IND application, depending on the goals of the proposed investigation, the specific human testing proposed, and the expected risks. The Agency believes that sponsors have not taken full advantage of that flexibility and often provide more supporting information in INDs than is required by regulations. This guidance is intended to clarify what manufacturing controls, preclinical testing, and clinical approaches can be considered when planning limited, early exploratory IND studies in humans.
>
> For the purposes of this guidance the phrase *exploratory IND study* is intended to describe a clinical trial that:
>
> — is conducted early in phase 1
> — involves very limited human exposure, and
> — has no therapeutic or diagnostic intent (e.g., screening studies, microdose studies).
>
> Such exploratory IND studies are conducted prior to the traditional dose escalation, safety, and tolerance studies that ordinarily initiate a clinical drug development program. The duration of dosing in an exploratory IND study is expected to be limited (e.g., 7 days). This guidance applies to early phase 1 clinical studies of investigational new drug and biological products that assess feasibility for further development of the drug or biological product.

The guidance describes a traditional IND approach in order to contrast it to an exploratory IND approach:

> Typically, during pharmaceutical development, large numbers of molecules are generated with the goal of identifying the most promising candidates for further development. These molecules are generally structurally related, but can differ in important ways. Promising candidates are often selected using in vitro testing models that examine binding to recep-

> tors, effects on enzyme activities, toxic effects, or other in vitro pharmacologic parameters; these tests usually require only small amounts of the drug. Candidates that are not rejected during these early tests are prepared in greater quantities for in vivo animal testing for efficacy and safety. Commonly, a single candidate is selected for an IND application and introduction into human subjects, initially healthy volunteers in most cases.
>
> Before the human studies can begin, an IND must be submitted to the Agency containing, among other things, information on any risks anticipated based on the results of pharmacologic and toxicological data collected during studies of the drug in animals (21 CFR 312.23(a)(8)). These basic safety tests are most often performed in rats and dogs. The studies are designed to permit the selection of a safe starting dose for humans, to gain an understanding of which organs may be the targets of toxicity, to estimate the margin of safety between a clinical and a toxic dose, and to predict pharmacokinetic and pharmacodynamic parameters. These early tests are usually resource intensive, requiring significant investment in product synthesis, animal use, laboratory analyses, and time. Many resources are invested in, and thus wasted on, candidate products that subsequently are found to have unacceptable profiles when evaluated in humans—less than 10 percent of INDs for new molecular entities (NME) progress beyond the investigational stage to submission of a marketing application (NDA). In addition, animal testing does not always predict performance in humans, and potentially effective candidates may not be developed because of resource constraints.
>
> Existing regulations allow a great deal of flexibility in terms of the amount of data that need to be submitted with any IND application, depending on the goals of the proposed investigation, the specific human testing proposed, and the expected risks. The Agency believes that sponsors have not taken full advantage of that flexibility. As a result, limited, early phase 1 studies, such as those described in this guidance, are often supported by a more extensive preclinical database than is required by the regulations.
>
> This guidance describes preclinical and clinical approaches, and the chemistry, manufacturing, and controls information that should be considered when planning exploratory IND studies in humans, including studies of closely related drugs or therapeutic biological products, under a single IND application (21 CFR 312).

The Exploratory IND approach is described thus:

> Exploratory IND studies usually involve very limited human exposure and have no therapeutic or diagnostic intent. Such studies can serve a number of useful goals. For example, an exploratory IND study can help sponsors
>
> - Determine whether a mechanism of action defined in experimental systems can also be observed in humans (e.g., a binding property or inhibition of an enzyme)
> - Provide important information on pharmacokinetics (PK)
> - Select the most promising lead product from a group of candidates designed to interact with a particular therapeutic target in humans, based on PK or pharmacodynamic (PD) properties
> - Explore a product's biodistribution characteristics using various imaging technologies

One benefit that the FDA identifies for the use of an exploratory IND is the ability to select "the most promising lead product from a group of candidates" based on PK or PD. In principle, this would be advantageous. However, in practice it is rare that a sponsor would have several potential clinical candidates that could be evaluated simultaneously in the clinic and for which there were no distinguishing features based on

nonclinical data. Under these circumstances, it would be beneficial to select the candidate that had the best PK or PD properties in humans. It should be noted that even if the compound with the best PK or PD properties were selected from such a "screening" IND, it would not necessarily mean that the selected compound would not fail in a traditional phase I study that would be conducted subsequently. Since the intent of studies conducted an exploratory IND is not to cause toxicity, the exploratory IND clinical trial could not "weed out" compounds that may show toxicity at doses higher than those studies in the exploratory IND study and could only select compounds that had suboptimal PK or PD properties. It should be noted that if a drug were to fail due to inappropriate PK properties, it would fail regardless of whether an exploratory or traditional IND approach was used. As mentioned earlier, it is unlikely that all of the compounds would be available at exactly the same time and that the filing of the exploratory IND or the clinical study that is to be conducted under the exploratory IND may be delayed until all of the compounds had been synthesized and the appropriate studies conducted. This may introduce a delay into the overall development plan. Even if the company intended to take, for example, three compounds into an exploratory IND, it may be likely that one or more of them drops out of contention based on data generated in the toxicology studies that are conducted to support the exploratory IND filing. If this were to happen, some of the benefits of conducting a clinical trial under and exploratory IND would be lost. Although the costs of taking one compound into a clinical trial to be conducted under an exploratory IND will be lower than the costs of taking one compound into a clinical trial to be conducted under a traditional IND, the cost savings will dwindle as more compounds are included in the exploratory IND. Therefore, the sponsor should only consider evaluating those compounds which may have properties that would suggest that they could become marketed drugs.

The guidance further states:

> Whatever the goal of the study, exploratory IND studies can help identify, early in the process, promising candidates for continued development and eliminate those lacking promise. As a result, exploratory IND studies may help reduce the number of human subjects and resources, including the amount of candidate product, needed to identify promising drugs. The studies discussed in this guidance involve dosing a limited number of subjects with a limited range of doses for a limited period of time. Existing regulations provide more flexibility with regard to the preclinical testing requirements for exploratory IND studies than for traditional IND studies. However, sponsors submitting the kinds of studies described in this guidance have not always taken full advantage of that flexibility. Sponsors often provide more supporting information in their INDs than is required by the regulations. Because exploratory IND studies involve administering either subpharmacologic doses of a product, or doses expected to produce a pharmacologic, but not a toxic, effect, the potential risk to human subjects is less than for a traditional phase 1 study that, for example, seeks to establish a maximally tolerated dose. Because exploratory IND studies present fewer potential risks than do traditional phase 1 studies that look for dose-limiting toxicities, such limited exploratory IND investigations in humans can be initiated with less, or different, preclinical support than is required for traditional IND studies.
>
> The Agency expects that this early phase 1 exploratory IND approach will apply to a number of different study paradigms. Although his guidance explores several potential applications, many others can be proposed. The Agency believes that, consistent with its Critical Path Initiative, clarifying Agency thinking about how much and what kind of

> testing is needed to support early studies in humans will facilitate the entry of new products into clinical testing and speed product development.
>
> Although exploratory IND studies may be used during development of products intended for any indication, it is particularly important for manufacturers to consider this approach when developing products to treat serious diseases. Because the approach can help identify promising candidates more quickly and precisely, exploratory IND studies could become an important part of the armamentarium when developing drug and biological products to treat a serious or life-threatening illness. The Agency has previously articulated its commitment to ensuring that appropriate flexibility is applied when patients with a serious disease and no satisfactory alternative therapies are enrolled in a trial with therapeutic intent.

The guidance described types of studies that can be conducted under the exploratory IND. "Potentially useful study designs include both single- and multiple-dose studies. In single-dose studies, a sub-pharmacologic or pharmacologic dose is administered to a limited number of subjects (healthy volunteers or patients). For example, microdose studies usually involve the single administration of a small dose with the goal of collecting pharmacokinetic information or performing imaging studies, or both." A footnote gives further clarification of a sub-pharmacologic study, "A radiolabeled candidate compound can be administered at doses that are known to have no pharmacologic effect in humans without an IND application in basic research studies when the compound has previously been studied in humans and the results published in the literature. These basic research investigations are conducted under the oversight of an institutional review board (IRB) and a radioactive drug research committee (21 CFR 361.1)." This statement could be interpreted as implying that a microdose study could be conducted without an IND, although to do so would require that the sponsor know that the doses would have no pharmacological effect.

Repeat-dose clinical studies can be designed with pharmacologic or pharmacodynamic end points. In exploratory IND studies, the duration of dosing should be limited (e.g. seven days). For escalating dose studies done under an exploratory IND, dosing should be designed to investigate a pharmacodynamic end point, not to determine the limits of tolerability.

A large part of the guidance is devoted to a description of the design of safety evaluation programs. A general statement is made regarding the nonclinical safety assessment program followed by specific examples for different clinical study designs:

> Pharmacology and toxicology information is derived from preclinical safety testing performed in animals and in vitro. Preclinical studies for small molecules are described in ICH M3 while those for biologics follow guidance described in ICH S6. Some of the toxicology tests described in this guidance may not be appropriate for biologics. The toxicology evaluation recommended for an exploratory IND application is more limited than for a traditional IND application. The basis for the reduced preclinical package is the reduced scope of an exploratory IND clinical study. Although exploratory IND studies in some cases are expected to induce pharmacologic effects, they are not designed to establish maximally tolerated doses. Furthermore, the duration of drug exposure in exploratory IND studies is limited. The level of preclinical testing performed to ensure safety will depend on the scope and intended goals of the clinical trials.
>
> There are a number of study objectives for which the preclinical safety programs may be tailored to the study design. Examples include: confirming that an expected mechanism of

action can be observed in humans; measuring binding affinity or localization of drug; assessing PK and metabolism; comparing the effect on a potential therapeutic target with other therapies.

*Clinical studies of pharmacokinetics or imaging*

Microdose studies are designed to evaluate pharmacokinetics or imaging of specific targets and are designed not to induce pharmacologic effects. Because of this, the risk to human subjects is very limited, and information adequate to support the initiation of such limited human studies can be derived from limited nonclinical safety studies. A *microdose* is defined as less than 1/100th of the dose of a test substance calculated (based on animal data) to yield a pharmacologic effect of the test substance with a maximum dose of <100 micrograms (for imaging agents, the latter criterion applies). Due to differences in molecular weights as compared to synthetic drugs, the maximum dose for protein products is ≤30 nanomoles.

FDA currently accepts the use of extended single-dose toxicity studies in animals to support single-dose studies in humans. For microdose studies, a single mammalian species (both sexes) can be used if justified by in vitro metabolism data and by comparative data on in vitro pharmacodynamic effects. The route of exposure in animals should be by the intended clinical route. In these studies, animals should be observed for 14 days post-dosing with an interim necropsy, typically on day 2, and endpoints evaluated should include body weights, clinical signs, clinical chemistries, hematology, and histopathology (high dose and control only if no pathology is seen at the high dose). The study should be designed to establish a dose inducing a minimal toxic effect, or alternatively, establishing a margin of safety. To establish a margin of safety, the sponsor should demonstrate that a large multiple (e.g., 100X) of the proposed human dose does not induce adverse effects in the experimental animals. Scaling from animals to humans based on body surface area can be used to select the dose for use in the clinical trial. Scaling based on pharmacokinetic/pharmacodynamic modeling would also be appropriate if such data are available.

Because microdose studies involve only single exposures to microgram quantities of test materials and because such exposures are comparable to routine environmental exposures, routine genetic toxicology testing is not needed. For similar reasons, safety pharmacology studies are also not recommended.

The mention that an "extended single-dose toxicology study" can be used to support a single-dose study in humans is worthy of note (US FDA, 1996). Since many of the studies conducted under an exploratory IND may be single-dose clinical trials, it may be important for the sponsor to use an "extended single-dose toxicology study" to support the clinical trial rather than the more extensive set of studies outlined in the exploratory IND guidance. The more extensive toxicology studies described in this guidance will support some limited duration of dosing. However, many of the questions that could be answered under an exploratory IND could be answered using a single-dose clinical trial rather than a multiple-dose clinical trial.

*Clinical trials to study pharmacologically relevant doses*

A second example involves clinical trials designed to study pharmacologic effects of candidate products. More extensive preclinical safety data would be needed to support the safety of such studies. However, since the goal would not include defining a maximally tolerated dose, the evaluation can still be less extensive than typically needed to support a traditional IND application. See the flow chart in the Attachment to this document. Repeat dose clinical trials lasting up to 7 days can be supported by a 2-week repeat dose

toxicology study in a sensitive species accompanied by toxicokinetic evaluations. The goal of such a study would be to select safe starting and maximum doses for the clinical trial. The rat is the usual species chosen for this purpose, but other species might be selected. In addition to studies in a rodent species, additional studies in nonrodents, most often dogs, can be used to confirm that the rodent is an appropriately sensitive species. If it is known that a particular species is most appropriate for a class of compounds, studies can be limited to that species. This confirmation can be approached in a number of ways. A lack of gender difference in the rodent study can serve as a basis for testing only a single sex in the second species if only a single sex will be studied in the clinical trial.

The numbers of animals used in the confirmatory study can be fewer than normally used to attain statistically meaningful comparisons, but of sufficient number to rule out any toxicologically significant difference in sensitivity compared with rodent (e.g. four nonrodents per treatment group). The confirmatory study could be a dedicated study involving repeat administrations of a single dose level approximating the rat NOAEL calculated on the basis of body surface area. Alternatively, the test in the second species could be incorporated as part of an exploratory, dose escalating study culminating in repeated doses equivalent to the rat NOAEL. The number of repeat administrations at the rat NOAEL should, at a minimum, be equal to the number of administrations, given with the same schedule, intended clinically. The route of administration should be the same as the expected clinical route, and toxicokinetic measurements should be used to assess exposure. The same endpoints assessed in the rodent study should be evaluated in the second species. If the data from the confirmatory study suggest that the rodent is not the more sensitive species, a 2-week repeated dose toxicity study should be performed in the second species to select doses for human trials. This study should include measurements of body weight, clinical signs, clinical chemistries, hematology, and histopathology.

In contrast to microdose studies, for clinical trials designed to evaluate higher or repeated doses, each candidate product to be tested should be evaluated for safety pharmacology. Evaluation of the central nervous and respiratory systems can be performed as part the rodent toxicology studies while safety pharmacology for the cardiovascular system can be assessed in the nonrodent species, generally the dog, and can be conducted as part of the confirmatory or dose-escalation study.

In general, each product in this type of exploratory IND should be tested for potential genotoxicity unless such testing is not appropriate for the population (e.g. terminally ill patients) or product to be studied. The genetic toxicology tests should include a bacterial mutation assay using all five tester strains with and without metabolic activation as well as a test for chromosomal damage either in vitro (cytogenetics assay or mouse lymphoma thymidine kinase gene mutation assay) or in vivo. The in vivo test can be a micronucleus assay performed in conjunction with the repeated dose toxicity study in the rodent species. The high dose in this case should be a maximally tolerated or limit dose.

The results from the preclinical program can be used to select starting and maximum doses for the clinical trials. The starting dose is anticipated to be no greater than 1/50 of the NOAEL from the 2-week toxicology study in the sensitive species on a $mg/m^2$ basis. The maximum clinical dose would be the lowest of the following:

- $\frac{1}{4}$ of the 2-week rodent NOAEL on a $mg/m^2$ basis
- Up to $\frac{1}{2}$ of the AUC at the NOAEL in the 2-week rodent study, or the AUC in the dog at the rat NOAEL, whichever is lower
- The dose that produces a pharmacologic and/or pharmacodynamic response or at which target modulation is observed in the clinical trial
- Observation of an adverse clinical response

Escalation from the proposed maximal clinical dose should only be performed after consultation with and concurrence of the FDA.

It is recognized that the studies described above are most appropriate for chemical drugs. Other animal models (e.g. nonhuman primates) may be more appropriate for biologics, and some tests may be inappropriate (e.g. genetic toxicology testing) for proteins.

These stopping rules have several implications. The use of an AUC as a stopping rule implies real-time PK analysis which may slow down trial conduct. The guidance does not give insight whether the stopping rule would be based on mean data or data from individual subjects. It should be noted that stopping rules (based on PK or PD) are not routinely used in clinical trials unless there was a toxicity observed in animals that could not be monitored in humans.

The use of pharmacologic and/or pharmacodynamic response as a stopping rule raises the implied availability of precise, real-time biomarker assay. Again, waiting for PD data before dose escalating could delay the conduct of the clinical trial. There may be challenges in using a PD assay for dose escalation/stopping rules. The FIH clinical trial may be the first time that the PD assay has been used in humans and therefore the ruggedness and robustness of the assay will not have been fully tested. Therefore, it may not be feasible to utilize the PD assay as part of the stopping rules.

Another stopping rule advocated by the exploratory IND guidance is the observation of an adverse clinical response. However, the guidance does not indicate whether the adverse clinical response would need to be a serious adverse event or whether any adverse event (including a headache) would halt the study. The guidance allows for escalation from the proposed maximal clinical dose. However, this should only be performed after consultation with and concurrence of the FDA which will require close, real-time communication with the FDA and slow the conduct of the clinical trial.

*Clinical studies of MOAs related to efficacy*

A third example involves clinical studies intended to evaluate mechanisms of action (MOAs). To support this approach, the FDA will accept alternative, or modified, pharmacologic and toxicological studies to select clinical starting doses and dose escalation schemes. For example, short-term, modified toxicity or safety studies in two animal species based on a dosing strategy to achieve a clinical pharmacodynamic endpoint can in some instances serve as the basis for selecting the safe clinical starting dose for a new candidate drug. These animal studies would incorporate endpoints that are mechanistically based on the pharmacology of the new chemical entity and thought to be important to clinical effectiveness. For example, if the degree of saturation of a receptor or the inhibition of an enzyme were considered possibly related to effectiveness, this parameter would be characterized and determined in the animal study and then used as an endpoint in a subsequent clinical investigation. The dose and dosing regimen determined in the animal study would be extrapolated for use in the clinical investigation. In some cases, a single species could be used if it were established as the most relevant species based on scientific evidence using the specific candidate intended for the clinical investigation.

Although the production of frank toxicity is not the primary intended goal of the nonclinical study, relevant informative endpoints (e.g., hematology and histopathology) selected as important for clinical safety evaluation should be investigated. For example, an antibody that binds with a high degree of selectivity to a tumor-associated antigen could be studied in accordance with this third category. The mechanism of action of antibody-based products is generally associated with their binding properties and the effect on functions associated with immunoglobulins. Pharmacology and toxicology studies provide information about

> the selection of doses used in clinical studies through evidence of both a safe upper and potentially efficacious lower limit of exposure. These doses might be consistent with target plasma levels of the drug based on animal models of disease. The upper safe levels could be established in animal studies that show a lack of toxicity at these levels.

The exploratory IND guidance describes the toxicology studies that are required and provides a flowchart for these studies. Although the sequence of studies described in the flowchart provides for a minimization of the amount of work that is conducted to support the exploratory IND submission, the sequential nature of the tasks and the small difference in cost of the studies if they are conducted in parallel rather than in sequence, means that the guidance, if followed as written, may delay the initiation of the clinical trial. It should be noted that many pharmaceutical companies may have more stringent criteria (vis-à-vis toxicology) for nominating a compound to enter into clinical trials than are required by the exploratory IND guidance.

The guidance does not make any concessions regarding compliance of the nonclinical studies to GLP practices. "It is expected that all preclinical safety studies supporting the safety of an exploratory IND application will be performed in a manner consistent with good laboratory practices (GLP) (21 CFR Part 58). The GLP provisions apply to a broad variety of studies, test articles, and test systems. Sponsors are encouraged to discuss any need for an exemption from GLP provisions with the FDA prior to conducting safety related studies, for example, during a pre-IND meeting. Sponsors must justify any nonconformance with GLP provisions (21 CFR 312.23(a)(8)(iii))."

The content of the IND submissions is also addressed. In essence, there is little, if any, difference between the format and content of a traditional and an exploratory IND, although the type of information required may vary between these two types of IND as outlined in the guidance. "To begin any kind of testing in humans, applicants must submit an IND application to the Agency with certain types of information (see 21 CFR 312.23 IND Content and Format). The primary purpose of the IND submission is to ensure that subjects will not face undue risk of harm. The major information that must be submitted includes:

- Information on a clinical development plan
- Chemistry, manufacturing, and controls information
- Pharmacology and toxicology information
- Previous human experience with the investigational candidate or related compounds, if there is any.

Because the exploratory IND studies addressed by this guidance will be first in human studies, previous human experience is not pertinent."

With respect to clinical information, the guidance states:

> A traditional IND application describes the rationale for the proposed clinical trial program and discusses the potential outcome of the clinical investigation. The exploratory IND studies discussed here focus on a circumscribed study or group of studies, and plans for further development cannot be formulated without the results of these studies. Therefore, an exploratory IND application should articulate the rationale for selecting a compound (or compounds) and for studying them in a single trial or related trials, as this represents all that is known about the overall development plan at this stage. This section should also make it clear that the IND is intended to be withdrawn after completion of the outlined

> study or studies. The withdrawn, or inactive, IND can be referenced in any subsequent traditional IND.

This point is an important one, and one that may lead to subsequent delays in the development of any drug candidate that has shown promise in an exploratory IND study. More chemical will need to be conducted for subsequent toxicology and clinical trials and a two-week non-rodent toxicology study. The clinical study report from the study or studies conducted under the exploratory IND will need to be written, the investigator's brochure will need to be updated and submitted with updated IND summaries to a new IND. The exploratory IND will need to be closed out per the FDA guidance. This will introduce a 6- to 12-month delay which could be stressful time within the pharmaceutical company if the drug had performed well in the clinical study conducted under the exploratory IND.

The Chemistry, Manufacturing and Controls (CMC) Information that is required for an exploratory IND is less than that is required for a traditional IND. The guidance describes in detail the CMC requirements, describing two scenarios under which CMC information can be provided to an IND application. The guidance states that:

> In the first scenario, the ***same batch*** of candidate product is used in both the toxicology studies and clinical trials. This material will be qualified for human use based on the CMC information (see III.B.1, above) and results of the toxicology studies described elsewhere in this guidance. Although we recommend establishing the impurity profile to the extent possible for future reference and/or comparison, not all impurities of the candidate product may need characterization at this stage of product development. If an issue arises during the toxicology qualification of the product, the appropriate parameters can be studied further, on an as-needed basis. Impurities (e.g., chemical and microbiological) should be characterized in accordance with recommendations in Agency guidance, if, and when, the sponsor files a traditional IND for further clinical investigation.
>
> In the second scenario, the batch of candidate drug product to be used in the clinical studies may not be the same as that used in the nonclinical toxicology studies. In such a case, the sponsor should demonstrate by analytical testing that the batch to be used is ***representative*** of batches used in the nonclinical toxicology studies. To achieve this, relevant analytical quality test results should be sufficient to enable comparison of different batches of the product. Tests to accomplish this include:
>
> - Identity
> - Structure (e.g., optical rotation (for chiral compounds), reducing/non-reducing electrophoresis (for proteins))
> - Assay for purity
> - Impurity profile (e.g., product- and process-related impurities, residual solvents, heavy metals)
> - Assay for potency (biologic)
> - Physical characteristics (as appropriate)
> - Microbiological characteristics (as appropriate).

The cost and time savings for an exploratory IND are predominantly found in the CMC area, and to a lesser extent in the toxicology area.

Despite the encouragement of the FDA, very few exploratory INDs have been received by the FDA (Anon, 2006). Those that have been received were not for the purposes originally envisioned by the FDA (Jacobson-Kram, 2007). The majority of the INDs have been submitted for imaging studies (National Cancer Institute, 2007). It has

been stated that the exploratory IND is a new paradigm that is available to an established industry which has been slow to adopt this new paradigm. The slow adoption may be due to the fact that microdose studies may not be predictive of pharmacological dose studies and that the exploratory IND protracts the development timelines. Additionally, the exploratory IND is designed to kill drugs early that are likely to fail. Since no development team thinks that their drug is a loser, they will therefore push for full-blown development of the drug and not take advantage of the exploratory IND. Small companies may find that the exploratory IND approach may offer some benefit if the company will derive additional value in the eyes of the investment community if they can show that their compound has promise based on acceptable PK in humans or some evidence of a pharmacodynamic response. Another factor that may favor the use of a traditional approach rather than an exploratory IND is the fact that most companies will only have on compound/project that can be taken into clinical trials and that part of the decision to proceed into clinical trials was that the compound was predicted to have acceptable PK properties in humans. In this scenario, the risk–benefit to the company (as measured by overall cost and net present value based on time to market) will most likely favor the traditional IND approach.

One of the driving factors for the exploratory IND was to try to ensure that drugs that would fail due to unacceptable PK or PD could "fail fast," thereby reducing the overall cost of developing a drug.

The causes of attrition of drugs in the drug development process have garnered much interest over the past 10 years. In 2004, Kola and Landis published an article that indicated that PK/bioavailability accounted for approximately 40% of all causes of attrition in 1991, but only 10% of all causes of attrition in 2000 (Kola and Landis, 2004). This was taken by the ADME community as a validation that the high-throughput ADME screens that had been put in place in the early to mid-1990s had paid great dividends. While these screens no doubt did serve to identify drugs that had poor ADME properties, it is highly unlikely that these screens were the cause of such a large drop in the attrition rate due to PK/bioavailability issues. In 2005, Schuster *et al.* called attention to a little known article by Kennedy that was published in 1997 (Kennedy, 1997; Schuster, Laggner and Langer, 2005). Kennedy noted that the original data set generated by the Center for Medicines Research contained a large number of anti-infective drugs. When these drugs were excluded from the data set, the attrition due to PK/bioavailability issues dropped from 39% to 7%. This could be taken to indicate that the attrition rate due to PK/bioavailability issues did not change over the period 1991 to 2000, although a clear understanding of the true situation is not possible due to the lack of availability of the data sets used for these analyses to the scientific community.

Therefore, it can be seen that the exploratory IND approach will address only a relatively minor cause of drug attrition (i.e. unacceptable PK properties). Ironically, the use of an exploratory IND to evaluate the relative PK properties of several compounds will inevitably increase the attrition due to unacceptable PK properties. For example, if all INDs submitted to the FDA were exploratory INDs for the purpose of the selection of the best of three compounds, the attrition rate due to unacceptable PK properties would be 66%. This attrition rate will be decreased by the other sources of failure at other stages of development but will still be higher than the present attrition rate due to unacceptable PK properties of approximately 11%.

In September 2007, the NCI held a workshop on "Phase 0 Trials in Oncologic Drug Development" (National Cancer Institute, 2007). This meeting was held in order to raise the awareness of phase 0 studies and how they can contribute to the development

of a novel anticancer agent. Despite the push to utilize phase 0 studies in anticancer drug development, concern has been raised regarding the appropriateness of phase 0/exploratory IND studies for drugs to treat cancer. One concern is that since reliable and validated assays are not currently available even for most approved targeted cancer drugs, many experimental drugs will fail phase 0 testing. There is a chance that potentially active drugs could be misidentified as being inactive (Marchetti and Schellens, 2007). Ethical concerns regarding conducting phase 0 studies in cancer patients have been that the doses that will be administered will be subtherapeutic and the patient will derive no benefit from that study. Participation in such a study may deny them the possibility of taking part in a clinical trial in which they may receive a treatment that could provide therapeutic benefit (Hill, 2007; Marchetti and Schellens, 2007). Despite these concerns, phase 0 studies for imaging purposes in cancer patients may be appropriate.

Recently, the results of a phase 0 study conducted by the NCI were reported (Kummar *et al.*, 2007). The objective of the study was to determine a dose range at which ABT-888 inhibits PARP in tumor tissue and in peripheral blood mononuclear cells (PBMCs); and the PK of ABT-888. Patients with advanced solid tumors refractory to at least one line of therapy were eligible; patients with CLL or follicular lymphomas were also eligible if standard therapy was not currently indicated. A single oral dose of ABT-888 was administered per patient; dose escalations were planned in cohorts of three patients each (10, 25, 50, 100, and 150 mg). PBMC and tumor sampling were performed before and after drug administration for real-time PK and PD analyses. All patients underwent PBMC sampling; tumor biopsies were planned once significant inhibition of PARP activity in PBMCs was seen in one of three patients in a cohort or plasma $C_{max}$ of 210 nM was achieved in at least one patient. Tumor biopsies were performed at baseline in the week prior to drug administration and then three to six hours post drug administration. Significant inhibition of PARP activity was defined as at least 0.69 reduction on the log scale, which also satisfied statistical significance. A total of six patients have been studied so far, three each for the 10- and 25-mg cohorts. No treatment-related adverse events have been observed. Target $C_{max}$ was exceeded in the first cohort; all patients in the next cohort underwent tumor biopsies in addition to PBMC sampling. A trend toward inhibition of PARP activity in PBMCs was observed in the first cohort. Significant inhibition of PAR levels was observed in tumor biopsies from all three patients in the second cohort (92%, 99%, 100% reductions respectively, as compared with baseline). Greater than 85% reduction of PAR levels was observed in PBMCs from two of the three patients in the second cohort (one patient was not evaluable). ABT-888 is orally bioavailable and inhibits PARP activity in PBMCs and tumor cells. Target assay feasibility was established in human samples.

It is unclear whether this study was conducted under an exploratory IND or a traditional IND. It may be assumed that the latter approach was taken since the researchers did not discuss the use of stopping rules and that a maximum biologic effect was observed in the tumor biopsy. This study could be viewed as being a phase 0 study in the sense that one of the objectives was to evaluate the utility of the PD assay and that a single dose was administered (i.e. the patient would receive limited, if any, therapeutic benefit).

Microdosing studies are one example of a study that can be conducted under an exploratory IND. This type of study has garnered much interest over the past decade since regulatory agencies and CROs have proposed that such studies would enable a

rapid identification of compounds with suboptimal PK properties. In the United States, microdose studies could be conducted under an exploratory IND or under a traditional IND, although strong emphasis has been placed upon conducting such studies under an exploratory IND. In either situation, the format of the IND submission would be identical and would include information on a clinical development plan; chemistry, manufacturing, and controls (CMC) information; pharmacology, toxicology, and ADME information; and previous human experience with the investigational candidate or related compounds, if there is any.

The CHMP issued "Position Paper on Non-Clinical Safety Studies to Support Clinical Trials with a Single Microdose" which came into effect in January 2003, with Revision 1 coming into effect in June 2004 (EMEA, 2004). The Position Paper defined "common standards of the non-clinical safety studies needed to support human clinical trials of a single dose of a pharmacologically active compound using microdose techniques." The term "microdose" was defined as being "less than 1/100th of the dose calculated to yield a pharmacological effect of the test substance based on primary pharmacodynamic data obtained in vitro and in vivo (typically doses in, or below, the low microgram range) and at a maximum dose of ≤100 microgram." Examples of the types of clinical trial included "the early characterization of a substance's pharmacokinetic-/distribution properties or receptor selectivity profile using positron emission tomography (PET) imaging, accelerator mass spectrometry (AMS) or other very sensitive analytical techniques."

As can be anticipated, the calculation of the dose required to yield a pharmacological effect will be based on many assumptions including the prediction of the pharmacokinetics of the drug in humans. It is also unclear as to how a pharmacological effect in humans would be defined (i.e. a full effect or a 10% effect) and implies that a PD assay would be available to measure that effect in humans in the microdose study. It should also be noted that a microdose study can be supported using a sensitive LC-MS/MS assay and does not necessarily have to use AMS as the bioanalytical tool.

The clinical trials that were covered in this Position Paper will be exploratory in nature (pre-phase 1) and may be conducted with a single test substance or with a number of closely related pharmaceutical candidates to choose the preferred candidate or formulation for further development. The CPMP proposed that "certain deviations from the existing CPMP/ICH notes for guidance to support pre-phase 1 clinical trials may be scientifically justified." The Position Paper noted that "In the European Union, repeated dose toxicity studies in two species (one non-rodent) for a minimum of 2 weeks are required to support a single, first human dose. However, in the United States of America, single dose acute toxicity studies are in some cases considered sufficient to support a single dose human clinical trial." The Position Paper noted that under the ICH M3 guidance, safety pharmacology, single-dose toxicity studies, and repeated-dose toxicity studies would be required for a single-dose human clinical trial. The Position Paper stated that "This set of studies may be replaced by an extended single-dose toxicity study in only one mammalian species if the choice in species could be justified based on comparative in vitro metabolism data and by comparative data on in vitro primary pharmacodynamics/biological activity." The design of the single-dose toxicity is described in some detail. Pertinent points include "The extended single-dose toxicity study should include a control group, and a sufficient number of treatment groups to allow the establishment of the dose inducing a minimal toxic effect. For compounds with low toxicity a limited dose approach could be used. Allometric scaling from animal species to man and using a safety factor of 1000 should be used to set the limit dose.

If a toxic effect is observed at the limit dose, the non-toxic dose level should be established." The Position Paper refers the reader to CPMP/ICH/283/95 for allometric scaling factors. It is interesting to note that the reference is to ICH Topic Q3C(R3) Impurities: Residual Solvents (ICH, 1997b). It should be noted that the allometric scaling factors and the assumptions on body weights differ between the ICHQ3C(R3) guidance and the FDA's guidance "Estimating the Maximum Safe Starting Dose in Initial Clinical Trials for Therapeutics in Adult Healthy Volunteers" that was issued in 2005. The FDA's guidance focuses on the use of toxicology data to establish the maximum safe starting dose. The guidance does mention metabolism and pharmacokinetic data in the following manner:

> Although the process outlined in this guidance uses administered doses, observed toxicities, and an algorithmic approach to calculate the MRSD (Maximum Recommended Starting Dose), an alternative approach could be proposed that places primary emphasis on animal pharmacokinetics and modeling rather than dose. In a limited number of cases, animal pharmacokinetic data can be useful in determining initial clinical doses. However, in the majority of investigational new drug applications (INDs), animal data are not available in sufficient detail to construct a scientifically valid, pharmacokinetic model whose aim is to accurately project an MRSD. Toxicity should be avoided at the initial clinical dose. However, doses should be chosen that allow reasonably rapid attainment of the phase 1 trial objectives (e.g., assessment of the therapeutic's tolerability, pharmacodynamic or pharmacokinetic profile). All of the relevant preclinical data, including information on the pharmacologically active dose, the full toxicologic profile of the compound, and the pharmacokinetics (absorption, distribution, metabolism, and excretion) of the therapeutic, should be considered when determining the MRSD. Starting with doses lower than the MRSD is always an option and can be particularly appropriate to meet some clinical trial objectives.

Interestingly, one of the references cited for the use of PK data to establish the starting dose for a clinical trial was published by FDA employees (Mahmood, Green and Fisher, 2003).

In 1Q 2006, the CHMP adopted a concept paper, "Concept Paper on the Development of a CHMP Guideline on the Non-Clinical Requirements to Support Early Phase I Clinical Trials with Pharmaceutical Compounds" (EMEA, 2006b). The Concept Paper states that:

> The requirements for non-clinical safety studies to support the conduct of human clinical trials for pharmaceuticals were largely addressed by the International Conference on Harmonization of Technical Requirements for the Registration of Pharmaceuticals for Human Use (ICH) Topic M3: Note for Guidance on Non-clinical Safety Studies for the Conduct of Human Clinical Trials for Pharmaceuticals (CPMP/ICH/286/95). This guidance aimed to facilitate the timely conduct of clinical trials and reduce the unnecessary use of animals and other resources. This was designed to promote safe and ethical development and availability of new pharmaceuticals. The development of a pharmaceutical is a stepwise process involving an evaluation of both the animal and human safety information. The overall goals of the non-clinical safety evaluation include a characterization of toxic effects with respect to target organs, dose dependence, relationship to exposure, and potential reversibility. This information is important for the estimation of an initial safe starting dose for the human trials and for the identification of parameters for clinical monitoring for potential adverse effects. The non-clinical safety studies at any stage should be adequate to characterise potential adverse effects under the conditions of the proposed clinical trial.

> It is recognised that significant advances in harmonization of the timing of non-clinical safety studies for the conduct of human clinical trials for pharmaceuticals have already been achieved. However, differences remain in a few areas, including toxicity studies to support first entry into man. In 2003, the CPMP released a position paper on non-clinical studies to support clinical trials with a single microdose (CPMP/SWP/2599/02). More recently, the FDA have released guidance on conducting clinical trials under an exploratory IND.

The Concept Paper further states that:

> The CHMP Safety Working Party (SWP) have discussed a revised toxicology package designed to support early clinical investigations in man. The objectives of these clinical studies would differ from the more classical phase I approach, in that they would investigate low doses and would not be designed to investigate tolerability. Instead, the trial would generate pharmacokinetic (PK)-related or pharmacodynamic (PD)-related data, possibly based on a sensitive biomarker of efficacy. These data could then be used for early decision making on whether to progress a particular compound or work on a potential therapeutic target. The clinical studies would require the administration of one or two doses that would achieve an exposure in the predicted pharmacological range, but which would be below the tolerability range. The initial dose administered to man would be very low, and the dose would be slowly escalated (using real time pharmacokinetic monitoring) until a predefined maximum exposure was achieved. Using this approach, a good understanding of dose exposure relationship in man could be obtained. In order to maintain international harmonization, the revised toxicology package will follow the general principles laid down by the ICH M3 guideline. However, the dose levels used in the nonclinical studies could be lower than usual, as they are only required to support the low intended clinical doses. Human safety will not be compromised by the revised package.

The current schedule calls for the draft guideline detailing what nonclinical data are required to be included in a clinical trials application for an early phase I study in humans to be issued in 1Q 2008. This guideline should allow for a flexibility of approaches, including those outlined in the EU microdose guideline or the FDA exploratory IND guideline. How this guideline will relate to the recently released EMEA guideline or the planned ICH guidance is presently unknown.

It should be recognized that as of November 2007, despite what is reported by several journals and CROs, there are no guidances or guidelines in the United States or EU that pertain solely to microdose studies. In the EU, there is a Position Paper and a Concept Paper but no guideline. In the United States, there is a guidance on exploratory INDs in which microdose studies are mentioned but are not the main topic of this guidance.

The utility of microdosing to predict the pharmacokinetics of a drug when that drug is administered at a pharmacologic dose has been much debated. Lappin *et al.* conducted a study (the CREAM study) in which the PK of five drugs (warfarin, ZK253, diazepam, midazolam, and erythromycin) were evaluated at a microdose and pharmacological dose level (Lappin *et al.*, 2006). This study found that the PK after a microdose of warfarin was different to that obtained at a pharmacological dose, possibly due to high-affinity, low-capacity tissue binding of warfarin. The oral microdose of erythromycin failed to provide detectable plasma levels as a result of possible acid lability in the stomach. The absolute bioavailability for the other three compounds showed good concordance between microdose and pharmacologic dose levels. These researchers concluded that "Overall, when used appropriately, microdosing offers the potential

to aid early drug candidate selection." In 2007, the American College of Clinical Pharmacology published a position statement on the use of microdosing in the drug development process (Bertino, Greenberg and Reed, 2007a, b). This publication stated that:

> At present, it would appear that studies using microdosing methodology should not be relied on as the primary or sole approach to screen new drug candidates, as the potential exists with current methodologies to possibly reject important new drugs while possibly accepting drugs that could result in significant safety issues. Until more information is available and has undergone appropriate scrutiny, it would appear that a microdosing strategy could complement standard animal-to-human allometric scaling, refining current phase I study designs. Microdosing methodology appears to be one of the many new viable "tools" in the drug development "toolbox." The exact role and impact that microdosing methodologies will have on new drug development is yet to be fully realized but will surely be continuously refined with continued experience and, above all, data sharing through the publication of studies employing this novel methodologic approach.

The conclusions of the ACCP position paper have been challenged by Rowland (2007). Other researchers have stated that "microdose studies will have limited impact on the overall efficiency of drug development" (Boyd and Lalonde, 2007).

It has been reported that new paradigms are being considered in the maintenance of ICHM3R "Timing of preclinical studies in relation to clinical trials" (Jacobson-Kram, 2007). "These new paradigms include:

- Microdose up to 5 doses each not to exceed 1/100th the NOAEL determined in the toxicology study or 1/100th the anticipated pharmacodynamically active dose or a total dose of 100 μg whichever is lower.
- Microdose up to 5 doses of 100 μg each not to exceed 1/100th the NOAEL determined in the toxicology study or 1/100th the anticipated pharmacodynamically active dose and with a washout of six half-lives between doses. Repeat-dose exploratory studies up to 14 days in the therapeutic range but not to an MTD. Supported by safety studies using large multiples of clinical exposure but not based on toxicity."

It will be interesting to see how the revision to ICHM3 progresses (ICH, 2006c).

### 5.4 Ethnic Factors

The ICHE5 (R1) guidance addresses "Ethnic Factors in the Acceptability of Foreign Clinical Data" (ICH, 1998). The basis for this guidance is described thus:

> To assess a medicine's sensitivity to ethnic factors it is important that there be knowledge of its pharmacokinetic and pharmacodynamic properties and the translation of those properties to clinical effectiveness and safety. Some properties of a medicine (chemical class, metabolic pathway, pharmacologic class) make it more or less likely to be affected by ethnic factors. Characterization of a medicine as "ethnically insensitive", that is unlikely to behave differently in different populations, would usually make it easier to extrapolate data from one region to another and need less bridging data.

The guidance provides details of the characteristics of an "ethnically sensitive" and an "ethnically insensitive" drug. It states that "the following properties of a compound make it less likely to be sensitive to ethnic factors:

- Linear pharmacokinetics (PK)
- A flat pharmacodynamic (PD) (effect-concentration) curve for both efficacy and safety in the range of the recommended dosage and dose regimen (this may mean that the medicine is well-tolerated)
- A wide therapeutic dose range (again, possibly an indicator of good tolerability)
- Minimal metabolism or metabolism distributed among multiple pathways
- High bioavailability, thus less susceptibility to dietary absorption effects
- Low potential for protein binding
- Little potential for drug-drug, drug-diet and drug-disease interactions
- Non-systemic mode of action
- Little potential for inappropriate use

The following properties of a compound make it more likely to be sensitive to ethnic factors:

- Non-linear pharmacokinetics
- A steep pharmacodynamic curve for both efficacy and safety (a small change in dose results in a large change in effect) in the range of the recommended dosage and dose regimen
- A narrow therapeutic dose range
- Highly metabolized, especially through a single pathway, thereby increasing the potential for drug-drug interaction
- Metabolism by enzymes known to show genetic polymorphism
- Administration as a prodrug, with the potential for ethnically variable enzymatic conversion
- High inter-subject variation in bioavailability
- Low bioavailability, thus more susceptible to dietary absorption effects
- High likelihood of use in a setting of multiple co-medications
- High likelihood for inappropriate use, e.g., analgesics and tranquilizers."

Based on the given definitions, it would appear that most drugs would be viewed as being ethnically sensitive. The concept that "metabolism distributed among multiple pathways" is an interesting concept since it would imply that such drugs would have a low potential for DDIs based on the existence of the "safety valve" of alternative routes of metabolism. This is a concept that was proposed by Vickers *et al.* in 1999 but, to date, this concept has not been extensively validated by *in vivo* data (Vickers *et al.*, 1999). A Q&A document for this guidance was issued in 2006 (ICH, 2006e).

### 5.5 Pharmacogenomics

In recent years, the genotyping of DMEs in clinical trials for the purpose of understanding interindividual variability in pharmacokinetics and effect (either efficacy or side effects) has become more common. This information has been included in product labels and has been the subject of several regulatory guidances. Additionally, several genotyping tests have been developed and commercialized. The Amplichip CYP450 test is the first FDA-cleared test for analysis of CYP2D6 and CYP2C19 ( Roche, 2008). The test identifies a patient's genotype and, based on this analysis, provides their predicted

phenotype—either poor, intermediate, extensive, or ultrarapid metabolizer. These two genes code for enzymes that metabolize many antidepressants, antipsychotics, and ADHD drugs, as well as other medications. Patients seeking therapy for depression and other mood disorders may not respond to or may have adverse reactions to many antidepressants, antipsychotics, and ADHD drugs due to their ability to metabolize these drugs. Physicians have traditionally used a "start low, go slow" dosing approach with their patients to mitigate efficacy and side-effect issues. This test can help physicians adjust dosing and select drugs by predicting a phenotype based on a genotype so that patient treatment can be individualized to get the best therapeutic results possible.

The Invader® UGT1A1 Molecular Assay is cleared for use to identify patients who may be at increased risk of adverse reaction to the chemotherapy irinotecan (Camptosar®, Pfizer) (Third Wave Technologies, 2008). Irinotecan is indicated for use in the treatment of metastatic colorectal cancer. While irinotecan is effective, irinotecan-associated toxicity can result in both neutropenia (reduced white blood cell count) and severe diarrhea. The UGT1A1*28 allele that the Invader UGT1A1 Molecular Assay detects, homozygous in approximately 10% of the North American population, has been shown to be an effective genetic marker for predicting irinotecan-induced toxicity. In July 2005, the FDA revised the safety labeling for irinotecan, recommending that treatment be altered for individuals who are homozygous for the UGT1A1*28 allele. One month later, the FDA approved the genotyping test for this allele.

In August, FDA approved updated labeling for Coumadin (warfarin), explaining that people with variations of the genes CYP2C9 and VKORC1 may respond differently to the drug (US FDA, 2007g). Manufacturers of generic warfarin are adding similar information to their products' labeling. The next month, the FDA approved the Nanosphere Verigene Metabolism Nucleic Acids Test detects particular variations in two genes, CYP2C9 and VKORC1, which are involved in the metabolism and mechanism of action of warfarin, respectively (Nanosphere Inc., 2008). The Nanosphere test is not intended to be a stand-alone tool to determine optimum drug dosage, but should be used along with clinical evaluation and other tools, including INR, to determine the best treatment for patients. The FDA cleared the test based on results of a study conducted by the manufacturer of hundreds of DNA samples as well as on a broad range of published literature. In a three-site study, the test was accurate in all cases where the test yielded a result; 8% of the tests could not identify which genetic variants were present.

Although these tests will be primarily used for therapeutic decisions, it can be seen that such tests may have a role in the selection of patients for inclusion in a particular clinical trial. Clinical Research Organizations (CROs) may offer populations of volunteers that have been prescreened for certain genetic polymorphisms.

Prior to the creation of regulatory guidances on pharmacogenomics, a report of an FDA-PWG-PhRMA-DruSafe Workshop on Pharmacogenetics and Pharmacogenomics in Drug Development and Regulatory Decision Making was published in 2003 (Lesko *et al.*, 2003). This workshop laid the foundations for the FDA guidances that followed. The dialog between the FDA and industry has continued through subsequent workshops (Frueh *et al.*, 2006).

Guidances that pertain to genotyping include:

- ICH Topic E15. Establish definitions for genomic biomarkers, pharmacogenomics, pharmacogenetics, genomic data and sample coding categories (Step 5; approved in EU) (EMEA, 2007g)

- Reflection Paper on the Use of Pharmacogenetics in the Pharmacokinetic Evaluation of Medicinal Products (EMEA/128517/2006) (EMEA, 2007i)
- Pharmacogenomic Data Submissions (FDA; 2005) (US FDA, 2005f)
- Pharmacogenomic Data Submissions—Companion Guidance (FDA; Draft 2007) (US FDA, 2007m).

These documents make several general points on drug metabolism and genotyping. ICHE15 states that "The definition for a genomic biomarker does not include the measurement and characterization of proteins or low molecular weight metabolites."

The EMEA reflection paper was adopted by the CHMP in May 2007. The purpose of this reflection paper is to "target the place of PG in the clinical PK evaluation of medicinal products during drug development and use." Like most guidelines or related documents, the reflection paper notes that it should be read in conjunction with other guidelines. Among the topics addressed by this reflection paper are:

- in which situations should the effect of PG on PK of new or existing medicinal products be studied;
- at what stage in the clinical development program should PG/PK studies be performed;
- study design and methodology;
- evaluation of the clinical consequences of genetic differences in drug substance exposure;
- special considerations related to drug–drug interactions and impaired or immature organ functions.

The reflection paper addresses the situations in which the effect of PG on PK should be studied. It states:

> Studies of the effect of PG on PK are required for PK evaluation of a new chemical entity if the genetic variation is likely to translate into important differences in the systemic and/or local exposure to this substance or its active or toxic metabolites, thereby potentially affecting safety and efficacy of the treatment. PG variants may impact on Absorption, Distribution, Metabolism and Excretion of the compound.
>
> Furthermore, combined PG/PK studies that may contribute to the identification of novel polymorphic loci are encouraged if the compound exhibits important inter-individual PK variability, likely to affect clinical efficacy and/or safety. Technology has advanced to the point that analysis of genetic factors affecting safety and efficacy of medicines is now fast and reliable. If there is clinically important variability in PK and genetic causes for this may not be excluded, it is advisable to carry out genetic analysis of loci likely to be responsible for this variation. In all cases where unexplained PK variation has been identified, samples for pharmacogenetic purpose should be collected. This would allow a critical evaluation of the clinical and, if population PK is used, PK consequences of this polymorphism at a later stage.
>
> Studies of PG differences in the activity and expression of transport proteins involved in drug distribution (efflux or influx) to target organs such as the central nervous system, possibly explaining adverse events or lack of therapeutic effect, are encouraged if there are indications of clinically important differences in these respects and if a relevant polymorphism can be studied using cohorts with sufficient power to reach conclusions.

> Although some polymorphisms are extremely rare, the applicant should always consider studying the PK and clinical consequences of all potentially clinically relevant polymorphisms. If this is not feasible, the applicant should justify the lack of data and discuss the possible safety (or efficacy) consequences based on prior knowledge on the effect of the polymorphisms on protein activity. Lack of data should be reflected in the Summary of Product Characteristics (SPC) if considered clinically relevant.

The timing of PG/PK studies is also addressed thus:

> In general, the importance of PG for the PK of a drug substance may be indicated by *in vitro* data where the enzymes involved in drug metabolism as well as formation and metabolism of pharmacologically active metabolites have been identified. Preclinical animal studies should be interpreted with great caution, as there are usually marked species differences. If human *in vitro* data suggest major involvement of a protein known to be subject to functionally important genetic polymorphism, inclusion of genotyping directed to the candidate gene is warranted in early phase I studies. When the involvement of the polymorphic gene has been verified, *in vivo* studies of the effects of specific polymorphisms on the PK of the pharmacologically active compounds likely to contribute to clinical efficacy and/or safety are recommended.

It should be noted that the reflection paper does not give insight into exactly how the importance of a polymorphic DME is determined.

Polymorphisms in drug transporters are also addressed:

> Involvement of transporters may also be indicated by *in vitro* data but presently the knowledge in this field is not mature enough for early genotyping to be required only on the basis of *in vitro* data. However, if the *in vitro* data together with other data such as ADME or renal excretion data indicate that a polymorphic transporter has a major role in the PK of a drug, genotyping in the subsequent PK studies and clinical studies is encouraged.

The collection and analysis of samples is also addressed: "It is recommended that samples from the early phase I studies are stored to allow retrospective analysis, when more experience has been gained or new proteins have been shown to play a major role in the PK of the active substances." It is not always possible to obtain informed consent from the patients/subjects and the cost of storage until the samples are analyzed can become burdensome. When such analyses are performed at a later date, consideration should be given to providing the results of the genotyping to the patient/subject.

The reflection paper further states that:

> When the genotype is predicted or known to markedly affect the PK of pharmacologically active compounds contributing to *in vivo* efficacy and/or safety of a medicinal product, genotyping is encouraged in as many of the phase I, II and III clinical studies as possible to increase the amount of data that will support the recommendations for use in the genetic subpopulation(s). Dose-response studies or other clinical studies covering the exposure of active substances obtained can be used to support safety and efficacy in a specific genetic subpopulation.

The effect of genotype in certain special populations is also recognized:

> The enzymes and transport proteins involved in the PK of a drug substance may be quantitatively and qualitatively different in paediatric patients than in adults as a consequence

> of developmental changes in gene expression. The most marked differences are expected in newborn infants, infants and toddlers (0–2 year-old children). In the very elderly patients PG related differences in metabolism may occur and have an impact due to impaired functionality of system/organs; limited knowledge is available at present.

The role of PG in interpreting DDIs is another topic addressed by this reflection paper. The paper states:

> Genotyping of the population included in an interaction study is recommended when PG is expected to affect the PK of any of the active substances. Depending on the question investigated, directed inclusion or exclusion of specific genotypes may be useful.
>
> If well known major elimination pathways are absent in a subpopulation, the consequences of an inhibition of parallel pathways should be considered. The magnitude of such an interaction can be difficult to determine in the absence of an in vivo interaction study. However, a "worst case" scenario can usually be estimated. If the safety consequences of the interaction are predicted to give an unacceptable risk-benefit, or if sufficient safety data is lacking at predicted exposures, an interaction study in the subpopulation may be needed to reach satisfactory treatment recommendations.
>
> The effect of active substances, that are enzyme or protein inhibitors or inducers, may also be different if the contribution of an enzyme is absent in a subpopulation. The consequences of inhibition of a protein in a population with a lower but not abolished activity of that protein should also be considered. In case of induction, the net result will depend on the degree of induction -if any- of the parallel pathway. This should be considered especially when a dose recommendation for a certain drug-combination is studied and evaluated.
>
> An increased systemic exposure in genetic subpopulations may lead to more pronounced effects of the investigated drug on other drugs. This should be considered and if necessary reflected in the SPC.

The contribution of genetic polymorphisms to inter-individual variability in drug response was again emphasized in a recent reflection paper on the use of genomics in cardiovascular clinical intervention trials (EMEA, 2007j). This reflection paper provided an example of the role of genetic polymorphisms for cardiovascular drugs, stating that "Genes coding for enzymes involved in the metabolism of warfarin and clopidogrel, and those coding for platelet components involved in the antiplatelet action of drugs have been extensively studied in relationship to pharmacokinetic issues, and anti-clotting effect."

The FDA has a deep interest in pharmacogenomics and toxicogenomics which is reflected on the amount of information available on these topics on its web site (US FDA, 2008e). Toxicogenomics, which is the study of drug-related safety/toxicity and pathology at the gene expression level in preclinical studies, is also of great interest to the FDA. At the present, there are no guidances specific to toxicogenomics, although the expectation is that the data from toxicogenomic studies would be submitted in a similar manner to the data from pharmacogenomic studies. Since DMEs or transporters are not a common feature of toxicogenomic studies, there will be no further discussion of toxicogenomics in this chapter.

The FDA issued a guidance on "Pharmacogenomic Data Submissions" in 2005 (US FDA, 2005f). This guidance notes that "some pharmacogenetic tests—primarily those related to drug metabolism—have well-accepted mechanistic and clinical significance and are currently being integrated into drug development decision making and clinical

practice." For the purposes of this guidance, a pharmacogenomic test result may be considered a valid biomarker if "(1) it is measured in an analytical test system with well-established performance characteristics and (2) there is an established scientific framework or body of evidence that elucidates the physiologic, pharmacologic, toxicologic, or clinical significance of the test results."

For example, the consequences for drug metabolism of genetic variation in the human enzymes CYP2D6 and thiopurine methyltransferase are well understood in the scientific community and are reflected in certain approved drug labels. The results of genetic tests that distinguish allelic variants of these enzymes are considered to be well established and, therefore, valid biomarkers.

The fact that most PG tests currently involve DMEs is illustrated by the examples given in this guidance. The guidance states:

> If a pharmacogenomic test shows promise for enhancing the dose selection, safety, or effectiveness of a drug, a sponsor may wish to fully integrate pharmacogenomic data into the drug development program. … The pharmacogenomic data may be intended to be included in the drug labeling in an informational manner. For example, such data might be used to describe the potential for dose adjustment by drug metabolism genotype (e.g., CYP2D6*5) or to mention the possibility of a side effect of greater severity or frequency in individuals of a certain genotype or gene expression profile. In such cases, the pharmacogenomic test result would be considered a known valid biomarker. However, an FDA-approved pharmacogenomic test may not be available or required to be available, or a commercial pharmacogenomic test may not be widely available. Given this level of complexity, at the current time, sponsors should consult the relevant FDA review division for advice on how to proceed in a specific case.

The guidance provides a definition of a pharmacogenetic test: "An assay intended to study interindividual variations in DNA sequence related to drug absorption and disposition (pharmacokinetics) or drug action (pharmacodynamics), including polymorphic variation in the genes that encode the functions of transporters, metabolizing enzymes, receptors, and other proteins." It is interesting to note that drug transporters are only mentioned in the definition of a pharmacogenetics test, but not in the main body of the guidance.

The recent draft guidance, "Pharmacogenomic Data Submissions—Companion Guidance" (US FDA, 2007m) states that:

> This guidance is intended to be used as a companion to the guidance *Pharmacogenomic Data Submissions* (March 2005). It reflects experience gained since the issuance of that guidance with voluntary genomic data submissions as well as with review by the FDA of numerous protocols and data submitted under investigational new drug (IND) applications, new drug applications (NDAs), and biologics license applications (BLAs). The recommendations are intended to facilitate scientific progress in the field of pharmacogenomics and to facilitate the use of pharmacogenomic data in drug development. The FDA believes that the recommendations made in this companion guidance, together with the recommendations in the March 2005 guidance, will benefit sponsors considering the submission of either voluntary genomic data submissions or marketing submissions containing genomics data. As technology changes and more experience is gained, these recommendations may be updated.

The draft guidance discusses genotyping methods:

> Genetic differences among individuals occur in a variety of forms, from alterations in chromosomal arrangement or copy number to single base-pair changes. Much of the

> genetic variation currently used in pharmacogenetics occurs at the level of individual genes (e.g., drug metabolizing enzymes) on a scale ranging from single base-pair changes to entire gene duplications or deletions. Examining genomic DNA is often the most reliable and practical method for characterizing genetic variation, although methods based on protein or mRNA expression levels can be preferable in some situations, such as when determining treatment-sensitivity of cancer or viral infection. Many methods are currently available for characterizing DNA variations, and new methods are rapidly being developed.

The reporting of the data in a genotyping report is also addressed: "We recommend that the following information be included in the genotyping report, regardless of the genomics submission type (see the *Pharmacogenomic Data Submissions* guidance for regulatory requirements):

- Description of assay platform or methodology
- Samples studied, including demographics and sample size justification for genotype/clinical phenotype correlation and adequate coverage for ethnic/racial groups; include expected allele frequency in different populations
- Alleles measured and correlation with metabolic status designation
- For metabolizing enzymes, how EM (extensive metabolizer), PM (poor metabolizer), IM (intermediate metabolizer), or UM (ultra rapid metabolizer) are determined
- Sample test report
- For new genes, correlation between gene variant and encoded protein activity.
- Whether the assay was performed in a CLIA-certified lab or research lab."

The reporting of genomic data in clinical reports is described:

> There are many possible sources of data for genomic data submissions. Genomic data from clinical studies may result from microarray expression profiling experiments, genotyping or single-nucleotide polymorphism (SNP) experiments, or from other evolving analytical methodologies pertaining to drug dosing or metabolism, safety assessments, or efficacy evaluations. Genomic data may also be reported from studies where other data are also reported, such as with efficacy or safety data from clinical or nonclinical studies. However, these data can be reviewed only if the content of the clinical data report included in the submission contains *sufficient detail* regarding the sample selection.

The draft companion guidance does not appear to provide guidance that is markedly different from that previously provided but simply serves to emphasize the FDA's interest and commitment to the use of genomic data in the evaluation of the safety and efficacy of drugs.

The FDA has authored several papers on the implementation of the guidance on pharmacogenomics data submission, in which its members have summarized their experiences to date (Frueh *et al.*, 2006; Goodsaid and Frueh, 2007).

Although pharmacogenetics is important in the disposition of certain drugs, it should be noted that a large degree of variability can arise due to other factors other than genotype, such as induction and inhibition of the enzyme systems or transporters. Additionally, there may be ethical concerns regarding genotyping for polymorphisms in CYP isoforms since some polymorphisms have been associated with increased risk of certain diseases, including cancer (Agundez, 2004; Danko and Chaschin, 2005; Rodriguez-Antona and Ingelman-Sundberg, 2006). PhRMA has created a white paper

describing the current best practices with respect to genotyping (Williams *et al.*, 2008).

### 5.6 Drug–Drug Interactions

The topic of unexpected drug–drug interactions first came to the forefront in the 1970s and 1980s. The history of the withdrawal of drugs from the market due to drug–drug interactions and the development of regulatory guidances has been summarized by Huang *et al.* (2008). Their summary is presented in abbreviated form in Table 3.

The example of Posicor (mibefradil) provides some insight into the process of the withdrawal of a drug from the market to the potential (and real) safety consequences of DDIs. Posicor was withdrawn from the market in June 1998, when it was learned that more than 25 drugs were found to be dangerous when used with Posicor. When Posicor entered the market in August of 1997, its enzyme-inhibiting properties were described in the labeling. The labeling specifically listed three drugs (astemizole, cisapride, and terfenadine) that could be expected to accumulate to dangerous levels if Posicor was coadministered (US FDA, 1998a). In less than a year, it became apparent that the liabilities caused by the inhibition of the metabolism of concomitant medications by Posicor meant that it would be extremely difficult, if not impossible, to safely prescribe Posicor.

It is clear that the FDA has a keen interest in DDIs and they have created a web page that contains a great deal of useful information on this topic (US FDA, 2007f). The vast majority of the information on DDIs relates to the interaction between two chemical drugs (NCEs). Comparatively little information is available regarding the potential for interaction between an NCE and a biologic, between two biologics, and between herbs and NCEs or biologics. The identification of herbs that interact with drugs can be of great importance given the extensive use of herbal medication by patients and the safety concerns that have arisen regarding the use of certain herbal supplements (Seeff, 2007; Zhou *et al.*, 2007a). The amount of literature that is available regarding the interaction between herbs and drugs is growing greatly (Hu *et al.*, 2005; Skalli, Zaid and Soulaymani, 2007; Zhou *et al.*, 2007b).

**TABLE 3 Timeline of events related to the development of FDA guidances for the evaluation of the potential for DDI studies.**

| Year | Event |
|---|---|
| 1970–1980s | Reports of unanticipated DDIS |
| 1990 | Terfenadine case study |
| 1997 | First *in vitro* DDI guidance posted by the FDA |
| 1998 | Withdrawal of terfenadine, mibefradil and bromfenac from US market |
| 1999 | First *in vivo* DDI guidance posted by the FDA |
| 1999 | Withdrawal of astemizole and grepafloxacin from US market |
| 2000 | Withdrawal of alosteron, cisapride and troglitazone from US market |
| 2001 | Withdrawal of cerivastatin and rapacuronium from US market |
| 2003 | RhRMA White white Paper paper on the conduct of *in vitro* and *in vivo* DDI studies |
| 2006 | FDA internet site devoted to DDIs launched |
| 2006 | Draft guidance on DDI studies issued by the FDA |

There are several guidances on the topic of drug–drug interactions, one of which was issued by the FDA in 1997, "Drug Metabolism/Drug Interaction Studies in the Drug Development Process: Studies In Vitro" (US FDA, 1997a). The initial elements of the guidance provide some framework for the later points by stating that:

> This FDA guidance to industry provides suggestions on current approaches to studies in vitro of drug metabolism and interactions. The guidance is intended to encourage routine, thorough evaluation of metabolism and interactions in vitro whenever feasible and appropriate. The FDA also recognizes that clinical observations can address some of the same issues identified in this document as being susceptible to in vitro study. The suggested approaches delineated in this document, however, are efficient and inexpensive considering the breadth of information they can provide, and often can reduce or eliminate the need for further clinical investigations. This particular guidance is directed toward a broad class of drugs: molecules with a molecular weight below 10 kiloDaltons.

It is unclear how the cutoff of 10 kDa was derived or what types of molecules that the FDA was imagining would be of the range of 2–10 kDa. It may be an attempt to indicate that such *in vitro* studies are not required for biotechnology-derived products such as monoclonal antibodies, oligonucleotide drugs, or aptamers.

The level of interest of the FDA in DDIs is clear in this guidance. The guidance states:

> Review scientists at the FDA have long been interested in the impact of drug metabolism and drug–drug interactions on drug safety and efficacy. As a result, discussion of this topic also is contained in other FDA guidance documents, including *General Considerations for the Clinical Evaluation of Drugs* (FDA 77-3040), *Guideline for Studying Drugs Likely to be Used in the Elderly* (11/89), and *Guideline for the Study and Evaluation of Gender Differences in the Clinical Evaluation of Drugs* (58 FR 39406, July 22, 1993).

The guidance notes that DDIs may be two-sided, that is, the drug could be the perpetrator or the victim. The guidance states that "Even for drugs that are not substantially metabolized, the potential effect of that drug on the metabolism of concomitant drugs could be important." The effect of genetic polymorphisms on exposure and the increase in drug levels as a consequence of a DDI is also discussed:

> Large differences in blood levels can occur because of individual differences in metabolism. Some drugs, such as tricyclic antidepressants, exhibit order of magnitude differences in blood concentrations depending on the enzyme status of patients. Drug-drug interactions can have similarly large effects when one drug inhibits the metabolism of another. For example, ketoconazole greatly increases concentrations of parent terfenadine, leading to QT prolongation and *torsades de pointes*.

The guidance notes that *in vitro* studies should form only one part of the evaluation of the potential for a DDI:

> The studies in vitro described in this document are one set of approaches to developing information about drug metabolism and drug-drug interactions. Mechanistic and empirical clinical study approaches are available as well to provide further information. As always, a carefully designed mix of approaches is likely to yield optimal results in the shortest time and at the least cost.

The timing of DDI studies is also addressed:

> Metabolic effects and drug-drug interactions should be considered as early as possible as well as later in the drug development process. Appropriately designed pharmacokinetic/phase 1 studies could provide important information about drug metabolism, relevant metabolites, and actual or potential drug interactions. Blood level data obtained during phase 2 and 3 clinical trials, for example, via a pharmacokinetic screen, also could reveal interactions or marked inter-individual differences. Because clinical trial protocols sometimes limit concomitant drug use, some later studies may not be optimally informative about possible drug interactions. Decreasing exclusions of concomitant drug treatment and measurement of blood levels before and after treatment with a test drug (interaction screen), as well as testing drug blood levels more frequently, could make later phase clinical studies more useful. All of these studies could be more informative if significant metabolites and prodrugs could be identified and their pharmacological properties described.

The term "significant metabolites" is now outdated and does not fit with the current terminology used in the final guidance, "Drug Safety Testing of Metabolites" (US FDA, 2008j). It is also unclear exactly what is meant by the phrase "prodrugs could be metabolized." It could be speculated that this refers to the observation that a sponsor may have initially believed that the parent drug was active (i.e. terfenadine) but later discovered that a metabolite was actually responsible for the majority of the pharmacological activity (i.e. fexofenadine).

The guidance gives an overview of the techniques and approaches for studies *in vitro* of drug metabolism and drug interactions. The guidance states that "The goals in evaluating in vitro drug metabolism are:

(1) to identify all of the major metabolic pathways that affect the test drug and its metabolites, including the specific enzymes responsible for elimination and the intermediates formed; and
(2) to explore and anticipate the effects of the test drug on the metabolism of other drugs and the effects of other drugs on its metabolism."

The guidance makes the interesting point that:

> Knowledge that a particular drug is not a substrate for certain metabolic pathways is helpful. For example, if it is learned early in drug development that a molecule is not a substrate for CYP450 3A4 or that this pathway represents only a minor contribution to overall metabolism, then concern is lessened or eliminated for possible inhibition of 3A4 metabolism by drugs such as ketoconazole and erythromycin or possible induction of metabolism by drugs such as rifampin and anticonvulsants. Studies in vitro also could indicate whether a drug itself is or is not an inhibitor of common metabolic pathways. The potential for a drug inhibiting the metabolism of other drugs is almost always present for drugs metabolized by the same pathway, but can also be present for entirely separate pathways, including the principal metabolic route for a compound. This potential was first appreciated for quinidine, which is a substrate for metabolism by CYP450 3A4 and is also a very potent inhibitor of CYP450 2D6.

Study designs are also addressed in this guidance. The guidance gives some very basic information regarding the cytochrome P450 system. This information is not presented here due to its basic nature and the fact that the information is almost 10 years old and may no longer represent state of the art information. The guidance addresses

the identification of the CYP isoforms involved in the metabolism of the drug as well as the ability of the drug to inhibit the activity of select CYP isoforms.

Although the main focus of the guidance is CYP mediated metabolism, the guidance does note that other drug metabolizing enzymes may play a role in the metabolism of a particular drug as may cytosolic enzymes.

The guidance acknowledges that the principal site of drug metabolism but also recognizes that "For particular drugs, however, other tissues may predominate (e.g., the kidney or gastrointestinal mucosa)." Furthermore, the guidance states: "Because most drugs are given orally, interest has been increasing in the effect of gastrointestinal mucosal enzymes on drug entry to the systemic circulation. Drugs susceptible to metabolism via CYP450 3A4 may exhibit low and/or variable bioavailability. Thus, determining the susceptibility of a drug to metabolism by CYP450 3A4 may be important not only in identifying routes of elimination but also in predicting the likelihood of significant first-pass metabolism."

The guidance notes that:

> When a difference arises between findings in vitro and in vivo, the results in vivo should always take precedence over studies in vitro. In many cases, however, studies in vitro, which are inexpensive and readily carried out, will serve as an adequate screening mechanism that can rule out the importance of a metabolic pathway and make in vivo testing unnecessary.

For example:

> If investigations in vitro suggest that the answer to the question "Does CYP450 2D6 metabolize this drug?" is "no," clinical studies to identify the impact of the slow metabolizer phenotype or to study the effect of CYP450 2D6 inhibitors will not be needed. Because studies in vitro, however, cannot adequately define the importance of a metabolic pathway, if the in vitro study answer is "yes," additional clinical studies will be important to answer whether CYP450 2D6 is clinically important to the elimination of the drug.

The guidance makes the important point that "although a drug may be extensively metabolized in vitro, a mass balance study in vivo may demonstrate that metabolism is less important than urinary or biliary excretion."

In 1999, the FDA issued the guidance, "In Vivo Drug Metabolism/Drug Interaction Studies—Study Design, Data Analysis, and Recommendations for Dosing and Labeling" (US FDA, 1999a).

In a manner similar to other guidances, the rationale for the conduct of DDI studies is laid out:

> Examples of substantially changed exposure associated with administration of another drug include (1) increased levels of terfenadine, cisapride, or astemizole with ketoconazole or erythromycin (inhibition of CYP3A4); (2) increased levels of simvastatin andits acid metabolite with mibefradil or itraconazole (inhibition of CYP3A4); (3) increased levels of desipramine with fluoxetine, paroxetine, or quinidine (inhibition of CYP2D6); and (4) decreased carbamazepine levels with rifampin (induction of CYP3A4). These large changes in exposure can alter the safety and efficacy profile of a drug and/or its active metabolites in important ways. This is most obvious and expected for a drug with a narrow therapeutic range (NTR), but is also possible for non-NTR drugs as well (e.g., HMG CoA reductase inhibitors). Depending on the extent and consequence of the interaction, the fact that a drug's metabolism can be significantly inhibited by other drugs and that the drug

itself can inhibit the metabolism of other drugs can require important changes in either its dose or the doses of drugs with which it interacts, that is, on its labeled conditions of use. Rarely, metabolic drug-drug interactions may affect the ability of a drug to be safely marketed.

Additionally, the guidance describes the underlying concepts for the guidance:

- Adequate assessment of the safety and effectiveness of a drug includes a description of its metabolism and the contribution of metabolism to overall elimination.
- Metabolic drug-drug interaction studies should explore whether an investigational agent is likely to significantly affect the metabolic elimination of drugs already in the marketplace and, conversely, whether drugs in the marketplace are likely to affect the metabolic elimination of the investigational drug.
- Even drugs that are not substantially metabolized can have important effects on the metabolism of concomitant drugs. For this reason, metabolic drug-drug interactions should be explored, even for an investigational compound that is not eliminated significantly by metabolism.
- In some cases, metabolic drug-drug interaction studies cannot be informative unless metabolites and prodrugs have been identified and their pharmacological properties described.
- Identifying metabolic differences in patient groups based on genetic polymorphism, or on other readily identifiable factors, such as age, race, and gender, can aid in interpreting results.
- The impact of an investigational or approved interacting drug can be either to inhibit or induce metabolism.
- A specific objective of metabolic drug-drug interaction studies is to determine whether the interaction is sufficiently large to necessitate a dosage adjustment of the drug itself or the drugs it might be used with, or whether the interaction would require additional therapeutic monitoring.
- In some instances, understanding how to adjust dosage in the presence of an interacting drug, or how to avoid interactions, may allow marketing of a drug that would otherwise have been associated with an unacceptable level of toxicity. Sometimes a drug interaction may be used intentionally to increase levels or reduce elimination of another drug. Rarely, the degree of interaction caused by a drug, or the degree to which other drugs alter its metabolism, may be such that it cannot be marketed safely.

The guidance discusses the importance of "the development of sensitive and specific assays for a drug and its key metabolites." However, the guidance does not define what a "key metabolite" is. This terminology is no longer in vogue in current guidances.

The guidance makes some general statements about the magnitude of DDIs and the development of DDIs:

Unlike relatively fixed influences on metabolism, such as hepatic function or genetic characteristics, metabolic drug-drug interactions can lead to abrupt changes in exposure. Depending on the nature of the drugs, these effects could potentially occur when a drug is initially administered, when it has been titrated to a stable dose, or when an interacting drug is discontinued. Interactions can occur after even a single concomitant dose of an inhibitor.

The timing of understanding the potential for drug to be involved in DDIs (either as a perpetrator or as a victim) is also addressed:

> The effects of an investigational drug on the metabolism of other drugs and the effects of other drugs on an investigational drug's metabolism should be assessed relatively early in drug development so that the clinical implications of interactions can be assessed as fully as possible in later clinical studies.

The design of *in vivo* DDI studies is addressed in detail. This guidance advises that "If in vitro studies and other information suggest a need for in vivo metabolic drug-drug interaction studies, the following general issues and approaches should be considered." However, at times it may be prudent for the sponsor to conduct DDI studies to confirm the prediction of a lack of DDI. Such studies will be important when drugs will be administered in combination for an intended therapeutic effect or for marketing purposes to show that two drugs will not interact with one another:

> In vivo metabolic drug-drug interaction studies generally are designed to compare substrate levels with and without the interacting drug. Because a specific study may consider a number of questions and clinical objectives, no one correct study design for studying drug–drug interactions can be defined. A study can use a randomized crossover (e.g., S followed by S + I, S + I followed by S), a one-sequence crossover (e.g., S always followed by S + I or the reverse), or a parallel design (S in one group of subjects and S + I in another).

The point that there is no "one size fits all" approach is an important one and the guidance is consistent with the regulatory agencies point of view that drugs should be developed on a "case by case" basis:

> The following possible dosing regimen combinations for a substrate and interacting drug may also be used: single dose/single dose, single dose/multiple dose, multiple dose/single dose, and multiple dose/multiple dose. The selection of one of these or another study design depends on a number of factors for both the substrate and interacting drug, including (1) acute or chronic use of the substrate and/or interacting drug; (2) safety considerations, including whether a drug is likely to be an NTR (narrow therapeutic range) or non-NTR drug; (3) pharmacokinetic and pharmacodynamic characteristics of the substrate and interacting drugs; and (4) the need to assess induction as well as inhibition. The inhibiting/inducing drugs and the substrates should be dosed so that the exposure of both drugs are relevant to their clinical use.

The guidance then further describes selected situations, including:

> When both substrate and interacting drug are likely to be given chronically over an extended period of time, administration of the substrate to steady state with collection of blood samples over one or more dosing intervals could be followed by multiple dose administration of the interacting drug, again with collection of blood for measurement of both the substrate and the interacting drug (as feasible) over the same intervals. This is an example of a one-sequence crossover design.

At times an extended dosing period may be required as noted in the guidance:

> The time at steady state before collection of endpoint observations depends on whether inhibition or induction is to be studied. Inducers can take several days or longer to exert

> their effects, while inhibitors generally exert their effects more rapidly. For this reason, a more extended period of time after attainment of steady state for the substrate and interacting drug may be necessary if induction is to be assessed.

However, repeat-dose administration is not always required. The guidance states that "For a rapidly reversible inhibitor, administration of the interacting drug either just before or simultaneously with the substrate on the test day might be the appropriate design to increase sensitivity."

In other circumstances, a nontraditional approach may be required:

> When attainment of steady state is important and either the substrate or interacting drugs and/or their metabolites exhibit long half-lives, special approaches may be useful. These include use of a loading dose to achieve steady state conditions more rapidly and selection of a one-sequence crossover or a parallel design, rather than a randomized crossover study design.

The achievement of steady state should not be assumed, but should be proved, as indicated in this guidance:

> When a substrate and/or an interacting drug is to be studied at steady state, documentation that near steady state has been attained is important both for each drug and its metabolites of interest. This documentation can be accomplished by sampling over several days prior to the periods when samples are collected. This is important for both metabolites and the parent drug, particularly when the half-life of the metabolite is longer than the parent, and is especially important if both parent drug and metabolites are metabolic inhibitors or inducers.

The guidance does not define what a "metabolite of interest" is. The design of DDI studies can become complicated for drugs that will be administered in combination. The guidance states:

> If the drug interaction effects are to be assessed for both agents in a combination regimen, the assessment can be done in two separate studies. If the pharmacokinetic and pharmacodynamic characteristics of the drugs make it feasible, the dual assessment can be done in a single study. Some design options are randomized three-period crossover, parallel group, and one-sequence crossover.

These statements introduce the concept of DDIs as a consequence of induction of drug metabolizing agents, as is the case with St. John's wort. The prediction of the ability of a drug to induce DMEs is not a trivial task. In nonclinical studies, the induction of DMEs is usually only evaluated if auto-induction of the drug's metabolism is observed when the drug is repeatedly administered to the animal. Even if auto-induction of metabolism is observed, sponsors do not frequently identify which enzyme systems have been induced. The reason for this approach will vary from company to company, but is most likely due to the differences in substrate specificities between animal and human DMEs. Auto-induction of a drug's metabolism is a common phenomenon in rats, but rarely occurs in monkeys or dogs. Alternative methods for the prediction of induction of DMEs or transporters include human hepatocyte systems and assays to determine the binding of the candidate drug to certain nuclear receptors. Both of these techniques have their limitations. Even if induction of DMEs or transporters is predicted, this knowledge may not lead to the discontinuation of the development of a

drug. However, such knowledge may lead to an assessment of the induction of DMEs or transporters in early clinical development. Such studies may involve a simple evaluation of the pharmacokinetics of the drug after multiple-dose administration, an evaluation of urinary cortisol/6β-OH cortisol ratio (as a crude marker for CYP3A activity), or by the use of a cocktail of substrates of different CYP isoforms. The urinary cortisol/6β-OH cortisol ratio assay is not widely accepted as a robust assay for the evaluation of CYP3A activity, although it can provide some useful information. The limitations of this assay have been described in the literature (Ohno *et al.*, 2000; Galteau and Shamsa, 2003; Micuda *et al.*, 2007). The different composition of the CYP cocktails, and the pros and cons of the use of a CYP cocktail, to evaluate the potential for drug–drug interactions have been debated in the literature (Chainuvati *et al.*, 2003; Zhou, Tong and McLeod, 2004; Ryu *et al.*, 2007).

The subject of the patient population to be used is also addressed:

> Clinical drug-drug interaction studies may generally be performed using healthy volunteers or volunteers drawn from the general population, on the assumption that findings in this population should predict findings in the patient population for which the drug is intended. Safety considerations, however, may preclude the use of healthy subjects. In certain circumstances, subjects drawn from the general population and/or patients for whom the investigational drug is intended offer certain advantages, including the opportunity to study pharmacodynamic endpoints not present in healthy subjects and reduced reliance on extrapolation of findings from healthy subjects. In either patient or healthy/general population subject studies, performance of phenotype or genotype determinations to identify genetically determined metabolic polymorphisms is often important in evaluating effects on enzymes with polymorphisms, notably CYP2D6 and CYP2C19.

Examples of where safety considerations preclude the use of healthy subjects include drugs for the treatment of cancer. However, it should be noted that non-cytotoxic drugs can, and have, been safely administered to humans. For example, the interaction of Sutent (sunitinib) with ketoconazole and rifampin was evaluated in normal volunteers (US FDA, 2006l). It should be noted that conducting DDI studies in patient populations may be fraught with difficulties and the level of rigor that may be achieved in studies in normal healthy volunteers may not be able to be achieved in patients (e.g. cancer patients). The design of such studies in cancer patients may involve an abbreviated study design and the collection of fewer samples. It may also not be ethical to conduct a true DDI study since this would involve the withdrawal of all medications other than the drugs to be evaluated in the DDI study. It should also be noted that the use of nonprescription, herbal medications by cancer patients is not uncommon. Such substances would have a pronounced effect on the outcome of the DDI studies, as well as therapeutic response (Meijerman, Beijnen and Schellens, 2006; Marchetti *et al.*, 2007). However, many cancer (or other) patients do not think of these substances as having a potential effect and may not indicate their use of these substances when being screened for the study. Recent estimates suggest an overall prevalence for herbal preparation use of 13% to 63% among cancer patients and that 72% of the cancer patients do not inform their physicians of their use of complementary and alternative medicines (Sparreboom *et al.*, 2004).

It should also be noted that smoking and the ingestion of certain foods such as charbroiled meats or cruciferous vegetables may also impact the outcome of DDI (or other) studies. Food-derived chemicals, including constituents of cruciferous vegetables and fruits, modulate the expression of certain CYP isoforms and phase II enzymes. For

example, some dietary indoles and flavonoids activate CYP1A expression. Apart from altered CYP regulation, a number of dietary agents also inhibit CYP enzyme activity, leading to pharmacokinetic interactions with coadministered drugs. A well-described example is that of grapefruit juice, which contains psoralens and possibly other chemicals, which inactivate intestinal CYP3A4. Chemicals in teas and cruciferous vegetables may also inhibit human CYP enzymes that have been implicated in the bioactivation of chemical carcinogens. Thus, food constituents modulate CYP expression and function by a range of mechanisms, with the potential for both deleterious and beneficial outcomes (Murray, 2006).

It is interesting to note that certain clinical trials or therapies may take advantage of DDIs. Kaletra is a co-formulation of lopinavir and ritonavir. The ritonavir inhibits the CYP3A-mediated metabolism of lopinavir, thereby providing increased plasma levels of lopinavir (Abbott Laboratories, 2007). Presently, there is an ongoing clinical trial of PKC412 with or without itraconazole in treating patients with acute myeloid leukemia or myelodysplastic syndrome (U.S. National Institutes of Health, 2008). Although not overtly stated, the rationale for this combination appears to be the need to maintain high levels of PKC412 which has previously been shown to induce its own metabolism upon repeated administration to cancer patients (Propper *et al.*, 2001). In this treatment paradigm, PKC412 would induce its own metabolism but the metabolism of PKC412 would be inhibited by itraconazole, leading to elevated plasma concentrations of PKC412 compared with patients that had received PKC412 alone.

The choice of substrate and interacting drugs is also dealt with in great detail (US FDA, 1999a). For clarity, the majority of this section of the guidance is repeated here verbatim. The choice of the substrates for an investigational drug is addressed thus:

> In contrast to earlier approaches that focused mainly on a specific group of approved drugs (digoxin, hydrochlorothiazide) where coadministration was likely or the clinical consequences of an interaction were of concern, improved understanding of the metabolic basis of drug-drug interactions enables more general approaches to and conclusions from specific drug-drug interaction studies. In studying an investigational drug as the interacting drug, the choice of substrates (approved drugs) for initial in vivo studies depends on the P450 enzymes affected by the interacting drug. In testing inhibition, the substrate selected should generally be one whose pharmacokinetics is markedly altered by coadministration of known specific inhibitors of the enzyme systems (i.e., a very sensitive substrate should be chosen) to assess the impact of the interacting investigational drug. Examples of substrates include, but are not limited to, (1) midazolam, buspirone, felodipine, simvastatin, or lovastatin for CYP3A4; (2) theophylline for CYP1A2; (3) S-warfarin for CYP2C9; and (4) desipramine for CYP2D6. If the initial study is positive for inhibition, further studies of other substrates may be useful, representing a range of substrates based on the likelihood of coadministration. For example, possible substrates for further study of a CYP3A4 interacting investigational drug might include dihydropyridine calcium channel blockers and triazolobenzodiazepines, or for a CYP2D6 inhibiting investigational drug might include metoprolol. If the initial study is negative with the most sensitive substrates, it can be presumed that less sensitive substrates will also be unaffected.

The subject of the investigational drug as a substrate is addressed in the following manner:

> In testing an investigational drug for the possibility that its metabolism is inhibited or induced (i.e., as a substrate), selection of the interacting drugs should be based on in vitro or other metabolism studies identifying the enzyme systems that metabolize the drug. The

> choice of interacting drug should then be based on known, important inhibitors of the pathway under investigation. For example, if the investigational drug is shown to be metabolized by CYP3A4 and the contribution of this enzyme to the overall elimination of this drug is substantial, the choice of inhibitor and inducer could be ketoconazole and rifampin, respectively, because of the substantial effects of these interacting drugs on CYP3A4 metabolism (i.e., they are the most sensitive in identifying an effect of interest). If the study results are negative, then absence of a clinically important drug-drug interaction for the metabolic pathway could be claimed. If the clinical study of the most potent specific inhibitor/inducer is positive and the sponsor wishes to claim lack of an interaction between the test drug and other less potent specific inhibitors, or give advice on dosage adjustment, further clinical studies would generally be recommended. Certain approved drugs are not optimal selections as the interacting drug. For example, cimetidine is not considered an optimal choice to represent drugs inhibiting a given pathway because its inhibition affects multiple metabolic pathways as well as certain drug transporters.

This statement focuses on "class effects" where a class would be all substrates of a given CYP isoform. Although studies of this type are often conducted, studies using substrate or inhibitor drugs that would potentially be commonly coadministered with the test agent may be appropriate. When designing the DDI component of the clinical pharmacology plan, the sponsor should pay attention to the use of agents that may not routinely be recognized as concomitant medications. For example, if the test agent is a CYP3A substrate, the effect of St. John's wort or grapefruit juice on the metabolism of the test agent should be considered. Conversely, if the test agent will be used in women of childbearing potential, the effect of the test agent on the metabolism of oral contraceptives could be considered.

The guidance notes that the "for an investigational agent used as either an interacting drug or substrate, the route of administration should generally be the one planned for in product labeling. When multiple routes are being developed, the necessity for doing metabolic drug-drug interaction studies by all routes should be based on the expected mechanism of interaction and the similarity of corresponding concentration-time profiles for parent and metabolites." It further states that the dose selected for the DDI study should "maximize the possibility of finding an interaction. For this reason, the maximum planned or approved dose and shortest dosing interval of the interacting drug (as inhibitors or inducers) should be used. Doses smaller than those to be used clinically may be needed for substrates on safety grounds and may be more sensitive to the effect of the interacting drug." This latter point is one the sponsor should play close attention to.

The end points of the study, in terms of pharmacokinetic parameters are primarily measurements of exposure, such as AUC, $C_{max}$, time to $C_{max}$ ($T_{max}$), and others as appropriate. Additionally, pharmacokinetic parameters such as clearance, volumes of distribution, and half-lives may be calculated. Additional measures may help in steady-state studies (e.g. trough concentration, $C_{min}$) to demonstrate that dosing strategies were adequate to achieve near steady state before and during the interaction. The guidance gives additional details regarding the calculation of PK parameters and the collection of samples.

The focus of the guidance is clearly illustrated in the following statement: "Because this guidance focuses on metabolic drug-drug interactions, protein binding determinations are considered unnecessary except for data interpretation." Additionally, this guidance does not address transporter-mediated DDIs. Since many drugs are substrates or inhibitors of both DMEs and transporters, it may be difficult to determine whether

an observed interaction is due to an effect on the metabolism or transport of the drug. Although it would be ideal to be able to deconvolute the individual contributions of metabolism and transport to a given DDI, it is not essential to do so. The magnitude of effect should be reported in the label, regardless of whether the interaction is due to a metabolic, transport or plasma protein binding DDI.

The guidance has not addressed some factors (e.g. gender, age, the use of multiple inhibitors, effect on GI microflora, effect of GI motility, or the effect of pH) that could play a role in the extent of a DDI.

DDIs may arise as a consequence of the physicochemical properties of the drug. For drugs that show marked pH-dependent solubility, an evaluation of the effect of pH on the exposure to the drug should be conducted. This type of DDI may occur when the drug is taken with antacids. The interaction can be "traditional" (i.e. inhibition of CYP-mediated metabolism) or "nontraditional" (change in pH resulting in change of solubility of the drug or physical adsorption of the drug to the antacid) (Neuvonen and Kivisto, 1994; Sadowski, 1994). Additionally, it should be noted that the pH of the stomach can change with age. For example, approximately 25% of elderly subjects suffer from achlorhydria, a condition in which the pH of the stomach is elevated (Mallet, 2008). Achlorhydria can interfere with the absorption of ketoconazole (Hurwitz *et al.*, 2003). This effect can be extremely pronounced and it has been noted that a manufacturer of ketoconazole has recommended that "in cases of achlorhydria, the patients should be advised to dissolve each tablet in 4 ml aqueous solution of 0.2 N HCl" and to "use a drinking straw so as to avoid contact with the teeth." On a related note, acidic beverages (e.g. Coca-Cola) have been shown to enhance the exposure to carbamazepine and itraconazole (Lange *et al.*, 1997; Malhotra, Dixit and Garg, 2002). It is possible that the change in stomach pH (from pH 1.4–2.1 to pH 3–7) when food is administered may be a contributing factor to observed effect of food upon the absorption of a given drug (Dressman *et al.*, 1998).

The guidance (US FDA, 1999a) notes that the desired goal of the DDI study is to "determine the clinical significance of any increase or decrease in exposure to the substrate in the presence of the interacting drug" and that:

> Results of drug-drug interaction studies should be reported as 90% confidence intervals about the geometric mean ratio of the observed pharmacokinetic measures with (S + I) and without the interacting drug (S). Confidence intervals provide an estimate of the distribution of the observed systemic exposure measure ratio of S + I versus S alone and convey a probability of the magnitude of the interaction. In contrast, tests of significance are not appropriate because small, consistent systemic exposure differences can be statistically significant ($p < 0.05$) but not clinically relevant.

The guidance further addresses this topic in great detail:

> When a drug-drug interaction is clearly present (e.g., comparisons indicate twofold or greater increments in systemic exposure measures for S + I) the sponsor should be able to provide specific recommendations regarding the clinical significance of the interaction based on what is known about the dose-response and/or PK/PD relationship for either the investigational agent or the approved drugs used in the study. … The sponsor may wish to make specific claims in the package insert that no drug-drug interaction is expected. In these instances, the sponsor should be able to recommend specific no effect boundaries, or clinical equivalence intervals, for a drug-drug interaction. No effect boundaries define the interval within which a change in a systemic exposure measure is considered not clinically meaningful.

There are three approaches to define no effect boundaries.

*Approach 1:* No effect boundaries can be based on population (group) average dose and/or concentration-response relationships, PK/PD models, and other available information for the substrate drug. If the 90% confidence interval for the systemic exposure measurement in the drug-drug interaction study falls completely within the no effect boundaries, the sponsor may conclude that no clinically significant drug-drug interaction was present.

*Approach 2:* No effect boundaries may also be based on the concept that a drug-drug interaction study addresses the question of switchability between the substrate given in combination with an interacting drug (test) versus the substrate given alone. Based on this concept, the sponsor may wish to use an individual equivalence criterion to allow scaling of the no effect boundary and to determine other useful information as well. Sponsors who wish to use this approach are encouraged to contact the Office of Clinical Pharmacology and Biopharmaceutics to discuss approaches to study design and data analysis.

*Approach 3:* In the absence of no effect boundaries defined in (1) or (2) above, a sponsor may use a default no effect boundary of 80–125% for both the investigational drug and the approved drugs used in the study. When the 90% confidence intervals for systemic exposure ratios fall entirely within the equivalence range of 80–125%, standard Agency practice is to conclude that no clinically significant differences are present.

The guidance notes that:

The selection of the number of subjects for a given drug-drug interaction study will depend on how small an effect is clinically important to detect, or rule out, the inter- and intrasubject variability in pharmacokinetic measurements, and possibly other factors or sources of variability not well recognized. In addition, the number of subjects will depend on how the results of the drug-drug interaction study will be used, as described above. This guidance should not be interpreted by sponsors as generally recommending the inclusion of some number of subjects in a drug-drug interaction study such that the 90% confidence interval for the ratio of pharmacokinetic measurements falls entirely within the no effect boundaries of 80–125%. This approach, however, could be deemed appropriate by a sponsor, after considering the expected outcome of a drug-drug interaction study, the anticipated magnitude of variability in pharmacokinetic measurements, and the desired label claim that no clinically significant drug-drug interaction was present.

Since this guidance was issued in 1999, our knowledge of the design, conduct, and interpretation of DDIs studies has evolved. This has led to the FDA issuing a draft guidance "Drug Interaction Studies—Study Design, Data Analysis, and Implications for Dosing and Labeling" in 2006 (US FDA, 2006g). This draft guidance mainly focuses on *in vivo* studies but does also address some aspects of *in vitro* studies. It is a substantial guidance, totaling 52 pages in its draft form and containing a great deal of information in terms of basic information regarding substrates and inhibitors as well as detailed description of potential study designs. A large amount of this information can be found in the Appendices which account for over 66% of the page count for this draft guidance.

The major advance in this guidance is the recognition of the importance of drug transporters in DDIs. The guidance makes this point thus, "Furthermore, not every drug-drug interaction is metabolism-based, but may arise from changes in pharmacokinetics caused by absorption, distribution, and excretion interactions. Drug-drug interactions related to transporters are being documented with increasing frequency and are important to consider in drug development." The guidance notes the interrelated nature of guidances by stating that "Discussion of metabolic and other types

of drug-drug interactions is also provided in other guidances, including the International Conference on Harmonization (ICH) E7 Studies in Support of Special Populations: Geriatrics, and E3 Structure and Content of Clinical Study Reports, and FDA guidances for industry on Studying Drugs Likely to be Used in the Elderly and Study and Evaluation of Gender Differences in the Clinical Evaluation of Drugs."

The guidance addresses several different types of DDIs. Firstly, the draft guidance addresses metabolism-based drug–drug interactions. The draft guidance provides a general introduction to this topic:

> Many metabolic routes of elimination, including most of those occurring through the P450 family of enzymes, can be inhibited or induced by concomitant drug treatment. Observed changes arising from metabolic drug-drug interactions can be substantial – an order of magnitude or more decrease or increase in the blood and tissue concentrations of a drug or metabolite – and can include formation of toxic and/or active metabolites or increased exposure to a toxic parent compound. These large changes in exposure can alter the safety and efficacy profile of a drug and/or its active metabolites in important ways. This is most obvious and expected for a drug with a narrow therapeutic range (NTR), but is also possible for non-NTR drugs as well (e.g., HMG CoA reductase inhibitors).

The purpose of the evaluation of metabolism-based DDIs is addressed:

> It is important that metabolic drug-drug interaction studies explore whether an investigational agent is likely to significantly affect the metabolic elimination of drugs already in the marketplace and likely in medical practice to be taken concomitantly and, conversely, whether drugs in the marketplace are likely to affect the metabolic elimination of the investigational drug. Even drugs that are not substantially metabolized can have important effects on the metabolism of concomitant drugs. For this reason, metabolic drug-drug interactions should be explored, even for an investigational compound that is not eliminated significantly by metabolism.

In this draft guidance, the applicability of the evaluation of metabolism-based DDIs for therapeutic biologics is described:

> Classical biotransformation studies are not a general requirement for the evaluation of therapeutic biologics (ICH guidance *S6 Preclinical Safety Evaluation of Biotechnology-Derived Pharmaceuticals*), although certain protein therapeutics modify the metabolism of drugs that are metabolized by the P450 enzymes. Type I interferons, for example, inhibit CYP1A2 production at the transcriptional and post-translational levels, inhibiting clearance of theophylline. The increased clinical use of therapeutic proteins may raise concerns regarding the potential for their impacts on drug metabolism. Generally, these interactions cannot be detected by in vitro assessment. Consultation with FDA is appropriate before initiating metabolic drug-drug interaction studies involving biologics.

It is interesting to see the mention of the potential effect of biologics on CYP expression. This topic has not previously been addressed in guidances related to DDIs. Additionally, the potential for DDIs between therapeutic monoclonal antibodies is one that has not been discussed in great depth by regulatory agencies. However, this subject was recently extensively reviewed by Seitz and Zhou (2007).

The possibility of a DDI-based on the effect of drugs on the degradation of CYP isoforms has not been addressed. Although the mechanisms of degradation of the CYP isoforms have been described, the relevance of this mechanism as a potential source of DDIs is unknown (Correia, 2003; Correia, Sadeghi and Mundo-Paredes, 2005).

The role of genotype on metabolism-based DDIs is also addressed:

> Identifying metabolic differences in patient groups based on genetic polymorphism, or on other readily identifiable factors, such as age, race, and gender, can aid in interpreting results. The extent of interactions may be defined by these variables (e.g., CYP2D6 genotypes). Further, in subjects who lack the major clearance pathway, remaining pathways become important and should be understood and examined. A specific objective of metabolic drug-drug interaction studies is to determine whether the interaction is sufficiently large to necessitate a dosage adjustment of the drug itself or the drugs with which it might be used, or whether the interaction would require additional therapeutic monitoring. In some instances, understanding how to adjust dose or dosage regimen in the presence of an interacting drug, or how to avoid interactions, may allow marketing of a drug that would otherwise have been associated with an unacceptable level of toxicity. Sometimes a drug interaction can be used intentionally to increase levels or reduce elimination of another drug (e.g., ritonavir and lopinavir). Rarely, the degree of interaction caused by a drug, or the degree to which other drugs alter its metabolism, can be such that it cannot be marketed safely.

As can be seen, some of the wording of this draft guidance is taken verbatim from the 1999 guidance. It will be interesting to see whether the 1999 guidance is withdrawn once this draft guidance is finalized.

Transporter-based drug–drug interactions are mentioned, albeit with a lesser extent of detail than was provided for metabolism-based DDIs:

> Transporter-based interactions have been increasingly documented. Examples of these include the inhibition or induction of transport proteins, such as P glycoprotein (P-gp), organic anion transporter (OAT), organic anion transporting polypeptide (OATP), organic cation transporter (OCT), multidrug resistance-associated proteins (MRP), and breast cancer resistant protein (BCRP). Examples of transporter-based interactions include the interactions between digoxin and quinidine, fexofenadine and ketoconazole (or erythromycin), penicillin and probenecid, and dofetilide and cimetidine. Of the various transporters, P-gp is the most well understood and may be appropriate to evaluate during drug development.

The draft guidance provides a useful table that lists some of the major human transporters and known substrates, inhibitors, and inducers.

This draft guidance provides some input into the strategy of the evaluation of DDI studies. Guidances have traditionally not provided much input into strategy and therefore the fact that this draft guidance provides thoughts on this topic is novel and refreshing. The draft guidance states:

> To the extent possible, drug development should follow a sequence in which early in vitro and in vivo investigations can either fully address a question of interest or provide information to guide further studies. Optimally, a sequence of studies could be planned, moving from in vitro studies to in vivo human studies, including those employing special study designs and methodologies where appropriate. In many cases, negative findings from early in vitro and early clinical studies can eliminate the need for later clinical investigations. Early investigations should explore whether a drug is eliminated primarily by excretion or metabolism, with identification of the principal metabolic routes in the latter case. Using suitable in vitro probes and careful selection of interacting drugs for early in vivo studies, the potential for drug-drug interactions can be studied early in the development process, with further study of observed interactions assessed later in the process, as needed. These

> early studies can also provide information about dose, concentration, and response relationships in the general population, specific populations, and individuals, which can be useful in interpreting the consequences of a drug-drug interaction. Once potential drug-drug interactions have been identified, based on in vitro and/or in vivo studies, sponsors are encouraged to design and examine the safety and efficacy databases of larger clinical studies, as feasible, to (1) permit confirmation/discovery of the interactions predicted from earlier studies and/or (2) verify that dosage adjustments or other prescribing modifications made in response to the potential interaction(s) have been adequate to avoid undesired consequences of the drug-drug interaction.

The guidance clearly indicates that an early understanding of the major route of clearance would be useful, but does not provide insight as to how the relative contributions of metabolism and excretion are determined for an orally administered drug.

The strategy for *in vitro* studies is outlined in the following manner:

> A complete understanding of the quantitative relationship between the in vitro findings and in vivo results of metabolism/drug-drug interaction studies is still emerging. Nonetheless, in vitro studies can frequently serve as a screening mechanism to rule out the importance of a metabolic pathway and the drug-drug interactions that occur through this pathway so that subsequent in vivo testing is unnecessary. This opportunity should be based on appropriately validated experimental methods and rational selection of substrate/interacting drug concentrations. For example, if suitable in vitro studies at therapeutic concentrations indicate that CYP1A2, CYP2C8, CYP2C9, CYP2C19, CYP2D6, or CYP3A enzyme systems do not metabolize an investigational drug, then clinical studies to evaluate the effect of CYP2D6 inhibitors or CYP1A2, CYP2C8, CYP2C9, CYP2C19, or CYP3A inhibitors/inducers on the elimination of the investigational drug will not be needed. Similarly, if in vitro studies indicate that an investigational drug does not inhibit CYP1A2, CYP2C8, CYP2C9, CYP2C19, CYP2D6, or CYP3A metabolism, then corresponding in vivo inhibition-based interaction studies of the investigational drug and concomitant medications eliminated by these pathways are not needed. Figure 1 in Appendix B shows a decision tree on when in vivo interaction studies are indicated based on in vitro metabolism, inhibition, and induction and in vivo metabolism data. The CYP2D6 enzyme has not been shown to be inducible. Recent data have shown co-induction of CYP2C, CYP2B and ABCB1 (P-gp) transporter with CYP3A. CYP3A appears to be sensitive to all known co-inducers. Therefore, to evaluate whether an investigational drug induces CYP1A2, CYP2C8, CYP2C9, CYP2C19, or CYP3A, the initial in vitro induction evaluation may include only CYP1A2 and CYP3A. If in vitro studies indicate that an investigational drug does not induce CYP3A metabolism, then in vivo induction-based interaction studies of the investigational drug and concomitant medications eliminated by CYP2C/CYP2B and CYP3A may not be needed. Drug interactions based on CYP2B6 are emerging as important interactions. When appropriate, in vitro evaluations based on this enzyme can be conducted. Other CYP enzymes, including CYP2A6 and CYP2E1, are less likely to be involved in clinically important drug interactions, but should be considered when appropriate.

This section identifies the inter-relatedness of the induction of selected CYP isoforms and drug transporters and a refinement in the understanding of the relative importance of the CYP isoforms (notably the CYP 2C family and CYP2B6). The induction of CYP isoforms is a topic that has not been well developed in previous guidances. The Appendices for this draft guidance describe general considerations in the *in vitro* evaluation of CYP-related metabolism and interactions; considerations in the experimental design, data analysis, and data interpretation in DME identification, including CYP enzymes (new drug as a substrate), CYP inhibition (new drug as an inhibitor), and CYP induc-

tion (new drug as an inducer), respectively. Additionally, the Appendices contain general considerations in the *in vitro* evaluation of P-gp substrates and inhibitors, including decision trees on when *in vivo* P-gp-based interaction studies are indicated based on *in vitro* evaluation. *In vivo* clinical investigations are also addressed in detail. The draft guidance notes that:

> In addition to in vitro metabolism and drug-drug interaction studies, appropriately designed pharmacokinetic studies, usually performed in the early phases of drug development, can provide important information about metabolic routes of elimination, their contribution to overall elimination, and metabolic drug-drug interactions. Together with information from in vitro studies, these in vivo investigations can be a primary basis of labeling statements and can often help avoid the need for further investigations.

The last statement is interesting, since it allows for the possibility that *in vitro* data can be used for labeling statements and that *in vivo* studies may not be required in all situations. The product label for Inspra (eplerenone) states that "Eplerenone is not an inhibitor of CYP1A2, CYP3A4, CYP2C19, CYP2C9, or CYP2D6. Eplerenone did not inhibit the metabolism of chlorzoxazone, diclofenac, methylphenidate, losartan, amiodarone, dexamethasone, mephobarbital, phenytoin, phenacetin, dextromethorphan, metoprolol, tolbutamide, amlodipine, astemizole, cisapride, 17α-ethinyl estradiol, fluoxetine, lovastatin, methylprednisolone, midazolam, nifedipine, simvastatin, triazolam, verapamil, and warfarin in vitro" (G.D. Searle LLC, a Division of Pfizer Inc. 2008). One point that is not addressed by this draft guidance is the assessment of the stability of the test article under the conditions that the *in vitro* study is performed. The Clinical Pharmacology Biopharmaceutics review of an *in vitro* induction study performed with Inspra in human hepatocytes notes that the data provided by the sponsor was not sufficient to determine whether the induction was caused by eplerenone or one of its metabolites (US FDA, 2002d). The implication was that it may be prudent to determine the metabolite profile in the *in vitro* incubations that are performed for the assessment of induction of CYP isoforms in human hepatocytes.

The draft guidance provides some general issues and approaches that should be considered if *in vitro* studies suggest that *in vivo* studies would be helpful. Consultation with the FDA regarding study protocols is recommended. In the following discussion, the term substrate (S) is used to indicate the drug studied to determine whether its exposure is changed by another drug, termed the interacting drug (I). Depending on the study objectives, the substrate and the interacting drug can be the investigational agents or approved products. Study designs are addressed as follows in the draft guidance:

> In vivo drug-drug interaction studies generally are designed to compare substrate concentrations with and without the interacting drug. Because a specific study can consider a number of questions and clinical objectives, many study designs for studying drug-drug interactions can be considered. A study can use a randomized crossover (e.g., S followed by S + I, S + I followed by S), a one-sequence crossover (e.g., S always followed by S + I or the reverse), or a parallel design (S in one group of subjects and S + I in another). The following possible dosing regimen combinations for a substrate and interacting drug can also be used: single dose/single dose, single dose/multiple dose, multiple dose/single dose, and multiple dose/multiple dose. The selection of one of these or another study design depends on a number of factors for both the substrate and interacting drug, including (1) acute or chronic use of the substrate and/or interacting drug; (2) safety considerations, including whether a drug is likely to be an NTR (narrow therapeutic range) or non-NTR

drug; (3) pharmacokinetic and pharmacodynamic characteristics of the substrate and interacting drugs; and (4) assessment of induction as well as inhibition. The inhibiting/inducing drugs and the substrates should be dosed so that the exposures of both drugs are relevant to their clinical use, including the highest doses likely to be used. Simulations can be helpful in selecting an appropriate study design. The following considerations may be useful:

- When attainment of steady state is important and either the substrate or interacting drugs and/or their metabolites have long half-lives and a loading dose to reach steady state promptly cannot be used, special approaches may be needed. These include the selection of a one-sequence crossover or a parallel design, rather than a randomized crossover study design.
- When it is important that a substrate and/or an interacting drug be studied at steady state because the effect of an interacting drug is delayed, as is the case for inducers and certain inhibitors, documentation that near steady state has been attained for the pertinent drug and metabolites of interest is critical. This documentation can be accomplished by sampling over several days prior to the periods when test samples are collected. This is important for both metabolites and the parent drug, particularly when the half-life of the metabolite is longer than the parent, and is especially important if both parent drug and metabolites are metabolic inhibitors or inducers.
- Studies can usually be open label (unblinded), unless pharmacodynamic endpoints (e.g., adverse events that are subject to bias) are critical to the assessment of the interaction.
- For a rapidly reversible inhibitor, administration of the interacting drug either just before or simultaneously with the substrate on the test day might increase sensitivity. For a mechanism-based inhibitor (a drug that requires metabolism prior to its inactivation of the enzyme; examples include erythromycin), administration of the inhibitor prior to the administration of the substrate drug can maximize the effect. If the absorption of an interacting drug (e.g., an inhibitor or an inducer) may be affected by other factors (e.g., the gastric pH), it may be appropriate to control the variables and confirm the absorption through plasma level measurements of the interacting drug.
- When the effects of two drugs on one another are of interest, the potential for interactions can be evaluated in a single study or two separate studies. Some design options are randomized three-period crossover, parallel group, and one-sequence crossover.
- To avoid variable study results because of uncontrolled use of dietary supplements, juices, or other foods that may affect various metabolizing enzymes and transporters during in vivo studies, it is important to exclude their use when appropriate.

Examples of statements in a study protocol could include "Participants will be excluded for the following reasons: Use of prescription or over-the-counter medications, including herbal products, or alcohol within two weeks prior to enrollment" and "For at least two weeks prior to the start of the study until its conclusion, volunteers will not be allowed to eat any food or drink any beverage containing alcohol, grapefruit or grapefruit juice, apple or orange juice, vegetables from the mustard green family (e.g., kale, broccoli, watercress, collard greens, kohlrabi, brussels sprouts, mustard) and charbroiled meats."

The last statement is a clear reminder that "natural" components in foods can have an effect on the metabolism and/or absorption of a drug.

The draft guidance notes that:

Clinical drug-drug interaction studies can generally be performed using healthy volunteers. Findings in this population should predict findings in the patient population for which

> the drug is intended. Safety considerations may preclude the use of healthy subjects, however, and in certain circumstances, subjects drawn from the population of patients for whom the investigational drug is intended offer advantages, including the opportunity to study pharmacodynamic endpoints not present in healthy subjects. Performance of phenotype or genotype determinations to identify genetically determined metabolic polymorphisms is important in evaluating effects on enzymes with polymorphisms, notably CYP2D6, CYP2C19, and CYP2C9. The extent of drug interactions (inhibition or induction) may be different depending on the subjects' genotype for the specific enzyme being evaluated. Subjects lacking the major clearance pathway, for example, cannot show metabolism and remaining pathways can become important and should be understood and examined.

The last statement could be taken as implying that the metabolism *in vivo* in humans may need to be determined for subjects of varying genotypes if the *in vitro* data indicate that the major route of metabolism of the drug is via a polymorphic DME.

The choice of substrate and interacting drugs for several different questions is also addressed. The section on "Investigational Drug as an Inhibitor or an Inducer of CYP Enzymes" updates the traditional thinking on the selection of drugs for use in DDI studies by noting that:

> In contrast to earlier approaches that focused mainly on a specific group of approved drugs (digoxin, hydrochlorothiazide) where co-administration was likely or the clinical consequences of an interaction were of concern, improved understanding of the mechanistic basis of metabolic drug-drug interactions enables more general approaches to and conclusions from specific drug-drug interaction studies. In studying an investigational drug as the interacting drug, the choice of substrates (approved drugs) for initial in vivo studies depends on the P450 enzymes affected by the interacting drug. In testing inhibition, the substrate selected should generally be one whose pharmacokinetics are markedly altered by co-administration of known specific inhibitors of the enzyme systems to assess the impact of the interacting investigational drug. Examples of substrates include (1) midazolam for CYP3A; (2) theophylline for CYP1A2; (3) repaglinide for CYP2C8; (4) warfarin for CYP2C9 (with the evaluation of S-warfarin); (5) omeprazole for CYP2C19; and (6) desipramine for CYP2D6. Additional examples of substrates, along with inhibitors and inducers of specific CYP enzymes, are listed in Table 2 in Appendix A. If the initial study determines an investigation drug either inhibit or induce metabolism, further studies using other substrates, representing a range of substrates, based on the likelihood of co-administration, may be useful. If the initial study is negative with the most sensitive substrates (for sensitive substrates, see Tables 3 and 4 in Appendix A), it can be presumed that less sensitive substrates will also be unaffected.
>
> CYP3A inhibitors can be classified based on their in vivo fold-change in the plasma AUC of oral midazolam or other CYP3A substrate, when given concomitantly. For example, if an investigational drug increases the AUC of oral midazolam or other CYP3A substrates by 5-fold or higher (>5-fold), it can be labeled as a strong CYP3A inhibitor. If an investigational drug, when given at the highest dose and shortest dosing interval, increases the AUC of oral midazolam or other sensitive CYP3A substrates by between 2- and 5-fold (>2- and <5-fold) when given together, it can be labeled as a moderate CYP3A inhibitor. Similarly, if an investigational drug, when given at the highest dose and shortest dosing interval, increases the AUC of oral midazolam or other sensitive CYP3A substrates by between 1.25- and 2-fold (>1.25- and <2-fold), it can be labeled as a weak CYP3A inhibitor. When an investigational drug is determined to be an inhibitor of CYP3A, its interaction with sensitive CYP3A substrates or CYP3A substrates with narrow therapeutic range (see Table 3 in Appendix A for a list) can be described in various sections of the labeling, as

> appropriate. Similar classifications of inhibitors of other CYP enzymes are discussed in section V.
>
> When an in vitro evaluation cannot rule out the possibility that an investigational drug is an inducer of CYP3A (see Appendix C-3), an in vivo evaluation can be conducted using the most sensitive substrate (e.g., oral midazolam, see Table 3 in Appendix A). When midazolam has been co-administered orally following administration of multiple doses of the investigational drug, as may have been done as part of an in vivo inhibition evaluation, and the results are negative, it can be concluded that the investigational drug is not an inducer of CYP3A (in addition to the conclusion that it is not an inhibitor of CYP3A). In vivo induction evaluation has often been conducted with oral contraceptives. However, as they are not the most sensitive substrates, negative data may not exclude the possibility that the investigational drug may be an inducer of CYP3A.

The use of a CYP cocktail approach is also considered by this draft guidance:

> Simultaneous administration of a mixture of substrates of CYP enzymes in one study (i.e., a "cocktail approach") in human volunteers is another way to evaluate a drug's inhibition or induction potential, provided that the study is designed properly and the following factors are present: (1) the substrates are specific for individual CYP enzymes; (2) there are no interactions among these substrates; and (3) the study is conducted in a sufficient number of subjects (see section IV.G). Negative results from a cocktail study can eliminate the need for further evaluation of particular CYP enzymes. However, positive results can indicate the need for further in vivo evaluation to provide quantitative exposure changes (such as AUC, $C_{max}$), if the initial evaluation only assessed the changes in the urinary parent to metabolite ratios. The data generated from a cocktail study can supplement data from other in vitro and in vivo studies in assessing a drug's potential to inhibit or induce CYP enzymes.

The use of CYP cocktails has been described previously in this chapter.

The evaluation of the investigational drug as a substrate of CYP enzymes is also a topic addressed by this comprehensive draft guidance. The draft guidance states:

> In testing an investigational drug for the possibility that its metabolism is inhibited or induced (i.e., as a substrate), selection of the interacting drugs should be based on in vitro or in vivo studies identifying the enzyme systems that metabolize the drug. The choice of interacting drug can then be based on known, important inhibitors of the pathway under investigation. For example, if the investigational drug is shown to be metabolized by CYP3A and the contribution of this enzyme to the overall elimination of this drug is either substantial (>25% of the clearance pathway) or unknown, the choice of inhibitor and inducer could be ketoconazole and rifampin, respectively, because they are the most sensitive in identifying an effect of interest. If the study results are negative, then absence of a clinically important drug-drug interaction for the metabolic pathway would have been demonstrated. If the clinical study of the strong, specific inhibitor/inducer is positive and the sponsor wished to determine whether there is an interaction between the test drug and other less potent specific inhibitors or inducers, or to give advice on dosage adjustment, further clinical studies would generally be needed (see Table 2, Appendix A, for a list of CYP inhibitors and inducers; see Table 5, Appendix A, for additional 3A inhibitors). If a drug is metabolized by CYP3A and its plasma AUC is increased 5-fold or higher by CYP3A inhibitors, it is considered a sensitive substrate of CYP3A. The labeling can indicate that it is a "sensitive CYP3A substrate" and its use with strong or moderate inhibitors may call for caution, depending on the drug's exposure-response relationship. If a drug is metabolized by CYP3A and its exposure-response relationship indicates that increases in the exposure levels by the concomitant use of CYP3A inhibitors may lead to serious

> safety concerns (e.g., Torsades de Pointes), it is considered as a "CYP3A substrate with narrow therapeutic range" (see Table 3 of Appendix A for a list). Similar classifications of substrates of other CYP enzymes are discussed in section V and listed in Table 6, Appendix A.

The draft guidance then expands the concept of DDIs beyond interactions with other pharmaceuticals and addresses the potential for interaction with food constituents and herbal medications:

> If an orally administered drug is a substrate of CYP3A and has low oral bioavailability because of extensive presystemic extraction contributed by enteric CYP3A, grapefruit juice may have a significant effect on its systemic exposure. Use of the drug with grapefruit juice may call for caution, depending on the drug's exposure-response relationship (see section V for labeling implications). If a drug is a substrate of CYP3A or P-gp and co-administration with St. John's wort can decrease the systemic exposure and effectiveness, St John's wort may be listed in the labeling along with other known inducers, such as rifampin, rifabutin, rifapentin, dexamethasone, phenytoin, carbamazepine, or phenobarbital, as possibly decreasing plasma levels.

The role of polymorphic DMEs on DDIs is also covered:

> If a drug is metabolized by a polymorphic enzyme (such as CYP2D6, CYP2C9, or CYP2C19), the comparison of pharmacokinetic parameters of this drug in poor metabolizers versus extensive metabolizers may indicate the extent of interaction of this drug with strong inhibitors of these enzymes, and make interaction studies with such inhibitors unnecessary. When the above study shows significant interaction, further evaluation with weaker inhibitors may be necessary.

In recognition that polypharmacy is a common situation in many patients the guidance states:

> There may be situations when an evaluation of the effect of multiple CYP inhibitors on the drug can be informative. For example, it may be appropriate to conduct an interaction study with more that one inhibitor if all of the following conditions are met: (1) the drug exhibits blood concentration-dependent safety concerns; (2) multiple CYP enzymes are responsible for the metabolic clearance of the drug; (3) the residual or non-inhibitable drug clearance is low. Under these conditions, the effect of multiple, CYP-selective inhibitors on the blood AUC of a drug may be much greater than the product of the fold AUC changes observed when the inhibitors are given individually with the drug. The degree of uncertainty will depend on the residual fractional clearance (the smaller the fraction, the greater the concern) and the relative fractional clearances of the inhibited pathway. However, if results from a study with a single inhibitor trigger a safety concern (i.e., contraindication), no multiple inhibitor studies will be necessary. Additional considerations may include the likelihood of co-administration of the drug with multiple inhibitors. Before investigating the impact of multiple inhibitors on drug exposure, it is important to first characterize the individual effects of the CYP inhibitors and to estimate the combined effect of the inhibitors based on computer simulation. For safety concerns, lower doses of the investigational drug may be appropriate for evaluating the fold increase in systemic exposure when combined with multiple inhibitors.

Since many interacting agents may have an impact on both drug metabolism and transport of the drug, the guidance states:

> The implications of simultaneous inhibition of a dominant CYP enzyme(s) and an uptake or efflux transporter that controls the availability of the drug to CYP enzymes can be just as profound as that of multiple CYP inhibitors. For example, the large effect of co-administration of itraconazole and gemfibrozil on the systemic exposure (AUC) of repaglinide may be attributed to collective effects on both enzyme and transporters. Unfortunately, current knowledge does not permit the presentation of specific guidance. The sponsor will need to use appropriate judgement when considering this situation.

The evaluation of transporter-based DDIs is also addressed. However, as can be seen by the relative brevity of the text, the importance of such interactions is not yet fully understood. Additionally, most focus (from both industry and regulatory agencies) has been on P-gp, although the importance of other transporters for certain classes of drugs is acknowledged. The draft guidance addresses investigational drug as an inhibitor or an inducer of P-gp transporter thus: "In testing an investigational drug for the possibility that it may be an inhibitor/inducer of P-gp, selection of digoxin or other known substrates of P-gp may be appropriate." In a similar brief manner, the evaluation of the investigational drug as a substrate of P-gp transporter is addressed as follows: "In testing an investigational drug for the possibility that its transport may be inhibited or induced (as a substrate of P-gp), an inhibitor of P-gp, such as ritonavir, cyclosporine, or verapamil, or an inducer, such as rifampin should be studied. In cases where the drug is also a CYP3A substrate, inhibition should be studied by using a strong inhibitor of both P-gp and CYP3A, such as ritonavir." The evaluation of interactions of investigational drugs as a substrate of other transporters is addressed: "In testing an investigational drug for the possibility that its disposition may be inhibited or induced (i.e., as a substrate of transporters other than or in addition to P-gp), it may be appropriate to use an inhibitor of many transporters (e.g., P-gp, OATP), such as cyclosporine. Recent interactions involving drugs that are substrates for transporters other than or in addition to P-gp include some HMG Co-A reductase inhibitors, rosuvastatin, and pravastatin." However, the draft guidance does not address the possibility that the investigational drug may be an inhibitor of drug transporters other than P-gp.

As was described in the previous (1999) guidance, "the route of administration chosen for a metabolic drug-drug interaction study is important. For an investigational agent, the route of administration should generally be the one planned for clinical use." The guidance makes additional statements that are similar, or identical to, those made in the 1999 guidance. The most pertinent of these statements are:

> Sometimes certain routes of administration can reduce the utility of information from a study. For example, intravenous administration of a substrate drug may not reveal an interaction for substrate drugs where intestinal CYP3A activity markedly alters bioavailability. For an approved agent used either as a substrate or interacting drug, the route of administration will depend on available marketed formulations.

Dose selection for DDI studies is extremely important in order to be able to position the findings of the DDI study:

> For both a substrate (investigational drug or approved drug) and interacting drug (investigational drug or approved drug), testing should maximize the possibility of finding an interaction. For this reason, we recommend that the maximum planned or approved dose and shortest dosing interval of the interacting drug (as inhibitors or inducers) be used. For example, when using ketoconazole as an inhibitor of CYP3A, dosing at 400 mg QD for

> multiple days would be preferable to lower doses. When using rifampin as an inducer, dosing at 600 mg QD for multiple days would be preferable to lower doses. In some instances, doses smaller than those to be used clinically may be recommended for substrates on safety grounds. In such instances, any limitations of the sensitivity of the study to detect the drug-drug interaction due to the use of lower doses should be discussed by the sponsor in the protocol and study report.

The changes in pharmacokinetic parameters that may occur as a result of a DDI study can be used to assess the clinical importance of drug–drug interactions. The draft guidance notes that:

> Interpretation of findings from these studies will be aided by a good understanding of dose/concentration and concentration/response relationships for both desirable and undesirable drug effects in the general population or in specific populations. In certain instances, reliance on endpoints in addition to pharmacokinetic measures/parameters may be useful. Examples include INR measurement (when studying warfarin interactions) or QT interval measurements.

The draft guidance reiterates the 1999 guidance in terms of the selection of PK end points for a DDI study:

> The following measures and parameters of substrate PK should be obtained in every study: (1) exposure measures such as AUC, $C_{max}$, time to $C_{max}$ ($T_{max}$), and others as appropriate; and (2) pharmacokinetic parameters such as clearance, volumes of distribution, and half-lives. In some cases, these measures may be of interest for the inhibitor or inducer as well, notably where the study is assessing possible effects on both study drugs. Additional measures may help in steady state studies (e.g., trough concentration) to demonstrate that dosing strategies were adequate to achieve near steady state before and during the interaction. In certain instances, an understanding of the relationship between dose, blood concentrations, and response may lead to a special interest in certain pharmacokinetic measures and/or parameters. For example, if a clinical outcome is most closely related to peak concentration (e.g., tachycardia with sympathomimetics), $C_{max}$ or another early exposure measure might be most appropriate. Conversely, if the clinical outcome is related more to extent of absorption, AUC would be preferred. The frequency of sampling should be adequate to allow accurate determination of the relevant measures and/or parameters for the parent and metabolites. For the substrate, whether the investigational drug or the approved drug, determination of the pharmacokinetics of important active metabolites is important.

The guidance also notes that pharmacodynamic measures can sometimes provide additional useful information, although in most instances PK measurements will suffice:

> Pharmacodynamic measures may be indicated when a pharmacokinetic/pharmacodynamic relationship for the substrate endpoints of interest is not established or when pharmacodynamic changes do not result solely from pharmacokinetic interactions (e.g., additive effect of quinidine and tricyclic antidepressants on QT interval). In most cases, when an approved drug is studied as a substrate, the pharmacodynamic impact of a given change in blood level ($C_{max}$, AUC) caused by an investigational interaction should be known from other data. If a PK/PD study is needed, it will generally need to be larger than the typical PK study (e.g., a study of QT interval effects).

Sample size and statistical considerations are also addressed in this draft guidance:

> The goal of the interaction study is to determine whether there is any increase or decrease in exposure to the substrate in the presence of the interacting drug. If there is, its implications must be assessed by an understanding of PK/PD relations both for $C_{max}$ and AUC. Results of drug-drug interaction studies should be reported as 90% confidence intervals about the geometric mean ratio of the observed pharmacokinetic measures with (S + I) and without the interacting drug (S alone). Confidence intervals provide an estimate of the distribution of the observed systemic exposure measure ratio of (S + I) versus (S alone) and convey a probability of the magnitude of the interaction. In contrast, tests of significance are not appropriate because small, consistent systemic exposure differences can be statistically significant ($p < 0.05$) but not clinically relevant.
>
> When a drug-drug interaction of potential importance is clearly present (e.g., comparisons indicate twofold (or lower for certain NTR drugs) or greater increments in systemic exposure measures for (S + I)), the sponsor should provide specific recommendations regarding the clinical significance of the interaction based on what is known about the dose-response and/or PK/PD relationship for either the investigational agent or the approved drugs used in the study. For a new drug, the more difficult issue is the impact on the investigational drug as substrate. For inhibition or induction by the investigational drug, the main consequence of a finding will be to add the drug to the list of inhibitors or inducers likely already present in labeling of the older drug. This information can form the basis for reporting study results and for making recommendations in the package insert with respect to either the dose, dosing regimen adjustments, precautions, warnings, or contraindications of the investigational drug or the approved drug. FDA recognizes that dose-response and/or PK/PD information can sometimes be incomplete or unavailable, especially for an older approved drug used as S.
>
> The sponsor may wish to make specific claims in the package insert that no drug-drug interaction of clinical significance occurs. In these instances, it would be helpful for the sponsor to recommend specific no effect boundaries, or clinical equivalence intervals, for a drug-drug interaction. No effect boundaries represent the interval within which a change in a systemic exposure measure is considered not clinically meaningful.

Two approaches for defining no effect boundaries are presented.

In 2008, the FDA published an article that provided an overview of the drug interaction guidances and the steps that the FDA took to revise the original drug interaction guidance documents (Huang *et al.*, 2008). The article also summarizes and highlights updated sections in the current draft guidance (US FDA, 2006g). PhRMA has also made contributions on the subject of the prediction of drug–drug interactions by publishing an article in 2003 that described PhRMA's perspective on the conduct of *in vitro* and *in vivo* DDI studies (Bjornsson *et al.*, 2003). The content of this position paper did not engender the kind of response that was created by PhRMA's MIST document since the FDA and PhRMA seem to be pretty much aligned on the topic of DDIs. More recently, Hewitt *et al.* conducted a survey of the *in vitro* methodologies that are used by industry to predict the potential for the induction of DMEs (Hewitt, de Kanter and LeCluyse, 2007). The survey reviewed the methodologies, and the interpretation of the data generated, for compliance with the FDA's recommendations. The survey showed that, although the basic methods used were similar, no two companies or CROs performed and interpreted the data in the exactly the same manner. They also noted that

no single method was superior to another, but that all methods used enzyme activities as the major end point for the assessment of induction. Currently, the use of human hepatocyte cultures appears to provide the best indication of the *in vivo* induction potential of an NCE. It was concluded that the use of a twofold induction threshold and a percent of positive control criteria may not be the best methods to accurately assess the *in vivo* induction potential of a drug.

## 5.7 Special Populations

The evaluation of the safety and ADME properties of a drug in special populations such as patients with impaired renal or hepatic function, the elderly, or pediatrics is of great importance in the clinical evaluation of a new drug. Therefore, special populations have been the topic of several regulatory guidances.

***5.7.1 Cardiovascular Safety*** The corollary to the guidances on the nonclinical evaluation of cardiovascular safety pharmacology is the ICH guidance, "ICH E14 The Clinical Evaluation Of QT/QTc Interval Prolongation and Proarrhythmic Potential for Non-Antiarrhythmic Drugs" (ICH, 2005e). Among other things, this guidance describes study designs for the evaluation of the potential for QT prolongation. The guidance states "Parallel group studies might be preferred under certain circumstances:

- For drugs with long elimination half-lives for which lengthy time intervals would be required to achieve steady-state or complete washout
- If carryover effects are prominent for other reasons, such as irreversible receptor binding or long-lived active metabolites"

"The timing of the collection of ECGs and the study design (e.g., single or multiple dose, duration) of the 'thorough QT/QTc study' should be guided by the available information about the pharmacokinetic profile of the drug. For drugs with short half-lives and no metabolites, a single dose study might be sufficient. Studies should characterize the effect of a drug on the QT/QTc throughout the dosing interval. While the peak serum concentration does not always correspond to the peak effect on QT/QTc interval, care should be taken to perform ECG recordings at time points around the $C_{max}$." As noted in previous guidance, the exact words used in the guidance may not be critical. For example, this guidance mentions "serum concentrations" but does not necessarily advocate that measurements be made on serum samples rather than on plasma samples.

Although one would not suspect a prominent role of drug metabolism in this guidance based on the title of the guidance, drug metabolism does indeed play a significant role in this guidance. However, this is not surprising due to the role of drug–drug interactions in unexpected sudden cardiac death in a small number of overtly healthy patients who were taking the antihistamine drug, terfenadine (Seldane) (Monahan *et al.*, 1990). It was recognized that terfenadine alone could have a modest QT prolonging effect, but when taken together with certain antifungal agents that inhibit the hepatic metabolism of terfenadine, toxic levels of terfenadine could develop with marked QT prolongation and an increased probability for life-threatening ventricular tachyarrhythmias. Within a few years of these findings, terfenadine was removed from the marketplace and replaced with its metabolite, fexofenadine (Allegra), a drug with good antihistamine properties but no QT prolonging effects.

The guidance states:

> An adequate drug development program should ensure that the dose-response and generally the concentration-response relationship for QT/QTc prolongation have been characterized, including exploration of concentrations that are higher than those achieved following the anticipated therapeutic doses. Data on the drug concentrations around the time of ECG assessment would aid this assessment. If not precluded by considerations of safety or tolerability due to adverse effects, the drug should be tested at substantial multiples of the anticipated maximum therapeutic exposure. Alternatively, if the concentrations of a drug can be increased by drug-drug or drug-food interactions involving metabolizing enzymes (e.g., CYP3A4, CYP2D6) or transporters (e.g., P-glycoprotein), these effects could be studied under conditions of maximum inhibition. This approach calls for an understanding of the pharmacokinetic and pharmacodynamic properties of the parent and significant human metabolites. In general, the duration of dosing or dosing regimen should be sufficient to characterize the effects of the drug and its active metabolites at relevant concentrations.

This approach to evaluating the safety of high levels of drug is an interesting one and may be applicable to drugs that are extensively metabolized. However, the extent of use of such an approach is not yet known. Although not addressed by the guidance, the use of inducers to increase the exposure to metabolites could be required to assess the potential of such metabolites to prolong the QT interval.

Metabolism is further mentioned in this guidance in the following statement: "If the 'thorough QT/QTc study' is positive, analyses of the ECG and adverse event data from certain patient sub-groups are of particular interest, such as:

- Patients with electrolyte abnormalities (e.g., hypokalemia);
- Patients with congestive heart failure
- Patients with impaired drug metabolizing capacity or clearance (e.g., renal or hepatic impairment, drug interactions)
- Female patients
- Patients aged <16 and over 65 years"

The potential for QT prolongation is routinely evaluated as part of the single and multiple ascending dose clinical trails that are conducted in normal healthy volunteers. This initial evaluation should be followed up by a "thorough QT/QTc study" which is intended to determine whether the drug has a threshold pharmacologic effect on cardiac repolarization, as detected by QT/QTc prolongation. The guidance states that:

> The threshold level of regulatory concern, discussed further below, is around 5 ms as evidenced by an upper bound of the 95% confidence interval around the mean effect on QTc of 10 ms. The study is typically carried out in healthy volunteers (as opposed to individuals at increased risk of arrhythmias) and is used to determine whether or not the effect of a drug on the QT/QTc interval in target patient populations should be studied intensively during later stages of drug development. It is not intended to identify drugs as being pro-arrhythmic. Although data are limited, it is not expected that the results of the "thorough QT/QTc study" would be affected by ethnic factors.
>
> The "thorough QT/QTc study" would typically be conducted early in clinical development to provide maximum guidance for later trials, although the precise timing will depend on the specifics of the drug under development. It would usually not be the first study, as it is

important to have basic clinical data for its design and conduct, including tolerability and pharmacokinetics. Some drugs might not be suitable for study in healthy volunteers because of issues related to tolerability (e.g., neuroleptic agents, chemotherapeutics). The results of the "thorough QT/QTc study" will influence the amount of information collected in later stages of development:

- A negative "thorough QT/QTc study" will almost always allow the collection of on-therapy ECGs in accordance with the current practices in each therapeutic area to constitute sufficient evaluation during subsequent stages of drug development
- A positive "thorough QT/QTc study" will almost always call for an expanded ECG safety evaluation during later stages of drug development.

There could be very unusual cases in which the "thorough QT/QTc study" is negative but the available nonclinical data are strongly positive (e.g., hERG (human ether-a-go-go-related gene) positive at low concentrations and in vivo animal model results that are strongly positive). If this discrepancy cannot be explained by other data, and the drug is in a class of pharmacological concern, expanded ECG safety evaluation during later stages of drug development might be appropriate.

These statements have been included here since they raise the timing of the conduct of the "thorough QT/QTc study." The guidance clearly advocates that this study should be conducted early in clinical development and implies that extensive QT monitoring may be required in clinical trials if there are no data to support the assumption that the risk for QT prolongation is low (i.e. absence of evidence is not evidence of absence). Since extensive QT monitoring in a large clinical trial is expensive and logistically challenging, the sponsor should strongly consider conducting the "thorough QT/QTc study" in the early stages of the clinical program. Since one of the goals of this study is to determine a concentration–effect relationship, the sponsor should ensure that they have identified the correct analytes to be measured. Therefore, it may be necessary to identify the major metabolites in humans prior to conducting the "thorough QT/QTc study." This would imply that the human ADME study should also be conducted in early clinical development. However, arguments can be put forth for conducting the "thorough QT/QTc study" at a later stage of development in order to ensure that an appropriate concentration range has been identified and in order to avoid incurring the very high costs of the "thorough QT/QTc study" until the sponsor has some evidence of efficacy in humans. The E14 guidance has been reviewed by Darpo *et al.* (Darpo, Nebout and Sager, 2006).

Another challenge that the sponsor will face when conducting a "thorough QT/QTc study" is that the regulatory agencies will want to determine the concentration–effect relationship, regardless of the relationship of the concentration that causes an effect to the concentrations associated with efficacy. Such knowledge will be important for generating an understanding of high concentrations of drug (e.g. due to polymorphisms in metabolism; DDIs or overdose) on QT prolongation. However, the sponsors' goal may be to show that their drug does not cause QT prolongation at a certain concentration that is a certain fold greater than the $C_{max}$ at the predicted efficacious dose. The sponsor may be reluctant to push to an effect since once such data have been generated, the stigma of QT prolongation will forever be associated with the drug, regardless of the therapeutic index of the drug. Therefore, it is clear that there may be a gap between the goals of the regulatory agencies and the sponsors.

The issue of drug–drug interactions has generated significant concern within the pharmaceutical industry and among regulatory authorities since the 1990s. This has

arisen with respect to early termination of clinical development (e.g. furafylline), refusal of approval (e.g. mibefradil in Sweden), severe prescribing restrictions and withdrawal from the market (e.g. sorivudine, terfenadine, mibefradil, astemizole, and cisapride), and threatened litigation.

***5.7.2 Oncology*** The EMEA issued a guideline in 2006 entitled "Guideline on the Evaluation of Anticancer Medicinal Products in Man" (EMEA, 2006f). This guideline replaces "NfG on Evaluation of Anticancer Medicinal Products in Man (CPMP/EWP/205/95, Rev. 2)" (EMEA, 2003). This guideline states that "Consideration should be given to study high-risk patients (e.g. high risk with respect to target organ toxicity or compromised metabolic or excretory mechanisms for the experimental compound) separately." Although organ impairment is covered in other guidances or guidelines, this subject is dwelt upon in this guideline:

> Studies in patients with decompensated liver function are rarely indicated, but patients with liver metastases should normally be included in the development program. For compounds metabolized by the liver, PK studies are expected, exploring the relationship between, e.g. enzyme levels, or bilirubin increase and exposure. For compounds developed for use in late line therapies, the need for dose reductions in patients with impaired bone marrow reserve due to prior chemo/radio therapy may need special attention. Exploratory studies, including PK, in patients with malignant ascites or other third space conditions are encouraged. Effects of renal impairment should be studied as appropriate. If justified by the target indication, the studies referred to above may be conducted post licensing.

The latter point is a pertinent one that is discussed later in this chapter. It should also be noted that many cancer patients receive combinations of drugs. These combinations may be for the treatment of their disease, for the treatment of comorbidities, or as prophylactic treatment for the side effects of the therapeutic treatment. Riechelmann *et al.* recently reported that the average number of drugs prescribed to a cancer patient was 5, with a range of 0–23 (Riechelmann *et al.*, 2005, 2007; Riechelmann and Saad, 2006). This polypharmacy obviously will have an impact on the design and interpretation of drug–drug interaction studies. Haddad *et al.* have also recently reviewed the role of CYP3A4 in DDIs in cancer patients (Haddad, Davis and Langman, 2007). The integral role of CYP activity on toxicity and efficacy of anticancer drugs in patients has been shown by Li *et al.* (2006), who have used CYP3A phenotyping to predict the exposure to EGFR tyrosine kinase inhibitors and Alexandre *et al.* (2007), who have shown a link between CYP3A activity and docetaxel-induced febrile neutropenia due to cancer-related inflammation. Transcriptional repression of hepatic CYP3A4 expression by cancer has also been reported by Charles *et al.* (2006), who determined that the repression of CYP3A4 expression was due to cancer-associated inflammation. They noted that targeted therapy to reduce inflammation may provide clinical benefit for chemotherapy drugs metabolized by CYP3A4 by improving their PK profile. Since many chemotherapeutic drugs have a narrow therapeutic index, it may be necessary to "personalize" the dosing for each subject by using therapeutic drug monitoring (TDM) and dose adaptation. This topic has been reviewed by Dumez *et al.* (2004) and de Jonge *et al.* (2005).

An "Addendum on Paediatric Oncology" that was issued to CPMP/EWP/205/95, Rev. 2 (i.e. the version of the guideline "Evaluation of Anticancer Medicinal Products in Man" that was updated in the 2006 guideline) was issued in 2004 (EMEA, 2003). It is curious to note that this addendum is listed on the EMEA guidance page, although

the guideline that it is an addendum to is no longer listed (EMEA, 2008h). It is unclear as to why the addendum was not included in the updated guideline on "Evaluation of Anticancer Medicinal Products in Man" that was issued in 2006 (EMEA, 2006f). This guideline illustrates the point that regulatory documents undergo periodic updating and that the drug development scientist needs to be continually aware of the changes of any updates or revisions to guidances as well as keeping abreast of the guidances that are under development or being planned.

***5.7.3 Rheumatoid Arthritis*** The FDA's guidance on "Clinical Development Programs for Drugs, Devices, and Biological Products for the Treatment of Rheumatoid Arthritis (RA)" that was issued in 1999, introduces the timing of the conduct of drug–drug interaction studies (US FDA, 1999b). This guidance states:

> Before starting phase 3 trials, an evaluation of the test product's interaction with other agents likely to be used by the target population should be performed. Initial information can be established based on metabolic pathways, studies of in vitro systems, animal or human pharmacology studies, or drug interaction studies. This type of information helps in directing areas in need of clinical evaluation. When products are intended to be tested as combination therapy with the investigational agent, substantial information on interactions and safety of co-administration should be developed in phase 2.

The guidance clearly implies that the relevant DDI studies should be performed before the initiation of phase 3 clinical trials. The guidance also reminds the reader that DDIs may not only arise as a consequence of metabolic interactions but potentially as a consequence of displacement of a drug from plasma proteins. The guidance states: "Because polypharmacy is common during the treatment of rheumatic disorders, in vitro binding studies with blood from patients with active disease should be used as a preliminary screening tool for potential displacement reactions."

Many reports in the biomedical literature have documented that hepatic disease can alter the absorption and disposition of drugs (PK) as well as their efficacy and safety (PD). These reports have been based on studies in patients with common hepatic diseases, such as alcoholic liver disease and chronic infections with hepatitis viruses B and C, and less common diseases, such as acute hepatitis D or E, primary biliary cirrhosis, primary sclerosing cholangitis, and alpha1-antitrypsin deficiency. Liver disease may also alter kidney function, which can lead to accumulation of a drug and its metabolites even when the liver is not primarily responsible for elimination. Liver disease may also alter PD effects (e.g. increased encephalopathy with certain drugs in patients with hepatic failure). The specific impact of any disease on hepatic function is often poorly described and highly variable, particularly with regard to effects on the PK and PD of a drug.

***5.7.4 Diabetes*** The draft FDA guidance on "Diabetes Mellitus: Developing Drugs and Therapeutic Biologics for Treatment and Prevention" was recently issued (February 2008) (US FDA, 2008i). This draft guidance has been in development since 2004. It addresses the pharmacokinetics of drugs for the treatment of diabetes mellitus by stating that:

> In general, pharmacokinetic parameters of noninsulin therapeutics should be evaluated in phase 1 studies. These studies can be performed in healthy volunteers to determine the basic pharmacokinetic parameters (e.g., absolute bioavailability, area under the curve (AUC), $C_{max}$, $T_{max}$, $T_{1/2}$). Additionally, pharmacokinetic studies also may be appropriate in

> the intended patient population. We recommend that exposure-response data be obtained during the phase 2 dose-finding studies. (See the guidance for industry *Exposure-Response Relationships: Study Design, Data Analysis, and Regulatory Applications*.)
>
> In patients with diabetes, the high prevalence of altered glomerular filtration rates, delayed or deficient gastrointestinal transit and absorption, and the potential for interactions with commonly used medications usually dictate the need for the evaluation of the pharmacokinetics of new agents in the target population, beyond investigations in healthy volunteers. It is important to evaluate the in vivo and in vitro mechanisms of drug absorption and disposition. This information will provide the basis for the design of the drug interaction studies addressing the class effects of oral antidiabetic drugs (e.g., addressing the induction potential of CYP enzymes by thiazolidinediones, CYP2C-based interactions with sulfonylureas, and interactions with renal tubular secretion of metformin). We also recommend interaction studies with drugs that have a narrow therapeutic index and with drugs likely to be co-administered in the diabetic population. (See the draft guidance for industry *Drug Interaction Studies—Study Design, Data Analysis, and Implications for Dosing and Labeling* for details.) Effects of food on pharmacokinetics should be evaluated in the development of therapeutic products that are intended to be administered orally in temporal proximity to meals (e.g., agents designed to exert effects on glycemia peri- or postprandially, such as meglitinides). Because patients with diabetes may be a particularly sensitive population in terms of polypharmacy and underlying, often subclinical, cardiac disease, we also encourage sponsors to address the effect of the drug on the QT interval by conducting a thorough QT study.

The statements in this draft guidance appear to be consistent with existing guidances and best practices.

***5.7.5 Hepatic Impairment*** In 2003, the FDA issued a guidance called "Pharmacokinetics in Patients with Impaired Hepatic Function: Study Design, Data Analysis, and Impact on Dosing and Labeling" (US FDA, 2003e). This guidance provides a general background on the subject of hepatic impairment, including how it is defined and measured. However, it notes that:

> Despite extensive efforts, no single measure or group of measures has gained widespread clinical use to allow estimation in a given patient of how hepatic impairment will affect the PK and/or PD of a drug. Even though clinically useful measures of hepatic function to predict drug PK and PD are not generally available, clinical studies in patients with hepatic impairment, usually performed during drug development, can provide information that may help guide initial dosing in patients. This information can be appropriately used with the understanding that careful observation and dose titration are critical to achieve the optima.

This document provides guidance on deciding whether to conduct a study in patients with impaired hepatic function. This guidance recommends a PK study in patients with impaired hepatic function if:

> hepatic metabolism and/or excretion accounts for a substantial portion (>20 percent of the absorbed drug) of the elimination of a parent drug or active metabolite. The guidance also recommends a hepatic impairment study even if the drug and/or active metabolite is eliminated to a lesser extent (<20 percent), if its labeling or literature sources suggest that it is a narrow therapeutic range drug. If the metabolism of the drug is unknown and other information is lacking to suggest that hepatic elimination routes are minor, the Agency recommends that the drug be considered extensively metabolized.

Conversely, for some drugs, hepatic functional impairment is not likely to alter PK sufficiently to require dosage adjustment. The guidance notes that, "In such cases, a study to confirm the prediction is generally not important. The following drug properties may support this conclusion:

- The drug is excreted entirely via renal routes of elimination with no involvement of the liver.
- The drug is metabolized in the liver to a small extent (<20 percent), and the therapeutic range of the drug is wide, so that modest impairment of hepatic clearance will not lead to toxicity of the drug directly or by increasing its interaction with other drugs.
- The drug is gaseous or volatile, and the drug and its active metabolites are primarily eliminated via the lungs."

In an additional caveat, the guidance states that "For drugs intended only for single-dose administration, a hepatic impairment study will generally not be useful, unless clinical concerns suggest otherwise."

The design of the study is addressed in the guidance which notes that the study can be a single-dose or multiple-dose study with PK assessment of the parent drug and any active metabolite(s). In a multiple-dose study, PK assessment is appropriately carried out at steady state. The guidance notes that:

> A single-dose study may be satisfactory for cases where prior evidence indicates that multiple-dose PK is accurately predicted by single-dose data for both parent drug and active metabolites. This would be the case when the drug and active metabolites exhibit linear and time-independent PK at the concentrations anticipated in the patients to be studied. A multiple-dose study is desirable when the drug or an active metabolite is known to exhibit nonlinear or time-dependent PK.

The mention of the measurement of metabolites clearly indicates that the sponsor should have identified the important metabolites prior to the conduct of a study in patients with hepatic impairment. This guidance appears to use the phrase "active metabolite" to mean metabolites that may have either pharmacological or toxicological activity.

The guidance provides some quite detailed information on sample collection and data analysis:

> The blood sampling duration should be adequate to determine the terminal half-life of the drug and its active metabolite(s), with the expectation that these times may be extended in the patient compared to the control population. For drugs that are highly extracted by the liver (extraction ratio > 0.7) and that are extensively bound to plasma proteins (fraction unbound < 10 percent), the Agency recommends that the unbound fraction be determined at least at trough and maximum plasma concentration. The clearance and volume parameters are appropriately expressed in terms of both unbound and total concentrations of drug in plasma/serum/blood.

The EMEA has also issued a guideline on this topic, "Guideline on the Evaluation of the Pharmacokinetics of Medicinal Products in Patients with Impaired Hepatic Function" (EMEA, 2005d). This guideline was issued in August 2005. Like the FDA guidance, this guideline provides some background information on hepatic function

and hepatic disease, albeit with a different level of detail and emphasis. The guideline states:

> Hepatic function decreases with age, but due to the high capacity of the liver this is considered not to change the pharmacokinetics to a clinically relevant extent. Liver disease, however, is known to be a common cause of altered pharmacokinetics of drugs. Hepatic function can be decreased through different pathophysiological mechanisms. Worldwide, chronic infections with hepatitis B or C are the most common causes of chronic liver disease, whereas in the western world, chronic and excessive alcohol ingestion is one of the major causes of liver disease. Other causes are uncommon diseases such as primary biliary cirrhosis, primary sclerosing cholangitis and autoimmune chronic active hepatitis. Ongoing destruction of the liver parenchyma in chronic liver diseases ultimately leads to liver cirrhosis and the development of portal hypertension. However, even if liver cirrhosis is established, the residual metabolic function of the liver may be rather well preserved for many years because of regeneration of hepatocytes. Clinical symptoms related to hepatocellular failure and portal hypertensions are most importantly ascites, oesophageal varices and encephalopathy. Serum markers of liver failure are low serum albumin and a prothrombin deficiency. Serum bilirubin as well as other liver tests may or may not be affected to a varying degree, e.g. depending on the liver disease (cholestatic versus hepatocellular). Liver cirrhosis is irreversible in nature, but progression can be modified by e.g. abstinence of alcohol in alcohol liver cirrhosis. The pharmacokinetics and pharmacodynamics of drugs may be altered by liver disease through different mechanisms. The effect most often depends on the severity of hepatic impairment. The effects on pharmacokinetics can be difficult to predict due to consequences of shunting of blood past the liver (both portosystemic and intra-hepatic), impaired hepatocellular function, impaired biliary excretion and decreased protein binding. Factors that influence the need for pharmacokinetic data in patients with hepatic impairment, and interpretation of these data, are the intended use of the drug, pharmacokinetic characteristic features in otherwise healthy individuals and PK/PD relationships. Based on this, the major concern (side effects or lack of efficacy) should be identified.

The guideline provides some input into when to perform pharmacokinetic studies in patients with impaired hepatic function. It states: "A pharmacokinetic study in subjects with impaired hepatic function is recommended when:

- The drug is likely to be used in patients with impaired hepatic function and
- Hepatic impairment is likely to significantly alter the pharmacokinetics (especially metabolism and biliary excretion) of the drug and/or its active metabolites and a posology adjustment may be needed for such patients taking into account the PK/PD relationship.

Furthermore:

> If no study is performed in patients with hepatic impairment, a justification should be given. Lack of data may be justified if the drug is not intended to be used in patients with hepatic impairment. If the drug is likely to be used in these patients, the applicant should discuss the potential for hepatic impairment to influence the pharmacokinetics (of parent drug, active and "inactive" metabolites) and should include relevant information in the SPC. Lack of data may lead to restriction in the use of the drug (not only warnings but also contraindications).

The guidance provides input into methodologies to be used for the confirmation that the subjects to be studied actually have an impaired metabolic capacity. It may be useful to administer, for instance, a CYP3A4 probe drug (if the drug under investigation is a CYP3A4 substrate) to the subjects to be included to observe if the pharmacokinetics of the probe drug is altered (like a "positive control" known to be specially sensitive to liver impairment). This probe would have to be sensitive enough to identify a range of severity in hepatic dysfunction. The guideline also describes other approaches:

> Exogenous markers that have been used to assess different hepatic drug elimination mechanisms are antipyrine, MEGX (lidocaine metabolite), ICG (indocyanine green) and galactose. Such markers may be used in parallel with the Child-Pugh classification and a justification for the choice of marker(s) should be given.
>
> In conclusion, until optimal markers have been found, the Child-Pugh classification system can be used to categorise the degree of hepatic impairment of subjects included in a pharmacokinetic study and can, together with its individual components, be used when evaluating the pharmacokinetic results. The sponsor should submit all individual scores of the subjects included in the study, as well as other information on subjects characteristics, for example results of standard laboratory tests.

The collection and analysis of samples is also addressed:

> "Plasma (or whole blood, as appropriate) samples should be analysed for parent drug and any metabolites with known or suspected activity (therapeutic or adverse). Metabolites, identified as toxic in preclinical studies, which could be affected by hepatic function should be evaluated. Also, metabolites that are considered relatively inactive in patients with normal hepatic function may reach active/toxic levels if the accumulation of the metabolites is substantial. Hence, evaluation of such metabolites should be considered. The frequency and duration of plasma sampling should be sufficient to accurately estimate relevant pharmacokinetic parameters for parent drug and metabolites.
>
> If the drug or metabolites exhibit a high extent of plasma protein binding, the pharmacokinetics should be described and analysed with respect to the unbound concentrations of the drug and active metabolites in addition to total concentration.
>
> For chiral drugs, the analysis of the enantiomers should be considered as the metabolic profile for each enantiomer may be different in subjects with hepatic impairment. … Consideration should also be given to possible consequences of altered importance of other elimination pathways and the interaction with concomitantly administered drugs. For prodrugs (i.e., drugs with activity predominantly due to a hepatically generated metabolite), the plasma levels of the active substance may be decreased in patients with hepatic impairment and adjustments of the dose and/or dosing interval may be needed."

The importance of dose adaptation for liver disease in patients with antineoplastic drugs has been reviewed by Tchambaz *et al.* (2006). Dose adaptation may be required due to the high prevalence of impaired liver function in cancer patients and the narrow therapeutic index and the potential for serious adverse events with antineoplastic drugs. Recently, Nakai *et al.* (2008) have shown a correlation of decreased expression of certain CYP isoforms and drug transporters with the progression of liver fibrosis in chronic hepatitis C patients.

***5.7.6 Renal Impairment*** The effect of renal function on the pharmacokinetics of a drug has been addressed by the FDA in a guidance issued in 1998, entitled "Pharmacokinetics in Patients with Impaired Renal Function—Study Design, Data Analysis, and Impact on Dosing and Labeling" (US FDA, 1998b).

The format and content of this guidance is analogous to the FDA's guidance on the effect of hepatic impairment on the pharmacokinetics of a drug. The guidance provides some background information:

> Although the most obvious type of change arising from renal impairment is a decrease in renal excretion, or possibly renal metabolism, of a drug or its metabolites, renal impairment has also been associated with other changes, such as changes in absorption, hepatic metabolism, plasma protein binding, and drug distribution. These changes may be particularly prominent in patients with severely impaired renal function and have been observed even when the renal route is not the primary route of elimination of a drug. Thus, for most drugs that are likely to be administered to patients with renal impairment, PK characterization should be assessed in patients with renal impairment to provide rational dosing recommendations.

The guidance provides suggestions on when studies in patients with impaired renal function may be important:

> A PK study in patients with impaired renal function is recommended when the drug is likely to be used in these patients and (1) renal impairment is likely to significantly alter the PK of a drug and/or its active/toxic metabolites and (2) a dosage adjustment is likely to be necessary for safe and effective use in such patients. In particular, a study in patients with impaired renal function is recommended when the drug or its active metabolites exhibit a narrow therapeutic index and when excretion and/or metabolism occurs primarily via renal mechanisms (excretion or metabolism). A study also should be considered when a drug or an active metabolite exhibits a combination of high hepatic clearance (relative to hepatic blood flow) and significant plasma protein binding. In this setting, renal impairment could induce a significant increase in the unbound concentrations after parenteral administration due to a decreased plasma protein binding coupled with little or no change in the total clearance (decrease in unbound clearance).

Conversely, such studies may not be important when, "For some drugs, renal impairment is not likely to alter PK enough to justify dosage adjustment. In such cases, a study to confirm that prediction may be helpful but is not necessary. If a study is not conducted, the labeling should indicate that the impact of renal impairment was not studied, but that an effect requiring dosage adjustment is unlikely to be present. Current knowledge suggests that the following drug properties may justify this approach:

- Drug and active metabolites with a relatively wide therapeutic index and that are primarily eliminated via hepatic metabolism or biliary excretion;
- Gaseous or volatile drug and active metabolites that are primarily eliminated via the lungs;
- Drugs intended only for single-dose administration unless clinical concerns dictate otherwise.

Controversy exists regarding the impact of severe renal impairment on hepatic metabolism. For this reason, a renal impairment study is still considered desirable for

a drug eliminated primarily via hepatic metabolism unless it also has a relatively wide therapeutic index."

The guidance provides input on the design of the study:

> A single-dose study is satisfactory for cases where there is clear prior evidence that the multiple-dose PK is accurately predictable from single-dose data for all chemical species of interest (drug and potentially active metabolites). A multiple-dose PK is predictable from a single-dose PK when the drug and active metabolites exhibit linear and time-independent PK at the concentrations anticipated in the patients to be studied. A multiple-dose study is desirable when the drug or an active metabolite is known to exhibit nonlinear or time-dependent PK. … For multiple-dose studies, lower or less frequent doses as renal function decreases may be important to prevent accumulation of drug and metabolites to unsafe levels. The dosage regimen may be adjusted based on the best available prestudy estimates of the PK of the drug and its active metabolites in patients with impaired renal function.

As was the case with the guidance on studies in hepatically impaired patients, this guidance provides input into sample collection, analysis, and data interpretation:

> Plasma or whole blood, if appropriate, (and optionally urine) samples should be analyzed for parent drug and any metabolites with known or suspected activity (therapeutic or adverse). This is particularly important in patients with impaired renal function since renally excreted metabolites can accumulate to a much higher degree in such patients. The frequency and duration of plasma sampling and urine collection should be sufficient to accurately estimate the relevant pharmacokinetic parameters for the parent drug and its active metabolites.
>
> Plasma protein binding is often altered in patients with impaired renal function. For systemically active drugs and metabolites, the unbound concentrations are generally believed to determine the rate and extent of delivery to the sites of action. This leads to the recommendation that the PK should be described and analyzed with respect to the unbound concentrations of the drug and active metabolites. Although unbound concentrations should be measured in each plasma sample, if the binding is concentration-independent and is unaffected by metabolites or other time-varying factors, the fraction unbound may be determined using a limited number of samples or even a single sample from each patient. The unbound concentration in each sample is then estimated by multiplying the total concentration by the fraction unbound for the individual patient. For drugs and metabolites with a relatively low extent of plasma protein binding (e.g., extent of binding less than 80%), alterations in binding due to impaired renal function are small in relative terms. In such cases, description and analysis of the PK in terms of total concentrations should be sufficient.

It should be noted that the direct measurement of unbound drug concentrations is rarely performed and can be extremely expensive. It should be noted that the guidance also mentions that it may be appropriate to measure "potentially active metabolites." This seems to imply that if one knew that a metabolite existed, but did not know whether it was active or not, that metabolite should be quantitated. This statement appears to be out of context of the exposure to this metabolite. The guidance does not appear to address the potential impact of metabolic genotype on the exposure to parent drug and/or metabolites in patients with renal impairment.

Patients with impaired renal function may undergo dialysis as a treatment for their impaired renal function. Dialysis may significantly affect the PK of a drug to an extent

that dosage adjustment is appropriate. The guidance states: "The need for dosage adjustment results when a significant fraction of the drug or active metabolites in the body is removed by the dialysis process. In such cases, a change in the dosage regimen, such as a supplemental dose following the dialysis procedure, may be required." The guidance further states that:

> In general, a study of the effect of dialysis on PK may be omitted if the dialysis procedure is unlikely to result in significant elimination of drug or active metabolites. This is arguable for drugs and active metabolites that have a large unbound volume of distribution ($V_u$)or a large unbound nonrenal clearance ($CL_{NR,u}$).
>
> If the drug and metabolites have a large unbound volume of distribution, only a small fraction of the amounts in the body will be removed by dialysis. For example, if $V_u$ were greater than 360 L, less than 10 percent of the amount initially in the body could be removed by 3 hours of high flux hemodialysis with an unbound dialysis clearance of 200 mL/min.
>
> If the drug and metabolites have a large unbound nonrenal clearance, dialysis contributes a relatively small amount to the overall unbound clearance. For example, if $CL_{NR,u}$ were greater than 125 mL/min, 3 hours of high-flux hemodialysis with an unbound dialysis clearance of 200 mL/min administered every 2 days would contribute less than 10 percent to the overall clearance.

Unlike the guidance on the evaluation in hepatically impaired patients, this guidance emphasizes the importance of pharmacodynamics. The guidance states:

> Whenever appropriate, pharmacodynamic assessment should be included in the studies of renal impairment. The selection of the pharmacodynamic endpoints should be discussed with the appropriate FDA review staff and should be based on the pharmacological characteristics of the drug and metabolites (e.g., extent of protein binding, therapeutic index, and the behavior of other drugs in the same class in patients with renal impairment).

The Data Analysis section of the guidance states that "the primary intent of the data analysis is to assess whether dosage adjustment is required for patients with impaired renal function, and, if so, to develop dosing recommendations for such patients based on measures of renal function. The data analysis typically consists of the following steps:

- Estimation of PK parameters;
- Mathematical modeling of the relationship between measures of renal function and the PK parameters;
- Development of dosing recommendations including an assessment of whether dosage adjustment is warranted in patients with impaired renal function."

Each of these points is further elaborated upon. For example, the guidance states that for Parameter Estimation:

> Plasma concentration data (and urinary excretion data if collected) should be analyzed to estimate various parameters describing the PK of the drug and its active metabolites. The PK parameters of a drug can include the area under the plasma concentration curve (AUC), peak concentration ($C_{max}$), apparent clearance (CL/F), renal clearance (CLR), apparent volume of distribution ($V_{z/F}$ or $V_{ss/F}$), terminal half-life ($t_{1/2}$). The PK parameters

> of active metabolites can include the area under the plasma concentration curve (AUC), peak concentration ($C_{max}$), terminal half-life ($t_{1/2}$). If possible, parameters are preferably expressed in terms of unbound concentrations; for example, apparent clearance relative to the unbound drug concentrations ($CL_{u/F} = D/AUC_u$) where the subscript 'u' indicates unbound drug. Noncompartmental and/or compartmental modeling approaches to parameter estimation can be employed.

The modeling of the relationship between renal function and PK is described thus:

> The objective of this step is to construct mathematical models for the relationships between the RF, the measures of renal function, particularly creatinine clearance ($CL_{cr}$) and relevant PK parameters. The PK parameters of greatest interest are usually the apparent unbound clearance ($CL_u/F$), or the dose-normalized area under the unbound concentration curve ($AUC_u/D$), and the dose-normalized peak unbound concentration ($C_{max,u}/D$) for the drug and active metabolites. The intended result is a model that can successfully predict K behavior given information about renal function.

Further details of this approach are given in the guidance but are not presented here.

Finally, the guidance addresses the development of dosing recommendations:

> Specific dosing recommendations should be constructed based on the study results using the aforementioned model for the relationships between RF and relevant PK parameters. Typically the dose is adjusted to produce a comparable range of unbound plasma concentrations of drug or active metabolites in both normal patients and patients with impaired renal function. Simulations are encouraged as a means to identify doses and dosing intervals that achieve that goal for patients with different levels of renal function.
>
> For some drugs, even severe renal impairment may not alter PK sufficiently to warrant dosage adjustment. A sponsor could make this claim by providing an analysis of the study data to show that the PK measurements most relevant to therapeutic outcome in patients with severe renal impairment are similar, or equivalent, to those in patients with normal renal function.

The guidance then proceeds to describe how prior to the conduct of the studies, specific "no effect" boundaries for the ratio of a PK measurement from patients with severe and normal renal functions respectively can be used.

***5.7.7 Geriatrics*** ICHE7, "Studies in Support of Special Populations: Geriatrics" was issued in 1994 (ICH, 1994c). The vast majority of this guidance was devoted to pharmacokinetics, pharmacodynamics, and DDIs. On the subject of pharmacokinetics, the guidance states that:

> Most of the recognized important differences between younger and older patients have been pharmacokinetic differences, often related to impairment of excretory (renal or hepatic) function or to drug-drug interactions. It is important to determine whether or not the pharmacokinetic behavior of the drug in elderly subjects or patients is different from that in younger adults and to characterize the effects of influences, such as abnormal renal or hepatic function, that are more common in the elderly even though they can occur in any age group. Information regarding age-related differences in the pharmacokinetics of

the drug can come, at the sponsor's option, either from a Pharmacokinetic Screen (as described subsequently) or from formal pharmacokinetic studies, in the elderly and in patients with excretory functional impairment. It is recognized that for certain drugs and applications (e.g., some topically applied agents, some proteins) technical limitations such as low systemic drug levels may preclude or limit exploration of age-related pharmacokinetic differences.

Formal Pharmacokinetic Studies are described as:

Formal PK studies can be done either in healthy geriatric subjects or in patient volunteers with the disease to be treated by the drug.

The initial PK study can be a pilot trial of limited size conducted under steady-state conditions to look for sizable differences between older and younger subjects or patients. A larger, single-dose PK study of sufficient size to permit statistical comparisons between geriatric and younger subjects' or patients' pharmacokinetic profiles is also acceptable. In either case, if large (i.e., potentially medically important) age-related differences are found, the initial PK study may need to be followed by a multiple-dose PK study of sufficient size to permit statistical comparisons (geriatric vs. younger) at steady-state.

The Pharmacokinetic Screening Approach is also addressed:

Sponsors may opt, instead of conducting a separate PK evaluation of the elderly, to utilize a pharmacokinetic screen in conjunction with the main Phase 3 (and Phase 2, if the sponsor wishes) clinical trials program. This screening procedure involves obtaining, under steady-state conditions, a small number (one or two) of drug blood level determinations at "trough" (i.e., just prior to the next dose) or other defined times from sufficient numbers of Phase 2/3 clinical trials patients, geriatric and younger, to detect age-associated differences in pharmacokinetic behavior, if they are present. It is important to record time of dosing prior to blood concentration measurements, and relation of dosing to meals, and to examine the influence of demographic and disease factors, such as gender renal function, presence of liver disease, gastrointestinal disease or heart disease, body size and composition, and concomitant illnesses.

Small differences are unlikely to be of medical importance. Where the screen detects large differences, formal pharmacokinetic studies may be indicated unless the screen's results are sufficiently informative.

The advantage of a pharmacokinetic screen is that it can assess the effects, not only of age itself, but also of other factors associated with age (altered body composition, other drugs, concomitant illness) and their interactions.

The topic of renally or hepatically impaired geriatric patients is also addressed. The guidance states:

Renal impairment is an aging-associated finding that can also occur in younger patients. Therefore, it is a general principle, not specific to these guidelines, that drugs excreted (parent drug or active metabolites) significantly through renal mechanisms should be studied to define the effects of altered renal function on their pharmacokinetics. Such information is needed for drugs that are the subject of this guideline but it can be obtained in younger subjects with renal impairment.

Similarly, drugs subject to significant hepatic metabolism and/or excretion, or that have active metabolites, may pose special problems in the elderly. Pharmacokinetic studies should be carried out in hepatically impaired young or elderly patient volunteers.

If a Pharmacokinetic Screen approach is chosen by the sponsor (Section VI, see above), and if patients with documented renal impairment or hepatic impairment (depending on the drug's elimination pattern) are included and the results indicate no medically important pharmacokinetic difference, that information may be sufficient to meet this Geriatric Guideline's purpose.

Pharmacodynamics is also addressed. The guidance notes that "The number of age-related pharmacodynamic differences (i.e., increased or decreased therapeutic response, or side effects, at a given plasma concentration of drug) discovered to date is too small to necessitate dose response or other pharmacodynamic studies in geriatric patients as a routine requirement. Separate studies are, however, recommended in the following situations:

- Sedative/hypnotic agents and other psychoactive drugs or drugs with important CNS effects, such as sedating antihistamines;
- Where subgroup comparisons (geriatric versus younger) in the Phase 2/3 clinical trials database indicate potentially medically significant age-associated differences in the drug's effectiveness or adverse reaction profile, not explainable by PK differences."

Drug–drug interactions are of particular importance to geriatric patients, who are more likely to be using concomitant medications than younger patients. The guidance states that "Therefore it is a general principle, not specific to these guidelines, that in cases where the therapeutic range (i.e., range of toxic to therapeutic doses) of the drug or likely concomitant drugs is narrow, and the likelihood of the concomitant therapy is great, that specific drug-drug interaction studies be considered. The studies needed must be determined case-by-case, but the following are ordinarily recommended:

- Digoxin and oral anticoagulant interaction studies, because so many drugs alter serum concentrations of these drugs, they are widely prescribed in the elderly, and they have narrow therapeutic ranges;
- For drugs that undergo extensive hepatic metabolism, determination of the effects of hepatic-enzyme inducers (e.g., phenobarbital) and inhibitors (e.g., cimetidine);
- For drugs metabolized by cytochrome P-450 enzymes, it is critical to examine the effects of known inhibitors, such as quinidine (for cytochrome P-450 2D6) or ketoconazole and macrolide antibiotics (for drugs metabolized by cytochrome P-450 3A4). There is a rapidly growing list of drugs that can interfere with other drugs that metabolize, and sponsors should remain aware of it;
- Interaction studies with other drugs that are likely to be used with the test drug (unless important interactions have been ruled out by a pharmacokinetic screen)."

The aforementioned changes in PK and metabolism with age may have a pronounced impact on the safety of narrow therapeutic index drugs (e.g. anticancer drugs). These effects have been reviewed by Wildiers *et al.* (2003). Other safety concerns in the elderly may be due to DDIs that arise due to the polypharmacy that is common in the elderly (Mallet, Spinewine and Huang, 2007).

In August 2006, the European Medicines Agency (EMEA) received a request from the European Commission (EC) for the Committee for Human Medicinal Products

(CHMP) to provide for an opinion on the adequacy of guidance on the elderly regarding medicinal products for human use, on the basis of Article 5(3) of Regulation (EC) No 724/2004. This led to the creation of a report, "Adequacy of Guidance on the Elderly Regarding Medicinal Products for Human Use" that made several recommendations (EMEA, 2006a). Although the exact recommendations contained in this report are not pertinent to the topic of drug metabolites in regulatory guidances, clinical trials, and product labeling, the report does discuss drug metabolism in several places.

> The "Note for Guidance on Studies in Support of Special Populations: Geriatrics" (CPMP/ICH/379/95) states that the information regarding age-related differences in the PK of a drug can come, at the sponsor's option, either from a PK screen in conjunction with the main phase 3 studies or from formal PK studies in the elderly (EMEA, 1994a). Where the screen detects large differences, then formal PK studies may be indicated. The formal PK studies can be done in healthy geriatric subjects or in patient volunteers with the disease to be treated by the drug. More restrictive in this sense, the guideline on "Pharmacokinetic Studies in Man" (EudraLex 3CC3a) also recommends that kinetics should be studied in the extreme ages but in patients and not in healthy subjects (European Economic Community, 1987). However, it acknowledges that this requires multiple, long, and expensive studies which cannot all be performed before authorization. The ICH E7 considers that specific drug-drug interaction studies should be planned in the cases where the therapeutic index of the drug or concomitant drugs is narrow (ICH, 1994c). Moreover, ICHE8 "General considerations for clinical trials" recommends that for drugs which are frequently co-administered, it is usually important that drug–drug interaction testing are performed in non-clinical and, if appropriate, in human studies (ICH, 1997a). The "Note for guidance on the investigations of drug interactions" (CPMP/EWP/560/95) outlines the requirements for interaction studies on new chemical entities (NCE) on the basis of their physico-chemical, pharmacokinetic and pharmacodynamic properties, but not on the basis of their clinical indication or the likelihood of the targeted population of having concomitant treatments (EMEA, 1997). Nevertheless, this is done intentionally, as mechanistic studies can be done to extrapolate to a variety of situations. Knowledge from adequate absorption/distribution/metabolism/excretion (ADME) studies and mechanistic interaction studies may help in predicting possible problems due to change in $C_{max}$, AUC or interaction. The fact that specific PK studies would probably be conducted in healthy elderly, and therefore may be less relevant for the targeted population, is also discussed. It is agreed that the place of population PK should be further explored.

***5.7.8 Pediatrics*** The administration of drugs to the pediatric population is covered in several guidances or guidelines issued by the FDA, EMEA, Health Canada, and ICH. Many of these documents include information of relevance to the drug metabolism scientist, indicating the importance of drug metabolism in the pediatric population. In 1998, the FDA issued a draft guidance entitled "General Considerations for Pediatric Pharmacokinetic Studies for Drugs and Biological Products" (US FDA, 1998c). As of October 2007, this guidance has not been finalized and the reasons for it remaining in the draft state are unknown. The guidance provides extensive literature citations to support the points made in this draft guidance. However, the literature citations may now be slightly outdated. This guidance provides some background information, including:

> In the pediatric population, growth and developmental changes in factors influencing ADME also lead to changes in pharmacokinetic measures and/or parameters. To achieve AUC and $C_{max}$ values in children similar to values associated with effectiveness and safety

in adults, it may be important to evaluate the pharmacokinetics of a drug over the entire pediatric age range in which the drug will be used. Where growth and development are rapid, adjustment in dose within a single patient over time may be important to maintain a stable systemic exposure.

The guidance then provides information on special areas of importance in planning pediatric pharmacokinetic studies:

Developmental changes in the pediatric population that can affect absorption include effects on gastric acidity, rates of gastric and intestinal emptying, surface area of the absorption site, gastrointestinal enzyme systems for drugs that are actively transported across the gastrointestinal mucosa, gastrointestinal permeability, and biliary function. Similarly, developmental changes in skin, muscle, and fat, including changes in water content and degree of vascularization, can affect absorption patterns of drugs delivered via intramuscular, subcutaneous, or percutaneous absorption. … Drug metabolism usually occurs in the liver, but may also occur in the blood, gastrointestinal wall, kidney, lung, and skin. Developmental changes in metabolizing capacity can affect both absorption and elimination, depending on the degree to which intestinal and hepatic metabolic processes are involved. Although developmental changes are recognized, information on drug metabolism of specific drugs in newborns, infants, and children is limited. In general, it can be assumed that children will form the same metabolites as adults via pathways such as oxidation, reduction, hydrolysis, and conjugation, but rates of metabolite formation can be different. In vitro studies performed early in drug development may thus be useful in focusing attention on metabolic pathways of elimination in both adults and children. … Protein binding may change with age and concomitant illness. In certain circumstances, an understanding of protein binding may be needed to interpret the data from a blood level measurement and to determine appropriate dose adjustments. In vitro plasma protein binding studies can determine the extent of binding of the parent and the major active metabolite(s) and identify specific binding proteins, such as albumin and alpha-1 acid glycoprotein. Optimal estimates of the degree to which protein binding is linear may be obtained by testing maximum and minimum observed concentrations.

Further information is provided upon factors to consider in designing a study in the pediatric population:

Because there may be limited information on the safety of the dose to be administered to a neonate or infant, doses in initial studies require careful consideration. Factors for consideration include (1) the relative bioavailability of the new formulation compared to the adult formulation; (2) the age of the pediatric population; (3) the therapeutic index of the drug; (4) pharmacokinetic data from the adult population; and (5) body size of the pediatric study population. Initial doses should be based on mg/kg of body weight or mg/m$^2$ of body surface area, extrapolated from adult doses. Knowledge of ADME in an adult population should be combined with an understanding of the physiologic development of the intended pediatric study population to modify the initial dose estimate. Consideration should initially be given to administering a fraction of the dose calculated from adult exposure, depending on the factors mentioned above and depending on whether there is any pediatric experience. Subsequent clinical observations and prompt assay of biological fluids for the drug and/or its metabolites should permit subsequent dose adjustment. The analytical method used to quantify the drug and metabolite(s) in the biological fluid of interest should be accurate, precise, sensitive, specific, and reproducible. Ideally, the method should be relatively rapid, readily adaptable, and use only minimum sample volumes. Protein binding studies may be performed if considered important.

The use of small sample volumes may require a revalidation of the bioanalytical assay.

In 2001, an ICH guidance (E11) called "Clinical Investigation of Medicinal Products in the Pediatric Population" came into operation (ICH, 2000c). This guidance makes some general statements regarding the role of metabolism in the clearance of drugs in the pediatric population. "Knowing the pathways of clearance (renal and metabolic) of the medicinal product and understanding the age-related changes of those processes will often be helpful in planning pediatric studies." The development of clearance mechanisms in different age groups is also described as follows.

> *Infants and toddlers (28 days to 23 months)*
>
> This is a period of rapid CNS maturation, immune system development and total body growth. Oral absorption becomes more reliable. Hepatic and renal clearance pathways continue to mature rapidly. By 1 to 2 years of age, clearance of many drugs on a mg/kg basis may exceed adult values. The developmental pattern of maturation is dependent on specific pathways of clearance. There is often considerable inter-individual variability in maturation.
>
> *Children (2 to 11 years)*
>
> Most pathways of drug clearance (hepatic and renal) are mature, with clearance often exceeding adult values. Changes in clearance of a drug may be dependent on maturation of specific metabolic pathways.
>
> *Puberty*
>
> The onset of puberty is highly variable and occurs earlier in girls, in whom normal onset of puberty may occur as early as 9 years of age. Puberty can affect the apparent activity of enzymes that metabolize drugs, and dose requirements for some medicinal products on a mg/kg basis may decrease dramatically (e.g., theophylline). In some cases, it may be appropriate to specifically assess the effect of puberty on a medicinal product by studying pre- and postpubertal pediatric patients.

The guidance also notes that "the volume of blood withdrawn should be minimized in pediatric studies." One suggestion is the use of sensitive assays for parent drugs and metabolites to decrease the volume of blood required per sample. However, it should be noted that the volume of plasma used for the bioanalytical assay should be small and that a sensitive assay does not necessarily mean that the volume of plasma required for the assay will be low.

In 2003, Health Canada issued an addendum to ICH E11 that clarifies that the Canadian research and regulatory environment is meant to more precisely define regulatory considerations for the timing and conduct of clinical studies in the pediatric population.

In 2007, the EMEA issued a guideline called "Guideline on the Role of Pharmacokinetics in the Development of Medicinal i the Paediatric Population" (EMEA, 2006h). In a similar manner to the draft FDA guidance, this guideline provides some general and specific information. The guideline states that:

> The design of pharmacokinetic and PK/PD studies in paediatric patients should be based on a number of factors including the known pharmacokinetic characteristics (dose- and time-dependency of pharmacokinetics, route of elimination, presence of active metabolites, protein binding, etc), route of administration, therapeutic index, paediatric group investi-

> gated, possibility to collect blood samples, sensitivity of the analytical method, method for the pharmacokinetic data analysis, desired use of the pharmacokinetic data, etc.

It further states that:

> If there are pharmacologically active metabolites significantly contributing to the efficacy or safety, the pharmacokinetics of such metabolites should be studied, unless it may be assumed that the exposure ratio of metabolite to parent drug is similar to the ratio in the "reference age group". Blood sampling may be difficult and the number of samples is usually limited, especially in younger age groups. Effort should be put into optimising study design and use of the available data as well as further developing the analytical methods to allow for small sample volumes to be used. If a medicinal product is not metabolised and elimination is predominantly renal, or if the drug is partly eliminated through renal excretion in a dose-linear fashion, data on urinary excretion may be used to describe elimination capacity.

The guideline provides some specific considerations for preterm and term newborn infants, infants and toddlers since these groups present the largest pharmacokinetic challenges. The guideline notes that "rapid developmental changes in absorption, distribution, metabolism and excretion, combined with all possible disease processes that might interfere with the developmental changes, necessitate a study design that is tailored to these populations." Two points of interest are the statements that "determining the protein binding of highly bound drugs and active metabolites should be considered when studying newborns" and "The effect of genetic polymorphisms in genes coding for drug metabolising enzymes, as well as hepatic or renal impairment on the pharmacokinetics within the paediatric age group, may be different from the effect in adults." How these points are addressed is not discussed in this guideline. It should be noted that in 2008, a corrigendum was issued to this guideline. However, it is not possible to determine what was corrected in this version since neither the guideline itself nor the EMEA web page gives any indication as to what was corrected and the uncorrected version of the document is no longer available on the EMEA web site.

One of the challenges in conducting pharmacokinetic or other studies in the pediatric population is the selection of the dose formulation. Small children may not be able to swallow the capsules or tablets that are administered to adults. Additionally, the dose strengths of such capsules or tablets may be inappropriate for pediatric studies. This topic was addressed in a reflection paper (Formulations of Choice for the Paediatric Population) issued by the EMEA in 2006 (EMEA, 2006k).

This reflection paper notes that "Gastric pH is relatively high in the neonatal period; gastric emptying increases; gut motility matures during early infancy and there are changes to splanchnic blood flow, intestinal drug-metabolising enzymes, microflora and transporters. There are few published bioavailability studies but in general, rate of absorption is slower in neonates and infants than in older children." The mention of differences in microflora between the pediatric and adult population acknowledges the role of the microflora as a drug-metabolizing system. The role of the GI microflora in the metabolism of certain drugs represents an understudied area. The guideline does not address some of the logistical challenges in the administration of drugs to the pediatric population. Often, drugs are dissolved in apple (or other juice) or mixed with food in order to ensure that the drug can be injested by the subject. Therefore, the stability of the drug in the dose formulation (e.g. apple juice or infant cereal) should be determined. Additionally, the effect of the juice

components on the PK of the drug and the potential food effect when the drug is administered in food should be taken into account when interpreting data from pediatric studies that utilize these modes of drug administration. It should be noted that some of these modes of administration may also be required for administration to an elderly population.

The guideline further notes that "In general, clearance of substances metabolised in the liver is greater in children than in adults, requiring higher doses per kg body weight." The ontogeny of the development of DMEs is also described:

> The main pathway for phase 1 reactions is oxidation using the cytochrome P450 dependent (CYP) enzymes and these enzymes are generally immature at birth reaching maximum values at about 2 years of age. Hepatic clearance of some substances will be greater on a per kg body weight basis than for adults e.g. carbamazepine, theophylline. The different CYP families of enzymes mature at different rates and there may be significant inter-individual variation. Phase 2 reactions include glucuronidation and sulphation. Many different enzymes are involved and they develop at different rates such that metabolism of substances may vary considerably in infancy both qualitatively and quantitatively. Neonates are unable to conjugate benzoic acid efficiently and this is of great importance to the use of benzyl alcohol as an excipient in this age group since its metabolite benzoic acid can accumulate and is toxic. The ontogeny and pharmacogenomics of metabolising enzymes for active drugs is receiving increasing attention but there is little information about the effect on excipients. They may not be able to metabolise or eliminate an ingredient in a pharmaceutical product in the same manner as an adult. Several EU guidelines related to the use and declaration of excipients have been published and should be consulted. Additional information may be found in documents published by the US Food and Drug Administration (FDA).

In addition to making several important and interesting scientific points, this guideline is one of the few regulatory documents that explicitly advocates the use of guidances from other regulatory agencies.

The guideline provides examples of excipients whose use may be problematic in infants. Examples include benzyl alcohol/benzoic acid/sodium benzoate. "Benzyl alcohol is often used as a preservative in injectable medicines. It can be toxic in neonates and pre-term neonates due to their immature metabolism. In developing pharmaceutical preparations for use in pre-term infants, neonates and young children benzyl alcohol should be carefully evaluated and may best be avoided. It may cause pain on injection" and propylene glycol:

> Propylene glycol is used as a solvent in oral, topical and injectable medications, often for substances which are not highly soluble in water, e.g. phenobarbital, phenytoin and diazepam. It is also commonly used in injectable multivitamin concentrates. Children below 4 years have a limited metabolic pathway (alcohol dehydrogenase), therefore accumulation of propylene glycol can occur in the body. For example, it has been shown that neonates have a longer propylene glycol half-life (16.9 hours) compared to adults (5 hours). Products containing high levels of propylene glycol should not be administered to children below the age of 4 years. Main toxic action is depression of the central nervous system. High osmotic pressure may cause laxative effects. Topical administration has been reported to cause contact dermatitis.

The CHMP released a draft guideline on "The Investigation of Medicinal Products in the Term and Preterm Neonate" for review in September 2007, with final comments being due in May 2008 (EMEA, 2007k).

This draft guideline provides some useful information related to the physiology of organ systems such as the heart/lung, CNS, liver, kidney, and GI tract of the neonate, and how they relate to PK and metabolism in the neonate. The draft guideline states that:

> The post-natal cardiopulmonary system adaptation marks the most dramatic changes during and after birth. Some of these changes occur instantaneous with the first breath, whereas others occur within hours or days after birth. In general, the impact of lung and heart maturation on PK/PD relationship (e.g., closure of the ductus arteriosus) has to be considered. As adequate cardiopulmonary function is paramount to maintain organ function in general (e.g., renal blood flow, brain perfusion, liver function), any potential impact on either cardiac or pulmonary function needs to be carefully monitored in neonatal clinical trials. The influence of cardiopulmonary function as the basis to maintain hepatic drug metabolism and excretion as well as renal excretion has to be considered.
>
> Transport across the blood brain barrier by both passive diffusion and by active transporters is age-related and undergoes constant maturational changes in the neonate. This may contribute to a significantly altered distribution of active substances or metabolites into the CNS with a potential impact on both clinical efficacy and adverse effects. Medicinal products known or expected to be substrate for specific transporters (e.g., P glycoprotein, Pgp) require specific consideration.
>
> Hepatic blood flow, plasma protein binding and intrinsic clearance determining hepatic clearance undergo significant post-natal changes. Most enzymatic microsomal systems responsible for drug metabolism are present at birth and their activities increase with advancing post-natal and gestational age. Rapid maturational changes occur during the first weeks of life. Hepatic clearance may be influenced by premature birth, pathologic conditions of the neonate or administration of drugs to the mother or to the neonate.
>
> To predict the exact nature of these consequences requires an understanding of post-natal maturation and main involved enzymes. The development of specific enzymes is partly described in the scientific literature and may allow estimations of drug metabolism in the neonate. These data should be considered when planning neonatal studies.
>
> The main pathway responsible for metabolism may be different in neonates as compared to adults. The applicant should consider this when assessing exposure margins of metabolites to the animals used in preclinical studies and also when comparing human safety data obtained in adults and older children. The relevant hepatic phase I and II metabolic pathways should be identified.
>
> If pharmacologically active metabolites are known to be formed, potential differences in exposure of such metabolites should be considered. If feasible, the applicant is encouraged to perform studies investigating drug metabolism in vitro in neonatal hepatic material (microsomes, hepatocytes etc.).
>
> In utero exposure to enzyme inducing agents (e.g., antiepileptic drugs, barbiturates, glucocorticoids) and the potential to temporarily alter post-natal drug disposition need to be considered when planning a study in neonates and in the interpretation of data.
>
> If the drug investigated is likely to be eliminated mainly through hepatic metabolism, markers of hepatic function could be included as covariates in the pharmacokinetic data analysis (e.g., in population PK analysis) as well as included in the safety assessment. Monitoring could include standard laboratory and imaging procedures.
>
> Renal clearance mechanisms include glomerular filtration (GFR), tubular secretion and reabsorption. Glomerular filtration matures faster than the tubular function, and both depend not only on age and maturational status but also on adverse factors occurring in

the pre- and post-natal period, including for example intrauterine growth retardation or administration of nephrotoxic drugs to the mother and the neonate.

Data concerning maturational changes of the neonatal gastrointestinal tract that may influence drug bioavailability are still limited.

Gastrointestinal absorption is influenced by factors such as tissue perfusion, surface area, gastric and intestinal pH, intestinal mobility and transit time as well as maturation of transporters and receptors. In principle, all these factors are reduced or immature in the neonate. The post-natal developmental pattern of these factors may additionally be highly variable due to environmental factors (i.e., diet, drug administration), genetic factors and underlying pathophysiology. Changes in bioavailability during the early post-natal period have to be considered and need to be predicted as accurate as possible in clinical trials including drugs administered orally.

Gastric pH is neutral at birth with gastric acid secretory capacity appearing after the first 24 to 48 hours of life. Post-natal increases in gastric acid production generally correlate with post-natal age and adult levels are reached by approximately 2 years of age. High gastric pH in the neonate may lead to increased bioavailability of weakly basic compounds and reduced bioavailability of weakly acidic compounds. Additionally, in premature infants, gastric pH may remain elevated due to immature acid secretion. This may lead to higher serum concentrations of acid-labile drugs in the premature neonate.

As pancreatic and biliary functions are immature at birth, bioavailability of drugs requiring pancreatic exocrine and biliary function may have reduced bioavailability. Both functions develop rapidly in the neonatal period, requiring careful consideration of increased bioavailability of orally administered drugs in neonatal clinical trials.

Reduced gastrointestinal motility may have unpredictable effects on drug availability in neonates. It may reduce the rate of drug absorption or conversely improve drug bioavailability due to longer retention times in the small intestine. Additionally, maturation of intestinal metabolising enzymes and transport proteins remains largely unknown, further leading to the unpredictability of oral bioavailability and intestinal first-pass effect of orally administered drugs in the neonate. Drugs undergoing secondary metabolism and secretion into the gut, especially when glucuronidation with enterohepatic recirculation occurs in adults and older children, may have different bioavailability and exposure because of reduced glucuronidation and bacterial activity in the intestine of neonates. Reduced gastrointestinal mobility that is often present in sick neonates is therefore particularly important to consider.

## Formulations and route of administration are also addressed as follows:

The choice of formulation and route of administration should depend on the condition to be treated and the clinical state of the neonate. Age-appropriate formulations using appropriate excipients must be developed to avoid extemporaneous preparations, even more so for neonates. Novel formulations should be evaluated through preclinical studies and in adults or older children as appropriate before consideration for administration to neonates.

Medication errors in neonatal practice are commonly due to use of inappropriate formulations requiring calculation and measurement of very small volumes or multiple dilutions. Prescribing software may not be appropriate for neonatal use. Excipients used for adults and older children may be toxic in neonates because of immature metabolism and elimination. The salt of the active ingredient and the chemical nature of the preparation must be carefully considered to avoid administration of excessive amounts of electrolytes.

In general, the IV route will normally be used in clinically unstable term and preterm neonates.

PK and PK/PD studies are described in detail. Reference is made to the "Guideline on the Role of Pharmacokinetics in the Development of Medicinal Products in the Paediatric Population," especially section 4.1. The current draft guideline states that:

Pharmacokinetic information is important to support adequate dosing in subpopulations of the clinically studied population and to assess the potential for clinical relevance of toxicity findings in the preclinical studies. In neonates, however, pharmacokinetics alone is of limited value in neonates for extrapolating efficacy and safety from other patient groups and extrapolation of efficacy will in general need PK/PD monitoring.

A population PK approach is preferable due to the importance of finding covariates related to dose-individualization between individuals and over time in the maturating individual. The analysis can be made on rich and/or sparse data depending on the number of patients available and the possibility of developing highly sensitive analytical methods where very small sample volumes could be used. The initial model could be based on rich data of a limited number of individuals and on other prior information, followed by a population PK approach.

It should be noted that population PK and modelling of oral administration require extra cautious consideration in the neonatal population as there may be marked absorption differences in neonates as compared to other age groups as well as very prolonged absorption in a subgroup of individuals.

In cases where $C_{max}$ is clinically important for safety or efficacy reasons, efforts should be made to characterise this parameter satisfactorily due to the differences in volume of distribution between neonates and older children. If possible, the protein binding of highly protein bound active substances should be assessed to enable the measurement of free plasma concentrations. Immature expression of carrier proteins should also be considered. Special consideration should be given to drugs which are highly protein bound and fast metabolized in adults, since major differences can be assumed in newborns, as synthesis of binding proteins such as albumin could be lower in the neonate with consequences on drug binding and free bilirubin. The need for differentiation between a loading dose (large $V_d$) and smaller maintenance doses (low total body clearances) as important, e.g., for methylxanthines, aminoglycosides, and anticonvulsants, has to be identified.

Effort should be made to include the determination of potential covariates in the studies (PNA, PMA, GA, weight, body surface area [BSA], renal function, concomitant use of drugs, S-bilirubin, repeated feeding and feeding patterns etc.) for allowing covariates to be identified which may allow satisfactory dose individualization. Adjustments of the dose by covariates (e.g. bodyweight, BSA) should usually be based on the covariate with the highest correlation to the relevant PK parameters. However, the difficulties in determining the covariate should be taken into account. The determination of BSA is difficult in neonates and other covariates should be considered if their use gives an adequate dosing. Titration based on plasma concentration or a clinical safety or efficacy marker should also be considered. This is further described in the Guideline on the Role of Pharmacokinetics in the Development of Medicinal Products in the Paediatric Population.

Neonates treated in hospital and especially on NICUs often receive multiple drugs to treat different conditions. Therefore, any known or potential interactions of the medicinal product investigated should be carefully considered when planning a clinical study as well as during data analysis. Concomitantly used drugs should be included in the population pharmacokinetic analysis. In general, formal interaction studies should be performed in adults. However, if the main enzymes involved in the elimination of the drug are different

> in the neonate, results of adult interaction studies investigating effects of other drugs on the investigated medicinal product can not be directly extrapolated to neonates. In these cases, estimations based on *in vitro* metabolism data as well as other sources of information should if possible be performed. If a dosing recommendation is needed for a commonly used drug combination and if an interaction is expected, specific pharmacokinetic interaction studies should be considered.

Related guidances include the 2005 FDA draft guidance "Clinical Lactation Studies —Study Design, Data Analysis, and Recommendations for Labeling" (US FDA, 2005g). The draft guidance states that "Such a study usually enrolls mother-infant pairs who are planning to or are currently receiving study medication. Its hallmark is the frequent collection of corresponding maternal blood and milk samples as well as sampling of infant blood and/or urine. Infant sampling provides information regarding the fraction of drug that is systemically available to the breast-fed child. Total clearance of the drug or metabolite by the breast-fed child can be estimated as well. This design can be considered if information is already known about the extent of drug transfer into breast milk, but the amount absorbed by the breast-fed child is not known. Other drugs that can be considered for a mother-infant pair design include drugs already approved and known to be used by lactating women who continue to breast-feed and drugs used to treat chronic maternal conditions. Drug or metabolite characteristics that favor selection of this study design include:

- High lipophilicity (weak bases)
- Potential for accumulation in breast milk
- Likelihood of being well absorbed by the breast-fed child
- Wide distribution to multiple organs
- Long half-life"

An alternative study design is also described: "In a sequential or step-wise approach to lactation studies, the lactating women (plasma and milk) study design might be considered before the infant is exposed to drug via breast milk in a more complex study. Situations that might favor use of this design include newly approved drugs (especially for drugs with no pediatric data), short-term or acute maternal dosing, and unknown risk of exposure to the breast-fed child. Drug and metabolite characteristics that favor selection of this study design include:

- High lipophilicity (weak bases)
- Presence in milk
- Predictions that drug is present in milk
- Knowledge of a class effect"

The draft guidance states that "For drugs that are hepatically metabolized and known to exhibit genetic polymorphism (e.g., CYP2D6 or CYP2C19), the metabolic status of the enrolled subjects (maternal and infant) can be important factors when analyzing the results of the study." However, how such polymorphisms (which can be different in parent and child) are used in evaluating the data is not discussed.

The guidance makes the point that:

> Because of varying lipophilicity among drugs, it is also important to assay milk samples for milk fat. Alternative, noninvasive pediatric sampling strategies (e.g., saliva, tears) might also be used to estimate drug levels in infants. However, drug concentrations obtained from

alternative fluids (e.g., saliva, tears) might not be equivalent to those obtained from plasma. Sponsors are, therefore, encouraged to demonstrate the relationship of the drug concentration between plasma and alternative fluids in adults. Estimating infant drug exposure via breast milk solely from excretion of unchanged drug in infant urine can be of limited utility because of the difficulty with urine collection and the variability of renal clearance and urine production in infants.

As can be seen, the design, conduct and analysis of such studies will be complicated.

***5.7.9 Reproduction and Pregnancy*** A draft guideline (Guideline on Risk Assessment of Medicinal Products on Human Reproduction and Lactation: From Data to Labeling) was issued by the EMEA in 2006 which covers similar topics to the two FDA guidances described earlier (EMEA, 2006c). This draft guideline states that "Information about the excretion into milk of the active substance and/or its metabolites should be available" and "The exposure in pregnant animals measured by plasma concentrations of the compound and/or metabolites should be assessed." These points imply that the secretion of drug and/or metabolites should be evaluated in animals and/or humans and that the metabolism of the drug in animals be determined prior to the determination of the exposure of the pregnant animal to the drug and/or metabolites. In most cases, the toxicology study is performed prior to the determination of the metabolism in the pregnant animal, if the metabolism is described in the pregnant animal. If the metabolism of the drug is determined in the pregnant animal, it is often only done in one species (usually rabbit). It is assumed that the metabolism in the nonpregnant and pregnant rat are identical, although that is not always the case. Additionally, distribution of drug and/or metabolites to the fetus is only occasionally performed and is often only performed in one species (it is assumed that the placental transfer is similar in both the rat and rabbit). The timing of when (in the course of the pregnancy) to address the metabolism of the drug and fetal exposure to the drug and/or metabolites is a subject that can engender great debate. It is well known that the metabolic capacity and capabilities of the fetus will change with the stage of pregnancy; that pregnancy can affect the clearance of the drug by the mother and that the placental transfer of the drug and/or metabolites can change with the stage of the pregnancy. Therefore, it can be seen that the description of the metabolism in pregnant animals is a complicated matter that can easily become resource- and time-intensive.

In 2004, the FDA issued a draft guidance entitled "Pharmacokinetics in Pregnancy—Study Design, Data Analysis, and Impact on Dosing and Labeling" (US FDA, 2004c). The CHMP issued a draft guideline (Guideline on Risk Assessment of Medicinal Products on Human Reproduction and Lactation: From Data to Labelling) in 2006 that covers similar topics to the two FDA guidances described earlier (EMEA, 2006c). The guideline is expected to be finalized in 4Q 2007/1Q 2008. The FDA guidance provides far more details on impact of pregnancy on the pharmacokinetics of a drug. For example, the guidance states: "Extrapolation of PK data from studies performed in nonpregnant adults fails to take into account the impact of the many physiologic changes that occur during pregnancy. Most of the physiologic changes manifest during the first trimester and peak during the second trimester of pregnancy. Physiologic changes are not fixed throughout pregnancy but rather reflect a continuum of change as pregnancy progresses, with return to baseline at various rates in the postpartum period. The physiologic changes have the potential to alter the PK and/or PD of drugs. Some of these changes include:

- Changes in total body weight and body fat composition.
- Delayed gastric emptying and prolonged gastrointestinal transit time.
- Increase in extra cellular fluid and total body water.
- Increased cardiac output, increased stroke volume, and elevated maternal heart rate.
- Decreased albumin concentration with reduced protein binding.
- Increased blood flow to the various organs (e.g., kidneys, uterus).
- Increased glomerular filtration rate.
- Changed hepatic enzyme activity, including phase I CYP450 metabolic pathways (e.g., increased CYP2D6 activity), xanthine oxidase, and phase II metabolic pathways (e.g., N-acetyltransferase)."

It should be noted that the induction of CYP2D6 in pregnancy is somewhat unique and unexpected, since CYP2D6 does not appear to be inducible by xenobiotics.

The guidance further states that:

> A significant amount of pharmacologic research has been conducted to improve the quality and quantity of data available for other altered physiologic states (e.g., in patients with renal and hepatic disease) and for other patient subpopulations (e.g., pediatric patients). The need for PK/PD studies in pregnancy is no less than for these populations, nor is the need for the development of therapeutic treatments for pregnant women.
>
> For drugs that are metabolized by enzymes known to exhibit genetic polymorphism (e.g., CYP2D6 or CYP2C19), the FDA recommends that the investigator consider the metabolic status of the enrolled subjects when analyzing the results of the study. Genotype has been shown to have an effect on pregnancy-related changes in metabolism. … Plasma protein binding, like renal function, is often altered in pregnancy. For example, albumin and alpha-1-acid glycoprotein levels are reduced in pregnancy, consequently the protein binding of drugs can be affected. With systemically active drugs and metabolites, the unbound concentrations are generally believed to determine the rate and extent of delivery to the sites of action. For drugs and metabolites with a relatively low extent of plasma protein binding (e.g., the extent of binding is less than 80 percent), alterations in binding due to pregnancy are small in relative terms. In such cases, description and analysis of the PK in terms of total concentrations would be sufficient. For drugs where the extent of protein binding is greater than 80 percent, primarily to albumin, it is recommended that the PK be described and analyzed with respect to the unbound concentrations of the drug and active metabolites. Although unbound concentrations should be measured in each plasma sample, if the binding is concentration-independent and unaffected by metabolites or other time-varying factors, the fraction unbound can be determined using a limited number of samples or even a single sample from each patient during each trimester. The unbound concentration in each sample should then be estimated by multiplying the total concentration by the fraction unbound for the individual patient.

The design of pharmacokinetic studies, as well as ethical concerns, is addressed in depth in this draft FDA guidance. Selected points of interest are as follows. The draft guidance notes that:

> Ethical issues are important when considering studying drugs in pregnant women. Given the large number of pregnant women who need prescription medicines to maintain their health, some have argued that it is unethical not to obtain dosing information in this subpopulation. Others recommend that only pregnant women who need a drug for therapeutic

reasons be included in clinical studies, citing that drug studies cannot be done in "normal pregnant volunteers."

It further notes that:

Although PK studies in pregnancy can be considered in Phase 3 development programs depending on anticipated use in pregnancy and the results of reproductive toxicity studies, the FDA anticipates that most PK studies in pregnant women will occur in the postmarketing period and will be conducted using pregnant women who have already been prescribed the drug as therapy by their own physician. An example of a minimal risk study would be one to determine PK/PD of an antihypertensive medication in pregnant women who are taking that medication to treat hypertension during pregnancy. The decision to use the antihypertensive medication is made by the patient and her physician independent of participation in the PK/PD study.

Generating a full understanding of the effect of pregnancy on a given drug is a far from trivial task as noted in the FDA's draft guidance:

Study design considerations are important when conducting a study in pregnant women to determine if the PK and/or PD are altered enough to require an adjustment from the established dosage. Ideally, PK studies in pregnancy would be done pre-pregnancy (for baseline comparison) and during all three trimesters, especially for chronically administered drugs. Given the constraints of a study design that enrolls women prior to pregnancy, an alternative can be to determine PK/PD in the second and third trimesters, with the baseline assessment for comparison to the pregnant state done in the postpartum period. The Agency recommends care be taken to select the most appropriate postpartum time for PK/PD determination, if known. Cardiovascular and renal changes do not return to the pre-pregnancy state until 3 months postpartum. Optimally, postpartum PK/PD assessments for comparative purposes to PK/PD in pregnancy would be done when the woman is neither pregnant nor lactating.

The guidance notes that:

It is possible to study drugs that have no intended direct therapeutic benefit to the pregnant woman provided that the risk to the fetus is minimal (45 CFR 46). For example, probe substrates can be used to investigate drug metabolism (e.g., cytochrome P-450 activity) or drug transporter status (e.g., p-glycoprotein). Data from these studies offer generalizable information to other pregnant women but do not offer direct therapeutic benefit to study participants. The Agency encourages sponsors or investigators to explore additional safeguards for human subject protection for this type of study. To minimize exposure to a nontherapeutic drug, each pregnant woman can be exposed to the drug once during pregnancy and in the postpartum period employing a nonlongitudinal design (e.g., one cohort of women sampled in second trimester and postpartum and another cohort of women sampled in third trimester and postpartum). Examples of additional safeguards include administering only products with a long or known record of safety in pregnancy, administering products using only a single dose of the drug, using lower doses of the drug, decreasing the number of drugs (probe substrates) used in any study subject, and limiting study participants to pregnant women only in second or third trimester.

Although some researchers (e.g. Hodge and Tracey, 2007) have summarized the literature related to the effect of pregnancy on the disposition on drugs, it appears fair to state that, at the present, there appear to be gaps in our knowledge regarding how

to determine the metabolism of drugs during pregnancy and that this is an area that may warrant further research and understanding.

***5.7.10 Human ADME Studies*** As can be gathered from all of the guidances and guidelines discussed earlier, the determination of the metabolism of a drug in humans is of pivotal importance. However, there are no specific regulatory guidances or guidelines that pertain to the design and conduct of ADME studies in humans. However, legislation exists in the United States that is aimed at ensuring "safe and effective use of the radioactive drug" and in the EU that is aimed at "protection of individuals against the dangers of ionizing radiation."

Unlike the guidelines and guidances, the topic of radiation safety is written into law and therefore has a different level of enforcement than do the guidances or guidelines. The EU legislation enables individual member states to compose national regulations regarding exposure to radioactivity in medical research. There are two European Atomic Energy Community (EURATOM) directives that are pertinent to the conduct of human ADME studies. The first is 96/29/EURATOM which imposes the ALARA principle ("As Low as Reasonably Achievable"), the intent of which is to minimize the exposure of human subjects to radiation (The Council of the European Union, 1996). This directive is aimed at workers and the general public. The administration of radioactivity to subjects in clinical trials is addressed in 97/43/EURATOM (The Council of the European Union, 1997). This directive supplements the previous directive and addresses issues such as the role of the Ethics Committee and quality of the radioactive material, but does not address the level of radioactivity that can be administered.

The US legislation (21 CFR 361) addresses the use of drugs in humans "intended to obtain basic information regarding the metabolism (including kinetics, distribution and localization) of a radioactively labeled drug" (US FDA, 2007d). The legislation places the responsibility for the assessment of the safety of the radiolabeled material to be used in the clinical trial on the local Radioactive Drug Research Committee (RDRC). Echoing the EU directive, the US legislation requires that the administered dose is as low as possible without jeopardizing the feasibility of the study. Maximum allowable exposures (in terms of a single dose and annual exposure) are defined for the whole body, active blood-forming organs, the lens of the eye, and the gonads. For subjects that are less than 18 years of age, the limits are 10-fold lower.

Neither the EU nor US legislation addresses other elements of the design and conduct of human ADME studies. Additionally, there are no guidances or guidelines that address these topics. The design and conduct of human ADME (mass balance) studies have been reviewed in detail by Beumer, Beijnen and Schellens (2006) and Roffey *et al.* (2007). The majority of studies are conducted in the United States, some studies are conducted in the EU (e.g., United Kingdom, the Netherlands, and Sweden), and no studies are conducted in Japan. The studies tend to be conducted in normal healthy male volunteers, although some studies are conducted in normal healthy female volunteers and the *in vivo* metabolism/mass balance of cytotoxic oncology drugs are conducted in cancer patients (if conducted at all). Studies tend to be conducted in a "young" population 18–60 and are not conducted in the pediatric or geriatric populations. Additionally, the number of subjects per study tends to be low (6–8). The studies tend to be conducted at one dose level only, with the dose approximating to the maximum dose for which approval will be sought. Serial blood samples (for processing to plasma) are collected. Urine and feces are also collected over a period of 7–10 days. Subjects are usually not discharged until a certain mass balance has been achieved or

the amount of radioactivity that is present in the collected samples has dropped below a threshold level. Additionally, the human ADME study tends to be conducted after the administration of a single dose, as is usually the case for these studies when they are performed in animals. However, it should be noted that alternative approaches should be considered for drugs that show nonlinear PK due to saturation of metabolism, for drugs that either induce or inhibit their own metabolism, for drugs that show polymorphic metabolism, and for drugs that show age-dependent differences in metabolism. A standard human ADME study will collect blood (which is subsequently processed to plasma), urine, and feces. In some circumstances, it may be useful to collect other matrices such as semen, hair, nails, and saliva. An understanding of the excretion of a drug in the semen may help in the risk–benefit evaluation with respect to reproductive toxicity. If drugs are excreted into the hair, nails, or saliva, it may be possible to utilize these matrices as means to noninvasively measuring exposure to drug and/or metabolites. For drugs that are administered orally, it will not be possible to determine whether parent drug that is found in the feces was never absorbed or absorbed and then eliminated in the bile without metabolism. Occasionally, it is possible to determine the extent of biliary elimination of a drug if the subject has an indwelling biliary T-tube or may be about to undergo surgery and will consent to the collection of bile samples. As can be anticipated, such opportunities rarely arise.

Roffey *et al.* (2007) reviewed 171 human mass balance studies and found that the average mass balance was 89 ± 11% (range = 39–113%). Eleven studies showed recoveries of >100%, which is scientifically impossible and most likely was due to errors in the determination of the amount of radioactivity in the dose or excreta or as a result of misdosing of the subject. Forty-two percent of the studies had recoveries that were <90%, and 15% of the studies had recoveries that were <80%. Usually, a recovery of 100 ± 10% is regarded as being acceptable. Low recovery of radioactivity can be due to methodological, analytical, and biological factors.

The biological factors include:

- Long elimination half-life
- Binding to specific proteins in tissues
- Binding to melanin
- Binding to phospholipids
- Other non-covalent binding to tissues
- Covalent binding
- Loss through expiration ($^{14}C$) or exchange with water ($^{3}H$).

Methodological factors include:

- Compliance
- Inadequate homogenization of feces.

Metabolite profiling tends to primarily be conducted on plasma samples, although profiling is also performed on urine and feces. Since the risk–benefit analysis for a drug will be based on the systemic (plasma) exposure to drug and/or metabolites, it can be seen that the primary focus should be on understanding the plasma metabolite profile. The number of samples that are profiled can vary from study to study and from sponsor to sponsor. Some sponsors will profile each plasma sample from each sample.

The concentration of parent and metabolite will be determined and the exposure (as measured by AUC) determined for parent drug and each metabolite and subject. The mean exposure to parent drug and each metabolite will then be calculated. Other sponsors may choose to pool samples across subjects at a given time point and then profile the pooled samples as indicated. Another approach is to pool samples across subjects and subjects. If the samples are pooled on a time-weighted basis, an estimate of the mean exposure (AUC) to the drug and metabolites can be determined based on the profiling of a single sample (Hop *et al.*, 1998; Cheung *et al.*, 2005; Han *et al.*, 2006). It may be appropriate to create two pooled samples: one that represents the samples collected over the proposed dosing interval and one based on the duration of time that the plasma samples contained measurable levels of radioactivity. Obviously, this approach means that it is not possible to make an estimate of the interindividual variability in exposure to a given metabolite. Therefore, another alternative is to create the "pseudo-AUC" pooled sample for each subject and profile that sample. This approach will allow the sponsor to begin to understand the interindividual variability in metabolic profile, albeit in a small population. A different approach is taken for the profiling of urine and fecal samples. For these matrices, a common practice is to determine the samples that cumulatively account for 90% or more of the radioactivity that is excreted in that matrix. The samples are then pooled (either per subject or across subjects) on a relative volume basis. The cutoff of 90% is somewhat arbitrary, but does ensure that the pooled sample is not diluted unnecessarily by samples that contain low amounts of radioactivity. The approaches described here are also used to varying extent for the analysis of samples from animal mass balance studies. As can be seen, there are many acceptable approaches to the profiling of samples. The manner in which the profiling is performed is more of a matter of a company's standard practices rather than an accepted required or best practice. Regardless of how the plasma profiling is performed, the outcome should be an estimate of the exposure to the parent drug, the contribution of the exposure to parent drug to exposure to total radioactivity, the contribution of a particular metabolite to the exposure to total radioactivity (for the purpose of identifying major or important metabolites), and an assessment of the exposure to selected metabolites. The FDA has not provided guidance, nor appears to have a position on, on whether a metabolite should be designated as a major metabolite based on mean or individual data (D. Jacobson-Kram, 2006, personal communication). Assuming that none of the subjects are outliers due to a genetic polymorphism in a DME involved in the metabolism of a drug, it is most probably prudent that the designation is made based upon mean data. The assessment of whether a subject is an outlier or not can be made based on exposure to parent drug or total radioactivity, or based upon the contribution of exposure to parent drug to the exposure to total radioactivity. Additionally, no insight has been provided regarding how to report the exposure to the parent or metabolite (i.e. number of significant figures and rounding of the data).

The amount of radioactivity that can be administered in a human ADME study can be contentious. Both the US and EU legislation state that the lowest amount of radioactivity that can be used to achieve the objective of the study should be used. This amount of radioactivity is determined based on the distribution of radioactivity in an animal study. These studies are most commonly conducted in pigmented rats, although other species have been used (Solon and Kraus, 2001). There is no guidance or legislation that addresses the conduct of these tissue distribution studies. From the data generated in the tissue distribution study, calculations are performed to determine

the exposure of humans to a given amount of radioactivity (Solon and Lee, 2002). Interestingly, most human ADME studies use a dose of 100 μCi and very rarely, if ever, are calculations performed to determine whether it would be necessary to administer this amount of radioactivity to achieve the objectives of the study (Dain, Collins and Robinson, 1994; Beumer, Beijnen and Schellens, 2006; Roffey *et al.*, 2007). In many cases, if the calculations were performed to determine the minimum amount of radioactivity that would be required to achieve the objectives of the study, the amount of radioactivity required would be lower than 100 μCi.

Over the past decade, AMS (accelerator mass spectrometry) has increasingly been used to determine the amount of radioactivity in biological samples, including samples obtained from a human ADME study. Since this technology can detect far lower levels of radioactivity than traditional techniques such as LSC, the amount of radioactivity that can be used in a human ADME study can be reduced even further. It is a useful technique for measuring $^{14}C$, but not for measuring $^{3}H$. The use of AMS is described in detail in a previous chapter of this book.

To date, very few human ADME studies have been performed using AMS as the detection tool. Recently, ixabepilone was approved for the treatment of breast cancer. A human ADME study was performed using AMS. However, the rationale for the use of AMS was the instability of [$^{14}C$] ixabepilone. Due to autoradiolysis, ixabepilone proved to be very unstable when labeled with conventional [$^{14}C$]-levels (100 μCi in a typical human radiotracer study). This necessitated the use of much lower levels of [$^{14}C$]-labeling and an ultrasensitive detection method, AMS (Beumer *et al.*, 2007). Regardless of the rationale for the use of AMS, it is clear that the use of AMS was acceptable to the FDA. This is consistent with statements that have been made that "As long as the approach used is scientifically valid, the FDA does not care what type of detection is used" (D. Jacobson-Kram, 2006, personal communication).

The adoption of the use of microradioactivity human ADME studies is slow. This is due partially due to institutional inertia (i.e. it takes a long time to change the traditional approaches of a pharmaceutical company), the cost of the AMS equipment, concerns about sample contamination, the cost of sample analysis ($/sample), the lack of general availability of the equipment, and a misunderstanding as to how AMS will be used for a microradioactivity study.

AMS has mainly been used primarily as a sensitive tool for the detection of the very low plasma concentrations of drug that occur during a course of a microdosing study, although it has also been used to measure femtomolar concentrations of doxorubicin in tumor cells and levels of zidovudine in peripheral blood monocytes (DeGregorio *et al.*, 2006; Vuong *et al.*, 2007). AMS may enable cancer patients to participate in a human ADME study while they are at home. Since microdose studies use μg amounts of drug, it has been mistakenly assumed that such low doses of drug would be used for a microradioactivity human ADME study. With such low levels of drug, concerns can be raised regarding the linearity in metabolism between these low doses and the pharmacologically relevant doses. However, a microradioactivity human ADME study will use pharmacologically relevant doses of drug but very low amounts of radioactivity (<200 nCi).

With respect to cost, an analysis of a traditional and microradioactivity study will most likely find that the costs for the two studies are similar in the long run. Although the analysis costs for a microradioactivity study will be higher than the costs for a traditional ADME study, there will be significant cost savings in other areas. For example, the microradioactivity study will require a smaller amount of radiolabeled drug.

Additionally, the low amount of radiolabeled drug that will be administered will be a trace amount and may not need to be subject to synthesis or release under GMP conditions. Another cost saving may be that it is likely that a tissue distribution study (for the purpose of dosimetry) will not be required before a human microradioactivity ADME study is conducted. In addition to the cost savings, there will also be time savings that would enable a microradioactivity dose human ADME study to start sooner than a traditional human ADME study could.

The US legislation states that the RDRC will determine the safety of the administered dose based on the data that they feel is appropriate. Using very conservative assumptions, it has been calculated that a dose of 200 nCi will represent 0.1% of the US whole body dose limit of 3000 mRem for a single dose. Therefore, it is likely that RDRCs will not require dosimetry studies for human ADME studies in which the dose of radioactivity is ≤200 nCi. This change in approach will potentially reduce the cost of conducting a human ADME study and give greater flexibility in the timing of such studies. It also opens the possibility that low doses of isotopically labeled drug can be administered as part of the initial human clinical trial and thereby enables a very rapid understanding of the metabolism of the NCE *in vivo* in humans. Having not to conduct the tissue distribution study also would enable sponsors to avoid some of the unintended consequences that often arise when attempting to interpret tissue distribution studies that utilize the determination of total radioactivity as the analyte. In a tissue distribution study, the concentrations of radioactivity represent the sum of the concentration of parent drug plus the concentration of all metabolites. Therefore, a tissue distribution study may be able to provide information as to which tissues the drug did not distribute to (at least, to a minimum level) but the study will not be able to provide information as to whether the radioactivity was present as parent drugs and/or metabolites. Therefore, it can be easy to assume that the organs that contain high levels of radioactivity will contain high levels of parent drug when, in reality, the radioactivity may represent metabolites and not parent drug. Another pitfall of tissue distribution studies is the assumption that organs that show high levels of radioactivity will show toxicity in animal studies. Not only do the aforementioned caveats regarding the lack of ability to discriminate between parent drug and metabolites apply, it should be recognized that the presence of parent drug and/or metabolites in a given tissue does not mean that there will be a toxic insult in that organ. Finally, a common finding in tissue distribution studies conducted in pigmented animals is the retention of radioactivity by melanin in the eye. The correlation of binding of drugs or their metabolites to melanin and subsequent toxicity is equivocal and it is possible that the retained radioactivity represents an animal-specific metabolite and therefore be a false-positive result (Roffey *et al.*, 2007). Another consequence of the detection of binding of parent drug and/or metabolites to melanin is the need to conduct a photosensitivity study (see earlier discussion). Since tissue distribution studies are not specifically recommended by guidances/guidelines or legislation for purposes other than dosimetry calculations in support of a radiolabeled study in humans and the technology exists to conduct such studies in humans at doses of radioactivity that do not require dosimetry studies, such studies may play a far lesser role in future development programs.

One element of the organizational inertia that may be responsible for the lack of adoption of the use of microradioactivity human ADME studies is not specific to this particular study design, but is a reflection of a reluctance to embark on a costly evaluation of the *in vivo* metabolism of a drug in humans until some signs of efficacy have been seen in clinical trials or until the latest possible time in the drug development

plan. Another element that should not be underestimated is that of a desire not to be the department that was responsible for the discontinuation of the development of a drug.

Although it is recognized that these studies are costly and labor-intensive, this mindset no longer seems to be consistent with the message that is being delivered by regulatory agencies (particularly the FDA) through guidances and guidelines that an understanding of the role of metabolites in efficacy and toxicity is of paramount importance. A lack of knowledge regarding the safety of metabolites may result in potential delays to clinical trials and a lack of knowledge of regarding the contribution of metabolites to the efficacy of a drug may result in inaccurate PK/PD models being created and incorrect doses being selected for phase III clinical trials.

The following is a proposal that may enable a sponsor to understand the metabolism of a new drug at a very early stage of clinical development and then design the clinical pharmacology portion of the clinical plan based on this understanding. This proposal could be applicable to non-cytotoxic drugs.

The FIH study would be conducted, in large part, in the same manner as a traditional FIH study. However, each dose would include a small (<200 nCi) amount of radiolabeled drug. This is easily achievable if the dose is administered as "powder in a bottle" (i.e. the dose is dissolved at the clinical site and administered as an oral solution) and the compound has adequate solubility. Serial blood samples would be collected and plasma prepared. Urine and/or fecal samples could also be collected in order to evaluate the mass balance as necessary. Plasma concentrations of the parent drug will be determined using an appropriate validated method (e.g. LC-MS/MS). Once the pharmacokinetics of the drug and the safety of the drug have been reviewed and determined to be acceptable for further development, the plasma profiling can then occur. Based on the parent drug exposure, the dose that most closely approximates to the predicted efficacious dose is selected for further analysis. The plasma samples are pooled across subjects and across time points on a time-weighted basis for all samples collected through 24 hours post dose (for a drug that is intended to be administered qd). Total radioactivity is determined in these pooled samples. The $AUC_{0-24h}$ of the parent drug is compared with that of total radioactivity determined in the pooled sample. If the AUC of parent drug represents >90% of the AUC of total radioactivity, then no profiling of the sample is warranted at this point in time since no single metabolite can represent >10% of the exposure to total radioactivity and hence be classified as a major metabolite as defined in the FDA's guidance on "Safety Testing of Drug Metabolites" (US FDA, 2008j). If the AUC of the parent drug is <90% of the exposure to total radioactivity, profiling of the plasma samples can be performed. The metabolites can be separated by chromatography and the fractions pooled for analysis by AMS. The exposure to the metabolites in humans can be then compared with the exposure in animals and decisions made regarding the need to perform any further testing based on the FDA's guidance on "Safety Testing of Drug Metabolites" (US FDA, 2008j). If nonlinear pharmacokinetics of the parent drug is observed, samples from additional dose levels may be analyzed. With this approach it will be feasible to very rapidly determine the extent of metabolism of a novel drug and therefore determine the optimal design for the clinical pharmacology program for this drug. It should be noted that the data generated in this early type of study will not necessarily generate all of the data that would be generated from a traditional human ADME study. The remaining data could be generated from the samples conducted in this early study or by a dedicated human ADME study (either microradioactivity or traditional study).

In order to describe the metabolism in special populations (e.g. renally impaired, hepatically impaired, geriatric, and pediatric patients), the metabolism may be first described in normal healthy volunteers and the exposure to the major/important metabolites determined subsequently in the special populations. An identical approach can be taken to estimating the effect of environmental factors (e.g. diet, alcohol, and smoking), the effect of food, and the effect of concomitant medications on the metabolism of a given drug.

## 6 BIOLOGICS

The development of a biological drug differs from the development of an NCE in several respects. The standard testing paradigms that are used for an NCE are not appropriate, and a customized approach should be taken. Several guidances and guidelines have been issued on the topic of safety testing of biological products. The primary guidance is the ICHS6 Guidance (Preclinical Safety Evaluation of Biotechnology-Derived Pharmaceuticals) that was issued in 1997 (ICH, 1996b). In 2006, a draft guidance entitled "Nonclinical Safety Evaluation of Biotechnology-Derived Pharmaceuticals" was made available to some pharmaceutical industry researchers. This guidance has not been officially released for review by the FDA but does appear on its list of ongoing or planned guidances for 2008 (US FDA, 2008f). However, this guidance has fostered much debate between the FDA and sponsors on the development of biologics. This "unofficial" draft guidance from the FDA states: "Unlike small molecule drugs, proteins are not metabolized by hepatic cytochrome P450 mechanisms. Instead, they are catabolized into individual amino acids which may then be excreted or reused for protein synthesis or energy production. This poses a problem for studies that rely on radiolabeling, such as mass balance assessments and tissue distribution studies typically performed for small molecules. The radiolabelled amino acid of the biopharmaceutical will be released through catabolism and possibly recycled into normal endogenous protein synthesis. Thus, measurements will likely reflect normal protein turnover. Studies assessing metabolism and excretion are therefore not generally warranted nor are they typically informative for biopharmaceuticals." This latter statement is even more strongly enforced in the ICHS6 guidance (Preclinical Safety Evaluation of Biotechnology-Derived Pharmaceuticals) that was issued in 1997 (ICH, 1996b). This guidance states: "The expected consequence of metabolism of biotechnology-derived pharmaceuticals is the degradation to small peptides and individual amino acids. Therefore, the metabolic pathways are generally understood. Classical biotransformation studies as performed for pharmaceuticals are not needed." The ICHS6 guidance also emphasizes some of the challenges of conducting radiolabeled ADME studies for biologics. The guidance states: "When using radiolabeled proteins, it is important to show that the radiolabeled test material maintains activity and biological properties equivalent to that of the unlabeled material." This is a concept that is not often discussed for small molecules that may be labeled with carbon-14 or tritium. However, it should be noted that the incorporation of a deuterium atom for a hydrogen atom in a small molecule could potentially change the metabolism of a drug and therefore caution should be employed in the use of certain stable isotopes in the assessment of the metabolism of a drug. The cautionary note in ICHS6 arises due to the fact that a common means of inserting a radiolabel into a biologic is by the incorporation of an iodine isotope such as Iodine-131. The effect of iodination on the PK of proteins has been shown by Bauer *et al.* (1996).

ICH S6 goes on to introduce a note of caution regarding the interpretation of radiolabeled studies of biologics by stating: "Tissue concentrations of radioactivity and/or autoradiography data using radiolabeled proteins may be difficult to interpret due to rapid *in vivo* metabolism or unstable radiolabeled linkage. Care should be taken in the interpretation of studies using radioactive tracers incorporated into specific amino acids because of recycling of amino acids into non-drug related proteins/peptides."

The "unofficial" draft guidance from the FDA states that "tissue distribution studies that provide insight into the localization of the administered biopharmaceutical may be useful. For example, tissue distribution studies for imaging agents may be informative since the interest is in identifying organ systems at risk from the radiolabel moiety, not the protein component of the molecule. Thus, distribution of the radiolabel is important because it retains its pharmacological activity, regardless of whether it is maintained within the intact molecule or released as a consequence of protein degradation." The example given relates to a specific subclass of biologics. The vast majority of biologics that are in development at this point in time are monoclonal antibodies. For such agents, the identification of potential organs of toxicity is determined by an evaluation of the ability of the antibody to bind to a range of tissues rather than by an evaluation of the distribution of the monoclonal antibody. This is performed on histopathology slides and allows the relatively easy assessment of the binding of the antibody across a range of species. This is the preferred way in which to evaluate the potential tissue distribution of a monoclonal antibody, primarily due to its ease, low cost, and relatively high throughput. If a traditional tissue distribution study were to be performed for a monoclonal antibody, the study should be performed in the species selected for toxicity testing (i.e. monkey). Conducting a tissue distribution study in the monkey (particularly one that involves an evaluation of the tissue distribution over time) would require a large number of monkeys and would be extremely expensive. Additionally, the equipment that would allow the visualization of the tissue distribution (i.e. quantitative whole body autoradiography [QWBA]) is uncommon. Therefore, for ethical, economic, and practical reasons, classical tissue distribution studies for monoclonal antibodies are not conducted unless the monoclonal antibody is to be used as an imaging agent and an assessment of the potential radiation exposure for humans is required.

ICHS6 also addresses a parameter that is commonly determined for "small molecules" but is often underlooked for biologics. The guidance states: "Understanding the behaviour of the biopharmaceutical in the biologic matrix, (e.g., plasma, serum, cerebral spinal fluid) and the possible influence of binding proteins is important for understanding the pharmacodynamic effect." Exactly how one determines this binding is not addressed and the determination of such binding would not be a trivial task. The ICHS6 guidance is being considered for revision by ICH.

Since the responsibility for the review of biologics (e.g. replacement proteins and antibodies) has been moved from CBER to CDER, many pharmaceutical companies have encountered challenges when interacting with reviewers that have little, if any, experience in developing biologics. These challenges have included duration of chronic non-rodent toxicology studies (which tend to be six months in duration for a biologic per ICH S6 but may be 9 or 12 months for an NCE), the desire for the toxicity of the biologic to be determined in two animal species (even if only one species shows a pharmacological response and if both species that show a pharmacological response are rodents), the need for tissue distribution studies in order to determine potential target organs of toxicity (this is addressed by other means), the need for metabolism studies, and the need for certain *in vitro* studies (i.e. hERG binding and CYP inhibition).

The FDA guidance, "Points to Consider in the Manufacture and Testing of Monoclonal Antibody Products for Human Use" that was issued in 1997 raises the point that "For immunoconjugates containing radionuclides, animal biodistribution data may be used for initial human dose estimation" (US FDA, 1997b).

It further states that "There should be complete accounting of the metabolism of the total dose of administered radioactivity and an adequate number of time points to determine early and late elimination phases. The expected biodistribution and routes of clearance of the administered radiolabeled antibody dose fractions in tissues/organs should be defined. The expected biodistribution and routes of clearance that might occur in the presence of diseases in organs that are critical in radioimmunoconjugate metabolism or excretion should be described. The expected biodistribution and routes of clearance that might occur in the presence of immune responses (e.g., HAMA, HAHA, HARA) should be described. With reference to the radioactive fractions of the administered radiolabeled antibody dose and the patterns of biodistribution, the following issues should be addressed:

- From the biodistribution estimation, the expected residence time of the radiolabeled antibody fractions in the target tissues/organs and non-target tissues/organs should be determined.
- Based on the estimated residence times in each organ, the radiation exposure for each tissue/organ should be estimated.
- Based on the radiation exposure for each tissue/organ, the potential toxicity should be described.
- Based on the potential radiation toxicity to tissues/organs, toxicity monitoring protocols should be developed and incorporated into the clinical trial. If the study has increasing doses of radioactive materials (e.g., a study of the maximum tolerated dose), the radiation exposure for tissues/organs and the associated potential toxicities should be estimated for each radiation dose level."

In January 2007, the EMEA issued the following guideline: "Guideline on the Clinical Investigation of the Pharmacokinetics of Therapeutic Proteins" (EMEA, 2007e). To date, this is the only guidance on this topic, although it should be noted that a book was published in 2006 entitled *Clinical Pharmacology of Therapeutic Proteins* (Mahmood, 2006). The editor of the book was Iftekhar Mahmood of the FDA. All of the authors were employed by the FDA.

The EMEA guideline emphasized the importance of bioanalytical methods by stating:

> One of the key elements of a pharmacokinetic study is the analytical method and its capability to detect and follow the time course of a given analyte (the parent compound and/or metabolites) in a complex biological matrix that contains many other proteins and with satisfactory specificity, sensitivity and a range of quantification with adequate accuracy and precision. The ability to distinguish the therapeutically applied protein from endogenously produced equivalents is an important criteria in selecting the analytical method, although it is recognized that developing an assay that distinguishes between the therapeutic agent and the endogenous molecule may not always be technically feasible.

The guideline again reiterated the message form ICHS6 regarding the catabolism of protein drugs by stating:

> The main elimination pathway should be identified. However, for therapeutic proteins this could be predicted, to a large extent, from the molecular size and specific studies may not be necessary. Catabolism of proteins occurs, usually, by proteolysis. Small proteins of MW < 50,000 Da are eliminated through renal filtration (renal filtration becomes increasingly important the lower the molecular weight) followed by tubular re-absorption and subsequent metabolic catabolism. For larger protein molecules, elimination in other tissues and/or in target cells through for example receptor-mediated endocytosis followed by catabolism is more important relative to renal filtration. Mass-balance studies are not useful for determining the excretion pattern of the drug and drug-related material. Excreted proteins are not necessarily recovered in urine or faeces as intact substance, but are instead metabolised and reabsorbed as amino acids and incorporated in the general protein synthesis.

However, the EMEA guideline went one step further and made several additional points about the metabolism of therapeutic proteins and the consequences of their metabolism. For example, the guideline stated: "The need for, and the feasibility of, specific studies of the route of elimination and metabolism (e.g. microsomal, whole cell or tissue homogenate studies) and identification of metabolites in vitro should be considered and discussed on a case-by-case basis." The guideline did define metabolite as encompassing *in vivo* degradation products and other truncated forms of the protein. The subject of metabolites was further addressed in the following statement:

> Metabolites that have pharmacodynamic activity should preferably be measured, e.g. through chromatographic separation, collection and further *in vivo* bioassay quantification. The metabolites may have different pharmacokinetic profiles compared with the parent compound. However, in cases where measurement of separate active metabolites or peptide fragments is not technically feasible, pharmacokinetics of the active moiety could be determined. Also, measurement of complexes between the protein and other components present in plasma should be considered (Section 2.2). The activity of a therapeutic protein may not only be related to the unbound fraction in plasma but also to bound fractions and to the binding kinetics. Thus, when interpreting the data it is important to understand what fraction is detected in the bioanalysis. It may even be important to analyse the bound fraction per se. Bioassays should be considered, especially if selective immunoassays for metabolites are lacking. Lack of relevant methods should be justified.

The guidance also states:

> It is advantageous if both immunoassay and bioassay are used. If the activity is generated by several species (e.g. metabolites, isoforms), each with different activity, their relative content might change with the degree of renal function due to different degrees of renal clearance. If the different species have similar affinity in the immunoassay, a bioassay allows a more relevant interpretation of the data regarding total activity.
>
> Hepatic impairment: Reduced hepatic function may decrease the elimination of a protein for which hepatic degradation is an important elimination pathway. The lack of studies should be justified by the Applicant (see guideline CPMP/EWP/2339/02).

There are several implications of the statements that have been made in this guidance. These include:

1. Therapeutic proteins may be metabolized to active metabolites and these metabolites should be quantitated.

2. Enhanced bioanalytical techniques may be required since ELISA assays cannot distinguish between parent drug and metabolites of the drug.
3. Methods to measure both drug (and metabolite) concentrations as well as total biological activity (bioassay) are required.
4. The protein therapeutic should be evaluated in similar manner in both humans and animals.
5. Bioanalytical methods may be required that can distinguish between parent drug and truncated versions of the parent drug (present in the drug that is administered) in biological samples.
6. A more expanded set of clinical pharmacology studies may be required (e.g. renal and hepatic impairment). At present, such studies are not routinely conducted.
7. Sponsors may be required to understand how antibodies are catabolized (which organs, which enzyme systems).
8. The focus on hepatic microsomal metabolism studies may lead to a need to assess the potential for these antibodies to inhibit DMEs.
9. The discussion of the need to discuss whether certain studies are required implies that the sponsor will request scientific advice on this topic. The scientific advice process is a long, drawn-out and expensive process.
10. Determination of the binding of the protein therapeutic agent to plasma (or other) proteins should be addressed. Again, this is not routinely done.

The impact of the molecular, biologic, and pharmacokinetic parameters of monoclonal antibodies on early clinical development of mAbs has recently been reviewed by Mascelli *et al.* (2007).

Interestingly, this guideline does not address the possibility that protein drugs, or the consequences of their administration, can affect the regulation of DMEs and hence the pharmacokinetics of the substrates of these enzymes. For example, cytokines have been shown to affect the expression of certain CYP isoforms (Renton, 2001; Aitken, Richardson and Morgan, 2006; Charles *et al.*, 2006). The administration of some protein therapeutics may cause a temporary increase in cytokine levels, resulting in a change in expression of CYP isoforms. However, this phenomenon is described in the FDA draft guidance "Drug Interaction Studies—Study Design, Data Analysis, and Implications for Dosing and Labeling" which was issued in 2006. This draft guidance states: "certain protein therapeutics modify the metabolism of drugs that are metabolized by the P450 enzymes. Type I interferons, for example, inhibit CYP1A2 production at the transcriptional and post-translational levels, inhibiting clearance of theophylline. The increased clinical use of therapeutic proteins may raise concerns regarding the potential for their impacts on drug metabolism." The selection of theophylline as an example is a good one since increased theophylline toxicity was observed in children in King County, Washington, during the 1980 Influenza B outbreak (Kraemer *et al.*, 1982). The role of viral infections on the induction of adverse drug interactions has been reviewed by Levy (1997). Data on the impact of vaccination on the metabolism of drugs is equivocal (Stults and Hashisaki, 1983; Meredith *et al.*, 1985; Hamdy *et al.*, 1995). The effect of disease on hepatic CYP expression has also been observed in human and in a rat model of inflammatory pain (Projean *et al.*, 2005; Charles *et al.*, 2006). Additionally, decreased intestinal CYP3A and p-glycoprotein activities have been observed in rats

with adjuvant-induced arthritis (Uno *et al.*, 2007). The observation that the disease can affect the PK of a drug that is being administered to treat that disease raises the interesting possibility that the PK of the drug will change as the drug works to cure the disease. If this were to occur, it would appear that the animal (or patient) is becoming resistant to the drug.

The impact of interferons on CYP-mediated drug metabolism has been evaluated for some marketed drugs, including PegIntron (Mahmood and Green, 2007). The following statement is taken from the drug label for Pegintron (Schering-Plough Research Institute, 2006):

> The pharmacokinetics of representative drugs metabolized by CYP1A2 (caffeine), CYP2C8/9 (tolbutamide), CYP2D6 (dextromethorphan), CYP3A4 (midazolam), and N-acetyltransferase (dapsone) were studied in 22 patients with chronic hepatitis C who received PegIntron™ (1.5 μg/kg) once weekly for 4 weeks. PegIntron™ treatment resulted in a 28% (mean) increase in a measure of CYP2C8/9 activity. PegIntron™ treatment also resulted in a 66% (mean) increase in a measure of CYP2D6 activity; however, the effect was variable as 13 patients had an increase, 5 patients had a decrease, and 4 patients had no significant change.
>
> No significant effect was observed on the pharmacokinetics of representative drugs metabolized by CYP1A2, CYP3A4, or N-acetyltransferase. The effects of PegIntron™ on CYP2C19 activity were not assessed.

It should be noted that the impact of interferons on the clearance of drugs will likely extend beyond their effects on CYP-mediated metabolism. The impact of interferon-alpha on p-glycoprotein activity in rats has been shown by Ben Reguiga *et al.* and it is not unreasonable to expect that such effects will also be observed in humans (Ben Reguiga *et al.*, 2005; Ben Reguiga, Bonhomme-Faivre and Farinotti, 2007).

DDI studies for therapeutic proteins or monoclonal antibodies have recently been reviewed by Mahmood and Green of the FDA (Mahmood and Green, 2007). They have noted that few DDI studies have been conducted for these classes of drugs, mainly because most macromolecules are not metabolized by CYPs and their mechanism of elimination is complex. The review does describe some DDI studies for monoclonal antibodies. These studies include adalimumab-methotrexate, etanacept-paclitaxel, etanacept-digoxin, etanacept-warfarin, etanacept-methotrexate, trastuzumab-paclitaxel, cetuximab-irinotecan, and panitumumab-irinotecan.

However, this review did not include the study of the effect of Avonex (interferon beta 1-a) on the pharmacokinetics of Tysabri (natalizumab) in MS patients. There was an increase in the $C_{max}$ and half-life of natalizumab by 20% and 70%, respectively, following the administration of Avonex. The clearance of natalizumab decreased by 35% (US FDA, 2004f). However, based on the similarity of the side-effect profile of Tysabri either with or without coadministration of Avonex, no dose adjustment was recommended. This DDI was noted in the original label for Tysabri but is not present in the latest version of the label (Biogen-Idec, 2004, 2006). The reason for this change is unknown.

Over the past decade, the types of materials that may be classed as biologics have expanded from being mainly replacement proteins and antibodies to include antisense RNA, RNAi, and aptamers. The regulatory guidances do not specifically cover these agents with the exception of a reflection paper that was generated by the CHMP SWP. The reflection paper was issued in 2005 and was titled "CHMP SWP Reflection Paper on the Assessment of the Genotoxic Potential of Antisense Oligodeoxynucleotides"

(EMEA, 2005a). ICHS6 does mention these novel agents in passing by stating "The principles outlined in this guidance may also be applicable to recombinant DNA protein vaccines, chemically synthesised peptides, plasma derived products, endogenous proteins extracted from human tissue, and oligonucleotide drugs," although no details are provided as to specific testing that may be required for these agents (ICH, 1996b).

As can be seen, this reflection paper addresses only one aspect of the development of this new class of agents. The reflection paper states: "With regard to oligodeoxynucleotides at least two issues need to be assessed which may indicate a cause for concern, i.e. (a) degradation products of the phosphorothioate oligodeoxynucleotides being nucleotide analogues might lead to mispairing and thus induction of point mutation when integrated into newly synthesized DNA and (b) site-specific mutations might be induced by triplex formation of the oligodeoxynucleotides with the DNA fiber." In this instance, the word "degradation" can be taken to mean both chemical degradation and biological degradation (i.e. metabolism). Therefore, this reflection paper implies that metabolites of oligonucleotide drugs may cause toxicity.

The reflection paper goes on to state: "Although the major metabolic pathways eliminate nucleotides through catabolism, it is possible that phosporothioate mononucleotides may be substrates for various kinases, and thus be incorporated in the triphosphate nucleotide pool. As part of this pool, nucleotide thiotrisposphates could be incorporated into newly synthesized DNA with the consequence that binding to complementary nucleotides occurs with reduced fidelity thus leading to mispairing and eventually point mutations."

Again, the reflection paper alerts the reader to the importance of metabolism in the evaluation of the safety of this type of agent. ICHS6 clearly implies that if you have a biologic drug that degrades into molecules that exist in nature, there will be a low possibility of toxicity. However, many of the oligonucleotide drugs may have been chemically modified and therefore will not be metabolized into natural products. Additionally, oligonucleotide drugs will be metabolized by nucleases which will gradually remove nucleotides from the structure. It is feasible that a metabolite of a parent oligonucleotide drug will simply consist of the parent drug minus one nucleotide residue. This metabolite will likely still retain pharmacological activity and could contribute to overall efficacy.

Similar to the chemically modified oligonucleotide drugs, there exists a class of drugs called "immunoconjugates." These drugs may consist of antibodies to which a toxic agent is conjugated. Mylotarg is an example of an immunoconjugate. Often, the antibody part of the immunoconjugate must be chemically modified before the toxin can be conjugated to the antibody. This is often achieved by attaching a chemical linker to the antibody. Therefore, the protein is no longer a "pure" biotechnology-derived agent and may no longer degrade to compounds that would be regarded as being endogenous substances. Additionally, it is very difficult to control the number of toxin molecules that are conjugated to the protein and the immunoconjugates may consist of a mixture of different conjugates, each containing a different number of toxin molecules per antibody molecule. The toxin molecules will be released in a sequential, albeit random, manner (i.e. a conjugate containing four toxin molecules will be deconjugated to a conjugate containing three toxin molecules). Since the pharmacological activity of the immunoconjugate may depend upon the number of toxin molecules, and that the rate of degradation of the immunoconjugate in the blood will affect the safety profile of the immunoconjugate, it is clear that metabolism plays a hitherto unexplored and critical role in the safety and efficacy of immunoconjugates. Until recently, the existing

bioanalytical technology has not been available to determine the metabolism of immunoconjugates. ELISA assays are traditionally used to quantitate protein-based drugs and are not capable of distinguishing between immunoconjugates that contain one toxin and an immunoconjugate that contains four toxins. However, they can be configured to be able to discriminate between immunoconjugates that contain a toxin moiety and those that do not. Recent advances in MALDI/TOF mass spectrometry has offered the possibility of being able to determine the relative levels of the immunoconjugates (i.e. levels of immunoconjugates containing three or four toxins). However, until technology is developed that allows for the generation of pure immunoconjugate standards (i.e. a conjugate that contains only one toxin molecule), this technology will not be able to readily provide absolute concentration data.

From the examples given, it can be seen that there exists an area in which metabolism plays an important role but for which there is little, if any, regulatory guidance.

## 7 MISCELLANEOUS TOPICS

There are several guidances that mention drug metabolism in passing that do not fit in the previous subsections of this chapter.

### 7.1 CMC-Related

For example, metabolites are mentioned in several CMC guidances, including "Impurities in New Drug Substances Q3A(R)" that states: "Impurities that are also significant metabolites present in animal and/or human studies are generally considered qualified" and "Impurities in New Drug Products Q3B(R2)" that states that "Degradation products that are also significant metabolites present in animal and/or human studies are generally considered qualified" (ICH, 2006a, b). It is clear that an early understanding of the metabolism profile of a drug can aid the qualification of impurities and eliminate the potential for unnecessary toxicology studies to qualify certain impurities or degradation products.

In December 2006, the EMEA has issued a guideline called "Guideline on the Environmental Risk Assessment of Medicinal Products for Human Use" (EMEA, 2006e). This guideline indicates that the environmental impact assessment will require an understanding of the metabolic fate of the drug.

### 7.2 Nanotechnology

The FDA's Nanotechnology Task Force issued a report in July 2007 that indicated the importance of an understanding of the ADME properties on nanoscale materials (US FDA, 2007q). The report stated that, for regulated products:

> There may be a fundamental difference in the kind of uncertainty associated with nanoscale materials compared to conventional chemicals, both with respect to knowledge about them and the way that testing is performed. For conventional chemicals, there is a relatively long history of exploring, and a correspondingly relatively robust understanding, of interactions of molecular classes (such as compounds with particular structures or functional groups) with biological systems. In some cases, screening test methods are used to define what

additional tests may need to be performed to gain sufficient knowledge about safety and/or effectiveness. For example, there are screening tests available to help identify whether DNA damage is a possible outcome from exposure to a certain chemical. Other tests can tell how the chemical is distributed in the body and in what forms it is present in various tissues. FDA has an expectation relevant to molecular forms of materials used in products that FDA regulates that if the molecule does not cause DNA damage during *in vitro* testing, or if it is metabolized quickly and does not reach sensitive organs, or if it is not absorbed, then it is less likely to present certain kinds of health hazards. This expectation is based on long experience with, and consequent understanding of, basic biological interactions of molecular forms of chemicals and of how these interactions correlate with the results of current testing methods.

## 7.3 Liposomes

Another guidance that provides some statements on how to determine the metabolism of a drug is the draft guidance issued by the FDA in 2002 called "Liposome Drug Products Chemistry, Manufacturing, and Controls; Human Pharmacokinetics and Bioavailability; and Labeling Documentation" (US FDA, 2002). Liposomes can be regarded as a nanoengineered product since their average diameter will be 60–100 nm. This draft guidance states that:

The Agency recommends a comparative mass-balance study be performed to define and assess the differences in systemic exposure and pharmacokinetic measures or parameters between liposome and nonliposome drug products when (1) the two products have the same active moiety, (2) the two products are given by the same route of administration, and (3) one of the products is already approved for marketing. The disposition and pathways of elimination (including metabolism and excretion) and several important pharmacokinetic measures ($C_{max}$, AUC) and parameters (clearance, volume, half-life) of a liposomal formulation administered intravenously can be different from that of a nonliposomal formulation given by the same route of administration. Although no examples currently exist, absorption could also be altered for liposome drug product when given via non-intravenous routes. For these reasons, if satisfactory mass balance information is already available for the approved drug product, a limited mass balance study can be undertaken for the proposed drug product. In such a study, the quantity of the drug substance excreted via the major route should be compared in sufficient subjects by giving the liposomal and the nonliposomal formulations, using a crossover or a parallel study design. Comparison of the absorption, distribution, metabolism, and excretion (ADME) of the liposome and nonliposome drug product form should be made, using a crossover or noncrossover study design that employs an appropriate number of subjects. Depending on the drug substance under investigation, the dose of the liposome and nonliposome drug product may be different. The mass balance study should be based on drug substance tagged with a radioactive label (e.g., $^{14}C$, $^{3}H$) before its incorporation into liposomes to allow for sensitive monitoring of radioactive label after administration. Blood (plasma or serum as appropriate), urine, and fecal samples should be collected and assayed for radioactive label.

Other routes of elimination should be monitored as appropriate. Both parent drug substances and any metabolites present should be quantitated. If feasible, mass balance studies can use nonlabeled drug moieties and ingredients. However, CDER recommends that a applicant contact the appropriate review division before conducting studies using nonlabeled drug substance. … Rarely, historical pharmacokinetic data for comparative purposes can be considered on a case-by-case basis in lieu of formal comparative mass balance and/or pharmacokinetic study, taking into account the following factors: (1) when and how the historical data was obtained, (2) similarities of study populations (e.g., disease

condition), (3) analytical procedures, and (4) data analysis. The appropriate CDER review division should be consulted to determine whether historical data can be relied upon.

### 7.4 Racemates

The topic of the development of stereoisomeric drugs is one that is very rarely discussed at the present, although this was a topic of concern in the past. This historical perspective is mentioned in the "FDA's Policy Statement for the Development of New Stereoisomeric Drugs" that was issued in 1992 (US FDA, 1992):

> When stereoisomers are biologically distinguishable, they might seem to be different drugs, yet it has been past practice to develop racemates (i.e., compound with 50:50 proportion of enantiomers). The properties of the individual enantiomers have not generally been well studied or characterized. Whether separated enantiomers should be developed was largely an academic question because commercial separation of racemates was difficult. Now that technological advances (large scale chiral separation procedures or asymmetric syntheses) permit production of many single enantiomers on a commercial scale, it is appropriate to consider what FDA's policy with respect to stereoisomeric mixtures should be. Development of racemates raises issues of acceptable manufacturing control of synthesis and impurities, adequate pharmacologic and toxicologic assessment, proper characterization of metabolism and distribution, and appropriate clinical evaluation. It should be appreciated that toxicity or unusual pharmacologic properties might reside not in the parent isomer, but in an isomer-specific metabolite.

Over the past decade or so, the vast majority of drugs that have entered clinical development have been single enantiomers. However, it should be recognized that the fact that the parent drug is a single enantiomer does not mean that the issue of the relative ADME, pharmacological or toxicological properties of the enantiomers is of no concern. For example, the parent drug can undergo enantiomeric conversion (e.g. ibuprofen) and that the parent drug could be metabolized to different enantiomers. This point is described in the Health Canada guidance (Stereochemical Issues in Chiral Drug Development) (Health Canada, 2000):

> The in vivo stability of the enantiomer must be established. If the antipode is formed in vivo, it should be considered to be a metabolite, and addressed as such during the drug development process. The metabolism and disposition of the enantiomer should be followed using enantioselective methods in each species used in preclinical studies, and in the Phase I studies in humans. If it is established that racemization or inversion does not occur, enantioselective methods may not be needed in all subsequent studies.

The 1994 EU Directive "Investigation of Chiral Active Substances. Legislative basis Directive 75/318/EEC as amended" covers very similar ground as the FDA and Health Canada guidances (European Union, 1994). This Directive describes several scenarios related to the development of enantiomers or racemates. For the development of a single enantiomer as a new active substance, the Directive states:

> An application for marketing authorisation of a single enantiomer as a new active substance should be considered and documented in the same way as any application for a new active substance. Studies should be carried out with the single enantiomer. However, where development studies commence with a racemate, the studies conducted up to the decision to develop the enantiomer may be taken into account to determine the necessity for further

> studies. The possibility of the formation of the other enantiomer "in vivo" should be considered in relation to the chemical structure at an early stage in order to justify the need for any enantiospecific bioanalysis. If the other enantiomer is formed "in vivo" it should be evaluated with its biotransformation products in the same way as for other biotransformation products. In the case of endogenous human chiral compounds, it is considered that data on disposition (ADME) should not necessarily be based upon enantiospecific methods.
>
> The enantiomeric purity of the active substance used in pre-clinical studies should be defined.

With respect to the development of a racemate as a new active substance, the EU Directive states that

> An application for marketing authorization of a new racemic active substance should be considered and documented in the same way as any application for a new active substance.
>
> The choice of the racemate instead of a single enantiomer should be explained. In practice one of the following two situations may occur:
>
> *Rapid interconversion "in vivo"*
>
> - if the interconversion rate "in vivo" is appreciably higher than the apparent distribution and elimination rates of the enantiomers, only the racemate should be studied as the active substance.
>
> No or slow interconversion "in vivo," allowing for separate enantiomer effects on and fate in the organism in most situations, the following studies, in principle, will be needed for evaluation.
>
> - Pharmacodynamics. The profile of the effects related to the therapeutic use should be provided for the racemate and each enantiomer. The racemate data related to the general pharmacodynamic properties should be extended with enantiomer data if necessary from the point of view of safety.
> - Pharmacokinetics. The effective exposure to the enantiomers after administration of the racemate in pivotal studies should be measured by enantiospecific analytical methods allowing extrapolation to the human exposure.
> - Toxicology. It is ordinarily sufficient to carry out toxicity studies on the racemate. If toxicity other than that predicted from the pharmacological properties of the medicinal product occurs at relatively low multiples of the exposure planned for clinical trials, relevant toxicity studies should be repeated with the individual enantiomers when possible.

The EU Directive clearly implies that the sponsor should develop a bioanalytical method that can distinguish between enantiomers and use the assay to evaluate the potential for chiral inversion *in vivo*. Interestingly, the FDA's guidance "Bioanalytical Method Validation" does not appear to address enantiomers and does not seem to require the sponsor to demonstrate the selectivity of the assay with respect to enantiomers (US FDA, 2001b).

The subject of the development of single enantiomers and racemates has been reviewed by several authors (Andersson, 2004; Mansfield, Henry and Tonkin, 2004; Brocks, 2006; Agrawal *et al.*, 2007).

## 8 PRODUCT LABELING

When a drug is approved for marketing, the available safety and efficacy information is summarized in materials that are intended to provide the basis of information for health professionals on how to use the medicinal product safely and effectively. They do not give general advice on the treatment of particular medical conditions. This information is included in either the product label (FDA terminology) or SPC (Summary of Product Characteristics; EU terminology). There are several guidances regarding the level of detail and the format of the information that will be included in the product label or SPC. The FDA has created a set of guidances (US FDA, 2008a) that address selected elements of the product label (e.g. Adverse Reaction sections; Warnings and Precautions, Contraindications, and Boxed Warning Sections; and Clinical Studies sections) or product labeling for selected classes of drugs (e.g. oral contraceptives). The EU has created a high-level guideline that describes the format and content of the SPC (European Commission, 2005). Additionally, the EU has guidelines for the SPC for a range of different classes of drugs (particularly blood products).

Information related to drug metabolism of metabolites can appear in many different places in the product label or SPC. These sections include (but are not limited to):

*Posology and method of administration (SPC, Section 4.2)*
This section will contain information regarding dosage adjustments for special populations such as hepatic/renal impairment or pediatrics.

*Contraindications (SPC, Section 4.3)*
This section will include information on situations where the medicinal product must not be given for safety reasons. These reasons could include demographic factors (e.g. age and gender) or predispositions (e.g. metabolic factors).

Interaction with other medicinal products and other forms of interaction (SPC). This section will also include information specific to a special age group.

*Special warnings and precautions for use (SPC, Section 4.4)*
This section will contain information regarding dosage adjustments for special populations such as hepatic/renal impairment or pediatrics.

*Interactions with other medicinal products and other forms of interaction (SPC, Section 4.5)*
This is one of two sections of the SPC that will contain the greatest extent of drug metabolism-related information. The EU guideline states:

> This section should provide information on the potential for clinically relevant interactions based on the pharmacodynamic properties and *in vivo* pharmacokinetic studies of the medicinal product, with a particular emphasis on the interactions, which result in a recommendation regarding the use of this medicinal product. This includes *in vivo* interaction results which are important for extrapolating an effect on a marker ('probe') substance to other medicinal products having the same pharmacokinetic property as the marker. Interactions affecting the use of this medicinal product should be given first, followed by those interactions resulting in clinically relevant changes on the use of others.

The following information should be given for each clinically relevant interaction:

a. Recommendations: these might be
   - contraindications of concomitant use (cross-refer to section 4.3),
   - concomitant use not recommended (cross-refer to section 4.4), and
   - precautions including dose adjustment (cross-refer to sections 4.2 and 4.4), mentioning specific situations where these may be required; for the actual dose recommendation, cross-refer to section 4.2.

b. Any clinical manifestations and effects on plasma levels and AUC of parent compounds or active metabolites and/or on laboratory parameters.

c. Mechanism, if known.

Interactions not studied *in vivo* but predicted from *in vitro* studies or deducible from other situations or studies should be described if they result in a change in the use of the medicinal product, crossreferring to sections 4.2 or 4.4.

This section should mention the duration of interaction when a medicinal product with clinically important interaction (e.g., enzyme inhibitor or inducer) is discontinued. Adjustment of dosing may be required as a result. The implication for the need for a washout period when using medicines consecutively should also be mentioned. Information on other relevant interactions such as with herbal medicinal products, food or, pharmacologically active substances not used for medical purpose, should also be given. With regard to pharmacodynamic effects where there is a possibility of a clinically relevant potentiation or a harmful additive effect, this should be stated. Results demonstrating an absence of interaction should only be mentioned here if this is of likely major interest to the prescriber. If no interaction studies have been performed, this should be clearly stated.

This section will contain additional information for special populations such as hepatic/renal impairment, elderly or pediatrics in which the impact of the interaction is more severe, or the magnitude of the interaction is expected to be larger.

### *Pregnancy and lactation (SPC, section 4.6)*

The guideline includes the following statements regarding lactation:

"If available, clinical data should be mentioned including the conclusions of the studies on the transfer of the active substance and/or its metabolite(s) into human milk (positive/negative excretion, milk/serum ratio)." "Conclusion on animal studies on the transfer of the active substance and/or its metabolite(s) into milk should be given only if no human data are available."

### *Pharmacokinetic Properties (SPC, section 5.2)*

This section will contain an extensive amount of information regarding the ADME properties of the drug. The guideline states:

Pharmacokinetic properties of the active substance(s) relevant for the advised dose, strength and the pharmaceutical formulation marketed should be given in this section. If these are not available, results obtained with other administration routes, other pharmaceutical forms or doses can be given as alternative.

Basic primary pharmacokinetic parameters, for instance bioavailability, clearance and half-life, should be given as mean values with a measure of variability.

Pharmacokinetics items, which could be included in this section when relevant, are given below.

a. General introduction, information about whether the medicinal product is a pro-drug or whether there are active metabolites, chirality, solubility etc.

b. General characteristics of the active substance(s) after administration of the medicinal product formulation to be marketed.

- Absorption: complete or incomplete absorption; absolute and/or relative bioavailability; first pass effect; $T_{max}$; the influence of food; in case of locally applied medicinal product the systemic bioavailability.
- Distribution: plasma protein binding; volume of distribution; tissue and/or plasma concentrations; pronounced multi-compartment behaviour.
- Biotransformation: degree of metabolism; which metabolites; activity of metabolites; enzymes involved in metabolism; site of metabolism; results from in vitro interaction studies that indicate whether the new compound can induce/inhibit metabolic enzymes.
- Elimination: elimination half-lives, the total clearance; inter and/or intra-subject variability in total clearance; excretion routes of the unchanged substance and the metabolites.
- Linearity/non-linearity: linearity/non-linearity of the pharmacokinetics of the new compound with respect to dose and/or time; if the pharmacokinetics are nonlinear with respect to dose and/or time, the underlying reason for the non-linearity should be presented.

Additional relevant information should be included here.

c. Characteristics in patients

- Variations with respect to factors such as age, gender, smoking status, polymorphic metabolism and concomitant pathological situations such as renal failure, hepatic insufficiency, including degree of impairment. If this influence on the pharmacokinetics is considered to be clinically relevant, it should be described here in quantitative terms (cross-referral to 4.2 when applicable).

d. Pharmacokinetic/pharmacodynamic relationship(s)

- Relationship between dose/concentration/pharmacokinetic parameter and effect (either true endpoint, validated surrogate endpoint or a side effect).
- Contribution (if any) of metabolite(s) to the effect.

The Annex to this guideline provides some generic statements that can be applied to these sections. It should be noted that the annex is currently under revision.

The FDA has provided guidance related to the product label in a different manner than has the EU. As mentioned previously, the FDA has created several guidances that relate to specific sections of the product label. In addition, labeling requirements are listed in 21 CFR Part 201 (US FDA, 2007b).

For example, the FDA issued a draft guidance, "Warnings and Precautions, Contraindications, and Boxed Warning Sections of Labeling for Human Prescription Drug and Biological Products—Content and Format" in January 2006 (US FDA, 2006j). This draft guidance provides information on when to contraindicate the administration of a particular drug. Several examples are provided, including "Likely Clinical Situations." The draft guidance states:

> A contraindication usually involves one or more of the following clinical situations: Comorbid condition or coexistent physiological state (e.g., existing hepatic disease, renal disease, congenital long QT syndrome, hypokalemia, pregnancy or childbearing potential, CYP 2D6 poor metabolizer).

> Coadministered drug where the combination is dangerous (e.g., MAO inhibitor and sympathomimetic drug, a drug known to prolong the QT interval and a drug known to interfere with the metabolism of that drug).

In 2007, the FDA issued a draft guidance entitled "Dosage and Administration Section of Labeling for Human Prescription Drug and Biological Products—Content and Format" (US FDA, 2007n). It should be noted that the title listed in the index of guidances indicates that the title is "Content and Format of the Dosage and Administration Section of Labeling for Human Prescription Drug and Biological Products."

This guidance states that

> When it is important to maintain specific therapeutic blood levels of a drug or its metabolites, whether for effectiveness or safety reasons, the section must identify desirable levels (§ 201.57(c)(3)(i)(J)). The section should describe the monitoring needed to assess levels and how to adjust dose based on observed levels.
>
> Conversely, if a drug interaction is suspected based on a shared metabolic pathway, but there is not enough information to support a specific dosage adjustment recommendation, the interaction should ordinarily not be discussed in the DOSAGE AND ADMINISTRATION section.

Certain elements of drug labeling can be found in other guidances that are not solely devoted to labeling. Such guidances include "Drug Metabolism/Drug Interaction Studies in the Drug Development Process: Studies In Vitro (1997)" which states that

> Each year, large numbers of new drug-drug interactions are discovered, precluding the possibility that any prescriber could memorize them all. Based on the increasing amount of valuable information that is available, it is now possible to label for class effects for various enzymes, and the ability to extrapolate from partial data is growing. Standardized approaches to labeling are likely to emerge and be helpful, in a manner analogous to the class labeling used for certain categories of drugs. For example, certain powerful inhibitors (quinidine for CYP450 2D6, ketoconazole for CYP450 3A4) are likely to affect all drugs metabolized by these pathways. For this reason, if a new drug is found to be a substrate for certain CYP450 enzymes, then certain interactions may be anticipated, even though specific data are lacking. This understanding relies on knowledge about the activity of the drug and its metabolites.
>
> Similarly, it would be helpful to know what metabolic pathways are not involved in the elimination of a drug. When generalizations are made from studies *in vitro*, the conditions of extrapolation should be explicitly stated. Thus, conclusions based on data gained from *in vitro* studies that are extrapolated to the clinical situation should be identified and distinguished from conclusions based on clinical observations *in vivo*. Under these circumstances, the best advice available at any given time may be provided, and class effects may be updated as new information is obtained.

The guidance provides an example of class labeling based on studies *in vitro*:

> Although clinical studies have not been conducted, on the basis of this drug's metabolism by CYP450 3A4, ketoconazole, itraconazole, erythromycin, and grapefruit juice are likely to inhibit its metabolism. Furthermore, rifampin, dexamethasone, and certain anticonvulsants (phenytoin, phenobarbital, carbamazepine) may induce this drug's metabolism. Thus, if a patient has been titrated to a stable dosage on this drug, and then begins a course of treatment with one of these inducers or inhibitors, it's reasonable to expect that a dose adjustment may be necessary to prevent toxicity or therapeutic failure.

The guidance also provides an example of where the class effects would be inserted and also where information on the drug's inhibitory effects would be stated:

> This drug is metabolized by CYP450 3A4 ⟨insert current statement⟩. At clinical doses, the drug itself does not inhibit the metabolism of other 3A4 substrates, but does inhibit the metabolism of substrates metabolized via the CYP450 2D6 pathway.

The guidance makes the point that:

> Given the tendency to include many potential interactions, it is sometimes unclear if anything is noninteracting. In such a circumstance, labeling statements that denote both positive and negative expectations may be helpful.
>
> For example:
>
> This drug is a substrate for CYP450 1A2. Although inhibition of its metabolism by ciprofloxacin is observed, quinidine, erythromycin, ketoconazole, and itraconazole are not inhibitors.

In 1999, the FDA issued a complementary guidance (In Vivo Drug Metabolism/Drug Interaction Studies—Study Design, Data Analysis, and Recommendations for Dosing and Labeling) to address *in vivo* studies.

This guidance states that:

> All relevant information on the metabolic pathways and metabolites and pharmacokinetic interaction should be included in the CLINICAL PHARMACOLOGY section of the labeling.
>
> The consequences of metabolism and interactions should be placed in PRECAUTIONS/WARNINGS, CONTRAINDICATIONS, and DOSAGE AND ADMINISTRATION sections, as appropriate.

It further states that:

> Relevant in vitro and in vivo metabolic drug-drug interaction data describing the drug's effects on substrates and the effects of inhibitors and inducers on the drug should be presented in the DRUG-DRUG INTERACTIONS section of the labeling in the CLINICAL PHARMACOLOGY section, including both positive and important negative findings. The types of studies on which statements are based should be identified briefly in the labeling. If findings indicate a known or potential interaction of clinical significance, or lack of an important interaction that might have been expected, these should be mentioned briefly in the clinical pharmacology interactions section and described more fully in the interaction section under PRECAUTIONS, with advice on how to adjust treatment placed in WARNINGS/PRECAUTIONS, DOSAGE AND ADMINISTRATION, and CONTRAINDICATIONS, as appropriate. In certain cases, information based on clinical studies not using the labeled drug under investigation can be described with an explanation that similar results may be expected for the labeled drug. For example, a strong inhibitor of CYP3A4 does not need to be tested with all 3A4 substrates to warn against an interaction.

Examples of appropriate labeling language are provided in this guidance for various situations. For example, when *in vivo* metabolic drug–drug interaction studies indicate little or no pharmacokinetic effect, the following phraseology should be used:

Data from a drug-drug interaction study involving (drug) and (probe drug) in _____ patients/healthy individuals indicate that the PK disposition of (probe drug) is not altered when the drugs are coadministered. This indicates that (drug) does not inhibit CYP3A4 and will not alter the metabolism of drugs metabolized by this enzyme.

When *in vivo* metabolic drug–drug interaction studies indicate a clinically significant pharmacokinetic interaction:

The effect of (drug) on the pharmacokinetics of (probe drug) has been studied in _____ patients/healthy subjects. The $C_{max}$, AUC, half-life and clearances of (probe drug) increased/decreased by ____ % (90% Confidence Interval: ____ to ____ %) in the presence of (drug). This indicates that (drug) can inhibit the metabolism of drugs metabolized by CYP3A4 and can increase blood concentrations of such drugs. (See PRECAUTIONS, WARNINGS, DOSAGE AND ADMINISTRATION, or CONTRAINDICATIONS sections.)

When specific enzymes have been identified as metabolizing the test drug, but no *in vivo* or *in vitro* drug interaction studies have been conducted:

In vitro drug metabolism studies reveal that (drug) is a substrate of the CYP _____ enzyme. No in vitro or clinical drug interaction studies have been performed to evaluate interactions. However, based on the in vitro data, blood concentrations of (drug) are expected to increase in the presence of inhibitors of _____ such as _____, _____, or_____.

When neither *in vivo* nor *in vitro* drug–drug interaction studies have been conducted and there is no significant metabolism of the drug:

In vivo or in vitro drug-drug interaction studies have not been conducted. The drug interaction potential resulting in changes of PK of (drug) is expected to be low because approximately 90% of the recovered dose of (drug) is excreted in the urine as unchanged drug. However, the role of other pathways of drug elimination, including drug transport systems, is not known. In addition, whether (drug) can inhibit or induce metabolic enzymes is not known. There is potential for drug interactions mediated via modulation of various CYP enzymes.

When *in vitro* interaction has been studied but no *in vivo* studies have been conducted to confirm or refute a finding:

In vitro interaction demonstrated: "In vitro drug interaction studies reveal that the metabolism of (drug) is by CYP3A4 and can be inhibited by the CYP3A4 inhibitor ketoconazole. No clinical studies have been performed to evaluate this finding. Based on the in vitro findings, it is likely that ketoconazole, itraconazole, ritonavir, and other 3A4 inhibitors may lead to substantial increase of (drug) blood concentrations. Refer to PRECAUTIONS, as appropriate.

*In vitro* interaction demonstrated and the substrate drug has substantial first-pass elimination:

In vitro drug interaction studies reveal that the metabolism of (drug) is by CYP3A4 and can be inhibited by the CYP3A4 inhibitor ketoconazole. No clinical studies have been performed to evaluate this finding. Based on the in vitro findings, it is likely that ketoconazole, itraconazole, ritonavir, grapefruit juice, and other 3A4 inhibitors may lead

to substantial increase of (drug) blood concentrations. Refer to PRECAUTIONS, as appropriate.

*In vitro* interaction not demonstrated:

In vitro drug interaction studies reveal no inhibition of the metabolism of (drug) by the CYP3A4 inhibitor ketoconazole. No clinical studies have been performed to evaluate this finding. However, based on the in vitro findings, a metabolic interaction with ketoconazole, grapefruit juice, and other 3A4 inhibitors is not anticipated. Refer to PRECAUTIONS, as appropriate.

Information regarding DDIs may be included in PRECAUTIONS and/or WARNINGS when an interacting drug causes increased concentrations of the substrate but the administration of both drugs may continue with appropriate dosage adjustment. Results of the studies are described in CLINICAL PHARMACOLOGY, DRUG–DRUG INTERACTIONS, PRECAUTIONS and/or WARNINGS and may state:

Drug _____ /class of drug causes significant increases in concentrations of _____ when coadministered, so that dose of _____ must be adjusted (see DOSAGE AND ADMINISTRATION). If there is an important interaction, information for patients should point this out also.

When an interacting drug causes increased risk because of increased concentrations of the substrate and the interacting drug should not be used with the substrate. After describing the interaction in the CLINICAL PHARMACOLOGY section, there should be a CONTRAINDICATIONS section and possibly a boxed warning if the risk is serious.

Drug_____ /class of drug can cause significant increases in concentrations of drug_____ when coadministered. The two drugs should not be used together.

When no *in vitro* or *in vivo* drug interactions were conducted. These are described in CLINICAL PHARMACOLOGY, DRUG–DRUG INTERACTIONS, PRECAUTIONS and/or WARNINGS and may state:

There is potential for drug interactions mediated via modulation of various CYP enzymes.

Additional statement related to DDIs can be found in the section, DOSAGE AND ADMINISTRATION. When an interacting drug causes increased risk because of increased concentrations of the substrate, but the administration for both drugs may continue with suitable monitoring:

Drug _____ /class of drug leads to significant increases in blood concentrations of _____ by _____%. The dose of _____ should be decreased by _____% when the patient is also taking _____. Patients should be closely monitored when taking both drugs.

In the CONTRAINDICATIONS section, DDIs are addressed when an interacting drug causes increased risk because of increased concentrations of the substrate and should not be coadministered:

> Drug _____ /class of drug leads to significant increases in blood concentrations of _____, with potentially serious adverse events. Administration of _____ to patients on drug _____ /class of drug is contraindicated.

As mentioned previously, in 2006 the FDA issued a draft guidance (Drug Interaction Studies—Study Design, Data Analysis, and Implications for Dosing and Labeling) that updates certain aspects of DDI studies.

This draft guidance states that:

> It is important that all relevant information on the metabolic pathways and metabolites and pharmacokinetic interactions be included in the PHARMACOKINETICS subsection of the CLINICAL PHARMACOLOGY section of the labeling. The clinical consequences of metabolism and interactions should be placed in DRUG INTERACTIONS, WARNINGS AND PRECAUTIONS, BOXED WARNINGS, CONTRAINDICATIONS, or DOSAGE AND ADMINISTRATION sections, as appropriate. Information related to clinical consequences should not be included in detail in more than one section, but rather referenced from one section to other sections, as appropriate. When the metabolic pathway or interaction data results in recommendations for dosage adjustments, contraindications, or warnings (e.g., co-administration should be avoided) that are included in the BOXED WARNINGS, CONTRAINDICATIONS, WARNINGS AND PRECAUTIONS, or DOSAGE AND ADMINISTRATION sections, these recommendations should also be included in HIGHLIGHTS.

The guidance then states: "Refer to the guidance for industry on Labeling for Human Prescription Drug and Biological Products – Implementing the New Content and Format Requirements, and Clinical Pharmacology and Drug Interaction Labeling for more information on presenting drug interaction information in labeling."

However, the latter guidance does not seem to be located on the FDA's guidance page.

Finally, this draft guidance addresses the subject of class effects:

> In certain cases, information based on clinical studies not using the labeled drug can be described, with an explanation that similar results may be expected for that drug. For example, if a drug has been determined to be a strong inhibitor of CYP3A, it does not need to be tested with all CYP3A substrates to warn about an interaction with sensitive CYP3A substrates and CYP3A substrates with narrow therapeutic range. An actual test involving a single substrate would lead to labeling concerning use with all sensitive and NTR substrates.

The draft guidance provides examples of sensitive CYP3A substrates and CYP3A substrates with narrow therapeutic range and examples of strong, moderate, and weak CYP3A inhibitors.

The guidance indicates that:

> If a drug has been determined to be a sensitive CYP3A substrate or a CYP3A substrate with a narrow therapeutic range, it does not need to be tested with all strong or moderate inhibitors of CYP3A to warn about an interaction with strong or moderate CYP3A inhibitors, and it might be labeled in the absence of any actual study if its metabolism is predominantly by the CYP3A route. Similarly, if a drug has been determined to be a sensitive CYP3A substrate or a CYP3A substrate with a narrow therapeutic range, it does not need to be tested with all CYP3A inducers to warn about an interaction with CYP3A inducers.

Examples of CYP3A inducers include rifampin, rifabutin, rifapentin, dexamethasone, phenytoin, carbamazepine, phenobarbital, and St. John's wort.

A similar classification system can be used for inhibitors of other CYP enzymes. (US FDA, 2006d, 2007f)

Labeling requirements for special populations are also addressed in selected FDA guidances.

In 2003, the FDA issued a guidance called "Pharmacokinetics in Patients with Impaired Hepatic Function: Study Design, Data Analysis, and Impact on Dosing and Labeling" (US FDA, 2003d).

The guidance notes that "A general approach in developing dosage recommendations is appropriately based on the following considerations:

- If the effect of hepatic impairment on the PK of the drug is obvious (e.g., two-fold or greater increase in AUC), dosage adjustments should be recommended in labeling. It should be noted that for prodrugs (i.e., drugs with activity predominantly due to hepatically generated metabolite), it is possible that the dose would be increased, or the dosing interval shortened, in hepatically impaired patients.
- A conclusion that there is no effect (really, no clinically important effect) of hepatic impairment on the drug's PK, would usually be supported by the establishment of one of the following: (1) delineation of no effect boundaries, prior to the conduct of the studies, based on information available for the investigational drug (e.g., dose- and/or concentration-response studies), or (2) in the absence of other information to determine a different equivalence interval, the employment of a standard 90 percent confidence interval of 80–125 percent for AUC and $C_{max}$. FDA recognizes that documentation that a PK parameter remains within an 80–125 percent no effect boundary would be very difficult given the small numbers of subjects usually entered into hepatic impairment studies. If a wider boundary can be supported clinically, however, it may be possible to conclude that there is no need for dose adjustment."

Statements regarding the effect of hepatic impairment on the pharmacokinetics of the drug can be found throughout the product label, including the Clinical Pharmacology, Pharmacokinetics Section. The guidance states that "Information in this section of the labeling should include:

- The mechanism of hepatic elimination (e.g., enzyme pathways, glucuronidation, biliary excretion)
- The percent of drug that is eliminated by these mechanisms (e.g., metabolism, biliary excretion)
- The disposition of active metabolites in patients with impaired hepatic function, if applicable
- The effects of hepatic impairment on protein binding of parent drug and metabolites, if applicable"

These statements raise an important question, namely, "What is the appropriate route for the determination of the mass balance/*in vivo* metabolism of a drug in humans?" Studies to address the enzyme pathways involved in the hepatic elimination

of the drug may be performed *in vitro*. An evaluation of the contribution of biliary elimination to hepatic elimination and the percent of drug that is eliminated by metabolism or biliary excretion may require the collection of bile samples if the drug is administered orally or the collection of bile and/or fecal samples if the drug is administered intravenously. For an orally administered drug, determination of the hepatic elimination of the drug is not a trivial task.

The FDA provides examples of labeling for different outcomes of the study in hepatically impaired patients. If studies show an effect of altered hepatic function, the guidance states:

> For drugs in which PK or PD is influenced by hepatic impairment, the following statement can be modified as appropriate and in accordance with what is known about the drug (e.g., racemate with different activity of stereoisomers, active or toxic metabolite) and from the studies performed in accordance with this guidance.
>
> *The disposition of _____ was compared in patients with hepatic impairment and subjects with normal hepatic function. Total body clearance of [unbound, if applicable] _____ / metabolite was reduced by _____ % in patients with moderate (as indicated by the Child-Pugh method) hepatic impairment. The half-life of _____ /metabolite is prolonged by ____ in patients with moderate hepatic impairment. Protein binding of _____ /metabolite [is/is not] affected by impaired hepatic function. The drug/metabolite accumulates to the extent of _____ in patients with impaired hepatic function on chronic administration. The dosage should be reduced in patients with mild and moderate hepatic impairment receiving _____. _____ should be [contraindicated/used with great caution] in severe hepatic impairment (see WARNINGS/PRECAUTIONS, CONTRAINDICATION and DOSAGE AND ADMINISTRATION).*

For the situation of "unknown hepatic elimination," the guidance states that the sponsor "consider the compound as extensively metabolized and use the above format."

The FDA guidance, "Pharmacokinetics in Patients with Impaired Renal Function—Study Design, Data Analysis, and Impact on Dosing and Labeling" provides clear information on the labeling of drugs for use in this patient population (US FDA, 1998b). It should be noted that the FDA's guidance page lists the title of this guidance as being "Pharmacokinetics in Patients with Impaired Renal Function." The guidance provides some general thoughts, including: "Specific dosing recommendations should be constructed based on the study results using the aforementioned model for the relationships between renal function and relevant PK parameters. Typically the dose is adjusted to produce a comparable range of unbound plasma concentrations of drug or active metabolites in both normal patients and patients with impaired renal function. Simulations are encouraged as a means to identify doses and dosing intervals that achieve that goal for patients with different levels of renal function." It is interesting to note that this is one of the few guidances that clearly encourages the use of simulations, a topic that is gaining importance at the FDA.

In addition to general thoughts, the guidance makes some specific recommendations. "The labeling should reflect the data pertaining to the effect of renal function on the pharmacokinetics and pharmacodynamics (if known) obtained from studies conducted. The various permutations of intrinsic drug characteristics and the effect of renal impairment on drug performance preclude precise specification of how such drugs should be labeled."

The guidance provides general suggestions on which sections of the labeling should include standardized information and how such information should be structured. Such information includes:

"The pharmacokinetics subsection should include information on the:

- Mechanism of renal elimination (e.g., filtration, secretion, active reabsorption)
- Percentage of drug that is eliminated by renal excretion and whether it is eliminated unchanged or as metabolites;
- Disposition of metabolites in patients with impaired renal function (if applicable);
- Effects of renal impairment on protein binding of parent drug and metabolites (if applicable);
- Effects of changes in urinary pH or other special situations that should be mentioned (e.g., tubular secretion inhibited by probenecid);
- If applicable, the effects of impaired renal function on stereospecific disposition of enantiomers of a racemic drug product should be described if there is evidence of differential stereoisomeric activity or toxicity."

The Special Populations subsection should "recapitulate, in brief, the pharmacokinetic changes found in various degrees of renal impairment and, if necessary, dosing adjustments for patients with varying degrees of renal impairment. This information should be based on the studies performed as described in this guidance. Reference should be made to the PRECAUTIONS/WARNINGS and the DOSAGE AND ADMINISTRATION sections." The guidance provides examples of appropriate wording for these sections.

The simplest situation involves drugs for which impaired renal function has little or no effect on PK:

*Impaired renal function has little or no influence on _____ pharmacokinetics and no dosing adjustment is required.*

Similarly, for drugs whose PK is influenced by renal impairment, the following statement may be modified as appropriate and in accordance with what is known about the drug (e.g., racemate with different activity of stereoisomers, active or toxic metabolite) and from the studies performed in accordance with this guidance:

*The disposition of _____ was studied in patients with varying degrees of renal function. Elimination of the drug (and metabolite, if applicable) is significantly correlated with the creatinine clearance. Total body clearance of (unbound, if applicable) _____ /metabolite was reduced in patients with impaired renal function by _____ % in mild (CLcr = _____–_____ mL/min), _____ % in moderate (CLcr = _____–_____ mL/min) and _____ % in severe renal impairment (CLcr = _____–_____ mL/min), and _____ % in patients under dialysis compared to normal subjects (CLcr >_____mL/min). The terminal half-life of _____/metabolite is prolonged by _____, _____ , and _____ fold in mild, moderate, and severe renal impairment, respectively.*

[Alternatively, the relationship between renal function and the PK parameters may be described in terms of equations, e.g., a linear equation relating unbound clearance and CLcr.]

*Protein binding of _____/metabolite is/is not affected by decreasing renal function. The drug/metabolite accumulates in patients with impaired renal function on chronic administra-*

*tion. The pharmacologic response is/is not affected by renal function. Approximately _____ % of the drug/metabolite in the body was cleared from the body during a standard 4-hour hemodialysis procedure. The dosage should be reduced in patients with impaired renal function receiving _____ and supplemental doses should/should not be given to patients after dialysis. (See DOSAGE AND ADMINISTRATION).*

Information on the effect of renal impairment on PK should also be included in the Precautions/Warnings section. The Dosage and Administration section will also include statements (as appropriate), including a statement on the relationship between drug clearance and endogenous creatinine clearance; the need for dosage adjustment and special considerations for combination drug products. Additionally, information may be included in the Overdosage section. The guidance states:

Although the primary objective of a hemodialysis study is to evaluate the need for dosing adjustments in ESRD (end stage renal disease), additional information regarding the value of hemodialysis in overdose situations may reasonably be garnered from such studies (if performed). In situations in which this information is known, the following wording may be adapted as appropriate:

*_____ is not eliminated to a therapeutically significant degree by hemodialysis.*

or

*Standard hemodialysis procedures result in significant clearance of _____ and should be considered in cases of life-threatening overdose.*

A guidance called "Content and Format for Geriatric Labeling" was issued by the FDA in 2001. However, this guidance does not provide any novel points related to drug metabolism and pharmacokinetics that are not already covered in other guidances.

In 2005, the FDA issued a draft guidance entitled "Clinical Lactation Studies – Study Design, Data Analysis, and Recommendations for Labeling" (US FDA, 2005g). This draft guidance made the following statements regarding labeling:

Clinical Pharmacology, Pharmacokinetics Subsection

This section would include information pertinent to lactation on the:

- Disposition of parent drug and metabolites, if applicable
- Effects of lactation on protein binding, if applicable.

*Similarly, for drugs whose PK is influenced by lactation, the following statement can be modified in accordance with what is known about the drug (e.g., active or toxic metabolite) and from the studies performed in accordance with this guidance:*

*The disposition of [Drug X] was studied in [number of] lactating women from [a through b months postpartum]. Elimination of the drug (and metabolite, if applicable) is significantly changed during lactation. Total body clearance of (unbound, if applicable) [Drug X]/metabolite was [reduced/increased] in lactating women compared to non-lactating women. The terminal half-life of [Drug X]/metabolite is [prolonged/decreased] by [Y-fold]. (See DOSAGE AND ADMINISTRATION.)*

Information should also be included in the Precautions/Nursing Mothers section. "In addition to standard labeling for use in lactation, if studies performed during lactation demonstrate clinically important changes, the Agency recommends that such infor-

mation be included in the PRECAUTIONS/NURSING MOTHERS section with cross-reference to DOSAGE AND ADMINISTRATION and CLINICAL PHARMACOLOGY sections. It is recommended that labeling contain information, to the extent possible, based on the lactation study conducted, including:

- PK/PD in lactation
- The effect of drug on milk production (e.g., quality and quantity of milk including milk production and composition)
- The presence of drug or metabolite in milk, including the limitation of the assay used if drug/metabolites are not detected in milk
- The amount of drug or metabolite in breast milk over a 24-hour period
- The amount of drug or metabolite consumed daily by the breast-fed infant
- The percent of maternal dose delivered via breast milk and consumed daily by the breast-fed infant (i.e., daily dose in human milk compared to the usual adult dose, or pediatric dose, if known)
- Possible ways to minimize exposure in the breast-fed child to drug via breast milk taking into account drug kinetics such as half-life in milk (e.g., timing of maternal dose relative to breast-feeding, the duration to discard breast milk relative to maternal dose, and how long to wait until resuming breast-feeding relative to maternal dose)
- Effects of drug exposure via breast milk in the breast-fed infant
- PK of drug in the breast-fed infant."

In 2004, the FDA issued a guidance called "Pharmacokinetics in Pregnancy—Study Design, Data Analysis, and Impact on Dosing and Labeling" (US FDA, 2004c). This guidance included recommendations related to labeling of the use of the drug in pregnancy. For the Clinical Pharmacology section (Pharmacokinetics subsection) the guidance stated that:

> It is recommended that this section include information pertinent to pregnancy such as:
>
> - Disposition of parent drug and metabolites, if applicable
> - Effects of pregnancy on protein binding of parent drug and metabolites, if applicable
> - Effects of changes in urinary pH or other special situations (e.g., tubular secretion inhibited by probenecid).
>
> Similarly, for drugs whose PK is influenced by pregnancy, the statement similar to the following can be modified as appropriate and in accordance with what is known about the drug (e.g., active or toxic metabolite) and from the studies performed in accordance with this guidance:
>
> *The disposition of [Drug X] was studied in [number of] pregnant patients [in y trimester or from a through b weeks gestation]. Elimination of the drug (and metabolite, if applicable) is significantly changed during pregnancy. Total body clearance of (unbound, if applicable) [Drug X]/metabolite was reduced/increased in pregnant patients compared to [healthy postpartum women, the same women prior to pregnancy or c weeks postpartum]. The terminal half-life of [Drug X]/metabolite is [prolonged/decreased] by Y-, and Z- fold in second and third trimesters, respectively. Protein binding of [Drug X]/metabolite [is/is not] affected by pregnancy. The [drug/metabolite accumulates/does not accumulate] in pregnant patients on*

*chronic administration resulting in increased/decreased plasma levels of drug/metabolite. The pharmacologic response [is/is not] affected by pregnancy. The dosage/dosing interval should be [decreased/increased] in pregnant patients receiving [Drug X] (see DOSAGE AND ADMINISTRATION).*

The EMEA issued a draft guideline (Guideline on Risk Assessment of Medicinal Products on Human Reproduction and Lactation: From Data to Labelling) in 2006 that covers similar topics to the two FDA guidances described earlier (EMEA, 2006c). The CHMP work plan indicates that this guideline will be finalized in 3Q2007/1Q2008.

The FDA issued a guidance called "Pharmacogenomic Data Submissions" in 2007 (US FDA, 2005f). This guidance states that "The pharmacogenomic data and resulting test or tests may be intended to be included in the drug labeling to choose a dose and dose schedule, to identify patients at risk, or to identify patient responders. Inclusion of a pharmacogenomic test in the labeling would be contingent upon its performance characteristics. For example:

- Patients will be tested for drug metabolism genotype and dosed according to the test results.
- Patients will be selected as potential responders for an efficacy trial (or deselected because of a high risk) based on genotype (e.g., of either the patient or the patient's tumor) or gene expression profile.
- Patients will be excluded from a clinical trial based on genotype or gene expression profile (e.g., biomarker for risk of an adverse event)."

In all of these cases, FDA recommends co-development of the drug and the pharmacogenomic tests, if they are not currently available, and submission of complete information on the test/drug combination to the Agency. The FDA plans to issue further guidance on co-development of pharmacogenomic tests and drugs.

In October 2007, the EMEA issued a draft guideline called "Guideline on the Clinical Development of Medicinal Products for the Treatment of HIV Infection" (EMEA, 2007f). Annex A (Presentation of pharmacokinetic interaction data in the SPC) to this draft guideline provides recommendations that refer to antiretroviral drugs with high propensity for pharmacokinetic (PK) interactions (EMEA, 2007a). The draft guideline states that "The principles guiding data presentation should take the following into account:

The SPC is a tool to be used by the clinicians. For compounds with complex interaction potential, the most user-friendly way to present data is by therapeutic areas.

- The aim should be to provide clear recommendations as regards use/non-use and, for essential drugs, which dose to be used.
- If major interactions with a specific compound have been identified, there may be alternatives within the same therapeutic area without this interaction propensity. Therefore absence of PK interactions is informative and should be provided for therapeutic areas where (potentially) problematic interactions have been identified.
- For some compounds (e.g. substrates of CYP3A), the number of possible combinations of interacting compounds might be high in clinical practice. For such compounds therapeutic drug monitoring (TDM) might be useful. Information as regards target concentrations may be put forward in sections 5.2 and/or 4.2,

depending on the robustness of data and foreseen need for TDM in clinical practice."

Although all relevant ADME information required to aid the physician in appropriately prescribing the drug should be included in the product label at the time of marketing approval, this is not always the case. However, not all of the information for the latter may be available as is often the case with accelerated approvals. The FDA may give approval for the lawful marketing of the drug but with certain caveats. These caveats may be statements in the label that no information is available regarding the administration of a given drug with the marketed drug or in the form of a letter to the sponsor indicating a list of additional studies that will be required in the post-marketing (phase IV) phase.

Such phase IV commitments are shown in the approval letter issued by the FDA to Pfizer regarding the marketing approval for maraviroc (a CCR5 co-receptor antagonist that is used for the treatment of HIV infections) (US FDA, 2007r). The FDA approved maraviroc based on its risk–benefit analysis and the unmet medical need. In this letter, the FDA requested that Pfizer conduct the following studies:

- Conduct a study to evaluate the effect of renal impairment on the pharmacokinetics of maraviroc.
  a) at a dose of 150 mg when combined with a boosted protease inhibitor (e.g. saquinavir/ritonavir) in subjects with mild and moderate renal impairment and subjects with End-Stage Renal Disease (ESRD) that require dialysis.
  b) at a dose of 300 mg alone in subjects with severe renal impairment and subjects with end stage renal disease who require dialysis.

Protocol Submission: December 30, 2007

Final Report Submission: December 30, 2008

- Conduct a study to evaluate the potential for maraviroc metabolite(s) to inhibit CYP2D6 enzymes at a maraviroc dose of 600 mg.

Protocol Submission: December 30, 2007

Final Report Submission: June 30, 2008

- Conduct a study to evaluate the potential of maraviroc to inhibit P-gp.

Protocol Submission: December 30, 2007

Final Report Submission: June 30, 2008

- Conduct a study to investigate the potential for maraviroc to induce CP1A2 [sic].

Protocol Submission: December 30, 2007

Final Report Submission: June 30, 2008

- Conduct and submit a clinical study to evaluate the potential for pharmacodynamic interaction between maraviroc and inhibitors of phosphodiesterase type 5 (PDE5).

Protocol Submission: December 2007

Final Report Submission: June 2008

The FDA clearly found that there was a need for maraviroc in the treatment of HIV and that the information that would be generated by the post-marketing commitment

studies would not presently preclude physicians from prescribing the drug, but that the information from these studies would enhance their ability to safely prescribe the drug.

The FDA has expressed concern regarding the number of post-marketing (phase IV) commitments that have not been completed. As of September 30, 2006 only 11% of the commitments for NDAs/ANDAs have been completed. Of the remainder, 71% are pending, 15% are ongoing, 3% are delayed, and <1% are terminated (US FDA, 2007s). Of the studies that concluded in the period October 1, 2005 to September 30, 2006, 83% of the studies met the commitment, 5% did not meet the commitment, and for 12% of the studies, the study was no longer required or was not feasible.

## 9 PROCEDURAL GUIDANCES

Since knowledge of the ADME characteristics of a drug is an integral part of the risk–benefit analysis for that drug, it is not surprising to find drug metabolism mentioned, albeit often in passing or by implication, in procedural guidances. Such guidances mainly relate to the format and content of submissions, rather than providing input into how the ADME characteristics of the drug should be determined.

The FDA guidance called "Content and Format of Investigational New Drug Applications (INDs) for Phase 1 Studies of Drugs, Including Well-Characterized, Therapeutic, Biotechnology-derived Products" was issued in 1995 and is an example of the lack of visibility for ADME in regulatory guidances (US FDA, 1995). This guidance provides a clear road map for the creation of an IND for submission to the US FDA. Section G of this guidance is titled "Pharmacology and Toxicology Information [21 CFR 312.23(a)(8)]." There is no mention of ADME in the title of this section and there is no section in this guidance that solely deals with ADME-related topics. This section also states that the following pharmacology and toxicology guidance is applicable to all phases of IND development of products covered by this guidance. Although ADME is not specifically mentioned in the title of section G, it is mentioned in subsections of this section. Subsection 1 is titled "Pharmacology and Drug Distribution [21 CFR 312.23(a)(8)(i)]" and states that this section should contain, if known:

1) a description of the pharmacologic effects and mechanism(s) of actions of the drug in animals, and
2) information on the absorption, distribution, metabolism, and excretions of the drug. This latter sentence is the sole mention of ADME in this guidance.

This subsection further states:

> The regulations do not further describe the presentation of these data, in contrast to the more detailed description of how to submit toxicologic data. A summary report, without individual animal records or individual study results, usually suffices. In most circumstances, five pages or less should suffice for this summary. If this information is not known, it should simply be so stated.
>
> To the extent that such studies may be important to address safety issues, or to assist in evaluation of toxicology data, they may be necessary; however, lack of this potential effectiveness information should not generally be a reason for a Phase 1 IND to be placed on clinical hold.

It is of interest to note that pharmacology and drug distribution have been linked together, giving the impression that these two items should be summarized together in the nonclinical section of the IND. This interpretation is enforced by the fact that subsection 2 refers to "Toxicology: Integrated Summary [21 CFR 312.23(a)(8)(ii)(a)]." This subsection also describes the submission of individual study reports as part of the IND, and raised the possibility of submitting either "final fully quality assured" individual study reports or earlier draft toxicological reports of the completed study(ies). It should be noted that the regulation does not specifically require individual toxicology study reports to be submitted. The regulations state that an integrated summary of the toxicological findings should be submitted along with a full tabulation of data from each study in order to allow for a detailed review. In practice, full study reports are submitted since the integrated assessment is often based on the findings of individual studies and therefore a draft or final report is available for submission to the IND. Although the guidance only refers to the submission of toxicology study reports, it should be borne in mind that the Pharm/Tox reviewer at the FDA will be reviewing the pharmacology, toxicology, and ADME data. It is reasonable to expect that the reviewer would have similar expectations regarding the level of documentation for the three components of the nonclinical section. Since the objective of the sponsor is to have the IND to be opened on schedule. The "approval" of an IND is essentially a passive process. If the FDA does not raise objections within 30 days, the IND will be viewed as being open and clinical trials can begin. If the FDA raises questions, these will need to be answered within the 30-day window or the IND will have to be withdrawn or the IND be placed on clinical hold by the FDA. The latter should be avoided at all costs. Therefore, the standard practice of sponsors is to provide the Pharm/Tox reviewer with all of the information that they may reasonably feel would be required to adequately review the application. To this end, the sponsor will not only provide an integrated summary of the pharmacology and ADME information but will also provide the full study reports for the studies that are summarized in the integrated summary. Similarly, although the guidance implies that it would be acceptable to submit no information on the ADME properties on the IND candidate if no such information was known, the practice is to provide a reasonable level of characterization of the ADME properties of the candidate drug. The guidance does not contain any details regarding what kind of information would be required to be summarized. If this guidance was to be written in today's regulatory environment, it would be reasonable to expect that the guidance would contain a far greater emphasis on the characterization of the ADME properties of the IND candidate and would not prescribe exactly what information should be included but would simply say that the level of characterization should be addressed on a "case-by-case" basis.

The subject of the level of documentation that should be submitted to the IND, as well as when the information should be submitted, has long been a topic of great debate and confusion. The level of documentation has been discussed earlier. Related to the timing of the submission of the documentation, the guidance states that full individual toxicology study reports should be available to the FDA, upon request, and individual study reports should be available to FDA, upon request, as final, fully quality-assured documents within 120 days of the start of the human study for which the animal study formed part of the safety conclusion basis. These final reports should contain in the introduction any changes from those reported in the integrated summary. If there are no changes, that should be so stated clearly at the beginning of the final, fully quality-assured report.

If the integrated summary is based upon unaudited draft reports, sponsors should submit an update to their integrated summary 120 days after the start of the human study(ies), identifying any differences found in the preparation of the final fully quality-assured study reports and the information submitted in the initial integrated summary. If there were no differences found, that should be stated in the integrated summary update.

In addition, any new finding discovered during the preparation of the final, fully quality-assured individual study reports that could affect subject safety must be reported to the FDA under 21 CFR 312.32.

The timing of the start of the 120-day clock has caused much debate. The guidance states in one place "within 120 days of the start of the human trial" and in another place states "120 days after the start of the human study(ies)." Additionally, the guidance states that the report should be available "upon request" which implies that the FDA could request the final report without notice. The sponsor would be obligated to submit the report or risk having the IND candidate placed on clinical hold. Since there is no mechanism for the FDA to determine when the clinical trial has started, the time frame suggested in the guidance is unenforceable.

This discrepancy was recognized by the FDA and a Q&A document for this guidance was issued in October 2000 (US FDA, 2000c). In this document, the FDA clarified that the 120-day period is measured based on the date of receipt stamped on the IND submission. In reality, this means that the finalized reports (and updated integrated summary) should be available to the FDA (either submitted to the IND or available upon request) 90 days after approval of the IND (i.e. after as 30-day review period).

Despite this clarification, debate exists to date on the definition of the "120-day rule" since the Q&A document cannot be found on the main FDA guidance page (US FDA, 2008g). The subsection of this web page that relates to investigational new drugs only lists the original guidance on format and content. The Q&A document can be found by searching the FDA web site for "120-day." The search will yield the location of the guidance (US FDA, 2000c). The existence of this document is not widely known in the pharmaceutical industry.

The creation of ADME-related reports for submission to the IND is a topic of debate between discovery and development scientists. If knowledge related to the ADME properties of a drug has been generated and was used by the company in decision making, it is most prudent to ensure that these studies have been written up and are submitted to the IND. This is critically important if the sponsor intends to utilize any of this information as a basis for the design or interpretation of future studies and for inclusion in the investigator's brochure. However, the studies that are conducted in the discovery phase (i.e. as part of a compound nomination package) are often conducted with a small number of animals or replicates. Therefore, consideration should be given to repeating these studies for inclusion in the IND, if appropriate. Some sponsors chose not to repeat these studies until later in the development timeline, once efficacy and safety has been confirmed in the clinic. However, if the studies are not repeated until late in development, it may be difficult to explain discrepancies between the early, discovery-mode studies and the later development studies.

The Pharm/Tox reviewer will review the integrated summary of nonclinical information, the more extensive and detailed summary of ADME information, and potentially individual study reports. All of these documents should contain a level of detail that is sufficient to allow the reviewer to come to the same conclusions that the author did.

In practice, this means that in addition to the appropriate mean data tables, the author should include appropriate individual data points. The intent of these summaries is to clearly and concisely convey the main points that the author is trying to make and is not to show how verbose or clever the author can be. To this end, graphics should be used wherever possible in order to convey a concept. This point is emphasized in the Format and Content guidance. Additionally, the author should appreciate that although the Pharm/Tox will have a good understanding of the science of ADME, this is unlikely to be their main area of expertise. Therefore, a simple and clear writing style is encouraged. Similarly, the author should acknowledge that English may not be the first language of the reviewer and hence the report or summary should be written with the bare minimum, if any, use of jargon or confusing words or phrases. Along the same lines, English may not be the first language of the author of the scientific reports or summaries that form part of the regulatory submissions. Most, if not all, companies employ scientific writing groups to either write or edit these documents. It should be noted that peer-reviewed literature can be submitted to the IND as required and may replace the need for the creation of study reports. This is an interesting loophole, given the difference in the extent of detail that is given in a peer-reviewed journal article and a study report that contains all individual data. The FDA has also issued "Guidance for Reviewers: Pharm/Tox Review Format" (US FDA, 2001c). This guidance does not indicate which specific ADME studies should be conducted, but does imply that ADME information should be available. This guidance contains tables and checklists to help facilitate the review for the Pharm/Tox reviewer.

The format and content for an exploratory IND are very similar to those of a traditional IND and have been described in depth earlier in this chapter.

The process for the approval of clinical trials is somewhat different in the EU than it is in the United States. In the United States, an IND application is made that contains the investigator's brochure; a general investigational plan; the clinical trial protocol for the initial clinical trial; a summary of CMC, nonclinical and clinical data (as applicable) and individual study reports (primarily toxicology reports as noted earlier). For subsequent clinical trials, only the clinical protocol and investigator's brochure need to be submitted. For each clinical trial in Europe, the applicant will provide the investigator's brochure, the clinical trial protocol, and a summary of CMC, nonclinical and clinical information (the investigational medicinal product dossier or IMPD). Individual study reports are not submitted.

In many other guidances, ADME may be mentioned in passing. For example, in the EMEA guidance, "Detailed guidance for the request for authorization of a clinical trial on a medicinal product for human use to the competent authorities, notification of substantial amendments and declaration of the end of the trial" that was issued in October 2005, the only mention of metabolism is that there should be a subsection that describes "Pharmacokinetics of active metabolites" (European Commission, 2003). The Health Canada draft guidance, "Quality (Chemistry and Manufacturing) Guidance: New Drug Submissions (NDSs) and Abbreviated New Drug Submissions (ANDSs)" simply states that "It should also be indicated if the impurity is a metabolite of the drug substance" (Health Canada, 2001).

In 1984, the 505(b)(2) application route for a New Drug Application was created by Congress to allow applicants to create innovative medicines using currently available products without performing a full complement of safety and efficacy studies. This article reviews the history of the approach, provides examples, and considers some of the scientific and technical challenges associated with documenting safety and efficacy

relative to the proposed change. The approach does not appear to have been used extensively in almost 18 years since it was created. The explanation for this is not fully apparent, but may relate to the limited exclusivity, usually three years, allowed for a 505(b)(2) application (Johnston and Williams, 2002). In 1999, the FDA issued a draft guidance, "505(b)(2)" (US FDA, 1999c). The guidance states that "In some cases a new molecular entity may have been studied by parties other than the applicant and published information may be pertinent to the new application. This is particularly likely if the NME is the prodrug of an approved drug or the active metabolite of an approved drug. In some cases, data on a drug with similar pharmacologic effects could be considered critical to approval."

Recently (March 2007), the FDA issued a draft guidance on "Target Product Profile—A Strategic Development Process Tool" (US FDA, 2007h). The Target Product Profile (TPP) is a tool that has been used for many years by pharmaceutical companies as a strategic communication tool that allows for continual dialog between different groups and allows for a common understanding of the desired characteristics of the drug under development and the steps that will be taken to achieve these goals. The FDA is interested in using this tool for the purposes listed earlier, with the purpose of improving the communications between the FDA and the sponsor. The TPP could potentially be a very useful tool that summarizes what is known about a compound, the development plan for the project, and how those goals will be achieved. Although the former is summarized in the investigator's brochure, the latter two topics are not presently gathered in an "evergreen" document. The guidance states: "Describe clinically significant pharmacokinetics of a drug or active metabolites (i.e., pertinent absorption, distribution, metabolism, and excretion parameters). Include results of pharmacokinetic studies that establish the absence of an effect, including pertinent human studies and in vitro data." The guidance also states: "Include a concise factual summary of the clinical pharmacology and actions of the drug in humans. Data that describe the drug's pharmacologic activity can be included in this section, including biochemical or physiological mechanism of action, pharmacokinetic information, degree of absorption, pathway for biotransformation, percent dose unchanged, metabolites, rate of half-lives including elimination concentration in body fluids at therapeutic and toxic levels, degree of binding to plasma, degree of uptake by a particular organ or fetus, and passage across the blood-brain barrier."

In an attempt to standardize the documents that would be submitted for marketing approval throughout the world, the ICH has issued several guidances, including "The Common Technical Document for the Registration of Pharmaceuticals for Human Use: Safety—M4S" (ICH, 2000d).

This guidance describes the sections that should be included in the CTD. These include:

"2.6.4 Pharmacokinetics Written Summary

The sequence of the Pharmacokinetics Written Summary should be as follows:

- Brief Summary
- Methods of Analysis
- Absorption
- Distribution
- Metabolism
- Excretion

- Pharmacokinetic Drug Interactions
- Other Pharmacokinetic Studies
- Discussion and Conclusions
- Tables and Figures (either here or included in text)

2.6.4.5 Metabolism (interspecies comparison)
The following data should be summarised in this section:

- Chemical structures and quantities of metabolites in biological samples
- Possible metabolic pathways
- Pre-systemic metabolism (GI/hepatic first-pass effects)
- In vitro metabolism including P450 studies
- Enzyme induction and inhibition.

2.6.5 Pharmacokinetics
2.6.5.1 Pharmacokinetics: Overview
2.6.5.2 Analytical Methods and Validation Reports*
2.6.5.3 Pharmacokinetics: Absorption after a Single Dose
2.6.5.4 Pharmacokinetics: Absorption after Repeated Doses
2.6.5.5 Pharmacokinetics: Organ Distribution
2.6.5.6 Pharmacokinetics: Plasma Protein Binding
2.6.5.7 Pharmacokinetics: Study in Pregnant or Nursing Animals
2.6.5.8 Pharmacokinetics: Other Distribution Study
2.6.5.9 Pharmacokinetics: Metabolism In Vivo
2.6.5.10 Pharmacokinetics: Metabolism In Vitro
2.6.5.11 Pharmacokinetics: Possible Metabolic Pathways
2.6.5.12 Pharmacokinetics: Induction/Inhibition of Drug-Metabolizing Enzymes
2.6.5.13 Pharmacokinetics: Excretion
2.6.5.14 Pharmacokinetics: Excretion into Bile
2.6.5.15 Pharmacokinetics: Drug-Drug Interactions
2.6.5.16 Pharmacokinetics: Other"

The FDA has created a web page for small businesses (US FDA, 2008l). The information provided includes a "Frequently Asked Questions" document on "Drug Development and Investigational New Drug Applications" (US FDA, 2001a). This document includes the FAQ, "What are the FDA requirements for pre-clinical studies?" The FDA's answer to this question is:

> Under FDA requirements, a sponsor must first submit data showing that the drug is reasonably safe for use in initial, small-scale clinical studies. During preclinical drug development, a sponsor evaluates the drug's toxic and pharmacologic effects through *in vitro* and *in vivo* laboratory animal testing. Genotoxicity screening is performed, as well as investigations on drug absorption and metabolism, the toxicity of the drug's metabolites, and the speed with which the drug and its metabolites are excreted from the body. At the preclinical stage, the FDA will generally ask, at a minimum, that sponsors: (1) develop a pharmacological profile of the drug; (2) determine the acute toxicity of the drug in at least two species of animals, and (3) conduct short-term toxicity studies ranging from 2 weeks to 3 months, depending on the proposed duration of use of the substance in the proposed clinical studies.

It is interesting to note that ADME studies are not mentioned in this latter list. The small business assistance web page also includes a document on "Frequently Asked Questions on the Pre-Investigational New Drug (IND) Meeting" (US FDA, 2005a).

It is of interest to note that even though drug metabolism and metabolites are mentioned, albeit often in passing, in many guidances, there is very little, if any, information included in these guidances regarding how to determine the metabolism of a drug, particularly *in vivo*. Occasionally, guidances or guidelines will provide some tips or insights. For example, the European guideline "Pharmacokinetics and Metabolic Studies in the Safety Evaluation of New Medicinal Products in Animals, Legislative basis Directive 75/318/EEC as amended" states that:

> When a labelled substance is used the position of the label in the molecule and the specific activity of the material must be stated. Consideration should be given when selecting the position of the label to its likely metabolic fate. When using labelled substances attention must be given to the fact that the measured label in body fluids may not correspond to that of the unmodified substance, but may include labelled metabolites and conjugates. Attention should be given to the possibility of isotope exchange with endogenous compounds. (EMEA, 1994b)

This may appear to be an obvious point, but one that is often overlooked. An example of this phenomenon is the exchange of tritium between a radiolabeled pharmaceutical and water.

## 10 INITIATIVES FROM REGULATORY AGENCIES

Guidances or guidelines are not the only tool that is used by regulatory agencies to communicate their expectations to sponsors. These expectations, or concerns, are also communicated by the participation of representatives from the regulatory agencies at scientific conferences; the publication of articles in peer-reviewed journals; the publication of book chapters and the creation of initiatives.

One such example of a regulatory initiative is the FDA's "Critical Path Initiative (Innovation or stagnation)" that was issued in 2004 (US FDA, 2004a). The original document and subsequent updates and related documents can be found on the FDA's web page on the Critical Path Initiative (US FDA, 2008d). In May 2007, the FDA expanded the scope of the Critical Path Initiative and created a document on Critical Path Opportunities for Generic Drugs (US FDA, 2007e).

The Critical Path Initiative document states that:

> Many of FDA's recent targeted efforts have involved working with the scientific community to define more reliable methods to predict and detect significant safety problems. For example, in the past, failure to predict unfavorable human metabolism of candidate drugs has led to costly failures in the clinic as well as multiple drug market withdrawals. FDA recommendations on the use of human cell lines to characterize drug metabolic pathways provide a straightforward in vitro method for prediction of human metabolism, allowing developers to eliminate early on compounds with unfavorable metabolic profiles (e.g., drug-drug interaction potential). Failures in the clinic due to drug interaction problems are now far less likely.

The FDA's Critical Path Opportunities List and Critical Path Opportunities Report that were issued in May 2007 identify several areas of interest to the drug metabolism

scientist. (US FDA, 2006a, b). The Opportunities list includes Biomarker Qualification, Modernizing Predictive Toxicology, Noninvasive Therapeutic Monitoring, Improving Extrapolation from Animal Data to Human Experience, Identifying Safety Effects of Excipients, Characterizing and Qualifying Nanotechnologies, and Drug Metabolism and Therapeutic Response for Pediatrics. The Opportunities list represents an initial list that will be updated as necessary. There are 76 specific scientific opportunities, or priority topics, in the Opportunities report and list.

The FDA also issued a report called "Critical Path Opportunities Initiated During 2006" (US FDA, 2006c). This document lists more than 40 Critical Path collaborations and research activities that currently are under way with FDA participation. It can be seen that there are a large number of initiatives that have been created that will lead to the creation of a concept paper, followed by the creation of a draft guidance, and ultimately a final guidance. Therefore, there will be an increased amount of documentation that the drug development scientists will need to become familiar with. However, it should be noted that the Critical Path Initiative enables the drug development scientists to be part of the process that will ultimately lead to new, or updated, regulatory guidances. This is an opportunity that has not really existed until now.

The Critical Path Opportunities for Generics includes several topics of interest including Model Development and In Vitro-In Vivo Correlations, bioequivalence for novel delivery technologies, bioequivalence for pharmacokinetic profiles with multiple peaks, bioequivalence for transdermal products, bioequivalence for highly variable drugs, bioequivalence for locally acting and targeted delivery drugs, and bioequivalence of liposome-based formulations (US FDA, 2007e; Lionberger, 2008).

The Innovative Medicines Initiative (IMI) issued a report entitled "Strategic Research Agenda Creating Biomedical R&D Leadership for Europe to Benefit Patients and Society" in 2004 (EFPIA, 2004). This report noted that:

> the regulatory bodies hold data on the pharmacokinetics of a large number of drugs. Collective analysis of the data for all substrates of a particular metabolising enzyme e.g. cytochrome 2D6 or 3A4, should provide information not only on the inherent functional variability of these enzymes within the patient population, but also allow one to determine quantitatively the contribution of such factors as age, gender, disease, and inhibitors of these enzymes, to the variability. Armed with this generic information, one should be able to predict a priori the likely variability of the pharmacokinetics of a new drug within the patient population, under a variety of situations, thereby facilitating future design of clinical studies and subsequent product labeling, and also improving the cost-efficiency of such studies. This proposal will require not only inter-company collaboration but also the agreement of EMEA and national bodies to release these data.

If such data were to become available, the regulatory agencies may be able to help facilitate drug development. Such a close partnership would be relatively novel but would greatly benefit the patient in the long run. In 2007, the FDA announced that the formation of the Reagan-Udall foundation which is a private and independent nonprofit organization will advance FDA's mission to modernize medical, veterinary, food, food ingredient, and cosmetic product development, accelerate innovation, and enhance product safety. The 14-member board consists of four representatives from the general pharmaceutical, device, food, cosmetic, and biotechnology industries; three from academic research organizations; two from patient or consumer advocacy groups; one representing health-care providers; and four at-large representatives with expertise or experience relevant to the foundation's purpose. Hopefully, close interaction between

industry and regulators will expedite the development of safe medicines, although concern has been expressed that such a foundation will enable the pharmaceutical industry to exert influence over the FDA.

## 11 SUMMARY

Drug metabolism is playing an ever-increasing role in the drug development process and this is being reflected in the number of guidances that include mention of drug metabolism and the increased role of drug metabolism in clinical trials and the product label. The importance of drug metabolism will continue to increase as we learn more about the role of drug metabolism in special populations, the role of drug metabolism in sources of inter- and intra-individual variability, and the role of drug metabolism in certain toxicities. It is therefore of critical importance that the drug metabolism scientist become aware of the relevant regulatory guidances and take an active role in applying the existing guidances and the creation of new guidances. This is not an easy task, but is a task that will reap great rewards for the pharmaceutical industry in the long term.

## REFERENCES

Abbott Laboratories (2007). *Kaletra Package Insert*, http://www.rxabbott.com/pdf/kaletrapi.pdf (accessed 7 March 2008).

Agrawal, Y.K. *et al.* (2007). Chirality—a new era of therapeutics. *Mini Reviews in Medicinal Chemistry*, *7*, 451–460.

Agundez, J.A.G. (2004). Cytochrome P450 gene polymorphism and cancer. *Current Drug Metabolism*, *5*, 211–224.

Aitken, A.E., Richardson, T.A., Morgan, E.T. (2006). Regulation of drug-metabolizing enzymes and transporters in inflammation. *Annual Review of Pharmacology and Toxicology*, *46* (1), 123–149.

Alexandre, J. *et al.* (2007). Relationship between cytochrome 3A activity, inflammatory status and the risk of docetaxel-induced febrile neutropenia: a prospective study. *Annals of Oncology*, *18* (1), 168–172.

Amgen (2005). *Comments on Draft Guidance for Industry on Safety Testing of Metabolites*, 3 August 2005, http://www.fda.gov/ohrms/dockets/dockets/05d0203/05D-0203-EC3-Attach-1.pdf (accessed 3 March 2008).

Andersson, T. (2004). Single-isomer drugs: true therapeutic advances. *Clinical Pharmacokinetics*, *43* (5), 279–285.

Ando, Y., Fuse, E., Figg, W.D. (2002). Thalidomide metabolism by the CYP2C subfamily. *Clinical Cancer Research*, *8* (6), 1964–1973.

Anon (2006). Slow start to phase 0 as researchers debate value. *Journal of the National Cancer Institute*, *98* (12), 804–805.

Ariyoshi, N. *et al.* (2001). Comparison of the levels of enzymes involved in drug metabolism between transgenic or gene-knockout and the parental mice. *Toxicologic Pathology*, *29* (Suppl.), 161–172.

Arlen, R.R., Wells, P.G. (1996). Inhibition of thalidomide teratogenicity by acetylsalicylic acid: evidence for prostaglandin H synthase-catalyzed bioactivation of thalidomide to a teratogenic reactive intermediate. *The Journal of Pharmacology and Experimental Therapeutics*, *277* (3), 1649–1658.

Association of the British Pharmaceutical Industry (2007). Guidelines for Phase I Clinical Trials. 2007 Edition.

Baillie, T.A. *et al.* (2002). Drug metabolites in safety testing. *Toxicology and Applied Pharmacology*, *182* (3), 188–196.

Baillie, T.A. *et al.* (2003). Reply. *Toxicology and Applied Pharmacology*, *190* (1), 93–94.

Baldrick, P. (2001). Preclinical considerations for regulatory submissions. *Drug Information Journal*, *35*, 99–105.

Bauer, R.J. *et al.* (1996). Alteration of the pharmacokinetics of small proteins by iodination. *Biopharmaceutics and Drug Disposition*, *17* (9), 761–774.

Ben Reguiga, M. *et al.* (2005). Modification of the P-glycoprotein dependent pharmacokinetics of digoxin in rats by human recombinant interferon-alpha. *Pharmaceutical Research*, *22* (11), 1829–1836.

Ben Reguiga, M., Bonhomme-Faivre, L., Farinotti, R. (2007). Bioavailability and tissular distribution of docetaxel, a P-glycoprotein substrate, are modified by interferon-alpha in rats. *Journal of Pharmacy and Pharmacology*, *59* (3), 401–408.

Bertino, J.S., Greenberg, Jr, H.E., Reed, M.D. (2007a). American college of clinical pharmacology position statement on the use of microdosing in the drug development process. *Journal of Clinical Pharmacol*, *47* (4), 418–422.

Bertino, J.S., Greenberg, Jr, H.E., Reed, M.D. (2007b). Response to comments on ACCP position statement on microdosing. *Journal of Clinical Pharmacol*, *47* (12), 1597–1598.

Beumer, J.H., Beijnen, J.H., Schellens, J.H.M. (2006). Mass balance studies, with a focus on anticancer drugs. *Clinical Pharmacokinetics*, *45* (1), 33–58.

Beumer, J.H. *et al.* (2007). Human mass balance study of the novel anticancer agent ixabepilone using accelerator mass spectrometry. *Investigational New Drugs*, *25*, 327–334.

Biogen-Idec (2004). *Tysabri. Prescribing Information*, http://www.fda.gov/cder/foi/label/2006/125104s015LBL.pdf (accessed 8 March 2008).

Biogen-Idec (2006). *Tysabri. Prescribing Information*, http://www.fda.gov/cder/foi/label/2004/125104lbl.pdf (accessed 8 March 2008).

Bjornsson, T.D. *et al.* (2003). The conduct of *in vitro* and *in vivo* drug–drug interaction studies: a pharmaceutical research and manufacturers of America (PhRMA) perspective. *Drug Metabolism and Disposition*, *31* (7), 815–832.

Bonate, P.L., Howard, D.R. (2004). *Pharmacokinetics in Drug Development: Regulatory and Development Paradigms*, AAPS Press.

Bowe, C. (2007). Wyeth Says Regulator Creating Monopolies. *Financial Times*, London. http://www.ft.com/cms/s/0/4aa84080-8b10-11dc-95f7-0000779fd2ac.html?nclick_check=1.

Boyd, R.A., Lalonde, R.L. (2007). Nontraditional approaches to first-in-human studies to increase efficiency of drug development: will microdose studies make a significant impact? *Clinical Pharmacol Therapy*, *81* (1), 24–26.

Brent, R.L. (2004). Utilization of animal studies to determine the effects and human risks of environmental toxicants (drugs, chemicals, and physical agents). *Pediatrics*, *113* (4), 984–995.

Brocks, D.R. (2006). Drug disposition in three dimensions: an update on stereoselectivity in pharmacokinetics. *Biopharmaceutics and Drug Disposition*, *27* (8), 387–406.

Brunner, M., Langer, O. (2006). Microdialysis versus other techniques for the clinical assessment of *in vivo* tissue drug distribution. *The AAPS Journal*, *8* (2), E263–E271.

Buehler, G.J. (2007). Letter to Sponsor of Pending Abbreviated New Drug Application(s) (ANDAs).

Cao, Y.J. *et al.* (2007). Effect of semen sampling frequency on seminal antiretroviral drug concentration. *Clinical Pharmacol Therapy*, *83*, 848–856.

Chainuvati, S. *et al.* (2003). Combined phenotypic assessment of cytochrome P450 1A2, 2C9, 2C19, 2D6, and 3A, N-acetyltransferase-2, and xanthine oxidase activities with the [ldquo]Cooperstown 5+1 cocktail [rdquo] [ast]. *Clinical Pharmacol Therapy*, *74* (5), 437–447.

Charles, K.A. *et al.* (2006). Transcriptional repression of hepatic cytochrome P450 3A4 gene in the presence of cancer. *Clinical Cancer Research*, *12* (24), 7492–7497.

Chaurand, P., Corentt, D., Caprioli, R. (2006). Molecular imaging of thin mammalian tissue sections by mass spectrometry. *Current Opinion in Biotechnology*, *17*, 431–436.

Cheung, B. *et al.* (2005). The application of sample pooling methods for determining AUC, AUMC and mean residence times in pharmacokinetic studies. *Fundamental and Clinical Pharmacology*, *19*, 347–354.

Chung, N.S., Wasan, K.M. (2004). Potential role of the low-density lipoprotein receptor family as mediators of cellular drug uptake. *Advanced Drug Delivery Reviews*, *56* (9), 1315–1334.

Correia, M.A. (2003). Hepatic cytochrome P450 degradation: mechanistic diversity of the cellular sanitation brigade. *Drug Metabolism Reviews*, *35* (2), 107–143.

Correia, M.A., Sadeghi, S., Mundo-Paredes, E. (2005). Cytochrome P450 ubiquitination: branding for the proteolytic slaughter? *Annual Review of Pharmacology and Toxicology*, *45* (1), 439–464.

Dain, J., Collins, J., Robinson, W. (1994). A regulatory and industrial perspective of the use of carbon-14 and tritium isotopes in human ADME studies. *Pharmaceutical Research*, *11* (6), 925–928.

Danko, I., Chaschin, N. (2005). Association of CYP2E1 gene polymorphism with predisposition to cancer development. *Experimental Oncology*, *27* (4), 248–256.

Dannan, H. (2003). Applying good laboratory practice regulations, in *Pharmaceutical Technology Europe*, Advanstar Publications.

Darpo, B., Nebout, T., Sager, P.T. (2006). Clinical evaluation of QT/QTc prolongation and proarrhythmic potential for nonantiarrhythmic drugs: the international conference on harmonization of technical requirements for registration of pharmaceuticals for human use E14 guideline. *Journal of Clinical Pharmacol*, *46* (5), 498–507.

Davis-Bruno, K.L., Atrakchi, A. (2006). A regulatory perspective on issues and approaches in characterizing human metabolites. *Chemical Research in Toxicology*, *19* (12), 1561–1563.

Davit, B. *et al.* (2008). Highly variable drugs: observations from bioequivalence data submitted to the FDA for new generic drug applications. *The AAPS Journal*, *10* (1), 148–156.

DeGeorge, J.J. *et al.* (1997). Regulatory considerations for preclinical development of anticancer drugs. *Cancer Chemotherapy and Pharmacology*, *41* (3), 173–185.

DeGregorio, M. *et al.* (2006). Accelerator mass spectrometry allows for cellular quantification of doxorubicin at femtomolar concentrations. *Cancer Chemotherapy and Pharmacology*, *57* (3), 335–342.

Dodds, P.F. (1991). Incorporation of xenobiotic carboxylic acids into lipids. *Life Sciences*, *49* (9), 629–649.

Dodds, P.F. (1995). Xenobiotic lipids: the inclusion of xenobiotic compounds in pathways of lipid biosynthesis. *Progress in Lipid Research*, *34* (3), 219–247.

Dressman, J. *et al.* (1998). Dissolution testing as a prognostic tool for oral drug absorption: immediate release dosage forms. *Pharmaceutical Research*, *15* (1), 11–22.

Dumez, H. *et al.* (2004). The relevance of therapeutic drug monitoring in plasma and erythrocytes in anti-cancer drug treatment. *Clinical Chemistry and Laboratory Medicine*, *42* (11), 147–173.

European Commission (2001). Directive 2001/20/EC of the European Parliament and of the Council of 4 April 2001 on the Approximation of the Laws, Regulations and Administrative Provisions of the Member States Relating to the Implementation of Good Clinical Practice in the Conduct of Clinical Trials on Medicinal Products for Human Use.

European Commission (2003). Detailed Guidance for the Request for Authorisation of a Clinical Trial on a Medicinal Product for Human use to the Competent Authorities, Notification of Substantial Amendments and Declaration of the End of the Trial.

European Commission (2005). A Guideline on Summary of Product Characteristics Rev 1.

European Commission (2008). *EUDRALEX. Volume 1—Pharmaceutical Legislation: Medicinal Products for Human Use*, http://ec.europa.eu/enterprise/pharmaceuticals/eudralex/homev1.htm (accessed 4 March 2008).

European Economic Community (1987). Pharmacokinetic Studies in Man. Legislative Basis Directive 75/318/EEC as Amended.

European Federation of Pharmaceutical Industries and Associations (EFPIA) (2004). Strategic Research Agenda Creating biomedical R&D Leadership for Europe to Benefit Patients and Society.

European Medicines Agency (EMEA) (1994a). Note for Guidance on Studies in Support of Special Populations: Geriatrics" (CPMP/ICH/379/95).

European Medicines Agency (EMEA) (1994b). Pharmacokinetics and Metabolic Studies in the Safety Evaluation of New Medicinal Products in Animals.

European Medicines Agency (EMEA) (1997). Note for Guidance on the Investigations of Drug Interactions (CPMP/EWP/560/95).

European Medicines Agency (EMEA) (1998). Note for Guidance on the Pre-Clinical Evaluation of Anticancer Medicinal Products (CPMP/SWP/997/96).

European Medicines Agency (EMEA) (1999). Note for Guidance on Modified Release Oral and Transdermal Dosage Forms: Section II (Pharmacokinetic and Clinical Evaluation) (CPMP/EWP/280/96).

European Medicines Agency (EMEA) (2001). Note for Guidance on the Investigation of Bioavailability and Bioequivalence (CPMP/EWP/QWP/1401/98).

European Medicines Agency (EMEA) (2003). Note for Guidance on Evaluation of Anticancer Medicinal Products in Man. Addendum on Paediatric Oncology (EMEA/CPMP/EWP/569/02).

European Medicines Agency (EMEA) (2004). Position Paper on Non-Clinical Safety Studies to Support Clinical Trials with a Single Microdose (CPMP/SWP/2599/02/Rev 1).

European Medicines Agency (EMEA) (2005a). CHMP SWP Reflection Paper on the Assessment of the Genotoxic Potential of Antisense Oligodeoxynucleotides (EMEA/CHMP/SWP/199726/2004).

European Medicines Agency (EMEA) (2005b). Guideline on Risk Management Systems for Medicinal Products for Human Use (EMEA/CHMP/96268/2005).

European Medicines Agency (EMEA) (2005c). Guideline on the Evaluation of Control Samples in Nonclinical Safety Studies: Checking for Contamination With the Test Substance (CPMP/SWP/1094/04).

European Medicines Agency (EMEA) (2005d). Guideline on the Evaluation of the Pharmacokinetics of Medicinal Products in Patients With Impaired Hepatic Function (CPMP/EWP/2339/02).

European Medicines Agency (EMEA) (2005e). Guideline on the Need for Non-Clinical Testing in Juvenile Animals on Human Pharmaceuticals for Paediatric Indications (Draft) (EMEA/CHMP/SWP/169215/2005).

European Medicines Agency (EMEA) (2005f). Guideline on the Non-Clinical Development of Fixed Combinations of Medicinal Products (Draft) (CHMP/EMEA/CHMP/SWP/258498/2005).

European Medicines Agency (EMEA) (2005g). Guideline on the Non-Clinical Documentation for Mixed Marketing Authorisation Applications (CPMP/SWP/799/95).

European Medicines Agency (EMEA) (2006a). Adequacy of Guidance on the Elderly Regarding Medicinal Products for Human Use (EMEA/498920/2006).

European Medicines Agency (EMEA) (2006b). Concept Paper on the Development of a CHMP Guideline on the Non-Clinical Requirements to Support Early Phase I Clinical Trials With Pharmaceutical Compounds (EMEA/CHMP/SWP/91850/2006).

European Medicines Agency (EMEA) (2006c). Guideline on Risk Assessment of Medicinal Products on Human Reproduction and Lactation: from Data to Labelling (Draft) (EMEA/CHMP/203927/2005).

European Medicines Agency (EMEA) (2006d). Guideline on the Development of a CHMP Guideline on the Non-Clinical Requirements to Support Early Phase I Clinical Trials With Pharmaceutical Compounds (EMEA/CHMP/SWP/91850/2006).

European Medicines Agency (EMEA) (2006e). Guideline on the Environmental Risk Assessment of Medicinal Products for Human Use (EMEA/CHMP/SWP/4447/00).

European Medicines Agency (EMEA) (2006f). Guideline on the Evaluation of Anticancer Medicinal Products in Man (CPMP/EWP/205/95/Rev.3/Corr.2).

European Medicines Agency (EMEA) (2006g). Guideline on the Non-Clinical Investigation of the Dependence Potential of Medicinal Products (EMEA/CHMP/SWP/94227/2004).

European Medicines Agency (EMEA) (2006h). Guideline on the Role of Pharmacokinetics in the Development of Medicinal Products in the Paediatric Population (EMEA/CHMP/EWP/147013/2004/Corr).

European Medicines Agency (EMEA) (2006i). Nexavar. European Public Assessment Report: Scientific Discussion.

European Medicines Agency (EMEA) (2006j). Questions & Answers on the Bioavailability and Bioequivalence Guideline (EMEA/CHMP/EWP/40326/2006).

European Medicines Agency (EMEA) (2006k). Reflection Paper: Formulations of Choice for the Paediatric Population (EMEA/CHMP/PEG/194810/2005).

European Medicines Agency (EMEA) (2007a). Annex a to Guideline on Clinical Development of Medicinal Products for Treatment of HIV Infection. Presentation of Pharmacokinetic Interaction Data in the SPC (Draft).

European Medicines Agency (EMEA) (2007b). Concept Paper on BCS-Based Biowaiver (EMEA/CHMP/EWP/213035/2007).

European Medicines Agency (EMEA) (2007c). EMEA Guidance for Companies Requesting Scientific Advice or Protocol Assistance (EMEA-H-4260-01-Rev. 4).

European Medicines Agency (EMEA) (2007d). Guideline on Strategies to Identify and Mitigate Risks for First-In Human Clinical Trials With Investigational Medicinal Products (EMEA/CHMP/SWP/28367/07).

European Medicines Agency (EMEA) (2007e). Guideline on the Clinical Investigation of the Pharmacokinetics of Therapeutic Proteins (CHMP/EWP/89249/2004).

European Medicines Agency (EMEA) (2007f). Guideline on the Clinical Development of Medicinal Products for the Treatment of HIV Infection (Draft) (EMEA/CPMP/EWP/633/02 Rev. 2).

European Medicines Agency (EMEA) (2007g). Note for Guidance on Definitions for Genomic Biomarkers, Pharmacogenomics, Pharmacogenetics, Genomic Data and Sample Coding Categories (EMEA/CHMP/ICH/437986/2006).

European Medicines Agency (EMEA) (2007h). Recommendation on the Need for Revision of "Note for Guidance on the Investigation of Bioavailability and Bioequivalence (CPMP/EWP/QWP/1401/98).

European Medicines Agency (EMEA) (2007i). Reflection Paper on the Use of Pharmacogenetics in the Pharmacokinetic Evaluation of Medicinal Products (EMEA/128517/2006).

European Medicines Agency (EMEA) (2007j). Reflection Paper on the Use of Genomics in Cardiovascular Clinical Intervention Trials (EMEA/CHMP/PGxWP/278789/2006).

European Medicines Agency (EMEA) (2007k). The Investigation of Medicinal Products in the Term and Preterm Neonate (Draft) (EMEA/267484/2007).

European Medicines Agency (EMEA) (2008a). *EPARs for Authorised Medicinal Products for Human Use*, http://www.emea.europa.eu/htms/human/epar/a.htm (accessed 3 March 2008).

European Medicines Agency (EMEA) (2008b). Guideline on Fixed Combination Medicinal Products (Draft) (CPMP/EWP/240/95 Rev 1).

European Medicines Agency (EMEA) (2008c). Guideline on Repeated Dose Toxicity (Draft) (EMEA/CHMP/SWP/488313/2007).

European Medicines Agency (EMEA) (2008d). Guideline on the Need for Non-Clinical Testing in Juvenile Animals of Pharmaceuticals for Paediatric Indications (EMEA/CHMP/SWP/169215/2005).

European Medicines Agency (EMEA) (2008e). Guideline on the Non-Clinical Development of Fixed Combinations of Medicinal Products (EMEA/CHMP/SWP/258498/2005).

European Medicines Agency (EMEA) (2008f). *Marketing Authorisation Withdrawals and Suspensions*, http://www.emea.europa.eu/htms/human/withdraw/withdraw.htm (accessed 3 March 2008).

European Medicines Agency (EMEA) (2008g). Non-Clinical Guideline on Drug-Induced Hepatotoxicity (Draft) (EMEA/CHMP/SWP/150115/2006).

European Medicines Agency (EMEA) (2008h). *Scientific Guidelines for Human Medicinal Products*, http://www.emea.europa.eu/htms/human/humanguidelines/background.htm (accessed 3 March 2008).

European Medicines Agency (EMEA) (2008i). *Work Plan for the Safety Working Party (SWP) 2008–2009*, 20 September 2007, http://www.emea.europa.eu/pdfs/human/swp/swpworkprogramme.pdf (accessed 4 March 2008).

European Parliament (2004). *Directive 2004/10/EC of the European Parliament and of the Council of 11 February 2004 on the Harmonisation of Laws, Regulations and Administrative Provisions Relating to the Application of the Principles of Good Laboratory Practice and the Verification of Their Applications for Tests on Chemical Substances (Codified Version) (Text with EEA Relevance)*, 20 February 2004, http://europa.eu.int/eur-lex/lex/LexUriServ/LexUriServ.do?uri=CELEX:32004L0010:EN:HTML (accessed 4 March 2008).

European Union (1994). Investigation of Chiral Active Substances. Legislative Basis Directive 75/318/EEC as Amended.

Expert Group on Phase One Clinical Trials (Chairman: Professor Gordon W. Duff) (2006). Expert Group on Phase One Clinical Trials: Final Report.

Farrell, A.T. *et al.* (2003). How oncology drug development differs from other fields, in *Handbook of Anticancer Drug Development* (eds D.R. Budman, A.H. Calvert and E.K. Rowinsky), Lippincott Williams & Wilkins, pp. 3–12.

Frueh, F.W. *et al.* (2006). Experience with voluntary and required genomic data submissions to the FDA: summary report from track 1 of the third FDA-DIA-PWG-PhRMA-BIO pharmacogenomics workshop. *The Pharmacogenomics Journal*, *6* (5), 296–300.

Fura, A. (2006). Role of pharmacologically active metabolites in drug discovery and development. *Drug Discovery Today*, *11* (3–4), 133–142.

G.D. Searle LLC, a Division of Pfizer Inc. (2008). *Inspra. Eplerenone Tablets*, http://www.fda.gov/cder/foi/label/2008/021437s005lbl.pdf (accessed 8 March 2008).

Galteau, M.M., Shamsa, F. (2003). Urinary 6ß-hydroxycortisol: a validated test for evaluating drug induction or drug inhibition mediated through CYP3A in humans and in animals. *European Journal of Clinical Pharmacology*, *59* (10), 713–733.

Goodsaid, F., Frueh, F. (2007). Implementing the U.S. FDA guidance on pharmacogenomic data submissions. *Environmental and Molecular Mutagenesis*, *48* (5), 354–358.

Haddad, A., Davis, M., Langman, R. (2007). The pharmacological importance of cytochrome CYP3A4 in the palliation of symptoms: review and recommendations for avoiding adverse drug interactions. *Support Care Cancer*, *15*, 251–257.

Haidar, S.H. *et al.* (2008). Bioequivalence approaches for highly variable drugs and drug products. *Pharmaceutical Research*, *25* (1), 237–241.

Hamdy, R. *et al.* (1995). Influenza vaccine may enhance theophylline toxicity. A case report and review of the literature. *Journal of the Tennessee Medical Association*, *88* (12), 463–464.

Han, H.-K. *et al.* (2006). An efficient approach for the rapid assessment of oral rat exposures for new chemical entities in drug discovery. *Journal of Pharmaceutical Sciences*, *95* (8), 1864–1692.

Hastings, K.L. *et al.* (2003). Letter to the editor. *Toxicology and Applied Pharmacology*, *190* (1), 91–92.

Health Canada (2000). Stereochemical Issues in Chiral Drug Development.

Health Canada (2001). Draft Guidance for Industry: Quality (Chemistry and Manufacturing) Guidance: New Drug Submissions (NDSs) and Abbreviated New Drug Submissions (ANDSs).

Health Canada (2003). Discussion Paper. Bioequivalence Requirements: Highly Variable Drugs and Highly Variable Drug Products: Issues and Options.

Health Canada (2007a). Clinical Assessment of Abuse Liability for Drugs with Central Nervous System Activity.

Health Canada (2007b). Withdrawal of Market Authorization for Prexige.

Hewitt, N.J., de Kanter, R., LeCluyse, E. (2007). Induction of drug metabolizing enzymes: a survey of *in vitro* methodologies and interpretations used in the pharmaceutical industry—do they comply with FDA recommendations? *Chemico-Biology Interactions*, *168* (1), 51–65.

Hill, T.P. (2007). Phase 0 trials: are they ethically challenged? *Clinical Cancer Research*, *13* (3), 783–784.

Hodge, L.S., Tracy, T.S. (2007). Alterations in drug disposition during pregnancy: implications for drug therapy. *Expert Opinion on Drug Metabolism and Toxicology*, *3* (4), 557.

Holmes, E., Nicholson, J. (2005). Variation in gut microbiota strongly influences individual rodent phenotypes. *Toxicological Sciences*, *87* (1), 1–2.

Hop, C. *et al.* (1998). Plasma-pooling methods to increase throughput for *in vivo* pharmacokinetic screening. *Journal of Pharmaceutical Sciences*, *87* (7), 901–903.

Hu, Z. *et al.* (2005). Herb-drug interactions: a literature review. *Drugs*, *65* (9), 1239–1282.

Huang, S.-M. *et al.* (2008). New era in drug interaction evaluation: US food and drug administration update on CYP enzymes, transporters, and the guidance process. *Journal of Clinical Pharmacol*, *48* (6), 662–670.

Humphreys, W.G., Unger, S.E. (2006). Safety assessment of drug metabolites: characterization of chemically stable metabolites. *Chemical Research in Toxicology*, *19* (12), 1564–1569.

Hurwitz, A. *et al.* (2003). Gastric function in the elderly: effects on absorption of ketoconazole. *Journal of Clinical Pharmacol*, *43* (9), 996–1002.

International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH) (1994a). *ICH Harmonised Tripartite Guideline. Pharmacokinetics: Guidance for Repeated Dose Tissue Distribution Studies. S3B*, http://www.ich.org/LOB/media/MEDIA496.pdf (accessed 6 March 2008).

International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH) (1994b). *ICH Harmonised Tripartite Guideline. Note for Guidance on Toxicokinetics: The Assessment of Systemic Exposure in Toxicity Studies. S3A*, 27 October 1994, http://www.ich.org/LOB/media/MEDIA495.pdf (accessed 4 March 2008).

International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH) (1994c). *ICH Harmonised Tripartite Guideline. Studies in*

*Support of Special Populations: Geriatrics. E7*, http://www.ich.org/LOB/media/MEDIA483.pdf (accessed 8 March 2008).

International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH) (1995). *ICH Harmonised Tripartite Guideline. Guidance on Specific Aspects of Regulatory Genotoxicity Tests for Pharmaceuticals. S2A*, http://www.ich.org/LOB/media/MEDIA493.pdf (accessed 6 March 2008).

International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH) (1996a). *ICH Harmonised Tripartite Guideline. Good Clinical Practice: Consolidated Guideline. E6(R1)*, 10 June 1996, http://www.ich.org/LOB/media/MEDIA482.pdf (accessed 6 March 2008).

International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH) (1996b). *ICH Harmonised Tripartite Guideline. Preclinical Safety Evaluation of Biotechnology-Derived Pharmaceuticals. S6*, http://www.ich.org/LOB/media/MEDIA503.pdf (accessed 8 March 2008).

International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH) (1997a). *ICH Harmonised Tripartite Guideline. General Considerations for Clinical Trials. E8*, 14 July 1997, http://www.ich.org/LOB/media/MEDIA484.pdf (accessed 6 March 2008).

International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH) (1997b). *ICH Harmonised Tripartite Guideline. Impurities: Guideline for Residual Solvents. Q3C(R3)*, http://www.ich.org/LOB/media/MEDIA423.pdf (accessed 7 March 2008).

International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH) (1998). *ICH Harmonised Tripartite Guideline. Ethnic Factors in the Acceptability of Foreign Clinical Data. E5(R1)*, http://www.ich.org/LOB/media/MEDIA481.pdf (accessed 7 March 2008).

International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH) (2000a). *ICH Harmonised Tripartite Guideline. Maintenance of the ICH Guideline on Non-Clinical Safety Studies for the Conduct of Human Clinical Trials for Pharmaceuticals. M3(R1)*, 9 November 2000, http://www.ich.org/LOB/media/MEDIA506.pdf (accessed 4 March 2008).

International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH) (2000b). *ICH Harmonised Tripartite Guideline. Safety Pharmacology Studies for Human Pharmaceuticals. S7A*, 8 November 2000, http://www.ich.org/LOB/media/MEDIA504.pdf (accessed 5 March 2008).

International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH) (2000c). *ICH Harmonised Tripartite Guideline. Clinical Investigation of Medicinal Products in the Pediatric Population. E11*.

International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH) (2000d). *ICH Harmonised Tripartite Guideline. The Common Technical Document for the Registration of Pharmaceuticals for Human Use: Safety—M4S(R2). Nonclinical Overview and Nonclinical Summaries of Module 2 Organisation of Module 4*, http://www.ich.org/LOB/media/MEDIA559.pdf (accessed 10 March 2008).

International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH) (2005a). *ICH Harmonised Tripartite Guideline. The Nonclinical Evaluation of the Potential for Delayed Ventricular Repolarization (QT Interval Prolongation) By Human Pharmaceuticals. S7B*, 12 May 2005, http://www.ich.org/LOB/media/MEDIA2192.pdf (accessed 5 March 2008).

International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH) (2005b). *ICH Harmonised Tripartite Guideline. Detection*

*of Toxicity to Reproduction for Medicinal Products & Toxicity to Male Fertility. S5(R2)*, November 2005, http://www.ich.org/LOB/media/MEDIA498.pdf (accessed 5 March 2008).

International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH) (2005c). *ICH Harmonised Tripartite Guideline. Dose Selection for Carcinogenicity Studies of Pharmaceuticals & Limit Dose. S1C(R1)*, http://www.ich.org/LOB/media/MEDIA491.pdf (accessed 6 March 2008).

International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH) (2005d). *ICH Harmonised Tripartite Guideline. Immunotoxicology Studies for Human Pharmaceuticals. S8*, http://www.ich.org/LOB/media/MEDIA1706.pdf (accessed 6 March 2008).

International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH) (2005e). *ICH Harmonised Tripartite Guideline. The Clinical Evaluation of QT/QTc Interval Prolongation and Proarrhythmic Potential for Non-Antiarrhythmic Drugs. E14*, http://www.ich.org/LOB/media/MEDIA1476.pdf (accessed 7 March 2008).

International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH) (2006a). *ICH Harmonised Tripartite Guideline. Impurities in New Drug Substances. Q3A(R2)*, 25 October 2006, http://www.ich.org/LOB/media/MEDIA422.pdf (accessed 4 March 2008).

International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH) (2006b). *ICH Harmonised Tripartite Guideline. Impurities in New Drug Products. Q3B(R2)*, 2 June 2006, http://www.ich.org/LOB/media/MEDIA421.pdf (accessed 4 March 2008).

International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH) (2006c). *Final Concept Paper: M3(R2): Revision of ICH M3(R1): Maintenance of the ICH Guideline on Non-Clinical Safety Studies for the Conduct of Human Clinical Trials for Pharmaceuticals*, http://www.ich.org/LOB/media/MEDIA3303.pdf (accessed 7 March 2008).

International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH) (2006d). *Final Concept Paper S2(R1): Guidance on Genotoxicity Testing and Data Interpretation for Pharmaceuticals Intended for Human Use*, http://www.ich.org/LOB/media/MEDIA3304.pdf (accessed 18 March 2008).

International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH) (2006e). *E5: Ethnic Factors in the Acceptability of Foreign Clinical Data. Questions & Answers (R1)*, http://www.ich.org/LOB/media/MEDIA1194.pdf (accessed 7 March 2008).

International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH) (2007). *Final Concept Paper. S9: Preclinical Guideline on Oncology Therapeutic Development*, http://www.ich.org/LOB/media/MEDIA3922.pdf (accessed 6 March 2008).

International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH) (2008). www.ich.org (accessed 4 March 2008).

Jacobson-Kram, D. (2007). Overview of exploratory IND (phase 0 trial): differences from the traditional IND, 2007 AAPS National Biotechnology Conference, San Diego, CA.

Japan Pharmaceutical Manufacturers Association (2007). *Pharmaceutical Administration and Regulations in Japan*, March 2007, http://www.nihs.go.jp/mhlw/jouhou/yakuji/yakuji-e0703.pdf (accessed 4 March 2008).

Johnston, G., Williams, R.L. (2002). 505(b)(2) applications: history, science, and experience. *Drug Information Journal*, *36*, 319–323.

de Jonge, M.E. *et al.* (2005). Individualized cancer chemotherapy: strategies and performance of prospective studies on therapeutic drug monitoring with dose adaptation. *Clinical Pharmacokinetics*, *44* (2), 147–173.

Joukhadar, C., Müller, M. (2005). Microdialysis: current applications in clinical pharmacokinetic studies and its potential role in the future. *Clinical Pharmacokinetics*, *44* (9), 895–913.

Kamimura, H. (2006). Genetic polymorphism of cytochrome P450s in beagles: possible influence of CYP1A2 deficiency on toxicological evaluations. *Archives Toxicology*, *80* (11), 732–736.

Kennedy, T. (1997). Managing the drug discovery/development interface. *Drug Discovery Today*, *2* (10), 436–444.

Kola, I., Landis, J. (2004). Can the pharmaceutical industry reduce attrition rates? *Nature Reviews. Drug Discovery*, *3* (8), 711–716.

Kraemer, M.J. *et al.* (1982). Altered theophylline clearance during an influenza B outbreak. *Pediatrics*, *69* (4), 476–480.

Kummar, S. *et al.* (2007). Inhibition of poly (ADP-ribose) polymerase (PARP) by ABT-888 in patients with advanced malignancies: results of a phase 0 trial. *Journal of Clinical Oncology*, *25* (18S), 3518.

Lacko, A.G. *et al.* (2007). Prospects and challenges of the development of lipoprotein-based formulations for anti-cancer drugs. *Expert Opinion on Drug Delivery*, *4* (6), 665–675.

Lange, D. *et al.* (1997). Effect of a cola beverage on the bioavailability of itraconazole in the presence of H2 blockers. *Journal of Clinical Pharmacol*, *37* (6), 535–540.

Lantz, R.J. *et al.* (2003). Metabolism, excretion, and pharmacokinetics of duloxetine in healthy human subjects. *Drug Metabolism and Disposition*, *31* (9), 1142–1150.

Lappin, G. *et al.* (2006). Use of microdosing to predict pharmacokinetics at the therapeutic dose: experience with 5 drugs. *Clinical Pharmacol Therapy*, *80* (3), 203–215.

Lappin, G., Temple, S. (2006). *Radiotracers in Drug Development*, CRC Press, Taylor & Francis Group.

Larson, J.L. (2005). Toxicity evaluations: ICH guidances and current practice, in *Preclinical Drug Development* (eds M.C. Rogge and D.R. Taft), Taylor and Francis, pp. 349–414.

Leblanc, B. *et al.* (1998). Binding of drugs to eye melanin is not predictive of ocular toxicity. *Regulatory Toxicology and Pharmacology*, *28* (2), 124–132.

Lesko, L.J. *et al.* (2003). Pharmacogenetics and pharmacogenomics in drug development and regulatory decision making: report of the first FDA-PWG-PhRMA-DruSafe workshop. *Journal of Clinical Pharmacol*, *43* (4), 342–358.

Levy, M. (1997). Role of viral infections in the induction of adverse drug reactions. *Drug Safety*, *16* (1), 1–8.

Li, J. *et al.* (2006). CYP3A phenotyping approach to predict systemic exposure to EGFR tyrosine kinase inhibitors. *Journal of the National Cancer Institute*, *98* (23), 1714–1723.

Li, M. *et al.* (2008). Symbiotic gut microbes modulate human metabolic phenotypes. *Proceedings of the National Academy of Sciences of the United States of America*, *105* (6), 2117–2122.

Lionberger, R. (2008). FDA critical path initiatives: opportunities for generic drug development. *The AAPS Journal*, *10* (1), 103–109.

Lytton, S.D. *et al.* (2002). Autoantibodies against cytochrome P450s in sera of children treated with immunosuppressive drugs. *Clinical and Experimental Immunology*, *127* (2), 293–302.

Mahmood, I. (2005). The correction factors do help in improving the prediction of human clearance from animal data. *Journal of Pharmaceutical Sciences*, *94* (5), 940–945.

Mahmood, I. (2006). *Clinical Pharmacology of Therapeutic Proteins*, Pine House Publishers.

Mahmood, I., Green, M.D. (2007). Drug interaction studies of therapeutic proteins or monoclonal antibodies. *Journal of Clinical Pharmacol*, *47* (12), 1540–1554.

Mahmood, I., Green, M.D., Fisher, J.E. (2003). Selection of the first-time dose in humans: comparison of different approaches based on interspecies scaling of clearance. *Journal of Clinical Pharmacol*, *43* (7), 692–697.

Malhotra, S., Dixit, R.K., Garg, S.K. (2002). Effect of an acidic beverage (Coca-Cola) on the pharmacokinetics of carbamazepine in healthy volunteers. *Methods and Findings in Experimental Clinical Pharmacology*, *24* (1), 31–33.

Mallet, L. (2008). *Drug Problems in the Elderly*, http://www.med.mcgill.ca/geriatrics/education/undergrad/ttag_pom_site/pharmacology.htm (accessed 8 March 2008).

Mallet, L., Spinewine, A., Huang, A. (2007). The challenge of managing drug interactions in elderly people. *Lancet*, *370* (9582), 185–191.

Mann, B.S. *et al.* (2007). FDA approval summary: Vorinostat for treatment of advanced primary cutaneous T-cell lymphoma. *Oncologist*, *12* (10), 1247–1252.

Mansfield, P., Henry, D., Tonkin, A. (2004). Single-enantiomer drugs: elegant science, disappointing effects. *Clinical Pharmacokinetics*, *43* (5), 287–290.

Marchetti, S., Schellens, J. (2007). The impact of FDA and EMEA guidelines on drug development in relation to Phase 0 trials. *British Journal of Cancer*, *97* (5), 577–581.

Marchetti, S. *et al.* (2007). Concise review: clinical relevance of drug–drug and herb–drug interactions mediated by the ABC transporter ABCB1 (MDR1, P-glycoprotein). *Oncologist*, *12* (8), 927–941.

Mascelli, M.A., Zhou, H., *et al.* (2007). Molecular, biologic, and pharmacokinetic properties of monoclonal antibodies: impact of these parameters on early clinical development. *J. Clin. Pharmacol.*, *47* (5), 553–65.

Meijerman, I., Beijnen, J.H., Schellens, J.H.M. (2006). Herb-drug interactions in oncology: focus on mechanisms of induction. *Oncologist*, *11* (7), 742–752.

Merck/Schering-Plough Pharmaceuticals (2008). *Zetia. Labelling information*, 7 February 2008, http://www.fda.gov/cder/foi/label/2008/021445s019lbl.pdf (accessed 4 March 2008).

Meredith, C. *et al.* (1985). Effects of influenza virus vaccine on hepatic drug metabolism. *Clinical Pharmacology and Therapeutics*, *37* (4), 396–401.

Micuda, S. *et al.* (2007). Diurnal variation of 6beta-hydroxycortisol in cardiac patients. *Physiological Research*, *57* (3), 307–313.

Ministry of Health and Welfare, Pharmaceutical Affairs Bureau (1995). *Japanese Guidelines for Nonclinical Studies of Drugs Manual 1995*, Yakuji Nippo, Ltd, Tokyo.

Miyakawa, H. *et al.* (2000). Immunoreactivity to various human cytochrome P450 proteins of sera from patients with autoimmune hepatitis, chronic hepatitis B, and chronic hepatitis C. *Autoimmunity*, *33* (1), 23–32.

Monahan, B.P. *et al.* (1990). Torsades de pointes occurring in association with terfenadine use. *The Journal of the American Medical Association*, *264* (21), 2788–2790.

Morita, K. *et al.* (1998). Strain differences in CYP3A-mediated C-8 hydroxylation (1,3,7-trimethyluric acid formation) of caffeine in wistar and dark agouti rats: rapid metabolism of caffeine in debrisoquine poor metabolizer model rats. *Biochemical Pharmacology*, *55* (9), 1405–1411.

Murray, M. (2006). Altered CYP expression and function in response to dietary factors: potential roles in disease pathogenesis. *Current Drug Metabolism*, *7*, 67–81.

Nagilla, R., Ward, K.W. (2004). A comprehensive analysis of the role of correction factors in the allometric predictivity of clearance from rat, dog, and monkey to humans. *Journal of Pharmaceutical Sciences*, *93* (10), 2522–2534.

Nagilla, R., Ward, K.W. (2005a). Erratum: a comprehensive analysis of the role of correction factors in the allometric predictivity of clearance from rat, dog, and monkey to humans. *Journal of Pharmaceutical Sciences*, *94* (1), 231–232.

Nagilla, R., Ward, K.W. (2005b). Correspondence. *Journal of Pharmaceutical Sciences*, *94* (5), 946–947.

Naito, S. *et al.* (2007). Current opinion: safety evaluation of drug metabolites in development of pharmaceuticals. *The Journal of Toxicological Sciences*, *32* (4), 329–341.

Nakai, K. *et al.* (2008). Decreased expression of cytochrome P450s 1A2, 2E1, and 3A4; and drug transporters Na+-taurocholate cotransporting polypeptide, organic cation transporter 1, and

organic anion-transporting peptide-C correlates with the progression of liver fibrosis in chronic hepatitis C patients. *Drug Metabolism and Disposition*, *36* (9), 1786–1793.

Nanosphere Inc. (2008). *Verigene® Warfarin Metabolism Nucleic Acid Test (IVD)*, http://www.nanosphere-inc.com/VerigeneWarfarinMetabolismNucleicAcidTest_4472.aspx (accessed 7 March 2008).

National Cancer Institute (2007). *Phase 0 Trials in Oncologic Drug Development*, National Cancer Institute, Bethesda, MD.

Neuvonen, P., Kivisto, K. (1994). Enhancement of drug absorption by antacids: an unrecognised drug interaction. *Clinical Pharmacokinetics*, *27* (2), 120–128.

Nicholson, J.K., Holmes, E., Wilson, I.D. (2005). Gut microorganisms, mammalian metabolism and personalized health care. *Nature Reviews. Microbiology*, *3* (5), 431–438.

Novartis Australia Pty Limited (2007). *Novartis Withdraws Prexige® (lumiracoxib) in Australia in Response to Decision from Therapeutic Goods Administration (TGA)*, http://www.novartis.com.au/Prexige%20press%20release%2011%20August.pdf (accessed 3 March 2008).

Ohno, M. *et al.* (2000). Circadian variation of the urinary 6beta-hydroxycortisol to cortisol ratio that would reflect hepatic CYP3A activity. *European Journal of Clinical Pharmacology*, *55* (11–12), 861–865.

Organization for Economic Co-Operation and Development (OECD) (2007). *OECD Series on Principles of Good Laboratory Practice and Compliance Monitoring*, http://www.oecd.org/document/63/0,2340,en_2649_34381_2346175_1_1_1_37465,00.html (accessed 4 March 2008).

Organization for Economic Co-Operation and Development (OECD) (2008). *Good Laboratory Practice*, http://www.oecd.org/topic/0,2686,en_2649_34381_1_1_1_1_37465,00.html (accessed 4 March 2008).

Paulson, S.K. *et al.* (1999). Evidence for polymorphism in the canine metabolism of the cyclooxygenase 2 inhibitor, celecoxib. *Drug Metabolism and Disposition*, *27* (10), 1133–1142.

Pichini, S., Zuccaro, P., Pacifici, R. (1994). Drugs in semen. *Clinical Pharmacokinetics*, *26* (5), 356–373.

Projean, D. *et al.* (2005). Selective downregulation of hepatic cytochrome P450 expression and activity in a rat model of inflammatory pain. *Pharmaceutical Research*, *22* (1), 62–70.

Propper, D.J. *et al.* (2001). Phase I and pharmacokinetic study of PKC412, an inhibitor of protein kinase C. *Journal of Clinical Oncology*, *19* (5), 1485–1492.

Prueksaritanont, T., Lin, J.H., Baillie, T.A. (2006). Complicating factors in safety testing of drug metabolites: kinetic differences between generated and preformed metabolites. *Toxicology and Applied Pharmacology*, *217* (2), 143–152.

Qin, X. (2008). Gut microbiota: a new aspect that should be taken into more consideration when assessing the toxicity of chemicals or the adverse effects and efficacy of drugs. *Regulatory Toxicology and Pharmacology*, *51* (2), 251–251.

Renton, K.W. (2001). Alteration of drug biotransformation and elimination during infection and inflammation. *Pharmacology and Therapeutics*, *92*, 147–163.

Riechelmann, R., Saad, E. (2006). A systematic review on drug interactions in oncology. *Cancer Investigation*, *24*, 704–712.

Riechelmann, R. *et al.* (2005). Potential for drug interactions in hospitalized cancer patients. *Cancer Chemotherapy and Pharmacology*, *56*, 286–290.

Riechelmann, R.P. *et al.* (2007). Potential drug interactions and duplicate prescriptions among cancer patients. *Journal of the National Cancer Institute*, *99* (8), 592–600.

Robosky, L.C. *et al.* (2005). Metabonomic identification of two distinct phenotypes in Sprague-Dawley (Crl:CD(SD) rats. *Toxicological Sciences*, *87* (1), 277–284.

Robosky, L.C. *et al.* (2006). Communication regarding metabonomic identification of two distinct phenotypes in Sprague-Dawley (Crl:CD(SD) rats. *Toxicological Sciences*, *91* (1), 309.

Rocci, M.L. Jr, Shah, V.P., Rose, M.J., Sailstad, J.M. (2006). Bioanalytical method validation and implementation: best practices for chromatographic and ligand binding assays. *The AAPS Journal*, http://www.aapsj.org/theme_issues/theme_issue21.asp (accessed 4 March 2008).

Roche Diagnostics. (2008). *Amplichip*, http://www.amplichip.us/ (accessed 27 June 2008).

Rodriguez-Antona, C., Ingelman-Sundberg, M. (2006). Cytochrome P450 pharmacogenetics and cancer. *Oncogene*, *25* (11), 1679–1691.

Roffey, S.J. *et al.* (2007). What is the objective of the mass balance study? A retrospective analysis of data in animal and human excretion studies employing radiolabeled drugs. *Drug Metabolism Reviews*, *39* (1), 17–43.

Rogge, M.C., Taft, D.R. (2005). *Preclinical Drug Development. Drugs and the Pharmaceutical Sciences*, Vol. *152*, Taylor and Francis.

Rohner, T.C., Staab, D., Stoeckli, M. (2005). MALDI mass spectrometric imaging of biological tissue sections. *Mechanisms of Ageing and Development*, *126* (1), 177–185.

Rowland, M. (2007). Commentary on ACCP position statement on the use of microdosing in the drug development process. *Journal of Clinical Pharmacol*, *47* (12), 1595–1596.

Ryu, J.Y. *et al.* (2007). Development of the inje cocktail for high-throughput evaluation of five human cytochrome P450 isoforms *in vivo*. *Clinical Pharmacol Therapy*, *82* (5), 531–540.

Sadowski, D. (1994). Drug interactions with antacids. Mechanisms and clinical significance. *Drug Safety*, *11* (6), 395–407.

Salazar-Bookaman, M., Wainer, I., Patil, P. (1994). Relevance of drug-melanin interactions to ocular pharmacology and toxicology. *Journal of Ocular Pharmacology*, *10* (1), 217–239.

Sanders, J.M. *et al.* (2001). Comparative xenobiotic metabolism between Tg.AC and p53+/– genetically altered mice and their respective wild types. *Toxicological Sciences*, *61* (1), 54–61.

Sanderson, J., Naisbitt, D., Park, B. (2006). The role of bioactivation in drug-induced hypersensitivity reactions. *The AAPS Journal*, *8* (1), E55–E64.

Sarapa, N. (2007). Exploratory IND: a new regulatory strategy for early clinical drug development in the United States. *Ernst Schering Research Foundation Workshop*, *59*, 151–163.

Scharberg, D. (2007). *Critical Recommendations from the Third AAPS/FDA Bioanalytical Workshop*, http://www.pharmoutsource.com/publications/bp/articles/scharberg/critical_recommendations.pdf (accessed 4 March 2008).

Scharberg, D. (2008). *Four Lessons from MDS Montreal*, http://www.pharmoutsource.com/publications/bp/articles/scharberg/MDS_Best_Practice.pdf (accessed 4 March 2008).

Schering-Plough Research Institute (2006). *Product Information: PegIntron*, http://www.fda.gov/cder/foi/label/2006/103949s5124LBL.pdf (accessed 8 March 2008).

Schuster, D., Laggner, C., Langer, T. (2005). Why drugs fail—a study on side effects in new chemical entities. *Current Pharmaceutical Design*, *11*, 3545–3559.

Seeff, L. (2007). Herbal hepatotoxicity. *Clinics in Liver Disease*, *11* (3), 577–596.

Seitz, K., Zhou, H. (2007). Pharmacokinetic drug–drug interaction potentials for therapeutic monoclonal antibodies: reality check. *Journal of Clinical Pharmacol*, *47* (9), 1104–1118.

Senior, J. (2007). Drug hepatotoxicity from a regulatory perspective. *Clinics in Liver Disease*, *11* (3), 507–524.

Shah, V.P. *et al.* (1992). Analytical methods validation: bioavailability, bioequivalence, and pharmacokinetic studies. *Journal of Pharmaceutical Sciences*, *81* (3), 309–312.

Shah, V.P. *et al.* (2000). Bioanalytical method validation—a revisit with a decade of progress. *Pharmaceutical Research*, *17* (12), 1551–1557.

Skalli, S., Zaid, A., Soulaymani, R. (2007). Drug interactions with herbal medicines. *Therapeutic Drug Monitoring*, *29* (6), 679–686.

Smith, D.A., Obach, R.S. (2005). Seeing through the mist: abundance versus percentage. Commentary on metabolites in safety testing. *Drug Metabolism and Disposition*, *33* (10), 1409–1417.

Smith, D.A., Obach, R.S. (2006). Metabolites and safety: what are the concerns, and how should we address them? *Chemical Research in Toxicology*, *19* (12), 1570–1579.

Solon, E., Kraus, L. (2001). Quantitative whole-body autoradiography in the pharmaceutical industry. Survey results on study design, methods, and regulatory compliance. *Journal of Pharmacological and Toxicological Methods*, *46* (2), 73–81.

Solon, E.G., Lee, F. (2002). Methods determining phosphor imaging limits of quantitation in whole-body autoradiography rodent tissue distribution studies affect predictions of 14C human dosimetry. *Journal of Pharmacological and Toxicological Methods*, *46*, 83–91.

Sparreboom, A. *et al.* (2004). Herbal remedies in the united states: potential adverse interactions with anticancer agents. *Journal of Clinical Oncol*, *22* (12), 2489–2503.

Stephens, T., Brynner, R. (2001). *Dark Remedy: The impact of Thalidomide and Its Revival as a Vital Medicine*, Perseus Publishing, Cambridge, MA.

Stults, B.M., Hashisaki, P.A. (1983). Influenza vaccination and theophylline pharmacokinetics in patients with chronic obstructive lung disease. *Clinical Investigation*, *139* (5), 651–654.

Stumpf, W.E. (2005). Drug localization and targeting with receptor microscopic autoradiography. *Journal of Pharmacological and Toxicological Methods*, *51* (1), 25–40.

Tang, W. (2007). Drug metabolite profiling and elucidation of drug-induced hepatotoxicity. *Expert Opinion on Drug Metabolism and Toxicology*, *3* (3), 407–420.

Tchambaz, L. *et al.* (2006). Dose adaptation of antineoplastic drugs in patients with liver disease. *Drug Safety*, *29* (6), 509–522.

The Council of the European Union (1996). Council Directive 96/29/EURATOM of 13 May 1996 Laying Down Basic Safety Standards for the Protection of the Health of Workers and the General Public Against the Dangers Arising from Ionizing Radiation.

The Council of the European Union (1997). Council Directive 97/43/EURATOM of 30 June 1997 on Health Protection of Individuals Against the Dangers of Ionizing Radiation in Relation to Medical Exposure, And Repealing Directive 84/466/EURATOM.

Third Wave Technologies (2008). *Third Wave Technologies*, http://www.twt.com/ (accessed 7 March 2008).

U.S. Environmental Protection Agency (2008). *Good Laboratory Practices Standards*, http://www.epa.gov/Compliance/monitoring/programs/fifra/glp.html (accessed 4 March 2008).

U.S. Food and Drug Administration (US FDA) (1992). FDA's Policy Statement for the Development of New Stereoisomeric Drugs.

U.S. Food and Drug Administration (US FDA) (1995). Guidance for Industry: Content and Format of Investigational New Drug Applications (INDs) for Phase 1 Studies of Drugs, Including Well-Characterized, Therapeutic, Biotechnology-derived Products.

U.S. Food and Drug Administration (US FDA) (1996). Guidance for Industry: Single Dose Acute Toxicity Testing for Pharmaceuticals.

U.S. Food and Drug Administration (US FDA) (1997a). Guidance for Industry: Drug Metabolism/Drug Interaction Studies in the Drug Development Process: Studies In Vitro.

U.S. Food and Drug Administration (US FDA) (1997b). Points to Consider in the Manufacture and Testing of Monoclonal Antibody Products for Human Use.

U.S. Food and Drug Administration (US FDA) (1998a). *FDA Talk Paper. Roche Laboratories Announces Withdrawal of Posicor from the Market*, http://www.fda.gov/bbs/topics/ANSWERS/ANS00876.html (accessed 4 March 2008)

U.S. Food and Drug Administration (US FDA) (1998b). Guidance for Industry: Pharmacokinetics in Patients with Impaired Renal Function—Study Design, Data Analysis, and Impact on Dosing and Labeling.

U.S. Food and Drug Administration (US FDA) (1998c). Guidance for Industry: General Considerations for Pediatric Pharmacokinetic Studies for Drugs and Biological Products (Draft).

U.S. Food and Drug Administration (US FDA) (1999a). Guidance for Industry: in vivo Drug Metabolism/Drug Interaction Studies—Study Design, Data Analysis, and Recommendations for Dosing and Labeling.

U.S. Food and Drug Administration (US FDA) (1999b). Guidance for Industry: Development Programs for Drugs, Devices, and Biological Products for the Treatment of Rheumatoid Arthritis (RA).

U.S. Food and Drug Administration (US FDA) (1999c). Guidance for Industry: Applications Covered By Section 505(b)(2) (Draft).

U.S. Food and Drug Administration (US FDA) (2000a). Guidance for Industry: FDA guidance for industry on Formal Meetings with Sponsors and Applicants for PDUFA Products.

U.S. Food and Drug Administration (US FDA) (2000b). Guidance for Industry: Waiver of *in vivo* Bioavailability and Bioequivalence Studies for Immediate-Release Solid Oral Dosage Forms Based on a Biopharmaceutics Classification System.

U.S. Food and Drug Administration (US FDA) (2000c). Guidance for Industry: Q & A. Content and Format of INDs for Phase 1 Studies of Drugs, Including Well-Characterized, Therapeutic, Biotechnology-Derived Products.

U.S. Food and Drug Administration (US FDA) (2001a). *Frequently Asked Questions on Drug Development and Investigational New Drug Applications*, http://www.fda.gov/cder/about/smallbiz/faq.htm (accessed 10 March 2008).

U.S. Food and Drug Administration (US FDA) (2001b). Guidance for Industry: Bioanalytical Method Validation.

U.S. Food and Drug Administration (US FDA) (2001c). Guidance for Reviewers: Pharm/Tox Review Format.

U.S. Food and Drug Administration (US FDA) (2001d). Manual of Policies and Procedures. INDs: Screening INDs (MAPP 6030.4), Center for Drug Evaluation and Research (Offices of Review Management and Pharmaceutical Sciences) Editor.

U.S. Food and Drug Administration (US FDA) (2001e). Reviewer Guidance: Integration of Study Results to Assess Concerns about Human Reproductive and Developmental Toxicities (Draft).

U.S. Food and Drug Administration (US FDA) (2002a). Guidance for Industry: Carcinogenicity Study Protocol Submissions.

U.S. Food and Drug Administration (US FDA) (2002b). Guidance for Industry: Food-Effect Bioavailability and Fed Bioequivalence Studies.

U.S. Food and Drug Administration (US FDA) (2002c). Guidance for Industry: Liposome Drug Products Chemistry, Manufacturing, and Controls; Human Pharmacokinetics and Bioavailability; and Labeling Documentation (Draft).

U.S. Food and Drug Administration (US FDA) (2002d). Inspra (Epleronone) Tablets: Clinical Pharmacology Biopharmaceutics Review, pp. 70–72.

U.S. Food and Drug Administration (US FDA) (2002e). Zetia—Summary Basis of Approval—Pharmacology Review.

U.S. Food and Drug Administration (US FDA) (2003a). Guidance for Industry: Photosafety Testing.

U.S. Food and Drug Administration (US FDA) (2003b). Guidance for Industry: Exposure-Response Relationships—Study Design, Data Analysis, and Regulatory Applications.

U.S. Food and Drug Administration (US FDA) (2003c). Guidance for Industry: Bioavailability and Bioequivalence Studies for Orally Administered Drug Products—General Considerations.

U.S. Food and Drug Administration (US FDA) (2003d). Guidance for Industry: Pharmacokinetics in Patients with Impaired Hepatic Function: Study Design, Data Analysis, and Impact on Dosing and Labeling.

U.S. Food and Drug Administration (US FDA) (2004a). Challenges and Opportunity on the Critical Path to New Medical Products.

U.S. Food and Drug Administration (US FDA) (2004b). Guidance for Industry: IND Exemptions for Studies of Lawfully Marketed Drug or Biological Products for the Treatment of Cancer.

U.S. Food and Drug Administration (US FDA) (2004c). Guidance for Industry: Pharmacokinetics in Pregnancy—Study Design, Data Analysis, and Impact on Dosing and Labeling (Draft).

U.S. Food and Drug Administration (US FDA) (2004d). Letter from Joanne Rhoads to Gilbert Godin on 26 Apr 2004.

U.S. Food and Drug Administration (US FDA) (2004e). Letter from Joanne Rhoads to Gilbert Godin on 21 Dec 2004.

U.S. Food and Drug Administration (US FDA) (2004f). Natalizumab: Clinical Pharmacology and Biopharmaceutics Review(s).

U.S. Food and Drug Administration (US FDA) (2005a). *Frequently Asked Questions on the Pre-Investigational New Drug (IND) Meeting*, http://www.fda.gov/cder/about/smallbiz/pre_IND_qa.htm (accessed 3 March 2008).

U.S. Food and Drug Administration (US FDA) (2005b). Guidance for Industry, Investigators, and Reviewers: Exploratory IND Studies (Draft).

U.S. Food and Drug Administration (US FDA) (2005c). Guidance for Industry: Safety Testing of Drug Metabolites (Draft).

U.S. Food and Drug Administration (US FDA) (2005d). Guidance for Industry: Nonclinical Studies for the Safety Evaluation of Pharmaceutical Excipients.

U.S. Food and Drug Administration (US FDA) (2005e). Guidance for Industry: Estimating the Maximum Safe Starting Dose in Initial Clinical Trials for Therapeutics in Adult Healthy Volunteers.

U.S. Food and Drug Administration (US FDA) (2005f). Guidance for Industry: Pharmacogenomic Data Submissions.

U.S. Food and Drug Administration (US FDA) (2005g). Guidance for Industry: Clinical Lactation Studies—Study Design, Data Analysis, and Recommendations for Labeling (Draft).

U.S. Food and Drug Administration (US FDA) (2005h). Nexavar. Summary Basis of Approval: Pharmacology Review.

U.S. Food and Drug Administration (US FDA) (2006a). Critical Path Opportunities Report.

U.S. Food and Drug Administration (US FDA) (2006b). Critical Path Opportunities List.

U.S. Food and Drug Administration (US FDA) (2006c). Critical Path Opportunities Initiated During 2006.

U.S. Food and Drug Administration (US FDA) (2006d). *Drug Development and Drug Interactions*, http://www.fda.gov/cder/drug/drugInteractions/ (accessed 8 March 2008).

U.S. Food and Drug Administration (US FDA) (2006e). *Food and Drug Administration. Center for Drug Evaluation and Research. Advisory Committee for Pharmaceutical Science (ACPS) October 6, 2006. Highly Variable Drugs—Bioequivalence Issues*, http://www.fda.gov/ohrms/dockets/ac/cder06.html#PharmScience (accessed 7 March 2008).

U.S. Food and Drug Administration (US FDA) (2006f). Guidance for Industry, Investigators and Reviewers: Exploratory IND Studies.

U.S. Food and Drug Administration (US FDA) (2006g). Guidance for Industry: Drug Interaction Studies—Study Design, Data Analysis, and Implications for Dosing and Labeling (Draft).

U.S. Food and Drug Administration (US FDA) (2006h). Guidance for Industry: Nonclinical Safety Evaluation of Drug or Biologic Combinations.

U.S. Food and Drug Administration (US FDA) (2006i). Guidance for Industry: Nonclinical Safety Evaluation of Pediatric Drug Products.

U.S. Food and Drug Administration (US FDA) (2006j). Guidance for Industry: Warnings and Precautions, Contraindications, and Boxed Warning Sections of Labeling for Human Prescription Drug and Biological Products—Content and Format (Draft).

U.S. Food and Drug Administration (US FDA) (2006k). Letter from Joseph Salewski to Stephen DeFalco on 31 Aug 2006.

U.S. Food and Drug Administration (US FDA) (2006l). Sutent: Summary Basis of Approval. Clinical Pharmacology and Biopharmaceutics Review.

U.S. Food and Drug Administration (US FDA) (2007a). Code of Federal Regulations, 21 CFR Part 58. Good Laboratory Practice for Nonclinical Laboratory Studies.

U.S. Food and Drug Administration (US FDA) (2007b). Code of Federal Regulations, 21 CFR part 201. Product Labeling.

U.S. Food and Drug Administration (US FDA) (2007c). Code of Federal Regulations, 21 CFR Part 320. Bioavailability and Bioequivalence Requirements.

U.S. Food and Drug Administration (US FDA) (2007d). Code of Federal Regulations, 21 CFR Part 361.1. Radioactive Drugs for Certain Research Uses.

U.S. Food and Drug Administration (US FDA) (2007e). Critical Path Opportunities for Generic Drugs.

U.S. Food and Drug Administration (US FDA) (2007f). *Drug Development and Drug Interactions: Presentations*, http://www.fda.gov/CDER/drug/drugInteractions/presentations.htm (accessed 8 March 2008).

U.S. Food and Drug Administration (US FDA) (2007g). FDA Clears Genetic Lab Test for Warfarin Sensitivity.

U.S. Food and Drug Administration (US FDA) (2007h). Guidance for Industry and Review Staff: Target Product Profile—A Strategic Development Process Tool (Draft).

U.S. Food and Drug Administration (US FDA) (2007i). Guidance for Industry: Clinical Trial Endpoints for the Approval of Cancer Drugs and Biologics.

U.S. Food and Drug Administration (US FDA) (2007j). Guidance for Industry: Good Laboratory Practices Questions and Answers.

U.S. Food and Drug Administration (US FDA) (2007k). Guidance for Industry: Role of HIV Resistance Testing in Antiretroviral Drug Development.

U.S. Food and Drug Administration (US FDA) (2007l). Guidance for Industry: Drug-Induced Liver Injury: Premarketing Clinical Evaluation (Draft).

U.S. Food and Drug Administration (US FDA) (2007m). Guidance for Industry: Pharmacogenomic Data Submissions—Companion Guidance (Draft).

U.S. Food and Drug Administration (US FDA) (2007n). Guidance for Industry: Dosage and Administration Section of Labeling for Human Prescription Drug and Biological Products—Content and Format (Draft).

U.S. Food and Drug Administration (US FDA) (2007o). *Investigational New Drug (IND) Application Process*, http://www.fda.gov/cder/regulatory/applications/ind_page_1.htm (accessed 3 March 2008).

U.S. Food and Drug Administration (US FDA) (2007p). Janumet—Summary Basis of Approval—Pharmacology Review.

U.S. Food and Drug Administration (US FDA) (2007q). Nanotechnology. A Report of the U.S. Food and Drug Administration Nanotechnology Task Force.

U.S. Food and Drug Administration (US FDA) (2007r). NDA Approval (Maraviroc). Letter from Edward Cox (FDA) to Pfizer, Inc.

U.S. Food and Drug Administration (US FDA) (2007s). Report on the performance of drug and biologics firms in conducting postmarketing commitment studies availability. *Federal Register*, *72* (22), 5069–5070.

U.S. Food and Drug Administration (US FDA) (2008a). Center for Drug Evaluation and Research. List of Guidance Documents.

U.S. Food and Drug Administration (US FDA) (2008b). *Drug-Induced Liver Toxicity*, http://www.fda.gov/cder/livertox/default.htm (accessed 6 March 2008).

U.S. Food and Drug Administration (US FDA) (2008c). *FDA Approved Drug Products*, http://www.accessdata.fda.gov/scripts/cder/drugsatfda/index.cfm (accessed 3 March 2008).

U.S. Food and Drug Administration (US FDA) (2008d). *FDA's Critical Path Initiative*, http://www.fda.gov/oc/initiatives/criticalpath/ (accessed 7 March 2008).

U.S. Food and Drug Administration (US FDA) (2008e). *Genomics at FDA*, http://www.fda.gov/cder/genomics/ (accessed 7 March 2008).

U.S. Food and Drug Administration (US FDA) (2008f). Guidance Agenda: Guidances CDER is Planning to Develop During Calendar Year 2008.

U.S. Food and Drug Administration (US FDA) (2008g). *Guidance Documents*, http://www.fda.gov/cder/guidance/index.htm (accessed 8 March 2008).

U.S. Food and Drug Administration (US FDA) (2008h). Guidance for Industry and Review Staff: Nonclinical Safety Evaluation of Reformulated Drug Products Intended for Administration by an Alternate Route (Draft).

U.S. Food and Drug Administration (US FDA) (2008i). Guidance for Industry: Diabetes Mellitus: Developing Drugs and Therapeutic Biologics for Treatment and Prevention (Draft).

U.S. Food and Drug Administration (US FDA) (2008j). Guidance for Industry: Safety Testing of Drug Metabolites.

U.S. Food and Drug Administration (US FDA) (2008k). *Guidances We Develop and Enforcement Information*, http://www.fda.gov/cder/regulatory/default.htm#Legislation (accessed 18 March 2008).

U.S. Food and Drug Administration (US FDA) (2008l). *Small Business Assistance*, http://www.fda.gov/cder/about/smallbiz/default.htm (accessed 10 March 2008).

U.S. Food and Drug Administration, Center for Drug Evaluation and Research (US FDA CDER) (2006). Oncologic Drugs Advisory Committee Transcript: Pre-Clinical Requirements and Phase 1 Trial Design Issues for the Development of Oncologic Products.

U.S. National Institutes of Health (2008). *PKC412 With or Without Itraconazole in Treating Patients With Acute Myeloid Leukemia or Myelodysplastic Syndrome*, http://clinicaltrials.gov/ct2/show/NCT00045578?intr=%22Itraconazole%22&rank=18 (accessed 7 March 2008).

Uno, S. *et al.* (2007). Decreased intestinal CYP3A and P-glycoprotein activities in rats with adjuvant arthritis. *Drug Metabolism and Pharmacokinetics*, *22* (4), 313–321.

Vickers, A.E.M. *et al.* (1999). Multiple cytochrome P-450s involved in the metabolism of terbinafine suggest a limited potential for drug–drug interactions. *Drug Metabolism and Disposition*, *27* (9), 1029–1038.

Villeneuve, J., Pichette, V. (2004). Cytochrome P450 and liver disease. *Current Drug Metabolism*, *5* (3), 273–282.

Viswanathan, C.T. *et al.* (2007). Workshop/conference report—quantitative bioanalytical methods validation and implementation: best practices for chromatographic and ligand binding assays. *The AAPS Journal*, *9* (1), E30–E42.

Vuong, L.T. *et al.* (2007). Use of accelerator mass spectrometry to measure the pharmacokinetics and peripheral blood mononuclear cell concentrations of zidovudine. *Journal of Pharmaceutical Sciences*, *97* (7), 2833–2843.

Ward, K.W., Smith, B.R. (2004). A comprehensive quantitative and qualitative evaluation of extrapolation of intravenous pharmacokinetic parameters from rat, dog, and monkey to humans. I. clearance. *Drug Metabolism and Disposition*, *32* (6), 603–611.

Ward, K.W., Smith, B.R. (2004). A comprehensive quantitative and qualitative evaluation of extrapolation of intravenous pharmacokinetic parameters from rat, dog, and monkey to humans. II. volume of distribution and mean residence time. *Drug Metabolism and Disposition*, *32* (6), 612–619.

Wasan, K.M., Cassidy, S.M. (1998). Role of plasma lipoproteins in modifying the biological activity of hydrophobic drugs. *Journal of Pharmaceutical Sciences*, *87* (4), 411–424.

Wasan, K.M. *et al.* (2008). Impact of lipoproteins on the biological activity and disposition of hydrophobic drugs: implications for drug discovery. *Nature Reviews. Drug Discovery*, *7* (1), 84–99.

Weiner, M., Kotkoskte, L. (2007). *Excipient Toxicity and Safety. Drugs and the Pharmaceutical Sciences*, Vol. 103, Informa Healthcare, New York.

Wikipedia (2008a). *TeGenero*, 31 May 2007, http://en.wikipedia.org/wiki/TeGenero (accessed 3 March 2008).

Wikipedia (2008b). *Paracelsus*, 2 Mar 2008, http://en.wikipedia.org/wiki/Paracelsus (accessed 4 Mar 2008).

Wildiers, H. *et al.* (2003). Pharmacology of anticancer drugs in the elderly population. *Clinical Pharmacokinetics*, *42* (14), 1213–1242.

Williams, J.A. *et al.* (2008). PhRMA white paper on ADME pharmacogenomics. *Journal of Clinical Pharmacology*, *48* (7), 849–889.

Zhou, H., Tong, Z., McLeod, J.F. (2004). "Cocktail" approaches and strategies in drug development: valuable tool or flawed science? *Journal of Clinical Pharmacol*, *44* (2), 120–134.

Zhou, S. *et al.* (2007a). Metabolic activation of herbal and dietary constituents and its clinical and toxicological implications: an update. *Current Drug Metabolism*, *8* (6), 526–553.

Zhou, S. *et al.* (2007b). Identification of drugs that interact with herbs in drug development. *Drug Discovery Today*, *12* (15–16), 664–673.

# *Drug Development Process and Global Pharmaceutical Marketplace*

Much more so then when the last edition of this book was written, pharmaceuticals[1] are global in impact, their regulation, and market.

This volume focuses on the assessment of the safety of new drugs. In the broadest sense, this means it must address not only the traditional "small molecules" that have dominated the field for the last century and the large therapeutic molecules derived from biotechnology sources but also vaccines, biologics such blood and blood products, and excipients. The globalization of the regulation of the safety, efficacy, and manufacture of their products comes from the success of the ICH (International Conference on Harmonisation) process. But, as will be seen, the same globalization of the industry and continuous our advance of science have also led to market diversification of the types and use of drugs and, with this, a fragmentation of regulatory drug safety evaluation requirements which has made things more complex rather than simpler.

[1]The term *pharmaceuticals* is here used in the broadest sense of man-made therapeutics: small molecules and large pills and vaccines, blood products, and, as must be, their attendant components (excipients, impurities, and all) to different degrees and in different products.

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

## 1 THE MARKETPLACE

The world marketplace for drugs is large, although the majority of sales are in three regions: about 51% of the pharmaceutical market resides in the United States, about 25% in Europe, and 15% in Japan. The balance of sales is spread across the globe. This does not mean, however, that the marketing applicant should ignore the requirements of other countries (e.g., Indonesia). Approval processes in these countries can, at times, be as rigorous as in any other regulatory authority domain.

Pharmaceuticals in all their forms compete today as part of a global market, though one which serves (and is available to) different parts of the world's population too.

According to the Intercontinental Marketing Service (IMS) 2008 global pharmaceutical market and therapy forecast, the global market for regulated drugs (as differentiated from dietary supplements, herbal products, and nutraceuticals) is estimated to be some $690 billion in 2007 (U.S. dollars). The same forecast projected growth to $735–745 billion in 2008 (Goozner, 2004). In 2006, there were 109 individual products with annual sales excess of $1 billion (Greider, 2003) which have tended to be the focus of pharmaceutical development until recently, but the impending demise of patents on which is changing the industry (Table 1).

This concentration of total sales in a limited number of products (e.g., there are currently more than 21,000 approved drugs in the United States) is widely held to have distorted the therapeutic aspects of new drug development but is now starting to undergo change (back to) a paradigm that looks at a decreased emphasis on the $1 billion "blockbuster" drugs.

Widely misunderstood is the extent of the pharmaceutical research and development (R&D) sector. While precise numbers are unavailable (and meaningless, as companies are continuously starting, merging, or going out of business though the overall trend is to increased numbers), best estimates place the number of companies directly involved in discovering and developing new drugs in the United States and Canada at 3700. There are an equal number in Europe and significant numbers in many other parts of the world (China, Australia, India, and Israel, to name a few). While most of the public focuses on the very large companies, such as those in Table 2, there are many more midsize and small companies.

One factor to consider in the regulatory requirements for early development of new therapeutic entities is the degree of barrier which costs may present to the smaller, innovative companies. This is commonly overlooked by many who also do not recognize that such small companies (most of which fail) are the primary initial source of new therapeutics.

A second complicating factor in considering the "pharmaceutical" market sector is the diversity of products involved. The most basic expression of this is the division of drugs into "small molecules" [which currently constitute about two-thirds of both investigational new drugs (INDs)—applications for clinical

**TABLE 1 Top 20 Selling Pharmaceuticals (2006)**

| Rank (2005) | Medicine | 2005 Sales ($M) | 2004 Sales ($M) | Company | Primary Diseased Medical Use | First Approval Date(s) | Route(s) |
|---|---|---|---|---|---|---|---|
| 1 | Lipitor | 12,986 | 11,587 | Pfizer, Astellas Pharma | Cholesterol | Dec.17, 1996 | Oral |
| 2 | Plavix/Iscover | 6,345 | 5,434 | Bristol-Myres Squibb, Sanofi-Aventis | Thrombotic events | Nov. 17, 1997 | Oral |
| 3 | Advair/Seretide | 5,465 | 4,503 | GlaxoSmithKline | Asthma | Nov. 23, 1999/ Sept. 7, 1998 | Inhalation |
| 4 | Norvasc | 4,706 | 4,463 | Pfizer | Hypertension | July 31, 1992 | Oral |
| 5 | Nexium | 4,633 | 3,883 | AstraZeneca | Gastrointestinal disorders | Mar. 2000 | Oral, parenteral |
| 6 | Zocor | 4,382 | 5,197 | Merck & Co. | Cholesterol | Dec. 23, 1991 | Oral |
| 7 | Zyprexa | 4,202 | 4,420 | Eli Lilly | Schizophrenia | Sept. 27, 1996 | Oral, injection |
| 8 | Prevacid/ Takepro | 3,996 | 4,050 | Tap Pharmaceutical, Takeda Pharmaceutical | Gastrointestinal disorders | May 10, 1995 | Oral, injection |
| 9 | Diovan group | 3,676 | 3,093 | Novartis | Hypertension | Dec. 23, 1996 | Oral |
| 10 | Enbrel | 3,657 | 2,580 | Amgen, Wyeth | Rheumatoid arthritis | Nov. 2, 1998 | Injection |
| 11 | Risperdal | 3,552 | 3,050 | Johnson & Johnson | Schizophrenia | Dec. 29, 1993 | Oral, injection |
| 12 | Remicade | 3,547 | 2,920 | Johnson & Johnson, Schering-Plough, Tanabe | Rheumatoid arthritis | Aug. 24, 1998 | Injection |
| 13 | Effexor | 3,459 | 3,347 | Wyeth | Depression | Dec. 28, 1993 | Oral |
| 14 | Protonix/ Pantozol | 3,428 | 3,105 | Wyeth, Atlana | Gastrointestinal disorders | Feb. 2, 2000/1994 | Oral, injection |
| 15 | Rituxan/ MabThera | 3,334 | 2,711 | Roche, Genentech | Non-Hodgkin's lymphoma | Nov. 26, 1997 | Injection |
| 16 | Procrit/Eprex | 3,324 | 3,589 | Johnson & Johnson | Anemia | Dec. 31, 1990/ May 4, 1995 | Injection |
| 17 | Aranesp | 3,273 | 2,473 | Amgen | Anemia | June 11, 2001 | Injection |
| 18 | Zoloft | 3,256 | 3,361 | Phizer | Depression | Dec. 30, 1991 | Oral |
| 19 | Fosamax | 3,191 | 3,160 | Merck & Co. | Osteoporosis | Sept. 29, 1995 | Oral |
| 20 | Cozaar, Hyzaar | 3,037 | 2,824 | Merck & Co. | Hypertension | Apr. 14, 1995, Apr. 28, 1995 | Oral |

**TABLE 2 Top 25 Drug Companies by Sales (2006)**

| Company | Pharma Sales ($ million) | % Change | Pharma Sales as % of Total Sales |
|---|---|---|---|
| Pfizer | 45,083 | 1.8 | 95.9 |
| GlaxoSmithKline | 40,156 | 25.1 | 86.5 |
| Sanofi-Aventis | 38,555 | 29 | 100 |
| Roche | 27,290 | 31.8 | 79.2 |
| AstraZeneca | 26,475 | 10.5 | 100 |
| Johnson & Johnson | 23,267 | 4.2 | 43.6 |
| Novartis | 22,576 | 11.4 | 62.7 |
| Merck & Co | 20,375 | –6.6 | 90 |
| Wyeth | 16,884 | 10.2 | 83 |
| Lilly | 15,691 | 13.9 | 100 |
| Bristol-Myers Squibb | 13,861 | –9.1 | 77.4 |
| Boehringer Ingelheim | 13,860 | 27.6 | 96.5 |
| Amgen | 13,858 | 15.3 | 100 |
| Abbott Laboratories | 12,395 | –8.9 | 55.2 |
| Bayer | 10,162 | –9 | 25.8 |
| Takeda | 8,716 | –3.9 | 88.6 |
| Schering-Plough | 8,561 | 13.2 | 80.8 |
| Teva | 7,821 | 65.5 | 93.1 |
| Genentech | 7,640 | 39.2 | 100 |
| Astellas | 7,390 | 71.2 | 98.9 |
| Novo Nordisk | 7,087 | 32.3 | 100 |
| Daiichi Sankyo | 6,790 | 6.7 | 90.1 |
| Baxter International | 6,461 | 67.7 | 62.3 |
| Merck KGaA | 5,643 | 22.4 | 65.8 |
| Eisai | 4,703 | 8.3 | 96.4 |

evaluation of a new drug in humans–and current new drug approvals] and biotechnology products (which constitute the other third). The challenges in both developing and assessing the safety of these are very different. As will also be seen, if one considers further division into therapeutic claim areas [e.g., oncology, anti-infectives, cardiovascular, central nervous system (CNS)], the differences become even more marked. Most of what will be presented and discussed in this volume speaks to regulatory requirements for nonclinical safety assessment in the general case for either small molecules or protein therapeutics. It should be kept in mind that this general case development model never applies.

Additionally, there is now a significant hybrid area—combination products, which include both device and drug (small-molecule or biological) components. These will be addressed in a separate chapter of the book, though there is no single dedicated regulatory arm [such as a center within the U.S. Food and Drug Administration (FDA) dedicated to only their regulation] in any major market country. For that reason, more exploration of regulatory considerations will be provided in the chapter on these products.

The extent of regulations and practices for drug approval cause pharmaceutical companies to spend an enormous amount of resources on developing

applications, following different standards for preclinical and nonclinical programs for specific therapeutic areas as well as time and resources to satisfy the regulatory processes for clinical trials. Because of the regulatory diversity that existed, representatives from the regulatory authorities and trade associations came together in the late 1980s and early 1990s to attempt at harmonizing the process for drug approvals. Clearly this was a daunting task. With time, however, the International Conference on Harmonisation (ICH) of Technical Requirements for Registration of Pharmaceutical for Human Use has become increasingly more effective. Japan, Europe, and the United States represent the major pharmaceutical markets for the world, and these regions have the most influence on developments within the ICH and tend to follow the guidance documents that are prepared. However, other countries [rest of the world (ROW)] follow the developments within the ICH and tend to follow the guidance offered by the ICH. However, when seeking registration of pharmaceuticals, it remains important to be aware of local country regulations. For example, China is becoming a major economic force in many aspects. Placement of pharmaceutical manufacturing facilities and the marketing of drugs in China may potentially represent a significant marketing advantage to companies. With this new market area in Asia, regulatory processes are being developed sometimes—it seems at the whim of the government. With time it is hoped that China aligns itself more with the processes and guidance that have been developed by the ICH, FDA, and other developed countries.

## 2 HISTORY OF MODERN THERAPEUTICS

Although prior to the nineteenth century preventive medicine had made some spectacular advances, for example, through nutrition (scurvy), controlling infectious diseases (vaccination), public health through sanitation, and controlling childbirth fever and surgical infections using antiseptic techniques, truly therapeutic medicine was virtually nonexistent until the end of the nineteenth century (Mathieu, 2000; Rang, 2006).

Oliver Wendell Holmes (a physician and U.S. Supreme Court justice) wrote in 1860: "I firmly believe that if the whole material medica, as now used, could be sunk to the bottom of the sea, it would be all the better for mankind—and the worse for the fishes." While there were a few effective medicines—digitalis, extract of willow bark, and quinine—on balance Holmes was quite correct—medicines did more harm than good.

An idea of the state of therapeutics at the time comes from the first edition of the British Pharmacopoeia, in 1864, which listed 311 preparations. Of these, 187 were plant-derived materials, only 9 of which were purified substances. Most of the plant products—lemon juice, rose hips, yeasts, and so on—lacked any components we would now regard as therapeutically relevant, but some—digitalis, castor oil, ergot, colchicum—were pharmacologically active. Of the 311 preparations, 103 were truly synthetic inorganic chemicals, such as iodine,

ferrous sulfate, sodium bicarbonate, and toxic salts of bismuth, arsenic, lead, and mercury, with but a few synthetic chemicals (diethyl ether and chloroform). The remainder was miscellaneous materials and a few animal products, such as lard, cantharidin, and cochineal.

For the pharmaceutical industry, the transition from theoretical to actual an history occurred late in the nineteenth century, when three essential technologies came together—biomedicine (especially pharmacology), synthetic organic chemistry, and the development of a chemical industry in Europe—coupled with development of a medical supplies products-trade.

Science began to be applied wholeheartedly to medicine—as to almost every other aspect of life—only late in the nineteenth century. Among the most important milestones from the point of view of drug discovery was the elaboration in 1858 of cell theory. This tremendous reductionist leap of cell theory gave biology—and the pharmaceutical industry—the fundamental scientific underpenning it required. Only by thinking of living systems in terms of the function of their cells can one begin to understand how molecules affect them.

A second milestone was the birth of pharmacology as a scientific discipline when the world's first Pharmacological Institute was set up in 1874 at Dorpat by Rudolf Buchheim—literally by Buchheim himself, as the institute was in his own house and funded by him personally. This was advanced by pioneers such as Magendie and Claude Bernard, who linked pharmacology to therapeutics.

Another vital spark on this road came with Louis Pasteur's germ theory of disease, proposed in Paris in 1878. A chemist by training, Pasteur's initial interest was in the process of fermentation of wine and beer and the souring of milk. He showed, famously, that airborne infection was the underlying cause and concluded that the air was alive with microorganisms. Particular types, he argued, were pathogenic to humans and accounted for many forms of disease, including anthrax, cholera, and rabies. Pasteur successfully introduced several specific immunization procedures to give protection against infectious diseases. Robert Koch, Pasteur's rival and near contemporary, clinched the infection theory by observing anthrax and other bacilli in the blood of infected animals.

The founder of chemotherapy—some would says the father of molecular pharmacology—was Paul Ehrlich. He invented "vital staining"—staining by dyes injected into living animals—and described how the chemical properties of the dyes, particularly their acidity and lipid solubility, influenced the distribution of dye to particular tissues and cellular structures. Thence came the idea of specific binding of molecules to particular specific binding of molecules to particular cellular components. This not only led to Ehrlich's study of chemotherapeutic agents but also became the basis of pharmacological thinking to the present day. "Receptors" and "magic bullets" were Ehrlich's terms, though he envisaged receptors as targets for toxins, rather than physiological mediators. Working at Koch's Institute, Ehrlich developed diphtheria antitoxin

for clinical use and put forward a theory of antibody action based on specific chemical recognition of microbial molecules, for which he won the 1908 Nobel Prize.

The first synthetic organic chemicals to be used for medical purposes were not therapeutic agents at all, but rather anesthetics. Diethyl ether ("sweet oil of vitriol") was first made and described in 1540. It and nitrous oxide (prepared by Humphrey Davy in 1799 and found, by experiments on himself, to have stupor-inducing properties) had their usefulness as surgical anesthetics demonstrated only in the 1840s, by which time chloroform had also made its appearance. Synthetic chemistry at the time could deal only with very simple molecules, made by recipe rather than reason, as our understanding of chemical processes and molecular structure was still in its infancy. The first therapeutic drug to truly come from synthetic chemistry was amyl nitrite, prepared in 1859 by Guthrie and used to treat angina by Brunton in 1864. This was the first example of a drug born in a recognizably "modern" way, through the application of synthetic chemistry, physiology, and clinical medicine. This was a landmark indeed, for it was nearly 40 years before synthetic chemistry made any further significant contribution to therapeutics, and it was not until well into the twentieth century that physiological and pharmacological knowledge began to be applied to the invention of new drugs.

During the latter half of the nineteenth century the foundations of synthetic organic chemistry were laid, the impetus coming from work on aniline, a copious byproduct of the coal-tar industry, with the discovery of how to produce a purple dye. This discovery gave birth to the synthetic dyestuff industry, which played a major part in establishing the commercial potential of synthetic organic chemistry—a technology which became the underpinning of the evolving pharmaceutical industry for the next century. A systematic approach to organic synthesis went hand in hand with improved understanding of chemical structure.

Despite the limited efficacy of the pharmaceutical preparations available in the nineteenth century ("patent medicines") the pharmacists' trade flourished; then, as now, physicians felt obligated to issue prescriptions to satisfy the expectations of their patients for some therapeutic action—or at least cause for hope. Early in the nineteenth century, a few enterprising chemists undertook the task of isolating the active substances from the plants extracts. The trend began with Friedrich Serturner, a junior apothecary in Westphalia, who in 1805 isolated and purified morphine, barely surviving a test of its potency on himself. This was the first "alkaloid," so named because of its ability to neutralize acids and form salts. This discovery in turn led to the isolation of other plant alkaloids, including strychnine, caffeine, and quinine. The recognition that medicinal plants owed their properties to their individual chemical constituents, rather than to some intangible property associated with their living nature, marks a critical point in the history of the pharmaceutical industry which can be recognized as the point of origin of two of the three roads from which the industry grew—the beginnings of the "industrialization" of the

pharmaceutical trade. It hinted at the future possibility of making drugs artificially.

The first local apothecary business to move into large-scale production and marketing of pharmaceuticals was the old-established Darmstadt firm Merck, founded in 1668. This development, in 1827, was stimulated by the advances in purification of natural products. Merck was closely followed in this astute business move by other German- and Swiss-based apothecary businesses, giving rise to some which later also became giant pharmaceutical companies, such as Schering and Boehringer. The American pharmaceutical industry emerged in the middle of the nineteenth century. Squibb began in 1858, with ether as its main product. The move to pharmaceuticals was also followed by several chemical companies, such as Bayer, Hoechst, Agfa, Sandoz, Geigy, and others which began as dyestuff manufacturers. The dyestuff industry at that time was also based largely on plant products, which had to be refined and were sold in relatively small quantities, so the commercial parallels with the pharmaceutical industry were plain.

After 1870, with the crucial discovery by Kekule of the structure of benzene, the dyestuff industry turned increasingly to synthetic chemistry as a source of new compounds, starting with aniline-based dyes. A glance through any modern pharmacopeia will show the overwhelming preponderance of synthetic aromatic compounds, based on the benzene ring structure, among the list of useful drugs. Understanding the nature of aromaticity was critical.

Thus the beginning of the pharmaceutical industry as we now know it dates from the 1800s, with origins in the apothecary and patent medicine trades on the one hand and the dyestuff industry on the other. Unfortunately, these firms had few effective products to sell (mainly inorganic compounds of varying degrees of toxicity and others most charitably described as concoctions).

Entering the 1900s, synthetic drugs were made and tested, including the "antipyretics" and various CNS depressants. Chemical developments based on chloroform produced chloral hydrate, the first nonvolatile CNS depressant, which was in clinical use for many years as a hypnotic drug. Independently, various compounds based on urea were found to act similarly, and von Mering followed this lead to produce the first barbiturate, barbitone (since renamed barbital), which was introduced in 1903 by Bayer and gained widespread clinical use as a hypnotic, tranquillizer, and antiepileptic drug—the first blockbuster. Barbitone and procaine were triumphs for chemical ingenuity but owed little or nothing to physiology or indeed pharmacology. The physiological sites of action of barbiturates remain unclear to this day, and their mechanism of action at the molecular level was unknown until the 1980s.

The pattern of drug discovery driven by synthetic chemistry—with biology often struggling to keep up—became the established model in the early part of the twentieth century and prevailed for at least 50 years. The balance of research in the pharmaceutical industry up to the 1970s clearly placed chemistry as the key discipline in drug discovery, the task of biologists being mainly

to devise and perform assays capable of revealing possible useful therapeutic activity among the many anonymous white powders that arrived for testing. Research management in the industry was largely in the hands of chemists. This strategy produced many successes, including benzodiazepine tranquillizers, several antiepileptic drugs, antihypertensive drugs, antidepressants, and antipsychotic drugs. The surviving practice of classifying many drugs on the basis of their chemical structure rather than on the more logical basis of their site or mode of action (therapeutic class) stems from this era. We have mentioned the early days of pharmacology, with its focus on plant-derived materials, such as atropine, tubocurafine, strychnine, digitalis, and ergot alkaloids, which were almost the only drugs that existed until well into the twentieth century. Despite the rise of synthetic chemistry natural products remain a significant source of new drugs, particularly in the field of chemotherapy, but also in other applications. Following the discovery of penicillin by Fleming in 1929—and its development as an antibiotic for clinical use by Chain and Florey in 1938—an intense search was undertaken for antibacterial compounds produced by fungi and other microorganisms which yielded many useful antibiotics, including chloramphenicol (1947), tetracyclines (1948), and streptomycin (1949). The same fungal source that yielded streptomycin also produced actinomycin D, used in cancer chemotherapy Higher plants have continued to yield useful drugs, including vincristine and vinblastine (1958), and paclitaxel (or taxol, 1971).

Outside the field of chemotherapy, successful drugs derived from natural products include cyclosporin (1972) and tacrolimus (1993), both of which come from fungi and are used to prevent transplant rejection. Soon after came mevastatin (1976), another fungal metabolite, which was the first of the "statin" series of cholesterol-lowering drugs which act by inhibiting the enzyme human menopausal gonadotrophin (HMG) coenzyme A (CoA) reductase.

Overall, the pharmaceutical industry continues to have something of a on-again, off-again relationship with natural products. They often have weird and wonderful structures that cause hardened chemists to turn pale, they are often near-impossible to synthesize and troublesome to produce from natural sources, and "optimizing" such molecules to make them suitable for therapeutic use is prone to frequent failure. But nature continues to unexpectedly provide some of our most useful drugs, and most of its potential remains untapped.

Although chemistry was the preeminent discipline in drug discovery until at least the 1970s, the seeds of the biological revolution were sown long before. It started foremost in the field of chemotherapy, where Ehrlich defined the principles of drug specificity in terms of a specific interaction between the drug molecule and a target molecule—the "receptor site"—in the organism. Although we now take it for granted that in almost all cases a highly specific chemical target molecule, as well as the "pharmacophore" or an outline portion of the drug molecule, determines what effects a therapeutic will yield, before Ehrlich no one had envisaged drug action in this way. By

**TABLE 3 Examples of Drugs from Different Sources**

| Natural Products | Synthetic Chemistry | Biopharmaceuticals Produced by Recombinant DNA Technology |
|---|---|---|
| Antibiotics (penicillin, streptomycin, tetracyclines, cephalosporins, etc.) | Early successes include: | Human insulin (the first biotech product, registered 1982) |
| Anticancer drugs (doxorubicin, bleomycin, actinomycin, vincristine, vinblastine, taxol etc.) | Antiepileptic drugs | Human growth hormone |
| Atopine, hyoscine | Antihypertensive drugs | α-Interferon, γ-interferon |
| Ciclosporin | Antimetabolites | Hepatitis B vaccine |
| Cocaine | Barbiturates | Tissue plasminogen activator (t-PA) |
| Colchicine | Bronchodilators | Hirudin |
| Digitalis (digoxin) | Diuratics | Blood-clotting factors |
| Ephedrine | Local anaesthetics | Erythropoietin |
| Heparin | Sulfonamides | Granulocyte and granulocyte–monocyte colony-stimulating factor (G-CSF, GM-CSF) |
| Human growth hormone[a] | (Since ca.1950, synthetic chemistry has accounted for the great majority of new drugs) | |
| Insulin (porcine, bovine)[a] | | |
| Opium alkaloids (morphine, papaverine) | | |
| Physostigmine | | |
| Rauwolfia alkaloids (reserpine) | | |
| Statins | | |
| Streptokinase | | |
| Tubocurarine | | |
| Vaccines | | |

[a]Now largely or entirely replaced by material prepared by recombinant DNA technology.

linking chemistry and biology, Ehrlich defined the parameters of modern drug discovery.

Despite these discoveries in Ehrlich's field, chemotherapy remained empirical rather than target directed. For many years, Ehrlich's preoccupation with the binding of chemical dyes, as exemplified by biologicals in the 1950s onward, steadily shifted the industry's focus from chemistry to biology (Lednicer, 1993). The history of successes in the field of chemotherapy prior to the antibiotic era (Table 3) demonstrates the diversity of sources of new therapeutic entities. The popular image of magic bullets (a phrase invented by Ehrlich) is the essence of today's target-directed approaches to drug discovery.

## 3 DRUG DEVELOPMENT PROCESS

While processes for the discovery of new potential therapeutic drugs are very diverse (Gad, 2007), once the decision is made to move a candidate com-

pound forward to (hopefully) market approval, the process is well defined in the components of its regulatory requirements. It has many components which are beyond the scope of safety assessment and therefore of this volume (including chemical development, clinical evaluation, and a host of regulatory actions.)

The process generally proceeds by way of getting regulatory concurrences for entering clinical trials, then proceeding through three (not strictly defined) stages of clinical trials (phases I–III), followed by submission of a full set of documents, data, and a proposed label seeking regulatory approval for a marketing application.

The metrics of this process as it now operates makes cancer the most prevalent therapeutic target for new drugs, with perhaps as many as one-third of all new drug candidates being in this claim area. Heart diseases, CNS diseases, nervous system diseases, and immune system disorders follow in order of current popularity (Table 4).

According to Pharma & BioIngredients (www.pharmabioingredients.com), the more than 16,000 different drugs in development in 2006 were spread across the entire course of the development process (Table 5).

At the same time, the metrics of regulatory applications for the development of new drugs in the United States (where the best data are available) show a continued increase in the number of candidates entering the development process, as indicated by the number of new (or original) INDs filed, with the proportion of these that are commercial (or traditional INDs) continuing to increase (see Table 6).

**TABLE 4 Potential New Drugs in U.S. Clinical Trials, 2005–2006**

| | |
|---|---|
| Cancer | 5468 |
| Heart disease | 2342 |
| Mental and behavioral disorders | 2397 |
| Bacterial and fungal diseases | 1591 |
| Blood and lymph conditions | 1654 |
| Digestive system diseases | 1527 |
| Nervous system diseases | 2928 |
| Rare diseases | 5765 |
| Respiratory tract diseases | 1548 |
| Viral diseases | 1168 |
| Injuries, poisonings, and occupational diseases | 832 |
| Immune system disorders (not including HIV/AIDS) | 2578 |
| Disease abnormalities at or before birth | 1090 |
| Gland- and hormone-related diseases | 1216 |
| Muscle, bone, and cartilage diseases | 699 |
| Nutritional and metabolic diseases | 1296 |
| Skin and connective tissue diseases | 1727 |
| Symptoms and general pathology | 4227 |
| Urinary tract and sexual organs and pregnancy | 1756 |

**TABLE 5 2006 Status of Drug in Development**

| Stage | Drugs |
|---|---|
| New drug application (NDA)/Biological license application (BLA) filed | 482 |
| Phase III | 1,179 |
| Phase II | 2,622 |
| Phase I/IND filed | 2,415 |
| Preclinical/discovery | 7,569 |
| Recent product launches | 2,002 |
| Total | 16,269 |

**TABLE 6 INDs Received and Active at Center for Drug Evaluation and Research**

| Calendar Year Received | Original INDs Received | Number of Active INDs at Year End | NDAs |
|---|---|---|---|
| 1998 | 2,419 | 12,723 | 121 |
| 1999 | 1,763 | 12,584 | 139 |
| 2000 | 1,812 | 11,838 | 115 |
| 2001 | 1,872 | 10,873 | 98 |
| 2002 | 2,374 | 11,544 | 105 |
| 2003 | 2,120 (426 commercial) | 12,661 (4544 commercial) | 109 |
| 2004 | 1,837 (621 commercial) | 12,778 (4827 commercial) | 115 |
| 2005 | 1,936 (637 commercial) | 13,360 (5029 commercial) | 116 |
| 2006 | 1,863 (713 commercial) | 14,117 (5445 commercial) | 123 |

At the same time the rate of approval of new molecular entities has decreased (and stayed stable at) 17 or 18 a year for the last three years. This has finally caused recognition that the traditional/existing system of development that focused on blockbusters is irretrievably broken.

## 4 STRATEGIES FOR DEVELOPMENT: LARGE VERSUS SMALL COMPANY OR SHORT VERSUS LONG GAME

While harmonization and societal concern for safety are driving the changes in regulatory processes for device and drug development to become less confused, strategies for product development and the associated nonclinical safety assessment can still be viewed in broad trends.

The truths driving strategies in developing new drugs are as follows:

1. Most molecules will fail. While the true success rate is certainly greater than the often-quoted 1 in 10,000, it is clear that only 3–5% of those that enter initial clinical evaluation (that is, for which an IND "opens") become marketed drugs. This rate varies depending on the therapeutic class (oncology drugs dosing at a rate as low as 1–2% and CNS therapeutics being only somewhat higher) (Czerepak and Rysef, 2008; Choerghade, 2006).

2. The cost of developing drugs is high. While not the currently quoted "average" of $1.4 billion, just getting to the point of an IND opening will cost a minimum of $2 million. One can spread out the rate of expenditure over time or shorten the required time by spending money more rapidly, but there are fixed minimums for cost and time.

And costs of development go up sharply with time/progress. Subsequent to a plain vanilla first-in-man (FIM) trial, outlays come to be spoken of first in tens of millions and (frequently), before a marketing approval filing, in the hundreds of millions. Once the decision is made to develop a molecule into a drug, the process takes years. One can dispute how many (from 5 to 16 years covers the extreme range), and at no point up to the end is success (achieving marketing approval and economically successful therapeutic use) assured. These truths conspire to produce the principal general goals behind a drug development strategy:

1. Kill the losers as early as possible, before too much money is spent on them.
2. Do all you can to minimize the time spent in developing a drug.

These principles produce a spectrum of strategies in the nonclinical safety assessment of drugs, best illustrated by looking at the two extreme cases.

*Do only what you must*. Driven by financial limitations and the plan that, at an optimal point in development (most commonly after either FIM/phase I trials or a "proof-of-concept" phase II trial), the candidate therapeutic will be licensed to or partnered with a large company, only the technical and regulatory steps necessary to get a molecule to this point are to be performed. For those pursuing this case, the guidance provided by this book should prove essential (though not generally completely sufficient). This approach in summarized in Figure 1.

*Minimize the risk of subsequent failure*. This is considered the traditional big company model. Studies and technical tasks are not limited to the minimum but rather are augmented by additional components. Development proceeds through a series of well-defined and carefully considered "go/no-go" decision points. This approach is summarized in Figure 2. Many of the additional components are either limited, non-GLP (good laboratory practice) forms of studies which will be required later [such as Ames, acute toxicity, human ether-a-gogo related gene (hERG) at only one concentration and 7 days to 4 weeks repeat-dose studies] or studies which are inexpensive and could be done later [cytochrome p. 450 (CYP) inhibitors and induction, metabolic stability, and longer than required repeat-dose toxicity studies before proceeding into phase II]. Exactly which "extra" components are included vary from company to company and frequently reflect past experiences of the organization or individuals involved.

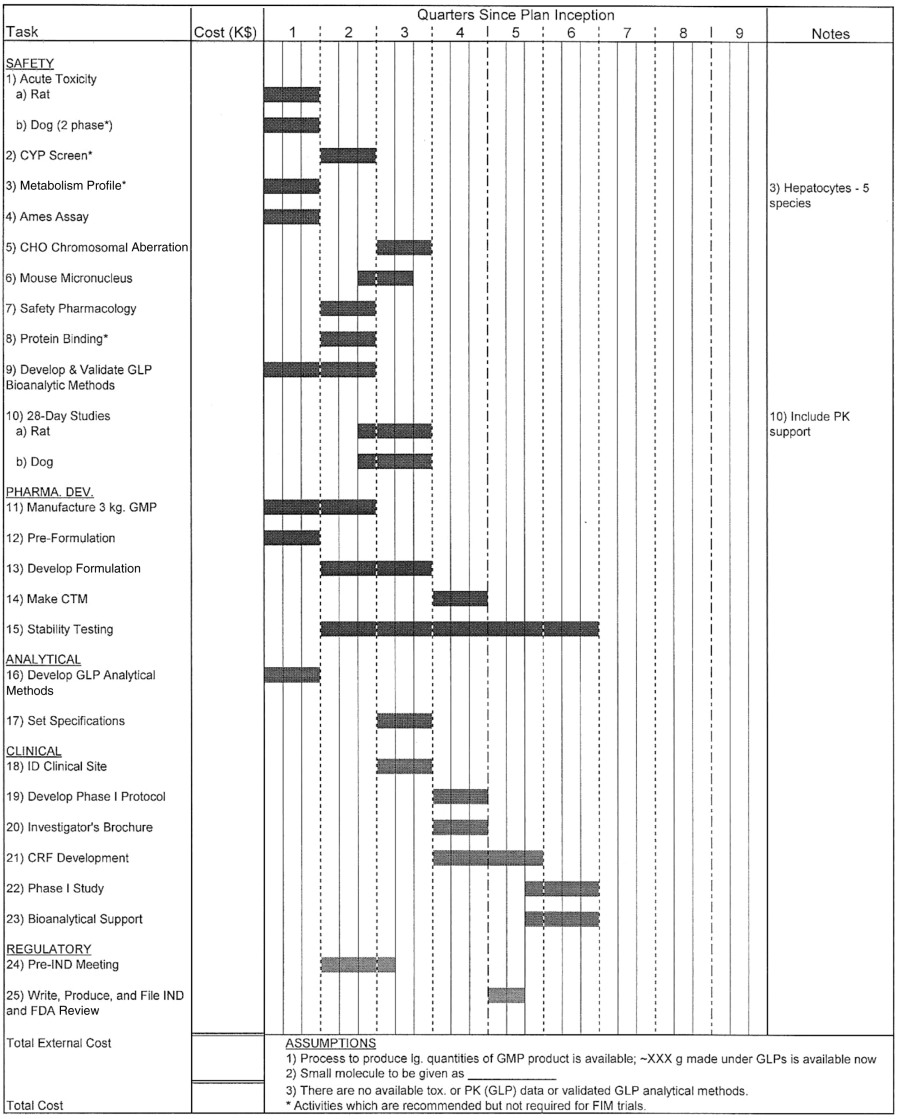


***Figure 1*** *General case oral drug: lead through phase 1 (do only what you must).*

The studies performed to meet regulatory nonclinical safety assessment requirements (which must be considered to include all of the supportive toxicokinetic and metabolism activities and studies) can be thought of as belonging to three major categories.

(a) Those necessary to support the successful filing/opening of an IND, clinical trial application (CTA), or equivalent application and of the subsequent FIM clinical studies.

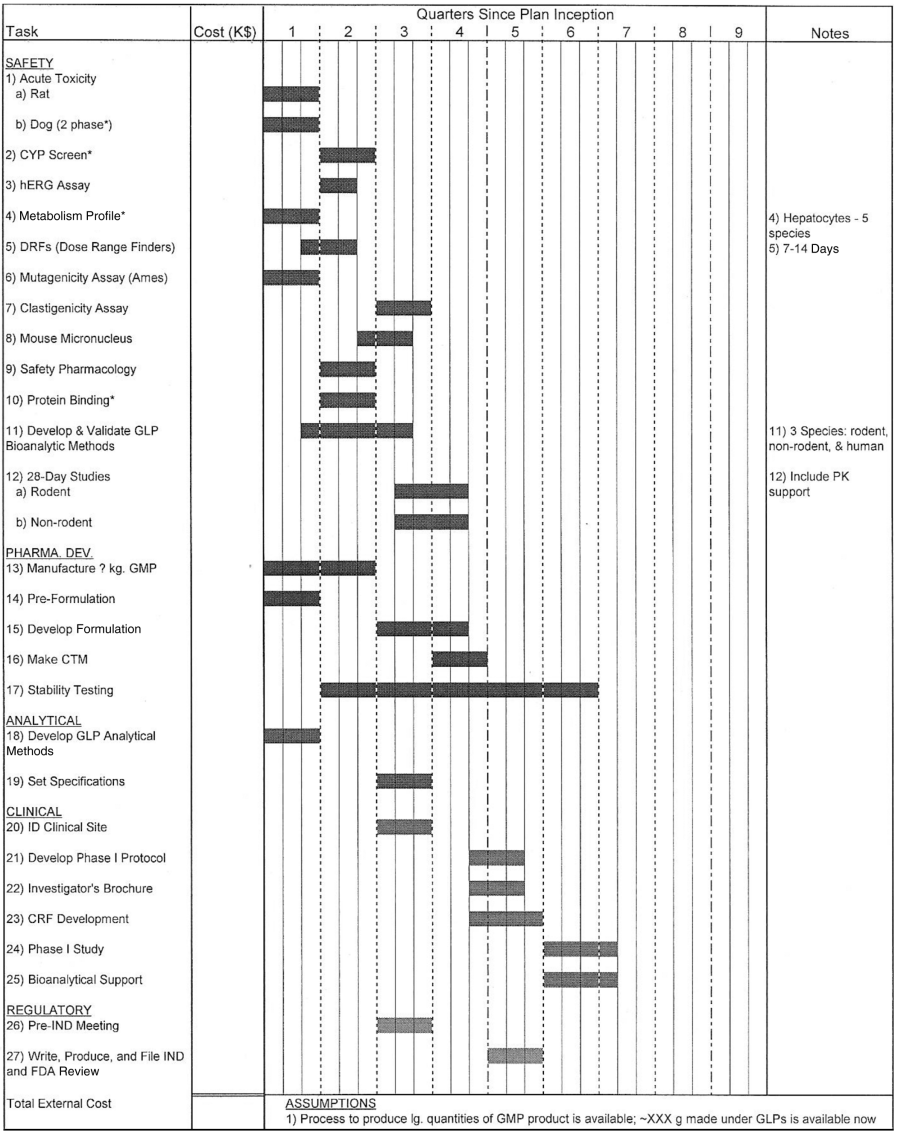


***Figure 2*** *General case oral drug: lead through phase 1 (minimize risk).*

(b) Those required to support continuation of clinical evaluation and development of up to and through successful phase III studies.

(c) Those studies required to support a successful marketing, approval application (NDA, BLA, or equivalent), but only required as such. This group is typically exemplified for carcinogenicity studies and the formal reproductive (as opposed to developmental) toxicity studies.

Which studies fit into what category is somewhat fluid and influenced by what patient population will be served (therapeutic claim) and the mechanism of action of the drug.

## 5 SAFETY ASSESSMENT AND EVOLUTION

In the mid-nineteenth century restrictions on the sale of poisonous substances were imposed in the United States and United Kingdom, but it was not until the early 1900s that system of "prescription-only" medicines was introduced, requiring approval of purchase by a licensed medical practitioner. Soon afterward, restrictions began to be imposed on what "cures" could be claimed in advertisements for pharmaceutical products and what information had to be given on the label; legislation evolved at a leisurely pace. Most of the concern was with controlling poisonous or addictive substances or contaminants, not with the efficacy and possible harmful effects of new drugs.

In 1937, the use of diethylene glycol as a solvent for a sulfonamide preparation caused the deaths of 107 children in the United States, and a year later the 1906 Food and Drugs Act was revised, requiring safety to be demonstrated before new products could be marketed as well as federal inspection of manufacturing facilities. The requirement for proven efficacy as well as safety was added in the Kefauver–Harris amendment in 1962 (said amendment being brought about largely by a safety issue—the thalidomide disaster in Europe).

In Europe, preoccupied with the political events in the first half of the century, matters of drug safety and efficacy were a minor concern, and it was not until the mid-1960s, in the wake of the thalidomide disaster—a disaster averted in the United States by an officer who used the provisions of the 1938 Food and Drugs Act to delay licensing approval—that the United Kingdom began to follow the U.S. lead in regulatory laws. Until then, the ability of drugs to do harm—short of being poisonous or addictive—was not really appreciated, most of the concern having been about contaminants. In 1959, when thalidomide was first put on the market by the German company Chemie Grumenthal, it was up to the company to decide how much research was needed to satisfy itself that the drug was safe and effective. Grumenthal made a disastrously wrong judgment [see Sjostrom and Nilsson (1972) for a full account], which resulted in an estimated 10,000 cases of severe congenital malformation following the company's specific recommendation that the drug was suitable for use by pregnant women. This single event caused an urgent reappraisal on a global scale, leading to the introduction of much tighter government controls.

By the end of the 1960s, the primary planks in the regulatory platform—evidence of safety, efficacy, and chemical purity—were in place in most developed countries. Subsequently, the regulations were adjusted in various minor ways and were adopted with local variations in most countries.

In 1988, Alder and Zbinden published national and international drug safety guidelines which set forth the wide differences in safety assessment requirements between the different nations of the world, at the time global development of a drug required multiple safety assessment programs, with a great number of repetition of studies and attendant extra costs and increased usage of test animals.

The solution to this was the ICH paradigm, which starting in the late 1980s sought to have a harmonized set of global requirements for all aspects of drug development (not just assessment). The safety assessment aspects were embodied primarily in the S series ICH guidelines (M4, which sets forth the overall structure of nonclinical requirements, being an exception). This did serve to largely regularize global requirements with minor differences.

As the rest of this book will make clear, this system is now fraying a bit of the edges. Recent additions of new guideline topic areas (e.g., immunotoxicology) and revisions to existing guideline (on genotoxicity and biotechnology) as well as recent regional guideline responses to recent occurrences [the case in point being the failed TGN-412 FIM trial and the resulting two European Medicines Agency (EMEA) special guidances issued in response to it] and differences in requirements for different therapeutic classes have reversed the harmonization trend.

## 6 THREE STAGES OF DRUG SAFETY EVALUATION ON GENERAL CARE

Nonclinical safety assessment studies fall into three categories, as will be examined in detail in the remainder of this book:

1. IND enabling ("FIM"): The studies necessary to support the initiation of clinical trials in human beings. These are generally as specified in ICH M3, and this is most commonly performed of the three sets.
2. To support continued clinical development: As clinical development proceeds, longer repeat drug studies must be performed, reproductive and developmental toxicology studies must be done, and other ancillary studies are required.
3. To support filing for marketing approval: The final studies generally required to support marketing of drugs—such as carcinogenicity.

Which studies fall into each of these categories and exactly what studies must be done to support the development of a drug for a specific therapeutic claim are extremely variable. The general case—much as specified in M3—gives us a starting place for understanding what must be done.

But the general case really applies to the simplest oral drug intended for chronic use and more often than not does not apply. In fact, it may never fully apply.

## REFERENCES

Alder, S., and Zbinden, G. (1988). *National and International Drug Safety Guidelines*. M.T.C Verlag Zollikon, Zollikon, Switzerland.

Choerghade, M. S. (2006). *Drug Discovery and Development*, Vol. I: *Drug Discovery*. Wiley, Hoboken, NJ.

Czerepak, A. F., and Ryser, S. (2008). Drug approvals and failures: implications for alliances. *Nature Rev. Drug Discov.* 7:197–198.

Gad, S. C. (2007). *Handbook of Drug Discovery*. Wiley, Hoboken, NJ.

Goozner, M. (2004). *The $800 Million Pill: The Truth Behind the Cost of New Drugs*. University of California Press, Berkley.

Greider, K. (2003). *The Big Fix: How the Pharmaceutical Industry Rips Off American Consumers*. Public Affairs, New York.

Lednicer, D. (Ed.) (1993). *Chronicles of Drug Discovery*, Vol. 3. American Chemical Society, Washington, DC.

Mathieu, M. (2000). *New Drugs Development: A Regulatory Overview*. Darexel, International, Waltham, MA.

Rang, H. P. (2006). *Drug Discovery and Development*. Churchill Livingstone, New York.

Sjostrom, H., and Nilsson, R. (1972). *Thalidomide and the Power of the Drug Companies*. Penguin, Harmondsworth.

# *Regulation of Human Pharmaceutical Safety*

## 1 INTRODUCTION

The safety of pharmaceutical agents, medical devices, and food additives are the toxicology issues of the most obvious and longest-standing concern to the public. A common factor among the three is that any risk associated with a lack of safety of these agents is likely to affect a very broad part of the population, with those at risk having little or no option as to undertaking this risk. Modern drugs are essential for life in our modern society, yet there is a consistent high level of concern about their safety.

This chapter examines the regulations which establish how the safety of human pharmaceutical products are evaluated and established in the United States and the other major international markets. As a starting place, the history of this regulation will be reviewed and the current organizational structure of the U.S. Food and Drug Administration (FDA) will be briefly reviewed along with the other quasi-governmental bodies that also influence the regulatory processes. The current structure and context of the regulations in the United States and overseas will also be presented. From this point the general case of regulatory product development and approval will be presented. Toxicity assessment study designs will be presented. The broad special case of biotechnology-derived therapeutic products and environmental concerns associated with the production of pharmaceuticals will be briefly addressed. The significant changes in regulation brought about by harmonization are also reflected.

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

As an aid to the reader, appendices are provided at the end of this book: a codex of acronyms that are used in this field followed by a glossary which defines some key terms.

## 2 BRIEF HISTORY OF U.S. PHARMACEUTICAL LAW

A synopsis of the history of U.S. drug legislation is presented in Table 1. Here we will review the history of the three major legislative acts covering pharmaceuticals.

### 2.1 1906: Pure Food and Drug Act

As so eloquently discussed by Temin (1991), the history of health product legislation in the United States largely involves the passage of bills in Congress which were primarily in response to public demand. In 1902, for example, Congress passed the Biologics Act in response to a tragedy in St. Louis where 10 children died after being given contaminated diphtheria toxins. Interestingly, the background that led to the passage of the first Pure Food and Drug Act in 1906 had more to do with food processing than with drugs. The conversion from an agrarian to an urban society fostered the growth of a food-processing industry that was rife with poor practice. Tainted and adulterated food was commonly sold. Such practices were sensationalized by the muckraking press, including books such as *The Jungle* by Sinclair Lewis.

In the early debates in the U.S. Congress on the Pure Food and Drug Act (passed in 1906), there was little mention of toxicity testing. When Harvey Wiley, chief of the Bureau of Chemistry. Department of Agriculture, and driving force in the enactment of this early law, did his pioneering work (beginning in 1904) on the effects of various food preservatives on health, he did so using only human subjects and with no prior experiments with animals (Anderson, 1988). Ironically, work that led to the establishment of the FDA would probably not have been permitted under the current guidelines of the agency. Wiley's studies were not double blinded, so it is also doubtful that his conclusions would have been accepted by the present agency or the modern scientific community. Legislation in place in 1906 consisted strictly of a labeling law prohibiting the sale of processed food or drugs that were misbranded. No approval process was involved and enforcement relied on postmarketing criminal charges. Efficacy was not a consideration until 1911, when the Sherley amendment outlawed fraudulent therapeutic claims.

### 2.2 1938: Food, Drug and Cosmetic Act

The present regulations are largely shaped by the law passed in 1938. It will therefore be discussed in some detail. The story of the 1938 Food, Drug and Cosmetic Act (FDCA) actually started in 1933. Franklin D. Roosevelt had

**TABLE 1 Important Dates in U.S. Federal Drug Law**

| Year | Event |
|---|---|
| 1902 | Passage of the Virus Act, regulating therapeutic serums and antitoxins. Enforcement by the Hygienic Laboratory (later to become the National Institute of Health), Treasury Department. |
| 1906 | Passage of Pure Food Act, including provisions for the regulations of drugs to prevent the sale of misbranded and adulterated products. Enforcement by the Chemistry Laboratory, Agriculture. |
| 1912 | Passage of the Sherley Amendment. Specifically outlawed any false label claims as to curative effect. |
| 1927 | Bureau of Chemistry renamed the Food, Drug and Insecticide Administration. |
| 1931 | Renamed again to Food and Drug Administration. |
| 1938 | Passage of the Food, Drug and Cosmetic Act. Superseded the law of 1906. Required evidence of safety, e.g., studies in animals. Included coverage of cosmetics and medical devices. Specifically excluded biologics. |
| 1944 | Administrative Procedures Act, codifying public health laws: included provision that for a biological license to be granted, a product must meet standards for safety, purity, and potency. The National Institutes of Health (NIH) also given the responsibility for developing biologics not developed by the private sector. |
| 1945 | Amendment to the 1936 act requiring that the FDA examine and certify for release each batch of penicillin. Subsequently amended to include other antibiotics. |
| 1949 | Publication of the first set of criteria for animal safety studies. Following several revisions, guidelines published in 1959 as *Appraisals Handbook*. |
| 1951 | Passage of Durham–Humphrey Amendment. Provided the means for manufacturers to classify drugs as over the counter (not requiring prescription). |
| 1953 | Transfer of FDA to the Department of Health, Education and Welfare from Agriculture (now the Department of Health and Human Services). |
| 1962 | Passage of major amendments (the Kefauver Bill) to the 1938 FDCA, which required proof of safety and effectiveness (efficacy) before granting approval of NDAS. Required affirmative FDA approval. |
| 1968 | FDA placed under the Public Health Service of Health, Education and Welfare (HEW). |
| 1970 | Controlled Substances Act and Controlled Substances Import and Export Act. Removed regulation of drug abuse from FDA (transferred to the Drug Enforcement Agency) and provided for stringent regulation of pharmaceuticals with abuse potential. |
| 1972 | Transfer of authority to regulate biologics transferred from NIH to FDA. The NIH retained the responsibility of developing biologics. |
| 1973 | Consumer Product Safety Act, leading to the formation of separate Consumer Product Safety Commission, which assumes responsibilities once handled by the FDA's Bureau of Product Safety. |
| 1976 | Medical Device Amendment to the FDCA requiring for devices that not only effectiveness be proven but also safety. |
| 1979 | Passage of the Good Laboratory Practices Act. |
| 1983 | Passage of the first Orphan Drug Amendment to encourage development of drugs for small markets. |
| 1984 | Drug Price Competition and Patent Term Restoration Act intended to allow companies to recover some of the useful patent life of a novel drug lost due to the time it takes the FDA to review and approve. Also permits the marketing of generic copies of approved drugs. |

**TABLE 1** ***Continued***

| Year | Event |
|---|---|
| 1985 | The "NDA rewrite" final rule. An administrative action streamlining and clarifying the NDA process. Now embodied in 21 CFR 314. |
| 1986 | The United States Drug Export Amendment Act of 1986. Permitted the export of drugs outside the United States prior to approval for the U.S. market. |
| 1987 | The "IND rewrite" final rule: "to encourage innovation and drug development while continuing to assure the safety of (clinical) test subjects."*Federal Register* 52:8798, 1987. Now embodied in 21 CFR 312. |
| 1990 | Safe Medical Device Act, providing additional authority to the FDA for regulation of medical devices. |
| 1992 | Safe Medical Device Amendments requiring more extensive testing of devices. |
| 1992 | Prescription Drug User Fee Act. Established the payment of fees for the filing of applications (e.g., IND, NDA, PLA). |
| 1994 | Orphan Drug Amendment. |
| 1997 | The Food and Drug Administration Modernization Act: to streamline the drug and device review and approval process. |
| 2002, 2007 | Food and Drug Administration Modernization Act Amendments. |

*Note*: Laws and amendments that have covered other aspects of FDA law, such as those governing food additives [e.g., Food Quality Production Act (FQPA)], are not included in this table.

just won his first election and installed his first cabinet. Walter Campbell was the chief of the FDA, reporting to Rexford Tugwell, the undersecretary of agriculture. The country was in the depths of its greatest economic depression. This was before the therapeutic revolution wrought by antibiotics in the 1940s, and medicine and pharmacy as we know it in the 1990s were not practiced. Most medicines were, in fact, self-prescribed. Only a relatively small number of drugs were sold via physicians' prescription. The use of so-called patent (because the ingredients were kept secret) preparations was rife, as was fraudulent advertising. Today, for example, it is difficult to believe that in the early 1930s a preparation such as Radithor (nothing more than a solution of radium) was advertised for treatment of 160 diseases. It is in this environment that one day in the winter of 1933 Campbell delivered a memo to Tugwell on an action level of an insecticide (lead arsenite) used on fruits. Tugwell briskly asked why, if the chemical was so toxic, was it not banned outright. He was amazed to find out from Campbell that the agency had no power to do so.

The 1906 law was designed to control blatantly misbranded and/or adulterated foods and drugs that relied on post facto criminal charges for enforcement. Safety and efficacy were not an issue so long as the product was not misbranded with regard to content. Premarketing review of a drug was an unknown practice. Thus, attempts at rewriting the old 1906 law to include control of bogus therapeutic claims and dangerous preparations proved to be unsatisfactory. Paul Dunbar of the FDA suggested to Campbell that an entirely new law was needed. A committee of FDA professionals and outside academic consultants drafted a new bill which immediately ran into trouble because no one in Congress was willing to sponsor it. After peddling the bill

up and down the halls of Congress, Campbell and Tugwell convinced Senator Royal Copeland of New York to sponsor the bill. Unknowingly at the time, Copeland put himself in the eye of a hurricane that would last for five years.

The forces that swirled around Copeland and the Tugwell Bill (Senate bill S.1944) were many. First was the immediate and fierce opposition from the patent medicine lobby. Flyers decried S.1944 as everything from a communist plot to un-American, stating it "would deny the sacred right of self-medication." In opposition to the patent trade organizations were two separate but unlikely allies: a variety of consumer advocacy and women's groups (such as the American Association of University Women, whose unfaltering support for the bill eventually proved critical to passage) and the mainline professional organizations. Interestingly, many of these organizations at first opposed the bill because it was not stringent enough. There were also the mainline professional pharmacy and medical organizations [such as the American Medical Association (AMA) and the American Association of Colleges of Pharmacy] whose support for the bill ranged from neutral to tepid but did grow over the years from 1933 to 1938.

Second, there was the basic mistrust on the part of Congress toward Tugwell and other "New Dealers." At the same time, Roosevelt gave the measure only lukewarm support at best (tradition has it that if it had not been for the First Lady, Eleanor, he would have given it no support at all) because of his political differences with Royal Copeland.

Third, there was a considerable bureaucratic turf war over the control of pharmaceutical advertising. Finally, despite the efforts of the various lobbying groups, there was no popular interest or support for the bill. At the end of the congressional period, S.1944 died for lack of passage.

The next five years would see the introduction of new bills, amendments, competing measures, committee meetings and hearings, lobbying, and House/Senate conferences. The details of this parliamentary in-fighting make for fascinating history but are outside the scope of this book. The reader is referred to the excellent history of this period by Jackson (1970).

The FDA was surprised by the force and depth of the opposition to the bill. The proposed law contained a then-novel idea that a drug was misbranded if its labeling made any therapeutic claim which was contrary to general medical practice and opinion. The definition of a drug was broadened to include devices used for medical purposes.[1] *Adulteration* was defined as any drug product dangerous to health when used according to label directions. The patent manufacturers charged that no bill granted too much discretionary power to a federal agency—that no manufacturer could stay in business except by the grace of the Department of Agriculture, a charge that may have been

[1]The use of a broad definition of what constitutes a drug for regulatory purposes is a precedent that remains in place today. For example, the computer software used in diagnostic systems is considered to be a pharmaceutical for purposes of regulation.

correct. In response to the patent trade lobbying effort, the FDA launched its own educational drive of radio spots, displays (such as the sensationalized chamber-of-horrors exhibition, in which the toxicity of a variety of useless medicines was clearly displayed), mimeographed circulars, speaking engagements, posters, and so on.

Ruth Lamb, FDA information officer at the time, was perhaps one of the hardest working and most quotable of the FDA staffers working the street at the time. For example, in reference to one of the counterbills that had language similar to the original Copeland Bill, but with extremely complicated enforcement provisions, Ruth Lamb called it "an opus for the relief of indigent and unemployed lawyers" (Jackson, 1970). She once described the Bailey amendment, which would have made proprietary drugs virtually immune to multiple seizures, as permitting the "sale of colored tap water as a cure for cancer … unless arsenic was added to each dose making [it] immediately dangerous." After 1934, however, the educational efforts of the FDA were greatly attenuated by federal laws prohibiting lobbying by federal agencies.

The fall of 1937 witnessed the beginning of the often-told elixir-of-sulfanilamide incident, which remains one of the nation's worst drug tragedies. The Massengil Company was not one of the industry giants, but neither was it a "snake oil peddler." The company's chief chemist, Harold Watkins, was simply trying to develop a product and, in fact, did so in a manner consistent with the norms of the time. There was a perceived need for a liquid form of sulfanilamide, but it was difficult to dissolve. Watkins hit upon diethylene glycol. No toxicity tests were performed on the finished product, although the product did pass through the "control lab" where it was checked for appearance, fragrance, and consistency.

The first reports of human toxicity occurred in October 1937 when Dr. James Stevenson of Tulsa requested some information from the AMA because of six deaths in his area that were attributable to the elixir. At the time, no product of Massengil stood accepted by the Council on Pharmacy and Chemistry, and the council recognized no solution of sulfanilamide. The AMA telegraphed Massengil, requesting samples of the preparation for testing. Massengil complied. The test revealed the diethylene glycol to be the toxic agent and the AMA issued a general warning to the public on October 18, 1937. In the meantime, the FDA had become aware of the deaths and launched an investigation through its Kansas City station. By October 20, when at least 14 people had died, Massengil wired the AMA to request an antidote for their own product. By the end of October, at least 73 people had died and another 20 suspicious deaths were linked to the drug. Had it not been for the response of the FDA, more deaths may have occurred. The agency put its full force of field investigators (239 members) on the problem and eventually recovered and accounted for 99.2% of the elixir produced. Massengil fully cooperated with the investigation and in November published a public letter expressing regret over the matter but further stating that no law had been broken. In fact, the company was eventually convicted on a

long list of misbranding charges and fined a total of $26,000 (the largest fine ever levied under the 1906 law).

The Massengil incident made the limits of the 1906 law quite clear. Because there were no provisions against dangerous drugs, the FDA could move only on the technicality of misbranding. The term *elixir* was defined by the U.S. Pharmacopeia (USP) as "a preparation containing alcohol," which elixir of sulfanilamide was not. It was only this technicality that permitted the FDA to declare the "elixir" misbranded, to seize the inventory, and to stop the sale of this preparation. If it had been called *solution of sulfanilamide*, no charges could have been brought.

The extensive press coverage of the disaster became part of the national dialogue. Letters poured in to congressmen demanding action to prevent another such tragedy. Medical and pharmacy groups and journals insisted that a new law was required. Congress was in special session in November 1937 and did not need to be told about the tragedy. Copeland and Representative Chapman (of Kentucky) pressed resolutions calling for a report from the FDA on the tragedy. When issued, the FDA report stunned Congress, not only because of the human disaster but also because it made apparent that, even had the bill then before Congress been law, the entire tragedy would still have occurred because there were no provisions for toxicity testing before new drugs entered the market. By December 1937 a new bill, S.3037, was introduced which stated that manufacturers seeking to place new drugs on the market would be required to supply records of testing, lists of components, descriptions of each manufacturing process, and sample labels. Drugs would require certification by the FDA before sale was permitted. A similar bill was introduced in the House by Chapman, although the issues of which agency was to control advertising of drugs was still festering in the House. In January 1938, debate started on the Wheeler–Lea Bill, which would ensure that all controls over drug advertising would remain with the Federal Trade Commission (FTC). Despite strong opposition by the FDA, the Wheeler–Lea Bill was signed into law March 1938. While the loss of advertising control was a blow to the FDA, the Wheeler–Lea Bill did facilitate the passage of the new Food and Drug Law.

With the issue of advertising controls settled, the Copeland–Chapman Bill faced one last hurdle. Section 701, which had been added in committee, provided for appeal suits that could be entered in any federal district court to enjoin the agency from enforcing new regulations promulgated as a result of the act. Interestingly, this issue had more to do with foods than with drugs, as its major focus was on acceptable tolerance limits for insecticides in food. The new bill defined an *adulterated food* as one containing any poison. However, because efforts to remove insecticides from fresh fruits and vegetables had never been completely successful, the secretary of agriculture needed this power to set tolerance levels. Allies of food producers tried to introduce provisions in the new bill that provided methods for stalling a tolerance regulation with rounds of appeals. The bill passed the House despite such provisions

(Section 701) and the resistance of consumer groups and the FDA and went into joint committee. Roosevelt, in one of his rare efforts to support the FDA, made it clear that he would not accept the bill with such a cumbersome appeals process. The resulting compromise was an appeals process which limited the new evidence that could be introduced into one of the 10 circuit courts. Other provisions regarding labeling were also rectified in joint committee. In May 1938, S.3073 passed by unanimous vote. Both chambers ratified the joint committee report, and Roosevelt signed the new law in June of 1938.

A historical note to this story was that Royal Copeland did not live to see his measure passed. In May 1938, he collapsed on the Senate floor. His death occurred one month before President Roosevelt signed his bill into law.

### 2.3 1962: Major Amendment

The 1938 law very much changed the manner in which Americans purchased pharmaceutical agents. In effect, it changed the pharmaceutical industry from a traditional consumer product industry to one in which purchases were made as directed by a third party (the physician). In 1929, ethical pharmaceuticals (prescription drugs) comprised only 32% of all medicines, while by 1969 this was up to 83% (Termini, 1980). This led to a peculiar lack of competition in the ethical market. In 1959, Senator Estes Kefauver initiated his now-famous hearings on the drug industry. Interestingly, almost 30 years later, Senator Edward Kennedy had hearings on exactly the same matter. In 1961, Kefauver submitted a proposed legislation to amend the 1938 act in such a way as to increase FDA oversight of the drug industry. The proposed amendment contained two novel propositions. The first was compulsory licensing, which would have required, for example, company A to license (with a royalty of no greater than 8% of sales) company B to market a drug patented by company A. Company A would have only three years' exclusivity with its patent. The second novel provision was that new drugs had to be not only "safe" but also "efficacious." There was not a groundswell of support for this legislation. When it was reported out of committee, it had been rewritten (including the removal of the licensing requirement) to the point that even Kefauver refused to support it. The Kennedy administration wanted new legislation but did not specifically support the Kefauver Bill; rather it introduced its own legislation, sponsored by Representative Orren Harris of Arkansas. It also had little support.

As in 1938, a tragic incident would intercede in the legislative process: 1961 would see the development of the thalidomide tragedy. An antianxiety agent marketed in Europe, thalidomide was prescribed for pregnancy-related depression and taken by countless numbers of women. At about the same time, phocomelia, a birth defect marked by the imperfect development of arms and legs, appeared in Europe. Thalidomide was eventually determined to be the causative teratogen in 1961 and subsequently taken off the market in

Europe. The William S. Merrill Company had applied for a new drug application (NDA) for thalidomide in the United States in 1960. It was never approved because the FDA examiner, Dr. Frances Kelsey, had returned the application for lack of sufficient information. Eventually, the company withdrew the application. Senator Kefauver's staff had uncovered the thalidomide story as it was unfolding and had turned its findings over to the *Washington Post*. The *Post* reported the episode under the headline "Heroine of the FDA Keeps Bad Drug Off the Market" in July 1962, three days after the Kefauver Bill was reported out of committee. Needless to say, the news created public support for the bill, which was sent back to committee and reported out again with new language in August 1962. The Kefauver–Harris Bill was signed into law in October 1962. It was demonstrated after the fact that thalidomide was teratogenic in the rabbit; out of the episode grew the current practice that new human pharmaceuticals are tested for teratogenicity in two species, one generally being the rabbit.

The 1962 Drug Amendment made three major changes in the manner in which new drugs could be approved (Merrill, 1994). First, and perhaps the most important, was that it introduced the concept of effectiveness into the approval process. An NDA had to contain evidence that the drug was not only safe but also effective. The 1938 law contained no such specification. The effectiveness requirement necessitated that a drug company had to do more extensive clinical trials. The new law required that companies apply to the FDA for approval of its clinical testing plan under an investigational new drug application (INDA). No response from the FDA was deemed to be acceptance. As each level of clinical testing came to require FDA review and approval, the new law made the FDA an active partner in the development of all drugs.

The second major change enacted under the 1962 amendment was the change in the approval process from premarket notification to a premarket approval system. Under the terms of the 1938 law, an NDA would take effect automatically if the FDA did not respond. For example, the only reason thalidomide was not approved was because Dr. Kelsey returned the application to the sponsor with a request for more information. In contrast, the 1962 law required affirmative FDA action before a drug could be put on the market. Under the terms of the 1962 amendments, the FDA was also empowered to withdraw NDA approval and remove the drug from the market for a variety of reasons, including new evidence that the product was unsafe or that the sponsor had misrepresented or underreported data.

The third major change enlarged the FDA's authority over clinical testing of new drugs. Thus, not only was evidence of effectiveness required, but Section 505(d) of the act specified the types of studies required: "Substantial evidence consisting of adequate and well-controlled investigations, including clinical investigations by qualified expert." In meeting the statutory requirement for setting standards of clinical evidence, the FDA has become highly influential in the design of drug-testing regimens (Merrill, 1994). Interestingly,

discussed in detail by Hutt (1987), the FDA was initially quite unprepared for this new level of responsibility. It was not until 1973 that audited regulations on the determination of safety and effectiveness were put into place (these were, in fact, approved by the Supreme Court). While there have been several procedural changes [e.g., the 1985 investigational new drug (IND) rewrite] and additions (e.g., the 1988 IND procedures for life-threatening disease treatment), there have actually been no major changes in the law through 1992 with the Prescription Drug User Fee Act (PDUFA) and 1997 with the Food and Drug Administration Modernization Act (FDAMA).

We must interject an interesting historical sidelight at this point. Despite its reputation, thalidomide made a bit of a comeback in the 1990s (Blakeslee, 1994). Among other properties, thalidomide has been shown to have good anti-inflammatory properties due to the fact that it apparently decreases the synthesis and/or release of tissue necrosis factor.

## 2.4 1992, 1997, 2002, and 2007: PDUFA and FDAMA

The history of pharmaceutical regulations has been dominated by two often-opposing schools of thought: the need to provide the citizenry with effective medicaments and the need to protect the consumer from unsafe and misbranded products. The reader is referred to Peter B. Hutt's in-depth reviews (1983a,b) on the subject. For example, the first federal drug legislation in the United States was the Vaccine Act of 1813, which mandated the provision of the smallpox vaccine to the general public. In the modern era, legislative debate could be further defined as the constant swing back and forth on these two issues (see Hutt, 1983a,b), that is, safety versus development costs. In 1963, for example, Senator Hubert Humphrey presided over hearings on the FDA's implementation of the Drug Amendment of 1962. The FDA came under substantial criticism for failure to take strong action to protect the public from dangerous drugs. Eleven years later (1974), Senator Edward Kennedy conducted hearings addressing exactly the same issue. Commissioner Schmidt pressed the point that the FDA is under constant scrutiny regarding the approval of "dangerous" drugs, but no hearing had ever been conducted (up to that time) on the failure of the FDA to approve an important new therapy.

The next decade and a half saw a proliferation of work that analyzed the impact of regulation on competitiveness and the introduction of new therapies [see Hutt (1983b) for a complete review]. This included, for example, Grabowski and Vernon's work (1983), which concluded that regulation had significant adverse effect on pharmaceutical innovation. This examination of the cost of regulation continued into the 1990s. In a meticulous and well-research study DiMasi et al. (1994) reported that throughout the 1980s the number of INDAS were decreasing while the NDA success rate was also dropping, and the length of time between discovery and approval was

increasing. Clearly this is a situation that could not go on forever. The cost of new drug development rose from $54 million (U.S.) in 1976 to $359 million (U.S.) in 1990 (Anonymous, 1998a). Members of the pharmaceutical industry and the biotechnology industry were becoming increasingly alarmed by the negative synergy caused by increased costs and increased time to market. In 1991, Dranove published an editorial examining the increased costs and decreased product flow that resulted from the 1962 amendment. He made the observation that European requirements are less stringent than those of the United States yet the Europeans did not seem to be afflicted by a greater number of dangerous drugs (see Table 1.2). Yet, if one looks at an analysis of worldwide withdrawals for safety from 1960 to 1999 (Fung et al., 2001), one sees that of 121 products identified 42.1% were withdrawn from European markets alone, 5% from North America, 3.3% from Asia Pacific, and 49.6% from multiple markets. The top five safety reasons for withdrawal were hepatic (26.2%), hematological (10.5%), cardiovascular (8.7%), dermatological (6.3%), and carcinogenic (6.3%) issue.

In an age of decreasing regulatory recourses, the FDA (as well as Congress) was under increasing pressure to review and release drugs more quickly. In response, Congress passed the 1992 PDUFA. Under the terms of this act, companies would pay a fee to the agency to defray costs associated with application review. They would supposedly provide the FDA with the resources available to decrease application review time. In return, companies were guaranteed a more rapid review time. By all accounts, PDUFA has been successful. In 1992 (the year PDUFA was passed), 26 NDAs were approved, requiring on average 29.9 months for data review, while in 1996, 53 new drug (or biological) products were approved, each requiring an average of 17.8 months of review time. PDUFA was successful in decreasing review times but has not really streamlined the procedures.

The AIDS activist community was particularly vocal and effective in demanding more rapid approvals and increased access to therapies. There was also demand for FDA reform on a number of other fronts (e.g., medical devices, pediatric claims, women and minority considerations, manufacturing changes, etc.). In 1993 the House Commerce Committee on Oversight and Investigations, chaired by John Dingel (D-MI), released a comprehensive investigation and evaluation of the FDA entitled *Less Than the Sum of Its Parts*. The report was highly critical of the FDA and made a number of recommendations (Pilot and Waldemann, 1998). The mid-1990s also saw the reinventing government initiatives (RIGO) chaired by Vice President AL Gore. Under RIGO, the FDA sought to identify and implement administrative reform. The RIGO report issued was entitled *Reinventing Regulation of Drugs and Medical Devices.* The 104th Congress started hearings on FDA reform again in the winter of 1995. Two bills were introduced that provided the essential outline of what would become FDAMA. Senator Nancy Kassebaum (R-KS), chair of the Senate Committee on Labor and Human Resources, introduced S-1477. The second was H.R. 3201, introduced by Rep. Joe Barton

(R-TX). Other bills were introduced by Senator Paul Wellstone (D-MN) and Rep. Ron Weyden (D-OR), which focused more on medical devices but still paved the way for bipartisan support of FDA reform (Pilot and Wladerman, 1998@@). Eventually, the 105th Congress passed FDAMA, which was signed into law by President Clinton in November 1997. The various sections of FDAMA are listed in Table 2. By any measure, it was a very broad and complex, if not overdeep, piece of legislation. In 1998, Marwick (1998, p. 815) observed, "a measure of the extent of the task is that implementation of the act will require 42 new regulations, … 23 new guidance notices, and 45 reports and other tasks." The FDA has identified these various tasks, regulations, and guidances necessary for the implementation of FDAMA. (The FDAMA implementation chart is available at http://www.fda.gov/po/modact97.html, and the reader is urged to explore this site. There is an FDAMA icon on the FDA home page, and both the Center for Biologics Evaluation and Research (CBER) and the Center for Drug Evaluation and Research (CDER) have issued various guidance documents. Some of the more interesting sections of the act that may be of interest to toxicologists included:

- Two successive renewals of PDUFA for another five years.
- Fast track for break-through products.
- Change in the fashion biologicals are regulated [elimination of the establishment and product licenses, both replaced with a biological license application or (BLA)].
- Change in the fashion antibiotics are developed and regulated.
- Incentives for the development of pediatric claims.
- Companies will be permitted to disseminate information about approved uses for their products.
- FDAMA requires that the FDA establishes a clinical trials database for drugs used to treat serious and life-threatening diseases, other than AIDS and cancers (databases for these diseases have already been established).

The full impact of FDAMA in the pharmaceutical industry in general and on toxicology within this industry in particular remains to be determined.

This is a debate that has continued to the present and has been highlighted by the demands for anti-HIV chromotherapeutic agents.

While it is not possible to review the history of regulations worldwide, it is possible to point out some differences. We will highlight specific differences where appropriate throughout the remainder of the text.

The strength of the United States regulatory system was highlighted at the BioEurope 1993 Conference. David Holtzman stated: "the main subject of the conference was regulation, and the U.S. was perceived to have the superior regulatory agency. It may be more difficult to satisfy but it is more predictable and scientifically based" (Holtzman, 1993, p. 87). This predictability has not

**TABLE 2 Summary of Contents of 1997 Food and Drug Administration Modernization Act**

| Title/Subtitle | Section |
|---|---|
| I. Improving Regulatory Drugs | |
| A. Fees Relating to Drugs | 101. Findings |
| | 102. Definitions |
| | 103. Authority to assess and use drug fees |
| | 104. Annual reports |
| | 105. Savings |
| | 106. Effective date |
| | 107. Termination of effectiveness |
| B. Other Improvements | 111. Pediatric studies of drugs |
| | 112. Expanding study and approval of fast-track drugs |
| | 113. Information program on trials for serious disease |
| | 114. Health care economic information |
| | 115. Manufacturing changes for drugs |
| | 116. Streamlining clinical research for drugs |
| | 118. Data requirements for drugs and biologics |
| | 119. Content and review of applications |
| | 120. Scientific advisory panels |
| | 121. Positron emission tomography |
| | 122. Requirements for radiopharmaceuticals |
| | 123. Modernization of regulation |
| | 124. Pilot- and small-scale manufacture |
| | 125. Insulin and antibiotics |
| | 126. Elimination of certain labeling requirements |
| | 127. Application of federal law to pharmacy compounding |
| | 128. Reauthorization of clinical pharmacology program |
| | 129. Regulation of sunscreen products |
| | 130. Report of postmarketing approval studies |
| | 131. Notification of discontinuance of a life-saving product |
| II. Improving Regulation of Devices | 201. Investigational device exemptions |
| | 202. Special review for certain devices |
| | 203. Expanding humanitarian use of devices |
| | 204. Device standards |
| | 205. Collaborative determinations of device data requirements |
| | 206. Premarket notification |
| | 207. Evaluation of automatic class III designation |
| | 208. Classification panels |
| | 209. Certainty of review time frames |
| | 210. Accreditation of person for review of premarket notification reports |

**TABLE 2** ***Continued***

| Title/Subtitle | Section |
|---|---|
| | 211. Device tracking |
| | 212. Postmarket notification |
| | 213. Reports |
| | 214. Practice of medicine |
| | 215. Noninvasive blood glucose meter |
| | 216. Data relating to premarket approval: product development protocol |
| | 217. Number of required clinical investigations for approval |
| III. Improving Regulation of Food | 301. Flexibility for regarding claims |
| | 302. Petitions for claims |
| | 303. Health claims for food products |
| | 304. Nutrient content claims |
| | 305. Referral statements |
| | 306. Disclosure of radiation |
| | 307. Irradiation petition |
| | 308. Glass and ceramic ware |
| | 309. Food contact substance |
| IV. General Provisions | 401. Dissemination of information new uses |
| | 402. Expanded access of investigational therapies and diagnostics |
| | 403. Approval of supplemental applications for approved products |
| | 404. Dispute resolution |
| | 405. Informal agency statements |
| | 406. FDA mission and annual report |
| | 407. Information system |
| | 408. Education and training |
| | 409. Centers for education and research on therapeutics |
| | 410. Mutual recognition of agreements and global harmonization |
| | 411. Environmental impact review |
| | 412. National uniformity for nonprescription drugs and cosmetics |
| | 413. FDA study of mercury in drugs and foods |
| | 414. Interagency collaboration |
| | 415. Contracts for expert review |
| | 416. Product classification |
| | 417. Registration of foreign establishments |
| | 418. Clarification of seizure authority |
| | 419. Interstate commerce |
| | 420. Safety report disclaimers |
| | 421. Labeling and advertising compliance with statutory requirements |
| | 422. Rule of construction |
| V. Effective Date | 501. Effective date |

stultified the growth and biotechnology industry in the United States. In fact, it has made the United States a more inciting target for investment than Europe. It is also a system that, while not perfect, has permitted very few unsafe products on the market.

## 2.5 FDAMA: Summary—Consequences and Other Regulations

In summary, federal regulation of the safety of drugs has had three major objectives:

- Requiring testing to establish safety and efficacy
- Establishing guidelines as to which tests are required and how they are designed
- Promulgating requirements of data recording and reporting

The first of these objectives was served by the 1906 act, which required that agents be labeled appropriately. This was amended in 1938, in response to the tragedies associated with elixir of sulfanilamide and Lash Lure, to require that drugs and marketed formulations of drugs be shown to be safe when used as intended. In the aftermath of the thalidomide tragedy, the 1962 Kefauver–Harris Amendment significantly tightened requirements for preclinical testing (the INDA) and premarket approval (the NDA) of new drugs. The regulations pertaining to INDAs and NDAs have been modified (most recently in 1988) but essentially remain as the backbone of regulations of the toxicity evaluation of new human pharmaceutical agents.

The Good Laboratories Practice (GLP) Act, which specifies standards for study planning, personnel training, data recording, and reporting, came out in 1978 in response to perceived shoddy practices of the operations of laboratories involved in the conduct of preclinical safety studies. It was revised in 1985 and is discussed elsewhere in this book.

The final major regulatory initiative on how drugs will be preclinically evaluated for safety arose out of the AIDS crisis. To that point, the process of drug review and approval had very generally been perceived as slowing down, the FDA pursuing a conservative approach to requiring proof of safety and efficacy before allowing new drugs to become generally available. In response to AIDS, in 1988 the Expedited Delivery of Drugs for Life-Threatening Diseases Act established a basis for less rigorous standards (and more rapid drug development) in some limited cases.

In the United Kingdom, the Committee on Safety of Medicines (reporting to the minister of health) regulates drug safety and development under the Medicines Act of 1968 (which has replaced the Therapeutic Substances Act of 1925). Details on differences in drug safety regulations in the international marketplace can be found in Alder and Zbinden (1988), but key points are presented in this chapter.

## 3 OVERVIEW OF U.S. REGULATIONS

### 3.1 Regulations: General Considerations

The U.S. federal regulations that govern the testing, manufacture, and sale of pharmaceutical agents and medical devices are covered in Chapter 1, Title 21, of the Code of Federal Regulations (21 CFR). These comprise nine 6 × 8-in. (printing on both sides of the pages) volumes which stack 8 in. high. This title also covers foods, veterinary products, and cosmetics. As these topics will be discussed elsewhere in this book, here we will briefly review those parts of 21 CFR that are applicable to human health products and medicinal devices.

Of most interest to a toxicologist working in this arena would be Chapter 1, Subchapter A (Parts 1–78), which cover general provisions, organization, and so on. The GLPs are codified in 21 CFR 58.

General regulations that apply to drugs are in Subchapter C (Parts 200–299). This covers topics such as labeling, advertising, commercial registration, manufacture, and distribution. Of most interest to a toxicologist would be a section on labeling (Part 201, Subparts A–G, which covers Sections 201.1–201.317 of the regulations) as much of the toxicological research on a human prescription drug goes toward supporting a label claim. For example, specific requirements on content and format of labeling for human prescription drugs are covered in Section 201.57. Directions for what should be included under the "Precautions" section of a label are listed in 201.57(f). This included 201.57(f)(6), which covers categorization of pregnancy risk, and the reliance upon animal reproduction studies in making these categorizations is made quite clear. For example, a drug is given a pregnancy category B if "animal reproduction studies have failed to demonstrate a risk to the fetus." The point here is not to give the impression that the law is most concerned with pregnancy risk. Rather, we wish to emphasize that much basic toxicological information must be summarized on the drug label (or package insert). This section of the law is quite detailed as to what information is to be presented and the format of the presentation. Toxicologists working in the pharmaceutical arena should be familiar with this section of the CFR.

### 3.2 Regulations: Human Pharmaceuticals

The regulations specifically applicable to human drugs are covered in Subchapter D, Parts 300–399. The definition of a new drug is covered in Part 310(g):

> A new drug substance means any substance that when used in the manufacture, processing or packaging of a drug causes that drug to be a new drug but does not include intermediates used in the synthesis of such substances.

The regulation then goes on to discuss "newness with regard to new formulations, indications, or in combinations." For toxicologists, the meat of the

regulations can be found in Section 312 (INDA) and Section 314 (applications for approval to market a new drug or antibiotic drug or NDA). The major focus for a toxicologist working in the pharmaceutical industry is on preparing the correct toxicology "packages" to be included to "support" these two types of applications. (The exact nature of these packages will be covered below.)

In a nutshell, the law requires solid scientific evidence of safety and efficacy before a new drug will be permitted in clinical trials or (later) on the market. The INDA (covered in 21 CFR 310) is for permission to proceed with clinical trials on human subjects. Once clinical trials have been completed, the manufacturer or "sponsor" can then proceed to file an NDA (covered in 21 CFR 314) for permission to market the new drug.

As stated in 321.21, "A sponsor shall submit an IND if the sponsor intends to conduct a clinical investigation with a new drug … [and] shall not begin a clinical investigation until … an IND … is in effect." Similar procedures are in place in other major countries. In the United Kingdom, for example, a clinical trials certificate (CTC) must be filed or a CTX (clinical trial exemption) obtained before clinical trials may proceed. Clinical trials are divided into three phases, as described in 312.21. Phase I trials are initial introductions into healthy volunteers primarily for the purposes of establishing tolerance (side effects), bioavailability, and metabolism. Phase II clinical trials are "controlled studies … to evaluate effectiveness of the drug for a particular indication or disease." The secondary objective is to determine common short-term side effects; hence the subjects are closely monitored. Phase III studies are expanded clinical trials. It is during this phase that the definitive, large-scale, double-blind studies are performed.

The toxicologist's main responsibilities in the IND process are to design, conduct, and interpret appropriate toxicology studies (or "packages") to support the initial IND and then design the appropriate studies necessary to support each additional phase of investigation. Exactly what may constitute appropriate studies are covered elsewhere in this chapter. The toxicologist's second responsibility is to prepare the toxicology summaries for the (clinical) investigator's brochure [described in 312.23(a)(8)(ii)]. This is an integrated summary of the toxicological effects of the drug in animals and in vitro. The FDA has prepared numerous guidance documents covering the content and format of INDs. It is of interest that in the guidance for industry (Lumpkin, 1995) an in-depth description of the expected contents of the pharmacology and toxicology sections was presented. The document contains the following self-explanatory passage:

> Therefore, if final, fully quality-assured individual study reports are not available at the time of IND submission, an integrated summary report of toxicological findings based on the unaudited draft toxicologic reports of the completed animal studies may be submitted.

If unfinalized reports are used in an initial IND, the finalized report must be submitted within 120 days of the start of the clinical trial. The sponsor must

also prepare a document identifying any differences between the preliminary and final reports and the impact (if any) on interpretation.

Thus, while the submission of fully audited reports is preferable, the agency does allow for the use of incomplete reports.

Once an IND or CTC/X is opened, the toxicologists may have several additional responsibilities. First is to design, conduct, and report the additional tests necessary to support a new clinical protocol or an amendment to the current clinical protocol (Section 312.20). Second is to bring to the sponsor's attention any finding in an ongoing toxicology study in animals "suggesting a significant risk to human subjects, including any finding of mutagenicity, teratogenicity or carcinogenicity," as described in 21 CFR 312.32. The sponsor has a legal obligation to report such findings within 10 working days. Third is to prepare a "list of the preclinical studies … completed or in progress during the past year" and a summary of the major preclinical findings. The sponsor is required (under Section 312.23) to file an annual report (within 60 days of the IND anniversary date) describing the progress of the investigation. INDs are never "approved" in the strict sense of the word. Once filed, an IND can be opened 30 days after submission unless the FDA informs the sponsor otherwise. The structure of an IND is outlined in Table 3. Complete and thorough reports on all pivotal toxicological studies must be provided with the application.

If the clinical trials conducted under an IND are successful in demonstrating safety and effectiveness [often established at a pre-NDA meeting, described in 21 CFR 312.47(b)(2)], the sponsor can then submit an NDA. Unlike an IND, the NDA must be specifically approved by the agency. The toxicologist's responsibility in the NDA/marketing authorization application (MAA) process is to prepare an integrated summary of all the toxicology and/or safety studies performed and be in a position to present and review the toxicology findings to the FDA or its advisory bodies. The approval process can be exhausting, including many meetings, hearings, appeals, and so on. The ground

**TABLE 3 Composition of Standard INDA**

| | |
|---|---|
| 1. | IND cover sheets (form FDA-1571) |
| 2. | Table of contents |
| 3. | Introductory statement |
| 4. | General (clinical) investigation plan |
| 5. | (Clinical) investigators brochure |
| 6. | (Proposed) clinical protocol(s) |
| 7. | Chemistry, manufacturing, and control information |
| 8. | Pharmacology and toxicology information (includes metabolism and pharmacokinetic assessments done in animals) |
| 9. | Previous human experience with the investigational drug |
| 10. | Additional information |
| 11. | Other relevant information |

rules for all of these are described in Part A of the law. For example, all NDAs are reviewed by an "independent" (persons not connected with either the sponsor or the agency) scientific advisory panel which will review the findings and make recommendations as to approval. MAAs must be reviewed by and reported on by an expert recognized by the cognizant regulatory authority. Final statutory approval in the United States lies with the commissioner of the FDA. It is hoped that few additional studies will be requested during the NDA review and approval process. When an NDA is approved, the agency will send the sponsor an approval letter and will issue a summary basis of approval (SBA) (312.30), which is designed and intended to provide a public record of the agency's reasoning for approving the NDA while not revealing any proprietary information. The SBA can be obtained through Freedom of Information and can provide insights into the precedents for which types of toxicology studies are used to support specific types of claims.

### 3.3 Regulations: Environmental Impact

Environmental impact statements, while once important only for animal drugs, must now accompany all MDAs. This assessment must also be included in the drug master file (DMF). The procedures, formats, and requirements are described in 21 CFR 2531. This requirement has grown in response to the National Environmental Policy Act, the heart of which required that federal agencies evaluate every major action that could effect the quality of the environment. In the INDs, this statement can be a relatively short section claiming that relatively small amounts will post little risk to the environment. The European Economic Community (EEC) has similar requirements for drug entities in Europe, though data requirements are more strenuous. With NDAs, this statement must be more substantial, detailing any manufacturing and/or distribution process that may result in release into the environment. Environmental fate (e.g., photohydrolysis) and toxicity (e.g., fish, daphnia, and algae) studies will be required. While not mammalian toxicology in the tradition of pharmaceutical testing, preparing an environmental impact statement will clearly require toxicological input. The FDA has published a technical bulletin covering the tests it may require (FDA, 1987).

### 3.4 Regulations: Antibiotics

The NDA law (safety and effectiveness) applies to all drugs, but antibiotic drugs were treated differently until the passage of FDAMA in 1997. Antibiotic drugs had been treated differently by the FDA since the development of penicillin revolutionized medicine during World War II. The laws applicable to antibiotic drugs were covered in 21 CFR 430 and 431. Antibiotics such as penicillin or doxorubicin are drugs derived (in whole or in part) from natural sources (such as molds or plants) which have cytotoxic or cytostatic properties. They were treated differently from other drugs as the applicable laws required

a batch-to-batch certification process. Originally passed into law in 1945 specifically for penicillin, this certification process was expanded by the 1962 amendment (under Section 507 of the FDCA) to require certification of all antibiotic drugs, meaning that the FDA would assay each lot of antibiotic for purity, potency, and safety. The actual regulations were covered in 21 CFR Subchapter D, Parts 430–460 (over 600 pages), which describes the standards and methods used for certification for all approved antibiotics. Section 507 was repealed by FDAMA (Section 125). As a result of the repeal of Sections 507, the FDA is no longer required to publish antibiotic monographs. In addition, the testing, filing and reviewing of antibiotic applications are now handled under Section 505 of the act like any other new therapeutic agent. The FDA has published a guidance document to which the reader is referred for more details (Anonymous, 1998a,b).

### 3.5 Regulations: Biologics

Biological products are covered in Subchapter F, Parts 600–680. As described in 21 CFR 600.3(h), "biological product means any virus, therapeutic serum, toxin, antitoxin or analogous product applicable to the prevention, treatment or cure of diseases or injuries of man." In other words, these are vaccines and other protein products derived from animal sources. Clearly the toxicological concerns with such products are vastly different than those involved with low-molecular-weight synthetic molecules. There is little rational basis, for example, for conducting a one-year, repeated dose toxicity study with a vaccine or a human blood product. The FDA definition for safety with regard to these products is found in 21 CFR 603.1(p): "Relative freedom from harmful effect to persons affected, directly or indirectly, by a product when prudently administered." Such safety consideration has more to do with purity, sterility, and adherence to good manufacturing standards than with the toxicity of the therapeutic molecule itself. The testing required to show safety is stated in the licensing procedures 21 CFR 601.25(d)(1): "Proof of safety shall consist of adequate test methods reasonably applicable to show the biological product is safe under the prescribed conditions." Once a license is granted, each batch or lot of biological product must be tested for safety, and the methods of doing so are written into the law. A general test for safety (i.e., required in addition to other safety tests) is prescribed using guinea pigs, as described in 610.11. Additional tests are often applied to specific products. For example, 21 CFR 630.35 describes the safety tests required for measles vaccines, which include tests in mice and in vitro assays with tissue culture. Many new therapeutic entities produced by biotechnology are seeking approval as biologics with the results being FDA approval of a product license application (PLA). Table 4 presents general guidance for the basis of deciding if an individual entity falls under CDER or the CBER authority for review.

The International Conferences on Harmonisation (ICH) has published its document S6, *Preclincial Safety Evaluation of Biotechnology-Derived*

**TABLE 4 Product Class Review Responsibilities**

| |
|---|
| *Center for Drug Evaluation and Review* |
| Natural products purified from plant or mineral sources |
| Products produced from solid tissue sources (excluding procoagulants, venoms, blood products, etc.) |
| Antibiotics, regardless of method of manufacture |
| Certain substances produced by fermentation |
| Disaccharidase inhibitors |
| HMG–CoA inhibitors |
| Synthetic chemicals |
| Traditional chemical synthesis |
| Synthesized mononuclear or polynuclear products including antisense chemicals |
| Hormone products |
| *Center for Biologics Evaluation and Review* |
| Vaccines, regardless of manufacturing method |
| In vivo diagnostic allergenic products |
| Human blood products |
| Protein, peptide, and/or carbohydrate products produced by cell culture (other than antibiotics and hormones) |
| Immunoglobulin products |
| Products containing intact cells or microorganisms |
| Proteins secreted into fluids by transgenic animals |
| Animal venoms |
| Synthetic allergens |
| Blood banking and infusion adjuncts |

*Pharmaceuticals*. The FDA (the CDER and CBER jointly) has published the document as a guidance for industry (Anonymous, 1997a,b).

A current list of regulatory documents available by email (including the most recent PTCs, or points to consider) can be found at doc_list@a1.fda.gov.

### 3.6 Regulations versus Law

A note of caution must be inserted here. The law (the document passed by Congress) and the regulations (the documents written by the regulatory authorities to enforce the laws) are separate documents. The sections in the law do not necessarily have numerical correspondence. For example, the regulations on the NDA process is described in 21 CFR 312, but the law describing the requirement for an NDA process is in Section 505 of the FDCA. Because the regulations rather than the laws themselves have a greater impact on toxicological practice, greater emphasis is placed on regulation in this chapter. For a complete review of FDA law, the reader is referred to the monograph by the Food and Drug Law Institute in 1999.

Laws authorize the activities and responsibilities of the various federal agencies. All proposed laws before the U.S. Congress are referred to committees for review and approval. The committees responsible for FDA

**TABLE 5 Congressional Committees Responsible for FDA Oversight**

| | |
|---|---|
| Authorization | |
| Senate | All public health service agencies are under the jurisdiction of the Labor and Human Resources Committee. |
| House | Most public health agencies are under the jurisdiction of Health and the Environmental Subcommittee of the House Energy and Commerce Committee. |
| Appropriation | |
| Senate | Unlike most other public health agencies, the FDA is under the jurisdiction of the Agriculture, Rural Development, and Related Agencies Subcommittee of the Senate Appropriations Committee. |
| House | Under the jurisdiction of the Agriculture, Rural Development, and Related Agencies Subcommittee of the House Appropriations Committee. |

oversight are summarized in Table 5. This table also highlights the fact that authorizations and appropriations (the funding necessary to execute authorizations) are handled by different committees.

## 4 ORGANIZATIONS REGULATING DRUG AND DEVICE SAFETY IN THE UNITED STATES

The agency formally charged with overseeing the safety of drugs in the United States is the FDA. It is headed by a commissioner who reports to the secretary of the Department of Health and Human Services (DHHS) and has a tremendous range of responsibilities. Drugs are overseen primarily by the CDER (though some therapeutic or health care entities are considered biologics and are overseen by the corresponding CBER). Figure 1 presents the organization of CDER. The organization of CBER is shown in Figure 2.

Most of the regulatory interactions of toxicologists are with the two offices of Drug Evaluation, which have under them a set of groups focused on areas of therapeutic claim (cardiorenal, neuropharmacological, gastrointestinal and coagulation, oncology and pulmonary, metabolism and endocrine, anti-infective and antiviral). Within each of these are chemists, pharmacologists/toxicologists, statisticians, and clinicians. When an INDA is submitted to the offices of Drug Evaluation, it is assigned to one of the therapeutic groups based on its area of therapeutic claim. Generally, it will remain with that group throughout its regulatory approval "life." When allowed, INDs grant investigators the ability to go forward into clinical (human) trials with their drug candidate in a predefined manner, advancing through various steps of evaluation in human (and in additional preclinical or animal studies) until an NDA can be supported, developed, and submitted. Likewise for biological products, the PLA or other applications (INDA, IND) are handled by the offices of Biological Products Review of the CBER.

For drugs, there is at least one nongovernmental body which must review and approve various aspects—the USP (established in 1820), which maintains

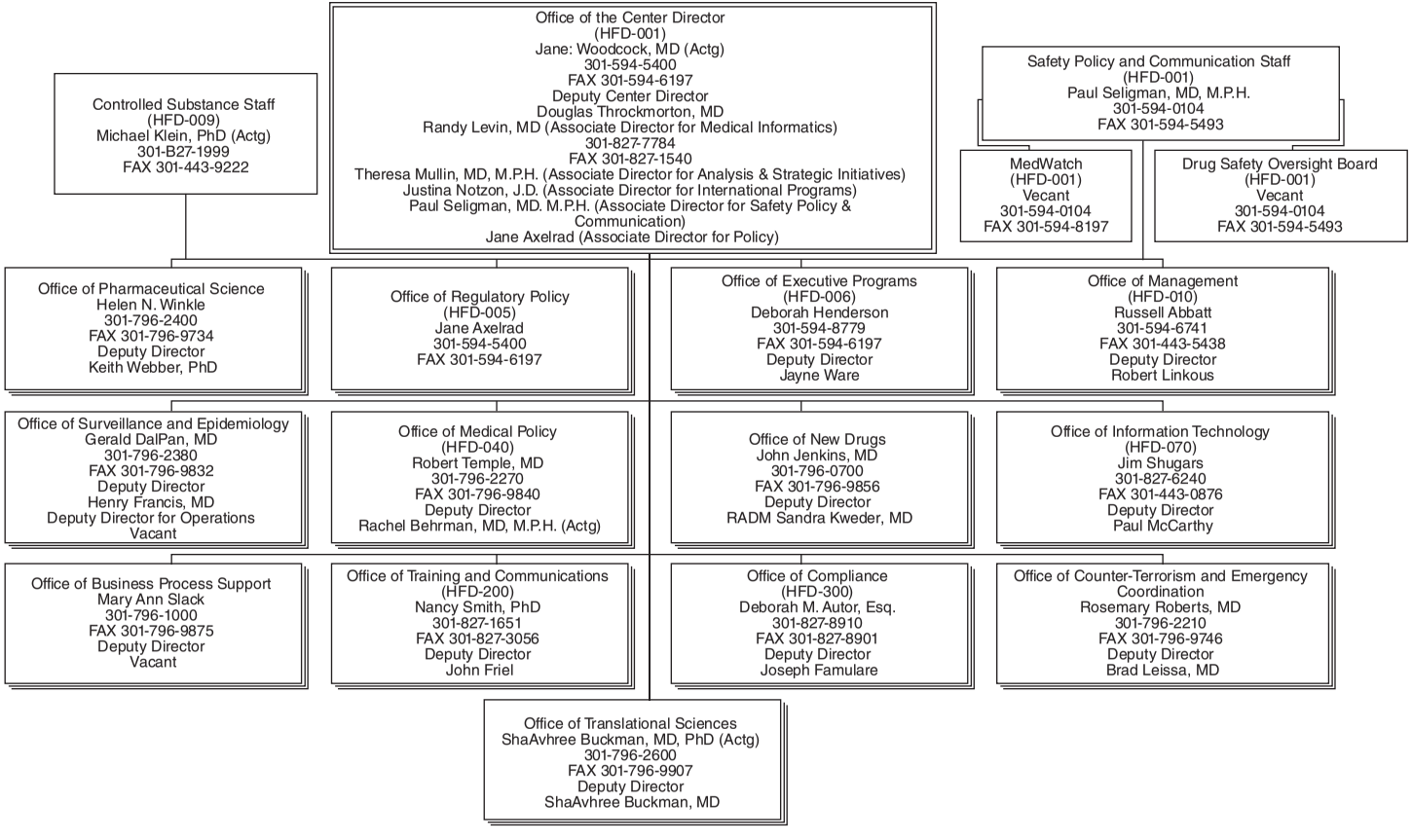


***Figure 1*** *Organizational Chart: Center for Drug Evaluation and Research.*

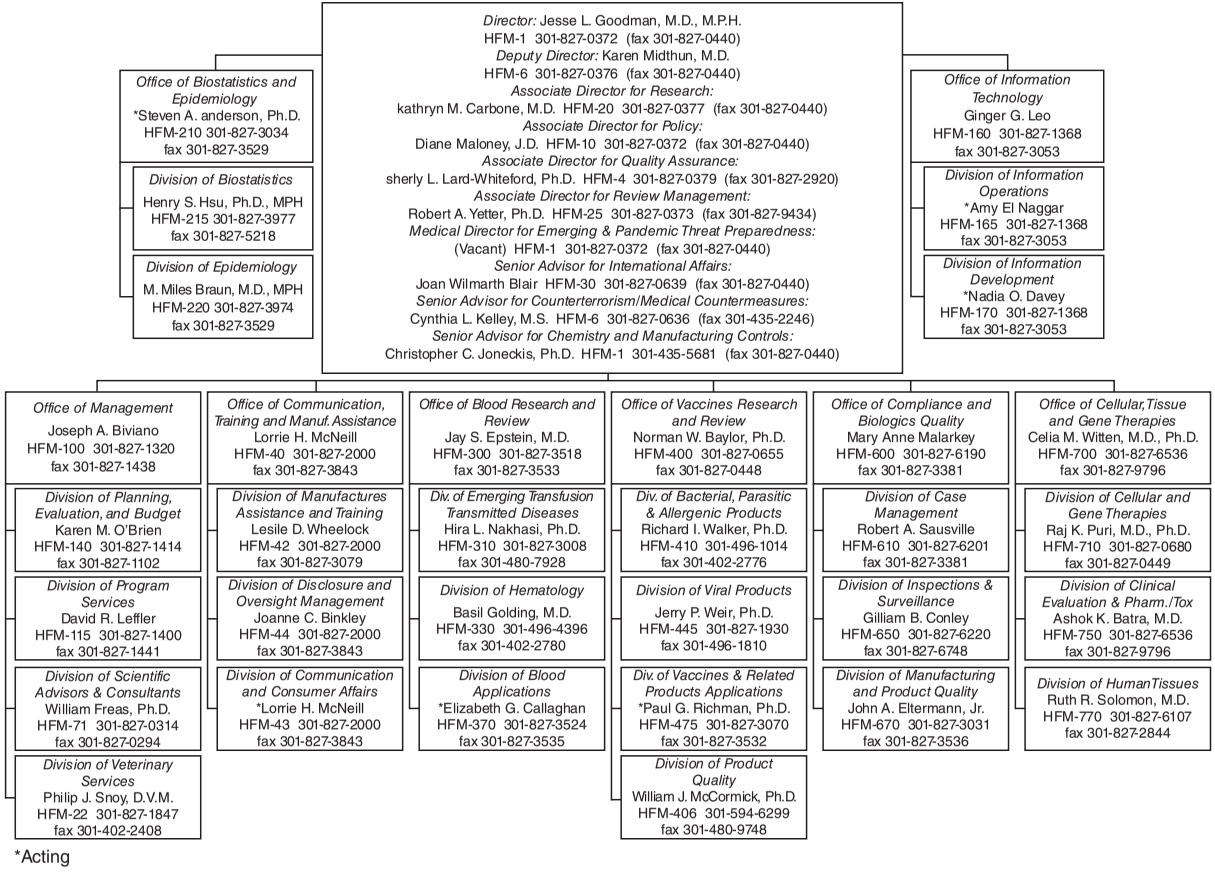


***Figure 2*** *Organizational Chart: Center for Biologics Evaluation and Research.*

(and revises) the compendia of the same name, and the National Formulary, which sets drug composition standards (Ember, 2001). This volume sets forth standards for purity of products in which residues may be present and tests for determining various characteristics of drugs, devices, and biologics. The USP also contains significant "guidance" for the evaluation.

## 5 PROCESS OF PHARMACEUTICAL PRODUCT DEVELOPMENT AND APPROVAL

Except for a very few special cases (treatments for life-threatening diseases such as cancer or AIDS), the safety assessment of new drugs as mandated by regulations are seemingly determined in a rather fixed manner. The IND is filed to support this clinical testing. An initial set of studies [typically, studies of appropriate length by the route intended for humans are performed in both a rodent (typically rat) and a nonrodent (usually a dog or a primate)] is required to support phase I clinical testing. Such phase I testing is intended to evaluate the safety ("tolerance" in clinical subjects), pharmacokinetics, and general biological effects of a new drug and is conducted in normal volunteers (almost always males).

Successful completion of phase I testing allows, with the approval of the FDA, progression into phase II clinical testing. Here, selected patients are enrolled to evaluate therapeutic efficacy, dose ranging, and more details about the pharmacokinetics and metabolism. Longer term systemic toxicity studies must be in conformity with the guidelines that are presented in the next section. Once a sufficient understanding of the actions, therapeutic dose response, and potential risk-to-benefit ratio of the drug is in hand (once again, with FDA approval), trials move into phase III testing.

Phase III tests are large, long, and expensive. They are conducted using large samples of selected patients and are intended to produce proof of safety and efficacy of the drug. Two studies providing statistically significant proof of the claimed therapeutic benefit must be provided. All the resulting data from preclinical and clinical animal studies are organized in a specified format in the form of a NDA, which is submitted to the FDA.

By the time that phase III testing is completed, some additional preclinical safety tests must also generally be in hand. These include the three separate reproductive and developmental toxicity studies (segments I and III in the rat and segment II in the rat and rabbit) and carcinogenicity studies in both rats and mice (unless the period of therapeutic usage is intended to be very short). Some assessment of genetic toxicity will also be expected.

The ultimate product of the pharmaceutical toxicologist will thus generally be the toxicology summaries of the IND and NDA (or PLA). For medical devices, the equivalents are the investigational device exemption (IDE) and product development notification (PDN). Data required to support each of these documents is specified in a series of guidelines, as will be discussed below.

Acceptance of these applications is contingent not only upon adherence to guidelines and good science but also adherence to GLPs.

## 6 TESTING GUIDELINES

### 6.1 Toxicity Testing: Traditional Pharmaceuticals

Although the 1938 act required safety assessment studies, no consistent guidelines were available. Guidelines were first proposed in 1949 and published in the *Food, Drug and Cosmetic Law Journal* that year (Burns, 1983). Following several revisions, these guidelines were issued as *The Appraisal Handbook* in 1959. While never formally called a guideline, it set the standard for preclinical toxicity test design for several years. The current basic guidelines for testing required for safety assessment in support of the phases of clinical development of drugs were first outlined by Goldenthal (1968) and later incorporated into a 1971 FDA publication entitled *FDA Introduction to Total Drug Quality*.

### 6.2 General or Systematic Toxicity Assessment

Table 6 presents an overview of the current FDA toxicity testing guidelines for human drugs. Table 7 presents the parallel ICH guidances (ICH, 2000), which are now largely supplanting the FDA guidelines. They are misleading in their apparent simplicity, however. First, each of the systemic toxicity studies in these guidelines must be designed and executed in a satisfactory manner. Sufficient animals must be used to have confidence in finding and characterizing any adverse drug actions that may be present. In practice, as the duration of the study increases, small doses are administered and larger numbers of animals must be employed per group. These two features—dosage level and group size—are critical to study designs. Table 8 presents general guidance on the number of animals to be used in systemic studies. These and other technical considerations for the safety assessment of pharmaceuticals are presented in detail in Gad (1994).

The protocols discussed thus far have focused on general or systemic toxicity assessment. The agency and, indeed, the lay public have a special set of concerns with reproductive toxicity, fetal/embryo toxicity, and developmental toxicity (also called *teratogenicity*). Collectively, these concerns often go by the acronym DART (developmental and reproductive toxicity) or RTF (reproduction, teratogenicity, fertility). Segment II studies are more designed to detect developmental toxicity. Only pregnant females are dosed during critical period of organogenesis. Generally, the first protocol DART test (exclusive of range-finding studies) is a segment I study of rats in fertility and general reproductive performance. This is generally done while the drug is in phase II clinical trials. Alternatively, many companies are now performing the segment II teratology study in rats before the segment I study because the

**TABLE 6 Synopsis of General Guidelines for Animal Toxicity Studies for Drugs**

| Category | Duration of Human Administration | Clinical Phase | Subacute or Chronic Toxicity | Special Studies |
|---|---|---|---|---|
| Oral or parenteral | Several days | I, II, III, NDA | 2 Species; 2 weeks | For parentally administered drugs |
| | Up to 2 weeks | I | 2 Species; 4 weeks | |
| | | II | 2 Species; up to 4 weeks | |
| | | III, NDA | 2 Species; up to 3 months | Compatibility with blood where applicable |
| | Up to 3 months | I, II | 2 Species; 4 weeks | |
| | | III | 2 Species; 3 months | |
| | | NDA | 2 Species; up to 6 months | |
| | 6 Months to unlimited | I, II | 2 Species; 3 months | |
| | | III | 2 Species; 6 months or longer | |
| | | NDA | 2 Species; 9 months (nonrodent) and 6 months (rodent) +2 rodent species for CA; □18 months (mouse), 24 months (rat). Mouse may be replaced with an allowable transgenic mouse study. | |
| Inhalation□(General □anesthetics), dermal | | I, II, III, NDA | 4 Species; 5 days (3 hours/day) | |
| | Single application | I | 1 Species; single 24-h exposure followed by 2-week observation | Sensitization |
| | Single or short-term application | II | 1 Species; 20-day repeated exposure (intact and abraded skin) | |
| | Short-term application | III | As above | |
| | Unlimited application | NDA | As above, but intact skin study extended up to 6 months | |
| Ophthalmic | Single application | I | | Eye irritation tests with graded doses |
| | Multiple application | I, II, III | 1 Species; 3 weeks daily applications, as in clinical use | |

**TABLE 6** ***Continued***

| Category | Duration of Human Administration | Clinical Phase | Subacute or Chronic Toxicity | Special Studies |
|---|---|---|---|---|
| | | NDA | 1 Species; duration commensurate with period of drug administration | |
| Vaginal or rectal | Single application | I | | Local and systematic toxicity after vaginal or rectal application in 2 species |
| | Multiple application | I, II, III, NDA | 2 Species; duration and number of applications determined by proposed use | |
| Drug combinations | | I, II, III, NDA | 2 Species; up to 3 months | Lethality by appropriate route, compared to components run concurrently in 1 species |

**TABLE 7 Duration of Repeated-Dose Toxicity Studies to Support Clinical Trials and Marketing**[a]

| Duration of Clinical Trials | Minimum Duration of Repeated-Dose Toxicity Studies[b] | | Duration of Clinical Trials | Minimum Duration of Repeated-Dose Toxicity Studies[c] | |
|---|---|---|---|---|---|
| | Rodents | Nonrodents | | Rodents | Nonrodents |
| Single dose | 2 Weeks[d] | 2 Weeks | Up to 2 Weeks | 1 Month | 1 Month |
| Up to 2 weeks | 2 Weeks[d] | 2 Weeks | Up to 1 Month | 3 Months | 3 Months |
| Up to 1 month | 1 Month | 1 Month | Up to 3 Months | 6 Months | 3 Months |
| Up to 6 months | 6 Months | 6 Months[e] | >3 Months | 6 Months | Chronic[d] |
| >6 Months | 6 Months | Chronic[e] | | | |

[a]In Japan, if there are no phase II clinical trials of equivalent duration to the planned phase III trials, conduct of longer duration toxicity studies is recommended as given above.
[b]Data from 6 months of administration in nonrodents should be available before the initiation of clinical trials longer than 3 months. Alternatively, if applicable, data from a 9-month nonrodent study should be available before the treatment duration exceeds that which is supported by the available toxicity studies.
[c]The table also reflects the marketing recommendations in the three regions except that a chronic nonrodent study is recommended for clinical use >1 month.
[d]In the United States, as an alternative to 2-week studies, single-dose toxicity studies with extended examinations can support single-dose human trials (4).
[e]To support phase I and II trials in the EU and phase I, II, and III trials in the United States and Japan.

**TABLE 8 Numbers of Animals per Dosage Group in Systemic Toxicity Studies**

| Study Duration (per Sex) | Rodents (per Sex) | Nonrodents |
|---|---|---|
| 2–4 Weeks | 5 | 3 |
| 13 Weeks | 20[a] | 6 |
| 26 Weeks | 30 | 8 |
| Chronic | 50 | 10 |
| Carcinogenicity | 60[b] | Applies only to contraceptives |
| Bioassays | | Applies only to contraceptives |

[a]Starting with 13-week studies, one should consider adding animals (particularly to the high dose) to allow evaluation of reversal of effects.
[b]In recent years there have been decreasing levels of survival in rats on 2-year studies. What is required is that at least 20–25 animals/sex/group survive at the end of the study. Accordingly, practice is beginning to use 70 or 75 animals per sex per group.

former is less time and resource intensive. One or both should be completed before including women of child-bearing potential in clinical trials. The FDA requires teratogenicity testing in two species—a rodent (rat or mouse) and the rabbit. The use of the rabbit was instituted as a result of the finding that thalidomide was a positive teratogen in the rabbit but not in the rat. On occasion, when a test article is not compatible with the rabbit, teratogenicity data in the mouse may be substituted. There are also some specific classes of therapeutics (e.g., the quinalone antibiotics) where segment II studies in primates are effectively required prior to product approval. Both should be completed before entering phase III clinical trials. The most complicated of the DART protocols—segment III—is generally commenced during phase III trials and should be part of the NDA. There are differences in the different national guidelines (as discussed later with international considerations) regarding the conduct of these studies. The large multinational drug companies try to design their protocols to be in compliance with as many of the guidelines as possible to avoid duplication of testing while allowing the broadest possible approval and marketing of therapeutics.

### 6.3 Genetic Toxicity Assessment

Genetic toxicity testing generally focuses on the potential of a new drug to cause mutations (in single-cell systems) or other forms of genetic damage. The tests, generally short in duration, often rely on in vitro systems and generally have a single endpoint of effect (point mutations, chromosomal damage, etc.). For a complete review of protocols, technology, and so on, the reader is referred to Brusick (1989). It is of interest that the FDA has no standard or statutory requirement for genetic toxicity testing but generally expects to see at least some such tests performed and will ask for them if the issue is not addressed. If one performs such a study, any data collected, of course, must be sent to the agency as part of any INDA, PLA, or NDA. These studies have

yet to gain favor with the FDA (or other national regulatory agencies) as substitutes for in vivo carcinogenicity testing. However, even with completed negative carcinogenicity tests, at least some genetic toxicity assays are generally required. Generally, pharmaceuticals in the United States are evaluated for mutagenic potential (e.g., the Ames assay) or for chromosomal damage (e.g., the in vivo mouse micronucleus test). In general, in the United States, pharmaceutical companies apply genetic toxicity testing in the following fashion:

- *As Screen* An agent that is positive in one or more genetic toxicity tests may be more likely than one that is negative to be carcinogenic and, therefore, may not warrant further development.
- *As Adjunct* An agent that is negative in carcinogenicity testing in two species and also negative in a genetic toxicity battery is more likely than not to be noncarcinogenic in human beings.
- *To Provide Mechanistic Insight* For example, if an agent is negative in a wide range of genetic toxicity screens but still produces tumors in animals, then one could hypothesize that an epigenetic mechanism was involved.

While not officially required, the FDA does have the authority to request, on a case-by-case basis, as specific tests it feels may be necessary to address a point of concern. A genetic toxicity test could be part of such a request. In general, therefore, companies deal with genetic toxicity (after "screening") on a case-by-case basis, as dictated by good science. If more than a single administration is intended, common practice is to perform the tests prior to submitting an IND.

### 6.4 Toxicity Testing: Biotechnology Products

As mentioned, the regulation of traditional pharmaceuticals (small molecules such as aspirin or digitalis) and the regulation of biologicals (proteins such as vaccines and antitoxins derived from animal sources) have very different histories. See the discussion on biologics in Section 3.5. Until 1972, the National Institutes of Heath (NIH; or its forerunning agency, the Hygienic Laboratory of the Department of the Treasury) was charged with the responsibilities of administering the Virus Act of 1902. With the passage of the Food and Drug Laws of 1906, 1938, and 1962, there was reoccurring debate about whether these laws applied or should apply to biologics (Pendergast, 1983). This debate was resolved when the authority for the regulation of biologics was transferred to the FDA's new Bureau of Biologics (now the CBER) in 1972. Since then, there appears to have been little difference in the matter of regulation for biologics and pharmaceuticals. The FDA essentially regulates biologics as described under the 1902 act but then uses the rule-making authority granted under the Food and Drug Act to "fill in the gaps."

The Bureau of Biologics was once a relatively "sleepy" agency, primarily concerned with the regulation of human blood products and vaccines used for mass immunization programs. The authors of the 1902 law could hardly have foreseen the explosion in biotechnology that occurred in the 1980s. New technology created a welter of new biological products, such as recombinant-DNA-produced proteins (e.g., tissue plasminogen activator), biological response modifiers (cytokinins and colony-stimulating factors), monoclonal antibodies, antisense oligonucleotides, and self-directed vaccines (raising an immune response to self-proteins such as gastrin for therapeutic reasons). The new products raised a variety of new questions on the appropriateness of traditional methods of evaluating drug toxicity that generated several PTC documents. For the sake of brevity, this discussion will focus on the recombinant DNA (rDNA) proteins. Some of the safety issues that have been raised over the years are as follows:

- The appropriateness of testing a human-specific peptide hormone in non-human species
- The potential that the peptide could break down due to nonspecific metabolism, resulting in products that had no therapeutic value or even a toxic fragment
- The potential sequelae to an immune response (formation of neutralizing antibodies, provoking an autoimmune or a hypersensitivity response), pathology due to immune precipitation, and so on
- The presence of contamination with oncogenic virus DNA (depending on whether a bacterial or mammalian system was used on the synthesizing agent) or endotoxins
- The difficulty interpreting the scientific relevance of response to supra-physiological systemic doses of potent biological response modifiers.

The intervening last few years have shown that some of these concerns were more relevant than others. The "toxic peptide fragment" concern, for example, has been shown to be without merit. The presence of potentially oncogenic virus DNA and endotoxins is a quality assurance concern and is not truly a toxicological problem. Regardless of the type of synthetic pathway, all proteins must be synthesized in compliance with good manufacturing practices (GMPs). Products must be as pure as possible, free of not only rDNA but also other types of cell debris (endotoxin). Batch-to-batch consistency with regard to molecular structure must also be demonstrated using appropriate methods (e.g., amino acid). The regulatory thinking and experience over the last 15 years has come together in the document *S6 Preclincial Safety Evaluation of Biotechnology-Derived Pharmaceuticals*, prepared by the ICH. The FDA (the CDER and CBER jointly) has published the document as a guidance for industry (Anonymous, 1997a,b). The document intended to provide basic guidance for the preclinical evaluation of biotechnology-derived products,

including proteins and peptides, either produced by cell culture using rDNA technology (but did not cover antibiotics, allergenic extracts, heparin, vitamins, cellular drug products vaccines) or as other products regulated as biologics. Items covered are summarized as follows:

- *Test Article Specifications* In general, the product that is used in the definitive pharmacology and toxicology studies should be comparable to the product proposed for the initial clinical studies.
- *Animal Species/Model Selection* Safety evaluation should include the use of relevant species, in which the test article is pharmacologically active due, for example, to the expression of the appropriate receptor molecule. These can be screened with in vitro receptor-binding assays. Safety evaluation should normally include two appropriate species if possible and/or feasible. The potential utility of gene knockout and/or transgenic animals in safety assessment is discussed.
- *Group Size* No specific numbers are given, but it does state that a small sample size may lead to failure to observe toxic events.
- *Administration* The route and frequency should be as close as possible to that proposed for clinical use. Other routes can be used when scientifically warranted.
- *Immunogenicity* It has also been clearly demonstrated in the testing of rDNA protein products that animals will develop antibodies to foreign proteins. This response has been shown to neutralize (rapidly remove from circulation) the protein, but no pathological conditions have been shown to occur as a sequelae to the immune response. Bear in mind, however, that interleukins have powerful effects on immune response, but these are due to their physiological activity and not to an antigen–antibody response. The first has to do with "neutralizing antibodies"; that is, is the immune response so great that the test article is being removed from circulation as fast as it is being added? If this is the case, does long-term testing of such a chemical make sense? In many cases, it does not. The safety testing of any large molecule should include the appropriate assays for determining whether the test system has developed a neutralizing antibody response. Depending on the species, route of administration, intended therapeutic use, and development of neutralizing antibodies (which generally takes about two weeks), it is rare for a toxicity test on an rDNA protein to be longer than four weeks duration. However, if the course of therapy in humans is to be longer than two weeks, formation of neutralizing antibodies must be demonstrated or longer term testing performed. The second antigen–antibody formation concern is that a hypersensitivity response will be elicited. Traditional preclinical safety assays are generally adequate to guard against this if they are two weeks or longer in duration and the relevant endpoints are evaluated.

- *Safety Pharmacology* It is important to investigate the potential for unwanted pharmacological activity in appropriate animal models and to incorporate monitoring for these activities in the toxicity studies.
- *Exposure Assessment* Single- and multiple-dose pharmacokinetics, toxicokinetics, and tissue distribution studies in relevant species are useful. Proteins are not given orally, demonstrating absorption and mass balance are not typically primary considerations. Rather, this segment of the test should be designed to determine the half-life (and other appropriate pharmacokinetic descriptor parameters), the plasma concentration associated with biological effects, and potential changes due to the development of neutralizing antibodies.
- *Reproductive Performance and Developmental Toxicity Studies* These will be dictated by the product, clinical indication, and intended patient population.
- *Genotoxicity Studies* The S6 document states that the battery of genotoxicity studies routinely conducted for traditional pharmaceuticals are not appropriate for biotechnology-derived pharmaceuticals. In contrast to small molecules, genotoxicity testing with a battery of in vitro and in vivo techniques of protein molecules has not become common U.S. industry practice. Such tests are not formally required by the FDA but, if performed, have to be reported. They are required by European and Japanese regulatory authorities. This has sparked a debate as to whether or not genotoxicity testing is necessary or appropriate for rDNA protein molecules. It is the author's opinion that such testing is scientifically of little value. First, large protein molecules will not easily penetrate the cell wall of bacteria or yeast, and (depending on size, charge, lipophilicity, etc.) penetration across the plasma lemma of mammalian cells will be highly variable. Second, if one considers the well-established mechanism(s) of genotoxicity of small molecules, it is difficult to conceive how a protein can act in the same fashion. For example, proteins will not be metabolized to be electrophilic active intermediates that will crosslink guanine residues. In general, therefore, genotoxicity testing with rDNA proteins is a waste of resources. It is conceivable, however, that some proteins, because of their biological mechanism of action, may stimulate the proliferation of transformed cells. For example, it is a feasible hypothesis that a colony-stimulating factor could stimulate the proliferation of leukemic cells (it should be emphasized that this is a hypothetical situation, presented here for illustrative purposes). Again, this is a question of a specific pharmacological property, and such considerations should be tested on a case-by-case basis.
- *Carcinogenicity Studies* These are generally inappropriate for biotechnology-derived pharmaceuticals; however, some products may have the potential to support or induce proliferation of transformed cells, possibly leading to neoplasia. When this concern is present, further studies in relevant animal models may be needed.

These items are covered in greater detail in the S6 guidance document.

So, given the above discussion, what should the toxicology testing package of a typical rDNA protein resemble? Based on the products that have successfully wound their way through the regulatory process, the following generalizations can be drawn:

- The safety tests look remarkably similar to those for traditional tests. Most have been done on three species: the rat, the dog, or the monkey. The big difference has to do with the length of the test. It is rare for a safety test on a protein to be more than 13 weeks long.
- The dosing regimens can be quite variable and at times very technique intensive. These chemicals are almost always administered by a parenteral route of administration—normally intravenously or subcutaneously. Dosing regimens have run the range from once every 2 weeks for an antihormone "vaccine" to continuous infusion for a short-lived protein.
- As reviewed by Ryffel (1996) most side effects in man of a therapy with rDNA therapy may be predicted by data from experimental toxicology studies, but there are exceptions. Interleukin 6 (IL-6), for example, induced a sustained increase in blood platelets and acute phase proteins, with no increase in body temperature. In human trials, however, there were increases in temperature.
- The S6 document also mentions monoclonal antibody products. Indeed, many of the considerations for rDNA products are also applicable to monoclonal antibodies (including hybridized antibodies). With monoclonal antibodies, there is the additional concern of cross-reactivity with nontarget molecules.

As mentioned, the rapid development in the biotechnology industry has created some confusion as to what arm of the FDA is responsible for such products. In October 1992, the two major reviewing groups, CBER and CDER, reached a series of agreements to explain and organize the FDA's position on products that did not easily fall into its traditional classification schemes. CDER will continue to have responsibility for traditional chemically synthesized molecules as well as those purified from mineral or plant sources (except allergenics), antibiotics, hormones (including insulin, growth hormone, etc.), most fungal or bacterial products (disaccharidase inhibitors), and most products from animal or solid human tissue sources. CBER will have responsibility for products subject to licensure (BLA), including all vaccines, human blood or blood-derived products (as well as drugs used for blood banking and transfusion), immunoglobulin products, products containing intact cells, fungi, viruses, proteins produced by cell culture or transgenic animals, and synthetic allergenic products. This situation was further simplified by the introduction of the concept of "well-characterized biologics." When introduced during the

debate on FDA reform in 1996, the proposed section of S-1447 stated: "Biological products that the secretary determines to be well-characterized shall be regulated solely under the Federal Food, Drug and Cosmetic Act." Under this concept, highly purified, well-characterized therapeutic rDNA proteins would be regulated by CDER, regardless of therapeutic target (Anonymous, 1996).

### 6.5 Toxicity/Safety Testing: Cellular and Gene Therapy Products

Human clinical trials of cellular and gene therapies involve administration to patients of materials considered investigational biological, drug, or device products. Somatic cell therapy refers to the administration to humans of autologous, allogenic, or xenogenic cells which have been manipulated or processed ex vivo. Gene therapy refers to the introduction into the human body of genes or cells containing genes foreign to the body for the purposes of prevention, treatment, diagnosing, or curing disease.

Sponsors of cellular or gene therapy clinical trials must file an INDA or in certain cases an IDE with the FDA before initiation of studies in humans. It is the responsibility of the CBER to review the application and determine if the submitted data and the investigational product meet applicable standards. The critical parameters of identity, purity, potency, stability, consistency, safety, and efficacy relevant to biological products are also relevant to cellular and gene therapy products.

In 1991, the FDA first published "Points to Consider on Human Somatic Cell and Gene Therapy." At this time virtually all gene therapies were retroviral and were prepared as ex vivo somatic cell therapies. This was subsequently reviewed by Kessler et al. (1993). While the data for certain categories of information such as the data regarding the molecular biology were defined in previous guidance documents relating to recombinant DNA products, the standards for preclinical and clinical development were less well defined. Over the past five years, the field has advanced to include not only new vectors but also novel routs of administration. "Points to Consider on Human Somatic Cell and Gene Therapy" has thus been recently amended (1996) to reflect both the advancements in product development and more importantly the accumulation of safety information over the past five years.

FDA regulations state that the sponsor must submit, in the IND, adequate information about pharmacological and toxicological studies of the drug, including laboratory animals or in vitro studies on the basis of which the sponsor has considered that it is reasonably safe to conduct the proposed clinical investigation. For cellular and gene therapies, designing and conducting relevant preclinical safety testing have been a challenge to both the FDA and the sponsor. For genes delivered using viral vectors, the safety of the vector system per se must be considered and evaluated.

The preclinical knowledge base is initially developed by designing studies to answer fundamental questions. The development of this knowledge base is

generally applicable to most pharmaceuticals as well as biopharmaceuticals and includes data to support (1) the relationship of the dose to the biological activity, (2) the relationship of the dose to the toxicity, (3) the effect of route and/or schedule on activity or toxicity, and (4) identification of the potential risks for subsequent clinical studies. These questions are considered in the context of indication and/or disease state. In addition there are often unique concerns relating to the specific category or product class.

For cellular therapies safety concerns may include development of a database from studies specifically designed to answer questions relating to growth factor dependence, tumorigenicity, local and systemic toxicity, and effects on host immune responses, including immune activation and altered susceptibility to disease. For viral-mediated gene therapies, specific questions may relate to the potential for overexpression of the transduced gene, transduction of normal cells/tissues, genetic transfer to germ cells and subsequent alterations to the genome, recombination/rescue with endogenous virus, reconstitutions of replication competence, potential for insertional mutagenesis/malignant transformation, altered susceptibility to disease, and/or potential risk(s) to the environment.

To date cellular and gene therapy products submitted to the FDA have included clinical studies indicated for bone marrow marking, cancer, cystic fibrosis, AIDS, and inborn errors of metabolism and infectious diseases. Of the current active INDs approximately 78% have been sponsored by individual investigators or academic institutions and 22% have also been industry sponsored. In addition to the variety of clinical indications the cell types have also been varied. Examples include tumor-infiltrating lymphocytes (TILs) and lymphocyte-activated killer (LAK) cells, selected cells from bone marrow and peripheral blood lymphocytes (e.g., stem cells), myoblasts, tumor cells, and encapsulated cells (e.g., islet cells and adrenal chromaffin cells).

***Cellular Therapies*** Since 1984 CBER has reviewed close to 300 somatic cell therapy protocols. Examples of the specific categories include manipulation, selection, mobilization, tumor vaccines, and others:

*Manipulation* Autologous, allogenic, or xenogenic cells which have been expanded, propagated, manipulated, or had their biological characteristics altered ex vivo (e.g., TIL or LAK cells; islet cells housed in a membrane).

*Selection* Products designed for positive or negative selection if autologous or allogenic cells are intended for therapy (e.g., purging of tumor from bone marrow, selection of $CD34^+$ cells).

*Mobilization* In vivo mobilization of autologous stem cells intended for transplantation.

*Tumor Vaccines* Autologous or allogenic tumor cells which are administered as vaccine (e.g., tumor cell lines, tumor cell lysates, primary

explant). This group also includes autologous antigen-presenting cells pulsed with tumor-specific peptides or tumor cell lysates.

*Other* Autologous, allogenic, and xenogenic cells which do not specifically fit above. This group includes cellular therapies such as extracorporeal liver assist devices.

***Gene Therapies*** The types of vectors that have been used, or proposed, for gene transduction include retrovirus, adenovirus, adeno-associated viruses, other viruses (e.g., herpes, vaccinia), and plasmid DNA. Methods for gene introduction include ex vivo replacement, drug delivery, marker studies, and others and in vivo, viral vectors, plasmid vectors, and vector producer cells.

**Ex Vivo**

*Replacement* Cells transduced with a vector expressing a normal gene in order to correct or replace the function of a defective gene.

*Drug Delivery* Cells transduced with a vector expressing a gene encoding a therapeutic molecule which can be novel or native to the host.

*Marker Studies* Cells (e.g., bone marrow, stem cells) transduced with a vector expressing a marker or reporter gene used to distinguish it from other similar host tissues.

*Other* Products which do not specifically fit above (e.g., tumor vaccines in which cells are cultured or transduced ex vivo with a vector).

**In Vivo**

*Viral Vectors* The direct administration of a viral vector (e.g., retrovirus, adenovirus, adeno-associated virus, herpes, vaccinia) to patients.

*Plasmid Vectors* The direct administration of plasmid vectors with or without other vehicles (e.g., lipids) to patients.

*Vector Producer Cells* The direct administration of retroviral vector producer cells [e.g., murine cells producing the Hidden Markov Model ToolKit (HTK) vector] to patients.

***Preclinical Safety Evaluation*** The goal of preclinical safety evaluation includes recommendation of an initial safe starting dose and safe dose escalation scheme in humans, identification of potential target organ(s) of toxicity, identification of appropriate parameters for clinical monitoring, and identification of "at-risk" patient population(s). Therefore, when feasible, toxicity studies should be performed in relevant species to assess a dose-limiting toxicity. General considerations in study design include selection of the model (e.g., species, alternative model, animal model or disease), dose (e.g., route, frequency, and duration), and study endpoint (e.g., activity and/or toxicity).

The approach to preclinical safety evaluation of biotechnology-derived products, including novel cellular and gene therapies, has been referred to as

the "case-by case" approach. This approach is science based, data driven, and flexible. The major distinction from past practices from traditional pharmaceuticals is that the focus is directed at asking specific questions across various product categories. Additionally, there is a consistent reevaluation of the knowledge base to reassess real or theoretical safety concerns and hence reevaluation of the need to answer the same questions across all product categories. In some cases there may even be conditions which may not need specific toxicity studies, for example, when there is a strong efficacy model which is rationally designed to answer specific questions and/or there is previous human experience with a similar product with respect to dose and regimen.

## *Basic Principles for Preclinical Safety Evaluation of Cellular and Gene Therapies*

### Biotechnology-Derived Products in General

- Use of product in animal studies that is comparable or the same as the product proposed for clinical trial(s)
- Adherence to basic principles of GLP to ensure quality of the study, including a detailed protocol prepared prospectively
- Use of the same or similar route and method of administration as proposed for clinical trials (whenever possible)
- Determination of appropriate doses delivered based upon preliminary activity obtained from both in vitro and in vivo studies (i.e., finding a dose likely to be effective and not dangerous, at no observed adverse effect level, and a dose causing dose-limiting toxicity)
- Selection of one or more species sensitive to the endpoint being measured, for example, infections or pathological sequelae and/or biological activity or receptor binding
- Consideration of animal model(s) of disease, which may be better to assess the contribution of changes in physiological or underlying physiology to safety and efficacy
- Determination of affect on host immune response
- Localization/distribution studies—evaluation of target tissue, normal surrounding tissue, and distal tissue sites and any alteration in normal or expected distribution
- Local reactogenicity at target tissue

### Additional Considerations for Cellular Therapies

- Evaluation of cytopathogenicity
- Evaluation of signs of cell transformation/growth factor dependence effect on animal cells, normal human cells, and cells prone to transform easily

- Determination of alteration in cell phenotype, altered cell products, and/or function
- Tumorigenicity

**Additional Considerations for Gene Therapies**

- Determination of phenotype/activation state of effector cells
- Determination of vector/transgene toxicity
- Determination of potential transfer to germ line
- In vitro challenge studies—evaluation of recombination or complementation, potential for "rescue" for subsequent infection with wild-type virus
- Determination of persistence of cells/vector
- Determination of potential for insertional mutagenesis (malignant transformation)
- Determination of environmental spread (e.g., viral shedding)

## 7 TOXICITY TESTING: SPECIAL CASES

On paper, the general-case guidelines for the evaluation of the safety of drugs are relatively straightforward and well understood. However, there are also a number of special-case situations under which either special rules apply or some additional requirements are relevant. The more common of these are summarized below.

### 7.1 Oral Contraceptives

Oral contraceptives are subject to special testing requirements. These have recently been modified so that, in addition to those preclinical safety tests generally required, the following are also required (Berliner, 1974):

- A three-year carcinogenicity study in beagles (this is a 1987 modification in practice from earlier FDA requirements and the 1974 publication)
- A rat reproductive (segment I) study, including a demonstration of return to fertility

### 7.2 Life-Threatening Diseases (Compassionate Use)

Drugs to treat life-threatening diseases are not strictly held to the sequence of testing requirements as put forth in Table 3 because the potential benefit on any effective therapy in these situations is so high. In the early 1990s, this situation applied to AIDS-associated diseases and cancer. The development of more effective HIV therapies (protease inhibitors) has now made cancer therapy more the focus of these considerations. Though the requirements

for safety testing prior to initial human trials are unchanged, subsequent requirements are flexible and subject to negotiation and close consultation with the FDA's Division of Oncology (within CDER) (FDA, 1988c). The more recent thinking on anticancer agents has been reviewed by DeGeorge et al. (1998). The preclinical studies that will be required to support clinical trials and marketing of new anticancer agents will depend on the mechanism of action and the target clinical population. Toxicity studies in animals will be required to support initial clinical trials. These studies have multiple goals:

- To determine a starting dose for clinical trials
- To identify target organ toxicity and assess recovery
- To assist in the design of clinical dosing regimens

The studies should generally conform to the protocols recommended by the National Cancer Institute. as discussed by Grieshaber (1991). In general, it can be assumed that most antineoplastic cytotoxic agents will be highly toxic. Two studies are essential to support initial clinical trials (IND phase) in patients with advanced disease. These studies are 5–14 days long with longer recovery periods. A study in rodents is required that identifies those doses that produce either life-threatening or non-life-threatening toxicity. Using the information from this first study, a second study in nonrodents (generally the dog) is conducted to determine if the tolerable dose in rodents is life threatening. Doses are compared on a milligram-per-square-meter basis. The staring dose in initial clinical trials is generally one-tenth of that required to produce severe toxicity in rodents (STD10) or one-tenth the highest dose in nonrodents that does not cause severe irreversible toxicity. While not required, information on pharmacokinetic parameters, especially data comparing the plasma concentration associated with toxicity in both species, is very highly regarded. Special attention is paid to organs with high cell division rates, bone marrow, testes, lymphoid tissue testing, and gastrointestinal tract. As these agents are almost always given intravenously, special attention needs to be given relatively early in development to intravenous (IV) irritation and blood compatibility study. Subsequent studies to support the NDA will be highly tailored, depending on the following:

- Therapeutic indication and mechanism of action
- Results of the initial clinical trials
- Nature of the toxicity
- Proposed clinical regimen

Even at the NDA stage, toxicity studies with more than 28 days of dosing are rarely required. While not required for the IND, assessment of genotoxicity and developmental toxicity will need to be addressed. For genotoxicity, it will

be important to establish the ratio between cytotoxicity and mutagenicity. In vivo models (e.g., the mouse micronucleus test) can be particularly important in demonstrating the lack of genotoxicity at otherwise subtoxic doses. For developmental toxicity, ICH stage C–D studies (traditionally known as segment II studies for teratogenicity in rat and rabbits) will also be necessary.

The emphasis of this discussion has been on purely cytotoxic neoplastic agents. Additional considerations must be given to cytotoxic agents that are administered under special circumstances: those that are photoactivated, delivered as liposomal emulsions, or delivered as antibody conjugates. These types of agents will require additional studies. For example, a liposomal agent will need to be compared to the free agent and a blank liposomal preparation. There are also studies that may be required for a particular class of agents. For example, anthracyclines are known to be cardiotoxic, so comparison of a new anthracycline agent to previously marketed anthracyclines will be expected.

In addition to antineoplastic, cytotoxic agents, there are cancer therapeutic or preventative drugs that are intended to be given on a chronic basis. This includes chemopreventatives, hormonal agents, and immunomodulators. The toxicity assessment studies on these will more closely resemble those of more traditional pharmaceutical agents. Chronic toxicity, carcinogenicity, and full developmental toxicity (ICH A–B, C–D, E–F) assessments will be required. For a more complete review, the reader is referred to DeGeorge et al. (1998).

### 7.3 Optical Isomers

The FDA [and like regulatory agencies, as reviewed by Daniels et al. (1997)] has become increasingly concerned with the safety of stereoisomeric or chiral drugs. Stereoisomers are molecules that are identical to one another in terms of atomic formula and covalent bonding but differ in the three-dimensional projections of the atoms. Within this class are those molecules that are non-superimposable mirror images of one another. These are called enantiomers (normally designated as R or S). Enantiomeric pairs of a molecule have identical physical and chemical characteristics except for the rotation of polarized light. Drugs have generally been mixtures of optical isomers (enantiomers) because of the difficulties in separating the isomers. It has become apparent in recent years, however, that these different isomers may have different degrees of both desirable therapeutic and undesirable toxicological effects. Technology has also improved to the extent that it is now possible to perform chiral specific syntheses, separations, and/or analyses. It is now highly desirable from a regulatory (FDA, 1988a–c; Anonymous, 1992a) basis to develop a single isomer unless all isomers have equivalent pharmacological and toxicological activity. The FDA has divided enantiomeric mixtures in the following categories:

- Both isomers have similar pharmacologic activity, which could be identical, or they could differ in the degrees of efficacy.
- One isomer is pharmacologically active, while the other is inactive.
- Each isomer has completely different activity.

During preclinical assessment of an enantiomeric mixture, it may be important to determine to which of these three classes it belongs. The pharmacological and toxicological properties of the individual isomers should be characterized. The pharmacokinetic profile of each isomer should be characterized in animal models with regard to disposition and interconversion. It is not at all unusual for each enantiomer to have a completely different pharmacokinetic behavior.

If the test article is an enantiomer isolated from a mixture that is already well characterized (e.g., already on the market), then appropriate bridging guides need to be performed which compare the toxicity of the isomer to that of the racemic mixture. The most common approach would be to conduct a subchronic (three-month) and a segment-II-type teratology study with an appropriate "positive" control group which received the racemate. In most instances no additional studies would be required if the enantiomer and the racemate did not differ in toxicity profile. If, on the other hand, differences are identified, then the reasons for this difference need to be investigated and the potential implications for human subjects need to be considered.

## 7.4 Special Populations: Pediatric and Geriatric Claims

Relatively few drugs marketed in the United States (approximately 20%) have pediatric dosing information available. Clinical trials had rarely been done specifically on pediatric patients. Traditionally, dosing regimens for children have been derived empirically by extrapolating on the basis of body weight or surface area. This approach assumes that the pediatric patient is a young adult, which simply may not be the case. There are many examples of how adults and children differ qualitatively in metabolic and/or pharmacodynamic responses to pharmaceutical agents. In their review, Schacter and DeSantis (1998, p. 300) state:

> The benefit of having appropriate usage information in the product label is that health care practitioners are given the information necessary to administer drugs and biologics in a manner that maximizes safety, minimizes unexpected adverse events, and optimizes treatment efficacy. Without specific knowledge of potential drug effects, children may be placed at risk. In addition, the absence of appropriate proscribing information, drugs and biologics that represent new therapeutic advances may not be administered to the pediatric population in a timely manner.

In response to the need for pediatric information, the FDA had developed a pediatric plan. This two-phase plan called first for the development of

pediatric information on marketed drugs. The second phase focused on new drugs. The implementation of the plan was to be coordinated by the Pediatric Subcommittee of the Medical Policy Coordinating Committee of CDER. The Pediatric Use Labeling Rule was a direct result of phase 1 in 1994 (Anonymous, 1998b). Phase 2 resulted in 1997 from a proposed rule entitled "Pediatric Patients; Regulations Requiring Manufacturers to Assess the Safety and Effectiveness of New Drugs and Biologics." Soon after this rule was proposed, the FDA Modernization Act of 1997 was passed. FDAMA contained provisions that specifically addressed the needs and requirements for the development of drugs for the pediatric population.

The FDAMA bill essentially codified and expanded several regulatory actions initiated by the FDA during the 1990s. Among the incentives offered by the bill, companies will be offered an additional six months of patent protection for performing pediatric studies (clinical trials) on already approved products. In fact, the FDA was mandated by FDAMA to develop a list of over 500 drugs for which additional information would produce benefits for pediatric patients. The FDA is supposed to provide a written request for pediatric studies to the manufacturers (Hart, 1999).

In response to the pediatric initiatives, the FDA has published policies and guidelines and conducted a variety of meetings. CDER has established a website (http//www.fda.gov/cder/pediatric) which lists three pages of such information. Interestingly, the focus has been on clinical trials, and almost no attention has been given to the preclinical toxicology studies that may be necessary to support such trials. There are three pages of documents on the pediatric website. None appear to address the issue of appropriate testing. This situation is just now being addressed and is in a great deal of flux.

In the absence of any guidelines from the agency for testing drugs in young or "pediatric" animals, one must fall back on the maxim of designing a program that makes the most scientific sense. As a guide, the FDA-designated levels of postnatal human development and the approximate equivalent ages (in the author's considered opinion) in various animal models are given in Table 9. The table is somewhat inaccurate, however, because of differences in the stages of development at birth. A rat is born quite underdeveloped when compared to a human being. A one-day-old rat is not equivalent to a one-day-old full-term human infant. A four-day-old rat would be more appropriate. In terms of development, the pig may be the best model of those listed; however, one should bear in mind that different organs have different developmental schedules in different species.

Table 9 can be used as a rough guide in designing toxicity assessment experiments in developing animals. In designing the treatment period, one needs to consider not only the dose and proposed course of clinical treatment but also the proposed age of the patient and whether or not an equivalent dosing period in the selected animal model covers more than one developmental stage. For example, if the proposed patient population is human infants, initiating a toxicity study of the new pharmaceutical agent in

**TABLE 9 Comparison of Postnatal Development Stages**

| Stage | Human | Rat | Dog | Pig |
|---|---|---|---|---|
| Neonate | Birth to 1 month | Birth to 1 week | Birth to 3 weeks | Birth to 2 weeks |
| Infant | 1 Month to 2 years | 1–3 Weeks | 3–6 Weeks | 2–4 Weeks |
| Child | 2 Years to 12 years | 3–9 Weeks | 6 Weeks–5 months | 4 Weeks–4 months |
| Adolescent | 12 Years to 16 years | 9–13 Weeks | 5 Moths–9 months | 4–7 Months |
| Adult | Over 16 years | Over 13 weeks | Over 9 months | Over 7 months |

three-day-old rats is not appropriate. Furthermore, if the proposed course of treatment in adult children is two weeks, it is unlikely that this would cross over into a different developmental stage. A two-week treatment initiated in puppies, however, might easily span two developmental stages. Thus, in designing an experiment in young animals one must carefully consider the length of the treatment period balancing the developmental age of the animal model and the proposed length of clinical treatment. Where appropriate (infant animals), one needs to also assess changes in standard developmental landmarks (e.g., eye opening, pinnae eruption, external genitalia development) as well as the more standard indicators of target organ toxicity. The need for maintaining the experimental animals past the dosing period, perhaps into sexual maturity, to assess recovery or delayed effects needs also to be carefully considered.

To summarize, the current status of assessment of toxicity in postnatal mammals, in response to the pediatric initiatives covered in FDAMA, is an extremely fluid situation. One needs to carefully consider a variety of factors in designing the study and should discuss proposed testing programs with the appropriate office at CDER.

Drugs intended for use in the elderly, like those intended for the very young, may also have special requirements for safety evaluation, but geriatric issues were not addressed in the FDAMA of 1997. The FDA has published a separate guidance document for geriatric labeling. As was the case with pediatric guidance, this document does not address preclinical testing. With the elderly, the toxicological concerns are quite different than the developmental concerns associated with pediatric patients. With the elderly, one must be concerned with the possible interactions between the test article and compromised organ function. The FDA had previously issued a guidance for clinically examining the clinical safety of new pharmaceutical agents in patients with compromised renal and/or hepatic function (CDER, 1989). The equivalent ICH guideline (S5A) was issued in 1994. Whether this type of emphasis will require toxicity testing in animal models with specifically induced organ insufficiency remains to be seen. In the interim, we must realize that there is

tacit evaluation of test article–related toxicity in geriatric rodents for those agents that undergo two-year carcinogenicity testing. As the graying of America continues, labeling for geriatric use may become more of an issue in the future.

### 7.5 Orphan Drugs

The development of sophisticated technologies coupled with the rigors and time required for clinical and preclinical testing has made pharmaceutical development very expensive. In order to recoup such expenses, pharmaceutical companies have tended to focus on therapeutic agents with large potential markets. Treatment for rare but life-threatening diseases have been "orphaned" as a result. An orphan product is defined as one targeted at a disease which affects 200,000 or fewer individuals in the United States. Alternatively, the therapy may be targeted for more than 200,000 but the developer would have no hope of recovering the initial investment without exclusivity. The Orphan Drug Act (ODA) of 1983 was passed in an attempt to address this state of affairs. Currently applicable regulations were put in place in 1992 (Anonymous, 1992b). In 1994, there was an attempt in Congress to amend the act, but it failed to be passed into law. The current regulations are administered by the office of Orphan Product Development (OPD). The act offers the following incentives to encourage the development of products to treat rare diseases:

- Seven years exclusive market following the approval of a product for an orphan disease
- Written protocol assistance from the FDA
- Tax credits for up 50% of qualified clinical research expenses
- Available grant to support pivotal clinical trials

As reviewed by Haffner (1998), other developed countries have similar regulations.

The ODA did not change the requirements of testing drug products. The nonclinical testing programs are similar to those used for more conventional products. They will undergo the same FDA review process. A major difference, however, is the involvement of the OPD. A sponsor must request OPD review. Once the OPD determines that a drug meets the criteria for orphan drug status, it will work with the sponsor to provide the assistance required under the act. The ODA does not review a product for approval. The IND/NDA process is still handled by the appropriate reviewing division (e.g., cardiovascular) for formal review. The act does not waive the necessity for submission of an IND, not for the responsibility of toxicological assessment. As always, in cases where there is ambiguity, a sponsor may be well served to request a pre-IND meeting at the appropriate division to discuss the acceptability of a toxicology assessment plan.

### 7.6 Botanical Drug Products

There is an old saying, "What goes around, comes around": and so it is with botanicals. At the beginning of the twentieth century, most marketed pharmaceutical agents were botanical in origin. For example, aspirin was first isolated from willow bark. These led the way to modern drug development in the middle part of the century, for reasons having to do with patentability, manufacturing costs, standardization, selectivity, and potency. The dawning of the twenty-first century has seen a grass-roots return to botanical preparations (also sold as herbals or dietary supplements). These preparations are being marketed to the lay public as "natural" supplements to the nasty synthetic chemicals now proscribed as pharmaceutical products. In 1994, the Dietary Supplement Health and Education Act was passed which permitted the marketing of dietary supplements (including botanicals) with limited submissions to the FDA (Wu et al., 2000). If a producer makes a claim that an herbal preparation is beneficial to a specific part of the body (e.g., enhanced memory), then it may be marketed after a 75-day period of FDA review but without formal approval. On the other hand, if any curative properties are claimed, then the botanical will be regulated as a drug and producers will be required to follow the IND/NDA process. In 1997 and 1998 combined, some 26 INDs were filed for botanical products (Wu et al., 2000).

The weakness in the current regulation has to do with its ambiguity. The line between a beneficial claim and a curative claim is sometimes difficult to draw. What is the difference, for example, between an agent that enhances memory and one that prevents memory loss? Given the number of products and claims hitting the shelves every day, this situation will probably demand increased regulatory scrutiny in the future.

## 8 INTERNATIONAL PHARMACEUTICAL REGULATION AND REGISTRATION

### 8.1 International Conference on Harmonisation

The International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use was established to make the drug regulatory process more efficient in the United States, Europe, and Japan. The U.S. involvement grew out of the fact that the United States is party to the General Agreement on Tariffs and Trade, which included the Agreement on Technical Barriers to Trade, negotiated in the 1970s, to encourage reduction of nontariff barriers to trade (Barton, 1998). The main purpose of ICH is, through harmonization, to make new medicines available to patients with a minimum of delay. More recently, the need to harmonize regulation has been driven, according to ICH, by the escalation of the cost of R&D. The regulatory systems in all countries have the same fundamental concerns about safety, efficacy, and quality, yet sponsors had to repeat many time-consuming

**TABLE 10 ICH Representation**

| Country/Region | Regulatory | Industry |
|---|---|---|
| European Union | European Commission (2) | European Federation of Pharmaceutical Industries Associations (2) |
| Japan | Ministry of Health and Welfare (2) | Japanese Pharmaceutical Manufactures Association (2) |
| United States | Food and Drug Administration (2) | Pharmaceutical Research and Manufacturers of America (2) |
| Observing organizations | World Health Organization, European Free Trade Area, Canadian Health Protection Branch | International Federation of Pharmaceutical Manufactures Associations (2): also provides the secretariat |

*Note*: Numbers in parentheses are number of representatives on the ICH steering committee.

and expensive technical tests to meet country-specific requirements. In addition, there was a legitimate concern over the unnecessary use of animals. Conference participants include representatives from the drug regulatory bodies and research-based pharmaceutical industrial organizations of three regions; the European Union, the United States, and Japan comprised over 90% of the world's pharmaceutical industry. Representation is summarized in Table 10. The biannual conference has met four times, beginning in 1991, rotating between sites in the United States, Europe, and Japan. The next meeting is scheduled for the year 2001 and will be held on the West Coast of the United States. The precise venue has yet to be named.

The ICH meets its objectives by issuing guidelines for the manufacturing, development, and testing of new pharmaceutical agents that are acceptable to all three major parties. For each new guideline, the ICH steering committee establishes an expert working group with representation from each of the six major participatory ICH bodies. Each new draft guideline goes through the five various steps of review and revision, summarized in Table 11. So far, the ICH has proposed or adopted over 40 safety, efficacy, and quality guidelines (listed in Table 12) for use by drug regulatory agencies in the United States, Europe, and Japan. The guidelines are organized under broad categories: the E series having to do with clinical trials, the Q series having to do with quality (including chemical manufacturing and control as wells as traditional GLP issues), and the S series having to do with safety. Guidelines may be obtained from the ICH secretariat, IFPMA, 30 rue de St.-Jean, PO Box 9, 1211 Geneva 18, Switzerland, or may be downloaded from a website set up by Nancy McClure (http://www.mcclurenet.com/index.html). They are also published in the *Federal Register*. It is the guidelines of the S series that will have the most impact on toxicologists. The biggest changes, having to do with toxicological assessment, are summarized as follows:

**TABLE 11 Steps in ICH Guideline Development and Implementation**

| |
|---|
| 1. Building scientific consensus in joint regulatory/industry expert working groups |
| 2. Agreement by the steering committee to release the draft consensus text for wider consultation |
| 3. Regulatory consultation in the three regions; consolidation of the comments |
| 4. Agreement on a harmonized ICH guideline; adopted by the regulators |
| 5. Implementation in the three ICH regions |

1. *Carcinogenicity Studies* Carcinogenicity studies are covered in guidelines S1A, S1B, and S1C. The guidelines are almost more philosophical than they are technical. In comparison to the EPA guidelines, for example, the ICH guidelines contain little in the way of concrete study criteria (e.g., the number of animals, the necessity for clinical chemistry). There is discussion on when carcinogenicity studies should be done, whether two species are more appropriate than one, and how to set dosages on the basis of human clinical pharmacokinetic (PK) data. The major changes wrought by these guidelines are:

- Only one two-year carcinogenicity study should be generally required. Ideally, the species chosen should be the one most like humans in terms of metabolic transformations of the test article.
- The traditional second long-term carcinogenicity study can be replaced by a shorter term alternative model. In practical terms, this guideline is beginning to result in sponsors conducting a two-year study in the rat and a six-month study in an alternative mouse model, such as the P53 or the TG.AC genetically manipulated mouse strains.
- In the absence of target organ toxicity with which to set the high dose at the maximally tolerated dose, the high dose can be set at the dose that produces an area under the curve (AUC). This is 25-fold higher than that obtained in human subjects.

2. *Chronic Toxicity* Traditionally, chronic toxicity of new pharmaceuticals in the United States was assessed in studies of one year duration in both rodent and nonrodent species of choice. The European view was that studies of six months are generally sufficient. The resulting guideline (S4A) was a compromise. Studies of six months duration were recommended for the rodent, as rodents would also be examined in two-year studies. For the nonrodent (dog, nonhuman primate, and pig) studies of nine months duration were recommended.

3. *Developmental and Reproductive Toxicity* This was an area in which there was considerable international disagreement and the area in which ICH has promulgated the most technically detailed guidelines (S5A and S5B). Some of the major changes include:

- The traditional segment I, II, and III nomenclature has been replaced with different nomenclature, as summarized in Table 13.

**TABLE 12 International Conference on Harmonisation Guidelines**

| References | Guideline | Date |
|---|---|---|
| E1 | The Extent of Population Exposure to Assess Clinical Safety | Oct. 94 |
| E2A | Clinical Safety Data Management: Definitions and Standards for Expedited Reporting | Oct. 94 |
| E2B | Clinical Safety Data Management: Data Elements for Transmission of Individual Case Safety Reports | May 05 |
| E2C | Clinical Safety Data Management: Periodic Safety Update Reports for Marketed Drugs | May 97 |
| E2D | Definitions and Standards for Expedited Reporting | Nov. 03 |
| E2E | Pharmocovigilance Planning | Nov. 04 |
| E3 | Structure and Content of Clinical Study Reports | Nov. 95 |
| E4 | Dose Response Information to Support Drug Registration | Mar. 94 |
| E5 | Ethnic Factors in the Acceptability of Foreign Clinical Data | Feb. 98 |
| E6 | Good Clinical Practice: Consolidated Guideline; Notice of Availability | May 96 |
| E6A | GCP Addendum on Investigator's Brochure | Mar. 95 |
| E7 | Studies in Support of Special Populations: Geriatrics | June 93 |
| E8 | Guidance on General Considerations for Clinical Trials; Notice | July 97 |
| E9 | Draft Guideline on Statistical Principles for Clinical Trials; Notice of Availability | Feb. 98 |
| E10 | Choice of Control Group and Related Issues in Clinical Trials | July 00 |
| E11 | Clinical Investigation of Medicinal Products in the Pediatric Population | July 00 |
| E12 | Principles for Clinical Evaluation of New Antihypertensive Drugs | |
| E14 | The Clinical Evaluation of QT/QTc Interval Prolongation and Proarrhythmic Potenial for Non-Antiarrhythmic Drugs | May 05 |
| E15 | Definitions for Genomic Biomakers, Pharmacogenomics, Pharmacogenetics, Genomic Data and Sample Coding Categories | Nov. 07 |
| M3 | Guidance on Nonclinical Safety Studies for the Conduct of Human Clinical Trials for Pharmaceuticals; Notice | Nov. 97 |
| Q1A | Stability Testing of New Drug Substances and Products | Feb. 03 |
| Q1B | Stability Testing of New Drug Substances and Products | Nov. 96 |
| Q1C | Stability Testing for New Dosage Forms | Nov. 96 |
| Q1D | Bracketing and Matrixing Designs for Stability Testing of Drug Substances and Drug Products | Feb. 02 |
| Q1E | Evaluation of Stability Data | Feb. 03 |
| Q1F | Stability Data Package for Registration Applications in Climatic Zones III and IV | June 06 |
| Q2 | Validation of Analytical Procedures: Text and Methodology | Oct. 94 |
| Q3A | Guideline on Impurities in New Drug Substances | Oct. 06 |
| Q3B | Guideline on Impurities in New Drug Products | June 06 |
| Q3C | Guideline on Impurities: Guideline for Residual Solvents | July 97 |
| Q4 | Pharmacopeias | Nov. 07 |
| Q4A | Pharmacopeias Harmonisation | Nov. 07 |
| Q4B | Evaluation and Recommendation of Pharmacopoeial Texts | Nov. 07 |
| Q4B Annex 1 | Evaluation and Recommendation of Pharmacopoeial Texts: Residue on Ignition/Sulphated Ash General Chapter | Nov. 07 |
| Q4B Annex 2 | *Evaluation and Recommendation of Pharmacopoeial Texts: Test for Extractable Volume of Parenteral Preparations General Chapter* | Nov. 07 |

**TABLE 12** ***Continued***

| References | Guideline | Date |
|---|---|---|
| Q4B Annex3 | *Evaluation and Recommendation of Pharmacopoeial Texts: Test for Particulate Contamination: Sub-Visible Particles General Chapter* | Nov. 07 |
| Q5A | Quality of Biotechnological Products Viral Safety Evaluation of Biotechnology Products Derived from Cell Lines of Human or Animal Origin | Mar. 97 |
| Q5B | Quality of Biotechnology Products Analysis of the Expression Construct in Cells Used for Production of r-DNA Derived Protein Product | Nov. 95 |
| Q5C | Quality of Biotechnological Products: Stability Testing of Biotechnological/Biology Products | Nov. 95 |
| Q5D | Availability of Draft Guideline on Quality of Biotechnological/ Biological Products: Derivation and Characterization of Cell Substrates Used for Production of Biotechnological/Biological Products | July 97 |
| Q5E | *Comparability of Biotechnological/Biological Products Subject to Changes in Their Manufacturing Process* | Nov. 04 |
| Q6A | *Specifications: Test Procedures and Acceptance Criteria for New Drug Substances and New Drug Products: Chemical Substances (including Decision Trees)* | Oct. 99 |
| Q6B | *Specifications: Test Procedures and Acceptance Criteria for Biotechnological/Biological Products* | Mar. 99 |
| Q7 | *Good Manufacturing Practice Guide for Active Pharmaceutical Ingredients* | Nov. 00 |
| Q8 | *Pharmaceutical Development* | Nov. 05 |
| Annex to Q8 | *Pharmaceutical Development Annex* | Nov. 07 |
| Q9 | *Quality Risk Management* | Nov. 05 |
| Q10 | *Pharmaceutical Quality System* | May 07 |
| Q6A | Draft Guidance on Specifications: Test Procedures and Acceptance Criteria for New Drug Substances and New Drug Products: Chemical Substances; Notice | Nov. 97 |
| Q6B | Specifications: Test Procedures and Acceptance Criteria for Biotechnology Products | Feb. 98 |
| S1A | Guidance on the Need for Carcinogenicity Studies of Pharmaceuticals | Nov. 95 |
| S1B | Draft Guideline on Testing for Carcinogenicity of Pharmaceuticals; Notice | July 97 |
| S1C | Dose Selection for Carcinogenicity Studies of Pharmaceuticals | Oct. 94 |
| S1Ca | Guidance on Dose Selection for Carcinogenicity Studies of Pharmaceuticals: Addendum on a Limit Dose and Related Notes; Availability; Notice | Dec. 97 |
| S2A | Genotoxicity: Guidance on Specific Aspects of Regulatory Genotoxicity Tests for Pharmaceuticals | July 95 |
| S2B | Guidance on Genotoxicity: A Standard Battery for Genotoxicity Testing of Pharmaceuticals; Availability; Notice | July 97 |
| S3A | Toxicokinetics: Guidance on the Assessment of Systemic Exposure in Toxicity Studies | Oct. 94 |
| S3B | Pharmacokinetics: Guidance for Repeated Dose Tissue Distribution Studies | Oct. 94 |
| S4 | Single Dose Acute Toxicity Testing for Pharmaceuticals; Revised Guidance; Availability; Notice | Sept. 98 |

**TABLE 12** ***Continued***

| References | Guideline | Date |
|---|---|---|
| S4A | Draft Guidance on the Duration of Chronic Toxicity Testing in Animals (Rodent and Nonrodent Toxicity Testing); Availability; Notice | Nov. 97 |
| S5A | Detection of Toxicity to Reproduction for Medicinal Products | Nov. 95 |
| S5B | Reproductive Toxicity to Male Fertility | Nov. 95 |
| S6A | Guidance on Preclinical Safety Evaluation of Biotechnology-Derived Pharmaceuticals; Availability | July 97 |
| S7A | *Safety Pharmacology Studies for Human Pharmaceuticals* | Nov. 00 |
| S7B | *The Non-Clinical Evaluation of the Potential for Delayed Ventricular Repolarization (QT Interval Prolongation) by Human Pharmaceuticals* | May 05 |
| S8 | *Immunotoxicity Studies for Human Pharmaceuticals* | Sept. 05 |

- The dosing period of the pregnant animals during studies on embryonic development (traditional segment II studies) has been standardized.
- New guidelines for fertility assessment (traditional segment I) studies have shortened the premating dosing schedule (e.g., in male rats from 10 to 4 weeks). There has been an increased interest in assessment of spermatogenesis and sperm function.
- The new guidelines allow for a combination of studies in which the endpoints typically assessed in the traditional segment II and III studies are now examined under a single protocol.

For a more complete review of the various study designs, the reader is referred to Manson (1994).

While not quite as sweeping in approach as the aforementioned guidelines, a toxicologist working in pharmaceutical safety assessment should become familiar with all the other ICH guidelines in the S series.

In an interesting recent article, Ohno (1998) discussed not just the harmonization of nonclinical guidelines but also the need to harmonize the timing of nonclinical tests in relation to the conduct of clinical trials. For example, there are regional differences in the inclusion of women of childbearing potential in clinical trials. In the United States, including women in such trials is becoming more important, and therefore evaluation of embryo–fetal development will occur earlier in the drug development process than in Japan. Whether or not such timing or staging of nonclinical tests becomes part of an ICH guideline in the near future remains to be established.

### 8.2 Other International Considerations

The United States is the single largest pharmaceutical market in the world. But the rest of the world [particularly but not limited to the second and third

**TABLE 13 Comparison of Traditional and ICH Guidelines for Reproductive and Developmental Toxicology**

| Traditional Protocol | Stages Covered | ICH Protocol | Dosing Regimen |
|---|---|---|---|
| Segment I (rats) | A. Premating to conception<br>B. Conception to implantation | Fertility and early embryonic development, including implantation | Males: 4 weeks premating, mating (1–3 weeks) plus 3 weeks postmating<br>Females: 2 weeks premating, mating through day 7 of gestation |
| Segment II (rabbits) | C. Implantation to closure of hard palate<br>D. Closure of hard palate to end of pregnancy | Embryo–fetal development | Female rabbits: day 6 to day 20 of pregnancy |
| **Study Title** | **Termination** | **Endpoints: In Life** | **Endpoints: Postmortem** |
| Fertility and early embryonic development, including implantation | Females: Days 13–15 of pregnancy<br>Males: Day after completion of dosing | Clinical signs and mortality<br>Body weights and feed intake<br>Vaginal cytology | Macroscopic examination plus histology on gross lesions<br>Collection of reproductive organs for possible histology<br>Quantitation of corpora lutea and implantation sites<br>Seminology (count, motility, and morphology) |
| Embryo–fetal development | | Clinical signs and mortality<br>Body weights and changes<br>Feed intake | Macroscopic examination plus histology on gross lesions<br>Quantitation of corpora lutea and implantation sites<br>Fetal body weights<br>Fetal abnormalities |
| Pre- and postnatal development, including maternal function | | Clinical signs and mortality<br>Body weights and changes<br>Feed intake<br>Duration of pregnancy<br>Parturition | Macroscopic examination plus histology on gross lesions<br>Implantation<br>Abnormalities (including terata)<br>Live/dead offspring at birth<br>Pre- and postweaning survival and growth ($F_1$)<br>Physical development ($F_1$)<br>Sensory functions and reflexes ($F_1$)<br>Behavior ($F_1$) |

largest markets, Japan and the European Union (EU)] represents in aggregate a much larger market, so no one develops a new pharmaceutical for marketing in just the United States. The effort at harmonization (exemplified by the ICH) has significantly reduced differences in requirements for these other countries but certainly not obliterated them. Though a detailed understanding of their regulatory schemes is beyond this volume, the bare bones and differences in toxicology requirements are not.

***European Union*** The standard EU toxicology and pharmacologic data requirements for a pharmaceutical include:

Single-dose toxicity
Repeat-dose toxicity (subacute and chronic trials)
Reproduction studies (fertility and general reproductive performance, embryotoxicity, and peri/postnatal toxicity)
Mutagenic potential (in vitro and in vivo)
Carcinogenicity
Pharmacodynamics
- Effects related to proposed drug indication
- General pharmacodynamics
- Drug interactions

Pharmacokinetics
- Single dose
- Repeat dose
- Distribution in normal and pregnant animals
- Biotransformation

Local tissue tolerance
Environmental toxicity

In general, the registration process in the EU allows one apply to either an overall medicines authority or an individual national authority. Either step is supposed to lead to mutual recognition by all members.

***Japan*** In Japan, the Koseisho is the national regulatory body for new drug. The standard median lethal dose ($LD_{50}$) test is no longer a regulatory requirement for new medicines in the United States, EU, or Japan. The Japanese guidelines were the first to be amended in accordance with this agreement, with the revised guidelines becoming effective in August 1993. The Japanese may still anticipate that single-dose (acute) toxicity studies should be conducted in at least two species, one rodent and one nonrodent (the rabbit is not accepted as a nonrodent). Both males and females should be included from at least one of the species selected: if the rodent, then a minimum of five per sex; if the nonrodent, at least two per sex. In nonrodents, both the oral

and parenteral routes should be used, and normally the clinical route of administration should be employed. In nonrodents, only the intended route of administration need be employed; if the intended route of administration in humans is intravenous, then use of this route in both species is acceptable. An appropriate number of doses should be employed to obtain a complete toxicity profile and to establish any dose–response relationship. The severity, onset, progression, and reversibility of toxicity should be studied during a 14-day follow-up period, with all animals being necropsied. When macroscopic changes are noted, the tissue must be subjected to histological examination.

Chronic and subchronic toxicity studies are conducted to define the dose level, when given repeatedly, that cause toxicity and the dose level that does not lead to toxic findings. In Japan, such studies are referred to as repeated-dose toxicity studies. As with single-dose studies, at least two animal species should be used, one rodent and one nonrodent (rabbit not acceptable). In rodent studies, each group should consist of at least 10 males and 10 females; in nonrodent species, 3 of each sex is deemed adequate. Where interim examinations are planned, however, the number of animals employed should be increased accordingly. The planned route of administration in human subjects is normally explored. The duration of the study will be dictated by the planned duration of clinical use (Table 14).

At least three different dose groups should be included, with the goals of demonstrating an overtly toxic dose and a no-effect dose and establishing any dose–response relationship. The establishment of a nontoxic dose within the framework of these studies is more rigorously adhered to in Japan than elsewhere in the world. All surviving animals should also be necropsied, either at the completion of the study or during its extension recovery period, to assess reversal of toxicity and the possible appearance of delayed toxicity. Full histological examination is mandated on all nonrodent animals used in a chronic toxicity study; at a minimum, the highest dose and control groups of rodents must be submitted to a full histological examination.

**TABLE 14 Required Duration of Dosing in Nonclinical Study to Support Clinical Dosing**

| Duration of Dosing in Toxicity Study | Duration of Human Exposure |
|---|---|
| 1 Month | Single dose or repeated dosage not exceeding 1 week |
| 3 Months | Repeated dosing exceeding 1 week and to a maximum of 4 weeks |
| 6 Months | Repeated dosing exceeding 4 weeks and to a maximum of 6 months |
| 12 Months[a] | Repeated dosing exceeding 6 months or where this is deemed to be appropriate |

[a]Where carcinogenicity studies are to be conducted, the Koseisho had agreed to forego chronic dosage beyond 6 months. *Source*: New Drugs Division Notification No. 43, June 1992.

While the value of repeated-dose testing beyond 6 months has been questioned (Lumley and Walker, 1992), such testing is a regulatory requirement for a number of agencies, including the FDA and the Koseisho. In Japan, repeated-dose testing for 12 months is required only for new medicines expected to be administered to humans for periods in excess of 6 months (Yakuji Nippo, 1994). At the first ICH held in Brussels, the consensus was that 12-month toxicity studies in rodents could be reduced to 6 months where carcinogenicity studies are required. While not yet adopted in the Japanese guidelines, 6-month repeated-dose toxicity studies have been accepted by the agencies of all three regions. Japan—like the EU—accepts 6 months duration if accompanied by a carcinogenicity study. The United States still requires a 9-month nonrodent study.

With regard to reproductive toxicology, as a consequence of the first ICH, the United States, EU, and Japan agreed to recommend mutual recognition of their respective current guidelines. A tripartite, harmonized guideline on reproductive toxicology has achieved ICH step 4 status and should soon be incorporated into the local regulations of all three regions. This agreement represents a very significant achievement that should eliminate many obstacles to drug registration.

*Preclinical Male Fertility Studies* Before conducting a single-dose male volunteer study in Japan, it is usually necessary to have completed a preclinical male fertility study (segment I) that has an in-life phase of 10 or more weeks (i.e., 10 weeks of dosing plus follow-up). Although government guidelines do not require this study to be completed before phase I trials begin, the responsible institutional review board or the investigator usually imposes this condition. Japanese regulatory authorities are aware that the segment I male fertility study is of poor predictive value. The rat, which is used in this study, produces a marked excess of sperm. Many scientists therefore believe that the test is less sensitive than the evaluation of testicular weight and histology that constitute part of the routine toxicology assessment

*Female Reproductive Studies* Before entering a female into a clinical study, it is necessary to have completed the entire reproductive toxicology program, which consists of the following studies:

- *Segment I* Fertility studies in the rat or mouse species used in the segment II program
- *Segment II* Teratology studies in the rat or mouse and the rabbit
- *Segment III* Late gestation and lactation studies in a species used in the segment II studies

Such studies usually take approximately two years. Although the U.S. regulations state the need for completion of segments I and II and the demonstration of efficacy in male patients, where appropriate, before entering females into a clinical program, the current trend in the United States is toward

relaxation of the requirements to encourage investigation of the drug both earlier and in a larger number of females during product development. Growing pressure for the earlier inclusion of women in drug testing may encourage selection of this issue as a future ICH topic. The trend in the United States and the EU toward including women earlier in the critical program has not yet been embraced in Japan, however.

The three tests required in Japan for genotoxicity evaluation are a bacterial gene mutation test, in vitro cytogenetics, and in vivo tests for genetic damage. The Japanese regulations state these tests to be the minimum requirement and encourage additional tests. Currently, Japanese guidelines do not require a mammalian cell gene mutation assay. Harmonization will likely be achieved by the Koseisho recommending all four tests, which will match requirements in the United States and the EU; at present, this topic is at step 1 in the ICH process. The mutagenicity studies should be completed before the commencement of phase II clinical studies.

Guidelines presented at the second ICH are likely to alter the preclinical requirements for registration in Japan; they cover toxicokinetics and when to conduct repeated-dose tissue distribution studies. The former document may improve the ability of animal toxicology studies to predict possible adverse events in humans; currently, there are not toxicokinetic requirements in Japan, and their relevance is questioned by many there. Although there is general agreement on the registration requirement for single-dose tissue distribution studies, implementation of the repeated-dose study requirement has been inconsistent across the three ICH parties.

## 8.3 Safety Pharmacology

Japan was the first major country to require extensive pharmacological profiling on all new pharmaceutical agents as part of the safety assessment profile. Prior to commencement of initial clinical studies, the drug's pharmacology must be characterized in animal models. In the United States and Europe, these studies have been collectively called safety pharmacology studies. For a good general review of the issues surrounding safety pharmacology the reader is referred to Hite (1997). The Japanese guidelines for such characterizations were published in 1991 (New Drugs Division Notification No. 4, January 1991):

- Effects on general activity and behavior
- Effects on the central nervous system
- Effects on the autonomic nervous system and smooth muscle
- Effects on the respiratory and cardiovascular systems
- Effects on the digestive system
- Effects on water and electrolyte metabolism
- Other important pharmacological effects

In the United States, pharmacological studies in demonstration of efficacy have always been required, but specific safety pharmacological studies have never been required. Special situational or mechanistic data would be requested on a case-by-case basis. This is a situation that is changing. In the United States the activities of the Safety Pharmacology Discussion Group, for example, have helped bring attention to the utility and issues surrounding safety pharmacology data. In 1999 and 2000, the major toxicological and pharmacological societal meetings had symposia on safety pharmacological testing. Many major U.S. pharmaceutical companies are in the process of implementing programs in safety pharmacology. The issue has been taken up by ICH and the draft guideline is currently at the initial stages of review. This initial draft (guideline S7) includes core tests in the assessment of CNS, cardiovascular, and respiratory function. Studies will be expected to be performed under GLP guidelines.

## 9 COMBINATION PRODUCTS

Recent years have seen a vast increase in the number of new therapeutic products which are not purely drug, device, or biologic but rather a combination of two or more of these. This leads to a problem of deciding which of the three centers shall have ultimate jurisdiction.

The Center for Devices and Radiological Health (CDRH) is designated the center for major policy development and for the promulgation and interpretation of procedural regulations for medical devices under the act. The CDRH regulates all medical devices, inclusive of radiation-related device, that are not assigned categorically or specifically to CDER. In addition, CDRH will independently administer the following activities (references to "sections" are the provisions of the act):

1. A. Small business assistance programs under Section 10 of the amendments [see Public Law (PL) 94-295]. Both CDER and CDRH will identify any unique problems relating to medical device regulation for small business.
   B. Registration and listing under Section 510, including some CDER-administered device applications. The CDER will receive printouts and other assistance, as requested.
   C. Color additives under Section 706, with review by CDER, as appropriate.
   D. Good Manufacturing Practices Advisory Committee. Under Section 520(f)(3), CDER will regularly receive notices of all meetings, with participation by CDER, as appropriate.
   E. Medical device reporting—The manufacturers, distributors, importers, and users of all devices, including those regulated by CDER, shall

report to CDRH under Section 519 of the act as required. The CDRH will provide monthly reports and special reports as needed to CDER for investigation and follow-up of those medical devices regulated by CDER.

### 9.1 Device Programs That CDER and CDRH Each Will Administer

Both CDER and CDRH will administer and, as appropriate, enforce the following activities for medical devices assigned to their respective centers (references to "sections" are the provisions of the act):

1. A. Surveillance and compliance actions involving general controls violations, such as misbranded or adulterated devices under Sections 301, 501, and 502
   B. Warning letters, seizures, injunctions, and prosecutions under Sections 302, 303, and 304
   C. Civil penalties under Section 303(f) and administrative restraint under Section 304(g)
   D. Nonregulatory activities, such as educational programs directed at users, participation in voluntary standards organizations, and so on
   E. Promulgation of performance standards and applications of special controls under Section 514
   F. Premarket notification, investigational device exemptions including humanitarian exemptions, premarket approval, product development protocols, classification, device tracking, petitions for reclassification, postmarket surveillance under Sections 510(k), 513, 515, 519, 520(g) and (m), and 522, and the advisory committees necessary to support these activities
   G. Banned devices under Section 516
   H. FDA-requested and firm-initiated recalls whether under Section 518 or another authority and other Section 518 remedies such as recall orders
   I. Exemptions, variances, and applications of CGMP regulations under Section 520(f)
   J. Governmentwide quality assurance program
   K. Requests for export approval under Sections 801(e) and 802

***Coordination*** The centers will coordinate their activities in order to assure that manufacturers do not have to independently secure authorization to market their product from both centers unless this requirement is specified in Section VII.

***Submissions*** Submissions should be made to the appropriate center, as specified herein, at the following addresses: Food and Drug Administration, Center for Drug Evaluation and Research, Central Document Room (Room 2-14), 12420 Parklawn Drive, Rockville, Maryland 20852 or Food and Drug

Administration, Center for Devices and Radiological Health, Document Mail Center (HFZ-401), 1390 Piccard Drive, Rockville, Maryland 20850.

For submissions involving medical devices and/or drugs that are not clearly addressed in this agreement, sponsors are referred to the product jurisdiction regulations (21 CFR Part 3). These regulations have been promulgated to facilitate the determination of regulatory jurisdiction but do not exclude the possibility for a collaborative review between the centers.

## 9.2 Center Jurisdiction

The following subsections provide details concerning status, market approval authority, special label/regulatory considerations, investigational options, and intercenter consultations for the categories of products specified. Section VII provides the general criteria that CDRH and CDER will apply in reaching decisions as to which center will regulate a product.

A. 1. (a) Device with primary purpose of delivering or aiding in the delivery of a drug that is distributed without a drug (i.e., unfilled)

*Examples*

Devices that calculate drug dosages

Drug delivery pump and/or catheter infusion pump for implantation iontophoresis device

Medical or surgical kit (e.g., tray) with reference in instructions for use with specific drug (e.g., local anesthetic)

Nebulizer

Small-particle aerosol generator (SPAG) for administering drug to ventilated patient

Splitter block for mixing nitrous oxide and oxygen

Syringe, jet injector, storage and dispensing equipment

*Status* Device and drug as separate entities

*Market Approval Authority* CDRH and CDER, respectively, unless the intended use of the two products, through labeling, creates a combination product

*Special Label/Regulatory Considerations* The following specific procedures will apply depending on the status of the drug delivery device and drugs that will be delivered with the device:

(i) It may be determined during the design or conduct of clinical trials for a new drug that it is not possible to develop adequate performance specifications data on those characteristics of the device that are required for the safe and effective use of the drug. If this is the case, then drug labeling cannot be written to contain information that makes it possible for the user to substitute a generic, marketed device for the device used during development to use with the marketed drug. In these situations, CDER will be the lead center for regulation of the device under the device authorities.

(ii) For a device intended for use with a category of drugs that are on the market, CDRH will be the lead center for regulation for the device under the device authorities. The effects of the device use on drug stability must be addressed in the device submission, when relevant. An additional showing of clinical effectiveness of the drug when delivered by the specific device will generally not be required. The device and drug labeling must be mutually conforming with respect to indication, general mode of delivery (e.g., topical, IV), and drug dosage/schedule equivalents.

(iii) For a drug delivery device and drug that are developed for marketing to be used together as a system, a lead center will be designated to be the contact point with the manufacturer(s). If a drug has been developed and marketed and the development and study of device technology predominate, the principal mode of action will be deemed to be that of the device, and CDRH would have the lead. If a device has been developed and marketed and the development and study of drug predominate, then, correspondingly, CDER would have the lead. If neither the drug nor the device is on the market, the lead center will be determined on a case-by-case basis.

*Investigation Options* IDE or IND as appropriate

*Intercenter Consultation* CDER, when the lead center, will consult with CDRH if CDER determines that a specific device is required as part of the NDA process. CDRH as lead center will consult with CDER if

(a) the device is intended for use with (but not already containing) a marketed drug and the device creates a significant change in the intended use, mode of delivery (e.g., topical, IV), or dose/schedule of the drug, or

(b) the device with primary purpose of delivering or aiding in the delivery of a drug and distributed containing a drug (i.e., "pre-filled delivery system").

*Examples*

Nebulizer

Oxygen tank for therapy and over-the-counter (OTC) emergency use

Prefilled syringe

Transdermal patch

*Status* Combination product

*Market Approval Authority* CDER using drug authorities and device authorities as necessary

*Special Label/Regulatory Considerations* None

*Investigation Options* IND

*Intercenter Consultations* Optional

2. Device incorporating a drug component with the combination product having the primary intended purpose of fulfilling a device function

   *Examples*
   Bone cement containing antimicrobial agent
   Cardiac pacemaker lead with steroid-coated tip
   Condom, diaphragm, or cervical cap with contraceptive or antimicrobial agent (including virucidal) agent
   Dental device with fluoride
   Dental wood wedge with hemostatic agent
   Percutaneous cuff (e.g., for a catheter or orthopedic pin) coated/impregnated with antimicrobial agent
   Skin closure or bandage with antimicrobial agent
   Surgical or barrier drape with antimicrobial agent
   Tissue graft with antimicrobial or other drug agent
   Urinary and vascular catheter coated/impregnated with antimicrobial agent
   Wound dressing with antimicrobial agent

*Status* Combination product

*Market Approval Authority* CDRH using device authorities

*Special Label/Regulatory Considerations* These products have a drug component that is present to augment the safety and/or efficacy of the device.

*Investigation Options* IDE

*Intercenter Consultation* Required if a drug or the chemical form of the drug has not been legally marketed in the United States as a human drug for the intended effect

3. Drug incorporating a device component with the combination product having the primary intended purpose of fulfilling a drug function

   *Examples*
   Skin-prep pads with antimicrobial agent
   Surgical scrub brush with antimicrobial agent

*Status* Combination product

*Market Approval Authority* CDER using drug authorities and, as necessary, device authorities

*Special Label/Regulatory Considerations* Marketing of such a device requires a submission of an NDA with safety and efficacy data on the drug component or that it meet monograph specifications as generally recognized as safe (GRAS) and generally recognized as effective (GRAE). Drug requirements (e.g., CGMPs, registration and listing, experience reporting) apply to products.

*Investigation Options* IND

*Intercenter Consultation* Optional

4. (a) Device used in the production of a drug either to deliver directly to a patient or for use in the producing medical facility (excluding use in a registered drug manufacturing facility)

*Examples*
Oxygen concentrators (home or hospital)
Oxygen generator (chemical)
Ozone generator

*Status* Device

*Market Approval Authority* CDER, applying both drug and device authorities

*Special Label/Regulatory Consideration* May also require an NDA if the drug produced is a new drug. Device requirements (e.g., CGMPs, registration and listing, experience reporting) will apply to products.

*Investigation Options* IDA or NDA as appropriate

*Intercenter Consultation* Optional

(b) Drug/device combination product intended to process a drug into a finished package form

*Examples*
Device that uses drug concentrates to prepare large-volume parenterals
Oxygen concentrator (hospital) output used to fill oxygen tanks for use within that medical facility

*Status* Combination product

*Market Approval Authority* CDER, applying both drug and device authorities

*Special Label/Regulatory Considerations* Respective drug and device requirements (e.g., CGMPs, registration and listing, experience reporting) will apply.

*Investigation Options* IDE or NDA as appropriate

*Intercenter Consultation* Optional but will be routinely obtained

B. 1. Device used concomitantly with a drug to directly activate or to augment drug effectiveness

*Examples*
Biliary lithotriptor used in conjunction with dissolution agent
Cancer hyperthermia used in conjunction with chemotherapy
Current generator used in conjunction with an implanted silver electrode (drug) that produces silver ions for an antimicrobial purpose
Materials for blocking blood flow temporarily to restrict chemotherapy drug to the intended site of action

Ultraviolet and/or laser activation of oxsoralen for psoriasis or cutaneous T-cell lymphoma

*Status* Device and drug as separate entities

*Market Approval Authority* CDRH and CDER, respectively

*Special Label/Regulatory Considerations* The device and drug labeling must be mutually conforming with respect to indications, general mode of delivery (e.g., topical, IV), and drug dosage/schedule equivalence. A lead center will be designated to be the contact point with the manufacturer. If a drug has been developed and approved for another use and development and study of device technology predominate, then CDRH would have the lead. If a device has been developed and marketed for another use and development and study of drug action predominate, then CDER would have the lead. If neither the drug nor the device is on the market, the lead center will be determined on a case-by-case basis. If the labeling of the drug and device create a combination product, as defined in the combination product regulations, then the designation of the lead center for both applications will be based upon a determination of the product's primary mode of action.

*Investigation Options* IDE or IND as appropriate

*Intercenter Consultations* Required

2. Device kits labeled for use with drugs that include both device(s) and drug(s) as separate entities in one package with the overall primary intended purpose of the kit fulfilling a device function

   *Examples*

   Medical or surgical kit (e.g., tray) with drug component

*Status* Combination product

*Market Approval Authority* CDRH, using device authorities, is responsible for the kit if the manufacturer is repackaging a market drug. Responsibility for overall packaging resides with CDRH. CDER will be consulted as necessary on the use of drug authorities for the repackaged drug component.

*Special Label/Regulatory Consideration* Device requirements (e.g., CGMPs, registration and listing, experience reporting) apply to kits. Device manufacturers must assure that manufacturing steps do not adversely affect drug components of the kit. If the manufacturing steps do affect the marketed drug (e.g., the kit is sterilized by irradiation), an ANDA or NDA would also be required with CDRH as the lead center.

*Investigation Options* IDA or IND as appropriate

*Intercenter Consultation* Optional if ANDA or NDA not required

C. Liquids, gases, or solids intended for use as devices (e.g., implanted, or components, parts, or accessories to devices)

*Examples*

Dye for tissues used in conjunction with laser surgery to enhance absorption of laser light in target tissue
Gas mixtures for pulmonary function-testing devices
Gases used to provide "physical effects"
Hemodialysis fluids
Hemostatic devices and dressings
Injectable silicon, collagen, and Teflon
Liquids functioning through physical action applied to the body to cool or freeze tissues for therapeutic purposes
Liquids intended to inflate, flush, or moisten (lubricate) indwelling device (in or on the body)
Lubricants and lubricating jellies
Ophthalmic solutions for contact lenses
Organ/tissue transport and/or perfusion fluid with antimicrobial or other drug agent, that is, preservation solutions
Powders for lubricating surgical gloves
Sodium hyaluronate or hyaluronic acid for use as a surgical aid
Solution for use with dental "chemical drill"
Spray on dressings not containing a drug component

*Status* Device

*Market Approval Authority* CDRH

*Special Label/Regulatory Considerations* None

*Investigation Options* IDE

*Intercenter Consultation* Required if the device has direct contact with the body and the drug or the chemical form of the drug has not been legally marketed as a human drug

D. Products regulated as drugs

*Examples*

Irrigation solutions
Purified water or saline in prefilled nebulizers for use in inhalation therapy
Skin protectants (intended for use on intact skin)
Sun screens
Topical/internal analgesic-antipyretic

*Status* Drug

*Market Approval Authority* CDER

*Special Label/Regulatory Considerations* None

*Investigation Options* IND

*Intercenter Consultations* Optional

E. Ad hoc jurisdictional decisions

*Examples*

| | Status | Center |
|---|---|---|
| Motility marker constructed of radiopaque plastic | Device | CDRH |
| Brachytherapy capsules, needles, etc., that are radioactive and may be removed from the body after radiation therapy has been administered | Device | CDRH |
| Skin markers | Device | CDRH |

*Status* Device or drug

*Market Approval Authority* CDRH or CDER as indicated

*Special Label/Regulatory Considerations* None

*Investigation Options* IDE or IND as appropriate

*Intercenter Consultation* Required to assure agreement on drug/device status

**General Criteria Affecting Drug/Device Determination**

The following represent the general criteria that will apply in making device/drug determinations.

A. *Device Criteria*

1. A liquid, powder, or other similar formulation intended only to serve as a component, part, or accessory to a device with a primary mode of action that is physical in nature will be regulated as a device by CDRH.
2. A product that has the physical attributes described in 201(h) (e.g., instrument, apparatus) of the act and does not achieve its primary intended purpose through chemical action within or on the body or by being metabolized will be regulated as a device by CDRH.
3. The phrase "within or on the body" as used in 201(h) of the act does not include extra corporeal systems or the solutions used in conjunction with such equipment. Such equipment and solutions will be regulated as devices by CDRH.
4. An implant, including an injectable material, placed in the body for primarily a structural purpose even though such an implant may be absorbed or metabolized by the body after it has achieved its primary purpose will be regulated as a device by CDRH.
5. A device containing a drug substance as a component with the primary purpose of the combination being to fulfill a device function is a combination product and will be regulated as a device by CDRH.

6. A device (e.g., machine or equipment) marketed to the user, pharmacy, or licensed practitioner that produces a drug will be regulated as a device or combination product by CDER. This does not include equipment marketed to a registered drug manufacturer.
7. A device whose labeling or promotional materials make reference to a specific drug or generic class of drugs unless it is prefilled with a drug ordinarily remains a device regulated by CDRH. It may, however, also be subject to the combination products regulation.

B. *Drug Criteria*

1. A liquid, powder, tablet, or other similar formulation that achieves its primary intended purpose through chemical action within or on the body or by being metabolized, unless it meets one of the specified device criteria, will as regulated as a drug by CDER.
2. A device that serves as a container for a drug or a device that is a drug delivery system attached to the drug container where the drug is present in the container is a combination product that will be regulated as a drug by CDER.
3. A device containing a drug substance as a component with the primary purpose of the combination product being to fulfill a drug purpose is a combination product and will be regulated as a drug by CDER.
4. A drug whose labeling or promotional materials makes reference to a specific device or generic class of devices ordinarily remains a drug regulated by CDER. It may, however, also be subject to the combination products regulation.

## 10 CONCLUSIONS

In summary, we have touched upon the regulations that currently control the types of preclinical toxicity testing done on potential human pharmaceuticals and medical device products. We have reviewed the history, the law, the regulations themselves, the guidelines, and common practices employed to meet regulatory standards. Types of toxicity testing were discussed, as were the special cases pertaining to, for example, biotechnology products.

## REFERENCES AND SUGGESTED READINGS

Alder, S., and Zbinden, G. (1988). *National and International Drug Safety Guidelines*. MTC Verlag Zoblikon, Switzerland.

Anderson, O. (1988). *The Health of a Nation; Harvey Wiley and the Fight for Pure Food*. University of Chicago, Chicago, IL.

Anonymous (1992a). FDA's policy statement for the development of new stereoisomeric drugs. Available: http://www.fda.gov/cder/guidance/stereo.htm.

Anonymous (1992b). Orphan Drug Regulations 21 CFR PART 316. Available: http://www.fda.gov/orphan/about/odreg.htm.

Anonymous (1996), Well-characterized biologics would be regulated under FD&C Act by Kassebaum FDA Reform Bill. *F-D-C Reports (Pink Sheets)* 58:11–12.

Anonymous (1997a). Guidance for industry and reviewers: Repeal of Section 507 of the Federal Food, Drug and Cosmetic Act. Available: http://www.fda.gov/cder/guidance/index.htm.

Anonymous (1997b). Points to consider in the manufacture and testing of monoclonal antibody products for human use. Available: http://www.fda.gov/cber/cberftp.html.

Anonymous (1998a). Annual report of the Pharmaceutical Research and Manufacturer Association (Priority #2: Improved FDA regulation of drug development), PhRMA, Washington, DC.

Anonymous (1998b). Guidance for industry: Content and format for geriatric labeling (draft guidance). Available: http://www.fda.gov/cder/guidance/2527dft.pdf.

Anonymous (1998). http://www.usajobs.org/jobs/66743900/Equipment%20Specialist%20-%20ORF.htm8a.

Barton, B. (1998). International Conference on Harmonization—Good clinical practices update. *Drug Inform. J.* 32:1143–1147.

Berliner, V. R. (1974). U.S. Food and Drug Administration requirements for toxicity testing of contraceptive products. In M. H. Briggs and E. Diczbalusy (Eds.). Pharmacological models in contraceptive development. *Acta Endocrinol. (Copenhagen)* suppl. 185:240–253.

Blakeslee, D. (1994). Thalidomide, a background briefing. JAMA HIV/AIDS Information Center Newsline. Available: http://www.ama-assn.org/special/hiv/newsline/briefing/briefing.htm.

Brusiek, D. (1987). *Principles of Genetic Toxicology*, 2nd ed., Plenum Press, New York.

Buesing, M. McSweegan (1999). Submitting biologics applications to the Center for Biologics Evaluation and Research electronically. *Drug Inform. J.* 33:1–15.

Burns, J. (1983). Overview of safety regulations governing food, drug and cosmetics. In F. Homberger (Ed.), *The United States in Safety and Evaluation and Regulation of Chemicals 3: Interface between Law and Science*. Karger, New York.

Cadden, S. (1999). *New Drug Approval in Canada*. Parexel, Waltham, MA.

Center for Drug Evaluation and Research (CDER) (Nov. 1989). Guideline for the study of drugs likely to be used in the elderly. U.S. Food and Drug Administration, Washington, DC.

Center for Drug Evaluation and Research (CDER) (1990). Premarket notifications: 510 (K) regulatory requirements for medical devices. Health and Human Servises, U.S. Food and Drug Administration, Washington, DC, p. 4158.

Currie, W. J. C. (1995). *New Drug Approval in Japan*. Parexel, Waltham, MA.

Daniels, J., Nestmann, E., and Kerr, A. (1997). Development of stereoisomeric (chiral) drugs: A brief review Of the scientific and regulatory considerations. *Drug Inform. J.* 41:639–646.

DeGeorge, J., Ahn, C., Andrews, P., Bower, M., Giorgio, D., Goheer, M., Lee-Yam, D., McGuinn, W., Schmidt, W., Sun, C., and Tripathi, S. (1998). Regulatory considerations for the preclinical development of anticancer drugs. *Cancer Chemother. Pharmacol.* 41:173–185.

DiMasi, J., Seibring, M., and Lasagna, L. (1994). New drug development in the United States from 1963 to 1992. *Clin. Pharmacol. Ther.* 55:609–622.

Ember, L. F. (2001). Raising the bar for quality of drugs. C&EN Mar. 19:26–30.

Evers, P. T. (1985). *New Drug Approval in the European Union*. Parexel, Waltham, MA.

Food and Drug Law Institute (1999). *Compilation of Food and Drug Laws*, Washington, DC.

Fung, M., Thornton, A., Mybeck, K., Ksiao-Hui, J., Hornbuckle, K., and Muniz, E. (2001). Evaluation of the characteristics of safety withdrawal of prescription drugs from worldwide pharmaceutical markets—1960 to 1999. *Drug Inform. J.* 35: 293–317.

Gad, S. C. (1994). *Safety Assessment for Pharmaceuticals*. Van Nostrand Reinhold, New York.

Goldenthal, E. (1968). Current view on safety evaluation of drugs. *FDA Papers* May:13–18.

Grabowski, H. G., and Vernon, J. M. (1983). *The Regulation of Pharmaceuticals*. Enterprise Institute, Washington DC.

Grieshaber, C. (1991). Prediction of human toxicity of new antineoplastic drugs from studies in animals. In G. Powis and M. Hacker (Eds.), *The Toxicity of Anticancer Drugs*. Pergamon, New York, pp. 10–24.

Haffner, M. (1998). Orphan drug development—International program and study design issues. *Drug Inform. J.* 32:93–99.

Hart, C. (1999). Getting PSArt the politics and paperwork of pediatric drug studies. *Modern Drug Discov.* 2:15–16.

Hayes, T., and Ryffel, B. (1997). Symposium in writing: Safety considerations of recombinant protein therapy; introductory comments. *Clin. Immunol. Immunother.* 83:1–4.

Hess, G. (1999). FDA accelerates its approval of new drugs. *Chem Market Newslett.* 255:21.

Hite, M. (1997). Safety pharmacology approaches. *Int.J. Toxicol.* 16:23–31.

Holtzman, D. C. (1994). European biotech firms invest in US facilities. *Bioworld Today*, June 2.

Hutt, P. B. (1983a). Investigations and reports respecting FDA regulations of new drugs (Part I). *Clin. Pharmacol. Ther.* 33:537–548.

Hutt, P. B. (1983b). Investigations and reports respecting FDA regulations of new drugs (Part II). *Clin. Pharmacol. Ther.* 33:174–187.

Hutt, P. B. (1987). Progress in new drug regulation. *Clin. Res. Prac. Reg. Affairs* 5:307–318.

International Conference on Harmonisation (ICH) (1997). *International Conference on Harmonization Safety Steps 4/5 Documents*. Interpharm, Buffalo Grove, IL.

International Conference on Harmonisation (ICH) (2000). *Non-Clinical Safety Studies for the Conduct of Human Clinical Trials for Pharmaceuticals*.

Jackson, S. (1970). Evolution of the regulation of ethical drugs. *American Economic Review* 60:107–124.

Kessler, D., Siegel, J., Noguchi, P., Zoon, K., Feidon, K., and Woodcock, J. (1993). Regulation of somatic-cell therapy and gene therapy by the Food and Drug Administration. *New Engl. J. Med.* 329:1169–1173.

Lumley, C. E., and Walker, S. E. (1992). An international appraisal of the minimum duration of chronic animal toxicity studies. *Hum. Exp. Toxicol.* 11:155–162.

Lumpkin, M. (1995). Guidance for industry: Content and format of investigational new drug applications (INDs) for phase 1 studies of drugs, including well-characterized, therapeutic, and biotechnology-derived products. Available: http://www.fda.gov/cder/guidance/phase1.pdf.

Manson, J. (1994). Testing of pharmaceutical agents for reproductive toxicity in developmental toxicology. In C. Kimmel and J. Buelke-Sam (Eds.), *Reproductive Toxicology*, Raven, New York, pp. 379–402.

Marwick, C. (1998). Implementing the FDA Modernization Act (Medical News and Perspectives). *J. Am. Med. Assoc.* 279:815–816.

Merrill, R. (1994). Regulation of drugs and devices: An evolution. *Health Affairs* 13:47–69.

Ohno, Y. (1998). Harmonization of the timing of the non-clinical tests in relation to the conduct of clinical trials. *J. Controlled Release* 62:57–63.

Pilot, L., and Waldemann, D. (1998). Food and Drug Modernization Act of 1997: Medical device provisions. *Food Drug Law J.* 53:267–295.

Ryffel, B. (1996). Unanticipated human toxicology of recombinant proteins. *Arch. Toxicol.* Suppl. 18:333–341.

Ryffel, B. (1997). Safety of human recombinant proteins. *Biomed. Environ. Sci.* 10:65–71.

Schacter, E., and DeSantis, P. (1998). Labeling of drug and biologic products for pediatric use. Drug Inform. *J.* 32:299–303.

Termini, R. B. (2001). *Food, Drug, and Medical Device Law: Topics and Cases*. FOI Services, Gaithersburg, MD.

United States Pharmacopeia (USP) (2000). *United States Pharmacopeia XXIV and National Formulary 19*. USP, Philadelphia, PA.

U.S. Food and Drug Administration (FDA) (1971). *FDA Introduction to Total Drug Quality*. U.S. Government Printing Office, Washington, DC.

U.S. Food and Drug Administration (FDA) (1985). Points to consider in the production and testing of new drugs and biologicals produced by recombinant DNA technology. FDA, Washington, DC.

U.S. Food and Drug Administration (FDA) (1987). *Environmental Assessment Technical Assistance Handbook*. NTIS No. PB87-175345. National Technical Information Service, Springfield, VA.

U.S. Food and Drug Administration (FDA) (1988a). Good laboratory practices. CFR 21 Part 58, April 1988.

U.S. Food and Drug Administration (FDA) (1988b). FDA Perspective on development of stereoisomers AAPS, May 16. FDA, Washington, DC.

U.S. Food and Drug Administration (FDA) (1988c). Investigational new drug, antibiotic, and biological drug product regulations, procedures for drug intended to treat life-threatening and severely debilitating illnesses. *Federal Register* 53(204), 1988:41516–41524.

U.S. Food and Drug Administration (FDA) (1991). Points to consider in human somatic cell therapy and gene therapy. FDA, Washington, DC.

U.S. Food and Drug Administration (FDA) (1993). Points to consider in the characterization of cell lines used to produce biologicals. FDA, Washington, DC.

Wessinger, J. (1989). Pharmacologic and toxicologic considerations for evaluating biologic products. *Reg. Tox. Pharmacol.* 10:255–263.

Wu, K.-M., DeGeorge, J., Atrachi, A., Barry, E., Bigger, A., Chen, C., Du, T., Freed, L., Geyer, H., Goheer, A., Jacobs, A., Jean, D., Rhee, H., Osterburg, R., Schmidt, W., and Farrelly, J. (2000). Regulatory science: A special update from the United States Food and Drug Administration. Preclinical issues and status of investigations of botanical drug products in the United States. *Toxicol. Lett.* 111:199–202.

Yakuji Nippo (1991). *Drug Registration Requirements in Japan*, Yakuji Nippo, Tokyo.

Yakuji Nippo (1994). *Guidelines for Toxicity Study of Drugs Manual*. Yakuji Nippo, Tokyo.

Young, F., Nightingale, S., Mitchell, W., and Beaver, L. (1989). The United States Drug Export Amendment Act of 1986: Perspectives from the Food and Drug Administration. *Int. Dig. Health Legislation* 40:246–252.

# *Prior Art and Its Use in Safety Assessment Process*

The appropriate starting place for the safety assessment of any new chemical entity, particularly a potential new drug, is to first determine what is already known about the molecule, its structural and therapeutics class analogues (pharmacological analogues being agents with assumed similar pharmacological mechanisms), and the disease one seeks to treat. Such a determination requires the fullest possible access and review of the available literature. Here we try to at least overview the range of approaches to gathering such data (Table 1). In using this information, one must keep in mind that there is both an initial requirement to build a data file or database and a continuing need to update such a database on a regular basis, seving as part of the project record. Updating a database requires not merely adding to what is already there but also discarding out-of-date (i.e., now known to be incorrect) information and reviewing the entire structure for connections and organization.

Such data are first used in selecting which possible compounds should be carried forward in development as a possible new drug [as illustrated in Figure 1 and explored in detail in Gad (2005)].

## 1 CLAIMS

Claims are what is said in labeling and advertising and may be either of a positive (therapeutic or beneficial) or negative (lack of an adverse effect)

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

**TABLE 1 Sources of Prior Art**

| Internet |
|---|
| FDA: Inactive Ingredients for Currently Marketed Drug Products, http://www.accessdata.fda.gov/scripts/cder/iig/index.cfm |
| Proprietory databases |
| Medline/toxline/journals |
| Book (monographs and edited) |
| Personal network/meetings |
| Obscure databases |

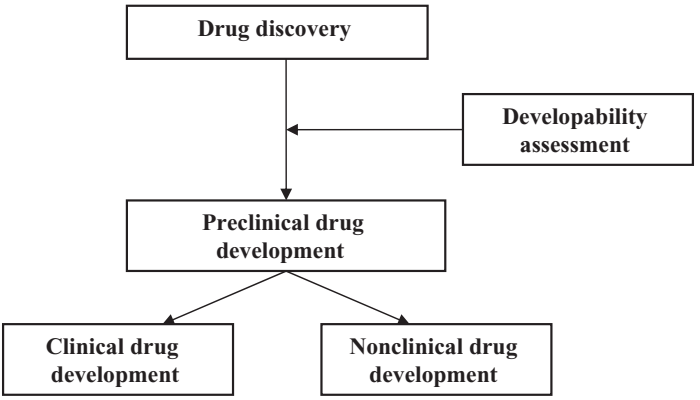


***Figure 1*** *Prior art in assessing pharmaceutical developability.*

nature. The positive or efficacy claims are not usually the direct concern of the toxicologist, though it must be kept in mind that such claims both must be proved and can easily exceed the limits of the statutory definition of a device, turning the product into a drug or combination product. Negative claims such as "nonirritating" or "hypoallergenic" also must be proved and are generally the responsibility of the product safety professional to provide proof for. There are special tests for such claims.

## 2 TIME AND ECONOMIES

The final factors of influence or arbitrator of test conduct and timing are the requirements of the marketplace, the resources of the organization, and the economic worth of the product. Plans for filings with regulatory agencies and for market launches are typically set before actual testing (or final stage development) is undertaken, as the need to be in the marketplace in a certain time frame is critical. Such timing and economic issues are beyond the scope of this volume but must be considered.

## 3 PRIOR KNOWLEDGE

The first step in any new literature review is to obtain as much of the following information as possible:

1. Correct chemical identity, including molecular formula, Chemical Abstracts Service (CAS) registry number, common synonyms, trade names, and a structural diagram. Gosselin et al. (1984) and Ash and Ash (1995, 2007) are excellent sources of information on existing commercial products and their components and uses.
2. Chemical composition (if a mixture) and major impurities.
3. Production and use information.
4. Chemical and physical properties (physical state, vapor pressure, pH, solubility, chemical reactivity, etc.).
5. Any structurally related chemical substances that are already on the market or in production.
6. Known or presumed pharmacological properties.

Collection of the above information is not only important for hazard assessment (high vapor pressure would indicate high inhalation potential, just as high and low pH would indicate high irritation potential), but the prior identification of all intended use and exposure patterns may provide leads to alternative information sources; for example, drugs to be used as antineoplastics or antibiotics may already have extensive toxicology data obtainable from government or private sources. A great deal of the existing toxicity information (particularly information on acute toxicity) is not available in the published or electronic literature because of concerns about the proprietary nature of this information and the widespread opinion that it does not have enough intrinsic scholarly value to merit publication. This unavailability is unfortunate, as it leads to a lot of replication of effort and expenditure of resources that could be better used elsewhere. It also means that an experienced toxicologist must use an informal search of the unpublished literature of his colleagues as a supplement to searches of the published and electronic literature.

There are now numerous published texts that should be considered for use in literature-reviewing activities. An alphabetic listing of 24 of the more commonly used hard-copy sources for safety assessment data is provided in Table 2. Obviously, this is not a complete listing and consists of only the general multipurpose texts that have a wider range of applicability for toxicology. Texts dealing with specialized classes of agents (e.g., disinfectants) or with specific target organ toxicity (neurotoxins and teratogens) are generally beyond the scope of this text. Parker (1988) should be consulted for details on the use of these texts. Wexler (2008), Parker (1988), and Sidhu et al. (1989) should be consulted for more extensive listings of the literature and computerized databases. Such sources can be off direct (free) Internet sources (where one must

**TABLE 2 Key Safety Assessment Reference Texts**

Abraham, Donald J. (Ed.) (2003). *Burger's Medicinal Chemistry and Drug Discovery*. John Wiley & Sons, New York.
American Conference of Government Industrial Hygienists (2000). *Documentation of the Threshold Limit Values and Biological Exposure Indices*, 7th ed.—2002 Supplement. ACGIH.
Ash, M., and Ash, I. (1995). *Food Additives, Electronic Handbook*. Gower, Brookfield, VT.
Ash, M., and Ash, I. (2007). *Pharmaceuticals Additives*, 2nd ed., *Electronic Handbook*. Gower, Brookfield, VT.
Barnhart, E. R. (2007). *Physician's Desk Reference*. Medical Economics Company, Oradell, NJ.
Bingham, E., et al. (Eds.) (2001). *Patty's Toxicology*, 5th ed. John Wiley & Sons, New York.
Budavari, E., et al. (Eds.) (2006). *The Merck Index*, 15th ed. Merck and Company, Inc., Rahway, NJ.
Center for Disease Control (2003). *Registry of Toxic Effects of Chemical Substances (RTECS)*. National Institute for Occupational Safety and Health.
Cronin, E. (1980). *Contact Dermatitis*. Churchill Livingston, Edinburgh.
Dart, R. C., et al. (Eds.) (2004). *Medical Toxicology*, 3rd ed., Lippincott Williams & Wilkins, Baltimore, MD.
Deichmann, W., and Gerard, H. (1996). *Toxicology of Drugs and Chemicals*. Academic, New York.
Dipiro, J. T., et al. (Eds.) (2002). *Pharmacotherapy, a Pathophysiologic Approach*, 5th ed., McGraw Hill, New York.
Finkel, A. J. (1983). *Hamilton and Hardy's Industrial Toxicology*, 4th ed. John Wright PSG, Boston.
Ford, M. D., et al. (2001). *Clinical Toxicology*. W.B. Saunders Company, Philadelphia.
Gosselin, R. E., Smith, R. P., Hodge, H. C. (1984). *Clinical Toxicology of Commercial Products*, 5th ed. Wilkins and Williams, Baltimore, MD.
Grant, Morton W., and Schuman, Joel S. (1993). *Toxicology of the Eye*. Charles C. Thomas, Springfield, IL.
Haddad, Lester M., et al. (1998). *Clinical Management of Poisoning and Drug Overdose*, 3rd ed. Saunders, Philadelphia.
Klaassen, Curtis D. (Ed.) (2007). *Casarett & Doull's Toxicology: The Basic Science of Poisons*, 6th ed. McGraw-Hill, New York.
Lewis Sr., Richard J. (Ed.) (1991). *Carcinogenically Active Chemicals: A Reference Guide*. Van Nostrand Reinhold, New York.
Lewis Sr., Richard J. (Ed.) (2000). *Sax's Dangerous Properties of Industrial Materials*, 10th ed., 3 Volume Set. John Wiley & Sons, New York.
National Toxicology Program (2000). Nineteenth Annual Report on Carcinogens. PB 85-134633. Department of Health and Human Services, Washington, DC.
O'Neil, M. J., et al. (Eds.) (2006). *The Merck Index*, 14th ed. Merck & Co., Whitehouse Station, NJ.
Proctor, N. H., and Hughes, J. P. (1978). *Chemical Hazards of the Workplace*, J. B. Lippincott, Philadelphia.
Sax, N. I. (2000). *Dangerous Properties of Industrial Materials*, 10th ed. Van Nostrand Reinhold, New York.
Schardein, J. (2000). *Chemically Induced Birth Defects*. Marcel Dekker, NY.
Shephard, Thomas J. (2007). *Catalog of Teratogenic Agents (Shepard's Catalog)*, 11th ed. Johns Hopkins University Press, Baltimore, MD.
Sweetman, S. C. (2007). *Martindale: The Complete Drug Reference*, Pharmaceutical Press, Chicago.
U.S. Department of Health and Human Services (2002). *Report on Carcinogens*, 10th ed. Public Health Service, National Toxicology Program. (http://ehp.niehs.nih.gov/roc/toc10.html)
Thomas Healthcare (2007). Physician's Desk Reference, 61st ed. Thomson Healthcare, Montvale, NJ. (http://www.pdr.net)
Wexler, Philip (Ed.) (2005). *Encyclopedia of Toxicology*. Elsevier Ltd., Oxford.
Wiley-Interscience (2007). *Wiley Handbook of Current and Emergency Drug Therapies*, 8 Volumes. John Wiley & Sons, Hoboken, N.J.

beware of GIGO: garbage in, garbage out), commercial databases, and package products, to mention just the major categories. Appendix C provides addresses for major free Internet sources.

## 4 MISCELLANEOUS REFERENCE SOURCES

There are some excellent published information sources covering some specific classes of chemicals, for example, heavy metals, plastics, resins, or petroleum hydrocarbons. The National Academy of Science series *Medical and Biologic Effects of Environment Pollutants* covers 10–15 substances considered to be environmental pollutants. *CRC Critical Reviews in Toxicology* is a well-known scientific journal that over the years has compiled over 20 volumes of extensive literature reviews of a wide variety of chemical substances. A photocopy of this journal's topical index will prevent one from overlooking information that may be contained in this important source. Trade organizations such as the Fragrance Industry Manufacturers Association and the Chemical Manufacturers Association have extensive toxicology databases from their research programs that are readily available to toxicologists of member companies. Texts that deal with specific target organ toxicity—neurotoxicity, hepatotoxicity, or hematotoxicity—often contain detailed information on a wide range of chemical structures. Published information sources like the *Target of Organ Toxicity* series (Taylor and Francis, now halfway through a second round of new editions) or a few examples of publications that contain important information on many industrial chemicals may be useful either directly or by analogy. Upon discovery that the material one is evaluating may possess target organ toxicity, a cursory review of these types of texts is warranted.

In the last decade, for many toxicologists the online literature search has changed from an occasional, sporadic activity to a semicontinuous need. Usually, non-toxicology-related search capabilities are already in place in many companies. Therefore, all that is needed is to expand the information source to include some of the databases that cover the types of toxicology information one desires. However, if no capabilities exist within an organization, one can approach a university, consultant, or private contract laboratory and utilize their online system at a reasonable rate. It is even possible to access most of these sources from home using a personal computer. The major available online databases are as follows:

A. *National Library of Medicine* The National Library of Medicine (NLM) information retrieval service contains the well-known and frequently used Medline, Toxline, and Cancerlit databases. Databases commonly used by toxicologists for acute data in the NLM service are the following:
   1. Toxline (Toxicology Information Online) is a bibliographic database covering the pharmacological, biochemical, physiological, environmen-

tal, and toxicological effects of drugs and other chemicals. It contains approximately 1.7 million citations, most of which are complete with abstract, index terms, and CAS registry numbers. Toxline citations have publication dates of 1981 to the present. Older information is on Toxline 65 (pre-1965 through 1980).

2. Medline (Medical Information Online) is a database containing approximately 7 million references to biomedical journal articles published since 1966. These articles, usually with an English abstract, are from over 3000 journals. Coverage of previous years (back to 1966) is provided by back files, searchable online, that total some 3.5 million references.
3. Toxnet (Toxicology Data Network) is a computerized network of toxicologically oriented data banks. Toxnet offers a sophisticated search and retrieval package that accesses the following three subfiles:
   a. Hazardous Substances Data Bank (HSDB) is a scientifically reviewed and edited data bank containing toxicological information enhanced with additional data related to the environment, emergency situations, and regulatory issues. Data are derived from a variety of sources, including government documents and special reports. This database contains records for over 4100 chemical substances.
   b. Toxicology Data Bank (TDB) is a peer-reviewed data bank focusing on toxicological and pharmacological data, environmental and occupational information, manufacturing and use data, and chemical and physical properties. References have been extracted from a selected list of standard source documents.
   c. Chemical Carcinogenesis Research Information System (CCRIS) is a National Cancer Institute–sponsored database derived from both short- and long-term bioassays on 2379 chemical substances. Studies cover carcinogenicity, mutagenicity, promotion, and cocarcinogenicity.
4. Registry of Toxic Effects of Chemical Substances (RTECS) is the NLM's online version of the National Institute for Occupational Safety and Health's (NIOSH) annual compilation of substances with toxic activity. The original collection of data was derived from the 1971 Toxic Substances Lists. RTECS data contain threshold limit values, aquatic toxicity ratings, air standards, National Toxicology Program carcinogenesis bioassay information, and toxicological/carcinogenic review information. The NIOSH is responsible for the file content in RTECS and for providing quarterly updates to NLM: RTECS currently covers toxicity data on more than 106,000 substances.

B. *The Merck Index* *The Merck Index* is now available online for up-to-the minute access to new chemical entities.

## 5 SEARCH PROCEDURE

As mentioned in the introduction, chemical composition and identification information should already have been obtained before the chemical is to be

searched. With most information retrieval systems this is a relatively straightforward procedure. Citations on a given subject may be retrieved by entering the desired free text terms as they appear in titles, key words, and abstracts of articles. The search is then initiated by entering the CAS number and/or synonyms. If you are only interested in a specific target organ effect—for instance, carcinogenicity—or specific publication years, searches can be limited to a finite number of abstracts before requesting the printout.

Often it is unnecessary to request a full printout (author, title, abstract). You may choose to review just the author and title listing before selecting the abstracts of interest. In the long run, this approach may save you computer time, especially if the number of citations being searched is large.

Once you have reviewed the abstracts, the last step is to request photocopies of the articles of interest. Extreme caution should be used in making any final health hazard determination based solely on an abstract or nonprimary literature source.

## 6 MONITORING PUBLISHED LITERATURE AND OTHER RESEARCH IN PROGRESS

Although there are a few other publications offering similar services, the *Life Sciences* edition of *Current Contents* is the publication most widely used by toxicologists for monitoring the published literature. *Current Contents* monitors over 1180 major journals and provides a weekly listing by title and author. Selecting out those journals you wish to monitor is one means of selectively monitoring the major toxicology journals.

Aids available to the toxicologist for monitoring research in progress are quite variable. The National Toxicology Program's (NTP) *Annual Plan for Fiscal Year XXXX* highlights all the accomplishments of the previous year and outlines the research plans for the coming year. *Annual Plan* contains all projects in the president's proposed fiscal year budget that occur within the National Cancer Institute/National Institutes of Health, National Institute of Environmental Health Sciences/National Institutes of Health, National Center for Toxicological Research/Food and Drug Administration, and NIOSH/Centers for Disease Control. this report includes a list of all the chemicals selected for testing in research areas that include but are not limited to mutagenicity, immunotoxicity, teratoly/reproduction, neurotoxicity, pharmacokinetics, subchronic toxicity, and chronic toxicity/carcinogenicity.

*Annual Plan* also contains a bibliography of NTP publications from the previous year. A companion publication is the 1999 NTP *Review of Current DHHS, DOE, and EPA Research Related to Toxicology*. Similar to *Annual Plan*, this document provides detailed summaries of both proposed and ongoing research.

Another mechanism for monitoring research in progress is by reviewing abstracts presented at the annual meetings of professional societies such as the Society of Toxicology, Teratology Society, Environmental Mutagen Society,

and American College of Toxicology. These societies usually have their abstracts prepared in printed form; for example, the current *Toxicologist* contains over 1700 abstracts presented at the annual meeting. Copies of the titles and authors of these abstracts are usually listed in the societies/respective journals, which, in many cases, would be reproduced and could be reviewed through *Current Contents*.

## 7 NEW SOURCES

Scientists today are more aware than ever before of the existence of what has been called the "information revolutions." At no other time in recent history has so much information become available from so many different "traditional" resources—including books, reviews, journals, and meetings—as well as personal computer–based materials such as databases, alerting services, optical-disk-based information, and news media.

The good news for toxicologists interested in the safety of chemical entities of all types is that numerous new computer-based information products are available that can be extremely useful additions to current safety and toxicology libraries. These tools enable one to save considerable time, effort, and money while evaluating the safety of chemical entities.

The primary focus of this section is on the description and applications of the recent innovations of newly emerging information services based on the personal computer (PC).

## 8 KINDS OF INFORMATION

The kinds of information described here are found on three types of PC media—floppy, CD-ROM, and laser disks. The products run the gamut of allowing one to assess current developments on a weekly basis as well as carry out more traditional reviews of historical information. The general types of information one can cover include basic pharmacology, preclinical toxicology, competitive products, and clinical safety.

The specific products discussed are as follows: two floppy disk–based products called Current Contents on Diskette and Focus On Global Change; five CD-ROM products called Toxic Release Inventory, Material Safety Data Sheets, CCINFOdisc, Pollution/Toxicology, and Medline Ondisc; and a laser disk product entitled the Veterinary Pathology Slide Bank. We provide a brief synopsis of the major features of each as well as a description of their integration into a functional, PC-based toxicology information center (TIC).

When such a TIC is established, one will find that some unusual benefits accrue. One now has immediate and uninterrupted access to libraries of valuable and comprehensive scientific data. This access is free of "online" constraints and designed to be user friendly, with readily retrievable information

available 24 hours a day, 7 days a week. The retrieved information can also usually be manipulated in electronic form, so one can use it in reports and/or store it in machine-readable form as ASCII files.

The minimal hardware requirements, which are certainly adequate for all items discussed here, are an IBM or IBM-compatible PC equipped with at least 640K RAM, a single floppy disk drive, at least a 40-Mbyte hard-disk drive, a CD-ROM drive, a VGA color monitor, and a printer. The basic point here is that hardware requirements are minimal and readily available. In the case of the laser disk products, a laser disk drive and high-resolution (VGA) monitor are also required.

### 8.1 PC-Based Information Products: Floppy Disk Based

We currently have ready access to a rapidly growing variety of relevant information resources. From a current awareness perspective, an excellent source of weekly information is the floppy disk–based product called Current Contents on Diskette (CCOD). Several versions are available; however, the Life Sciences version is most appropriate for this review because of its coverage, on a weekly basis, of over 1200 journals describing work in the biological sciences. One will note that the product has several useful features, including very quick retrieval of article citations as well as several output options (including either hard-copy or electronic storage of references as well as reprint requests).

### 8.2 PC-Based Information Products: CD-ROM Media

The gradual emergence of this technology during the past several years has recently blossomed with the introduction of several CD-ROM products that deal with safety issues surrounding the toxicology and safety of chemicals. CD-ROM media with such information can generally be characterized by two major advantages: They are relatively easy to use and are amazingly quick in retrieving data of interest.

***Toxic Release Inventory (TRI)*** Before embarking on a discussion of products describing health, toxicology, and safety issues, it is well to be aware of a new, pilot CD-ROM version of the U.S. Environmental Protection Agency's (EPA) 1987 Toxic Chemical Release Inventory and Hazardous Substances Fact Sheets. This TRI resource, which contains information regarding the annual inventory of hundreds of named toxic chemicals from certain facilities (since 1987) as well as the toxicological and ecological effects of chemicals, is available from the National Technical Information Service (NTIS), U.S. Department of Commerce, Springfield, Virginia 22161.

The list of toxic chemicals subject to reporting was originally derived from those designed for similar purposes by the states of Maryland and New Jersey. As such, over 300 chemicals and categories are noted. (After appropriate rule

making, modifications to the list can be made by the EPA.) The inventory is designed to inform the public and government officials about routine and accidental releases of toxic chemicals to the environment.

The CD-ROM version of the database can be efficiently searched with a menu-driven type of software called Search Express. It allows one to search with Boolean expressions as well as individual words and/or frequency of "hits" as a function of the number of documents retrieved on a given topic. Numerous searchable fields have been included, allowing one to retrieve information by a variety of means—for example, the compound name; the chemical registry number; the amount of material released into the air, water, or land; the location of the site of release; and the Standard Industrial Classification (SIC) code of the releasing party. One can also employ ranging methods with available numeric fields and sorting of output.

It is hoped that this shared information will help to increase the awareness, concern, and action by individuals to ensure a clean and safe environment. The TRI database is a significant contribution to that effort and the CD-ROM version is a superb medium with which to widely publicize and make accessible the findings.

***Material Safety Data Sheets (MSDSs)*** The MSDS CD-ROM is a useful resource that contains over 33,000 MSDSs on chemicals submitted to the Occupational Safety and Health Administration (OSHA) by chemical manufacturers. This resource contains complete MSDS information as well as other important information such as the chemical formula, structure, physical properties, synonyms, registry number, and safety information.

Users can easily search the CD-ROM by employing the Aldrich catalog number, CAS number, chemical name, or molecular formula. One can also export the chemical structures to some supported software for subsequent inclusion into work-processing programs. The product is available from Aldrich Chemical Company, 940 West St. Paul Avenue, Milwaukee, Wisconsin 54233.

***Canadian Centre for Occupational Health and Safety (CCINFO)*** This set of four CD-ROM disks contains several valuable databases of information that are updated on a quarterly basis: MSDS, CHEM Data, Occupational Health and Safety (OHS) Source, and OHS Data. The MSDS component currently contains over 60,000 MSDSs supplied by chemical manufacturers and distributors. It also contains several other databases [RIPP, RIPA, Pest Management Research Information System (PRIS)], one of which (PRIS) even includes information on pest management products, including their presence and allowable limits in food.

A second disk in the series (CHEM Data) contains comprehensive information from CHEMINFO, Registry of Toxic Effects of Chemical Substances (RTECS) and Chemical Evaluation Search and Retrieval System (CESARS) data bases as well as recommendations on Transport of Dangerous Goods (TDG)/Hazardous Materials (49 CFR).

The third and fourth disks include OHS information. These disks contain databases on resource organizations, resource people, case law, jurisprudence, fatalities, mining incidents, and ADISCAN. Furthermore, information on noise levels, and National Institute for Occupational Safety and Health (NIOSHTEC) nonionizing radiation levels, and a Document Information Directory System is readily retrievable. These CD-ROM materials are available from the Canadian Center for Occupational Health and Safety, 250 Main Street East, Hamilton, Ontario L8N 1H6.

***Pollution and Toxicology (POLTOX)*** This CD-ROM library also focuses our attention on environmental health and safety concerns. Scientists working in any industry or capacity that deals with toxic or potentially toxic chemicals will find it very useful. It allows one access to seven major databases in this field in a single search through its use of "linking" features in its software. The distributors of this product have provided us with a spectrum of information dealing with toxic substances and environmental health.

The collection of these databases include five that are available exclusively from Cambridge Scientific Abstracts (CSA)—Pollution Abstracts, Toxicology Abstracts, Ecology Abstracts, Health and Safety Science Abstracts, and Aquatic Pollution and Environmental Quality. The abstracts come from journals or digests published by CSA on important issues, including environmental pollution, toxicological studies of industrial chemicals, ecological impacts of biologically active chemicals, as well as health, safety, and risk management in occupational situations. The POLTOX CD-ROM contains over 200,000 records from these sources since 1981.

POLTOX also contains two other useful databases—Toxline (described earlier) and the Food Science and Technology Abstracts (FSTA) libraries. The FSTA component is a reasonably comprehensive collection of information regarding toxicological aspects of compounds found in food, including contamination, poison, and carcinogenic properties. The CD-ROM product is available from Compact Cambridge, 7200 Wisconsin Avenue, Bethesda, Maryland 20814.

***Medline*** The Medline database, which comes from the NLM, is a superb, indispensable reference library that is particularly strong in its wide coverage of research activities in the biomedical literature. It also encompasses the areas of clinical medicine, health policy, and health care services. Each year, over 300,000 articles are reviewed and indexed into the database. The full bibliographic citations of these articles, usually including the abstract of the published work, are available from numerous vendors in CD-ROM format and are usually updated on a monthly basis.

Information can be accessed from Medline in a variety of ways: by author, title, subject, CAS registration number, keyword, publication year, and journal title. Medline Ondisc is the CD-ROM product we employ (from Dialog Information Services, 3460 Hillview Ave., Palo Alto, CA 94304). It allows one access

to the full Medline files back to 1984. Each year from that time until 1988 is covered on a single CD-ROM disk; starting in 1989, each disk covers only a six-month time period. The information is accessed through either an easily employed "menu-driven" system or a more standard online type of "command language."

Gower Publishing (Brookfield, VT) has published a series of "electronic handbooks" providing approved ingredient information on materials used in cosmetics, personal care additives, food additives, and pharmaceuticals. Academic Press, through its Sci-Vision branch, launched (in 2000) an ambitious service of CD ROM-based toxicity database products which are structure and substructure searchable.

It is worth nothing that the CD-ROM-based system has been seamlessly integrated with both (proprietary) recordkeeping and communications software so that one can optionally monitor the use of the online services and easily continue searching in the Dialog "online" environment after using the CD-ROM-based Medline library. Another very useful feature includes the storage of one's search logic so that repetitive types of searches, over time, for example, can be done very easily.

### 8.3 PC-Based Information Products: Laser Disk

***International Veterinary Pathology Slide Bank (IVPSB)*** This application represents an important complementary approach toward training and awareness using laser disk technology. The IVPSB provides a quality collection of transparencies, laser videodisks, and interactive computer/videodisk training programs. In particular, the videodisk contains over 21,000 slides from over 60 contributors representing 37 institutions from 6 countries. These slides are accessible almost instantaneously because of the tremendous storage capacity and rapid random-search capabilities of the videodisk through the interactive flexibility of the computer. The information available is of particular interest to toxicologists and pathologists because the visuals illustrate examples of gross lesions of infectious diseases, regional diseases, clinical signs or external microscopy, histopathology, normal histology, cytology and hematology, and parasitology.

The laser disk, a catalog of the entrees, a computer database, and selected interactive programs can be obtained from Dr. Wayne Crowell, Department of Veterinary Pathology, University of Georgia, Athens, Georgia 30602.

## 9 CONCLUSIONS

This brief overview of some of the readily available PC-based information resources will, hopefully, encourage more widespread use of this type of technology. Toxicologists and pathologists, in particular, can avail themselves of these useful resources in a way that was simply not possible just a few years

ago. The information one needs to make decisions is now far more accessible to many more of us for relatively reasonable expenditures of money for software and hardware.

An effective approach to provide maximal access to these resources is to set up a TIC, which consists of the earlier noted PC hardware and single, centrally available copies of the noted floppy disk, CD-ROM-based, and laser disk products. By employing a menu-based system (available commercially or by shareware) to access the respective products, one can usually provide entry into each of the products discussed here with a single keystroke.

As time goes on, one can grow with the system by considering networking the CD-ROM-based resources and/or setting up other, strategically located TICs on one's campus. The important concept here is that we wish to make the superb "new" PC-based information products as available as we can to interested scientists.

A critical part of the strategy for delivery of information to the end user is that one can anticipate marked increased usage of the more traditional, hard-copy-based resources of the centralized library. The tools described here are frequently complementary to the pivotal library-based information center. What one can anticipate, however, is a much more focused use of hard-copy-based information.

## REFERENCES

Abraham, D. J. (Ed.) (2003). *Burger's Medicinial Chemistry and Drug Discovery*. Wiley, New York.

American Conference of Governmental Industrial Hygienists (ACGIH) (2000). *Documentation of the Threshold Limit Values for Substances in Workroom Air*, 5th ed. ACGIH, Cincinnati, OH.

Ash, M., and Ash, I. (1995). *Food Additives, Electronic Handbook*. Gower, Brookfield, VT.

Ash, M., and Ash, I. (2007). *Pharmaceutical Additives,* 2nd ed., *Electronic Handbook*. Gower, Brookfield, VT.

Barnhart, E. R. (2007). *Physician's Desk Reference*. Medical Economics Company, Oradell, NJ.

Bingham, E., et al. (Eds.) (2001). *Patty's Toxicology*, 5th ed. Wiley, New York.

Budavari, S. (2006). *The Merck Index*, 15th Ed. Merck and Company, Inc., Rahway, NJ.

Cronin, E. (1980). *Contact Dermatitis*. Churchill Livingston, Edinburgh.

Dart, R. C., et al. (Eds.) (2004). *Medical Toxicology*, 3rd ed. Lippincott Williams & Wilkins, Philadelphia, PA.

Deichmann, W., and Gerard, H. (1996). *Toxicology of Drugs and Chemicals*. Academic, New York.

Dipiro, J. T., et al. (Eds.) (2002). *Pharmacotherapy, a Pathophysiologic Approach*, 5th ed. McGraw-Hill, New York.

Finkel, A. J. (1983). *Hamilton and Hardy's Industrial Toxicology*, 4th ed. John Wright PSG, Boston.

Ford, M. D., et al. (2001). *Clinical Toxicology*. W.B. Saunders, Philadelphia.

Gad, S. C. (2005). *Drug Discovery Handbook*. Wiley-Interscience, Hoboken, NJ.

Gosselin, R. E., Smith, R. P., and Hodge, H. C. (1984). *Clinical Toxicology of Commercial Products*, 5th ed. Williams & Wilkins, Baltimore, MD.

Grant, W. M. (1993). *Toxicology of the Eye*, 4th ed. Charles C. Thomas, Springfield, IL.

Haddad, L. M., et al. (1998). *Clinical Management of Poisoning and Drug Overdose*, 3rd ed. Saunders, Philadelphia.

Klassen, C. D. (Ed.) (2007). *Casarett and Doull's Toxicology*, 6th ed. McGraw Hill, New York.

Lewis, R. J. (Ed.) (1991). *Carcinogenically Active Chemicals: A Reference Guide*. Van Nostrand Reinhold, New York.

National Toxicology Program (2000). *Nineteenth Annual Report on Carcinogens*. PB 85-134633. Department of Health and Human Services, Washington, DC.

O'Neil, M. J., et al. (Eds.) (2006). *The Merck Index*, 14th ed. Merck & Co., Whitehouse Station, NJ.

Parker, C. M. (1988). Available toxicology information sources and their use. In S. C. Gad (Ed.), *Product Safety Evaluation Handbook*. Marcel Dekker, New York, pp. 23–41.

Proctor, N. H., and Hughes, J. P. (1978). *Chemical Hazards of the Workplace*. J. B. Lippincott, Philadelphia.

Sax, N. I. (2000). *Dangerous Properties of Industrial Materials*, 10th ed. Van Nostrand Reinhold, New York.

Schardein, J. L. (1999). *Chemically Induced Birth Defects*, 3rd ed. Marcel Dekker, New York.

Shepard, T. H. (1998). *Catalog of Teratogenic Agents*, 9th ed. Johns Hopkins University Press, Baltimore, MD.

Sidhu, K. S., Stewart, T. M., and Netton, E. W. (1989). Information sources and support networks in toxicology. *J. Am. Coll. Toxicol.* 8:1011–1026.

Sittig, M. (1985). *Handbook of Toxic and Hazardous Chemicals*, 2nd ed. Noyes Publications, Park Ridge, NJ.

Sweetman, S. C. (2007). *Martindale: The Complete Drug Reference*. Pharmaceutical Press, Chicago.

Thomas Healthcare (2007). *Physician's Desk Reference*, 61st ed. Thomson Healthcare, Montvale, NJ.

Turco, S., and Davis, N. M. (1973). Clinical significance of particular matter: A review of the literature. *Hosp. Pharm.* 8:137–140.

Wexler, P. (2004). *Encyclopedia of Toxicology*, 2nd ed. Academic, San Diego, CA.

Wexler, P. (2008). *Information Resources in Toxicology*, 3rd ed. Elsevier, New York.

Wiley (2007). *Wiley Handbooks of Current and Emerging Drug Therapies* (8 vols.). (2007). Wiley-Interscience, Hoboken, NJ.

# *Screens in Safety and Hazard Assessment*

## 1 INTRODUCTION

In biological research, screens are tests designed and performed to identify agents or organisms having a certain set of characteristics that will either exclude them from further consideration or cause them to be selected for closer attention. In pharmaceutical safety assessment, our use of screens is usually negative (i.e., no activity is found)—agents or objects possessing certain biochemical activities are considered to present enough of a hazard that they are not studied further or developed as potential therapeutic agents without compelling reasons (in cases of extreme benefit such as life-saving qualities).

In the broadest terms what is done in preclinical (and, indeed, in phase I clinical) studies can be considered a form of screening (Zbinden et al., 1984). What varies is the degree of effectiveness of (or our confidence in) each of the tests used. As a general rule, though we think of the expensive and labor-intensive "pivotal" studies required to support regulatory requirements (e.g., 4-week-to-1-year toxicity, carcinogenicity, and segment I–III studies) as definitive, in fact, they are highly effective (or at least we generally so believe) but not necessarily efficient screens.

Though toxicologists in the pharmaceutical industry are familiar with the broad concepts of screening, they generally do not recognize the applicability of screens. The principles underlying screening are also not generally well

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

recognized or understood. The objective behind the entire safety assessment process in the pharmaceutical industry is to identify those compounds for which the risk of harming humans does not exceed the potential benefit to them.

In most cases this means that if a test or screen identifies a level of risk that we have confidence in (our "activity criterion"), then the compound that was tested is no longer considered a viable candidate for development. In this approach, what may change from test to test is the activity criterion (i.e., our basis for and degree of confidence in the outcome). We are interested in minimizing the number of false negatives in safety assessment. Anderson and Hauck (1983) should be consulted for statistical methods to minimize false-negative results.

Figure 1 illustrates how currently decisions are more likely to be made on a multidimensional basis, which creates a need for balance among (1) degree of benefit, (2) confidence that there *is* a benefit (efficacy is being evaluated in "models" or screens at the same time safety is), (3) type of risk (with, e.g., muscle irritation, mutagenicity, acute lethality, and carcinogenicity having various degrees of concern attached to them), and (4) confidence in and degree of risk. This necessity for balance is commonly missed by many who voice opposition to screens because "they may cause us to throw out a promising compound based on a finding in which we have only (for example) 80% confidence." Screens, particularly those performed early in the research and development process, should be viewed as the biological equivalent of exploratory data analysis. They should be very sensitive, which by definition means that they will have a lot of "noise" associated with them. Screens generally do not establish that an agent is (or is not) a "bad actor" for a certain endpoint. Rather, they confirm that if interest in a compound is sufficient, a more definitive test (a confirmatory test) is required, which frequently will provide a basis for selecting between multiple candidate compounds.

## 2 CHARACTERISTICS OF SCREENS

The terminology involved in screen design and evaluation and the characteristics of a screen should be clearly stated and understood. The characteristics of screen performance are defined as:

- Sensitivity: the ratio of true positives to total actives
- Specificity: the ratio of true negatives to total inactives
- Positive accuracy: the ratio of true to observed positives
- Negative accuracy: the ratio of true to observed negatives
- Capacity: the number of compounds that can be evaluated
- Reproducibility: the probability that a screen will produce the same results at another time (and, perhaps, in some other laboratory)

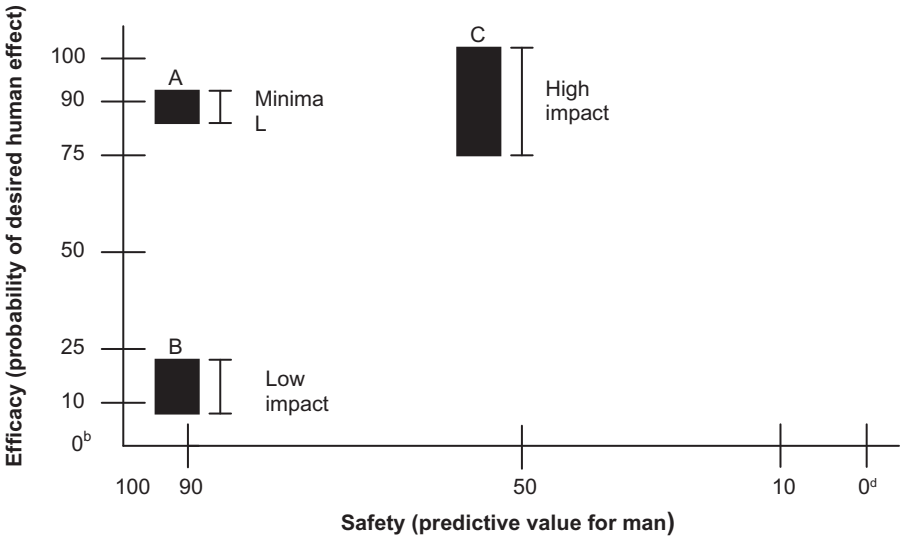


***Figure 1*** *Decision making for pharmaceutical candidates based on outcome of screening tests. (a) A 100% probability of efficacy means that every compound that has the observed performance in the model(s) used has the desired activity in man. (b) A 0% probability of efficacy means that every compound that has the observed performance in the model(s) used does not have the desired activity in man. (c) A 100% probability of a safety finding means that such a compound would definitely cause this toxicity in man. (d) A 0% probability means this will never cause such a problem in man. Note: These four cases (*a, b, c, *and* d*) are almost never found.*

*The height of the "impact" column refers to the relative importance ("human risk") of a safety finding. Compound A has a high probability of efficacy but also a high probability of having some adverse effect in man. But if that adverse effect is of low impact—say, transitory muscle irritation for a life-saving antibiotic—A should go forward. Likewise, B, which has a low probability of efficacy and a high probability of having an adverse effect with moderate impact, should not be pursued. Compound C is at a place where the high end of the impact scale should be considered. Though there is only a 5% probability of this finding (say, neurotoxicity or carcinogenicity) being predictive in man, the adverse effect is not an acceptable one. Here a more definitive test is called for or the compound should be dropped.*

These characteristics may be optimized for a particular use if we also consider the mathematics underlying them and "errors."

A brief review of the basic relationships between error types and power starts with considering each of five interacting factors (Gad, 1982a, 1982b, 1999) that serve to determine power and define competing error rates:

$\alpha$: the probability of our committing a type I error (a false positive)

$\beta$: the probability of our committing a type II error (a false negative)

$\Delta$: the desired sensitivity in a screen (such as being able to detect an increase of 10% in mutations in a population)

$\sigma$: the variability of the biological system and the effects of chance errors

$n$: the necessary sample size needed to achieve the desired levels of each of these factors

We can, by our actions, generally change only this portion of the equation, since $n$ is proportional to $\sigma/(\alpha, \beta, \text{and } \Delta)$.

The implications of this are therefore that (1) the greater $\sigma$ is, the larger $n$ must be to achieve the desired levels of $\alpha$, $\beta$, and/or $\Delta$, and (2) the smaller the desired levels of $\alpha$, $\beta$, and/or $\Delta$, if $n$ is constant, the larger $\sigma$ must be.

What are the background response level and the variability in our technique? As any good toxicologist will acknowledge, matched concurrent control (or standardization) groups are essential to minimize within-group variability as an "error" contributor. Unfortunately, in in vivo toxicology test systems, large sample sizes are not readily attainable, and there are other complications to this problem that we shall consider later.

In an early screen, a relatively large number of compounds will be tested. It is unlikely that one will stand out so much as to have greater statistical significance than all the other compounds (Bergman and Gittins, 1985). A more or less continuous range of activities will be found instead. Compounds showing the highest (beneficial) or lowest (adverse) activity will proceed to the next assay or tier of tests in the series and may be used as lead compounds in a new cycle of testing and evaluation.

The balance between how well a screen discovers activities of interest versus other effects (specificity) is thus critical. Table 1 presents a graphic illustration of the dynamic relationship between discovery and discrimination.

Both discovery and discrimination in screens hinge on the decision criterion that is used to determine if activity has or has not been detected. How sharply such a criterion is defined and how well it reflects the working of a screening system are two of the critical factors driving screen design.

An advantage of testing many compounds is that it gives the opportunity to average activity evidence over structural classes or to study quantitative structure–activity relationships (QSARs). Quantitative structure–activity relationships can be used to predict the activity of new compounds and thus reduce the chance of in vivo testing on negative compounds. The use of

**TABLE 1 Discovery and Discrimination of Toxicants**

| | Actual Activity of Agent Tested | |
|---|---|---|
| Screen Outcome | Positive | Negative |
| Positive | *a* | *b* |
| Negative | *c* | *d* |

*Notes*: Discovery (sensitivity) $= a/(a+c)$, where $a$ = all toxicants found positive
$a + c$ = all toxicants tested
Discrimination (specificity) $= d/(b+d)$, where $d$ = all nontoxicants found negative
$b + d$ = all nontoxicants tested

QSARs can increase the proportion of truly active compounds passing through the system.

It should be remembered that maximization of the performance of a series of screening assays requires close collaboration among the toxicologist, chemist, and statistician. Screening, however, forms only part of a much larger research and development context. Screens thus may be considered the biological equivalent of exploratory data analysis (EDA). In fact, EDA methods, provide a number of useful possibilities for less rigid and yet utilitarian approaches to the statistical analysis of the data from screens and are one of the alternative approaches presented and evaluated here (Tukey, 1977; Redman, 1981; Hoaglin et al., 1983, 1985). Over the years, the author has published and consulted on a large number of screening studies and projects. These have usually been directed at detecting or identifying potential behavioral toxicants or neurotoxicants, but some have been directed at pharmacological, immunotoxic, and genotoxic agents (Gad, 1988, 1989a).

The general principles or considerations for screening in safety assessments are as follows:

1. Screens almost always focus on detecting a single point of effect (such as mutagenicity, lethality, neurotoxicity, or developmental toxicity) and have a particular set of operating characteristics in common.
2. A large number of compounds are evaluated, so ease and speed of performance (which may also be considered efficiency) are very desirable characteristics.
3. The screen must be very sensitive in its detection of potential effective agents. An absolute minimum of active agents should escape detection; that is, there should be very few false negatives (in other words, the type II error rate or $\beta$ level should be low). Stated yet another way, the signal gain should be way up.
4. It is desirable that the number of false positives be small (i.e., there should be a low type I error rate or $\alpha$ level).
5. Items 2–4, which to some degree are contradictory, require that involved researchers agree on a set of compromises, starting with the acceptance of a relatively high $\alpha$ level (0.10 or more), that is, a higher noise level.
6. In an effort to better serve item 1, safety assessment screens frequently are performed in batteries so that multiple endpoints are measured in the same operation. Additionally, such measurements may be repeated over a period of time in each model as a means of supporting item 2.
7. The screen should use small amounts of compound to make item 1 possible and should allow evaluation of materials that have limited availability (such as novel compounds) early in development.
8. Any screening system should be validated initially using a set of blind (positive and negative) controls. These blind controls should also be

evaluated in the screening system on a regular basis to ensure continuing proper operation of the screen. As such, the analysis techniques used here can then be used to ensure the quality or modify the performance of a screening system.

9. The more that is known about the activity of interest, the more specific the form of screen that can be employed. As specificity increases, so should sensitivity. However, generally the size of what constitutes a meaningful change (that is, $\Delta$) must be estimated and is rarely truly known.
10. Sample (group) sizes are generally small.
11. The data tend to be imprecisely gathered (often because researchers are unsure what they are looking for) and therefore possess extreme within-group variability or modify test performance.
12. Proper dose selection is essential for effective and efficient screen design and conduct. If insufficient data are available, a suitably broad range of doses must be evaluated (however, this technique is undesirable on multiple grounds, as has already been pointed out).

Much of the mathematics involved in calculating screen characteristics came from World War II military-based operations analysis and research, where it was important in the design of radar, antiair, and antisubmarine warfare systems and operations (Garrett and London, 1970).

## 2.1 Uses of Screens

The use of screens first occurs most to pharmaceutical scientists in pharmacology (Martin et al., 1988). Early experiences with the biological effects of a new molecule are almost always in some form of efficacy or pharmacology screen. The earliest of these tend to be with narrowly focused models, not infrequently performed in vitro. The later pharmacology screens, performed in vivo to increase confidence in the therapeutic potential of a new agent or to characterize its other activities [cardiovascular, central nervous system (CNS), etc.], can frequently provide some information of use in safety assessment also (even if only to narrow the limits of doses to be evaluated), and the results of these screens should be considered in early planning. In the new millennium, requirements for specific safety pharmacology screens have been promulgated. Additionally, since the late 1990s two new areas of screening have become very important in pharmaceutical safety assessment. The first is the use of screens for detecting compounds with the potential to cause fatal cardiac arrhythmias. These are almost always preceded by the early induction of a prolongation of the Q–T interval. While this should be detected in electrocardiograms (ECGs) performed in repeat-dose canine studies, several early screens [such as the human Ether-à-go-go Related Gene (hERG)] are more rapid and efficient (though not conclusive) for selecting candidate compounds for further development.

The other area is the use of microassays in toxicogenomic screening—early detection of the potential for compounds to alter gene expressions with adverse consequences (Pennie, 2000; Nuwaysir et al., 1999).

Safety assessment screens are performed in three major settings—discovery support, development (what is generally considered the "real job" of safety assessment), and occupational health/environmental assessment testing. Discovery support is the most natural area of employment of screens and is where effective and efficient screen design and conduct can pay the greatest long-range benefits. If compounds with unacceptable safety profiles can be identified before substantial resources are invested in them—and structures modified to maintain efficacy while avoiding early safety concerns—then long-term success of the entire research and development effort is enhanced. In the discovery support phase, one has the greatest flexibility in the design and use of screens. Here screens truly are used to select from among a number of compounds.

Examples of the use of screens in the development stage are presented in some detail in the next section.

The use of screens in environmental assessment and occupational health is fairly straightforward. On the occupational side the concerns (as addressed in Chapter Developmental and Reproductive Toxicity Testing of this volume) address the potential hazards to those involved in making the bulk drug. The need to address potential environmental concerns covers both true environmental items (e.g., aquatic toxicity) and potential health concerns for environmental exposure of individuals. The resulting work tends to be either regulatorily defined tests (for aquatic toxicity) or defined endpoints such as dermal irritation and sensitization, which have been (in a sense) screened for already in other nonspecific tests.

The most readily recognized examples of screens in toxicology are those that focus on a single endpoint. The traditional members of this group include genotoxicity tests, lethality tests (particularly recognizable as a screen when in the form of limit tests), and tests for corrosion, irritation (both eye and skin), and skin sensitization. Others that fit this same pattern, as will be shown, include the carcinogenicity bioassay (especially the transgenic mouse models) and developmental toxicity studies.

The "chronic" rodent carcinogenicity bioassay is thought of as the "gold standard" or definitive study for carcinogenicity, but, in fact, it was originally designed (and functions) as a screen for strong carcinogens (Page, 1977). It uses high doses to increase its sensitivity in detecting an effect in a small sample of animals. The model system (be it rats or mice) has significant background problems of interpretation. As with most screens, the design has been optimized (by using inbred animals, high doses, etc.) to detect one type of toxicant—strong carcinogens. Indeed, a negative finding does not mean that a material is not a carcinogen but rather than it is unlikely to be a potent one.

Many of the studies done in safety assessment are multiple-endpoint screens. Such study types as a 90-day toxicity study or immunotox/neurotox screens are designed to measure multiple endpoints with the desire of increasing both

sensitivity and reliability (by correspondence/correlation checks between multiple data sets).

## 2.2 Types of Screens

There are three major types of screen designs: single stage, sequential, and tiered. Both the sequential and tiered are multistage approaches, and each of these types also varies in terms of how many parameters are measured. These three major types have the following characteristics:

***Single Stage*** A single test will be used to determine acceptance or rejection of a test material. Once an activity criterion (such as $X$ score in a righting reflex test) is established, compounds are evaluated based on being less than $X$ (i.e., negative) or equal to or greater than $X$ (i.e., positive). As more data are accumulated, the criterion should be reassessed.

***Sequential*** Two or more repetitions of the same test are performed, one after the other, with the severity of the criterion for activity being increased in each sequential stage. This procedure permits classification of compounds into a set of various ranges of potencies. As a general rule, it appears that a two-stage procedure, by optimizing decision rules and rescreening compounds before declaring compounds "interesting," increases both sensitivity and positive accuracy; however, efficiency is decreased (or is throughput rate).

***Tier (or Multistage)*** In this procedure, materials found active in a screen are reevaluated in one or more additional screens or tests that have greater discrimination. Each subsequent screen or test is both more definitive and more expensive.

For purposes of our discussion here, we will primarily focus on the single-stage system, which is the simplest. The approaches presented here are appropriate for use in any of these screening systems, although establishment of activity criteria becomes more complicated in successive screens. Clearly, the use of multistage screens presents an opportunity to obtain increased benefits from the use of earlier (lower order) screening data to modify subsequent screen performance and the activity criterion.

## 2.3 Criterion: Development and Use

In any early screen, a relatively large number of compounds will be evaluated with the expectation that a minority will be active. It is unlikely that any one will stand out so much as to have greater statistical significance that all the other compounds based on a formal statistical test. A more or less continuous range of activities will be found. Compounds displaying a certain degree of activity will be identified as "active" and handled as such. For safety screens, those which are "inactive" go on to the next test in a series and may be used

as lead compounds in a new cycle of testing and evaluation. The single most critical part of the use of screens is how to make the decision that activity has been found.

Each test or assay has an associated activity criterion. If the result for a particular test compound meets this criterion, the compound is active and handled accordingly. Such a criterion could have a statistical basis (e.g., all compounds with observed activities significantly greater than the control at the 5% level could be tagged). However, for early screens, a statistical criterion may be too strict, given the power of the assay, resulting in a few compounds being identified as active. In fact, a criterion should be established (and perhaps modified over time) to provide a desired degree of confidence in the predictive value of the screen.

A useful indicator of the efficiency of an assay series is the frequency of discovery of truly active compounds. This is related to the probability of discovery and to the degree of risk (hazard to health) associated with an active compound passing a screen undetected. These two factors in turn depend on the distribution of activities in the series of compounds being tested and the chances of rejecting and accepting compounds with given activities at each stage.

Statistical modeling of the assay system may lead to the improvement of the design of the system by reducing the interval between discoveries of active compounds. The objectives behind a screen and considerations of (1) costs for producing compounds and testing and (2) degree of uncertainty about test performance will determine desired performance characteristics of specific cases. In the most common case of early toxicity screens performed to remove possible problem compounds, preliminary results suggest that it may be beneficial to increase the number of compounds tested, decrease the numbers of animals (or other test models) per assay, and increase the range and number of doses. The result will be less information on more structures, but there will be an overall increase in the frequency of discovery of active compounds (assuming that truly active compounds are entering the system at a random and steady rate).

The methods described here are well suited to analyzing screening data when the interest is truly in detecting the absence of an effect with little chance of false negatives. There are many forms of graphical analysis methods available, including some newer forms that are particularly well suited to multivariate data (the type that are common in more complicated screening test designs). It is intended that these aspects of analysis will be focused on in a later publication.

The design of each assay and the choice of the activity criterion should, therefore, be adjusted, bearing in mind the relative costs of retaining false positives and rejecting false negatives (Bickis, 1990). Decreasing the group sizes in the early assays reduced the chance of obtaining significance at any particular level (such as 5%), so that the activity criterion must be relaxed, in a statistical sense, to allow more compounds through. At some stage, however, it becomes too expensive to continue screening many false positives, and the

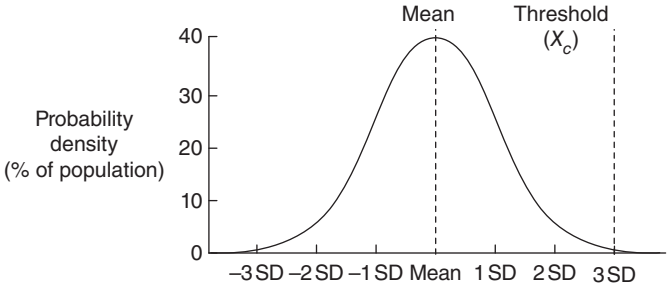


***Figure 2*** *Setting thresholds using historical control data. The figure shows a Gaussian ("normal") distribution of screen parameters; 99.7% of the observations in the population are within three standard deviations (SD) of the historic mean. Here the threshold (i.e., the point at which a datum is outside of normal) was set at* $X_c$ *= mean + 3 SD. Note that such a screen is one sided.*

criteria must be tightened accordingly. Where the criteria are set depends on the acceptable noise levels in a screening system.

Criteria can be as simple (lethality) or as complex (a number of clinical chemical and hematological parameters) as required. The first step in establishing them should be an evaluation of the performance of test systems that have not been treated (i.e., negative controls). There will be some innate variability in the population, and understanding this variability is essential to selling some "threshold" for "activity" that has an acceptably low level of occurrence in a control population. Figure 2 illustrates this approach.

What endpoints are measured as inputs to an activity criterion are intrinsic in the screen system but may be either direct (i.e., having some established mechanistic relationship to the endpoint that is being predicted in humans, such as gene mutations as predictive of carcinogenicity) or correlative. Correlated variables (such as many of those measured in in vitro systems) are "black-box" predictors—compounds causing certain changes in these variables have a high probability of having a certain effect in humans, though the mechanisms (or commonality of mechanism) is not established. There is also, it should be noted, a group of effects seen in animals the relevance of which in humans is not known. This illustrates an important point to consider in the design of a screen—one should have an understanding (in advance) of the actions to be taken given each of the possible outcomes of a screen.

## 2.4 Analysis of Screening Data

Screening data present a special case that, due to their inherent characteristics, is not well served by traditional approaches (Gad, 1988, 1989a,b,c).

Why? First consider which factors influence the power of a statistical test. Gad (1988) established the basic factors that influence the statistical performance of any bioassay in terms of its sensitivity and error rates. Recently, Healy (1987) presented a review of the factors that influence the power of a

study (the ability to detect a dose-related effect when it actually exists). In brief, the power of a study depends on six aspects of study design:

- Sample size
- Background variability (error variance)
- Size of true effect to be detected (i.e., objective of the study)
- Type of significance test
- Significance level
- Decision rule (the number of false positives one will accept)

There are several ways to increase power—each with a consequence:

| Action | Consequence |
|---|---|
| Increase the sample size | Greater resources required |
| Design test to detect larger differences | Less useful conclusions |
| Use a more powerful significance test | Stronger assumptions required |
| Increase the significance level | Higher statistical false-positive rate |
| Use one-tailed decision rule | Blind to effects in the opposite direction |

Timely and constant incorporation of knowledge of test system characteristics and performance will reduce background variability and allow sharper focus on the actual variable of interest. There are, however, a variety of nontraditional approaches to the analysis of screening data.

### *Univariate Data*

*Control Charts* The control chart approach (Montgomery, 1985), commonly used in manufacturing quality control in another form of screening (for defective product units), offers some desirable characteristics.

By keeping records of cumulative results during the development of screen methodology, an initial estimate of the variability (such as standard deviation) of each assay will be available when full-scale use of the screen starts. The initial estimates can then be revised as more data are generated (i.e., as we become more familiar with the screen).

The following example shows the usefulness of control charts for control measurements in a screening procedure. Our example test for screening potential muscle strength suppressive agents measures reduction of grip strength by test compounds compared with a control treatment. A control chart was established to monitor the performance of the control agent to (1) establish the mean and variability of the control and (2) ensure that the results of the control for a given experiment are within reasonable limits (a validation of the assay procedure).

As in control charts for quality control, the mean and average range of the assay were determined from previous experiments. In this example, the screen had been run 20 times previous to collecting the data shown. These initial data showed a mean grip strength $X$ of 400 g and a mean range $R$ of 90 g. These values were used for the control chart (Figure 3). The subgroups are of size 5. The action limits for the mean and range charts were calculated as follows:

$$X \pm 0.58R = 400 \pm 0.58 \times 90 = 348\text{–}452 \quad \text{(from } X \text{ chart)}$$

Then, using the upper limit ($du$) for $n = 5$,

$$2.11R = 2.11 \times 90 = 190 \quad \text{(upper limit for range)}$$

Note that the range limit, which actually established a limit for the variability of our data, is in fact a "detector" for the presence of outliers (extreme values).

Such charts may also be constructed and used for proportion or count types of data. By constructing such charts for the range of control data, we may then

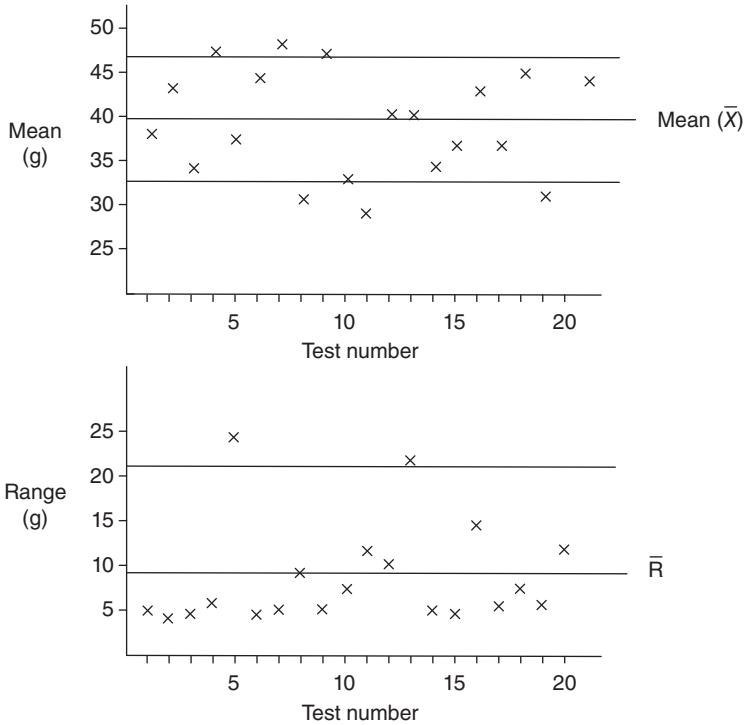


***Figure 3*** *Example of a control chart used to "prescreen" data (actually, explore and identify influential variables) from a portion of a functional observational battery.*

use them as rapid and efficient tools for detecting effects in groups being assessed for that same activity endpoint.

*Central Tendency Plots* The objective behind our analysis of screen data is to have a means of efficiently, rapidly, and objectively identifying those agents that have a reasonable probability of being active. Any materials that we so identify may be further investigated in a more rigorous manner, which will generate data that can be analyzed by traditional means. In other words, we want a method that makes out-of-the-ordinary results stand out. To do this we must first set the limits on "ordinary" (summarize the control case data) and then overlay a scheme that causes those things that are not ordinary to become readily detected ("exposed," in EDA terms) (Velleman and Hoaglin, 1981; Tufte, 1983). One can then perform "confirmatory" tests and statistical analysis (using traditional hypothesis-testing techniques), if so desired.

If we collect a set of control data on a variable (say scores on our observations of the righting reflex) from some number of ordinary animals, we can plot it as a set of two histograms (one for individual animals and the second for the highest total score in each randomly assigned group of five animals), such as those shown in Figure 4 (the data for which came from 200 actual control animals).

Such a plot identifies the nature of our data, visually classifying them into those that will not influence our analysis (in the set shown, clearly scores of zero fit into this category) and those that will critically influence the outcome of an analysis. In so doing, the position of control ("normal") observations is

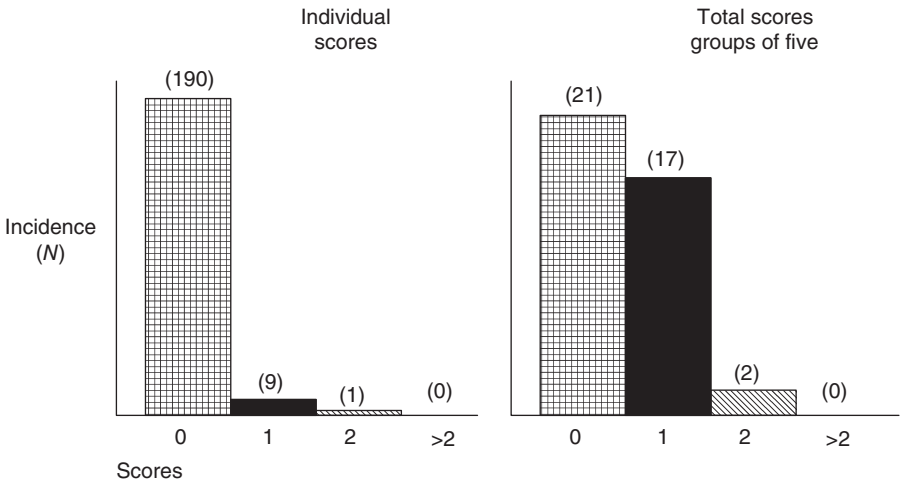


***Figure 4*** *Plotting central tendency. Possible individual scores for righting reflexes may range from 0 to 8 (Gad, 1982a). Group total scores would thus range from 0 to 40. (Shown are the number of groups that contain individual scores in the individual categories.)*

readily revealed as a "central tendency" in the data (hence the name for this technique).

We can (and should) develop such plots for each of our variables. Simple inspection makes clear that answers having no discriminatory power (zero values in Figure 4) do not interest us or influence our identifying of an outlier in a group and should simply be put aside or ignored before continuing with the analysis. This first stage, summarizing the control data, thus gives us a device for identifying data with discriminatory power (extreme values), allowing us to set aside the data without discriminatory power.

Focusing our efforts on the remainder, it becomes clear that although the incidence of a single, low, nonzero observation in a group means nothing, total group scores of 2 or more occurred only 5% of the time by chance. So we can simply perform an extreme-value screen on our "collapsed" data sets, looking for total group values or individual values that are beyond our acceptance criteria.

The next step in this method is to develop a histogram for each ranked or quantal variable by both individual and group. "Useless" data (those that will not influence the outcome of the analysis) are then identified and dropped from analysis. Group scores may then be simply evaluated against the baseline histograms to identify those groups with scores divergent enough from control to be either true positives or acceptably low incidence false positives. Additional control data can continue to be incorporated in such a system over time, both increasing the power of the analysis and providing a check on screen performance.

***Multivariate Data*** The traditional acute, subchronic, and chronic toxicity studies performed in rodents and other species also can be considered to constitute multiple-endpoint screens. Although the numerically measured continuous variables (body weight, food consumption, hematology values) generally can be statistically evaluated individually by traditional means, the same concerns of loss of information present in the interrelationship of such variables apply. Generally, traditional multivariate methods are not available, efficient, sensitive, or practical (Young, 1985).

*Analog Plot* The human eye is extremely good at comparing the size, shape, and color of pictorial symbols (Anderson, 1960; Andrews, 1972; Davison, 1983; Schmid, 1983; Cleveland and McGill, 1985). Furthermore, it can simultaneously appreciate both the minute detail and the broad pattern.

The simple way of transforming a table of numbers to a sheet of pictures is by using analog plots. Numbers are converted to symbols according to their magnitude. The greater the number, the larger the symbol. Multiple variables can be portrayed as separate columns or as differently shaped or colored symbols (Wilk and Gnanadesikan, 1986).

The conversion requires a conversion chart from the magnitude of the number to the symbol size. The conversion function should be monotonic (e.g., dose, and the measured responses should each change in one direction accord-

ing to a linear, logarithmic, or probit function). Log conversion will give more emphasis to differences at the lower end of the scale, whereas a probit will stabilize the central range of response (16–84%) of a percentage variable. For example, for numbers $x$, symbol radius $r$, and plotting scaling factor $k$, a log mapping will give

$$x = \begin{cases} 1 & r = k \\ 10 & r = 2k \\ 100 & r = 3k \end{cases}$$

To compare different variables on the same sheet requires some form of standardization to put them on the same scale. Also, a choice must be made between displaying the magnitude of the numbers and their significance (Kruskal, 1964; Kass, 1980). Two possibilities are:

- Express each mean as a percentage change from a control level or overall mean (*means plot*)
- Calculate effects for meaningful contrasts (*contrasts plot*)

The analog plot chart in Figure 5 shows relationships for five measures on a time-versus-dose basis, allowing ready evaluation of interrelationships and patterns.

A study using 50 rats of each sex in each of five groups (two controls and three increasing doses) measured body weight and food and liquid consumption every week or month for two years. This resulted in 3 variables × 2 sexes × 5 groups × 53 times × 50 animals. Means alone constituted some 1600 four-digit numbers.

Body weight gains from the period immediately preceding each consumption measurement were used since these were less correlated. For each variable and at each time, the sums of squares for group differences were divided into four meaningful contrasts:

Control A versus control B

Control A + B versus low

Control A + B + low versus medium

Control A + B + low + medium versus high

To make the variables comparable, the sums of squares were standardized by the within-group standard deviations. Contrast involving doses can be compared with the contrast for the difference between the controls, which should be random. The clearest feature is the high-dose effect for food consumption. However, this seems not to be closely correlated with changes in body weight gains. Certain changes can be seen at the later measurement times, probably because of dying animals.

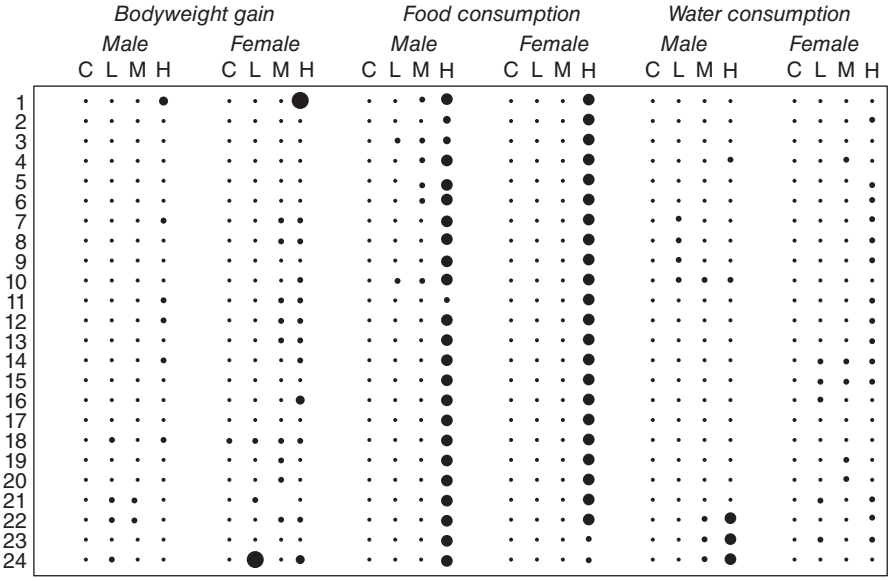


Note: Square root relationship

Key: C = control A vs. B.
L = low vs. A + B.
M = medium vs. A + B + low.
H = high vs. A + B + low + medium

***Figure 5*** *Analog plot for dose–response contrasts. One of many possible approaches to graphically presenting multidimensional data. In this case, various effects—day of dosing, dose response, and magnitude of response—are simultaneously portrayed, with the size of each circle being proportional to the magnitude of the measured value.*

There are numerous approaches to the problem of capturing all the information in a set of multi-endpoint data. When the data are continuous in nature, approaches such as the analog plot can be used (Chernoff, 1973; Chambers et al., 1983). A form of control chart also can be derived for such uses when detecting effect rather than exploring relationships between variables is the goal. When the data are discontinuous, other forms of analysis must be used. Just as the control chart can be adapted to analyzing attribute data, an analog plot can be adapted. Other methods are also available.

## REFERENCES

Anderson, E. (1960). A semigraphical method for the analysis of complex problems. *Technometrics* 2:387–391.

Anderson, S., and Hauck, W. (1983). A new procedure for testing equivalence in comparative bioavailability and other clinical trials. *Commun. Statist. Ther. Meth.* 12:2663–2692.

Andrews, D. F. (1972). Plots of high dimensional data. *Biometrics* 28:125–136.

Bergman, S. W., and Gittins, J. C. (1985). Screening procedures for discovering active compounds. In K. Peace (Ed.), *Statistical Methods for Pharmaceutical Research Planning*. Marcel Dekker, New York.

Bickis, M. G. (1990). Experimental design. In D. L. Arnold, H. C. Grice, and D. R. Krewski (Eds.), *Handbook of In Vivo Toxicity Testing*. Academic, San Diego, pp. 128–134.

Chambers, J. M., Cleveland, W. S., Kliner, B., and Tukey, P. A. (1983). *Graphical Methods for Data Analysis*. Duxbury, Boston.

Chernoff, H. (1973). The use of faces to represent points in *K*-dimensional space graphically. *J. Am. Stat. Assoc.* 68:361–368.

Cleveland, W. S., and McGill, R. (1985). Graphical perception and graphical methods for analyzing scientific data. *Science* 229:828–833.

Davison, M. L. (1983). *Multidimensional Scaling*. Wiley, New York.

Gad, S. C. (1982a). A neuromuscular screen for use in industrial toxicology. *J. Toxicol. Environ. Health* 9:691–704.

Gad, S. C. (1982b). Statistical analysis of behavioral toxicology data and studies. *Arch. Toxicol.* 5(Suppl.):256–266.

Gad, S. C. (1988). An approach to the design and analysis of screening studies in toxicology. *J. Am. Coll. Toxicol.* 8:127–138.

Gad, S. C. (1989a). Principles of screening in toxicology with special emphasis on applications to neurotoxicology. *J. Am. Coll. Toxicol.* 8:21–27.

Gad, S. C. (1989b). Screens in neurotoxicity: Objectives, design and analysis, with the observational battery as a case example. *J. Am. Coll. Toxicol.* 8:1–18.

Gad, S. C. (1989c). Statistical analysis of screening studies in toxicology with special emphasis on neurotoxicology. *J. Am. Coll. Toxicol.* 8:171–183.

Gad, S. C. (1999). *Statistics and Experimental Design for Toxicologists*, 3rd ed. CRC Press, Boca Raton, FL.

Garrett, R. A., and London, J. P. (1970). *Fundamentals of Naval Operations Analysis*. U.S. Naval Institute, Annapolis, MD.

Healy, G. F. (1987). Power calculations in toxicology. *A.T.L.A.* 15:132–139.

Hoaglin, D. C., Mosteller, F., and Tukey, J. W. (1983). *Understanding Robust and Exploratory Data Analysis*. Wiley, New York.

Hoaglin, D. C., Mosteller, F. D., and Tukey, J. W. (1985). *Exploring Data Tables, Trends, and Shapes*. Wiley, New York.

Kass, G. V. (1980). An exploratory technique for investigating large quantities of categorical data. *Appl. Stat.* 29:119–127.

Kruskal, J. B. (1964). Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika* 29:1–27.

Martin, Y. C., Kutter, E., and Austel, V. (1988). *Modern Drug Research*. Marcel Dekker, New York, pp. 31–34, 155, 265–269, 314–318.

Montgomery, D. C. (1985). *Introduction to Statistical Quality Control*. Wiley, New York.

Nuwaysir, E. F., Bittner, M., Trent, J., Barrett, J. C., and Afshari, C. A. (1999). Microassays and toxicology: The advent of toxicogenomics. *Mol. Carcinog.* 24:153–159.

Page, N. P. (1977). Concepts of a bioassay program in environmental carcinogenesis. In H. F. Kraybill and M. A. Mehlman (Eds.), *Environmental Cancer*. Hemisphere Publishing, New York, pp. 87–171.

Pennie, W. D. (2000). Use of cDNA microassays to probe and understand the toxicological consequences of altered gene expression. *Toxicol. Lett.* 112/113:473–477.

Redman, C. (1981). Screening compounds for clinically active drugs. In C. R. Buncher and J. Tsya (Eds.), *Statistics in the Pharmaceutical Industry*. Marcel Dekker, New York, pp. 19–42.

Schmid, C. F. (1983). *Statistical Graphics*. Wiley, New York.

Tufte, E. R. (1983). *The Visual Display of Quantitative Information*. Graphic Press, Cheshire, CT.

Tukey, J. W. (1977). *Exploratory Data Analysis*. Addison-Wesley, Reading, PA.

Velleman, P. F., and Hoaglin, D. C. (1981). *Applications, Basics and Computing of Exploratory Data Analysis*. Duxbury, Boston.

Wilk, M. B., and Gnanadesikan, R. (1986). Probability plotting methods for the analysis of data. *Biometrics* 55:1–17.

Young, F. W. (1985). Multidimensional scaling. In S. Katz and N. L. Johnson (Eds.), *Encyclopedia of Statistical Sciences*, Vol. 5. Wiley, New York, pp. 649–659.

Zbinden, G., Elsner, J., and Boelsterli, U. A. (1984). Toxicological screening. *Regul. Toxicol. Pharmacol.* 4:275–286.

# *Formulations, Routes, and Dosage Design*

The perfect drug would be along the lines of Paul Eurlich's "magic bullet": As illustrated in Figure 1, a drug molecule is readily administered, completely absorbed, moves to the desired therapeutic target site, does what it is supposed to, and is completely eliminated. The most pressing (and rewarding) area for current drug development is optimizing the therapeutic target delivery part of this process. One of the key steps in the nonclinical and clinical formulation of the drug is the choice of the inactive ingredients (excipients). Excipients are essential components of drug products in the United States, and one must adequately address the safety of the proposed exposure to the excipients in those products. The specific safety data that may be needed will vary depending upon the clinical situation, including such factors as the duration, level, and route of exposure (i.e., means of patient drug administration).

Many guidances exist to aid in the development of pharmaceutical drugs, but very few guidances exist to aid in the safety evaluation of pharmaceutical excipients. The U.S. Food and Drug Administration (FDA)/Center for Drug Evaluation and Research (CDER) adopted, in 2005, the guidance for industry "Nonclinical Studies for Development of Pharmaceutical Excipients," which focuses on the development of safety profiles to support use of new excipients as components of drug or biological products.

A similar guidance was published by the International Pharmaceutical Excipients Council (IPEC), "Excipient Safety Evaluation Guidance," in 1995

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

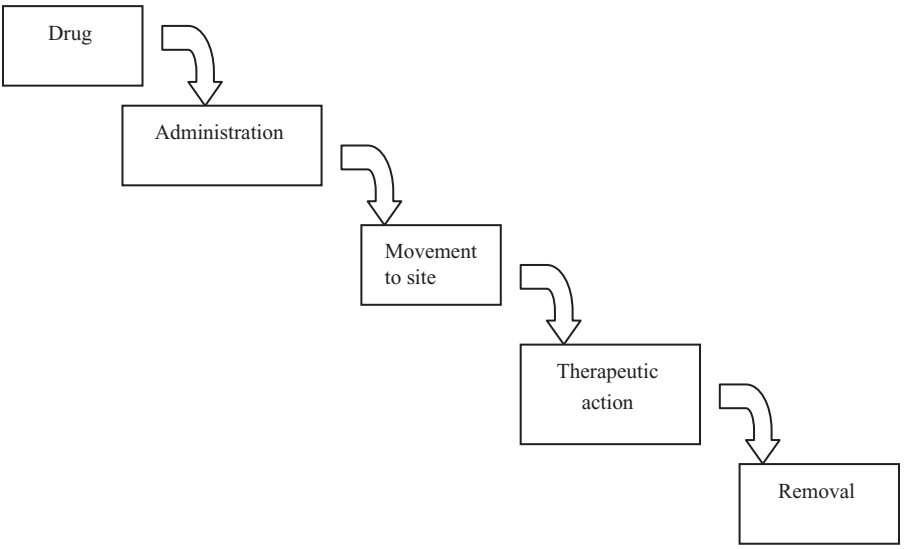


***Figure 1*** *Magic bullet: desired course of drug therapeutic development cycle.*

(updated in 2002). These guidelines are presented in a tiered approach of recommended data that should be available on an excipient to provide a pharmaceutical formulator with a rational basis for including a new excipient in a drug formulation.

The objective of the current proposal is to provide a rational approach to cover the field between "nothing is needed" and "full testing." The final aim of these safety evaluation guidelines for excipients is to provide an important element in the acceptability of a new excipient by health authorities independently of the approval of a specific drug formulation.

The three essential requirements of the active pharmaceutical ingredients (API) principles are compared with those of excipients. Fundamental for both are quality and safety. The requirement of therapeutic efficacy for drugs is replaced by that of functionality for excipient, defined as "the physical, physicochemical and biopharmaceutical properties" of the same.

Throughout the development process for pharmaceuticals, formulation development is proceeding with several objectives in mind. The importance of each of these factors changes over time (Monkhouse and Rhodes, 1998; illustrated in Figure 2). First is optimizing the bioavailability of the therapeutic target organ site by the intended clinical route. Clinical route(s) are selected on a number of grounds (nature of the drug, patient acceptance, issues of safety). Second is minimizing any safety concerns. This means not just systemic toxicity but also local tissue tolerance at the site of application. Third is optimizing stability of the drug active ingredient. Its activity and integrity must be maintained for long enough to be made effectively available to patients. Early

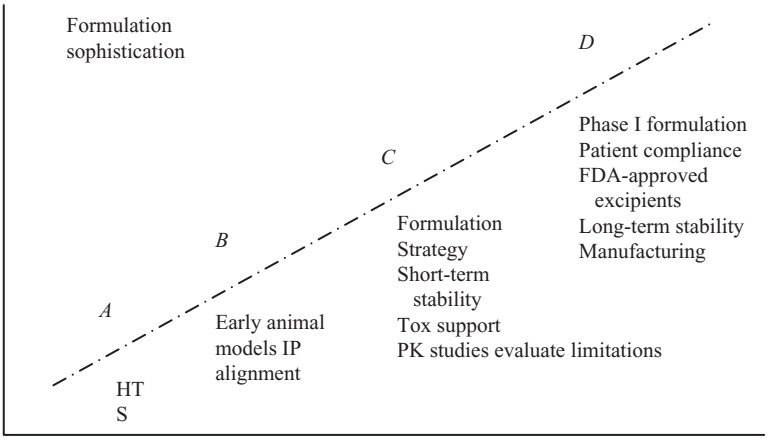


***Figure 2*** *Formulation development during course of drug development.*

on in preclinical development, simplicity and maximized bioavailability are essential. Early single-dose studies in animals are the starting place and usually bear no relationship to what is used later.

Formulations used to administer potential drugs undergoing development occupy an unusual place in pharmaceutical safety assessment compared to the rest of the industrial toxicology. Eventually, a separate function in the pharmaceutical company developing a drug will develop a specific formulation that is to be administered to people—a formulation that optimizes the conditions of absorption and stability for the drug entity (Racy, 1989). The final formulation will need to be assessed to see if it presents any unique local or short-term hazards, but as long as its nonactive constituents are drawn from the approved formulary lists, no significant separate evaluation of their safety is required preclinically. These short-term hazards can, of course, alter the toxicity of the drug under study.

Simultaneous with this development of an optimized clinical formulation, however, preclinical evaluations of the safety of the drug moiety must be performed. Separate preclinical formulations (which generally are less complex than the clinical ones) are developed, sometimes by a formulation group and other times by the toxicology group itself. These preclinical formulations will frequently include much higher concentrations of the drug moiety being tested than do any clinical formulations. The preclinical formulations are developed and evaluated with the aim of reproducibly delivering the drug (if at all possible by the route intended in humans), maintaining drug stability through an optimum period of time and occluding the observed effects of the drug with vehicle effects to the minimum extent possible. And these preclinical formulations are not restricted to materials that will (or even can) be used in final clinical formulations.

In pivotal studies, the actual blood levels of active moiety that are achieved will be determined so that correlations to later clinical studies can be made.

The formulations that are developed and used for preclinical studies are sometimes specific for the test species to be employed, but their development always starts with consideration of the route of exposure that is to be used clinically and, if possible, in accordance with a specified regimen of treatment (mirroring the intended clinical protocol as much as possible). One aspect of both nonclinical and clinical formulation and testing which prevents an important but often overlooked aspect of pharmaceutical safety assessment is the special field of excipients. These will be considered at the end of this chapter.

Among the cardinal principles of both toxicology and pharmacology is that the means by which an agent comes in contact with or enters the body (i.e., the route of exposure or administration) does much to determine the nature and magnitude of an effect. However, a rigorous understanding of formulations, routes, and their implications to the design and analysis of safety studies is not widespread. And in the day-to-day operations of performing studies in animals, such an understanding of routes, their manipulation, means and pitfalls of achieving them, and the art and science of vehicles and formulations is essential to the sound and efficient conduct of a study.

As presented in Table 1, there are at least 26 potential routes of administration, of which 10 are commonly used in safety assessment and therefore are addressed here.

## 1 MECHANISMS

There are three primary sets of reasons why differences in formulations and the route of administration are critical in determining the effect of an agent of the biological system: (1) local effects, (2) absorption and distribution, and (3) metabolism.

***Local Effects*** Local effects are those that are peculiar to the first area or region of the body to which a test material gains entry or that it contacts. For the dermal route, these include irritation, corrosion, and sensitization. For the parenteral routes, these include irritation, pyrogenicity, sterility, and blood compatibility. In general, the same categories of possible adverse effects (irritation, immediate immune response, local tissue/cellular compatibility, and physicochemical interactions) are the mechanisms of or the basis for concern.

In general, no matter what the route, certain characteristics will predispose a material to have local effects (and, by definition, if not present, tend to limit the possibility of local effects). These factors include pH, redox potential, high molar concentration, and the low flexibility and sharp edges of certain solids.

**TABLE 1 Potential Routes of Administration**

A. Oral routes
   1. Oral (PO)[a]
   2. Inhalation[a]
   3. Sublingual
   4. Buccal
B. Placed into natural orifice in body other than mouth
   1. Intranasal
   2. Intra-auricular
   3. Rectal
   4. Intravaginal
   5. Intrauterine
   6. Intraurethral
C. Parenteral (injected into body or placed under skin)
   1. Intravenous (IV)[a]
   2. Subcutaneous (SC)[a]
   3. Intramuscular (IM)[a]
   4. Intra-arterial
   5. Intradermal (ID)[a]
   6. Intralesional
   7. Epidural
   8. Intrathecal
   9. Intracisternal
   10. Intracardial
   11. Intraventricular
   12. Intraocular
   13. Intraperitoneal (IP)[a]
D. Topical routes
   1. Cutaneous[a]
   2. Transdermal (also called percutaneous)[a]
   3. Ophthalmic[a]

[a]Commonly used in safety assessment.

These characteristics will increase the potential for irritation by any route and, subsequent to the initial irritation, other appropriate regional adaptive responses (for orally administered materials, e.g., emesis and diarrhea).

***Absorption and Distribution*** For a material to be toxic, it must be absorbed into the organism (local effects are largely not true toxicities by this definition).

There are characteristics that influence absorption by the different routes, and these need to be understood by any person trying to evaluate and/or predict the toxicities of different moieties. Some key characteristics and considerations are summarized below by route.

Table 2 presents the normal pH ranges for human physiological fluids. These need to be considered in terms of the impact on solubility and stability of a formulation and active drug.

**TABLE 2 Normal pH Range for Human Physiological Fluids**

| Medium | Normal pH Range |
|---|---|
| Tears | 7.35–7.45 |
| Saliva | 6.0–8.0 |
| Gastric juice | 1.5–6.5 |
| Intestinal juice | 6.5–7.6 |
| Blood | 7.35–7.45 |
| Skin (sweat) | 4.0–6.8 |

A. Oral and rectal routes [gastrointestinal (GI) tract]
   1. Lipid-soluble compounds (nonionized) are more readily absorbed than water-soluble compounds (ionized).
      a. Weak organic bases are in the nonionized, lipid-soluble form in the intestine and tend to be absorbed there.
      b. Weak organic acids are in the nonionized, lipid-soluble form in the stomach and one would suspect that they would be absorbed there, but absorption in the intestine is greater because of time and area of exposure.
   2. Specialized transport systems exist for some moieties: sugars, amino acids, pyrimidines, calcium, and sodium.
   3. Almost everything is absorbed—at least to a small extent (if it has a molecular weight below 10,000).
   4. Digestive fluids may modify the structure of a chemical.
   5. Dilution increases toxicity because of more rapid absorption from the intestine unless stomach contents bind the moiety.
   6. Physical properties are important; for example, dissolution of metallic mercury is essential to allow its absorption.
   7. Age is important; for example, neonates have a poor intestinal barrier.
   8. Effect of fasting on absorption depends on the properties of the chemical of interest.
B. Inhalation (lungs)
   1. Aerosol deposition
      a. Nasopharyngeal—5 μm or larger in humans, less in common laboratory animals
      b. Tracheobronchial—1–5 μm
      c. Alveolar—1 μm
   2. If inhalant is a solid, mucociliary transport from lungs to GI tract may clear it out.
   3. Lungs are anatomically good for absorption.
      a. Large surface area (50–100 $m^2$)
      b. Blood flow high
      c. Close to blood (10 μm between gas media and blood)

4. Absorption of gases is dependent on solubility of the gas in blood.
   a. Chloroform, for example, has high solubility and is all absorbed, though respiration is limited.
   b. Ethylene has low solubility and only a small percentage is absorbed—blood flow limits absorption.

C. Parenteral routes
1. Irritation at the site of injection is influenced by solubility, toxicity, temperature, and pH of injected solution.
2. Pyrogenicity and blood compatibility are major concerns for intravenously administered materials.
3. Solubility of test material in an aqueous or modified aqueous solution is the chief limitation on how much material may be given intravenously.

D. Dermal routes
1. In general, any factor that increases absorption through the stratum corneum will also increase the severity of an intrinsic response. Unless this factor mirrors potential exposure conditions, it may, in turn, adversely affect the relevance of test results.
2. The physical nature of solids must be carefully considered both before testing and in interpreting results. Shape (sharp edges), size (small particles may abrade the skin due to being rubbed back and forth under the occlusive wrap), and rigidity (stiff fibers or very hard particles will be physically irritating) of solids may all enhance an irritation response and alter absorption.
3. The degree of occlusion (in fact, the tightness of the wrap over the test site) also alters percutaneous absorption and therefore irritation. One important quality control issue in the laboratory is achieving a reproducible degree of occlusion in dermal wrappings.
4. Both the age of the test animal and the application site (saddle of the back vs. flank) can markedly alter test outcome. Both of these factors are also operative in humans, of course, but in dermal irritation tests, the objective is to remove all such sources of variability. In general, as an animal ages, sensitivity to irritation decreases. And the skin on the middle of the back (other than directly over the spine) tends to be thicker (and therefore less sensitive to irritations) than that on the flanks.
5. The sex of the test animals can also alter study results, because both regional skin thickness and surface blood flow vary between males and females.

As a generalization, there is a pattern of relative absorption rates that characterizes the different routes that are commonly employed. This order of absorption (by rate from fastest to slowest and, in a less rigorous manner, by degree of absorption from most to least is intravenous (IV) > inhala-

tion > intramuscular (IM) > intraperitoneal (IP) > subcutaneous (SC) > oral > intradermal (ID) > other.

***Metabolism*** Metabolism is directly influenced both by the region a material is initially absorbed into and by distribution (both the rate and the pattern). Rate determines whether the primary enzyme systems will handle the entire xenobiotic dose or if these are overwhelmed. Pattern determines which routes of metabolism are operative.

Absorption (total amount and rate), distribution, metabolism, and species similarity in response are the reasons for selecting particular routes in toxicology in general. In safety assessment of pharmaceuticals, however, the route is usually dictated by the intended clinical route and dosing regimen. If this route of human exposure is uncertain or if there is the potential for either the number of routes or the human absorption rate and pattern being greater, then common practice becomes that of the most conservative approach. This approach stresses maximizing potential absorption in the animal species (within the limits of practicality) and selecting from among those routes commonly used in the laboratory the ones that get the most material into the animal's system as quickly and completely as possible to evaluate the potential toxicity. Under this approach, many compounds are administered intraperitoneally in acute testing, though there is little or no real potential for human exposure by this route.

Assuming that a material is absorbed, distribution of a compound in early preclinical studies is usually of limited interest. In so-called heavy acute studies (Gad et al., 1984), where acute systemic toxicity is intensive and evaluated to the point of identifying target organs, or in range-finder-type study results, for refining the design of longer term studies, distribution would be of interest. Some factors that alter distribution are listed in Table 3.

The first special case is the parenteral route, where the systemic circulation presents a peak level of the moiety of interest to the body at one time, tempered only by the results of a single pass through the liver.

The second special case arises from inhalation exposure. Because of the arrangement of the circulatory system, inhaled compounds (and those administered via the buccal route) enter the full range of systemic circulation without any "first-pass" metabolism by the liver. Kerberle (1971) and O'Reilly (1972) have previously published reviews of absorption, distribution, and metabolism that are relevant to acute testing.

## 2 COMMON ROUTES

Each of the 10 routes most commonly used in safety assessment studies has its own peculiarities, and for each there are practical considerations and techniques ("tricks") that should be either known or available to the practicing toxicologist.

**TABLE 3 Selected Factors That May Affect Chemical Distribution to Various Tissues**

A. Factors relating to chemical and its administration
   1. Degree of binding of chemical to plasma proteins (i.e., agent affinity for proteins) and tissues
   2. Chelation to calcium, which is deposited in growing bones and teeth (e.g., tetracyclines in young children)
   3. Whether chemical distributes evenly throughout body (one-compartment model) or differentially between different compartments (models of two or more compartments)
   4. Ability of chemical to cross blood–brain barrier
   5. Diffusion of chemical into tissues or organs and degree of binding to receptors that are and are not responsible for drug's beneficial effects
   6. Quantity of chemical given
   7. Route of administration or exposure
   8. Partition coefficients (nonpolar chemicals are distributed more readily to fat tissues than are polar chemicals)
   9. Interactions with other chemicals that may occupy receptors and prevent the drug from attaching to receptor, inhibit active transport, or otherwise interfere with drug's activity
   10. Molecular weight of chemical

B. Factors relating to test subject
   1. Body size
   2. Fat content (e.g., obesity affects distribution of drugs that are highly soluble in fats)
   3. Permeability of membranes
   4. Active transport for chemicals carried across cell membranes by active processes
   5. Amount of proteins in blood, especially albumin
   6. Pathology or altered homeostasis that affects any of the other factors (e.g., cardiac failure, renal failure)
   7. Presence of competitive binding substances (e.g., specific receptor sites in tissues bind drugs)
   8. pH of blood and body tissues
   9. pH of urine[a]
   10. Blood flow to various tissues or organs (e.g., well-perfused organs usually tend to accumulate more chemical than less well perfused organs)

[a]The pH of urine is usually more important than the pH of blood.

## 2.1 Dermal Route

For all agents of concern in occupational toxicology (except therapeutics), the major route by which the general population is most frequently exposed is the percutaneous (dermal) route. Brown (1980) has previously reviewed background incidence data on pesticides, for example, that show such exposures to be common. Dermal (or topical) drugs are not as common but are certainly numerous.

Percutaneous entry into the body is really by separate means (Marzulli, 1962; Scheuplein, 1965, 1967):

- Between the cells of the stratum corneum
- Through the cells of the stratum corneum

- Via the hair follicles
- Via the sweat glands
- Via the sebaceous glands

Certain aspects of the material of interest as well as those of the test animals, including absorption (Blank and Scheuplein, 1964), are as follows:

1. Small molecules penetrate skin better than large molecules.
2. Undissociated molecules penetrate skin better than do ions.
3. Preferential solubility of the toxicant in organic solvents indicates better penetration characteristics than preferential solubility in water.
4. The less viscous or more volatile the toxicant, the greater its penetrability.
5. The nature of the vehicle and the concentration of the toxicant in the vehicle both affect absorption (vehicles are discussed later in this chapter).
6. Hydration (water content) of the stratum corneum affects penetrability.
7. Ambient temperature can influence the uptake of toxicant through the skin. The warmer it is, the greater the blood flow through the skin and, therefore, the greater the percutaneous absorption.
8. Molecular shape (particularly symmetry) influences absorption (Medved and Kundiev, 1964).

There are at least two excellent texts on the subject of percutaneous absorption that go into great detail (Brandau and Lippold, 1982; Bronaugh and Maibach, 1985).

## 2.2 Parenteral Route

The parenteral routes include three major ones—IV, IM, and SC—and a number of minor routes (such as intra-arterial) that are not considered here. Administration by the parenteral routes raises a number of special safety concerns in addition to the usual systemic safety questions. These include irritation (vascular, muscular, or subcutaneous), pyrogenicity, blood compatibility, and sterility (Avis, 1985). The background of each of these, along with the underlying mechanisms and factors that influence the level of occurrence of such an effect, is discussed in Chapter Developmental and Reproductive Toxicity Testing.

The need for a rapid onset of action (and/or clearance) usually requires that an IV route be used, although at a certain stage of cardiopulmonary resuscitation (for example), the need for an even more rapid effect may require the use of an intracardiac injection. The required site of action may influence the choice of route of administration (e.g., certain radiopaque dyes are given intra-

arterially near the site being evaluated; streptokinase is sometimes injected experimentally into the coronary arteries close to coronary vessel occlusion during a myocardial infarction to cause lysis of the thrombus and therefore reestablish coronary blood flow).

The characteristics of the fluid to be injected will also influence the choice of parenteral routes. The drug must be compatible with other fluids (e.g., saline, dextrose, Ringer's lactate) with which it may be combined for administration to the patient as well as with the components of the blood itself.

There are certain clinical situations in which a parenteral route of administration is preferred to other possible routes. These include the following:

1. When the amount of drug given to a subject must be precisely controlled (e.g., in many pharmacokinetic studies), it is preferable to use a parenteral (usually IV) route of administration.
2. When the "first-pass effect" of a drug going through the liver must be avoided, a parenteral route of administration is usually chosen, although a sublingual route or dermal patch will also avoid the first-pass effect.
3. When one requires complete assurance that an uncooperative subject has actually received the drug and has not rejected it (e.g., via forced emesis).
4. When subjects are in a stupor, coma, or otherwise unable to take a drug orally.
5. When large volumes (i.e., more than a liter) of fluid are injected (such as in peritoneal dialysis, hyperalimentation, fluid replacement, and other conditions). Special consideration of fluid balance must be given to patients receiving large volumes as well as careful consideration of the systemic effects of injection fluid components (e.g., amino acids and their nephrotoxicity).

The three significant parenteral routes we are concerned with here and their specific set of advantages and disadvantages or specific considerations that must be kept in mind will be discussed next.

***Intravenous Route*** The IV route is the most common method of introducing a drug directly into the systemic circulation (Lukas et al., 1971). It has the following advantages:

1. Rapid onset of effect
2. Usefulness in situations of poor GI absorption
3. Avoidance of tissue irritation that may be present IM or other routes (e.g., nitrogen mustard)
4. More precise control of levels of drug than with other routes, especially of toxic drugs, where the levels must be kept within narrow limits

5. Ability to administer large volumes over time by a slow infusion
6. Ability to administer drugs at a constant rate over a long period of time

It also suffers from disadvantages:

1. Higher incidence of anaphylactic reactions than with many other routes
2. Possibility of infection or phlebitis at site of injection
3. Greater pain to patients than with many other routes
4. Possibility that embolic phenomena may occur—either air embolism or vascular clot—as a result of damage to the vascular wall
5. Impossibility of removing or lavaging drug after it is given, except by dialysis
6. Inconvenience in many situations
7. Possibility that rapid injection rates may cause severe adverse reactions
8. Patient dislike of and psychological discomfort with the injection procedure

For IV fluids, it must be determined how the dose will be given (i.e., by bolus or slow injection, intermittent or constant infusion, or constant drip) and whether special equipment will be used to control and monitor the flow. Drugs with short half-lives are usually given by a constant drip or infusion technique. All IV fluids given immediately subsequent to an IV drug must be evaluated for their compatibility with the study drug. Suspensions are generally not given intravenously because of the possibility of blocking the capillaries.

In the IV route, anaphylactic reactions (caused by administration of an agent to an animal previously sensitized to it or to a particularly sensitive species such as a guinea pig) may be especially severe—probably because of sudden, massive antigen–antibody reactions. When the drug is given by other routes, its access to antibody molecules is necessarily slower; moreover, its further absorption can be retarded or prevented at the first sign of a serious allergic reaction.

Embolism is another possible complication of the IV route. Particulate matter may be introduced if a drug intended for IV use precipitates for some reason or if a particular suspension intended for IM or SC use is inadvertently given into a vein. Hemolysis or agglutination of erythrocytes may be caused by injection of hypotonic hypertonic solutions or by more specific mechanisms (Gray, 1978).

***Bolus versus Infusion*** Technically, for all the parenteral routes (but in practice only for the IV route), there are two options for injecting a material into the body. The bolus and infusion methods are differentiated on the single basis of rate of injection, but they actually differ on a wide range of characteristics.

The most commonly exercised option is the bolus, "push," injection, in which the injection device (syringe or catheter) is appropriately entered into the vein and a defined volume of material is introduced through the device. The device is then removed. In this operation, it is relatively easy to restrain an experimental animal and the stress on the animal is limited. Though the person doing the injection must be skilled, it takes only a short amount of time to become so. And the one variable to be controlled in determining dosage is the total volume of material injected (assuming dosing solutions have been properly prepared) (Theeuwes and Yum, 1976). See Chapter Immunotoxicology in Drug Development for a more complete discussion.

***Subcutaneous Route*** Drugs given by the SC route are forced into spaces between connective tissues, as with IM injections. Vasoconstrictors and drugs that cause local irritation should not be given subcutaneously under usual circumstances, since inflammation, abscess formation, or even tissue necrosis may result. When daily or even more frequent SC injections are made, the site of injection should be continually changed to prevent local complications. Fluids given subcutaneously must have an appropriate tonicity to prevent pain. Care must be taken to prevent injection of the drug directly into veins.

The absorption of drugs from a SC route is influenced by blood flow to the area, as with IM injections. The rate of absorption may be retarded by cooling the local area to cause vasoconstriction, adding epinephrine to the solution for the same purpose (e.g., with local anesthetics), decreasing blood flow with a tourniquet, or immobilizing the area. The opposite effect may be achieved by warming the injection region or by using the enzyme hyaluronidase, which breaks down mucopolysaccharides of the connective tissue matrix to allow the injected solution to spread over a larger area and thus increase its rate of absorption.

Absorption from SC injection sites is affected by the same factors that determine the rate of absorption from IM sites (Schou, 1971). Blood flow through these regions is generally poorer than in muscles, so the absorption rate is generally slower.

The rate of absorption from an SC injection site may be retarded by immobilization of the limb, local cooling to cause vasoconstriction, or application of a tourniquet proximal to the injection site to block the superficial venous drainage and lymphatic flow. In small amounts, adrenergic stimulants, such as epinephrine, will constrict the local blood vessels and therefore slow systemic absorption. Conversely, cholinergic stimulants (such as methacholine) will induce very rapid systemic absorption subcutaneously. Other agents may also alter their own rate of absorption by affecting local blood supply or capillary permeability.

A prime determinant of the absorption rate from an SC injection is the total surface area over which the absorption can occur. Although the SC tissues are somewhat loose and moderate amounts of fluid can be administered, the normal connective tissue matrix prevents indefinite lateral spread

of the injected solution. These barriers may be overcome by agents that break down mucopolysaccharides of the connective tissue matrix; the resulting spread of injected solution leads to a much faster absorption rate.

In addition to fluids, solid forms of drugs may be given by SC injection. This has been done with compressed pellets of testosterone placed under the skin which are absorbed at a relatively constant rate over a long period.

***Intramuscular Route*** The IM route is frequently used for drugs dissolved in oily vehicles or for those in a microcrystalline formulation that are poorly soluble in water (e.g., procaine or penicillin G). Advantages include rapid absorption (often in under 30 min), the opportunity to inject a relatively large amount of solution, and a reduction in pain and local irritation compared with SC injections. Potential complications include infections and nerve damage. The latter usually results from the choice of an incorrect site for injection.

Although the time to peak drug concentration is often on the order of 1–2 h, depot preparations given by IM injection are absorbed extremely slowly. Numerous physiochemical properties of a material given intramuscularly will affect the rate of absorption from the site within the muscle (e.g., ionization of the drug, lipid solubility, osmolality of the solution, volume given). The primary sites used for IM injections in people are the gluteal (buttocks), deltoid (upper arm), and lateral vastus (lateral thigh) muscles, with the corresponding sites in test animals being species specific. The rate of drug absorption and the peak drug levels obtained will often differ between sites used for IM injections because of differences in blood flow between muscle groups. The site chosen for an IM injection in humans and some animals may be a critical factor in whether or not the drug exhibits an effect (Schwartz et al., 1974). Agents injected into the larger muscle masses are generally absorbed rapidly.

Blood flow through muscles in a resting animal is about 0.02–0.07 mL $min^{-1}$ $g^{-1}$ of tissue, and this flow rate may increase many times during exercise, when additional vascular channels open. Large amounts of solution can be introduced intramuscularly, and there is usually less pain and local irritation than is encountered by the SC route. Ordinary aqueous solutions of chemicals are usually absorbed from an IM site within 10–30 min, but faster or slower absorption is possible, depending on the vascularity of the site, the ionization and lipid solubility of the drug, the volume of the injection, the osmolality of the solution, animal temperature, and other variables. Small molecules are absorbed directly into the capillaries from an IM site, whereas large molecules (e.g., proteins) gain access to the circulation by way of the lymphatic channels (Ballard, 1968). Radiolabeled compounds of widely differing molecular weights (maximum 585) and physical properties have been shown to be absorbed from rat muscle at virtually the same rate, about 16% per minute (i.e., the absorption process is limited by the rate of blood flow).

Drugs that are insoluble at tissue pH or that are in an oily vehicle form a depot in the muscle tissue, from which absorption proceeds very slowly.

***Intraperitoneal Route*** Kruger et al. (1962) demonstrated the efficiency of absorption of some chemicals injected intraperitoneally, while Lukas et al. (1971) showed that compounds administered intraperitoneally are absorbed primarily through the portal circulation.

A prime practical consideration in the use of the IP route for acute testing should be the utilization of aseptic techniques to preclude bacterial or viral contamination. If these are not exercised, the resulting infected and compromised animals cannot be expected to produce either valid or reproducible indications or actual chemical toxicity.

Compounds that are very lipophilic will be quickly absorbed systemically by the IP route but not by the IM or SC route.

### 2.3 Oral Route

The oral route is the most commonly used route for the administration of drugs both because of ease of administration and because it is the most readily accepted route of administration. Although the dermal route may be as common for occupational exposure, it is much easier to accurately measure and administer doses by the oral route.

Enteral routes technically include any that will put a material directly into the GI tract, but the use of enteral routes other than oral (such as rectal) is rare in toxicology. Though there are a number of variations of technique and peculiarities of animal response that are specific to different animal species, there is also a great deal of commonality across species in methods, considerations, and mechanisms.

***Mechanisms of Absorption*** Ingestion is generally referred to as oral or peroral (PO) exposure and includes direct intragastric exposure in experimental toxicology. The regions for possible agent action and absorption from PO absorption should, however, be considered separately (Hogben et al., 1959; Bates and Gibaldi, 1970; Gad 2007, 2008).

Because of the rich blood supply to the mucous membranes of the mouth (buccal cavity), many compounds can be absorbed through them. Absorption from the buccal cavity is limited to nonionized, lipid-soluble compounds. Buccal absorption of a wide range of aromatic and aliphatic acids and basic drugs in human subjects has been found to be parabolically dependent on log $P$, where $P$ is the octanol–water partition coefficient. The ideal lipophilic character (log $P_0$) for maximum buccal absorption has also been shown to be in the range 4.2–5.5 (Schranker et al., 1957; Lien et al., 1971). Compounds with large molecular weights are poorly absorbed in the buccal cavity, and, since absorption increases linearly with concentration and there is generally no difference between optical enantiomorphs of several compounds known to be absorbed from the mouth, it is believed that uptake is by passive diffusion rather than by active transport chemical moieties.

A knowledge of the buccal absorption characteristics of a chemical can be important in a case of accidental poisoning. Although an agent taken into the

mouth will be voided immediately on being found objectionable, it is possible that significant absorption can occur before any material is swallowed.

Unless voided, most materials in the buccal cavity are swallowed. No significant absorption occurs in the esophagus and the agent passes on to enter the stomach. It is common practice in safety assessment studies to avoid the possibility of buccal absorption by intubation (gavage) or by the administration of the agent in gelatin capsules designed to disintegrate in the gastric fluid.

Absorption of chemicals with widely differing characteristics can occur at different levels in the GI tract (Schranker, 1960). The two factors primarily influencing this regional absorption are (1) the lipid–water partition characteristics of the undissociated toxicant and (2) the dissociation constant $pK_a$, which determines the amount of toxicant in the dissociated form.

Therefore, weak organic acids and bases are readily absorbed as uncharged lipid-soluble molecules, whereas ionized compounds are absorbed only with difficulty, and nonionized toxicants with poor lipid solubility characteristics are absorbed slowly. Lipid-soluble acid molecules can be absorbed efficiently through the gastric mucosa, but bases are not absorbed in the stomach.

In the intestines the nonionized form of the drug is preferentially absorbed and the rate of absorption is related to the lipid–water partition coefficient of the toxicant. The highest $pK_a$ value for a base compatible with efficient gastric absorption is about 7.8 and the lowest $pK_a$ for an acid is about 3.0, although a limited amount of absorption can occur outside these ranges (Share et al., 1971). The gastric absorption and the intestinal absorption of a series of compounds with different carbon chain lengths follow two different patterns. Absorption from the stomach increases as the chain lengthens from methyl to *n*-hexyl, whereas intestinal absorption increases over the range methyl to *n*-butyl and then diminishes as the chain length further increases. Houston et al. (1974) concluded that to explain the logic of optimal partition coefficients for intestinal absorption it was necessary to postulate a two-compartment model with a hydrophilic barrier and a lipoidal membrane and that if there is an acceptable optimal partition coefficient for gastric absorption it must be at least 10 times greater than the corresponding intestinal value.

Because they are crucial to the course of an organism's response, the rate and extent of absorption of biologically active agents from the GI tract also have major implications for the formulation of test material dosages and also for how production (commercial) materials may be formulated to minimize potential accidental intoxications while maximizing the therapeutic profile.

There are a number of separate mechanisms involved in absorption from the GI tract, and these will be discussed below.

*Passive Absorption* The membrane lining of the tract has a passive role in absorption. As toxicant molecules move from the bulk water phase of the intestinal contents into the epithelial cells, they must pass through two membranes in series, one the layer of water and the other the lipid membrane of the microvillar surface (Wilson and Dietschy, 1974). The water layer may be

the rate limiting factor for passive absorption into the intestinal mucosa, but it is not rate limiting for active absorption. The concentration gradient and the physiochemical properties of the drug and the lining membrane are the controlling factors. Chemicals that are highly lipid soluble are capable of passive diffusion, and they pass readily from the aqueous fluids of the gut lumen through the lipid barrier of the intestinal wall and into the bloodstream. The interference in the absorption process by the water layer increases with increasing absorbability of the substances in the intestine (Winne, 1978).

Aliphatic carbamates are rapidly absorbed from the colon by passive uptake (Wood et al., 1978), and it is found that there is a linear relationship between log $k_a$ and log $P$ for absorption of these carbamates in the colon and the stomach, whereas there is a parabolic relationship between these two values for absorption in the small intestine. The factors to be considered are

Octanol–buffer partition coefficient $P$
Absorption rate constant $k_a$
Time $t$
Half-life $t^{1/2} = \ln 2/k_a$

Organic acids that are extensively ionized at intestinal pH's are absorbed primarily by simple diffusion.

*Facilitated Diffusion* Temporary combination of the chemical with some form of "carrier" occurs in the gut wall, facilitating the transfer of the toxicant across the membranes. This process is also dependent on the concentration gradient across the membrane, and there is no energy utilization in making the translocation. In some intoxications, the carrier may become saturated, making this the rate-limiting step in the absorption process.

*Active Transport* As above, the process depends on a carrier but differs in that the carrier provides energy for translocation from regions of lower concentration to regions of higher concentration.

*Pinocytosis* This process, by which particles are absorbed, can be an important factor in the ingestion of particulate formulations of chemicals (e.g., dust formulations, suspensions of wettable powders); however, it must not be confused with absorption by one of the above processes, where the agent has been released from particles.

*Absorption via Lymphatic Channels* Some lipophilic chemicals dissolved in lipids may be absorbed through the lymphatics.

*Convective Absorption* Compounds with molecular radii of less than 4 nm can pass through pores in the gut membrane. The membrane exhibits a molecular sieving effect.

Characteristically, within certain concentration limits, if a chemical is absorbed by passive diffusion, then the concentration of toxicant in the gut and the rate of absorption are linearly related. However, if absorption is mediated by active transport, the relationship between concentration and rate of absorption conforms to Michaelis–Menten kinetics and a Lineweaver–Burk plot (i.e., reciprocal of rate of absorption plotted against reciprocal of concentration), which graphs as a straight line.

Differences in the physiological chemistry of GI fluids can have a significant effect on toxicity. Both physical and chemical differences in the GI tract can lead to species differences in susceptibility to acute intoxication. The antihelminthic pyrvinium chloride has an identical median lethal dose ($LD_{50}$) value when administered intraperitoneally to rats and mice (approximately 4 mg $kg^{-1}$); when administered orally, however, the $LD_{50}$ value in mice was found to be 15 mg $kg^{-1}$, while for the rat, the $LD_{50}$ values were 430 mg $kg^{-1}$ for females and 1550 mg $kg^{-1}$ for males. It is thought that this is an absorption difference rather than a metabolic difference (Ritschel et al., 1974).

Most exogenous chemical absorbed from the GI tract must pass through the liver via the hepatic–portal system (leading to the so-called first-pass effect) and, as mixing of the venous blood with arterial blood from the liver occurs, consideration and caution are called for in estimating the amounts of chemical in both the systemic circulation and the liver itself.

Despite the GI absorption characteristics discussed above, it is common for absorption from the alimentary tract to be facilitated by dilution of the toxicant. Borowitz et al. (1971) have suggested that the concentration effects they observed in atropine sulfate, aminopyrine, sodium salicylate, and sodium pentoparbital were due to a combination of rapid stomach emptying and the large surface area for absorption of the drugs.

Major structural or physiological differences in the alimentary tract (e.g., species differences or surgical effects) can give rise to modifications of toxicity. For example, ruminant animals may metabolize toxicants in the GI tract in a way that is unlikely to occur in nonruminants.

The presence of bile salts in the alimentary tract can affect absorption of potential toxicants in a variety of ways, depending on their solubility characteristics.

***Factors Affecting Absorption*** Test chemicals are given most commonly by mouth. This is certainly the most convenient route, and it is the only one of practical importance for self-administration. Absorption, in general, takes place along the whole length of the GI tract, but the chemical properties of each molecule determine whether it will be absorbed in the strongly acidic stomach or in the nearly neutral intestine. Gastric absorption is favored by an empty stomach, in which the chemical, in undiluted gastric juice, will have good access to the mucosal wall. Only when a chemical would be irritating to the gastric mucosa is it rational to administer it with or after a meal. However, the antibiotic griseofulvin is an example of a substance with poor water solubility

the absorption of which is aided by a fatty meal. The large surface area of the intestinal villi, the presence of bile, and the rich blood supply all favor intestinal absorption of griseofulvin and physiochemically similar compounds.

The presence of food can impair the absorption of chemicals given by mouth. Suggested mechanisms include reduced mixing, complexing with substances in the food, and retarded gastric emptying. In experiments with rats, prolonged fasting has been shown to diminish the absorption of several chemicals, possibly by deleterious effects upon the epithelium of intestinal villi.

Chemicals that are metabolized rapidly by the liver cannot be given for systemic effect by the enteral route because the portal circulation carries them directly to the liver. For example, lidocaine, a drug of value in controlling cardiac arrhythmias, is absorbed well from the gut but is completely inactivated in a single passage through the liver.

The principles governing the absorption of drugs from the GI lumen are the same as for the passage of drugs across biological membranes elsewhere. Lower degree of ionization, high lipid–water partition coefficient of nonionized form, and small atomic or molecular radii of water-soluble substances all favor rapid absorption. Water passes readily in both directions across the wall of the GI lumen. Sodium ion is probably transported actively from lumen into blood. Magnesium ion is very poorly absorbed and therefore acts as a cathartic, retaining an osmotic equivalent of water as it passes down the intestinal tract. Ionic iron is absorbed as an amino acid complex at a rate usually determined by the body's need for it. Glucose and amino acids are transported across the intestinal wall by specific carrier systems. Some compounds of high molecular weight (polysaccharides and large proteins) cannot be absorbed until they are degraded enzymatically. Other substances cannot be absorbed because they are destroyed by GI enzymes—insulin, epinephrine, and histamine are examples. Substances that form insoluble precipitates in the GI lumen or that are insoluble either in water or in lipid clearly cannot be absorbed.

*Absorption of Weak Acids and Bases* Human gastric juice is very acid (about pH 1), whereas the intestinal contents are nearly neutral (actually very slightly acid). The pH difference between plasma (pH 7.4) and the lumen of the GI tract plays a major role in determining whether a drug that is a weak electrolyte will be absorbed into plasma or excreted from plasma into the stomach or intestine. For practical purposes, the mucosal lining of the GI tract is impermeable to the ionized form of a weak acid or base, but the nonionized form equilibrates freely. The rate of equilibration of the nonionized molecule is directly related to its lipid solubility. If there is a pH difference across the membrane, then the fraction ionized may be considerably greater on one side than on the other. At equilibrium, the concentration of the nonionized moiety will be the same on both sides, but there will be more total drug on the side where the degree of ionization is greater. This mechanism is known as *ion trapping*. The energy for sustaining the unequal chemical potential of the acid

or base in question is derived from whatever mechanism maintains the pH difference. In the stomach, this mechanism is the energy-dependent secretion of hydrogen ions.

Consider how a weak electrolyte is distributed across the gastric mucosa between plasma (pH 7.4) and gastric fluid (pH 1.0). In each compartment, the Henderson–Hasselbalch equation gives the ratio of acid–base concentrations. The negative logarithm of the acid dissociation constant is designated here by the symbol $pK_a$ rather than the more precisely correct $pK$[1]:

$$\mathrm{pH} = \mathrm{p}K_a + \log\left(\frac{\text{base}}{\text{acid}}\right)$$

$$\log\left(\frac{\text{base}}{\text{acid}}\right) = \mathrm{pH} - \mathrm{p}K_a$$

$$\frac{\text{Base}}{\text{Acid}} = \text{antilog}(\mathrm{pH} - \mathrm{p}K_a)$$

The implications of the above equations are clear. Weak acids are readily absorbed from the stomach. Weak bases are not absorbed well; indeed, they would tend to accumulate within the stomach at the expense of agent in the bloodstream. Naturally, in the more alkaline intestine, bases would be absorbed better, acids more poorly.

It should be realized that although the principles outlines here are correct, the system is dynamic, not static. Molecules that are absorbed across the gastric or intestinal mucosa are removed constantly by blood flow; thus, simple reversible equilibrium across the membrane does not occur until the agent is distributed throughout the body.

Absorption from the stomach, as determined by direct measurements, conforms, in general, to the principles outlined above. Organic acids (as indeed many of the drug molecules) are absorbed well since they are all almost completely nonionized at the gastric pH; indeed, many of these substances are absorbed faster than ethyl alcohol, which had long been considered one of the few compounds that were absorbed well from the stomach. Strong acids whose $pK_a$ values lie below 1, which are ionized even in the acid contents of the stomach, are not absorbed well. Weak bases are absorbed only negligibly, but their absorption can be increased by raising the pH of the gastric fluid.

As for bases, only the weakest are absorbed to any appreciable extent at normal gastric pH, but their absorption can be increased substantially by neutralizing the stomach contents. The quaternary cations, however, which are charged at all pH values, are not absorbed at either pH.

The accumulation of weak bases in the stomach by ion trapping mimics a secretory process; if the drug is administered systemically, it accumulates in the stomach. Dogs given various drugs intravenously by continuous infusion to maintain a constant drug level in the plasma had the gastric contents

sampled by means of an indwelling catheter. The results showed that stronger bases ($pK_a > 5$) accumulated in stomach contents to many times their plasma concentrations; the weak bases appeared in about equal concentrations in gastric juice and in plasma. Among the acids, only the weakest appeared in detectable amounts in the stomach. One might wonder why the strong bases, which are completely ionized in gastric juice and whose theoretical concentration ratios (gastric juice/plasma) are very large, should nevertheless attain only about a 40-fold excess over plasma. Direct measurements of arterial and venous blood show that essentially all the blood flowing through the gastric mucosa is cleared of these agents; obviously, no more chemical can enter the gastric juice in a given time period than is brought there by circulation. Another limitation comes into play when the base $pK_a$ exceeds 7.4; now a major fraction of the circulating base is cationic and a decreasing fraction is nonionized, so the effective concentration gradient for diffusion across the stomach wall is reduced.

The ion-trapping mechanism provides a method of some forensic value for detecting the presence of alkaloids (e.g., narcotics, cocaine, amphetamines) in cases of death suspected to be due to overdosage of self-administered drugs. Drug concentrations in gastric contents may be very high even after parenteral injection.

Absorption from the intestine has been studied by perfusing drug solutions slowly through rat intestine in situ and by varying the pH as desired. The relationships that emerge from such studies are the same as those for the stomach, the difference being that the intestinal pH is normally very near neutrality. As the pH is increased, the bases are absorbed better, the acids more poorly. Detailed studies with a great many drugs in unbuffered solutions revealed that in the normal intestine acids with $pK_a > 3.0$ and bases with $pK_a < 7.8$ are very well absorbed; outside these limits the absorption of acids and bases falls off rapidly. This behavior leads to the conclusion that the "virtual pH" in the microenvironment of the absorbing surface in the gut is about 5.3; this is somewhat more acidic than the pH in the intestinal lumen is usually considered to be.

Absorption from the buccal cavity has been shown to follow exactly the same principles as those described for absorption from the stomach and intestine. The pH of human and canine saliva is usually about 6. Bases in people are absorbed only on the alkaline side of their $pK_a$, that is, only in the nonionized form. At normal saliva pH, only weak bases are absorbed to a significant extent.

***Bioavailability and Thresholds*** The difference between the extent of availability (often designated solely as bioavailability) and the rate of availability is illustrated in Figure 3, which depicts the concentration–time curve for a hypothetical agent formulated into three different dosage forms. Dosage forms A and B are designed so that the agent is put into the blood circulation at the same rate but twice as fast as for dosage form C. The times at which

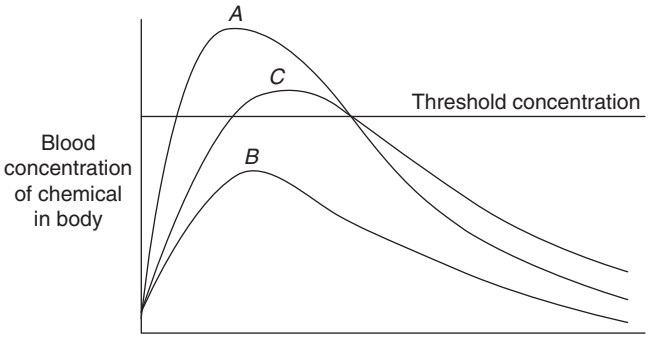


***Figure 3*** *Blood concentration–time curves illustrating how changes in rate and extent of chemical availability can influence duration of action and efficacy of dose of agent. The designated line indicates the threshold concentration ($T_c$) of the agent in the body that will evoke a response. Case* A *is absorbed rapidly and completely. This product produces a prompt and prolonged response. The agent in case* B *is absorbed at the same rate as that in case* A *but is only 50% as available. There will be no response from this dose of the agent because* $T_c$ *is not reached. The agent in case* C *is absorbed at one-half the rate seen in cases* A *and* B *but is 100% available.*

agent concentrations reach a peak are identical for dosage forms A and B and occur earlier than the peak time for dosage form C. In general, the relative order of peak times following the administration of different dosage forms of the drug corresponds to the rates of availability of the chemical moiety from the various dosage forms. The extent of availability can be measured by using either chemical concentrations in the plasma or blood or amounts of unchanged chemical in the urine. The area under the blood concentration–time curve for an agent can serve as a measure of the extent of its availability. In Figure 3, the areas under curves *A* and *C* are identical and twice as great as the area under curve *B*. In most cases, where clearance is constant, the relative areas under the curves or the amount of unchanged chemical excreted in the urine will quantitatively describe the relative availability of the agent from the different dosage forms. However, even in nonlinear cases, where clearance is dose dependent, the relative areas under the curves will yield a measurement of the rank order of availability from different dosage forms or from different routes of administration.

Because there is usually a critical concentration of a chemical in the blood that is necessary to elicit either a pharmacological or toxic effect, both the rate and extent of input or availability can alter the toxicity of a compound. In the majority of cases, the duration of effects will be a function of the length of time the blood concentration curve is above the threshold concentration; the intensity of the effect for many agents will be a function of the elevation of the blood concentration curve above the threshold concentration.

Thus, the three different dosage forms depicted in Figure 3 will exhibit significant differences in their levels of "toxicity." Dosage form B requires that twice the dose be administered to attain blood levels equivalent to those for

dosage form A. Differences in the rate of availability are particularly important for agents given acutely. Dosage for A reaches the target concentration earlier than chemical from dosage form C; concentrations from A reach a higher level and remain above the minimum effect concentration for a longer period of time. In a multiple-dosing regimen, dosage forms A and C will yield the same average blood concentrations, although dosage form A will show somewhat greater maximum and lower minimum concentrations.

For most chemicals, the rate of disposition or loss from the biological system is independent of rate of input once the agent is absorbed. Disposition is defined as what happens to the active molecule after it reaches a site in the blood circulation where concentration measurements can be made (the systemic circulations, generally). Although disposition processes may be independent of input, the inverse is not necessarily true, because disposition can markedly affect the extent of availability. Agents absorbed from the stomach and the intestine must first pass through the liver before reaching the general circulation (Figure 4). Thus, if a compound is metabolized in the liver or excreted in bile, some of the active molecule absorbed from the GI tract will be inactivated by hepatic processes before it can reach the systemic circulation and be distributed to its sites of action. If the metabolizing or biliary excreting capacity of the liver is great, the effect on the extent of availability will be substantial. Thus, if the hepatic blood clearance for the chemical is large relative to hepatic blood flow, the extent of availability for this chemical will be low when it is given by a route that yields first-pass metabolic effects. This

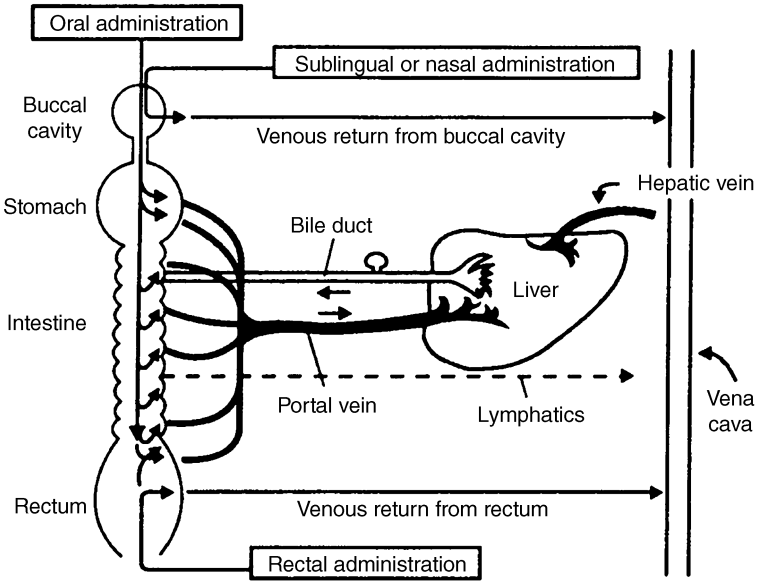


***Figure 4*** *Diagrammatic representation of entry of drug moieties into body from variety of routes, with subsequent passage into bloodstream and out of body.*

decrease in availability is a function of the physiological site from which absorption takes place, and no amount of modification to dosage form can improve the availability under linear conditions. Of course, toxic blood levels can be reached by this route of administration if larger doses are given.

It is important to realize that chemicals with high extraction ratios (i.e., greater extents of removal by the liver during first-pass metabolism) will exhibit marked intersubject variability in bioavailability because of variations in hepatic function or blood flow or both. For the chemical with an extraction ratio of 0.90 that increases to 0.95, the bioavailability of the agent will be halved, from 0.10 to 0.05. These relationships can explain the marked variability in plasma or blood drug concentrations that occurs among individual animals given similar doses of a chemical that is markedly extracted. Small variations in hepatic extraction between individual animals will result in large differences in availability and plasma drug concentrations.

The first-pass effect can be avoided, to a great extent, by use of the sublingual route and by topical preparations (e.g., nitroglycerine ointment), and it can be partially avoided by using rectal suppositories. The capillaries in the lower and middle sections of the rectum drain into the interior and middle hemorrhoidal veins, which in turn drain into the inferior vena cava, thus bypassing the liver. However, suppositories tend to move upward in the rectum into a region where veins that lead to the liver predominate, such as the superior hemorrhoidal vein. In addition, there are extensive connections between the superior and middle hemorrhoidal veins, and thus probably only about 50% of a rectal dose can be assumed to bypass the liver. The lungs represent a good temporary clearing site for a number of chemical (especially basic) compounds by partition into lipid tissues as well as serve a filtering function for particulate matter that may be given by IV injection. In essence, the lung may cause first-pass loss by excretion and possible metabolism for chemicals input into the body by the non-GI routes of administration.

Biological (test subject) factors that can influence absorption of a chemical from the GI tract are summarized in Table 4.

There are also a number of chemical factors that may influence absorption from the GI tract. These are summarized in Table 5.

***Techniques of Oral Absorption*** There are three major techniques for oral delivery of drugs to test animals. The most common way is by gavage, which requires that the material be in a solution or suspension for delivery by tube to the stomach. Less common materials may be given as capsules (particularly to dogs) or in diet (for longer term studies). Rarely, oral studies may also be done by inclusion of materials in drinking water.

Test materials may be administered as solutions or suspensions as long as they are homogeneous and delivery is accurate. For traditional oral administration (gavage), the solution or suspension can be administered with a suitable stomach tube or feeding needle ("Popper" tube) attached to a syringe. If the dose is too large to be administered at one time, it can be divided into

**TABLE 4 Test Subject Characteristics That Can Influence GI Tract Absorption[a]**

A. General and inherent characteristics
   1. General condition of subject (e.g., starved versus well fed, ambulatory versus supine)
   2. Presence of concurrent diseases (i.e., diseases may either speed or slow gastric emptying)
   3. Age
   4. Weight and degree of obesity
B. Physiological function
   1. Status of subject's renal function
   2. Status of subject's hepatic function
   3. Status of subject's cardiovascular system
   4. Status of subject's GI motility and function (e.g., ability to swallow)
   5. pH of gastric fluid (e.g., affected by fasting, disease, food intake, drugs)
   6. Gastrointestinal blood flow to area of absorption
   7. Blood flow to areas of absorption for dose forms other than those absorbed through GI routes
C. Acquired characteristics
   1. Status of subject's anatomy (e.g., previous surgery)
   2. Status of subject's GI flora
   3. Timing of drug administration relative to meals (i.e., presence of food in GI tract)
   4. Body position of subject (e.g., lying on side slows gastric emptying)
   5. Psychological state of subject (e.g., stress increases gastric emptying rate and depression decreases rate)
   6. Physical exercise of subject may reduce gastric-emptying rate
D. Physiological principles
   1. Food enhances gastric blood flow, which should theoretically increase rate of absorption
   2. Food slows the rate of gastric emptying, which should theoretically slow the rate of passage to the intestines where the largest amounts of most agents are absorbed. This should decrease the rate of absorption for most agents. Agents absorbed to a larger extent in the stomach will have increased time for absorption in the presence of food and should be absorbed more completely than in fasted patients
   3. Bile flow and secretion are stimulated by fats and certain other foods. Bile salts may enhance or delay absorption depending on whether they form insoluble complexes with drugs or enhance the solubility of agents
   4. Changes in splanchnic blood flow as a result of food depend on direction and magnitude of the type of food ingested
   5. Presence of active (saturable transport mechanisms places a limit on the amount of a chemical that may be absorbed

[a]The minimization of variability due to these factors rests on the selection of an appropriate animal model, careful selection of healthy animals, and use of proper techniques.

equal subparts with 2–4 h between each administration; however, this subdivided dosing approach should generally be avoided.

Test chemicals placed into any natural orifice exert local effects and, in many instances, systemic effects as well. The possibility of systemic effects occurring when local effects are to be evaluated should be considered.

For routes of administration in which the chemical is given orally or placed into an orifice other than the mouth, clear instructions about the correct administration of the chemical must be provided. Many cases are known of oral pediatric drops for ear infections being placed into the ear, and vice versa (ear drops being swallowed) in humans. Errors in test article administration

**TABLE 5 Chemical Characteristics of Drug That May Influence Absorption**

A. Administration of chemical and its passage through the body
   1. Dissolution characteristics of solid dosage forms, which depend on formulation in addition to the properties of the chemical itself (e.g., vehicle may decrease permeability of suspension or capsule to water and retard dissolution and diffusion).
   2. Rate of dissolution in GI fluids. Chemicals that are inadequately dissolved in gastric contents may be inadequately absorbed.
   3. Chemicals that are absorbed into food may have a delayed absorption.
   4. Carrier-transported chemicals are more likely to be absorbed in the small intestine.
   5. Route of administration.
   6. Chemicals undergo metabolism in the GI tract.

B. Physiochemical properties of chemicals
   1. Chemicals that chelate metal ions in food may form insoluble complexes and will not be adequately absorbed.
   2. pH of dosing solutions—weakly basic solutions are absorbed to a greater degree in the small intestine.
   3. Salts used.
   4. Hydrates or solvates.
   5. Crystal form of chemical (e.g., insulin).
   6. "Pharmaceutical" form (e.g., fluid, solid, suspension).
   7. Enteric coating.
   8. Absorption of quaternary compounds (e.g., hexamethonium, amiloride) is decreased by food.
   9. Molecular weight of chemical (e.g., when the molecular weight of a drug is above about 1000, absorption is markedly decreased).
   10. p$K_a$ (dissociation constant).
   11. Lipid solubility (i.e., a hydrophobic property relating to penetration through membranes).
   12. Particle size of chemical in solid dosage form—smaller particle sizes will increase the rate and/or degree of absorption if dissolution of the chemical is the rate-limiting factor in absorption. Chemicals that have a low dissolution rate may be made in a micronized form to increase their rate of dissolution.
   13. Particle size of the dispersed phase in an emulsion.
   14. Type of disintegrating agent in the formulation.
   15. Hardness of a solid (granule, pellet, or tablet) (i.e. related to amount of compression used to make tablet) or capsule if they do not disintegrate appropriately.

are especially prevalent when a chemical form is being used in a nontraditional manner (e.g., suppositories that are given by the buccal route).

Administration of a drug in capsules is a common means of dosing larger test animals (particularly dogs). It is labor intensive (each capsule must be individually prepared, though robotic systems are now available for this), but capsules offer the advantages that neat drug may be used (no special formulation need be prepared, and the questions of formulation or solution stability are avoided), the dogs are less likely to vomit, and the actual act of dosing requires less labor than using a gavage tube. Capsules may also be used with primates, though they are not administered as easily.

Incorporation of a drug in the diet is commonly used for longer term studies (particularly carcinogenicity studies, though the method is not limited to these). Dosing by diet is much less labor intensive than any other oral dosing

methodology, which is particularly attractive over the course of a long (13-week, 1-year, 18-month, or 2-year) study.

The most critical factor to dietary studies is the proper preparation of the test chemical–diet admixtures. The range of physical and chemical characteristics of test materials requires that appropriate mixing techniques be determined on an individual basis. Standard practices generally dictate the preparation of a premix to which is added appropriate amounts of feed to achieve the proper concentrations.

Dietary preparation involving liquid materials frequently results in either wet feed in which the test article does not disperse or formation of "gumballs"—feed and test material that form discernible lumps and chemical "hotspots." Drying and grinding of the premix to a free-flowing form prior to mixing the final diets may be required; however, these actions can affect the chemical nature of the test article.

Solid materials require special techniques prior to or during addition to diets. Materials that are soluble in water may be dissolved and added as described above for liquids. Non-water-soluble materials may require several preparatory steps. The test chemical may be dissolved in corn oil, acetone, or other appropriate vehicle prior to addition to the weighed diet. When an organic solvent such as acetone is used, the mixing time for the premix should be sufficient for the solvent to evaporate. Some solids may require grinding in a mortar and pestle with feed added during the grinding process.

Prior to study initiation, stability of the test chemical in the diet must be determined over a test period at least equivalent to the time period during which animals are to be exposed to a specific diet mix. Stability of test samples under the conditions of the proposed study is preferable. Labor and expense can be saved when long-term stability data permit mixing of several weeks (or a month) of test diet in a single mixing interval.

Homogeneity and concentration analysis of the test article–diet admixture are performed by sampling at three or four regions within the freshly mixed diet (e.g., samples from the top, middle, and bottom of the mixing bowl or blender).

A variety of feeders are commercially available for rats and mice. These include various-sized glass jars and stainless steel or galvanized feed cups, which can be equipped with restraining lids and food followers to preclude significant losses of feed due to animals digging in the feeders. Slotted metal feeders are designed so that animals cannot climb into the feed, and they also contain mesh food followers to prevent digging.

Another problem sometimes encountered is palatability—the material may taste so strongly that animals will not eat it. As a result, palatability, stability in diet, and homogeneity of mix must all be ensured prior to the initiation of an actual study.

Inclusion in drinking water is rarely used for oral administration of human drugs to test animals, though it sees more frequent use for the study of environmental agents.

Physiochemical properties of the test material should be a major consideration in selecting drinking water as a dosing matrix. Unlike diet preparation or preparation of gavage dose solutions and suspensions where a variety of solvents and physical processes can be utilized to prepare a dosable form, preparations of drinking water solutions are less flexible. Water solubility of the test chemical is the major governing factor and is dependent on factors such as pH, dissolved salts, and temperature. The animal model itself sets limitations for these factors (acceptability and suitability of pH and salt-adjusted water by the animals as well as animal environmental specifications such as room temperature).

Stability of the test chemical in drinking water under study conditions should be determined prior to study initiation. Consideration should be given to conducting stability tests on test chemical–drinking water admixtures presented to some test animals. Besides difficulties of inherent stability, changes in chemical concentrations may result from other influences. Chemicals with low vapor pressure can volatilize from the water into the air space located above the water of an inverted water bottle; thus, a majority of the chemical may be found in the "dead space," not in the water.

Certain test chemicals may be degraded by contamination with microorganisms. A primary source of these microorganisms is the oral cavity of rodents. Although rats and mice are not as notorious as the guinea pig in spitting back into water bottles, significant bacteria can pass via the sipper tubes and water flow restraints into the water bottles. Sanitation and sterilization procedures for water bottles and sipper tubes must be carefully attended to.

Many technicians may not be familiar with terms such as sublingual (under the tongue), buccal (between the cheek and gingiva), otic, and so on. A clear description of each of these nontraditional routes (i.e., other than gavage routes) should be discussed with technicians, and instructions may also be written down and given to them. Demonstrations are often useful to illustrate selected techniques of administration (e.g., to use an inhaler or nebulizer). Some chemicals must be placed by technicians into body orifices (e.g., medicated intrauterine devices such as Progesterset).

## 2.4 Minor Routes

The minor routes see some use in safety assessment and four are briefly presented here.

***Perocular Route*** The administration of drugs or accidental exposure of chemicals to the eyes is not commonly a concern in systemic toxicity due to the small surface area exposed and the efficiency of the protective mechanisms (i.e., blink reflex and tears). As long as the epithelium of the eyes remains intact, it is impermeable to many molecules, but, if the toxicant has a suitable polar–nonpolar balance, penetration may occur (Kondrizer et al., 1959; Swan and White, 1972).

Holmstedt (1959) and Brown and Muir (1971) have reviewed perocular absorption of pesticides. More recently, Sinow and Wei (1973) have shown that the quarternary herbicide paraquat can be lethal to rabbits if applied directly to the surface of the eyes. Parathion, in particular, is exceedingly toxic when administered via the eye—a concern that must be kept in mind for the protection of pesticide applicators.

***Rectal Administration*** Since a number of therapeutic compounds are given in the form of suppositories, an indication of the toxicity after rectal administration is sometimes required. Toxicity studies and initial drug formulations of such compounds are usually performed by the oral route and the rectal formulation comes late in development and marketing. In view of the difference between laboratory animals and humans in the anatomy and microflora of the colon and rectum, animal toxicity studies late in drug development are of limited value. However, in cases where an indication of potential rectal hazard or bioavailability is required, the compound may be introduced into the rectum of the rat using an oral dosing needle to prevent tissue damage. To avoid the rapid excretion of the unabsorbed dose, anesthetized animals should be used and the dose retained with an inert plug or bung (such as a cork).

Drugs (and therefore test chemicals) are occasionally administered by rectum, but most are not as well absorbed here as they are from the upper intestine. Aminophylline, used in suppository form for the management of asthma, is one of the few drugs routinely given in this way. Inert vehicles employed for suppository preparations include cocoa butter, glycerinated vehicles, gelatin, and polyethylene glycol. Because the rectal mucosa is irritated by nonisotonic solutions, fluids administered by this route should always be isotonic with plasma (e.g., 0.9% NaCl).

***Vaginal Administration*** Though not a common one, some materials do have routine exposure by this route (e.g., spermicides, tampons, douches, and antibiotics) and therefore must be evaluated for irritation and toxicity. The older preferred models used rabbits and monkeys (Eckstein et al., 1969), but more recently a model that uses rats has been developed (Staab et al., 1987). McConnell (1973) clearly described the limitations, particularly of volume of test material, involved in such tests.

***Nasal Administration*** A route that has gained increasing popularity of late for pharmaceutical administration in humans is the intranasal route. The reasons for this popularity are the ease of use (and therefore ready patient acceptance and high compliance rate), the high degree and rate of absorption of many substances (reportedly for most substances up to 1000 molecular weight; McMartin et al., 1987), and avoidance of the highly acid environment in the stomach and first-pass metabolism in the liver (particularly important for some of the newer peptide moieties) (Attman and Dittmer, 1971). The only special safety concerns are the potential for irritation of the mucous

**TABLE 6 General Guidelines for Maximum Dose Volumes by Route**

| Route | Volume (mL kg$^{-1}$) Should Not Exceed | Notes |
|---|---|---|
| Oral | 20 | Fasted animals |
| Dermal | 2 | Limit is accuracy of dosing per available body surface |
| Intravenous | 1 | Over 5 min |
| Intramuscular | 0.5 | At one site |
| Perocular | 0.01 mL | |
| Rectal | 0.5 | |
| Vaginal | 0.2 mL in rat<br>1 mL in rabbit | |
| Inhalation | 2 mg L$^{-1}$ | |
| Nasal | 0.1 mL per nostril in monkey or dog | |

*Source*: Baker et al., 1979; Garramone, 1986.

membrane and the rapid distribution of administered materials to the central nervous system (CNS).

A number of means may be used to administer materials nasally—nebulizers and aerosol pumps being the most attractive first choices. Accurate dose administration requires careful planning, evaluation of the administration device, and attention to technique.

***Volume Limitations by Route*** In the strictest sense, absolute limitations on how much of a dosage form may be administered by any particular route are determined by specific aspects of the test species or dosage form. But there are some general guidelines (determined by issues of humane treatment of animals, accurate deliver of dose, and such) that can be put forth. These are summarized in Table 6. The chapter Appendix and Section 3 should, of course, be checked to see if there is specific guidance due to the characteristics of a particular vehicle.

## 2.5 Route Comparisons and Contrasts

The first part of this chapter described, compared, and contrasted the various routes used in toxicology and presented guidelines for their use. There are, however, some exceptions to the general rules that the practicing toxicologist should keep in mind.

The relative ranking of efficacy of routes that was presented earlier in the chapter is not absolute; there can be striking exceptions. For example, though materials are usually much quicker acting and more potent when given by the oral route than by the dermal one, this is not always the case. In the literature, Shaffer and West (1960) reported that tetram as an aqueous solution was more toxic when applied dermally than when given orally to rats. The $LD_{50}$ values reported were as follows [$LD_{50}$ (mg kg$^{-1}$) of tetram; 95% confidence limits]:

| Rat | Oral ($mg\,kg^{-1}$) | Percutaneous ($mg\,kg^{-1}$) |
|---|---|---|
| Male | 9 (7–13) | 2 (1–3) |
| Female | 8 (6–11) | 2 (1–3) |

The author has in the past experienced this same phenomenon. Several materials that were found to be relatively nontoxic orally were extremely potent by the dermal route (differences in potency of more than an order of magnitude have been seen at least twice).

A general rule applicable to routes and vehicles should be presented here:

*Vehicles can mask the effects of active ingredients.* Particularly for clinical signs, attention should be paid to the fact that a number of vehicles (e.g., propylene glycol) cause transient neurobehavioral effects that may mask similar short-lived (though not necessarily equally transient and reversible) effects of test materials.

## 3 FORMULATION OF TEST MATERIALS

One of the areas that is overlooked by virtually everyone in toxicology testing and research yet is of crucial importance is the use of vehicles in the formulation of test chemicals for administration to test animals. For a number of reasons, a drug of interest is rarely administered or applied as is ("neat"). Rather, it must be put in a form that can be accurately given to animals in such a way that it will be absorbed and not be too irritating. Most laboratory toxicologists come to understand vehicles and formulation, but to the knowledge of the author, guidance on the subject is limited to a short chapter on formulations by Fitzgerald et al. (1983). There is also a very helpful text on veterinary dosage forms by Blodinger (1982).

Regulatory toxicology in the United States can be said to have arisen, due to the problem of vehicles and formulation, in the late 1930s, when attempts were made to formulate the new drug sulphanilamide. This drug is not very soluble in water, and the U.S. firm Massengill produced a clear, syrupy elixir formulation that was easy to take orally. The figures illustrate how easy it is to be misled. The drug sulphanilamide is not very soluble in glycerol, which has an $LD_{50}$ in mice of $31.5\,g\,kg^{-1}$, but there are other glycols that have the characteristic sweet taste and a much higher solvent capacity. Ethylene glycol has an $LD_{50}$ of $13.7\,g\,kg^{-1}$ in mice and $8.5\,g\,kg^{-1}$ in rats, making it slightly more toxic than diethylene glycol, which has an $LD_{50}$ in rats of $20.8\,g\,kg^{-1}$, similar to that for glycerol. The drug, which is itself inherently toxic, was marketed in a 75% aqueous diethylene glycol–flavored elixir. Early in 1937 came the first reports of deaths, but the situation remained obscure for about six months until it became clear that the toxic ingredient in the elixir was the diethylene glycol. Even as late as March 1937, Haag and Ambrose were reporting that the glycol was excreted substantially unchanged in dogs, suggesting that it was likely to

be safe (Hagenbusch, 1937). Within a few weeks, Holick (1937) confirmed that a low concentration of diethylene in drinking water was fatal to a number of species. Hagenbusch (1937) found that the results of necropsies performed on patients who had been taking 60–70 mL of the solvent per day were similar to those of rats, rabbits, and dogs taking the same dose of solvent with or without the drug. This clearly implicated the solvent, although some authors considered that the solvent was simply potentiating the toxicity of the drug. Some idea of the magnitude of this disaster may be found in the paper of Calvary and Klump (1939), who reviewed 105 deaths and a further 2560 survivors who were affected to varying degrees, usually with progressive failure of the renal system. It is easy to be wise after the event, but the formulator fell into a classic trap, in that the difference between acute and chronic toxicity had not been adequately considered. In passing, the widespread use of ethylene glycol itself as an antifreeze has led to a number of accidental deaths, which suggests that the lethal dose in humans is around 1.4 mL $kg^{-1}$, or a volume of about 100 mL. In the preface to the first United States Pharmacopeia (USP), published in 1820, there is the statement that "it is the object of the Pharmacopoeia to select from among substances which possess medical power, those, the utility of which is most fully established and best understood; and to form from them preparations and compositions, in which their powers may be exerted to the greatest advantage." This statement suggests that the influence that formulation and preparation may have on the biological activity of a drug (and on nonpharmaceutical chemicals) has been appreciated for a considerable time.

Available and commonly used vehicles and formulating agents are reviewed, along with basic information on their characteristics and usages, in the appendix at the end of the chapter. There is a general presumption that those excipients and formulating agents listed in the USP or the *Inactive Ingredient Guide* prepared by the FDA (www.accessdata.fda.gov) are safe to use and without biological effect. This may not always be the case in either experimental animals (see appendix to this chapter) or humans (see Weiner and Bernstein, 1989) either directly or in how they alter absorption of and response to the active ingredient.

There are some basic principles to be observed in developing and preparing test material formulations. These are presented in Table 7.

*Bioavailability* is defined as the fraction of the dose reaching either the therapeutic target organ or tissue or the systemic circulation as unchanged compound following administration by any route. For an agent administered orally, bioavailability may be less than unity for several reasons. The chemical may be incompletely absorbed. It may be metabolized in the gut, the gut wall, the portal blood, or the liver prior to entry into the systemic circulation (see Figure 4). It may undergo enterohepatic cycling with incomplete reabsorption following elimination into the bile. Biotransformation of some chemicals in the liver following oral administration is an important factor in the pharmacokinetic profile, as will be discussed further. Bioavailability measures following oral administration are generally given as the percentage of the dose available to the systemic circulation.

**TABLE 7 Basic Principles in Developing Formulations**

| | |
|---|---|
| A. | Preparation of the formulation should not involve heating of the test material anywhere near the point where its chemical or physical characteristics are altered. |
| B. | If the material is a solid and it is to be assessed for dermal effects, its shape and particle size should be preserved. If intended for use in humans, topical studies should be conducted with the closest possible formulation to that to be used on humans. |
| C. | Multicomponent test materials (mixtures) should be formulated so that the administered form accurately represents the original mixture (i.e., components should not be selectively suspended or taken into solution). |
| D. | Formulation should preserve the chemical stability and identity of the test material. |
| E. | The formulation should be such as to minimize total test volumes. Use just enough solvent or vehicle. |
| F. | The formulation should be easy to administer accurately. |
| G. | pH of dosing formulations should be between 5 and 9, if possible. |
| H. | Acids or bases should not be used to divide the test material (for both humane reasons and to avoid pH partitioning in either the gut or the renal tubule). |
| I. | If a parenteral route is to be employed, final solutions should be as nearly isotonic as possible. Do not assume a solution will remain such upon injection into the bloodstream. It is usually a good idea to verify that the drug stays in solution upon injection by placing some drops into plasma. |
| J. | Particularly if use is to be more than a single injection, steps (such as filtration) should be taken to ensure suitable sterility. |

As the components of a mixture may have various physiochemical characteristics (solubility, vapor pressure, density, etc.), great care must be taken in preparing and administering any mixture so that what is actually tested is the mixture of interest.

Examples of such procedures are making dilutions (not all components of the mixture may be equally soluble or miscible with the vehicle) and generating either vapors or respirable aerosols (not all the components may have equivalent volatility or surface tension, leading to a test atmosphere that contains only a portion of the components of the mixture).

By increasing or decreasing the viscosity of a formulation, the absorption of a toxicant can be altered (Ritschel et al., 1974). Conversely, the use of absorbents to diminish absorption has been used as an antidote therapy for some forms of intoxication. Using the knowledge that rats cannot vomit, there have been serious attempts at making rodenticides safer to nontarget animals by incorporating emetics into the formulations, but this has had only a limited success. Gaines (1960) used in vivo liver perfusion techniques to investigate the apparent anomaly that the carbamate Isolan was more toxic when administered to rats percutaneously than when administered orally. It has been shown that these results, a manifestation of different formulations, have been used for the two routes of exposure (oral and percutaneous) in estimating the $LD_{50}$ values using a common solvent, *n*-octanol. It was found that Isolan was significantly more toxic by the oral route than by the percutaneous route; by regression analysis it was found that at no level of lethal dose values was the reverse correct.

Although the oral route is the most convenient, there are numerous factors that make it unpredictable. Absorption by this route is subject to significant variation from animal to animal and even in the same individual animal at different times. Considerable effort has been spent by the pharmaceutical industry to develop drug formulations with absorption characteristics that are both effective and dependable. Protective enteric coatings for pharmaceuticals were introduced long ago to retard the action of gastric fluids and then disintegrate and dissolve after passage of a tablet into the human intestine. The purposes of these coatings for drugs are to protect the active ingredient, which would be degraded in the stomach, to prevent nausea and vomiting caused by local gastric irritation (also a big problem in rodent studies, where over a long time period gastric irritation frequently leads to forestomach hyperplasia), to obtain higher local concentrations of the active ingredient intended to act locally in the intestinal tract, to produce a delayed biological effect, or to deliver the active ingredient to the intestinal tract for optimal absorption there. Such coatings are generally fats, fatty acids, waxes, or other such agents, and all of these intended purposes for drug delivery can readily be made to apply for some toxicity studies. Their major drawback, however, is the marked variability in time for a substance to be passed through the stomach. In humans, this gastric emptying time can range from minutes to as long as 12 h. One would expect the same for animals, as the limited available data suggest is the case. Similar coating systems, including microencapsulation (see Melnick et al., 1987), are available for and are currently used in animal toxicity studies.

The test chemical is unlikely to be absorbed or excreted unless it is first released from its formulation. It is this stage of the process that is the first and most critical step for the activity of many chemicals. If the formulation does not release the chemical, the rest of the process becomes somewhat pointless.

It might be argued that the simplest way around the formulation problem is to administer any test as a solution in water, thereby avoiding the difficulties altogether. However, since multiple, small, accurately measured doses of a chemical are required repeatedly, reproducible dilutions must be used. Also, the water itself is to be regarded as the formulation vehicle, and the test substance must be water soluble and stable in solution, which many are not. If we take into account the need for accuracy, stability, and optimum performance in vivo, the problem can become complex.

Direct connection between observed toxicity and formulation components is uncommon and it is usually assumed that vehicles and other nontest chemical components are innocuous or have only transitory pharmacological effects. Historically, however, this has certainly not been the case. Even lactose may have marked toxicity in individual test animals (or humans) who are genetically incapable of tolerating it.

The initial stage of drug release from the formulation, in terms of both the amount and the rate of release, may exercise considerable influence at the clinical response level. A close consideration of the formulation parameters

of any chemical is therefore essential during the development of any new drug, and, indeed, there are examples where formulations of established drugs also appear to require additional investigation.

The effects of formulation additives on chemical bioavailability from oral solutions and suspensions have been well reviewed by Hem (1973). He pointed out how the presence of sugars in a formulation may increase the viscosity of the vehicle. However, sugar solutions alone may delay stomach-emptying time considerably when compared to solutions of the same viscosity prepared with celluloses, which may be due to the sugar's effect on osmotic pressure. Sugars of different types may also have an effect on fluid uptake by tissues and this, in turn, correlates with the effect of sugars such as glucose and mannitol on drug transport.

Surfactants have been explored widely for their effects on drug absorption, in particular using experimental animals (Gibaldi and Feldman, 1970; Gibaldi, 1976). Surfactants alter dissolution rates (of lipid materials), surface areas of particles and droplets, and membrane characteristics, all of which affect absorption.

Surfactants may increase the solubility of the drug via micelle formation, but the amounts of material required to increase solubility significantly are such that at least orally the laxative effects are likely to be unacceptable. The competition between the surfactant micelles and the absorption sites is also likely to reduce any useful effect and make any prediction of net overall effect difficult. However, if a surfactant has any effect at all, it is likely to be in the realm of agents that help disperse suspensions of insoluble materials and make them available for solution. Natural surfactants, in particular bile salts, may enhance absorption of poorly soluble materials.

The effective surface area of an ingested chemical is usually much smaller than the specific surface area that is an idealized in vitro measurement. Many drugs whose dissolution characteristics could be improved by particle size reduction are extremely hydrophobic and may resist wetting by GI fluids. Therefore, the GI fluids may come in intimate contact with only a fraction of the potentially available surface area. The effective surface area of hydrophobic particle can often be increased by the addition of a surface-active agent to the formulation, which reduces the contact angle between the solid and the GI fluids, thereby increasing effective surface area and dissolution rate.

Formulations for administering dermally applied toxicants present different considerations and problems. The extent of penetration and speed with which a biologically active substance penetrates the skin or other biological membrane depends on the effect that the three factors—vehicle, membrane, and chemical—exert on the diffusion process. It is now accepted that they together represent a functional unit that controls the penetration and location of the externally applied chemicals in the deeper layers of the skin or membrane layer. The importance of the vehicle for the absorption process has been neglected until recently. One of the few requirements demanded of the vehicle has been that it act as an inert medium that incorporates the test chemical in the most homogeneous distribution possible. In addition, chemical stability

and good cosmetic appearance have been desirable. Most formulation in toxicology are based on empirical experience.

The chemical incorporated in a vehicle should reach the surface of the skin at a suitable rate and concentration. If the site of action lies in the deeper layers of the epidermis or below, the substance must cross the stratum corneum if the skin is intact. Both processes, diffusion from the dosage form and diffusion through the skin barriers, are inextricably linked. They should be considered simultaneously and can be influenced by the choice of formulation.

The thesis that all lipid-soluble compounds basically penetrate faster than water-soluble ones cannot be supported in this absolute form. A lipophilic agent can penetrate faster or slower or at the same rate as a hydrophilic agent, depending on the vehicle used.

Disregarding such chemical-specific properties as dissociation constants (in the case of ionic compounds), particle size, and polymorphism, as well as side effects of viscosity, binding to vehicle components, complex formulation, and the like, the following formulation principles arise:

(a) Optimization of the concentration of chemical capable of diffusion by testing its maximum solubility
(b) Reduction of the proportion of solvent to a degree that is adequate to keep the test material still in solution
(c) Use of vehicle components that reduce the permeability barriers

These principles lead to the conclusion that each test substance requires an individual formulation. Sometimes different ingredients will be required for different concentrations to obtain the maximum rate of release. No universal vehicle is available for any route, but a number of approaches are. Dosage preparation laboratories should be equipped with glassware, a stirring hot plate, a sonicator, a good homogenizer, and a stock of the basic formulating material, as detailed at the end of this chapter.

### 3.1 Dermal Formulations

Preparing formulations for application to the skin has special considerations associated with it, which, in the case of human pharmaceuticals, has even led to a separate book (Barry, 1983).

The physical state of the skin is considerably affected by external factors such as relative humidity, temperature, and air movement at the skin surface. If this contact is broken (e.g., by external applications of ointments or creams), it is reasonable to assume that the new skin will change in some way, sometimes to an extent that creates new conditions of permeability for the test material. This would be the case, for example, if the stratum corneum becomes more hydrated than normal due to the topical delivery form. Temperature might also have an effect, as is the case when any constituents of the vehicle

affect the inner structure of the skin through interactions with endogenous skin substances. Often several of these processes occur together. Figure 5 shows the relationship between water content and relative humidity.

Since this is a question of interaction between the vehicle and the skin (and the latter cannot be viewed as an inert medium), the composition of the vehicle itself may be altered (e.g., by incorporation of skin constituents or through loss of volatile components).

The first contact between vehicle and skin occurs on the skin's surface. The first phase of interaction undoubtedly begins with the lipid mantle in the case of so-called normal skin. If the skin has been damaged by wounds, the surface can form a moist milieu of serious exudate, resulting in abnormal wetting properties. Normally it is impregnated with oily sebaceous secretions and horny fat, presenting a hydrophobic surface layer. Water will not spread out as a film but will form droplets, while bases with a high affinity to the skin surface constituents spread spontaneously into a film and can wet. In the case of a base low viscosity, the degree of wetting can often be determined by measuring the angle of contact. If the preparation wets the skin surface, is drawn by capillary action from the visible area into the large inner surface of the stratum corneum, and is transported away into the interior, then it is said that the ointment or cream penetrates well. Spreading and wetting are purely surface phenomena, not penetration in the strict sense. If the skin shows a high content of its own lipids, spreading is limited. It is also reduced if the value of the surface tension of the skin ($o_s$) decreases compared to the value of the interfacial force between the skin and subject liquid ($y_{s/1}$) and the surface tension of the subject liquid ($o_1$), as is the case with aqueous bases. Addition of amphiphilic compounds decreases $o_1$ and $y_{s/1}$ and thus spreadability increases.

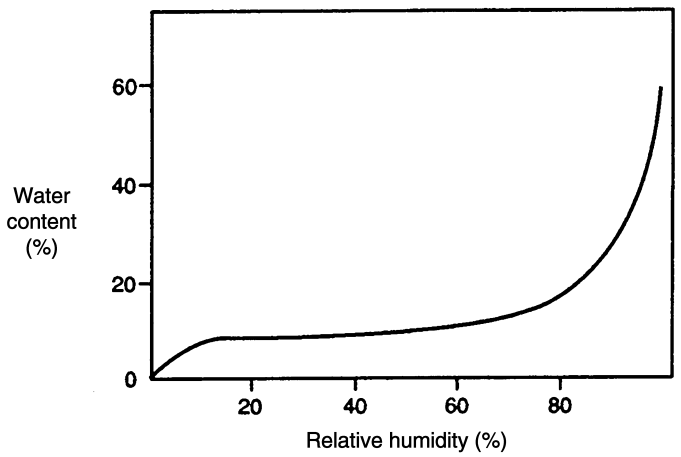


***Figure 5*** *Sorption isotherms of water vapor as function of relative humidity, composition of constituents, and water content in stratum corneum.*

How much the endogenous emulsifying substances of the fatty film, such as cholesterol esters and fatty acid salts, affect this spreading process is not clear. They can probably promote the emulsification of hydrophobic substances with water. Whether the sebaceous and epidermal lipids alone are sufficient to emulsify water and so form a type of emulsive film remains controversial. However, it is assumed that they, together with appropriate vehicle components, improve the spreading of the applied vehicle and that this effect can be potentiated by mechanical means such as intensive rubbing in. Good spreadability ensures that the active ingredient is distributed over a large area.

High local concentrations are avoided and, at the same time, close contact is made between the chemical and the upper layers of the skin.

In grossly simplified terms, hydrogels, suspensions, and oil–water emulsions behave on the skin surface similarly to aqueous solutions. By contrast, pastes and water–oil emulsions act like oil. The ability of an organic solvent to stick or wet depends on its specific properties (e.g., its viscosity and its surface tension).

At present, information concerning alterations in vehicle composition on the skin surface is sparse. However, two possible extremes are conceivable. On the one hand, if the vehicle has a high vapor pressure, it often completely evaporates shortly after application. On the other hand, the vehicle may remain on the skin surface in an almost completely unchanged composition (e.g., highly viscous Vaseline or similar thick covering systems). Between these two extremes lie the remaining types of vehicles.

The first situation applies for the short-chain alcohols, acetone, or ether. After their evaporation, the drug remains finely dispersed on or in the skin at 100% concentration.

If individual components evaporate, the structure of the vehicle changes and, under certain circumstances, also the effective drug concentration. Oil–water emulsions lose water rapidly, giving rise to the well-known cooling effect. If evaporation continues, the dispersed oil phase coalesces and forms a more or less occlusive film on the skin, together with the emulsifier and the drug. Of course, it is possible that a certain hydrophilic proportion of the drug is then present in suspended form or at least can react with charged molecules and is thus removed from the diffusion process at the start. At the same time, it is to be expected that soluble constituents of the skin are incorporated so that a new system can be formed on the surface and the adjoining layers of skin. Comparable transformations probably also occur after application of water–oil emulsions, providing one realizes that the water evaporates more slowly, the cooling effect is less strong, and, due to the water–oil character of the molecule, the occlusive effect can be more marked because of the affinity of the oily components for the skin.

By contrast, Vaseline and similar highly viscous, lipid bases from the outset form an impenetrable layer, virtually unaffected by external factors or effects emanating from the skin itself. Interactions with the skin lipids are only likely at the boundary between ointment and skin.

The evaporation of the water from the skin into the atmosphere is a continuous process. It can be increased or decreased by the use of suitable vehicles. An evaporation increase will always occur if the water vapor from the vehicle is taken away more quickly than water can diffuse from the deeper layers into the stratum corneum. This applies in principle to all hydrophilic bases, particularly for systems with an oil–water character that, after loss of most of their own water, develop a true draining effect that can lead to the drying out of the underlying tissue. How much the penetration of hydrophilic drugs can be proved with the help of oil–water systems depends on the solution properties of the rest of the components in the skin. Generally, such compounds can only seldom reach deeper layers. It is equally difficult to show an adequate release of water from hydrophilic systems to a dry skin. If any such effects do occur, they are short term and are quickly overtaken by opposing processes. The same seems to apply to most traditional moisturizers such as glycerin and propylene glycol (Powers and Fox, 1957; Rieger and Deems, 1974). They can also cause a large rise in the rate of evaporation, depending on the relative humidity, and thus increase the transepidermal loss of water. It is probably impossible to prevent this drying out without preparations having some occlusive properties.

In contrast, vehicles that are immiscible with water and those with a high proportion of oils have occlusive effects. They reduce both insensible perspiration and the release of sweat. The sweat collects as droplets at the opening of the glands but does not spread as a film between the hydrophobic skin surface and the lipophilic base because the free surface energy of the vehicle–skin interface is smaller than that between water and skin. If a lipophilic layer of vehicle is present, this is not spontaneously replaced by the water–skin layer if sweat is secreted.

The horny layer consists of about 10% extracellular components such as lipids, proteins, and mucopolysaccharides. Around 5% of the protein and lipids form the cell wall. The majority of the remainder is present in the highly organized cell contents, predominantly as keratin fibers, which are generally assigned an α-helical structure. They are embedded in a sulfur-rich amorphous matrix, enclosed by lipids that probably lie perpendicular to the protein axis. Since the stratum corneum is able to take up considerably more water than the amount that corresponds to its volume, it is assumed that this absorbed fluid volume is mainly located in the region of these keratin structures.

Some insight into where on the relative humidity continuum water molecules are absorbed can be gained from equilibrium isotherms (Ziegenmeyer, 1982) (Figure 5), which show a characteristic sigmoidal shape. At low relative humidity, water is first absorbed at specific skin sites, probably in the region of the peptide compounds and the various polar side chains. At higher moisture content, layers of water form on the skin. By using Zimm–Lundberg cluster theory (Zimm and Lundberg, 1956), additional information can be obtained about the nature of the absorbed water.

Because of thick intertwining protein fibers in the cell and in the area of the cell membrane, cell structure is rigid and remains so but is altered by the osmotic effect of the penetrating water. The uptake of water entails a continual shifting of the cell matrix, which gradually develops elastic opposing forces that increasingly resist further expansion. An equilibrium is reached if both forces balance each other. In the case of water, it takes quite a long time to completely hydrate the cell. This process can, however, be shortened if there are components present with a solvent effect diffusing out of a basic vehicle. The duration and degree of swelling depends on the affinity of all the dissolved substances for the tissue and on the size of the maximum possible elastic reaction, which stabilizes cell structure.

### 3.2 Interactions between Skin, Vehicle, and Test Chemical

The diffusion coefficient of the hydrated stratum corneum is larger than that of dry skin. Therefore, hydration increases the rate of passage of all substances that penetrate the skin. If the hydrated keratin complex is represented by a biphasix system, then it can be considered to exist as a continuous region covered with layers of water and intervening layers of lipids. Nonpolar compounds are predominantly dissolved in the nonpolar lipid matrix and diffuse through it. Polar substances, by contrast, pass through the aqueous layers. The diffusion of water and low-molecular-weight, hydrophilic molecules through these layers of water is more difficult than a corresponding free diffusion in an aqueous solution. This could, under certain circumstances, be due to a higher degree of organization of water in the protein structures (than in plasma or the free state), in the sense that this water is only available as the driving force of the diffusion process to a limited degree.

The degree of hydration can be controlled by the choice of vehicle. Lipophilic paraffin bases are available, but vehicles such as water–oil emulsions are more acceptable since they are less occlusive and offer ease of formulation.

In principle, temperature can also have an effect on penetration, which may be exerted on the basic vehicle if it contains temperature-sensitive components (e.g., nonionic tensides). Room and body temperatures can be enough to change the hydrophilic–lipophilic balance and thus possibly change the entire system. It has long been known that increasing temperature can considerably reduce diffusional resistance and thereby increase the rate of penetration of substances. In practice, however, this effect is of no importance. Of course, skin temperature will be increased a few degrees by occlusion because of the prevention of sweating and restriction of heat radiation. However, compared to the increase in penetration achieved by the simultaneous hydration process, this effect is insignificant.

Additives aimed at accelerating penetration always attempt to enable diffusion of pharmacologically active compounds into or through the stratum corneum without damaging it and without causing undesirable systemic effects. Although attempts have been made to limit these effects, this goal has not

been achieved as yet. There are numerous substances that decrease the diffusional resistance of the skin, such as propylene glycol, tensides, parotic substances such as urea, dimethylsulfoxide (DMSO), dimethylformamide (DMF), and various other organic solvents, mostly of medium chain length. They all improve the penetration of dissolved agents, but only at the cost of the integrity of skin structure, raising the question of the degree of damage and reversibility.

If the substances have passed the stratum corneum, they also generally diffuse into the living part of the epidermis, reach the circulation, and then have systemic effects depending on the amount absorbed. Because these are often constituents of formulations, one generally expects them to have little direct influence on skin penetration. However, their amphiphilic properties allow them to form new systems with the body's constituents and even to change the physical state of water in the skin. By this means, a pathway is cleared for other hydrophilic substances to gain entry into the general circulation.

Most of a permeability enhancer (such as a tenside) is bound to the stratum corneum. It is assumed that the underlying mechanism of the process involves interactions with keratin structures. Positively and negatively charged ionic groups of proteins have been suggested as binding sites for ionic substances. Ion pairs could also form. On the other hand, hydrophobic areas are present that bind with the uncharged part of the enhancers. The total free binding energy of molecules to keratin is made up of the contributions arising from electrostatic and nonpolar interactions. Nonpolar interactions increase with the chain length of the molecule. This would be the reason why predominantly anionic molecules of medium chain length exert stronger effects on the keratin structure than those of shorter chain length (Dominguez et al., 1977).

In order to reach the interior of the tightly enmeshed keratin, the molecule must overcome the elastic energy of the polypeptide matrix. The energy necessary to do this is proportional to the volume of the penetrating molecule. The larger the volume, the more difficult it will be for the molecule to approach the various binding sites of proteins in the interior of the keratin complex. Thus, the size of the penetrating molecule is subject to certain limits. If more molecules are present than can become bound, it is possible that a few of them will reach the living layers of the epidermis, as has been described for several anionic, mostly medium-chain enhancer molecules such as tensides. It remains unclear whether this is a consequence of pure saturation or if other interactions are involved (e.g., with structural lipids or hydrophilic materials from the intercellular lipids).

The extent to which the vehicle can affect the entire diffusion process can be shown by an example. In a four-component system of 40% oil, 40% water, and 20% of an emulsifying agent and coemulsifier, alteration of only the proportion of emulsifier to coemulsifier leads to systems of completely different colloidal–chemical structures, which can be labeled as creams, gels, or microemulsions.

Dermal administration presents fewer logistic difficulties than oral administration. Liquids can be administered as supplied and powders or solids can be moistened with saline to form a thick paste or slurry or can be applied dry and moistened with saline. Solid materials (e.g., sheets of plastic, fabric) can also be administered dermally. Liquid materials or slurries are applied directly to the skin, taking care to spread the material evenly over the entire area or as much of the area as can reasonably be covered and then covering with a strip of gauze. If a large amount of material is being administered and the abdominal skin will be exposed, it is sometimes necessary to apply material to the gauze and to the skin. Dry materials are weighed out, then placed on the gauze strip and moistened with physiological saline (generally 15 mL) so that they adhere to the gauze. The gauze is then wrapped around the animal. This porous gauze dressing is then held in place by an additional wrapping, generally of an impervious material, to create an "occlusive" covering. This occlusion enhances penetration and prevents ingestion or evaporation of the test material.

Another recently developed approach is the use of plastic containment capsules (modified Hill Top Chambers) for administration of well-measured doses in a moisturized microenvironment (Derelanko et al., 1987).

Finally, it should be noted that for some agents (contrary to the general rule) decreasing the concentration of chemical in a vehicle may increase its apparent intrinsic toxicity.

### 3.3 Oral Formulations

The physical form of a material destined for oral administration often presents unique challenges. Liquids can be administered as supplied or diluted with an appropriate vehicle, and powders or particulates can often be dissolved or suspended in an appropriate vehicle. However, selection of an appropriate vehicle is often difficult. Water and oil (such as vegetable oils) are used most commonly. Materials that are not readily soluble in either water or oil can frequently be suspended in a 1% aqueous mixture of methylcellulose. Occasionally, a more concentrated methylcellulose suspension (up to 5%) may be necessary. Materials for which appropriate solutions or suspensions cannot be prepared using one of these three vehicles often present major difficulties.

Limited solubility or suspendability of a material often dictates preparation of dilute mixtures that may require large volumes to be administered. The total volume of liquid dosing solution or suspension that can be administered to a rodent is limited by the size of its stomach. However, because rats lack a gagging reflex and have no emetic mechanism, any material administered will be retained. Guidelines for maximum amounts to be administered are given in Table 6.

Limitations on total volume therefore present difficulties for materials that cannot easily be dissolved or suspended. The most dilute solutions that can be administered for a limit-type test (5000 mg $kg^{-1}$), using the maximum volumes

shown in Table 6, generally are 1% for aqueous mixtures and 50% for other vehicles.

Although vehicle control animals are not required for commonly used vehicles (water, oil, methylcellulose), most regulations require that the biological properties of a vehicle be known and/or that historical data be available. Unfortunately, the best solvents are generally toxic and thus cannot be used as vehicles. Ethanol and acetone can be tolerated in relatively high doses but produce effects that may complicate interpretation of toxicity associated with the test material alone. It is sometimes possible to dissolve a material in a small amount of one of these vehicles and then dilute the solution in water or in oil.

Gels and resins often present problems because of their viscosity at room temperature. Warming these materials in a water bath to a temperature of up to 50 °C will frequently facilitate mixing and dosing. However, it is important to ascertain that no thermal degradation occurs and that actually administered formulations be at or near body temperature.

Other possibilities for insoluble materials are to mix the desired amount of material with a small amount of the animal's diet or to use capsules. The difficulty with the diet approach is the likelihood that the animal will not consume all of the treated diet or that it may selectively not consume chunks of test material. Use of capsules, meanwhile, is labor intensive. In rare cases, if all of these approaches fail, it may not be possible to test a material by oral administration. In capsules, particle size is generally inversely related to solubility and bioavailability. However, milling of solids may adversely affect their chemical nature and/or pose issues of safety.

If necessary, the test substance should be dissolved or suspended as a suitable vehicle, preferably in water, saline, or an aqueous suspension such as 0.5% methylcellulose in water. If a test substance cannot be dissolved or suspended in an aqueous medium to form a homogenous dosage preparation, corn oil or another solvent can be used. The animals in the vehicle control group should receive the same volume of vehicle given to animals in the highest dose group.

The test substance can be administered to animals at a constant concentration across all dose levels (i.e., varying the dose volume) or at a constant dose volume (i.e., varying the dose concentration). However, the investigator should be aware that the toxicity observed by administration in a constant concentration may be different from that observed when given in a constant dose volume. For instance, when a large volume of corn oil is given orally, GI motility is increased, causing diarrhea and decreasing the time available for absorption of the test substance in the GI tract. This situation is particularly true when a highly lipid-soluble chemical is tested.

If an organic solvent is used to dissolve the chemical, water should be added to reduce the dehydrating effect of the solvent within the gut lumen. The volume of water or solvent–water mixture used to dissolve the chemical should be kept low, since excess quantities may distend the stomach and cause rapid gastric emptying. In addition, larger volumes of water may carry the chemical

through membrane pores and increase the rate of absorption. Thus, if dose-dependent absorption is suspected, it is important that the different doses are given in the same volume of solution.

Larger volumes than those detailed earlier may be given, although nonlinear kinetics seen under such circumstances may be due to solvent-induced alteration of intestinal function. The use of water-immiscible solvents such as corn oil (which are sometimes used for gavage doses) should be avoided, since it is possible that mobilization from the vehicle may be rate limiting. Magnetic stirring bars or homogenizers can be used in preparing suspensions. Sometimes a small amount of a surfactant such as Tween 80, Span 20, or Span 60 is helpful in obtaining a homogenous suspension.

A large fraction of such a material may quickly pass through the GI tract and remain unabsorbed. Local irritation by a test substance generally decreases when the material is diluted. If the objective of the study is to establish systemic toxicity, the test substance should be administered in a constant volume to minimize GI irritation that may, in turn, affect its absorption. If, however, the objective is to assess the irritation potential of the test substance, then it should be administered undiluted.

### 3.4 Parenteral Formulations

Parenteral dose forms include aqueous, aqueous organic, and oily solutions, emulsions, suspensions, and solid forms for implantation. These parenterals need to be sterile and pyrogen free; they are, if possible, buffered close to normal physiological pH and preferably are isotonic with the body fluids.

The preparation of parenteral dosage forms of approved and potential drugs for animals is the same as for humans. Turco and King (1974) provide a comprehensive review of the subject, which though written with human therapeutics in mind contains very little that is not applicable to animals. Sterility, lack of pyrogenicity, blood compatibility, and low to no irritation at the point of injection are biological requirements; there are also a corresponding set of physicochemical requirements.

Parenteral products are usually given to humans when an immediate effect is needed, when a patient is unable to accept medication by the oral route, or when the drug will be ineffective by the oral route. These conditions apply to animals used in safety evaluation.

Parenteral products can be easily administered to confined or restrained animals, leaving no doubt that the animal received its medication.

To be acceptable, a SC or IM formulation should cause only a minimum of irritation and no permanent damage to the tissues and be systemically distributed and active when administered by this route. The ideal parenteral product is an aqueous solution isotonic with the body fluids with a pH between 7 and 8. When the drug lacks sufficient aqueous solubility, a suspension may be considered; however, in most cases, the bioavailability of the drug may be affected

and encapsulation by the body at the site of injection is extremely likely. The solubility of the drug in water may be improved by the addition of cosolvents such as alcohol, propylene glycol, polyethylene glycol, dimethylacetamide, DMSO, or DMF. The resulting solution must have additional tolerance for water so that the drug will not precipitate at the site of injection when the solution is diluted by body fluids. If precipitation occurs at the site of injection, the absorption of the drug may be delayed or even completely inhibited.

Water-miscible solvents alone can be used when the drug is chemically unstable in the presence of water. The number of solvents available for this purpose is extremely limited. The classic review of this subject was made in 1963 (Spiegel and Noseworthy), and some 30 years later, no additional solvents are available. This is unlikely to change in the near future due to the extensive effort necessary to determine the safety of a solvent used as a vehicle. When a nonaqueous vehicle is used, one can invariably expect some degree of pain upon injection, and subsequent tissue destruction is possible. This damage may be due to the heat of solution as vehicle mixes with body fluids, it may be associated with tissues rejecting the solvent, or it may be an inherent property of the solvent.

Fixed oils of vegetable origin and their esters may be used as parenteral vehicles for some drugs, particularly steroidal hormones. While an oleaginous vehicle may delay or impair absorption of the drug, this characteristic has been used to advantage with some drugs where a small dose is desired over a long period of time. The formulator must know which species will receive the formulation and the type of equipment used in its administration. A product intended for a dog or primate is usually given to a single animal at a time. Conventional glass or disposable syringes will be used with a 20- or 22-gauge needle, which may impede the flow of the liquid, especially when an oleaginous vehicle is used. Impedance is usually compensated for by using small animals, since the volume of injection is small and no more than one injection is normally given at one time.

The viscosity of the solution will influence its acceptability when automatic injection equipment is used. If many animals are injected at one time, a viscous solution that requires a great deal of force to eject will rapidly tire the user. When the automatic injector is refilled from a reservoir, a viscous solution will be slow to fill the volumetric chamber. The subjective aspect of measuring the ease of expelling a dose can be eliminated by constructing an apparatus that will measure the pressure needed to expel a dose (Groves, 1966). An objective means of measuring ease will allow the formulator to vary the composition of the injection and measure any improvement in injectability. For example, the addition of a wetting agent can be investigated and, if improvement is seen, the level of use can be optimized.

A parenteral product in a multidose vial must contain a preservative to protect the contents of the vial against contamination during repeated withdrawal of dose aliquots.

## 4 DOSING CALCULATIONS

One of the first things a new technician (or graduate student) must learn is how to calculate dose. Generally, administered doses in systemic toxicity studies are based on the body weight of the animal (expressed as either weight or volume—for liquids—of the test substance per kilogram of body weight of the animal), although some would maintain that surface area may be a more appropriate basis on which to gauge individual dose. The weight (or dose) of the test substance is often expressed in milligrams or grams of active ingredient if the test substance is not pure (i.e., if it is not 100% active ingredient).

Ideally, only the 100% pure sample should be tested; however, impurity-free samples are difficult to obtain and preparation of formulations (as previously discussed) is frequently essential. The toxicity of impurities or formulation components should be examined separately if the investigator feels that they may contribute significantly to the toxicity of the test substance.

If the test substance contains only 75% active ingredient and the investigator chooses a constant dose volume of $10\,\mathrm{mL\,kg^{-1}}$ body weight across all dose levels, it will be more convenient to prepare a stock solution such that, when $10\,\mathrm{mL\,kg^{-1}}$ of this stock solution is given to the animal, the dose will be the desired one (say $500\,\mathrm{mg\,kg^{-1}}$ of active ingredient). The concentration of this stock solution would be (500 mg/10 mL)/0.75 = 66.7 mg of the test substance per milliliter of diluent.

Aliquots of the test substance for other dose levels can then be prepared by dilution of the stock solution. For example, the solution concentration for a dose level of $250\,\mathrm{mg\,kg^{-1}}$ is (200 mg/10 mL)/0.75 = 26.7 mg of the test substance per milliliter of diluent.

This solution can be prepared by diluting the stock solution 25 times; that is, for each milliliter of the 26.7-$\mathrm{mg\,mL^{-1}}$ solution to be prepared,

$$\frac{(26.7\,\mathrm{mg\,mL^{-1}})(1\,\mathrm{mL})}{66.7\,\mathrm{mg\,mL^{-1}}} = 0.400\,\mathrm{mL}\text{ of stock solution}$$

This amount should be diluted to a final volume of 1 mL with the vehicle.

The other way to express a relative dose in animals or humans is to do so in terms of body surface area. There are many reasons for believing that the surface area approach is more accurate for relating doses between species (Schmidt-Nielsen, 1984)—and especially between test animals and humans—but this is still a less common approach in safety assessment, although it is the currently accepted norm in a couple of areas—carcinogenesis and chemotherapy, for example.

## 5 CALCULATING MATERIAL REQUIREMENTS

One of the essential basic skills for the efficient design and conduct of safety assessment studies is to be able to accurately project compound requirements

for the conduct of a study. In theory, this simply requires plugging numbers into a formula such as

$$(A \times B \times C \times D) \times 1.1 = \text{total compound requirement}$$

where $A$ = number of animals in each study group
$B$ = *sum* of doses of the dose groups (such as $0.1 + 0.3 + 1.0\,\text{mg kg}^{-1} = 1.4\,\text{mg kg}^{-1}$)
$C$ = number of doses to be delivered (usually length of study in days)
$D$ = average body weight per animal (assuming dosing is done on per-body-weight basis)
1.1 = safety factor (in effect, 10%) to allow for spillage, etc.

As an example of this approach, consider a study that calls for 10 dogs per sex per group ($A = 10 \times 2 = 20$) to receive 0, 10, 50, or $150\,\text{mg kg}^{-1}\,\text{day}^{-1}$ ($B = 10 + 50 + 150 = 210\,\text{mg kg}^{-1}$) for 30 days ($C = 30$). On average, our dogs of the age range used weigh 10 kg ($D = 10$ kg). Our compound need is then $(20 \times 210\,\text{mg kg}^{-1} \times 30 \times 10\,\text{kg}) \times 1.1 = 1.386$ kg.

The real-life situation is a bit more complicated, since animal weights change over time, diet studies have doses dependent on daily diet consumption, the material may be a salt but dosage should be calculated on the basis of the parent compound, and not all animals may be carried through the entire study.

For rats and mice (where weight change is most dramatic and diet studies most common), Table 8 presents some reliable planning values for compound requirements during diet studies.

## 6 EXCIPIENTS

Excipients are usually thought of as inert substances (such as gum arabic and starch) that form the vehicle or bulk of the dosage form of a drug. They are, of course, both much more complicated than this and not necessarily inert. A better definition would be that of the USP (2007) and National Formulary (NF), which defined excipients as any component other than the active substances (i.e., drug substances or DSs) intentionally added to the formulation of a dosage form. These substances serve a wide variety of purposes: enhancing stability, adding bulking, increasing and/or controlling absorption, providing or masking flavor, coloring, and serving as a lubricant in the manufacturing process. They are, in fact, essential for the production and delivery of marketed drug products. As will soon be made clear, they are regulated both directly and as part of the drug product (DP). For the pharmaceutical manufacturers, using established and accepted excipients [such as can be found in Hawley, 1971; Budavari, 1989; Smolinske 1992 or American Pharmaceutical Association (APhA), 1994—though these lists are not complete] is much preferred.

**TABLE 8 Standardized Total Compound Requirements for Rodent Diet Studies[a]**

| | Total Compound Requirement (g) per dose ($mg\,kg^{-1}\,day^{-1}$) | | | | | |
|---|---|---|---|---|---|---|
| Length of Study | 1 | 3 | 10 | 30 | 100 | 300 |
| Rat[b] | | | | | | |
| 2 Weeks | 0.2 | 0.4 | 1.2 | 4 | 10.6 | 32 |
| 4 Weeks | 0.43 | 0.7 | 2.5 | 7.5 | 25 | 75 |
| 13 Weeks | 0.8 | 2.6 | 8.5 | 25.5 | 85 | 260 |
| 52 Weeks | 7 | 21 | 70 | 210 | 0.7[c] | 2.1[c] |
| 2 Years | 15 | 45 | 150 | 450 | 1.5[c] | 4.5[c] |
| Mouse | | | | | | |
| 2 Weeks | 0.03 | 0.06 | 0.22 | 0.65 | 2.2 | 6.4 |
| 4 Weeks | 0.08 | 0.14 | 0.8 | 1.4 | 8 | 14 |
| 13 Weeks | 0.14 | 0.42 | 1.4 | 4.2 | 14 | 42 |
| 18 Months | 0.85 | 2.5 | 8.5 | 25 | 85 | 250 |

[a]Based on 10 animals per sex per group for the length of the study that are 6–8 weeks old at study initiation. Animals are weighed to determine body weights.
[b]Sprague-Dawley rats (body weights and compound requirements for Fischer's would be less).
[c]In kilograms.

However, both pharmaceutical manufacturers and the companies which supply excipients must from time to time utilize (and therefore develop, evaluate for safety, and get approved) new excipients.

## 6.1 Regulation of Excipients

Table 9 lists the relevant sections of Code of Federal Regulations (CFR) 21 which govern excipients. Under Section 201(g)(1) of the Federal Food, Drug, and Cosmetic Act (FDCA; 1), the term *drug* is defined as:

> (A) articles recognized in the official *United States Pharmacopeia*, official *Homeopathic Pharmacopeia of the United States*, or official *National Formulary*, or any supplement to any of them; and (B) Articles intended for use in the diagnosis, cure, mitigation, treatment, or prevention of disease in man or other animals; and (C) Articles (other than food) intended to affect the structure of any function of the body of man or other animals; and (D) Articles intended for use as a component of any articles specified in clause (A), (B), or (C).

An excipient meets the definitions as listed in (A) and (D) above.

In 21 CFR 210.3(b)(8)(2), an "inactive ingredient means any component other than an active ingredient." According to the CFR, the term *inactive ingredient* includes materials in addition to excipients. According to 21 CFR 201.117:

> Inactive ingredients: A harmless drug that is ordinarily used as an inactive ingredient, such as a coloring, emulsifier, excipient, flavoring, lubricant, preservative,

**TABLE 9 U.S. Code of Federal Register References to Excipients**

| Subject | Reference | Content |
|---|---|---|
| General | 21 CFR 210.3(b)(8) | Definitions |
| | 21 CFR 201.117 | Inactive ingredients |
| | 21 CFR 210.3(b)(3) | Definitions |
| Over-the-counter drug products | 21 CFR 330.1(e) | General conditions for general recognition as safe, effective, and not misbranded |
| | 21 CFR 328 | Over-the-counter drug products intended for oral ingestion that contain alcohol |
| Drug master files | 21 CFR 314.420 | Drug master files |
| Investigational new drug application | 21 CFR 312.23(a)(7) | IND content and format |
| New drug application | 21 CFR 312.31 | Information amendments |
| | 21 CFR 314.50(d)(1)(ii)(a) | Content and format of an application |
| | 21 CFR 314.70 | Supplements and other changes to an approved application |
| Abbreviated new drug application | 21 CFR 314.94(a)(9) | Content and format of an abbreviated application |
| | 21 CFR 314.127 | Refusal to approve an abbreviated new drug application |
| | 21 CFR 314.127(a)(8) | Refusal to approve an abbreviated new drug application |
| Current good manufacturing practice | 21 CFR 211.84(d) | Testing an approval or rejection of components, drug product containers, and closures |
| | 21 CFR 211.165 | Testing and release for distribution |
| | 21 CFR 211.180(b) | General requirements |
| | 21 CFR 211.80 | General requirements |
| | 21 CFR 211.137 | Expiration dating |
| Listing of drugs | 21 CFR 207 | Registration of procedures of drugs and listing of drugs in commercial distribution |
| | 21 CFR 207.31(b) | Additional drug listing information |
| | 21 CFR 207.10(e) | Exceptions for domestic establishments |
| Labeling | 21 CFR 201.100(b)(5) | Prescription drugs for human use |
| | 21 CFR 201.20 | Declaration of presence of FD&C yellow no. 5 and/or FD&C yellow no. 6 in certain drugs for human use |
| | 21 CFR 201.21 | Declaration of presence of phenylalanine as component of aspartame in over-the-counter and prescription drugs for human use |
| | 21 CFR 201.22 | Prescription drugs containing sulfites; required warning statements |

> or solvent in the preparation of other drugs shall be exempt from Section 502(f)(1) of the Act. This exemption shall not apply to any substance intended for a use which results in the preparation of a new drug, unless an approved new-drug application provides for such use.

Excipients also meet the definition of component in the good manufacturing practice (GMP) regulations in 21 CFR 210.3(b)(3): "Component means any ingredient intended for use in the manufacture of a drug product, including those that may not appear in such drug product."

The NF admissions policy in the *United States Pharmacopeia 30/National Formulary 25* defines the word *excipient* (3): "An excipient is any component other than the active substance(s), intentionally added to the formulation of a dosage form. It is not defined as an inert commodity or an inert component of a dosage form."

Similar to all other drugs, excipients must comply with the adulteration and misbranding provisions of the FDCA (Katdare and Chaubal, 2006). Under Section 501(a), an excipient shall be deemed to be adulterated if it consists in whole or in part of any filthy, putrid, or decomposed substance or if it has been prepared, packed, or held under unsanitary conditions whereby it may have been contaminated with filth or whereby it may have been rendered injurious to health. An excipient is adulterated if the methods used in or the facilities or controls used for its manufacture, processing, packing, or holding do not conform to or are not operated or administered in conformity with current GMPs to assure that such drug meets the requirements of the act as to safety and has the identity and strength and meets the quality and purity characteristics which it purports or is represented to possess. In addition, under Section 501(b), an excipient shall be deemed to be adulterated if it purports to be or is represented as a drug the name of which is recognized in an official compendium, and its strength differs from or its quality or purity falls below the standards set forth in such compendium.

In 2005, the FDA promulgated new guidance on the selection and use of excipients in nonclinical and clinical studies. FDA compliance officials require the use of inactive ingredients that meet compendial standards when standards exist and either have previous use in FDA-approved pharmaceuticals or may be qualified as "novel" excipients (with studies as summarized in Table 10). The CDER maintains an inactive ingredient committee whose charter includes the evaluation of the safety of inactive ingredients on an as-needed basis, preparation of recommendations concerning the types of data needed for excipients to be declared safe for inclusion in a drug product, and other related functions.

From a regulatory standpoint, the FDA's concern regarding safety involves the toxicity, degradants, and impurities of excipients, as discussed in other chapters in this book. In addition, other chapters of this book address types of toxicity concerns, toxicity-testing strategies, and exposure and risk assessment of excipients.

**TABLE 10 Summary of Toxicological Studies Recommended for New Pharmaceutical Excipients Based on Route of Exposure**

| Tests | Oral | Mucosal | Transdermal | Dermal/Topical | Parenteral | Inhalation/Intranasal | Ocular |
|---|---|---|---|---|---|---|---|
| Appendix 1—base set | R | R | R | R | R | R | R |
| Acute oral toxicity | R | R | R | R | R | R | R |
| Acute dermal toxicity | C | C | C | C | C | R | C |
| Acute inhalation toxicity | R | R | R | R | R | R | R |
| Eye irritation | R | R | R | R | R | R | R |
| Skin irritation | R | R | R | R | R | R | R |
| Skin sensitization | R | R | R | R | R | R | R |
| Acute parenteral toxicity | — | — | — | — | R | — | — |
| Application site evaluation | — | — | R | R | R | R | — |
| Pulmonary sensitization | — | — | — | — | — | R | — |
| Phototoxicity/photoallergy | — | — | R | R | — | — | — |
| Ames test | R | R | R | R | R | R | R |
| Micronucleus test | R | R | R | R | R | R | R |
| ADME—intended route | R | R | R | R | R | R | R |
| 28-Day toxicity (2 species)—intended route | R | R | R | R | R | R | R |
| Appendix 2 | | | | | | | |
| 90-Day toxicity (most appropriate species) | R | R | R | R | R | R | R |
| Developmental toxicity (rat and rabbit) | R | R | R | R | R | R | R |
| Additional assays | C | C | C | C | C | C | C |
| Genotoxicity assays | R | R | R | R | R | R | R |
| Appendix 3 | | | | | | | |
| Chronic toxicity (rodent, nonrodent) | C | C | C | C | C | C | C |
| Photocarcinogenicity | — | — | C | C | — | — | — |
| Carcinogenicity | C | C | C | C | C | C | — |

*Note*: R, required; C, conditionally required.

Excipients must be safe for their intended use. Under 21 CFR 330.1(e), over-the-counter (OTC) human drugs that are generally recognized as safe and effective and not misbranded may only contain inactive ingredients if they are suitable and if the amounts administered are safe and do not interfere with the effectiveness of the drug or with required tests or assays. Color additives may be used in accordance with the provisions of the FDCA and the regulations of 21 CFR Parts 70–82. The FDA proposed that, to be considered as suitable within the meaning of 21 CFR 330.1(e), each inactive ingredient in an OTC human drug product should perform a specific function (5). The proposed regulation defined *safe and suitable* to mean that the inactive ingredient meets various conditions as mentioned in the foregoing. Over-the-counter drug manufacturers are responsible for assuring that these conditions are met. There is no formal approval mechanism (Levi, 1963).

In the United States, the safety and suitability of excipients used in new drugs are considered as part of the new drug application (NDA) process. There is no separate and independent review and approval system for excipients. There are no specific regulations or guidelines that specify the requirements needed to gain approval of a new drug that contains a new excipient. Generally, pharmaceutical companies choose excipients that previously have been approved for commercial use in other NDAs. The FDA's *Inactive Ingredient Guide*, discussed later in this chapter, contains a listing of inactive ingredients present in approved drug products. There is currently no way of gaining a listing for an excipient in the guide independent of the NDA route. The FDA reviews the status of an excipient in food as information to support its use in drug products. Factors relative to the use of an excipient, such as dosing regimen and route of administration, are also reviewed. Advances in excipient technology and drug dosage from technology have created a need for a separate regulatory approval process for new excipients. The USP published IPEC's Excipient Safety Evaluation Guidelines as Information Chapter 1074, Excipient Biological Safety Evaluation Guideline.

Information on existing or new excipients can be described and provided to the FDA in an NDA directly. Alternatively, the manufacturers of excipients may prepare and submit type IV drug master files (DMFs) to support the use of an excipient in one or more NDAs. The DMFs are discussed in FDA regulations under 21 CFR 314.420 and the FDA-issued Guidance for Drug Master Files. When authorized by the DMF submitter (i.e., the excipient manufacturer) and cross-referenced by an NDA submitter, the FDA reviews the DMF to make determinations on the safety, manufacture, and quality of the excipient use in the new drug that is the subject of the then-pending NDA. The DMF becomes active when reviewed in conjunction with the review and approval of an NDA.

The USP/NF provides a listing of excipients by categories in a table according to the function of the excipient in a dosage form, such as tablet binder, disintegrant, and such. An excellent reference for excipient information is the *Handbook of Pharmaceutical Excipients* (APhA, 2007). Additionally, Gad et al. (2006) provide an excellent and extensive database of nonclinical formulation

components and either acceptable maximum usage levels by species route and duration of study.

Excipients have historically not been subjected to extensive safety testing because they have been considered a priori to be biologically inactive and therefore nontoxic. Many, if not most, excipients used are approved food ingredients, the safety of which has been assured by a documented history of safe use or appropriate animal testing. Some of the excipients are generally recognized as safe (GRAS) food ingredients. The excipient is an integral component of the finished drug preparation and, in most countries, is evaluated as part of this preparation. There has been no apparent need to develop specific guidelines for the safety evaluation of excipients, and most developed countries do not have specific guidelines. However, as drug development has become more complex and/or new dosage forms have developed, improved drug bioavailability has become more important. It was noted that the available excipients were often inadequate, new pharmaceutical excipients specifically designed to meet the challenges of delivering new drugs were needed, and these are being developed. The proper safety evaluation of new excipients has now become an integral part of drug safety evaluation.

In the absence of official regulatory guidelines, safety committees of the IPEC in the United States, Europe, and Japan developed guidelines for the proper safety evaluation of new pharmaceutical excipients (IPEC, 1997). The committees critically evaluated guidelines for the safety evaluation of food ingredients, cosmetics, and other products as well as textbooks and other appropriate materials. Before initiating a safety evaluation program for a new pharmaceutical excipient, it is advisable to address the following:

1. Chemical and physical properties and functional characterization of the test material
2. Analytical methods that are sensitive and specific for the test material and that can be used to analyze for the test material in animal food used in the feeding studies or in the vehicle used for other studies
3. Available biological, toxicological, and pharmacological information on the test material and related materials (which involves a thorough search of the scientific literature)
4. Intended conditions of use, including reasonable estimates of exposure
5. Potentially sensitive segments of the population

As discussed in Chapter Drug Development Process and Global Pharmaceutical Marketplace, a comprehensive and critical search of the scientific literature on the test material and related materials is essential before the start of any testing program.

As pharmaceutical excipients are assumed to be biologically nonreactive, dose–response relations cannot always be established. An acceptable alternative is to use a maximum attainable or maximum feasible dose. This is the

highest dose possible that will not compromise the nutritional or health status of the animal. Table 11 summarizes the maximum or limit doses for various types of studies by different routes of exposure. For example, 2000 mg $kg^{-1}$ body weight of an orally administered test material is the maximum dose recommended for a testing strategy that has been developed for new pharmaceutical excipients that takes into consideration the physicochemical nature of the product and the potential route(s) and duration of exposures, both through its intended use as part of a drug product and through workplace exposure during manufacturing. The number and types of studies recommended in this tiered approach are based on the duration and routes of potential human exposure. Thus, the longer the exposure to the new pharmaceutical excipient, the more studies are necessary to assure safety. Table 10 summarizes the entire set of toxicological studies recommended for new pharmaceutical excipients (Wiener and Katkoskie, 1999; IPEC, 1997).

Tests have been outlined for each exposure category to assure safe use of the time period designated. The tests for each exposure category assure the safe use of the new pharmaceutical excipient of the time frame specified for the specific exposure category. Additional tests are required for longer exposure times.

The base set required for all excipients is detailed in Table 12. These are sufficient, however, only for those excipients intended for use for up to two weeks in humans.

**TABLE 11 Limit Doses for Toxicological Studies**

| Nature of Test | Species | Limit Dose[a] |
|---|---|---|
| Acute oral | Rodent | 2000 mg $kg^{-1}$ bw |
| Acute dermal | Rabbit<br>Rat | 2000 mg $kg^{-1}$ bw |
| Acute inhalation[b] | Rat | 5 mg $L^{-1}$ air for 4 h or maximum attainable level under conditions of study |
| Dermal irritation | Rabbit | 0.5 mL liquid<br>0.5 g solid |
| Eye irritation | Rabbit | 0.1 mL liquid<br>100 mg solid |
| 14-day/28-day oral repeated dosing; 90-day subchronic | Rodent, nonrodent | 1000 mg $kg^{-1}$ $bw^{-1}$ day |
| 14-day/28-day oral repeated dosing; 90-day subchronic | Rat, rabbit | 1000 mg $kg^{-1}$ $bw^{-1}$ day |
| Chronic toxicity, carcinogenicity | Rats, mice | 5% maximum dietary concentration for nonnutrients |
| Reproduction | Rats | 1000 mg $kg^{-1}$ $bw^{-1}$ day |
| Developmental toxicity (teratology) | Mice, rats, rabbits | 1000 mg $kg^{-1}$ $bw^{-1}$ day |

[a]mg $kg^{-1}$ bw, milligrams of test material dosed per kilogram of body weight to test species.
[b]Acute inhalation guidelines that indicate this limit dose are U.S. Environmental Protection Agency Toxic Substance Health Effect Test Guidelines, Oct. 1984; (PB82-232984) Acute Inhalation Toxicity Study; the OECD Guidelines of the Testing of Chemicals, Vol. 2, Section 4; Health Effects, 403, Acute Inhalation Toxicity Study, May 12, 1982; and the *Official Journal of the European Communities*, L383A, Vol. 35, Dec. 29, 1992, Part B.2 (adapted from Wiener and Katkoskie, 1999).

**TABLE 12 Base Set Studies for Single Dose Up to 2 Weeks Exposure in Humans**

| Test | Purpose |
|---|---|
| Acute oral toxicity | To determine potential acute toxicity–lethality following single oral dose |
| Acute dermal toxicity | To determine potential acute toxicity–lethality following single dermal dose |
| Acute inhalation toxicity | To determine potential acute toxicity–lethality following single 4-h inhalation exposure to test atmosphere containing new pharmaceutical excipient (aerosol, vapor, or particles) |
| Eye irritation | To determine potential to produce acute irritation or damage to eye |
| Skin irritation | To determine potential to produce acute irritation or damage to skin |
| Skin sensitization | To determine potential to induce skin sensitization reactions |
| Ames test | To evaluate potential mutagenic activity in bacterial reverse mutation system with and without metabolic activation |
| Micronucleus test | To evaluate clastogenic activity in mice using polychromatic erythrocytes |
| ADME—intended route | To determine extent of absorption, distribution, metabolism, and excretion by intended route of exposure following single dose and repeated doses |
| 28-Day toxicity—intended route | To assess repeated-dose toxicity in male and female animals of two species following dosing for 28 days by intended route of exposure |

**TABLE 13 Studies for Intermediate Duration (28 Day to 3 Months) Exposure to Humans**

| Test | Purpose |
|---|---|
| 90-Day Toxicity—intended route (Rodent and nonrodent) | To assess the repeated-dose toxicity in male and female animals of two species following daily dosing for 90 days by the intended route of administration |
| Developmental Toxicity | To assess the effects of dosing of pregnant female animals by the intended route during the period of organogenesis |

If exposure to the new pharmaceutical excipient is expected to occur for longer than two but no more than six weeks, additional toxicological studies are required, as shown in Table 13. The longer the expected human exposure, the more extensive will be the toxicological studies to assure safety. A tiered approach assures that those tests necessary to ensure safety for the expected duration of human exposure are conducted. Thus, to assure safe use for greater than two weeks but no more than six weeks in humans, subchronic toxicity and developmental toxicity studies are required. To assure safe use for greater than six continuous weeks, chronic or oncogenicity studies are conditionally required, as per Table 14. This means long-term studies should be considered for prolonged human exposures but may not be absolutely required. A thorough scientific review of the data generated in the base set and Appendix 2 studies should be undertaken. From a critical evaluation by

**TABLE 14 Appendix 3 Studies for Repeated Chronic Exposure in Humans**

| Test | Purpose |
|---|---|
| Chronic toxicity | To assess toxicity following chronic (lifetime) exposure by route of intended exposure |
| Oncogenicity | To assess potential to induce tumors by intended route of exposure |
| One-generation reproduction | To assess potential reproductive and developmental toxicity in males and females by intended route of exposure |

a competent toxicologist, the results of the physicochemical properties of the test material, the 28- and 90-day tests, the ADME-PK (absorption distribution metabolism and excretion pharmacokinetics) acute and repeated-dose tests, and the developmental toxicity test(s), a final determination can be made on the value of chronic toxicity or oncogenicity studies.

For example, if no toxicity is observed at a limit dose of 1000 mg $kg^{-1}$ body weight per day following the 90-day toxicity study, no genotoxicity was found, and the ADME-PK profile indicates that the material is not absorbed and is completely excreted unchanged in the feces, then it is likely that a chronic study is not necessary. The decision to conduct chronic studies should be determined on a case-by-case basis using scientific judgment. It will be interesting to observe how this scheme may change in light of the International Conference on Harmonisation (ICH).

## APPENDIX PRIMARY POTENTIAL FORMULATION COMPONENTS FOR NONCLINICAL TOXICITY STUDIES

| Excipient/Vehicle | Chemical Abstract Service Number | Chemical Name | Animal Studies |
|---|---|---|---|
| 2-Hydroxypropyl-β-cyclodextrin | 128446-35-5 | | Rat, primate, mouse, rabbit, dog |
| Acacia | 9000-01-5 | Acaciae gummi | Rat, primate |
| Acetate, sodium | 127-09-3 | Acetic acid sodium salt | Rat |
| Acetic acid | 64-19-7 | Ethanolic acid | Rat, mouse |
| Acetone | 67-64-1 | 2-Propanone | Rat, mouse, guinea pig, rabbit |
| Acetylmethylamine in water | 79-16-3 | *N*-Methylacetamide | |
| Alginic acid | 9005-32-7 | Norgine | Rat |

| Excipient/Vehicle | Chemical Abstract Service Number | Chemical Name | Animal Studies |
|---|---|---|---|
| Anecortave acetate | 7753-60-8 | — | Rat |
| Benzoic acid | 65-85-0 | Benzoic acid | Rat |
| β-Cyclodextrin | 7585-39-9 | β-Dextrin | Rat, primate |
| BHT | 128-37-0 | Butylated hydroxytoluene | |
| Canola oil | 120962-03-0 | Canbra oil | Dog |
| Capryol 90 | 31565-12-5 | Propylene glycol monocaprylate | Rat, dog, rabbit |
| Captisol | 182410-00-0 | β-Cyclodextrin sulfobutyl ether, sodium salt (CDSBE) | Rat, primate, mouse |
| Carboxymethylcellulose (CMC) | 9000-11-7 | acetic acid; 2,3,4,5,6-pentahydroxyhexanal | Primate, rat |
| Carboxymethylcellulose calcium | 9050-04-8 | Calcium CMC | Dog |
| Carboxymethylcellulose sodium | 9004-32-4 | Carmellose sodium | Rabbit |
| Cavasol W7 | 128446-35-5 | 2-Hydroxypropyl cycloheptaamylose | |
| Cetyl alcohol | 36653-82-4 | Hexadecan-1-ol | Mouse |
| Citrate buffer | 77-92-9 | sodium citrate–citric acid buffer | Dog, rat |
| Citric acid buffer | 77-92-9 | | Rat |
| CMC with dimethicone | 9004-32-4<br>9006-65-9 | Carboxymethylcellulose sodium<br>trimethyltrimethylsilyloxysilane | |
| Coconut oil | 8001-31-8 | N/A | |
| Collagen matrix | 9007-34-5 | Collagen human | Primate, rabbit |
| Corn oil | 8001-30-7 | corn germ oil, glyceridic | Dog, rat, mouse, rabbit, chick embryo |
| Cremophore EL | 61791-12-6 | Polyoxyl castor oil | Dog, rat |
| Cyclohexane | 110-82-7 | Hexahydrobenzene; hexamethylene; hexanaphthene | Rat, rabbit |
| DAM PEG (polyethyllene Glycol) | | | Dog, rat |
| Dextrose | 50-99-7 | D-Glucose, anhydrous; dextrosol | Dog, rat |
| Diethyleneglycolmonoethylether | 111-90-0 | | Primate |
| DMSO | 67-68-5 | Dimethylsulfoxide | Dog, rat, guinea pig, primate, mouse, rabbit |
| Dulbecco's modified PBS | | | Rat |
| EDTA | 60-00-4 | Ethylenediamineetraacetic acid | |

| Excipient/Vehicle | Chemical Abstract Service Number | Chemical Name | Animal Studies |
|---|---|---|---|
| Ethanol | 64-17/5 | Ethyl alcohol | Dog, rat, primate, mouse |
| Gelucire 44/14 | 121548-04-7 | PEG-32 glyceryl laurate | Rabbit, rat, dog |
| Gelucire 50/13 | 121548-05-8 | G-50-13 | Rabbit, rat, dog |
| Glucose | 50-99-7 | Dextrose | Dog, rat, primate |
| Glycerol | 56-81-5 | Glycerine | Rat, guinea pig, mouse, rabbit |
| Gum tragacanth | 9000-65-1 | | Mouse |
| Gum xanthane | 11138-66-2 | | |
| Hydroxypropyl β-cyclodextrin | 94035-02-6 | | Dog, rat |
| Hydroxypropyl cellulose | 9004-64-2 | Methocel | Rat |
| Hydroxypropyl methylcellulose | 9004-65-3 | Benecel MHPC, hypromellose | Dog, rat, mouse |
| Isopropyl alcohol | 67-63-0 | *sec*-Propyl alcohol | Rabbit |
| Isopropyl myristate | 110-27-0 | Crodamol IPM | Rabbit |
| Labrafil M1944 | 62563-68-2 | Labrafil | Dog |
| Labrasol | 85536-07-8 | Polyglycolyzed glycerides | Rat, dog, rabbit |
| Lauroglycol | 27194-74-7 | Lauric acid, monoester with propane-1,2-diol | Rabbit, rat |
| Lactose | 63-42-3(anhy) | *O*-β-D-Galactopyranosyl-(1->4)-α-D-glucopyranose | Primate |
| Lanolin | 8006-54-0 | Lanolin | Rabbit |
| L-Ascorbic acid | 50-81-7 | Cevatine, Cevex, Cevital | Rat |
| Maltitol solution | 9053-46-7 | Liquid maltitol | Rat |
| Maltol | 118-71-8 | 3-Hydroxy-2-methyl-4*H*-pyran-4-one | Guinea pig, rabbit |
| Mannitol | 69-65-8 | D-Mannitol | Primate |
| Methane sulfonic acid | 75-75-2 | Methylsulfonic acid | |
| Methyl cellulose | 9004-67-5 | Cellulose methyl ester | Rat, guinea pig, primate, mouse, rabbit, dog |
| Miglyol 810 | 85409-09-2 | Caprylic, capric triglycerides | |
| Mineral oil | 8012-95-1 | Liquid paraffin | Rat, mouse, dog |
| Neobee 1053 | 73398-61-5 | Medium-chain triglycerides | |
| *N*-Methylpyrrolidone (Pharmasolv) | 872-50-4 | 1-Methyl-2-pyrrolidinone | |

| Excipient/Vehicle | Chemical Abstract Service Number | Chemical Name | Animal Studies |
|---|---|---|---|
| PBS (phosphate-buffered saline) | | | Rat, primate, mouse |
| Peanut oil | 8002-03-7 | Arachis oil, Fletcher's | Rat |
| PEG 300 | 25322-68-3 | Polyethylene glycol #300 | Guinea pig, mouse, rabbit |
| PEG 400 | 25322-68-3 | Polyethylene glycol #400 | Rat, minipig, guinea pig, mouse |
| Petrolatum | 8009-03-8 | Yellow soft paraffin | Rabbit |
| Poloxamer | 9003-11-6 | Lutrol | Rat, mouse |
| Polysorbate 80 | 9005-65-6 | Polyoxyethylene (20) sorbitan monooleate | |
| Povidone | 9080-59-5 | 2-methoxy-6-methyl-phenol | Rat |
| Propylene glycol | 57-55-6 | 1,2-Dihydroxypropane | Rat, minipig, mouse, dog |
| Rameb 7.5% | | Randomly methylated-β-cyclodextrins | Primate |
| Sesame oil | 8008-74-0 | Sesame oil | Rat, mouse, rabbit, dog |
| Sodium acetate trihydrate buffer | 6131-90-4 | | Primate |
| Sodium chloride | 7647-14-5 | Salt, halite | Dog, rat, primate, mouse, rabbit |
| Sodium phosphate | 7558-80-7 | | Dog, rat |
| Solutol® HS15/purified water | 70142-34-6 | Polyethylene glycol-15-hydroxystearate | |
| Succinate, sodium | 150-90-3 | Succinic acid Sodium salt | |
| Tartaric acid | 87-69-4 | D-Tartaric acid; 2,3-dihydroxybutanedioic acid | Rat, rabbit |
| Transcutol | 111-90-0 | 2-(2-Ethoxyethoxy)ethanol | Cat, rabbit, rat |
| Trisodium citrate dihydrate | 6132043 | | Dog, rat, mouse |
| Tween 20 | 9005-64-5 | Polysorbate 20 NF | Rat, mouse |
| Tween 80 | 9005-65-6 | armotan pmo-20, Tween(R) 80 | Rat, primate, mouse, dog |
| Xylitol | 87-99-0 | Xylite | Primate |

*Source*: Gad et al., 2006.

## REFERENCES

American Pharmaceutical Association (APhA) (1994). *Handbook of Pharmaceutical Excipients*, 2nd ed. APhA, Washington, DC.

American Pharmaceutical Association (APhA) (2007). *Handbook of Pharmaceutical Excipients*, 5th ed. APhA, Washington, DC.

Attman, P. L., and Dittmer, D. S. (Eds.) (1971). *Respiration and Circulation*. FASEB, Bethesda, MD, pp. 56–59.

Avis, K. E. (1985). Parenteral preparations. In A. R. Gennaro (Ed.), *Remington's Pharmaceutical Sciences*. Mack Publishing, Easton, PA, pp. 1518–1541.

Baker, H. J., Lindsey, J. R., and Weisbroth, S. H. (1979). *The Laboratory Rat*, Vol. 1. Academic, New York, pp. 411–412.

Ballard, B. E. (1968). Biopharmaceutical considerations in subcutaneous and intramuscular drug administration. *J. Pharm. Sci.* 57:357–378.

Barry, B. W. (1983). *Dermatological Formulations*. Marcel Dekker, New York.

Bates, T. R., and Gibaldi, M. (1970). Gastrointestinal absorption of drugs. In J. Swarbrick (Ed.), *Current Concepts in the Pharmaceutical Sciences: Biopharmaceutics*. Lea & Febiger, Philadelphia, pp. 58–99.

Blank, 1. H., and Scheuplein, R. J. (1964). The epidermal barrier. In W. Rook and 1. Champion (Eds.), *Progress in the Biological Sciences in Relation to Dermatology*, Vol. 2. Cambridge University Press, Cambridge.

Blodinger, J. (1982). *Formulation of Veterinary Dosage Forms*. Marcel Dekker, New York.

Borowitz, J. L., Moore, P. F., Yim, G. K. W., and Miya, T. S. (1971). Mechanism of enhanced drug effects produced by dilution of the oral dose. *Toxicol. Appl. Pharmacol.* 19:164–168.

Brandau, R., and Lippold, B. H. (1982). *Dermal and Transdermal Absorption*. Wissenschaftliche, Verlagsgesellschaft mbH, Stuttgart.

Bronaugh, R. L., and Maibach, H. I. (1985). *Percutaneous Absorption*. Marcel Dekker, New York.

Brown, V. K. (1980). *Acute Toxicity in Theory and Practice*. Wiley, New York.

Brown, V. K., and Muir, C. M. C. (1971). Some factors affecting the acute toxicity of pesticides to mammals when absorbed through skin and eyes. *Int. Pest Control* 13:16–21.

Budavari, S. (1989). *The Merck Index*, 11th ed. Merck & Co., Rahway, NJ.

Calvary, H. O., and Klump, T. G. (1939). The toxicity for human beings for diethylene glycol with sulfanilimide. *Southern Med. J.* 32:1105.

Derelanko, M. J., Gad, S. C., Gavigan, F. A., Babich, P. C., and Rinehart, W. E. (1987). Toxicity of hydroxylamine sulfate following dermal exposure: Variability with exposure method and species. *Fund. Appl. Toxicol.* 8:425–431.

Dominguez, J. G., Parra, J. L., Infante, M. R., Pelejero, C. M., Balaguer, F., and Sastre, T. (1977). *J. Soc. Cosmet. Chem.* 28:165.

Eckstein, P., Jackson, M. C. N., Millman, N., and Sobrero, A. J. (1969). Comparison of vaginal tolerance tests and spermicidal preparations in rabbits and monkeys. *J. Reprod. Fertil.* 20:85–93.

Fitzgerald, J. M., Boyd, V. F., and Manus, A. G. (1983). Formulations of insoluble and immiscible test agents in liquid vehicles for toxicity testing. In C. W. Jameson and D. B. Walters (Eds.), *Chemistry for Toxicity Testing*. Butterworth, Boston, pp. 83–90.

Gad, S.C. (2007). *Target Organ Toxicity: The GI Tract*. Taylor & Francis, Philadelphia.

Gad, S. C. (Ed.) (2008). *Preclinical Development Handbook: ADME and Biopharmaceutical Properties*. Wiley, Hoboken, NJ.

Gad, S. C., Cassidy, C., Aubert, N., Spainhour, B., and Robbe, H. (2006). Nonclinical vehicle use in studies by multiple routes in multiple species. *Int. J. Toxicol.* 25:499–522.

Gad, S. C., Smith, A. C., Cramp, A. L., Gavigan, F. A., and Derelanko, M. J. (1984). Innovative designs and practices for acute systemic toxicity studies. *Drug Chem. Toxicol.* 7(5):423–434.

Gaines, T. B. (1960). The acute toxicity of pesticides to rats. *Toxicol. Appl. Pharmacol.* 2:88–99.

Garramone, J. P. (1986). *Vascular Access Port Model SLA Users Manual*. Norfolk Medical Products, Skokie, IL.

Gibaldi, M. (1976). Peptide formulation. In L. Lachman, H. A. Lieberman, and J. L. Kanig (Eds.), *Biopharmaceutics in the Theory and Practice of Industrial Pharmacy*, 2nd ed. Lea & Febiger, Philadelphia.

Gibaldi, M., and Feldman, S. (1970). Mechanisms of surfactant effects on drug absorption. *J. Pharm. Sci.* 59:579.

Gray, J. E. (1978). Pathological evaluation of injection injury. In J. Robinson (Ed.), *Sustained and Controlled Release Drug Delivery System*. Marcel Dekker, New York, pp. 351–405.

Groves, M. (1966). The influence of formulation upon the activity of thermotherapeutic agents. *Reps. Progr. Appl. Chem.* 12:51–151.

Hagenbusch, O. E. (1937). Elixir of sulfanilamide massengill. *J. Am. Med. Assoc.* 109:1531.

Hawley, G. G. (1971). *The Condensed Chemical Dictionary*. Van Nostrand Reinhold, New York.

Hem, S. L. (1973). The effect of formulation additives on drug bioavailability from oral solutions and suspensions. In J. Swarbrick (Ed.), *Current Concepts in Pharmaceutical Sciences: Dosage Form, Design and Bioavailability*. Lea & Febiger, Philadelphia.

Hogben, C. A. M., Tocco, D. 1., Brodie, B. B., and Schranker, L. S. (1959). On the mechanism of intestinal absorption of drugs. *J. Pharmacol. Exp. Ther.* T25:275.

Holick, H. G. O. (1937). Glycerine, ethylene glycol, propylene glycol and diethylene glycol. *J. Am. Med. Assoc.* 109:1517.

Holmstedt, B. (1959). Pharmacology of organophosphorus cholinesterase inhibitors. *Pharmacol. Revs.* 11:567–688.

Houston, J. B., Upshall, D. G., and Bridges, J. W. (1974). The re-evaluation of the importance of partition coefficients in the gastrointestinal absorption of nutrients. *J. Pharmacol. Exp. Ther.* 189:244–254.

International Pharmaceutical Excipients Council (IPEC) (1997). Europe Safety Committee: The proposed new guidelines for safety evaluation of new excipients, *Eur. Pharm. Rev.* Nov.: 13–20.

Katdare, A., and Chaubal, M. V. (2006). *Excipient Development for Pharmaceutical Biotechnology and Drug Delivery Systems*. Informa Healthcare, New York.

Kerberle, H. (1971). Physiocochemical factors of drugs affecting absorption, distribution and excretion. *Acta Pharmacol. Toxicol.* 29(Suppl 3):30–47.

Kondrizer, A. A., Mayer, W. H., and Zviblis, P. (1959). Removal of sarin from skin and eyes. *Arch. Ind. Health* 20:50–52.

Kruger, S., Greve, D. W., and Schueler, F. W. (1962). The absorption of fluid from the peritoneal cavity. *Arch. Int. Pharmacodyn.* 137:173–178.

Levi, G. (1963). In J. B. Sprowls (Ed.), *Prescription Pharmacy*. Lea & Febiger, Philadelphia.

Lien, E., Koda, R. T., and Tong, G. L. (1971). Buccal and percutaneous absorptions. *Drug Intel. Clin. Pharm.* 5:38–41.

Lukas, G., Brindle, S. D., and Greengard, P. (1971). The route of absorption of intraperitoneally administered compounds. *J. Pharmacol. Exp. Ther.* 178:562–566.

Marzulli, F. N. (1962). Barriers to skin penetration. *J. Invest. Dermatal* 39:387–393.

McConnell, R. F. (1973). Special requirements for testing spermicides. In *Meeting on Pharmaceutical Models to Assess Toxicity and Side Effects of Fertility Regulating Agents*. International Life Science Institute, Geneva, pp. 373–385.

McMartin, C., Hutchinson, L. E. F., Hyde, R., and Peters, G. E. (1987). Analysis of structural requirements for the absorption of drugs and macromolecules from the nasal cavity. *J. Pharm. Sci.* 76:535–540.

Medved, L. L., and Kundiev, Y. I. (1964). On the methods of study of penetration of chemical substances through the intact skin (Russian). *Gig. Sanit.* 29:71–76.

Melnick, R. L., Jameson, C. W., Goehl, T. J., and Kuhn, G. O. (1987). Application of microencapsulation for toxicology studies. *Fund. Appl. Toxicol.* 11:425–431.

Monkhouse, D. C., and Rhodes, C. T. (1998). *Drug Products for Clinical Trials*. Marcel Dekker, New York.

O'Reilly, W. J. (1972). Pharmacokinetics in drug metabolism and toxicology. *Can. J. Pharm. Sci.* 7:6–77.

Powers, D. H., and Fox, C. (1957). A study of the effect of cosmetic ingredients, creams and lotions on the rate of moisture loss. *Proc. Sci. Sect. Toilet Goods Skin Assoc.* 28:21.

Racy, I. (1989). *Drug Formulation*. Wiley, New York.

Rieger, M. M., and Deems, D. E. (1974). Skin moisturizers. II. The effects of cosmetic ingredients on human stratum comeum. *J. Soc. Cosmet. Chem.* 25:253.

Ritschel, W. A., Siegel, E. G., and Ring, P. E. (1974). Biopharmaceutical factors influencing LDP: Pt. 1 Viscosity. *Arzneim.-Forsch.* 24:907–910.

Scheuplein, R. J. (1965). Mechanism of percutaneous absorption. (i) Routes of penetration and the influence of solubility. *J. Invest. Dermatol.* 45:334–346.

Scheuplein, R. J. (1967). Mechanism of percutaneous absorption. (ii) Transient diffusion and the relative importance of various routes of skin penetration. *J. Invest. Dermatol.* 48:79–88.

Schmidt-Nielsen, K. (1984). *Scaling: Why Is Animal Size So Important?* Cambridge University Press, New York.

Schou, J. (1971). Subcutaneous and intramuscular injection of drugs. In B. B. Brodie and J. R. Gillette (Eds.), *Handbook of Experimental Pharmacology*. Springer-Verlag, Berlin, Chapter 4.

Schranker, L. S. (1960). On the mechanism of absorption of drugs from the gastrointestinal tract. *J. Med. Pharm. Chem.* 2:343–359.

Schranker, L. S., Shore, P. A., Brodie, B. B., and Hogben, C. A. M. (1957). Absorption of drugs from the stomach, I. The rat. *J. Pharmacol. Exp. Ther.* 120:528.

Schwartz, M. L., Meyer, M. B., Covino, B. G., and Narang, R. M. (1974). Antiarrhythmic effectiveness of intramuscular lidocaine: Influence of different injection sites. *J. Clin. Pharmacol.* 14:77–83.

Shaffer, C. B., and West, B. (1960). The acute and subacute toxicity of technical *o,o*-diethyl *s*-2-diethyl-aminoethyl phosphorothioate hydrogen oxalate (TETRAM). *Toxicol. Appl. Pharmacol.* 2:1–13.

Share, P. A., Brodie, B. B., and Hogben, C. A. M. (1971). The gastric secretion of drugs: A pH partition hypothesis, *J. Pharmacol. Exp. Ther.* 119:361.

Sinow, J., and Wei, E. (1973). Ocular toxicity of paraquat. *Bull. Environ. Contam. Toxicol.* 9:163–168.

Smolinske, S. C. (1992). *Handbook of Food, Drug and Cosmetic Excipients*. CRC Press, Boca Raton, FL.

Spiegel, A. J., and Noseworthy, M. M. (1963). Use of nonaqueous solvents in parenteral products. *J. Pharm. Sci.* 52:917–927.

Staab, R. J., Palmer, M. A., Auletta, C. S., Blaszcak, D. L., and McConnell, R. F. (1987). A relevant vaginal irritation/subacute toxicity model in the rabbit and ovariectomized rat. *Toxicologist* 7A:1096.

Swan, K. C., and White, N. G. (1972). Corneal permeability: (1) factors affecting penetration of drugs into the cornea. *Am. J. Ophthalmol.* 25:1043–1058.

Theeuwes, F., and Yum, S. L. (1976). Principles of the design and operation of generic osmotic pumps for the delivery of semisolid or liquid drug formulations. *Ann. Biomed. Eng.* 4:343–353.

Turco, S., and King, R. E. (1974). *Sterile Dosage Forms*. Lea & Febiger, Philadelphia.

Unites States Pharmacopeia (USP) (2007). USP 30/NF 25. USP, Washington, DC.

U.S. Food and Drug Administration (FDA) (May 2005). Guidance for industry: Nonclinical studies for the safety evaluation of pharmaceutical excipients. FDA, Washington, DC.

Wagner, J. G. (1971). *Biopharmaceutics and Relevant Pharmacokinetics*. Drug Intelligence Publications, Hamilton, IL.

Weiner, M., and Bernstein, I. L. (1989). *Adverse Reactions to Drug Formulation Agents*. Marcel Dekker, New York.

Wiener, M. L., and Katkoskie, L. A. (1999). *Excipient Toxicity and Safety*, Marcel Dekker, New York.

Wilson, F. A., and Dietschy, J. M. (1974). The intestinal unstirred layer—Its surface area and effect on active transport kinetics. *Biochem. Biophys. Acta* 363:112–126.

Winne, D. (1978). Dependence of intestinal absorption *in vivo* in the unstirred layer. *Nauyn-Schmiedelberg's Arch. Pharmacol.* 304:175–181.

Wood, S. G., Upshall, D. G., and Bridges, J.W. (1978). The absorption of aliphatic carbamates from the rat colon. *J. Pharm. Pharm.* 30:638–641.

Ziegenmeyer, J. (1982). The influence of the vehicle on the absorption and permeation of drugs. In R. Brandan and P. Reisen (Eds.), *Dermal and Transdermal Absorption*. Wissenschaftliche, Verlagsgesellschaft, Stuttgart, pp. 73–89.

Zimm, B. H., and Lundberg, J. L. (1956). Sorption of vapors by high polymers. *J. Phys. Chem.* 60:425.

# *Single-Dose (Acute) and Pilot (DRF) Toxicity Testing in Drug Safety Evaluation*

## 1 INTRODUCTION

Acute toxicity testing is the defining and evaluation of the toxic syndrome (if any) produced by a single or a few doses over the course of a day, such as twice or three times per day (bid or tid, in the case of continuously infused intravenous formulation in a 24-h course of treatment) of a drug. Historically, the main focus of these tests has been lethality determinations and the identification of overt signs and symptoms of overdosage. For a complete historical perspective, see Deichmann and Gerarde (1969), Piegorsh (1989), Auletta (1998), Gad and Chengelis (1999), or Rhodes (2000). A more enlightened and modern view holds that, especially for pharmaceutical agents, lethality in animals is a relatively poor predictor of hazard (other than lethality) in humans (Gad and Chengelis, 1999). The current trend is toward gaining increasing amounts of more sophisticated data from these tests. The various types of acute study designs, their utility in pharmaceutical product testing, their limitation, and the resultant sample data are discussed in this chapter.

For new product approvals (and first in human clinically trials), single-dose toxicity studies are required by regulatory authorities though this requirement

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

is being challenged in the European communities [Osterberg, 1983; U.S. Food and Drug Administration (FDA), 1996, 2006]. In the pharmaceutical industry, acute toxicity testing has uses other than simply product safety determinations. First, as in other industries, acute toxicity determinations are part of industrial hygiene or occupational health environmental impact assessments. These requirements demand testing not only for finished products but frequently of intermediates as well. These issues and requirements, however, are discussed in that content in Chapter Regulation of Human Pharmaceutical Safety and are not directly addressed here.

Another use, now almost abandoned except in natural product–derived drugs (Pendergast, 1984), is in quality control testing or batch release testing. The latter was once a mandated part of the standardization process for antibiotics, digoxin, and insulin in the U.S. Pharmacopeia. While, perhaps, this type of testing is part of a broad safety picture, it is not typically part of a "preclinical" safety package used to make decisions on whether to market a new chemical entity or on what the allowable clinical dosage shall be. These uses also therefore are not discussed here. The emphasis in this volume, rather, is on tests used to elucidate the toxicity of new chemical entities, not the safely of finished drug preparations. These tests fall into three general categories: (1) range-finding studies, used primarily to set dosages for initial subchronic or acute testing; (2) complete "heavy" or expanded acute toxicity tests, used to thoroughly describe the single-dose toxicity of a chemical or to support the opening of an explanatory or phase-zero investigational new drug (IND); and (3) screening tests, used to select candidates for development.

## 2 RANGE-FINDING STUDIES

Range finders or pilots [now also commonly called dose range finders (DRFs)] are not normally done completely under the auspices of the Good Laboratory Practices Act. They are not used to generate data to support decisions on human safety; rather, they are used to allow successful dose selection for definitive toxicity studies. These dosage-level determinations can be used in acute studies, in in vivo genotoxicity studies, or subchronic studies. As discussed by Gad and Chengelis (1999), however, there can be a great deal of difference between the acute toxic dosage and subchronic daily dosage of a drug. Therefore, acute range-finding studies currently most common include a component (or second phase) whereby a second set of animals will receive a short-term treatment (up to seven days) with the drug in question. Accordingly, the definition of "acute" in this chapter is stretched to include "subacute" (lower than acute level) dosing of very short duration.

### 2.1 Lethality Testing

Often, in range-finding tests, the endpoint is simply to determine the maximum dosage of a drug that can be given without killing an animal. There are numer-

ous designs available for obtaining this information that minimize the expenditure of animals and other resources.

***Classical $LD_{50}$*** The median lethal dose ($LD_{50}$) test has a rich and controversial history (it is one of a number of tests that raises the ire of the animal welfare movement) (Trevan, 1927; Rowan, 1981; LeBeau, 1983). In pharmaceutical development, however, there is rarely a need or requirement for an $LD_{50}$. In general, a complete, precisely calculated $LD_{50}$ consumes more resources than is generally required for range-finding purposes. The reader is referred to Chapter Genotoxicity of Gad and Chengelis (1999) for a complete discussion of this test.

***Dose Probes*** Dose probe protocols (see Figure 1) are of value when one needs the information supplied by a traditional protocol but has no preliminary data from which to choose dosages. In this acute protocol, one animal is dosed at each of three widely spaced dosages, where the top dosage is generally the maximum deliverable. The method works best if the dosages are separated by constant multiples (e.g., 3000, 300, and 30 mg kg$^{-1}$—a logarithmic progression). Subsequent dosages are selected on the basis of the results from these probe animals. If none of these animals dies, the protocol defaults to a limit test (described below), and two more animals are dosed at the top dosage to confirm the limit.

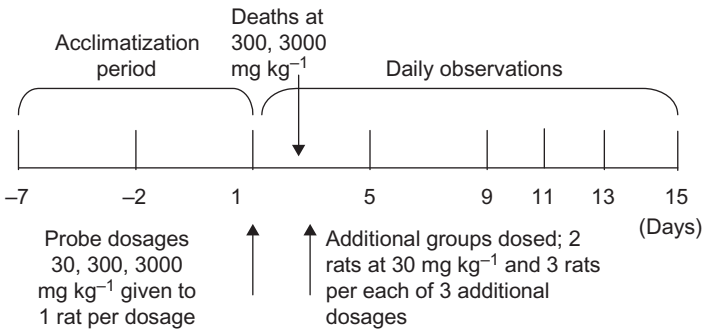


Results

| Dosage | Mortality |
|---|---|
| 30 mg kg$^{-1}$ | 0/3 |
| 60 | 1/3 |
| 120 | 2/3 |
| 240 | 2/3 |

$LD_{50}$ = 115 mg kg$^{-1}$ (moving-average method)

***Figure 1*** *Example of typical dosage probe protocol.*

A dose probe can develop into a more thorough lethality determination. If one or two animals die, then two additional dosages between the lethal and nonlethal dosages are chosen and three animals are treated per dosage for defining acute lethality. Selection of these dosages is often a matter of personal judgment. If, for example, one wishes to apply the moving-average method of calculation, these subsequent dosages can be either even fractions of the top dosage or even multiples of the low dosage. In either case, two to three animals are dosed at the initial dose and three to four animals are dosed at each of the two to three new dosages. The results should be three to four groups of three to four animals each, which should probably provide sufficient data for calculating the $LD_{50}$ and the slope of the curve. Probing can also be used to define the dosages for subchronic tests. Instead of selecting additional doses for an acute study, one can use the results from the probe to select two dosages for a short (e.g., five-day) daily dosing regimen (see the later section entitled "'Rolling' Acute Tests").

In a few instances, all the animals may die following the first day of dosing. In that case, the probe activity continues on day 2 with two more animals dosed at two widely spaced lower dosages (i.e., 3 and 0.3 $mg\,kg^{-1}$). This regimen could continue daily until a nonlethal dosage is identified. Unless one has grossly misestimated the toxicity of the test substance, it is unlikely that the probing process would take more than three days. Carrying our example into three days of dosing would have resulted in probing in the range 3 $\mu g\,kg^{-1}$–3 $g\,kg^{-1}$, and it is a rare chemical that is lethal at less than 3 $\mu g\,kg^{-1}$. Once a nonlethal dosage is identified, additional animals and/or dosages can be added, as discussed above.

There are two disadvantages to dose probe studies. First, delayed deaths pose difficulties. Hence, all animals should be observed for at least seven days after dosing (though most deaths occur within three days). Second, if the follow-up dosages are not lethal, the next decision point is ill defined. Should more animals be dosed at some different dosage? The resulting data sets may be cumbersome and difficult to analyze by traditional statistical methods. Alternatively (and this is true regardless of protocol design), if no "partial response" (mortality greater than zero but less than 100%) dosage is identified, one can simply conclude that the $LD_{50}$ is between two dosages, but the data do not permit the calculation of the $LD_{50}$ or the slope of the curve. This can happen if the dosage response is fairly steep.

Lörke (1983) has developed a similar protocol design. His probe (or dose range) experiment consists of three animals per dosage at 10, 100, and 1000 $mg\,kg^{-1}$. The results of the experiment dictate the dosages for the second round of dosing, as shown in Table 1. Animals were observed for 14 days after dosing. Lörke (1983) compared the results obtained when one to five animals were used per dosage group for the second test. He concluded that using only one animal per group gives unreliable results in only 7% of chemicals tested. Hence, the Lörke design can produce reasonable estimates of lethal dosages using 14 or fewer animals. Schutz and Fuchs (1982) have proposed a dose

**TABLE 1 Dosage Selection for Two-Step Dose-Probing Protocol Design**

| Mortality by Dose[a] | | | Dosages ($mg\,kg^{-1}$) for Definitive Experiment as Determined by Results of Probe | | | |
|---|---|---|---|---|---|---|
| $10\,mg\,kg^{-1}$ | $100\,mg\,kg^{-1}$ | $1000\,mg\,kg^{-1}$ | | | | |
| 0/3 | 0/3 | 0/3 | | 1600 | 2900 | 5000 |
| 0/3 | 0/3 | 1/3 | 600 | 1000[b] | 1600 | 2900 |
| 0/3 | 0/3 | 2/3 | 200 | 400 | 800 | 1600 |
| 0/3 | 0/3 | 3/3 | 140 | 225 | 370 | 600 |
| 0/3 | 1/3 | 3/3 | 50 | 100[b] | 200 | 400 |
| 0/3 | 2/3 | 3/3 | 20 | 40 | 80 | 160 |
| 0/3 | 3/3 | 3/3 | 15 | 25 | 40 | 60 |
| 1/3 | 3/3 | 3/3 | 5 | 10[b] | 20 | 40 |
| 2/3 | 3/3 | 3/3 | 2 | 4 | 8 | 16 |
| 3/3 | 3/3 | 3/3 | 1 | 2 | 4 | 8 |

[a]Number of animals that died/number of animals used.
[b]The results from the probe is inserted for these doses.
*Source*: Lörke, 1983.

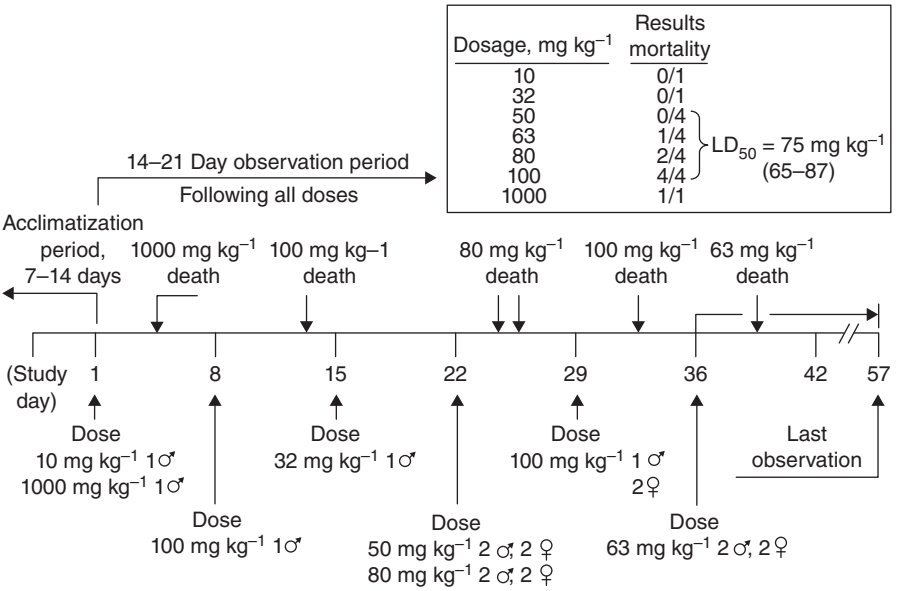


***Figure 2*** *Example of dose probe method with delayed deaths. From Schultz and Fuchs (1982).*

probe protocol that adequately deals with delayed deaths (Figure 2). All animals are observed for seven days before subsequent dosages are given. Dosing is initiated at two widely delivered dosages using one rate for each dosage. A third probe dosage is determined pending the outcome of the first two probes. A fourth may also be used. After that groups of three to four

animals are used at subsequent dosages either as part of a "para-acute" dosing regimen to select or confirm dosages for a subchronic study or to continue with the definition of an acute lethality curve.

***Up/Down Method*** Using classical or traditional acute lethality protocols, 15–30 animals per curve may be required to calculate a singe $LD_{50}$. This is because the method relies on the analysis of group responses. The up/down method can provide lethality information by analyzing the responses on an individual animal basis using appropriate statistical maximum-likelihood methods (Bruce, 1985). Deichmann and LeBlanc (1943) published an early method that provided an estimate of lethality using no more than 6 animals. All animals were dosed at the same time. The dosage range was defined as 1.5 times a multiplication factor (e.g., 1.0, 1.5, 2.2, 3.4, 5.1 mL $kg^{-1}$). The approximate lethal dose (ALD), as they defined it, was the highest dose that did not kill the recipient animal. The resultant ALD differed from the $LD_{50}$ (as positive calculated by the probit method from more complete data sets) by –22 to +33%.

The Deichmann method proved to be too imprecise (Muller and Kley, 1982). Later, Dixon and Wood (1948), followed by Brownlee et al. (1953), developed the method in which one animal was exposed per dosage, but subsequent dosages were adjusted up or down by some constant factor depending on the outcome of the previous dosage. In this method (Figure 3), which has been developed more extensively by Bruce (1985), individual animals are dosed at different dosages on successive days. If an animal dies, the dosage for the next animal is decreased by a factor of 1.3. Conversely, if an animal lives, the next dosage is increased by a factor of 1.3. The process is continued until

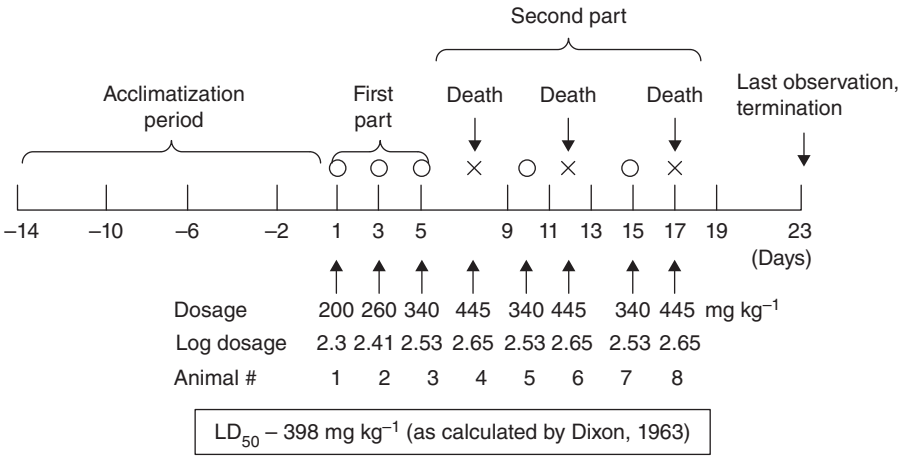


***Figure 3*** *Example of typical up/down acute lethality protocol.*

five animals have been dosed after a reversal of the first observation. Alternatively, one can use the tables developed by Dixon (1965). This design can be used not only for range-finding purposes but also to define an $LD_{50}$ if this value is needed. In general, only six to nine animals are required—unless the initial dosages are grossly high or low. When compared to the $LD_{50}$ obtained by other more classical protocols, excellent agreement is obtained with the up/down method (Bruce, 1985). As with classical protocols, sexes should be tested separately. However a further reduction in the numbers of animals used can be accomplished if one is willing to accept that females are of the same or increased sensitivity as males, as is the case approximately 85–90% of the time (Gad and Chengelis, 1999).

There are three main disadvantages to using the up/down method. The first is regulatory, the second procedural, and the third scientific. First, many regulatory guidelines simply have a requirement for the use of traditional protocols. Some also specify the method of calculation. Second, the sequential dosing design is inappropriate for substances that cause delayed deaths. As reported by various authors (Gad et al., 1984; Bruce, 1985), delayed deaths (beyond two days after dosing) are rare but not known. They are most prevalent when animals are dosed by the intraperitoneal route with a chemical that causes peritonitis. Death secondary to severe liver or gastrointestinal damage may also take over two days to occur. To guard against possible spurious results, all animals should be maintained and observed for at least seven days after dosing. If delayed deaths occur, the original data set must be corrected and the $LD_{50}$ recalculated. A substantial number of delayed deaths could result in a data set from which an $LD_{50}$ cannot be calculated, in which case the test should be rerun.

***"Pyramiding" Studies*** Using this type of design (Figure 4), one can obtain information about lethality with the minimum expenditure of animals. A minimum of two animals are dosed throughout the study, usually on alternate days (e.g., Monday, Wednesday, and Friday), but the dosage at session may be 1, 3, 10, 30, 100, 300, 1000, and 3000 $mg\,kg^{-1}$ or 10, 20, 40, 80, 160, 320, 640, and 1280 $mg\,kg^{-1}$. One is literally stepping up, or pyramiding, the lethality–dosage curve. Dosing continues in this fashion until one or both animals die or until some practical upward limit is reached. For drugs, there is no longer a need to go higher than 1000 $mg\,kg^{-1}$ for rodents or nonrodents. An alternative, but similar, design is the "leapfrog" study (Figure 5). This consists of two groups of two animals each. They are dosed on alternating days, but the dosages are increased each day. Extending the example of the pyramiding regiment, group 1 would receive 10, 60, and 120 $mg\,kg^{-1}$, while group 2 would be given 30, 100, and 120 $mg\,kg^{-1}$. This design is of value when one has to complete the range-finding activity in a short period of time. Because these designs utilize few animals, they are commonly used for assessing lethality in nonrodent species. An exploratory study typically uses an animal of each sex.

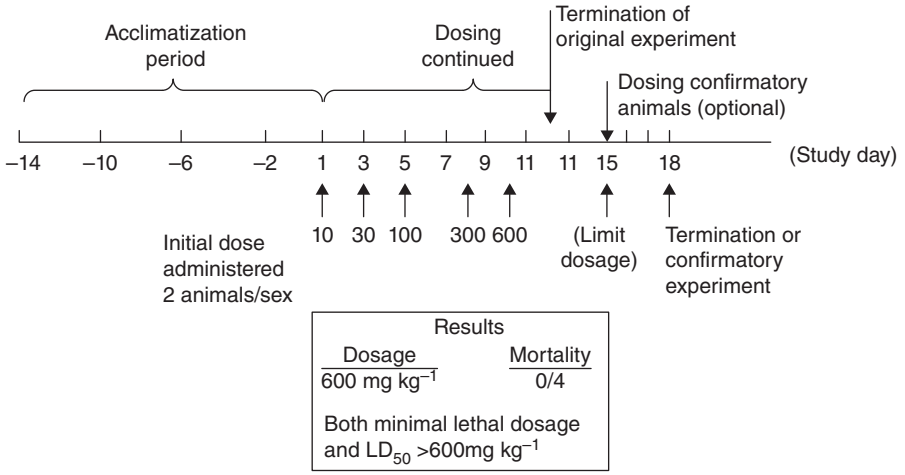


***Figure 4*** *Example of typical pyramiding dose protocol.*

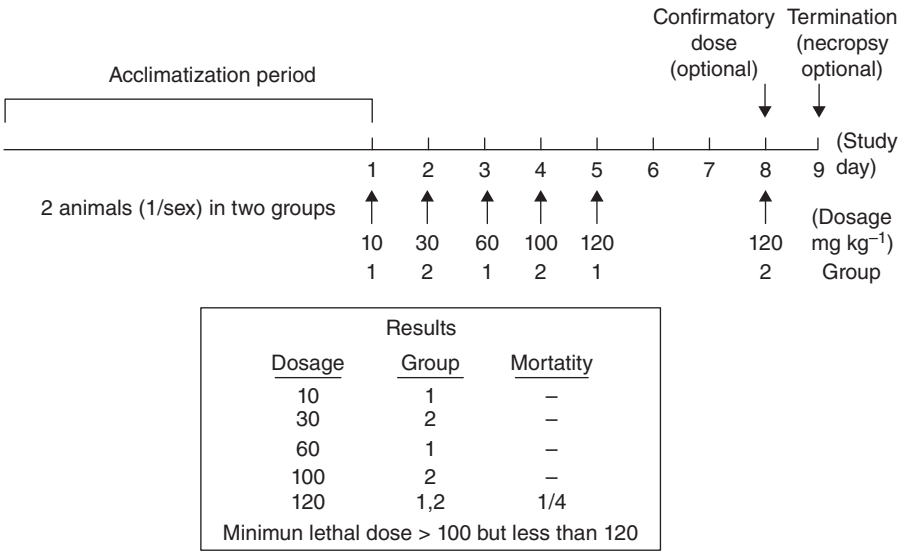


***Figure 5*** *Example of typical "leapfrog" dosing protocol.*

There are three conclusions that can be reached on the basis of data from a pyramiding dosage study. First, if none of the animals die, then both the threshold or minimum lethal dosage (MLD) and the $LD_{50}$ are greater than the top or limit dosage. Second, if all animals die at the same dosage, then both the MLD and the $LD_{50}$ are reported as being between the last two dosages given. This not uncommon finding is an indication that the lethality curve has a steep slope. Third, one animal may die at one dosage, and remaining deaths

occur at a subsequent dosage. In this care, the MLD is between the lowest nonlethal dosage and the dosage at which the first death occurred, while the $LD_{50}$ is reported as being between this latter dosage and the dosage at which the last animal dies. A frequently employed variation with nonrodents is, if lethality is not observed, the animals are dosed for five or seven consecutive days at the highest observed tolerated dose. This "phase B" study portion serves to provide more confidence in selecting the top dose in subsequent repeat-dose studies.

There are some disadvantages to the pyramiding dose protocol. First, it cannot produce a lethality curve or provide for the calculation of an $LD_{50}$. Second, this method cannot identify delayed deaths. If an animal, for example, dies 1 h after the second dosage, one has no way of determining whether it was actually the second dosage or a delayed effect of the first. For this reason it is of little value to observe the animals for any more than a few days after the last dosage. Third, if the test article has an unusually long half-life, bioaccumulation can lead to an underestimation of the acute lethal dosage. By contract, the pharmacological accommodation can lead to a spuriously high estimate of lethality. Depending on the importance of the finding, one may want to confirm that the results obtained at the highest dosage administered were dosing two naïve animals at the same dosage. Fortunately, the minimum 48-h period between dosing sessions will minimize such effects. Because of this design feature, it may take as long as three weeks to complete the dosing sequence. However, as there is generally no need for a one- to two-week postdosing observation or holding period, the actual study may not take significantly more time than a test of more traditional design.

Keep in mind that the objective of such studies is to gain information about lethality and gross tolerance. For nonrodents (especially monkeys), if none of the animals die or demonstrate obvious signs of toxicity, little would be gained by euthanizing and necropsying such animals. They can be saved and used again, following a reasonable "washout" period, to assess the lethality, toxicity, or safety pharmacology of a different chemical. In the hands of a skilled toxicologist, such adaptive reuse of animals is a cost-effective way to minimize overall usage.

***Limit Tests*** There are relatively innocuous drugs that are simply not potently lethal. The limit test (Figure 6) provides the simplest protocol for determining the lethality or such substances. The limit test is designed to obtain clearance at a specific dosage based on the assumption that what may occur at a higher dosage is not of practical relevance. Thus, one dosage only is studied. This limit "dosage" can be set on the basis of the chemical or physical properties of the test article (or vehicle) or on the basis of an upward safety margin. If the preparation is highly acidic (pH $< 3$), large intravenous dose would be expected to cause systemic acidosis as well as local irritation but will yield little relevant toxicology information, as such a preparation would never be approved for clinical use. Alternately, if the anticipated human dosage of a

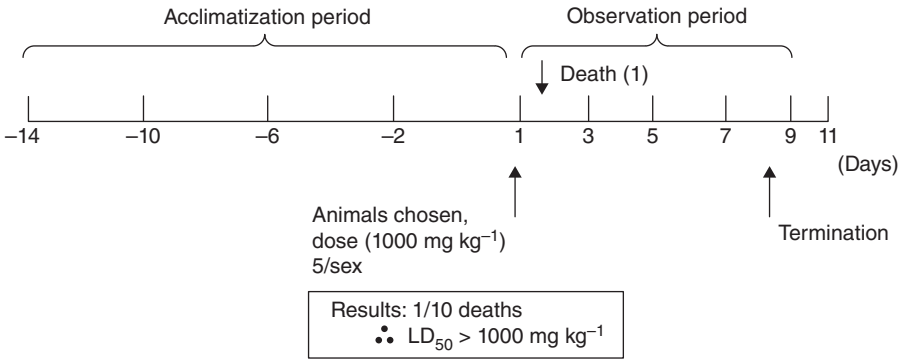


***Figure 6*** *Example of typical limit test protocol.*

drug is 0.3 mg kg$^{-1}$, there is probably little reason to test dosages in excess of 300 mg kg$^{-1}$ (1000 times the expected human dosage). In general, there is never any reason to use dosages of 5 g kg$^{-1}$ or greater and rarely any reason to exceed 3 g kg$^{-1}$.

There are three possible outcomes to a limit test. If none of the animals die, then the conclusion is that the MLD is greater than the limit dosage. If fewer than 50% of the animals die, then the conclusion is that the $LD_{50}$ is greater than the limit dosage. If more than 50% of the animals die, then one has a problem. Depending on the reasons for performing the test, one could reset the limit and repeat the study or one could assess lethality by a different protocol. Alternatively, the change in the limit could reflect a change in the chemical or biological properties of the test substance that should be evaluated further.

***Fixed-Dose Procedure*** The fixed-dose design (Figure 7) was proposed by the British Toxicology Society (1984). It is designed to supply the data needed for classification or labeling purposes. It is essentially a three-step limit test.

Five rats per sex are given 50 mg kg$^{-1}$. If survival is less than 90%, a second group of animals is given 5 mg kg$^{-1}$. If survival is again less than 90%, the substance is classified as "very toxic"; otherwise, it is classified as "toxic."

If, after the 50-mg kg$^{-1}$ dose, survival is 90% but there is evident toxicity, no further dosages are given and the substance is classified as "harmful." If, on the other hand, there is no evident toxicity at 50 mg kg$^{-1}$, another group of rats is given 500 mg kg$^{-1}$. If there is again 90% survival and no evident toxicity, the substance is given "unclassified" or "slightly toxic" status.

The fixed-dose procedure is relatively new and apparently results in a large decrease in animal usage. It is also noteworthy in that it utilizes not only lethality but also "evident toxicity," which, in all likelihood, refers to obvious signs of central nervous system (CNS) effect, such as seizures or prostration. Whether or not this protocol design becomes widely accepted by various regulatory agencies remains to be established.

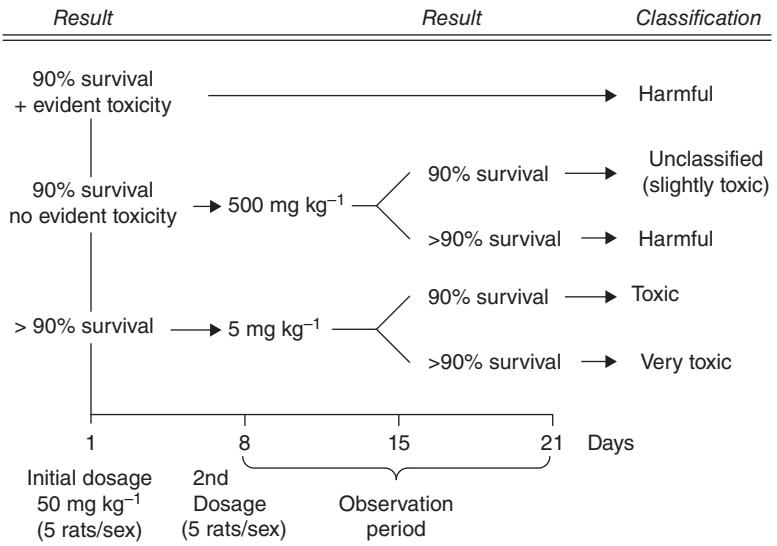


***Figure 7*** *British Toxicology Society fixed-dose procedure.*

The potential utility of the fixed-dose procedure was demonstrated in an international validation study in which the acute oral toxicity of 20 different chemicals was evaluated using both the fixed-dose and classical $LD_{50}$ procedures. Thirty-three laboratories in 11 different countries were involved in the validation project, and the results have been published (van den Heuvel et al., 1990). The results demonstrated that the fixed-dose procedure produced consistent evaluations of acute toxicity that were not subject to significant interlaboratory variation and provided sufficient information for hazard identification and risk assessment based on signs of toxicity (clinical signs, time to onset, duration, outcome, etc.). The fixed-dose procedure used fewer animals than the classical $LD_{50}$ tests and generally required less time to complete. Because of the emphasis on toxicity (rather than mortality) and the use of fewer animals, the fixed-dose procedure could be considered a more "humane" or animal-sparing design than the classical $LD_{50}$ test. When the results of the fixed-dose and $LD_{50}$ tests were compared for hazard-ranking purposes (Table 2), comparable results were obtained. Thus, it would appear that the fixed-dose procedure has utility and has been recommended late in 2000 for broad regulatory adaptation by ICVAM (Interagency Coordinating Committee on the Validation of Alternative Methods).

***"Rolling" Acute Test*** The rolling acute test is a combination protocol that is designed to find a tolerated dose to use for a subchronic toxicity test. The first segment can be either a dose probe or an up/down or pyramiding type of study to define the MLD. In the second segment, three to five animals are dosed for a short period of time—five to seven days. The objective of this

**TABLE 2 Comparison of Toxicity Classification Based on $LD_{50}$ versus Fixed-Dose Procedure**

| Test Chemical | Toxicity Classification Based on $LD_{50}$ | Fixed Dose: Number of Laboratories Classifying Chemical | | | |
|---|---|---|---|---|---|
| | | Very Toxic | Toxic | Harmful | Unclassified |
| Nicotine | Toxic | — | 23 | 3 | — |
| Sodium | Harmful | — | 1 | 25 | — |
| Ferrocene | Harmful/unclassified | — | — | 3 | |
| 2-Chloroethyl alcohol | Toxic | — | 19 | 7 | — |
| Sodium arsenite | Toxic | — | 25 | 1 | — |
| Phenyl mercury acetate | Toxic | 2 | 24 | — | — |
| *p*-Dichlorobenzene | Unclassified | — | — | — | 26 |
| Fentin hydroxide | Toxic | — | 8 | 17 | 1 |
| Acetanilide | Harmful | — | — | 4 | 22 |
| Quercetin dihydrate | Unclassified | — | — | — | 26 |
| Tetrachlorvinphos | Unclassified | — | — | 1 | 25 |
| Piperidine | Harmful | — | 2 | 24 | — |
| Mercuric chloride | Toxic | — | 25 | 1 | — |
| 1-Phenyl-2-thiourea | Toxic/harmful | 12 | 12 | 2 | — |
| 4-Aminophenol | harmful | — | — | 17 | 9 |
| Naphthalene | Unclassified | — | — | — | 26 |
| Acetonitrile | Harmful | — | — | 4 | 22 |
| Aldicarb (10%) | Very toxic | 22 | — | — | — |
| Resorcinol | Harmful | — | — | 25 | 1 |
| Dimethyl formamide | Unclassified | — | — | — | 26 |

*Source*: van der Heuvel et al., 1990.

design is to compensate for the fact that cumulative toxicity can occur at substantial differences in acute and subchronic toxic dosages. One can be easily misled by selecting subchronic dosages based entirely on acute lethality data. An example is a drug tested where it was found that 360 mg kg$^{-1}$ was acutely nonlethal and the MLD was 970 mg kg$^{-1}$. The dosages selected for the four-week subchronic study were 50, 100, 200, and 400 mg kg$^{-1}$ day$^{-1}$. The top-dose animals all died within a week. Substantial mortality occurred at 200 mg kg$^{-1}$ and evident toxicity was present at 50 mg kg$^{-1}$. A no-effect dosage was not identified, so the entire test had to be repeated with a different dosage structure. The rolling acute structure is a quick and relatively simple "sanity" check that permits one to avoid making such mistakes.

## 2.2 Using Range-Finding Lethality Data in Drug Development: Minimum Lethal Dosage

Range-finding data are often used early in drug development to make preliminary safety estimates. The $LD_{50}$ is simply a calculated point on a curve. The

shape or slope of this curve is also an important characteristic of the test substance. However, unless one does a great deal of acute toxicity testing, the difference between a slope of 1.5 and a slope of 4 has very little meaning. Further, for safety considerations, the dosage that kills 50% of the animals is not as important as the dosage at which lethality first becomes apparent (i.e., the threshold dosage or MLD). For example, if the oral $LD_{50}$s of two different drugs (A and B) were 0.6 and 2.0 g kg$^{-1}$, respectively, what would we conclude about the relative safety of these compounds? Further, let us assume that the estimated human dosage of drug A is 0.5 mg kg$^{-1}$ and of drug B is 5 mg kg$^{-1}$. Do our conclusions concerning the relative safety of these two drugs change? In fact, the $LD_{50}$s of both drugs are so high that both are considered only slightly toxic (0.5–5.0 g kg$^{-1}$). One can also compute the lethality safety margin or index (LSI, equal to $LD_{50}$/EHD, where EHD is the estimated human dose) for these two drugs; both indices are so large (1200 for A and 400 for B) that there is still no toxicologically relevant difference between the two drugs. Let us now assume that the lethality curve for substance A is very steep, such that 0.4 g kg$^{-1}$ causes death in a very small percentage of animals—it is, in fact, the lowest dose administered that causes death. This is the MLD or estimated MLD (EMLD). Let us now assume that the lethality curve for B is very shallow, such that its MLD is also 0.4 g kg$^{-1}$. Does this change our safety considerations of these two drugs? One can calculate a new more conservative safety index (MLD/EHD) of 800 for A and 80 for B. As a very general rule of thumb, an index for lethality of less than 100 is cause for mild concern, one less than 10 is cause for caution, and one less than 1 should be cause for extreme caution. In the case of our two hypothetical drugs, the development of drug B should be approached with more caution than that of drug A, despite the fact that B has a higher $LD_{50}$. This is demonstrated in Figure 8. There are drugs sold over the counter, however, that have lethality safety indices of less than 10. For example, the MLD of indomethacin in rats is 3.7 mg kg$^{-1}$ (from data reported by Schiantarelli and Cadel, 1981), while the maximum recommended human dose is 200 mg (2.9 mg kg$^{-1}$ for a 70-kg person); hence, indomethacin has an LSI of 1.3. Such a finding is only cause for some caution but does not in and of itself justify restricting the use or sale of a drug. Hence, because it results in a more conservative safety factor and also takes into consideration the slope of the lethality curve, the use of the MLD rather than the $LD_{50}$ is recommended in calculating acute safety indices.

A number of different safety factors and therapeutic indices have been proposed in the literature. Despite their similarity, some distinction should be made between these two. A therapeutic index applies only to drugs and is the ratio between a toxic dosage (TD or LD: the toxic endpoint does not always have to be death) and the pharmacologically effective dosage (ED) in the same species. A safety index can be calculated for all xenobiotics, not just drugs. A safety index is the ratio of likely human exposure (or dosage) and the dosage that causes death or other forms of toxicity in the most sensitive experimental animal species. The most conservative (lethality) safety index (LSI) is obtained by dividing the maximum estimated human dosage or

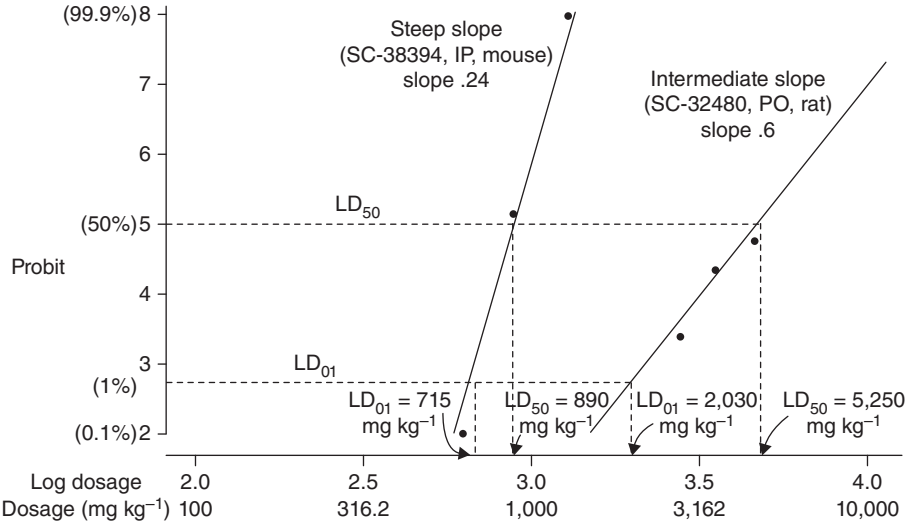


***Figure 8*** *Examples of probit-log dosage–response curves illustrating differences in slope curves and relationship between slope, $LD_{50}$, and $LD_{01}$.*

exposure by either the minimum lethal dosage or the maximum nonlethal dosage.

***Minimum Lethal Dosage Protocols*** Stating that the MLD is preferable to the $LD_{50}$ for safety considerations is one thing; trying to determine what a specific MLD may be or could be is another. There are no commonly used experimental designs that have the MLD as an endpoint. Assuming a log dose response, the MLD may become a function of group size. Theoretically, if enough animals are dosed, at least one animal could die at any reasonable dosage. There are, however, practical considerations that can and should be applied to determining an MLD. As a practical rule of thumb, we recommend that the estimated $LD_{01}$—the dose that would be expected to kill 1% of the experimental animals exposed—be used as an estimate of the MLD. If one already has sufficient data to describe a lethality curve, an $LD_{01}$ can be calculated as easily as the $LD_{50}$. This is often the case with acute toxicity data obtained to support regulatory submission.

How is the MLD calculated without a complete lethality curve? A modified pyramiding dosage design may be the most appropriate approach. With this design, groups of animals are treated with stepwise increases in dosage until death occurs or a limit dosage is attained. If one has no idea as to what the initial dosage should be or how to graduate the dosages, a dose-probing experiment can be conducted. If the dose-probing experiment produces no deaths, two to three more animals can be dosed at the limit dose to confirm the results; the lethality determination is now complete. If the probe experiment does

produce death, then the additional dosages can be graduated between the lowest lethal and the highest nonlethal dosages. A typical progression may proceed as follows (Figure 9): On day 1 of the study, three probe animals are dosed at 10, 100, and 1000 mg kg$^{-1}$. The animal at 100 mg kg$^{-1}$ dies within a few hours of dosing. The two remaining animals are dosed at 300 mg kg$^{-1}$ on day 3. Neither dies. They are then dosed at 500 mg kg$^{-1}$ on day 5. One dies. Three additional animals should be dosed on day 7 or 8 at a dosage in between (i.e., 400 mg kg$^{-1}$ is a good estimate of the maximum nonlethal dosage, or MNLD). While different by definition, there is usually not a great deal of distance between the MLD and the MNLD, as this example illustrates. In fact, even for a well-characterized lethality curve, the confidence limits for the $LD_{01}$ will be quite broad and encompass both the MLD and MNLD.

Malmfors and Teiling (1983) have proposed a similar method for determining what they also termed the MNLD. Rather than initiating the study with probe animals, their design calls for three consecutive pyramiding-type studies with the steps becoming increasingly smaller. For example, two animals will be sequentially dosed at 2, 200, and 2000 mg kg$^{-1}$. If death occurs at 2000 mg kg$^{-1}$, a new pair of animals is initiated at 200 mg kg$^{-1}$, and sequential dosages are increased by a factor of 1.8 until death occurs. Then another pair of animals is initiated at the highest nonlethal dosage, and successive dosages are increased by a factor of 1.15. The result of this exercise will be two dosages, one apparently nonlethal and the other lethal. Six animals are dosed at each dosage. If none die at the lower dosage and one dies at the higher dose, then the lower dose is considered to be the MNLD. At least 24 h between dosing rounds are recommended. While this method may have some utility, there are some disadvantages. First, the recommended limiting dosage of 5 g kg$^{-1}$ is too high. Second, 24 h between doses may be too short a period to allow for recovery. Third, even with only 24 h between doses, this is a time-consuming procedure—it may take up to two weeks to complete the dosing. Finally, it does not decrease the number of animals needed, since it may use 18–20 animals.

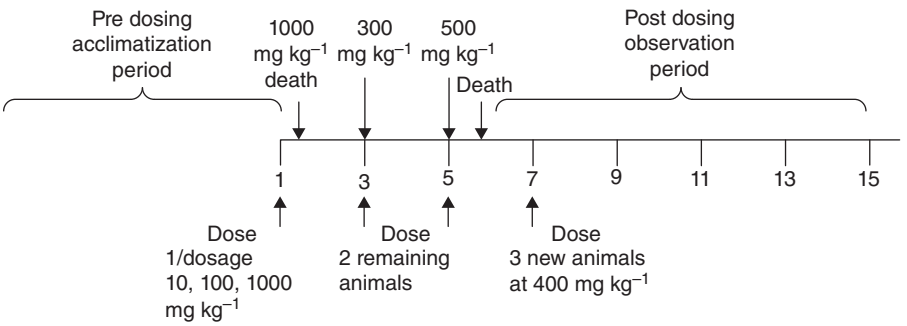


***Figure 9*** *Example of MLD pyramiding dose design.*

Dose probing is not generally used for nonrodents (rather, the pyramiding dose scheme is) and the initiating dosage is normally in the range of 1–5 times the projected human clinical dosage. The limit is generally in the area of 1 g $kg^{-1}$ or 100–200 times the human dosage, whichever is less. The normal study will include two animals of each sex treated with the test article. For simple lethality, there is seldom any need to include control animals. If the projected human dosage is 4 mg $kg^{-1}$, for example, the initial dosage in an MLD range finder in dogs will be 20 mg $kg^{-1}$ and succeeding dosages will increase stepwise at half-log intervals; thus, 20-, 60-, 200-, and 600-mg $kg^{-1}$ doses are separated by at least 48 h. The MLD is simply reported as being between the highest observable nonlethal and the lowest lethal dosages, or at greater than the limit dosage—in this case, 600 mg $kg^{-1}$. Studies should not be done with nonrodents solely for determining lethality, because this would not be an appropriate use of time and animals. Generally, these studies should also include some combination of extensive physical examinations, such as electrocardiograms (ECGs) and rectal temperatures, careful observations of behavior and activity, and extensive clinical laboratory workups after each dose.

The pyramiding dose study is not without disadvantages. The small number of animals used can cause simple random variation resulting in misestimation of lethality. It is a well-accepted statistical maxim that the smaller the sample size, the greater the impact of any random variation (error or outlier) on the population characteristic. This may be especially true for a nonrodent species where experimental animals are drawn from an outbred population. Second, the pyramiding dose regimen can permit the development of tolerance. For example, pyramiding dosage studies were conducted to range find dosages for a two-week study on 1,4-benzodiazepine. Lethality in dogs was observed at 600 mg $kg^{-1}$ in the pyramiding study. For the subsequent subchronic study, the top dose was set at 300 mg $kg^{-1}$; both dogs died of CNS depression of the first day of dosing.

## 3 ACUTE SYSTEMIC TOXICITY CHARACTERIZATION

Acute systemic toxicity studies are performed to more completely define the acute toxicity of a drug. They are more extensive and time consuming than range-finding tests or screens and are normally the type of study done to satisfy regulatory requirements or to provide a more thorough early characterization or prediction of toxicity (McClain, 1983). In pharmaceutical development, rarely would an acute test be sufficient to support registration, but it could support exploratory INDs first in a human single-dose study (FDA, 2006) or a single human dose study of an imaging agent and it may be required as part of an overall package. These protocols may resemble range-finding tests, but they call for collection of more data. A list of the types of data that can be obtained in well-conducted acute toxicity tests is given in Table 3. Given that these studies usually include control groups, the classical or traditional

**TABLE 3 Information, Including Lethality, That Can be Gained in Acute Toxicity Testing**

| |
|---|
| Lethality/mortality |
| $LD_{50}$ with confidence limits |
| Shape and slope of lethality curves |
| Estimation of maximum nonlethal dose or minimum lethal dose ($LD_{01}$) |
| Time to dose estimates |
| Clinical signs |
| Times of onset and recovery |
| Thresholds |
| Agonal vs. nonagonal (i.e., do signs occur only in animals that die?) |
| Specific vs. general responses |
| Separation of dose–response curves from lethality curves |
| Body weight changes |
| Actual loss vs. decreased gain |
| Recovery |
| Accompanied by changes in feed consumption |
| Changes in animals that die vs. those that survive |
| Target organ identification |
| Gross examinations |
| Histological examinations |
| Clinical chemical changes |
| Hematological changes |
| Specialized function tests |
| Immunocompetency |
| Neuromuscular screening |
| Behavioral screening |
| Pharmacokinetic considerations |
| Different routes of administration yielding differences in toxicity |
| Plasma levels of test article |
| Areas under the curves, volume of distribution, half-life |
| Metabolic pattern of test article |
| Distribution to key organs |
| Relationship between plasma levels and occurrence of clinical signs |

design is the most common because it allows for the most straightforward statistical analyses. In addition, while the use of staggered dosing days for different groups is still a fairly common practice, data analyses may be more sensitive if all animals are dosed on the same day, requiring that one have preliminary range finder data that permit selection of appropriate dosages. Studies of more than one species and/or more than one route should be limited to those instances where they are required by statute.

In general, traditionally designed acute toxicity tests can be divided into three types that can be called the minimal acute toxicity test, the complete acute toxicity test, and the supplemented acute toxicity test. Of these, the minimal protocol is by far the most common and is discussed first. The other two represent increasing orders of complexity as additional parameters of measurement are added to the basic minimal study.

### 3.1 Minimal Acute Toxicity Test

An example of a typical minimal acute toxicity test protocol is shown in Figure 10. This study resembles a traditional lethality test in terms of the number of groups and the number of animals per group. Standard protocols consist of three or four groups of treated animals and one group of control animals, each group consisting of five animals per sex per dosage (OECD, 1991). Traditionally, the emphasis in these types of studies was on determining the $LD_{50}$, time to death, slope of the lethality curve, and prominent clinical signs, as illustrated by the data reported by Jenner et al. (1964). More recent designs specify, in addition to lethality and clinical observations, that body weights be recorded during the study and gross necropsies performed at the end of the postdosing observation period. For an excellent example of a well-performed acute toxicity evaluation the reader is referred to the paper by Peterson et al. (1987) of the acute toxicity of the alkaloids of *Lupinus angustifolius*, in which the $LD_{50}$s, time to death, clinical signs, body weight effect, and gross necropsy findings were all discussed. For pharmaceuticals, where acute toxicity data for more than one species are often required, these studies will be done as batteries on both rats and mice. In addition, because many drugs will be given by more than one route to human patients, these batteries will include groups treated by two different routes. Thus, an acute study on a pharmaceutical agent will often result in eight "curves"—one per route per species per sex. For tests on nonrodent species, as required for pharmaceuticals, a different design is used (discussed later).

The animals should be acclimated to laboratory conditions for 7–14 days prior to dosing. For acute toxicity tests, this pretreatment period should be more than just a holding period. Animals should be checked daily for signs of ill health and/or abnormal behavior. Body weights may also be determined. These data should be used to exclude abnormal animals from the test. Such data also provide an additional basis for interpreting the data gathered during

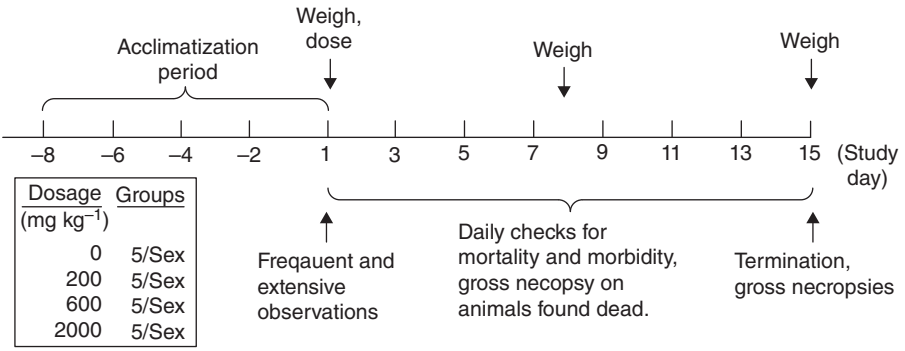


***Figure 10*** *Example of minimal acute toxicity protocol.*

the postdosing period. Finally, these activities acclimate the animals to the frequent handling that is a necessary part of an acute toxicity test.

In selecting dosages for an acute systemic toxicity study, the same general guidelines apply as with lethality testing:

1. There is little to be gained from testing dosages that are so high that the physical rather than biological properties become prominent. Generally, little additional information is gained by pushing dosages past $2\,g\,kg^{-1}$. The usual regulatory limit for pharmaceuticals is now $1.5\,g\,kg^{-1}$.
2. The highest dosage should be no larger than 100–300 times the anticipated human dosage.
3. Widely spaced dosages are better than narrowly spaced dosages.

This latter point is particularly true in an acute toxicity test on a drug, because pharmacologically based clinical signs may occur at dosages considerably lower than those that cause death. Also, as discussed by Sperling (1976) and Gad (1982), the effects at high dosages may mask the effects that would be observed at low dosages. As human beings are more likely to be exposed to lower dosages than experimental animals, these low-dosage effects may be important parameters to define.

Historically, it has been stated in various regulatory communications that a well-conducted acute toxicity test should contain sufficient data to calculate an $LD_{50}$. This is no longer necessarily the case. Simpler, less resource-intensive range-finding protocols should be used for defining lethality. Because it is rare that an extensive acute protocol would be attempted without preliminary lethality data, the lethality objectives of acute systemic testing are not always critical. Ideally, the highest dosage should elicit marked toxicity (such as lethality), but it does not need to kill all of the animals to satisfy one's need to show due diligence in stressing the test system. If one already has sufficient preliminary data to suspect that the top dosage will be nonlethal or otherwise innocuous, the test can be conducted as a limit test, consisting of one treated group and one control group.

***Clinical Signs*** The nonlethal parameters of acute toxicity testing have been extensively reviewed by Sperling (1976) and Balazs (1970, 1976). Clinical observations or signs of toxicity are perhaps the most important aspect of a minimal acute toxicity test because they are the first indicators of drug- or chemical-related toxicity or morbidity, and they are necessary in the interpretation of other data collected. For example, body weight loss (or a reduction in body weight gain) would be expected if an animal had profound CNS depression lasting several hours.

With regard to clinical signs and observations, there are some basic definitions that should be kept in mind. Symptomatology is the overall manifestation of toxicity. Signs are overt and observable events (Brown, 1983). Symptoms

are the subjective impressions of a human patient (e.g., headache) and cannot be described or reported by speechless animals (Balazs, 1970). Clinical signs can be reversible or irreversible. Reversible signs are those that dissipate as the chemical is cleared from the body or tolerance develops (Chan et al., 1982) and are generally not accompanied by permanent organ damage. Irreversible signs are those that do not dissipate and are generally accompanied by organ or tissue damage. Signs can also represent a normal biological or pharmacological response (Chan et al., 1982). For example, an antidepressant would be expected to cause decreased activity and some ataxia. These symptoms are generally reversible and can lead to secondary, nonspecific signs—nonspecific in that any number of agents or stimuli can evoke the same response and secondary in that they are probably not due (at least, one has no evidence to determine otherwise) to the direct action of the test article. Responses can also be abnormal in that they are not due to a homeostatic process. The increases in serum urea and creatinine due to kidney damage, for example, are abnormal responses. These are often irreversible, but this is not always the case, depending on the repair capacity or functional reserves of the target organ. These abnormal responses may also be called primary effects because they reflect the direct action of a test article. Agonal signs are those occurring immediately prior to or concomitantly with death. They are obviously irreversible, but not necessarily reflective of a specific effect of a test article. For example, regardless of the cause, labored breathing will occur in a moribund animal. It is therefore important to distinguish between signs that occur in animals that die and those that do not. It should also be kept in mind that agonal signs may mask (make it difficult or impossible) to observe other signs, including those clearly seen at lower doses.

In their simplest form, clinical observations are those done on an animal in its cage or, preferably, in an open plane, such as on the top of a counter or laboratory cart. These are considered passive observations. One can gain even more information by active examination of the animal, such as the animal's response to stimulation. Fowler and Rutty (1983) divide their clinical evaluation of toxicity into those signs scored by simple observations (e.g., ataxia), those scored by provocation (e.g., righting reflex), those scored in the hand (e.g., mydriasis), and those scored by monitoring (e.g., rectal temperature). Cage pans should always be examined for unusually large or small amounts of excreta or excreta of abnormal color or consistency. A list of typical observations is summarized in Table 4. A more extensive table has been prepared by Chan et al. (1982). Given the fact that the number of different signs displayed is not infinite and that some signs are simply easier to discern than others, most clinical signs are referable to the CNS (e.g., lack of activity), the gastrointestinal (GI) tract (e.g., diarrhea), or the general autonomic nervous system (e.g., increased salivation or lacrimation). This is illustrated by an actual example set of data from acute toxicity studies summarized in Table 5.

Other signs can be detected by a well-trained observer but are nonetheless less common than those described above. Respiratory distress can be diag-

**TABLE 4 Clinical Observations in Acute Toxicity Tests**

| Organ System | Observation and Examination | Common Signs of Toxicity |
|---|---|---|
| CNS and somatomotor | Behavior | Unusual aggressiveness, unusual vocalization, restlessness, sedation |
| | Movements | Twitch, tremor, ataxia, catatonia, paralysis, convulsion |
| | Reactivity to various stimuli | Irritability, passivity, anesthesia, hyperesthesia |
| | Cerebral and spinal reflexes | Sluggishness, absence of reflex |
| | Muscle tone | Rigidity, flaccidity |
| Autonomic nervous system | Pupil size | Miosis, mydriasis |
| Respiratory | Nostrils | Discharge (color vs. uncolored) |
| | Character and rate | Bradypnea, dyspnea, Cheyne–Stokes breathing, Kussmaul breathing |
| Cardiovascular | Palpation of cardiac region | Thrill, bradycardia, arrhythmia, stronger or weaker beat |
| Gastrointestinal | Events | Diarrhea, constipation |
| | Abdominal shape | Flatulence, contraction |
| | Feces consistency and color | Unformed, black or clay colored |
| Genitourinary | Vulva, mammary glands | Swelling |
| | Penis | Prolapse |
| | Perineal region | Soiled |
| Skin and fur | Color, turgor, integrity | Reddening, flaccid skinfold, eruptions, piloerection |
| Mucous membranes | Conjunctiva, mouth | Discharge, congestion, hemorrhage, cyanosis, jaundice |
| Eye | Eyelids | Ptosis |
| | Eyeball | Exophthalmos, nystagmus |
| | Transparency | Opacities |
| Others | Rectal or paw skin temperature | Subnormal, increased |
| | Injection site | Swelling |
| | General condition | Abnormal posture, emaciation |

*Source*: Balazs, 1970.

nosed by examining the animal's breathing motions and listening for breathing noises. Cardiovascular signs are generally limited to pallor, cyanosis, and/or hypothermia. Changes in cardiac function can be difficult to detect in small animals and generally consist of "weak" or "slow" breathing. Arrhythmias can be difficult to detect because the normal heart rate in a rodent is quite rapid. ECGs are difficult to record from rodents on a routine basis. Therefore, the assessment of potential acute cardiovascular effect of a drug or chemical is usually restricted to a nonrodent species, usually the dog.

Given the subjective nature of recognizing clinical signs, careful steps must be taken to ensure uniformity (is the animal depressed or prostrated?) of

**TABLE 5 Summary of Clinical Observations from Actual Acute Toxicity Tests**

| Drug (Route) | Indication | Acute Clinical Signs[a] |
|---|---|---|
| SC-37407 (PO) | Analgesic (opiate) | Reduced motor activity, mydriasis, reduced fecal output, hunched posture, convulsions (tonic), ataxia |
| SC-35135 (PO) | Arrhythmias | Reduced motor activity, lost righting reflex, tremors, dyspnea, ataxia, mydriasis |
| SC-32840 (PO) | Intravascular thrombosis | Reduced motor activity, ataxia, lost righting reflex, closed eyes, red/clear tears |
| SC-31828 (PO) | Arrhythmias | Reduced activity, dyspnea, ataxia, lost righting reflex, red/clear tears |
| SC-25469 (PO) | Analgesic (nonopiate) | Reduced motor activity, ataxia, lost righting reflex, dyspnea, convulsions (clonic) |

[a]The five or six most frequent signs in descending order of occurrence.

observation so that the data can be analyzed in a meaningful fashion. There are three ways of achieving this. First, signs should be restricted to a predefined list of simple descriptive terms, such as those listed in Table 4 or in Appendix B. Second, if a computerized data acquisition system is unavailable, the use of standardized forms will add uniformity to the observation and recording processes. An example of such a form is shown in Figure 11. Third, technicians should be trained in studies (not intended for regulatory submission) using material of known toxicity, so that all personnel involved in such evaluations are using the same terminology to describe the same signs.

Animals should be observed continuously for several hours following dosing. Times of observation should be recorded as well as the actual observations. After the first day of the study, observations generally need only to consist of brief checks for sign remission and the development of new signs of morbidity. Data should be collected in such a way that the following could be concluded for each sign: (1) estimated times of onset and recovery, (2) the range of threshold dosages, and (3) whether signs are directly related (primary) to the test article. An example of clinical signs provoked by a specific drug is given in Table 6. Incidences are broken down by dosage group and sex. These data illustrate the fact that mortality can censor (preclude) the occurrence of clinical signs. Note that reduced fecal output was a more frequent observation at the intermediate dosages because most of the animals died at the higher dosages.

Therapeutic ratios are traditionally calculated using the dose of the lowest observed adverse effect. A more sensitive therapeutic ratio could be calculated using the $ED_{50}$ (effective dose) for the most prominent clinical sign. However, while it may be possible to describe a dosage–response curve (which may, in fact, have a different slope than the lethality curve) for a clinical sign and calculate the $ED_{50}$, in practice this is rarely done. It is more common for the approximate threshold dosages or no observable effect levels (NOELs) to be reported. A typical minimal acute toxicity study can be summarized as shown in Table 7.

Acute observation record

(Days, other than Study Day 1, on which no signs are observed are recorded on the Log of Animal Observations)

Species ____ Sex ____ Route ____ Dose level ____ Animals coded* ____ Date dosed ____

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Study day | | | | | | | | | | | |
| Observations: Time | | | | | | | | | | | |
| Date | | | | | | | | | | | |
| No signs observed | | | | | | | | | | | |
| Reduced motor activity | | | | | | | | | | | |
| Ataxia | | | | | | | | | | | |
| Lost righting reflex | | | | | | | | | | | |
| Convulsions ( ) | | | | | | | | | | | |
| Mydriasia | | | | | | | | | | | |
| | | | | | | | | | | | |
| | | | | | | | | | | | |
| | | | | | | | | | | | |
| | | | | | | | | | | | |
| | | | | | | | | | | | |
| | | | | | | | | | | | |
| | | | | | | | | | | | |
| Death | | | | | | | | | | | |
| Observer | | | | | | | | | | | |

Page of notes:

*An. Code ____ *An. ID ____

*Animal code for recording observations

Read and understood

____ Date

***Figure 11*** *Example of form for recording clinical observations in acute systemic toxicity studies.*

**TABLE 6 Example of Clinical Observations Broken Down by Dosage Group and Sex in Acute Toxicity Study of Drug SC-37407[a]**

| | Dose Levels (mg kg$^{-1}$) by Sex | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | 0 | | 50 | | 160 | | 500 | | 1600 | |
| Signs Observed | M | F | M | F | M | F | M | F | M | F |
| Reduced motor activity | — | — | — | — | — | — | 5/5 | 5/5 | 4/5 | 4/5 |
| Mydriasis | — | — | — | — | 3/5 | 4/5 | 4/5 | 5/5 | 5/5 | 5/5 |
| Reduced fecal output | — | — | 5/5 | 5/5 | 3/5 | 5/5 | — | 1/5 | — | — |
| Hunched posture | — | — | — | — | — | 1/5 | 3/5 | 3/5 | — | — |
| Convulsions (tonic) | — | — | — | — | — | — | 5/5 | 1/5 | 5/5 | 3/5 |
| Ataxia | — | — | — | — | — | — | 5/5 | 4/5 | 2/5 | 1/5 |
| Tremors | — | — | — | — | — | — | 1/5 | 2/5 | 1/5 | — |
| Death | 0/5 | 0/5 | 0/5 | 0/5 | 0/5 | 0/5 | 5/5 | 4/5 | 5/5 | 5/5 |

[a]Signs observed in rats treated orally (no. exhibiting sign within 14 days after treatment/no. treated). A dash indicates the sign was not observed at that dose level.

**TABLE 7 Minimal Acute Toxicity Study Summary of Drug SC-34871**

| Species (Route) | Dose ($mg\,kg^{-1}$) | Dead/ Dosed | $LD_{50}$ ($mg\,kg^{-1}$) | Signs Observed | Treatment to Death Intervals |
|---|---|---|---|---|---|
| Rat (PO) | 2400 | 0/10 | >2400[a] | None | None |
| Rat (IV) | 16 | 0/10 | Approximately 67 | Reduced motor activity at $50\,mg\,kg^{-1}$; convulsions, dyspnea, lost righting reflex at $160\,mg\,kg^{-1}$ | 0–2 h |
| | 50 | 2/10 | | | |
| | 160 | 10/10 | | | |
| Mouse (PO) | 500 | 0/10 | >2400 | None | None |
| | 1600 | 0/10 | | | |
| | 2400 | 0/10 | | | |
| Mouse (IV) | 50 | 1/10 | 120 (75–200)[b] | Reduced motor activity, ataxia at $160\,mg\,kg^{-1}$; tremors, convulsions, dyspnea at $500\,mg\,kg^{-1}$ | 0–2 h |
| | 160 | 6/10 | | | |
| | 500 | 10/10 | | | |

[a]Limit dosage.
[b]Fiducial limits.

## 3.2 Complete Acute Toxicity Testing

An example of the next-level test, the complete acute toxicity test, is given in Figure 12. As stated by Dayan (1983), the value of doing more than the minimal test will depend on the nature of subsequent testing. The complete protocol is designed to provide for a more in-depth search for target organs than the minimal protocol. This type of study, which has been well described by Gad and co-workers (1984), is similar in design to a minimal acute toxicity study but includes feed consumption data, more frequent body weight determinations, and more detailed and frequent clinical sign assessment. Groups should consist of at least 10 animals per group; 5 per sex per dosage should then be sacrificed 24–48 h for more immediate examination of any pathological changes induced by the test article. The remaining animals will be sacrificed at the end of the two-week period and examined for pathological changes. Blood will be collected at both sacrifices for clinical chemistry and/or hematology determinations. It should be noted that this design bears a striking resemblance to the design specified for an "expanded acute" study as required under the exploratory IND guidance (FDA, 2006), as shown in Figure 13.

***Body Weight Considerations*** Body weight and feed consumption are frequently determined parameters in toxicity testing. To an extent, the ability of

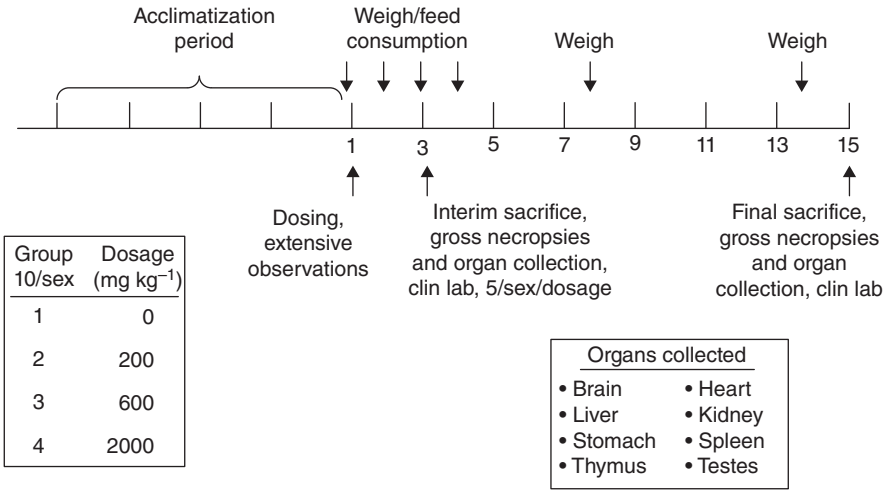


***Figure 12*** *Example of complete acute toxicity protocol.*

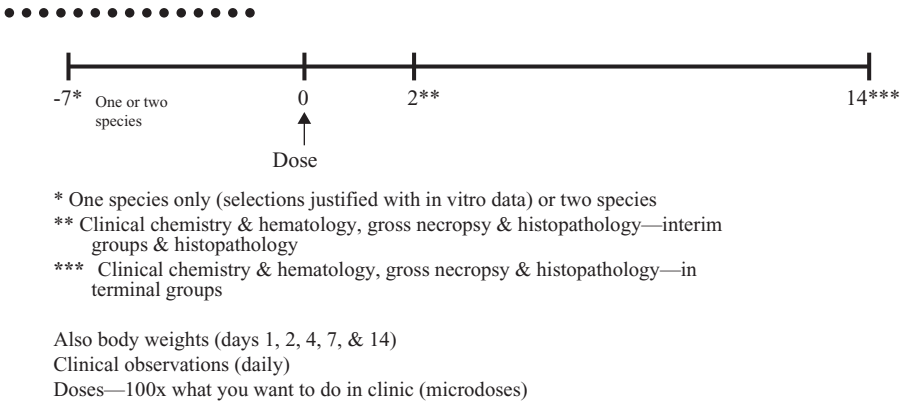


***Figure 13*** *Exploratory IND enabling acute study. From Gad et al. (1984).*

an animal to gain or maintain weight may be considered a sensitive but nonspecific indicator of health. While this is true in subchronic or chronic studies, its relevance in acute studies must be carefully considered. In most protocols, body weights are determined on day 1 (prior to dosing), day 7, and day 14, which are the days mandated by most regulatory guidelines. Despite being common, the design is not well founded: If an animal has not died within seven days postdosing, it has probably recovered and its body weight may not be noticeably different from controls by day 14. A complete protocol addresses this problem by specifying more frequent body weight determinations (daily for the first three to five days of the observation period) so that not only can

initial decreases (if they occur) be detected, but recovery can also be charted. Feed consumption measurements should be made at the same times, because it is difficult to determine the causes behind body weight changes in the absence of feed consumption data. Body weight loss accompanied by normal feed consumption implies something very different than body weight loss (or lack of gain) accompanied by lack of feed consumption. In the absence of feed consumption data, however, changes in body weight should still be considered indicative of a change in an animal's health status.

Yet another reason why body weight determinations are of questionable value in acute studies has to do with the statistical analysis of the data. Deaths may substantially alter group size and complicate analysis. The death of two of five animals causes a 40% decrease in group size and a substantial diminution of the power of any statistical test. In addition, the resulting data sets are censored: Comparisons will often be between the control group, a dosage group where all the animals survive, and a high-dosage group where less than 50% of the animals survive to the end of the observation period. One has to question the utility of body weight changes if they occur at dosages that are acutely lethal. The data in Table 8 illustrate this point. Body weight changes tended to occur only at dosages that were acutely lethal. Additionally, one would suspect that the censoring of body weights in groups where death occurs is not random; that is, the animals that die are most likely those that are most sensitive, while those that survive are the most resistant or robust. This problem can be addressed by building exclusionary criteria into a protocol. For example, one could statistically analyze body weight data in groups that only had less than 50% mortality.

Minimal rather than complete protocols tend to be more common in the acute testing of pharmaceutical agents. Drugs will almost always be subjected to at least one subchronic study. Body weight and feed consumption determinations are a standard feature of such studies. Additionally, changes in body weight and feed consumption are more likely in a subchronic than an acute study because the animals are dosed continuously between body weight determinations.

***Pathology Considerations*** One of the objectives of any well-conducted toxicity study is to identify target organs. There is some question, however, concerning the utility of extensive pathological assessments as part of an acute study. Gross necropsies are generally the minimum assessments requested by most regulatory bodies. Hence, minimal protocols will include necropsies on all animals found dead and those sacrificed following the postdosing observation period. An example of necropsy findings is given in Table 9. This table illustrates that gross necropsy observations on acute studies rarely predict the toxicity that will be seen when the chemical is given for longer periods of time. This is not surprising, because most drug-related histological lesions are the result of chronicity; that is, discernible lesions tend to result from the cumulative effect of dosages that are acutely well tolerated.

**TABLE 8 Examples of Body Weight Changes in Rats from Minimal Acute Toxicity Studies**

| Drug (Route) | Dosage (mg kg$^{-1}$) | BWT Change (g)[a] | Mortality |
|---|---|---|---|
| SC-32561 | | | |
| PO | 0 | 45 ± 4 | 0/10 |
| | 5000 | 39 ± 10 | 0/10 |
| IP | 0 | 43 ± 4 | 0/10 |
| | 500 | 43 ± 9 | 0/10 |
| | 890 | 44 ± 11 | 0/10 |
| | 1600 | 6 ± 14[b] | 2/10 |
| | 2800 | 24 ± 20[b] | 3/10 |
| SC-36250 | | | |
| PO | 0 | 38 ± 10 | 0/10 |
| | 5000 | 34 ± 10 | 0/10 |
| *IP* | 0 | 34 ± 6 | 0/10 |
| | 670 | 50 ± 8[b] | 2/10 |
| | 890 | 46 ± 8[b] | 3/10 |
| | 1200 | 45 ± 4 | 4/10 |
| | 1400 | 35[c] | 9/10 |
| SC-36602 | | | |
| IV | 0 | 38 ± 9 | 0/10 |
| | 58 | 38 ± 3 | 0/10 |
| | 67 | 36 ± 7 | 2/10 |
| | 77 | 49 ± 5[b] | 3/10 |
| | 89 | 41 ± 7 | 7/10 |
| PO | 0 | 38 ± 5 | 0/10 |
| | 2100 | 41 ± 5 | 3/10 |
| | 2800 | 38 ± 5 | 7/10 |
| | 3700 | 26 ± 6 | 7/10 |

[a]Mean ± standard deviation. Body weight (BWT) changes in grams for each group during the first week of the postdosing observation period.
[b]Statistically different from control (0 dosage group), $p \leq 0.05$.
[c]Only one animal survived, so there is no standard deviation.

The data in Table 9 also demonstrate that substantial gross macroscopic findings are rare in minimal acute studies and seldom suggestive of a specific effect. There are several reasons for the lack of specificity. The first is the rather limited nature of gross observations, in that they are limited to broad descriptive terms (size, shape, color, etc.). Second, for animals found dead, it is difficult to separate the chemically associated effects from agonal and/or autolytic changes. Finally, it is difficult to come to a conclusion about the nature of a gross lesion without histological assessment.

If there are any identifiable gross lesions, they often differ between animals that die and those that survive to the end of the observation period. The reason for these differences is very simple. An animal that dies less than 24 h after chemical exposure probably has not had sufficient time to develop a well-defined lesion. As mentioned earlier, most deaths occur within 24 h. Animals that survive for the two-week observation period have probably totally recov-

**TABLE 9 Examples of Gross Necropsy Findings from Acute Toxicity Studies**

| Drug | Acute Gross Pathology | Subchronic Target Organs[a] |
|---|---|---|
| SC-36602 | Distended stomach and intestine, bloody fluid in intestine, congested lung, pale liver | None |
| SC-38394 | None | Liver, testes, bone marrow, thymus, kidney |
| SC-32840 | None | Heart, stomach, kidney, bladder |
| SC-25469 | Peritonitis (IP route only) | None |
| SC-36250 | Peritonitis (IP route only) | Adrenal, liver, thyroid |
| SC-27166 | None | Liver |

[a]Organs that showed any evidence of test-article-related changes in repeated-dose studies of two weeks or longer duration.

ered and rarely have apparent lesions. Hence, the animals that provide the best chance to identify test-article-specific lesions are those that die in the region of 24–96 h postdosing. This is, in fact, one of the problems with acute pathology data—that is, comparing animals found dead with those sacrificed at a different time and comparing both to controls. As mentioned, a complete protocol, where groups of animals are sacrificed 24 to 96 h after dosing, at least partially solves this problem.

Many guidelines suggest microscopic confirmation of gross lesions "when necessary"; however, these are seldom done because of the autolytic nature of many of the tissues collected from animals found dead. Additionally, the practice of collecting and examining only gross lesions is difficult to justify because it does not permit in-depth comparisons. Pathological findings are most easily interpreted when the same organs are collected and examined from all animals on a test regardless of the circumstances of death. Elsberry (1986) recommends that the GI tract, kidney, heart, brain, liver, and spleen be specifically examined routinely in acute studies. Given the timing issues discussed in the previous paragraph, the amount of effort may not be worth the result. In an attempt to address these problems, Gad and co-workers (1984) have developed a complete protocol that includes groups of satellite animals that are sacrificed 48 h after exposure, necropsied, and a standardized organ list collected, weighed, and prepared for histological assessment. This list routinely includes the "first-line" organs: brain, thyroid, liver, kidneys, heart, and adrenals. The same organs are collected from all other animals, that is, those that die as a result of the toxicity as well as control animals. Additional tests can be included if one has a specific concern. For example, the structure of a test article may suggest that it has anticholinesterase potential. Therefore, one could include serum pseudocholinesterase determinations in the clinical laboratory package, as is frequently done for organophosphate and carbamate structures.

### 3.3 Supplemented Acute Studies

An example of the third-level acute toxicity test, a supplemented study, is given in Figure 14. Such tests are rarely performed but are of use when one wishes to obtain data other than descriptive toxicity data. For example, the addition of satellite groups of animals to be dosed with a radiolabeled compound to gain pharmacokinetic information will turn a "complete" study into a "supplemented" one. Another common practice is the addition of other examinations or measurements to gain more information about a potential target organ. An example of this would be recording ECGs in rats, which is too complicated and time consuming to do on a routine basis but should be considered if the heart is a potential target organ. One way of describing such a study is that it is a complete toxicity study carrying a specific screen "piggyback."

An excellent example of a supplemented protocol is that described by Gad and colleagues (1984). A neuromuscular screen was developed (Gad, 1982) and incorporated into their routine acute toxicity protocol for testing nonpharmaceuticals. Doing so allowed for the more systematic and quantifiable examination of effects of the CNS than reliance on simple clinical observations. The neuromuscular screen consists of a battery incorporating standard clinical observations plus some behavioral assessment techniques already described in the literature. These are summarized in Table 10. This screen has been further developed to become the now regulatorily required FOB (functional observational battery). An advantage of this screen is that it uses noninvasive techniques and therefore will require the use of no additional animals. If an animal is displaying signs of severe CNS depression 2 h postdosing, little useful data will be gathered by examining behavior. In testing a pharmaceutical it is probably better practice to apply the neuromuscular screen on days 2, 7, and 14 postdosing in an attempt to identify more subtle or lingering effects and to

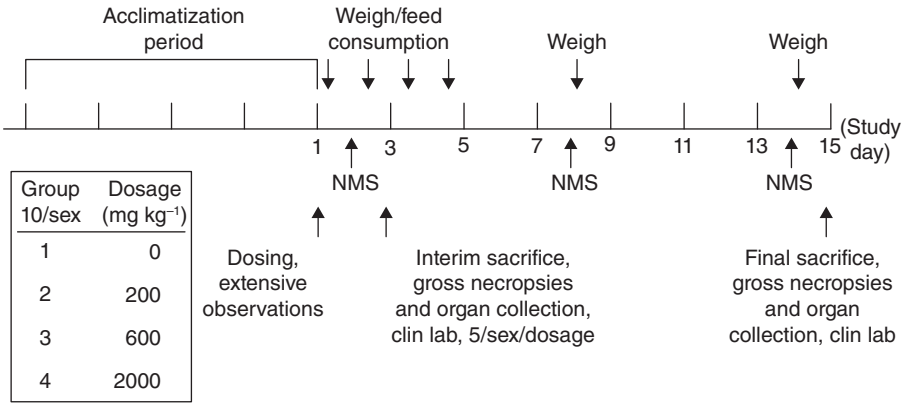


***Figure 14*** *Design and conduct of supplemented (or "heavy") acute systemic toxicity study. The figure illustrates the approach to such a study when it is to serve as the definitive systemic toxicity study for some period of time.*

**TABLE 10 Neuromuscular Screen Observations**

| Observation | Nature of Data Generated[a] | Correlates to Which Neutral Component[b] |
|---|---|---|
| Locomotor activity | S/N | M/C |
| Righting reflex | S | C/M |
| Grip strength (forelimb) | N | M |
| Body temperature | N | C |
| Salivation | Q | P |
| Startle response | Q | S/C |
| Respiration | S | M/P/C |
| Urination | S | P/M |
| Mouth breathing | Q | S |
| Convulsions | S | C |
| Pineal response | Q | Reflex |
| Piloerection | Q | P/C |
| Diarrhea | S | GI tract/P/M |
| Pupil size | S | P/C |
| Pupil response | Q | P/C |
| Lacrimation | Q | S/P |
| Impaired gait | S | M/C |
| Stereotypy | Q | C |
| Toe pinch | S | S (surface pain; spinal reflex) |
| Tail pinch | S | S (deep pain) |
| Wire maneuver | S | C/M |
| Hind-leg splay | N | P/M |
| Positional passivity | S | S/C |
| Tremors | S | M/C |
| Extensor thrust | S | C/M |
| Positive geotropism | Q | C |
| Limb rotation | S | M/C |

[a]Data quantal (Q), scalar (S), or interval (N). Quantal data are characterized by being of an either/or variety, such as dead/alive or present/absent. Scalar data allow one to rank something as less than, equal to, or greater than other values, but one cannot exactly quantitate the difference between such rankings. Interval data are continuous data where one can assign (theoretically) an extremely accurate value to a characteristic that can be precisely related to other values in a quantitative fashion.
[b]Peripheral (P), sensory (S), muscular (M), or central (C).

chart recovery from these effects. For drugs that produce no observable CNS effect following dosing, the neuromuscular screen can be done a few hours postdosing. The more extensive and detailed nature of the data generated by the neuromuscular screen permits more confidence in the conclusion that the test article had no effect on the CNS. Any suspect target organ can be investigated in a similar fashion. Depending on the invasiveness of the supplementary techniques, satellite groups may or may not need to be added to the study. Care must be taken in this regard to prevent the study from becoming too cumbersome and too complicated to conduct. It may be better to address some questions as separate studies. For this reason, one should not attempt to address more than one supplemental question in any one study.

### 3.4 Acute Toxicity Testing with Nonrodent Species

The designs described thus far for acute toxicity testing generally assume that the test species being used is a rodent. Nonrodent species are also used for acute toxicity testing. Many regulatory bodies require acute testing in at least one nonrodent species. The animals most often used are the dog, pig, or monkey. Veterinary products will also be tested in the target species. For example, a flea collar intended for cats must be tested in cats. While the rabbit is not technically a rodent, it is the species of choice for a variety of tests for assessing acute oral or intravenous toxicity and is considered a rodent for regulatory purposes. The section is written with the dog and monkey in mind. Clearly, there are some profound differences between these species and rodents with regard to handling, husbandry, and dosing. Here we focus on the design differences in toxicity testing in large species.

For financial, procurement, and ethical reasons, acute systemic toxicity tests on nonrodents are not performed using traditionally designed animal-intensive protocols. The minimal acute study requires 30–50 animals. Complete and supplemented studies will usually require even more. At a cost of $1500 per beagle dog, $1200 for minipig, or $3500 per monkey, the animal costs alone are enough to make such studies with these species prohibitively expensive. Vivarium space and husbandry costs are also much higher with nonrodent species that with rodents. Nonrodents also require a much longer prestudy quarantine period than rodents: at least 6–8 weeks for dogs and pigs and 18–24 weeks for monkeys. Treatment during the quarantine period is more extensive than that given rodents. The animals should be given frequent physical examinations, including complete clinical laboratory panels and appropriate tests for common illnesses and parasites. Special care must be taken with monkeys not only because they can be vectors of human disease but also because they can contract human diseases and a sick animal can compromise study outcome. All these factors dictate that these animals should be used sparingly. Hence, it is most common to study acute systemic toxicity in nonrodent animals using a pyramiding dosage design. The typical study will consist of two treated animals per sex and two control animals per sex for a total of eight animals. A typical protocol is shown in Figure 15.

The use of fewer but larger animals permits more extensive observations of each individual. Following each dose, animals can be given complete physical examinations that include palpations, behavioral checks, spinal reflex checks, pupillary light reflexes, respiration rate, ECG recording, and rectal temperature measurement. Blood samples can also be collected following each dose to determine standard clinical chemistry and hematology profiles. Hence, while fewer animals are used with the pyramiding dosage protocol, more information per animal is collected.

The small number of animals used in a pyramiding dosage study makes it difficult to do standard statistical comparisons. This difficulty can be overcome to a certain extent by taking advantage of two design aspects of the pyramiding

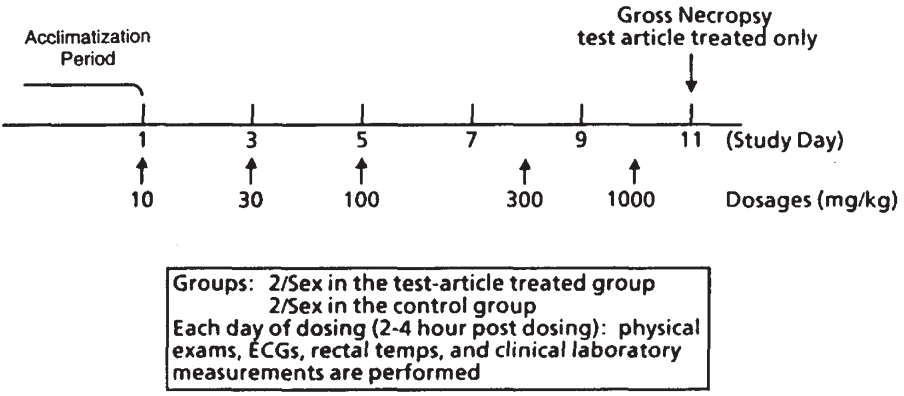


***Figure 15*** *Example of pyramiding dose study for acute toxicity testing in nonrodent species.*

protocol. First, pretreatment data can and should be obtained on all animals for all parameters examined or determined. In-study comparisons should be made both to pretreatment data and to concurrent control animals. Such comparisons can be made not only on the basis of absolute numbers but also on the magnitude of any changes from pretreatment values. Second, all animals should be measured repeatedly throughout the study. Hence, to reflect a true drug-related effect, the magnitude of change should increase following each dose (though one must be aware of the potential for the development of tolerance as induction of metabolism). This is in fact the only way one can make any dosage–response or threshold conclusions using the pyramiding protocol.

Seldom are drugs tested in nonrodent animals via routes other than the intended or likely routes of human exposure. Hence, the most common routes in these types of protocols are oral, intravenous, and respiratory. Rarely is a test article given to nonrodent species by the intraperitoneal route. Routes are discussed elsewhere in detail (Gad and Chengelis, 1999), but some discussion is appropriate here because of design considerations. Test articles are normally given orally by capsule to dogs and pigs and by gavage to monkeys. Nonrodents have to be restrained if dosed by gavage, making the process very labor intensive. This is minimized by the small number of animals specified by the pyramiding protocol. In contrast, because of the differences in size, it is much easier to deliver a test article intravenously to nonrodents than to rodents. For topical studies, the rabbit is the nonrodent choice because it is easier to prevent a rabbit from grooming the delivery site, and considerably less material is required to deliver a comparable dose to a rabbit than a dog or pig. Acute dermal studies are not, however, usually done with a pyramiding study design but rather as a limit dose study

The biggest problem with the pyramiding protocol is the development of tolerance. If no toxicity is observed, the chemical could be innocuous or animals could have developed tolerance during the study. The escalating dosage feature of the pyramiding protocol is an excellent vehicle for fostering the development of tolerance. One can check this by dosing additional naïve animals at the limit dosage to confirm, as it were, a negative result. Another problem, which is most peculiar to the dog, is emesis. Oral administration of a large amount of almost any material will cause a dog to vomit. This is always somewhat of a surprise to toxicologists whose prior experience is primarily with rodents, which cannot vomit. One should pay close attention to dogs the first hour after capsule delivery. If the dog vomits up most of the dose, the actual dosage absorbed could be grossly overestimated. This can be a particular problem if one is using the results of a pyramiding dosage study to set the dosages for a repeated-dose study. Dogs can develop tolerance to the emetic effect of a set dosage. When this occurs, absorption and resulting blood concentrations of a test article can increase dramatically, resulting in more florid toxicity than expected on the basis of the pyramiding study. Another problem is that emesis can result in secondary electrolyte changes—especially decreases in chloride—that may be mistaken for a direct effect of the test article. If emesis is a severe problem, one can study toxicity in a different nonrodent species or divide larger dosages into two or three divided dosages on the day of dosing.

As with traditionally designed rodent studies, the pathology component of pyramiding studies usually consists of gross necropsies followed by (when appropriate and necessary) histological assessment of gross lesions. Unfortunately, this study design does not permit the establishment of a dose–response relationship with regard to gross necropsy findings. In addition, the small number of animals makes definitive conclusions difficult. Usually, gross lesions are defined in absolute terms with few comparisons to control animals. Suspected target organs should be further investigated in subsequent subchronic studies or in rigorous and specific mechanistic studies. Because of the limited value of the pathology data generated by the pyramiding protocol, control animals should not be terminated but rather should be saved for reuse.

### 3.5 Factors That Can Affect Acute Tests

Many investigations into the sources of variability in acute toxicity testing have been conducted, and these have been reviewed by Elsberry (1986). The factors causing the greatest interstudy variation included lack of specifications for sex, strain, age, and weight range. When clearly defined, detailed protocols were used and interlaboratory variation was found to be minimal. Hence, it is equally important that the details of the protocol be well described and followed. It is not appropriate to draw dosage–response conclusion by comparing

groups that differ substantially in age or that have been fed, fasted, or otherwise manipulated differently. Guidelines for standardization of acute toxicity testing were proposed by the interagency regulatory liaison group [Interagency Regulatory Liaison Group (IRLG), 1981; Elsberry, 1986]. These do not differ markedly from those mandated by the Toxic Substance Control Act of 1986 (Gad and Chengelis, 1999).

***Number, Size, and Sex of Dosage Groups*** The precision with which lethality and signs of toxicity are described will depend on the number of groups (hence, dosages) and the number of animals in each group. Between 1940 and 1980, the standard was to use from four to six dosages with 10 animals per dosage. The current emphasis is on limiting the number of animals used for acute testing, particularly with recognition of the limited value of "precise" lethality data (Gad and Chengelis, 1999). Retrospective analyses by DePass (1989) and Olson et al. (1990) have demonstrated that decreasing group size to two to three animals generally has little impact on overall study results. Hence, the number and size of dosage groups will depend, to an extent, on the methods of statistical analysis. The classic statistical methods for analyzing lethality data (or, indeed, any quantal dosage–response data) were published between 1930 and 1960 and have been extensively reviewed by Armitage and Allen (1959) and Morrison et al. (1968). These methods are mentioned here with regard to the demand they make on protocol design—specifically, the number of dosage groups, the spacing of the dosages, and the number of animals per dosage group. The probit and moving-average methods are most commonly used today. In general, all methods of calculation and computation are more precise if the designs are symmetrical (i.e., the dosages are evenly spaced and the group sizes are equal). The probit method, first developed by Bliss (1935, 1957) and later refined by Finney (1971, 1985), is considered to be the most precise, but it requires at least two groups of partial responses (i.e., mortality greater than zero but less than 100%). This may require dosing more than three groups until this criterion is met. It also deals ineffectively with groups that had either zero or 100% mortality. (The most common correction for these groups is to substitute 0.1% for zero and 99.7% for 100%.) The moving-average method, first described by Thompson and Weil (1952), does not require partial responses, deals effectively with complete responses, and therefore can produce an acceptable estimate of an $LD_{50}$ with as few as three groups of three to five animals each. The moving-average method can also be used to design the experiment. Groups can be dosed in a sequential fashion as in a pyramiding study, with each step dictated by the moving-average method. Once evidence of toxicity is observed, further dosing is discontinued. This method requires that the dosages be separated by a constant geometric factor (e.g., 2, 4, and 8 $mg\,kg^{-1}$) and that groups be of equal size. Weil (1952) and later Gad and Chengelis (1999) and Gad (2007) published tables that allow for the easy calculation of the $LD_{50}$ using $K = 3$ (where $K$ is the number of dosage groups minus 1). The $LD_{50}$ for $K < 3$ can be easily calculated without

the aid of tables. In addition, methods for estimating the confidence limits of this calculated $LD_{50}$ have also been published (Gad and Chengelis, 1999). Traditionally, the moving-average method has not been extensively used because, while it yielded an estimate of the $LD_{50}$, it did not give the slope of the (probit-transformed) lethality curve. However, Weil (1983) has published a method for calculating a slope from the same data. Hence, an estimate of the $LD_{50}$ and slope can be obtained from as few as three groups of three to five animals per group, provided that at least one group shows a response less than 50% and another shows a response greater than 50%.

The Litchfield and Wilcoxon (1949) plotting method was once commonly used. It is certainly a valid method, and it poses no more restrictions on study design than those imposed by the probit method. The Litchfield–Wilcoxon method has become a victim of technology as modern, hand-held calculators and the ready availability of simple computer programs have made other methods more convenient to run. However, at least one software company has adopted the Litchfield–Wilcoxon method for its acute toxicity protocol package.

The normit–chi square, developed by Berkson (1955), is also sometimes used. Like the probit method, the normit–chi square does not absolutely require equally spaced dosages or equal group sizes, but it does require at least one partial response. Hence, fewer dosage groups may be needed with the normit–chi square method than with the probit method. According to Waud (1972), the correction for including complete responses is better than that used for probit analysis but is still "tainted." His method supposedly deals adequately with complete responses, but it is extremely complex and, probably for this reason, is rarely used.

In an early paper, Karber (1931) published a simple method (often described but rarely cited) for calculating $LD_{50}$. It does not require that dosages be equally spaced, but widely divergent dosages will lead to a biased result. The method was originally described for groups of equal size, but groups of slightly varying sizes can be used, provided they do not differ by more than a few animals each. In this case, mean group size can be inserted into Karber's formula with little change in accuracy. The formula is very simple, and one can calculate an acceptable estimate of $LD_{50}$ quickly with only a few arithmetic computations. This method, unlike those mentioned above, does not allow for calculating the confidence limit or slope of the probit response curve. Hence, if these calculated parameters are not sought, the Karber method allows one a bit more freedom in picking dosages.

While much has been written about the influence of gender on acute lethality, most authors now agree that there are seldom any substantial differences in $LD_{50}$ due to sex (DePass et al., 1984; Gad and Chengelis, 1999). In those instances where there is a sex-related difference, females tend to be more sensitive than males (approximately 85% of the time). If one is willing to accept this amount of uncertainty, only one sex needs to be tested. Alternatively, as few as two to three animals per sex per dosage can be used. Schutz

and Fuchs (1982) have demonstrated that, by pooling sexes, there are seldom any substantial differences in the $LD_{50}$ calculations between groups consisting of five per sex versus three per sex. If there are no substantial differences between sexes (i.e., 70% mortality for males and 80% for females at a dosage), the results from pooling the sexes can provide a pooled $LD_{50}$. For most safety considerations, an $LD_{50}$ derived on this basis will be acceptable and will result in the use of fewer animals.

### 3.6 Selection of Dosages

In setting dosages for acute studies a few commonsense rules have to be applied. First, the intrinsic biological and chemical activity of the test article must be considered. Zbinden and Flury-Roversi (1981) have documented several cases where lethality was of no biological relevance. The oral lethality of tartaric acid, for example, is due to the caustic action of a high concentration of acid in the GI tract. In these instances, limit tests are more appropriate tests. Additionally, it is uncommon that a completely unknown chemical will be tested. Factors such as known pharmacological profile; chemical or physical characteristics including molecular weight, particient coefficient, and the like; and the toxicity of related chemicals should be considered. For example, it is likely that a polymeric, poorly soluble molecule will not be bioavailable at an initial dosage of $100\,mg\,kg^{-1}$. A full understanding of all available data will permit one to pick dosages with more confidence and thereby save both time and resources.

Second, no protocol will yield high-quality data if all dosages given cause 100% lethality. Therefore, one is best advised to pick widely spaced, rather than closely spaced, dosages. In general, the best dosage regimen includes a dose that will definitely produce a high incidence of severe toxicity, another that will produce marginal toxicity, and one in between. If this pattern is obtained, adding more groups does not generally change the results. This point is illustrated by the data in Table 11. For two drugs, an $LD_{50}$ of $300\,mg\,kg^{-1}$ was obtained using six groups of 10 mice each. Essentially the same result was obtained if the second, fourth, and sixth groups were eliminated and not used in the calculations. Behrens (1929) noted this phenomenon almost 60 years ago.

Widely spaced dosages also decrease the likelihood of nonnormotonic data, where mortality does not necessarily increase with dosage (see Table 12). This can occur when the test chemical has a shallow dose–response curve and the group size is small (three to four animals). Wile it is possible to calculate an $LD_{50}$ from such data, the slope and confidence limits will be inaccurate. Nonmonotonic data can also occur if the lethality is indeed biphasic. If one suspects that this is occurring, additional dosages should be examined. For safety considerations, only the first part of the curve, the lowest $LD_{50}$, is of importance.

**TABLE 11 Sample Data Sets: $LD_{50}$ Calculations Using Fewer Dosages**[a]

| SC-27166 | | Theophylline | |
|---|---|---|---|
| Dosage (mg kg$^{-1}$) | Mortality | Dosage (mg kg$^{-1}$) | Mortality |
| 100 | 0/10 | 280 | 0/10 |
| 180 | 0/10 | 320 | 3/10 |
| 240 | 4/10 | 370 | 5/10 |
| 320 | 7/10 | 430 | 9/10 |
| 560 | 9/10 | 500 | 10/10 |
| 1000 | 10/10 | 670 | 10/10 |
| $LD_{50}$ = 300 | | $LD_{50}$ = 300 | |
| *Using Every Other Dosage* | | | |
| 100 | 0/10 | 280 | 0/10 |
| 240 | 4/10 | 370 | 5/10 |
| 560 | 9/10 | 500 | 10/10 |
| $LD_{50}$ = 290 | | $LD_{50}$ = 290 | |

[a]Adult male mice; drugs given by gavage.

**TABLE 12 Sample Data Sets: Homogeneous versus Heterogeneous Data**

| Homogeneous[a] (Normotonic) | | Hetereogeneous[b] (Nonnormotonic) | |
|---|---|---|---|
| Dosage (mg kg$^{-1}$) | Mortality | Dosage (mg kg$^{-1}$) | Mortality |
| 300 | 0/20 | 620 | 0/10 |
| 600 | 1/20 | 1600 | 2/10 |
| 800 | 10/20 | 2100 | 8/10 |
| 1000 | 17/20 | 2800 | 5/10 |
| | | 3700 | 8/10 |
| | | 5000 | 8/10 |

[a]Data from study of SC-31828 using adult rats of both sexes.
[b]Data from study of SC-3894 using adult male rats.

***Timing*** The greatest precision in any lethality curve is obtained when the number of experimental variables is kept to a minimum. Hence, it is best if all the animals used for determining a specific curve are dosed on the same day and, if possible, at the same time of day, which limits age-related and diurnal effects. If a total of only 15 animals are being dosed, this is not a difficult task for a single well-trained technician. However, if the test substance is of unknown lethality, it is imprudent to deliver all doses on the same day. It is common practice for a single dosage group to be treated on the first day of an experiment and the dosages for the second and third groups to be adjusted pending the results of the first group. Generally, most acute deaths will occur within 24 h of dosing. Delayed deaths (those occurring more than 24 h after

dosing) are relatively rare and generally restricted to the 72-h period following dosing (Gad et al., 1984; Bruce, 1985). Hence, waiting for 24 h between doses will generally yield sufficient data to allow the choice of the next dosage. For example, if all but one of the animals dosed in the first groups dies, there is no doubt that the next dosage should be adjusted downward considerably, whether or not the final animal eventually dies. All the dosing for a single curve can be completed in three days. If a test article is being tested in traditional protocols (with two species, two routes, separate sexes), the two initial groups by a route can be treated on the first day of the dosing period and the second route initiated on the next day. Subsequent dosages can be adjusted on alternate days. Little real impact on the results will occur if there are two to three days between dosing sets. After that, however, the increasing age of the animals may result in a change in sensitivity. As reviewed by Balazs (1976), for example, the ratios of the $LD_{50}$s obtained in adult animals to the $LD_{50}$s obtained in neonates can vary from 0.002 to 160. One can use longer observation periods between dosing days if separate animal orders are timed for delivery to ensure that all animals dosed are closer in age. As a rule of thumb, the animals should not differ in age by more that 15%; hence, the younger the animals, the small the age window.

## 4 SCREENS

Screens are generally not safety studies in the regulatory sense. These are the studies done, as the name implies, to examine several chemicals in order either to select those with the most desirable properties for development or to eliminate those that have undesirable properties. There is nothing novel about screening; the process has been an integral part of pharmaceutical research for decades (Irwin, 1962). In a pioneering paper, Smyth and Carpenter (1944) described a screening process for gathering preliminary toxicity data for a new chemical. In their discussion they clearly state the underlying rationale for toxicity screening:

> Opinions upon the toxicity, hazards of manufacture, and fields for safe use must be expressed regarding many chemicals which will never be produced in quantity. Large expenditures of time and money upon securing these basic conclusions is not justified. Later, when a few of the new compounds are obviously going to be made commercially, more detailed studies can be undertaken.

Screens are designed for speed, simplicity, and minimal resource expenditure. They are designed to answer positive, single-sided questions. For example, the lack of an effect in an initial screen does not mean that toxicity will not be manifested with a different formulation or in a different species. It is for this reason that screen should not, as stated by Zbinden et al. (1984), be seen as replacements for thorough safety testing. An acute toxicity screen can be

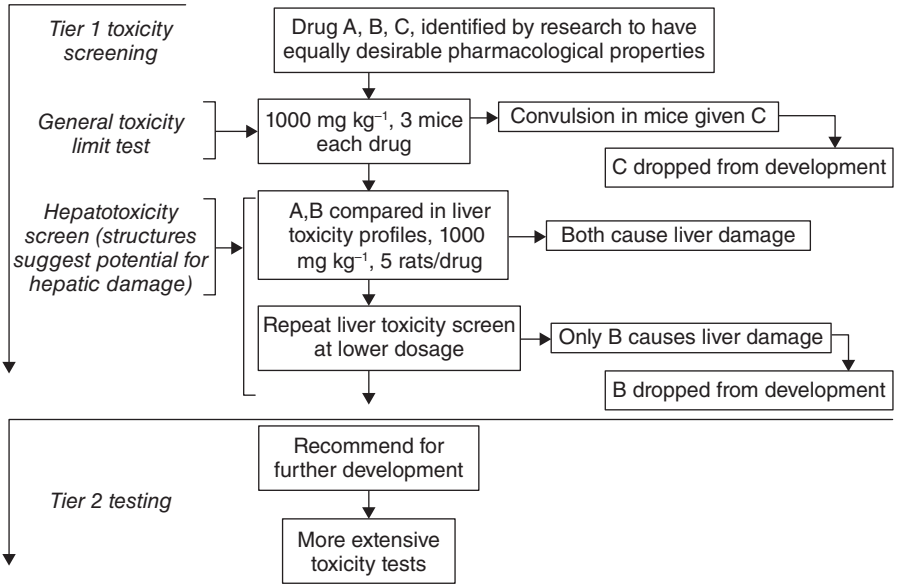


***Figure 16*** *Example of use of screens in selecting drug candidates for development.*

the first leg in a decision tree or tier-testing process for selecting a chemical or drug candidate for development. An example of this process is given in Figure 16.

## 4.1 General Toxicity Screens

There are two types of acute toxicity screens. In the general toxicity screen, animals (often, for economic reasons, mice) are exposed to two or three predefined dosages of chemical. No more than three mice per dosage are necessary and no control group is required. An example of this type of protocol is shown in Figure 17. The animals are carefully observed for mortality and obvious signs of toxicity, such as convulsions, but no attempt should be made to quantify the severity of a response. There is seldom any need to have an observation period of more than four to five days. Because of the quantal nature of the data, interpretation is straightforward. There are four possible outcomes: (1) no death or signs of toxicity seen at dosages up to $X$ mg kg$^{-1}$; (2) no deaths but evident signs of toxicity seen at $X$ mg kg$^{-1}$; (3) evident signs of toxicity at $X$ mg kg$^{-1}$; (4) deaths and evident signs of toxicity both occurring at $X$ mg kg$^{-1}$. General toxicity screens may also provide the preliminary information for picking the dosages for more definitive acute studies.

There are two ways to apply the data from toxicity screens to the development of a drug or chemical. On a relative basis, the drugs under consideration can be ranked according to screen results and the one that appears to be the

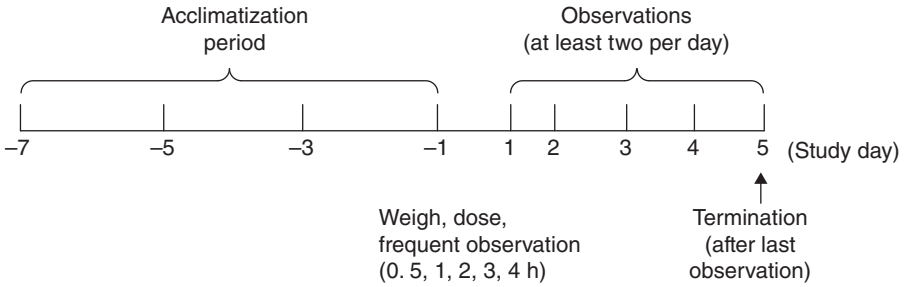


| | | Results | |
|---|---|---|---|
| *Group* | *Dosage* | *Deaths* | *Signs summary* |
| 3 mice | 100 mg $kg^{-1}$ | 0/3 | None |
| 3 mice | 300 | 0/3 | CNS depression |
| 3 mice | 1000 | 0/3 | CNS depression, agonal changes |

***Figure 17*** *Example of general toxicity screen.*

least toxic can be chosen for future development. Alternatively, decisions can be made on an absolute basis. All candidates that are positive below a certain dosage are dropped, and all those that are negative at or above that dosage will continue to the next tier of testing. If absolute criteria are used, the screen need be done only at the critical dosage. If only one dosage is examined, the test is a limit test. A limit test of this kind is the simplest form of toxicity screen and, depending on the nature of subsequent testing, it is highly recommended.

Fowler and his colleagues (1979) have described a rat toxicity screen (illustrated in Figure 18) that is more extensive and detailed than the one shown in Figure 17. It includes two rounds of dosing. In the first round, up to 12 rats are (singly) exposed to six different dosages by two different routes for the purpose of defining the maximally tolerated dose (MTD). In the second round of dosing, 16 rats are dosed at two-thirds (0.66) of the MTD and sacrificed on a serially timed basis for blood sample collections to determine test article concentrations and for clinical laboratory tests. These features make this design too complicated, time consuming, and expensive to run as an initial screen. This design is better suited as a second-tier screen to provide a more extensive follow-up study for a more limited screen. Fowler et al. contend that their screen disclosed most toxicity uncovered by more conventional studies. This screen was most successful in defining acute CNS, liver, or kidney toxicity (Fowler et al., 1979). Lesions that require long-term exposure, such as those generally involving the eyes, may not be detected in this type of screen.

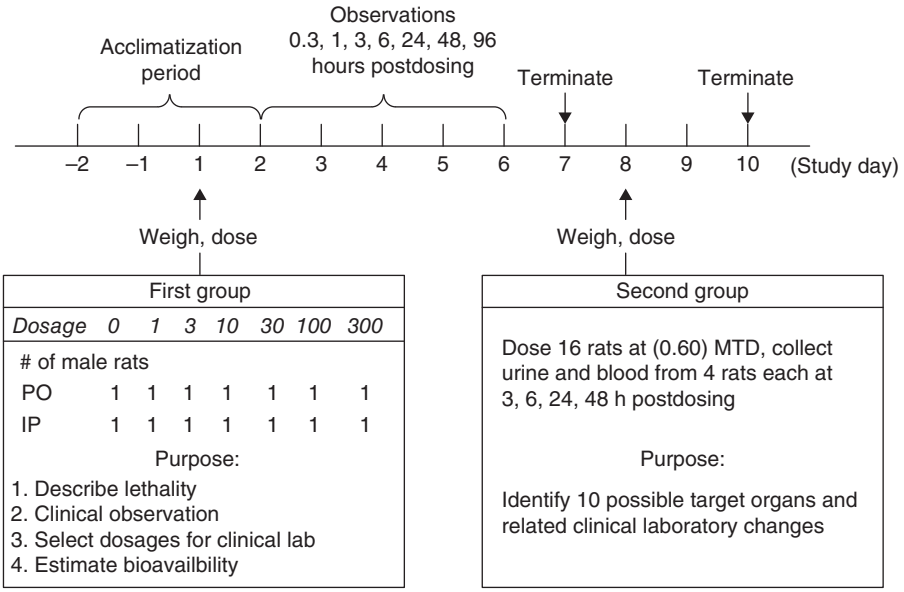


***Figure 18*** *Example of rat toxicity screen for drugs.*

Up/down or pyramiding designs can be used for general toxicity screens, but this is not a common approach because of the time involved. In addition, if several chemicals are being compared, an up/down study where death occurs at different dosages can be too complicated to run. It is much easier to test several chemicals at the same time using a limit test design. Because only individual animals are dosed, these designs can be used when there is a very limited amount of test article available and/or there are few prior data on which to base an expected toxic dosage.

Hazelette and colleagues (1987) have described a rather novel pyramiding dosage screen that they term the rising-dose tolerance (RDT) study (illustrated in Figure 19). The study, which uses a subacute rather than an acute dosing regimen, can also be used as a range-finding study design. The rats are exposed for four days to the initial dosage followed by three days of recovery before the next four-day dosing period at the next highest dosage. This process is repeated for the three dosing cycles. Plasma and urine samples are collected for clinical chemistry and urinalysis as well as test article determinations. Necropsies and microscopic examinations are performed. While this study design is novel, it appears to provide considerable acute data. It is also possible that this design could generate sufficient data to plan a pivotal subchronic study and therefore replace a traditional two-week study, resulting in considerable savings of time and animals. This is not a simple study and, therefore, is

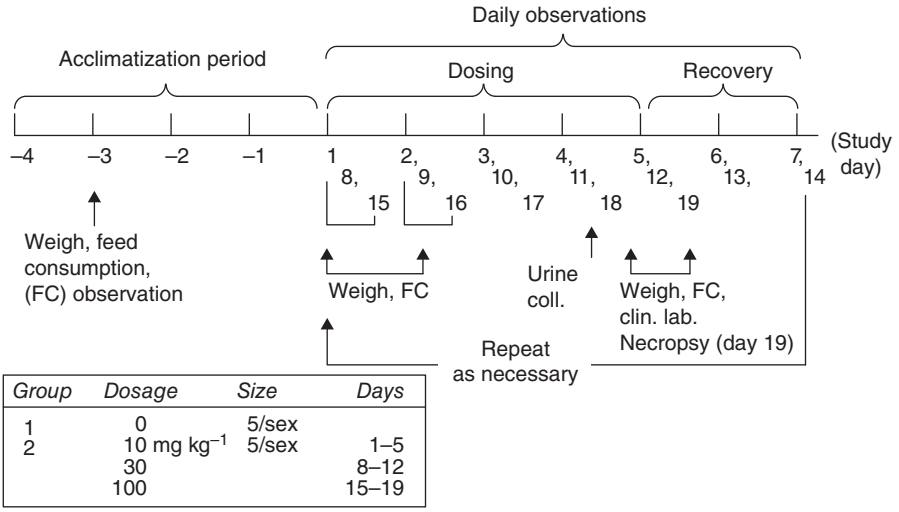


| Group | Dosage | Size | Days |
|---|---|---|---|
| 1 | 0 | 5/sex | |
| 2 | 10 mg kg$^{-1}$ | 5/sex | 1–5 |
| | 30 | | 8–12 |
| | 100 | | 15–19 |

***Figure 19*** *Example of rising-dose tolerance test [no pharmacokinetic (PK) groups].*

inappropriate as an initial screen, but it would appear to be appropriate for a second-tier test.

## 4.2 Specific Toxicity Screening

The second type of acute toxicity screen is the specific toxicity test. This type of test is done when one has a specific toxicological concern, for example, when prior structure–activity data suggest that a family of chemicals can be hepatotoxic. A screen to select the chemical with the least hepatotoxic potential is then in order. These tests are also done, as described by Zbinden (1984), to look for a specific toxicological effect that may be easily overlooked in a routine safety study. Zbinden gives, as an example, screens that are designed to detect specific lesions to the hemostatic process. As pointed out by Irwin (1962) over three decades ago, such tests have their greatest power if more than one measure of specific target organ toxicity is used. Dayan (1983) refers to this technique as a matrix of measurements. In a liver toxicity screen, for example, liver weights (both absolute and relative), gross necropsy examinations, and a battery of serum enzyme assays should all be part of the protocol. As a general rule, because of the time and expense involved, screens should be designed to minimize the use of histopathological techniques. The dosages can be standardized or set on the basis of the results of a generalized toxicity screen. Specific toxicity screens can be the next-level test in the decision

tree process for selecting a candidate for development (as illustrated in Figure 16).

The number of animals and the number of dosages are highly dependent on the type of data gathered. A few rules of thumb should be followed: (1) Keep it lean. Each additional group, animal, or test article added to a protocol makes the study exponentially more difficult to conduct; simplicity is one of the most important features of a screen. (2) The more parameters examined, the fewer the number of animals required. (3) If normal limits of a test parameter are relatively broad (e.g., serum glucose), more animals will be required than if the parameter is normally tightly controlled (e.g., prothrombin time). In general, 3 is the minimum and 10 is the maximum number of animals required per group. Further, if a single chemical is examined per study, no more than three groups will be required. If more than one chemical is included in the study, then a single dosage (limit) group per chemical is the best design.

Strictly speaking, an acute toxicity study is conducted to examine the effect of a single dose of a single compound. In designing specific toxicity screens, however, deviation from this principle is permissible if it increases screen sensitivity. For example, the sensitivity of mice to many indirect hepatotoxins will be enhanced by prior treatment with phenobarbital. Hence, the sensitivity of a hepatotoxicity screen will be enhanced if the mice are pretreated for three days with phenobarbital.

The screen should be validated for consistency of response with both positive and negative control articles. A positive control article is one that is known to reliably produce the toxic syndrome the screen is designed to detect. Concurrent control groups are not required with each replicate. Rather, control groups should be evaluated on some regular basis to ensure that screen performance is stable. Because a screen relies on a biological system, it is not a bad idea to test the control benchmarks, particularly the positive ones, on a routine period basis. Not only does that give one increased confidence in the screen, but it also provides a historical base against which to compare the results of new test articles. Zbinden and colleagues (1984) refer to the positive control as the reference compound, and they have discussed some of the general criteria to be applied in the selection of these compounds. Any changes to the design should trigger revalidation. Any analytical methods should be subjected to PASS (precision, accuracy, sensitivity, and selectivity) validation.

Interpretation of specific toxicity screen data is not as straightforward as that of a general toxicity screen. This is because the data will often be continuous, following a Gaussian, or normal, distribution. This has two ramifications. First, for results around the threshold, it may be very difficult to differentiate between positive and negative responses. Second, for any one parameter, there is a real chance of false statistical significance (type I errors), especially if small numbers of animals are used. This occurrence is one of the reasons why specific toxicity screens should include the determination of more than one variable, since it is unlikely for multiple false positives to occur in the same group of animals. An undetected false positive could lead to the dropping of a promis-

ing candidate in error. False negatives, by contrast, may not be as critical (other than the time lost and the resources spent), because extensive subsequent tests should lead to the more complete description of the test article's toxic properties.

The problems described in the preceding paragraph assume that the screen will include a traditional ("negative," or vehicle) control group and the data from the treated groups will be compared to those of the control group by standard methods. These problems will be minimized if no control group and therefore no traditional statistical comparisons are included. In addition, a decrease in the number of animals used simplifies the study. Data can be interpreted by comparison to a historical control database as described by Zbinden (1984). The threshold, or test criterion $X_c$, is calculated according to the formula

$$X_c = m + (z)s$$

where $m$ is the population mean, $s$ is the standard deviation, and $z$ is an arbitrary constant. This formula is essentially a method of converting continuous data to quantal data: It is used to determine if individual animals are over the test threshold, not if the group mean is over the threshold. Analysis of screening data by comparison to experience (i.e., historical control data) and an activity criterion are discussed in greater detail in Chapter Screens in Safety and Hazard Assessment and by Gad (1989, 2007). The higher the $z$ value, the lower the probability of a false positive, but the lower the sensitivity of the screen. Again, including multiple parameters in the screen helps alleviate this problem. Zbinden has proposed a ranking procedure in which various levels of suspicion (LOS) or a level of certainty (LOC) is assigned to the result of a toxicity screen. This is simply a formalized fashion of stating that the more animals that respond and the greater the severity of the response, the more certainty one has in drawing a conclusion. If relative comparisons are being made, this system provides a framework for ranking test articles and selecting those to continue to the next tier of testing.

With regard to specific toxicity screening, behavioral toxicity screening is an area currently generating a great deal of interest. As reviewed by Hopper (1986), there are several reasons for this interest. First, the Toxic Substance Control Act of 1976 legislatively recognized behavioral measures as essential to examining chemicals for neurotoxic potential. Second, the structure and function of the CNS are not amenable to traditional methods of examination, in that profound behavioral changes can be induced in the absence of any detectable morphological lesions. This large and somewhat controversial subject is outside the scope of this chapter. Specific screening strategies are presented and critically discussed by Hopper (1986). Other recommended references to consult for different perspectives on acute toxicity testing are Rhodes (2000), Brown (1980), and Arnold et al. (1990).

## 5 PILOT AND DRF STUDIES

Between acute studies and the repeat-dose (14- or more typically 28-day) studies (described in Chapter Repeat-Dose Toxicity Studies), there are those which are necessary to be able to set doses required by regulation to be able to initiate initial clinical studies.

The DRF-style studies are not required by regulation but rather by proper scientific practice. As such, it should be noted that while the results must be included in the documents submitted to regulatory authorities, they do not have to be performed in compliance with good laboratory practices (GLPs). The studies should be conducted with dosing by the route and regimen (dose frequencies) intended for clinical studies. The formulations used should be those to be employed in the 14- or 28-day studies, but the actual drug substance [active pharmaceutical ingredient (API)] lot does not need to be.

For nonrodent species (dog, pig, or primate), dose levels are set based on the maximum tolerated dose identified in the acute pyramiding dose study. The design of the study is presented in Table 13.

The design for the rodent pilot study is shown in Table 14. Dose levels are set based on results seen in the acute study with the middose typically being the maximum tolerated dose seen in that study.

The hematology, clinical chemistry, and urinalysis measures are as described in Chapter Repeat-Dose Toxicity Studies.

**TABLE 13 Seven-Day Nonrodent Pilot Toxicology Study**

| | |
|---|---|
| Animal: 9 Males, 9 Females | |
| Study Design: | |
| Vehicle control | 3 males, 3 females |
| Low dose | 3 males, 3 females |
| High dose | 3 males, 3 females |
| Dosing: Once daily all animals for 7 days | |
| Observations: (Mortality/moribundity) Twice daily | |
| Clinical examination: Daily after dosing | |
| Body weights: Before start and on days 1, 2, 3, 5, and 7 | |
| Food consumption: Before start and on days 1, 2, 3, 5, and 7 | |
| Physical examinations: Conducted by staff veterinarian on all animals prior to initiation of compound administration and at study termination | |
| Electrocardograms: All animals prior to initiation of compound administration and at study termination | |
| Clinical pathology: Hematology, clinical chemistry, and urinalysis evaluations on all animals pretest and at study termination | |
| Necropsy: All animals on day after last dose | |
| Organ weights: Adrenals, brain, heart, kidneys, liver, lungs, ovaries with oviducts, pituitary, prostate, salivary glands, seminal vesicles, spleen, thyroid with parathyroid, thymus, testes, uterus | |

**TABLE 14 7-Day Rodent Pilot Toxicology Study**

| | |
|---|---|
| Study design: main study | |
| Vehicle Control | 5 males, 5 females |
| Low Dose | 5 males, 5 females |
| Mid Dose | 5 males, 5 females |
| High Dose | 5 males, 5 females |
| Dosing: Once daily all animals | |
| Observations: (Mortality/moribundity) Twice daily | |
| Detailed clinical observation: Daily | |
| Functional observational batter: After first dose | |
| Body consumption: Weekly | |
| Clinical pathology: Hematology, clinical chemistry, and urinalysis evaluations on all surviving main study animals at termination (see attachment) | |
| Toxicokinetics: Blood may be collected after first and last dose at six time points | |
| Necropsy: All main study animals, toxicokinetics animals euthanized and discarded | |
| Organ weights: Adrenals, brain, heart, kidneys, liver, lungs, ovaries with oviducts, pituitary, prostate, salivary glands, seminal vesicles, spleen, thyroid with parathyroid, thymus, testes, uterus | |

## REFERENCES


Armitage, P., and Allen, J. (1959). Methods of estimating the $LD_{50}$ in quantal response data. *J. Hyg.* 48:298–322.

Arnold, D. L., Grice, H. C., and Krewski, D. R. (1990). *Handbook of in Vivo Toxicity Testing*. Academic, San Diego, CA.

Auletta, C. (1998). Acute systemic toxicity testing. In S. Gad (Ed.), *Handbook of Product Safety Assessment*, 2nd ed. Marcel Dekker, New York.

Balazs, R. (1970). Measurement of acute toxicity. In G. Paget (Ed.), *Methods of Toxicology*. F. A. Davis, Philadelphia, pp. 49–81.

Balazs, R. (1976). Assessment of the value of systemic toxicity studies in experimental animals. In M. Mahlman, R. Shapiro, and H. Blumenthal (Eds.), *Advances in Modern Toxicity*, Vol. 1 Part1: New Concepts in Safety Evaluation Hemisphere Publishing, Washington, DC, pp. 141–153.

Behrens, B. (1929). Evaluation of *Digitalis* leaves in frog experiments. *Arch. Exp. Pathol. Pharmacol.* 140:236–256 (in German).

Berkson, J. (1955). Estimate of the integrated normal curve by minimum normit chi-square with particular reference to bioassay. *J. Am Stat. Assoc.* 50:529–549.

Bliss, C. (1935). The calculation of the dosage mortality curve. *Anal. Appl. Biol.* 22:134–167.

Bliss, C. (1957). Some principles of bioassay. *Am. Sci.* 45:449–466.

British Toxicology Society (1984). A new approach to the classification of substances and preparations on the basis of their acute toxicity. A report by the British Toxicology Society Working Party on Toxicology. *Human Toxicol.* 3:85–92.

Brown, V. K. (1980). *Acute Toxicity*. Wiley, New York.

Brown, V. K. (1983). Acute toxicity testing. In M. Bals, R. Riddell, and A. Worden (Eds.), *Animals and Alternatives in Toxicity Testing*. Academic, New York, pp. 1–13.

Brownlee, K., Hodges, J., and Rosenblatt, M. (1953). The up-and-down method with small samples. *J. Am. Stat. Assoc.* 48:262–277.

Bruce, R. (1985). An up-and-down procedure for acute toxicity testing. *Fund. Appl. Tox.* 5:151, 157.

Chan, P., O'Hara, G., and Hayes, A. (1982). Principles and methods for acute and subchronic toxicity. In A. Hayes (Ed.), *Principles and Methods of Toxicology*. Raven, New York, pp. 1–51.

Dayan, A. (1983). Complete program for acute toxicity testing—not only $LD_{50}$ determination. *Acta Pharmacol. Toxicol.* 52(Suppl. 2):31–51.

Deichmann, W., and Gerarde, H. (1969). *Toxicology of Drugs and Chemicals*. Academic, New York.

Deichmann, W., and LeBlanc T. (1943). Determination of the approximate lethal dose with about six animals. *J. Ind. Hyg. Toxicol.* 25:415–417.

DePass, L. (1989). Alternative approaches in median lethality ($LD_{50}$) and acute toxicity testing. *Toxicol. Lett.* 49:159–170.

DePass, L., Myers, R., Weaver, E., and Weil, C. (1984). An assessment of the importance of number of dosage levels, number of animals per dosage level, sex, and method of $LD_{50}$ and slope calculations in acute toxicity studies. In A. Goldberg (Ed.), *Acute Toxicity Testing: Alternative Approaches*, Vol. 2. Mary Ann Liebert, New York, pp. 139–153.

Dixon, W. (1965). The up-and-down method for small samples. *Am. Statist. Assoc. J.* 60:967–978.

Dixon, W. J., and Wood, A. M. (1948). A method for obtaining and analyzing sensitivity data. *J. Am. Stat. Assoc.* 43:109–126.

Elsberry, D. (1986). Screening approaches for acute and subacute toxicity studies. In W. Lloyd (Ed.), *Safety Evaluation of Drugs and Chemicals*. Hemisphere Publishing, Washington, DC, pp. 145–151.

Finney, D. (1985). The median lethal dose and its estimation. *Arch. Toxicol.* 56:215–218.

Finney, D. J. (1971). *Probit Analysis*, 3rd ed. Cambridge University Press, New York.

Fowler, J., Brown, J., and Bell, H. (1979). The rat toxicity screen. *Pharmacol. Ther.* 5:461–466.

Fowler, J., and Rutty, D. (1983). Methodological aspects of acute toxicity testing particularly $LD_{50}$ determinations: Present use in development of new drugs. *Acta Pharmacol. Toxicol.* 52(Suppl. 2):20–30.

Gad, S. (1982). A neuromuscular screen for use in industrial toxicology. *J. Toxicol. Environ. Health* 9:691–704.

Gad, S. C. (1989). Statistical Analysis of Screening Studies in Toxicology With Special Emphasis on Neurotoxicology. *J. Amer. Coll. Toxicol.* 8(1):171–183.

Gad, S. (2007). Statistics for toxicologists. In A. Hayes (Ed.), *Principles and Methods in Toxicology*, 5th ed. Taylor & Francis, Philadelphia.

Gad, S., and Chengelis, C. P. (1999). *Acute Toxicology Testing*, 2nd ed. Academic, San Diego, CA.

Gad, S., Smith, A., Cramp, A., Gavigan, F., and Derelanko, M. (1984). Innovative designs and practices for acute systemic toxicity studies. *Drug Chem. Toxicol.* 7:423–434.

Hazelette, J., Thompson, T., Mertz, B., Vuolo-Schuessler, L., Gree, J., Tripp, S., Robertson, P., and Triana, V. (1987). Rising dose tolerance (RDT) study: A novel scenario for obtaining preclinical toxicology/drug metabolism data. *Toxicologist* 7, Abstract #846.

Hopper, D. (1986). Behavioral measures in toxicology screening. In W. Lloyd (Ed.), *Safety Evaluation of Drugs and Chemicals*. Hemisphere Publishing, New York, pp. 305–321.

Interagency Regulatory Liaison Group (IRLG), Office of Consumer Affairs (1981). *Testing Standards and Guidelines Work Group (HFE-88)*. IRLG, Washington, DC.

Irwin, S. (1962). Drug screening and evaluative procedures. *Science* 136:123–128.

Jenner, P., Hagan, E., Taylor, J., Cook, E., and Fitzhugh, O. (1964). Food flavorings and compounds of related structure. I. Acute oral toxicity. *Fd. Cosmet. Toxicol.* 2:327–343.

Karber, G. (1931). Contribution to the collective treatment of pharmacological serial experiments. *Arch. Exp. Path. Pharmacol.* 162:480–483.

LeBeau, J. (1983). The role of the $LD_{50}$ determination in drug safety evaluation. *Reg. Tox. Pharmacol.* 3:71–74.

Litchfield, J., and Wilcoxon, F. (1949). A simplified method of evaluating dose-effect experiments. *J. Pharmacol. Exp. Ther.* 96:99–113.

Lörke, D. (1983). A new approach to practical acute toxicity testing. *Arch. Toxicol.* 54:275–287.

Malmfors, T., and Teiling, A. (1983). $LD_{50}$—Its value for the pharmaceutical industry in safety evaluation of drugs. *Acta Pharmacol. Toxicol.* 52(Suppl. 2):229–246.

McClain, R. (1983). Generating, interpreting and reporting information in toxicity studies. *Drug Info. J.* 17:245–255.

Morrison, J., Quinton, R., and Reinert, H. (1968). The purpose and value of $LD_{50}$ determinations. In E. Boyland and R. Goulding (Eds.), *Modern Trends in Toxicology*, Vol. I. Appleton-Century-Crofts, London, pp. 1–17.

Muller, H., and Kley, H. (1982). Retrospective study on the reliability of an "approximate $LD_{50}$" determined with a small number of animals. *Arch. Toxicol.* 51:189–196.

Office of Science Coordination, U.S. Food and Drug Administration (1984). *Final Report on Acute Studies Workshop Sponsored by the U.S. Food and Drug Administration*. U.S. Government Printing Office, Washington, DC.

Olson, H., Fabian, R., Greener, Y., Pack, F., Zelinger, D., and Dean, J. (1990). Reduction in animals used for acute toxicity testing based on retrospective analysis. *Toxicologist* 141 (Abstract).

Organization for Economic Co-operation and Development (OECD) (1991). *OECD Guidelines for Testing of Chemicals*, No. 401: Acute Oral Toxicity. OECD, Paris.

Osterberg, R. (1983). Today's requirements in food, drug, and chemical control. *Acta Pharmacol. Toxicol.* 52(Suppl. 2):201–228.

Pendergast, W. (1984). Biological drug regulation. In *The Seventy-Fifth Anniversary Commemorative Volume of Food and Drug Law*. Edited and published by the Food and Drug Law Institute, Washington, DC, pp. 293–305.

Peterson, D., Ellis, Z., Harris, D., and Spadek, Z. (1987). Acute toxicity of the major alkaloids of cultivated *Lupinus angustifolius* seeds to rats. *J. Appl. Toxicol.* 7:51–53.

Piegorsh, W. (1989). Quantification of toxic response and the development of the median effective dose ($ED_{50}$)—a historical perspective. *Toxicol. Indust. Health* 5:55–62.

Rhodes, C. (2000). Principles of Testing for Acute Toxic Effects. In B. Ballantyne, T. Marrs, and T. Szversen (Eds.), *General and Applied Toxicology*. Macmillan References, London, pp. 33–54.

Rowan, A. (Apr. 1981). The $LD_{50}$ test: A critique and suggestions for alternatives. *Pharm. Technol.*, pp. 65–92.

Schiantarelli, P., and Cadel, S. (1981). Piroxicam pharmacological activity and gastrointestinal damage by oral and rectal route: Comparison with oral indomethacin and phenylbutazone. *Arzneim. Forsch./Drug Res.* 31:87–92.

Schultz, E., and Fuchs, H. (1982). A new approach to minimizing the number of animals in acute toxicity testing and optimizing the information of the test results. *Arch. Toxicol.* 51:197–220.

Smyth, H., and Carpenter, C. (1944). The place of the range-finding test in the industrial toxicology laboratory. *J. Indust. Hyg. Toxicol.* 26:269–273.

Sperling, F. (1976). Nonlethal parameters as indices of acute toxicity: Inadequacies of the acute $LD_{50}$. In M. Mehlman, R. Shapiro, and H. Blumenthal (Eds.), *Advances in Modern Toxicology, Vol. 1, Part 1: New Concepts in Safety Evaluation*. Hemisphere Publishing, Washington, DC, pp. 177–191.

Thompson, W., and Weil, C. (1952). On the construction of a table for moving average interpolation. *Biometrics* 8:51–54.

Trevan, J. (1927). The error of determination of toxicity. *Proc. Roy. Soc.* B101:483–514.

U.S. Food and Drug Administration (FDA) (1996). *Guidance for industry: Single dose acute toxicity testing for pharmaceuticals*. FDA, Washington, DC.

U.S. Food and Drug Administration (FDA) (2006). *Guidance for industry investigations and reviewers: Exploratory IND studies*. FDA, Washington, DC.

van den Heuvel, M., Clark, D., Fielder, R., Koundakijian, P., Oliver, G., Pelling, D., Tomlinson W., and Walker, A. (1990). The international validation of a fixed-dose procedure as an alternative to the classical $LD_{50}$ test. *Fd. Chem. Toxicol.* 28:469–482.

Waud, D. (1972). Biological assays involving quantal responses. *J. Pharmacol. Exp. Ther.* 183:577–607.

Weil, C. (1952). Table for convenient calculation of median effective dose ($LD_{50}$ or $ED_{50}$) and instructions in their use. *Biometrics* 8:249–263.

Weil, C. (1983). Economical $LD_{50}$ and slope determination. *Drug Chem. Toxicol.* 6:595–603.

Zbinden, G. (1984). Acute toxicity testing, purpose. In A. Goldberg (Ed.), *Acute Toxicity Testing: Alternative Approaches*. Mary Ann Liebert, New York, pp. 5–22.

Zbinden, G., Elsner, J., and Boelsterli, U. (1984). Toxicological screening. *Regul. Toxicol. Pharmacol.* 4:275–286.

Zbinden, G., and Flury-Roversi, M. (1981). Significance of the $LD_{50}$ test for the toxicological evaluation of chemical substances. *Arch. Toxicol.* 47:77–99.

# *Genotoxicity*

Genotoxicity encompasses all the potential means by which the genetic material of higher organisms may be damaged, with resulting serious consequences. Most forms of genotoxicity are expressions of mutagenicity—the induction of deoxyribonucleic acid (DNA) damage and other genetic alterations, with changes in one or a few DNA bare pairs (gene mutations). Others are clastogenicity, with gross changes in chromosomal structure (i.e., chromosomal aberrations) or in chromosome numbers. Clearly the potential of any pharmaceutical to cause such damage is a concern.

It has been known for several hundred years that exposure to particular chemicals or complex mixtures can lead to cancer in later life (Doll, 1977), and it has been postulated more recently that chemicals can also induce heritable changes in humans, leading to diseases in the next generation (ICEMC, 1983). There has been accumulating evidence that such changes can arise following damage to DNA and resulting mutations (see, e.g., Bridges, 1976). Therefore, it has become necessary to determine whether widely used drugs or potentially useful new drugs possess the ability to damage DNA. In pharmaceutical development, such information may be used to discard a new candidate drug from further work, to control or eliminate human exposure for a mutagenic industrial compound, or, for a drug, to proceed with development if benefits clearly outweigh risks. Data concerning the genotoxicity of a new drug have become part of the safety package, though the timing of the performance of the tests may vary. They are needed for decision making and to reduce risks that might otherwise be unforeseen.

---

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

**TABLE 1 Genotoxicity tests recommended by ICH (ICH, 1996)**

| Genotoxicity Test—ICH | Mutation | Cell Type | Method |
|---|---|---|---|
| A test for gene mutation in bacteria | gene | bacterial | in vitro |
| In vitro cytogenetic assay using mouse lymphomas tk cells | chromosome | mammalian | in vitro |
| In vivo test for chromosomal damage using rodent hematopoietic cells | gene | mammalian | in vivo |

*Note*: The International Conference on Harmonization recommends a specific profile of genotoxicity tests for drugs, which is different from those of OECD and ISO. They want to see an in vivo test conducted.

The International Conference on Organisation (ICH) S2B clear guidance in testing requirements is summarized in Table 1. Currently (2005) an early stage (step 2) revision S2 is under consideration. This revision, if adapted, would fundamentally change many aspects of how candidate drugs are evaluated for potential genotoxicity. The U.S. Food and Drug Administration (FDA) follows the operative ICH guidance.

The ICH recommends a rather different profile of genotoxicity tests for drugs. It wants to see an in vivo test conducted.

## 1 DNA STRUCTURE

With the exception of certain viruses, the blueprint for all other organisms is contained in code by DNA, a giant macromolecule whose structure allows a vast amount of information to be stored accurately. We have all arisen from a single cell, the fertilized ovum containing two sets of DNA (packaged with protein to form chromatin), one set from our mother, resident in the nucleus of the unfertilized ovum, the second set from our father via the successful sperm. Every cell in the adult has arisen from this one cell and (with the exception of the germ cell and specialized liver cells) contains one copy of these original chromosome sets.

The genetic code is composed of four "letters"—two pyrimidine nitrogenous bases, thymine and cytosine, and two purine bases, guanine and adenine—which can be regarded functionally as arranged in codons (or triplets). Each codon consists of a combination of three letters; therefore, $4^3$ (64) different codons are possible. Sixty-one codons code for specific amino acids (three produce stop signals), and as only 20 different amino acids are used to make proteins, one amino acid can be specified by more than one codon. The bases on one strand are connected together by a sugar (deoxyribose) phosphate backbone. DNA can exist in single- or double-stranded form. In the latter state, the two strands are held together by hydrogen bonds between the bases. Hydrogen bonds are weak electrostatic forces involving oxygen and nitrogen atoms. As a strict rule, one fundamental to mutagenesis, the adenine bases on one strand always hydrogen bond to the thymine bases on the sister

strand. Similarly, guanine bases pair with cytosine bases. Adenine and thymine form two hydrogen bonds, and guanine and cytosine form three.

Double-stranded DNA has a unique property in that it is able to make identical copies of itself when supplied with precursors, relevant enzymes, and cofactors. In simplified terms, two strands begin to unwind and separate as the hydrogen bonds are broken. This produces single-stranded regions. Complementary deoxyribonucleotide triphosphates then pair with the exposed bases under the control of a DNA polymerase enzyme.

A structural gene is a linear sequence of codons which codes for a functional polypeptide, that is, a linear sequence of amino acids. Individual polypeptides may have a structural, enzymatic, or regulatory role in the cell. Although the primary structure of DNA is the same in prokaryotes and eukaryotes, there are differences between the genes of these two types of organism, in internal structure, number, and mechanism of replication. In bacteria, there is a single chromosome, normally a closed circle, which is not complexed with protein, and replication does not require specialized cellular structures. In plant and animal cells, there are many chromosomes, each present as two copies, as mentioned earlier, and the DNA is complexed with protein. Replication and cell division require the proteinaceous spindle apparatus. The DNA of eukaryotic cells contains repeated sequences of some genes. Also, eukaryotic genes, unlike prokaryotic genes, have noncoding DNA regions called introns between coding regions called exons. This property means that eukaryotic cells have to use an additional processing step at transcription.

### 1.1 Transcription

The relationship between the DNA in the nucleus and proteins in the cytoplasm is not direct. The information in the DNA molecule is transmitted to the protein-synthesizing machinery of the cell via another informational nucleic acid, called messenger RNA (mRNA), which is synthesized by an enzyme called RNA polymerase. Although similar to DNA, mRNAs are single stranded and possess the base uracil instead of thymine and the sugar ribose rather than deoxyribose. These molecules act as short-lived copies of the genes being expressed.

In eukaryotic cells, the initial mRNA copy contains homologues of both the intron and exon regions. The intron regions are then removed by enzymes located in the nucleus of the cell. Further enzymes splice the exon regions together to form the active mRNA molecules. In both groups of organisms mature mRNA molecules then pass out of the nucleus into the cytoplasm.

### 1.2 Translation

The next process is similar in both eukaryotes and prokaryotes and involves the translation of mRNA molecules into polypeptides. This procedure involves

many enzymes and two further types of RNA: transfer RNA (tRNA) and ribosomal RNA (rRNA). There is a specific tRNA for each of the amino acids. These molecules are involved in the transportation and coupling of amino acids into the resulting polypeptide. Each tRNA molecule has two binding sites, one for the specific amino acid, the other containing a triplet of bases (the "anticodon") which is complementary to the appropriate codon on the mRNA.

Ribosomal RNA is complexed with protein to form a subcellular globular organelle called a ribosome. Ribosomes can be regarded as the "reading head", which allows the linear array of mRNA codons each to base pair with an anticodon of an appropriate incoming tRNA–amino acid complex. The polypeptide chain forms as each tRNA–amino acid comes in to register with the RNA codon and with specific sites on the ribosome. A peptide bond is formed between each amino acid as it passes through the reading head of the ribosome (Venitt and Parry, 1984).

## 1.3 Gene Regulation

Structural genes are regulated by a special set of codons, in particular "promoter" sequences. The promoter sequence is the initial binding site for RNA polymerase before transcription begins. Different promoter sequences have different affinities for RNA polymerase. Some sets of structural genes with linked functions have a single promoter and their coordinate expression is controlled by another regulatory gene called an operator. A group of such genes is called an operon. The activity of the operator is further controlled by a protein called a repressor, since it stops the expression of the whole operon by binding to the operator sequence, preventing RNA polymerase from binding to the promoter. Repressors can be removed by relevant chemical signals or in a time-related fashion.

In the ways described above, only the genes required at a given moment are expressed. This not only helps to conserve the energy of the cell but also is critical for correct cellular differentiation, tissue pattern formation, and formation of the body plan.

## 1.4 DNA Repair

All living cells appear to possess several different major DNA repair processes [for reviews see Walker (1984) and Rossman and Klein (1988)]. Such processes are needed to protect cells from the lethal and mutating effects of heat-induced DNA hydrolysis, ultraviolet light, ionizing radiation, DNA-reactive chemicals, free radicals, and so on. In single-celled eukaryotes such as the yeast *Saccharomyces cerevisiae*, the number of genes known to be involved in DNA repair approaches 100 (Friedberg, 1988). The number in mammalian cells is expected to be at least equal to this and emphasizes the importance of correction of DNA damage.

***Excision Repair*** Some groups of enzymes (light independent) are apparently organized to act cooperatively to recognize DNA lesions, remove them, and correctly replace the damaged sections of DNA. The most comprehensively studied of these is the excision repair pathway.

Briefly, the pathway can be described as follows:

1. *Preincision Reactions* UvrA protein dimers are formed which bind to the DNA at a location distant from the damaged site. The UvrB protein then binds to the DNA–UvrA complex to produce an energy-requiring topological unwinding of the DNA via DNA gyrase. This area of unwinding is then translocated, again using adenosine triphosphate (ATP) as an energy source, to the site of the damaged DNA.
2. *Incision Reactions* The UvrC protein binds to the DNA–UvrA,B complex and incises DNA at two sites—seven bases to the 5′ end and three bases to the 3′ end of the damage.
3. *Excision Reactions* UvrD protein and DNA polymerase 1 excise the damaged bases and then resynthesize the strand using the sister strand as a template. The Uvr complex then breaks down, leaving a restored, but nicked, strand.
4. *Ligation Reaction* The nick in the phosphate backbone is repaired by DNA ligase. A similar excision repair mechanism exists in mammalian cells (see, e.g., Cleaver, 1983). In both cases, the process is regarded as error free and does not lead to the generation of mutations. However, this pathway can become saturated with excessive numbers of damaged DNA sites, forcing the cell to fall back on other repair mechanisms.

## 1.5 Error-Prone Repair

Exposure of *Escherichia coli* to agents or conditions that either damage DNA or interfere with DNA replication results in the increased expression of the so-called SOS regulatory network (Walker, 1984). Included in this network is a group of at least 17 unlinked DNA damage-inducible (*din*) genes. The *din* gene functions are repressed in undamaged cells by the product of the *lexA* gene (Little and Mount, 1982) and are induced when the LexA protein is cleaved by a process that requires modified RecA protein (RecA*), which then acts as a selective protease (Little, 1984). The *din* genes code for a variety of functions, including filamentation and cessation of respiration. Included are the *umuDC* gene products, which are required for so-called error-prone or mutagenic DNA repair (Kato and Shinoura, 1977). The precise biochemical mechanism by which this repair is achieved is still not fully understood. Bacterial polymerase molecules have complex activities, including the ability to "proofread" DNA—that is, to ensure that the base-pairing rules of double-stranded DNA are met. It is hypothesized that Umu proteins may suppress this proofreading activity, so that base mismatches are tolerated (Villani et al.,

1978). Recent evidence suggests that DNA lesions are bypassed, and this bypass step required UmuDC proteins and RecA* protein (Bridges et al., 1987). The net result is that random base insertion occurs opposite the lesion which may result in mutation.

Analogues of the *umuDC* genes can be found in locations other than the bacterial chromosome—for example, plasmid pKM101 (Walker and Dobson, 1979), a derivative of the drug resistance plasmid R46 (Mortelmans and Strocker, 1979), which carried *mucAB* genes (Shanabruch and Walker, 1980; see pp. 879–880). Mutagenic repair, as controlled by *umuDC*, is not universal even among enterobacteria (Sedgwick and Goodwin, 1985). For instance, *Salmonella typhimurium* LT2 does not appear to express mutagenic repair (Walker, 1984). Thus, the usefulness of strains of this species is greatly enhanced by using derivatives containing plasmids with genes coding for error-prone repair (MacPhee, 1973).

## 1.6 Mismatch Repair

Mismatched pairs that break the normal base-pairing rules can arise spontaneously due to DNA biosynthetic errors, events associated with genetic recombination, and the deamination of methylated cytosine (Modrich, 1987). With the latter, when cytosine deaminates to uracil, an endonuclease enzyme, *N*-uracil-DNA glycosylase (Lindahl, 1979), excises the uracil residue before it can pair with adenine at the next replication. However, 5-methyl cytosine deaminates to form thymine and will not be excised by a glycosylase. As a result, thymine exits on one strand paired with guanine on the sister strand, that is, a mismatch. This will result in a spontaneous point mutation if left unrepaired. For this reason, methylated cytosines form spontaneous mutation "hot spots" (Miller, 1985). The cell is able to repair mismatches by being able to distinguish between the DNA strand that exists before replication and a newly synthesized strand.

The mechanism of strand-directed mismatch correction has been demonstrated in *E. coli* (see, e.g., Wagner and Meselson, 1976). In this organism, adenine methylation of d(G–A–T–C) sequences determines the strand on which repair occurs. Thus, parental DNA is fully methylated, while newly synthesized DNA is undermethylated, for a period sufficient for mismatch correction. By this means the organism preserves the presumed correct sequence—that is, that present on the original DNA strand—and removes the aberrant base on the newly synthesized strand. Adenine methylation is achieved in *E. coli* by the *dam* methylase, which is dependent on *S*-adenosylmethionine. Mutants (*dam*) lacking this methylase are hypermutable, as would be expected by this model (Marinus and Morris, 1974).

## 1.7 Adaptive Repair Pathway

The mutagenic and carcinogenic effects of alkylating agents such as ethyl methane sulfonate are due to the generation of $O^6$-alkylguanine residues in

DNA, which result in point mutations. Bacterial and mammalian cells can repair a limited number of such lesions before DNA replication, thus preventing mutagenic and potentially lethal events taking place.

If *E. coli* are exposed to low concentrations of a simple alkylating agent, a repair mechanism is induced that causes increased resistance to subsequent challenge with a high dose. This adaptation response was first described by Samson and Cairns (1977) and has recently been reviewed by Lindahl et al. (1988). The repair pathway is particularly well understood.

## 1.8 Plasmids

Plasmids are extrachromosomal genetic elements that are composed of circular double-stranded DNA. In bacteria some can mediate their own transfer from cell to cell by conjugation—that is, they contain a set of *tra* genes coding for tubelike structures, such as pili, through which a copy of plasmid DNA can pass during transfer.

Plasmids range in size from 1.5 to 200 million daltons. The number of copies per cell differs from plasmid to plasmid. Copy number relates to control of replication and this correlates with size—that is, small plasmids tend to have large copy numbers per cell. This may relate to a lack of replication control genes (Mortelmanns and Dousman, 1986).

## 1.9 Plasmids and DNA Repair

Many plasmids are known to possess three properties: (1) increased resistance to the bactericidal effects of UV and chemical mutagens, (2) increased spontaneous mutagenesis, and (3) increased susceptibility to UV and chemically induced mutagenesis. Some plasmids possess all three properties; others may possess just one, for example, increased susceptibility to mutagenesis [for a review see Mortelmanns and Dousman (1986)]. Often the profile of activity depends on the DNA repair status of the host cell (Pinney, 1980). Plasmid pKM101 carries DNA repair genes and has been widely in strains in bacterial mutagenicity tests.

## 1.10 Nature of Point Mutations

The word "mutation" can be applied to point mutations, which are qualitative changes involving one or a few bases in base sequences within genes, as described below, as well as to larger changes involving whole chromosomes (and thus many thousands of genes), and even to changes in whole chromosome sets (described in Section 2).

Point mutations can occur when one base is substituted for another (base substitution). Substitution of another purine for a purine base or of another pyrimidine for a pyrimidine base is called a transition, while substitutions of purine for pyrimidine or pyrimidine for purine are called transversions. Both types of base substitution have been identified within mutated genes. These

changes lead to a codon change which can cause the "wrong" amino acid to be inserted into the relevant polypeptide and are known as missense mutations. Such polypeptides may have dramatically altered properties if the new amino acid is close to the active center of an enzyme or affects the three-dimensional makeup of an enzyme or a structural protein. These changes in turn can lead to change or reduction in function, which can be detected as a change in phenotype of the affected cells.

A base substitution can also result in the formation of a new inappropriate terminator (or nonsense) codon and is thus known as nonsense mutation. The polypeptide formed from such mutated genes will be shorter than normal and is most likely to be inactive. Owing to the redundancy of the genetic code, about a quarter of all possible base substitutions will not result in an amino acid replacement and will be silent mutations.

Bases can be deleted or added to a gene. As each gene is of a precisely defined length, these changes, if they involve a number of bases that is not a multiple of 3, result in a change in the "reading frame" of the DNA sequence and are thus known as frameshift mutations. Such mutations tend to have a dramatic effect on the polypeptide of the affected gene, as most amino acids will differ from the point of the insertion or deletion of bases onward. Very often a new terminator codon is produced, so, again, short inactive polypeptides will result.

Both types of mutation result in an altered polypeptide, which in turn can have a marked effect on the phenotype of the affected cell. Much use of phenotypic changes is made in mutagenicity tests.

Base substitutions and frameshift changes occur spontaneously and can be induced by radiations and chemical mutagens. It is apparent that the molecular mechanisms resulting in these changes are different in each case, but the potential hazards associated with mutagens capable of inducing the different types of mutation are equivalent.

## 1.11 Suppressor Mutations

In some instances a mutation within one gene can be corrected by a second mutational event at a separate site on the chromosome. As a result, the first defect is suppressed and the second mutation is known as a suppressor mutation. Most suppressor mutations have been found to affect genes encoding for RNAs. Usually the mutation causes a change in the sequence of the anticodon of the tRNA. Thus, if a new terminator or nonsense codon is formed as the first mutation, this can be suppressed by a second mutation forming a tRNA species that now has an anticodon complementary to a termination codon. Thus, the new tRNA species will supply an amino acid at the terminator site on the mRNA and allow translation to proceed. Surprisingly most suppressors of this type do not adversely affect cell growth, which implies that the cell can tolerate translation proceeding through termination signals, producing abnormal polypeptides. An alternative explanation is that the particular DNA

sequences surrounding normal terminator codons result in a reduced efficiency of suppressor tRNAs (Bossi, 1985).

Frameshift suppression is also possible. This can be achieved by a second mutation in a tRNA gene such that the anticodon of a tRNA molecule consists of four bases rather than three—for example, an extra C residue in the CCC anticodon sequence of a glycine tRNA gene. This change will allow correction of a +1 frameshift involving the GGG codon for glycine (Bossi, 1985).

## 1.12 Adduct Formation

The earlier discussion of adaptive repair made reference to the fact that some unrepaired alkylated bases are lethal, owing to interference with DNA replication, while others, such as $O^6$-methylguanine, lead to mutation if unrepaired. These differences indicate that not all DNA adducts (i.e., DNA bases with additional chemical groups, not associated with normal DNA physiology) are equivalent. In fact, some adducts appear not to interfere with normal DNA functions or are rapidly repaired; others are mutagenic and yet others are lethal. Chemicals that form electrophilic species readily form DNA adducts. These pieces of information are hard won, and the reader is recommended to read reviews of the pioneering work of Brooks and Lawley (see Lawley, 1989) summarizing work identifying the importance of DNA adduct formation with polycyclic hydrocarbons and the importance of "minor" products of base alkylation such as $O^6$-methylguanine and, in addition, the work of the Millers in linking attack of nucleophilic sites in DNA by electrophiles to mutagenesis and carcinogenesis (Miller and Miller, 1971).

If a DNA adduct involves the nitrogen or oxygen atoms involved in base pairing and the adducted DNA is not repaired, base substitution can result. Adducts can be small, such as the simple addition of methyl or ethyl groups, or they can be very bulky, owing to reaction with multiringed structures. The most vulnerable base is guanine, which can form adducts at several of its atoms (e.g., $N^7$, $C^8$, $O^6$, and exocyclic $N^2$) (Venitt and Parry, 1984). Adducts can form links between adjacent bases on the same strand (intrastrand crosslinks) and can form interstrand crosslinks between each strand of double-stranded DNA.

The induction of frameshift mutation does not necessarily require covalent adduct formation. Some compounds that have a flat, planar structure, such as particular polycyclic hydrocarbons, can intercalate between the DNA strands of the DNA duplex. The intercalated molecules interfere with DNA repair enzymes or replication and cause additions and deletions of base pairs. The precise mechanism is still unclear, although several mechanisms have been proposed. Hot spots for frameshift mutation often involve sections of DNA where there is a run of the same base—for example, the addition of a guanine to a run of six guanine residues. Such information led to a "slipped mispairing" model for frameshift mutation (Streisinger et al., 1966; Roth, 1974). In this scheme single-strand breaks allow one strand to slip and loop out one or more base pairs, the configuration being stabilized by complementary base pairing

at the end of the single-stranded region. Subsequent resynthesis results ultimately in additions or deletions of base pairs (Miller, 1985).

## 1.13 Mutations Due to Insertion Sequences

The subject of mutations due to insertion sequences is reviewed in Cullum (1985). Studies of spontaneous mutation in *E. coli* detected a special class of mutations that were strongly polar, reducing the expression of downstream genes (Jordan et al., 1967). These genes mapped as point mutations and reverted like typical point mutations. However, unlike point mutations, mutagens did not increase their reversion frequency. Further studies showed that these mutations were due to extra pieces of DNA that can be inserted into various places in the genome. They are not just random pieces of DNA but are "insertion sequences" 0.7–1.5 kb long that can "jump" into other DNA sequences. They are related to transposons, which are insertion sequences carrying easily detected markers such as antibiotic resistance genes, and Mu phages (bacterial viruses).

## 1.14 Link between Mutation and Cancer

The change in cells undergoing normal, controlled cell division and differentiation to cells that are transformed, dividing without check, and are undifferentiated or abnormally differentiated does not appear to occur as a single step—that is, transformation is multistage. Evidence for this comes from in vitro studies, animal models, and clinical observations—in particular, the long latent period between exposure to a carcinogen and the appearance of a tumor in the target tissue. There is much evidence for the sequence of events shown in Figure 1 (tumor initiation, promotion, malignant conversion, and progression). Such a scheme provides a useful working model but clearly does not apply to all "carcinogens" in all circumstances.

Study of Figure 1 shows that there are several points where genetic change appears to play a role. Such change may occur spontaneously, due to rare errors at cell division such as misreplication of DNA or spindle malfunction, or may be induced by exposure to viruses (e.g., acute transforming retroviruses), ionizing and nonionizing radiations absorbed by DNA (e.g., X-rays, UVC), or particular chemical species capable of covalently interacting with DNA (as discussed earlier) or with vital proteins, such as tubulin, that polymerize to form the cell division spindle apparatus.

## 1.15 Genetoxic versus Nongenotoxic Mechanisms of Carcinogenesis

The previous discussions of oncogene activation and human DNA repair deficiencies provide strong evidence for carcinogenesis via genotoxic mechanisms. However, it has been recognized for many years that cancers can arise

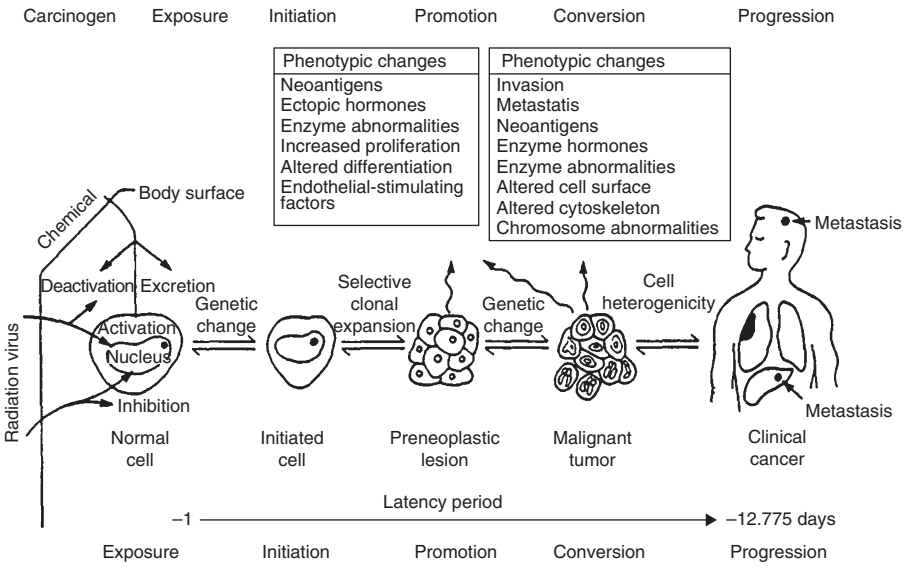


***Figure 1*** *Schematic representation of events leading to neoplasia. [Source: Adapted from Harris et al. (1987).]*

without biologically significant direct or indirect interaction between a chemical and cellular DNA (see, e.g., Gatehouse et al., 1988). The distinction between nongenotoxic and genotoxic carcinogens has recently been brought into a sharper focus following the identification of a comparatively large number of "nongenotoxic" carcinogens by the United States National Toxicology Program (Tennant et al., 1987). These include a wide range of chemicals acting via a variety of mechanisms, including augmentation of high "spontaneous" tumor yields; disruption of normal hormonal homeostasis in hormone-responsive tissues; peroxisome proliferation; and proliferation of urothelial cells following damage via induced kidney stones (Clayson, 1989). This author points out that a major effort is under way to determine whether many of these compounds can elicit similar effects in humans.

Ashby and Tennant (1988) and Ashby et al. (1989) stress the significance of their observations that 16 tissues are apparently sensitive to genotoxic carcinogens, while a further 13 tissues are sensitive to both genotoxic and nongenotoxic carcinogens (Table 2). Also, genotoxic carcinogens tend to induce tumors in several tissues of both males and females in both rats and mice. This contrasts with nongenotoxic carcinogens, which may induce tumors at high doses, in one tissue, of one sex, of one species. Although it is most unlikely that all nongenotoxic carcinogens will prove to be irrelevant in terms of human risk, it appears from the analysis above that a proportion of carcinogens identified by the use of near-toxic levels in rodent bioassays are of dubious relevance to the induction of human cancer. For further discussion, see Butterworth and Slaga (1987).

**TABLE 2 Tissues Sensitive to Genotoxic and/or Nongenotoxic Carcinogens**

| Tissues Sensitive Primarily to Genotoxins | Tissues Sensitive to Both Genotoxins and Nongenotoxins |
|---|---|
| Stomach | Nose |
| Zymbal gland | Mammary gland |
| Lung | Pituitary gland |
| Subcutaneous tissue | Integumentary system |
| Circulatory system | Kidney |
| Clitoral gland | Urinary bladder |
| Skin | Liver |
| Intestine/colon | Thyroid gland |
| Uterus | Hematopoietic system |
| Spleen | Adrenal gland |
| Tunica vaginalis | Pancreas |
| Bile duct | Seminal vesicle |
| Ovary | Urinary tract |
| Haderian gland | Lymphatic system |
| Preputial gland | |
| Multiple organ sites | |

## 1.16 Genetic Damage and Heritable Defects

Concern about the effects of radiations and chemicals on the human gene pool and the resulting heritable malformations and syndromes have steadily risen during this century. The recognition that changes in morphology would result from changes in the hereditary material due to mutations (from the Latin word *mutare*, "to change") was adopted by de Vries following observations on the evening primrose, *Oenothera* (deVries, 1901). Muller went on to demonstrate that X-rays could induce mutations in the germ cells of the fruit fly *Drosophila melanogaster* (Muller, 1927).

The human gene pool is known to carry many deleterious genes acquired from preceding generations which result in numerous genetic disease. It is clear that these arise as a result of DNA changes affecting particular chromosomes or genes. They can be grouped as follows:

1. Chromosome abnormalities, small changes in either number or structure
2. Autosomal-dominant gene mutations, in which a change in only one copy of the pair of genes is sufficient for the condition to be expressed
3. Autosomal-recessive gene mutations in which both copies of a gene must be mutated for the trait to become manifest
4. Sex-linked conditions, which may also be recessive or dominant, where the mutant gene is on an X chromosome and will be expressed at high frequency in males (XY) and at a much lower frequency in females (XX) if the gene acts in a recessive manner

5. Polygenic mutations, in which the condition results from the interaction of several genes and may include an environmental component

### 1.17 Reproductive Effects

If a potent genotoxin is able to cross the placental barrier, it is very likely to interfere with differentiation of the developing embryo and thus possess teratogenic potential. Indeed, many of the better studied teratogens are also mutagenic (Kalter, 1977). However, mutagens form only one class of teratogens and a large proportion of teratogens are not mutagenic. Alternative mechanisms of teratogenesis include cell death, mitotic delay, retarded differentiation, vascular insufficiency, and inhibited cell migration (Beckman and Brent, 1986).

It is known that more fetal wastage and many spontaneous abortions arise as a result of the presence of dominant lethal mutations in the developing embryo, many of which appear to be due to major chromosomal damage. In addition, impairment of male fertility may also be a consequence of exposure to mutagens.

## 2 CYTOGENETICS

There are various types of cytogenetic change which can be detected in chromosomes. These are structural chromosome aberrations (Cas), numerical changes which could result in aneuploidy, and sister chromatid exchanges (SCEs). Chromosome aberration assays are used to detect the induction of chromosome breakage (clastogenesis) in somatic or germinal cells by direct observation of chromosome damage during metaphase analysis or by indirect observation of micronuclei. Chromosome damage detected in these assays is mostly lethal to the cell during the cell cycle following the induction of the damage. Its presence, however, indicates a potential to induce more subtle chromosome damage which survives cell division to produce heritable cytogenetic changes. Cytogenetic damage is usually accompanied by other genotoxic damage such as gene mutation.

### 2.1 Cytogenetic Damage and Its Consequences

Structural and numerical chromosomal aberrations in somatic cells may be involved in the etiology of neoplasia and in germ cells can lead to perinatal mortality, dominant lethality or congenital malformations in the offspring (Chandley, 1981), and some tumors (Anderson, 1990).

Chromosome defects arise at the level of the individual chromosome or at the level of the chromosomal set, so affecting chromosomal number.

## 2.2 Individual Chromosome Damage

Damage to individual chromosomes consists of breakage of chromatids, which must result from a discontinuity of both strands of the DNA in a chromatid. How mutagens produce chromosome breakage is not totally understood, but DNA lesions which are not in themselves discontinuities will produce breakage of a chromosome as a consequence of their interference with the normal process of DNA replication. In haploid microorganisms and prokaryotes chromosome breaks are usually lethal, but not in diploid eukaryotes. According to Bender et al. (1974), in these organisms chromosome breaks may reconstitute in the same order, probably as a result of an enzyme repair process, resulting in no apparent cytogenetic damage; they may remain unjoined as fragments, which could result in cell death at the next or following mitoses—if, for example, unrejoined fragments are introduced into the zygote via treated germ cells, the embryo may die at a very early stage from a dominant lethal mutation or they may rejoin in a different order form the original one, producing chromosomal rearrangements. There are various types of chromosomal rearrangements:

*Reciprocal translocations* can result from the exchange of chromosomal segments between two chromosomes and, depending on the position of the centromeres in the rearranged chromosomes, different configurations will result.

1. Asymmetrical exchanges arise when one of the rearranged chromosomes carries both centromeres and is known as dicentric while the other carries none and is acentric. The cell or zygote carrying this anomaly usually dies, death being caused by segregation difficulties of the dicentric or the loss of the acentric fragment at cell division. Such a translocation contributes to dominant lethality.
2. Symmetrical exchanges occur when each rearranged chromosome carries just one centromere. This allows the zygote to develop normally, but when such heterozygotes form germ cells at meiosis, about half of their gametes will be genetically unbalanced, since they have deficiencies and duplications of chromosomal material. The unbalanced gametes which survive produce unbalanced zygotes, which results in death shortly before and after birth, or congenital malformations.

*Centric fusions* involve the joining together of two chromosomes, each of which has a centromere at or near one end, to produce a single metacentric or submetacentric chromosome. When such translocations are produced in a germ cell and result from breakage and rejoining in the short arms of the two chromosomes, as a consequence of loss of the derived acentric fragments, a genetic deficiency can result. Some Robertsonian translocations are able to survive but others pose a risk. In heterozygotes the two arms of the translocation chromosome may pair with the two separate homologous chromosomes

at meiosis but segregate in a disorderly manner. Some of the resultant germ cells lack copies (nullisomy) or carry two copies (disomy) of one or other of the two chromosomes involved, which results in monosomic or trisomic embryos. Monosomics die early but trisomic embryos, which carry three copies of a chromosome, can survive to birth or beyond. If chromosome 21 is involved in the translocation, it can form a translocation trisomy and produce inherited Down syndrome (this differs from nondisjunctional Down syndrome trisomy).

*Deletions and deficiencies* are produced when two breaks arise close together in the same chromosome. The two ends of the chromosome join when the fragment between the breaks becomes detached. At the next cell division the unattached piece of chromosome is likely to be lost. Large deletions may contribute to dominant lethality. Small deletions are difficult to distinguish from point mutations. Deletions may uncover preexisting recessive genes. If one gene that is essential for survival is uncovered, it can act as a lethal in a homozygote and as a partial dominant in a heterozygote.

*Inversions* occur when two breaks occur in the same chromosome. The portion between them is detached and becomes reinserted in the opposite way to its original position, that is, the gene order is reversed. This need not cause a genetic problem, but imbalanced gametes could result in congenital malformation or fetal death.

### 2.3 Chromosome Set Damage

Accuracy of chromosome replication and segregation of chromosomes to daughter cells requires accurate maintenance of the chromosome complement of a eukaryotic cell. Chromosome segregation in meiosis and mitosis is dependent upon the synthesis and functioning of the proteins of the spindle apparatus and upon the attachment and movement of chromosomes on the spindle. The kinetochores attach the chromosomes to the spindle and the centrioles are responsible for the polar orientation of the division apparatus. Sometimes such segregation events proceed incorrectly and homologous chromosomes separate, with deviations from the normal number (aneuploidy) into daughter cells or as a multiple of the complete karyotype (polyploidy). When both copies of a particular chromosome move into a daughter cell and the other cell receives none, the event is known as nondisjunction.

Aneuploidy in live births and abortions arises from aneuploid gametes during germ cell meiosis. Trisomy or monosomy of large chromosomes leads to early embryonic death. Trisomy of the smaller chromosomes allows survival but is detrimental to the health of an affected person—for example, Down syndrome (trisomy 21), Patau syndrome (trisomy 13), and Edward syndrome (trisomy 18). Sex chromosome trisomies (Klinefelter's and XXX syndromes) and the sex chromosome monosomy (XO), known as Turner syndrome, are also compatible with survival.

Aneuploidy in somatic cells is involved in the formation of human tumors. Up to 10% of tumors are monosomic and trisomic for a specific chromosome

as the single observable cytogenetic change. Most common among such tumors are trisomy 8, 9, 12, and 21 and monosomy for chromosomes 7, 22, and Y.

## 2.4 Test Systems

In vivo and in vitro techniques are available to test mutagenic properties to demonstrate presence or lack of ability of the test material to cause mutation or chromosomal damage or cause cancer, as summarized in Table 3. The material intended for intimate contact and long exposure should not have any genotoxic properties. The presence of unpolymerized materials and traces of monomers, oligomers, additives, or biodegration products can cause mutations. Mutation can be a point mutation or chromosomal rearrangement caused by DNA damage. Therefore, the material's ability to cause point mutation, chromosomal change, or evidence of DNA damage is tested. As we have seen, correlations exist between mutagenic and carcinogenic properties. Most carcinogens are mutagens, but not all mutagens are human carcinogens.

The Ames salmonella/microsome test is a principal sensitive mutagen screening test. Compounds are tested on the mutants of *S. typhimurium* for reversion from a histidine requirement back to prototrophy. A positive result is seen by the growth of revertant bacteria (which do not require an external

**TABLE 3 Fifteen Common Assays Described by OECD**

| | In Vitro | In Vivo |
|---|---|---|
| *Assays for Gene Mutations* | | |
| *Salmonella typhimurium* reverse mutation assay (Ames test, bacteria) (OECD 471) | ✓ | |
| *Escherichia coli* reverse mutation assay (bacteria) (OECD 472) | ✓ | |
| Gene mutation in mammalian cells in culture (OECD 476) | ✓ | |
| *Drosophila* sex-linked recessive lethal assay (fruit fly) (OECD 477) | | ✓ |
| Gene mutation in *Saccharomyces cerevisiae* (yeast) (OECD 480) | ✓ | |
| Mouse spot test (OECD 484) | | ✓ |
| *Assays for Chromosomal and Genomic Mutations* | | |
| In vitro cytogenetic assay (OECD 473) | ✓ | |
| In vivo cytogenetic assay (OICD 475) | | ✓ |
| Micronucleus test (OECD 474) | | ✓ |
| Dominant lethal assay (OECD 478) | | ✓ |
| Heritable translocation assay (OECD 485) | | ✓ |
| Mammalian germ cell cytogenetic assay (OECD 483) | | ✓ |
| *Assays for DNA Effects* | | |
| DNA damage and repair: unscheduled DNA synthesis in vitro (OECD 482) | ✓ | |
| Mitotic recombination in *S. cerevisiae* (yeast) (OECD 481) | ✓ | |
| In vitro sister chromatid exchange assay (OECD 479) | ✓ | |

histidine source). A microsomal activation system should be included in this assay. The use of five different bacterial test strains are generally required.

Two mammalian mutagenicity tests are generally required to support the lack of mutagenic or carcinogenic potential. Some well known tests are:

- The L5178Y mouse lymphoma assay (MLA) for mutants at the TK locus
- Induction of recessive lethals in *D. melanogaster*
- Metaphase analysis of cultured mammalian cells and of treated animals
- SCE assay
- Unscheduled DNA synthesis (UDS) assay
- Cell transformation assay
- SOS repair system assay
- Gene mutation in cultured mammalian cells such as Chinese hamster V79 cell/hypoxanthine guanine phosphoribosyl transferase (HGPRT) mutation system

The ICH guidelines specifically require three genotoxicity assays for all drugs (see Table 4). The assays should preferably evaluate DNA effects, gene mutations, and chromosomal aberrations, and two of the assays should preferably use mammalian cells. Guidance for providing tests for selection to meet these needs is contained in the Organisation for Economic Co-operation and Devolpment (OECD) guidelines, which include eight in vitro and seven in vivo assays.

## 2.5 In Vitro Test Systems

The principal tests can be broadly categorized into microbial and mammalian cell assays. In both cases the tests are carried out in the presence and

**TABLE 4 Alternative Test under ICH**

| Aspect | Comet Assay | DNA Adducts | UDS Test (Liver) | Transgenic Gene Mutation |
|---|---|---|---|---|
| Test definition (accepted protocol) | Yes | No | Yes | Yes |
| Regulatory acceptance/ use | Yes | Yes | Yes | Yes |
| Relevance of endpoint | Moderate | Moderate | Moderate | High |
| Technical demands | Low moderate | Moderate high | Moderate | High |
| Widespread use | Yes | No | Yes | No |
| Applicable to most tissues | Yes | Yes | No | Yes |
| Dependence of cell turnover | No | No | No | Yes |
| Cost | Low moderate | Moderate high | Low | high |

absence of in vitro metabolic activation enzymes, usually derived from rodent liver.

***In Vitro Metabolic Activation*** The target cells for in vitro mutagenicity tests often possess a limited (often overlooked) capacity for endogenous metabolism of xenobiotics. However, to simulate the complexity of metabolic events that occur in the whole animal, there is a critical need to supplement this activity.

***Choice of Species*** A bewildering variety of exogenous systems have been used for one purpose or another in mutagenicity tests. The choice begins with plant or animal preparations. The attraction of plant systems has stemmed from a desire to avoid the use of animals, where possible, in toxicity testing. In addition, plant systems have particular relevance when certain chemicals are being tested, for example, herbicides.

If animal systems are chosen, preparations derived from fish (see, e.g., Kada, 1981) and birds (Parry et al., 1985) have been used. However, by far the most widely used and validated are those derived from rodents—in particular, the rat. Hamsters may be preferred as a source of metabolizing enzymes when particular chemical classes are being screened—for example, aromatic amines, heterocyclic amines, *N*-nitrosamines, and azo dyes (Prival and Mitchell, 1982; Haworth et al., 1983).

***Choice of Tissue*** The next choice is that of source tissue. Preparations derived from liver are the most useful, as this tissue is a rich source of mixed-function oxygenases capable of converting procarcinogens to genetically active electrophiles. However, many extrahepatic tissues (e.g., kidney, lung) are also known to possess important metabolic capacity which may be relevant to the production of mutagenic metabolites in the whole animal.

***Cell-Free versus Cell-Based Systems*** Most use has been made of cell-free systems—in particular, crude homogenates such as 9000 g supernatant (S9 fraction) from rat liver. This fraction is composed of free endoplasmic reticulum, microsomes (membrane-bound packets of "membrane-associated" enzymes), soluble enzymes, and some cofactors. Hepatic S9 fractions do not necessarily completely reflect the metabolism of the whole organ, in that they mainly possess phase I metabolism (e.g., oxygenases) and are deficient in phase II systems (e.g., conjugation enzymes). The latter are often capable of efficient detoxification, while the former are regarded as "activating." This can be a strength in that S9 fractions are used in screening tests as a surrogate for all tissues in an animals, some of which may be exposed to reactive metabolites in the absence of efficient detoxification. Many carcinogens are organ specific in extrahepatic tissues, yet liver S9 fraction will reveal their mutagenicity. The deficiency of S9 fractions for detoxification can also be a weakness, in that detoxification may predominate in the whole animal, such that the potential carcinogenicity revealed in vitro is not realized in vivo.

Cell-free systems, when supplemented with relevant cofactors, are remarkably proficient, despite their crudity in generating reactive electrophiles from most procarcinogens. However, they provide at best a broad approximation of in vivo metabolism and can fail to produce sufficient quantity of a particular reactive metabolite to be detectable by the indicator cells or they can produce inappropriate metabolites that do not play a role in vivo [for a discussion see Gatehouse and Tweats (1987)].

Some of these problems can be overcome by the use of cell-based systems—in particular, primary hepatocytes. Hepatocytes closely simulate the metabolic systems found in the intact liver and do not require additional cofactors for optimal enzyme activity. However, apart from greater technical difficulties in obtaining hepatocytes as opposed to S9 fraction, hepatocytes can effectively detoxify particular carcinogens and prevent their detection as mutagens. Despite these difficulties, hepatocytes have a role to play in mutagenicity screening in both bacterial and mammalian-based systems (Tweats and Gatehouse, 1988).

***Inducing Agents*** The final choice considered here is whether to use "uninduced" liver preparations or those derived from animals pretreated with an enzyme inducer to promote high levels of metabolic activity. If induced preparations are preferred, which inducer should be used?

It appears that uninduced preparations are of limited use in screening assays, as they are deficient in particular important activities such as cytochrome P-$450_{IA1}$ cytochrome oxygenases. In addition, species and organ differences are most divergent with uninduced enzyme preparations (Brusick, 1987a).

The above differences disappear when induced microsomal preparations are used. A number of enzyme inducers have been used, the most popular being Aroclor 1254, which is a mixture of polychlorinated bipheynyls (as described by Ames et al., 1975). However, concern about the toxicity, carcinogenicity, and persistence of these compounds in the environment has led to the use of alternatives, such as a combination of phenobarbitone (phenobarbital) and β-naphthoflavone (5,6-benzoflavone). This combination results in the induction of a range of monooxygenases similar to that induced by Aroclor 1254 (see, e.g., Ong et al., 1980). More selective inducers such as phenobarbitone (cytochrome P-$450_{IIa1}$, P-$450_{IIB1}$) or 3-methylcholanthrene (cytochrome P-$450_{IA1}$) have also been used.

In summary, genetic toxicity tests with both bacterial and mammalian cells are normally carried out with rat liver cell-free systems (S9 fraction) from animals pretreated with enzyme inducers. However, investigations should not slavishly follow this regimen: There may be sound scientifically based reasons for using preparations from different species or different organs or for using whole cells such as hepatocytes.

***Standard Method of S9 Fraction Preparation*** The following method describes the production of hepatic S9 mix from rats induced with a combina-

tion of phenobarbitone and β-naphthoflavone and is an adaptation of the method described by Gatehouse and Delow (1979).

Male albino rats within the weight range 150–250 g are treated with phenobarbitone sodium 16 mg mL$^{-1}$, 2.5 mL kg$^{-1}$ in sterile saline and β-naphthoflavone 20 mg mL$^{-1}$ in corn oil. A fine suspension of the latter is achieved by sonicating for 1 h. These solutions are dosed by intraperitoneal injection on days 1, 2, and 3.

Phenobarbitone sodium is normally administered between 0.5 and 2 h prior to β-naphthoflavone.

The animals are killed on day 4 by cervical dislocation and the livers removed as quickly as possible and placed on ice-cold KCI buffer (0.01 M $Na_2HPO_4$ + KCl 1.15%). The liver is cleaned, weighted, minced, and homogenized (in an Ultra Turrx homogenizer) in the above buffer to give a 25% (w/v) liver homogenate. The homogenate is stored at 4 °C until it can be centrifuged at 9000*g* for 15 min. The supernatant is decanted, mixed, and divided into 2-mL volumes in cryotubes. These are then snap frozen in liquid nitrogen. Storage at –196 °C for up to three months results in no appreciable loss of most P-450 isoenzymes (Ashwood-Smith, 1980).

Quality control of S9 batches is usually monitored by ability to activate compounds known to require metabolism to generate mutagenic metabolites. This is a rather crude approach and more accurate data can be obtained by measuring biochemical paramets—for example, protein, cytochrome P-450 total activity (from crude S9), and related enzyme activities (from purified microsomes) such as 7-ethoxyresorufin-*O*-deethylase and 7-methoxycoumarin-*O*-demethylase—to give an indication of S9 batch-to-batch variation and to set standards for rejecting suboptimal batches (Hubbard et al., 1985). For further details on critical features affecting the use and limitations of S9 fraction, see Gatehouse and Tweats (1987).

***S9 Mix*** The S9 fraction prepared as described above is used as a component in "S9 mix" along with buffers and various enzyme cofactors. The amount of S9 fraction in the S9 mix can be varied, but a "standard" level of 0.1 mL mL$^{-1}$ of S9 mix (or 10% S9) is often recommended for general screening.

No single concentration of S9 fraction in the S9 mix will detect all classes of genotoxic carcinogen with equal efficiency (Gatehouse et al., 1990). Some mutagens, including many polycyclic aromatic hydrocarbons, are activated to mutagens by higher than normal levels of S9 fraction in the S9 mix (see, e.g., Carver et al., 1985).

The mixed-function oxidases in the S9 fraction require NADPH (the reduced form of nicotinamide adenine dinucleotide phosphate), normally generated from the action of glucose-6-phosphate dehydrogenase acting on glucose-6-phosphate and reducing NADP, both of which are normally supplied as cofactors. As an alternative, isocitrate can be substituted for glucose-6-phosphate (to be used as a substrate by isocitrate dehydrogenase) (Linblad and Jackim, 1982). Additional cofactors may be added (e.g., flavin mono-

**TABLE 5 Composition of Standard S9 Mix**

| Constituent | Final Concentration in Mix (mM) |
|---|---|
| Glucose-6-phosphate | 5 |
| Nicotinamide adenine dinucleotide phosphate | 4 |
| Salt solution | |
| $MgC1_26H_2O$ | 8 |
| KC1 | 33 |
| Phosphate buffer (90.2 M) | 100 |
| Distilled water to make up to required volume | |
| S9 fraction added at 0.1 mL mL$^{-1}$ S9 mix | |

*Note*: For assays using cultured mammalian cells, phosphate buffer and distilled water are replaced by tissue culture medium as high concentrations of Na and K salts are toxic to such cells. The concentration of S9 fraction in the S9 mix varies, depending on the relevant assay (see individual sections). Once prepared, S9 mix should be used as soon as possible, and should be stored on ice until required. S9 fraction, once thawed, should not be refrozen for future use.

nucleotide) when particular classes of compound such as azo dyes are being tested (Prival et al., 1984) or acetyl coenzyme A when aromatic animes such as benzidine are being tested (Kennelly et al., 1984). The composition of a standard S9 mix is given in Table 5.

## 2.6 Bacterial Mutation Tests

The study of mutation in bacteria (and bacterial viruses) has had a fundamental role in the science of genetics in the twentieth century. In particular, the unraveling of biochemical anabolic and catabolic pathways, the identification of DNA as the hereditary material, the fine structure of the gene, the nature of gene regulation, and so on, have all been aided by bacterial mutants.

As an offshoot of studies of genes concerned with the biosynthesis of amino acids, a range of *E. coli* (see, e.g., Yanofsky, 1971) and *S. typhimurium* strains (see, e.g., Ames, 1971) with relatively well-defined mutations in known genes became available. Thus, bacteria already mutant at an easily detectable locus are treated with a range of doses of the test material to determine whether the compound can induce a second mutation that directly reverses or suppresses the original mutations. Thus, for amino acid auxotrophs, the original mutation has resulted in loss of ability to grow in the absence of the required amino acid. The second mutation restores prototrophy—that is, the affected cell is now able to grow in the absence of the relevant amino acid if provided with inorganic salts and a carbon source. This simple concept in fact underlines the great strength of these assays, for it provides enormous selective power which can identify a small number off the chosen mutants from a population of millions of unmutated cells and cells mutated in other genes. The genetic target—that is, the mutated DNA bases in the gene in question (or bases in the relevant tRNA genes; see the discussion of suppressor mutations)—can thus be very small, just one or a few bases in length.

An alternative approach is to use bacteria to detect "forward mutations." Genetic systems which detect forward mutations have an apparent advantage in that a wide variety of genetic changes may lead to a forward mutation—for example, point mutation, deletions, and insertions. In addition, forward mutations in a number of different genes may lead to the same change in phenotype; thus, the genetic target is much larger than that seen in most reverse mutation assays. However, if a particular mutagen causes rare specific changes, these changes may be lost against the background of more common events (Gatehouse et al., 1990). Spontaneous mutation rates tend to be relatively high in forward mutation systems. Acquisition of resistance to a toxic chemical (e.g., an amino acid analogue or antibiotic) is a frequently used genetic marker in these systems. For instance, the use of resistance to the antibiotic streptomycin preceded the reversion assays in common use today.

***Reversion Tests: Background*** There are several excellent references describing the background and use of bacteria for reversion tests (Brusick, 1987a,b; Gatehouse et al., 1990). Three different protocols have been widely used: plate incorporation assays, treat and plate tests, and fluctuation tests. These methods are described in detail in the following sections. Fundamental to the operation of these tests is the genetic compositions of the tester strains selected for use.

***Genetic Make-Up of Tester Strains*** The most widely used strains are those developed by Bruce Ames and colleagues which are mutant derivatives of the organism *S. typhimurium*. Each strain carries one of a number of mutations in the operon coding for histidine biosynthesis. In each case the mutation can be reverted either by base change or by frameshift mutations. The genotype of the commonly used strains is shown in Table 6.

**TABLE 6 Genotype of Commonly Used Strains of *Salmonella typhimurium* LT2 and Their Reversion Events**

| Strain | Genotype | Reversion Events |
|---|---|---|
| TA1535 | $hisG_{46}$ rfa ◑ gal chlD bio uvrB | Subset of base-pair substitution events |
| TA100 | $hisG_{46}$ ◑ rfa gal chlD bio uvrB (pKM101) | Subset of base-pair substitution events |
| TA1537 | $hisC_{3076}$ ◑ rfa gal chlD bio uvrB | Frameshifts |
| TA1538 | $hisD_{3052}$ ◑ rfa gal chlD bio uvrB | Frameshifts |
| TA98 | $hisD_{3052}$ ◑ rfa gal chlD bio uvrB pKM101) | Frameshifts |
| TA97 | $hisD_{6610}$ $hisO_{1242}$ rfa ◑ gal $chl^D$ bio uvrB (pKM101) | Frameshifts |
| TA102 | his · $(\dot{G})_{8476}$ rfa galE (pAQ1) (pKM101) | All possible transitions and transversions; small deletions |

***Use of Plasmid pKM101*** *Salmonella typhimurium* LT2 strains do not appear to possess classical "error-prone" repair as found in *E. coli* strains and some other members of the Enterobacteria (Walker, 1984; Sedgwick and Goodwin, 1985). This is due to a deficiency in *umuD* activity in these *Salmonella* strains (Herrera et al., 1988; Thomas and Sedgewick, 1989). One way to overcome this deficiency and to increase sensitivity to mutagens is to use strains containing a plasmid carrying analogues to the *umuDC* genes, such as are present in the pKM101 pasmid.

***Ames Salmonella/Plate Incorporation Method*** The following procedure is based on that described by Ames and colleagues (Maron and Ames, 1983), with additional modifications:

1. Each selected test strain is grown for 10 h at 37 °C in nutrient broth (Oxoid no. 2) or supplemented minimal media (Vogel-Bonner) on an orbital shaker. A timing device can be used to ensure that cultures are ready at the beginning of the working day.
2. Two-milliliter aliquots of soft agar overlay medium are melted just prior to use and cooled to 50 °C, and relevant supplements added—that is, L-histidine, final concentration 9.55 μg mL$^{-1}$, and D-biotin, 12 μg mL$^{-1}$. (*Note:* If *E. coli* WP2 tester strains are used, the only supplement required is tryptophan 3.6 μg mL$^{-1}$.) The medium is kept semimolten by holding the tubes containing the medium in a hot aluminum dry block held at 45 °C. It is best to avoid water baths as microbial contamination can cause problems.
3. The following additions are made to each tube of top agar: the test article (or solvent control) in solution (10–200 μL), the test strain (100 μL), and, where necessary, S9 mix (500 μL). The test is carried out in the presence and absence of S9 mix. The exact volume of test article or solvent may depend on toxicity or solubility, as described in the preceding section.
4. There should be at least three replicate plates per treatment with at least five test doses plus untreated controls. Duplicate plates are sufficient for the positive and sterility control treatments. The use of twice as many negative control plates as used in each treatment group will lead to more powerful tests from a statistical standpoint (Mahon et al., 1989).
5. Each tube of top agar is mixed and quickly poured onto dried prelabeled Vogel-Bonner basal agar plates.
6. The soft agar is allowed to set at room temperature and the plates are inverted and incubated (within 1 h of pouring) at 37 °C in the dark. Incubation is continued for two to three days.
7. Before scoring the plates for revertant colonies, the presence of a light background lawn of growth (due to limited growth of nonrevertant colonies before the trace of histidine or tryptophan is exhausted) should be

confirmed for each concentration of test article by examination of the plate under low power of a light microscope. At concentrations that are toxic to the test strains, such a lawn will be deplated and colonies may appear that are not true revertants but surviving, nonprototrophic cells. If necessary, the phenotype of any questionable colonies (pseudorevertants) should be checked by plating on histidine or tryptophan-free medium.

8. Revertant colonies can be counted by hand or with an automatic colony counter. Such machines are relatively accurate in the range of colonies normally observed (although regular calibration against manual counts is a wise precaution). Where accurate quantitative counts of plates with large numbers of colonies are required, only manual counts will give accurate results.

## 2.7 Controls

***Positive Controls*** Where possible, positive controls should be chosen that are structurally related to the test article. This increases the confidence in the results. In the absence of structurally related mutagens, the set of positive controls given in Table 7 can be used. The use of such controls validates each test run and helps to confirm the nature of each strain. Pagano and Zeiger (1985) have shown that it is possible to store stock solutions of most routinely used positive controls (sodium azide, 2-aminoanthracene, benzo[*a*]phyene, 4-nitroquinoline oxide) at –20 to –80 °C for several months without loss of

**TABLE 7 Positive Controls for Use in Plate Incorporation Assays**

| Species | Strain | Mutagen | Concentration (μg plate$^{-1}$)[a] |
|---|---|---|---|
| (a) In absence of S9 mix | | | |
| *S. typhimurium* | TA1535, TA100 | Sodium azide | 1–5 |
| | TA1538, TA98 | Hycanthone methane sulfonate | 5–20 |
| | TA1537 | ICR 191 | 1 |
| *E. coli* | WP2 uvrA | Nifuroxime | 5–15 |
| (b) In presence of S9 mix | | | |
| *E. coli* | WP2 uvraA (pKM101) | | |
| *S. typhimurium* | TA1538, TA1535, TA100, TA90 | 2-Aminoanthracene | 1–10 |
| | TA1537 | Neutral red | 10–20 |

[a]The concentration given above will give relatively small increases in revertant count above the spontaneous level. There is little point in using large concentrations of reference mutagens which invariably give huge increases in revertant counts. This would give little information on the day-to-day performance of the assay.

activity. This measure can help reduce potential exposure to laboratory personnel.

***Untreated/Vehicle Controls*** Untreated controls omit the test article but are made up to volume with buffer. The vehicle control is made up to volume with the solvent used to dissolve the test substance. It is preferable to ensure that each of the treated plates contains the same volume of vehicle throughout.

As detailed by Gatehouse and Tweats (1987), the nature and concentration of solvent may have a marked effect on the test result. Dimethylsulfoxide (DMSO) is often used as the solvent of choice for hydrophobic compounds. However, there may be unforeseen effects, such as an increase in mutagenicity of some compounds—for example, *p*-phenylenediamine (Burnett et al., 1982)—or a decrease in mutagenicity of others, such as simple aliphatic nitrosoamines (Yahagi et al., 1977). It is essential to use fresh batches of the highest purity grade available and to prevent decomposition/oxidation on storage. The products after oxidation, for example, both are toxic and can induce base-pair substitutions in both bacterial and mammalian assays. Finally, DMSO and other organic solvents can inhibit the oxidation of different premutagens by microsomal monooxygenases (Wolff, 1977a,b). To reduce the risk of artifactual results, it is essential to use the minimum amount of organic solvent (e.g., <2% w/w/) compatible with adequate testing of the test chemical.

It is important to keep a careful check of the number of mutant colonies present on untreated or vehicle control plates. These numbers depend on the following factors:

1. The repair status of the cell—that is, excision repair-deficient strains tend to have more "spontaneous mutants" than repair-proficient cells.
2. The presence of mutator plasmids. Excision-deficient strains containing pKM101 have a higher spontaneous mutation rate at both base substitution and frameshift loci than excision-proficient strains.
3. The total number of cell divisions that take place in the supplemented top agar. This is controlled by the supply of nutrients—in particular, histidine. Rat liver extracts may also supply trace amounts of limiting nutrients, resulting in a slight increase in the spontaneous yield of mutants in the presence of S9 mix.
4. The size of the initial inoculum. During growth of the starting culture, mutants will arise. Thus, if a larger starting inoculum is used, more of these "preexisting" mutants will be present per plate. In fact, the "plate mutants" arising as described in point 3 predominate.
5. The intrinsic mutability of the mutation in question. In practice the control mutation values tend to fall within a relatively precise range for each strain. Each laboratory should determine the normal range of revertant colonies per plate for each strain.

Deviations in background reversion counts from the normal range should be investigated. It is possible that cross-contamination, variations in media quality, and so on, have occurred that may invalidate particular experiments.

Frequent checks should also be made on the sterility of S9 preparations, media, and test articles. These simple precautions can prevent loss of valuable time and resources.

***Evaluation of Results*** At least two independent assays are carried out for each test article. The criterion for positive response is a reproducible and statistically significant result at any concentration for any strain. When positive results are obtained, the test is repeated using the strain(s) and concentration range with which the initial positive results were observed. This range may be quite narrow for toxic agents.

Several statistical approaches have been applied to the results of plate incorporation assays (Mahon et al., 1989). These authors make a number of important suggestions to maximize the power of statistical analyses; those that relate to the method of analysis are reproduced below:

1. Unless it is obvious that the test agent has had no effect, the data should be plotted to give a visual impression of the form of any dose response and the pattern of variability.
2. Three methods of analysis—linear regression (Gad, 1999; Steel and Torrie, 1960); a multiple-comparison analysis, Dunnett's (1955) method; and a nonparametric analysis, such as Kruskal–Wallis (Gad, 1999)—can all be recommended. Each has its strengths and weaknesses, and other methods are not excluded.
3. Linear regression assumes that variance across doses is constant and that the dose response is linear. If the variance is not approximately constant, then a transformation may be applied or a weighted analysis may be carried out. If the dose scale tends to a plateau, then the dose scale may be transformed. If counts decline markedly at high doses, then linear regression is inappropriate.
4. Dunnett's method, perhaps with a transformation, is recommended when counts decline markedly at one or two high doses. However, when the dose response shows no such decline, other methods may be more powerful.
5. Kruskal–Wallis's nonparametric method avoids the complications of transformations of weighting and is about as powerful as any other method. However, it is inappropriate when the response declines markedly at high dose.

***Preincubation Tests*** Some mutagens are poorly detected in the standard plate incorporation assay, particularly those that are metabolized to short-

lived reactive electrophiles—for example, short-chain aliphatic *N*-nitroso compounds (Bartsch et al., 1984). It is also possible that some metabolites may bind to components within the agar. Such compounds can be detected by using a preincubation method first described by Yahagi et al. (1975) in which the bacteria, test compound, and S9 mix are incubated together in a small volume at 37 °C for a short period (30–60 min) before adding the soft agar and pouring as for the standard assay. In this variation of the test, during the preincubation step, the test compound, S9 mix, and bacteria are incubated in liquid at higher concentrations than in the standard test, and this may account for the increased sensitivity with relevant mutagens. In the standard method the soluble enzymes in the S9 mix, cofactors, and test agent may diffuse into the bottom agar. This can interfere with the detection of some mutagens—a problem that is overcome in the preincubation method (Forster et al., 1980; Gatehouse and Wedd, 1984).

The test is carried out as follows:

1. The strains are cultured overnight, and the inocula and S9 mix are prepared as in the standard Ames test.
2. The soft agar overlays are prepared and maintained at 45 °C prior to use.
3. To each of three to five tubes maintained at 37 °C in a Driblock are added 0.5 mL of S9 mix, 0.1 mL of the tester strain (10–18-h culture), and a suitable volume of the test compound to yield the desired range of concentrations. The S9 mix is kept on ice prior to use.
4. The reaction mixtures are incubated for 1 h at 37 °C.
5. Two milliliters of soft agar is added to each tube. After mixing, the agar and reaction mixture are poured onto previously labeled, dried Vogel–Bonner plates.
6. Once the agar has set, the plates are incubated for two to three days before revertant colonies are scored.

The use of controls is as described for the plate incorporation assay. It is crucial to use the minimum amount of organic solvent in this assay, as the total volume of the incubation mixture is small relative to the solvent component.

This procedure can be modified to provide optimum conditions for particular chemical classes. For instance, preincubation times greater than 60 min plus aeration have been found necessary in the detection of allyl compounds (Neudecker and Henschler, 1985).

***E. coli* Tester Strains** Ames and colleagues have made an impressive contribution to mutagenicity testing by the development of the *Salmonella*/ microsome test and, in particular, its application in the study of environmental mutagens. In genetic terms, *Salmonella* strains are, in some ways, not the best choice (see, e.g., Venitt and Croften-Sleigh, 1981). Unlike the *Salmonella*

strains, *E. coli* B strains such as the WP2 series developed by Bridges, Green, and colleagues (Bridges, 1972; Green and Muriel, 1976) inherently possess the *umuDC*$^+$ genes involved in generating mutations; they are also part rough and thus allow many large molecules to enter the cell.

In addition to being effective general strains for mutagen detection, studies by Wilcox et al. (1990) have shown that a combination of *E. coli* WP2 *trp E* (pKM101), which has a functioning excision repair system for the detection of crosslinking agents, and *E. coli* WP2 *trp E uvrA* (pKM101) can be used as alternatives to *Salmonella* TA102 for the detection of oxidative mutagens. The *E. coli* strains have the advantage of lower spontaneous mutation rate and are somewhat less difficult to use and maintain. The *Salmonella* strains are, however, more commonly employed.

***Storage and Checking of Tester Strains*** Detailed instructions for maintenance and confirmation of the phenotypes of the various tester strains are given in Maron and Ames (1983) and Gatehouse et al. (1990). Permanent master cultures of tester strains should be stored in liquid nitrogen or in dry ice. Such cultures are prepared from fresh nutrient broth cultures to which DMSO is added as a cryopreservative. These cultures are checked for the various characteristics before storage as described below. Cultures for use in individual experiments should be set up by inoculation from the master culture or from a plate made directly from the master culture, not by passage from a previously used culture. Passage in this way will inevitably increase the number of preexisting mutants, leading to unacceptably high spontaneous mutation rates (Gatehouse et al., 1990).

The following characteristics of the tester strains should be confirmed at monthly intervals or if the internal controls of a particular experiment fail to meet the required limits:

- Amino acid requirement.
- Sensitivity to the lethal effects of the high-molecular-weight dye crystal violet for those strains carrying the *rfaE* mutation.
- Increased sensitivity to UV irradiation for those strains carrying the *uvrA* or *uvrB* mutations.
- Resistance to ampicillin for strains carrying pKM101 and resistance to tetracycline for strains carrying pAQ1.
- Sensitivity to diagnostic mutagens. This can be measured very satisfactorily by testing pairs of strains—one giving a strongly positive response, the partner a weak response.

The importance of these checks together with careful experiment-to-experiment controls of spontaneous mutation rates and response to reference mutation rates and response to reference mutagens cannot be overstressed; failure to apply them can result in much wasted effort.

## 2.8 Plate Incorporation Assay

***Protocol for Dose Ranging and Selection*** Before carrying out the main tests, it is necessary to carry out a preliminary toxicity dose-ranging test. This should be carried out following the same basic protocol as the mutation test, except that instead of scoring the number of mutants on, for example, minimal media plates with limiting amounts of a required amino acid, the number of survivors is scored on fully supplemented minimal media. A typical protocol is outlined below:

1. Prepare a stock solution of the test compound at a concentration of 50 mg $mL^{-1}$ in an appropriate solvent. It may be necessary to prepare a lower concentration of stock solution, depending on the solubility of the test compound.
2. Make dilutions of the stock solution.
3. To 2.0-mL aliquots of soft agar overlay medium (0.6% agar and 0.5% sodium chloride in distilled water) containing a trace of histidine and excess biotin and maintained at 45 °C in a dry block, add 100 μL of a solution of the test article. Use only one plate per dilution.
4. Mix and pour onto dried Vogel and Bonner minimal medium plates as in an Ames test, including an untreated control and a solvent control, if necessary. The final concentrations of test compound will be 5000, 1500, 500, 150, and 50 μg $plate^{-1}$.
5. Repeat step 3 using 0.5 mL of 8% S9 mix per 2.0-mL aliquot of soft agar in addition to the test compound and tester strain. The S9 mix is kept on ice during the experiment.
6. Incubate the plates for two days at 37 °C and examine the background lawn of growth with a microscope (8-eyepiece lens, 10-objective lens). The lowest concentration giving a depleted background lawn is regarded as a toxic dose.

This test will also demonstrate excess growth, which may indicate the presence of histidine or tryptophan or their precursors in the test material, which could make testing for mutagenicity impracticable by this method.

When setting the maximum test concentration, it is important to test into the milligram-per-plate range where possible (Gatehouse et al., 1990), as some mutagens are only detectable when tested at high concentrations. However, for nontoxic, soluble mutagens an upper limit of 5 mg $plate^{-1}$ is recommended (DeSerres and Shelby, 1979). For less soluble compounds at least one dose exhibiting precipitation should be included.

***Forward Mutation Tests*** Forward mutation is an endpoint that may arise from various events, including base substitutions, frameshifts, DNA deletions, and so on, as mentioned earlier.

## 2.9 Eukaryotic Mutation Tests

Prokaryotic systems, as described, have proved to be quick, versatile, and in many cases surprisingly accurate in identifying potential genetic hazards to humans. However, there are intrinsic differences between eukaryotic and prokaryotic cells in the organization of the genome and the processing of the genetic information. Thus, there is a place for test systems based on mammalian cells for fundamental studies to understand the mutation process in higher cells and for the use of such tests for screening for genotoxic effects.

The early work of Muller showed the usefulness of the fruit fly *D. melanogaster* as a higher system for measuring germ line mutations in a whole animal. The *Drosophila* sex-linked recessive lethal test has yielded much useful information and in the 1970s was a popular system for screening chemicals for mutation, but this test failed to perform well in international collaborative trials to study the utility of such tests to detect carcinogens and popularity wanted. Another *Drosophila* test devised in the 1980s, the SMART assay (somatic mutation and recombination test) shows much promise and has revived the popularity of *Drosophila* for screening for genotoxic agent.

A number of test systems that use cultured mammalian cells from both established and primary lines now have a large database of tested chemicals in the literature, are relatively rapid, and are feasible to use for genetic toxicity screening. These are discussed in the next section.

## 2.10 In Vitro Tests for Detection of Mammalian Mutation

There have been a variety of in vitro mutation systems described in the literature, but only a small number have been defined adequately for quantitative studies (Cole et al., 1990). These are based on the detection of forward mutations in a similar manner to the systems described earlier for bacteria. A defined large number of cells are treated with the test agent and then, after a set interval, exposed to a selective toxic agent, so that only cells that have mutated can survive. As cultured mammalian cells are diploid (or near diploid), normally there are two copies of each gene. Recessive mutations can be missed if a normal copy is present on the homologous chromosome. As mutation frequencies for individual genes are normally very low, an impossibly large population of cells would need to be screened to detect cells in which both copies are inactivated by mutation. This problem is overcome by measuring mutation in genes on the X chromosome in male cells where only one copy of the gene will be present or using heterozygous genes where two copies of a gene may be present but one copy is already inactive through mutation or deletion.

Many genes are essential for the survival of the cell in culture, and thus mutations in such genes would be difficult to detect. However, use has been made of genes that are not essential for cell survival but allow the cell to

salvage nucleotides from the surrounding medium. This saves the cell energy, as it does not have to make these compounds from simpler precursors by energy-expensive catabolism. These enzymes are located at the cell membrane. If the cell is supplied with toxic nucleotides, the "normal" unmutated cells will transport these into the cell and kill the cell. However, if the cells have lost the enzyme as a result of mutation (or chromosomal deletion, rearrangement, etc.), then they will not be able to "salvage" the exogenous toxic nucleotides and will survive. The surviving mutant cells can be detected by the formation of colonies on tissue culture plates or, in some cases, in the wells of microtiter plates.

One factor to take into account with these tests is that of expression time. Although a gene may be inactivated by mutation, the mRNA existing before the mutational event may decay only slowly, so that active enzyme may be present for some time after exposure to the mutagen. Thus, the cells have to be left for a period before challenging with the toxic nucleotide: This is the expression time and differs between systems.

***Chinese Hamster Lines*** Chinese hamster cell lines have given much valuable data over the past 15 years, but their use for screening is limited by lack of sensitivity, as only a relatively small target cell population can be used, owing to metabolic cooperation (see Cole et al., 1990); however, they are still in use, so a brief description follows.

Chinese hamster ovary (CHO) and V79 lines have high plating efficiencies and short generation times (less than 24 h). These properties make the lines useful for mutagenicity experiments. Both cell lines have grossly rearranged chromosomal complements, which has an unknown effect on their responsiveness to mutagens (Tweats and Gatehouse, 1988). There is some evidence that Chinese hamster lines are undergoing genetic drift in different culture collections (Kirkland and Garner, 1987).

***V79 System*** The Chinese hamster V79 line was established in 1958 (Ford and Yerganian, 1958). Publication of the use of the line for mutation studies (by measuring resistance to purine analogues due to mutation of the *hgprt* locus) occurred 10 years later (Chu and Malling, 1968). The V79 line was derived from a male Chinese hamster; hence, V79 cells possess only a single X chromosome.

V79 cells grow as a cell sheet or monolayer on glass or plastic surfaces. If large numbers of cells are treated with a mutagen, when plated out, cells in close contact can link via intracellular bridges. These allow the transfer of cellular components between cells such as mRNA. Thus, if a cell carries a mutation in the *hgprt* gene resulting in the inactivation of the relevant mRNA, it can receive viable mRNA or intact enzyme from adjacent nonmutated cells. Therefore, when the mutated cell is challenged with a toxic purine, it is lost, owing to the presence of active enzyme derived from the imported mRNA. This phenomenon is termed *metabolic cooperation* and severely limited the

sensitivity of lines such as V79 for mutagen detection. This drawback can be overcome to an extent by carrying out the detection of mutant clones in semi-solid agar (see, e.g., Oberly et al., 1987) or by using the "respreading technique" (Fox, 1981).

The preferred expression time for *hgprt* mutants is six to eight days, although care needs to be taken when testing chemicals well into the toxic range, where the "expression time" needs to be extended to allow recovery.

***Preliminary Cytotoxicity Testing*** An essential first step is to carry out a preliminary study to evaluate the toxicity of the test material to the indicator cells under the conditions of the main mutagenicity test. When selecting dose levels, the solubility of the test compound, the resulting pH of the media, and the osmolality of the test solutions all need to be considered. The latter two parameters have been known to induce false-positive effects in in vitro mammalian tests (Brusick, 1986). The experimental procedure is carried out as follows:

1. Seek T75 plastic tissue culture flasks with a minimum of $2.5 \times 10^6$ cells in 120 mL of Eagle's medium containing 20 mM L-glutamine: 0.88 g $L^{-1}$ sodium bicarbonate, 20 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES), 50 μg $mL^{-1}$ streptomycin sulfate, 50 IU $mL^{-1}$ benzylpenicillin, and 7.5% of fetal bovine serum. Incubate the flasks for 18–24 h at 37 °C in a $CO_2$ incubator to establish monolayer cultures.
2. Prepare treatment medium containing various concentrations of test compound—for example, 19.7 mL of Eagle's medium (without serum) plus 300 μL of stock concentration of compound in a preferred solvent (e.g., water, ethanol, DMSO, etc.). The final concentration of solvent other than water should not exceed 1% v/v. Normally a range of 0–5000 μg $mL^{-1}$ (final concentration) is covered. For a sparingly soluble compound, the highest concentration will be the lowest at which visible precipitation occurs. Similarly, if a compound has a marked effect on osmolality, concentrations should not be used that exceed 500 milliosmoles (mOsm) per kilogram. In addition, a pH range of 6.5–7.5 should be maintained.
3. Rinse each cell monolayer with a minimum of 20 mL phosphate-buffered saline (PBS) and then carefully add 20 mL of treatment medium. Incubate the flasks for 3 h at 37 °C in a $CO_2$ incubator.
4. After treatment, carefully discard the medium from each flask and wash each monolayer twice with PBS. Care needs to be taken to safely dispose of contaminated solutions.
5. To each flask add 10 mL of trypsin solution (0.025% trypsin in PBS). Once the cells have rounded up, neutralize the trypsin by adding 10 mL of complete medium. A cell suspension is obtained by vigorous pipetting to break up cell clumps.

6. Count the trypsinized cell suspension and dilute in complete media before assessing for survival. For each treatment set up five Petri dishes containing 200 cells per dish.
7. Incubate at 37 °C in a $CO_2$ incubator for 7–10 days.
8. Remove the medium and fix and stain the colonies using 5% Giemsa in buffered formalin. Once the colonies are stained, remove the Giemsa and count the colonies.

The method can be repeated including 20% v/v S9 mix.

To calculate percentage survival, the following formula is used:

$$\frac{\text{Cell titer in treated culture}}{\text{Cell titer in control culture}} \times \frac{\text{mean no. of colonies on treated plates}}{\text{mean no. of colonies on control plates}} \times 100$$

The cloning efficiency (CE) of the control culture is calculated as

$$\text{CE} = \frac{\text{mean no. of colonies per plate}}{\text{no. of cells per plate (i.e., 200)}} \times 100$$

In the absence of precipitation or effects on pH or osmolality, the maximum concentration of the main mutagenicity study is a concentration that reduces survival to approximately 20% of the control value.

***Procedure for the Chinese Hamster V79/Hgprt Assay*** The assay usually comprises three test concentrations, each in duplicate, and four vehicle control replicates. Suitable positive controls are ethylmethane sulfonate (–S9) and dimethyl benzanthracene (+S9). V79 cells with a low nominal passage number should be used from frozen stocks to help minimize genetic drift. The procedure described includes a reseeding step for mutation expression.

Steps 1–5 are the same as the cytotoxicity assay. As before, tests can be carried out in the presence and in the absence of S9 mix:

6. Count the trypsinized cultures and assess a sample for survival as for the cytotoxicity assay. In addition, reseed an appropriate number of cells for estimation of mutation frequency at day 8 of the expression time. Transfer the cells to roller bottles (usually 490 $cm^2$) for this stage. Gas the bottles with pure $CO_2$, tighten the tops, and incubate the bottles at 37 °C on a roller machine (approximate speed 0.5–1.0 rev $min^{-1}$). Usually $10^6$ viable cells are reseeded in 50 mL of Eagle's medium containing serum, but more cells are required at the toxic dose levels.
7. Subculture the bottles as necessary throughout the expression period to maintain subconfluency. This involves retrypsinization and determining the cell titer for each treatment. For each culture a fresh roller bottle is reseeded with a minimum of $10^6$ cells.

8. On day 8, trypsinize, count, and dilute each culture again so that a sample cell population can be assessed for cloning efficiency and a second sample can be assessed for the induction of 6TG-resistant cells.
9. Dilute the cell suspension in complete medium and add $2 \times 10^5$ cells per Petri dish (10 Petri dishes per treatment). Add 6-thioguanine to the medium to a final concentration of $10\,\mu g\,mL^{-1}$.
10. Incubate the Petri dishes for 7–10 days and then remove the medium. Fix and stain the colonies are as previously. Then count the colonies (>50 cells per clone).

Mutation frequency in each culture is calculated as

$$\frac{\text{Mean no. colonies on thioguanine plates}}{1000 \times \text{mean no. colonies on survival plates}}$$

***Data Analysis (Arlett et al., 1989)*** A weighted analysis of variance is performed on the mutation frequencies, as the variation in the number of mutations per plate usually increases as the mean increases. Each dose of test compound is compared with the corresponding vehicle control by means of a one-sided Dunnett's test and, in addition, the mutation frequencies are examined to see whether there is a linear relationship with dose.

The criterion employed for a positive response in this assay is a reproducible statistically significant increase in mutation frequency (weighted mean for duplicate treated cultures) over the concurrent vehicle control value (weighted mean for four independent control cultures). Ideally, the response should show evidence of a dose–response relationship. When a small isolated significant increase in mutation frequency is observed in only one of the two duplicate experiments, then a third test should be carried out. If the third test shows no significant effects, the initial increase is likely to be a chance result. In cases where an apparent treated-related increase is thought to be a result of unusually low variability or a low control frequency, comparison with the laboratory historical control frequency may be justified.

***CHO/Hgprt System*** The CHO cells have 21 or 22 chromosomes with one intact X chromosome and a large acrocentric marker chromosome (Natarajan and Obe, 1982). The use of these cells in mammalian mutation experiments was first reported by Hsie et al. (1975) and was refined into a quantitative assay for mutagenicity testing by O'Neill. The performance of this system has been reviewed by the U.S. Environmental Protection Agency (EPA) Gene-Tox Program. The experimental procedure for this assay is similar to the V79/Hgprt system already described, and for more detailed descriptions the reader is referred to Li et al. (1987).

***Mouse Lymphoma L5178Y TK$^{+/-}$ Assay*** Whereas the Chinese hamster cell systems are regarded as relatively insensitive, the mouse lymphoma L5178Y

$TK^{+/-}$ test is undoubtedly more sensitive. Unfortunately, there are persistent doubts regarding its specificity—that is, the ability to distinguish between carcinogens and noncarcinogens (see, e.g., Tennant et al., 1987). However, a great advantage is the ability of these cells to grow in suspension culture in which intracellular bridges do not occur. Thus, the problems of metabolic cooperation are avoided, which allows a large number of cells to be treated for optimum statistical analysis of results.

A candid historical overview of the development of the mouse lymphoma $TK^{(+/-}$ mutagenicity assay is given by its originator, Clive (1987). Initially methodologies were developed for producing the three TK genotypes ($TK^{+/+}$ and $TK^{-/-}$ homozygotes and the $TK^{+/-}$ heterozygotes (Clive et al., 1972). This first heterozygote was lost; however, it was recognized that subsequent heterozygotes produced distinctly bimodal distributions of mutant colony sizes, owing to differences in growth rate. These were interpreted in terms of single-gene (large-colony mutants) and viable chromosomal mutations (small-colony mutants). A period of diversification of the mouse lymphoma assay followed with controversy over the significance of small-colony mutants (Amacher et al., 1980).

Following this, a series of cytogenetic studies confirmed the cytogenetic interpretation for small-colony mutants (see, e.g., Hozier et al., 1982). Molecular studies showed that most mutations resulting in small-colony mutants involve large-scale deletions (Evans et al., 1986). A current theory states that, for many compounds, deletion mutants are induced by binding of the compound to complexes between topoisomerase II and DNA (Clive, 1989). Topoisomerases are enzymes that control supercoiling via breakage and reunion of DNA strands; it is the latter strep that is disrupted, which leads to chromosome damage and deletions. Further molecular studies (Applegate et al., 1990) have shown that a wide variety of genetic events can result in the formation of $TK^{+/-}$ genotype from the heterozygote, including recombinations and mitotic nondisjunction.

The $TK^{+/-}$ line was originally isolated as a spontaneously arising revertant clone from a UV-induced $TK^{-/-}$ clone. The parental $TK^{+/+}$ cell and the heterozygote were then the only TK-competent mouse lymphoma cells that could be maintained in THMG medium ($3\,\mu g\,mL^{-1}$ thymidine, $5\,\mu g\,mL^{-1}$ hypoxanthine, $0.1\,\mu g\,mL^{-1}$ methotrexate, and $7.5\,\mu g\,mL^{-1}$ glycine) (Clive, 1987). Thus, like most established lines, these cells are remote from wild-type cells. The karyotype of the $TK^{+/-}$–3.7.2C line has a modal chromosome number of 40, like wild type, but has a variety of chromosomal rearrangements and centromeric heteromorphisms (Blazak et al., 1986).

Two main protocols have been devised for carrying out mutation assays with mouse lymphoma L5178Y cells—that is, plating the cells in soft agar or a fluctuation test approach. It is the latter that is described in the following section, based on Cole et al. (1986). The reader is referred to Clive et al. (1987) for a full description of the soft-agar method.

***Preliminary Cytotoxicity Assay*** The cells are maintained in RPMI 1640 medium containing 2.0 mM glutamine, 20 mM HEPES, 200 μg $mL^{-1}$ sodium pyruvate, 50 IU $mL^{-1}$ benzylpenicillin, 50 μg $mL^{-1}$ streptomycin sulfate, and 10% donor horse serum (heat inactivated for 30 min at 56 °C). This medium is designated CM10. Conditioned medium is CM10 in which cells have grown exponentially for at least one day. Treatment medium contains 3% horse serum and 30% conditioned media (CM3). Medium without serum is known as incomplete medium (ICM). If treatment time exceeds 3 h, treatment is carried out in CM10.

The method is as follows:

1. The cell titer of an exponentially growing culture of cells in CM10 is determined with a Coulter counter. The cell suspension is centrifuged at 70 *g* for 5 min and the supernatant is reduced such that 3 mL contains approximately $5 \times 10^6$ cells (3 h treatment) or $2 \times 10^6$ (treatment >3 h).

2a. For tests in the absence of S9 mix, treatment groups are prepared by mixing 3 mL of solution of test compound and 6.9 mL of ICM (3 h treatment) or 6.9 mL of CM10 (treatment >3 h).

2b. Tests in the presence of S9 mix are carried out in the same way, except the treatment medium contains 10% v/v S9 mix at the expense of ICM—that is, 3 mL cell suspension, 5.9 mL ICM, 1 mL S9 mix, and 0.1 mL test compound solution per vehicle. The composition of the S9 mix is as described earlier. It is prepared immediately before required and kept on ice until it is added to the test system. For the vehicle controls, if an organic solvent is used, it should not exceed 1% v/v.

3. After the treatment period, cells are spun down at 70 *g* for 5 min and the supernatant is transferred for assessment of pH and osmolality. The cell pellet is washed twice in PBS and then resuspended in 10 mL CM10. (All contaminated material and waste should be disposed of safely.)

4. The cell titer of each culture is counted and a sample diluted in CM10 for assessment of posttreatment survival. For this two 96-well microtiter plates are charged with 200 μL of a diluted cell suspension using a multichannel pipette such that each well contains on average one cell.

5. Plates are incubated for seven to eight days at 37 °C and 5% $CO_2$ in 95 ± 3% relative humidity.

6. The plates are removed from the incubator and 20 μL of MTT [3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide] at 5 mg $mL^{-1}$ (in PBS) is added to each well with a multichannel pipette. The plates are left to stand for 1–4 h and are then scored for the presence of colonies with a Titertek mirror-box, which allows direct viewing of the bottom surface of the plates.

7. Cytotoxicity can also be determined posttreatment as follows: T25 flasks are set up after treatment containing $0.75 \times 10^5$ cells $mL^{-1}$ in 5 mL

CM10. Flasks are incubated with loose lids at 37 °C with 5% $CO_2$ in 95 ± 3% relative humidity. Two days later the cell titer of each culture is determined with a Coulter counter.

8. Following this procedure, various calculations are carried out to aid selection of dose levels for the main mutation assay.

   (a) *Cloning Efficiency* In microtiter assays calculations are based on the Poisson distribution:

$$P(o) = \frac{\text{no. of wells without colony}}{\text{total no. of cells}}$$

   (b) *Relative Survival* Relative survival ($S$) is calculated as

$$S = \frac{\text{CE of treated group}}{\text{CE of control group}}$$

   (c) *Growth* Growth in suspension (SG) is calculated as

$$\text{SG} = \frac{\text{cell count after 3 days}}{0.75 \times 10^5}$$

   Relative suspension growth (RSG) is calculated as

$$\text{RSG} = \frac{\text{SG of treated group}}{\text{SG of control group}} \times 100\%$$

***Selection of Dose Levels*** The highest test concentration is selected from one of the following options, whichever is lowest:

- A concentration which reduces survival to about 10–20% of the control value.
- A concentration which reduces RSG to 10–20% of the control value.
- The lowest concentration at which visible precipitation occurs.
- The highest concentration which does not increase the osmolality of the medium to greater than 400 mmol $kg^{-1}$ or 100 mmol above the value for the solvent control.
- The highest concentration that does not alter the pH of the treatment medium beyond the range 6.8–7.5.
- If none of these conditions are met, 5 mg $mL^{-1}$ should be used.

Lower test concentrations are selected as fractions of the highest concentration, usually including one dose which causes 20–70% survival and one dose which causes >70% survival.

***Main Mutation Assay*** The assay normally comprises three test concentrations, a positive control, and a vehicle control. All treatment groups are set up in duplicate. The expression time is two days, unless there are indications that the test agent inhibits cell proliferation, where an additional or possibly alternative expression time should be employed.

Stock cultures are established from frozen ampoules of cells that have been treated with thymidine, hypoxanthine, methotrexate, and glycine for 24 h, which purges the culture of preexisting $TK^{-/-}$ mutants. This cell stock is used for a maximum of two months.

Treatment is normally carried out in 50-mL centrifuge tubes on a roller machine. During the expression time the cells are grown in T75 plastic tissue culture flasks. For estimation of cloning efficiency and mutant induction, cells are plated out in 96-well microtiter plates. Flasks and microtiter plates are incubated at 37 °C in a $CO_2$ incubator as in the cytotoxicity assays.

Cell titers are determined by diluting the cell suspension in Isoton and counting an appropriate volume (usually 0.5 mL) with a Coulter counter. Two counts are made per suspension.

The experimental procedure is carried out as follows:

1. On the day of treatment stock solutions for the positive control and the various concentrations of test compound (selected as per the previous selection) are prepared.
2. Treatment is carried out in 30% conditioned media. The serum concentration is 3% (3 h treatment) or 10% (treated >3 h).
3. Cell suspensions of exponentially growing cells are prepared as in the cytotoxicity assay, except that 6 mL of media required for each treatment culture contains $10^7$ cells (3 h treatment) or $3 \times 10^6$ cells (>3 h treatment). The number of cells per treatment may be increased if marked cytotoxicity is expected to allow enough cells to survive (e.g., if 20% survival or less is expected, $2 \times 10^7$ cells may be treated).
4. For tests in the absence of S9 mix, 6 mL of cell suspension, 0.2 mL test compound/vehicle, and 13.8 mL ICM (3 h treatment) or 13.8 mL CM10 (treatment >7 h) are mixed in the presence of S9 mix and 0.2 mL of test compound per vehicle is prepared.
5. After treatment the cells are centrifuged at 70 *g* for 5 min, supernatant is discarded, and the cell pellet is resuspended in PBS (pH 7). This washing procedure is repeated twice, and finally the cell pellet is resuspended in CM10.
6. Each culture is counted so that a sample of cells can be assessed for posttreatment survival, and the remaining cell population is assessed for estimation of mutation frequency.
7. For survival estimation, cells are placed into 96-well microtiter trays at a cell density of one cell per well as per the cytotoxicity assay.

8. For mutation estimation, the cells are diluted to a cell density of $2 \times 10^5$ cells $mL^{-1}$ with CM10 in tissue culture flasks and the culture is incubated at 37 °C in a $CO_2$ incubator. On day 1 each culture is counted and diluted with fresh medium to a cell density of $2 \times 10^5$ cells $mL^{-1}$ in a maximum of 100 mL of medium.
9. On day 2 each culture is counted again and an aliquot of cells taken so that: (a) A sample of the cell population can be assessed for cloning efficiency. Plates are incubated at 37 °C in a $CO_2$ incubator for seven days. (b) A sample of the cell population can be assessed for the induction of TFT (tetrafluorotoluene)-resistant cells (mutants). For this $2 \times 10^3$ cells are plated per well in 200 μL CM10 containing 4 μg $mL^{-1}$ TFT. TFT and TFT-containing cultures must not be exposed to bright light, as the material is light sensitive. The plates are incubated for 10–12 days at 37 °C in a $CO_2$ incubator.
10. At the end of incubation 20 μL MTT is added to each well. The plates are left to develop for 1–4 h at 37 °C and then scored for colony-bearing wells. Colonies are scored by eye and are classified as small or large.

The calculation for cloning efficiency is made as for the cytotoxicity assay.

Relative total growth (RTG) is a cytotoxicity parameter which considers growth in suspension during the expression time and the cloning efficiency of the end of the expression time as follows:

$$\mathrm{SG} = \frac{24\,\mathrm{h\ cell\ count}}{2 \times 10^4} \times \frac{48\,\mathrm{h\ cell\ count}}{2 \times 10^5}$$

$$\mathrm{RTG} = \frac{\mathrm{SG\ treated\ culture}}{\mathrm{SG\ control\ culture}} \times \frac{\mathrm{CE\ of\ treated\ culture}}{\mathrm{CE\ of\ control\ culture}}$$

Mutation frequency (MF) is calculated as

$$\mathrm{MF} = \frac{\mathrm{InPo\ for\ mutation\ plates}}{\mathrm{no.\ of\ cells\ per\ well} \times \mathrm{CE}/100}$$

***In Vivo Genotoxicity Tests for Assessment of Primary DNA Lesions*** Primary DNA lesions are detected with so-called indicator tests. These tests do not directly measure consequences of DNA interaction (i.e., mutation) but do detect effects related to the process of mutagenesis, such as DNA damage, recombination, and repair. Results from indicator tests can provide additional useful information in the context of extended genotoxicity testing. However, primary DNA lesions may be repaired error free and do not necessarily result in formation of mutations. The most commonly utilized assays in pharmaceutical development are the phosphorus-postlabeling assay

**TABLE 8 Genotoxicity Tests Recommended by ICH**

| Genotoxicity Test | Mutation | Cell Type | Method |
|---|---|---|---|
| Test for gene mutation in bacteria | Gene | Bacterial | In vitro |
| In vitro cytogenetic assay using mouse lymphomas tk cells | Chromosome | Mammalian | In vitro |
| In vivo test for chromosomal damage using rodent hematopoietic cells | Gene | Mammalian | In vivo |

and the comet assay. A comparison of different aspects of the methods described in the text is depicted in Table 8. Basic aspects regarding optimal study design for in vivo micronucleus assays are largely applicable to the design of supplemental in vivo assays. Specific or unique aspects on study protocols are described more extensively where appropriate.

***Comet Assay*** The in vivo comet assay (single-cell gel electrophoresis) is increasingly being used as a supplement genotoxicity test for drug candidates (Hartman et al., 1993; Brendler-Schwaab et al., 2005) There are general review articles on the comet assay (Tice, 2000; Sheit and Hartman, 2005) and a general guideline for test conductance has been published as a result of the International Workshop on Genotoxicity Test Procedures (IWGTP).

More specific recommendations with the goal of gaining more formal regulatory acceptance of the comet assay were published following the fourth International Comet Assay Workshop (Hartman et al., 2007). An updated position paper on specific aspects of tests conditions and data interpretation was prepared following the IWGTP in 2005 (Burlinson et al., 2006). See Table 2.

*Principle of Method* The basic principle of the comet assay is the migration of DNA in an agarose matrix under electrophoretic conditions. When viewed through the microscope, a cell has the appearance of a comet, with a head (the nuclear region) and a tail containing DNA fragments or strands migrating in the direction of the anode. Among the various versions of the comet assay, the alkaline (pH of the unwinding and electrophoresis buffer at least 13) method enables detection of the broadest spectrum of DNA damage and is therefore urgently recommended (in the first instance) for regulatory purposes (Tice et al., 2000). The alkaline version detects DNA damage such as strand breaks, alkali-labile sites (ALS), and single-strand breaks associated with incomplete excision repair. Under certain conditions, the assay can also detect DNA–DNA and DNA–protein crosslinking, which (in the absence of other kinds of DNA lesions) appears as a relative decrease in DNA migration compared to concurrent controls. In contrast to other DNA alterations, crosslinks may stabilize chromosomal DNA and inhibit DNA migration (Merks et al., 1998). Thus, reduced DNA migration in comparison to the negative control (which

should show some degree of DNA migration) may indicate the induction of crosslinks, which are relevant lesions with regard to mutagenesis and should be further investigated. Increased DNA migration indicated the induction of DNA strand breaks and/or ALS. Furthermore, enhanced activity of excision repair may result in increased DNA migration. DNA excision repair can influence comet assay effects in a complex way (Speit and Hartmann, 1995). While DNA repair generally reduces DNA migration by eliminating DNA lesions, ongoing excision repair may increase DNA migration to incision-related DNA strand breaks. Thus, the contribution of excision repair to the DNA effects seen in the comet assay depends on the types of induced primary DNA damage and the time point of analysis (Collins et al., 1993). Test procedure aspects regarding test animals, test substance, use of concurrent negative and positive control animals, as well as dose selection for the design of a cytogenetic assay, as described in detail previously, are largely applicable to the design of an in vivo comet assay. In addition, more specific details can be found in an earlier publication. A single treatment or repeated treatments (generally at 24-h intervals) are equally acceptable. In both experimental designs, the study is acceptable as long as a positive effect has been demonstrated or, for a negative result, as long as an appropriate level of animal or tissue toxicity has been demonstrated or the limit dose with appropriate level tissue exposure has been used. For repeated treatment schedules, dosing must be continued until the day of sampling. On a daily basis, test substances may be administrated as a split dose (i.e., two treatments separated by no more than a few hours) to facilitate administering a large volume of material. The test may be performed in two ways. If animals are treated with the test substances once, then tissue/organ samples are obtained at 2–6 and 16–26 h after dosing. The shorter sampling time is considered sufficient to detect rapidly absorbed as well as unstable or direct-acting compounds. In contrast, the late sampling time is intended to detect compounds that are more slowly absorbed, disturbed, and metabolized. When a positive response is identified at one sampling time, data from the other sample time need not be collected. Alternatively, if multiple treatments at 24-h intervals are used, tissue/organ samples need be collected only once. The sampling time should be 2–6 h after the last administration of the test substance. Alternative sampling times may be used when justified on the basis of toxicokinetic data.

*Selection of Tissues and Cell Preparation* In principle, any tissue of the experimental animal, provided that a high-quality single-cell/nucleus suspension can be obtained, can be used for a comet assay. Selection of the tissue(s) to be evaluated should be based, wherever possible, on data from absorption, distribution, metabolism, excretion studies, and/or other toxicological information. A tissue should not be evaluated unless there is evidence of or support for exposure of the tissue to the test substance and/or its metabolite (s). In the absence of such information and unless scientifically justified, two tissues should be examined. Recommended tissues are liver, which is the major organ

for the metabolism of absorbed compounds, and a site of first-contact tissue—for example, gastrointestinal for orally administered substances, respiratory tract for substances administrated via inhalation, or skin for dermally applied substances. Which tissue is evaluated first is at the discretion of the investigator and both tissues need not be evaluated if a positive response is obtained in the first tissue evaluated.

Single-cell suspension can be obtained from solid tissue by mincing briefly with a pair fine scissors (Tice et al., 1991), incubation with digestive enzymes such as collagenase trypsin (Brendler Schwaab et al., 1994), or pushing the tissue sample through a mesh membrane. Cell nuclei can also be obtained by homogenization. During mincing or homogenization, ethylenediaminetetraacetic acid (EDTA) can be added to the processing solution to chelate calcium/magnesium and prevent endonuclease activation. In addition, radical scavengers (e.g., DMSO) can be added to prevent oxidant-induced DNA damage. Any cell dissociation method is acceptable as long as it can be demonstrated that the process is not associated with inappropriate background levels of DNA damage.

*Cytotoxicity: Potential Confounding Factor* A general issure with DNA strand break assays such as the comet is that indirect mechanisms related to cytoxicity may lead to enhanced strand break formation. However, since DNA damage in the comet assay is assessed on the level of individual cells, dead or dying cells may be identified on a microscopic slide by their specific image. Necrotic or apoptotic cells can result in comets with small or nonexistent head and large, diffuse tails as observed in vitro upon treatment with cytotoxic, nongenotoxic articles. However, such microscopic images are not uniquely diagnostic for apoptosis or necrosis since they may also be detected after treatment with high doses of radiation or high concentrations of strong mutagens. For the in vivo comet assay, only limited data are available to establish whether cytotoxicity results in increased DNA migration in tissues of experimental animals. Despite necrosis or apoptosis in target organs of rodents such as kidneys, testes, liver, or duodenum, no elevated DNA migration was observed. However, enhanced DNA migration was seen in homogenized liver tissue of mice dosed with carbon tetrachloride when histopathological examination showed evidence of necrosis in the liver. Therefore, to avoid potential false-positive effects resulting from cytotoxicity, recommendations regarding a concurrent assessment of target organ toxicity have been made, including dye viability assays, histopathology, and a neutral diffusion assay (Tice et al., 2000; Hartmann et al., 2001).

*Biological Significance of Lesions Detected* DNA lesions leading to effects in the comet assay can be strand breaks which may be relevant to the formation of chromosome aberrations or DNA modifications such as abasic (AP) sites with relevance to the induction of gene mutations. However, primary lesions detected by the comet assay may also be correctly repaired without resulting

in permanent genetic alterations. Neither the magnitude of DNA migration in the comet assay nor the shape of the comet can reveal the types of DNA damage causing the effect other biological significance, that is, their mutagenic potential. Therefore, conclusions regarding the mutagenicity of a test compound cannot be made solely on the basis of comet assay effects. There are a few limitations of the comet assay with regard to its application and interpretation of test results. For example, short-lived primary DNA lesions such as single-strand breaks, which may undergo rapid DNA repair, could be missed when using inadequate sampling times. However, an appropriate study design including only early preparation time points (i.e., at 3–6 h) is considered sufficient to ensure that these lesions are captured—in particular at higher dose levels, where DNA repair may be significantly delayed or even overwhelmed. In any case, it should be kept in mind that a negative comet result can be considered a strong indicator for the absence of a mutagenic potential.

*Advantages* The advantages of this assay for use in genotoxicity testing of drug candidates include applicability to various tissues and/or special cell types, sensitivity for detecting low levels of DNA damage, the requirement for small numbers of cells per sample, the general ease of test performance, the short time needed to complete a study, and relatively low cost. The comet assay can be applied to any tissue in the given in vivo model, provided that a single cell/nuclei suspension can be obtained. Therefore, the comet assay has potential advantages over other in vivo genotoxicity test methods, which are reliably applicable to rapidly proliferating cells only or have been validated preferentially in a single tissue only. The comet assay may detect a broader spectrum of primary DNA lesions, including single-strand breaks and oxidative-base damage, which may not be detected in the UDS (unscheduled DNA synthesis) test because they are not repaired by nucleotide excision repair. The advantages of the comet assay over the alkaline elution test include the detection of DNA damage on a single-cell level and the requirement for only small numbers of cells per sample. In contrast, when using the alkaline elution assay, large quantities of cells are necessary for the determination of genotoxic effects, and, therefore, only a limited number of organs/tissues can be evaluated using this technique. In particular, this seems important for investigation of suspected tissue-specific genotoxic activity, which includes "site-of-contact" genotoxicity (cases of high local versus low systemic exposure).

*Limitations* Experimental variability is an important issue and shoud be kept to a minimum to ensure reliable interpretation and comparability of the data obtained with other in vivo comet experiments. Experimental variability may result from shortcomings with regard to number of doses tested, number of animals per dose, number of slides per animal, number of cells analyzed, lack of sufficient DNA migration in cells of concurrent controls, and deviation from minimum time for treatment of slides with alkaline buffer. Considering

these discrepancies, data of comprehensive study as well as other study reports not in agreement with current recommendations should be interpreted with caution. This point was highlighted recently in a position paper on the use and status of the in vivo comet assay in genotoxicity testing which critically assessed published data produced under test conditions not fully in agreement with the minimal requirements for an acceptable test. For example, it was noted that positive comet assay data were published for compounds that have been assessed before to be neither genotoxic nor carcinogenic, such as food additives. Such isolated positive comet assay results should be critically evaluated in light of current recommendations to exclude methodological shortcomings and potential artifacts. In cases where negative carcinogenicity data are already available and the in vivo comet assay result represents an isolated positive finding in the context of existing genotoxicity data, the biological significance of the effect seen in the comet assay should be assessed with caution.

*Data Analysis* Data from the fluctuation test described above are analyzed by an appropriate statistical method as described in Robinson et al. (1989). Data from plate assays are analyzed as described in Arlett et al. (1989) for treat and plate tests.

***Status of Mammalian Mutation Tests*** At present the only practical assays for screening new chemical entities for mammalian mutation are the mammalian cell assays described above. The protocols are well defined, and mutant selection and counting procedures are simple and easily quantified. In general, the genetic endpoints are understood and relevant to deleterious genetic events in humans. For these reasons the assays are still regarded as valuable in safety evaluation (Li et al., 1991). It is, however, recognized that there are still unknown factors and molecular events that influence test results. This can be illustrated by the conclusions of the third UKEMS (United Kingdom Environmental Mutagen Society) collaborative trial, which focused on tests with cultured mammalian cells. The following points were made:

- The number of cells to be cultured during expression imposes a severe limitation in the use of surface-attached cells.
- A careful determination of toxicity is important.
- S9 levels may need to be varied.
- The aromatic amine benzidine is mutagenic only at the TK locus in L5178Y $TK^{+/-}$ cells. The most disturbing finding was that benzidine (detectable without metabolism by S9 mix) did not produce detectable DNA adducts (as shown by $^{32}$P-postlabelling) in L5178Y cells. Thus, the mechanism for mutagenesis in L5178Y cells benzidine remains to be elucidated (Arlett and Cole, 1990).

## 2.11 In Vivo Mammalian Mutation Tests

Mammalian mutation studies of chemicals in the whole animal have provided fundamental information on mutation parameters in germ cells such as dose response, dose fractionation, and sensitivity of various stages in gametogenesis, just as is known for ionizing radiation (Russell, 1989). This has led to estimations of the possible impact chemical mutagens may have on heritable malformation, inborn errors of metabolism, and so on. Today germ cell studies are still required when estimating the heritable damage a mutagen may inflict on exposed human populations.

The existing tests tend to be cumbersome and are not used for routine genetic toxicology screening, and thus only brief descriptions will follow. Reviews of existing data, particularly by Holden (Holden, 1982; Adler and Ashby, 1989), have indicated that most, if not all, germ cell mutagens also induce DNA damage in somatic cells, as detected by well-established assays such as the rodent micronucleus test. The converse is not true—that is, some mutagens/clastogens can induce somatic cell damage but do not induce germ cell changes, which probably reflects the special protection afforded to the germ cells, such as that provided by the blood–testis barrier. In other words, it appears that germ cell mutagens are a subset of somatic cell mutagens.

In vivo mammalian mutation tests are not restricted to germ cell tests. The mouse spot test described below first is, again, a test used for studying radiation-induced mutation but also has been used for screening chemicals for in vivo mutagenic potential. This test has had several proponents but compared with in vivo chromosomal assays is not widely used.

***Mouse-Specific Locus Test*** The mouse somatic spot test is a type of specific locus test. The classical specific locus test was developed independently by Russell at Oak Ridge in the late 1940s (Russell, 1951, 1989) and Carter in Edinburgh (Carter et al., 1956). The test consists of treatment of parental mice homozygous for a wild-type set of marker loci. The targets for mutation are the germ cells in the gonads of the treated mice. These are mated with a tester stock that is homozygous recessive at the marker loci. The $F_1$ offspring that result are normally heterozygous at the marker loci and thus express the wild-type phenotype. In the event of a mutation from the wild-type allele at any of these loci, the $F_1$ offspring express the recessive phenotype.

The test marker strain (T) developed by Russell uses seven recessive loci: *a* (nonagouti), *b* (brown), $c^{ch}$ (chinchilla), *d* (dilute), *p* (pink-eyed dilution), *s* (piebald), and *se* (short ear). As for the mouse spot test, these genes control coat pigmentation, intensity or pattern, and, for the *se* gene, the size of the external ear.

As the occurrence of mutation is rare even after mutagen treatment, the specific locus test is the ultimate study of mutation, requiring many thousands

of offspring to be scored plus significant resources of time, space, and animal husbandry. Because of these constraints, it is often difficult to define a negative result, as insufficient animals are scored or all stages of spermatogenesis are not covered. Of the 25 compounds tested in the assay, as reviewed by Ehling et al. (1986), 17 were regarded as "inconclusive" and 8 positive. The scale studies can reach is illustrated by the test of ethylene oxide described by Russell et al. (1984), where exposures of 101,000 and 150,000 ppm were used over 16–23 weeks. A total of 71,387 offspring were examined. The spermatogonial stem cell mutation rate in the treated animals did not differ significantly from the historical control frequency!

With regard to the design of the test, mice are mated when seven to eight weeks old. By this age all germ cell stages are present. The test compound is normally administered by the IP route to maximize the likelihood of germ cell exposure. The preferred dose is just below the toxic level so long as fertility is not compromised. One lower dose should also be included.

In males spermatogonia are most at risk, but it is also desirable that later stages also be exposed. Thus, the mice are mated immediately after treatment to two to four females. This is continued each week for seven weeks. Then the first group has completed its rearing of the first set of offspring and is remated. This cycle can be continued for the lifetime of the males. Tests can also be carried out by dosing females, when treatment is carried out for three weeks to cover all stages of oogenesis.

The offspring are examined immediately after birth for identification of malformations (dominant visibles) and then at weaning for the specific locus mutations. Presumptive mutant mice are checked by further crosses to confirm their status (Searle, 1984).

Comparison of mutation frequencies is made with the historical database. For definition of a positive result the same principles are recommended as for the mouse spot test (Selby and Olson, 1981). A minimum size of 18,000 offspring per group is recommended by those authors for definition of a negative result.

## 3 IN VITRO CYTOGENETIC ASSAYS

The in vitro cytogenetic assay is a short-term mutagenicity test for detecting chromosomal damage in cultured mammalian cells.

Cultured cells have a limited ability metabolically to activate some potential clastogens. This can be overcome by adding an exogenous metabolic activation system such as S9 mix to the cells (Ames et al., 1975; Madle and Obe, 1980; Natarajan and Obe, 1982; Maron and Ames, 1983).

Observations are made in metaphase cells arrested with a spindle inhibitor such as colchicine or colcemid to accumulate cells in a metaphase-like stage of mitosis (c-metaphase) before hypotonic treatment to enlarge cells and fixation with alcohol–acetic acid solution. Cells are then dispersed onto

microscope slides and stained and slides are randomized, coded, and analyzed for chromosome aberrations with high-power light microscopy. Details of the procedure are given in Dean and Danford (1984) and Preston et al. (1981, 1987). The UKEMS guidelines (Scott et al., 1990) recommend that all tests be repeated regardless of the outcome of the first test and that, if a negative or equivocal result is obtained in the first test, the repeat test should include an additional sampling time. In the earlier version of the guidelines (Scott et al., 1983) a single sampling at approximately 1.5 normal cycle times (–24 h for a 1.5-cell cycle) from the beginning of treatment was recommended, provided that a range of concentrations was used which induced marginal to substantial reductions in mitotic index, usually an indicator of mitotic delay. However, Ishidate (1988a) reported a number of chemicals which gave negative responses with a fixation time of 24 h but which were positive at 48 h. This was when a Chinese hamster fibroblast line (CHO) with a doubling time of 15 h was used. It would appear, therefore, that there are chemicals which can induce extensive mitotic delay at clastogenic doses and may be clastogenic only when cells have passed through more than one cell cycle since treatment (Thust et al., 1980). A repeat test should include an additional sample at approximately 24 h later, but it may only be necessary to score cells from the highest dose at this later fixation time. When the first test gives a clearly positive result, the repeat test need only utilize the same fixation time. The use of other sampling times is in agreement with other guidelines (European Community EEC Directive—OECD, 1983; American Society for Testing and Materials—Preston et al., 1987; Japanese guidelines—Japan Ministry of Health and Welfare (JMHW), 1984; Joint Directives, 1987; Ishidate, 1988b).

### 3.1 Cell Types

Established cell lines, cell strains, or primary cell cultures may be used. The most often used are Chinese hamster cell lines and human peripheral blood lymphocytes. The merits of these two cell lines have been reported (Ishidate and Harnois, 1987; Kirkland and Garner, 1987). The cell system must be validated and consistently sensitive to known clastogens.

### 3.2 Chinese Hamster Cell Lines

Chinese hamster ovary cells in which there has been an extensive rearrangement of chromosome material and the chromosome number may not be constant from cell to cell are frequently used. Polyploidy, endoreduplication, and high spontaneous chromosome aberration frequencies can sometimes be found in these established cell lines, but careful cell culture techniques should minimize such effects. Cells should be treated in exponential growth when cells are in all stages of the cell cycle.

### 3.3 Human Peripheral Blood Lymphocytes

Blood should be taken from healthy donors not known to be suffering from viral infections or receiving medication. Staff handling blood should be immunized against hepatitis B and regular donors should be shown to be hepatitis B antigen negative. Donors and staff should be aware of AIDS implications, and blood and cultures should be handled at containment level 2 (Advisory Committee on Dangerous Pathogens, 1984).

Peripheral blood cultures are stimulated to divide by the addition of a T-cell mitogen such as phytohemagglutinin (PHA) to the culture medium. Mitotic activity is at a maximum at about three days but begins at about 40 h after PHA stimulation and the chromosome constitution remains diploid during short-term culture (Evans and O'Riordan, 1975). Treatments should commence at about 44 h after culture initiation. This is when cells are actively proliferating and cells are in all stages of the cell cycle. They should be sampled about 20 h later. In a repeat study the second sample time should be about 92 h after culture initiation. Morimoto et al. (1983) report that the cycle time for lymphocytes averages about 12–14 h except for the first cycle.

Female donors can give higher yields of chromosome damage (Anderson et al., 1989).

### 3.4 Positive and Negative Controls

When the solvent is not the culture medium or water, the solvent, liver enzyme activation mixture and solvent, and untreated controls are used as negative controls.

Since cultured cells are normally treated in their usual growth medium, the solubility of the test material in the medium should be ascertained before testing. As pointed out earlier, extremes of pH can be clastogenic (Cifone et al., 1987), so the effect of the test material on pH should also be determined, but buffers can be utilized.

Various organic solvents are used, such as DMSO, dimethylformamide, ethanol, and acetone. The volume added must not be toxic to cells. Greater than 10% v/v water can be toxic because of nutrient dilution and osmolality changes.

A known clastogen should always be included as a positive control. When metabolic activation is used, a positive-control chemical known to require metabolic activation should also be used to ensure that the system is functioning properly. Without metabolic activation, a direct-acting positive-control chemical should be used. A structurally related positive control can also be used. Appropriate safety precautions must be taken in handling clastogens [International Agency for Research on Cancer (IARC), 1979; Medical Research Council (MRC), 1981].

Positive-control chemicals should be used to produce relatively low frequencies of aberrations so that the sensitivity of the assay for detecting weak clastogens can be established (Preston et al., 1987).

Aberration yields in negative and positive controls should be used to provide a historical database.

### 3.5 Treatment of Cells

When an exogenous activation system is employed, short treatments (about 2 h) are usually necessary because S9 mix is often cytotoxic when used for extended lengths of time. However, cells may be treated with chemicals either continuously up to harvest time or for a short time followed by washing and addition of fresh medium to allow cell cycle progression. Continuous treatment avoids centrifugation steps required with washing of cells and optimizes the endogenous metabolic capacity of the lymphocytes.

When metabolic activation is used, S9 mix should not exceed 1–10% of the culture medium by volume. It has been shown that the S9 mix is clastogenic in CHO cells and mouse lymphoma cells (Cifone et al., 1987; Kirkland et al., 1989) but not in human lymphocytes, where blood components can inactivate active oxygen species which could cause chromosome damage. When S9 mix from animals treated with other enzyme-inducing agents such as phenobarbitone/β-naphthoflavone is used, clastogenesis may be minimized (Kirkland et al., 1989).

Prior to testing, it is necessary to determine the cytotoxicity of the test material in order to select a suitable dose range for the chromosome assay both with and without metabolic activation. The range most commonly used determines the effect of the agent on the mitotic index (MI), that is, the percentage of cells in mitoses at the time of cell harvest. The highest dose should inhibit mitotic activity by approximately 50% (EEC Annex V) or 75% (UKEMS: Scott et al., 1990) or exhibit some other indication of cytotoxicity. If the reduction in MI is too great, insufficient cells can be found for chromosome analysis. Cytotoxicity can also be assessed by making cell counts in the chromosome aberration test when using cell lines. In the lymphocyte assay total white cell counts can be used in addition to MI. A dose which induces 50–75% toxicity in these assays should be accompanied by a suitable reduction in mitotic index.

If the test material is not toxic, it is recommended that it be tested up to $5\,mg\,mL^{-1}$. The UKEMS recommends that chemicals be tested up to their maximum solubility in the treatment medium and not just their maximum solubility in stock solutions.

For highly soluble nontoxic agents, concentrations above 10 mM may produce substantial increases in the osmolality of the culture medium, which could be clastogenic by causing ionic imbalance within the cells (Ishidate et al., 1984; Brusick, 1987a). At concentrations exceeding 10 mM the osmolality

of the treatment media should be measured, and if the increase exceeds 50 mmol $kg^{-1}$, clastogenicity resulting from high osmolality should be suspected and, according to the UKEMS, is unlikely to be of relevance to human risk. The UKEMS also does not recommend the testing of chemicals at concentrations exceeding their solubility limits as suspensions or precipitate.

A minimum of three doses of the test material should be used—the highest chosen as described above, the lowest on the borderline of toxicity, and an intermediate one. Up to six doses can be managed satisfactorily, and this ensures the detection of any dose response and that a toxic range is covered. MIs are as required for the preliminary study (at lease 1000 cells per culture). It is also useful to score endoreduplication and polyploidy for historical data. Cells from only three doses need to be analyzed.

The range of doses used at the repeat fixation time can be those which induce a suitable degree of mitotic inhibition at the earlier fixation time, but if the highest dose reduces the MI to an unacceptably low level at the second sampling time, the next highest dose should be chosen for screening.

A complete assay requires the test material to be investigated at a minimum of three doses together with a positive (untreated) and solvent-only control that can be omitted if tissue culture medium is used as a solvent. When two fixation times are used in repeat tests, the positive control is necessary at only one time but the negative or solvent control is necessary at both times.

Duplicates of each test group and quadruplicates of solvent or negative controls should be set up. The sensitivity of the assay is improved with larger numbers scored in the negative controls (Richardson et al., 1989).

### 3.6 Scoring Procedures

Prior to scoring, slides should be coded, randomized, and then scored "blind." Metaphase analysis should only be carried out by an experienced observer. Metaphase cells should be sought under low-power magnification and those with well-spread (i.e., non-overlapping), clearly defined nonfuzzy chromosomes examined under high power with oil immersion. It is acceptable to analyze cells with total chromosome numbers or that have lost one or two chromosomes during processing. In human lymphocytes ($2n - 46$) 44 or more centromeres and in CHO cells ($2n - 22$; range 21–24) 20 or more centromeres can be scored. Chromosome numbers can be recorded for each cell to give an indication of aneuploidy. Only cells with increases in numbers (above 46 in human lymphocytes and 24 in CHO cells) should be considered in this category, since decreases can occur through processing.

Recording microscope coordinates of cells is necessary and allows verification of abnormal cells. A photographic record is also useful of cells with aberrations. Two hundred cells (100 from each of two replicates) should be scored per treatment group. When ambiguous results are obtained, there may be further "blind" reading of these samples.

### 3.7 Data Recording

The classification and nomenclature of the International System for Human Cytogenetic Nomenclature (ISCN, 1985) as applied to acquired chromosome aberrations are recommended. Score sheets giving the slide code, microscope scorer's name, date, cell number, number of chromosomes, and aberration types should be used. These should include chromatid and chromosome gaps, deletions, exchanges, and others. A space for the vernier reading for comments and a diagram of the aberration should be available.

From the score sheets, frequencies of various aberrations should be calculated and each aberration should be counted only once. To consider a break as one event and an exchange as two events is not acceptable, since unfounded assumptions are made about mechanisms involved (Revell, 1974).

### 3.8 Presentation of Results

The test material, test cells used, method of treatment, harvesting of cells, cytotoxicity assay, and so on, should be clearly stated as well as the statistical methods used. Richardson et al. (1989) recommend that comparison be made between the frequencies in control cells and those at each dose level using Fisher's exact test.

In cytogenetic assays the absence of a clear positive dose–response relationship at a particular time frequently arises. This is because a single common sampling time may be used for all doses of a test compound. Chromosome aberration yields can vary markedly with posttreatment sampling time of an asynchronous population, and increasing doses of clastogens can induce increasing degrees of mitotic delay (Scott et al., 1990). Additional fixation times should clarify the relationship between dose and aberration yield.

Gaps are by tradition excluded from quantification of chromosome aberration yields. Some gaps have been shown to be real discontinuities in DNA (e.g., Heddle and Bodycote, 1970). Where chromosome aberration yields are on the borderline of statistical significance above control values, the inclusion of gaps could be useful. Further details on this approach may be found in the UKEMS guidelines (Scott et al., 1990).

Since chromosome exchanges are relatively rare events, greater biological significance should be attached to their presence than to gaps and breaks.

Chemicals which are clastogenic in vitro at low doses are more likely to be clastogenic in vivo than those where clastogenicity is detected only at high concentrations (Ishidate et al., 1988). Negative results in well-conducted in vitro tests are a good indication of a lack of potential for in vivo clastogenesis, since almost all in vivo clastogens have given positive results in vitro when adequately tested (Thompson, 1986; Ishidate et al., 1988).

## 4 IN VIVO CYTOGENETICS ASSAYS

Damage induced in whole animals can be detected in in vivo chromosome assays in either somatic or germinal cells by examination of metaphases or the formation of micronuclei. The micronucleus test can also detect whole chromosome loss or aneuploidy in the absence of clastogenic activity and is considered comparable in sensitivity to chromosome analysis (Tsuchimoto and Matter, 1979).

Rats and mice are generally used for in vivo studies, with the mouse being employed for bone marrow micronucleus analysis and the rat for metaphase analysis, but both can be used for either. Mice are cheaper and easier to handle than rats, and only a qualitative difference in response has been found between the species (Albanese et al., 1988). Chinese hamsters are also widely used for metaphase analysis because of their low diploid chromosome number of 22. However, there are few other historical toxicological data for this species.

### 4.1 Somatic Cell Assays

***Metaphase Analysis*** Metaphase analysis can be performed in any tissue with actively dividing cells, but bone marrow is the tissue most often examined. Cells are treated with a test compound and are arrested in metaphase by the administration of colcemid or colchicine at various sampling times after treatment. Preparations are examined for structural chromosome damage. Because the bone marrow has a good blood supply, the cells should be exposed to the test compound or its metabolites in the peripheral blood supply, and the cells are sensitive to S-dependent and S-independent mutagens (Topham et al., 1983).

Peripheral blood cells can be stimulated to divide even though the target cell is relatively insensitive (Newton and Lilly, 1986). It is necessary to stimulate them with a mitogen since the number of lymphocytes which are dividing at any one time is very low. Cells are in $G_0$ when exposure is taking place, so they may not be sensitive to cell cycle stage specific mutagens and any damage might be repaired before sampling.

***Micronuclei*** The assessment of micronuclei is considered simpler than the assessment of metaphase analysis (Collaborative Study Group, 1986, 1988). This assay is most often carried out in bone marrow cells, where polychromatic erythrocytes are examined. Damage is induced in the immature erythroblast and results in a micronucleus outside the main nucleus, which is easily detected after staining as a chromatid-containing body. When the erythroblast matures, the micronucleus, whose formation results from chromosome loss during cell division or from chromosome breakage forming centric and acentric fragments, is not extruded with the nucleus. Micronuclei can also be detected in peripheral blood cells (MacGregor et al., 1980). In addition, they can be detected in liver

(Tates et al., 1980; Braithwaite and Ashby, 1988) after partial hepatectomy or stimulation with 4-acetylaminofluorene or they can be detected in any proliferating cells.

### 4.2 Germ Cell Assays

The study of chromosome damage is highly relevant to the assessment of heritable cytogenetic damage. Many compounds which cause somatic cell damage have not produced germ cell damage (Holden, 1982) and, so far, all germ mutagens have also produced somatic damage.

Germ cell data, however, are needed for genetic risk estimation, and testing can be performed in male or female germ cells. The former are most often used, owing to their systemic effects in females. Testing in the male is performed in mitotically proliferating premeiotic spermatogonia, but chromosomal errors in such cells can result in cell death or prevent the cell from passing through meiosis. Damage produced in postmeiotic cells, the spermatids, or sperm are more likely to be transmitted to the $F_1$ progeny (Albanese, 1987). In females it is during early fetal development of the ovary that oocyte stage is the most commonly tested in the adult female. To test other stages during the first or second meiotic divisions demands the use of oocytes undergoing ovulation, which occurs naturally or is hormone stimulated. It is thus more difficult technically to test female germ cells.

### 4.3 Heritable Chromosome Assays

Damage may be analyzed in the heritable translocation test, which involves the examination in male $F_1$ animals if diakinesis metaphase 1 spermatocytes for multivalent association fall within the acceptable range for the laboratory for a substance to be considered positive or negative under the conditions of the study.

### 4.4 Germ Cell Cytogenetic Assays

Either mouse or rat can be used, but the mouse is generally the preferred species. Normally such assays are not conducted for routine screening purposes.

Spermatogonial metaphases can be prepared by the air-drying technique of Evans et al. (1964) for the first and second meiotic metaphase (MI and MII) in the male mouse. This method is not so suitable for rat and hamster. The numbers of spermatogonial metaphases can be boosted if, prior to hypotonic treatment, the testicular tubules are dispersed in trypsin solution (0.25%). At least one month between treatment and sample should be allowed to pass in the mouse to allow treated cells to reach meiosis. Brook and Chandley (1986) established that 11 days and 4 h was required for spermatogonial cells to reach preleptotene and 8 days and 10 h to reach zygotene. It takes 4 h for cells to

move from MI to MII, but test compounds can alter this rate. A search for multivalent formation can be made at MI for the structural rearrangements induced in spermatogonia. Cawood and Breckon (1983) examined the synaptonemal complex at pachytene using electron microscopy. Errors of segregation should be searched for at the first meiotic division in the male mouse, MII cells showing 19 (hypoploid) and 21 (hyperploid) chromosomes (Brook and Chandley, 1986). Hansmann and El-Nahass (1979), Brook (1982), and Brook and Chandley (1985) describe assays in the female mouse and procedures used for inducing ovulation by hormones and treatment of specific stages of meiosis.

## 5 SISTER CHROMATID EXCHANGE ASSAYS

SCEs are reciprocal exchanges between sister chromatids. They result in a change in morphology of the chromosome, but breakage and reunion are involved although the exact mechanism is unclear. They are thought to occur at homologous loci.

In 1958 Taylor demonstrated SCEs using autoradiographic techniques to detect the disposition or labeled DNA following incorporation of [$^3$H]-thymidine. 5-Bromo-2′-deoxyuridine (drdU) has now replaced [$^3$H]-thymidine and various staining methods have been used to show the differential incorporation of BrdU between sister chromatids: fluorescent—Hoechst 33258 (Latt, 1973), combined fluorescent and Giemsa (Perry and Wolff, 1974), and Giemsa (Korenberg and Freedlender, 1974). The fluorescent plus Giemsa procedure is recommended in view of the fact that stained slides can be stored and microscope analysis is simpler.

So that SCEs can be seen at metaphase, cells must pass through S phase (Kato, 1973; 1974; Wolff and Perry, 1974). SCEs appear to occur at the replication point, since SCE induction is maximal at the beginning of DNA synthesis but drops to zero at the end of the S phase (Latt and Loveday, 1978).

For SCE analysis in vitro, any cell type that is replicating or can be stimulated to divide is suitable. The incorporation of BrdU into cells in vivo allows the examination of a variety of tissues (Latt et al., 1980). Edwards et al. (1993) suggest that it is necessary to standardize protocols measuring SCE since different responses can be obtained depending on the extent of simultaneous exposure of test compound and BrdU.

### 5.1 Relevance of SCE in Terms of Genotoxicity

SCEs do not appear to be related to other cytogenetic events, since potent clastogens such as bleomycin and ionizing radiation induce low levels of SCE (Perry and Evans, 1975). The mechanisms involved in chromosome aberrations and SCE formation are dissimilar (e.g., Galloway and Wolff, 1979). There is not evidence that SCEs are in themselves lethal events, since there is little relationship to cytotoxicity (e.g., Bowden et al., 1979). It was suggested by

Wolff (1977a,b) that they relate more to mutational events due to a compatibility with cell survival. However, there are examples of agents that induce significant SCE increases in the absence of mutation (Bradley et al., 1979) as well as the converse (Connell, 1979; Connell and Medcalf, 1982).

The SCE assay is particularly sensitive for alkylating agents and base analogues, agents causing single-strand breaks in DNA, and compounds acting through DNA binding (Latt et al., 1981). The most potent SCE inducers are S-phase dependent. Painter (1980) reports that agents such as X-irradiation, which inhibits replicon initiation, are poor SCE inducers, whereas mitomycin C, which inhibits replication fork progression, is a potent SCE inducer.

## 5.2 Experimental Design

Established cell lines, primarily cell cultures of rodents, may be used. Detailed information on in vitro and in vivo assays may be obtained in reviews of SCE methods by Latt et al. (1977, 1981), Perry and Thomson (1984), and Perry et al. (1984). The in vitro methods will be briefly explored here.

Either monolayer or suspension cultures or human lymphocytes can be employed. Human fibroblasts are less suitable because of their long cell cycle duration.

The concentration of organic solvents for the test compound should not exceed 0.8% v/v, as higher concentrations could lead to slight elevations in the SCE level (Perry et al., 1984).

For monolayer cultures, the cultures are set up the day before BrdU treatment so that the cells will be in exponential growth before the addition of BrdU or the test compound. After BrdU addition the cells are allowed to undergo the equivalent of two cell cycles before cell harvest. A spindle inhibitor such as colchicine or colcemid is introduced for the final 1–2 h of culture to arrest cells in metaphase, after which the cells are harvested and chromosome preparations are made by routine cytogenetic techniques.

In the absence of metabolic activation, BrdU and the test agent can be added simultaneously and left for the duration of BrdU labeling. Shorter treatments should be used in the presence of metabolic activation or to avoid synergistic effects with BrdU, when cells can be pulse treated for, for example, 1 h before BrdU addition (see Edwards et al., 1993).

Peripheral blood cultures are established in medium containing BrdU and PHA. Colcemid is added 1–2 h before harvest and the cells are harvested between 60 and 70 h post-PHA-stimulation. Cell harvest and slide preparations are conducted according to routine cytogenetic methods.

Heparinized blood samples may be stored at 4 °C for up to 48 h without affecting the SCE response (Lambert et al., 1982). If the test agent is known to react with serum or red blood cells, the mononuclear lymphocytes may be isolated by use of a Ficoll/Hypaque gradient (Boyum, 1968).

If metabolic activation is not required, treatment is best conducted over the whole of the final 24 h of culture, or if metabolic activation is required, a pulse exposure may be employed to treat cultures at the first S phase at around 24–30 h or at 48 h for an asynchronous population.

Exposure of cells to fluorescent light during the culture period leads to photolysis of BrdU-containing DNA and a concomitant increase in SCE frequency (Wolff and Perry, 1974). Consequently, SCE cultures should be kept in the dark and manipulated under subdued light conditions such as yellow safe light. Furthermore, media used in SCE assays should be stored in the dark, since certain media components produce reactive SCE-inducing intermediates on exposure to fluorescent light (Monticone and Schneider, 1979).

Coded and randomized slides should be read. All experiments should be repeated at least once (Perry et al., 1984) with higher and lower concentrations of S9 mix if a negative response is achieved. Even for an apparently unambiguous positive response with a greater than twofold increase in SCEs over the background level at the highest dose and with at least two consecutive dose levels with an increased SCE response, a repeat study is necessary to show a consistent response.

The quality of differential staining will determine the ease and accuracy of SCE scoring, and, to eliminate variation, results from different observers should occasionally be compared. Furthermore, to avoid observer bias, scorers should have slides from different treatment groups equally distributed among them, as with all cytogenetic studies.

***Issues in Assay Interpretation and Relevance*** The reason for consideration of a significant revision to S2 (such as is now under consideration) ties into issues and considerations as to perceived unacceptable error rates in test performance and difficulties in interpreting the relevance of findings. Supplement in vivo genotoxicity studies are used to (1) follow up on positive finding in one or more tests of the standard genotoxicity battery, (2) elucidate a potential contribution of genotoxicity to the induction of preneoplastic and/or neoplastic changes detected in long-term tests in rodents, and (3) elucidate mechanisms of micronucleus formation to differentiate clastogenic from aneuogenic effects since aneugenicity is well accepted to result from mechanisms of action for which thresholds exist, demonstrating that micronucleus formation as a result of chromosome loss should allow an acceptable level of human exposure to be defined. No matter the trigger for conducting supplemental in vivo genotoxicity testing, it is critical that the approach utilized—for example, the endpoint and target tissue assessed—is scientifically valuable such that the results will aid in interpreting the relevance of the initial finding of concern. Ultimately, the goal of supplemental genotoxicity testing is to determine if a genotoxic risk is posed to patients under the intended condition of treatment.

*Follow-Up Testing of Drug Candidates Positive in Standard Genotoxicity Test Battery* It has been reported that approximately 30–50% of pharmaceuticals produce positive genotoxicity results in vitro (Kirkland and Muller, 2000). In contrast, results from bone marrow cytogenetic assays are frequently negative, even for those compounds that produce positive results in vitro. This discrepancy may result from a number of major differences that exist when testing in cultured cells versus intact animals. For example, differing metabolic pathways can exist in vitro and in vivo, metabolic inactivation can occur in the intact animal, parent compound or active metabolite may not reach the target cell in vivo, rapid detoxification and elimination may occur, or plasma levels in vivo may not be comparable to concentrations that generate positive responses in the in vitro assay, which is often accompanied by high levels of cytotoxicity. It is also worth nothing that positive results generated in vitro may be secondary to effects, such as cytotoxicity, which may never be achieved under in vivo exposure conditions. Data from in vivo experiments are therefore essential before definitive conclusions are drawn regarding the potential mutagenic hazard to humans from chemicals that produce positive results in one or more in vitro tests.

*Follow-Up Testing of Tumorigenic Drug Candidates Negative in Standard Genotoxicity Test Battery* In carcinogenicity testing of pharmaceutical drug candidates of tumorigenic response in rodents, the ICH guidance S2B currently stipulates that such tumorigenicity is not clearly based on a nongenotoxic mechanism. Typically, supplemental in vivo genotoxicity tests should be performed with cells of the respective tumor target organ to distinguish between genotoxic and nongenotoxic mechanisms of tumor induction.

*Endpoints Assessed in Supplemental Assays* Commonly applied test systems are described that are used as supplemental genotoxicity assays. These assays differ with respect to the endpoints assessed:

1. Induction of primary DNA lesions, that is, measurement of exposure, uptake, and reactivity to DNA via the comet assay or P-postlabeling assay
2. Measurement of the repair of DNA lesion using the in-scheduled DNA synthesis (UDS) test
3. Measurement of induction of genetic drug transgenic animal assays for point mutations or the mouse spot test

The comet assay is the most commonly applied of these.

The issue of how to assess the relevance of a finding of genotoxicity in a candidate drug is a complex one. If the judgment is that such a finding is not relevant to human risks, there are two approaches to assessing and defend-

ing such a finding: the WOE (weight of evidence) or the MOA (made of action).

In the case of impurities in a marketed or candidate drug, the approach is more straightforward. One must reduce the level to or below the TTC (toxicological threshold of concern), the level at which no patient would receive more than 1.5 μg $day^{-1}$ (EMEA, 2004).

## REFERENCES


Adler, I. D., and Ashby, J. (1989). The present lack of evidence for unique rodent germ-cell mutagens. *Mutat. Res.* 212:55–66.

Albanese, R. (1987). Mammalian male germ cell cytogenetics. *Mutagenesis* 2:79–85.

Albanese, R., Mirkova, E., Gatehouse, D., and Ashby, J. (1988). Species-specific response to the rodent carcinogens 1,2-dimethylhydrazine and 1,2-dibromochloropropane in rodent bone marrow micronucleus assays. *Mutagenesis* 3:35–38.

Amacher, D. E., Paillet, S. C., Turner, G. N., Ray, V. A., and Salsburg, D. S. (1980). Point mutations at the thymidine kinase locus in L5178Y mouse lymphoma cells. 2. Test validation and interpretation. *Mutat. Res.* 72:447–474.

Ames, B. N. (1971). The detection of chemical mutagens with enteric bacteria. In A. Hollaender (Ed.), *Chemical Mutagens, Principles and Methods for Their Detection*, Vol. 1. Plenum, New York, pp. 267–282.

Ames, B. N., McCann, J., and Yamasaki, E. (1975). Methods for detecting carcinogens and mutagens with the Salmonella/mammalian-microsome mutagenicity test. *Mutat. Res.* 31:237–364.

Anderson, D. (1990). Male mediated $F_1$ abnormalities, *Mutat. Res.* 229:103–246.

Anderson, D., Jenkinson, P. C., Dewdney, R. S., Francis, A. T., Godbert, P., and Butterworth, K. R. (1989). Chromosome aberrations, mitogen induced blastogenesis and proliferative rate index in peripheral lymphocytes from 106 control individuals in the UK population. *Mutat. Res.* 204:407–420.

Applegate, M. L., Moore, M. M., Broder, C. B., et al. (1990). Molecular dissection of mutations at the heterozygous thymidine canise locus in mouse lymphoma cells. *Proc. Natl. Acad. Sci. USA* 87:51–55.

Arlett, C. F., and Cole, J. (1990). The third United Kingdom Environmental Mutagen Society collaborative trial: Overview, a summary and assessment. *Mutagenesis* 5(Suppl):85–88.

Arlett, C. F., Smith, D. M., Clark, G. M., Green, J. H. L., Cole, J., McGregor, D. B., and Asquith, J. C. (1989). Mammalian cell assays based upon colony formation. In D. J. Kirkland (Ed.), *UKEMS Subcommittee on Guidelines for Mutagenicity Testing. Report Part III: Statistical Evaluation of Mutagenicity Test Data*. Cambridge University Press, Cambridge, pp. 66–101.

Ashby, J., and Tennant, R. W. (1988). Chemical structure, *Salmonella* mutagenicity and extent of carcinogenicity as indices of genotoxic carcinogens among 222 chemicals tested in rodents by the US NCI/NTP. *Mutat. Res.* 204:17–115.

Ashby, J., Tennant, R. W., Zeiger, E., and Stasiewiczs, S. (1989). Classification according to chemical structure, mutagenicity to *Salmonella* and level of carcinogenicity of a further 42 chemicals tested for carcinogenicity by the U.S. National Toxicology Program. *Mutat. Res.* 223:73–104.

Ashwood-Smith, M. J. (1980). Stability of frozen microsome preparations for use in the Ames Salmonella mutagenicity assay. *Mutat. Res.* 69:199–200.

Bartsch, H., Camus, A.-M., and Malaveille, C. (1984). Comparative mutagenicity of *N*-nitrosamines in a semi-solid and in a liquid incubation system in the presence of rat or human tissue fractions. *Mutat. Res.* 37:149–162.

Beckman, D. A., and Brent, R. L. (1986). Mechanism of known environmental teratogens: Drugs and chemicals. *Clin. Perinatol.* 13:649–687.

Bender, M. A., Griggs, H. G., and Bedford, J. S. (1974). Mechanisms of chromosomal aberration production. III. Chemicals and ionizing radiation. *Mutat. Res.* 23:197–212.

Blazak, W. F., Steward, B. E., Galperin, I., Allen, K. L., Rudd, C. J., Mitchell, A. D., and Caspary, W. J. (1986). Chromosome analysis of trifluorothymidine-resistant L5178Y mouse lymphoma cells colonies. *Environ. Mutagen.* 8:229–240.

Bossi, L. (1985). Information suppression. In J. Scaife, D. Leach, and A. Galizzi (Eds.), *Genetics of Bacteria*. Academic, New York, pp. 49–64.

Bowden, G. T., Hsu, I. C., and Harris, C. C. (1979). The effect of caffeine on cytotoxicity, mutagenesis and sister chromatid exchanges in Chinese hamster cells treated with dihydrodiol epoxide derivatives of benzo(a)pyrene. *Mutat. Res.* 63:361–370.

Boyum, A. (1968). Separation of lymphocytes and erythrocytes by centrifugation. *Scand. J. Clin. Invest.* 21:77–85.

Bradley, M. O., Hsu, I. C., and Harris, C. C. (1979). Relationships between sister chromatid exchange and mutagenicity, toxicity and DNA damage. *Nature* 282:318–320.

Braithwaite, I., and Ashby, J. (1988). A non-invasive micronucleus assay in rat liver. *Mutat. Res.* 203:23–32.

Brendler-Schwaab, S. Y., Schmezer, P., Liegibel, U., Weber, S., Michalek, K., Tompa, A., and Pool-Zobel, B. L. (1994). Cells of different tissues for in vitro and in vivo studies in toxicology: compilation of isolation methods. *Toxicol. In Vitro* 8:1285–1302.

Brendler-Schwaab, S., Hartmann, A., Pfuhler, S., and Speit, G. (2005). The in vivo comet assay: Use and status in genotoxicity testing. *Mutagenesis* 20:245–254.

Bridges, B. A. (1972). Simple bacterial systems for detecting mutagenic agents. *Lab. Pract.* 21:413–419.

Bridges, B. A. (1976). Short-term screening tests for carcinogens. *Nature* 261:195–200.

Bridges, B. A., Woodgate, R., Ruiz-Rubio, M., Sharif, F., Sedgwick, S. G., and Hubschere, U. (1987). Current understanding of UV-induced base pair substitution mutation in *E. coli* with particular reference to the DNA polymerase III complex. *Mutat. Res.* 181:219–226.

Brook, J. D. (1982). The effect of 4CMB on germ cells of the mouse. *Mutat. Res.* 100:305–308.

Brook, J. D., and Chandley, A. C. (1985). Testing of 3 chemical compounds for aneuploidy induction in the female mouse. *Mutat. Res.* 157:215–220.

Brook, J. D., and Chandley, A. C. (1986). Testing for the chemical induction of aneuploidy in the male mouse. *Mutat. Res.* 164:117–125.

Brusick, D. (1986). Genotoxic effects in cultures mammalian cells produced by low pH treatment conditions and increased ion concentrations. *Environ. Mutagen.* 8:879–886.

Brusick, D. (1987a). Genotoxicity produced in cultures mammalian cell assays by treatment conditions. Special issue. *Mutat. Res.* 189:1–80.

Brusick, D. (1987b). *Principles of Genetic Toxicology*, 2nd ed. Plenum, New York, London.

Burlinson, B., Tice, R. R., Speit, G., Agurell, E., Brendler-Schwaab, S. Y., Collins, A. R., Escobar, P., Honma, M., Kumaravel, T. S., Nakajima, M., Sasaki, Y. F., Thybaud, V., Uno, Y., Vasquez, M., Hartmann, A. In Vivo Comet Assay Workgroup, part of the Fourth International Workgroup on Genotoxicity Testing: results of the in vivo Comet Assay.

Burnett, C., Fuchs, C., Corbett, J., and Menkart, J. (1982). The effect of dimethylsulphoxide on the mutagenicity of the hair-dye, *p*-phenylenediamine. *Mutat. Res.* 103:1–4.

Butterworth, B. E., and Slaga, T. J. (1987). *Nongenotoxic Mechanisms in Carcinogenesis*. Banbury Report No. 25. Cold Spring Harbor Laboratory, New York.

Carter, T. C., Lyon, M. F., and Philips, R. J. S. (1956). Induction of mutations in mice by chronic gamma irradiation; interim report. *Br. J. Radiol.* 29:106–108.

Carver, J. H., Machado, M. L., and MacGregor, J. A. (1985). Petroleum distillates suppress in vitro metabolic activation: Higher (S9). Required in the *Salmonella*/microsome mutagenicity assay. *Environ. Mutagen.* 7:369–380.

Cawood, A. D., and Breckon, G. (1983). Synaptonemal complexes as indicators of induced structural change in chromosomes after irradiation of spermatogonia. *Mutat. Res.* 122:149–154.

Chandley, A. C. (1981). The origin of chromosomal aberrations in man and their potential for survival and reproduction in the adult human population. *Ann. Genet.* 24:5–11.

Chu, E. H. Y., and Malling, H. U. (1968). Mammalian cell genetics. II. Chemical induction of specific locus mutations in Chinese hamster cells *in vitro*. *Proc. Natl. Acad. Sci. USA* 61:1306–1312.

Cifone, M. A., Myhr, B., Eiche, A., and Bolisfodi, G. (1987). Effect of pH shifts on the mutant frequency at the thymidine kinase locus in mouse lymphoma L5178Y TK+/– cells. *Mutat. Res.* 189:39–46.

Clayson, D. B. (1989). ICPEMC publication No. 17: Can a mechanistic rationale be provided for non-genotoxic carcinogens identified in rodent bioassays? *Mutat. Res.* 221:53–67.

Cleaver, J. E. (1983). Xeroderma pigmentosum. In J. B. Stanbury, J. B. Wyngaarden, D. S. Fredrickson, J. C. Goldstein, and M. S. Brown (Eds.), *The Metabolic Basis of Inherited Disease*. McGraw-Hill, New York, pp. 1227–1248.

Clive, D. (1987). Historical overview of the mouse lymphoma $TK^{+/-}$ mutagenicity assay. In M. M. Moore, D. M. Demarini, F. J. De Serres, and K. R. Tindall (Eds), *Mammalian Cell Mutagenesis*. Banbury Report 28. Cold Spring Harbor Laboratory, New York, pp. 25–36.

Clive, D. (1989). Mammalian cell genotoxicity: A major role for non-DNA targets? *Mutat. Res.* 223:327–328.

Clive, D., Caspary, W., Kirkby, P. E., Krehl, R., Moore, M., Mayo, J., and Oberly, T. J. (1987). Guide for performing the mouse lymphoma assay for mammalian cell mutagenicity. *Mutat. Res.* 189:145–146.

Clive, D., Flamm, W. G., and Patterson, J. B. (1972). A mutational assay system using the thymidine kinase locus in mouse lymphoma cells. *Mutat. Res.* 16:77–87.

Cole, J., Fox, M., Garner, R. C., McGregor, D. B., and Thacker, J. (1990). Gene mutation assays in cultured mammalian cells. In D. J. Kirkland (Ed.), *UKEMS Subcommittee on Guidelines for Mutagenicity Testing. Report Part I Revised*. Cambridge University Press, Cambridge, pp. 87–114.

Cole, J., Muriel, W. J., and Bridges, B. A. (1986). The mutagenicity of sodium fluoride to L5178Y (wildtype and $TK^{+/-}$ 3.7.2c) mouse lymphoma cells. *Mutagenesis* 1:157–167.

Collaborative Study Group for the Micronucleus Test (1986). Sex differences in the micronucleus test. *Mutat. Res.* 172:151–163.

Collaborative Study Group for the Micronucleus Test (1988). Strain differences in the micronucleus test. *Mutat. Res.* 204:307–316.

Collins, A. R., Duthie, S. J., Dobson, V. L. (1993). Direct enzymic detection of endogenous oxidative base damage in human lymphocyte DNA. *Carcinogenesis* 14:1733–1735.

Connell, J. R. (1979). The relationship between sister chromatid exchange, chromosome aberration and gene mutation induction by several reactive polycyclic hydrocarbon metabolites in cultured mammalian cells. *Int. J. Cancer* 24:485–489.

Connell, J. R., and Medcalf, A. S. (1982). The induction of SCE and chromosomal aberrations with relation to specific base methylation of DNA in Chinese hamster cells by *N*-methyl-*n*-nitrosourea and dimethyl sulphate. *Carcinogenesis* 3:385–390.

Cullum, J. (1985). Insertion sequences. In Scarfe, J., Leach, D., and Galizzi, A. (Eds.), *Genetics of Bacteria*, Academic Press, New York, pp. 85–94.

Dean, B. J., and Danford, N. (1984). Assays for the detection of chemically-induced chromosome damage in cultured mammalian cells. In S. Venitt and J. M. Parry (Eds.), *Mutagenicity Testing. A Practical Approach*. IRL Press, Oxford, pp. 187–232.

DeSerres, F. J., and Shelby, M. D. (1979). Recommendations on data production and analysis using the *Salmonella*/microsome mutagenicity assay. *Mutat. Res.* 64:159–165.

deVries, J. H. (1901). *The Mutation Theory*. Verlag von Veit & Co., Leipzig.

Doll, R. (1977). Strategy for detection of cancer hazards to man. *Nature* 265:589–596.

Dunnett, C. W. (1955). A multiple comparison procedure for comparing several treatments with a control. *J. Am. Stat. Assoc.* 50:1096–1121.

Edwards, A. J., Moon, E. Y., anderson, D., and McGregor, D. B. (1993). The effect of simultaneous exposure to bromodeoxyuridine and methyl methanesulphonate on sister chromatid exchange frequency in culture human lymphocytes and its mutagenicity. *Mutat. Res.* 247:117–125.

Ehling, U. H., Chu, E. H. Y., DeCarli, L., Evans, H. J., Hayashi, M., Lambert, B., Neubert, D., Thilly, W. G., and Vainio, H. (1986). Report 8. Assays for germ-cell mutations in mammals. In R. Montesano, H. Bartsch, H. Vainio, J. Wilbourn, and H. Yamasaki (Eds.), *Long-Term and Short-Term Assays for Carcinogens. A Critical Appraisal*. No. 83. IARC Scientific Publications, Lyon, pp. 245–265.

EMEA, (2004) Guideline on the Limits of Genotoxic Impurities in Pharmaceuticals.

Evans, H. H., Mencl, J., Horng, M. F., Ricanti, M., Sanchez, D., and Hozier, J. (1986). Locus specificity in the mutability of L5178Y mouse lymphoma cells: The role of multilocus lesions. *Proc. Natl. Acad. Sci. USA* 83:4379–4385.

Evans, H. J., and O'Riordan, M. L. (1975). Human peripheral blood lymphocytes for the analysis of chromosome aberrations in mutagen tests. *Mutat. Res.* 31:135–148.

Ford, D. K., and Yerganian, G. (1958). Observations on the chromosomes of Chinese hamster cells in tissue culture. *J. Natl. Cancer Inst.* 21:393–425.

Forster, R., Green, M. H. L., and Priestley, A. (1980). Optimal levels of S9 fraction in Ames and fluctuation tests: Apparent importance of diffusion of metabolites from top agar. *Carcinogenesis* 2:1081–1085.

Fox, M. (1981). Some quantitative aspects of the response of mammalian in vitro to induced mutagenesis. In J. J. Marchelonis and M. G. Hanna (Eds.), *Cancer Biology Reviews* Vol. 3. Marcel Dekker, New York, pp. 23–62.

Friedberg, E. C. (1988). DNA repair in the yeast *Saccharomyces cerevisiae*. *Microb. Rev.* 52:70–102.

Gad, S. C. (1999). *Statistics and Experimental Design for Toxicologists*, 3rd ed. CRC Press, Boca Raton, FL.

Galloway, S. M., and Wolff, S. (1979). The relation between chemically induced SCE's and chromatid breakage. *Mutat. Res.* 61:297–307.

Gatehouse, D. G., and Delow, G. F. (1979). The development of a "Microtitre$^{R}$" fluctuation test for the detection of indirect mutagens and its use in the evaluation of mixed enzyme induction of the liver. *Mutat. Res.* 60:239–252.

Gatehouse, D. G., and Tweats, D. J. (1987). Letter to the Editor. *Mutagenesis* 1:307–308.

Gatehouse, D., and Wedd, D. J. (1984). The differential mutagenicity of isoniazid in fluctuation assays and *Salmonella* plate tests. *Carcinogenesis* 5:391–397.

Gatehouse, D. G., Wedd, D. J., Paes, D., Delow, G., Burlinson, B., Pascoe, S., Brice, A., Stemp, G., and Tweats, D. J. (1988). Investigations into the genotoxic potential of Ioxtidine, a long-acting $H_2$-receptor antagonist. *Mutagenesis* 3:57–68.

Gatehouse, D. G., Wilcox, P., Forster, R., Rowland, I. R., and Callander, R. D. (1990). Bacterial mutation assays. In D. J. Kirkland (Ed.), *Basic Mutagenicity Tests: UKEMS Recommended Procedures*. Cambridge University Press, Cambridge, pp. 13–61.

Green, M. H. L., and Muriel, W. J. (1976). Mutagen testing using $TRP^{+}$ reversion in *E. coli*. *Mutat. Res.* 38:3–32.

Hansmann, I., and El-Nahass, E. (1979). Incidence of non-disjunction in mouse öocytes. *Cytogenet. Cell Genet.* 24:115–121.

Harris, C. C., Weston, A., Willey, J. C., Trivers, G. E., and Mann, D. L. (1987). Biochemical and molecular epidemiology of human cancer: Indicators of carcinogen exposure, DNA damage and genetic predisposition. *Environ. Health Perspect.* 75:109–119.

Hartmann, A., Kiskinis, E., Fjaellman, A., and Suter, W. (2001). Influence of cytotoxicity and compound precipitation on test results in the alkaline comet assay. *Mutat. Res.* 497:199–212.

Hartmann, A., Agurell, E., Beevers, C., Brendler-Schwaab, S., Burlinson, B., Clay, P., Collins, A., Smith, A., Speit, G., Thybaud, V., Tice, R. R. (2007). 4th workgroup. *Mutat. Res.* 627:31–35.

Haworth, S., Lawlor, T., Mortelmanns, K., Speck,W., and Zeiger, E. (1983). Salmonella mutagenicity results for 250 chemicals. *Environ. Mutagen. Suppl.* 1:3–142.

Heddle, J. A., and Bodycote, D. J. (1970). On the formation of chromosomal aberrations. *Mutat. Res.* 9:117–126.

Herrera, G., Urios, A., Alexandre, V., and Blanco, M. (1988). UV light induced mutability in *Salmonella* strains containing the *umu DC* or the *muc AB* operon: Evidence for a *umu C* function. *Mutat. Res.* 198:9–13.

Holden, H. E. (1982). Comparison of somatic and germ cell models for cytogenetic screening. *J. Appl. Toxicol.* 2:196–200.

Hozier, J., Sawger, D., Clieve, D., and Moore, M. (1982). Cytogenetic distinction between the $TK^+$ and $TK^-$ chromosomes in L5178Y/$TK^{+/-}$-3.7.2.C mouse lymphoma cell line. *Mutat. Res.* 105:451–456.

Hsie, A. W., Brimer, P. A., Mitchell, T. J., and Gosslee, D. G. (1975). The dose-response relationship for ethyl methane sulfonate-induced mutation at the hypoxanthine-guanine phosphoribosyl transferase locus in Chinese hamster ovary cells. *Somatic Cell Genet.* 1:247–261.

Hubbard, S. A., Brooks, T. M., Gonzalez, L. P., and Bridges, J. W. (1985). Preparation and characterization of S9-fractions. In J. M. Parry and C. F. Arlett (Eds.), *Comparative Genetic Toxicology*. Macmillan, London, pp. 413–438.

International Agency for Research on Cancer (IARC) (1979). *Handling Chemical Carcinogens in the Laboratory; Problems of Safety*. Scientific Publications, No. 33. IARC, Lyons, France.

International Commission for Protection Against Environmental Mutagens and Carcinogens (ICEMC) (1983). Committee & Final Report: Screening strategy for chemicals that are potential germ-cell mutagens in mammals. *Mutat. Res.* 114:117–177.

International Conference on Harmonization (ICH) (1996). Technical requirements for registration of pharmaceuticals for human use. Guidance on specific aspects of regulatory genotoxicity tests for pharmaceuticals. S2A document recommended for adoption at step 4 of the ICH process on July 19, 1995. *Federal Register* 61(Apr. 24):18198–18202.

International Conference on Harmonization (ICH) (1997). Technical requirements for registration of pharmaceuticals for human use. Genotoxicity: A standard battery for genotoxicity testing of pharmaceuticals. S2B.

International Conference on Harmonization (ICH) (2008). Genotoxicity leveling and data interpretation for human use (Step 2, March 6, 2008). S2(R1).

International System for Human Cytogenetic Nomenclature (ISCN) (1985). An international system for human cytogenetic nomenclature. In D. G. Harnden and H. P. Klinger (Eds.), *Report of the Standing Committee on Human Cytogenetic Nomenclature*. Karger, Basel, Switzerland.

Ishidate, M. Jr. (1988a). *Data Book of Chromosomal Aberration Tests in Vitro*. Elsevier, Amsterdam, The Netherlands.

Ishidate, M. Jr. (1988b). A proposed battery of tests for the initial evaluation of the mutagenic potential of medicinal and industrial chemicals. *Mutat. Res.* 205:397–407.

Ishidate, M. Jr., and Harnois, M. C. (1987). The clastogenicity of chemicals in mammalian cells. Letter to the editor. *Mutagenesis* 2:240–243.

Ishidate, M. Jr., Sofuni, T., Yoshikawa, K., et al. (1984). Primary mutagenicity screening of food additives currently used in Japan. *Fd. Chem. Toxicol.* 22:623–636.

Ishidate, M., Jr, Harnois, M. C., and Sofuni, T. (1988). A comparative analysis of data on the clastogenicity of 951 chemical substances tested in mammalian cell cultures. *Mutat. Res.* 195:151–213.

Japanese Ministry of Health and Welfare (JMHW) (1984). *Guidelines for Testing of Drugs for Toxicity*. Pharmaceutical Affairs Bureau, Notice No. 118. JMHW, Japan.

Japanese Ministry of Health and Welfare and Japanese Ministry of International Trade and Industry (1987, Mar. 31). Joint Directives of the Japanese Environmental Protection Agency, Tokyo.

Jordan, E., Saedler, H., and Starlinger, P. (1967). Strong polar mutations in the transferase gene of the galactose operon in *E. coli*. *Molec. Gen. Genet.* 100:296–306.

Kada, T. (1981). The DNA damaging activity of 42 coded compounds in the Rec-assay. In F. de Serres and J. Ashby (Eds.), *Evaluation of Short-term Tests for Carcinogens—report of the International Collaborative Program*. Elsevier/North Holland, Amsterdam, The Netherlands, pp. 175–182.

Kagunaga, T., and Yamasaki, H. (Eds.). (1984). *Transformation Assays of Established Cell Lines: Mechanisms and Application*. Proceedings of a Workshop Organized by IARC in Collaboration with the U.S. National Cancer Institute and the U.S. Environmental Protection Agency, February 15–17, 1984, Lyon. IARC Scientific Publication No. 67.

Kalter, K. (1977). Correlation between teratogenic and mutagenic effects of chemicals in mammals. In A. Hollanender (Ed.), *Chemical Mutagens: Principles and Methods for Their Detection*, Vol. 6, Plenum, New York, pp. 57–82.

Kato, H. (1973). Induction of sister chromatid exchanges by UV light and its inhibition by caffeine. *Exp. Cell Res.* 82:383–390.

Kato, H. (1974). Induction of sister chromatid exchanges by chemical mutagens and its possible relevance to DNA repair. *Exp. Cell Res.* 85:239–247.

Kato, T., and Shinoura, Y. (1977). Isolation and characterization of mutants of *Escherichia coli* deficient in induction of mutation by ultraviolet light. *Molec. Gen. Genet.* 156:121–132.

Kennelly, J. C., Stanton, C., and Martin, C. N. (1984). The effect of acetyl-CoA supplementation on the mutagenicity of benzidines in the Ames assay. *Mutat. Res.* 137:39–45.

Kirkland, D. J., and Garner, R. C. (1987). Testing for genotoxicity-chromosomal aberrations in vitro—CHO cells or human lymphocytes? *Mutat. Res.* 189:186–187.

Kirkland, D. J., and Muller, L. (2000). Interpretation of the biological relevance genotoxicity test results the importance of thresholds. *Mutat. Res.* 464:137–1147.

Kirkland, D. J., Marshall, R. R., McEnaney, S., Bidgwood, J., Rutter, A., and Mullineux, S. (1989). Aroclor-1254 induced rat liver S-9 causes chromosome aberrations in CHO cells but not in human lymphocytes: a role for active oxygen? *Mutat. Res.* 214:115–122.

Korenberg, J. R., and Freedlender, E. F. (1974). Giesma technique for the detection of sister chromatid exchanges. *Chromasoma* 48:355–360.

Lambert, B., Lindblad, A., Holmberg, K., and Francesconi, D. (1982). The use of sister chromatid exchange to monitor human populations for exposure to toxicologically

harmful agents. In S. Wolff (Ed.), *Sister Chromatid Exchange*. Wiley, New York, pp. 149–182.

Latt, S., Allen, J. W., Bloom, S. E., et al. (1981). Sister chromatid exchanges: A report of the gene-tox program. *Mutat. Res.* 87:17–62.

Latt, S. A. (1973). Microfluorometric detection of deoxyribonucleic acid replication in human metaphase chromosomes. *Proc. Natl. Acad. Sci. USA* 770:3395–3399.

Latt, S. A., Allen, J. W., Rogers, W. E., and Jurgens, L. A. (1977). In vitro and in vivo analysis of sister chromatid exchange formation. In B. J. Kilbey, M. Legator, W. Nichols, and C. Ramel (Eds.), *Handbook of Mutagenicity Test Procedures*. Elsevier, Amsterdam, The Netherlands, pp. 275–291.

Latt, S. A., and Loveday, K. S. (1978). Characterization of sister chromatid exchange induction by 8-methoxypsoralen plus UV light. *Cytogenet. Cell Genet.* 21:184–200.

Latt, S. A., Schreck, R. R., Loveday, K. S., Dougherty, C. P., and Shuler, C. F. (1980). Sister chromatid exchanges. *Adv. Human Genet.* 31:291–298.

Lawley, P. (1989). Mutagens as carcinogens: Development of current concepts. *Mutat. Res.* 213:3–26.

Li, A. P., Aaron, C. S., Aueltta, A. E., Dearfield, K. L., Riddle, J. C., Slesinski, R. S., and Stankowski, L. F., Jr. (1991). An evaluation of the roles of mammalian cell mutation assays in the testing of chemical genotoxicity. *Regul. Toxicol. Pharmacol.* 14:24–40.

Li, A. P., Carver, J. H., Choy, W. N., Bupta, R. S., Loveday, K. S., O'Neill, J. P., Riddle, J. C., Stankowski, L. F., and Yang, L. C. (1987). A guide for the performance of the Chinese Hamster ovary cell/hypoxanthine guanine phosphoribosyl transferase gene mutation assay. *Mutat. Res.* 189:135–141.

Linblad, W. J., and Jackim, E. (1982). Mechanism for the differential induction of mutation by kS9 activated benzo(*a*)pyrene employing either a glucose-6-phosphate dependent NADPH-regenerating system or an isocitrate dependent system. *Mutat. Res.* 96:109–118.

Lindahl, T. (1979). DNA glycosylases, endonucleases for apurinic/apyrimidinic sites and base excision repair. *Proc. Nucl. Acid Res. Mol. Biol.* 22:109–118.

Lindahl, T., Sedwick, B., Sekiguchi, M., and Nakabeppu, Y. (1988). Regulation and expression of the adaptive response to alkylating agents. *Ann. Rev. Biochem.* 57:133–157.

Little, J. W. (1984). Autodigestion of *lex A* and phage T repressors. *Proc. Natl. Acad. Sci. USA* 81:1375–1379.

Little, J. W., and Mount, D. W. (1982). The SOS regulatory system of *Escherichia coli*, *Cell* 29:11–22.

MacGregor, J. T., Wehr, C. M., and Gould, D. H. (1980). Clastogen-induced micronuclei in peripheral blood erythrocytes: The basis of an improved micronucleus test. *Environ. Mutagen.* 2:509–514.

MacPhee, D. G. (1973). *Salmonella typhimurium* hisG46 (R-Utrecht): Possible use in screening mutagens and carcinogens. *Appl. Microbiol.* 26:1004–1005.

Madle, S., and Obe, G. (1980). Methods for the analysis of the mutagenicity of indirect mutagens/carcinogens in eukaryotic cells. *Human Genet.* 56:7–20.

Mahon, G. A. T., Green, M. H. L., Middleton, B., Mitchell, I. DeG., Robinson, W. D., and Tweats, D. J. (1989). Analysis of data from microbial colon assays. In D. J.

Kirkland (Ed.), *Statistical Evaluation of Mutagenicity Test Data*. Cambridge University Press, Cambridge, pp. 26–65.

Marinus, M. G., and Morris, R. N. (1974). Biological function for the 6-methyladenine residues in the DNA of *Escherichia coli* K12. *J. Molec. Biol.* 85:309–322.

Maron, D. M., and Ames, B. N. (1983). Revised methods for the *Salmonella* mutagenicity test. *Mutat. Res.* 113:173–215.

Medical Research Council (MRC) (1981). *Guidelines for Work with Chemical Carcinogens in Medical Research Council Establishments*, MRC, London.

Merk, O., and Speit, G. (1998). Significance of formaldehyde-induced DNA–protein crosslinks for mutagenesis. *Environ. Mol. Mutagen.* 32:260–268.

Miller, E. C., and Miller, J. A. (1971). The mutagenicity of chemical carcinogens: Correlations, problems and interpretations. In A. Hollaender (Ed.), *Chemical Mutagens, Principles and Methods for Their Detection*, Vol. 1. Plenum, New York, pp. 83–120.

Miller, J. H. (1985). Pathways of mutagenesis revealed by analysis of mutational specificity. In J. Scaife, D. Leach, and A. Galizzi (Eds.), *Genetics of Bacteria*. Academic, New York, pp. 25–40.

Modrich, P. (1987). DNA mismatch correction. *Ann. Rev. Biochem.* 56:435–466.

Monticone, R. E., and Schneider, E. L. (1979). Induction of SCEs in human cells by fluorescent light. *Mutat. Res.* 59:215–221.

Morimoto, K., Sato, M., and Koizumi, A. (1983). Proliferative kinetics of human lymphocytes in culture measure by autoradiography and sister chromatid differential staining. *Epithel. Cell Res.* 145:249–356.

Mortelmanns, K. E., and Dousman, L. (1986). Mutagenesis and plasmids. In F. J. de Serres (Ed.), *Chemical Mutagens, Principles and Methods for Their Detection*. Vol. 10, Plenum, New York, pp. 469–508.

Mortelmanns, K. E., and Strocker, B. A. D. (1979). Segregation of the mutator property of plasmid R46 from its ultraviolet-protecting property. *Molec. Gen. Genet.* 167: 317–328.

Muller, H. J. (1927). Artificial transmutation of the gene. *Science* 66:84–87.

Natarajan, A. T., and Obe, G. (1982). Mutagenicity testing with cultured mammalian cells: Cytogenetic assays. In J. A. Heddle (Ed.), *Mutagenicity, New Horizons in Genetic Toxicology*. Academic, New York, pp. 172–213.

Neudecker, T., and Henschler, D. (1985). Allyl isothiocyanate is mutagenic in *Salmonella typhimurium*. *Mutat. Res.* 30:143–148.

Newton, M. F., and Lilly, L. J. (1986). Tissue specific clastogenic effects of chromium and selenium salts in vivo. *Mutat. Res.* 169:61–69.

Oberly, T. J., Bewsey, B. J., and Probst, G. S. (1987). A procedure for the CHO/HGPRT mutation assay involving treatment of cells in suspension culture and selection of mutants in soft agar. *Mutat. Res.* 182:99–111.

Organisation for Economic Co-operation and Development (OECD) (1983). *OECD Guidelines for the Testing of Chemicals*. No. 475. Genetic toxicology: *In vivo* mammalian bone marrow cytogenetic test—chromosomal analysis. Adopted April 4, 1984. OECD, Paris.

Ong, T., Mukhtar, M., Wolf, C. R., and Zeiger, E. (1980). Differential effects of cytochrome P450-inducers on promutagen activation capabilities and enzymatic activities of S-9 from rat liver. *J. Environ. Pathol. Toxicol.* 4:55–65.

Pagano, D. A., and Zeiger, E. (1985). The stability of mutagenic chemicals tested in solution. *Environ. Mutagen.* 7:293–302.

Painter, R. B. (1980). A replication model of sister-chromatid exchange. *Mutat. Res.* 70:337–341.

Parry, J. M., Arlett, C. F., and Ashby, J. (1985). An overview of the results of the *in vivo* and *in vitro* test systems used to assay the genotoxicity of BZD, DAT, DAB and CDA in the second UKEMS study. In J. M. Parry and C. F. Arlett (Eds.), *Comparative Genetic Toxicology: The Second UKEMS Collaborative Study*. Macmillan, London, pp. 597–616.

Perry, P., Henderson, L., and Kirkland, D. (1984). Sister chromatid exchange in cultured cells. In *UKEMS Subcommittee on Guidelines for Mutagenicity Testing. Report Part IIA*, pp. 89–121.

Perry, P. E., and Evans, H. J. (1975). Cytotological detection of mutagen/carcinogen exposure by sister chromatid exchange. *Nature* 258:121–125.

Perry, P. E., and Thomson, E. J. (1984). Sister chromatid exchange methodology. In B. J. Kilbey, M. Legator, W. Nichols, and C. Ramel (Eds.), *Handbook of Mutagenicity Test Procedures*. Elsevier, Amsterdam, Netherlands, pp. 495–529.

Perry, P. E., and Wolff, S. (1974). New Giemsa method for the differential staining of sister chromatids. *Nature* 251:156–158.

Pinney, R. J. (1980). Distribution among incompatibility groups of plasmids that confer UV mutability and UV resistance. *Mutat. Res.* 72:155–159.

Preston, R. J., Au, W., Bender, M., et al. (1981). Mammalian in vivo and in vitro cytogenetic assays. *Mutat. Res.* 87:143–188.

Preston, R. J., San Sebastian, J. R., and McFee, A. F. (1987). The *in vitro* human lymphocyte assay for assessing the clastogenicity of chemical agents. *Mutat. Res.* 189:175–183.

Prival, M. J., Bell, S. J., Mitchell, V. D., Peiperi, M. D., and Vaughn, V. L. (1984). Mutagenicity of benzidine and benzidine-congener dyes and selected monoazo dyes in a modified *Salmonella* assay. *Mutat. Res.* 136:33–47.

Prival, M. J., and Mitchell, V. D. (1982). Analysis of a method for testing azo-dyes for mutagenic activity in *S. typhimurium* in the presence of FMN in hamster liver S9. *Mutat. Res.* 97:103–116.

Revell, S. H. (1974). The breakage-and-reunion theory and the exchange theory for chromosome aberrations induced by ionizing radiations: A short history. In J. T. Lett and M. Zelle (Eds.), *Advances in Radiation Biology*, Vol. 4, Academic, New York, pp. 367–415.

Richardson, C., Williams, D. A., Allen, J. A., Amphlett, G., Changer, D. O., and Phillips, B. (1989). Analysis of data from *in vitro* cytogenetic assays. In D. J. Kirkland (Ed.), *UKEMS Sub-Committee on Guidelines for Mutagenicity Testing. Report. Part III. Statistical Evaluation of Mutagenicity Test Data*. Cambridge University Press, Cambridge, pp. 141–154.

Robinson, W. D., Green, M. H. L., Cole, J., Healy, M. J. R., Garner, R. C., and Gatehouse, D. (1989). Statistical evaluation of bacterial mammalian fluctuation tests. In D. J. Kirkland (Ed.), *Statistical Evaluation of Mutagenicity Test Data*. Cambridge University Press, Cambridge, pp. 102–140.

Rossman, T. G., and Klein, C. B. (1988). From DNA damage to mutation in mammalian cells: A review. *Environ. Molec. Mutagen.* 11:119–133.

Roth, J. R. (1974). Frameshift mutations. *Ann. Rev. Genet.* 8:319–346.

Russell, L. B. (1984). Procedures and evaluation of results of the mouse spot test. In B. J. Kilbey, M. Legator, W. Nichols, and C. Ramel (Eds.), *Handbook of Mutagenicity Test Procedures*. Elsevier, Amsterdam, The Netherlands, pp. 393–403.

Russell, L. B., Cumming, R. B., and Hunsicker, P. R. (1984). Specific-locus mutation rates in the mouse following inhalation of ethylene oxide and the application of the results to estimation of human genetic risk. *Mutat. Res.* 129:381–388.

Russell, W. L. (1951). X-ray induced mutations in mice. *Cold Spring Harbor Symp. Quant. Biol.* 16:327–336.

Russell, W. L. (1989). Reminiscences of a mouse specific-locus addict. *Environ. Molec. Mutagen.* 14(Suppl. 16):16–22.

Samson, L., and Cairns, J. (1977). A new pathway for DNA repair in *E. coli*. *Nature* 267:281–282.

Scott, D., Danford, N., Dean, B., Kirkland, D., and Richardson, C. (1983). In vitro chromosome aberration assays. In B. J. Dean (Ed.), *UKEMS Subcommittee on Guidelines for Mutagenicity Testing, Report Part I, Basic Test Battery*. UKEMS, Swansea, pp. 63–86.

Scott, D., Dean, B. J., Danford, N. D., and Kirkland, D. J. (1990). Metaphase chromosome aberration assays in vitro. In D. J. Kirkland (Ed.), *UKEMS Subcommittee on Guidelines for Mutagenicity Testing, Report. Part I: Revised Basic Mutagenicity Tests, UKEMS Recommended Procedures*. Cambridge University Press, Cambridge, pp. 63–86.

Searle, A. G. (1984). The specific locus test in the mouse. In B. J. Kilbey, M. Legator, W. Nichols, and C. Ramel (Eds.), *Handbook of Mutagenicity Test Procedures*. Elsevier, Amsterdam, The Netherlands, pp. 373–391.

Sedgwick, S. G., and Goodwin, P. A. (1985). Differences in mutagenic and recombinational DNA repair in enterobacteria. *Proc. Natl. Acad. Sci. USA* 82:4172–4176.

Selby, P. B., and Olson, W. H. (1981). Methods and criteria for deciding whether specific-locus mutation-rate data in mice indicates a positive, negative or inconclusive result. *Mutat. Res.* 83:403–418.

Shanabruch, W. G., and Walker, G. C. (1980). Localization of the plasmid (pKM101) gene(s) involved in $recA^+lexA^+$-dependent mutagenesis. *Molec. Gen. Genet.* 129:289–297.

Speit, G., and Hartmann, A. (1999). The comet assay (single-cell gel test). A sensitive genotoxicity test for the detection of DNA damage and repair. *Methods Mol. Biol.* 113:203–212.

Steel, R. G. D., and Torrie, J. H. (1960). *Principles and Procedures of Statistics*. McGraw-Hill, New York.

Streisinger, G., Okada, T., Emrich, J., et al. (1966). Frameshift mutations and the genetic code. *Cold Spring Harbor Symp. Quant. Biol.* 31:77–84.

Tates, A. D., Neuteboom, I., Hofker, M., and den Engelese, L. (1980). A micronucleus technique for detecting clastogenic effects of mutagens/carcinogens (DEN,DMN) in hepatocytes of rat liver in vivo. *Mutat. Res.* 74:11–20.

Taylor, J. H. (1958). Sister chromatid exchanges in tritium labeled chromosomes. *Genetics* 43:515–529.

Tennant, R. W., Margolin, B. H., Shelby, M. D., Zeiger, E., Haseman, J. K., Spalding, J., Caspary, W., Resnick, M., Stasiewicz, S., anderson, B., and Minor, R. (1987). Prediction of chemical carcinogenicity in rodents from *in vitro* genetic toxicity assays. *Science* 236:933–941.

Thomas, S. M., and Sedgwick, S. G. (1989). Cloning of *Salmonella typhimurium* DNA encoding mutagenic DNA repair. *J. Bacteriol.* 171:5776–5782.

Thompson, E. D. (1986). Comparison of in vivo and in vitro cytogenetic assay results. *Environ. Mutagen.* 8:753–767.

Thust, R., Mendel, J., Schwarz, H., and Warzoki, R. (1980). Nitrosated urea pesticide metabolites and other nitrosamides. Activity in clastogenicity and SCE assays, and aberration kinetics in Chinese hamster V79-E cells. *Mutat. Res.* 79:239–248.

Tice, R. R., Andrews, P. W., Hirai, O., and Singh, N. P. (1991). The single cell gel (SCG) assay: an electrophoretic technique for the detection of DNA damage in individual cells. In Witmer, C. R., Snyder, R. R., Jollow, D. J., Kalf, G. F., Kocsis, J. J., and Sipes, J. G. (eds.), *Biological Reactive Intermediates IV, Molecular and Cellular Effects and their Impact on Human Health*. Plenum Press, New York, pp. 157–164.

Tice, R. R., Agurell, E., Anderson, D., Burlinson, B., Hartmann, A., Kobayashi, H., Miyamae, Y., Rojas, E., Ryu, J. C., Sasaki, Y. F. (2000). Single cell gel/Comet Assay: guidelines for in vitro and in vivo genetic toxicology testing. Environ. *Mol. Mutagen.* 35:206–221.

Topham, J., Albanese, R., Bootman, J., Scott, D., and Tweats, D. (1983). *In vivo* cytogenetic assays. In B. Dea (Ed.), *Report of UKEMS Sub-committee on Guidelines for Mutagenicity Testing. Part I*, pp. 119–141.

Tsuchimoto, T., and Matter, B. E. (1979). *In vivo* cytogenetic screening methods for mutagens with special reference to the micronucleus test. *Arch. Toxicol.* 42:239–248.

Tweats, D. J., and Gatehouse, D. G. (1988). Discussion forum: Further debate of testing strategies. *Mutagenesis* 3:95–102.

Venitt, S., and Crofton-Sleigh, C. (1981). Mutagenicity of 42 coded compounds in a bacterial assay using *Escherichia coli* and *Salmonella typhimurium*. In F. J. de Serres and J. Ashby (Eds.), *Evaluation of Short-Term Tests for Carcinogens. Report of the International Collaborative Program. Progress in Mutational Research*, Vol. 1. Elsevier, New York, pp. 351–360.

Venitt, S., and Parry, J. M. (1984). Background to mutagenicity testing. In S. Venitt and J. M. Parry (Eds.), *Mutagenicity Testing, a Practical Approach*. IRL Press, Oxford, pp. 1–24.

Villani, G., Boiteux, S., and Radman, M. (1978). Mechanisms of ultraviolet-induced mutagenesis: Extent and fidelity of in vitro DNA synthesis on irradiated template. *Proc. Natl. Acad. Sci. USA* 75:3037–3041.

Wagner, R., and Meselson, M. (1976). Repair tracts in mismatched DNA heteroduplexes. *Proc. Natl. Acad. Sci. USA* 73:4135–4139.

Walker, G. C. (1984). Mutagenesis and inducible responses to deoxyribonucleic acid damage in *Escherichia coli*. *Microbiol. Rev.* 48:60–93.

Walker, G. C., and Dobson, P. P. (1979). Mutagenesis and repair deficiencies of *Escherichia coli umu C* mutants are suppressed by the plasmid pKM101. *Proc. Gen. Genet.* 172:17–24.

Wilcox, P., Naidoo, A., Wedd, D. J., and Gatehouse, D. G. (1990). Comparison of *Salmonella typhimurium* TA102 with *Escherichia coli* WP2 tester strains. *Mutagenesis* 5:285–291.

Wolff, S. (1977a). Lesions that lead to SCEs are different from those that lead to chromosome aberrations. *Mutat. Res.* 46:164.

Wolff, S. (1977b). In vitro inhibition of mono-oxygenase dependent reactions by organic solvents. Paper presented at the International Conference on Industrial and Environmental Xenobiotics, Prague, Czechoslovakia.

Wolff, S., and Perry, P. (1974). Differential staining of sister chromatids and the study of sister chromatid exchange without autoradiography. *Chromosomes* 48:341–353.

Yahagi, T., Degawa, M., Seino, Y., et al. (1975). Mutagenicity of carcinogen azo dyes and their derivatives. *Cancer Lett.* 1:91–96.

Yahagi, T., Nagao, M., Seino, Y., Matsushima, T., Sugimura, T., and Okada, M. (1977). Mutagenicities of *N*-nitrosamines in *Salmonella*. *Mutat. Res.* 48:120–130.

Yanofsky, C. (1971). Mutagenesis studies with *Escherichia coli* mutants with known amino acid and base-pair) changes. In A. Hollaender (Ed.), *Chemical Mutagens, Principles and Methods for Their Detection*, Vol. 1. Plenum, New York, pp. 283–287.

# *Repeat-Dose Toxicity Studies*

In the broadest sense, subchronic and chronic studies for pharmaceutical products can incorporate any of the routes used to administer a therapeutic agent, use any of a number of animal models, and conform to a broad range of experimental designs. They can be two weeks long (what used to be called "subacute" studies because they were conducted at dose levels below those employed for single-dose or acute studies) or last up to a year. Another name for these studies is repeat-dose studies (Ballantyne, 2000; Wilson et al., 2001; Gad, 2008a)—that is, those studies whereby animals have a therapeutic agent administered to them on a regular and repeated basis by one or more routes over a period of one year or less. There is great flexibility and variability in the design of such studies.

This chapter seeks to provide a firm grasp of the objectives for repeat-dose studies, the regulatory requirements governing them, the key factors in their design and conduct, and the interpretation of their results.

## 1 OBJECTIVES

As with any scientific study or experiment (but especially for those in safety assessment), the essential first step is to define and understand the reason(s) for the conduct of the study—that is, its objectives. There are three major (scientific) reasons for conducting subchronic and chronic studies, but a basic

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

characteristic of all but a few subchronic studies needs to be understood. The subchronic study is (as are most other studies in whole-animal toxicology) a broad screen. It is not focused on a specific endpoint; rather, it is a broad exploration of the cumulative biological effects of the administered agent over a range of doses. In fact, it is so broad an exploration that it can be called a "shotgun" study.

The objectives of the typical subchronic and chronic studies fall into three categories. The first is to broadly define the toxicity (and, if one is wise, the pharmacology and hyperpharmacology) of repeated doses of a potential therapeutic agent in an animal model (Traina, 1983). This definition is both qualitative (what are the target organs and the nature of the effects seen) and quantitative (at what dose levels, or, more importantly, at what plasma and tissue levels, are effects definitely seen and not seen).

The second objective (and the one that in the pharmaceutical industry laboratory usually compels both timing and compromising of design and execution) is to provide support for the initiation and/or continued conduct of clinical trials in humans (O'Grady and Linet, 1990; Smith, 1992). As such, subchronic studies should provide not only adequate clearance (therapeutic margin) of initial dose levels and duration of dosing but also guidance for any special measures to be made or precautions to be taken in initial clinical trials. Setting inadequate dose levels (either too low or too high) may lead to the failure of a study. A successful study must both define a safe, or "clean," dose level (one that is as high as possible, to allow as much flexibility as possible in the conduct of clinical studies) and demonstrate and/or characterize signs of toxicity at some higher dose. The duration-of-dosing issue is driven by a compromise between meeting regulatorily established guidelines (as set out in Table 1) and the economic pressure to initiate clinical trials as soon as possible.

The third objective is one of looking forward to later studies. The subchronic study must provide sufficient information to allow a prudent setting of doses for later, longer studies (including, ultimately, carcinogenicity studies). At the same time, the subchronic study must also provide guidance for the other (than dose) design features of longer term studies (such as what parameters to measure and when to measure them, how many animals to use, and how long to conduct the study).

**TABLE 1 Duration of Treatment Supported by Preclinical Studies**

| Animal Study Length | Generally Allowed Human Dosing |
|---|---|
| 2 Weeks | Up to 3 doses |
| 1 Month | 10 Days |
| 3 Months | 1 Month |
| 1 Year (rodent); 9 months (dog)—(U.S.)/6 months (European Economic Community, Japan) | Unlimited |

*Source*: ICH, 2000.

These objectives are addressed by the usual subchronic study. Some subchronic studies, however, are unusual in being conceived, designed, and executed to address specific questions raised (or left unanswered) by previous preclinical or early clinical studies. Such a special purpose is addressed separately.

Chronic studies (those that last six or nine months or a year) may also be conducted for the above purposes but are primarily done to fulfill registration requirements for drugs that are intended for continuous long-term (lifetime) use or frequent intermittent use.

## 2 REGULATORY CONSIDERATIONS

Much of what is done (and how it is done) in repeat-dose studies is a response to a number of regulations. Three of these have very broad impact. These are the good laboratory practiced (GLP) requirements, Animal Welfare Act requirements, and regulatory requirements that actually govern study design.

### 2.1 Good Laboratory Practices

Since 1978, the design and conduct of preclinical safety assessment studies for pharmaceuticals in the United States (and, indeed, internationally) have been governed and significantly influenced by GLPs. Strictly speaking, these regulations cover qualifications of staff and facilities, training, record-keeping, documentation, and actions required to ensure compliance with and effectiveness of these steps. Though the initial regulations were from the U.S. Food and Drug Administration (FDA, 1983), they have always extended to cover studies performed overseas (FDA, 1988a). Most other countries have adopted similar regulations. A discussion of these regulations is beyond the scope of the current chapter, but several aspects are central to this effort. Each technique or methodology to be employed in a study (such as animal identification, weighing and examination, blood collection, and data recording) must be adequately described in a standard operating procedure (SOP) before the study begins. Those who are to perform such procedures must be trained in them beforehand. The actual design of the study, including start date and how it is to be ended and analyzed, plus the principal scientists involved (particularly the study director) must be specified in a protocol that is signed before the study commences. Any changes to these features must be documented in amendments once the study has begun. It is a good idea that the pathologist who is to later perform or oversee histopathology be designated before the start of the study and that the design be a team effort involving the best efforts of the toxicologist, pathologist, and (usually for subchronic studies) the drug metabolism scientist.

## 2.2 Animal Welfare Act

Gone are the days when the pharmaceutical scientist could conduct whatever procedures or studies that were desired using experimental animals. The Animal Welfare Act (APHIS, 1989, amended every five years since) (and its analogues in other countries) rightfully requires careful consideration of animal usage to ensure that research and testing use as few animals as possible in as humane a manner as possible. As a start, all protocols must be reviewed by an institutional animal care and use committee. Such review takes time but should not serve to hinder good science. When designing a study or developing a new procedure or technique, the following points should be kept in mind:

1. Will the number of animals used be sufficient to provide the required data yet not constitute excessive use? (It ultimately does not reduce animal use to utilize too few animals to begin with and then have to repeat the study.)
2. Are the procedures employed the least invasive and traumatic available? This practice not only is required by regulations but also is sound scientific practice, since any induced stress will produce a range of responses in test animals that can mask or confound the chemically induced effects.

## 2.3 Regulatory Requirements for Study Design

The first consideration in the construction of a study is a clear statement of its objectives, which are almost always headed by meeting regulatory requirements to support drug development and registration. Accordingly, the relevant regulatory requirements must be analyzed, which is complicated by the fact that new drugs are no longer developed for registration and sale in a single-market country. The expense is too great, and the potential for broad international sales too appealing. While each major country has its own requirements as to study designs and studies required (with most of the smaller countries adhering to the regulations of one of the major players), harmonization has done much to smooth these differences (Adler and Zbinden, 1988). Meeting these regulatory requirements is particularly challenging for several reasons. First, the only official delineation of general requirements in the Untied States is dated (FDA, 1971), and recently special cases have arisen (anti-HIV agents, biotechnologically derived agents, therapeutic agents for neonates and the very elderly, etc.) that try the utility of these requirements. These needs have led to a stream of points to consider which seek to update requirements. Second, the term "guidelines" means different things in different countries (in the United States it means "requirements," in Japan "suggestions").

Agents intended to treat or arrest the progress of rapidly spreading life-threatening diseases (such as AIDS) are subject to less stringent safety assessment requirements prior to initial clinical evaluations than are other drugs. However, even though approval (if clinical efficacy is established) for

marketing can be granted with preclinical testing still under way, all applicable safety assessments (as with any other class of drugs) must still be completed (FDA, 1988b).

Drugs intended for use in either the elderly or the very young have special additional requirements for safety evaluation, in recognition of the special characteristics and potential sensitivities of these populations. For the elderly, these requirements call for special consideration of renal and hepatic effects [Center for Drug Evaluation and Research (CDER), 1989]. Likewise, drugs intended for the young require special studies to be performed in neonates and juvenile animals (usually of two or four weeks duration in rats).

In the last five to six years, a number of potentially important drugs have been produced by recombinant DNA technology. These biomacromolecules, which are primarily endogenously occurring proteins, present a variety of special considerations and concerns, including the following:

- Because they are endogenously occurring molecules, assessing their pharmacokinetics and metabolism presents special problems.
- Is the externally commercially produced molecule biologically equivalent to the naturally occurring one?
- As proteins, are they immunogenic or do they provoke neutralizing antibodies that will limit their usefulness?
- Because they are available only in very small quantities, the use of traditional protocols (such as those that use ever-increasing doses until an adverse effect is achieved) is impractical.
- Agents with such specific activity in humans may not be appropriately evaluated in rodents or other model species.

Each of these points must be addressed in any safety-testing plan (Weissinger, 1989). The requirements set out in this chapter are designed to do this (for repeat-dose testing).

## 3 STUDY DESIGN AND CONDUCT

### 3.1 Animals

In all but a few rare cases, for pharmaceutical safety assessment, separate studies in at least two species are required. Regulations require that both species be mammalian, and one of these must be a nonrodent; practice and economics dictate that the other species will be a rodent. With extremely rare exception, the rodent species employed is the rat (though the mouse also sees significant use). There is considerably more variability in the nonrodent species, with a range of factors determining whether the dog (most common choice), a primate species (typically the rhesus or cynomolgus, though some others are used in particular cases), the pig (particularly in Europe), or some other animal

(e.g., the ferret) is selected. The factors that should and do govern species selection are presented in detail in Gad (2008b). The use of multiple species is a regulatory requirement arising from experience and the belief (going back to 1944, at least) that it will provide a better chance of detecting the full range of biological responses (adverse and otherwise) to the new molecular entity being evaluated. This belief has come under fire in recent years (Zbinden, 1993) but is unlikely to be changed soon. Along the same lines, unless an agent is to be used by only one sex or the other of humans, equal numbers of both sexes of an animal species are utilized in the studies, with the sexes being treated as unrelated for purposes of statistical analysis. Also, except in rare cases, the animals used are young, healthy adults in the logarithmic phase of their growth curve. (The FDA specifies that rodents be less than six weeks of age at the initiation of dosing.)

Numbers of animals to be used in each dose group of a study are presented in Table 2. Though the usual practice is to use three different dose groups and at least one equal-sized control group, this number is not fixed and should be viewed as a minimum (see the section on study design later in this chapter). Use of more groups allows for a reduction in the risk of not clearly defining effects and establishing the highest possible safe dose at a modest increase in cost. There must be as many control animals as there are in the largest size test group to optimize statistical power.

Animals are assigned to groups (test and control) by one or another form of statistical randomization. Prior to assignment, animals are evaluated for some period of time after being received in house (usually at least one week for rodents and two for nonrodents) to ensure that they are healthy and have no discernible abnormalities. The randomization is never pure; it is always "blocked" in some form or another (by initial body weight, at least) so that each group is not (statistically) significantly different from the others in terms of the blocked parameters.

Proper facilities and care for test animals are not only a matter of regulatory compliance (and a legal requirement) but also essential for a scientifically sound and valid study. Husbandry requires clean cages of sufficient size and continuous availability of clean water and food (unless the protocol requires some restriction on their availability). Environmental conditions (temperature, humidity, and light–dark cycle) must be kept within specified limits. All

**TABLE 2 Numbers of Animals for Chronic and Subchronic Study per Test Group**

| Study Length | Rats per Sex | Dogs per Sex | Primates per Sex |
|---|---|---|---|
| 2–4 Weeks | 5–10 | 3–4 | 3 |
| 3 Months[a] | 20 | 6 | 5 |
| 6 Months | 30 | 8 | 5 |
| 1 Year | 50 | 10 | 10 |

[a]Starting with 13-week studies, one should consider adding animals (particularly to the high-dose group) to allow evaluation of reversal (or progression) of effects.

of these must in turn be detailed in the protocols of studies. The limits for these conditions are set forth in relevant National Institutes of Health and U.S. Department of Agriculture publications.

### 3.2 Routes and Setting Doses

Route (how an agent is administered to a test animal) and dose (how much of and how frequently an agent is administered) are inseparable in safety assessment studies and really cannot be defined independently. The selection of both begins with an understanding of the intended use of the drug in humans. The ideal case is to have the test material administered by the same route, at the same frequency [once a day, three times a day (t.i.d.), etc.], and for the same intervals (e.g., continuously if the drug is an intravenously infused agent) as the drug's eventual use in people. Practical considerations such as the limitations of animal models (i.e., there are some things you cannot get a rat to do), limitations on technical support,[1] and the like and regulatory requirements (discussed below as part of dose setting) frequently act or interact to preclude this straightforward approach.

Almost 30 routes exist for administration of drugs to patients, but only a handful of these are commonly used in preclinical safety studies (Gad, 1994). The most common deviation from what is to be done in clinical trials is the use of parenteral (injected) routes such as intravenous and subcutaneous deliveries. Such injections are loosely characterized as bolus (all at once or over a very short period, such as 5 min) and infusion (over a protracted period of hours, days, or even months). The term continuous infusion implies a steady rate over a protracted period, requiring some form of setup such as an implanted venous catheter or infusion port.

It is rare that the raw drug itself is suitable (in terms of stability, local tissue tolerance, and optimum systemic absorption and distribution) for direct use as a dosage form. Either it must be taken into a solution or suspension in a suitable carrier or a more complex formulation (a prototype of the commercial form) must be developed. Gad (2007) should be consulted for a more complete discussion of dose formulation for animals or humans. One formulation or more must be developed (preferably the same one for both animals and humans) based on the specific requirements of preclinical dosage formulation. For many therapeutic agents, limitations on volumes that can be administered and concentrations of active ingredient that can be achieved impact heavily on dose setting.

Setting of doses for longer term toxicity studies is one of the most difficult tasks in study design. The doses administered must include one that is devoid of any adverse effect (preferably of any effect) and yet still high enough to "clear" the projected clinical dose by the traditional or regulatory safety

[1] Many antiviral agents, particularly some anti-HIV agents, have rather short plasma half-lives, which requires frequent oral administration of the agent. Thirteen-week studies have been conducted with t.i.d. dosing of rats and monkeys, requiring around-the-clock shift work for technical staff of the laboratory.

factors (10 times for rodents, 5 times for nonrodents). At the same time, if feasible, at least one of the doses should characterize the toxicity profile associated with the agent (for some biotechnologically derived agents, particularly those derived from endogenous human molecules, it may only be possible to demonstrate biological effects in appropriate disease models and impossible to demonstrate toxicity). Because of limitations on availability of protodrugs, it is generally undesirable to go too high to achieve this second (toxicity) objective.

Traditionally, studies include three or more dose groups to fulfill these two objectives. Based on earlier results (generally, single-dose or two-week studies), doses are selected. It is, by the way, generally an excellent idea to observe the "decade rule" in extrapolation of results from shorter to longer studies; that is, do not try to project doses for more than an order-of-magnitude-longer study (thus the traditional progression from single-dose to 14-day to 90-day studies). Also, one should not allow the traditional use of three dose groups plus a control to limit designs. If there is a great deal of uncertainty, it is much cheaper in every way to use four or five dose groups in a single study than to have to repeat the entire study. Finally, remember that different doses may be appropriate for the different sexes.

It should also be kept in mind that formulating materials may have effects of their own, and a "vehicle" control group may be required in addition to a negative control group. Additionally the limitations on volumes of dose administration should be kept in mind as presented in Table 3.

### 3.3 Parameters to Measure

As was stated earlier, subchronic studies are shotgun in nature; that is, they are designed to look at a very broad range of endpoints with the intention of screening as broadly as possible the indications of toxicity. Meaningful findings

**TABLE 3 Guidance on Volumes of Administration**

| | Routes and Volumes (mL $kg^{-1}$ Administration Site) | | | | | |
|---|---|---|---|---|---|---|
| Species | Oral | SC | IP | IM | IV Bolus | IV (Slowing) |
| Mouse | 10 (50) | 10 (40) | 20 (80) | 0.5[a] (0.1)[a] | 5 | (25) |
| Rat | 10 (40) | 5 (10) | 10 (20) | 0.1[a] (0.2)[a] | 5 | (20) |
| Rabbit | 10 (15) | 1 (2) | 5 (20) | 0.25 (0.5) | 2 | (10) |
| Dog | 5 (15) | 1 (2) | 1 (20) | 0.25 (0.5) | 2.5 | (10) |
| Macaque | 5 (15) | 2 (5) | – (10) | 0.25 (0.05) | 2 | — |
| Marmoset | 10 (15) | 2 (5) | – (20) | 0.25 (0.5) | 2.5 | (10) |
| Minipig | 10 (15) | 1 (2) | 1 (20) | 0.25 (0.5) | 2.5 | (5) |

*Notes*: —, data not available; SC, subcutaneous; IP, intraperitoneal; IM, intramuscular. For nonaqueous injectates consideration must be given to time of absorption before redosing. No more than two IM sites should be used per day. Subcutaneous sites should be limited to two to three per day. Figures in parentheses are total per day with multiple administrations.

[a]Milliliters.

*Sources*: Adapted from ECVAM, 2001.

are rarely limited to a single endpoint—rather, what typically emerges is a pattern of findings. This broad search for components of the toxicity profile is not just a response to regulatory guidelines intended to identify potentially unsafe drugs. An understanding of all the indicators of biological effect can also frequently help one to understand the relevance of findings, to establish some as unrepresentative of a risk to humans, and even to identify new therapeutic uses of an agent.

Parameters of interest in the repeat-dose study can be considered as sets of measures, each with its own history, rationale, and requirements. It is critical to remember, however, that the strength of the study design as a scientific evaluation lies in the relationships and patterns of effects that are seen in looking at each of these measures (or groups) not simply as independent findings but rather as integrated profiles of biological effects.

***Body Weight*** Body weight (and the associated calculated parameter of body weight gain) is a nonspecific, broad screen for adverse systemic toxicity (Ellaben and Hart, 1998). Animals are initially assigned to groups based on a randomization scheme that includes having each group vary insignificantly from one another in terms of body weight. Weights are measured prior to the initial dose, then typically 1, 3, 5, 7, 11, and 14 days thereafter. The frequency of measurement of weights goes down as the study proceeds—after 2 weeks, weighing is typically weekly through 6 weeks, then every other week through 3 months, and monthly thereafter. Because the animals used in these studies are young adults in the early log phase of their growth, decreases in the rate of gain relative to control animals is a very sensitive (albeit nonspecific) indicator of systemic toxicity.

***Food Consumption*** Food consumption is typically measured with one or two uses in mind. First, it may be explanatory in the interpretation of reductions (either absolute or relative) in body weight. In cases where administration of the test compound is via diet, it is essential to be able to adjust dietary content so as to accurately maintain dose levels. Additionally, the actual parameter itself is a broad and nonspecific indicator of systemic c toxicity. Food consumption is usually measured over a period of several days, first weekly and then on a once-a-month basis. Water consumption, which is also sometimes measured, is similar in interpretation and use. Additionally, particularly in nonrodents, loss of appetite is an early indicator of adverse effects in animals.

***Clinical Signs*** Clinical signs are generally vastly underrated in value, probably because insufficient attention is paid to care in their collection. Two separate levels of data collection are actually involved here. The first is the morbidity and mortality observation, which is made twice a day. This generally consists of a simple cageside visual assessment of each animal to determine if it is still alive, and, if so, whether it appears in good (or at least stable) health. Historically, this regulatory required observation was intended to ensure that tissues

from intoxicated animals were not lost for meaningful histopathological evaluation due to autolysis (Arnold et al., 1990).

The second level of clinical observation is the detailed hands-on examination analogous to the human physical examination. It is usually performed against a checklist [see Gad (2007) for an example], and evaluation is of the incidence of observations of a particular type in a group of treated animals compared to controls. Observations range from being indicative of nonspecific systemic toxicity to fairly specific indicators of target organ toxicity. These more detailed observations are typically taken after the first week of a study and on a monthly basis thereafter.

Ophthalmological examinations are typically made immediately prior to initiation of a study (and thus serve to screen out animals with preexisting conditions) and toward the end of a study.

Particularly when the agent under investigation either targets or acts via a mechanism likely to have a primary effect on a certain organ for which functional measures are available, an extra set of measurements of functional performance should be considered. The organs or organ systems that are usually of particular concern are the kidneys, liver, cardiovascular, nervous, and immune. Special measures (such as creatinine clearance as a measure of renal function) are combined with other data already collected (organ weights, histopathology, clinical pathology, etc.) to provide a focused "special" investigation or evaluation of adverse effects on the target organ system of concern. In larger animals (dogs and primates) some of these measures [such as electrocardiograms (ECGs)] are made as a matter of course in all studies.

***Clinical Chemistry and Pathology*** Clinical pathology covers a number of biochemical and morphological evaluations based on invasive and noninvasive sampling of fluids from animals that are made periodically during the course of a subchronic study. These evaluations are sometimes labeled as clinical (as opposed to anatomical) pathology determinations. Table 4 presents a summary of the parameters measured under the headings of clinical chemistry, hematology, and urinalysis using samples of blood and urine collected at predetermined intervals during the study. Conventionally, these intervals are typically at three points evenly spaced over the course of the study, with the first being one month after study initiation and the last being immediately prior to termination of the test animals. For a three-month study, this means that samples of blood and urine would be collected at one, two, and three months after study initiation (i.e., after the first day of dosing of the animals). There are some implications of these sampling plans that should be considered when the data are being interpreted. Many of the clinical chemistry (and some of the hematological) markers are really the result of organ system damage that may be transient in nature (see Table 5 for a summary of interpretations of clinical chemistry findings and Table 6 for a similar summary for hematological findings). The samples on which analysis is performed are from fixed points in time, which may miss transient changes (typically, increases) in some enzyme levels.

**TABLE 4 Clinical Pathology Measures**

| Clinical Chemistry | Hematology | Urinalysis |
|---|---|---|
| Albumin | Erythrocyte count (RBC) | Chloride |
| Alkaline phosphatase (ALP) | Hemoglobin (HGB) | Bilirubin |
| Blood urea nitrogen (BUN) | Hematocrit (HCT) | Glucose |
| Calcium | Mean corpuscular hemoglobin (MCH) | Occult blood |
| Chloride | Mean corpuscular volume (MCV) | pH |
| Creatine | Platelet count | Phosphorus |
| Creatine phosphokinase (CPK) | Prothrombin time | Potassium |
| Direct bilirubin | Reticulocyte count | Protein |
| γ-Glutamyltransferase (GGT) | White cell count (WBC) | Sodium |
| Globulin | White cell differential count | Specific gravity |
| Glucose | | Volume |
| Lactic dehydrogenase (LDH) | | |
| Phosphorus | | |
| Potassium | | |
| Serum glutamic–oxaloacetic transaminase (SGOT) | | |
| Serum glutamic–pyruvic transaminase (SGPT) | | |
| Sodium | | |
| Total bilirubin | | |
| Total cholesterol | | |
| Total protein | | |
| Triglycerides | | |

There is now a concerted effort to improve predictive power of nonclinical safety assessment studies for major classes of issues that are "discovered" in clinical trials and marketed use of new drugs. In particular, there are efforts to identify biomarkers which can be measured in animal studies and have relevance and more similarity for clinical finding.

Particular focus has been on biomakers for the heart (see Braunwald, 2008), kidney, and liver. The kidney guidance is quite recent with European Medicines Agency (EMEA) and FDA announcements in late May 2008 that they have accepted seven biomarkers (Kim-1, albumin, total protein, Cystatin C, $\beta_2$-microglobulin, urinary clusterin, and urinary trefoil factor 3) for use in preclinical studies. The liver guidance is recent (EMEA, 2008), though this reflects an ongoing effort (see Kaplowitz, 2005) to try and reduce the occurrence of "idiosyncratic" liver toxicity in clinical trials. Table 7 presents a list of available clinical chemistry measures that are considered preclinical in the three major categories of hepatic toxicity.

It would seem likely that rather than these being adapted whole cloth into the standard design of repeat-dose studies, they are likely to be used in a more considered manner when there is cause for specific concern. This, however, must still be worded out.

***Pharmacokinetics and Metabolism*** Pharmaceutical subchronic toxicity studies are always accompanied by a parallel determination of the

**TABLE 5 Association of Changes in Biochemical Parameters with Actions at Particular Target Organs**

| | Organ System | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Parameter | Blood | Heart | Lung | Kidney | Liver | Bone | Intestine | Pancreas | Notes |
| Albumin | | | | ↓ | ↓ | | | | Produced by the liver. Very significant reductions indicate extensive liver damage. |
| ALP (alkaline phosphatase) | | | | | ↑ | ↑ | ↑ | | Elevations usually associated with cholestasis. Bone ALP tends to be higher in young animals |
| ALT (alanine aminotransferase), also called SGPT (serum glutamic pyruvic transaminase) | | | | | ↑ | | | | Evaluations usually associated with hepatic damage or disease. Prime clinical hepatic damage marker—Hy's law sets a boundary of a 3× increase of significant damage. |
| Bilirubin (total) | ↑ | | | | ↑ | | | | Usually elevated due to cholestasis either due to obstruction or hepatopathy. |
| BUN (blood urea nitrogen) | | | | ↑ | ↓ | | | | Estimates blood-filtering capacity of the kidneys. Does not become significantly elevated until kidney function is reduced 60–75%. |
| Calcium | | | | ↑ | | | | | Can be life threatening and result in acute death. |
| Cholinesterase | | | | ↑ | ↓ | | | | Found in plasma, brain, and RBC. |
| CPK (creatinine phosphokinase) | | ↑ | | | | | | | Most often elevated due to skeletal muscle damage but can also be produced by cardiac muscle damage. Can be more sensitive than histopathology. |
| Creatine | | | | ↑ | | | | | Also estimates blood-filtering capacity of kidney as BUN does. More specific than BUN. |

**TABLE 5** ***Continued***

| Parameter | Blood | Heart | Lung | Kidney | Liver | Bone | Intestine | Pancreas | Notes |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | Organ System | | | | |
| Glucose | | | | | | | | ↑ | Alterations other than those associated with stress are uncommon and reflect an effect on the pancreatic islets or anorexia. |
| GGT (γ-glutamyltransferase) | | | | | ↑ | | | | Elevated in cholestasis. This is microsomal enzyme and levels often increase in response to microsomal enzyme induction. |
| HBDH (hydroxybutyric dehydrogenase) | | ↑ | | | ↑ | | | | Most prominent in cardiac muscle tissue. |
| LDH (lactic dehydrogenase) | | ↑ | ↑ | ↑ | ↑ | | | | Increase usually due to skeletal muscle, cardiac muscle, and liver damage. Not very specific unless isozymes are evaluated. |
| Protein (total) | | | | ↑ | ↑ | | | | Absolute alterations are usually associated with decreased production (liver) or increased loss (kidney). |
| SGOT (serum glutamic oxaloacetic transaminase); also called AST (aspartate amino transferase) | | ↑ | | ↑ | ↑ | | | ↑ | Present in skeletal muscle and heart and most commonly associated with damage to these. |
| SDH (sorbitol dehydrogenase) | | | | | ↑ or ↓ | | | | Liver enzyme which can be quite sensitive but is fairly unstable.QQSamples should be processed as soon as possible. |

*Source*: Gad, 2007.

**TABLE 6 Some Probable Conditions Affecting Hematological Changes**

| Parameter | Elevation | Depression |
|---|---|---|
| Red blood cells | 1. Vascular shock<br>2. Excessive diuresis<br>3. Chronic hypoxia<br>4. Hyperadrenocorticism | 1. Anemias<br>a. Blood loss<br>b. Hemolysis<br>c. Low RBC production |
| Hematocrit | 1. Increased RBC<br>2. Stress<br>3. Shock<br>a. Trauma<br>b. Surgery<br>4. Polycythemia | 1. Anemias<br>2. Pregnancy<br>3. Excessive hydration |
| Hemoglobin | 1. Polycythemia (increase in production of RBCs) | 1. Anemias<br>2. Lead poisonings |
| Mean cell volume | 1. Anemias<br>2. $B_{12}$ deficiency | 1. Iron deficiency |
| Mean corpuscular hemoglobin | 1. Reticulocytosis | 1. Iron deficiency |
| White blood cells | 1. Bacterial infections<br>2. Bone marrow stimulation | 1. Bone marrow depression<br>2. Cancer chemotherapy<br>3. Chemical intoxication<br>4. Splenic disorders |
| Platelets | — | 1. Bone marrow depression<br>2. Immune disorder |
| Neutrophils | 1. Acute bacterial infections<br>2. Tissue necrosis<br>3. Strenuous exercise<br>4. Convulsions<br>5. Tachycardia<br>6. Acute hemorrhage | 1. Viral infections |
| Lymphocytes | 1. Leukemia<br>2. Malnutrition<br>3. Viral infections | |
| Monocytes | 1. Protozoal infections | |
| Eosinophils | 1. Allergy<br>2. Irradiation<br>3. Pernicious anemia<br>4. Parasitism | |
| Basophils | 1. Lead poisoning | |

*Source*: Gad, 2007.

pharmacokinetics of the material of interest administered by the same route as that used in the safety study. This parallel determination consists of measuring plasma levels of the administered agent and its major metabolites either in animals that are part of the main study or in a separate set of animals (in parallel with the main study) that are dosed and evaluated to determine just these endpoints. The purpose of these determinations is both to allow a better interpretation of the findings of the study and to encourage the most accurate

**TABLE 7 Clinical Chemistry Measures Considered Useful in Identifying Liver Toxicity**

| Parameters | Hepatocellular | Hepatobiliary | Mitochondrial |
|---|---|---|---|
| Alanine aminotransferase (ALT) | X | | |
| Aspartate aminotransferase (AST) | X | | |
| Sorbitol dehydrogenase (SDH) | X | | |
| Glutamate dehydrogenase (GLDH) | X | | X |
| Total bile acid (TBA) | X | X | |
| Alkaline phosphatase (ALP) | | X | |
| γ-Glutamyltransferase (GGT) | | X | |
| 5′-nucleotidase (5-NT) | | X | |
| Total bilirubin (TBILI) | | X | |
| Potential ancillary markers | | | |
| Lactate | | | X |
| Lactate dehydrogenase (LHD) | X | | |
| Ornithine carbamyltransferase | X | | X |
| Unconjugated bilirubin (UBILI) | X | | |

possible extrapolation to humans. The first data of interest are the absorption, distribution, and elimination of the test material, but a number of other types of information can also be collected (Yacobi et al., 1989; Tse and Jaffe, 1991). For nonparenteral routes it is essential to demonstrate that systemic absorption and distribution of the test material did occur; otherwise, it is open to question whether the potential safety of the agent in humans has been adequately addressed (not to mention the implication for potential human therapeutic efficacy). A complication, however, is that there are limits as to how much blood may be collected from specific species at one time (see Table 8), particularly as samples must be drawn to allow evaluation of clinical chemistry, clinical pathology, and pharmacokinetics. This is even worse in the case of biologics, where adequate samples must also be drawn to allow the evaluation of antibody levels.

### 3.4 Other In-Life Endpoints for Evaluation

***Ophthalmology*** Ophthalmological examination of all animals in a study (particularly nonrodents) should be performed both before study initiation and at the completion of the period at which the drug is administered. This should be performed by an experienced veterinary ophthalmologist.

***Cardiovascular Function*** Particularly in light of recent concerns with drug-induced arrhythmias, careful consideration must be given to incorporating adequate evaluation of drug-induced alterations on cardiovascular function. This is usually achieved by measuring blood pressure, heart rate, and an ECG prestudy and periodically during the course of the study (usually at at least one intermediate period and at the end of the study) in the nonrodent species being employed.

**TABLE 8 Total Blood Volumes and Recommended Maximum Blood Sample Volumes for Species of Given Body Weight**

| Species | Blood Volume (mL) | 7.5% (mL) | 10% (mL) | 15% (mL) | 20% (mL) |
|---|---|---|---|---|---|
| Mouse (25 g) | 1.8 | 0.1 | 0.2 | 0.3 | 0.4 |
| Rat (250 g) | 16 | 1.2 | 1.6 | 2.4 | 3.2 |
| Rabbit (4 kg) | 224 | 17 | 22 | 34 | 45 |
| Dog (10 kg) | 850 | 64 | 85 | 127 | 170 |
| Macaque (Rhesus) (5 kg) | 280 | 21 | 28 | 42 | 56 |
| Macaque (Cynomologus) (5 kg) | 325 | 24 | 32 | 49 | 65 |
| Marmoset (350 kg) | 25 | 2.0 | 2.5 | 3.5 | 5 |
| Mini pig (15 kg) | 975 | 73 | 98 | 146 | 195 |

**TABLE 9 FDA Draft Criteria for Neurotoxicity Screen as Component of Short-Term and Subchronic Studies**

Histopathological examination of tissues representative of nervous system, including brain, spinal cord, and peripheral nervous system

Quantitative observations and manipulative test to detect neurological, behavioral, and physiological dysfunctions. These may include:
- General appearance
- Body posture
- Incidence and severity of seizure
- Incidence and severity of tremor, paralysis, and other dysfunction
- Level of motor activity and arousal
- Level of reactivity to stimuli
- Motor coordination
- Strength
- Gait
- Sensorimotor response to primary sensory stimuli
- Excessive lacrimation or salivation
- Pilorection
- Diarrhea
- Ptosis
- Other signs of neurotoxicity deemed appropriate

***Neurotoxicology*** Table 9 presents the FDA's current draft criteria (FDA, 1993, 2000) for endpoints to be incorporated in studies as a screen for neurotoxicity. In practice, a functional observation battery is employed at several endpoints (usually one and three months into the study) to fill these requirements.

***Immunotoxicology*** In response to concerns about potential effects of drugs on the immune system, the ICH and FDA (2006) have promulgated a guidance

**TABLE 10 FDA Draft Recommendation for Type I Immunotoxicity Test That Can be Included in Repeated-Dose Toxicity Studies**

| | |
|---|---|
| Hematology | Clinical Chemistry |
| White blood cell counts | Total serum production |
| Differential white blood cell counts | Albumin |
| Lymphocytosis Lymphopenia | Albumin-to-globulin ratio |
| Eosinophilia | Serum transaminases |
| Histopathology | |
| Lymphoid tissues | |
| Spleen | |
| • Lymph nodes | |
| • Thymus | |
| • Peyer's patches in gut | |
| • Bone Marrow | |
| Cytology (if needed)[a] | |
| • Prevalence of activated macrophages | |
| • Tissue prevalence and location of lymphocytes | |
| • Evidence of B-cell germinal centers | |
| • Evidence of T-cell germinal centers | |
| Necrotic or proliferative changes in lymphoid tissues | |

[a]More comprehensive cytological evaluation of the tissues would not be done unless there is evidence of potential immunotoxicity from the preceding evaluations.

calling for a basic set of potential indicators of immunotoxicity (Table 10) to be evaluated and considered in standard repeat-dose studies. Most of these endpoints are, it should be noted, already collected in traditional subchronic designs.

***Pharmacokinetics*** All subchronic and chronic toxicity studies now incorporate (either in the study itself or in a parallel study) evaluation of the basic pharmacokinetics of a compound. This is discussed in detail in Chapter Pharmacokinetics and Toxicokinetics in Drug Safety Evaluation.

## 3.5 Histopathology

Histopathology is generally considered the single most significant portion of data to come out of a repeat-dose toxicity study. It actually consists of three related sets of data (gross pathology observations, organ weights, and microscopic pathology) that are collected during the termination of the study animals. At the end of the study, a number of tissues are collected during termination of all surviving animals (test and control). Organ weights and terminal body weights are recorded at study termination, so that absolute and relative (to body weight) values can be statistically evaluated.

These tissues, along with the organs for which weights are determined, are listed in Table 11. All tissues collected are typically processed for microscopic observation, but only those from the high-dose and control groups

**TABLE 11 Tissues for Histopathology**

| Adrenals[a] | Mainstream Bronchi |
|---|---|
| Body and cervix | Major salivary glad |
| Brain, all three levels[a] | Mesenteric lymph nodes |
| Cervical lymph nodes | Ovaries and tubes |
| Cervical spinal cord | Pancreas |
| Duodenum | Pituitary |
| Esophagogastric junction | Prostate |
| Esophagus | Skeletal muscle from proximal hind limb |
| Eyes with optic nerves | Spleen[a] |
| Femur with marrow | Sternebra with marrow |
| Heart | Stomach |
| Ileum | Testes with epididymides[a] |
| Kidneys[a] | Thymus and mediastinal contents[a] |
| Large bowel | Thyroid with parathyroid[a] |
| Larynx with thyroid and parathyroid | Trachea |
| Liver[a] | Urinary bladder |
| Lungs[a] | Uterus including horns |

[a]Organs to be weighed.

are necessarily evaluated microscopically. If a target organ is discovered in the high-dose group, then successively lower dose groups are examined until a "clean" (devoid of effect) level is discovered (Haschek and Rousseaup, 1991).

In theory, all microscopic evaluations should be performed blind (without the pathologist knowing from which dose group a particular animal came), but this is difficult to do in practice and such an approach frequently degrades the quality of the evaluation. Like all the other portions of data in the study, proper evaluation benefits from having access to all data that address the relevance, severity, timing, and potential mechanisms of a specific toxicity. Blind examination is best applied in peer review or consultations on specific findings.

In addition to the "standard" set of tissues specified in Table 8, observations during the course of the study or in other previous studies may dictate that additional tissues be collected or special examinations (e.g., special stains, polarized light or electron microscopy, immunocytochemistry, or quantitative morphometry) be undertaken to evaluate the relevance of or understand the mechanisms underlying certain observations.

Histopathology testing is a terminal procedure, and, therefore, sampling of any single animal is a one-time event (except in the case of a tissue collected by biopsy). Because it is a regulatory requirement that the tissues from a basic number of animals be examined at the stated end of the study, an assessment of effects at any other time course (most commonly, to investigate recovery from an effect found at study termination) requires that satellite groups of animals be incorporated into the study at startup. Such animals are randomly

assigned at the beginning of the study and are otherwise treated exactly the same as the equivalent treatment (or control) animals.

### 3.6 Study Designs

The traditional design for a repeat-dose toxicity study is very straightforward. The appropriate number of animals of each sex are assigned to each of the designated dose and control groups. Unfortunately, this basic design is taken by many to be dogma, even when it does not suit the purposes of the investigator. There are many possible variations to study design, but four basic factors should be considered: controls, the use of interval and satellite groups, balanced and unbalanced designs, and staggered starts.

Classically, a single control group of the same size as each of the dose groups is incorporated into each study. Some studies incorporate two control groups (each the same size as the experimental groups) to guard against having a statistically significant effect due to one control group being abnormal for one or more parameters (a much more likely event when laboratory animals were less genetically homogeneous than they are now). The belief is that a "significant" finding that differs from one (but not both) of the concurrent control groups and does not differ from historical control data can be considered as not biologically significant. This is, however, an indefensible approach. Historical controls have value, but it is the concurrent control group(s) in a study that is of concern.

Interval or satellite groups have been discussed at two earlier points in this chapter. They allow measurement of termination parameters at intervals other than at termination of the study. They are also useful when the manipulation involved in making a measurement (such as the collection of an extensive blood sample), while not terminal, may compromise (relative to other animals) the subject animals. Another common use of such groups is to evaluate recovery from some observed effect at study termination.

Usually, each of the groups in a study is the same size, with each of the sexes being equally represented. The result is called a balanced design, with statistical power for detection of effects optimized for each of the treatment groups. If one knows little about the dose–toxicity profile, this is an entirely sound and rational approach. However, there are situations when one may wish to utilize an unbalanced design—that is, to have one or more dose groups larger than the others. This is usually the case when either greater sensitivity is desired (typically in a low-dose group) or an unusual degree of attrition of test animals is expected (usually due to mortality in a high-dose group) or as a guard against a single animal's idiopathic response being sufficient to cause "statistical significance."

As it is the normal practice to have a balanced design, it is also traditional to initiate treatment of all animals at the same time. This may lead to problems at study termination, however. It is a very uncommon toxicology laboratory that can "bring a study down" on a single day. In fact, there are no laboratories

that can collect blood and perform necropsies in a single day on even the 48–80 dogs involved in a study, much less the 160–400+ rats in the rodent version. Starting all animals on a study the same day presents a number of less than desirable options. The first is to terminate as many animals as can be done each day, continuing to dose (and therefore further affect) the remaining test animals. Assuming that the animals are being terminated in a random, balanced manner, this means that the last animals terminated will have received from 3 to 10 additional days of treatment. At the least, this is likely to cause some variance inflation (and therefore both decrease the power of the study design and possibly confound interpretation). If the difference in the length of treatment of test animals is greater than 3% of the intended length of the study, one should consider alternative designs.

An alternative approach to study design that addresses this problem employs one of several forms of staggered starts. In these, distinct groups of animals have their dosing initiated at different times. The most meaningful form recognizes that the two sexes are in effect separate studies anyway (they are never compared statistically, with the treatment groups being compared only against the same-sex control group). Thus if the termination procedure for one sex takes three to five days, then one sex should be initiated on dosing one week and the other on the following week. This maximizes the benefits of common logistical support (such as dose formulation) and reduces the impact of differential length of dosing on study outcome.

A variation on this is to stagger the startup either of different dose groups or of the satellite and main study portions of dose groups. The former is to be avoided (it will completely confound study outcome), while the latter makes sense in some cases (pharmacokinetics and special measures) but not others (recovery and interval sacrifice).

## 4 STUDY INTERPRETATION AND REPORTING

For a successful repeat-dose study, the bottom line is the clear demonstration of a no-effect level, characterization of a toxicity profile (providing guidance for any clinical studies), enough information on pharmacokinetics and metabolism to scale dosages to human applications, and at least a basic understanding of the mechanisms involved in any identified pathogenesis. The report that is produced as a result of the study should clearly communicate these points—along with the study design and experimental procedures, summarized data, and their statistical analysis—and it should be GLP compliant, suitable for FDA submission format.

Interpretation of the results of a study should be truly scientific and integrative. It is elementary to have the report state only each statistically and biologically significant finding in an orderly manner. The meaning and significance of each in relation to other findings as well as the relevance to potential human effects must be evaluated and addressed.

The author of the report should ensure that it is accurate and complete but also that it clearly tells a story and concludes with the relevant (to clinical development) findings.

## REFERENCES


Alder, S., and Zbinden, G. (1988). *National and International Drug Safety Guidelines*. M.T.C. Verlag, Zollikon, Switzerland.

Allaben, W. T., and Hart, R. W. (1998). Nutrition and toxicity modulation: The impact of animal body weight on study outcome. *Int. J. Toxicol.* 17(Suppl. 2):1–3.

Animal and Plant Health Inspection Service (APHIS) (1989). United States Department of Agriculture (August 31, 1989). *Fed. Reg.* 54(168):36112–36163.

Arnold, D. L., Grice, H. C., and Krawksi, D. R. (1990). *Handbook of In Vivo Toxicity Testing*. Academic, San Diego.

Ballantyne, B. (2000). Repeated exposure toxicity. In B. Ballantyne, T. Marrs, and T. Syversen (Eds.), *General and Applied Toxicology*. MacMillan, London, pp. 55–66.

Braunwald, E. (2008). Biomarkers in heart failure. *New Engl. J. Med.* 358:2148–2155.

Center for Drug Evaluation and Research (CDER) (1989, Nov.). *Guideline for the Study of Drugs Likely to be Used in the Elderly*. FDA, Washington, DC.

ECVAM (2000). A good practice guide to the administration of substances and removal of blood, including routes and volumes. *J. Appl. Toxicol.* 21:15–23.

Ellaben, W. T., and Hart, R. W. (1998). Nutrition and toxicity modulation: The impact of animal body weight on study outcome. *Int. J. Toxicol* 17(Suppl. 2):1–3.

European Medicines Ageny (EMEA) (2008). *Non-Clinical Guidance on Drug Included Hepatoxicology*. EMEA, Brussels, Belgium.

Gad, S. C. (1994). Routes in safety assessment studies. *J. Am. Coll. Toxicol.* 13:1–17.

Gad, S. C. (2007). *Acute Toxicology*, 2nd ed. Academic, San Diego, CA.

Gad, S. C. (2008a) Repeat dose toxicity studies. In S. C. Gad (Ed.), *Preclinical Development Handbook: Toxicology.* Wiley-Interscience, Hoboken, NJ.

Gad, S. C. (2008b). *Animal Models in Toxicology*, 2nd ed. Taylor & Francis, Philadelphia, PA.

Haschek, W. M., and Rousseaup, C. G. (1991). *Handbook of Toxicology Pathology*. Academic, San Diego.

International Conference on Harmonisation (ICH) (2008). M3(R2) Non-clinical safety studies for the conduct of human clinical trials for pharmaceuticals. Available online.

Kaplorwitz, N. (2005). Idiosyncratic drug hepatotoxicology. *Nature Rev. Drug Discov.* 4:489–499.

O'Grady, J., and Linet, O. I. (1990). *Early Phase Drug Evaluation in Man*. CRC Press, Boca Raton, FL.

Smith, C. G. (1992). *The Process of New Drug Discovery and Development*. CRC Press, Boca Raton, FL.

Traina, V. M. (1983). The role of toxicology in drug research and development. *Med. Res. Rev.* 3:45–72.

Tse, F. L. S., and Jaffe, J. M. (1991). *Preclinical Drug Disposition*. Marcel Dekker, New York.

U.S. Food and Drug Administration (FDA) (1971). *FDA Introduction to Total Drug Quality*. U.S. Government Printing Office, Washington, DC.

U.S. Food and Drug Administration (FDA) (1983, Mar.). Good laboratory practices for nonclinical laboratory studies. Code of Federal Regulations 21 Part 58.

U.S. Food and Drug Administration (FDA) (1988a, Apr.). Good laboratory practices. Code of Federal Regulations 21 Part 58.

U.S. Food and Drug Administration (FDA) (1988b). Investigational new drug, antibiotic, and biological drug product regulations, procedures for drugs intended to treat life-threatening and severely debilitating illness. *Fed. Reg.* 53(204):41516–41524.

U.S. Food and Drug Administration (FDA) (1993). *Toxicological Principles for the Safety Assessment of Direct Food Additives and Color Additives Used in Food, Redbook II*. Center for Food Safety and Applied Nutrition, FDA, Washington, DC, p. 86.

U.S. Food and Drug Administration (FDA) (2000). *Toxicological Principles for the Safety of Food Ingredients, Redbook 2000.* Center for Food Safety and Applied Nutrition, FDA, Washington, DC. Code of Federal Regulations.

Weissinger, J. (1989). Nonclinical pharmacologic and toxicologic considerations for evaluating biologic products. *Regul. Toxicol. Pharmacol.* 10:255–263.

Wilson, N. H., Hardisty, J. F., and Hayes, J. R. (2001). Short-term, subchronic and chronic toxicity studies. In A. W. Hayes (Ed.), *Principles and Methods of Toxicology*, 4th ed. Taylor & Francis, Philadelphia, PA, pp. 917–958.

Yacobi, A., Skelly, J. P., and Batra, V. K. (1989). *Toxicokinetics and New Drug Development*. Pergamon, New York.

Zbinden, G. (1993). The concept of multispecies testing in industrial toxicology. *Regul. Toxicol. Pharmacol.* 17:84–94.

# *Immunotoxicology in Drug Development*

## 1 INTRODUCTION

Unlike many of the other areas covered in this volume, immunotoxicity evaluation of drugs has undergone fundamental changes since the first edition. Before, there was no specific guidance. Now, with the International Conference on Harmonisation (ICH) S8A in place and in force, there is both a requirement and a roadmap for immunotoxicity evaluations, though biologics are excluded from coverage.

This is in particular a concern as biologics have attained their therapeutic promise with their hyperpharmacology providing the largest portion of safety concern for them [from the "cytokine storm" of TGN-1412 to the unintended immune issues associated with most monoclonal antibodies (mAbs)]. With immune modulation of some form being the primary intended therapeutic effects and one-third of all new approved therapeutics being biologics, this is particularly a concern, while specifically not covered by ICH guidances, approaches and requirements for immunotoxicity evaluations of these moieties are addressed, however, in this chapter.

All three ICH regions have made strong efforts to harmonize the immunotoxicity risk assessment for investigational new drugs. These efforts culminated in the release of the ICH S8 guideline, which was adopted by the CHMP (Committee for Medicinal Products for Human Use) in October 2005 and came into force in the United States in April 2006, as well as the MHLW (Ministry of Health, Labor and Welfare, Japan) in April 2006. According to this current

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

guideline, initial immunotoxicity assessment should be based on the evaluation of data already available from standard toxicity studies and other characteristics of the drug substance, such as pharmacological properties, the intended patient population, known drug class effects, clinical data for the drug, and its disposition. The need for additional immunotoxicity testing should be decided on the basis of a weight of evidence assessment, taking into account all available information. Testing should thus be a tiered approach, triggered and determined by concerns from the weight of evidence assessment.

Immunotoxicity is defined in the ICH S8 guideline as unintended immunosupression or enhancement. It must be noticed that drug-induced hypersensitivity and autoimmunity are not in the scope of ICH S8 (Pattels and Taylor, 2008). Surprisingly, the ICH S8 guideline excludes biological and biotechnology-derived products, probably the larger area of concern for pharmaceuticals. EMEA (2007) has promulgated specific guidance for immunogenicity evaluation of proteins.

The immune system is a highly complex system of organ systems, cells, and soluble factors distributed throughout the body and involved in a multitude of functions, including antigen presentation and recognition, amplification, and cell proliferation with subsequent differentiation and secretion of lymphokines and antibodies. In health, these are in a state of balance, and there are extensive mechanisms to maintain this balance. The resulting integrated system is responsible for defense against foreign pathogens and spontaneously occurring neoplasms and is readily triggered to response. To be effective, the immune system must be able to both recognize and destroy foreign antigens. To accomplish this, cellular and soluble components of diverse function and specificity circulate through blood and lymphatic vessels, thus allowing them to act at remote sites and tissues. For this system to function in balance and harmony requires regulation through cell-to-cell communications and precise recognition of self versus nonself. There are multiple opportunities for immunotoxicants to upset this balance by selectively disabling one or more of the cell types or alter membrane morphology and receptors. There are several undesired immune system responses that may occur upon repeated therapeutic administration of a pharmaceutical that may ultimately present barriers to its development, including:

- Down-modulation of the immune response (immunosuppression), which may result in an impaired ability to deal with neoplasia and infections. This is of particular concern if the therapeutic agent is intended to be used in patients with preexisting conditions such as cancer, severe infection, or immunodeficiency diseases.
- Up-modulation of the immune system (i.e., autoimmunity).
- Direct adverse immune responses to the agent itself in the form of hypersensitivity responses (anaphylaxis and delayed contact hypersensitivity).
- Direct immune responses to the agent that limit or nullify its efficacy (i.e., the development of neutralizing antibodies).

Iammune modulated responses to drugs ("drug allergies") are a major problem and a cause of discontinuance of use by patients who need access to the therapeutic benefits (Patterson et al., 1986), and there remains no adequate preclinical methodology for identifying/predicting these responses to orally administered small-molecule drugs (Hastings, 2001).

It is the intent of this chapter to provide an understanding of these adverse immunological effects, the types of preclinical tests that may be used to detect them, and approaches for testing and interpreting test results.

Immunotoxicology has evolved over the last 20 years as a specialty within toxicology that brings together knowledge from basic immunology, molecular biology, microbiology, pharmacology, and physiology. As a discipline, immunotoxicology involves the study of adverse effects that xenobiotics have on the immune system. As listed above, several different types of adverse immunological effects may occur, including immunosuppression, autoimmunity, and hypersensitivity. Although these effects are clearly distinct, they are not mutually exclusive. For example, immunosuppressive drugs that suppress suppressor-cell activity can also induce autoimmunity (Hutchings et al., 1985), and drugs that are immunoenhancing at low doses may be immunotoxic at high doses. Chemical xenobiotics may be in the form of natural or man-made environmental chemicals—pharmaceuticals and biologics that are pharmacologically, endocrinologically, or toxicologically active. Although, in general, xenobiotics are not endogenously produced, immunologically active biological response modifiers that naturally occur in the body should also be included, since many are not known to compromise immune function when administered in pharmacologically effective doses (Koller, 1987).

Although the types of immunological responses to various xenobiotics may be similar, the approach taken for screening potential immunological activity will vary depending on the application of the compound. Thus, this chapter will primarily focus on the immunotoxicology of pharmaceuticals. In contrast to potential environmental exposures, pharmaceuticals are developed with intentional but restricted human exposure and their biological effects are extensively studied in surveillance. Pharmaceuticals are developed to be biologically active, and, in some cases, intentionally immunomodulating or immunosuppressive. Many will react with biological macromolecules or require receptor binding in order to be pharmacologically active. By their nature, these interactions may result in toxicity to the cells of the immune system, may adversely alter the appearance of "self" to produce an autoimmune response, or may form a hapten, which may then elicit a hypersensitivity response. Because of the fast-expanding development of new drugs that can potentially impact the immune responsiveness of humans, immunotoxicity testing of new pharmaceutical products has become a growing concern.

Until recently, immunotoxicology in pharmaceutical safety assessment has been poorly addressed by both regulatory requirements/guidelines and existing practice. Notable exceptions are the testing requirements for delayed contact hypersensitivity for dermally administered agents and antigenicity/anaphylaxis testing for drugs to be registered in Japan. The most recently

announced regulatory expectation for parenterally administered protein or peptide agents produced by biotechnology is that the development of antibodies (neutralizing and otherwise) should be evaluated in at least one (preferably two) of the animal models used to assess general systemic toxicity.

Unanticipated immunotoxicity is infrequently observed with drugs that have been approved for marketing. With the exception of drugs that are intended to be immunomodulatory or immunosuppressive as part of their therapeutic mode of action, there is little evidence that drugs cause unintended functional immunosuppression in humans (Gleichmann et al., 1989). However, hypersensitivity (allergy) and autoimmunity are frequently observed and are serious consequences of some drug therapies (DeSwarte, 1986; Patterson et al., 1986; Vos et al., 1989; Choquet-Kastylevsky et al., 2001; Pieters, 2001; Luebke et al., 2007). An adverse immune response in the form of hypersensitivity is one of the most frequent safety causes for withdrawal of drugs that have already made it to the market (see Table 1) and accounts for approximately 15% of adverse reactions to xenobiotics (deWeck, 1983). In addition, adverse immune responses such as this (usually urticaria and frank rashes) are the chief "unexpected" finding in clinical studies. These findings are unexpected in that they are not predicted by preclinical studies because there is a lack of good preclinical models for predicting systemic hypersensitivity responses, especially to orally administered agents. As a consequence, the unexpected occurrence of hypersensitivity in the clinic may delay, or even

**TABLE 1 Drugs Withdrawn from Market Due to Dose- and Time-Unrelated Toxicity Not Identified in Animal Experiments**

| Compound | Adverse Reaction | Year of Introduction | Years on Market |
|---|---|---|---|
| Aminopyrine | Agranulocytosis | ca. 1900 | 75 |
| Phenacetin | Interstitial nephritis | ca. 1900 | 83 |
| Dipyrone | Agranulocytosis | ca. 1930 | 47 |
| Clioquinol | Subacute myelo-optic neuropathy | ca. 1930 | 51 |
| Oxyphenisatin | Chronic active hepatitis | ca. 1955 | 23 |
| Nialamide | Liver damage | 1959 | 19 |
| Phenoxypropazine | Liver damage | 1961 | 5 |
| Mebanazine | Liver damage | 1963 | 3 |
| Ibufenac | Hepatotoxicity | 1966 | 2 |
| Practolol | Oculo-mucocutaneous syndrome | 1970 | 6 |
| Alcolofenace | Hypersensitivity | 1972 | 7 |
| Azaribine | Thrombosis | 1975 | 1 |
| Ticrynafen | Nephropathy | 1979 | 1 |
| Benoxaprofen | Photosensitivity, hepatotoxicity | 1980 | 2 |
| Zomepirac | Urticaria, anaphylactic shock | 1980 | 3 |
| Zirnelidine | Hepatotoxicity | 1982 | 2 |
| Temafloxacin | Hepato- and renal toxicity | 1990 | 2 |
| Tronan | Hepato- and renal toxicity | 1997 | 3 |
| Renzalin | Hepatotoxicity | 1996 | 4 |

*Source*: Adapted from Bakke et al. (1984) and updated.

preclude, further development and commercialization. Thus, a primary purpose for preclinical immunotoxicology testing is to help us detect these adverse effects earlier in development, before they are found in clinical trials.

## 2 REGULATORY POSITIONS

The pharmaceutical and medical device industries are increasingly concerned with whether preclinical testing of their products should include routine immunotoxicological screening or be done on an "as-needed basis," triggered by the toxicological profile of the xenobiotic established in routine preclinical safety testing (Bloom et al., 1987). Although the U.S. Food and Drug Administration (FDA) has not as yet officially released guidelines for immunotoxicity testing of pharmaceuticals, recent drug development efforts in the areas of biotechnology, prostaglandins, interleukins, and recombinant biological modifiers have elicited the expectation that the development of antibodies (neutralizing and otherwise) should be evaluated in at least one of the animal models used to assess general systemic toxicity. More to the point, draft guidelines have been released for devices [Center for Devices and Radiological Health (CDRH), 1997]. The other available guidance had been the draft guidelines in the revision of the "Redbook" (FDA, 1993). The ICH S8 guidance (2006) supersedes all of them (for pharmaceuticals) (Table 2).

The FDA had drafted a similar two-leveled approach (Hinton, 1992) for assessing immunotoxicity of food colors, additives, and drugs, and these are reflected in the S8 guidance. In all of these testing schemes, the initial tier generally includes a fundamental standard toxicity assessment with emphasis on histopathology of the major components of the immune system. Additional tiers are then added to more precisely evaluate the functionality of the components that appeared to be adversely affected in the first tier of tests. These test strategies are primarily geared toward the detection of small-molecule pharmaceutical-induced immunosuppression; thus the effectiveness of these test schemes for detecting immunostimulation has still not been determined (Spreafico, 1988). Table 3 presents items that should be considered in such an initial evaluation.

The FDA, ICH, and European Medicines Agency (EMEA) guidelines are not currently entirely aligned. The ultimate immune test would be to examine the effects of xenobiotics on the intact animal's response to challenge by viral, bacterial, or parasitic pathogens or neoplastic cells. The ability of the immune system to compensate or, conversely, its inability to compensate for loss or inhibition of its components is fully examined through host resistance mechanisms. This tiered test approach has been validated with 50 selected compounds, and results from these studies have shown that the use of only two or three immune tests are sufficient to predict known immunotoxic compounds in rodents with a >90% concordance (Luster et al., 1992a,b). Specifically the use of either a humoral response assay for plaque-forming colonies (PFC

**TABLE 2 Comparison of Current ICH and Former European Union (EU) and U.S. Immunotoxicity**

| Parameter | ICH S8 (in Operation in All ICH Regions Since 2006) | FDA CDER (Still in Operation Since 2002) | EU CHMP (Still in Operation Since 2000) |
|---|---|---|---|
| Specific immunotoxicity guideline | Yes | Yes | No, included in guidance on repeat-dose toxicity. |
| Drug-induced hypersensitivity, immunogenicity, and autoimmunity excluded | Yes | No, these categories are included in the guideline. | Yes. (Note: Skin sensitizing potential addressed in CPMP Note for Guidance on Non-Clinical Local Tolerance Testing, 2001 [42].) |
| Screening study(ies) required | Yes, the initial screen for potential immunotoxicity involves standard toxicity studies (STSs) from short-term to chronic repeat-dose studies in rodents and nonrodents. | Yes, including all standard repeat-dose toxicology studies that have been performed. | Yes, screening required for all new active substances in at least one repeat-dose toxicity study (duration ideally should be 28 days). Rats or mice are species of choice. |
| Screening study(ies) immunotoxicity parameters | Changes in hematology, lymphoid organ weights, histopathology of immune system, and serum globulins and increased incidences of infections and tumors should be evaluated for signs of immunotoxic potential in the STSs. | Changes in hematology lymphoid organ weights, gross pathology and histopathology of immune system, and serum globulins and increased incidences of infections and tumors should be evaluated for signs of immunotoxic potential | Hematology, lymphoid organ weights, histopathology of lymphoid tissues, bone marrow, cellularity, distribution of lymphocyte subsets, and NK-cell activity (if latter two unavailable, primary antibody response to T-cell-dependent antigen). |
| Other factors to consider in evaluation of potential immunotoxicity and need for additional immunotoxicity studies | Pharmactological properties of drug; patient population; structural similarities to known immunomodulators; drug disposition; clinical data. | Patients population; known drug class effects [including structure-activity relationships (SARs)]; drug pharmacokinetics; clinical data. If drug intended for HIV, immune function studies required. | None specifically included in the guideline. |
| "Follow-on"/"additional" immunotoxicity studies | Additional studies may be required depending on the "weight of evidence review" of STSs and "other factors." Additional studies addressed in 3.2, 3.3, and Appendix of guideline. | Not specified. | Not specified. |

**TABLE 3 Typical Indicators of Immunotoxcity Which May Be Observed During Regulatory Repeat-Dose Toxicity Studies**

| Findings | Possible Indicator of |
|---|---|
| *During In-Life Phase* | |
| Increased frequencies of infectious disease | Immunosuppression |
| Increased frequencies of tumors in long-term studies in absence of genotoxicity or nongenotoxic indicators of tumorigenicity (e.g., endocrine) | Immunosuppression |
| Unexpected pathological symptoms or deaths shortly after administration | Hypersensitivity |
| Strong inflammatory reactions at site of administration | Hypersensitivity |
| *Gross Necropsy* | |
| Significant increase or decrease of size and weight of lymphatic organs | Unintended immunostimulation or immunosuppression |
| *Hematology* | |
| Changes in total or differential blood counts | Unintended immunostimulation or mmunosuppression |
| Anemia | Type II hypersensitivity |
| Altered frequencies of lymphocyte subsets (flow cytometry)[a] | Unintended immunostimulation or immunosuppression |
| *Clinical Chemistry* | |
| Altered total globulin levels or albumin: globulin ratio | Unintended immunostimulation or immunosuppression |
| Changes of immunoglobulin isotype levels[a] | Unintended immunostimulation or immunosuppression |
| Reduction of hemolytic complement activity[a] | Unintended immunostimulation type III hypersensitivity |
| Antinuclear or anticytoplasmic antibodies[a] | Unintended immunostimulation or autoimmunity |
| *Histopathology* | |
| Changes of cellularity and/or microanatomy of lymphatic organs | Unintended immunostimulation or immunosuppression |
| Vasculitis, glomerulonephritis | Type III hypersensitivity |

[a]This parameter is normally not measured during standard toxicity studies but may be integrated when a focus is drawn on immunotoxicity assessment.

response) or determination of surface marker expression in combination with almost any other parameter significantly increased the ability to predict immunotoxicity when compared to the predictivity of any assay alone.

The FDA guidelines for immunotoxicity testing of food additives start with a type 1 battery of tests. Type 1 tests can be derived from the routine measurements and examinations performed in short-term and subchronic rodent toxicity studies, since they do not require any perturbation of the test animals (immunization or challenge with infectious agents). These measurements

include hematology and serum chemistry profiles, routine histopathological examinations of immune-associated organs and tissues, and organ and body weight measurements, including thymus and spleen. If a compound produces any primary indicators of immunotoxicity from these measurements, more definitive immunotoxicity tests, such as those indicated in the preceding paragraph, may be recommended on a case-by-case basis.

The following is a brief explanation of some of the indicators that may be used to trigger additional definitive testing and a description of some of the most commonly used assays to assess humoral, cell-mediated, or nonspecific immune dysfunction, which are common to most immunotoxicology test strategies.

### 2.1 CDER Guidance for Investigational New Drugs

The Center for Drug Evaluation and Research (CDER, 2001) promulgated draft guidance for pre-INDA (investigation new drug application) immunotoxicity clearly established the framework for the FDA's approach, which was finalized under ICH S8 (2006). It begins by characterizing five adverse-event categories:

- Immunosuppression
- Antigenicity
- Hypersensitivity
- Autoimmunity
- Adverse immunostimulation

Specific tests are proposed for each of these categories. The CDER notes that immune system effects in noclinical toxicology studies are often attributed and written off as due to stress (Ader and Cohen, 1993). Such effects are frequently reversible with repeat dosing and tend not to be dose related. It is also proposed that, when possible, dose extrapolations to those in clinical use be based on relative body area. Specific recommendations are made for when to conduct specific testing (as opposed to the broader general evaluations integrated into existing repeat-dose testing) (Figure 1) and for follow-up studies for exploring mechanisms (Figure 2).

## 3 OVERVIEW OF IMMUNE SYSTEM

A thorough review of the immune system is not the intent of this chapter, but a brief description of the important components of the system and their interactions is necessary for an understanding of how xenobiotics can affect immune function. A breakdown at any point in this intricate and dynamic system can lead to immunopathology.

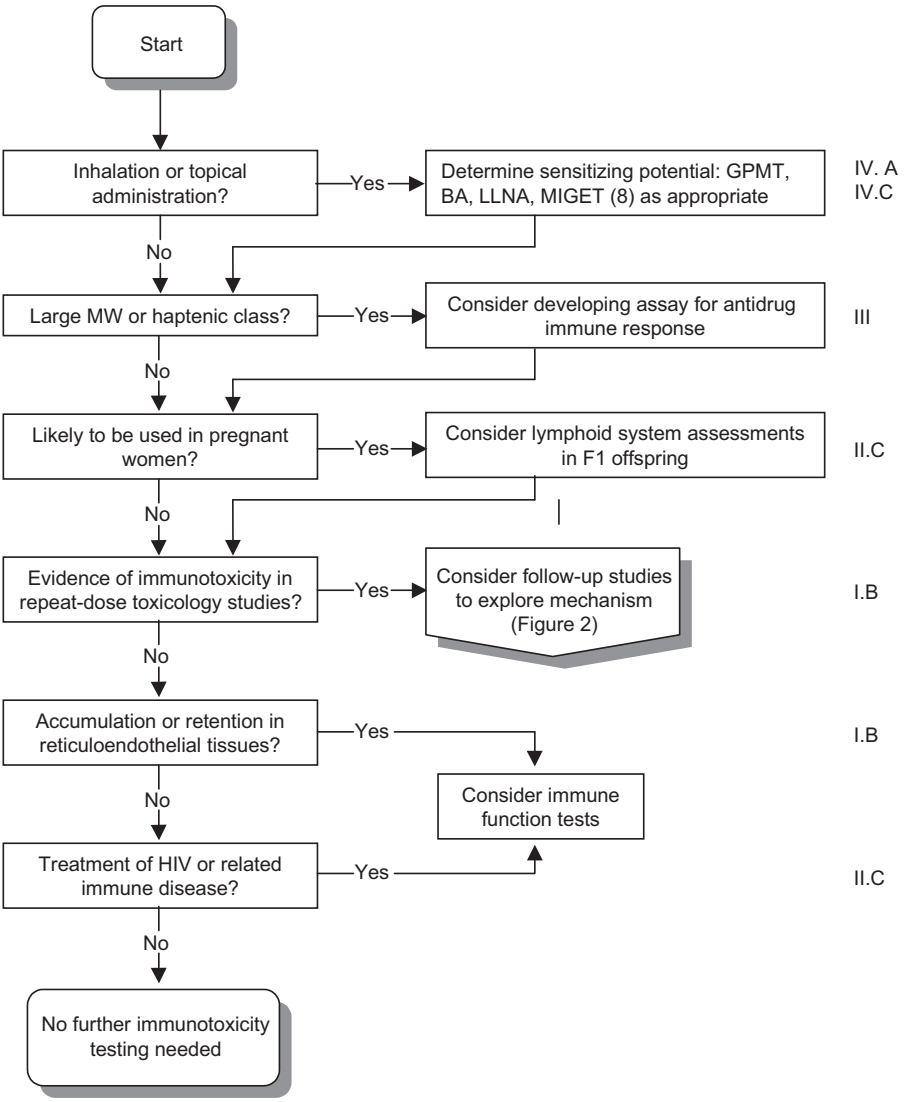


***Figure 1*** *CDER flowchart for determining when to conduct specific immunotoxicity testing. Annotations in right margin indicate location of text describing specific advice GPMT, guinea pig maximization test; BA, Buehler assay (Buehler patch test); LLNA, local lymph node assay; MIGET, mouse IgE test. (There is only a relatively small database available for assessing the usefulness of the MIGET for drug regulatory purposes.)*

The immune system is divided into two defense mechanisms: nonspecific, or innate, and specific, or adaptive, mechanisms that recognize and respond to foreign substances. Some of the important cellular components of nonspecific and specific immunity are described in Table 4. The nonspecific immune system

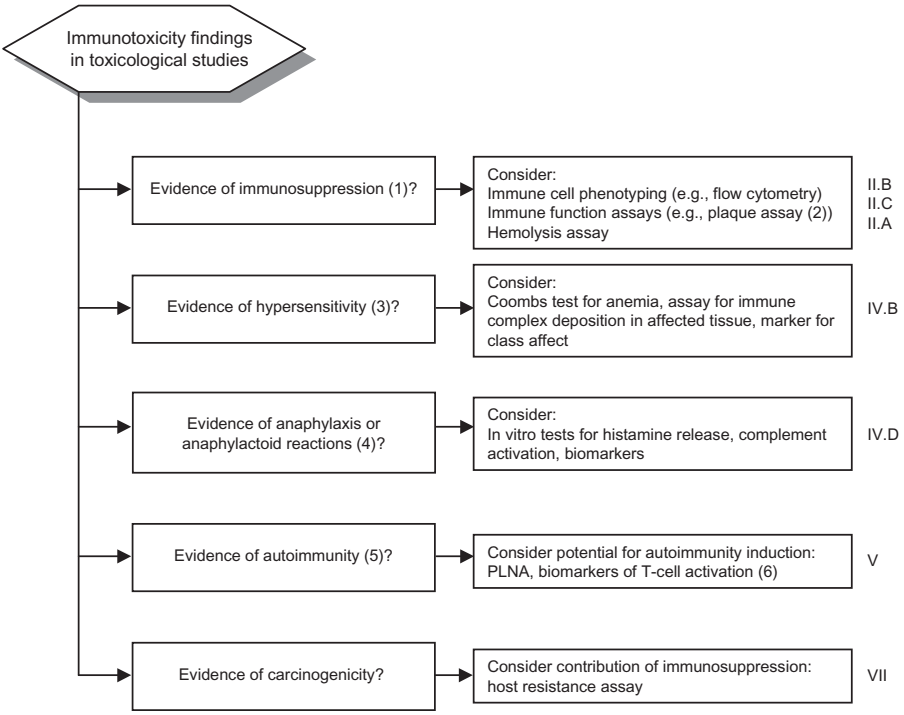


***Figure 2*** *Follow-up studies to consider for exploring mechanisms of immunotoxicity. Annotations in right margin indicate location of text describing specific advice. (1) Examples include myelosuppression, histopathology in immune-associated tissues, increased infection, tumors, decreased serum Ig, phenotypic changes in immune cells. (2) Other acceptable assays include drug effect on NK-cell function in vitro bastogenesis, cytotoxic T-cell function, cytokine production, DTH, host resistance to infections or implanted tumors. (3) Examples include anemia, leukopenia, thrombocytopenia, pnuemonitis, vasculitis, lupuslike reactions, glomerulonephritis. (4) Examples include cardiopulmonary distress, rashes, flushed skin, swelling of face or limbs. (5) Examples include vasculitis, lupuslike reactions, glomerulonephritis, hemolytic anemia. (6) There are no established assays that reliably assess potential for autoimmunity and acute systemic hypersensitivity. The popliteal lymph node assay (PLNA) has only a relatively small database available for assessing its usefulness for drug regulatory purposes.*

is the first line of defense against infectious organisms. Its cellular components are the phagocytic cells such as the monocytes, macrophages, and polymorphic neutrophils (PMNs).

The innate immune system has been identified in immune defense in insects and alone in the phylogenetic sequence and serves the same purpose in all these organisms, including humans. The innate immune system encompasses all physical, chemical, and cellular barriers that protect the individual from microbial infections without the need to learn to discriminate self from nonself. The body protects itself from dangerous actions of the innate immune system by the lack of expression of molecular patterns of microorganisms and by the

**TABLE 4 Cellular Components of Immune System and Their Functions**

| Cell Subpopulations | Markers[a] | Functions |
|---|---|---|
| | *Nonspecific Immuity* | |
| Granulocytes | — | Degranulate to release mediators |
| Neutrophils (blood) | | |
| Basophils (blood) | | |
| Eosinophils (blood) | | |
| Mast cells (connective tissue) | | |
| NK cells | — | Nonsensitized lymphocytes; directly kill target cells |
| Reticuloendothelial | CD14; HLA-DR | Antigen processing, presentation, and phagocytosis (humoral and some cell-mediated responses) |
| Macrophage (peritoneal, pleural, alveolar spaces) | | |
| Histiocytes (tissues) | | |
| Monocytes (blood) | | |
| | *Specific immunity* | |
| Humoral immunity | CD19; CD23 | Proliferate; form plasma cells |
| Activated B cells | — | Secrete antibody; terminally differentiated |
| Plasma cells | — | |
| Resting | — | Secrete IgM antibodies (primary response) |
| Memory | | Secrete IgG antibodies (secondary response) |
| | *Cell-Mediated Immunity* | |
| T-Cell types | CD4; CD25 | Assists in humoral immunity; required for antibody production |
| Helper ($T_h$) | CD8; CD25 | |
| Cytotoxic ($T_k$) | CD8; CD25 | Targets lysis |
| Suppressor ($T_s$) | | Suppresses/regulates humoral and cell-mediated responses |

[a]Activation surface markers detected by specific monoclonal antibodies; can be assayed with flow cytometry.

abundant expression of inhibitors. Most components of the innate immune system can work independently and in parallel to destroy microorganisms. Due to its redundancy, the innate immune system is rather robust regarding its actions.

The specific, or adaptive, immune system is present only in vertebrates and is characterized by memory, specificity, and the ability to distinguish self from nonself. The important cells of the adaptive immune system are the lymphocytes and antigen-presenting cells (APCs) that are part of nonspecific immunity. The lymphocytes, which originate from pluripotent stem cells located in the hematopoietic tissues of the liver (fetal) and bone marrow, are composed of two general cell types: T and B cells. The T cells differentiate in the thymus and are made up of three subsets: helper, suppressor, and cytotoxic. The B cells, which have the capacity to produce antibodies, differentiate in the bone

marrow or fetal liver. The various functions of the T cells include presenting antigen to B cells, helping B cells to make antibody, killing virally infected cells, regulating the level of the immune response, and stimulating cytotoxic activity of other cells such as macrophages (Male et al., 1987).

Activation of the immune system is thought to occur when APCs such as macrophages and dendritic cells take up antigen via $F_c$ or complement receptors, process the antigen, and present it to T cells (see Figure 3). Macrophages release soluble mediators such as interleukin 1 (IL-1), which stimulate T cells to proliferate. Antigen-presenting cells must present antigen to T cells in

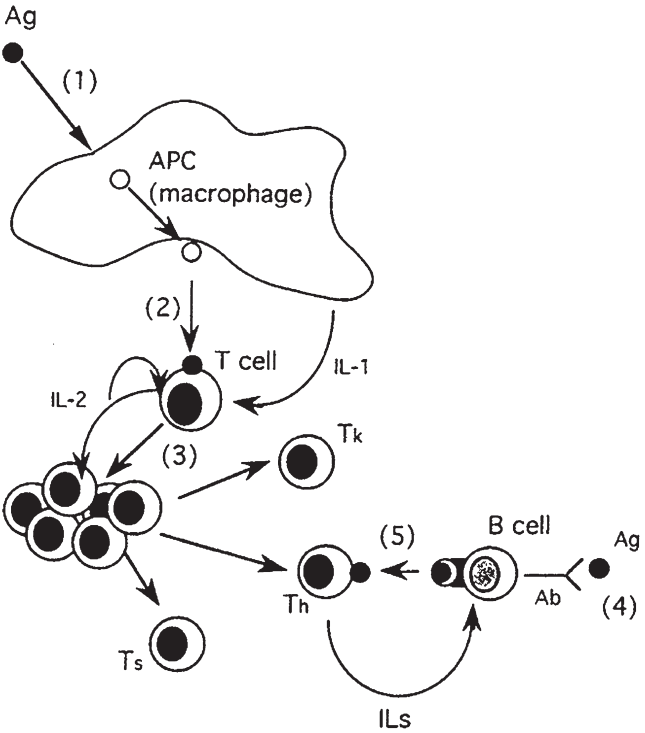


***Figure 3*** *Simplified schematic of immunoregulatory circuit that regulates activation of T cells and B cells involved in humoral (T-cell-dependent) and cell-mediated immunity. (1) Antigen (Ag) is processed by the APCs expressing class II MHC molecules. (2) Antigen plus class II MHC is then presented to antigen-specific T helper cells ($CD4^+$), which stimulates secretion of IL-2. (3) IL-2 in turn stimulates proliferation (clonal expansion) of T cells and differentiation into T suppressor ($T_s$), T killer ($T_k$), and T helper ($T_h$) effector cells. The expanded clone has a higher likelihood of finding the appropriate B cell that has the same antigen and class II molecules on its surface. (4) Next, the antigen binds to an antibody (Ab) on the surface of a specific B cell. (5) The B cell in turn processes the antigen and presents it (plus class II MHC) to the specific $T_h$ cell. The $T_h$ cell is then stimulated to secrete additional ILs that stimulate clonal expansion and differentiation of the antigen-specific B cell.*

conjunction with the class II major histocompatibility complex (MHC) proteins that are located on the surfaces of T cells. The receptor on the T cell is a complex of the Ti molecule that binds antigen, the MHC proteins, and the T3 molecular complex, which is often referred to as the CD3 complex. Upon stimulation, T cells proliferate, differentiate, and express interleukin 2 (IL-2) receptors. T cells also produce and secrete IL-2, which in turn acts on antigen-specific B cells, causing them to proliferate and differentiate into antibody-forming (plasma) cells.

The adaptive arm of the immune system, which depends on delicate tolerance mechanisms to shape the repertoire of the randomly produced antigen receptors, makes this part of the immune system susceptible to the induction of autoimmunity. In contrast to innate immune responses, induction of an adaptive immune response depends on a well-regulated temporal and spatial sequence of activation events, which requires the cooperation of several cell types and the migration of cells between different organs. As a result, activation of adaptive immune responses is a rather slow process, as compared with activation of innate defense mechanisms. All cellular interactions during the induction of an adaptive immune response are based on differentiated intercellular communication pathways involving a host of cytokines and surface receptors.

Antibodies circulate freely in the blood or lymph and are important in neutralizing foreign antigens. The various types of antibodies involved in humoral immunity and their functions are described in Table 5. There are multiple genes (polymorphisms) that encode diversity to the variable region

**TABLE 5 Antibodies Involved in Humoral Immune Response**

| Antibodies | Serum Concentration | | Characteristics/Functions |
|---|---|---|---|
| | $mg\,mL^{-1}$ | % | |
| IgG | 10–12 | 80% | Monomeric structure (γ-globulin); secreted from B cells during secondary response; binds complement; can cross placenta |
| IgM | 1–2 | 5–10% | Pentameric structure; secreted from B cells during primary response; potent binder of complement; high levels indicative of SLE or rheumatoid arthritis; cannot cross placenta |
| IgA | 3–4 | 10–15% | Dimeric or monomeric structures; found in seromucous secretions (breast milk); secreted by B cells associated with epithelial cells in gastrointestinal tract, lung, etc. |
| IgD | 0.03 | <1% | Monomer; extremely labile; functions not well known |
| IgE | <0.0001 | | Reaginic antibody involved in immediate hypersensitivity; antihelminthic; does not bind complement |

*Source*: Adapted from Clark, 1983.

of the antibody. B cells are capable of generating further diversity to antibody specificity by a sequence of molecular events involving somatic mutations, chromosomal rearrangements during mitosis, and recombination of gene segments (Roitt et al., 1985).

The immune system is regulated in part by feedback inhibition involving complex interactions between the various growth and differentiation factors listed in Table 6. Since antigen initiates the signal for the immune response, elimination of antigen will decrease further stimulation (Male et al., 1987). T suppressor ($T_s$) cells also regulate the immune response and are thought to be important in the development of tolerance to self-antigens. In addition to the humoral immune system or the branch that is modulated by antibody, cell-mediated immunity and cytotoxic cell types play a major role in the defense against virally infected cells, tumor cells, and cells of foreign tissue transplants. Cytotoxic $T_k$ (T killer) cells recognize antigen in association with class I molecules of the MHC, while natural killer (NK) cells are not MHC restricted. Cell killing results in a sequence of events following activation of the effector cell, lysosomal degranulation, and calcium influx into the targeted cell. The various types of cells involved in cell-mediated cytotoxicity and their mechanisms of action are outlined in Table 7.

## 4 IMMUNOTOXIC EFFECTS

The immune system is a highly integrated and regulated network of cell types that requires continual renewal to achieve balance and immunocompetence. The delicacy of this balance makes the immune system a natural target for cytotoxic drugs or their metabolites. Since renewal is dependent on the ability of cells to proliferate and differentiate, exposure to agents that arrest cell division can subsequently lead to reduced immune function or immunosuppression. This concept has been exploited in the development of therapeutic drugs intended to treat leukemias, autoimmune disease, and chronic inflammatory diseases and to prevent transplant rejection. However, some drugs may adversely modulate the immune system secondarily to their therapeutic effects.

Two broad categories of immunotoxicity have been defined on the basis of suppression or stimulation of normal immune function. Immunosuppression is a down-modulation of the immune system characterized by cell depletion, dysfunction, or dysregulation that may subsequently result in increased susceptibility to infection and tumors. By contrast, immunostimulation is an increased or exaggerated immune responsiveness that may be apparent in the form of a tissue-damaging allergic hypersensitivity response or pathological autoimmunity. However, as knowledge of the mechanisms involved in each of these conditions has expanded, the distinction between them has become less clear. Some agents can cause immunosuppression at one dose or duration of exposure and immunostimulation at others. For instance, the chemotherapeutic drug cyclophosphamide is in most cases immunosuppressive; however, it

**TABLE 6 Growth and Differentiation Factors of Immune System**

| Factors | Cell of Origin | Primary Immune Functions |
|---|---|---|
| Interleukins[a] | | |
| IL-1 | Macrophage, B and T cells | Lymphocyte-activating factor; enhances activation of T and B cells, NK cells, and macrophages |
| IL-2 | T cells ($T_h$) | T-cell growth factor; stimulates T-cell growth and effector differentiation; stimulates B-cell proliferation/differentiation |
| IL-3 | T cells ($T_h$) | Mast- cell growth factor; stimulates proliferation/ differentiation of mast cells, neutrophils, and macrophages |
| IL-4 | T cells ($T_h$), mast cells, B cells | B-cell growth factor; induces proliferation/ differentiation of B cells and secretion of IgA, $IgG_1$, and IgE; promotes T-cell growth; activates macrophages |
| IL-5 | T cells ($T_h$) | Stimulates antibody secretion (IgA), proliferation of B cells, and eosinophil differentiation |
| IL-6 | T cells, fibroblasts, monocytes | Stimulates growth/differentiation of B cells and secretion of IgG; promotes IL-2-induced growth of T cells |
| IL-7 | Bone marrow stromal cells | Stimulates pre-B- and pre-T-cell growth/ differentiation; enhances thymocyte proliferation |
| IL-8 | Monocytes, fibroblasts | Neutrophil chemotaxis |
| IL-9 | T cells | Stimulates T cells and mast cells |
| IL-10 | T cells | Stimulates mast cells and thymocytes; induction of class II MHC |
| Interferons (INFs) | | |
| A-INF | Leukocytes, mast cells | Antiviral; increases NK-cell function, B-cell differentiation, potentiates macrophage production of IL-1 |
| B-INF | Fibroblasts, epithelial cells | Antiviral; potentiates macrophage production of IL-1; increases NK-cell function |
| $\Gamma$-INF | T cells ($T_h$), cytotoxic T cells | Antiviral; activates macrophages; induces MHC class II expression on macrophages, epithelial, and endothelial cells |
| Tumor necrosis factors (TNFs) | | |
| TNF-$\alpha$ | Macrophage, B and T cells | Catectin; promotes tumor cytotoxicity; activates macrophages and neutrophils; enhances IL-2 receptor expression on T cells; inhibits antibody secretion |
| TNF-$\beta$ | T cells ($T_h$) | Lymphotoxin; promotes T-cell-mediated cytotoxicity |
| | NK cells | B-cell activation |

| | Stem Cells | Promotes Growth and Differentiation of |
|---|---|---|
| Colony stimulating factors (CSFs) | | |
| Granulocyte CSF | Myeloid | Granulocytes and macrophages |
| Macrophage CSF | Myeloid | Macrophages and granulocytes |
| Granulocyte–macrophage CSF | Myeloid | Granulocytes, macrophages, eosinophils, mast cells, and pluripotent progenitor cells |

[a]Includes lymphokines, monokines, and cytokines produced by T cells, macrophages, and other cells, respectively.

*Source*: Adapted from Golub and 2Green, 1991.

**TABLE 7 Cells and Mechanisms Involved in Cell-Mediated Cytotoxicity**

| Cell Type | Mechanism of Cytotoxicity |
|---|---|
| $T_k$ cells | $T_k$ cells that are specifically sensitized to antigens on target cells interact directly with target cells to lyse them. |
| $T_D$ | Cells involved in delayed hypersensitivity that act indirectly to kill target cells; $T_D$ cells react with antigen and release cytokines that can kill target cells. |
| NK cells | Nonspecific T cells that react directly with target cells (tumor cells) without prior sensitization. |
| Null cells | Antibody-dependent cell-mediated cytotoxicity (ADCC) involving non-T/non-B cells (null cells) with $F_c$ receptors specific for antibody-coated target cells. |
| Macrophages | Nonspecific, direct killing of target by phagocytosis; also involved in presenting antigen to specific $T_k$ cells that can then mediate cytotoxicity as described above. |

can also induce autoimmunity (Hutchings et al., 1985). Likewise, dimethylnitrosamine, a nitrosamine detected in some foods, has been shown to have both suppressing and enhancing effects on the immune system (Yoshida et al., 1989).

## 5 IMMUNOSUPPRESSION

The various cells of the immune system may differ in their sensitivity to a given xenobiotic. Thus, immunosuppression may be expressed as varying degrees of reduced activity of a single cell type of multiple populations of immunocytes. Several lymphoid organs such as the bone marrow, spleen, thymus, and lymph nodes may be affected simultaneously or the immunodeficiency may be isolated to a single tissue, such as the Peyer's patches of the intestines. The resulting deficiency may in turn lead to an array of clinical outcomes of varying ranges of severity. These outcomes include increased susceptibility to infections, increased severity or persistence of infections, or infections with unusual organisms (e.g., systemic fungal infections). Immunosuppression can be induced in a dose-related manner by a variety of therapeutic agents at dose levels lower than those required to produce overt clinical signs of general toxicity. In addition, immunosuppression can occur without regard to genetic predisposition given that a sufficient dose level and duration of exposure have been achieved.

Humoral immunity is characterized by the production of antigen-specific antibodies that enhance phagocytosis and destruction of microorganisms through opsonization. Thus, deficiencies of humoral immunity (B lymphocytes) may lead to reduced antibody titers and are typically associated with acute gram-positive bacterial infections (i.e., *Streptococcus*). Although chronic infection is usually associated with dysfunction of some aspect of cellular

immunity, chronic infections can also occur when facultative intracellular organisms such as *Listeria* or *Mycobacterium* evade antibodies and multiply within phagocytic cells.

Since cellular immunity results in the release of chemotactic lymphocytes that in turn enhance phagocytosis, a deficiency in cellular immunity may also result in chronic infections. Cellular immunity is mediated by T cells, macrophages, and NK cells involved in complex compensatory networks and secondary changes. Immunosuppressive agents may act directly by lethality to T cells or indirectly by blocking mitosis, lymphokine synthesis, lymphokine release, or membrane receptors to lymphokines. In addition, cellular immunity is involved in the production and release of interferon, a lymphokine that ultimately acts in blockage of viral replication (Table 6). Viruses are particularly susceptible to cytolysis by T cells since they often attach to the surface of infected cells. Thus, immunosuppression of any of the components of cellular immunity may result in an increase in protozoan, fungal, and viral infections as well as opportunistic bacterial infections.

Immune depression may result unintentionally as a side effect of cancer chemotherapy or intentionally from therapeutics administered to prevent graft rejection. In fact, both transplant patients administered immunosuppressive drugs and cancer patients treated with chemotherapeutic agents have been shown to be at high risk of developing secondary cancers, particularly of lymphoreticular etiology (Penn, 1977). Most of these drugs are alkylating or crosslinking agents that by their chemical nature are electrophilic and highly reactive with nucleophilic macromolecules (protein and nucleic acids). Nucleophilic sites are quite ubiquitous and include amino, hydroxyl, mercapto, and histidine functional groups. Thus, immunotoxic agents used in chemotherapy may induce secondary tumors through direct genotoxic mechanisms (i.e., DNA alkylation).

Reduced cellular immunity may result in increased malignancy and decreased viral resistance through indirect mechanisms as well by modulating immunosurveillance of aberrant cells. T lymphocytes, macrophage cells, and NK cells are all involved in immunosurveillance through cytolysis of virally infected cells or tumor cells, each by a different mechanism (Table 6) (Burnet, 1970). In addition to the common cell types described in Table 7, at least two other types of cytotoxic effector cells of T-cell origin have been identified, each of which has a unique lytic specificity phenotype and activity profile (Merluzzi, 1985). Of these, both lymphocyte-activated killer (LAK) and tumor-infiltrating lymphocyte (TIL) cells have been shown to lyse a variety of different tumor cells. However, TIL cells have 50–100 times more lytic activity than LAK cells. Most tumor cells express unique surface antigens that render them different from normal cells. Once detected as foreign, they are presented to the T helper cells in association with MHC molecules to form an antigen–MHC complex. This association elicits a genetic component to the immunospecificity reaction. T helper cells subsequently direct the antigen complex toward the cytotoxic T lymphocytes, which possess receptors for antigen–MHC complexes. These

cells can then proliferate, respond to specific viral antigens or antigens on the membranes of tumor cells, and destroy them (Yoshida et al., 1989).

In contrast, the macrophages and NK cells are involved in nonspecific immunosurveillance in that they do not require prior sensitization with a foreign antigen as a prerequisite for lysis and are not involved with MHC molecules. The enhancement of either NK-cell function or macrophage function has been shown to reduce metastasis of some types of tumors. Macrophage cells accumulate at the tumor site and have been shown to lyse a variety of transformed tumor cells (Volkman, 1984). NK cells are involved in the lysis of primary autochthonous tumor cells. Migration of NK cells to tumor sites has been well documented. Although not clearly defined, it appears that they can recognize certain proteinaceous structures on tumor cells and lyse them with cytolysin.

### 5.1 Immunosuppressive Drugs

Table 8 lists numerous types of drugs that are immunosuppressive and describes their immunotoxic effects. Several classes of drugs that characteristically depress the immune system are further discussed below.

***Antimetabolites*** This class of drugs includes purine, pyrimidine, and folic acid analogues that have been successfully used to treat various carcinomas, autoimmune diseases, and dermatological disorders such as psoriasis. Because of their structural similarities to normal components of DNA and RNA synthesis, they are capable of competing with the normal macromolecules and alkylating biological nucleophiles.

Thioguanine and mercaptopurine are purine analogues structurally similar to guanine and hypoxanthine that have been used to treat malignancies. Azathioprine, an imidazolyl derivative of mercaptopurine, has been used as an immunosuppressive therapeutic in organ transplants and to treat severe refractory rheumatoid arthritis (Hunter et al., 1975) and autoimmune disorders, including pemphigus vulgaris and bullous pemphigoid. These drugs act as antimetabolites to block de novo purine synthesis through the erroneous incorporation of thioinosinic acid into the pathway in place of inosine. The antimetabolite can bind to the inosine receptor, which in turn will inhibit the synthesis of DNA, RNA, and protein and ultimately T-cell differentiation (Hadden et al., 1984). For example, both thioguanine and mercaptopurine can act as substrates for the hypoxanthine guanine phosphoribosyl transferase (HGPRT) enzyme to produce T-IMP (thioinosine monophosphate) and T-GMP (thioguanine monophosphate), respectively. Thioinosine monophosphate is a poor substrate for guanylyl kinase, which would normally catalyze the conversion of GMP to guanosine diphosphate (GDP) (Calabresi and Chabner, 1990). Thus T-IMP can accumulate in the cell and inhibit several vital metabolic reactions. At high doses, these drugs can suppress the entire immune system. However, at clinical dosages, only the T-cell response is affected,

**TABLE 8 Immunosuppressive Drugs and Their Effects**

| Drugs | Biological Activity and Indications | Immunotoxic Effects |
|---|---|---|
| | *Hormones and Antagonists* | |
| Corticosteroids (prednisone) | Anti-inflammatory; SLE; leukemias; rheumatoid arthritis; breast cancer | Depresses T- and B-cell function; reduces lymphokines; alters macrophage function; increases infections |
| Diethylstilbestrol | Synthetic estrogen; cancer chemotherapy | Depletes or functionally impairs T cells; enhances macrophage suppressor cell; increases infections and tumorigenesis |
| Estradiol | Synthetic estrogen; dysmenorrhea; osteoporosis | Decreases $T_h$ cells and IL-2 synthesis; increases $T_s$ cell function, infections, and tumorigenesis |
| | *Antibiotics* | |
| Cephalosporins | β-Lactam antimicrobial | Granulocytopenia; cytopenia |
| Chloramphenol | Wide-spectrum antimicrobial | Pancytopenia, leukopenia (idiosyncratic) |
| Penicillins | β-Lactam antimicrobial | Granulocytopenia; cytopenia |
| Rifampin | Macrocyclic antibiotic | Suppresses T-cell function |
| Tetracyclines | Antimicrobial | Decreased migration of granulocytes |
| | *Chemotherapeutics and Immunomodulators* | |
| Arabinoside (AraA, AraC) | Antimetabolites; antivirals; leukemias; lymphomas | Leukopenia; thrombocytopenia |
| Azathioprine | Antimetabolite; leukemia; arthritis; transplant rejection | Inhibits protein synthesis; bone marrow suppression |
| Busulfan | Alkylating agent; chronic granulocytic leukemia | Leukopenia; myelosuppressive; granulocytopenia |
| Carmutin, Lomustin (BCNU, CCNU) | Alkylating agents; Hodgkin's disease; lymphomas | Delayed hematopoietic depression; leukopenia; thrombocytopenia |
| Chlorambucil | Alkylating agent; leukemia; lymphomas; vasculitis | Bone marrow suppression; myelosuppressive |
| Cyclophosphamide (cytotoxin) | Alkylating agent; cancer chemotherapy; transplant rejection; rheumatoid arthritis | Decreased $T_s$ cells, B cells, and NK cells |
| Cyclosporin A | Transplant rejections | Depresses T cells; inhibits IL-2 production |
| Interferon | Immunomodulator; antiviral, hairy cell leukemia | Bone marrow suppression; granulocytopenia; leukopenia |

**TABLE 8** ***Continued***

| Drugs | Biological Activity and Indications | Immunotoxic Effects |
|---|---|---|
| Melphalan (L-PAM) | Alkylating agent; breast and ovarian cancer | Leukopenia; bone marrow suppression; granulocytopenia; pancytopenia |
| 6-Mercaptopurine | Antimetabolite; acute leukemias; arthritis | Decreased T-cell function; bone marrow suppression |
| Methotrexate | Folic acid analogue; cancer chemotherapy, arthritis | Inhibits proliferation; T-cell suppression; granulocytopenia; lymphocytopenia |
| Penostatin | Adenosine analog; T-cell leukemia | Inhibits adenosine deaminase; suppresses T and B cells |
| Zidovudine (AZT) | Antiviral (HIV) | Decreases $T_h$ cells and granulocytes |
| | *Miscellaneous* | |
| Colchicine | Antimitotic; gout; anti-inflammatory | Inhibits migration of granulocytes; leukopenia; agranulocytosis |
| Diphenylhydantoin (phenytoin) | Antiepileptic | Leukocytopenia; neutrapenia |
| Indomethacin (indocin) | Nonsteroidal anti-inflammatory; analgesic; antipyretic | Neutrapenia |
| Procainamide | Antiarrhythmic | Agranulocytosis; leukopenia (rare) |
| Sulfasalazine | Antimicrobial anti-inflammatory; ulcerative colitis/inflammatory bowel diseases | Suppresses NK cells; impaired lymphocyte function |

*Source*: Adapted primarily from Brunton et al., 2006.

without an apparent decrease in T-cell numbers (Spreafico and Anaclerio, 1977).

Pentostatin (2′-deoxycoformycin) is an adenosine analogue that is a potent inhibitor of adenosine deaminase. Pentostatin is particularly useful for treating T-cell leukemia since malignant T cells have higher levels of adenosine deaminase than most cells. Similar to individuals that are genetically deficient in adenosine deaminase, treatment with pentostatin produces immunosuppression of both T and B lymphocytes, with minimal effect on other tissues. As a result, severe opportunistic infections are often associated with its clinical use.

5-Fluorouracil (5-FU), adenosine arabinoside (AraA), and cytosine arabinoside (AraC) are pyrimidine analogues to uracil, adenine, and cytosine, respectively. 5-FU is used primarily to treat cancer of the breasts and gastrointestinal tract as well as severe recalcitrant psoriasis (Alper et al., 1985). AraC is predominantly indicated for the treatment of acute leukemia and non-Hodgkin's lymphomas. Although high-dose therapy with AraC has a good

likelihood of producing complete remission, it is often accompanied by severe leukopenia, thrombocytopenia, and anemia (Barnett et al., 1985). Likewise, myelosuppression is the major toxicity associated with bolus-dose regimens of 5-FU.

***Glucocorticosteroids*** Corticosteroids are commonly used to reduce inflammation, to treat autoimmune diseases such as systemic lupus erythematosus (SLE), and as a prophylactic measure to prevent transplant rejection. The adrenocorticosteroid prednisone is often coadministered with other immunosuppressives such as cyclosporine and azathioprine (Elion and Hitchings, 1975). Glucocorticosteroids act pharmacologically by modulating the rate of protein synthesis. The molecule reacts with specific receptors to form a complex that crosses into the nucleus of the cell and regulates transcription of specific messenger RNA (mRNA). The corticosteroid complex releases inhibition of transcription, thus enhancing protein synthesis (Hollenberg et al., 1987). This may lead to the initiation of de novo synthesis of the phospholipase A2–inhibiting protein lipocortin, which blocks the synthesis of arachidonic acid and its prostaglandin and leukotriene metabolites (Haynes and Murad, 1985; Wallner et al., 1986). Glucocorticosteroids induce immunosuppression and anti-inflammation as a result of the inhibition of specific leukocyte functions such as lymphokine activity. Glucocorticoids can also inhibit recruitment of leukocytes and macrophages into the site of inflammation. In addition, amplification of cell-mediated immunity can be suppressed by inhibiting the interaction of IL-2 with its T-cell receptors. However, the immunosuppression is reversible and immune function recovers once therapy has ceased.

***Cyclosporine*** Cyclosporin A (cyclosporine) is an 11-amino-acid cyclic peptide residue of fungal origin isolated from the fermentation products of *Trichoderma polysporum* and *Cylindrocarpon lucidum.* In addition to having a very narrow range of antibiotic activity, it was also found to inhibit proliferation of lymphocytes, which made it unsuitable as an antibiotic. Cyclosporine inhibits the early cellular response of helper T cells to antigens (Kay and Benzie, 1984) primarily by inhibiting production of IL-2 (Elliot et al., 1984), and at higher doses it may inhibit expression of IL-2 receptors (Herold et al., 1986). Cyclosporine does not prevent the stimulation of helper-T-cell clonal expansion by IL-2, only its activation. Since it is not myelosuppressive at therapeutic dosages, the incidence of secondary infection is lower than that induced by other classes of immunosuppressives. Thus, cyclosporine is ideal as an immunosuppressive agent to prevent transplant rejection and graft-host disease (Kahan and Bach, 1988). Cyclosporine has also been used as an anti-helminthic and as an anti-inflammatory agent to treat rheumatoid arthritis and other autoimmune-type diseases.

***Nitrogen Mustards*** Nitrogen mustards characteristically consist of a bis(2-chloroethyl) group bonded to nitrogen. These molecules are highly reactive

bifunctional alkylating agents that have been successfully used in cancer chemotherapy. Included in this group are mechlorethamine, L-phenylalanine mustard (melphalan), chlorambucil, ifosfamide, and cyclophosphamide. The cytotoxic effects of each on the bone marrow and lymphoid organs are similar; however, their pharmacokinetic and toxic profiles can vary on the basis of the substituted side group. For example, the side group may consist of a simple methyl group, as is the case of mechlorethamine, or substituted phenyl groups, in the cases of melphalan and chlorambucil.

Cyclophosphamide, which contains a cyclic phosphamide group bonded to the nitrogen mustard, is representative of this class. The parent compound itself is not active in vitro unless treated in conjunction with an exogenous P450 microsomal enzyme system (Colvin, 1982) such as rat liver S9 homogenate, which metabolizes it to a highly reactive alkylating agent (4-hydroxycyclophosphamide). Thus, in vivo, cyclophosphamide is not toxic until it is metabolically activated in the liver. Cyclophosphamide has been the most widely used nitrogen mustard where it has been effective as a cancer chemotherapeutic and to treat autoimmune-type diseases including SLE, multiple sclerosis, and rheumatoid arthritis (Calabresi and Parks, 1985). Treatment with cyclophosphamide suppresses all classes of lymphoid cells, which may result in reduced lymphocyte function as well as lymphopenia and neutropenia (Webb and Winkelstein, 1982). Thus, it has also been administered as a large single dose prior to bone marrow transplants to suppress cellular immunity and subsequently inhibit rejection (Shand, 1979).

***Estrogens*** β-Estradiol (Luster et al., 1984; Pung et al., 1984) and therapeutics with estrogenic activity, such as diethylstilbestrol (DES), have also been shown to be immunosuppressive (Luster et al., 1985). Estrogens have been shown to increase T-suppressor-cell activity in splenocytes, decrease numbers of T helper cells, inhibit IL-2 synthesis, and modulate production of immunoregulatory factors (Luster et al., 1987). These effects have been particularly characterized in studies with DES, a nonsteroidal synthetic estrogen used widely in the treatment of prostate and breast cancers, as well as administered to pregnant women as a "morning after" contraceptive. Decreased mitogenicity of human peripheral blood lymphocytes has been observed in men treated with DES for prostate cancer and women exposed in utero (Haukass et al., 1982; Ways et al., 1987). In mice, thymic involution and atrophy with depletion of the cortical lymphocytes have been observed histologically. Function is also modulated, as evident by depressed mixed-lymphocyte responses, mitogenicity, and T-cell release of IL-2 (Pung et al., 1985). Dean et al. (1980) speculated that DES treatment selectively depletes or functionally impairs T cells and/or the induction of suppressor macrophages, resulting in immunosuppression. Macrophage suppressor cell activity is enhanced (Luster et al., 1980) and PMN cells accumulate following bacterial challenge. Although macrophage functions of phagocytosis and tumor growth inhibition are potentiated, defects in macrophage migration and decreased bactericidal activity contribute to

decreased host resistance with resulting increased susceptibility to bacterial infections.

***Heavy Metals*** Some heavy metals such as gold and platinum are used pharmacologically as immunomodulators to treat rheumatoid arthritis and as antineoplastic drugs, respectively. Most heavy metals inhibit mitogenicity, antibody responses, and host resistance to bacterial or viral challenge and tumor growth. Platinum has been shown to suppress humoral immunity, lymphocyte proliferation, and macrophage function (Lawrence, 1985). Clinically, mild to moderate myelosuppression may also be evident with transient leukopenia and thrombocytopenia.

Likewise, injectable gold salts such as gold sodium thiomalate affect a variety of immune responses in humans (Bloom et al., 1987). Severe thrombocytopenia occurs in 1% of patients as a result of an immunological disturbance that accelerates the degradation of platelets. Leukopenia, agranulocytosis, and fatal aplastic anemia may also occur. Although better tolerated than parenteral preparations, administered orally the organic gold compound auranofin is also immunosuppressive. In a dog study, auranofin was shown to produce thrombocytopenia similar to that described in humans administered parenteral preparations (Bloom et al., 1985b). Long-term toxicity studies with these compounds in dogs show evidence of immune-modulating activity, possible drug-induced immunotoxicity, and treatment-related changes in immune function (e.g., lymphocyte activation).

***Antibiotics*** β-Lactam-containing antibiotics such as the cephalosporins may also induce significant immunosuppressive effects (Caspritz and Hadden, 1987) in a small percentage of human patients. Adverse effects including anemia, neutropenia, thrombocytopenia, and bone marrow depression were observed in dogs administered high doses of cefonicid for six months (Bloom et al., 1985a). A similar syndrome has been characterized in cefazedone-treated dogs expressing an agglutinating red cell antibody. Further studies with this drug indicated that both cytopenia (Bloom et al., 1985a) and suppression of bone marrow stem cell activity appear to be antibody mediated (Deldar et al., 1985).

## 6 IMMUNOSTIMULATION

A variety of drugs as well as environmental chemicals have been shown to have immunostimulatory or sensitizing effects on the immune system and these effects are well documented in humans exposed to drugs (DeSwarte, 1986). The drug or metabolite can act as a hapten and covalently bind to a protein or other cellular constituent of the host to appear foreign and become antigenic. Haptens are low-molecular-weight substances that are not in themselves immunogenic but will induce an immune response if conjugated with

nucleophilic groups on proteins or other macromolecular carriers. In both allergy and autoimmunity, the immune system is stimulated or sensitized by the drug conjugate to produce specific pathological responses. An allergic hypersensitivity reaction may vary from one which results in an immediate anaphylactic response to one which produces a delayed hypersensitivity reaction or immune complex reaction. Allergic hypersensitivity reactions result in a heightened sensitivity to nonself-antigens, whereas autoimmunity results in an altered response to self-antigens. Unlike immunosuppression, which nonspecifically affects all individuals in a dose-related manner, both allergy and autoimmunity have a genetic component that creates susceptibility in those individuals with a genetic predisposition. Susceptible individuals, once sensitized, can respond to even minute quantities of the antigen. Several examples of drugs that can stimulate the immune system are presented in Table 9.

### 6.1 Hypersensitivity (or Allergenicity)

The four types of hypersensitivity reactions as classified by Coombs and Gell (1975) are outlined in Table 10. The first three types are immediate antibody-mediated reactions, whereas the fourth type is a cellular-mediated delayed-type response that may require one to two days to occur after a secondary exposure. Type I reactions are characterized by an anaphylaxis response to a variety of compounds, including proteinaceous materials and pharmaceuticals such as penicillin. Various target organs may be involved depending on the route of exposure. For example, the gastrointestinal tract is usually involved with food allergies, the respiratory system with inhaled allergens, the skin with dermal exposure, and smooth muscle vasculature with systemic exposure. The type of response elicited often depends on the site of exposure and includes dermatitis and urticaria (dermal), rhinitis and asthma (inhalation), increased gastrointestinal emptying (ingestion), and systemic anaphylactic shock (parenteral) (Steele et al., 1989).

***Type I Hypersensitivity*** During an initial exposure, immunoglobulin E (IgE) antibodies are produced and bind to the cell surface of mast cells and basophils. Upon subsequent exposures to the antigen, reaginic IgE antibodies bound to the surface of target cells at the $F_c$ region (mast cells and basophils) become crosslinked (at the $F_{ab}$ regions) by the antigen. Crosslinking causes distortion of the cell surface and IgE molecule, which in turn activates a series of enzymatic reactions, ultimately leading to degranulation of the mast cells and basophils. These granules contain a variety of pharmacological substances (Table 11), such as histamines, serotonins, prostaglandins, bradykinins, and leukotrienes [slow-reacting substance of anaphylaxis (SRS-A) and ecdysone receptor (ECR)-A]. Upon subsequent challenge exposures, these factors are responsible for eliciting an allergic reaction through vasodilation and increased vascular permeability. The nasal passages contain both mast cells and plasma cells that secrete IgE antibodies. Allergic responses localized in the nasal mucosa result in dilation of

**TABLE 9 Drugs That Produce Immunostimulation**

| Drug | Type of Response |
|---|---|
| | *Hypersensitivity* |
| Antibiotics | |
| Cephalosporins | Anaphylaxis, urticaria, rash, granulocytopenia |
| Chloramphenicol | Rash, dermatitis, urticaria |
| Neomycin | Dermal exposure—rash, dermatitis |
| Sulfathiazole | Rash, dermatitis, urticaria |
| Spiramycin | Rash, dermatitis, urticaria |
| Quinolones | Photosensitivity |
| Tetracyclines | Photosensitivity, anaphylaxis, asthma, dermatitis |
| Others | |
| Allopurinol | Rash, urticaria, fever, eosinophilia |
| Avridine | DTH; increases NK cells, T cells, IL-1, and IL-2 |
| Isoprinosine | DTH; increases T lymphocytes |
| Indomethacin | Rash, urticaria, asthma, granulocytopenia |
| Quinidine | Fever, anaphylaxis, asthma |
| Salicylates | Rash, urticaria |
| | *Autoimmunity* |
| Amiodarone | Thyroiditis |
| Captopril | Autoimmune hemolytic anemia, pemphigus, granulocytopenia |
| Chlorpromazine | Granulocytopenia |
| Halothane | Autoimmune chronic active hepatitis |
| Hydralizine | Autoimmune hemolytic anemia, drug-induced SLE, myasthenia gravis, pemphigus, glomerulonephritis, Goodpasture's disease |
| Methyldopa | Autoimmune hemolytic anemia, leukopenia, drug-induced SLE, pemphigus |
| Nitrofurantoin | Peripheral neuritis |
| D-Penicillamine | Autoimmunity, drug-induced SLE, myasthenia gravis, pemphigus, glomerulonephritis, Goodpasture's disease |
| Propranolol | Autoimmunity |
| Procainamide | Autoimmunity, drug-induced SLE, rash, vasculitis, myalgias |
| Pyrithioxine | Pemphigus |
| | *Hypersensitivity and Autoimmunity* |
| Antibiotics | |
| Isoniazid | Rash, dermatitis, vasculitis, arthritis, drug-induced SLE |
| Penicillins | Anaphylaxis, dermatitis; vasculitis, serum sickness, hemolytic anemia |
| Sulfonamides | Dermatitis, photosensitivity, pemphigus, hemolytic anemia, serum sickness, drug-induced SLE |
| Others | |
| Acetazolamide | Rash, fever, autoimmunity |
| Lithium | Dermatitis, autoimmune thyroiditis, vasculitis |
| Thiazides | Hypersensitivity, photosensitivity, autoimmunity (diabetes) |
| Phenytoin | Rash, drug-induced SLE, hepatitis |

**TABLE 10 Types of Hypersensitivity Responses**

| Type and Designation | Agents: Clinical Manifestations | Components | Effects | Mechanism |
|---|---|---|---|---|
| I: Immediate (reaginic) | Food additives (GI allergies; anaphylactic)<br>Penicillin: urticaria and dermatitis | Mast cells; IgE | Anaphylaxis, asthma, urticaria, rhinitis, dermatitis | IgE binds to mast cells to stimulate release of humoral factors |
| II: Cytotoxic | Cephalosporine: hemolytic anemia<br>Quinidine: thrombocytopenia | IgG, IgM | Hemolytic anemia, Goodpasture's disease | IgG and IgM bind to cells (e.g., RBCs), fix complement (opsonization), then lyse cells |
| III: Immune complex (arthus) | Methicillin: chronic glomerulonephritis | Antigen–antibody complexes (Ag–Ab) | SLE, rheumatoid arthritis, glomerular nephritis, serum sickness, vasculitis | Ag–Ab complexes deposit in tissues and may fix complement |
| IV: Delayed hypersensitivity | Penicillin: contact dermatitis | $T_D$ cells; macrophages | Contact dermatitis, tuberculosis | Sensitized T cells induce a delayed-hypersensitivity response upon challenge |

*Source*: Based on classification system of Coombs and Gell, 1975.

**TABLE 11 Proteins and Soluble Mediators Involved in Hypersensitivity**

| Factor | Origin | Characteristics/Functions |
|---|---|---|
| Histamine | Mast cells, basophils | Contraction of smooth muscle; increases vascular permeability |
| Serotonin | Mast cells, basophils | Contraction of smooth muscle; leukotriene |
| SRS-A | Lung tissue | (Slow-reacting substance of anaphylaxis); Contraction of smooth muscle; acidic polypeptide |
| ECF-A | Mast cells | (Eosinophilic chemotactic factor of anaphylaxis); attracts eosinophils; small peptide |
| Prostaglandins | Various tissues | Modifies release of histamine and serotonin from mast cells and basophils |

*Source*: Extracted and modified from Clark, 1983.

the local blood vessels, tissue swelling, mucous secretion, and sneezing. Reactions localized in the respiratory tract, also rich in mast cells and IgE, result in an allergic asthma response. This condition is triggered by the release of histamine and SRS-A, which induce constriction of the bronchi and alveoli, pulmonary edema, and mucous secretions that block the bronchi and alveoli, together resulting in severe difficulty in breathing. In the case of a challenge dose of a drug administered systemically, the reactive patient may have difficulty breathing within minutes of exposure and may experience convulsions, vomiting, and low blood pressure. The effects of anaphylactic shock and respiratory distress, if severe, may ultimately result in death.

Antibiotics containing β-lactam structures, such as penicillin and cephalosporins, are the most commonly occurring inducers of anaphylactic shock and drug hypersensitivity in general. Other hypersensitivity reactions may include urticarial rash, fever, bronchospasm, serum sickness, and vasculitis with reported incidences of all types varying from 0.7 to 10% (Idsøe et al., 1968) and the incidence of anaphylactoid reactions varying from 0.04 to 0.2%. When the β-lactam ring is opened during metabolism, the penicilloyl moiety can form covalent conjugates with nucleophilic sites on proteins. The penicilloyl conjugates can then act as haptens to form the determinants for antibody induction. Although most patients that have received penicillin produce antibodies against the metabolite benzylpenicilloyl, only a fraction experience allergic reaction (Garratty and Petz, 1975), which suggests a genetic component to susceptibility.

***Type II Hypersensitivity*** Type II cytolytic reactions are mediated by IgG and IgM antibodies that can fix complement, opsonize particles, or induce antibody-dependent cellular cytolysis reactions. Erythrocytes, lymphocytes, and platelets of the circulatory system are the major target cells that interact with the cytolytic antibodies causing depletion of these cells. Hemolytic anemia (penicillin, methyldopa), leukopenia, thrombocytopenia (quinidine), and/or granulocytopenia (sulfonamide) may result. Type II reactions involving the

lungs and kidneys occur through the development of antibodies (autoantibodies) to the basement membranes in the alveoli or glomeruli, respectively. Prolonged damage may result in Goodpasture's disease, an autoimmune disease characterized by pulmonary hemorrhage and glomerulonephritis. Several other autoimmune-type diseases have been associated with extended treatments with D-penicillamine and other pharmaceuticals. Various types of autoimmune responses and examples of drug-induced autoimmunity are discussed in further detail later in this section.

***Type III Hypersensitivity*** Type III reactions (arthus) are characterized as immediate hypersensitivity reactions initiated by antigen–antibody complexes that form freely in the plasma instead of at the cell surface. Regardless of whether the antigens are self or foreign, complexes mediated by IgG can form and settle into the tissue compartments of the host. These complexes can then fix complement and release C3a and C5a fragments that are chemotactic for phagocytic cells. Polymorphonuclear leukocytes are then attracted to the site, where they phagocytize the complexes and release hydrolytic enzymes into the tissues. Additional damage can be caused by binding to and activating platelets and basophils, which, in the end, results in localized necrosis, hemorrhage, and increased permeability of local blood vessels. These reactions commonly target the kidney, resulting in glomerulonephritis through the deposition of the complexes in the glomeruli.

Some antibiotics (β-lactam) have been reported to produce glomerular nephritis in humans that has been attributed to circulating immune complexes. These complexes have also been observed in preclinical toxicology studies with baboons treated with a β-lactam antibiotic prior to the appearance of any biochemical or clinical changes (Descotes and Mazue, 1987). In addition, immunoglobulin complexes have been observed in rats treated with gold and autologus immune complex nephritis has been observed in guinea pigs (Ueda et al., 1980). Similar evidence of immunomediated nephrotoxicity has been reported in rheumatoid arthritis patients administered long-term treatments with gold compounds; proteinuria has been observed in approximately 10% of these patients.

Other target organs such as the skin with lupus, the joints with rheumatoid arthritis, and the lungs with pneumonitis may be affected. The deposition of antigen–antibody complexes through the circulatory system results in a syndrome referred to as serum sickness, which was quite prevalent prior to 1940 (Clark, 1983), when serum therapy for diphtheria was commonly used. Serum sickness occurs when the serum itself becomes antigenic as a side effect from passive immunization with heterologous antiserum produced from various sources of farm animals. The antitoxin for diphtheria was produced in a horse and administered to humans as multiple injections of passive antibody. As a consequence, these people often became sensitized to the horse serum and developed a severe form of arthritis and glomerulonephritis caused by deposition of antigen–antibody complexes. Clinical symptoms of serum sickness

present as urticarial skin eruptions, arthralgia or arthritis, lymphadenopathy, and fever. Drugs such as sulfonamides, penicillin, and iodides can induce a similar type of reaction. Although uncommon today, transplant patients receiving immunosuppressive therapy with heterologous antilymphocyte serum or globulins may also exhibit serum sickness.

***Type IV Delayed-Type Hypersensitivity (DTH)*** Delayed-type hypersensitivity reactions are T-cell mediated with no involvement of antibodies. However, these reactions are controlled through accessory cells, suppressor T cells, and monokine-secreting macrophages, which regulate the proliferation and differentiation of T cells. The most frequent form of DTH manifests itself as contact dermatitis. The drug or metabolite binds to a protein in the skin or the Langerhans cell membrane (class II MHC molecules) where it is recognized as an antigen and triggers cell proliferation. After a sufficient period of time for migration of the antigen and clonal expansion (latency period), a subsequent exposure will elicit a dermatitis reaction. A 24–48-h delay often occurs between the time of exposure and onset of symptoms to allow time for infiltration of lymphocytes to the site of exposure. The T cells ($CD4^+$) that react with the antigen are activated and release lymphokines that are chemotactic for monocytes and macrophages. Although these cells infiltrate to the site via the circulatory vessels, an intact lymphatic drainage system from the site is necessary since the reaction is initiated in drainage lymph nodes proximal to the site (Clark, 1983). The release (degranulation) of enzymes and histamines from the macrophages may then result in tissue damage. Clinical symptoms of local dermal reactions may include a rash (not limited to sites of exposure), itching, and/or burning sensations. Erythema is generally observed in the area around the site, which may become thickened and hard to the touch. In severe cases, necrosis may appear in the center of the site followed by desquamation during the healing process. The immune-enhancing drugs isoprinosine and avridine have been shown to induce a DTH reaction in rats (Exon et al., 1986).

A second form of DTH response is similar to that of contact dermatitis in that macrophages are the primary effector cells responsible for stimulating $CD4^+$ T cells; however, this response is not necessarily localized to the epidermis. A classical example of this type of response is demonstrated by the tuberculin diagnostic tests. To determine if an individual has been exposed to tuberculosis, a small amount of fluid from tubercle bacilli cultures is injected subcutaneously. The development of induration after 48 h at the site of injection is diagnostic of prior exposure.

Shock, similar to that of anaphylaxis, may occur as a third form of a delayed systemic hypersensitivity response. However, unlike anaphylaxis, IgE antibodies are not involved. This type of response may occur 5–8 h after systemic exposure and can result in fatality within 24 h following intravenous or intraperitoneal injection.

A fourth form of delayed hypersensitivity results in the formation of granulomas. If the antigen is allowed to persist unchecked, macrophages and

fibroblasts are recruited to the site to proliferate, produce collagen, and effectively "wall off" the antigen. A granuloma requires a minimum of one to two weeks to form.

### 6.2 Photosensitization

Regardless of the route of exposure, some haptens (photoantigens) that are absorbed locally into the skin or reach the skin through systemic absorption can be photoactivated by ultraviolet (UV) light between 320 and 400 nm. Once activated, the hapten can bind to the dermal receptors to initiate sensitization (photoallergy). Subsequent exposures to the hapten in the presence of UV light can result in a hypersensitivity response. Clinical symptoms of photoallergy may occur within minutes (immediate hypersensitivity) of exposure to sunlight or 24 h or more after exposure (DTH). Symptoms may range from acute urticarial reactions to eczematous or papular lesions. Although both phototoxic and photoallergic reactions require the compound to be exposed to sunlight in order to elicit a response, their mechanisms of action are quite different. Since photosensitization is an immune-mediated condition, repeated exposures with a latency period between the initial exposure and subsequent exposures is required, the response is not dose related (small amounts can produce a response once sensitized), and not all individuals exposed to the compound will necessarily respond (genetic component to susceptibility). Although both conditions can present similar symptoms (erythema), phototoxicity is limited mainly to erythema, whereas photoallergy can result in erythema, edema, and dermatitis as described above.

Several drug classes, including tetracycline, sulfonamide, and quinolone antibiotics, as well as chlorothiazide, chlorpromazine, and amiodarone hydrochloride, have been shown to be photoantigens. Photosensitivity may persist even after withdrawal of the drug, as has been observed with the antiarrhythmic drug amiodarone hydrochloride, since it is lipophilic and can be stored for extended periods in the body fat (Unkovic et al., 1984). In addition, it is quite common for cross-reactions to occur between structurally related drugs of the same class.

### 6.3 Autoimmunity

In autoimmunity, as with hypersensitivity, the immune system is stimulated by specific responses that are pathogenic, and both tend to have a genetic component that predisposes some individuals more than others. However, as is the case with hypersensitivity, the adverse immune response of drug-induced autoimmunity is not restricted to the drug itself but also involves a response to self-antigens.

Autoimmune responses directed against normal components of the body may consist of antibody-driven humoral responses and/or cell-mediated DTH responses. T cells can react directly against specific target organs, or B cells

can secrete autoantibodies that target "self." Autoimmunity may occur spontaneously as the result of a loss of regulatory controls that initiate or suppress normal immunity causing the immune system to produce lymphocytes reactive against its own cells and macromolecules such as DNA, RNA, or erythrocytes.

Although autoantibodies are often associated with autoimmune reactions, they are not necessarily indicative of autoimmunity (Russel, 1981). Antinuclear antibodies can occur normally with aging in some healthy women without autoimmune disease, and all individuals have B cells with the potential of reacting with self-antigens through Ig receptors (Dighiero et al., 1983). The presence of an antibody titer to a particular immunogen indicates that haptenization of serum albumin has occurred as part of a normal immune response. However, if cells are stimulated to proliferate and secrete autoantibodies directed against a specific cell or cellular component, a pathological response may result. The tissue damage associated with autoimmune disease is usually a consequence of type II or III hypersensitivity reactions that result in the deposition of antibody–antigen complexes.

Several diseases have been associated with the production of autoantibodies against various tissues. For example, an autoimmune form of hemolytic anemia can occur if the antibodies are directed against erythrocytes. Similarly, antibodies that react with acetylcholine receptors may cause myasthenia gravis, those directed against glomerular basement membranes may cause Goodpasture's syndrome, and those that target the liver may cause hepatitis. Other forms of organ-specific autoimmunity include autoimmune thyroiditis (as seen with amiodarone) and juvenile diabetes mellitus, which can result from autoantibodies directed against the tissue-specific antigens thyroglobulin and cytoplasmic components of pancreatic islet cells, respectively. In contrast, systemic autoimmune diseases may occur if the autoantibodies are directed against an antigen that is ubiquitous throughout the body, such as DNA or RNA. For example, SLE occurs as the result of autoimmunity to nuclear antigens that form immune complexes in the walls of blood vessels and basement membranes of tissues throughout the body.

The etiology of drug-induced autoimmunity is not well established and is confounded by factors such as age, sex, and nutritional state as well as genetic influences on pharmacological and immune susceptibility. Unlike idiopathic autoimmunity, which is progressive or characterized by an alternating series of relapses and remissions, drug-induced autoimmunity is thought to subside after the drug is discontinued. However, this is not certain since a major determining factor for diagnosis of a drug-related disorder is dependent on the observation of remission upon withdrawal of the drug (Bigazzi, 1988).

One possible mechanism for xenobiotic-induced autoimmunity involves xenobiotic binding to autologus molecules, which then appear foreign to the immunosurveillance system. If a self-antigen is chemically altered, a specific T helper ($T_h$) cell may see it as foreign and react to the altered antigenic determinant portion, allowing an autoreactive B cell to react to the unaltered

hapten. This interaction results in a carrier–hapten bridge between the specific $T_h$ and autoreactive B cell, bringing them together for subsequent production of autoantibodies specific to the self-antigen that was chemically altered (Weigle, 1980). Conversely, a xenobiotic may alter B cells directly, including those that are autoreactive. Thus, the altered B cells may react to self-antigens independent from $T_h$-cell recognition and in a non-tissue-specific manner.

Another possible mechanism is that the xenobiotic may stimulate nonspecific mitogenicity of B cells. This could result in a polyclonal activation of B cells with subsequent production of autoantibodies. Alternatively, the xenobiotic may stimulate mitogenicity of T cells that recognize self, which in turn activate B-cell production of antibodies in response to "self" molecules. There is also evidence to suggest that anti-DNA autoantibodies may originate from somatic mutations in lymphocyte precursors with antibacterial or antiviral specificity. For example, a single amino acid substitution resulting from a mutation in a monoclonal antibody to polyphorylcholine was shown to result in a loss of the original specificity and an acquisition of DNA reactivity similar to that observed for anti-DNA antibodies in SLE (Talal, 1987).

The mechanisms of autoimmunity may also entail interaction with MHC structures determined by the human leukocyte antigen (HLA) alleles. Individuals carrying certain HLA alleles have been shown to be predisposed to certain autoimmune diseases, which may account in part for the genetic variability of autoimmunity. In addition, metabolites of a particular drug may vary between individuals to confound the development of drug-induced autoimmunity. Dendritic cells, such as the Langerhans cells of the skin and B lymphocytes that function to present antigens to $T_h$ cells, express class II MHC structures. Although the exact involvement of these MHC structures is unknown, Gleichmann et al. (1989) have theorized that self-antigens rendered foreign by drugs such as D-penicillamine may be presented to $T_h$ cells by MHC class II structures. An alternate hypothesis is that the drug or a metabolite may alter MHC class II structures on B cells, making them appear foreign to $T_h$ cells.

A number of different drugs have been shown to induce autoimmunity in susceptible individuals. A syndrome similar to that of SLE was described in a patient administered sulfadiazine in 1945 by Hoffman (see Bigazzi, 1988). Sulfonamides were one of the first classes of drugs identified to induce an autoimmune response, while to date, over 40 other drugs have been associated with a similar syndrome.

Autoantibodies to red blood cells and autoimmune hemolytic anemia have been observed in patients treated with numerous drugs, including procainamide, chlorpropamide, captopril, cefalexin, penicillin, and methyldopa (Logue et al., 1970; Kleinman et al., 1984). Hydralazine- and procainamide-induced autoantibodies may also result in SLE. Approximately 20% of patients administered methyldopa for several weeks for the treatment of essential hypertension developed a dose-related titer and incidence of autoantibodies to erythrocytes, 1% of which presented with hemolytic anemia. Methyldopa does

not appear to act as a hapten but appears to act by modifying erythrocyte surface antigens. IgG autoantibodies then develop against the modified erythrocytes.

D-Penicillamine is used to treat patients with rheumatoid arthritis, to reduce excess cystine excretion in patients with cystinurias, and as a chelating agent for copper in patients with Wilson's disease. D-Penicillamine can cause multiple forms of autoimmunity, including SLE, myasthenia gravis, pemphigus, and autoimmune thyroiditis. This drug is thought to act as immunomodulator in patients by initiating or even potentiating anti-DNA antibody synthesis (Mach et al., 1986). The highly reactive thiol group may react with various receptors and biological macromolecules to induce autoantibodies. Long-term (many months) treatment has been shown to induce autoimmunity resulting in myasthenia gravis in 0.5% of patients (Bigazzi, 1988) and SLE in approximately 2% of patients as exhibited by varying degrees of joint pain, synovitis, myalgia, malaise, rash, nephritis, pleurisy, and neurological effects. In patients exhibiting myasthenia gravis, D-penicillamine may act to alter the acetylcholine receptors. Autoantibodies to acetylcholine receptors have been detected in these patients and have been shown to decrease gradually after drug withdrawal concomitant with reversibility of the clinical syndrome. However, myasthenia gravis may persist for long periods of time after D-penicillamine therapy has ceased.

Although rare, cases of renal lupus syndrome and pemphigus blisters have also been reported as a consequence of D-penicillamine-induced immune complexes (Ntoso et al., 1986; Bigazzi, 1988) as well as with other drugs. With renal lupus syndrome, secondary glomerulonephritis may result if granular IgG antibodies are produced and deposited on the basement membranes. In patients with pemphigus blisters, autoantibodies to the intercellular substance of the skin have been recovered from the sera, and dermal biopsies have demonstrated intracellular deposits or immunoglobulin deposits on the basement membranes. Pemphigus has also been observed in patients treated with sulfhydryl compounds such as captopril and pyrithioxine (Bigazzi, 1988).

Some metals that are used therapeutically have also been shown to induce autoimmune responses. Gold salts used to treat arthritis may induce formation of antiglomerular basement membrane antibodies, which may lead to glomerulonephritis similar to that seen in Goodpasture's disease (see type II hypersensitivity). Since gold is not observed at the site of the lesions (Druet et al., 1982), it has been hypothesized that the metal elicits an antiself response. Lithium, used to treat manic depression, is thought to induce autoantibodies against thyroglobulin, which in some patients results in hypothyroidism. In studies with rats, levels of antibodies to thyroglobulin were shown to increase significantly in lithium-treated rats compared to controls immediately after immunization with thyroglobulin; however, rats that were not immunized with thyroglobulin did not produce circulation antithyroglobulin antibodies upon receiving lithium, and there was no effect of lithium on lymphocytic infiltration of the thyroid in either group (Hassman et al., 1985).

Some drugs such as penicillin have been shown to induce autoimmunity as well as anaphylaxis (Gleichman et al., 1989). The carbonyl of the β-lactam ring of penicillin can form a covalent penicilloyl conjugate with nucleophilic sites on proteins, particularly the amino groups of lysine residues. This conjugate, which acts as the major immunogenic determinant, may become biotransformed to other isomeric forms of clinical relevance (Batchelor et al., 1965).

A genetic predisposition to drug-induced development of SLE has been shown to occur in some individuals treated with the drugs hydralazine, isoniazid, procainamide, and sulphamethazine. A polymorphism, which is known to exist for the genes responsible for expression of hepatic *N*-acetyl transferase enzymes, determines the rate of acetylation of these drugs to regulate the rate of drug inactivation. Individuals that are relatively slow acetylators of these drugs are more likely to develop antinuclear antibodies and are at a higher risk for developing SLE (Perry et al., 1970). Other predisposing factors, such as HLA phenotype (HLA-DR4 and/or C4 allele), may also play a genetic role in determining susceptibility to hydralazine-induced SLE (Spears and Batchelor, 1987).

In addition, silicone-containing medical devices, particularly breast prostheses, have been reported to cause serum-sickness-like reactions, scleroderma-like lesions, and an SLE-like disease termed human adjuvant disease (Kumagai et al., 1984; Guillaume et al., 1984). Some patients may also present with granulomas and autoantibodies. Human adjuvant disease is a connective tissue or autoimmune disease similar to that of adjuvant arthritis in rats and rheumatoid arthritis in humans. Autoimmune-disease-like symptoms usually develop two to five years after implantation in a small percentage of people that receive implants, which may indicate that there is a genetic predisposition similar to that for SLE. An early hypothesis is that the prosthesis or injected silicone plays an adjuvant role by enhancing the immune response through increased macrophage and T-cell helper function. There is currently controversy as to whether silicone, as a foreign body, induces a nonspecific inflammation reaction, a specific cell-mediated immunological reaction, or no reaction at all. However, there is strong support to indicate that silicone microparticles can act as haptens to produce a delayed hypersensitivity reaction in a genetically susceptible population of people. It should be noted that there are currently no known drug-induced type I autoimmunities.

## 7 EVALUATION OF IMMUNE SYSTEM

The ICH/FDA CDER guidelines for immunotoxicity testing of small-molecule pharmaceuticals (ICH S8) start with evaluation of parameters evaluated in repeat-dose (typically 28-day) systemic toxicity studies. There tests can generally be derived from the routine measurements and examinations performed in short-term and subchronic rodent and nonrodent toxicity

studies, since they do not require any perturbation of the test animals (immunization or challenge with infectious agents). These measurements include hematology and serum chemistry profiles, routine histopathological examinations of immune-associated organs and tissues, and organ and body weight measurements including thymus and spleen. If a compound produces any primary indicators of immunotoxicity from these measurements, more definitive immunotoxicity tests, such as those indicated in the preceding paragraph, may be recommended on a case-by-case basis.

The following is a brief explanation of some of the indicators that may be used to trigger additional definitive testing and a description of some of the most commonly used assays to assess humoral, cell-mediated, or nonspecific immune dysfunction, which are common to most immunotoxicology test strategies.

### 7.1 Immunopathological Assessments

Various general toxicological and histopathological evaluations of the immune system can be made as part of routine preclinical safety testing to obtain a preliminary assessment of potential drug-related effects on the immune system. At necropsy, various immunological organs of the immune system such as thymus, spleen, and lymph nodes are typically observed for gross abnormalities and weighed in order to detect decreased or increased cellularity. Bone marrow and peripheral blood samples are also taken to evaluate abnormal types and/or frequencies of the various cellular components. Tables 12 and 13 summarize the observations and interpretations.

***Organ and Body Weights*** Changes in absolute weight, organ-to-body weight ratios, and organ-to-brain weight ratios of tissues such as thymus and spleen are useful general indicators of potential immunotoxicity. However, these measures are nonspecific for immunotoxicity since they may also reflect general toxicity and effects on endocrine function that can indirectly affect the immune system.

### 7.2 Humoral Immune Response and Possible Entry Points for Immunotoxic Actions

***Hematology*** Hemacytometers or electronic cell counters can be used to assess the numbers of lymphocytes, neutrophils, monocytes, basophils, and eosinophils in the peripheral blood, while changes in relative ratios of the various cell types can be assessed by microscopic differential evaluation. Similar evaluations can be performed with bone marrow aspirates, where changes may reflect immunotoxicity to the pluripotent stem cells and newly developing lymphoid precursor cells. Potential hematological indicators of immunotoxicity include altered white blood cell counts or differential ratios, lymphocytosis, lymphopenia, or eosinophilia. Changes in any of these

**TABLE 12 Examples of Antemortem and Postmorten Findings That May Include Potential Immunotoxicity If Treatment Related**

| Parameter | Possible Observation (Cause) | Possible State of Immune Competence |
|---|---|---|
| | *Antemortem* | |
| Mortality | Increased (infection) | Depressed |
| Body weight | Decreased (infection) | Depressed |
| Clinical signs | Rales, nasal discharge (respiratory infection) | Depressed |
| | Swollen cervical area (sialodacryoadenitis virus) | Depressed |
| Physical examinations | Enlarged tonsils (infection) | Depressed |
| Hematology | Leukopenia/lymphopenia | Depressed |
| | Leukocytosis (infection/cancer) | Enhanced/depressed |
| | Thrombocytopenia | Hypersensitivity |
| | Neutropenia | Hypersensitivity |
| Protein electrophoresis | Hypogammaglobulinemia | Depressed |
| | Hypergammaglobulinemia (ongoing immune response or infection) | Enhanced/activated |
| | *Postmortem* | |
| Organ weights | | |
| Thymus | Decreased | Depressed |
| Histopathology | | |
| Adrenal glands | Cortical hypertrophy (stress) | Depressed (secondary) |
| Bone marrow | Hypoplasia | Depressed |
| Kidney | Amyloidosis | Autoimmunity |
| | Glomerulonephritis (immune complex) | |
| Lung | Pneumonitis (infection) | Depressed |
| Lymph node | Atrophy | Depressed |
| Spleen | Hypertrophy/hyperplasia | Enhanced/activated |
| | Depletion of follicles | Depressed B cells |
| | Hypocellularity of periarteriolar sheath | Depressed T cells |
| | Active germinal centers | Enhanced/activated |
| Thymus | Atrophy | Depressed |
| Thyroid | Inflammation | Autoimmunity |

parameters can be followed up with more sophisticated flow cytometric analyses or immunostaining techniques that are useful for phenotyping the various types of lymphocytes (B cell, T cell) and the T-cell subsets ($CD4^+$ and $CD8^+$) on the basis of unique surface markers. Decreases or increases in the percentages of any of the cell populations relative to controls or in the ratios of B cells to T cells or $CD4^+$ to $CD8^+$ cells may be indicators of immunotoxicity.

***Clinical Chemistry*** Nonspecific clinical chemistry indicators of potential immune dysfunction include changes in serum protein levels in conjunction

with changes in the albumin-to-globulin (A/G) ratio. Immunoelectrophoretic analysis of serum proteins can then be performed to quantify the relative percentages of albumin and the α-, β-, and γ-globulin fractions. To perform these assays, a drop of serum (antigen) is placed into a well cut in a gel, then the gel is subjected to electrophoresis so that each molecule in the serum moves in the electric field according to its charge. This separation is then exposed to specific antiserum, which is placed in a trough cut parallel to the direction in which the components have moved. By passive diffusion, the antibody reaches the electrophoretically separated antigen and reacts to form Ag–Ab complexes. The γ-globulin fractions can be separated and further quantified for the relative proportions of IgG, IgM, IgA, and IgE using similar techniques.

Serum concentrations of immunoglobulin classes and subclasses can also be measured using various techniques such as radioimmunoassays (RIAs) or enzyme-linked immunosorbent assays (ELISAs). In the ELISA, antigens specific for each class of immunoglobulin can be adsorbed onto the surfaces for microtiter plates. To determine the quantity of each antibody in a test sample, an aliquot of antiserum is allowed to react with the adsorbed antigens. Unreacted molecules are rinsed off and an enzyme-linked anti-Ig is then added to each well. Next, substrate is added and the amount of color that develops is quantified using a spectrophotometric device. The amount of antibody can then be extrapolated from standard curves since the amount of color is proportional to the amount of enzyme-linked antibody that reacts. Variations in levels of a given antibody may indicate the decreased ability of B cells of decreased numbers of B cells producing that antibody. In addition, serum autoantibodies to DNA, mitochondria, and parietal cells can be used to assess autoimmunity. Serum cytokines (IL-1, IL-2, and γ-interferon) can also be evaluated using immunochemical assays to evaluate macrophage, lymphocyte, and lymphokine activity; prostaglandin $E_2$ can also be measured to evaluate macrophage function.

CH50 determinations can be used to analyze the total serum complement and are useful for monitoring immune complex diseases (Sullivan, 1989); activation of complement (Table 13) in the presence of autoantibodies is indicative of immune complex diseases and autoimmunity. The various components of the complement system (C3, C4) can also be measured to assess the integrity of the system. For instance, low serum concentrations of C3 and C4 with a concomitant decrease in CH50 may indicate activation of complement, while a low C4 alone is a sensitive indicator of reduced activation of the complement system. Since C3 is used as an alternate complement pathway, it usually measures high. Therefore, a low C3 with a normal C4 may indicate an alternate pathway of activation.

***Histopathology*** Histopathological abnormalities can be found in lymphoid tissues during gross and routine microscopic evaluations of the spleen, lymph nodes, thymus, bone marrow, and gut-associated lymphoid tissues such as Peyer's patches and mesenteric lymph nodes. Microscopic evaluations should

**TABLE 13 Immune System Components in Organ Sites**

| Organ | Antigen-Presenting Cell | B Cell | CD4* T Cell | Possible Immunotoxic Entry Point | Outcome |
|---|---|---|---|---|---|
| Bone marrow | Differentiation from pluripotent stem cells | Differentiation from pluripotent stem cells; rearrangement of antigen receptors, central repertoire selection, homing, antibody production | Differentiation from pluripotent stem cells; rearrangement of antigen receptors | Inhibition of cell proliferation, interference with signal transduction, genotoxicity, interference with gene recombination, interference with DNA repair, interference with signal transduction, interference with apoptosis, interference with signal transduction, inhibition of protein synthesis | Leukopenia, malignancies, lymphopenia, selective lymphopenia, autoimmunity, impaired adaptive immune response |
| Thymus | — | — | Proliferation and positive/negative repertoire selection | Interference with signal transduction, interference with apoptosis | Selective lymphopenia, autoimmunity |
| Peripheral tissue | Antigen uptake and processing, maturation | — | — | Inhibition of cell motility, interference with phagolysosomal degradation, interference with intracellular transport, inhibition of cell motility, interference with signal transduction | Lack of antigen presentation, impaired adaptive immune response, lack of antigen presentation, impaired adaptive immune response |
| Lymphatic organ | Antigen presentation | Antigen uptake and presentation, cross-comunication with T cell, activation, proliferation, differentiation | Antigen recognition (antigen-presenting cell) and activation, proliferation, differentiation, cross-comunication with B cell | Inhibition of cell motility, interference with phagolysosomal degradation or intracellular transport, interference with signal transduction, inhibition of cytokine release, interference with signal transduction, inhibition of cell proliferation, interference with signal transduction | Impaired adaptive immune response, impaired adaptive immune response, autoimmunity, allergy, impaired adaptive immune response |

Adapted from Haley et al., 2005.

include descriptive qualitative changes such as types of cells, densities of cell populations, proliferation in known T- and B-cell areas (e.g., germinal centers), relative numbers of follicles and germinal centers (immune activation), and the appearance of atrophy or necrosis. In addition, unusual findings such as granulomas and scattered, focal mononuclear cell infiltrates in nonlymphoid tissues may be observed as indicators of chronic hypersensitivity or autoimmunity. A complete histopathological evaluation should also include a quantitative assessment of cellularity through direct counts of each cell type in the various lymphoid tissues. In addition, changes in cellularity of the spleen can be more precisely quantitated from routine H&E (hematoxylin and eosin) sections using morphometric analysis of the germinal centers (B cells) and periarteriolar lymphocyte sheath (T cells). Similar morphometric measurements can be made of the relative areas of the cortex and medulla of the thymus. If changes in cellularity are apparent from routinely stained histopathology sections, special immunostaining (immunoperoxidase or immunofluorescence) of B cells in the spleen and lymph nodes using polyclonal antibodies to IgG, or immunostaining of the T cells and their subsets in the spleen using mono- or polyclonal antibodies to their specific surface markers, can be used to further characterize changes in cellularity.

Numerous physiological and environmental factors such as age, stress, nutritional deficiency, and infections may affect the immune system (Sullivan, 1989). Thus, adverse findings in animal studies may reflect these indirect immunotoxic effects rather than the direct immunotoxic potential of a chemical or drug. Indirect immunotoxic effects may be assessed through histopathological evaluations of endocrine organs such as the adrenals and pituitary.

It is also well known that the functional reserves of the immune system can allow biologically significant, immunotoxic insults to occur without the appearance of morphological changes. In addition, there is some built-in redundancy in the system in that several mechanisms may produce the same outcome. For instance, cytotoxic T cells may alone be sufficient to protect the organism against a bacterial infection; however, the body will also produce antibodies for future protection. Thus, if one mechanism is insufficient to fight off infection, the second mechanism can serve as a backup. Because of this functional reserve, adverse effects may remain subclinical until the organism is subjected to undue stress or subsequent challenge (Bloom et al., 1987). Therefore, routine immunopathological assessments as part of standard preclinical toxicity tests may not be sufficient to detect all immunotoxins. Although changes detected in routine toxicological and pathological evaluations are nonspecific and of undetermined biological significance to the test animal, they can be invaluable as flags for triggering additional testing.

### 7.3 Humoral Immunity

As described previously, the humoral immune response results in the proliferation, activation, and subsequent production of antibodies by B cells

following antigenic exposure and stimulation. The functionality and interplay between the three primary types of immune cells (macrophage, B cells, and T cells) required to elicit a humoral response can be assessed through various in vitro assays using cells from the peripheral blood or lymphoid tissues. (Burleson et al., 1995).

***Antibody Plaque-Forming Cell (PFC) Assay*** The number of B cells producing antibody (PFC) to a T-dependent antigen such as sheep red blood cells (SRBCs) can be assessed in vitro following in vivo exposure to the test article and antigen (ex vivo tests). The PFC response to a T-dependent antigen is included as a tier I test by the National Toxicology Program (NTP) since it appears to be the most commonly affected functional parameter of exposure to immunosuppressants. However, this test is designated as a type II test in the FDA Redbook since it requires an in vivo immunization of the animals with antigen and thus cannot be evaluated as part of an initial toxicity screen.

Although this assay requires that B cells be fully competent in secreting antibodies, T cells and macrophage cells are also essential for the proper functioning of humoral immunity. However, this assay is nonspecific in that it cannot determine which cell type(s) is responsible for dysfunction. Macrophage cells are needed to process antigen and produce IL-1. T cells are needed for several functions including antigen recognition of surface membrane proteins and B-cell maturation through the production of various lymphokines that stimulate growth and differentiation. SRBCs are most commonly used as the T-dependent antigen, although T-cell-independent antigens may also be useful to rule out T helper dysfunction as a cause of immunodysfunction.

The PFC assay has evolved from methodology originally developed as a hemolytic plaque assay (or Jerne plaque assay) by Nils Jerne to quantitate the number of antibody-forming cells in a cell suspension plated with red blood cells (RBCs) onto agar plates (Jerne and Nordin, 1963). In its present form, animals are treated in vivo with the test compound, immunized with approximately $5 \times 10^8$ SRBCs administered intravenously within two to three days posttreatment, and then sacrificed four days (IgM) or six days (IgG) later. Antibody-producing spleen cell suspensions are then mixed in vitro with SRBCs, placed onto covered slides, and incubated for a few hours in the presence of complement. During incubation, antibody diffuses from the anti-SRBC-producing cells and forms Ag–Ab complexes on the surfaces of nearby SRBCs. In the presence of complement, the Ag–Ab complexes cause lysis of the SRBCs, resulting in the formation of small, clear plaques on the slide. Plaques are then counted and expressed as PFCs/$10^6$ spleen cells. A dose-related reduction in PFCs is indicative of immunosuppression.

***B-Cell Lymphoproliferation Response*** The NTP has classified this assay as a tier I test since mitogenesis can be performed easily in tandem with other tests to provide an assessment of the proliferative capacity of the cells (Luster et al., 1988). Since this assay is performed ex vivo with peripheral blood (or

spleen) and is well characterized for use in various animal species, it has also been included as an expanded type I test in the revised Redbook.

The proliferation of peripheral blood or splenic B cells following stimulation with lipopolysaccharide (LPS) or other mitogens (pokeweed mitogen extract) is another measurement of humoral immunity. LPS (a bacterial lipopolysaccharide) is a B-cell-specific mitogen that stimulates polyclonal proliferation (mitosis) as part of the natural sequence of antigen recognition, activation, and clonal expansion. The mitogen interacts not with just one particular antigen-specific clone but with all cells bearing the carbohydrate surface marker for which it is specific. Since mitogens are both polyclonal and polyfunctional, they can stimulate a wider spectrum of antigenic determinants than antigens, which can only stimulate a low number ($10^{-6}$) of specific cells.

In this assay, lymphocytes from animals are treated in vivo and cultured in vitro in microtiter plates in the presence of tritiated [$^{3}$H]thymidine (or uridine) using a range of at least three concentrations of mitogen to optimize the response. Lymphocytes can be obtained aseptically from peripheral blood or from single-cell suspensions of spleen cells that are prepared by pushing the tissue through sterile gauze or 60-mesh wire screens. A decrease in DNA synthesis (incorporation of $^{3}$H) as compared to the unexposed cells of control animals may indicate that the B cells were unable to respond to antigenic stimulation. Alternative methodology employs 18–20 h incubation with $^{125}$I-labeled iododeoxyuridine ([$^{125}$I]IudR) and fluorodeoxyuridine (FudR) (White et al., 1985). After incubation, the cells are collected onto filter disks and then counted with a gamma counter.

Assays such as this that use polyclonal mitogens for activation may not be as sensitive as specific antigen-driven systems (Luster et al., 1988). In addition, suppression of the mitogen response does not always correlate with the PFC response. Since mitogenesis represents only a small aspect of B-cell function and maturation, this endpoint is not sensitive to early events that may affect activation or later events that may affect differentiation of B cells into antibody-secreting cells (Klaus and Hawrylowicz, 1984).

### 7.4 Cell-Mediated Immunity

***T-Cell Lymphoproliferation Response*** This assay is analogous to the B-cell lymphoproliferative response assay described above. Thus, this assay is also classified as a tier I test by the NTP and as an expanded type I test in the revised draft of the Redbook.

T cells from the peripheral blood or spleen undergo blastogenesis and proliferation in response to specific antigens that evoke a cell-mediated immune response. T-cell proliferation is assessed using T-cell-specific mitogens such as the plant lectins, concanavalin A (Con A), and phytohemagglutinin (PHA) or T-cell-specific antigens (i.e., tuberin, *Listeria*). Uptake of $^{3}$H as an indicator of DNA synthesis is used as described above for evaluating B-cell

proliferation. T-cell mitogens stimulate not only the synthesis of DNA but also the expression of cell-specific function. For instance, Con A can trigger the expression of T helper, suppressor, and cytotoxic effector cells, and either mitogen may induce the expression (or reexpression of memory cells) of differentiated function (Clark, 1983). Since cell populations responsive to Con A are thought to be relatively immature compared to those that are stimulated with PHA, the parallel usage of both mitogens may be useful for distinguishing the affected subset (Tabo and Paul, 1973). A secondary response to T-cell antigens such as purified protein derivative of tuberculin (PPD) or tetanus toxoid can also be assessed.

***Mixed-Lymphocyte Response (MLR) Assay*** This assay has been shown to be sensitive for the detection of chemical-induced immunosuppression and is a recommended tier I assay by the NTP (Luster et al., 1988). In addition, it has been shown to be predictive of host response to transplantation and of general immunocompetence (Harmon et al., 1982).

The MLR assay assesses the ability of T cells to recognize foreign antigens on allogenic lymphocytes and, thus, is an indirect measure of the cell-mediated ability to recognize graft or tumor cells as foreign. Responder lymphocytes from animals treated in vivo with the test compound are mixed with allogenic stimulator lymphocytes that have been treated in vitro with mitomycin C or irradiated to render them unable to respond (Bach and Voynow, 1966). Both cell types are cultured in vitro for three to five days, then incubated with $^{3}H$ for an additional 6 h. Once the radiolabel is incorporated into the DNA of the responding cells, the DNA is extracted and the amount of radioactive label is measured to quantitate proliferation of the responder cells of drug-treated animals compared to those of the controls.

***Cytotoxic T-Lymphocyte (CTL)–Mediated Assay*** This assay is similar to the MLR assay and can be performed in parallel or as a tier II follow-up to the MLR assay.

The CTL assay ascertains the ability of cytotoxic T cells to lyse an allogenic target cell or the specific target cell type with which they were immunized. In general, the cytolytic response of activated effector cells is assessed by measuring the amount of radioactivity ($^{51}Cr$) that is released from the target cell. When performed in conjunction with the MLR assay, lymphoid cells of the two strains are cultured together in vitro as described above; however, $^{51}Cr$ is added to the culture after four to five days (instead of $^{3}H$). Both responder and target cells are labeled with the $^{51}Cr$, which is taken up rapidly by the cells through passive diffusion but is released slowly as long as the cell membrane is intact. Furthermore, since chromium is reduced from $Cr^{6+}$ to $Cr^{3+}$, and since $Cr^{3+}$ enters the cells at a much slower rate than $Cr^{6+}$, the $^{51}Cr$ released from the damaged target cells is not significantly reincorporated into undamaged cells (Clark, 1983), which would reduce the sensitivity of the assay. Thus, the amount of chromium released into the medium and recovered in the super-

natant of the mixture of the cells is directly proportionate to the extent of lysis of the target cells by the sensitized responder cells.

In a capillary tube assay developed in 1962 by George and Vaughan, the inhibition of migration of macrophage cells can be used to access normal T-cell function (see Clark, 1983). T cells are obtained from the peripheral blood of animals treated in vivo with a test article and injected with antigen (e.g., tuberculin). These T cells are functioning normally and should release migration inhibition factor (MIF). As a consequence, the macrophages, which generally show a propensity for migration upon stimulation with the antigen, should show a MIF-induced reduction in migratory behavior.

***Delayed-Type Hypersensitivity Response*** The DTH response assay is considered to be a comprehensive tier II assay for cell-mediated immunity by the NTP.

To express a DTH inflammatory response, the immune system must be capable of recognizing and processing antigen, blastogenesis and proliferation of T cells, migration of memory T cells to the challenge site of exposure to antigen, and subsequent production of inflammatory mediators and lymphokines that elicit the inflammatory response. Thus, by measuring a DTH response to an antigen, these assays assess the functional status of both the afferent (antigen recognition and processing) and efferent (lymphokine production) arms of cellular immunity. Various antigens have been used for assessing DTH, including keyhole limpet hemocyanin (KLH), oxazolone, dinitrochlorobenzene, and SRBCs (Vos, 1977; Godfrey and Gell, 1978; Luster et al., 1988).

In one such assay described by White et al. (1985), mice previously treated with the test article are sensitized to SRBCs by inoculation of SRBCs into the hind footpad and four days later are challenged in the same footpad. Seventeen hours following challenge, they are injected intravenously with $^{125}$I-labeled human serum albumin (HSA), then sacrificed 2 h later. Both hind feet are then radioassayed in a gamma counter (the second foot serves as a control for background infiltration of the label). With a normal functioning cell-mediated response, $^{125}$I-labeled HSA will extravasate into the edematous area produced by the DTH response (Paranjpe and Boone, 1972). In general, a decrease in the extravasation of $^{125}$I-labeled HSA is indicative of immunosuppression of the efferent arm of the cell-mediated immune system.

To assay specific functionality of the afferent arm of the DTH response, proliferation of the popliteal lymph node cells to SRBCs can also be measured (White et al., 1985). As described above, mice treated with the test article are sensitized to SRBCs by inoculation of SRBCs into the hind footpad. However, 1.5 h later they are challenged intraperitoneally with FUdR and 2 h later they are administered [$^{125}$I]IUdR intravenously (instead of $^{125}$I-labeled HSA). Mice are sacrificed 24 h after challenge and both popliteal lymph nodes are removed and counted in a gamma counter.

Similar assays for DTH have been traditionally performed with the antigen *Mycobacterium tuberculosis*, which preferentially elicits a cell-mediated

response. In this assay a small amount of antigen contained in the supernatant fluid from the medium in which the pathogen was grown is injected into the footpad. Upon challenge, a visible and palpable lump should appear by 48 h. The amount of swelling is then measured and compared with the footpad that did not receive the challenge. Alternatively, methods used by the NTP employ a modified $^{125}$I-labeled uridine (UdR) technique to measure the monocyte influx at the challenge site (ear) injected with KLH antigen. This assay has been shown to correlate well with decreased resistance to infectious disease (Luster et al., 1988). However, one should note that regardless of which technique is used, anti-inflammatory drugs may produce false-positive results in this type of assay.

### 7.5 Nonspecific Immunity Function Assay

***Natural Killer Cell Assays*** This assay is a tier I test for nonspecific immunity in the NTP testing scheme (Luster et al., 1988) and is proposed as an additional type I test in the draft Redbook.

Natural killer cells, like cytotoxic T cells, have the ability to attack and destroy tumor cells or virus-infected cells. However, unlike T cells, they are not antigen specific, do not have unique, conally distributed receptors, and do not undergo clonal selection. In in vitro or ex vivo tests, target cells (e.g., YAC-1 tumor cells) are radiolabeled in vitro or in vivo with $^{51}$Cr and incubated in vitro with effector NK cells from the spleens of animals that had been treated with a xenobiotic. This assay can be run in microtiter plates over a range of various ratios of effector/target cells. Cytotoxic activity is then measured by the amount of radioactivity released from the damaged tumor cells, as was previously described for cytotoxic T cells. This assay can also be performed in vivo, where YAC-1 cells labeled with [$^{125}$I]IUdR are injected directly into mice and NK-cell activity is correlated with its level of radioactivity (Riccardi et al., 1979). Immunotoxicity observed as reduced NK-cell activity is correlated with increased tumorigenesis and infectivity.

***Macrophage Function*** Several assays are available to measure various aspects of macrophage function, including quantitation of resident peritoneal cells, antigen presentation, cytokine production, phagocytosis, intracellular production of oxygen free radicals (used to kill foreign bodies), and direct tumor-killing potential. Techniques for quantitation of peritoneal cells and functional assays for phagocytic ability are classified as comprehensive tier II tests by the NTP and as additional type I tests in the draft Redbook.

Macrophage cells and other PMNs contribute to the first-line defense of nonspecific immunity through their ability to phagocytize foreign materials, including pathogens, tumor cells, and fibers (e.g., silica, asbestos). Xenobiotics can affect macrophage function by direct toxicity to macrophages or by modulating their ability to become activated. Differential counts of resident perito-

neal cells can be made as a rapid, preliminary assessment of macrophage function for xenobiotics that are not administered parenterally.

Numerous in vitro assays can be employed to assess common functions of macrophages and PMNs, including adherence to glass, migration inhibition, phagocytosis, respiratory activity (chemiluminescent assays or nitroblue tetrazolium), and target cell killing. In one such assay, the chemotactic response to soluble attractants is evaluated using a Boyden chamber with two compartments that are separated by a filter. Macrophage cells or PMNs from treated animals are placed in one side and a chemotactic agent in the other. Chemotaxis is then quantified by counting the number of cells that pass through the filter. In another assay, the ability of the macrophages to phagocytize foreign materials can be evaluated by adding fluorescent latex beads to cultures containing macrophage cells, then determining the proportion of cells that have phagocytized the beads using a fluorescent microscope or by flow cytometry (Duke et al., 1985). Similar functions can be evaluated by incubating the cells with known amounts of bacteria. The cells are then removed by filtration or centrifugation, the remaining fluid is plated onto bacterial nutrient agar, and, after a few days of incubation, the bacterial colonies are counted. Furthermore, the efficiency of the cells to kill the bacteria once phagocytized can be assayed by lysing the cells and plating the lysate onto bacterial agar.

Various in vivo assessments of macrophage function have also been used. For example, peritoneal exudate cell (PEC) recruitment can be assessed using eliciting agents such as *Corynebacterium parvum*, MVE-2, or thioglycolate (Dean et al., 1984). In one such assay (White et al., 1985), mice are injected intraperitoneally with thioglycolate and sacrificed five days later, and the peritoneal cavity is flushed with culture medium. The cell suspension is then counted, the cell concentration is adjusted to a known density ($2 \times 10^5\,mL^{-1}$), and the cells are cultured for 1 h in 24-well culture dishes. Adherent cells are then washed with medium and aliquots of $^{51}Cr$-labeled SRBCs that were opsonized with mouse IgG are added to each well and incubated for various times. This same system can be used to assess adherence and chemotaxis of the PECs (Laskin et al., 1981). Phagocytosis can also be evaluated in vivo by measuring the clearance of injected particles from the circulation and the accumulation of the particles in lymphatic tissues such as the spleen.

***Mast Cell/Basophil Function*** The function of mast cells and basophils to degranulate can be evaluated using a passive cutaneous anaphylaxis test (Cromwell et al., 1986). Serum containing specific anaphylactic (IgE) antibodies from donor animals previously exposed to a known antigen is first administered by intradermal (or subcutaneous) injection into unexposed host animals. After a sufficient latency period to allow binding of the donor IgE to the host tissue mast cells, the animals are administered a second intravenous injection of the antigen. The anaphylactic antibodies present in the serum will stimulate normally functioning mast cells to degranulate (release histamines) and produce a marked inflammatory response. Using similar in vitro assays

with mast cells and basophils, the quantities of histamines that are released from the cells can be measured directly in the culture medium.

## 7.6 Host Resistance Assays

Host resistance assays can be used to assess the overall immunocompetence of the humoral or cell-mediated immune systems of the test animal (host) to fend off infection with pathogenic microbes or to resist tumorigenesis and metastasis. These assays are performed entirely in vivo and are dependent on all of the various components of the immune system to be functioning properly. Thus, these assays may be considered to be more biologically relevant than in vitro tests that only assess the function of cells from one source and of one type. Since these assays require that the animal be inoculated with a pathogen or exogenous tumor cell, they cannot be performed as part of a general preclinical toxicity assessment and are thus classified as type II tests in the revised Redbook. These assays are also included as tier II tests by the NTP.

Several host resistance assays have been developed using various infectious agents, including bacteria (*Listeria monocytogenes*, *Streptococcus*, and *Escherichia coli*), viruses (influenza, cytomegalovirus, and herpes), yeast (*Candida albicans*), and parasites (*Trichinella spiralis* and *Plasmodium berghei*). These assays have been described in the NTP guidelines (Luster et al., 1988). In general, animals previously treated with a xenobiotic are injected with the pathogen at a target dose that is estimated to kill 10–30% of control animals (LD10–30). After a period of time, the animals receive a challenge dose at a much higher concentration (LD60–80) and by a different route to determine if animals are resistant to reinfection. Although these assays are similar in their mechanisms of resistance to different pathogens, they have been shown to differ with regard to varying degrees of susceptibility by the same drug (Morahan et al., 1979). Thus, for screening purposes, it is recommended that at least two tests be used (Descotes and Mazue, 1987). Although these tests are relatively easy to perform, those involving the use of pathogens require special handling, containment, and decontamination procedures to prevent infection to humans and spread throughout the animal colony.

Similar host resistance assays are used to evaluate the immunosurveillance of spontaneous tumors, which is assessed as the capacity of the organism to reject grafted syngeneic tumors. Various animal-bearing tumor models (Pastan et al., 1986) and host resistance models have been used to assess immunotoxicity. Several of the host resistance assays utilize cultured tumor cell lines such as PYB6 sarcoma and B16F10 melanoma cells that are used with C57/BL/6 mice or the MADB106 lung tumor cell lines that are used with Fischer 344 rats. For example, the PYB6 sarcoma model uses death as an endpoint. In this assay, syngeneic mice are injected with the PYB6 sarcoma cells and death due to tumor is recorded daily. In another routinely used assay, animals that have been treated with a xenobiotic are injected with either B16F10 melanoma cells

or Lewis lung carcinoma cells; then approximately 20 days later they are sacrificed and pulmonary tumors are measured and counted.

### 7.7 T-Cell-Dependent Antibody Response (TDAR)

The TDAR should be performed using a recognized T-cell-dependent antigen like SRBCs, BSA, or KLH that results in a robust antibody response. For the SRBC assay, IgM measurement is considered the most appropriate endpoint, whereas IgG measurement is considered to be most appropriate for BSA or KLH. Antibody can be measured by using an ELISA or other immunoassay method. One advantage of ELISA over the traditional PFC assay (Ladics, 2005) is that samples can be collected serially during the study, if necessary. Since immunization is likely to have effects on the hematology, clinical chemistry, and histology of lymphatic organs, TDAR studies should always be performed as separate studies, or at least in satellite groups of repeat-dose toxicity studies.

Brief details of the assay are as follows:

**Treatment**

1. Use a suitable SPF mouse strain like BALB/c or C57BL/6 × C3H F1 (B6C3F1).
2. Allocate 120 animals in six groups of 10 males and 10 females each. The group sizes may be reduced to 5 males and 5 female animals per group when a substantial immunosuppressive effect can be expected. The larger group size should be chosen when immunosuppressives are excluded, since otherwise the statistical power of the assay might be insufficient to prove a lack of immunosuppression.
3. Allow acclimatization for 7 days before sampling of pretest serum (day –7) from the test groups, and allow an additional 7 days of rest before first dosing. Pretest serum of recovery groups may be taken on day 35, which is 1 week postdosing.
4. Administer the test substance and vehicle daily over a period of 28 days (day 1 to day 28) to all animals using an appropriate route of administration.
5. Use a low, intermediate, and high dose level, whereby the high dose level should be above the no-observable-effect level (NOEL) and below a dose level that causes stress, if possible. The intermediate (or low) dose level should ideally represent the intended clinical dose level.
6. Immunize all mice of the test groups on day 14 and mice of the recovery groups on day 42 by intraperitoneal injection of 100 μg KLH per mouse without the use of adjuvant.
7. Sample immune serum from all animals of the test groups on day 29 and from all of the recovery groups on day 57.
8. Store serum at –20 °C until ELISA testing.

## 7.8 Hypersensitivity

***Type I Hypersensitivity*** Although here are acceptable systems for evaluating type I (immediate) reactions following systemic exposure, there are no reliable animal models for predicting type I reactions following dermal applications or oral administrations of drug. Repeated exposure of a xenobiotic is required to produce a type I response. A drug in the form of a hapten must covalently bind to macromolecules (proteins, nucleic acids) before it can initiate a primary antibody response. Once sensitized, even the smallest exposure to the xenobiotic can elicit a rapid, intensive IgE antibody-mediated inflammatory response. With the exception of antivirals and chemotherapeutic drugs, most drugs should not be reactive with biological nucleophiles since these drugs are usually screened out as mutagens or carcinogens in preclinical safety studies. However, type I hypersensitivity is a particular problem with biotechnology products themselves (e.g., insulin, growth hormones, interleukins), trace impurities from the producing organisms (e.g., *E. coli* proteins, mycelium), or the vehicles used to form emulsions (Matori et al., 1985).

The production of neutralizing antibodies to recombinant DNA protein products or their contaminants may be assayed using ELISAs or RIAs. A suitable animal model used to evaluate the potential for a type I response to protein hydrolysates is detailed in the U.S. Pharmacopeia. This test is very sensitive for testing proteins administered by the parenteral route but is of little value for low-molecular-weight drugs and those that are administered orally (Descotes and Mazue, 1987). Active systemic anaphylaxis can be assessed in guinea pigs following systemic exposure to the test compound. For dermal exposures, however, rabbits or guinea pigs must be exposed to the test article by intradermal injections and then evaluated for their ability to mount a systemic anaphylactic response. The passive cutaneous anaphylaxis test (as described above for mast cells) can also be used to assess a potential anaphylactic response to a test compound. The serum containing potential anaphylactic (IgE) antibodies from donor animals previously exposed to the test compound is first administered by intradermal (or subcutaneous) injection into unexposed host animals. After a latency period, the animals are administered an intravenous injection of the test compound together with a dye. If anaphylactic antibodies are present in the serum, the subsequent exposure to the test compound will cause a release of vasoactive amines (degranulation of mast cells), ultimately resulting in the migration of the dye to the sites of the intradermal serum injections.

***Types II and III Hypersensitivity*** No simple animal modes are currently available to assess type II (antibody-mediated cytotoxicity) hypersensitivity reactions. IgE antibodies and immune complexes in the sera of exposed animals can be assayed using ELISA or RIA techniques that require the use of specific antibodies to the drug.

Type III (immune-complex-related disease) reactions have been demonstrated by the presence of proteinuria and immune complex deposits in the kidneys of the Brown-Norway, Lewis, and PVG/C rat strains. However, susceptibility to the deposition and the subsequent lesions (glomerulonephritis) are often variable and dependent on the strain (Bigazzi, 1985). For example, despite the appearance of clinical signs and proteinuria, after two months administration of mercuric chloride, detectable levels of circulating antinuclear autoantibodies can no longer be observed in the Brown-Norway strain (Bellon et al., 1982). By contrast, in PVG/C rats administered mercuric chloride, immune complex deposition and antinuclear autoantibodies are present for longer periods of time; however, proteinurea is not observed (Weening et al., 1978).

***Type IV Hypersensitivity*** There are several well-established preclinical models for assessing type IV (delayed-type) hypersensitivity reactions following dermal exposure, but not for predicting this response after systemic exposure.

Type IV hypersensitivity responses are elicited by T lymphocytes and are controlled by accessory cells and suppressor T cells. Macrophages are also involved in that they secrete several monokines, which results in proliferation and differentiation of T cells. Thus, there are numerous points along this intricate pathway in which drugs may modulate the final response. To achieve a type IV response, an initial high-dose exposure or repeated lower dose exposures are applied to the skin; the antigen is carried from the skin by Langerhans cells and presented to cells in the thymus to initiate T-cell proliferation and sensitization. Once sensitized, a second "challenge" dose will elicit an inflammatory response. Thus, before sensitivity can be assessed, each of the models used to evaluate dermal hypersensitivity requires as a minimum:

- Initial induction exposure
- Latency period for expression
- Challenge exposure

A preliminary test for acute irritancy is also required to ensure that the initial dose is sufficient to stimulate sensitization and that the challenge dose is sufficient to ensure expression of the response without producing irritation, which would confound the response. To confirm suspected sensitization or determine a threshold dose, each assay may also include a second challenge dose one to two weeks after the first challenge, at the same or lower concentrations. To increase penetration of the test article, various methods of abrasion (e.g., tape stripping) and occlusive coverings may also be used.

Several systems are used routinely to test compounds for dermal hypersensitivity. The two most commonly used, the modified Buehler test and the guinea pig maximization test (GPMT), are briefly reviewed. More detailed methodology and a description of alternative test systems can be found in Gad

and Chengelis (1998). Although either rabbits or guinea pigs are sensitive test species, guinea pigs have traditionally been the animal of choice. Guinea pig models of skin sensitization have been (and remain) widely used and have been valuable in assessing human risk (Andersen and Maibach, 1985).

***Modified Buehler*** Buehler (1964) developed the first test system to use an occlusive patch to maximize dermal exposure and to increase the test sensitivity (Buehler, 1964). Although this assay is still insensitive for some xenobiotics that may not sufficiently traverse the epidermis, it is particularly useful for compounds that are either highly irritating by intradermal injection or cannot be dissolved or suspended in a form that is conducive to injection. Other advantages are that the test produces few false positives, rarely overpredicts the potency of sensitizers, and is less likely to produce limiting system toxicity or ulceration at the induction sites. Figure 4 shows the test design in its current [Organisation for Economic Co-operation and Development (OECD)] form. The assay is no longer accepted in Europe due to a belief that it has an unacceptable rate of false-negative outcomes and is appropriate only for true topical (dermal) exposures.

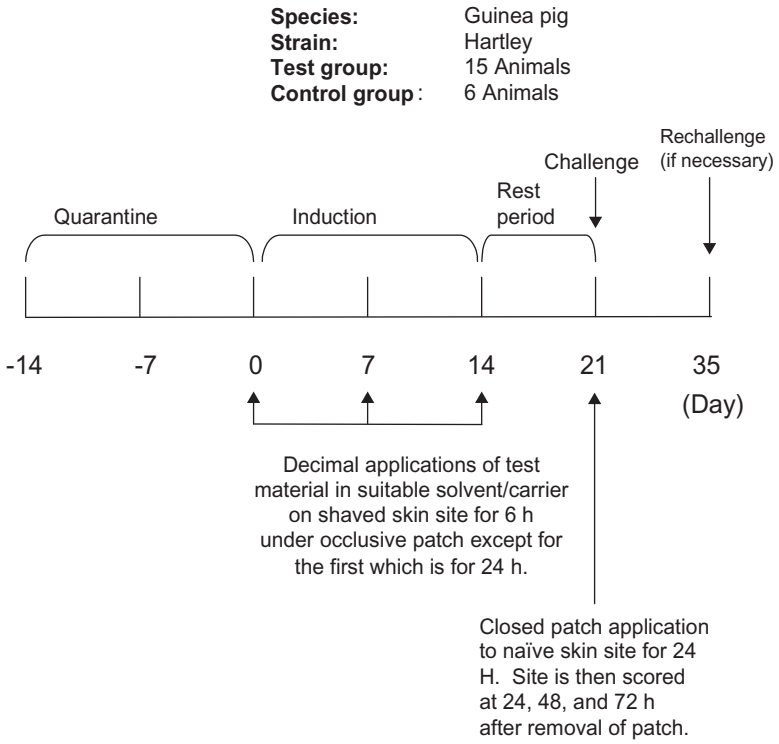


***Figure 4*** *Line chart for modified Buehler test for delayed contact dermal sensitization in guinea pig.*

During the induction phase, the test compound is applied to a cotton patch (1 in. × 1 in.) or placed in a Hill Top–style occlusive chamber. The patch is then placed onto a shaven area of epidermis on the left flank of a guinea pig and secured firmly in place for 24 h, after which time the patch is removed and the area is observed and scored for irritation (i.e., edema, erythema). A fresh patch is then reapplied for 6 h every other day during the induction period for a total of 10 treatments while continuing to score the application site at 24 and 48 h from the start of each treatment. Two weeks after the last induction exposure, the animals receive a challenge exposure for 24 h in the form of a patch applied to a shaven area of epidermis on the other flank (opposite the one used for induction). The challenge dose should be the highest concentration that does not produce dermal irritation after a single, 24-h exposure. The challenge site is observed for evidence of inflammation 24, 48, and 72 h after the patch is removed. Both the intensity and duration of a response to the test article compared to that of the vehicle are used to determine the potential and severity of sensitization.

***Guinea Pig Maximization Test*** This assay, as developed by Magnusson and Kligman (1969), differs from the Buehler test in that the compound is administered by intradermal injection during the first stage of induction and coadministered with an adjuvant (Freund's complete adjuvant) during the induction phase to further stimulate the immune system. This test system is more sensitive (fewer false negatives) than the Buehler test; however, it may overpredict the potency for many sensitizers. Figures 5 and 6 illustrate the study design.

Prior to induction, a 4 × 6-cm area of fur is clipped from the shoulder region of each guinea pig. On day 0, three pairs of intradermal injections are made along opposite sides of the dorsal midline of the animal. The first pair (closest to the head) is administered as the test substance in the vehicle, the second pair is administered proximal to the first pair and consists only of Freund's

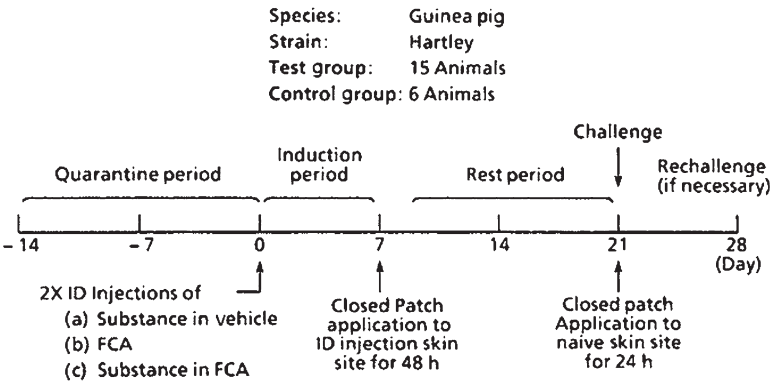


***Figure 5*** *Line chart for guinea pig maximization test for dermal sensitization.*

Outline of guinea pig maximization test

| State | Induction | | Challenge | Rechallenge |
|---|---|---|---|---|
| Day | 0 | 7 | 21 | 28 |
| Test group (15) | A. 0.1 mL substance ID<br>B. 0.1 mL FCA ID<br>C. 0.1 mL substance + FCA ID | Closed patch-48H application of substance | Closed patch-24H substance vehicle | Closed patch-24H vehicle |
| Test group (15) | A. 0.1 mL substance ID<br>B. 0.1 mL FCA ID<br>C. 0.1 mL vehicle + FCA ID | Closed patch-48H application of vehicle | Closed patch-24H substance vehicle | Closed patch-24H substance |

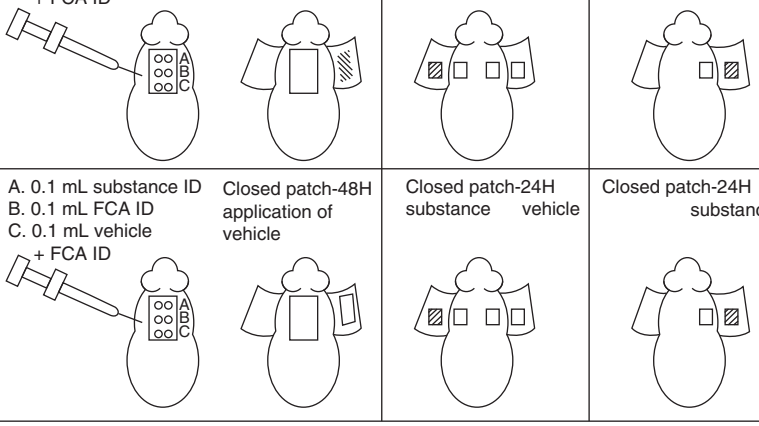


***Figure 6*** *Illustrative figures for injection and patching of animals in GPMT.*

complete adjuvant (FCA), and the third pair (spaced most posteriorly) is administered as the test substance in FCA. Seven days later (day 7), a mild to moderately irritating dose of the test article is spread onto a 1 × 2-in. filter paper patch, secured, and occluded for 48 h on the epidermal site that received the initial injections. On day 21, an area of fur on each flank is shaved and a 1 × 1-in. patch containing a nonirritating concentration of the test article is applied to one flank and a patch containing vehicle alone is applied to the other flank. The patches are secured and occluded for 24 h, and the challenge sites are scored for inflammation 24 and 48 h after removal of the patches. The incidence of animals that respond as well as the intensity and duration of a response to the test article are used to determine the potential and severity of sensitization.

## 7.9 Local Lymph Node Assay

This method has developed out of the work of Ian Kimber and associates (Kimber et al., 1986, 1991; Kimber and Weisenberger, 1989). It has the advantage over the other methods discussed in this chapter in that it provides an objective and quantifiable endpoint. The method is based on the fact that dermal sensitization requires the elicitation of an immune response. This immune response requires proliferation of a lymphocyte subpopulation. The local lymph node assay (LLNA) relies on the detection of increased DNA

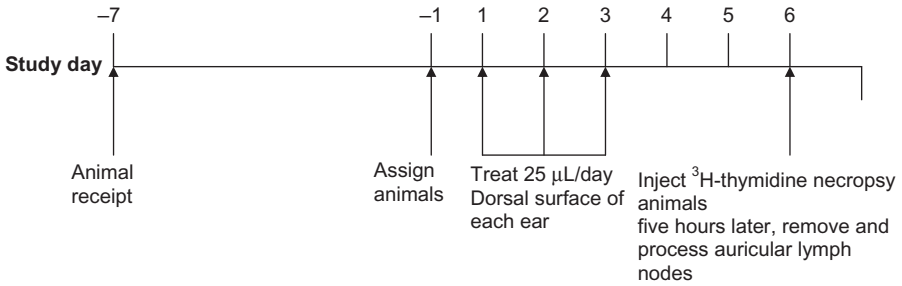


***Figure 7*** *Mouse LLNA.*

synthesis via tritiated thymidine incorporation. Sensitization is measured as a function of lymph node cell proliferative responses induced in a draining lymph node following repeated topical exposure of the test animal to the test article. Unlike the other tests discussed in this chapter, this assay looks only at the induction phase, as there is neither a challenge (elicitation) phase nor a sufficient period of evaluation for development of the underlying clonal expansion in response. Additionally, acute clinical formulations cannot be evaluated in this test system—meaning the all-important question of whether there is actually the potential for the drug to react with the immune system in clinical use remains unaddressed. Rather, in the case of a positive outcome, such interaction is presumed to occur.

The typical test (illustrated in Figure 7) is performed using mice—normally female CBA mice 6–10 weeks of age. Female BALB/c and ICR mice have also been used. After animal receipt, they are typically acclimated to standard laboratory husbandry conditions for 7–10 days. The usual protocol will consist of at least two groups (vehicle control and test article treated) of five mice each. They are treated on the dorsal surface of both ears with 25 μL (on each ear) of test article solution for three consecutive days. Twenty-four to 48 h after the last test article exposure, the animals are given a bolus (0.25-mL) dose of [$^3$H]thymidine (20 μCi with a specific activity of 5.0–7.0 Ci/mmol) in phosphate-buffered saline (PBS) via a tail vein. Five hours after the injection, the animals are euthanized by $CO_2$ asphyxiation and the auricular lymph nodes removed.

After removal, the lymph nodes can either be pooled by group or processed individually. Single-cell suspensions are prepared by gentle mechanical disaggregation through a nylon (100-μm) mesh. Cells are washed twice by centrifugation in an excess of PBS. After the final supernatant wash is removed, the cells are precipitated with cold 5% trichloroacetic acid (TCA) and kept at 4 °C for 12–18 h. The precipitate is then pelleted by centrifugation and resuspended in 1 mL 5% TCA, and the amount of radioactivity is determined by liquid scintillation counting using established techniques for tritium.

The data are reduced to the stimulation index (SI):

$$\mathrm{SI} = \frac{H(\mathrm{dpm})\text{treated group}}{H(\mathrm{dpm})\text{control group}}$$

An SI of 3 or greater is considered a positive response, that is, the data support the hypothesis that the test material is a sensitizer.

The test article concentration is normally the highest nonirritating concentration. Several concentrations could be tested at the same time should one wish to establish a dose–response curve for induction. The test is easiest to perform if the vehicle is a standard nonirritating organic, such as acetone, ethanol, or dimethylformamide, or a solvent–olive oil blend. Until a laboratory develops its own historical control base, it is also preferable to include a positive control group. Either 0.25% dinitricholorobenzene or 0.05% oxazalone is recommended as a positive control. If the vehicle for the positive control is different than the vehicle for the test material, then two vehicle control groups may be necessary.

This method has been extensively validated in two international laboratory exercises (Basketter et al., 1991; Loveless et al., 1996). In the earlier work (Basketter et al., 1991), there was good correlation between the results obtained with guinea pig tests and those obtained with the LLNA. In the 1996 report, for example, five laboratories correctly identified dinitrochlorobenzene and oxazalone as sensitizers and the fact that *p*-aminobenzoic acid was not (Loveless et al., 1996). Arts and colleagues (1996) demonstrated that rats could be used as well as mice. Interestingly, they validated their assay (for both rats and mice) using BrDU uptake and immunohistochemical staining (rather than [$^3$H] thymidine) to quantitate lymph node cell proliferation.

This method is relatively quick and inexpensive because it uses relatively few mice (which are much less expensive than guinea pigs) and takes considerably less time than traditional guinea pig assays (Basketter and Scholes, 1991). It has an advantage over other methods in that it does not depend on an arguably subjective scoring system and produces a quantifiable endpoint. It does require a radiochemistry laboratory and license. Unless one already has an appropriately equipped laboratory used for other purposes (most likely metabolism studies), setting one up for the sole purpose of running the LLNA does not make economic sense. The standard version of the test has been adopted by the OECD (OECD429), ICCVAM (Interagency Coordination Committee on the Validation of Alternative Methods), European Medicines Agency (EMEA), and FDA (see Figure 8) but also has been shown to have a modest false-positive rate (misidentifying strong irritants as sensitizers).

Indeed, it has become clear that certain classes of structures (e.g., surfactants, fatty acids, fatty alcohols, siloxanes, and polyols) yield high incidences of false-positive outcomes. Though initially thought to be associated with (and due to) strong irritant responses, this is now clearly not the case (Kreiling et al., 2008).

The other possibility, of course, is clinical evaluation using the human repeat insult path test (HRIPT). This five-week hazard test is rarely taken due to cost,

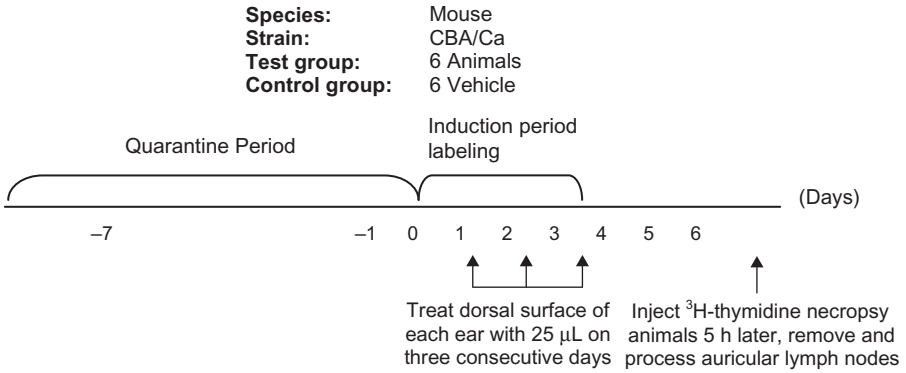


***Figure 8*** *Mouse LLNA (ICVAM protocol) Modification using flow cytometry instead of radiolabeling is preferable.*

though actual clinical formulations are effectively evaluated clinically in the normal course of clinical development of a drug.

## 7.10 Photosensitization

Some compounds can act as photoantigens that require exposure to UV light to become photoactive haptens. The physiochemical characteristics of compounds can sometimes reveal them as potentially photoactive, particularly if they are photounstable to light in the UV range. There are several in vivo tests that are used for determining photosensitization. The two assays described here are similar to those previously described for DTH with the primary exception that the dermal test sites are exposed to a light source during the induction and challenge phases. Like the DTH assays, these assays may also include a second challenge dose or the use of various methods of abrasion and occlusion to increase dermal penetration of the test article. The methods outlined below are more thoroughly described in Gad and Chengelis (1998).

***Harber and Shalita Method*** This method (Harber and Shalita, 1975) is similar to the Buehler test in that the compound is applied topically to guinea pigs without the use of adjuvants; however, the test site is not occluded during exposure. During the induction phase, the compound is applied on alternate days during a 12-day period for a total of six applications. Thirty minutes after each application, the test sites are exposed to a sunlamp for 30 min and then to a black light for 30 min. The challenge dose is applied 21 days after the last induction exposure. Thirty minutes after application, the challenge site is shielded with a 3-mm-thick piece of glass, while the site is exposed to black light for an additional 30 min. The glass filters out erythrogenic (causing redness) radiation of less than 320 nm that may confound scoring the reaction. The challenge sites are observed and scored for inflammatory reactions 24 h later.

***Armstrong Method*** This test (Ichikawa et al., 1981) resembles the guinea pig maximization test in that the FCA is injected intradermally at the test sites; however, covered Hill Top chambers are used to apply and occlude the test article at the test site, as was described for the Buehler test. During induction, four intradermal injections of FCA are administered at the test site; the test article is applied to the Hill Top chamber, which is then applied over the test site and occluded. After 4 h, the patches are removed and the test site is exposed to UV-A light (320–400 nm) for 30 min. Five additional applications of the test article (without FCA) with subsequent exposure to light are made on alternate days throughout the 11-day induction period. Nine to 13 days after the last induction exposure, the animals are challenged for 2 h with a nonirritating concentration of the test article on an occluded Hill Top chamber. The patches are then removed, and the sites are exposed to the UV-A light. Each site is graded for inflammation 24 and 48 h after challenge.

Despite its extreme clinical importance, the evaluation of small-molecule pharmaceuticals for allergenic potential is extremely unreliable in nonclinical toxicology studies. First, a drug that has been proved to be immunogenic in a laboratory animal species may not necessarily be immunogenic in humans, and vice versa. Second, the subtle factors that determine whether an individual responds to an antigen with an IgG or an IgE response can hardly be extrapolated from one species to another. Thus, an immune response that manifests as an allergic response in a laboratory animal may not necessarily manifest as an allergic response in humans, and vice versa.

With regard to antibody-mediated hypersensitivity, three methods have been used extensively to assess the induction of drug-specific (type I) anaphylactic reactions:

- Passive cutaneous anaphylaxis (PCA) assay
- Active cutaneous anaphylaxis (ACA) assay
- Active systemic anaphylaxis (ASA) assay

All three assays are normally conducted in guinea pigs, which is the only rodent species that actually develops symptoms of severe anaphylactic reactions and even fatal allergic shock. However, the usefulness of these assays for the safety assessment of drugs is considered limited. Since IgE as well as IgG antibodies can cause anaphylactic reactions in guinea pigs, a positive result in any of the three assays can only be weighted as proof of immunogenicity, but not allergenicity of a drug. The PCA, ACA, and ASA assays are therefore not requested or recommended for the routine evaluation of allergenicity of investigational new drugs by any regulatory agency.

The situation is as unsatisfactory for the prediction of a type II and III allergenic potential of drugs. Although there are examples of drugs that are associated with type II and III hypersensitivity reactions, there are no standard nonclinical methods for predicting these effects. Manifestations of both kinds of immunopathies are often indistinguishable from direct, non-immune-

mediated drug toxicity. Thus, in some instances of hemolytic anemia, vasculitis, or glomerulonephritis, which may be observed during standard toxicity studies, follow-up studies should be considered to determine if antibody-mediated immune mechanisms are involved.

Since all available nonclinical assays to assess the antibody-based allergenic potential of drugs have a limited predictivity for the human situation, detection of drug immunogenicity should already be considered to be a potential safety alert. Whether proven immunogenicity in a nonclinical test does in fact lead to allergic implications in patients can only be convincingly demonstrated (or excluded) in clinical trials or even later still during postmarketing surveillance of the approved drug.

The most robust and predictive procedures available for assessment of allergenicity are those measuring the skin-sensitizing potential of topically administered drug substances. In these cases, a drug has to permeate the keratinized skin, bind to MHC molecules of dermal APCs, and stimulate $CD4^+$ T cells for proliferation and $T_H1$ differentiation. Any drug that is able to induce the above sequence of events will inevitably induce an inflammatory reaction, a so-called DTH reaction, after subsequent challenge exposure to the skin. Thus, in this special situation of dermal sensitization, a proof of immunogenicity is also a proof of allergenicity.

When a drug is intended for topical administration (dermal, ocular, vaginal, rectal), the skin-sensitizing potential of the drug should be determined using an appropriate assay based on sensitization and challenge as part of nonclinical safety evaluation. The most common methods for evaluating the dermal sensitizing potential of drugs have been the Buehler assay (BA) and the GPMT. Both in vivo guinea pig–based methods are reliable and have demonstrated a high correlation with known human skin sensitizers. Techniques using mice, like the mouse ear swelling test, which uses an induction and challenge pattern similar to the traditional guinea pig tests, or the murine LLNA, correlate well with traditional guinea pig tests. Especially the LLNA, which is designed to detect lymphoproliferation in draining lymph nodes of the exposition area instead of inflammation following challenge, gives quantitative results. Furthermore, the assay is now accepted by most regulatory agencies with regard to reduction, refinement, and replacement of animal experimentation.

Pseudoallergic (anaphylactoid) reactions, which are independent of antigen-specific immune responses, result from direct drug-mediated histamine release or complement activation. Anaphylactoid reactions can be differentiated from true IgE-mediated anaphylaxis by in vitro testing of drug-induced histamine release from mast cell lines or by the detection of activated complement products in serum of animals showing signs of anaphylaxis.

## 8 APPROACHES TO COMPOUND EVALUATION

As outlined above, there are numerous assays available to assess the various endpoints that are relevant to immunotoxicity. Early in the development

process, a new compound should be evaluated with regard to various factors that may flag it as a potential immunotoxin, including chemical, structural, or physiochemical properties (e.g., photoallergin) and therapeutic class (i.e., immunomodulators, anti-inflammatory drugs, and antimetabolites). Compounds from therapeutic or structural classes that are known to be potential immunotoxins or immunomodulators should be evaluated for the effects in question on a case-by-case basis. With the exception of immunomodulators, protein products, and products of biotechnology, the majority of pharmaceuticals can be assessed for most forms of immunotoxicity during routine preclinical toxicity tests. In general, a well-conducted preclinical toxicity study can detect most serious immunotoxins in the form of altered clinical, hematological, or histological endpoints. For example, possible effects on humoral immunity may be indicated from clinical observation of gastrointestinal or respiratory pathology, changes in serum total protein and globulin, and histological changes in lymphoid cellularity. Likewise, effects on the cell-mediated response may be observed as increases in infections and tumor incidences and by changes in the T-cell compartments of lymphoid tissues. In the case of immunosuppressive drugs such as cyclophosphamide and cyclosporin A, the immune effects seen in rodents are similar to those observed in the clinic (Dean et al., 1987).

If perturbations are observed in any hematological or histopathological indicators of immunotoxicity, it is then prudent to follow up these findings with one or more of the following:

- Use of special immunochemical and cytological assays that can be performed retrospectively on samples taken from the animals in question.
- Use of more specific in vitro assays to further assess effects on the pertinent target system and potential mechanism of activity.
- Use of more specific in vitro and ex vivo assays to determine toxicological significance.
- Inclusion of additional nonroutine parameters for immunotoxicity assessment in subsequent (longer term) toxicity assays. This can also include additional satellite groups for functional tests that may require coadministration of adjuvants, pathogens, or tumor cells.

### 8.1 Use of In Vivo Tests

In vivo tests are more relevant indicators than are in vitro tests of immunotoxicity since the dynamic interactions between the various immunocomponents as well as the pertinent pharmacokinetic (absorption, distribution, plasma concentrations) and metabolic factors are taken into consideration. However, it is important to select the appropriate animal model and to design the protocol such that it will accurately reflect drug (or relevant metabolite) exposure to humans. For example, one should consider species variability

when selecting the animal model since biological diversity may further obscure the ability to accurately predict human toxicity.

***Species Selection*** When possible, the species selected should demonstrate similar pharmacology and toxicity profiles to those anticipated in the clinic. Thus, the test animals should metabolize the drug and express the same target organ responses and toxic effects as humans. Although the rat and dog are the most common species used in preclinical safety tests, they are not as well characterized and validated as the mouse for assessing effects on immune function. For most immunosuppressive drugs, rodent data on target organ toxicities and comparability of immunosuppressive doses have been reflective of what was later observed in the clinic. Immunosuppressive effects and the doses that produced them have been shown to be similar in the various species that are typically used in preclinical safety tests (Dean and Thurmond, 1987). An exception has been seen with glucocorticosteroids, which are lympholytic in rodents but not in primates (Claman, 1972; Haynes and Murad, 1985). Although some compounds may show different pharmacokinetics and pharmacological effects in rodents than in humans, rodents still appear to be the most appropriate animals for assessing immunotoxicity of non-species-specific compounds (Dean and Thurmond, 1987).

The appropriate animal model is also important when performing follow-up testing or additional mechanistic tests to further investigate findings observed as part of the routine preclinical safety tests. When possible, these studies should employ the same animals or animal model in which the change was initially observed for several reasons, as outlined by Bloom et al. (1987), including:

- The incidence of adverse effect may be low and not easy to reproduce.
- Another species may not be genetically susceptible to the toxic effect.
- The biological significance of the change is well defined in that model.
- If the change follows long-term exposure to the drug, reproducing the effect in another model may be costly and impractical.

***Route and Treatment Regimen*** When possible, it is important to administer the compound by the route and treatment regimen most appropriate for demonstrating the specific response and/or reflecting the intended clinical route of administration. It is also necessary for the compound to be in the same dosage form (i.e., salt form, excipients, solubilizers) that will be used clinically. With the exception of tests for contact hypersensitivity, most of the in vivo tests can be carried out with a minimum of three dose levels, which are needed to assess a dose–response relationship. Dose levels should range from the proposed clinical dose, or one that approximates the no-effect level, to a maximum-tolerated (or limit) dose that is lower than the $LD_{10}$ but that produces some evidence of general toxicity (e.g., reduced body weight). A wide

dose interval may be necessary to detect immune changes that show a nonlinear dose response. Proper dose selection is crucial for a meaningful interpretation of test results since severe stress and malnutrition may produce indirect immunotoxic effects that would confound a clear interpretation of the data.

For compounds such as antibiotics, with a relatively short duration of therapeutic exposure, a short treatment period of one to two weeks in the animal model is generally appropriate. Longer treatments may not be suitable for these drugs since animals can adapt to toxic doses or develop a tolerance by inducing enzymes that increase metabolism of the drug. However, for compounds with intended chronic or prolonged usage, animals should be treated at least daily for at least a month to assess the cumulative effects of the drug.

## 8.2 Use of In Vitro Tests

In vitro tests are useful as sensitive follow-up tests to determine potential effects or mechanisms of effects on specific cell types at the cellular and molecular levels. In addition, most are relatively simple to perform and ex vivo tests can be performed in conjunction with preclinical in vivo tests. There are several advantages to using in vitro tests:

- Specific cell types of humoral components of the immune system can be isolated and studied.
- Cells can be stimulated with various mitogens to assess their proliferative functions in vitro.
- For mechanistic studies, cells and their secretory products can be systematically studied in isolation and in various combinations to assess their interactions and cell-to-cell communications.

However, for general preclinical assessments and screening purposes, in vitro tests should be well validated and used cautiously for several reasons:

- They may over- or underestimate an effect or give contradictory results compared to in vivo tests.
- Most immunotoxic responses express a clear dose–response relationship that can be used for human risk assessment. However, it is more difficult to extrapolate in vitro concentrations than in vivo animal doses (plasma concentrations) to the clinical dose.
- It is difficult to simulate in vitro the interaction of all of the various cell types and modulators of immune function that make up the in vivo system.
- Cells can be harvested from a variety of sources and each source may have a different sensitivity since they may be at various stages of maturation or activation.

### 8.3 Assessment of Immunotoxicity and Immunogenicity/Allergenicity of Biotechnology-Derived Drugs

This is an area outside of current ICH S8 guidance and yet of clinical importance to the safety of current new therapeutics. The decision on a suitable species for preclinical immunotoxicity assessment of biotechnology-derived drugs must be made on a case-by-case basis. In any case, the limits of predictivity should be clearly stated in the rationales for choosing a certain assay protocol. The biological activity together with species and/or tissue specificity of many biotechnology-derived pharmaceuticals (e.g., recombinant cytokines, therapeutic antibodies) often preclude standard toxicity testing designs in commonly used species (e.g., rats and dogs). The same holds for immunotoxicity testing. The design of immunotoxicity testing programs for biotechnology-derived drugs should include the use of relevant species. A relevant species is one in which the test material is pharmacologically active due to expression of the receptor or an epitote (in the case of monoclonal antibodies). A variety of techniques (e.g., immunochemical or functional in vitro tests) can be used to identify a relevant species. In some cases nonhuman primates may be the only suitable species available. When no relevant species exists, the use of transgenic mice expressing the human receptor or epitope may be accepted by regulatory agencies. The information gained from use of a transgenic mouse model expressing the human receptor is optimized when the interaction of the product and the humanized receptor has similar physiological consequences to those expected in humans.

In other cases, the use of the homologous animal protein instead of the human counterpart may be considered. While useful information may also be gained from the use of homologous proteins, it should be noted that the production process, range of impurities/contaminants, pharmacokinetics, and exact pharmacological mechanism(s) may differ between the homologous form and the product intended for clinical use. Thus, from a formalistic point of view, the test item used in such protocols is not identical to the drug substance to be assessed. Results from such studies can therefore only be weighed as "supportive data." In such situations, it is highly recommended to discuss the testing strategy with the responsible regulatory agency for scientific advice. Where it is not possible to use transgenic animal models of homologous proteins, it may be advisable to assess certain aspects of potential immunotoxicity in vitro using human material like peripheral blood mononuclear cells (PBMCs), monocyte-derived macrophages, or long-term cultivated cell lines of hematopoietic origin.

Most biotechnology-derived pharmaceuticals intended for human use are per se immunogenic in animals. The induction of antibody formation in animals is therefore not predictive of a potential for antibody formation in humans. In this regard, the results of, for example, a guinea pig anaphylaxis test, which is usually positive for xenogenic protein products, is not predictive for reactions in humans. Such studies are therefore considered of little value

for the routine evaluation of these types of products. It must be kept in mind that even humanized proteins may be immunogenic in humans. In most cases, reliable information on immunogenicity of biotechnology-derived drugs can therefore only be obtained during clinical studies. However, immunogenicity studies in animals using biotechnology-derived drugs may yield valuable information when comparing the immunogenic potential of a test compound with a biosimilar reference compound or between different production batches.

Even if immunogenicity assessment of biotechnology-derived pharmaceuticals has limited predictivity for the human situation, measurement of antibodies associated with the administration of biotechnology-derived drugs should always be included in the design of a repeat-dose toxicity study (Wierda et al., 2001; Vohr, 2005). Antibody responses should be characterized with regard to titer, number of responding animals, and neutralizing or nonneutralizing antibodies. Furthermore, the detection and quantization of antibodies should be correlated with any pharmacological and/or toxicological changes. Specifically, the effects of antibody formation on pharmacokinetic/pharmacodynamic parameters, incidence and/or severity of adverse effects, complement activation, or the emergence of new toxic effects should be considered when interpreting the data. Attention should also be paid to the evaluation of possible pathological changes related to immune complex formation and deposition.

## 8.4 Suggested Approaches to Evaluation of Results

Several rodent toxicity studies have shown impaired host resistance to infectious agents or tumor cells at exposure levels of drugs that did not cause overt signs of toxicity (Vos, 1977; Dean et al., 1982). One serious limitation to the incorporation of specific immunotoxicological evaluations into general use in safety assessment for pharmaceuticals is a lack of clarity in how to evaluate and use such findings. This problem is true for all new diagnostic techniques in medicine and for all the new and more sensitive tools designed to evaluate specific target organ toxicities. Ultimately, as we have more experience and a reliable database that allows us to correlate laboratory findings with clinical experience, the required course of action will become clearer. However, some general suggestions and guidance can be offered:

- First, it is generally agreed that adverse effects observed above a certain level of severity should be given the same importance as any other life-threatening events when assessing biological significance. These effects are so severe that they are detected as part of the routine evaluations made in safety assessment studies. Such findings may include death, severe weight loss, early appearance of tumors, and the like. Findings such as significantly increased mortalities in a host resistance assay would also fit into this category.
- Second, there are specific endpoint assays for which an adverse outcome clearly dictates the action to be taken. These endpoints include either

immediate or delayed hypersensitivity reactions because once the individual is sensitized, a dose–response relationship may not apply.

- Third, as with most toxicological effects, toxic effects to the immune system are dependent upon dose to the target site. The dose–response curve can be used to determine no-effect and low-effect levels for immunotoxicity. These levels can then be compared to the therapeutic levels to assess whether there is an adequate margin of safety for humans.

If we consider both the specific immunotoxicity assays surveyed earlier in the chapter and the arrays of endpoints evaluated in traditional toxicology studies, which may be indicative of an immune system effect, these guidelines leave many potential questions unanswered. As additional data on individual endpoints indicative of immune system responses are collected, the pharmaceutical toxicologist is challenged with various issues regarding assay interpretation and relevance to proposed (or future) clinical trials. For example, what do significant but non-life-threatening decreases in antibody response, lymphocyte numbers, macrophage functions, or host resistance in an animal mean about the clinical use of a drug in a patient? The intended patient population is clearly relevant here—if the disease is one in which the immune system is already challenged or incorrectly modulated, any immune system effect other than an intended one should be avoided. There are several additional considerations and questions that should be answered when evaluating the biological and clinical significance of a statistically significant immune response:

1. *Is there a dose response?* The dose response should be evaluated as a dose-related trend in both incidence and severity of the response. If there is a dose-related response, is the lowest dose (preferably plasma level) at which the effect is seen near or below the target clinical dose (plasma level), and is there an adequate therapeutic margin of safety?
2. *Does the finding stand alone?* Is a change observed in only one parameter, or are there correlated findings that suggest a generalized, biologically significant effect? For example, are there changes in lymph node and spleen weights and morphological changes in these tissues to accompany changes in lymphocyte numbers?
3. *Is the effect a measure of function or a single endpoint measurement?* Functional measures such as host resistance of phagocytosis involve multiple cells and immunocomponents and, therefore, are considered to be more biologically relevant than a significant change in a single endpoint measurement (e.g., T-cell number).
4. *Is the effect reversible?* Reversibility of a response is dependent on the drug itself, exposure levels/duration, and factors related to the test animal (metabolic capability, genetic susceptibility, etc.). Most effects produced by immunosuppressive drugs have been shown to be reversible after cessation of therapy, such as those produced during cancer chemother-

apy. However, if a tumor develops before the immune system is restored, the effect is not reversible, as is the case of secondary tumors related to chemotherapy.

5. *Is there sufficient systematic toxicity data available at levels that demonstrate adequate exposure?* If a study was designed such that there was insufficient exposure or duration of exposure to potential lymphoid target tissues, the test protocol may not be adequate to demonstrate an adverse effect.

In general, a well-conducted long-term study in two species with no indication of immunotoxicity, based on the considerations outlined above, should be adequate to evaluate the potential for drug-induced immunotoxicity. If the results from these studies do not produce evidence if immune-specific toxicity after examination of standard and/or additional hematological, serum chemical, and histopathological parameters, then additional testing should not be indicated. However, if there are structure–activity considerations that may indicate a potential for concern, of if significant abnormalities are observed that cannot be clearly attributed to other toxicities, then it is important to perform additional tests to fully assess the biological significance of the findings.

## 9 PROBLEMS AND FUTURE DIRECTIONS

There are some very pressing problems for immunotoxicology, particularly in the context of pharmaceuticals and biological therapeutics and the assessment of their safety. Unlike industrial chemicals, environmental agents, or agricultural chemicals, pharmaceutical products are intended for human exposure, are usually systemically absorbed, and have intentional biological effects on humans—some of which are intentionally immunomodulating (interleukins, growth factors) or immunotoxic (cyclosporin, cyclophosphamide).

***Data Interpretation*** The first major issue was presented and explored in the preceding section. This is how to evaluate and utilize the entire range of data that current immunotoxicological methodologies provide to determine the potential for immunotoxicity and how to interpret the biological significance of minor findings.

***Appropriate Animal Models*** As previously addressed, most routine preclinical toxicology tests are performed with rats and dogs; therefore, toxicity, pharmacokinetic, and pharmacology data are most abundant for these species. However, most immunological parameters are best characterized and validated with mice. In addition, the NTP test battery was developed for the mouse, and some of these assays cannot be readily transferred to the rat. Over the last few years, several laboratories have begun adapting tests to both the rat and the dog (Bloom et al., 1987; Thiem et al., 1988); however, efforts need

to continue along these lines to further our understanding of the immune responses in these species and how they correlate with other animal models and humans.

***Indirect Immunotoxic Effects*** A problem related to data interpretation is how to distinguish secondary effects that may indirectly result in immunotoxicity from the primary effects of immunotoxicity in preclinical toxicity studies. Various factors may produce pathology similar to that of an immunotoxin, including:

- Stress in a chronically ill animal as related to general toxicity, such as lung or liver damage, can result in immune suppression.
- Malnutrition in animals with drug-induced anorexia or malabsorption can trigger immune suppression.
- Infections and/or parasites may also modulate immune parameters.

These indirect factors must be systematically ruled out, and additional mechanistic studies may be necessary to address this problem. The potential for some indirect effects may be assessed through histopathological evaluation of endocrine organs such as the adrenals and pituitary.

***Hypersensitivity Tests*** Probably the largest immunotoxicity concern in clinical studies is unexpected hypersensitivity reactions. While the available guinea-pig- and mouse-based tests for delayed contact hypersensitivity resulting from dermal exposure are generally good predictors, there are currently no well-validated models for either immediate or (dermal) delayed hypersensitivity responses resulting from either oral ingestion or parenteral administration. Yet these two situations are the largest single cause for discontinuing clinical trials.

One assay that may hold some promise for delayed hypersensitivity is an adoptive transfer-popliteal lymph node assay (Klinkhammer et al., 1988). This assay, based on the techniques previously described for the popliteal lymph node assay, allows assessment of hypersensitivity following systemic exposure of the drug. Donor mice are first injected with drug for five consecutive days. After a four-week latency period, potentially sensitized T cells obtained from the spleen are injected into the footpad of a syngenic mouse together with a subcutaneous challenge dose of the drug. Two to five days after the cell transfer, the popliteal lymph nodes are measured and observed for evidence of a response (enlargement). Once this assay is validated, it should allow for a more relevant assessment of hypersensitivity for drugs that are administered systemically (Gleichmann et al., 1989).

***Autoimmunity*** Traditional methods for assessing immunotoxicity as part of routine preclinical toxicity tests are primarily geared toward the detection of immunosuppressive effects. Although it is possible to incorporate clinical

methods for detecting immune complexes and autoantibodies into the preclinical test protocols, the significance of adverse findings is ambiguous. Since these effects have a genetic component to their expression, the relevance of findings in animals is of questionable significance, particularly since these findings in the clinic do not always correlate with pathological effects.

***Functional Reserve Capacity*** As previously discussed, the immune system has a tremendous reserve capacity that offers several levels of protection and backups to the primary response. As a consequence, this functional reserve can allow biologically significant, immunotoxic insults to occur without the appearance of morphological changes. Furthermore, adverse effects may remain subclinical until the organism is subjected to undue stress or subsequent challenge. Thus, there is some concern that routine immunopathological assessments by themselves may not be sufficiently sensitive to detect all immunotoxins, particularly when testing is conducted in a relatively pathogen-free, stress-free laboratory environment.

***Significance of Minor Perturbations*** Although the immune system has a well-developed reserve capacity, some of these systems may act synergistically rather than independently. For instance, a macrophage can recognize and kill bacteria coated with antibodies more effectively than can either the macrophage or antibodies alone. Thus, even minor deficiencies and impairments may have some impact on the organism's ability to fend off infection or tumors, particularly if the organism is very young, old, ill, stressed, genetically predisposed to certain cancers, or otherwise immunocompromised. These considerations lead to some additional questions that must be addressed:

- What level of immunosuppression will predispose healthy or immunocompromised individuals to increased risk of infections or tumors?
- Will slight disturbances or immunosuppression lead to a prolonged recovery from viral or bacterial infections?
- Will slight up-modulation for extended periods result in autoimmune diseases or increased susceptibility to allergy?
- Are individuals that are slightly immunosuppressed at higher risk of developing AIDS after exposure to HIV?

***Biotechnology Products*** Immunotherapeutics such as interferons and interleukins hold tremendous promise for those diseases where malfunctioning of the immune system is not the root of pathogenesis. Likewise, many of the new approaches to therapy of yet untreatable diseases are aimed at modulating the body's own immune system. Many of the new therapeutics coming from biotechnology are proteins of human origin. As such, they can evoke antibody responses in nonhuman species that are not indicative of what will be seen in patients. Meaningful evaluations must allow the toxicologist to

discriminate between those responses that are relevant to clinical development/utilization and those that are not.

In summary, it is the role of preclinical immunotoxicology testing to allow us to identify potential immune hazards early in development, before they are found in the clinic, and to provide us with a mechanistic understanding for the basis of these effects so that we may direct the development of alternative agents and/or treatment regimens to avoid them. The challenge for the toxicologist is to determine the appropriate course of action for evaluating each unique drug and to differentiate the desired therapeutic effects from the undesired and potentially adverse effects.

## REFERENCES

Ader, R., and Cohen, N. (1993). Psychoneuroimmunology: Conditioning and stress. *Ann. Rev. Psychol.* 44:53–85.

Alper, J. C., Wieman, M. C., Rueckl, F. S., McDonald, C. J., and Calabresi, P. (1985). Rationally designed combination chemotherapy for the treatment of patients with recalcitrant psoriasis. *J. Am. Acad. Dermatol.* 13:567–577.

Andersen, K. E., and Maibach, H. I. (Eds.) (1985). Contact allergy predictive test in guinea pigs. In *Current Problems in Dermatology*, Vol. 14, pp. 59–106. S. Karger, Basel.

Arts, J. H. E., Droge, S. C. M., Bloksma, N., and Kuper, C. F. (1996). Local lymph node activation in rats after dermal application of the sensitizers 2,4-dinitrochlorobenzene and trimellitic anhydride. *Food Chem. Toxicol.* 34:55–62.

Bach, F. H., and Voynow, N. K. (1966). One-way stimulation in mixed leukocyte cultures. *Science* 153:545–547.

Bakke, O. M., Wardell, W. M., and Lasagna, L. (1984). Drug discontinuations in the United Kingdom and United States, 1964–1983: Issues of safety. *Clin. Pharmacol. Ther.* 35:559–567.

Barnett, M. J., Waxman, J. H., Richards, M. A., Ganesan, T. S., Bragman, K. S., Rohatiner, A. Z. S., and Lister, T. A. (1985). High-dose cytosine arabinoside in the initial treatment of acute leukemia. *Semin. Oncol.* 12:133–138.

Basketter, D. A., and Scholes, E. W. (1991). Comparison of the local lymph node assay with the guinea pig maximization test for the detection of a range of contact allergens. *Fd. Chem. Toxicol.* 30:65–69.

Basketter, D. A., Scholes, E. W., Kimber, I., Botham, P. A., Hilton, J., Miller, K., Robbins, M. C., Harrison, P. T. C., and Waite, S. J. (1991). Interlaboratory evaluation of the local lymph node assay with 25 chemicals and comparison with guinea pig test data. *Toxicol. Methods* 1:30–43.

Batchelor, F. R., Dewdney, J. M., and Cazzard, D. (1965). Penicillin allergy: The formation of penicilloyl determinant. *Nature (London)* 206:362–364.

Bellon, B., Capron, M., Druet, E., Verroust, P., Wial, M. C., Sapin, C., Girard, J. F., Foidart, J. M., Mahieu, P., and Druet, P. (1982). Mercuric chloride induced autoimmune disease in Brown-Norway rats: Sequential search for anti-basement membrane antibodies and circulating immune complexes. *Eur. J. Clin. Invest.* 12:127–133.

Bigazzi, P. E. (1985). Mechanisms of chemical-induced autoimmunity. In J. H. Dean, M. I. Luster, A. E. Munson, and H. Amos (Eds.), *Immunotoxicology and Immunopharmacology*. Raven, New York, p. 277.

Bigazzi, P. E. (1988). Autoimmunity induced by chemicals. *Clin. Toxicol.* 26:125–126.

Bloom, J. C., Blackmer, S. A., Bugelski, P. J., Sowinski, J. M., and Saunders, L. Z. (1985a). Gold-induced immune thrombocytopenia in the dog. *Vet. Pathol.* 22:492–499.

Bloom, J. C., Thiem, P. A., and Morgan, D. G. (1987). The role of conventional pathology and toxicology in evaluating the immunotoxic potential of xenobiotics. *Toxicol. Pathol.* 15:283–293.

Bloom, J. C., Thiem, P. A., Sellers, T. S., Deldar, A., and Lewis, H. B. (1985b). Cephalosporin-induced immune cytopenia in the dog—Demonstration of cell-associated antibodies. *Blood* 66:1232.

Brunton, L. L., Lazo, J. S., and Parker, K. L. (2006) *Goodman & Gilman's The Pharmacologic Basis of Therapeutics*, 11th ed. McGraw-Hill, New York.

Buehler, E. V. (1964). A new method for detecting potential sensitizers using the guinea pig. *Toxicol. Appl. Pharmacol.* 6:341.

Burleson, G. R., Dean, J. H., and Munson, A. E. (1995). *Methods in Immunotoxicology*. Wiley-Liss, New York.

Burnet, F. M. (1970). The concept of immunological surveillance. *Prog. Exp. Tumor. Res.* 13:1–27.

Calabresi, P., and Chabner, B. A. (1990). Antineoplastic agents. In A. G. Goodman, T. W. Rall, A. S. Nies, and P. Taylor (Eds.), *The Pharmacological Basis of Therapeutics*. Pergamon, New York, pp. 1209–1263.

Calabresi, P., and Parks, R. (1985). Antiproliferative agents and drugs used for immunosuppression. In A. G. Gilman, L. S. Goodman, T. W. Rall, and F. Murad (Eds.), *Goodman and Gilman's The Pharmacological Basis of Therapeutics*, 7th ed. Macmillan, New York, pp. 1247–1306.

Caspritz, G., and Hadden, J. (1987). The immunopharmacology of immunotoxicology and immunorestoration. *Toxicol. Pathol.* 15:320–322.

Center for Devices and Radiological Health (CDRH) (1997). Immunotoxicology testing framework. Draft document. U.S. Department of Health and Human Services, Washington, DC.

Center for Drug Evaluation and Research (CDER) (2001). Guidance for industry: Immunotoxicology evaluation of investigational new drugs. U.S. Department of Health and Human Services, Washington, DC.

Choquet-Kastylevsky, G., Vial, T., and Descotes, J. (2001). Drug allergy diagnosis in humans: Possibilities and pitfalls. *Toxicology* 158:1–10.

Claman, H. (1972). Corticosteroids and lymphoid cells. *New Engl. J. Med.* 287:388–397.

Clark, W. R. (1983). *The Experimental Foundations of Modern Immunology*, 2nd ed. Wiley, New York.

Colvin, M. (1982). The alkylating agents. In B. A. Chabner (Ed.), *Pharmacologic Principles of Cancer Treatment*. W. B. Saunders, Philadelphia, pp. 276–308.

Coombs, R. R. A., and Gell, P. G. H. (1975). Classification of allergic reactions responsible for clinical hypersensitivity and disease. In P. G. H. Gell, R. R. A. Coombs, and D. J. Lachman (Eds.), *Clinical Aspects of Immunology*. Blackwell Scientific, Oxford, p. 761.

Cromwell, O., Durham, S. R., Shaw, R. J., Mackay, J. A., and Kay, A. B. (1986). Provocation tests and measurements of mediators from mast cells and basophils in asthma and allergic rhinitis. In D. M. Weir, L. A. Herzenberg, and C. Blackwell (Eds.), *Handbook of Experimental Immunology*, 4th ed. Blackwell, Oxford, pp. 127.1–127.51.

Dean, J. H., Boorman, G. A., Luster, M. I., Adkins, B. J., Lauer, L. D., and Adams, D. O. (1984). Effect of agents of environmental concern on macrophage functions. In A. Volkman (Ed.), *Mononuclear Phagocyte Biology*. Dekker, New York, pp. 473–485.

Dean, J. H., Luster, M. I., and Boorman, G. A. (1982). Immunotoxicology. In P. Sirois and M. Rola-Pleszczynski (Eds.), *Immunopharmacology*. Elsevier Biomedial, Amsterdam, pp. 349–397.

Dean, J. H., Luster, M. I., Boorman, G. A., Luebke, R. W., and Lauer, L. D. (1980). The effect of adult exposure to diethylstilbestrol in the mouse: Alterations in tumor susceptibility and host resistance parameters. *J. Reticuloendothelial. Soc.* 28:571–583.

Dean, J. H., and Thurmond, L. M. (1987). Immunotoxicology: An overview. *Toxicol. Pathol.* 12:265–271.

Dean, J. H., Thurmond, L. M., Lauer, L. D., and House, R. V. (1987). Comparative toxicology and correlative immunotoxicology in rodents. In E. J. Burger, R. G. Tardiff, and J. A. Bellanti (Eds.), *Environmental Chemical Exposure and Immune System Integrity*. Princeton Scientific, Princeton, pp. 265–271.

Deldar, A., Lewis, H. B., Bloom, J. C., Apostoli, A., and Weiss, L. (1985). Residual stem cell defects associated with cephalosporin therapy in dogs. *Blood* 66:1202.

Descotes, G., and Mazue, G. (1987). Immunotoxicology. *Adv. Vet. Sci. Compar. Med.* 31:95–119.

DeSwarte, R. D. (1986). Drug allergy: An overview. *Clin. Rev. Allergy* 4:143–169.

DeWeck, A. L. (1983). Immunopathological mechanisms and clinical aspects of allergic reactions to drugs. In A. L. deWeck and H. Bundgaard (Eds.), *Handbook of Experimental Pharmacology: Allergic Reactions to Drugs*. Springer-Verlag, New York, pp. 75–133.

Dighiero, G., Lymberi, J., Marie, J. C., Rouyse, S., Butler-Browne, G. S., Whalen, R. G., and Avrameas, S. (1983). Murine hybridomas secreting natural monoclonal antibodies reacting with self antigens. *J. Immunol.* 135:2267–2271.

Druet, P., Bernard, A., Hirsch, F., Weening, J. J., Gengoux, P., Mahieu, P., and Brikenland, S. (1982). Immunologically mediated glomerulonephritis induced by heavy metals. *Arch. Toxicol.* 50:187–194.

Duke, S. S., Schook, L. B., and Holsapple, M. P. (1985). Effects of *N*-nitrosodimethylamine on tumor susceptibility. *J. Leukocyte Biol.* 37:383–394.

Elion, G. B., and Hitchings, G. H. (1975). Azathioprine. In A. C. Sartorelli and D. G. Johns (Eds.), *Antineoplastic and Immunosuppressive Agents*. Springer-Verlag, Berlin, pp. 403–425.

Elliot, J. F., Lin, Y., Mizel, S. B., Bleackley, R. C., Harnish, D. G., and Peaettian, V. (1984). Induction of interleukin 2 messenger RNA inhibited by cyclosporin A. *Science* 226:1439–1441.

European Medicines Agency (EMEA) (2007, Jan.). Immunogenicity assessment of biotechnology derived therapeutic proteins. EMEA, London.

Exon, J. H., Koller, L. D., Talcott, P. A., O'Reilly, C. A., and Henningsen, G. M. (1986). Immunotoxicology testing: An economical multiple assay approach. *Fund. Appl. Toxicol.* 7:387–397.

Gad, S. C., and Chengelis, C. P. (1998). *Acute Toxicology Testing*, 2nd ed. Academic, San Diego, CA.

Garratty, G., and Petz, L. D. (1975). Drug-induced immune hemolytic anemia. *Am. J. Med.* 58:398–407.

Gleichmann, E., Kimber, I., and Purchase, I. F. H. (1989). Immunotoxicology: Suppressive and stimulatory effects of drugs and environmental chemicals on the immune system. *Arch. Toxicol.* 63:257–273.

Godfrey, H. P., and Gell, P. G. H. (1978). Cellular and molecular events in the delayed-onset hypersensitivities. *Rev. Physiol. Biochem. Pharmacol.* 84:2.

Golub, E. S., and Green, D. R. (1991). *Immunology: A Synthesis*. Sinauer, Sunderland, MA.

Guillaume, J. C., Roujeau, J. C., and Touraine, R. (1984). Lupus systémique après protheses mammaires. *Ann. Derm. Verner.* 111:703–704.

Hadden, J. W., Cornaglia-Ferraris, P., and Coffey, R. G. (1984). Purine analogs as immunomodulators. In Y. Yamamura and T. Tada (Eds.), *Progress in Immunology*, Vol. 4. Academic, London, pp. 1393–1407.

Haley, P., et al. (2005) STP position paper: Best practice guideline for the routine pathology evaluation of the immune system. *Toxicol. Pathol.* 33:404–407.

Harber, L. C., and Shalita, A. R. (1975). The guinea pig as an effective model for the demonstration of immunologically-mediated contact photosensitivity. In H. Maibach (Ed.), *Animal Models in Dermatology*. Churchill Livingstone, New York, pp. 90–102.

Harmon, W. E., Parkman, R., Gavin, P. T., Grupe, W. E., Ingelfunger, J. R., Yunis, E. J., and Levey, R. H. (1982). Comparison of cell-mediated lympholysis and mixed lymphocyte culture in the immunologic evaluation for renal transplantation. *J. Immunol.* 129:1573–1577.

Hassman, R. A., Lazarus, J. H., Dieguez, C., Weetman, A. P., Hall, R., and McGregor, A. M. (1985). The influence of lithium chloride on experimental autoimmune thyroid disease. *Clin. Exp. Immunol.* 61:49–57.

Hastings, K. L. (2001). Pre-clinical methods for detecting the hypersensitivity potential of pharmaceuticals: Regulatory consideration. *Toxicology* 158:85–89.

Haukaas, S. A., Hoisater, P. A., and Kalland, T. (1982). *In vitro* and *in vivo* effects of diethylstilbestrol and estramustine phosphate (Estracyte) on the mutagen responsiveness of human peripheral blood lymphocytes. *Prostate* 3:405–414.

Haynes, R. C., and Murad, F. (1985). Adrenocorticotropic hormone; adrenocortical steroids and their synthetic analogs; inhibitors of adrenocortical steroid biosynthesis. In A. G. Goodman, L. S. Goodman, T. W. Rell, and F. Murad (Eds.), *Goodman and Gilman's The Pharmacological Basis of Therapeutics*. Macmillan, New York, pp. 1459–1489.

Herold, K. C., Lancki, D. W., Moldwin, R. L., and Fitch, F. W. (1986). Immunosuppressive effects of cyclosporin A on cloned T cells. *J. Immunol.* 136:1315–1321.

Hinton, D. M. (1992). Testing guidelines for evaluation of the immunotoxic potential of direct food additives. *Crit. Rev. Fd. Sci. Nutrit.* 32:173–190.

Hollenberg, S. M., Giguere, V., Segui, P., and Evans, R. M. (1987). Colocalization of DNA-binding and transcriptional activation functions in the human glucocorticoid receptor. *Cell* 49:39–46.

Hunter, T., Urowitz, M. B., Gordon, D. A., Smythe, H. A., and Ogryzlo, M. A. (1975). Azathioprine in rheumatoid arthritis. A long-term follow-up study. *Arthr. Rheum.* 18:15–20.

Hutchings, P., Nador, D., and Cooke, A. (1985). Effects of low doses of cyclophosphamide and low doses of irradiation on the regulation of induced erythrocyte autoantibodies in mice. *Immunology* 54:97–104.

ICH (2005). Immunotoxicity Studies for Human Pharmaceuticals, Sept.

Ichikawa, H., Armstrong, R. B., and Harber, L. C. (1981). Photoallergic contact dermatitis in guinea pigs: Improved induction technique using Freund's complete adjuvant. *J. Invest. Dermatol.* 76:498–501.

Idsøe, O., Guthe, T., Willcox, R. R., and DeWeck, A. L. (1968). Nature and extent of penicillin side-reactions, with particular reference to fatalities from anaphylactic shock. *Bull. WHO* 38:159–188.

Jerne, N. K., and Nordin, A. A. (1963). Plaque formation in agar by single antibody-producing cells. *Science* 140:405.

Kahan, B. D., and Bach, J. F. (1988). Proceedings of the Second International Congress on Cyclosporine. *Transplant. Proc.* 20(Suppl.):11131.

Kay, J. E., and Benzie, C. R. (1984). Rapid loss of sensitivity of mitogen-induced lymphocyte activation to inhibition by cyclosporin A. *Cell. Immunol.* 87:217–224.

Kimber, I., and Weisenberger, C. (1989). A murine local lymph node assay for the identification of contact allergens: assay development and results of an initial validation study. *Arch Toxicol.* 63:274–282.

Kimber, I., Mitchell, J. A., and Griffin, A. C. (1986). Development of a murine local lymph node assay for the determination of sensitizing potential. *Food Chem. Toxicol.* 24:481–494.

Kimber, I., Hilton, J., Botham, P. A., et al. (1991). The murine local lymph node assay: results of an inter-laboratory trial. *Toxicol. Lett.* 55:203–213.

Klaus, G. G. B., and Hawrylowicz, C. M. (1984). Cell-cycle control in lymphocyte stimulation. *Immunol. Today* 5:15–19.

Kleinman, S., Nelson, R., Smith, L., and Goldfinger, D. (1984). Positive direct antiglobulin tests and immune hemolytic anemia in patients receiving procainamide. *New Engl. J. Med.* 311:809–812.

Klinkhammer, C., Popowa, P., and Gleichmann, H. (1988). Specific immunity to the diabetogen streptozotocin: Cellular requirements for induction of lymphoproliferation. *Diabetes* 37:74–80.

Koller, L. D. (1987). Immunotoxicology today. *Toxicol. Pathol.* 15:346–351.

Kreiling, R., Hollinagel, H. M., Hareng, L., et al. (2008). Comparison of the skin sensitizing potential of unsaturated compounds as assessed by the murine local lymph node assay (LLNA) and the guinea pig maximization test (GPMT). *Food Chem. Toxicol.* 46(6):1896–1904.

Kumagai, Y., Shiokawa, Y., Medsger, T. A. J., and Rodnan, G. P. (1984). Clinical spectrum of connective tissue disease after cosmetic surgery. Observations of eighteen patients and a review of the Japanese literature. *Arthr. Rheum.* 27:1–12.

Ladics, G. (2005) Assays for antibody production. In H.-W. Voher (Ed.), *Encyclopedia Reference of Immunotoxicity*. Springer, New York, p. SC-53.

Laskin, D. L., Laskin, J. D., Weinstein, I. B., and Carchman, R. A. (1981). Induction of chemotaxis in mouse peritoneal macrophages by phorbolester tumor promoters. *Cancer Res.* 41:1923.

Lawrence, D. A. (1985). Immunotoxicity of heavy metals. In J. H. Dean, M. I. Luster, A. E. Munson, and H. A. Amos (Eds.), *Immunotoxicology and Immunopharmacology*. Raven, New York, pp. 341–353.

Logue, G. L., Boyd, A. E., and Rosse, W. F. (1970). Chlorpropamide-induced immune hemolytic anemia. *New Engl. J. Med.* 283:900–904.

Loveless, S. E., Ladics, G. S., Gerberick, G. F., Ryan, C. A., Basketter, D. A., Scholes, E. W., House, R. V., Hilton, J., Dearman, R. J., and Kimber, I. (1996). Further evaluation of the local lymph node assay in the final phase of an international collaborative trial. *Toxicology* 108:141–152.

Luebke, R., House, R., and Kimber, I. (2007) *Immuntoxicology and Immunopharmacology*, 3rd ed. CRC Press, Boca Raton, FL.

Luster, M. L., Blank, J. A., and Dean, J. H. (1987). Molecular and cellular basis of chemically induced immunotoxicity. *Ann. Rev. Pharmacol. Toxicol.* 27:23–49.

Luster, M. L., Boorman, G. A., Dean, J. H., Luebke, R. W., and Lawson, L. D. (1980). The effect of adult exposure to diethylstilbestrol in the mouse. Alterations in immunological function. *J. Reticuloendothel. Soc.* 28:561–569.

Luster, M. L., Hayes, H. T., Korach, K., Tucker, A. N., and Dean, J. H. (1984). Estrogen immunosuppression is regulated through estrogenic responses in the thymus. *J. Immunol.* 133:110–116.

Luster, M. L., Munson, A. E., Thomas, P. T., Holsapple, M. P., Fenters, J. D., White, K. L., Jr., Lauer, L. D., Germolec, D. R., Rosenthal, G. J., and Dean, J. H. (1988). Development of a testing battery to assess chemical-induced immunotoxicity: National Toxicology Program's guidelines for immunotoxicity evaluation in mice. *Fund. Appl. Toxicol.* 10:2–19.

Luster, M. L., Pait, D. G., Portier, C., Rosenthal, G. J., Dermolec, D. R., Comment, C. E., Munson, A. E., White, K., and Pollock, P. (1992a). Qualitative and quantitative experimental models to aid in risk assessment for immunotoxicology. *Toxicol. Lett.* 64/65:71–78.

Luster, M. L., Pfeifer, R. W., and Tucher, A. N. (1985). Influence of sex hormones on immunoregulation with specific reference to natural and synthetic estrogens. In J. A. McLachlin, K. Korach, and J. Thomas (Eds.), *Endocrine Toxicology*. Raven, New York, pp. 67–83.

Luster, M. L., Portier, C., Pait, D. G., White, K. L., Jr., Gennings, C., Munson, A. E., and Rosenthal, G. J. (1992b). Risk assessment in immunotoxicology. I. Sensitivity and predictability of immune tests. *Fund. Appl. Toxicol.* 18:200–210.

Mach, P. S., Brouilhet, H., and Smor, B. (1986). D-Penicillamine: A modulator of anti-DNA antibody production. *Clin. Exp. lmmunol.* 63:41418.

Magnusson, B., and Kligman, A. M. (1969). The identification of contact allergens by animal assay. *J. Invest. Dermatol.* 52:268–276.

Male, D., Champion, B., and Cooke, A. (1987). *Advanced Immunology*. J. B. Lippincott, Philadelphia.

Matori, Y. L., Chang, A. E., Lipford, E. H., Braziel, R., Hyatt, C. L., McDonald, H. D., and Rosenberg, S. A. (1985). Toxicity of recombinant human interleukin-2 in rats following intravenous infusion. *J. Biol. Response Mod.* 4:377–390.

Merluzzi, V. J. (1985). Comparison of murine lymphokine, activated killer cells, natural killer cells, and cytotoxic T lymphocytes. *Cell Immunol.* 95:95–104.

Morahan, P. S., Klykken, P. C., Smith, S. H., Harris, L. S., and Munson, A. (1979). *Infect. lmmunol.* 53:670–674.

Ntoso, K. A., Tomaszewski, J. E., Jimenez, S. A., and Neilson, E. G. (1986). Penicillamine-induced rapidly progressive glomerulonephritis in patients with progressive systemic sclerosis: Successful treatment of two patients and a review of the literature. *Am. J. Kidney Dis.* 8:159–163.

Paranjpe, M. S., and Boone, C. W. (1972). Delayed hypersensitivity to simianvirus 40 tumor cells in BALB/c mice demonstrated by a radioisotopic footpad assay. *J. Natl. Cancer. lnst.* 48:563.

Pastan, L., Willingham, M. C., and FitzGerald, D. J. (1986). Immunotoxins. *Cell* 47:64148.

Pattels, H.-G., and Taylor, J. (2008). Immunotoxicity testing: ICH guidelines S8 and related aspects. In S. C. Gad (Ed.), *Preclinical Development Handbook: Toxicology*. Wiley, New York, pp. 323–362.

Patterson, R., De Ewarte, R. D., Greenberger, P. A., and Grammer, L. C. (1986). Drug allergies and protocols for management of drug allergies. *NER Allergy Proc.* 7:325–242.

Penn, I. (1977). Development of cancer as a complication of clinical transplantation. *Transplant. Proc.* 9:1121–1127.

Perry, H. M., Tane, M., and Camody, S. (1970). Relationship of acetyl transferase activity to antinuclear antibodies and toxic symptoms in hypertensive patients treated with hydralazine. *J. Lab. Clin. Med.* 76:114–125.

Pieters, R. (2001). The popliteal lymph node assay: A tool for predicting drug allergies. *Toxicology* 158:65–69.

Pung, O. J., Luster, M. L, Hayes, H. T., and Rader, J. (1984). Influence of steroidal and nonsteroidal sex hormones on host resistance in the mouse: Increased susceptibility to *Listeria monocytogenes* following exposure to estrogenic hormones. *Infect. lmmun.* 46:301–307.

Pung, O. J., Tucker, A. N., Yore, S. J., and Luster, M. I. (1985). Influence of estrogen on host resistance: Increased susceptibility of mice to *Listeria monocytogenes* correlates with depressed production of interleukin 2. *Infect. Immun.* 50:91–96.

Riccardi, C., Puccetti, P., Santoni, A., and Herberman, R. B. (1979). Rapid *in vivo* assay of mouse natural killer cell activity. *J. Natl. Cancer Inst.* 63:1041–1045.

Roitt, I. M., Brostoff, J., and Male, D. K. (1985). *Immunology*. C. V. Mosby, St. Louis, 1985.

Russel, A. S. (1981). Drug-induced autoimmune disease. *Clin. Immun. Allergy* 1:57.

Shand, F. L. (1979). Review/commentary: The immunopharmacology of cyclophosphamide. *Int. J. Immunopharm.* 1:165–171.

Spears, C. J., and Batchelor, J. R. (1987). Drug-induced autoimmune disease. *Adv. Nephrol.* 16:219–230.

Spreafico, F. (1988). Immunotoxicology in 1987: Problems and challenges. *Fund. Clin. Pharmacol.* 2:353–367.

Spreafico, F., and Anaclerio, A. (1977). Immunosuppressive agents. In J. Hadden, R. Coffey, and R. Spreafico (Eds.), *Immunopharmacology*, Vol. 3. Plenum Medical Book Company, New York, pp. 245–278.

Steele, R. W., Williams, L. W., and Beck, S. A. (1989). Immunotoxicology. *Ann. Allergy* 63:168–174.

Sullivan, J. (1989). Immunological alterations and chemical exposure. *Clin. Toxicol.* 27:311–343.

Tabo, J. D., and Paul, W. E. (1973). Functional heterogeneity of murine lymphoid cells. III. Differential responsiveness of T cells to phytohemagglutinin and concanavalin A for T cell subsets. *J. Immunol.* 110:362–369.

Talal, N. (1987). Autoimmune mechanisms in patients and animal models. *Toxicol. Pathol.* 15:272–275.

Thiem, P. A., Halper, L. K., and Bloom, J. C. (1988). Techniques for assessing canine mononuclear phagocyte function as part of an immunotoxicologic evaluation. *Int. J. Immunopharm.* 10:765–771.

Ueda, S., Wakahim, Y., Takei, L, Mori, T., and Lesato, K. (1980). Autologous immune complex nephritis in gold injected guinea pigs. *Nippon Jinzo Gakkai Shi* 22:1221–1230.

Unkovic, J., Combes, M., Mazue, G., and Roncucci, R. (1984). Poster. Presented at the Annual Meeting of the American Society of Dermatology, Washington, DC.

U.S. Food and Drug Administration (FDA), Center for Safety and Applied Nutrition (1993). Draft: Toxicological principles for the safety assessment of direct food additives and color additives used in food. *Fed. Reg.* 58:10536.

U.S. Food and Drug Administration (FDA) (2006). ICH S8: Guidance for industry: Immunotoxicity studies for human pharmaceuticals, CBER. FDA, Washington, DC.

Vohr, H.-W. (2005). *Encyclopedia Reference of Immuno-Toxicology*. Springer, New York.

Volkman, A. (1984). *Mononuclear Phagocyte Function*. Marcel Dekker, New York.

Vos, J., Van Loveren, H., Wester, P., and Vethaak, D. (1989). Toxic effects of environmental chemicals on the immune system. *TIPS* 10:289–292.

Vos, J. G. (1977). Immune suppression as related to toxicology. *CRC Crit. Rev. Toxicol.* 5:67.

Wallner, B. P., Mattaliano, R. J., Hession, C., Cate, R. L., Tizard, R., Sinclair, L. K., Foeller, C., Chow, E. P., Browning, J. L., Ramachandran, K. L., and Pepinsky, R. B. (1986). Cloning and expression of human lipocortin, a phospholipase A2 inhibitor with potential anti-inflammatory activity. *Nature* 320:77–80.

Ways, S. C., Mortola, J. F., Zvaifler, N. J., Weiss, R. J., and Yen, S. S. C. (1987). Alterations in immune responsiveness in women exposed to diethylstilbestrol in utero. *Fertil. Steril.* 48:193–197.

Webb, D. R., and Winkelstein, A. (1982). Immunosuppression, immunopotentiation and anti-inflammatory drugs. In D. P. Stites, J. D. Stobo, H. H. Fudenberg, and J. V. Wells (Eds.), *Basic and Clinical Immunology*, 4th ed. Lange Medical, Los Altos, CA, pp. 277–292.

Weening, J. J., Fleuren, G. J., and Hoedemaeker, J. (1978). Demonstration of antinuclear antibodies in mercuric chloride-induced glomerulopathy in the rat. *Lab. Invest.* 39:405–411.

Weigle, W. O. (1980). Analysis of autoimmunity through experimental models of thyroiditis and allergic encephalomyelitis. *Adv. Immunol.* 30:159–275.

White, K. L., Jr., Sanders, V. M., Barnes, D. W., Shopp, J. G. M., and Munson, A. E. (1985). Immunotoxicological investigations in the mouse: General approach and methods. *Drug. Chem. Toxicol.* 8:299–331.

Wierda, D., Smith, H. W., and Zevieki, C. M. (2001). Immunogenicity of biopharmaceuticals in laboratory animals. *Toxicology* 58:71–74.

Yoshida, S., Golub, M. S., and Gershwin, M. E. (1989). Immunological aspects of toxicology: Premises not promises. *Reg. Toxicol. Pharm.* 9:56–80.

# *Nonrodent Animal Studies*

## 1 INTRODUCTION

Most safety assessment studies are conducted in rodents (rats, mice, and hamsters) or their close "cousins" rabbits and guinea pigs. Outside of the pharmaceutical, medical device, and veterinary product industries, it has become rare for the practicing toxicologist to have close familiarity with the nonrodent animal species addressed in this chapter. Indeed, it is unlikely that a toxicologist has received any significant academic experience or training with these species. Additionally, use of these species in the European Union (EU) even for therapeutics has become very difficult at best. However, the proper use of nonrodent species is essential in the evaluation of potential new therapeutic entities, on both scientific and regulatory grounds. Indeed, there are numerous studies showing significantly better concordance between humans and nonrodents than humans and rodents for detection of adverse responses to pharmaceuticals (Olson et al., 2000). This is tacitly recognized in regulatory practice for in those new product development plans where a single species is deemed appropriate and sufficient (medical devices, protein therapeutic, and 505(b)(2) approval candidates), the single species is overwhelming a nonrodent.

In addition to rodent studies, regulatory guidelines for pharmaceuticals require that repeated-dose safety studies of up to nine months (in the United States, six months elsewhere) in duration be conducted in a nonrodent species. The most commonly used nonrodent species is the dog followed by an NHP

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

(nonhuman primate) and then the pig. Another nonrodent model used to a limited extent in systemic safety evaluation is the ferret. The major objectives of this chapter are (1) to discuss differences in rodent and nonrodent experimental design, (2) to examine the feasibility of using the dog, NHP, pig, and ferret in safety assessment testing, and (3) to identify the advantages and limitations associated with each species.

## 2 COMPARISON BETWEEN RODENT AND NONRODENT EXPERIMENTAL DESIGN: NUMBER OF ANIMALS

One of the main differences in experimental design between rodent and nonrodent safety studies is the number of animals used (Table 1). In general, approximately nine times as many rodents are used in toxicity studies as compared to nonrodents. This difference is reflected in the 2005 estimates of overall usage of animals in the United States (10 million animals total) as published by the National Research Council (NRC, 1998), which showed that only about 10% of the animals used in general research for that year were NHPs (2%), pigs (1%), or dogs (7%). The smaller number of nonrodents used is related in part to the higher costs associated with their purchase, housing, and maintenance and in part to their limited use in other areas of research.

## 3 DIFFERENCES IN STUDY ACTIVITIES

### 3.1 Blood Collection

In rodent studies, large numbers of satellite animals (often close to the number used in the main study phase) are usually needed for pharmacokinetic blood sampling, whereas with most nonrodent species adequate blood samples can be collected from the main study animals without compromising their health status (Diehl et al., 2001).

**TABLE 1 Comparison of Rodent and Nonrodent Experimental Design**

| | Total Numbers of Animals in Study (No./Group/Sex) | | | |
|---|---|---|---|---|
| Duration of Study | Rat | Dog | Pig | Monkey |
| 4 weeks | 360 (20)[a] | 40 (4) | 40 (4) | 32 (4) |
| 13 weeks | 280 (20)[a,b] | 48 (6)[b] | 48 (6)[b] | 48 (6)[b] |
| 52 weeks/9 months | 360 (10)[a,b] | 64 (8)[b] | 64 (8)[b] | 48 (6)[b] |

[a]Includes satellite animals for pharmacokinetic evaluation.
[b]Number of animals/group includes several animals.

### 3.2 Dosing

The most appropriate route of oral administration is probably capsule dosing for dogs and gavage for monkeys and ferrets. Nasal gastrogavage is also occasionally used for dogs and primates.

### 3.3 Handling of Animals

Once rodents are acclimated to handling, they are generally relatively easy to work with. In contrast, some nonrodent species, such as NHPs and pigs, are often difficult to handle because of their size, strength, emotionality, and aggressiveness. This can make the conduct of routine study activities [such as dosing, blood collection, and recording electrocardiograms (ECGs)] relatively time consuming as well as stressful to the animals.

### 3.4 Behavioral Evaluation

Behavioral assessment of nonrodents is generally more difficult than evaluation of rodents because of their larger size, difficulties associated with handling and manipulation, and greater awareness of and reactivity to the experimenter. Such factors can confound detection and/or interpretation of more subtle test compound–related behavioral changes. However, there is now a published and validated functional observations battery (FOB) for neurological exams of dogs and it is widely used (Gad et al., 2003).

## 4 NONRODENT MODELS

This section is devoted to the definition and comparison of the four nonrodent animal models (dog, ferret, pig, and NHP) in terms of experimental procedures, environmental and dietary requirements, as well as advantages and disadvantages of use in safety assessment testing.

### 4.1 Dog

***Environmental and Dietary Requirements*** Typical housing for laboratory dogs consists of stainless steel or fiberglass cages (of dimensions appropriate to the dog's size) or indoor pens (typical dimensions are 3 ft, 8 in. wide, 8 ft high, and 10 ft long). Two important aspects of the laboratory dog's environment are the need for exercise and socialization. Recent amendments to the U.S. Department of Agriculture's (USDA's) Animal Welfare Act require that an exercise and socalization program be established for dogs maintained in a laboratory environment. Difficulty often arises in establishing a program that will be truly beneficial to the animals. One important consideration is whether dogs should be group or individually exercised. Studies have demonstrated that dogs exercised alone tend to spend most of their time walking or inves-

tigating the area rather than jumping or running (Campbell et al., 1988), which suggests that group exercise is more beneficial.

The need for a certain degree of socialization is also important, in terms of both dog–dog and dog–human contact. If at all possible, dogs should share a cage or pen with another animal. One study approach undertaken by some laboratories has been to allow dogs of the same sex and treatment group to have daily contact with each other, usually from early evening to early morning. If study dictates do not make this approach feasible, efforts should be made to ensure the animals are housed in such a way that they have visual, auditory, and olfactory access to each other.

Recommended dry-bulb temperatures and relative humidity ranges for dogs are 64.4–84.2 °F (16–27 °C) and 30–70%, respectively [U.S. Department of Health and Human Services (DHHS), 1985]. Increases in temperature and high humidity are of particular concern because of the dog's limited capacity to dissipate heat (primarily through panting and, to a lesser extent, through radiation and conduction). Dogs would likely not survive exposure for extended periods of time to environments where the temperature is in excess of 40 °C and 40% relative humidity (Norris et al., 1968).

While the dog is a carnivore, it is able to adapt to an omnivorous diet. Requirements for dietary sources of energy, amino acids, glucose precursors, fatty acids, minerals, vitamins, and water have been established based on recommendations by the NRC (1985). Adult beagles maintained in a laboratory environment function well with one feeding of standard laboratory chow per day. In safety assessment testing, however, some compounds may induce serious dietary deficiencies through induced loss of appetite, malabsorption, or vomiting, and, in these cases, it may be advisable to provide a dietary supplement.

The dog's requirement for water appears to be self-regulated and depends on factors such as the type of feed consumed, ambient temperature, amount of exercise, and physiological state; therefore, in most cases, dogs should have free access to water.

***Common Study Protocols*** The dog is the most commonly used nonrodent species in safety assessment testing (i.e., acute, subchronic, and chronic studies and cardiovascular safety pharmacology studies). The exception to this is its use in developmental toxicity and reproductive studies. For developmental toxicity studies, the dog does not appear to be as sensitive an indicator of teratogens as other nonrodent species such as the monkey (Earl et al., 1973) or the ferret (Gulamhusein et al., 1980), and, for reproductive studies, the dog is not the species of choice because fertility testing is difficult to conduct (due to prolonged anestrus and the inability to predict the onset of proestrus) and there is no reliable procedure for induction of estrus or ovulation.

Examples of experimental designs and suggested timing of various study activities for 4- and 13-week dog studies are shown in Tables 2 and 3, respectively. Beagles are generally in the age range of 6–9 months at study start, and

**TABLE 2 Four-Week Dog or Primate Toxicity Study**

| *Experimental Design* |
|---|
| Four to five groups (including a control group)—4/sex/group main groups, plus extra in high-dose and control groups to evaluate recovery |
| Repeated daily dosing for 28 or 29 days |
| Necropsy starting on day 29 |
| *Study Activities* |
| Daily observations: pretreatment and twice daily during study period |
| Physical examinations: pretreatment and after dosing during weeks 2 and 4 |
| ECG: pretreatment and after dosing during weeks 2 and 4 |
| Ophthalmic examinations: pretreatment and during week 4 |
| Body weight: pretreatment, weekly, and prior to scheduled necropsy |
| Feed consumption: pretreatment and weekly |
| Clinical lab: twice before first dosing day, before dosing on day 2, during week 2, and prior to scheduled necropsy |
| Urine collection: pretreatment and during weeks 2 and 4 |
| Pharmacokinetic: blood collected at specified times after dosing on days 1 and 28 |

*Source*: Adapted from a table in *Animal Models in Toxicology* (Gad, 2006).

**TABLE 3 Thirteen-Week Dog Toxicity Study**

| *Experimental Design* |
|---|
| Four groups (including control group)—6/sex/group (plus 3/sex extra in high-dose and control groups for assessing potential recovery) |
| Repeated daily dosing for 91–93 days |
| Necropsy of main group (4/sex/group) on week 14 |
| Necropsy of reversal group (2/sex/group) on week 18 |
| *Study Activities* |
| Daily observations: pretreatment, twice daily during treatment, and once daily during reversal |
| Physical examinations: pretreatment, after dosing during weeks 4, 8, and 13 of treatment, and during week 4 of reversal |
| ECG: pretreatment, after dosing during weeks 4, 8, and 13 of treatment, and during week 4 of reversal |
| Ophthalmic examinations: pretreatment, during weeks 6 and 13 of treatment and during week 4 of reversal |
| Body weight: pretreatment (3 times), weekly during treatment and reversal periods, and prior to scheduled necropsy |
| Feed consumption: pretreatment, weekly through first month, bimonthly during remainder of treatment period, and weekly during reversal |
| Clinical lab: pretreatment, during weeks 4 and 8 of treatment, prior to scheduled necropsy, and during weeks 1 and 4 of reversal |
| Urine collection: pretreatment, monthly during treatment, and during week 4 of reversal. |
| Pharmacokinetic samples: blood collected at specified times after dosing on day 1 and during weeks 6 and 12 |

*Source*: Adapted from a table in *Animal Models in Toxicology* (Gad, 2006).

the number of animals per sex per treatment group (*N*) will depend on the duration of the study. For a 2–4-week study without a reversal phase, *N* will likely be 4, whereas for a 26-week or 1-year study, *N* will be larger (e.g., *N* may be 9 including 3 per sex per group for the reversal phase); it should be noted that it has become a U.S. Food and Drug Administration (FDA) expectation for a high-dose and control group to include animals to allow the assessment of recovery.

Dogs should be selected for study use on the basis of acceptable body weights, urinalysis, and clinical pathology findings as well as physical, ophthalmic, and electrocardiographic evaluations. To minimize familial effects, efforts should be made to ensure that no two littermates of the same sex are assigned to the same treatment group.

Because most, if not all, study-related activities are conducted in the same dogs, the stress induced by repeated manipulation of dogs for activities such as blood collection, ECG, and physical examinations needs to be taken into consideration. Efforts should be made wherever possible to separate study activities by several days.

### General Study Activities

*Dosing Techniques* The most frequently used route of administration in dog safety assessment studies is oral. Dosing by capsule is usually the preferred oral route in the dog. Gavage is also used, but it is a more labor intensive technique, and there is always the possibility of gavage error or aspiration. Since dogs have a natural tendency to vomit, it is recommended that they be sham dosed with empty capsules or gavaged with a water solution for several days prior to starting a study so that they can become acclimated to the dosing procedure. If the test substance in irritating (and therefore proemetic), the use of enterically coated capsules should be considered. Attention should be paid, however, to the time course of emesis. If it follows dose administration by more than 4–6 h, the cause will not have been due to gastrointestinal (GI) tract irritation, as the material will have already moved from the stomach.

Next to oral dosing, the most common dosing route for dogs is intravenous. For bolus or limited infusion intravenous dosing, the femoral, cephalic, and saphenous veins are commonly used. For continuous infusion, the jugular is often the vein of choice, and the procedure will require surgical preparation either for a direct line catheterization or subcutaneous insertion of a vascular access port (a rigid, multipuncturable reservoir equipped with an indwelling catheter).

Other routes of administration used less commonly in dog safety studies are subcutaneous, intramuscular, intraperitoneal, rectal, and vaginal and dermal.

*Clinical Observations and Physical Examinations* Daily clinical observations in dog safety studies, usually conducted pretreatment (prior to cage

cleaning) and at a specified time(s) after dosing, consist of a home cage observation with notation of clinical signs indicative of poor health (such as salivation, weight loss, abnormal feces, and vomitus) or abnormal behavior (such as reduced activity or increased aggression).

Physical examinations are conducted less frequently and generally involve the evaluation of gait, mobility, demeanor, and reflexes (pupillary light, corneal patellar, wheelbarrowing, hopping, etc.) as well as an examination of the head (eyes, ears, mouth, teeth, gums, and tongue), body (palpation for signs of masses and nodal swellings), and urogenital and anal regions.

*Feed Consumption* Feed consumption is relatively easy to measure in the dog since dogs do not usually spill much of their feed. Generally, the full feed bowl is weighed at the beginning and the empty bowl at the end of the feeding period (usually a 4-h period). This is repeated over two or three consecutive days and the average daily feed consumption is calculated from the numbers.

*Electrocardiograms* The collection of recorded ECGs in test dogs has become increasingly of interest due to the concurrence to the QT prolonguative by drugs and International Conference on Harmonisation (ICH)/FDA/European Medicines Agency (EMEA) mandated evaluation studies. Traditionally, a 10-lead system consisting of bipolar leads (I, II, III), augmented unipolar leads (aVR, aVL, and aVF), and unipolar precordial leads [V10, CV6LL(V2), CV6LU(V4), CV5RL(rV2)] has been recommended for dogs in the conscious state (Detweiler et al., 1979; Detweiler, 1980). For toxicity studies, dog ECGs are usually recorded by technical personnel and read at a later time by a veterinary cardiologist. Depending on the length of the study and the pharmacological–toxicological profile of the test compound, ECGs may be recorded as frequently as every day or as infrequently as every three months. Dog ECGs can also be highly variable. Factors that can affect the quality of the tracing include the positioning of the electrodes, the positioning of the dog, and the degree of nervousness and excitability of the animal. Conditioning the dogs to the electrode clips and the recording position (usually sphinx or right-lateral recumbency) during the pretreatment period will help improve the quality of the recording.

The use of surgically implanted sensors to be able to remotely monitor electrophysiology, blood pressure, and blood gases has become very common for cardiovascular safety pharmacology studies. It should always be the case using such dogs that a concurrent evaluation of baseline and vehicle effects be performed prior to that of acute drug grounds.

*Blood and Urine Collection* As mentioned previously, serial blood samples can be fairly easily collected from the dog. The jugular vein is probably the most commonly used vein because of its size and accessibility. Other veins used less frequently are the cephalic, femoral, brachial, and saphenous.

Due to the difficulty in obtaining sufficient volumes of urine in dogs over short collection periods, urine is usually collected overnight (approximately a 16–17-h period) in stainless steel metabolism cages. It is recommended that a sample for urinalysis be taken early in the collection process and that all samples be collected in light-resistant containers to help avoid problems such as dissolution of urine casts, increased bacterial activity, and breakdown of bilirubin with exposure of the sample to light.

***Advantages and Disadvantages*** Some of the advantages and disadvantages of using the dog in safety assessment studies are listed in Table 4. With respect to its medium size and even temperament, the beagle is certainly a desirable nonrodent model. The relative ease in handling beagles makes them suitable for activities such as serial collection of blood samples and recording of electrocardiograms.

Disadvantages include an often wide variation in size and body weight and a loud, penetrating bark. The large amount of space required to house dogs and the current emphasis on regular exercise may also be disadvantages. Test compound requirements are generally higher for the dog than for either the NHP or ferret when these are alternative nonrodents species used in safety testing. This may be a problem in the early period of drug development when compound availability is often limited. Other problems center around the dog's tendency to vomit, which can be a disadvantage when compounds are orally administered, and the fact that, unlike rodents, studies requiring large numbers of dogs need careful advance planning to ensure that sufficient numbers of animals of the appropriate age can be obtained in a timely manner.

## 4.2 Ferret

The ferret, *Mustela putorius furo*, is a small carnivore that has become an increasingly popular species in various areas or research, including anatomy,

**TABLE 4 Use of Beagle in Safety Assessment Studies**

| *Advantages* |
|---|
| Medium size |
| Moderate length of hair coat |
| Adaptability to living in group housing |
| Ease of handling (e.g., dosing, blood collection, ECG) |
| *Disadvantages* |
| Variation in size and body weight |
| Loud, penetrating bark |
| Greater test compound requirements than smaller nonrodent species |
| Availability |
| Exercise and housing requirements |

*Source*: Adapted from a table in *Animal Models in Toxicology* (Gad, 2006).

virology, bacteriology, physiology (GI, pulmonary, and cardiovascular), pharmacology, neurology, teratology, and, to some extent, toxicology. The reader is referred to the excellent review by Fox (1988) on the biology and diseases of the ferret and to the chapter in Gad (2006) or the recent paper by Gad (2000a) on the ferret as an animal model in toxicology. Since 1990, the literature has reported on work done by Pfizer, Hoffman LaRoche, Gilead, Bristol Myers Squibb, Merck, Yamanouchi, Proctor and Gamble, Abbot, and Glaxo Wellcome using ferrets in pharmaceutical development. The ferret is also the species of choice for respiratory virus (especially influenza) efficacy studies.

***Environmental and Dietary Requirements*** For reasons of environmental control, ferrets used in safety assessment studies should be housed indoors. It has been suggested that an optimal temperature range for the ferret is 40–65 °F (4–18 °C), while relative humidity should be maintained in the range of 40–65% (Fox, 1988). The ferret does not tolerate heat well due to its lack of well-developed sweat glands; the primary method of regulating heat loss appears to be through panting (Moody et al., 1985).

Since ferrets are seasonal breeders, the female being monestrus and an induced ovulator, the breeding cycle can be controlled by varying the length of exposure to artificial light. For safety studies, it is desirable to prevent both estrus in females and increased sexual activity in males; thus it has been recommended that the light period be kept short (Fox, 1988). In this laboratory, a 9-h light–15-h dark cycle has been used successfully for this purpose.

Ferrets should be housed in well-ventilated rooms that provide at least 10–15 air changes per hour. Good ventilation is important since ferrets are susceptible to respiratory viral infections. Additionally, there is a need to dissipate the musky odor of the animals. While housing standards for ferrets are not specified in the NRC (2004) *Guide to the Care and Use of Laboratory Animals*, space requirements of 49 × 46 × 46 cm have been defined by other groups (Wilson and Donnoghue, 1982). Stainless steel cat or rabbit cages equipped with a drop pan to catch feces and urine are a suitable form of primary housing for ferrets. Ferrets are more content when they have access to a small secluded nesting area within their cage in which they can sleep, such as hammocks. The use of paper to line the cage or drop pan is not recommended, since the ferrets are likely to eat it. For socialization purposes, ferrets should be housed as a group or have visual access to neighboring ferrets if housed individually.

Since ferrets eat only their caloric requirements, and since their GI transit time is short (3–5 h), it is recommended that they receive diet ad libitum. Only one of the available standardized ferret chows commercially available should be fed. The most important dietary variable is the quality of the protein, and ferrets appear to do best with a high percentage of animal protein in their diet (Morton and Morton, 1985). Feed consumption will be higher in the fall and

winter and lower in the spring and summer. Hairball laxative is essential during the spring and summer months when the animals experience considerable hair loss. Water should be available at all times.

***Study Protocols*** Historically, the ferret was used more often in teratology (Hoar, 1984), reproductive (Hoar, 1984), and acute safety studies than in repeated-dose studies (4–52 weeks in duration). This has changed since 2000, however, and the use of the ferret in pivotal repeated-dose safety assessment testing (Thornton et al., 1979; Hart, 1986; Haggerty et al., 1989) has increased, particularly for vaccines.

An example of the experimental design for a 4-week pivotal study in ferrets is shown in Table 5. Young adult ferrets are usually in the age range of 9–11 months at study start, and there should be sufficient numbers of animals in each group for statistical confidence (generally in the range of six to eight animals per sex per group reflecting the high degree of intergroup variability in the species). For longer term studies, the number of animals per group would be increased to include reversal group animals (three to four animals per sex per dose group).

Assignment of ferrets to a study should be based on evaluation of pretreatment clinical signs and body weights as well as physical, electrocardiographic, and opthalmological findings.

For longer term studies, females should be spayed to avoid the development of aplastic anemia, which will occur if the animals go into heat and are not bred (Morton and Morton, 1985).

As with dogs, efforts should be made to separate study activities as much as possible to minimize the stress of multiple activities being performed in the same animals.

**TABLE 5 Four-Week Ferret Toxicity Study**

| *Experimental Design* |
|---|
| Five groups (including control group)—7/sex/group |
| Repeated daily dosing for 28 or 29 days |
| Necropsy starting on day 29 |
| *Study Activities* |
| Daily observations: pretreatment and twice daily during study period |
| Physical examinations: pretreatment and after dosing during weeks 1, 2, and 4 |
| ECG: pretreatment and after dosing during weeks 2, and 4 |
| Ophthalmic examinations: pretreatment and during week 4 |
| Body weight: pretreatment, twice weekly, and prior to scheduled necropsy |
| Feed consumption: pretreatment and weekly during the study |
| Clinical lab: pretreatment, week 2, and prior to scheduled necropsy |
| Urine collection: pretreatment and during weeks 2 and 4 |
| Pharmacokinetic samples: blood collected at specified times after dosing on days 1 and 28 |

### General Study Activities

*Dosing Techniques* Oral dosing of ferrets is usually done by gavage. One method is to hold the ferret perpendicular to the floor and insert the appropriate size stainless steel gavage needle into the animal's mouth, back into the esophagus, and down toward the stomach. Confirmation of correct positioning of the tube can be determined by visual inspection of the aspirate. As with dogs, ferrets have a tendency to retch or vomit, and daily gavaging with a water solution for several days prior to starting a study (for adaptation purposes) is recommended.

Repeated daily intravenous dosing in the ferret is generally considered to be technically difficult and time consuming; the use of an indwelling catheter is recommended (Moody et al., 1985). There are, however, reports in the literature of subchronic intravenous dosing (three times weekly for three months) of the ferret via the caudal vein (Mclain et al., 1987).

Dosing techniques such as intramuscular, intradermal, subcutaneous, and intraperitoneal administration can be used for the ferret. Care needs to be taken, however, when administering lipophilic compounds by the subcutaneous or intradermal routes to avoid inadvertently injecting compounds into the ferret's thick layer of subcutaneous fat, which can result in poor absorption (Moody et al., 1985).

*Clinical Observations and Examinations* Daily clinical observations will usually begin the week prior to study start and continue twice daily (pre- and postdosing) throughout the study. Ferrets are observed in their home cage for signs of physical debilitation (such as abnormal feces or vomitus), behavioral abnormalities, hair loss, swelling of the vulva (females), and testicular prominence (males). A physical examination should periodically be made and should include measurement of rectal temperature; observation of general demeanor and activity; palpation of the head, thorax, and abdomen; examination of eyes, ears, and body orifices; and testing of the pupillary and patellar reflexes.

*Feed Consumption* Feed consumption can be measured over two to three consecutive days and the average daily intake calculated. A problem with measuring feed intake in ferrets is their tendency to dig through their feed bowl, which often results in an unacceptable amount of spillage. Use of a feed follower may help reduce the spillage.

*Electrocardiograms* Most electrocardiographic evaluation in the ferret has been previously done in the anesthetized animal, though sensors may be surgically implanted. This allows electrocardiograms to be recorded using the limb (I, II, and III) and augmented (aVR, aVL, and aVF) leads. It is quite possible to obtain fairly good quality ECGs in the conscious ferret using leads I, II, and III. The standard position used for recording ECGs in the conscious or anesthetized ferret is right-lateral incumbency. ECGs have also been measured in

the ferret using surface electrodes placed between two points on the chest, with the signals being led off to an amplifier by a long, flexible cable and recorded on magnetic tape for later analysis (Andrews et al., 1979). The advantage of such a system is that the animals are allowed to move freely during the recording.

*Blood and Urine Collection* About 5–10 mL of blood can be collected from adult ferrets using retroorbital blood collection techniques. Other methods of blood collection include cephalic and jugular veins and caudal tail venipuncture as well as bleeding via the ventral tail artery. Cardiac puncture is also used, but in the opinion of this and other (Hart, 1986) laboratories, the procedure is traumatic and can cause myocardial scarring. Blood collection from the tail can be difficult because the ferret tail is short and the tail veins and arteries cannot be seen. For all the above-mentioned collection techniques, some form of pharmacological or mechanical restraint is required. To facilitate serial blood collection, methodology has been developed for a tethered restraint system with an implanted indwelling venous jugular catheter, which does not interfere with the normal activities of the ferrets and allows blood sampling to occur from outside the cage (Jackson et al., 1988).

For urine collection, glass or plastic rat metabolism cages work well for short-term or overnight collection. Care needs to be taken to avoid contamination of the urine with feces.

***Advantages and Disadvantages*** Two advantages to using the ferret are its cost and its size (Table 6). The cost of the ferret is approximately one-tenth that of the dog. The ferret's smaller size means that it is easier to maintain and more economically housed and fed than the dog (Hart, 1986). Smaller size also means that test compound requirements for the ferret will be considerably less than those for larger nonrodent species (e.g., on the order of one-tenth of that needed for the dog). Another advantage is that if exercise requirements are

**TABLE 6 Use of Ferret in Safety Assessment Studies**

| *Advantages* |
|---|
| Small size |
| Significantly lower cost than most other nonrodents |
| Lower test material requirements (relative to larger nonrodents) |
| Adaptability to exercise program |
| *Disadvantages* |
| Pervading musky odor |
| Rodents' inherent fear of ferrets |
| Can be difficult to handle |
| Background disease profile with resulting "background noise" and increased variability in clinical and anatomical pathology |

ever established for the ferret, it will be an easier species than the dog for which to design an acceptable exercise program.

Disadvantages associated with the ferret include its pervading musky odor and its background disease profile. While the scent glands can be removed, 90% of the animal's odor is derived from sebaceous secretions onto the skin. However, neutering the males and spaying the females, in addition to descenting, will markedly reduce the odor. Rats should be housed as far away from ferrets as possible because of their inherent fear of ferrets (triggered by olfactory stimulation), which can interrupt breeding cycles or disturb other physiological functions (Fox, 1988). In this laboratory's experience, ferrets, which are generally less docile than dogs, can be difficult to handle and prone to bite, especially when restrained for activities such as ophthalmic and ECG examinations. The lack of easily accessible veins for intravenous dosing and serial blood collection is also a disadvantage.

The major disadvantage in the use of the ferret in safety studies is the profile of diseases associated with the species and the resulting variability in background clinical and anatomical pathology. Pneumonitis and hepatic lymphoid accumulation, associated with chronic parvovirus infection, have been observed in ferrets in this laboratory (Haggerty et al., 1989). Submucosal lymphoid nodules of the intestines are also a common finding (Hart, 1986). Additionally, a relatively high incidence of electrocardiographic (atrial or ventricular premature depolarization, atrial and ventricular extrasystoles) and ophthalmological (optic nerve hypoplasia and cataracts) anomalies have been found in ferrets in this laboratory. While it may be possible to work with the animal suppliers to reduce the chances of receiving animals with background ECG or ocular abnormalities, at the present time there is no supplier of a disease- and viral-free ferret.

### 4.3 Pig

***Background*** The use of pigs (*Sus scrufa*) in biomedical research is well established. In toxicology, whereas the use of pigs in the United States is largely limited to dermal studies, in Europe they have become very popular for pharmaceutical studies in place of dogs and primates. They have been extensively used for surgical (Swindle et al., 1988) and physiological (primarily cardiovascular, renal, and digestive) research (Khan, 1984; Clausing et al., 1986) for years. Until relatively recently, their use in toxicity testing was uncommon except in the testing of veterinary or herd management drugs intended for use in swine or in dermal toxicity and absorption studies. Because of their well-accepted physiological similarities to humans, minipigs are becoming increasingly attractive toxicological models (Table 7). In fact, they are already more frequently used in nutritional toxicology studies (Clausing et al., 1986). Among the more common experimental animals, pigs are the only ones whose use is on the increase (Khan, 1984). Their expense (both in procurement and maintenance) and their relatively large size have mitigated against their

**TABLE 7 Minipig in Toxicity Testing**

| |
|---|
| Due to many advantages, mini- and micropigs are real alternatives to use of nonrodents (dogs, ferrets, and primates) |
| Minnesota minipig introduced in 1949 |
| Body weights at age 2 years |
| Yucatan minipig: 70–90 kg |
| Yucatan micropig 40–45 kg |
| Göttinger micropig: 35–40 kg |
| Use in general toxicity testing and reproduction, teratological and behavioral toxicity (aspects of public acceptance as a species for testing) |

**TABLE 8 Main Advantages of Minipig**

| |
|---|
| Similarity to humans in |
| • Cardiovascular anatomy and physiology |
| Ventricular performance |
| Electrophysiology |
| Coronary artery distribution |
| • Human skin |
| Thickness and permeability |
| Pigmentation |
| Allergic reaction |
| Reaction to burning and distress |
| • Gastrointestinal system and digestion |
| • Renal system |
| • Immune system (FDA: "better than rodents") |
| • P450 total enzyme activity (especially CYP2E1, CYP3A4) |

use in more general toxicity testing. The development of minipigs has resulted in a strain of more manageable size. In addition, the increase in expense in the use of dogs as well as the perceived lay opposition to their uses make minipigs (in Europe and Israel) even more attractive as a nonrodent species for general toxicity studies. The dog is a far more common companion animal and many of the recent developments in animal care and use laws have made specific provisions about the care of dogs. Minipigs have been shown to be more sensitive to a wide variety of drugs and chemicals (e.g., carbaryl, methylmercury) than dogs (Khan, 1984). The FDA has kept its own breeding colony of minipigs since the early 1960s. In short, there are scientific, economic, and sociological reasons that make minipigs good toxicological models. The reader is referred to an excellent short review by Phillips and Tumbleson (1986) that puts the issue of minipigs in biomedical research into the context of modeling in general. Table 8 presents the advantages of the minipig.

Several breeds of miniature swine have been developed. In the United States, these include the Yucatan micro- and minipigs, the Handford, the Sinclair, the Pitman-Moore, and the Hormel. The Yucatan and the Sinclair tend to be the most commonly used, though the Göttinger (widely used in Europe) is seeing increasing use (Ellegaard et al., 1995). Panepinto and Phillips (1986) have discussed the characteristics, advantages, and disadvantages of the

Yucatan minipig in some detail. In Europe, the Göttinger minipig is extensively used. At sexual maturity (4–6 months) the typical minipig weighs 20–40 kg, as compared to 102 kg for the more common pig, 8–15 kg for the dog. Micropigs weigh about 14–20 kg at sexual maturity. The minipig and the dog have comparable life spans; for example, Peggins et al. (1984) reported that the average life span for miniature swine is 15–17 years. The average beagle dog may have a life span of 8–12 years. Most of this discussion will focus on the purpose-bred minipigs, primarily the Yucatan and the Sinclair.

The greatest area of use of the pig in pharmaceutical safety assessment is for dermal agents. It should be noted that there are differences in skin thicknesses in different species, and these differences are not well characterized.

### *Husbandry*

*Housing* A general review of handling and husbandry is given by Panepinto (1986) and Swindle et al. (1988). Young weanling pigs can be kept for short periods of time (up to one month) in standard dog cages with the floor modified with narrow mesh to account for the smaller foot of the pig. After that, however, their rapid growth generally makes such caging inappropriate. Larger stainless steel cages would be extremely expensive. Standard dog runs could have enough floor space to be converted for pigs, but smooth flooring does not provide appropriate footing for pigs and needs to be covered with wood chip bedding (Swindle et al., 1988). Although pigs are very social, they do not have to be group housed, as discussed by Barnett and Hensworth (1986); individually housed swine show little evidence of a chronic stress response. Insufficient space, on the other hand, can cause chronic stress in pigs. Hunsaker et al. (1984) have described an inexpensive caging system for miniature swine that is appropriate for toxicology studies. The flooring and walls are constructed of 0.50-cm welded wire coated with polyvinyl chloride polymer. As described, the unit has sufficient room for two pigs separated by a partition. These units are relatively inexpensive and provide more than sufficient floor space (about 17 $ft^2$ per pig) to meet the recommendations for pigs.

*Water and Feed* Like all animals, pigs should be permitted free access to potable water, preferably from a municipal water supply intended for human consumption. Drinking water intended for pigs does not have to be filtered or deionized. Various diets have been described. Because of their size (i.e., high maintenance charges and test article demands), pigs have seldom been used for chronic studies where the possibility of waterborne environmental contaminants could influence a study.

For miniature swine, the consistent use of a certified chow from a major manufacturer is recommended (Swindle et al., 1988). Free access to feed is not recommended as pigs will eat to excess and grow quite large. Available feed should be restricted to approximately 4% of body weight per day to prevent the animals from becoming obese.

*Restraint and Dosing* In general, minipigs are docile and easily socialized and trained. Barnett and Hensworth (1986) recommended a socialization regimen of 2 min of gentle interaction (e.g., striking). Pigs, like most experimental animals, are rarely simply kept and fed but have to be occasionally restrained so samples can be taken and other measurements made. Restraint methods designed for commercial swine should not be used for laboratory swine. Panepinto (1986) have described a sling method that provides restraint with minimal stress. The most frequently mentioned dosing routes in the literature are dietary admix, dermal (topical), gavage, and intravenous injections. Generally, minipigs are restrained in a sling while beginning dosing by the active route such as gavage. If the experiment requires the implantation of, for example, an indwelling catheter, minipigs can be anesthetized with ketamine ($20\,mg\,kg^{-1}$ IM) as described by Swindle et al. (1988).

***Clinical Laboratory*** Clinical chemical and hematological parameters for minipigs have been studied. Ranges for some of the more commonly examined parameters from Yucatan minipigs are summarized in Tables 9 and 10 (from Radin et al., 1986). Parsons and Wells (1986) have published similar data on the Yucatan minipig. Brechbuler et al. (1984), Oldigs (1986), Ellegaard et al. (1995), and Koch et al. (2001) have published on the Göttinger minipig. Middleton and co-workers have published extensive lists (organized by age and sex) on the hematological parameters (Burks et al., 1977) and serum electrolytes (Hutcheson et al., 1979) for the Sinclair minipig. In general, the clinical laboratory picture of the various strains are quite similar. No real

**TABLE 9 Minipig Clinical Chemistry Parameters in Different Strains**

| Parameter | Yucatan | Göttinger |
|---|---|---|
| Glucose ($mmol\,L^{-1}$) | 3.75 ± 0.64 | 5.98 ± 1.01 |
| Urea ($mmol\,L^{-1}$) | 7.84 ± 2.64 | 3.19 ± 1.15 |
| Creatinine ($\mu mol\,L^{-1}$) | 115 ± 16 | 52.2 ± 11.1 |
| Total protein ($g\,L^{-1}$) | 74 ± 9 | 54.0 ± 4.6 |
| Albumin ($g\,L^{-1}$) | 50 ± 6 | 26.2 ± 6.0 |
| Bilirubin total ($\mu mol\,L^{-1}$) | 3.42 ± 1.37 | — |
| Triglycerides ($mg\,L^{-1}$) | 267 ± 134 | 565 ± 250 |
| Total cholesterol ($mmol\,L^{-1}$) | 1.85 ± 0.38 | 1.65 ± 0.38 |
| γ-Glutamyl transpeptidase ($U\,L^{-1}$) | 61.6 ± 11.2 | — |
| Alanine aminotransferase ($U\,L^{-1}$) | 72.5 ± 13.6 | — |
| Aspartate aminotrasnferase ($U\,L^{-1}$) | 40.3 ± 5.9 | — |
| $Na^+$ ($mmol\,L^{-1}$) | 140.5 ± 4.2 | 142.3 ± 3.00 |
| $K^+$ ($mmol\,L^{-1}$) | 4.1 ± 0.3 | 3.94 ± 0.32 |
| $CL^-$ ($mmol\,L^{-1}$) | 103.1 ± 4.3 | 101.3 ± 3.6 |
| $Ca^{2+}$ ($mmol\,L^{-1}$) | 2.62 ± 0.18 | 2.58 ± 0.16 |
| $PO_4^{2-}$ ($mmol\,L^{-1}$) | 2.41 ± 0.26 | 1.61 ± 0.30 |

*Note*: Data are mean ± SD. From Parsons and Wells, 1986; Brechbuler, Kaeslin, and Wyler, 1984; Oldigs, 1986.

**TABLE 10 Minipig Hematological Parameters in Different Strains**

| Parameter | Yucatan | Göttinger |
|---|---|---|
| Red blood cells ($10^6\,mm^{-3}$) | 7.61 ± 0.15 | 7.0 ± 0.80 |
| Hemoglobin ($g\,dL^{-1}$) | 14.87 ± 0.18 | 14.9 ± 1.20 |
| Hematocrit (%) | 44 ± 0.5 | 44.6 ± 4.1 |
| Mean corpuscular volume (fL) | 58.5 ± 0.8 | 64.4 ± 3.7 |
| Mean corpuscular hemoglobin (pg) | 19.8 ± 0.3 | 21.4 ± 1.3 |
| Mean corpuscular hemoglobin concentration ($g\,dL^{-1}$) | 33.9 ± 0.3 | 33.2 ± 0.8 |
| White blood cells ($10^3\,mm^{-3}$) | 12.73 ± 0.41 | 12.6 ± 3.0 |
| Lymphocytes ($10^3\,mm^{-3}$) | 7.25 ± 0.24 | 5.75 ± 1.52 |
| Neutrophils ($mm^{-3}$) | 4.47 ± 0.24 | 5.27 ± 1.29 |
| Eosinophils ($mm^{-3}$) | 534 ± 57 | 517 ± 31 |
| Monocyte ($mm^{-3}$) | 422 ± 35 | 945 ± 71 |
| Basophils ($mm^{-3}$) | 89 ± 15 | 63 ± 1.3 |
| Platelets ($10^3\,mm^{-3}$) | — | 441 ± 119 |

*Source*: From Burks et al. 1977 (12 months old, sexes pooled); Radin et al., 1986.

differences between sexes have been identified, but age can be very much a factor. For example, serum creatinine can be 33% higher in a three-month-old as compared to 18-month-old Sinclair minipigs (based on data reported by Burks et al., 1977). As with other species, health status, feed composition, feeding regimen, fasting state, season, time of day, and so on, can affect clinical laboratory results in the minipig. Toxicological experiments should not be run without concurrent controls.

***Xenobiotic Metabolism*** Some critical parameters of hepatic microsomal drug metabolism in the minipig, common swine, and rats are given in Table 11. As most investigators tend to use younger minipigs, the values reported in this table are for young (less than four-year-old) minipigs. Relatively few papers have examined the mitochondrial mixed functional oxidase (MMFO) in a broad age range (10 months to 12 years) of Hanford minipigs. They identified definite age-related differences. The amounts of cytochrome P-450 (CYP), the MMFO activity with aniline and *p*-chloro-*N*-methylaniline, and glucoronosyl transferase activity were all significantly higher in middle-aged (5–8-year) versus young (less-than-4-year) minipigs. Freudenthal et al. (1976) examined Hanford minipigs in the age range of 2–8-months and obtained somewhat different CYP (approximately $0.95\,nmol\,mg^{-1}$) values than did Peggins et al. (1984) (approximately 0.50 nmol). The reported ranges for aniline hydroxylase (about $0.70\,nmol\,min^{-1}\,mg^{-1}$) and uridine 5′-diphospho (UDP)-glucoronosyl transferase (about $50\,nmol\,min^{-1}\,mg^{-1}$) were similar in the two papers. Hence, the available data on the MMFO of young Hanford minipigs are fairly consistent. CYP isoenzymes 1A, 2A, 2B, 2C, 2D, 2E, and 3A have all been well characterized in the pig (Gad, 2006).

The flavin adenosine dinucleotide (FAD) containing monooxygenase (FMFO) has traditionally been studied in hog liver obtained from slaughter-

**TABLE 11 Comparison of Xenobiotic Metabolism Systems in Rat and Pig**

| Enzyme | Rat[a] | Minipig[b] | Common swine[a] |
|---|---|---|---|
| Cytochrome P-450[c] | 0.59 ± 0.04 | 0.95 ± 0.02 | 0.30 ± 0.04 |
| MMFO activity[d] | | | |
| Ethylmorphine | 5.09 ± 0.34 | 8.53 ± 0.51 | 1.39 ± 0.16 |
| Ethoxyresorufin | 0.134 ± 0.022 | | 0.88 ± 0.02 |
| Epoxide hydrolase styrene oxide | 8.36 ± 2.48 | — | 11.4 ± 1.67 |
| UDP-glucoronosyl transferae | | | |
| 1-Naphthol | 6.43 ± 1.66 | — | 5.50 ± 0.89 |
| 4-Nitrophenol | 4.51 ± 0.50 | 5.5 ± 1.5 | 9.38 ± 1.07 |
| Glutathione *S*-transferase | | | |
| DNCB | 2659 ± 168 | — | 2746 ± 499 |
| DCNB | 118 ± 8.8 | — | 2.44 ± 0.23 |
| PAPS sulfotransferase | | | |
| 2-Naphthol | 0.785 ± 0.066 | — | 0.095 ± 0.025 |
| Acetyltransferase | | | |
| *p*-Aminobenzoate | 0.77 ± 0.23 | — | 0.621 ± 0.111 |

[a]From Mueller et al., 1980; Smith et al., 1984; Watkins and Klaassen, 1986.
[b]From Freundenthal et al., 1976; Peggins et al., Shipley, and Weiner, 1984.
[c]nmol/mg microsomal protein.
[d]All enzyme activities; nmol/min/mg (either microsomal or cytosolic) protein.

houses (Tynes and Hodgson, 1984). Interestingly, when FMFO activity is compared between species, substrate specificities are found to be generally very similar (Tynes and Hodgson, 1984). Rettie et al. (1990) isolated and studied the FMFO from Yucatan minipig liver. As with the enzyme studied from other species, the hepatic enzyme exists as a single isozymic species, is active with both dimethylanaline (*N*-oxide formation) and alkyl *p*-tolyl sulfides (sulfoxidation), and is enantioselective in metabolite formation. It would thus appear that the minipig does not differ appreciably from regular swine in the presence or activity of FMFO.

Perhaps some aspects of minipig xenobiotic metabolism can be inferred from studies in regular swine. For example, Rendic et al. (1984) demonstrated that cimetidine and ranitidine are excellent inhibitors of the porcine MMFO in vitro and is probably also inhibitory in microsomal preparations from minipigs (Van Ryzin and Trapold, 1980). Walker et al. (1978) reported on epoxide hydratase activity in various species, including the pig. Depending on the substrate, the pig had activities equivalent to or greater than that of the rat. This was confirmed by Smith et al. (1984) and Watkins and Klaassen (1986). The MMFO, epoxide hydrolase, UDP-glucoronosyl transferase, *N*-acetyl transferase, glutathione *S*-transferase, and sulfotransferase activities in regular swine may be used to help infer the expected activity in minipigs until more complete and specific information appears in the literature on minipigs.

There are relatively few papers that compare in vivo pharmacokinetic behavior of a specific chemical in the minipig versus another animal. Schneider et al. (1977) reported on the toxicology and pharmacokinetics of cyclotri-

methylenetrinitramine in the rat and minipig. Rats convulsed within the first several hours after receiving this chemical, whereas minipigs convulsed 12–14 h later. This is consistent with the observation that at 24 h postdosing (100 mg $kg^{-1}$ PO), the plasma levels were 3.0 μg $mL^{-1}$ in rats and 4.7 μg $mL^{-1}$ in minipigs. Other differences in pharmacokinetics and metabolism between the two species were described. The latent period for convulsion development was more similar between minipigs and humans than between rats and humans. The implication in this paper is that the minipig is a more suitable model for the study of the toxicity and metabolism of the nitramines than rats.

***Dermal Toxicity*** Although rabbits are commonly used for the assessment of primary (and even cumulative) dermal irritation, pigs are now generally considered to be better models for the more sophisticated study of dermal permeability and toxicity. As reviewed by Sambuco (1985), human and porcine skin are similar with regard to sparsity of the pelage, thickness and general morphology, epidermal cell turnover time, size, and orientation, and distribution of vessels in the skin. The particularly thin haircoat and lack of pigments of the Yucatan minipig makes it particularly ideal for dermal studies. The size of the animal also provides the additional practical advantage of abundant surface area for multiple site testing.

Sambuco (1985) has described the sunburn response of the Yucatan minipig to ultraviolet (UV) light, suggesting that this species would also make a good model in phototoxicity as well as photocontact dermatitis studies. Thirty 12-cm sites were demarcated, permitting the study of 15 different dermal dosages of UV radiation.

Mannisto and co-workers (1984) have published a series of articles on the dermal toxicity of the anthralins in the minipig. In one experiment, 24 sites per minipig were used to assess the acute dermal irritation of various concentrations to four different chemicals per site. The range of concentrations tested permitted them to calculate the median erythema concentration and median irritation concentration with relatively few animals. They were able to show clear differences between anthralin congeners (antipsoriatic drugs) with regard to irritation. When compared to other species (mouse and guinea pig) the response of the minipig was the most similar to humans in that in both species these chemicals are delayed irritants, and several days postexposure may pass before the maximal irritant response is presented.

Likewise, Hanhijarvi et al. (1985) studied the chronic, cumulative dermal effects of anthralin chemicals in minipigs. Using only 12 animals they were able, by having 32 sites per animal, to study the effects of two different chemicals (dithranol and butantrone, both anthralins) in three different formulations at three different concentrations each. The protocol also included observations for systemic toxicity, clinical laboratory measurements, plasma drug analyses, and gross and histopathological examinations.

In a third report (very similar to the second), Hanhijarvi et al. (1985) clearly demonstrated that the type of vehicle can greatly influence irritation in that

dithranol was clearly more irritating when applied in paraffin that when applied in a gel. They were also able to demonstrate that although dithranol was less irritating than butantrone acutely, the cumulative irritations (mean scores at the end of six months of six times per week applications) were quite similar (Mannisto et al., 1984). There was no evidence of systemic toxicity or of test article in plasma with either species.

***Cardiovascular Toxicity*** In general, the published literature consistently maintains that the cardiovascular systems of swine and humans are very similar. For example, as reviewed by Lee (1986), swine, including minipigs, have a noticeable background incidence of atherosclerotic lesions and swine fed high-lipid diets will develop even more extensive atherosclerotic lesions. High-lipid diets will produce lesions similar to advanced atheromatous lesions seen in humans. Although few drugs or chemicals have been shown to cause atherosclerosis, this information has three general applications to toxicology and pharmacology. First, the feeding regimen of minipigs should be carefully controlled in general toxicity studies to minimize the incidence of arterial disease, especially in long-term studies. Second, the pathologist should be aware of the natural background of this disease when preparing a diagnosis. Third, the minipig could provide a convenient model for the study of atherosclerotic disease and the screening of potential therapies.

The minipig has been used to study cardiotoxicity, particulary with medical device and drug/device combination products. Van Vleet et al. (1984) reported that minipigs were the only species studied other than dogs to develop cardiac damage in response to large doses of minoxidil. In both pig and the dod, minoxidil cardiotoxicity is characterized by vascular damage (with hemorrhage in the arterial epicardium) and myocardial necrosis (mostly of the left ventricular papillary muscles). Interestingly, in the dog the atrial lesion is largely restricted to the right atrium, whereas in the pig it is restricted to the left atrium. These lesions can be produced in roughly 50% of the minipigs given $10\,mg\,kg^{-1}$ of minoxidil for two days and sacrificed 48 h after the last dose (Herman et al., 1988, 1989). Herman and colleagues have published extensive descriptions of minoxidil-induced lesions in minipigs in comparison to those produced in dogs (Herman et al., 1988, 1989). The right versus left arterial difference is believed to be due to differences in the anatomical pattern of coronary circulation between two species (Herman et al., 1988).

Minipigs are also sensitive to the cardiotoxic effect of doxorubicin. When given six intravenous injections of either 1.6 or $2.4\,mg\,kg^{-1}$ of doxorubicin at three-week intervals, minipigs develop cardiac lesions similar to those seen in dogs, rabbits, and other experimental animals (Herman et al., 1989). The lesion is characterized by cytoplasmic vacuolation and varying degrees of myofibrillar degeneration and loss. Thus, the minipig is sensitive to the cardiotoxic effect of two well-known and extensively studied chemicals. Therefore, it is a suitable nonrodent species for the general assessment of cardiotoxicity.

***Advantages and Disadvantages*** There is one disadvantage to the use of minipigs: their size. Although minipigs are smaller than regular swine, at maturity they are generally larger than beagle dogs. Among the advantages are the facts that they are long-lived, cooperative animals with well-defined physiological and metabolic characteristics. As they are not either popular companion animals (like dogs) or physically resemble humans (like monkeys), minipigs are not specifically discussed in animal "welfare" laws like the other two species. Depending on their final form, new animal welfare regulations could make the space and maintenance costs for dogs and monkeys very prohibitive. This may make minipigs increasingly more attractive as a nonrodent species for general toxicity testing.

### 4.4 Nonhuman Primates

Nonhuman primates are often the nonrodent species of choice for safety assessment studies. There are over 500 species of NHPs that differ widely from each other in size and physical characteristics. Most of the monkeys used in experimental research belong to the suborder Anthropoidea and especially to the superfamilies of Ceboidea (marmoset, squirrel monkey) and Cercopitcoidea (macaque, papio species, rhesus). These have been popular because of (1) assumed better concordance of effects seen to those in humans and (2) smaller weights (and therefore reduced compound requirement). However, predominant factors leading to a decision whether or not to select primates as the nonrodent species for safety evaluation are summarized as follows (Hobson, 2000).

**Primates Are Selected for Safety Studies Because**

- They are the only species which exhibit the human response to the test article.
- Due to smaller body size they conserve rare or expensive test articles.
- They do not form neutralizing antibodies to the test article.
- They are physiologically more similar to humans.
- Regulatory agencies require their use.
- Prior development history dictates species choice.
- Known class effects have previously been seen in primates.

**Primates May Not be Selected for Safety Studies Because**

- Perceived expense
- Facility and logistic concerns
- Limited supplies
- Biosafety concerns
- Perceived animal rights or animal welfare concerns or pressures

- Tradition and prior development history
- Regulatory agency direction
- Data suggesting that other animal models are the "most sensitive species"

Tradition and cost are the two most frequently quoted reasons for selecting dogs as the second toxicology species instead of NHPs. Many pharmaceutical companies, especially those that primarily work with small molecules, have many years of background data in dogs and do not choose to venture into primate research without a compelling reason to do so. Secondary concerns often center on perceived biosafety or animal rights issues. Contrary to conventional wisdom, primate studies are often more cost efficient than studies in dogs. Although the purchase cost for primates is approximately twice that of dogs, many other factors suggest that the total cost of a primate safety study may be less than the cost for a similar-sized dog study. Husbandry costs are higher in dogs because of the USDA requirement for exercise. Approximately four-fold more building space is required for a dog study due to the larger cages needed. Perhaps most importantly, the smaller body size of macaques means that the requirement for expensive or scarce test articles in a primate study is approximately a third of that of a dog study.

COMPARISON OF COSTS FOR TYPICAL 90-DAY STUDIES CONDUCTED WITH NHPS OR DOG

| | |
|---|---|
| Animal cost (assume 40 animals) | |
| Cost of dogs | $40 \times \$900 = \$36{,}000$ |
| Cost of primates | $40 \times \$2200 = \$88{,}000$ |
| Per diem | |
| Dogs | $90 \text{ days} \times \$11 \times 40 = \$39{,}600$ |
| Primates | $90 \text{ days} \times \$7 \times 40 = \$25{,}200$ |
| Test article (at $\$400\,\text{mg}^{-1}$) | |
| Dogs | $10\,\text{kg} \times 90 \text{ days} \times 40 \text{ animals} \times 100\,\mu\text{g}\,\text{kg}^{-1}\,\text{day}^{-1} = \$1{,}440{,}000$ |
| Primates | $4\,\text{kg} \times 90 \text{ days} \times 40 \text{ animals} \times 100\,\mu\text{g}\,\text{kg}^{-1}\,\text{day}^{-1} = \$576{,}000$ |

Clearly, the amount and cost of the test article and the length of the study determine which species is most cost efficient.

Because NHPs are phylogenetically closer to humans than other species, there is less chance that they will recognize human protein, peptide, or antibody-based biopharmaceuticals as foreign. Thus, they are often selected for safety studies of these materials. Although highly conserved proteins may not be immunogenic in lower species, clearly the formation of neutralizing antibodies to less conserved proteins during a safety study can confound experimental results (Dean et al., 1990). The formation of neutralizing antibodies to human biopharmaceuticals almost never occurs in chimpanzees (which no longer may be used in pharmaceutical research) but occurs more

and more frequently as the primate phylogenetic tree is descended. It is generally believed that the NHP phylogenetic difference from humans is ranked as follows: great apes, baboons, other Old World primates (including macaques), and New World primates. Clearly, as proteins are modified, they can become immunogenic in all primate species, including humans.

The physiological similarity and phylogenetic proximity of NHPs to humans are often cited as a rationale for primate selection for safety studies, especially when mechanisms of toxicity or pharmacological action are expected to be closely related to potential physiological reactions in humans. Likewise, species selection is often based on the demonstration of pharmacological activity of the test article. Many biopharmaceuticals do not exhibit their intended activity in nonprimate species, whereas small molecules may have activity across all species.

Regulatory agencies sometimes suggest (read "dictate") use of primates for certain study designs or drug classes. These requirements are often a surprise to companies when they are first presented. Usually they are derived from confidential data that the regulatory agencies have previously reviewed. Often the regulatory bodies are privy to data that suggest that a class effect is seen in primates and not in other species or that primates are the most sensitive species. An example was regulatory agency encouragement to perform cardiovascular evaluations of oligonucleotide pharmaceutical candidates in primates (Black et al., 1993, 1994). This was based on background information that suggested that oligonucleotides induced complement activation and the attendant hemodynamic and cardiovascular changes in primates but not in other species (Galbraith et al., 1994).

Animal welfare and conservative issues have frequently led to decisions to avoid primate use. Through the mid-1980s many NHPs used in medical research came from wild populations. This led to strong conservationist concerns with the use of monkeys in research. Now, however, almost all NHPs used in research are purpose bred and the conservationist concern has abated. Although there is some animal rights pressure specifically directed against primate use, it is not as formidable as the well-financed and sophisticated efforts to prevent the use of cast-off dogs (pound or shelter dogs or dogs from other class B sources in research). As a consequence, a few pharmaceutical companies are considering switching to NHPs for their second toxicology species.

***Environmental and Dietary Requirements*** For most NHP species, room temperatures should be maintained in the range of 75 ± 5 °F with a relative humidity of 40% or greater. These temperature and humidity ranges have been found to be beneficial to the prevention of pneumonia and bloody nose syndrome. Rooms in which monkeys are housed should have 10–15 air changes per hour and be kept under negative pressure in relation to other parts of the building. Where there is significant risk of airborne infection, it is necessary to contain infected animals in units designed to remove the air away from personnel (Mazue and Richez, 1982).

Physical comfort should be an important consideration when determining the appropriate housing for NHPs. For individually housed NHPs, the floor area and height of cages should be about $0.28\,cm^2 \times 76.20\,cm$ for animals in the weight range of 1–3 kg and $0.40\,cm^2 \times 76.20\,cm$ for monkeys weighing 3–10 kg (DHHS, 1985). Probably the most common form of commercially available housing is mobile stainless steel rack–mounted cages. Group housing of NHPs used in safety studies is likely to become more common in the future as a result of the USDA current animal welfare regulations, which require that NHPs have the opportunity for socialization.

Another requirement of the new animal welfare regulations is that any cage or pen in which NHPs are housed must also contain toys, food, or other objects that animals can manipulate as they would objects in their natural environment. From experience, laboratories have found that toys in themselves are not sufficient since the animals quickly lose interest. Effective enrichment materials include foraging boards (fur-covered objects under which food is buried) and puzzle feeders for more complex foraging.

In many laboratories, monkeys are often fed commercial pelleted chow ad libitum supplemented with fresh fruits and bread. Like the human and guinea pig, the monkey cannot synthesize vitamin C and, thus, has a dietary requirement for this vitamin. Powdered chow is an inefficient form for feeding NHPs because a high percentage of the diet is wasted. Also, dust associated with the chow can cause respiratory problems in some species (NRC, 1978). Even with pelleted or extruded food, monkeys will waste about 50% of the ration sorting through the pellets (Mazue and Richez, 1982). Monkeys should have ad libitum access to water, and it is important that the device (either a water bottle equipped with a sipper tube or an automatic watering system) be fixed securely to the cage to avoid detachment by the animal.

***Common Study Protocols*** Group sizes and numbers of animals per group for NHP toxicology studies vary slightly from country to country and from company to company; however, with the movement for international harmonization there is trend toward less variation in study design. Selection of group size is a compromise among regulatory guidelines, cost, statistical power, and conservation of animals.

An example of a protocol for a four-week safety study in cynomolgus monkeys is shown in Table 2. Cynomolgus monkeys are generally in the age range of one to three years at study start. A two- or four-week study will usually have about four animals per sex per group, not including the mandatory extra animal for evaluating recovery. For the longer term studies, the number of animals per group will be larger in order to include reversal animals. As with dogs and ferrets, monkeys should be selected for study use based on acceptable pretreatment body weights, clinical laboratory profiles, and physical, ECG, and ophthalmic examinations.

One aspect of study design in NHPs that is not well understood is caused by the variability in the age at which monkeys undergo puberty. Although age

at the onset of puberty is highly variable within macaque species, there is a remarkable correlation between body weight and sexual maturity in macaques. *Rhesus* females undergo menarche at 3000 ± 200 g irrespective of age, whereas males tend to become sexually mature around 4500 g. This means that a "typical" study is initiated with sexually mature females and sexually immature males. This practice is debatable and is certainly not universally adopted. A few pharmaceutical companies require mature animals of both sexes. Because sexual maturation in males occurs many months later than in females, rearing costs are higher for males and animal numbers may be limited because the younger males may have already been sold with their female birth-year counterparts.

### General Study Activities

*Common Dosing Techniques* Dosing routes and permissible volumes for NHPs vary between laboratories. The volume limitations from our laboratory are presented in Table 12.

Primates offer all of the possible dosing routes available in humans, but body size often limits dosing volumes. If volumes for subcutaneous or intramuscular injections exceed those suggested above, enzyme elevations [particularly alanine aminotransferase (ALT) and aspartate aminotransferase (AST)] are frequently observed (unpublished results). Continuous infusion techniques in alert animals are available in some laboratories either through use of programmable backpack pumps or jacket-and-tether systems (Perkin and Stejskal, 1994).

Probably the most common oral route is gavage. This procedure usually requires some degree of physical restraint of the animal (by one or more persons) while a stomach tube for dosing is inserted either orally or intranasally. Other oral dosing methods include buccal, capsule, or addition of the test compound to the drinking water. It is also possible to prepare a modified diet admixture consisting of test compound, diet meal, water, agar, and a jelling agent. This type of preparation will reduce both the feed spillage and the dust normally associated with powdered chow; however, it is susceptible to microbial growth and must be kept frozen or refrigerated (NRC, 1978).

Bolus intravenous, intramuscular, or subcutaneous injections can be administered by a single person by securing the animal's arm through the cage bars

**TABLE 12 Permissible Dosing Volumes for Nonhuman Primates**

| Route | Maximum Permissible Dose[a] |
|---|---|
| Intravenous | Varies with duration of administration and character of test article |
| Subcutaneous | 2 mL site$^{-1}$ and 5 mL kg$^{-1}$ |
| Intramuscular | 0.25 mL site$^{-1}$ and 0.5 mL kg$^{-1}$ |
| Oral/nasogastric | 5 mL kg$^{-1}$ |

[a]Values are given for single or infrequent administration. Smaller volumes are appropriate for repeated dosing.

(Mazue and Richez, 1982). For safety considerations, many investigators prefer to have the animal physically restrained by a second person before the injection is given. Arterial injections (via the femoral artery) as well as limited or continuous intravenous infusions (via catheterization of the femoral or jugular vein) are other less commonly used parenteral routes in the monkey.

Other routes of administration sometimes used in monkey safety assessment studies are intravaginal dosing, topical application, inhalation, and nasal administration.

*Clinical Observations and Examinations* As with other species, it is important to have a good understanding of the types of normative behaviors and clinical signs that can be seen in normal, untreated monkeys before attempting to make observations in drug-treated animals. Cage-side observations in the monkey should be conducted at least two times daily to monitor general health and behavior. The first observation should be made before cages are cleaned in the morning, and the floors of the cages should be critically examined for signs of blood, abnormal feces, or vomitus. Clinical signs to which investigators should pay particular attention include reduced activity and lethargy, excessive excitation, reduced feed consumption, vomiting, and abnormal feces. If at all possible, the same people should work on a study for its entirety. The behavior of more timorous monkeys can be affected by the presence of unfamiliar personnel, resulting in undesirable clinical signs such as a loss of appetite and lethargy (Evans et al., 1982). To circumvent these kinds of problems, isolated observation using a video camera system may be a preferable approach.

Physical examinations of monkeys are usually conducted no more than once a week and generally consist of the measurement of rectal temperature, observation of general demeanor, palpation of the head, thorax, and abdomen, examination of eyes, ears, and bodily orifices, as well as testing of the pupillary and patellar reflexes.

*Feed Consumption* As mentioned previously, monkeys tend to scatter their feed, which can make feed consumption difficult to measure. It may be possible to successfully monitor feed consumption in monkeys by using the larger chow biscuits and counting the number of biscuits (or fractions of biscuits) consumed over two consecutive 24-h periods.

*Electrocardiograms and Cardiovascular Measurements* The availability of excellent good laboratory practices (GLP)–validated telemetry systems has led to recent increases in the number of cardiovasular safety pharmacology studies conducted in primates. In addition, telemetry is now sometimes included as a design element in standard safety studies. Because of the ability to collect large amounts of high-quality data over an extended time, total numbers of animals can often be reduced by appropriate application of telemetry. Indeed, it is often difficult to avoid statistical and reporting problems

caused by the temptation to collect too much data using telemetry. Implanted transmitters can function continuously for up to a year without battery replacement while providing data such as blood pressure, heart rate, ECG, body temperature, and activity.

For safety assessment studies, it is preferable to record monkey ECGs in the conscious animal, which, if using standard ECG techniques, requires chairing the animal. Electrocardiographic leads used in this laboratory include II, aVL, and V10. To help reduce emotional tachycardia, it is recommended that there be pretreatment habituation (no more than 10–30 min at least twice before study start) to the chairing and attachment of the surface electrodes. Probably the best and least stressful approach to monitoring ECG activity in conscious monkeys is automatic monitoring using a biotelemetry system. With this system, a transmitter surgically implanted subcutaneously along the dorsal midline broadcasts a radio signal encoding the ECG to a receiver mounted on top of the animal's cage and a computer records the signal at 2-min intervals (Line et al., 1989).

*Blood and Urine Collections* For blood collection, the rhesus can be bled from the saphenous or femoral vein. For female rhesus monkeys, it may not be possible to use the saphenous vein because of the swelling of the sex skin (i.e., the edematous thickening and reddening of the skin over the external genital region, rump, and tail that often extend down the leg to the knee). For the cynomolgus and squirrel monkeys, the veins are very small, and the femoral vein is usually the one of choice. Depending on the species, 2 (marmoset) to 24 (cynomolgus) milliliters may be collected. However, experimental designs in primates are often constrained by the limitations in the amount of blood that can be safely and humanely obtained during the course of a study (Fuller et al., 1992). With increasing emphasis on obtaining toxicokinetic data during safety studies, these constraints have become more vexing. A guideline for maximum blood withdrawal is $10\,mL\,kg^{-1}\,day^{-1}$ (Heiser, 1970). More blood can be collected, but hematocrits should be monitored (Schalm, 1975). These amounts do not approach maximum amounts allowable for humane considerations but do represent the maximum that can be collected without causing more than slight decreases in hematological parameters (notably hematocrit, hemoglobin, and red cell count). Currently used catheter material and vascular access ports long-term frequent blood collection without the catheter clotting and emboli problems experienced in the past. The new vascular access ports remain patent for over a year with routine maintenance. Vascular access ports are particularly useful in primates where frequent samples are required because blood sample collection through ports appears to be far less stressful than collection by needle stick. We have also found them useful when evaluating anticoagulant test articles because venipuncture is contraindicated. Sample quality is also superior with ported collections.

Urine collection in NHPs can be measured using either a metabolism cage or a collection pan (equipped with a screen to catch the feces), which is

inserted under the floor grid of the home cage. The advantage of the latter system is that the animals do not need to be removed from their home cage; however, care needs to be taken to avoid contamination of the urine with drinking water.

***Advantages and Disadvantages*** Advantages of using NHPs in safety assessment studies include their phylogenetic proximity as well as their physiological, behavioral, and, often, metabolic similarities to humans (Table 13). An example is the similarity between the ovarian cycle of female monkeys and women (Mazue and Richez, 1982), which makes the monkey the ideal animal model for reproductive studies. Another advantage associated with most species of monkeys used in safety assessment studies is that they are much smaller than nonrodents such as the dog and thus, like the ferret, require less test compound.

The most significant disadvantage to working with monkeys is the serious spontaneous diseases they can carry that are transmissible, and often life threatening, to humans. An example of one such disease is herpesvirus simiae (B virus). B virus is widespread, especially among wild-caught and to some extent laboratory-bred rhesus monkeys, including cynomolgus monkeys. Human exposure to B virus occurs during handling of monkeys and monkey tissues (via contact with tears, blood, or saliva of infected animals) and is associated with a high incidence of human mortality (DiGiacomo and Shah, 1972). Other serious to very serious diseases that can be transmitted from monkey to human are Marburg disease, viral hepatitis, tuberculosis, and monkeypox.

## 5 STATISTICS IN LARGE-ANIMAL STUDIES

Large-animal toxicology studies, typically ranging from 14 days to generally a maximum of 52 weeks, pose different types of statistical problems and open up new possibilities in terms of statistical evaluations. Standard statistical

**TABLE 13 Use of Nonhuman Primate in Safety Assessment Studies**

| *Advantages* |
|---|
| Small size of many species |
| Less test material needed than for other nonrodent species |
| Physiological, behavioral, and, often, metabolic similarities to humans |
| *Disadvantages* |
| Limited availability |
| Cost |
| Need to develop environmental enrichment program |
| High potential for spontaneous diseases |

methods used for chronic toxicology studies, such as one-way designs, often do not provide any meaningful insights because of small sample sizes used in large-animal studies. The designs for such studies are, generally speaking, nonoptimal. As a consequence, an investigator must attempt to use optimal statistical methods to evaluate such studies. Fortunately, for many of the relevant parameters for such studies, there are fewer to no dropouts (if one is careful) and there are repeated measurements on the same parameters of interest, at both pre- and posttreatment intervals. Optimality of statistical methods for such studies is then achieved by making use of the longitudinal observations in the analysis. The optimality can be further enhanced by introducing sex as a factor in the evaluation of the data in many such studies.

Many of the standard assumptions in both parametric and distribution-free statistical methods cannot be meaningfully tested in large-animal studies because of extremely small sample sizes (which is dictated not necessarily by scientific doctrine but by economic and minimum regulatory requirements). Fortunately, by making use of solid biological as well as statistical judgments, we seem to have made many discoveries in terms of human safety and efficacy in large-animal toxicology.

Instead of conventional textbook-type layout, this discussion will try to focus on various issues in large-animal toxicology experiments with plausible examples. One word of caution before we get deeper into our discussion: As in most areas of applied statistics, there really is no gospel in what we will be discussing today. Many statisticians may have variations of the theme to be brought out here.

### 5.1 Reasons for Small Sample Sizes in Large-Animal Toxicology

The following are some of the main reasons for having only three to five dogs or monkeys per sex in a typical large-animal study:

1. These studies are very expensive. A typical full-fledged study may cost as much as $30,000–$500,000 (for 26 weeks).
2. There is tremendous pressure from animal rights groups to look for alternatives, rather than using dogs and monkeys for investigative purposes.
3. Regulatory agencies throughout the world recognize these two facts and recommend such small sample sizes as minimum requirements. As a consequence, the pharmaceutical and chemical industries are reluctant to expand the scopes of such studies.

### 5.2 Cross-Sectional or Longitudinal Analysis?

Many of the studies we deal with have various parameters, such as body weight, food consumption, clinical chemistry, and hematology, that are collected repeatedly at various pre- and posttreatment intervals. Unfortunately, many investigators in the field do not take advantage of this important design

feature of such studies. Instead, the literature is full of simple parametric or distribution-free one-way techniques such as Student's *t* test, Wilcoxon–Mann–Whitney rank test, and one-way analysis-of-variance (ANOVA) methods that are being widely used sometimes without satisfaction. The argument then is given that "although there is apparent biological effect (or lack of it), because of small sample sizes and poor statistics, no significant effects can be determined from these data," or something like that. If truth be known, the small-sample-size part of this arguments may be correct; however, no attempts were made to optimize the statistical methods above using the various pieces of the particular design. The repeated-sampling part of the design (repeated measures) is very important for such studies and therefore should be incorportated in the analysis of the data. After all, design of experiment and analysis of data are inseperable. There are advantages and disadvantages of such analyses (the advantages generally outweigh the disadvantages) as described below.

## 5.3 Repeated Measures

**Advantages**

1. Between-subject variations are excluded from the experimental and stochastic errors.
2. Only the within-subject variation is included in the mean-square error (MSE) term.
3. Each subject becomes its own control.
4. The number of subjects in an experiment are kept to a minimum.
5. Both type I (false-positive) and type II (false-negative) error rates are minimized, thereby increasing power of the test statistic to be employed while decreasing inconsistent significant effects.

**Disadvantages**

1. Order of the treatment may cause interference, which can be avoided by appropriate randomization.
2. There is the possibility of carry-over effects. This is more crucial in Latin square and other cross-over designs. Knowledge of pharmacokinetics and metabolism of a compound under study generally helps in avoiding this problem.
3. Exact permutation and distribution-free techniques are not as widely developed as in the cases of one-way methods.
4. Power and sample size computations are a little more difficult to compute than for one-way designs.
5. Generally computers are required for performing the analyses using specialized software (not a major issue in most societies nowadays).
6. The results are a little more difficult to interpret than their one-way counterparts.

## 5.4 Common Practices in Large-Animal Toxicology

Older (and some newer) literature in large-animal toxicology is full of two-sample, one-way parametric, and distribution-free techniques. Some of the newer works use repeated-measures and even multivariate techniques. The following is a brief exposé of various methods used in the field:

1. One-way analysis of variance/covariance/regression and preplanned and post hoc group comparisons.
2. Two-sample Student's *t* test, Wilcoxan–Mann–Whitney rank test, and so on
3. Graphical display of response over time (as two- or three-dimensional plots)
4. Univariate repeated-measures analysis of variance/covariance techniques
5. Multivariate analysis of variance/covariance (MANOVA/MANCOVA) techniques

Methods 1 and 2 above should not be preferred in global analyses. Graphical displays have tremendous values as exploratory data analysis (EDA) techniques with the type of data one encounters in these studies. For formal analyses, one could weigh univariate repeated and other factorial designs against their true multivariate counterparts.

## 5.5 Univariate (Repeated-Measures) Techniques

**Advantages**

1. Easier to compute
2. Less susceptible to violation of normality
3. Exact and distribution-free tests easier to compute
4. Require smaller sample sizes; there is more power
5. Very few test statistics to deal with: classical ANOVA *F*; Greenhouse–Geisser and Huynh–Feldt adjusted degrees of freedom (df), and ANOVA *F*
6. Biologically meaningful and easier to resolve contrasts and multiple comparison tests
7. Missing values easily handled

**Disadvantages**

1. Susceptible to heteroscedasticity (heterogeneity of variances)
2. Less fancy compared to multivariate techniques

## 5.6 Multivariate Techniques

**Advantages**

1. Less susceptible to heteroscedasticity
2. Handles multiple dependent variables
3. Real fancy compared to univariate ANOVA/ANCOVA techniques

**Disadvantages**

1. More susceptible to violation of normality
2. Less power than univariate ANOVA, particularly with small sample sizes
3. Contrasts and multiple comparisons difficult to construct
4. Missing values more difficult to handle
5. Computationally more difficult (a mute point nowadays with personal computers)
6. Too many test statistics to deal with, sometimes giving contradictory answers

## 5.7 Some Other Design Factors to be Considered in Analysis

Most of the toxicological studies are designed to evaluate efficacy and safety in both sexes. With small sample sizes, one can increase the power efficiency of the particular test statistic by including sex as a factor in a full factorial analysis (not combining the two sexes) where appropriate. The factorial analysis will reveal whether there is any need to separate the two sexes. The other design fact that should be weighed carefully is the presence of any concomitant variables or covariates. For example, most large-animal studies will involve collection of data both prior to the beginning of the experiments as well as after. Thus pre-treatment values and other characteristic control variables (e.g., body weights) may be important covariates in the analysis of the data. There are both advantages and disadvantages in including covariates in the analysis:

**Advantages**

1. Increases precision of an analysis (indirect or statistical control of variability)
2. Correction of bias
3. Elimination of extraneous variation in the data

**Disadvantages**

1. Unequal intra- and intergroup covariate slopes—may actually introduce bias as a consequence.
2. Nonlinearity of covariate slopes—may have the same effect as above.
3. In some cases the covariates may be affected by treatment.

An example is shown in Table 14. A two-factor ANOVA for the covariate, as shown in Table 15, clearly indicates that the two sexes started with approximately the same means ($p = 0.5598$). Moreover, there were no differences between the group means in either sex, as indicated by the large tail probabilities for treatment ($p = 0.8823$) and sex–treatment interaction ($p = 0.6532$). These facts justify using sex as a factor in the analysis, as was done here.

There are various other ways of examining the variate in question in this case. Let us first examine a simple one-way ANOVA of the variate by sex as in Table 16. In neither of the two cases was there any indication of significant treatment differences at any reasonable level. Because the two sexes did not show any pretreatment differences based on the two-factor analysis of the covariate, let us combine the two sexes and analyze the data by one-way ANOVA as in Table 17. In this case, because of the increased sample sizes for combining the two sexes, there was an indication of some treatment differences ($p = 0.0454$). Unfortunately, this analysis assumes that because there was no pretreatment difference between the two sexes, that pattern will hold

**TABLE 14 Example 1**

| | Control | | Treatment 1 | | Treatment 2 | |
|---|---|---|---|---|---|---|
| Sex | Covariate | Variate | Covariate | Variate | Covariate | Variate |
| Male | 40 | 95 | 30 | 85 | 50 | 90 |
| | 35 | 80 | 40 | 100 | 40 | 85 |
| | 40 | 95 | 45 | 85 | 40 | 90 |
| | 50 | 105 | 40 | 90 | 30 | 80 |
| | 45 | 100 | 40 | 90 | 40 | 85 |
| Raw mean | 42.0 | 95.0 | 39.0 | 90.0 | 40.0 | 86.0 |
| SD | 5.7 | 9.4 | 5.8 | 6.1 | 7.1 | 4.2 |
| Female | 50 | 100 | 50 | 100 | 45 | 95 |
| | 30 | 95 | 30 | 90 | 30 | 85 |
| | 35 | 95 | 40 | 95 | 25 | 75 |
| | 45 | 110 | 45 | 90 | 50 | 105 |
| | 30 | 88 | 40 | 95 | 35 | 85 |
| Raw mean | 38.0 | 97.6 | 41.0 | 94.0 | 37.0 | 89.0 |
| SD | 9.1 | 8.1 | 7.4 | 4.2 | 10.4 | 11.4 |

**TABLE 15 Two-Factor Analysis of Variance for Covariate**

| Source | Sum of Squares | DF | Mean Squares | *F* | Tail Probability |
|---|---|---|---|---|---|
| Mean | 46,807.50000 | 1 | 46,807.50000 | 785.58 | 0.0000 |
| Sex | 20.83333 | 1 | 20.83333 | 0.35 | 0.5598 |
| Treatment | 15.00000 | 2 | 7.50000 | 0.13 | 0.8823 |
| Sex × treatment | 51.66667 | 2 | 25.83333 | 0.43 | 0.6532 |
| Error | 1,430.00000 | 24 | 59.58333 | | |

**TABLE 16 One-Way Analysis of Variance of Variable of Sex**

| Source | Sum of Squares | DF | Mean Squares | *F* | Tail Probability |
|---|---|---|---|---|---|
| | | | *Males* | | |
| Mean | 122,401.66667 | 1 | 122,401.66667 | 2576.88 | 0.0000 |
| Treatment | 203.33333 | 2 | 101.66667 | 2.14 | 0.1603 |
| Error | 570.00000 | 12 | 47.50000 | | |
| | | | *Females* | | |
| Mean | 131,227.26667 | 1 | 131,227.26667 | 1841.36 | 0.0000 |
| Treatment | 186.53333 | 2 | 93.26667 | 1.31 | 0.3061 |
| Error | 855.20000 | 12 | 71.26667 | | |

**TABLE 17 One-Way Analysis of Variance for Combined Sexes**

| Source | Sum of Squares | DF | Mean Squares | *F* | Tail Probability |
|---|---|---|---|---|---|
| Mean | 253,552.13333 | 1 | 253,552.13333 | 4549.999 | 0.0000 |
| Treatment | 387.26667 | 2 | 193.63333 | 3.47 | 0.454 |
| Error | 1,504.60000 | 27 | 55.72593 | | |

**TABLE 18 Two-Factor Analysis of Variance with Sex as Factor**

| Source | Sum of Squares | DF | Mean Squares | *F* | Tail Probability |
|---|---|---|---|---|---|
| Mean | 253,552.13333 | 1 | 253,552.13333 | 4269.75 | 0.0000 |
| Sex | 76.80000 | 1 | 76.80000 | 1.29 | 0.2667 |
| Treatment | 387.26667 | 2 | 193.63333 | 3.26 | 0.0559 |
| Sex × treatment | 2.60000 | 2 | 1.30000 | 0.02 | 0.9784 |
| Error | 125.20000 | 24 | 59.38333 | | |

during the posttreatment period. That often may not be the case because of biological reasons.

The above analysis establishes that there was no significant sex difference, as indicated by the tail probabilities for sex ($p = 0.2667$) and sex–treatment interaction ($p = 0.9784$). There was also some indication that there may have been some treatment effect across the treatment groups in both sexes ($p = 0.0559$) (Table 18). Examination of the variate means indicated that both sexes seemed to have lower means than their respective controls. The picture was clouded by the fact that there was such a slightly lower tendency, though not very consistent, in the covariate means as well. Under this circumstance, it is more appropriate to take both the covariate and the variate into any optimal analysis. Table 19 shows an ANCOVA for the factorial model.

As the ANCOVA table indicates, there was definite significant treatment effect ($p = 0.0104$), but this effect was not sex specific because there was no significant sex–treatment interaction ($p = 0.7613$). Furthermore, there was a signficinat difference between the two sexes in terms of magnitude but not in

**TABLE 19 Analysis of Covariance of Factorial Model**

| Source | Sum of Squares | DF | Mean Squares | *F* | Tail Probability |
|---|---|---|---|---|---|
| Mean | 147.42310 | 1 | 147.42310 | 5.65 | 0.0262 |
| Sex | 292.81064 | 2 | 146.40532 | 5.61 | 0.0104 |
| Treatment | 14.41235 | 2 | 7.20617 | 0.28 | 0.7613 |
| Sex × treatment | 824.75245 | 1 | 824.75245 | 31.59 | 0.0000 |
| Error | 600.44755 | 23 | 26.10642 | | |

| | Adjusted Cell Means and Standard Errors | | |
|---|---|---|---|
| | Control | Treatment 1 | Treatment 2 |
| | *Males* | | |
| Mean | 93.10140 | 90.37972 | 85.62028 |
| Standard error | 2.30985 | 2.28601 | 2.28601 |
| | *Females* | | |
| Mean | 98.73916 | 92.86084 | 90.89860 |
| Standard error | 2.29398 | 2.29398 | 2.30985 |

the direction of the effect. These findings are apparent in the covariate-adjusted means in all groups in both sexes. The magnitude of the treatment effect became amplified by introducing the covariate in the model. As can be seen from the two ANOVA and ANCOVA tables above, despite the fact that the ANCOVA error term lost one degree of freedom (df = 23) as opposed to the ANOVA error term (df = 24), the former gains some edge over the latter because of increased precision. Precision in this context is defined as the ratio between the MSEs of ANOVA and ANCOVA. For this example:

$$\begin{aligned}\text{Precision} &= \frac{1/\text{MSE}_{\text{ANCOVA}}}{1/\text{MSE}_{\text{ANOVA}}} \\ &= \frac{\text{MSE}_{\text{ANOVA}}}{\text{MSE}_{\text{ANCOVA}}} \\ &= \frac{59.38333}{26.10642} \sim 2.3\end{aligned}$$

In other words, we have gained about 2.3-fold precision by ANCOVA over ANOVA in resolving treatment effect.

With the advent of powerful personal computers and the availability of sophisticated "do-it-all" statistical packages, there is a trend among nonstatisticians (even some statisticians) to accept the results from these packages without contemplating twice. Many of these packages have flexible features that allow one to perform different types of analyses with the same data set, inappropriately or appropriately sometimes. What popular statistical packages give is not necessarily correct statistics or they may not be correct under spe-

cific designs. Some programs, for example, BMDP's 2V (1992), have "intelligence" built into them whereby they can identify the design based on the data matrix. By correctly following the data matrix setup specified in the manual, one can simply press the button and get the appropriate analysis needed. On the other hand, incorrect specification of the data matrix will produce incorrect results (although some programs, such as 2V, will often give an error message or prompt to make sure one wants what one is asking for; some, such as SAS's PROC GLM, may not and give results that are not even remotely related to the design). In other words, one must know some statistics and must be well versed in the features of the particular package before using it. The one-time famous mathematician–statistician–composer–pianist–singer–producer–recording artist Tom Leher (1959), in one of his famous monologues, said, "Life is a sewer; what one gets out of it depends on what one puts into it." Statistical packages are exactly like that.

## 5.8 Missing Values

All investigators know that missing values are a nuisance. They also create statistical nightmares. Classical statistical techniques were not geared toward having missing values in experiments. Unfortunately, in real-life situations, it just happens. Animals may die or are censored for various reasons. There are various techniques of calculating missing values for specific designs (Miller, 1981) just like there are for extreme values or outliers [Statistical Analysis System (SAS), 1996]. In neither case is there any unique way of handling them that is completely agreed upon by statisticians. One should remember that every time a missing value is computed and used in statistical analyses, one loses a degree of freedom. In large-animal toxicology, with small sampling sizes, one must be very careful about dealing with missing values. In a repeated-measures analysis, if one observation is missing from an animal during one interval, classical techniques automatically will exclude observations from that animal for all remaining intervals. Newer techniques based on regression or imputation have been developed in recent years and have been implemented in popular packages such as BMDP (5V) or SAS (PROC MIXED). Within a single package, there may be various techniques based on assumptions on covariance structures (unstructured, compound symmetry, etc.) and statistical algorithms (maximum likelihood, restricted minimum likelihood, etc.). The results sometimes could be very different under the same assumptions and algorithms. As a result, given the same compound symmetry assumption and using the same restricted maximum-likelihood (REML) algorithm, two well-known programs give different quantitative results. These methods are still experimental in nature and should not be taken for granted. Actually, the BMDP manual clearly warns users about the nature of this method. Consequently, the best way to avoid confusion is to try to make sure that missing values do not occur in key parameters in *large-animal* studies (Thakur, 2000).

## 6 SUMMARY

While there are advantages and disadvantages associated with all three nonrodent species, the dog is probably the nonrodent species most frequently used in safety assessment studies. This is because dogs are relatively docile and even tempered, they are generally more easy to obtain and relatively less expensive than monkeys, they carry less serious diseases than the ferret and the monkey, and they have a more extensive historical database in safety studies. It should be noted, however, that if the technical and health problems associated with the ferret can be overcome, its small size in terms of compound requirements, cost, and housing may make it an ideal nonrodent species for future use in safety assessment studies.

## REFERENCES

Andrews, P. L. R., Bower, A. J., and Illman, O. (1979). Some aspects of the physiology and anatomy of the cardiovascular system of the ferret, *Mustela putorius furo*. *Lab. Anim.* 13:215–220.

Barnett, J., and Hensworth, P. (1986). The impact of handling and environmental factors in the stress response and its consequences in swine. *Lab. Anim. Sci.* 36:366–369.

Black, L. E., DeGeorge, J. J., Cavagnaro, J. A., Jordan, A., and Ahn, C. H. (1993). Regulatory considerations for evaluating the pharmacology and toxicology of antisense drugs. *Antisense Res. Dev.* 3:399–404.

Black, L. E., Farrelly, J. G., Cavagnero, J. A., Ahn, C. H., DeGeorge, J. J., Taylor, A. S., DeFelice, A. F., and Jordan, A. (1994). Regulatory considerations for oligonucleotide drugs: Updated recommendations for pharmacology and toxicology studies. *Antisense Res. Dev.* 4:299–301.

Brechbuler, T., Kaeslin, M., and Wyler, F. (1984). Reference values of various blood constituents in young minipigs. *J. Clin. Chem. Clin. Biochem.* 22:301–304.

Burks, M., Ymbleson, M., Hicklin, K., Hutcheson, D., and Middleton, C. (1977). Age and sex related changes of hematologic parameters in sinclair (S-1) miniature swine. *Growth* 41:51–62.

Campbell, S. A., Hughes, H. C., Griffin, H. E., Landis, M. S., and Mallon, F. M. (1988). Some effects of limited exercise on purpose-bred beagles. *Am. J. Vet. Res.* 49:1298–1301.

Clausing, P., Beitz, H., Gericke, S., and Solecki, S. (1986). On the usefulness of minipigs in toxicology testing of pesticides. *Arch. Toxicol.* 9(Suppl.):225–271.

Dean, J. H., Cornacoff, J. B., Labrie, T., and Barbolt, T. A. (1990). Assessment of immune responses in rodents and non-human primates—Implications in preclinical evaluation of proteins. In A. Sundwall et al. (Eds.), *Preclinical Evaluation of Peptides and Recombinant Proteins*. Malmo. Skogs Grafiska AB, Stockholm, Sweden, pp. 23–34.

Diehl, K. H., Hull, R., Morton, D., Pfister, R., Rabmampianina, Y., Smith, D., Vidal, J.-M., and vortembosch, C. (2001). A good practice guide to the administration of substances and removal of blood, including routes and volumes, *J. Appl. Toxicol.* 21:21–23.

Detweiler, D. K. (1980). The use of electrocardiography in toxicological studies with beagle dogs. In S. L. Baskin (Ed.), *Cardiac Toxicology*. CRC Press, Boca Raton, FL, pp. 82–104.

Detweiler, D. K., Patterson, D. F., Buchanan, J. N., and Knight, D. H. (1979). The cardiovascular system. In Vol. 2. E. J. Catcott (Ed.), *Canine Medicine*, American Veterinary Publications, Santa Barbara, CA.

DiGiacomo, R. F., and Shah, K. V. (1972). Virtual absence of infection with *Herpesvirus simiae* in colony-reared rhesus monkeys (*Macaca mulatta*) with a literature review on antibody prevalence in natural and laboratory populations. *Lab. Anim. Sci.* 27:61–67.

Earl, F. L., Miller, E., and Van Loon, E. J. (1973). Teratogenic research in beagle dogs and miniature swine. In H. Spiegel (Ed.), *The Laboratory Animal in Drug Testing: Fifth Symposium*. Internal Committee on Laboratory Animals, pp. 233–247.

Ellegaard, L., Jorgensen, K. D., Klastbruck, S., Hansen, A. K., and Svendsen, O. (1995). Haematologic and clinical chemical values in three- and six-month old Göttinger minipigs. *Scand. J. Lab. Anim. Sci.* 22:239–248.

Evans, R. H. (1982). *Nonhuman Primates*. Ralston Purina Co., St. Louis, MO.

Fox, J. G. (1988). *Biology and Diseases of the Ferret*. Lea and Febiger, Philadelphia.

Freudenthal, R., Leber, P., Emmerling, D., Kerchner, G., and Campbell, D. (1976). Characterization of the hepatic microsomal mixed-function oxidase system in miniature pigs. *Drug Metab. Dispos.* 4:25–27.

Fuller, G. B., Hobson, W. C., Renquist, D. M., Port, C. D., and Chengelis, C. P. (1992). Nonhuman primates. In S. C. Gad and C. P. Cehngelis (Eds.), *Animal Models in Toxicology*, Marcel Dekker, New York, pp. 675–735.

Gad, S. C. (2000a). Pigs and ferrets as models in toxicology and biological safety assessment. *Int. J. Toxicol.* 19:149–168.

Gad, S. C. (2000b). Large animal toxicology: Introduction and general principles. *Int. J. Toxicol.* 19:129–132.

Gad, S. C. (2006). *Animal Models in Toxicology*, 2nd ed. Taylor and Francis, Philadelphia, PA.

Gad, S. C., Gad, S. D., and Gad, S. E. (2003). The dog functional observational battery (FOB) for use in pharmaceutical safety evaluation studies. *Int. J. Toxicol.* 22:1–8.

Galbraith, W. M., Hobson, W. C., Giclas, P. C., Schechter, P. J., and Agrawal, S. (1994). Complement activation and hemodynamic changes following intravenous administration of phosphorothioate oligonucleotides in the monkey. *Antisense Res. Dev.* 4:201–206.

Gulamhusein, A. P., Harrison-Sage, C., Beck, F., and Al-Alousi, A. (1980). Salicylate-induced teratogenesis in the ferret. *Life Sci.* 27:1799–1805.

Haggerty, G. C., Miller, G., Port, C., and Gad, S. (1989). Four-week oral toxicity study of diclofenac in the ferret. Paper presented at the National Meeting of the American College of Toxicology, Williamsburg, VA.

Hanhijarvi, H., Nevalainen, T., and Mannisto, P. (1985). A six-month dermal irritation test with anthralins in the Göttinger miniature swine. *Arch. Toxicol.* 8(Suppl.): 463–468.

Hart, J. E. (1986). The ferret as a replacement for the dog in toxicity studies. *Anim. Technol.* 37:201–206.

Heiser, H. J. (1970). *Atlas of Comparative Hematology*. Academic, New York.

Herman, E., Ferrans, V., and Young, R. (1988). Examination of minoxidil-induced acute cardiotoxicity in miniature swine. *Toxicology* 48:41–51.

Herman, E., Ferrans, V., Young, R., and Balazs, T. (1989). A comparative study in minoxidil-induced myocardial lesions in Beagle dogs and miniature swine. *Toxicol. Pathol.* 17:189–191.

Hoar, R. M. (1984). Use of ferrets in toxicity testing. *J. Am. Coll. Toxicol.* 3:325–329.

Hobson, W. (2000). Safety assessment studies in nonhuman primates. *Int. J. Toxicol.* 19:141–147.

Hunsaker, H., Norden, S., and Allen, K. (1984). An inexpensive caging method for miniature swine suitable for trace-element studies. *Lab. Anim. Sci.* 2:386–387.

Hutcheson, D., Tumbleson, M., and Middleton, C. (1979). Serum electrolyte concentrations in sinclair (S-1) miniature swine from 1 through 36 months of age. *Growth* 43:62–70.

Jackson, R. K., Kieffer, V. A., Sauber, J. J., and King, G. L. (1988). A tether-restraint system for blood collection from ferrets. *Lab. Anim. Sci.* 38:625–628.

Khan, M. (1984). Minipig: Advantages and disadvantages as a model in toxicity testing. *J. Am. Coll. Toxicol.* 3:337–342.

Koch, W., Windt, H., Walles, M., Borlak, J., and Clausing, P. (2001). Inhalation studies with the Göttinger minipig. *Inhalation Toxicol.* 13:249–259.

Lee, K. (1986). Swine as animal models in cardiovascular research. *Swine Biomed Res.* 3:1481–1496.

Lehrer, T. (1959). *An evening wasted with Tom Lehrer*. Lehrer Records TL202.

Line, S. W., Morgan, K. N., Markowitz, H., and Strong, S. (1989). Influence of cage size on heart rate and behavior in rhesus monkeys. *Am. J. Vet. Res.* 50:1523–1526.

Mannisto, P., Havas, A., Haasio, K., Hahnijarvi, H., and Mustakallio, K. (1984). Skin irritation by dithranol (anthralin) and its 10-acetyl analogues in three animal models. *Contact Dermatitis* 10:140–145.

Mazue, G., and Richez, P. (1982). Problems in utilizing monkeys in toxicology. In I. Bartosek (Ed.), *Animals in Toxicological Research*. Raven, New York, pp. 147–163.

Mclain, D., McCartney, M., Giovanetto, S., Martis, L., Greener, Y., and Youkilis, E. (1987). Assessment of the subchronic intravenous toxicity and disposition of ($^{14}$C) Acrolein in the rat and the acute and subchronic toxicity in ferrets. *Toxicologist* 7:208.

Miller, R. G., Jr. (1981). *Simultaneous Statistical Inference*, 2nd ed. Springer-Verlag, New York.

Moody, K. D., Bowman, T. A., and Lang, C. M. (1985). Laboratory management of the ferret for biomedical research. *Lab. Anim. Sci.* 35:272–279.

Morton, C., and Morton, F. (1985). In M. M. Vriends (Consult. Ed.), *Ferrets: A Complete Pet Owner's Manual*. Barron's Educational Series, Hauppauge, NY.

Mueller, D., Nelles, J., Deparde, E., and Arni, P. (1980). The activity of S-9 liver fractions from seven species in salmonella/mammalian-microsome mutagenicity tests. *Mutat. Res.* 70:279–300.

National Research Council (NRC) (1978). *Nutrient Requirements of Nonhuman Primates*. Panel on Nonhuman Primate Nutrition, NRC, Washington, DC.

National Research Council (NRC) (1985). *Nutrient Requirements of Dogs*. Subcommittee on Dog Nutrition, NRC, National Academy Press, Washington, DC.

National Research Council (NRC) (2004). *Use of Laboratory Animals in Biomedical and Behavioral Research*. Commission on Life Sciences, NRC, National Academy Press, Washington, DC.

Norris, W. P., Poole, C. M., Frye, R. J., and Kretz, N. D. (1968). *A Study of Thermoregulatory Capabilities of Normal, Aged and Irradiated Beagles*. Argonne National Laboratory, Chicago, IL., pp. 166–169.

Oldigs, B. (1986). Effects of internal factors upon hematological and clinical chemical parameters in the Göttinger miniature pig. *Swine Biomed. Res.* 2:809–813.

Olson, H., Betton, G., Robinson, D., Karluss, T., Monro, A., Kolaja, G., Lilly, P., Sanders, J., Sipes, G., Bracken, W., Dorato, M., Van Deun, K., Smith, P., Berger, B., and Heller, A. (2000). Concordance of the toxicity of pharmaceuticals in humans and in animals. *Reg. Tox. Pharmacol.* 32:56–67.

Panepinto, L. (1986). Laboratory methodology and management of swine in biomedical research. *Swine Biomed. Res.* 1:97–109.

Panepinto, L., and Phillips, R. (1986). The Yucatan miniature pig: Characterization and utilization in biomedical research. *Lab. Anim. Sci.* 36:344–347.

Parsons, A., and Wells, R. (1986). Serum biochemistry of healthy Yucatan miniature pigs. *Lab. Anim. Sci.* 36:428–430.

Peggins, J., Shipley, L., and Weiner, M. (1984). Characterization of age-related changes in hepatic drug metabolism in miniature swine. *Drug Metab. Dispos.* 12:379–381.

Perkin, C. J., and Stejskal, R. (1994). Intravenous infusion in dogs and primates. *J. Am. Coll. Toxicol.* 13:40–47.

Phillips, R., and Tumbleson, M. (1986). Models. *Swine Biomed. Res.* 1:437–440.

Radin, M., Weiser, M., and Frettman, M. (1986). Hematologic and serum biochemical values for Yucatan miniature swine. *Lab. Anim. Sci.* 36:425–427.

Rendic, S., Ruf, S., Weber, P., and Kajfez, F. (1984). Cimetidine and ranitidine: Their interaction with human and pig liver microsomes and with purified cytochrome P-450. *Eur. J. Drug Metab. Pharmacokinet.* 9:195–200.

Rettie, A., Bogucki, B., Lim, I., and Meier, P. (1990). Stereoselective sulfadioxidation of a series of alkyl P-tolyl sulfides by microsomal and purified flavin-containing monooxygenases. *Mol. Pharmacol.* 37:643–651.

Sambuco, P. (1985). Miniature swine as an animal model in photodermatology: Factors influencing sunburn cell formation. *Photodermatology* 2:144–150.

Schalm, D. W. (1975). Materials and methods for the study of the blood including brief comments on factors to be considered in interpretation. In N. C. Jain and E. J. Carroll (Eds.), *Veterinary Hematology*. Lea and Febiger, Philadelphia, p. 24.

Schneider, N., Bradley, S., and andersen, M. (1977). Toxicology of cyclomethylenetrinitramine: Distribution and metabolisms in the rat and the miniature swine. *Toxicol. Appl. Pharmacol.* 39:531–541.

Smith, G., Watkins, J., Thompson, T., Rozman, K., and Klassen, C. (1984). Oxidative and conjugative metabolism of xenobiotics by livers of cattle, sheep, swine and rats. *J. Anim. Sci.* 58:386–395.

Statistical Analysis System (SAS) (1996). SAS Institute, Cary, NC.

Swindle, M., Smith, A., and Hepburn, B. (1988). Swine as models in experimental surgery. *J. Invest. Surg.* 1:65–79.

Thakur, A. K. (2000). Statistical issues in large animal toxicology. *Int. J. Toxicol.* 19:133–140.

Thornton, P. C., Wright, P. A., Sacra, P. J., and Goodier, T. E. W. (1979). The ferret, *Mustela putorius furo*, as a new species in toxicology. *Lab. Anim.* 13:119–124.

Tynes, R., and Hodgson, E. (1984). The measurement of FAD-containing oxygenase activity in microsomes containing cytochrome P-450. *Xenobiotica* 14:515–520.

U.S. Department of Health and Human Services (DHHS) (1985). *Guide for the Care and Use of Laboratory Animals*. NIH Publication No. 86–23. DHHS, Washington, DC.

Van Ryzin, R., and Trapold, J. (1980). The toxicology profile of the anti-inflammatory drug proquazone in animals. *Drug Chem. Toxicol.* 3:361–379.

Van Vleet, J., Herman, E., and Ferrans, V. (1984). Cardiac morphologic alterations in acute minoxidil cardiotoxicity in miniature swine. *Exp. Mol. Pathol.* 41:10–25.

Walker, C., Bentley, P., and Oesch, P. (1978). Phylogenetic distribution of epoxide hydratase in different vertebrate species, strains and tissues measured using three substrates. *Biochem. Biophys. Acta* 539:427–434.

Watkins, J., and Klaassen, C. (1986). Xenobiotic biotransformation in livestock: Comparison to other species commonly used in toxicity testing. *J. Anim. Sci.* 63: 933–942.

Wilson, M. S., and Donnoghue, P. N. D. (1982). A mobile rack of cages for ferrets (Mustela putorius furo). *Lab. Anim.* 16:278–288.

# *Developmental and Reproductive Toxicity Testing*

## 1 INTRODUCTION

The goal of testing the developmental and reproductive toxicity of drug candidates in laboratory animals is to predict which agents would adversely affect the ability to achieve and maintain pregnancy and for normal development of offspring in humans and to allow evaluation of the potential risks to patients. This testing involves an extensive battery of studies based historically on guidelines promulgated by the U.S. Food and Drug Administration (FDA) in 1966 (see FDA, 1966, 1982, 1984; D'Aguanno, 1973) and subsequently modified by the International Conference on Harmonisation (ICH). These guidelines established three basic types of studies, segments I, II, and III, that are based on dosing during sequential phases of the reproductive cycle. These guidelines represented a dramatic increase in the extent and sophistication of testing expected of new drug candidates. The impetus for this intensified interest was the tragic epidemic of phocomelia and other congenital malformations caused in the early 1960s by the exposure of pregnant women to the sedative thalidomide. (For an excellent discussion of the history of the thalidomide tragedy, see pp. 228–249 in Schardein (1993)]. Table 1 presents the most recent guidelines.

The types of developmental and reproductive toxicity studies performed prior to 1993 and the methods used have been extensively documented (see Palmer, 1981; Christian, 1983; Heinrichs, 1985; Heywood and James, 1985;

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

**TABLE 1 Current Regulatory Guidelines—ICH, FDA: Medical Agents**

| | |
|---|---|
| ICH(2000) | Detection of toxicity to reproduction for medicinal products (ICH S5) |
| ICH(2000) | Detection of toxicity to reproduction for medicinal products (ICH S5) |
| FDA | International Conference on Harmonisation: Guideline on detection of toxicity to reproduction for medicinal products. *Fed.* Reg., September 22, 1994, Vol. 59, No. 183 |
| FDA | International Conference on Harmonisation: Guideline on detection of toxicity to reproduction for medicinal products; Addendum on Toxicity to make fertility. *Fed.* Reg., April 5, 1996, Vol. 61, No. 67 |

Persaud, 1985; Schardein, 1988; Tyl, 1988; Christian and Hoberman, 1989; and Manson and Kang, 1989; Khera, 1991). Since June 20, 1979, the FDA has required that these studies be conducted according to good laboratory practice (GLP) regulations (see FDA, 1978, 1987). The conduct of these studies had been complicated by the need to satisfy worldwide regulatory guidelines that varied from country to country. As a result, studies were conducted for regulatory purposes that, from a scientific viewpoint, were redundant, superfluous, and/or unnecessarily complex. This situation was changed in 1993 when the International Conference on Harmonisation of Technical Requirements for the Registration of Pharmaceuticals for Human Use standardized worldwide requirements in the guideline "Detection of Toxicity to Reproduction for Medicinal Products" (ICH, 1994, 2000a, b).

This chapter briefly describes the current standard study designs and then focuses on current issues in developmental and reproductive toxicity testing (also see Hood, 2006).

## 2 ICH STUDY DESIGNS

The ICH S5 A and B guideline allows for various combinations of studies. The studies conducted must include evaluation of the following components:

1. Male and female fertility and early embryonic development to implantation
2. Embryo–fetal development
3. Pre- and postnatal development, including maternal function

These components would normally be evaluated in a rodent species (preferably the rat), and, in addition, embryo–fetal development would be evaluated in a second species, typically the rabbit. The "most probable option" in the ICH guideline is the case where three rodent studies would be conducted that separately addressed each of the components listed above. These study designs are described below. Table 2 presents a comparison of ICH, FDA, European, and Japanese guidelines. The day of insemination or detection of evidence of mating is considered day 0 of gestation and the day of birth is

**TABLE 2 Comparison of ICH Stages and Study Types with Similar Observations Made**

| ICH Stage | FDA Guidelines | Great Britain and EEC Guidelines | Japanese Guidelines | EPA OPPTS, OECD, and FDA Redbook Guidelines |
|---|---|---|---|---|
| **A—Premating to conception:** reproductive functions in adult animals, including development and maturation of gametes, mating behavior, and fertilization | Segment I | Segment I | Segment I | Multigeneration One generation |
| **B—Conception to implantation:** reproductive functions in adult female, preimplantation and implantation stages of conceptus | Segment I | Segment I | Segment I | Multigeneration One generation Developmental toxicity |
| **C—Implantation to closure of the hard palate:** adult female reproductive functions and development of embryo through major organ formation | Segments I, II | Segments I, II | Segment II | Multigeneration One generation Developmental toxicity Developmental neurotoxicity |
| **D—Closure of the hard palate to the end of pregnancy:** adult female reproductive function, fetal development, and growth and organ development and growth | Segments I, II, III | Segments I, II | Segment II | Multigeneration One generation Developmental toxicity Developmental neurotoxicity |
| **E—Birth to weaning:** adult female reproduction function, adaptation of neonate to extrauterine life, including preweaning development and growth (postnatal age optimally based on postcoital age) | Segments I, III, pediatric | Segments I, II, III | Segments II, III | Multigeneration One generation Developmental toxicity Developmental neurotoxicity |
| **F—Weaning to sexual maturaity:** (pediatric evaluation when treated) postweaning development and growth, adaptation to independent life and attainment of full sexual development | Pediatric | Segment I | Segments II, III | Multigeneration Developmental neurotoxicity Developmental immunotoxicity |

*Note*: Bolded information indicates treatment interval.

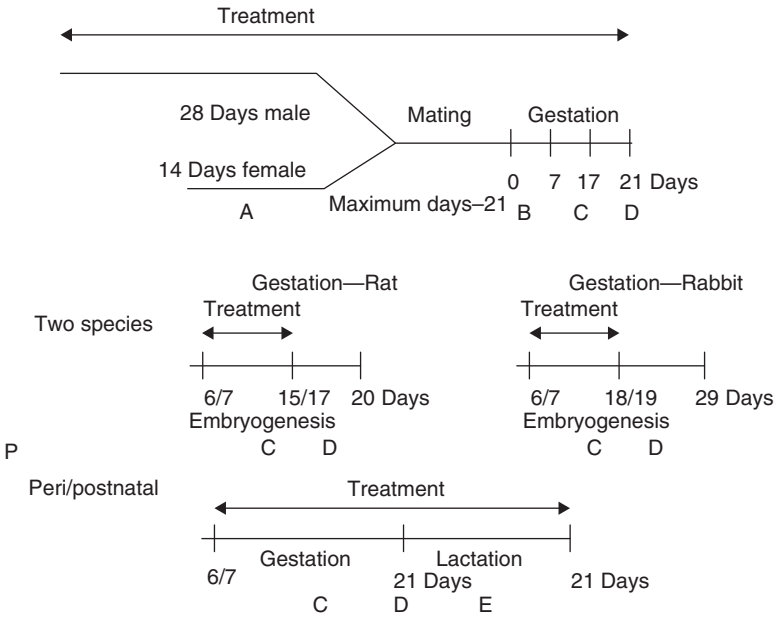


***Figure 1*** *Line charts for ICH stage study designs.*

considered postpartum and postnatal day 0. Figure 1 presents line charts for the ICH stage study designs.

## 2.1 Male and Female Fertility and Early Embryonic Development to Implantation

The purpose of this component is to assess the effects that result from treatment during maturation of gametes, during cohabitation, and, in females, during gestation up through the time of embryo implantation (typically last dose on day 6 of gestation). Assuming that the findings from a toxicity study of at least one month in duration do not contraindicate, the treatment period begins in males four weeks before male–female cohabitation and, in females, two weeks prior to cohabitation. A group size of 16–24 litters would generally be considered acceptable.

Minimal in-life observations include:

1. Clinical signs and mortality daily
2. Body weight twice weekly
3. Food consumption weekly
4. Vaginal cytology daily during cohabitation
5. Valuable target effects seen in previous toxicity studies

Females are sacrificed after the middle of the gestation period. Males are sacrificed at any time after the end of the cohabitation period, but it is generally advisable to retain the males until after the outcome of the first mating is known to ensure that a repeat cohabitation with untreated females will not be needed to determine if an observed effect on mating performance is a male effect. Males are treated until termination. Terminal examination of adults includes:

1. Necropsy
2. Preservation of organs with gross changes and sufficient control organs for comparison
3. Preservation of testes, epididymides, ovaries, and uteri
4. Sperm count and sperm viability
5. Count of corpora lutea and implantation sites
6. Count of live and dead conceptus

Among the study designs conducted before the ICH guidelines, the segment I fertility study conducted according to Japanese guidelines is most similar to this ICH study design. The major differences are the shortening of the treatment period of males prior to cohabitation from the duration of spermatogenesis (60–80 days) to 4 weeks and the addition of sperm evaluation. The justifications given for shortening the treatment period of males are:

1. Careful organ weight and histopathological evaluation of testes in general toxicity studies will detect most testicular toxins.
2. Fertility is an insensitive measure of testicular effects.
3. Compounds known to affect spermatogenesis generally exert their effects during the first four weeks of treatment.

Sperm counts can be performed with sperm from either the testis or the epididymis. Sperm motility is commonly being treated as a measure of sperm viability. The addition of sperm evaluation greatly increases the sensitivity of the study to detect effects on sperm maturation and the current study design will likely detect more male effects than previous designs even though the treatment period has been shortened.

### 2.2 Embryo–Fetal Development

The purpose of this component is to detect anatomical effects on the developing conceptus by treatment during the period of organogenesis from implantation to closure of the secondary palate. The study design is very similar to the historical segment II developmental toxicity study. A group size of 16–24 litters would generally be considered acceptable. The following is recommended:

| | Rat | Rabbit | Mouse |
|---|---|---|---|
| Treatment period, gestational days | 6–17 | 6–18 | 6–15 |
| Group size, mated or inseminated | 25 | 20 | 25 |

Minimal in-life observations include:

1. Clinical signs and mortality daily
2. Body weight twice weekly
3. Food consumption weekly
4. Valuable target effects seen in previous toxicity studies

Females are sacrificed at the end of the gestation period, about one day prior to parturition (day 20 or 21 for rats, day 28 or 29 for rabbits, and day 17 or 18 for mice). Terminal examinations include:

1. Necropsy
2. Preservation of organs with gross changes and sufficient control organs for comparison
3. Count of corpore lutea and live and dead implantations
4. Fetal body weight
5. External, visceral, and skeletal examination of fetuses
6. Gross evaluation of placenta

A minimum of 50% of fetuses are to be examined for visceral alterations and a minimum of 50% for skeletal abnormalities. When a fresh tissue microdissection technique is being used for the visceral examination of rabbit fetuses, all fetuses should be examined for both visceral and skeletal abnormalities.

Interpretation of results requires understanding and utilizing the following definitions:

*Malformation* Structural change that is likely to be permanent and detrimental to the survival or well-being of the fetus in the species/strain of animal being tested.

*Alternation* Change that is, in isolation, unlikely to be detrimental to the survival or well-being of the fetus in the species/strain of animal being tested.

*Variant* Observation occurring frequently in a particular strain of animal.

### 2.3 Adverse Effects

The following definitions should be referred to when considering whether an observed effect of treatment is adverse or not:

1. Treatment-related trend in incidence of specific or related malformations
2. Treatment-related increase in alterations, the cumulative effect of which is considered to be detrimental to the well-being of the fetus
3. Treatment-related increase in alterations, which are related in nature or derivation to treatment-related malformations evident on the study
4. Treatment-related marked change in the incidence of a group of alterations, which although their form is normal for a previous or future stage of development, that is, their occurrence suggests precocious or delayed development, their presence in a marked degree suggests some permanent change in the rate of development of the fetus and could be detrimental to its future development
5. Marked treatment-related increase in the occurrence of a specific alteration, in which the form is not predictive of the normal chronological order of development (e.g., bent scapula)

### 2.4 Pre- and Postnatal Development

The purpose of this component is to detect effects of treatment from implantation through lactation on the pregnant and lactating female and on the development of the conceptus and offspring through sexual maturity. The study design is similar to the previous segment III study design except that dosing begins on day 6 of gestation instead of day 15. A group size of 16–24 litters would generally be considered acceptable (with 25 mated females being recommended).

Minimal in-life observations for parental ($F_0$ generation) females include:

1. Clinical signs and mortality daily
2. Body weight twice weekly
3. Food consumption weekly
4. Valuable target effects seen in previous toxicity studies
5. Length of gestation
6. Parturition

Parental females are sacrificed after weaning of the $F_1$ generation. The age of sacrifice of the $F_1$ generation animals is not specified in the ICH guideline and varies among laboratories. Typically, they are sacrificed intermittently, with some laboratories reducing litter size on postnatal day 0, 3, or 4, on postnatal day 21 or at weaning, at male–female cohabitation to produce an $F_2$ generation, and the terminal sacrifice after production of the $F_2$ generation. Terminal examinations for maternal animals and offspring include:

1. Necropsy of all parental and $F_1$ adults
2. Preservation of organs with gross changes and sufficient control organs for comparison
3. Count of implantations

Additional observations of the $F_1$ generation include:

1. Abnormalities
2. Live and dead offspring at birth
3. Body weight at birth
4. Pre- and postnatal survival, growth, maturation, and fertility
5. Physical development, including vaginal opening and preputial separation
6. Sensory function, reflexes, motor activity, learning, and memory

## 2.5 Single- and Two-Study Designs for Rodents

Except for the embryo–fetal development component in rabbits, the components described above can be combined into fewer, larger studies instead of conducting each component separately. Acceptable alternatives include the "single-study design" and "two-study design." The choice may be made based on when study results are needed (how soon are females to be incorporated in clinical studies) and compound availability.

In the "single-study design," all of the above components are combined into one study. The dosing period, extending from before mating through lactation, is a combination of that for the fertility study together with that for the pre- and postnatal development study. Subgroups of animals are terminated at the end of gestation for fetal examination.

There are a variety of possible "two-study designs." One is to conduct the single study described above except that, instead of having subgroups for fetal examination, a separate embryo–fetal development study in rodents is conducted. Another two-study design consists of combining the embryo–fetal development study with the pre- and postnatal development study such that the two studies to be conducted would be (1) the fertility study and (2) the pre- and postnatal development study with subgroups terminated at the end of gestation for fetal examination. A third possible two-study design is to combine the fertility study with the embryo–fetal development study. In the first study, treatment would extend through the end of organogenesis and then, at termination at the end of gestation, there would be a complete fetal examination. The second study would be the pre- and postnatal development study.

For all the options described above, effects on male and female fertility can be evaluated separately by conducting separate studies in which only one sex is treated. The treatment periods are the same, but the treated animals are

cohabited with untreated animals of the opposite sex. In the male fertility study, the untreated females are terminated after the middle of gestation and terminal observations include embryo survival and possibly external examination of fetuses (if terminated at the end of gestation) (Tanimura, 1990). The advantage of conducting separate male and female studies is that, if there are effects, is it clear which sex was affected by treatment. Often when effects are seen in a combined male and female study, additional work is required to resolve which sex was affected. Either a second cohabitation of the treated males with untreated females is added or studies with only one sex treated must then be conducted.

With the possible exception of combining the female fertility component with the embryo–fetal development component, the combined-study approach is used often. The female fertility and embryo–fetal development components are needed to support clinical trials in women of childbearing potential in most countries and thus will be conducted early in the development of a drug. However, since the pre- and postnatal development component is not routinely required for clinical studies of women of childbearing potential and represents a large commitment of resources, it will not generally be conducted until late in the drug development process.

## 2.6 Preliminary Studies

According to the ICH guideline, "some minimal toxicity is to be expected to be induced in the high dose dams" in the reproductive toxicity studies. In some cases, particularly for the fertility and early embryonic development study, available information from general toxicity studies in the selected rodent species may be sufficient to allow the selection of dosage levels for a reproductive toxicity study with the goal of achieving minimal toxicity in high-dose dams. However, pregnant females sometimes respond differently to toxins than nonpregnant females, the duration of dosing for reproductive toxicity studies is different than for general toxicity studies, and toxicity may not have been achieved in the subacute toxicity studies. Thus, it is often necessary to conduct a range-finding study in pregnant rodents prior to the embryo–fetal development study. A range-finding study in rabbits is almost always required since only rarely are results available from other toxicity studies.

The range-finding study in pregnant animals (rodents or rabbits) is similar to the embryo–fetal development study discussed above except that there may be more dosage groups, group size is smaller (6–10 inseminated or mated females per group), and there is no need to examine fetuses for visceral or skeletal abnormalities. Evaluating litters from range-finding studies for resorption, fetal weight, and external abnormalities is valuable for providing an early indication of marked developmental toxicity. This is particularly important if conceptus survival at a particular dosage level would be inadequate to evaluate effects on development in the subsequent embryo–fetal development study. Once it has been determined during a range-finding study that a particu-

lar dosage level causes toxicity exceeding the minimal toxicity desired for the embryo–fetal development study, it is best to terminate that dosage group since continued treatment and evaluation unnecessarily expose animals to toxicity, any subsequent data collected are not useful for risk assessment (since it is known that excessive maternal toxicity itself causes developmental toxicity), and investment of resources is therefore unwarranted.

## 2.7 Toxicokinetics

The ICH guidelines do not require that toxicokinetic studies be conducted except that "at the time of study evaluation further information on kinetics in pregnant or lactating animals may be required according to the results obtained." In addition, the guidelines state that "it is preferable to have some information on kinetics before initiating reproduction studies." In practice, however, at least some degree of toxicokinetic sampling and analysis is performed to verify exposure and allow risk assessment.

The major toxicokinetic issue for reproductive toxicity studies is whether systemic exposure in the selected species and route is adequate relative to the systemic exposure with the clinical regimen. Often, this information is available for the selected rodent species from studies conducted independently from the reproductive toxicity studies. For rabbits, though, there is rarely toxicokinetic information available from other studies. Accordingly, it is advisable to conduct at least a crude evaluation of systemic exposure in the rabbit. It is best if theses data are available prior to the embryo–fetal development study so that, if the rabbit is found to have inadequate systemic exposure, an alternative species may be selected before the investment of resources in a large rabbit study. The collection of blood samples for toxicokinetic evaluations may be incorporated into the range-finding study in pregnant rabbits. However, rabbits stressed by multiple bleedings should not be retained for evaluation of developmental toxicity and satellite groups of toxicokinetic animals for bleeding only may be needed.

It would be ideal to have data from five to eight time points following the first and last doses to examine accumulation and other changes in kinetic parameters during pregnancy and, since physiology changes rapidly during gestation, to have data periodically during gestation as well. However, from a practical point of view, the question being asked (what is the approximate systemic exposure?) does not justify a comprehensive kinetic evaluation. When circumstances dictate that a toxicokinetic evaluation be performed, determining maternal plasma levels at a few postdosing intervals during a single 24-h period of gestation, preferably during the period when serious adverse effects are most likely to be induced (days 9–12 of gestation), will generally provide adequate information.

Only in special circumstances will the determination of embryo levels of drug add meaningfully to the assessment of human risk from a drug. In such studies, even if it is found that the embryo is not exposed, the lack of exposure

of the embryo would not necessarily indicate an invalid study or increased human risk since there may also be no exposure in human embryos. When embryo level studies are conducted, the selection of day(s) of gestation to harvest embryos is severely restricted by the sensitivity of the assay. Often, the earliest day that allows the collection of sufficient tissue for assay is gestational day 10 or 11.

## 2.8 Timing of Studies

The definition of which studies need to be performed in advance of clinical trials has not been addressed yet by the ICH process and is currently monitored by the regulatory agencies of individual countries and institutional review boards (IRBs). Embryo–fetal development studies in two species are almost universal prerequisites for clinical studies in women of childbearing potential. Some regulatory agencies also request that a fertility study in female rodents be conducted before clinical trials in women of childbearing potential. A fertility study in male rodents is required before clinical trials in men in Japan. Some pharmaceutical companies have internal guidelines that specify compliance with all the guidelines listed above, regardless of the location of the clinical trials.

The most conspicuous exception to the policy described above is the position of the FDA (1993). The FDA withdrew the restriction on the participation of women of childbearing potential in early clinical trials, citing "(1) exclusion of women from early trials is not medically necessary because the risk of fetal exposure can be minimized by patient behavior and laboratory testing, and (2) initial determinations about whether that risk is adequately addressed are properly left to patients, physicians, local IRBs and sponsors with appropriate review and guidance by FDA, as are all other aspects of the safety of proposed investigations." The policy of excluding women has been replaced by one that specifies that "the patients included in clinical trials should, in general, reflect the population that will receive the drug when it is marketed." In fact, inclusion of women at the earliest possible stages is frequently mandated. This led to the current situation where requirements as to when developmental and reproductive testing must be performed are very different between ICH regions, with FDA requirements generally not mandating any such studies until after phase I unless a potential risk/concern is indicated by a drug mechanism or finding in other studies (most commonly the initial repeat-dose toxicity studies).

To comply with FDA policy, at least for the conduct of clinical trials in the United States, pharmaceutical companies have a few choices. They can conduct the standard battery of reproductive studies prior to enrolling women of childbearing potential in early clinical trials. The possible negative impact would be a delay in the initiation of clinical trials. Alternatively, pharmaceutical companies can enroll women of childbearing potential in early clinical trials without having conducted any reproductive toxicity studies and accept

the additional risk resulting from exposure to untested drugs during inadvertent or undetected pregnancy. In either case, the incidence of pregnancy during clinical trials can be decreased by pregnancy testing and/or assurances of contraception.

## 3 METHODOLOGICAL ISSUES

### 3.1 Control of Bias

An important element to consider when designing developmental and reproductive toxicity studies is the control of bias. For example, animals should be assigned to groups randomly and preferably blocked by body weight. This can be accomplished by first ranking the animals in order of body weight and then, starting with the lightest or heaviest, assigning by rank to groups based on a list of sets of random permutations of numbers (e.g., 1, 2, 3, and 4 if there are four groups, where 1 represents the control group, 2 represents the low-dose group, etc.). Housing of treatment groups should also be unbiased. This can be done by a "Latin square" design where each block of four cages (if there are four groups) includes an animal from each group. It is often an acceptable compromise to have animals from different groups in alternating vertical columns with all the animals in a column from the same group. This provides equal vertical balancing for all groups. Historically, it has proven unwise to have groups segregated on separate racks.

The order of sacrifice on the day of cesarean sectioning should be balanced by group (again using random permutations) since fetuses continue to grow during the day and an unbalanced time of sacrifice would bias fetal weights, particularly for rodents. Alternatively, all animals can be killed at about the same time in the morning and the fetuses stored for examination later the same day. Fetal examinations should be conducted blind, that is, without knowledge of treatment group.

### 3.2 Diet

It is known that rodents require a diet relatively rich in protein and fats for successful reproduction (Zeman, 1967; Chow and Rider, 1973; Turner, 1973; Mulay et al., 1982). Consequently, rodents are fed high-protein, high-fat diets ad libitum for reproductive toxicity studies and also generally as a maintenance diet for all toxicity studies. Female rats fed in this manner begin to show decreases in fertility, litter size, and the incidence of normal estrus cycling at the age of six months (Matt et al., 1986, 1987). The disadvantage of this feeding practice is that the animals more quickly acquire age-related diseases and sexual dysfunction and die sooner than if they are fed a restricted amount of calories (for review, see Weindruch and Walford, 1988). In relatively short-term studies (such as standard ICH studies), this rapid aging does not present a

problem. However, for male breeding colonies or multigeneration studies with multiple litters per generation, it could be advantageous to restrict caloric intake, at least when the animals are not being bred. Restriction of food intake to achieve a 30% decrease in body weight gain compared to ad libitum–fed controls has no adverse effect on male rat reproduction (Chapin et al., 1991), although it does affect reproduction in mice (Gulati et al., 1991) and female rats (Chapin et al., 1991).

Dietary restriction is even more important for rabbits. Rabbits fed ad libitum fare very poorly. Some laboratories restrict New Zealand white rabbits to 150–180 g per day of a high- (at least 13.5%) fiber diet. However, even this regimen results in some rabbits going off feed late in gestation. It has been observed that by restricting New Zealand white rabbits to only 125 g of food per day nearly all control animals retain appetite throughout gestation and fewer of these animals abort (Clark et al., 1991). More uniform food consumption late in gestation is associated with greater uniformity in maternal body weight change and fetal weight. This decreased variability makes these measures more sensitive indicators of maternal and developmental toxicity. Thus, 125 g is the preferred daily ration for New Zealand white rabbits.

### 3.3 Clinical Pathology

Regulatory guidelines require that there be maternal toxicity at the highest dosage level in embryo–fetal developmental toxicity studies. It is important to avoid excessive toxicity in these studies since it is known that marked maternal toxicity can cause secondary developmental toxicity (see discussion in Section 4.3). This secondary developmental toxicity is irrelevant to the assessment of the developmental hazard of the test agent and thus simply confounds the interpretation of the data.

The traditional indicators of maternal toxicity in range-finding studies in pregnant animals (mortality, body weight, food consumption, and clinical signs) do not always provide a sensitive measure of toxicity. This insensitivity is a particular problem for rabbit studies since typically no other toxicity studies are conducted in rabbits and body weight change in rabbits is very variable (typically −100 to +400 g during gestation), making it a particularly insensitive indicator of toxicity.

Thus, it is desirable to improve the assessment of toxicity in range-finding studies in pregnant animals. Complete histopathological examination is not practical. However, it is often feasible to perform hematological and serum biochemical analyses that can significantly increase the changes of detecting significant toxicity and provide important information for selecting an appropriate highest dosage level for the embryo–fetal developmental toxicity study.

Based on more than 20 years of experience, body weight effects most often provided the basis for the selection of dosage levels in the segment II study. However, there have been cases where clinical pathology was or would have been useful to justify dosage selection. For example, the nonsteroidal anti-

inflammatory drug diflunisal caused a decrease in erythrocyte count from $6.0 \times 10^6$ to $2.9 \times 10^6\,mm^{-3}$ at a dosage level ($40\,mg\,kg^{-1}\,day^{-1}$) that caused only a 1% decrease in body weight in pregnant rabbits. The severe hemolytic anemia caused by this excessively high dosage level in turn caused secondary axial skeletal malformations in the fetuses (Clark et al., 1984). Also, the angiotensin-converting enzyme (ACE) inhibitor enalapril caused an increase in serum urea nitrogen from 16 to $46\,mg\,dL^{-1}$ (highest value 117) at a dosage level ($10\,mg\,kg^{-1}\,day^{-1}$) that had no apparent effect on body weight but caused a significant ($p < 0.05$) increase in resorptions (Minsker et al., 1990). Serum urea nitrogen concentration was used to select dosage levels for a subsequent ACE inhibitor, lisinopril. Likewise, the routine use of clinical pathology in range-finding studies has previously been proposed (Wise et al., 1988). The animals can be bled on the day after the last dose or sooner to detect transient effects or to allow an evaluation of the data prior to cesarean section.

### 3.4 Gravid Uterine Weights

Effects of treatment on maternal body weight gain are commonly evaluated as indicators of maternal toxicity. However, maternal body weight gain is influenced by fetal parameters such as live fetuses per litter and fetal body weight. Thus, effects indicative of developmental toxicity could contribute to decreased maternal body weight gain and confound the interpretation of maternal toxicity. In addition, other maternal but pregnancy-related parameters, such as volume of intrauterine fluid, could be affected by treatment and contribute to effects on overall body weight gain.

In an attempt to correct this complication, some laboratories weigh the gravid uterus at cesarean section and then subtract the weight of the gravid uterus from the body weight gain to obtain an adjusted weight gain that is more purely maternal. This adjustment is imprecise but not inappropriate for rats for which gravid uterine weight is correlated with and generally substantially less than maternal body weight change during gestation (e.g., see Figure 2 for which the correlation coefficient $r = 0.63$ and $p < 0.001$). However, the subtraction of gravid uterine weight from maternal weight gain is an overadjustment for rabbits. The maternal body weight gain of rabbits during gestation is generally less than the weight of the gravid uterus (see Figure 3). Moreover, gravid uterine weight is correlated with maternal body weight change in some but not all studies. For example, in the 53 untreated rabbits from the study shown in Figure 3, $r = 0.54$ and $p < 0.001$. However, in a study of 32 rabbits treated with a methylcellulose vehicle, $r = 0.21$ and $p = 0.25$. Thus, subtracting the gravid uterine weight from the maternal weight gain is not always appropriate. A preferred method for adjusting maternal body weight gain for possible developmental effects is to test and, if appropriate, use gravid uterine weight as a covariate (J. Antonello, personal communication, 1990). This method can be used for both rats and rabbits and for body weight change intervals in addition to those ending at study termination.

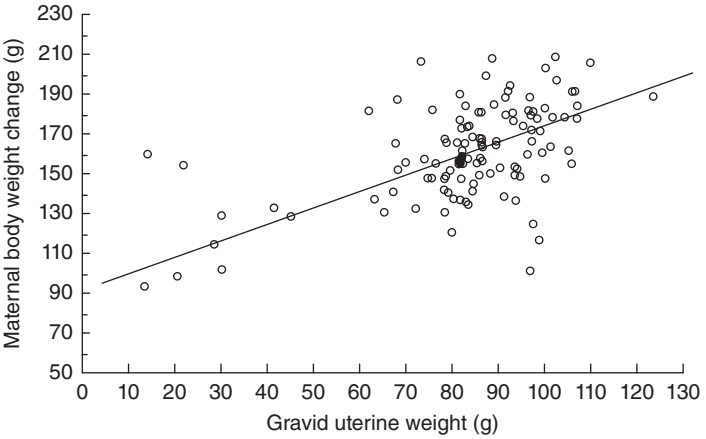


***Figure 2*** *Relationship between gravid uterine weight and maternal body weight change in control rats between days 0 and 20 of gestation. One hundred and twenty pregnant Sprague-Dawley [Crl:CD(SD)BR] rats were dosed orally with 0.5% aqueous methylcellulose on days 6–17 of gestation and cesarean sectioned on day 20 of gestation. The gravid uterus from each animal was removed and weighed.*

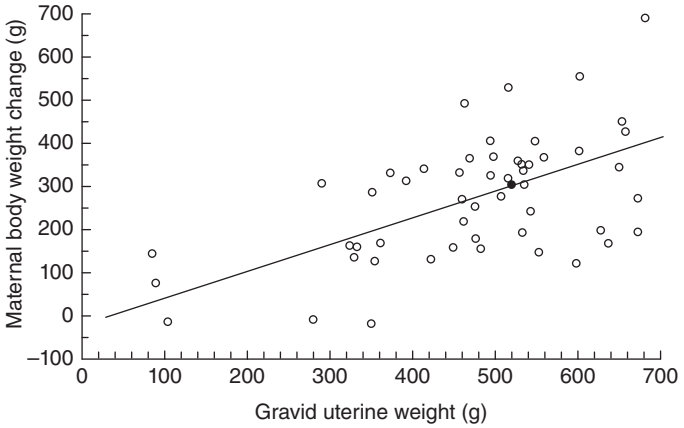


***Figure 3*** *Relationship between gravid uterine weights and maternal body weight change in untreated rabbits between days 0 and 28 of gestation. Fifty-three pregnant New Zealand white rabbits that had not been treated with control article or test agent were cesarean sectioned on day 28 of gestation. The gravid uterus from each animal was removed and weighed.*

Alternatively, to avoid weighing the uterus (or if the analysis is being performed retrospectively and uterine weights are unavailable) or if a more purely fetal adjustment is desired, one can use the sum of the live fetal weights within the litter (total live fetal weight) as the covariate instead of gravid uterine weight. As expected, total live fetal weight is very highly correlated with gravid uterine weight in control animals ($r$ is 0.99 in control rats and 0.95

in control rabbits; J. Antonello, personal communication, 1990). Thus, in general, using either gravid uterine weight or total live fetal weight as the covariate will yield similar results. However, if treatment was to have an effect on gravid uterine weight that was not reflected in total live fetal weight (e.g., if the volume of amniotic, extracoelomic, or intrauterine fluid was affected), then total live fetal weight may not be highly correlated with gravid uterine weight and, hence, not interchangeable as a covariate. In that case, only weighing the gravid uterus would allow the detection of these effects not revealed by total live fetal weight.

### 3.5 Implant Counts and Determination of Pregnancy

Two observations suggest that the remnants of embryos that die soon after implantation are not apparent at gross examination of the uterus near term. First, embryos that were observed to be resorbing at laparotomy early in gestation left no readily visible trace near term (Staples, 1971). Second, occult implantation sites can be revealed near term by staining the uterus with ammonium sulfide or sodium hydroxide (Salewski, 1964; Yamada et al., 1988). It is not known if the uterine staining techniques reveal all implantation sites. It is clear, though, that when uterine staining techniques are not used, very early resorptions may not be included in what is termed the "resorption rate" but instead may contribute to the apparent "preimplantation loss" or, if no implantation sites were detected, the rate of "nonpregnant" animals.

In normal circumstances, probably very few implantation sites are not detected without staining. However, cases have occurred in which treatment effects were probably detected only as a result of uterine staining. For example, in one rabbit study with drug treatment starting on day 6 of the gestation, a drug-treated group had four litters that had implantation sites that were seen only after staining with ammonium sulfide, indicating very early drug-induced resorption. For critical studies in rabbits designed to determine early effects on resorption and abortion rates, it would be advantageous to measure plasma levels of progesterone on day 6 of gestation since low levels indicate nonpregnant animals (Adams et al., 1989, 1990).

### 3.6 Fetal Examinations

Many fetal anomalies, such as cleft palate, exencephaly, ectrodactyly, and missing vertebra, are discrete and distinct and therefore easy to recognize objectively. Some anatomical structures, though, occur along a continuous gradation of size and shape and are only considered anomalous if the deviation from the typical exceeds a somewhat arbitrarily selected threshold. These anomalies are observed in all examination types and include, for example, micrognathia, reduced gallbladder, enlarged heart, distended ureter, wavy rib, and incomplete ossification at many sites. In many cases, it cannot be said with certainty whether a specific degree of variation from normal would have

resulted in an adverse consequence to the animal and should therefore be considered abnormal. In the absence of certainty about outcome, the best approach is to uniformly apply a single criterion within a study (and preferably among studies) so that all treatment groups are examined consistently. The subjectivity (and hence fetus-to-fetus variability) of the examination can be minimized by having the criteria be as clear and objective as possible. For example, when examining for incompletely ossified thoracic centra or supraoccipitals, it can be required that the ossification pattern be absent (unossified), unilateral, or bipartite (which are objective observations) before recording as an observation. Subjective criteria such as being dumbbell or butterfly shaped would not be applied.

***Examination of External Genitalia*** One aspect of external anatomy that is largely overlooked in the examination of offspring exposed in utero to test agents is the external genitalia, even though major malformations can occur in those structures. For example, hypospadias are malformations in the male in which the urethra opens on the underside of the penis or in the perineum. Hypospadias can occur in the male rat following in utero exposure to antiandrogens (e.g., Neumann et al., 1970), testosterone synthesis inhibitors (e.g., Bloch et al., 1971), or finasteride, a 5α-reductase inhibitor (Clark et al., 1990b). However, it is impractical to detect hypospadias in fetuses or young pups. Although the genital tubercle of the normal male rat fetus is grossly distinguishable from that of the normal female as early as day 21 of gestation (the female has a groove on the ventral side), the difference is very subtle and partial feminization of the male genital tubercle would be very difficult to ascertain. Routine histological examination is obviously too labor intensive to be considered. Hypospadias can readily be determined, though, by expressing and examining the penis of the adult. Thus, it is recommended that adult $F_1$ males be examined for hypospadias. If the timing of the separation of the balano–preputial membrane is being included in the pre- and postnatal development study as a developmental sign (see Korenbrot et al., 1977), the examination of the penis for hypospadias can be conducted at the same time.

The critical period for the induction of hypospadias by finasteride in rats is days 16–17 of gestation (Clark et al., 1990a). It is unlikely that other agents would have a much earlier critical period since testosterone synthesis, which is required for the development of the penile urethra, begins in the rat on day 15 of gestation (Habert and Picon, 1984). Thus, if treatment in the embryo–fetal development study terminates on day 15 of gestation (as is done in some laboratories), it is doubtful that hypospadias could be induced. However, hypospadias could be induced in the pre- and postnatal development study. Since the formation of the penile urethra in the rat is not completed until day 21 of gestation (Anderson and Clark, 1990), it could be argued that "major organogenesis" continues until that time.

One parameter that is readily and commonly measured as an indicator of effects on differentiation of the external genitalia in rodent fetuses is the sexu-

ally dimorphic distance between the anus and the genital tubercle (anogenital distance). However, it should not be assumed that anogenital distance is synonymous with hypospadias, since effects on anogenital distance are not necessarily predictive of hypospadias. Finasteride caused both hypospadias and decreased anogenital distance in male offspring but with very different dose–response relationships and only a slight tendency for animals with hypospadias to have a shorter anogenital distance (Clark et al., 1990b). Also, the effects on anogenital distance were largely reversible, whereas hypospadias were not. Another agent, triamcinolone acetonide, caused dramatic (reversible) decreases in anogenital distance in male rat fetuses on day 20 of gestation but did not affect the development of the genital tubercle and did not cause hypospadias (Wise et al., 1990b). Thus, decreased anogenital distance per se does not necessarily indicate a serious congenital anomaly.

When evaluating effects of treatment on fetal anogenital distance, it is obviously important to correct for effects on fetal weight. One approach is to calculate "relative" anogenital distance, the ratio between anogenital distance and another linear measure, for example, biparietal diameter (head width). The cube root of fetal weight simulates a linear measure (Wise et al., 1990b) and can also be used to normalize anogenital distance. Another approach is to compare the anogenital distance in a weight-reduced treatment group to that in a weight-matched control group at a younger age.

***Visceral Fetal Examinations*** The examination of the abdominal and thoracic viscera of fetuses is performed either fresh without fixation ("Staples technique") or after Bouin's fixation by making freehand razor blade sections ("Wilson's technique"; Wilson, 1965). Both techniques have advantages. The fresh examination technique, which may require less training to become thoroughly proficient, provides a more easily interpreted view of heart anomalies. The examination must be performed on the day the dam is terminated, though, so having a large number of litters to examine in one day requires that a large team of workers be committed to the task.

With both techniques, the heads of one-half of the fetuses can be fixed in Bouin's fixative for subsequent freehand sectioning and examination. A common artifact induced by fixation in rabbit fetal heads is retinal folding.

Whether or not the kidneys are sliced transversely to examine the renal pelvis varies among laboratories. Hydronephrosis, delayed papillary development, and distended renal pelvis are most readily detected in this manner. However, it is not necessary to slice the kidneys to detect the urinary retention that can lead to distended renal pelvis and hydronephrosis. This point was demonstrated in a study in which 200,000 IU $kg^{-1}$ $day^{-1}$ of vitamin A administered orally on days 8–10 of gestation induced hydronephrosis and/or distended renal pelvis in 29 fetuses (R. Clark, personal communication). In all of these 29 fetuses (and two others), a distended ureter also occurred. Thus, a distended ureter may be a more sensitive indicator of urinary retention than a distended renal pelvis.

***Skeletal Fetal Examination*** There is variability in the development of the fetal skeleton, including numbers of vertebrae and ribs, patterns of sternebral ossification, alignment of ribs with sternebrae, and alignment of ilia with lumbar and sacral vertebrae. There is also extensive plasticity in the development of the skeleton beyond the fetal stage. For example, it is known that markedly wavy ribs in fetuses can resolve so that the ribs in the adult are normal (Saegusa et al., 1980; Nishimura et al., 1982) and supernumerary ribs can be resorbed (Wickramaratne, 1988). This variability and plasticity complicates the classification of anomalies as true malformations as opposed to variations of normal. There is no unanimity on terminology, but, in general, a variation tends to be an alteration that occurs at relatively high spontaneous incidence (>1%), is often reversible, and has little or no adverse consequence for the animal.

When tabulating and interpreting fetal skeleton data, a distinction is made between alterations in the pattern of development and simple delays in development that are considered to be less serious. A delay in skeletal development is usually apparent as a delay in ossification, as evidenced by an increased incidence of specific, incompletely ossified sites or decreases in counts of ossified bones in specific regions (e.g., sacrocaudal vertebrae). These delays are normally associated with decreases in fetal weight and commonly occur at dosage levels of the test agent that also cause decreased maternal body weight gain.

When determining the criteria for recording skeletal alterations, particularly sites of incomplete ossification, it is legitimate to consider the resulting incidences. For example, including an unossified fifth sternebra in the criteria for recording incomplete sternebral ossification may increase the control incidence to a very high proportion (over 95%) of fetuses affected, which would then reduce the sensitivity for detecting treatment effects. The additional effort expended in recording the extra observations due to sternebra 5 would be wasted. In addition, recording high incidences of incomplete ossification at many sites is not worth the effort involved. The ossification at various sites is highly correlated, so recording at multiple sites is redundant. In some cases, the incidences can be reduced to reasonable levels (1 to 20% of control fetuses) and the criteria simultaneously being made more objective by requiring that the bone be entirely unossified before recording.

### 3.7 Developmental Signs

The postnatal evaluation of $F_1$ pups includes the observation of developmental signs in two or more pups per sex per litter. In general, the acquisition of these developmental landmarks, including anatomical changes (e.g., ear pinna detachment, incisor eruption, hair growth, and eye opening) and reflexes (negative geotaxis, surface righting, and free-fall righting), are highly correlated with body weight, but as indicators of developmental toxicity they are not as sensitive as body weight (Lochry et al., 1984; Lochry, 1987) and thus have

minimal value. Possible exceptions to this generality are the ontogeny of the auditory startle reflex and the markers of sexual maturation (vaginal patency, testes descent, and balano–preputial separation in males).

The examinations for developmental signs should be performed daily starting before and continuing until the criterion is achieved. The separation of the balano–preputial membrane of the penis (occurring at postnatal weeks 6–7; Korenbrot et al., 1977) is becoming the preferred landmark of sexual maturation in males. The timing of the testes descent is more variable and very dependent on the achievement criteria used. Another advantage of determining the time of the balano–preputial separation is that anomalies of the penis may be observed at the same time (as noted above).

### 3.8 Behavioral Tests

The trend within reproductive toxicology is to move from simple determinations of developmental landmarks and reflexes to more sophisticated and sensitive behavioral tests. This process was accelerated by the U.S. Environmental Protection Agency (EPA), which issued guidelines requiring a "developmental neurotoxicity" study of compounds that meet any of several broad criteria (EPA, 1991). The behavioral tests to be performed in this study are extensive and rigidly defined. As laboratories become equipped and trained to meet these guidelines, they are adding such tests to their evaluations of pharmaceuticals. The suggestions for routine testing made below are considered reasonable for pre- and postnatal development studies intended as routine screens. It is suggested that testing be conducted on one or two adults per sex per litter, keeping the range of actual ages as tight as possible.

Measurement of motor activity is commonly performed in the dark in cages or plastic boxes (open field) or residential mazes in which movement is quantitated by infrared detectors or by recording the interruption of light beams as the test subject moves through a horizontal grid of light beams. Possible parameters to evaluate include horizontal activity (light beams interrupted), number of movements, and time spent in the middle of the cage. The test period is selected to be long enough (normally 30–50 min) to allow the activity of the animals to decrease to an approximately constant level (asymptote). Testing of young pups (e.g., 13 days of age) is not recommended as their activity level is fairly constant during the test period and young unweaned pups should not be separated from their mothers for extended periods of time.

Another test paradigm for detecting treatment effects on brain functioning in $F_1$ offspring measures auditory startle habituation. In this test, the animal is placed in a chamber with a floor that detects movement. The animal is exposed to a sequence of 50–60 auditory stimuli, each at 110–120 decibels for 20–50 s and separated by 5–20 s. The gradual diminution of the animal's movement response is indicative of normal habituation.

There is not a consensus about the procedures to use to test for effects on learning and memory. The two most commonly used techniques are the water-

filled maze, which is preferred for measuring learning, and passive avoidance, which is preferred for measuring memory (see Buelke-Sam et al., 1985). Retention is tested in a repeat test conducted approximately one week later.

### 3.9 Detecting Effects on Male Reproduction

Male fertility studies with typical group sizes (15–30 males per group) are very insensitive for detecting effects on male fertility. If the control fertility rate is 80%, even a group size of 30 will only detect (at the 5% significance level) a 38% decrease in fertility 80% of the time and a 50% decrease 95% of the time (J. Antonello, personal communication, 1990). To detect slight effects on male fertility would require enormous group sizes. Mating each male with more than one female provides a more precise estimate of the reproductive capacity of each male but does not greatly increase statistical power. If multiple matings are to be done, it is recommended that the cohabitations with multiple females be sequential rather than concurrent.

It is difficult to detect effects on male fertility not only because of group size considerations, but also those mediated by decreased sperm production because of the normally huge excess of sperm included in a rat ejaculate. Sperm production can be decreased by up to 90% without effect on fertility (either pregnancy rate or litter size) in the rat. This is not the case for men, so the sperm excess in the rat represents a serious flaw in the rat model (see Working, 1988). To address this deficiency and improve the sensitivity of the model, it is advisable to determine the effects of the test agent on testes weights, testicular spermatid counts, and histopathology of the testes (preferably plastic sections) in the male fertility study and/or the 14-week toxicity study. In some cases, these parameters may be more predictive of possible effects on male fertility in humans than the fertility rate in rats.

## 4 DATA INTERPRETATION

### 4.1 Use of Statistical Analyses

Statistical analysis is a very useful tool for evaluating the effects of treatment on many developmental and reproductive toxicity parameters. For some parameters, such as maternal body weight changes, fetal weight, and horizontal activity in an open field, the comparison to the concurrent control is the primary consideration and, assuming adequate group size, the investigator relies heavily on the results of appropriate statistical analyses to interpret differences from control.

For other parameters, though, statistical analysis is just one of several considerations that include historical control data and other relevant information about the test agent and related test agents. For example, statistical analysis of a low incidence of an uncommon fetal malformation will usually not be significant ($p > 0.05$) even if treatment related, due to the low power for detect-

ing such effects with typical group sizes. In such cases, examination of the historical control data becomes paramount. If two fetuses with a particular malformation occur in separate litters only in a high-dose group, the finding is of much more concern if it is a very rare malformation than if recent historical control groups have had a few fetuses with that malformation.

Other known effects of the test agent or related agents also sometimes contribute to data interpretation. For example, a low incidence of a malformation may be considered treatment related if it is at the low end of a typical dose–response curve or if it is in a high-dose group and that malformation is an expected effect of the test agent. In general, though, a single occurrence of a fetal malformation in a treatment group (with none in control) is not cause for alarm, since this occurs in almost every study (together with occurrences of some malformations only in the control group).

Statistical methods exist to appropriately analyze most developmental and reproductive toxicity parameters. Exceptions to this are the "*r/m*" litter parameters in which, for each litter, there is a number affected divided by the number in the litter. These parameters include preimplantation loss ($r$ = corpora lutea without implants, $m$ = corpora lutea), resorption rate ($r$ = resorptions, $m$ = implants), and the family of alteration rates ($r$ = affected fetuses, $m$ = fetuses). There are two factors complicating the statistical analysis of these data that have heretofore been inadequately handled (Clark et al., 1989). One is that almost all of these parameters have a strong dependence on $m$. For example, both preimplantation loss (Figure 4) and resorption rate (Figure 5) are normally higher at both the low and high extremes of $m$. In contrast, supernumerary rib tends to occur at higher incidences in average-size litters (Figure 6). The second factor that complicates the statistical analysis of *r/m* data is that affected implants tend to occur in clusters within litters ("litter effects"); that is, the intralitter correlation is greater than the interlitter correlation. For example, the total number of litters affected with anasarca, missing vertebra, and supernumerary rib is much less than would be expected by change based on the number of affected fetuses (Table 3 and Figure 7).

These problems have been resolved for analysis of resorption rate (and preimplantation loss) in Sprague-Dawley rats using a three-step process (Soper and Clark, 1990). First, based on an analysis of data from 1379 control rat litters examined since 1978, a likelihood score was derived for each ($r,m$) couplet based on the incidence of that couplet given that value of $m$. These scores were approximately equal to $r$. Second, an analysis of 136 litters from groups with slight effects on resorption rate revealed that, at low-effect doses of embryocidal test agents, the increases in resorptions tended to occur as increased numbers of resorptions within affected litters rather than as an increased proportion of affected litters. To maximize the difference in scores between control and affected litters, the scores for controllike litters ($r$ = 1, 2, 3) were downgraded from $r$ (1, 2, and 3) to 0.4, 1, and 2.4, respectively. Third, to arrive at the final score for each litter, the modified $r$ score for each litter

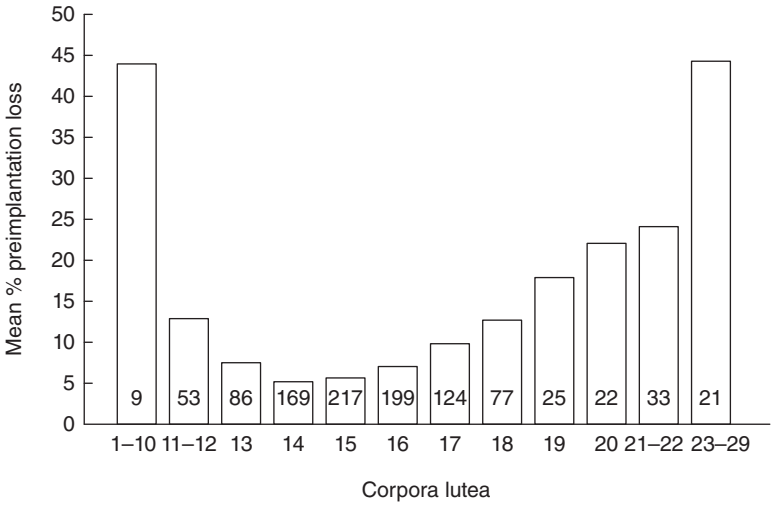


***Figure 4*** *Effect of litter size on mean percentage preimplantation loss in 1035 control rat litters. Between 1970 and 1988, 1035 control rats were cesarean sectioned on day 20 of gestation and the numbers of resorptions and implants were counted. Numbers within the bars indicate number of litters.*

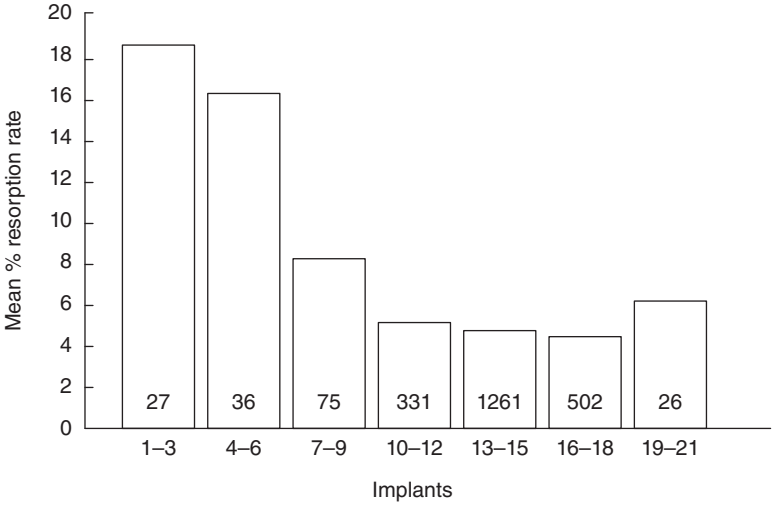


***Figure 5*** *Effect of litter size on mean percentage resorption rate in 2258 control rat litters. Between 1970 and 1988, 2258 control rats were cesarean sectioned on day 20 of gestation and the numbers of resorptions and implants were counted. Numbers within the bars indicate number of litters.*

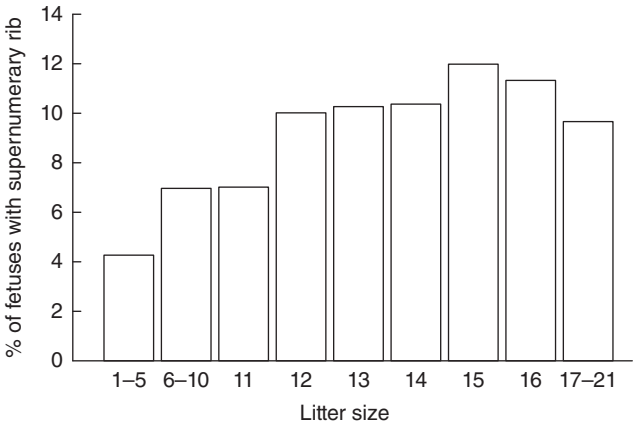


***Figure 6*** *Effect of litter size (live fetuses per litter) on incidence of supernumerary rib in 1379 control rat litters. Between 1978 and 1988, fetal skeletons from 1379 litters of control rats were stained with alizarin red and examined for supernumerary rib.*

**TABLE 3 Pregnancy Categories**

| |
|---|
| A: Adequately tested in humans, no risk (0.7% of approved drugs) |
| B/C/D: (Increasing levels of concern) (92.3% of approved drugs—66% in C) |
| X: Contraindicated for use in pregnancy (7.0% of approved drugs) |
| A: Animal studies and well-controlled studies in pregnant women failed to demonstrate a risk to the fetus. |
| B: Animal studies have failed to demonstrate risk to fetus; no adequate and well-controlled studies in pregnant women. |
| C: Animal studies showed adverse effect on fetus; no well-controlled human studies. |
| D: Positive evidence of human fetal risk based upon human data, but potential drug benefit outweighs risk. |
| X: Studies show fetal abnormalities in animals and humans; drug is contraindicated in pregnant women. |

*Note*: Use in-pregnancy Ratings; FDA, 1979.

was divided by the expected control value for that value of $m$. This last step makes the litter score immune to spontaneous or treatment-related effects on $m$. The final "robust" scores have more power for detecting effects than various other measures (raw $r/m$, affected litters/litters, $r$, $\Sigma r/\Sigma n$, and the likelihood score) and has a lower false-positive rate with fluctuations in $m$.

A covariance analysis (Snedecor and Cochran, 1980) can be used to reduce variability in a parameter and thereby increase sensitivity. For example, much of the variability in fetal weight data is due to variable litter size and, for rats, litters being sacrificed at different times during the workday. The variability due to these sources can be reduced by using litter size and time of sacrifice as potential covariates. Similarly, litter size and length of gestation can be used as covariates for neonatal pup weights and body weight at the beginning of

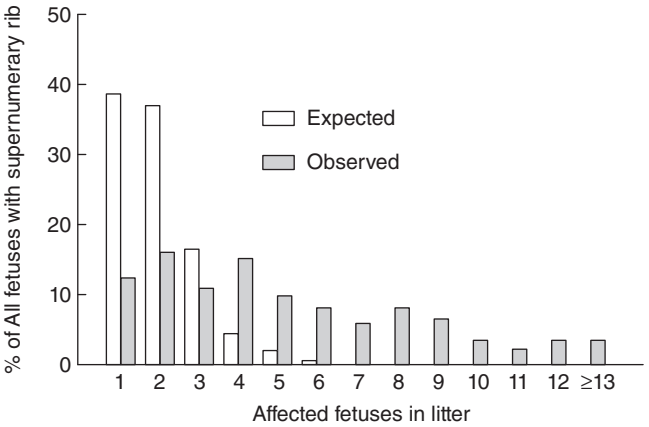


***Figure 7*** *Litter effect with supernumerary rib in 1379 control litters. Between 1978 and 1988, fetal skeletons from 1379 litters of control rats were stained with alizarin red and examined for supernumerary rib in addition to other anomalies. The calculation of the expected number of fetuses with supernumerary rib in each litter was based on the assumption that each fetus had an equally likely chance of having supernumerary rib independent of the incidence among littermates (K. Soper, personal communication, 1990).*

treatment can be used as a covariate for maternal body weight changes during the treatment period of an embryo–fetal development study.

### 4.2 Potential Hazard Categories of Developmental Toxins

It is generally agreed that an agent that causes developmental toxicity in laboratory animals at dosage levels that cause no maternal toxicity (i.e., "selective" developmental toxins) is potentially more hazardous to humans than an agent that causes developmental toxicity only at maternotoxic dosages ("nonselective" developmental toxins; e.g., see Johnson 1981; Schwetz, 1981; Fabro et al., 1982; Johnson, 1984; Johnson and Christian, 1984). This position is based on the supposition that pregnant women will avoid being exposed to toxic dosages of pharmaceuticals (which is usually but not always true). Developmental toxins can also be categorized as acting directly on the embryo or indirectly via an effect on the mother. All selective developmental toxins are presumably direct acting. Nonselective developmental toxins can either act directly or indirectly.

Direct-acting developmental toxins may be potentially more hazardous to humans than indirectly acting ones even if the direct developmental toxicity occurred only at maternotoxic dosages in the species of laboratory animals tested. When the developmental toxicity of an agent is secondary to maternal toxicity in all species tested, the dose–response curves for developmental and maternal toxicity in various species may be invariably linked and developmental toxicity would never occur at nonmaternotoxic dosages. However, when an agent acts directly on the embryo to cause developmental toxicity, the dose–

response curves may not be linked and, although they may be superimposed in the species of laboratory animals tested, they may not be superimposed in other species including humans. Thus, a direct-acting developmental toxin that is nonselective in one species may be selective in another species.

The ranking of potential developmental hazard in terms of selective, direct/nonselective, and indirect is more meaningful than the use of the terminology of specific/nonspecific and malformation/variation. However, when it cannot be determined if observed developmental toxicity is a direct or indirect effect, the alternative terminology becomes useful. A nonspecific effect (including in some cases decreased fetal weight, supernumerary rib, cleft palate in mice, and abortion in rabbits) is one that occurs commonly in response to high toxic dosages of a test agent. What makes a nonspecific effect generally less important than a specific effect is that nonspecific effects commonly occur only at maternally toxic dosages ("coeffective") and may be secondary to maternal toxicity—that is, indirect. However, when an apparently nonspecific adverse developmental effect is selective (direct), that is, it occurs at nonmaternotoxic dosages, it may nevertheless be indicative of a potential developmental hazard.

In general, an agent that induces a malformation (i.e., a teratogen) is considered to be more of a potential hazard than one that induces only a minor variation. Also, there has traditionally been more of a stigma associated with an agent that induces malformations than one that causes resorptions, even though embryo death is obviously a seriously adverse outcome. The point that makes the distinction among malformations, variations, or resorptions less important is that an agent that perturbs development to cause one effect in one species may cause a different effect in another species. Thus, any developmental toxic effect at nonmaternotoxic dosages should be considered carefully.

### 4.3 Associations between Developmental and Maternal Toxicity

The developmental toxicity of many pharmaceuticals occurs only at maternally toxic dosages (Khera, 1984, 1985; Schardein, 1987). Also, there are several compounds for which there is evidence that their developmental toxicity is secondary to their maternal toxicity. The decreased uterine blood flow associated with hydroxyurea treatment of pregnant rabbits may account for the embryotoxicity observed (Millicovsky et al., 1981). The teratogenicity of diphenylhydantoin in mice may be secondary to decreased maternal heart rate (Watkinson and Millicovsky, 1983) as supported by the amelioration of the teratogenicity by hyperoxia (Millicovsky and Johnston, 1981) and the dependence on maternal genotype in genetic crosses between sensitive and resistant strains (Johnston et al., 1979; Hansen and Hodes, 1983). The hemolytic anemia caused in pregnant rabbits by diflunisal was severe enough to explain the concomitant axial skeletal malformations (Clark et al., 1984). Acetazolamide-induced fetal malformations in mice are apparently related to maternal hypercapnia (Weaver and Scott, 1984a,b) and hypokalemia (Ellison and Maren,

1972). The increased resorption rate induced in rabbits by the antibiotic norfloxacin depends on exposure of the maternal gastrointestinal tract (Clark et al., 1991).

In addition, various treatments that simulate effects that can result from pharmaceutical treatment have been shown to cause developmental toxicity. Food deprivation can cause embryo–fetal toxicity and teratogenicity in mice (Szabo and Brent, 1975; Hemm et al., 1977) and rats (Ellington, 1980) and fetal death, decreased fetal weight, and abortions in rabbits (Matsuzawa et al., 1981; Clark et al., 1991). Treatments that result in maternal hypoxia, such as hypobaric exposure (Degenhardt and Kladetzky, 1955) and blood loss (Grote, 1969), have been shown to be teratogenic. Also, the results from testing with numerous agents suggest that supernumerary rib in mice is caused by maternal stress (Kavlock et al., 1985; Beyer and Chernoff, 1986).

Thus, in any case where developmental toxicity occurs at dosage levels with only moderate to severe maternal toxicity, the possibility of the developmental toxicity being secondary to the maternal toxicity can be considered. That is not to say, however, that it can be concluded that the developmental toxicity is secondary any time there is coincident maternal toxicity. To the contrary, it is usually very difficult to establish a causal relationship. Superficially similar types of maternal toxicity do not always cause the same pattern of developmental toxicity (Chernoff et al., 1990). This may be because the developmental toxicity is secondary to maternotoxicity, but, since typical developmental toxicity studies include only a very cursory evaluation of maternal toxicity, the developmental toxicity may be secondary to an aspect of maternotoxicity that is not even being measured.

To demonstrate that a developmental effect is secondary to a particular parameter of maternal toxicity, it is necessary but not sufficient to show that all mothers with developmental toxicity also had maternal toxicity and that the severity of the developmental effect was correlated with the maternal effect. An example of such a correlation is shown in Figure 8, in which a drug-induced effect on maternal body weight change in rabbits is correlated ($r = 0.45, p < 0.05$) with a drug-induced decrease in fetal weight. Other examples where this approach has been used to evaluate the relationship between maternal and developmental toxicity include (1) the negative correlation between resorption rate and maternal body weight change in norfloxacin-treated rabbits (Clark et al., 1991), supporting the contention that the developmental toxicity was secondary, and (2) the lack of correlation between embryotoxicity and maternal body weight change in pregnant mice treated with caffeine and L-phenylisopropyladenosine (Clark et al., 1987), suggesting no causal relationship between developmental and maternal toxicity may be required.

### 4.4 Assessment of Human Risk

Most test agents can be demonstrated to be developmentally toxic if tested under extreme conditions. This fact has been popularized as Karnofsky's law:

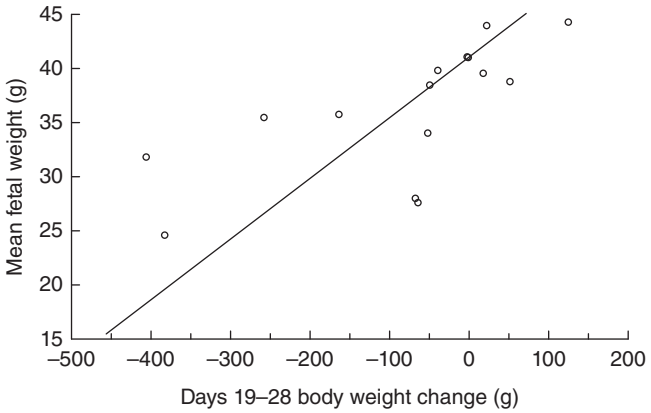


***Figure 8*** *Correlation between drug-induced effects on maternal body weight change and fetal weight in rabbits. The data were collected from the high-dosage group of a developmental toxicity study of a prospective drug candidate. The rabbits were dosed orally with the test agent from days 6–18 of gestation. On day 28 of gestation, the rabbits were cesarean sectioned and the live fetuses weighed.*

"Any drug administered at the proper dosage, at the proper stage of development, to embryos of the proper species … will be effective in causing disturbances in embryonic development" (Karnofsky, 1965, p. 185). In practice, about 37% of 3301 chemicals tested have been found to be teratogenic according to one tabulation (Schardein, 1993, p. viii; see also Shepard, 1998). Contributing to this high rate is the practice of testing maternotoxic doses (to satisfy regulatory guidelines) that in some cases result in developmental toxicity secondary to maternal toxicity. Despite the high rate of positives in animal tests, very few xenobiotics are known to cause developmental toxicity in humans as commonly used. Thus, simply the induction of developmental toxicity by a test agent in animals does not necessarily indicate that that test agent will be a developmental hazard to human conceptus under normal exposure conditions.

When a prospective drug under development for use in women of childbearing potential is determined to cause developmental toxicity in laboratory animals, the first question to be considered is whether that agent would cause developmental toxicity in humans at the anticipated therapeutic dosage level. This assessment and the related decision of whether to continue development of the drug candidate are currently based on the following:

1. The ratio between the estimated systemic exposure at the lowest effect level [or highest no-observed-effect level (NOEL)] and the estimated systemic exposure at the anticipated therapeutic dosage level (the "safety factor")
2. Whether the effect is selective, direct, and/or specific

3. The potential benefit to the patient population (compared to other available therapies)

The most common finding is that minor, nonselective, nonspecific developmental toxicity (e.g., decreased fetal weight) is observed at dosages at least 10-fold above the anticipated therapeutic dosage level. In this situation, development of the agent would normally proceed even if the "safety factor" were only 3–5. This is the case since (1) many pharmaceuticals cause maternal toxicity in laboratory animals at low multiples (e.g., 10) of the clinical exposure, (2) nonspecific developmental toxicity commonly accompanies maternal toxicity, and (3) pharmaceuticals fitting this pattern do not usually cause developmental effects as used clinically (which often includes the practice of not prescribing for women known to be pregnant).

In contrast, a drug candidate that selectively causes major malformations at a dosage threefold higher than the clinical dosage would likely not be developed to treat a non-life-threatening disease. However, it might be developed if the disease to be treated was particularly debilitating, no other effective therapy was available, and it was felt that the exposure of pregnant women could be largely avoided.

Until this year, once a new pharmaceutical is approved by the FDA, it is placed in one of five pregnancy categories (A, B, C, D, or X) based on the results of animal developmental toxicity studies and, when available (usually not), information from human usage experience (see Table 2 and Frankos, 1985). Note that the categorization does not depend on the safety factor for a developmental effect or whether the effect is major, selective, direct, or specific (although these factors may be considered when determining if a drug is to be approved). Most often, there are positive findings in animals and no experience in pregnant women and the drug is placed in pregnancy category C, indicating that it is to be used in pregnancy only if the potential benefit justifies the risk to the conceptus. Thus, it is left to the prescribing physician to regulate the exposure of pregnant women to the drug. If animals studies were negative and there is no information on effects on pregnant women, the agent is placed in pregnancy category B, indicating that the agent is to be used in pregnancy only if clearly needed. If developmental toxicity has been established in women (or, in some cases, if only strongly suspected), the agent is placed in category D or X. With category D, women may be prescribed the drug if the benefit outweighs the risk and the patient is informed of the potential hazard to the conceptus. Category X drugs are contraindicated in women who are or may become pregnant. Table 4 summarizes these categories and the proportions of drugs in each in 2000 [according to the *Physician's Desk Reference* (PDR)].

The FDA has adopted (2009) a change in these labeling requirements first proposed in May 2008. A new FDA pregnancy labeling guidance was proposed for labeling of drugs for reproductive and pregnancy risks. The proposed rule would remove the letter categories from the pregnancy section. The new format would have three sections:

**TABLE 4 Pregnancy Categories**[a]

| | Outcome of Animal Studies | | |
|---|---|---|---|
| Outcome of Human studies | + | – | Not Available |
| + | X or D | X or D | X or D |
| – | B | A | A or B |
| Not available | $C_1$ | B | $C_2$ |

[a]A, B, $C_2$: Use during pregnancy only if clearly needed.
$C_1$: Use during pregnancy only if the potential benefit justifies the potential risk to the fetus.
D: If used during pregnancy, the patient should be apprised of the potential hazard to the fetus.
X: Contraindicated in women who are or may become pregnant.

1. *Fetal Risk Summary* This section would begin with a one-sentence risk conclusion that characterizes the likelihood that the drug increases the risk of four types of developmental abnormalities: structural anomalies, fetal and infant mortality, impaired physiological function, and alterations to growth. An example of a risk conclusion based on human data is: "human data do not indicate that Drug X increases the overall risk of structural anomalies." Many of the risk conclusions in the proposed rule are standardized statements that must be used. This would be followed by a summary of significant effects.
2. *Clinical Considerations* This component would address three main topics important when counseling with and prescribing for women who are pregnant, lactating, or of childbearing age.
3. *Data* This section would have a more detailed discussion of available data. Human data would appear before animal data.

The pregnancy section would also include information about whether there is a pregnancy registry for the drug. Pregnancy exposure registries collect and maintain data on the effects of approved drugs that are prescribed to and used by pregnant women. The lactation section of prescription drug labeling would provide information on use of the drug while breastfeeding and would use the same format as the pregnancy section.

Using the highest NOEL for determining a safety factor has the following flaws:

1. The determination of a true no-effect level (should one actually exist, which is debatable in some cases) is impossible given the statistical power associated with the group sizes typically used; thus, the reported NOEL is very dependent on the selected group size.
2. The NOEL depends greatly on the selection of dosage levels; unless the selected dosage is just below the threshold for detectable effects, the reported NOEL is an underestimate; thus, tightly spaced dosage levels favor the determination of a higher NOEL.

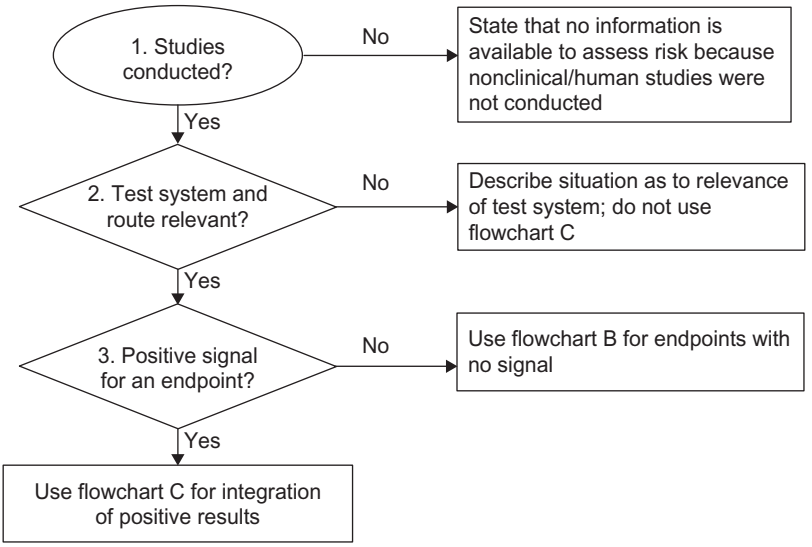


***Figure 9*** *Flowchart A. Overall decision tree for evaluation on reproduction/developmental toxicity risk from Wedge Document, 1999, distributed through www.FDA.gov (June 1999).*

Accordingly, the FDA has developed a sequential method of evaluating and dealing with reproductive and developmental analysis. This is called wedge analysis and is demonstrated in Figures 9–11.

## 5 IN VITRO TESTS FOR DEVELOPMENTAL TOXICITY

Many in vitro systems have been proposed as tests for developmental toxicity (for review, see Brown, 1984; Lin, 1987; In Vitro Teratology Task Force, 1987; and Gad, 2000). Various uses have been suggested for these in vitro tests, including the following:

1. A general prescreen to select likely developmental toxins for subsequent whole-animal studies
2. A prescreen to select among possible backups to a lead drug candidate that had been found to be developmentally toxic
3. To study mechanisms of developmental toxicity
4. To provide supplementary information about developmental toxicity in addition to that provided by whole-animal studies
5. To replace whole animals for evaluating developmental toxicity

Uses 1 and 5 above are very unlikely to be applicable to the pharmaceutical industry. One problem with in vitro systems for these purposes is that the

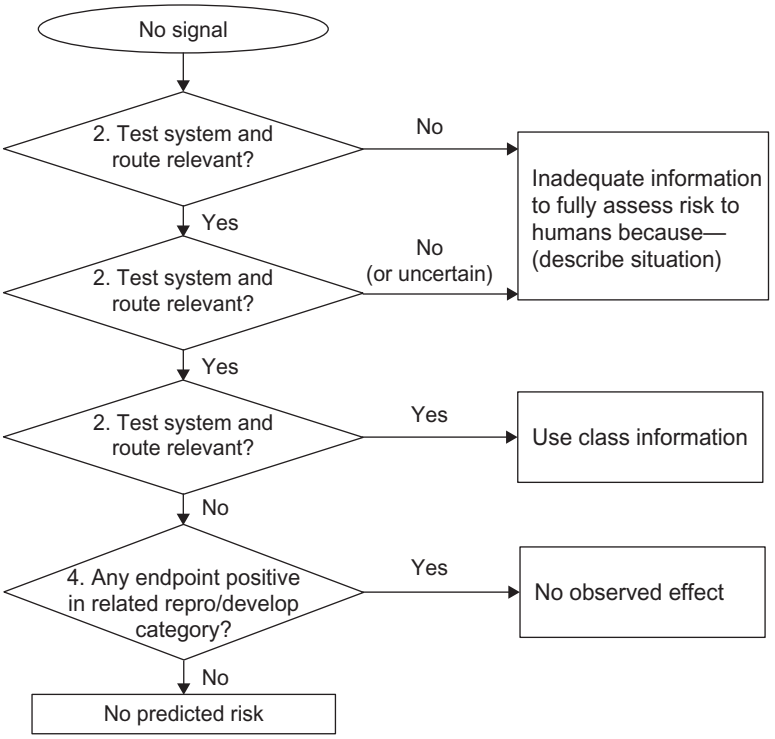


***Figure 10*** *Flowchart B. Decision tree for endpoints with no signal from Wedge Document, 1999, distributed through www.FDA.gov (June 1999).*

percentages of the agents that are positive are very high, for example, 69% of agents tested in the mouse ovarian tumor (MOT) cell attachment test and 72% of agents tested in the mouse limb bud (MLB) assay. High correlations between in vivo and in vitro results have been reported based on the limited number of validation work completed. But these correlations have compared an in vitro end point to teratogenicity in laboratory animals without regard to maternotoxicity. Thus, the question that these screens seem to be answering is: Can this agent be teratogenic or developmentally toxic in laboratory animals at any dosage level? However, as discussed above, it is not important for the purpose of safety assessment if an agent can be developmentally toxic in laboratory animals at high, maternotoxic dosages. The important question for prospective screens to answer is this: Is the agent a selective or direct developmental toxin? For these reasons, a promising drug candidate would not be dropped from development due to a positive result in a current in vitro test and a negative result would not preclude the need for whole-animal studies.

To relate a positive finding in an in vitro test to the in vivo situation, one must either compare the concentration that caused the positive developmental effect in vitro to the exposure level of the embryo in vivo or compare the in

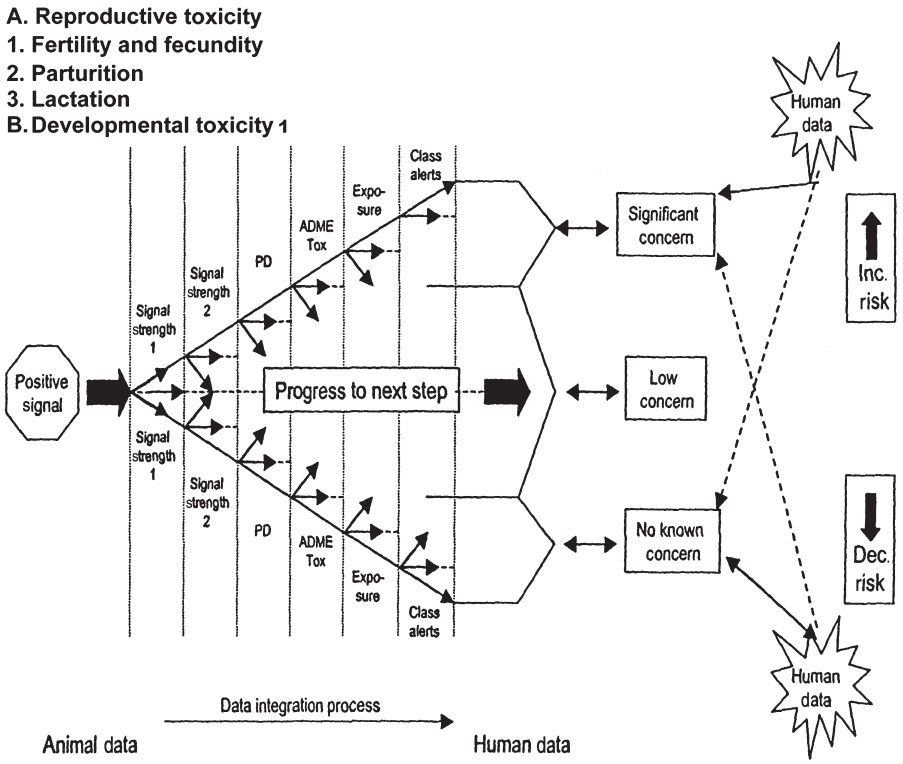


***Figure 11*** *Flowchart C. Integration of positive reproduction/ancillary study results from Wedge Document, 1999, distributed through www.FDA.gov (June 1999).*

vitro concentration for a developmental effect to the maternotoxicity that would be associated with exposure at that concentration in vivo. To do the necessary pharmacokinetic studies in vivo would defeat the purpose of using an in vitro test. It would be very desirable and may be possible, though, to have an endpoint in an in vitro test that would correlate with maternal toxicity.

Currently, only the *Hydra* system incorporates a measurement of "toxicity" to the adult to provide a comparison of the sensitivity of the "embryo" with that of the adult (Johnson et al., 1988). However, the *Hydra* screen has not been fully validated as being predictive of results in mammals and has fallen from favor. Thus, a major goal of research directed toward developing an in vitro teratogen screen should be to find a simple yet appropriate measure of toxicity unrelated to development. This would allow the comparison of the dose for a 50% effect ($ED_{50}$) of "developmental toxicity" as measured in vitro to an $ED_{50}$ for "adult" toxicity in vitro. The validation of such a dual in vitro system would involve comparing the developmental selectivity in vitro to that in vivo for a large number of compounds. In a preliminary effort in this regard,

effects on cell division in the rat limb bud micromass assay were considered as a possible correlate of maternal toxicity (Wise et al., 1990a).

Another possible use of in vitro developmental toxicity tests would be to select the least developmentally toxic backup from among a group of structurally related compounds with similar pharmacological activity (use 2 above), for example, when a lead compound causes malformations in vivo and is also positive in a screen that is related to the type of malformation induced. However, even for this limited role for a developmental toxicity screen, it would probably also be desirable to have a measure of the comparative maternotoxicity of the various agents and/or information on the pharmacokinetics and distribution of the agents in vivo.

In vitro developmental toxicity systems have clearly been useful for studies of mechanisms of developmental effects (e.g., Daston et al., 1989)—use 3 above. It is unclear, though, whether in vitro developmental toxicity tests will provide useful information about developmental toxicity that is not derived from whole-animal studies (use 4). As is true for a possible use as a prescreen, the interpretation of a positive finding in an in vitro test will depend on knowing the exposure level in vivo. When this is known, the in vitro information could be helpful. The results of in vivo studies, though, would still likely be considered definitive for that species.

## 6 APPRAISAL OF CURRENT APPROACHES FOR DETERMINING DEVELOPMENTAL AND REPRODUCTIVE HAZARDS

The current system for testing new pharmaceuticals for developmental and reproductive toxicity has been largely intact since 1966. In that time, no thalidomide-like disasters have occurred. It cannot be proven, but there is a good chance that these two statements are linked, that is, that the testing system has prevented potent, selective, human teratogens from being marketed. Indeed, the development of many compounds has been terminated because of positive findings in standard developmental toxicity studies. We do not know for certain if any of these agents would have been developmental hazards in humans, but it seems very likely. Due to the limited information on developmental toxicity of chemical agents in humans and the obvious inability to conduct controlled human studies, the correlation between animal studies and human findings is uncertain and it is difficult to extrapolate precisely from animals to humans (see Frankos, 1985). However, the worst hazards—the few dozen selective developmental toxins that are known to be teratogens in humans—are generally also selective teratogens in animals. Thus, although the current battery of animal studies is not perfect, it appears to have been adequate and effective in performing the important task of preventing the widespread exposure of pregnant women to dangerous developmental toxins. In the few cases where new pharmaceuticals have been shown to cause malformations in humans, animal studies had been positive and provided an early warning to the potential problem.

## REFERENCES


Adams, M. J., Hicks, P. G., and York, M. J. (1989). Detection of pregnancy in the rabbit by assay of plasma progesterone. *Teratology* 40:275.

Adams, M. J., Hicks, P. G., and York, M. J. (1990). Detection of pregnancy in the rabbit by assay of progesterone: An update. *Teratology* 42:25A–26A.

Anderson, C., and Clark, R. L. (1990). External genitalia of the rat: Normal development and the histogenesis of 5α-reductase inhibitor-induced abnormalities. *Teratology* 42:483–496.

Beyer, P. E., and Chernoff, N. (1986). The induction of supernumerary ribs in rodents: The role of maternal stress. *Teratogen. Carcinogen. Mutagen.* 6:149–429.

Bloch, E., Lew, M., and Klein, M. (1971). Studies on the inhibition of fetal androgen formation. Inhibition of testosterone synthesis in rat and rabbit fetal testes with observations on reproductive tract development. *Endocrinology* 89:16–31.

Brown, N. A., and Freeman, S. J. (1984). Alternative tests for teratogenicity. *ATLA* 12:7–23.

Buelke-Sam, J., Kimmel, C. A., and Adams, J. (1985). Design considerations in screening for behavioral teratogens: Results of the collaborative behavioral teratology study. *Neurobehav. Toxicol. Teratol.* 7:537–789.

Chapin, R. E., Gulati, D. R., and Barnes, L. H. (1991). The effects of dietary restriction on reproductive endpoints in Sprague-Dawley rats. *Toxicologist* 11:112.

Chernoff, N., Setzer, R. W., Miller, D. B., Rosen, M. B., and Rogers, J. M. (1990). Effects of chemically induced maternal toxicity on prenatal development in the rat. *Teratology* 42:651–658.

Chow, B. F., and Rider, A. A. (1973). Implications of the effects of maternal diets in the various species. *J. Anim. Sci.* 36:167–173.

Christian, M. S. (1983). Assessment of reproductive toxicity: State of the art. In M. S. Christian, M. Galbraith, P. Voytek, and M. A. Mehlman (Eds.), *Assessment of Reproductive and Teratogenic Hazards*. Princeton Scientific, Princeton, NJ, pp. 65–76.

Christian, M. S., and Hoberman, A. M. (1989). Current *in vivo* reproductive toxicity and developmental toxicity (teratology) test methods. In J. A. Marquis and A. W. Hayes (Eds.), *A Guide to General Toxicology*, 2nd ed. S. Karger, Basel, Switzerland, pp. 91–100.

Clark, R. L., Anderson, C. A., Prahalada, S., Leonard, Y. M., Stevens, J. L., and Hoberman, A. M. (1990a). 5α-Reductase inhibitor-induced congenital abnormalities in male rat external genitalia. *Teratology* 41:544.

Clark, R. L., Antonello, J. M., Grossman, J. T., Wise, L. D., anderson, C., Bagdon, W. J., Prahalada, S., MacDonald, J. S., and Robertson, R. T. (1990b). External genitalia abnormalities in male rats exposed *in utero* to finasteride, a 5α-reductase inhibitor. *Teratology* 42:91–100.

Clark, R. L., Antonello, J. M., Soper, K. A., Bradstreet, T. C., Heyse, J. F., and Ciminera, J. L. (1989). Statistical analysis of developmental toxicity data. *Teratology* 39:445–446.

Clark, R. L., Antonello, J. M., Wenger, J. D., Deyerle-Brooks, K., and Duchai, D. M. (1991). Selection of food allotment for New Zealand white rabbits in developmental toxicity studies. *Fund. Appl. Toxicol.* 17:584–592.

Clark, R. L., Robertson, R. T., Minsker, D. H., Cohen, S. M., Tocco, D. J., Allen, H. L., James, M. L., and Bokelman, D. L. (1984). Association between adverse maternal and embryo-fetal effects in norfloxacin-treated and food-deprived rabbits. *Fund. Appl. Toxicol.* 7:272–286.

Clark, R. L., Eschbach, K., Cusick, W. A., and Heyse, J. F. (1987). Interactions between caffeine and adenosine agonists in producing embryo resorptions and malformations in mice. *Toxicol. Appl. Pharmacol.* 91:371–385.

D'Aguanno, W. (1973). Guidelines of reproduction studies for safety evaluation of drugs for human use. In *FDA Introduction to Total Drug Quality*. DHEW Publication No. (FDA) 74–3006. Department of Health, Education and Welfare/Public Health and Safety/Food and Drug Administration, Washington, DC.

Daston, G. P., Yonker, J. E., Powers, J. F., and Heitmeyer, S. A. (1989). Difference in teratogenic potency of ethylenethiourea in rats and mice: Relative contribution of embryonic and maternal factors. *Teratology* 40:555–556.

Degenhardt, K., and Kladetzky, J. (1955). Spinal deformation and chordal attachment. *Z. Menschl. Vererb. Konstitutionsl.* 33:151–192.

Ellington, S. (1980). *In vivo* and *in vitro* studies on the effects of maternal fasting during embryonic organogenesis in the rat. *J. Reprod. Fertil.* 60:383–388.

Ellison, A. C., and Maren, T. H. (1972). The effect of potassium metabolism on acetazolamide-induced teratogenesis. *Johns Hopkins Med. J.* 130:105–115.

Fabro, S., Shull, G., and Brown, N. A. (1982). The relative teratogenic index and teratogenic potency: Proposed components of developmental toxicity risk assessment. *Teratogen. Carcinogen. Mutagen.* 2:61–76.

Frankos, V. (1985). FDA perspectives on the use of teratology data for human risk assessment. *Fund. Appl. Toxicol.* 5:615–625.

Gad, S. C. (2000). *In Vitro Toxicology*, 2nd ed. Taylor & Francis, Philadelphia, PA.

Grote, W. (1969). Trunk skeletal malformations following blood loss in gravid rabbits. *Z. Anat. Entwicklungsgesch.* 128:66–74.

Gulati, K., Hope, E., and Chapin, R. E. (1991). The effects of dietary restriction on reproductive endpoints in Swiss mice. *Toxicologist* 11:112.

Habert, R., and Picon, R. (1984). Testosterone, dihydrotestosterone and estradiol-17β levels in maternal and fetal plasma and in fetal testes in the rat. *J. Steroid Biochem.* 21:183–198.

Hansen, D. K., and Hodes, M. E. (1983). Comparative teratogenicity of phenytoin among several inbred strains of mice. *Teratology* 28:175–179.

Heinrichs, W. L. (1985). Current laboratory approaches for assessing female reproductive toxicity. In R. L. Dixon (Ed.), *Reproductive Toxicology*. Raven, New York, pp. 95–108.

Hemm, R., Arslanoglou, L., and Pollock, J. (1977). Cleft palate following prenatal food restriction in mice: Association with elevated maternal corticosteroids. *Teratology* 15:243–248.

Heywood, R., and James, R. W. (1985). Current laboratory approaches for assessing male reproductive toxicity: Testicular toxicity in laboratory animals. In R. L. Dixon (Ed.), *Reproductive Toxicology*. Raven, New York, pp. 147–160.

Hood, R. D. (2006). *Developmental and Reproductive Toxicology: A Practical Approach*, 2nd ed. Taylor & Francis, Philadelphia.

International Conference on Harmonisation (ICH) (2000a). S5A Detection of reproduction for medicinal products.

International Conference on Harmonisation (ICH) (2000b). S5B Detection of toxicity to reproduction for medicinal products: Addendum on toxicity to male fertility.

International Conference on Harmonisation (ICH) (1994). Detection of toxicity to reproduction for medicinal products. *Fed. Reg.* 59(183):48746–48752.

*In Vitro* Teratology Task Force (1987). Report. *Environ. Health Perspect.* F72:201–249.

Johnson, E. M. (1981). Screening for teratogenic hazards: Nature of the problems. *Ann. Rev. Pharmacol. Toxicol.* 21:417–429.

Johnson, E. M. (1984). A prioritization and biological decision tree for developmental toxicity safety evaluations. *J. Am. Coll. Toxicol.* 3:141–147.

Johnson, E. M., and Christian, M. S. (1984). When is a teratology study not an evaluation of teratogenicity? *J. Am. Coll. Toxicol.* 3:431–434.

Johnson, E. M., Newman, L. M., Gabel, B. E. G., Boerner, T. F., and Dansky, L. A. (1988). An analysis of the *Hydra* assay's applicability and reliability as a developmental toxicity prescreen. *J. Am. Coll. Toxicol.* 7:111–126.

Johnston, M. C., Sulik, K. K., and Dudley, K. H. (1979). Genetic and metabolic studies of the differential sensitivity of A/J and C57BL/6J mice to phenytoin ("Dilantin")-induced cleft lip. *Teratology* 19:33A.

Karnofsky, D. (1965). Mechanisms of action of certain growth-inhibiting drugs. In J. Wilson and J. Warkany (Eds), *Teratology: Principles and Techniques*. University of Chicago Press, Chicago, pp. 185–213.

Kavlock, R. J., Chernoff, N., and Rogers, E. H. (1985). The effect of acute maternal toxicity on fetal development in the mouse. *Teratogen. Carcinogen. Mutagen.* 5:3–13.

Khera, K. S. (1984). Maternal toxicity—A possible factor in fetal malformations in mice. *Teratology* 29:411–416.

Khera, K. S. (1985). Maternal toxicity: A possible etiological factor in embryo-fetal deaths and fetal malformations of rodent-rabbit species. *Teratology* 31:129–153.

Khera, K. S. (1991). Chemically induced alterations in maternal homeostasis and histology of conceptus: their etiologic significance in rat fetal anomalies. *Teratology* 44:259–297.

Korenbrot, C. C., Huhtaniemi, I. T., and Weiner, R. I. (1977). Preputial separation as an external sign of pubertal development in the male rat. *Biol. Reprod.* 17:298–303.

Lin, G. H. Y. (1987). Prediction of teratogenic potential and a proposed scheme for teratogenicity screening of industrial research and development materials. *In Vitro Toxicol.* 1:203–217.

Lochry, E. A. (1987). Concurrent use of behavioral/functional testing in existing reproductive and developmental toxicity screens: Practical consideration. *J. Am. Coll. Toxicol.* 6:433–439.

Lochry, E. A., Hoberman, A. M., and Christian, M. S. (1984). Positive correlation of pup body weight with other commonly used developmental landmarks. *Teratology* 29:44A.

Manson, J. M., and Kang, Y. J. (1989). Test methods for assessing female reproductive and developmental toxicology. In A. W. Hayes (Ed.), *Principles and Methods of Toxicology*, 2nd ed. Raven, New York, pp. 311–359.

Matsuzawa, T., Nakata, M., Goto, I., and Tsushima, M. (1981). Dietary deprivation induces fetal loss and abortions in rabbits. *Toxicology* 22:255–259.

Matt, D. M., Sarver, P. L., and Lu, J. K. H. (1987). Relation of parity and estrous cyclicity to the biology of pregnancy in aging female rats. *Biol. Reprod.* 37:421–430.

Matt, D. W., Lee, J., Sarver, P. L., Judd, H. L., and Lu, J. K. H. (1986). Chronological changes in fertility, fecundity, and steroid hormone secretion during consecutive pregnancies in aging rats. *Biol. Reprod.* 34:478–487.

Millicovsky, G., DeSesso, J. M., Kleinman, L. I., and Clark, K. E. (1981). Effects of hydroxyurea on hemodynamics of pregnant rabbits: A maternally mediated mechanism of embryotoxicity. *Am. J. Obstet. Gynecol.* 140:747–752.

Millicovsky, G., and Johnston, M. C. (1981). Maternal hyperoxia greatly reduces the incidence of phenytoin-induced cleft lip and palate in A/J mice. *Science* 212:671–672.

Minsker, D., Bagdon, W., MacDonald, J., Robertson, R., and Bokelman, D. (1990). Maternotoxicity and fetotoxicity of an angiotensin-converting enzyme inhibitor, enalapril, in rabbits. *Fund. Appl. Toxicol.* 14:461–479.

Mulay, S., Varma, D. R., and Solomon, S. (1982). Influence of protein deficiency in rats on hormonal status and cytoplasmic glucocorticoid receptors in maternal and fetal tissues. *J. Endocrinol.* 95:49–58.

Neumann, F., von Berswordt-Wallrabe, R., Elger, W., Steinbeck, H., Hahn, J. D., and Kramer, M. (1970). Aspects of androgen-dependent events as studied by antiandrogens. *Recent Prog. Horm. Res.* 26:337–410.

Nishimura, M., Iizuka, M., Iwaki, S., and Kast, A. (1982). Repairability of drug-induced "wavy ribs" in rat offspring. *Arzneim.-Forsch./Drug Res.* 32:1518–1522.

Palmer, A. K. (1981). Regulatory requirements for reproducing toxicology: Theory and practice. In C. A. Kimmel and J. Buelke-Sam (Eds.), *Developmental Toxicology*. Raven, New York, pp. 259–287.

Persaud, T. V. N. (1985). Teratogenicity testing. In T. V. N. Persaud, A. E. Chudley, and R. G. Skalko (Eds.), *Basic Concepts in Teratology*. Alan R. Liss, New York, pp. 155–181.

Saegusa, T., Kaneko, Y., Sato, T., Nrama, I., and Segima, Y. (1980). BD40A-induced wavy ribs in rats. *Teratology* 22:14A.

Salewski, V. E. (1964). Färbemethode zum makroskopischen Nachweis von Implantationsstellen am Uterus der Ratte. Nauyn-Schmiedebergs. *Arch. Exp. Path. U. Pharmak.* 247:367.

Schardein, J. (1987). Approaches to defining the relationship of maternal and developmental toxicity. *Teratogen. Carcinogen. Mutagen.* 7:255–271.

Schardein, J. (1988). Teratologic testing: Status and issues after two decades of evolution. *Rev. Environ. Contam. Toxicol.* 102:1–78.

Schardein, J. (1993). *Chemically Induced Birth Defects*, 2nd ed. Marcel Dekker, New York.

Schwetz, B. A. (1981). Monitoring problems in teratology. In E. J. Gralla (Ed.), *Scientific Considerations in Monitoring and Evaluating Toxicological Research*. Hemisphere, Washington, DC, pp. 179–192.

Shepard, T. (1998). *Catalog of Teratogenic Agents*, 9th ed. Johns Hopkins University Press, Baltimore, MD.

Snedecor, G. W., and Cochran, W. G. (1980). *Statistical Methods*, 7th ed., Chapter 10. Iowa State University Press, Ames.

Soper, K. A., and Clark, R. L. (1990). Exact permutation trend tests for fetal survival data. *Proc. Biopharm. Sect. Am. Stat. Assoc.*, pp. 263–268.

Staples, R. E. (1971). Blastocyst transplantation in the rabbit. In J. C. Daniel, Jr. (Eds.), *Methods in Mammalian Embryology*. W. H. Freeman, San Francisco, pp. 290–304.

Szabo, K., and Brent, R. (1975). Reduction of drug-induced cleft palate in mice. *Lancet* June:1296–1297.

Tanimura, T. (1990). The Japanese perspectives on the reproductive and developmental toxicity evaluation of pharmaceuticals. *J. Am. Coll. Toxicol.* 9:27–38.

Turner, M. R. (1973). Perinatal mortality, growth, and survival to weaning in offspring of rats reared on diets moderately deficient in protein. *Br. J. Nutr.* 29:139–147.

Tyl, R. W. (1988). Developmental toxicity in toxicological research and testing. In B. Ballantyne (Ed.), *Perspectives in Basic and Applied Toxicology*. Butterworth, Cambridge, UK, pp. 206–241.

U.S. Environmental Protection Agency (EPA) (1991). *Pesticide Assessment Guidelines—Subdivision F, Hazard Evaluation: Human and Domestic Animals*. Addendum 10—Neurotoxicity Series 81, 82, and 83. Publication No. PB-91-154617. National Technical Information Service, Washington, DC.

U.S. Food and Drug Administration (FDA) (1966). *Guidelines for Reproduction Studies for Safety Evaluation of Drugs for Human Use*. Drug Review Branch, Division of Toxicological Evaluation, Bureau of Science, FDA, Washington, DC.

U.S. Food and Drug Administration (FDA) (1978). Good laboratory practice regulations. *Federal Register*, December 22, 1978. Title 21, Part II, Part 58.

U.S. Food and Drug Administration (FDA) (1979). Pregnancy Category Classification of Fetal Risks. *FDA Drug Bulletin* 9(4):23–24.

U.S. Food and Drug Administration (FDA) (1984). *Final Report of Task Force on Reproductive and Developmental Toxicity. Review of Current Recommendations and Requirements for Reproductive and Developmental Toxicity Studies*, Department of Health and Human Services, Washington, DC.

U.S. Food and Drug Administration (FDA) (1987). Good laboratory practice regulations: Final rule. *Federal Register*, September 4, 1987. Part IV, Vol. 52, No. 172.

U.S. Food and Drug Administration (FDA) (1993). Guideline for the study and evaluation of gender differences in the clinical evaluation of drug. *Federal Register*, July 22, 1993. Vol. 58, No. 139, pp. 39405–39416.

U.S. Food and Drug Administration (FDA), Bureau of Foods (1982). *Toxicological Principles for Safety Assessment of Direct Food Additives and Color Additives and Color Used in Food*. No. PB83-170696, pp. 80–117. National Technical Information Service, Washington, DC. [Also cited as (Red Book). *Guidelines for a Three-Generation Reproduction Toxicity Study with Optional Teratology Phase. Guidelines for Reproduction Testing with a Teratology Phase.*]

Watkinson, W. P., and Millicovsky, G. (1983). Effect of phenytoin on maternal heart rate in A/J mice: Possible role in teratogenesis. *Teratology* 28:1–8.

Weaver, T. E., and Scott, W. J. (1984a). Acetazolamide teratogenesis: Association of maternal respiratory acidosis and ectrodactyly in C57Bl/6J mice. *Teratology* 30:187–193.

Weaver, T. E., and Scott, W. J. (1984b). Acetazolamide teratogenesis: Interaction of maternal metabolic and respiratory acidosis in the induction of ectrodactyly in C57Bl/6J mice. *Teratology* 30:195–202.

Weindruch, R., and Walford, R. L. (1988). *The Retardation of Aging and Disease by Dietary Restriction*. Charles C. Thomas, Springfield, IL.

Wickramaratne, G. A. de S. (1988). The post-natal fate of supernumerary ribs in rat teratogenicity studies. *J. Appl. Toxicol.* 8:91–94.

Wilson, J. (1965). Methods for administering agents and detecting malformations in experimental animals. In J. G. Wilson and J. Warkany (Eds.), *Teratology, Principles and Techniques*. The University of Chicago Press, Chicago, IL, pp. 262–277.

Wise, L. D., Clark, R. L., Minsker, D. H., and Robertson, R. T. (1988). Use of hematology and serum biochemistry data in developmental toxicity studies. *Teratology* 37:502–503.

Wise, L. D., Clark, R. L., Rundell, J. O., and Robertson, R. T. (1990a). Examination of a rodent limb bud micromass assay as a prescreen for developmental toxicity. *Teratology* 41:341–351.

Wise, L. D., Vetter, C. M., anderson, C., Antonello, J. M., and Clark, R. L. (1990b). Reversible effects on external genitalia in rats exposed *in vitro* to triamcinolone acetonide. *Teratology* 41:600.

Working, P. K. (1988). Male reproductive toxicology: Comparison of the human to animal models. *Environ. Health Perspect.* 77:37–44.

Yamada, T., Ohsawa, K., and Ohno, H. (1988). The usefulness of alkaline solutions for clearing the uterus and staining implantation sites in rats. *Exp. Anim. (Tokyo)* 37:325–332.

Zeman, F. J. (1967). Effect on the young rat of maternal protein restriction. *J. Nutr.* 93:167–173.

# *Carcinogenicity Studies*

## 1 INTRODUCTION

In the experimental evaluation of substances for carcinogenesis based on experimental results of studies in a nonhuman species at some relatively high dose or exposure level, an attempt is made to predict the occurrence and level of tumorogenesis in humans at much lower levels. In this chapter we will examine the assumptions involved in this undertaking and review the aspects of design and interpretation of traditional long-term (lifetime) animal carcinogenicity studies as well as some alternative short-term models. It should be noted that these are required of the majority of marketed drugs but are only performed on the minority of drugs which reach a stage of development where a marketing application is likely and to occur within three or so years.

At least in a general way, we now understand what appear to be most of the mechanisms underlying chemical- and radiation-induced carcinogenesis. The most recent regulatory summary on identified carcinogens [National Institutes of Health (NIH), 2000] lists 44 agents classified as known to be human carcinogens. Several hundred other compounds are also described as having lesser degrees of proof. A review of these mechanisms is not germane to this chapter [readers are referred to Miller and Miller (1981) for a good short review], but it is now clear that cancer as seen in humans is the result of a multifocal set of causes.

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

**Mechanisms and Theories of Chemical Carcinogenesis**

1. Genetic (all due to some genotoxic event)
2. Epigenetic (no mutagenic event)
3. Oncogene activation
4. Two-step (induction/promotion)
5. Multistep (combination of above)

Looked at another way, the four major carcinogenic mechanisms are DNA damage, cell toxicity, cell proliferation, and oncogene activation. Any effective program to identify those drugs which have the potential to cause or increase the incidence of neoplasia in humans must effectively screen for these mechanisms (Kitchin, 1999; McGregor, 2000; Powell and Berry, 2000; Williams and Iatropoulos, 2001).

The single most important statistical consideration in the design of bioassays in the past was based on the point of view that what was being observed and evaluated was a simple quantal response (cancer occurred or it did not) and that a sufficient number of animals needed to be used to have reasonable expectations of detecting such an effect. Though the single fact of whether or not the simple incidence of neoplastic tumors is increased due to an agent of concern is of interest, a much more complex model must now be considered. The time to tumor, patterns of tumor incidence, effects on survival rate, and age of first tumor all must now be captured in a bioassay and included in an evaluation of the relevant risk to humans.

The task of interpreting the results of any of the animal-based bioassays must be considered from three different potential perspectives as to organ responsiveness:

I. Those organs with high animal and low human neoplasia rates
II. Those organs with high neoplasia rates in both animals and humans
III. Those organs with low animal but high human neoplasia rates

Note that not considered is the potential other case—where the neoplasia rates are low for both animals and humans. This is a very rare case and one for which our current bioassay designs probably lack sufficient power to be effective.

In group I, the use of animal cancer data obtained in the liver, kidney, forestomach, and thyroid gland are perceived by some as being hyperresponsive, too sensitive, and of limited value and utility in the animal cancer data obtained in group I organs. The liver is such a responsive and important organ in the interpretation of cancinogenesis data that the discussion of this subject area has been broken up into three chapters for human, rat, and mouse data. Peroxisome proliferation in the liner, particularly in mice, is an area of interpretive battle as in many cases the metabolism and mechanisms involved are not relevant to humans.

Group II organs (mammary gland, hematopoietic, urinary bladder, oral cavity, and skin) are less of an interpretive battleground than group I organs. For group II organs, all four major mechanisms of carcinogenesis (electrophile generation, oxidation of DNA, receptor–protein interactions, and cell proliferation) are known to be important. The high cancer rates for group B organs in both experimental animals and humans may at first give us a false sense of security about how well the experimental animal models are working. As we are better able to understand the probable carcinogenic mechanism(s) in the same organ in the three species, we may find that the important differences between the three species are more numerous than we suspect. This is particularly true for receptor-based and for cell-proliferation-based carcinogenic mechanisms.

Animal cancer data of group III organs are the opposite of group I organs. Group III organs have low animal cancer rates and high human cancer rates. In contrast to the continuing clamor and interpretive battleground associated with group A organs, there is little debate over group III organs. Few voices have questioned the adequacy of the present-day animal bioassay to protect the public health from possible cancer risks in these group III organs. Improved efforts must be made toward the development of cancer-predictive systems or short-term tests for cancer of the prostate gland, pancreas, colon/rectum, and cervix/uterus.

Carcinogenicity bioassays are the longest and most expensive of the extensive battery of toxicology studies required for the registration of pharmaceutical products in the United States and in other major countries. In addition, they are often the most controversial with respect to interpretation of their results. These studies are important because, as noted by the International Agency for Research on Cancer (IARC, 1987), "in the absence of adequate data on humans, it is biologically plausible and prudent to regard agents for which there is sufficient evidence of carcinogenicity in experimental animals as if they presented a carcinogenic risk to humans."

In this chapter, we consider the major factors involved in the design, conduct, analysis, and interpretation of carcinogenicity studies as they are performed in the pharmaceutical industry.

## 2 REGULATORY REQUIREMENTS AND TIMING

The prior U.S. Food and Drug Administration (FDA) guidance on the need for carcinogenicity testing of pharmaceuticals presented a dual criteria: that such studies were required to support registration of a drug that was to be administered for a period of three months or more (in Japan and Europe this was stated to be six months or more) and such testing had to be completed before filing for registration in such cases. International Conference on Harmonisation (ICH, 1996) guidelines now fix this triggering human exposure period at six months, excluding agents given infrequently through a lifetime

or for shorter periods of exposure unless there is reason for concern (such as positive findings in genotoxicity studies, structure–activity relationships suggesting such risk, evidence of preneoplastic lesions in repeat-dose studies, or previous demonstration of carcinogenic potential in the product class that is considered relevant to humans). Such studies are still only required to be completed before filing for registration. Most developers conduct carcinogenicity studies in parallel with phase III clinical studies.

Endogenous peptides, protein substances, and their analogues are generally not required to be evaluated for carcinogenicity. There are three conditions which call the need into question however:

- Products where there are significant differences in biological effects to the natural counterparts
- Products where modifications lead to significant changes in structure compared to the natural substance
- Products resulting in humans having a significant increase over the existing local or systemic concentration.

The ICH has also given guidance on design, dose selection, statistical analysis, and interpretation for such studies (1996, 1997a,b). The FDA has also offered guidance, the most recent form (FDA, 2001) in a 44-page document available online.

There has been extensive debate and consideration on the relevance and value of the traditional long-term rodent bioassays. The FDA looked at rat and mouse studies for 282 human pharmaceuticals, resulting in the conclusion that "sufficient evidence is now available for some *alternative in vivo carcinogenicity models to support their application as complimentary studies in combination with a single 2-year carcinogenicity study*" (italics added) to identify transspecies tumorigens (Contrera et al., 1997, pp. 130–131).

The Europeans, meanwhile, have focused on the need for better care in study design, conduct, and interpretation (Spindler et al., 2000), aiming to incorporate these in the revision of the CPMP (Center for Proprietary Medicinal Products) carcinogenicity guidelines (Weaver and Brunden, 1998).

## 3 SPECIES AND STRAIN

Two rodent species are routinely used for carcinogenicity testing in the pharmaceutical industry—the mouse and the rat. Sprague Dawley–derived rats are most commonly used in American pharmaceutical toxicology laboratories. However, the Wistar and Fischer 344 strains are favored by some companies, while the Long Evans and CFE (Carworth) strains are rarely used [Pharmaceutical Manufacturers Association (PMA), 1988].

With respect to mice, the CD-1 is by far the most commonly used strain in the pharmaceutical industry. Other strains used less frequently are the B6C3F1,

CF-1, NMRI, C57B1, Balb/c, and Swiss (PMA, 1988; Rao et al., 1988). Swiss is the generic term since most currently used inbred and outbred strains were originally derived from the "Swiss" mouse.

If either the mouse or the rat is considered to be an inappropriate species for a carcinogenicity study, the hamster is usually chosen as the second species.

The use of two species in carcinogenicity studies is based on the traditional wisdom that no single species can be considered an adequate predictor of carcinogenic effects in humans. Absence of carcinogenic activity in two different species is thought to provide a greater level of confidence that a compound is "safe" for humans than data derived from a single species.

One may question this reasoning on the basis that data from two "poor predictors" may not be significantly better than data from a single species. It is also reasonable to expect that the ability of one rodent species to predict a carcinogenic effect in a second rodent species should be at least equal to, if not better than, its ability to predict carcinogenicity in humans. The concordance between mouse and rat carcinogenicity data has been investigated and a summary of the results is presented in the next paragraph.

A review of data from 250 chemicals found 82% concordance between results of carcinogenicity testing in the mouse and the rat (Purchase, 1980). Haseman et al. (1984a) reported a concordance of 73% for 60 compounds studied in both species. However, 30–40% of 186 National Cancer Institute (NCI) chemicals were found to be positive in one species and negative in the other (Gold et al., 1984). It is reasonable to conclude that neither rodent species will always predict the results in the other rodent species or in humans and that the use of two species will continue until we have a much better understanding of the mechanisms of carcinogenesis.

The choice of species and strain to be used in a carcinogenicity study is based on various criteria, including susceptibility to tumor induction, incidence of spontaneous tumors, survival, existence of an adequate historical database, and availability.

Susceptibility to tumor induction is an important criterion. There would be little justification for doing carcinogenicity studies in an animal model that did not respond when treated with a "true" carcinogen. Ideally, the perfect species/strain would have the same susceptibility to tumor induction as the human. Unfortunately, this information is usually unavailable, and the tendency has been to choose animal models that are highly sensitive to tumor induction to minimize the probability of false negatives.

The incidence of spontaneous tumors is also an important issue. Rodent species and strains differ greatly in the incidence of various types of spontaneous tumors. The Sprague Dawley stock, although preferred by most pharmaceutical companies, has a very high incidence of mammary tumors in aging females, which results in substantial morbidity during the second year of a carcinogenicity study. If one chooses the Fischer 344 (F344) strain, the female mammary tumor incidence will be lower, but the incidence of testicular tumors will be higher (close to 100%) than that in Sprague Dawley rats.

A high spontaneous tumor incidence can compromise the results of a carcinogenicity study in two ways. If a compound induces tumors at a site that already has a high spontaneous tumor incidence, it may be impossible to detect an increase above the high background "noise." Conversely, if a significant increase above control levels is demonstrated, one may question the relevance of this finding to humans on the basis that the species is "highly susceptible" to tumors of this type.

The ability of a species/strain to survive for an adequate period is essential for a valid assessment of carcinogenicity. Poor survival has caused regulatory problems for pharmaceutical companies and is, therefore, an important issue (PMA, 1988). The underlying concept is that animals should be exposed to the drug for the greater part of their normal life span to make a valid assessment of carcinogenicity. If animals in a study die from causes other than drug-induced tumors, they may not have been at risk long enough for tumors to have developed. The sensitivity of the bioassay would be reduced and the probability of a false-negative result would be increased.

The availability of an adequate historical database is often cited as an important criterion for species/strain selection. Historical control data can sometimes be useful in evaluating the results of a study. Although such data are not considered equal in value to concurrent control data, they can be helpful if there is reason to believe that the concurrent control data are "atypical" for the species/strain.

Although outbred stocks (e.g., Sprague Dawley rats and CD-1 mice) are generally favored in the pharmaceutical industry, inbred strains are also used (e.g., F344 rats and B6C3F1 mice). Inbred strains may offer greater uniformity of response, more predictable tumor incidence, and better reproducibility than outbred strains. However, their genetic homogeneity may also result in a narrower range of sensitivity to potential carcinogens than exists in random-bred animals. In addition, extrapolation of animal data to humans is the ultimate goal of carcinogenicity studies, and the human population is anything but genetically homogenous.

The ideal species for carcinogenicity bioassays should absorb, metabolize, and excrete the compound under study exactly as humans do. Unfortunately, because of the small number of species that meet the other criteria for selection, there is limited practical utility to this important scientific concept as applied to carcinogenicity studies.

Before concluding this discussion of species/strain selection, it may be worthwhile to take a closer look at the animals preferred by pharmaceutical companies to determine to what extent they meet the conditions described above. Advantages of the CD-1 mouse are (1) a good historical database including various routes of exposure, (2) demonstrated susceptibility to induction of tumors, and (3) relatively low spontaneous incidence of certain tumors to which other strains are highly susceptible, especially mammary and hepatic tumors. Disadvantages are (1) lack of homogeneity, (2) relatively low survival, (3) moderate to high incidence of spontaneous pulmonary tumors and

leukemias, and (4) high incidence of amyloidosis in important organs, including the liver, kidney, spleen, thyroid, and adrenals (Sher et al., 1982).

There has recently been a reduction in survival of Sprague Dawley rats and rats of other strains (FDA, 1993). This reduction may be the result of ad libitum feeding, as preliminary results suggest that caloric restriction may improve survival. Leukemia appears to be the major cause of decreasing survival in the F344 rat. The problem of reduced survival may necessitate a reevaluation of the survival requirements for carcinogenicity studies by regulatory agencies.

## 4 ANIMAL HUSBANDRY

Because of the long duration and expense of carcinogenicity studies, the care of animals used in these studies is of paramount importance. Various physical and biological factors can affect the outcome of these studies. Some important physical factors include light, temperature, relative humidity, ventilation, atmospheric conditions, noise, diet, housing, and bedding (Rao and Huff, 1990). Biological factors include bacteria and viruses that may cause infections and diseases.

The duration, intensity, and quality of light can influence many physiological responses, including tumor incidence (Greenman et al., 1984; Wiskemann et al., 1986). High light intensity may cause eye lesions, including retinal atrophy and opacities (Bellhorn, 1980; Greenman et al., 1982). Rats housed in the top row and the side columns of a rack may be the most severely affected.

The influence of light on the health of animals may be managed in several ways. The animals may be randomly assigned to their cages on a rack such that each column contains animals of a single dose group. The location of the columns on the rack may also be randomized so that the effect of light is approximately equal for all dose groups. In addition, the cages of each column of the rack may be rotated from top to bottom when the racks are changed.

Room temperature has been shown to influence the incidence of skin tumors in mice (Weisbrode and Weiss, 1981). Changes in relative humidity may alter food and water intake (Fox, 1977). Low humidity may cause "ringtail," especially if animals are housed in wire mesh cages (Flynn, 1960).

Diets for rodents in carcinogenesis studies should ideally be nutritionally adequate while avoiding excesses of nutrients that may have adverse effects.

Types of caging and bedding have been shown to affect the incidence and latency of skin tumors in mice. In a study by DePass et al. (1986), benzo[*a*] pyrene-treated mice were housed either in stainless steel cages or polycarbonate shoebox cages with hardwood bedding. The mice housed in shoebox cages developed tumors earlier and with higher frequency than those housed in steel cages.

Housing of rats in stainless steel cages with wire mesh floors may result in decubitous ulcers on the plantar surfaces. This condition may be a significant

clinical problem associated with high morbidity and may affect survival of the animals if euthanasia is performed for humane reasons. Ulcers are particularly frequent and severe in older male Sprague Dawley rats, perhaps because of their large size and weight compared with females and rats of other strains.

Common viral infections may affect the outcome of carcinogenicity studies by altering survival or tumor incidence. Nevertheless, viral infections did not cause consistent adverse effects on survival or tumor prevalence in control F344 rats from 28 NCI/National Toxicalogy Program (NTP) studies, though body weights were reduced by Sendai and pneumonia viruses of mice (Rao et al., 1989). The probability of such infections can be minimized by using viral-antibody-free animals, which are readily available.

## 5 DOSE SELECTION

### 5.1 Number of Dose Levels

In the pharmaceutical industry, most carcinogenicity studies have employed at least three dose levels in addition to the controls, but four levels have occasionally been used (PMA, 1988). The use of three or four dose levels satisfies regulatory requirements (Speid et al., 1990) as well as scientific and practical considerations. If a carcinogenic response is observed, information on the nature of the dose–response relationship will be available. If excessive mortality occurs at the highest dose level, a valid assessment of carcinogenicity is still possible when there is adequate survival at the lower dose levels.

### 5.2 Number of Control Groups

Pharmaceutical companies have most frequently favored the use of two control groups of equal size (PMA, 1988). A single control group of the same size as the treated groups is also used and, less frequently, one double-sized control group may be used. The diversity of study designs reflects the breadth of opinion among toxicologists and statisticians on this issue.

Use of two control groups has the advantage of providing an estimate of the variation in tumor incidence between two groups of animals in the absence of a drug effect. If there are no significant differences between the control groups, the data can be pooled, and the analysis is identical to that using a single, double-sized group. When significant differences occur between the control groups, one must compare the data from the drug-treated groups separately with each control group.

There will be situations in which the incidence of a tumor in one or more drug-treated groups is significantly higher than that of one control group but similar to that of the other control group. In such a situation, it is often helpful to compare the tumor incidences in the control groups to appropriate historical control data. One may often conclude that, for this tumor, one of the

control groups is more "typical" than the other and should therefore be given more weight in interpreting the differences in tumor incidence.

In spite of its current popularity in the pharmaceutical industry, the use of two control groups is opposed by some statisticians on the grounds that a significant difference between the two groups may indicate that the study was compromised by excessive, uncontrolled variation. Haseman et al. (1986), however, analyzed tumor incidence data from 18 color additives tested in rats and mice and found that the frequency of significant pairwise differences between the two concurrent control groups did not exceed that which would be expected by chance alone.

The use of one double-sized group is sometimes preferred because it may provide a better estimate of the true control tumor incidence than that provided by a smaller group. Nevertheless, more statistical power would be obtained by assigning the additional animals equally to all dose groups rather than to the control group only if power is a primary consideration.

### 5.3 Criteria for Dose Selection

Dose selection is one of the most important activities in the design of a toxicology study. It is especially critical in carcinogenicity studies because of their long duration. Whereas faulty dose selection in an acute or subchronic toxicity study can easily be corrected by repeating the study, this situation is much less desirable in a carcinogenicity study, especially since such problems may not become evident until the last stages of the study.

The information used for dose selection usually comes from subchronic toxicity studies, but other information about the pharmacological effects of a drug and its metabolism and pharmacokinetics may also be considered. The maximum recommended human dose (MRHD) of the drug may be an additional criterion if this is known when the carcinogenicity studies are being designed.

For most pharmaceutical companies, the doses selected are as follows. The highest dose is selected to be the estimated maximum tolerated dose (MTD). The lowest dose is usually a small multiple (1–5 times) of the MRHD, and the middose approximates the geometric mean of the other two doses (PMA, 1988; McGregor, 2000).

The MTD is commonly estimated to be the maximum dose that can be administered for the duration of the study that will not compromise the survival of the animals by causes other than carcinogenicity. It should be defined separately for males and females. ICH (1997b) states that the MTD is "that dose which is predicted to produce a minimum toxic effect over the course of the carcinogenicity study, usually predicted from the results of a 90-day study." Factors used to define minimum toxicity include no more than a 10% decrease in body weight gain relative to controls, target organ toxicity, and/or significant alterations in clinical pathology parameters. If the MTD has been chosen appropriately, there should be no adverse effect on survival, only a modest

decrement in body weight gain and minimal overt signs of toxicity. The procedures for dose selection described above are generally consistent with major regulatory guidelines for carcinogenicity studies (Speid et al., 1990; FDA, 1993).[1] There are, however, exceptions to the general approach described above. For example, for nontoxic drugs, the difference between the high and the low doses may be many orders of magnitude if the high dose is set at the estimated MTD and the low dose is a small multiple of the clinical dose. Some guidelines request that the low dose be no less than 10% of the high dose (Speid et al., 1990). In this situation, it may be acceptable to set the high dose at 100 times the MRHD even if the MTD is not achieved (Speid et al., 1990). Similarly, when a drug is administered in the diet, the highest concentration should not exceed 5% of the total diet whether or not the MTD is achieved (Japanese Ministry of Health and Welfare, 1989).

Metabolism and/or pharmacokinetic data, when available, should also be considered in the dose selection process. It is desirable that a drug not be administered at such a high dose that it is excreted in a different manner than at lower doses, such as the MRHD. Similarly, the high dose should not lead to the formation of metabolites other than those formed at lower (clinical) doses. If data show that a given dosage produces maximum plasma levels, administration of higher doses should be unnecessary. These considerations may be very useful when interpreting the results of the study or attempting to extrapolate the results to humans.

## 6 GROUP SIZE

The minimum number of animals assigned to each dose group in pharmaceutical carcinogenicity studies is 50 of each sex (PMA, 1988). Most companies, however, use more than the minimum number, and some use up to 80 animals per sex per group. The most important factor in determining group size is the need to have an adequate number of animals for a valid assessment of carcinogenic activity at the end of the study. For this reason, larger group sizes are used when the drug is administered by daily gavage because this procedure may result in accidental deaths by perforation of the esophagus or aspiration into the lungs. Larger group sizes are also used when the carcinogenicity study is combined with a chronic toxicity study in the rat. In this case, serial sacrifices are performed at 6 and 12 months to evaluate potential toxic effects of the drug.

In the final analysis, the sensitivity of the bioassay for detecting carcinogens is directly related to the sample size. Use of the MTD has often been justified based on the small number of animals at risk compared to the potential human

[1] Note the *FDA Redbook* applies, strictly speaking, only to food additives. It is cited here because it is a well-known toxicology guideline routinely applied to animal pharmaceuticals to which humans may be exposed. The *Redbook* has recently been updated by the FDA (1993).

population, in spite of the difficulties inherent in extrapolating effects at high doses to those expected at much lower clinical doses. A reasonable compromise may be the use of doses lower than the MTD combined with a larger group size than the 50 per sex minimum accepted by regulatory agencies.

## 7 ROUTE OF ADMINISTRATION

In the pharmaceutical industry, the two most common routes of administration are via diet and gavage (PMA, 1988). Some compounds are given by drinking water, topical (dermal) application, or injection, depending on the expected clinical exposure route, which is the primary criterion for determining the route of administration in carcinogenicity studies. When more than one clinical route is anticipated for a drug, the dietary route is often chosen for practical reasons.

Dietary administration is often preferred over gavage because it is far less labor intensive. Another advantage is that the MTD has rarely been overestimated in dietary studies, whereas it has often been overestimated in gavage studies, according to data from the NTP (Haseman, 1985). The dietary route is unsuitable for drugs that are unstable in rodent chow or unpalatable. The dietary route is also disadvantaged by the fact that dosage can only be estimated based on body weight and food intake data, in contrast with gavage, by which an exact dose can be given. Disadvantages of gavage testing are the likelihood of gavage-related trauma, such as puncture of the trachea or esophagus, and possible vehicle (e.g., corn oil) effects.

When doing studies by the dietary route, the drug may be administered as a constant concentration at each dose level or the concentration may be increased as body weight increases to maintain a constant dose on a milligram-per-kilogram basis. The latter method allows greater control of the administered dose and avoids age- and sex-related variations in the dose received, which occur with the former method. Both methods are acceptable to regulatory agencies.

## 8 STUDY DURATION

The duration of carcinogenicity studies for both rats and mice is 2 years in most pharmaceutical laboratories (PMA, 1988). Occasionally, rat studies are extended to 30 months, while come companies terminate mouse studies at 18 months. The difference in duration between mouse and rat studies is based on the belief that rats have a longer natural life span than mice. Recent data indicate, however, that this is not the case. The most commonly used strains, the Sprague Dawley rat and the CD-1 mouse, have approximately equal survival at 2 years based on industry data (PMA, 1988). The same is true for the most popular inbred strains, the F344 rat and the B6C3F1 mouse (PMA, 1988).

Data from NCI studies confirm that the 2-year survival of the B6C3F1 mouse is at least equal to, if not greater than, that of the F344 rat (Cameron et al., 1985)

## 9 SURVIVAL

As stated earlier, adequate survival is of primary importance in carcinogenicity studies because animals must be exposed to a drug for the greater part of their life span to increase the probability that late-occurring tumors can be detected. Early mortality, resulting from causes other than tumors, can jeopardize the validity of a study because dead animals cannot get tumors.

In general, the sensitivity of a carcinogenicity bioassay is increased when animals survive to the end of their natural life span because weak carcinogens may induce late-occurring tumors. The potency of a carcinogen is often inversely related to the time to tumor development. By analogy, as the dose of a carcinogen is reduced, the time to tumor occurrence is increased (Littlefield et al., 1979; DePass et al., 1986).

Why do we not allow all animals on a carcinogenicity study to live until they die a natural death if by so doing we could identify more drugs as carcinogens? In fact, the sensitivity of a bioassay may not be improved by allowing the animals to live out their natural life span because the incidence of spontaneous tumors tends to increase with age. Thus, depending on the tumor type, the ability of the bioassay to detect a drug-related increase in tumor incidence may actually decrease, rather than increase, with time. Therefore, the optimum duration of a carcinogenicity study is that which allows late-occurring tumors to be detected but does not allow the incidence of spontaneous tumors to become excessive.

Reduced survival in a carcinogenicity study may or may not be drug related. Sometimes, the MTD is exceeded and increased mortality occurs at the highest dose level and, occasionally, at the middose level as well. This situation may not necessarily invalidate a study; in fact, the protocol may be amended to minimize the impact of the drug-induced mortality. For example, cessation of drug treatment may enhance the survival of the animals in the affected groups and allow previously initiated tumors to develop. As shown by Littlefield et al. (1979) in the National Center for Toxicological Research (NCTR) ED01 study, liver tumors induced by 2-acetylaminofluorene, which appeared very late in the study, were shown to have been induced much earlier and not to require the continuous presence of the carcinogen to develop. By contrast, bladder tumors that occurred in the same study were dependent on the continued presence of the carcinogen.

Whether drug treatment is terminated or not, drug-related toxicity may also be managed by performing complete histopathology on animals in the lower dose groups rather than on high-dose and control animals only. If there is no increase in tumor incidence at a lower dose level that is not compromised by

reduced survival, the study may still be considered valid as an assessment of carcinogenicity.

When reduced survival is related to factors other than excessive toxicity, the number of animals at risk for tumor development may be inadequate, and the validity of the study may be compromised even in the absence of a drug effect on survival. Obviously, the adjustments described above for excessive, drug-related toxicity are not relevant to this situation.

There is no unanimity of opinion among regulatory agencies as to the minimum survival required to produce a valid carcinogenicity study or as to the best approach for dealing with survival problems. Even within a single agency such as the FDA, different opinions exist on these issues. For example, the recently issued *FDA Redbook II Draft Guideline* (FDA, 2000) requires that rats, mice, or hamsters be treated for 24 months. Early termination due to decreased survival is not recommended. The European Economic Commission (EEC) guidelines differ in that they suggest termination of the study when survival in the control group reaches 20%, while the Japanese guideline suggests termination at 25% survival in the control or low-dose groups (Speid et al., 1990). These provisions make good sense in that they do not request termination of the study when drug-related mortality may be present only at the highest dose.

## 10 ENDPOINTS MEASURED

A carcinogenicity study is more focused than a chronic toxicity study—fewer endpoints are evaluated, and as such it is a simpler study. The key endpoints are actually few:

- Pathology (limited to neoplastic and preneoplastic tissue transformations
- Body weight (to ensure that toxicity is not so great as to invalidate the assays and also that it is just sufficient to validate the assay)
- Survival (key to determining when to terminate the study)
- Clinical pathology (limited to evaluating the morphology of white blood cells, and usually this is actually deferred until there are indications that such data are needed)
- Food consumption (actually measured to ensure that dietary administration doses are accurate)

Only pathology will be considered in detail.

The primary information for carcinogenicity evaluation is generated by pathologists. Table 1 lists the tissues normally collected, processed, and evaluated. These professionals, like any other group of professionals, vary in their training and experience, and these are characteristics which may influence the evaluation in a number of ways. Some of these are listed below:

**TABLE 1 Standard Tissue List**

| | | |
|---|---|---|
| Kidney | Urinary bladder | Aorta |
| Heart | Trachea | Lungs |
| Liver | Gall bladder | Pancreas |
| Fat | Salivary gland | Spleen |
| Cervical lymph node | Mesenteric lymph node | Thymus |
| Tongue | Esophagus | Stomach |
| Duodenum | Jejunum | Ileum |
| Cecum | Colon | Mammary gland |
| Skin | Skeletal muscle | Sciatic nerve |
| Parathyroid | Thyroid | Adrenal |
| Pituitary | Prostate | Seminal vesicles |
| Testes | Epididymides | Ovaries |
| Oviducts | Uterine horns | Uterine body |
| Cervix | Vagina | Brain |
| Spinal cord | Sternum | Rib/bone |
| Eyes | Harderian glands | BM smear |
| Nares | Clitoral/preputial gland | Zymbal's gland |
| Gross lesions | | |

1. Differences in terminology may be important when considering controversial lesions.
2. Lack of consistency throughout a study is likely when a pathologist has only recently become involved with rodent carcinogenicity. Training is often in a clinical situation (especially in Europe), where each animal or person is unique and there is in a rodent carcinogenicity study consisting of 500 animals.
3. Unfamiliarity with the observed lesion in a particular species may cause problems in interpretation.

Possible bias introduced by knowledge of treatment can be corrected in several ways, but the use of a two-stage process would seem to be most efficient:

1. An initial evaluation is performed with full knowledge of the animal's history, including treatment.
2. A second evaluation of specific lesions is then carried out. This should be done blind, either by the same pathologist or, preferably, by the same and a second pathologist.

Differences in evaluation between pathologists should always be discussed by them to resolve the differences; they may be due to subtle differences in diagnosis and do not indicate incompetence in one of the pathologists. It is unacceptable for a study sponsor to shop around until he or she finds a pathologist who gives, for whatever reason, the expected result without giving

an opportunity for interaction with all of the other evaluators. Sometimes these diagnoses are given years apart, during which time understanding of the pathogenesis of lesions may change, and even the first pathologist may not arrive at the same conclusion as he or she did some years ago.

Evaluation of the data is not purely a statistical exercise. A number of important factors should be considered: (1) dose–effect relationship; (2) a shift toward more anaplastic tumors in organs where tumors are common; (3) earlier appearance of tumors; and (4) presence of preneoplastic lesions.

The language used to describe the carcinogenic response has masked its complexity and presents a stumbling block to its understanding among non-histopathologists. Benign or malignant neoplasms do not arise without some precursor change within normal tissue. An important concept in carcinogenicity evaluation is that of neoplastic progression, which was derived from studies on skin tumors (Berenblum and Shubik, 1947) and expanded to a number of other tissues (Foulds, 1969, 1975). There is, on many occasions, a far from clear distinction between hyperplastic and "benign" neoplasia and between benign and malignant neoplasia.

Hyperplasia and benign and malignant neoplasia are convenient medical terms with prognostic significance. Hyperplasia can occur either as a regenerative response to injury, with no neoplastic connotations, or as a sustained response to a carcinogenic agent. It is an increase in the number of normal cells retaining normal intercellular relationships within a tissue. This normally may break down, resulting in altered growth patterns and altered cellular differentiation—a condition which may be described as atypical hyperplasia or presumptively as preneoplastic lesions. Possible sequelae to hyperplasia are (1) persistence without qualitative change in either structure or behavior; (2) permanent regression; (3) regression with later reappearance; and (4) progression to develop new characteristics indicating increased probability of malignancy. The last of these in the least likely to occur in experimental multistage models, such as in mouse skin or rat liver, where large numbers of hyperplastic lesions may occur but notably fewer carcinomas develop from them.

Benign neoplasms in most rodent tissues apparently arise in hyperplastic foci, for example, squamous cell papillomas of the skin and forestomach. Furthermore, these papillomas seldom demonstrate autonomous growth and even fewer progress to squamous cell carcinomas (Burns et al., 1976; Colburn, 1980). This decisive progression to carcinoma, when it occurs, provides powerful evidence for the multistage theory of carcinogenesis: the new, malignant cells arising as a focus within the papilloma or even in an area of hyperplasia since the papilloma is not a necessary intermediate stage. In other organs, benign neoplasia is usually characterized by well-differentiated cell morphology, a fairly uniform growth pattern, clear demarcation from surrounding tissues, and no evidence of invasion. The progression toward malignancy involves anaplasia (loss of differentiation) and pleomorphism (variety of phenotypic characteristics within the neoplasm). These changes may be focal in an otherwise

**TABLE 2 Tumor-Bearing Animals in Control Groups from Rodent Studies**

| Control Animals for 2-Year NTP Bioassay | No. of Animals | Percent with Tumors | | |
|---|---|---|---|---|
| | | Malignant | Benign | Total |
| B6C3F1 mice | | | | |
| Male | 1692 | 42 | 35 | 64 |
| Female | 1689 | 45 | 33 | 64 |
| F344 rats | | | | |
| Male | 1596 | 55 | 95 | 98 |
| Female | 1643 | 38 | 76 | 88 |
| Osborne–Mendel rats | | | | |
| Male | 50 | 26 | 68 | 78 |
| Female | 50 | 12 | 80 | 88 |
| Sprague Dawley rats | | | | |
| Male | 56 | 9 | 36 | 39 |
| Female | 56 | 30 | 68 | 79 |

*Source*: Haseman, unpublished summary of U.S. NTP data.

benign neoplasm and may vary in degree and extent. Evidence of invasion of the surrounding tissues or of metastasis is not an essential characteristic of malignancy, although their presence strengthens the diagnosis.

The grouping together of certain tumor types can aid statistical analysis, but it must be done carefully, with full appreciation of the biology and whatever is known of the pathogenesis of the lesions. Grouping for analysis of all animals showing neoplasia, irrespective of the tumor type, is inappropriate because the incidence in most treatment control groups can be very high and, in U.S. NTP studies, approaches 100% in rats and 50–70% in mice (Table 2).

There may be similar incidences of tumors in aging people, but the real prevalence of tumors in human populations is uncertain. In the United States, where autopsies are uncommon, over one-third reveal previously undiagnosed cancers when they are conducted (Silverberg, 1984). A single type of neoplasm, renal adenoma, is present in 15–20% of all adult kidneys (Holm-Nielson and Olsen, 1988), although it is unclear whether these 2–6-mm foci of proliferating tubular and papillary epithelium represent small carcinomas or benign precursors of renal cell carcinomas. Irrespective of the significance of these lesions in human pathology, the presence of similar foci in a rodent carcinogenicity experiment would trigger the recording of renal tumor–bearing animals and, hence, their consideration in the statistical and pathological evaluation processes. Evaluation is further complicated by the increased background incidences of tumors as animals get older.

The independent analysis of every different diagnosis in rodent studies would also mask significant effects in many cases while enhancing them in others. Benign and malignant neoplasms of a particular histogenesis are often grouped because the one is seen as a progression from the other. However, this grouping may result in a nonsignificant difference from the controls

because there has been an acceleration of progression toward malignancy, the incidence of benign neoplasms decreasing while the malignant neoplasms increase. Guidelines are available for "lumping" or "splitting" tumor types, but in using them, the basis for the classification of neoplastic lesions should be clarified, especially when data generated over several or many years are coupled, since diagnostic criteria and ideas regarding tumor histogenesis may have changed. Reliance on tabulated results alone can lead to serious misinterpretation by those not closely connected with a particular study. For this very important reason, the pathology and toxicology narrative should be full and clear. If it is not, then there will always be doubts about future interpretations, even if these doubts are not, in reality, justified.

## 11 STATISTICAL ANALYSIS

Irrespective of the specific protocols used, all carcinogenicity studies end with a statistical comparison of tumor proportions between treated and control groups. This analysis is necessary because the control incidence of most tumor types is rarely zero. In the unlikely case that a type of tumor is found in treated animals but not in concurrent or appropriate historical controls, it is reasonable to conclude that the tumor is drug related without statistical analysis.

Most pharmaceutical companies analyze tumor data using mortality-adjusted methods (PMA, 1988). Peto/International Agency for Research on Cancer (IRC) methodology is most commonly used, perhaps because this method is currently favored by the FDA (Peto et al., 1980). The use of life table methods is most appropriate for "lethal" tumors, that is, those that cause the death of the animals. Various statistical methods are available for analyzing the incidence of the lethal and nonlethal tumors (e.g., Gart et al., 1979, 1986; Chu et al., 1981; Dinse and Lagakos, 1983; McKnight, 1988; Portier and Bailer, 1989; Gaylor and Kodell, 2001). These methods are especially useful when there are drug-related differences in mortality rates. When there is no drug effect on survival, unadjusted methods will generally give the same results.

As a general approach, most pharmaceutical statisticians begin by testing for the presence of a dose-related trend in tumor proportions. If the trend test is significant, that is the $p$ value is less than or equal to 0.05, pairwise comparisons are performed between the treated and control groups. Trend and pairwise analyses may be adjusted for mortality as stated earlier or performed without mortality adjustment using such simple methods as chi-square or Fisher's exact tests.

Although in most cases the use of trend tests is appropriate since most biological responses are dose related, there are exceptions to this rule. Certain drugs, especially those with hormonal activity, may not produce classical dose responses and may even induce inverse dose–response phenomena. In these cases, a pairwise comparison may be appropriate in the absence of a significant positive trend.

Most (70%) pharmaceutical companies use one-tailed comparisons, and a substantial number use two-tailed methods (PMA, 1988). Since regulatory agencies are primarily interested in identifying carcinogenic drugs, as opposed to those that inhibit carcinogenesis, the use of one-tailed tests is generally considered more appropriate. Some companies prefer two tailed comparisons because, in the absence of a true carcinogenic effect, there is an equal probability of seeing significant decreases as well as significant increases by chance alone.

One of the most important statistical issues in the analysis of carcinogenicity data is the frequency of "false positives," or type I errors. Because of the multiplicity of tumor sites examined and the number of tests employed, there is concern that noncarcinogenic drugs may be erroneously declared carcinogens. If an $p < 0.05$ increase in tumor incidence is automatically regarded as a biologically meaningful result, then the false-positive rate may be as high as 47–50% (Haseman et al., 1986).

Several statistical procedures designed to correct for the multiplicity of significance tests have been published (and reviewed by Haseman, 1990). One approach to the problem of multiple tumor site/type testing is a procedure attributed to Tukey by Mantel (1980). This method is used to adjust a calculated $p$ value based on the number of tumor types/sites for which there are minimum number of tumors in the particular study. The reasoning here is that, for tumor sites, the number of tumors found is so small that it is impossible to obtain a significant result for that tumor site no matter how the tumors might have been distributed among the dose groups. Only those sites for which a minimum number of tumors is present can contribute to the false-positive rate for a particular study.

A method proposed by Schweder and Spjotvoll (1982) is based on a plot of the cumulative distribution of observed $p$ values. Farrar and Crump (1988) have published a statistical procedure designed not only to control the probability of false-positive findings but also to combine the probabilities of a carcinogenic effect across tumor sites, sexes, and species.

Another approach to controlling the false-positive rate in carcinogenicity studies was proposed by Haseman (1983). Under this "rule," a compound would be declared a carcinogen if it produced an increase significant at the 1% level in a common tumor or an increase significant at the 5% level in a rare tumor. A rare neoplasm was defined as a neoplasm that occurred with a frequency of less than 1% in control animals. The overall false-positive rate associated with this decision rule was found to be not more that 7–8% based on control tumor incidences from NTP studies in rats and mice. This false-positive rate compares favorable with the expected rate of 5%, which is the probability at which one would erroneously conclude that a compound was a carcinogen. The method is notable for its simplicity and deserves serious consideration by pharmaceutical statisticians and toxicologists. Without resorting to sophisticated mathematics, this method recognizes the fact that tumors differ in their spontaneous frequencies and, therefore, in their contribution to

the overall false-positive rates in the carcinogenicity studies. False-positive results are much less likely to occur at tissue sites with low spontaneous tumor incidences than at those with high frequencies.

As a final point that has special relevance to pharmaceutical carcinogenicity studies, one may question whether the corrections for multiple comparisons and their effect on the overall false-positive rate are appropriate for all tumor types. For example, if a compound is known to bind to receptors and produce pharmacological effects in a certain organ, is it justified to arbitrarily correct the calculated $p$ value for the incidence of tumors in that organ using the methods described above? It is difficult to justify such a correction considering that the basis for correcting the calculated $p$ value is that the true probability of observing an increased incidence of tumors at any site by chance alone may be much higher than the nominal $\alpha$ level (usually 0.05). It is reasonable to expect that, when a drug has known pharmacological effects on a given organ, the probability of observing an increased tumor incidence in that organ by chance alone is unlikely to be higher than the nominal 5% $\alpha$ level.

Although most pharmaceutical statisticians and toxicologists agree on the need to control the probability of false-positive results, there is no consensus as to which method is most appropriate or most acceptable to regulatory agencies. The FDA and other such agencies will accept a variety of statistical procedures but will often reanalyze the data and draw their own conclusions based on their analyses.

## 12 TRANSGENIC MOUSE MODELS

Sine the early 1970s, the standard for adequate evaluation of the carcinogenic potential of a candidate pharmaceutical has been the conduct of lifetime, high-dose assays in two species—almost always the rat and the mouse.

The relevance (and return on investment) for the bioassays performed in mice has been questioned for some time. In 1997, ICH opened the possibility for the substitution of some form of short- or medium-term mouse test as an alternative to the traditional lifetime mouse bioassay. The FDA has subsequently stated that it would accept "validated" forms of a set of medium-term mouse studies based on transgenic models, and significant effort has since gone into such validation.

The huge advances made in molecular biology since the late 1980s have provided the possibility of approaches to evaluating chemicals and potential drugs for carcinogenic potential in approaches which are different, less expensive, and take a shorter period of time than traditional long-term bioassays. This work has also been stimulated by dissatisfaction with the performance of traditional test systems.

The traditional long-term bioassays use highly inbred animals developed with the goal of reducing the variability in background tumor incidences as a means of increasing the statistical sensitivity of the bioassays. This inbreeding

has led to narrowing of the allelic pool in the strains of animals that are currently used for testing, as opposed to the wild-type populations (of humans) that the tests are intended to protect (Festing, 1979). Transgenic models should serve to improve the identification of carcinogens by providing the gene-specific mechanistic data, minimizing the influence of spontaneous tumors and strain-specific effects, and reducing time required. Cost and animal usage should also be reduced (Eastin et al., 1998).

As it has become possible to transfer new or engineered genes to the germ lines of mammals, the results have been transgenic mice that can be used in shorter term in vivo assays for carcinogenicity and which are also useful for research into the characterization of genotoxic events and mechanisms in carcinogenesis. By coupling reporter phenotypes (such as papilleomas in the Tg·AC mouse, the task of "reading" results in test animals is made much less complex.

There are four transgenic mouse models that have been broadly evaluated—the TSPp53$^{+/-}$, the TgAC, the Hras2, and the XPA$^{-/-}$. Each of these has its own characteristics. Each of these merits some consideration. They are each made by either zygote injection or specific gene targeting in embryonic cells (McAnulty, 2000; French et al., 1999).

## 12.1 Tg·AC Mouse Model

This was the earliest of the models to be developed, and its use in mouse skin carcinogenicity studies was first reported in 1990. The mice have four copies of the v-H-*ras* oncogene in tandem on chromosome 11, and the transgene is fused with a fetal ξ-globin gene which acts as a promoter. The transgene codes for a switch protein which is permanently "on," and this results in the mice having genetically initiated skin. The application of tumor promoters to the surface of the skin causes the rapid induction of pedunculate papillomas that arise from the follicular epithelium. This is despite the fact that the transgene is not expressed in the skin, although it is present in the papillomas that form and also in the focal follicular hyperplastic areas that are the precursors to the papillomas. In about 40% of the mice, the papillomas become malignant skin tumors—mainly squamous cell carcinomas and sarcomas.

The first assessments of this model as an alternative to traditional carcinogenicity studies were performed by the U.S. National Institute of Environmental Health Sciences (NIEHS) and NTP, and the results with over 40 chemicals have been published. The majority of studies were performed by skin painting, regardless of whether the product was a dermal or systemic carcinogen. However, a good correlation was found with the known carcinogenicity of the test compounds, and both mutagens and nonmutagens were identified. It was found that great care had to be taken with the skin because damage could also induce papillomas, which means that these animals cannot be identified using transponder chips. This sensitivity may also explain some of the false positive results that have occurred with resorcinol and rotenone. Of more concern is

that there have been false-negatives with known carcinogens, namely ethyl acrylate and *N*-methyl-*O*-acrylamide. The model was designed for use in the context of the two-stage model of carcinogenesis with the underlying mechanistic pathway involving specific transcription factors, hypomethylation, and cell-specific expression of the results. Along with the p53, this model has seen the widest use and evaluation (in terms of number of agents evaluated) so far. The carrier mouse strain employed, the FVB/N, is not commonly employed in toxicology and is prone to sound-induced seizures. It may be that the dermal route is not suitable for all systemic carcinogens, and this is the reason that in the International Life Sciences Institute (ILSI) program both the dermal and systemic routes are being investigated in this model.

Another problem with this model was the occurrence of a nonresponder genotype to positive control agents. This was found to be attributable to a rearrangement of the ζ-globin promoter region, but it is claimed that this problem has been resolved. However, this has considerably delayed the ILSI studies with this model, but all data should be available in time for the November meeting. It is already clear that the model gives a robust response to the positive control agent, 12-*O*-tetradecanoylphorbol 13-acetate (TPA).

### 12.2 Tg.rasH2 Mouse Model

This model was developed at the Central Institute for Experimental Animals (CIEA) in Japan, and the first information about the mouse was published in 1990. The mice have five or six copies of the human H-*ras* protooncogene inserted in tandem into their genome surrounded by their own promoter and enhancer regions. This transgene has been very stabile, with no loss of responsiveness sine the model was developed. The transgene codes for a molecular switch protein in the same way as the previous model, but the transgene is expressed in all organs and tissues. Thus the response endpoint is not primarily dermal.

The initial studies with this model revealed a rapid appearance of forestomach papillomas with *N*-methyl-*N*-nitrosourea (MNU), and this compound has already been used as the positive control agent in subsequent studies with this strain. A study duration of six months is sufficient to obtain a positive response, and longer periods should be avoided because the mice start to develop various spontaneous tumors, such as splenic hemangiosarcomas, forestomach and skin papillomas, lung and Harderian gland adenocarcinomas, and lymphomas. It has a high level of constitutive expression and some spontaneous tumors even when the animals are younger. It is, however, very responsive to carcinogens—one gets a rapid onset after exposure and a higher response incidence than with the other models. The underlying mechanism is still not certain.

A large number of studies have been run in this strain in Japan in advance of the ILSI program. The model is sensitive to both mutagenic and nonmutagenic carcinogens, although cyclophosphamide and furfural have

given equivocal results in each category, respectively. The majority of noncarcinogens have also been identified correctly, although, again, there are a small number of compounds that have given equivocal results. In the ILSI program, 24 of the 25 studies will have been completed in time for the November meeting, and the final studies will be completed during 2001.

## 12.3 p53$^{+/-}$ Mouse Model

The thrombospondin (TSP) *p53*$^{+/-}$, hereafter referred to as the *p53* (the designation of the tumor suppressor gene involved), is a heterozygous knockout with (up to seven or so months of age) a low spontaneous tumor incidence. It is responsive to the genotoxic carcinogens by a mechanism based on the fact that many (but not all) tumors show a loss of the wild-type allele. The p53 has been extensively worked on by Tennant's group at NIEHS (Tennant et al., 1995, 1999). This model was developed in the United States and carries a hemizygous knockout of the *p53* gene which was developed by integrating a mutated copy of the gene into the genome of mice. The *p53* gene is known as a tumor suppressor gene, and it is the most commonly mutated gene in human malignancies. It searches for a protein transcription factor which activates multiple genes when damage to DNA strands occurs, and this in turn leads to either the arrest of the cell cycle while DNA repair occurs or to apoptosis (programmed cell death), which removes the damaged cell. The heterozygote is used because homozygotes show a very high incidence of spontaneous tumors within a few months of birth. The heterozygotes have a low background incidence of tumors up to 12 months, but during this time there is a high chance of a second mutagenic event occurring—following exposure to a carcinogen, for example—and this would result in a loss of suppressor function or an increase in transforming activity.

The initial studies with this model as an alternative in traditional carcinogenicity testing were performed at the NIEHS, and these suggested that it was sensitive to mutagenic carcinogens such as benzene and *p*-cresidine within six months. Nonmutagenic carcinogens were negative in the assay, as were mutagenic noncarcinogens. However, subsequent studies and some parts of the ILSI program have shown clear indications that a six-month duration may be insufficient. In particular, benzene has given negative or equivocal results within six months, although positive results have been obtained by extending the study to nine months. It will be very important to assess the results of the ILSI program when deciding the best study duration for this model. This is the most popular model in the United States.

## 12.4 XPA$^{-/-}$ Mouse Model

This was the last of the models to be developed and was created using a knockout technique after the *XPA* gene had been cloned. The first data were published by the National Institute for Public Health and the Environment

(RIVM) in the Netherlands in 1995. Both alleles of the *XPA* gene have been inactivated by a homologous recombination in ES cells, resulting in a homozygous deletion of the gene spanning exons 3 and 4. The protein coded by this gene is essential for the detection and repair of DNA damage using the nucleotide excision repair (NER) pathway. This model only has between 2 and 5% of residual NER activity.

The initial studies at RIVM demonstrated that exposure of these mice to UV-B radiation or 7,12-dimethylbenz[*a*]anthracene resulted in the rapid induction of skin tumors. It was also shown that various internal tumors could be induced following oral administration of mutagenic carcinogens such as benzo[*a*]pyrene (B[a]P) and 2-acetylaminofluorine (2-AAF). The early studies suggested that this response could occur within six months, but further experience has indicated that a nine-month treatment period is essential in order to obtain a response with positive control agents such as B[a]P, 2-AAF, and *p*-cresidine.

All of the 13 studies that have been undertaken with this model were available for review at the November 2000 meeting. The model is sensitive to both UV and genotoxic carcinogens and also to some nonmutagenic carcinogens, such as diethylstilbestrol (DES), Wy-14,643, and cyclosporin A. There have been no false positives with noncarcinogens. Some laboratories have also investigated a double transgenic $XPA^{-/-}$ $p53^{+/-}$ model, and this seems to increase the sensitivity of the assay. For example, in a DES study, seven animals with metastasizing osteosarcomas were found in the double transgenic group, compared with one in the *XPA* group and none among the wild-type animals. There remains concern (as with any new model) that these models may be overly sensitive or (put another way) that the relevance of positive findings to risk in humans may not be clear. The results of the ILSI Health and Environmental Sciences Institute (HESI) workshop seem to minimize these concerns.

It is generally proposed that while such models can improve the identification of carcinogens in three ways (providing gene-specific mechanistic data, minimizing the influence of spontaneous tumors and strain-specific effects, and reducing the time, cost, and animal usage involved), they have two potential uses in pharmaceutical development. These are in lieu of either the mouse two-year cancer bioassay or subchronic toxicity assessments prior to making a decision to commit to a pair of two-year carcinogenicity bioassays.

As performance data have become available on these strains, ICH (1997a) has incorporated their use into pharmaceutical testing guidelines in lieu of the second rodent species tests (that is, to replace the long-term mouse bioassay when the traditional rat study has been performed). The FDA has stated that it would accept such studies when "performed in a validated code." In fact, the Center for Biologics Evaluation and Research (CBER) has accepted such studies as a sole carcinogenicity bioassay in some cases where there was negative traditional genotoxicity data and strong evidence of a lack of a mechanistic basis for concern.

A joint ILSI and HESI validation program has been completed looking at the results of the four prime candidate models in identifying carcinogens as compared to the results of traditional long-term rodent bioassays. This validation program involved 51 different laboratories and imposed protocol standards to allow comparison of results. Three dose levels were studied per chemical, with 15 males and 15 females being used for each dose group. A vehicle and high-dose control in wild-type animals was also included, with information from NTP bioassays and 4-week range-finding assays being used to help set doses. Animals were dosed for 26 weeks. The issues coming in and out of these validation programs bear consideration (Tennant et al., 1999):

- Is there proper comparator data for evaluating the performance of human or rodent bioassays? It should be kept in mind that there are sets of rodent bioassay data (particularly those involving liver tumors in mice) that are widely accepted as irrelevant in the prediction of human risk.
- How will the data from these assays be incorporated into any weight-of-evidence approach to assessing human health risk?
- What additional mechanistic research needs to be undertaken to improve our understanding of the proper incorporation and best use of the data from these assays?
- How can the results of current validation efforts be best utilized in the timely evaluation of the next generation of assays?
- Given that, at least under some conditions, assays using these models tend to "blow up" (have high spontaneous tumor rates) once the animals are more than eight or nine months of age, how critical are age and other not currently apprehended factors to optimizing both sensitivity and specificity?
- How wide and unconditional will FDA (and other regulatory bodies) acceptance be of these models in lieu of the traditional two-year mouse bioassay?

## 13 INTERPRETATION OF RESULTS: CRITERIA FOR POSITIVE RESULT

There are three generally accepted criteria for a positive result in a carcinogenicity study. The first two are derived directly from the results of the statistical analysis: (1) a statistically significant increase in the incidence of a common tumor and (2) a statistically significant reduction in the time to tumor development. The third criterion is the occurrence of very rare tumors, that is, those not normally seen in control animals, even if the incidence is not statistically significant. Table 3 presents an evaluation matrix for these factors.

**TABLE 3 Interpretation of Analysis of Tumor Incidence and Survival Analysis (Life Table)**

| Outcome Type | Tumor Association with Treatment[a] | Mortality Association with Treatment | Interpretation |
|---|---|---|---|
| A | – | + | Unadjusted test[b] may underestimate tumorigenicity of treatment. |
| B | + | + | Unadjusted test gives valid picture of tumorigenicity of treatment. |
| C | + | – | Tumors found in treated groups may reflect longer survival of treated groups. Time-adjusted analysis is indicated. |
| D | – | + | Apparent negative findings on tumors may be due to the shorter survival in treated groups. Time-adjusted analysis and/or a retest at lower doses is indicated. |
| E | – | 0 | Unadjusted test gives a valid picture of the possible tumor-preventive capacity of the treatment. |
| F | – | – | Unadjusted test may underestimate the possible tumor-preventive capacity of the treatment. |
| G | 0 | + | High mortality in treated groups may lead to unadjusted test missing a possible tumorigen. Adjusted analysis and/or retest at lower doses is indicated |
| H | 0 | 0 | Unadjusted test gives a valid picture of lack of association with treatment. |
| I | 0 | – | Longer survival in treated groups may mask tumor-preventive capacity of treatment. |

[a]+ = Yes, – = No, and 0 = no bearing on discussion.
[b]The unadjusted test referred to here is a contingency table type of analysis of incidence, such as Fisher's exact test.

## 14 STATISTICAL ANALYSIS

The actual statistical techniques used to evaluate the results of carcinogenicity bioassays basically utilize four sets of techniques, three of which have been presented earlier in this book:

- Exact tests
- Trend tests
- Life tables (such as log rank techniques)
- Peto analysis

These are then integrated into the decision-making schemes discussed earlier in this chapter. The methods themselves and alternatives are discussed elsewhere in detail (Chow and Lin, 1998; Gad, 2006).

## 14.1 Exact Tests

The basic forms of these (the Fisher exact test and chi square) have previously been presented, and the reader should review these. Carcinogenicity assays are, of course, conducted at doses that are at least near those that will compromise mortality. As a consequence, one generally encounters competing toxicity producing differential mortality during such a study. Also, often, particularly with certain agricultural chemicals, latency of spontaneous tumors in rodents may shorten as a confounded effect of treatment with toxicity. Because of such happenings, simple tests on proportions, such as $\chi^2$ and Fisher–Irwin exact tests on contingency tables, may not produce optimal evaluation of the incidence data. In many cases, however, statisticians still use some of these tests as methods of preliminary evaluation. These are unadjusted methods without regard for the mortality patterns in a study. Failure to take into account mortality patterns in a study sometimes causes serious flaws in interpretation of the results. The numbers at risk are generally the numbers of animals histopathologically examined for specific tissues.

Some gross adjustments on the numbers at risk can be made by eliminating early deaths or sacrifices by justifying that those animals were not at risk to have developed the particular tumor in question. Unless there is dramatic change in tumor prevalence distribution over time, the gross adjusted method provides fairly reliable evidence of treatment effect, at least for nonpalpable tissue masses.

## 14.2 Trend Tests

Basic forms of the trend tests [such as that of Tarone (1975)] have previously been presented in this text.

Group comparison tests for proportions notoriously lack power. Trend tests, because of their use of prior information (dose levels), are much more powerful. Also, it is generally believed that the nature of true carcinogenicity (or toxicity for that matter) manifests itself as dose–response. Because of the above facts, evaluation of trend takes precedence over group comparisons. In order to achieve optimal test statistics, many people use ordinal dose levels (0,1,2, …) instead of the true arithmetic dose levels to test for trend. However, such a decision should be made a priori. Example 1 demonstrates the weakness of homogeneity tests.

As is evident from this example, often group comparison tests will fail to identify significant treatment but trend tests will. The same arguments apply to survival-adjusted tests on proportions as well. In an experiment with more than one dose group ($K > 1$), the most convincing evidence for carcinogenicity

**Example 1** Trend Versus Heterogeneity

| Number at Risk | Number with Tumor | Dose Level |
|---|---|---|
| 50 | 2 | 0 |
| 50 | 4 | 1 |
| 50 | 6 | 2 |
| 50 | 7 | 3 |

*Cochran–Armitage Test for Trend*

| | Calculated $\chi^2$ Subgroup | Degrees of Freedom | $\alpha$ | Two-tailed *p* |
|---|---|---|---|---|
| Trend | 3.3446 | 1 | 0.0500 | 0.0674 |
| Departure | 0.0694 | 2 | 0.0500 | 0.9659 |
| Homogeneity | 3.4141 | 3 | 0.0500 | 0.3321 |

*One-Tail Tests for Trend*

| Type | Probability[a] |
|---|---|
| Uncorrected | 0.0337 |
| Continuity corrected | 0.0426 |
| Exact | 0.0423 |

*Multiple Pairwise Group Comparisons by Fisher–Irwin Exact Test*

| Groups Compared | $\alpha$ | One-Tailed Probability |
|---|---|---|
| 1 vs. 2 | 0.0500 | 0.33887 |
| 2 vs. 3 | 0.0500 | 0.13433 |
| 1 vs. 4 | 0.0500 | 0.07975 |

[a]Direction = +.

is given by tumor incidence rates that increase with increasing dose. A test designed specifically to detect such dose-related trends is Tarone's (1975) trend test.

Letting $\mathbf{d}=(O, d_1, d_2 \ldots d_k)^{\mathrm{T}}$ be the vetor of dose levels in *all* $K + 1$ groups and letting

$$(\mathbf{O}-\mathbf{E}) = (O_o - E_0, , , O_k - E_k)^{\mathrm{T}} \quad \text{and} \quad \mathbf{V} = \begin{pmatrix} V_{00} \ldots V_{0K} \\ \cdot \quad \cdot \\ \cdot \qquad \cdot \\ V_{K0} \ldots V_{KK} \end{pmatrix}$$

contain elements as described in the previous section but for *all* $K$+1 *groups*, the trend statistic is given by

$$X_{\frac{2}{T}} = \frac{\left[\mathbf{d}^{\mathrm{T}}(\mathbf{O}-\mathbf{E})\right]^2}{\mathbf{d}^{\mathrm{T}}\mathbf{V}\mathbf{d}}$$

The statistic $X\frac{2}{T}$ will be large when there is evidence of a dose-related increase or decrease in the tumor incidence rates and small when there is little difference in the tumor incidence between groups or when group differences are not dose related. Under the null hypothesis of no differences between groups, $X\frac{2}{T}$ has approximately a chi-squared distribution with one degree of freedom.

Tarone's trend test is most powerful at detecting dose-related trends when tumor onset hazard functions are proportional to each other. For more power against other dose-related group differences,weighted versions of the statistic are also available; see Breslow (1984) or Crowley and Breslow (1984) for details.

These tests are based on the generalized logistic function (Cox, 1972). Specifically one can use the Cochran–Armitage test (or its parallel, Mantel–Haenszel version) for monotonic trend as the heterogeneity test.

## 14.3 Life Table and Survival Analysis

These methods are essential when there is any significant degree of mortality in a bioassay. They seek to adjust for the differences in periods of risk individual animals undergo. Life table techniques can be used for those data where there are observable or palpable tumors. Specifically, one should use Kaplan–Meier product limit estimates from censored data graphically, Cox–Tarone binary regression (log rank test), and the Gehan–Breslow modification of Kruskal–Wallis test (Thomas et al., 1977) on censored data.

The Kaplan–Meier estimates produce a step function for each group and are plotted over the lifetime of the animals. Planned, accidentally killed, and lost animals are censored. Moribund deaths are considered to be treatment related. A graphical representation of Kaplan–Meier estimates provide excellent interpretation of survival-adjusted data except in the cases where the curves cross between two or more groups. When the curves cross and change direction, no meaningful interpretation of the data can be made by any statistical method because the proportional-odds characteristic is totally lost over time. This would be a rare case where treatment initially produces more tumor or death and then, due to repair or other mechanisms, becomes beneficial.

In Cox–Tarone binary regression (Tarone, 1975; Thomas et al., 1977) censored survival and tumor incidence data are expressed in a logistic model in dose over time. The log rank test (Peto, 1974), tests based on the Weibull distribution, and the Mantel–Haenszel (1952) test are very similar to this test when there are no covariates or stratifying variables in the design. The logistic regression–based Cox–Tarone test is preferable because one can easily incorporate covariates and stratifying variables which one cannot in the IARC methods.

The Gehan–Breslow modification of the Kruskal–Wallis test is a nonparametric test on censored observations. It assigns more weight to early incidences compared to the Cox–Tarone test.

For survival-adjusted tests on proportions, as mentioned earlier, in the case of survival-adjusted analyses, instead of having a single $2k$ table, one has a

series of such $2k$ tables across the entire lifetime of the study. The numbers at risk for such analyses will depend on the type of tumor one is dealing with. These are shown below:

1. Palpable or Lethal Tumors Number at risk at time $t$ = number of animals surviving at the end of time $t$–1
2. Incidental Tumors Number at risk at time $t$ = number of animals that either died or were sacrificed whose particular tissue was examined histopathologically

The methods of analyzing the incidences once the appropriate numbers at risk are assigned for these tumors are rather similar, either binary regression based or by pooling evidence from individual tables (Gart et al., 1986).

## 14.4 Peto Analysis

The Peto method of analysis of bioassay tumor data is based on careful classification of tumors into five different categories as defined by IARC:

1. Definitely incidental
2. Probably incidental

*Comment:* Combine 1 and 2:

3. Probably lethal
4. Definitely lethal

*Comment:* These categories may be combined into one (otherwise a careful cause-of-death determination is required):

5. Mortality independent (such as mammary, skin, and other observable or superficial tumors)

### *Interval Selection for Occult (Internal Organ) Tumors*

1. *FDA* 0–50, 51–80, 81–104 weeks, interim sacrifice, terminal sacrifice
2. *NTP* 0–52, 53–78, 79–92, 93–104 weeks, interim sacrifice, terminal sacrifice
3. *IARC* ad hoc selection method (Peto et al., 1980)

*Comment*: Any of the above may be used. Problems with IARC selection method include two sexes and two or more strains that will have different intervals for the same compound. Different interval selection methods will produce different statistical significance levels. This may produce bias and requires an isotonic tumor prevalence for ready analysis.

***Logistic Regression Method for Occult (Internal Organ) Tumors (Dinse, 1985)*** Tumor prevalence is modeled as a logistic function of dose and polynomial in age.

*Comment:* The logistic tumor prevalence method is unbiased. It requires maximum-likelihood estimation and allows for covariates and stratifying variables. It may be time consuming and have convergence problems with sparse tables (low tumor incidences) and clustering of tumors.

## 14.5 Methods To Be Avoided

The following methods and practices should be avoided in evaluation of carcinogenicity:

1. Use of only the animals surviving after one year in the study
2. Use of a two-strata approach; separate analyses for animals killed during the first year of the study and the ones thereafter
3. Exclusion of all animals in the study that died on test and analysis of only the animals that are sacrificed at the end of the study
4. Exclusion of interim sacrifice animals from statistical analyses
5. Evaluation of number of tumors of all sites as opposed to number of animals with tumors for specific sites of specific organs

Another issue is subjectivity in slide reading by most pathologists who do not want to read them in a coded fashion whereby they will not know the dose group an animal is coming from. This is not under the statisticians' control but they should be aware of it in any case.

Often a chemical being tested is both toxic as well as potentially carcinogenic. When competing toxicity causes extreme differences in mortality or there is a clustering effect in tumor prevalence in a very short interval of time, none of the adjusted methods works. One then must use biological intuition to evaluate the tumor data.

Use of historical control incidence data for statistical evaluation is controversial. There are too many sources of variation in these data. For example, different pathologists use different criteria for categorizing tumors (in fact, the same pathologist may change his or her opinion over time), there is laboratory-to-laboratory variation, there may be genetic drift over time; the location of suppliers may make a difference, and, finally, these data are not part of randomized concurrent control. Regulatory agencies and pathologists generally use these data for qualitative evaluation. My personal view is that is where they belong.

## 14.6 Use of Historical Controls

When the study is over, the data analyzed, and the $p$ values corrected, as appropriate, one may find that one or more tumor types increased in drug-

treated groups relative to concurrent controls. Although the FDA and other regulatory agencies play down the importance of historical control data, it is common practice in the pharmaceutical industry to use historical data in the interpretation of tumor findings. The first and most appropriate comparison of a treated group is with concurrent control group(s), but it is of interest to see how tumor incidences in the treated groups compare with the historical incidence, and such a comparison is an accepted practice in toxicology and biostatistics (Gart et al., 1979; Hajian, 1983; Haseman et al., 1984b). A treated group may have a tumor incidence significantly higher than that of the concurrent control groups(s) but comparable to or lower than the historical incidence. Occasionally, a small number of tumors may be found in a treated group and the incidence may be significant because of the absence of this tumor in the concurrent controls. Review of appropriate historical control data may reveal that the low tumor incidence in the treated group is within the "expected" range for this tumor.

The role of historical control data in interpreting carcinogenicity findings depends on the "quality" of the historical data. Ideally, the data should be derived from animals of the same age, sex, strain, and supplier and housed in the same facility, and the pathology examinations should have been performed by the same pathologist or using the same pathological criteria for diagnosis. Since genetic drift occurs even in animals of a given strain and supplier, recent data are more useful than older data. The value of historical control data is directly proportional to the extent to which these conditions are fulfilled.

Although methods are available for including historical control data in the formal statistical analysis (Tarone, 1982; Dempster et al., 1983), this is usually not done and for good reason. The heterogeneity of historical data requires that they be used qualitatively and selectively to aid in the final interpretation of the data after completion of the formal statistical analysis. Table 4 presents a summary of background tumor incidences for the most commonly employed rodent strains.

## 14.7 Relevance to Humans

After statistical analyses have been performed and historical data consulted, the final interpretation may be that a drug appears to cause tumors at one or more tissue sites in the mouse or the rat. But what does this mean for the species to which the drug will be administered, namely, the human? Extrapolation of rodent carcinogenicity data to humans remains one of the greatest challenges of modern toxicology. There is no simple formula, and each case must be evaluated on its own merits. Very generally speaking, the FDA and other major regulatory agencies consider compounds that are tumorigenic in one or more animal species to be "suspect" tumorigens in humans. The actual impact of this conclusion on the approval of a drug depends on the target population and the indication. For example, even a suspicion of carcinogenic

**TABLE 4 Comparative Percent Incidence of Pertinent Neoplasia in Different Strains of Rats and Mice (104 weeks old)**

| | F344 Rats | | SD Rats | | Wistar Rats | | $B6C3F_1$ Mice | | CD-1 Mice | |
|---|---|---|---|---|---|---|---|---|---|---|
| Types of Neoplasia | Males | Females | Males | Females | Males | Females | Males | Females | Males | Females |
| Hepatocellular adenoma | 4 | <1 | 5 | <1 | 1 | 2 | 29 | 30 | 26 | 5 |
| Hepatocellular carcinoma | 2 | 0 | 2 | 0 | <1 | <1 | 26 | 16 | 10 | 1 |
| Pancreas islet adenoma | 12 | 2 | 8 | 9 | 4 | 2 | 2 | 0 | <1 | <1 |
| Pancreas islet carcinoma | 3 | 0 | <1 | 5 | <1 | <1 | 0 | 0 | 0 | 0 |
| Pancreas acinar adenoma | 6 | 0 | 1 | 0 | 13 | <1 | 2 | 0 | <1 | 0 |
| Pheochromocytoma | 21 | 4 | 23 | 5 | 10 | 2 | 0 | 2 | <1 | <1 |
| Adrenocortical adenoma | 0 | 2 | 3 | 0 | 8 | 9 | <1 | 0 | 1 | <1 |
| Pituitary adenoma | 49 | 42 | 62 | 85 | 34 | 55 | 2 | 8 | 0 | 5 |
| Thyroid C-cell adenoma | 17 | 8 | 7 | 6 | 6 | 8 | 0 | 0 | 0 | 0 |
| Thyroid follicular adenoma | 0 | 0 | 4 | 2 | 2 | 1 | 2 | 6 | 1 | <1 |
| Mammary gland fibroadenoma | 4 | 57 | 2 | 54 | 3 | 36 | 0 | 0 | <1 | 1 |
| Mammary gland carcinoma | 0 | 4 | <1 | 26 | 1 | 13 | 0 | 0 | 0 | 6 |
| Skin fibroma | 10 | 2 | 2 | <1 | 5 | 1 | 1 | 2 | <1 | <1 |
| Skin papilloma | 6 | 0 | 2 | 0 | 2 | <1 | 0 | 0 | <1 | 0 |
| Pulmonary adenoma | 4 | 4 | <1 | <1 | <1 | 0 | 22 | 6 | 15 | 15 |
| Preputial gland neoplasia | 10 | NA | >1 | NA | <1 | NA | <1 | NA | <1 | NA |
| Leydig cell neoplasia | 89 | NA | 7 | NA | 11 | NA | 0 | NA | 1 | NA |
| Clitoral gland neoplasia | NA | 14 | NA | <1 | NA | <1 | NA | <1 | NA | 0 |
| Uterine polyps | NA | 14 | NA | 6 | NA | 16 | NA | 1 | NA | <1 |
| Ovarian neoplasia | NA | 6 | NA | 1 | NA | 8 | NA | 6 | NA | 1 |
| Mononuclear cell leukemia | 62 | 42 | 0 | 0 | <1 | <1 | 0 | 0 | 2 | 2 |
| Lymphoma | 0 | 0 | 2 | 1 | 3 | 5 | 14 | 24 | 8 | 22 |
| Forestomach papilloma | 0 | 2 | <1 | <1 | 0 | <1 | 4 | 2 | <1 | <1 |
| Scrotal mesothelioma | 5 | NA | 1 | NA | 2 | NA | 0 | NA | 0 | NA |

*Note*: F344, Fischer 244; SD, Sprague Dawley; $B6C3F_1$, $(C57BL/6N+C3H/HeN)F_1$; CD-1, 1CRCr: CD-1 mice; NA, not applicable; the average number used by species/strain/gender was in excess of 750 animals.

activity may be fatal for a potential contraceptive drug intended for use in a very large population of healthy people. By contrast, clear evidence of carcinogenic activity may be overlooked in a drug being considered for use in a restricted population with a life-threatening disease.

Regardless of the target population and indication, the FDA and other agencies have, in recent years, attempted to consider the mechanism of tumor induction in rodents and its relevance for humans. If a drug is known to cause tumors in a rodent via a mechanism that does not exist in humans, the importance of the tumor findings may be markedly reduced. For example, drugs that cause tumors by a secondary hormonal mechanism shown to be inapplicable to humans may be given special consideration. It is the sponsor's responsibility to provide pertinent data on the mechanism of tumor induction and its relevance, or irrelevance, for humans. If the sponsor can show that an apparently drug-related tumor is species specific, the importance of the tumor in the overall evaluation of the drug will be greatly minimized. Table 5 presents a list of neoplastic /tumorigenic responses seen in rodents which have limited relevance to human safety. Part of the consideration must also be recognition of the main characteristics of nongenotoxic carcinogens. These are recognized to be dose-dependent responses with operative thresholds. The major characteristics are (Spindler et al., 2000):

- Specificity (of species, sex, and organ).
- A threshold is operative and must be exceeded for cell proliferation and tumor development to occur.
- There is a stepwise dose–response curve/relationship between exposure, cell proliferation, and tumor development.
- The response is reversible with a cessation of dosing unless a point of no return has been passed.

## 15 CONCLUSIONS

The design, conduct, and interpretation of carcinogenicity studies is one of the major challenges for the pharmaceutical toxicologist, pathologist, biostatistician, and regulator. This is a rapidly changing field generating more questions than answers. The largest question continues to be the extrapolation of rodent data to humans, especially when data on mechanisms of tumor induction are unavailable or controversial. Much has been written on the difficulties inherent in extrapolating results from rodents treated with MTDs of a compound to humans who will be exposed to much lower doses and often for shorter periods. A discussion of these and other aspects of carcinogenic risk assessment is beyond the scope of this chapter.

Regulatory agencies are very aware of these challenges and deserve credit for attempting to respond to changes in the state of knowledge while still

**TABLE 5 Examples of Neoplastic Effects in Rodents with Limited Significance for Human Safety**

| Neoplastic Effect | Pathogenesis (Agents) |
|---|---|
| Renal tubular neoplasia in male rats | $\alpha_{2\mu}$-Globulin nephropathy/hydrocarbons (*d*-limonene, *p*-dichlorobenzene) |
| Hepatocellular neoplasia in rats and mice | Peroxisome proliferation (clofibrate, phthalate esters, phenoxy agents)<br>Phenobarbital-like promotion |
| Urinary bladder neoplasia in rats | Crystalluria, carbonic anhydrase inhibition, urine pH extremes, melamine, saccharine, carbonic anhydrase inhibitors, dietary phosphates |
| Hepatocellular neoplasia in mice | Enzymatic-metabolic activation (in part unknown)/phenobarbital-like promotion |
| Thyroid follicular cell neoplasia in rats | Hepatic enzyme induction, thyroid enzyme inhibition/axazepam, amobarbital, sulfonamides, thioureas |
| Gastric neuroendocrine cell neoplasia mainly in rats | Gastric secretory suppression, gastric atrophy induction (climetidine, omeprazole, butachlor) |
| Adenohypophysis neoplasia in rats | Feedback interference/neuroleptics (dopamine inhibitors) |
| Mammary gland neoplasia in female rats | Feedback interference/neuroleptics, antiemetics, antihypertensives (calcium channel blockers), serotonin agonists, anticholinergics, exogenous estrogens |
| Pancreatic islet cell neoplasia in rats | Feedback interference/neuroleptics |
| Harderian gland neoplasia in mice | Feedback interference/misoprostol ($PGE_1$), nalidixic acid, aniline dyes |
| Adrenal medullary neoplasia in rats | Feedback interference (lactose, sugar alcohols) |
| Forestomach neoplasia in rats and mice | Stimulation of proliferation/butylated hydroxyanisole, phthalate esters, proprionic acid |
| Lymphomas in mice | Immunosuppression/cyclosporin |
| Mononuclear cell leukemia in rats (mainly F344) | Immunosuppression (in part unknown)/furan, iodinated glycerol |
| Splenic sarcomas in rats | Methemoglobinemia (in part unknown)/dapsone |
| Osteomas in mice | Feedback interference/lactose, sugar alcohols, $H_2$ antagonists, carbamazepine, vidarabine, isradipine, dopaminergics, finasteride |
| Leydig cell testicular neoplasia in mice | Feedback interference (proestrogens, finasteride, methoxychlor, cadmium) |
| Endometrial neoplasia in rats | Feedback interference (proestrogens, dopamine agonists) |
| Uterine leiomyoma in mice | Feedback interference ($\beta_1$ antagonists) |
| Mesovarial leiomyoma in rats (occasionally in mice) | Feedback interference ($\beta_2$ agonists) |
| Ovarian tubulostromal neoplasia in mice | Feedback interference (cytotoxic agents, nitrofurantoin) |

discharging their responsibility to protect the public health. For example, the latest version of the Japanese guidelines (Speid et al., 1990) acknowledges that the highest does in a carcinogenicity study may be set at 100 times the clinical dose, instead of requiring that the MTD be achieved. It is also noteworthy that the FDA Center for Drug Evaluation and Research has announced the formation of a carcinogenicity assessment committee representing all drug review divisions. This group will advise all the divisions on issues related to carcinogenicity. Creation of such a group reflects the importance that the agency places on carcinogenicity data in evaluating the safety of new drug candidates.

## REFERENCES

Bellhorn, R. W. (1980). Lighting in the animal environment. *Lab Anim. Sci.* 30:440–450.

Berenblum, I., and Shubik, P. (1947). The role of croton oil applications associated with a single painting of a carcinogen, in tumour induction in the mouse's skin. *Br. J. Cancer* 1:379–383.

Breslow, N. (1984). Comparison of survival curves. In M. E. Buyse, M. J. Staquet, and R. J. Sylvester (Eds.), *Cancer Clinical Trials: Methods and Practice*. Oxford University Press, Oxford, pp. 381–406.

Burns, F. J., Vanderlan, M., Snyder, E., and Albert, R. E. (1976). Induction and progression kinetics of mouse skin papillomas. In T. J. Slaga, A. Sivak, and R. K. Boutwell (Eds)., *Carcinogenesis*, Vol. 2: *Modifiers of Chemical Carcinogenesis*. Raven Press, New York, pp. 91–96.

Cameron, T. P., Hickman, R. L., Kornreich, M. R., and Tarone, R. E. (1985). History, survival, and growth patterns of B6C3F1 mice and F344 rats in the National Cancer Institute Carcinogenesis Testing Program. *Fund. Appl. Toxicol.* 5:526–538.

Chow, S. C., and Liu, J.-P. (1998). *Design and Analysis of Animal Studies in Pharmaceutical Development*. Marcel Dekker, New York.

Chu, K., Cueto, C., and Ward, J. (1981). Factors in the evaluation of 20 NCI carcinogenicity bioassays. *J. Toxicol. Environ. Health* 8:251–280.

Contrera, J. F., Jacobs, A. C., and DeGeorge, J. J. (1997). Carcinogenicity testing and the evaluation of regulatory requirements for pharmaceuticals. *Reg. Tox. Pharmacol.* 25:130–145.

Cox, D. R. (1972). Regression models and life-tables. *J. Roy. Stat. Soc.* 34B:187–220.

Crowley, J., and Breslow, N. (1984). Statistical analysis of survival data. *Ann. Rev. Public Health* 5:385–411.

Dempster, A. P., Selivyn, M. R., and Weeks, B. J. (1983). Combining historical and randomized controls for assessing trends in proportions. *J. Am. Stat. Assoc.* 78:221–227.

DePass, L. R., Weil, C. S., Ballantyne, B., Lewis, S. C., Losco, P. E., Reid, J. B., and Simon, G. S. (1986). Influence of housing conditions for mice on the results of a dermal oncogenicity bioassay. *Fund. Appl. Toxicol.* 7:601–608.

Dinse, G. E. (1985). Estimating tumor prevalence, lethality and mortality. Paper presented at the Symposium on Long-term Animal Carcinogenicity Studies: A Statistical Perspective, Bethesda, MD, March 4–6, 1985.

Dinse, G. E., and Lagakos, S. W. (1983). Regression analysis of tumor prevalence data. *Appl. Stat.* 32:236–248.

Eastin, W. C., Haseman, J. K., Mahler, J. F., and Bucher, J. R. (1998). The National Toxicology Program evaluation of genetically altered mice predictive models for identifying carcinogens. *Toxicol. Pathol.* 26:461–584

Farrar, D. B., and Crump, K. S. (1988). Exact statistical tests for any carcinogenic effect in animal bioassays. *Fund. Appl. Toxicol.* 11:652–663.

Festing, M. W. (1979). Properties of inbreed strains and outbreed stocks, with special reference to toxicity testing. *J Toxicol. Environ. Health* 5:53–68.

Flynn, R. J. (1960). Studies on the aetiology of ringtail in rats. *Proc. Anim. Care Panel* 9:155–160.

Foulds, L. (1969). *Neoplastic Development*, Vol. 1. Academic Press, New York.

Foulds, L. (1975). *Neoplastic Development*, Vol. 2. Academic Press, New York.

Fox, J. G. (1977). Clinical assessment of laboratory rodents on long term bioassay studies. *J. Environ. Pathol. Toxicol.* 1:199–226.

French, J. E., Spalding, J. W., Dunnick, J. K., Tice, R. R., Furedi-Marchacck, M., and Tennant, R. W. (1999). The use of transgenic animals in cancer testing. *Inhal. Toxicol.* 11:541–544.

Gad, S. C. (2006). *Statistics and Experimental Design for Toxicologists*, 4th ed. CRC Press, Boca Raton, FL.

Gart, J. J., Chu, K. C., and Tarone, R. E. (1979). Statistical issues in interpretation of chronic bioassay tests for carcinogenicity. *J. Natl. Cancer Inst.* 62:957–974.

Gart, J. J., Krewski, D., Lee, P. N., Tarone, R. E., and Wahrendorf, J. (1986). The design and analysis of long-term animal experiment. In *Statistical Methods in Cancer Research*, Vol. III. IARC Scientific Publication No. 79. International Agency for Research on Cancer, Lyon.

Gaylor, D. W., and Kodell, R. L. (2001). Dose-response trend tests for tumorogenesis adjusted for differences in survival and body weight across doses. *Toxicol. Sci.* 59:219–225.

Gold, L. S., Sawyer, C. B., Magaw, R., Blackman, G., deVeciana, M, Levenson, R., Hooper, N. K., Havender, W. R., Bernstein, L., Peto, R., Pike, M. C., and Ames, B. N. (1984). A carcinogenic potency data base of the standardized results of animal bioassays. *Environ. Health Perspect.* 58:9–319.

Greenman, D. L., Bryant, P., Kodell, R. L., and Sheldon, W. (1982). Influence of cage shelf level on retinal atrophy in mice. *Lab. Anim. Sci.* 32:353–356.

Greenman, D. L., Kodell, R. L., and Sheldon, W. G. (1984). Association between cage shelf level and spontaneous induced neoplasms in mice. *J. Natl. Cancer Inst.* 73:107–113.

Hajian, G. (1983). Statistical issues in the design and analysis of carcinogenicity bioassays. *Toxicol. Pathol.* 11:83–89.

Haseman, J. K. (1983). A reexamination of false-positive rates for carcinogenicity studies. *Fund. Appl. Toxicol.* 3:334–339.

Haseman, J. K. (1985). Issues in carcinogenicity testing: Dose selection. *Fund. Appl. Toxicol.* 5:66–78.

Haseman, J. K. (1990). Use of statistical decision rules for evaluating laboratory animal carcinogenicity studies. *Fund. Appl. Toxicol.* 14:637–648.

Haseman, J. K., Crawford, D. D., Huff, J. E., Boorman, G. A., and McConnell, E. E. (1984a). Results for 86 two-year carcinogenicity studies conducted by the National Toxicology Program. *J. Toxicol. Environ. Health* 14:621–639.

Haseman, J. K., Huff, J., and Boorman, G. A. (1984b). Use of historical control data in carcinogenicity studies in rodents. *Toxicol. Pathol.* 12:126–135.

Haseman, J. K., Winbush, J. S., and O'Donnell, M. W. (1986). Use of dual control groups to estimate false positive rates in laboratory animal carcinogenicity studies. *Fund. Appl. Toxicol.* 7:573–584.

Holm-Nielsen, P., and Olsen, T. S. (1988). Ultrastructure of renal adenoma. *Ultrastruct. Pathol.* 12:27–39.

International Conference on Harmonisation (ICH) (1996). *S1A: The Need for Long-Term Rodent Carcinogenicity Studies of Pharmaceuticals*. ICH, Geneva.

International Conference on Harmonisation (ICH) (1997a). *S1B: Testing for Carcinogenicity of Pharmaceuticals*. ICH, Geneva.

International Conference on Harmonisation (ICH) (1997b). *S1C(R1): Dose Selection for Carcinogenicity Studies of Pharmaceuticals*. ICH, Geneva.

International Agency for Research on Cancer (IARC) (1987). IARC monographs on the evaluation of carcinogenic risks to humans—preamble. IARC Internal Technical Report 87/001. IARC, Lyon. ICH, Geneva.

Japanese Ministry of Health and Welfare (MHW) (1989). *Revised Guidelines for Toxicity Studies Required for Application for Approval of Manufacturing/Importing Drugs*. MHW, Tokyo, pp. 37–48.

Kitchin, K. T. (1999). *Carcinogenicity*. Marcel Dekker, New York.

Littlefield, N. A., Farmer, J. H., Gaylor, D. W., and Sheldon, W. G. (1979). Effects of dose and time in a long-term, low-dose carcinogenic study. In J. A. Staffa and M. A. Mehlman (Eds.), *Innovations in Cancer Risk Assessment (ED01Study)* Pathotox Publishers, Chicago, IL, pp. 17–34.

Mantel, N. (1980). Assessing laboratory evidence for neoplastic activity. *Biometrics* 36:381–399.

Mantel, N., and Haenszel, W. (1952). Statistical aspects of the analysis of data from the retrospective studies of disease. *J. Natl. Cancer Inst.* 22:719–748.

McAnulty, P. A. (2000). Transgenic mouse models in carcinogenicity testing. *Eur. Pharm. Contractor* Nov.

McGregor, D. (2000). Carcinogenicity and genotoxic carcinogens. In B. Ballantyne, T. Marrs, and T. Syversen (Eds.), *General and Applied Toxicology*, 2nd eds. MacMillan, London.

McKnight, B., and Crowley, J. (1984). Tests for differences in tumor incidence based on animal carcinogenesis experiments. *J. Am. Stat. Assoc.* 79:639–648.

Miller, E. C., and Miller, J. A. (1981). Mechanisms of chemical carcinogenesis. *Cancer* 47:1055–1064.

National Institutes of Health (NIH) (2000). *Ninth Report on Carcinogens*. U.S. Department of Health and Human Services, Washington, DC.

Peto, R. (1974). Guidelines on the analysis of tumor rates and death rates in experimental animals. *Br. J. Cancer* 29:101–105.

Peto, R., Pike, M. C., Day, N. E., Gray, R. G., Lee, P. N., Parish, S., Peto, J., Richards, S., and Wahrendorf, J. (1980). Guidelines for simple, sensitive significance tests for carcinogenic effects in long-term animal experiments. In *IARC Monographs on the Evaluation of the Carcinogenic Risk of Chemicals to Humans*. International Agency for Research on Cancer, Lyon.

Pharmaceutical Manufacturers Association (PMA) (1988). Results of a questionnaire involving the design of and experience with carcinogenicity studies. Unpublished document. PMA, Washington, DC.

Portier, C. J., and Bailer, A. J. (1989). Testing for increased carcinogenicity using a survival-adjusted quantal response test. *Fund. Appl. Toxicol.* 12:731–737.

Powell, C. J., and Berry, C. L. (2000). Non-genotoxic or epigenetic carcinogenesis. In B. Ballantyne, T. Marrs, and T. Syversen (Eds.), *General and Applied Toxicology*, 2nd eds. MacMillan, London.

Purchase, I. F. H. (1980). Interspecies comparisons of carcinogenicity. *Br. J. Cancer* 41:454–468.

Rao, G. N., Birnbaum, L. S., Collins, J. J., Tennant, R. W., and Skow, L. C. (1988). Mouse strains for chemical carcinogenicity studies: Overview of a workshop. *Fund. Appl. Toxicol.* 10:385–394.

Rao, G. N., Haseman, J. K., and Edmondson, J. (1989). Influence of viral infections on body weight, survival, and tumor prevalence in Fisher 344 rats on two-year studies. *Lab. Anim. Sci.* 39:389–393.

Rao, G. N., and Huff, J. (1990). Refinement of long-term toxicity and carcinogenesis studies. *Fund. Appl. Toxicol.* 15:33–43.

Schweder, T., and Spjotvoll, E. (1982). Plots of *p*-values to evaluate many tests simultaneously. *Biometrika* 69:493–502.

Sher, S. P., Jensen, R. D., and Bokelman, D. L. (1982). Spontaneous tumors in control F344 and Charles River CD rats and Charles River CD-1 and B6C3F1 mice. *Toxicol. Lett.* 11:103–110.

Silverberg, S. G. (1984). The autopsy and cancer. *Arch. Pathol. Lab. Med.* 108:476–478.

Speid, L. H., Lumley, C. E., and Walker, S. R. (1990). Harmonization of guidelines for toxicity testing of pharmaceuticals by 1992. *Reg. Toxicol. Pharmacol.* 12:179–211.

Spindler, P., Loan, J., Ceuppens, P., Harling, R., Eittlin, R., and Lima, B. S. (2000). Carcinogenicity testing of pharmaceuticals in the European Union: A workshop report. *Drug Inform. J.* 34:821–828.

Tarone, R. E. (1975). Tests for trend in life table analysis. *Biometrika* 62:679–682.

Tarone, R. E. (1982). The use of historical control information in testing for a trend in proportions. *Biometrics* 38:215–220.

Tennant, R. W., French, J. E., and Spalding, J. W. (1995). Identification of chemical carcinogens and assessing potential risks in short-term bioassays using transgenic mouse models. *Environ. Health. Perspect.* 103:942–950.

Tennant, R. W., Stasiewicz, S., Mennear, J., French, J. E., and Spalding, J. W. (1999). Genetically altered mouse models for identifying carcinogens. In D. B. McGregor, J. M. Rice, and S. Venith, Eds. *The Use of Short and Medium-Term Tests for Car-*

*cinogens and Data on Genetic Effects in Carcinogenic Hazard Evaluation*. IARC Science Publications No. 146. Lyon, France, pp. 23–148.

Thomas, D. G., Breslow, N., and Gart, J. J. (1977). Trend and homogeneity analyses of proportions and life table data. *Comput. Biomed. Res.* 10:10:373–381.

U.S. Food and Drug Administration (FDA) (2000). Toxicological principles for the safety assessment of direct food additives and color additives used in food. In *Redbook II* (draft). FDA, Washington, DC, pp. 111–115.

U. S. Food and Drug Administration (FDA) (2001). Guidance for industry: statistical aspects of the design, analysis and interpretation of chronic rodent carcinogenicity studies of pharmaceuticals. U.S. Department of Health, Education and Welfare, Washington, DC.

Weaver, R. J., and Brunden, M. N. (1998). The design of long-term carcinogenicity studies. In S. Chow and J. Lide (Eds.), *Design and Analysis of Animal Studies in Pharmaceutical Development*. Marcel Dekker, New York.

Weisbrode, S. E., and Weiss, H. S. (1981). Effect of temperature on benzo(*a*)pyrene induced hyperplastic and neoplastic skin lesions in mice. *J. Natl. Cancer Inst.* 66:978–981.

Williams, G. M., and Iatropoulos, M. J. (2001). Principles of testing for carcinogenic activity. In A. W. Hayes (Ed.), *Principles and Methods of Toxicology*, 5th ed. Taylor and Francis, Philadelphia, PA.

Wiskemann, A., Sturm, E., and Klehr, N. W. (1986). Fluorescent lighting enhances chemically induced papilloma formation and increases susceptibility to tumor challenge in mice. *J. Cancer Res. Clin. Oncol.* 112:141–143.

# *Histopathology in Nonclinical Pharmaceutical Safety Assessment*

## 1 INTRODUCTION

Toxicological pathology is the study of the molecular, cellular, tissue, and/or organ responses of a living organism when exposed to injurious chemical or physical agents. These responses represent a spectrum of cellular changes ranging from cell death to malignant transformations, tissue and organ responses (including regeneration inflammation), and organization and overall response as identified by clinical changes and alterations in body fluids (Arnold et al., 1990; MHLW, 1990; EEC, 1992; EMEA, 2000; Rousseaux et al., 2002; Prince and Wilson, 2003; ICH, 2005, 2006). It starts from recognition of the fact that the cell constitutes the basic unit of life. Accordingly, morphological changes in organs and tissues arise as a result of injury beginning with the responses of underlying cells to the toxic insult. A proper evaluation and understanding of related pathology must start at the cellular level. Some cellular components whose alterations have been reported to be critically associated with cell injury include the plasma membrane, site of osmotic, electrolyte,

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

and water regulation as well as signal transduction; the mitochondrion (site of energy storage and aerobic respiration); the endoplasmic reticulum (site of much protein synthesis); and the nucleus, which contains the genetic information and where transcription of the genetic code takes place (Hamm, 1974; Wallig, 2002).

Biochemical changes such as enzyme induction and gene expression occur at the earlier stages of the exposure–disease continuum. The degrees of cellular injury in different target tissue depend on the metabolic rate. Cells with high metabolic rates such as neurons, myocardial cells, and renal proximal convoluted tubule epithelial cells frequently suffer from injury more quickly than low-metabolizing ones. These high-metabolism cells depend on a continuous flow of oxygen to conduct the aerobic metabolism necessary to provide required energy in the form of adenosine triphosphate (ATP) for the maintenance of membrane polarity and membrane integrity (neurons), for the continual muscular contraction/relaxation and calcium transport (myocardium), and for the transport of fluids, electrolytes, and metabolites [renal proximal tubule cells (PTCs)]. Hence, any depletion in oxygen supply is likely to have a significant impact on their survival.

In contrast, cells with low metabolic activity such as fibroblasts and adenocytes are less affected by the low supply of oxygen, and they have a prominent role in regeneration and scarring. Homeostasis is one of the most remarkable and most typical properties of highly complex biological systems. It defines the ultimate environment under which cells maintain the physiochemical conditions (intracellular pH, cytosolic osmolarity, ion gradients) necessary to perform their biological functions. In biological systems, homeostasis is maintained by means of a multiplicity of dynamic equilibriums rigorously controlled by interdependent regulation mechanisms. Hence, the homeostatic system reacts to changes or disturbances in response to various insults by exerting a series of modifications or adjustments to maintain the internal balances and conditions within tolerable limits.

Pathology [including all the aspects of anatomic (histopathology) clinical chemistry and clinical pathology (hematology)] is generally considered the single most significant portion of data to come out of systemic toxicity studies (particularly the repeat dose, with versions going from 14 days to 2 years in duration) (Roz and Andrews, 2004; Perez and Barthold, 2007; Frame and Mann, 2008; Tehounwan and Centeno, 2008). Anatomic pathology evaluations actually consist of three related sets of data (gross pathology observations, organ weights, and microscopic pathology) that are collected during termination of the study animals. At the end of the study, a number of tissues are collected during termination of all surviving animals (test and control). Organ weights and terminal body weights are recorded at study termination, so that absolute and relative (to body weight) values can be statistically evaluated. Bindhu et al. (2007) have provided a review of practices of how such organ weight information is evaluated and utilized in the overall evaluation of pathology and adverse effects. In general, with the exception of brain weights,

**TABLE 1 Tissues for Histopathology**

| | |
|---|---|
| Adrenals[a] | Mainstream bronchi |
| Body and cervix | Major salivary gland |
| Brain, all three levels[a] | Mesenteric lymph nodes |
| Cervical lymph nodes | Ovaries and tubes |
| Cervical spinal cord | Pancreas |
| Duodenum | Pituitary |
| Esophagogastric junction | Prostate |
| Esophagus | Skeletal muscle from proximal hind limb |
| Eyes with optic nerves | Spleen[a] |
| Femur with marrow | Strenebrae with marrow |
| Heart[a] | Stomach |
| Ileum | Testes with epididymides[a] |
| Kidneys[a] | Thymus and mediastinal contents[a] |
| Large bowel | Thyroid with parathyroid[a] |
| Larynx with thyroid and parathyroid | Trachea |
| Liver[a] | Urinary bladder |
| Lungs[a] | Uterus including horns |

[a]Organs to be weighed.

relative (to body weight) changes are considered more relevant to identifying target organ toxicities.

These "standard list" tissues, along with the organs for which weights are determined, are listed in Table 1. All tissues collected are typically processed for microscopic observation, but only those from the high-dose and control groups are necessarily evaluated microscopically. If a target organ is discovered in the high-dose group, then the effect is "followed" in successively lower dose groups until a NOEL (no observed effect level) is determined.

In theory, all microscopic evaluations should be performed in a blind manner (without the pathologist knowing from which dose group a particular animal came), but this is difficult to do in practice and such an approach frequently actually limits (degrades) the quality of the evaluation. Like all the other portions of data in the study, proper evaluation benefits from having access to all data that address the relevance, severity, timing, and potential mechanisms of a specific toxicity. Blind examination is best applied in peer review or consultations on specific findings after a primary evaluation.

In addition to the "standard" set of tissues specified in Table 1, observations during the course of the study or in other or previous studies may dictate that special examination or tissue preparation procedures such as polarized light or electron microscopy, immunocytochemistry, or quantitative morphometry be undertaken to evaluate the relevance of such findings and help or understand the mechanisms underlying certain observations.

The evaluation of the pathological alterations induced in laboratory animals by new drugs represents the cornerstone of their safety assessment before they can be first tried in patients. This preliminary assessment, which is based largely on conventional histopathological techniques, represents a major con-

**TABLE 2 Principles of Drug Testing before Trials in Humans**

1. Exact composition of drug should be known; if not, method of preparation
2. Acute toxicity studies in animals of different species
3. Chronic toxicity experiments at varying doses in different species for cumulative effects
4. Careful and frequent observations of animals to develop a composite picture of clinical effects
5. Careful pathological examination of tissues with appropriate stains
6. Effects of drugs on excretory or detoxifying organs, especially kidney and liver
7. Rate of absorption and elimination, path and manner of excretion, concentration in blood and tissues at varying times
8. Possible influence of other drugs and foodstuffs
9. Careful examination for any syncrasies or untoward reactions

*Note*: As Defined by Geiling and Cannon (1938).

tribution to the development of new treatments for both human and animal diseases.

Although there have been many changes over the past few decades in the details of study design and conduct, the principles of drug testing prior to trial in humans are the same as those expounded by Geiling and Cannon (1938) after they studied the pathological effects and causes of death of patients treated with toxic elixir of sulphanilamide over 60 years ago (Table 2). The basic paradigm of dosing laboratory animals with various doses of new drug for increasing periods of time accompanied by careful clinical observations, biochemical and hematological monitoring, followed by histopathological examination of the tissue remains essentially unaltered. The pathologist is required to not only evaluate alterations to organs and tissues and any relationship that they might have to drug treatment but also to assess the likely relevance any treatment-related findings might have for patients.

The use of animals to study the pathological effects of chemicals and therapeutic agents has a long history. In the eighteenth century Morgagni reported his attempts to compare pathological changes produced by accidental ingestion by people of chemicals such as arsenic. A thorough and systematic review of pathology induced by toxins in humans and animals was published by Orfila as long ago as 1815 (Orfila, 1815). Although in the modern era drug safety evaluation has been widely practiced in rodent and nonrodent species since before World War II, there have been very few critical comparisons of the effects of drugs in humans and those seen in laboratory animals. Much potentially useful information still awaits data mining in the archives of pharmaceutical companies and government agencies. Nevertheless, the available data suggest that the traditional approach using experimental pharmacology alongside conventional toxicology studies with pathology is usually sufficient to predict important adverse effects and to support the safe conduct of the first clinical studies in humans (Turton and Hoasen, 1998; Greaves et al., 2004). Such a degree of concordance varies between different organs and tissues. Therefore each observed drug-induced pathological finding needs to be assessed on a case-by-case basis for its likely clinical relevance. For some

systems, histopathology remains critical for some organ systems but of lesser importance for others. Traditional animal studies are poor predictors of subjective neurological symptoms, but histopathological examination of the nervous system in laboratory animals treated with cancer drugs does well at identifying potential serious clinical neurotoxic effects. Results from nonclinical studies frequently fail to predict renal and hepatic toxicity (largely because of a "formula" approach to evaluation), but there is generally a good correlation for gastrointestinal effects, and histopathology still seems to represent one of the most sensitive techniques to detect effects on the reproductive system. Though the relevance of such findings can be confounded by general systemic toxicity, the pathologist also needs to be aware that some minor inflammatory alterations in certain organs, such as the liver, may have greater significance for the use of a drug in humans than other types of severe damage such as subendocardial necrosis in the myocardium mediated by exaggerated hemodynamic effects.

Treatment-induced findings in conventional toxicity studies found in different laboratory animal species also seem to possess different degrees of relevance for humans. Although the data are fragmentary, findings in beagle dog studies are not often better predictors of human adverse effects than data from rodents or, surprisingly, from primates (Greaves et al., 2004). Dog gastrointestinal and cardiovascular physiology appear to model particularly well for humans, though the pig is generally better yet.

Another long-standing issue most recently recognized again due to findings with cyclooxygenase 2 (COX-2) inhibitors is the importance of evaluating the adverse effects of some therapies with specific human diseases. COX-2 inhibitors were used for inflammatory disorders because of their perceived lower adverse effect profile on the gastrointestinal tract compared with conventional drugs, but this benefit is outweighed by an increased incidence of cardiovascular disease in some patients. Such effects are difficult, if not impossible, to predict from nonclinical safety studies in "normal" healthy animals. Unfortunately the detection of an increased incidence of a common event such as heart attack or stroke is difficult in patients for it requires careful collection and analysis of data, even though it may have a big impact on public health (Dragen, 2005). Such interactions usually require randomized controlled trials specifically designed to look for such risks. It has to be remembered that aspirin was in use for over 100 years before it became generally acknowledged about 30 years ago to be associated with Reye's syndrome, a devastating toxicity in children (Monto, 1999). While the actual mechanism involved in Reye's syndrome is still unknown, it is often preceded by a viral infection and displays a strong correlation with the subsequent ingestion of aspirin.

Histopathology testing is a terminal procedure, and, therefore, sampling of any single animal is a one-time event (except in the case of a tissue collected by biopsy). Because it is a regulatory requirement that the tissues from a specific minimum number of animals be examined at the stated end of the study, any evaluation of effects in tissues at another time course (most commonly, to investigate recovery from effects at study termination) requires that

additional numbers of animals be incorporated into that study at startup. Such animals are randomly assigned at the beginning of the study and are otherwise treated exactly the same as the rest of their group/cohort animals.

Anatomic pathology evaluation occurs only after the in-life portion of a study is complete and is typically the rate-limiting step in producing a report on the study. At the end of the study animals are euthanized, final blood and urine samples are taken, and tissues are collected with a prespecified set weighed while still wet and evaluated grossly (see Table 1) as to whether they are other than normative in appearance. They are then preserved in appropriate manners (Olsen et al., 2000; Haschek et al., 2002; Greaves, 2000, 2007) and processed so as to optimize evaluation (Gray, 1964). Experimental design may call for an interim necropsy (to allow for evaluation of progression of lesions or observation of indications of adaptive change by animals) and/or a recovery group (usually additional numbers of high-dose and control animals in which treatment is discontinued at the time of the main necropsy). Such recovery animals are maintained without further manipulation or treatment for a period of time after the termination of the main study animals (usually a month), allowing for an assessment of treatment-free regression or progression of conditions seen at the end of the main study. Once gathered, tissues must be processed, mounted, stained, and examined with great care. The steps involved in anatomic pathology represent a significant portion of the time required to complete a study and add from 30 to 50% to the cost of the shorter (14-, 28-, and 90-day) studies.

Nevertheless, the available data suggest that the traditional approach using experimental pharmacology alongside conventional toxicology studies with pathology are usually sufficient to predict important adverse effects and to support the safe conduct of the first clinical studies in humans. (Greaves et al., 2004). Indeed, the dosing rodent and nonrodent species with a new drug up to one month identifies over 90% of adverse effects that will ever be detected in the usual nonclinical safety assessment studies. However, more generally these studies do not detect all adverse drug effects that can occur in clinical practice and there remains significant over- and underprediction of human toxicity. Overall, the true positive concordance rate (sensitivity) is on the order of 70% with perhaps 30% of human adverse effects not predicted by safety pharmacology or conventional toxicity studies (Olsen et al., 2000). Moreover, this concordance varies between different organs and tissues. Therefore each drug-induced pathological finding needs to be assessed on a case-by-case basis for its likely clinical relevance.

## 2 PATHOLOGICAL TECHNIQUES

Over the past few years a number of excellent reviews of standardized techniques for use in the histopathology evaluation of toxicology studies have been produced covering tissue selection, blocking and sectioning procedures, immunocytochemical stains for laboratory animals, and other basic techniques

(Bregman et al., 2003; Mikaelian et al., 2004). In addition, the scientific literature is full of interesting techniques and novel reagents that can be applied to tissue sections. Some of these can be very useful in the analysis of pathological alterations in toxicity studies and some fail to work in routinely fixed material. However, it is important that these techniques are used in a judicious manner with clear aims following careful analysis of conventional hematoxylin and eosin–stained sections. This is particularly true for the application of microarray and bioinformatics technology. While undoubtedly useful in toxicology, these techniques should not be applied in isolation but in combination with other information, particularly pathology.

## 3 ORGAN WEIGHTS

Regulatory guidelines indicate that certain organs should be weighed during the course of the necropsy in repeat-dose toxicity studies (Alder and Zbinden, 1988). The extent to which organs are weighed varies between laboratories, but organ weighing is a useful adjunct to macroscopic assessment. Therefore, the selection of organs for weighing is the primary responsibility of the study pathologist. Weighing helps to focus the histopathological examination on key target organs, such as the liver and kidney, the weights of which are frequently altered upon administration of xenobiotics (Peters and Boyd, 1996).

Heart weight is a guide to potential cardiac alterations and is especially important in the assessment of cardiovascular drugs. Likewise, the lungs are weighed in inhalation studies as this can provide a useful indication of the extent of edoma or accumulation of exudate. Brain weight is employed as a stable reference point in adult animals as it is fairly independent of body weight changes. The weights of endocrine organs are useful guides to alterations in the endocrine status of laboratory animals (Pfeiffer, 1998). However, weighing a small and firmly attached organ such as the thyroid can severely disrupt its quality and orientation in the sections and thus offset any apparent advantage (Sellers et al., 2007).

Testicular weights correlate with testicular toxicity and weights can be compared with in-life measurement of testicular size (Heywood and James, 1978; Creasy, 2003). Weighing the testes is a useful precaution at the early phase of development of a novel drug prior to any assessment of male fertility. By contrast, ovarian weight is highly variable as a consequence of cyclical ovarian development and is therefore a less sensitive indicator of treatment-induced changes in the female reproductive system (Long et al., 1998).

## 4 CLINICAL PATHOLOGY

Clinical pathology is the evaluation of changes (or lack of changes) in the formed blood elements and their characteristics—the most common of these

**TABLE 3 Examples of Basic Tests Applicable to Most Rat, Dog, and Monkey Studies**

| Hematology and Coagulation | Clinical Chemistry | Urinalysis |
|---|---|---|
| Red blood cell (RBC) count | Glucose | Color and clarity |
| Hemoglobin | Urea nitrogen (or urea) | Overnight volume (e.g., 16 h) |
| Hematocrit | Creatinine | Urine specific gravity |
| Mean corpuscular volume | Total protein | Reagent strip test, pH, protein, glucose, ketones, bilirubin, urobilinogen, blood |
| Mean corpuscular hemoglobin | Albumin | Microscopic examination of sediment: cells, casts, crystal, bacteria, sperm |
| Mean corpuscular hemoglobin concentration | Globulin (calculated) | |
| RBC morphology | Albumin/Globulin ratio (calculated) | |
| White blood cell (WBC) count | Cholesterol | |
| WBC differential count | Total bilirubin | |
| Platelet count | Alanine aminotransferase | |
| Blood and bone marrow smears | Aspartate aminotransferase | |
| Prothrombin time (PT) | Alkaline phosphatase | |
| Activated partial thromboplastin time (APTT) | γ-Glutamyltransferase | |
| | Creatine kinase | |
| | Calcium | |
| | Inorganic phosphorus | |
| | Sodium | |
| | Potassium | |
| | Chloride | |

parameters are listed in Table 3. These parameters reflect the homeostasis and function of both the hematopoietic system and associated metabolic systems. Beutler et al. (1995) provides an extensive and detailed overview of these systems, though primarily from the perspective of the human system. These measures have the advantage that samples can be taken (and therefore evaluations made) at multiple points over the course of drug administration to drugs and at points subsequent to the discontinuation of such administration (that is, during "recovery").

At the same time, there is the disadvantage that these are frequently indirect measures of what is happening at specific target sites (primarily the bone marrow). The actual target organ effects can generally only be evaluated after termination of the test animals.

Actual evaluation of meaning and relevance and mechanism of cause of observed changes requires, of course, consideration of not individual parameters in isolation but rather of the entire set of measures and the relationships/correlations of these changes. This evaluation is discussed in the earlier chapter on repeat-dose toxicity studies.

## 5 CLINICAL CHEMISTRY

One of the portions of the information employed in a pathological evaluation which is not terminally collected comes from samples of blood and urine.

These can be analyzed for the presence and activity of enzyme and endogenous physiological components (such as electrolytes), including those listed in Table 3. Which are collected and how they are analyzed largely simply follow what is done in human beings during clinical evaluation. Some adaptations have been made, and indeed the measurement methods (and their validation) and interpretation are essentially modified for the specific species in question.

The interpretation of these parameters is primarily addressed in Chapter Repeat-Dose Toxicity Studies on repeat-dose studies, though there are some excellent references on the field (Loeb and Quimby, 1999; Lewandrowski, 2003; Burtis et al., 2005). The last of these best addresses modern biomasses for organ damage. The issue of sampling in nonclinical safety studies is a multifaceted one. First, each species presents limitations on how much (and how often) sample can be drawn. Mice are the most limiting and large nonrodent species (dogs, primates, and pigs) the least. The proximity of sample collection, the time of drug to dosing, or organ damage means that there is a strong emphasis on a need for frequent sample collection. One must not lose sight of the fact that we only see where we look, that is, only where we sacrifice.

## 6 TARGET ORGAN TOXICITY BIOMARKERS

As this is written, the U.S. Food and Drug Administration (FDA) and European Medicines Agency (EMEA) are qualifying organ toxicities. These serve to significantly improve the performance of nonclinical safety studies in identifying potential drug-related toxicities.

The prototype set is for nephrotoxicity, where the two agencies have qualified a set of seven biomarkers—Kim-1, albumin, total protein, cystatin C, $B_2$-microglobulin, urinary clusterin, and urinary trefoil factor. Two of these (albumin and total protein) have been part of the clinical chemistry parameter set for decades. As with traditional clinical chemistry and clinical pathology measures, these biomarkers should be considered as sets, not in isolation.

## REFERENCES

Alder, S. A., and Zbinden, G. (1988). *National and International Drug Safety Guidelines*. M.T.C. Verlag Zollion, Switgerlad.

Arnold, D. L., Dice, H. C., and Krewshi, D. R. (1990). *Handbook of In-Vivo Toxicity Testing*. Academic, New York.

Beutler, E., Lichtman, M. A., Coller, B. S., and Kipps T. J. (1995). *Williams Hematology*, 5th ed. McGraw-Hill, New York.

Bindhu, M., Yano, B., Sellers, R. S., Perry, R., Morton, D., Roome, N., Johnson, J. K., and Schafer, K. (2007). Evaluation of organ weights for rodent and non-rodent toxicity studies: A review of regulatory guidelines and a survey of current practices. *Toxicol. Pathal.* 35:742–750.

Bregman, C. L., Adler, R. R., Morton, D. G., et al. (2003). Recommended tissue list for histopathologic examination in repeat dose toxicity and carcinogenicity studies: A proposal of the society of toxicologic pathology (STP). *Toxicol. Pathol.* 31:252–253.

Burtis, C. A., Ashwood, E. R., and Bruns, D. E. (2005). *Tietz Textbook of Clinical Chemistry and Molecular Diagnosis*, 4th ed. W. B. Saunders, Philadelphia.

Creasy, D. M. (2003). Evaluation of testicular toxicology: a synopsis and discussion of the recommendations proposed by the society of toxicologic pathology. *Birth Defects, Research Part B—Developmental and Reproductive Toxicology* 68: 408–415.

Dragen, J. M. (2005). COX-2 inhibitors—A lesson in unexpected problems. *N. Engl. J. Med.* 352:1131–1132.

EMEA, European Agency for the Evaluation of Medicinal Products Committee for Proprietary Medicinal Products. (2000). Note for guidance on repeated dose toxicity. Available: http://www.emea.eu.int/pdfs/human/swp/104229en.pdf.

European Economic Community (EEC). (1992). Toxicology and chemical substances guidelines. European Chemicals Bureau. Available: http://ecb.jrc.it/testng-methods.

Frame, S. R., and Mann, P. C. (2008). Principles of pathology for toxicology studies. In A. W. Hayes, (Ed.), *Principles and Methods of Toxicology*, 5th ed. CRC Press, Boca Raton, FL, pp. 591–609.

Geiling, E. M. K., and Cannon, P. R. (1938). Pathological Effects of elixir and sulphanilamide (diethylene, glycol) poisoning. *JAMA* 111:919–926.

Gray, P. (1964). *Handbook of Basic Microtechnique*. McGraw-Hill, New York.

Greaves, P. (2000). Pathological techniques in toxicology. In B. Ballantyne, T. Marrs, and T. Synersen (Eds.), *General and Applied Toxicology*, 2nd ed. March Millian, London, pp. 333–353.

Greaves, P. (2007). *Histopathology of Preclinical Toxicity Studies*, 3rd ed. Academics, San Diego.

Greaves, P., Williams, A., and Ever, M. (2004). First dose of potential medicines to humans; how animals help. *Nature Rev. Drug Discov.* 3:226–236.

Hamm, A.W. (1974). *Histology*, 7th ed. J. B., Lippincott, Philadelphia.

Haschek, W. M., Rousseaux, C. G., and Wallig, M. A. (2002). *Handbook of Toxicology Pathology*, 2nd ed. Academic, San Diego.

Heywood, R., and James, R. W. (1978). Assessment of testicular toxicity in lab animals. *Environmental Health Perspectives* 24:73–80.

International Conference of Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH). (2005). Topic S5. Dectection of toxicity to reproduction for medicinal products and toxicity to male fertility. Available: http://www.ich.org/cache/compo/276-254-1.html.

International Conference of Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH). (2006). Topic S8. Immunotoxicity studies for human pharmaceuticals, note for guidance on immunotoxicity studies for human pharmaceuticals. Available: https://www.icb.org/cache/compo/276-254-1.html.

Lewandrowski, K. (2003). *Clinical Chemistry*. Lippincott, Williams and Wilkins, Philadelphia, PA.

Loeb, W. F., and Quimby, F. W. (1999). *The Clinical Chemistry of Laboratory Animals*, 2nd ed. Taylor & Francis, Philadelphia, PA.

Long, G. G., et al. (1998). Proliferative lesions of ovarian granulosa cells and reversible hormonal changes induced in rats by a selective estrogen receptor modulator. *Dev. Reproduct. Toxicol.* 63:380–392.

Mikaelian, I., Nanney, L. B., Parman, K. S., et al. (2004). Antibodies that label paraffin embedded mouse tissues: A collaborative endeavor. *Toxicol. Pathol.* 32:181–191.

Ministry of Health, Labour and Welfare (MHLW), Pharmaceutical Affairs Bureau, New Drugs Division (1990). *Guidelines for Toxicity Studies of Drugs Manual*. MHLW, Tokyo, Japan.

Monto, A. S. (1999). The disappearance of Reyes syndrome—A public health triumph. *N. Engl. J. Med.* 340:1423–1424.

Olsen, H., Betton, G., Robinson, K., et al. (2000). Concordance of the toxicity of pharmaceuticals of humans and animals. *Reg. Toxical. Pharmacol.* 32:56–67.

Orfila, M. J. B. (1815). Traite des poisons Tires des regens mineral, vegetal et animal, on toxicologic generale, connidlhee sons les repports de la physiologic, de la pathologic et de la medicine legale Crochard, Paris.36

Organization for Economic Cooperation and Development Enviroment Directorate. (1992). Guideline for the testing of chemicals. Available: http://www.oecd.og/department/0,2688,en_2649_34377_1_1_1_1_1,00.html.

Perez, D. H., and Barthold, S. W. (2007). *Pathology of Laboratory Rodents and Rabbits*, 3rd ed. Blackwell Publishing, Ames, IA.

Peters, J. M., and Boyd, E. M. (1996). Organ weights and water levels of the rat following reduced food intake. *J. Nutr.* 90(4):345–360.

Pfeiffer, C. J. (1998). A mathematical evaluation of the thymic weight parameter. *Toxicol. Appl. Pharmacol.* 13(2):220–227.

Prince, E. A., and Wilson, L. M. (2003). *Pathophysiology*, 6th ed. Mosley, Saint Louis, MD.

Rousseaux, C. G., Haschek, W. M., and Wallig, M. A. (2002). Toxicologic pathology: an introduction. In *Handbook of Toxicologic Pathology*, 2nd ed., Haschek, W. M., Rousseaux, C. G., and Wallig, M. A., Eds., Academic Press, San Diego, CA, chap. 1.

Roy, D., and Andrews, P. A. (2004). *Nonclinical* testing: From theory to practice. In B. A. Teicher and P. A. Andrews (Eds.), *Anticancer Drug Development Guide*. Humana, Totowa, NJ, p. 306.

Sellers, R. S., et al. (2007). Society of Toxicologic Pathology position paper: organ weight recommendations for toxicology studies. *Toxicol. Pathol.* 35:751–755.

Tehounwan, P. B., and Centeno, J. A. (2008). Toxicologic pathology. In S. C. Gad (Ed.), *Preclinical Development Handbook: Toxicology*. Wiley, Hoboken, NJ, pp. 551–580.

Turton, J., and Hoasen, J. (1998). *Target Organ Pathology*. Taylor and Francis, Philadelphia.

Wallig, M. A. (2002). Mophologic manifestations of toxic cell Injury. In W. M. Haschek, C. G. Rousseaux, and M. A. Wallig (Eds.), *Handbook of Toxicologic Pathology*. Academic, New York, pp. 39–65.

# *Irritation and Local Tissue Tolerance in Pharmaceutical Safety Assessment*

Both irritation and local tolerance studies assess the short-term hazard of pharmaceutical agents in the immediate region of their application or installation. In particular, these studies are done (expected) to assess topically or parenterally administered drug formulations (Gad and Chengelis, 1998). Note that these are hazard tests properly performed using the clinical formulation.

Topical local tolerance effects are almost entirely limited to irritation. Though this usually means dermal irritation, it can also be intracutaneous, mucosal, penile, perivascular, vaginal, bladder, rectal, nasal, or ocular, depending on the route of drug administration. All but ocular irritation use some version of a common subjective rating scale (see Table 1) to evaluate responses. The outcome of all of these tests primarily evaluates the response of the first region of tissue (which is exposed to the highest concentration) to an administered drug substance. In general, any factor which enhances absorption through the contacted tissue is likely to decrease tissue tolerance. Zhai et al. (2008) should be referred to for a more detailed coverage of the subject of topical tissue toxicology.

For the skin, this scale is used in the primary dermal irritation test, which is performed for those agents that are to be administered to patients by application

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

**TABLE 1 Evaluation of Local Tissue Reactions in Tissue Irritation Studies**

| Skin Reaction | Value |
|---|---|
| Erythema and eschar formation | |
| No erythema | 0 |
| Very slight erythema (barely perceptible) | 1 |
| Well-defined erythema | 2 |
| Moderate to severe erythema | 3 |
| Severe erythema (beet redness) to slight eschar formation (injuries in depth) | 4 |
| Necrosis (death of tissue) | +N |
| Eschar (sloughing or scab formation) | +E |
| Edema formation | |
| No edema | 0 |
| Very slight edema (barely perceptible) | 1 |
| Slight edema (edges of area well defined by definite raising) | 2 |
| Moderate edema (raised approximately 1 mm) | 3 |
| Severe edema (raised more than 1 mm and extending beyond area of exposure) | 4 |
| Total possible score for primary irritation | 8 |

to the skin. As with all local tolerance tests, it is essential that the material be evaluated in "condition of use"—that is, in the final formulated form, applied to test animals in the same manner that the agent is to be used clinically.

## 1 PRIMARY DERMAL IRRITATION TEST

A. *Rabbit Screening Procedure*

1. A group of at least four to six New Zealand white rabbits are screened for the study.
2. All rabbits selected for the study must be in good health; any rabbit exhibiting sniffles, hair loss, loose stools, or apparent weight loss is rejected and replaced.
3. One day (at least 18 h) prior to application of the test substance, each rabbit is prepared by clipping the hair from the back and sides using a small animal clipper. A size no. 10 blade is used to remove long hair and then a size no. 40 blade is used to remove the remaining hair.
4. Six animals with skin sites that are free from hyperemia or abrasion (due to shaving) are selected. Skin sites that are in the telogen phase (resting stage of hair growth) are used; those skin sites that are in the anagen phase (stage of active growth, indicated by the presence of a thick undercoat of hair) are not used.

B. *Study Procedure*

1. As many as four areas of skin, two on each side of the rabbit's back, can be utilized for sites for administration.
2. Separate animals are not required for an untreated control group. Each animal serves as its own control.

3. Besides the test substance, a positive control substance (a known skin irritant—1% sodium lauryl sulfate in distilled water) and a negative control (untreated patch) are applied to the skin. When a vehicle is used for diluting, suspending, or moistening the test substance, a vehicle control patch is required—especially if the vehicle is known to cause any toxic dermal reactions or if there is insufficient information about the dermal effects of the vehicle.
4. The intact (free-of-abrasion) sites of administration are assigned a code number. Up to four sites can be used as follows:
   1: Test substance
   2: Negative control
   3: Positive control
   4: Vehicle control (if required)
5. Application sites should be rotated from one animal to the next to ensure that the test substance and controls are applied to each position at least once.
6. Each test or control substance is held in place with a 1 × 1-in. 12-ply surgical gauze patch. The gauze patch is applied to the appropriate skin site and secured with 1-in.-wide strips of surgical tape at the four edges, leaving the center of the gauze patch nonoccluded.
7. If the test substance is a solid or a semisolid, a 0.5-g portion is weighed and placed on the gauze patch. The test substance patch is placed on the appropriate skin site and secured. The patch is subsequently moistened with 0.5 mL of physiological saline.
8. When the test substance is in flake, granule, powder, or other particulate form, the weight of the test substance that has a volume of 0.5 mL (after compacting as much as possible without crushing or altering the individual particles, such as by tapping the measuring container) is used whenever this volume is less than 0.5 g. When applying powders, granules, and the like, the gauze patch designated for the test sample is secured to the appropriate skin site with one of the four strips of the tape at the most ventral position of the animal. With one hand, the appropriate amount of sample measuring 0.5 mL is carefully poured from a glycine weighing paper onto the gauze patch that is held in a horizontal (level) position with the other hand. The patch containing the test sample is then carefully placed into position onto the skin and the remaining three edges secured with tape. The patch is subsequently moistened with 0.5 mL of physiological saline.
9. If the test substance is a liquid, a patch is applied and secured to the appropriate skin site. A 1-mL tuberculin syringe is used to measure and apply 0.5 mL of test substance to the patch.
10. The negative control site is covered with an untreated 12-ply surgical gauze patch (1 × 1 in.).
11. The positive control substance and vehicle control substance are applied to a gauze patch in the same manner as a liquid test substance.

12. The entire trunk of the animal is covered with an impervious material (such as saran wrap) for a 24-h period of exposure. The saran wrap is secured by wrapping several long strips of athletic adhesive tape around the trunk of the animal. The impervious material aids in maintaining the position of the patches and retards evaporation of volatile test substances.
13. An Elizabethan collar is fitted and fastened around the neck of each test animal. The collar remains in place for the 24-h exposure period. The collars are utilized to prevent removal of wrappings and patches by the animals while allowing the animals food and water ad libitum.
14. The wrapping is removed at the end of the 24-h exposure period. The test substance skin site is wiped to remove any test substance still remaining. When colored test substances (such as dyes) are used, it may be necessary to wash the test substance from the test site with an appropriate solvent or vehicle (one that is suitable for the substance being tested). This is done to facilitate accurate evaluation for skin irritation.
15. Immediately after removal of the patches, each 1 × 1-in. test or control site is outlined with indelible marker by dotting each of the four corners. This procedure delineates the site for identification.

C. *Observations*

1. Observations are made of the test and control skin sites 1 h after removal of the patches (25 h postinitiation of application). Erythema and edema are evaluated and scored on the basis of designated values presented in Table 1.
2. Observations are again performed 48 and 72 h after application and scores are recorded.
3. If necrosis is present or the dermal reaction is unusual, the reaction should be described. Severe erythema should receive the maximum score (4), and +N should be used to designate the presence of necrosis and +E the presence of eschar.
4. When a test substance produces dermal irritation that persists 72 h postapplication, daily observations of test and control sites are continued on all animals until all irritation caused by the test substance resolves or until day 14 postapplication.

D. *Evaluation of Results*

1. A *subtotal irritation value* for erythema or eschar formation is determined for each rabbit by adding the values observed at 25, 48, and 72 h postapplication.
2. A subtotal irritation value for edema formation is determined for each rabbit by adding the values observed at 25, 48, and 72 h postapplication.
3. A total irritation value is calculated for each rabbit by adding the subtotal irritation value for erythema or eschar formation to the subtotal irritation value for edema formation.

4. The *primary dermal irritation index* (PDII) is calculated for the test substance or control substance by dividing the sum of the total irritation scores by the number of observations (three days × three animals = nine observations).
5. The categories of the PDII are as follows [this categorization of dermal irritation is a modification of the original classification described by Draize et al. (1944)]:

| PDII | Rating |
|---|---|
| 0.0 | Nonirritant |
| >0.0–0.5 | Negligible irritant |
| >0.5–2.0 | Mild irritant |
| >2.0–5.0 | Moderate irritant |
| >5.0–8.0 | Severe irritant |

Other abnormalities, such as atonia or desquamation, should be noted and recorded.

## 2 OTHER NONPARENTERAL ROUTE IRRITATION TESTS

The design of vaginal, rectal, and nasal irritation studies is less formalized but follows the same basic pattern as the primary dermal irritation test. The rabbit is the preferred species for vaginal and rectal irritation studies, but the monkey and dog have also been used for these (Eckstein et al., 1969). Both the rabbit and rat have commonly seen use for nasal irritation evaluations. Defined quantities (typically 1.0 mL) of test solutions or suspensions are instilled into the orifice in question. For the vagina or rectum inert bungs are usually installed immediately thereafter to continue exposure for a defined period of time (usually the same period of hours as future human exposure). The orifice is then flushed clean, and 24 h after exposure it is examined and evaluated (graded) for irritation using the scale in Table 1.

## 3 FACTORS AFFECTING IRRITATION RESPONSES AND TEST OUTCOME

The results of local tissue irritation tests are subject to considerable variability due to relatively small differences in test design or technique. Weil and Scala (1971) arranged and reported on the best known of several intralaboratory studies to clearly establish this fact. Though the methods presented above have proven to give reproducible results in the hands of the same technicians over a period of years (Gad et al., 1986) and contain some internal controls (the positive and vehicle controls in the PDII) against large variabilities in

results or the occurrence of either false positives or negatives, it is still essential to be aware of those factors that may systematically alter test results. These factors are summarized below:

A. In general, any factor that increases absorption through the stratum corneum or mucous membrane will also increase the severity of an intrinsic response. Unless this factor mirrors potential exposure conditions, it may, in turn, adversely affect the relevance of test results.

B. The physical nature of solids must be carefully considered both before testing and in interpreting results. Shape (sharp edges), size (small particles may abrade the skin due to being rubbed back and forth under the occlusive wrap), and rigidity (stiff fibers or very hard particles will be physically irritating) of solids may all enhance an irritation response.

C. Solids frequently give different results when they are tested dry than if wetted for the test. As a general rule, solids are more irritating if moistened (going back to item A, wetting is a factor that tends to enhance absorption). Care should also be taken as to moistening agent—some (few) batches of U.S. Pharmacopeia (USP) physiological saline (used to simulate sweat) have proven to be mildly irritating to the skin and mucous membrane on their own. Liquids other than water or saline should not be used.

D. If the treated region on potential human patients will be a compromised skin surface barrier (e.g., if it is cut or burned), some test animals should likewise have their application sites compromised. This procedure is based on the assumption that abraded skin is uniformly more sensitive to irritation. Experiments, however, have shown that this is not necessarily true; some materials produce more irritation on abraded skin while others produce less (Guillot et al., 1982; Gad et al., 1986).

E. The degree of occlusion (in fact, the tightness of the wrap over the test site) also alters percutaneous absorption and therefore irritation. One important quality control issue in the laboratory is achieving a reproducible degree of occlusion in dermal wrappings.

F. Both the age of the test animal and the application site (saddle of the back vs. flank) can markedly alter test outcome. Both of these factors are also operative in humans, of course (Mathias, 1983), but in dermal irritation tests, the objective is to remove all such sources of variability. In general, as an animal ages, sensitivity to irritation decreases. For the dermal test, the skin middle of the back (other than directly over the spine) tends to be thicker (and therefore less sensitive to irritations) than that on the flanks.

G. The sex of the test animals can also alter study results because both regional skin thickness and surface blood flow vary between males and females.

H. Finally, the single most important (yet also most frequently overlooked) factor that influences the results and outcome of these (and in fact most) acute studies is the training of the staff. In determining how test materials are prepared and applied and in how results are "read" against a subjective scale, both

accuracy and precision are extremely dependent on the technicians involved. To achieve the desired results, initial training must be careful and all-inclusive. As important, some form of regular refresher training must be exercised, particularly in the area of scoring results. Use of a set of color photographic standards as a training and reference tool is strongly recommended; such standards should clearly demonstrate each of the grades in the Draize dermal scale.

I. It should be recognized that the dermal irritancy test is designed with a bias to preclude false negatives and, therefore, tends to exaggerate results in relation to what would happen in humans. Findings of negligible irritancy (or even in the very low mild irritant range) should therefore be of no concern unless the product under test is to have large-scale and prolonged dermal contact.

## 4 PROBLEMS IN TESTING (AND THEIR RESOLUTIONS)

Some materials, by either their physicochemical or toxicological natures, generate difficulties in the performance and evaluation of dermal irritation tests. The most commonly encountered of these problems are presented below:

A. *Compound Volatility* One is sometimes required or requested to evaluate the potential irritancy of a liquid that has a boiling point between room temperature and the body temperature of the test animal. As a result, the liquid portion of the material will evaporate off before the end of the testing period. There is no real way around the problem; one can only make clear in the report on the test that the traditional test requirements were not met, though an evaluation of potential irritant hazard was probably achieved (for the liquid phase would also have evaporated from a human that it was spilled on).

B. *Pigmented Material* Some materials are strongly colored or discolor the skin at the application site. This makes the traditional scoring process difficult or impossible. One can try to remove the pigmentation with a solvent; if successful, the erythema can then be evaluated. If use of a solvent fails or is unacceptable, one can (wearing thin latex gloves) feel the skin to determine if there is warmth, swelling, and/or rigidity—all secondary indicators of the irritation response.

C. *Systemic Toxicity* On rare occasions, the dermal irritation study is begun only to have the animals die very rapidly after test material is applied.

## 5 OCULAR IRRITATION TESTING

Ocular irritation is significantly different from the other local tissue irritation tests on a number of grounds. For the pharmaceutical industry, eye irritation testing is performed when the material is intended to be put into the eye as a

means or route of application for ocular therapy. There are a number of special tests applicable to pharmaceuticals or medical devices that are beyond the scope of this volume since they are not intended to assess potential acute effects or irritation. In general, however, it is desired that an eye irritation test that is utilized by this group be both sensitive and accurate in predicting the potential to cause irritation in humans. Failing to identify human ocular irritants (lack of sensitivity) is to be avoided, but of equal concern is the occurrence of false positives.

The primary eye irritation test was originally intended to predict the potential for a single splash of chemical into the eye of a human being to cause reversible and/or permanent damage. Since the introduction of the original Draize test more than 60 years ago (Draize et al., 1944), ocular irritation testing in rabbits has both developed and diverged. Indeed, clearly there is no longer a single test design that is used, and different objectives are pursued by different groups using the same test. This lack of standardization has been recognized for some time, and attempts have been made to address standardization of at least the methodological aspects of the test, if not the design aspects.

One widely used study design, which begins with a screening procedure as an attempt to avoid testing severe irritants or corrosives in animals, goes as follows:

A. *Test Article Screening Procedure*
   1. Each test substance will be screened in order to eliminate potentially corrosive or severely irritating materials from being studied for eye irritation in the rabbit.
   2. If possible, the pH of the test substance will be measured.
   3. A primary dermal irritation test will be performed prior to the study.
   4. The test substance will not be studied for eye irritation if it is a strong acid (pH of 2.0 or less) or strong alkali (pH of 12.0 or greater) and/or if the test substance is a severe dermal irritant (with a PDII of 5–8) or causes corrosion of the skin.
   5. If it is predicted that the test substance does not have the potential to be severely irritating or corrosive to the eye, continue to Section B, Rabbit Screening Procedure.

B. *Rabbit Screening Procedure*
   1. A group of at least six New Zealand white rabbits of either sex are screened for the study. The animals are removed from their cages and placed in rabbit restraints. Care should be taken to prevent mechanical damage to the eye during this procedure.
   2. All rabbits selected for the study must be in good health; any rabbit exhibiting sniffles, hair loss, loose stools, or apparent weight loss is rejected and replaced.
   3. One hour prior to installation of the test substance, both eyes of each rabbit are examined for signs of irritation and corneal defects with a

hand-held slit lamp. All eyes are stained with 2.0% sodium fluorescein and examined to confirm the absence of corneal lesions.

*Fluorescein staining:* Cup the lower lid of the eye to be tested and instill one drop of a 2% (in water) sodium fluorescein solution onto the surface of the cornea. After 15 s, thoroughly rinse the eye with physiological saline. Examine the eye, employing a hand-held long-wave ultraviolet illuminator in a darkened room. Corneal lesions, if present, appear as bright yellowish-green fluorescent areas.

4. Only three of the six animals are selected for the study. The three rabbits must not show any signs of eye irritation and must show either a negative or minimum fluorescein reaction (due to normal epithelial desquamation).

C. *Study Procedure*

1. At least 1 h after fluorescein staining, the test substance is placed in one eye of each animal by gently pulling the lower lid away from the eyeball to form a cup (conjunctival cul-de-sac) into which the test material is dropped. The upper and lower lids are then gently held together for 1 s to prevent immediate loss of material.
2. The other eye remains untreated and serves as a control.
3. For testing liquids, 0.01 mL of the test substance is used.
4. For solids or pastes, 100 mg of the test substance is used.
5. When the test substance is in flake, granular, powder, or other particulate form, the amount that has a volume of 0.01 mL (after gently compacting the particles by tapping the measuring container in a way that will not alter their individual form) is used whenever this volume weighs less than 10 mg.
6. For aerosol products, the eye should be held open and the substance administered in a single 1-s burst at a distance of about 4 in. directly in front of the eye. The velocity of the ejected material should not traumatize the eye. The dose should be approximated by weighing the aerosol can before and after each treatment. For other liquids propelled under pressure, such as substances delivered by pump sprays, an aliquot of 0.01 mL should be collected and instilled in the eye as for liquids.
7. The treated eyes of the three rabbits are not washed following the instillation of the test substance.
8. The treated eyes of the remaining three rabbits are irrigated for 1 min with room temperature tap water starting 20 s after instillation.
9. To prevent self-inflicted trauma by the animals immediately after instillation of the test substance, the animals are not immediately returned to their cages. After the test and control eyes are examined and graded at 1 h postexposure, the animals are returned carefully to their respective cages.

D. *Observations*

1. The eyes are observed for any immediate signs of discomfort after instilling the test substance. Blepharospasm and/or excessive tearing are

indicative of irritating sensations caused by the test substance, and their duration should be noted. Blepharospasm does not necessarily indicate that the eye will show signs of ocular irritation.

2. Grading and scoring or ocular irritation is performed in accordance with Table 2. The eyes are examined and grades of ocular reactions are recorded.

**TABLE 2 Scale of Weighted Scores for Grading Severity of Ocular Lesions**[a]

| Reaction Criteria | Score |
|---|---|
| I. Cornea | |
| A. Opacity degree of density (area that is most dense is taken for reading) | |
| 1. Scattered or diffuse area, details of iris clearly visible | 1 |
| 2. Easily discernible translucent area, details of iris slightly obscured | 2 |
| 3. Opalescent areas, no details of iris visible, size of pupil barely discernible | 3 |
| B. Area of cornea involved | |
| 1. One-quarter (or less) but not zero | 1 |
| 2. Greater than one-quarter, less than one-half | 2 |
| 3. Greater than one-half, less that whole area | 3 |
| 4. Greater than three-quarters up to whole area | 4 |
| Scoring equals $A \times B \times 5$; total maximum = 80[b] | |
| II. Iris | |
| A. Values | |
| 1. Folds above normal, congestion, swelling, circumcorneal ingestion (any one or all of these or combination of any thereof), iris still reacting to light (sluggish reaction possible) | 1 |
| 2. No reaction to light, hemorrhage, gross destruction (any one or all of these) | 2 |
| Scoring equals $A \times B$ (where $B$ is the area of the iris involved, graded as "under cornea"); total maximum = 10 | |
| III. Conjunctivae | |
| A. Redness (refers to palpebral conjunctivae only) | |
| 1. Vessels definitely injected above normal | 1 |
| 2. More diffuse, deeper crimson red, individual vessels not easily discernible | 2 |
| 3. Diffuse beefy red | 3 |
| B. Chemosis | |
| 1. Any swelling above normal (includes nictitating membrane) | 1 |
| 2. Obvious swelling with partial eversion of lids | 2 |
| 3. Swelling with lids about half closed | 3 |
| 4. Swelling with lids about half closed to completely closed | 4 |
| C. Discharge | |
| 1. Any amount different from normal (does not include small amount observed in inner canthus of normal animals) | 1 |
| 2. Discharge with moistening of lids and hair just adjacent to lids | 2 |
| 3. Discharge with moistening of lids and considerable area around eye | 3 |
| Scoring $(A + B + C) \times 2$; total maximum = 20 | |

[a]The maximum total score is the sum of all scores obtained for cornea, iris, and conjunctivae.
[b]All $A \times B = \sum(1\text{–}3) \times \sum(1\text{–}4)$ for three animals.

3. If signs or irritation persists at day 7, readings are continued on days 10 and 14 after exposure or until all signs of reversible toxicity are resolved.
4. In addition to the required observation of the cornea, iris, and conjunctiva, serious effects (such as pannus, rupture of the globe, or blistering of the conjunctivae) indicative of a corrosive action are reported.
5. Whether or not toxic effects are reversible depends on the nature, extent, and intensity of damage. Most lesions, if reversible, will heal or clear within 21 days. Therefore, if ocular irritation is present at the 14-day reading, a 21-day reading is required to determine whether the ocular damage is reversible or nonreversible.

## 6 VAGINAL IRRITATION

Few, if any, products are administered via the vagina that are intended for systemic absorption. Thus, this route has not been as widely studied and characterized as others. On the other hand, large numbers of different products (douches, spermicides, antiyeast agents, etc.) have been developed that require introduction into the vagina in order to assert their localized effects. Increased research into different birth control and antiviral prophylaxis will result in more vaginal products in the future. All these must be assessed for vaginal irritation potential, and this serves as an example of the other tissue tolerance issues.

Considerable research (Eckstein et al., 1969; Auletta, 1994) has indicated that the rabbit is the best species for assessing vaginal irritation. There are those investigators, however, who consider the rabbit too sensitive and recommend the use of ovariectomized rats. Ovariectomy results in a uniformly thin, uncornified epithelium which is more responsive to localized effects. This model is used when the results from a study with rabbits are questionable (Auletta, 1994). The routine progression of studies consists of first doing an acute primary vaginal irritation study, then a 10-day repeated dose study in rats. These protocols are summarized below. Longer term vaginal studies have been conducted in order to assess systemic toxicity of the active agents when administered by these routes (while the intended effects may be local, one cannot assume that there will be no systemic exposure).

### 6.1 Acute Primary Vaginal Irritation Study in Female Rabbit

1. *Overview of Study Design* One group of three adult rabbits received a single vaginal exposure and were observed for three days with periodic examination (1, 24, 48, and 72 h postdosing) of the genitalia. Animals are then euthanized and the vagina is examined macroscopically.
2. *Administration*
   *Route* The material is generally introduced directly into the vagina using a lubricated 18 French rubber catheter attached to a syringe for

quantification and delivery of the test material. Gentle placement of the catheter is important because one needs to ensure complete delivery of the dose without mechanical trauma. For rabbits, the depth of insertion is about 7.5 cm and the catheter should be marked to about that depth. After deliver is completed, the tube is slowly withdrawn. No attempt is made to close or seal the vaginal orifice. Alternative methods may be used to administer more viscous materials. The most common is to backload a lubricated 1-mL tuberculine syringe, then warm the material to close to body temperature. The syringe is then inserted into the vagina and the dose administered by depressing the syringe plunger.

*Dosage* The test material should be one (concentration, vehicle, etc.) that is intended for human application.

*Frequency* Once.

*Duration* One day.

*Volume* One milliliter per rabbit.

3. *Test System* Species, age, and weight range: Sexually mature New Zealand white rabbits are generally used weighing between 2 and 5 kg. The weight is not as important as the fact that the animals need to be sexually mature.

Selection: Animals should be multiparous and nonpregnant. Animals should be healthy and free of external genital lesions.

Randomization: Because there is only one group of animals, randomization is not a critical issue.

4. *In-Life Observations*

*Daily Observations* At least once daily for clinical signs.

*Detailed Physical Examination* Once during the week prior to dosing.

*Body Weight* Day of dosing.

*Vaginal Irritation* Scored at 1, 24, 48, and 72 h postdosing. Scoring criteria are shown in Table 3.

5. *Postmortem Procedures* Rabbits are euthanized by lethal dose of a barbiturate soon after the last vaginal irritation scores are collected. The vagina is opened by longitudinal section and examined for evidence of mucosal damage such as erosion, localized hemorrhage, and so on. No other tissues are examined. No tissues are collected. After the macroscopic description of the vagina is recorded, the animal is discarded.

### 6.2 Repeated-Dose Vaginal Irritation in Female Rabbit

1. *Overview of Study Design* Four groups of three adult rabbits each receive a single vaginal exposure daily for 10 days. The genitalia are examined daily. Animals are then euthanized and the vagina is examined macroscopically and microscopically.

**TABLE 3 Scoring Criteria for Vaginal Irritation**

| Value | |
|---|---|
| | *Erythema* |
| 0 | No erythema |
| 1 | Very slight erythema (barely perceptible) |
| 2 | Slight erythema (pale red in color) |
| 3 | Moderate to severe erythema (definite red in color) |
| 4 | Severe erythema (beet or crimson red) |
| | *Edema* |
| 0 | No edema |
| 1 | Very slight edema (barely perceptible) |
| 2 | Slight edema (edges of area well defined by definite raising |
| 3 | Moderate edema (raised approximately 1 mm) |
| 4 | Severe edema (raised more than 1 mm and extending beyond area of exposure) |
| | *Discharge* |
| 0 | No discharge |
| 1 | Very slight discharge |
| 2 | Slight discharge |
| 3 | Moderate discharge |
| 4 | Severe discharge (moistening of considerable area around vagina) |

2. *Administration*

   *Route* The test materials are introduced directly into the vagina with a lubricated 18 French rubber catheter using the techniques described previously for acute studies.

   *Dosage* The test material should be one (concentration, vehicle, etc.) that is intended for human application. There will also be a sham-negative control (catheter in place but nothing administered), a vehicle control, and a positive control (generally 2% nonoxynol-9).

   *Frequency* Once daily.

   *Duration* Ten days.

   *Volume* One milliliter per rabbit for each material.

3. *Test System*

   *Species, Age, and Weight Range* Sexually mature New Zealand white rabbits are generally used weighing between 2 and 5 kg. The weight is not as important as the fact that the animals need to be sexually mature.

   *Selection* Animals should be nulliparous and nonpregnant. Animals should be healthy and free of external genital lesions.

   *Randomization* At least 14 animals should be on pretest. Randomization to treatment groups is best done using a computerized blocking by body weight method or a random-number generation method.

4. *In-Life Observations*

   *Daily Observations* At least once daily for clinical signs.

*Detailed Physical Examination* Once during the week prior to dosing and immediately prior to necropsy.

*Body Weight* First, fifth, and last day of dosing.

*Vaginal Irritation* Scored once daily. Scoring criteria are shown in Table 3.

5. *Postmortem Procedures* Rabbits are euthanized by lethal dose of a barbiturate soon after the last vaginal irritation scores are collected. The vagina is isolated using standard prosection techniques and then opened by longitudinal section and examined for evidence of mucosal damage such as erosion, localized hemorrhage, and so on. No other tissues are examined. The vagina and cervix are collected and fixed in 10% neutral buffered formalin. Standard hematoxylin/eosin-stained, paraffin-embedded histological glass slides are prepared by routine methods. Three levels of the vagina (low, mid, and upper) are examined and graded using the scoring system shown in Table 4. Each level is cored separately and an average is calculated. Irritation is rated as follows:

**TABLE 4 Microscopic Scoring Procedure for Vaginal Sections**

| Section | Value |
|---|---|
| Epithelium | |
| Intact—normal | 0 |
| Cell degeneration or flattening of epithelium | 1 |
| Metaplasia | 2 |
| Focal erosion | 3 |
| Erosion or ulceration, generalized | 4 |
| Leukocytes | |
| Minimal: <25 per high-power field | 1 |
| Mild: 25–50 per high-power field | 2 |
| Moderate: 50–100 per high-power field | 3 |
| Marked: >100 per high-power field | 4 |
| Injection | |
| Absent | 0 |
| Minimal | 1 |
| Mild | 2 |
| Moderate | 3 |
| Marked with disruption of vessels | 4 |
| Edema | |
| Absent | 0 |
| Minimal | 1 |
| Mild | 2 |
| Moderate | 3 |
| Marked | 4 |

*Source*: Eckstein et al., 1969.

| Score | Rating |
|---|---|
| 0 | Nonirritating |
| 1–4 | Minimal irritation |
| 5–8 | Mild irritation |
| 9–11 | Moderate irritation |
| 12–16 | Marked irritation |

The score for each rabbit is then averaged and acceptability ratings are given as follows:

| Average Score | Acceptability Rating |
|---|---|
| 0–8 | Acceptable |
| 9–10 | Marginal |
| 11 or greater | Unacceptable |

### 6.3 Repeated-Dose Vaginal Irritation in Ovariectomized Rats

This study is very similar in design to that described previously for rabbits, with the following (sometimes obvious) exceptions. Mature ovariectomized female rats can be obtained from a commercial breeder. A 15% surplus should be obtained. Ten animals per group should be used (40 total for the study). The vaginal catheter is placed to a depth of approximately 2.5 cm and the treatment volume should be 0.2 mL.

## 7 PARENTERAL IRRITATION/TOLERANCE

There are a number of special concerns about the safety of materials that are routinely injected (parenterally administered) into the body. By definition, these concerns are all associated with materials that are the products of the pharmaceutical and (in some minor cases) medical device industries. Such parenteral routes include three major ones—IV (intravenous), IM (intramuscular), and SC (subcutaneous)—and a number of minor routes (such as intra-arterial) that are not considered here.

These unusual concerns include irritation (vascular, muscular, or subcutaneous), pyrogenicity, blood compatibility, and sterility (Avis, 1985). The background of each of these, along with the underlying mechanisms and factors that influence the level of occurrence of such an effect, are briefly discussed below.

***Irritation*** Tissue irritation upon injection and the accompanying damage and pain are concerns that must be addressed for the final formulation, which is to be either tested in humans or marketed, rather than for the active ingredient. This is because most irritation factors are either due to or influenced by

aspects of formulation design [see Avis (1985) for more information or parenteral preparations]. These factors are not independent of the route (IV, IM, or SC) that will be used and, in fact (as discussed later), are part of the basis for selecting between the various routes (USP 1995a).

The lack of irritation and tissue damage at the injection site is sometimes called *tolerance*. Some of the factors that affect tolerance are not fully under the control of an investigation and are also unrelated to the material being injected. These include body movement, temperature, and animal age. Factors that can be controlled but that are not inherent to the active ingredient include solubility, tonicity, and pH. Finally, the active ingredient and vehicle can have inherent irritative effects and factors such as solubility (in the physiological milieu into which they are being injected), concentration, volume molecular size, and particle size. Gray (1978) and Ballard (1968) discuss these factors and the morphological consequences that may occur if they are not addressed.

***Pyrogenicity*** Pyrogenicity is the induction of a febrile (fever) response induced by the parenteral (usually IV or IM) administration of exogenous material (USP 1995b). Pyrogenicity is usually associated with microbiological contamination of a final formulation, but it is now of increasing concern because of the growing interest in biosynthetically produced materials. Generally, ensuring sterility of product and process will guard against pyrogenicity for traditional pharmaceuticals. For biologically produced products, the U.S. Food and Drug Administration (FDA) has promulgated the general guideline that no more than 5.0 units of endotoxin may be present per milligram of drug substance.

***Blood Compatibility*** It is important that cellular components of the blood are not disrupted and that serum- or plasma-based responses are not triggered by parenteral administration. Therefore, two mechanisms must be assessed regarding the blood compatibility of component materials. These include the material's effect on cellular components that cause membrane destruction and hemolysis and the activation of the clotting mechanism resulting in the formation of the thromboemboli.

Many of the nonactive, ingredient-related physicochemical factors that influence irritation (tonicity, pH, and particle size, for example) also act to determine blood compatibility. But the chemical features of a drug entity itself—its molecular size and reactivity—can also be of primary importance.

***Sterility*** Sterility is largely a concern to be answered in the process of preparing a final clinical formulation, and it is not addressed in detail in this chapter. However, it should be clear that it is essential that no viable microorganisms are present in any material to be parenterally administered (except for vaccines).

### 7.1 Parenteral Routes

There are at least 13 different routes by which to inject material into the body:

1. Intravenous
2. Subcutaneous
3. Intramuscular
4. Intra-arterial
5. Intradermal
6. Intralesional
7. Epidural
8. Intrathecal
9. Intracisternal
10. Intracardiac
11. Intraventricular
12. Intraocular
13. Intraperitoneal

Only the first three are discussed in any detail here. Most of these routes of administration place a drug directly or indirectly into systemic circulation. There are a number of these routes, however, by which the drug exerts a local effect, in which case most of the drug does not enter systemic circulation (e.g., intrathecal, intraventricular, intraocular, intracisternal). Certain routes of administration may exert both local and systemic effects depending on the characteristics of the drug and excipients (e.g., subcutaneous).

The choice of a particular parenteral route will depend on the required time of onset of action, the required site of action, and the characteristics of the fluid to be injected, among other factors.

***Bolus versus Infusion*** Technically, for all the parenteral routes (but in practice only for the IV route), there are two options for injecting a material into the body. The bolus and infusion methods are differentiated on the single basis of rate of injection, but they actually differ on a wide range of characteristics.

The most commonly exercised option is the bolus, or "push," injection, in which the injection device (syringe or catheter) is appropriately entered into the vein and a defined volume of material is introduced through the device. The device is then removed. In this operation, it is relatively easy to restrain an experimental animal and the stress on the animal is limited. Though the person doing the injection must be skilled, it takes only a short amount of time to become so. And the one variable to be controlled in determining dosage is the total volume of material injected (assuming dosing solutions have been properly prepared).

There are limitations and disadvantages to the bolus approach, however. Only a limited volume may be injected, which may prohibit the use of bolus when volumes to be introduced are high (due to, e.g., low active compound solubility or a host of other reasons). Only two devices (syringe and catheter) are available for use in the bolus approach. If a multiple-day course of treatment is desired (say, every day for 15 days), separate injections must be made at discrete entry sites.

The infusion approach involves establishing a fixed entry point into the vein, then slowly passing the desired test material through that point over a

period of time (30 min is about minimum, while continuous around-the-clock treatment is at least therapeutically possible). There are a number of devices available for establishing their entry point: catheter, vascular port (Garramone, 1986), or osmotic pump (Theeuwes and Yum, 1976). Each of these must, in turn, be coupled with a device to deliver the dosing solution at a desired rate. The osmotic pump, which is implanted, is also its own delivery device. Other options are gravity driven "drips," hand-held syringes (not practical or accurate over any substantial period of time), or syringe pumps. Very large volumes can be introduced by fusion over a protracted period of time, and only a single site need be fitted with an entry device.

However, infusions also have their limitations. Skilled labor is required, and the setup must be monitored over the entire period of infusion. Larger animals must be restrained, while there are special devices that make this requirement unnecessary for smaller animals. Restraint and protracted manipulation are very stressful on animals. Over a period of time, one must regularly demonstrate patency of a device—that is, that entry into the vascular system continues to exist. Finally, one is faced with having to control two variables in controlling the dose—both total volume and rate.

When are the two approaches (bolus and infusion) interchangeable? And why select one over the other? The selection of infusion is usually limited to two reasons: (1) when a larger volume must be introduced than is practical in a bolus injection or (2) when tolerance is insufficient if the dose is given all at once (i.e., an infusions will "clear" a higher daily dose than will a bolus injection). For safety studies, when a bolus can be used to clear a human, infusion dosing is a matter of judgment. If the planned clinical infusion will take less than half an hour, practicality dictates that the animal studies be accomplished by bolus. In other situations, pharmacokinetics (in particular, the half-life of the drug entity) should be considered in making the decision.

### 7.2 Test Systems for Parenteral Irritation

There are no regulatory guidelines or suggested test methods for evaluating agents for muscular or vascular irritation. Since such guidelines are lacking but the evaluation is necessary, those responsible for these evaluations have tried to develop and employ the most scientifically valid procedures.

Hagan (1959) first suggested a method for assessing IM irritation. His approach, however, did not include a grading system for evaluation of the irritation, and the method used the sacrospinalis muscles, which are somewhat difficult to dissect or repeatedly inject.

Shintani et al. (1967) developed and proposed the methodology that currently seems to be more utilized. It uses the lateral vastus muscle and includes a methodology for evaluation, scoring, and grading of irritation. Additionally, Shintani et al. investigated the effects of several factors, such as pH of the solution, drug concentration, volume of injection, effect of repeated injections, and time to maximum tissue response.

### *Acute Intramuscular Irritation in Male Rabbit (USP 1985)*

1. *Overview of Study Design* Each rabbit is injected as follows:

| Site | Treatment |
|---|---|
| Muscle vastus lateralis | 1.0 mL/site |
| Left | Test article |
| Right | Vehicle |

Day 1: Injection of all treatment groups—nine rabbits
Day 2: Sacrifice and evaluation: 24 h posttreatment group—three rabbits
Day 3: Sacrifice and evaluation: 48 h posttreatment group—three rabbits
Day 4: Sacrifice and evaluation: 72 h posttreatment group—three rabbits

2. *Administration*
   2.1. Route: The test article is injected into the vastus lateralis of each rabbit.
   2.2. Dose: The dose selected is chosen to evaluate the severity of irritation and represents a concentration that might be used clinically. This volume has been widely used in irritation testing.
   2.3. Frequency: Once only.
   2.4. Duration: One day
   2.5. Volume: One milliliter per site.
3. *Test System*
   3.1. Species, age, and weight range: Male New Zealand white rabbits weighing 2–5 kg are used. The New Zealand white rabbit has been widely used in muscle irritation research for many years and is a reasonable sized, even-tempered animal that is well adapted to the laboratory environment.
   3.2. Selection: Animals to be used in the study are selected on the basis of acceptable findings from physical examinations and body weights.
   3.3. Randomization: Animals are ranked by body weight and assigned a number between 1 and 3. The order of number assigned (e.g., 1–3–2) is chosen from a table of random numbers. Animals assigned number 1 are in the 24-h-posttreatment group; those assigned number 2 are in the 48-h-posttreatment group; and those assigned number 3 are in the 72-h-posttreatment group.
4. *In-Life Observations*
   4.1. Daily observations: Once daily following dosing.
   4.2. Physical examinations: Once within the two weeks before the first dosing day.
   4.3. Body weight: Should be determined once before the start of the study.
   4.4. Additional examinations may be done by the study director to elucidate any observed clinical signs.

5. *Postmortem Procedures*
   5.1. Irritation is evaluated as follows: Three rabbits are sacrificed by a lethal dose of barbiturate at approximately 24, 48, or 72 h after dosing. The left and right lateral vastus of each rabbit are excised. The lesions resulting from injection are scored for muscle irritation on a numerical scale of 0–5 as follows (Shintani et al., 1967):

| Reaction Criteria | Score |
|---|---|
| No discernable gross reaction | 0 |
| Slight hyperemia | 1 |
| Moderate hyperemia and discoloration | 2 |
| Distinct discoloration in comparison with color of surrounding area | 3 |
| Brown degeneration with small necrosis | 4 |
| Widespread necrosis with appearance of "cooked meat" and occasionally an abscess involving major portions of muscle | 5 |

The average score for the nine rabbits is then calculated, and a category of irritancy is then assigned based on the following table:

| Average Score | Grade |
|---|---|
| 0.0–0.4 | None |
| 0.5–1.4 | Slight |
| 1.5–2.4 | Mild |
| 2.5–3.4 | Moderate |
| 3.5–4.4 | Marked |
| 4.5 or greater | Severe |

***Acute Intravenous Irritation in Male Rabbit*** The design here is similar to the intramuscular assay, except that injections are made into the veins in specific muscle masses.

1. *Overview of Study Design* Rabbits will be injected as follows:

| Group | No. of Animals | Treatment Site | Evaluation |
|---|---|---|---|
| 1 | 2 | Muscles vastus lateralis (left) and cervicodorsal subcutis (left) | 24 h |
| | | Muscles vastus lateralis (right) and cervicodorsal subcutis (right) | 24 h |
| 2 | 2 | Muscles vastus lateralis (left) and cervicodorsal subcutis (left) | 72 h |
| | | Muscles vastus lateralis (right) and cervicodorsal subcutis (right) | 72 h |
| 3 | 2 | Auricular vein, left and right (both evaluated) | 24 and 72 h |

Day 1: Injection of all groups (six rabbits).
Day 2: Evaluation of group 3 (two rabbits). Sacrifice and evaluation of group 1 (two rabbits).
Day 4: Evaluation of group 3 (two rabbits). Sacrifice and evaluation of group 2 (two rabbits).

2. *Administration*
   2.1. Intramuscular: Vastus lateralis.
   2.2. Subcutaneous: Cervicodorsal subcutis.
   2.3. Intravenous: Auricular vein.
   2.4. Dose: The doses and concentration selected are chosen to evaluate the severity of irritation. The dose volumes have been widely used in irritation testing.
   2.5. Frequency: Once only.
   2.6. Duration: One day.
   2.7. Volume: Muscles vastus lateralis and cervicodorsal subcutis: 1.0 mL per site; auricular vein: 0.5 mL per site.
3. *Test System*
   3.1. Species, age, and weight range: Male New Zealand white rabbits weighing 2–5 kg are used.
   3.2. Selection: Animals to be used in the study are selected on the basis of acceptable findings from physical examinations.
   3.3. Randomization: Animals are ranked by body weight and assigned a number between 1 and 3. The order of numbers assigned (e.g., 1–3–2) is chosen from a table of random numbers. Animals assigned number 1 are in group 1, those assigned number 2 are in group 2, and those assigned number 3 are in group 3.
4. *In-Life Observations*
   4.1. Daily observations: Once daily following dosing.
   4.2. Physical examinations: Once within the two weeks before the first dosing day.
   4.3. Body weight: Determined once before the start of the study.
   4.4. Additional examinations may be done by the study director to elucidate any observed clinical signs.
5. *Postmortem Procedures*
   5.1. Intramuscular irritation is evaluated as follows: Rabbits are sacrificed by lethal dose of barbiturate approximately 24 and 72 h after dosing. The left and right lateral vastus muscles of each rabbit are excised. The reaction resulting from injection is scored for muscle irritation using the scale shown in section 1 above.
   5.2. Subcutaneous irritation is evaluated as follows: Rabbits are scarified by a lethal dose of barbiturate approximately 24 and 72 h after dosing. The subcutaneous injection sites are exposed by dissection, and the reaction is scored for irritation on a scale of 0–5 as follows (Shintani et al., 1967):

| Reaction Criteria | Score |
|---|---|
| No discernible gross reaction | 0 |
| Slight hyperemia and discoloration | 1 |
| Moderate hyperemia and discoloration | 2 |
| Distinct discoloration in comparison with color of surrounding area | 3 |
| Small areas of necrosis | 4 |
| Widespread necrosis, possibly involving underlying muscle | 5 |

| Average Score per Site | Irritancy Grade |
|---|---|
| 0.0–0.4 | None |
| 0.5–1.4 | Slight |
| 1.5–2.4 | Mild |
| 2.5–3.4 | Moderate |
| 3.5–4.4 | Marked |
| 4.5 or greater | Severe |

5.3. Intravenous irritation is evaluated as follows: Rabbits are sacrificed by a lethal dose of barbiturate following the 72-h irritation evaluation. The injection site and surrounding tissue are grossly evaluated at approximately 24 and 72 h after dosing on a scale of 0–3 as follows:

| Reaction Criteria | Score |
|---|---|
| No discernible gross reaction | 0 |
| Slight erythema at injection site | 1 |
| Moderate erythema and swelling with some discoloration of vein and surrounding tissue | 2 |
| Severe discoloration and swelling of vein and surrounding tissue with partial or total occlusion of vein | 3 |

| Average Score per Site | Irritancy Grade |
|---|---|
| 0.0–0.4 | None |
| 0.5–1.4 | Slight |
| 1.5–2.4 | Moderate |
| 2.5 or greater | Severe |

5.4. Additional examinations may be done by the study director to elucidate the nature of any observed tissue change.

## 7.3 Alternatives

Intramuscular and IV injection of parenteral formulations of pharmaceuticals can produce a range of discomfort including pain, irritation, and/or damage to muscular or vascular tissue. These are normally evaluated for prospective formulations before use in humans by histopathological evaluation of damage in intact animal models, usually the rabbit. Attempts have been made to make this in vivo methodology both more objective and quantitative based on measuring the creatinine phosphokinase released in the tissue surrounding the

injection site (Sidell et al., 1974). Currently, a protocol utilizing a cultured skeletal muscle cell line (L6) from the rat as a model has been evaluated in an interlaboratory validation program among 11 pharmaceutical laboratories. This methodology (Young et al., 1986) measures creatine kinase levels in media after exposure of the cells to the formulation of interest and predicts in vivo IM damage based on this endpoint. It is reported to give excellent rank-correlated results across a range of antibiotics (Williams et al., 1987). The current multilaboratory evaluation covers a broader structural range of compounds and has shown a good quantitative correlation with in vivo results for antibiotics and a fair correlation for a broader range of parenteral drug products. Likewise, Kato et al. (1992) have proposed a model that uses cultured primary skeletal muscle fibers from the rat. Damage is evaluated by the release of creatinine phosphokinase. An evaluation using six parenterally administered antibiotics [ranking their median effective concentration ($EC_{50}$) values] showed good relative correlation with in vivo results.

Another proposed in vitro assay for muscle irritancy for injectable formulations is the red blood cell hemolysis assay (Brown et al., 1989). Water-soluble formulations in a 1:2 ratio with freshly collected human blood are gently mixed for 5 min. The percentage of red blood cell survival is then determined by measuring differential absorbance at 540 nm; this value is then compared to values for known irritants and nonirritants. Against a very small group of compounds (four), this assay reportedly accurately predicts muscle irritation.

## 8 PHOTOTOXICITY

The potential for sunlight (or selected other light frequencies) to transform a drug or product is both a useful tool for activating some drugs and a cause of significant adverse effects for others [such as the quinolone antibiotics (Horio et al., 1995; Lambert et al., 1996)].

### 8.1 Theory and Mechanisms

The portion of the solar spectrum containing the biologically most active region is from 290 to 700 mm.

The ultraviolet (UV) part of the spectrum includes wavelengths from 200 to 400 nm. Portions of the UV spectrum have distinctive features from both the physical and biological points of view. The accepted designations for the biologically important parts of the UV spectrum are UVA, 400–315 nm; UVB, 315–280 nm; and UVC, 280–220 nm. Wavelengths less than 290 nm (UVC) do not occur at the earth's surface because they are absorbed, predominantly by ozone in the stratosphere. The most thoroughly studied photobiological reactions that occur in skin are induced by UVB. The quinolones, for example, absorb light strongly in the 300–400-nm wavelength range. Although UVB wavelengths represent only approximately 1.5% of the solar energy received

at the earth's surface, they elicit most of the known chemical phototoxic and photoallergic reactions. The visible portions of the spectrum, representing about 50% of the sun's energy received at sea level, include wavelengths from 400 to 700 nm. Visible light is necessary for such biological events as photosynthesis, the regulation of circadian cycles, vision, and pigment darkening. Furthermore, visible light in conjunction with certain chromophores (e.g., dyes, drugs, and endogenous compounds which absorb light and therefore "give" color) and molecular oxygen induces photodynamic effects.

Understanding the toxic effects of light impinging on the skin requires knowledge of both the nature of sunlight and the skin's optical properties. Skin may be viewed as an optically heterogeneous medium composed of three layers that have distinct refractive indices, chromophore distributions, and light-scattering properties. Light entering the outermost layer, the stratum corneum, is in part reflected—4–7% for wavelengths between 250 and 3000 nm (Anderson and Parrish, 1981)—due to the difference in refractive index between air and the stratum corneum. Absorption by urocanic acid (a deamination product of histidine), melanin, and proteins containing the aromatic amino acids tryptophan and tyrosine in the stratum corneum produces further attenuation of light, particularly at shorter UV wavelengths. Approximately 40% of UVB is transmitted through the stratum corneum to the viable epidermis. The light entering the epidermis is attenuated by scattering and, predominantly, absorption. Epidermal chromophores consist of proteins, urocanic acid, nucleic acids, and melanin. Passage through the epidermis results in appreciable attenuation of UVA and UVB radiation. The transmission properties of the dermis are largely due to scattering, with significant absorption of visible light by melanin, β-carotene, and the blood-borne pigments bilirubin, hemoglobin, and oxyhemoglobin. Lightly traversing these layers of the skin is extensively attenuated, most drastically for wavelengths less than 400 nm. Longer wavelengths are more penetrating. It has been noted that there is an "optical window"—that is, greater transmission—for light at wavelengths of 600–1300 nm, which may have important biological consequences.

Normal variations in the skin's optical properties frequently occur. The degree of pigmentation may produce variations in the attenuation of light, particularly between 300 and 400 nm, by as much as 1.5 times more in blacks than in Caucasians (Pathak, 1967). Alterations in the amount or distribution of other natural chromophores account for further variations in skin optical properties. Urocanic acid deposited on the skin's surface during perspiration (Anderson and Parrish, 1981) and UV-absorbing lipids excreted in sebum may significantly reduce UV transmission through the skin. Epidermal thickness, which varies over regions of the body and increases after exposure to UVB radiation, may significantly modify UV transmission.

Certain disease states also produce alterations in the skin's optical properties. Alterations of the skin's surface, such as by psoriatic plaques, decrease

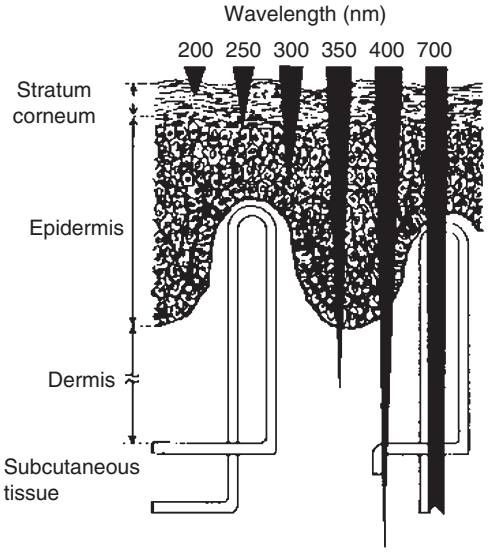


***Figure 1*** *Schematic penetration of light of varying wavelengths into skin.*

transmitted light. The effect may be lessened by application of oils whose refractive index is similar to that of skin (Anderson and Parrish, 1981). Disorders such as hyperbilirubinemia, porphyrias, and blue skin nevi result in increased absorption of visible light due to accumulation or altered distribution of endogenous chromophoric compounds.

The penetration of light into and through dermal tissues has important consequences. This penetration is demonstrated in Figure 1. Skin, as the primary organ responsible for thermal regulation, is overperfused relative to its metabolic requirements (Anderson and Parrish, 1981).

It is estimated that the average cutaneous blood flow is 20–30 times that necessary to support the skin's metabolic needs. The papillary boundaries between epidermis and dermis allow capillary vessels to lie close to the skin's surface, permitting the blood and important components of the immune system to be exposed to light. The equivalent of the entire blood volume of an adult may pass through he skin and potentially be irradiated in 20 min. This corresponds to the time required to receive one or two MEDs (the MED is defined as the minimal dose of UV irradiation that produces definite, but minimally perceptible, redness 24 h after exposure). The accessibility of incidence radiation to blood has been exploited in such regimens and phototherapy of hyperbilirubinemia in neonates, where light is used as a therapeutic agent. However, in general, there is a potential for light-induced toxicity due to irradiation of blood-borne drugs and metabolites.

## 8.2 Factors Influencing Phototoxicity/Photosensitization

There are a number of factors which can influence an agent acting as either a phototoxin or a photoallergen. In addition to the factors reviewed in Chapter Formulations, Routes, and Dosage Design, there are the following:

1. The quantity and location of photoactive material present in or on the skin
2. The capacity of the photoactive material to penetrate into normal skin by percutaneous absorption as well as into skin altered by trauma, such as maceration, irritation, and sunburn
3. The pH, enzyme presence, and solubility conditions at the site of exposure
4. The quantity of activating radiation to which the skin is exposed
5. The capacity of the spectral range to activate the materials on or within the skin
6. The ambient temperature and humidity
7. The thickness of the horny layer
8. The degree of melanin pigmentation of the skin
9. The inherent "photoactivity" of the chemical, whether it weakly or strongly absorbs light

Basically, any material that has both the potential to absorb ultraviolet light (in the UVA or UVB region) and the possibility of dermal exposure or distribution into the dermal region should be subject to some degree of suspicion as to potential phototoxicity. As shown in Table 5, a large number of agents have been identified as phototoxic or photoallergenic agents. Of these, tetrachlorosalicylanilide (TCSA) is the most commonly used as a positive control in animal studies.

## 8.3 Predictive Tests for Phototoxicity

Before we start on our description of the different methods, we will first cover some basics on light dosimetry. The intensity of the irradiation used in phototoxicity testing is determined with a light meter, which provides output as watts per square meter. The shelves on which the animals rest during the exposure periods are normally adjustable in order to control the dose of light to the exposure area. The irradiation from fluorescent lights will vary somewhat from day to day, depending on, for example, temperature and variations in line current. The dose the animals receive is generally represented as joules per square centimeter. A joule is equal to one watt per second. Therefore, the dose of light is dependent on the time of exposure. For example, in their review, Lambert et al. (1996) discuss dosages of UVA light of 9 or 10 $J\,cm^{-2}$ in the UVA spectral region. If the irradiation from the light is found to be

**TABLE 5 Known Phototoxic Agents**

| In Humans | | In Animals | |
|---|---|---|---|
| Compounds | Route | Compound | Route |
| Aminobenzoic acid derivatives | Topical | Acradine | Topical |
| | | Amiodarone | Oral |
| Amyldimethylamino benzoate, mixed ortho and para isomers | Topical | Anthracine | Topical |
| | | Bergapten (5-methoxypsolaren) | Topical |
| Anthracene acridine | Topical | Bithionol | Topical |
| Bergapten (5-methoxypsoralen) | Topical | Chlordiazepoxide | IP |
| Cadmium sulfide | Tattoo | Chlorprothiazide | IP |
| Chlorothiazides | Oral | Chlorpromazine | Topical |
| Coal tar (multicomponent) | Topical | Demeclocycline | IP |
| Dacarbazine | Infusion | Griseofulvin | IP |
| Disperse blue 35 (anthaquinone-base dye) | Topical | Kynuremic acid | Oral |
| Nalidixic acid | Oral | Nalidixic acid | Oral |
| Padimate A or Escolol 506 (amyl-*p*-dimethylamino benzoate) | Topical | Prochlorperazine | IP |
| | | Quinokine methanol | IP |
| | | Quinolone (antibacterial) | Oral |
| Psoralens | Oral, topical | Tetracyclines | IP, topical |
| Quinolone (antibacterial) | Oral | Xanthotoxin (8-methoxypsoralen) | Oral, IP, IM |
| Tetracyclines | Oral | | |
| Xanthotoxin (8-methoxypsoralen) | Topical, oral | | |

$20\,\mathrm{W\,m^{-2}}$ at the exposure site, then the time of exposure required to obtain the target dose of light (in joules) is calculated as

$$\text{Time of exposure} = \frac{\mathrm{Ws}}{\mathrm{J}}\,\frac{9\,\mathrm{J}}{\mathrm{cm}^2}\,\frac{\mathrm{m}^2}{20\,\mathrm{W}}\,\frac{10^4\,\mathrm{cm}^2}{\mathrm{m}^2}\,\frac{\mathrm{min}}{60\,\mathrm{s}} = 75\,\mathrm{min}$$

If, with the same set of lights, two weeks later the irradiation is determined to be $19\,\mathrm{W\,m^{-2}}$, then the exposure period would have to be 79 min.

There are four recommended protocols for assessing topical phototoxicity potential—3T3 cells, rabbit, guinea pig, and mouse.

The first is an in vitro alternative success story—the 3T3 NRU (neutral red uptake) phototoxicity test: uses the BALB/c 3T3 (murine) cell line with cytotoxicity determination based on neutral red uptake to measure cell viability. While not a direct replacement alternative (as there is no in vivo equivalent test), it is an accepted screen for phototoxicity potential by the FDA.

The next test uses the rabbit. The traditional methodology for a predictive test for phototoxicity has been an intact rabbit test (Marzulli and Maibach, 1970). This test is conducted as follows (and illustrated diagrammatically in Figure 2):

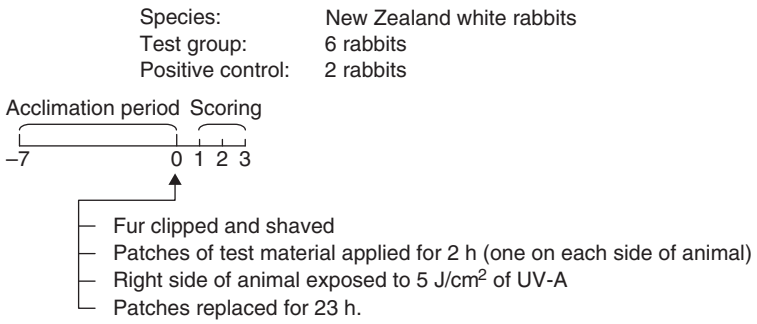


***Figure 2*** *Line chart for design and conduct of phototoxicity assay using rabbits.*

A. *Animals and Animal Husbandry*
   1. Strain/species: Female New Zealand white rabbits
   2. Number: Six rabbits per test; two rabbits for positive control
   3. Age: Young adult
   4. Acclimation period: At least seven days prior to study
   5. Food and water: Freely available commercial laboratory feed and water

B. *Test Article*
   1. A dose of 0.5 mL of liquid or 500 mg of a solid or semisolid will be applied to each test site.
   2. Liquid substances will be used undiluted.
   3. For solids, the test article will be moistened with water (500 mg test article/0.5 mL water or another suitable vehicle) to ensure good contact with the skin.
   4. The positive-control material will be a lotion containing 1% 8-methoxypsoralen.

C. *Experimental Procedures*
   1. Animals will be weighed on the first day of dosing.
   2. On the day prior to dosing, the fur of the test animals will be clipped from the dorsal area of the trunk using a small animal clipper, then shaved clean with a finer bladed clipper.
   3. On the day of dosing, the animals will be placed in restraints.
   4. One pair of patches (approximately 2.5 × 2.5 cm) per test article will be applied to the skin of the back, with one patch on each side of the backbone.
   5. A maximum of two pairs of patches may be applied to each animal and the patches must be at least 2 in. apart.
   6. The patches will be held in contact with the skin by means of an occlusive dressing for the 2-h exposure period.

7. After the 2-h exposure period, the occlusive dressing as well as the patches on the right side of the animal will be removed (aluminum foil).
8. The left side of the animal will be covered with opaque material.
9. The animal will then be exposed to approximately 5 J $cm^{-2}$ of UVA (320–400 nm).
10. After exposure to the UVA light, the patches on the right side of the animal as well as the occlusive dressing will be replaced.
11. The dressing will again be removed approximately 23 h after the initial application of the test article. Residual test article will be carefully removed, where applicable, using water (or another suitable vehicle).
12. Animals will be examined for signs of erythema and edema and the responses scored at 24, 48, and 72 h after the initial test article application according to the Draize reaction grading system previously presented in this volume.
13. Any unusual observation and mortality will be recorded.

D. *Analysis of Data* The data from the irradiated and nonirradiated sites are evaluated separately. The scores from erythema and eschar formation and edema at 24, 48, and 72 h, are added for each animal (six values). The six values are then divided by 3, yielding six individual scores. The mean of the six individual animal irritation scores represents the mean primary irritation score (maximum score = 8, as in the primary dermal irritation study). This method was developed after a human model had been developed.

## 8.4 Guinea Pig

Recently, a standardized protocol for using the guinea pig for phototoxicity testing has been proposed (Nilsson et al., 1993) and has been the subject of an international validation exercise. This is detailed in Figure 3.

A. Animals and animal husbandry
   1. Strain/species: Male Hartley guinea pigs
   2. Number: At least 10 (two groups)
      Irradiation control: 4 animals
      Test material treated: 6 animals
   3. Age: Young adult, 300–500 g
   4. Acclimation period: At least five days
   5. Feed/water: Ad libitum
B. Test material
   1. Vehicle: Test assumes that material will be in solution. Use the most volatile, nonirritating organic solvent possible, for example, ethanol, acetone, dimethylacetamide, or some combination.
   2. Treatment: There can be up to four sites per animal, each measuring 1.5 × 1.5 cm (2.25 $cm^2$). In general, one side should be for a vehicle

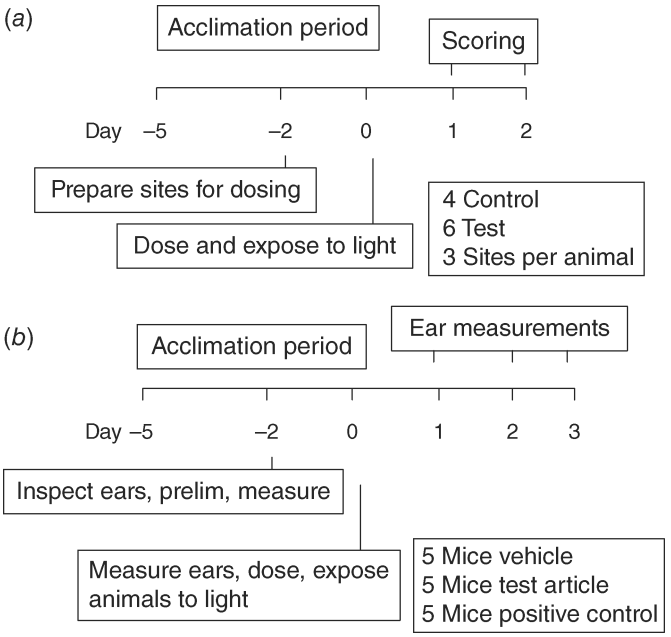


***Figure 3*** *(a) Guinea pig and (b) mouse for phototoxicity testing.*

control and another for a positive control [8-methoxypsoralen (8-MOP), 0.005% in ethanol].

3. Dosage: A dose of 0.025–0.050 mL is applied using a micropipette to each site.

C. Experimental procedure

1. Animals will be weighed on the first day of dosing.
2. Preparation: Approximately 48 h prior to treatment, the hair is removed from a 6 × 8-cm area on the back with a fine clipper. On the day of dosing, animals are dosed as described previously. Tests are situated to prevent mixing of test solutions after application. No patches or wraps are used.
3. Immediately after the dose application, the animals are placed in a restraint while keeping the test sites uncovered. Prior to irradiation the heads are covered to prevent ocular damage from light exposure.
4. Thirty minutes after dosing, animals are exposed to a nonerythmogenic dose of light in the UVA band (should have peak intensity between 335 and 365 nm). The dose of light should be 9 or 10 J $cm^{-2}$ for UVA and 0.1–0.3 J/$cm^{-2}$ for UVB.
5. Immediately after light exposure, the animals are wiped clean if necessary and returned to their home cages.

6. Animals are inspected and scored at 24 and 48 h postexposure according to the following:
   0: No reaction
   1: Slight erythema
   2: Moderate erythema
   3: Severe erythema, with or without edema

   The reader should note that this scoring scheme is the same one used for dermal sensitization scoring, whereas the scoring method for the rabbit model discussed previously is that used for dermal irritation studies.
7. Any unusual clinical signs noted during exposure should be noted. The following descriptive parameters can be calculated from the data.

$$\text{Phototoxic irritation index (PTII)} = \frac{\text{number of positive sites} \times 100}{\text{number of exposure sites}}$$

$$\text{Phototoxicity severity index (PSI)} = \frac{\text{total of scores}}{\text{total of observations}}$$

Lovell and Sanders (1993) had previously proposed a similar model of assessing topical phototoxicity potential in the guinea pig. Their model differed from the one proposed by Nilsson et al. (1993) with regard to the following:

- Only one test site per animal was used.
- Test sites were smaller (about 1.6 $cm^2$).
- Amounts applied were less (about 10 μL).
- Light intensity was set at 15 J $cm^{-2}$.
- Their paper made no reference to the use of a restrainer.
- Assessments were conducted at 4, 24, 48, and 72 h.

The scoring system was as follows:

0: Normal
2: Faint/trace erythema
4: Slight erythema
6: Definite erythema
8: Well-developed erythema

(Intermediate scores were indicated by odd numbers.) They recommended the use of acidine (weak phototoxin) or anthracene (strong phototoxin) for positive controls.

## 8.5 Pyrogenicity

The USP describes a pyrogen test using rabbits as a model (1995b). This test, which is the standard for limiting risks of a febrile reaction to an acceptable level, involves measuring the rise in body temperature in a group of three rabbits for 3 h after injection of 10 mL of test solution.

1. *Apparatus and Diluents* Render the syringes, needles, and glassware free of pyrogens by heating at 250 °F for not less than 30 min or by any other suitable method. Treat all diluents and solutions for washing and rinsing of devices or parenteral injection assemblies in a manner that will ensure that they are sterile and pyrogen free. Periodically perform control pyrogen tests on representative portions of the diluents and solutions that are used for washing or rinsing of the apparatus.

2. *Temperature Recording* Use an accurate temperature-sensing device, such as a clinical thermometer or thermistor or similar probe, that has been calibrated to ensure an accuracy of ±0.1 ° and has been tested to determine that a maximum reading is reached in less than 5 min. Insert the temperature-sensing probe into the rectum of the test rabbit to a depth of not less than 7.5 cm and, after a period of time not less than that previously determined as sufficient, record the rabbit's temperature.

3. *Test Animals* Use healthy, mature rabbits. House the rabbits individually in an area of uniform temperature (between 20 and 23 °C) free from disturbances likely to excite them. The temperature should vary no more than ±3 °C from the selected temperature. Before using a rabbit for the first time in a pyrogen test, condition it for not more than seven days before use by a sham test that includes all of the steps as directed in item 4 below, except injection. Do not use a rabbit for pyrogen testing more frequently than once every 48 h or prior to two weeks following a maximum rise in its temperature of 0.6 °C or more while being subjected to the pyrogen test or following its having been given a test specimen that was adjusted to be pyrogenic.

4. *Procedure* Perform the test in a separate area designated solely for pyrogen testing and under environmental conditions similar to those under which the animals are housed. Withhold all food from the test rabbits during the period of the test. Access to water is allowed at all times but may be restricted during the test. If probes measuring rectal temperature remain inserted throughout the testing period, restrain the rabbits with loose-fitting Elizabethan collars that allow the rabbits to assume a natural resting posture. Not more than 30 min prior to the injection of the test dose, determine the "control temperature" of each rabbit; this is the base for the determination of any temperature increase resulting from the injection of a test solution. In any one group of test rabbits, use only those rabbits whose control temperatures do not vary by more than 1 °C from each other and do not use any rabbit having a temperature exceeding 39.8 °C.

Unless otherwise specified in the individual protocol, inject 10 mL of the test solution per kilogram of body weight into an ear vein of each of three rabbits, completing each injection within 10 min after the start of administration. The test solution is either the product, constituted if necessary as directed in the labeling, or the material under test. For pyrogen testing of devices or injection assemblies, use washings or rinsings of the surfaces that come in contact with the parenterally administered material or with the injection site or internal tissues of the patient. Ensure that all test solutions are protected from contamination. Perform the injection after warming the test solution to a temperature of 37 ± 2 °C. Record the temperature at 1, 2, and 3 h subsequent to the injection.

5. *Test Interpretation and Continuation* Consider any temperature decreases as zero rise. If no rabbit shows an individual rise in temperature of 0.6 °C or more above its respective control temperature and if the sum of the three individual maximum temperature rises does not exceed 1.4 °C, the product meets the requirements for the absence of pyrogens. If any rabbit shows an individual temperature rise of 0.6 °C or more or if the sum of the three individual maximum temperature rises exceeds 1.4 °C, continue the test using five other rabbits. If not more than three of the eight rabbits show individual rises in temperature of 0.6 °C or more and if the sum of the eight individual maximum temperature rises does not exceed 3.7 °C, the material under examination meets the requirements for the absence of pyrogens.

***In Vitro Pyrogenicity*** In vitro pyrogenicity testing (or bacterial endotoxin testing) is one of the great success stories for in vitro testing. Some 15 years ago, the limulus amebocyte lysate (LAL) test was developed, validated, and accepted as an in vitro alternative (Cooper, 1975; Wearly and Baker, 1977) to the rabbit test. An in vitro test for estimating the concentration of bacterial endotoxins that may be present in or on a sample of the article(s) to which the test is applied uses LAL that has been obtained from aqueous extracts of the circulating amebocytes of the horseshoe crab, *Limulus polyphemus*, and that has been prepared and characterized for use as an LAL reagent for gel-clot formation. The test's limitation is that it detects only the pyrogens of gram-negative bacteria. This is generally not significant, since most environmental contaminants that gain entrance to sterile products are gram negative (Develeeshouwer et al., 1985).

Where the test is conducted as a limit test, the specimen is determined to be positive or negative to the test judged against the endotoxin concentration specified in the individual monograph. Where the test is conducted as an assay of the concentration of endotoxin, with calculation of confidence limits of the result obtained, the specimen is judged to comply with the requirements if the result does not exceed (1) the concentration limit specified in the individual monograph and (2) the specified confidence limits for the assay. In either case the determination of the reaction endpoint is made with parallel dilutions of redefined endotoxin units.

Since LAL reagents have also been formulated to be used for turbidimetric (including kinetic) assays or colorimetric readings, such tests may be used if shown to comply with the requirements for alternative methods. These tests require the establishment of a standard regression curve and the endotoxin content of the test material is determined by interpolation from the curve. The procedure includes incubation for a preselected time of reacting endotoxin and control solutions with LAL reagent and reading the spectrophotometric light absorbance at suitable wavelengths. In the case of the turbidimetric procedure the reading is made immediately at the end of the incubation period. In the kinetic assays, the absorbance is measured throughout the reaction period and rate values are determined from those readings. In the colorimetric procedure the reaction is arrested at the end of the preselected time by the addition of an appropriate amount of acetic acid solution prior to the readings. A possible advantage in the mathematical treatment of results, if the test is otherwise validated and the assay suitably designed, could be the confidence interval and limits of potency from the internal evidence of each assay itself.

## 8.6 Blood Compatibility

The standard test (and its major modifications) currently used for this purpose is technically an in vitro one, but it requires a sample of fresh blood from a dog or other large donor animal. The test was originally developed by the National Cancer Institute for use in evaluating cancer chemotherapeutic agents (Prieur et al., 1973) and is rather crude, though definitive.

The variation described here is one commonly utilized. It uses human blood from volunteers, eliminating the need to keep a donor colony of dogs. The test procedure is described below.

1. *Test System* Human blood. Collect 30 mL heparinized blood for whole blood and plasma (three tubes) and 30 mL clotted blood for serum (two tubes) from each of six donors.
2. *Precipitation Potential*
   - 2.1. For each donor, set up and label eight tubes 1–8.
   - 2.2. Add 1 mL serum to tubes 1–4.
   - 2.3. Add 1 mL plasma to tubes 5–8.
   - 2.4. Add 1 mL formulation to tubes 1–5.
   - 2.5. Add 1 mL vehicle to tubes 2 and 6.
   - 2.6. Add 1 mL physiological saline to tubes 3 and 7 (negative control).
   - 2.7. Add 1 mL 2% nitric acid to tubes 4 and 8 (positive control).
   - 2.8. Observe tubes 1–8 for qualitative reactions (e.g., precipitation or clotting) before and after mixing.
   - 2.9. If a reaction is observed in the formulation tubes (1 and/or 5), dilute the formulation in an equal amount of physiological saline (½ dilution) and test 1 mL of the dilution with an equal amount of plasma

and/or serum. If a reaction still occurs, make serial dilutions of the formulation in saline (i.e., ¼, ⅛, etc.).

2.10. If a reaction occurs in the vehicle tubes (2 and/or 6), repeat in a manner similar to that in step 2.9.

3. *Hemolytic Potential*
   3.1. For each donor, set up and label eight tubes 1–8.
   3.2. Add 1 mL whole blood to each tube.
   3.3. Add 1 mL formulation to tube 1.
   3.4. Add 1 mL vehicle to tube 2.
   3.5. Add 1 mL of ½ dilution of formulation in saline to tube 3.
   3.6. Add 1 mL of ½ dilution of vehicle in saline to tube 4.
   3.7. Add 1 mL of ¼ dilution of formulation in saline to tube 6.
   3.8. Add 1 mL of ¼ dilution of vehicle in saline to tube 6.
   3.9. Add 1 mL of physiological saline to tube 7 (negative control).
   3.10. Add 1 mL of distilled water to tube 8 (positive control).
   3.11. Mix by gently inverting each tube three times.
   3.12. Incubate tubes for 45 min at 37 °C.
   3.13. Centrifuge 5 min at 1000 *g*.
   3.14. Separate the supernatant from the sediment.
   3.15. Determine hemoglobin concentrations to the nearest 0.1 g $dL^{-1}$ on the supernatant (plasma).
   3.16. If hemoglobin concentrations of the above dilutions are 0.2 g $dL^{-1}$ (or more) greater than the saline control, repeat the procedure, adding 1 mL of further serial dilutions (⅛, 1/16, etc.) of formulation or vehicle to 1 mL of blood until the hemoglobin level is within 0.2 g $dL^{-1}$ of the saline control.

There are two proposed, true in vitro alternatives to this procedure (Mason et al., 1974; Kambic et al., 1976), but neither has been either widely evaluated or accepted.

## REFERENCES

Anderson, R. R., and Parrish, J. A. (1981). The optics of skin. *J. Invest. Dermatol.* 77:13–19.

Auletta, C. (1994). Vaginal and rectal administration. *J. Am. Coll. Toxicol.* 13:48–63.

Avis, K. E. (1985). Parenteral preparations. In A. R. Gennaro (Ed.), *Remington's Pharmaceutical Sciences*. Mack Publishing Company, Easton, PA, pp. 1518–1541.

Ballard, B. E. (1968). Biopharmaceutical considerations in subcutaneous and intramuscular drug administration. *J. Pharm. Sci.* 57:357–378.

Brown, S., Templeton, D., Prater, D. A., and Potter, C. C. (1989). Use of an in vitro hemolysis test to predict tissue irritancy in an intramuscular formation. *J. Parenter. Sci. Technol.* 43:117–120.

Cooper, J. F. (1975). Principles and applications of the limulus test for pyrogen in parenteral drugs. *Bull. Parenter. Drug Assoc.* 3:122–130.

Develeeshouwer, M. J., Cornil, M. F., and Dony, J. (1985). Studies on the sensitivity and specificity of the limulus amebocyte lysate test and rabbit pyrogen assays. *Appl. Environ. Microbiol.* 50:1509–1511.

Draize, J. H., Woodard, G., and Calvery, H. O. (1944). Method for the study of irritation and toxicity of substances applied topically to the skin and mucous membranes. *J. Reprod. Fert.* 20:85–93.

Eckstein, P., Jackson, M., Millman, N., and Sobero, A. (1969). Comparison of vaginal tolerance test of spermical preparations in rabbits and monkeys. *J. Reprod. Fertil.* 20:85–93.

Gad, S. C., and Chengelis, C. P. (1998). *Acute Toxicity: Principles and Methods*, 2nd ed. Academic, San Diego, CA.

Gad, S. C., Walsh, R. D., and Dunn, B. J. (1986). Correlation of ocular and dermal irritancy of industrial chemicals. *Ocul. Dermal Toxicol.* 5(3):195–213.

Garramone, J. P. (1986). *Vascular Access Port Model SLA. Users Manual*. Norfolk Medical Products, Skokie, IL.

Gray, J. E. (1978). Pathological evaluation of injection injury. In J. Robinson (Ed.), *Sustained and Controlled Release Drug Delivery Systems*. Marcel Dekker, New York, pp. 351–405.

Guillot, J. P., Gonnet, J. F., Clement, C., Caillard, L., and Trahaut, R. (1982). Evaluation of the cutaneous-irritation potential of 56 compounds. *Food Chem. Toxicol.* 201:563–572.

Hagan, E. C. (1959). *Appraisal of the Safety of Chemicals in Foods, Drugs and Cosmetics*. Association of Food and Drug Officials of the United States, Austin, TX, p. 19.

Horio, T., Kohachi, K., Ogawa, A., Inoue, T., and Ishirara, M. (1995). Evaluation of photosensitizing ability of quinolones in guinea pigs. *Drugs* 49(Suppl. 2):283–285.

Kambic, H. E., Kiraly, R. J., and Yukihiko, N. (1976). A simple in vitro screening test for blood compatibility of materials. *J. Biomed. Mater. Res. Symp.* 7:561–570.

Kato, I., Harihara, A., and Mizushima, Y. (1992). An in vitro method for assessing muscle irritation of antibiotics using rat primary cultured skeletal muscle fibers. *Toxicol. Appl. Pharmacol.* 117:194–199.

Lambert, L., Warmer, W., and Kornhauser, A. (1996). Animal models for phototoxicity testing. *Toxicol. Methods* 2:99–114.

Lovell, W. W., and Sanders, D. J. (1992). Phototoxicity testing in guinea pigs. *Food Chem. Toxicol.* 30:155–160.

Marzulli, F. M., and Maibach, H. K. (1970). Perfume phototoxicity. *J. Soc. Cosmet. Chem.* 21:685–715.

Mason, R. G., Shermer, R. W., Zucker, W. H., Elston, R. C., and Blackwelder, W. C. (1974). An in vitro test system for estimation of blood compatibility of biomaterials. *J. Biomed. Mater. Res.* 8:341–356.

Mathias, C. G. T. (1983). Clinical and experimental aspects of cutaneous irritation. In F. M. Margulli and H. T. Maiback (Eds.), *Dermatotoxicology*. Hemisphere Publishing, New York, pp. 167–183.

Nilsson, R., Maurer, T., and Redmond, N. (1993). A standard protocol for phototoxicity testing: Results from an interlaboratory study. *Contact Dermatitis* 28:285–290.

Pathak, M. A. (1967). Photobiology of melanogenesis: Biophysical aspects. In W. Montagna and F. Uh (Eds.), *Advances in Biology of Skin,* Vol. 8: *The Pigmentary System*. Pergamon, New York, pp. 400–419.

Prieur, D. J., Young, D. M., Davis, R. D., Cooney, D. A., Homan, E. R., Dixon, R. L., and Guarino, A. M. (1973). Procedures for preclinical toxicologic evaluation of cancer chemotherapeutic agents: Protocols of the laboratory of toxicology. *Cancer Chemother. Repts.* Part 3, 4:1–30.

Shintani, S., Yamazaki, M., Nakamura, M., and Nakayama, I. (1967). A new method to determine the irritation of drugs after intramuscular injection in rabbits. *Tox. Appl. Pharm.* 11:293–301.

Sidell, F. R., Calver, D. L., and Kaminskis, A. (1974). Serum creatine phosphokinase activity after intramuscular injection. *JAMA* 228:1884–1887.

Theeuwes, F., and Yum, S. I. (1976). Principles of the design and operation of generic osmotic pumps for the delivery of semisolid or liquid drug formulations. *Ann. Biomed. Eng.* 4:343–353.

U.S. Pharmacopeia (USP) (1985). Intramuscular irritation test. *United States Pharmacopeia*. USP Convention, Rockville, MD, pp. 1180–1183.

U.S. Pharmacopeia (USP) (1995a). Intracutaneous test. *United States Pharmacopeia*. USP Convention, Rockville, MD, pp. 1201–1202.

U.S. Pharmacopeia (USP) (1995b). Bacterial endotoxins test. *United States Pharmacopeia*. USP Convention, Rockville, MD, pp. 1696–1697.

Wearly, M., and Baker, B. (1977). Utilization of the limulus amebocyte lysage test for pyrogen testing of large-volume parenterals, administration sets and medical devices. *Bull. Parenter. Drug Assoc.* 31:127–133.

Weil, C. S., and Scala, R. A. (1971). Study of intra- and interlaboratory variability in the results of rabbit eye and skin irritation tests. *Toxicol. Appl. Pharmacol.* 19:276–360.

Williams, P. D., Masters, B. G., Evans, L. D., Laska, D. A., and Hattendorf, G. H. (1987). An in vitro model for assessing muscle irritation due to parenteral antibiotics. *Fund. Appl. Toxicol.* 9:10–17.

Young, M. F., Tobretta, L. D., and Sophia, J. V. (1986). Correlative in vitro and in vitro study of skeletal muscle irritancy. *Toxicologist* 6(1):1225.

Zhai, H., Wilhelm, K.-P., and Maibach, H. I. (2008). *Dermatoxicology*, 7th ed. CRC Press, Boca Raton, FL.

# *Pharmacokinetics and Toxicokinetics in Drug Safety Evaluation*

## 1 INTRODUCTION

Among the cardinal principles of both toxicology and pharmacology is that the means by which an agent comes in contact with or enters the body (i.e., the route of exposure or administration) does much to determine the nature and magnitude of its effects (Goldstein et al., 1974; Pratt and Taylor, 1990). Accordingly, an understanding of route(s) of administration and their implications for absorption is essential. The therapeutic index (calculated based on plasma and therefore absorbed levels) is the ratio between what levels cause adverse effects and the levels which have the desired therapeutic effect.

Safety assessment studies usually involve a control group of animals (untreated and/or dosed with formulation only) and at least three treated groups receiving "low," "intermediate," and "high" dose levels of the drug entity of interest via a route approximately that used in humans (as closely as possible). Frequently there will also be recovery groups to determine if any observed effects are reversible (and if so, to what extent). In most instances the high dose level is expected to elicit some toxic effects in the animals, often expressed as decreased food consumption and/or below-normal body weight gain, and has been selected after consideration of earlier data, perhaps from dose range-finding studies, or at least to a dose as high as possible by the intended route. The other two dose levels are anticipated not to cause toxic effects. Generally, but not always (e.g., nonsteroidal anti-inflammatory drugs

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

in rodents), the low dose level is a several-fold multiple of the expected human therapeutic or exposure level (generally more so in rodents than in nonrodents). However, without knowing the true relationship of these dose levels to each other with respect to the absorption, distribution, and elimination of the new chemical entity as reflected by its pharmacokinetics, it is difficult to see how meaningful extrapolations concerning safety margins can be made from the toxicity data obtained. Also, without pharmacokinetic data from the positive-control group, its inclusion is of limited value and the results obtained could lead to erroneous conclusions (Yacobi et al., 1989).

Pharmacokinetic studies can provide information on several aspects, knowledge of which greatly facilitates assessment of safety of the chemical entity. Six such aspects can be mentioned:

(i) Relationship between the dose levels used and the relative extent of absorption of the test compound
(ii) Relationship between the protein binding of the test compound and the dose levels used
(iii) Relationship between pharmacological or toxicological effects and the kinetics of the test compound
(iv) Effect of repeated doses on the kinetics of the test compound
(v) Relationship between the age of the animal and the kinetics of the test compound
(vi) Relationship between the dose regimens of the test compound used in the toxicity studies and those employed clinically in humans

International Conference on Harmonisation (ICH) guidelines (ICH, 2000a,b) dictate a clearly defined set of objectives for toxicokinetic studies:

- *Primary* To describe the systemic exposure achieved in animals and its relationship to dose level and the time course of toxicity studies.
- *Secondary*
  1. To relate the exposure achieved in toxicity studies to toxicological findings and contribute to the assessment of the relevance of these findings to clinical safety.
  2. To support the choice of species and treatment regimen in nonclinical toxicity studies.
  3. To provide information which, in conjunction with the toxicity findings, contributes to the design of subsequent nonclinical toxicity studies.

These data may be obtained from all animals on a toxicity study, from representative subgroups, from satellite groups, or from separate studies.

If toxicology can be described as being the study of the effects of a chemical on an organism, metabolism can be described as the opposite—the effects of the organism on the chemical. Metabolism refers to a process by which a drug

(xenobiotic) is chemically modified by an organism. It is part of the overall process of disposition of xenobiotic (ADME)—the process by which a chemical gains access to the inner working of an organism (absorption), how it moves around inside an organism (distribution), how it is changed by the organism (metabolism), and how it is eventually eliminated from the organism (elimination). The U.S. Environmental Protection Agency (EPA) definition of biotransformation or metabolism is "the sum of processes by which a xenobiotic (foreign chemical) is handled by a living organism." The mathematical formulae used to describe and quantify these processes are collectively known as pharmacokinetics. The EPA definition of pharmacokinetics is the "quantitation and determination of the time course and dose dependency of the absorption, distribution, biotransformation and excretion of chemicals." The acronym ADME (for absorption, distribution, metabolism, and excretion) has been used to describe the multifaceted biological process. The term metabolism has also come into common jargon to describe the entire process. This science has long played a central role in pharmaceutical development but has played less of a role in the development of other types of products. The purpose of this chapter is to introduce the basic concepts of ADME and practices of studies conducted to study it as described in regulations of the EPA and Organisation for Economic Co-operation and Development (OECD), which require such data for nonpharmaceutical products and to give some real-world examples.

## 2 REGULATIONS

The FDA believes that data from studies on the ADME of a chemical provide insight into mechanisms of toxicity and are essential in the design and evaluation of results from other toxicity studies (FDA, 2005). Such data should be provided for all drugs and significant impurities. Recommendations for obtaining data on the metabolism and pharmacokinetics of these substances are presented in ICH guidelines and the FDA *Redbook II* (2000). In general, it is required that this information be obtained as part of initial and subsequent repeat-dose studies with a drug. The European Medicines Agency (EMEA) has promulgated separate guidelines (EMEA, 2006) for evaluating the pharmacokinetics of protein therapeutics.

## 3 PRINCIPLES

An understanding of the design and analysis of pharmacokinetic studies requires a broad understanding of the underlying concepts and principles inherent in the ADME process and in our current technology for studying such. Each of these four principal areas is overviewed from a practical basis as it relates to toxicology. First, however, one should consider the fundamental terminology used in pharmacokinetic studies (Table 1).

**TABLE 1 Fundamental Terms Used in Pharmacokinetic Studies**

| | |
|---|---|
| Absolute bioavailability | Bioavailability of dosage form relative to intravenous administration |
| Absorption | Process by which xenobiotic and its metabolites are transferred from site of absorption to blood circulation |
| Accumulation | Progressive increase of chemical and/or metabolites in body. Accumulation is influenced by the dosing interval and half-life of the chemical. The process can be characterized by an "accumulation factor," which is the ratio of the plasma concentration at steady state to that following the first dose in a multiple dosing regimen. |
| Analyte | Drug entity assayed in biological samples |
| Area under the curve (AUC) | Concentration of chemical and/or metabolites in blood (or plasma/serum) integrated over time. This is typically considered the best indicator of exposure. |
| Bioavailability | Rate and extent to which xenobiotic entity enters systemic circulation intact following oral or dermal administration. It is sometimes expanded to include therapeutically active metabolites. Also known as comparative bioavailability. |
| Biotransformation | Process by which xenobiotic is structurally and/or chemically changed in body by either enzymatic or nonenzymatic reactions. The product of the reaction is a different composition of matter or different configuration than the original compound. |
| Clearance | Volume of biological fluid which is totally cleared of xenobiotic in unit time |
| $C_{max}$ | Maximum mean concentration of chemical in plasma. Also known as peak plasma concentration. |
| Concomitant toxicokinetics | Toxicokinetic measurements performed in toxicity study either in all animals or in representative subgroups or in satellite groups |
| Disposition | All processes and factors involved from time chemical enters body to time when it is eliminated from body either intact or in metabolite form |
| Distribution | Process by which absorbed xenobiotic and/or its metabolites partition between blood and various tissues/organs in body |
| Dosage form | Formulation (diet, lotion, capsule, solution, etc.) administered to animals or humans |
| Dose proportionality | Relationship between doses of chemical and measured parameters, usually including tests for linearity |
| Enterohepatic circulation | Process by which xenobiotics are emptied via bile into small intestine and then reabsorbed into hepatic circulation |
| Enzyme induction | Increase in enzyme content (activity and/or amount) due to xenobiotic challenge which may result in more rapid metabolism of chemical |
| Enzyme inhibition | Decrease in enzymatic activity due to effect of xenobiotic challenge |
| Excretion | Process by which administered compound and/or its biotransformation product(s) are eliminated from body |

**TABLE 1** ***Continued***

| Absolute bioavailability | Bioavailability of dosage form relative to intravenous administration |
|---|---|
| Exposure | Exposure is represented by pharmacokinetic parameters demonstrating the local and systemic burden on the test species with the test compound and/or its metabolites. The area under the matrix level concentration–time curve (AUC) and/or the measurements of matrix concentrations at the expected peak-concentration time $C_{max}$ or at some other selected time $C_{(time)}$ are the most commonly used parameters. Other parameters might be more appropriate in particular cases. |
| First-order kinetics | Kinetic processes the rate of which is proportional to concentration |
| First-pass effect | Phenomenon whereby xenobiotics may be extracted or metabolized following enteral absorption before reaching systemic circulation |
| Flux | Term (that takes area into consideration) used to describe movement of chemical across a barrier. Most typically used to describe the absorption of a chemical across the skin as micrograms per square centimeter per hour. |
| Half-life | Time elapsed for given chemical entity concentration or amount to be reduced by factor of 2 |
| Hepatic clearance | Rate of total body clearance accounted for by liver |
| $K_{el}$ | Elimination constant for chemical in plasma. Typically calculated using the formula $K_{el} = -1n[10] \times b$, where $b$ is the slope of the linear regression line of the log of the mean plasma concentrations vs. time from $t_{max}$ to 24 h. |
| Lag time | Interval between compound administration and when compound concentration is measurable in blood |
| Metabolite characterization | Determination of physiochemical characteristics of biotransformation product(s) |
| Metabolite identification | Structural elucidation of biotransformation product(s) |
| Metabolite profile | Chromatographic pattern and/or aqueous/nonaqueous partitioning of biotransformation products of administered compound |
| Monitor | To take a small number of matrix samples (e.g., 1–3) during dosing interval to estimate $C_{(time)}$ and/or $C_{max}$. |
| Nonlinear kinetics (saturation kinetics) | Kinetic processes the rate of which is not directly proportional to concentration |
| Presystemic elimination | Loss of that portion of dose that is not bioavailable. This would include, among others, loss through intestinal and gut-wall metabolism, lack of absorption, and first-pass hepatic metabolism. |
| Profile | To take (e.g., 4–8) matrix samples during dosing interval to estimate $C_{max}$ and/or $C_{(time)}$ and area under matrix concentration–time curve (AUC) |
| Protein binding | Complexation of xenobiotic and/or its metabolite(s) with plasma or tissue proteins |
| Relative bioavailability | Bioavailability relative to reference or standard formulation or agent |

**TABLE 1** ***Continued***

| Absolute bioavailability | Bioavailability of dosage form relative to intravenous administration |
|---|---|
| Renal clearance | Rate of total body clearance accounted for by kidney. Its magnitude is determined by the net effects of glomerular filtration, tubular secretion and reabsorption, renal blood flow, and protein binding. |
| Satellite | Groups of animals included in design and conduct of toxicity study treated and housed under conditions identical to those of main study animals but used primarily for toxicokinetics |
| Steady state | Equilibrium state where rate of chemical input is equal to rate of elimination during given dose interval |
| Support | In context of toxicity study to ratify or confirm design of toxicity study with respect to pharmacokinetic and metabolic principles. This process may include two steps:<br>a. Confirmation using toxicokinetic principles that the animals on a study were exposed to appropriate systemic levels of the administered compound and/or its metabolite(s).<br>b. Confirmation that the metabolic profile in the species used was acceptable; data to support this will normally be derived from metabolism studies in animals and humans. |
| $T_{max}$ | Sampling time point at which $C_{max}$ occurs. |
| Total clearance | Volume of biological fluid totally cleared of xenobiotic per unit time and usually includes hepatic clearance and renal clearance |
| Toxicokinetics | Study of kinetics of absorption, distribution, metabolism, and excretion of toxic or potentially toxic chemicals |
| Validate | In context of analytical method to establish accuracy, precision, reproducibility, response function, and specificity of analytical method with reference to biological matrix to be examined and analyte to be quantified |
| Volume of distribution ($V_d$) | Hypothetical volume of body fluid into which chemical distributes. It is not a "real" volume but is a proportionality constant relating the amount of chemical in the body to the measured concentration in blood or plasma. |

## 3.1 Absorption

Absorption describes the process by which a chemical crosses a biological membrane to gain access to the inner workings of an organism. For mammals, this process results in the entry of the chemical into the bloodstream, or systemic circulation. In this case the process is also called systemic absorption. Pharmaceutical products, procedures, and devices, such as hypodermic needles or catheters, can be used to by-pass biological barriers. Other products gain access to the systemic circulation via the oral, dermal, buccal, or inhalatory route of administration.

For a material to be toxic (local tissue effects are largely not true toxicities by this definition), the first requirement is that it be absorbed into the organism [for which purpose being in the cavity of the gastrointestinal (GI) tract does not qualify]. Most pharmaceuticals are intended to gain such access.

There are characteristics which influence absorption by the different routes, and these need to be understood by any person trying to evaluate and/or predict the toxicities of different moieties. Some key characteristics and considerations are summarized below by route:

A. Oral and rectal routes (gastrointestinal tract)
   1. Lipid-soluble compounds (nonionized) are more readily absorbed than water-soluble compounds (ionized).
      a. Weak organic bases are in the nonionized, lipid-soluble form in the intestine and tend to be absorbed there.
      b. Weak organic acids are in the nonionized, lipid-soluble form in the stomach and one would suspect they would be absorbed there, but the intestine is more important because of time and area of exposure.
   2. Specialized transport systems exist for some moieties: sugars, amino acids, pyrimidines, calcium, and sodium.
   3. Almost everything is absorbed—at least to a small extent (if it has a molecular weight below 10,000).
   4. Digestive fluids may modify the structure of a drug.
   5. Dilution increases toxicity because of more rapid absorption from the intestine, unless stomach contents bind the moiety.
   6. Physical properties are important—for example, dissolution of metallic mercury is essential to allow absorption.
   7. Age—neonates have a poor intestinal barrier.
   8. The effect of fasting on absorption depends on the properties of the chemical of interest.

B. Inhalation (lungs)
   1. Aerosol deposition:
      a. Nasopharyngeal—5 μm or larger in humans, less in common laboratory animals
      b. Tracheobronchiolar—1–5 μm
      c. Alveolar—1 μm
   2. If a solid, mucociliary transport may serve to clear from the lungs to the GI tract.
   3. Lungs are anatomically good for absorption.
      a. Large surface area (50–100 $m^2$)
      b. High blood flow
      c. Close to blood (10 μm between gas media and blood)
   4. Absorption of gases is dependent on solubility of the gas in blood.
      a. Chloroform, for example, has high solubility and is all absorbed; respiration rate is the limiting factor.
      b. Ethylene has low solubility and only a small percentage is absorbed—blood flow limited absorption.
      c. Parenteral routes.
      d. Dermal routes.

As a generalization, there is a pattern of relative absorption rates which extends between the different routes that are commonly employed. This order of absorption (by rate from fastest to slowest and, in a less rigorous manner, in degree to absorption from most to least) is intravenous (IV) > inhalation > intramuscular (IM) > intraperitoneal (IP) > subcutaneous (SC) > oral > intradermal (ID) > other dermal.

Absorption (total amount and rate), distribution, metabolism, and species similarity in response are the reasons for selecting particular routes in toxicology. In acute studies, however, these things are rarely known to us. So the cardinal rule for selecting routes of use in acute testing is to use those routes which mirror the intended route for human exposure. If this route of human exposure is uncertain or if there is the potential for either a number of routes or the human absorption rate and pattern being greater, then the common practice becomes that of the most conservative approach. This approach stresses maximizing potential absorption in the animal species (within the limits of practicality) and selecting from among those routes commonly used in the laboratory that which gets the most material into the animal's system as quickly and completely as possible to evaluate the potential toxicity.

In general, drugs cross biological barriers by one of three mechanisms: active transport, facilitative transport, and passive transport. In active transport, the chemical is specifically recognized by the organism, which then expends energy to take the chemical up, even against a concentration gradient. In facilitative transport, the organism produces a carrier molecule which reacts with the target molecule to form a complex which more easily traverses the membrane, but no energy is expended to take up the complex. Such complexes do not flow against a concentration barrier. The simplest mechanism is passive transfer or diffusion. Here, a drug flows down a concentration gradient (from high concentration to a lower concentration) and must passively (no energy expended by organism) cross a biological membrane. Passive transfer or diffusion is the most common (if not the only) mechanism involved in the absorption of the vast majority of approved drugs. It should be remembered that for purposes here "concentration gradient" must be considered in relationship to partition coefficient. That is, a gradient will reflect the relative solubilities of drug in polar (water) and nonpolar (lipid) matrices or tissues. The other mechanisms involved in absorption will not be further discussed here.

Drugs in solution have a natural tendency (more rigorously defined by the laws of thermodynamics) to move down a concentration gradient. That is to say, the individual molecules of solute tend to move from a region of high concentration toward regions of lower concentration. Also, the movement of a chemical across a permeable barrier, such as a biological membrane, is a process called diffusion, as illustrated by Figure 1. For most products, these biological barriers are the wall of the GI tract, the lining of the pulmonary system, and/or the skin.

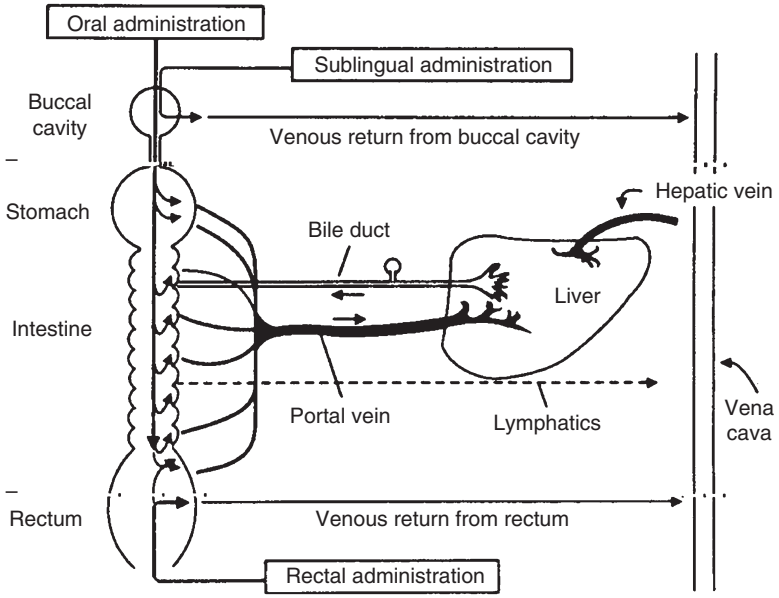


***Figure 1*** *Passage of chemical moieties from GI tract into bloodstream.*

Absorption from the GI tract is controlled by a variety of factors. These include the acid–base characteristics of the chemical (described as the p$K_a$), the solubility, the nature of the delivery (e.g., diet vs. gavage), the nature of any vehicle (suspensions vs. solution or aqueous vs. nonaqueous), and the GI tract of the species under study. Gad (2007b) provides much greater detail on this subject.

Ionized or charged organic moieties do not readily pass through the lipophilic cell membranes of the epithelial cells that line the GI tract. Thus more acidic molecules tend to be more readily absorbed from the stomach while more alkaline materials tend to be absorbed from the small intestine. This is because at the acidic pH of the stomach acidic chemicals tend to be nonionized. More alkaline chemicals tend to be more ionized in the stomach and less ionized in the gut. The equilibrium reaction for acidic dissociation can be represented by the equation

$$\begin{array}{l} \text{X-C-OH} + H_2O == \text{X-C-O} + H_3O^+ \\ \quad\;\; // \qquad\qquad\qquad\quad // \\ \quad\;\; O \qquad\qquad\qquad\quad O \end{array}$$

Like all chemical equations, this one has an equilibrium constant. The discussion of basic chemistry is outside the purview of this book. Readers who may need a refresher are referred to Tse and Jaffe (1991). For every chemical,

**TABLE 2 Receptors Slowing Gastric Emptying**

| Receptor (Site) | Stimulus | Example | Sensitivity |
|---|---|---|---|
| Osmoreceptor (Duodenum) | Osmotic Effect (except tryptophan) | Glucose Amino Acids Electrolytes | Least |
| Acid Receptor (Proximal Duodenum & Jejunum) | Acids with pKa < 5 | Citric Acid HCl | Intermediate |
| Fat Receptor (Jejunum) | Fatty Acids | Sodium Myristate | Most |
| Tryptophan Receptor (Duodenum & Jejunum) | Tryptophan | Tryptophan | — |

Adapted from Hunt & Knox, 1968 and Minami & MacCallum, 1984.

a $pK_a$ can be calculated based on the equilibrium constant, which represents the proportion of ionized and un-ionized material in solution. The lower the $pK_a$ of a chemical the more likely it is to be nonionized in the stomach.

One can manipulate movement of a drug through the GI tract (and particularly the stomach) by formulation and/or by feeding concurrent with drug administration. Some of the factors which slow gastric emptying (and therefore increase drug absorption) are presented in Table 2.

***Absorption from Pulmonary System*** Of the three routes discussed here, absorption from the pulmonary system is perhaps the most rapid. Systemic absorption of inhaled materials is highly dependent on the physical properties of the inhaled materials which dictate how easily the materials reach the alveoli of the deep lung. Proteins may be readily absorbed when instilled in the deep lungs—as an example, note the inhalable insulins. Gases and vapors easily penetrate into the deep lung. For mists and dusts, absorption will be highly dependent on particle size. In general, the larger the particles, the less they will penetrate the pulmonary system. The term impaction describes the deposition of particles in the respiratory tract. Particles of less than 0.2 μm are preferentially deposited in the pulmonary portion of the respiratory system and particles over 2 μm do not reach the alveolar epithelium in great number. Particles from 1 to 4 μm tend to be distributed over the length of the system and particles over 4 μm tend to be deposited in the nasal region. Aerosolized particles of greater than 20 μm do not commonly occur in nature. Tidal volume will also influence impaction. In general, the larger the tidal volume, and thus the more forceful the inhalatory process, the more deeply particles of all sizes tend to be driven into the lung.

Once deposited, materials must be in solution before they can be absorbed. Hence, materials in an aerosolized solution will be more readily absorbed than materials that are delivered as solid (e.g., dusts) particles. Solid materials must be able to go into solution in situ in order to be absorbed. Particle size influences dissolution rate. Large particles dissolve more slowly (for any given material) than small particles due to the differences in surface area. Once in

solution, the same laws of passive diffusion apply to materials in the lung as apply to material in the GI tract. The large surface area and the rich blood flow at the alveoli make for ideal conditions for rapid absorption into the systemic circulation. Absorption across the mucosa lining the upper airways is less rapid. Materials that do not dissolve are ingested by pulmonary macrophages and are either broken down there or moved out of the lungs by the upward movement of the bronchociliary tree.

For gases and vapors, the amount absorbed is highly dependent on the partial pressure of the gas and the solubility of the gas in blood. Take the simple case of a gas that is not metabolized and is excreted by exhalation (e.g., an anesthetic gas or a Halon-type fire-extinguishing agent). At any given concentration (or partial pressure) in the atmosphere, the concentration in the blood will reach a steady state in the blood. Accordingly, prolonged exposure does not lead to continual buildup.

At equilibrium, the concentration in the blood is depicted by the formula (also known as the Ostwald coefficient) $X_b/X_a = S$, where $X_b$ is the concentration in the blood and $X_a$ is the concentration in the inspired air. Thus, if one knows $S$ for a given chemical and the target concentration for a given exposure, one can predict what the resulting concentration may be at equilibrium. Additionally, the lower the $S$ value (i.e., the lower the solubility in blood), the more rapidly the chemical will achieve equilibrium.

***Absorption across Skin*** An aqueous carrier may be used for a variety of dermal products. In fact, carriers can be designed to limit the transportation of the penetration of the active ingredient (such as an insect repellent) if the desired effect is to keep the active ingredient on the surface of the skin. Once again, however, only those materials that are dissolved will be available for penetration across the skin to gain access to the systemic circulation. For almost all drugs in or about to enter clinical trials, dermal penetration is a passive process. The relative thickness of the skin makes absorption (into the systemic circulation) slower than the absorption across the GI or pulmonary barriers. This is compounded by the fact that the stratum corneum function is to be impervious to the environment. One of the skin's major functions is protection from infection. Once a drug penetrates into the dermis, it may partition into the subcutaneous fat. Essentially, absorption across the skin is a two-step process with the first being penetration and deposition into the skin and the second being release from the skin into the systemic circulation. The pattern of blood levels obtained via dermal penetration is generally one with a delayed-absorption, slow buildup to more of a plateau than a peak. Blood levels of chemicals absorbed via the dermal route are generally low (Garner and Matthews, 1998).

Given the overwhelming influence of the physical properties of skin in determining bioavailabilities via the dermal route, assessment of dermal penetration is one area in metabolism and toxicology where the use of in vitro methods can be effectively used to predict in vivo results and to screen chemi-

cals. Apparatus and equipment exist that one can use to maintain sections of skin (obtained from euthanized animals or from human cadavers or surgical discard) for such experiments (Holland et al., 1984; Bronaugh, 1998). These apparatus are set up to maintain the metabolic integrity of the skin sample between two reservoirs: the one on the stratum corneum side, called the application reservoir and the one on the subcutaneous side, called the receptor reservoir. One simply places radiolabeled test material in the application reservoir and collects samples at various time points from the receptor fluid.

The rate of penetration can be presented by the traditional kinetic formulas to obtain a penetration rate constant. Given that exposed surface area also plays a role in the amount of material absorbed, the concept of flux is also important.

Determining the quantity of material that is absorbed into the skin and eventually released into the systemic circulation is primarily dependent upon three factors: the surface area exposed, the volume of material applied, and the concentration of the material applied:

*Surface Area* All things being equal, it is clear that the greater the surface exposed, the higher the achieved internal dose.

*Volume* The volume of material will obviously play a role in total dose, but it is not as straightforward as the relationship to surface area. Theoretically, the maximum absorption is obtained when the material is spread as thin and uniform as possible; piling material on so that it is literally rolling off the animal serves no practical purpose. In fact, it is not sound practice when dealing with an in vivo animal experiment as it makes it more likely for the material to be available for oral ingestion.

*Concentration* The higher the concentration in a formulation, the higher the flux achieved of drug molecules across the skin.

Of course, the nature of the vehicle that the drug is being carried in may also have a profound influence on absorption.

***Parameters Controlling Absorption*** The absorption of a chemical into the skin is a function of the nature of the molecule, the behavior of the vehicle, and the status of the skin. Three major variables account for differences in the rate of absorption or flux of different topical chemicals or of the same molecule in different vehicles: the concentration of the molecule in the vehicle, the partition coefficient of chemical between the stratum corneum and the vehicle, and the diffusion coefficient of the molecule in the stratum corneum.

The rate of diffusion is proportional to the concentration of molecule in the vehicle. The relationship is linear only at low molecule concentrations and only applies to the soluble molecule in the vehicle. The latter factor may explain the variable therapeutic effects of different formulations of the same drug molecule. The partition coefficient is a measure of the molecule's ability to escape from the vehicle and is defined as the equilibrium solubility of the

molecule in the surface of the stratum corneum relative to its solubility in the vehicle. Increased lipid solubility favors penetration of the molecule through the skin by increasing the solubility in the relatively lipophilic stratum corneum. The diffusion coefficient indicates the extent to which the matrix of the barrier restricts the mobility of the molecule. Increases in molecular size of the molecule will increase frictional resistance and decrease the diffusion coefficient (Bronaugh, 1998); molecules over 1000 Da usually will not be absorbed easily into normal adult skin.

Finally, intact stratum corneum is an excellent barrier, but in disease states that compromise the skin barrier, the resistance to absorption is rapidly lost and absorption can be facilitated. Such compromised skin can be humanly simulated by using either a dermatome or tape striping the skin site in question.

### 3.2 Distribution

Once the chemical gains access to the body, it is carried by the bloodstream and distributed to the different organs. The preferential organ of deposition is determined by a variety of factors: The two most important are blood flow to the organ and the affinity of the chemical for that organ. Affinity is governed by two general characteristics. First, the product may be designed to have a specific affinity for a specific molecular entity in a target cell. For example, an anticholinesterase insecticide will tend to accumulate in the cells that have the highest concentration of cholinesterase. Second, the product may have a nonspecific or general chemical attraction for a specific cell type. The more highly lipophilic a chemical, the more likely it is to distribute and remain in adipose tissue. Blood flow will also have a major impact on distribution, as chemicals will be distributed more readily to those organs that are more highly perfused. A highly lipophilic chemical may first be deposited in the brain due to the fact that it is richly perfused and then be distributed to body fat with time.

Once a material is absorbed, distribution of a compound in most early toxicology studies is usually of limited interest. This is unfortunate, as it is the preferential distribution of the drug to the therapeutic target that is desired. Some factors which can serve to alter distribution are listed in Table 3.

For most drugs, the rate of disposition or loss from the biological system is independent of rate and input once the agent is absorbed. Disposition is defined as what happens to the active molecule after it reaches a site in the blood circulation where concentration measurements can be made (the systemic circulations, generally). Although disposition processes may be independent of input, the inverse is not necessarily true because disposition can markedly affect the extent of availability. Agents absorbed from the stomach and the intestine must first pass through the liver before reaching the general circulation (Figure 1). Thus, if a compound is metabolized in the liver or excreted in bile, some of the active molecule absorbed from the GI tract will be inactivated by hepatic processes before it can reach the systemic circulation

**TABLE 3 Selected Factors That May Affect Chemical Distribution to Various Tissues**

| |
|---|
| Factors relating to chemical and its administration |
| • Degree of binding of chemical to plasma proteins (i.e., agent affinity for proteins) and tissues |
| • Chelation to calcium, which is deposited in growing bones and teeth (e.g., tetracyclines in young children) |
| • Whether chemical distributes evenly throughout body (one-compartment model) or differentially between different compartments (two-or-more-compartment model) |
| • Ability of chemical to cross blood–brain barrier |
| • Diffusion of chemical into tissues or organs and degree of binding to receptors that are and are not responsible for drug's beneficial effects |
| • Quantity of chemical given |
| • Route of administration/exposure |
| • Partition coefficients (nonpolar chemicals are distributed more readily to fat tissues than are polar chemicals) |
| • Interactions with other chemicals that may occupy receptors and prevent drug from attaching to receptor, inhibit active transport, or otherwise interfere with drug's activity |
| • Molecular weight of chemical |
| Factors relating to test subject |
| • Body size |
| • Fat content (e.g., obesity affects distribution of drugs that are highly soluble in fats) |
| • Permeability of membranes |
| • Active transport for chemicals carried across cell membranes by active processes |
| • Amount of proteins in blood, especially albumin |
| • Pathology or altered homeostasis that affects any of the other factors (e.g., cardiac failure and renal failure) |
| • Presence of competitive binding substances (e.g., specific receptor sites in tissues bind drugs) |
| • pH of blood and body tissues |
| • pH of urine[a] |
| • Blood flow to various tissues or organs (e.g., well-perfused organs usually tend to accumulate more chemical than less well perfused organs) |

[a]The pH of urine is usually more important than the pH of blood.

and be distributed to its sites of action. If the metabolizing or biliary excreting capacity of the liver is great, the effect on the extent of availability will be substantial. Thus, if the hepatic blood clearance for the chemical is large, relative to hepatic blood flow, the extent of availability for this chemical will be low when it is given by a route that yields first-pass metabolic effects.

Likewise, metabolism is generally of only limited concern in most acute studies. There are some special cases, however, in which metabolic considerations must be factored in seeking to understand differences between routes and the effects which may be seen.

The first special case is parenteral routes, where the systemic circulation presents a peak level of the moiety of interest to the body at one time tempered only by the results of a single pass through the liver.

The second special case arises from inhalation exposures. Because of the arrangements of the circulatory system, inhaled compounds enter the full range of systemic circulation without any "first-pass" metabolism by the liver. Keberle et al. (1971) and O'Reilly (1972) have published reviews of absorption, distribution, and metabolism that are relevant.

***Protein Binding*** The degree to which a drug binds to plasma proteins will highly influence its distribution. Albumin, the most prominent of the many proteins found in mammalian plasma, carries both positive and negative charges with which a polar compound can associate by electrostatic attraction. As with all such reactions it can be described by the following equations. The more avidly bound the material, the less will be distributed to surrounding fluids as part of a solution and only that portion that is free in solution will be available for diffusion into the tissues.

***Water Solubility*** The solubility of a chemical has a direct bearing on its distribution. Recall that only molecules that are in solution will be available for absorption.

As mentioned above, only that portion that is free in solution will be available for diffusion into the tissues. Hence, the more material that is in solution, the more will be available for diffusion.

***Volume of Distribution*** If one takes the dose administered (in milligrams) and divides it by the plasma concentration of the test material (milligrams per milliliter), the result is a volume number:

$$\frac{\text{Dose}}{\text{Concentration}} = \text{volume}$$

One can take this process a step further and extrapolate back from a plasma time curve to the $y$ axis. This is theoretically the plasma concentration ($C_0$) that would occur if, upon being administered, the material is instantly distributed throughout the body. The volume number obtained with the above equation becomes

$$\frac{\text{Dose}}{C_0} = V_D$$

where $V_D$ represents the apparent volume of distribution, a proportionality constant that reflects the relation of the concentration of a xenobiotic in plasma to the total amount of the entity in the body. Materials that are avidly bound to plasma proteins will have a high volume of distribution, while materials that are avidly taken by the tissues (deposit fat, for example) will have a low one. The $V_D$ is a parameter that is simple to calculate yet gives an important piece of information about the distribution of the chemical under investigation.

**TABLE 4 Volume and Half-Life of Body Water in Selected Species**

| Species | Sex | Exchangeable Body Water (% of Body Weight) | Half-Life (Days) |
|---|---|---|---|
| Mouse | F | 58.5 | 1.13 |
| Rat | M | 59.6 | 2.53 |
| Rabbit | F | 58.4 | 3.87 |
| Dog | M | 66.0 | 5.14 |
| Cynomolgus monkey | M | 61.6 | 7.80 |
| Rhesus monkey | M | 61.6 | 7.80 |
| Humans | M, F | 55.3 | 9.46 |

**TABLE 5 Typical Organ Weights in Adult Laboratory Animals**

| | Percent of Body Weight | | | | |
|---|---|---|---|---|---|
| Organ | Rat | Mouse | Dog | Rabbit | Monkey |
| Liver | 3.5 | 6 | 3.5 | 3 | 2.5 |
| Kidney | 0.8 | 1.6 | 0.5 | 0.8 | 0.5 |
| Heart | 0.4 | 0.4 | 0.8 | 0.3 | 0.4 |
| Spleen | 0.3 | 0.5 | 0.3 | 0.04 | 0.1 |
| Brain | 0.5 | 0.6 | 0.8 | 0.4 | 3 |
| Adrenals | 0.02 | 0.01 | 0.01 | 0.02 | 0.03 |
| Lung | 0.6 | 0.6 | 1 | 0.6 | 0.7 |

The available volumes and masses for distribution vary from species to species, as summarized in Tables 4 and 5.

### 3.3 Metabolism/Biotransformation

Metabolism describes the process by which chemicals are changed by the body. In fact, very few foreign chemicals that come to enter the body are excreted unchanged. Most are chemically modified. In general, metabolism results in chemicals that are more polar and water soluble and more easily excreted (La Du et al., 1972). Examples of more common metabolic conversions are shown in Table 6. In general, the vast majority of lipophilic chemicals are first oxidized via the cytochrome P-450 (CYP)–dependent mixed-function oxidase system of the liver. This is the process classically called phase I metabolism. Cytochrome P-450 exists as a family of isozymes (the CYP gene superfamily) with varying but overlapping substrate affinity and responses to different inducing agents. For a review of the molecular biology of the CYP gene superfamily the reader is referred to Meyer (1994). Induction is the process whereby exposure to a chemical leads to increased activity of the mitochondrial mixed-function oxidase (MMFO) due to an increase in CYP. The isoenzymes induced by a variety of different chemicals are given in Table 7, and example compounds which inhibit specific CYPs are presented in

**TABLE 6 Summary of Prominent Phase I Biotransformation Reactions**

| Reaction | Enzyme | Location | Example/Comments |
|---|---|---|---|
| Hydrolysis | Carboxylesterase | Ubiquitous | Vinyl acetate to acetate and acetaldehyde |
| | Peptidase | Blood, lysomes | Amino-, carboxy-, and endopeptidases which cleave peptides at specific amino acid linkages |
| | Epoxide hydroplase | Microsomes, cytosol | Conversion of styrene 7,8-epoxide to styrene 7,8-glycol |
| Reductions | Azo and nitro reduction | Gut microflora | Sequential conversion of nitrobenzene to aniline |
| | Carbonyl reductase | Cytosol | Conversion of haloperidol to reduced haloperidol (secondary alcohol) |
| | Disulfide reduction | Cytosol | Glutathione-dependent reduction of disulfiram to diethyldithiocarbamate |
| | Sulfoxide reduction | Cytosol | Thioredoxin-dependent sulindac to sulindac sulfide |
| | Quinone reduction | Cytosol, microsomes | DT diaphorase reduction of menadione to hydroquinone |
| | Reductive dehalogenation | Microsomes | Conversion of pentabromoethane to tetrabromoethane (releasing free bromide ion) |
| Oxidation | Alcohol dehydrogenase | Cytosol | Conversion of ethanol to acetaldehyde [Drosophila alcohol dehydrogenase (DAD)/DADH-dependent reversible reaction] |
| | Aldehyde dehydrogenase | Mitochondria/ cytosol | Conversion of acetaldehyde to acetate |
| | Aldehyde oxidase | Liver cytosol | Flavin adenine dinucleotide (FAD)–dependent metalloenzyme, oxidation of benzaldehyde to benzoic acid |
| | Xanthene oxidase | Cytosol | Oxidation of purine derivative, conversion of allopurinol to alloxanthene |
| | Monoamine oxidase | Mitochondria | FAD-dependent oxidative deamination of monoamines, e.g., primaquine |
| | Diamine oxidase | Cytosol | Pyridoxal-dependent, copper-containing enzyme; conversion of allylamine to acrolein |
| | Prostaglandin oxidase | Microsomes | Cooxidation reaction, can "activate" chemical in tissues low in cytochrome P-450, e.g., nephrotoxicity of acetaminophen, oxidation of phenylbutazone |
| | Flavin-monooxygenase | Microsomes | FAD-dependent oxidation of nucleophilic nitrogen, sulfur, and phosphorus heteroatoms, e.g., conversion of nicotine to nicotine 1′-*N*-oxide, cimetidine to cimetidine *S*-oxide |
| | Cytochrome P-450 | Microsomes | |

**TABLE 7 Examples of Xenobiotics Metabolized by Human P450**

| | |
|---|---|
| CYP1A1: Benzo[*a*]pyrene and other polycyclic aromatic hydrocarbons | CYP2E1 |
| CYP1A2 | Acetaminophen |
| Acetaminophen | Acrylonitrile |
| 2-Acetylaminofluorene | Benzene |
| 4-Aminobiphenyl | Carbon tetrachloride |
| 2-Aminofluorene | Chloroform |
| 2-Naphthylamine | Chlorzoxzone |
| Amino acid pyrolysis products (DiMeQx, MelQ, MelQx, Glu P-1, Glu P-2, IQ, PhIP, Trp P-1, Trp P-2) | Dichloromethane |
| CYP2A6 | 1,2-Dichloropropane |
| 6-Aminochursene | Ethylene dibromide |
| Cyclophosphamide | Ethylene dichloride |
| Isosphamine | Ethyl carbamate |
| *N*-Nitrosodiethylamine | *N*-Nitrosodimethylamine |
| CYP2B6 | Styrene |
| 6-Aminochrysene | Trichloroethylene |
| Cyclophosphamide | Vinyl chloride |
| Ifosphamine | CYP3A4 (50% of all marketed drugs are marketed here) |
| CYP2C8, 9, 18 (Note: 2C9 is absent in 15–30% of Asians) | Acetaminophen |
| Talbutamide | Aflatoxin $B_1$ and $G_1$ |
| Taxol | 6-Aminochrysene |
| CYC2C19 | Benzo[*a*]pyrene 7,8-dihydrodiol |
| Diazepam | Cyclophosphamide |
| Diphenylhydantoin | Ifosphamide |
| Hexaburbitol | Nidazolam |
| Propanolol | Nifedipine |
| CYP2D6 (Note: absent in 7% of Caucasians): Bufuralol | Testosterone |
| | 1-Nitropyrene |
| | Sterigmatocystin |
| | Senecionine |
| | Tris(2,3-dibromopropyl) phosphate |
| | CYP4A9/11: None known |

Table 7. In a practical sense, a drug can induce its own metabolism. Hence, repeated dosing with a chemical may lead to lower blood levels at the end, for example, of a 13-week study than at the beginning. There could also be alterations in the spectrum of metabolites produced such that an agent could become more or less toxic with repeated dosing depending on the nature of the metabolites. It is not unusual during a subchronic or chronic toxicity test for tolerance to occur. There may be signs of toxicity early in the study, but even with continued daily dosing, the signs abate. This phenomenon, particularly in rodents, is frequently due to microsomal induction, whereby the chemical has induced its own metabolism and more rapid clearance of the parent chemical occurs. It should be noted the CYP system is not the only drug-metabolizing system. As Table 8 summarizes, there are at least five major metabolic systems in mammals.

**TABLE 8 A Comparison of the key *in vitro* drug metabolizing experimental systems (liver microsomes), liver postmitochondrial supernatant (S9), liver cytosol (cytosol) and hepatocytes in their contents of the major drug metabolizing enzymes (cytochrome P450 isoforms (P450); monoamine oxidase (MAO); UDP-glucuronsyl transferase (UGT); sulfotransferase (ST); and glutathione-S-transferase (GST)**

| *In Vitro* System | P450 | MAO | UGT | ST | GST |
|---|---|---|---|---|---|
| Microsomes | + | – | +* | – | +** |
| S9 | + | – | +* | +* | + |
| Cytosol | – | – | –* | +* | +*** |
| Hepatocytes | + | + | + | + | + |

*activity of this drug metabolizing enzyme requires the addition of specific cofactors, for instance UDP-glucuronic acid (UDPGA) for UGT activity, and 3′-phosphoadenosine 5′-phosphosulfate (PAPS) for ST activity.
**membrane-bound GST but not the soluble GST are found in the microsomes.
***soluble GST but not membrane-bound GST are found in the cytosol.

After the drug has been metabolically oxidized, it can in fact be further metabolized. In fact, it is possible for the metabolites to also be substrates of the MMFO and to be metabolized themselves.

The route of metabolic activation of the classic carcinogen benzo[*a*]pyrene is due to such a mechanism. The biology of these reactive intermediates has been extensively studied. Glutathione is among the most common organic intracellular chemicals in all mammalian species, being present at a concentration of up to 10 mm and glutathione *S*-transferase is very active. Glutathione is a tripeptide (glutamine–cysteine–glycyne). The sulfhydryl group of cysteine is the business end of the molecule where the reaction with the nucleophilic reactive intermediate takes place. After that, the glutathione conjugate is further metabolized to a cytinyl-acetyl moiety. These moieties are called mercapturic acids and are generally found in the urine. The relative predominance of mercapturic acid over other metabolites may be considered a rough indication of how "reactive" the intermediates may have been. Teleologically, it is tempting to speculate that it is a very well designed protective mechanism. So long as intracellular glutathione concentrations remain above a critical level, the destructive actions of active metabolites can be held in check. Thus, a small dose of a chemical (bromo-benzene is a good example) may cause no liver damage while a large dose may. This is also a good example of one of the aspects of toxicokinetics versus pharmacokinetics where a high dose of a chemical will become toxic due to saturation of a detoxification pathway.

The glutathione *S*-transferase pathway is sometimes in biochemical competition with the epoxide hydratase pathway, in that both deactivate intermediates of the MMFO. Epoxide hydratase is a microsomal enzyme that acts specifically to deactivate epoxide intermediates by the addition of water across the C–O bond to form a diol. As a very broad generality, the glutathione *S*-transferase pathway tends to be more prominent in rodents, while the epoxide hydratase pathway tends to be more dominant in nonrodents.

The hydroxyl- or diol-containing metabolites of the MMFO can be further metabolized by so-called phase II (synthetic) metabolism whereby they are conjugated to/from glucuronides and/or sulfates (so-called etherial sulfates). Amines can also be substrates. The net effect of phase II reactions is to create a more polar molecule that is more readily excretable. While there are species differences, glucuronides are actively transported and excreted in the bile into the GI tract. Sulfates are excreted more predominantly in the urine. Both glucuronides and sulfates, however, can be found in both the urine and the feces. Like the MMFO pathway, glutathione 3-transferase, uridine 5′-diphospho (UDP)-glucuronyl transferase, and eopoxide hydratase are inducible, that is, treatment with exogenous chemicals will increase the amount of enzyme protein present.

Outside of the MMFO-mediated (phase I) reactions there are a few other major reactions that are worthy of note. The two major ones involve ester hydrolysis and alcohol and aldehyde dehydrogenases. All mammalian species have an extensive ability to hydrolyze the ester bond. The products of the reactions then can go on to be further metabolized. In the pharmaceutical industry, this property has been utilized to synthesize prodrugs, that is, chemicals that have desirable pharmaceutical properties (generally increased water solubility) that are not converted to their active moiety until hydrolyzed in the body.

The activity of alcohol dehydrogenase is one with which we should all be familiar. It oxidizes alcohols to aldehydes. The aldehydes produced by this reaction can go on to be further metabolized to a carbocylic acid if they are not sterically hindered. Side-chain constituents of aromatic compounds can also be a substrate for this reaction sequence, producing side-chain carboxylates. The oxidation of alcohols to aldehydes can also be a form of metabolic activation as aldehydes can have potent physiological actions. Fortunately, aldehyde dehydrogenase has a very high activity when compared to alcohol dehydrogenase, so that the aldehydes do not accumulate. Inhibition of aldehyde dehydrogenase by disulfiram (Antabuse) leads to the accumulation of acetaldehyde, causing nausea, dizziness, and flushing. Like disulfiram, some pesticides contain dithiocarbamates and have the potential of causing this type of reaction.

Hopefully, this brief description of the major metabolic pathways has given one some appreciation of the richness of the processes. The different sites of oxidation, the possibility of additional oxidative metabolism of metabolites, and differences in phase II reactions all lead to a multiplicity of possible metabolites. Over 100 different metabolites of the human pharmaceutical chlorpromazine have been isolated and identified. When analyzed by high-performance liquid chromatography (HPLC), for example, the parent chemical and the different (detectable) metabolites will form a pattern of different peaks. This is referred to as the metabolic fingerprint or profile of a chemical. Different species will have different profiles. Ideally, in doing a risk assessment, one would like to know the similarity in this pattern between the animals used

in the toxicology studies and that produced by human beings. This is only infrequently available for most nonpharmaceutical products, as pesticides (for example) are rarely given intentionally to human subjects for the purposes of study. The technology now exists, however, to address this potential problem. Cell lines with human CYP have been developed that can provide some indication of the similarities of human metabolism of a chemical to that of experimental animals. At least they may be able to assist in identifying the major oxidative metabolite. For nonpharmaceutical products, it may be an unusual circumstance that would require one to identify potential human metabolites as part of a marketing application; however, it may be useful for one to know that the technology exists to do so.

The processes of metabolic conversion are frequently involved in the mechanisms of toxicity and carcinogenicity.

***Metabolic Activation*** As mentioned, most nonnutritive chemicals pass through the GI tract by passive absorption and then enter the mesenteric circulation. The venous circulation from the mesentery flows through the portal vein into the liver. The metabolic action of the liver literally sits between the GI tract and the general systemic circulation. Thus, even chemicals that may be highly absorbed from the GI tract could appear only sparingly in systemic circulation if they are highly metabolized by the liver. The combination of absorption from the GI tract and metabolism from the liver leads to what is called the first-pass effect. An extension of this is the fact that the gut flora contain glucuronidases that can cleave glucuronides of chemicals and/or metabolites that are then available to be reabsorbed. This process is called enterohepatic circulation.

***Induction of CYP Metabolism and Isoenzymes*** When organisms are exposed to certain xenobiotics, their ability to metabolize a variety of chemicals is increased. This phenomenon can produce either a transitory reduction in the toxicity of a drug or an increase (if the metabolite is the more toxic species). However, this may not be the case with compounds that require metabolic activation. The exact toxicological outcome of such increased metabolism is dependent on the specific xenobiotic and its specific metabolic pathway. Since the outcome of a xenobiotic exposure can depend on the balance between those reactions that represent detoxification and those that represent activation, increases in metabolic capacity may at times produce unpredictable results.

The ability of different drugs to differentially inhibit and/or induce individual CYP isoenzymes has become critical in assessing the potential safety of drug molecules. Table 9 presents a summary overview of some of what we have come to know about differential metabolism by CYP isoenzymes. Draper et al. (1998) have published work on the use of human liver microsomes for determining the levels of activity or inhibition a drug has the formation of 6-testosterone as a model for CYP3A activity (1) and chlorzoxazone for

**TABLE 9 Examples of Xenobiotics Activated by Human Cytochrome P-450 Isoenzymes**

| | |
|---|---|
| CYP1A1 | Codeine |
| Benzo[*a*]pyrene and other polycyclic aromatic hydrocarbons | Timolol |
| CYP1A2 | Metoprolol |
| Acetaminophen | CYP2E1 |
| 2-Acetylaminofluorine | Acetaminophen |
| 4-Aminobiphenyl | Acrylonitrile |
| 2-Aminofluorene | Benzene |
| 2-Naphthylamine | Carbon tetrachloride |
| CYP2A6 | Chloroform |
| *N*-Nitrosodiethylamine | Dichloromethane |
| Butadiene | 1,2-Dichloropropane |
| Coumarin | Ethylene dibromide |
| CYP2B6 | Ethylene dichloride |
| 6-Aminochrysene | Ethyl carbamate |
| Cyclophosphamide | *N*-Nitrosodimethylamine |
| Ifosphamine | Styrene |
| CYP2C8: Taxol | Trichloroethylene |
| CYP2C9 | Vinyl chloride |
| Diclofenac | CYP3A4 |
| Phenytoin | Acetaminophen |
| Piroxicam | Aflatoxin $B_1$ and $G_1$ |
| Tolbutamide | 6-Aminochrysene |
| CYP2C19 | Benzo[*a*]pyrene 7,8-dihydrodiol |
| Diazepam | Cyclophosphamide |
| Diphenylhydantoin | Ifosphamide |
| Hexabarbitol | 1-Nitropyrene |
| Propanolol | Sterigmatocystin |
| CYP2D6 | Senecionine |
| Buforolol | Tris(2,3-dibromopropyl) phosphate |
| | CYP4A9/11: None known |

*Source*: Adapted in part from Parkinson, 1996.

CYP2E1 activity (2). If, for example, a chemical under study competitively inhibits the metabolism of these model substrates in these systems, then it is a substrate for that human isozyme. Using these more recently available in vitro systems, it is much easier to perform cross-species comparisons with regard to biotransformation. It is now easier to determine how similar the routes of metabolism are in the experimental animals with comparison to that in humans without having to administer the chemical to human subjects. Human and animal model microsome preparations may be used as models to identify patterns of metabolites in vitro, allowing for better selection of model species for safety studies, and competition for or inhibition activation of specific isoenzymes can be evaluated to identify potential problems of drug–drug interaction in patients (Levy et al., 2000).

***Species Differences*** Species differences in metabolism are among the principal reasons that there are species differences in toxicity. A difference in CYP

is one of the most common reasons for a difference in metabolism. For example, Monostory et al. (1997) recently published a paper comparing the metabolism of panomifene (a tamoxifen analogue) in four different species. These data serve to address that the rates of metabolism in the nonhuman species was most rapid in the dog and slowest in the mouse. Thus, one should not a priori make any assumptions about which species will have the more rapid metabolism. Of the seven metabolites, only one was produced in all four species. Both the rat and the dog produced the two metabolites (M5 and M6) produced by human microsomes. So how does one decide which species best represents the humans? One needs to consider the chemical structure of the metabolites and the rates at which they are produced. In this particular case, M5 and M6 were relatively minor metabolites in the dog which produced three other metabolites in larger proportion. The rat produced the same metabolites at a higher proportion, with fewer other metabolites than the dog. Thus, in this instance the rat, rather than the dog, was a better model. Likewise, Table 10 offers a comparison of excretion patterns between three species for a simple inorganic compound. Table 11 presents a summary of interspecies differences between species in drug metabolism and pharmacokinetics.

**TABLE 10 Differences in Disposition of 2,4-Dichlorophenoxyacetic acid**

| Species | Sex | Urine | Feces |
|---|---|---|---|
| Rat | M | 31.2 | 2.7 |
| | F | 16.5 | 1.1 |
| Mouse | M | 12.7 | 2.8 |
| | F | 26.8 | 6.7 |
| Hamster | M | 4.9 | 2.5 |
| | F | 33.9 | 14.5 |

*Note*: All animals dosed orally with radiolabeled 2.4-D, 200 mg/kg. Results are expressed as percent of 14C dose recovered. Urine was collected for 8 h and feces for 24 h.

**TABLE 11 "General Rules" on interspecies differences in DMPK**

| Species | DMPK Characteristics |
|---|---|
| Human | Polymorphisms (e.g., CYP 2C9, CYP 2C19, CYP 2D6, NAT1, NAT2) |
| Dog | Low acetylation, high capacity for deacetylation |
| | Different absorption due to higher pH in gastro-intestinal tract than in humans (consider use of synthetic gastric fluid to mimic human situation) |
| Rat | Often gender differences which are not observed in other species |
| | Abundant tetrahydrofolate (protects, e.g., against methanol ocular damage) |
| Rabbit | Low sulfation |
| (Mini)Pig | Low sulfation |
| | Gastro-intestinal conditions similar to humans |
| Cat | Low glucuronidation |
| | High sulfation |

More thorough reviews on species differences in pharmacokinetics have been presented by Smith (1991), Gad and Chengelis (1997), and Gad (2007a).

***Sex-Related Differences in Rodents*** Not only are there differences in absorption, distribution, biotransformation, and metabolism between species, there may also be differences between sexes within a species (Mugfor and Kidderis, 1998). Griffin et al. (1997), for example, has demonstrated sex-related differences in the metabolism of 2,4-dichlorophenoxyacetic acid (2,4-D). They noted that while there were differences between sexes they tended to be quantitative (rates), not qualitative (metabolites). Differences between species were greater than sex-related differences. With regard to sex-related differences, it is noteworthy that males do not always have the higher rates, as Griffen et al. have shown; in hamsters, the female metabolizes 2,4-D more rapidly than males. In general, male rats tend to have higher activity than female rats, especially with regard to CYP-dependent activity.

In the case of 2,4-D, the only urinary metabolite is 2,4-D glucuronide, but the half-life of 2,4-D was 138 min in males and 382 min in females.

## 3.4 Excretion

Excretion encompasses the process by which chemicals or their metabolites are transported out of the body. There are three possible major routes of excretion and a handful of minor ones. The major routes of excretion for chemicals, in particular their metabolites, are as follows:

***Urine*** The kidneys filter the entire cardiac output multiple times each day and thus provide a large opportunity for the removal of chemicals from the bloodstream. How much of a xenobiotic is actually excreted is dependent on three factors or processes:

1. The glomerular membrane has pores of 70–80 Å; under the positive hydrostatic conditions in the glomerulus, all molecules smaller than about 20,000 Da are filtered. Proteins and protein-bound compounds thus remain in the plasma, and about 20% of the nonbound entity is carried with 20% of the plasma water into the glomerular filtrate.
2. Because the glomerular filtrate contains many important body constituents (e.g., glucose), there are specific active uptake processes for them. Also, lipid-soluble chemicals diffuse back from the tubule into the blood, especially as the urine becomes more concentrated because of water reabsorption. The pH of the urine is generally lower than that of the plasma, and therefore pH partitioning tends to increase the reabsorption of weak acids. The pH of the urine can be altered appreciably by treatment with ammonium chloride (decreases pH) or sodium carbonate (increases pH); the buffered plasma shows little change.

3. Xenobiotics may be secreted actively into the renal tubule against a concentration gradient by anion and cation carrier processes. These processes are saturable and of relatively low specificity; many basic or acidic compounds and their metabolites (especially conjugation products) are removed by them. Because the dissociation rate for the chemical–albumin complex is rapid, it is possible for highly protein-bound compounds to be almost completely cleared at a single passage through the kidney.

***Feces*** The most important mechanism allowing circulating foreign compounds to enter the gut is in the bile. The biological aspects of this mechanism have been reviewed, and certain pertinent points have emerged. The bile may be regarded as a complementary pathway to the urine, with small molecules being eliminated by the kidney and large molecules in the bile. Thus the bile becomes the principal excretory route for many drug conjugates. Species differences exist in the molecular weight requirement for significant biliary excretion, which has been estimated as 325 ± 50 in the rat, 440 ± 50 in the guinea pig, and 475 ± 50 in the rabbit. In the rat, small molecules (less than 350 Da) are not eliminated in the bile or large molecules (more than 450 Da) in the urine, even if the principal excretory mechanism is blocked by ligation of the renal pedicles or bile duct, respectively. Compounds of intermediate molecular weight (350–450 Da) are excreted by both routes, and ligation of one pathway results in increased use of the other.

Foreign compounds may also enter the gut by direct diffusion or secretion across the gut wall, elimination in the saliva, pH partitioning of bases into the low pH of the stomach, and elimination in the pancreatic juice.

***Expired Air*** Volatile compounds or metabolites can be extensively excreted by passage across pulmonary membranes into the airspace of the lungs, then expulsion from the lungs in expired air.

Minor routes for excretion can include tears, saliva, sweat, exfoliated keratinocytes, hair, and nasal discharge. These are of concern or significance only in rare cases. Accordingly, quantitation of excretion typically requires collection of urine and feces (and occasionally expired air) over a period of time.

### 3.5 Pharmacokinetics

The interplay of the processes of ADME result in changes in concentration of the test chemical in different organs with time (Shargel and Yu, 1999; Renwick, 2000). With regard to the practical concerns of monitoring human exposure, the organ of interest is the blood. Blood is generally considered the central compartment. Determining the concentration of the chemical in plasma gives one an assessment of exposure. Mathematical formulas are used to quantitatively describe this exposure (Bauer, 2001).

***Physiologically Based Pharmacokinetic Modeling*** Pharmacokinetic parameters are descriptive in nature. They quantitatively describe the manner in which a test material is absorbed and excreted such that a specific blood or tissue level is achieved or maintained. In the past, experiments had to be done by every route of administration to gather the data appropriate for describing the pharmacokinetic behavior of a chemical administered by different routes. The development of more sophisticated and readily accessible computers has led to the development of a different approach, that of pharmacokinetic modeling (Connally and Anderson, 1991). In this computerized model, different compartments are represented as shown in boxes and the movement of the material in and out of the compartments is defined by the rate constants. These can be determined either in vivo or in vitro. Other physiological parameters are brought into play as well, such as octanol–water partition coefficient, blood flow through an organ, respiration rate (for the inhalation route of exposure), rate of microsomal metabolism, and so on.

## 4 LABORATORY METHODS

The actual means by which pharmacokinetic information is collected is through the conduct of one or more specific studies employing a wide range of available analytical techniques. Administered therapeutic molecules can be identified and quantified in relevant samples collected in accordance with carefully designed and executed protocols.

### 4.1 Analytical Methods

There are three broad categories of analytical techniques now available—instrumental (cold chemical), radiolabeled, and immunological. Each of these have advantages and disadvantages. Only an overview of these techniques will be given here—detailed explanations are beyond the scope of this text. These methodologies are all directed at being able to identify and/or quantify a chemical (and/or its metabolites) in various biological matrices.

***Instrumental Methods*** These bioanalytical methods are also sometimes called cold-chemistry methods. These generally start from a place of isolating the compound or compounds of interest, for which the work horse methodology is HPLC. A wide variety of specialized columns are used to achieve desired separation. At the end of the column, where separation of molecular entities has been achieved, the outflow of the column can be directed to any of a wide variety of detection instruments, including various forms of detectors intrinsic to the HPLC. In general, all of the cold-chemistry methodologies have less sensitivity (higher detection limits) than do radiochemical or immunological methods (Caldwell et al., 1994).

Mass spectrometry (MS); nuclear magnetic resonance (NMR); electron skin resonance (ESR); ultraviolet, infrared, and visible spectrophotometry; and mass spectroscopy are all well-established detection methodologies.

***Radiochemical Methods*** The massive expansion of our understanding of toxicokinetics since the late 1970s is to a large degree a reflection of the wide use of radioactive isotopes as tracers of chemical and biological processes. Appropriately radiolabeled test compounds are commonly used in toxicokinetic studies, providing a simple means of following the administered dose in the body. This is particularly important when specific analytical methods are unavailable or too insensitive. The use of total radioactivity measurements allows an estimation of the total exposure to drug-related material and facilitates the achievement of material balance.

The most commonly used radionuclides in drug metabolism and disposition studies are carbon-14 ($^{14}C$) and tritium ($^{3}H$), both of which are referred to as beta emitters. Since these beta-emitting isotopes have relatively long half-lives, their radioactive decay during an experiment is insignificant. Additionally, they provide sufficient emission energy for measurement and are relatively safe to use, as indicated by the data in Table 12. Although individual beta particles can have any energy up to the maximum, $E_{max}$, the basic quantity in determining the energy imparted to tissues by beta emitters is the average energy $E_{\beta}$. The range is the maximum thickness the beta particles can penetrate. Beta particles present virtually no hazard when they originate outside the body (Shapiro, 1981).

During the synthesis of radiolabeled compounds, the label is usually introduced as part of the molecular skeleton in a metabolically stable and, with tritium, nonexchangeable position. The in vivo stability of $^{14}C$ labels is often reflected by the extent of [$^{14}C$] carbon dioxide formation. The biological stability of $^{3}H$ labels can be estimated by the extent of tritiated water formation. The tritiated water concentration (dpm $mL^{-1}$) in urine samples collected during a designated time interval after dosing, assumedly after equilibrium is reached between urine and the body water pool, is determined. This value is extrapolated from the midpoint of the collection interval to zero time based on the known half-life of tritiated water in the given species. The percentage of the radioactive dose that is transformed to tritiated water (% $^{3}H_2O$) can be calculated using the equation

**TABLE 12 Properties of Primary Radioisotopes Employed in Pharmacokinetics**

| Property | $^{3}H$ | $^{51}Cr$ | $^{14}C$ | $^{125}I$ |
|---|---|---|---|---|
| Half-life | 12.3 years | 27.8 days | 5730 years | 13 days |
| Maximum beta energy (MeV) | 0.0186 | 0.752 | 0.156 | 2.150 |
| Average beta energy (MeV) | 0.006 | 0.049 | | |
| Range in air (mm) | 6 | 300 | | |
| Range in unit density material (mm) | 0.0052 | 0.29 | | |

$$\text{Percent } {}^{3}H_2O = \frac{{}^{3}H_2O \text{ concentration at zero time} \times \text{exchangeable body water volume}}{\text{radioactivity dose}} \times 100\%$$

Values for the exchangeable body water content as well as the half-life of tritiated water in some mammalian species that can be applied to the above equation were shown earlier in Table 3. If the molecule is likely to or is known to fragment into two major portions, it may be desirable to monitor both fragments by differential labeling ($^{3}H$ and $^{14}C$).

The chemical and radiochemical purity of the labeled compound must be ascertained prior to use. In practice a value of 95% or greater is usually acceptable. The desired specific activity of the administered radioactive compound depends on the dose to be used as well as the species studied. Doses of $^{14}C$ on the order of $5\,\mu Ci\,kg^{-1}$ for the dog and $20\,\mu Ci\,kg^{-1}$ for the rat have been found adequate in most studies, while doses of $^{3}H$ are usually two to three times higher owing to lower counting efficiency of this isotope.

Liquid scintillation counting is the most popular technique for the detection and measurement of radioactivity. In order to count a liquid specimen such as plasma, urine, or digested blood or tissues directly in a liquid scintillation spectrometer, an aliquot of the specimen is first mixed with a liquid scintillant. Aliquots of blood, feces, or tissue homogenates are air-dried and ash-free filter papers and combusted in a sample oxidizer provided with an appropriate absorption medium and a liquid scintillant prior to counting. The liquid scintillant plays the role of an energy transducer, converting energy from nuclear decay into light. The light generates electrical signal pulses which are analyzed according to their timing and amplitude and are subsequently recorded as a count rate, for example, counts per minute (cpm). Based on the counting efficiency of the radionuclide used, the count rate is then converted to the rate of disintegration, for example, disintegrations per minute (dpm), which is a representation of the amount of radioactivity present in the sample.

***Immunoassay Methods*** Radioimmunoassay (RIA) allows measurement of biologically active materials which are not detectable by traditional cold-chemistry techniques. RIAs can be used to measure molecules that cannot be radiolabeled to detectable levels in vivo. They also are used for molecules unable to fix complement when bound to antibodies or they can be used to identify cross-reacting antigens that compete and bind with the antibody.

Competitive inhibition of radiolabeled hormone antibody binding by unlabeled hormone (either as a standard or an unknown mixture) is the principle of most RIAs. A standard curve for measuring antigen (hormone) binding to antibody is constructed by placing known amounts of radiolabeled antigen and the antibody into a set of test tubes. Varying amounts of unlabeled antigen are added to the test tubes. Antigen–antibody complexes are separated from the antigen and the amount of radioactivity from each sample is measured to detect how much unlabeled antigen is bound to the antibody. Smaller amounts

of radiolabeled antigen–antibody complexes are present in the fractions containing higher amounts of unlabeled antigen. A standard curve must be constructed to correlate the percentage of radiolabeled antigen bound with the concentration of unlabeled antigen present.

Two methods are commonly employed in RIAs to separate antigen–antibody complexes. The first, the double-antibody technique, precipitates antigen–antibody complexes out of solution by utilizing a second antibody, which binds to the first antibody. The second most commonly used method is the dextran-coated activated charcoal technique. Addition of dextran-coated activated charcoal to the sample followed immediately by centrifugation absorbs free antigen and leaves antigen–antibody complexes in the supernatant fraction. This technique works best when the molecular weight of the antigen is 30 kDa or less. Also, sufficient carrier protein must be present to prevent adsorption of unbound antibody.

Once a standard curve has been constructed, the RIA can determine the concentration of hormone in a sample (usually plasma or urine). The values of hormone levels are usually accurate using the RIA, but certain factors (e.g., pH or ionic strength) can affect antigen binding to the antibody. Thus similar conditions must be used for the standard and the sample.

Problems of RIAs include lack of specificity. This problem is usually due to nonspecific cross-reactivity of the antibody. RIA represents an analytical approach of great sensitivity. Unlike assays that often require large amounts of tissue (or blood), the greater sensitivity of the RIAs or monoclonal antibody techniques can be achieved using small samples of biological fluids. Some of these RIA methodologies are more useful than others and to some extent depend on the degree of hormonal cross-reactions or, in the case of monoclonal antibody methods, their degree of sensitivity.

Enzyme-linked immunosorbent assay (ELISA) is comparable to the immunoradiometric assay except that an enzyme tag is attached to the antibody instead of a radioactive label. ELISAs have the advantage of no radioactive materials and produce an end product that can be assessed with a spectrophotometer. The molecule of interest is bound to the enzyme-labeled antibody, and the excess antibody is removed for immunoradiometric assays. After excess antibody has been removed or the second antibody containing the enzyme has been added (two-site assay), the substrate and cofactors necessary are added in order to visualize and record enzyme activity. The level of molecule of interest present is directly related to the level of enzymatic activity. The sensitivity of the ELISAs can be enhanced by increasing the incubation time for producing substrate.

Immunoradiometric assays (IRMAs) are like RIAs in that a radiolabeled substance is used in an antibody–antigen reaction, except that the radioactive label is attached to the antibody instead of the hormone. Furthermore, excess of antibody, rather than limited quantity, is present in the assay. All the unknown antigen becomes bound in an IRMA rather than just a portion, as in a RIA; IRMAs are more sensitive. In the one-site assay, the excess antibody

that is not bound to the sample is removed by addition of a precipitating binder. In a two-site assay, a molecule with at least two antibody-binding sites is adsorbed onto a solid phase, to which one of the antibodies is attached. After binding to this antibody is completed, a second antibody labeled with $^{125}I$ is added to the assay. This antibody reacts with the second antibody-binding site to form a "sandwich" composed of antibody-hormone-labeled antibody. The amount of hormone present is proportional to the amount of radioactivity measured in the assay.

With enzyme-multiplied immunoassay technique (EMIT) assays, enzyme tags are used instead of radiolabels. The antibody binding alters the enzyme characteristics, allowing for measurement of target molecules without separating the bound and free components (i.e., homogeneous assay). The enzyme is attached to the molecule being tested. This enzyme-labeled antigen is incubated with the sample and with antibody to the molecule. Binding of the antibody to the enzyme-linked molecule either physically blocks the active site of the enzyme or changes the protein conformation so that the enzyme is no longer active. After antibody binding occurs, the enzyme substrate and cofactor are added, and enzyme activity is measured. If the sample contains subject molecules, it will compete with enzyme-linked molecules for antibody binding, enzyme will not be blocked by antibody, and more enzyme activity will be measurable.

Most protein drug entities can now be assessed using monoclonal antibody (MAb) techniques. It is possible to produce antisera containing a variety of polyclonal antibodies that recognize and bind many parts of the molecule. Polyclonal antisera can create some nonspecificity problems such as cross-reactivity and variation in binding affinity. Therefore it is oftentimes desirable to produce a group of antibodies that selectively bind to a specific region of the molecule (i.e., antigenic determinant). In the past, investigators produced antisera to antigenic determinants of the molecule by cleaving the molecule and immunizing an animal with the fragment of the hormone containing the antigenic determinant of interest. This approach solved some problems with cross-reactivity of antisera with other similar antigenic determinants, but problems were still associated with the heterogeneous collection of antibodies found in polyclonal antisera.

The production of MAbs offers investigators a homogenous collection of antibodies that could bind selectively to a specific antigenic determinant with the same affinity. In addition to protein isolation and diagnostic techniques, MAbs have contributed greatly to RIAs.

While MAbs offer a highly sensitive, specific method for detecting antigen, sometimes increasing MAb specificity compromises affinity of the antibody for the antigen. In addition, there is usually decreased complement fixation, and costs are usually high for preparing and maintaining hybridomas that produce MAbs (Table 13).

The monoclonal antibody techniques provide a means of producing a specific antibody for binding antigen. This technique is useful for studying protein

**TABLE 13 Advantages and Disadvantages of Monoclonal Antibodies Compared to Polyclonal Antisera**

| Advantages | Disadvantages |
|---|---|
| Sensitivity | Overly specific |
| Quantities available | Decreased affinity |
| Immunologically defined | Diminished complement fixation |
| Detection of neoantigens on cell membrane | Labor intensive; high cost |

structure relations (or alterations) and has been used for devising specific RIAs.

***Metabolism and Elimination*** Biologics are usually not excreted unchanged in urine. They are degraded to small peptides and individual amino acids with pathways equally and generally understood for endogenous compounds. Their metabolites (amino acids) are reutilized in the endogenous amino acid pool for the de novo biosynthesis of structural or functional body proteins.

The metabolism of biologics is highly dependent on structure (including sugars), charge (density and distribution), size, and hydrophilicity/lipophilicity. Sites of metabolism of biologics are the liver, the kidneys, and the blood and the extravascular sites of administration. In the liver, hepatocytes are mainly responsible for the catabolism of biologics using carrier-mediated membrane transport as well as endocytosis/pinocytosis for transport process. Kidneys play a major role in the catabolism of many small polypeptides. After being filtered by the glomeruli, some proteins are reabsorbed by the proximal tubule by endocytosis, while small amino acid chains are hydrolyzed at the brush border. Controversy exists surrounding glomerular filtration selectivity regarding size, molecular conformation, and charge of the protein (Tang et al., 2004). The often-observed incomplete bioavailability of biologics after extravascular injection can be attributed to local metabolism. Catabolism at extravascular sites has been observed, for example, for insulin, calcitonin, and interferon-β (for review see, Mohler et al., 1992).

Several biologics, especially antibodies, show high interindividual variation of pharmacokinetics parameters, which is mainly clearance driven. It is now clear that MAbs which target cellular antigens have far more complex, non-linear pharmacokinetics such that the half-life of these drugs can be both dose and time dependent (Tobo et al., 2004). When antigen concentration is high, half-life is short because the MAb is rapidly cleared from the blood through antigen–MAb interaction. As the antigen is depleted, clearance decreases and half-life is consequently prolonged. As the MAb accumulates, a new steady state is reached. Eventually, the target is totally depleted, at which time the clearance of MAb will be at its slowest. At this point, half-life will be at its longest, approaching the half-life of endogenous immunoglobulin G (IgG)

(~21 days). More probable than total target depletion is saturation of the target–MAb binding with similar consequences.

Due to catabolism of proteins to (mostly) endogenous amino acids, classical biotransformation studies for small molecules are not needed. Additionally, limitations of current analytical methods to detect and distinguish metabolites and the putative lack of pharmacological or toxicological activity of the metabolites remain obstacles. Similarly, mass balance studies usually used to determine the excretion pathways of small molecules (and their metabolites) are not used for biologics.

The majority of therapeutic biologics, especially after chronic administration, elicit an immune response in test animal species and often also in humans. This is an inherent property when administering nonhuman sequence proteins of sufficient size. Initial success in reducing immunogenicity has been achieved by replacing biologics obtained from nonhuman sources with human sequences. Antibody formation can also occur in immunocompetent recipients after treatment with products derived from human sera and tissues and also with recombinant human proteins that are identical or nearly identical in sequence to native human proteins. The mechanism for generation of antibodies of recombinant human proteins is not well understood. In most cases, the underlying mechanism is the breaking of immune tolerance that typically exists in self-antigens. Other reasons for immunogenicity relate to manufacturing, formulation, and storage (e.g., aggregates). These are especially addressed when modifications of these processes are performed which might influence the physicochemical properties of the product. An immune response to a product does not mean it cannot be developed. However, the development and use of a product may be complicated and, in rare cases, also impossible. In certain cases, some patients develop antibodies which neutralize the biological activity of the therapeutic product and become unresponsive to treatment. Alterations in the pharmacokinetic profile due to immune-mediated clearance mechanisms may affect the pharmacokineticprofiles and the interpretation of the preclinical toxicity data. Last but not least, safety issues like immunomediated toxicity may be raised. Detection and characterization of the immune response in patients are expected by the authorities [U.S. Food and Drug Administration (FDA) (2002)].

The development and usefulness of appropriate animal models for testing immunogenicity are still unclear. Conventional animal models have poorly predicted immunogenicity problems in humans. One limitation of traditional animal models is that tolerance, a key aspect of the immune response, is highly species specific. However, characterization of the immune response in research and preclinical development is necessary to get a valid interpretation of the preclinical efficacy and safety data.

Not only the compound structure itself but also the administration route may affect immunogenicity. Extravascular injection is known to stimulate antibody formation more than IV application. This is most likely due to

increased immunogenicity of protein aggregates and precipitates formed at the injection site.

To lower the systemic clearance and increase elimination half-lives, several strategies have been developed, including polyethylene glycol (PEG) attachment (PEGylation), glycosylation, or fusion to proteins with decreasing clearance and prolonged serum half-lives. PEGylation improves the pharmacokinetic behaviors by increasing the effective size of the protein, with most significant effects for proteins smaller than 70 kDa. However, conjugation of the protein may also alter receptor affinity and biodistribution, changing the concentration–response profile for the protein independent of effects on pharmacokinetics. PEGylation can also reduce immunogenicity and aggregation.

### 4.2 Sampling Methods and Intervals

1. *Blood* Since blood (plasma and serum) is the most easily accessible body compartment, the blood concentration profile is most commonly used to describe the time course of drug disposition in the animal. With the development of sensitive analytical methods that require small volumes (100–200 μL) of blood, ADME data from individual rats can be obtained by serial sample collection. Numerous cannulation techniques have been utilized to facilitate repeated blood collection, but the animal preparation procedures are elaborate and tedious and are incompatible with prolonged sampling periods in studies involving a large number of animals. In contrast, noncannulation methods such as collection from the tail vein, orbital sinus, or jugular vein are most practical. Significant volumes of blood can be obtained from the intact rat by cardiac puncture, although this method can cause shock to the animal system and subsequent death.

Blood collection from the tail vein is a simple and rapid, nonsurgical method which does not require anesthesia. A relatively large number of serial samples can be obtained within a short period of time. However, this method is limited to relatively small sample volumes (~250 μL per sample). Although larger volumes can be obtained by placing the rat in a warming chamber, this procedure could significantly influence the disposition of the test compound and therefore is not recommended for routine studies. Blood collected from the cut tail has been shown to provide valid concentration data for numerous compounds.

The rat is placed in a suitable restrainer with the tail hanging freely. The tail is immersed in a beaker of warm water (37–40 °C) for 1–2 min to increase the blood flow. Using surgical scissors or a scalpel, the tail is completely transected approximately 5 mm above the tip. The tail is then gently "milked" by sliding the fingers down the tail from its base. It should be noted that excessive "milking" could cause damage to the blood capillaries or increase the white cell count in the blood. A heparinized micropipette of desired capacity (25–

250 μL) is held at a 30°–45° downward angle in contact with the cut end of the tail. This allows blood to fill the micropipette by capillary action. Application of gentle pressure with a gauze pad for approximately 15 s is sufficient to stop bleeding. A sufficient number of serial blood samples may be obtained to adequately describe the blood level profile of a compound.

If plasma is required, the blood may be centrifuged after sealing one end of the filled micropipette and placing it in a padded centrifuge tube. The volume of plasma is determined by measuring the length of plasma as a fraction of the length of the micropipette multiplied by the total capacity of the pipette. The tube is then broken at the plasma–red blood cell interface and the sample is expelled using a small bulb. If serum is needed, the blood should be collected without using anticoagulants in the sampling tube.

Serial blood samples can also be collected from the orbital sinus, permitting rapid collection of larger (1–3-mL) samples.

2. *Excreta* Excretion samples commonly collected from the rat include urine, feces, bile, and expired air. By using properly designed cages and techniques, the samples can be completely collected so that the mass balance is readily determined. These samples also serve to elucidate the biotransformation characteristics of the compound.

These samples can be easily collected through the use of suitable metabolism cages. Since rodents are coprophagic, the cage must be designed to prevent the animal from ingesting the feces as it is passed. Other main features of the cage should include the ability to effectively separate urine from feces with minimal cross-contamination, a feed-and-water system that prevents spillage and subsequent contamination of collected samples, and collection containers that can be easily removed without disturbing the animal. Also, the cage should be designed so that it can be easily disassembled for cleaning or autoclaving.

Following dose administration, rats are placed in individual cages. The urine and feces that collect in containers are removed at predetermined intervals. The volume of urine and the weight of feces are measured. After the final collection, the cage is rinsed, normally with ethanol or water, to assure complete recovery of excreta. If the rats are also used for serial blood sampling, it is important that bleeding be performed inside the cage to avoid possible loss of urine or feces.

3. *Bile* The bile is the pathway through which an absorbed compound is excreted in the feces. In order to collect this sample, surgical manipulation of the animal is necessary (Wang and Reuning, 1994).

4. *Expired Air* For $^{14}C$-labeled chemicals, the tracer carbon may be incorporated in vivo into carbon dioxide, a possible metabolic product. Therefore, when the position of the radiolabel indicates the potential for biological instability, a pilot study to collect expired air and monitor its radioactivity content should be conducted prior to initiating a full-scale study. Expired air studies should also be performed in situations where the radiolabel has been

postulated to be stable but analyses of urine and feces from the toxicokinetic study fail to yield complete recovery (mass balance) of the dose.

Following drug administration, the rat is placed in a special metabolism cage. Using a vacuum pump, a constant flow of room air (~500 mL min$^{-1}$) is drawn through a drying column containing anhydrous calcium sulfate impregnated with a moisture indicator (cobalt chloride) and passed into a second column containing Ascarite II, where it is rendered carbon dioxide free. The air is then drawn in through the top of the metabolism cage. Exhaled breath exiting the metabolism cage is passed through a carbon dioxide adsorption tower, where the expired $^{14}CO_2$ is trapped in a solution such as a mixture of 2-ethoxyethanol and 2-aminoethanol (2:1). The trapping solution is collected, replaced with fresh solution, and assayed at designated times postdose so that the total amount of radioactivity expired as labeled carbon dioxide can be determined.

5. *Milk* The study of passage of a xenobiotic into milk serves to assess the potential risk to breast-fed infants in the absence of human data. The passage into milk can be estimated as the milk–plasma ratio of drug concentrations at each sampling time or that of the area under the curve (AUC). Approximately 30 rats in their first lactation are used. The litter size is adjusted to about 10 within 1–2 days following parturition. The test compound is administered to the mothers 8–10 days after parturition. The rats are then divided into groups for milk and blood collection at designated times postdose. All sucklings are removed from the mother rats several hours before milking. Oxytocin, 1 IU per rat, is given intramuscularly 10–15 min before each collection of milk to stimulate milk ejection. The usual yield of milk is about 1 mL from each rat. Blood is obtained immediately after milking. In order to minimize the number of animals used, the sucklings can be returned to the mother rat, which can then be milked again 8–12 h later.

In all the fluid-sampling techniques above, the limitations of availability should be kept in mind. Table 14 presents a summary of such availability for the principal model species.

For topical exposures, determining absorption (into the skin and into the systemic circulation) requires a different set of techniques. For determining how much material is left, skin washing is required. There are two components

**TABLE 14 Approximate Volumes of Pertinent Biological Fluids in Adult Laboratory Animals**

| Fluid | Rat | Mouse | Dog | Rabbit | Monkey |
|---|---|---|---|---|---|
| Blood (mL kg$^{-1}$) | 75 | 75 | 70 | 60 | 75 |
| Plasma (mL kg$^{-1}$) | 40 | 45 | 40 | 30 | 45 |
| Urine (mL kg$^{-1}$ day$^{-1}$) | 60 | 50 | 30 | 60 | 75 |
| Bile (mL kg$^{-1}$ day$^{-1}$) | 90 | 100 | 12 | 120 | 25 |

to skin washing in the recovery of chemicals. The first component is the physical rubbing and removal from the skin surface. The second component is the surfactant action of soap and water. However, the addition of soap affects the partitioning. Some compounds may require multiple successive washing with soap and water applications for removal from skin.

Skin tape stripping can be used to determine the concentration of drug in the stratum corneum at the end of a short application period (30 min) and by linear extrapolation predicts the percutaneous absorption of that chemical for longer application periods. The chemical is applied to the skin of animals or humans, and after a 30-min skin contact application time, the stratum corneum is blotted and then removed by successive tape applications. The tape strippings are assayed for chemical content. There is a linear relationship between this stratum corneum reservoir content and percutaneous absorption. The major advantages of this method are (1) the elimination of urinary and fecal excretion to determine absorption and (2) the applicability to nonradiolabeled determination of percutaneous absorption because the skin strippings contain adequate chemical concentrations for nonlabeled assay methodology.

Finally, a complete determination of the distribution and potential departing of a chemical and its metabolites requires some form of measurement or sampling of tissues/organs. Autoradiography provides a nonquantitative means of doing such, but quantitation requires actual collection and sampling of tissues. Table 7 provided guidance as to the relative percentage of total body mass that the organs constitute in the common model species.

***Sampling Interval*** To be able to perform valid toxicokinetic analysis, it is necessary to properly collect not only samples of appropriate biological fluids but also a sufficient number of samples at the current intervals. Both of these variables are determined by the nature of the answers sought. Useful parameters in toxicokinetic studies are $C_{max}$, which is the peak plasma test compound concentration; $T_{max}$, which is the time at which the peak plasma test compound concentration occurs; $C_{min}$, which is the plasma test compound concentration immediately before the next dose is administered; AUC, which is the area under the plasma test compound concentration–time curve during a dosage interval; and $t½$, which is the half-life for the decline of test compound concentrations in plasma. The samples required to obtain these parameters are shown in Table 15. The $C_{min}$ requires one blood sample immediately before a dose is given and provides information on accumulation. If there is no accumulation in plasma, the test compound may not be detected in this sample.

Several $C_{min}$ samples are required at intervals during the toxicity study to check whether accumulation is occurring. Assume $C_T$ is a blood sample taken at a chosen time after dosing and provides proof of absorption as required by the good labor practice (GLP) regulations, but little else. In addition, $C_{max}$ requires several blood samples to be taken for its accurate definition, as does

**TABLE 15 Blood Samples Required so That Certain Toxicokinetic Parameters Can be Obtained and Calculated**

| Parameter | Blood Sample Required | Information Obtained |
|---|---|---|
| $C_{min}$ ($C_{24}$) | 24 h | Accumulation |
| CT | T hours | Proof of absorption |
| $C_{max}$ ($C$ peak) | Several[a] | Rate of absorption |
| $T_{max}$ ($T$ peak) | Several[a] | Rate of absorption |
| AUC | Several[a] | Extent of absorption |
| $t_{½}$ | Several[a] | Various |
| Accumulation ratio | Several after first and repeated doses | Extent of accumulation |

[a]Several samples to define concentration–time profile.

$T_{max}$: these two parameters provide information on rate of absorption. The AUC also requires several blood samples to be taken so that it can be calculated: It provides information on extent of absorption. The half-life t½ requires several samples to be taken during the terminal decline phase of the test compound concentration–time curve: This parameter provides information on various aspects, such as change in the kinetics of the test compound during repeated doses or at different dose levels. Depending on the other parameters obtained, the accumulation ratio can be calculated from $C_{min}$, $C_{max}$, and/or AUC when these are available after the first dose and after several doses to steady state.

Operational and metabolic considerations generally make urine sampling and assay of limited value for toxicokinetic purposes.

***Study Type*** Metabolic and pharmacokinetic data from a rodent species and a nonrodent species (usually the dog) used for repeat-dose safety assessments (14, 28, or 90 days or 6 months) are recommended. If a dose dependency is observed in metabolic and pharmacokinetic or toxicity studies with one species, the same range of doses should be used in metabolic and pharmacokinetic studies with other species. If human metabolism and pharmacokinetic data also are available, this information should be used to help select test species for the full range of toxicity tests and may help to justify using data from a particular species as a human surrogate in safety assessment and risk assessment.

Metabolism and pharmacokinetic studies have greater relevance when conducted in both sexes of young adult animals of the same species and strain used for other toxicity tests with the test substance. The number of animals used in metabolism and pharmacokinetic studies would be sufficient to reliably estimate population variability. This usually means a separate (but parallel) set of groups of animals in rodent studies. A single set of IV and oral dosing results from adult animals, when combined with some in vitro kinetic results, may provide an adequate data set for the design and interpretation of short-term, subchronic and chronic toxicity studies.

Studies in multiple species may clarify what appear to be contradictory findings in toxicity studies (i.e., equal milligram-per-kilogram body weight doses having less effect in one species that in another). If disposition and metabolite profiles are found to be similar, then differences in responses among species could more reliably be attributed to factors other than differences in metabolism. Studies of the pharmacokinetics and metabolism of a substance in neonatal and adolescent animals provide information about any changes in metabolism associated with tissue differentiation and development. Animals with fetuses of known gestational age should be used for determining the disposition of the test substance in the fetus. Dosage is by (to the maximum extent possible) the intended clinical route.

An acute IV study can provide accurate rates of metabolism—without interference from intestinal flora—plus rates of renal and biliary elimination if urine and bile are collected. This route also avoids the variability in delivered dose associated with oral absorption and ensures that the maximum amount of radiolabel is excreted in the urine or bile for purposes of detection. Once IV data and parameters are available, they can be used with plasma concentrations from limited oral studies to compute intestinal absorption via the ratio of areas under the (plasma and/or urine) curves or via simulations of absorption with GI absorption models.

In single-dose pharmacokinetic studies of oral absorption, the primary concerns are with the extent of absorption and peak plasma or target tissue concentrations of the test substance. If the test vehicle affects gastric emptying, it may be necessary to use both fasted and nonfasted animals for pharmacokinetic studies.

Blood [red blood cells (RBCs), plasma, and serum], urine, and feces are the most commonly collected samples. In addition, a few representative organ and tissue samples should be taken, such as liver, kidney, fat, and suspected target organs. Sampling times should depend on the substance being tested and the route of administration. In general, an equal number of blood samples should be taken in each phase of the concentration-versus-time curve. Intravenous studies usually require much shorter and more frequent sampling than is required for oral dosing. Time spacing of samples will depend on the rates of uptake and elimination. In a typical IV study, blood and tissue samples are taken in a "powers of 2" series, that is, samples at 2, 4, 8, 16, and 30 min and 1, 2, 4, 8, and 16 h. Similar coverage could be obtained with only seven time points by using a "powers of 3" series: 3, 9, and 30 min and 1, 3, 9, and 24 h. Oral dosing studies usually extend to at least 72 h. Such a sampling scheme would provide data coverage for evaluation of absorption, elimination, enterohepatic recirculation, and excretion processes.

The number of animals used in metabolism and pharmacokinetic studies should be large enough to reliably estimate population variability. In the case for rats and mice, tissue and/or blood sample size is usually the limiting factor: Analysis of the substance may require 1 mL or more blood, but it is difficult to obtain multiple blood samples of this size from one animal. As a consequence,

a larger number of animals is required (three to four per time point, seven to nine time points) when small rodents are used. Such an approach has the advantage of allowing limited sampling of critical tissues (e.g., liver, fat) at each time point, an option which is usually unavailable with large animals. The use of humans and large animals generally permits collection of multiple (serial) blood samples. For outcrossing populations like humans and large animals, individual differences in the rates of biotransformation are likely to be greater than those of inbred rodent populations; under these circumstances, more samples per sex per group may be needed to reliably estimate variability.

Individual metabolism cages are recommended for collecting urine and feces in oral dosing studies. Excreta should be collected for at least five elimination half-lives of the test substance. When urine concentrations will be used to determine elimination rates, sampling times should be less than one elimination half-life (taken directly from the bladder in IV studies); otherwise, samples should be taken at equal time intervals.

The results of the preliminary biotransformation/kinetic study together with the current regulatory metabolism studies and the 28- and 90-day studies should allow the selection of a relatively small number of appropriate tissues and/or fluids for monitoring purposes. Satellite groups of animals will provide the material for analysis. Methods must be developed to analyze non-radioactive test chemical. Obviously it is important to monitor blood. It is accessible and convenient, and in certain circumstances sequential sampling from the same animal may be important. The most useful aspect of blood is that the results can be compared with those obtained in humans (see below). It is important, however, not to be constrained by this aspect. The most relevant tissues and body fluids should also be analyzed. These are target organs (if known) and indicator organs, tissues, or fluids, that is, those in which the concentration of pesticide or metabolite is a measure of that in the whole animal. In cases where distribution varies with dose (if shown in the preliminary study), a larger number of organs/tissues would be chosen for monitoring.

Whether the parent drug or metabolite (or both) is chosen for analysis depends on the preliminary study. In principle, analysis for the parent compound should always be carried out; however, there are situations (e.g., rapid metabolism) when this is quite futile and a major retained metabolite should be used. Covalently bound metabolites are addressed below.

Four occasions may be adequate for monitoring:

(i) One month (equilibrium between intake of chemical and elimination of metabolites should be established; the time relates to the 28-day preliminary study)
(ii) Three months (confirmation of results at one month; relates to the 90-day study)
(iii) One year (coincides with the interim kill)
(iv) Two years (effects of age; coincides with termination of study)

Consideration should be given to the analysis of moribund animals.

***In Vitro Studies*** In vitro measurements employing enzymes, subcellular organelles, isolated cells, and perfused organs may be used to augment the dose–response information available from less extensive metabolic and pharmacokinetic studies. Because in vitro systems generally are less complex than whole animals, elucidation of a test compound's metabolic pathways and the pathways' kinetic characteristics may be facilitated. Such systems can be used to measure binding, adduct, and conjugate formation, transport across cell membranes, enzyme activity, enzyme substrate specificity, and other singular objectives. Biochemical measurements that can be made using in vitro systems include intrinsic clearances of enzymes in an organ or tissue, kinetic constants for an enzyme, binding constants, and the affinity of the test compound and its metabolites for the target macromoecules. The activity of a hepatic drug-metabolizing enzyme in vivo may be approximated by kinetic constants that are calculated from in vitro studies; when a first-order approximation is used, the ratio of $V_{max}$ to $K_m$ is equal to the intrinsic clearance of the drug. In vitro measurements made using readily accessible tissues and body fluids from animals and humans may also be useful in elucidating mechanisms of toxicity.

***Analysis of Data*** Data from all metabolism and pharmacokinetic studies should be analyzed with the same pharmacokinetic model and results should be expressed in the same units. Concentration units are acceptable if the organ or sample size is reported, but percent of dose/organ is usually a more meaningful unit. In general, all samples should be analyzed for metabolites that cumulatively represent more than 1% of the dose.

A variety of rate constants and other parameters can be obtained from IV and oral dosing data sets provided that good coverage of the distribution, elimination, and absorption (oral dose) phase is available. Typical parameters calculated to characterize the disposition of a test substance are half-lives of elimination and absorption; area under the concentration-versus-time curve for blood; total body, renal, and metabolic clearance(CL); volume of distribution ($V_d$); bioavailability ($F$); and mean residence and absorption times (MAT, MRT). Some of these parameters, such as half-lives and elimination rates, are easily computed from one another; the half-life is more easily visualized than the rate constant.

Computation of oral absorption ($k_a$) and elimination ($E$) rates is often complicated by the "flip-flop" of the absorption and elimination phases when they differ by less than a factor of 3. Because of these analysis problems, computation of absorption and elimination rates should not be attempted on the basis of oral dosing results alone.

Blood/tissue uptake rates ($k_{jl}$) can often be approximated from data at early ($t < 10$ min) time points in IV studies provided that the blood has been washed from the organ (e.g., liver) or the contribution from blood to the tissue residue is subtracted (fat). High accuracy is not usually required since these parameters can be optimized to fit the data when they are used in more complex models. Tissue/blood recycling rates ($k_{lj}$) and residence times can

be computed from partition coefficients if estimates of uptake rates are available.

Tissue/blood partition coefficients ($R_{jl}$) should be determined when a steady state has been achieved. Estimates based on samples obtained during the elimination phase following a single dose of the test substance may lead to underestimates of this ratio in both eliminating and noneliminating tissues unless its half-life is very long. Correction of these values for elimination has been described by several authors.

It may be important to determine the degree of plasma protein and RBC binding of the test substance; calculation of blood clearance rates using plasma or serum concentrations of the substance that have not been adjusted for the degree of binding may under- or overestimate the true rate of clearance of the test substance from the blood. This is usually done through experiments in vitro.

Two classical methods used in the analysis of pharmacokinetic data are the fitting of sums of exponential functions (two- and three-compartment mammillary models) to plasma and/or tissue data and, less frequently, the fitting of arbitrary polynomial functions to the data (noncompartmental analysis).

Noncompartmental analysis is limited in that it is not descriptive or predictive; concentrations must be interpolated from data. The appeal of noncompartmental analysis is that the shape of the blood concentration-versus-time curve is not assumed to be represented by an exponential function and, therefore, estimates of metabolic and pharmacokinetic parameters are not biased by this assumption. In order to minimize errors in parameter estimates that are introduced by interpolation, a large number of data points that adequately define the concentration-versus-time curve are needed (Gabrielsson and Weiner, 1997).

Analysis of data using simple mammillary, compartmental models allows the estimation of all of the basic parameters mentioned above if data for individual tissues are analyzed with one- or two-compartment models and combined with results from two- to three-compartment analyses of blood data. "Curve-stripping" analysis can be applied to such simple models through the use of common spreadsheet programs (i.e., LOTUS 1-2-3 or Excel) as long as a linear regression function is provided in the program. Optimization of the coefficients and exponents estimated may require the use of more sophisticated software: A number of scientific data analysis packages such as RS/1 and SigmaPlot have the necessary capabilities. Specialized programs such as NONLIN, CONSAM, or SIMUSOLV will be needed when more complex models must be analyzed. Coefficients and exponents from mammillary models can be used to calculate other parameters; however, they should not be taken too literally, since mammillary models assume that all inputs are to a central pool (blood) which communicates without limitation into other compartments. This approach does not include details such as blood flow limitations, anatomical volumes, or other physiological limits in the animal.

Physiologically based pharmacokinetic (PBPK) models were developed to overcome the limitations of simple mammillary models. Physiologically based models describe the disposition of test substances via compartmental models which incorporate anatomical, biochemical, and physiological features of specific tissues in the whole animal. The types of information added include organ-specific blood flows, volumes, growth models, and metabolism rates. Metabolic parameters often are obtained from in vitro studies (e.g., enzyme reaction rates in cultured hepatocytes, and plasma protein binding), while other parameters are becoming available as standard parameters in the literature. Parameters from mammillary models can be used to compute the value of parameters used in physiological pharmacokinetic models using tissue-specific blood flows, anatomical volumes, and other information (literature values). Estimation of parameters for a simple mammillary model is often the first data reduction step in creating a physiological model.

Because PBPK models are based on physiological and anatomical measurements and all mammals are inherently similar, they provide a rational basis for relating data obtained from animals to humans. Estimates of predicted disposition patterns for test substances in humans may be obtained by adjusting biochemical parameters in models validated for animals; adjustments are based on experimental results of animal and human in vitro tests and by substituting appropriate human tissue sizes and blood flows. Development of these models requires special software capable of simultaneously solving multiple (often very complex) differential equations, some of which were mentioned above. Several detailed descriptions of data analysis have been reported.

***Use of Data from Metabolism and Pharmacokinetic Studies*** Information from metabolism and pharmacokinetic studies can be used in the design and analysis of data from other toxicity studies. Some examples are described below.

***Design of Toxicity Studies*** The concentration-versus-time curve and peak and steady-state concentrations of the test substance in blood or plasma provide information on the distribution and persistence of the substance in the animal which may suggest essential elements in the design of the toxicity studies. For example, when metabolic and pharmacokinetic studies indicate that the test compound accumulates in the bone marrow, long-term toxicity tests should include evaluation of the test compound's effect on hematopoietic function and morphology. If a test compound is found to accumulate in milk, an investigator may need to plan to perform reproductive toxicity studies with in utero exposure and a nursing phase (cross-fostering study). In addition, information from metabolic and pharmacokinetic studies can be used to predict the amount of test compound that enters biological compartments (tissues, organs, etc.) that may not suffer a toxic insult but may serve as depots for indirect or secondary exposure.

### 4.3 Whole-Body Autoradiography

Autoradiography is the production of an image in a photographic emulsion by the emission from a radioactive element. The term autoradiography is preferred to that of radioautography. Prefixes are added to words to further classify the concept. Therefore, the process is "auto-" radiography for a "self-" radiograph and not a "radio-" autograph for one's transmitted signature (Waddell, 1972).

Whole-body autoradiography (WBA) has been used with increasing frequency as a means of identifying tissues which concentrate test substances. This technique allows a small number of animals (5–10) to be used for screening purposes with a minimal investment in manual labor. The FDA encourages the use of WBA with IV dosing as a means of screening and selecting tissues of greatest relevance for later oral dosing studies. Animals used for WBA should generally be sacrificed during primary consideration in selecting specific tissues.

The most comprehensive technique currently available for the initial survey of the distribution of a drug is that of WBA. The species of animals used include mice, rats, hamsters, monkeys, pigs, dogs, and ferrets. The most widely used animal has been the mouse, which has the advantages of requiring less isotope and being easier to section.

The animals are anesthetized and then frozen by immersion at various times after administration of the labeled compound in hexane or acetone cooled with dry ice. Since the freezing in the interior of the animals occurs slowly, large ice crystals form within these tissues, and hence subcellular localization of compounds is not possible.

The selection of times for freezing an animal after injection of a drug must be based on the information available on the rate of elimination of the compound from the animal by metabolism and excretion. In general, a geometric increase in time intervals is most useful. In order to have time intervals for comparison, we routinely have employed freezing times which are approximately multiples of three, namely 2, 6.5, and 20 min and 1, 3, 9, and 24 h. In certain cases, rapid elimination of the drug by the kidneys must be circumvented by ligation of the renal pedicles to avoid apparent localization from failure of the agent to reach equilibrium. An example would be the clearance of urea-$^{14}$C in pregnant mice.

The frozen animal is frozen into a block of carboxymethylcellulose ice on the microtome stage. Although the Jung, type K, microtome has been used, the Leitz, model 1300, sledge microtome is more suitable, for its smaller size allows it to be mounted in an ordinary commercial freezer instead of a walk-in freezer. The microtome stage must be designed for mounting in the vice of the front end of the stage.

Sections from 5 μm to approximately 80 μm thick are taken onto No. 800 Scotch tape (Minnesota Mining & Mfg. Co.). Before removal from the freezer, the sections must be allowed to dry thoroughly so that no ice remains which melts and allows movement of the isotope. After drying, if covered to prevent

condensation of moisture on the sections, the sections may be transferred from the freezer to room temperature.

X-ray films which produce the most satisfactory autoradiograms are Kodak industrial type AA and Gevaert Structurix D-7. Both are fine-grain films which have been demonstrated not to produce chemical artifacts. Approximately six times faster, Kodak No Screen and Kodirex may be used for rapid screening and timing of autoradiograms. However, they occasionally produce artifacts and should not be relied on for interpretation. Some investigators have used photographic emulsions such as Ilford G-5, 10 μm thick, preapplied to glass plates. The increased cost and likelihood of breakage, however, hardly justify the small improvement in resolution for whole-body sections.

Exposure of the photographic emulsion by the radioactivity of the tissue section should be at freezer temperatures to prevent autolysis of the tissue. After exposure of the X-ray film, sections with isotopes which have a long half-life may be placed against fresh X-ray film for additional sets of autoradiograms with either a longer or shorter exposure time. This procedure is useful for revealing relative concentrations of radioactivity for areas that have either very high or very low concentrations after the first exposure. When no further autoradiograms are needed, the section can be stained with histological dyes to verify localizations of radioactivity.

Compounds that fluoresce under ultraviolet light can be visualized in the tissue sections and their locations recorded with color film. Whole-body tissue sections can be used for histochemical localizations for comparison with the autoradiograms. Furthermore, the areas can be removed and extracted and the extract chromotographed to identify the chemical nature of the radioactivity revealed by the autoradiogram.

Although the whole-body technique will allow localization of an increased concentration of an isotope in a tissue or occasionally a cell type, other techniques must be used for single cells and subcellular localization. A nuclear tract plate is prepared by dipping the plate in a 12% solution of glycerine in absolute ethyl alcohol and allowing it to drain for 10 min in a vertical position before approximating the section on tape. After the emulsion is exposed, soaking in xylene removes the tape but leaves the section attached to the nuclear tract plate. The Ilford G-5 nuclear tract plates with 10-μm emulsions are most satisfactory. The increased resolution gained by the finer grained Ilford K and L emulsions is warranted only for tissues that are well preserved and relatively free of ice crystal artifacts. Kodak NTB emulsions seem to produce more pressure artifacts than the Ilford plates.

Comparison of various techniques of autoradiography for diffusible compounds clearly demonstrates that no solutions can be used in processing the tissue. These investigators have dried thin sections of liver and uterus at temperatures below –60 °C. These freeze-dried sections were dry mounted on microscope slides which had been precoated with either Kodak NTB-3 or NTB-10 emulsion. Other techniques which thawed the frozen section, embedded the tissue in paraffin, or dipped the section in liquid emulsion were

demonstrated to translocate diffusible compounds. Many other similar attempts have been and are currently being made to localize diffusible compounds by autoradiography at the electron microscope level.

## 5 PHYSIOLOGICALLY BASED PHARMACOKINETIC (PBPK) MODELING

Pharmacokinetic modeling is the process of developing mathematical explanations of absorption, distribution, metabolism, and excretion of chemicals in organisms. Two commonly used types of compartmental pharmacokinetic models are (a) data based and (b) physiologically based. The data-based pharmacokinetic models correspond to mathematical descriptions of the temporal change in the blood/tissue level of a xenobiotic in the animal species of interest. This procedure considers the organism as a single homogeneous compartment or as a multicompartmental system with elimination occurring in specific compartments of the model. The number, behavior, and volume of these hypothetical compartments are estimated by the type of equation chosen to describe the data and not necessarily by the physiological characteristics of the model species in which the blood/tissue concentration data were acquired.

Whereas these data-based pharmacokinetic models can be used for interpolation, they should not be used for extrapolation outside the range of doses, dose routes, and species used in the study on which they were based. In order to use the data-based models to describe the pharmacokinetic behavior of a chemical administered at various doses by different routes, extensive animal experimentation would be required to generate similar blood–time course data under respective conditions. Even within the same species of animal, the time-dependent nature of critical biological determinants of the disposition (e.g., tissue glutathione depletion and resynthesis) cannot easily be included or evaluated with the data-based pharmacokinetic modeling approach. Further, due to the lack of actual anatomical, physiological, and biochemical realism, these data-based compartmental models cannot easily be used in interspecies extrapolation, particularly to predict pharmacokinetic behavior of chemicals in humans. These various extrapolations, which are essential for the conduct of dose–response assessment of chemicals, can be performed more confidently with a physiologically based pharmacokinetic modeling approach. This chapter presents the principles and methods of physiologically based pharmacokinetic modeling as applied to the study of toxicologically important chemicals.

PBPK modeling is the development of mathematical descriptions of the uptake and disposition of chemicals based on quantitative interrelationships among the critical biological determinants of these processes. These determinants include partition coefficients, rates of biochemical reactions, and physiological characteristics of the animal species. The biological and mechanistic basis of the PBPK models enable them to be used, with limited animal experimentation, for extrapolation of the kinetic behavior of chemicals from high

dose to low dose, from one exposure route to another, and from test animal species to people.

The development of PBPK models is performed in four interconnected steps: model representation, model parameterization, model stimulation, and model validation. Model representation involves the development of conceptual, functional, and computational descriptions of the relevant compartments of the animal as well as the exposure and metabolic pathways of the chemical. Model parameterization involves obtaining independent measures of the mechanistic determinants, such as physiological, physicochemical, and biochemical parameters, which are included in one or more of the PBPK model equations. Model simulation involves the prediction of the uptake and disposition of a chemical for defined exposure scenarios using a numerical integration algorithm, simulation software, and a computer. Finally, the model validation step involves the comparison of the a priori predictions of the PBPK model with experimental data to refute, validate, or refine the model description and the characterization of the sensitivity of tissue dose to changes in model parameter values. After appropriate testing and validation PBPK models can be used to conduct extrapolations of the pharmacokinetic behavior of chemicals from one exposure route/scenario to another, from high dose to low dose, and from one species to another.

The PBPK model development for a chemical is preceded by the definition of the problem, which in toxicology may often be related to the apparent complex nature of toxicity. Examples of such apparent complex toxic responses include nonlinearity in dose–response, sex/species differences in tissue response, differential response of tissues to chemical exposure, qualitative and/or quantitative difference responses for the same cumulative dose administered by different routes/scenarios, and so on. In these instances, PBPK modeling studies can be utilized to evaluate the pharmacokinetic basis of the apparent complex nature of toxicity induced by the chemical. One of the values of PBPK modeling, in fact, is that accurate description of target tissue dose often resolves behavior that appears complex at the administered dose level.

The principal application of PBPK models is in the prediction of the target tissue dose of the toxic parent chemical or its reactive metabolite. Use of the target tissue dose of the toxic moiety of a chemical in risk assessment calculations provides a better basis of relating to the observed toxic effects than the external or exposure concentration of the parent chemical. Because PBPK models facilitate the prediction of target tissue dose for various exposure scenarios, routes, doses, and species, they can help reduce the uncertainty associated with the conventional extrapolation approaches. Direct application of modeling includes:

- High-dose/low-dose extrapolation
- Route–route extrapolation
- Exposure scenario extrapolation
- Interspecies extrapolation

## 6 POINTS TO CONSIDER

Stereoisomerism will influence metabolism and toxicity. For example, Lu et al. (1998) reported a comparison of (*S*)-(–)ifosfamide and (*R*)-(+)-ifosfamide. They demonstrated that there were significant differences between the two stereoisomers with regard to pharmacokinetic behavior and major metabolite formation, as shown in Table 16.

In addition, treatment of animals with phenobarbital not only increased overall rates of metabolism and clearance but also shifted the metabolite patterns. One of the more common methods used for determining an exposure to (or the amount of a metabolite produced) is to determine an AUC for the metabolite. Further, one of the more common methods for representing a racemically preferred metabolite is to calculate the ratio of *R* to *S*. For example, the 3-decholoro metabolite of ifosfamide was produced in higher amounts from the *R* enantiomer while the 2-decholorometabolite was the major metabolite produced from the *R* enantiomer in naive animals. Treatment with phenobarbital shifted the metabolism so that the 3-dechloro metabolite was no longer the major metabolite for the *S* enantiomer.

## 7 BIOLOGICALLY DERIVED MATERIALS

The progress and products of biotechnology have brought some new challenges to the assessment of pharmacokinetics and toxicokinetics. While the reasons for needing these data (demonstrating exposure, displaying dose dependency, correlating any findings of toxicity to exposure, and determining steady state for systemic agent levels) are certainly as compelling as with traditional drugs, there are a whole set of special problems involved (Baumann, 2006; Dennis et al., 2002).

These special concerns for biologically derived products are:

**TABLE 16 Example of Stereoselective Differences in Metabolism (*R*) versus (*S*) Ifosfamide**

| Parameter | Phenobarb | *R* | *S* | *R*/*S* |
|---|---|---|---|---|
| Term half-life | – | 34.3 | 41.8 | .820 |
| (min) | + | 19.8 | 19.41 | 1.02 |
| AUC | – | 4853 | 6259 | .820 |
| (μM·min) | + | 1479 | 1356 | 1.03 |
| 2-Dehloro metabolite | – | 799 | 2794 | .287 |
| AUC | + | 229 | 1205 | .186 |
| 3-Dehloro metabolite | – | 1380 | 996 | 1.41 |
| AUC | + | 192 | 1175 | .159 |

*Note*: Animals were pretreated with phenobarbital (80 mg·kg) for four days.

*Source*: Adapted from Lu et al., 1998.

**Assay Sensitivity/Specificity**

- Needs to be at 1 ng $mL^{-1}$ or lower.
- Cross-reactivity to native protein may confound results.
- If test article is the same as native protein, how do you tell the difference?
- Western blot can be used to demonstrate specificity.
- Antibody interference may occur with assay.

**Low Systemic Levels**

- Rapid metabolism: Metabolites may be endogenous proteins or amino acids.
- Extensive metabolism: Metabolites may be incorporated into cell structures rapidly.
- Rapid distribution
- Rapid hepatic clearance
- Route of administration may bypass first-pass metabolism:

  Subcutaneuos

  Intracerebroventricular, intrathecal

  Buccal

**Endogenous Protein**

- May cross react and lead to false-positive blood levels.
- Can radiolabel to tell difference between administered molecule and endogenous molecule:

  However, the label may lead to different distibution.

  What is the specific activity if diluted with unlabeled endogenous material?

**Sample Volume**

- May need to be large to increase sensitivity.
- May also need to be small because of competing assays:

  Immune factors (antibodies, globulins)

  Hormones

  Disease state modifiers

  In humans, concomitant medications

**Distribution**

- Rapidly cleared from blood
- Frequently distributed via lymphatics

- Target and off-target receptor found portion of drug molecule population in the body rapidly predominate
- Pharmacodynamic very different from pharmacokinetics
- Delivery rarely by oral route
- Available test material supply very limited in early development

The upshot of these points is that it may not be practical to follow established guidelines for ADME evaluation. Binding proteins, immunoreactive metabolites, and antibodies could interfere with the immunoassays used to measure the activity of biotechnologically derived pharmaceuticals. The link between immunoreactivity and pharmacological activity may be difficult to establish, making the data difficult to interpret. In radiolabeled distribution studies, if the label alters the physicochemical and biological properties of the test material, its pharmacokinetic behavior may change. These analytical difficulties may preclude accurate characterization of the distribution, metabolism, and excretion of a protein.

AUC and $C_{max}$ are commonly measured to identify safety ratios for new chemical entities. Since the analytical methods used for biotechnologically derived pharmaceuticals may lack specificity, a clinical marker of biological activity or efficacy may sometimes be more appropriate than exposure data.

It is therefore essential that before pivotal (repeat-dose) preclinical studies are initiated, bioanalytical assay development must be completed. This has to cover potential test species, normal (and diseased) humans. The assays must be validated in the sampling matrix of the toxicity test species, and one should also develop suitable assays for antibodies to the test article.

## REFERENCES

Bauer, L. A. (2001). *Applied Clinical Pharmacokinetics*. McGraw-Hill, New York.

Baumann, A. (2006). Early development of therapeutic biologics—pharmacokinetics. *Curr. Drug Metab.* 7:15–21.

Bronaugh, R. L. (1998). Methods for in vitro percutaneous absorption. In F. N. Marzulli and H. I. Maibach (Eds.), *Dermatotoxicology Methods*. Taylor & Francis, Philadelphia.

Caldwell, W. S., Byrd, C. D., DeBethizz, J. D., and Brooks, P. A. (1994). Modern instrumental methods for studying mechanisms of toxicity. In A. W. Hayes (Ed.), *Principles and Methods of Toxicology*. Raven, New York.

Connally, R., and Anderson, M. (1991). Biologically based pharmacokinetic models: Tools for toxicological research risk assessment. *Annu. Rev. Pharmacol. Toxicol.* 31:503–523.

Dennis, M. S., Zhang, M., Meng, Y. G., Kadkhodayan, M., Kirchokofer, D., Combs, D., and Damicot, L. A. (2002). Albumin binding as a general strategy for improving the pharmacokinetics of proteins. *J. Biol. Chem.* 277:35035–35043.

Draper, A., Madan, A., Smith, K., and Parkinson, A. (1998). Development of a non-high pressure liquid chromatography assay to determine testosterone hydroxylase (CYP3A) activity in human liver microsomes. *Drug Metab. Disp.* 26:299–304.

European Medicines Agency (EMEA) (2006, Jan.). Guideline on the clinical investigation of the pharmacokinetics of therapeutic proteins. EMEA, London.

Gabrielsson, J., and Weiner, D. (1997). *Pharmacokinetic/Pharmacodynamic Data Analysis: Concepts and Applications*. Apotekarsociateten, Stockholm.

Gad, S. C. (Ed.) (2007a). *Animal Models in Toxicology*, 2nd ed. Marcel Dekker, New York.

Gad, S. C. (Ed.) (2007b). *Toxicology of the Gastrointestinal Tract*. CRC Press, Boca Raton, FL.

Gad, S. C., and Chengelis, C. P. (1997). *Acute Toxicology Testing*, 2nd ed. Academic, San Diego, CA.

Garner, C., and Matthews, H. (1998). The effect of chlorine substitution on the dermal absorption of polychlorinated biphenyls. *Toxicol. Appl. Pharmacol.* 149:150–158.

Goldstein, A., Aronow, L., and Kalman, S. (1974). *Principles of Drug Action: The Basis of Pharmacology*. Wiley, New York.

Griffin, R., Godfrey, V., Kim, Y., and Burka, L. (1997). Sex-dependent differences in the disposition of 2,4-dichlorophenoxyacetic acid in Sprague-Dawley rats, B6C3F1 mice and Syrian hamsters. *Drug Metab. Disp.* 25:1065–1071.

Holland, J., Kao, M., and Whitaker, M. J. (1984). A multisample apparatus for kinetic evaluation of skin penetration in vitro: The influence and metabolic status of the skin. *Toxicol. Appl. Pharmacol.* 72:272–280.

International Conference on Harmonisation (ICH) (2000a). S3A: Toxicokinetics: The assessment of systemic exposure in toxicity studies. ICH, Brussels.

International Conference on Harmonisation (ICH) (2000b). S3B: Pharmacokinetics: Guidance for repeated dose tissue distribution studies. ICH, Brussels.

Keberle, G., Brindle, S. D., and Greengard, P. (1971). The route of absorption of intraperitoneally administered compounds. *J. Pharmacol. Exp. Ther.* 178:562–566.

La Du, B., Mandel, H., and Way, E. (1972). *Fundamentals of Drug Metabolism and Drug Disoposition*. Williams & Wilkins, Baltimore, MD.

Levy, R. H., Thummel, K. D., Tauger, W. F., Hansten, P. D., and Meichelbaum, E. (2000). *Metabolic Drug Interactions*. Lippincott, Williams & Wilkins, Philadelphia, PA.

Lu, H., Wang, J., Chan, K., and Young, D. (1998). Effects of phenobarbital of stereoselective metabolism of ifosfamide in rats. *Drug Metab. Disp.* 26:476–482.

Meyer, U. A. (1994). The molecular basis of genetic polymorphisms of drug metabolism. *J. Pharm. Pharmacol.* 83(Suppl. 1):409–415.

Mohler, A. M., Cook, J. E., and Baumann, G. (1992). In B. L. Ferraiolo, M. A. Mohler, and C. A. Gloff (Eds.), *Protein Pharmacokinetics and Metabolism*. Plenum, New York, 35–71.

Monostory, K., Jemnitz, K., Vereczkey, L., and Czira, G. (1997). Species differences in metabolism of panomifene, an analogue of tamoxifen. *Drug Metab. Disp.* 25:1370–1378.

Mugfor, C., and Kidderis, G. (1998). Sex-dependent metabolism of xenobiotics. *Drug Metab. Rev.* 34:441–498.

O'Reilly, W. J. (1972). Pharmacokinetics in drug metabolism and toxicology. *Can. J. Pharm. Sci.* 7:66–77.

Parkinson, A. (1996). Biotransformation of xenobiotics. In C. Klaassen (Ed.), *Cassarett & Doull's Toxicology: The Basic Science of Poisons*, 5th ed. McGraw-Hill, New York.

Pratt, W. B., and Taylor, P. (1990). *Principles of Drug Action: The Basis of Pharmacology*, 3rd ed. Wiley, New York.

Renwick, A. G. (2000). Toxicokinetics. In B. Ballantyne, T. Marrs, and T. Syversen (Eds.), *General and Applied Toxicology*. Grove's Dictionaries, New York.

Shapiro, J. (1981). *Radiation Protection*, 2nd ed. Harvard University Press, Cambridge, MA, pp. 12–18.

Shargel, L., and Yu, A. B. C. (1999). *Applied Biopharmaceutics & Pharmacokinetics*. Appleton & Lange, New York.

Smith, D. (1991). Species differences in metabolism and pharmacokinetics: Are we close to an understanding? *Drug Metab. Rev.* 23:355–373.

Tang, L., Persky, A. M., Hochhaus, G., and Meibohm, B. (2004). Pharmacokinetic aspects of biotechnology products. *J. Pharm. Sci.* 93:2184–2204.

Tse, F. L. S., and Jaffe, J. M. (1991). *Preclinical Drug Disposition*. Marcel Dekker, New York.

U.S. Food and Drug Administration (FDA) (2000). *Redbook II*. FDA, Washington, DC. pp. 138–148.

U.S. Food and Drug Administration (2002, Oct.). Guidance for industry: Immunotoxicological evaluation of investigational new drugs. Center for Drug Evaluation and Research, Washington, DC.

U.S. Food and Drug Administration (2005). Guidance for industry: Safety testing of drug metabolizers. Center for Drug Evaluation and Research, Washington, DC.

Waddell, W. J. (1972). Autoradiography in drug dispositions studies. In B. N. La Du, H. G. Mandel, and E. L. Way (Eds.), *Fundamentals of Drug Metabolism and Drug Disposition*. Williams & Wilkins, New York, pp. 505–514.

Wang, Y. M., and Reuning, R. (1994). A comparison of two surgical techniques for the preparation of rats with chronic bile duct canulae for the investigation of enterohepatic circulation. *Lab. Anim. Sci.* 44:479–485.

Yacobi, A., Skelly, J. P., and Batera, V. K. (1989). *Toxicokinetics and New Drug Development*. Pergamon, New York.

# *Safety Pharmacology*

Safety pharmacology is the evaluation and study of the potentially life threatening pharmacological effects of a potential drug which is unrelated to the desired therapeutic effect and therefore may present a hazard—particularly in individuals who already have one or more compromised or limited organ system functions. Unlike other nonclinical evaluations of the safety of a drug, these evaluations are usually conducted at doses not too much in excess of the intended clinical dose. This topic is another which has had to undergo significant change since the last edition because of a changed regulatory environment.

General/safety pharmacology was an emerging discipline within the pharmaceutical industry in 2002, when the International Conference on Harmonisation (ICH) guidance was promulgated and it became a major area of both concern and activity. It seeks to identify unanticipated effects of new drug candidates on major organ function (i.e., secondary pharmacological effects) and ensure that they are critically assessed in a variety of animal models. A survey was conducted to obtain customer input on the role and strategies of this emerging discipline. Overlooked in importance by all but a few (Zbinden, 1966, 1984) for many years, the Japanese clearly became the leaders in developing and requiring such information, while the United States was in a position behind Japan and the European Union (EU) in both having formal requirements and implementing industrial programs. While major companies were aware and largely addressing the need by the mid-1990s (Kintner et al.,

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

1994; Murphy et al., 1995; Sullivan and Kinter, 1995) and the European Medicines Agency (EMEA) promulgated guidelines in 2000, it was only with the ICH S7 guidance that U.S. and global regulatory interest came to play.

While historically companies have conducted evaluations of cardiovascular and central nervous system (CNS) functions, less evaluate respiratory, gastrointestinal, and renal functions; a few conduct a ligand-binding/activity panel as part of their pharmacological profiling. Since 2001, studies to evaluate the cardiovascular, pulmonary, and CNS safety pharmacological aspects of all but a few new drugs have been required (these exceptions will be discussed shortly) before they are evaluated clinically (i.e., had human exposure).

It is important that the tests employed detect bidirectional drug effects and that the tests performed be validated in both directions with appropriate reference (control) substances and be sensitive in the acute therapeutic range (Folke, 2000; Redfern et al., 2002). This requirement is less appropriate for multiparameter procedures. Blind testing could be an advantage. Ethical considerations are important, but the ultimate ethical criterion is the assessment of risk for humans. Safety pharmacology studies should not be overinclusive but should be performed to the most exacting standards, including good laboratory practice (GLP) compliance. It is important that safety pharmacological data be available during the planning stage for phase I studies, but this is often not the case. In part this arises from the viewpoint that human tolerance (particularly in a well-designed and executed phase I study in normal volunteers) is, in itself, an adequate assessment of safety pharmacology. This is, of course, backward—such human tolerance is rather, properly, an extension (and expression) of the nonclinical safety pharmacology.

The other point of view in the past has been that properly executed repeat-dose preclinical safety studies meeting the current design requirements will (or could) fill these needs. In addition, undesired pharmacological activities of novel drugs or biologicals are seen as limiting development of a therapeutic agent prior to the characterization of any toxicological effects. In rodent species, general pharmacological assays have traditionally been used to screen new agents for pharmacological effects on the central and peripheral nervous systems, the autonomic nervous system and smooth muscles, the respiratory and cardiovascular systems, the digestive system, and the physiological mechanisms of water and electrolyte balance. In large-animal species, such as dogs and nonhuman primates, smaller numbers of animals per study limit their use for screening assays, but these species may play an important role in more detailed mechanistic studies. For drugs and biologicals that must be tested in nonhuman primates because of species-specific action of the test agent, functional pharmacological data are often collected during acute or subacute toxicity studies. This requires careful experimental design to minimize any impact that pharmacological effects or instrumentation may have on the assessment of toxicity. In addition, with many new therapies targeted at immunological diseases, the pharmacological effect of therapeutics on the immune system presents new challenges for pharmacological profiling. The applications of pharmacological

assays by organ systems in both rodent and large-animal species are discussed as well as practical issues in assessing pharmacological endpoints in the context of toxicity studies (Martin et al., 1997; Matsuzawa et al., 1997).

Pharmacoepidemiological studies in Europe and the United States show that adverse drug reactions now may account for up to 10% of the admissions of patients to hospitals at a cost of hundreds of millions of U.S. dollars annually (Sjoquist, 2000). This represents a considerable increase compared to 20 years ago. A partial explanation is the many shortcomings of clinical trials and their relevance for health care. Adverse drug reactions are often poorly studied and documented in these studies and very seldom included in health economical analyses of the value of new drugs. Pharmacovigilance is product—rather than utilization—oriented and quite invisible in clinical medicine. This is regrettable, since up to 50% of adverse drug reactions (ADRs) are dose dependent and thus preventable. Hopefully, the rapid progress in molecular and clinical pharmacogenetics will provide new tools for clinicians to choose and dose drugs according to the individual needs of patients. A good starting point for those not well versed in pharmacology and the range of potential mechanisms of action and of interaction can be found in *Goodman & Gilman* (Brunton et al., 2006).

## 1 REGULATORY REQUIREMENTS

While the ICH guidelines promulgated in November 2000 (implemented in Europe in the three regions in June 2001) are the announced international standards for regulation, the actual situation in some countries remains very mixed (Hite, 1997; Olejiniczak, 1999; Fujimori, 1999).

Japan continues to operate in conformance to the Ministry of Health and Welfare (MHW) 1999 revision of its "Guidelines for Safety Pharmacology Studies." The basic principle of the revision is to harmonize the guideline with the international concepts. The working group decided to change the title "General Pharmacology" to "Safety Pharmacology" because the objective of this guideline is to assess the safety of a test substance in humans by examining the pharmacodynamic properties of the substance. The proposed guideline includes studies on vital functions as essential studies that should be performed prior to human exposure. Studies are also required to be conducted when predictable or unexpected observed effects are concerned. The working group recommends a case-by-case approach to select the necessary test items in consideration of the variable information available.

In the EU, the Committee for Proprietary Medicinal Products (CPMP) issued the draft "Note for Guidance on Safety Pharmacology Studies in Medicinal Product Development" in 1998, but it has not yet been finalized or put in force, and as of the middle of 2001, the U.S. Food and Drug Administration (FDA) has remained mute on guidelines.

The actual requirements of the initial November 8, 2000, ICH guidelines provided only broad outlines of requirements. They called for the conduct of

studies in a core battery to assess effects on the cardiovascular (Table 1), respiratory (Table 2), central nervous system (Table 3), and secondary organ system (Table 4) effects. Follow-up studies for the care battery are also required on a case-by-case basis for the three main organ systems:

- *Central Nervous System* Behavioral pharmacology, learning and memory, specific ligand binding, neurochemistry, visual, auditory, and/or electrophysiological examinations, and so on.
- *Cardiovascular System* Behavioral pharmacology, learning and memory, specific ligand binding, neurochemistry, visual, auditory, and/or electrophysiological examinations, and so on.

**TABLE 1 Cardiovascular System Safety Pharmacology Evaluations: Core**

| |
|---|
| Hemodynamics (blood pressure, heart rate) |
| Autonomic function (cardiovascular challenge) |
| Electrophysiology (ECG in dog) |
| *QT Prolongation: Noncore* |
| An additional guideline, ICH S7B, is in preparation which will address the assessment of potential for QT prolongation. In the meantime, CPMP 986/96 indicates the following preclinical studies should be conducted prior to first administration to humans: |
| • Cardiac action potential in vitro |
| • ECG (QT measurements) in cardiovascular study which would be covered in core battery |
| • hERG channel interactions (hERG expressed in HEK293 cells) |

**TABLE 2 Respiratory System Safety Pharmacology Evaluation**

| |
|---|
| Respiratory functions: Measurement of rate and relative tidal volume in conscious animals |
| Pulmonary function: Measurement of rate, tidal volume, and lung resistance and compliance in anesthetized animals |

**TABLE 3 Central Nervous System Safety Pharmacology Evaluation**

| |
|---|
| Irwin test: General assessment of effects on gross behavior and physiological state[a] |
| Locomotor activity: Specific test for sedative, excitatory effects of compounds |
| Neuromuscular function: Assessment of grip strength |
| Rotarod: Test of motor coordination |
| Anesthetic interactions: Test for central interaction with barbiturates |
| Anti-/proconvulsant activity: Potentiation or inhibition of effects of pentylenetetrazole |
| Tail flick: Tests for modulation of nociception (also hot plate, Randall Selitto, tail pinch) |
| Body temperature: Measurement of effects on thermoregulation |
| Autonomic function: Interaction with autonomic neurotransmitters in vitro or in vivo |
| Drug dependency: Test for physical dependence, tolerance, and substitution potential |
| Learning and memory: Measurement of learning ability and cognitive function in rats |

[a]Usually a functional observational battery (FOB) is integrated into rodent (rat) repeat-dose toxicity studies to meet this requirement.

**TABLE 4 Secondary Organ System Safety Pharmacology Evaluation**

| |
|---|
| Renal system |
| Renal function—Measurement of effects on urine excretion in saline-loaded rats |
| Renal dynamics—Measurement of renal blood flow, glomerular filtration rate (GFR), and clearance |
| GI system |
| GI function—Measurement of gastric emptying and intestinal transit |
| Acid secretion—Measurement of gastric acid secretion (Shay rat) |
| GI irritation—Assessment of potential irritancy to gastric mucosa |
| Emesis—Nausea, vomiting |
| Immune system |
| Passive cutaneous anaphylaxis (PCA)—Test for potential antigenicity of compounds |
| Other |
| Blood coagulation |
| In vitro platelet aggregation |
| In vitro hemolysis |

- *Respiratory System* Tidal volume, bronchial resistance, compliance, pulmonary arterial pressure, blood gases.

No testing is deemed necessary for:

- Locally applied agents (e.g., dermal or ocular) where systemic exposure or distribution to the vital organs is low
- Cytotoxic agents for treatment of end-stage cancer patients
- Biotechnology-derived products that achieve *highly specific* receptor targeting (refer to toxicological studies)
- New salts having similar pharmacokinetics and pharmacodynamics to already well-characterized drugs

## 2 STUDY DESIGNS AND PRINCIPALS

As a starting place, unlike older pharmacology studies, safety pharmacology studies are conducted as GLP studies unless performed as screens. At the same time, unlike other safety assessment studies, these do not need to vastly exceed intended therapeutic doses so as to identify signs of toxicity. In this sense, they are closer to hazard tests.

General guidance for dose (or concentration) section for such studies is as follows (ICH, 1997, 2001; Gad, 2004):

*In vivo studies* are designed to define the dose–response curve of the adverse effects:

- Doses should include and exceed primary pharmacodynamic or therapeutic range.

- In the absence of safety pharmacology parameters, the highest doses equal or exceed some adverse effects (toxic range).

*In vitro studies* are generally designed to establish an effect–concentration relationship (range of concentrations).

Considerations in the selection and design of specific studies are straightforward:

The following factors should be considered (selection):

- Effects related to the therapeutic class
- Adverse effects associated with members of the chemical/therapeutic class
- Ligand-binding or enzyme data suggesting a potential for adverse effects
- Data from investigations that warrant further investigation

A hierarchy of organ systems can be developed:

- Importance with respect to life-supporting functions: cardiovascular, respiratory, central nervous system
- Functions which can be transiently disrupted without causing irreversible harm

The absence of observed activity may represent either a true- or a false-negative effect. If an assay is valid for the particular test article and fails to indicate activity, it is an appropriate indicator of future events (Green, 1997). However, if the assay is insensitive or incapable of response, the test represents a form of bias, albeit unconscious. Many biological products demonstrate a specificity of response that limits the utility of commonly employed safety studies. Specificity for many biologicals arises from both their physicochemical properties and their similarity to endogenous substances which are regulated in a carefully controlled manner. To overcome the issue of a lack of predictive value, various approaches may be used. For example, a multiple-testing strategy of mutually reinforcing studies may be employed or safety studies may be adaptively fit to the biological circumstance.

A separate issue is how and when to consider isomers, metabolites, and the actual finished product:

- Generally a parent compound and its major metabolite(s) that achieve systemic exposure should be evaluated.
- It may be important to test active metabolites from humans.
- Testing of individual isomers should also be considered.
- Studies with the finished product are necessary only if kinetics/dynamics are substantially altered in comparison to the active substance previously tested.

There are also special considerations as to how to statistically evaluate specific aspects of these studies. Specifically, analysis of time to event becomes very important (Anderson et al., 2000).

## 3 ORGAN-SYSTEM SPECIFIC TESTS

### 3.1 General Considerations in Selection and Design of Safety Pharmacology Studies

The following factors should be considered (selection):

- Effects related to the therapeutic class (e.g., proarrhythmia is a common feature of antiarrhythmic drugs)
- Adverse effects associated with members of the chemical/therapeutic class (e.g., antipsychotics and QT prolongation)
- Receptor/enzyme/ion channel binding data suggesting a potential for adverse effects
- Any data from previous studies that warrant further investigation

A hierarchy of organ systems is considered:

- Importance with respect to life-supporting functions: cardiovascular, respiratory, central nervous
- Functions which can be transiently disrupted without causing irreversible harm (e.g., urinary system, GI tract)

### 3.2 Studies on Metabolites, Isomers, and Finished Products

Generally the parent compound and its major metabolite(s) that achieve systemic exposure need to be evaluated. This means that either the test species must be metabolically comparable to humans or that human metabolites must also be evaluated. The testing of individual isomers also needs to be considered.

## 4 CARDIOVASCULAR

While the initial greatest concern for cardiovascular risks was associated with QT prolongation (first raised by the CAST trial in 1989), it has become clear since then that there are a range of potentially life threatening cardiovascular pharmacological drug effects which must be evaluated (see Gad, 2008;

Braunwald, 2008) by drugs not intended to have cardiac effects (see Table 5). The cardiovascular system is one of the primary vital functions which has to be examined during safety pharmacology studies. Cardiovascular system functioning is maintained by cardiac electrical activity and by pump-muscle function, which contribute to hemodynamic efficacy.

The aim of cardiovascular safety pharmacology is to evaluate the effects of test substances on the most pertinent components of this system in order to detect potentially undesirable effects before engaging in clinical trials (Lacroix and Provost, 2000). In the basic program, a detailed hemodynamic evaluation is carried out in the anesthetized dog. It is completed by cardiac and/or cellular electrophysiology investigations in order to assess the arrhythmogenic risk. The basic program can be preceded by rapid and simple testing procedures during the early drug discovery stage. It should be completed, if necessary, by specific supplementary studies, depending on the data obtained during the early clinical trials. The current gold standard study is performed using unrestrained radiotelemetrized dogs (Gauvin et al., 2006) and can measure multiple endpoints continuously.

**TABLE 5 Noncardiac Drugs Known to Induce or Worsen Heart Failure According to Suggested Mechanism(s) Implicated**

| Drug Class | Drug |
|---|---|
| Cardiomyopathy | |
| Cytotoxic drugs | Doxorubicin, epitubicin, and other anthracyclines; mitoxantrone, cyclophosphamide, 5-fluorouracil, capecitabine |
| Immunomodulating drugs/antibodies | Trastuxumab, interferon-α, interleukin-2, infliximab, etanercept |
| Antifungal drugs | Itraconazole, amphotericin B |
| Antipyschotic drugs | Clozapine |
| Pulmonary hypertension | |
| Antimigraine drugs | Methysergide, ergotamine |
| Appetite suppressants | Fenfluramine, fluramine, phentermine |
| Heart valve abnormalities | |
| Antimigraine drugs | Methysergide, ergotamine |
| Appetite suppressants | Fenfluramine, fluramine, phentermine |
| Antiparkinsonian drugs | Pergolide |
| Fluid overload | |
| Nonsteroidal anti-inflammatory drugs (NSAIDs), including cyclooxygenase-2 inhibitors | All |
| Antidiabetic drugs | Rosiglitazone, pioglitazone, troglitazone |
| Glucocorticoids | All |
| Herbal drugs | Herbal drugs containing liquorice or adulterated with NSAIDs |

### 4.1 Hemodynamics, Electrocardiograms, and Respiration in Anesthetized Dogs or Primates

Anesthetized studies using Modular Instruments data capture systems are done to record six-lead electrocardiograms (ECGs) (I, II, III, $aV_l$, and $aV_f$), left ventricular pressure variables, arterial blood pressure and respiratory measurement of arterial blood flow in selected vascular beds, cardiac output, and arterial blood gas measurement. ECG intervals are measured from the lead II ECG and the QT interval can be corrected for heart rate using Bazett's, Friderecia's, or Van De Water's formulas. Different formulas are appropriate for different species (Solovien et al., 2006).

### 4.2 Cardiac Conduction Studies

In addition to the above hemodynamic measurements, intraventricular, intra-arterial, and atrioventricular conduction times and velocities can be measured using epicardial electrodes in the anesthetized and thoracotomized dog.

### 4.3 Conscious Dog, Primate, or Pig Telemetry Studies

Effects on blood pressure, heart rate, lead II ECG, core body temperature, and locomotor activity can be explored using DataSciences telemetry (or similar) implanted devices in guinea pigs, dogs, pigs, and primates. Effects on behavior can be captured on video using CCTV for dog and primate studies. Repeated administration and interaction studies may also be performed. This approach has the advantages of avoiding the effects of both anesthesia and restraint of the animal.

### 4.4 Six-Lead ECG Measurement in Conscious Dog

Conscious studies using integrated telemetry systems devices for measurement of blood pressure and six-chest-lead ECG measurements (V2, V4, V6, V10, rV2, rV4) (Hamlin, 2008). ECG interval analysis is performed on the V2 lead (RR, PR, QT, QTc intervals, QRS duration). QT dispersion can also be measured. Locomotor activity can be monitored and behavior captured on video using CCTV:

- In addition to validated systems for automatic measurement of ECG parameters, ECGs can be reviewed by veterinary cardiology services to detect any transparent abnormalities.
- Colonies of telemetered animals can be set up and maintained for repeat use.
- Respiration rate measurements can be taken from dogs in slings using a pneumograph system.

- An animal-specific correction of the QT interval can also be derived for each dog/primate based on individual variability of the QT interval with rate using the Framingham equation.

Recent concerns over the arrhythmogenic effects of a number of marketed compounds have resulted in the issue of the Points to Consider document CPMP 986/96 by the EMA (http://www.eudra/org/humandocs/PDFs/SWP/098696en.pdf).

Studies to assess the effects of a compound and any known metabolites on the ECG and cardiac action potentials are recommended. Changes in action potential duration and other parameters measured are a functional consequence of effects on the ion channels which contribute to the action potential. This in vitro test is considered to provide a reliable risk assessment of the potential for a compound to prolong the QT interval in humans.

### 4.5 Systems for Recording Cardiac Action Potentials

These include a range of currently available methodologies, some of which can be incorporated into existing study designs:

- Isolated ventricular Purkinje fibers from dog or sheep
- Isolated right ventricular papillary muscle from guinea pig
- Continuous intracellular recording of action potentials and online analysis of resting membrane potential, maximum rate of depolarization, upstroke amplitude, and action potential duration using the Notocord HEM data acquisition system
- Assessment of use-dependent and inverse-use-dependent actions by stimulation at normal, bradycardic, and tachycardic frequencies (e.g., see the discussion below of the inverse-use-dependent properties of sotalol in dog Purkinje fibers)

### 4.6 Special Case (and Concern)—QT Prolongation

Drugs that alter ventricular repolarization (generally recognized as drugs that prolong the QT interval) have been associated with malignant ventricular arrhythmias (especially the distinctive polymorphic ventricular tachycardia called torsade de pointes) and death. Many of the drugs now known to alter ventricular repolarization were developed as antiarrhythmics (e.g., dofetelide, sotalol), but others (e.g., cisapride, terfenadine) were developed without the expectation of any effect upon electrically excitable membranes. This has led to the ICH promulgating S7B (ICH, 2005) with specific guidance for evaluation.

The QT interval of the ECG is reflected in three main ways. First, electrophysiologically, it reflects the depolarization and repolarization phase of ven-

tricular myocytes. Second, mechanically, it represents the time of contraction of the ventricles. And last, physiologically, its duration is a function of numerous variables (heart rate, diseases, nutrition, diurnal cycle, etc.) Diagnostically, a prolongation of its duration indicates an enhanced risk for ventricular arrhythmias (torsade de pointes) and sudden cardiac death.

The association between abnormalities of repolarization and life-threatening arrhythmias is stronger than some other associations between laboratory abnormalities and clinical events. For example, there are drugs (tacrine) and inborn errors of metabolism (Gilbert's syndrome) that cause wild excursions in liver function tests but with no adverse consequences. In contrast, although the severity or proarrhythmia at a given QT duration varies from drug to drug and from patient to patient, no drug is known to alter ventricular repolarization without inducing arrhythmias,[1] and each of the several congenital long-QT syndromes is associated with an elevated incidence of malignant arrythmias.

With any given repolarization-altering drug, the risk of malignant arrhythmia seems to increase with increasing QT interval, but there is no well-established threshold duration below which a prolonged QT interval is known to be harmless. The extent of QT prolongation seen with a given drug and patient may be nonlinearly related to patient factors (sex, electrolyte levels, etc.) and to serum levels of the drug and/or its metabolites. The actual incidence of malignant arrhythmias, even in association with the drugs most known to induce them, is relatively low, so failure to observe malignant arrhythmias during clinical trials of ordinary size and duration does not provide substantial reassurance.

Abnormal repolarization and the associated arrhythmias are the end results of a causative chain that starts with alternations in the channels of ionic flux through cell membranes. Some cells (e.g., those of the Purkinje system or midmyocardium) seem especially susceptible to these changes. At a substrate level, the links on the chain are alterations in the time course of the action potential, alteration in the propagation of action potentials within a given cell, and alterations in the propagation of action potentials from cell to cell within syncitia and from tissue to tissue within the heart. At a higher level of aggregation, one sees "afterdepolarizations" in the terminal portion of the action potential, spontaneous beats triggered by afterdepolarizations, propagation of these beats to other cells, and reentrant excitation.

With these considerations in mind, the problem of altered repolarization should be integrated into drug development by (Malik and Camm, 2001):

[1] Some QT-prolonging drugs [e.g., amiodarone; see Hohnloser et al. (1994)] are not reported to have caused many arrhythmic deaths, but this observation must be interpreted carefully. In a population with a high incidence of life-threatening arrhythmias, a drug with both proarrhythmic and antiarrhythmic effects might cause a net reduction in arrhythmias, and the arrhythmias that it had induced might not be attributed to it. In a population whose native arrhythmias were not life threatening, the same drug might result in a net decrease in mortality.

- In vitro screening of the drug and its metabolites for effects on ion channels (especially $I_{Kr}$)
- In vitro screening of the drug and its metabolites for effects on action potential duration
- Screening of the drug and its metabolites for altered repolarization in animal models
- Focused preclinical studies for proarrhythmia if altered repolarization is seen in preclinical screening or in patients

Some specific techniques which can be employed are as follows:

1. *Cloned Human Potassium Channels* Assessment of effects on cloned human ether-á-go-go (hERG) $K^+$ channels stably expressed in a cell line by measurement of whole-cell K current ($I_{Kr}$) using a voltage clamp. Other cloned human ion channels (e.g., KvLQT1/minK-IKs currents) are also possible.
2. *Cardiac Action Potential In Vitro—Purkinje Fibers* Intracellular recording of action potentials from cardiac Purkinje fibers isolated from dog or sheep ventricle. Measurement of maximum rate of depolarization and action potential duration to detect sodium and potassium channel interactions, respectively, according to recommendations in EMA Points to Consider document CPMP 986/96 (1997).
3. *Monophasic Action Potential in Anesthetized Guinea Pigs* Epicardia monophasic action potential recording using suction/contact pressure electrodes according to Carlsson et al. (1997). Simultaneous measurement of ECG.
4. *ECG by Telemetry in Conscious Guinea Pigs* Lead II ECG recording using DataSciences telemetry device. Repeated administration and interaction studies can be performed.
5. *Hemodynamics and ECG in Anesthetized or Conscious Dogs or Primates*
   - Conscious studies using DataSciences telemetry for blood pressure and lead II ECG or the ITS system for blood pressure and six-chest-lead ECG measurements (including QT dispersion)
   - Anaesthetized studies using $MI^2$ data capture system with additional measurement of blood flow in selected vascular beds, cardiac output, respiratory, and left ventricular function

## 4.7 Relevance of hERG to QT Prolongation

Compounds which are associated with ADRs of QT prolongation, arrhythmias such as torsade de pointes, and sudden death predominantly have a secondary pharmacological interaction with the rapidly activating delayed rectifier potassium channel $I_{Kr}$. The gene encoding this channel has been identified as *hERG*.

Testing of compounds for interactions with the hERG channel allows the identification of the potential risk of QT prolongation in humans and can be used as a screen in development candidate selection.

***Expression and Recording Systems*** HEK-293 cells have been transfected with cDNA for hERG-1 to produce a stable expression system. The cell line has been obtained under license for the laboratory of Craig January at the University of Wisconsin (Mohammad et al., 1997).

## 5 CENTRAL NERVOUS SYSTEM

The primary screening tool for CNS safety pharmacology evaluation is the functional observational screen, which seeks to use objective but noninvasive methods for evaluating the pharmacological effects of a drug on peripheral and nervous system effects.

Initially, the starting basis for such screens were the Irwin screen (Irwin, 1968), used to screen for effects in mice and still one of the ICH-designated primary screens for fulfilling the regulatory requirements for S7 evaluation of new drugs.

More commonly used is the rat functional observational batter, initially developed by Gad (1982) and subsequently further modified (Haggerty, 1991; Mattson et al., 1996). Moscardo et al. (2007) present an excellent description of the rodent functional observation battery (FOB) as currently performed in the pharmaceutical industry. Other modifications/versions which are included under the ICH screening guidelines cover the use of the dog (Gad et al., 2003).

The neurobehavioral screens which meet the primary regulatory requirement for CNS safety pharmacology evaluation of necessity actually evaluate both central and peripheral nervous system function. Such evaluations used, to the maximum extent possible, semiquantitative evaluations of a wide range of endpoints which serve to determine if there are effects on the primary functional domains of the CNS. The methods are noninvasive and use basic instruments to get quantitative data where possible—an electronic thermometer (to measure rectal/core body temperature), a strain gauge with T-bar animal grip (to measure forelimb grip strength), a sand table (to measure hindlimb splay), an activity stage (to measure locomotor activity), and a rotarod (to evaluate motor coordination).

The only other required equipment usually includes a dicher for auditory startle, a pen light (for pupil response), and a blunt probe (for various touch-based reflexes). The complete screen should be performed at at least two (and preferably three) time intervals after a single dose of the drug.

There are four broad classes of approaches to any subsequent and more detailed assessment of nervous system effects of drugs in animals.

### 5.1 Isolated Tissue Assays

The classic approach to screening for nervous system effect is a series of isolated tissue preparation bioassays conducted with appropriate standards to determine if the material acts pharmacologically directly on neural receptor sites or transmission properties. Though these bioassays are normally performed by a classical pharmacologist, a good technician can be trained to conduct them. The required equipment consists of a Mangus (or similar style) tissue bath (Turner, 1965; Offermeier and Ariens, 1966; Nodine and Seigler, 1964), physiograph or kymograph, force transducer, glassware, stimulator, and bench spectrophotometer. The assays utilized in the screening battery are listed in Table 6 along with the original reference describing each preparation and assay. The assays are performed per the original author's descriptions with only minor modifications, except that control standards (as listed in Table 6) are always used. Only those assays that are appropriate for the neurological/muscular alterations observed in the screen are utilized. Note that all these are intact organ preparations, not minced tissue preparations as others (Bondy, 1979) have recommended for biochemical assays.

The first modification in each assay is that, where available, both positive and negative standard controls (pharmacological agonists and antagonists, respectively) are employed. Before the preparation is utilized to assay the test material, the tissue preparation is exposed to the agonist to ensure that the preparation is functional and to provide a baseline dose–response curve against which the activity of the test material can be quantitatively compared. After the test material has been assayed (if a dose–response curve has been generated), one can determine whether the antagonist will selectively block the activity of the test material. If so, specific activity at that receptor can be considered as established. In this assay sequence, it must be kept in mind that a test material may act to either stimulate or depress activity, and therefore the roles of the standard agonists and antagonists may be reversed.

Commonly overlooked when performing these assays is the possibility of metabolism to an active form that can be assessed in this in vitro model. The test material should be tested in both original and "metabolized" forms. The metabolized form is prepared by incubating a 5% solution (in aerated Tyrodes) or other appropriate physiological salt solution with strips of suitably prepared test species liver for 30 min. A filtered supernatant is then collected from this incubation and tested for activity. Suitable metabolic blanks should also be tested.

### 5.2 Electrophysiology Methods

There are a number of electrophysiological techniques available which can be used to detect and/or assess neurotoxicity. These techniques can be divided into two broad general categories: those focused on CNS function and those focused on peripheral nervous system function (Seppalainen, 1975).

**TABLE 6 Isolated Tissue Pharmacological Assays**

| Assay System | Endpoint | Standards (Agonist/ Antagonist) | References |
|---|---|---|---|
| Rat ileum | General activity | None (side-spectrum assay for intrinsic activity) | Domer, 1971 |
| Guinea pig vas deferens | Muscarinic nicotinic or muscarinic | Methacholine/atropine<br>Methacholine/ hexamethonium<br>Methacholine/atropine | Leach, 1956 |
| Rat serosal strip | Nicotinic | Methacholine/hexamethonium | Khayyal et al., 1974 |
| Rat vas deferens | Alpha adrenergic | Norepinephrine/ phenoxybenzamine | Rossum, 1965 |
| Rat uterus | Beta adrenergic | Epinephrine/propranol | Levy and Tozzi, 1963 |
| Rat uterus | Kinin receptors | Bradykinin/none | Gecse et al., 1976 |
| Guinea pig tracheal chain | Dopaminergic | Dopamine/none | Domer, 1971 |
| Rat serosal strips | Tryptaminergic | 5-Hydroxytryptamine (serotonin)/dibenzyline or lysergic acid dibromide | Lin and Yeoh, 1965 |
| Guinea pig tracheal chain | Histaminergic | Histamine/benadryl | Castillo and De Beer, 1947a,b |
| Guinea pig ileum (electrically stimulated) | Endorphin receptors | Methenkephaline/none | Cox et al., 1975 |
| Red blood cell hemolysis | Membrane stabilization | Chloropromazine (not a receptor-mediated activity) | Seeman and Weinstein, 1966 |
| Frog rectus abdominis | Membrane depolarization | Decamethonium iodide (not a receptor-mediated activity) | Burns and Paton, 1951 |

First, however, the function of the individual components of the nervous system, how they are connected together, and how they operate as a complete system should be very briefly overviewed.

Data collection and communication in the nervous system occur by means of graded potentials, action potentials, and synaptic coupling of neurons. These electrical potentials may be recorded and analyzed at two different levels depending on the electrical coupling arrangements: individual cells (i.e., intracellular and extracellular) or multiple cells [e.g., electroencephalogram (EEG), evoked potentials (EPs), slow potentials]. These potentials may be recorded in specific central or peripheral nervous system areas (e.g., visual cortex, hippocampus, sensory and motor nerves, muscle spindles) during various behavioral states or in in vitro preparations (e.g., nerve muscle, retinal photoreceptor, brain slice).

### 5.3 CNS Function: Electroencephalography

The EEG is a dynamic measure reflecting the instantaneous integrated synaptic activity of the CNS, which most probably represents, in coded form, all ongoing processes under higher nervous control. Changes in frequency, amplitude, variability, and pattern of the EEG are thought to be directly related to underlying biochemical changes, which are believed to be directly related to defined aspects of behavior. Therefore, changes in the EEG should be reflected by alterations in behavior and vice versa.

The human EEG is easily recorded and readily quantified, is obtained noninvasively (scalp recording), samples several regions of the brain simultaneously, requires minimal cooperation from the subject, and is minimally influenced by prior testing. Therefore, it is a very useful and recommended clinical test in cases in which exposure to drugs produces symptoms of CNS involvement and in which long-term exposures to high concentrations are suspected of causing CNS damage.

Since the EEG recorded using scalp electrodes is an average of the multiple activity of many small areas of cortical surface beneath the electrodes, it is possible that in situations involving noncortical lesions the EEG may not accurately reflect the organic brain damage present. Noncortical lesions following acute or long-term low-level exposures to toxicants are well documented in neurotoxicology (Norton, 1980). The drawback mentioned earlier can be partially overcome by utilizing activation or evocative techniques, such as hyperventilation, photic stimulation, or sleep, which can increase the amount of information gleaned from a standard EEG.

As a research tool, the utility of the EEG lies in the fact that it reflects instantaneous changes in the state of the CNS. The pattern can thus be used to monitor the sleep–wake cycle activation or deactivation of the brainstem and the state of anesthesia during an acute electrophysiological procedure. Another advantage of the EEG, which is shared by all CNS electrophysiological techniques, is that it can assess the differential effects of toxicants (or drugs) on various brain areas or structures. Finally, specific CNS regions (e.g., the hippocampus) have particular patterns of afterdischarge following chemical or electrical stimulation which can be quantitatively examined and utilized as a tool in neurotoxicology.

The EEG does have some disadvantages or, more correctly, some limitations. It cannot provide information about the effects of toxicants on the integrity of sensory receptors or of sensory or motor pathways. As a corollary, it cannot provide an assessment of the effects of toxicants on sensory system capacities. Finally, the EEG does not provide specific information at the cellular level and therefore lacks the rigor to provide detailed mechanisms of action.

Rats represent an excellent model for this as they are cheap, resist infection during chronic electrode and cannulae implantation, and are relatively easy to train so that behavioral assessments can be made concurrently.

Depending on the time of drug exposure, the type of scientific information desired, and the necessity of behavioral correlations, a researcher can perform acute and/or chronic EEG experiments. Limitations of the former are that most drugs that produce general anesthesia modify the pattern of EEG activity and thus can complicate subtle effects of toxicants. However, this limitation can be partially avoided if the effect is robust enough. For sleep–wake studies, it is also essential to monitor and record the electromyogram (EMG).

Excellent reviews of these electrophysiological approaches can be found in Fox et al. (1982) and Takeuchi and Koike (1985).

### 5.4 Neurochemical and Biochemical Assays

Though some very elegant methods are now available to study the biochemistry of the brain and nervous system, none has yet discovered any generalized marker chemicals which will serve as reliable indicators or early warnings of neurotoxic actions or potential actions. There are, however, some useful methods. Before looking at these, however, one should understand the basic problems involved.

Normal biochemical events surrounding the maintenance and functions of the nervous system centers around energy metabolism, biosynthesis of macromolecules, and neurotransmitter synthesis, storage, release, uptake, and degradation. Measurement of these events is complicated by the sequenced nature of the components of the nervous system and the transient and labile nature of the moieties involved. Use of measurements of alternations in functionality as indicators of neurotoxicity is further complicated by our lack of a complete understanding of the normal operation of these systems and the multitude of day-to-day occurrences (such as diurnal cycle, diet, temperature, age, sex, and endocrine status) which are constantly modulating the baseline system. For detailed discussions of these difficulties, the reader is advised to see Johnson (1975) and Damstra and Bondy (1980, 1982).

## 6 RESPIRATORY/PULMONARY SYSTEM

The known effects of drugs from a variety of pharmacological/therapeutic classes on the respiratory system and worldwide regulatory requirements support the need for conducting respiratory evaluations in safety pharmacology. The objective of the studies is to evaluate the potential for drugs to cause secondary pharmacological or toxicological effects that influence respiratory function. Changes in respiratory function can result from either alterations in the pumping apparatus that controls the pattern of pulmonary ventilation or changes in the mechanical properties of the lung that determine the transpulmonary pressures (work) required for lung inflation and deflation. Defects in the pumping apparatus are classified as hypo- or hyperventilation syndromes and are evaluated by examining ventilatory parameters in a conscious animal

model. The ventilatory parameters include respiratory rate, tidal volume, minute volume, peak (or mean) inspiratory flow, peak (or mean) expiratory flow, and fractional inspiratory time. Defects in the mechanical properties of the lung are classified as obstructive or restrictive disorders and can be evaluated in animal models by performing flow–volume and pressure–volume maneuvers, respectively. The parameters used to detect airway obstruction include peak expiratory flow, forced expiratory flow at 25 and 75% of forced vital capacity, and a timed forced expiratory volume, while the parameters used to detect lung restriction include total lung capacity, inspiratory capacity, functional residual capacity, and compliance. Measurement of dynamic lung resistance and compliance, obtained continuously during tidal breathing, is an alternative method for evaluating obstructive and restrictive disorders, respectively, and is used when the response to drug treatment is expected to be immediate (within minutes postdose). The species used in the safety pharmacology studies conducted in our laboratory are the same as those used in toxicology studies since pharmacokinetic and toxicological/pathological data are available in these species. These data can be used to help select test measurement intervals and doses and to aid in the interpretation of functional change. The techniques and procedures for measuring respiratory function parameters are well established in guinea pigs, rats, and dogs (Murphy, 1994):

**Respiration and Pulmonary Function in Rats**

Study design includes three doses and controls with eight animals per group and must be GLP compliant and a single dose by oral (depending on clinical route of administration) gavage.

Evaluations include respiratory rate, peak inspiratory flow, peak expiratory flow, inspiration time, expiration time, tidal volume, airway resistance, and blood oxygenation.

## 7 SECONDARY ORGAN SYSTEM

The kidneys are an important target for toxic effects of drug candidates. It is mandatory to select accurate, clinically relevant parameters in order to be in a position to detect putative nephrotoxic effects during the safety pharmacology program. The glomurelar filtration rate appears to be of major interest since it is associated with the definition of acute renal failure. Measurement of, for example, renal blood flow, proteinuria, enzymuria, and fractional excretion of sodium is also highly useful to detect any possible renal impact of a new compound. Although the rat is, by far, the most widely used animal species, there are no specific (clinically relevant) reasons to choose it. Various parameters may vary according to the species, sex, strain, age, and so on. Since in most cases acute renal failure occurs following administration of drugs in patients with preexisting risk factors, it is suggested that sensitized animal

models be validated and used (salt depletion, dehydration, coadministration of pharmacological agents, etc.).

The potential effects of new drugs on the digestive system can be examined in a number of model systems of which intestinal motility in the mouse and/or gastric emptying in the rat are examples recommended for safety pharmacology evaluation. Intestinal motility, assessed by the transit of carmine dye in the mouse, and gastric motility, assessed by stomach weight in the rat, were examined using a range of clinical drugs or potent pharmacological agents known to affect gastrointestinal function. Assessment of both models in the guinea pig was also evaluated. Activity was demonstrated with codeine, diazepam, atropine, and CCK-8 (all of which inhibited gastric function). However, neither model gave consistent and reliable results with the remaining reference compounds, namely metoclopramide, bethanechol, cisapride, deoxycholate, carbachol, and domperidone. In conclusion, this investigation questions the usefulness of simple models of gastrointestinal transport in the rodent as a means of detecting potential effects of a new drug on the digestive system. This finding should be of concern to the pharmaceutical industry as these simple models are routinely used as part of a regulatory safety pharmacology "package" of studies.

A number of classic assays have been designed to examine the effects of a test article on gastrointestinal function. Gastrointestinal transit rate is most often measured with a test employing a forced meal of an aqueous suspension of activated charcoal (Janssen and Jageneau, 1957). The test article is given via the appropriate route at a preset time prior to the charcoal meal. For example, a compound intended for use via intravenous injection would be injected intravenously in mice 30 min prior to delivering a charcoal meal by gavage. The distance traveled from the stomach by the black-colored charcoal meal to a specific anatomic location within the intestine is measured at a fixed time after this meal, usually 20 or 30 min later. In validating this procedure at Mason Laboratories, we tested the ability of a parasympatholytic agent, intravenous atropine sulfate, to inhibit gastrointestinal transit. In a dose-dependent fashion, 30 and 50 mg $kg^{-1}$ atropine sulfate significantly decreased the distance traveled by the charcoal meal.

Another important safety assay of the gastrointestinal system is the influence of test article on the formation of ulcers (Shay et al., 1945). After overnight fasting, young rats are given the test article and euthanized 4 or 6 h later. The mucosal surface of the stomach and duodenum is scored for the presence of hyperemia, hemorrhage, and ulcers. The dose-dependent ulcerative properties of nonsteroidal anti-inflammatory drugs (NSAIDs) are clearly demonstrated in this assay, making it important in the development of other NSAIDs that are not as caustic to the gastrointestinal mucosa (Bramm et al., 1981; Cashin et al., 1977; Diadone et al., 1994; Darias et al., 1994).

Additional digestive system safety pharmacology tests include effects of test articles on gastric emptying rate and gastric secretion. Gastric emptying rate is measured in rats using a solution of phenol red (or Evans blue) deliv-

ered via oral gavage a preset time after administration of the test article (Megens et al., 1991). The dilution of phenol red after 30 min in the rat's stomach is determined colorimetrically at 558 nm in a spectrophotometer. This is compared to a group of control rats that are euthanized immediately after phenol red administration. The influence of test articles on gastric juice secretion is accomplished by ligating the pyloric sphincter under anesthesia in rats following a fasting period (Graf et al., 1982; Shay et al., 1945; Takasuna et al., 1992). Immediately after recovery from anesthesia, each rat is given a preset dose of the test article. The fluid content of the rat's stomach is recovered after a set period of time, usually 4 h. The volume and contents of the stomach are measured to determine the effect of the test article on gastric secretions. Electrolyte concentrations, pH, and protein content of gastric secretions can be measured in this assay (Takasuna et al., 1992).

### 7.1 Gastric Emptying Rate and Gastric pH Changes: New Model

Sometimes new technologies for safety pharmacology can come from clinical settings. The Heidelberg pH capsule (HC) was developed over 30 years ago at Heidelberg University in West Germany. H. G. Noller invented and first tested this device on over 10,000 adult patients over a three-year period. The HC is a pill-sized device containing an antimony–silver chloride electrode for measuring pH and a high-frequency transmitter operating at an average frequency of 1.9 MHz. The transmitter in the HC is activated by immersion in physiological saline by a permeable membrane enclosing the battery compartment. Thus, when a patient swallows the HC, the fluid contents of the stomach activate the transmitter. Transmitted signals are picked up via a belt receiver and can be displayed and recorded. The profile of changes in pH over time correlate with the movement of the HC through the different regions of the gastrointestinal tract (Mojaverian et al., 1989). The pH of the fasted human stomach is very acidic, on average about pH 1. When the HC moves through the pyloric sphincter and into the duodenum, there is a rapid increase in pH of over 4 pH units. Thus, one can get a fairly precise measure of gastric emptying rate in humans with this noninvasive technique. Additional pH changes have been correlated with transition of the HC through the duodenum, jejunum, and colon.

Mojaverian and colleagues (1989, 1991) have used the HC extensively to examine the influence of gender, posture, age, and content and frequency of food ingestion on the gastric emptying rate (or gastric residence time) in healthy volunteers. While developed for clinical use in people, the HC may be a useful tool for measuring important digestive system parameters in laboratory animals. The size of the HC, approximately the size of a No. 1 gelatin capsule (7 mm diameter, 20 mm long), prohibits its use in small animals (Mojaverian et al., 1989). It may be useful in studies with dogs and possibly in nonhuman primates. In particular, the HC could be used to measure gastric emptying rate in a totally noninvasive manner in dogs (Itoh et al., 1986). Dogs

are readily trainable to accept pills and to wear a receiver belt and could be tested after administration of a test compound (Lui et al., 1986; Vashi and Meyer, 1988). This technique for measuring gastric emptying rate in dogs is also advantageous in that it is not a terminal procedure. The influence of test articles on the pH within different portions of the gastrointestinal system could also be measured with the HC (Youngberg et al., 1985). The major drawback for using the HC for safety pharmacology screening is the price of the capsules and the receiver system.

## 8 SUMMARY

The initiative to all mandated safety pharmacology studies to the drug development process is overdue in arriving. However, its actual implementation and the use of the resulting data in risk–benefit decision will take some time to be worked out.

## REFERENCES

Aderson, H., Spliid, H., Larsen, S., and Dall, V. (2000). Statistical analysis of time to event data from preclinical safety pharmacology studies. *Toxical. Methods* 10:111–125.

Bondy, S. C. (1979). Rapid screening of neurotoxic agents by in vivo means. In R. M. Gryder and V. H. Frankos (Eds.), *Effects of Food and Drugs on the Development and Function of the Nervous System: Methods for Predicting Toxicity*. Office of Health Affairs, U.S. Food and Drug Administration, Washington, DC, pp. 133–143.

Bramm, E., Binderup, L., and Arrigoni-Martelli, E. (1981). An unusual profile of activity of a new basic anti-inflammatory drug, timegadine. *Agents Actions* 11:402–409.

Braunwald, E. (2008). Biomarkers in heart failure. *N. Engl. J. Med.* 358:2148–2159.

Brunton, L. L., Lazo, J. S., and Parker, K. L. (2006). *Goodman & Gilman's The Therapeutic Basis of Therapeutics*. McGraw-Hill, New York.

Burns, B. D., and Paton, W. D. M. (1951). Depolarization of the motor end-plate by decamethonium and acetylcholine. *J. Physiol.* (*London*) 115:41–73.

Carlsson, L., Amos, G. J., Andersson, B., Drews, L., and Duker, G. (1997). Electrophysiological characterization of the prokinetic agents cisapride and mosapride in vivo and in vitro. *J. Pharmacol. Exp. Ther.* 82:220–227.

Cashin, C. H., Dawson, W., and Kitchen, E. A. (1977). The pharmacology of benoxaprofen (2-[4-chlorophyll]-α-methyl-5-benzoxazole acetic acid), LRCL 3794, a new compound with anti-inflammatory activity apparently unrelated to inhibition of prostaglandin synthesis. *J. Pharm. Pharmacol.* 29:330–336.

Cardiac Arrhythmia Suppression Trial (CAST) (1989). Preliminary report: Effects of encainide and flecainide on morbidity in a randomized trial of arrhtythmia suppression after myocardial infaction. *N. Engl. J. Med.* 321:406–414.

Castillo, J. C., and De Beer, E. J. (1947a). The guinea pig tracheal chain as an assay for histamine agonists. *Fed. Proc.* 6:315.

Castillo, J. C., and De Beer, E. J. (1947b). The tracheal chain. *J. Pharmacol. Exp. Ther.* 90:104.

Cox, B. M., Opheim, K. E., Teschemach, H., and Goldstein, A. (1975). A peptide-like substance from pituitary that acts like morphine 2. Purification and properties. *Life Sci.* 16:1777–1782.

Committee for Proprietary Medicinal Products (CPMP) (1998). Note for guidance on safety pharmacology studies in medicinal product development. CPMP, London, UK.

Damstra, T., and Bondy, S. C. (1980). The current status and future of biochemical assays for neurotoxicity. In P. S. Spencer and H. H. Shaumburg (Eds.), *Experimental and Clinical Neurotoxicology*. Williams and Wilkins, Baltimore, pp. 820–833.

Damstra, T., and Bondy, S. C. (1982). Neurochemical approaches to the deletion of neurotoxicity. In C. L. Mitchell (Ed.), *Nervous System Toxicology*. Raven, New York, pp. 349–373.

Darias, V., Abdallah, S. S., Tello, M. L., Delgado, L. D., and Vega, S. (1994). NSAI activity study of 4-phenyl-2-thioxo-benzo[4,5]thieno[2,3-*d*]pyrimidine derivatives. *Arch. Pharm. (Weinheim)* 327:779–783.

Diadone, G., Maggio, B., Raff, D., Plescia, S., Bajardi, M. L., Caruso, A., Cutuli, V. M. C., and Amico-Roxas, M. (1994). Synthesis and pharmacological study of ethyl 1-methyl-5-[2-substituted-4-oxo-3(4*H*0-quinazolinyl]-1*H*)-pyrazole-4-acetates. *Eur. J. Med. Chem.* 29:707–711.

Domer, F. R. (1971). *Animal Experiments in Pharmacological Analysis*. Charles C. Thomas, Springfield, IL, pp. 98, 115, 155, 164, 220.

Fenichel, R. P., and Koerner, J. (1999). Development of drugs that alter ventricular repolarization. Available online.

Folke, S. (2000). Drug safety in relation to efficacy: The view of a clinical pharmacologist. *Pharmacol. Toxicol.* 86:30–32.

Fox, D. A., Lowndes, H. E., and Bierkamper, G. G. (1982). Electrophysiological techniques in neurotoxicology. In C. L. Mitchell (Ed.), *Nervous System Toxicology*. Raven, New York, pp. 299–336.

Fujimori, K. (1999). The role of general pharmacological studies and pharmacokinetics in the evolution of drugs (1): The role of general/safety pharmacology studies in the development of pharmaceuticals: International harmonization guidelines. *Folia Pharmacol. Jpn.* 13:31–39.

Gad, S. C. (1982). A neuromuscular screen for use in industrial toxicology. *J. Toxicol. Environ. Health* 9:691–704.

Gad, S. C. (2004) *Safety Pharmacology*. CRC, Boca Raton, FL.

Gad, S. C. (2008). Nonclinical safety assessment of the cardiovascular toxicity of drugs and combination medical devices. In D. Acosta (Ed.), *Target Organ Toxicity: The Cardiovascular System*. Taylor & Francis, Philadelphia.

Gad, S. C., Gad, S. D., and Gad, S. E. (2003). The dog functional observation battery (FOB) for use in pharmaceutical safety evaluation studies. *Int. J. Toxicol.* 22:1–8.

Gauvin, D. V., Tilley, L. P., Smith, F. W. K., and Baird, T. J. (2006). Electrocardiogram, hemodynamics, and core body temperatures of the normal freely moving laboratory beagle dog by remote telemetry. *J. Pharmacol. Toxicol. Methods* 53:128–139.

Gecse, A., Zsilinsky, E., and Szekeres, L. (1976). Bradykinin antagonism. In F. Sicuteri, N. Back, and G. Haberland (Eds.), *Kinins; Pharmacodynamics and Biological Roles*. Lenum, New York, pp. 5–13.

Green, M. D. (1997). Problems associated with the absence of activity in standard models of safety pharmacology used to assess biological products. *Int. J. Toxicol.* 16:33–40.

Graf, E., Sieck, A., Wenzl, H., and Winkelmann, J. 1982. Animal experiments on the safety pharmacology of lofexidine. *Arzneim.-Forsch./Drug Res.* 32(II) (8a):931–940.

Haggerty, G. C. (1991). Strategies for and experience with neurotoxicity testing of new pharmaceuticals. *J. Am. Coll. Toxicol.* 10:677–687.

Hamlin, R. L. (2008). How many ECG leads are required for in vivo safety pharmacology? *J. Pharmacol. Toxicol. Meth.* 57:161–168.

Hite, M. (1997). Safety pharmacology approaches. *Int. J. Toxicol.* 16:23–31.

Hohnloser, S. H., Klingenheben, T., and Singh, B. N. (1994). Amiodarone-associated proarrhythmic effects. *Ann. Intern. Med.* 121(7):529–535.

International Conference on Harmonisation (ICH) (1997). S6A: Preclinical safety evaluation of biotechnology-derived pharmaceuticals. ICH, Brussels.

International Conference on Harmonisation (ICH) (2001). S7A: Safety pharmacology studies for human pharmaceuticals. ICH, Brussels.

International Conference on Harmonisation (ICH) (2005). S7 B(R): The non-clinical evaluation of the potential for delayed ventricular prolongation by human pharmaceuticals. ICH, Brussels.

International Conference on Harmonisation (ICH) (2002). Safety pharmacology studies for assessing the potential For Delayed Ventricular Repolarization (QT Interval Prolongation) By Human Pharmaceuticals. ICH, Brussels.

Irwin, S. (1968). Comprehensive observational assessment: la. A systematic, quantitative procedure for assessing the behavioral and physiologic state of the mouse. *Psychopharmacologia* (*Berlin*) 13:222–257.

Itoh, T., Higuchi, T., Gardner, C. R., and Caldwell, L. 1986. Effect of particle size and food on gastric residence time of nondisintegrating solids in beagle dogs. *J. Pharm. Pharmacol.* 38:801–806.

Janssen, P. A. J., and Jageneau, A. H. (1957). A new series of potent analgesics: *Dextro* 2:2-diphenyl-3-methyl-4-morpholino-butyrylpyrrolidine and related amides. *J. Pharm. Pharmacol.* 9:381–400.

Johnson, M. K. (1975). The delayed neuropathy caused by some organophosphorus esters: Mechanism & challenge. *Crit. Rev. Toxicol.* 31:289–316.

Khayyal, M. T., Tolba, N. M., El-Hawary, M. B., and El-Wahed, S. A. (1974). A sensitive method for the bioassay of acetylcholine. *Eur. J. Pharmacol.* 25:287–290.

Kinter, L. B., Gossett, K. A., and Kerns, W. D. (1994). Status of safety pharmacology in the pharmaceutical industry. *Drug Dev. Res.* 32:208–216.

Lacroix, P., and Provost, D. (2000). Basic safety pharmacology: The cardiovascular system. *Therapie* 55:63–69.

Leach, G. D. H. (1956). Estimation of drug antagonisms in the isolated guinea pig vas deferens. *J. Pharm. Pharmacol.* 8:501.

Levy, B., and Tozzi, S. (1963). The adrenergic receptive mechanism of the rat uterus. *J. Pharmacol. Exp. Ther.* 142:178.

Lin, R. C. Y., and Yeoh, T. S. (1965). An improvement of Vane's stomach strip preparation for the assay of 5-hydroxy-tryptamine. *J. Pharm. Pharmacol.* 17:524–525.

Lui, C. Y., Amidon, G. L., Berardi, R. R., Fleisher, D., Youngberg, C. A., and Dressman, J. B. (1986). Comparison of gastrointestinal pH in dogs and humans: Implications on the use of the beagle dog as a model for oral absorption in humans. *J. Pharm. Sci.* 75:271–274.

Malik, M., and Camm, A. J. (2001). Evaluation of drug-induced QT interval prolongation: Implications for drug approval and labeling. *Drug Safety* 24:323–351.

Martin, L. I., Horvath, C. J., and Wyand, M. S. (1997). Safety pharmacology screening: Practical problems in drug development. *Int. J. Toxicol.* 16:41–65.

Matsuzawa, T., Hashimoto, M., Nara, H., Yoshida, M., Tamura, S., and Igarashi, T. (1997). Current status of conducting function tests in repeated dose toxicity studies in Japan. *J. Toxicol. Sci.* 22:374–382.

Mattson, J. L., Spencer, P. J., and Albee, R. R. (1996). A performance standard for clinical and functional observational battery examination of rats. *J. Am. Coll. Toxicol.* 15:239.

Megens, A. A. H. P., Awouters, F. H. L., and Niemegeers, C. J. E. (1991). General pharmacology of the four gastrointestinal motility stimulants bethanechol, metoclopramide, trimebutine, and cisapride. *Arzneim.-Forsch./Drug Res.* 41(I)(6):631–634.

Mohammad, A. H., Jabed, C. W., and Jones, R. W. (1997). Blockage of the HERG human cardiac $K^+$ channel by gastrointestinal prokinetic agent cisapride. *Am. J. Physiol.* 273:H2534–H2538.

Mojaverian, P., Chan, K., Desai, A., and Vivian, J. (1989). Gastrointestinal transit of a solid indigestible capsule as measured by radiotelemetry and dual gamma scintigraphy. *Pharm. Res.* 6:717–722.

Mojaverian, P., Reynolds, J. C., Ouyang, A., Wirth, F., Kellner, P. E., and Vlasses, P. H. (1991). Mechanism of gastric emptying of a nondisintegrating radiotelemetry capsule in man. *Pharm. Res.* 8:97–100.

Moscardo, E., Maurin, A., Dorigatti, R., Champeraux, P., and Richard, S. (2007). An optimized methodology for the neurobehavioral assessment in rodents. *J. Pharmacol. Toxicol. Meth.* 56:239–255.

Murphy, D. J. (1994). Safety pharmacology of the respiratory system: Techniques and study design. *Drug Dev. Res.* 32:237–246.

Murphy, D. J., Joran, M. E., and Grando, J. C. (1995). A non-invasive method for distinguishing central from peripheral nervous system effects of respiratory depressant drugs in conscious rats. *Gen. Pharmacol.* 26(3):569–575.

Nodine, J. H., and Sieger, P. E. (1964). *Animal and Clinical Pharmacologic Techniques in Drug Evaluation*. Year Book Medical Publishers, Chicago, IL, pp. 36–38.

Norton, S. (1980). Toxic responses of the central nervous system. In J. Doull, C. D. Klaassen, and M. O. Amdur (Eds.), *Toxicology: The Basic Science of Poisons*, 2nd edn. Macmillan, New York.

Offermeier, J., and Ariens, E. J. (1966). Serotonin I. Receptors involved in its action. *Arch. Int. Pharmacodyn. Ther.* 64:92–215.

Olejiniczak, K. (1999). Development of a safety pharmacology guideline. *Human Exp. Toxicol.* 18:502.

Redfern, W. S., Wakefiel, I. D., Prior, H., Pollard, C. E., Hammond, T. G., and Valenlin, J. P. (2002). Safety pharmacology—A progressive approach. *Fund. Clin. Pharmacol.* 16:161–173.

Rossum, J. M. van (1965). Different types of sympathomimetic β-receptors. *J. Pharm. Pharmacol.* 17:202.

Seeman, P., and Weinstein, J. (1966). Erythrocyte membrane stabilization by tranquilizers and antihistamines. *Biochem. Pharmacol.* 15:1737–1752.

Seppalainen, A. M. (1975). Applications of neurophysiological methods in occupational medicine: A review. *Scand. J. Work Environ. Health* 1:1–14.

Shay, H., Komarox, S. A., Fels, S. S., Meranze, D., Gruenstein, M., and Siplet, H. (1945). A simple method for the uniform production of gastric ulceration in the rat. *Gastroenterology* 5:43–61.

Sjoquist, F. (2000). Drug-related hospital admissions. *Ann. Pharmacol.* 34:832–839.

Solovien, M. V., Hamlin, R. L., Barrett, R. M., Chengelis, C. P., and Schaefer, G. J. (2006). Different species require different correction factors for the QT interval. *Cardiovasc. Toxical.* 6:145–157.

Sullivan, A. T., and Kinter, R. B. (1995). Status of safety pharmacology in the pharmaceutical industry. *Drug Dev. Res.* 35:166–172.

Takasuna, K., Kasai, Y., Usui, C., Takahashi, M., Hirohashi, M., Tamura, K., and Takayama, S. (1992). General pharmacology of the new quinolone anti-bacterial agent levofloxacin. *Arzneim.-Forsch./Drug Res.* 42(I)(3a):408–418.

Takeuchi, Y., and Koike, Y. (1985). Electrophysiological methods for the in vivo assessment of neurotoxicology. In K. Blum and L. Manzo (Eds.), *Neurotoxicology*. Marcel Dekker, New York, pp. 613–629.

Turner, R. A. (1965). *Screening Methods in Pharmacology*, Vols. I and II. Academic, New York, pp. 27–128.

Vashi, V. I., and Meyer, M. C. (1988). Effect of pH on the in vitro dissolution and in vivo absorption of controlled-release theophylline in dogs. *J. Pharm. Sci.* 77:760–764.

Youngberg, C. A., Wlodyga, J., Schmaltz, S., and Dressman, J. B. (1985). Radio-telemetric determination of gastrointestinal pH in four healthy beagles. *Am. J. Vet. Res.* 46:1516–1521.

Zbinden, G. (1966). The significance of pharmacologic screening tests in the preclinical safety evaluation of new drugs. *J. New Drugs* 6:1–7.

Zbinden, G. (1984). Neglect of function and obsession with structure in toxicity testing. In W. Paton, J. Mitchell, and P. Turner (Eds.) *Proceedings of the Ninth International Congress on Pharmacology*, Vol. 1. Macmillan, New York, pp. 43–49.

# *Special Concerns for Preclinical Evaluation of Biotechnology Products*

Although many assume that biotechnology is a recent concept, the application of this science has been understood for many years and utilized in its simplest form in the fermentation of beer, wine, and bread by microbial agents. Modern biotechnology can be divided into three research and development areas: recombinant DNA technology (rDNA), monoclonal antibody technology, and bioprocess technology (Mackett, 1993; Malinowski, 1999). The commercialization of these three processes is based on the premise that biotechnology can cost-effectively produce large quantities of a highly purified product.

Biotechnology as a promising source of new and more efficient source of more targeted therapeutics has been with us since the mid-1980s (Maulik and Patel, 1997). While (as one should expect) some of the early promises have not quite been met, biotechnology has turned out to be a valuable source of new and valuable therapeutics and currently accounts for one-third of all new therapeutics entering the marketplace. Table 1 lists the proteins approved in 2007 that have more than a billion dollars a year in global sales.

Protein and other biotechnology-derived therapeutics have some fundamental differences from traditional small (synthetic organic) molecules and so require modified or different approaches to characterize their toxicity and evaluate their safety. Table 2 presents a comparative summary of these differences.

---

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

**TABLE 1 Blockbuster Biotechnology Approvals (2007)**

| Company | Product | Indication | 2007 Approval | Status | Peak Sales Potential ($ billions) |
|---|---|---|---|---|---|
| Cephalon | Nuvigil (amordafinil) | Daytime sleepiness | Approved | Approved June 18 | $1–2 |
| Onyx/Bayer | Nexavar (sorafenib) | Hepatocellular carcinoma | Likely | PDUFA* Dec. 28 | $1+ |
| ImClone | Erbitux (cetuximab) | Refractory colorectal cancer (survival claim) | Approved | Approved Oct. 2 | $1 |
| Genzyme | Renvela (sevelamer carbonate) | Serum phosphorus control in dialysis patients with chronic kidney disease | Approved | Approved Oct. 21 | $1 |
| Alexion | Soliris (eculizumab) | Paroxysmal nocturnal hemoglobinuria | Approved | Approved Mar. 19 | $0.5–1+ |
| Speedel/ Novartus | Tekturna | Hypertension | Approved | Approved Mar. 6 | $0.5–1 |

*Prescription Drug User Fee Act.

**TABLE 2 Comparison of Protein Therapeutic Agents with Small-Molecule Drugs**

| Parameter | Proteins | Small Molecules |
|---|---|---|
| Drug substance | Heterogeneous mixture; broad specifications during development; specifications may change during development | Single entity; high chemical purity; exception: racemic mixtures; specifications well defined early in development |
| Drug product | Usually intravenously or subcutaneously | Generally oral; few formulations during development |
| Impurities | Difficult to standardize | Purity standards well established |
| Bridging requirements | Significant for drug substance | Bioequivalence procedures |
| Biological activity | May mimic naturally occurring molecules; primary mechanism of toxicity; predictive based on mechanism | Less predictive |
| Nonspecificity | Variable significance | Usually significant; drug–drug interactions |
| Chronic toxicity | Lack of models because of species-determined biological specificity and antigenicity | Models sometimes relevant |
| Impurities | Toxicity not a major issue; may impact immunogenicity | May be significant; purity standards well established |

Among all the other aspects of increasing understanding of what is involved in the evaluation and development of biologically derived therapeutics has been a very real evolution of what is needed to evaluate the safety of these products. In the beginning, there was a stark duality of expectations. On one

side, early advocates of biotechnology held that there were unlikely to be any safety concerns other than those due to hyperpharmacology overactivity at the target receptors (Thomas and Myers, 1998; Weissinger, 1989; Vallbracht et al., 1982) and contamination (such as occurred with the Cutter product early on in the history of the polio vaccine; Offit, 2005). On the other hand, there were those that cautioned against the possibility of extreme and unforeseeable toxicities. The truth, as is usually the case, has turned out to be in between. We have come to understand that the toxicity of protein moieties primarily arises from either overexpression of their desired therapeutic effects (i.e., largely disordering of the immune system) such as seen with interferons and interleukins (Fent and Zbinden, 1997) or "off-target effects" that are at other than the intended receptor. Additionally, there are cases of immune responses to therapeutics and of antibody neutralization products (Weissinger, 1989; Vallbracht et al., 1982).

The principal purposes of preclinical safety evaluation in this context remain:

(a) To detect harmful (toxic) effects
(b) To exclude other potentially harmful effects
(c) To determine their relationship to dose and duration of treatment
(d) If possible to discover their mechanism or at least pathogenesis

The information from (a)–(d) should be used to predict possible adverse actions in the target species in order to:

- Warn clinicians about unacceptable risks
- Warn clinicians about risks that should be monitored
- Remind clinicians and others of the possibility of toxic effects not detected because they could not be displayed by the test systems used (e.g., headache in a nonhuman species or carcinogenicity in a one-month experiment) or were not sought.

In addition, the toxicologist as a general biological scientist should always be alert to physiological and pharmacological effects manifested in his or her experiments because they may illuminate mechanisms of health and disease of both academic and practical importance (Folb, 2006). Table 3 points out the common differences between use and large-molecule safety assessment programs.

The objectives of the preclinical safety studies on biologically derived therapeutics are to identify the pharmacological and toxicological effects that are likely to be encountered throughout the course of clinical development and beyond. The selection and design of such studies should first consider what may be known about other products which are structurally and/or

**TABLE 3 Differences between Chronic Use Nonclinical Safety Assessment Plans for Large and Small Molecules**

| Biopharmaceuticals | Small-molecule Pharmaceuticals |
|---|---|
| • "Case-by-case" approach to nonclinical safety evaluation | • Similar nonclinical safety studies for most products |
| • Pharmacology used to select species | • Metabolism used to select species |
| • Single species common for repeat-dose studies beyond 28 days | • Rodent and nonrodent typical |
| • Dosing sequences may not be daily | • Dosing daily unless justified |
| • Immunogenicity important | • Generally not immunogenic |
| • Genetic toxicology not required | • Genetic toxicology required |
| • Metabolism studies not appropriate | • Metabolism studies required |
| • Longest repeat dose typically 6 months nonrodent | • Nonrodent may be up to 12 months |
| • Two-year carcinogenicity rare | • Two-year carcinogenicity required |
| • Developmental toxicity in relevant species may be required | • Developmental toxicity in two species required |

pharmacologically similar. The program and study design should then proceed to consider:

1. Intended manner of use, including dose, route of administration, and particulars of dosing regimen
2. Age of intended patient population
3. Selection of relevant model species
4. Stability of formulated drug substance under conditions of use
5. Physiological (disease) state of intended patients

In the area of bioengineered products, many of which are complex proteins of potent but sparsely studied activities in living systems, the investigative responsibilities of the toxicologist are likely to be very important because he or she may be the first observer able to study the effects of repeated administration of a range of doses on a living system (Griffiths and Lumley, 1998). It is now possible to frame a classification of the types of biologically derived therapeutic products (Table 4).

Each type of product has some specific considerations. The range of materials is enormous. The deciding factors for the toxicologist should be the precision with which the material can be characterized by physicochemical means, as that should be inversely related to the burden of repeated biological testing necessary to assure safety, and the extent of prior knowledge of its biological properties. The greater our ignorance of the latter, the more searching should the toxicologist's studies be in order to discern the biological (pharmacological) properties of the substance. Exposure of the individual must also be considered, as different criteria may apply to deliberate administration of a living organism, which could spread in the community.

**TABLE 4 Classification of Bioengineered Products on Practical Grounds**

| Type | Bioengineering Involvement | Pharmacological Properties | PhysicoChemical Characterization | Example |
|---|---|---|---|---|
| 1. Low-molecular-weight substance | New production route | Well known | Rigorous | Amino acid 6-APA |
| 2. High-molecular-weight substance | New production route | Fairly well known | Extensive | Human hormones, e.g., HGH, hPTH (humanized parathyroid hormone) |
| 3a. Endogenous high-molecular-weight substance | First ever production | Some knowledge | Moderate | IFN |
| 3b. Endogenous high-molecular-weight substance | First ever production, perhaps gene splicing to make hybrid molecule | Scanty to limited knowledge | Moderate | Other lymphokines, tumor necrosis factor, etc. |
| 4. Engineered antigen | Partly or totally synthetic antigen + rDNA production | Probably predictable | Rigorous | Synthetic vaccine for poliomyelitis or hepatitis B |
| 5. Monoclonal antibody (or component) | Hybridoma human cell line | Probably predictable antigenicity | Moderate | Antitumor antibody for imaging anti-idiotype antibody as vaccine |
| 6. Living organism | Removal of pathogenicity by genetic manipulation | Uncertain | Limited | As vaccine immunogen, e.g., *Salmonella typhimurium* TY21a, modified herpes, or to carry antigen, e.g., vaccinia |

As in any safety evaluation, the planned work should be related to the intended use and treatment of humans, for example, one dose in a few gravely ill patients or multiple doses of the entire healthy community as prophylaxis against a trivial condition. Contrast, say, what might be appropriate for tumor necrosis factor (TNF), as in an experimental trial in a few sufferers from late-stage cancer, with the requirements for a candidate vaccine against dental caries to be widely administered to healthy children.

Because of the rapid development of new biotechnology products, toxicology and safety assessment departments at most chemical or pharmaceutical companies are presently or soon will be confronted with the development of testing protocols for the safety evaluation of rDNA products. Routine

toxicology assessment as performed in the past using standard protocols may not apply and may in fact represent unnecessary or inappropriate studies. Because of the relatively nontoxic nature and species specificity of many of the new biotechnology products, less evaluation in rodent species may be required than for some of the chemicals of the past. What is needed in dealing with the products of this new technology is the rethinking of traditional toxicology-testing approaches.

## 1 REGULATION

The regulation of biologically derived therapeutics actually has a long history and has also continued to evolve (see Table 5). This history led to the Public Health Service (PHS) Act providing a somewhat mixed description of the products under its authority, which in turn serves to define biologicals for the U.S. Food and Drug Administration (FDA) Center for Biologics Evaluation and Research (CBER): "[A biologic is] any virus, therapeutic serum, toxin, antitoxin, vaccine, blood, blood component or derivative, allergenic product, or analogous product, or arsphenamine or its derivative (or any other trivalent organic arsenic compound), applicable to the prevention, treatment, or cure of disease, or injuries in man. …" The Code of Federal Regulations (CFR), Title 21, Section 600.3 then goes on to state: "(h) Biological product means any virus, therapeutic serum, toxin, antitoxin, or analogous product applicable to the prevention, treatment or cure of diseases or injuries in man." Confusion

**TABLE 5 Historical Perspectives of Biological Therapeutics Regulation**

| | |
|---|---|
| 1902 | Federal Virus, Serum and Toxin Act / PHS (after tetanus-contaminated diphtheria antitoxin led to deaths of 10 children); intent was to ensure safety, purity, and potency |
| 1906 | Pure Food and Drugs Act (Upton Sinclair's *The Jungle*) |
| 1937 | Division of Biologics Control/NIH |
| 1938 | FD&C Act (sulfanilamide elixir), biologicals exempt! |
| 1955 | Division of Biological Standards Established (poliovirus) |
| 1962 | FD&C Amendments (thalidomide) |
| 1972 | Bureau of Biologics / FDA |
| 1978 | Good manufacturing practices regulations |
| 1979 | Good laboratory practices regulations |
| 1982 | Bureau of Biologics merged with Bureau of Drugs ⇒ National Center for Drugs and Biologics |
| 1983 | Orphan Drug Act |
| 1984 | Drug Price Competition and Patent Term Restoration Act |
| 1992 | User fees |
| 1997 | ICH harmonized guidelines |
| 2002 | Responsibility for regulation of biological substances divided between CDER and CBER |

of authority and responsibility between the three human health product centers of the FDA [Center for Drug Evaluation and Research (CDER), CBER, and Center for Devices and Radiological Health (CDRH)] led in 1992 to the promulgation of three intercenter agreements. The agreement between CDER and CBER states that the following biological products require licensure and come under the CBER's authority (Mathieu, 1997):

- Vaccines, regardless of the method of manufacture (vaccines were defined as agents administered for the purpose of eliciting an antigen-specific cellular or humoral response) (CBER, 1997)
- In vivo diagnostic allergenic products and allergens intended for use as "hyposensitization agents"
- Human blood or human blood–derived products, including placental blood–derived products, animal-derived procoagulant products, and animal- or cell culture-derived hemoglobin-based products intended to act as red blood cell substitutes
- Immunoglobulin products
- Products composed of or intended to contain intact cells or intact microorganisms
- Proteins, peptides, or carbohydrate products produced by cell culture, excluding antibiotics, hormones, and products previously derived from human or animal tissue regulated as drugs
- Protein products made in animal body fluid by genetic alteration of transgenic animals, animal venoms, or constituents of venoms

Other classes of products identified as CBER-regulated products include:

- Synthetically produced allergenic products intended to specifically alter the immune response to a specific antigen or allergen
- Certain drugs used in conjunction with blood banking or transfusion

As will be seen at the end of this chapter in the discussion of gene therapy products [where the National Institutes of Health (NIH) also has some regulatory role], there is still some ambiguity of authority in areas where technology has outrun regulatory foresight.

Within the United States, the regulation of therapeutics is split on relatively arbitrary grounds. This is presented in Table 5 of Regulation of Human Pharmaceutical Safety. This chapter reflects both current FDA practices and the International Conference on Harmonisation (ICH) guidelines. European Medicines Agency (EMEA) guidances started with that of the Committee on Proprietary Medicinal Products (CPMP, 1989) but subsequently has come from the EMEA. Most recently, subsequent to the TGN1412 near disaster, the EMEA has issued new guidances meant to issue a more conservative approach to clinical trials (EMEA, 2007).

## 2 PRECLINICAL SAFETY ASSESSMENT

Because of the complexity and wide diversity of biological products, their safety is evaluated on a case-by-case basis (in accordance with CBER's promulgated points to consider as summarized in Table 6) until such time as enough data on either specific products or a class of products are available.

Generally, in vivo nonclinical studies should be designed to include a sufficient number of animals per group to permit a valid estimation of a drug's toxicological and pharmacological effects in terms of incidence, severity, and the dose–response relationships involved. The latter point requires, as pointed

**TABLE 6 Points to Consider in Preclinical Safety Assessment of Biologicals**

Rationale
- In vitro or in vivo studies
- Potency assays
- Receptor characteristics (across species)
- Physiological modeling
- Scientific literature
- Scientific speculation

Indication
- Replacement therapy (long term)
- Nonpharmacodynamic treatment (prophylactic or diagnostic)
- Pharmacodynamic treatment (short or long term)

Pharmacological activity (pharmacodynamics)
- Primary endpoints
- Secondary endpoints

In vivo model selection
- Species-specific effects
- Effects independent of species
- Animal model of disease

Pharmacokinetics and absorption, distribution, metabolism, and excretion (ADME): correlation with pharmacodynamics
- Low dose
- High dose

General toxicity
- Single dose (acute)
- Repeated dose (subacute or subchronic)

Specific toxicity (may include one or more of the following studies)
- Local irritation (local reactogenicity)
- Antigenicity
- Chronic toxicity
- Reproduction toxicity, including teratogenic potential
- Mutagenicity
- Tumorigenicity
- Carcinogenicity
- Other toxicity concerns (e.g., neurotoxicity, immunotoxicity)

out throughout this text, thoughtful selection of doses. Comparable formulation, routes and regimens of administration, duration of exposure, and suitable time to allow expression of expected response are also important proper design features. Tables 7–9 summarize the basic testing requirements for three subsets of biologically derived drugs.

**TABLE 7 Biotechnology-Derived Drug Test Matrix**

| Test Requirement | Species |
|---|---|
| Initial clinical trial/investigational new drug (IND) requirement | |
| 1. Acute toxicity in rodents (oral and IV[a]) | R/M |
| 2. Acute toxicity in nonrodents (oral) | D/S/P |
| 3. Seven-day dose range finder (DRF) toxicity in rodents (oral) | R/M |
| 4. Seven-day DRF toxicity in nonrodents (oral) | D/S/P |
| 5. Genotoxicity only if appropriate (species cases) | |
| 6. Safety pharmacology: cardiovascular in vivo | D/P/S |
| 7. Safety pharmacology: respiratory—rodent | R |
| 8. Pivotal/repeat dose in rodents (14–28 days oral) | R/M |
| 9. Pivotal/repeat dose | D/P/S |
| 10. Five species microsome metabolic panel[a] | In vitro |
| 11. Develop bioanalytical for three species (human/rodent/nonrodent) | NA |
| 12. Antibody-based assay to select appropriate species | |
| To support continued clinical development | |
| 13. Immunotoxicity | TBD |
| 14. Pivotal/repeat dose in rodents (3/6 months oral)[b] | R/M |
| 15. Pivotal/repeat dose in nondents (3/9–12 months oral) | D/P/S |
| To support marketing approval | |
| 16. Reproductive toxicity, segment I[b] | R |
| 17. Reproductive toxicity, segment III[b] | R |

*Note*: Species: R = rat, M = mouse, D = dog, S = pig, P = primate, B = rabbit; TBD = to be determined. All studies described must be performed with GLP.
[a]Recommended.
[b]May be required.

**TABLE 8 Vaccine Test Matrix**

| Test Requirement | Species |
|---|---|
| Initial clinical trial/IND requirement | |
| 1. Acute toxicity in rodents (oral and IV[a]) | R/M |
| 2. Acute toxicity in nonrodents (oral) | D/S/P |
| 8. Pivotal/repeat dose in rodents (14–28 days oral)[b] | R/M |
| 9. Pivotal/repeat dose in nonrodents (14–28 days oral) | D/P/S |
| To support continued clinical development and to support marketing approval | |
| 13. Immunotoxicity | TBD |

*Note*: Abbreviations as in Table 7. All studies described must be performed with GLP.
[a]Recommended.
[b]May be required.

**TABLE 9 Biological Test Matrix**

| Test Requirment | Species |
|---|---|
| Initial clinical trial/IND requirement | |
| 1. Acute toxicity in rodents (oral and IV[a]) | R/M |
| 2. Acute toxicity in nonrodents (oral) | D/S/P |
| 3. Seven-day DRF toxicity in rodents (oral) | R/M |
| 4. Seven-day DRF toxicity in nonrodents (oral) | D/S/P |
| 5. Genotoxicity: bacterial mutagenicity (Ames)[b] | In vitro |
| 6. Pivotal/repeat dose in rodents (28 days oral) | R/M |
| 7. Pivotal/repeat dose in nonrodents (28 days oral) | D/P/S |
| 8. Develop bioanalytical for three species (human/rodent/nonrodent) | NA |
| To support continued clinical development | |
| 9. Developmental toxicity (segment II), rat and rabbit pilots and rat and rabbit studies[b] | R/B |
| 10. Neoantigenicity | B |

*Note*: Abbreviations as in Table 7. All studies described must be performed with GLP.
[a]Recommended.
[b]May be required.

The number of species necessary in preclinical testing programs varies. However, there is no specific requirement for the routine use of two species (e.g., one rodent and one nonrodent) in toxicology studies of biological products. Proper species selection for use in these trials is essential. Table 10 summarizes the key considerations for species selection.

In each stage of product development, it is important to determine exposure by measuring pharmacokinetic (including ADME) or pharmacodynamic endpoints. This includes the following: (1) measurements of the biological in plasma or target organs; (2) distribution and persistence of cells for cellular therapies; (3) measurements of viral shedding and recovery of certain values; (4) localization of targeted novel delivery systems; and (5) tissue tropism, including germ line tissue, of vectors used in gene therapies.

Such studies provide important information for a better interpretation of the toxicity observed in animals and aid in the selection of not only the proposed initial human dose but also the dose escalation scheme and the frequency of dosing in the clinical trial(s). Further, once such exposure data are available in humans, the data can be used to better correlate the human and animal findings. Toxicity studies should be performed in the same species used to assess exposure. Oftentimes, exposure and toxicity are measured in the same study, particularly when nonrodents are used.

Toxicity studies should be designed to identify not only a safe dose but also a toxic dose(s) to anticipate the product's safety and to better define the therapeutic index in humans. Specific product considerations that may complicate the process of defining a toxic dose may include limits based on

**TABLE 10 Factors to Consider in Species Selection for Protein Therapeutic Development**

| Cross-Reactivity | Immunogenicity |
|---|---|
| • Cross-reactivity alone is not sufficient to identify a relevant species. Suitable affinity and potency to give valid results is also necessary. How close does the level of potency have to be to the situation in humans to be meaningful?<br>• How can alternative approaches be "front loaded" into the development pathway without leading to a standard two-species approach?<br>• What is the possibility of assigning greater value to nonconventional preclinical studies, such as genetically altered rodents and surrogate antibodies, that are scientifically relevant?<br>• If, after a variety of cross-reactivity testing that includes binding studies, functional activity in cell-based systems, sequence homology, and tissue cross-reactivity studies, the only relevant species is chimpanzees, is it justified, scientifically and ethically, to use the chimpanzees to study the effects of the MAb? | • An emerging issue as more MAb is developed for chronic use is the impact of neutralizing antibodies on repeat-dose studies. This is a significant scientific problem that might be partially overcome by the use of the surrogate antibodies. What is the potential for the regulatory acceptability of surrogate antibodies for immunogenicity reasons.<br>• To what extent is immunogenicity considered in species selection for safety and toxicology studies? Should this be given more priority in the selection of relevant species for long-term toxicity studies for MAb intended for chronic indications? |

formulation, lack of significant systemic absorption, or the amount of the product available. The lack of significant toxicity in animals does not necessarily mean that the product is safe. The margin of safety for the initial starting dose, however, will likely be adequate.

Historically, the goal of acute toxicity studies was to define a lethal dose range following a single administration or the administration of a few closely spaced doses. More recently, these studies have been designed to evaluate a high dose that causes significant toxicity but not necessarily lethality. If deaths occur, rarely are such studies expected to provide sufficient information to determine the cause of death.

Studies are often one to two weeks in duration and routinely include body weight determinations, clinical observations, and gross necropsy findings. Additional antemortem studies may be performed as appropriate, especially in large animals (e.g., observation of local reactogenicity, pharmacokinetic evaluations, hematological, and/or clinical biochemistry measurements). Histological evaluations may also be performed.

The duration of repeat-dose studies should be at least as long as the proposed clinical study. These studies are designed to establish a dose–response relationship, define target organ(s) of toxicity, and determine whether observed toxicities are reversible. Evaluation parameters should include not only those routinely performed in the acute studies but also those performed in the additional studies as well. Special tests, such as ophthalmoscopic, electrocardiograph,

body temperature, and blood pressure monitoring, are often included. Depending on the study duration, sampling at multiple time points may be necessary to better characterize the kinetics of response. As mentioned, a group of animals will be examined at term, and some may be reserved for at treatment-free or recovery period to evaluate the reversibility of any findings.

Specific (local tissue tolerance) toxicity studies may be necessary due to special characteristics of the product or the clinical indication. Adjuvant vaccines are routinely evaluated for local (injection site) reactions, and cellular therapies are routinely screened for tumorigenic potential. Research is also needed to better predict the sensitizing potential of biological products and to determine the relevance of serum antibody levels following repeat dosing in animals and humans.

While carcinogenicity studies have not been performed routinely for biological products, they may be appropriate for products proposed for chronic use. Reproductive toxicology studies will probably become more common, especially as more women of child-bearing potential participate in early clinical trials. In the past, such studies have not been conducted for biologicals. Reproductive toxicology studies have been performed with many of the recently approved therapeutics (e.g., interferons, interleukins, cytokines, growth factors). Such studies also have been conducted in the development of AIDS vaccines intended for use in pregnant women. The standard protocol designs were modified to address specific vaccine-related concerns, including dosing in relationship to immunological effects.

The recent development of biologicals to treat various nervous system diseases has involved additional, specific neurotoxicological studies on these products. However, despite the fact that most products regulated as biologicals have an immune component or impact directly or indirectly on the immune system, standardized immunotoxicity tests that are potentially useful in screening large numbers of chemicals for their ability to adversely affect the immune system have not proven essential in assessing the safety of biological products.

Throughout the various phases of product development, additional preclinical safety studies may be necessary due to unexpected toxicity, significant changes in manufacturing process or the final formulation, or changes in clinical indications. In some cases, the ideal assessment of the safety of novel biological therapies may require alternative approaches, such as in vitro or in vivo organogenesis model systems, animal models of tolerance, animal models of disease, or transgenic animal models.

## 3 RECOMBINANT DNA TECHNOLOGY

The concept of recombinant DNA technology is based on the premise that a gene sentence may be taken from an animal or human gene responsible for the production of a particular protein and inserted into the DNA of

Escherichia coli, a single-cell bacterium. The bacterial cells then divide very rapidly, making billions of copies of themselves, including a replica of the gene that has been inserted.

There are unique ways to insert human genes into bacteria. In addition to chromosomal DNA, bacteria have numerous copies of extrachromosomal circular DNA called plasmids, which are not attached to the bacterial chromosome. These plasmids can be transferred from one bacterium to another by conjugation (e.g., mating) and can be isolated from bacteria and easily purified. Through use of restriction enzymes (i.e., a family of enzymes that can cut DNA at specific base sequences), the gene sentence to be inserted can be isolated and the plasmid DNA can be opened. While in the open state, the desired piece of animal or human DNA can be inserted. Through the use of ligase enzymes, complementary ends of the plasmid can be connected, thereby producing a recombinant plasmid recombined by joining two heterologous pieces of DNA. This recombinant plasmid can then be put back into the bacteria, and the bacteria will express the new gene function that has been inserted.

A unique characteristic of plasmids is that thousands are produced within each bacterium, to the point that up to 40% of the total DNA of the bacterium may in fact be plasmid DNA. Hence, a single piece of human DNA that heretofore could only be obtained in low concentrations can be recombined with a plasmid and the DNA sequence multiplied a million- or a billionfold (i.e., cloned). Use of cloning techniques may produce many grams of a particular human protein, instead of the few molecules that are produced in normal cells.

Examples of the early application of recombinant DNA technology in medicine are the development of recombinant human growth hormone; human insulin; human interferons (IFNs), thought to have anticancer activity in addition to antiviral activity; interleukins (regulatory proteins from lymphocytes that are believed to be important in the treatment of immunodeficiency diseases and cancer); tumor necrosis factor; epidermal and bone marrow progenitor cell growth factors; and the production of vaccines (Table 1).

Human growth hormone and insulin produced by rDNA technology are already registered with the FDA for therapeutic use. The applications of rDNA technology agriculture should improve the quality of domesticated animals through the production of new and improved vaccines, growth--promoting hormones, and less expensive food additives. Seed crops will be produced that offer improved yields and better resistance to environmental conditions. Further applications may include the insertion of genes into plants or bacteria for production of toxins that can act as biochemical pesticides or allelophatic agents (chemicals that act as natural herbicides to prevent the growth of other plant species in the same geographical area).

### 3.1 General Safety Issues

Recombinant DNA technology represents one of the most innovative achievements in biology in the last century. Although the new technology has

generated much enthusiasm for its potential applications, it has also raised concerns among both scientists and the public in general. Many of the early fears of the inadvertent development of an "Andromeda strain" during the genetic engineering of a specific microbe have long since vanished. However, other concerns remain. Can a gene cloned for toxic production from an rDNA microbe be transferred into normal bacterial flora? Could antibiotic-resistant genes be cloned and inadvertently inserted into clinically relevant pathogens not presently antibiotic resistant? To reduce these possibilities, the NIH has only certified nonconjugative plasmids (e.g., nonmating) for use in recombinant DNA microbes.

Studies by Levine et al. (1983) have addressed the issue of plasmid mobilizations, the movement of plasmids between different host cells. Human volunteers fed tetracycline along with *E. coli* HS-4 (typical of the normal intestinal flora of humans) bearing highly mobilized plasmids (e.g., pJBK5) that carried resistance to chloramphenicol and tetracycline became cocolonized with *E. coli* HS-4 bearing the antibiotic-resistant plasmid. However, the use of a poorly mobilizable plasmid (pBR325) did not result in plasmid transfer.

Taken as a whole, these studies establish the safety of recombinant DNA research when poorly mobilizable cloning vectors are used, while supporting the rationale for biological containment of highly mobilizable plasmids. They also point out the need to protect laboratory workers on antibiotic therapy from potential exposure to DNA organisms carrying any sort of antibiotic-resistant genes. A reassuring point is the relatively poor survival of rDNA strains of *E. coli* in the intestinal environment. For example, in most successful studies, 50 billion *E. coli* HS-4 organisms were required to ensure survival within the harsh environment of a human's stomach.

### 3.2 Specific Toxicological Concerns

While rDNA techniques offer exciting possibilities, there are many unanswered questions about the potential toxicity that each new product represents. For example, acute clinical toxicities of IFNs include flulike syndrome, fever, chills, malaise, anorexia, fatigue, and headache. Chronic dose-limiting toxicities include neutropenia, thrombocytopenia, impairment of myeloid maturation, reversible dose-related hepatotoxicity, some neurological toxicity (stupor, psychosis, peripheral neuropathy), and gastrointestinal toxicity. Some of these toxicities would be difficult to ascertain in rodents and, in fact, may be species specific.

A particular toxicity associated with the administration of IFN to humans and experimental animals has been depression of the cytochrome P-450 monooxygenase (MFO) metabolizing enzymes. As a consequence of MFO inhibition following treatment with IFN, the sleep time of mice treated with hexabarbital is increased, as is the toxicity of acetaminophen (Stebbing and

Weck, 1984). Possible effects on the metabolism of chemotherapeutic agents or other drugs processed by the P-450 MFOs should be anticipated.

The in vivo antitumor effects of IFNs are believed to be related to both augmentation of natural killer cell activity and antiproliferative effects. Antiproliferative activity probably also accounts for the bone marrow suppression observed in some individuals given IFN and could potentially produce effects in a routine preclinical reproduction or teratology evaluation. Dosing studies performed in newborn mice with homologous IFN have resulted in death at high doses and a marked wasting syndrome when given over an extended period (Gresser and Bourali, 1970). Both effects were attributed to the antiproliferative activity of IFNs. Inhibition of proliferation and metabolism represent potential dose-limiting toxicities of this family of rDNA molecules.

## 4 IMMUNOGENICITY

Human biopharmaceuticals are commonly immunogenic (elicit and antibody response) in nonhuman species. Immunogenicity should be evaluated in repeat-dose nonclinical safety studies to help determine whether antibody may have influenced pharmacology, toxicity, or exposure. If immunogenicity decreases exposure or neutralizes the activity of the biopharmaceutical, it may not be appropriate to continue the study or conduct studies of longer duration in that species. It is important to confirm biological activity of administered biological in "no-effect" studies where the top dose is the no-observable-effect level (NOEL). The immunogenicity of a biopharmaceutical in nonclinical species is not necessarily predictive of a potential for antibody formation in humans, but the potential consequences (summarized in Table 11) must be considered. Table 12 summarizes the strategies to conduct a bioanalytical evaluation of immunogenicity responses.

**TABLE 11 Consequences of Immunogenicity**

| Loss of efficacy | Enhancement of efficacy |
|---|---|
| • Insulin | • Growth hormone |
| • Salmon Calcitonin | |
| • Factor VIII | Neutralization of native protein |
| • IFN-$\alpha_2$ | • MDGF |
| • IFN-$\beta$ | • EPO |
| • IL-2 | |
| • GnRH | General immune effects |
| • TNFR55/IgG1 | • Allergy |
| • HCG | • Anaphylaxis |
| • GM-CSF/IL3 | • Serum sickness, etc. |

*Note*: Immunogenicity in humans may be a major safety concern.

**TABLE 12 Immunogenicity Bioanalytical Strategy for Animal Studies**

| Bioanalytical Scheme for Lower Risk Products | Bioanalytical Scheme for Medium- and Higher Risk Products |
|---|---|
| *Frequency of Sampling within Study* | |
| • Planned assessment of baseline and an appropriate, drug-free, end-of-study immunogenicty sampling time point; contingent analysis of dosing phase samples if required to support pharmacokinetic profiles | • Planned assessment of baseline and an appropriate, drug-free, end-of-study immunogenicty sampling time point; contingent analysis of dosing phase samples if required to support pharmacokinetic profiles |
| *Assessment of ADAs* | |
| • Detection of ADAs through screen and confirmatory immunoassays<br>• Consideration of characterization of titer/relative concentration of ADAs | • Detection of ADAs through screen and confirmatory immunoassays<br>• Characterization of titer/relative concentration of ADAs<br>• Detection of cross-reactivity to endogenous counterpart through screen and confirm (and maybe titer) immunoassays<br>• If antibody reactivity to endogenous counterpart is detected, characterization of neutralizing ability using target binding inhibition-based neutralizing antibody immunoassay or cell-based neutralizing antibody bioassay |

*Note*: ADAs, anti-drug antibodies.

## 5 MONOCLONAL ANTIBODY TECHNOLOGY

Offering an impressive potential for human therapy, monoclonal antibodies have become the first commercialized products of the new biotechnology. They are now becoming widely used in diagnostic medicine and are very successful as therapeutic agents in cancer (Oldham, 1983). In clinical diagnostic medicine they have provided us with the sensitivity not heretofore available for specific and rapid diagnosis of a particular drug level or infectious disease process.

Antibodies are important in the body as defense against infectious agents. They are extremely specific proteins that are produced in response to a foreign material, or antigen, by lymphoid cells of the immune system and share the property of being able to bind specifically to the inducing antigenic epitope (a single antigenic determinant; that portion of the antigen which combines with the antibody paratope). Unfortunately, under most conditions of antigenic stimulation, a family of antibodies is produced, each with a slightly different antigenic specificity.

In 1975, Kohler and Milstein observed that if an antibody-producing cell was fused with a myeloma tumor cell, a rapidly dividing hybrid was produced

that synthesized a monospecific antibody. Each hybridoma formed then became a "factory," producing antibodies monospecific to a particular sensitizing antigenic epitope. Cell cloning allows selection of hybrids producing antibody with the desired characteristics.

Monoclonal antibodies are thought to represent a major advance in cancer therapy because they have a very high therapeutic-to-toxic index when compared with anticancer drugs or radiation therapy and should provide a greater degree of specificity for the tumor cell than other forms of therapy. The conjugation of toxins with monoclonal antibodies is theoretically very exciting because a high specific toxin activity could be achieved at the tumor target cell.

Clinically, monoclonal antibodies are also proposed as drug delivery vehicles in certain tumors where specific tumor-associated antigens are expressed. In this context, investigators have found that by conjugating toxins such as the A-chain polypeptide of the plant protein ricin or the bacterial toxin from *Corynebacterium diphtheriae* to monoclonal antibodies specific for certain tumor type, as few as one or two molecules of antibody–toxin conjugate can destroy a tumor cell in vitro. Some success has also been obtained in clinical trials with monoclonal antibody–toxin conjugates.

Monoclonal antibodies have also been proposed for detoxification of individuals suffering from drug overdose or chemical intoxication as well as for radioimaging of tumor burden or metastatic foci. In veterinary medicine, monoclonal antibodies are already being used to develop new rapid methods for diagnosis of infections in poultry, cattle, and other animals.

### 5.1 Toxicological Concerns with Monoclonal Antibodies

It is already clear that there are certain problems implicit in the use of monoclonal antibodies in therapeutic trials. First, for example, there may be modulation of the antigenic determinant on the target cell, so that the monoclonal antibody cannot recognize its appropriate antigenic epitope. Second, the tumor cell may release free antigens so that the monoclonal antibody is effectively neutralized before it can reach the target cell. Third, antibodies to mouse epitopes on the monoclonal antibody could be induced (this may be overcome in the future by the use of human–human hybrids or the use of immunosuppressive agents to prevent the development of antibodies). Fourth, monoclonal antibodies have an extremely short half-life in systemic circulation, which would require that they be intermittently infused to provide the beneficial effect. Last, there may be an unwanted release of the toxin from its conjugate or specificity problems may develop whereby the antibody–toxin conjugates end up in an inappropriate organ.

The clinical toxicology findings associated with the use of monoclonal antibodies in therapeutic trials have included fever, chills, flushing, dyspnea, hypotension and tachycardia, anaphylactic and anaphylactoid reactions, urticaria, rash, nausea, elevated creatinine levels, headache, bronchial spasm, and serum

sickness (Oldham, 1983). Few of these reactions might be predicted from safety evaluation in rodents. A major problem with using the intact ricin or diphtheria toxin molecule, containing both the A and B polypeptide chains, has been the dissociation of the parent molecule from the monoclonal antibody, leading to toxicity of the reticuloendothelial system. A promising solution to this problem comes from separating the A chain (toxic moiety) from the B chain (cell association moiety) and preparing only A-chain conjugates. This results in much lower toxicity if the A chain should become dissociated from the antibody conjugate because cellular association does not occur.

Since a monoclonal antibody is a fusion product of a malignant mouse cell and an antibody-producing cell, there is some concern about the safety of the production process itself (Petricciani, 1983). Methods for the production of monoclonal antibodies raise two general safety issues: (1) the theoretical risk of transferring in the product factors associated with malignancy (e.g., oncogene factors) and (2) the use of animals for antibody production that are known to harbor a number of microbial agents some of which can produce diseases in humans.

Preclinical studies should address the potential toxicity due to inappropriate release of the conjugated toxin. Preclinical toxicology of monoclonal antibodies may not require extensive animal studies but should examine for cross-reactivity with antigenic epitopes present on normal cells in vitro and for the presence of human or rodent viruses. Early clinical trial should involve biodistribution studies with radiolabelled material.

The level of regulatory concern with the safety of the products of biotechnology underwent a sea change in early 2006 with the near catastrophe with TGN1412, a superagonist anti-CD28 monoclonal antibody that induces the production of anti-inflammatory cytokines by directly simulating T cells.

A German company, Tegenero, had the first-in-man clinical trial of the product initiated by a clinical research organization (CRO) at a clinic within a hospital in London. Eight healthy young males were enrolled, and six received a dose of the drug while two received placebo. Within 90 min after receiving a single intravenous dose of the drug, all six volunteers had a systemic inflammatory response characterized by a rapid induction of proinflammatory cytokines and accompanied by headache, myalgias, nausea, diarrhea, erythema, vasodilatation, and hypotension. Within 12–16 h after infusion, they became critically ill, with pulmonary infiltrates and lung injury, renal failure, and disseminated intravascular coagulation. Severe and unexpected depletion of lymphocytes and monocytes occurred within 24 h after infusion. All six patients were transferred to the care of the authors at an intensive care unit at a public hospital, where they received intensive cardiopulmonary support (including dialysis), high-dose methylprednisolone, and an anti-interleukin-2 (IL-2) receptor antagonist antibody. Prolonged cardiovascular shock and acute respiratory distress syndrome developed in two patients, who required intensive organ support for 8–16 days. Despite evidence of multiple cytokine release syndrome (CRS), all six patients survived (Goodyear, 2006; Suntharalingam

et al., 2008). This response has been characterized as a "cytokine storm," similar to that seen earlier with OKT3.

The preclinical evaluation which occurred before the trial was conducted was a set piece approach to the then guideline requirements. Studies were conducted in rats, mice, and primates, with the repeat-dose studies in the rats and mice being once a week for four weeks at doses up to 50 mg $kg^{-1}$, with these doses being characterized as well tolerated with transient increases in CD4 and CD8 (which were expected) and of IL-2, IL-5, and IL-6 but with no signs of first-dose CRS. There was cross-reactivity in stained lymphoid tissue and astrocytes in both human and primate tissues, but no histopathology signs of central nervous system (CNS) toxicity were seen in safety studies. The no-observable-adverse-effect level (NOAEL) was set at 50 mg $kg^{-1}$ in the primate (cynomolgus monkey), and with FDA HED (human equivalent doses) style allometric scaling and safety factor of 10 and then additional safety factors were applied, proposed 0.1 mg $kg^{-1}$ for first-in-man clinical trials, which was the dose used in the trial.

What was not done for this highly humanized monoclonal antibody (MAb) was to use the appropriate model (either a knockout mouse responsive to the humanized molecule or evaluation of a suitable homologue molecule).

The EMEA response has been to put in place the MABEL (minimum active biological effective level) approach to setting first doses in clinical trials (EMEA, 2007). In the case of TGN1412, based on biological activity in the rat, a safe dose would have been set at 0.005 mg $kg^{-1}$.

## 6 BIOPROCESS TECHNOLOGY

In the chemical and pharmaceutical industry, DNA technology will allow for the synthesis of chemical that can only be practically achieved through a bioprocess. For example, methylation of a particular carbon in a chemical structure might be done quite easily with a recombinant-engineered bacteria. This technology will allow the synthesis of a family of isomers and the development of a synthetic process that cannot be achieved by strict physical chemical processes.

This area of DNA technology also has application in the degradation of solid waste materials: in wastewater recovery, in leaching minerals from ore containing rock, in improved oil recovery, and in the decontamination of chemical waste dumps through the engineering of microorganisms that can destroy specific toxic contaminants.

## 7 GENE THERAPY PRODUCTS

Gene therapy products, while holding tremendous promise, have so far delivered but limited (two cases as of this writing) positive outcomes. The concept involved—inserting functioning genes in place (or places) where

nonfunctional or malfunctioning genes have produced a disease state—is stunning. But the public outcry over the death of Jesse Gelsinger, an 18-year-old in a clinical trial at the University of Pennsylvania's Institute of Human Gene Therapy, in 2000 has led to a significant slowdown in the rate of evaluation. This tragic event, probably due to an innate immune response to a protein in the vector's protein coat (Stephenson, 2001), has led to increased restrictions.

Five aspects specific to gene therapy need to be evaluated to assess the safety of a therapy—DNA/RNA biodistribution, gene transfer and biological activity, risk of vertical transmission of the gene, the safety of the vector (the means of delivering the gene to the intended site), and the safety of the product protein:

1. Evaluate DNA/RNA biodistribution:
   - Radiolabeling
   - Souther blot
   - Polymerase chain reaction (PCR)
   - Real-time PCR
   - In situ PCR
2. Evaluate gene transfer and biological activity:
   - Immunohistochemistry
   - Western blot
   - Enzyme-linked immunosorbent assay (ELISA)
   - Flow cytometry
3. Evaluate risk of vertical transmission:
   - To gonads
   - If yes to risk to gonads, then semen/germ cells (which is probably required anyway)
   - To circulating blood; if so, how long (persistence)
4. Assess safety of the vector:
   - Identify a suitable model species.
   - Assess the acute toxicity of the vector particle (in rabbits): at high dose, potential for anaphylactic response (not seen in mice). Also look for neutrophil proliferation.
5. Assess safety of the product protein:
   - Use data from preclinical pharmacology toxicology ("safety") studies to support the safety of clinical trials.
   - Those not appropriate for genotox. Instead, assess integration/insertion frequency in a mammalian cell.
   - Standard rodent carcinogenicity is *probably* not appropriate.

Regulatory authority for gene therapy products is unique in that overlapping responsibilities extend to both the CBER and NIH:

| | |
|---|---|
| CBER | Division of Cell and Gene Therapies: manufacturing<br>Division of Clinical Trial Design and Analysis<br>• Preclinical pharmacology and toxicology<br>• Clinical trial design, safety and efficacy<br>• 20% of CBER clinical protocols now for gene therapy |
| NIH | Recombinant DNA Advisory Committee (RAC): no authority for approval for clinical trials, but all adverse events must be reported |

This duplicating authority has led to both misunderstanding and problems in trials, as investigators must report adverse responses to both. But each has a different definition of what constitutes a reportable adverse response.

Currently, gene therapy is restricted to life-threatening and severely disabling diseases, but when a larger safety database has been accumulated, there should be expanded opportunities for therapy. It is not possible or desirable to identify a uniform "recipe" for safety studies that should be conducted with gene therapy products to support either the first dose in humans or extended clinical evaluation. Each product should be treated on a case-by-case basis, taking into consideration a number of important factors, such as the clinical indication, the duration of expression of the gene, and whether DNA transfer will be in vivo or ex vivo. For example, elimination of a tumor may require short-term treatment such as the transient expression of a suicidal gene. On the other hand, treatment may be long term, such as the replacement of a missing enzyme in the liver, where the goal may be lifetime expression. Gene therapy is currently an area of limited but rapidly advancing knowledge, and study design should be based on previous experience together with ongoing feedback from the clinic throughout development. Considerable early thought should also be given to appropriate assays and their sensitivities. The choice of assay will need to be justified and the basic "tooklit" of assays properly considered and evaluated in advance.

### 7.1 Vectors

In gene therapy, genes typically are delivered using a vector which may be nonviral or viral. The complete construct should be tested; separate safety evaluation studies of vectors per se are not generally recommended except to explore mechanisms of action if potentially harmful effects have been demonstrated in a previous investigation, for example, red cell agglutination on intravenous administration. If a novel nonviral vector is to be used, evidence of its lack of toxicity and information on its basic pharmacokinetics will be an essential component of the preclinical package. The interaction between the vector and the gene is quite important, perhaps more so with nonviral gene therapeutics, where the physicochemical properties of the particles themselves very much determine which tissues take up the gene. For viral vectors it is important to have sufficient knowledge of how they replicate, how to render

the viruses replication incompetent, any inherent pathogenicity and immunogenicity, and any risk of recombination with wild-type virus.

Conventional pharmaceutical quality assurance procedures should be applied to gene therapy products as well as appropriate infectivity tests for self-replicating/living vectors.

### 7.2 Studies to Support First Dose in Humans

The most scientific approach is to replicate, in an appropriate animal, the type of dosing that would be expected to be used in humans, employing the dose for dose–animal to human principle. A single, suitable animal species should suffice. If viral vectors are used, the animal species should be sensitive to infection by the wild-type virus. Studies should not automatically be done in primates but, initially, the commonly used laboratory species should be utilized. Only if those are demonstrated to be unsuitable should the next step be to consider the use of a primate.

Based on these general principles, the first dose in humans should be supported by a single-dose study in an appropriate animal species by the intended clinical route. Several dose levels should be explored, as some gene therapy expression products have a narrow therapeutic index.

There are many circumstances when a single-dose intravenous (IV) study can provide useful information; for example, if the intended treatment route is intraperitoneal (IP) or if the product will be administered to an open wound or injected into a muscle or a tumor, it might accidentally enter a blood vessel, and the knowledge gained from an IV study would be of value as well as one by the clinical route. Hence, if the intended clinical route is not IV, the absence of an additional study with IV dosing would require specific justification.

The physiological consequences of the gene product should be explored in these studies, particularly with totally novel gene products. In addition, all the standard toxicological evaluations should be carried out, including examination of functional endpoints in vivo, including cardiovascular and respiratory effects.

### 7.3 Distribution of Gene and Gene Product

The distribution of the gene must be evaluated carefully in time point assays. The choice of assay used, with regard to specificity and sensitivity, must be justified. The objectives are to identify the tissues in which the gene is present, demonstrate whether or not the gene product is expressed in particular tissues, and demonstrate the time course of gene expression, that is, how long it persists.

Some regulatory authorities, including those in the United Kingdom, are particularly concerned that the possibility of alteration to germ line cells should be excluded. Both male and female gonads should therefore be examined. If the gene is found there, then it is necessary to examine the gonads at

a more detailed level to ascertain whether the gene is present in the actual germ cells. Where gene persistence is short as, for example, in nonintegrating nonreplicating vectors, assay of the gonads at an appropriate time point will minimize false-positive findings.

### 7.4 Studies to Support Multiple Doses in Humans

The animal studies should parallel the intended treatment regimen in humans. At their present stage of development, a single dose of a gene therapeutic may not be totally curative, so multiple cycles of treatment may be used clinically rather than a single period of administration. It is appropriate to explore this in animal safety work; that is, the cycle regimen should be paralleled in the animal up to a maximum of three cycles. The duration of follow-up in the test animals after completion of the last test cycle should be based on the duration of gene expression up to a maximum of six months. There may be situations, particularly if long-term gene expression is the goal, when there could be an argument for a longer follow-up, but that should be considered on a case-by-case basis.

There should not be blind adoption of a checklist of assays and observations, but appropriate investigations should be selected based on earlier findings in the single-dose studies. Increasing the number of doses raises more concern about the immune response. There may be indications of this, such as lymphocyte infiltration at the site of administration, and there are many markers from single-dose studies that would indicate when it might be appropriate to examine the immune response to the gene product and selection markers, for example, immunity to adenoviral vectors, or to expressed proteins, resulting in accelerated loss of the transgene.

### 7.5 Unnecessary Studies

Genotoxicity and carcinogenicity studies are not generally recommended or required. Eliminating the possibility of the gene being inserted in the germ line ensures that some elements of concern about reproduction toxicology have already been addressed; hence, classical reproduction and developmental toxicity studies are not generally recommended. They should be considered on a case-by-case basis, for example, if the treatment were to be intended to manage a long-term metabolic disease and the patients would then survive to reach reproductive competence.

Drug interaction studies are not generally appropriate, with the exception of gene-directed enzyme prodrug therapy (involving a gene expressing an enzyme that activates a prodrug given subsequently). In the latter case, it is necessary to demonstrate the presence of the prodrug as well as the gene in the animal and to consider the potential toxicity of the active metabolite(s) both locally (which is the desired pharmacological effect) and systemically (the undesirable effects).

### 7.6 Ex Vivo Procedures

Ex vivo procedures involve removing cells and transfecting or transducing them. The cells should be checked to confirm that they are all healthy and are still expressing their normal surface markers, for example; observations of normal growth characteristics can also be reassuring. Animal studies are of limited value to test the safety of transfected or transduced human cells.

### 7.7 Change of Gene or Vector

Currently, only a limited number of vectors are available, although there is a large array of inserted genes. If the therapy involves developing a construct of a new gene in a well-characterized vector, it is important to use existing information on the vector. Rather than regenerating data on the vector itself, bridging studies of the construct should be carried out, that is, additional studies involving a limited toxicology evaluation to specifically characterize the nature of the new gene. Since the safety of the vector is already known, this should drive more exploration of the effects of the gene and the gene product rather than the vector itself.

If the vector is changed, a full safety evaluation may be required. However, if the changes are minor compared to the structure of a fully evaluated vector, it is appropriate for safety to be addressed by bridging studies. For example, if there is only a minor change on one of the condensation peptides of a nonviral, self-assembling vector, then some simple bridging work, rather than a full evaluation, may be appropriate.

The possibility of abbreviated testing is referred to in a guidance document on somatic cell and gene therapy published by the FDA (1991). According to this guideline, if changes are made to the vector backbone which do not alter the safety properties of the vector and the same route of administration and a similar dosing regimen is used to that employed previously, then truncated testing may be appropriate, depending upon the gene being expressed. When a promoter sequence or a targeting sequence in a viral or nonviral vector is changed or the expression of viral gene products is considerably altered, the vector should be considered as a new vector, even though it may have the same gene as the previous version of that vector.

### 7.8 Change of Route

It is quite possible that a treatment may be initiated using, for example, intratumoral injections to deliver a gene, which may subsequently be administered systemically. As a considerable amount of relevant information will have already been generated to support the intratumoral route, this would be another case for doing bridging studies. Comparative distribution studies will help to identify how much more safety evaluation may be required.

### 7.9 Insertional Mutagenesis

The long-term management of genetic disorders will require integration of the therapeutic DNA into the host genome or the maintenance of a stable episomal gene. The target of homologous recombination is still not achievable and, until that time, the problem of insertional mutagenesis—that is, inappropriate insertion of DNA into the host genome—must be addressed. How can the risk be quantified? Characterization of gene expression over time is more important than the copy number of the gene. An increased copy number can equate to an increased risk of insertional mutagenesis, but it also equates to an increase in the desired product. Insertional mutagenesis is a safety problem, and it is important to advise and warn patients who receive genes which will become integrated into the genome of this potential risk associated with their treatment.

## 8 VACCINES

Vaccination against viral and bacterial diseases has been one of the success stories of human and veterinary medicine. Probably the most outstanding example of the effectiveness of vaccination is the eradication of smallpox. In 1967 between 10 and 15 million cases of smallpox occurred annually in some 33 countries. By 1977 the last naturally occurring case was reported in Somalia. Polio too has been controlled in developed countries; for example, the number of cases in the United States was reduced from over 40,000 per year in the early 1950s, before a vaccine was available, to only a handful of cases in the 1980s. Diphtheria is now almost unheard of, whereas over 45,000 cases in 1940 led to 2480 deaths from diphtheria in the United Kingdom (similar numbers to those who died from AIDS in the United Kingdom in the entire 1980s). This has been reduced in the United Kingdom to only 13 cases and no deaths from the bacterium between 1986 and 1991. The scale of the problem is enormous—over 10 million deaths worldwide per year are due to infectious disease. United Nations figures suggest that cancers, circulatory problems, and injuries cause fewer deaths in developing countries than infectious diseases.

The process of developing vaccines is becoming increasingly complex due both to the nature of the infections being protected against and the nature of the cultures in which the affected individuals live. Kaufman (1996) provided an excellent review of this process and the inherent problems.

### 8.1 Approaches to Vaccination

There are two classical strategies for vaccination. One involves vaccination with either killed pathogenic organisms or subunits of the pathogenic organism. The other utilizes live attenuated viruses or bacteria that do not cause disease but have been derived from the pathogenic parent organism.

Inactivated vaccines are made from virulent pathogens by destroying their infectivity, usually with β-propiolactone or formalin to ensure the retention of full immunogenicity. Vaccines prepared in this way are relatively see and stimulate circulating antibody against the pathogen's surface proteins thereby conferring resistance to disease. Two or three vaccinations are usually required to give strong protection and booster doses are often required a number of years later to boost flagging immunity.

Subunit vaccines can be seen as a subcategory of inactivated vaccines because similar considerations apply to subunits and whole organisms. Doses, routes, duration of immunity, and efficacy of these vaccines are all very comparable. In this case a part of the pathogen, such as a surface protein, is used to elicit antibodies that will neutralize the infectivity of the pathogenic agent. The widespread use of hepatitis B virus surface antigen purified from the blood of carriers (or more recently from recombinant yeast) shows that this can be a very effective way to immunize. Hepatitis B virus surface antigen, the product of a single gene, assembles into a highly antigenic 22-nm particle which if used in three 40-μg doses at zero, one, and six months gives virtually complete protection against infection with hepatitis B virus.

Another example that can be included in the subunit vaccine class is the use of bacterial toxoids. Many bacteria produce toxins which play an important role in the development of the disease caused by a particular organism. Thus, vaccines against some agents, for example, tetanus and diphtheria, consist of the toxin inactivated with formaldehyde conjugated to an adjuvant. Immunization protects from disease by stimulating antitoxin antibody which neutralizes the effects of the toxin.

A further type of vaccine included in the subunit category is the capsular polysaccharide vaccine, for example, those against *Haemophilus influenzae* and meningococcal meningitis. In this case an extract of the polysaccharide outer capsule of the bacterium is used as a vaccine and is sometimes conjugated to protein to improve immunogenicity. Antibody persists for several years and is able to protect against the bacterium.

About half of all vaccines have traditionally been from live attenuated mutants of parent pathogenic organisms (Walker and Gingold, 1993). In effect live vaccines mimic natural infection yet produce subclinical symptoms and elicit long-lasting immunity, often giving rise to resistance at the portal of entry. Most of today's attenuated vaccine strains have been derived by a tortuous, often emperical route involving passage in culture until the pathogen is found to lose its virulence. This loss of virulence is tested in animal model systems before being tested in human volunteers. For example, the vaccine used to immunize against tuberculosis was derived after 13 years' passage in bile-containing medium by Calmette and Guerin (hence the name BCG—bacille Calmette–Guerin).

There has been much debate over the past 40 years as to the relative merits of live and killed vaccines often generating more heat than light! The evidence is that both routes will give adequate vaccines that can be used to protect

**TABLE 13 Relative Merits of Live versus Killed Vaccines**

| | Step | Live | Killed/Subunit |
|---|---|---|---|
| Production | Purification[a] | Relatively simple | More complex |
| | Cost | Low[b] | Higher |
| | Route | Natural or injection | Injection |
| | Dose | Low, often single | High, multiple |
| Administration | Adjuvant | None | Required[c] |
| | Heat lability | Yes | No |
| | Need for refrigeration[d] | Yes | Yes |
| | Antibody response | IgG; IgA | IgG |
| Efficacy | Duration of immunity | Many years | Often less |
| | Cell-mediated response | Good | Poor |
| | Interference | Occasional OPV only[e] | No |
| Safety | Reversion to virulence | Rarely[f] | |
| | Side effects | Low level[g] | No |

[a]Increasing safety standards mean that for new vaccines some of the older methodologies would not be acceptable.
[b]The price for new vaccines will approach that of killed subunit vaccines as safety standards are increased.
[c]Very few adjuvants for human use are acceptable.
[d]The need for refrigeration increases the costs significantly.
[e]Especially in the Third World.
[f]At very low levels (less than 1 case per $10^6$ vaccinations).
[g]This varies from occasional mild symptoms with rubella and measles vaccines to possible brain damage with pertussis vaccine.

against disease under the appropriate conditions. Table 13 summarizes the major points of debate. Many factors, including cost, safety, number of immunizations, ease of access to vaccines, politics, and social acceptance, will determine whether there is a high uptake of a particular vaccine and whether it is ultimately successful in eradicating the target disease. Even if a perfectly viable, relatively safe vaccine is available, uptake may be limited. For example, it has been estimated that vaccination against measles within the World Health Organization Extended Program on Immunisation (WHO EPI) has prevented over 60 million cases and 1.37 million deaths. Despite these efforts there are still some 70 million cases of measles annually, resulting in nearly 1.5 million deaths. Consequently a recent WHO congress adopted the following goals:

(i) Increasing immunization coverage
(ii) Improving surveillance
(iii) Developing laboratory services and improving vaccine quality
(iv) Training
(v) Promoting social mobilization
(vi) Developing rehabilitation services
(vii) Research and development

This again also serves to illustrate the importance of factors other than the efficacy of the vaccine itself in disease prevention.

The single most important issue in developed countries is the safety of a vaccine; a single death in a million vaccinations for a new vaccine would be unacceptable (except possibly if it were an effective AIDS vaccine). While this is obviously important in a third world country, other issues such as cost and how to deliver the vaccine are of paramount importance.

## 8.2 Genetic Engineering and Vaccine Development

Not all protective antigens are as simple to identify, clone, and express as the surface antigen gene of hepatitis. The entire sequence of the hepatitis B virus (HBV) genome became available, and as it is less than 10 kb, it was relatively simple to establish which open reading frame to express. It has been known for many years that irradiated malarial sporozoites protect against malaria. As the sporozoite stage in the life cycle of the malarial parasite can only be grown in small quantities, it was left to DNA technology to identify, clone, and express components of the sporozoite that might be of use in vaccine production. The genome of the malarial parasite is many thousands of times larger than the genome of HBV and therefore provides a different scale of problem. Not only was there little sequence data available, but there was also no idea of which gene products may be protective.

The starting point of any DNA work is to generate a library of DNA in *E. coli* which is representative of the organism under study. Once there is a complementary DNA (cDNA) bank or a genomic library, there are three basic ways of identifying and isolating a gene of interest.

***DNA/Oligonucleotide Hybridization*** If there is some preexisting knowledge of the nucleic acid sequence or where purified messenger RNA (mRNA) is available, it is possible to detect recombinant clones by hybridization of $^{32}$P-labeled DNA or RNA to bacterial colonies or bacteriophage plaques. Often a protein has been purified and some amino acid sequence is available which allows a corresponding nucleic acid sequence to be synthesized. Due to the degeneracy of the genetic code, a complex mixture of oligonucleotides is required to ensure that all possible sequences are represented. Labeling this mixture of oligonucleotides yields a probe that can be used to screen a cDNA (or possibly genomic) library that might be expected to contain the gene of interest.

***Hybrid Selection and Cell-Free Translation*** A second approach is to use hybrid selection of mRNA coupled with cell-free translation. DNA clones from a library, either individually or in pools of clones, can be immobilized by binding to a solid support and mRNA hybridized to them. Only the mRNA that corresponds to the clones will bind, and this can then be eluted and translated to protein in a cell-free system. The protein can then be immunoprecipitated

with antisera to the gene product of interest or assayed for activity. An example that encompasses both this approach and the sequence route is in the development of a vaccine for Epstein–Barr virus (EBV). It had been known since 1980 that antibody to the major membrane antigen of the virus (gp350/220) would neutralize the virus. Around 1983 a fragment of the virus genome was cloned and sequenced; using computer predictions the gp340/220 gene was identified. The experimental evidence that confirmed this prediction was published in 1985 and came from experimental work that managed to hybrid select EBV mRNA using genomic DNA clones. This was followed by cell-free translation of the eluted mRNA and immunoprecipitation of gm350/220 with a high-titer antibody. The DNA clone that hybridized with the gm340/220 mRNA was the one predicted to encode the gp340/220 gene by computer analysis. The hybrid selection approach is rather labor intensive and has for the most part been superseded by one of the forms of expression cloning.

***Expression Cloning*** This approach is invaluable when the only means of identification are antisera against the protein or pathogen of interest.

Probably the most laborious form of this approach is its use in conjunction with a biological assay. Complementary DNA libraries are cloned into a plasmid that will allow expression in eukaryotic cells, for example, simian virus 40 (SV40) or EBV vectors. Clones or pools of clones are then transferred to appropriate cell types; for example, COS cells for SV40 vectors and cell extracts or cell supernatant is assayed for biological activity. If a pool of clones gives the biological activity, then the individual clones can be reassayed and the desired cDNA clone identified. This methodology, although tedious, has allowed many of the interleukin genes to be cloned probably because the assays for these proteins are very sensitive.

Other gene products or vaccine antigens may require an enrichment step. For example, many genes expressed on the cell surface (e.g., receptors, adhesion molecules) have been cloned by "panning" techniques where the cells expressing the gene of interest are selected out either with antibody or by interaction with other cells. Complementary DNA libraries are constructed in *E. coli* and the library is transferred to eukaryotic cells. Those cells expressing the gene of interest are enriched for and the library transferred back to *E. coli*. This can be done for several rounds of expression and eventually individual clones conferring the selected phenotype will be isolated.

The most extensively used form of expression cloning involves the use of plasmid or bacteriophage vectors in *E. coli* and identification of DNA clones using antisera to the protein of interest. Here a vector such as the bacteriophage λgt11 is set up so that, when cDNA fragments are cloned into sites adjacent to the β-galactosidase gene, bacteria will express a β-galactosidase fusion protein containing epitopes present in the cDNA. Recombinant phage is detected with antisera. The cDNA insert is then sequenced and the whole gene can then be isolated in a more traditional way. The antisera used can be

monoclonal antibodies, polyclonal monospecific antisera, or even polyclonal antisera with many antibody specificities present. A variation on this method allowed the initial cloning of the malarial sporozoite surface antigen. The malaralial sporozoite stage cDNAs were introduced into the ampicillin resistance gene of the plasmid pBR322. Low levels of expression of the sporozoite surface antigen were detected by solid-phase radioimmunoassay using a monoclonal antibody specific for the protein. In this way a cDNA clone coding for the antigen was isolated and subsequently sequenced. This information was then used to design peptide vaccines which have already been tested in humans.

The λgt11 system is a more sophisticated bacteriophage version of the plasmid system described above and has been used to isolate many different antigens from various stages in the life cycle of the human malarial parasite using human immune sera as well as antigens from pathogens.

***Expression of Potential Vaccine Antigens*** In general, in the future eukaryotic cell culture is likely to be the method of choice for the production of subunit vaccine antigens where the organism to be vaccinated against replicates in eukaryotic cells. *Eschirichia coli* is unable to posttranslationally modify some vaccine candidates; for example, bacterial systems cannot add carbohydrate, which is important in the antigenicity and structure of many protective antigens.

Since 1986, the FDA has approved 22 vaccines (Table 14), half of them from a genetic engineering (and all, of course, from a biotechnology source). The cells used for such genetic engineering production of vaccine can be mammalian, insect, or bacterial.

The CBER has provided broad guidelines on the evaluation and production of vaccines (CBER, 2000). In general, the center's requirements have paralleled those for other biotechnology products. Beyond establishing sterility, lack of pyrogenicity, and lack of viral contaminants, a single GLP toxicity study in an appropriate species (one that has been established, if possible, to be immune responsive to the vaccine) is required. If the vaccine is to be used in pregnant women or women of childbearing potential, a segment I style reproductive study should be performed in an appropriate animal species.

Regulatory guidance for the conduct of clinical trials on vaccines is specific. Traditional phase I trials in normal volunteers are not conducted. Rather, all trials assess not only safety but also efficacy (or at least immunogenicity). Trials may well be challenge trials; that is, after immunization subjects are purposely challenged with exposure to the infective agent of concern.

In any case, injection site responses (erythema, edema, pain, and tenderness) and systemic responses are both evaluated in subjects (Mathieu, 1997).

The FDA also has specific guidance on the tracking and reporting of adverse clinical responses to vaccines.

Any adverse events of product problems with vaccines should be sent not to MedWatch but to the Vaccine Adverse Event Reporting System (VAERS), operated jointly by FDA and the Centers for Disease Control and Prevention.

**TABLE 14 Vaccines Approved by FDA since 1986**

| Vaccine | Indication | Date Approved | Company |
|---|---|---|---|
| Recombivax HB | Hepatitis B | June 23, 1986 | Merck, Chiron |
| ProHIBIT | *Haemophilus influenza* B | 1988 | Connaught |
| Pedvax | HIB | 1989 | Merck |
| Engerix-B | Hepatitis B | 1989 | SmithKlineBeecham |
| Tetramune, Hib TITR Haemophilis B, diphtheria CRM 197 protein conjugate | Bacterial meningitis | Jan. 1991 | Lederle-Praxis Biologics/American Cyanamid |
| IPOL | Poliovirus vaccine inactivated-injected | 1991 | Inst. Merieux |
| Acel-Imune | Diphtheria, tetanus toxoids and acellular pertussis vaccine | Jan. 6, 1992 | Takeda Chemical Industries/Amercan Cyanamid |
| Tripedia | Diphtheria, tetanus toxoids and acellular pertussis | Aug. 20, 1992 | Connaught Laboratories, Inc. |
| JE-VAX | Japanese encephalitis | Dec. 18, 1992 | Connaught/Biken |
| Enzon | Bubonic plague | 1994 | Green Labs |
| Typhim Vi | Typhoid | 1994 | Laboratories, Inc. |
| Havrix | HAV | Mar. 1995 | SmithKlineBeecham |
| Varivax, Varicella Virus Vaccine Live | Chicken pox | Apr. 10, 1995 | Merck |
| VAQTA | Hepatitis A | Mar. 29, 1996 | Merck |
| COMVAX | Haemophilis B and hepatitis B | Oct. 2, 1996 | Merck |
| Infanivir | Diphtheria, tetanus, and pertussis (DTP) | Jan. 27, 1997 | SmithKlineBeecham |
| Rabovert | Rabies (pre- and postexposure) | Oct. 27, 1997 | Chiron/Behring |
| Certina | DTP | Jul. 29, 1998 | North Amercan Vaccine |
| RotaShield[a] | Rotavirus | Aug. 31, 1998 | Wyeth |
| LYMEriv | Lyme disease | Dec. 21, 1998 | SmithKlineBeecham |
| Prevnar | Pneumococcal disease | Feb. 17, 2000 | Lederle |
| TWINRIX | Hepatitis A and B | May 11, 2001 | SmithKlineBeecham |

[a]Subsequently withdrawn.

For a copy of the VAERS form, call 1-800-822-7967 or download the form (in PDF format) from www.fda.gov/cber/vaers/vaers1.pdf.

## 9 SPECIFIC CHALLENGES

The problem with using a classical toxicological approach for evaluating an rDNA product or species-specific protein is that standard protocols are probably inappropriate and nonrelevant in most cases. In the traditional approach

to toxicology, a standard protocol or battery of tests is performed followed by an estimation of the types and degree of hazard or risk to humans. For example, conventional toxicity testing of a new rDNA product might lead to evaluation at excessively high doses in two rodent species. The production of antibody in the test species during preclinical toxicology testing may inactivate the test compound and thus invalidate the toxicity evaluation.

This approach appears somewhat irrational and without much scientific merit since many of these new molecules are minimally toxic or nontoxic by this sort of acute evaluation. As in the case of IFNs or monoclonal antibodies, the toxic effects observed in humans might not be predicted from safety assessments in rodents. An appropriate test species should be selected. Is the rat or mouse the appropriate species to evaluate a species-specific rDNA protein such as human growth hormone or IFNs or would nonhuman primates be more suitable? Does the nonhuman primate really offer any advantages? There is some consensus that the nonhuman primate may be a more appropriate species for testing some rDNA human proteins.

In contrast, in the "pharmacological approach" to toxicology, the potential targets of toxicity are first identified (Zbinden, 1986). The criteria for relevant effects are established, usually based on experience with reference substances, and appropriate in vivo or in vitro experimental models are selected to assess the pertinent toxicological responses.

Doses should be selected that are reasonable multiples of the proposed therapeutic dose to be employed, especially since in many cases the amount of material available for testing may be limited and not available in kilogram amounts. Preclinical rodent or primate studies should merely provide the flags to monitor during phase I clinical trials. Reason should prevail, not only in the selection of methods and models for assessing the potential toxicity of the new agents, but also in the use of these data for extrapolation to humans. Whether U.S. industry succeeds or fails in the biotechnology arena will depend on the quick resolution of issues such as selection of appropriate toxicological tests, fermentation scale-up of the rDNA microbe, product purity, and expedition of regulatory pathways.

### 9.1 Purity and Homology

Major concerns in the production of a species-specific protein by rDNA technology are the purity of the product, the amount and type of contaminants present, and the homology of the product to the native molecule (Table 15). The toxicologist should be concerned about the acceptability and toxicity of intentional or inadvertent contaminants introduced during fermentation or isolation of the product (e.g., DNA, chemicals, *E. coli* proteins). Other issues concern the introduction of amino acid residues that might alter the three-dimensional structure or antigenicity of the molecule, partial denaturation of the product during isolation and recovery, genetic stability of the rDNA clone during production (mutation could result in altered amino acid sequence), and

**TABLE 15 Issues in Safety Evaluation of Species-Specific rDNA Products**

| |
|---|
| Purity |
| Homology to native molecule (amino acid sequence, extra amino acids, three-dimensional structure) |
| Type and amounts of contaminants (chemicals, *E. coli* proteins, fermentation products, foreign DNA) |
| Stability of clone |
| Immunogenicity |
| Toxicities (direct or secondary to therapeutic effect) |

the level of foreign DNA present. Although these are issues of analytical biochemistry, their impact on the potential toxicity and overall safety of the finished product is of some toxicological concern.

## 9.2 Immunogenicity

The problem of the immunogenic nature of many human recombinant DNA proteins and the potential to generate antibodies to a normal human protein are of special interest to the immunotoxicologist. For example, 3 of 16 patients administered the rDNA-derived interferon-α (IFN-α) (clone A) developed antibodies of the immunoglobulin G (IgG) class that were undetectable prior to or during therapy (Gutterman et al., 1982). These antibodies were capable of in vitro neutralization of IFN activity, although in vivo neutralization of IFN was not documented, since there are several different subtypes of IFN-α-containing epitopes not present on their own IFN subtype. Similarly, two patients treated with IFN-β for many months developed high-titered antibody, which in one case was correlated with an inability of the patient's fibroblasts to produce IFN (Vallbracht et al., 1982).

Virtually all patients treated with conventional porcine insulin develop circulating anti-insulin antibodies (Klaff et al., 1978) that are less frequent and in lower titer in individuals treated with more highly purified porcine (Falholt, 1982) or rDNA human insulin (Fineberg et al., 1983). In the study by Fineberg and associates, 44% of the patients developed antibodies to rDNA human insulin over a 12-month period compared to a 60% antibody frequency with porcine insulin. Human growth hormone (HGH; Genentech) prepared by rDNA technology was observed to produce a frequency of immunogenicity similar to that seen with human insulin (approximately 40% of the children developed antibody, according to the product insert). The ultimate goal is to develop rDNA products that will be less immunogenic than purified animal sources of these therapeutic agents.

The exact mechanism of the immunogenicity of species-specific rDNA proteins is unknown but is believed to be attributable to (1) the addition of extra amino acid residues during synthesis, which the host reads as foreign;

(2) denaturing of the native molecules; or (3) contamination by *E. coli* polypeptides or lipopolysaccharides.

A second unanswered concern is whether the antibody induced by the recombinant protein has any discernible health effect. Other than some reports of neutralization of biological activity, little pathology has been attributed to the presence of antibodies in patients given recombinant protein therapy. It should also be noted that the question of antibody specificity has not been well studied, so that it is entirely conceivable that autoimmune pathology or even an anaphylaxis response could be induced. Equally important is the concern that induced antibody might neutralize the endogenous hormone or protein that it is intended to replace or supplement.

A third consideration is that certain routes of administration may favor immunogenicity of recombinant proteins. In early trials, rDNA proteins introduced by subcutaneous or intramuscular injections (procedures known to improve the immunogenicity of proteins) resulted in a higher frequency of antibody responses than in the intravenous route.

In summary, these are the clinically relevant questions about the immunogenicity of rDNA species-specific proteins: Will antibody be induced in the recipient that will neutralize the therapeutic effect or lead to immune complex disease? What is the class (e.g., IgG or IgE) and specificity (i.e., reactivity against specific protein or contaminant) of the antibody induced? The former antibody type could potentially neutralize the product and produce immune complex disease, while the latter could result in an anaphylaxis response. It is possible that the antibody induced is of insignificant health consequence, and its presence is known only because of improvements made in the sensitivity of detection methods with the introduction of ELISA.

## 10 PLANNING SAFETY EVALUATION PROGRAM

Safety evaluation of a candidate product should start with a consideration of the specific nature and consequential hazards of the three P's:

- Producing system
- Process
- Product

The need under each heading is to decide what data are required, then how to obtain them with the greatest efficiency and economy, and last whether the toxicologist is necessarily the person with the appropriate skills and experimental techniques to do so. There will often be a trade-off between precise control by other means and possibly cheaper or more familiar, old-fashioned toxicological studies (Dorato and Vodicnik, 2001). The inventor of a new product or process, too, may often have to do a great deal of work to show

safety by excluding hypothetical hazards, which subsequent manufacturers can afford to ignore.

***Producing System*** The questions of particular concern here are the nature of the system used to manufacture the desired substance and the precision with which it is controlled.

If the system consists of prokaryotic cells, then how well defined is their provenance and how is their consistency demonstrated? If mammalian cells are employed, their lineage must be considered. In both instances, it is important to ensure that extraneous virus, infections, DNA, and less well-defined factors such as "slow viruses" are excluded by the origins and history of the producer strain or because the physical (e.g., filtration) or chemical (pH, solvents, affinity separation) nature of the production process can be relied upon to exclude passage of an infectious agent.

If the degree of safety arising from these factors is weak, the toxicologist should consider appropriate studies in vivo to exclude contaminating agents, oncogenic factors, and so on, but there is no point in doing short-term or prolonged animal experiments or other types of tests unless the desired endpoint has first been clearly defined.

The aspect to which far more attention has been directed is the nature of the inserted gene(s) and promoters in rDNA products. Again, the toxicologist should ask how well the nucleotide sequence is known, whether there is only one reading frame, and how any introns are handled. Again, toxicity-type testing would appear to be an inefficient and expensive way to study molecular biology and biochemistry.

Last under this heading, for intact infectious organisms to be used directly in humans, is assessment of pathogenicity to the range of individuals that make up our populations, the possibility of reversion to a wild and more dangerous strain, the hazard of an allergenic reaction to the organism (e.g., vaccinia), and the possibility of spread from subject to subject in a naive or incompletely immune population.

There may be some role for animal experimentation here, if there is a suitable model, because it gives the chance to study the organism under intense pressure from commensals and the rising immune response. Possible hazards in the manufacturing plant also need to be evaluated.

In general, conventional toxicity procedures seem to have little to offer here, except in specific instances of helping to exclude certain infection factors, perhaps ruling out oncogenicity and examining the stability of engineered organisms for direct infection of humans.

***The Process*** The toxicologist has the least to offer here. In fact, only his or her intellectual analysis and review of the literature should be required to asses the manufacturing process and any residues of its chemicals and so to set analytical limits on purity and residues in the final preparation.

***The Product*** There are two distinct and probably divergent forces affecting the way in which the toxicologist regards the final product: (1) whether there is a need to learn about the biology and pharmacology of the product and (2) to define the minimum scientific concern, clinical caution, and industrial path before clinical trial or marketing.

***Biology of Bioengineered Products*** This may not be a useful concept scientifically, but it represents the practical point that the pace of development often forces the rapid sequence—interesting biological property—identification of responsible molecule in very small amounts (e.g., tumor necrosis factor or erythropoietin)—cloning, for example, resulting in large-scale production, perhaps even before the full structure is known. The clinical interest in administering a substance to humans for investigative or therapeutic purposes must be balanced against the total lack of knowledge of its general effects on the body or of the consequences of prolonged high-level exposure to it.

An example is the history of interferon: discovered through its antiviral actions, subsequently found to modulate mitosis and certain immune functions, capable of producing fever, and probably ECG and EEG changes as evidence of membrane effects in excitable tissues. If interferon were a novel discovery, now just being produced for the first time, then investigation of its general biological effects on repeated administration to responsive animals would be important prior to study in humans. The same arguments apply to lymphokines, for example, IFN-α, IL-2.

The planning of this type of investigation as an empirical, open study of responses must be carefully related to the nature and what is known of the product:

(i) It is necessary to work in a species capable of responding to the principal activity. Interferons are notorious for their species specificity, but most other lymphokines at least are more generally active. Work in a primate may be required, but it depends on the substance to be tested. There may be no point in using more than one species in pivotal studies.

(ii) Any test should be as broad and as general as possible, that is, monitor many variables clinically, in the laboratory and by pathology, until enough is known for there to be confidence in a focused approach.

(iii) Relate any testing to the clinical circumstances of probable use.

Thus, if a new synthetic antigen or engineered antigen is for testing, for administration only a few times to humans, there would be no point in a multidose experiment. It should suffice to show that it was antigenic in the intended preparation. Unless there were a prior reason to do so, a special search for, say, autoimmune reactions seems unnecessary. Similarly, testing a monoclonal antibody for activity is likely to be difficult, if not impossible, because of species specificity and the antigenicity of the preparation.

(iv) The toxicologist should be prepared to do nothing if the material is well known, its properties are understood, and there is adequate characterization of the nature of the preparation supplied; for example, human insulin or growth hormone produced by genetic engineering should not be submitted to prolonged safety tests in animals provided that the molecular forms present are sufficiently well understood.

It may be useful, however, to consider limited animal studies to examine the pharmacokinetics and duration of action even of a well-known material made by a new route unless physicochemical analyses show that to be pointless.

### 10.1 Animal Models

Species selection is probably one of the most important considerations when designing a preclinical safety program; for a biotechnolically derived pharmaceutical, it requires an understanding of the biology of the product. Since most of these either are human proteins or target human receptors, they tend to be species specific. Studies in rodents and dogs, the species commonly used in traditional toxicity studies, may not provide scientifically meaningful data. However, nonhuman primates are not necessarily the most appropriate species either, despite their phylogenetic similarity to human beings.

Some approaches that offer guidance in selecting relevant species are a literature review, determining the extent of homology between the endogenous animal protein and the human recombinant protein, determining the activity of the protein in pharmacological models, and in vivo assays of the receptor/tissue binding.

A literature review may provide useful information about the physiological properties of the protein in animals and how they compare with those of the human protein in humans. For example, prior to recombinant DNA technology, growth factors and/or hormones were purified from biological fluids. Although the quantities obtained were limited, they were nevertheless sufficient to allow investigation of the physiological properties of these proteins. Computer programs are now available for online searching of databases which hold information not only on the sequences of various animal and human proteins but also on the extent of homology between an animal protein and its human equivalent, including common amino acid sequences. It should be remembered, however, that a protein showing a high degree of honology to the human protein may not necessarily share similar pharmacological activities. Evaluation of activity or lack of activity in pharmacological animal models, if available, certainly would aid species selection. Finally, in vitro assays which analyze receptor and/or tissue binding are commonly used to determine the appropriate species for preclinical safety evaluation.

Some biotechnogically derived pharmaceuticals will cross-react with species that can be evaluated toxicologically, while others cross-react only with nonhuman primates such as the chimpanzee—a protected species. In this case, a

**TABLE 16 Alternative Models for Toxicity Assessment**

| Model | Example | Caveat |
|---|---|---|
| Nontraditional animal model | Transgenic mice carrying appropriate human receptor | Antibody formation would need to be monitored, as it is probable that a large human protein would produce an immune response |
| Homologous proteins and/or systems | Testing purified animal protein in the same species or, for monoclonal antibodies, testing an antibody directed against the receptor in the animal | Data should be interpreted with caution as the biological properties of the animal protein may differ from those of the human protein |
| In vitro methods | Tissue-binding assays | If no in vivo models are available, in vitro; methods combined with in vivo testing in a pharmacologically nonreactive may suffice |

well-designed "safety," or "Phase 0," study at doses higher than the proposed clinical dose may provide valuable safety information. However, a lack of cross-reactivity with any nonhuman species does not necessarily make preclinical safety evaluation impossible, nor does it limit toxicity testing to species in which the protein lacks relevant pharmacological activity. Some alternative possibilities are summarized in Table 16.

Toxicity studies traditionally are conducted using "normal" animals. However, studies in animal disease models may provide additional safety information regarding the possibility of disease exacerbation. For example, the administration of human recombinant erythropoietin was associated with hypertension in patients with chronic renal failure and also in uremic dogs but not normal dogs.

Species differences must be considered when choosing a model and, in particular, species-specific immunological differences between the human and the test animal. For example, in humans, an anti-CD4 MAb will bind to CD4 expressed on monocytes, with subsequent fixing of complement and destruction of antigen-presenting cells. However, since CD4 molecules are not expressed on murine monocytes, these effects would not be evident in a murine model.

For highly humanized proteins, the approach to proper safety evaluation starts with identification of an appropriate nonclinical safety model.

First, evaluate for comparative tissue binding. Then if no appropriate species is identified, one can either:

- Preparing and testing a homologue
- Testing the molecule itself in a humanized mouse model

A tissue cross-reactivity study is required by the FDA points-to-consider document for monoclonal antibodies prior to the first clinical trial. Its usefulness is debated, but its purpose is to evaluate the potential for binding to

nontarget tissues. Positive and negative controls are important for interpreting results. Tissues from humans and all nonclinical species used in safety studies should be evaluated. At minimum, the evaluation should include the 32 tissues recommended by the FDA "Points to Consider in the Manufacturing and Testing of Monoclonal Anitbody Products for Human Use."

When is it appropriate to use a homologue? According to the ICH S6 (1996) guidance: "When no relevant species exists, the use of homologous proteins should be considered. While useful information may also be gained from the use of homologous proteins, it should be noted that the production process range of impurities/contaminants, pharmacokinetics, and exact pharmacological mechanism(s) may differ between the homologous form and the product intended for clinical use."

Comparability of the homologue with the clinical candidate is critical:

Characterize pharmacology

- Literature: What's known about the target in the test species compared to humans?
- In vitro binding: Similar affinity or neutralization?
- Functional assays: in vitro cells in vivo bioassays (if possible relevant).
- Similar tissue distribution (tissue cross-reactivity for MAb).

Pharmacokinetics

Is Fc activity important and similar?

**Challenges of Homologues**

They represent a second test article. The decision must be made early in development:

- Sometimes, it is not possible to make a homologue.
- If possible, months to years needed to develop construct, clones, manufacture material, characterize pharmacology, and/or establish bioanalytical support.
- May be immunogenic, thus limiting usefulness.

How do you interpret the data?

- No "validation" can be performed that homologue is predictive of human toxicities.
- What if findings are different from the clinical candidate in an appropriate toxicology species?
- How do we extrapolate safety margins to the clinical candidate?

**Regulatory Challenges of Homologues**

No common criteria for what's expected.

- How much comparison with the clinical candidates is enough?
- Expectations for analytical characterization and does this need to be comparable?

- Do all aspects of testing need to be good laboratory practices (GLPs)?

Can studies with a homologue replace studies with the clinical candidate?

- Development and reproductive testing.
- What if results are more severe than with the clinical candidate?

Are negative findings meaningful?

- It's not your clinical candidate, so do the data impact risk assessment?

Homologues have been used to support registration for infliximab (Remicade) (anti-TNF) and efalizumab (Raptiva) (anti-CD11a). In both cases, there was not an appropriate species for the clinical candidate. Chimpanzees were the only pharmacologically responsive species, but they are not acceptable for toxicity testing due to humane reasons.

**"Humanizing" Mice**

- Isolate CD34+ stem cells [human hematopoietic progenitor cells (hHPCs)] from human cord blood.
- Breed NOD. Cg-Prkdc$^{scid}$//2rg$^{tmWj1}$.
- PND1 pup irradiated 1 cGy.
- hHPCs ($10^5$) injected in 25 μL phosphate-buffered saline (PBS) into liver.
- Assay tail blood for reconstitution of human immune cells.

Humanized mice can be used to:

- Assess the in vivo influence of stressors and drugs on the development of immune cells.
- Evaluate how toxicants can modify in vivo human immune responses.
- Investigate whether hematopoietic stem cells from cord blood can be induced to develop into nonimmune cell types
- Determine the influence of mouse effects on human immune cell longevity and human immune cells on mouse longevity.

## 10.2 Study Design

It is questionable whether traditional toxicological paradigms are applicable to biological or protein agents. If they are not, then how can the clinician gain reassurance to administer the first dose to humans, move into multidose trials, and even assess the agent in combination with other established medicines or biological agents? Monoclonal antibodies, soluble cytokine receptors, and growth factors all have been used in patients for nearly a decade, providing a wealth of experience in this area from which to learn. One of the most striking lessons is that pharmacodynamic effects may appear long after dosing of the agent has been discontinued.

As a class, biotechnologically derived pharmaceuticals share certain characteristics which have influenced their preclinical development. They are proteins and therefore toxicity was expected to be minimal and limited to an exaggeration of their desired pharmacological effects, a myth which was ultimately exploded. These agents are designed to perturb specific molecular or cell-to-cell interactions, sometimes with minimal effect on the pathophysiology of the target disease. Owing to the species-specific nature of these agents, preclinical toxicology is usually limited. For example, if a primatized anti-CD4 MAb cross-reacts only with chimpanzee and human CD4, the species of choice for toxicity tests is the chimpanzee, the use of which is restricted by its limited availability.

These characteristics of protein agents give rise to problems in clinical development such that the traditional paradigm for preclinical testing may not be appropriate. The dose in animals may not be predictive of an appropriate starting dose for humans. A surrogate marker (e.g., CD4 cell counts in the preclinical chimpanzee model) may be useful in setting the initial human dose but may only serve to indicate a no-effect dose. Once in the clinic, trials conducted early in development are usually not sufficiently powered to distinguish effects due to the toxicity of the test agent from those due to, for example, the underlying disease and concomitant or previous medications. Finally, short-term (three- to six-month) preclinical studies do not necessarily predict the long-term effects of these agents.

The long-term toxicities of concern are opportunistic infections, lymphoproliferative disorders, and immunogenicity, manifesting as tachyphylaxis and/or allergic reactions. Preclinical approaches which serve to identify these as potential hazards to humans of a biological drug moiety are thus needed.

The choice of toxicity studies and the design of individual studies will depend on the proposed clinical program. Important issues to consider are:

- Frequency and route of administration, including use of novel delivery systems
- Duration of dosing
- Special toxicity testing

***Frequency and Route of Administration*** Clinical trials for biotechnologically derived pharmaceuticals may be more complex than those for conventional pharmaceuticals and so the route and frequency of test drug administration should, if possible, mirror the proposed clinical use, even if that route employs a novel delivery system.

***Duration*** Traditionally, the duration of a toxicity study depends on the intended clinical use and disease duration. The potential immunogenicity of the human protein is a significant issue since antibody binding can partially or completely inhibit the biological activity of that protein, affect its catabolism, or alter its distribution and clearance. Any multiple-dose study therefore should include evaluation of the impact of antibody formation, including their neutralizing capacity. However, antibody formation in itself should not be a

reason for termination of a toxicity study, particularly if the antibodies are not neutralizing or do not alter the pharmacodynamics of the protein.

Multiple-dose toxicity studies are usually conducted before single-dose administration to volunteers. Many of the clinical trials for biological agents target life-threatening illnesses, and it has therefore been suggested that single-dose toxicity studies are sufficient to support single-dose "proof-of-concept" clinical studies. While this approach promotes faster introduction into the clinic, it may be of limited use since there may be a tendency to overlook the preclinical data. Clinical development may not progress without interruption if relevant preclinical data are missing.

***Special Toxicity Testing*** In addition to multiple-dose studies, information on potential functional changes—as obtained from safety pharmacology studies—and the potential for genotoxicity, reproductive toxicity, and carcinogenicity may be required for registration. Once again, the species specificity of recombinant proteins may preclude the use of traditional animal species such as rodents and/or rabbits for safety pharmacology, reproductive toxicity, and carcinogenicity studies. Functional evaluations of cardiovascular and pulmonary systems could be incorporated into a nonhuman primate multidose toxicity study. If appropriate, potential reproductive toxicity can be evaluated in a nonhuman primate.

There may be situations which warrant an assessment of carcinogenic potential, but immunogenicity and species specificity may preclude a two-year rodent bioassay. It may be necessary to develop in vitro assays to address a particular concern. For example, growth factors which may have the potential to support or stimulate the growth of transformed cells should be assessed for their ability to promote growth of either malignant or normal cells.

Large-molecular-weight compounds are unlikely to react with DNA or other chromosomal material and therefore a genotoxicity evaluation may be of little value. However, genotoxicity studies may provide useful information about the safety of products containing organic linkers.

***Program Design Considerations*** The standard toxicological data package for any new drug entity typically evaluates:

- Potential toxicity following single and multiple dosing
- Genotoxic potential
- Functional changes, that is, safety pharmacology studies

In addition, depending on the proposed clinical plan, the following may need evaluation:

- Toxicity following chronic dosing
- Carcinogenic potential
- Possible reproductive toxicity

Although biotechnologically derived pharmaceuticals often need customized preclinical development programs, certain issues are common to all. These include species specificity, potential immunogenicity and its impact on the duration of dosing, and the need for special toxicity testing.

## REFERENCES

Center for Biologics Evaluation and Research (CBER) (1997). Guidance for industry for the evaluation of combination vaccines for preventable diseases: Production, testing and clinical studies. CBER, Washington, DC.

Center for Biologics Evaluation and Research (CBER) (2000). Guidance for industry—Consideration for reproductive toxicity studies for preventive vaccines for infectious disease indications. CBER, Washington, DC.

Committee on Proprietary Medicinal Products (CPMP) (1989). Guidelines on the preclinical biological safety testing of medicinal products derived from biotechnology. *TIBTECH* 7:613–617.

Dorato, M. A., and Vodicnik, M. J. (2001). The toxicological assessment of pharmaceutical and biotechnology products. In A. W. Hayes (Ed.), *Principles and Methods of Toxicology*. Taylor & Francis, Philadelphia, PA.

European Medicines Agency (EMEA) (2007, Sept.). Guideline on strategies to identify and mitigate risks for first-in-human clinical trials with investigational medicinal products. EMEA, London.

Falholt, K. (1982). Determination of insulin specific IgE in serum of diabetic patients by solid phase radioimmunoassay. *Diabetologia* 22:254–257.

Fent, K., and Zbinden, G. (1997). Toxicity of interferon and interleukin. *TIPS* 8:100–105.

Fineberg, S. E., Galloway, J. A., Fineberg, N. S., Rathbun, M. J., and Hufferd, S. (1983). Immunogenicity of recombinant DNA human insulin. *Diabetologia* 25:465–469.

Folb, P. T. (2006). Animal tests as predictors of human response. In C-J. Lee, L. A. Lee, C. L. Wu, B. R. Lee, and M-L. Chen (Eds.), *Clinical Trials of Drugs and Biopharmaceuticals*. Taylor & Francis, Boca Raton, FL.

Goodyear, M. (2006). Editorial: Learning from the TGN1412 trial. *BMJ* 332:677–678.

Gresser, I., and Bourali, C. (1970). Development of newborn mice during prolonged treatment of interferon. *Eur. J. Cancer* 6:553–556.

Griffiths, S. A., and Lumley, C. E. (1998). *Safety Evaluation of Biotechnology-Derived Pharmaceuticals: Facilitating a Scientific Approach*. Kluwer Academic, Boston.

Gutterman, J. U., Fine, S., Quesada, J., Horning, S. J., Levine, J. F., Alexanian, R., Bernhardt, L., Kramer, M., Speigal, H., Colburn, W., Trown, P., Merigan, T., and Dziewanawska, Z. (1982). Recombinant leukocyte A interferon: Pharmacokinetics, single-dose tolerance, and biological effects in cancer patients. *Ann. Intern. Med.* 96:549–556.

International Conference on Harmonization (ICH) (1996). *Preclinical Safety Evaluation of Biotechnology-Derived Pharmaceuticals*.

International Conference on Harmonization (ICH) (1997). *Preclinical Safety Evaluation of Biotechnology-Designed Pharmaceuticals*. ICH, Brussels.

Kaufman, G. H. E. (1996). *Concepts in Vaccine Development*. Walter de Geruptor, New York.

Klaff, L. J., Vinik, A. I., Berelowitz, M., and Jackson, W. P. U. (1978). Circulating antibodies in diabetics treated with conventional and purified insulins. *S. Afr. Med. J.* 54:149–153.

Kohler, G., and Milstein, C. (1975). Continuous culture of fused cells secreting antibody of predefined specificity. *Nature* 256:495–497.

Levine, M. M., Kaper, J. B., Lockman, H., Black, R. E., Clements, M. L., and Falkow, S. (1983). Recombinant DNA risk assessment studies in man: Efficacy of poorly mobilizable plasmids in biologic containment. *J. Infect. Dis.* 148:699–709.

Mackett, M. (1993). Vaccination and gene manipulation. In J. M. Walker and E. B. Gingold (Eds.), *Molecular Biology and Biotechnology*, 3rd ed. Royal Society of Chemists, Cambridge.

Malinowski, M. J. (1999). *Biotechnology: Law, Business and Regulation*. Aspen Law & Business, New York.

Mathieu, M. (1997). *Concepts in Vaccine Development*. Walter de Geruptor, New York.

Maulik, S., and Patel, S. D. (1997). *Molecular Biotechnology: Therapeutic Applications and Strategies*. Wiley-Liss, New York.

Offit, P. A. (2005). *The Cutter Incident*. Yale University Press, New Haven, CT.

Oldham, R. K. (1983). Monoclonal antibodies in cancer therapy. *J. Clin. Oncol.* 1:582–590.

Petricciani, J. C. (1983). An overview of safety and regulatory aspects of the new biotechnology. *Regul. Toxicol. Pharmacol.* 3:428–433.

Stebbing, N., and Weck, P. K. (1984). Preclinical assessment of biological properties of recombinant DNA-derived human interferons. In A. P. Bollon (Ed.), *Recombinant DNA Products: Insulin, Interferon and Growth Hormone*. CRC Press, Boca Raton, FL.

Stephenson, J. (2001). Studies illuminate cause of total reaction in gene-therapy trial. *JAMA* 285:2570–2571.

Suntharalingam, G., Perry, M. R., Ward, S., Brett, S. J., Castello-Cortes, A., Brunner, M. D., and Panoskaltsis, N. (2008). Cytokine storm in a phase 1 trial of the anti-CD28 monoclonal antibody TGN1412. *N. Engl. J. Med.* 355:1018–1028.

Thomas, J. A., and Myers, L. A. (1998). *Biotechnology and Safety Assessment*. Taylor & Francis, Philadelphia, PA.

U.S. Food and Drug Administration (FDA). (1991). Points to consider in human somatic cell and gene therapy. FDA, Washington, DC.

Vallbracht, A., Treuner, J., Manncke, K. H., and Niethammer, D. (1982). Autoantibodies against human *beta* interferon following treatment with interferon. *J. Interferon Res.* 2:197.

Walker, J. M., and Gingold, E. B. (1993). *Molecular Biology and Biotechnology*. Royal Society of Chemists, Cambridge.

Weissinger, J. (1989). Nonclinical pharmacologic and toxicologic considerations for evaluating biologic products. *Regul. Toxicol. Pharmacol.* 10:255–263.

Zbinden, G. (1986). A toxicologist's view of immunotoxicology. In *Proceedings of an International Seminar in the Immunological System as a Target for Toxic Damage*.

# *Safety Assessment of Inhalant and Dermal Route Drugs*

While the two most common routes for administering a drug are either orally or parenterally (by injection), there are many other routes that can be used. The two next most common are inhalation (with all its variations) and dermally.

Each route has special considerations as to the evaluation of their safety—both regulatory and technological—which this chapter will consider in turn.

## 1 INHALED THERAPEUTICS

Drugs and medicinal agents administered by the inhalation route include the gaseous and vaporous anesthetics, coronary vasodilators, the aerosols of bronchodilators, corticosteroids, mucolytics, expectorants, antibiotics, and peptides and proteins where there is significant nasal absorption (Cox et al., 1970; Williams, 1974; Paterson et al., 1979; Hodson et al., 1981; Lourenco and Cotromanes, 1982). Concerns with the environmental effects of chlorofluorocarbons has also led to renewed interest in dry-powder inhalers, which have additionally shown promise for better tolerance and absorption for some new drugs. Excessive inhalation of a drug into the pulmonary system during therapy or manufacturing may result in adverse local and/or systemic effects. Consequently, safety assessment of inhaled medicinal preparations with respect to

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

pulmonary toxicity and the therapeutic-to-toxicity ratio are essential. The data generated are essential for charting the course of evaluation and development of a potential therapeutic agent. The art of evaluating their toxicity is a specialty area (Newton, 2000; Gad, 2006).

## 1.1 Pulmonary System

An average man inhales approximately 7.5, 28.6, and 42.9 L of air per minute during resting, light work, and heavy work periods, respectively, the corresponding mean tidal volumes being 750, 1673, and 2030 mL [National Academy of Sciences (NAS), 1958]. Each breath is distributed between 300 and 400 million alveoli, where gas exchange takes place. The total alveoli surface area is approximately 75 $m^2$, which is penetrated by approximately 200 km of capillary blood vessels (Hatch and Gross, 1964). The high vascularity and large surface area of the lung ensure rapid gas exchange and entry of an inhaled drug into the bloodstream. A drug is then quickly carried to the heart and brain before reaching the liver, where first-pass metabolism occurs. The pulmonary system is therefore a very effective portal through which gases, vapors, and aerosols can enter the body to exert desirable therapeutic effects and undesirable side effects locally and/or systemically.

Anatomically, the pulmonary system is divided into extrathoracic and thoracic regions. The extrathoracic, or head, region includes the nasal and pharyngeal passages. The thoracic region is subdivided into tracheobronchial (TB) and alveolar (AL) regions. The TB region consists of the trachea, primary and secondary bronchi, and primary through-ciliated bronchioles. The alveolar region consists of nonciliated terminal bronchioles, alveolar ducts, and alveoli (Lippmann, 1970). The anatomical structure of the pulmonary system maximizes gas exchange but minimizes the penetration of extraneous particulate matter into the lungs. The formalized anatomy (Davis, 1961; Weibel, 1963; Horsfield and Cumming, 1968; Parent, 1991) of the branchings, the dimensions of the airways, the penetrability by particles of certain sizes, and the distribution of cell types in the respiratory tract and lungs are summarized in Figure 1.

## 1.2 Penetration and Absorption of Inhaled Gases and Vapors

Pulmonary dynamics, the dimension and geometry of the respiratory tract and the structure of the lungs together with the solubility and chemical reactivity of the inhalants greatly influence the magnitude of penetration, retention, and absorption of inhaled gases, vapors (Dahl, 1990), and aerosols (Raabe, 1982; Phalen, 1984). The quantity of an inhalant effectively retained in the pulmonary system constitutes the inhaled "dose" that causes pharmacotoxic responses. Acute inhalation toxicity of vapors and gases is subject to concentration/time rules (Carpenter et al., 1949), but not repeat dose exposures.

| Regions | | | Generations | Total Cross section ($cm^2$) | Particle size penetration limits (μm) | Distribution | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | | Cell types | Mucous glands | Smooth muscle | Parasympathetic innervation |
| Extrathoracic | Head | Nasopharynx | | | 60 | | | | |
| Intrathoracic | Tracheo-bronchial | Trachea | 0 | 2.54 | | Ciliated | | | |
| | | Primary bronchi | 1 | 2.33 | 20 | Goblet | | | |
| | | Secondary bronchi | 2 | 2.13 | 10 | Brush Border Squamous | | | |
| | | | 3 | 2.00 | | | | | |
| | | | 4 | 2.48 | | | | | |
| | | | 10 | 13.4 | | Columnar | | | |
| | | Bronchioles | 11 | 19.6 | 6 | | | | |
| | | | 15 | 113 | | | | | |
| | | Terminal bronchioles | 16 | 180 | 4 | Cuboidal Less ciliated | | | |
| | Alveolar | Respiratory bronchioles | 17 | 300 | 3 | Less goblet Noncilicated clara | | | |
| | | | 18 | 534 | | | | | |
| | | | 19 | 944 | | | | | |
| | | Alveoli ducts | 10 | 1.60K | <3 | Type I | | | |
| | | | 21 | 3.22K | | | | | |
| | | | 22 | 5.88K | | Type II | | | |
| | | Alveoli | 23 | 11.8K | <3 | | | | |

***Figure 1*** *The distribution of cell types in the respiratory tract and lungs.*

Highly reactive and soluble gaseous or vaporous drugs react and dissolve readily in the mucosal membrane of the nasopharynx and the upper respiratory tract (URT), thereby exerting pharmacological effects or causing local irritation and/or adverse effects on the ciliated, goblet, brush border columnar, and squamous cells of the epithelium (Weibel, 1983). The dissolved drug is also absorbed into the bloodstream and transported to the target organ where it exerts systemic effects. Less reactive and less soluble gaseous or vaporous drugs are likely to penetrate beyond the URT and reach the bronchial and alveolar regions, causing local and systemic effects. The unabsorbed gases or vapors are then exhaled. For example, ammonia gas generated from a 10% ammonia water may be inhaled for reflex respiratory stimulation purposes (Budavari, 1989). Ammonia is extremely soluble in water at a concentration of 715 mL of ammonia per milliliter of water (Phalen, 1984) and is readily solubilized in the mucous lining, causing URT irritation. By contrast, oxygen is only sparingly soluble in water (0.031 $cm^3$ of oxygen per milliliter of water) and capable of penetrating deeply into the alveoli where gas exchange takes place. Oxygen that binds reversibly with the hemoglobin of erythrocytes is unloaded at the target tissues, while the unbound oxygen is exhaled. Inhalation of properly humidified oxygen is life supporting, but inhalation of unhumidified oxygen may cause a reduction in the mucociliary clearance of secretions in the trachea of animals (Pavia, 1984) and humans (Lichtiger et al., 1975; Gamsu et al., 1976). Gases or vapors of low lipid solubility are also poorly absorbed in the lungs, with much of the inhaled vapor exhaled. Other pharmacological gases and vapors, such as the anesthetics (nitrous oxide, halothane, enflurane, isoflurane, etc.) and the coronary vasodilators (amyl nitrite), likewise affect the epithelium of the respiratory tract and the lungs. The absorbed drugs exert local effects on various types of epithelial cells of the respiratory tract and on types I and II cells and the alveolar macrophages (AMs) in the alveoli. Repeated inhalation of some halogenated hydrocarbon anesthetics will result in accumulation of the vapors and systemic toxicity (Chenoweth et al., 1972). By contrast, vapors such as the fluorocarbons FC 11 and FC 12, which are used extensively as propellants for bronchodilator and corticosteroid aerosols, are absorbed rapidly but are not accumulated in the body even upon repeated inhalation (Aviado and Micozzi, 1981).

In general, dissolved gases or vapors at a nontoxic concentration are absorbed and metabolized locally by the lungs and systemically by the liver. The unchanged parent drug and its metabolites may be excreted to some extent via exhalation but mainly via the renal system. A dissolved gas or vapor at a toxic concentration, however, is likely to exert local effects such as altering the surface tension of the alveoli linings or disrupting the normal functions of the epithelial cells, the pneumocytes, and the AMs. The disrupted AMs in turn release their intracellular enzymes, causing destruction of the alveolar septa and contributing to histopathological changes of the respiratory tract and the lungs. Again, the magnitude of the adverse effects is dependent on pulmonary

dynamics and the solubilities of the inhalants in the mucous membrane of the URT and in the plasma or lipids of the erythrocytes.

## 1.3 Deposition of Inhaled Aerosols

For inhaled aerosols, particle size is the major factor affecting the penetration, deposition, and hence "dose" and site of pharmacological action (Dautrebande, 1962a,b; Agnew, 1984). Particle size is expressed in terms of *aerodynamic diameter* (AD), defined as the diameter of a spherical particle of unit density (1 g/cm$^3$) that has the same terminal settling velocity as the particle in question, regardless of its shape and density (Marple and Rubow, 1980). The unit for AD is the micrometer. A sample of aerosol particles having ADs within a narrow size range is considered to be a monodisperse aerosol, whereas a sample of aerosols with a wide range of ADs is a heterodisperse, or polydisperse, aerosol. The pattern of particle size distribution is usually bell shaped, with smaller and larger particles on both sides of the mean AD. An aerosol sample with a high proportion of particles of similar size has a narrow particle size distribution, or small geometric standard deviation (GSD). An aerosol sample with a GSD of less than 2 is considered to be a monodisperse aerosol. Thus, both the AD and GSD of 2 or less are considered to be optimal for pulmonary penetration and distribution in the respiratory tract and the lungs. For example, in nose breathing, aerosol particles with ADs > 15 μm are likely to be trapped in the nasopharynx (extrathoracic, or head, region) by filtration and impaction. Particles deposited in the nasopharynx are considered to be "noninhalable" (Lippmann, 1977; Miller et al., 1979).

In mouth breathing, only 10–15% of 15-μm particles penetrate through the larynx to the intrathoracic TB region. Particles reaching the TB region are considered to be "inhalable" (Lippmann, 1977; Miller et al., 1979).

In natural nose and mouth breathing, only a negligible proportion of aerosol particles of AD > 10 μm reach the lungs (Swift and Proctor, 1982). Aerosol particles of 3–4 μm in AD are considered to be optimal sizes for TB deposition. The mechanisms of deposition are by impaction along the trachea and at bronchial branchings where the direction of airflow changes and by gravity settlement in the fine airways in amounts proportional to the particle-settling velocity and the time available for settlement (Hatch and Gross, 1964; Heyder et al., 1980). Aerosol particles of 1–2 μm in AD, however, decrease in TB deposition because the particles are too small for effective impaction and sedimentation (Lippmann, 1977; Chan and Lippmann, 1980; Stahlhofen et al., 1980). Consequently, the majority of the very fine particles are exhaled. However, the deposition of the ultrafine particles of approximately 0.5 μm in AD on the walls of the finest bronchioles and the alveoli increases again due to molecular diffusion processes. Even so, some 90% of the inhaled 0.5-μm particles will still be exhaled during quiet tidal breathing and much more under forced exhalation (Davis et al., 1972; Taulbee et al., 1978). Those fine particles reach-

ing the finest bronchioles and alveoli are considered to be "respirable" (Lippmann, 1970).

In general, particles of AD > 10 μm deposit mainly in the URT (upper respiratory tract), whereas particles of 1–5 μm AD, with a GSD of less than 2, are likely to reach the lower respiratory tract, which includes the TB region and the alveoli, with small oropharyngeal loss.

The proportion of an aerosol sample suitable for inhalation can also be determined on the basis of mass median aerodynamic diameter (MMAD), which is defined as the percentage (50%) by weight of an aerosol sample having ADs equal to or less than the stated median AD. For example, a sample with an MMAD of 5 μm means that 50% by weight of that sample has ADs of 5 μm and smaller. The MMAD is therefore a good index for determining the proportion of an aerosol sample that is "noninhalable," "inhalable," or "respirable." An aerosol sample with an MMAD of 5 μm and a GSD of less than 2 is considered to be optimal for pulmonary deposition and retention (Task Group on Lung Dynamics, 1966).

In addition to AD and GSD, the pulmonary dynamics of a subject also greatly influence the distribution of aerosol particles in various regions of the respiratory tract (Agnew, 1984). For example, the velocity of airflow in the respiratory tract significantly influences the pattern of TB deposition. An increase in airflow velocity in the airways increases the effectiveness of particle impaction at the bifurcations of the large airways (Dennis, 1961; Hatch and Gross, 1964; Parent, 1991). As a result, spots impacted with a high concentration of particles (hot spots) are frequently present at the carina and the bifurcations of the airways (Lee and Wang, 1977; Bell, 1978; Stahlhofen et al., 1981). Furthermore, the depth of each breath (tidal volume) also influences the distribution of aerosols. A small tidal volume permits greater impaction in the proximal conducting airways and less sedimentation in the distal airways.

In general, slow, deep inhalation followed by a period of breath holding increases the deposition of aerosols in the peripheral parts of the lungs, whereas rapid inhalation increases the deposition in the oropharynx and in the large central airways. Thus, the frequency of respiration (the flow velocity) and the depth of breath (tidal volume) influence the pattern of pulmonary penetration and deposition of inhaled aerosols. Therefore, an aerosol of ideal size will penetrate deeply into the respiratory tract and the lungs only when the aerosols are inhaled in the correct manner.

## 1.4 Absorption and Clearance of Inhaled Aerosols

Soluble aerosols deposited on the epithelial linings of the respiratory tract are absorbed and metabolized in the same way as soluble gases and vapors.

Insoluble medicinal aerosols are few in number. Sodium cromoglycate (SCG) is probably the only insoluble powder to be administered as a prophylactic antiasthmatic (Wanner, 1979). Insoluble particles deposited on the ciliated linings of the URT are removed by a mucociliary clearance mechanism.

Particles deposited on a terminal airway devoid of ciliated cells may be endocytosed into the epithelial cells (Jones, 1984; Newhouse et al., 2000). At a toxic concentration, the cells die and the debris is then phagocytosed and transported into the interstitial space for removal via the lymph or vascular drainages or reenters the ciliated zone of the airway. Particles deposited in the alveolar walls are phagocytosed by the AMs and transported from the low-surface-tension surfactant in the alveolar lining to the high-surface-tension bronchial fluid of the ciliate airways for elimination by the mucociliary clearance mechanism (Lauweryns and Baert, 1977). The particle sizes optimal for phagocytosis are 2–3 μm, while particles smaller than 0.26 μm are less effective in activating the macrophages (Holma, 1967). In any case, AMs can phagocytose only a small fraction of a large number of deposited particles. The non-phagocytosed particles are translocated to the lymphatic system for elimination (Ferin, 1977).

Like the inhaled gases or vapors, soluble and insoluble aerosol particles can directly exert desirable and undesirable local effects at the site of deposition and/or systemic effects after solubilization, absorption, and metabolization (Sackner et al., 1975; Sackner, 1978).

## 1.5 Pharmacotoxicity of Inhaled Aerosols

The inhalation route for administering drugs into the pulmonary system for treatment of respiratory diseases eliminates many bioavailability problems such as plasma binding and "first-pass" metabolism, which are encountered in parenteral or oral administration. Consequently, a small inhalation dose is adequate for achieving the desirable therapeutic response without inducing many undesirable side effects. Furthermore, the direct contact of the drug with the target site ensures rapid action. Nevertheless, the effects from inhaled drug aerosols also depend on the pharmacological properties of the aerosols and the location of their deposition in the respiratory system. For example, the classic experiments on bronchodilation drugs (Dautrebande, 1962a,b) showed that fine aerosol particles of isoproterenol penetrate deeply in to the lower respiratory airways (LRA). In this way, a high concentration of the drug aerosol can reach the beta-adrenergic receptors of the bronchial smooth muscles. Stimulation of the receptors causes relaxation of the smooth muscle fibers and results in brochodilation (Weiner, 1984; McFadden, 1986). Such rapid bronchial responses can be produced in healthy and asthmatic subjects without inducing any cardiac effects. By contrast, the same dose of isoproterenol of large particle sizes deposits mainly along the URT, with a minimal amount reaching the smooth muscles of the LRA. The drug is quickly absorbed into the tracheal and bronchial veins and delivered immediately to the left ventricle of the heart. A high plasma concentration of the drug in the heart causes prominent cardiovascular effects such as tachycardia and hypertension. Other aerosols of beta-adrenergic drugs, such as epinephrine, isoprenaline, terbutaline, and salbutamol, induce bronchodilation effects in animals and

humans (Pavia, 1984) via inhalation and stimulate ciliary beat frequency and mucous production at the site of deposition in the trachea (Wanner, 1981). Thus the tracheobronchial mucociliary clearance mechanism is also stimulated. By contrast, anticholinergic bronchodilators, such as atropine and ipratropium bromide, cause mucous retention in the lungs (Pavia et al., 1983a,b). Therefore, in pharmacological or safety assessments of inhalant beta-adrenergic bronchial dilatation drugs, aerosols should be of small particle sizes suitable for deposition in the peripheral airways to minimize side effects. However, anticholinergic agents should be of larger particle sizes suitable for deposition in the large airways (Ingram et al., 1977; Hensley et al., 1978).

Other therapeutic aerosols—such as beclomethasone dipropionate, betamethasone valerate, and budesonide corticosteroid (Williams, 1974); the carbenicillin and gentamicin antibiotics (Hodson et al., 1981); the 2-mercaptoethane-sulfonate (Pavia et al., 1983b) and *N*-acetylcysteine (Hollinger, 1985) mucolytics; and even vaccines for the prevention of influenza and tuberculosis (Lourenco and Cotromanes, 1982)—are active by inhalation and/or oral administration. When these drugs are administered as aerosols, certain particle sizes may be targeted to a specific region or to multiple regions of the pulmonary system depending on the therapeutic target site(s). In any case, when aerosols are delivered as fine particles, the rate of absorption is increased because of an increase in the distribution area per unit mass of the drug. Thus, an effective aerosol dose of corticosteroid for treatment of asthma and bronchitis is merely a fraction of an oral dose (Williams, 1974). An aerosol of SCG dry powder, a prophylactic for preventing the onset of bronchoconstriction in asthmatic attacks (Cox et al., 1970), is effective mainly by local inhibition of the release of chemical mediators from mast cells in bronchial smooth muscle. Therefore, SCG particle sizes should be approximately 2 μm in AD for the most effective penetration into the bronchial regions (Godfrey et al., 1974; Curry et al., 1975). Likewise, therapeutic aerosols of local anesthetics and surfactants may require appropriate particle sizes to be targeted to a specific region of the pulmonary system.

Other than undesirable pharmacological effects, toxic concentrations of soluble or insoluble aerosol particles may lead to adverse physiological and/or histophathological responses. For example, irritating aerosols cause dose-related reflex depression of the respiratory rate (Alarie, 1966, 1981a), while phagocytosed particles cause chemotaxis of AMs and neutrophils to the site of deposition (Brain, 1971). The maximum response usually occurs at 24 h and returns to normal in approximately three days postexposure (Kavet et al., 1978). Furthermore, a toxic quantity of phagocytosed particles may interact with the lysosomal membrane within a macrophage, releasing cytotoxic lysosomal enzymes, proteases, and free radicals that in turn damage the adjacent lung tissue (Hocking and Golde, 1979).

In general, a specific category of drug delivered to a specific site of the pulmonary system will exert a specific pharmacological or toxicological action locally or systemically. Therefore, in safety assessments of inhalants, a drug

should be delivered to the target sites of the pulmonary system according to the toxicological information required.

Finally, there are many drugs in the categories of amphetamines, anorectics, antihistamines, antipsychotics, tricyclic antidepressants, analgesics and narcotics, and beta-adrenergic blocking agents that are known to accumulate in the lung (Wilson, 1982; Hollinger, 1985) even though these drugs are not administered via the inhalation route. Therefore, in safety assessments of these drugs, their pulmonary toxicity should also be evaluated.

## 1.6 Methods for Safety Assessment of Inhaled Therapeutics

Methods for evaluation of inhalation toxicity should be selected according to the pharmacological and/or toxicological questions asked, and the design of experiments should specify the delivery route of a drug to the target sites in the pulmonary system (Gad, 2006). For example, if an immunological response of the lungs to a drug is in question, then the lymphoid tissues of the lungs should be the major target of evaluation. The following are some of the physiological, biochemical, and pharmacological tests that are applicable for safety assessment of inhaled medicinal gases, vapors, or aerosols.

Upper respiratory tract irritation can occur from inhalation of a medicinal gas, vapor, or aerosol. For assessing the potential of an inhalant to cause URT irritation, the mouse body plethysmographic technique (Alarie, 1966, 1981a,b) has proven to be extremely useful. This technique operates on the principle that respiratory irritants stimulate the sensory nerve endings located at the surface of the respiratory tract from the nose to the alveolar region. The nerve endings in turn stimulate a variety of reflex responses (Alarie, 1973; Widdicombe, 1974) that result in characteristic changes in inspiratory and expiratory patterns and, most prominently, depression of respiratory rate. Both the potency of irritation and the concentration of the irritant are positively related to the magnitude of respiratory rate depression. The concentration response can be quantitatively expressed in terms of the $RD_{50}$, defined as the concentration (in logarithmic scale) of the drug in the air that causes a 50% decrease in respiratory rate. The criteria for positive URT irritation in intact mice exposed to the drug atmosphere are depression in breathing frequency and a qualitative alteration of the expiratory patterns. Numerous experimental results have shown that the responses of mice correlated almost perfectly with those of humans (Alarie et al., 1980; Alarie and Luo, 1986). Thus, this technique is useful for predicting the irritancy of airborne medicinal compounds in humans. From the drug-formulating point of view, an inhalant drug with URT-irritating properties indicates the need for an alternate route of administration. From the industrial hygiene point of view, the recognition of the irritant properties is very important. If a chemical gas, vapor, or aerosol irritates, it has a "warning property." With an adequate warning property a worker will avoid inhaling damaging amounts of the airborne toxicant; without

a warning property a worker may unknowingly inhale an injurious amount of the toxicant.

Inhalation of a cardiovascular drug, such as an aerosol of propranolol (a beta-adrenergic receptor agonist), may affect the respiratory cycle of a subject. For evaluating the cardiopulmonary effects of an inhalant, the plethysmograph technique using a mouse or a guinea pig model is useful. The criteria for a positive response in intact mice or guinea pigs are changes in the duration of inspiration and expiration and the interval between breaths (Schaper et al., 1989).

Pulmonary sensitization may occur from inhalation of drug vapors such as enflurane (Schwettmann and Casterline, 1976) and antibiotics such as spiramycin (Davies and Pepys, 1975) and tetracycline (Menon and Das, 1977). To detect pulmonary sensitization from inhalation of drug and chemical aerosols, the body plethysmographic technique using a guinea pig model has been shown to be useful (Patterson and Kelly, 1974; Paterson, 1977; Karol, 1988; Karol and Thorne, 1988; Karol et al., 1989; Thorne and Karol, 1989). The criteria for positive pulmonary sensitization in intact guinea pigs are changes in breathing frequency and their extent and the time of onset of an airway constrictive response after induction and after a challenge dose of the test drug (Karol et al., 1989).

The mucociliary transport system of the airways can be impaired by respiratory irritants, local analgenesics and anesthetics, and parasympathetic stimulants (Pavia, 1984). Any one of the above agents will retard the beating frequency of the cilia and the secretion of the serous fluid of the mucous membranes. As a result, the propulsion of the inhaled particles, bacteria, or endogenous debris toward the oral pharynx for expectoration or swallowing will be retarded. Conversely, inhalation of adrenergic agonists increases the activity of the mucociliary transport system and facilitates the elimination of noxious material from the pulmonary system. Laboratory evaluation of the adverse drug effects on mucociliary transport in animal models can be achieved by measuring the velocity of the linear flow of mucus in the trachea of surgically prepared animals (Rylander, 1966; Oyarzun and Clements, 1977, 1978). Clinically, the transportation of markers placed on the tracheal epithelium of normal human subjects can also be observed using a fiber-optic bronchoscopic technique (Pavia et al., 1980; Mussatto et al., 1988). The criteria of a positive response are changes in the transport time over a given distance of markers placed on the mucus or changes in the rate of mucous secretion (Davis et al., 1976; Johnson et al., 1983, 1987; Webber and Widdicombe, 1987). More comprehensive discussion on mucociliary clearance can be found in several reviews (Last, 1982; Pavia, 1984).

Cytological studies on the bronchial alveolar lavage fluid (BALF) permit the evaluation of the effects of an inhaled drug on the epithelial lining of the respiratory tract. This fluid can be obtained form intact animals or from excised lungs (Henderson, 1984, 1988, 1989). Quantitative analyses of fluid constituents such as neutrophils, antibody-forming lymphocytes, and antigen-specific IgG provide information on the cellular and biochemical responses of the

lungs to the inhaled agent (Henderson, 1984; Henderson et al., 1985, 1987). For example, BALF parameters were found to be unperturbed by the inhalation of halothane (Henderson and Lowrey, 1983). The criteria of a positive response are increase in protein content, increase in number of neutrophils and macrophages for inflammation, increase in number of lymphocytes and alteration of lymphocyte profiles for immune response, increase in cytoplasmic enzymes (lactate dehydrogenease) for cell lysis (Henderson, 1989), and presence of antigen-specific antibodies for specific immune responses (Bice, 1985).

Morphological examination of the cellular structure of the pulmonary system is the foundation of most inhalation toxicity studies. Inhalation of airborne drug vapors or aerosols at harmful concentrations results mainly in local histopathological changes in the epithelial cells of the airways, of which there are two types: nonciliated and ciliated cells. The nonciliated cells are the Clara cells, which contain secretory granules and smooth endoplasmic reticulum (SER); the ciliated cells are the brush cells, which have secretory granules and stubby microvilli and numerous cytoplasmic fibers on their free surfaces but lack SER. If the concentration gradient of the drug in the lung is high enough to reach the alveoli, the type I alveoli cells will also be affected (Evans, 1982). Drugs that affect the lungs via the bloodstream, such as bleomycin (Aso et al., 1976), cause changes to the endothelial cells of the vascular system that result in diffuse damage to the alveoli. The criteria of cellular damage are loss of cilia, swelling, and necrosis and sloughing of cell debris into the airway lumina. Tissues recovering from injuries are characterized by increases in the number of dividing progenitor cells followed by increases in intermediate cells that eventually differentiate into normal surface epithelium.

Pulmonary drug disposition studies are essential in the research and development of new inhalant drugs. Inhaled drugs are usually absorbed and metabolized to some extent in the lungs because the lungs, like the liver, contain active enzyme systems. A drug may be metabolized to an inactive compound for excretion or to a highly reactive toxic metabolite(s) that causes pulmonary damage. In most pulmonary disposition studies, a gas or vapor is delivered via whole-body exposure (Paustenbach et al., 1983) or head-only exposure (Hafner et al., 1975). For aerosols, over 90% of a dose administered by mouth breathing is deposited in the oropharynx and swallowed. Consequently, the disposition pattern reflects that of ingestion in combination with a small contribution from pulmonary metabolism. For determining the disposition of inhaled drugs by the pulmonary system alone, a dosimetric endotracheal nebulization technique (Leong et al., 1988) is useful. In this technique, microliter quantities of a radiolabeled drug solution can be nebulized within the trachea using a miniature air–liquid nebulizing nozzle. Alternatively, a small volume of liquid can be dispersed endotrachially using a microsyringe. In either technique, an accurate dose of a labeled drug solution is delivered entirely into the respiratory tract and lungs. Subsequent radioassay of the excreta thus reflects only the pulmo-

nary disposition of the drug without complication from aerosols deposited in the oropharyngeal regions if the drug had been delivered by mouth inhalation. For example, in a study of the antiasthmatic drug lodoximide tromethamine, the urinary metabolites produced by beagle dogs after receiving a dose of the radiolabeled drug via endotracheal nebulization showed a high percentage of the intact drug. However, metabolites produced after oral administration were mainly nonactive conjugates. The differences were due to the drug's escape from first-pass metabolism in the liver when it was administered through the pulmonary system. The results thus indicated that the drug had to be administered by inhalation to be effective. This crucial information was extremely important in the selection of the most effective route of administration and formulation of this antiasthmatic drug (Leong et al., 1988).

Cardiotoxicity of inhalant drugs should also be evaluated. For example, adverse cardiac effects may be induced by inhaling vapors of fluorocarbons, which are used extensively as propellants in drug aerosols. Inhalation of vapors of anesthetics also has been shown to cause depression of the heart rate and alteration of the rhythm and blood pressure (Merin, 1981; Leong and Rop, 1989). More important, inhalation of antiasthmatic aerosols of beta-receptor agonists delivered in a fluorocarbon propellant has been shown to cause marked tachycardia, electrocardiogram (ECG) changes, and sensitization of the heart to arrhythmia (Aviado, 1981; Balazs, 1981). Chronic inhalation of drug aerosols can also result in cardiomyopathy (Balazs, 1981). For detection of cardiotoxicity, standard methods of monitoring arterial pressures, heart rate, and ECGs of animals during inhalation of a drug or at frequent intervals during a prolonged treatment period should be useful in safety assessments of inhalant drugs.

Since the inhalation route is just a method for administering drugs, other nonpulmonary effects, such as behavioral effects (Ts'o et al., 1975) and renal and liver toxicity, should also be evaluated. In addition, attention should also be given to drugs that are not administered via the inhalation route but that accumulate in the lungs where they cause pulmonary damage (Wilson, 1982; Hollinger, 1985). Such inhaled organics can have a wide range of target organ effects—even on the eyes (Leong et al., 1987).

## 1.7 Parameters of Toxicity Evaluation

Over 400 years ago Paracelsus stated: "All substances are poison. The right dose differentiates a poison and a remedy." Thus, in safety assessments of inhaled drugs, the "dose," or magnitude of inhalation exposure, in relation to the physiological, biochemical, cytological, or morphological response(s) must be determined. Toxicity information is essential to establishing guidelines to prevent the health hazards of acute or chronic overdosage during therapy or of unintentional exposure to the bulk drugs and their formulated products during manufacturing and industrial handling.

***Inhaled "Dose"*** Most drugs are designed for oral or parenteral administration in which the dose is calculated in terms of drug weight in milligrams divided by the body weight in kilograms:

$$\text{Dose} = \frac{\text{drug weight (mg)}}{\text{body weight (kg)}} = \frac{\text{mg}}{\text{kg}}$$

For inhalant drugs, the inhaled dose has been expressed in many mathematical models (Dahl, 1990). However, the practical approach is based on exposure concentration and duration rather than on theoretic concepts. Thus, an inhaled dose is expressed in terms of the exposure concentration ($C$) in milligrams per liter or milligrams per cubic meter or parts per million (ppm) of air, the duration of exposure ($t$) in minutes, the ventilatory parameters including the respiratory rate ($R$) in number of breaths per minute and the tidal volume (Tv) in liters per breath, and a retention factor $\alpha$ which is related to the reactivity and the solubility of the drug. The product of these parameters divided by the body weight in kilograms gives the dose:

$$\text{Dose} = \frac{CtR\text{Tv}\ \alpha}{\text{body weight}} = \frac{\text{mg}}{\text{kg}}$$

In critical evaluation of the effect of a gas, vapor, or aerosol inhaled into the respiratory tract of an animal, the dosimetric method has been recommended (Oberst, 1961). However, due to the complexity of measuring the various parameters simultaneously, only a few studies on gaseous drugs or chemicals have employed the dosimetric method (Weston and Karel, 1946; Leong and MacFarland, 1965; Landy et al., 1983; Stott and McKenna, 1984; Dallas et al., 1986, 1989). For studies on liquid or powdery aerosols, modified techniques such as intratracheal instillation (Brain et al., 1976) or endotracheal nebulization (Leong et al., 1988) were used to deliver an exact dose of the test material into the lower respiratory tract (LRT) while bypassing the URT and ignoring the ventilatory parameters.

In routine inhalation studies, it is generally accepted that the respiratory parameters are relatively constant when the animals are similar in age, sex, and body weight. This leaves only $C$ and $t$ to be the major variables for dose consideration:

$$\text{Dose} = Ct = \text{mg} \cdot \text{min/L}$$

The product $Ct$ is not a true dose because its unit is milligrams per minute per liter rather than milligrams per kilogram. Nevertheless, $Ct$ can be manipulated as though it were a dose, an approximated dose (MacFarland, 1976).

**TABLE 1 Respiratory Parameters for Common Experimental Species and Humans**

| Species | Body Weight (kg) | Lung Volume (mL) | Minute Volume ($mL\,min^{-1}$) | Alveolar Surface Area ($m^2$) | Lung Volume % Surface Area | Minute Volume % Lung Volume | Minute Volume % Surface Area |
|---|---|---|---|---|---|---|---|
| Mouse | 0.023 | 0.74 | 24 | 0.068 | 10.9 | 32.4 | 353 |
| Rat | 0.14 | 6.3 | 84 | 0.39 | 16.2 | 13.3 | 215 |
| Monkey | 3.7 | 184 | 694 | 13 | 14.2 | 3.77 | 53 |
| Dog | 22.8 | 1501 | 2923 | 90 | 16.7 | 1.95 | 33 |
| Human | 75 | 7000 | 6000 | 82 | 85.4 | 0.86 | 73 |

*Source*: Altman and Dittmar, 1974.

The respiratory parameters of an animal will dictate the volume of air inhaled and hence the quantity of test material entering the respiratory system. Commonly used parameters for a number of experimental species and humans are given in Table 1 to illustrate this point and include the alveolar surface area because this represents the target tissue for most inhaled materials. It can be seen that, by taking the ratios of these parameters and comparing the two extremes, that is, the mouse and human, (1) a mouse inhales approximately 30 times its lung volume in 1 min whereas a human at rest inhales approximately the same volume as that of his or her lung. This can increase with heavy work up to the same ratio as the mouse but is not sustained for long periods. This means that the dose per unit lung volume is up to 30 times higher in the mouse than in humans at the same inhaled atmospheric concentration. (2) The minute volume of the mouse is in contact with five times less alveolar surface area than humans and hence the dose per unit area is up to five times greater in the mouse. (3) The lung volume in comparison with the alveolar surface area in experimental animals is less than in humans, meaning that the extent of contact of inhaled gases with the alveolar surface is greater in experimental animals.

While it is possible, and common, to refer to standard respiratory parameters for different species in order to calculate inhaled dose and deposited dose with time, it is usually the case that inhaled materials influence the breathing patterns of test animals. The most common examples of this are irritant vapors, which can reduce the respiratory rate by up to 80%. This phenomenon results from a reflexive pause during the breathing cycle due to stimulation by the inhaled material of the trigeminal nerve endings situated in the nasal passages. The duration of the pause and hence the reduction in the respiratory rate are concentration related, permitting concentration–response relationships to be plotted. This has been investigated extensively by Alarie (1981a) and forms the basis of a test screen for comparing quantitatively the irritancy of different materials and has found application in assessing appropriate exposure limits for human exposure when respiratory irritancy is the predominant cause for concern.

While irritancy resulting from the above reflex reaction is one cause of altered respiratory parameters during exposure, there are many others. These include other types of reflex response, such as bronchoconstriction, the narcotic effects of many solvents, the development of toxic signs as exposure progresses, or simply a voluntary reduction in respiratory rate by the test animal due to the unpleasant nature of the inhaled atmosphere. The extent to which these affect breathing patterns and hence inhaled dose can only be assessed by actual measurement.

By simultaneous monitoring of tidal volume and respiratory rate, or minute volume, and the concentration of an inhaled vapor in the bloodstream and the vapor in the exposure atmosphere, pharmacokinetic studies on the $Ct$ relationship have shown that the effective dose was nearly proportional to the exposure concentration for vapors such as 1,1,1-trichloroethane (Dallas et al., 1989), which has a saturable metabolism. However, it was also found that the steady-state plasma concentrations were disproportionally greater at higher exposure concentrations.

Acknowledging the possible existence of deviations, this simplified approach of using $C$ and $t$ for dose determination provides the basis for dose–response assessments in practically all inhalation toxicological studies.

***Dose–Response Relationship*** The first principle of dose–response determination in inhalation toxicology is based on Haber's rule, which states that responses to an inhaled toxicant will be the same under conditions where $C$ varies in complementary manner to $t$ (Haber, 1924). For example, if $Ct$ elicits a specific magnitude of the same response, that is, $Ct = K$, where $K$ is a constant for the stated magnitude of response.

This rule holds reasonably well when $C$ or $t$ varies within a narrow range for acute exposure to a gaseous compound (Rinehart and Hatch, 1964) and for chronic exposure to an inert particle (Henderson et al., 1991). Excursion of $C$ or $t$ beyond these limits will cause the assumption $Ct = K$ to be incorrect (Adams et al., 1950, 1952; Sidorenko and Pinigin, 1976; Andersen et al., 1979; Uemitsu et al., 1985). For example, an animal may be exposed to 1000 ppm of diethyl ether for 420 min or 1400 ppm for 300 min without incurring any anesthesia. However, exposure to 420,000 ppm for 1 min will surely cause anesthesia or even death of the animal. Furthermore, toxicokinetic study of liver enzymes affected by inhalation of carbon tetrachloride (Uemitsu et al., 1985), which has a saturable metabolism in rats, showed that $Ct = K$ does not correctly reflect the "toxicity value" of this compound. Therefore, the limitations of Haber's rule must be recognized when it is used in interpolation or extrapolation of inhalation toxicity data.

***Exposure Concentration versus Response*** In certain medical situations (e.g., a patient's variable exposure duration to a surgical concentration of an inhalant anesthetic or the repeated exposures of surgeons and nurses to subanesthetic concentrations of an anesthetic in the operating theater), it is neces-

sary to know the duration of safe exposure to a drug. Duration safety can be assessed by determining a drug's median effective time ($ET_{50}$) or median lethal time ($LT_{50}$). These statistically derived quantities represent the duration of exposure required to affect or kill 50% of a group of animals exposed to a specified concentration of an airborne drug or chemical in the atmosphere.

The graph in Figure 2 is the probit plot of percent response to exposure concentrations. The log/probit transformation( Litchfield and Wilcoxon, 1949) is used to linearize the dose response data for graphing data. Figure 3,

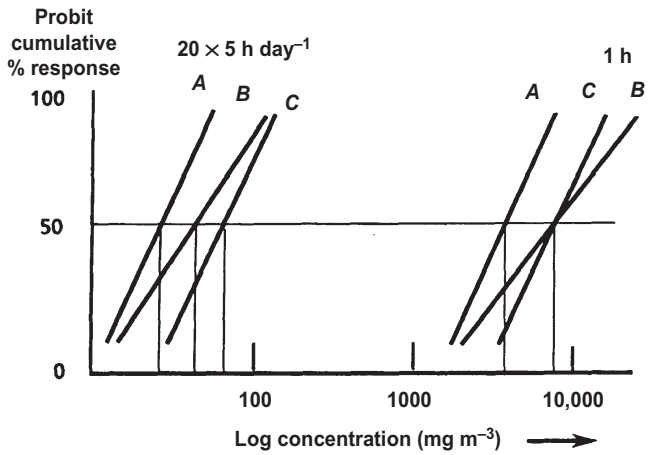


***Figure 2*** *Dose–response plot in terms of probit of cumulative percentage response to logarithm of exposure concentrations.*

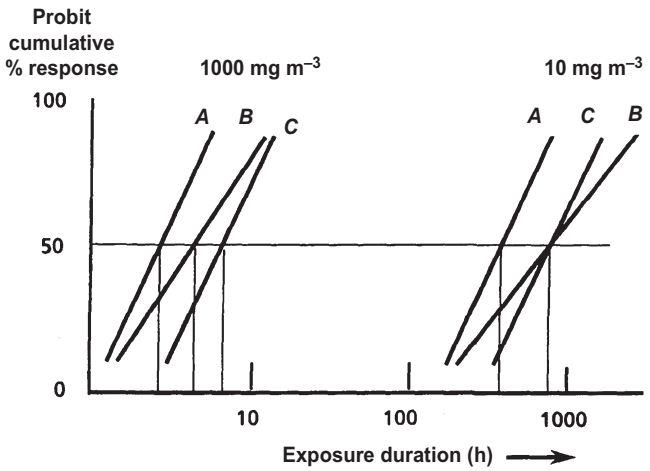


***Figure 3*** *Dose–response plot in terms of probit of cumulative percentage response to logarithm of exposure duration.*

meanwhile, is the probit plot of cumulative percentage response to logarithm of exposure duration. It shows the 1000 mg m$^{-3}$ for 10 h or to 10 mg m$^{-3}$ for 1000 h, each with a *Ct* (an approximated dose) of ~ 10,000 h mg m$^{-3}$. Similar to concentration–response graphs, the slopes indicate the differences in the mechanism of action and the margins of safe exposure of the three drugs. The ratio of the $ET_{50}$ or $LT_{50}$ of two drugs indicates their relative toxicity, and the ratio of $ET_{50}$ over $LT_{50}$ of the same drug is the therapeutic ratio.

***Product of Concentration and Duration (Ct) versus Responses*** To evaluate inhalation toxicity in situations where workers are exposed to various concentrations and durations of a drug vapor, aerosol, or powder in the work environment during manufacturing or packaging, a more comprehensive determination of $E(Ct)_{50}$ or $L(Ct)_{50}$ values are used. The $E(Ct)_{50}$ or $L(Ct)_{50}$ values are statistically derived values that represent the magnitude of exposure, expressed as a function of the product of *C* and *t*, that is expected to affect or kill <50% and >50% of the animals. The other curve represents exposures that kill 50% or >50% of each group of animals (Irish and Adams, 1940).

The graph in Figure 4 illustrates inhalation exposures to a drug using various combinations of *C* and *t* that kill 50% of the animals. For example, a 50% mortality occurs when a group of animals is exposed to drug A at a concentration of 1000 mg m$^{-3}$ for a duration of approximately 2 h or at a

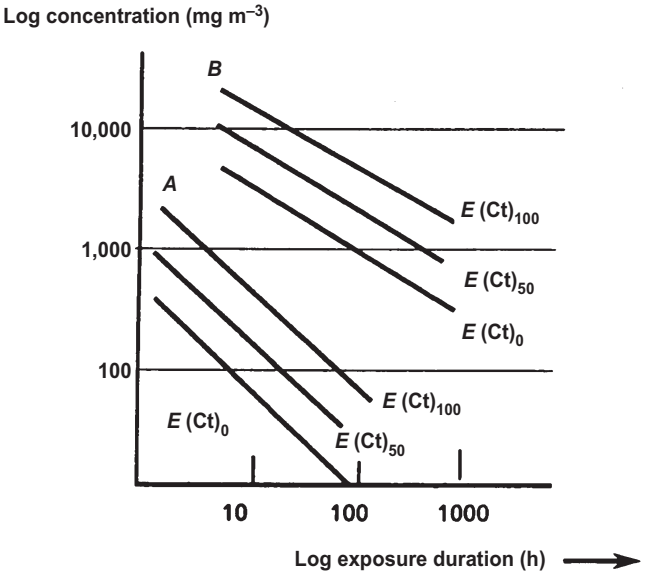


***Figure 4*** *Dose–response plot in terms logarithms of exposure concentration and durations.*

concentration of 100 mg m$^{-3}$ for a duration of approximately 20 h. Furthermore, the graph also illustrates that the inhalation toxicity of drug A is more than one order of magnitude higher than that of drug B. For example, an exposure to drug A at the concentration of 100 mg m$^{-3}$ for 100 h kills 100% of the animals whereas an exposure to drug B at the concentration of 1000 mg m$^{-3}$ for 100 h does not kill any animals.

***Units for Exposure Concentration*** For gases and vapors, exposure concentrations are traditionally expressed in parts per million (ppm). The calculation for the ppm of a gas or vapor in an air sample is based on Avogadro's law, which states that equal volumes contain equal numbers of molecules under the same temperature and pressure. In other words, under standard temperature and pressure (STP), one gram-molecular weight (mole) of any gas under a pressure of one atmosphere (equivalent to the height of 760 mm mercury) and a temperature of 273 K has the same number of molecules and occupies the same volume of 22.4 L. However, under ambient conditions, the volume of 22.4 L has to be corrected to a larger volume based on Charles's law, which states that at constant pressure the volume of gas varies directly with the absolute temperature. Thus, at a room temperature of 25 °C, 1 mol of a gas occupies a volume of 24.5 L:

$$22.4\text{L} \times \frac{298\text{K}}{273\text{K}} = 24.5\text{L}$$

Further correction of volume for an atmospheric pressure deviation from one atmosphere may be done by applying Boyle's law, which states that the volume of a gas without change of temperature varies inversely with the pressure applied to it:

$$24.5\text{L} \times \frac{758\text{mm Hg}}{760\text{mm Hg}} = 24.4\text{L}$$

In practice, atmospheric pressure in most animal experimental environments usually varies only a few millimeters of mercury, so little or no correction is required.

Using the aforementioned principles, the volume of a vapor generated from a given weight of a liquid can be calculated. For example, 1 mol of water weighs 18 g, while 1 mol of ethanol weighs 46 g. When 1 mol of each liquid is totally vaporized, each will occupy the same volume of 24.5 L at room temperature (25 °C) and pressure (760 mm Hg). In an inhalation experiment, if the volume of test liquid and the rate of airflow being mixed in the animal exposure chamber are known, the vapor concentration in the chamber atmosphere can be calculated in parts per million or milligrams per liter. A conversion table published by the U.S. Bureau of Mines enables quick conversion between parts

per million and milligrams per liter for compounds with molecular weights up to 300 g (Fieldner et al., 1921; Patty, 1958).

For aerosols of nonvolatile liquid and powdery compounds, the concentration of the mist or dust atmosphere must be expressed in terms of milligrams per liter or milligrams per cubic meter of air. With advances in biotechnology, Many pharmacological testing techniques are based on specific receptor bindings, in which the ratio of the number of molecules to those of the receptors are considered, in which case the exposure concentration may be more appropriately expressed in micromoles per unit volume of air.

## 1.8 Inhalation Exposure Techniques

Many inhalation exposure techniques, such as the whole-body, nose-only, mouth-only, or head-only technique (Drew and Laskin, 1973; MacFarland, 1976; Leong et al., 1981; Smith et al., 1981; Phalen, 1984), the intranasal exposure technique (Elliot and DeYoung, 1970), the endotracheal nebulization technique (Leong et al., 1985, 1988; Schreck et al., 1986), and the body plethysmographic techniques (Alarie, 1966; Thorne and Karol, 1989), have been developed for inhalation toxicity studies. Table 2 provides a summary of the advantages and disadvantages of each of the major inhalation exposure methodologies.

The main criteria for the design and operation of any dynamic (as opposed to static) inhalation exposure system are as follows:

- The concentration of the test atmosphere must be reasonably uniform throughout the chamber and should increase and decrease at a rate close to theoretical at the start or end of the exposure. Silver (1946) showed that the time taken for a chamber to reach a point of equilibrium was proportional to the flow rate of atmosphere passing through the chamber and the chamber volume. From this, the concentration–time relationship during the "run-up" and "run-down" phase could be expressed by the equation

$$t_x = k\frac{V}{F}$$

where $t_x$ = time required to reach $x$ percent of the equilibrium concentration, $k$ = a constant determined by the value of $x$, $V$ = chamber volume, and $F$ = chamber flow rate. The $t_{99}$ value is frequently quoted for exposure chambers, representing the time required to reach 99% of the equilibrium concentration and providing an estimate of chamber efficiency. Thus, at maximum efficiency, the theoretical value of $k$ at $t_{99}$ is 4.605, and the closer to this that the results of evaluation of actual chamber performance fall, the greater the efficiency and the better the design of the chamber.

**TABLE 2 Advantages, Disadvantages, and Considerations Associated with Patterns of Inhalation Exposure**

| Mode of Exposure | Advantages | Disadvantages | Design Considerations |
|---|---|---|---|
| Whole body | Variety and number of animals; chronic studies possible; minimum restraint; large historical database; controllable environment; minimum stress; minimum labor | Messy; multiple routes of exposure: skin, eyes, oral; variability of "dose"; cannot pulse exposure easily; poor contact between animals and investigators; capital intensive; inefficient compound usage; difficult to monitor animals during exposure | Cleaning effluent air; inert materials; losses of test material; even distribution in space; sampling; animal care; observation; noise, vibration, humidity; air temperature; safe exhaust; loading; reliability |
| Head only | Good for repeated exposure; limited routes of entry into animal; more efficient dose delivery | Stress to animal; losses can be large; seal around neck; labor in loading/unloading | Even distribution; pressure fluctuations; sampling and losses; air temperature, humidity; animal comfort; animal restraint |
| Nose/mouth only | Exposure limited to mouth and respiratory tract; uses less material (efficient); containment of material; can pulse exposure | Stress to animal; seal about face; effort to expose large number of animals | Pressure fluctuations; body temperature; sampling; airlocking; animals' comfort; losses in plumbing/ masks |
| Lung only (tracheal administration) | Precision of dose; one route of exposure; uses less material (efficient); can pulse exposure | Technically difficult; anesthesia or tracheostomy; limited to small numbers; bypasses nose; artifacts in deposition and response; technically more difficult | Air humidity/ temperature; stress to animal; physiological support |
| Partial lung | Precision of total dose; localization of dose; can achieve very high local doses; unexposed control tissue from same animal | Anesthesia; placement of dose; difficulty in interpretation of results; technically difficult; possible redistribution of material within lung | Stress to animal; physiological support |

*Source*: Gad and Chengelis, 1998.

- Flow rates must be controlled in such a way that they are not excessive, which might cause streaming effects within the chamber, but must be adequate to maintain normal oxygen levels, temperature, and humidity in relation to the number of animals being exposed. A minimum of 10 air changes per hour is frequently advocated and is appropriate in most cases. However, the chamber design and housing density also need to be taken into account and some designs, such as that of Doe and Tinston (1981), function effectively at lower air change rates.
- The chamber or exposure manifold materials should not affect the chemical or physical nature of the test atmosphere.

The whole-body exposure technique is useful for acute and chronic toxicity studies of gases and vapors. For acute whole-body exposure, a few animals are exposed for 1–4 h to a gas, vapor, or aerosol of a drug or chemical in a simple glass jar. The gaseous drug is metered with a precision flow meter into the stream of filtered room air being drawn through the glass jar or chamber. For vapor generation from a volatile liquid, a stream of clean air is bubbled at a constant rate onto the walls of a temperature-regulated flask, which vaporizes the liquid droplets rapidly and continuously. In either method, the vapor emerging from the vaporizer is directed into the filtered air stream being drawn through the glass jar or chamber. For the generation of drug aerosols from liquids or powders, various types of atomizers or nebulizers and dust generators are available (Drew and Laskin, 1973; Drew and Lippmann, 1978; Leong et al., 1981; Phalen, 1984). For more critical and precision studies, an adequate number of animals per group is calculated by an appropriate statistical method (Gad, 2007) and the exposure is carried out in an elaborate dynamic airflow chamber with precision control of the chamber airflow, temperature, and humidity.

Regardless of the exposure apparatus used, the most important aspect of an exposure study is the generation of a constant concentration of the airborne drug vapor or aerosol in the chamber atmosphere has to be sampled (Drew and Lippmann, 1978) and analyzed using an appropriate analytical instrument, such as an infrared spectrophotometer for halogenated propellants or a gas chromatograph for other gases and vapors. The concentration of the drug as detected by the analyzer is the "analytical concentration." For characterizing the aerosol atmosphere, particle sizing, in addition to concentration analysis, is essential. Because the breathing patterns of the experimental animals cannot be regulated, it is extremely important to generate aerosols of the appropriate size for bioavailability.

For critical laboratory studies on inhaled drugs, a monodisperse aerosol of a specified range of MMAD should be used to increase the probability that the aerosol reaches the specified target area of the lungs. The Dautrebande aerosol generators (Dautrebande, 1962c) and the DeVilbiss nebulizer (Drew and Lippmann, 1978) are the classic single-reservoir generators for short-duration inhalation studies. For long-duration inhalation studies, the

multiple-reservoir nebulizer (Miller et al., 1981) or the continuous syringe metering and elutriating atomizer (Leong et al., 1981) are frequently used. The nebulizers generate a polydisperse droplet aerosol either by the shearing force of a jet of air over a fine stream of liquid or by ultrasonic disintegration of the surface liquid in a reservoir (Drew and Lippmann, 1978). The aerosols emerging from a jet nebulizer generally have MMADs ranging between 1.2 and 6.9 μm with GSDs of 1.7–2.2, and aerosols from an ultrasonic nebulizer have MMADs ranging between 3.7 and 18.5 μm with GSDs of 1.4–2.0 (Mercer, 1981).

For testing therapeutic formulations, the liquid aerosols are usually generated by the pressurized metered-dose inhaler (Newman, 1984; Newton, 2000; Gad and Chengelis, 1998). The pressurized metered-dose inhaler (MDI) generates a bolus of aerosols by atomizing a well-defined quantity of a drug that is solubilized in a fluorocarbon propellant. The aerosols thus consist of the drug particles with a coating of the propellant. As the aerosols emerge from the orifice, the mean particle size may be as large as 30 μm (Moren, 1981). After traveling through a tubular or cone-shaped spacer, the propellant may evaporate, reducing the MMADs to a range of 2.8–5.5 mm with GSDs of 1.5–2.2 (Hiller et al., 1978; Sackner et al., 1981; Newman, 1984) and making the aerosols more stable for inhalation studies. In a prolonged animal exposure study, multiple metered-dose inhalers have to be actuated sequentially with an electromechanical gadget (Ulrich et al., 1984) to maintain a slightly pulsatile but relatively consistent chamber concentration.

For generating an aerosol from dry powders, various dust generators, such as the Wright dust feed, air elutriator or fluidized-bed dust generator, and air impact pulverizer, have been developed for acute and chronic animal inhalation studies and described in many articles (Hinds, 1980; Leong et al., 1981; Phalen, 1984; Gad and Chengelis, 1998; Valentine and Kennedy, 2001; Hext, 2000). For generating powdery therapeutic agents, a metered-dose dry-powder inhaler, spinhaler, or rotahaler is used (Newman, 1984). The particle size of the drug powder is micronized to a specific size range during manufacture and the spinhaler or the rotahaler only disperses the powders.

More recently, another approach for administering dry powders to both humans and test animals has arisen. Dry powders, while less frequently used in nasal drug delivery, are becoming more popular. Powders can be administered from several devices, the most common being the insufflator. Many insufflators work with predosed powder in gelatin capsules. To improve patient compliance, a multidose powder inhaler has been developed which has been used to deliver budesonide. These devices can also be used for administration to test animals delivery, in terms of both amount and aerodynamic size of the particles. While early dry-powder inhalers such as the Rotohaler used individual capsules of micronized drug which were difficult to handle, modern devices use blister packs (e.g., Diskus) or reservoirs (e.g., Turbuhaler). The dry-powder inhalers rely on inspiration to withdraw drug from the inhaler to the lung, and hence the effect of inhalation flow rate through various devices

has been extensively studied. The major problem to be overcome with these devices is to ensure that the finely micronized drug is thoroughly dispersed in the airstream. It has been recommended that patients inhale as rapidly as possible from these devices in order to provide the maximum force to disperse the powder. The quantity of drug and deposition pattern vary enormously depending on the device, for example, the Turbuhaler produces significantly greater lung delivery of salbutamol than the Diskus. Vidgren and co-workers (1987) demonstrated by gamma scintigraphy that a typical dry-powder formulation of sodium cromoglycate suffers losses of 44% in the mouth and 40% in the actuator nozzle itself.

It must also be emphasized that the major mass of a heterodispersed aerosol may be contained in a few relatively large particles since the mass of a particle is proportional to the cube of its diameter. Therefore, the particle size distribution and the concentration of the drug particles in the exposure atmosphere should be sampled using a cascade impactor or membrane filter sampling technique, monitored using an optical or laser particle size analyzer, and analyzed using optical or electron microscopy techniques.

In summary, many techniques have been developed for generating gas, vapor, and aerosol atmospheres for inhalation toxicology studies. By proper regulation of the operating conditions of the nebulizers and the formulation of metered-dose inhalers, together with the use of spacer or reservoir attachments to MDIs, more particles within the respirable range can be generated for inhalation. An accurately controlled exposure concentration is essential to an accurate determination of the dose–response relationship in a safety assessment of an inhalant drug.

Finally, comparisons of various techniques for animal exposures indicate that the whole-body exposure technique is the most suitable for safety assessment of gases and vapors and permit simultaneous exposure of a large number of animals to the same concentration of a drug; however, this technique is not suitable for aerosol and powder exposures because the exposure condition represents the resultant effects from inhalation, ingestion, and dermal absorption of the drug (Phalen, 1984; Gad and Chengelis, 1998).

## 1.9 Utility of Toxicity Data

Regardless of the type of test and the parameters to be monitored, the ultimate goal is to interpolate or extrapolate from the dose–response data to find a no-observable-adverse-effect level (NOAEL) or a no-observable-effect level (NOEL). By applying a safety factor of 1–10 to the NOAEL, a safe single-exposure dose for a phase I clinical trial may be obtained. By applying a more stringent safety factor, a multiple-exposure dose for a clinical trial may also be obtained. After the drug candidate has successfully passed all the drug safety evaluations and entered in the production stage, more toxicity tests may be needed for the establishment of a threshold limit value–time-weighted

average (TLV–TWA). A TLV–TWA is defined as "the time weighted average concentration for a normal 8-hour workday and a 40-hour workweek, to which nearly all workers may be repeatedly exposed, day after day, without adverse effect" [American Conference of Governmental Industrial Hygienists (ACGIH), 1991]. Using TLVs as guides, long-term safe occupational exposures during production and industrial handling of a drug may be achieved. Appropriate safety assessments of pharmaceutical chemicals and drugs will ensure the creation and production of a safe drug for the benefit of humans and animals. Further, inhalation toxicity data are needed for compliance with many regulatory requirements of the U.S. Food and Drug Administration (FDA), the Occupational Health and Safety Administration, and the Environmental Protection Agency (Gad and Chengelis, 1998).

More comprehensive descriptions and discussions on inhalation toxicology and technology may be found in several recent monographs, reviews, and textbooks (Willeke, 1980; Leong et al., 1981; Witschi and Nettesheim, 1982; Clarke and Pavia, 1984; Phalen, 1984; Witschi and Brain, 1985; Barrow, 1986; McFadden, 1986; Menzel and Amdur, 1986; Salem, 1986; Gardner et al., 1988; Gad and Chengelis, 1998; Valentine and Kennedy, 2001; McClellan and Henderson, 1989; Hext, 2000).

## 2 THERAPEUTIC DRUG DELIVERY BY DERMAL ROUTE

The dermal route has both some significant attractions and some significant drawbacks for use in drug administration. The ability to utilize the former and avoid (at least the relevance of) the latter is the key to successful development of new drugs administered by this route (Ozhai et al., 2008).

On the advantage side, dermal administration:

- Avoids first-pass metabolism
- Avoids acidic environment of the stomach
- Can be designed to deliver controlled amounds of drug over a prolonged period outside of a clinical setting
- Can be designed to achieve desired local tissue exposure to the therapeutic area while avoiding (or minimizing) systemic exposure

On the disadvantage side, dermal administration historically:

- Cannot be used for systemic delivery of large molecules (usually with molecular weights in excess of 1000)
- Achieves a lower systemic bioavailability than oral routes
- Uses formulation components which may irritate or damage the application site (Zhai et al., 2008) and are perceived as "messy" or aesthetically undesirable

New approaches to formulation or delivery systems (see Touitou and Barry, 2007) have overcome (or minimized) a number of the disadvantages and continue to evolve. In addition to the traditional gels, lotions, and creams (for which new chemical permeation enhancers and vesicular carriers such as liposomes have and are being added), new technology such as iontophoresis, electrophoresis, and ultrasound is also evolving. Additionally, in some cases, the barrier layer in the skin (stratum corneum) can be bypassed (using such approaches as microneedles) or removed by tape-stripping laser, adhesion, and ballistic methods (such as used for mass vaccine administration in the military).

Bioavailability is typically defined as the rate and extent at which a drug reaches the general circulation from an administered dosage form. Dermatological drug products include preparations which are designed to exert a local effect in diseased skin following topical application on the skin surface. The objective is to maximize drug concentration at the site of action within the skin with, ideally, a minimal systemic uptake. Thus, systemic availability may not properly reflect local cutaneous bioavailability (as it does for transdermal products which are designed to deliver drug into the systemic circulation). Moreover, topical doses tend to be so small (typically 2–5 mg of product per square centimeter) that serum and/or urine concentrations are often undetectable using conventional assay techniques. Further complicating this is the lack of knowledge of the drug concentration needed at the skin target site (with the exception of antifungal and antibacterial agents whose target site is the SC surface). Topical bioavailability has been more properly defined as the temporal pattern of free drug, but this approach remains largely theoretical due to the difficulty of quantifying drug within the skin.

Available options include estimating a drug's permeability coefficient through human skin from the molecular weight and octanol–water partition coefficient. This information is not really sufficient to estimate topical bioavailability. The algorithms available are only able to approximate values for drug bioavailability.

Alternatives include collecting samples by tape stripping or biopsy and then measuring the actual drug (and metabolites) present in the tissue samples. There are numerous complications to this and approaches to solving them (see Herkenne et al., 2008).

There are no direct guidance documents focused on nonclinical safety assessment for topical route drugs or on what must be done before taking such a drug into clinical trials. Rather, the FDA regulatory expectations come from ICH/M4, the Center for Drug Evaluation and Research (CDER, 2008) reformulation document, and current practice. The results are summarized in Table 3.

The CDER (2008) document presents the following dermal route specific expectations:

- In the United States delayed hypersensitivity of any previously not evaluated topical drug or new formulation means performing the LLNA (local

**TABLE 3 Test Requirement Matrix for Topical Agents**

| Test Requirement | Species |
|---|---|
| *Initial Clinical Trial/IND Requirements* | |
| 1. Acute toxicity in rodents (IV[a]) | R/M |
| 2. Acute toxicity in nonrodents | D/S/P |
| 3. Seven-day DRF toxicity in rodents | R/M |
| 4. Seven-day DRF toxicity in nonrodents | D/S/P |
| 5. Genotoxicity: bacterial mutagenicity (Ames)[b] | In vitro |
| 6. Genotoxicity: in vitro clastogenicity (CHO chromosome aberration)[b] | In vitro |
| 7. Genotoxicity: in vivo (mouse or rat micronucleus)[b] | R/M |
| 8. Safety pharmacology: CV-Herg[a] | In vitro |
| 9. Safety pharmacology: CV in vivo[b] | D/P/S |
| 10. Safety pharmacology: FOB/Irwin[b] | R/M |
| 11. Safety pharmacology: respiratory—Rodent[b] | R |
| 12. Pivotal/repeat dose in rodents (14–28 days intended route[c]) | R/M |
| 13. Pivotal/repeat dose in nonrodents (14–28 days intended route[c]) | D/P/S |
| 14. CYP induction/inhibition[a] | In vitro |
| 15. Five species microsome metabolic panel[a] | In vitro |
| 16. Develop bioanalytical for three species (human/rodent/nonrodent) | NA |
| 17. Local irritation[c] (clinical formulation) | R |
| 18. Of dermal—sensitization | G |
| *To Support Continued Clinical Development* | |
| 19. Developmental toxicity (segment II)—rat and rabbit pilots and rat and rabbit studies | |
| 20. Immunotoxicity[c] (if immune modulatory claim or there are findings in 14/28-dog studies | |
| 21. Pivotal/repeat dose in rodents (3/6 months oral) | |
| 22. Pivotal/repeat dose in nonrodents (3/9–12 months oral) | |
| *To Support Marketing Approval* | |
| 23. Reproductive toxicity—segment I | |
| 24. Reproductive toxicity—segment III | |
| 25. Tumorgenicity/carcinogenicity—rat | |
| 26. Tumorgenicity/carcinogenicity—mouse | |

*Note*: Species: R = rat, M = mouse, D = dog, S = pig, P = primate, B = rabbit, G = guinea pig, TBD = to be determined. All studies described must be performed according to good labor practices.
[a]Recommended.
[b]May be required.
[c]Dermal/opthalmic/vaginal/rectal.

lymph node assay), guinea pig maximization test (GPMT), or Buehler. As the Buehler is not accepted in Europe, use of one of the other two (see Chapter Immunotoxicology in Drug Development on immunotoxicology for details on these and their limitations) should be used.

- Photoirritation should be evaluated if the new formulation absorbs ultraviolet or visible radiation (290–700 nm) and if the product is applied to

sun-exposed skin. If the new formulation is a patch, then photoirritation should be considered if the patch is permeable to light and is applied to sun-exposed skin (see details in local tissue tolerance). A screen using three T cells is acceptable in place of the traditional rabbit study.

- If the new formulation contains an active ingredient that has not been used by the dermal route, the repeat-dose local toxicity study mentioned earlier should be conducted in a nonrodent species (preferably the pig). This study should be of at least the same duration as clinical use (up to nine months) and include both local and systemic evaluation. For NCEs (new chemical entities), repeat-dose studies in both rodents and nonrodents (preferably the pig) are required. In one of these (typically the rodent), it should be ensured that sufficient systemic exposure of the NCE is achieved usually by conducting such a study by the oral or parenteral route.
- The skin dose from topically applied drug products can be orders of magnitude larger than the skin dose after systemic administration. Therefore, a dermal carcinogenicity study might be recommended for drugs with a chronic indication even if systemic carcinogenicity studies are available.
- The photocarcinogenic potential should be evaluated if the new formulation is used chronically on sun-exposed skin. Evaluation of photocarcinogenicity generally is not recommended for path products (see the guidance for industry on photosafety testing) (FDA, 2006).
- Nonclinical dermal studies generally should be conducted with untreated control, vehicle control, and complete formulation groups.

A key point is if significant systemic exposure is or is not achieved. If it is not, then genotoxicity and safety pharmacology testing is not required. Such testing is required if it cannot be shown that systemic absorption is minimal or not detectable.

For the repeat-dose systemic toxicity studies, added to the usual details of study design is the need to evaluate local tissue effects of application sites and to collect tissue and histopathologically evaluate such sites.

As a last point, it should be added that topical administration is not limited to the dermal route. It also includes the body surfaces (vaginal and rectal) for which evaluation of irritation and hypersensitivity is required and for which systemic toxicology repeat-dose studies are performed by administration to these sites (but there is no phototoxicity concern) and ocular for which it is expected that:

- The dermal irritation and delayed contact hypersensitivity potential of the new formulation should be evaluated (because dermal exposure will occur with such drugs).
- The local tissue (eye) irritation must be evaluated.

## REFERENCES

Adams, E. M., Spencer, H. C., Rowe, V. K., and Irish, D. D. (1950). Vapor toxicity of 1,1,1,-trichloroethane (methylchloroform) determined by experiments on laboratory animals. *Arch. Ind. Hyg. Occup. Med.* 1:225–236.

Adams, E. M., Spencer, H. C., Rowe, V. K., McCollister, D. D., and Irish, D. D. (1952). Vapor toxicity of carbon tetrachloride determined by experiments on laboratory animals. *Arch. Ind. Hyg. Occup. Med.* 6:50–66.

Agnew, J. E. (1984). Physical properties and mechanisms of deposition of aerosols. In S. W. Clarke and D. Pavia (Eds.), *Aerosols and the Lung, Clinical Aspects*. Butterworth, London, pp. 49–68.

Alarie, Y. (1966). Irritating properties of airborne material to the upper respiratory tract. *Arch. Environ. Health* 13:433–449.

Alarie, Y. (1973). Sensory irritation by airborne chemicals. *CRC Crit. Rev. Toxicol.* 2:299–363.

Alarie, Y. (1981a). Toxicological evaluation of airborne chemical irritants and allergens using respiratory reflex reactions. In B. K. J. Leong (Ed.), *Inhalation Toxicology and Technology*. Ann Arbor Science, Ann Arbor, MI, pp. 207–231.

Alarie, Y. (1981b). Bioassay for evaluating the potency of airborne sensory irritants and predicting acceptable levels of exposure in man. *Food Cosmet. Toxicol.* 19:623–626.

Alarie, Y., Kane, L., and Barrow, C. (1980). Sensory irritation: The use of an animal model to establish acceptable exposure to airborne chemical irritants. In A. L. Reeves (Ed.), *Toxicology: Principles and Practice*, Vol. 1. Wiley, New York.

Alarie, Y., and Luo, J. E. (1986). Sensory irritation by airborne chemicals: A basis to establish acceptable levels of exposure. In C. S. Barrow (Ed.), *Toxicology of the Nasal Passages*. Hemisphere, New York, pp. 91–100.

Altman, P. L., and Dittmer, D. S. (1974). *Biological Data Book*, Vol. III. Federation of American Societies for Experimental Biology, Bethesda, MD.

American Conference of Governmental Industrial Hygienists (ACGIH) (1991). *Documentation of the Threshold Limit Values and Biological Exposure Indices*, 6th ed. ACGIH, Cincinnati, OH.

Andersen, M. E., French, J. E., Gargas, M. L., Jones, R. A., and Jenkins, L. J., Jr. (1979). Saturable metabolism and the acute toxicity of 1,1-dichloroethylene. *Toxicol. Appl. Pharmacol.* 47:385–393.

Aso, Y., Yoneda, K., and Kikkawa, Y. (1976). Morphologic and biochemical study of pulmonary changes induced by bleomycin in mice. *Lab. Invest.* 35:558–568.

Aviado, D. M. (1981). Comparative cardiotoxicity of fluorocarbons. In T. Balazs (Ed.), *Cardiac Toxicology*, Vol. II. CRC Press, Boca Raton, FL, pp. 213–222.

Aviado, D. M., and Micozzi, M. S. (1981). Fluorine-containing organic compounds. In G. D. Clayton and F. E. Clayton (Eds.), *Patty's Industrial Hygiene and Toxicology*, Vol. 2B. Wiley, New York, pp. 3071–3115.

Balazs, T. (1981). Cardiotoxicity of adrenergic bronchodilator and vasodilating antihypertensive drugs. In T. Balazs (Ed.), *Cardiac Toxicology*, Vol. II. CRC Press, Boca Raton, FL, pp. 61–73.

Barrow, C. S. (1986). *Toxicology of the Nasal Passages*. Hemisphere, New York.

Bell, K. A. (1978). Local particle deposition in respiratory airway models. In D. T. Shaw (Ed.), *Recent Developments in Aerosol Science*. Wiley, New York, pp. 97–134.

Bice, D. E. (1985). Methods and approaches to assessing immunotoxicology of the lower respiratory tract. In J. H. Dean, M. I. Luster, A. E. Munson, and H. A. Amos (Eds.), *Immunotoxicology and Immunopharmacology*. Raven, New York, pp. 145–157.

Brain, J. D. (1971). The effects of increased particles on the number of aveolar macrophages. In M. D. Attfield and E. D. Kuempfel (Eds.), *Inhaled Particles III, Proceeding of BOHS Symposium*. Unwin Brothers, London, pp. 200–233.

Brain, J. D., Knudson, D. E., Sorokin, S. P., and Davis, M. A. (1976). Pulmonary distribution of particles given by intratracheal instillation or by aerosol inhalation. *Environ. Res.* 11:13–33.

Budavari, S. (Ed.) (1989). *Merck Index*, 11th ed. Merck & Co., Rahway, NJ, p. 82.

Carpenter, C. P., Smyth, H. F., Jr., and Pozzani, U. C. (1949). The assay of acute vapor toxicity, and grading and interpretation of results on 96 chemical compounds. *J. Ind. Hyg. Toxicol.* 31:343–346.

Center for Drug Evaluation and Research (CDER) (2008). *Nonclinical Safety Evaluation of Reformulated Drug Products and Products Intended for Administration by an Alternate Route*. CDER, Washington, DC.

Chan, T. L., and Lippmann, M. (1980). Experimental measurements and empirical modelling of the regional deposition of inhaled particles in humans. *Am. Ind. Hyg. Assoc. J.* 41:399–409.

Chenoweth, M. B., Leong, B. K. J., Sparschu, G. L., and Torkelson, T. R. (1972). Toxicities of methoxyflurane, halothane and diethyl ether in laboratory animals on repeated inhalation at subanesthetic concentrations. In B. R. Fink (Ed.), *Cellular Biology and Toxicity of Anesthetics*. Williams & Wilkins, Baltimore, MD, pp. 275–284.

Clarke, S. W., and Pavia, D. (Eds.) (1984). *Aerosol and the Lung: Clinical and Experimental Aspects*. Butterworth, London.

Cox, J. S. G., Beach, J. E., Blair, A. M. J. N., Clarke, A. J., King, J., Lee, T. B., Loveday, D. E. E., Moss, G. F., Orr, T. S. C., Ritchie, J. T., and Sheard, P. (1970). Disodium cromoglycate. *Adv. Drug. Res.* 5:115–196.

Curry, S. H., Taylor, A. J., Evans, S., Godfrey, S., and Zeidifard, E. (1975). Disposition of disodium cromoglycate administered in three particle sizes. *Br. J. Clin. Pharmacol.* 2:267–270.

Dahl, A. R. (1990). Dose concepts for inhaled vapors and gases. *Toxicol. Appl. Pharmacol.* 103:185–197.

Dallas, C. E., Bruckner, J. V., Maedgen, J. L., and Weir, F. W. (1986). A method for direct measurement of systemic uptake and elimination of volatile organics in small animals. *J. Pharmacol. Methods* 16:239–250.

Dallas, C. E., Ramanathan, R., Muralidhara, S., Gallo, J. M., and Bruckner, J. V. (1989). The uptake and elimination of 1,1,1-trichloroethane during and following inhalation exposure in rats. *Toxicol. Appl. Pharmacol.* 98:385–397.

Dautrebande, L. (1962a). Importance of particle size for therapeutic aerosol efficiency. In L. Dautrebande (Ed.), *Microaerosols*. Academic, New York, pp. 37–57.

Dautrebande, L. (1962b). Practical recommendation for administering pharmacological aerosols. In L. Dautrebande (Ed.), *Microaerosols*. Academic, New York, pp. 86–92.

Dautrebande, L. (1962c). production of liquid and solid micromicellar aerosols. In L. Dautrebande (Ed.), *Microaerosols*. Academic, New York, pp. 1–22.

Davies, R. J., and Pepys, J. (1975). Asthma due to inhaled chemical agents—The macrolide antibiotic spiramycin. *Clin. Allergy* 5:99–107.

Davis, B., Marin, M. G., Fischer, S., Graf, P., Widdicombe, J. G., and Nadel, J. A. (1976). New method for study of canine mucus gland secretion *in vivo*: Cholinergic regulation. *Am. Rev. Respir. Dis.* 113:257 (abstract).

Davis, C. N. (1961). A formalized anatomy of the human respiratory tract. In C. N. Davis (Ed.), *Inhaled Particles and Vapours*. Pergamon, London, pp. 82–91.

Davis, C. N., Heyder, J., and Subba, Ramv, M. C. (1972). Breathing of half micron aerosols. I. Experimental. *J. Appl. Physicol.* 32:591–600.

Dennis, W. L. (1961). The discussion of a paper by C. N. Davis: A formalized anatomy of the human respiratory tract. In C. N. Davis (Ed.), *Inhaled Particles and Vapours*. Pergamon, London, p. 88.

Doe, J. E., and Tinston, D. J. (1981). Novel chamber for long-term inhalation studies. In K. J. Leong (Ed.), *Inhalation Toxicology and Technology*. Ann Arbor Science: Ann Arbor, MI.

Drew, R. T., and Laskin, S. (1973). Environmental inhalation chambers. In *Methods of Animal Experimentation*, Vol. IV. Academic, New York, pp. 1–41.

Drew, R. T., and Lippmann, M. (1978). Calibration of air sampling instruments. In *Air Sampling Instruments for Evaluation of Atmospheric Contaminants*, 5th ed. American Conference of Governmental Industrial Hygienists, Cincinnati, OH, Section I, pp. 1–32.

Elliot, G. A., and DeYoung, E. N. (1970). Intranasal toxicity testing of antiviral agents. *Am. New York Acad. Sci.* 173:169–175.

Evans, M. J. (1982). Cell death and cell renewal in small airways and alveoli. In H. Witschi and P. Nettesheim (Eds.), *Mechanisms in Respiratory Toxicology*, Vol. 1. CRC, Boca Raton, FL, pp. 189–218.

Ferin, J. (1977). Effect of particle content of lung on clearance pathways. In C. L. Sanders, R. P. Schneider, G. E. Dagle, and H. A. Ragan (Eds.), *Pulmonary Macrophage and Epithelial Cells*. Technical Information Center, Energy Research & Development Administration, Springfield, VA, pp. 414–423.

Fieldner, A. C., Kazt, S. H., and Kinney, S. P. (1921). Gas masks for gases met in fighting fires. Technical Paper No. 248. U.S. Bureau of Mines, Denver, CO.

Gad, S. C. (2006). Safety assessment of therapeutic agents administration by the respiratory route. In D. E. Darlner (Ed.), *Toxicology of the Lung*, 4th ed. Taylor & Francis, New York.

Gad, S. (2007). *Statistics and Experimental Design for Toxicologists*, 4th ed. CRC Press, Boca Raton, FL.

Gad, S., and Chengelis, C. P. (1998). *Acute Toxicology Testing: Perspectives and Horizons*, 2nd ed. Academic, San Diego, CA, pp. 404–466.

Gamsu, G., Singer, M. M., Vincent, H. H., Berry, S., and Nadel, J. A. (1976). Postoperative impairment of mucous transport in the lung. *Am. Rev. Respir. Dis.* 114:673–679.

Gardner, D. E., Crapo, J. D., and Massaro, E. J. (1988). *Toxicology of the Lung*. Raven, New York.

Godfrey, S., Zeidifard, E., Brown, K., and Bell, J. H. (1974). The possible site of action of sodium cromoglycate assessed by exercise challenge. *Clin. Sci. Mol. Med.* 46:265–272.

Haber, F. R. (1924). Funf Vortage aus den jahren 1920–23, No. 3. *Die Chemie im Kriege*. Julius Springer, Berlin.

Hafner, R. E., Jr., Watanabe, P. G., and Gehring, P. J. (1975). Preliminary studies on the fate of inhaled vinyl chloride monomer in rats. *Am. New York Acad. Sci.* 246:135–148.

Hatch, T. F., and Gross, P. (1964). *Pulmonary Deposition and Retention of Inhaled Aerosols*. Academic, New York, pp. 16–17, 51–52, 147–168.

Henderson, R. (1988). Use of bronchoalveolar lavage to detect lung damage. In D. E. Gardner, J. D. Crapo, and E. J. Massaro (Eds.), *Toxicology of the Lung*. Raven, New York, pp. 239–268.

Henderson, R. (1989). Bronchoalveolar lavage: A tool for assessing the health status of the lung. In R. O. McClellan and R. F. Henderson (Eds.), *Concepts in Inhalation Toxicology*. Hemisphere, Washington DC, pp. 414–442.

Henderson, R. F. (1984). Use of bronchoalveolar lavage to detect lung damage. *Environ. Health Persp.* 56:115–129.

Henderson, R. F., Barr, E. B., and Hotchkiss, J. A. (1991). Effect of exposure rate on response of the lung to inhaled particles. *Toxicologists* 11:234 (abstract).

Henderson, R. F., Benson, J. M., Hahn, F. F., Hobbs, C. H., Jones, R. K., Mauderly, J. L., McClellan, R. O., and Pickrell, J. A. (1985). New approaches for the evaluation of pulmonary toxicity: Bronchoalveolar lavage fluid analysis. *Fund. Appl. Toxicol.* 5:451–458.

Henderson, R. F., and Lowrey, J. S. (1983). Effect of anesthetic agents on lavage fluid parameters used as indicators of pulmonary injury. *Lab. Anim. Sci.* 33:60–62.

Henderson, R. F., Mauderly, J. L., Pickrell, J. A., Hahn, F. F., Muhle, H., and Rebar, A. H. (1987). Comparative study of bronchoalveolar lavage fluid: Effect of species, age, method of lavage. *Exp. Lung Res.* 1:329–342.

Hensley, M. J., O'Cain, C. F., McFadden, E. R., Jr., and Ingram, R. H., Jr. (1978). Distribution of bronchodilatation in normal subjects: Beta agonist versus atropine. *J. Appl. Physiol.* 45:778–782.

Herkenne, C., Alberti, I., Naik, A., Yogeshvar, N. K., Mathy, F. X., Preat, V., and Guy, R. H. (2008). *In vivo* method for the assessment of topical drug bioavailability. *Pharm Res.* 25(1):87–103.

Hext, P. M. (2000). Inhalation toxicology. In B. Ballantyne, T. Marrs, and T. Syversen (Eds.), *General and Applied Toxicology*. Macmillan: London, pp. 587–601.

Heyder, J., Gebhart, J., and Stahlhofen, W. (1980). Inhalation of aerosols: Particle deposition and retention. In K. Willeke (Ed.), *Generation of Aerosols and Facilities for Exposure Experiments*. Ann Arbor Science, Ann Arbor, MI, pp. 80–99.

Hiller, F. C., Mazunder, M. K., Wilson, J. D., and Bone, R. C. (1978). Aerodynamic size distribution of metered dose bronchodilator aerosols. *Am. Rev. Respir. Dis.* 118:311–317.

Hinds, W. C. (1980). Dry-dispersion aerosol generators. In K. Willeke (Ed.), *Generation of Aerosols and Facilities for Exposure Experiments*. Ann Arbor Science, Ann Arbor, MI, pp. 171–187.

Hocking, W. G., and Golde, D. W. (1979). The pulmonary alveolar macrophage. *New Engl. J. Med.* 310:580–587, 639–645.

Hodson, M. E., Penketh, A. R., and Batten, J. C. (1981). Aerosol carbenicillin and gentamicin treatment of *Pseudomonas aeruginosa* infection I patients with cystic fibrosis. *Lancet* 2:1137–1139.

Hollinger, M. A. (1985). Mucokinetic agents, radionuclides, and anticoagulant-thrombolytic agents. In *Respiratory Pharmacology and Toxicology*. W. B. Saunders, Philadelphia, PA, pp. 133–134.

Holma, B. (1967). Lung clearance of mono- and di-disperse aerosols determined by profile scanning and whole body counting: A study on normal and $SO_2$ exposed rabbits. *Acta Med. Scand.* 473(Suppl.):1–102.

Horsfield, K., and Cumming, G. (1968). Morphology of the bronchial tree in man. *J. Appl. Physiol.* 24:373–383.

Ingram, R. H., Wellman, J. J., McFadden, E. R., Jr., and Mead, J. (1977). Relative contributions of large and small airways to flow limitation in normal subjects before and after atropine and isoproterenol. *J. Clin. Invest.* 59:696–703.

Irish, D. D., and Adams, E. M. (1940). Apparatus and methods for testing the toxicity of vapors. *Ind. Med. Surg.* 1:1–4.

Johnson, H. G., McNee, M. L., and Braughler, J. M. (1987). Inhibitors of metal catalyzed lipid peroxidation reactions inhibit mucus secretion and 15 HETE levels in canine trachea. *Prostagland. Leukotr. Med.* 30:123–132.

Johnson, H. G., McNee, M. L., Johnson, M. A., and Miller, M. D. (1983). Leukotriene $C_4$ and dimethylphenylpiperazinium-induced responses in canine airway tracheal muscle contraction and fluid secretion. *Int. Arch. Allergy Appl. Immunol.* 71:214–218.

Jones, J. G. (1984). Clearance of inhaled particles from alveoli. In S. W. Clarke and D. Pavia (Eds.), *Aerosols and the Lung: Clinical and Experimental Aspects*. Butterworth, London, pp. 172–196.

Karol, M. H. (1988). Immunologic responses of the lung to inhaled toxicants. In R. O. McClellan and R. Henderson (Eds.), *Concepts in Inhalation Toxicology*. Hemisphere Publishing, Washington, DC, pp. 403–413.

Karol, M. H., and Thorne, P. S. (1988). Hypersensitivity and hyperreactivity. In D. E. Gardner, J. D. Erapo, and E. J. Massaro (Eds.), *Toxicology of the Lung*. Raven Press, New York, pp. 427–448.

Karol, M. H., Hillebrand, J. A., and Thorne, P. S. (1989). Characteristics of weekly pulmonary hypersensitivity responses elicited in the guinea pig by inhalation of ovalbumin aerosols. In D. E. Gardner, J. D. Crapo, and E. J. Massaro (Eds.), *Toxicology of the Lung*. Raven, New York, pp. 427–448.

Kavet, R. I., Brain, J. D., and Levens, D. J. (1978). Characteristics of weekly pulmonary macrophages lavaged from hamsters exposed to iron oxide aerosols. *Lab. Invest.* 38:312–319.

Landy, T. D., Ramsey, J. C., and McKenna, M. J. (1983). Pulmonary physiology and inhalation dosimetry in rats: Development of a method and two examples. *Toxicol. Appl. Pharmacol.* 71:72–83.

Last, J. A. (1982). Mucus production and ciliary escalator. In H. Witschi and P. Nettesheim (Eds.), *Mechanisms in Respiratory Toxicology*, Vol. 1. CRC Press, Boca Raton, FL, pp. 247–268.

Lauweryns, J. M., and Baert, J. H. (1977). Alveolar clearance and the role of the pulmonary lymphatics. *Am. Rev. Resp. Dis.* 115:625–683.

Lee, W.-C., and Wang, C.-S. (1977). Particle deposition in systems of repeated bifurcating tubes. In W. H. Walton (Ed.), *Inhaled Particles*, Vol. 4. Oxford, Pergamon, pp. 49–60.

Leong, B. K. J., Coombs, J. K., Petzold, E. N., and Hanchar, A. J. (1988). A dosimetric endotracheal nebulization technique for pulmonary metabolic disposition studies in laboratory animals. *Inhal. Toxicol.* Premier issue: 37–51.

Leong, B. K. J., Coombs, J. K., Petzold, E. N., Hanchar, A. H., and McNee, M. L. (1985). Endotracheal nebulization of drugs into the lungs of anesthetized animals. *Toxicologist* 5:31 (abstract).

Leong, B. K. J., Lund, J. E., Groehn, J. A., Coombs, J. K., Sabaitis, C. P., Weaver, R. J., and Griffin, L. (1987). Retinopathy from inhaling 4,4′-methylenedianiline aerosols. *Fund. Appl. Toxicol.* 9:645–658.

Leong, B. K. J., and MacFarland, H. N. (1965). Pulmonary dynamics and retention of toxic gases. *Arch. Environ. Health* 11:555–563.

Leong, B. K. J., Powell, D. J., and Pochyla, G. L. (1981). A new dust generator for inhalation toxicological studies. In B. K. J. Leong (Ed.), *Inhalation Toxicology and Technology.* Ann Arbor Science, Ann Arbor, MI, pp. 157–168.

Leong, B. K. J., and Rop, D. A. (1989). The combined effects of an inhalation anesthetic and an analgesic on the electrocardiograms of beagle dogs. In *Abstract-475. International Congress of Toxicology*. Taylor & Francis, London, p. 159.

Lichtiger, M., Landa, J. F., and Hirsch, J. A. (1975). Velocity of tracheal mucus in anesthetized women undergoing gynaecologic surgery. *Anesthesiology* 42:753–756.

Lippmann, M. (1970). "Respirable" dust sampling. *Am. Ind. Hyg. Assoc. J.* 31:138–159.

Lippmann, M. (1977). Regional deposition of particles in the human respiratory tract. In D. H. K. Lee, L. K. Falk, and S. D. Murphy (Eds.), *Handbook of Physiology. Section 9: Reactions of Environmental Agents*. American Physiology Society, Bethesda, MD, pp. 213–232.

Lippmann, M., Yeates, D. B., and Albert, R. E. (1980). Deposition, retention and clearance of inhaled particles. *Br. J. Ind. Med.* 37:337–362.

Litchfield, J. T., Jr., and Wilcoxon, F. (1949). A simplified method of evaluating dose-effect experiments. *J. Pharmacol. Exp. Ther.* 96:99–113.

Lourenco, R. V., and Cotromanes, E. (1982). Clinical aerosols. II. Therapeutic aerosols. *Arch. Intern. Med.* 142:2299–2308.

MacFarland, H. N. (1976). Respiratory toxicology. In W. J. Hayes (Ed.), *Essays in Toxicology*, Vol. 7. Academic, New York, pp. 121–154.

Marple, V. A., and Rubow, K. L. (1980). Aerosol generation concepts and parameters. In K. Willeke (Ed.), *Generation of Aerosols and Facilities for Exposure Experiments*. Ann Arbor Science, Ann Arbor, MI, p. 6.

McClellan, R. O., and Henderson, R. F. (1989). *Concepts in Inhalation Toxicology*. Hemisphere, Washington, DC.

McFadden, E. R., Jr. (1986). *Inhaled Aerosol Bronchodilators*. Williams & Wilkins, Baltimore, MD, pp. 40–41.

Menon, M. P. S., and Das, A. K. (1977). Tetracycline asthma—A case report. *Clin. Allergy* 7:285–290.

Menzel, D. B., and Amdur, M. O. (1986). Toxic responses of the respiratory system. In C. D. Klaassen, M. O. Amdur, and J. Doull (Eds.), *Casarett and Doull's Toxicology*, 3rd ed. Macmillan, New York, pp. 330–358.

Mercer, T. T. (1981). Production of therapeutic aerosols; principles and techniques. *Chest* 80(Suppl. 6):813–818.

Merin, R. G. (1981). Cardiac toxicity of inhalation anesthetics. In T. Balazs (Ed.), *Cardiac Toxicology*, Vol. II. CRC Press, Boca Raton, FL, pp. 4–10.

Miller, F. J., Gardner, D. E., Graham, J. A., Lee, R. E., Jr., Wilson, W. E., and Bachmann, J. D. (1979). Size considerations for establishing a standard for inhalable particles. *J. Air Pollat. Control. Assoc.* 29:610–615.

Miller, J. L., Stuart, B. O., Deford, H. S., and Moss, O. R. (1981). Liquid aerosol generation for inhalation toxicology studies. In B. K. J. Leong (Ed.), *Inhalation Toxicology and Technology*. Ann Arbor Science, Ann Arbor, MI, pp. 121–207.

Moren, F. (1981). Pressurized aerosols for oral inhalation. *Int. J. Pharm.* 8:1–10.

Mussatto, D. J., Garrad, C. S., and Lourenco, R. V. (1988). The effect of inhaled histamine on human tracheal mucus velocity and bronchial mucociliary clearance. *Am. Rev. Respir. Dis.* 138:775–779.

National Academy of Sciences (NAS). (1958). *Handbook of Respiration*. NAS National Research Council. W. B. Saunders, Philadelphia, p. 41.

Newhouse, M., Sanchis, J., and Bienenstock, J. (1976). Lung defense mechanisms. *New Engl. J. Med.* 295:1045–1052.

Newman, S. P. (1984). Therapeutic aerosols. In S. W. Clarke and D. Pavia (Eds.), *Aerosols and the Lung, Clinical and Experimental Aspects*. Butterworth & Company, London, pp. 197–224.

Newton, P. E. (2000). Techniques for evaluating hazards of inhaled products. In *Product Safety Evaluation Handbook*, 2nd ed. Marcel Dekker, New York, pp. 243–298.

Oberst, F. W. (1961). Factors affecting inhalation and retention of toxic vapours. In C. N. Davis (Ed.), *Inhaled Particles and Vapours*. Pergamon, New York, pp. 249–266.

Oyarzun, M. J., and Clements, J. A. (1977). Ventilatory and cholinergic control of pulmonary surfactant in rabbit. *J. Appl. Physiol.* 43:39–45.

Oyarzun, M. J., and Clements, J. A. (1978). Control of lung surfactant by ventilation, adrenergic mediators and prostaglandins in the rabbits. *Am. Rev. Respir. Dis.* 117:879–891.

Ozhai, H., Wilhelm, K.-P., and Maibach, H. I. (2008). *Dermatoxicity*, 7th ed. CRC Press, Boca Raton, FL.

Parent, R. A. (1991). *Comparative Biology of the Normal Lung*. CRC Press, Boca Raton, FL.

Paterson, J. W. (1977). Bronchodilators. In T. J. H. Clark and S. Godfrey (Eds.), *Asthma*. Chapman and Hall, London, pp. 251–271.

Paterson, J. W., Woocock, A. J., and Shenfield, G. M. (1979). Bronchodilator drugs. *Am. Rev. Respir. Dis.* 120:1149–1188.

Patterson, R., and Kelly, J. F. (1974). Animal models of the asthmatic state. *Ann. Rev. Med.* 25:53–68.

Patty, F. A. (1958). *Industrial Hygiene and Toxicology*, 2nd ed. Interscience, New York.

Paustenbach, D. J., Carlson, G. P., Christian, J. E., Born, G. S., and Rausch, J. E. (1983). A dynamic closed-loop recirculating inhalation chamber for conducting pharmacokinetic and short-term toxicity studies. *Fund. Appl. Toxicol.* 3:528–532.

Pavia, D. (1984). Lung mucociliary clearance. In S. W. Clarke and D. Pavia (Eds.), *Aerosols and the Lung, Clinical and Experimental Aspects*. Butterworth, London, pp. 127–155.

Pavia, D., Bateman, J. R. M., Sheahan, N. F., Agnew, J. E., Newman, S. P., and Clarke, S. W. (1980). Techniques for measuring lung mucociliary clearance. *Eur. J. Respir. Dis.* 67(Suppl. 110):157–177.

Pavia, D., Sutton, P. P., Agnew, J. E., Lopez-Vidriero, M. T., Newman, S. P., and Clarke, S. W. (1983a). Measurement of bronchial mucociliary clearance. *Eur. J. Respir. Dis.* 64(Suppl. 127):41–56.

Pavia, D., Sutton, P. P., Lopez-Vidriero, M. T., Agnew, J. E., and Clarke, S. W. (1983b). Drug effects on mucociliary function. *Eur. J. Respir. Dis.* 64(Suppl. 128):304–317.

Phalen, R. F. (1984). *Inhalation Studies: Foundations and Techniques*. CRC Press, Boca Raton, FL, pp. 35–46, 51–57.

Raabe, O. G. (1982). Deposition and clearance of inhaled aerosols. In H. Witschi and P. Nettesheim (Eds.), *Mechanisms in Respiratory Toxicology,* vol. 1. CRC Press, Boca Raton, FL, pp. 35–46, 51–57.

Rinehart, W. E., and Hatch, T. (1964). Concentration–time product (*Ct*) as an expression of dose in sublethal exposures to phosgene. *Am. Ind. Hyg. Assoc. J.* 25:545–553.

Rylander, R. (1966). Current techniques to measure alterations in the ciliary activity of intact respiratory epithelium. *Am. Rev. Respir. Dis.* 93(Suppl):67–85.

Sackner, M. A. (1978). Effect of respiratory drugs on mucociliary clearance. *Chest* 73(Suppl. 6):958–966.

Sackner, M. A., Brown, L. K., and Kim, C. S. (1981). Basis of an improved metered aerosol delivery system. *Chest* 80(Suppl. 6):915–918.

Sackner, M. A., Landa, J., Hirsch, J., and Zapata, A. (1975). Pulmonary effects of oxygen breathing. A 6-hour study in normal men. *Ann. Intern. Med.* 82:40–43.

Salem, H. (1986). Principles of inhalation of inhalation toxicology. In *Inhalation Toxicology: Research Methods, Applications, and Evaluation*. Marcel Dekker, New York, pp. 1–33.

Schaper, M., Detwiler, K., and Alarie, Y. (1989). Alteration of respiratory cycle timing by propranolol. *Toxicol. Appl. Pharmacol.* 97:538–547.

Schreck, R. W., Sekuterski, J. J., and Gross, K. B. (1986). Synchronized intratracheal aerosol generation for rodent studies. In *Aerosols, Formation and Reactivity, Second International Aerosol Conference, Berlin*. Pergamon, New York, pp. 37–51.

Schwettmann, R. S., and Casterline, C. L. (1976). Delayed asthmatic response following occupational exposure to enflurane. *Anesthesiology* 44:166–169.

Sidorenko, G. I., and Pinigin, M. A. (1976). Concentration-time relationship for various regimens of inhalation of organic compounds. *Environ. Health Perspect.* 13:17–21.

Silver, S. D. (1946). Constant flow gassing chambers: Principles influencing design and operation. *J. Lab. Clin. Med.* 31:1153–1161.

Smith, D. M., Ortiz, L. W., Archuleta, R. F., Spalding, J. F., Tillery, M. I., Ettinger, H. J., and Thomas, R. G. (1981). A method for chronic nose-only exposure of laboratory animals to inhaled fibrous aerosols. In B. K. J. Leong (Ed.), *Inhalation Toxicology and Technology*. Ann Arbor Science, Ann Arbor, MI, pp. 89–105.

Stahlhofen, W., Gebhart, J., and Heyder, J. (1980). Experimental determination of the regional deposition of aerosol particles in the human respiratory tract. *Am. Ind. Hyg. Assoc. J.* 41:385–398.

Stahlhofen, W., Gebhart, J., and Heyder, J. (1981). Biological variability of regional deposition of aerosol particles in the human respiratory tract. *Am. Ind. Hyg. Assoc. J.* 42:348–352.

Stott, W. T., and McKenna, M. J. (1984). The comparative absorption and excretion of chemical vapors by the upper, lower and intact respiratory tract of rats. *Fund. Appl. Toxicol.* 4:594–602.

Swift, D. B., and Proctor, D. F. (1982). Human respiratory deposition of particles during oronasal breathing. *Atmos. Environ.* 16:2279–2282.

Task Group on Lung Dynamics (1966). Deposition and retention models for internal dosimetry of the human respiratory tract. *Health Phys.* 12:173–207.

Taulbee, D. B., Yu, C. P., and Heyder, J. (1978). Aerosol transport in the human lung from analysis of single breaths. *J. Appl. Physiol.* 44:803–812.

Thorne, P. S., and Karol, M. H. (1989). Association of fever with late-onset pulmonary hypersensitivity responses in the guinea pig. *Toxicol. Appl. Pharmacol.* 100:247–258.

Touitou, E., and Barry, B. W. (Eds.). (2007). *Enhancement of Drug Delivery*. CRC Press, Boca Raton, FL.

Ts'o, T. O. T., Leong, B. K. J., and Chenoweth, M. B. (1975). Utilities and limitations of behavioral techniques in industrial toxicology. In B. Weiss and V. G. Laties (Eds.), *Behavioral Toxicology*. Plenum, New York. pp. 265–291.

Uemitsu, N., Minobe, Y., and Nakayoshi, H. (1985). Concentration-time-response relationship under conditions of single inhalation of carbon tetrachloride. *Toxicol. Appl. Pharmacol.* 77:260–266.

Ulrich, C. E., Klonne, D. R., and Church, S. V. (1984). Automated exposure system for metered-dose aerosol pharmaceuticals. *Toxicologist* 4:48.

U.S. Food and Drug Administration (FDA). (2006). Guidance for industry and review staff: Photosafety toxicity testing for adverse effects of drugs.

Valentine, R., and Kennedy, G. L., Jr. (2001). Inhalation toxicology. In W. Hayes (Ed.), *Principles and Methods of Toxicology*, 4th ed. Raven, New York, pp. 1085–1143.

Vidgren, M. T., Karkkainen, A., Paronen, T. P., and Karjalainen, P. (1987). Respiratory tract deposition of $^{99}$Tc-labeled drug particles administered via a dry powder inhaler. *Int. J. Pharmaceut.* 39:101–105.

Wanner, A. (1979). The role of mucociliary dysfunction in bronchial asthma. *Am. J. Med.* 67:477–485.

Wanner, A. (1981). Alteration of tracheal mucociliary transport in airway disease. Effect of pharmacologic agents. *Chest* 80(Suppl. 6):867–870.

Webber, S. E., and Widdicombe, J. G. (1987). The actions of methacholine, phenylephrine, salbutamol and histamine on mucus secretion from the ferret in-vitro trachea. *Agents and Actions* 22:82–85.

Weibel, E. R. (1963). *Morphometry of the Human Lung*. Springer-Verlag, Heidelberg.

Weibel, E. R. (1983). How does lung structure affect gas exchange? *Chest* 83:657–665.

Weiner, N. (1984). Norepinephrine, epinephrine, and the sympathomimetic amines. In A. G. Gilman, L. S. Goodman, T. W. Rall, and F. Murad (Eds.), *Goodman and Gilman's, The Pharmacological Basis of Therapeutics*, 7th ed. Macmillan, New York, pp. 145–180.

Weston, R., and Karel, L. (1946). An application of the dosimetric method for biologically assaying inhaled substances. *J. Pharmacol. Exp. Ther.* 88:195–207.

Widdicombe, J. G. (1974). Reflex control of breathing. In J. G. Widdicombe (Ed.), *Respiratory Physiology*, Vol. 2. University Park Press, Baltimore, MD.

Willeke, K. (1980). *Generation of Aerosols and Facilities for Exposure Experiments*. Ann Arbor Science, Ann Arbor, MI.

Williams, M. H. (1974). Steroids and antibiotic aerosols. *Am. Rev. Respir. Dis.* 110:122–127.

Wilson, A. G. E. (1982). Toxicokinetics of uptake, accumulation, and metabolism of chemicals by the lung. In H. Witschi and P. Nettesheim (Eds.), *Mechanisms in Respiratory Toxicology*, Vol. 1. CRC Press, Boca Raton, FL, pp. 162–178.

Witschi, H. P., and Brain, J. D. (1985). *Toxicology of Inhaled Materials*. Springer-Verlag, New York.

Witschi, H. P., and Nettesheim, P. (1982). *Mechanisms in Respiratory Toxicology*, Vol. 1. CRC Press, Boca Raton, FL.

Zhai, H., Wilhelm, K. P., and Maibach, H. I. (Eds.) (2008). *Marzulli and Maibach's Dermatoxicology*. CRC Press, Boca Raton, FL.

# *Special-Case Products: Imaging Agents and Oncology Drugs*

## 1 INTRODUCTION

Special cases in the assessment of nonclinical safety arise from a variety of situations or for a number of reasons. We have already explored the major cases that arise as a result of different routes of administration—inhalation and topical routes with their variations. These are the most common variations from the general cases of oral and parenteral administration.

Likewise, the major cases associated with different physical forms (small molecules, being the general case, the major variation being biotechnology sourced materials) have been discussed.

In this chapter, two special cases which arise from the nature of the therapeutic use are considered: imaging agents (in vivo diagnostic agents) and oncology drugs.

## 2 IMAGING AGENTS

Medical imaging agents used for diagnosis or monitoring serve a variety of different modes of activity such as radiography, computed tomography (CT), ultrasonography, magnetic reasonance imaging (MRI), and radionuclide

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

imaging (Szabo, 2004). These imaging agents can be classified into at least two general categories—contrast agents and diagnostic radiopharmaceuticals (Van Brocklin, 2008).

Historically, imaging agents (and the subset of radiopharmaceuticals) start as a real exception to the general case in that they do not require an investigational new drug (IND) before an initial clinical evaluation. Rather, under the RDRC [Radioactive Drug Research Committee, a special type of institutional review board (IRB)] process as specified in Code of Federal Regulations (CFR), Title 21, Part 36.1, single-dose studies may be conducted in one of the few academic medical centers with only RDRC review and with use limited to a small number of individuals (up to 30) (CDER, 2000).

The RDRC is an institutional body that reviews research protocols for scientific and technical merit. An overview of the RDRC program from 1975 to 2004 has been recently reported (Van Brocklin, 2008). As of 2003, there were 84 active RDRCs in the United States.

In order to be GRAS (generally regarded as safe), a radiopharmaceutical is limited in terms of pharmacological and radiation dose. The mass associated with the radiopharmaceutical must "be known not to cause any clinically detectable pharmacological effect in human beings" (CDER, 2000). As a result of this limitation, FIH (first in human) studies must be limited to "microdosing" under an RDRC protocol. Typically, RDRCs require published human studies involving the tracer to be evaluated before approving a protocol at near-therapeutic dose. The dose limitation requires that the smallest radiation dose needed to obtain meaningful data from the study be administered to the study subject. The maximum allowable single dose to the whole body, blood-forming organs, lens of the eye, and gonads is 30 mSv (3 rem) with a maximum annual or total dose to all other organs of 50 mSv (5 rem). The maximum single dose and total annual dose to all other organs is 50 mSv (5 rem) and 150 mSv (15 rem), respectively. There is also a significant radiation dose limit on studies involving research subjects that are less than 18 years of age. The dose may not exceed 10% of the adult doses reported above. Additionally, all radiation doses associated with the study must be included in the total study dose. This means that the CT dose from a positron emission tomography (PET)/CT study must be included in the total and this total may not exceed the maximum limits set forth in the regulations.

The types of studies that may be conducted under an RDRC-approved protocol are also regulated. The research must be basic in nature and may include the evaluation of the radiopharmaceutical pharmacoKinetic (PK), metabolism, and excretion. The distribution of a radiopharmaceutical to evaluate human physiology, pathophysiology, or biochemistry is permitted as long as the studies are not for diagnostic or therapeutic benefit. Safety and efficacy studies are not permitted under these regulations. An example of a study that is permissible under RDRC would be the brain distribution of [$^{18}$F] fluoror DOPA relative to subject age or neurodegenerative disease.

The RDRC may not approve protocols that require more than 30 subjects. If more than 30 subjects need to be studied and may be justified by the

researcher then a special summary form is submitted to the U.S. Food and Drug Administration (FDA) for review and a formal IND must be opened. A pediatric consultant to the RDRC must review studies involving minors under 18 years old and a special summary must be submitted to the FDA. In addition, all adverse reactions "attributable to the use of the radioactive drug" must be reported immediately to the FDA (CDER, 2000). It is interesting to note, however, that over 30 years since the inception of the RDRC regulations with an estimated 60,000 subjects enrolled in the studies not one adverse event has been reported.

The regulations also stipulate the constitution of the RDRC with appropriate expertise to review the protocol applications. The committee must have at least five members. Three of those members must be a nuclear medicine physician, a qualified individual with radiopharmaceutical preparation experience, and a radiation dosimetry/radiation safety expert. The remaining members must have experience and qualifications in disciplines related to nuclear medicine.

### 2.1 Contrast Agents

As used in the guidance, a contrast agent is a medical imaging agent used to improve the visualization of tissues, organs, and physiological processes by increasing the relative difference of imaging signal intensities in adjacent regions of the body. Types of contrast agents include but are not limited to (1) iodinated compounds used in radiography and CT; (2) paramagnetic metallic ions (such as ions of gadolinium, iron, and manganese) linked to a variety of molecules and microparticles (such as superparamagnetic iron oxide) used in MRI; and (3) microbubbles, microaerosomes, and related microparticles used in diagnostic ultrasonography.

### 2.2 Diagnostic Radiopharmaceuticals

As used in the guidance, a diagnostic radiopharmaceutical is (1) an article that is intended for use in the diagnosis or monitoring of a disease or a manifestation of a disease in humans and that exhibits spontaneous disintegration of unstable nuclei with the emission of nuclear particles or photons or (2) any nonradioactive reagent kit or nuclide generator that is intended to be used in the preparation of such an article. As stated in the preamble to the FDA's proposed rules on regulations for in vivo radiopharmaceuticals used for diagnosis and monitoring, the agency interprets this definition to include articles that exhibit spontaneous disintegration leading to the reconstruction of unstable nuclei and subsequent emission of nuclear particles or photons (*Federal Register*, Vol. 63, 28301 to 28303, May 22, 1998).

Diagnostic radiopharmaceuticals are generally radioactive drugs or biological products that contain a radionuclide that typically is linked to a ligand or carrier. These products are used in nuclear medicine procedures, including

planar imaging, single-photon emission computed tomography (SPECT), PET, or in combination with other radiation detection probes.

Diagnostic radiopharmaceuticals used for imaging typically have two distinct components:

- A radionuclide that can be detected in vivo (e.g., technetium-99m, iodine-123, indium-111): The radionuclide typically is a radioactive atom with a relatively short physical half-life that emits radioactive decay photons having sufficient energy to penetrate the tissue mass of the patient. The photons can then be detected with imaging devices or other detectors.
- A nonradioactive component to which the radionuclide is bound that delivers that radionuclide to specific areas within the body: This nonradionuclide portion of the diagnostic radiopharmaceutical often is an organic molecule such as a carbohydrate, lipid, nucleic acid, peptide, small protein, or antibody.

As technology advances, new products may emerge that do not fit into these traditional categories (e.g., agents for optical imaging, magnetic resonance spectroscopy, combined contrast and functional imaging). It is anticipated that the general principles discussed here could apply to these new diagnostic products. Developers of these products should contact the appropriate reviewing FDA division for advice on product development.

## 2.3 Medical Imaging Agent Characterisitics Relevant to Safety

The following sections discuss the special characteristics of a medical imaging agent that can lead to a more focused safety evaluation. Characteristics include its radiation absorbed dose, mass dose, route of administration, frequency of use, biodistribution, and biological, physical, and effective half-lives in the serum, the whole body, and critical organs.

***Mass Dose*** Some medical imaging agents can be administered at low mass doses. For example, the mass dose of a single administration of a diagnostic radiopharmaceutical can be small because device technologies can typically detect relatively small amounts of radionuclide (e.g., radiopharmaceuticals for myocardial perfusion imaging). When a medical imaging agent is administered at a mass dose that is at the low end of the dose–response curve, safety concerns are minimal.

***Route of Administration*** Some medical imaging agents are administered by routes that decrease the likelihood of systemic adverse events. For example, medical imaging agents that are administered as contrast media for radiographic examination of the gastrointestinal tract (e.g., barium sulfate) can be administered orally, through an oral tube, or rectally. In patients with normal gastrointestinal tracts, many of these products are not absorbed, so systemic

adverse events are less likely to occur. In general, nonradiolabeled contrast agents pose safety issues similar to therapeutic drugs and generally should be treated as therapeutic agents for the purpose of conducting clinical safety assessments.

***Frequency of Use*** Many medical imaging agents, including both contrast and diagnostic radiopharmaceuticals, are administered infrequently or as single doses. Accordingly, adverse events that are related to long-term use or to accumulation are less likely to occur with these agents than with agents that are administered repeatedly to the same patient. Therefore, the nonclinical development programs for such single-use products usually can omit long-term (i.e., 3 months duration or longer), repeat-dose safety studies. In clinical settings where it is possible that the medical imaging agent will be administered to a single patient repeatedly (e.g., to monitor disease progression), we recommend that repeat-dose studies (of 14–28 days duration) be performed to assess safety.

Biological medical imaging agents are frequently immunogenic, and the development of antibodies after intermittent, repeated administration can alter the pharmacokinetics, biodistribution, safety, and/or imaging properties of such agents and, potentially, of immunologically related agents. We recommend studies in which repeat dosing of a biological imaging agent is planned to incorporate pharmacokinetic data, human antimouse antibody (HAMA), human antihumanized antibody (HAHA), or human antichimeric antibody (HACA) levels as well as whole-body biodistribution imaging to assess for alterations in the biodistribution of the imaging agent following repeat dosing. Studies of immunogenicity in animal models are generally of little value. Therefore, we recommend that human clinical data assessing the repeat use of a biological imaging agent be obtained prior to application to licensure of such an agent.

***Biological, Physical, and Effective Half-Lives*** Diagnostic radiopharmaceuticals often use radionuclides with short physical half-lives or that are excreted rapidly. The biological, physical, and effective half-lives of diagnostic radiopharmaceuticals are incorporated into radiation dosimetry evaluations that require an understanding of the kinetics of distribution and excretion of the radionuclide and its mode of decay. We recommend that biological, physical, and effective half-lives be considered in planning appropriate safety and dosimetry evaluations of diagnostic radiopharmaceuticals.

## 2.4 Performance of Nonclinical Safety Assessments

The FDA recommends that the nonclinical development strategy for an agent be based on sound scientific principles, the agent's unique chemistry (including, e.g., those of its components, metabolites, and impurities), and the agent's intended use. Because each product is unique, we encourage sponsors to

consult with us before submitting an IND application and during product development. The number and types of nonclinical studies recommended would depend in part on the phase of development, what is known about the agent or its pharmacological class, its proposed use, and the indicated patient population. If it is determined that nonclinical pharmacology or toxicology studies are not needed, the FDA is prepared to grant a waiver under 21 CFR 312.0 if adequate justification is provided.

In the discussion that follows, a distinction is made between drug products and biological products. The existing specific guidance for biological products is referenced but not repeated here (ICH S8).

## *Nonclinical Safety Assessments for Nonbiological Drug Products*

### Timing of Nonclinical Studies Submitted to IND application

- The FDA recommends that nonclinical studies be timed so that they help facilitate the timely conduct of clinical trials (including appropriate safety monitoring based on findings in nonclinical studies) and reduce the unnecessary use of animals and other resources. The recommended timing of nonclinical studies for medical imaging drugs is summarized in Table 1.

### Contrast Agents

- Because of the characteristics of contrast drug products (e.g., variable biological half-life) and the way they are used, the FDA recommends that nonclinical safety evaluations of such drug products be made more efficient with the following modifications:
  (a) Long-term (i.e., greater than 3 months), repeat-dose toxicity studies in animals usually can be omitted. (Exceptions are products with long residence time, e.g., >90 days.)
  (b) Long-term rodent carcinogenicity studies usually can be omitted.
  (c) Reproductive toxicology studies required under CFR 312.23(a)(8)(ii)(a) often can be limited to an evaluation of embryonic and fetal toxicities in rats and rabbits and to evaluations of reproductive organs in other short-term toxicity studies. If you determine that such reproductive studies are not needed, FDA is prepared to grant a waiver under CFR 312.10 if adequate justification is provided.

The FDA recommends that studies be conducted to address the effects of large mass dose and volume (especially for iodinated contrast materials administered intravenously); osmolality effects; potential transmetallation of complexes of gadolinium, manganese, or iron (generally MRI drugs); potential effects of tissue or cellular accumulation on organ function (particularly if the drug is intended to image a diseased organ system); and the chemical, physiological, and physical effects of ultrasound microbubble drugs (e.g., coalescence, aggregation, margination, and cavitation).

**TABLE 1 Timing of Nonclinical Studies for Nonbiological Products Submitted to IND**

| Study Type | Before Phase 1 | Before Phase 2 | Before Phase 3 | Before NDA |
|---|---|---|---|---|
| Safety pharmacology | Major organs[a] and organ systems the drug is intended to visualize | | | |
| Toxicokinetic pharmacokinetic | See ICH guidances | | | |
| Expanded single-dose toxicity | Expanded acute single dose[b] | | | |
| Short-term (2–4 weeks) multiple-dose toxicity | — | Repeat-dose toxicity[c] | | |
| Special toxicology | Conduct as necessary based on route irritancy, blood compatibility, protein flocculation, misadministration, extravasation | | | |
| Radiation dosimetry | If applicable | | | |
| Genotoxicity | In vitro[d] | Complete standard battery | | |
| Immunotoxicity | — | — | May be needed based on molecular structure, biodistribution pattern, class concern, or clinical or nonclinical signal | |
| Reproductive and developmental toxicity | — | — | Needed or waiver obtained[d] | |
| Drug interaction | — | — | — | As needed |
| Other based on data results | — | — | — | As needed |

[a]See guidances S7: A safety pharmacology studies for human pharmaceutical and S7B: Safety pharmacology studies for assessing the potential for delayed ventricular repolarization (QT interval prolongation) by human pharmaceuticals (note that S7B allows for phase evaluation of required studies).
[b]See the guidance on single-dose acute testing for pharmaceuticals.
[c]When repeat-dose toxicity studies have been performed but single-dose toxicology studies have not, dose selection for initial human studies will likely be based on the results of the no-adverse-effect level (NOAEL) obtained in repeat-dose study. The likely result will be a mass dose selection for initial human administration that is lower than if the dose selection had been based on the results of acute, single-dose toxicity studies.
[d]See radiopharmaceutical discussion in Section III.B.1.c of CDER, 2000.

***Diagnostic Radiopharmaceuticals (Nonbiological Products)*** Because of the characteristics of diagnostic radiopharmaceuticals and the way they are used, we recommend that nonclinical safety evaluations of these drugs be made more efficient by the following modifications:

- Long-term, repeat-dose toxicity studies in animals typically can be omitted.
- Long-term rodent carcinogenicity studies typically can be omitted.
- Reproductive toxicology studies can be waived when adequate scientific justification is provided.
- Genotoxicity studies should be conducted on the nonradioactive component because the genotoxicity of the nonradioactive component should be identified separately from that of the radionuclide. Genotoxicity studies can be waived if adequate scientific justification is provided.

The FDA recommends that special safety considerations for diagnostic radiopharmaceuticals include verification of the mass dose of the radiolabeled and unlabeled moiety; assessment of the mass, toxic potency, and receptor interactions for any unlabeled moiety; assessment of potential pharmacological or physiological effects due to molecules that bind with receptors or enzymes; and evaluation of all components in the final formulation for toxicity (e.g., excipients, reducing drugs, stabilizers, antioxidants, chelators, impurities, and residual solvents). It is recommended that the special safety considerations include an analysis of particle size (for products containing particles) and an assessment of instability manifested by aggregation or precipitation. It is also recommended that an individual component be tested if specific toxicological concerns are identified or if toxicological data for that component are lacking. However, if toxicological studies are performed on the combined components of a radiopharmaceutical and no significant toxicity is found, toxicological studies of individual components are seldom required.

***Nonclinical Safety Assessments for Biological Products*** Many biological products raise relatively distinct nonclinical issues such as immunogenicity and species specificity. We recommend the following agency documents be reviewed for guidance on the preclinical evaluation of biological medical imaging agents:

- S6 Preclinical Safety Evaluation of Biotechnology-Derived Pharmaceuticals
- Points to Consider in the Manufacture and Testing of Monoclonal Antibody Products for Human Use

Table 2 seeks to integrate these nonclinical safety testing requirements into a test matrix.

Unfortunately, most recently it has come to light that imaging agents used for cardiovascular effects (Definity) are associated with patient deaths (*Health Imaging News*, 2008). The Optison agent also has come under suspicion.

**TABLE 2 Special Case: Imaging Agents**

| Test Requirment | Species[a] |
|---|---|
| **Initial clinical trial/IND Requirements**[b] | |
| 1. Acute toxicity in rodents (IV), expanded | R/M |
| 2. Acute toxicity in nonrodents (IV), expanded | D/S/P |
| 3. Genotoxicity: bacterial mutagenicity (Ames) | In vitro |
| 4. Genotoxicity: in vitro clastogenicity (CHO chromosome aberration) | In vitro |
| 5. Genotoxicity: in vivo (mouse or rat micronucleus) | R/M |
| 6. Safety pharmacology: CV in vivo | D/P/S |
| 7. Safety pharmacology: respiratory, rodent | R |
| 8. Pivotal/repeat dose in nonrodents (14–28 days oral) | D/P/S |
| 9. Develop bioanalytical for 2 species (human/nonrodent) | NA |
| 10. Hemolysis | In vitro |

[a]R = rat, M = mouse, D = dog, S = pig, P = primate.
[b]Initial single-dose clinical study at single academic center may not require IND.
*Note*: All studies described must be performed according to good laboratory practices.
*Source*: Center for Biologics Evaluation and Research (CBER), 2004.

## 3 ONCOLOGY DRUGS

Traditional oncology drugs now have a specific proposed International Conference on Harmonisation (ICH) guidance for their nonclinical safety assessment requirements (ICH, 2008). This chapter reflects that guidance.

These drugs are intended for use in individuals with a serious and life-threatening disease. Initial clinical trials are conducted in patients who have already failed other forms of therapy. Because of this, as summarized in Tables 3 and 4, the requirements are not as strenuous as the general case. Indeed, genotoxicity and safety pharmacology are usually not required (Baguley and Kerr, 2002).

The other unusual feature in such trials, whether the drug is a small or large (protein) molecule, is that clinical dosing regiments are not daily but rather in accordance with a schedule set as much by tradition and clinical operations as by drug pharmacokinetics.

Administrations are usually two or three times a week on a three- or four-week sequence of dosings followed by a period without dosing. It is normal practice for an initial clinical tolerance trial to be conducted with the dose level being escalated with each three- or four-week dosing series. The 28-day two-species nonclinical safety assessment studies are taken to serve for the entire multiseries set of administrations.

Dose scaling as we now practice it across the range of pharmaceuticals arose from the practice of oncology. Clinicians in this therapeutic area think of dosing in terms of milligrams per square meter, which leads to the need to perform conversions of the expression of dose in safety studies (milligrams per kilogram) to the dose for body surface area form.

**TABLE 3 Special Case: Oncology Agents (Cytotoxic)**

| Test Requirment | Species[a] |
|---|---|
| **Initial clinical trial/IND Requirements** | |
| 1. Acute toxicity in rodents (IV) | R/M |
| 2. Acute toxicity in nonrodents (IV) | D/S/P |
| 3. 7-day DRF toxicity in rodents (IV) | R/M |
| 4. 7-day DRF toxicity in nonrodents (IV) | D/S/P |
| 5. Safety pharmacology: CV-hERG[b] | In vitro |
| 6. Pivotal/repeat dose in rodents (14–28 days IV) | R/M |
| 7. Pivotal/repeat dose in nonrodents (14–28 days IV) | D/P/S |
| 8. CYP induction/inhibition[b] | In vitro |
| 9. Five species microsome metabolic panel[b] | In vitro |
| 10. Develop bioanalytical for 3 species (human/rodent/nonrodent) | NA |
| **To support continued clinical development** | |
| 10. Pivotal/repeat dose in rodents (3/6 months oral) | R/M |
| 11. Pivotal/repeat dose in nonrodents (3/9–12 months oral) | D/P/S |

[a]R = rat, M = mouse, D = dog, S = pig, P = primate.
[b]Recommended.
*Note*: All studies described must be performed according to good laboratory practices.

**TABLE 4 Special Case: Oncology Agents (Protein-Targeted Molcules)**

| Test Requirment | Species |
|---|---|
| **Initial clinical trial/IND Requirements** | |
| 1. Acute toxicity in rodents (IV) | R/M |
| 2. Acute toxicity in nonrodents (IV) | D/S/P |
| 3. 7-day DRF toxicity in rodents (IV) | R/M |
| 4. 7-day DRF toxicity in nonrodents (IV) | D/S/P |
| 5. Safety pharmacology: CV in vivo | D/P/S |
| 6. Safety pharmacology: FOB/Irwin | R/M |
| 7. Safety pharmacology: respiratory, rodent | R |
| 8. Pivotal/repeat dose in rodents (14–28 days IV) | R/M |
| 9. Pivotal/repeat dose in nonrodents (14–28 days IV) | D/P/S |
| 10. CYP induction/inhibition[b] | In vitro |
| 11. Five species microsome metabolic panel[b] | In vitro |
| 12. Develop Bioanalytical for 3 species (human/rodent/nonrodent) | NA |
| **To support continued clinical development** | |
| 13. Developmental toxicity (segment II), rat and rabbit pilots and rat and rabbit studies | R/B |
| 14. Immunotoxicity | TBD |
| 15. Pivotal/repeat dose in nonrodents (3/6 months oral) | D/P/S |
| **To support marketing approval** | |
| 16. Reproductive toxicity, segment I | R |
| 17. Reproductive toxicity, segment III | R |

[a]R = rat, M = mouse, D = dog, S = pig, P = primate, B = rabbit, TBD = To be determined.
[b]Recommended.
[c]May be required.
*Note*: All studies described must be performed according to good laboratory practices.

The toxicity of cytotoxic anticancer drugs correlates more closely with body surface area (BSA) than with body weight (BW). Thus, it is standard practice to administer and compare such drugs on the basis of BSA. To convert between milligrams per kilogram and milligrams per square meter, it is necessary to know either BSA or BW. Accurate values for BW are easy to obtain. In contrast, BSA is difficult to measure and therefore is usually calculated from a known BW. The relationship between BW and BSA is described by the formula: $BSA(m^2) = b[BW/(kg)]^{2/3}$, where $b$ is a species-specific factor derived from empirical data. In practice, the values for $b$ vary between laboratories (e.g., values for $b$ for the mouse have been found to vary from 0.090 to 0.096). Confusion and miscommunication can occur as a result of the use of different species-specific factors. For example, difficulties have arisen during the dose selection process for a novel cytotoxic anticancer agent with a sleep dose–response curve. In this case, if different investigators use $b$ values of 0.099, 0.101, and 0.111, then animal BSA values for a 10-kg dog will likewise vary, depending on the choice of $b$ (0.926, 0.945, and 1.038 mg, respectively, for a 2-mg/m$^2$ dose). The use of a single set of species-specific factors by all investigators is proposed as it would be beneficial in removing confusion from discussions of study designs and results. Such a proposal has an even more immediate benefit in the case of compounds with very steep dose–response curves, where even small variations in dosing can profoundly influence the results obtained and their interpretation.

### 3.1 Dose Conversions: Perspective

Dosages of pharmaceuticals are typically calculated on the basis of BW. Interspecies comparisons of toxicological effects of drugs and chemicals are commonly based on exposures normalized with respect to body weight or systemic exposure [i.e., area under the curve (AUC) and $C_{max}$ values].

### 3.2 Use of mg/m$^2$ Dose Unit

***Calculations of Drug Dosages for Treatment*** Cytotoxic anticancer drugs are typically administered in dose units of milligrams per square meter. However, BSA is difficult to measure. In contrast, values for height and body weight are easy to obtain. Two solutions present themselves:

1. In a clinic, estimates of BSA can be accomplished from height and weight data using established nomograms.
2. In the toxicology laboratory, BSA can be calculated from animal body weight data *using the appropriate formula.*

***Conversion of mg/kg Body Weight Doses to Units of mg/m$^2$*** Conversions from milligrams per kilogram body weight to milligrams per square meter occur often in the design of toxicity studies, in the interpretation of data

**TABLE 5 Factors Used for Milligram-per-Kilogram to Milligram-per-Square-Meter Conversions**

| Species | Conversion Factor |
|---|---|
| Mouse | 3 |
| Rat | 6 |
| Guinea pig | 7.7 |
| Hamster | 4.1 |
| Rabbit | 11.8 |
| Dog | 20 |
| Monkey | 12 |
| Human (60 kg) | 37 |

from a number of different studies, including old studies, and in comparisons of animal and human data where dose units are not similar. Table 5 provides conversion factors for this.

Another unusual feature in oncology is that the maximum utilized clinical dose is the toxic-limited dose (TLD). Patients are typically titrated to toxicity and actual clinical treatment practice doses patients up to a frankly tolerated dose.

Somewhat in parallel, what constitutes a dose-limiting adverse effect in a nonclinical study [a no-observable-adverse-effect level (NOAEL)] is usually taken to be a frank toxicity and not a simple intolerance (such as sporadic emesis).

## REFERENCES

Baguley, B. C., and Kerr, D. J. (2002). *Anticancer Drug Development*. Academic, San Diego, CA.

Center for Biologics Evaluation and Research (CBER) (2004). Developing medical imaging drug and biological products Part I: conducting safety assessments, guidance for industry. CBER, Washington, DC.

Center for Drug Evaluation and Research (CDER) (2000). Biological products: Developing medical imaging drugs and biological products, guidance for industry. CDER, Washington, DC.

*Health Imaging News* (2008). FDA warns that deaths continue with heart imaging agents. June 23, 2008.

International Conference on Harmonization (ICH) S9 (2008). Guideline on Nonclinical Evaluation for Anticancer Pharmaceuticals. December.

Szabo, T. L. (2004). *Diagnostic Ultrasound Imaging: Inside Out*. Elsevier, San Diego.

Van Brocklin, H. F. (2008). Radiopharmaceuticals for drug development: United States regulatory perspective. *Curr. Radiopharm.* 1:2–6.

# *Occupational Toxicology in Pharmaceutical Industry*

## 1 INTRODUCTION

Most of the assessment of toxicology and safety of therapeutics is focused on the patients who are to benefit from the new medicine. However, there are two other groups of individuals (each of which has different exposure profiles) that one must be concerned about: health care providers (nurses, pharmacists, and physicians) who provide and/or administer the drugs and individuals involved in manufacturing them. The concerns here are in the realm of occupational toxicology.

Modern toxicology has its roots in the occupational environment. The earliest recorded observations relating exposure to chemical substances and toxic manifestations were made about workers. These include Agricola's identification of the diseases of miners and Pott's investigation of scrotal cancer incidence among chimney sweeps. Occupational toxicology, as its name implies, concerns itself with the toxicological implications of exposure to chemicals in the work environment.

In this chapter we will examine occupational toxicology as it applies to and is currently practiced in the pharmaceutical industry. This industry, which by definition involves biologically active compounds, has been a driving force in the development of the science of toxicology. The need for a thorough safety evaluation of potential therapeutics prior to marketing approval has driven

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

the continued evolution of toxicological testing methods and the identification of mechanisms of toxic action. The area of occupational toxicology gained great momentum in the pharmaceutical industry from the early 1980s through the 1990s. This is made clear by the increased number of companies that have implemented occupational toxicology programs during this period. Yet, occupational-related activities generally represent only a small fraction of activity in safety assessment in the pharmaceutical and biotechnology industries. It is difficult to gauge the level of activity in this occupational areadue to the paucity of publications on the subject (Teichman et al., 1988). This is probably the result of the fact that most occupational toxicologists function in an administrative environment and thus experience less pressure to publish and that they in general deal with information relating to new chemical processes that may not be protected by patents. The lack of general knowledge about the function of the occupational toxicologist that has resulted could lead one to conclude that (1) the thorough evaluation of drugs to obtain marketing approval makes an investigation of their potential hazards to manufacturing employees unnecessary and (2) because they are used therapeutically, pharmaceutical agents are safe under any circumstances in occupational settings.

## 2 OCCUPATIONAL TOXICOLOGY VERSUS DRUG SAFETY EVALUATION

While pharmaceutical products are indeed created to treat disease, they cannot always be considered nonhazardous. The clinician must evaluate the benefits to the patient in light of any side effects or adverse reactions that may be associated with drug usage and any other toxicological properties uncovered in animal studies. Examples range from antibiotics, which have clastogenic properties in mice (for which, to the patient, its activity in suppressing life-threatening infections presents a clear and overriding benefit), to an antineoplastic that has extreme renal toxicity but which effectively kills established tumors. The occupational toxicologist must look at pharmaceutical agents in a completely different light. These same risk–benefit analyses do not apply in an occupational setting. Even agents with minimal clinical adverse reactions and whose pharmacological activity could be considered generally beneficial in the clinical setting may present certain employees with health hazards in manufacturing or health care provider settings.

Table 1 presents some of the basic differences in the way preclinical and occupational toxicologists must approach their work. Preclinical development of a pharmaceutical product requires exhaustive testing of drug candidates under the requirements of the U.S. Food and Drug Administration (FDA) or equivalent national agency to help predict and evaluate clinical findings and to preclude serious or chronic hazards that would not ordinarily be observed in clinical studies limited in duration and population (FDA, 1990). By contrast, the occupational toxicologist must evaluate the potential of a compound to

**TABLE 1 Comparison of Occupational and Preclinical Toxicology**

| Parameter | Occupational | Preclinical |
|---|---|---|
| Purpose | Potential for effects from unintended exposure | Predict and evaluate clinical findings and preclude serious hazards from clinical use |
| Routes of administration | Inhalation, direct contact with skin or eyes | Oral and/or parenteral most likely |
| Dose level | Not really predictable | Relative to estimated therapeutic dose or maximum tolerated dose |
| Duration of exposure | Extremely variable; depends on campaign/batch, procedure, etc.; may be short daily exposure for working lifetime | Dependent on therapeutic use and test model |

cause toxicity from unintended exposures via a variety of routes of administration and a wide range of exposure levels of varying lengths. In drug safety evaluations, studies are designed to approximate the clinical setting, particularly in terms of routes of administration and dosage. Thus, most preclinical studies generally focus on oral and/or parenteral administration with dosages that either are comparable to or exceed therapeutic levels. However, neither oral nor parenteral administration is a likely route of exposure among employees. Rather, during manufacturing operations employees are more likely to be exposed via inhalation or direct contact with the skin or eyes. In addition to the route of exposure itself, the effects occurring following direct contact or inhalation exposure may be of a nature not predictable by the studies undertaken for preclinical safety evaluation. Most important among these effects are irritation and sensitization. Dermal, ocular, and respiratory irritation potential generally cannot be predicted from preclinical studies utilizing oral or parenteral administration. Similarly, sensitization, which has the potential to significantly add to the difficulty of conducting manufacturing operations safely, is difficult to evaluate even with current specific methods and models. For most pharmaceutical agents, testing to ascertain the potential to induce dermal sensitization reactions is not conducted during a typical preclinical development program.

Other important differences lie in the length of treatment and the dosages involved. Therapeutic use of pharmacological agents may be of acute or limited duration, such as in the administration of anti-infectives, or chronic, as with antihypertensive agents. Occupational exposure may also be of varying length, limited by shift or batch manufacturing methods. It is possible, however, that the manufacture of certain high-volume products such as antibiotics may result in daily exposure, if only in limited doses, over a significant portion of a working lifetime. The levels to which employees may be exposed are, in

general, potentially lower than those that are used therapeutically, although exposures will vary with the type of operations performed.

The area of occupational toxicology has received a great deal of attention in the chemical industry. Historically, the chemical industry has focused on the occupational environment and developed many of our current toxicological methods to address health and safety concerns. However, since the mid-1970s the chemical industry has increasingly become subject to testing requirements relevant to the protection of the environment and the public at large, as mandated by U.S. Environmental Protection Agency regulations (EPA, 1976, 1979). Data development for occupational health hazard evaluation has seldom been sought by the Occupational Safety and Health Administration (OSHA). Consequently, few new test methods have been developed and those in current use are generally modifications of methods introduced in the 1930s and 1940s. Some of the differences in the issues addressed by occupational toxicologists in the pharmaceutical and chemical industries are highlighted in Table 2. Among these, a major difference lies in the physical and functional nature of the substances involved. Pharmaceutical agents are generally handled as solids, while chemical industry products are generally processed as liquids or vapors. Although there is a greater focus on the consequences of occupational exposure, the occupational toxicologist in the chemical industry rarely has available the wealth of information that exists in the

**TABLE 2 Toxicological Testing Requirements under EC Seventh Amendment (Directive 92/32/EC—Notification of New Substances)**

| Quantity Imported to or Manufactured in EC | Test Results to be Submitted |
|---|---|
| <1000 kg yr$^{-1}$ or <5000 kg total | None needed unless compound is considered toxic (oral $LD_{50}$ 25–200 mg kg$^{-1}$) or very toxic (oral $LD_{50}$<25 mg kg$^{-1}$) |
| >1000 kg yr$^{-1}$ or 5000 kg total | Acute oral/dermal $LD_{50}$<br>Acute inhalation $LC_{50}$<br>Skin irritation<br>Eye irritation<br>Skin sensitization<br>28-day subacute toxicity<br>Mutagenicity (bacteriological and nonbacteriological tests)<br>Acute toxicity to fish ($LC_{50}$)<br>Acute toxicity to *Daphnia* |
| >10,000 kg yr$^{-1}$ or 50,000 kg total | Additional tests may be required depending on results, including:<br>• Fertility—1 or 2 generations (males or females)<br>• Teratology—additional species<br>• Subchronic or chronic—90 days to 2 years<br>• Carcinogenicity<br>• Acute and subacute on additional species |

*Source*: EEC, 1992.

pharmaceutical industry. New drug dossiers include toxicological information as well as data on the pharmacology, pharmacokinetics, and mechanism of action, and, most important, much of this information has been gathered from clinical trials on human beings. Even though the data are not developed for the purpose of evaluating the occupational environment, they can be invaluable for this purpose. Clinical studies, even if the doses and routes of administration may be different from those used in clinical trials/therapeutically, provide insight into the unique responses of the human body. Another important difference in the parameters involved in the chemical and pharmaceutical industries concerns biological activity. The chemical industry strives to minimize it, while pharmaceutical agents are specifically designed to be biologically active. Recent advances in our understanding of molecular and cellular processes have led to the development of agents with improved specificity for unique receptor or molecular targets. These potent agents may present increased hazards for employees and a great challenge to the occupational toxicologist in the pharmaceutical industry.

At the same time, several types of data necessary to ensure proper management of occupational risks associated with a drug substance are not generally useful in evaluating potential patient risks. So the necessary tests—eye and skin irritation, sensitization and inhalation toxicity, as well as assessment of the hazards of by-products and impurities that do not get incorporated into the final therapeutic product—are not performed in the normal course of development.

## 3 REGULATORY PRESSURES IN UNITED STATES AND EUROPEAN COMMUNITY

The safety and health of workers in the United States is regulated under the Occupational Health and Safety Act of 1970, which established OSHA. Since its inception, OSHA has promulgated a variety of health standards, including compound-specific regulations, permissible exposure limits (PELs), and rules for providing access to medical records (OSHA, 1986) and for communicating to employees the hazards of the materials they handle. This last regulation, the hazard communication standard (OSHA, 1987), is a standard that specifically requires manufacturers or importers to carry out an evaluation of the toxicological properties of chemicals. This standard outlines specific criteria for evaluating substances as hazardous or nonhazardous. However, there still is no U.S. regulatory requirement for testing a compound of unknown toxicity (Gad, 2001). Rather, such a compound could be classified as nonhazardous based on the unavailability of data. Pharmaceutical agents are generally considered hazardous under the standard since they meet the criterion of having a biological effect on humans. Any adverse reaction observed during clinical use will be construed as toxicity, however irrelevant to the occupational environment. The main result of classification as hazardous is a requirement

to develop a material safety data sheet (MSDS) as the main vehicle for providing to employees. The model MSDS suggested for use in complying with the OSHA hazard communication standard contains a great deal of information about the physical properties and hazards and the procedures necessary to deal with the accidental spill, fire, explosion, or accidental contact with hazardous material. The standard requires that all information regarding adverse effects in human beings and most animal toxicity data be included in the MSDS, however irrelevant this information may be to the work environment. The resulting MSDS can be a highly technical document that may not be the optimal vehicle for conveying this type of information to manufacturing employees handling pharmaceutical agents.

In the environmental area, the EPA's TSCA (Toxic Substances Control Act) (EPA, 1976) regulations for filing of premanufacture notification (PMN) (EPA, 1979) have resulted in the development of toxicological information on many new industrial chemicals. New chemical entities generated for use as pharmaceutical agents are exempted from PMN requirements. This exemption may also extend to all intermediates generated during chemical synthesis. Many pharmaceutical companies have instituted toxicological testing of these compounds even though there is no specific U.S. regulatory impetus to develop such information. The European Community (EC) has implemented several directives that parallel and exceed the OSHA hazard communication standard and TSCA regulations. European Community Directive 80/1107 [European Economic Community (EEC), 1979] requires employee communication of hazards of chemical substances as well as biological materials. Another EEC directive, which was the impetus for the development of notification of new substances regulations in several member nations, requires the development of toxicological data on new compounds and does not exempt pharmaceutical agents or isolated intermediates (EEC, 1979). Notification and testing must be conducted in accordance with the amount of the substance manufactured in or imported into Europe yearly (Table 2). These regulations will support the development of toxicological data on entities manufactured or processed in Europe that can in turn be applied to occupational health hazard evaluations. Pharmaceutical companies are, as demonstrated in Table 3, very different from chemical companies in their handling of occupational toxicology.

## 4 ORGANIZATIONAL STRUCTURE

The occupational toxicology function is organized and structured in very different ways across the industry, particularly with so much of the sector now composed of small companies that are not in possession of any internal development resources. The function exists in many of the major PrMA (Pharmaceutical Manufacturing Association) member companies. In most of these companies the occupational toxicology function is located within the employee safety/industrial hygiene area, while in some it resides within research and

**TABLE 3 Comparison of Occupational Toxicology in Pharmaceutical and Chemical Industries**

| Parameter | Pharmaceutical | Chemical |
|---|---|---|
| Compounds | | |
| Physical state | Generally solids | Liquids, vapors, polymers, solids |
| Biological activity | Designed for biological activity | Strive for biological inertness |
| Toxicology data | | |
| Development | Focus on preclinical evaluation | Focus on occupational and general environment |
| Study length | Acute to chronic for final products; acute for intermediates | Acute to chronic (depending on volume) |
| Human data | | |
| ADME[a] | Generally available for oral/parenteral routes | Not generally available |
| Mechanism of action | Targeted during drug development | Not generally studied |
| Adverse effects | Extensive clinical trial studies for final products from oral or parenteral route | Generally only known as result of overexposure, accident, etc. |

[a]ADME: absorption, distribution, metabolism, and excretion.

development (R&D), toxicology, or employee health/medical services areas. How the function fits into the organization greatly depends on its mission. Occupational toxicology will function well with the R&D environment if evaluation of occupational health hazards is considered an integral requirement in the development and approval process. In such an organization there would likely be a greater emphasis on developing toxicological data on novel compounds and their synthetic intermediates, rather than on existing processes, or such other activities as training. Most occupational toxicology departments are located within the employee safety/industrial hygiene areas. This organization provides great opportunity for interaction and cooperation with those disciplines that are charged with implementing the toxicologists' recommendations. Good interaction between the occupational toxicologist and the industrial hygienist can be particularly useful in developing and implementing solutions to potential health hazards. However, poor understanding of the limitations of scientific data by the more engineering-oriented safety specialists may lead to unrealistic expectations for easy solutions or answers. The last existing arrangement is for the occupational toxicologist to report into the medical services area. This arrangement provides perhaps the easiest interactions for the toxicologist, who shares a common language and understanding of biological systems with the occupational physician. However, in order to effect any changes in the work environment, it is necessary to enlist the aid of the employee safety/industrial hygiene group, an act that may incur the potential problems mentioned above. Clearly, wherever the function is located, the occupational toxicologist must be able to interact well with a

variety of disciplines, including R&D, safety, industrial hygiene, medical services, legal services, regulatory affairs, technical services, and, of course, operations management.

Staffing of industrial toxicology programs varies among the different programs, including groups with two or three full-time Ph.D.'s who spend all of their time on occupational issues and those with one or two Ph.D.'s or masters-level staff members who may have part-time responsibility for occupational-related issues along with R&D responsibilities. The level of staffing depends, of course, on the activities assigned to the occupational toxicology group and these may vary from one organization to another. It is impossible to generalize or recommend an adequate staffing level, since that will be dictated by the emphasis placed on specific activities. Whatever the mission of the occupational toxicology function, a high level of education or professional credentials is desirable. A doctoral degree and/or board certification in toxicology should be imperative to be able to interact effectively with many of the other disciplines mentioned above, particularly R&D management.

## 5 ACTIVITIES

The scope of activities of occupational toxicologists may be quite different from one organization to another, depending on its specific mission, resources available, and corporate culture. In general, their activities can be divided into four broad areas: data development, data evaluation and dissemination, hazard assessment, and employee training.

### 5.1 Data Evaluation and Dissemination

It is important to establish who will use the toxicological information provided and how this information will be applied. Unlike the preclinical toxicologist who provides information to other toxicologists, to the regulatory agencies, or to physicians for evaluation of potential therapeutic liabilities, the occupational toxicologist is providing information to a variety of individuals and functions. First, the information will be provided to the industrial hygienist or safety specialist who must evaluate the quality of the work environment and the appropriateness of personal protective equipment. Second, the information will be given to the occupational physician who must evaluate the potential causes of any symptoms reported by employees who may have been exposed to the material. Third, R&D, plant management, and/or manufacturing services must evaluate the need to implement engineering or other controls and weigh these costs against the commercial viability of the product. Last, but not least, the information must be provided to production employees who will be handling the compound and who need to know of its hazards. Clearly, there is a need to provide the necessary information in such a way

that it can be clearly understood by nonscientists. With an audience of such potentially wide-ranging educational levels and understanding, multiple communication vehicles may be necessary.

To be effective the toxicology evaluation must meet several criteria: it must (1) be thorough, (2) be clear and concise, (3) be in a form appropriate to its target audience, and (4) include a conclusion or recommendation.

Thoroughness can be achieved through an exhaustive search of the published literature using the available computerized databases. There is a risk, particularly when dealing with pharmaceutical agents, that the most relevant information to occupational toxicology can be overlooked in the great number of clinical case reports, many of which are not relevant to the work environment. In general, little information has been published on the occupational hazards of pharmaceutical agents. A thorough review does not mean a listing of every reported clinical adverse reaction. This type of information is more likely to confuse the reader and my lead him to her to ignore the important occupational hazards. The toxicologist must therefore be extremely selective in performing this evaluation. A review of the available clinical information, however, may yield data that can be used in the evaluation, particularly if the product has been tested for dermal administration. An integral part of assuring the thoroughness of the evaluation must be a process of updating the information on a regular basis. In general, it is unlikely that new clinical data will significantly change a review for an established pharmaceutical agent. However, new therapeutic entities should be reviewed more frequently since new data may be published on potential adverse reactions not identified in clinical trials, and these data may impact the occupational evaluation.

When providing information to technical personnel, it is best to use language that does not require the use of a medical dictionary. It is tempting to use medical terms, particularly when quoting from the clinical literature. However, use of these terms may result in poor understanding of the information and may also evoke unnecessary anxiety in the reader. A good rule of thumb is to think of what the reader will do with the information: If the biological effect will require more than a few words to be clearly explained in plain language and it is irrelevant or unimportant to the work environment, it is best left off any communication to the field.

It is not always possible to reach a conclusion regarding the degree of hazard of exposure to a compound, particularly if the data are not directly relevant to the work environment. There is often a temptation to provide a thorough evaluation, setting out all necessary information in plain language but leaving the formulation of a conclusion to the reader. However, if it is difficult for the trained toxicologist to reach such a conclusion, it must be even more difficult for the layperson. If an estimate of the hazard cannot be reached, then the evaluation must be concluded with a recommendation of the type of exposures that may increase hazard or the type of effect that is most likely to occur should there be an overexposure. These may at least give the industrial hygienist or physician a useful reference point. At the same time, it is

important to express to the reader the inherent limitations of such a conclusion. The audience may expect black-and-white answers; if this is not possible, they should be made to understand why.

Perhaps the most difficult part of the communication equation is the ability to match the information to the audience. This may be best illustrated using an example. Over the past 10 years one company has developed several vehicles for communicating information to various audiences. One instrument is the toxicology review. In general, this is a one- to two-page document that reviews the published literature on the compound. A reference list is prepared and maintained on file for future reference. This review is provided to safety, medical, and industrial hygiene personnel and, if appropriate, research chemists. These individuals have received training to help them understand the terms used and the effects outlined. A second method of communication involves a computerized database. This personal computer–based system, which provides only bottom-line information, is available online via a modem to safety, industrial hygiene, nursing, and research personnel (Sussman and Gáler, 1990). It includes only that information specifically relevant to the work environment and necessary for compliance with OSHA hazard communication standard or EC directive 80/1107. A third vehicle was developed jointly with an industrial hygiene department and consists of a short paragraph highlighting the specific hazards of the compound followed by safe handling recommendations. A fourth commonly used method is the MSDS. The toxicology department prepares the toxicology section of the MSDS. The appropriate other disciplines complete the remaining sections, and the completed MSDS is then reviewed and approved by a committee. Last, for certain compounds, on-site training programs, such as those described below, can be presented by the toxicologist on the hazards of the chemical. These various formats for the same information were developed to serve the informational needs and educational levels of various audiences. This is one approach to filling the need to communicate toxicological information to a variety of groups. The appropriate vehicle for each company will, of course, depend on the available resources and corporate culture. Even a large number of formats may not suffice. The occupational toxicologist should determine, through discussions and follow-up communications, how the information is received and if it is understood. The communication of toxicological information may represent approximately 50% of the occupational toxicologists' responsibilities, thus explaining the level of commitment to developing appropriate formats. The MSDS alone is often insufficient for the successful communication of health hazard information to employees.

### 5.2 Data Development

The motivation for conducting toxicological tests for pharmaceutical, chemical intermediates and impurities arises from the need to ensure the health of employees by preventing the occurrence of adverse reactions from occupa-

tional exposure. Employers thus secondarily minimize the associated potential for work interruption. Programs in place at many larger companies routinely test new drug candidates and/or isolated synthetic intermediates for the purpose of occupational health hazard evaluation.

The development of a toxicological testing program for occupational health hazard evaluation requires consideration of (1) the compounds to be tested, (2) the stage of drug development at which testing occurs, (3) the specific tests to be conducted, and (4) the means for funding.

These four issues are, of course, interdependent, and it is not always possible to deal with one without affecting the others. As indicated previously, drug candidates undergo extensive toxicological testing to ensure an adequate margin of safety for patients. Additional tests are usually required to obtain information specific to the work environment. By contrast, synthetic intermediates are generally not subject to testing for drug safety evaluation. These compounds, if they have the potential to present an exposure hazard to employees, may warrant evaluation. Clearly, it is neither feasible nor necessary to conduct the same level of testing required for drug marketing approval. However, a toxicological assessment can often be developed to determine whether these isolated intermediates have the potential to elicit toxicity from exposures that could occur during work.

Compounds should be selected for testing on the basis of an evaluation of potential exposure and likelihood of their causing adverse effects. The first evaluation is best achieved by including the research chemist, industrial hygienist, and/or safety specialist in the decision-making process. They are in the best possible position for judging potential sources of employee exposure. Including these disciplines in the pretesting stage ensures not only their commitment to the program but also that the studies will be designed with careful consideration of the work experience. The second evaluation, an estimate of toxic effects, may be obtained from a comparison of the compounds in question to known toxic agents, also known as a structure–activity relationship (SAR) evaluation. Software programs are currently available for obtaining a quantitative estimate of toxicity using SAR models. However, it is most likely that an SAR evaluation will be achieved by simple comparison to the final product, similar pharmacological agents, or raw materials that have known toxic properties. Information on potential exposures and toxic effects can thus be utilized to decide which compounds to test or to assign priorities to compounds selected for testing.

The timing of these studies depends greatly on the developmental track for the test compounds and may vary for intermediates and final products. Discovery early in the development process that an isolated intermediate poses a significant health hazard may prompt a change in the chemical synthesis or process or in the implementation of engineering controls or personal protective equipment. This it is generally useful to test intermediates at an early stage. This approach presents several practical problems. First, in a long development program, such as occurs in the pharmaceutical industry, there are many oppor-

tunities for changing the synthetic route for reasons other than toxicity. Thus, a large percentage of the intermediates tested during the early development stages may be replaced in the ultimate manufacturing process. Second, only a fraction of new drug candidates actually reach the drug approval process. Therefore, the majority of intermediates tested early in the development process will likely never reach large-scale manufacture. Conducting a toxicological assessment of intermediates at a later stage in the development process presents a comparable set of advantages and disadvantages: It is more likely that the compounds tested will be manufactured on a large scale, but the ability to make fundamental changes in the chemical process will be greatly diminished. Testing of new drug candidates for occupational health hazards can be an integral part of the drug safety evaluation process. Acute oral toxicity is frequently evaluated as the first step in the drug's safety assessment. Adding acute dermal toxicity and thus skin irritation evaluation at the same time can often be accomplished with minimum impact on the development schedule. This additional information can then be used to protect not only employees manufacturing supplies of the chemical but also laboratory employees handling test doses of the substance. Eye irritation testing requires minimal amounts of test compounds and could also be accomplished at the same time. Sensitization testing requires a greater commitment in terms of time and the quantity of compounds needed. Therefore, investigation of a compound's allergenic properties is often postponed until sufficient toxicological information is available to permit a decision as to whether the compounds will advance to the next stage in the development process.

Practical considerations of funding and the selection of the testing laboratory need to be addressed when developing an occupational toxicology testing program. As indicated above, if the activity is located within the R&D department, it may be simple to include the cost of conducting these tests within the new drug's development budget. There is a possible risk in this situation, however, that the safety and industrial hygiene communities may be inadvertently omitted from the prioritization process and the information loop. Explaining the necessity of testing programs to nonscientific management personnel may be challenging. Solutions to these barriers may be found with R&D funding of testing or designation of testing cost, thus possibly including these programs in research funds.

While there are no set regulations on what tests should be done or when these should be conducted and, indeed, no activity to preclude such testing on animal welfare grounds, there is general agreement as to the type of effects that need to be addressed: skin and eye irritation, sensitization, and acute oral and dermal toxicity. Testing for these effects generally involves studies of short duration. Thus, results can often be obtained relatively quickly. Additional tests for inhalation toxicity and/or sensory irritation are conducted by several companies.

Although there is general agreement on the effects to be investigated, the methods used have not necessarily been consistent. Several companies have

developed protocols uniquely tailored to the needs of their workplace health hazard evaluation and their in-house testing resources. The most common protocols utilized for occupational health hazard evaluation are briefly described in Table 4 (Gad and Chengelis, 1998).

These modifications included a combined protocol to assess acute dermal toxicity as well as skin irritation in rabbits and a stepwise approach to acute oral toxicity determination rather than a classic $LD_{50}$ (Gáler, 1989). Doses are selected based on regulatory criteria, such as those that are required for classification as a toxic under the OSHA hazard communication standard and/or EEC 80/1107. Testing for eye irritation involves a modification of current methods using rabbits. While in some views there is no justification for testing cosmetic produces in live animals, eye irritation information pertaining to unique pharmacological chemicals is important to protect employees from accidental exposures. There are currently no acceptable alternatives to the rabbit eye irritation test (Society of Toxicology, 1989); therefore, the rabbit eye irritation test is used by the occupational toxicologist. Current protocols include refinements to the original method, including a reduction in the number of animals used, the application of topical anesthetics to decrease animals' sensation, and rinsing with distilled water following the instillation of the test compound to allow evaluation of the benefits of washing the eye as a first aid measure. Another refinement that may be utilized is a reduction in the amount of material instilled into the eye (Griffith and Yam, 1989). In general, modifications of this type have not affected the reliability of this test (Hatoum et al., 1990) and may, in fact, better simulate possible workplace accidents and provide additional information.

The battery of tests shown in Table 4 can provide useful information to complete a workplace hazard assessment. However, they are not the only tools that may be used to determine the toxic potential of a workplace contaminant. Additional tests may be required to provide more rigorous recommendations. Depending on the results of initial tests, a second stage of testing may be initiated to address specific needs. For example, sensory irritation tests may be conducted for compounds that are found to have irritant properties. The sensory irritation test, developed by Alarie (1966), is used to develop a parameter—the $RD_{50}$—that has been directly correlated with threshold limit values (TLVs) for a certain class of compounds (Kane et al., 1979; Alarie, 1981). However, the usefulness of this test for solid compounds, which include most pharmaceutical agents, has not been determined. Results of genotoxicity tests may present a need for testing in additional systems to assess genotoxic potential. Mechanistic studies may also be appropriate for certain compounds, such as intermediates, in the synthesis of inhibitors of specific enzymes or receptor agonists/antagonists. The information available from clinical, pharmacology, or pharmacokinetic studies on the final drug can be useful in determining possible avenues for investigation. A particularly interesting type of study, yet to be developed, might involve determination of the absorption and bioavailability of compounds from occupational exposures that could then be related

**TABLE 4 Summary of Protocols used for Current Test Methods**

| Test | Species | Method | Dose | Data Application | Additional Data |
|---|---|---|---|---|---|
| Acute toxicity | | | | | |
| Oral | Rat or mouse | *N* = 5/sex/dose; 14-day observation period; necropsy with/without histopathology | Chosen as limit (0.5 or 5 g kg$^{-1}$) or for $LD_{50}$ | Classify compounds as harmful/toxic/highly toxic | |
| Dermal | Rabbit | *N* = 5/sex/dose; 24-h dermal application under occlusion; 14-day observation period; necropsy with/without histopathology; dermal irritation scores | Chosen as limit (2 g kg$^{-1}$) or for $LD_{50}$ | Classify compounds as harmful/toxic/highly toxic; selection of protective equipment | Irritation potential class; target organ information |
| Inhalation | Rat or mouse | *N* = 5/sex/dose; 4-h nose only or whole-body exposure; 14-day observation period; necropsy with/without histopathology | Chosen as limit (20 g m$^{-3}$) or for $LD_{50}$ | Classify compounds as harmful/toxic/highly toxic; selection of protective equipment | Respiratory irritation potential; target organ information |
| Skin irritation | Rabbit | *N* = 3; 4-h application under semiocclusive binder to abraded and nonabraded skin; irritation scores at ½, 1, 24, 48, and 72 h | 500 mg/site | Classify irritation potential | |
| Eye irritation | Rabbit | *N* = 3/group; into right eye, compare to untreated eye; test only compounds with pH <12 or >2; score at 1, 24, 48, and 72 h and up to 21 days for corneal opacity, conjunctivitis, iritis: may use a rinse with some animals | 100 mg or 0.1 mL in standard protocol or 10 mg or 0.01 mL in low-volume protocol | Classify eye irritation potential; selective equipment | Evaluate first aid methods; ocular toxicity |
| Sensitization | | | | | |
| Buehler method | Guinea pigs | *N* = 10–15; topical applications 1–3 times per week for 3 weeks, 2-week rest, then challenge at a naive site; concurrent negative/vehicle and positive controls | Up to 500 mg or 0.5 mL per dose | Classify as sensitizer—most sensitive for moderate to strong sensitizers; selection of protective equipment; evaluation of allergic reactions | Repeated dermal dosing; additional data on skin irritation and dermal absorption |
| Maximization | Guinea pigs | *N* = 10–15; combines 2 intradermal ± adjuvant and 1 topical occlusive administration for induction, 2-week rest, then topical application for challenge at naive site; concurrent positive and negative/vehicle controls | 0.1 mL of compound in solution for induction; nonirritating concentration for challenge | May be sensitive to mild sensitizing agents but may also overpredict severity | |
| Local lymph node | Mouse (female) | *N* = 5; combines dermal exposure on ear with IV challenge by tail vein on day 6 with tritiated methyl thymidine | 25 μL of compound in solvent on each ear for 3 days | Does produce some false positives | No rechallenge possible |

to similar parameters developed in clinical or preclinical pharmacokinetic studies. The need to conduct additional testing will depend on the application of the information by the individual toxicologist and the resources available. The cost of additional tests should be weighed against the cost of applying the most conservative interpretation of the data to the work environment. In some cases implementing stricter controls based on preliminary tests may be less costly than conducting more extensive confirmatory testing.

There has recently been increasing pressure from governmental agencies and animal rights advocates to reduce the number of animals used in toxicological testing. As alternative toxicological methods become more accurate and sophisticated, they should be considered for incorporation into the occupational toxicology battery. Additional tools such as computer-aided quantitative structure–activity relationship (QSAR) evaluations may also be considered as additions or alternatives to animal tests (Jurs et al., 1985; Klopman, 1985; Frierson et al., 1986; Enslein, 1988). As indicated previously, QSAR methods may be particularly well suited to aid in the selection and/or prioritization of chemicals for testing, particularly in the case of intermediates. Alternative test methods currently under investigation, such as those being proposed for replacement of the Draize eye irritation test, do not appear to be well suited to the testing of pharmaceuticals or their synthetic intermediates (Booman et al., 1988, 1989). An intensive program of testing the available alternative models with compounds in this class is required to determine the ultimate usefulness of these alternative testing methods.

Occupational toxicologists from several companies supported a program to evaluate several experimental models as alternatives to the rabbit eye irritation test (Gáler et al., 1993; Sina et al., 1994). As a result of this cooperative study, several of the participating companies have implemented the routine use of several of these alternative models in their test batteries (R. G. Sussman and J. Sina, personal communications), thus effectively increasing the number and classes of compounds evaluated in alternative models.

### 5.3 Hazard Assessment

This is quite possibly the most difficult and controversial activity for the occupational toxicologist. Just as there is no blueprint for conducting toxicological testing, there is no formula for performing an occupational risk assessment. I have indicated above the need to include an indication of the degree of hazard from occupational exposure in informational communications. A more specific form of hazard assessment is the development of occupational exposure limits (OELs). Occupational exposure limits, such as the American Conference of Governmental Industrial Hygienists (ACGIH) TLVs, have been available for approximately 50 years (Cook, 1987). Similar workplace limits developed by OSHA are known as PELs (OSHA, 1989). The majority of the substances for which TLVs or PELs have been developed are large-volume industrial chemicals, often encountered in the workplace as vapors. The process of developing

TLVs and PELs has been documented by ACGIH and OSHA, respectively (OSHA, 1987; ACGIH, 1995). Unfortunately, the process is not always consistent or straightforward. Some TLVs or PELs are intended to prevent chronic effects yet are developed using acute reactions as reference points. Others are based on an existing TLV developed for a third compound. Clearly, these are not examples that can be easily followed by the pharmaceutical industry.

Individual companies have, nonetheless, developed methodologies for developing OELs based on the type of data and other resources available. One method involves a formula for extrapolating to an 8-h time-weighted average from the therapeutic dose of the drug using safety factors (Sargent and Kirk, 1988). A group composed of occupational toxicologists from several companies presented a monograph at the second annual Occupational Toxicology Roundtable, held in November 1989, regarding the development of OELs (Gáler et al., 1989, 1992). Several methods are available for developing OELs: analogy, correlation, safety and uncertainty factors, and low-dose extrapolation (Table 5). The appropriate method must be selected on the basis of the appropriateness of the available data. For example, low-dose extrapolation may be appropriate only if sufficient pharmacokinetic data are available to build a suitable physiologically based pharmacokinetic (PBPK) model. Analogy is a method by which the OEL for one compound is adopted for a second compound, based on the two compounds' structural and functional similarity. This method is suitable only if the two compounds are similar in every aspect, including therapeutic or toxic dose and physical properties. Correlation is similar to analogy in that it compares similar compounds. However, the OEL is chosen based on a key property of the chemical that influences its toxicological properties. An example would be the use of the relative potency of two drugs as the key property used to adjust the reference OEL. The most commonly used method is that of applying safety and/or uncertainty factors to a reference dose, which may be the lowest therapeutic dose. In using this method, it is important to choose the reference dose and endpoint with great care. In general, the most sensitive endpoint should be chosen. Uncertainty factors are selected to approximate levels from effective doses and to account

**TABLE 5 Methods for Setting Occupational Exposure Limits**

| Method | Formula[a] |
|---|---|
| Analogy | $OEL_i = OEL_j$ |
| Correlation | $OEL_j = (PP_i/PP_j) \times OEL_j$ |
| Safety factors | $OEL = \text{reference dose}/UF_1 \times UF_2 \times SF \times BR$ |
| Low-dose extrapolation | $OEL = [\text{rodent RSD} \times (BW_H/BW_R)^{-1/3}]/BR$ |
| | Or, if PBPK available, |
| | OEL = (human reference dose)/BR |

[a]OEL, occupational exposure limit; PP, physical property; $UF_1$, uncertainty in extrapolation to NOEL; $UF_2$, uncertainty from interspecies extrapolation; SF, safety factor; BR, breathing rate for 8-h workday; PBPK, physiologically based pharmacokinetic model; RSD, risk-specific dose; BW, body weight.

*Source*: Adapted from Gáler et al., 1989.

for interspecies differences. A safety factor is selected based on the overall toxicological evaluation of the compound. Because it is necessary to look at the complete toxicological profile, it is generally inappropriate to assign specific values to each type of toxic effect.

Most commonly employed now is the approach suggested initially by Sargent and Kirks in 1988, where the OEL for airborne pharmaceutical materials is calculated as

$$\mathrm{OEL} = \frac{\mathrm{NOAEL\ (mg/kg)(BW\ in\ kg)}}{(\mathrm{UF})(V\ \mathrm{in\ nm}^3)(\propto)(s)}$$

where NOAEL is the no effect level in the most sensitive species, BW is body weight, UF is an uncertainty factor (usually 10), $V$ is the volume of air breathed by a worker in 8 h (usually 10 m is used), $\propto$ is an adjustment for bioavailability between routes (if the animal data are from a route other than inhalation), and $s$ is for already known plasma levels if known (Binks, 2003). The FDA HED (human equivalent dose) factor is commonly used in place of plasma level if the steady state is unknown.

The OEL process has essentially become an industry standard. While U.S. OSHA regulations do not require that manufacturers establish OELs, European governmental agencies do. The first country to require this activity is the United Kingdom, under its Control of Substances Hazardous to Health (COSHH) regulations (Health and Safety Executive, 1988; Agius, 1989). A similar requirement is included in the EEC directive 80/1107, which was promulgated into law by other EC member nations.

In creating a program for establishing OELs, several disciplines will generally be included in the development or approval process. In those programs that are currently in place, safety, industrial hygiene, manufacturing or technical services, medical services, legal services, R&D, and, of course, occupational toxicology may take part in the process. While the ability to make the OEL level in the workplace does not drive the process, the OEL may often be issued as an interim guideline to provide manufacturing locations an opportunity to bring their operations into compliance and for the development of a suitable industrial hygiene sampling and analysis method.

### 5.4 Employee Training

The ultimate client for the services of the occupational toxicologist is the manufacturing or research employee who must be informed of and protected from the potential hazards of chemicals present in the work environment. Providing employees with health hazard information directly through presentations or training programs accomplishes this task better than most written communications and also provides an excellent way to build confidence in the organization and its safety and health programs. The trust gained in this manner can be an invaluable asset when a company is challenged with the

manufacture of especially toxic or potent compounds. Face-to-face communication will also promote discussions with line employees and give the occupational toxicologist the opportunity to learn of those adverse health effects that might otherwise go unnoticed or uninvestigated.

There are several areas for which it may be useful to consider developing specific training programs. The OSHA hazard communication standard requires that employees be trained to understand the hazards of chemicals as they are outlined in the MSDS. There is an obvious need for the occupational toxicologist to be involved in the development of an internal training program or the selection of a commercial program to address this need. In addition to this required training, it may be useful to consider a more in-depth program on basic concepts involved in health hazard evaluation, particularly the dose–response relationship and the different types of chronic health hazards. It may be particularly important to promote an understanding of health hazard information obtained at work as well as through the news media. There are several commercial training programs available that may be useful for this purpose, including computer-based self-training programs and videos.

Table 6 presents the currently most commonly used classification system for pharmaceutical occupational risks. Based on the assigned categorization from this risk, appropriate worker protection methods are selected (Olsen et al., 2002; Binks, 2003).

As most pharmaceutical manufacturing is now performed on contract (in "toll" manufacturers), categorization is usually performed on contract (by a firm such as a safe bridge) or by a consultant.

Specific training on compounds of interest can also be useful, particularly before the beginning of a manufacturing campaign, and is particularly effective if coupled with industrial hygiene training on appropriate safe handling techniques. If a testing program is in place, it is good policy to present an evaluation of the information gained in the compound's testing program to the research or manufacturing chemists involved in its manufacture.

Issues continuing to gain in importance for pharmaceutical industry occupational toxicologists are those relating to the new, more potent drugs currently being designed. Current drugs have therapeutic dose levels ranging down to the submicrogram (and in some cases even nanogram) levels. Hazard assessment and OEL development, already difficult, may be nearly impossible or inappropriate when dealing with drugs active at pico- or femtogram levels. Alternative methods of evaluating occupational exposures and assuring a safe work environment may need to be developed. Then there are biotechnology products and their associated issues.

A third of all new drugs come from the the biotechnology pipeline, most commonly peptides and proteins with significant allergenic potential in an occupational setting. The potential occupational health hazards of this class of potent but high-molecular-weight products have not been fully evaluated. Because of the inherent functional and structural differences, the extrapolation of testing methods from traditional pharmaceutical products to

**TABLE 6 Categorization Code Enrollment Criteria**

| Criteria | Categorization Code | | | | |
|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 |
| OEL | >1 mg m$^{-3}$ | 0.10 to <1 mg m$^{-3}$ | 10 to <100 μg m$^{-3}$ | 1 to <10 μg m$^{-3}$ | <1 μg m$^{-3}$ |
| Potency (mg day$^{-1}$) | >100 low | 100–10 low/moderate | 10–0.01 moderate | 0.1–0.01 moderate/ high | <0.01 high |
| Acute toxicity | Slightly toxic | Moderately toxic | Highly toxic | Extremely toxic | Supertoxic |
| Severity of acute effects | Low | Low/moderate | Moderate | Moderate high | High |
| Chronic toxicity | None | None | Slight to moderate | Moderate | Severe |
| Absorption by inhalation/skin | Minimal inhalation or skin adsorption | Moderate inhalation or skin absorption | Significant inhalation or skin absorption | Significant inhalation or skin absorption | Significant inhalation or skin absorption |
| Irritation potential | | | | | |
| Skin | None | Mild to moderate | Moderate to severe | Severe to extreme | Extreme |
| Eye | Mild to moderate | Mild to moderate | Severe | Severe to extreme | Extreme |
| Sensitization | Not a sensitizer | Mild sensitizer | Moderate sensitizer | Strong sensitizer | Extreme sensitizer |
| Mutagenicity | None | Yes | Yes | Severe | Severe |
| Carcinogenicity | Negative | Suspected | Suspected–confirmed animal | Defined medical case studies | Defined medical case studies |
| Reproductive disorders | None | None | Slight | Moderate | Known |
| Teratogenicity | None | None | Yes | Severe | Severe |
| Neurotoxicity | Minimal | Minimal to moderate | Moderate to immediate and delayed | Life threatening | Life threatening |
| Epidemiology | Confirmed, no effect | Case studies/ ongoing definitive data | Defined medical case studies | Defined medical case studies | Defined medical case studies |
| Genotoxicity | None | None –(+) Ames test | (+) In battery of genotox studies | (+) In a battery of genotox studies | (+) In a battery of genotox studies |
| Warning properties | Good | Fair | Fair/poor | Poor | None |

biotechnology-derived compounds may be fraught with many difficulties. Hypersensitivity and other immunologically based toxicities are particularly of concern for protein- and peptide-based therapeutics.

## 6 CONCLUSION

The field of occupational toxicology in the pharmaceutical industry presents continuing challenges to the industry, particularly as it shifts increasingly to an outsourced function in companies. The occupational toxicologist finds that he or she must become an "expert" in several fields and not be limited to the scientific area. Unlike the preclinical toxicologist, the occupational practitioner functions under less stringent regulatory requirements and minimal precedents. Additionally, as new classes of therapeutic agents enter development and commerce, new concerns and challenges will accompany them.

## REFERENCES

Agius, R. (1989). Occupational exposure limits for therapeutic substances. *Ann. Occup. Hyg.* 33:555–562.

Alarie, Y. (1966). Irritating properties of airborne materials to the upper respiratory tract. *Arch. Environ. Health* 13:433–449.

Alarie, Y. (1981). Bioassay for evaluating the potency of airborne sensory irritants and predicting acceptable levels of exposure in man. *Food Cosmet. Toxic.* 19:623–626.

American Conference of Governmental Industrial Hygienists (ACGIH) (1995). *Documentation of Threshold Limit Values and Biological Exposure Indices*, 6th ed. With 1996, 1997, and 1998 supplements. ACGIH, Cincinnati, OH.

Binks, S. P. (2003). Occupational toxicology and the control of exposure to pharmaceutical agents at work. *Occup. Med.* 53:363–370.

Booman, K. A., Cascieri, T. M., Demetrulias, J., Diedger, A., Griffith, J. F., Grochosky, G. T., Kong, B., McCormick, W. C., North-Root, H., Rozen, M. G., and Sedlak, R. I. (1988). *In vitro* methods for estimating eye irritancy of cleaning products. Phase I: Preliminary assessment. *J. Toxicol. Cut. Ocul. Toxicol.* 7:173–185.

Booman, K. A., DeProspo, J., Demetrulias, J., Diedger, A., Griffith, J. F., Grochosky, G., Kong, B., McCormick, W. C., North-Root, Hl, Rozen, M. G., and Sedlak, R. I. (1989). The SDA alternatives program: Comparison of *in vitro* data with Draize test data. *J. Toxicol. Cut. Ocul. Toxicol.* 8:35–49.

Cook, W. A. (1987). *Occupational Exposure Limits—Worldwide*. American Industrial Hygiene Association, Cincinnati, OH.

Enslein, K. (1988). An overview of structure-activity relationships as an alternative to testing in animals for carcinogenicity, mutagenicity, dermal and eye irritation and acute oral toxicity. *Toxicol. Indust. Health* 4:479–498.

European Economic Community (EEC) (1979). Council Directive 67/548/EEC as amended by Directive 79/831/EEC. *Official J. Eur. Commun.* No. L259, 10.

European Economic Community (EEC) (1992). Council Directive 92/32/EC of November 27, 1980, on the protection of workers from the risks related to exposure to chemical, physical, and biological agents at work. *Official J. Eur. Commun.* No. L327, 8.

Frierson, M. R., Klopman, G., and Rosenkranz, H. S. (1986). Structure-activity relationships (SARs) among mutagens and carcinogens: A review. *Environ. Mutagen.* 8:283–327.

Gad, S. C. (2001). *Regulatory Toxicology*, 2nd ed. Taylor & Francis, Philadelphia, PA.

Gad, S. C., and Chengelis, C. P. (1998). *Acute Toxicology Testing*, 2nd ed. Academic, San Diego, CA.

Gáler, D. M. (1989). A testing battery for evaluating occupational health hazards. *J. Am. Coll. Toxicol.* 8:1215.

Gáler, D. M., Curren, R., Gad, S. C., Gautheron, P., Leong, B., Miller, K., Sargent, E., Shah, P. V., Sina, J., and Sussman, R. G. (1993). A 10-company collaborative evaluation of alternatives to the eye irritation test using chemical intermediates. *Alternative Methods Toxicol.* 9:237.

Gáler, D. M., Leung, H.-W., Sussman, R. G., and Trzos, R. J. (1989). Scientific and practical considerations for developing occupational exposure limits for chemical substances. Presentation to the second annual Occupational Toxicology Roundtable, November 21, 1989, Rahway, NJ.

Gáler, D. M., Leung, H. W., Sussman, R. G., and Trzos, R. J. (1992). Scientific and practical considerations for the development of occupational exposure limits (OELs) for chemical substances. *Regul. Toxicol. Pharmacol.* 15:291–306.

Griffith, J. F., and Yam, J. (1989). The low-volume eye irritation test. Current status of use and acceptance. *J. Am. Coll. Toxicol.* 8:1215.

Hatoum, N. S., Leach, C. L., Talmsa, D. M., Gibbons, R. D., and Garvin, P. F. (1990). A statistical basis for using fewer rabbits in dermal irritation testing. *J. Am. Coll. Toxicol.* 9:49–59.

Health and Safety Executive, U.K. (1988). The control of substances hazardous to health regulations 1988. *Statutory Instruments* 1988, No. 1657.

Jurs, P. C., Stouch, T. P., Czerwinski, M., and Narvaez, J. N. (1985). Computer-assisted studies of molecular structure-activity relationships. *J. Chem. Inf. Comput. Sci.* 25:296–308.

Kane, L. E., Barrow, C. S., and Alarie, Y. (1979). A short-term test to predict acceptable levels of exposure to airborne sensory irritants. *Am. Ind. Hyg. J.* 40:207–229.

Klopman, G. (1985). Predicting toxicity through a computer automated structure evaluation program. *Environ. Health Perspect.* 61:269–274.

Occupational Safety and Health Administration (OSHA) (1986). Administrative Rule concerning OSHA access to employee medical records. 29 CFR Part 1913.

Occupational Safety and Health Administration (OSHA) (1987). Hazard communication standard. 29 CFR Part 1910, Subpart Z, Section 1910.1200.

Olsen, M. J., Slaman, C. W., and Guerieo, F. J. (2002). An integrated approach to hazard evaluation for worker safety assessment of pharmaceuticals and synthetic intermediates. *Toxicol. Sci.* 66:328–340.

Sargent, E. V., and Kirk, G. D. (1988). Establishing airborne exposure control limits in the pharmaceutical industry. *Am. Ind. Hyg. Assoc. J.* 49:309–313.

Sina, J. F., Gáler, D. M., Sussman, R. G., Gautheron, P. D., Sargent, E. V., Leong, B., Shah, P. V., Curren, R., and Miller, K. (1995). A collaborative evaluation of seven alternatives to the Draize eye irritations test using pharmaceutical intermediates. *Fund. Appl. Toxicol*. 26:20–31.

Society and Toxicology (1989). SOT position paper, comments on the $LD_{50}$ and acute eye and skin irritation tests. *Fund. Appl. Toxicol.* 13:621–623.

Sussman, R., and Gáler, D. (1990). A database for communicating toxicology information to safety personnel. *Toxicologist* 10:87.

Teichman, R. F., Fallon, L. F., and Brandt-Rauf, P. (1988). Health effects of workers in the pharmaceutical industry: A review. *J. Soc. Occup. Med.* 38:55–57.

U.S. Environmental Protection Agency (EPA) (1976). Toxic Substances Control Act, Public Law 94-469, 94th Congress.

U.S. Environmental Protection Agency (EPA) (1979). Toxic Substances Control Act premanufacture testing of new chemical substances. Guidance for premanufacture testing: Discussion of policy issues, alternative approaches and tests methods. *Fed. Reg.* 44:16240–16292.

U.S. Food and Drug Administration (FDA) (1990). Applications for FDA approval to market a new drug or an antibiotic drug. 21 CFR Part 314.

# *Strategy and Phasing for Nonclinical Drug Safety Evaluation in Discovery and Development of Pharmaceuticals*

## 1 INTRODUCTION

The preclinical assessment of the safety of potential new pharmaceuticals represents a special case of the general practice of nonclinical safety assessment (Gad, 1996, 2000; Meyer, 1989) possessing its own peculiarities and special considerations and differing in several ways from the practice of toxicology in other fields—for some significant reasons. Because of the economics involved and the essential close interactions with other activities (e.g., clinical trials, chemical process optimization, formulation development, regulatory reviews), the development and execution of a crisp, timely, and flexible, yet scientifically sound, program are prerequisites for success. The ultimate aim of preclinical assessment also makes it different. A good pharmaceutical safety assessment program seeks to efficiently and effectively move safe, potential therapeutic agents into, and support them through, the clinical evaluation, then to registration, and, finally, to market. This requires the quick identification of those agents that are not safe. At the same time, the very biological activity which makes a drug efficacious also acts to complicate the design and interpretation of safety studies.

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

Such evaluations occur on different time scales because of different objectives. There is the traditional big pharma case and different versions used predominantly by the much more numerous small pharmaceutical cases. These small pharma cases may be either the short case (do only what is required to initiate and support initial clinical trials) or the midgame (a variety where studies to support further clinical trials are performed but spread out).

Pharmaceuticals, unlike industrial chemicals, agricultural chemicals, and environmental agents, are intended to have human exposure and biological activity. And, unlike these materials and food additives, pharmaceuticals are intended to have biological effects on the people that receive them. Frequently, the interpretation of results and the formulation of decisions about the continued development and eventual use of a drug are based on an understanding of both the potential adverse effects of the agent (its safety) and its likely benefits as well as the dose separation between these two (the "therapeutic index"). This makes a clear understanding of dose–response relationships critical, so that the actual risk–benefit ratio can be identified. It is also essential that the pharmacokinetics be understood and that "doses" (plasma tissue levels) at target organ sites are known (Scheuplein et al., 1990). Integral evaluation of pharmacokinetics is essential to any effective safety evaluation program.

The development and safety evaluation of pharmaceuticals have many aspects specified by regulatory agencies, and this has also tended to make the process more complex [until recently, as the International Conference on Harmonisation (ICH) has tended to take hold] as markets have truly become global. An extensive set of safety evaluations is absolutely required before a product is ever approved for market. There are even novels on the subject (see Zbinden, 1992). Regulatory agencies have increasingly come to require not only the establishment of a "clean dose" in two species with adequate safety factors to cover potential differences between species but also an elucidation of the mechanisms underlying such adverse effects as are seen at higher doses and are not well understood. These regulatory requirements are compelling for the pharmaceutical toxicologist (Traina, 1983; Guarino, 1987; Smith, 1992). There was not, however, a set menu of what must be done. Rather, much (particularly in terms of the timing of testing) is open to professional judgment and is tailored for the specific agent involved and its therapeutic claim. ICH M3(R2) guidance (ICH, 2009) has acted to improve this situation significantly, but significant "regulatory drift" leading to the need for experience, judgment, and interaction with FDA remains.

The discovery, development, and registration of a pharmaceutical is an immensely expensive operation and represents a rather unique challenge. For every 9000–10,000 compounds specifically synthesized or isolated as potential therapeutics, one (on average) will actually reach the market. This process is illustrated diagrammatically in Figure 1. Each successive stage in the process is more expensive, making it of great interest to identify as early as possible those agents that are likely to not go the entire distance, allowing a concentration of effort on the compounds that have the highest probability of

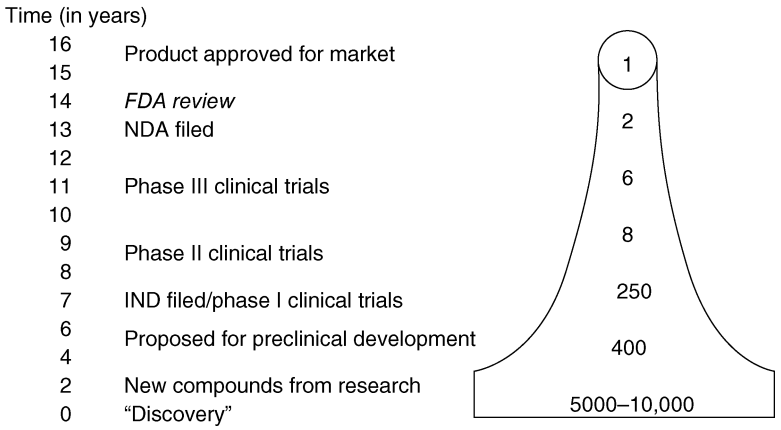


***Figure 1*** *Attrition during development of new molecules with promise of therapeutic potential. Over the course of taking a new molecular entity through scaleup, safety and efficacy testing, and, finally, to market typically only one out of every 9000–10,000 will go to the marketplace.*

reaching the market. Compounds "drop out" of the process primarily for three reasons:

1. Toxicity (or lack of toxicity tolerance)
2. (Lack of) efficacy
3. (Lack of) bioavailability of the active moiety in humans.

Early identification of poor or noncompetitive candidates in each of these three categories is thus extremely important (Fishlock, 1990), forming the basis for the use of screening in pharmaceutical discovery and development. How much and which resources to invest in screening and each successive step in support of the development of a potential drug are matters of strategy and phasing that are detailed in Section 5 of this chapter. In vitro methods are increasingly providing new tools for use in both early screening and the understanding of mechanisms of observed toxicity in preclinical and clinical studies (Gad, 1989b, 2001), particularly with the growing capabilities and influence of genomic and proteomic technologies. This is increasingly important as the societal concern over drug prices has grown (Littlehales, 1999). Additionally, the marketplace for new drugs is exceedingly competitive. The rewards for either being introduced early (first or second) into the marketplace ("first movers advantage") or achieving a significant therapeutic advantage are enormous in terms of eventual market share. Additionally, the first drug approved sets agency expectations for those drugs which follow. In mid-2008, there were 254 pharmaceutical products awaiting approval (86 of these biotechnological products)—the "oldest" having been in review nine years) and some 16,000 additional agents in the investigational new drug (IND) stage [U.S. Food and

Drug Administration (FDA) website]. Not all of these (particularly the oldest) will be economically successful.

The successful operation of a safety assessment program in the pharmaceutical industry requires that four different phases of the product-related operation be simultaneously supported. These four phases of pharmaceutical product support (discovery support, IND support, clinical and registration support, and product support) constitute the vast majority of what is done by the safety assessment groups in the pharmaceutical industry. The constant adjustment of balance of resources between these four areas is the greatest management challenge in pharmaceutical safety assessment. An additional area, occupational toxicology, is conducted in a manner similar to that for industrial environments and is the subject of Chapter Irritation and Local Tissue Tolerance in Pharmaceutical Safety Assessment of this volume. In most companies, occupational toxicology is the responsibility of a separate group.

The usual way in which transition (or "flow") between the different phases is handled in safety assessment is to use a tiered testing approach. Each tier generates more specific data (and costs more to do so) and draws on the information generated in earlier tiers to refine the design of new studies. Different tiers are keyed to the support of successive decision points (go/no-go points) in the development process, with the intent of reducing risks as early as possible.

The first real critical decisions concerning the potential use of a compound in humans are the most difficult. They require an understanding of how well particular animal models work in predicting adverse effects in humans (usually very well, but there are notable lapses; e.g., giving false positives and false negatives) and an understanding of what initial clinical trials are intended to do. Though an approved IND grants one entry into limited evaluations of drug effects in humans, flexibility in the execution and analysis of these studies offers a significant opportunity to also investigate efficacy (O'Grady and Linet, 1990).

Once past the discovery and initial development stages, the safety assessment aspects of the process become extremely tightly connected with the other aspects of the development of a compound, particularly the clinical aspects. These interconnections are coordinated by project management systems. Many times during the early years of the development process, safety assessment constitutes the rate-limiting step—it is, in the language of project management, on the critical path.

Another way in which pharmaceutical safety assessment varies from toxicology as practiced in other in industries is that it is a much more multidisciplinary and integrated process. This particularly stands out in the incorporation of the evaluation of ADME (absorption, distribution, metabolism, and excretion) aspects in the safety evaluation process. These pharmacokinetic/metabolism (PKM) aspects are evaluated for each of the animal model species (most commonly the rat and dog or primate) utilized to evaluate the preclinical systemic toxicity of a potential drug prior to evaluation in humans. Frequently, in vitro characterizations of metabolism for model (or potential model) species and humans are performed to allow optimal model selection and understanding of findings. This allows for an early appreciation of both the potential bioavail-

ability of active drug moieties and the relative predictive values of the various animal models. Such data early on are also very useful (in fact, sometimes essential) in setting dose levels for later animal studies and in projecting safe dose levels for clinical use. Unlike the case in most other areas of industrial toxicology, one is not limited to extrapolating the relationships between administered dose and systemic effects. Rather, one has significant information on systemic levels of the therapeutic moiety—typically, total area under the curve (AUC), peak plasma levels ($C_{max}$), and plasma half-lives at a minimum. Chapter Safety Assessment of Inhalant and Dermal Route Drugs looks at these aspects in detail.

The state of the art for preclinical safety assessment has now developed to the point where the resulting products of the effort [reports, IND/new drug application (NDA) summaries, and the overall professional assessment of them] are expected to reflect and integrate the best effort of all the available scientific disciplines. Actual data and discussion should thus come from toxicology, pharmacology, pathology, and metabolism at a minimum. The success of current premarket efforts to develop and ensure that only safe drugs make it to market are generally good but clearly not perfect. This is reflected in popular (Arnst, 1998; Raeburn, 1999) and professional (Moore et al., 1998; Lazaron et al., 1998) articles looking at both the number of recent marketed drug withdrawals for safety (summarized in Table 1) and rates of drug-related adverse drug events and deaths in hospital patients. It is hoped that this system can be improved, and there are continuing efforts to improve or optimize drug candidate selection and development (Lesko et al., 2000). Indeed, the entire pharmaceutical development paradigm is clearly in need of a complete change—a synthesis, as opposed to a continual application of small corrections.

## 2 REGULATORY REQUIREMENTS

Minimum standards and requirements for safety assessment of new pharmaceuticals are established by the need to meet regulatory requirements for developing and eventually gaining approval to market the agent. Determining what these requirements are is complicated by (1) the need to compete in a global market, which means gaining regulatory approval in multiple countries that do not have the same standards or requirements, and (2) the fact that the requirements are documented as guidelines, the interpretation of which is subject to change as experience alters judgments. The ICH process has much improved this situation, as detailed in Chapter Regulation of Human Pharmaceutical Safety. The ICH (M3) (2000) clearly denotes what nonclinical studies are required to support clinical drug development. Unfortunately, since then additional requirements (safety pharmacology and immunotoxicology) have been promulgated by additional guidelines. Accordingly, M3 is not all current or inclusive in its guidance.

Standards for the performance of studies (which is one part of regulatory requirements) have as their most important component good laboratory practices (GLPs). Good laboratory practices largely dictate the logistics of

**TABLE 1 Post-approval Adverse Side Effects and Related Drug Withdrawals Since 1990**

| Year | Drug | Indication/Class | Causative Side Effect |
|---|---|---|---|
| 1991 | Enkaid (4 years on market) | Antiarrhythmic | Cardiovascular (sudden cardiac death) |
| 1992 | Temafloxacin | Antibiotic | Blood and kidney |
| 1997 | Fenfuramine[a]/ dexafluramine (combo used since 1984) | Diet pill | Heart valve abnormalities |
| 1998 | Posicor (Midefradil) (1 year on market) | $Ca^{2+}$ channel blocker | Lethal drug interactions (inhibited liver enzymes) |
| | Duract (Bronfemic Sodium) (early preapproval warnings of elevated liver enzymes) | Pain relief | Liver damage |
| 1999 | Tronan (use severely restricted) | Antibiotic | Liver/kidney damage |
| | Raxar | Quinolone antibiotic | QT internal prolongation/ ventricular arrhythmias (deaths) |
| | Hismanal | Antihistamine | Drug–drug interactions |
| | Rotashield | Rotavirus vaccine | Bowel obstruction |
| 2000 | Renzulin (approved Dec. 1996) | Type II diabetes | Liver damage |
| | Propulsid | Heartburn | Cardiovascular irregularities/death |
| | Lotonex | Irritable bowel syndrome | Ischemic colitis/death |
| 2001 | Phenylpropanolamine (PPA) | OTC ingredient | Hemorrhagic stroke |
| | Baychlor | Cholesterol reducing (statin) | Rhabdomyolysis (muscle weakening) (deaths) |
| 2002/2003 | None | | |
| 2004 | Serazone | Antidepressant | Liver failure and injury |
| | Vioxx | Arthritis (CoOX-2 inhibitor) | Heart attack/ cardiovascular (thrombosis) |
| 2005 | Tysabri | Multiple sclerosis (MS) | Progressive multifocal leukoencephalopathy (PML) |
| | Bextra | Arthritis (CoOX-2 inhibitor) | Skin reaction (sometimes fatal) |
| 2006 | Dolophine (methadone hydrochloride) | Treatment of moderate to severe pain | Respiratory depression and cardiac arrhythmias |
| 2007 | Zelnorm | Constipation | Cardiovascular safety |
| | Permax | Parkinson's disease | Heart valve damage |

*Note*: 51% of approval drugs had serious postapproval identified side effects. FDAMA (Food and Drug Administration Modernization Act) passed in 1997.
[a]Twenty four years on market.

safety assessment—training, adherence to other regulations (such as those governing the requirements for animal care), and (most of all) the documentation and recordkeeping that are involved in the process. There are multiple sets of GLP regulations [in the United States alone, agencies such as the FDA and the Environmental Protection Agency (EPA) each have their own] that are not identical; however, adherence to FDA (1987a) GLPs will rarely lead one astray.

Not all studies that are done to assess the preclinical safety of a new pharmaceutical need be done in strict adherence to GLPs. Those studies that are "meant to support the safety of a new agent" (FDA, 1987a) (i.e., are *required* by regulatory guidelines) must be so conducted or run a significant risk of rejection. However, these are also many other studies of an exploratory nature (such as range finders and studies done to understand the mechanisms of toxicity) that are not required by the FDA and which may be done without strict adherence to GLPs. A common example is those studies performed early on to support research in selecting candidate agents. Such studies do not meet the requirements for having a validated analytical method to verify the identity, composition, and stability of materials being assayed, yet they are essential to the processes of discovery and development of new drugs. All such studies must eventually be reported to the FDA if an IND application is filed, but the FDA does not in practice "reject" such studies (and therefore the IND) because they are "non-GLP."

There is a second set of "standards" of study conduct that are less well defined. These are "generally accepted practice" and, though not written down in a regulation, are just as important as GLPs for studies to be accepted by the FDA and the scientific community. These standards, which are set by what is generally accepted as good science by the scientific community, include techniques, instruments utilized, and interpretation of results. Most of the chapters in this book will reflect these generally accepted practices in one form or another.

Guidelines establish which studies must be done for each step in the process of development. Though guidelines supposedly are suggestion (and not requirements), they are in fact generally treated as minimums by the promulgating agency. The exceptions to this are special cases where a drug is to meet some significant need [a life-threatening disease such as AIDS or amyotrophic lateral sclerosis (ALS)] or where there are real technological limitations as to what can be done [as with many of the new biologically derived (or biotechnology) agents, where limitations on compound availability and biological responses make traditional approaches inappropriate].

There are some significant differences in guideline requirements between the major countries [see Alder and Zbinden (1988) for an excellent country-by-country review of requirements], though this source is now becoming dated. The core of what studies are generally done are those studies conducted to meet FDA requirements. These are presented in Table 2. As will be discussed in Chapter Regulation of Human Pharmaceutical Safety, these guidelines are giving way to the ICH guidelines. However, while the length and details of studies have changed, the nature and order of studies remain the same.

**TABLE 2 Synopsis of General Guidelines for Animal Toxicity Studies (U.S. FDA, *Total Drug Quality*)**

| Category | Duration of Human Administration[a] | Phase[b] | Subacute or Chronic Toxicity[c] | Special Studies |
|---|---|---|---|---|
| Oral or parenteral | Several days (up to 3) | I, II, III, NDA | 2 species: 2 weeks | For parenterally administered drugs; compatibility with blood and local tolerance at injection site where applicable |
| | Up to 2 weeks | I | 2 species: 4 weeks | |
| | | II | 2 species: up to 4 weeks | |
| | | III, NDA | 2 species: up to 3 months | |
| | Up to 3 months | I, II | 2 species: 4 weeks | |
| | | III | 2 species: 3 months | |
| | | NDA | 2 species: up to 6 months | |
| | 6 months to unlimited | I, II | 2 species: 3 months | |
| | | III | 2 species: 6 months or longer | |
| | | NDA | 2 species: 12 months in rodents, 9 months in nonrodents + 2 rodent species for CA; 18 months (mouse)—may be met by use of transgenic model 24 months (rat) | |
| Inhalation (general anesthetics) | Single administration | I, II, III, NDA | 4 species: 5 days (3 h/day) | |
| Dermal | Single application | I | 1 species: single 24-h exposure followed by 2-week observation | Sensitization |
| | Single or short-term application | II | 1 species: 20-day repeated exposure (intact and abraded skin) | |
| | Short-term application | III | As above | |
| | Unlimited application | NDA | As above, but intact skin study extended up to 6 months | |
| Ophthalmic | Single application | I | | Eye irritation tests with graded doses |
| | Multiple application | I, II, III | 1 species: 3 weeks, daily applications, as in clinical use | |
| | | NDA | 1 species: duration commensurate with period of drug administration | |

**TABLE 2** ***Continued***

| Category | Duration of Human Administration[a] | Phase[b] | Subacute or Chronic Toxicity[c] | Special Studies |
|---|---|---|---|---|
| Vaginal or rectal | Single application | I | | Local and systematic toxicity after vaginal or rectal application in 2 species |
| | Multiple application | I, II, III, NDA | 2 species: duration and no. of applications determined by proposed use | Lethality by appropriate route, compared to components run concurrently in 1 species |
| Drug combinations[d] | I, II, III, NDA | 2 species: up to 3 months | | |

[a]Phases I dosing of females if childbearing potential requires a Segment II study in at least one species; Phase III dosing of this population requires a Segment I study and both Segment II studies.
[b]Phases I, II, and III are defined in Section 130.0 of the New Drug Regulations.
[c]Acute toxicity should be determined in three species; subacute or chronic studies should be by the clinical route to be used. Suitable mutagenicity studies should also be performed.
Observations:
Body weights
Metabolic studies
Gross and microscopic examination
Coagulation tests
Food consumption
Ophthalmological examination
Hemogram
Others as appropriate
Behavior
Fasting blood sugar
Liver and kidney function tests
[d]Where toxicity data are available on each drug individually.
*Source*: FDA, *Total Drug Quality*, 1971.

The major variations in requirements for other countries still tend to be in the area of special studies. The United States does not formally require any genotoxicity studies, but common practice for U.S. drug registration is to perform at least a bacterial gene mutation assay (Ames test), a mammalian cell mutation assay, and a clastogenicity assay, while Japan requires specific tests, including a gene mutation assay in *Escherichia coli*. Likewise, the European Economic Community (EEC) has a specified set of requirements, while individual countries have additional special requirements (Italy, for example, requires a mutagenicity assay in yeast). As detailed in Chapter Single-Dose (Acute) and Pilot (DRF) Toxicity Testing in Drug Safety Evaluation, the new ICH genotoxicity guidelines have come to meet multinational requirements.

Japan maintains a special requirement for an antigenicity assay in guinea pigs. The new safety pharmacology requirements are likely to be adopted over a period of time by different adherents.

It is possible to interact with the various regulatory agencies (particularly the FDA) when peculiarities of science or technology leave one with an unclear understanding of what testing is required. It is best if such discussions directly involve the scientists who understand the problems, and it is essential that the scientists at the FDA be approached with a course of action (along with its rationale) that has been proposed to the agency in advance.

The actual submissions to a regulatory agency that request permission either to initiate (or advance) clinical trials of a drug or to market a drug are not just bundles of reports on studies. Rather, they take the form of summaries that meet mandated requirements for format accompanied by the reports discussed in these summaries. In the United States, these summaries are the appropriate section of the IND and NDA. The formats for these documents have recently been revised (FDA, 1987b). The EEC equivalent is the expert report, as presented in EEC Directive 75/319. Similar approaches are required by other countries. In each of these cases, textual summaries are accompanied by tables that also serve to summarize significant points study design and study findings.

All of these approaches have in common that they are to present integrated evaluations of the preclinical safety data that are available on a potential new drug. The individual studies and reports are to be tied together to present a single, cohesive overview of what is known about the safety of a drug.

Leber (1987) presents an excellent overview of the regulatory process involved in FDA oversight of drug development and gives the historical perspective for the evolution of the conservative process that is designed to ensure that any new pharmaceutical is both safe and efficacious.

There are other regulatory, legal, and ethical safety assessment requirements beyond those involved in the selection and marketing of a drug as a product entity. The actual drug product must be manufactured and transported in a safe manner and any waste associated with this manufacture disposed of properly. Chapter Irritation and Local Tissue Tolerance in Pharmaceutical Safety Assessment of this volume specifically addresses this often overlooked aspect of safety assessment programs.

## 3 ESSENTIAL ELEMENTS OF PROJECT MANAGEMENT

It is important to keep in mind that safety assessment is only one of many components involved in the discovery and development of new pharmaceuticals. The entire process has become enormously expensive, and completing the transit of a new drug from discovery to market has to be as efficient an expedition process as possible. Even the narrow part of this process (safety assessment) is dependent on many separate efforts. Compounds must be made, analytical and bioanalytical methods developed, and dosage formulations

developed, to name a few. One needs only to refer to Beyer (1978), Hamner (1982), Matoren (1984), Sneader (1986) (a good short overview), or Spilker (1994) for more details on this entire process and all of its components.

The coordination of this complex process is the province of project management, the objective of which is to ensure that all the necessary parts and components of a project match up. This discipline in its modern form was first developed for the Polaris missile project in the 1960s. Its major tool, which is familiar to pharmaceutical scientists, is the "network" or PERT (program evaluation review technique) chart, as illustrated in Figure 2. This chart is a tool that allows one to see and coordinate the relationships between the different components of a project. One outcome of the development of such a network is identification of the rate-limiting steps, which in aggregate comprise the critical path (see Table 3 for a lexicon of the terms used in project management).

A second graphic tool from project management is the Gantt chart, as illustrated in Figure 3. This chart allows one to visualize the efforts underway in any one area, such as safety assessment, for all projects that are currently being worked on.

Figure 4 is a hybrid from of the PERT and Gantt charts, designed to allow one to visualize all the resources involved in any one project.

An understanding of the key concepts of project management and their implications is critical in strategic planning and thinking for safety assessment. Kliem (1986) and Knutson (1980) offer excellent further reading in the area of project management.

## 4 SCREENS: THEIR USE AND INTERPRETATION IN SAFETY ASSESSMENT

Much (perhaps even most) of what is performed in safety assessment can be considered screening—trying to determine if some effect is or is not (to an acceptable level of confidence) present (Zbinden et al., 1984). The general concepts of such screens are familiar to toxicologists in the pharmaceutical industry because the approach is a major part of the activities of the pharmacologists involved in the discovery of new compounds. But the principles underlying screening are not generally well recognized or understood. And such understanding is essential to the proper use, design, and analysis of screens (Gad, 1988, 1989a). Screens are the biological equivalent of exploratory data analysis (EDA) (Tukey, 1977).

Each test or assay has an associated activity criterion, that is, a level above which the activity of interest is judged to be present. If the result for a particular test compound meets this criterion, the compound may pass to the next stage. This criterion could be based on statistical significance (e.g., all compounds with observed activities significantly greater than the control at the 5% level could be tagged). However, for early screens, such a formal criterion may be too strict, resulting in a few compounds begin identified as "active."

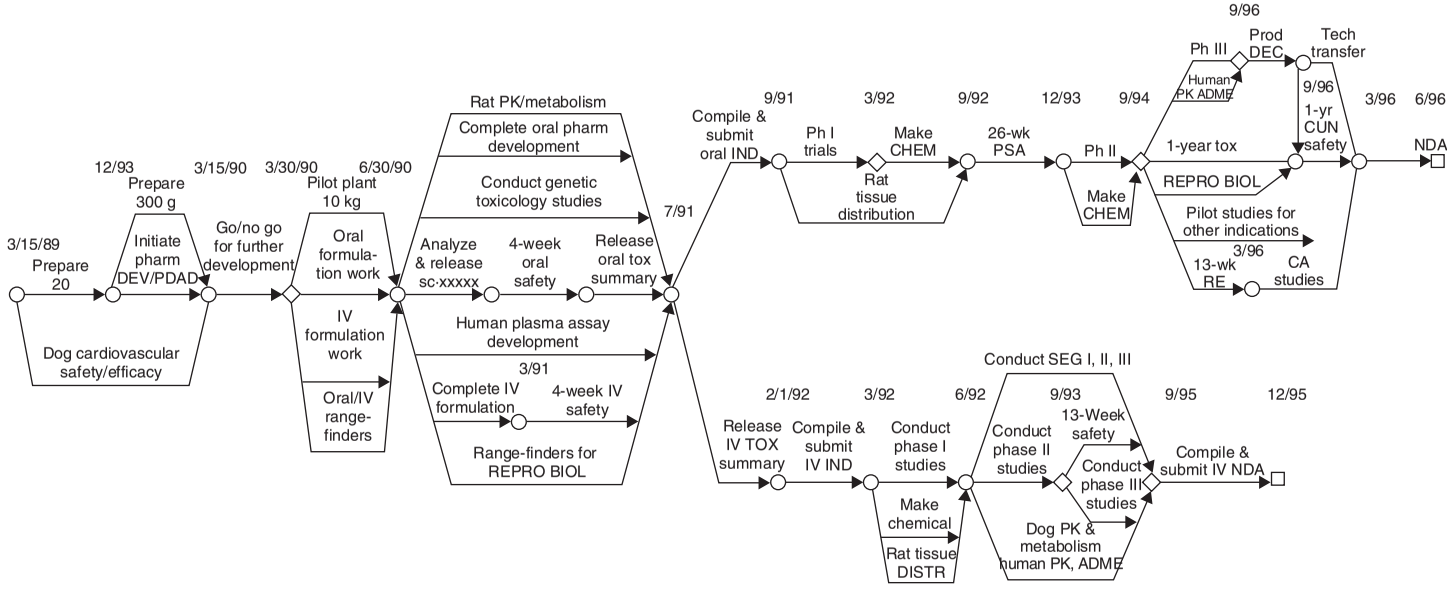


***Figure 2*** *An example PERT (program evaluation review technique) chart of the development of a new pharmaceutical through to the filing of an NDA (new drug application). Circles are "nodes" indicating completion of activities. Diamonds are initiation points for tasks that have starting points independent of others. This "network" serves to illustrate the relationships between different activities and to evaluate effects of changes on project timing.*

**TABLE 3 Glossary of Project Management Terms**

| | |
|---|---|
| Activity | The work or effort needed to complete a particular event. It consumes time and resources. |
| Average daily resource requirement | The likely amount of resources required to complete an activity or several activities on any workday during a project. The average daily labor requirement is one example. |
| CPM | Acronym for critical path method. A network diagramming technique that places emphasis on time, cost, and the completion of events. |
| Critical path | The longest route through a network that contains activities absolutely crucial to the completion of the project. |
| Dummy arrow | A dashed line indicating an activity that uses no time or resources. |
| Duration | The time it takes to complete an activity. |
| Earliest finish | The earliest time an activity can be completed. |
| Earliest start | The earliest an activity can begin if all activities before it are finished. It is the earliest time that an activity leaves its initiation node. |
| Event | A synonym for node. A point in time that indicates the accomplishment of a milestone. It consumes neither time nor resources and is indicated whenever two or more arrows intersect. |
| Free float | The amount of time that an activity can be delayed without affecting succeeding activities. |
| Gannt chart | A bar chart indicating the time interval for each of the major phases of a project. |
| Histogram | A synonym for bar chart. |
| Latest finish | The latest time an activity can be completed without extending the length of a project. |
| Latest start | The latest time an activity can begin without lengthening a project. |
| Leveling | The process of "smoothing" out labor, material, and equipment requirements to facilitate resource allocation. The project manager accomplishes this by "rescheduling" noncritical activities so that the total resource requirements for a particular day match the average daily resource requirements. |
| Most likely time | Used in PERT diagramming. The most realistic time estimate for completing an activity or project under normal conditions. |
| Node | A synonym for event. |
| Optimistic time | Used in PERT diagramming. The time the firm can complete an activity or project under the most ideal conditions. |
| PERT | Acronym for program evaluation and review technique. A network diagramming technique that places emphasis on the completion of events rather than cost or time. |
| Pessimistic time | Used in PERT diagramming. The time the firm can complete an activity or project under the worst conditions. |
| Project | The overall work or effort begin planned. It has only one beginning node and ending node. Between those nodes are countless activities and their respective nodes. |
| Project phase | A major component, or segment, of a project. It is determined by the process known as project breakdown structuring. |
| Total float | The total amount of flexibility in scheduling activities on a noncritical path. Hence, it provides the time an activity could be prolonged without extending a project's final completion date. |

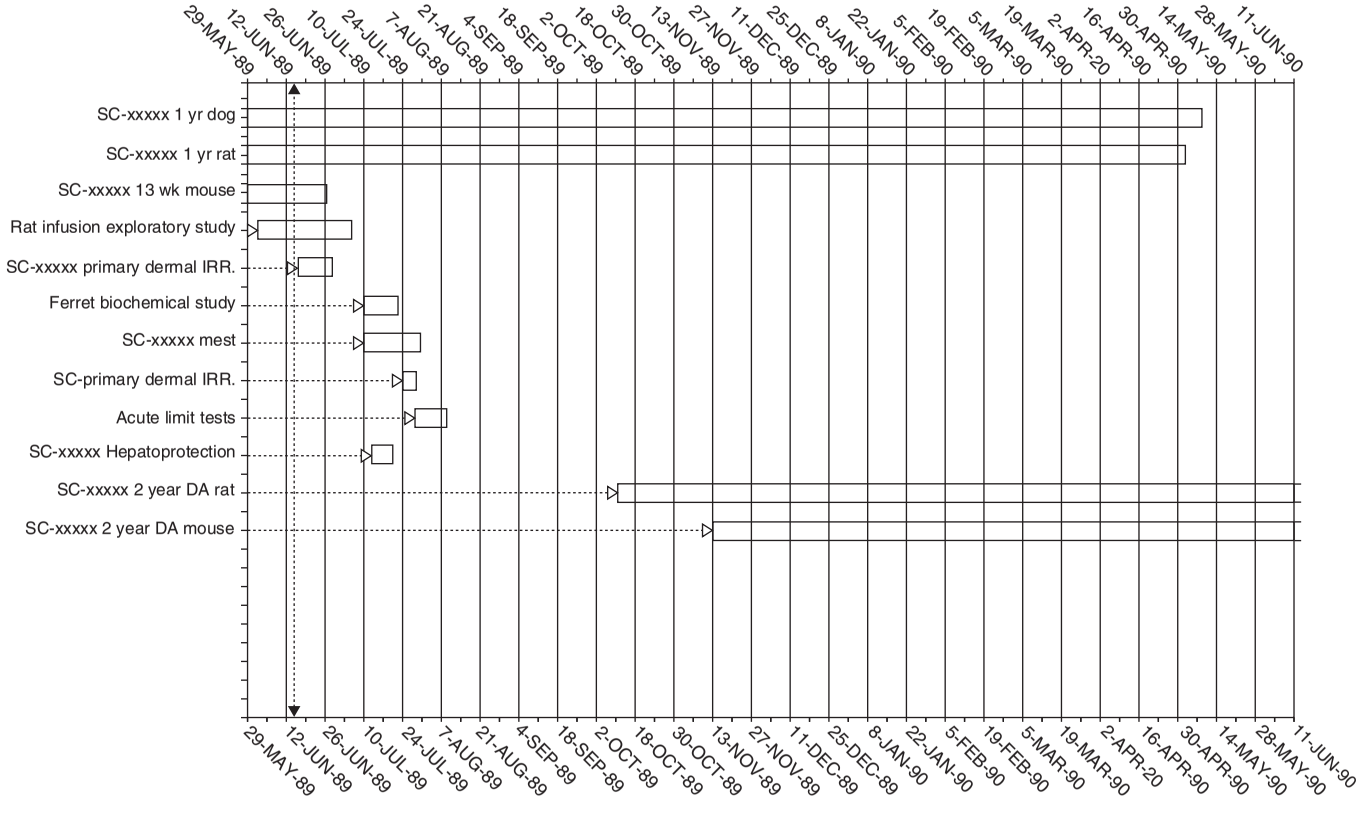


***Figure 3*** *Gantt, or bar, chart showing scheduling of major safety assessment activities (studies) involved in pharmaceutical development project.*

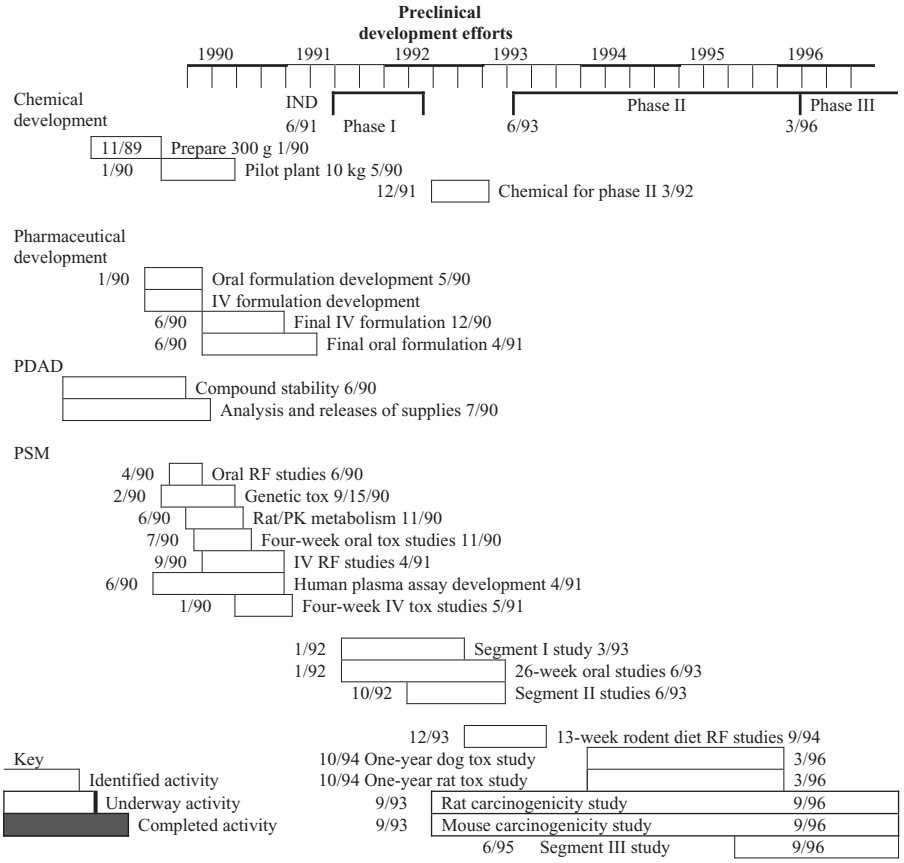


***Figure 4*** *Hybrid project Gantt chart which identifies work of each development function ("line operation") in development of new compound and how it matches phase of development.*

A useful indicator of the efficacy of an assay series is the frequency of discovery of truly active compounds. The frequency is related to the probability of discovery and to the degree of risk (hazard to health) associated with an active compound passing a screen undetected. These two factors in turn depend on the distribution of activities in the series of compounds being tested and the chances of rejecting or accepting compounds with given activities at each stage.

Statistical modeling of the assay system may lead to the improvement of the design of the system by reducing the interval between discoveries of active compounds. The objectives behind a screen and considerations of (1) costs for producing compounds and testing and (2) the degree of uncertainly about test performance will determine the desired performance characteristics of specific cases. In the most common case of early toxicity screens performed to remove possible problem compounds, preliminary results suggest that it may be beneficial to increase the number of compounds tested, decrease the number of animals per group, and increase the range and number of doses. The result

will be less information on more structure, but there will be an overall increase in the frequency of discovery of active compounds (assuming that truly active compounds are entering the system at a steady rate).

The methods described here are well suited to analyzing screening data when the interest is truly in detecting the absence of an effect with little chance of false negatives. There are many forms of graphical analysis methods available, including some newer forms that are particularly well suited to multivariate data (the type that are common in more complicated screening test designs). It is intended that these aspects of analysis will be focused on in a later publication.

The design of each assay and the choice of the activity criterion should therefore be adjusted, bearing in mind the relative costs of retaining false positives and rejecting false negatives. Decreasing the group sizes in the early assays reduces the chance of obtaining significance at any particular level (such as 5%), so that the activity criterion must be relaxed, in a statistical sense, to allow more compounds through. At some stage, however, it becomes too expensive to continue screening many false positives, and the criteria must be tightened accordingly. Where the criteria are set depends on the acceptable noise levels in a screening system.

### 4.1 Characteristics of Screens

An excellent introduction to the characteristics of screens is Redman's (1981) interesting approach, which identifies four characteristics of an assay. Redman assumes that a compound is either active or inactive and the proportion of activities in a compound can be estimated from past experience. After testing, a compound will be classified as positive or negative (i.e., possessing or lacking activity). It is then possible to design the assay so as to optimize the following characteristics:

1. Sensitivity: the ratio of true positives to total activities.
2. Specificity: the ratio of true negatives to total inactives.
3. Positive accuracy: the ratio of true to observed positives.
4. Negative accuracy: the ratio of true to observed negatives.
5. Capacity: the number of compounds that can be evaluated.
6. Reproducibility: the probability that a screen will produce the same result at another time (and, perhaps, in some other lab).

An advantage of testing many compounds is that it gives the opportunity to average activity evidence over structural classes or to study quantitative structure–activity relationships (QSARs). Quantitative structure–activity relationships can be used to predict the activity of new compounds and thus reduce the chance of in vivo testing on negative compounds. The use of QSARs can increase the proportion of truly active compounds passing through the system.

To simplify this presentation, data sets drawn only from neuromuscular screening activity were used. However, the evaluation and approaches should be valid for all similar screening data sets, regardless of source. The methods are not sensitive to the biases introduced by the degree of interdependence found in many screening batteries that use multiple measures (such as the neurobehavioral screen):

1. Screens almost always focus on detecting a single endpoint of effect (such as mutagenicity, lethality, neurotoxicity, or development toxicity) and have a particular set of operating characteristics in common.
2. A large number of compounds are evaluated, so ease and speed of performance (which may also be considered efficiency) are very desirable characteristics.
3. The screen must be very sensitive in its detection of potential effective agents. An absolute minimum of active agents should escape detection; that is, there should be very few false negatives (in other words, the type II error rate or beta level should be low). Stated yet another way, the signal gain should be way up.
4. It is desirable that the number of false positives be small (i.e., there should be a low type I error rate or alpha level).
5. Items 2–4, which are all to some degree contradictory, require the involved researchers to agree on a set of compromises, starting with the acceptance of a relatively high alpha level (0.10 or more), that is, an increased noise level.
6. In an effort to better serve item 2, safety assessment screens are frequently performed in batteries so that multiple endpoints are measured in the same operation. Additionally, such measurements may be repeated over a period of time in each model as a means of supporting item 3.
7. This screen should use small amounts of compound to make item 1 possible and should allow evaluation of materials that have limited availability (such as novel compounds) early on in development.
8. Any screening system should be validated initially using a set of blind (positive and negative) controls. These blind controls should also be evaluated in the screening system on a regular basis to ensure continuing proper operation of the screen. As such, the analysis techniques used here can then be used to ensure the quality or modify performance of a screening system.
9. The more that is known about the activity of interest, the more specific the form of screen that can be employed. As specificity increases, so should sensitivity.
10. Sample (group) sizes are generally small.
11. The data tend to be imprecisely gathered (often because researchers are unsure of what they are looking for) and therefore possess extreme

within-group variability. Control and historical data are not used to adjust for variability or modify test performance.

12. Proper dose selection is essential for effective and efficient screen design and conduct. If insufficient data are available, a suitably broad range of doses must be evaluated (however, this technique is undesirable on multiple grounds, as has already been pointed out).

The design, use, and analysis of screens is covered in detail in Chapter Screens in Safety and Hazard Assessment of this volume.

## 5 STRATEGY AND PHASING

Regulatory requirements and our understanding of the pharmacology, marketing, and clinical objectives for a potential product provide a framework of requirements for the safety assessment of potential new pharmaceuticals. How one meets these requirements is not fixed, however. Rather, exactly what is done and when are reflections of the philosophy and managerial climate of the organization that is doing the discovery and development. It should be kept in mind that establishing and maintaining an excellent information base on the biological basis for a compound's expected therapeutic activity and safety are essential but often left undone. This subject is addressed in Chapter Regulation of Human Pharmaceutical Safety of this volume.

There are multiple phases involved in the safety assessment portion of the discovery, development, and marketing process. The actual conduct of the studies in each phase forms the basis of the bulk of the chapters in this book. However, unless the pieces are coordinated well and utilized effectively (and completed at the right times), success of the safety assessment program is unlikely or very expensive.

First, support needs to be given to basic research (also called discovery, biology, or pharmacology in different organizations) so that it can efficiently produce a stream of potential new product compounds with as few overt toxicity concerns as possible. This means that there must be early and regular interaction between the individuals involved and safety assessment must provide screening services to rank the specific safety concerns of the compounds. These screens may be in vitro (both for genetic and nongenetic endpoints) or in vivo (designed on purpose for a single endpoint, such as effects on reproductive performance, promotion activity, etc.). There must also be ongoing work to elucidate the mechanisms and structure–activity relationships behind those toxicities that are identified (Gad, 1989b).

Second is the traditional core of safety assessment that is viewed as development. Development includes providing the studies to support compounds getting into the clinic (an IND application being filed and accepted), evaluating a compound to the point at which it is considered safe, able to be absorbed, and effective (clinical phase II), and, finally, registration (filing an NDA and

having it approved). Various organizations break this process up differently. Judgments are generally made on the likelihood of compounds failing ("dying") at different stages in the clinical development process, and the phasing of preclinical support is selected and/or adjusted accordingly. If an organization has a history of many compounds failing early in the clinic (such as in the initial phase I tolerance trials, where there may be only 3–10 days of human dosing), then initial "pivotal" preclinical studies are likely to be only 4-week-long studies. If compounds tend to fail only in longer efficacy trials, then it is more efficient to run longer initial preclinical trials. Figure 5 shows several variations on these approaches. Additionally, the degree of risk involved in study design (particularly in dose selection) is also an organizational characteristic. Pivotal studies can fail on two counts associated with dose selection. Either they cannot identify a "safe" (no-effect) dose or they can neglect to find a dose that demonstrates a toxic effect (and therefore allows identification of potential target organs). Therefore, picking the doses for such studies is an art that has been risky because, traditionally, only three different dose groups have been used, and before clinical trials are conducted, there is

Plan 1: Clinical decision point is short term tolerance or human pharmacokinetics

| Range finder | | Pivotal study | | Phase I | | | Phase II | |
|---|---|---|---|---|---|---|---|---|
| "Pyramid" or "rolling acute" (see Chapter Screens in Safety and Hazard Assessment) | → | Two or four weeks in two species by the intended route | → | Tolerance and pharmacokinetics (PK) with up to 3–14 days human dosing | → | Delay*** | | → |

Plan 2: Clinical decision point is an indication of efficacy in humans

| Range finder | | Pivotal study | | Phase I/II | | | Phase II/III | |
|---|---|---|---|---|---|---|---|---|
| "Pyramid" or "rolling acute" and/or two- or four-week study | → | Thirteen weeks in two species | → | Tolerance, PK and efficacy with human dosing up to one month in length | → | Delay*** | | → |

Plan 3: Plan for success or resources are not a constraint

| Range finder | Pivotal study | Phase I/II | Phase III | |
|---|---|---|---|---|
| "Pyramid" or "rolling acute: and/or two week study | Four weeks in two species | Preclinical | Preclinical | |
| | | Thirteen weeks in two species | One year in two species | Carcinogenicity in two rodents (if required) |

***Figure 5*** *Three different approaches to matching preclinical safety efforts to support clinical development of new drug. Which is the best one for any specific case depends on considerations of resource availability and organizational tolerance of "risk." In plan 1, little effort will be "wasted" on projects that fail during early (phase I) clinical trials, but if phase I trials are successful, there will be major delays. In plan 3, clinical development will never be held up waiting for more safety work, but a lot of effort will go into projects that never get past phase I. Plan 2 is a compromise. Delays are to allow additional preclinical (animal safety) studies to support longer clinical trials in accordance with FDA or other applicable guidelines.*

at best a guess as to what clinical dose will need to be cleared. The use of four (or five) dose groups only marginally increases study cost and, in those cases where the uncertainty around dose selection is great, provides a low-cost alternative to repeating the study.

Pivotal studies can also be called *shotgun tests* because it is unknown in advance what the endpoints are being aimed at. Rather, the purpose of the study is to identify and quantitate all potential systemic effects resulting from a single exposure to a compound. Once known, specific target organ effects can then be studied in detail if so desired. Accordingly, the generalized design of these studies is to expose groups of animals to controlled amounts or concentrations of the material of interest and then to observe for and measure as many parameters as practical over a period past or during the exposure. Further classification of tests within this category would be the route by which test animals are exposed/dosed or by the length of dosing. "Acute," for example, implies a single exposure interval (of 24 h or less) or dose of test material. Using the second scheme (length of dosing), the objectives of the successive sets of pivotal studies could be defined as follows:

Acute or dose range-finding (DRF) studies:

1. Set maximum doses for next studies.
2. Identify very or unusually toxic agents.
3. Estimate upper limit of tolerability.
4. Identify organ system affected.

Two-week studies:

1. Set doses for next studies.
2. Identify organ toxicity.
3. Identify very or unusually toxic agents.
4. Estimate lethality potential.
5. Evaluate potential for accumulation of effects.
6. Get estimate of kinetic properties (blood sampling/urine sampling).

Four-week studies:

1. Set doses for next studies.
2. Identify organ toxicity.
3. Identify very or unusually toxic agents.
4. Estimate lethality potential.
5. Evaluate potential for accumulation of effects.
6. Get estimate of kinetic properties (blood sampling/urine sampling).
7. Elucidate nature of specific types of target organ toxicities induced by repeated exposure.

Thirteen-week studies (now commonly included as an interim necropsy and report in chronic studies):

1. Set doses for next studies.
2. Identify organ toxicity.
3. Identify very or unusually toxic agents.
4. Evaluate potential for accumulation of effects.
5. Evaluate pharmacokinetic properties.
6. Elucidate nature of specific types of target organ toxicities induced by repeated exposure.
7. Evaluate reversibility of toxic effects.

Chronic studies:

1. Elucidate nature of specific types of target organ toxicities induced by prolonged repeated exposure or
2. identify potential carcinogens.

The problems of scheduling and sequencing toxicology studies and entire testing programs have been minimally addressed in print. Though there are several books and many articles available that address the question of scheduling multiple tasks in a service organization (French, 1982) and an extremely large literature on project management (as briefly overviewed earlier in this chapter), literature specific to a research testing organization does not exist.

For all the literature on project management, however, a review will quickly establish that it does not address the rather numerous details that affect study/program scheduling and management. There is, in fact, to my knowledge, only a single article (Levy et al., 1977) in the literature that addresses scheduling, and it describes a computerized scheduling system for single studies.

There are commercial computer packages available for handling the network construction, interactions, and calculations involved in what, as will be shown below, is a complicated process. These packages are available for use on both mainframe and microcomputer systems.

Scheduling for the single-study case is relatively simple. One should begin with the length of the actual study and then factor in the time needed before the study is started to secure the following resources:

- Animals must be on hand and properly acclimated (usually for at least two weeks prior to the start of the study).
- Vivarium space, caging, and animal care support must be available.
- Technical support for any special measurements such as necropsy, hematology, urinalysis, and clinical chemistry must be available on the dates specified in the protocol.
- Necessary and sufficient test material must be on hand.
- A formal written protocol suitable to fill regulatory requirements must be on hand and signed.

The actual study (from first dosing or exposure of animals to the last observation and termination of the animals) is called the in-life phase, and many people assume the length of the in-life phase defines the length of a study. Rather, a study is not truly completed until any samples (blood, urine, and tissue) are analyzed, slides are prepared and microscopically evaluated, data are statistically analyzed, and a report is written, proofed, and signed off. Roll all of this together, and if you are conducting a single study under contract in an outside laboratory, an estimate of the least time involved in its completion should be equal to (other than in the case of an acute or single-endpoint study) no more than:

$$L + 6 \text{ weeks} + \tfrac{1}{2} L$$

where $L$ is the length of the study. If the study is a single-endpoint study and does not involve pathology, then the least time can be shortened to $L + 6$ weeks. In general, the best that can be done is $L + 10$ weeks.

When one is scheduling an entire testing program on contract, it should be noted that, if multiple tiers of tests are to be performed (such as acute, 2-week, 13-week, and lifetime studies), then these must be conducted sequentially, as the answer from each study in the series defines the design and sets the doses for the subsequent study.

If, instead of contracting out, one is concerned with managing a testing laboratory, then the situation is considerably more complex. The factors and activities involved are outlined below. Within these steps are rate-limiting factors that are invariably due to some critical point or pathway. Identification of such critical factors is one of the first steps for a manager to take to establish effective control over either a facility or a program.

Before any study is actually initiated, a number of prestudy activities must occur (and, therefore, these activities are currently underway—to one extent or another—for the studies not yet underway but already authorized or planned for this year for any laboratory):

- Test material procurement and characterization
- Development of formulation and dosage forms for study
- If inhalation study, development of generation and analysis methodology, chamber trials, and verification of proper chamber distribution
- Development and implementation of necessary safety steps to protect involved laboratory personnel
- Arrangement for waste disposal
- Scheduling to assure availability of animal rooms, manpower, equipment, and support services (pathology and clinical)
- Preparation of protocols
- Animal procurement, health surveillance, and quarantine
- Preparation of data forms and books

- Conduct of prestudy measurements on study animals to set baseline rates of body weight gain and clinical chemistry values

After completion of the in-life phase (i.e., the period during which live animals are used) of any study, significant additional effort is still required to complete the research. This effort includes the following:

- Preparation of data forms and books, preparation of tissue slides, and microscopic evaluation of these slides
- Preparation of data tables
- Statistical analysis of data
- Preparation of reports

There are a number of devices available to a manager to help improve the performance of a laboratory involved in these activities. One such device (cross training) is generally applicable enough to be particularly attractive.

Identification of rate-limiting steps in a toxicology laboratory over a period of time usually reveals that at least some of these are variable (almost with the season). At times, there is too much work of one kind (say, inhalation studies) and too little of another (say, dietary studies). The available staff for inhalation studies cannot handle this peak load, and since the skills of these two groups are somewhat different, the dietary staff (which is now not fully occupied) cannot simply relocate down the hall and help out. However, if, early on, one identifies low- and medium-skill aspects of the work involved in inhalation studies, one could cross train the dietary staff at a convenient time so that it could be redeployed to meet peak loads.

It should be kept in mind that there are a number of common mistakes (in both the design and conduct of studies and in how information from studies is used) that have led to unfortunate results, ranging from losses in time and money and the discarding of perfectly good potential products to serious threats to people's health. Such outcomes are indeed the great disasters in product safety assessment—especially since many of them are avoidable if attention is paid to a few basic principles.

It is quite possible to design a study for failure. Common shortfalls include:

1. Using the wrong animal model.
2. Using the wrong route or dosing regimen.
3. Using the wrong vehicle or formulation of test material.
4. Using the wrong dose level. In studies where several dose levels are studied, the worst outcome is to have an effect at the lowest dose level tested (i.e., the safe dosage in animals remains unknown). The next worst outcome is to have no effect at the highest dose tested (generally meaning that the signs of toxicity remain unknown, invalidating the study in the eyes of many regulatory agencies).

5. Making leaps of faith. An example is to set dosage levels based on others' data and to then dose all test animals. At the end of the day, all animals in all dose levels are dead. The study is over; the problem remains.
6. Using the wrong concentration of test materials in a study. Many effects (including both dermal and gastrointestinal irritation, for example) are very concentration dependent.
7. Failing to include a recovery (or rebound) group. If one finds an effect in a 90-day study (say, gastric hyperplasia), how does one interpret it? How does one respond to the regulatory question, "Will it progress to cancer?" If an additional group of animals were included in dosing and, then were maintained for a month after dosing had been completed, recovery (reversibility) could be both evaluated and (if present) demonstrated.

Additionally, there are specialized studies designed to address endpoints of concern for almost all drugs (carcinogenicity, reproductive or developmental toxicity) or concerns specific to a compound or family of compounds (local irritation, neurotoxicity, or immunotoxicity, for example). When these are done (timing) also requires careful consideration. It must always be kept in mind that the intention is to ensure the safety of people in whom the drug is to be evaluated (clinical trials) or used therapeutically. An understanding of special concerns for both populations should be considered essential.

Safety evaluation does not cease being an essential element in the success of the pharmaceutical industry once a product is on the market. It is also essential to support marketed products and ensure that their use is not only effective but also safe and unclouded by unfounded perceptions of safety problems. This requires not only that clinical trials be monitored during development (Spector et al., 1988) but also that experience in the marketplace be monitored.

The design and conduct of safety assessment studies and programs also require an understanding of some basic concepts:

1. The studies are performed to establish or deny the safety of a compound, rather than to characterize the toxicity of a compound.
2. Because pharmaceuticals are intended to affect the functioning of biological systems and safety assessment characterizes the effects of higher than therapeutic doses of compounds, it is essential that one be able to differentiate between hyperpharmacology and true (undesirable) adverse effects.
3. Focus of the development process for a new pharmaceutical is an essential aspect of success but is also difficult to maintain. Clinical research units generally desire to pursue as many or as broad claims as possible for a new agent and frequently also apply pressure for the development of multiple forms for administration by different routes. These forces

must be resisted because they vastly increase the work involved in safety assessment, and they may also produce results (in one route) that cloud evaluation [and impede institutional review board (IRB) and regulatory approval] of the route of main interest.

## 6 CRITICAL CONSIDERATIONS

In general, at the beginning of a project the management of a pharmaceutical development enterprise wants to know three things: what are the risks (and how big are they), how long will it take, and how much (money and test compound) will it take?

The risk question is beyond the scope of this volume. The time question was addressed earlier in this chapter. How much money is needed is also beyond the scope of this volume. But calculating projected compound needs for studies is a fine challenge in the design and conduct of a safety evaluation program. The basic calculation is simple. The amount needed for a study is equal to

$$NWILD$$

where $N$ = number of animals per group
$W$ = mean weight per animal during course of study (kg)
$I$ = total number of doses to be delivered (such as, in a 28-day study, 28 consecutive doses)
$L$ = loss or efficiency factor (to allow for losses in formulation and dose delivery a 10% factor is commonly employed, meaning a value of 1.1 is utilized)
$D$ = total dose factor, i.e., the sum of all dose levels; e.g., if groups are to receive 1000, 300, 100, and 30 mg kg$^{-1}$, then total dose factor is 1000 + 300 + 100 + 30, or 1430 mg kg$^{-1}$.

As an example, take a 28-day study in rats where there are 10 males and 10 females per group and the dose levels employed are 1000, 3000, 100, and 30 mg kg$^{-1}$. Over the course of the 28 days the average weight of the rats is likely to be 300 g (or 0.3 kg). This means our values are

$$N = 20 \quad W = 0.3\ \text{kg} \quad I = 28 \quad L = 1.1 \quad D = 1430\ \text{mg kg}^{-1}$$

and therefore our total compound needs will be

$$(20)(0.3)(28)(1.1)(1430\ \text{mg}) = 2642.64\ \text{mg or } 2.642\ \text{g}$$

This is the simplest case but shows the principles.

A governing principle of pharmaceutical safety assessment is the determination of safety factors—the ratio between the therapeutic dose (that which

achieves the desired therapeutic effect) and the highest dose which evokes no toxicity. This grows yet more complex (but has less uncertainty) if one bases these ratios on plasma levels rather than administered doses. Traditionally based on beliefs of differences of species sensitivity, it has been held that based on toxicity findings a minimum 5-fold safety factor should be observed in nonrodents and 10-fold in rodents.

The desire to achieve at least such minimal therapeutic indices and to also identify levels associated with toxicity (and the associated toxic effects) forms the basis of dose selection for systemic (and most other in vivo) toxicity studies.

## 7 SPECIAL CASES IN SAFETY ASSESSMENT

It may seem that the course of preclinical safety assessment (and of other aspects of development) of a pharmaceutical is a relatively linear and well-marked route within some limits. This is generally the case, but not always. There are a number of special cases where the pattern and phasing of development (and of what is required for safety assessment) do not fit the usual pattern. Four of these cases are:

1. When the drug is intended to treat a life-threatening disease, such as acquired immunodeficiency syndrome (AIDS)
2. When the drug is a combination of two previously existing drug entities
3. When the drug consists of two or more isomers
4. When the drug is a peptide produced by a biotechnology process

Drugs intended to treat a life-threatening disease for which there is no effective treatment are generally evaluated against less rigorous standards of safety when making decisions about advancing them into and through clinical testing. This acceptance of increased risk (moderated by the fact that the individuals involved will die if not treated at all) is balanced against the potential benefit. These changes in standards usually mean that the phasing of testing is shifted: Animal safety studies may be done in parallel or (in the case of chronic and carcinogenicity studies) after clinical trials and commercialization. But the same work must still eventually be performed.

Combination drugs, at least in terms of safety studies up to carcinogenicity studies, are considered by regulatory agencies as new drug entities and must be so evaluated. Accordingly the required safety tests must be performed on a mixture with the same ratio of components as is to be a product. Any significant change in ratios of active components means one is again evaluating a new drug entity.

Now that it is possible to produce drugs that have multiple isomers in the form of single isomers (as opposed to racemic mixtures), for good historical

reasons, regulatory agencies are requiring at least some data to support any decision to develop the mixture as opposed to a single isomer. One must, at a minimum, establish that the isomers are of generally equivalent therapeutic activity and, if there is therapeutic equivalence, that any undesirable biological activity is not present to a greater degree in one isomer or another.

## 8 SUMMARY

It is the belief of this author that the entire safety assessment process that supports pharmaceutical research and development is a multistage process of which no single element is overwhelmingly complex. These elements must be coordinated and their timing and employment carefully considered on a repeated basis. Focus on the objective of the process, including a clear definition of the questions being addressed by each study, is essential, as is the full integration of the technical talents of each of the many disciplines involved. A firm understanding of the planned clinical development of the drug is essential. To stay competitive requires that new technologies be identified and incorporated effectively into safety assessment programs as they become available. It is hoped that this volume will provide the essential knowledge of the key elements to allow these goals to be realized.

## REFERENCES

Alder, S., and Zbinden, G. (1988). *National and International Drug Safety Guidelines*. MTC Verlag, Zollikon, Switzerland.

Arnst, C. (1998). Danger: Read the label. *Business Week*, May 11, p. 100.

Beyer, K. (1978). Discovery, development and delivery of new Drugs. *Monographs in Pharmacology and Physiology*, No. 12. Spectrum, New York.

Fishlock, D. (1990, Apr. 24). Survival of the fittest drugs. *Financial Times*, pp. 16–17.

French, S. (1982). *Sequencing and Scheduling*. Halsted, New York.

Gad, S. C. (1988). An approach to the design and analysis of screening studies in toxicology. *J. Am. Coll. Toxicol.* 7(2):127–138.

Gad, S. C. (1989a). Principles of screening in toxicology: With special emphasis on applications to neurotoxicology. *J. Am. Coll. Toxicol.* 8(1):21–27.

Gad, S. C. (1989b). A tier testing strategy incorporating in vitro testing methods for pharmaceutical safety assessment. *Humane Innovations and Alternatives in Animal Experimentation* 3:75–79.

Gad. S. C. (1996). Preclinical toxicity testing in the development of new therapeutic agents, *Scand. J. Lab. Anim. Sci.* 23:299–314.

Gad, S. C. (2000). *Product Safety Evaluation Handbook*, 2nd ed. Marcel Dekker, New York.

Gad, S. C. (2001). *In Vitro Toxicology*, 2nd ed. Taylor and Francis, Philadelphia, PA.

Guarino, R. A. (1987). *New Drug Approval Process*. Marcel Dekker, New York.

Hamner, C. E. (1982). *Drug Development*. CRC Press, Boca Raton, FL, pp. 53–80.

International Conference on Organisation (ICH) (2009). M3(R2): Non-clinical safety studies for the conduct of human clinical trials for pharmaceuticals. ICH, Brussels.

Kliem, R. L. (1986). *The Secrets of Successful Project Management*. Wiley, New York.

Knutson, J. R. (1980). *How to Be a Successful Project Manager*. American Management Associations, New York.

Lazaron, J., Parmeranz, B. H. and Corey, P. N. (1998). Incidence of adverse drug reactions in hospitilized patients. *JAMA* 279:1200–1209.

Leber, P. (1987). FDA: The federal regulations of drug development. In H. Y. Meltzer (Ed.), *Psychopharmacology: The Third Generation of Progress*. Raven, New York, pp. 1675–1683.

Lesko, L. J., Rowland, M., Peck, C. C., and Blaschke, T. F. (2000). Optimizing the science of drug development: Opportunities for better candidate selection and accelerated evaluation in humans. *Pharm. Res.* 17:1335–1344.

Levy, A. E., Simon, R. C., Beerman, T. H., and Fold, R. M. (1977). Scheduling of toxicology protocol studies. *Comput. Biomed. Res.* 10:139–151.

Littlehales, C. (1999). The price of a pill. *Modern Drug Discov.* Jan./Feb. 1999, pp. 21–30.

Matoren, G. M. (1984). *The Clinical Process in the Pharmaceutical Industry*. Marcel Dekker, New York, pp. 273–284.

Meyer, D. S. (1989). Safety evaluation of new drugs. In Y. C. Martin, E. Kutter, and V. Austel (Eds.), *Modern Drug Research*. Marcel Dekker, New York, pp. 355–399.

Moore, T. J., Patsy, D., and Furnberg, J. (1998). Time to act on drug safety. *JAMA* 279:1971–1976.

O'Grady, J., and Linet, O. I. (1990). *Early Phase Drug Evaluation in Man*. CRC Press, Boca Raton, FL.

Raeburn, P. (1999). Drug safety needs a second opinion. *Business Week*, Sept. 20, pp. 72–74.

Redman, C. (1981). Screening compounds for clinically active drugs. In C. R. Buncher and J. Tsay (Eds.), *Statistics in the Pharmaceutical Industry*. Marcel Dekker, New York, pp. 19–42.

Scheuplein, R. J., Shoal, S. E., and Brown, R. N. (1990). Role of pharmacokinetics in safety evaluation and regulatory considerations. *Ann. Rev. Pharmacol. Toxicol.* 30:197–218.

Smith, C, G. (1992). *The Process of New Drug Discovery and Development*. CRC Press, Boca Raton, FL.

Sneader, W. (1986). *Drug Development: From Laboratory to Clinic*. Wiley, New York.

Spector, R., Park, G. D., Johnson, G. F., and Vessell, E. S. (1988). Therapeutic drug monitoring. *Clin. Pharmacol. Ther.* 43:345–353.

Spilker, B. (1994). *Multinational Drug Companies*, 2nd ed. Raven, New York.

Traina, V. M. (1983). The role of toxicology in drug research and development. *Med. Res. Rev.* 3:43–72.

Tukey, J. W. (1977). *Exploratory Data Analysis*. Addison-Wesley, Reading, MA.

U.S. Food and Drug Administration (FDA) (1987a). Good laboratory practice regulations: Final rule. 21 CFR Part 58, *Fed. Reg.*, Sept. 4, 1987.

U.S. Food and Drug Administration (FDA) (1987b). New drug, antibiotic, and biologic drug produce regulations. 21 CFR Parts 312, 314, 511, and 514, *Fed. Reg.* 52(53):8798–8857.

Zbinden, G. (1992). *The Source of the River Po*. Haag and Herschen, Frankfurt, Germany.

Zbinden, G., Elsner, J., and Boelsterli, U. A. (1984). Toxicological screening. *Reg. Toxicol. Pharmacol.* 4:275–286.

# *Application of In Vitro Techniques in Drug Safety Assessment*

## 1 INTRODUCTION

The key assumptions underlying modern toxicology are (1) that other organisms can serve as accurate predictive models of toxicity in humans, (2) that selection of an appropriate model to use is the key to accurate prediction of potential hazard in humans, and (3) that understanding the strengths and weaknesses of any particular model is essential to translating potential hazards identified in these models to assess relevant hazards in humans and in the subsequent management of actual risks. The nature of models and their selection in toxicological research became the subject of critical scientific review starting in the 1980s. Usually in toxicology, when we refer to "models," we really have meant test organisms or systems, although, in fact, the manner in which parameters are measured (and which parameters are measured to characterize an endpoint of interest) is also a critical part of the model (or, indeed, may actually constitute the model).

Although there have been accepted principles for test organism selection, these have not generally been the actual final basis for such selection. It is a fundamental hypothesis of both historical and modern toxicology that adverse effects caused by chemical entities in higher animals are generally the same as those induced by those entities in humans. There are many who point to individual exceptions to this and conclude that the general principle is false.

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

Yet, as our understanding of molecular biology advances and we learn more about the similarities of structure and function of higher organisms at the molecular level, it becomes clear that the mechanisms of chemical toxicity are largely identical in all higher life forms, including humans. The target sites are molecular, and differences in responses are all about similarities in receptor populations, receptor population distribution in organ systems, and the manner and means of getting toxicophores to these sites or preventing them from reaching these sites. In this sense, it is now the age of translational toxicology. This increased understanding has caused some of the same people who question the general principle of predictive value to in turn suggest that our state of knowledge is such that mathematical models or simple cell culture systems could be used just as well as intact higher animals to predict toxicities in humans. This last suggestion has unfortunately missed the point that the final expressions of toxicity in humans or animals are frequently the summations of extensive and complex interactions on cellular and biochemical levels. Zbinden (1987) and Gad (1996a) published extensively in this area, including a very advanced defense of the value of animal models. Lijinsky (1988) has reviewed the specific issues about the predictive value and importance of animals in carcinogenicity testing and research. Although it was once widely believed, and may still be believed by many animal rights activists, that in vitro mutagenicity tests would entirely replace animal bioassays for carcinogenicity, this is clearly not the case on either scientific or regulatory grounds (despite the limitations of the current bioassay models). Although there are differences in the responses of various species (including humans) to carcinogens, the overall predictive value of such results, when tempered by judgment, is clear. At the same time, a well-reasoned use of in vitro or other alternative test model systems is essential to the continued development of a product safety assessment program that is effective, efficient, and relevant to human safety (Gad, 1990, 1996a, 1998, 2000, 2007).

The subject of intact animal models and their proper selection and use has been addressed elsewhere (Gad, 2007) and will not be further addressed here. However, alternative models which use other than intact higher organisms are seeing increasing use in toxicology for a number of reasons.

The first and most significant factor behind the interest in so-called in vitro systems has clearly been philosophical and political—an unremitting campaign by a wide spectrum of individuals concerned with the welfare and humane treatment of laboratory animals (Singer, 1975), though some are also clearly simply antiscience and antitechnology. In 1959 Russell and Burch first proposed what have come to be called the 3 R's of humane animal use in research—replacement, reduction, and refinement. These have served as the conceptual basis for reconsideration of animal use in research.

Replacement means utilizing methods that do not use intact animals in place of those that do. For examples, veterinary students may use a canine cardiopulmonary resuscitation simulator, Resusci-Dog, instead of living dogs and cell cultures may replace mice and rats that are fed new products to

discover substances poisonous to humans. In addition, using the preceding definition of animal, an invertebrate (e.g., a horseshoe crab) could replace a vertebrate (e.g., a rabbit) in a testing protocol.

Reduction refers to the use of fewer animals. For instance, changing practices allow toxicologists to estimate the lethal dose of a chemical with as few as one-tenth the number of animals used in traditional tests. In biomedical research, long-lived animals, such as primates, may be used in multiple sequential protocols, assuming that they are not deemed inhumane or scientifically conflicting. Designing experimental protocols with appropriate attention to statistical inference can lead to decreases or increases in the number of animals used. Through coordination of efforts among investigators, several tissues may be simultaneously taken from a single animal. Reduction can also refer to the minimization of any unintentionally duplicative experiments, perhaps through improvements in information resources.

Refinement entails the modification of existing procedures so that animals are subjected to less pain and distress. Refinements may include administration of anesthetics to animals undergoing otherwise painful procedures, administration of tranquilizers for distress, humane destruction prior to recovery from surgical anesthesia, and careful scrutiny of behavioral indices of pain or distress followed by cessation of the procedure or the use of appropriate analgesics. Refinements also include the enhanced use of noninvasive imaging technologies that allow earlier detection of tumors, organ deterioration or metabolic changes, and the subsequent early euthanasia of test animals.

Progress toward these first three R's has been previously reviewed (Gad, 1990, 1993; Salem, 1995; Salem and Katz, 1998; Gribaldo, 2007). However, there is fourth R—responsibility—which was not in Russell and Burch's initial proposal. To toxicologists this is the cardinal R. They may be personally committed to minimizing animal use and suffering and to doing the best possible science of which they are capable, but at the end of it all, toxicologists must stand by their responsibility to be conservative in ensuring the safety of the people using or exposed to the drugs and chemicals produced by our society.

Since 1980, issues of animal use and care in toxicological research and testing have become one of the fundamental concerns of both science and the public. Are our results predictive of what may or may not be seen in humans? Are we using too many animals, and are we using them in a manner that gets the answer we need with as little discomfort to the animals as possible? How do we balance the needs of humans against the welfare of animals?

During the same time frame, interest and progress in the development of in vitro test systems for toxicity evaluations have also progressed. Early reviews by Hooisma (1982), Neubert (1982), and Williams et al. (1983) record the proceedings of conferences on the subject, but Rofe's (1971) review was the first found by this author. Although it is hoped that in the long term some of these (or other) in vitro methods will serve as definitive tests in place of those that use intact animals, at present it appears more likely that their use in most cases will be as screens. Frazier (1992), Gad (2000), and Gad and Chengelis

(1999) give recent overviews of the general concepts and status of in vitro alternatives.

Lastly, in viewing the progression of in vitro toxicology in the pharmaceutical industry, it is relevant to note that the advent of toxicology itself as a profession is a recent development in this century. This development can be reflected by the formation of the Society of Toxicology in 1961 and the establishment of drug safety units distinct from pharmacology departments within the industry in the 1950s and 1960s. Toxicology can therefore be regarded as a relatively young profession when compared to pharmacology, whose professional society, the American Society for Pharmacology and Experimental Therapeutics (ASPET), was established in 1908. Industrial laboratories have therefore been in the forefront in the incorporation of in vitro techniques in toxicology.

## 2 INDUSTRIAL APPLICATIONS

The number and types of in vitro toxicological models utilized in the pharmaceutical industry encompass virtually every major target organ of toxicological interest. A partial listing of representative test methods is provided in Table 1. The breadth of the systems available is impressive and again signifies a relatively rapid progression of in vitro test development in toxicology.

The fact that many of the in vitro test systems listed in Table 1 are utilized in toxicology underscores the commitment of industry to the principles of reduction, refinement, and replacement of whole-animal in vivo tests whenever possible. However, industrial toxicologist must appropriately balance this commitment with the fourth R—responsibility (Figure 1)—recently discussed by Gad (1990). The ethical and legal responsibility of the toxicologist is to assess the safety of new products and to protect, to the best of his or her ability, the public from harm. Thus, current test procedures cannot be abandoned unless the new tools can be adopted with the assurance that adverse properties will be reliably detected. A key factor in the application of in vitro techniques in toxicology involves the degree of correlation between events occurring in vitro and those that the toxicologist evaluates in the intact animal. This correlation determines the ultimate scientific value of the techniques and the level of confidence associated with a particular test in terms of its predictability from a safety perspective.

The criteria that determine the degree of correlation or level of confidence in a given test are summarized in Table 2. The first of these is predictability, both qualitative and quantitative. Qualitatively, do the rankings or order of toxicities in vitro correlate with the order of toxicities in vivo? From a quantitative aspect, what are the dose–response characteristics from which potency estimates and comparisons can be made? Finally, how do drug concentrations in vitro compare with those achieved in vivo? The second criterion to examine is that of test system identity. To what degree does the in vitro

**TABLE 1 In Vitro Toxicity Models**

| Organ | Models(s) | Applications |
|---|---|---|
| Liver | Hepatocytes | Hepatotoxicity |
| | Enzymes | Peroxisomal proliferation |
| | Isolated perfused liver | Enzyme inhibition/induction |
| Kidney | Tissue slices | Nephrotoxicity |
| | Cells | Renal transport |
| | Tubules | |
| | Membranes | |
| Brain | Slices | Receptor interactions |
| | Homogenates | Neurotoxicity |
| | Cells | |
| Heart | Myocytes | Cardiotoxicity |
| | Isolated atria | Receptor interactions |
| | Isolated perfused heart | |
| Muscle | Cells | Muscle irritation |
| | Smooth muscles | Receptor interactions and smooth muscle effects |
| | Skeletal muscle: phrenic nerve/diaphragm | Neuromuscular blockade |
| Blood | Red blood cells | Hemolytic potential |
| | Mast cells | Histamine release |
| | Platelets | Aggregation |
| Reproductive/endocrine | Sperm | Fertility |
| | Limb bud growth | Teratogenic potential |
| | Whole-embryo culture | Teratogenic potential |
| | Pituitary cultures | Prolactin, leuteinizing hormone, follicle-stimulating hormone release |
| | Testicular cultures | Reproductive toxicity |
| Immune | Mitogen assay | Immunomodulation |
| | Mixed-lymphocyte response | Immunomodulation |
| | Plaque-forming cell assay | Immunomodulation |
| | Macrophage phagocytosis assay | Immunomodulation |
| | Bone marrow colony-forming unit assay | Myelosuppression |
| Other | Yeast | Phototoxicity |
| | Bacterial/mammalian cells | DNA/chromosome damage |
| | Ocular cells/organ systems | Ocular irritation |
| | Dermal cells/organ systems | Dermal irritation |

system structurally and functionally mimic the in vivo organ? Third, the area of mechanisms of cellular injury is a key criterion to consider in the utilization of in vitro models for toxicological evaluation. The variety of mechanisms that can play a central role in cell injury underscores the need to recognize specifically how the response of the in vitro test system correlates with the response in the intact organ or organism. Lastly, the topic of compensatory factors needs to be considered. How does the ability of the in vitro system to scavenge toxic

***Figure 1*** *Four R's of in vitro testing.*

**TABLE 2 Criteria for Establishing In Vitro–In Vivo Correlations**

| |
|---|
| Predictability |
| Qualitative: Rankings (% maximum response) |
| Quantitative: Dose–response characteristics, relative drug concentrations in vitro–in vivo |
| Identity |
| Structure: Morphological correlates |
| Function: Tissue specificity (e.g., transport characteristics, metabolic pathways |
| Mechanisms of injury |
| Membrane damage (structural, functional) |
| Synthetic activity (protein, RNA, DNA) |
| Metabolic poisoning ($O_2$ utilization/consumption, glycolysis, gluconeogenesis) |
| Compensatory factors |
| Biochemical scavengers (glutathione, metalloproteins) |
| Detoxification pathways \ oxidation/reduction/hydrolysis/conjugation |
| Toxification pathways/ |

reaction products compare with in vivo abilities? What detoxification or toxification pathways relevant to the in vivo fate of toxins or chemicals are present in the in vitro test system? These questions must be addressed prior to utilization of in vitro procedures to establish a clear understanding of the assumptions and/or limitations of the data to be generated. While in vitro systems need not fulfill all the correlative criteria outlined to be useful, the successful achievement of these criteria will be required to replace current whole-animal tests.

## 2.1 In Vitro Testing in Pharmaceutical Safety Assessment

The preclinical assessment of the safety of potential new pharmaceuticals and new devices represents a special case of the general practice of toxicology (Gad, 1999; Meyer, 1989), possessing its own peculiarities and special considerations and differing in several ways from the practice of toxicology in other fields—for some significant reasons. Because of the economics involved and

the essential close interactions with other activities (e.g., clinical trials, chemical process optimization, formulation development, regulatory reviews), the development and execution of a crisp and flexible, yet scientifically sound, program is a prerequisite for success. The ultimate aim of preclinical safety and biocompatibility assessment also makes them different. A good safety assessment program seeks to efficiently and effectively move safe potential therapeutic agents or devices into the clinical evaluation, then to registration, and finally to market and to support them through this process. This requires the quick identification of those agents that are not safe so that effort (and limited resources) are not wasted on them.

Pharmaceuticals are intended to have human exposure. Furthermore, pharmaceuticals are intended to have biological effects on the people that receive them. Frequently, the interpretation of results and the formulation of decisions about the continued development and eventual use of a drug are based on an understanding of both the potential adverse effects of the agent and its likely benefits as well as the dose separation between these two. This makes a clear understanding of the dose–response relationship critical, so that the actual risk–benefit ratio can be identified. It is also essential that the pharmacokinetics be understood and that "doses" (plasma tissue levels) at target organ sites be known (Scheuplein et al., 1990). Integral pharmacokinetics are essential to such a safety program, especially now that there is wider recognition of the existence and importance of subpopulations with different metabolic competencies. As we have come to understand that pharmacogenetics underlie many of the subpopulation effects we see in both the safety and efficacy of drugs, we have also come to recognize that in vitro methods also offer some of the best and most efficient means of understanding the basis for these differences and for identifying members of specific subpopulations.

The development and safety evaluation of pharmaceuticals and medical devices have many aspects broadly or tightly specified by regulatory agencies (Gad, 2001a). An extensive set of defined safety evaluations is required before a product is ever approved for market. For pharmaceuticals, regulatory agencies have increasingly come to require not only the establishment of a "clean dose" in two species with adequate safety factors to cover potential differences between species but also an elucidation of the mechanisms underlying those adverse effects that are seen at higher doses and are not well understood. These regulatory requirements are compelling to the pharmaceutical toxicologist (Traina, 1983). There is not, however, a set menu of what must be done. Rather, many decisions (particularly in terms of the timing of testing) are open to professional judgment. Devices have tended to be more set piece in their testing approach but are beginning to likewise require more mechanistic understanding to allow for competitive positioning in the marketplace.

The discovery, development, and registration of a pharmaceutical or biological comprise an immensely expensive operation and represent a rather unique challenge. For every 9000–10,000 compounds specifically synthesized or isolated as potential therapeutics, one (on average) will actually reach the

market. The overall cost for each successful compound is currently estimated to be between $250 and $320 million (though those figures are, of course, burdened with the cost of all the unsuccessful compounds), with each successive stage in the development process being more expensive. This dynamic makes it of great interest to identify as early as possible those agents that are likely not to go the entire distance, allowing a concentration of effort on the compounds that have the highest probability of reaching the market (and of possessing therapeutic utility) to do so.

Compounds "drop out" of the process primarily for three reasons: (1) toxicity or (lack of tolerance), (2) (lack of) efficacy, and (3) (lack of) bioavailability of the therapeutic active moiety in humans. Early identification of "losers" in each of these three categories is thus extremely important (Fishlock, 1990), forming the basis for the use of screening in pharmaceutical discovery and development. How much and which resources to invest in screening and each successive step in support of the development of a potential drug are matters of strategy and phasing that are detailed elsewhere (Gad, 2000). A range of test systems is available to be used in screening and in the definitive testing that follows for selected promising compounds. Table 3 presents a summary of the levels of available model systems. Those test systems that involve in vitro methods are now providing new tools for use in both early screening and understanding the mechanisms of observed toxicity in preclinical and clinical studies (Gad, 1988b, 1992). Devices are generally less complicated in design and in their testing procedures and have a much lower rate of failure in the qualification and approval stages that precede going to market. The trend in devices, however, is for regulatory authorities to require more testing, to be more critical of results, and to take longer in the review and approval process.

The entire safety assessment process that supports new product research and development is a multistage effort in which none of the individual steps is overwhelmingly complex but for which the integration of the whole process involves fitting together a large and complex pattern of pieces. This chapter proposes an approach in which integration of in vitro test systems calls for a modification of the approach to the general safety assessment problem. This modification can be addressed by starting with the current general case and progressing to a means of changing the process in an iterative fashion as new tools become available. Particularly with an understanding of mechanisms of toxicity becoming increasingly important in both candidate drug selection and the design and evaluation of the relevance of findings, the integration of in vitro methodologies, particularly into the pharmaceutical safety assessment process, has become essential. Determining what information is needed calls for an understanding of the way in which the device or pharmaceutical is to be made and used as well as an understanding of the potential health and safety risks associated with exposure of humans who will either be using the drug/device or be associated with the processes involved in making it. This is on the basis of a hazard and toxicity profile. Once such a profile is established, the available literature is searched to determine what is already known. Much

**TABLE 3 Levels of Models for Safety Assessment and Toxicological Research**

| Level/Model | Advantages | Disadvantages |
|---|---|---|
| In vivo (intact higher organism) | Full range of organismic responses similar to target species | Cost<br>Ethical/animal welfare concerns<br>Species-to-species variability |
| Lower organisms (earthworms, fish) | Range of integrated organismic responses | Frequently lack responses typical of higher organisms<br>Animal welfare concerns |
| Isolated organisms | Intact yet isolated tissue and vascular system<br>Controlled environment and exposure conditions | Donor organism still required<br>Time consuming and expensive<br>No intact organismic responses<br>Limited duration of viability |
| Cultured cells | No intact animals directly involved<br>Ability to carefully manipulate system<br>Low cost<br>Ability to study a wide range of variables | Instability of system<br>Limited enzymatic capabilities and viability of system<br>No (or limited) integrated multicell and/or organismic responses |
| Chemical/biochemical systems | No donor organism problems<br>Low cost<br>Long-term stability of preparation<br>Ability to study wide range of variables<br>Specificity of response | No de facto correlation to in vivo systems<br>Limited to investigation of a single defined mechanism |
| Genomics and proteomics | Speed and broad scope | Much effort is still required to correlate to intact organism effects |
| Computer simulations | No animal welfare concerns<br>Speed and low per-evaluation cost | May not have predictive value beyond narrow range of structures<br>Expensive to establish |

of the necessary information for the support of safety claims in the registration of a new drug/device is regulatorily mandated. This is not the case at all, however, for those safety studies done (1) to select candidate products or materials for development, (2) to design pivotal safety studies to support registration, or (3) to pursue mechanistic questions about materials and products in development.

Taking into consideration this literature information and the previously defined exposure profile, investigators have traditionally used a tier approach to generate a list of tests or studies to be performed based on regulatory requirements. What goes into a tier system is determined by (1) regulatory requirements imposed by government agencies, (2) the philosophy of the parent organization, (3) economics, and (4) available technology. How such tests are actually performed is determined on one of the two bases. The first

(and most common) is the menu approach, which involves selecting a series of standard design tests as "modules" of data. This assumes that all drugs or devices are alike except for route and duration of administration. The second is an interactive/iterative approach, where strategies are developed and studies are designed based both on needs and on what has been learned to date about the product.

## 3 DEFINING TESTING OBJECTIVE

The initial and most important aspect of a product safety evaluation program is the series of steps that leads to an actual statement of the problem or of the objectives of testing and research programs. This definition of objectives is essential and, as proposed here, consists of five steps: (1) defining product or material use, (2) estimating or quantitating exposure potential, (3) identifying potential hazards, (4) gathering baseline data, and (5) designing and defining the actual research program to answer outstanding questions.

### 3.1 Objectives Behind Data Generation and Utilization

To understand how product safety and toxicity data are used and how the data generation process might be changed to better meet the product safety assessment needs of society, it is essential to understand that different regulatory organizations have different answers to these questions. The ultimate solution is in the form of a multidimensional matrix, with the three major dimension of the matrix being (1) the toxicity/biocompatibility data type (lethality, sensitization, corrosion, irritation, photosensitization, phototoxicity, etc.), (2) exposure characteristics (extent, population size, population characteristics, etc.), and (3) the stage in the research and development process we are dealing with.

What is called for is a careful zero-based consideration of what the optimum product safety assessment strategy for a particular development problem should be. Before formulating such a strategy and deciding what mix of tests should be used, it is first necessary to decide criteria for what would constitute an ideal (or at least acceptable) test system.

The ideal test should have an endpoint measurement that provides data such that dose–response relationships can be obtained where possible or necessary (and such are almost always necessary). Furthermore, any criterion of effect must be sufficiently accurate in the sense that it can be used to reliably resolve the relative toxicity of two compounds that produce distinct (in terms of hazard to humans) yet similar responses. In general, it may not be sufficient to classify compounds into generic toxicity categories, such as "intermediate" toxicity, since a candidate chemical that falls in a given category yet is borderline to the next more severe toxicity category should be treated with more concern than a second candidate that falls at the less toxic extreme of

the same category. Therefore it is useful for a test system to be able to accurately rank compounds with potentially similar uses within any common toxicity category.

The endpoint measurement of the "ideal" test system must be objective, so that a given compound will give similar results when tested using the standard test protocol in different laboratories. If it is not possible to obtain reproductive results in a given laboratory over time or between various laboratories, then the historical database against which new compounds are evaluated will be time and laboratory dependent. Along these lines, it is important for the test protocol to incorporate internal standards to serve as quality controls. Thus, test data could be represented utilizing a reference scale based on the test system response to the internal controls. Such normalization, if properly documented, could reduce intertest variability.

The test results from any given compound should be reproducible both intrinsically (within the same laboratory over time) and extrinsically (between laboratories). If these conditions are not satisfied, then there will be significant limitations on the application of the test system because it could potentially produce conflicting results at different times and places. Such a possibility would significantly reduce confidence in the outcome of any single assay or assay set. From a regulatory point of view, this possibility would be highly undesirable (and perhaps indefensible). Alternatives to current in vivo test systems basically should be designed to evaluate the subject toxic response in a manner as closely predictive of that occurring in humans as possible while also reducing animal use and avoiding inhumane treatments where possible.

From a practical point of view, several additional features of the ideal test should be satisfied. The test should be rapid so that the turnaround time for a given compound is reasonable. Obviously, the speed of the test and the ability to conduct tests on several candidate drugs or materials simultaneously will determine the overall productivity. The test should be inexpensive, so that it is economically competitive with current testing practices (in the pharmaceutical industry, any reduction in critical path time for decisions has great economic value, so speed is generally preferable to lower cost—within limits). Finally, the technology should be easily transferred from one laboratory to another without excessive capital investment specific to test implementation. Although some of these practical considerations may appear to present formidable limitations for a given test system at the present time, the possibility of future developments in testing technology could overcome these obstacles.

This brief discussion of the characteristics of the ideal test system provides a general framework for evaluation of alternative test systems in general. No test system is likely to be ideal, of course. Our current armamentarium of tests, primarily in vivo tests using mammals, has developed and been maintained because (1) the tests have generally performed well in preventing dangerous drugs and materials from reaching the marketplace (Gad, 1996b,c) and (2) we

are comfortable with them. A significant number of rationales exist for the use of current in vivo test systems:

1. They provide evaluation of actions/effects on intact animal and organ/tissue interactions.
2. Either pure chemical entities or complete formulated products (complex mixtures) can be evaluated.
3. Either concentration or diluted products can be tested.
4. They yield data on the recovery and healing processes.
5. They are required statutory tests for agencies such as the U.S. Food and Drug Administration (FDA) (for "pivotal" safety studies) and the European Economic Community (EEC).
6. They are quantitative and qualitative tests with scoring systems generally capable of ranking materials as to relative hazard.
7. They are amenable to modifications to meet the requirements of special situations (such as multiple dosing or exposure schedules).
8. They have extensive available database and cross-reference capability for evaluation of relevance to human situation.
9. They are easy to perform and relatively low in capital costs in many cases.
10. They are generally both conservative and broad in scope, providing for maximum protection by erring on the side of overprediction of hazard to humans.
11. They can be either single endpoint (such as lethality and pyrogenicity) or shotgun (also called multiple endpoint, including such test systems as a 13-week oral toxicity study).

At the same time, progress and critical examination over the last 15 years have led to the formulation of an equally impressive list of reasons for pursuing the development of in vitro test systems:

1. They avoid the complications (and potential confounding or masking findings) of animal and tissue/organ in vivo evaluation.
2. In vivo systems may assess only short-term sites of application or immediate structural alterations produced by agents. Note, however, that tests may be intended to evaluate only local effects.
3. Technician training and monitoring are critical in in vivo testing (particularly if the evaluation called for is subjective by nature).
4. If our objective is either the total exclusion of a particular type of agent or the identification of truly severe acting agents on an absolute basis (i.e., without false positives or false negatives), then in vivo tests in animals do not perfectly predict results in humans.
5. Structural and biochemical differences exist between test animals and humans that make extrapolation from one to the other difficult.

6. In vivo systems are not standardized.
7. In vitro tests provide variable correlation with human results.
8. Large biological variability exists between more complex experimental units (i.e., individual animals).
9. Large, diverse, and fragmented databases (which are not readily comparable) are generated by in vivo studies.

Therefore, it will be necessary to weigh the strengths and weaknesses of each proposed test system in order to reach a conclusion on how "good" any particular test is. The next section presents the basis for specific test evaluations.

## 4 DESIGNING TESTING PROGRAM AND BUILDING LIBRARY

The next step, given that no relevant data can be found from any literature sources or from any internal files (and that it has been determined what data are needed or most likely to allow selection of desirable candidate compounds), is to perform appropriate predictive tests. The bulk of this section addresses the specifics of performing such evaluations using in vitro models. Before considering how to design, develop the components of, and conduct such a testing program, we must first consider how the practice of safety assessment came to its current state of acceptance and utilization of such tests.

To understand how product safety and toxicity data are used and how the data generation process might be changed to better meet the safety assessment needs of both industry and society, it is essential to understand that different commercial and regulatory organizations have different questions to address and operate in different cultures. The ultimate answer as to whether a drug, biological, or device is safe requires consideration of a multidimensional matrix, whose four major dimensions are (1) the toxicity data type (lethality, sensitization, irritation, photosensitization, genetic toxicity, liver toxicity, etc.), (2) exposure characteristics (extent of use and routes of exposure, patient population size, patient population characteristics, etc.), (3) the benefit to be derived from the marketing and use of the drug, device, or biological, and (4) the type of commercial organization (what do they make and who regulates them, i.e., what is the community of interest?).

Medical devices and pharmaceuticals are two closely related communities. Their materials of concern are agents intended as therapeutics or as components of devices to be used in health care, where the production worker or health care provider (doctor, nurse, or pharmacist) may have a significant chance of exposure, but the major concern is for those patients who receive or use the drug or device. Various centers of the FDA are the primary U.S. regulators.

What is needed is a careful consideration of what the optimum product safety assessment (including safety pharmacology assessments for pharmaceu-

ticals and biologicals) strategy would be. The framework for such a strategy calls for considering each of the issues to be resolved or data points to be generated as a separate box or compartment in a flowchart. As in any flowchart, the individual components need to be arranged in a logical order so that work is not duplicated and the data from earlier "cells" (studies) are available and utilized to help better design, execute, and evaluate the results from subsequent cells. These component studies can each be considered a tool for generating required data, and the entire collection as arranged can be thought of as a data generation toolbox. Many (most) of the components that constitute each of the data generation toolboxes (screens, confirmatory tests, higher tier tests, and mechanistic evaluations) are common to all safety assessment programs in some form. But what is actually used for each of these tasks is not common to all of these programs, nor is how the decision points or notes in the chart operate (acceptance criteria and risk–benefit judgment for proceeding with the development of the candidate drug or device). The selection of these details is what constitutes the actual formulation of a strategy. Before formulating such a strategy and deciding what mix of tests should be used, it is first necessary to decide criteria for what would constitute and ideal (or acceptable) test program.

### 4.1 Considerations in Adopting New Test Systems

Conducting toxicological investigations in two or more species of laboratory animals is generally accepted as a benign, prudent, and responsible practice in developing a new chemical entity, especially one that is expected to receive widespread use and to have exposure potential over human lifetimes. Adding a second or third species to the testing regimen offers an extra measure of confidence to toxicological and other professionals who will be responsible for evaluating the associated risks, benefits, and exposure limitations or protective measures (Gad, 2000; Smith, 1992). Although it undoubtedly broadens and deepens a compound's profile of toxicity, the practice of enlarging on the number of test species is, as has been demonstrated in multiple points in the literature (Gad and Chengelis, 1999), an indiscriminate scientific generalization. Moreover, such a tactic is certain to generate the problem of species-specific toxicoses; that is, a toxic response or an inordinately low biological threshold for toxicity is evident in one species or strain, whereas all other species examined are either unresponsive or strikingly less sensitive. The investigator confronting such findings must be prepared to address the all-important question: Are humans likely to react positively or negatively to the test agent under similar circumstances?

Assuming that numerical odds prevail and that humans automatically fit into the predominant category, whether on the side of being safe or at risk, would be scientifically irresponsible. Far from being an irreconcilable nuisance, however, such a confounded situation can be an opportunity to advance more quickly into the heart of the search for predictive information. Species-specific toxicosis can frequently contribute toward better understanding of the general case if the

underlying biological mechanisms either causing or enhancing toxicity are defined, especially if it is discovered to uniquely reside in the sensitive species.

A mention of species-specific toxicosis usually implies that either different metabolic pathways for converting and excreting xenobiotics or anatomical differences are involved. The design of our current safety evaluation tests appear to serve society reasonably well (i.e., significantly more times than not) in identifying hazards that would be unacceptable in a confirmatory manner. However, the process can just as clearly be improved from the standpoints of both improving our protection of society and performing necessary screening and exploratory research in a manner that used fewer animals in a more humane manner.

### 4.2 In Vitro Models

In vitro models, at least as screening tests, have been used in toxicology for some 25 years. The years since 1980 have brought a great upsurge in interest in such models. This increased interest is due to economic and animal welfare pressures and technological improvements (Rowan and Stratmann, 1980; Borenfreund and Puerner, 1984; Tyson and Frazier, 1993; Salem and Baskin, 1993) and has led to the development and (in some cases) successful utilization of numerous new test methods. Tissues can now be engineered for specific model development, such as tubular heart tissue (Franchini et al., 2007).

In addition to potential advantages, in vitro systems also have a number of limitations that can contribute to their not being acceptable modes:

1. The chemical is not absorbed at all or is poorly absorbed in in vivo studies.
2. The chemical is well absorbed but is subject to "first-pass effect" in the liver.
3. The chemical is distributed so that less (or more) reaches the target tissue than would be predicted on the basis of its absorption.
4. The chemical is rapidly metabolized to an active or inactive metabolite that has a different profile of activity and/or different duration of action than the parent drug.
5. The chemical is rapidly eliminated (e.g., through secretory mechanisms).
6. Species of the two test systems used are different.
7. Experimental conditions of the in vitro and in vivo experiments differed and may have led to different effects than expected. These conditions include factors such as temperature or age, sex, and strain of animal.
8. Effects elicited in vitro and in vivo by the particular test substance in question differ in their characteristics.
9. Tests used to measure responses may differ greatly for in vitro and in vivo studies, and the types of data obtained may not be comparable.

10. The in vitro study may not use adequate controls (e.g., pH, vehicle used, volume of test agent given, samples taken from sham-operated animals), resulting in "artifacts" of methods rather than results.
11. In vitro data cannot predict the volume of distribution in central or peripheral compartments.
12. In vitro data cannot predict the rate constants for chemical movement between compartments.
13. In vitro data cannot predict the rate constants of chemical elimination.
14. In vitro data cannot predict whether linear or nonlinear kinetics will occur with a specific dose of a chemical in vivo.
15. Pharmacokinetic parameters (e.g., bioavailability, peak plasma concentration, half-life) cannot be predicted based solely on in vitro studies.
16. In vivo effects of a chemical are due to an alteration in the higher order integration of an intact animal system, which cannot be reflected in a less complex system.

At the same time, as pointed out in this chapter, there are substantial potential advantages in using in vitro systems. Using cell or tissue culture in toxicological testing results in (1) isolation of test cells or organ fragments from homeostatic and hormonal control, (2) accurate dosing, and (3) quantitation of results. It is important to devise a suitable model system that is related to the mode of toxicity of the compound. Tissue and cell cultures have the immediate potential to be used in two very different ways by industry: (1) to examine a particular aspect of the toxicity of a compound in relation to its toxicity in vivo (i.e., mechanistic or explanatory studies) and (2) as a form of rapid screening to compare the toxicity of a group of compounds for a particular form of response. Indeed, the pharmaceutical industry has used in vitro test systems in these two ways for years in the search for new potential drug entities.

The author has already addressed the theory and use of screens in toxicology (Gad, 1988a) and the general concepts associated with their integration into the pharmaceutical and device development process (Gad, 1995a). Mechanistic and explanatory studies are generally called for when a traditional test system gives a result that is unclear or whose relevance to the real-life human exposure is doubted. In vitro systems are particularly attractive for such cases because they can focus on defined single aspects of a problem or pathogenic response, free of the confounding influence of the multiple responses on an intact higher level organism. Note, however, that first one must know the nature (indeed the existence) of the questions to be addressed.

### 4.3 Current Case: Mixed Battery

The current situation reflects the significant advances made in toxicology since 1985. It is rare to see a pharmaceutical or device researched and developed

with the use of other than an extensively commingled in vivo and in vitro test battery. This is reflected in the use of what may be termed a mixed test battery. The principles behind the development of these batteries are as follows:

1. Pharmaceutical and device development (particularly the product safety assessment aspects of it) cannot continue to be performed as it has been traditionally (on ethical, economic, or competitive grounds).
2. While there are no generally accepted in vitro test systems immediately available to completely replace all (or, indeed, any) of the regulatorily mandated in vivo testing requirements, there are test systems that can replace distinct component in vivo tests or, just as important, preclude their having to be performed by providing information quickly that makes difficult "go–no go" decisions viable at an earlier (and cheaper) stage of development.
3. Some steps can be taken to move development and acceptance of additional in vitro systems along, including wider industry utilization of available test methodologies, increased public regulatory acceptance of in vitro data where appropriate, and continued multilab validation/evaluation studies of alternatives. The single most helpful step, however, would be the clear definition of what constitutes an acceptance criterion for new test designs by regulatory authorities.
4. Some modifications to current in vivo testing methods both can and should be adopted. A current example of this would be in medical devices where a substantial portion of the requirements under the governing regulatory body [International Organization for Standardization (ISO)] can be met with in vitro alternatives (cytogenicity, muscle cell implantation, the limulus test for pyrogens, and in vitro mutagenicity assays).

### 4.4 Continuing Incremental Advances: How to Get Them

Great progress has been made in conducting safety assessment tests in intact animals and in developing an array of promising in vitro replacements, supplements, and candidates for the in vivo test. We want to have in place (that is, accepted and used by industry and accepted without question by regulatory agencies) a battery of in vitro systems that would preclude or reduce the need for intact animal testing to necessary cases. We also want duplicate or unnecessary testing of materials to be reduced to a minimum. These goals are dictated as much by economic reasons and the need to do better science as they are by ethical and humane concerns. The efficient and effective safety assessment/toxicology laboratory of the very near future will have as its "front door" an in vitro screening shop that will draw validated specific target organ screens from a library as needed to perform the initial go–no go evaluations on new compounds (or at least provide guidance as to where further evaluation is

required). This same shop would also provide (again, from its established collection) in vitro system models to elucidate mechanistic questions later in the assessment process.

Some would say that this is the current state of the art. Much of the necessary library could be assembled from test systems that have been extensively evaluated and have already undergone extensive validation (Gad, 2000, 2001b). Three critical steps must be taken for the eventual fulfillment of these objectives: (1) acceptance of a scientific approach to the problem of safety assessment, (2) development of an operative validation and acceptance process for new test procedures, and (3) clear enunciation of an acceptance criterion for new test designs by regulatory authorities.

A scientific approach to safety assessment, such as the one presented in this chapter, does have proponents and adherents. Such an approach requires those involved in both the management and conduct of the safety assessment process to continually question (and test) both the efficacy and the validity of their evaluation systems and processes. More to the point, it requires recognition of the fact that "we have always done it this way" is not a reason for continuing to do so. This approach asks first what is the objective behind the testing and then how well our testing is meeting this objective.

Currently, the second step, a collaborative process involving industrial, academic, and regulatory agencies for the validation and "acceptance" for new test systems, is totally absent. The general model of peer recognition leading to acceptance by the scientific community is not working in this case, as should have been expected from a situation where politics, social policy, and litigation have as much influence as science itself.

## 4.5 Conclusion

The first principle in hazard assessment is to have the data correspond as closely as possible to the real-life situation; that is, the nearer the model to humnas, the better the quality of the prediction of any potential hazards. The second principle should now also be clear: To be able to translate toxicity to hazard and to be able to manage such hazards, it is essential to know how the agent is to be used and the marketplace it is to be part of. It is hoped that this section has made these relationships clear.

Finally, alternatives of both in vitro and in vivo types are in the process of development for almost all the different endpoints of concern in safety assessment. Many of these have promise and could be used as screens for many of the uses presented here or as mechanistic tools, but complete replacement is clearly not near at hand, particularly for the more complicated endpoints. How these tests can (and should) be integrated into strategies for product safety assessment is the key scientific and managerial challenge for the next decade. Not only are there strong reasons against continuing where we are, there is also the possibility of tremendous competitive advantage to those who successfully manage to integrate in vitro tools as both efficient screens and effec-

tive means of isolating and understanding the mechanistic underpinning for toxic and pathogenic processes. These advantages are not primarily a matter of each piece of testing (data generation point) being less expensive, but rather that multitudes of information can be on hand much earlier in the research (discovery) and development process (before the bulk of expenses associated with individual compounds has been incurred) to allow the elimination of noncompetitive candidates. At the same time, each practicing toxicologist should feel both a moral and ethical compulsion to reduce the number of animals used in research and testing to the fullest extent possible and to ensure that those that are used are maintained and used in as humane a manner possible.

## 5 LETHALITY

Many endpoints of interest in toxicology present a fundamental limitation to the development and use of an in vitro or nonmammalian system in place of established in vivo methods. While cytotoxicity is a component mechanism in many of these toxic responses, disruption or diminution of the integrated function of multiple cells and systems is just as important.

The evaluation of lethality [symbolized in the public mind by the median lethal dose ($LD_{50}$) test] would seem to offer a unique opportunity for the development and use of alternatives. Approaches to alternatives for lethality testing include no living materials at all [the structure–activity relationship (SAR) or computer model approaches], those that use no intact higher organisms (but rather cultured cells or bacteria), and those that use lower forms of animal life (e.g., invertebrates and fish). Each of these presents a different approach to the objective of predicting acute lethality in humans or, rarely, economic animals and will be examined in turn.

There are systems that do not directly use any living organisms but, rather, seek to predict the lethality (in particular, the $LD_{50}$) of a chemical on the basis of what is known about structurally related chemicals. Such SAR systems have improved markedly over the last 10 years (Enslein et al., 1983; Lander et al., 1984; Blagg, 2006) but are still limited. Accurate predictions are usually possible only for those classes of structures where data have previously been generated on several members of the classes. For new structural classes, the value of such predictions is minimal. Accordingly, this approach is valuable when working with analogues in a series but not for novel structures. It is also a strong argument for getting as much data as possible into the published literature.

A more extensive and once promising approach has been the use of various cultured cell systems. Kurack et al. (1986), for example, have developed and suggested a system based on cultured mammalian hepatocytes. The system does metabolize materials in a manner like mammalian target species and has shown promise in a limited battery of chemicals. Such mammalian cell culture and bacterial screening systems have significant weaknesses for assessing the

lethality of many classes of chemicals since they lack any of the integrative functions of a larger organism. Thus, they would miss all agents that act by disrupting functions such as the organophosphate pesticides, most other neurologically mediated lethal agents, and agents that act by modifying hormonal or immune systems.

Clive et al. (1979) have reported on the correlation of the median lethal concentration ($LC_{50}$) of a variety of chemicals in mouse lymphoma cell cultures with their oral $LD_{50}$ in mice, as shown in Figure 2. No linear correlation is present, but highly cytotoxic substances (in this group) are significantly more toxic orally. Given the impression of some $LD_{50}$ values, due to such factors as steepness of slope of the lethality curve, the lack of linear correlation should be no surprise. Most recently, Ekwall et al. (1989) have reported on the MEIC program system, which utilizes a battery of five cellular systems. For a group of 10 chemicals, the system provided good correlation with or predictive power of rat $LD_{50}$.

Recently Parce et al. (1989) reported on a biosensor technique in which cultured cells are confined to a flow chamber through which a sensor measures the rate of production of acidic metabolites. It is proposed to use this as a functional measure of cytotoxicity and as a screening technique for a number of uses, including in vivo lethality.

Three lower species of intact animals have been proposed for use in screening or testing of the lethal effects of chemicals. First, some researchers have shown a good correlation between the $LD_{50}$ of the same chemicals in rats. This correlation is nonlinear but still suggests that more toxic materials could be at least initially identified and classified in some form of screening system based

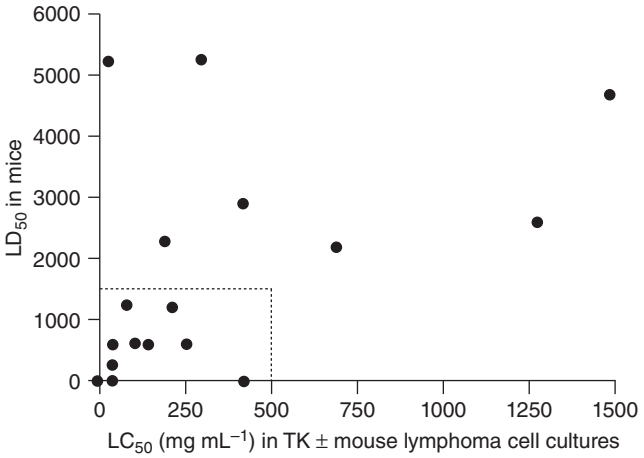


***Figure 2*** *Graph showing comparison of lethality of group of 18 drugs of diverse structure in in vivo (mouse) and in vitro (cultured mouse lymphoma cells) test systems. Correlation of these $LD_{50}/LC_{50}$ values is very poor, though extreme high- and low-scale values seem to be more closely associated in two systems.*

on *Daphnia*. A broader range of chemical structures will need to be evaluated, however, and some additional laboratories will need to confirm the finding. It must also be kept in mind that the metabolic systems and many of the other factors involved in species differences (as presented in Gad and Chengelis, 1999; Gad, 2007) contribute to a nonlinear correlation and may also make the confidence in prediction of human effects in cases somewhat limited.

Earthworms have been one of the more common species used to test chemicals for potential hazardous impact on the environment. The 48-h contrast test has proved to be a fast and resource-effective way of assessing acute toxicity of chemicals in earthworms and is outlined in Table 4. The standardized method, approved by the EEC, is discussed by Neuhauser et al. (1986). This test is for environmental impact assessment where cross-laboratory comparisons are important. If, however, one wishes to adopt this technology for the purpose of screening new chemicals or releasing batches of antibiotics, then variants of this method may be acceptable, as internal consistency is more important than interlaboratory comparisons. There are two important considerations. First, because of seasonal variation in the quality of earthworms obtained from suppliers, positive controls or comparator chemicals should be included on every assay run. Second, distilled water must be used, as worms are quite sensitive to contaminants that may occur in chlorinated water. The filter paper should completely cover the sides of the vessel; otherwise the worms will simply crawl up the sides to escape the adverse stimulus the chemical contact may provide.

Using these techniques, Roberts and Dorough (1984, 1985) and Neuhauser et al. (1986) have compared acute toxicity in a variety of organic chemicals in several earthworm species. While there are some obvious differences between worm species, in general the rank order of toxicity is about the same. *Lumbricus rubellus* tends to be the most sensitive species. All earthworms are very sensitive to carbofuran under the conditions of this test. Neuhauser et al.

**TABLE 4 Earthworm 48-h Contact Test—Acute Lethality**

1. Place filter paper of known size (9 cm, or 12 × 6.7 cm) in a Petri dish or standard scintillation vial.
2. Dilute test article in acetone or some other volatile solvent.
3. Slowly and evenly deposit known amounts of test article solution onto filter paper.
4. Dry thoroughly with air or nitrogen gentle stream.
5. Add 1.0 mL of distilled water to filter paper.
6. Add worm (*L. rubellus*). Use 400–500 mg body weight range.
7. Ten replicate vials per concentration.
8. Store/incubate in the absence of light at 15–20 °C for 48 h.
9. Examine for lethality (swollen, lack of movement upon warming to room temperature, lack of response to tactile stimulation).
10. Express dose as micrograms per square centimeter and mortality as usual. Calculate $LD_{50}$ using standard techniques.
11. Always include negative and positive (benchmark) controls.

(1985a,b) have proposed a toxicity-rating scheme based on acute lethality in the earthworms which is similar to the more familiar scheme based on acute toxicity in rodents (Table 5). Roberts and Dorough (1985) and Neuhauser et al. (1986) have published extensive compilations of acute lethality in worms and compared these with acute lethality in rats and mice. A selection of these is shown in Table 6. Applying the rating scheme of Neuhauser, most chemicals receive about the same toxicity rating based on results in *Eisenia foetida* and mice. This may suggest that replacing the $LD_{50}$ with the $LC_{50}$ for rating toxicity (e.g., for transportation permits) deserves serious consideration.

The main advantages of the 40-h contact test are the savings of time and money. The cost savings fall into three categories. First, earthworms are cheap. One hundred *L. rubellus* will cost about U.S. $2.00. The 100 mice they could replace in screens and quality control (QC) testing, for example, would cost $300–$400 (£150–£200) at an exchange rate of $2.00 = £1.00. Second, earthworms require no vivarium space, and their use could reduce the number of rodents used, resulting in a net decrease in vivarium use. Third, adapting the 48-h contact test would require little capital investment, other than a dedicated under-the-counter refrigerator set at 15–20 °C. Otherwise, the assay can be easily performed in a standard biochemistry laboratory. With regard to time savings, the standard lethality test with rodents requires 7–14 days of postdosing observations. The 48-h contact test is completed in 48 h. Not only is the turnaround time faster, but also the amount of time that technical personnel will have to spend observing animals and recording observations will be

**TABLE 5 Earthworm Toxicity: Toxicity Rating**

| Rating | Designation | Rat $LD_{50}$ (mg $kg^{-1}$) | *Eisenia foetida* $LC_{50}$ (μg $cm^{-2}$) |
|---|---|---|---|
| 1 | Supertoxic | <5 | <1.0 |
| 2 | Extremely toxic | 5–50 | 1.0–10 |
| 3 | Very toxic | 50–500 | 10–100 |
| 4 | Moderately toxic | 500–5000 | 100–1000 |
| 5 | Relatively nontoxic | >5000 | >1000 |

*Source*: From Neuhauser et al. (1985a,b).

**TABLE 6 Earthworm Acute Lethality: Comparative Values**

| Chemical | *Eisenia foetida* ($LC_{50}$) | Mouse ($LD_{50}$) |
|---|---|---|
| 2,4-Dinitrophenol | 0.6 | 45 |
| Carbaryl | 14 | 438 |
| Benzene | 75 | 4,700 |
| 1,1,1-Trichloroethane | 83 | 11,240 |
| Dimethylphthalate | 550 | 7,200 |

reduced. An incidental advantage of earthworms is that they are cold-blooded vertebrates and thus exempt from current animal welfare laws.

There are two main disadvantages to the use of earthworms in acute toxicity testing. First, there are a limited number of endpoints. Other than death and a few behavioral abnormalities (Stenersen, 1979; Drewes et al., 1984), the test does not yield much qualitative information. Second, there probably is some institutional bias. Because the test is basically low technology (no tissue culture) and uses a nonmammalian model, it may be easy to dismiss the utility of the test.

Finally, the use of smaller species of fish as a surrogate for humans has gained some supporters. Currently, the zebra fish has been shown to be a significant surrogate model for mammalian species toxicity (Hill et al., 2005; Chiu et al., 2008; Jeong et al., 2008; Tanguay and Reimers, 2008) and as a screen for carcinogens (Stern and Zon, 2003; Berghmans et al., 2005). There is certainly no reason why they could not be used for screening water-soluble compounds for extreme acute toxicity.

Although the intact organisms would seem to be the most utilitarian on the face of it, they still will not totally replace mammalian systems, owing to the need to be concerned about those systems that are significantly different in the higher organisms. Still, it would appear that for those compounds for which human exposure is not intentional, testing in an intact lower organism system (or perhaps even in a cell culture system) should be sufficient to identify agents of significant concern. In these cases, lethality testing in intact mammals is probably unwarranted.

### 5.1 Irritation of Parenterally Administered Pharmaceuticals

Intramuscular (IM) and intravenous (IV) injection of parenteral formulations of pharmaceuticals can produce a range of discomfort resulting in pain, irritation, and/or damage to the muscular or vascular tissue. These are normally evaluated for prospective formulations before use in humans by evaluation in intact animal models—usually the rabbit (Gad and Chengelis, 1999).

Currently, a protocol utilizing a cultured rat skeletal muscle cell line (the L6) as a model has undergone an interlaboratory validation program among more than 10 pharmaceutical company laboratories. This methodology (Young et al., 1986) measures creatine kinase levels in media after exposure of the cells to the formulation of interest and predicts in vivo intramuscular damage based on this endpoint. It is reported to give excellent rank-correlated results across a range of antibiotics (Williams et al., 1987) and in a recent multilaboratory evaluation a broader structural range of compounds (PMA, 1994).

Another proposed in vitro assay for muscle irritancy for injectable formulations is the red blood cell hemolysis assay (Brown et al., 1989). Water-soluble formulations are gently mixed at a 1:2 ratio with freshly collected human blood for 5 s, then mixed with a 5% w/v dextrose solution and centrifuged for 5 min. The percentage of red blood cell survival is then determined by

measuring differential absorbance at 540 nm, and this is compared with values for known irritants and nonirritants. Against a very small group of compounds (four), this is reported to be an accurate predictor of muscle irritation.

There is no current candidate alternative for the venous irritation test, but the in vitro alternative for pyrogenicity testing—the *Limulus* test—is one of the success stories for the alternatives movement. It has totally replaced the classical intact rabbit test in both research and product release testing. The test is based on the jelling or color development of a pyrogenic preparation in the presence of the lysate of the amebocytes of the horseshoe crab (*Limulus polyphemus*). It is simpler, more rapid, and of greater sensitivity than the rabbit test it replaced (Cooper, 1975).

### 5.2 Phototoxicity

ICI (Imperial Chemical Industries) has conducted studies on an in vitro phototoxicity assay which involves using three cultured cell lines: the A431 human epidermal cell line (a derived epidermal carcinoma), normal human epidermal keratinocytes (a primary cell line derived from cosmetic surgery), and the 3T3 Swiss mouse fibroblast cell line. The protocol for this assay involves subculturing the particular cell type into microtiter tissue culture grade plates and incubating them over a period of 24 h. Following incubation, the cultures are exposed to the test compound at a concentration predetermined as nontoxic. After 4 h exposure to the compound, the cell cultures are exposed to either ultraviolet (UV) A (320–400 nm) or UV A/B (280–400 nm) radiation for varying lengths of time. The degree of enhanced toxicity effected by either UV A or UV A/B radiation in the presence of the test compound relative to the control is assessed using the 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyl-tetrazolium bromide (MTT) assay, which undergoes a reduction reaction specific to mitochondrial dehydrogenases in viable cells. Work on validation of this test using 30 compounds of known phototoxic potential has shown a high degree of correlation between in vitro and in vivo results (Jackson and Goldner, 1989).

### 5.3 Hepatocyte

The liver extensively metabolizes and biotransforms xenobiotics, and the subsequent hepatotoxicity due to exposure to parent compound or its metabolites is a common clinical event. Isolated hepatocytes, in either suspension or monolayer culture, have been utilized to study the hepatotoxicity of many compounds (Klaasen and Stacey, 1982; Suolinna, 1982; Holme, 1985; Guillouzo, 1986). The utilization of monolayer cultures has several advantages: (1) monolayer cultures consist of viable cells while suspensions typically contain a mixed population of viable and dead cells; (2) monolayer cells can be maintained for a longer period of time, which is an important factor in studying chronic exposure of a compound; and (3) monolayer cultures allow cell-to-cell contact, which is important for studying the processes necessary for cellular

organization. The following section discusses the methods employed in the establishment of primary hepatocyte cultures and provides insights as to how hepatocyte cultures are a valuable model for short-term studies involving the safety assessment of xenobiotics.

***Methods*** The utility of primary hepatocyte cultures is most aptly illustrated by the number of species from which primary cultures can be derived. Isolation of hepatocytes from livers of small animals such as mice, hamsters, guinea pigs, and rats is generally performed in situ (Williams, 1976a,b; Dougherty et al., 1980; McQueen and Williams, 1981). In large animals such as rabbits, dogs, and monkeys, and even in human tissue, isolation of hepatocytes is performed by perfusion of biopsy liver specimen (Reese and Byard, 1981; Strom et al., 1982; Smolarek et al., 1990a,b).

*Preparation of Hepatocytes* In general, the harvesting of hepatocytes from the liver or liver specimen is performed by tissue perfusion using a peristaltic pump with two solutions: one consisting of 0.5 mM ethylene glycol bis (β-aminoethyl ether) *N,N′*-tetraacetic acid (EGTA) in Hank's balanced salt solution without $Ca^{2+}$ or $Mg^{2+}$ and the other consisting of 100 untis collagenase per milliliter Williams medium E (WME) buffered with 10 mM *N*-2-hydroxyethylpiperazine-*N*-1-ethanesulfonic acid (HEPES) and adjusted to pH 7.35 with 1 *N* sodium hydroxide (Maslansky and Williams, 1985; McQueen and Williams, 1985). The solutions are filter sterilized and kept at 37 °C. In situ perfusion involves cannulation of the portal or of the biopsy liver specimen; perfusion involves implantation of a large-gauge needle into the sinus cavities of the lobe followed by perfusion at a slow rate with the solution containing EGTA. The rate of perfusion and the volume of perfusate will vary depending on the size of the tissue. In the in situ method, the subhepatic inferior vena cava is severed to prevent excessive swelling of the tissue and to allow the perfusate to leave as waste. In the tissue specimen, perfusate diffuses freely out of the tissue. After uniform blanching of the liver is observed with the first solution, the solution containing collagenase is perfused through the tissue at a rate dependent of the size of the tissue. The liver is covered to keep it moist and a lamp is positioned above the liver to keep it warm. The liver is removed and placed into a sterile petri dish containing cold WME. Hepatocytes are dispersed into the medium by gentle teasing of the tissue. The cell suspension is then washed several times in WME containing 5–10% calf serum and 50 μg $mL^{-1}$ gentamycin sulfate. The viability is determined by trypan blue exclusion, and suspensions having viabilities of 9–100% are plated.

*Culture Conditions* As previously noted, the viability of freshly isolated hepatocytes in suspension does not usually exceed 5–6 h (Guguen-Guillouzo et al., 1988). Thus, the extended viability of hepatocytes requires culture techniques initially involving cell attachment to a support structure (Guguen-Guillouzo et al., 1988). There are a number of matrices that have been used

to maintain primary hepatocyte cultures: laminin (Ledbetter et al., 1984), collagen type IV derived from rat tail tendons (Seglen and Fossa, 1978), a preparation of collagen type IV and laminin in a solubilized basement membrane derived from a transplantable mouse tumor (Kleinman et al., 1986), commercially prepared plastic flasks that incorporate amide and amino function groups (Pittner et al., 1985), and cocultivation of hepatocytes with undifferentiated epithelial cells (Ratanasavahn et al., 1988). The success in maintaining primary hepatocyte cultures from different species on these matrices may vary considerably. In general, freshly derived hepatocytes are most easily maintained on collagen type IV–coated tissue culture flasks.

The tissue culture medium is another important factor in supporting survival and proliferation of primary hepatocyte cultures. Typically, hepatocytes are incubated at 37 °C, 5% $CO_2$–95% air in either WME media containing 50 μg mL$^{-1}$ gentamycin and 1–10% fetal calf serum (McQueen and Williams, 1985) or Eagle's minimum essential medium supplemented with Earle's salts, nonessential amino acids, 50 μg mL$^{-1}$ gentamycin, and 5% fetal calf serum (Seglen, 1976). The initial culture medium is removed 2–4 h postplating and replenished with fresh media with serum or without serum. Due to the significant losses of cytochrome P-450 content during the first day in culture, much effort has been devoted to developing culture media devoid of fetal calf serum and supplemented with a variety of substances. Supplements such as aminolevulinic acid (Hockin and Paine, 1983), ascorbic acid (Guzelian and Bissell, 1974), adenine (Guzelian and Bissell, 1974), nicotinamide (Decad et al., 1977), dimethylsufoxide (Isom et al., 1985), and selenium (Paine and Hockin, 1980) are reportedly important for the maintenance of P-450 activities in primary hepatocyte cultures. In addition, long-term exposure of the hepatocyte cultures to fetal calf serum has been found to be deleterious by inhibiting expression of liver-specific functions, being cytostatic at all plating densities and being cytotoxic at low seeding densities (Enat et al., 1984).

### *Specific Applications*

*Cytotoxicity* The liver is the primary target organ for a variety of drugs and chemicals (Haseman et al., 1984; Farland et al., 1985). The prevalence of drug- and chemical-induced liver injury is of concern because some xenobiotics can produce liver damage at dose levels that are magnitudes below that which causes cell death (Plaa, 1976). Environmental and commercial chemicals can increase this effect by as much as 100-fold (Plaa and Hewitt, 1982; Plaa, 1976). Studies of early cell injury caused by exposure to a toxicant can be undertaken easily in monolayer cultures of hepatocytes, whereas early cell injury is very difficult to assess in vivo.

The methods that have been used to determine cytotoxicity in primary hepatocyte cultures include morphological changes, assessment of enzyme leakage, and membrane blebbing (Jewell et al., 1982; Story et al., 1983; McQueen et al., 1984). Together, these endpoints depend on the compound

tested and, in general, at least two endpoints should be monitored. Commonly used indicators of hepatocyte toxicity that suggest impaired cell function but not necessarily cell death include leakage of lactate dehydrogenase (Story et al., 1983), changes in cell morphology (McQueen et al., 1984), and inhibition of protein or DNA synthesis (Shaw et al., 1975; Seglen et al., 1980). Indicators of cell death include exclusion of neutral red dye, inclusion of trypan blue (Williams, 1977), and the loss of cell attachment (Simmerman et al., 1974). Gomez-Lechon et al. (1988) studied the use of primary hepatocyte cultures in predicting the hepatotoxicity of xenobiotics. Four well-documented indirect hepatotoxins (α-amanitin, D-galactosamine, thioacetamide, and acetaminophen) were studied in cultured rat hepatocytes and the results compared with the toxicity in vivo (Gomez-Lechon et al., 1988). The data indicated that when functional parameters were utilized, toxicity occurred in vitro at concentrations at which no effects were seen in vivo. These results indicate the sensitivity of the assay and the necessity of selecting the correct indicators of toxicity.

Primary hepatocyte cultures have been used as a tool to predict the hepatotoxicity of many compounds such as nonsteroidal anti-inflammatory drugs (Castell et al., 1988), psychotropic drugs (Boelsterli et al., 1987), immunosuppressant drugs (Boelsterli et al., 1988), and salicylates (Tolman et al., 1978). Rat primary hepatocyte cultures have also been shown to be a good model for examining the mechanisms of metallothionein-induced tolerance to cadmium toxicity (Liu et al., 1990). Guillouzo et al. (1988) developed a coculture system of rat or human hepatocytes with rat liver epithelial cells that maintains the hepatocytes in a differentiated state for extended periods of time, thereby allowing studies involving chronic treatment with the test substance to be conducted. Primary cultures of hepatocytes can therefore provide a useful model for short- and long-term studies involving the safety assessment of xenobiotics.

*Comparative Metabolism* Since the liver is the major organ involved in the biotransformation of xenobiotics, primary hepatocyte cultures provide an excellent model for in vitro metabolism studies. Primary hepatocyte cultures provide useful tools with which to study the comparative metabolism of xenobiotics by both humans and laboratory animals.

Primary hepatocyte cultures undergo a significant decrease in the activities of phase I and II enzymes that correlates with time in culture. Croci and Williams (1985) compared, during the first 24 h in culture representative phase I and phase II biotransformation pathways in hepatocyte primary cultures isolated from male and female rats to freshly isolated hepatocytes. Hepatocytes lost 50% of cytochrome P-450 activity during the first 24 h in culture but maintained high mixed-function activities; 75% of aryl hydrocarbon hydroxylase and 65% of benzphetamine demethylase activities were preserved in hepatocytes from male rates. Uridine 5′-diphosphate (UDP)–glucuronosyl transferase activities were slightly increased during 24 h of culture to levels higher than present in liver tissue before perfusion. Glutathione transferase

activity after 24 h diminished to 20% of the initial enzyme activity for one form while another form was stable. Donato et al. (1990) showed that in human hepatocytes 3-methylcholanthrene, phenobarbital, and ethanol can increase the activity of cytochrome P-450 monooxygenases, aryl hydrocarbon hydroxylase, and 7-ethoxycoumarin *O*-de-ethylase.

Hepatocyte cultures are not affected by some of the variables that influence in vivo studies such as absorption and distribution; accordingly, the exact concentration of the parent compound is known, and the total number and amount of each metabolite formed over time can be accurately determined. Species-specific differences in the metabolism of xenobiotics may, in fact, play a major role in determining relative susceptibilities to chemical toxicants. As an example, studies were recently conducted utilizing primary hepatocyte cultures derived from the perfusion of livers from rats, dogs, and monkey to determine if the species-specific differences in acetaminophen (APAP)–induced cytotoxicity were correlated with species-specific differences in the amounts of APAP metabolized and the formation of APAP conjugates (Smolarek et al., 1990a). In vivo studies on the hepatotoxicity of APAP have shown a species-specific difference in susceptibility to its hepatotoxic effects (Davis et al., 1974). The dog, hamster, and mouse are very sensitive, but the rabbit, guinea pig, and rat are resistant. Rat, rabbit, dog, and monkey hepatocyte cultures were exposed to 2 mM APAP for 24 h (Smolarek et al., 1990a). Aliquots of hepatocyte culture media containing APAP and its metabolites were analyzed by reverse-phase high-pressure liquid chromatography (HPLC). The sensitivity of the dog hepatocytes to APAP was directly related to low conjugating enzyme activity. In contrast, monkey hepatocyte cultures had a very large capacity to transform APAP to glucuronide conjugates and a very high level of glutathione *S*-transferase activity, which correlated with their resistance to cytotoxicity. These studies indicate that the competing pathways of APAP conjugation in hepatocyte cultures from different species explain the differences observed in APAP-induced cytotoxicity (Smolarek et al., 1990a). Similar comparisons have also been made relating species-specific tetrahydroaminoacridine (THA)–induced hepatotoxicity to differences in the rate and extent of THA biotransformation in cultured hepatocytes from rat, dog, and monkey (Smolarek et al., 1990b). These studies demonstrated that the hepatocytes from these three species differed in their sensitivity to concentration-dependent cytotoxicity, with the monkey cells being most sensitive and the canine cells least sensitive to THA-associated cytotoxicity. Tetrahydroaminoacridine biotransformation also differed among the three species, with the canine hepatocytes most effective and monkey hepatocytes least effective in the conversion of THA to most polar metabolites (Smolarek et al., 1990b). These in vitro studies correlate differences in the cytotoxicity of THA to its biotransformation. They also suggest that the monkey would be a good animal model for in vivo THA toxicity testing (Smolarek et al., 1990b).

Green et al. (1986) compared the metabolism of amphetamine in isolated hepatocyte suspensions from rat, dog, squirrel, monkey, and human livers. The

metabolite profile of hepatocytes from each species corresponded to the profile of urinary metabolites identified previously. These results indicate that species-specific differences in the metabolic activation of compounds seen in vivo can be reproduced in vitro by the utilization of primary hepatocyte cultures.

Primary hepatocyte cultures have been used in vitro to metabolically activate toxins for evaluation with target tissues since the early 1980s. Cocultures of rat embryos with hepatocytes have been used to study the role of metabolism in teratogenesis (Oglesby et al., 1986). Lindahl-Kiessling et al. (1989), in an attempt to bring test conditions closer to in vivo conditions, developed an assay utilizing primary rat hepatocytes and human peripheral lymphocytes to detect metabolism-medicated mutagenesis. Doehmer (2006) has published on the development and use of genetically engineered cells as more metabolically relevant predictors for human drug metabolism.

*Genotoxicity* The genotoxicity of drugs and chemicals can be detected by measurement of their interactions with cellular DNA. Many test chemicals are not genotoxic by themselves and thus may require metabolic activation to a DNA-reactive metabolite. Primary hepatocyte cultures function (1) metabolic activating systems for chemicals and (2) target cells for the interaction of reactive metabolites with hepatocellular DNA. Indirect assessment of human genotoxicity can be implemented when primary human hepatocytes are used to assess a chemical genotoxicity. In primary hepatocyte cultures, studies to resolve the extent of xenobiotic–DNA interaction products formed as a result of exposure to a test chemical determine the total amount of xenobiotic bound per milligram DNA (Poirier et al., 1980) and quantify the number of single-strand beaks in DNA by alkaline elution (Bradley et al., 1982). An indirect method for determining the genotoxicity of xenobiotics from a wide variety of structural classes in primary hepatocyte cultures is to measure DNA repair by autoradiography (Williams, 1976a). McQueen and Williams (1983) tested a large number of chemicals utilizing this technique and reported that all of the known noncarcinogens tested were negative in eliciting DNA repair, while approximately 90% of the known carcinogens were positive. The suggested reasons for 10% of the carcinogens not eliciting DNA repair are that many carcinogens are of the epigenetic type and therefore will not produce DNA damage and, second, that the carcinogen may inhibit DNA repair. Audioradiographic measurement of DNA repair allows cells undergoing replicative DNA synthesis to be readily distinguished from those in repair, and it allows DNA repair to be studied at concentrations of compound that are not cytotoxic.

Species-specific differences in DNA repair by primary hepatocyte cultures have been demonstrated by McQueen and Williams (1983). The relative resistance of the mouse to aflatoxin $\beta_1$–induced carcinogenesis was shown by the 10–100 times higher concentration of aflatoxin $\beta_1$ necessary to cause maximum DNA repair in the mouse than in other species tested. Administration of safrole to primary hepatocyte cultures elicited a positive response in DNA repair for the hamster and mouse hepatocyte cultures but was negative in the

rat hepatocyte cultures, suggesting the importance of multispecies genotoxicity testing of a compound (McQueen and Williams, 1987). With respect to the toxicological testing tier, genotoxicity testing contributes in the overall safety assessment of new drug candidates, and it can also be applied at the preproject stages of evaluation.

***Summary*** It is clear from the preceding discussion that the primary hepatocyte in culture represents a versatile in vitro tool in the safety assessment process. The applications of the hepatocyte are summarized in Table 7. First, they provide a useful model for studying drug- and chemical-induced hepatotoxicity. Second, they afford the potential to generate and examine the toxicity of phase I and II metabolites to liver cells or other target organs in vitro. Third, they provide a mechanism to examine potential species differences in metabolism of drug candidates prior to in vivo studies. Early in the drug discovery process, primary hepatocyte cultures can therefore aid in the selection of a chemical series based on favorable metabolic and toxicological profiles. The choice of the animal species appropriate for in vivo toxicology studies could also be based on data obtained from primary hepatocyte cultures of several species. Lastly, utilization of primary hepatocyte cultures derived from livers of humans is becoming more commonplace. The toxicologist's ultimate goal of determining hepatotoxic (and genotoxic) risk to humans should be greatly enhanced through the utilization of primary human hepatocyte cultures.

***Needs for the Future*** Primary hepatocyte is a well-established technology for studying the pharmacology and toxicology of different classes of xenobiotics. However, since there are currently no well-defined experimental conditions for establishing primary hepatocyte cultures, it is difficult to compare experimental results from different laboratories. The major variable consists of the cell culture media used for maintaining hepatocytes. Although each specialized medium has been shown to be useful for a specific application, a real need exists to establish an optimal medium that can be standardized. A common objective of most of the varieties of cell culture media is to define a system that maintains cytochrome P-450 activity and the hepatocellular differentiated state. Conventional cell culture conditions have only maintained approximately 20–40% of cytochrome P-450 activity in rodent hepatocytes cultured for two days, while human hepatocyte culture seemed to be more stable and maintained approximately 50–60% of their initial P-450 activity after five to six days in culture (Guguen-Guillouzo et al., 1988). Development

**TABLE 7 In Vitro Testing Utilizing Hepatocyte Cultures**

| |
|---|
| Cytotoxicity mechanisms → Cellular damage and necrosis |
| Comparative metabolism → Species specificity |
| Teratogenic mechanisms → Proteratogens |
| Genotoxic mechanisms → DNA repair |

of a well-defined matrix upon which to grow primary hepatocytes also needs further consideration. The greater degree of loss of P-450 activity in primary hepatocytes in suspension relative to monolayer cultures suggests that cell-to-cell contact is an important issue. A procedure that should contribute some degree of uniformity to primary hepatocyte culture as a test system is to verify the status of the cultures prior to utilization in a particular way in a particular assay. This could be accomplished by routinely checking standard cytotoxic parameters such as leakage of cytoplasmic enzymes or cell survival as well as metabolic parameters such as protein synthesis and gluconeogenesis.

In light of the interest and value of studying the responses of human hepatocytes relative to other species, an important need for the future is to develop successful techniques to culture and freeze preparations of human hepatocytes which, upon thawing, maintain their viability and metabolic capabilities. Typically, human liver tissue is obtained in large quantities. Since only a small portion of the total tissue is needed to provide sufficient material for a given experiment, the remaining portion could be frozen for future use if the proper freezing techniques were established. As previously noted, there has been some success in coculturing human hepatocytes with liver epithelial cells for long-term cultures useful in chronic toxicity testing (Guillouzo et al., 1988) and for maintaining the viability of previously frozen human hepatocytes (Li et al., 1990). Recent advances in culturing techniques should provide a well-defined experimental system for culturing primary hepatocytes in the near future. The role of primary human hepatocyte cultures in safety assessment will become increasingly important, and therefore successful efforts in freezing hepatocytes will help to meet the increasing needs in pharmacological and toxicological studies.

### 5.4 Ocular Alternatives

In recent years, much attention and effort have been directed toward the search for non-whole-animal tests to predict ocular irritation of drugs and chemicals. A variety of in vitro assays as well as "nonexperimental" approaches have been proposed (Williams, 1985; Gad, 2000). These model systems run the gamut of responses observed in vivo using biological systems encompassing a range of organization (from cells to whole eye) and measure a multitude of responses. In addition to these experimental tests, one may also think about predilection of ocular irritation by means that require no additional animal work. The use of literature or computer databases, computer modeling, and prediction from physical or chemical parameters or other toxicity data (such as dermal or acute toxicity) are examples of such an approach. Given a plethora of possibilities, where does one begin in trying to develop a program for establishing ocular irritation potential without the use of live animals? What are the general strengths and weaknesses of the existing techniques and what developmental work needs to be done? How does one "validate" a model system and use it in the decision-making process? These issues are addressed in this section.

### *Methods*

*"Nonexperimental" Techniques* Perhaps the most obvious and practical starting point for irritation potential to the eye is to ask what is already known about a compound or related structures. If enough data exist, perhaps further experimental work will be unnecessary. Such data might include (1) information from the literature or other databases, (2) data generated by computer modeling, (3) physical and chemical characteristics of the test compound, or (4) previously obtained toxicity data such as acute and/or dermal toxicity. The use of each of these approaches has its advantages and disadvantages, which are discussed below.

PUBLISHED DATA One might assume that previously collected data appearing in the literature should be quite useful and reliable in evaluating potential ocular irritancy. In practice, however, two basic issues detract from the utility of the literature databases. The first is illustrated in a paper by Weil and Scala (1971). In their study analysis a number of laboratories tested the same set of compounds for ocular irritation potential in both a standard protocol and in the routine protocol of the particular laboratory. Comparisons of the results showed considerable variability between laboratories, and to some degree within the laboratories, in the ranking of test compounds. These findings are not surprising in light of the subjective nature of the scoring systems, which makes it difficult to achieve consistent results among individuals and laboratories. How, then, does one determine which literature data to accept and which to reject?

The second issue involves the nature of published material itself. Due to journal space limitations and/or other factors, it is rare to find enough detail presented for a reader to make an independent evaluation of the value of the irritation data. Sometimes the test methodology is not defined, making direst comparisons to results obtained in other laboratory situations tenuous at best. Often raw data are lacking, and the results are reduced to either a plus–minus score or a broad categorization (mild, moderate, severe), so that a more specific appraisal cannot be made. These factors preclude the resolution and evaluation of seemingly conflicting data that one often finds in the literature.

A possible means of overcoming these problems is to use information generated within one's own company. In-house data are usually more consistent since the methodology generally does not change substantially over time, and one usually has the further advantage that the original raw data are available. The in-house database is also more likely to contain information on compounds whose chemistry is similar to the unknowns to be tested. However, irrespective of the source of background data on related substances, such information can only provide a starting point for the investigation of an unknown compound. While the data may be potentially useful in raising a warning flag, they are unlikely to be sufficient for a definitive judgment.

COMPUTER MODELING Computer modeling shares essentially the same problems as literature data. A computer simulation can only be as reliable as the data used in its construction. A model system called TOPKAT is available commercially from Health Designs. According to its own literature (*HDI Toxicology Newsletter*, 1987; TOPKAT manual), eye irritation has been quite difficult to model due to the inherent variability of classification of compounds. Health Designs has, in fact, had to design two sets of equations: one to separate nonirritants from all other compounds and the second to separate severe irritants from the rest, leaving a fair number of compounds broadly classified between these extremes. Furthermore, the model is predicted to be unsuitable for approximately 30% of chemical structures, and the number of indeterminates is higher than the developers would like. It appears that computer modeling, while promising, is somewhat limited as a definitive predictive tool. As with data from the literature, one can probably use the computer model to obtain a rough idea of irritation potential, but more definitive categorization will likely require further testing.

PHYSICAL AND CHEMICAL PARAMETERS Knowledge of the physical and chemical parameters of a compound should be a useful point of reference in predicting irritation potential. In fact, many companies have already reduced animal testing by using a rule of thumb with regard to the pH of a material. Compounds with an exceedingly high or low pH (for instance <3 or >12) are presumed to be irritants and would simply be labeled as such without further testing. This practice has some support in the literature (Guillot et al., 1982b), and further work in this area is ongoing in a study sponsored by the Soap and Detergent Association (SDA). A preliminary finding in the SDA study is that the alkalinity of a compound (i.e., the strength of the acid or base) may be more important than the simple measure of pH (Booman et al., 1989).

The correlation (or lack of correlation) of other physiochemical characteristics has not yet been established. For instance, are all surfactants irritants? Can one classify severity by the size of the molecule? Can octanol–water partition coefficients predict irritation potential; does a propensity to partition out of the ocular fluid mean that a compound presents more of an irritation hazard than one which is more water soluble? Theoretically, these data should reflect the ability of a compound to penetrate the eye and cause an irreversible lesion. However, until definitive data are available, physical and chemical parameters will probably have limited utility in an overall assessment of irritation.

DATA FROM OTHER TOXICOLOGICAL TESTING Another question under study has been whether dermal or systemic (acute) toxicity is predictive of ocular irritant potential. Little work has been reported in the literature regarding the correlation between ocular irritation and acute toxicity, so one can only speculate as to this relationship. When one considers the multiplicity of physiological mechanisms involved in acute lethal reactions [e.g., at the cardiovascular and central nervous system (CNS) levels], it is difficult to envision homology with equally

complex responses of tissue damage and inflammation observed at the ocular site. Thus, from a correlative viewpoint, it is unlikely that mechanisms of lethality operating in acute toxicity tests would also occur in ocular injury, thereby making any predictive analysis strictly fortuitous.

However, more extensive information is available comparing dermal and ocular irritation. Gillman et al. (1983), for example, examined dermal and ocular irritation data on selected petrochemicals and consumer products and reported no reliable correlation between ocular and dermal irritation scores. Guillot et al. (1982a,b) examined ocular and dermal irritation of 56 compounds, comparing different protocols. If one combines the data from both papers, one finds that of 11 dermal irritants, all showed ocular irritation, but the extent of irritation was not predictable. Of the 45 compounds that were nonirritating to slightly irritating dermally, only 18 were nonirritating or slightly irritating to the eye. The remaining 27 compounds ranged from mild or moderate to extremely irritating upon ocular exposure. Similar results were reported by Williams (1984, 1985), who examined 60 severe dermal irritants and found 39 to be severe ocular irritants, while 6 were moderate irritants and 15 mild to nonirritant in the eye. Lastly, in a study by Gad et al. (1986), it was revealed that correlations between dermal and ocular data differed dramatically depending on the scale of irritation used for the comparison. When the compounds were classified simply as positive or negative in both dermal and ocular irritation, the correlation was better than when one attempted to predict a specific classification (i.e., negligible, mild, moderate, severe). The only generalization that can be made is that if a compound is a severe dermal irritant, it is likely to be a strong ocular irritant. However, as alluded to previously, severe irritancy can be predicted, in many cases, on the basis of chemical properties alone (e.g., pH). Nonetheless, it appears that dermal irritancy data can be predictive of ocular irritation for severe dermal irritants. However, as cited above, a number of false positives will occur. Caution is therefore advised in the use of this parameter.

To summarize the utility of "nonexperimental" methods, it is obvious that the more available information there is about a compound, the more likely one will be able to substantially reduce the amount of testing involved in prediction of ocular irritation potential. However, at this point in time, none of the individual methods, alone or in combination, are sufficiently predictive to provide a definitive assessment of in vivo ocular irritation. There is definitely a place, however, for consideration of the above factors in a battery of tests as well as for prioritizing compounds to be tested further.

*In Vitro Methods* In view of the limitations of extrapolation of irritancy by nonexperimental methods, it is clear that in most cases new experimental data will need to be generated. This new information can be obtained using in vitro approaches in conjunction with in vivo data, when necessary. With respect to selecting an in vitro procedure, the first step would be to evaluate what processes one is attempting to model followed by the development of

a methodology to mimic that particular biochemical effect. However, while the broad processes of inflammation, opacity, and the like are known to be involved in ocular injury and irritation, a detailed understanding of the underlying mechanisms of these processes in the eye is lacking. As previously mentioned, in examining criteria for in vitro test systems, it is extremely useful to know enough about the underlying mechanism(s) to identify and measure a specific endpoint that is (preferably) causal to the in vivo effect. For instance, is the process of opacification due to osmotic imbalances, protein coagulation, necrosis of cells within the epithelial organization, or stromal swelling due to ionic interference with cell-to-cell junctional complexes or (more likely) does the mechanism differ with the compound being tested? And what is really known about inflammation at the cellular and molecular levels? Which components of the arachidonic acid cascade are involved? Do these components act synergistically or antagonistically, and what about specificity of cell response? Should corneal epithelial or endothelial cells, conjunctivae, or stroma be examined as the primary tissue affected? Some of the answers are known, but our deficiencies in knowledge and understanding highlight the complexity of modeling ocular responses to toxins.

Given the mechanistic complexity of the ocular response to xenobiotics, attempting to predict irritation with one assay may not be realistic at this time. The type of approach more likely to be immediately fruitful, as suggested by a number of people [Fielder et al., 1987; European Centre for Exotoxicology and Toxicology of Chemicals (ECETOC), 1988], is to use a tiered approach, taking advantage of as wide a diversity of methods as is practical to get a comprehensive picture of irritation potential. Indeed, various regulatory agencies have unofficially stated that they would like to see all components of the Draize score addressed before they would consider in vitro data in lieu of animal data. Thus, a logical approach would be to examine an assay(s) for each component of ocular irritation (i.e., opacity, inflammation, necrosis, or toxicity).

Because the materials tested for ocular irritancy in the pharmaceutical industry are diverse and many are novel chemical entities, it is advantageous to identify an in vitro assay that closely mirrors the target tissue biochemistry and physiology and monitors the specific endpoint of interest. The expectation of such a system would be that the methodology applied would detect compounds exerting toxic effects by a variety of mechanisms. An extensive list of proposed alternatives has been compiled and categorized by Frazier et al. (1987). While this listing can serve as a starting point as to the available technology, many of these methods do not possess any obvious correlation with the target tissue of interest and/or are not well validated. The following sections provide an overview of one or two representative methodologies for each component of the Draize system along with reported advantages and disadvantages.

OPACITY Corneal opacity is the most heavily weighted of the components of the Draize eye score (80 out of 110 possible points) (Conquet et al., 1977).

Thus, an in vitro system that provides an accurate measure of opacification should contribute substantially toward in vitro modeling of the classical Draize test. Two assays attempting to model this process are discussed.

A bovine corneal opacity (BCO) assay described by Muir (1984, 1985) seems to be quite promising as a model for studying opacification. Recently, this technology has been further developed for the prediction of drug- and chemical-induced ocular irritation (Gautheron and Sina, 1990; Gautheron et al., 1992). Basically, the BCO assay uses freshly isolated bovine eyes obtained from an abattoir. Alternatively, whole corneas may be obtained for experimental purposes from other species such as the rabbit (Elgebaly, personal communication, 1990) or pig (Igarashi et al., 1989). Corneas are removed, allowed to equilibrate in medium for a period of time, then mounted in a special chamber in which both sides of the cornea are bathed in fluid. Test compound is added to the epithelial side of the cornea for a defined period of time and opacity is measured at intervals using an opacitometer. An evaluation of commercially available compounds (Table 8) as well as a substantial number of pharmaceutical candidates and process intermediates has shown an excellent correlation between the BCO assay and in vivo irritation scores. A few severe ocular irritants give a false-negative reading in this test due to the fact that they cause the entire epithelial layer to slough from the cornea, thereby producing a lower opacity reading than expected. This difficulty can be overcome by simple examination of the cornea postexposure. The only other drawback encountered has been observed with some insoluble compounds. These materials are generally not a problem (although the highest dose tested is limited), since the corneas can be washed prior to obtaining an opacity reading. However, in a few cases compound can adhere too tightly for effective rinsing, thus giving an artificially high opacity reading. Overall, the BCO assay has been found to be an excellent first-line test in a battery of alternatives, and interlaboratory evaluation studies are currently being initiated in Europe.

A commercially available methodology (Eytex) has recently been put forth as an alternative corneal opacification assay. The method, based upon the presumption that opacification is due to coagulation or denaturation of protein, measures the precipitation caused in a protein matrix by the introduction of a test compound (Gordon and Kelly, 1989). The developers have tested a number of materials, a substantial number of which are surfactants or surfactant-based products, and claim a high level of correlation with ocular irritation. However, evaluation of the methodology has been undertaken by various researchers with mixed results. For instance, on the one hand, Lawrence-Beckett and James (1990) and Soto et al. (1989) found the test useful for evaluating industrial chemicals. On the other hand, Bruner and Parker (1990) found that the Eytex assay was not very predictive of household product irritancy compared with five other in vitro assays. One potential reason for this type of discrepancy may be found in the evaluation performed by Thomson et al. (1989a). These researchers tested a series of compounds within a variety of groups of formulations and found good correlations with in vivo data when

**TABLE 8 Comparison between In Vivo Ocular Irritancy and Opacity Induced in Bovine Cornea**

| Compound | In Vivo Irritancy | Opacity at 100% concentration[a] |
|---|---|---|
| Tween 20 | Mild[b] | 1.7 |
| Dimethylsulfoxide | Mild | 11.7 |
| Ethanol | Mild/moderate[b] | 20.0 |
| Ethylene glycol monomethyl ether | Mild/moderate[b] | 24.0 |
| Isopropyl alcohol | Mild/moderate[b,c] | 30.1 |
| Methanol | Mild/moderate[b] | 32.1 |
| Ethylene glycol monoethyl ether | Mild/moderate[d] | 33.2 |
| Carbitol | Mild/moderate[b] | 36.5 |
| Formamide | Mild/moderate[b] | 51.8 |
| Acetone | Mild/moderate[b] | 65.5 |
| Tetrahydrofurfuryl alcohol | Moderate[c] | 71.2 |
| Ethylene glycol monobutyl ether | Moderate[d] | 72.3 |
| Tetramethyl ethylenediamine | Severe[b] | 80.7 |
| Solketal | Moderate[c] | 87.7 |
| Acetonitrile | Severe[e] | 88.5 |
| Pyridine | Severe[b] | 92.3 |
| Allyl alcohol | Severe[b,c] | 123 |
| Trichloroacetic acid | Severe[b] | 219 |

[a]Opacity scaling described in Gautheron et al., 1992.
[b]Grant, 1986.
[c]Clayton and Clayton, 1981.
[d]Smyth et al., 1951.
[e]Smyth et al., 1949.

the Eytex assay was used for surfactant blends and eye-area-use products. However, the correlation was not acceptable when alcohol-containing formulations were tested. This illustrates a recurring caveat in the field of in vitro toxicology: A technique must be evaluated with materials from a variety of chemical classes and preferably with compounds that are likely to be tested as unknowns before it can be considered a "validated" assay.

CYTOTOXICITY TESTING The majority of proposed alternative tests are assays for cytotoxicity in which a target cell is exposed to test compound and some endpoint of viability is measured. The advantages of such tests are that these methods are usually very easily and rapidly performed and can be readily transferred among different laboratories. Furthermore, some investigators report that the choice of target cell is relatively unimportant to the predictivity of the assay (Borenfreund and Borrero, 1984), suggesting that the choice of the target cell–ending combination is essentially unlimited. However, despite these advantages, cytotoxicity assays cannot, in general, be considered mechanisticaly based and may not be appropriate for every situation. While it is likely that a number of compounds, particularly severe to extreme irritants, damage the eye through overt cellular toxicity, it is equally clear that a number of compounds exert their effects by different mechanisms. The need for judicious

use of the cytotoxicity endpoint, preferably as part of a battery of approaches, is borne out by a number of evaluations reported in the literature (Bracher et al., 1987; Flower, 1987; Kennah et al., 1989).

Consider, for example, the neutral red assay developed by Borenfreund and Puerner (1984, 1987). This test has been evaluated with a number of surfactants and cosmetic ingredients and has reportedly given good correlations with in vivo ocular irritation data (Bracher et al., 1987; Shopsis, 1989; Bruner and Parker, 1990). However, Thomson et al. (1989b) used this assay to examine a variety of products and found high correlations with some types of materials and poor correlations with others. Again, the basis for this variation in assay performance seems to be the class of chemical being evaluated.

Recently, the cytotoxicity of various drug candidates and process intermediates, as well as commercially available compounds, was examined in the presumed target cell—rabbit corneal epithelium—and a nontarget cell, V79 fibroblasts (Sina and Gautheron, 1990; Sina et al., 1992). Two endpoints of toxicity were monitored: leucine incorporation into protein, a general measure of toxicity, and the tetrazolium dye assay (MTT), a measure of mitochondrial dysfunction. An $IC_{50}$ (concentration necessary to reduce endpoint to 50% of control value) was determined for each cell–endpoint combination, and correlations were drawn with the in vivo classifications (i.e., mild, mild/moderate, moderate, severe). The data showed such a broad overlapping of the $IC_{50}$ values that no definitive cutoff could be established between classifications (Figure 3). Subsequently, an $IC_{50}$ threshold was established simply to distinguish severe in vivo irritants. Again, such a value was difficult to determine. If one evaluates commercially available and in-house products separately, a threshold value that allows approximately 70–75% predictivity between these two broad categories can be established. However, this value changes dramatically with the types of compounds as well as with the different target cell–endpoint combinations. It is therefore of limited value in testing unknowns.

Thus, cytotoxicity assays are unlikely to provide an adequate predictor of in vivo irritation in every case. There is no doubt that some type of compounds (such as surfactants) will give (and have given) acceptable correlations with in vivo data. However, for the diversity of compounds common in the pharmaceutical industry, cytotoxicity assays alone are inadequate predictors of ocular irritation, though they may have a place in a battery of tests. For instance, if one is interested in a very quick assessment that will be corroborated later, cytotoxicity assays may be indicated. But each laboratory needs to make its own evaluation of the utility and value of these methods.

INFLAMMATION To complete the examination of the major components of the Draize score, one also needs to examine inflammation. Although inflammation is not a large portion of the Draize score (20 of 110 total points), most people appear to put a great deal of weight on this component due to the subjectivity inherent in the scoring. Furthermore, an evaluation of in-house historical data (J. Sina, 1994, unpublished) indicates that, of those compounds that cause

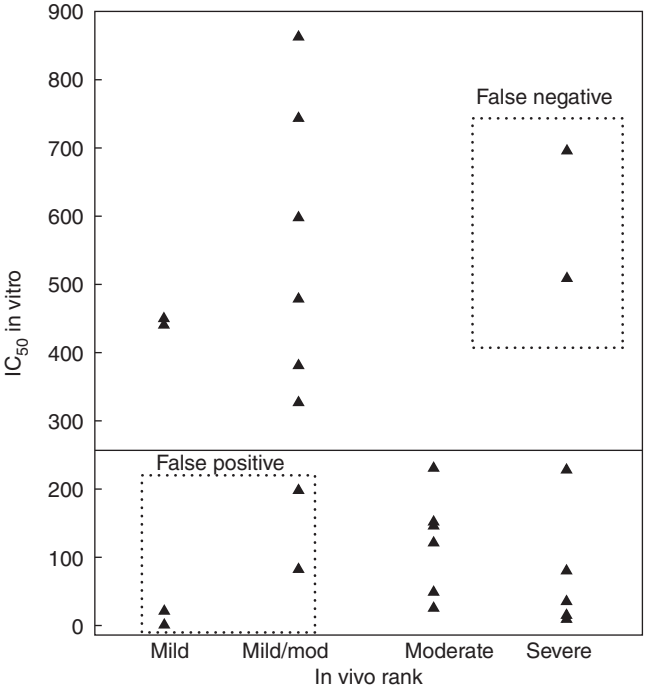


***Figure 3*** *Comparison of cytotoxicity in corneal cells (leucine incorporation) and ocular irritancy in vivo.*

opacity, the majority show moderate to severe inflammation first. This suggests that inflammation may be a more important and predictive component of ocular irritation than the Draize scoring system would suggest.

One approach to examining inflammation is the assay reported by Elgebaly et al. (1987) for the release of chemotactic factors. Earlier work has elegantly shown that neutrophil or macrophage infiltration from either the endothelial or epithelial surface can cause substantial damage to the eye (Elgebaly et al., 1985). Thus, it seems likely that release of inflammatory mediators will correlate with ocular irritation or, more important, with the potential for ocular damage. Briefly, the methodology involves the exposure of the isolated bovine cornea (or cornea from other species) to a compound, with the intact structure serving as a cup to contain the test material. After an indicated time, the exposure medium is harvested and analyzed for the release of chemotactic mediators by means of a chemotaxis assay. Conceptually, this assay is quite attractive and validation is currently underway.

Another approach that has been proposed under the category of inflammation assays [though it has been suggested (Lawrence, 1987) that the assay measures necrosis rather than the true inflammation] is the chick chorioallantoic membrane (CAM) test and its modifications (Leighton et al., 1983;

Luepke, 1985; Kong et al., 1987; Bagley et al., 1989). Basically the assay scores alterations (vascularity, necrosis, etc.) in the chorioallantoic membrane of chicken eggs upon exposure to test compounds. As with most of the other assays discussed here, the CAM assay has proven useful in some applications but inadequate in others. For instance, Bagley et al. (1989) report good predictivity of ocular irritation for surfactant-based materials with a CAM modification, the CAMVA method. By contrast, Price et al. (1986) and Lawrence et al. (1986) found the assay inadequate for their purposes. The reason(s) for these discrepancies may be due to the compound classes tested or to various methodological differences between laboratories. Another point worth mentioning with regard to this assay is that since a viable chicken embryo is part of the model, the CAM may not be truly considered an in vitro model, particularly in Europe (Lawrence, 1987).

***Specific Applications*** How are the above-proposed tests integrated into the safety assessment process? An important issue that arises (and impinges upon the number of assays or types of data required) relative to answering this question is the scheme for decision making. Some approaches have been suggested (Fielder et al., 1987; ECETOC, 1988). For instance, the most straightforward approach would be to gather as much data as possible by nonexperimental methods such as those described above, then proceed to in vitro testing, and finally, if required, perform an animal assay. But, do all steps in the tier need to be performed? Will nonexperimental methods be sufficient? If in vitro assays are needed, how many should be performed? And if different assays are inconsistent, how does one decide which data are more accurate? Unfortunately, the answers to questions such as these are elusive. The manner in which alternative methods will be used in decision making will likely depend both on the individual company, and, within a company, on the reason for testing (i.e., how precise a measure of irritation is required and does the answer need to be definitive, subjective, or conservative?).

For example, if testing is to be done on a new entity about which little is known, one would need to perform more assays than if testing a material that is essentially a reformulation of previously tested components. Additionally, if the reason for developing in vitro assays is to completely replace animals in irritation testing, then an extensive battery of methods examining a range of endpoints and potential mechanisms of action becomes essential since no confirming animal data will be obtained. If, on the other hand, the alternative methods are used as a prescreen to reduce the number of animals used, then fewer tests would be required, since one is only attempting to approximate the irritation potential. A decision would then be made whether to simply label a compound as irritating or to proceed with animal testing for a more accurate assessment.

In addition, the type and number of assays developed may depend on why the testing is being done. If, on the one hand, the testing is to be done for

worker safety within the industrial or pharmaceutical environment, where eye protection is mandatory for all employees, it may only be necessary to establish an irritant/nonirritant label rather than a qualitative ranking (nonirritating, slight, mild, moderate, severe, extreme). In these circumstances, only a single in vitro test or reliance on a database of information on similar compounds may be required to achieve this level of information. However, because the information obtained would provide a simple "yes/no," irritant versus nonirritant label, the methods used must be highly conservative; that is to say, false positives, while undesirable, would be more acceptable than false negatives, which could have drastic consequences upon accidental exposure. If, on the other hand, the testing is to be done on an ocular product where exposure to the eye would be deliberate or on a consumer product where a large, unprotected population may be potentially exposed, a more extensive and definitive irritancy evaluation would be necessary. Thus, it seems likely that no one approach to alternative testing can be consistently applied but that the objectives of each laboratory and situation need to be considered in deciding on an appropriate test battery.

Whatever the specific need or application, the use of a tier testing scheme can significantly reduce and in some cases eliminate the use of animals. An example of a decision-making tree and its application in ocular testing is provided in Figure 4. First, all available data about a test material (or related compounds), including chemical characteristics, historical data, other known toxicity, and the like, are collected. Analysis of these data could provide a strong indication of irritation potential, in which case the material would be labeled a presumed positive and handled as such. Alternatively, the data may be equivocal or insufficient, in which case a battery of in vitro assays could be performed.

The in vitro battery would ideally include measures of opacity, cytotoxicity, and inflammation. The actual test method(s) will vary depending upon the experience of the laboratory, types of compounds to be tested, and so on. If the measured endpoint(s) indicates that the test material is approximately equipotent with known irritants, one would presume the unknown to be an irritant and further testing would not generally be required. One should keep in mind, however, that in many cases in vitro assays are more sensitive than whole-animal testing, so a positive response in vitro may not always indicate an in vivo irritant. If the assays give equivocal results or responses similar to those seen with nonirritants or mild irritants, some type of animal testing may be indicated as confirmation.

If animal testing is required, a full-scale Draize test may not be necessary given the background established in the beginning of the tier approach. For instance, the compound could be tested in a single sentinel animal to obtain confirmation of in vitro data. In addition, other modifications could be used, such as the administration of appropriate anesthetics to the test animals or the use of the low-volume Draize modification (Falahee et al., 1982; Freeberg et al., 1984; Griffith, 1987).

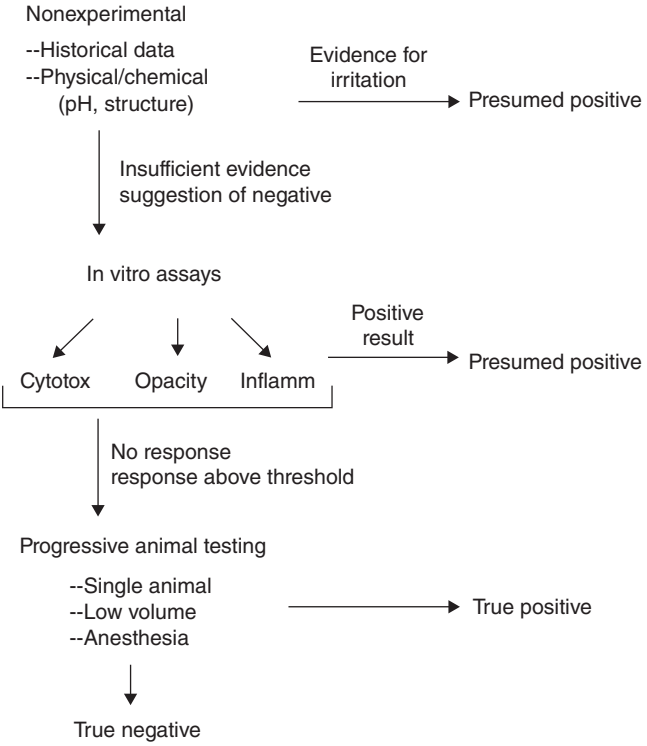


***Figure 4*** *Ocular decision tree.*

***Needs for the Future*** One significant need in the area of ocular alternatives is for the validation of current in vitro assays. The key point is that users of the assay(s) need to have confidence in their ability to interpret and reliably use the data generated. This confidence level can only be achieved by parallel testing in one's own laboratory with compounds similar to those likely to be evaluated as unknowns. The validation process will be a long, somewhat tedious project but will be necessary before in vitro alternatives can be used responsibly.

Another area where advances need to be made is in facing certain common technical issues in the in vitro methods themselves. A number of these potential pitfalls have been outlined by Frazier and Bradlow (1989). One of the chief issues is the testing of water-insoluble materials. Many assay methods are based on material maintained in aqueous media; therefore, dosing the target cells with insoluble or immiscible test agents is difficult or impossible. Agar overlays or partitioning of test compound from a membrane have been used in some cases, but these are not universally applicable. Application of test material in various solvents may or may not allow adequate doses to be achieved. However, the toxicity of the solvents alone and in combination with test compounds must be carefully examined. Until adequate solutions to these

solubility problems are found, some materials will simply not be able to be tested with in vitro methodologies.

Finally, as previously stated, if in vitro models are to be fully effective, the underlying mechanisms of ocular irritation need to be identified. Many of the in vitro assays proposed as alternatives to in vivo testing are based on correlations rather than mechanisms of irritation. The scoring or ranking of substances utilizing the in vitro endpoint may correlate with the severity of the in vivo response, but the reason for the agreement may be unclear and strictly fortuitous for the compounds evaluated. The ideal assay would monitor several biochemical or biological events specifically evoked in the whole animal by irritants. If events that are actually causal to irritation are measured, the probability of false results would decrease. Until we fully understand the mechanisms of the irritant response seen in the whole eye, a battery of tests aimed at evaluating as many biological events as possible is likely to be the best approach.

## 5.5 Nephrotoxicity In Vitro

The kidney is a frequent site of toxic insult due to drug and chemical exposure in experimental animals and humans. Broad classes of drugs and industrial chemicals are implicated in nephrotoxic reactions. The use of various in vitro models in studying renal toxicology is well documented in the literature (Johnson and Maack, 1977; Kacew and Hirsch, 1981; Hassall et al., 1983; Hook and Hewitt, 1986; Smith et al., 1987; Tay et al., 1988; Williams, 1989). A listing of the available in vitro models is provided in Table 9. Each model system possesses its own advantages and disadvantages, and all have demonstrated their usefulness and application in renal toxicology.

### *Methods*

*Isolated Perfused Kidney* As the name implies, the isolated perfused kidney consists of the intact organ maintained in a tissue bath apparatus. An excellent review of the methods involved in preparing this model has been published by Newton and Hook (1981). With respect to the functional integrity of the model, proximal tubule transport processes such as glucose uptake (Bowman and Maack, 1972), protein reabsorption (Maack, 1975), and amino acid uptake

**TABLE 9 In Vitro Nephrotoxicity Models**

| |
|---|
| Isolated perfused kidney |
| Isolated tubules |
| Renal cells |
| Cell lines |
| Primary cells |
| Kidney slices |
| Isolated organelles |

(De Mello and Maack, 1976) have been shown to be maintained. While this model has been used widely to examine renal physiological and drug disposition issues, it has not been extensively utilized as a toxicological tool (Bekersky, 1983). Cojocel et al. (1983) have used the isolated perfused rat kidney to examine the mechanisms of proteinuria induced by various aminoglycosides, while several other investigators have studied brush-border and basolateral aspects of renal accumulation and toxic reactions to aminoglycosides in the isolated perfused kidney (Collier et al., 1979; Williams et al., 1984). In another series of investigations, Newton et al. (1982a,b) examined the metabolism of acetaminophen in relationship to the nephrotoxicity of this agent. Using the isolated perfused kidney, they were able to show that the kidney was capable of metabolizing acetaminophen to the toxic chemical species *para*-aminophenol and of generating electrophilic intermediates capable of depleting glutathione.

*Kidney Slices* Renal tissue slice technology has been extensively exploited for renal pharmacology and toxicology assessments. This methodology involves the removal of longitudinal sections of kidney tissue of varying thickness and weight with a razor guided either free-hand or through the use of an apparatus known as a Stadie–Riggs (1944) microtome. Kidney slices from virtually all species can be prepared; however, rat and rabbit have been the most frequently utilized models. Renal slices maintain the architecture and cellular heterogeneity of the intact kidney, with tubular segments and surrounding interstitial and vascular elements present. Functionally, renal slices also exhibit organic ion transport, gluconeogenesis, and active maintenance of cellular sodium and potassium balance through the enzymatic activity of $Na^+K^+$–ATPase (Cross and Taggart, 1950; Berndt et al., 1984).

Typically, the functional and morphological integrity of renal slices has a relatively short life span of approximately 2 h. However, one laboratory has succeeded in prolonging the viability of tissue slices for as long as 24 h (Ruegg et al., 1987a).

It is contended that the renal slice technique measures primarily basolateral uptake of substrates or nephrotoxins based on histological evidence of collapsed tubular lumens. This results in the inaccessibility of brush-border surfaces for reabsorptive transport (Burg and Orloff, 1969; Cohen and Kamm, 1976). This observation limits the ability of this model to accurately reflect reactions to nephrotoxins that occur as the result of brush-border accumulation of an injurious agent. Ultrastructurally, a number of alterations, particularly in the plasma membrane and mitochondrial compartments, have been shown to occur over a 2-h incubation period (Ware et al., 1975; Martel-Pelletier et al., 1977). This deterioration in morphology is very likely a consequence of the insufficient diffusion of oxygen, metabolic substrates, and waste products in the innermost regions of the kidney slice (Cohen and Barac Neito, 1973; Cohen and Kamm, 1976). Such factors also limit the use of slices in studying renal metabolism and transport functions.

Renal cortical slices have been used to study numerous nephrotoxins including gentamicin (Hsu et al., 1977; Kluwe and Hook, 1978), cisplatin (Goldstein et al., 1981; Safirstein et al., 1981; Phelps et al., 1987), cephaloridine (Goldstein et al., 1986, 1987), chloroform (Smith and Hook, 1983; Smith et al., 1983), potassium dichromate (Kacew and Hirsch, 1981; Ruegg et al., 1987b), and the mycotoxin citrinin (Berndt et al., 1984; Baggett and Berndt, 1984), to name a few. While congruence between biochemical effects of nephrotoxins on in vitro and ex vivo renal slice function has been achieved, correlations of morphological features of nephrotoxin-induced damage in kidney slices with in vivo effects are largely lacking.

*Isolated Tubules* Isolated renal proximal tubule segments are most commonly prepared from either rats or rabbits by collagenase digestion of the kidneys followed by dispersion of the tissue into an appropriate buffer and filtration through sieves. Although this technique yields a tubule suspension that is highly contaminated with cellular debris, glomeruli, and distal tubules, the preparation can be further purified in a number of ways (Balaban et al., 1980; Vinay et al., 1981; Taub, 1984; Hatzinger and Stevens, 1989).

Renal tubular fragments appear to retain most of their in vivo functions and morphology. Balaban et al. (1980) reported that less than 3% of the tubular cells had any structural pathology except for the removal of the basement membrane. An important feature of this preparation was that the tubular lumina were open (Balaban et al., 1980). Functionally, isolated proximal tubular fragments transport *para*-aminohippurate, tetraethylammonium, α-methylglucoside, phosphate, and $^{86}$Rb in a fashion similar to that reported in vivo, demonstrating the functional integrity of membrane processes of the preparation (Vinay et al., 1981; Dantzler and Brokl, 1984; Tessitore et al., 1986; Dantzler et al., 1989). With respect to renal cortical glutathione (GSH), isolated tubules are approximately 50–70% GSH depleted when compared to control tissue. However, suspensions of isolated rabbit tubules will synthesize GSH from the constituent amino acids—glutamate, cysteine, and glycine.

Isolated proximal tubules have been utilized to study the mechanisms of nephrotoxicity induced by antibiotics (Sina et al., 1985, 1986), radiocontrast dyes (Humes et al., 1987), metals (Rylander et al., 1985), anoxia (Weinberg, 1985; Weinberg et al., 1987), cellular oxidants (Messana et al., 1988), cysteine conjugates (Rylander et al., 1985; Schnellman et al., 1987; Zhang and Stevens, 1989), and a variety of nephrotoxic bromobenzene metabolites (Schnellman and Mandel, 1986; Schnellman et al., 1987).

*Renal Cells* A variety of isolated cellular models exist for studying renal function and injury. These models can generally be divided into two categories: models derived from permanent renal cell lines and cellular models derived from freshly isolated renal tissue.

CELL LINES Cell lines, derived from tissue of various species, are commercially available from tissue culture banks. These cell populations are "immortalized" in that they possess the capacity to permanently proliferate in culture. Such cellular models can be studied in short-term suspension (hours) or longer term monolayer culture (days, weeks, months). Since cell lines have been extensively cultured or passaged for multiple generations, the degree or retention (or loss) of kidney-specific morphology and function is an important limitation that is not thoroughly addressed for a number of renal dell lines. One renal cell line that has been relatively well characterized is the pig kidney cell line LLC-$PK_1$.

The LLC-$PK_1$ (porcine kidney) cell line (Hull et al., 1976) exhibits a range of morphological and functional properties of proximal tubule epithelium. For example, $Na^+$-dependent glucose and amino acid transport has been demonstrated in LLC-$PK_1$ cells (Rabito and Ausiello, 1980; Rabito and Karish, 1982, 1983). LLC-$PK_1$ cells have also demonstrated the ability to transport organic cations such as tetraethylammonium (TEA) (Inui et al., 1985) but do not appear to possess the capacity for organic anion transport (Rabito, 1986). Morphologically, LLC-$PK_1$ cells have been shown to be polarized with apical microvilli, apical junctional complexes, desmosomes, and basolateral unfoldings (Cereijido et al., 1978), which is typical of transporting epithelial cells. The expression of brush-border morphology is striking in LLC-$PK_1$ cells, although the amount of brush border is less than that observed in the proximal tubule in vivo. High activities of proximal tubule brush-border membrane enzymes have also been observed in LLC-$PK_1$ cells (Gstraunthaler et al., 1985). LLC-$PK_1$ cells also possess significant levels of the basolateral enzyme marker $Na^+K^+$–adenosine triphosphatase (ATPase), which is thought to result in the formation of fluid-filled "blisters" or domes in monolayer culture via transepithelial sodium and water transport (Gstraunthaler et al., 1985).

Renal cell lines have been utilized to a limited extent for evaluation of nephrotoxins. A rabbit kidney cell line (LLC-$RK_1$) has been utilized for evaluating nephrotoxic antibiotics (Viano et al., 1983; Hottendorf et al., 1987; Williams et al., 1988). LLC-$PK_1$ cells have by far been the most widely employed cell line for studying drug-induced nephrotoxicity, specifically in the evaluation of aminoglycoside antibiotics (Hori et al., 1984; Schwertz et al., 1986; Williams et al., 1986b; Holohan et al., 1988). The morphological alterations induced by aminoglycosides in LLC-$PK_1$ cells correlated well with in vivo histological findings in the kidney, including the formation of secondary lysosomal inclusions referred to as myeloid bodies.

PRIMARY CELLS Cells from freshly isolated renal tissue can be obtained by explant cultures, cultures of isolated renal tubules discussed in the previous section, and digestion and isolation of individual renal cells. Cultures of isolated renal fragments such as the proximal tubule (Fine and Sakhrani, 1986) and collecting tubule (Grenier, 1986) have been the most extensively characterized models to date. Primary cultures of proximal tubules have been

isolated from rabbit (Chung et al., 1982), dog (Goligorsky et al., 1986), rat (Bellemann, 1980; Hatzinger and Stevens, 1989), and human kidneys (Detrisac et al., 1984; Wilson et al., 1985). Primary cultures prepared from rabbit proximal tubules have been shown to exhibit polarity and transport characteristics consistent with proximal tubule function. As with certain renal cell lines such as the LLC-$PK_1$ and MDCK, primary cultures also possess tight junctions and basolateral $Na^+K^+$–ATPase activity that results in the formation of fluid-filled "blisters" or domes in monolayer culture. Transport functions including phlorizin-sensitive $Na^+$-dependent glucose uptake (Sakhrani et al., 1984) and organic anion and cation transport (Yang et al., 1988) are present in rabbit primary cultures. Primary cultures of rabbit proximal tubules have also been shown to be responsive to parathyroid hormone (Chung et al., 1982), to maintain cellular glutathione levels (Aleo et al., 1987), and to exhibit brush-border enzyme activity (Ford et al., 1987).

Primary renal cell culture has been utilized to study a number of nephrotoxic agents, including mercuric chloride (Inamoto et al., 1976), cadmium (Cherian, 1982), lead (McLachlin et al., 1980), cisplatin (Tay et al., 1988), aminoglycoside antibiotics (Chatterjee et al., 1984; Sens et al., 1988), and cyclosporin (Trifillis et al., 1984). Studies reported by Tay et al. (1988) in rabbit proximal tubule cultures with cisplatin revealed biochemical effects upon DNA synthetic activty that correlated with in vivo histochemical effects of this antitumor agent in animals. With respect to studies involving mercuric chloride and aminoglycoside antibiotics in primary renal cultures, light and electron microscopy revealed similar patterns of cellular pathology in vitro as compared to in vivo exposure in animals (Chatterjee et al., 1984; Aleo et al., 1987).

Suspensions rather than cultures of renal cell lines and primary cells are also available for short-term evaluations. Procedures for the isolation and purification of specific renal cell types include density gradient centrifugation, flow cytometry, free flow electrophoresis, enzymatic digestion, and monoclonal antibodies (Dworzack and Grantham, 1975; Kreisberg et al., 1977; Heidrich and Dew, 1977; Eveloff et al., 1980; Endou et al., 1982; Smith and Garcia-Perez, 1985). While primary renal cell suspensions possess the advantage of being freshly isolated and derived from intact kidney tissue, a major disadvantage is their lack of cellular polarity that is exhibited with cells in culture. Relatively few reports of renal suspension cultures being used in nephrotoxic evaluations are in the literature at this time (Holohan et al., 1988).

One of the unique advantages of renal cell culture rests in the ability to study the directional aspects of drug exposure and cellular injury that operate in vivo. The technology to grow renal epithelial cells on filter inserts for this purpose became available in the late 1990s (Figure 5). This potential provides the opportunity to study compounds that interact or accumulate within the renal tubular epithelium in vivo via tubular reabsorption from the luminal surface or extraction from the basolateral or blood surface. Preliminary data involving the directional aspects of antibiotic toxicity to LLC-$PK_1$ cells have recently been obtained, indicating that LLC-$PK_1$ cells are significantly more

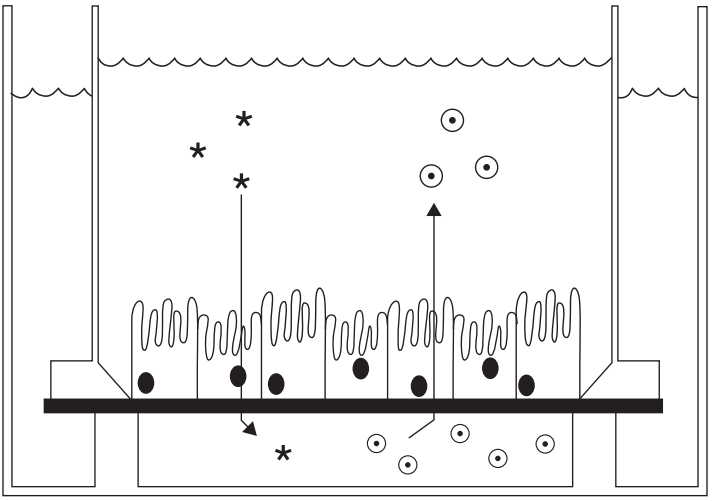

***Figure 5*** *Developing technology for bidirectional exposure/transport in polarized epithelial cells. Shown are kidney cells grown on porous membrane filters. Symbols represent chemical exposure from basolateral (⊙) and brush-border (*) surfaces.*

sensitive to the nephrotoxin cephaloridine when the antibiotic was applied to the basolateral surface of the kidney cell (Williams et al., 1993). These data are consistent with the evidence that cephaloridine is accumulated in renal cells predominantly through basolateral transport. This cellular accumulation is then proposed to precipitate cellular injury and death.

*Biochemical Systems* In addition to the isolated organ, tubule, and cellular models described above, biochemical systems derived from renal tissue represent valuable resources for in vitro evaluations of nephrotoxins. Particularly, the use of isolated cellular organelles such as mitochondria and plasma membranes has been applied to the study of nephrotoxins (Williams and Hottendorf, 1985, 1986; Tune et al., 1988, 1989). The examination of mechanisms of accumulation of nephrotoxins in the kidney has been accomplished utilizing purified renal brush-border and basolateral membrane vesicles for aminoglycosides (Williams and Hottendorf, 1986; Williams et al., 1987), cephalosporins (Kasher et al., 1983; Williams et al., 1985), and cisplatin (Williams and Hottendorf, 1985). Williams and co-workers (1987) reported that the affinity for renal membrane binding sites correlated well with the nephrotoxic potential of aminoglycosides. Thus, this membrane model could be utilized to assess the relative nephrotoxic potential of new aminoglycosides. Furthermore, isolated renal membranes were successfully employed in the identification of in vivo nephrotoxicity inhibitors for aminoglycosides (Williams et al., 1986a). The correlation between the ability of polyamino acids to inhibit in vitro binding and in vivo nephrotoxicity of aminoglycosides with

aminoglycosides coadministered with polyamino acids demonstrated the utility of in vitro models to contribute to the drug discovery process.

***Needs for the Future*** Challenges for the future involve the continued development of current in vitro models to provide physiological and morphological characteristics that correlate more closely with the kidney in vivo. These challenges will include, for example, defining conditions that facilitate and optimize the retention of renal transport and metabolizing capabilities in cell culture. Perhaps most challenging is the need to establish in vitro systems that reproduce the dynamic features of the nephron in vivo. The kidney, like other organs, exists in a dynamic rather than static environment. The fluid dynamics and vasculature of the renal nephron most certainly play a role in the sensitivity of the kidney in terms of both dose and time responses to nephrotoxins.

## 6 VALIDATION

Perhaps one of the biggest issues in developing alternative test strategies is the validation of the test methods. One reason that this topic has become controversial is that validation has different meanings to different people. In the strictest scientific and linguistic sense, validation is the process of proving adherence to principles, logic, and facts free from error or superficiality. As applied to in vitro test procedures designed to mimic in vivo responses, this definition would require adherence to the principles and facts of in vivo toxicity, according to the criteria previously outlined (Table 2). However, test reproducibility, simplicity, and transferability are frequently viewed as the critical ingredients to test validation at the expense of any mechanistic or scientific validity. It can be argued that such components, though important in test standardization and acceptance, may have little bearing on the true scientific validation and rigor of new test procedures. Despite divergent opinions on validation, much has been written about this topic, and a number of validation or evaluation projects are currently underway within various organizations such as the Soap and Detergent Association (Booman et al., 1988), the Fund for the Replacement of Animals in Medical Experiments (FRAME), the German Federal Health Office (Kalweit et al., 1987), and the Cosmetics, Toiletries, and Fragrances Association as well as within the laboratories of individuals working with specific assays.

What constitutes acceptable validation? As alluded to above, there is no general agreement on this point; the answer is likely to be different for each company or individual performing the tests. If one generalization is to be made concerning validation, it would be that one needs to test enough compounds with different characteristics and mechanisms of action to develop confidence in the in vitro data. And this in turn will depend upon the individual user. If one simply requires a toxicity classification for purposes of prescreening, the validation process may be less involved than if a more definitive classification

is required of a test system. Stated another way, the more one expects in terms of predictability from the in vitro data, the more testing that needs to be done to develop confidence in the methods. The cornerstone to this confidence lies in the fulfillment of the scientific criteria previously outlined (Table 2).

Additionally, the test materials used in the validation process should be as closely related as possible to the characteristics of the unknowns to be tested. It is clear from the literature, for instance, that many cytotoxicity assays give good correlations with the in vivo ocular irritancy data for surfactants but that the correlations fail when compounds from other chemical classes are tested. Since any particular assay may be used differently by individual safety assessment programs, each user must evaluate potential methods under conditions likely to be encountered in his or her own situation.

## 7 THE FUTURE

The future of in vitro techniques in toxicological assessment takes us back to our earlier discussion of the philosophical and scientific considerations operating in the evolution of alternative methods.

Scientifically, the future will depend on the level of confidence achieved that in vitro systems provide information that is representative of the in vivo processes the toxicologist seeks to model and predict. The degree to which this level of confidence will evolve is directly proportional to the fulfillment of the scientific criteria outlined in Table 2 which determine the strength of the in vitro–in vivo correlations obtained. Because gaps remain in our knowledge with respect to these important criteria, the future will also depend on advancements in available knowledge and technology with respect to biochemical or in vitro toxicological events. Specifically, challenges for the future involve the continued development of current in vitro models to provide psychological and morphological characteristics that correlate more closely with target organs in vivo. These challenges will include defining media constituents that will facilitate the retention of normal morphology and metabolizing capabilities in vitro as well as the establishment of systems that reproduce the dynamic features or organs in vivo. With in vitro–in vivo correlations playing a key role in in vitro test development, the growth of technologies that focus on differentiated functions, cellular relationships, human models, and other in vivo properties, such as fluid dynamics, will be critical to the future applications of in vitro techniques in toxicology. An example of a scientific and technological advancement that promises to contribute to the field of in vitro pharmacology and toxicology is the development of a human skin equivalent model. This commercially available material possesses the qualities of differentiation, cellular relationship, and incorporation of human tissue and has already shown usefulness in the study of dermal absorption and irritancy (Bell et al., 1981, 1983, 1989).

Philosophically, where will the future lead us? Assuming that the industrial toxicologist continues to play a central role in the evaluation and prediction of product safety, one can envision continuing involvement in performing extrapolations from animal data to potential human risk. At the same time, one may also expect increasing public pressure to reduce animal testing. With these toxicological and societal goals in mind, the value and application of human tissue in predicting human metabolism and target organ effects are likely to grow. This approach not only reduces animal use but also may significantly enhance our correlative abilities with the ultimate species of concern—humans. Perhaps challenges in risk assessment in the future will involve extrapolations from human tissues to human safety. However, as alluded to above, the ultimate value of using human tissues in vitro relies on the technological and intellectual advances in performing and interpreting in vitro biochemical data. Thus, the challenges in extrapolating in vitro data from human tissues to human responses in vivo are identical to current issues in the use of in vitro models to predict toxicity to animals. As predictability and correlation of in vitro models with specific in vivo events is achieved, toxicologists can look forward to performing meaningful species comparisons in vitro, involving direct preclinical assessments in humans (Figure 6).

In summary, an examination of the current state of the art in the pharmaceutical industry illustrates exciting and important roles that in vitro systems can play in toxicological assessments.

While in Europe there are numerous validated [by the European Centre for the Validation of Alternative Methods (ECVAM) standards] in vitro alternatives for personal care products and industrial and agricultural chemicals (though only two—the LAL assay for pyrogenicity and the 3T3 cell assay for phototoxicity—have acceptance for use with pharmaceutical regulators) with regulatory acceptance (as summarized in Table 10), this is not the case in the

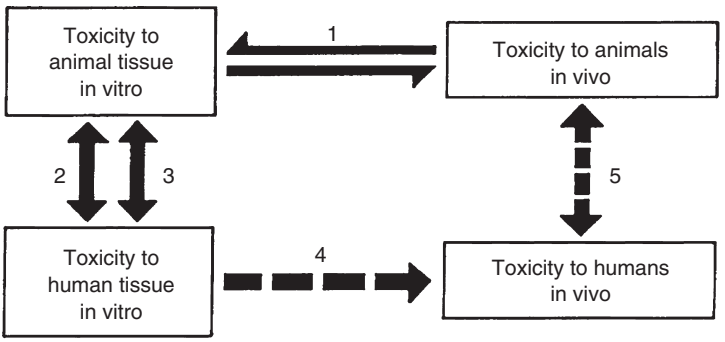


***Figure 6*** *Use of in vitro techniques in performing interspecies comparisons: 1, correlation of in vitro with in vivo observations; 2, development of parallel in vitro systems in humans; 3, comparison of in vivo sensitivity of different species; 4, potential toxicity in vivo; 5, comparison of in vivo sensitivity of different species.*

**TABLE 10 Status of Nonanimal Methods That Have Regulatory Standing**

| Test Method | Test System | Endpoint | OECD/TG (Test Group) or other Regulatory Comments |
|---|---|---|---|
| *In Vitro Test Methods for Which There Are OECD Health Effects Test Guidelines*[a] | | | |
| Transcutaneous electrical resistance test (TER) | Monitors changes in electrical resistance as measure of loss of corneum integrity and barrier function, involves skin disks from euthanized rats | Skin corrosion (topical agents) | TG 430 |
| Human skin models (EpiDerm, EPISKIN) | Reconstructed human epidermal equivalent (commercial system) used to assess cell viability, involving MTT reduction test | Skin corrosion (topical agents) | TG 431 |
| 3T3 NRU (neutral red uptake) phototoxicity test | BALB/c 3T3 (murine) cell line cytotoxicity based on neutral red uptake to measure cell viability; not direct replacement alternative, as there is no in vivo equivalent test | Phototoxicity | TG 432/FDA guidance |
| Corrositex membrane barrier test | Artificial barrier system coupled to pH-based chemical detection system | Skin corrosion (topical agents) | Draft TG 435 |
| Bacterial reverse mutation test (Ames) | Revertant bacteria detected by their ability to grow in absence of amino acid | Genotoxicity | TG471/ICH, ISO, FDA guidance |
| In vitro mammalian chromosome aberration test | Microscopic detection of chromosomal damage to cells in culture | Genotoxicity | TG473/ICH, ISO, FDA guidance |

**TABLE 10** ***Continued***

| Test Method | Test System | Endpoint | OECD/TG (Test Group) or other Regulatory Comments |
|---|---|---|---|
| In vitro mammalian cell gene mutation test | Functional bioassays to monitor mutations in enzyme encoding genes | Genotoxicity | TG476 |
| Sister chromatid exchange assay | Cells in culture are examined after two rounds of division by metaphase arrest and chromosomal preparation; chromatid exchange is monitored by microscopy | Genotoxicity | TG479/ICH, ISO, FDA guidance |
| Gene mutation assay in yeast | *Saccharomyces cerivisiae* exposed to test substance grown under different culture conditions used to monitor mutagenic potential (cf. Ames test) | Genotoxicity | TG 480 |
| Mitotic recombination assay in yeast | Crossover or gene conversion following exposure of yeast to test substance; relies on different growth requirements of mutated and wild-type yeast strains | Genotoxicity | TG 481 |
| Unsheduled DNA synthesis in mammalian cells | Measures DNA repair synthesis after deletions caused by test substance based on incorporation of radioactive nucleotides into newly synthesized DNA | Genotoxicity | TG 482/ICH, ISO, and FDA guidance |

**TABLE 10** ***Continued***

| Test Method | Test System | Endpoint | OECD/TG (Test Group) or other Regulatory Comments |
|---|---|---|---|
| In vitro micronucleus test | Cell-based assay; supplement to TG 474 (in vivo micronucleus test); detection of chromosome damage and formation of micronuclei in interphase | Genotoxicity | Draft TG 487/ICH guidance |
| Sex-linked recessive lethal test | *Drosophila* exposed to test substance; germline transmission of mutations is monitored through two successive generations | Reproductive toxicity | TG 477 |
| *Validated Methods Yet to be Introduced Into Regulatory Use* | | | |
| EpiOcular | Human keratinocyte–derived model of corneal epithelium barrier function | Eye irritation (topical application) | Retrospective (weight-of-evidence) validation (ECVAM) |
| In vitro micronucleus test | CHL/IU, CHO, SHE, or V79 cell lines commonly used, with or without metabolic activation, to monitor damage and formation of micronuclei in interphase | Mutagenicity | Retrospective (weight-of-evidence) validation (ECVAM) |
| Embryonic stem cell test | 3T3 cell cytotoxicity and differentiation of embryonic stem murine cell lines used to examine teratogenic potential | Developmental toxicity | Endorsed as screening test (EU) |
| Postimplantation rat whole-embryo test | Morphological assessment of rat embryos | Developmental toxicity | Endorsed as screening test (EU) |

**TABLE 10** ***Continued***

| Test Method | Test System | Endpoint | OECD/TG (Test Group) or other Regulatory Comments |
|---|---|---|---|
| Micromass test | Micromass cultures of rat limb are bud monitored for inhibition of cell proliferation and differentiation | Developmental toxicity | Endorsed as screening test (EU) |
| | *Methods Undergoing Valitation* | | |
| EPISKIN | Reconstructed human skin system used with MTT assay to monitor barrier function | Skin irritation | Report stage in EU |
| EpiDerm | Similar to EPISKIN | Skin irritation | Report stage in EU |
| | *Prevalidated Methods* | | |
| SkinEthic eye model | Epithelial corneal cell line used for cytotoxicity testing based on MTT reduction assay | Eye irritation | Appraisal stage in EU |
| | *Methods Undergoing Development, Prevalidation, or Evaluation* | | |
| Tissue culture models | Neutral red release and silicon microphysiometry or fluorescein leakage bioassays with human keratinocytes and MDCK cells, respectively; red blood cell (RBC) hemolysis test | Eye irritation | Being reviewed by ICCVAM for possible retrospective (weight-of-evidence) validation |
| Organotypic models | Bovine corneal opacity and permeability (BCOP) assay, with postmortem corneas; hen's egg test on the chorioallantoic membrane (HET-CAM assay); isolated rabbit and chicken eye tests (IRE and ICE) | Eye irritation | Being reviewed by ICCVAM for possible retrospective (weight-of-evidence) validation |

**TABLE 10** ***Continued***

| Test Method | Test System | Endpoint | OECD/TG (Test Group) or other Regulatory Comments |
|---|---|---|---|
| Modified Leydig cell line | Analysis of progesterone production as measure of test substance effects on steroid hormone production | One- or two-generation study | For use as part of test battery |
| Testis slices | Assessment of steroid production capacity of Leydig cells upon exposure of ex vivo rat tissue to toxicants | One- or two-generation study | For use as part of test battery |
| Human adrenocortical carcinoma cell line | Assay to allow entire steroid pathway effects to be mapped | One- or two-generation study | For use as part of test battery |
| Placental microsomal aromatase assay | Monitors ability of substances to affect steroid production; subcellular microsomal assay is used industrially | One- or two-generation study | For use as part of test battery (Stigson et al., 2007) |

*Note:* This is a comprehensive list of methods that have been validated or that are at various stages of development for toxicity testing. More information about these methods and how they can be applied is available from http://www2.defra.gov.uk/research/project_data/more?!=CB01067amp;M=KWSamp;V=reachamp;scope=0and (OECD, 2005).

[a]Including draft guidelines under review for acceptance, can be found at http://www.oecd.org/home/.

United States. Indeed, only four alternatives have been approved by the Interagency Coordinanating Committee for the Validation of Alternative Methods (ICVAM) to date (Gaul, 2008). Though there is considerable research continuing on the development (and putatively validation) of many more methods (Kuehn, 2008), progress in obtaining regulatory acceptance of such methods to take the place of animal-testing methods is very slow, and those methods that are accepted are limited in scope screens for eye and skin irritation pyrogenicity, phototoxicity, and genotoxicity.

This regulatory acceptance is, indeed, as it has been for 25 years, the final frontier and true obstacle to any further significant reduction in animal use.

## 8 SUMMARY

The tools are currently at hand (or soon will be) to provide the practicing toxicologist with unique opportunities for both identifying potentially toxic compounds in a much more rapid and efficient manner than before and teasing apart the mechanisms underlying such toxicities on an integrated basis (from the level of the molecule to that of the intact organism). The in vitro systems overviewed here, once understood (by investigators and regulators) in how they function and fail (just as in vivo systems have come to be understood), will allow this to happen while reducing the need to have recourse to intact mammalian test systems. However, the intact animal models—and, indeed, humans for pharmaceuticals—will still be an essential element in the safety assessment armamentarium for the foreseeable future.

## REFERENCES

Aleo, M. D., Wyatt, R. D., and Schnellmann, R. G. (1991). The role of the altered mitochondrial function in citrinin-induced toxicity to rat renal proximal tubule suspensions. *Toxicol. Appl. Pharmacol.* 109:455–463.

Baggett, J. M., and Berndt, W. O. (1984). Interaction of potassium dichromate with the nephrotoxins, mercuric chloride and citrinin. *Toxicology* 33(2):157–169.

Bagley, D. M., Dong, B. M., and De Salva, S. J. (1989). Assessing the eye irritation potential of surfactant-based materials using the chorioallantoic membrane vascular assay (CAMVA). In A. M. Goldberg (Ed.), *Alternative Methods in Toxicology*, Vol. 7. Mary Ann Liebert, New York, pp. 265–272.

Balaban, R. S., Soltoff, S. P., Storey, J. M., and Mandel, L. J. (1980). Improved renal cortical suspension: Spectrophotometric study of $O_2$ delivery. *Am. J. Physiol.* 238:F50–F59.

Bekersky, I. (1983). The isolated perfused kidney as a pharmacological tool. *TIBS* 4:6–7.

Bell, E., Ehrlich, H. P., Buttle, D. J., and Nakatusuji, T. (1981). Living tissue formed *in vitro* and accepted as skin-equivalent tissue of full thickness. *Science* 211:1052–1054.

Bell, E., Gay, R., Swiderek, M., Class, T., Kemp, P., Green, G., Haimes, H., and Bilbo, P. (1989). Use of fabricated living tissue and organ equivalents as defined higher order systems for the study of pharmacologic responses to test substances. Presented at NATO Advanced Research Workshop, Sept. 4–9, Bandol, France.

Bell, E., Sher, S., Hull, B., Merrill, C., Rosen, S., Chamson, A., Asselineau, D., Dubertret, L., Coulomb, B., Lapiere, C., Nusgens, B., and Neveus, Y. (1983). The reconstitution of living skin. *J. Invest. Dermatol.* 81:2s–10s.

Bellemann, P. (1980). Primary monolayer culture of liver parenchymal cells and kidney cortical tubules as a useful new model for biochemical pharmacology and experimental toxicology. Studies *in vitro* on hepatic membrane transport, induction of liver enzymes, and adaptive changes in renal cortical enzymes. *Arch. Toxicol.* 44:63–84.

Berghmans, S., Jette, C., Langenau, D., Hsu, K., Stewart, R., Look, T., and Kanki, J. P. (2005, Aug.). Making waves in cancer research: New models in the zebrafish. *Biotechniques* 39(2):227–237.

Berndt, W. O., Hayes, A. W., and Baggett, J. M. (1984). Effects of fungal toxins on renal slice calcium balance. *Toxicol. Appl. Pharmacol.* 74(1):78–85.

Blagg, J. (2006). Structure-activity relationships for in vitro and in vivo toxicity. *Annu. Reps. Med. Chem.* 41:353–368.

Boelsterli, U. A., Bouis, P., Brouillard, J. F., and Donatsch, P. (1988). *In vitro* toxicity assessment of cyclosporin A and its analogs in a primary rat hepatocyte culture model. *Toxicol. Appl. Pharmacol.* 96:212–221.

Boelsterli, U. A., Bouis, P., and Donatsch, P. (1987). Relative cytotoxicity of psychotropic drugs in cultured rat hepatocytes. *Cell Biol. Toxicol.* 3:231–250.

Booman, K. A., Cascieri, T. M., Demetrulias, J., Driedger, A., Griffith, J. F., Grochoski, G. T., Kong, B., McCormick, W. C., North-Root, H., Rozen, M. G., and Sedlak, R. I. (1988). *In vitro* methods for estimating eye irritancy of cleaning products phase I: Preliminary assessment. *J. Toxicol. Cut. Ocul. Toxicol.* 7:173–185.

Booman, K. A., DeProspo, J., Demetrulias, J., Driedger, A., Griffith, J. F., Grochoski, G., King, B., McCormick, W. C., North-Root, H., Rozen, M. G., and Sedlak, R. I. (1989). The SDA alternatives program: Comparison of *in vitro* data with Draize test data. *J. Toxicol. Cut. Ocul. Toxicol.* 8:35–49.

Borenfreund, E., and Borrero, O. (1984). *In vitro* cytotoxicity assays: Potential alternatives to the Draize ocular irritancy test. *Cell Biol. Toxicol.* 1:33–39.

Borenfreund, E., and Puerner, J. A. (1984). A simple quantitative procedure using monolayer cultures for cytotoxicity assays. *J. Tissue Culture Meth.* 9:7–9.

Borenfreund, E., and Puerner, J. A. (1987). Short-term quantitative *in vitro* cytotoxicity assay involving an S-9 activating system. *Cancer Lett.* 34:243–248.

Bowman, R. H., and Maack, T. (1972). Glucose transport by the isolated perfused rat kidney. *Am. J. Physiol.* 222:1499–1504.

Bracher, M., Faller, C., Spengler, J., and Reinhardt, C. A. (1987). Comparison of *in vitro* cell toxicity with *in vivo* eye irritation. *Mol. Toxicol.* 1:561–570.

Bradley, M. O., Dysart, G., Fitzsimmons, K., Harback, P., Lewin, J., and Wolf, G. (1982). Measurement by filter elution of DNA single- and double-strand breaks in rat hepatocytes: Effects of nitrosamines and γ-irradiation. *Cancer Res.* 42:2569–2597.

Brown, S., Templeton, L., Praten, D. A., and Potter, L. J. (1989). Use of an in vitro haemolysis test to predict tissue irritancy in an intramuscular formulation. *J. Parent. Sci. Technol.* 43:117–120.

Bruner, L. H., and Parker, R. D. (1990). Evaluation of six *in vitro* alternatives for ocular irritancy testing. *Toxicologist* 10:258.

Burg, M. B., and Orloff, J. (1969). *Para*-aminohippurate uptake and exchange by separated renal tubules. *Am. J. Physiol.* 217:1064–1068.

Castell, J. V., Larrauri, A., and Gomez-Lechon, M. J. (1988). A study of the relative hepatotoxicity *in vitro* of the non-steroidal anti-inflammatory drugs ibuprofen, flurbiprofen and butibufen. *Xenobiotica* 18:737–745.

Cereijido, M., Robbins, E. S., Doland, W. J., Rotunno, C. A., Sabatini, D. D. (1978). Polarized monolayers formed by epithelial cells on a permeable and translucent support. *J. Cell Biol.* 77:853–880.

Chatterjee, S., Trifillis, A., and Regec, A. (1984). Morphological and biochemical effects of gentamicin on cultured human-kidney renal tubular cells. *Hum. Toxicol.* 3:455.

Cherian, M. G. (1982). Studies on toxicity of metallothionein in rat kidney epithelial cell culture. *Dev. Toxicol. Environ. Sci.* 9:193–202.

Chiu, L. L., Cunningham, L. L., Raible, D. W., Rubel, E. W., and Ou, H. C. (2008). Using the zebrafish lateral line to screen for ototoxicity. *J. Assoc. Res. Otolaryngol.* 9:170–190.

Chung, S. D., Alavi, N., Livingston, D., Hiller, S., and Taub, M. (1982). Characterization of primary rabbit kidney cultures that express proximal tubule functions in a hormonally defined medium. *J. Cell Biol.* 95:118–126.

Clayton, G. D., and Clayton, F. E. (Eds.) (1981). *Patty's Industrial Hygiene and Toxicology*. Wiley, New York.

Clive, D., Johnson, K., Spector, J., Batson, A., and Brown, M. (1979). Validation and characterization of the L5178Y/TK mouse lymphoma mutagen assay system. *Mutat. Res.* 59:61–108.

Cohen, J. J., and Barac Neito, M. (1973). Renal metabolism of substrates in relation to renal function. In J. Orloff and R. W. Berliner (Eds.), *Handbook of Physiology*. American Physiological Society, Washington, DC, pp. 909–926.

Cohen, J. J., and Kamm, D. E. (1976). Renal metabolism: Relation to renal function. In B. M. Brenner and F. C. Rector (Eds.), *The Kidney*. Saunders, Philadelphia, pp. 126–214.

Cojocel, C., Dociu, N., Maita, K., Sleight, S. D., and Hook, J. B. (1983). Effects of aminoglycosides on glomerular permeability, tubular reabsorption, and intracellular catabolism of the cationic low molecular weight protein lysozyme. *Toxicol. Appl. Pharm.* 39:129–139.

Collier, V., Lietman, P., and Mitch, W. (1979). Evidence for luminal uptake of gentamicin in the perfused rat kidney. *J. Pharmacol. Exp. Ther.* 210:247–251.

Conquet, P., Durand, G., Lailler, J., and Plazonnet, B. (1977). Evaluation of ocular irritation in the rabbit: Objective versus subjective assessment. *Toxicol. Appl. Pharmacol.* 39:129–139.

Cooper, J. F. (1975). Principles and applications of the Limulus test for pyrogen in parenteral drugs. *Bull. Parent. Drug Assoc.* 3:122–130.

Croci, T., and Williams, G. M. (1985). Activities of several phase I and phase II xenobiotic biotransformation enzymes in cultured hepatocytes from male and female rats. *Biochem. Pharm.* 17(34):3029–3035.

Cross, R. J., and Taggart, J. L. (1950). Renal tubular transport: Accumlation of *p*-aminohippurate by rabbit kidney slices. *Am. J. Physiol.* 161:181–190.

Dantzler, W. H., and Brokl, O. H. (1984). Effects of low [$Ca^{2+}$] and $La^{3+}$ on PAH transport by isolated perfused renal tubules. *Am. J. Physiol.* 246:F175–F187.

Dantzler, W. H., Brokl, O. H., and Wright, S. H. (1989). Brush-border TEA transport in intact proximal tubules and isolated membrane vesicles. *Am. J. Physiol.* 256: F290–F297.

Davis, D. C., Potter, W. Z., Jollow, D. J., and Mitchell, J. R. (1974). Species differences in hepatic glutathione depletion, covalent binding and hepatic necrosis after acetaminophen. *Life Sci.* 14:2099–2109.

Decad, G. M., Hsieh, D. P. H., and Byard, J. L. (1977). Maintenance of cytochrome P-450 and metabolism of aflatoxin $B_1$ in primary hepatocyte cultures. *Biochem. Biophys. Res. Commun.* 78:279–287.

De Mello, G., and Maack, T. (1976). Nephron function of the isolated perfused rat kidney. *Am. J. Physiol.* 231:1699–1707.

Detrisac, C. J., Sens, M. A., Garvin, A. J., Spicer, S. S., and Sens, D. A. (1984). Tissue culture of human kidney epithelial cells of proximal tubule origin. *Kidney Int.* 25:383–390.

Doehmer, J. (2006). Predicting drug metabolism dependent toxicity for humans with a genetically engineered cell battery. *Altern. Lab. Animal* 34:561–573.

Donato, M. T., Gomez-Lechon, M. J., and Castell, J. V. (1990). Effect of xenobiotics on monooxygenase activities in cultured human hepatocytes. *Biochem. Pharmacol.* 8(39):1321–1326.

Dougherty, K. K., Spilman, S. D., Green, C. E., Steward, A. R., and Byard, J. L. (1980). Primary cultures of adult mouse and rat hepatocytes for studying the metabolism of foreign compounds. *Biochem. Pharmacol.* 29:2117–2124.

Drewes, C., Vining, E., and Callahan, C. (1984). Non-invasive electrophysiological monitoring: a sensitive method for detecting sublethal neurotoxicity in earthworms. *Environ. Toxicol. Chem.* 3:559–607.

Dworzack, D. L., and Grantham, J. J. (1975). Preparation of renal papillary collecting duct cells for study in vitro. *Kidney Int.* 8:191–194.

Ekwall, B., Bondesson, I., Castell, J. V., Gomez-Lechon, M. J., Hellberg, S., Hagberg, J., Jover, R., Ponsoda, X., Romert, L., Stenberg, K., and Watum, E. (1989). Cytotoxicity evaluation of the first ten MEIC chemicals: acute lethal toxicity in man predicted by cytotoxicity in five cellular assays and by oral LD50 tests in rodents. *J. Anim. Tech. Lab. Assoc.* 17:83–100.

Elgebaly, S. A., Gillies, C., Forouhar, F., Hahem, M., Baddour, M., O'Rourke, J., and Dreutzer, D. L. (1985). An *in vitro* model of leukocyte mediated injury to the corneal epithelium. *Curr. Eye Res.* 4:31–41.

Elgebaly, S. A., Forouhar, F., and Kreutzer, D. L. (1987). In vitro detection of cornea-derived leukocytic chemotactic factors as indicators of corneal inflammation. In A. M. Goldberg (Ed.), *Alternative Methods in Toxicology*, Vol. 5. Mary Ann Liebert, Inc., New York, pp. 257–268.

Enat, R., Jefferson, D. M., Ruiz-Opazo, N., Gatmaitan, Z., Leinwand, L. A., and Reid, L. M. (1984). Hepatocyte proliferation *in vitro*: Its dependence on the use of serum-free hormonally defined medium and substrata of extracellular matrix. *Proc. Natl. Acad. Sci U.S.A.* 81:1411–1415.

Endou, H., Koseki, C., Kimura, K., Yokokura, Y., Fukida, S., and Sakai, F. (1982). Use of a flow cytometer for the separation of isolated kidney cells. In F. Morel (Ed.), *Biochemistry of Kidney Functions*. Elsevier Biomedical, Amsterdam, pp. 69–78.

Enslein, K. (1995). Structure-activity predictions as alternatives to animal tests. In *Animal Test Alternatives*. Marcel Dekker, New York, pp. 33–35.

Enslein, K., Lander, T. R., Tomb, M. E., and Craig, P. N. (1983). *A Predictive Model for Estimating Rat Oral LD50 Values*. Princeton Scientific, Princeton, NJ.

European Centre for Exotoxicology and Toxicology of Chemicals (ECETOC) (1988). Eye irritation testing. Monograph No. 11. ECETOC, Brussles, pp. 1–65.

Eveloff, J., Haase, W., and Kinne, R. (1980). Separation of renal medullary cells: Isolation of cells from the thick ascending limb of Henle's loop. *J. Cell Biol.* 87:672–681.

Falahee, K. J., Rose, C. S., Seifried, H. F., and Sawhney, D. (1982). Alternatives in toxicity testing. In A. M. Goldberg (Ed.), *Alternative Methods in Toxicology*, Vol. 1: *Product Safety Evaluation*. Mary Ann Liebert, New York, pp. 137–162.

Farland, W. H., Tyson, C. A., and Sawhney, D. S. (1985). Rationale and use of functions tests in toxicity testing: A review. In C. A. Tyson and D. S. Sawhney (Eds.), *Organ Function Tests in Toxicity Evaluation*. Noyes Publications, Park Ridge, NJ, pp. 1–22.

Fielder, R. J., Gaunt, I. F., Rhodes, C., Sullivan, F. M., and Swanston, D. W. (1987). A hierarchical approach to the assessment of dermal and ocular irritancy: A report by the British Toxicological Society working party on irritancy. *Human Toxicol.* 6:269–278.

Fine, L. G., and Sakhrani, L. M. (1986). Proximal tubular cells in primary culture. *Mineral Electrolyte Metab.* 12:51–57.

Fishlock, D. (1990). Survival of the fittest drugs. *Financial Times*. Apr. 24, pp. 16–22.

Flower, C. (1987). Some problems in validating cytotoxicity as a correlate of ocular irritancy. In A. M. Goldberg (Ed.), *Alternative Methods in Toxicology*, Vol. 5. Mary Ann Liebert, New York, pp. 269–274.

Ford, S. M., Bennett, D., Laska, D. A., Tay, L. K., and Williams, P. D. (1987, Apr.). Biochemical comparison of fresh kidney homogenates, rabbit primary proximal tubule cultures (RPT), and a rabbit kidney epithelial cell line (LLC-$RK_1$). *FASEB J.* 9:526–533.

Franchini, J. L., Propat, J. T., Coner, G. R., and Yost, M. J. (2007). Novel tissue engineered tubular heart tissue for in vitro pharmaceutical toxicity testing. *Micmose Microavail.* 13:267–271.

Frazier, J. M. (1992). *In Vitro Toxicity Testing*. Marcel Dekker, New York.

Frazier, J. M., and Bradlow, J. A. (1989). Technical problems associated with in vitro toxicity testing systems. Technical Report No. 1. The Johns Hopkins Center for Alternatives to Animal Testing, Baltimore, MD.

Frazier, J. M., Gad, S. C., Goldberg, A. M., and McCulley, J. P. (1987). A critical evaluation of alternatives to acute ocular irritation testing. In A. M. Goldberg (Ed.), *Alternative Methods in Toxicology*, Vol. 4. Mary Ann Liebert, New York.

Freeberg, F. E., Griffith, J. F., Bruce, R. D., and Bay, P. H. S. (1984). Correlation of animal test methods with human experience for household products. *J. Toxicol. Cut. Ocul. Toxicol.* 1:53–64.

Gad, S. C. (1988a). Principles of screening in toxicology: With special emphasis on applications to neurotoxicology. *J. Am. Coll. Toxicol* 8:21–27.

Gad, S. C. (1988b). A tier testing strategy incorporating in vitro testing methods for pharmaceutical safety assessment. *Hum. Innovat. Alt. Anim. Exp.* 3:75–79.

Gad, S. C. (1990). Recent developments in replacing, reducing, and refining animal use in toxicologic research and testing. *Fund. Appl. Toxicol.* 15:8–16.

Gad, S. C. (1992). Industrial application for *in vitro* toxicity testing methods: A tier testing strategy for product safety assessment. In J. Frazier (Ed.), *In Vitro Toxicity Testing*. Marcel Dekker, New York, pp. 253–279.

Gad, S. C. (1993). In vivo and in vitro Tox. In J. A. Thomas and L. A. Myers (Eds.), *Biotechnology and Safety Assessment*. Raven, New York, pp. 97–129.

Gad, S. C. (1995a). Integration of alternative methods into the safety evaluation process in the United States. *Comments Toxicol.* 5:301–314.

Gad, S. C. (1995b). Challenges and needs in the safety assessment of new therapeutics. *Toxicol. Meth.* 5:viii–x.

Gad, S. C. (1996a). Strategies for the application and interpretation of *in vitro* methods to the development of new pharmaceuticals and medical devices. *Toxicol. Methods* 6:1–12.

Gad, S. C. (1996b). Histologic and clinical pathology in the safety assessment and development of new therapeutic agents. *Scand. J. Lab. Anim. Sci.* 12:77–86.

Gad, S. C. (1996c). Preclinical toxicity testing in the development of new therapeutic agents. *Scand. J. Lab. Anim. Sci.* 12:61–76.

Gad, S. C. (1998). Strategies and application of in vitro methods to the development of pharmaceuticals and devices. In H. Salem (Ed.), *Advances in Animal Alternatives for Safety and Efficacy Testing*. Taylor & Francis, Philadelphia, PA, pp. 293–302.

Gad, S. C. (1999). *Product Safety Evaluation Handbook*, 2nd ed. Marcel Dekker, New York.

Gad, S. C. (2000). *In Vitro Toxicology*, 2nd ed. Taylor & Francis, Philadelphia, PA.

Gad, S. C. (2001a). *Regulatory Toxicology*, 2nd ed. Taylor & Francis, Philadelphia, PA.

Gad, S. C. (2001b). Alternatives to in vivo studies in toxicology. In B. Ballantyne, T. Marvis, and P. Turner (Eds.), *General & Applied Toxicology*, 2nd eds. Macmillan, New York, pp. 401–424.

Gad, S. C. (2007). *Animal Models in Toxicology*, 2nd ed. Marcel Dekker, New York.

Gad, S. C., and Chengelis, C. P. (1999). *Acute Toxicology*, 2nd ed. Academic, San Diego, CA.

Gad, S. C., Walsh, R. D., and Dunn, B. J. (1986). Correlation of ocular and dermal irritancy of industrial chemicals. *J. Toxicol. Cut. Ocul. Toxicl.* 5:195–213.

Gaul, G. M. (2008). In US, few alternatives to testing in animals. *Washington Post*, April 12, p. A1.

Gautheron, P. D., Dukic, M., Alix, D., and Sina, J. F. (1992). Bovine corneal opacity and permeability test: An in vitro assay of ocular irritation. *Fund. Appl. Toxicol. Cut. Ocul. Toxicol.* 2:107–122.

Gautheron, P. D., and Sina, J. F. (1990). The bovine corneal opacity, an in vitro assay of ocular irritancy. *Toxicologist* 10:258.

Gilman, M. R., Jackson, E. M., Cervan, D. R., and Moreno, M. T. (1983). Relationship between the primary dermal irritation index and ocular irritation. *J. Toxicol. Cut. Ocular Toxicol.* 2:107–117.

Goldstein, R. S., Contardi, L. R., Pasino, D. A., and Hook, J. B. (1987). Mechanisms mediating cephaloridine inhibition of renal gluconeogenesis. *Toxicol. Appl. Pharmacol.* 87(2):297–305.

Goldstein, R. S., Noordewier, B., Bond, J. T., Hook, J. B., and Mayor, G. H. (1981). *Cis*-diaminedichloroplatinum nephrotoxicity: Time course and dose response of renal functional impairment. *Toxicol. Appl. Pharmacol.* 60:163–175.

Goldstein, R. S., Pasino, D. A., Hewitt, W. R., and Hook, J. B. (1986). Biochemical mechanisms of cephaloridine nephrotoxicity: Time and concentration dependence of peroxidative injury. *Toxicol. Appl. Pharmacol.* 83(2):261–270.

Goligorsky, M. S., Menton, D. M., and Hruska, K. A. (1986). Parathyroid hormone-induced changes of the brush border topography and cytoskeleton in cultured renal proximal tubular cells. *J. Membr. Biol.* 92:151–162.

Gomez-Lechon, M. J., Montoya, A., Lopez, P., Donato, T., Larrauri, A., and Castell, J. V. (1988). The potential use of cultured hepatocytes in predicting the hepatotoxicity of xenobiotics. *Xenobiotica* 18:725–735.

Gordon, V. C., and Kelly, C. P. (1989). An *in vitro* method for determining ocular irritation. *Cosmet. Toiletr.* 104:69–73.

Grant, W. M. (1986). *Toxicology of the Eye*. Charles C. Thomas: Springfield, IL.

Green, C. E., LeValley, S. E., and Tyson, C. A. (1986). Comparison of amphetamine metabolism using isolated hepatocytes from five species including human. *J. Pharm. Exp. Ther.* 237:931–936.

Grenier, F. C. (1986). Characteristics of renal collecting tubule cells in primary culture. *Mineral Electrolyte Metab.* 12:58–63.

Gribaldo, L. (2007). Overview of alternative methodologies in toxicology. *Current Protocols in Toxicology* 20:1–20.

Griffith, J. F. (1987, July). The low-volume eye irritation test: A case study in progress toward validation. *Chem. Times Trends*, pp. 19–40.

Gstraunthaler, G., Pfaller, W., and Kotanko, P. (1985). Biochemical characterization of renal epithelial cell cultures (LLC-$PK_1$ and MDCK). *Am. J. Physiol.* 248 (Renal Fluid Electrolyte Physiol.) 17:F536–544.

Guguen-Guillouzo, C., Gripon, P., Vandenbughe, Y., Lamballe, F., Rataanasavanh, D., and Guillouzo, A. (1988). Hepatotoxicity and molecular aspects of hepatocyte function in primary culture. *Xenobiotica* 18:773–783.

Guillot, J. P., Gonnet, J. F., Clement, C., Caillard, L., and Truhaut, R. (1982a). Evaluation of the cutaneous-irritation potential of 56 compounds. *Fd. Chem. Toxicol.* 20:563–572.

Guillot, J. P., Gonnet, J. F., Clement, C., Caillard, L., and Truhaut, R. (1982b). Evaluation of the ocular-irritation potential of 56 compounds. *Fd. Chem. Toxicol.* 20:573–582.

Guillouzo, A. (1986). Use of isolated and cultured hepatocytes for xenobiotic metabolism and cytotoxicity studies. In A. Guillouzo and C. Guguen-Guillouzo (Eds.), *Isolated and Cultured Hepatocytes.* John Libbey and Company, London, pp. 313–332.

Guillouzo, A., Begue, J. M., Ratanasavanh, D., and Chesne, C. (1988). Drug metabolism and cytotoxicity in long-term cultured hepatocytes. In A. Guillouzo (Ed.), *Liver Cells and Drugs*. Colloque Inserm, John Libbey Eurotext, Paris, pp. 235–244.

Guzelian, P. S., and Bissell, D. M. (1974). Metabolic factors in the regulation for cytochrome P-450. Studies in rat hepatocyte monolayer culture. *Fed. Proc.* 33: 1246.

Haseman, J. K., Crawford, D. D., Huff, J. E., Boorman, G. A., and McConnell, E. E. (1984). Results from 86 two-year carcinogenicity studies conducted by the National Toxicology Program. *J. Toxicol. Environ. Health* 14:621–639.

Hassall, C. D., Gandolfi, A. J., and Brendel, K. (1983). Correlation of the *in vivo* and *in vitro* renal toxicity of *S*-(1,2-dichlorovinyl)-L-cysteine. *Drug. Chem. Toxicol.* 6:507–520.

Hatzinger, P. B., and Stevens, J. L. (1989). Rat kidney proximal tubule cells in defined medium: The roles of cholera toxin, extracellular calcium and serum in cell growth and expression of γ-glutamyltransferase. *In Vitro Cell. Dev. Biol.* 25(2):205–212.

*HDI Toxicology Newsletter* (1987). HDI develops rabbit eye irritation model. No. 6 (K. Enslein, Ed.). Health Designs, Rochester, NY.

Heidrich, H. G., and Dew, M. E. (1977). Homogeneous cell populations from rabbit kidney cortex. *J. Cell Biol.* 74:780–788.

Hill, A. J., Teraka, H., Heideman, W., and Peterson, R. E. (2005) Zebrafish as a model vertebrate for investigating chemical toxicity. *Toxicol. Sci.* 86(1):6–19.

Hockin, L. J., and Paine, A. J. (1983). The role of 5-aminolevulinate synthetase, haem oxygenase and ligand formation in the mechanism of maintenance of cytochrome P-450 concentration in hepatocyte culture. *Biochem. Pharmacol.* 210:855–857.

Holme, J. (1985). Xenobiotic metabolism and toxicity in primary monolayer cultures of hepatocytes. *NIPH Ann.* 8:49–63.

Holohan, P. D., Sokol, P. P., Ross, C. R., Coulson, R., Trimble, M. E., Laska, D. A., Kracke, M. J., and Williams, P. D. (1988). Gentamicin-induced increases in cytosolic calcium in pig kidney cells (LLC-$PK_1$). *J. Pharm. Exp. Ther.* 247:349–354.

Hooisma, J. (1982). Tissue culture and neurotoxicology. *Neurobehav. Toxicol. Teratol.* 4:617–622.

Hook, J. B., and Hewitt, W. R. (1986). Toxic responses of the kidney. In C. D. Klaassen, M. O. Amdur, and J. Doull (Eds.), *Casarett & Doull's Toxicology. The Basic Science of Poisons*, 3rd eds. Macmillan, New York, pp. 310–329.

Hori, R., Yamamoto, Saito, H., Kohno, M., and Inui, K. (1984). Effect of aminoglycoside antibiotics on cellular functions of kidney epithelial cell line (LLC-$PK_1$): A model system for aminoglycoside nephrotoxicity. *Pharmacol. Exp. Ther.* 230:742–748.

Hottendorf, G. H., Laska, D. A., Ford, S. M., and Williams, P. D. (1987). The role of desacetylation in the detoxification of cephalothin in renal cells in culture. *J. Toxicol. Environ. Health* 22:101–111.

Hsu, C. H., Kurtz, T. W., and Weller, J. M. (1977). In vitro uptake of gentamicin by rat renal cortical tissue. *Antimicrob. Ag. Chemother.* 12:192–194.

Hull, R. N., Cherry, W. R., and Weaver, G. W. (1976). The origin and characteristics of a pig kidney cell strain LLC-$PK_1$. *In Vitro* 12:670–677.

Humes, H. D., Hunt, D. A., and White, M. D. (1987). Direct toxic influence of the radiocontrast agent diatrizoate on renal proximal tubule cells. *Am. J. Physiol.* 252: F246–F255.

Igarashi, H., Katsuta, Y., Matsumo, H., Nakazoto, Y., and Kawasaki, T. (1989). Opacification test by using the pig isolated cornea and its application to a test of corneal opacity induced by befunolol hydrochloride. *J. Toxicol. Sci. (Jpn.)* 14:91–103.

Inamoto, H., Ino, Y., Inamoto, N., Wada, T., Kihara, H., Watanabe, I., and Asano, S. (1976). Effect of $HgCl_2$ on rat kidney cells primary culture. *Lab. Invest.* 34: 489–493.

Inui, K., Saito, H., and Hori, R. (1985). $H^+$ gradient-dependent active transport of tetraethylammonium cation in apical-membrane vesicles isolated from kidney epithelial cell line LLC-$PK_1$. *Biochem. J.* 227:199–203.

Isom, H. C., Secott, T., Georgoff, I., Woodworth, C., and Mummaw, J. (1985). Maintenance of differentiated rat hepatocytes in primary culture. *Proc. Natl. Acad. Sci. U.S.A.* 82:3252–3256.

Jackson, E. M., and Goldner, R. (1989). *Irritant Contact Dermatitis*. Marcel Dekker, New York.

Jeong J. Y., Kwon, H. B., Ahn, J. C., Kang, D., Kwon, S. H., Park, J. A., and Kim, K. W. (2008). Functional and developmental analysis of the blood-brain barrier in zebrafish. *Brain Res. Bull.* 75(5): 619–628.

Jewell, S. A., Bellomo, G., Thor, H., Orrenius, S., and Smith, M. T. (1982). Bleb formation in hepatocytes during drug metabolism is caused by disturbances in thiol and calcium ion homeostasis. *Science* 217:1257–1258.

Johnson, V., and Maack, T. (1977). Renal extraction, filtration, absorption, and catabolism of growth hormone. *Am. J. Physiol.* 233(2):F185–196.

Kacew, S., and Hirsch, G. H. (1981). Evaluation of nephrotoxicity of various compounds by means of *in vitro* techniques and comparison to *in vivo* methods. In J. B. Hook (Ed.), *Toxicology of the Kidney*. Raven, New York, pp. 77–98.

Kalweit, S., Gerner, I., and Spielmann, H. (1987). Validation project of alternatives for the Draize eye test. *Mol. Toxicol.* 1:597–603.

Kasher, J. S., Holohan, P. D., and Ross, C. R. (1983). Effect of cephaloridine on the transport of organic ions in dog kidney plasma membrane vesicles. *J. Pharmacol. Exp. Ther.* 225(3):606–610.

Kennah, H. E., Albulescu, D., Hignet, S., and Barrow, C. S. (1989). A critical evaluation of predicting ocular irritancy potential from an in vitro cytotoxicity assay. *Fund. Appl. Toxicol.* 12:281–290.

Klaasen, C. D., and Stacey, N. H. (1982). Use of isolated hepatocytes in toxicity assessment. In G. Plaa and W. R. Hewitt (Eds.), *Toxicology of the Liver*. Raven, New York, pp. 147–179.

Kleinman, H. K., McGarvey, M. L., Hassell, J. R., Star, V. L., Cannon, F. B., Laurie, G. W., and Martin, G. R. (1986). Basement membrane complexes with biological activity. *Biochemistry* 25:312–318.

Kluwe, W. M., and Hook, J. B. (1978). Functional nephrotoxicity of gentamicin in the rat. *Toxicol. Appl. Pharmacol.* 45:163–175.

Kong, B. M., Viau, C. J., Rizvi, P. Y., and De Salva, S. J. (1987). The development and evaluation of the chorioallantoic membrane (CAM) assay. In A. M. Goldberg (Ed.), *Alternative Methods in Toxicology*, Vol. 5. Mary Ann Liebert, New York, pp. 59–73.

Kreisberg, J., Pitts, A. M., and Pretlow, T. G. (1977). Separation of proximal tubule cells from suspensions of rat kidney cells in density gradients of Ficoll in tissue culture medium. *Am. J. Pathol.* 86:591–601.

Kuehn, B. M. (2008). Health, environmental agencies team up to develop new plan for toxicity testing. *JAMA* 299:1415–1416.

Kurack, G., Vossen, P., Deboyser, D., Goethals, F., and Roberfubid, M. (1986). An in vitro model for acute toxicity screening using hepatocytes freshly isolated from adult mammals. In Goldberg, A. M. (Ed.), *In Vitro Toxicology*, pp. 235–242. Mary Ann Liebert, New York.

Lander, T., Enslein, K., Craig, P., and Tomb, N. (1984). Validation of a structure-activity model of rat oral LD50. In A. M. Goldberg (Ed.), *Acute Toxicity Testing: Alternative Approaches*. Mary Ann Liebert, New York. pp. 183–184.

Lawrence, R. S. (1987). The chorioallantoic membrane in irritancy testing. In C. K. Atterwill and C. E. Steele (Eds.), *In Vitro Methods in Toxicology*. Cambridge University Press, New York, pp. 263–278.

Lawrence, R. S., Groom, M. H., Ackroyd, D. M., and Parish, W. E. (1986). The chorioallantoic membrane in irritation testing. *Fd. Chem. Toxicol.* 24:497–502.

Lawrence-Beckett, E. M., and James, J. T. (1990). Initial experience with the Eyetex in vitro eye irritation test system. *Toxicologist* 20:259.

Ledbetter, S. R., Kleinman, H. K., Hassell, J. R., and Martin, G. R. (1984). Isolation of laminin. In D. W. Barnes, D. A. Sirbasku and G. H. Sato (Eds.), *Methods for Preparation of Media, Supplements and Substrata for Serum-Free Animal Cell Culture*. A. R. Liss, New York, pp. 231–238.

Leighton, J., Nassauer, J., Tchao, R., and Verdone, J. (1983). Development of a procedure using the chick egg as an alternative to the Draize rabbit test. In A. M. Goldberg (Ed.), *Alternative Methods in Toxicology*, Vol. 1. Mary Ann Liebert, New York, pp. 165–177.

Li, A. P., Merrill, J. C., and Beck, D. J. (1990). Future optimization of the cryopreservation procedures for rat and human hepatocytes. *Toxicologist* 10:63.

Lijinsky, W. (1988). Importance of animal experiments in carcinogenesis research. *Environ. Mol. Mutagen.* 11:307–314.

Lindahl-Kiessling, K., Karlberg, I., and Olofsson, A. M. (1989). Induction of sister-chromatid exchanges by direct and indirect mutagens in human lymphocytes, co-cultured with intact rat liver cells: Effect of enzyme induction and preservation of the liver cells by freezing in liquid nitrogen. *Mut. Res.* 211:77–87.

Liu, J., Kershaw, W. C., and Klaasen, C. D. (1990). Rat primary hepatocyte cultures are a good model for examining metallothionein-induced tolerance to cadmium toxicity. *In Vitro Cell. Dev. Biol.* 26:75–79.

Luepke, N. P. (1985). Hen's egg chorioallantoic membrane test for irritation potential. *Fd. Chem. Toxicol.* 23:287–291.

Maack, T. (1975). Renal handling of low molecular weight proteins. *Am. J. Med.* 58:57–64.

Martel-Pelletier, J., Guerette, D., and Bergeron, M. (1977). Morphologic changes during incubation of renal slices. *Lab Invest.* 36(5):509–518.

Maslansky, C. J., and Williams, G. M. (1985). Methods for the initiation and use of hepatocyte primary cultures from various rodent species to detect metabolic activation of carcinogens. In M. Webber and L. Sekely (Eds.), *In Vitro Models for Cancer Research*. CRC Press, Boca Raton, FL, pp. 43–60.

McLachlin, J. R., Goyer, R. A., and Cherian, M. G. (1980). Formation of lead-induced inclusion bodies in primary rat kidney epithelial cell cultures: Effect of actinomycin D and cycloheximide. *Toxicol. Appl. Pharmacol.* 56:418–431.

McQueen, C. A., Merrill, B. M., and Williams, G. M. (1984). Comparison of several indicators of cytotoxicity in rat hepatocytes in primary culture. *Toxicologist* 4:134.

McQueen, C. A., and Williams, G. M. (1981). Characterization of DNA repair elicited by carcinogens and drugs in the hepatocyte primary culture/DNA repair test. *J. Toxicol. Environ. Health* 8:463–477.

McQueen, C. A., and Williams, G. M. (1983). The use of cells from rat, mouse, hamster, and rabbit in the hepatocyte primary culture/DNA repair test. *Ann. N.Y. Acad. Sci.* 407:119–130.

McQueen, C. A., and Williams, G. M. (1985). Methods and modifications of the hepatocyte primary culture/DNA repair test. In H. A. Milman and E. K. Weisburger (Eds.), *Handbook of Carcinogenic Testing*. Noyes Publications, Park Ridge, NJ, pp. 116–129.

McQueen, C. A., and Williams, G. M. (1987). Toxicology studies in cultured hepatocytes from various species. In E. J. Rauckman and G. M. Padilla (Eds.), *The Isolated Hepatocyte: Use in Toxicology and Xenobiotic Biotransformation*. Academic, Orlando, FL, pp. 51–67.

Messana, J. M., Cieslinski, D. A., O'Connor, R. P., and Humes, H. D. (1988). Glutathione protects against exogenous oxidant injury to rabbit renal proximal tubules. *Am. J. Physiol.* 255:F874–F884.

Meyer, D. S. (1989). Safety evaluation of new drugs. In Y. C. Martin, E. Kutter, and V. Austel (Eds.), *Modern Drug Research*. Marcel Dekker, New York, pp. 355–399.

Muir, C. K. (1984). A simple method to assess surfactant-induced bovine corneal opacity *in vitro*: Preliminary findings. *Toxicol. Lett.* 22:199–203.

Muir, C. K. (1985). Opacity of bovine cornea *in vitro* induced by surfactants and industrial chemicals compared with ocular irritancy *in vivo*. *Toxicol. Lett.* 24:157–162.

Neubert, D. (1982). The use of culture techniques in studies on prenatal toxicity. *Pharmacol. Ther.* 18:397–434.

Neuhauser, E., Durkin, P., Malecki, M., and Antara, M. (1985a). Comparative toxicity of ten organic chemicals to four earthworm species. *Comp. Biochem. Physiol.* 83C:197–200.

Neuhauser, E., Loehr, C., Malecki, M., Milligan, D., and Durkin, P. (1985b). The toxicity of selected organic chemicals to the earthworm *Eisenia fetida*. *J. Environ. Qual.* 14:383–388.

Neuhauser, E., Loehr, C., and Malecki, M. (1986). Contact and artificial soil tests using earthworms to evaluate the impact of wastes in soil. In J. Petros, W. Lacy, and R. C. Conway (Eds.), *Hazardous and Industrial Solid Waste Testing: Fourth Symposium*. ASTM STP 886. American Society for Testing and Materials, Philadelphia, PA, pp. 192–202.

Newton, J. F., Braselton, W. E., Kuo, W., Kluwe, W. M., C. H., Gemborys, M. W., Mudge, G. H., and Hook, J. B. (1982a). Metabolism of acetaminophen by the isolated perfused kidney. *J. Pharmacol. Exp. Ther.* 221:76–79.

Newton, J. F., and Hook, J. B. (1981). Isolated perfused rat kidney. In W. B. Jacoby (Ed.), *Methods in Enzymology*, Vol. 77. Academic, New York.

Newton, J. F., Kuo, C. H., Gemborys, M. W., Mudge, G. H., and Hook, J. B. (1982b). Nephrotoxicity of *p*-aminophenol, a metabolite of acetaminophen, in the Fisher 344 rat. *Toxicol. Appl. Pharmacol.* 65:336–344.

Organisation for Economic Co-operation and Development (OECD) (2005). *Environmental, Health and Safety News*, No. 17, Apr. 2005.

Oglesby, L. A., Ebron, M. T., Beyer, P. E., Carver, B. D., and Kavlock, R. J. (1986). Co-culture of rat embryos and hepatocytes: *In vitro* detection of a proteratogen. *Teratogen. Carcinogen. Mutagen.* 6:129–138.

Paine, A. J., and Hockin, L. J. (1980). Nutrient imbalance causes the loss of cytochrome P-450 in liver cell culture: Formulation of culture media which maintain cytochrome P-450 at *in vivo* concentrations. *Biochem. Pharmacol.* 29:3215–3218.

Parce, J. W., Owicki, J. C., Kercso, D. M., Sigal, G. B., Wada, H. G., Muir, V. C., Bousse, L. J., Ross, K. L., Sikic, B. I., and McConnell, H. M. (1989). Detection of cell-affecting agents with a silicon biosensor. *Science* 246:243–247.

Phelps, J. S., Gandolfi, A. J., Berndel, K., and Dorr, R. T. (1987). Cisplatin nephrotoxicity: *In vitro* studies with precision-cut rabbit renal cortical slices. *Toxicol. Appl. Pharmacol.* 90:501–512.

Pittner, R. A., Fears, R., and Brindley, D. N. (1985). Effects of glucocorticoids and insulin on activities of phosphatidate phosphohydrolase, tyrosine aminotransferase and glycerol kinase in isolated rat hepatocytes in relation to the control of triacyglycerol synthesis and gluconeogenesis. *Biochem. J.* 225:455–462.

Plaa, G. L. (1976). Quantitative aspects in the assessment of liver injury. *Environ. Health Perspect.* 15:39–46.

Plaa, G. L., and Hewitt, W. R. (1982). Quantitative evaluation of indices of hepatotoxicity. In G. L. Plaa and W. R. Hewitt (Eds.), *Toxicology of the Liver*. Raven, New York, pp. 103–120.

PMA/Drug Safety Subsection (Drusafe) In Vitro Toxicology Task Force (1994). A collaborative evaluation of an in vitro muscle irritation assay. *Toxicol. Methods.* 4:215–223.

Poirier, M. C., Williams, G. M., and Yuspa, S. H. (1980). Effect of culture conditions cell type and species of origin on the distribution of acetylated and deacetylated deoxyguanosine C-8 adducts of *N*-acetoxy-2-acetylaminofluorene. *Mol. Pharmacol.* 18:581–587.

Price, J. B., Barry, M. P., and anderws, I. J. (1986). The use of the chick chorioallantoic membrane to predict eye irritants. *Fd. Chem. Toxicol.* 24:503–505.

Rabito, C. A. (1986). Occluding junctions in a renal cell line (LLC-$PK_1$). The basolateral systems. *J. Biol. Chem.* 257:6802–6808.

Rabito, C. A., and Ausiello, D. A. (1980). $Na^+$-dependent sugar transport in a cultured epithelial cell line from a pig kidney. *J. Membr. Biol.* 54:31–38.

Rabito, C. A., and Karish, M. V. (1982). Polarized amino acid transport by an epithelial cell line of renal origin (LLC-$PK_1$). The basolateral systems. *J. Biol. Chem.* 257: 6802–6808.

Rabito, C. A., and Karish, M. V. (1983). Polarized amino acid transport by an epithelial cell line of renal origin (LLC-$PK_1$). The apical systems. *J. Biol. Chem.* 258:2543–2547.

Ratanasavahn, D., Baffet, G., Latinier, M. F., Fissel, M., and Guillouzo, A. (1988). Use of hepatocyte co-cultures in the assessment of drug toxicity from chronic exposure. *Xenobiotica* 18:765–771.

Reese, J. A., and Byard, J. L. (1981). Isolation and culture of adult hepatocytes from liver biopsies. *In Vitro* 17:935–940.

Roberts, R., and Dorough, H. (1984). Relative toxicities of chemicals to the earthworm *Eisenia foetida*. *Environ. Toxicol.* 3:67–78.

Roberts, R., and Dorough, H. (1985). Hazards of chemicals to earthworms. *Environ. Toxicol. Chem.* 4:307–323.

Rofe, P. C. (1971). Tissue and culture toxicology. *Food Cosmet. Toxicol.* 9:685–696.

Rowan, A. N., and Stratmann, C. J. (1980). *The Use of Alternatives in Drug Research*. University Park, Baltimore, MD.

Ruegg, C. E., Gandolfi, A. J., Brendel, K., Nagle, R. B., and Krumdieck, C. L. (1987a). Preparation of positional renal slices for study of cell-specific toxicity. *J. Pharmacol. Methods* 12:111–123.

Ruegg, C. E., Gandolfi, A. J., Nagle, R. B., and Brendel, K. (1987b). Differential patterns of injury to the proximal tubule of renal cortical slices following in vitro exposure to mercuric chloride, potassium dichromate or hypoxic conditions. *Toxicol. Appl. Pharmacol.* 90:261–273.

Russel, W. M. S., and Burch, R. L. (1959). *The Principles of Humane Experimental Technique*. Methuen, London.

Rylander, L. A., GAndolfi, A. J., and Brendel, K. (1985). Inhibition of organic acid/base transport in isolated rabbit renal tubules by nephrotoxins. In A. M. Goldberg (Ed.), *In Vitro Toxicology*. A progress report from The Johns Hopkins Center for Alternatives to Animal Testing, Vol. 3. Mary Ann Liebert, New York, pp. 235–247.

Safirstein, R., Miller, P., Dikman, S., Lyman, S., and Shapiro, I. (1981). Cisplatin nephrotoxicity in rats. *Kidney Int.* 25:753–758.

Sakhrani, L. M., Badie-Dezfooly, B., Trizna, W., Mikhails, N., Lowe, A., Taub, M., and Fine, L. G. (1984). Transport and metabolism of glucose by renal proximal tubular cells in primary culture. *Am. J. Physiol.* 246:F757–F764.

Salem, H. (1995). *Animal Test Alternatives*. Marcel Dekker, New York.

Salem, H., and Baskin, S. I. (1993). *New Technologies and Concepts of Reducing Drug Toxicities*. CRC Press, Boca Raton, FL.

Salem, H., and Katz, S. A. (1998). *Advances in Animal Alternatives for Safety and Efficacy Testing*. Taylor & Francis, Washington, DC.

Scheuplein, R. J., Schoal, S. E., and Brown, R. N. (1990). Role of pharmacokinetics in safety evaluation and regulatory considerations. *Ann. Rev. Pharmacol. Toxicol.* 30:197–218.

Schnellman, R. G., Lock, E. A., and Mandel, L. J. (1987). A mechanism of *S*-(1,2,3,4,4-pentachloro-1,3-butadienyl)-L-cysteine toxicity to rabbit proximal tubules. *Toxicol. Appl. Pharmacol.* 90:513–521.

Schnellman, R. G., and Mandel, L. J. (1986). Cellular toxicity of bromobenzene and bromobenzene metabolites to rabbit proximal tubules: The role and mechanism of 2-bromohydroquinone. *J. Pharmacol. Exp. Ther.* 237:456–461.

Schwertz, D. W., Dreisberg, J. I., and Venkatachalam, M. A. (1986). Gentamicin-induced alterations in pig kidney epithelial (LLC-PK1) cells in culture. *J. Pharmacol. Exp. Ther.* 236(1):254–262.

Seglen, P. O. (1976). Preparation of isolated rat liver cells. *Methods Cell Biol.* 13:29–83.

Seglen, P. O., and Fossa, J. (1978). Attachment of rat hepatocytes in vitro to substrata of serum protein, collagen or concanavalin A. *Exp. Cell Res.* 116:199–206.

Seglen, P. O., Solheim, A. E., Grinde, P. B., Schwarze, P. E., Gjessing, R., and Poli, A. (1980). Amino acid control of protein synthesis and degradation in isolated rat hepatocytes. *Ann. N.Y. Acad. Sci.* 349:1–22.

Sens, M. A., Hennigar, G. R., Hazen-Martin, D. J., and Sens, D. A. (1988). Cultured human proximal tubule cells as a model for aminoglycoside nephrotoxicity. *Ann. Clin. Lab. Sci.* 18:204–214.

Shaw, J. L., Blanco, J., and Mueller, G. C. (1975). A simple procedure for isolation of DNA, RNA and protein factions from cultured animal cells. *Anal. Biochem.* 65:125–131.

Shopsis, C. (1989). Validation study: Ocular irritancy prediction with the total cell protein, uridine uptake, and neutral red assays applied to human epidermal keratinocytes and mouse 3t3 cells. In A. M. Goldberg (Ed.), *Alternative Methods in Toxicology*, Vol. 7. Mary Ann Liebert, New York, pp. 273–287.

Simmerman, H. J., Kendler, J., Libber, S., and Lukacs, L. (1974). Hepatocyte suspensions as a model for demonstration of drug hepatoxicity. *Biochem. Pharmacol.* 23:2187–2189.

Sina, J. F., Bean, C. L., Bland, J. A., MacDonald, J. J., Noble, C., Robertson, R. T., and Bradley, M. O. (1986). An *in vitro* assay for cytotoxicity to proximal tubule suspensions from rabbit kidney. *In Vitro Toxicol.* 1(1):12–22.

Sina, J. F., Bean, C. L., Noble, C., and Bradley, M. O. (1985). An *in vitro* nephrotoxicity assay utilizing proximal tubule suspensions from rabbit kidney. In *In Vitro Toxicology*, Vol 3. Mary Ann Liebert, New York, pp. 683–693.

Sina, J. F., and Gautheron, P. D. (1990). Assessment of *in vivo* ocular irritation using *in vitro* cytotoxicity assays. *Toxicologist* 10:259.

Sina, J. F., Ward, G. J., Laszek, M. A., and Gautheron, P. D. (1992). Assessment of cytotoxicity assays as predictors of ocular irritation of pharmaceuticals. *Fund. Appl. Toxicol.* 18:515–521.

Singer, P. (1975). *Animal Liberation: A New Ethic for Our Treatment of Animals*. Random House, New York.

Smith, C. G. (1992). *The Process of New Drug Discovery and Development*. CRC Press, Boca Raton, FL.

Smith, J. H., and Hook, J. B. (1983). Mechanism of chloroform nephrotoxicity. II. In vitro evidence for renal metabolism of chloroform in mice. *Toxicol. Appl. Pharmacol.* 70:480–485.

Smith, J. H., Maita, K., Sleight, S. D., and Hook, J. B. (1983). Mechanism of chloroform nephrotoxicity. I. Time course of chloroform toxicity in male and female mice. *Toxicol. Appl. Pharmacol.* 70:467–479.

Smith, M. A., Hewitt, W. R., and Hook, J. (1987). *In Vitro* methods in renal toxicity. In C. K. Atterwill and C. E. Steele (Eds.), *In Vitro Methods in Toxicology*. Cambridge University Press, Cambridge, pp. 13–35.

Smith, W. L., and Garcia-Perez, J. (1985). A. Immunodissection: Use of monoclonal antibodies to isolate specific types of renal cells. *Am. J. Physiol.* 248:F1–F7.

Smolarek, T. A., Higgings, C. V., and Amacher, D. E. (1990a). Metabolism and cytotoxicity of acetaminophen in hepatocyte cultures from rat, rabbit, dog and monkey. *Drug Metab. Dispos.* 18:659–663.

Smolarek, T. A., Higgings, C. V., and Amacher, D. E. (1990b). The biotransformation of tetrahydroaminoacidine in cultured hepatocytes as the cause for relative cytotoxicity in 3 species. *Toxicologist* 10:329.

Smyth, H. F., Carpenter, C. P., and Weil, C. S. (1949). Range-finding toxicity data, List III. *J. Ind. Hyg.* 31:60–62.

Smyth, H. F., Carpenter, C. P., and Weil, C. S. (1951). Range-finding toxicity data, List IV. *Arch. Ind. Hyg.* 4:119–122.

Soto, R. J., Servi, M. J., and Gordon, V. C. (1989). Evaluation of an alternative method for ocular irritation. In A. M. Goldberg (Ed.), *Alternative Methods in Toxicology*, Vol. 7. Mary Ann Liebert, New York, pp. 289–296.

Stadie, W. C., and Riggs, B. C. (1944). Microtome for the preparation of tissue slices for metabolic studies of tissues *in vitro*. *J. Biol. Chem.* 154:687–690.

Stenersen, J. (1979). Action of pesticides on earthworms. Part I: Toxicity of cholinesterase-inhibiting insecticides to earthworms as evaluated by laboratory tests. *Pestic. Sci.* 10:66–74.

Stern, H. M., and Zon, L. I. (2003). Cancer genetics and drug discovery in the zebrafish. *Nat. Rev. Cancer* 3(7):533–539.

Stigson, M., Kultima, K., Jergil, M., Scholz, B, Alm, H., Gastophon, A. L., and Dencher, L. (2007). Molecular targets and early response biomarkers for the prediction of developmental toxicity in vitro. *Alt. Test Anim.* 35:561–573.

Story, D. L., Gee, S. J., Tyson, C. A., and Gould, D. H. (1983). Response of isolated hepatocytes to organic and inorganic cytotoxins. *J. Toxicol. Environ. Health* 11: 483–501.

Strom, S. C., Jirtle, R. L., Jones, R. S., Novicki, D. L., Rosenberg, M. R., Novotny, A., Irons, G. P., McLain, J. R., and Michalopoulous, G. (1982). Isolation, culture, and transplantation of human hepatocytes. *J. Natl. Cancer Inst.* 68:771–778.

Suolinna, E. (1982). Isolation and culture of liver cells and their use in the biochemical research of xenobiotics. *Med. Biol.* 60:237–254.

Tanguay, R. L., and Reimers, M. J. (2008). Analysis of ethanol developmental toxicity in zebrafish. *Methods Mol. Biol.* 447:63–74.

Taub, M. (1984). Growth of primary and established kidney cell cultures in serum-free media. In D. W. Barnes, D. A. A. Sirbasku, and G. H. Sato (Eds.), *Methods for Serum-Free Culture of Epithelial and Fibroblastic Cells*. Alan R. Liss, New York, pp. 3–24.

Tay, L. K., Bregman, D. L., Maters, B. G., and Williams, P. D. (1988). Effects of *cis*-diaminedichloroplatinum (I) on rabbit kidney *in vivo* and on rabbit renal proximal tubule cells in culture. *Cancer Res.* 48:2538–2543.

Tessitore, N., Sakhrani, L. M., and Massary, S. G. (1986). Quantitive requirement for ATP for active transport in isolated renal cells. *Am. J. Physiol.* 251:C120–C127.

Thomson, M. A., Dickens, M. S., and Gordon, V. C. (1989a, Apr.). Evaluation of the Eytex biochemical assay for use in determining cosmetic product ocular irritancy. CAAT Symposium Poster, Baltimore, MD.

Thomson, M. A., Hearn, L. A., Smith, K. T., Teal, J. J., and Dickens, M. S. (1989b). Evaluation of the neutral red cytotoxicity assay as a predictive test for the ocular irritancy potential of cosmetic products. In A. M. Goldberg (Ed.), *Alternative Methods in Toxicology*, Vol. 7. Mary Ann Liebert, New York, pp. 297–305.

Tolman, K. G., Peterson, P., Gray, P., and Hammar, S. P. (1978). Hepatotoxicity of salicylates in monolayer cell cultures. *Gastroenterology* 74:205–208.

Traina, V. M. (1983). The role of toxicology in drug research and development. *Med. Res. Rev.* 3:43–72.

Trifillis, A., Regec, A., Hallcraggs, M., and Trump, B. (1984). Effects of cyclosporine on cultured human renal tubular cells. *Hum. Toxicol.* 3:454.

Tune, B. M., Fravert, D., and Hsu, C. Y. (1989). Oxidative and mitochondrial toxic effects of cephalosporin antibiotics in the kidney: A comparative study of cephaloridine and cephaloglycine. *Biochem. Pharmacol.* 38:795–802.

Tune, B. M., Sibley, R. K., and Hsu, C. Y. (1988). The mitochondrial respiratory toxicity of cephalosporin antibiotics. An inhibitory effect on substrate uptake. *J. Pharmacol. Exp. Ther.* 245:1054–1059.

Tyson, C. A., and Frazier, J. M. (1993). *Methods in Toxicology*: In *Vitro Biological Systems*. Academic, San Diego, CA.

Viano, I., Eandi, M., and Santiano, M. (1983). Toxic effects of some antibiotics on rabbit kidney cells. *Int. J. Tiss. Reac.* 2:11–186.

Vinay, P., Gougoux, A., and Lemieux, G. (1981). Isolation of a pure suspension of rat proximal tubules. *Am. J. Physiol.* 241:F403–F411.

Ware, R. A., Burkholder, P. M., and Chang, L. W. (1975). Ultrastructural changes in renal proximal tubules after chronic organic and inorganic mercury intoxication. *Environ. Res.* 10:121–140.

Weil, C. S., and Scala, R. A. (1971). Study of intra- and interlaboratory variability in the results of rabbit eye and skin irritation tests. *Toxicol. Appl. Pharm.* 19:276–360.

Weinberg, J. M. (1985). Oxygen deprivation-induced injury to isolated rabbit kidney tubules. *J. Clin. Invest.* 76:1193–2108.

Weinberg, J. M., Davis, J. A., Abarzua, M., and Rajan, T. (1987). Cytoprotective effects of glycine and glutathione against hypoxic injury to renal tubules. *J. Clin. Invest.* 80:1446–1454.

Williams, G. M. (1976a). Carcinogen-induced DNA repair in primary rat liver cell cultures; a possible screen for chemical carcinogens. *Cancer Lett. (Shannon, Irel.)* 1:231–236.

Williams, G. M. (1976b). The detection of chemical carcinogens by unscheduled DNA synthesis in rat liver primary cell cultures. *Cancer Res.* 37:1845–1851.

Williams, G. M., Dunkel, V. C., and Ray, V. A. (Eds.) (1983). Cellular systems for toxicity testing. *Ann. NY Acad. Sci*. 407.

Williams, P. D. (1977). The application of renal cells in culture in studying drug-induced nephrotoxicity. *In Vitro Cel. Dev. Biol.* 25:800–805.

Williams, P. D. (1989). The application of renal cells in culture in studying drug-induced nephrotoxicity. *In Vitro* 25:800–805.

Williams, P. D., Bennett, D. B., Gleason, C. R., and Hottendorf, G. H. (1987). Correlation between renal membrane binding and nephrotoxicity of aminoglycosides. *Antimicrob. Ag. Chemother.* 31(4):570–574.

Williams, P. D., Hitchcock, M. J. M., and Hottendorf, G. H. (1985). Effects of cephalosporins on organic ion transport in renal membrane vesicles from rat and rabbit kidney cortex. *Res. Commun. Chem. Path. Pharm.* 47(3):357–371.

Williams, P. D., and Hottendorf, G. H. (1985). Effects of *cis*-dichlorodiamineplatinum-II (Cisplatin) on organic ion transport in membrane vesicles from rat kidney cortex. *Cancer Treat. Rep.* 69:875–880.

Williams, P. D., and Hottendorf, G. H. (1986). $^{3}$H-Gentaminicin uptake in brush border and basolateral membrane vesicles from rat kidney cortex. *Biochem. Pharmacol.* 35:2253–2256.

Williams, P. D., Hottendorf, G. H., and Bennett, D. B. (1986a). Inhibition of renal membrane binding and nephrotoxicity of aminoglycosides. *J. Pharmacol. Exp. Ther.* 237:919–925.

Williams, P. D., Laska, D. A., Heim, R. A., and Rush, G. F. (1993). Differential toxicity of parenteral antibiotic drugs in renal cells (LLC-$PK_1$) grown on permeable membrane filters. *Toxicol. Methods* 3(2):130–141.

Williams, P. D., Laska, D. A., and Hottendorf, G. H. (1986b). Comparative toxicity of aminoglycoside antibiotics in cell cultures derived from human and pig kidney. *In Vitro Toxicol.* 1:23–32.

Williams, P. D., Laska, D. A., Tay, L. K., and Hottendorf, G. H. (1988). Comparative toxicity of cephalosporin antibiotics in a rabbit kidney cell line (LLC-$RK_1$). *Antimicrob. Ag. Chemother.* 32(3):314–318.

Williams, P. D., Trimble, M. E., Crespo, L., Holohan, P. D., Freedman, J. C., and Ross, C. R. (1984). Inhibition of renal $Na^+$, $K^+$-adenosine triphosphatase by gentimicin. *J. Pharmacol. Exp. Ther.* 231(2):919–925.

Williams, S. J. (1984). Prediction of ocular irritancy potential from dermal irritation test results. *Fd. Chem. Toxicol.* 22:157–161.

Williams, S. J. (1985). Changing concepts of ocular irritation evaluation: Pitfalls and progress. *Fd. Chem. Toxicol.* 23:189–193.

Wilson, P. D., Dillingham, M. A., Breckon, R., and anderson, R. J. (1985). Defined human renal tubular epithelia in culture: Growth, characterization, and hormonal response. *Am. J. Physiol.* 248:F436–F443.

Yang, I. S., Goldinger, J. M., Hong, S. K., and Taub, M. (1988). Preparation of basolateral membranes that transport *p*-aminohippurate from primary cultures of rabbit kidney proximal tubule cells. *J. Cell. Physiol.* 135:481–487.

Young, M. F., Trombetta, L. D., and Sophia, J. V. (1986). Correlative in vitro and in vivo study of skeletal muscle irritancy. *Toxicologist* 6:1225.

Zbinden, G. (1987). *Predictive Value of Animal Studies in Toxicology*. Centre for Medicines Research, Carshalton, UK.

Zhang, G., and Stevens, J. L. (1989). Tansport and activation of *S*-(1,2-dichlorovinyl)-L-cysteine and *N*-acetyl-*S*-(1,2-dichlorovinyl)-L-cysteine in rat kidney proximal tubules. *Toxicol. Appl. Pharmacol.* 100:51–61.

# *Evaluation of Human Tolerance and Safety in Clinical Trials: FIM Trials and Beyond*

## 1 PHARMACEUTICAL DEVELOPMENT PROCESS AND SAFETY

As introduced in the beginning of this volume, the pharmaceutical development process is a long (6–16 years from drug inception to market approval) and costly ($100 million to $1 billion, depending on how one allocates costs) process, even when successful. It is shaped by medical needs, regulatory requirements, economics, our understanding of sciences and diseases, and limitations of technology. All of these interact to shape a process which serves to iteratively reduce risks (both economic and human safety), with the probability of failure being reduced in a stepwise fashion (Matoren, 1984; PhRMA, 2000). Figure 1 briefly summarizes this process.

For our purposes (that is, for a safety assessment perspective), the purpose of all nonclinical (animal and in vitro) testing is to reduce the risks and probability of adverse events in humans. But between initial nonclinical testing (and concurrent with additional animal testing) and a drug reaching the marketplace, the potential for having adverse effects in the general patient population it is intended for is further guarded against by a scheme of increasingly more powerful human (or "clinical") trials (Piantadosi, 1997). How safety is

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

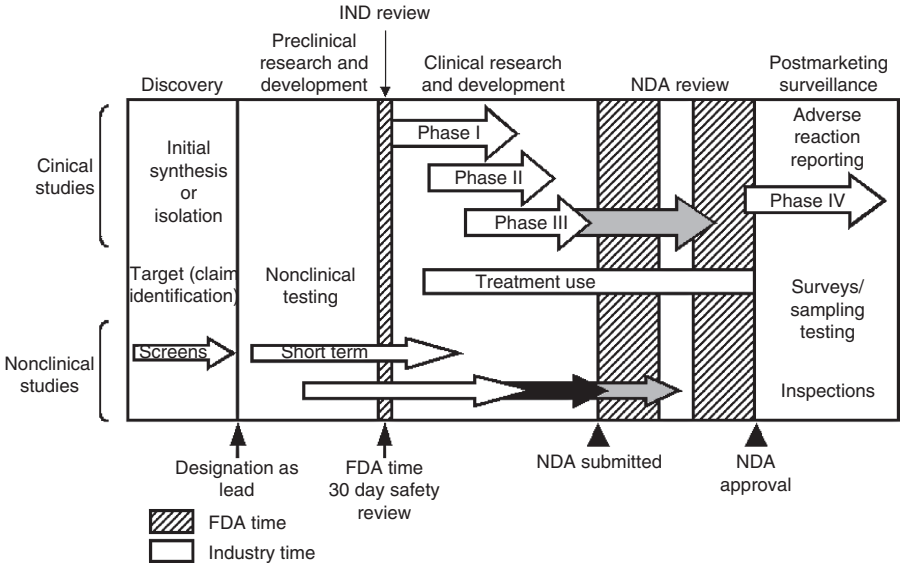


***Figure 1*** *Pharmaceutical development process viewed as four stages (discovery, preclinical development, clinical development, and NDA review) as well as important postmarket surveillance phase.*

evaluated in these is the subject of this chapter. The most common "unexpected" (from nonclinical trial results) safety findings in initial trials involve the skin (dermatitis of one form or another) and liver (Kaplowilz, 2001).

Except for those cases where there is substantial potential to save or extend lives [such as anticancer amyotrophic lateral sclerosis (ALS) and anti-AIDS drugs] or where the intended target diseases are chronic and severe [Parkinson's or multiple sclerosis (MS)] or the routes of administration are invasive (intrathecal), the initial evaluations in humans are performed in "normal," healthy volunteers with the primary objective being limited to defining the limits of tolerance (safety) of the potential drug and its pharmacokinetic characteristics. These trials may also seek to detect limited (usually surrogate—i.e., indirect) indicators of efficacy but are severely limited in doing so (Biomarkers Definitions, Working Group, 2001). Later trials look at the drug's actions on carefully defined groups of patients.

Something should be said about the special classes of studies (for cancer, HIV, ALS, and other life-threatening diseases) for which first-in-humans are not performed in normal healthy volunteers but rather in patients. Most frequently, these are patients that have failed other available forms of therapy. Even the design of such trials is different, though the monitored safety factors and pharmacokinetics are the same. As the current designs for oncology are proving to have exceptional failure rates, new "adaptive" study designs are actively being pursued (Meille et al., 2008).

With the number of drugs withdrawn from the marketplace since 1990[1] (or, perhaps, the degree of media coverage of such withdrawals), public concern with the workings of the drug safety evaluation aspects of the development process has risen sharply (Ganter, 1999; Wechsler, 2001). It is currently estimated that in the United States adverse drug reactions (ADRs) rank between the fourth to sixth leading cause of death (Eikelblom et al., 2001). While improvements in the nonclinical aspects of drug safety assessment (as covered in the first 19 chapters of this book) are possible and even likely, clearly the clinical aspects (the subject of this chapter) are likely to be where most improvement is likely. This will come from both better selection of subjects for inclusion in trials and a better understanding of individual or subpopulation differences in human responses to drugs.

While there is much press about the concern that the "increased pace of drug approval" has caused the release onto the market of less safe drugs (Willman, 2000), the causes are more mundane and of much longer standing. Additionally, the actual flow of new drugs entering the market is reduced. An important reason for the high incidence of serious and fatal ADRs is that the existing drug development paradigms do not generate adequate information on the mechanistic sources of marked variability in pharmacokinetics and pharmacodynamics of new therapeutic candidates, precluding treatments from being tailored for individual patients (Ozdemir et al., 2001).

Pharmacogenetics is the study of the hereditary basis of person-to-person variations in drug response. The focus of pharmacogenetic investigations has traditionally been unusual and extreme drug responses resulting from a single gene effect. The Human Genome Project and recent advancements in molecular genetics now present an unprecedented opportunity to study all genes in the human genome, including genes for drug metabolism, drug targets, and postreceptor second-messenger mechanisms, in relation to variability in drug safety and efficacy. In addition to sequence variations in the genome, high throughput and genomewide transcript profiling for differentially regulated messenger RNA (mRNA) species before and during drug treatment will serve as important tools to uncover novel mechanisms of drug action. Pharmacogenetic-guided drug discovery and development represent a departure for the conventional approach which markets drugs for broad patient populations, rather than smaller groups of patients in whom drugs may work more optimally.

Pharmacogenetics provides a rational framework to minimize the uncertainty in the outcome of drug therapy and clinical trials and thereby should significantly reduce the risk of drug toxicity. The reader is referred to the Internet sources in Table 1 for more details on pharmacogenetics and drug development. Potential improvements in patient inclusion criteria will be addressed later in this chapter.

[1] See Table 1.1.

**TABLE 1 Educational Sources on Pharmacogenetics and Drug Development Available on WorldWide Web**

| Source | Focus | Web Address |
|---|---|---|
| Affymetrix | DNA microarray technology | www.affymetrix.com |
| Celera Genomics | Human genome sequencing and variation | www.celera.com |
| Center for Drug Development Science | Drug development | www.dml.georgetown.edu/depts/pharmacology/cdds/index.html |
| Center for Ecogenetics and Environmental Health | Gene–environment interactions | depts.washington.edu/ceeh |
| Cold Spring Harbor Laboratory | Genetics education | www.cshl.org |
| Food and Drug Administration | Drug development and regulation | www.fda.gov |
| Genaissance Pharmaceuticals | Human genetic variation | www.genaissance.com |
| Genset Corporation | Genomics and drug development | www.genxy.com/index/html |
| Human Genic Bi-allelic Sequences Database | SNPs | http://hgbase.cgr.ki.se |
| Human Genome Project | Human genetic variation | www.ornl.gov/TechResources/Human_Genome/home.html |
| Karolinska Institute | Genetics of drug metabolism | www.imm.ki.se/CYPalleles |
| National Institutes of Health | Glossary of genetic terms | www.nhgri.nih.gov/DIR/VIP/Glossary |
| Nature Genetics | Genomics | http://www.nature.com/genomics |
| Orchid Biocomputer | SNPs | www.snps.com |
| Pharsight Corporation | Drug development | www.pharsight.com |
| SNP Consortium | Human genetic variation | Snp.cshl.org |
| Stanford University | Genome resources | www-genome.stanford.edu/index.html |
| Whitehead Institute | Genome resources | www-genome.wi.mit.edu |

*Note*: SNPs = single nucleotide polymorphisms.

## 1.1 Pharmacokinetics

Current costs and time pressures for developing a new therapeutic motivate companies to make the best possible decision as to whether to continue or abandon the development of a new drug based on a likelihood matrix for three factors: does it work (efficacy), is it acceptably tolerated at therapeutic doses (safety), and is it possible to deliver those therapeutic doses to the target sites/organs economically via the desired route (bioavailability). These likelihoods change most rapidly in early phase development and most compounds abandoned during clinical development are abandoned in this phase, with the most common reason begin unsuitable pharmacokinetics (is it absorbed and does it stay at therapeutic levels for an optimal or near-optimal period—Rolan,

1997). Hence the assessment of pharmacokinetics in early-phase drug development is strategically important. Some of the drug development issues which are likely to be answered at least in part by a thoughtful interpretation of pharmacokinetic data include the following:

1. Is the compound adequately absorbed to be likely to have a therapeutic effect?
2. Is the compound absorbed with a speed consistent with the desired clinical response?
3. Does the compound stay in the body long enough to be consistent with the desired duration of action?
4. Is the within- or between-subject variability acceptable given the likely therapeutic index of the compound?
5. Is there evidence of a formulation problem?
6. Is there a dose range which produces plasma (or tissue) concentrations which are likely to be associated with a desired clinical response or which gives rise to safety concerns?
7. Is there a relationship between plasma concentrations and a relevant measure of drug effect?
8. Are metabolites produced which may confound the therapeutic response or safety profile?
9. From the absorption, metabolism, and excretion profile, are there subsets of the target population which may behave differently from expected?
10. Considering the above issues, what is a suitable dosing regimen for clinical efficacy trials?

As most drugs are preferably given orally, absorption which is complete, consistent, and predictable is desirable. Although it may be possible from solubility, lipophilicity, $pK_a$, molecular size, and animal data to make some prediction about likely absorption, only a study in humans will give quantitative data as the mechanisms of drug absorption are complex and still incompletely understood (Washington et al., 2001). It may be helpful here to distinguish between the terms *absorption* and *bioavailability*.

Absorption refers to the fraction of the administered dose which is taken into the body. If a drug is taken up into intestinal cells but then extensively metabolized, it is still regarded as having been absorbed. However, for a drug to be bioavailable, unchanged drug must reach the systemic circulation. Hence a drug with a very high first-pass metabolism might be well absorbed but poorly bioavailable. Although in therapeutic terms poor absorption and poor bioavailability pose similar problems, it is important to distinguish between them because there are likely to be different possible solutions. Poor absorption might be approached by reformulation, change in the route of administration, or development of a prodrug; extensive presystemic metabolism might only be avoided by change in the route of administration or chemical modifica-

tion. Poor absorption is still frequently encountered in modern drug development because the rational drug discovery process often puts more emphasis on potency and selectivity (because these programs are run by biochemists and pharmacologists) than on factors likely to be associated with good absorption. This can result in lead compounds which perform very well in vitro but which may present major bioavailability and/or formulation problems (see discussion of this by Rolan et al., 1994).

Quantitative assessment of the extent of absorption (absolute bioavailability) is most rigorously obtained by comparison of the areas under the plasma concentration–time curves (after adjusting for dose) following intravenous (IV) and oral administration. However, even after oral administration alone some idea of absorption or bioavailability can be obtained in the following ways:

1. If a drug is not substantially metabolized, urinary excretion of unchanged drug may be a useful measure of absorption and bioavailability.
2. If a drug is substantially metabolized but it is reasonable to assume that metabolites are not produced in the gut lumen, urinary recovery of drug and metabolites might be a useful measure of absorption.
3. If the "apparent" plasma clearance (dose/area under the plasma concentration–time curve; equivalent to true clearance/fraction of dose absorbed) gives an implausibly high value of clearance (e.g., greater than hepatic and renal plasma flow), it is likely the bioavailability is low. However, this could be due to presystemic metabolism in addition to low absorption.
4. If there is a very large within- or between-subject variability in "apparent" clearance, this might indicate variable absorption or bioavailability, which in turn is often seen when absorption or bioavailability is low.

Whether absorption is related to the formulation or to an intrinsic property of the molecule can be determined by comparing absorption from a solid formulation and an oral solution, ideally with an IV solution as a reference.

Some idea of the rate of absorption can be obtained from examination of the plasma concentration–time profile. It should be remembered, however, that the time to maximum plasma concentration ($t_{max}$) is not when absorption is complete but when the rates of drug absorption and elimination are equal. Thus two drugs with the same absorption rate will differ in $t_{max}$ if elimination rates differ. Assessment of the rate of absorption can also be confounded by complex or slow drug distribution. For example, the calcium channel blocker amlodipine has a much later $t_{max}$ than other similar drugs. This is not due to slow absorption but to partitioning in the liver membrane with slow redistribution. A quantitative assessment of the rate of absorption can be obtained by deconvolution of plasma profiles following IV and oral administration.

***Relating Time Course of Plasma Concentrations to Time Course of Effect*** A critical decision to be made after the first human study is whether the compound's speed of onset and duration of action are likely to be consistent with the desired clinical response. Speed of onset is clearly of interest for treatments which are taken intermittently for symptom relief, for example, acute treatments for migraine, analgesics, or antihistamines for hay fever. Duration of action is particularly important when the therapeutic effect needs to be sustained continuously, for example, anticonvulsants. The first information on the probable time course of action often comes from the plasma pharmacokinetic profile. However, it has become increasingly evident that the profile alone may be misleading, with the concentration–time and the effect–time curves being substantially different. Some reasons for this, including examples, are as follows:

1. The effect may be delayed with respect to plasma concentration because of slow uptake into the target tissue from the plasma. A well-known example is digoxin, where there is a delay of several hours between peak plasma concentration and peak effect.
2. The effect may wane faster than the plasma elimination curve due to tolerance, for example, benzodiazepines and nitrates.
3. The effect may persist despite apparent elimination from plasma. This can occur with an irreversible effect of the drug (e.g., acetylation of platelet cyclooxygenase by aspirin). Another reason is very tight binding of the drug near the receptor (e.g., salmeterol) or concentration and trapping in the target tissue (omeprazole).
4. The formation of active metabolites may also contribute to a delay in onset and/or prolongation of action.

Some of these mechanisms may become apparent during animal pharmacology studies, but the clinical pharmacologist must always be aware of the possible discrepancy between concentration and effect–time curves. Clearly, if a relevant drug effect can also be measured in early human studies, establishing a relationship between plasma concentration and effect may be possible. If the desired clinical effect can be measured directly (e.g., blood pressure for an antihypertensive drug), the pharmacokinetic profile may not contribute greatly to the assessment of the time course of action, but these circumstances are the exception rather than the rule. Because of the many causes of discrepancies between the time course of drug concentrations and effect, and often the difficulty in measuring the clinical effects directly, a potentially useful approach comes from the use of surrogate markers of drug effect (discussed elsewhere in this book) combined with pharmacokinetic–pharmacodynamic (PK/PD) modeling to explore the relationships between dose, plasma concentrations, and effects.

## 1.2 Safety of Clinical Trial Subjects

While there continues to be increased interest in and concern about the safety of marketed ("approved") drugs [Ioannides and Lau (2001), e.g., have published a study showing that adverse safety findings are frequently low compared to actual numbers], as of this writing the public's confidence in the safety of participants in trials is at a low level (Lee et al., 2006). Certainly, the fialuridine (FIAU) tragedy (Meinert, 1996) and the case of the death of a healthy volunteer in a Johns Hopkins trial brought this issue to the forefront of the public mind. Amid estimates of as many as 5000 subject deaths per year in federally funded clinical trials (out of seven million individuals enrolled in such trials) (Shamoo, 2000; Wilson, 1998; Davis, 1998; Association of American Universities, 2000; Henney, 2000), the current guidelines and procedures should be clearly understood and carefully adhered to but are likely to be changed. As a starting place, Table 2 presents a glossary of key terms employed in this discussion.

There are international regulations which govern (on a country-by-country basis) the conduct of clinical trials, with many national governments expecting researchers to follow specific guidelines, such as the International Conference on Harmonization (ICH, 1997) Guideline for Good Clinical Practice. Regulations and guidelines are generally based on the principles of the Nuremberg Convention. The Nuremberg Code was written in 1946 in an effort to prevent recurrence of the human experimentation atrocities of World War II. This document states that all research in humans should be done with the well-being of the subject being the primary concern (Schmidt, 2001; O'Donnell, 2005).

The 1964 Declaration of Helsinki includes significant detail about clinical trial practices and the rights of potential subjects to be informed about risks, benefits, and alternative therapies (World Medical Association, 2001). It has been amended several times, most recently in 2000, when the use of placebos in trials employing patients was pronounced to be unethical (Mackintosh, 2001). Together, several parts of the U.S. Code of Federal Regulations (21 CFR 50, 21 CFR 54, 21 CFR 56, 21 CFR 312) constitute the good clinical practice (GCP) regulations for studies conducted in the United States. The regulations detail the responsibilities of sponsors, investigators, and IRBs (institutional review boards) and also outline monitoring practices to ensure regulatory and study design compliance and subject safety. Similarly, the ICH guidelines on GCPs provide detailed instructions for investigators, institutions, sponsors, and IRBs.

As with preclinical matters, during the 1990s, the ICH has brought together regulatory agencies and industry representatives from the United States, Europe, and Japan—and observers from all over the world—to agree to a single set of technical requirements for the registration of pharmaceuticals for human use. This process is now almost complete. The ICH Guideline for Good Clinical Practice has been adopted by the three lead regions and by many other countries (ICH, 1997). As developing nations begin establishing

**TABLE 2 Key Terms**

*Adverse event (or adverse experience):* Any untoward medical occurrence in a patient or clinical investigation subject administered a pharmaceutical product and which does not necessarily have to have a causal relationship with this treatment.

*Adverse drug reaction (ADR):* In the preapproval clinical experience with a new medicinal product or its new usages, particularly as the therapeutic dose(s) may not be established "all noxious and unintended responses to a medicinal product related to any dose should be considered adverse drug reactions."

*IND:* Investigational new drug application filed with FDA after preclinical testing is complete asking for permission to proceed with human tests.

*Efficacy pharmacology:* Evaluation of a drug's characteristics, effects, and uses with regard to the target illness and its interactions with living organisms.

*Healthy volunteer:* A healthy person who agrees to participate in a clinical trial for reasons other than medical and receives no direct health benefit from participating.

*Human subject:* An individual who is or becomes a participant in research either as a recipient of the test article or as a control. A subject may be either a healthy human or a patient (21 CFR 50.3).

*Nonclinical studies:* Studies in living systems (animals or cells) other than humans.

*Phase I:* Initial safety trials on a new medicine in which investigators attempt to establish the dose range tolerance for single and multiple doses in about 20–80 healthy volunteers.

*Phase II:* Pilot clinical trials to evaluate efficacy, safety, and therapeutic dose ranges in selected populations of about 100–300 subjects who have the disease or condition to be treated, diagnosed, or prevented.

*Phase III:* Multicenter studies in populations of perhaps 100–3000 subjects (or more) for whom the medicine is eventually intended.

*Phase IV:* Postmarketing trials to provide additional details about the product's safety, efficacy, and additional uses.

*Preclinical studies:* Animal studies that support phase I safety and tolerance studies and must comply with good laboratory practice (GLP). Other preclinical studies are done in discovery research laboratories to support drug efficiency claims.

*Serious adverse event:* A serious adverse event (experience) or reaction is any untoward medical occurrence that at any dose:

- Results in death.
- Is life threatening.
- Requires inpatient hospitalization or prolongation of existing hospitalization.
- Results in persistent or significant disability/incapacity.
- Is a congenital anomaly/birth defect.

*Subject/trial subject:* An individual who participates in a clinical trial, either as recipient of the investigational product(s) or as a control. (ICH) See also healthy volunteers, human subject (ICH 1.57).

*Unexpected adverse drug reaction:* An adverse reaction, the nature or severity of which is not consistent with the applicable product information (e.g., investigator's brochure for an unapproved investigational medicinal product).

*Source*: Machin et al., 2004; Fletcher et al., 2006.

practices for the testing and registration of new molecular entities, many are using ICH guidelines as standards.

Thus, during the past 50 years, the conduct of clinical drug research has improved because of regulations, guidelines, and policies put in place to protect subjects. Individual pharmaceutical companies have used these guidelines and

regulations as the basis for their standard operating procedures (SOPs), technical operations policies, and training programs to direct work processes and staff in their research. Most companies have created quality assurance (QA) units to oversee their researchers' adherence to agency guidelines and regulations and to their own company policies and practices.

In the United States, no clinical drug research can begin without prior Food and Drug Administration (FDA) review of the investigational new drug (IND) application, which includes the human testing protocol and associated preclinical testing results. While formal FDA approval of such an application is not legally a requirement, assent before proceeding is a prudent goal. Regulators in some countries require only notification of intent to initiate first-in-human studies. Before moving on to phase II or phase III studies, pharmaceutical companies and other sponsors must submit the information gathered to date for agency reviews.

For new chemical entities (NCEs), new indications or new formulation companies must file an IND which must be approved by the FDA before a drug can be used in humans. The requirements for ADR reporting for INDs are thus known as the IND regulations. Before a new product can be marketed, companies must file a new drug application (NDA) and have it approved by the FDA. The requirements for ADR reporting after marketing are thus known as the NDA regulations. Both sets of regulations can apply to a drug at the same time; for instance the NDA regulations apply to any marketed forms, but the IND regulations apply to a new indication or formulation. At the time of writing, the current regulations are in 21 CFR as follows:

- 21 CFR 312.32: Safety reports for investigational products subject to an IND application (published 1987)
- 21 CFR 314.80: Postmarketing reporting of ADEs (NDA) (published 1985)
- 21 CFR 600.80: Postmarketing reporting of adverse experiences for licensed biological products (includes vaccines) (published 1994)

See also CFR website: http://www.access.gpo.gov/nara/cfr/index.html. The August 1997 guideline "Post-marketing adverse experience reporting for human drug and licensed biological products: Clarification of what to report" defined the minimum data relevant for a safety report as:

- An identifiable patient
- An identifiable reporter
- A suspect drug or biological product
- An adverse event or fatal outcome

If any of these items remain unknown after being actively sought, a report should not be submitted to the FDA. The guideline also clarifies that adverse

experiences derived during planned contacts and active solicitation of information from patients (e.g., company-sponsored patient support programs, disease management programs) should be handled as safety information from a postmarketing study (i.e., for expedited reporting, events must be serious and unexpected and there must be a reasonable possibility that the drug caused the event).

See also the Center for Drug Evaluation and Research (CDER) guidance page http://www.fda.gov/cder/guidance/index.htm.

In the *Federal Register* of October 27, 1994, the FDA published a proposed rule to amend the regulations to provide consistency with certain standardized definitions, procedures, and formats developed by the ICH and the Council for International Organizations of Medical Sciences (CIOMS). The FDA received many comments on these proposals and finally published detailed amended expedited safety reporting regulations, implementing the ICH in October 1997 (*Federal Register*, October 7, 1997, Volume 62, Number 194, pages 52237–52253). A revision of the associated guideline has also been proposed by the FDA but was not available at the time of writing.

These new regulations for expedited reporting are effective 180 days later on April 6, 1998, but companies may comply with the provisions of this final rule before its effective date. Key points from the new regulations are described below. The amendments to the periodic postmarketing safety reporting regulations are delayed awaiting further consideration of the ICH E2C guideline. More recently, as of the start of 2008, all new drug clinical trials must be registered with ClinicalTrials.gov as part of the IND process. The form required to register, FDA-3674, is provided as Figure 2.

***IND Regulations*** In 1994 the FDA proposed to amend requirements for clinical study design, conduct, and annual sponsor reporting in the IND regulations as a result of events with FIAU. In the light of comments received, the FDA withdrew the proposed amendments and will develop a guidance document with recommendations on study design and monitoring of investigational drugs used to treat serious and potentially fatal illnesses, with particular attention to detection of adverse events similar to those caused by underlying disease.

***Increased Frequency Reports*** The requirement for increased frequency reports for serious expected ADRs with marketed products is revoked. This was also published in the *Federal Register*, June 25, 1997 (Volume 62, Number 122, pages 34166–34168). The rationale for this was that despite receiving many such reports, only a small number of drug safety problems were identified.

***Reporting Forms*** FDA form 3500/3500A (refer to Figure 3) is the standard form for notifying expedited reports and can also be used by companies to submit IND safety reports. Foreign cases may be reported on the CIOMS I form.

See OMB Statement on Reverse. Form Approved: OMB No. 0910-0616, Expiration Date: 09-30-2008

DEPARTMENT OF HEALTH AND HUMAN SERVICES
Food and Drug Administration

**Certification of Compliance, under 42 U.S.C. § 282(j)(5)(B), with Requirements of ClinicalTrials.gov Data Bank (42 U.S.C. § 282(j))**

(For submission with an application/submission, including amendments, supplements, and resubmissions, under §§ 505, 515, 520(m), or 510(k) of the Federal Food, Drug, and Cosmetic Act or § 351 of the Public Health Service Act.)

**SPONSOR / APPLICANT / SUBMITTER INFORMATION**

1. NAME OF SPONSOR/APPLICANT/SUBMITTER
2. DATE OF THE APPLICATION/SUBMISSION WHICH THIS CERTIFICATION ACCOMPANIES
3. ADDRESS *(Number, Street, State, and ZIP Code)*
4. TELEPHONE AND FAX NUMBER *(Include Area Code)*
   (Tel.) _____
   (Fax) _____

**PRODUCT INFORMATION**

5. **FOR DRUGS/BIOLOGICS:** Include Any/All Available Established, Proprietary and/or Chemical/Biochemical/Blood/Cellular/Gene Therapy Product Name(s)
   **FOR DEVICES:** Include Any/All Common or Usual Name(s), Classification, Trade or Proprietary or Model Name(s) and/or Model Number(s)
   *(Attach extra pages as necessary)*

**APPLICATION / SUBMISSION INFORMATION**

6. TYPE OF APPLICATION/SUBMISSION WHICH THIS CERTIFICATION ACCOMPANIES
   ☐ IND ☐ NDA ☐ ANDA ☐ BLA ☐ PMA ☐ HDE ☐ 510(k) ☐ PDP ☐ Other
7. INCLUDE IND/NDA/ANDA/BLA/PMA/HDE/510(k)/PDP/OTHER NUMBER *(If number previously assigned)*
8. SERIAL NUMBER ASSIGNED TO APPLICATION/SUBMISSION WHICH THIS CERTIFICATION ACCOMPANIES

**CERTIFICATION STATEMENT / INFORMATION**

9. CHECK ONLY ONE OF THE FOLLOWING BOXES *(See instructions for additional information and explanation)*
   ☐ A. I certify that the requirements of 42 U.S.C. § 282(j), Section 402(j) of the Public Health Service Act, enacted by 121 Stat. 823, Public Law 110-85, do not apply because the application/submission which this certification accompanies does not reference any clinical trial.
   ☐ B. I certify that the requirements of 42 U.S.C. § 282(j), Section 402(j) of the Public Health Service Act, enacted by 121 Stat. 823, Public Law 110-85, do not apply to any clinical trial referenced in the application/submission which this certification accompanies.
   ☐ C. I certify that the requirements of 42 U.S.C. § 282(j), Section 402(j) of the Public Health Service Act, enacted by 121 Stat. 823, Public Law 110-85, apply to one or more of the clinical trials referenced in the application/submission which this certification accompanies and that those requirements have been met.
10. IF YOU CHECKED BOX C, IN NUMBER 9, PROVIDE THE NATIONAL CLINICAL TRIAL (NCT) NUMBER(S) FOR ANY "APPLICABLE CLINICAL TRIAL(S)," UNDER 42 U.S.C. § 282(j)(1)(A)(i), SECTION 402(j)(1)(A)(i) OF THE PUBLIC HEALTH SERVICE ACT, REFERENCED IN THE APPLICATION/SUBMISSION WHICH THIS CERTIFICATION ACCOMPANIES *(Attach extra pages as necessary)*
    NCT Number(s): _____

The undersigned declares, to the best of her/his knowledge, that this is an accurate, true, and complete submission of information. I understand that the failure to submit the certification required by 42 U.S.C. § 282(j)(5)(B), section 402(j)(5)(B) of the Public Health Service Act, and the knowing submission of a false certification under such section are prohibited acts under 21 U.S.C. § 331, section 301 of the Federal Food, Drug, and Cosmetic Act.
**Warning:** A willfully and knowingly false statement is a criminal offense, U.S. Code, title 18, section 1001.

11. SIGNATURE OF SPONSOR/APPLICANT/SUBMITTER OR AN AUTHORIZED REPRESENTATIVE (Sign)
12. NAME AND TITLE OF THE PERON WHO SIGNED IN NO. 11
    (Name) _____
    (Title) _____
13. ADDRESS *(Number, Street, State, and ZIP Code) (of person identified in No. 11 and 12)*
14. TELEPHONE AND FAX NUMBER *(Include Area Code)*
    (Tel.) _____
    (Fax) _____
15. DATE OF CERTIFICATION

**FDA-3674 (1/08) (FRONT)** PSC Graphics (301) 443-1090 EF

***Figure 2*** *FDA form 3674.*

**Instructions for Completion of Form FDA 3674**

**Certification of Compliance, under 42 U.S.C. § 282(j)(5)(B), with Requirements of ClinicalTrials.gov Data Bank (42 U.S.C. § 282(j))**
Form 3674 must accompany an application/submission, including amendments, supplements, and resubmissions, submitted under §§ 505, 515, 520(m), or 510(k) of the Federal Food, Drug, and Cosmetic Act or § 351 of the Public Health Service Act.

1. **Name of Sponsor/Applicant/Submitter -** This is the name of the sponsor/applicant/submitter of the drug/biologic/device application/submission which the certification accompanies. The name must be identical to that listed on the application/submission.

2. **Date -** This is the date of the application/submission which the certification accompanies.

3. **& 4. -** Provide complete address, telephone number and fax number of the sponsor/applicant/submitter.

5. **Product Information - For Drugs/Biologics:** Provide the established, proprietary name, and/or chemical/biochemical/blood product/cellular/gene therapy name(s) for the product covered by the application/submission. Include all available names by which the product is known. **For Devices:** Provide the common or usual name, classification, trade or proprietary or model name(s), and/or model number(s). Include all available names/model numbers by which the product is known.

6. **Type of Application/Submission -** Identify the type of application/submission which the certification accompanies by checking the appropriate box. If the name of the type of application/submission is not identified, check the box labeled "Other."

7. **IND/NDA/ANDA/BLA/PMA/HDE/510(k)/PDP/Other Number -** If FDA has previously assigned a number associated with the application/submission which this certification accompanies, list that number in this field. For example, if the application/submission accompanied by this certification is an IND protocol amendment and the IND number has already been issued by FDA, that number should be provided in this field.

8. **Serial Number -** In some instances a sequential serial number is assigned to the application. If there is such a serial number, provide it in this field.

9. **Certification -** This section contains three different check-off boxes.
   **Box A** should be checked if the sponsor/applicant/submitter has concluded that the requirements of 42 U.S.C. § 282(j), section 402(j) of the Public Health Service Act, do not apply because no clinical trials are included, relied upon, or otherwise referred to, in the application/submission which the certification accompanies.

   **Box B** should be checked if the sponsor/applicant/submitter has concluded that the requirements of 42 U.S.C. § 282(j), section 402(j) of the Public Health Service Act, do not apply at the time of submission to any clinical trials that are included, relied upon, or otherwise referred to, in the application/submission which the certification accompanies. This means that, at the time the application/submission is being made, the requirements of 42 U.S.C. § 282(j), section 402(j) of the Public Health Service Act, do not apply to any of the clinical trials included, relied upon, or otherwise referred to, in the application/submission which this certification accompanies.

   **Box C** should be checked if the sponsor/applicant/submitter has concluded that the requirements of 42 U.S.C. § 282(j), section 402(j) of the Public Health Service Act, do apply at the time of submission to some or all of the clinical trials that are included, relied upon, or otherwise referred to, in the application/submission which the certification accompanies. This means that, at the time the application/submission is being made, the requirements of 42 U.S.C. § 282(j), section 402(j) of the Public Health Service Act, apply to one or more of the clinical trials included, relied upon, or otherwise referred to, in the application/submission which this certification accompanies.

10. **National Clinical Trial (NCT) Numbers -** If you have checked Box C in number 9 (Certification), provide the NCT Number obtained from www.ClinicalTrials.gov for each clinical trial that is an "applicable clinical trial" under 42 U.S.C. § 282(j)(1)(A)(i), section 402(j)(1)(A)(i) of the Public Health Service Act, and that is included, relied upon, or otherwise referred to, in the application/submission which the certification accompanies. Type only the number, as NCT will be added automatically before number. Include any and all NCT numbers assigned to the clinical trials included, relied upon, or otherwise referred to, in the application/submission which this certification accompanies. Multiple NCT numbers may be required for a particular certification, depending on the number of "applicable clinical trials" included, relied upon, or otherwise referred to, in the application/submission which the certification accompanies.

11. **Signature of Sponsor/Applicant/Submitter or an Authorized Representative -** The person signing the certification must sign in this field.

12. **Name and Title of Person Who Signed in number 11. -** Include the name and title of the person who is signing the certification. If the person signing the certification is not the sponsor/applicant/submitter of the application/submission, he or she must be an authorized representative of the sponsor/applicant/submitter.

13. **& 14. -** Provide the full address, telephone and fax number of the person who is identified in number 11 and signs the certification in number 11.

15. Provide the date the certification is signed. This date may be different from the date provided in number 2.

**Paperwork Reduction Act Statement**

Public reporting burden for this collection of information is estimated to average 15 minutes and 45 minutes (depending on the type of application/submission) per response, including time for reviewing instructions. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to the applicable address below.

Food and Drug Administration
Center for Drug Evaluation and Research
Central Document Room
Form No. FDA 3674
5901-B Ammendale Road
Beltsville, MD 20705-1266

Food and Drug Administration
Center for Biologics Evaluation and Research
1401 Rockville Pike
Rockville, MD 20852-1448

Food and Drug Administration
Center for Devices and Radiological Health
Program Operations Staff (HFZ-403)
9200 Corporate Blvd.
Rockville, MD 20850

*An agency may not conduct or sponsor, and a person is not required to respond to, a collection of information, unless it displays a currently valid OMB control number.*

**FDA-3674 (1/08) (BACK)**

***Figure 2*** *Continued*

Form Approved: OMB No. 0910-0291, Expires: 10/31/08
See OMB statement on reverse.

**U.S. Department of Health and Human Services**
Food and Drug Administration

For use by user-facilities, importers, distributors and manufacturers for MANDATORY reporting

Mfr Report #
UF/Importer Report #
FDA Use Only

## MEDWATCH

**FORM FDA 3500A (10/05)**

**Page ____ of ____**

PLEASE TYPE OR USE BLACK INK

### A. PATIENT INFORMATION

1. **Patient Identifier** — In confidence
2. **Age at Time of Event:** or **Date of Birth:**
3. **Sex** ☐ Female ☐ Male
4. **Weight** ____ lbs or ____ kgs

### B. ADVERSE EVENT OR PRODUCT PROBLEM

1. ☐ **Adverse Event** and/or ☐ **Product Problem** *(e.g., defects/malfunctions)*
2. **Outcomes Attributed to Adverse Event** *(Check all that apply)*
   ☐ Death: _____ *(mm/dd/yyyy)*
   ☐ Life-threatening
   ☐ Hospitalization - initial or prolonged
   ☐ Disability or Permanent Damage
   ☐ Congenital Anomaly/Birth Defect
   ☐ Other Serious (Important Medical Events)
   ☐ Required Intervention to Prevent Permanent Impairment/Damage (Devices)
3. **Date of Event** *(mm/dd/yyyy)*
4. **Date of This Report** *(mm/dd/yyyy)*
5. **Describe Event or Problem**
6. **Relevant Tests/Laboratory Data, Including Dates**
7. **Other Relevant History, Including Preexisting Medical Conditions** *(e.g., allergies, race, pregnancy, smoking and alcohol use, hepatic/renal dysfunction, etc.)*

**Submission of a report does not constitute an admission that medical personnel, user facility, importer, distributor, manufacturer or product caused or contributed to the event.**

### C. SUSPECT PRODUCT(S)

1. **Name** *(Give labeled strength & mfr/labeler)* #1 #2
2. **Dose, Frequency & Route Used** #1 #2
3. **Therapy Dates** *(If unknown, give duration) from/to (or best estimate)* #1 #2
4. **Diagnosis for Use** *(Indication)* #1 #2
5. **Event Abated After Use Stopped or Dose Reduced?** #1 ☐ Yes ☐ No ☐ Doesn't Apply; #2 ☐ Yes ☐ No ☐ Doesn't Apply
6. **Lot #** #1 #2
7. **Exp. Date** #1 #2
8. **Event Reappeared After Reintroduction?** #1 ☐ Yes ☐ No ☐ Doesn't Apply; #2 ☐ Yes ☐ No ☐ Doesn't Apply
9. **NDC# or Unique ID**
10. **Concomitant Medical Products and Therapy Dates** *(Exclude treatment of event)*

### D. SUSPECT MEDICAL DEVICE

1. **Brand Name**
2. **Common Device Name**
3. **Manufacturer Name, City and State**
4. **Model #** **Lot #** **Catalog #** **Expiration Date** *(mm/dd/yyyy)* **Serial #** **Other #**
5. **Operator of Device** ☐ Health Professional ☐ Lay User/Patient ☐ Other: _____
6. **If Implanted, Give Date** *(mm/dd/yyyy)*
7. **If Explanted, Give Date** *(mm/dd/yyyy)*
8. **Is this a Single-use Device that was Reprocessed and Reused on a Patient?** ☐ Yes ☐ No
9. **If Yes to Item No. 8, Enter Name and Address of Reprocessor**
10. **Device Available for Evaluation?** *(Do not send to FDA)* ☐ Yes ☐ No ☐ Returned to Manufacturer on: _____ *(mm/dd/yyyy)*
11. **Concomitant Medical Products and Therapy Dates** *(Exclude treatment of event)*

### E. INITIAL REPORTER

1. **Name and Address**
**Phone #**
2. **Health Professional?** ☐ Yes ☐ No
3. **Occupation**
4. **Initial Reporter Also Sent Report to FDA** ☐ Yes ☐ No ☐ Unk.

***Figure 3*** *FDA Form 3500A.*

**MEDWATCH**

**FORM FDA 3500A (10/05)** *(continued)* **Page ____ of ____**

FDA USE ONLY

## F. FOR USE BY USER FACILITY/IMPORTER *(Devices Only)*

1. **Check One** ☐ User Facility ☐ Importer
2. **UF/Importer Report Number**
3. **User Facility or Importer Name/Address**
4. **Contact Person**
5. **Phone Number**
6. **Date User Facility or Importer Became Aware of Event** *(mm/dd/yyyy)*
7. **Type of Report** ☐ Initial ☐ Follow-up # _____
8. **Date of This Report** *(mm/dd/yyyy)*
9. **Approximate Age of Device**
10. **Event Problem Codes** *(Refer to coding manual)*
    Patient Code ____ - ____ - ____
    Device Code ____ - ____ - ____
11. **Report Sent to FDA?** ☐ Yes _____ *(mm/dd/yyyy)* ☐ No
12. **Location Where Event Occurred** ☐ Hospital ☐ Home ☐ Nursing Home ☐ Outpatient Treatment Facility ☐ Other: _____ *(Specify)* ☐ Outpatient Diagnostic Facility ☐ Ambulatory Surgical Facility
13. **Report Sent to Manufacturer?** ☐ Yes _____ *(mm/dd/yyyy)* ☐ No
14. **Manufacturer Name/Address**

## G. ALL MANUFACTURERS

1. **Contact Office - Name/Address** ***(and Manufacturing Site for Devices)***
2. **Phone Number**
3. **Report Source** *(Check all that apply)* ☐ Foreign ☐ Study ☐ Literature ☐ Consumer ☐ Health Professional ☐ User Facility ☐ Company Representative ☐ Distributor ☐ Other: _____
4. **Date Received by Manufacturer** *(mm/dd/yyyy)*
5. (A)NDA # _____ IND # _____ STN # _____ PMA/510(k) # _____ Combination Product ☐ Yes Pre-1938 ☐ Yes OTC Product ☐ Yes
6. **If IND, Give Protocol #**
7. **Type of Report** *(Check all that apply)* ☐ 5-day ☐ 30-day ☐ 7-day ☐ Periodic ☐ 10-day ☐ Initial ☐ 15-day ☐ Follow-up # ____
8. **Adverse Event Term(s)**
9. **Manufacturer Report Number**

## H. DEVICE MANUFACTURERS ONLY

1. **Type of Reportable Event** ☐ Death ☐ Serious Injury ☐ Malfunction ☐ Other: _____
2. **If Follow-up, What Type?** ☐ Correction ☐ Additional Information ☐ Response to FDA Request ☐ Device Evaluation
3. **Device Evaluated by Manufacturer?** ☐ Not Returned to Manufacturer ☐ Yes ☐ Evaluation Summary Attached ☐ No *(Attach page to explain why not)* or provide code: _____
4. **Device Manufacture Date** *(mm/yyyy)*
5. **Labeled for Single Use?** ☐ Yes ☐ No
6. **Evaluation Codes** *(Refer to coding manual)*
   Method ____ - ____ - ____ - ____
   Results ____ - ____ - ____ - ____
   Conclusions ____ - ____ - ____ - ____
7. **If Remedial Action Initiated, Check Type** ☐ Recall ☐ Repair ☐ Replace ☐ Relabeling ☐ Notification ☐ Inspection ☐ Patient Monitoring ☐ Modification/Adjustment ☐ Other: _____
8. **Usage of Device** ☐ Initial Use of Device ☐ Reuse ☐ Unknown
9. **If action reported to FDA under 21 USC 360i(f), list correction/removal reporting number:**
10. ☐ **Additional Manufacturer Narrative** and / or 11. ☐ **Corrected Data**

The public reporting burden for this collection of information has been estimated to average 66 minutes per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to:

Department of Health and Human Services
Food and Drug Administration - MedWatch
10903 New Hampshire Avenue
Building 22, Mail Stop 4447
Silver Spring, MD 20993-0002

***Please DO NOT RETURN this form to this address.***

**OMB Statement:**
"An agency may not conduct or sponsor, and a person is not required to respond to, a collection of information unless it displays a currently valid OMB control number."

***Figure 3*** *Continued*

***Definitions*** The definition of "serious" has been revised to make it consistent with ICH E2A and is the same for INDs and NDAs (see above).

The definition of "unexpected" for IND reporting is as follows:

> Any adverse drug experience, the specificity or severity of which is not consistent with the current investigator brochure; or if an investigator brochure is not required or available, the specificity or severity of which is not consistent with the risk information described in the general investigational plan or elsewhere in the current application, as amended. For example, under this definition, hepatic necrosis would be unexpected (by virtue of greater severity) if the investigator brochure only referred to elevated hepatic enzymes or hepatitis. Similarly, cerebral thromboembolism and cerebral vasculitis would be unexpected (by virtue of greater specificity) if the investigator brochure only listed cerebral vascular accidents. "Unexpected," as used in this definition, refers to an adverse drug experience that has not been previously observed (e.g., included in the investigator brochure) rather than from the perspective of such experience not being anticipated from the pharmacological properties of the pharmaceutical product.

***Time Frames*** The time period for submitting written IND safety reports has been revised from 10 working days to 15 calendar days. For telephone reports (fatal and life-threatening unexpected reactions), it has been revised from 3 working days to 7 calendar days. Such reports can also be made by fax. Telephone reporting was previously restricted to clinical studies conducted under the IND, but under the new rule, telephone reporting within 7 calendar days applies to any unexpected fatal or life-threatening reaction from any source.

The time period for submitting NDA alert reports (serious and unexpected) has been revised from 15 working days to 15 calendar days.

Wherever human drug research is conducted, national regulations call for an independent ethical review of the study plan. In countries where a guideline on GCPs is used, ethics review bodies are made up of medical professionals from the institution, nonmedical personnel, and community members. Sponsor companies and involved investigators have no voting representation on these review boards, and they may not be present during the voting on the research approval. The investigator conducting the study may, however, present the protocol and answer questions at the IRB review meeting. The company sponsoring the trial is not allowed to participate in IRB meetings as a matter of routine, although a representative might be invited to explain or clarify the protocol to the ethics review body.

*Informed consent* must be obtained from study participants—in writing—before any study-related activities are performed. Regulations clearly describe the required elements of the consent document and the consent process to be followed. A good informed consent process can help ensure that potential subjects understand the nature of the studies they will enter, the type of treat-

ments they will undergo, alternative therapies currently available, and any particular hazards they might experience. They must be informed that they can withdraw from the study at any time without penalty. Subjects are to be asked for their consent to release information from their medical records and told that the medical information may be inspected by sponsor, company, and regulatory agency representatives. They are to be informed that the results of the trials may be used publicly, but anonymously.

*Drug supplies* must be accounted for throughout the trial and reconciled at the end of the trial. These practices are designed to prevent the misuse or inappropriate redistribution of the investigational drug and to help ensure compliance with the protocol.

*Adverse events*, unexpected drug reactions, and drug side effects experienced and reported by subjects or observed by clinical investigators are all to be recorded and promptly reported to sponsor companies. It is clear, however, that there are differences in reporting standards between companies and between countries (Hayachi and Walker, 1996). There is also, however, a marked difference in reporting standards and rates in U.S. clinical trials between different medical areas (Ioannides and Lau, 2001). The investigators involved with subject care and the pharmaceutical company sponsor are then to analyze each event for "causality" and "relatedness" to administration of the drug—did this reaction occur because of the drug or because of something else such as the progression of the disease symptoms, other medications being taken, or unrelated causes? Such safety information is then to be forwarded to the appropriate IRB or ethics committee and regulatory agencies. The sponsor company is to send periodic updates to investigators, alerting them to new serious or unexpected drug reactions.

Company physicians are expected to continuously analyze the adverse drug event data coming in from worldwide trials for trends and patterns that could foretell a drug safety problem. Drug companies frequently set up data and safety monitoring boards (DSMBs) composed of noncompany medical experts and statisticians who impartially evaluate safety as the study progresses and are responsible for alerting the sponsor to unanticipated problems. Regulatory agencies also watch for trends because they are often in the best position to see safety trends across classes of drugs from any different companies.

Sponsor companies use quality assurance units independent of the clinical research group to audit medical operations. Their role is to ensure that regulatory standards and company policies and procedures for clinical research are being followed in all countries where research is being conducted.

*Regulatory agencies* oversee sponsor organizations, clinical trial processes, and clinical trial sites to verify that sponsors are conducting trials appropriately. Existing FDA regulations are conforming to ICH guidelines. When deficiencies are noted, agency inspectorates can restrict and penalize the offending academic IRB, investigator, and/or sponsor company. Investigator sites (including, in extreme cases, entire universities) can be prohibited from

conducting clinical research. Company studies can be rejected by regulatory agencies.

***Continuous Safety Monitoring*** Medical staff members at the clinical trial site and at the sponsor company are expected to be continuously alerted to adverse drug reactions or unexpected and serious medical problems that might be attributed to the new medication begin tested. The investigator and the site staff examine subjects and take vital sign measurements on the schedule designated by the protocol, the guide for study conduct. Each drug, as shown in its preclinical studies, has unique characteristics and the potential for adverse events or side effects that investigators need to watch for during the clinical trial. Because of this, safeguard activities are built into the protocol, such as the time intervals between subject visits, how often subjects are to be questioned and examined, the specific medical tests to be run at various time points, and the special diagnostic tests or interviews to be conducted. Staff members at the clinical site must record the medical information from these tests and from interviews and medical histories. Site staff must also transfer information from the medical source documents to the case report forms (CRFs) specific to the study. The CRFs contain key information required for the protocol. Clinical research associates and physicians are required to review the information regularly and to immediately report anything alarming to the IRB and regulatory agencies for further evaluation.

***Sponsor Pharmacovigilance*** Dedicated departments in pharmaceutical companies—often called pharmacovigilance groups—receive, review, analyze, follow up on, and appropriately distribute safety-related information from new drug trials. These groups sometimes staff hot lines and question-and-answer services to provide up-to-date answers to drug-related questions. Safety information is to be reported to regulatory agencies at specified intervals and at milestone time points throughout all phases of drug development. The postapproval aspects of this effort are the core of the subject of Chapter Application of In Vitro Techniques in Drug Safety Assessment.

*Sponsor monitoring* is another important oversight process to ensure quality, compliance, and subject safety. Monitors may be employees of the sponsor's medical staff or a contract research organization or may be independent contractors. In each case, they represent the sponsor and visit investigator sites regularly, perhaps every four to eight weeks. They examine subject records in detail and verify that the correct information was transferred to the clinical trial case report forms—a process called source data verification. During their site visits, monitors also examine administrative and regulatory documents, including drug supply and dosing records, adverse-event documentation and reporting, the informed consent process and forms, and case report forms.

Ensuring subject safety must be of foremost concern during the entire drug development process, especially in clinical trials. The gradual dosing of a drug

candidate in healthy human volunteer subjects is performed under tightly controlled conditions and under the direction and scrutiny of physicians trained in clinical pharmacology. A drug candidate's progress toward broader testing in a population of individuals with the disease or condition to be treated also moves prudently and in well-defined steps.

### 1.3 Limitations on/of Clinical Trials

Before looking more closely at the definition, structure, and designs of trials, one should understand their limitations. These are regulatory, economic, legal, and due to custom.

First, the most recent (October 2000) revision of the Declaration of Helsinki (World Medical Association, 2000) called for discontinuing the use of placebo-controlled trials in patients. While this is not currently binding on U.S. trials (the FDA has specifically said that they will not mirror this as a requirement) and is intended to protect the health of participating patient subjects by precluding having some denied existing efficacious treatments (which would be the effect in most, but not all, cases). It will also likely cause the numbers of subjects required in a trial to increase. This will further stretch the economic aspects of limitations on the power of trials to assess potential drug safety in what will be the intended patient population. Trials are already very expensive—each additional subject enrolled costs $15,000 or more in a phase II or III trial.

The legal (or rather, litigation) limitations are that any adverse event in a trial (or resulting from it) exposes a sponsor to potential litigation. Accordingly, trials are designed to exclude not only those individuals who are not in the precisely designed subject disease population but also those who represent potential additional risk subpopulations (the elderly, the young, those currently taking other drugs, minorities and women who are or may become pregnant) who are likely to eventually use the drug when it enters the marketplace.

Custom (continuing to do things as they have previously been done) also limits the power of trials to identify safety issues. While there are now regulatory inducements to include more women, the young, and ethnic minorities in trials, the first two groups still are not proportionately incorporated because of both the perceived risks of adverse events that they represent and because historically they have not been included. Ethnic minorities, particularly African Americans, present a different problem in that there is a historically based resistance to participation in such trials.

## 2 CLINICAL TRIALS PROCESS

While the numbers of animals involved in research is tracked closely and is well known, such is not the case for human subjects involved in clinical trials.

We simply do not know how many are involved in such trials in the United States, much less worldwide. Though the National Institutes of Health (NIH) does track closely how many dollars and individuals are involved in research it funds ($18.0 billion and eight million subjects in 1997), the same is not true for privately funded research (where the numbers are greater). And while there is now a website where one can examine the numbers and types of efficacy trials open, the same is not true for phase I tolerance and pharmacokinetic trials—where most potential drugs cease development.

Clinical drug development is often described as consisting of four distinct phases (I–IV). It is important to recognize that the phase of development provides an inadequate basis for classification of clinical trials because one type of trial may occur in several phases (see Figure 2). Table 3 presents a preferable (objective-based) classification of trial types. It is important to appreciate that the phase label is a description, not a set of requirements. It is

**TABLE 3 Approach to Classifying Clinical Studies According to Objective**

| Type of Study | Objective of Study | Study Examples |
|---|---|---|
| Human pharmacology (phase I) | • Assess tolerance<br>• Define/describe pharmacokinetics and pharmacodynamics<br>• Explore drug metabolism and drug interactions<br>• Estimate activity | • Dose–tolerance studies<br>• Single- and multiple-dose PK and/or PD studies<br>• Drug interaction studies |
| Therapeutic exploratory (phase Ib/II) | • Explore use for targeted indication<br>• Estimate dosage for subsequent studies<br>• Provide basis for confirmatory study design, endpoints, methodologies | • Earliest trials of relatively short duration in well-defined narrow patient populations using surrogate or pharmacological endpoints or clinical measures<br>• Dose–response exploration studies |
| Therapeutic confirmatory (phase III) | • Demonstrate/confirm efficacy<br>• Establish safety profile<br>• Provide adequate basis for assessing benefit–risk relationship to support licensing<br>• Establish dose–response relationship | • Adequate and well-controlled studies to establish efficacy<br>• Randomized parallel dose–response studies<br>• Clinical safety studies<br>• Studies of mortality/morbidity outcomes<br>• Large simple trials<br>• Comparative studies |
| Therapeutic use (phase III/IV) | • Refine understanding of benefit–risk relationship in general or special populations and/or environment<br>• Identify less common adverse reactions<br>• Refine dosing recommendation | • Comparative effectiveness studies<br>• Studies of mortality/morbidity outcomes<br>• Studies of additional endpoints<br>• Large simple trials<br>• Pharmacoeconomic studies |

also important to realize that these temporal phases do not imply a fixed order of studies since for some drugs in a development plan the typical sequence will not be appropriate or necessary. For example, although human pharmacology studies are typically conducted during phase I, many such studies are conducted at each of the other three stages but nonetheless sometimes labeled as phase I studies. Figure 4 demonstrates this close but variable correlation between the two classification systems. The distribution of the points of the graph shows that the types of study are not synonymous with the phases of development.

Drug development is ideally a logical, stepwise procedure in which information from small early studies is used to support and plan later larger, more definitive studies. To develop new drugs efficiently, it is essential to identify characteristics of the investigational drugs in the early stages of development and to plan an appropriate development based on this profile.

Initial trials provide an early evaluation of short-term safety and tolerability and can provide pharmacodynamic and pharmacokinetic information needed to choose a suitable dosage range and administration schedule for initial exploratory therapeutic trials. Later confirmatory studies are generally larger and longer and include a more diverse patient population. Dose–response information should be obtained at all stages of development, from early tolerance studies, to studies of short-term pharmacodynamic effect, to large efficacy studies. Throughout development, new data may suggest the need for additional studies that may commonly be part of an earlier phase. For example, blood-level data in a late trial may suggest a need for a drug–drug interaction

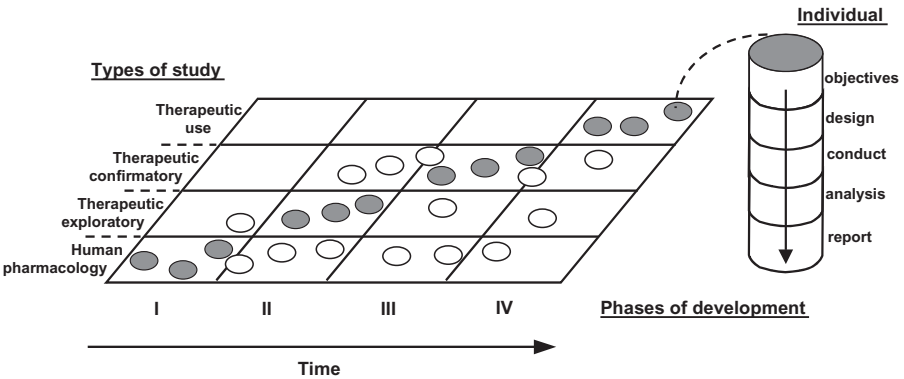


***Figure 4*** *Matrix illustrating relationship between phases of development and types of study by objective that may be conducted during each clinical development of new medicinal product. The shaded circles show the types of the study most usually conducted in a certain phase of development; the open circles show certain types of study that may be conducted in that phase of development but are less usual. Each circle represents an individual study. To illustrate the development of a single study, one circle is joined by a dotted line to an inset column that depicts the elements and sequence of an individual study.*

study or adverse effects may suggest the need for further dose finding and/or additional nonclinical studies. In addition, to support a new marketing application approval for the same drug, for example, for a new indication, pharmacokinetic or therapeutic exploratory studies are considered to be in phase I or II of development (Stone, 2006; Green et al., 2003; Gallin and Ognibene, 2007).

*Phase I* starts with the first-in-man administration of an IND. Although human pharmacology studies are typically identified with phase I, they may also be indicated at other points in the development sequence. Studies in this phase of development usually have nontherapeutic objectives and may be conducted in healthy volunteer subjects or certain types of patients, for example, patients with mild hypertension. Drugs with significant potential toxicity (cytotoxic drugs) or risks due to route of administration (such as intrathecal) are usually studied in patients. Studies in this phase can be open or baseline controlled or may use randomization and blinding to improve the validity of observations.

Studies conducted in phase I typically may involve one or a combination of the following:

(a) *Estimation of Initial Safety and Tolerability* The initial and subsequent administration of an IND into humans is usually intended to determine the tolerability of the dose range expected to be needed for later clinical studies and to determine the nature of adverse reactions that can be expected. These studies typically include both single- and multiple-dose administration.

(b) *Pharmacokinetics* Characterization of a drug's absorption, distribution, metabolism, and excretion continues throughout the development plan. Such preliminary characterization is an important goal of phase I. Pharmacokinetics may be assessed via separate studies or as a part of efficacy, safety, and tolerance studies. Pharmacokinetic studies are particularly important to assess the clearance of the drug and to anticipate possible accumulation of parent drug or metabolites and potential drug–drug interactions. Some PK studies are commonly conducted in later phases to answer more specialized questions. For many orally administered drugs, especially modified release products, the study of food effects on bioavailability is important. Obtaining PK information in subpopulations such as patients with impaired elimination (renal or hepatic failure), the elderly, children, women, and ethnic subgroups should be considered. Drug–drug interaction studies are important for many drugs and are generally performed in phases beyond phase I. But studies in animals and in vitro studies of metabolism and potential interactions may lead to doing such studies earlier.

(c) *Assessment of Pharmacodynamics* Depending on the drug and the endpoints studied, PD studies and studies relating drug blood levels to response (PK/PD studies) may be conducted in healthy volunteer subjects or in patients with the target disease. In patients, if there is an appropriate

measure, PD data can provide early estimates of activity and potential efficacy and may guide the dosage and dose regimen in later studies.

(d) *Early Measurement of Drug Activity* Preliminary studies of activity or potential therapeutic benefit may be conducted in phase I as a secondary objective. Such studies are generally performed in later phases but may be appropriate when drug activity is readily measurable with a short duration of drug exposure in patients at this early stage. Frequently such evaluations are done in what are called "phase Ib" studies.

*Phase II* is usually considered to start with the initiation of studies in which the primary objective is to explore therapeutic efficacy in patients.

Initial therapeutic exploratory studies may use a variety of study designs, including concurrent controls and comparison with baseline status. Subsequent trials are usually randomized and concurrently controlled to evaluate the efficacy of the drug and its safety for a particular therapeutic indication. Studies in phase II are typically conducted in a group of patients who are selected by relatively narrow criteria, leading to a relatively homogeneous population and are closely monitored.

An important goal for this phase is to determine dose levels and regimen for phase III trials. Early studies in this phase often utilize dose escalation designs to give an early estimate of dose response, and later studies may confirm the dose–response relationship for the indication in question by using recognized parallel dose–response designs (could also be deferred to phase III). Confirmatory dose–response studies may be conducted in phase II or deferred until phase III. Doses used in phase II are usually but not always less than the highest doses used in phase I.

Additional objectives of clinical trials conducted in phase II may include evaluation of potential study endpoints, therapeutic regimens (including concomitant medications) and target populations (e.g., mild versus severe disease) for further study in phase II or III. These objectives may be served by exploratory analyses, examining subsets of data, and including multiple endpoints in trials.

*Phase III* usually is considered to begin with the initiation of studies in which the primary objective is to demonstrate or confirm therapeutic benefit.

Studies in phase III are designed to confirm the preliminary evidence accumulated in phase II that a drug is safe and effective for use in the intended indication and recipient population. These studies are intended to provide an adequate basis for marketing approval. Studies in phase III may also further explore the dose–response relationship or explore the drug's use in wider populations, in different stages of disease, or in combination with another drug. For drugs intended to be administered for long periods, trials involving extended exposure to the drug are ordinarily conducted in phase III, although they may be started in phase III. ICH E1 and ICH E7 describe the overall clinical safety database considerations for chronically administered drugs and drugs used in the elderly. These studies carried out in phase III complete the

information needed to support adequate instructions for use of the drug (official product information).

*Phase IV* begins after drug approval. Once rare, there are now commonly required therapeutic use studies that go beyond the prior demonstration of the drug's safety, efficacy, and dose definition.

Studies in phase IV are all studies (other than routine surveillance) performed after drug approval and related to the approved indication. They are studies that were not considered necessary for approval but are often important for optimizing the drug's use. They may be of any type but should have valid scientific objectives. Commonly conducted studies include additional drug–drug interaction, dose response or safety studies, and studies designed to support use under the approved indication, for example, mortality/morbidity studies, epidemiological studies.

***Development of Application Unrelated to Original Approved Use*** After initial approval, drug development may continue with studies of new or modified indications, new dosage regimens, and new routes of administration or additional patient populations. If a new dose, formulation, or combination is studied, additional human pharmacology studies may be indicated, necessitating a new development plan.

The need for some studies may be obviated by the availability of data from the original development plan or from therapeutic use.

## 2.1 Special Considerations

A number of special circumstances and populations require consideration on their own when they are part of the development plan.

***Studies of Drug Metabolites*** Major active metabolite(s) should be identified and deserve detailed PK study. Timing of the metabolic assessment studies within the development plan depends on the characteristics of the individual drug.

***Drug–Drug Interactions*** If a potential for drug–drug interaction is suggested by metabolic profile, by the results of nonclinical studies, or by information on similar drugs, studies on drug interaction during clinical development are highly recommended. For drugs that are frequently coadministered it is usually important that drug–drug interaction studies be performed in nonclinical and, if appropriate, in human studies. This is particularly true for drugs that are known to alter the absorption or metabolism of other drugs or whose metabolism or excretion can be altered by effects by other drugs.

***Special Populations*** Some groups in the general population may require special study because they have unique risk–benefit considerations that need to be taken into account during drug development or because they can be

anticipated to need modification of use of the dose or schedule of a drug compared to general adult use. Pharmacokinetic studies in patients with renal and hepatic dysfunction are important to assess the impact of potentially altered drug metabolism or excretion. Specific ICH and FDA documents address such issues for geriatric patients and patients from different ethnic groups. The need for nonclinical safety studies to support human clinical trials in special populations is addressed in the ICH M3 document.

A key issue is thus when to perform kinetic studies in special patient groups (elderly, patients with renal or hepatic disease) and how. As the elderly are the majority users of many medicines, the subject of evaluating new drugs in the elderly is a major issue which is discussed in detail elsewhere. Unless a medication is unlikely to be used in the elderly, some data will be required by regulators for registration. However, an important issue in early phase development is whether to perform a separate elderly volunteer kinetic/tolerability study before elderly patients are included in later phase clinical trials. The major argument for doing so includes the possible reluctance of clinical investigators to enroll elderly patients without such data being available because of safety concerns; however, the utility of such studies has been questioned. The subjects in elderly volunteer studies are usually in much better health than the general population they are intended to represent. Also, the elderly may differ from the young not so much in terms of mean kinetic parameters but because the variability in the elderly may be much greater. The relatively small sample size (typically 12–18) may not allow a good estimation of the variability within the elderly population. For these reasons the FDA has recommended that information about the kinetics of a drug in the elderly should come from a larger group representative of the target population and this can be done in the efficacy clinical trials. Although these data are useful, it is not always an acceptable substitute for a specific elderly volunteer PK study because the information is only available after many patients have been exposed rather than before and clinical investigators may be reluctant to enroll patients without such data in advance. In practice, for a drug likely to be given to the elderly, an elderly volunteer study should be performed soon after a young healthy volunteer study to expand the potential population for efficacy studies as much as possible. If elderly patients are then included in the main efficacy/safety studies, the population approach can then be used to explore the pharmacokinetic variability in this subset of the population and whether this is associated with an altered clinical outcome.

A similar rationale can be used to decide whether special kinetic (and possibly dynamic) studies should be performed in patients with renal or hepatic disease. For example, if the compound is largely metabolized to inactive metabolites, renal function can reasonably be expected not to have a major effect on kinetics. However, regulators usually will want some information as some expectations do exist to the above assumption. An example is the "futile cycle" involving some nonsteroidal anti-inflammatory drugs (NSAIDs)—where prolonged residence of inactive acyl glucuronide metabolites in the

plasma in patients with renal disease allows breakdown back to the parent molecule resulting in accumulation (Sallustio et al., 1989). As outlined above, a population approach could be used to screen for an effect of disease on drug kinetics, but some investigators may need reassurance before enrolling patients in trials. A small study in patients with advanced renal disease may be able to provide this reassurance. Liver disease can be handled similarly for drugs which are primarily eliminated renally.

The safety of marketed drugs could be significantly improved if the subject groups involved in phase II and III trials better reflected the patient populations that will use drugs. By excluding "representatives" for what will clearly be subpopulations utilizing a drug (those using other drugs or with other diseases), many clear safety questions go unasked. The same is, to some degree, true about preclinical animal studies where only healthy young animals are employed rather than (perhaps) some disease model groups which might serve as better predictors of patient safety concerns.

## 3 INSTITUTIONAL REVIEW BOARDS (IRBS) IN CLINICAL TRIAL PROCESS

Clinical drug trials represent research with human subjects (Cato, 1988). All research involving human subjects that is supported by the federal government or the results of which are to be used in applications for drug or device approval must be conducted in accordance with regulations promulgated by the Department of Health and Human Services (DHHS) (45 CFR 46) and the FDA (21 CFR 56). The regulations of both the DHHS and the FDA require that an IRB "shall review and have authority to approve, required modifications in (to secure approval), or disapprove all research activities covered by [the] regulations" (45 CFR 46, 21 CFR 56).

The review of clinical drug trials by IRBs raises a number of interesting and difficult issues. These relate to the origin and sponsor of the proposed trial, the nature of the institution the IRB serves, and the manner in which the norms for determining ethical conduct in clinical trials can be applied to specific trials.

Here the ethical principles underlying research involving human subjects, the legal authority for IRBs, and the regulatory requirements affecting the operations of IRBs are reviewed. It will then discuss the role of IRBs in reviewing clinical trials by examining how IRBs can assess the scientific design of trials, the competency of the investigator, the manner of selecting subjects for the trial, the balance of risks and benefits, informed consent, and provisions for compensating for research-related injuries.

***Legal Authority for IRBs*** The legal authority for IRBs derives from two parallel sets of federal regulations. One set of regulations was promulgated by

the DHHS and implements the 1974 amendments to the Public Health Services Act (National Research Act, 1974). These regulations are codified in 45 CFR 46. The second set of regulations was promulgated by the FDA under the federal Food, Drug and Cosmetic (FD&C) Act. These regulations are codified in 21 CFR; regulations pertaining to IRBs are in Part 50 and those pertaining to informed consent are in Part 56.

The FDA has the legal authority to regulate clinical investigations in the United States when the investigational products move across state or national boundaries. Under the FDA regulations, review and approval by an IRB is required for any experiment that involves a test article and one or more human subjects, either patients or healthy persons, and that is subject to the requirements for prior submission to the FDA (21 CFR 50.3). Such review is also required for any experiment the results of which are intended to be submitted later to, or held for inspection by, the FDA.

The regulations of the FDA are identical or similar to those of the ICH and DHHS in nearly all essential respects. Such differences as do exist reflect the different statutory authority under which the separate sets of regulations were promulgated and the difference in mission between the FDA and the NIH, the agency within DHHS charged with overseeing the implementation and enforcement of the DHHS regulations. The difference in mission between the FDA and the NIH is reflected in the FDA's approach to compliance with its regulations utilizing its traditional tools of inspections and audits.

The FDA regulations specify requirements for IRB membership, function, and operation, and the criteria according to which approval may be given for conducting research. Since these requirements are similar, a single committee should be established to undertake the activities required by both sets of regulations. Additionally, the FDA regulations allow a wide variety of ways in which private practitioners not affiliated with an institution can obtain necessary IRB review of their clinical research activities. The basic ethical tenets governing the actions of an IRB should include (Sharp, 2001):

1. Risks to subjects are minimized: (i) by using procedures which are consistent with sound research design and which do not unnecessarily expose subjects to risk; and (ii) whenever appropriate, by using procedures already being performed on the subjects for diagnostic or treatment purposes.
2. Risks to subjects are reasonable in relation to anticipated benefits, if any, to subjects, and the importance of the knowledge that may reasonably be expected to result. …
3. Selection of subjects is equitable. …
4. Informed consent will be sought from each prospective subject or the subject's legally authorized representative. …
5. Informed consent will be appropriately documented. …

Additional requirements are that adequate provisions exist for monitoring the data collected, adequate provisions exist to protect the privacy of subjects and maintain the confidentiality of the data, and appropriate safeguards be included to protect the rights and welfare of subjects who are "vulnerable to coercion or undue influence … or persons who are economically or educationally disadvantaged" (45 CFR 46.1).

Before a trial initiates, formal review by an IRB that agrees to assume this additional function or by IRBs formed by a local or state health agency, a medical school, a medical society, a state licensing board, or a nonprofit or for-profit independent group is required. All IRBs, regardless of sponsorship, that are assuming responsibilities for reviewing and approving clinical research protocols subject to FDA authority must comply with the IRB regulations set out by the FDA.

***Duties of IRBs*** IRBs are required to review and have the authority to approve, require modifications in, or disapprove all research activities covered by the regulations [45 CFR 46.109(b)]. They must require that information given to subjects as part of informed consent is in accordance with the general requirements for informed consent that are set out in the regulations. Additionally, they may require that other information be given to subjects when they judge that such information would further protect the rights and welfare of the subjects [45 CFR 46.109(c)].

IRBs must require documentation of informed consent in all studies except those specified in the regulations in which documentation may be waived. Clinical drug trials are not among the classes of studies in which documentation of informed consent may be waived.

IRBs must provide written notification to investigators and institutions of their decisions to approve, require modifications in, or disapprove proposed research activities. Decisions to disapprove a proposed research proposal must be accompanied by a statement of reasons for the decision and provide the investigator an opportunity to respond in person or in writing.

IRBs must conduct continuing reviews of research they approve at least once each year. More frequent reviews may be required if the risk of a particular research project so warrants. IRBs have the authority to suspend or terminate approval of research that is not being conducted in accordance with their requirements or that has been associated with unexpected serious harm to subjects. Such action must be accompanied by a statement of reasons for it and be communicated to the investigator, appropriate institutional officials, and the secretary of DHHS.

The regulations require that IRBs must follow the written procedures that are set out in the assurances they have filed with DHHS, review proposed research at convened meetings at which a majority of the IRB members are present, vote approval by a majority of members present at the meeting, and be responsible for reporting to the appropriate institutional official and the

secretary of DHHS "… any serious or continuing noncompliance by investigators with the requirements and determination of the IRB" (45 CFR 46.109(b)).

Institutions that are cooperating in multiinstitutional studies, such as clinical drug trials, must each review and approve the proposed studies. Such institutions may, however, use joint review, rely on the review of another qualified IRB, or utilize similar arrangements to avoid duplication of efforts (Cato, 1988; Fletcher et al., 2002).

***Informed Consent*** Assuring that adequate provisions exist for securing informed consent is a central duty of IRBs and that which is seemingly the most visualized when it fails (Office of Inspector General, 2000a,b). The requirements for informed consent are specified in international guidelines and the federal regulations. These require that investigators "shall seek such consent only under circumstances that provide the prospective subject … sufficient opportunity to consider whether or not to participate and that minimize the possibility of coercion or undue influence. The information that is given to the subject … shall be in language that is understandable to [him]" (45 CFR 46.116). The regulations further stipulate that "no informed consent, whether oral or written, may include any exculpatory language through which the subject … is made to waive or appear to waive any of the subject's legal rights, or releases or appears to be release the investigator, the sponsor, the institution or its agents from liability from negligence."

The federal regulations specify the information that shall be provided to each subject:

1. A statement that the study involves research, an explanation of the purposes of the research and the expected duration of the subject's participation, a description of the procedures to be followed, and identification of any procedures which are experimental
2. A description of any reasonably foreseeable risks or discomforts to the subjects
3. A description of any benefits to the subject or to others which may reasonably be expected from research
4. A disclosure of appropriate alternative procedures or courses of treatment, if any, that might be advantageous to the subject
5. A statement describing the extent, if any, to which confidentiality of records identifying the subject will be maintained
6. For research involving more than minimal risk, an explanation as to whether any compensation and an explanation as to whether medical treatments are available if injury occurs and, if so, what they consist of or where further information may be obtained
7. An explanation of whom to contact for answers to pertinent questions about the research and research subject's rights and whom to contact in the event of a research-related injury to the subject

8. A statement that participation is voluntary, refusal to participate will involve no penalty or loss of benefits to which the subject is otherwise entitled, and the subject may discontinue participation at any time without penalty or loss of benefits to which the subject is otherwise entitled

In addition to these basic elements of informed consent, IRBs also require that information be provided, where indicated, to the effect that (1) the particular treatment or procedure being tested may involve risks to the subject that are currently unforeseeable; (2) foreseeable circumstances may exist under which continued participation by the subject may be terminated by the investigator without regard to the subject's consent; (3) additional costs to the subject may result from participation in the research; (4) the consequences of a decision to withdraw; and (5) significant findings that may influence a subject's continued participation will be related to the subject.

In addition to the elements enumerated in the federal regulations, IRBs must consider whether consent forms should include the fact of randomization in the case of prospective randomized clinical trials.

Those who feel that the fact of randomization need not be disclosed to prospective subjects argue that since the alternative treatments to be tested are not known to produce significantly different results and since the physician would have to make an arbitrary selection of one treatment or the other for a particular patient, notification that selection of treatment is by computer rather than by the patient's own physician does not provide additional protection for the subjects and is unnecessary. The response to this contention is that a subject's ability to exercise full autonomy over what will be done with his or her own body is best served by notifying the subject as to how the treatment will be selected and by whom, even if the selection process is equally arbitrary whatever process is used.

The weight of the arguments favors the notion that for consent to be fully informed subjects must be notified that their treatments will be allocated in a random manner, that is, selected by a process other than the judgment of their own physician. The meaning of the concept of randomization and the fact that it will be the manner by which treatment is selected is therefore considered to be an important and integral part of informed consent for participation in randomized clinical trials.

Implicit in the elements that comprise informed consent for subjects participating in clinical trials is that subjects will be notified of the nature of their disease. Current bioethical thinking views this to be essential in order for patients/subjects to give legally effective informed consent. The current practice in the United States is that informed consent to participate in clinical trials requires that patients be notified of their diagnosis. Accordingly, a statement regarding the diagnosis is required in consent forms for participation in clinical trials that are sponsored by national cooperative groups. The Tuskegee study on syphilis is an excellent example of the results of not informing patients

(Jones, 1993). It is of interest that other Western countries do not feel that it is necessary or even appropriate to inform patients of their diagnosis as part of the consent process.

Increased incidents in clinical trials have led to recognition of the weaknesses of informed consent procedures. Actually, the FDA has reported that such deficiencies have been the poorest area of GCP compliance for more than 12 years (FDC Reports, 2000).

## 4 DRUG FORMULATIONS AND EXIPIENTS

It should never be lost sight of that one of the major reasons for the 1938 FD&C Act was a public health disaster caused by a drug formulation mistake. In the 1930s, the Massengill Company's use of diethylene glycol in an elixir of sulfanilamide led to 105 deaths. This same disaster was, by the way, repeated in Haiti in 1995 and 1996 (O'Brien et al., 1998). Such considerations are also overlooked in clinical safety evaluations, though the history of them directly and indirectly causing problems, even to the current day, is extensive (Winek, 2000).

Preclinical animal studies are usually performed with simple formulations which are appropriate for the route investigated in the (nonhuman) species involved. While similar simple formulations or approaches (such as capsules) are also employed for first-in-man studies, as development proceeds, efforts are made to develop formulations which optimize bioavailability. This may lead to effects not seen in earlier animal (or, indeed, human) studies—a factor that should be kept in mind in both study design and interpretation.

It is essential that formulations used in clinical trials should be well characterized, including information on bioavailability wherever feasible. The formulation should be appropriate for the stage of drug development. Ideally, the supply of a formulation will be adequate to allow testing in a series of studies that examine a range of doses. During drug development, different formulations of a drug may be tested. Links between formulations established by bioequivalence studies or other means are important in interpreting clinical study results across the development program.

Safety limitations on formulations usually arise from local tissue tolerance concerns at the site of administration for drugs other than oral.

## 5 PHASE I DESIGNS

Phase I clinical trials are the first studies in which a new drug is administered to human subjects. The primary purpose of phase I studies of new drugs is to establish a safe dose and schedule of administration (O'Grady and Linet, 1990). Other purposes are to determine the types of side effects and toxicity and organ systems involved, to assess evidence for efficacy, and to

investigate basic clinical pharmacology of the drug. Not all of these goals can be met completely in any phase I trial in part because the number of patients treated is small. However, well-conducted phase I studies can achieve substantial progress toward each of these goals. Phase I trials are not synonymous with dose–response studies, but they have many characteristics in common.

The initial phase of clinical testing has the following objectives:

(a) Establish a dose–response pharmacodynamic profile by using initial doses projected to be therapeutic in humans. The dose required is predicted on the basis of blood levels found in animal screens.
(b) Determine the pharmacokinetic profile for initial titration and maintenance of steady state for chronically administered drugs.
(c) Design a safe dosage regimen for efficacy testing in adults, the pediatric, or the elderly.
(d) Estimate efficacy information necessary to make sample size determinations for phase II studies and establish adequate duration of treatment.
(e) Determine the drug interaction potential when concurrent medications are administered as well as food interaction, assess the enzyme induction potential, and assess the need for therapeutic drug monitoring during efficacy testing.
(f) Establish the requirements for the final formulation.

The initial strategy for phase I is to conduct a single-dose safety study in normal volunteers. The first trial demands close 24-h supervision in a clinical setting. Ethical considerations may, however, demand that only patients be used—for example, when evaluating an anticancer agent with predictable toxicity. A repeat-dose tolerance and pharmacokinetic study in normal or patient volunteers is then conducted for chronically administered drugs. These studies will provide the necessary safety information to support efficacy testing.

Sometimes investigators say that phase I studies are not "clinical trials" because there is no treatment comparison being made (except that frequently a placebo is employed). Such treatment comparisons are not a prerequisite for experiments. Because Phase I trials rely on investigator-controlled treatment administration and subsequent structured observations, they are clinical trials.

In the development of cytotoxic drugs in oncology, dose finding usually means establishing a maximum tolerated dose (MTD). This is the dose associated with serious but reversible side effects in a sizable proportion of patients and the one that offers the best chance for a favorable therapeutic ratio. Side effects from cytotoxic drugs tend to be serious and are referred to as toxicities.

Investigators are interested not only in the organ systems involved but also in the duration, reversibility, and probability of specific toxicities. In this setting, evidence of efficacy is usually weak or nonexistent because many patients receive what turn out to be subtherapeutic doses of the drug.

For all phase I studies, learning about basic pharmacokinetics (clinical pharmacology) is important and includes measuring drug uptake, metabolism, distribution, and elimination. This information is vital to the future development and use of the drug and is helpful in determining the relationship between blood levels and side effects, if any. These goals indicate that the major areas of concern in designing phase I trials will be selection of patients, choosing a starting dose, rules for escalating doses, and methods for determining the MTD or safe dose.

If basic pharmacology were the only goal of a phase I study, the patients might be selected from any underlying disease and without regard to functioning of specific organ systems. However, phase I studies are usually targeted for patients with the specific condition under investigation. For example, in phase I cancer trials, patients are selected from those with a disease type targeted by the new drug. Because the potential risks and benefits of the drug are unknown, patients often are those with relatively advanced disease. It is usually helpful to enroll patients with a normal cardiac, hepatic, and renal function. Because bone marrow suppression is a common side effect of cytotoxic drugs, it is usually helpful to have normal hematological function as well when testing new drugs in cancer patients. In settings other than cancer, the first patients to receive a particular drug might have less extensive disease or even be healthy volunteers.

### 5.1 First Administration: Single Dose

First-time administration of single doses of new drugs is undertaken using a wide range of study design, but essentially there are several basic designs available which are modified to meet the needs of a given study. Fundamental to all designs is that in the interests of safety successive subjects are exposed to increasing doses of the drug. The fact that doses are titrated upward either in the same subject or in groups of subjects and not randomized to remove the potential for bias can be argued as a design weakness, but there is no alternative. Nevertheless, an ordered dose response can be taken as reasonable evidence of a drug-related effect. In addition, the use of placebo, which enables studies to be conducted on either a single- or a double-blind basis, will help to minimize bias. For this reason, placebo control is an integral part of a phase I study. Unwanted feelings or sensations are common occurrences in every-day life; hence it is to be expected that adverse events will be encountered during phase I studies. Adverse events may be drug related, study related, or result from something which has nothing to do with the drug or the study. They may act singly or in combination. For example, headache, which is one of the com-

monest if not the commonest symptom reported by volunteers taking part in phase I studies, can result from any one of the following: fasting, caffeine withdrawal, feeling anxious about the study, an impending attack of influenza, or a combination of all four factors. Thus without placebo control it becomes difficult to differentiate between headache which is drug related and headache which is not drug related. But a placebo is not only of value in helping to distinguish between drug- and non-drug-related subjective effects, it also plays a role in the interpretation of results from laboratory and other safety tests and pharmacological tests which may be influenced by diverse factors such as diet, physical activity, mental state, circadian or other biological rhythms, and asymptomatic illnesses, for example, subclinical viral infections.

The different designs available for a first-in-man study all have their own advantages and disadvantages (Fleiss, 1986; Spilker, 1991; Nylen, 2000; Gallin and Ognibene, 2007; Rallin and Ognibene, 2007). At the end of the day it is up to the investigator to weigh the pros and cons of each and then to choose the design which best meets the aims of the study. In an attempt to examine their strengths and weaknesses, let us consider some designs open to an investigator who wishes to undertake a single rising-dose safety and tolerability study with a new drug. A typical protocol might require:

- Placebo control
- The dose be increased from $x$ (first dose) to $64x$ (top dose)
- Twofold increases in successive doses (within or between subjects)
- A seven-day within-subject washout period
- A minimum of four subjects to receive each dose level

Some design options are shown in Tables 4–8, while the implications for going with one or the other in terms of subject numbers, number of clinic visits, highest first dose given to a subject, biggest increment in dose, and time to complete the study are given in Table 9.

**TABLE 4 Phase I Study: Type A**

| | Number of Subjects Who Received Each Treatment | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Group | Placebo | $x$ | $2x$ | $4x$ | $8x$ | $16x$ | $32x$ | $64x$ |
| 1 | 2 | 3 | | | | | | |
| 2 | 2 | | 3 | | | | | |
| 3 | 2 | | | 3 | | | | |
| 4 | 2 | | | | 3 | | | |
| 5 | 2 | | | | | 3 | | |
| 6 | 2 | | | | | | 3 | |
| 7 | 2 | | | | | | | 3 |

*Note*: $x$ = first dose.

**TABLE 5 Phase I Study: Type B**

| | Number of Subjects Who Received Each Treatment | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Group | Placebo | *x* | 2*x* | 4*x* | 8*x* | 16*x* | 32*x* | 64*x* |
| 1 | 2 | 3 | 1 | | | | | |
| 2 | 2 | | 2 | 1 | | | | |
| 3 | 2 | | | 2 | 1 | | | |
| 4 | 2 | | | | 2 | 1 | | |
| 5 | 2 | | | | | 2 | 1 | |
| 6 | 2 | | | | | | 2 | 1 |
| 7 | 2 | | | | | | | 2 |

*Note*: *x* = first dose.

**TABLE 6 Phase I Study: Type C**

| No. Visits[a] | Volunteer No. | Treatment | | | |
|---|---|---|---|---|---|
| 2 | 1–4 | *x* | 2*x* | 4*x* | (P) |
| 2 | 5–8 | 4*x* | 8*x* | 16*x* | (P) |
| 2 | 9–12 | 16*x* | 32*x* | 64*x* | (P) |

*Note*: *x* = first dose; (P) = randomized placebo.
[a]Volunteers receive three doses of drug on one visit and placebo on the other visit.

**TABLE 7 Phase I Study: Type D**

| No. Visits[a] | Volunteer No. | Treatment | | | |
|---|---|---|---|---|---|
| 4 | 1–4 | *x* | 2*x* | 4*x* | (P) |
| 4 | 5–8 | 4*x* | 8*x* | 16*x* | (P) |
| 4 | 9–12 | 16*x* | 32*x* | 64*x* | (P) |

*Note*: *x* = first dose; (P) = randomized placebo.
[a]Volunteers receive single dose of drug on each of three visits and placebo on one visit.

**TABLE 8 Phase I Study: Type E**

| Visit | Volunteer No. | | Volunteer No. | | Volunteer No. | |
|---|---|---|---|---|---|---|
| 1 | 1 | P | 5 | *x* | 9 | 2*x* |
| | 2 | *x* | 6 | P | 10 | 2*x* |
| | 3 | *x* | 7 | 2*x* | 11 | P |
| | 4 | *x* | 8 | 2*x* | 12 | 4*x* |
| 2 | 1 | 4*x* | 5 | P | 9 | 4*x* |
| | 2 | 4*x* | 6 | 4*x* | 10 | P |
| | 3 | 4*x* | 7 | 4*x* | 11 | 8*x* |
| | 4 | P | 8 | 4*x* | 12 | 8*x* |
| 3 | 1 | 8*x* | 5 | 16*x* | 9 | P |
| | 2 | 8*x* | 6 | 16*x* | 10 | 16*x* |
| | 3 | P | 7 | 16*x* | 11 | 16*x* |
| | 4 | 16*x* | 8 | P | 12 | 16*x* |
| 4 | 1 | 16*x* | 5 | 32*x* | 9 | 64*x* |
| | 2 | P | 6 | 32*x* | 10 | 64*x* |
| | 3 | 32*x* | 7 | P | 11 | 64*x* |
| | 4 | 32*x* | 8 | 64*x* | 12 | P |

*Note*: *x* = first dose; P = placebo.

**TABLE 9 Comparisons of Study Types**

| | A | B | C | D | E |
|---|---|---|---|---|---|
| No. of volunteers | 12 | 12 | 12 | 42 | 42 |
| No. of clinic visits | 4 | 4 | 2 | 1 | 1 |
| Highest first dose to subject | 16*x* | 8*x* | 16*x* | 64*x* | 64*x* |
| Largest within-subject increment in dose | ×2 | ×16 | ×2 | — | — |
| Times to do study | 9 weeks | 4 weeks | 3 weeks | 3–6 weeks[a] | 3–6 weeks[a] |

[a]Dependent upon whether dosing takes place once or twice weekly.

Design types A and B require more than threefold the number of volunteers needed for the other designs but also give the clearest picture of the pharmacokinetics and tolerance of a single dose. This requirement is compounded by the fact that one-third of the volunteers will receive placebo. The number of volunteers on placebo in each group is open to the investigator's choice, but a balanced (even number of each) design again is easiest to interpret. Whichever way, types A and B require large numbers of subjects, which could present problems when recruiting suitable subjects. The situation is made more difficult when numbers on the volunteer panel are limited (which often is the case) and when one is attempting to recruit the best available volunteers who also satisfy the inclusion/exclusion criteria for a first-in-man study. It can also be argued that if the drug under test proved to be toxic, then more subjects would be exposed to its harmful effects. On the other hand, if the drug turns out to be well tolerated, it can be argued equally well that exposing a larger number of subjects is a better basis on which to proceed to the next study.

Types A and B, however, have two clear advantages over the other designs. First, as only one visit to the clinic is required, this will encourage the volunteer to take part in and complete the study. Second, they are ideal designs for drugs with long (or unclear) pharmacological, clinical, or chemical half-lives when a seven-day washout period is an inadequate time for the drug effects to disappear or for it to be cleared from the body.

In the interests of safety, the lower the dose the volunteer is given on the first exposure to the drug, the better. However, as it is impractical to start everyone off with dose $x$, the next best thing one can do is to keep the first dose given to a volunteer in each group as low as possible within the confines of the design of the study. In this respect, type E works best and types A and B do badly.

With types D and C, twofold increments in dose are uniformly made throughout the whole dose range. This is in contrast to type E, in which the size of dosage increments over the dose range within subjects varies between 2- and 16-fold. Thus type E might be an unwise choice for a drug anticipated to have a narrow therapeutic index or a steep dose response.

Assuming the study goes according to plan (which is often not the case in first-administration studies) and depending upon the study design used, it will take between three and nine weeks to complete. However, although type C (in which the dose is increased stepwise on the same study day) offers the advantage of speed, the fact that it can only really be used for drugs given by the intravenous route and for drugs with rapid onsets and offsets of action limits its usefulness in practice.

### 5.2 First Administration in Humans: Repeat Dose

In clinical practice, drugs are often prescribed for illnesses which require regular treatment for days, weeks, months, or years. For drugs used in this way (or for which off-label use is likely to be this way), testing on a repeat-dose basis in volunteers is required to evaluate safety and tolerability before treating patients. As with first-time single-dose studies, first-time repeat-dose studies can be undertaken using different designs but with the emphasis again on safety and tolerability. The cornerstone design is a randomized, rising-dose, placebo-controlled group comparative evaluation. Whichever design is used, the investigator has to decide upon an appropriate dosing schedule. The choice of a unit dose and dosing interval depends primarily upon the results from the single-dose study. To illustrate this point, if one assumes that the top dose (i.e., 64*x*) given in the previously described single-dose study proved to be well tolerated, then one might opt for the dosing schedule given in Table 10. Of course, the frequency of dosing will depend upon the pharmacodynamic and/or pharmacokinetic profile of the drug. Ideally, dosing should be continued until steady-state plasma concentrations of drug have been achieved, but this may not be practical for drugs with long half-lives. More often than not volunteers are dosed for 7–10 days, but in certain circumstances if toxicological clearance is available and there is a definite need to do so, volunteers may be dosed for 4 weeks. Even if the intent is to dose more than once daily [as in Table 10 where twice-daily (bid) dosing is required], giving single doses on the mornings of day 1 and the last day of dosing (i.e., day 7) offers certain advantages. For example, it allows for a longer period to assess tolerability before the second dose of drug is given to a volunteer who more than likely will not have been exposed to the drug previously. It also enables comparisons to be

**TABLE 10 Design and Dosing Schedule for First Repeat-Dose Phase I Study**

| Group | Day 1 | Day 2–6 | Day 7 |
|---|---|---|---|
| 1 | 8*x* | 8*x* bid | 8*x* |
| 2 | 16*x* | 16*x* bid | 16*x* |
| 3 | 32*x* | 32*x* bid | 32*x* |

*Note*: bid = twice daily.

made between drug plasma concentration–time profiles over 24 h and the elimination kinetics of the drug at the start and end of dosing.

In the interests of safety, doses are increased between groups sequentially and as a rule dosing is completed in the previous group before dosing is started in the next group. However, if groups are to be dosed for more than 7–10 days or a large number of increments in dose is planned, particularly if more than one dosing frequency is under test, the investigator might choose to overlap dosing between one group and the next, thus enabling the study to be completed in a reasonable time frame. Within each group, volunteers are randomly allocated to receive drug or placebo. The size of the groups usually varies between 6 and 12 with the numbers of subjects receiving drug and placebo in a group being subject to investigator preference.

***Number of Subjects*** In a phase I trial, a sufficient number of subjects must be included in a study if valid conclusions are to be drawn from the results. Studies in healthy volunteers and patients are inherently flawed when it comes to assessing safety and tolerability because of the small numbers of subjects involved, and only the most guarded of conclusions are possible. It is easier to draw valid conclusions with respect to drug action involving pharmacodynamic, surrogate, or clinical endpoints because one is able to specify before hand the magnitude of the difference which constitutes a useful drug effect and thus calculate the numbers of studies, except that regulatory authorities rather than the investigator specify the criteria which have to be met to enable different formulations to be judged bioequivalent.

### 5.3 Route of Administration

Just as new drugs must be tested in animals by the route to be used in humans, so must they be tested in volunteers using the intended route for patients. But there are clear benefits in testing all drugs when going into humans for the first time using intravenous infusions, even if systemic exposure in patients will be achieved by another route. These benefits relate primarily to the fact that intravenous infusion allows for precise control of drug administration:

- In the event of a serious or otherwise distressing adverse event during the infusion, drug delivery can be halted.
- As the drug is delivered directly into the bloodstream, this ensures 100% exposure and overcomes problems relating to bioavailability which may occur with other routes but in particular dosing by the oral route when the drug may be destroyed in the gastrointestinal (GI) tract or metabolized presystemically in the gut wall or in the liver.
- Delivery of the full dose into the bloodstream coupled with a uniform delivery rate results in less variability in plasma or tissue concentrations of drug than is possible using oral dosing, where not only the extent but also the rate of absorption from the GI tract can vary considerably

between subjects. Less intersubject variability in plasma concentrations of drug in turn enables the study to be done using smaller numbers of subjects and also offers advantages for drugs it is anticipated might have narrow therapeutic ratios.

- Intravenous dosing allows the true disposition kinetics of the drug to be evaluated and makes the assessment of PK/PD relationships easier to perform.
- Pharmacokinetic scaling between species, that is, animals to humans, is made simpler as fewer assumptions need to be made about extent and rate of exposure in humans. This in turn helps in further dose selection for human studies.
- Blinding of studies is made easier when intravenous dosing is used, that is, there is no need to produce matching placebos while intravenous dosing overcomes any problems relating to taste, which can make it difficult in blind studies involving oral dosing.

The primary disadvantage of using intravenous dosing for first-in-human studies is that additional resource will be needed to be spent in toxicology, establishing dosage form stability, and mutation development on a drug which might fail at the first hurdle in humans, as indeed many do. For this reason investigators often prefer to administer drugs for the first time in healthy volunteers using the route to be used in patients and dose intravenously to establish the drugs' pharmacokinetic profile only when they feel reasonably certain that it is likely to be a candidate for further development.

## 6 CLINICAL TRIAL SAFETY INDICATORS

One major purpose of preclinical (animal) toxicity studies of a potential new drug is to identify the toxic effects which most commonly occur at doses nearest to those to be used in humans. These observations serve to help ensure that care is taken to detect any such effects in humans. Additionally, a broad range of other indicators of adverse drug action may be identified to ensure that their occurrence is looked for. These are also commonly called safety parameters.

Because of the relatively small numbers of volunteers and patients involved, only the most common of drug-related adverse events are likely to be detected during early studies (O'Grady and Joubert, 1997). For example, to have a 95% chance of picking up three subjects who have experienced an adverse reaction (with no background incidence) which occurs in 1 in every 100 subjects treated with the drug, it would need to be given to 650 subjects. Matters are made worse when the adverse event in question also occurs in the general population, which is usually the case with the kind of symptoms reported by volunteers and patients taking part in drug studies. No matter how good the

study design is, nothing can compensate for this problem of inadequate numbers. In this respect, all of the study designs described earlier are more or less equally adequate or inadequate as the case may be.

Monitoring for drug-related adverse events employs the same or similar methods in both volunteers and patients. In both cases assessments of tolerability and safety are based upon symptom reports, routine laboratory safety screens, electrocardiogram (ECG) monitoring, and on occasion special tests designed to detect unwanted effects associated with a particular class of drug. The chances of obtaining reliable information on a drug's safety profile are enhanced by detailed and careful monitoring, such as special biomarkers for cardiotoxicity (Braunwald, 2008). Symptoms may be reported spontaneously or elicited in reply to standard questions. Open questions such as "how are you feeling?" are to be preferred to leading questions on the basis that they result in fewer reports of adverse events. If leading questions are used, they need to be carefully worded. A certain amount of basic information is required on all adverse events, that is, type, severity, time of onset in relation to time of dosing, duration, and causality. Attributing the cause of an unwanted effect to the drug or some other factor can be difficult, particularly when little is known about the drug, as is often the case at the state of initial studies in volunteers or patients. Rechallenge with the drug ideally using the same dose or, if need be (because the event caused a degree of discomfort), a reduced dose is probably the single best way of proving or disproving a causal relationship. But if done the rechallenge procedure must be designed using placebo as the comparator under double-blind conditions. Obviously rechallenges can be done only if the adverse event was reversible, did not cause excessive discomfort, and most importantly was not life threatening. The question of assessing attributions or causality is considered in detail later in this chapter.

## 6.1 Overall Approach to Assessing Safety

***Choosing Safety Parameters*** Choosing the appropriate safety parameters for a clinical trial depends on a number of factors. A selected list of examinations and tests commonly used to assess the safety of medicines is given in Table 11. The majority of these tests will not be conducted in most drug trials. An assessment of the quantity and quality of prior experience and previous data obtained with the therapeutic is essential to enable one to decide which specific safety tests to incorporate in a medicine trials. The choice of safety parameters requires both data in areas where there are indications of potential (or actual) safety problems to monitor and additional experience and data with a new drug. Until a sufficient body of safety data has accumulated, more laboratory parameters of safety are generally included than will be needed at a later date. The nature of the clinical trials and efficacy tests used may dictate that certain safety parameters should or should not be included (e.g., in testing a new anticancer medicine, it may be necessary to perform a bone marrow biopsy and smear to confirm the lack of toxicity, and

**TABLE 11 Selected List of Examinations and Tests Used to Evaluate Safety**

A. Clinical examinations
   1. Physical
   2. Vital signs (usually considered part of physical examination)
   3. Height and weight (state of dress is usually specified, e.g., socks)
   4. Neurological or other specialized clinical examinations

B. Clinical Laboratory examinations
   1. Hematology (see Table 15)
   2. Clinical chemistry (see Table 14)
   3. Urinalysis (see Table 15)
   4. Virology (viral cultures or viral serology)
   5. Immunology or immunochemistry (e.g., immunoglobulins, complement)
   6. Serology
   7. Microbiology (including bacteriology and mycology)
   8. Parasitology (e.g., stool for ova and protozoa)
   9. Pulmonary function tests (e.g., arterial blood gas)
   10. Other biological tests (e.g., endocrine, toxicology screen)
   11. Stool for occult blood (specify hemoccult or guaiac method)
   12. Skin tests for immunological competence
   13. Medicine screen (usually in urine) for detection of illegal or non-protocol-approved medicines
   14. Bone marrow examination
   15. Gonadal function (e.g., sperm count, sperm motility)
   16. Genetic studies (e.g., evaluate chromosomal integrity)
   17. Stool analysis using in vivo dialysis

C. Probe for adverse reactions

D. Psychological and psychiatric tests and examinations
   1. Psychometric and performance examinations
   2. Behavioral rating scales
   3. Dependence liability

E. Examinations requiring specialized equipment (selected examples)
   1. Audiometry
   2. Electrocardiogram (ECG)
   3. Electroencephalogram (EEG)
   4. Electromyography (EMG)
   5. Stress test
   6. Endoscopy
   7. Computed tomography (CT) scans
   8. Ophthalmological examination
   9. Ultrasound
   10. X-rays
   11. Others

in assessing an agent in anesthetized patients, the appropriate tests to ensure the patient's safety while under anesthesia must be performed). If, on the basis of preclinical pharmacological or toxicological data, any toxicity is either anticipated or considered possible, then an attempt should be made to evaluate patients for those possible problems. The anticipated use(s) of a therapeutic will also influence which safety parameters are chosen for evaluation (e.g., ophthalmological tests would be included for drugs intended for ocular use).

***Measuring Safety Parameters*** After specific safety parameters are chosen, it is necessary to determine how thorough an evaluation of each parameter should be conducted. It is also possible that different types of examinations would be suitable at different points of a clinical trial. For example, a physical examination may be specified to include more or fewer measurements or facets, and a complete examination may not be necessary or even suitable during some periods of clinical trial.

Vital signs may be measured with the patient in a supine, seated, and/or erect position. Both supine and erect positions are usually used if orthostatic changes are being evaluated. The need for such data will depend on the situation, but the position of the patients for this examination, as well as the period of time desired for stabilization, should be noted in the protocol.

***Parameters That Measure Either Safety or Efficacy*** Certain parameters may, of course, be either safety or efficacy parameters or both. The electroencephalogram (EEG) is an example. Blood pressure is another. It is thus important to clearly establish in the protocol whether each parameter is begin incorporated in the protocol for safety or efficacy evaluation. Almost any safety parameter can be used for measuring efficacy.

***Appropriateness of Each Parameter for Clinical Trial and Patient*** There are four categories of appropriateness of safety tests used in clinical trials:

1. Appropriate for patients but not necessary for the clinical trial. All of these tests should be included in the clinical trial. They indirectly benefit the trial because they may be monitored for progress or trends or they may simply ensure that patients are receiving appropriate care.
2. Appropriate for the clinical trial but not necessary for the patients. These tests should be included in the clinical trial if they do not place the patient at unacceptable risk or discomfort. If any tests are deemed unethical in the context of the trials and the patients enrolled, then they should be excluded.
3. Appropriate for both patients and the clinical trial. All of these tests should be included in the clinical trial.
4. Appropriate for neither patients nor the clinical trial. All of these tests should be identified and excluded from the clinical trial.

## 6.2 Precautions

Clinical laboratory parameters must be specified individually in the protocol. Abbreviations such as SMA-6 or SMA-12 are not acceptable, as different laboratories include different tests in their SMA-6 (or SMA-12) battery, and using these abbreviations without an explanation can adversely affect the clarity of the protocol and possibly lead to the collection of data on divergent

parameters at different sites. Other precautions to consider prior to initiating a clinical trial are to decide if (1) severely abnormal results should be routinely confirmed, (2) samples should be divided and sent to two separate laboratories when specified abnormalities are determined, (3) additional tests should be routinely requested if specified abnormalities are observed, (4) medical consultants should examine patients whenever severe abnormalities are observed, and (5) aliquots of known concentrations of standard drugs should be sent to laboratories for confirmatory measurements and interlaboratory evaluation.

***Summary of Tests*** Common dermatological tests are shown in Table 12 and ophthalmological tests in Table 13. Note that any of these tests could be utilized as measures of efficacy if they addressed the clinical trial objectives. Selected pointers are given in Table 14. Specific tests that may be used in hematology, clinical chemistry, and urinalysis are shown in Table 15, adult and pediatric behavioral rating scales in Tables 16 and 17, and psychometric and performance tests in Table 18.

***Choosing Laboratory Tests*** There is no standardized series of laboratory parameters that are evaluated in all clinical trials, nor is there a single standard for drugs in phases I, II, or III. There are, however, broad general guidelines for laboratory tests that are performed at each stage of clinical development.

***Tests in Phase I*** In Phase I clinical trials, there is the greatest need to obtain a wide variety of laboratory evaluations as part of developing the safety profile on a new medicine. This entails an evaluation of the basic hematology, clinical chemistry, and urinalysis parameters (Table 15). There will never be 100% agreement among investigators and/or clinical scientists as to which specific tests constitute a "basic" workup.

**TABLE 12 Deleterious Effects of Biomarkers of Inflammation in Heart Failure**

| |
|---|
| Known |
| Left ventricular dysfunction |
| Pulmonary edema |
| Cardiomyophathy |
| Decreased skeletal-muscle blood flow |
| Endothelial dysfunction |
| Anorexia and cachexia |
| Potential[a] |
| Receptor uncoupling from adenylate cyclase |
| Activation of the fetal-gene program |
| Apoptosis of cardiac myocytes |

[a]Effects shown in animals but not yet in humans.

*Source*: Adapted from Mann.

**TABLE 13 Biomarkers in Heart Failure**

| |
|---|
| Inflammation[a,b,c] |
| C-reactive protein |
| Tumor necrosis factor α |
| Fas (APO-1) |
| Interleukins 1, 6, and 18 |
| Oxidative stress[a,b,d] |
| Oxidized low-density lipoproteins |
| Myeloperoxidase |
| Urinary biopyrrins |
| Urinary and plasma isoprostanes |
| Plasma malondialdehyde |
| Extracellular matrix remodeling |
| Matrix metalloproteinases |
| Tissue inhibitors of metalloproteinases |
| Collagen propeptides |
| Propeptide procollagen type I |
| Plasma procollagen type III |
| Neurohormones[a,b,d] |
| Norepinephrine |
| Renin |
| Angiotensin II |
| Aldosterone |
| Arginine vasopressin |
| Endothelin |
| Myocyte injury[a,b,d] |
| Cardiac-specific troponins I and T |
| Myosin light-chain kinase I |
| Heart-type fatty acid protein |
| Creatin kinase MB fraction |
| Myocyte stress[b,c,d,e] |
| Brain natriuretic peptide |
| N-terminal probrain natriuretic peptide |
| Midregional fragment of proadrenomedullin |
| ST2 |
| New biomarkers[b] |
| Chromogranin |
| Galectin 3 |
| Osteoprotegerin |
| Adiponectin |
| Growth differentiation factor 15 |

[a]Biomarkers in this category aid in elucidating the parthenogenesis of heart failure.
[b]Biomarkers in this category provide prognostic information and enhance risk stratification.
[c]Biomarkers in this category can be used to identify subjects at risk for heart failure.
[d]Biomarkers in this category are potential targets of therapy.
[e]Biomarkers in this category are useful in the diagnosis of heart failure and in monitoring therapy.

*Source*: Adapted from Braunwald, 2008.

**TABLE 14 Selected Considerations Pertaining to Laboratory Data**

1. Ask the laboratory to maintain assayed samples that are of particular importance; if questions arise as to the accuracy of results, it might be possible to retest the original samples.
2. If laboratory problems are anticipated, divide the initial (and subsequent) samples and send them to two different laboratories or to the same laboratory at two different times.
3. If laboratory samples for a complete blood count are going to remain unexamined for a long period of time (e.g., sample obtained on Sunday), prepare a fresh smear so that a comparison may be made with one made 24 or more hours later because abnormalities may occur when a sample lies around even when it is kept at an appropriate temperature.

**TABLE 15 Hematology, Clinical Chemistry, and Urinalysis Parameters Usually Evaluated During Development of New Therapeutic Agent**

A. Hematology
   1. Red blood cell (RBC) count
   2. Hemoglobin
   3. Hematocrit
   4. White blood cell (WBC) count and differential
   5. Platelet estimate or platelet count
   6. Red blood cell indices (MCV, MCH, MCHC)[a]
   7. Prothrombin (PT) and partial thromboplastin (PTT) times
   8. Reticulocytes
   9. Fibrinogen
   10. Any additional tests suggested by previous data

B. Clinical chemistry
   1. Albumin
   2. Albumin/globulin ratio
   3. Alkaline phosphatase (and/or its isoenzymes)
   4. Amylase
   5. Bilirubin, total and direct
   6. Bicarbonate (carbon dioxide)
   7. BUN/creatinine ratio
   8. Calcium
   9. Chloride
   10. Cholesterol (and/or a lipid panel)
   11. Creatinine
   12. Creatine phosphokinase (CPK)
   13. γ-Glutamyl transferase (GGT)
   14. Globulin
   15. Glucose, nonfasting or fasting
   16. Glucose-6-phosphate dehydrogenase (G6PD)
   17. Glutamate oxaloacetic transaminase (SGOT), now frequently referred to as aspartate aminotransferase (AST)
   18. Glutamate pyruvate transaminase (SGPT), now frequently referred to as alanine aminotransferase (ALT)
   19. Iron (and/or other related parameters such as ferritin, total iron-binding capacity)
   20. Lactic acid dehydrogenase, total (LDH, and/or its isoenzymes)
   21. Inorganic phosphorus
   22. Potassium
   23. Sodium
   24. Total iron-binding capacity
   25. Total protein
   26. Triglycerides
   27. Urea nitrogen (BUN)
   28. Uric acid

**TABLE 15** ***Continued***

C. Hormones and/or other chemical substances in blood
D. Urinalysis[b]
   1. Appearance and color
   2. Specific gravity
   3. Acetone
   4. Protein
   5. Glucose
   6. PH
   7. Bile
   8. Irobilinogen
   9. Occult blood
   10. Microscopic evaluation of sediment
      a. Red blood cells (number per high-power field)
      b. White blood cells (number per high-power field)
      c. Casts (describe and give number per high- or low-power field)
      d. Crystals (describe and give number per high-power field)
      e. Bacteria (generally rated as few, many, or loaded)
      f. Epithelial cells (number per low-power field)
E. Other urine tests sometimes evaluated
   1. Creatinine (actual values are preferable to estimated values)
   2. Electrolytes (usually sodium, potassium, and chloride)
   3. Protein
   4. Specific hormones or chemicals
   5. 24-h collections for specific evaluations

[a]MCH, mean corpuscular hemoglobin = hemoglobin divided by RBC count; MCHC, mean corpuscular hemoglobin concentration = hemoglobin divided by hematocrit; MCV, mean corpuscular volume = hematocrit divided by RCB count.
[b]Sample codes used to quantify several parameters in the urinalysis are the following. Protein, glucose, ketones, bilirubin: 0, none or negative; 0.5, trace or positive (qualitative); 1, + or 1+; 2, ++ or 2+; 3, +++ or 3+; 4, ++++ or 4+. Epithelial cells, crystal, WBC, RBC, casts: 0, none or negative; 0.5, rare, occasional, few present, trace (1–5); 1, several, mild (6–10); 2, moderate (11–25); 3, many, much (26–50); 4, loaded, severe (>50). Bacteria: 0, none or negative; 0.5, rare, trace, occasional, few several (1–10); 1, mild (11–50); 2, moderate (51–75); 3, many, numerous (76–100); 4, loaded, severe (>100).

**TABLE 16 Procedures and Tests Performed in Ophthalmological Examination**

1. Ophthalmological history (attention is paid to patient family history plus patient's diseases and drug reactions
2. Visual acuity corrected (i.e., with glasses present)
3. External ocular examination (i.e., check for inflammation, ptosis, nystagmus, tearing, proptosis, and other abnormalities)
4. Extraocular muscle testing
5. Pupil size and evaluation (in darkened room with controlled illumination)
6. Slit-lamp biomicroscopy (with dilated pupils)
7. Tonometry (ocular pressure)
8. Ophthalmoscopy with fundus photographs
9. Visual field testing and color vision testing
10. Goniscopy[a]
11. Lacrimation[a] (Schirmer test)

[a]These tests are of minimal value in determining ocular toxicity and are not recommended for routine use in ophthalmological examination to detect drug toxicities.

**TABLE 17 Selected Examples of Safety Measurements and Tests for Specialized Dermatological Examination**

1. Biopsy
2. Erythema at site of lesion
3. Absorption of medications systemically (e.g., blood levels)
4. Signs and symptoms of absorption
5. Interactions with standard treatment (e.g., ultraviolet light)

**TABLE 18 Adult Behavioral Rating Scales**

| | Scale Rated By | | |
|---|---|---|---|
| Scale | Professional | | Subject |
| 1. Anxiety Status Inventory (ASI) | X | | |
| 2. Beck Depression Inventory (Beck) | | | X |
| 3. Brief Psychiatric Rating Scale (BPRS) | X | | |
| 4. Carroll Depression Scale | | | X |
| 5. Clinical Global Impression (CGI) | X | or | X |
| 6. Clyde Mood Scale | X | | |
| 7. Covi Anxiety Scale | X | | |
| 8. Crichton Geriatric Rating Scale | X | | |
| 9. Depression Status Inventory | X | | |
| 10. Hamilton Anxiety Scale (HAMA) | X | | |
| 11. Hamilton Depression Scale (HAMD) | X | | |
| 12. Hopkin Symptom Checklist (HSCL) | | | X |
| 13. Inpatient Multidimensional Psychiatric Scale (IMPS) | X | | |
| 14. Nurses Observation Scale for Inpatient Evaluation (NOSIE) | X | | |
| 15. Plutchik Geriatric Rating Scale (PLUT) | X | | |
| 16. Profile of Mood States (POMS) | | | X |
| 17. Sandoz Clinical Assessment—Geriatric | X | | |
| 18. Self-Report Symptom Inventory (SCL-90) | | | X |
| 19. Wittenborn Psychiatric Rating Scale (WITT) | X | | |
| 20. Zung Self-Rating Anxiety Scale (SAS) | | | X |
| 21. Zung Self-Rating Depression Scale (SDS) | | | X |

*Note*: Standard abbreviations are used (see Guy, 1976). Additional tests are described in Buros (1978).

***Tests in Later Phases*** The total number of normal laboratory values that is sufficient to collect on a new drug to demonstrate safety is impossible to specify. Numerous factors must be considered, such as the toxicological profile on other safety parameters and the expected use of the drug in patients. It is important to determine if a therapeutic agent is to be used topically or parenterally, whether it is to be used in generally healthy patients or in seriously ill patients, whether it is a "me-too" drug or a totally novel drug chemically, and whether it will be life saving or provide a minimal therapeutic effect. The number of laboratory tests performed usually decreases as an investigational drug moves closer to the market, but one or more tests may be added to the list in Table 15 and studied in great detail.

**TABLE 19 Pediatric Behavioral Rating and Diagnostic Scales**

1. Children's Behavior Inventory (CBI)
2. Children's Diagnostic Classification (CDC)
3. Children's Diagnostic Scale (CDS)
4. Children's Psychiatric Rating Scale (CPRS)
5. Clinical Global Impression (CGI)
6. Conners Parent Questionnaire (PO)
7. Conners Parent–Teacher Questionnaire (PTO)
8. Conners Teacher Questionnaire (TO)
9. Devereux Child Behavior Rating Scale
10. Devereux Elementary School Behavior Rating Scale
11. Dosage Record and Treatment Emergent Symptoms (DOTES)
12. Stereotyped Behavior in Retarded

**TABLE 20 Psychometric and Performance Tests**

| | For Use In | |
|---|---|---|
| Test | Adults | Children |
| 1. Bender–Gestalt Test | X | X |
| 2. Conceptual Clustering Memory Test | X | X |
| 3. Digital Symbol Substitution Test | X | X |
| 4. Embedded Figures Test | X | X |
| 5. Frostig Development Test of Visual Perception | | X |
| 6. Goodenough–Harris Figure-Drawing Test (GOOD) | | X |
| 7. Peabody Picture Vocabulary Test | | X |
| 8. Porteus Mazes | X | X |
| 9. Reaction Time | X | X |
| 10. Vigilance Tests | X | X |
| 11. Wechsler Adult Intelligence Scale (WAIS) | X | |
| 12. Wechsler Intelligence Scale for Children (WISC) | | X |
| 13. Wechsler Memory Scale (WMEN) | X | X |
| 14. Wide Range Achievement Test (WRAT) | | X |

*Note*: Additional tests are described in Buros (1978).

***Tests in Medical Practice*** The ordering of laboratory tests in medical practice (as opposed to phase I clinical trials) is extremely inefficient and often irrational. This suggests the need in some clinical situations to develop logical protocols and algorithms for physicians to follow in ordering tests, particularly when the technology is changing (e.g., hepatitis), in therapeutic areas in which an excessive number of tests are often ordered (e.g., thyroid tests), or when hospitals have developed their own approaches to diagnosis (e.g., use of cardiac isoenzymes in diagnosing a myocardial infarction).

***Less Commonly Used Methods*** Evaluations of virtually any biological fluid, tissue, or sense (taste, smell, hearing, sight, and touch) can be conducted to ascertain the safety of a drug (several have been reported to affect taste in some patients, and there are many other examples involving medicine-induced

effects on one of the other senses). The choice of tests will depend on experience with the medicine and suspicions about possible problems. Drugs should also be reviewed for teratogenic potential, drug dependence, liability, and carcinogenicity.

***Identifying Most Important Laboratory Analytes to Monitor in Clinical Trial*** A choice often must be made among the numerous laboratory analytes that could be measured in a clinical trial. This choice is based on (1) past experience with the treatment(s) being evaluated, (2) therapeutic claim, (3) cost of the tests, (4) convenience of obtaining samples, (5) resources available, (6) state-of-the-art concept of the data's importance, and (7) the ability of data obtained to convince both regulators and medical practitioners. Arriving at a decision given these and other previously discussed factors may be difficult.

***Uses of Specific Laboratory Tests to Discover, Confirm, and/or Exclude a Disease*** Some tests can confirm the diagnosis of a disease (e.g., tissue histology from a bronchoscopic biopsy to confirm lung cancer) but cannot be used to exclude the disease or discover the disease in routine screening. Other tests can be used both to confirm and to exclude the diagnosis of a disease (e.g., glucose tolerance test for diabetes mellitus) but are too inconvenient to be used to discover the disease in routine screening. The uses of each laboratory test to discover, confirm, or exclude a disease should be considered before a test is simply added to a clinical trial protocol. This ensures that the test is appropriate in the context of the planned clinical trial.

***Hematology*** A basis hematology evaluation usually includes determination of hemoglobin, hematocrit, red blood cell (RBC) count, white blood cell (WBC) count, RBC indices [mean corpuscular hemoglobin (MCH), mean corpuscular hemoglobin concentration (MCHC), and mean corpuscular volume (MCV)], and platelet count. The WBC differential count is usually not required as part of a basic hematological workup unless a specific parameter of the differential count is being evaluated. Nonetheless, a WBC differential count is often obtained in phase I and generally provides useful (though often negative) information. Other hematological parameters (some of which are indicated in Table 15) are not usually obtained unless there is a specific reason to do so.

### 6.3 Clinical Chemistry

A measurement of renal function [creatinine and/or blood urea nitrogen (BUN)] is an "essential" test for most clinical studies, as is the inclusion of a panel of liver function tests [serum glutamic–oxaloacetic acid transaminase (SGOT), serum glutamic pyruvic transaminase (SGPT), lactic dehydrogenase (LDH), creatine phosphokinase (CPK), gamma-glutamyl transferase (GGT),

and/or alkaline phosphatase]. The specific tests chosen to be included in a study are somewhat dependent on both the investigator's and/or clinical scientist's experiences and the characteristics of the drug. Other important parameters to measure include serum electrolytes and at least some of the tests listed in Table 15.

***Drug Levels in Plasma*** Drug levels may also be measured in a clinical trial. Such levels are usually part of a pharmacokinetic analysis but also provide important safety data. This information would be particularly relevant in cases of suspected or actual drug overdosage or drug interactions, to correlate medicine levels with toxic events, or in other situations. It must be clarified whether free levels of the drug and/or the protein bound will be measured by the laboratory.

***Total Blood That May be Taken from Patients*** The total amount of blood that may be taken from a subject in most therapeutic trials should be limited to one unit (about 460 mL) per eight-week period.

### 6.4 Urinalysis

Most clinical laboratories have established a standard battery of tests that includes most or all of the basic parameters listed in Table 15. If a dipstick is used to test the urine for several parameters, it is useful to use one that measures occult blood, even if a microscopic examination will count the number of RBCs per high-power field. The means of obtaining the specimen should be indicated (i.e., normal voiding sample, clean catch, midstream, catheterization, suprapubic tap, or cytoscopy), especially in clinical trials in which an antidiuretic or antibiotic (or other relevant drug) is being tested.

It is usually unnecessary to obtain a microscopic examination on all urinalyses unless there are reasons to believe that important information and data may be lost. This is particularly true after it has been demonstrated that the test treatment does not affect the parameters measured in the microscopic evaluation of urine.

### 6.5 Urine Screens

A urine screen can be used to confirm generally that patients being screened or entering the baseline period of the clinical trial are not using agents (legal or otherwise) contraindicated in the protocol. It can also be used on a scheduled or random basis during the study to confirm that patients are not using such agents. The urine screen is limited in that it is unable to detect positive compliance with the protocol and only measures certain aspects of compliance failure. If a urine test will be conducted at unannounced times in the clinical trial, then this point must be mentioned in the informed consent.

The number of agents tested in the urine screen is generally determined individually for each clinical trial, since there is a wide variety of possible drugs

that may be measured. The choice of drugs to screen will be based on their relative importance for the trial plus the cost and reliability of the methodology. Results of urine screens are usually best viewed in qualitative (i.e., present or absent) rather than quantitative terms. The identification of specific drugs in a patient's urine may help in explaining unusual adverse reactions, laboratory abnormalities, or other events. Urine screens may detect the presence of the therapeutic under study. If the urine screen is able to detect the presence of the study drug, and this is reported as an unknown drug that is present or as a false-positive for another drug, then it could essentially unblind a double-blind clinical trial. To prevent this situation from occurring, data from urine screens may be reported to a nonblinded monitor rather than to the investigator. If a sample of the study drug is put in urine at a physiological concentration and sent to the laboratory, the possibility of cross-reactivity with known agents may be assessed prior to initiation of the trial.

***Type of Container to be Used*** The specific type of container used to collect blood or urine samples is sometimes indicated in a protocol, especially if a special anticoagulant or additive is required or if other specific conditions of sample collection and handling are required. It is generally not necessary to provide this information for commonly requested laboratory tests.

### 6.6 Use of International System Units

Although the international system of laboratory analyte units is almost universally agreed upon, many people in the United States resist using it. Typically, these are physicians (and others) who desire to retain the system with which they were trained, which makes more sense to them.

### 6.7 Identifying New Diagnostic Laboratory Tests

Numerous laboratory tests are periodically performed as aids in the diagnosis of disease states. The standards that must be met before a new test is accepted are extremely high, particularly in terms of calculated rates of false-positive and false-negative results. A five-step process is proposed leading up to acceptance of a new diagnostic test.

### 6.8 Ophthalmological Examination

Various parts of the ophthalmological examination are shown in Table 13. The most important common ophthalmological test to evaluate patients for the occurrence of chronic drug-induced toxicity is slit-lamp examination. Specific types of drugs with known potential for ocular toxicity may require that special attention be directed to other evaluations shown in Table 13. Most drugs that are to be taken systemically require at least some evaluation of ocular safety prior to approval for marketing.

## 6.9 Dermatological Examinations

A few selected safety measurements and tests for specialized dermatological examination are listed in Table 12.

In evaluating the safety of drugs using laboratory or other tests, it is important to develop data that help establish the nature and magnitude of any issue or problem (real or potential) that arises with abnormal laboratory data. Data obtained must also measure the strength of the association between the drug and the event noted or of the serial trends that are observed. While this information is being collected, the definitive courses of action in dealing with the issue or problem can be developed and evaluated. These countermeasures may take the form of (1) periodic monitoring [i.e., prothrombin (PT) or partial thromboplastin (PTT) times for patients receiving anticoagulants], (2) cessation of medicine treatment, (3) decreasing the dose or changing the dose schedule, (4) initiating countertreatment, (5) specific antidotes that may be used to counter or reverse medicine effects, (6) increasing surveillance of the patient, or (7) various other alternatives.

## 6.10 Deaths in Clinical Trials

Certain ADRs may be sufficiently alarming so as to require very rapid notification to regulators in countries where the medicinal product or indication, formulation, or population for the medicinal product is still not approved for marketing because such reports may lead to consideration of suspension of or other limitations to a clinical investigation program. Fatal or life-threatening, unexpected ADRs occurring in clinical investigations qualify for very rapid reporting. Regulatory agencies should be notified (e.g., by telephone, by facsimile transmission, or in writing) as soon as possible, but no later than seven calendar days after first knowledge by the sponsor that a case qualifies, followed by a report that is as complete as possible within eight additional calendar days. This report must include an assessment of the importance and implication of the findings, including relevant previous experience with the same or similar medicinal products.

Determining the cause of death in clinical trials is extremely important, but this goal is often difficult or impossible to achieve. Investigators should be prepared to present reasons to family members to convince them of the importance of conducting an autopsy. Such an autopsy should include examination of the brain, whenever possible.

Any history of drug or alcohol abuse by a patient should trigger a request for appropriate blood and urine tests. Blood samples should always be taken to assess the levels of study drugs and any concomitant agents used. The drug containers should always be analyzed to confirm their contents. This usually entails sending these drugs to their manufacturer.

The circumstances surrounding the patients' death should be as well documented as possible, including a description of all possible influences of the

clinical trial procedures on the death, even influences that are clearly independent of the medicine(s) being tested. Even procedures in a clinical trial apparently unrelated to a patient's death may have contributed to the death in some way. For example, these procedures could include (1) the requirement for excessive physical exertion, (2) prolonged periods of psychologically difficult testing that lead to extreme fatigue, or (3) giving patients many (e.g., 30) large capsules to ingest per day that lead to choking or aspiration.

Evaluation of the data surrounding the death by physicians who are unassociated with the clinical trial lends additional credibility to the report and conclusions. Physician biases probably will strongly influence their decision regarding the association of a patient's death with the clinical trial, and this factor must be considered in interpreting their report. This is particularly true for developing survival curves in cancer or other often fatal diseases, when deaths unrelated to the disease or to the treatment are excluded from the analysis.

## 6.11 Behavioral Rating Scales, Performance, Personality, and Disability Tests

A number of behavioral rating scales and psychometric and performance tests, listed in Tables 18–20, are briefly summarized below, since many of these scales and tests may be used to evaluate safety as well as efficacy. The following comments on the tests provide only a few highlights; readers interested in more details are advised to obtain additional information before choosing the tests that appear most relevant to their particular protocol.

These scales may be used either as part of a clinical trial or as major endpoints in an efficacy trial. Here they are described as a means of obtaining ancillary data on psychological factors in a clinical trial. If these scales are used to demonstrate efficacy, it is mandatory to include only those scales known to be valid.

Unless otherwise noted, all of the adult and children's behavioral scales are given once pretreatment and at least once posttreatment (depending on the trial design, subject drug pharmacokinetics, and length of the trials). Investigators may schedule additional evaluations with these tests, but this is usually not done at less than weekly or biweekly intervals. Many tests provide data on both a total score and subtest (factor) scores. The times given to complete tests are subject to significant variation depending on the anxiety and characteristics of the patient and/or the experience of the professional. The times listed do not include either scoring or preliminary and/or necessary observations of the patient.

## 6.12 Adult Behavioral Rating Scales

***Anxiety Status Inventory*** The Anxiety Status Inventory (ASI) scale is the professional-rated version of the Zung Self-Rating Anxiety Scale (SAS). Both

tests (ASI and SAS) contain 20 items, each with a four-point scale, and are designed for use in adults diagnosed as having anxiety neurosis. Both assess anxiety as a clinical disorder rather than a "feeling state." The tests rate either the present time or the average status of the patient during the week preceding the evaluation. The ASI takes up to 15–20 min to complete and gives two scores: state anxiety and trait anxiety.

***Beck Depression Inventory*** The Beck Depression Inventory (Beck) test may be used to measure the depth of depression as a rapid screen for depressed patients. It is a self-rating scale of 21 items (13 in a shortened form), with each item rated on a four-point scale. It measures the immediate present and has been used in antidepressant medicine trials. The original 21-item scale can be completed in about 10 min and the test is able to discriminate between anxiety and depression. No subtests are present in the Beck.

***Brief Psychiatric Rating Scale*** The Brief Psychiatric Rating Scale (BPRS) is used primarily in adult inpatients to evaluate treatment response in medicine trials and in nonmedicine clinical treatment, but it is also used in some outpatient trials. Abbreviated instructions are printed on the form. Ratings are based on observations of patients. Originally developed for psychopharmacological research, this test contains 18 symptoms, each rated on a seven-point severity scale. It requires approximately 20 min to complete and rates the period of time since the last test. If the test is being used for the first time, it rates the previous week. Five separate subscales are obtained: anxiety–depression, anergia, thought disturbance, activation, and hostility–suspiciousness.

***Carroll Rating Scale for Depression*** The Carroll Rating Scale for depression (52-item self-rating scale) is scored with yes or no answers by patients. It was designed to match closely the information content and specific items included in the Hamilton rating scale. It has been validated by comparisons with both the Hamilton Depression Scale (HAMD) and Beck and requires approximately 20 min to complete. Seventeen components of depression are measured.

***Clinical Global Impressions*** Although the *ECDEU Assessment Manual for Psychopharmacology* (Guy, 1976) provides a formal test for the Clinical Global Impression (CGI) Scale, numerous investigators have modified the three major questions as well as the scales used in order to fit this test to their own clinical trials. The three questions, which may be applied in almost all phase II and III clinical trials, are:

1. *Severity of Illness* "Considering your total clinical experience with this particular population, how mentally ill [the investigator may substitute a more appropriate term if this is not applicable] is the patient at this time?"

2. *Global Improvement* "Rate total improvement, whether or not in your judgment it is due entirely to medicine treatment."
3. *Efficacy Index* "Rate this item on the basis of medicine effect only." This utilizes a rating of both efficacy and adverse reactions and divides the efficacy score by the adverse reaction score to form a ratio (*efficacy index*).

Severity of illness is the only one of these three that is rated pretreatment. All three questions may be rated posttreatment, and additional ratings are possible during a clinical trial. The CGI measure, which is widely used in all types of medicine trials, is generally well accepted.

A scale of two to nine gradations is usually used for questions 1 and 2, although five or so gradations are probably most common. A typical five-point scale for question 2 would be that the patient is rated as 1 (much worse), 2 (minimally worse), 3 (unchanged), 4 (minimally improved), or 5 (markedly improved).

***Clyde Mood Scale*** The Clyde Mood Scale test may be used as either a self-rated or observer-rated scale. It contains 48 items to measure mood and has been shown to be sensitive to medicine effects. The test takes 5–15 min to complete and measures the immediate present in a patient or normal individual. The test gives six scores: friendly, aggressive, clear thinking, sleepy, unhappy, and dizzy.

***Covi Anxiety Scale*** The Covi Anxiety Scale is a global observer's rating scale of patient anxiety. There were three items that are each rated on a 0–5 scale. The test is simple to use and requires only a few minutes to complete.

***Crichton Geriatric Rating Scale*** The Crichton Geriatric Rating Scale test measures the level of behavioral function in elderly psychiatric patients using a five-point scale on 11 items. It rates either the present or the period within the last week and takes 5–10 min to complete.

***Depression Status Inventory*** The Depression Status Inventory (DSI) scale is the professional's version of the Zung Self-Rating Depression Scale (SDS). Each of the two scales (DSI and SDS) consists of the same 20 items rated on a four-point scale and is applied to adults with depressive symptomatology. The DSI is completed by the professional, and the SDS is completed by the patient. Both tests take about 5–10 min to complete. The DSI rates either the present situation or the situation for the last week prior to the test, and a total score is obtained.

***Hamilton Anxiety Scale*** The Hamilton Anxiety (HAMA) scale was designed to be used in adult patients who already have a diagnosis of anxiety neurosis rather than for making a diagnosis of anxiety in patients who have

other problems. The test contains 14 items, each with a five-point scale, and is completed by a physician or psychologist. The test emphasizes the patient's subjective state. The two subscales determined are somatic anxiety and psychic anxiety.

***Hamilton Depression Scale*** The HAMD is one of the most widely used tests to quantitatively evaluate the severity of depressive illness in adults. The most widely used form of this test contains 21 items covering a broad range of symptomatology, with a three- to five-point scale for most items. The minimum time required to complete this test is usually 10–20 min, and it requires a skilled interviewer. Either the present time or the period within the last week is rated. Six subscales are obtained in the HAMD: anxiety/somatization, weight, cognitive disturbance, diurnal variation, retardation, and sleep disturbance.

***Hopkins Symptom Checklist*** The Hopkins Symptom Checklist (HSCL) is a scale that has been used to measure the presence and intensity of various symptoms in outpatient neurotic patients. It is a 58-item self-rating scale and has generally been replaced by the Self-Report Symptom Inventory (SCL-90). It measures the symptoms during the past week and requires approximately 20 min to complete. There are five subtests: somatization, obsessive–compulsive, interpersonal sensitivity, depression, and anxiety.

***Inpatient Multidimensional Psychiatric Scale*** The Inpatient Multidimensional Psychiatric Scale (IMPS) is used to measure psychotic syndromes in hospitalized adults capable of being interviewed. The 89 items are rated on the basis of a psychiatric interview. This test has been well validated and requires 10–15 min following a 35–45-min interview. There are 10 scores: excitement, hostile belligerence, paranoid projection, grandiose expansiveness, perceptual distortions, anxious intropunitiveness, retardation and apathy, disorientation, motor disturbances, and conceptual disorganization.

***Nurses Observation Scale for Inpatient Evaluation*** The Nurses Observation Scale for Inpatient Evaluation (NOSIE) (30-item test) is used by nursing personnel to rate a patient's behavior on the ward, with a five-point scale for each item. This test is widely used and is well accepted for adult inpatients. The test, which rates the most recent three days, is relatively easy to use and requires 3–5 min to complete.

***Plutchik Geriatric Rating Scale*** The Plutchik Geriatric Rating Scale (PLUT) (31-item test) is designed to measure the degree of geriatric functioning in terms of both physical and social aspects. The three-point scale for each item is completed on the basis of direct observation of the patient's behavior and takes 5–10 min to complete. The subscales measure overall dysfunction,

aggressive behavior, sleep disturbance, social isolation, sensory impairment, work and activities, and motor impairment.

***Profile of Mood States*** The Profile of Mood States (POMS) self-rating scale is used in both normal and psychiatric outpatients to evaluate feelings, affect, and mood. It has been widely used in medicine trials. The 65 adjectives included in this test may be used to rate the present and/or previous week. This test requires approximately 5–10 min to complete and provides scores for six subtests: tension–anxiety, depression–dejection, anxiety–hostility, vigor, fatigue, and confusion.

***Sandoz Clinical Assessment—Geriatric*** The Sandoz Clinical Assessment—Geriatric (SCAG) test measures 18 individual symptoms plus a global rating using a seven-point scale similar to those used in the Brief Psychiatric Rating Scale. It measures the present period or that within the last week, requires about 10–15 min to complete, and does not contain subtests.

***Self-Report Symptom Inventory*** Each of the 90 items in the SCL-90 uses a five-point scale of distress. It was designed as a general measure of symptomatology for use by adult psychiatric outpatients in either a research or a clinical setting. It rates either the present or previous week. It requires about 15 min for the patient to complete this form and about 5 min for a technician to verify identifying information. This test is sensitive to drug effects and may be used with inpatients. Nine subscales are measured: somatization, obsessive–compulsive, interpersonal sensitivity, depression, anxiety, anger–hostility, phobic anxiety, paranoid ideation, and psychoticism.

***Wittenborn Psychiatric Rating Scale*** The ECDEU version [Wittenborn Psychiatric Rating Scale (WITT)] is a 17-item test shortened from the original 72-item test. All but one item use a four-point scale, and the test takes 5–10 min to complete. It is used in both inpatients and outpatients and rates either the present or previous week. This test is not intended to make diagnoses but to reflect changes within one patient and to provide a basis for comparing different patients. This test provides descriptive, as opposed to etiological or prognostic, information on patients and includes the following subscales: anxiety, somatic–hysterical, obsessive–compulsive–phobic, depressive retardation, excitement, and paranoia.

***Zung Self-Rating Anxiety Scale*** The SAS test requires approximately 5–10 min to complete.

***Zung Self-Rating Depression Scale*** The SDS test requires approximately 5–10 min to complete.

## 6.13 Pediatric Behavioral Rating and Diagnostic Scales

Many of the behavioral rating scales described for adults are not suitable for use in the pediatric population. Special tests have been designed, and a number of pediatric behavioral rating scales are presented in Table 17. General comments on these tests are presented below. A further description of rating scales used in pediatric medicine trials is given in the *ECDEU Assessment Manual for Psychopharmacology* by Guy (1976). His article is a practical guide to identifying appropriate scales for a particular situation. Conners discusses the two broad approaches of many pediatric rating scales as either "rating current behaviors, symptoms or states; or … describing basic traits, dispositions, and personality characteristics." The choice of one of these two approaches depends in part on the purpose of using a scale in a medicine trial. Three general purposes have been suggested for using a behavioral test (prediction, measurement of change, and classification). The choice of one of these three purposes usually implies that one of the two specific approaches implicit in the pediatric behavioral scales will be more appropriate:

1. To be able to predict something about a patient, choose a scale that rates basic traits.
2. To measure change in a patient, choose a scale that rates current symptoms.
3. To assess a patient's classification, choose a scale that rates either basic traits or current symptoms, depending on the purpose of the classification.

The type of patient population and the desired format of the test to be used in a clinical trial also influence the particular scale(s) chosen.

An evaluation system that can be used in a wide variety of pediatric inpatients is the Children's Behavior Inventory.

***Children's Behavior Inventory*** The Children's Behavior Inventory (CBI) is an 139-item, two-point (yes–no) scale to record maladaptive behavior in children aged 1–15 years. Relatively little training is needed to administer this test. It is easily used by nurses, teachers, graduate students, psychologists, and others. This test usually requires at least 2 h of observation of the child, but better reliability is achieved if behavior is observed over an 8-h period. Nine subtest scores are provided: anger–hostility, conceptual dysfunctioning, fear and worry, incongruous behavior, incongruous ideation, lethargy–dejection, perceptual dysfunctioning, physical complaints, and self-deprecation.

***Children's Diagnostic Classification*** The Children's Diagnostic Classification (CDC) test may be used instead of the Children's Psychiatric Rating Scale (CPRS) to arrive at a diagnosis. This differs from the CPRS in that it is highly directed and leads the observer to a diagnosis. It rates the current status

of the child and may be used at pretreatment and/or the termination of the clinical trial.

***Children's Diagnostic Scale*** The Children's Diagnostic Scale (CDS) is used in children up to 15 years of age to assist in the diagnosis and classification of the child's condition. It contains 13 items, 8 of which have a seven-point scale. The others are specific diagnostic questions. It measures current status only and is mainly used at the start of a study, although it may be used at the termination of the study as well.

***Children's Psychiatric Rating Scale*** The CPRS is a comprehensive scale to assess a wide range of psychopathologies in children up to age 15. It contains 63 items, with a seven-point scale derived from the BPRS. This test rates 28 items by direct observation of the child based on behavior expressed during the interview and rates other items based on the child's reports of events that occurred either over the preceding week or during the interview. Scores of 15 separate clusters of the rated items are provided as well as the overall score.

***Clinical Global Impression*** See adult behavioral rating scale description of the CGI.

***Conners Parent Questionnaire*** The Conners Parent Questionnaire (PQ) is a 94-item checklist of symptoms that evaluates common behavior disorders using a four-point scale in children up to 15 years of age and takes 15–20 min to complete. It is used once pretreatment and may be repeated but is often replaced after the first use by the 11-item Conners Parent–Teacher Questionnaire (PTQ). There are eight subscales: conduct problem, anxiety, impulsive–hyperactive, learning problem, psychosomatic, perfectionism, antisocial, and muscular tension.

***Conners Parent–Teacher Questionnaire*** See descriptions above for Conners Parent Questionnaire and below for the Conners Teacher Questionnaire (TQ). The PTQ is used in conjunction with either the PQ to TQ and yields a total score only (i.e., no subscales are given). The PTQ takes about 5 min to complete and is not used pretreatment.

***Conners Teacher Questionnaire*** The TQ form was designed to obtain teacher evaluations of children up to age 15 in terms of their interactions with peers and their ability to cope with the school environment and requirements. There are 41 items, and the first 39 have a four-point scale. Question 40 deals with the teacher's evaluation of the child's severity of illness, and question 41 deals with global improvement in four different areas. This test is used once at pretreatment and as needed afterward. It takes about 15 min to complete and covers either the present or any interval period up to 1 month. A shorter 11-item PTQ is often used after the initial use of the 41-item TQ. The five

subscales included are conduct, inattentive–passive, tension–anxiety, hyperactivity, and social ability.

***Devereux Child Behavior Rating Scale*** The Devereux Child Behavior Rating Scale contains 97 items and is similar to the Devereux Teacher Scale. It is used for emotionally disturbed and mentally retarded children aged 8–12 years. Besides being easy to use, this scale is well researched and discussed in the literature. It requires 10–20 min to complete by clinicians, child care workers, parents, or others and gives 17 scores. There is a Devereux Adolescent Behavior Rating Scale for children from ages 13 to 18.

***Devereux Elementary School Behavior Rating Scale*** The Devereux Elementary School Behavior Rating Scale is a widely used test incorporating 47 items that have high test–retest reliability. It uses a checklist format and is easy to use (requires 10 min). There are 11 factor scores and 3 item scores.

## 6.14 Psychometric and Performance Tests

The psychometric and performance tests presented in Table 18 may be grouped as being applicable for use in either children or adults. In children, the tests measure intellect [GOOD, Porteus Mazes, Wechsler Intelligence Scale for Children (WISC), Peabody], achievement [Wide Range Achievement Test (WRAT)], and motor performance (vigilance tests, reaction time). There are other tests that may be used to measure learning, although many of these tests utilize equipment and are not described. All of these tests (unless otherwise noted) are given once pretreatment, at least once posttreatment, and at additional times if desired by the investigator. The contribution of learning in the scores obtained at second and third testings is usually unknown. The methods used to motivate patients to perform to the best of their ability in all tests must be standardized and reported.

***Bender–Gestalt Test*** The Bender–Gestalt is a nonverbal performance test in which the individual copies a design shown on a card. It is often used to identify a problem of visual perception and/or motor performance or minimal brain dysfunction in children.

The scoring used for children (age 4 or 5–11 years) differs from that used for adults (age 15 years to adult). This test measures perceptual maturity, possible neurological impairment, and emotional adjustment in children. It measures maturation, intelligence, psychological disturbance, and cortical impairment in adults. The test requires 10 min to complete. Scores may fluctuate from test to test and thus must be interpreted carefully.

***Conceptual Clustering Memory Test*** For the Conceptual Clustering Memory Test, patients are given a list of 24 specific words from a number of different categories such as birds, cars, or types of drinks. The words are pre-

sented one at a time over 2 min, after which patients are asked to recall as many of the specific words as possible. The test measures the total recall as well as the degree to which words of a specific category (e.g., animals) are recalled from the cluster of words given in that category (e.g., dog, cat, cow).

***Digital Symbol Substitution Test*** A subtest of the Wechsler Adult Intelligence Scale (WAIS), the Digital Symbol Substitution test, measures sensorimotor integration and learning relationships of symbols. It has been used in many psychopharmacological studies. Subjects are given different forms of this test at each session. The test requires the patient to match as many of 100 symbols to their respective numerals, found in a code key, as possible within 60 s.

***Embedded Figures Test*** For the Embedded Figures Test, patients are shown a complex design and must identify as quickly as possible a simple figure that is "embedded" within the design. Twenty-four embedded figures are included, and a maximum of 3 min is allowed for each one.

***Frostig Developmental Test of Visual Perception*** The Frostig Developmental Test of Visual Perception (FROST) measures the development of perceptual skills in children from four to eight years of age or in older children with learning difficulties. It may be administered individually (requires 30–45 min) or to groups (requires 40–60 min).

***Goodenough–Harris Figure-Drawing Test*** The Goodenough-Harris Figure-Drawing Test is a brief (10–15-min) easy-to-use test for children 4–15 years of age to measure intellectual maturity.

***Peabody Picture Vocabulary Test*** The Peabody Picture Vocabulary Test is a rapid 10–15-min intelligence test for children aged 2.5–18 years that is useful when there is inadequate time to give the WISC.

***Porteus Mazes*** The Porteus Mazes is a nonverbal test that has been shown to be sensitive to medicine effects in both children (over three years) and adults. The test has three series of mazes to prevent score improvement on retesting with the same test. It requires about 25 min and provides both a qualitative and a quantitative score.

***Reaction Time*** There are many different tests used to measure reaction time. These tests measure the period of time between the presentation of a stimulus to a patient and the onset of the resulting response. The signal is usually a visual or auditory stimulus, and the onset of a motor reaction, such as the lifting of a finger, arm, or leg or the pressing of a buzzer, is used to measure the speed of response.

In simple reaction times, a stimulus is presented that always requires the same response, even if the nature of the stimulus changes. A complex reaction time requires the patient to respond to some stimuli but not to others.

***Vigilance Tests*** Numerous tests have been designed to measure vigilance. In these tests, patients are requested to respond in some manner to certain stimuli or occurrences but not to others. The stimuli may be controlled to present minimally perceived signals that require vigilance on the part of the patient.

***Wechsler Adult Intelligence Scale*** The WAIS consists of 11 subtests—6 verbal tests and 5 performance tests. This provides an age-related IQ in adults from 16 to 75 years of age; that is, the test measures intelligence of the person in relation to his age group and not to the entire population. It may be used either as an initial assessment or as a tool to measure change. The test, which takes 40–60 minutes to complete, provides 13 scores in verbal and performance categories plus a total score.

***Wechsler Intelligence Scale for Children*** The WISC was extensively revised in 1974, and it became the WISC-R, which requires 40–60 minutes to complete. This widely used scale in children from 6 to 16 years of age may be used for either screening or baseline data or as a measure of change. There is a "preschool and primary scale of intelligence" version that may be used in children from 4 to 6½ years of age (requires 50–75 min). The WISC-R has six verbal and six performance subtests.

***Wechsler Memory Scale*** The Wechsler Memory (WMEM) Scale is a brief test that is used to measure memory deficits. There are two forms of the test, and they are generally alternated to avoid a training effect in children taking the test on two or more occasions.

***Wide Range Achievement Test*** The WRAT is used in children from age five years to adult in college. It assesses basic skills in reading, spelling, and mathematics. It is simple and easy to administer and requires 20–30 min to complete.

## 6.15 Personality Tests

In addition to the above behavioral and performance tests, there are a number of well-known tests of personality that may provide useful information in select clinical studies. The most well known of these tests is the Minnesota Multiphasic Personality Inventory (MMPI). This test consists of 550 affirmative statements to which a true or false response is given and requires about 1 h to complete. It is given to adults over the age of 16 and is scored for 10 scales: depression, hysteria, hypochondriasis, psychopathic deviate, masculinity–femininity, paranoia, hypomania, schizophrenia, psychasthenia, and social introversion.

## 7 ASSESSMENT OF UNWANTED DRUG EFFECTS

Because of the relatively small numbers of volunteers and patients involved, only the most common of drug-related adverse events are likely to be detected during early studies. For example, to have a 95% chance of picking up three subjects who have experienced an adverse reaction (with no background incidence) which occurs in 1 in every 100 subjects treated with the drug, it would need to be given to 650 subjects (Gad, 1998). Matters are made worse when the adverse event in question also occurs in the general population, which is usually the case with the kind of symptoms reported by volunteers and patients taking part in drug studies. No matter how good the study design is, nothing can compensate for this problem of inadequate numbers. In this respect, all of the study designs described earlier are more or less equally adequate or inadequate, as the case may be.

Monitoring for drug-related adverse events employs the same or similar methods in both volunteers and patients. In both cases assessments of tolerability and safety are based upon symptom reports, routine laboratory safety screens, ECG monitoring, and on occasion special tests designed to detect unwanted effects associated with a particular class of drug. The chances of obtaining reliable information on a drug's safety profile are enhanced by detailed and careful monitoring. Symptoms may be reported spontaneously or elicited in reply to standard questions. Open questions such as "how are you feeling?" are to be preferred to leading questions on the basis that they result in fewer reports of adverse events. If leading questions are used, they need to be carefully worded. A certain amount of basic information is required on all adverse events, that is, type, severity, time of onset in relation to time of dosing, duration, and causality. Attributing the cause of an unwanted effect to the drug or some other factor can be difficult, particularly when little is known about the drug, as is often the case at the stage of initial studies in volunteers or patients. Rechallenge with the drug ideally using the same dose or if need be (because the event caused a degree of discomfort) a reduced dose is probably the single best way of proving or disproving a causal relationship; but if done, the rechallenge procedure must be designed using placebo as a comparator under double-blind conditions. Obviously rechallenges can be done only if the adverse event was reversible, did not cause excessive discomfort, and most importantly was not life threatening.

***Separation of Adverse Reactions from Placebo Reactions*** Since adverse nondrug symptoms are common (Reidenberg and Lowenthal, 1968) and are not easily separated from drug-induced symptoms, both must be collected for analysis if a complete profile of adverse reactions is to be made. However, this technique can only be sued in controlled studies, ideally with placebo, as well as with other standard drugs. The temptation to subtract the number of the particular adverse reactions in the placebo group from the number in the active drug group as

$$\text{Drug group} - \text{placebo group} = \text{number of adverse reactions to drugs}$$

should be resisted because:

- The difference may not be statistically significant and may have arisen by chance.
- Although the total number of events may be statistically different in the two treatment groups, it is also necessary to establish whether the number of patients afflicted with the adverse event is different and vice versa.
- Having established that there is a significant difference between the two treatment groups for the number of events and the number of patients afflicted, the severity of the ADRs in the two groups should be compared.

A further problem is that due to classification; some terms may include more than one type of abnormality (e.g., the incidence of "blurred vision" may be equal in both groups, but there may be several cases of tunnel vision with the trial drug but because there is no code for tunnel vision it is coded under more general terms). Another problem is that the symptoms forming a syndrome are often coded separately and individually and there may be no difference between two drugs, but when the cases are examined there may be a combination of symptoms with one drug that warrant being called a syndrome. It is therefore essential to read the individual original description of the adverse events (AEs) before making a judgment. This area has been explored more fully by Bernstein and he has added bias to the equation:

$$\text{Attributable AEs} = \text{drug group AEs} - \text{placebo group AEs} \pm \text{bias}$$

where bias is equal to the baseline ($B$) frequency and severity of the AE multiplied by the pharmacological clinical activity of the drug (AD) minus the pharmacological clinical activity of the placebo (AP):

$$\text{Bias} = \text{B}(\text{AD} - \text{AP})$$

The argument is that the disease or a symptom or sign of the disease and the drug ADR may interact as follows:

- Compliance—Early improvement may cause the patient to stop the drug and the improvement of the ADR may be inappropriately assigned to tachyphylaxis of the ADR; failure of the disease to improve may persuade the patient to add a rescue drug or increase the dose of the study drug or even stop the drug; impaired mental or cognitive function due to the disease may affect compliance.
- The disease may alter the absorption, distribution, metabolism, or elimination of the drug (e.g., alteration of the blood–brain barrier by the disease may allow the drug to affect the brain).

- Observational bias of convalescence (e.g., severe pain causing insomnia may require morphine, causing compensatory hypersomnia in excess of that caused by morphine alone).
- Observational bias by halo effects. Perception of an ADR may be swamped by the symptoms of the disease; thus as the disease symptoms resolve, the ADR becomes apparent.
- Unblinding—If the patient or physician is unblinded due to rapid improvement of the disease or an ADR, he or she may be led to expect ADRs with the active treatment.
- Pharmacological clinical activity bias—An AE that is already present due to the disease may be increased if it is also an ADR of the drug or vice versa. For example, the diarrhea of gastroenteritis may be alleviated by codeine-containing preparations given to relieve pain while the inertia of a severely depressed patient may be sufficiently resolved by an antidepressant to enable the patient to commit suicide.

Adverse drug reactions that are similar to common nondrug AEs are rarely described or investigated sufficiently for a causal relationship for each individual event to be established. If they cannot be distinguished qualitatively, the correct quantitative procedure is to compare them using nonparametric statistics, giving the confidence limits for the incidences of ADRs. Small studies ($n < 30$) have little chance of separating ADRs from placebo or nondrug events unless they are very common and specific to the drug. The situation is worsened by the fact that members of a placebo group have a tendency to "catch" AEs from the active drug group, therefore changing a relatively specific ADR to a nonspecific event.

### 7.1 Base-Case Causality of Single-Event Adverse Drug Reactions

The analysis and evaluation of ADRs are major problems in both the development of new drugs and the postmarketing surveillance period. Just as there are standards and requirements established as guidelines in the chemical, pharmacological, and toxicological phases preceding the marketing of a new drug, there are also guidelines for causality assessment of individual human cases in both pre- and postapproval for new drugs.

The most common problem of assessment is the single-event ADR case. Presented here is an approach to such single-event ADR cases. This methodology, relating to the use of therapeutic, diagnostic, and prophylactic-type drugs in a clinical setting, should permit the diagnostician to make one of three responses after an assessment of an ADR case: an assured *yes*, a firm *no*, or a reasoned admission of *uncertainty*. The clinicopathological picture presented by the ADR case is often not readily distinguishable from non-drug-induced diseases. The clinical and morphological findings of ADRs have the same

limited number of final common paths that characterize these other (nondrug) human illnesses.

There are three major requirements for establishing the occurrence of an ADR:

1. The possibility and likelihood of a causal relationship between the drug and the ADR must be confirmed by establishing its eligibility.
2. Linkage of the drug with the clinicopathological findings.
3. The degree of certainty of this drug linkage should be determined.

As an initial background for developing this algorithm or methodology, Figure 4 is offered for consideration and orientation. This figure has the basic elements of a "time flow chart," which has considerable utility in evaluating ADR cases.

In this graphical representation of an ADR, the ordinate ($Q$) represents any of the findings of an ADR. *Specifically,* $Q$ may be a symptom (pain, nausea, etc.), a sign, a clinical laboratory result, a radiological finding, a morphological finding, or any combination of these. Synonyms for $Q$ include marker, disease marker, signal, indicator, parameter, detector, response, and effect.

The abscissa is the time element ($T$) related to both the time of drug administration and the dating of disease marker data. Both are usually plotted on the same time flow chart in a particular case.

This graphical representation of an ADR case will be sued frequently in the assessment of eligibility and linkage determinations of ADRs. The four eligibility criteria are also listed in Figure 5.

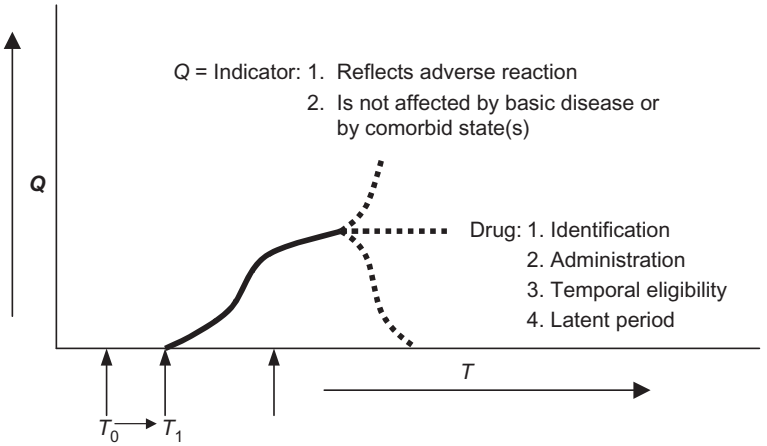


***Figure 5*** *Adverse drug reaction (curve* Q*) plotted against time (abscissa* T*). Dashed lines show three courses an ADR can take: increasing severity to death; leveling off to chronicity; or return to abscissa, indicating recovery. Four criteria that must be met before drug is eligible to be empiric correlate of* Q *(adverse drug reaction) are listed.*

***Administration of the Drug*** As it is with accurate identification of a drug, so it is that its "administration" must at times be held in question. Subject compliance with the study protocol is not a rare problem in clinical trials. Complete noncompliance sometimes occurs.

***Temporal Eligibility*** The time factor in assessment of ADRs is a very important one and in some cases is of critical diagnostic importance. This is true not only in establishing "eligibility" of the drug but also in linking the drug to the reaction. On the other hand, the time element may be equally important in denying eligibility and also makes linkage of the drug with the clinicopathological picture a most unlikely possibility.

It is quite apparent that a drug cannot be responsible for an ADR if the latter is already in progress before the drug is first administered. This dyssynchronicity is sometimes seen both in trials and later in the marketplace.

***Latent Period*** Latent period refers to the time interval between the initial administration of the drug and the onset of the ADR (in Figure 5, it is the interval $T_0$ to $T_1$). The latent period is not rigidly fixed or exactly predictable, but it tends to fall within certain limits.

Characteristically, strychnine deaths occur in seconds to minutes. Most anaphylactic deaths occur within 20–30 min after contact with the lethal antigen, while jaundice associated with most drugs has its onset within three days to three weeks after the beginning of therapy. The fatal pancytopenia following chloramphenicol appears in one to three months, while hepatic angiosarcoma related to aflotoxin has a latent period of one to several decades. The ultimate in length of latency is one to several generations from a drug-induced mutational germ cell change to its manifestation in a conceptus.

Consideration of the latent period in an ADR is of use in an ADR assessment in one of two ways: The latent period may be too long or too short.

In summary, identification, administration, temporal eligibility, and latent period are the four criteria for establishing the eligibility of a drug to have caused an ADR. Emphasis should be placed on obtaining sufficiently detailed time-related data on drug administration and the appearance of ADR markers. These data are a *sine qua non* in the assessment of drug eligibility.

***Linking Drug with Clinicopathological Findings*** The second major task in analyzing an ADR case is to establish a connection or linkage between the drug and the clinicopathological findings (making empiric correlates of the drug and these findings).

Figure 6 is a time flow chart representing an ADR that itemizes six ways of making this linkage.

***Exclusion*** Exclusion consists of selecting one drug from a group of drug candidates by the use of the time flow chart.

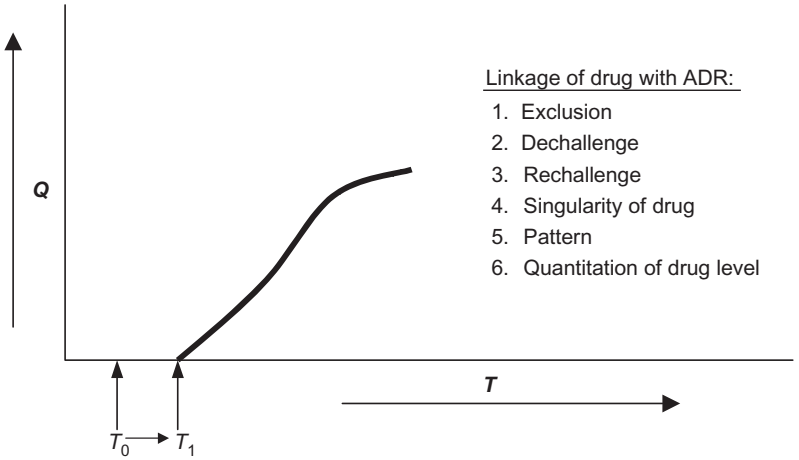


***Figure 6*** *Six methods of linking drug with adverse drug reaction.*

The exclusion method also includes instances in which drug candidates are themselves excluded from causation status because a nondrug etiology is clearly demonstrable (environmental or occupational factors, radiation injury, the underlying disease of the patient, or a comorbid state) and can reasonably account for the clinicopathology findings.

***Dechallenge*** The principle involved in the dechallenge method of linkage is that if there is a reversible effect present, then removing the cause will eliminate the effect.

***Rechallenge*** The principle involved in the rechallenge method of linkage of a therapeutic to an ADR is implied in the phrase *post hoc ergo propter hoc* ("after this therefore because of this").

As applied, if a drug has been incriminated with a reaction and the ADR disappeared when the drug was discontinued, a rechallenge with this drug followed by a return of the ADR would increase the probability that the drug and the ADR were empiric correlates. While intentional challenge is not often done, such a rechallenge may occur inadvertently.

***Singularity of Drug*** The principle involved in the singularity method of linking a drug with an ADR is based on two assumptions: Only one drug was administered and there was no basic disease or comorbid state that could be related to the ADR marker being used in the assessment.

***Pattern*** The pattern method of linking a drug with an ADR shifts the focus of attention to the clinicopathological findings in an ADR and away from the identification of the causative drug. This shift of emphasis is necessary when

detailed time-related drug and disease marker data are unavailable to the evaluator of the case. The site-process profile may then be used as a guideline for searching past experience and the literature for cases that have matching features. Matching features found in the literature may include associations with certain drugs or chemicals, which serves as a guideline for a focused examination of the patient's history for the causative agent.

This "pattern" method may also be used in excluding drugs. If the drugs or chemicals suggested by the morphological findings are not identified or disclosed by historical or toxicological efforts, then the morphological changes appear to remain non–drug or non–chemical related.

***Quantitation of Drug Level*** Assessing an ADR case by quantitation of drug level brings our focus back to the search for and identification of the causative agent by quantitative and objective data based on laboratory analysis of body fluids and/or viscera (Ozdemir et al., 2001). This method is applicable and strongest in the case of higher dose level. The feasibility of this approach is based on the availability of dependable information on lethal levels from past experience or preclinical work. Without this comparison information, there is no judgmental significance to toxicological levels in the case at hand.

Quantitated levels of drugs have limitations in diagnostic value. In adverse reactions in the hypersensitivity, idiosyncratic, and pharmacogenetic categories, drugs have been administered in therapeutic (not toxic) amounts, and blood and other body fluids and tissue levels have been found to lie within therapeutic ranges. Such analyses will confirm any prior administration of the drugs, but the problem of the etiological differential diagnosis will still remain.

In addition to a quantitative approach, qualitative identification can be of value in appropriate instances. In some cases of ADRs, more than one of the six methods of drug linkage that are listed in Figure 5 may be used in causation analysis. In fact, multiple methods in the same case strengthen the confirmation of the rejection of an ADR and its etiology.

***Difficulties in Assessing ADRs*** Requirements for establishing eligibility and methods of linking a drug with an illness have been presented in the preceding discussion. This algorithm should constitute a blueprint for solving many, if not most, of the ADR problems in this area of medical diagnostics.

However, in the hands-on practice of the assessment of ADR cases, there are at least four major difficulties that stand in the way of such high diagnostic expectations:

1. *Incomplete Information* Incomplete information is not unique to ADR evaluation but is common to all areas of medical practice. The lack of sufficiently detailed, time-related data on drug administration and disease markers may make it impossible to render a reasoned judgment on many ADR cases, leaving them in their original and unsatisfactory anecdotal status.

Denial to the evaluation of access to these required facts makes it impossible to make judgments on latent period and temporal eligibility; time flow charts cannot be utilized in exclusion, dechallenge, and rechallenge techniques. The diagnostic database should also include information on any other drugs being administered or taken, concurrent comorbid states, and the existence of any preexisting occupational and environmental hazards.

2. *Polypharmacy* In recent times, polypharmacy is the rule rather than the exception. Patients with complicated and prolonged illnesses may have 20–30 medications in their medical background. Cases of this sort may be of such complexity that even with ideally complete drug and disease marker information, diagnostic success may be elusive.
3. *Lack of Objective Means of Linking Drug to ADR* Tests and procedures that specifically and causally connect a drug to an illness are lacking. Our high-technology laboratory instrumentation is capable of identifying and quantifying extremely low levels of drugs and chemicals, but this type of information falls short of establishing causation.
4. *Limited Number of Toxicological Responses in Human Disease* There are a limited number of generic morphological reaction patterns that diseases fit into (inflammatory, congenital, neoplastic, degenerative, infiltrative, vascular, and functional). In parallel, there are also a rather limited number of clinical symptoms and signs (pain, nausea, fever, lumps, etc.) that come to the attention of the practicing physician. There are a multitude of causes and a multitude of clinical conditions that funnel into these clinicopathological "final common paths." The algorithm previously described is an attempt to move from the generic to the specific in analyzing ADR causation.

Of the above four difficulties, only the first (incomplete information) is subject to at least some degree of improvement.

***Degree of Certainty*** The third major task in analyzing and assessing ADRs is determining the degree of certainty one has as to the causal relationship between the drug and the clinicopathological findings. Interposed between the definitive causative and negative categories are three shades of certainty (probable, possible, and coincidental) that titrate between these two extremes. These degrees of certainty are defined as follows:

1. *Causative* Cases in this class are those in which there is no doubt that a drug has caused the reaction. This category is essentially limited to drug overdose cases or those cases in which the causative agent can be objectively identified (asbestos bodies, granuloma-encapsulated silica, etc.). Parenthetically, the overdose cases with drug levels in lethal ranges should have important negative findings: no anatomical cause of death at autopsy.

2. *Probable* This term is equivalent to the phrase "consistent with," and cases in this category of certainty fall short of the "causative" designation because they lack an objective and quantitative laboratory finding that is the sine qua non of the causative category. Cases placed in this category have the following characteristics:
   (a) The criteria of temporal eligibility and appropriateness of latent period have been met.
   (b) The clinicopathological features are consonant with previous experience and literature precedent for the drug in question.
   (c) Other causes (the basic disease, comorbid states, and other modalities of therapy) have been eliminated from consideration.
   (d) One or several means of linkage of the drug to the ADR have been utilized: exclusion, dechallenge, rechallenge, singularity of the drug, and pattern.
3. *Possible* Cases are put in this category when the relationship between the drug and the clinicopathological findings can be neither confirmed nor denied. There are three subdivisions in this category:
   (a) Cases with potential causes other than the drug in question. The clinicopathological picture could have been produced by the basic disease, a comorbid state, or some other modality of therapy.
   (b) Cases in which some of the criteria for eligibility and linkage have been met but some have not because of lack of adequate information. Such a case could be put in this category temporarily while awaiting more information or placed here permanently if it were evident that further data would not be forthcoming.
   (c) Cases that have met all the criteria of eligibility and linkage but for which there is no known precedent literature. Such a case might be a new and emergent ADR. It could be placed in the "possible" group, awaiting the appearance of similar cases for cluster studies at a later time.
4. *Coincidental* Cases in this category include those that were indeed exposed to the drug in question but in which assessment of the case clearly reveals only an anecdotal association.
5. *Negative* This category applies to those cases in which the alleged drug was not or could not have been in the patient's system at the time of the ADR. This circumstance could be related to noncompliance, mislabeling of the drug, or historical misinformation.

## 8 CONCLUSION

The roles of the toxicologist in clinical trials are as follows:

1. Prospectively and retrospectively evaluate, explain, and extrapolate from the relationships between nonclinical trial findings and adverse clinical trial events.

2. Provide mechanistic insight into the causes, treatment, and avoidance (of further) undesired effects in a clinical trial.
3. Guide selection and refinement of subject profiles for clinical trials.

## REFERENCES

Association of American Universities (2000). Task Force on Research Accountability: Report on university protection of human beings who are the subjects of research. Washington, DC, June 28, 2000. Available: www.aau.edu/HumSUbRpt06.28.00pdf.

Bernstein, D. I., Sack, D. A., Rothstein, E., Reisinger, K., Smith, V. E., and Sullivan, D. (1999). Efficacy of live, attenutated, human rotavirus vaccine 89-12 in infants. *Lancet* 917:417–430.

Biomarkers Definitions Working Group (2001). Biomarkers and surrogate endpoints: Preferred definitions and conceptual frameworks. *Clin. Pharmacol. Ther.* 69:89–95.

Braunwald, E. (2008). Biomarkers in heart failure. *N. Engl. J. Med.* 358:2148–2159.

Buros, O. K. (1978). *The Eight Mental Measurements Yearbook*. Gyrphon, Highland Park, NJ.

Cato, A. E. (1988). *Clinical Drug Trials and Tribulations*. Marcel Dekker, New York.

Code of Federal Regulations (CFR), Title 21, Parts 50, 54, 56, 312, 314. U.S. Government printing Office, Washington, DC. Available: www.access.gpo.gov/nara/cfr/.

Davis, R. (1998). Holes are growing in medical testing safety nets. *USA Today*, June 8, pp. 19A–20A.

Eikelbom, J. W., Mehta, S. R., Pogue, J., and Yusuf, S. (2001). Safety outcomes in meta-analyses of phase II versus phase III randomized trials. *JAMA* 285:444–450.

F-D-C Reports, Inc. (2000). FDA may request "intermediate" enforcement authority for clinical trials. *Pink Sheet* 62(19):16–17.

Fleiss, J. L. (1986). *The Design and Analysis of Clinical Experiments*. Wiley, New York.

Fletcher, A. J., Edwards, L. D., Fox A. W., and Stonier, P. (2002). *Principles and Practice of Pharmaceutical Medicine*. Wiley, Hoboken, NJ.

Gad, S. C. (1998). *Statistics and Experimental Design for Toxicologists*. CRC Press, Boca Raton, FL.

Gallin, J. I., and Ognibene, F. P (2007). *Principles and Practice of Clinical Research*, 2nd ed. Academic, Burlington, MA.

Ganter, J. (1999, Nov.). Responding to industry critics: If the industry doesn't address concerns raised by the consumer press, who will? *Appl. Clin. Trials* 10:18–24.

Green, S., Beneditti, J., and Crowley, J. (2003). *Clinical Trials in Oncology*, Chapman Hall, Boca Raton, FL.

Guy, W. (1976). *ECDEU Assessment Manual for Psychopharmacology*. U.S. DHEW Publication No. (ADM) 76-338. U.S. Government Printing Office, Washington, DC.

Hayachi, K., and Walker, A. M. (1996). Japanese and American reports of randomized trials: Difference in the reporting of adverse effects. *Controlled Clin. Trials* 17:99–110.

Henney, J. E. (2000). Remarks presented at the Association of American Medical Colleges Council on Teaching Hospitals Spring Meeting, May 11. Available: www.fda.gov/oc/speeches/2000/aamc511.html.

International Conference on Harmonisation (ICH) (1997). *General Considerations for Clinical Trials*. ICH, Brussels.

Ioannides, J. P. A., and Lau, J. (2001). Completeness of safety reporting in randomized trials: An evaluation of 7 medical areas. *JAMA* 24:437–443.

Jones, J. H. (1993). *Bad Blood: Tuskegee Syphilis Experiment*. Free Press, Berkeley.

Kaplowilz, N. (2001). Drug-induced liver disorders: Implications for drug development and regulation. *Drug Information Journal* 35:347–400.

Lee, C-J., Lee, L. H, Wu, C. L., Lee, B. R., and Chen, M-L. (2006). *Clinical Trials of Drugs and Biopharmaceuticals*. Taylor & Francis, Boca Raton, FL.

Machin, D., Dag, S., Green, S., Everitt, B., and Bard, G. S. (2004). *Textbook of Clinical Trials*. Academic, San Diego.

Mackintosh, D. R. (2001, Apr.). Bye-bye placebos. *Appl. Clin. Trials*, pp. 47–49.

Matoren, G. M. (1984). *The Clinical Research Process in the Pharmaceutical Industry*. Marcel Dekker, New York.

Meille, C., Gentet, J. C., Barbolon, D., andre, N., Doz, F., and Hiadis, A. (2008). New adaptive method for phase I trials in oncology. *Clin. Pharmacol. SC Ther.* 83:873–881.

Meinert, C. L. (1996). An open letter to the FDA regarding changes in reporting procedures for drugs and biologics proposed in the wake of the FIAU tragedy. *Controlled Clin. Trials* 17:273–284.

Nylen, R. A. (2000). *The Ultimate Step-By-Step Guide to Conducting Pharmaceutical Clinical Trials in the USA*. RAN Institute, Tampa, FL.

O'Brien, K. L., Selaniko, J. D., and Heedivert, C. (1998). Epidemic of pediatric deaths from acute renal failure caused by diethylene glycol poisoning. *JAMA* 279:1175–1180.

O'Donnell, P. (2005). A new GCP bible is born. *Applied Clinical Trials* 14(6):38–42.

O'Grady, J., and Joubert, P. H. (1997). *Phase I/II Clinical Drug Trials*. CRC Press, New York.

O'Grady, J., and Linet, O. I. (1990). *Early Phase Drug Evaluations in Man*. CRC Press, New York.

Office of Inspector General (2000a, June). Protecting human research subjects: Status of recommendations, OEI-01-97-00197. Department of Health and Human Services, Washington, DC. Available: www.dhhs.gov/progorg/oei/reports/a447.pdf.

Office of Inspector General (2000b, June). Recruiting human subjects: Pressures in industry-sponsored clinical research, OEI-01-97-00195. Department of Health and Human Services, Washington, DC. Available: www.hhs.gov/oig/oei/reports/a459.pdf.

Ozdemir, V., Shear, N. H., and Kalow, W. (2001). What will be the role of pharmacokinetics in evaluating drug safety and minimizing adverse effects? *Drug Safety* 24:75–85.

Pharmaceutical Research and Manufacturers Association (PhRMA) (2000). Available: www.phrma.org/publications/publications/brochure/questions/whycostmuch.phtml.

Piantadosi, S. (1997). *Clinical Trials: A Methodologic Perspective*. Wiley, New York.

Rallin, J. I., and Ognibene, F. P. (2007). *Principles and Practice of Clinical Research*, 2nd ed. Academic, Burlington, MA.

Reidenberg, M. M., and Lowenthal, D. T. (1968). Adverse non-drug reactions. *New Eng. J. Med.* 279:678–679.

Rolan, P. E. (1997). The assessment of pharmacokinetics in early phase drug evaluations. In J. C. Grady and P. H. Joubert (eds.), *Handbook of Phase I/II Clinical Drug Trials.* CRC Press, Boca Raton, FL, pp. 169–175.

Rolan, P. E., Mercer, A. J., Weatherly, B. C., Holdich, T., Meire, H., Peck, R. W., Ridout, G., and Posner, J. (1994). Examination of some factors responsible for a food-induced increase in absorption of atovaquone. *Br. J. Clin. Pharmacol.* 37:13–20.

Sallustio, B. C., Purdie, Y. J., Birkett, D. J., and Meffin, P. J. (1989). Effect of renal dysfunction on the individual components of the acyl-glucuronide futile cycle. *J. Pharmacol. Exp. Ther.* 25:288–294.

Schmidt, M. J. (2001, July). Human safety in clinical research. *Appl. Clin. Trials*, July, pp. 40–50.

Shamoo, A. E. (2000). Future challenges to human subject protection. *Scientist* 14(13):35. Available: www.the-scientist.com/yr2000/jun/opin_000626.html.

Sharp, S. M. (2001, Apr.). The ethical foundations of clinical trials. *Appl. Clin. Trials*, pp. 48–54.

Spilker, B. (1991). *Guide to Clinical Trials*. Raven, New York.

Stone, J. (2006). *Conducting Clinical Research*. Mountain side MD Press, Cumberland, MD.

Washington, N., Washington, C., and Wilson, C. G. (2001). *Physiological Pharmaceutics*. Taylor & Francis, London.

Wechsler, J. (2001, Jan.). Clinical trial safety and oversight top policy agenda. *Appl. Clin. Trials*, pp. 18–21. Available: www.actmagazine.com/articles/act/act0101_jill.pdf.

Willman, D. (2000). Quickened pace of drug approvals by FDA taking toll. *San Jose Mercury News*, Dec. 28.

Wilson, D. (1998). A cautionary tale for human "guinea pigs." *News and Observer*, May 8, p. 4E.

Winek, C. L. (2000). History of excipient safety and toxicity. In M. L. Werner and L. A. Kotkoskie (Eds.), *Excipient Toxicity and Safety*. Marcel Dekker, New York, pp. 59–72.

World Medical Association (2000). Declaration of Helsinki: Ethical principles for medical research involving human subjects. Adopted by the 18th WMA General Assembly, Helsinki, Finland, June 1964; amended by the 29th WMA, 1975; 35th WMA, 1983; 41st WMA, 1989; 48th WMA, Somerset West, Republic of South Africa, Oct. 1996; and 52nd WMA, Edinburgh, Scotland, Oct. 2000. Available: www.wma.net/e/policy/17-c_e.html.

# *Postmarketing Safety Evaluation: Monitoring, Assessing, and Reporting of Adverse Drug Responses*

Once a new drug is approved, it proceeds to market either with or without postmarketing evaluation requirements. While this represents the end of a long road, it is also the start of yet another. While careful work during development (both in animals and humans) serves to provide the tools to greatly reduce the potential safety issues around a new drug, it cannot totally eliminate them. One needs only to look at Table 1 to appreciate the history of market withdrawals due to safety issues in the modern era (1961–2001) or turn to Table 2 to verify that the problem is still present in the first decade of the twenty-first century and comparable to the past.

Tracking and continuing to evaluate the safety of a therapeutic agent once it is on the market represent a complex task. Manufacturers are legally required to collect, analyze, and report such data both nationally [by the U.S. Food and Drug Administration (FDA) and by equivalent organizations in other countries, but here the emphasis will be on the U.S. situation] and internationally

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

**TABLE 1 Therapeutic Products Withdrawn from Marketplace Due to Safety Reasons in United Kingdom and/or United States, 1961–1995**

| Drug (INN) | Trade Name | Therapeutic Class | Reason(s) for Withdrawal/Suspension of Product License | Launch Date | Country | License Suspension or Market Withdrawal | Years on Market |
|---|---|---|---|---|---|---|---|
| Acetylsalicylic acid (pediatric) | Aspirin | Analgesic | Reye's syndrome | 1899 | U.K. | 1986 | 87 |
| Alclofenac | Prinalgin | Nonsteroidal anti-inflammatory drug (NSAID) | Skin and renal reactions, mutagenic metabolite | 1972 | U.K. | 1979 | 8 |
| Alphoxalone | Althesin | Anesthetic | Allergic-type reaction | 1972 | U.K. | 1984 | 12 |
| Aminoglutethimide | Elipten | Anticonvulsant | Endocrinological, | 1960 | U.S. | 1966 | 6 |
| | | | Reintroduced—Cushing's syndrome | 1978 | U.S. | | |
| Aminopyrine | — | Analgesic | Hematological | 1900 | U.S. | 1970 | 70 |
| | | | | | U.K. | 1975 | 75 |
| Astemizole | Hismanal | Antihistamine | Drug–drug interaction | 1988 | U.S. | 1999 | 11 |
| Azaribine | Triazure | Antipsoriatic | Neuropsychiatric, coagulation disorders | 1975 | U.S. | 1976 | 1 |
| Benoxaprofen | Opren | NSAID | Cholestatic jaundice | 1980 | U.K. | 1982 | 2 |
| | | | Photosensitivity | 1982 | U.S. | 19882 | <1 |
| Benziodarone | Amplivix | Uricosuric, coronary dilator | Hepatic | 1962 | U.K. | 1964 | 2 |
| Bithionol | Actamer | Anthelminthic | Dermatological | ? | U.S. | 1967 | ? |
| Bronfemic sodium | Duract | NSAID | Liver damage | 1996 | U.S. | 1998 | 3 |
| Cerivastatin sodium | Baycol | Cholesterol lowerer | Muscle weakening | 1999 | U.S. | 2001 | 2 |
| Chlormadinone | Normenon | Hormone | Animal carcinogenicity | 1965 | U.S. | 1970 | 5 |
| | | | | 1966 | U.K. | 1970 | 4 |
| Cisapride | Propulsid | Heartburn | Cardiovascular irregularities | 1993 | U.S. | 2000 | |
| Clioquinol | Enterovioform | Antidiarrheal | Neuropsychiatric | 1930 | U.K. | 1981 | 51 |
| | | | | 1930 | U.S. | 1973 | 43 |
| Centoxin | HA-1A | Hu-anti-lipid A IgM monoclonal | Gram-negative septicemia | 1990 | U.K. | 1992 | 2 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Danthron | Dorbanex | Laxative | Animal carcinogenicity | 1959 | U.S. | 1987 | 28 |
| | | | Reintroduced with restrictions | 1964 | U.K. | 1987 | 23 |
| Desensitizing vaccines | Various | Vaccine | Allergic-type reactions | ? | U.K. | 1989 | |
| Diamthazole | Asterol | Antifungal | Neuropsychiatric | ? | U.S. | 1970 | 17 |
| Dihydrostreptomycin | — | Antibiotic | Neuropsychiatric | ? | U.S. | 1970 | ? |
| Dinoprostone | Propess | Hormone | Uterine hypertonus and fetal distress | 1989 | U.K. | 1990 | <1 |
| Dipyrone | — | Analgesic | Hematological | 1930 | U.S. | 1977 | 47 |
| | | | | 1930 | U.K. | 1977 | 47 |
| Dithiazanine | — | Anthelminthic | Metabolic, cardiovascular | ? | U.S. | 1964 | |
| Domperidone (injection) | Motilium | Antiemetic | Cardiovascular, risk of overdose | 1984 | U.K. | 1986 | 2 |
| Doxylamine | Bendectin | Antihistamine | Fear of teratogenicity—dysmorphogenicity? | 1956 | U.S. | 1983 | 27 |
| Dicyclomine | Debendox | | | 1957 | U.K. | 1983 | 26 |
| Encainide | Enkaid | Antiarrhythmic | Cardiovascular, excess mortality risk | 1987 | U.S. | 1991 | 4 |
| Factor VIII | Factorate | Coagulation factor | Manufacture problem, risk of AIDS transmission | 1972 | U.K. | 1986 | 14 |
| Fenclofenac | Flenac | NSAID | Multiple—especially skin reactions | 1978 | U.K. | 1984 | 6 |
| Feprazone | Methrazone | NSAID | Multiple | 1976 | U.K. | 1984 | 8 |
| | | | | | | 1984 | |
| Flosequinan | Manoplax | Heart failure | Increased mortality, lack of long-term efficacy | 1992 | U.K. | 1993 | 9 months |
| | | | | | | 1993 | |
| Grepafloxacin | Raxar | Quinolone antibiotic | QT-interval prolongation | 1997 | U.S. | 1999 | 23 months |
| Growth hormone (natural) | Crescormon | Hormone | Manufacture problem | 1970 | U.K. | 1985 | 15 |
| | | | Creutzfeldt–Jakob disease transmission | 1970 | U.S. | 1985 | |
| Guanethidine | Ganda (high dose) | Anti-glaucoma eye drops | Ophthalmological | 1977 | U.K. | 1986 | 9 |
| Ibufenac | Dytransin | NSAID | Hepatic | 1966 | U.K. | 1968 | 2 |

**TABLE 1** ***Continued***

| Drug (INN) | Trade Name | Therapeutic Class | Reason(s) for Withdrawal/Suspension of Product License | Launch Date | Country | License Suspension or Market Withdrawal | Years on Market |
|---|---|---|---|---|---|---|---|
| Indomethacin-R | Osmosin form | NSAID | Multiple gastrointestinal—36 fatal small intestine perforations | 1982 | U.K. | 1983<br>1983 | 9 months |
| Indoprofen | Flosint | NSAID | Gastrointestinal carcinogenicity | 1982 | U.K. | 1983<br>1984 | 1<br>?2 |
| Iodinated casein strophantin | Coratose | Anorexiant | Metabolic | ? | ? | 1964 | ? |
| — | Lotanex | Irritable bowel syndrome | Ischemic colitis | 2000 | U.S. | 2000 | 9 months |
| Mebanazine | Actomol | Antidepressant | Hepatic drug interactions | 1963 | U.K. | 1975 | 12 |
| Megestrol acetate | Volidan 21 | Hormone | Carcinogenicity | 1963 | U.K. | 1970<br>1969<br>1975 | 7 |
| Methandrostenolone | Dianabol | Hormone | Endocrinological | 1960 | U.S. | 1982 | 12 |
| Methapyrilene | — | H1 antihistamine | Carcinogenicity | 1947<br>1950 | U.S.<br>U.K. | 1979<br>1979 | 31<br>29 |
| Metipranolol | Glauline | Antiglaucoma eye drops | Ophthalmological—uveitis (high dose) | ? | U.K. | 1990 | ? |
| | | | Low-dose preparation | — | — | 1991 | |
| Metofoline | Versidyne | Analgesic | Experimental toxicity | ? | | 1965 | ? |
| Midefradil | Posiccor | Calcium channel blocker | Lethal drug interaction (inhibited liver enzymes)` | 1996 | U.S. | 1998 | 3 |
| Mumps vaccine | Pariorix | Vaccine | Neuropsychiatric | 1988 | U.K. | 1992 | 4 |
| Urabe AM9 strain | — | — | Meningitis | 1988 | U.S. | 1992 | 4 |
| Neomycin (injection) | — | Antibiotic | Misuse, irrigation of open wounds | ? | U.S. | 1989 | ? |
| Nialamide | Niamid | Antidepressant (MAOI) | Hepatic<br>Drug interactions | 1959<br>1959 | U.K.<br>U.S. | 1978<br>1974 | 19<br>15 |
| Normifensin | Merital | Antidepressant | Hemolytic anemia, hepatotoxicity-fatal hepatitis | 1977 | U.K. | 1986 | 9 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Oxphenbutazone | Tanderil | NSAID | Hematological | 1960 | U.S. | 1985 | 25 |
| | | | Multiple | 1962 | U.K. | 1984 | 22 |
| Oxyphenisatin | Veripaque | Laxative | Hepatic | 1955 | U.K. | 1978 | 23 |
| | | | | 1957 | U.S. | 1972 | 15 |
| Perhexiline maleate | Pexid | — | Hepatic damage, peripheral neuropathy | 1975 | U.K. | 1985 | 10 |
| Phenacetin | — | Analgesic | Renal carcinogenicity | <1900 | U.K. | 19080 | 80 |
| | | | | 1900 | U.S. | 1983 | 83 |
| Phen-fen | Fenfuramine/ dexafluramine | Diet aid | Heart valve abnormality | 1973 | U.S. | 1997 | 24 |
| Phenformin | Insoral, Dibotin | Antidiabetic | Metabolic | 1959 | U.K. | 1982 | 23 |
| | | | | 1959 | U.S. | 1977 | 18 |
| Phenoxypropazine | Drazine | Antidepressant (MAOI) | Hepatic | 1961 | U.K. | 1966 | 5 |
| | | | Drug Interactions | | | | |
| Phenylpropanolamine | PPA | CTC ingredient | Hemorrhagic stroke | 1936 | U.S. | 2001 | 65 |
| Pituitary chorionic hormone | — | Hormone | Hypersensitivity | ? | U.K. | 1972 | ? |
| Polidexide | Secholex | Antihyperlipidemic | Experimental toxicity | 1974 | U.K. | 1975 | 1 |
| | | | Toxic impurities | | | | |
| Practolol | Eraldin | Beta blocker | Oculomucocutaneous syndrome | 1970 | U.K. | 1989 | 16 |
| | | | Deafness | | | | |
| | | | Sclerosing peritonitis | | | | |
| Prenylamine | Segontin, Synadrin | Antianginal | Cardiovascular | 1973 | U.K. | 1989 | 16 |
| Pronethalol | Alderlin | Beta blocker | Animal Carcinogenicity | 1963 | U.K. | 1965 | 2 |
| Remoxipride | Roxiam | Antipsychotic | Hematological—aplastic anemia | 1991 | U.K. | 1994 | 3 |
| Rotashield | — | Rotavirus vaccine | Bowel obstruction | 1998 | U.S. | 1999 | 1 |
| Somatropin | Crescormone | Natural growth hormone | Creutzfeldt-Jakob disease | 1973 | U.K. | 1985 | 12 |
| Sulfamethoxy-pyridazine | Lederkyn | Anti-infective | Hematological | ? | U.K. | 1986 | ? |
| | | | Dermatological | | | | |
| Supofren | Suprol | NSAID | Renal | 1986 | U.S. | 1987 | 1 |
| | | | | | | 1987 | |

**TABLE 1** ***Continued***

| Drug (INN) | Trade Name | Therapeutic Class | Reason(s) for Withdrawal/Suspension of Product License | Launch Date | Country | License Suspension or Market Withdrawal | Years on Market |
|---|---|---|---|---|---|---|---|
| Temafloxacin | Omniflox, Teflox | Anti-infective | Hepatic dysfunction, hemolytic anemia, nephrological, anaphylaxis, metabolic | 1991<br>1992 | U.K.<br>U.S. | 1992<br>1992 | 1<br>4 months |
| Terolidine | Micturin | Urinary incontinence | Cardiac arrhythmia | 1986 | U.K. | 1991 | 5 |
| Tetracycline (pediatric form) | Achromycin V | Antibiotic | Teeth discoloration | 1952 | U.S. | 1979 | 27 |
| Thalidomide | Contergan| | Sedative | Teratogenicity Phocomelia | 1956 | U.K. | 1961 | 5 |
| Thenalidine | Distaval | $H_1$ antihistamine | Hematological | | | 1961 | ? |
| Ticrynafen | Selacryn, Diflurex | Diuretic | Hepatic | 1979 | U.S. | 1980 | 1 |
| Triazolam | Halcion | Hypnotic | Neuropsychiatric—memory loss, depression | 1979<br>1983 | U.K.<br>U.S. | 1991<br>NW | 12 |
| Triparanol | MER-29 | Antihyperlipidemic | Ophthalmological | 1959 | U.S. | 1962 | 3 |
| Troglitazone | Renzulin | Type II diabetes | Liver damage | 1996 | U.S. | 2000 | 5 |
| Trovafloxacin | Trovan | Antibiotic | Liver/kidney damage | 1992 | U.S. | 1999 (use severely restricted) | 7 |
| Tryptophan | Pacitron, Optimax | Low-protein diet | Eosinophilic myalgia syndrome | 1974 | U.S. | 1989<br>1990 | 15 |
| Vitamin E | E-Ferol | Vitamin | Hematological, hepatic, renal | 1983 | U.S. | 1984 | 1 |
| Zomepirac | Zomax | NSAID | Allergic-type reactions—fatal anaphylaxis | 1980<br>1981 | U.S.<br>U.K. | 1983<br>1983 | 3<br>2 |
| Zimeldine | Zelmid | Antidepressant | Hepatotoxicity, neurological—peripheral neuropathy, Guillain–Barré syndrome | 1982 | U.K. | 1983 | 1 |

**TABLE 2 Drugs Withdrawn since 1990**

| Year | Drug | Indication/Class | Causative Side Effect |
|---|---|---|---|
| 1991 | Enkaid (4 years on market) | Antiarrhythmic | Cardiovascular (sudden cardiac death) |
| 1992 | Temafloxacin | Antibiotic | Blood and kidney |
| 1997 | Fenfuramine/ dexafluramine (combo used since 1984) (24 years on market) | Diet pill | Heart valve abnormalities |
| | Seldane (terfenadine) | Antihistamine | Ventricular arrhythmias |
| 1998 | Posicor (midefradil) (1 year on market) | $Ca^{2+}$ channel blocker | Lethal drug interactions (inhibited liver enzymes) |
| | Duract (bronfemic sodium) (early approval warnings of elevated liver enzymes) | Pain relief | Liver damage |
| 1999 | Trovan (use severely restricted) | Antibiotic | Liver/kidney damage |
| 2000 | Hismanal | Antihistamine | Drug–drug interactions |
| | Rotashield | Rotavirus vaccine | Bowel obstruction |
| | Renzulin (approved Dec. 1996) | Type II diabetes | Liver damage |
| | Propulsid | Heartburn | Cardiovascular irregularities/death |
| | Lotronex[a] | Irritable bowel syndrome | Ischemic colitis/death |
| 2001 | Phenylpropanolamine (PPA) | OTC ingredient | Hemorrhagic stroke |
| | Baychlor (Baycol) | Cholesterol reducing (statin) | Muscular weakness/death |
| 2004 | Serazone | Antidepressant | Liver failure and injury |
| | Vioxx[b] | Arthritis (COX-2 inhibitor) | Heart attack/cardiovascular (thrombosis) |
| 2005 | Tysabri | MS (multiple sclerosis) | PML (progressive multifocal leukoencephalopathy) |
| | Bextra | Arthritis (CoOX-2 inhibitor) | Skin reaction (sometimes fatal) |
| 2006 | Dolophine (methadone hydrochloride) | Treatment of moderate to severe pain | Respiratory depression and cardiac arrhythmias |
| 2007 | Zelnorm | Constipation | Cardiovascular safety |
| | Permax | Parkinson's disease | Heart valve damage |

*Note*: None were withdrawn in 2002/2003.

[a]Reintroduced to market in 2003.

[b]Voluntary withdrawal.

[by the World Health Organization (WHO)]. There are regulatory reporting systems (i.e., where the reports go directly to government agencies) and organizational reporting systems [organized around method of distribution, such as hospital pharmacies—American Society of Hospital Pharmacy (ASHP), 1995; Hunziker et al., 1977—or by product type, such as radiopharmaceuticals]. Poison control centers also monitor adverse drug reaction (ADR) cases and rates (Chyka and McEommon, 2000; Chyka, 1999). The regulatory systems for such pharmacosurveillance in the United States are MedWatch (for human drugs), Vaccine Adverse Event Reporting System (VAERS, for human vaccines) (Niu et al., 1998, 1999; Varrincchio, 1998), and the FDA-CVM system (for veterinary drugs) (Bukowski and Wartenberg, 1996; Keller et al., 1998). One key difference of the U.S. systems from those in other countries is that they are voluntary (largely a reflection of the primarily private—i.e., not national government—health care system in America).

There is no compulsion for physicians, hospitals, or individuals to report adverse events to either the manufacturer or the government (though the marketing companies for a therapeutic are required to periodically summarize all adverse events that they know of and report them to the federal government). It is widely held that it is this voluntary (or "spontaneous") aspect which limits the effectiveness of the U.S. systems (Piazza-Hepp and Kennedy, 1995; Sharpe, 1998; White and Love, 1998; Goldman, 1998; Kennedy and Goldman, 1997; Brewer and Colditz, 1999). Studies have identified factors which influence (and limit) physician use of such systems (LaCalamita, 1995; Figueras et al., 1999), and newly marketed drugs are subject to a higher rate of underreporting of ADRs than are established drugs (Martin et al., 1998). It should be kept in mind that what we are considering here are adverse effects caused by the use of a drug as intended, and not by a medication error. Medication errors are at least as serious a problem (and complex an issue) as ADRs (Antonow et al., 2000) but are beyond the scope of this volume. Both companies and the regulatory agencies must collect reports of adverse events, evaluate them, and then decide on a correct course of action (ranging from doing nothing through improving labeling, then on to restricting access and/or requiring ongoing or increased medical surveillance of patients, to withdrawing the drug from the market).

While new medications can save or mostly improve lives, once on the market side effects—harmful ones—are virtually certain to occur and be recognized.

## 1 CAUSES OF SAFETY WITHDRAWALS

It would be comforting to be able to state that the causes of postmarketing withdrawals from drugs were substantially different from those of failure of drugs in clinical trials. While the last few years (refer back to Table 2) are seemingly somewhat different from those in the past, (carcinogenicity is no

**TABLE 3 Characteristics of Drug Safety Withdrawals (1960–August 2001)**

| | Drugs | % of Total |
|---|---|---|
| **Most common classes** | | |
| NSAIDs | 16 | 13 |
| Nonnarcotic analgesics | 10 | 8 |
| Antidepressants | 9 | 7 |
| Vasodilators | 7 | 6 |
| Anorexiants | 5 | 4 |
| CNS stimulants | 5 | 4 |
| Barbiturates | 5 | 4 |
| Anesthetics | 4 | 3 |
| Antihistamines | 4 | 3 |
| Antibiotics | 3 | 2 |
| **Most common causes of withdrawal** | | |
| Hepatic toxicity | | 26 |
| Hematological toxicity (bone marrow suppression) | | 10 |
| Cardiovascular toxicity | | 6 |
| Carcinogenicity | | 6 |
| Renal toxicity | | 5 |
| Drug interactions | | 4 |
| Neurotoxicity | | 4 |
| Behavioral effects | | 4 |
| Abuse potential | | 4 |

longer necessarily a major marketed drug problem) the historic causes for the modern era (the last 40 years) are by hepatic toxicity—also the primary cause for safety failures in early clinical trials (see Table 3).

Fung et al. (2001) have done an extensive assembly and analysis of safety withdrawal data through 1999, and Ajani et al. (2000) have also analyzed factors that increase the likelihood of safety problems. Table 2 presents this author's extension of their work through the time of this writing (late in the third quarter of 2001), which changes the results but a little. It should be noted that the rank order of these two lists is different than the rank orders based on numbers of adverse events (see Holland and DeGruz, 1997). Adverse events can have a wide range of causes which may be due not to unanticipated effects of a drug but to medication error or something as mundane as discrepancies between doses recommended in *The Physician's Desk Reference* and those recommended or reported in the medical literature (Cohen, 2001).

In 2005, Schuster et al . published an anaysis stating that cardiovascular toxicity accounted for 40% of current (2000–2005) safety withdrawals and hepatotoxicity for 27% (leaving all others to share 33% of the total). Failure to identify these largely predictable causes of failures in new therapeutic entities largely reflects both a continuing lack of recognition of the actual patient populations utilizing drugs with their existing pathophysiological

**TABLE 4 Factors That Increase Patient Risk for Adverse Drug Interactions**

| Factor | Group/Disorder |
|---|---|
| Age | Neonates, elderly |
| Gender | Women |
| Genetic phenotype | Slow metabolizers |
| Chronic disease | Moderate/severe renal or hepatic impairment, CHF, cirrhosis |
| Acute illness | Pneumonia, influenza |
| Metabolic disturbances | Hypothyroidism, hypoxia |
| Multiple drug use | Elderly, HIV patients |
| Multiple prescribing physicians | Elderly |
| Use of drugs with a low therapeutic index | |
| Use of drugs that are enzyme inhibitors or inducers | |

**TABLE 5 Limitations of FDA's Current Clinical Trials**

| | |
|---|---|
| 1. Too few | Prior to approval, most drugs are administered to 2000–3000 patients. (To obtain an 80% probability of detecting an adverse drug event that occurs in one out of every 10,000 recipients, 16,000 patients must receive the drug.) |
| 2. Too simple | Premarketing trials often exclude patients with complicated medical histories or medication regimens. It is easier to demonstrate efficacy without including these complex patients. |
| 3. Too median | Most premarketing trials exclude patient populations such as pediatric, geriatric, lactating, and pregnant patients. |
| 4. Too narrow | Premarketing trials are generally intended to investigate a drug for a single indication. After release to the market, the drug may be used to treat other conditions in different populations with varying medical histories. |
| 5. Too brief | Adverse drug events that occur only with chronic use will not be detected in the relatively short clinical trial. |

*Source*: From Rogers, 1987.

characteristics (Table 4) and limitations of the currently employed clinical trial scheme (Table 5). While such efforts as mandatory assessment of safety pharmacology features will serve to improve the situation, for the foreseeable future it remains vital to ensure that our pharmacovigilance systems identify problems as soon as possible.

## 2 REGULATORY REQUIREMENTS

Regulations and guidelines concerning pharmacovigilance have been in a continuous state of change and development in recent years. A discussion of

thse should start with the understanding of the individual incident, or "case."

A case is a basic unit of drug safety surveillance. It is used to ensure, to the greatest extent possible, the safety of approved drug products that are still in use. The basic unit of all postmarketing safety submissions is the adverse drug experience (AED) case, which is an individual adverse drug experience.

The FDA has explicit requirements for reporting adverse event cases for drugs. A postmarketing adverse drug experience source can be categorized into several sample categories: clinical trial, nonclinical trial/regulatory authority, nonclinical trial/literature, and nonclinical trial/all other. This chapter only deals with spontaneous experiences—nonclinical trial adverse drug experiences reported to the industry any time after a marketed drug product achieves marketing approval from the FDA (Adams et al., 1997).

A typical case folder will contain certain types of information:

- The outside of a folder is identified by a numeric or alphanumeric code.
- There is an "initial" report.
- There is either at least one letter requesting additional information regarding the initial report or documentation reflecting the failed attempts to obtain additional information.
- There is at least one "follow-up" report.
- The spontaneous report event is categorized as serious or nonserious, expected, or unexpected.
- The source is literature, regulatory authority, or spontaneous.
- There is at least one MedWatch form or Council for International Organization of Medical Sciences (CIOMS) I form for each report.

Everything in a given case folder is present because of an FDA regulation requirement or a related company-written standard operating procedure.

The designation on the outside of the case is required to be numeric or alphanumeric rather than the name of the patient. Patient names are not permitted to be publicly disclosed in the context of a MedWatch report per 21 Code of Federal Regulations (CFR) 21.63(f). The initial report is the first reported information received by the company about an individual's adverse drug experience. There must be a "prompt" attempt to obtain follow-up information about each initial report. The attempt(s) are made per the company's written procedures. If the written safety procedures are not followed, the safety reports are not appropriately submitted, or the safety records are not appropriately kept, the FDA has the authority under Section 80 of Part 315 to withdraw the market new drug application (NDA). The follow-up report is the format for submitting additional information about an experience. Each case regards only one individual unless the experience is both temporally and clinically unrelated to a second event experienced by the same person taking the same drug product.

**TABLE 6 How Spontaneous Drug Case Is First Submitted to FDA**

| Case Source/Case Type | Report Submitted |
|---|---|
| Foreign literature/not both serious and unexpected | Not 15-day, not periodic |
| Foreign literature/serious and unexpected | 15-day |
| U.S. consumer/not both serious and unexpected | Periodic |
| U.S. consumer/serious, unexpected | 15-day |
| Foreign consumer/not both serious and unexpected | Not 15-day, not periodic |
| Foreign consumer/serious and unexpected | 15-day |
| FDA, initial/serious and unexpected | 15-day |
| FDA, initial/serious and unexpected | Periodic, not 15-day |
| International regulatory authority/serious and unexpected | 15-day |
| International regulatory authority/not serious and unexpected | Not 15-day, not periodic |

Table 6 summarizes FDA reporting requirements of spontaneous reports in terms of how the case event is first submitted to the agency. The definitions of *serious*, *unexpected*, and so on, are in 21 CFR 314.80(a) (CFR, 1994).

***The 15-Day Report versus U.S. Periodic Report*** Postmarketing adverse drug experiences are reported to a drug company by the public via regulatory authorities, literature, attorneys, consumers, and health professionals. Sometimes an adverse drug experience is reported directly to the FDA and the FDA then submits the report to the company. When the FDA sends the applicant an initial MedWatch report, the information does not have to be resubmitted to the FDA in an initial 15-day report if the information is serious and unexpected. This is because the FDA already has knowledge of the report. However, MedWatch and its information are incorporated into the next periodic report of the product. Follow-up to a report obtained from a non-FDA source would be submitted as an expedited 15-day report (and should reference the source of the initial report). When the FDA sends the applicant an initial MedWatch that is not both serious and unexpected, the applicant incorporates the information into the next periodic report, under the normal procedure for submitting follow-up information.

If an initial 15-day report was submitted and the first follow-up information reflects that the event is no longer classified as suitable for a 15-day report (never was serious and unexpected), the first follow-up report describes the change in the report classification but is a (first) follow-up 15-day report. Subsequent additional information is not submitted in the form of a 15-day report.

A periodic report contains certain information, such as the event terms submitted during the period, the dates that events of the period were submitted, an event term count by body system, and labeling changes made due to the period's adverse experiences. In addition (and prior) to being incorporated into a periodic report, 15-day reports are submitted within 15 calendar

days of the date the applicant received the data. All 15-day reports contain serious, unexpected events. Non-15-day reports are submitted periodically in FDA periodic reports.

If on a given day a serious, unexpected domestic report is received, it is first submitted on FDA form 3500A within 15 calendar days of receipt via the 15-day report and subsequently is incorporated (not in the form of FDA form 3500A) into a periodic report. If a report is received that is domestic but not both serious and unexpected, it is not submitted in a 15-day report but rather in the U.S. periodic report. A U.S. periodic report is submitted quarterly for the first three years after the date the product was approved by the FDA for marketing (21 CFR 314.80). However, the March 2001 FDA guidance allows an applicant to request a 21 CFR 314.90 waiver of the U.S. periodic reporting period and base the report not on the date of FDA marketing approval but instead on the international birth date (the first date the product was approved in the international community). The request for such a waiver should be submitted to Director, Office of Postmarketing Drug Risk Assessment, Center for Drug Evaluation and Research (CDER), FDA, 5600 Fishers Lane, HFD-400, Rockville, Maryland 20857. The request should include the product's name, the date of FDA marketing approval, and the product's approved application number. In addition, an applicant may request a 21 CFR 314.90 waiver of the 21 CFR 314.80(2)(ii) format of the periodic report submitted. If the waiver is granted, the International Conference on Harmonisation (ICH) E2C Periodic Safety Update Report format may be used, provided that the content of the Section 80(2)(ii) information that is not in the body of the ICH E2C periodic report is found in appendices, that is, certain reports from consumers that are not in the body of the ICH E2C periodic report submission. Among other things, 21 CFR 314.90 states that the applicant may request that the FDA waive any of the postmarketing requirements under 21 CFR 314.80.

ICH E2C and the FDA March 2001 draft guidance "Post-Marketing Safety Reporting for Human Drug and Biological Products Including Vaccines" are available at www.fda.gov/cder/guidance/index.htm; case requirements are accessible in 21 CFR 314.80.

The major change over the past few years has been a significant attempt to harmonize regulations under the aegis of the ICH, and this is covered in some depth below. The latest response to the ICH in its three participating regions (Europe, United States, and Japan) is also described together with an update on the current U.K. regulations. The ICH potentially offers real advantages to the pharmaceutical industry, but the process takes time and countries have adopted and implemented the guidance in slightly different ways and at different times. National regulations and guidelines are therefore bound to change in the near future as each country embraces the ICH.

The ICH of Technical Requirements for Registration of Pharmaceuticals for Human Use has brought together as equal partners the regulatory authorities of Europe, Japan, and the United States and experts from the pharmaceutical industry in these regions to discuss scientific and technical aspects of

product registration. The WHO, European Free Trade Area (EFTA), and Canada are observers, and the International Federation of Pharmaceutical Manufacturers Association (IFPMA) ensures contact with the research-based industry outside the ICH regions.

The aim of the ICH is to achieve greater harmonization in the interpretation and application of technical guidelines and requirements for product registration and reduce or eliminate duplicate testing. This should result in better use of resources and eliminate unnecessary delay in the global development and availability of new medicines while maintaining safety guards on quality, safety, and efficacy.

There are four broad areas within the ICH:

S—Safety (animal toxicology and pharmacology)
Q—Quality (pharmaceutical and analytical)
E—Efficacy (clinical)
M—Multidisciplinary topics

Timely, complete reporting of ADRs and medical device problems is essential to an effective national system of postmarketing surveillance. Pharmaceutical manufacturers are required by federal regulations to report all ADRs of which they are aware to the FDA. However, many health care professionals do not report adverse events to either the manufacturers or the FDA. To encourage and facilitate the reporting of serious adverse events, the FDA launched the MedWatch reporting program in June 1993. The MedWatch reporting form is used by health care professionals to voluntarily report ADRs and other problems with all FDA-regulated products used in medical therapy (drugs, biologicals, medical devices, and special nutritional agents). The ADRs associated with vaccine products is the only exception, since reporting of these ADRs is mandatory. The form used for vaccines is the joint FDA/Centers for Disease Control and Prevention (CDC) VAERS form. For drugs and therapeutic biologicals, the MedWatch (3500A) form replaces the 1639 reporting form.

The FDA does not require reports on every adverse event observed, as this would not be practical for reporters or the FDA because of the sheer number of adverse-event reports already being sent to the agency each year (about 130,000 in 1994). While 80–85% of these reports are submitted by the manufacturer, 10–15% are received by MedWatch directly from physicians, pharmacists, other health care professionals, and consumers. MedWatch encourages reporters to be selective by limiting their reports to events for which the outcome was serious (see the definition on page 876 in the previous chapter). This enables the FDA to focus on those events with potentially the largest public health impact. Reporters are encouraged to fill out the reporting form as completely and accurately as possible.

From 1978 through 1990, the CDC and the FDA divided the responsibility for postmarketing surveillance of vaccines in the United States. The FDA

received reports of adverse events after vaccines were administered in the private sector; events occurring after the administration of vaccines purchased with public funds were reported to the Monitoring System for Adverse Events Following Immunization.

The monitoring system was a stimulated passive surveillance system. In other words, when vaccines purchased with federal funds were administered in the public sector, "Important Information" forms were given to recipients or their parents or guardians instructing them to report any illnesses requiring medical attention that occurred within four weeks of vaccination. System coordinators at each immunization project/grantee site and the state health department completed standardized forms that were reviewed for consistency and completeness and then forwarded to the CDC for data entry and analysis.

In response to the National Childhood Vaccine Injury Act of 1988, which required health workers to report vaccine adverse events, the CDC and the FDA collaborated in 1990 to implement the VAERS to monitor the safety of vaccines in both sectors. Health care professionals and parents/caretakers are *encouraged* to report all clinically significant vaccine adverse events. Narrative diagnostic reports are reviewed and assigned standard codes using Coding Symbols for a Thesaurus of Adverse Reaction Terms. The source of the vaccines (public versus private provider) is recorded on the form.

The WHO system, created in response to the thalidomide disaster, seeks to capture worldwide adverse events and identify problems (WHO, 1975; Olsson, 1998). It is proposed that first all such gathered reports be analyzed for mortality effects and trends (Rose and Elnis, 2000) and then the identified most critical trends be evaluated.

## 3 MANAGEMENT OF ADR AND ADE DATA

In monitoring the safety of products, pharmaceutical companies need to comply with worldwide regulations as well as the primary requirement of helping doctors to prescribe safely. The intent in this chapter is not to provide a comprehensive review but to provide an insight into the methods of managing ADR data.

***Sources of Data*** There can be an enormous variation in the nature and quality of data depending upon the source, and this must be considered when the data are processed, computerized, and analyzed. Safety data may come from any of the sources described below.

***Clinical Trials*** In phase I studies, good documentation and additional investigations should be standard practice. Serious reactions are pretty unusual in these studies, which will detect only very common ADRs, in particular those

that are pharmacologically mediated (e.g., bradycardia with beta-adrenergic receptor antagonists).

Good documentation and follow-up should be possible in phase II studies, but rare reactions will not be identified due to the small numbers of patients involved. The larger numbers in phase III trials can pose problems, but these can be minimized by careful choice of investigators, good case report form design, and procedures for follow-up. Phase IV studies are designed to test the efficacy and safety of the drug in clinical practice and often share the same constraints in patient numbers as premarketing trials.

***Postmarketing Surveillance Studies*** Any surveillance of the safety of a drug after marketing is postmarketing surveillance (PMS) (now often referred to as a postauthorization safety study). In practice the distinction between phase IV studies and PMS is blurred (e.g., German drug experience studies).

***Spontaneous Reports*** Spontaneous reports are the most effective means of identifying rare, serious adverse reactions (usually idiosyncratic or type B) after marketing despite the underreporting that exists. Spontaneous reports are an unsolicited communication to a company, regulatory authority, or other organization that describes an adverse event in a patient given one or more medical products. These reports do not originate from a study or from any organized data collection scheme. Unless indicated otherwise by the reporter, all spontaneous adverse events are assumed to be possible ADRs. The quality and completeness of spontaneous reports is often inadequate. Pharmaceutical companies or regulatory authorities can only achieve good case documentation through effective data collection, detailed follow-up, and use of field workers for complex cases. The quality of spontaneous reports also varies from country to country. Some countries do not have a regulatory reporting form for ADRs. There are differences among countries in publicity of drug safety issues and drug regulations differ regarding the format, content, and submission time frames for ADR reporting.

Reports received by companies via regulatory authorities are often edited and poorly documented, but they cannot be ignored and should be handled alongside reports received directly. The FDA implemented the Medical Products Reporting Program (MedWatch) in 1993, which encourages health care providers to regard reporting as a fundamental professional and public health responsibility and submit serious adverse-event reports directly to the FDA on the FDA 3500A form. The FDA forwards these reports to the manufacturer, who is obliged to follow up with those reporting such events and submit any relevant information obtained to the FDA and other regulatory agencies worldwide as required. It is currently proposed that regulatory agencies (FDA) should take a more direct hand in these activities (Snidermann, 2000).

***Literature*** The publication of case reports in medical and scientific journals is an important primary source of information on ADRs. Many ADRs are

noted in medical and scientific journals before they become well known. For example, the association of thalidomide with birth defects was first noted in a letter to *Lancet* in 1961. The quality of ADR reports in the published literature can be variable and has been the subject of much criticism and correspondence, though guidelines have been promulgated for these (Jones, 1982).

Despite the anecdotal nature and sometimes poor documentation, publication of case reports in journals remains one of the most useful primary sources of information on ADRs. ADR reports in the literature can be identified in several different ways. Prepublication manuscripts describing a spontaneous case report or an event from a clinical trial are sometimes provided by authors to the manufacturer of the drug and the regulatory authority in that country. Pharmaceutical companies are required to be aware of the literature as to the safety of their approved therapeutic products and are assumed (by law) to be cognizant of such.

***Searching for ADRs in Literature*** With the increasing number of scientific and biomedical journals there are more sources of ADR data on many drugs. Conversely, for some drugs, particularly those recently marketed, there is a scarcity of clinical publications and frequently there is an inadequate account of the adverse reaction profile. Searching for ADRs in the literature may be assisted by online databases such as MEDLINE (Index Medicus) and EMBASE (Excerpta Medica) and secondary sources such as SEDBASE (Meyler's side effects of drugs) and ADIS online services such as REACTIONS. Many journals contain relevant information, but some specific ADR-related journals may assist in the search for information. Increasingly, the use of high-capacity storage systems such as compact disks (CD-ROM) has led to stand-alone systems for storage and search of the literature other than online systems. Integrated dictionaries have allowed the development of user-friendly information; however, due to the anecdotal nature of these reports, pharmaceutical companies should have a clear policy on how to handle them.

***Information Required for Reports*** In order to draw a conclusion about the possible relationship between a drug and an adverse event, certain minimal information elements are required. Points considered essential for literature reports have been proposed (Jones, 1982) and some journals issue guidelines or checklists for potential authors. These can be adapted as a potential checklist for information that should be included in any ADR report as follows:

- Patient demography—age, sex, body weight, height, race, pregnancy.
- Medical history—previous medical history and concurrent conditions, known allergies (including ADRs with similar drugs), previous experience with drug.
- Timing—duration of treatment with the suspect drug before the adverse event.

- Concurrent medications—details of other drugs, including formulation, dose, and duration.
- Dechallenge—action taken with the suspect drug (stopped, continued, dose reduction).
- Outcome—outcomes of the adverse events.
- Alternative causes—what other factors could have accounted for the adverse event (diet, occupations exposure) and which were excluded?
- Rechallenge—was the patient rechallenged, and if so, what was the result?
- Relevant additional data—blood levels, laboratory data, biopsy data, and, where relevant, postmortem findings.

*ADR Forms and Form Design.* Many forms are used by different organizations to collect ADR information. Most regulatory authorities have their own form (see Figure 1). Although the content of these forms is similar, little attempt has been made to standardize the design other than by the CIOMS.

In order to design the best form for their needs, users must first define what data they wish to collect and which factors are of the greatest importance. In addition, all the usual factors in form design need to be considered (e.g., size, layout, color, print type, spacing, flow of questions, boxes, language, and instructions). A pilot to test the form should be carried out before formal introduction and use.

Consideration should be given to what happens to the form once it is returned. Form design will be affected depending upon whether it is intended to serve as a direct-entry document (i.e., the data elements closely match the data entry screens) or whether a transcription document will be used.

The key factor in ADR form design is the compatibility with other forms required for output, most importantly regulatory authority forms. The FDA, for example, required ADR reports to be submitted on form FDA 3500A (Figure 1). If the pharmaceutical company does not wish to collect data on an FDA 3500A but must submit reports to the FDA, it will need to design a form that collects the same information. Adverse-event report forms generally collect the basic data elements outlined below:

- Patient demography
- Relevant medical history and allergies
- Suspect and concurrent drugs, route, indication
- Adverse events
- Treatment and management of adverse event
- Dechallenge, rechallenge, outcome
- Relevant laboratory data
- Reporter's opinion of causality
- Report source of information

Approved by FDA on 3/27

MEDWATCH

THE FDA MEDICAL PRODUCTS REPORTING PROGRAM

Mfr report #

UF/Dist report #

FDA Use Only

**A. Patient information**

1. Patient identifier — in confidence
2. Age at time of event: or_____ Date of birth:
3. Sex: ☐ female ☐ male
4. Weight: _____lbs or _____kgs

**B. Adverse event or product problem**

1. ☐ Adverse event and/or ☐ Product problem (e.g., defects/malfunctions)
2. Outcomes attributed to adverse event (check all that apply)
   - death_____ (mo/day/yr)
   - ☐ life threatening
   - ☐ hospitalization – initial or prolonged
   - ☐ disability
   - ☐ congenital anomaly
   - ☐ required interventio0n to prevent permanent impairment/damage
   - ☐ other:_____
3. Date of event (mo/day/yr)
4. Date of this report (mo/day/yr)
5. Describe event or problem
6. Relevant tests/laboratory data, including dates
7. Other relevant history, including preexisting medical conditions (e.g. allergies, race, pregnancy, smoking and alcohol use, hepatic/renal dysfunction, etc.)

**C. Suspect medications(s)**

1. Name (give labeled strength & mfr/labeler, if known) #1_____ #2
2. Dose, frequency & route used #1_____ #2
3. Therapy date (if unknown, give duration) from/to (or best estimate) #1_____ #2
4. Diagnosis for use (indication) #1_____ #2
5. Event abated after use stopped or dose reduced #1 ☐ yes ☐ no ☐ doesn't apply #2 ☐ yes ☐ no ☐ doesn't apply
6. Lot # (if known) #1_____ #2
7. Ex. date (if known) #1_____ #2
8. Event reappeared after reintroduction #1 ☐ yes ☐ no ☐ doesn't apply #2 ☐ yes ☐ no ☐ doesn't apply
9. NDC # - for product problems only (if known) #1 #2
10. Concomitant medical products and therapy dates (exclude treatment of event) NI

**G. All manufacturers**

1. Contact office – name/address (& mfring site for devices)
2. Phone number
3. Report Source (check all that apply): ☐ foreign ☐ study ☐ literature ☐ consumer ☐ health professional ☐ user facility ☐ company representative ☐ distributor ☐ other:
4. Date received by manufacturer (mo/day/yr)
5. (A)NDA#_____ IND#_____ PLA#_____ pre-1938 ☐ yes OTC product ☐ yes
6. If IND, protocol #
7. Type of report (check all that apply): ☐ 5-day ☐ 15-day ☐ 10-day ☐ periodic ☐ initial ☐ follow-up
8. Adverse event term(s)
9. Mfr. report number

**E. Initial reporter**

1. Name, address & phone #
2. Health professional? ☐ yes ☐ no
3. Occupation
4. Initial reported also sent report to FDA ☐ yes ☐ no ☐ unk

FDA

Domain Facsimile of FDA Form 3500A

**Submission of a report does not constitute an admission that medical personnel, user facility, distributor, manufacturer or product caused or contributed to the event**

***Figure 1*** *FDA form 3500A.*

The form can be printed as a postage prepaid folding envelope for domestic use to encourage a reply. The pharmaceutical company must be able to demonstrate due diligence in seeking relevant follow-up information on each adverse-event report.

Within the next two years, several key regulatory authorities including the MCA (Medicines Control Authority) and FDA will require electronic data submission by companies for both expedited and nonexpedited case reports. The compatibility between the company's and the regulatory authority's data-

bases with regard to content and format of the key data elements for transmission is a critical factor to success of these initiatives. The adoption of internationally sanctioned standards such as a dictionary of medical terms, various code lists (e.g., countries, routes, units), file formats, and periodic safety update reports is essential to enable efficient and accurate transmission. The ICH guideline E2B defines data elements for transmission of individual case safety reports. The guideline aims to standardize the data elements for all individual case safety reports regardless of source and destination and covers reports for both preapproval and postapproval periods. It also defines the minimum information for a report and the requirements for proper processing of the report. The medium for electronic submissions will be Electronic Data Interchange (EDI)–encrypted transmissions over the Internet.

***Computerization of Drug Safety Data: Data Collection and Input*** "Rubbish in, rubbish out" applies to safety data as to any other computerized data. The enforced control of terms at entry can be linked to checking of data, which should form part of the quality control procedures. Such controls should be driven by the business so that clinical trial data, free from all errors and needed for statistical analysis, will probably involve double data entry whereas single data entry is generally considered adequate for adverse-event databases used for signal generation and regulatory reporting.

Data are still generally typed into a database rather than electronically loaded from other systems. The first step of any data entry process should involve a check for duplicate cases. The need for decision making at the data entry stage will depend upon the type of database design. In all cases, there should be clear rules on how data should be entered into each field to ensure consistency and aid subsequent searching and outputting. This is particularly important when there are multiple users distributed over a number of international sites. Use of electronically available field-specific lists of value and well-defined coding conventions will help with this.

In the future, data will increasingly be captured electronically. Image processing and developments in optical character recognition are already proving useful. Electronic data capture (using fax or pen-based methods) is used to collect data in some clinical trial.

With the increase in licensing agreements between pharmaceutical companies, safety data frequently need to be exchanged between one or more parties. If the case volume is sufficient, it is worth considering electronic data exchange between the databases involved. In addition to preventing rekeying of data, this minimizes discrepancies between the data sets. With the adoption of proposed ICH standards in the future, this will become a much simpler process.

***Medical and Drug Terminology*** Medical and drug terminology is at the heart of the ADR systems. Accurate and consistent input of terms is critical for retrieval and analysis of ADR information. An integrated dictionary allows the capture of original text, which is autoencoded against the dictionary

to retrieve the correct code for that piece of text. Coded information allows easy retrieval and analysis. The dictionary structure should allow different ways of grouping and analyzing data encompassing body systems at the highest level. Specific reporter's wording should meet the following needs:

- Acceptable to all users of the system.
- New terms can be easily added.
- Specificity of the reported term preserved.
- Hierarchical structure to group terms at various levels of specificity.
- Logical groupings so similar terms are not scattered.
- A default grouping for each term.
- Unambiguous to enable autoencoding on input.

***Dictionaries*** This section compares commonly used dictionaries in monitoring drug safety. As electronic exchange of ADR data between industry and regulatory authorities in different countries increases, so does the need for standardization of terminology (Benichou et al., 1991). MedDRA (Medical Dictionary for Drug Regulatory Affairs) has completed development with version 4.0 just being available and is discussed later in this chapter. Table 4 presents a summary of its structure (Brown et al., 1999; Gruchalla, 1995).

***Medical Term Coding Dictionaries*** It is logical to deal with adverse events, indications, diseases, surgeries, and procedures using one system for the following reasons:

- ADRs frequently mimic spontaneously occurring diseases; hence the same diagnosis or symptom could appear as an adverse event or disease.
- In the identification of new ADRs, it is important not to separate a possible side effect from a disease.
- Separate classifications can lead to confusion and add a layer of complexity when developing ADR systems.

Meaningful codes may or may not be needed for modern dictionaries. For example, the new Adverse Drug Reactions On-Line Information Tracking (ADROIT) dictionaries do not use meaningful codes but rely on linkage of related terms and effective text processing. Where codes are considered necessary, they should be as short as possible (Westland, 1991). Whenever a system is used for adverse events from the literature, spontaneous reports, clinical trials, or a combination of these, the needs of the users of the system will influence the selection of the dictionary.

***MedDRA (Medical Dictionary for Regulatory Activities)*** MedDRA is a medical dictionary encompassing terms relevant to pre- and postmarketing

phases of the regulatory process. It was developed by the MCA to support its information systems and has subsequently been further developed by the MCA to support its information systems and Medical Terminology Working Group. The objective is to harmonize standards for electronic submissions among regulatory authorities and between authorities and industry within and across regions. The aims of the dictionary are:

- To address pre- and postmarketing adverse-event reporting
- To cover multiple medical product areas
- To be available in multiple languages
- To be available in multiple formats and platforms
- To be well maintained

The guiding principles are:

- To build from existing terminologies to maximize compatibility
- To focus on the international community need rather than on individual countries
- To ensure worldwide use through collaboration and participation in development
- To ensure mechanisms and structures are in place for translation into many languages
- To ensure long-term maintenance

The scope of MedDRA is as follows:

- Disease
- Diagnoses

There is a dual classification for some terms (e.g., 573.1, "Hepatitis in viral diseases classified elsewhere"), but this is not extensive. The dictionaries are very comprehensive with the exception of symptoms, which tend to be scattered. They have been widely used in coding patient histories and hospital charts.

ICD-9 CM is a clinical modification of ICD-9 and offers some advantages, particularly the inclusion of synonyms, but is constrained by systems that have used the older versions of ICD-9. ICD-10 is more comprehensive than any ICD revision to date (see www.meddramsso.com/mssoweb). It extends well beyond the traditional causes of death and causes of hospitalization. The content has been expanded to include symptoms, signs, abnormal findings, factors related to lifestyle, and other factors causing contact with health services:

**TABLE 7 MedDRA Structure**

| Level of Hierarchy | Approximate Number of Terms | Definition | Example |
|---|---|---|---|
| System organ class (SOC) | 26 | Broadest collection of concepts for retrieval; grouped by anatomy or physiology | Cardiac disorders |
| High-level group term | 333 | Broad concepts for linking clinically related terms; can be linked to one or more SOCs | Cardiac rhythm disorders |
| High-level term | 1,685 | Groups of preferred terms related by anatomy, pathology, physiology, etiology, or function; can be linked to one or more high-level group terms or SOCs | Tachyarrhythmia |
| Preferred term | 14,287 | International level of information exchange; single, unambiguous clinical concept | Ventricular tachycardia |
| Lowest level term | 51,083 | Synonyms and quasi-synonyms; help define scope of preferred terms | Paroxysmal ventricular tachycardia |

- Signs and symptoms
- Therapeutic indications
- Investigation names and qualitative results
- Medical and surgical procedures
- Medical, social, and family history
- Terms from COSTART (Coding Symbols for a Thesaurus of Adverse Reaction Times), WHO-ART (World Health Organization Adverse Reaction Terminology), ICD-9, HARTS, J-ART

The current structure of MedDRA is defined in Table 7. There will be a central maintenance organization responsible for development, user support, implementation, and communication as well as an international user group. A management board will oversee the activities of the central maintenance organization with direction provided by the ICH steering committee. A standard medical dictionary will facilitate electronic data exchange between industry and regulatory authorities worldwide, as recommended by the ICH.

Table 8 presents current annual MedDRA fees.

***FDA*** Under the March 2001 draft guidance "Postmarketing Safety Reporting for Human Drug and Biological Products Including Vaccines" (FDA, 2001), the FDA will accept SAEs coded with MedDRA, COSTART, or WHO ART MedDRA has been implemented for SAE coding in FDA's Adverse Event Reporting System (AERS) program. While the FDA encourages

**TABLE 8 MedDRA Annual Fees**

| | |
|---|---|
| **Basic service**[a] | $3,000/year |
| System developer (no change requests) | $5,000/year |
| **Core service**[b] | |
| Under $10 million in annual revenue | $7,000/year |
| $10–500 million in annual revenue | $12,000/year |
| $500 million to $1 billion in annual revenue | $23,000/year |
| $1–5 billion in annual revenue | $62,000/year |
| More than $5 billion in annual revenue | $82,000/year |

[a]Reserved for nonprofit medical libraries, educational institutions, and direct patient care providers. No change requests are available for this service.
[b]Have the ability to send in proposed changes to MSSO. An individual core subscriber can send in up to 100 proposed changes a month. Any changes beyond that will be charged $325 per change. Collectively, no more than 9000 change requests from all core subscribers can be received per year. Anything beyond that will be charged $325 per request and distributed among all core subscribers. A change must be medically valid and internally acceptable.

companies to use MedDRA, the deadline for full MedDRA implementation is still pending.

***European Union (EU)*** The European Agency for the Evaluation of Medical Products (EMEA) established January 2002 as the deadline for all electronically filed single case reports to be coded in MedDRA. All ADR reporting must be coded in MedDRA by January 2003.

***Japan*** MedDRA/J, the Japanese version of MedDRA, was officially issued on December 28, 1999, and the Ministry of Health highly recommended its use for ADR reporting beginning at the end of March 2000. However, J-ART terms are still applicable and upon submission to the ministry are begin converted to MedDRA/J terms. No firm deadline for full implementation has been issued.

MedDRA is available in English and Japanese only. The Maintenance and Support Services Organization (MSSO) is working on translations in French, Portuguese, German, Greek, and Spanish. With their annual dues, subscribers can get MedDRA in English and one other European language (when available). Japanese or additional European languages will need to be purchased separately. The German and Portuguese translations were recently submitted to MSSO for review.

The most current version of MedDRA, 11.0, was released in March 2008. The cost of the dictionary depends on the type of organization and annual revenue. An annual subscription provides a company with all versions of MedDRA released during the year. All regulators are provided MedDRA free of charge. Otherwise, costs are as follows:

***Periodic Reports*** Many regulatory authorities require detailed summary reports on groups of cases on a regular basis. The FDA requires annual progress reports for investigational compounds and periodic reports for marketed

drugs either quarterly or annually depending upon the length of time the product has been on the U.S. market. CIOMS II guidelines recommend submission of line listings of serious, unlabeled spontaneous cases in conjunction with a summary of the drug safety profile on a six-month basis. These reports are well defined in format, content, and submission time frame. Most major pharmaceutical companies produce them electronically.

The regulatory requirements, particularly regarding frequency of submission and content, differ in the tree regions (Europe, Japan, and the United States). In order to avoid duplication of effort and to ensure that important data are submitted with consistency to regulatory authorities worldwide, the ICH3 Topic E2C Guideline on the Format and Content for Comprehensive Periodic Safety Update Reports (PSUR, 1996) of marketed medicine products has been developed. The general principles of this guideline include:

- One report is submitted for one active substance. All dosage forms as well as indications for a given active substance should be covered in one PSUR.
- The focus is on ADRs, which include all spontaneous reports and all drug-related clinical trial and literature reports.
- An international birthdate and frequency of review and reporting are defined. The international birthdate is the date of the first marketing authorization for the product granted to any company in any country in the world. Preparation of PSURs should be based on data sets of six months or multiples thereof. The PSUR should be submitted within 60 days of the data lock point.
- The reference safety information is the company core data sheet to determine whether an ADR is listed or unlisted.
- ADR data are presented in line listings and/or summary tabulations.

## 4 CAUSALITY ASSESSMENT

Decisions have to be made by pharmaceutical companies and regulatory authorities about whether a drug can cause a particular adverse event so that an appropriate action can be taken. What does "can cause" mean? Does it imply certainty? In many cases waiting for "certainty" before taking action would entail many patients suffering unnecessarily. The degree of certainty, or "probability," required will vary according to the situation.

There are nearly always many factors other than the administration of a drug that can cause an adverse event and will determine whether the adverse event will occur in a particular patient. The drug may be "the last straw that broke the camel's back." If an adverse event would not have occurred if not for the drug, then the drug caused the adverse event (Hutchinson, 1992). So with an adverse drug interaction both drugs caused the adverse event. Using this definition the drug may only be a minor factor.

Certainty is rarely obtainable; perhaps an adverse event with a positive rechallenge where there is objective evidence and an absence of confounders in an individual case would be considered as certainty due to the drug. In the majority of cases action is needed before there is absolute certainty that a drug can cause an adverse event. This lack of certainty in individual cases has been described using rather vague terms such as "almost certain," "probably," "possible," and "unlikely." These terms have also been defined, but each author has a slightly different definition (Venulet et al., 1982; Stephens et al., 1998).

Again, in epidemiological studies or clinical trials there is nearly always a degree of uncertainty due to bias, chance, and confounders. In these studies uncertainty is measured in terms of *p* values, odd ratios, and relative risks.

The differential diagnosis of adverse events associated with a drug or drug(s) is an everyday part of a practicing clinician's life (Rogers, 1987). However, the term *causality assessment* is reserved for a similar process performed at one or more stages removed from the patient and with some important differences. Clinicians do not necessarily need to find out whether a drug caused an adverse event in order to satisfy themselves and their patients. They will be more interested in resolving the event as quickly as possible. If there is a possibility that the event might be an ADR, it may be resolved either by reducing the dose or stopping the drug or by treating the ADR while waiting for tolerance to develop or it may resolve if any of the underlying factors are altered. The resolution of the adverse event might be because the event has been caused by the drug or it may have been a transient natural occurrence; either way the patient and doctor will welcome its disappearance. If, however, the doctor is interested in knowing whether it was an ADR, further investigations can be undertaken, as long as the patient is willing, until it is established or refuted.

When causality assessment is undertaken by a regulatory authority or a scientist/physician in industry, it is unlikely that the full details known to the clinic treating the patient will be reported, even after further inquiry is made. The only way to obtain all available data is usually by visiting the physician and, with permission, reading the notes and discussing the case with him or her.

***Aims of Causality Assessment*** Of the many similar events on an adverse-event database, only a few have sufficient and relevant data to enable the assessor to decide what caused the adverse event. A preliminary assessment (sometimes referred to as a "triage") can be made by placing the event into a category (e.g., probably, possibly, or unlikely or using the EEC classification of A, B, or O) (Mezboom and Rozer, 1992). This will enable the company to extract the probable cases at regular intervals in order to consider whether there is a "signal." The possible and unlikely cases will probably not contribute much to this signal.

This preliminary assessment will need to be updated as and when further information becomes available. It should favor sensitivity over specificity so

that a borderline possible/probable case is classified as probable rather than possible to make certain that the case is not lost when at a later stage the probable cases are picked out as a signal. A full assessment when all the information is available can then rectify any misclassifications.

## 5 COURSES OF CORRECTIVE ACTION

Identification of a safety issue with a marketed drug does not necessarily (or even usually) lead to the withdrawal of that therapeutic agent from the market. As noted at the beginning of this chapter, there are a range of possible actions:

- Change in dosage or dose form (reformation)
- Change in labeling (warnings)
- Restriction of situation of use (from open prescription to either clinician administration or hospital use only)
- Monitoring of patients during use
- Restriction on use (i.e., of patients allowed to use)
- Withdrawal from the market (usually a permanent step, but not always)

Which action(s) are taken depends on severity and incidence rates of the adverse response, technical details, the existence of alternative therapies, and the benefit of the use of the drug.

## 6 LEGAL CONSEQUENCES OF SAFETY WITHDRAWAL

Although in the context of personal injury claims an HMO and other parties (e.g., doctors, hospital) may all be the target of proceedings, it is usually the pharmaceutical company, perceived as having "deep pockets," that is the prime target for claimants. Claims for negligence based upon a failure to act with reasonable care (e.g., to obtain or act upon pharmacovigilance data) and/or the supply of a product that is "defective" in legal terms (e.g., because its labeling was not amended, pursuant to the receipt and review of pharmacovigilance data so as to give adequate warnings and precautions) is always possible.

Tables 9 and 10 set out in very simple terms the necessary "ingredients" for establishing product liability, either in negligence or under statute: so-called strict liability.

All of the elements of each of these legal wrongs must be present in a given situation for liability to be established. In negligence, therefore, where the claim is made against the person alleged to owe the duty of care (in the context of this chapter, this will be the company putting the product on the market), proof of causation, without a lack of reasonable care having occurred, will not

**TABLE 9 Criteria for Negligence—D + L + F + C = N**

| | |
|---|---|
| D: Duty of care | Owed to claimant; easy to establish in case of supplier/manufacturer vis á vis patient who uses product |
| L: Lack of reasonable care | Evidenced by failure to conduct operations according to accepted standards applicable at time—that is, breach of regulatory requirements or possibly failure to take account of or apply (industry) guidelines |
| F: Foreseeable injury | Of type likely to occur following failure (e.g., side effect of drug) |
| C: Causation | Lack of reasonable care must have caused/contributed to injury; if label would not have been read by patient in any event, omission from it might not have caused injury |

**TABLE 10 Criteria for Strict Liability—D + D + C = SL**

| | |
|---|---|
| D: Defect | Widely defined—product design defect, manufacturing error (so that product is less safe than persons generally would be entitled to expect), deficiency in "presentation" |
| D: Damage | To person or property flowing from defect |
| C: Causation | See Table 7 |

afford the claimant a remedy. However, the chief distinction between negligent liability and so-called strict liability is that in the case of the latter fault is not required to be shown. To establish strict liability, the claimant must establish against the "producer" (manufacturer/importer) that the product was "defective" (for the purpose of the law, this could refer to shortcomings in its presentation, design, or manufacture) and that it caused the injury suffered.

It would not be at all unusual for claimants in personal injury actions to look for a regulatory compliance failure on the part of a company defendant. The demonstration of a regulatory breach will significantly assist the plaintiff in establishing lack of reasonable care (i.e., conduct falling below acceptable standards). In fact, whether the failure is alleged to be directly relevant to the injury or not, it can be used to demonstrate a general lack of care in the operation of corporate systems with prejudicial effect. Failure to warn is a common element of many pharmaceutical product liability cases, where the pleadings (of negligence and strict liability) might be expected to assert that had the labeling accurately dealt with contraindications, precautions, and/or warnings, the patient would have avoided the injury allegedly suffered, because either the product would not have been used/administered at all or the patient would have been monitored, advised (by the treating doctor), or managed differently so as to avert injury.

In a case where pharmacovigilance omissions are identified that can be said to lead to no/or an insufficient response being adopted by the manufacturer (especially where the regulatory authorities have taken some form of action

or simply criticized a company), the plaintiff is a significant way toward establishing a case for lack of reasonable care in negligence, or that the product was defective in strict liability terms, because it was not presented accurately and was therefore less safe than persons were entitled to expect given the content of the labeling.

### 6.1 FDA Tools for Risk Management

The Food and Drug Administration Amendments Act of 2007 (PDUFA IV) gives the FDA an array of new regulatory tools to exert tighter controls over prescription drugs in the postmarketing setting (McCaughen, 2008).

The centerpiece of the new authorities is included in the section of the law allowing the FDA to impose risk management plans on new drugs under the new acronym REMS (Risk Evaluation and Mitigation Strategies). But the spirit of the REMS lives throughout the new law, providing the FDA with new authority on direct onset of consumer ads, a proposed active surveillance system, mandatory postmarketing study provisions, and new safety labeling procedures. Fundamentally, these arise from the belief that product labeling alone is not sufficient to ensure the safe, appropriate use of prescription drugs.

This belief that top FDA officials expressed publicly for a decade grew out of the agency's frustration over a series of earlier product withdrawals like the type 2 diabetes drug troglitazone (Rezulin) and the cholesterol agent cerivastatin (Baycol).

In both those cases (and several less prominent examples), FDA officials believed that the product could be safely marketed if only prescribers followed appropriate use instructions already included in the labeling. But such labeling actions did not occur and the drugs ultimately were taken off the market for safety reason.

Such concerns gave rise to the voluntary risk management plans that serve as the models for the REMS. The amendment makes the REMS the new de facto baseline for drug approval, setting a formal mechanism for the FDA to impose the programs.

The standard for the FDA to impose the REMS is quite low: It can do so if a program is necessary to ensure that the benefits of a drug outweigh the risks. That, of course, is the standard the FDA already applies for any new drug approval decisions.

That does not mean the FDA will impose the REMS for every new drug—but rather it can. So sponsors need to anticipate that every new drug review will involve a discussion of whether the REMS is necessary, in essence forcing sponsors to work with the FDA to define a program up front or prove that one is not necessary for a new drug.

Not all REMS plans are created equal. Instead, the new law allows the FDA to impose different degrees of control depending on the profile of the product under consideration.

## 6.2 Regulatory Pyramid

The FDA's new regulatory Risk Management Evaluation and Mitigation Strategies (REMS) under FDAA (Food and Drug Administration Amendments Act, Pub. L. No. 110-185, 121 Stat 823) can be envisioned as forming a pyramid (Figure 2), with the base representing a relatively small increase in the regulatory burdens already imposed on sponsors in the postmarketing setting and the tip of the pyramid representing the strictest distribution controls that essentially require patients to be identified and screened individually before the medicine is administered.

Though it is not explicitly defined in law, each tier of the pyramid layers new regulatory controls on top of those used in the lower tiers. The result: Products regulated on the lower tiers can reach a larger share of the potential patient population for the medicine, while products regulated on higher tiers are limited to narrower and narrower slices.

Following is one approach to arranging the FDA's new regulatory authorities that may be useful in drug development planning, along with some examples of currently marketed products that already have these kinds of restrictions in place.

***Tier 1: Mandatory Studies*** The lowest tier in the new regulatory system will be products for which the FDA does not impose the REMS at all but does mandate postmarketing studies. The FDA's authority to mandate studies is independent of the new REMS provisions in FDAAA, and it is broadly crafted. The FDA can mandate studies to (1) assess a known safety risk identified prior to marketing, (2) investigate a safety signal about a product,

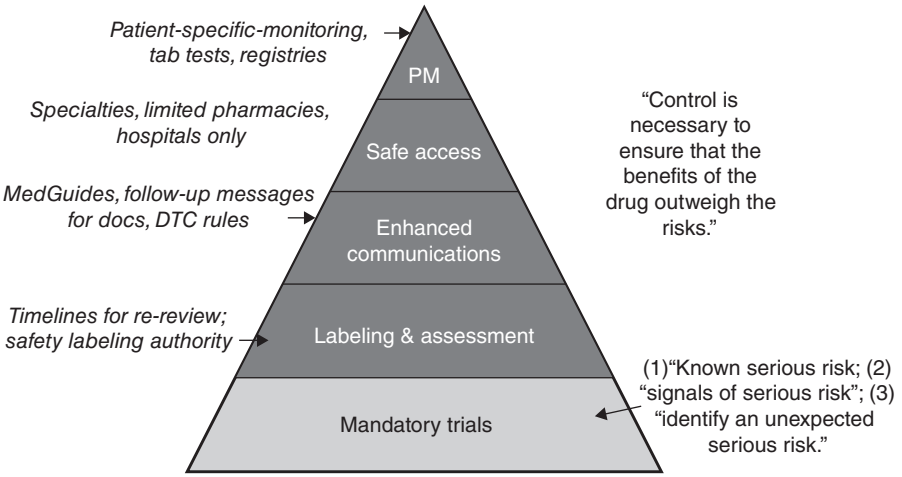


***Figure 2*** *FDA PDUFA IV risk management pyramid.*

or (3) identify "an unexpected serious risk when available data indicates the potential for a serious risk" (FDAAA).

These criteria suggest that the FDA will have the ability to argue for mandatory studies for essentially any drug.

The mandatory study authority also interacts with the separate provisions of the law directing the FDA to create an active surveillance system via a public/private partnership.

The law stipulates that the FDA is not to impose a clinical trial requirement if the active surveillance system can answer a safety question. On the other hand, the active surveillance network will be certain to generate plenty of safety signals that could serve as justification for the FDA to impose phase IV trials.

Still, products in the lowest tier of the FDA's new regulatory authority will have no special limits on their distribution or the potential access to patients. Indeed, there is nothing in the law that prevents sponsors from using studies of new indications—trials that could expand use of their medicines—to answer safety questions about existing uses.

Any product marketed with a voluntary phase IV commitment can serve as an example for this tier—in particular, products approved under the FDA's accelerated approval mechanism which face the threat of a streamlined withdrawal process if the sponsor fails to complete a postmarketing trial or if the trial fails to prove clinical benefit.

But even accelerated approval medicines are not perfect examples of the new regulatory reality, since the FDA has not so far used its enhanced power to pull a product from the market for failure to complete trials. In a pending case the FDA may be considering using that authority: The agency has written to generic drug manufacturer who market versions of Shire PLC's orthostatic hypertension therapy midrodine (ProAmatine), pointing out that the required postmarketing studies are not yet complete and inviting comments on proposals from some of the generic manufacturers to do the trials themselves.

One of the first high-profile examples of a sponsor preparing for the new reality of mandatory postmarketing commitments for already marketed products is Amgen, which will be further refining its postmarketing research program for darbepoetin (Aranesp).

***Tier 2: Labeling and Assessment*** The next tier up the pyramid is the first layer of the new REMS authority. The law says that the minimal requirement of the REMS is a periodic assessment of a product's postmarketing safety profile. In other words, everything about the drug would be the same as a non-REMS product, but there would be predefined timelines to assess how effectively the labeling is working to ensure safe use.

The FDA's separate authority to dictate labeling changes will also come into play in this tier. After 18 months on the market (the first predefined assessment point), the FDA and the sponsor would review adverse-event

reports and other new data (eventually including the active surveillance system). The new information would be added to labeling, and warning statements might be revised. But in the event of a disagreement about the interpretation of the data, the FDA now has the upper hand—the agency can force a change (subject to due process provisions in the law).

In principle, there is nothing new about the idea that the FDA will be analyzing postmarketing safety reports for new drugs. But the deadline for a formal rereview makes a profound difference in the impact of those analyses.

The FDA is already pilot testing this authority via a series of planned drug safety "report cards" for new molecular entities. The agency published its first review in September 2008, looking at spontaneous adverse-event reports for the oral chelating agent deferasirox (Exjade). The review prompted a new round of media coverage about acute renal failure and cytopenia associated with the drug; Novartis alerted prescribers to the issue in May 2008.

Going forward, sponsors will have to be cognizant of the deadlines for reassessment and wary of the FDA's ability to dictate new labeling.

At a minimum, sponsors will need to enhance their own postmarketing monitoring of products regulated under this tier to prevent any surprises in their interactions with the FDA. In some cases sponsors may decide to take voluntary steps to further restrict use if they believe the alternative will be facing an unworkable warning in labeling.

***Tier 3: Enhanced Communication*** The REMS authority allows the FDA to require consumer medication guides, enhanced communications for doctors, or other tools to communicate appropriate use information more effectively. Current examples of risk management tools that would fall under this authority might include special stickers to be affixed to prescriptions, agreements to limit sampling, enhanced patient compliance programs, or voluntary agreements not to advertise to consumers.

The FDA's new direct-to-consumer ad review user fee program and its separate authority to require disclaimers in broadcast ads also should be considered as part of this regulatory tier. Though not part of the REMS, the DTC (Direct to Consumer) provisions will work in the same way by giving the FDA a stronger hand in shaping the final version of safety messages included in broadcast ads.

The enhanced communication provisions obviously place a higher regulatory burden on the sponsor. Sponsors operating under an enhanced communication plan may find themselves in jeopardy if there is significant, measurable use of the product in a setting that the FDA considers inappropriate.

The agency will consider imposing a more restrictive REMS in that situation; it could also seek to hold the company accountable for the failure of the program. The new law makes REMS mandatory with potentially significant penalties for violations. A sponsor that willfully promotes a product in a manner that runs counter to the enhanced communication plan is obviously in jeopardy. But the FDA may also view the failure of the plan as suspicious

in itself, triggering a deeper investigation of a company's marketing activities.

There are a number of examples of products marketed with enhanced communication plans. Amylin Pharmaceuticals type 2 diabetes therapy pramlintide (Symlin) is noteworthy as an early example of a manufacturer voluntarily agreeing not to advertise directly to consumers.

Another instructive example is AstraZeneca PLC's rosuvastatin (Crestor). Developed as a "superstatin," Crestor had the misfortune to be pending at the FDA at the time of the Baycol withdrawal. In response to concerns about potential toxicity from high-potency statins, AstraZeneca agreed not to sample the high dose of Crestor. The product also underwent an unplanned safety reassessment after approval due in part to a petition filed by Public Citizen seeking Crestor's withdrawal. The result was a relabeling with stronger warnings followed by the resumption of direct-to-consumer advertising and the emergence of Crestor as a brand approaching $3 billion in annual sales.

Two recent Pfizer approvals also used enhanced communication plans. In the case of the inhaled insulin brand Exubera, the program was part of an effective regulatory strategy to resolve safety issues that long delayed its launch. (See "Pfizer's Exubera: Breathing New Life into Inhaled Insulin," In Vivo, October 2005). But getting the product to market turned out to be the easy part: After dismal uptake, Pfizer is giving up on the brands and writing off almost $3 billion as a result.

A happier example is the smoking cessation agent varenicline (Chantix). Pfizer markets the drug in conjunction with a voluntary patient support program and also delayed the launch of DTC advertising as part of its commitment to principles established by the Pharmaceutical Research & Manufacturers of America. Now the brand is one of the rising stars of the company's commercial portfolio, with sales jumping almost 10-fold to $241 million in the third quarter of 2007.

***Tier 4: Safe-Use Restriction, Defined by Provider*** The next level up the pyramid is when the REMS really start to pinch market sizes. The law gives the FDA the authority to set distribution restrictions on a drug if "necessary to assure safe use of the drug, because of its inherent toxicity or potential harmfulness" (FDAAA). The FDA is supposed to apply the restrictions only when a drug could not be marketed (or would be withdrawn) without them or if less restrictive REMS fail to assure appropriate use.

The safe-use authority is one section of the law, but it really represents two tiers of regulation.

The lower tier is a restriction based on the providers. That restriction could be to allow use only by providers who complete a special training program—a burden to the sponsor, to be sure, but one with a clear upside in promoting a new medicine.

Or the restriction could be defined by physician specialty or practice setting. The FDA could allow use of a medicine only in hospitals, for example, as

was the case for Praecis Pharmaceuticals' abarelix (Plenaxis). The restriction essentially killed the product; Praecis withdrew it and was itself acquired by GlaxoSmithKline.

In other cases, such as with Genentech omalizumab (Xolair), the distribution restriction can serve more as an opportunity. Xolair's labeling stipulates that the asthma drug must be administered in a physician's office to monitor for acute adverse reactions. The product did not initially meet Genentech's expectations but is now emerging as a prime example of the "minibuster" model that will be critical to industry in the future.

Still, restricted access provisions place a significant burden on sponsors, typically requiring specialized distribution systems and close coordination with providers. In other words, the primary care blockbuster model that built the modern big pharma business does not apply.

***Tier 5: Safe-Use Restiction, Defined by Patient*** The highest tier on the pyramid involves the other elements of a restricted distribution programs, ones that define the limits by patient-specific criteria.

The law allows the FDA to require distribution only to patients with a documented lab test or to require special monitoring of each patient given the drug or to mandate patient registries for recipients. This tier represents the tip of the regulatory pyramid, where the FDA's authority dictates the conditions for administering the drug to individual patients.

But even at this layer of regulation, commercial success is possible. Biogen Idee Tysabri is marketed under the most restrictive distribution plan approved by the FDA. The product is not at the level the sponsor hoped for when it was first launched for MS—but the company believes it could still be a billion dollar brand.

Or consider Celgene's thalidomide (Thalomid), the drug that ushered in the modern era of risk management plans when it was approved in 1998. The program has been so successful that it has essentially defined Celgene's business model—an approach that many other companies are likely to want to imitate post-FDAAA.

There is an even earlier example: Novartis's atypical antipsychotic clozapine (Clozaril), which was sold with a requirement for blood monitoring, initially tied to a single lab test provider (Caremark). Clozaril has been surpassed by many newer atypicals, but it is worth remembering that one of the biggest blockbuster classes of the past two decades began as a prototype for the era of restrictive distribution plans.

## REFERENCES

Adams, D. G., Cooper, R. M., and Kahan, J. S. (1997). *Fundamentals of Law and Regulation*, Vol. II. FDLI, Washington, DC.

Ajani, F. O., Sun, H., and Perry, J. (2000). Adverse drug reactions: A review of relevant factors. *J. Clin. Pharmacol.* 40:1093–1101.

American Society of Hospital Pharmacy (ASHP) (1995). ASHP Guidelines on adverse drug reaction monitoring and reporting. *Am. J. Health Systems Pharm.* 52:417–419.

Antonow, J. A., Smith, A. B, and Silver, M. P. (2000). Medication error reporting: A survey of nursing staff. *J. Nursing Care Quality* 15:42–48.

Benichou, C., Danan, G., and Solal-Celignz, P. (1991). Standardization of international databases: Definitions of drug-induced blood cytopenias. *Int. J. Clin. Pharmacol. Toxicol.* 29:75–81.

Brewer, T., and Colditz, G. A. (1999). Postmarketing surveillance and adverse drug reactions: Current perspectives and future needs. *JAMA* 281:824–829.

Brown, E. G., Wood, L., and Wood, S. (1999). The Medical Dictionary for Regulatory Activities (Med DRA). *Drug Safety* 20:109–117.

Bukowski, J. A., and Wartenberg, D. (1996). Comparison of adverse drug reaction reporting in veterinary and human medicine. *JAVMA* 209:40–45.

Chyka, P. A. (1999). Role of poison centers in adverse drug reaction monitoring. *Vet. Sci. Human Toxicol.* 41:400–402.

Chyka, P. A., and McEommon, S. W. (2000). Reporting of adverse drug reactions by poison control centers in the United States. *Drug Safety* 23:87–93.

Code of Federal Regulations (CFR) (1994). Postmarketing reporting of adverse drug experiences. CFR 314.80.

Cohen, J. S. (2001). Dose discrepancies between the *Physicians Desk Reference* and the medical literature and their possible roles in the high incidence of dose-related adverse drug events. *Arch. Intern. Med.* 161:957–964.

Figueras, A., Tato, F., Fontainas, J., and Gestal-Otero, J. J. (1999). Influence of physician's attitudes on reporting adverse drug events: A case-control study. *Med. Care* 37:809–814.

Fung, M., Thorton, A., Mazbeck, K., Wu, J. H., Hornbuckle, K., and Muniz, E. (2001). Evaluation of the characteristics of safety withdrawal of premarketing prescription drugs from worldwide pharmaceutical markets—1960 to 1999. *Drug Inform. J.* 35:293–317.

Goldman, S. A. (1998). Limitations and strengths of spontaneous reports data. *Clin. Ther.* 201(Suppl. C):CAO-58.

Gruchalla, R. S. (1995). A one-year perspective on MedWatch: The Food and Drug Administration's new medical products reporting program. *J. Aller. Clin. Immunol.* 95:1153–1157.

Holland, E. G., and DeGruz, F. V. (1997). Drug-induced disorders. *Am. Family Phys.* 56:1781–1788.

Hunziker, T., Kunyi, U. P., Braunschweig, S., Zehnder, D., and Hoigne, R. (1977). Comprehensive hospital drug monitoring (CHDM): Adverse skin reactions, a 20-year survey. *Allergy* 52:388–393.

Hutchinson, T. A. (1992). Causality assessments of suspected adverse drug reactions. In M. B. D. Stephens (Ed.), *Detection of New Adverse Drug Reactions*. Macmillan, London.

Jones, J. K. (1982). Criteria for journal reports of suspected adverse drug reactions. *Clin. Pharm.* 1:554–555.

Keller, W. C., Bataller, N., and Oeller, D. S. (1998). Processing and evaluation of adverse drug experience reports at the Food and Drug Administration Center for Veterinary Medicine. *JAVMA* 213:208–211.

Kennedy, D. L., and Goldman, S. A. (1997). Monitoring for adverse drug events. *Am. Family Phys.* 56:1748–1788.

LaCalamita, S. (1995). Top 10 reasons for not reporting adverse drug reactions. *Hosp. Pharm.* 30:245–246.

Martin, R. M., Kapoor, K. V., Wilton, L. V., and Mann, R. O. (1998). Underreporting of suspected adverse drug reactions to newly marketed ("black triangle") drugs in general practice: Observational study. *Br. J. Med.* 317:119–120.

McCaughen, M. (2008). Pyramid schemes and pipeline dreams: Adapting to a tiered regulatory system. RTM Report, Research Technology Management.

Mezboom, R. H. B., and Rozer, R. J. (1992). Causality classification at pharmacovigilance centers in the European Community. *Pharmacoepidemiol. Drug Safety* 1:87–97.

Niu, M. T., Rhodes, P., Saline, M., Livelz, T., Davis, D. M., Black, S., Shinefield, H., Chen, R. T., Ellenberg, S. S., and the VAERS and VSD Working Groups (1998). Comparative safety of two recombinant hepatitis B vaccines in children: Data from the Vaccine Adverse Event Reporting System (VAERS) and Vaccine Safety Data links (VSD). *J. Clin. Epidemiol.* 51:503–510.

Niu, M. T., Saline, M. E., and Ellenberg, S. S. (1999). Neonatal deaths after hepatitis B. Vaccine: The vaccine adverse event reporting system, 1991–1998. *Arch. Pediatr. Adolesc. Med.* 153:1279–1282.

Olsson, S. (1998). The Role of the WHO Programme on International Drug Monitoring in Coordinating Worldwide Drug Safety Efforts. *Drug Safety* 19:1–10.

Piazza-Hepp, T. D., and Kennedy, D. L. (1995). Reporting of adverse events to MedWatch. *AJHSP* 52:1436–1439.

Rogers, A. S. (1987). Adverse drug events: Identification and attribution. *Drug Intell. Clin. Pharm.* 21:915–920.

Rose, J. C., and Elnis, A. S. (2000). A mortality index for postmarketing surveillance of new medications. *Am. J. Emerg. Med.* 18(2):176–179.

Schuster, D., Laggner, C., and Langer, T. (2005). Why drugs fail—A study on side effects in new chemical entities. *Curr. Pharm. Des.* 11:3545–3559.

Sharpe, R. (1998). MedWatch system comes under fire. *Wall Street J.* June 24, p. B5.

Snidermann, O. A. (2000). The need for greater involvement of regulatory agencies in assessing adverse drug reactions. *CMAJ* 162:209–210.

Stephens, M. D. B., Talbot, J. C. C., and Routledge, P. A. (1998). *Detection of New Adverse Drug Reactions*, 4th ed. Grove Dictionaries, New York.

U.S. Food and Drug Administration (FDA) (2001, Apr.). Postmarketing safety reporting for human drug and biological products including vaccines. Draft Guidance. FDA, Washington, DC.

Varrincchio, F. (1998). The vaccine adverse event reporting system. *Clin. Toxicol.* 36:765–768.

Venulet, J., Blattner, R., Von Bülow, J., and Bernecker, G. C. (1982). How good are articles on ADR? *Br. Med. J.* 284:252–254.

Westland, M. M. (1991). Coding: The mortar in the bricks of data analysis. *Drug Inform. J.* 25:197–200.

White, G. G., and Love, L. (1998). The MedWatch Program. *Clin. Toxicol.* 36:145–149.

World Health Organization (WHO) (1975). *Requirements for Adverse Drug Reaction Reporting*. WHO, Geneva.

# *Statistics in Pharmaceutical Safety Assessment*

## 1 INTRODUCTION

This chapter is intended for both practicing and student toxicologists as a practical guide to the common statistical problems encountered in drug safety assessment and the methodologies that are available to solve them. The chapter has been enriched by the inclusion of discussions of why a particular procedure or interpretation is recommended, by the clear enumeration of the assumptions that are necessary for a procedure to be valid, and by discussion of problems drawn from the actual practice of toxicology and toxicological pathology.

Studies continue to be designed and executed to generate increased amounts of data. The resulting problems of data analysis have thus become more complex and toxicology has drawn more deeply from the well of available statistical techniques. Statistics has also been very active and growing during the last 35 years—to some extent, at least, because of the parallel growth of toxicology. These simultaneous changes have led to an increasing complexity of data and, unfortunately, to the introduction of numerous confounding factors which severely limit the utility of the resulting data in all too many cases.

A major difficulty is that there is a very real necessity to understand the biological realities and implications of a problem as well as the peculiarities

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

of toxicological data before procedures are selected and employed for analysis. These characteristics include the following:

1. The need to work with a relatively small sample set of data collected from the members of a population (laboratory animals) which is not actually our population of interest (i.e., humans or a target animal population).
2. Dealing frequently with data resulting from a sample which was censored on a basis other than by the investigator's own design. By censoring, of course, we mean that not all data points were collected, as might be desired. This censoring could be the result of either a biological factor (a test animal that is dead or too debilitated to manipulate) or a logistic factor (inoperative equipment or tissue that is missed in necropsy).
3. The conditions under which our experiments are conducted are extremely varied. In pharmacology (the closest cousin to at least classical toxicology), the possible conditions of interaction of a chemical or physical agent with a person are limited to a small range of doses via a single route over a short course of treatment to a defined patient population. In toxicology, however, all these variables (dose, route, time span, and subject population) are determined by the investigator.
4. The time frames available to solve our problems are limited by practical and economic factors. This frequently means that there is not time to repeat a critical study if the first attempt fails. So a true iterative approach is not possible.

The training of most pathologists in statistics remains limited to a single introductory course which concentrates on some theoretical basics. As a result, the armementarium of statistical techniques of most toxicologists is limited and the tools that are usually present [*t* tests, chi square, analysis of variance (ANOVA), and linear regression] are neither fully developed nor well understood. It is hoped that this chapter will help change this situation.

As a point of departure toward this objective, it is essential that any analysis of study results be interpreted by a professional who firmly understands three concepts: the difference between biological significance and statistical significance, the nature and value of different types of data, and causality.

For the first concept, we should consider the four possible combinations of these two different types of significance, for which we find the relationship shown below:

| | | **Statistical Significance** | |
|---|---|---|---|
| | | No | Yes |
| Biological significance | No | Case I | Case II |
| | Yes | Case III | Case IV |

Cases I and IV give us no problems, for the answers are the same statistically and biologically. But cases II and III present problems. In case II (the "false positive"), we have a circumstance where there is a statistical significance in the measured difference between treated and control groups, but there is no true biological significance to the finding. This is not an uncommon happening, for example, in the case of clinical chemistry parameters. This is called type I error by statisticians (for an example, see Portier and Hoel, 1984), and the probability of this happening is called the $\alpha$ level. In case III (the "false negative"), we have no statistical significance, but the differences between groups are biologically/toxicologically significant. This is called type II error by statisticians, and the probability of such an error happening by random chance is called the $\beta$ level. An example of this second situation is when we see a few of a very rare tumor type in treated animals. In both of these latter cases, numerical analysis, no matter how well done, is no substitute for professional judgment. Along with this, however, one must have a feeling for the different types of data and for the value or relative merit of each. Note that the two error types interact, and in determining sample size we need to specify both $\alpha$ and $\beta$ levels. Table 1 demonstrates this interaction in the case of tumor or specific lesion incidence.

The reasons that biological and statistical significance are not identical are multiple, but a central one is certainly causality. Through our consideration of statistics, we should keep in mind that just because a treatment and a change in an observed organism are seemingly or actually associated with each other does not "prove" that the former caused the latter. Though this fact is now widely appreciated for correlation (e.g., the fact that the number of storks' nests found each year in England is correlated with the number of human births that year does not mean that storks bring babies), it is just as true in the general case of significance. Timely establishment and proof that treatment

**TABLE 1 Sample Size Required to Obtain Specified Sensitivity at $p < 0.05$**

| | | Treatment Group Incidence | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Background Tumor Incidence | *P* | 0.95 | 0.90 | 0.80 | 0.70 | 0.60 | 0.50 | 0.40 | 0.30 | 0.20 | 0.10 |
| 0.30 | 0.90 | 10 | 12 | 18 | 31 | 46 | 102 | 389 | | | |
| | 0.50 | 6 | 6 | 9 | 12 | 22 | 32 | 123 | | | |
| 0.20 | 0.90 | 8 | 10 | 12 | 18 | 30 | 42 | 88 | 320 | | |
| | 0.50 | 5 | 5 | 6 | 9 | 12 | 19 | 28 | 101 | | |
| 0.10 | 0.90 | 6 | 8 | 10 | 12 | 17 | 25 | 33 | 65 | 214 | |
| | 0.50 | 3 | 3 | 5 | 6 | 9 | 11 | 17 | 31 | 68 | |
| 0.05 | 0.90 | 5 | 6 | 8 | 10 | 13 | 18 | 25 | 35 | 76 | 464 |
| | 0.50 | 3 | 3 | 5 | 6 | 7 | 9 | 12 | 19 | 24 | 147 |
| 0.01 | 0.90 | 5 | 5 | 7 | 8 | 10 | 13 | 19 | 27 | 46 | 114 |
| | 0.50 | 3 | 3 | 5 | 5 | 6 | 8 | 10 | 13 | 25 | 56 |

*Note*: *P* = power for each comparison of treatment group with background tumor incidence.

causes an effect require an understanding of the underlying mechanism and proof of its validity. At the same time, it is important that we realize that not finding a good correlation or suitable significance associated with a treatment and an effect likewise does not prove that the two are not associated—that a treatment does not cause an effect. At best, it gives us a certain level of confidence that under the conditions of the current test these items are not associated.

These points (p. 973) are discussed in greater detail in the "assumptions" section for each method along with other common pitfalls and shortcomings associated with the method. To help in better understanding the discussion, terms frequently used throughout this book should first be considered. These are presented in Table 2.

**TABLE 2 Some Frequently Used Terms and Their General Meanings**

| Term | Meaning |
|---|---|
| 95% confidence interval | Range of values (above, below, or above and below) the sample (mean, median, mode, etc.) has a 95% chance of containing the true value of the population (mean, median, mode). Also called the fiducial limit equivalent to the $P < 0.05$. |
| Bias | Systemic error as opposed to a sampling error. For example, selection bias may occur when each member of the population does not have an equal chance of being selected for the sample. |
| Degrees of freedom | Number of independent deviations, usually abbreviated df. |
| Independent Variables | Also known as predictors or explanatory variables. |
| P value | Another name for significance level; usually 0.005. |
| Power | Effect of experimental conditions on dependent variable relative to sampling fluctuation. When the effect is maximized, the experiment is more powerful. Power can also be defined as the probability that there will not be a type II error ($1 - \beta$). Conventionally, power should be at least 0.07. |
| Random | Each individual member of the population has the same chance of being selected for the sample. |
| Robust | Having inferences or conclusions little effected by departure from assumptions. |
| Sensitivity | Number of subjects experiencing each experimental condition divided by variance of scores in sample. |
| Significance level | Probability that a difference has been erroneously declared to be significant, typically 0.005 and 0.001 corresponding to 5% and 1% chance of error. |
| Type I error (false positives) | Concluding that there is an effect when there really is not an effect. Its probability is the $\alpha$ level. |
| Type II error (false negatives) | Concluding there is a particular effect when there really is another effect. Its probability is the $\beta$ level. |

*Source*: Marriott, 1991.

Each measurement we make—each individual piece of experimental information we gather—is called a datum. However, we gather and analyze multiple pieces at one time, the resulting collection being called data.

Data are collected on the basis of their association with a treatment (intended or otherwise) as an effect (a property) that is measured in the experimental subjects of a study, such as body weights. These identifiers (i.e., treatment and effect) are termed variables. Our treatment variables (those that the researcher or nature controls and which can be directly controlled) are termed independent, while our effect variables (such as weight, life span, and number of neoplasms) are termed dependent variables—their outcome is believed to depend on the "treatment" being studied.

All the possible measures of a given set of variables in all the possible subjects that exist is termed the population for those variables. Such a population of variables cannot be truly measured—for example, one would have to obtain, treat, and measure the weights of all the Fischer- 344 rats that were, are, or ever will be studied. Instead, we deal with a representative group—a sample. If our sample of data is appropriately collected and of sufficient size, it serves to provide good estimates of the characteristics of the parent population from which it was drawn.

### 1.1 Bias and Chance

Any toxicological study aims to determine whether a treatment elicits a response. An observed difference in response between a treated and a control group need not necessarily be a result of treatment. There are, in principle, two other possible explanations—*bias*, or systematic differences other than treatment between the groups, and *chance*, or random differences. A major objective of both experimental design and analysis is to try to avoid bias. Wherever possible, treated and control groups to be compared should be alike with respect to all other factors. Where differences remain, these should be corrected for in the statistical analysis. Chance cannot be wholly excluded, since identically treated animals will not respond identically. While even the most extreme difference might in theory be due to chance, a proper statistical analysis will allow the experimenter to assess this possibility. The smaller the probability of a "false positive", the more confident the experimenter can be that the effect is real. Good experimental design improves the chance of picking up a true effect with confidence by maximizing the ratio between "signal" and "noise."

### 1.2 Hypothesis Testing and Probability (*p*) Values

A relationship of treatment to some toxicological endpoint is often stated to be *statistically significant* ($p < 0.05$). What does this really mean? A number of points have to be made. *First*, statistical significance need not necessarily imply biological importance if the endpoint under study is not relevant to the ani-

mal's well-being. *Second*, the statement will usually be based only on the data from the study in question and will not take into account prior knowledge. In some situations, for example, when one or two of a very rare tumor type are seen in treated animals, statistical significance may not be achieved, but the finding may be biologically extremely important, especially if a similar treatment was previously found to elicit a similar response. *Third*, the $p$ value does not describe the probability that a true effect of treatment exists. Rather, it describes the probability of the observed response or one more extreme occurring on the assumption that treatment actually had no effect whatsoever. A $p$ value that is not significant is consistent with a treatment having a small effect not detected with sufficient certainty in this study. *Fourth*, there are two types of $p$ values. A *one-tailed* (or *one-sided*) *p value* is the probability of getting by chance a treatment effect in a specified direction as great as or greater than that observed. A *two-tailed p value* is the probability of getting, by chance alone, a treatment difference in either direction which is as great as or greater than that observed. By convention $p$ values are assumed to be two tailed unless the contrary is stated. Where, which is unusual, one can rule out in advance the possibility of a treatment effect except in one direction, a one-tailed $p$ value should be used. Often, however, two-tailed tests are to be preferred, and it is certainly not recommended to use one-tailed tests and *not* report large differences in the other direction. In any event, it is important to make it absolutely clear whether one- or two-tailed tests have been used.

It is a great mistake, when presenting results of statistical analyses, to mark, as do some laboratories, results simply as significant or not significant at one defined probability level (usually $p < 0.05$). This poor practice does not allow the reader any real chance to judge whether or not the effect is a true one. Some statisticians present the actual $p$ value for every comparison made. While this gives precise information, it can make it difficult to assimilate results from many variables. One practice the author recommends is to mark $p$ values routinely using plus signs to indicate positive differences (and minus signs to indicate negative differences) as follows: +++, $p = 0.001$; ++, $0.001 \leq p < 0.01$; +, $0.01\ p < 0.05$; ±, $0.05 \leq p < 0.1$. This highlights significant results more clearly and also allows the reader to judge the whole range from "virtually certain treatment effect" to "some suspicion." Note that using two-tailed tests, bracketed plus signs indicate findings that would be significant at the conventional $p < 0.05$ level using one-tailed tests but are not significant at this level using two-tailed tests. This "fiducal limit" ($p < 0.05$) implies a false-positive incidence of 1 in 20 and, though now imbedded in regulation, practice, and convention, was a somewhat arbitrary choice to begin with. In interpreting $p$ values it is important to realize they are only an aid to judgment to be used in conjunction with other available information. One might validly consider a $p < 0.01$ increase as chance when it was unexpected, occurred only at a low dose level with no such effect seen at higher doses, and was evident in only one subset of the data. In contrast, a $p < 0.05$ increase might be convincing if it occurred in the

top dose and was for an endpoint one might have expected to be increased from known properties of the chemical or closely related chemicals.

### 1.3 Multiple Comparisons

When a $p$ value is stated to be <0.05, this implies that, for that particular test, the difference could have occurred by chance less than 1 time in 20. Toxicological studies frequently involve making treatment–control comparisons for large numbers of variables and, in some situations, also for various subsets of animals. Some statisticians worry that the larger the number of tests, the greater is the chance of picking up statistically significant findings that do not represent true treatment effects. For this reason, an alternative "multiple-comparison" procedure has been proposed in which, if the treatment was totally without effect, then 19 times out of 20 *all* the tests should show nonsignificance when testing at the 95% confidence level. Automatic use of this approach cannot be recommended. Not only does it make it much more difficult to pick up any real effects, but also there is something inherently unsatisfactory about a situation where the relationship between a treatment and a particular response depends arbitrarily on which other responses happened to be investigated at the same time. It is accepted that in any study involving multiple endpoints that there will inevitably be a gray area between those showing highly significant effects and those showing no significant effects, where there is a problem distinguishing chance and true effects. However, changing the methodology so that the gray areas all come up as nonsignificant can hardly be the answer.

### 1.4 Estimating Size of Effect

It should be clearly understood that a $p$ value does not give direct information about the size of any effect that has occurred. A compound may elicit an increase in response by a given amount, but whether a study finds this increase to be statistically significant will depend on the size of the study and the variability of the data. In a small study, a large and important effect may be missed, especially if the endpoint is imprecisely measured. In a large study, on the other hand, a small and unimportant effect may emerge as statistically significant.

Hypothesis testing tells us whether an observed increase can or cannot be reasonably attributed to chance, but not how large it is. Although much statistical theory relates to hypothesis testing, current trends in medical statistics are toward confidence interval estimation with differences between test and control groups expressed in the form of a best estimate coupled with the 95% confidence interval (CI). Thus, if one states that treatment increases response by an estimated 10 units (95% CI 3–17 units), this would imply that there is a 95% chance that the indicated interval includes the true difference. If the lower 95% confidence limit exceeds zero, this implies the increase is statistically significant at $p < 0.05$ using a two-tailed test. One can also calculate, for

example, 99 or 99.9% confidence limits, corresponding to testing for significance at $p < 0.01$ or $p < 0.001$.

In screening studies of standard design, the tendency has been to concentrate mainly on hypothesis testing. However, presentation of the results in the form of estimates with confidence intervals can be a useful adjunct for some analyses and is very important in studies aimed specifically at quantifying the size of an effect.

Two terms refer to the quality and reproducibility of our measurements of variables. The first, accuracy, is an expression of the closeness of a measured or computed value to its actual or "true" value in nature. The second, precision, reflects the closeness or reproducibility of a series of repeated measurements of the same quantity.

If we arrange all of our measurements of a particular variable in order as a point on an axis marked as the values of that variable and if our sample were large enough, the pattern of distribution of the data in the sample would begin to become apparent. This pattern is a representation of the frequency distribution of a given population of data—that is, of the incidence of different measurements, their central tendency, and dispersion.

The most common frequency distribution—and one that has been talked about throughout this book—is the normal (or Gaussian) distribution. The normal distribution is such that two-thirds of all values are within one standard deviation of the mean (or average value for the entire population) and 95% are within 1.96 standard deviations of the mean. Symbols used are $\mu$ for the mean and $\sigma$ for the standard deviation. Other common frequency distributions include the binomial, Poisson, and chi square.

In all areas of biological research, optimal design and appropriate interpretation of experiments require that the researcher understand both the biological and technological underpinnings of the system being studied and of the data being generated. From the point of view of the statistician, it is vitally important that the experimenter both know and be able to communicate the nature of the data and understand its limitations. One classification of data types is presented in Table 3.

The nature of the data collected is determined by three considerations: the biological source of the data (the system being studied), the instrumentational and techniques being used to make measurements, and the design of the experiment. The researcher has some degree of control over each of these—least over the biological system (he or she normally has a choice of only one of several models to study) and most over the design of the experiment or study. In fact, such choices dictate the type of data generated by a study.

Statistical methods are based on specific assumptions. Parametric statistics—those that are most familiar to the majority of scientists—have more stringent underlying assumptions than do nonparametric statistics. Among the underlying assumptions for many parametric statistical methods (such as the ANOVA) is that the data are continuous. The nature of the data

**TABLE 3 Types of Variables (Data) and Examples of Each Type**

| Classified By | | Type | Example |
|---|---|---|---|
| Scale | Continuous | Scalar | Body weight |
| | | Ranked | Severity of lesion |
| | Discontinuous | Scalar | Weeks until first observation of tumor in carcinogenicity study |
| | | Ranked | Clinical observations in animals |
| | | Attribute | Eye colors in fruit flies |
| | | Quantal | Dead/alive or present/absent |
| Frequency Distribution | | Normal | Body weights |
| | | Bimodal | Some clinical chemistry parameters |
| | | Others | Measures of time to incapacitation |

associated with a variable (as described above) imparts a "value" to that data, the value being the power of the statistical tests which can be employed.

Continuous variables are those which can at least theoretically assume any of an infinite number of values between any two fixed points (such as measurements of body weight between 2.0 and 3.0 kg). Discontinuous variables, meanwhile, are those which can have only certain fixed values with no possible intermediate values (such as counts of five and six dead animals, respectively).

Limitations on our ability to measure constrain the extent to which the real-world situation approaches the theoretical, but many of the variables studied in toxicology are in fact continuous. Examples of these are lengths, weights, concentrations, temperatures, periods of time, and percentages. For these continuous variables, we may describe the character of a sample with measures of central tendency and dispersion that we are most familiar with: the mean, denoted by the symbol $\bar{x}$ and also called the arithmetic average, and the standard deviation, SD, which is denoted by the symbol $\sigma$ and is calculated as

$$\sigma = \sqrt{\frac{\sum X^2 - \left(\sum X\right)^2 / N}{N-1}}$$

where $X$ is the individual datum and $N$ is the total number of data in the group.

Contrasted with these continuous data, however, we have discontinuous (or discrete) data, which can only assume certain fixed numerical values. In these cases our choice of statistical tools or tests is, as we will find later, more limited.

## 1.5 Functions of Statistics

Statistical methods may serve to do any combination of three possible tasks. The one we are most familiar with is hypothesis testing—that is, determining if two (or more) groups of data differ from each other at a predetermined

level of confidence. A second function is the construction and use of models which may be used to predict future outcomes of chemical–biological interactions. This is most commonly seen in linear regression or in the derivation of some form of correlation coefficient. Model fitting allows us to relate one variable (typically a treatment or "independent" variable) to another. The third function, reduction of dimensionality, continues to be less commonly utilized than the first two. This final category includes methods for reducing the number of variables in a system while only minimally reducing the amount of information, therefore making a problem easier to visualize and to understand. Examples of such techniques are factor analysis and cluster analysis. A subset of this last function, discussed later under descriptive statistics, is the reduction of raw data to single expressions of central tendency and variability (such as the mean and standard deviation).

There is also a special subset of statistical techniques which is part of both the second and third functions of statistics. This is data transformation, which includes such things as the conversion of numbers to log or probit values.

## 1.6 Descriptive Statistics

Descriptive statistics are used to summarize the general nature of a data set. As such, the parameters describing any single group of data have two components. One of these describes the location of the data, while the other gives a measure of the dispersion of the data in and about this location. Often overlooked is the fact that the choice of which parameters are used to give these pieces of information implies a particular type of distribution for the data.

Most commonly, location is described by giving the (arithmetic) mean and dispersion by giving the SD or the standard error of the mean (SEM). The calculation of the first two of these has already been described. If we again denote the total number of data in a group as $N$, then the SEM would be calculated as

$$\mathrm{SEM} = \frac{\mathrm{SD}}{\sqrt{N}}$$

The use of the mean with either the SD or SEM implies, however, that we have reason to believe that the sample of data being summarized is from a population which is at least approximately normally distributed. If this is not the case, then we should rather use a set of statistical descriptions which do not require a normal distribution. These are the median for location and the semiquartile distance for a measure of dispersion. These somewhat less familiar parameters are characterized as follows.

***Median*** When all the numbers in a group are arranged in a ranked order (i.e., from smallest to largest), the median is the middle value. If there is an odd number of values in a group, then the middle value is obvious (in the case

of 13 values, for example, the seventh largest is the median). When the number of values in the sample is even, the median is calculated as the midpoint between the ($N$/2)th and the [($N$/2) + 1]th number. For example, in the series 7, 12, 13, 19 the median value would be the midpoint between 12 and 13, which is 12.5.

The SD and SEM are related to each other but yet are quite different.

The SEM is quite a bit smaller than the SD, making it very attractive to use in reporting data. This size difference is because the SEM actually is an estimate of the error (or variability) involved in measuring the means of samples, not an estimate of the error (or variability) involved in measuring the data from which means are calculated. This is implied by the *central limit theorem*, which tells us three major things:

- The distribution of sample means will be approximately normal regardless of the distribution of values in the original population from which the samples were drawn.
- The mean value of the collection will reflect that of a normal distribution.
- The standard deviation of the collection of all possible means of samples of a given size, called the standard error of the mean, depends on both the standard deviation of the original population and the size of the sample.

The SEM should be used only when the uncertainty of the estimate of the mean is of concern, which is almost never the case in toxicology. Rather, we are concerned with an estimate of the variability of the population, for which the standard deviation is appropriate.

***Semiquartile Distance*** When all the data in a group are ranked, a quartile of the data contains one ordered quarter of the values. Typically, we are most interested in the borders of the middle two quartiles $Q_1$ and $Q_3$, which together represent the semiquartile distance and which contain the median as their center. Given that there are $N$ values in an ordered group of data, the upper limit of the $j$th quartile ($Q_j$) may be computed as being equal to the [$jN$/(1)/4th] value. Once we have used this formula to calculate the upper limits of $Q_1$ and $Q_3$, we can then compute the semiquartile distance (which is also called the quartile deviation and abbreviated as QD) with the formula QD = ($Q_3 - Q_1$)/2.

For example, for the 15-value data set 1, 2, 3, 4, 4, 5, 5, 5, 6, 6, 6, 7, 7, 8, 9, we can calculate the upper limits of $Q_1$ and $Q_3$ as

$$Q_1 = \frac{1(15+1)}{4} = \frac{16}{4} = 4 \qquad Q_3 = \frac{3(15=1)}{4} = \frac{48}{4} = 12$$

The 4th and 12th values in this data set are 4 and 7, respectively. The semiquartile distance can then be calculated as

$$QD = \frac{1}{2}(7-4) = 1.5$$

There are times when it is desired to describe the relative variability of one or more sets of data. The most common way of doing this is to compute the coefficient of variation (CV), which is calculated simply as the ratio of the standard deviation to the mean:

$$CV = \frac{SD}{\overline{X}}$$

A CV of 0.2 or 20% thus means that the standard deviation is 20% of the mean. In toxicology the CV is frequently between 20 and 50% and may at times exceed 100%.

## 2 EXPERIMENTAL DESIGN

Toxicological experiments generally have a twofold purpose. The first question is whether or not an agent results in an effect on a biological system. The second question, never far behind, is how much of an effect is present. It has become increasingly desirable that the results and conclusions of studies aimed at assessing the effects of environmental agents be as clear and unequivocal as possible. It is essential that every experiment and study yield as much information as possible and that the results of each study have the greatest possible chance of answering the questions it was conducted to address. The statistical aspects of such efforts, so far as they are aimed at structuring experiments to maximize the possibilities of success, are called experimental design.

The four basic statistical principles of experimental design are replication, randomization, concurrent ("local") control, and balance. In abbreviated form, these may be summarized as follows.

***Replication*** Any treatment must be applied to more than one experimental unit (animal, plate of cells, litter of offspring, etc.). This provides more accuracy in the measurement of a response than can be obtained from a single observation since underlying experimental errors tend to cancel each other out. It also supplies an estimate of the experimental error derived from the variability among each of the measurements taken (or "replicates"). In practice, this means that an experiment should have enough experimental units in each treatment group (i.e., a large enough $N$) so that reasonably sensitive statistical analysis of data can be performed. The estimation of sample size is addressed in detail later in this chapter.

***Randomization*** This is practiced to ensure that every treatment shall have its fair share of extreme high and extreme low values. It also serves to allow the

toxicologist to proceed as if the assumption of "independence" is valid. That is, there is not avoidable (known) systematic bias in how one obtains data.

***Concurrent Control*** Comparisons between treatments should be made to the maximum extent possible between experimental units from the same closely defined population. Therefore, animals used as a "control" group should come from the same source, lot, age, and so on, as test group animals. Except for the treatment being evaluated, test and control animals should be maintained and handled in exactly the same manner.

***Balance*** If the effect of several different factors is being evaluated simultaneously, the experiment should be laid out in such a way that the contributions of the different factors can be separately distinguished and estimated. There are several ways of accomplishing this using one of several different forms of design, as will be discussed below.

There are 10 facets of any study which may affect its ability to detect an effect of a treatment. The first 6 concern minimizing the role of chance and the last four relate to avoidance of bias.

## 2.1 Choice of Species and Strain

Ideally, the responses of interest should be rare in untreated control animals but should be reasonably readily evoked by appropriate treatments. Some species or specific strains, perhaps because of inappropriate diets (Roe, 1989), have high background tumor incidences which make increases both difficult to detect and difficult to interpret when detected.

## 2.2 Sampling

Sampling—the selection of which individual data points will be collected, whether in the form of selecting which animals to collect blood from or to remove a portion of a diet mix from for analysis—is an essential step upon which all other efforts toward a good experiment or study are based.

There are three assumptions about sampling which are common to most of the statistical analysis techniques used in toxicology: that the sample is collected without bias, that each member of a sample is collected independently of the others, and that members of a sample are collected with replacements. Precluding bias, both intentional and unintentional, means that at the time of selection of a sample to measure each portion of the population from which that selection is to be made has an equal chance of being selected. Ways of precluding bias are discussed in detail in the chapter on experimental design.

Independence means that the selection of any portion of the sample is not affected by and does not affect the selection or measurement of any other portion.

Finally, sampling with replacement means that, in theory, after each portion is selected and measured, it is returned to the total sample pool and thus has the opportunity to be selected again. This is a corollary of the assumption of independence. Violation of this assumption (which is almost always the case in toxicology and all the life sciences) does not have serious consequences if the total pool from which samples are drawn is sufficiently large (say 20 or greater) that the chance of reselecting that portion is small anyway.

There are four major types of sampling methods—random, stratified, systematic, and cluster. Random is by far the most commonly employed method in toxicology. It stresses the fulfillment of the assumption of avoiding bias. When the entire poor of possibilities is mixed or randomized (procedures for randomization are presented in a later chapter), the members of the group are selected in the order drawn from the pool.

Stratified sampling is performed by first dividing the entire pool into subsets or strata, then doing randomized sampling from each strata. This method is employed when the total pool contains subsets which are distinctly different but in which each subset contains similar members. An example is a large batch of a powdered pesticide in which it is desired to determine the nature of the particle size distribution. Larger pieces or particles are on the top while progressively smaller particles have settled lower in the container, and at the very bottom the material has been packed and compressed into aggregates. To determine a timely representative answer, proportionally sized subsets from each layer or strata should be selected, mixed, and randomly sampled. This method is used more commonly in diet studies.

In systematic sampling, a sample is taken at set intervals (such as every fifth container of reagent or taking a sample of water from a fixed sample point in a flowing stream every hour). This is most commonly employed in quality assurance or (in the clinical chemistry lab) in quality control.

In cluster sampling, the pool is already divided into numerous separate groups (such as bottles of tablets), and we select small sets of groups (such as several bottles of tablets) and then select a few members from each set. What one gets then is a cluster of measures. Again, this is a method most commonly used in quality control or in environmental studies when the effort and expense of physically collecting a small group of units is significant.

In classical toxicology studies sampling arises in a practical sense in a limited number of situations. The most common of these are as follows:

1. Selecting a subset of animals or test systems from a study to make some measurement (which either destroys or stresses the measured system or is expensive) at an interval during a study. This may include such cases as doing interim necropsies in a chronic study or collecting and analyzing blood samples from some animals during a subchronic study.
2. Analyzing inhalation chamber atmospheres to characterize aerosol distributions with a new generation system.

3. Analyzing diet in which test material has been incorporated.
4. Performing quality control on an analytical chemistry operation by having duplicate analyses performed on some materials.
5. Selecting data to audit for quality assurance purposes.

### 2.3 Dose Levels

This is a very important and controversial area. In screening studies aimed at hazard identification it is normal, in order to avoid requiring huge numbers of animals, to test at dose levels higher than those to which humans will be exposed but not so high that marked toxicity occurs. A range of doses is usually tested to guard against the possibility of a misjudgment of an appropriate high dose and that the metabolic pathways at the high doses differ markedly from those at lower doses and, perhaps, to ensure no large effects occur at dose levels in the range to be used by humans. In studies aimed more at risk estimation, more and lower doses may be tested to obtain fuller information on the shape of the dose–response curve.

### 2.4 Number of Animals

This is obviously an important determinant of the precision of the findings. The calculation of the appropriate number depends on (1) the critical difference, that is, the size of the effect it is desired to detect; (2) the false-positive rate, that is, the probability of an effect being detected when none exists (equivalent to the "$\alpha$ level" or "type I error"); (3) the false-negative rate, that is, the probability of no effect being detected when one of exactly the critical size exists (equivalent to the "$\beta$ level" or "type II error"); and (4) some measure of the variability in the material.

Tables relating numbers of animals required to obtain values of critical size $\alpha$ and $\beta$ are given in Kraemer and Thiemann (1987) and Gad (1998) and software is also available for this purpose. As a rule of thumb, to reduce the critical difference by a factor $n$ for a given $\alpha$ and $\beta$, the number of animals required will have to increased by a factor $n^2$.

### 2.5 Duration of Study

It is obviously important not to terminate the study too early for fatal conditions, which are normally strongly age related. Less obviously, going on for too long in a study can be a mistake, partly because the last few weeks or months may produce relatively few extra data at a disproportionate cost and partly because diseases of extreme old age may obscure the detection of tumors and other conditions of more interest. For nonfatal conditions, the ideal is to sacrifice the animals when the average prevalence is around 50%.

## 2.6 Stratification

To detect a treatment difference with accuracy, it is important that the groups being compared are as homogeneous as possible with respect to other known causes of the response. In particular, suppose that there is another known important cause of the response for which the animals vary, so that the animals are a mixture of hyper- and hyporesponders from this cause. If the treated group has a higher proportion of hyperresponders, it will tend to have a higher response even if treatment has no effect. Even if the proportion of hyperresponders is the same as in the controls, it will be more difficult to detect an effect of treatment because of the increased between-animal variability.

Given that this other factor is known, it will be sensible to take it into account in both the design and analysis of the study. In the design, it can be used as a "blocking factor" so that animals at each level are allocated equally (or in the correct proportion) to control and treated groups. In the analysis, the factor should be treated as a stratifying variable, with separate treatment–control comparisons made at each level, and the comparisons combined for an overall test of difference. This is discussed later, where we refer to the factorial design as one example of the more complex designs that can be used to investigate the separate effect of multiple treatments.

## 2.7 Randomization

Random allocation of animals to treatment groups is a prerequisite of good experimental design. If not carried out, one can never be sure whether treatment–control differences are due to treatment or to "confounding" by other relevant factors. The ability to randomize easily is a major advantage animal experiments have over epidemiology.

While randomization eliminates bias (as least in expectation), simple randomization of all animals may not be the optimal technique for producing a sensitive test. If there is another major source of variation (e.g., sex or batch of animals), it will be better to carry out stratified randomization (i.e., carry out separate randomizations within each level of the stratifying variable).

The need for randomization applies not only to the allocation of the animals to the treatment but also to anything that can materially affect the recorded response. The same random number that is used to apply animals to the treatment group can be used to determine cage position, order of weighing, order of bleeding for clinical chemistry, order of sacrifice at terminations, and so on.

## 2.8 Adequacy of Control Group

While historical control data can, on occasion, be useful, a properly designed study demands that a relevant concurrent control group be included with which results for the test group can be compared. The principle that like should be compared with like, apart from treatment, demands that control animals

should be randomized from the same source as treatment animals. Careful consideration should also be given to the appropriateness of the control group. Thus, in an experiment involving treatment of a compound in a solvent, it would often be inappropriate to include only an untreated control group, as any differences observed could only be attributed to the treatment–solvent combination. To determine the specific effects of the compound, a comparison group given the solvent only, by the same route of administration, would be required.

It is not always generally realized that the position of the animal in the room in which it is kept may affect an animal's response. An example is the strong relationship between incidence of retinal atrophy in albino rats and closeness to the lighting source. Systematic differences in cage position should be avoided, preferably via randomization.

We have now become accustomed to developing exhaustively detailed protocols for an experiment or study prior to its conduct. A priori selection of statistical methodology (as opposed to the post hoc approach) is as significant a portion of the process of protocol development and experimental design as any other and can measurably enhance the value of the experiment or study. Prior selection of statistical methodologies is essential for proper design of other portions of a protocol, such as the number of animals per group or the sampling intervals for body weight. Implied in such a selection is the notion that the toxicologist has both an in-depth knowledge of the area of investigation and an understanding of the general principles of experimental design, for the analysis of any set of data is dictated to a large extent by the manner in which the data are obtained.

A second concept and its understanding are essential to the design of experiments in toxicology, that of censoring. Censoring is the exclusion of measurements from certain experimental units, or indeed of the experimental units themselves, from consideration in data analysis or inclusion in the experiment at all. Censoring may occur prior to initiation of an experiment (where, in modern toxicology, this is almost always a planned procedure), during the course of an experiment (when they are almost universally unplanned, resulting from, e.g., the death of animals in a test), or after the conclusion of an experiment (when usually data are excluded because of being identified as some form of outlier).

In practice, a priori censoring in toxicology studies occurs in the assignment of experimental units (such as animals) to test groups. The most familiar example is in the common practice of assignment of test animals to acute, subacture, subchronic, and chronic studies, where the results of otherwise random assignments are evaluated for body weights of the assigned members. If the mean weights are found not to be comparable by some preestablished criterion (such as a 90% probability of difference by ANOVA), then members are reassigned (censored) to achieve comparability in terms of starting body weights. Such a procedure of animal assignment to groups is known as a *censored randomization*.

The first precise or calculable aspect of experimental design encountered is determining sufficient test and control group sizes to allow one to have an adequate level of confidence in the results of a study (i.e., in the ability of the study design with the statistical tests used to detect a true difference—or effect—when it is present). The statistical test contributes a level of power to such detection. Remember that the power of a statistical test is the probability that a test results in rejection of a hypothesis, $H_0$ say, when some other hypothesis, $H$ say, is valid. This is termed the power of the test "with respect to the (alternative) hypothesis $H$."

If there is a set of possible alternative hypotheses, the power, regarded as a function of $H$, is termed the *power function* of the test. When the alternatives are indexed by a single parameter $\theta$, simple graphical presentation is possible. If the parameter is a vector $\theta$, one can visualize a *power surface.*

If the power function is denoted by $\beta(\theta)$ and $H_0$ specifies $\theta = \theta_0$, then the value of $\beta(\Pi)$—the probability of rejecting $H_0$ when it is in fact valid—is the significance level. A test's power is greatest when the probability of a type II error is the least. Specified powers can be calculated for tests in any specific or general situation.

Some general rules to keep in mind are:

- The more stringent the significance level, the greater the necessary sample size. More subjects are needed for a 1% level test than for a 5% level test.
- Two-tailed tests require larger sample sizes than one-tailed tests. Assessing two directions at the same time requires a greater investment.
- The smaller the critical effect size, the larger the necessary sample size. Subtle effects require greater efforts.
- Any difference can be significant if the sample size is large enough.
- The larger the power required, the larger the necessary sample seize. Greater protection from failure requires greater effort. The smaller the sample size, the smaller the power, that is, the greater the chance of failure.
- The requirements and means of calculating necessary sample size depend on the desired (or practical) comparative sizes of test and control groups.

This number ($N$) can be calculated, for example, for equal-sized test and control groups using the formula

$$N = \frac{(t_1 + t_2)^2}{d^2} S$$

where $t_1$ is the one-tailed $t$ value with $N - 1$ degrees of freedom corresponding to the desired level of confidence, $t_2$ is the one-tailed $t$ value with $N - 1$ degrees of freedom corresponding to the probability that the sample size will be

adequate to achieve the desired precision, and $S$ is the sample standard deviation, derived typically from historical data and calculated as

$$S = \sqrt{\frac{1}{N-1}\sum (V_1 - V_2)^2}$$

A number of aspects of experimental design are specific to the practice of toxicology. Before we look at a suggestion for step-by-step development of experimental designs, these aspects should first be considered as follows:

1. Frequently, the data gathered from specific measurements of animal characteristics are such that there is wide variability in the data. Often, such wide variability is not present in a control or low-dose group, but in an intermediate dosage group variance inflation may occur. That is, there may be a large standard deviation associated with the measurements from this intermediate group. In the face of such a set of data, the conclusion that there is no biological effect based on a finding of no statistically significant effect might well be erroneous.

2. In designing experiments, one should keep in mind the potential effect of involuntary censoring on sample size. In other words, though a study might start with five dogs per group, this provides no margin should any die before the study is ended and blood samples are collected and analyzed. Just enough experimental units per group frequently leave too few at the end to allow meaningful statistical analysis, and allowances should be made accordingly in establishing group sizes.

3. It is certainly possible to pool the data from several identical toxicological studies. One approach to this is meta-analysis, considered in detail later in this chapter. For example, after first having performed an acute inhalation study where only three treatment group animals survived to the point at which a critical measure (such as analysis of blood samples) was performed, we would not have enough data to perform a meaningful statistical analysis. We could then repeat the protocol with new control and treatment group animals from the same source. At the end, after assuring ourselves that the two sets of data are comparable, we could combine (or pool) the data from survivors of the second study with those from the first. The costs of this approach, however, would then be both a greater degree of effort expended (than if we had performed a single study with larger groups) and increased variability in the pooled samples (decreasing the power of our statistical methods).

4. Another frequently overlooked design option in toxicology is the use of an unbalanced design—that is, of different group sizes for different levels of treatment.

There is no requirement that each group in a study (control, low dose, intermediate dose, and high dose) have an equal number of experimental units assigned to it. Indeed, there are frequently good reasons to assign more experimental units to one group than to others, and, as we shall see later in this book,

all the major statistical methodologies have provisions to adjust for such inequalities within certain limits. The two most common uses of the unbalanced design have larger groups assigned to either the highest dose, to compensate for losses due to possible deaths during the study, or the lowest dose, to give more sensitivity in detecting effects at levels close to an effect threshold, or more confidence to the assertion that no effect exists.

5. We are frequently confronted with the situation where an undesired variable is influencing our experimental results in a nonrandom fashion. Such a variable is called a confounding variable; its presence, as discussed earlier, makes the clear attribution and analysis of effects at best difficult and at worst impossible. Sometimes such confounding variables are the result of conscious design or management decisions, such as the use of different instruments, personnel, facilities, or procedures for different test groups within the same study. Occasionally, however, such confounding variables are the result of unintentional and unrecognized factors or actions, in which as it is called a lurking variable. Examples of such variables are almost always the result of standard operating procedures being violated—water not being connected to a rack of animals over a weekend, a set of racks not being cleaned as frequently as others, or a contaminated batch of feed being used.

6. Finally, some thought must be given to the clear definition of what is meant by experimental unit and concurrent control.

The experimental unit in toxicology encompasses a wide variety of possibilities. It may be cells, plates of microorganisms, individual animals, litters of animals, and so on. The importance of clearly defining the experimental unit is that the number of such units per group is the *N*, which is used in statistical calculations or analyses and critically affects such calculations. The experimental unit is the unit which receives treatments and yields a response which is measured and becomes a datum.

A true concurrent control is one that is identical in every manner with the treatment groups except for the treatment being evaluated. This means that all manipulations, including gavaging with equivalent volumes of vehicle or exposing to equivalent rates of air exchanges in an inhalation chamber, should be duplicated in control groups just as they occur in treatment groups.

The goal of the four principles of experimental design is statistical efficiency and the economizing of resources. The single most important initial step in achieving such an outcome is to clearly define the objective of the study—get a clear statement of what questions are being asked.

For the reader who wants to further explore experimental design, there are a number of more detailed texts available which include more extensive treatment of the statistical aspects of experimental design (Cochran and Cox, 1975; Diamond, 1981; Federer, 1955; Hicks, 1982; Kraemer and Thiemann, 1987; Myers, 1972).

There are four basic experimental design types used in toxicology: randomized block, latin square, factorial design, and nested design. Other designs that

are used are really combinations of these basic designs and are very rarely employed in toxicology. Before examining these four basic types, however, we must first examine the basic concept of blocking.

Blocking is, simply put, the arrangement or sorting of the members of a population (such as all of an available group of test animals) into groups based on certain characteristics which may (but are not sure to) alter an experimental outcome. Such characteristics which may cause a treatment to give a differential effect include genetic background, age, sex, and overall activity levels. The process of blocking then acts (or attempts to act) so that each experimental group (or block) is assigned its fair share of the members of each of these subgroups.

We should now recall that randomization is aimed at spreading out the effect of undetectable or unsuspected characteristics in a population of animals or some portion of this population. The merging of the two concepts of randomization and blocking leads to the first basic experimental design, the randomized block. This type of design requires that each treatment group have at least one member of each recognized group (such as age), the exact members of each block being assigned in an unbiased (or random) fashion.

The second type of experimental design assumes that we can characterize treatments (whether intended or otherwise) as belonging clearly to separate sets. In the simplest case, these categories are arranged into two sets which may be thought of as rows (for, say, source litter of test animal, with the first litter as row 1, the next as row 2, etc.) and the secondary set of categories as columns (for, say, ages of test animals, with 6–8 weeks as column 1, 8–10 weeks as column 2, etc.). Experimental units are then assigned so that each major treatment (control, low dose, intermediate dose, etc.) appears once and only once in each row and each column. If we denote our test groups as A (control), B (low), C (intermediate), and D (high), such as assignment would appear as follows:

| | Age | | | |
|---|---|---|---|---|
| Source litter | 6–8 weeks | 8–10 weeks | 10–12 weeks | 12–14 weeks |
| 1 | A | B | C | D |
| 2 | B | C | D | A |
| 3 | C | D | A | B |
| 4 | D | A | B | C |

The third type of experimental design is the factorial design, in which there are two or more clearly understood treatments, such as exposure level to test chemical, animal age, or temperature. The classical approach to this situation (and to that described under the latin square) is to hold all but one of the treatments constant and at any one time to vary just that one factor. Instead, in the factorial design all levels of a given factor are combined with all levels of every other factor in the experiment. When a change in one factor produces a different change in the response variable at one level of a factor than at

other levels of this factor, there is an interaction between these two factors which can then be analyzed as an interaction effect.

The last of the major varieties of experimental design are the nested designs, where the levels of one factor are nested within (or are subsamples of) another factor. That is, each subfactor is evaluated only within the limits of its single larger factor.

## 3 DATA RECORDING

Two distinct sources of systematic bias may occur in data recording. One is that awareness of treatment may, consciously or subconsciously, affect the values recorded by the measurer. This can be avoided by organizing data recording so that observations are made blind. The second is that there is a systematic shift in the standard of measurement with time coupled with a tendency for different groups to be measured at different time points. This is particularly important when a pathologist grades a lesion for severity and when the control and high-dose animals are read before the intermediate-dose animals. In some situations it may be necessary to reread all the slides blind and in random order to be sure that diagnostic drift is avoided (Gad and Taulbee, 1996).

Valid analysis cannot be conducted unless one can distinguish animals which were examined and did not have the relevant response and animals which were not examined. It can also be important to clearly identify why data are missing. Table 4 identifies some basic rules for effective design of data collection forms.

## 4 GENERALIZED METHODOLOGY SELECTION

One approach for the selection of appropriate techniques to employ in a particular situation is to use a decision tree method. Figure 1 is a decision tree that leads to the choice of one of three other trees to assist in technique selection, with each of the subsequent trees addressing one of the three functions of statistics that was defined earlier in this chapter. Figure 2 is for the selection of hypothesis-testing procedures, Figure 3 for modeling procedures, and Figure 4 for reduction of dimensionality procedures. For the vast majority of situations, these trees will guide the user into the choice of the proper technique. The tests and terms in these trees will be explained subsequently.

## 5 STATISTICAL ANALYSIS: GENERAL CONSIDERATIONS

### 5.1 Variables to be Analyzed

Although some pathologist still regard their discipline as providing qualitative rather than quantitative data, it is abundantly clear that pathology, when

**TABLE 4 Rules for Form Design and Preparation**

1. Forms should be used when some form of repetitive data must be collected. They may be either paper or electronic.
2. If only a few (two or three) pieces of data are to be collected, they should be entered into a notebook and not onto a form. This assumes that the few pieces are not a daily event, with the aggregate total of weeks/months/years ending up as lots of data to be pooled for analysis.
3. Forms should be self-contained but should not try to repeat the content of the SOPs or method descriptions.
4. Column headings on forms should always specify the units of measurement and other details of entries to be made. The form should be arranged so that sequential entries proceed down a page, not across. Each column should be clearly labeled with a heading that identifies what is to be entered in the column. Any fixed part of entries (such at °C) should be in the column header.
5. Columns should be arranged from left to right so that there is a logical sequential order to the contents of an entry as it is made. An example would be date/time/animal number/body weight/name of recorder. The last item for each entry should be the name or unique initials of the individual who made the data entry.
6. Standard conditions that apply to all the data elements to be recorded on a form or the columns of the form should be listed as footnotes at the bottom of the form.
7. Entries of data on the form should not use more digits than are appropriate for the precision of the data being recorded.
8. Each form should be clearly titled to indicate its purpose and use. If multiple types of forms are being used, each should have a unique title or number.
9. Before designing the form, carefully consider the purpose for which it is intended. What data will be collected, how often, with what instrument, and by whom? Each of these considerations should be reflected in some manner on the form.
10. Those things which are common/standard for all entries on the form should be stated as such once. These could include such things as instrument used, scale of measurement (°C, °F, or K), or location where the recording is made.

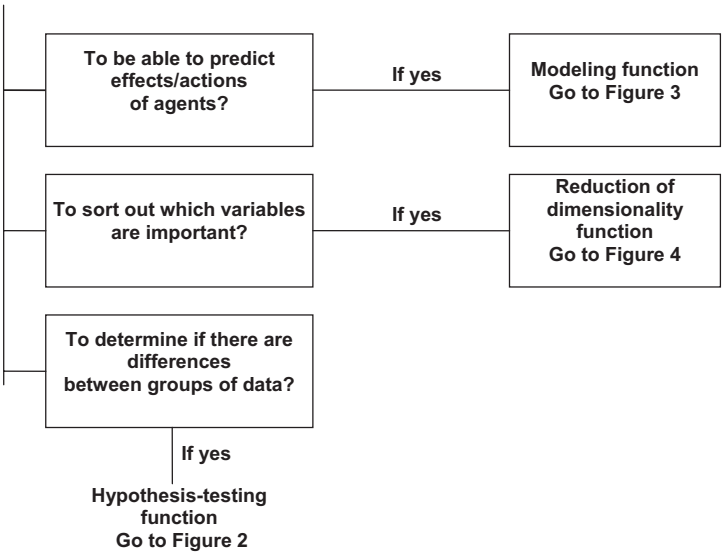


***Figure 1*** *Overall Decision Tree for Selecting Statistical Procedures.*

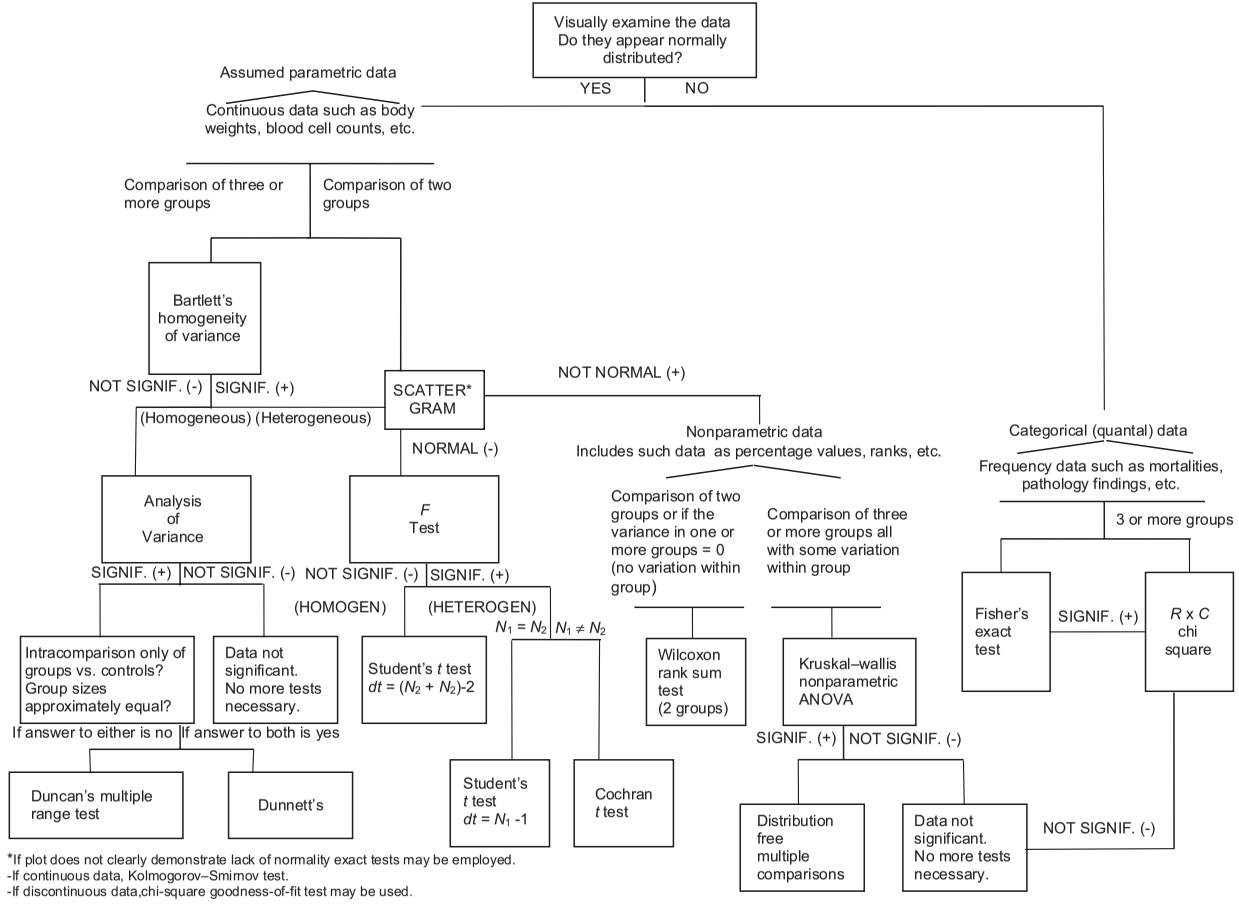


***Figure 2*** *Decision Tree for Selecting Hypothesis-testing Procedures.*

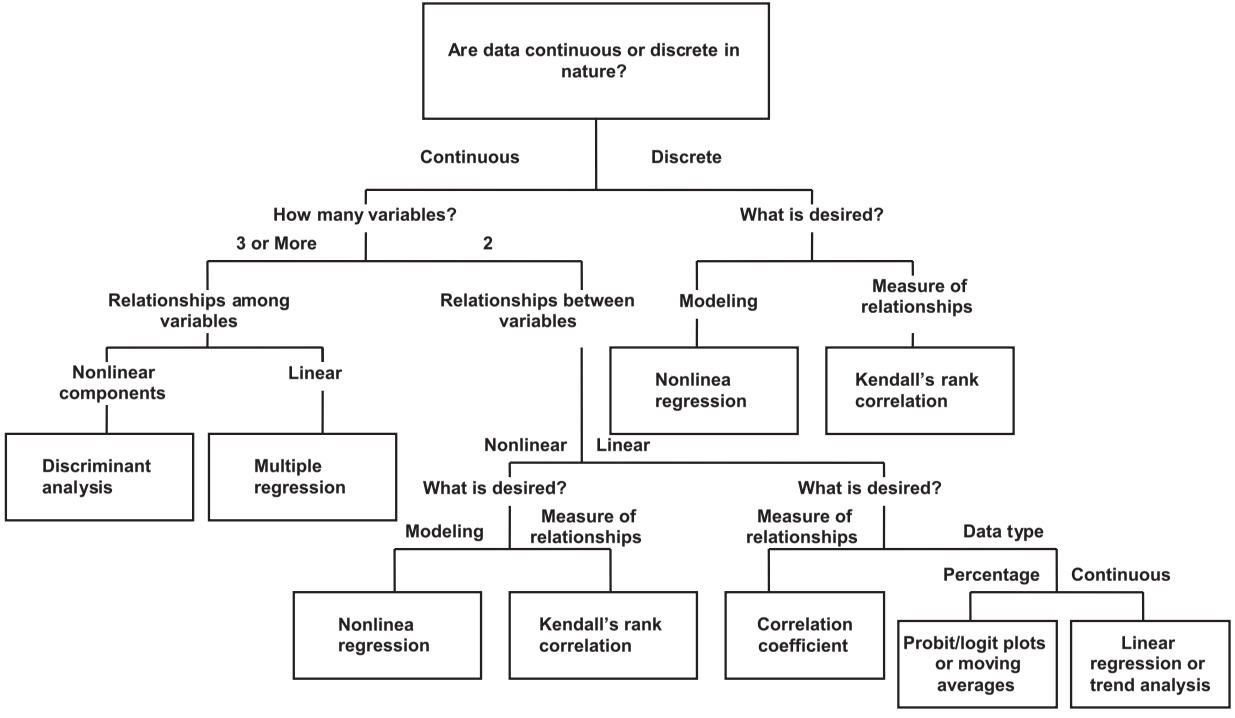


***Figure 3*** *Decision Tree for Selecting Modeling Procedures.*

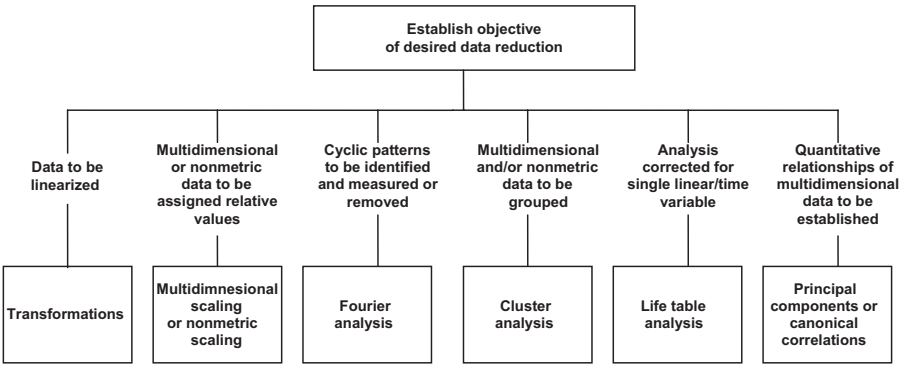


***Figure 4*** *Decision Tree for Selection of Reduction of Dimensionality Procedures.*

applied to routine screening of animal toxicity and carcinogenicity studies, has to be quantitative to at least some degree so that statistical inferences and statements can be made about possible treatment effects. Inevitably, there will be some descriptive text which will not be appropriate for statistical analysis. However, the main objective of the pathologist should be to provide information on the presence or absence (with severity grade or size where appropriate) of a list of conditions, consistently recorded from animal to animal and classified by well-defined criteria, which can be validly used in a statistical assessment.

Given that statistical analysis is worth doing and data are available that would be analyzed, should one then analyze all the endpoints recorded? Some arguments have been put forward against analyzing all the endpoints studied, none of which really holds water.

One argument is that some endpoints are not of interest. Perhaps the study is essentially a carcinogenicity study, so that nonneoplastic endpoints are not considered to be "background pathology" and almost per se unrelated to treatment. However, if the pathologist has gone to the trouble of recording the data, then, in general, they ought to be analyzed. The costs of the statistical analysis are much less than those of doing the study and the pathology. While one might justify failure to analyze nonneoplastic data where tumor analysis has already shown that the compound is clearly carcinogenic and no longer of market potential, the general rule ought to be to analyze everything that has been specifically investigated.

Another argument put forward against doing multiple analyses is that it may yield many chance significant $p$ values that have to be considered and evaluated for biological significance in the context of the entire set of available data. The whole context of dose response, as summarized in Table 5, must be kept in mind. A detailed look at the data can only aid interpretation provided one is not bound by the false argument that statistical significance necessarily equates with biological importance and definitely indicates a true effect of treatment.

**TABLE 5 Three Dimensions of Dose Response**

| As dose increases: |
|---|
| • Incidence of responders in an exposed population increases. |
| • Severity of response in effected individuals increases. |
| • Time to occurrence of response or of progressive stage of response decreases. |

Finally, some endpoints occur only very rarely. One must then be clear what "very rarely" is. For a typical study with a control and three dose groups of equal size, one would get a significant trend statistics if all three cases occurred at the top dose level or in the control group (two-tailed $p \approx 0.03$), so a total of three cases will normally be enough for statistical analysis. Endpoints occurring once or twice only are not worth analyzing formally, although, if only seen in the top dose group, they may be worth noting in the report. This is especially true if they are lesions that are rarely reported.

## 5.2 Combination of Pathological Conditions

There are four main situations where one might consider combining pathological conditions in a statistical analysis.

The first is when essentially the same pathological condition has been recorded under two or more different names or even under the same name in different places. Here failure to combine these conditions in the analysis may severely limit the chances of detecting a true treatment effect. It should be noted, however, that grouping together conditions which are actually different may also result in the masking of a true treatment effect, particularly if the treatment has a very specific effect.

The second is when separately recorded pathological conditions form successive steps on the pathway of the same process. The most important example of this is for the incidence of related types of malignant tumor, benign tumor, and focal hyperplasia. It will normally be appropriate to carry out analyses of (1) incidence of malignant tumor, (2) incidence of benign or malignant tumor, and, where appropriate, (3) incidence of focal hyperplasia and benign or malignant tumor. It will not normally be appropriate to carry out analyses of benign tumor incidence only or of the incidence of hyperplasia only.

The third situation for combining is when the same pathological condition appears in different organs as a result of the same underlying process. Examples of this are the multicentric tumors (such as myeloid leukemia, reticulum cell sarcoma, and lymphosarcoma) or certain nonneoplastic conditions (such as arteritis/periarteritis and amyloid degeneration). Here analysis will normally be carried out only of incidence at any site, although in some situations site-specific analyses might be worth carrying out.

The final situation where an analysis of combined pathological conditions is normal is for analyses of overall incidence of malignant tumor at any site, of benign or malignant tumor at any site, or of multiple tumor incidence. While

analyses of tumor incidence at specific sites are normally more meaningful, since treatments often affect only a few specific sites, these additional analyses are usually required to guard against the possibility that treatment had some weak but general tumor-enhancing effect not otherwise evident.

In some situations, one might also envisage analyses of other combinations of specific tumors, such as tumors at related sites (e.g., endocrine organs if the compound had a hormonal effect) or of similar histological type.

### 5.3 Taking Severity into Account

The same line or argument that suggests that if the pathologist records data they should be analyzed also suggests that if the pathologist chooses to grade a condition for severity the grade should be taken into account in the analysis. There are two ways to carry out analysis when the grade has to be taken into account. In one, analyses are carried out not only of whether or not the animal has a condition but also of whether or not the condition is at least gade 2, at least grade 3, and so on. In the other approach, nonparametric (rank) methods are used. The latter approach is more powerful, as it uses all the information in one analysis, although the output may not be so easily understood by those without some statistical training.

Note that the analyses based on grade can be carried out only if grading has been consistently applied throughout. If a condition has been scored only as present/absent for some animals but has been graded for others, it is not possible to carry out graded analyses unless the pathologist is willing to go back and grade the specific animals showing the condition.

### 5.4 Using Simple Methods Which Avoid Complex Assumptions

Different methods for statistical analysis can vary considerably in their complexity and in the number of assumptions they make. Although the use of statistical models has its place, more so for effect estimation than for hypothesis testing and more so in studies of complex design than in those of simple design, there are advantages in using, wherever possible, statistical methods that are simple and robust and make as few assumptions as possible. There are three reasons for this. First, such methods are more generally understandable to the toxicologist. Second, there are hardly ever extensive enough data in practice to validate any given formal model fully. Third, even if a particular model were known to be appropriate, the loss of efficiency in using appropriate simpler methods is often only very small.

The methods we advocate for routine use for the analysis of tumor incidence tend, therefore, not to be based on the use of formal parametric statistical models. For example, when studying the relationship of treatment to incidence of a pathological condition and wishing to adjust for other factors (in particular, age at death) that might otherwise bias the comparison, methods involving "stratification" are recommended rather than a multiple-regression

approach or time-to-tumor models. ANOVA methods can be useful in the case of continuously distributed data for estimating treatment effects. However, they involve underlying assumptions (normally distributed variables, variability equal in each group). If these assumptions are violated, nonparametric methods based on the rank of observations, rather than their actual value, may be preferable for hypothesis testing.

## 5.5 Using All Data

Often information is available about the relationship between treatment and a condition of interest for groups of animals differing systematically with respect to some other factor. Obvious examples are males and females, differing times of sacrifice, and differing secondary treatments. While it will be necessary, in general, to look at the relationship within levels of this other factor, it will also be advisable to try to come to some assessment of the relationship over all levels of the other factor and where a combined inference is not sensible, but in far more situations this is not the case, and using all the data in one analysis allows a more powerful test of the relationship under study. Some scientists consider that conclusions for males and females should always be drawn separately, but there are strong statistical arguments for a joint analysis.

## 5.6 Combining, Pooling, and Stratification

Suppose, in a hypothetical study of a toxic agent which induces tumors that do not shorten the lives of tumor-bearing animals, the data regarding the number of animals with tumor out of number examined are as follows:

| | Control | Exposed | Combined |
|---|---|---|---|
| Early deaths | 1/20 (5%) | 18/90 (20%) | 19/110 (17%) |
| Late deaths | 24/80 (30%) | 7/10 (70%) | 31/90 (34%) |
| Total | 25/100 (25%) | 25/100 (25%) | 50/200 (25%) |

It can be seen that if the time of death is ignored and the *pooled* data are studied, the incidence of tumors is the same in each group, resulting in the *false* conclusion that treatment had no effect. Looking within each time of death, however, an increased incidence in the exposed group can be seen. An appropriate statistical method would *combine* a measure of difference between the groups based on the early deaths and a measure of difference based on the late deaths and conclude *correctly* that incidence, after adjustment for time of death, is greater in the exposed groups.

In this example, time of death is the stratifying variable, with two strata—early deaths and late deaths. The essence of the methodology is to make comparisons only within strata (so that one is always comparing like with like except with respect to treatment) and then to combine the differences over

strata. Stratification can be used to adjust for any variable or indeed combinations of variables.

Some studies are of factorial design, in which combinations of treatments are tested. The simplest such design is one in which four equal-sized groups of animals receive (1) no treatment, (2) treatment A only, (3) treatment B only, and (4) treatments A and B. If one is prepared to assume that any effects of the two treatments are independent, one can use stratification to enable more powerful tests to be conducted of the possible individual treatment effects. Thus, to test for effects of treatment A, for example, one conducts comparisons in two strata, the first consisting of groups 1 and 2 not given treatment B and the second consisting of groups 3 and 4 given treatment B. Combination of results from the two strata is based on twice as many animals and is therefore markedly more likely to detect possible effects of treatment A than is a simple comparison of groups 1 and 2. There is also the possibility of identifying interactions, such as synergism and antagonism, between the two treatments.

In some routine long-term screening studies, the study design involved five groups of (usually) 50 animals of each sex, three of which are treated with successive doses of a compound and two of which are untreated controls. Assuming that there is no systematic difference between the control groups (e.g., the second control group in a different room or from a different batch of animals), it will be normal to carry out the main analyses with the control groups treated as a single group of 100 animals. It will usually be a sensible preliminary precaution to carry out additional analyses comparing incidences in the two control groups.

### 5.7 Trend Analysis, Low-Dose Extrapolation, and No-Observable-Effect level (NOEL) Estimation

While comparisons of individual treated groups with the control group are important, a more powerful test of a possible effect of treatment will be to carry out a test for a dose-related trend. This is because most true effects of treatment tend to result in a response which increases (or decreases) with increasing dose and because trend tests take into account all the data in a single analysis. In interpreting the results of trend tests, it should be noted that a significant trend does not necessarily imply an increased risk at lower doses. Nor, conversely, does a lack of increase at lower doses necessarily indicate evidence of a threshold (i.e., a dose below which no increase occurs).

Note that the testing for trend is seen as a more sensitive way of picking up a possible treatment effect than simple pairwise comparisons of treated and control groups. Attempting to estimate the magnitude of effects at low doses, typically below the lowest positive dose tested in the study, is a much more complex procedure and is heavils dependent on the assumed functional form of the dose–response relationship.

Deterministic trend models are based on the assumption that the trend of a time series can be approximated closely by simple mathematical functions of

time over the entire span of the series. The most common representation of a deterministic trend is by means of polynomials or of transcendental functions. The time series from which the trend is to be identified is assumed to be generated by a nonstationary process where the nonstationarity results from a deterministic trend. A classical model is the regression or error model (Anderson, 1971) where the observed series is treated as the sum of a systematic part or trend and a random or irregular part. This model can be written as

$$Z_t = Y_t + U_t'$$

where $U_t$ is a purely random process, that is, $U_t \sim$ i.i.d. $(0, 2/u)$ (independent and identically distributed with expected value zero and variance $2/u$.)

Trend tests are generally described as $k$-sample tests of the null hypothesis of identical distribution against an alternative of linear order; that is, if sample $I$ has distribution function $F_i$, $i = 1$, then the null hypothesis

$$\mathrm{H}_0 : \mathrm{F}_1 = F_2 - \cdots = F_k$$

is tested against the alternative

$$\mathrm{H}_1 : F_1 \geq F_2 \geq \cdots = F_k$$

(or its reverse), where at least one of the inequalities is strict. These tests can be thought of as special cases of tests of regression or correlation in which association is sought between the observations and its ordered sample index. They are also related to ANOVA except that the tests are tailored to be powerful against the subset of alternatives $H_1$, instead of the more general set $\{F_1 \neq F_j$, some $i \neq j\}$.

Different tests arise from requiring power against specific elements or subsets of this rather extensive set of alternatives.

The most popular trend test in toxicology is currently that presented by Tarone in 1975 because it is that used by the National Cancer Institute (NCI) in the analysis of carcinogenicity data. A simple, but efficient alternative is the Cox and Stuart (1955) test, which is a modification of the sign test. For each point at which we have a measure (such as the indidence of animals observed with tumors) we form a pair of observations—one from each of the groups we wish to compare. In a traditional NCI bioassay this would mean pairing control with low dose and low dose with high dose (to explore a dose-related trend) or each time period observation in a dose group (except the first) with its predecessor (to evaluate time-related trend). When the second observation in a pair exceeds the earlier observation, we record a plus sign for that pair. When the first observation is greater than the second, we record a minus sign for that pair. A preponderance of plus signs suggests a downward trend while an excess

of minus signs suggests an upward trend. A formal test at a preselected confidence level can then be performed.

More formally put, after having defined what trend we want to test for, we first match pairs as $(X_1 - X_{1+c}), (X_2, X_{2+c}), \ldots, (X_{n'-c}, X_{n'})$, where $c = n'/2$ when $n'$ is even and $c = (n' + 1)/2$ when $n'$ is odd (where $n'$ is the number of observations in a set). The hypothesis is then tested by comparing the resulting number of excess positive or negative signs against a sign test table such as are found in Beyer (1976a,b).

We can, of course, combine a number of observations to allow ourselves to actively test for a set of trends, such as the existence of a trend of increasing difference between two groups of animals over a period of time. This is demonstrated in Example 1.

**ASSUMPTION AND LIMITATION**

1. Trend tests seek to evaluate whether there is a monotonic tendency in response to a change in treatment. That is, the dose–response direction is absolute—as dose goes up, the incidence of tumors increases. Thus the test loses power rapidly in response to the occurrences of "reversals"—for example, a low-dose group with a decreased tumor incidence. There are methods which "smooth the bumps" of reversals in long data series. In toxicology, however, most data series are short (i.e., there are only a few dose levels).

Tarone's trend test is most powerful at detecting dose-related trends when tumor onset hazard functions are proportional to each other. For more power against other dose-related group differences, weighted versions of the statistic are also available (Breslow, 1984; Crowley and Breslow, 1984).

In 1985, the U.S. *Federal Register* recommended that the analysis of tumor incidence data is carried out with a Cochran–Armitage (Armitage, 1955; Cochran, 1954) trend test. The test statistic of the Cochran–Armitage test is defined as

$$T_{\mathrm{CA}} = \sqrt{\frac{N}{(N-r)r}, \frac{\sum_{i=0}^{k}[R_1 - (n_1/N)r]d_1}{\sum_{i=0}^{k}(n_i/N)d_i^2 - \left(\sum_{i=0}^{k}(n_i/N)d_1\right)^2}}$$

with dose scores $d_i$. Armitage's test statistic is the square of this term $(T_{\mathrm{CA}}^2)$. As one-sided tests are carried out for an increase of tumor rates, the square is not considered. Instead, the above-mentioned test statistic, which is presented by Portier and Hoel (1984), is used. This test statistic is asymptotically standard normally distributed. The Cochran–Armitage test is asymptotically efficient for all monotone alternatives (Tarone, 1975), but this result only holds asymptotically. And tumors are rare events, so the binominal

**Example 1** In a chronic feeding study in rats, we tested the hypothesis that, in the second year of the study, there was a dose-responsive increase in tumor incidence associated with the test compound. Below, we utilize, a Cox–Stuart test for trend to address this question. All groups start the second year with an equal number of animals.

| | Control | | Low Doses | | | High Doses | | |
|---|---|---|---|---|---|---|---|---|
| Month of Study | Total *X*, Animals with Tumors | Change [$X_{A-B}$] | Total *Y*, Animals with Tumors | Change [$Y_{A-B}$] | Compared to Control ($Y - X$) | Total *Z*, Animals with Tumors | Change [$Z_{a-b}$] | Compared to Control ($Z - X$) |
| 12 (*A*) | 1 | NA | 0 | NA | NA | 5 | NA | NA |
| 13 (*B*) | 1 | 0 | 0 | 0 | 0 | 7 | 2 | (+)2 |
| 14 (*C*) | 3 | 2 | 1 | 1 | (–)1 | 11 | 4 | (+)2 |
| 15 (*D*) | 3 | 0 | 1 | 0 | 0 | 11 | 0 | 0 |
| 16 (*E*) | 4 | 1 | 1 | 0 | (–)1 | 13 | 2 | (+)1 |
| 17 (*F*) | 5 | 1 | 3 | 2 | (+)1 | 14 | 1 | 0 |
| 18 (*G*) | 5 | 0 | 3 | 0 | 0 | 15 | 1 | (+)1 |
| 19 (*H*) | 5 | 0 | 5 | 2 | (+)2 | 18 | 3 | (+)3 |
| 20 (*I*) | 6 | 1 | 6 | 1 | 0 | 19 | 1 | 0 |
| 21 (*J*) | 8 | 2 | 7 | 1 | (–)1 | 22 | 3 | (+)1 |
| 22 (*K*) | 12 | 4 | 9 | 2 | (–)2 | 26 | 4 | 0 |
| 23 (*L*) | 14 | 2 | 12 | 3 | (+)1 | 28 | | |
| 6 + | | | | | | | | |
| 1 – | | | | | | | | |
| | | | | $Y - X = 0$ (no trend) | | | $Z - X = 5$ | |

Reference to a sign table is not necessary for the low-dose comparison (where there is no trend) but clearly shows the high dose to be significant at the $p \leq 0.5$ level.

proportions are small. In this situation approximations may become unreliable.

Therefore, exact tests which can be performed using two different approaches—conditional and unconditional—are considered. In the first case, the total number of tumors $r$ is regarded as fixed. As a result the null distribution of the test statistic is independent of the common probability $p$. The exact conditional null distribution is a multivariate hypergeometric distribution.

The unconditional model treats the sum of all tumors as a random variable. Then the exact unconditional null distribution is a multivariate binomial distribution. The distribution depends on the unknown probability.

Such low-dose extrapolation is typically only conducted for tumors believed to be caused by a genotoxic effect which some, but by no means all, scientists believe have no threshold. For other types of tumors and for many nonneoplastic endpoints a threshold cannot be estimated directly from data at a limited number of dose levels, and a NOEL can be estimated by finding the highest dose level at which there is no significant increase in effects.

A useful technique for determining if there is an effect of treatment on any toxicological parameter is the NOSTASOT method (Tukey et al., 1985; Antonello et al., 1993). This test is based on the principle that a possible toxicological effect of interest occurs with a normal dose response; that is, there is an increasing effect with increasing dosage. The data to be analyzed should be examined first to confirm that this principle is not violated. In this method, regression analysis is used to determine if there is an increased or decreased response in a parameter with increasing dosage. This method can be visualized as a plot of response versus dosage in which the analysis determines if the slope of the plotted line deviates significantly from zero.

This method can be used for essentially all parameters. Three analyses are performed each with different spacing between dosage levels. The spacing in the first analysis is based on the arithmetic values of the dosage levels. The spacing in the second, referred to as the ordinal scaling, has equal spacing between dosage levels; that is, the control through high dosage levels are assigned values of 0, 1, 2, and 3. In the third analysis, the log of the dosage level is used. Since the log of zero is impractical, the control group is assigned a value based on the spacing between the low and middle dosage levels according to a formula that assigns a log scale value to the control such that the ratio of the difference between the control and low-dose groups and the difference between the low- and middle-dose groups is equal both in absolute values and in log scale values. This places the control group at a reasonable distance from the low-dose group. The lowest $p$ value among the three analyses—arithmetic, ordinal, and logarithmic—is taken as the $p$ value of the overall analysis based on the assumption that, if there is a dosage-related effect, the method of analysis yielding the lowest value is the best model for that dosage response. A correction for the multiplicity of analyses can be applied. If none of the three analyses are significant at the 0.05 level, the analysis is complete and the

high dosage level is referred to as the "no statistical significance of trend dose," or the NOSTASOT dose. If there is a significant trend through the high dosage level, the data from the high dosage level is deleted and the trend test repeated. This process is repeated until a NOSTASOT dose is determined. Effects at dosage levels above the NOSTASOT dose are then considered to be statistically significant.

There are two major benefits of the NOSTASOT method. One is that spurious statistically significant results only at the low and/or middle dosage levels are eliminated, resulting in a reduction in false positives. A second benefit is that in some cases there may be real effects as multiple dosage levels that at any single dosage level are not statistically significant but will nevertheless result in a significant trend, thus providing increased sensitivity and reducing false negatives.

### 5.8 Need for Age Adjustment

Where there are marked differences in survival between treated groups, it is widely recognized that there is a need for an age adjustment (i.e., an adjustment for age at death or onset). This is illustrated in the example above, where, because of the greater number of deaths occurring early in the treated group, the true effect of treatment disappears if no adjustment is made. Thus, a major purpose of age adjustment is to avoid bias.

It is not so generally recognized, however, that, even where there are no survival differences, age adjustment can increase the power to detect between-group differences. This is illustrated in the example below:

| | Control | Exposed |
|---|---|---|
| Early deaths | 0/20 | 0/20 |
| Middle deaths | 1/10 | 9/10 |
| Late deaths | 20/20 | 20/20 |
| Total | 21/50 | 29/50 |

Here treatment results in a somewhat earlier onset of a condition which occurs eventually in all animals. Failure to age adjust will result in a comparison of 29/50 with 21/50, which is not statistically significant. Age adjustment will essentially ignore the early and late deaths, which contribute no comparative statistical information, and be based on the comparison of 9/10 with 1/10, which is statistically significant. Here age adjustment sharpens the contrast, rather than avoiding bias, by avoiding diluting data capable of detecting treatment effects with data that are of little or no value for this purpose.

### 5.9 Need to Take Context of Observation into Account

It is now widely recognized that age adjustment cannot properly be carried out unless the context of observation is taken into account. There are three

relevant contexts, the first two relating to the situation where the condition is only observed at death (e.g., an internal tumor) and the third where it can be observed in life (e.g., a skin tumor):

In the first context the condition is assumed to have caused the death of the animal, that is, to be *fatal*. Here the incidence rate for a time interval and a group is calculated by the number of animals dying in the interval because of the lesion divided by the number of animals alive at the start of the interval.

In the second context, the animal is assumed to have died of another cause, that is, the condition is *incidental*. Here the rate is calculated by the number of animals dying in the interval with the lesion divided by the total number of animals dying in the interval.

In the third context, where the condition is *visible*, the rate is calculated by the number of animals getting the condition in the interval divided by the number of animals without the condition at the start of the interval.

A problem with the method of Peto et al. (1980), which takes the context of observation into account, is that some pathologists are unwilling or feel unable to decide whether, in any given case, a condition is fatal or incidental. A number of points should be made here.

First, where there are marked survival differences, it may not be possible to conclude reliably whether a treatment is beneficial or harmful unless such a decision is made. This is well illustrated by the example in Peto et al. (1980), where assuming all pituitary tumors were fatal results in the (false) conclusion that *N*-nitrosodimethylamine (NDMA) was carcinogenic, while assuming they were all incidental resulted in the (false) conclusion that NDMA was protective. Using, correctly, the pathologist's best opinion as to which were and which were not likely to be fatal resulted in an analysis which (correctly) concluded NDMA had no effect. If the pathologist, in this case, had been unwilling to make a judgment as to fatality, believing it to be unreliable, no conclusion could have been reached. This state of affairs would, however, be a fact of life, and *not* a position reached because an inappropriate statistical method was being used.

Although it will normally be a good routine for the pathologist to ascribe "factors contributory to death" for each animal that was not part of a scheduled sacrifice, it is in fact not strictly necessary to determine the context of observation for all conditions at the outset. An alternative strategy is to analyze under differing assumptions: (1) no cases fatal, (2) all cases occurring in descendants fatal, and (3) all cases of the same defined severity occurring in descedants fatal, with, under each assumption, other cases incidental.

If the conclusion turns out the same under each assumption or if the pathologist can say, on general grounds, that one assumption is likely to be a close approximation to the truth, it may not be necessary to know the context

of observation for the condition in question for each individual animal. Using the alternative strategy might result in a saving of the pathologist's time by only having to make a judgment for a limited number of conditions where the conclusion seems to hand on correct knowledge of the context of observation.

Finally, it should be noted that, although many nonneoplastic conditions observed at death are never causes of death, it is, in principle, as necessary to know the context of observation for nonneoplastic conditions as it is for tumors.

## 5.10 Experimental and Observational Units

In many situations, the animal is both the "experimental unit" and the "observational unit", but this is not always so. For determining treatment effects by the methods of the next section, it is important that each experimental unit provides only one item of data for analysis, as the methods all assume that individual data items are statistically independent. In many feeding studies, where the cage is assigned to a treatment, it is the cage, rather than the animal, that is the experimental unit. In histopathology, observations for a tissue are often based on multiple sections per animal, so that the section is the observational unit. Multiple observations per experimental unit should be combined in some suitable way into an overall average for that unit before analysis.

## 5.11 Missing Data

In many types of analysis, animals with missing data are simply removed from the analysis. There are, however, some situations where this can be an inappropriate thing to do. One situation is when carrying out an analysis of a condition that is assumed to have caused the death of the animal. Although an animal dying at week 83 for which the section was unavailable for microscopic examination cannot contribute to the group comparison at week 83, one knows that it did not die because of any condition in previous weeks, so it should contribute to the denominator of the calculations in all previous weeks.

Another situation is when histopathological examination of a tissue is not carried out unless an abnormality is seen postmortem. In such an experiment one might have the following data for that tissue:

- Control group: 50 animals, 2 abnormal postmortem, 2 examined microscopically, 2 with tumor of specific type.
- Treated group: 50 animals, 15 abnormal postmortem, 15 examined microscopically, 14 with tumor of specific type.

Ignoring animals with no microscopic sections, one would compare 2/2 = 100% with 14/15 = 93% and conclude treatment nonsignificantly decreased incidence. This is likely to be a false conclusion, and it would be better here to compare the percentages of animals which had a postmortem

abnormality which turned out to be a tumor, that is, 2/50 = 4% with 14/50 = 28%. Unless some aspect of treatment made tumors much easier to detect postmortem, one could then conclude that treatment did have an effect on tumor incidence.

Particular care has to be taken in studies where the procedures for histopathological examination vary by group. In a number of studies conducted in recent years, the protocol demands full microscopic examination of a given tissue list in descedants in all groups and in terminally killed controls in high-dose animals. In other animals, that is, terminally killed low- and middose animals, microscopic examination of a tissue is only conducted if the tissue is found to be abnormal postmortem. Such a protocol is designed to save money but leads to difficulty in comparing the treatment groups validly. Suppose, for example, responses in terminally killed animals are 8/20 in the controls, 3/3 (with 17 unexamined) in the low-dose, and 5/6 (with 14 unexamined) in the middose animals. Is one supposed to conclude that treatment at the low- and middoses increased response based on a comparison of the proportions examined microscopically (40, 100, and 83%) or that it decreased response based on the proportion of animals in the group (40, 15, and 25%)? It could well be that treatment had no effect but some small tumors were missed postmortem. In this situation, a valid comparison can only be achieved by ignoring the low- and middose groups when carrying out the comparison for the age stratum "terminal kill." This, of course, seems wasteful of data, but these are data that cannot be usefully used owing to the inappropriate protocol.

## 5.12 Use of Historical Control Data

In some situations, particularly where incidences are low, the results from a single study may suggest an effect of treatment on tumor incidence but be unable to demonstrate it conclusively. The possibility of comparing results in the treated groups with those of control groups from other studies is then often raised. Thus, a nonsignificant incidence of 2 cases out of 50 in a treated group may seem much more significant if no cases have been seen in, say, 1000 animals representing controls from 20 similar studies. Conversely, a significant incidence of 5 cases out of 50 in a treated group as compared with 0 out of 50 in the study controls may seem far less convincing if many other control groups had incidences around 5 out of 50.

While not understating the importance of looking at historical control data, it must be emphasized that there are a number of reasons why variation between studies may be greater than variation within study. Differences in diet, in duration of the study, in intercurrent mortality, and in who the study pathologist is may all contribute. Statistical techniques that ignore this and carry out simple statistical tests of treatment incidence against a pooled control incidence may well give results that are seriously in error and are likely to overstate statistical significance considerably.

## 5.13 Methods for Data Examination and Preparation

The data from toxicology studies should always be examined before any formal analysis. Such examinations should be directed to determining if the data are suitable for analysis and, if so, what form the analysis should take (see Figure 2). If the data as collected are not suitable for analysis or if they are only suitable for low-powered analytical techniques, one may wish to use one of many forms of data transformation to change the data characteristics so that they are more amenable to analysis.

The above two objectives, data examination and preparation, are the primary focus of this chapter. For data examination, two major techniques are presented—the scattergram and Bartlett's test. Likewise, for data preparation (with the issues of rounding and outliers having been addressed in a previous chapter) two techniques are presented—randomization (including a test for randomness in a sample of data) and transformation. Exploratory data analysis (EDA) is presented and briefly reviewed later. This is a broad collection of techniques and approaches to "probe" data—that is, to both examine and perform some initial, flexible analysis of the data.

## 5.14 Scattergram

Two of the major points to be made throughout this volume are (a) the use of the appropriate statistical tests and (b) the effects of small sample sizes (as is often the case in toxicology) on our selection of statistical techniques. Frequently, simple examination of the nature and distribution of data collected from a study can also suggest patterns and results which were unanticipated and for which the use of additional or alternative statistical methodology is warranted. It was these three points which caused the author to consider a section on scattergrams and their use essential for toxicologists.

Bartlett's test may be used to determine if the values in groups of data are homogeneous. If they are, this (along with the knowledge that they are from a continuous distribution) demonstrates that parametric methods are applicable.

But, if the values in the (continuous data) groups fail Bartlett's test (i.e., are heterogeneous), we cannot be secure in our belief that parametric methods are appropriate until we gain some confidence that the values are normally distributed. With large groups of data, we can compute parameters of the population (kurtosis and skewness, in particular) and from these parameters determine if the population is normal (with a certain level of confidence). If our concern is especially marked, we can use a chi-square goodness of fit test for normality. But when each group of data consists of 25 or fewer values, these measures or tests (kurtosis, skewness, and chi-square goodness of fit) are not accurate indicators of normality. Instead, in these cases we should prepare a scattergram of the data, then evaluate the scattergram to estimate if the data

**Example 2** Suppose we have the two data sets below:

***Group 1:*** 4.5, 5.4, 5.9, 6.0, 6.4, 6.5, 6.9, 7.0, 7.1, 7.0, 7.4, 7.5, 7.5, 7.5, 7.6, 8.0, 8.1, 8.4, 8.5, 8.6, 9.0, 9.4, 9.5 and 10.4.

***Group 2:*** 4.0, 4.5, 5.0, 5.1, 5.4, 5.5, 5.6, 6.5, 6.5, 7.0, 7.4, 7.5, 7.5 8.0, 8.1, 8.5, 8.5, 9.0, 9.1, 9.5, 9.5, 10.1, 10.0 and 10.4.

Both of these groups contain 24 values and cover the same range. From them we can prepare the following scattergrams:

***Group 1:***

```
                                    x
                                x   x
                                x   x       x
                    x   x       x   x   x   x           x
    x       x       x   x       x   x   x   x   x       x
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |
___ ___ ___ ___ ___ ___ ___ ___ ___ ___ ___ ___ ___ ___
4       5       6       7       8       9       10
```

***Group 2:***

```
                x                   x
            x   x           x       x   x   x   x   x   x
x   x       x   x           x   x   x   x   x   x   x   x   x
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |
___ ___ ___ ___ ___ ___ ___ ___ ___ ___ ___ ___ ___ ___
4       5       6       7       8       9       10
```

are normally distributed. This procedure consists of developing a histogram of the data, then examining the histogram to get a visual appreciation of the location and distribution of the data.

The abscissa (or horizontal scale) should be in the same scale as the values and should be divided so that the entire range of observed values is covered by the scale of the abscissa. Across such a scale we then simply enter symbols for each of our values. Figure 2 shows such a plot.

Example 2 presents a traditional and rather limited form of scatterplot but such plots can reveal significant information about the amount and types of association between the two variables, the existence and nature of outliers, the clustering of data, and a number of other two-dimensional factors (Anscombe, 1973; Chambers et al., 1983).

Current technology allows us to add significantly more graphical information to scatterplots by means of graphic symbols (letters, faces, different shapes such as squares, colors, etc.) for the plotted data points. One relatively simple example of this approach is shown in Figure 5, where the simple case of dose (in a dermal study), dermal irritation, and white blood cell count are presented. This graph quite clearly suggests that as dose (variable $x$) is increased, dermal irritation (variable $y$) also increases, and as irritation becomes

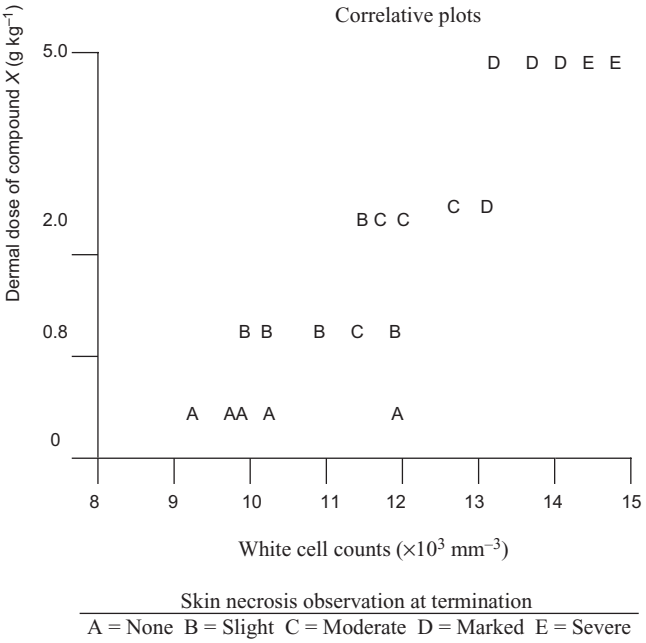


***Figure 5*** *Exploratory Data Analysis.*

more severe, white blood cell count (variable $z$), an indicator of immune system involvement, increases, suggesting infection or persistent inflammation, also increases. There is no direct association of variables $x$ and $z$, however (Cleveland and McGill, 1984; Cleveland, 1985; Tufte, 1990).

Group 1 can be seen to approximate a normal distribution (bell-shaped curve); we can proceed to perform the appropriate parametric tests with such data. But group 2 clearly does not appear to be normally distributed. In this case, the appropriate nonparametric technique must be used.

## 5.15 Bartlett's Test for Homogeneity of Variance

Bartlett's test (Sokal and Rohlf, 1994) is used to compare the variances (values reflecting the degree of variability in data sets) among three or more groups of data, where the data in the groups are continuous sets (such as body weights, organ weights, red blood cell counts, or diet consumption measurements). It is expected that such data will be suitable for parametric methods (normality of data is assumed) and Bartlett's is frequently used as a test for the assumption of equivalent variances.

Bartlett's is based on the calculation of the corrected $\chi^2$ (chi-square) value by the formula

$$\chi^2_{\text{corr}} = 2.3026 \frac{\sum \text{df}\left(\log_{10}\left\{\sum[\text{df}(S^2)]/\sum \text{df}\right\}\right) - \sum[\text{df}(\log_{10} S^2)]}{1 + \{1/[3(K-1)]\}\left[\sum(1/\text{df}) - 1/\sum \text{df}\right]}$$

where $S^2$ = variance, = $\{[n\Sigma X^2 - (\Sigma X)^2]/n\}/(n-1)$
$X$ = individual datum within each group
$n$ = number of data within each group
$K$ = number of groups being compared
df = degrees of freedom for each group, = $(N-1)$

The corrected $\chi^2$ value yielded by the above calculations is compared to the values listed in the chi-square table according to the numbers of degrees of freedom (Snedecor and Cochran, 1980).

If the calculated value is smaller than the table value at the selected $p$ level (traditionally 0.05), the groups are accepted to be homogeneous and the use of ANOVA is assumed proper. If the calculated $\chi^2$ is greater than the table value, the groups are heterogeneous and other tests (as indicated in Figure 2, the decision tree) are necessary.

**ASSUMPTIONS AND LIMITATIONS**

1. Bartlett's test does not test for normality, but rather it tests for homogeneity of variance (also called equality of variances or homoscedasticity).
2. Homoscedasticity is an important assumption for Student's $t$ test, ANOVA, and analysis of covariance.
3. The $F$ test (covered in Section 7.3) is actually a test for the two-sample (i.e., control and one test group) case of homoscedasticity. Bartlett's is designed for three or more samples.
4. Bartlett's is very sensitive to departures from normality. As a result, a finding of a significant chi-square value in Bartlett's may indicate non-normality rather than heteroscedasticity. Such a finding can be brought about by outliers, and the sensitivity to such erroneous findings is extreme with small sample sizes.

## 5.16 Statistical Goodness-of-Fit Tests

A goodness-of-fit test is a statistical procedure for comparing individual measurements to a specified type of statistical distribution. For example, a normal distribution is completely specified by its arithmetic mean and variance (the square of the standard deviation ). The null hypothesis, that the data represent a sample from a single normal distribution, can be tested by a statistical goodness-of-fit test. Various goodness-of-fit tests have been devised to

determine if the data deviate significantly from a specified distribution. If a significant departure occurs, it indicates only that the specified distribution can be rejected with some assurance. This does not necessarily mean that the true distribution contains two or more subpopulations. The true distribution may be a single distribution, based upon a different mathematical relationship (e.g., lognormal). In the latter case, logarithms of the measurement would not be expected to exhibit by a goodness-of-fit test a statistically significant departure from a lognormal distribution.

Everitt and Hand (1981) recommended use of a sample of 200 or more to conduct a valid analysis of mixtures of populations. Even the maximum-likelihood method, the best available method, should be used with extreme caution or not at all when separation between the means of the subpopulations is less than 3 SD and sample sizes are less than 300. None of the available methods conclusively establish bimodality, which may, however, occur when separation between the two means (modes) exceeds 2 SD. Conversely, inflections in probits or separations in histograms *less than* 2 SD apart may arise from genetic differences in test subjects.

Mendell et al. (1993) compared eight tests of normality to detect a mixture consisting of two normally distributed components with different means but equal variances. Fisher's skewness statistic was preferable when one component comprised less than 15% of the total distribution. When the two components comprised more nearly equal proportions (35–65%) of the total distribution, the Engelman–Hartigan (1969) test was preferable. For other mixing proportions, the maximum-likelihood ratio test was best. Thus, the maximum-likelihood ratio test appears to perform very well, with only small loss from optimality, even when it is not the best procedure.

The method of *maximum likelihood* provides estimators which are usually quite satisfactory. They have the desirable properties of being consistent, asymptotically normal, and asymptotically efficient for large samples under quite general conditions. They are often biased, but the bias is frequently removable by a simple adjustment. Other methods of obtaining estimators are also available, but the maximum-likelihood method is the most frequently used.

Maximum-likelihood estimators also have another desirable property: *invariance*. Let us denote the maximum-likelihood estimator of the parameter $\theta$ by $\hat{\sigma}$. Then, if $f(\theta)$ is a single-valued function of $\theta$, the maximum-likelihood estimator of $f(\theta)$ is $f(\hat{\sigma})$. Thus, for example,

$$\hat{\sigma} = (\hat{\sigma}^2)1/2 \ .$$

The principle of maximum likelihood tells us that we should use as our estimate that value which maximizes the likelihood of the observed event.

These maximum-likelihood methods can be used to obtain *point estimates* of a parameter, but we must remember that a point estimator is a random variable distributed in some way around the true value of the parameter. The true parameter value may be higher or lower than our estimate. It is often

useful therefore to obtain an interval within which we are reasonably confident the true value will lie, and the generally accepted method is to construct what are known as *confidence limits*.

The following procedure will yield upper and lower 95% confidence limits with the property that when we say that these limits include the true value of the parameter, 95% of all such statements will be true and 5% will be incorrect.

1. Choose a (test) statistic involving the unknown parameter and no other unknown parameter.
2. Place the appropriate sample values in the statistic.
3. Obtain an equation for the unknown parameter by equating the test statistic to the upper 2½% point of the relevant distribution.
4. The solution of the equation gives one limit.
5. Repeat the process with the lower 2½% point to obtain the other limit.

One can also construct 95% confidence intervals using unequal tails (e.g., using the upper 2% point and the lower 3% point). We usually want our confidence interval to be as short as possible, however, and with a symmetric distribution such as the normal or $t$, this is achieved using equal tails. The same procedure very nearly minimizes the confidence interval with other nonsymmetric distributions (e.g., chi square) and has the advantage of avoiding rather tedious computation.

When the appropriate statistic involves the square of the unknown parameter, both limits are obtained by equating the statistic to the upper 5% point of the relevant distribution. The use of two tails in this situation would result in a pair of nonintersecting intervals. When two or more parameters are involved, it is possible to construct a region within which we are reasonably confident the true parameter values will lie. Such regions are referred to as confidence regions. The implied interval for $p_1$ does not form a 95% confidence interval, however. Nor is it true that an 85.7375% confidence region for $p_1$, $p_2$, and $p_3$ can be obtained by considering the intersection of the three separate 95% confidence intervals because the statistics used to obtain the individual confidence intervals are not independent. This problem is obvious with a multiparameter distribution such as the multinomial, but it even occurs with the normal distribution because the statistic which we use to obtain a confidence interval for the mean and the statistic which we use to obtain a confidence interval for the variance are not independent. The problem is not likely to be of great concern unless a large number of parameters is involved.

## 5.17 Randomization

Randomization is the act of assigning a number of items (e.g., plates of bacteria or test animals) to groups in such a manner that there is an equal chance for any one item to end up in any one group. This is a control against any possible

bias in assignment of subjects to test groups. A variation on this is censored randomization, which ensures that the groups are equivalent in some aspect after the assignment process is complete. The most common example of a censored randomization is one in which it is ensured that the body weights of test animals in each group are not significantly different from those in the other groups. This is done by analyzing group weights for both homogeneity of variance and ANOVA after animal assignment, then rerandomizing if there is a significant difference at some nominal level, such as $p \leq 0.10$. The process is repeated until there is no significant difference.

There are several methods for actually performing the randomization process. The three most commonly used are card assignment, use of a random number table, and use of a computerized algorithm.

For the card-based method, individual identification numbers for items (e.g., plates or animals) are placed on separate index cards. These cards are then shuffled and placed one at a time in succession into piles corresponding to the required test groups. The results are a random group assignment.

The random-number table method requires only that one have unique numbers assigned to test subjects and access to a random-number table. One simply sets up a table with a column for each group to which subjects are to be assigned. We start from the head of any one column of numbers in the random table (each time the table is used, a new starting point should be utilized). If our test subjects number less than 100, we utilize only the last two digits in each random number in the table. If they number more than 99 but less than 1000, we use only the last three digits. To generate group assignments, we read down a column one number at a time. As we come across digits which correspond to a subject number, we assign that subject to a group (enter its identifying number in a column) proceeding to assign subjects to groups from left to right filling one row at a time. After a number is assigned to an animal, any duplication of its unique number is ignored. We use as many successive columns of random numbers as we may need to complete the process.

The third (and now most common) method is to use a random-number generator that is built into a calculator or computer program. Procedures for generating these are generally documented in user manuals.

## 5.18 Transformations

If our initial inspection of a data set reveals it to have an unusual or undesired set of characteristics (or to lack a desired set of characteristics), we have a choice of three courses of action. We may proceed to select a method or test appropriate to this new set of conditions, or abandon the entire exercise, or transform the variable(s) under consideration in such a manner that the resulting transformed variates ($X'$ and $Y'$, e.g., as opposed to the original variates $X$ and $Y$) meet the assumptions or have the characteristics that are desired.

The key to all this is that the scale of measurement of most (if not all) variables is arbitrary. Although we are most familiar with a linear scale of

measurement, there is nothing which makes this the "correct" scale on its own, as opposed to a logarithmic scale [familiar logarithmic measurements are that of pH values or earthquake intensity (Richter scale)]. Transforming a set of data (converting $X$ to $X'$) is really as simple as changing a scale of measurement.

There are at least four good reasons to transform data:

1. To normalize the data, making them suitable for analysis by our most common parametric techniques such as ANOVA. A simple test of whether a selected transformation will yield a distribution of data which satisfies the underlying assumptions for ANOVA is to plot the cumulative distribution of samples on probability paper (i.e., a commercially available paper which has the probability function scale as one axis). One can then alter the scale of the second axis (i.e., the axis other than the one which is on a probability scale) from linear to any other (logarithmic, reciprocal, square root, etc.) and see if a previously curved line indicating a skewed distribution becomes linear to indicate normality. The slope of the transformed line gives us an estimate of the standard deviation. And if the slopes of the lines of several samples or groups of data are similar, we accordingly know that the variance of the different groups are homogenous.
2. To linearize the relationship between a paired set of data, such as dose and response. This is the most common use in toxicology for transformations and is demonstrated in the section under probit/logit plots (Section 7.13).
3. To adjust data for the influence of another variable. This is an alternative in some situations to the more complicated process of analysis of covariance. A ready example of this usage is the calculation of organ weight–body weight ratios in in vivo toxicity studies, with the resulting ratios serving as the raw data for an ANOVA performed to identify possible target organs. This use is discussed in detail later in this chapter.
4. Finally, to make the relationships between variables clearer by removing or adjusting for interactions with third, fourth, and so on. Uncontrolled variables influence the pair of variables of interest. This case is discussed in detail under time series analysis (Section 8.5).

Common transformations are presented in Table 6.

## 5.19 Exploratory Data Analysis

Over the past 20 years, an entirely new approach has been developed to get the most information out of the increasingly larger and more complex data sets that scientists are faced with. This approach involves the use of a very diverse set of fairly simple techniques which comprise exploratory data analysis (EDA). As expounded by Tukey (1977), there are four major ingredients to EDA:

**TABLE 6 Common Data Transformations**[b]

| Transformation | How Calculated[a] | Example of Use |
|---|---|---|
| Arithmetic | $x' = x/y$ or $x' = x + c$ | Organ weight/body weight |
| Reciprocals | $x' = 1/x$ | Linearizing data, particularly rate phenomena |
| Arcsine (also called angular) | $x' = \arcsin\sqrt{x}$ | Normalizing dominant lethal and mutation rate data |
| Logarithmic | $x' = \log x$ | pH values |
| Probability (probit) | $x' =$ probability $X$ | Percentage responding |
| Square roots | $x' = \sqrt{x}$ | Surface area of animal from body weights |
| Box–Cox | $x' = \begin{cases} (x^v - l)v & \text{for } v \neq 0 \\ \ln x & \text{for } v = 0 \end{cases}$ | Family of transformsFor use when one has no prior knowledge of appropriate transformation to use |
| Rank transformations | Depends on nature of samples | As bridge between parametric and nonparametric statistics (Conover and Inman, 1981) |

[a]$x$ and $y$ are original variables, $x'$ and $y'$ transformed values; $c$ is a constant.
[b]Plotting a double reciprocal (i.e., $1/x$ vs. $1/y$) will linearize almost any data set. So will plotting the log transforms of a set of variables.

*Displays* These visually reveal the behavior of the data and suggest a framework for analysis. The scatterplot (presented earlier) is an example of this approach.

*Residuals* These are what remain of a set of data after a fitted model (such as a linear regression) or some similar level of analysis has been removed.

*Reexpressions* These involve questions of what scale would best simplify and improve the analysis of the data. Simple transformations, such as those presented earlier in this chapter, are used to simplify data behavior (e.g., linearizing or normalizing) and clarify analysis.

*Resistance* This is a matter of decreasing the sensitivity of analysis and summary of data to misbehavior, so that the occurrence of a few outliers, for example, will not complicate or invalidate the methods used to analyze the data. For example, in summarizing the location of a set of data, the median (but not the arithmetic mean) is high resistant.

These four ingredients are utilized in a process falling into two broad phases: an exploratory phase and a confirmatory phase. The exploratory phase isolates patterns in and features of the data and reveals them, allowing an inspection of the data before there is any firm choice of actual hypothesis testing or modeling methods.

Confirmatory analysis allows evaluation of the reproducibility of the patterns or effects. Its role is close to that of classical hypothesis testing but also often includes steps such as (a) incorporating information from an analysis of another, closely related set of data and (b) validating a result by assembling and analyzing additional data. These techniques are in general beyond the

scope of this text. However, Velleman and Hoaglin (1981) and Hoaglin et al. (1983) present a clear overview of the more important methods along with codes for their execution on a microcomputer (they have also now been incorporated into Minitab). A short examination of a single case of the use of these methods, however, is in order.

Toxicology has long recognized that no population—animal or human—is completely uniform in its response to any particular toxicant. Rather, a population is composed of a (presumably normal) distribution of individuals—some resistant to intoxication (hyporesponders), the bulk that respond close to a central value [such as the median lethal dose ($LD_{50}$)], and some that are very sensitive to intoxication (hyperresponders). This population distribution can, in fact, result in additional statistical techniques. The sensitivity of techniques such as ANOVA is reduced markedly by the occurrence of outliers (extreme high or low values, including hyper- and hyporesponders) which, in fact, serve to markedly inflate the variance (standard deviation) associated with a sample. Such variance inflation is particularly common in small groups that are exposed or dosed at just over or under a threshold level, causing a small number of individuals in the sample (who are more sensitive than the other members) to respond markedly. Such a situation is displayed in Figure 6, which plots the mean and standard deviations of methemoglobin levels in a series of groups of animals exposed to successively higher levels of a hemolytic agent.

Though the mean level of methemoglobin in group C is more than double that of the control group (A), no hypothesis test will show this difference to be significant because it has such a large standard deviation associated with it. Yet this "inflated" variance exists because a single individual has such a marked response. The occurrence of the inflation is certainly an indicator that the data need to be examined closely. Indeed, all tabular data in toxicology should be visually inspected for both trend and variance inflation.

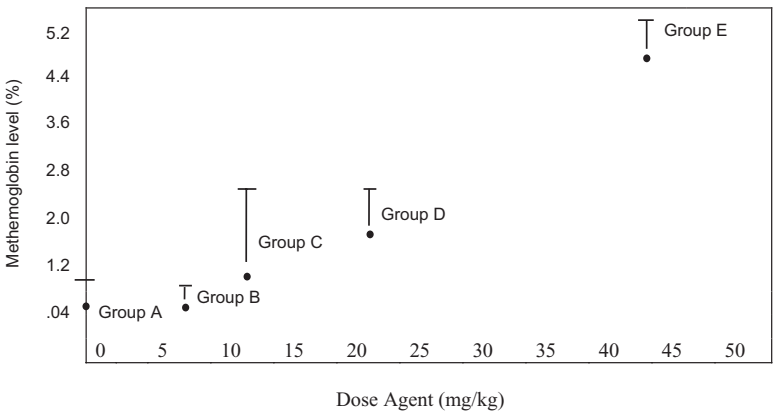


***Figure 6*** *Variance Inflation: points are means minus error bars plus one standard deviations.*

A concept related (but not identical) to resistance and EDA is that of robustness. Robustness generally implies insensitivity to departures from assumptions surrounding an underlying model, such as normality.

In summarizing the location of data the median, though highly resistant, is not extremely robust. But the mean is both nonresistant and nonrobust.

## 6 HYPOTHESIS TESTING OF CATEGORICAL AND RANKED DATA

Categorical (or contingency table) presentations of data can contain any single type of data, but generally the contents are collected and arranged so that they can be classified as belonging to treatment and control groups, with the members of each of these groups then classified as belonging to one of two or more response categories (such as tumor/no tumor or normal/hyperplastic/neoplastic). For these cases, two forms of analysis are presented: Fisher's exact test (for the $2 \times 2$ contingency table) and the $R \times C$ chi-square test (for large tables). It should be noted, however, that there are versions of both of these tests which permit the analysis of any size of contingency table.

The analysis of rank data—what is generally called nonparametric statistical analysis—is an exact parallel of the more traditional (and familiar) parametric methods. There are methods for the single-comparison case (just as Student's $t$ test is used) and for the multiple-comparison case (just as ANOVA is used) with appropriate post hoc tests for exact identification of the significance with a set of groups. Four tests are presented for evaluating statistical significance in rank data: the Wilcoxon rank-sum test, distribution-free multiple comparisons, Mann-Witney $U$ test, and Kruskall–Wallis nonparametric ANOVA. For each of these tests tables of distribution values for the evaluations of results can be found in any of a number of reference volumes (Gad, 1998).

It should be clearly understood that for data which do not fulfill the necessary assumptions for parametric analysis these nonparametric methods are either as powerful or in fact more powerful than the equivalent parametric test.

### 6.1 Fisher's Exact Test

Fisher's exact test should be used to compare two sets of discontinuous, quantal (all-or-none) data. Small sets of such data can be checked by contingency data tables, such as those of Finney et al. (1963). Larger sets, however, require computation. These include frequency data such as incidences of mortality or certain histopathological findings. Thus, the data can be expressed as ratios. These data do not fit on a continuous scale of measurement but usually involve numbers of responses classified as either negative or positive—that is, contingency table situation (Sokal and Rohlf, 1994).

The analysis is started by setting up a $2 \times 2$ contingency table to summarize the numbers of "positive" and "negative" responses as well as the totals of these as follows:

| | Positive | Negative | Total |
|---|---|---|---|
| Group I | $A$ | $B$ | $A + B$ |
| Group II | $C$ | $D$ | $C + D$ |
| Totals | $A + C$ | $B + D$ | $A + B + C + D = N_{total}$ |

Using the above set of symbols, the formula for $P$ appears as follows[1]:

$$P = \frac{(A+B)!(C+D)!(A+C)!(B+D)!}{N!A!B!C!D!}$$

The exact test produces a probability ($P$) which is the sum of the above calculation repeated for each possible arrangement of the numbers in the above cells (i.e., $A$, $B$, $C$, and $D$) showing an association equal to or stronger than that between the two variables.

The $P$ resulting from these computations will be the exact one- or two-tailed probability depending on which of these two approaches is being employed. This value tells us if the groups differ significantly (with a probability less than 0.05, say) and the degree of significance.

**ASSUMPTIONS AND LIMITATIONS**

1. Tables are available which provide individual exact probabilities for small-sample-size contingency tables. See Zar, 1974, pp. 518–542.
2. Fisher's exact test must be used in preference to the chi-square test when there are small cell sizes.
3. The probability resulting from a two-tailed test is exactly double that of a one-tailed test from the same data.
4. Ghent (1972) has developed and proposed a good (though, if performed by hand, laborious) method extending the calculation of exact probabilities to $2 \times 3$, $3 \times 3$, and $R \times C$ contingency tables.
5. Fisher's probabilities are not necessarily symmetric. Although some analysts will double the one-tailed $p$ value to obtain the two-tailed result, this method is usually overly conservative.

### 6.2 2 × 2 Chi Square

Though Fisher's exact test is preferable for analysis of most $2 \times 2$ contingency tables in toxicology, the chi-square test is still widely used and is preferable in a few unusual situations (particularly if cell sizes are large yet only limited computational support is available).

[1] $A!$ is A factorial. For example, for 4! this would be (4) (3) (2) (1) = 24.

The formula is simply

$$\chi^2 = \frac{(O_1 - E_1)^2}{E_1} + \frac{(O_2 - E_2)^2}{E_2}$$
$$= \sum \frac{(O_i - E_1)^2}{E_i}$$

where $O$ are observed numbers (or counts) and $E$ are expected numbers. The common practice in toxicology is for the observed figures to be test or treatment group counts. The expected figure is calculated as

$$E = \frac{(\text{column total})(\text{row total})}{\text{grand total}}$$

for each box or cell in a contingency table.

Our degrees of freedom are $(R - 1)(C - 1) = (2 - 1)(2 - 1) = 1$. Looking at a chi-square table (such as in Table C in Gad 1998) for one degree of freedom we see that this is greater than the test statistic at 0.05 (3.84) but less than that at 0.01 (6.64) so that $0.05 > p > 0.01$.

## ASSUMPTIONS AND LIMITATIONS

### Assumptions

1. Data are univariate and categorical.
2. Data are from a multinomial population.
3. Data are collected by random, independent sampling.
4. Groups being compared are of approximately the same size, particularly the for small group sizes.

### When to Use

1. When the data are of a categorical (or frequency) nature
2. When the data fit the assumptions above
3. To test goodness to fit to a known form of distribution
4. When cell sizes are large

### When Not to Use

1. When the data are continuous rather than categorical
2. When sample sizes are small and very unequal
3. When sample sizes are too small (e.g., when total $N$ is less than 50 of if any expected value is less than 5)
4. For any $2 \times 2$ comparison (use Fisher's exact test instead).

### 6.3 $R \times C$ Chi Square

The $R \times C$ chi-square test can be used to analyze discontinuous (frequency) data as in the Fisher's exact or $2 \times 2$ chi-square tests. However, in the $R \times C$ test ($R$ = row, $C$ = column) we wish to compare three or more sets of data. An example would be comparison of the incidence of tumors among mice on three or more oral dosage levels. We can consider the data as positive (tumors) or negative (no tumors). The expected frequency for any box is equal to row total times column total divided by $N_{\text{total}}$.

As in the Fisher's exact test, the initial step is setting up a table (this time a $R \times C$ contingency table). This table would appear as follows:

| | Positive | Negative | Total |
|---|---|---|---|
| Group I | $A_1$ | $B_1$ | $A_1 + B_1 = N_1$ |
| Group II | $A_2$ | $B_2$ | $A_2 + B_2 = N_2$ |
| | $\downarrow$ | $\downarrow$ | |
| Group $R$ | $A_{\text{R}}$ | $B_{\text{R}}$ | $A_{\text{R}} + B_{\text{R}} = N_{\text{R}}$ |
| Total | $N_{\text{A}}$ | $N_{\text{B}}$ | $N_{\text{total}}$ |

Using these symbols, the formula for chi square is

$$\chi^2 = \frac{N_{\text{tot}}^2}{N_A N_B N_k}\left(\frac{A_1^2}{N_1} + \frac{A_2^2}{N_2} + \cdots \frac{A_K^2}{N_k} - \frac{N_A^2}{N_{\text{tot}}}\right)$$

The resulting $\chi^2$ value is compared to table values (as in Snedecor and Cochran, 1994, pp. 470–471) according to the number of degrees of freedom, which is equal to $(R - 1)(C - 1)$. If $\chi^2$ is smaller than the table value at the 0.05 probability level, the groups are not significantly different. If the calculated $\chi^2$ is larger, there is some difference among the groups and $2 \times R$ chi-square or Fisher's exact tests will have to be compared to determine which group(s) differ from which other group(s).

**ASSUMPTIONS AND LIMITATIONS**

1. Based on data being organized in a table (such as below) so that there are *cells* (below, *A, B, C*, and *D* are cells).

| | | **Columns (*C*)** | | |
|---|---|---|---|---|
| | | **Control** | **Treated** | **Total** |
| | **No effect** | $A$ | $B$ | $A + B$ |
| Rows ($R$) | | | | |
| | **Effect** | $C$ | $D$ | $C + D$ |
| Total | | $A + C$ | $B + D$ | $A + B + C + D$ |

2. None of the "expected" frequency values should be less than 5.0.
3. The chi-square test is always one tailed.
4. Without the use of some form of correction, the test becomes less accurate as the differences between group sizes increases.
5. The results from each additional column (group) is approximately additive. Due to this characteristic, chi square can be readily used for evaluating any $R \times C$ combination.
6. The results of the chi-square calculation must be a positive number.
7. The test is weak with either small sample sizes or when the expected frequency in any cell is less than 5 (this latter limitation can be overcome by "pooling"—combining cells.).
8. Test results are independent of order of cells, unlike Kolmogorov–Smirnov.
9. It can be used to test the probability of validity of any distribution.

### 6.4 Wilcoxon Rank-Sum Test

The Wilcoxon rank-sum test is commonly used for the comparison of two groups of nonparametric (inteval or not normally distributed) data, such as those which are not measured exactly but rather as falling within certain limits (e.g., how many animals died during each hour of an acute study.) The test is also used when there is no variability (variance is zero) within one or more of the groups we wish to compare (Sokal and Rohlf, 1994).

The data in both groups being compared are initially arranged and listed in order of increasing value. Then each number in the two groups must receive a rank value. Beginning with the smallest number in either group (which is given a rank of 1.0), each number is assigned a rank. If there are duplicate numbers (called "ties"), then each value of equal size will receive the median rank for the entire identically sized group. Thus if the lowest number appears twice, both figures receive a rank of 1.5. This in turn means that the ranks of 1.0 and 2.0 have been used and that the next highest number has a rank of 3.0. If the lowest number appears three times, then each is ranked as 2.0 and the next number has a rank of 4.0. Thus, each tied number gets a "median" rank. This process continues until all of the numbers are ranked. Each of the two columns of ranks (one for each group) is totaled, giving the "sum of ranks" for each group being compared. As a check, we can calculate the value $\frac{1}{2}(N)(N+1)$, where $N$ is the total number of data in both groups. The result should be equal to the sum of ranks for both groups.

The sum-of-rank values are compared to table values (Diem, 1975; Beyer, 1976a,b, pp. 409–413) to determine the degree of significant differences, if any. These tables include two limits (an upper and a lower) that are dependent upon the probability level. If the number of data is the same in both groups ($N_1 \neq N_2$), then the lesser sum of ranks (smaller $N$) is compared to the table

limits to find the degree of significance. Normally the comparison of the two groups ends here and the degree of significant difference can be reported.

### 6.5 Distribution-Free Multiple Comparison

The distribution-free multiple-comparison test should be used to compare three or more groups of nonparametric data. These groups are then analyzed two at a time for any significant differences (Hollander and Wolfe, 1973, pp. 124–129). The test can be used for data similar to those compared by the rank-sum test. We often employ this test for reproduction and mutagenicity studies (such as comparing survival rates of offspring of rats fed various amounts of test materials in the diet).

Two values must be calculated for each pair of groups: the difference in mean ranks and the probability-level value against which the difference will be compared. To determine the difference in mean ranks we must first arrange the data within each of the groups in order of increasing values. Then we must assign rank values, beginning with the smallest overall figure. Note that this ranking is similar to that in the Wilcoxon test except that it applies to more than two groups.

The ranks are then added for each of the groups. As a check, the sum of these should equal $\frac{1}{2}N_{\text{tot}}(N_{\text{tot}}+1)$, , where $N_{\text{tot}}$ is the total number of figures from all groups. Next we can find the mean rank ($R$) for each group by dividing the sum of ranks by the numbers in the data ($N$) in the group. These mean ranks are then taken in those pairs which we want to compare (usually each test group vs. the control) and the differences are found ($|R_1 - R_2|$. This value is expressed as an absolute figure; that is, it is always a positive number.

The second value for each pair of groups (the probability value) is calculated lated from the expression $z[(a/K)(K-1)]\sqrt{\frac{1}{12}N_{\text{tot}}(N_{\text{tot}}+1)}\sqrt{(1/N_1)(1/N_2)}$, where $a$ is the level of significance for the comparison (usually 0.05, 0.01; 0.001, etc.), $K$ is the total number of groups, and $Z$ is a figure obtained from a normal probability table and determining the corresponding "$Z$ score."

The result of the probability value calculation for each pair of groups is compared to the corresponding mean difference $|R_1 - R_2|$. If $|R_1 - R_2|$ is smaller, there is no significant difference between the groups. If it is larger, the groups are different and $|R_1 - R_2|$ must be compared to the calculated probability values for $a = 0.01$ and $a = 0.001$ to find the degree of significance.

**ASSUMPTIONS AND LIMITATIONS**

1. As with the Wilcoxon rank sum, too many tied ranks inflate the false positive.
2. Generally, this test should be used as a post hoc comparison after Kruskall–Wallis.

### 6.6 Mann–Whitney *U* Test

This is a nonparametric test in which the data in each group are first ordered from lowest to highest values, then the entire set (both control and treated values) is ranked, with the average rank being assigned to tied values. The ranks are then summed for each group and $U$ is determined according to

$$U_t = n_c n_t + \frac{1}{2}[n_t(n_t+1)] - R_t \qquad U_c = n_c n_t + \frac{1}{2}[n_c(n_c+1)] - R_c$$

where $n_c$, $n_t$ are the sample sizes for the control and treated groups and $R_c$, $R_t$ are the sums of ranks for the control and treated groups.

For the level of significance for a comparison of the two groups, the larger value of $U_c$ or $U_t$ is used. This is compared to critical values as found in tables (Siegel, 1956).

The Mann–Whitney $U$ test is employed for the count data, but which test should be employed for the percentage variables should be decided on the same grounds as described later under reproduction studies.

**ASSUMPTIONS AND LIMITATIONS**

1. It does not matter whether the observations are ranked from smallest to largest or vice versa.
2. This test should not be used for paired observations.
3. The test statistics from a Mann–Whitney are linearly related to those of the Wilcoxon. The two tests will always yield the same result. The Mann–Whitney is presented here for historical completeness, as it has been much favored in reproductive and developmental toxicology studies. However, it should be noted that the author does not include it in the decision tree for method selection (Figure 2).

### 6.7 Kruskal–Wallis Nonparametric ANOVA

The Kruskal–Wallis nonparametric one-way ANOVA should be the initial analysis performed when we have three or more groups of data which are by nature nonparametric (either not a normally distributed population or of a discontinuous nature or all the groups being analyzed are not from the same population) but not of a categorical (or quantal) nature. Commonly these will be either rank-type evaluation data (such as behavioral toxicity observation scores) or reproduction study data. The analysis is initiated (Pollard, 1977, pp. 170–173] by ranking all the observations from the combined groups to be analyzed. Ties are given the average rank of the tied values (i.e., if two values which would tie for 12th rank—and therefore would be ranked 12th and 13th—both would be assigned the average rank of 12.5).

The sum of ranks of each group ($r_1, r_2, \ldots, r_k$) is computed by adding all the rank values for each group. The test value $H$ is then computed as

$$H = \frac{12}{n(n+1)} \sum \left( \frac{r_1^2}{n_1} + \frac{r_2^2}{n_2} + \cdots + \frac{r_k^2}{n_k} \right) - 3(n+1)$$

where $n_1, n_2, \ldots, n_k$ are the number of observations in each group. The test statistic is then compared with a table of $H$ values. If the calculated value of $H$ is greater than the table value for the appropriate number of observations in each group, there is a significant difference between the groups, but further testing (using the distribution-free multiple-comparison method) is necessary to determine where the difference lies.

**ASSUMPTIONS AND LIMITATIONS**

1. The test statistic $H$ is used for both small and large samples.
2. When we find a significant difference, we do not know which groups are different. It is not correct to then perform a Mann–Whitney $U$ test on all possible combinations—rather, a multiple-comparison method must be used, such as distribution-free multiple comparisons.
3. Data must be independent for the test to be valid.
4. Too many tied ranks will decrease the power of this test and also lead to increased false-positive levels.
5. When $k = 2$, the Kruskal–Wallis chi-square value has one degree of freedom (df). This test is identical to the normal approximation used for the Wilcoxon rank-sum Test. As noted in previous sections, a chi-square with one df can be represented by the square of a standardized normal random variable. In the case of $k = 2$, the $H$ statistic is the square of the Wilcoxon rank-sum $Z$ test (without the continuity correction).
6. The effect of adjusting for tied ranks is to slightly increase the value of the test statistic $H$. Therefore, omission of this adjustment results in a more conservative test.

## 6.8 Log Rank Test

The log rank test is a statistical methodology for comparing the distribution of time until the occurrence of the event in independent groups. In toxicology, the most common event of interest is death or occurrence of a tumor, but it could just as well be liver failure, neurotoxicity, or any other event which occurs only once in an individual. The elapsed time from initial treatment or observation until the *event* is the *event time*, often referred to as "survival time," even when the event is not "death."

The log rank Test provides a method for comparing "risk-adjusted" event rates, useful when test subjects in a study are subject to varying degrees of opportunity to experience the event. Such situations arise frequently in toxicology studies due to the finite duration of the study, early termination of the animal, or interruption of treatment before the event occurs.

Examples where use of the log rank test might be appropriate include comparing survival times in carcinogenity bioassay animals which are given a new treatment with those in the control group or comparing times to liver failure for several dose levels of a new nonsteroidal anti-inflammatory drug (NSAID) where the animals are treated for 10 weeks or until cured, whichever comes first.

If every animal were followed until the event occurrence, the event times could be compared between two groups using the Wilcoxon rank-sum test. However, some animals may die or complete the study before the event occurs. In such cases, the actual time of the event is unknown since the event does not occur while under study observation. The event times for these animals are based on the last known time of study observation and are called "censored" observations since they represent the lower bound of the true, unknown event times. The Wilcoxon rank-sum test can be highly biased in the presence of the censored data.

The null hypothesis tested by the log rank test is that of equal event time distributions among groups. Equality of the distributions of event times implies similar event rates among groups not only for the clinical trial as a whole but also for any arbitrary time point during the trial. Rejection of the null hypothesis indicates that the event rates differ among groups at one or more time points during the study.

The idea behind the log rank test for comparison of two life tables is simple: If there were no difference between the groups, the total deaths occurring at any time should split between the two groups at that time. So if the numbers at risk in the first and second groups in (say) the sixth month were 70 and 30, respectively, and 10 deaths occurred in that month we would expect

$$10 \times \frac{70}{70+30} = 7$$

of these deaths to have occurred in the first group and

$$10 \times \frac{30}{70+30} = 3$$

of the deaths to have occurred in the second group.

A similar calculation can be made at each time of death (in either group). By adding together for the first group the results of all such calculations, we obtain a single number, called the extent of exposure ($E_1$), which represents the "expected" number of deaths in that group if the two groups had the distribution of survival time. An extent of exposure ($E_2$) can be obtained for the

second group in the same way. Let $O_1$ and $O_2$ denote the actual total numbers of deaths in the two groups. A useful arithmetic check is that the total number of deaths $O_1 + O_2$ must equal the sum $E_1 + E_2$ of the extents of exposure.

The discrepancy between the $O$'s and $E$'s can be measured by the quantity

$$x^2 = \frac{\left(|O_1 - E_1| - \frac{1}{2}\right)^2}{E_1} + \frac{\left(|O_2 - E_2| - \frac{1}{2}\right)^2}{E_2}$$

For rather obscure reasons, $x^2$ is known as the log rank statistic. An approximate significance test of the null hypothesis of identical distributions of survival time in the two groups is obtained by referring $x^2$ to a chi-square distribution on one df.

The log rank test as presented by Peto et al. (1977) uses the product limit life table calculations rather than the actuarial estimators shown above. The distinction is unlikely to be of practical importance unless the grouping intervals are very coarse.

Peto and Pike (1973) suggest that the approximation in treating the null distribution of $\chi^2$ as a chi square is conservative, so that it will tend to understate the degree of statistical significance. In the formula for $\chi^2$ we have used the continuity correction of subtracting ½ from $|O_1 - E_1|$ and $|O_2 - E_2|$ before squaring. This is recommended by Peto et al. (1977) when, as in nonrandomized studies, the permutational argument does not apply. Peto et al. (1977) gives further details of the log rank test and its extension to comparisons of more than two treatment groups and to tests that control for categorical confounding factors.

## ASSUMPTIONS AND LIMITATIONS

1. The endpoint of concern is or is defined so that it is "right censored"—once it happens, it does not reoccur. Examples are death or a minimum or maximum value of an enzyme or physiological function (such as respiration rate).
2. The method makes no assumptions on distribution.
3. Many variations of the log rank test for comparing survival distributions exist. The most common variant has the form

$$\chi^2 = \frac{(O_1 - E_1)^2}{E_1} + \frac{(O_2 - E_2)^2}{E_2}$$

where $O_i$ and $E_i$ are computed for each group, as in the formulas given previously. This statistic also has an approximate chi-square distribution with one df under $H_0$.

A continuity correction can also be used to reduce the numerators by ½ before squaring. Use of such a correction leads to even further conservatism and may be omitted when sample sizes are moderate or large.

4. The Wilcoxon rank-sum test could be used to analyze the event times in the absence of censoring. A "generalized Wilcoxon" test, sometimes called the Gehan test, based on an approximate chi-square distribution has been developed for use in the presence of censored observations.

   Both the log rank and the generalized Wilcoxon tests are nonparametric tests and require no assumptions regarding the distribution of event times. When the event rate is greater early in the trial than toward the end, the generalized Wilcoxon test is the more appropriate test since it gives greater weight to the earlier differences.

5. Survival and failure times often follow the exponential distribution. If such a model can be assumed, a more powerful alternative to the log rank test is the likelihood ratio test.

   This parametric test assumes that event probabilities are constant over time. That is, the chance that a patient becomes event positive at time $t$ given that he or she is event negative up to time $t$ does not depend on $t$. A plot of the negative log of the event times distribution showing a linear trend through the origin is consistent with exponential event times.

6. Life tables can be constructed to provide estimates of the event time distributions. Estimates commonly used are known as the Kaplan–Meier estimates.

## 7 HYPOTHESIS TESTING: UNIVARIATE PARAMETRIC TESTS

Univariate case[2] data from normally distributed populations generally have a higher information value associated with them, but the traditional hypothesis-testing techniques (which include all the methods described in this chapter) are generally neither resistant nor robust. All the data analyzed by these methods are also effectively continuous—that is, at least for practical purposes, the data may be represented by any number and each such data number has a measurable relationship to other data numbers.

### 7.1 Student's *t* Test (Unpaired *t* Test)

Pairs of groups of continuous, randomly distributed data are compared via this test. We can use this test to compare three or more groups of data, but they must be compared by examination of two groups taken at a time and are preferentially compared by ANOVA. Usually this means comparison of a

[2]That is, where each datum is defined by one treatment and one effect variable.

test group versus a control group, although two test groups may be compared as well. To determine which of the three types of *t* tests described in this chapter should be employed, the *F* test is usually performed first. This will tell us if the variances of the data are approximately equal, which is a requirement for the use of the parametric methods. If the *F* test indicates homogeneous variances and the numbers of data within the groups (*N*) are equal, then the Student's *t* test is the appropriate procedure (Sokal and Rohlf, 1994). If the *F* is significant (the data are heterogeneous) and the two groups have equal numbers of data, the modified Student's *t* test is applicable (Cochran and Cox, 1975).

The value of *t* for the Student *t* test is calculated using the formula

$$t = \frac{\overline{X_1 - \overline{X_2}}}{\sum D_1^2 + \sum D_2^2} \sqrt{\frac{N_1 N_2}{N_1 + N_2}(N_1 + N_2 - 2)}$$

where $\Sigma D^2 = [N\Sigma X^2 - \Sigma X)^2]/N$.

The value of *t* obtained from the above calculation is compared to the values in a *t*-distribution table according to the appropriate number of degrees of freedom (df). If the *F* value is not significant (i.e., variances are homogeneous), df = $N_1 + N_2 - 2$. If the *F* was significant and $N_1 = N_2$, then df = $N - 1$. Although this case indicates a nonrandom distribution, the modified *t* test is still valid. If the calculated value is larger than the table value at $p = 0.05$, it may then be compared to the appropriate other table values in order of decreasing probability to determine the degree of significance between the two groups.

## ASSUMPTIONS AND LIMITATIONS

1. The test assumes that the data are univariate, continuous, and normally distributed.
2. Data are collected by randomly sampling.
3. The test should be used when the assumptions in 1 and 2 are met and there are only two groups to be compared.
4. Do not use when the data are ranked, when the data are not approximately normally distributed, or when there are more than two groups to be compared. Do not use for paired observations.
5. This is the most commonly misused test method, except in those few cases where one is truly only comparing two groups of data and the group sizes are roughly equivalent or not valid for multiple comparisons (because of resulting additive errors) or where group sizes are very unequal.
6. The test is robust for moderate departures from normality and, when $N_1$ and $N_2$ are approximately equal, robust for moderate departures from homogeneity of variances.
7. The main difference between the *Z* test and the *t* test is that the *Z* statistic is based on a known standard deviation σ while the *t* statistic uses

the sample standard deviation $s$ as an estimate of $\sigma$. With the assumption of normally distributed data, the variance $\sigma^2$ is more closely estimated by the sample variance $s^2$ as $n$ gets large. It can be shown that the $t$ test is equivalent to the $Z$ test for infinite degrees of freedom. In practice, a "large" sample is usually considered, $n \geq 30$.

### 7.2 Cochran *t* Test

The Cochran test should be used to compare two groups of continuous data when the variances (as indicated by the $F$ test) are heterogeneous and the numbers of data within the groups are not equal ($N_1 \neq N_2$). This is the situation, for example, when the data, though expected to be randomly distributed, were found not to be (Cochran and Cox, 1975, pp. 100–102).

Two $t$ values are calculated for this test, the "observed" $t$ ($t_{obs}$) and the "expected" $t$ ($t'$). The observed $t$ is obtained by

$$t_{obs} = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{W_1 + W_2}}$$

$$W = \text{SEM}^2 \quad \text{(standard error of the mean squared)}$$

$$= \frac{S^2}{N}$$

where $S$ (the variance) can be calculated from

$$S = \frac{\left[N\sum X^2 - \left(\sum X\right)^2\right]/N}{N-1}$$

The value for $t'$ is obtained from

$$t' = \frac{t_1' W_1 + t_2' W_2}{W_1 + W_2}$$

where $t_1'$ and $t_2'$ are values for the two groups taken from the $t$-distribution table corresponding to $N - 1$ degrees of freedom (for each group) at the 0.05 probability level (or such level as one may select).

The calculated $t_{obs}$ is compared to the calculated $t'$ value (or $t'$ values prepared for more than one probability level). If $t_{obs}$ is smaller than $t'$, the groups are not considered to be significantly different at that probability level.

**ASSUMPTIONS AND LIMITATIONS**

1. The test assumes that the data are univariate, continuous, and normally distributed and group sizes are unequal.
2. The test is robust for moderate departures from normality and very robust for departures from equality of variances.

### 7.3 *F* Test

This is a test of the homogeneity of variances between two groups of data (Sokal and Rohlf, 1994). It is used in two separate cases. The first is when Bartlett's indicates heterogeneity of variances among three or more groups (i.e., it is used to determine which pairs of groups are heterogeneous). Second, the *F* test is the initial step in comparing two groups of continuous data which we would expect to be parametric (two groups not usually being compared using ANOVA), the results indicating whether the data are from the same population and whether subsequent parametric comparisons would be valid.

The *F* is calculated by dividing the larger variance $(S_1^2)$ by the smaller one $(S_2^2)$, where $S^2$ is calculated as

$$S^2 = \frac{\left[N\sum X^2 - \left(\sum X\right)^2\right]/N}{N-1}$$

where $N$ is the number of data in the group and $X$ represents the individual values within the group. Frequently, $S^2$ values may be obtained from ANOVA calculations.

The calculated *F* value is compared to the appropriate number in an *F*-value table for the appropriate degrees of freedom $(N - 1)$ in the numerator (along the top of the table) and in the denominator (along the side of the table). If the calculated value is smaller, it is not significant and the variances are considered homogeneous (and the Student's *t* test would be appropriate for further comparison). If the calculated *F* value is greater, *F* is significant and the variances are heterogeneous (and the next test would be a modified Student's *t* test if $N_1 = N_2$ or the Cochran *t* test if $N_1 \neq N_2$; see Figure 2 to review the decision tree).

**ASSUMPTIONS AND LIMITATIONS**

1. This test could be considered as a two-group equivalent of the Bartlett's test.
2. If the test statistic is close to 1.0, the results are (of course) not significant.
3. The test assumes normality and independence of data.

### 7.4 Analysis of Variance

The ANOVA is used for comparison of three or more groups of continuous data when the variances are homogeneous and the data are independent and normally distributed.

A series of calculations are required for ANOVA, starting with the values within each group being added ($\Sigma X$) and then these sums being added ($\Sigma\Sigma X$).

Each figure within the groups is squared, and these squares are then summed ($\Sigma X^2$) and these sums added ($\Sigma\Sigma X^2$). Next the "correction factor" (CF) can be calculated from the formula

$$\mathrm{CF} = \frac{\left(\sum_1^K \sum_1^N X\right)^2}{N_1 + N_2 + \cdots + N_k}$$

where $N$ is the number of values in each group and $K$ is the number of groups. The total sum of squares (SS) is then determined as

$$\mathrm{SS}_{\mathrm{total}} = \sum_1^K \sum_1^N X^2 - \mathrm{CF}$$

In turn, the sum of squares between groups (bg) is found from

$$\mathrm{SS}_{\mathrm{bg}} = \frac{\left(\sum X_1\right)^2}{N_1} + \frac{\left(\sum X_2\right)^2}{N_2} + \cdots + \frac{\left(\sum X_k\right)^2}{N_k} - \mathrm{CF}$$

The within-group (wg) sum of squares is then the difference between the last two figures:

$$\mathrm{SS}_{\mathrm{wg}} = \mathrm{SS}_{\mathrm{total}} - \mathrm{SS}_{\mathrm{bg}}$$

Now, there are three types of degrees of freedom to determine. The first, total df, is the total number of data within all groups under analysis minus 1 ($N_1 + N_2 + \ldots + N_k - 1$). The second figure (the df between groups) is the number of groups minus 1($K - 1$). The last figure (the df within groups or "error df") is the difference between the first two figures ($\mathrm{df}_{\mathrm{total}} - \mathrm{df}_{\mathrm{bg}}$).

The next set of calculations requires determination of the two mean squares ($\mathrm{MS}_{\mathrm{bg}}$ and $\mathrm{MS}_{\mathrm{wg}}$). These are the respective sum-of-square values divided by the corresponding df figures (MS = SS/df). The final calculation is that of the $F$ ratio. For this, the MS between groups is divided by the MS within groups ($F = \mathrm{MS}_{\mathrm{bg}}/\mathrm{MS}_{\mathrm{wg}}$).

A table of the results of these calculations would appear as follows:

| | df | SS | MS | $F$ |
|---|---|---|---|---|
| bg | 3 | 0.04075 | 0.01358 | 4.94 |
| wg | 12 | 0.03305 | 0.00275 | |
| Total | 15 | 0.07380 | | |

For interpretation, the $F$-ratio value obtained in the ANOVA is compared to a table of $F$ values. If $F \leq 1.0$, the results are not significant and comparison

with the table values is not necessary. The df for the greater mean square ($MS_{bg}$) are indicated along the top of the table. Then one must read down the side of the table to the line corresponding to the df for the lesser mean square ($MS_{wg}$). The figure shown at the desired significance level (traditionally 0.05) is compared to the calculated $F$ value. If the calculated number is smaller, there is no significant differences among the groups being compared. If the calculated value is larger, there is some difference but further (post hoc) testing will be required before we know which groups differ significantly.

**ASSUMPTIONS AND LIMITATIONS**

1. What is presented here is the workhorse of toxicology—the one-way ANOVA. Many other forms exist for more complicated experimental designs.
2. The test is robust for moderate departures from normality if the sample sizes are large enough. Unfortunately, this is rarely the case in toxicology.
3. ANOVA is robust for moderate departures from equality of variances (as determined by Bartlett's test) if the sample sizes are approximately equal.
4. It is not appropriate to use a $t$ test (or a two-groups-at-a-time version of ANOVA) to identify where significant differences are within the design group. A multiple-comparison post hoc method must be used.

## 7.5 Post Hoc Tests

There is a wide variety of post hoc tests available to analyze data after finding significant result in an ANOVA. Each of these tests has advantages and disadvantages and proponents and critics. Four of the tests are commonly used in toxicology and will be presented or previewed here. These are Dunnett's $t$ test and Williams' $t$ test. Two other tests which are available in many statistical packages are Tukey's method and the Student–Newman–Keuls method (Zar, 1974, pp. 151–161).

If ANOVA reveals no significance, it is not appropriate to proceed to perform a post hoc test in the hope of finding differences. To do so would only be another form of multiple comparisons, increasing the type I error rate beyond the desired level.

## 7.6 Duncan's Multiple-Range Test

Duncan's (1955) is used to compare groups of continuous and randomly distributed data (body weights, organ weights, etc.). The test normally involves

three or more groups taken one pair at a time. It should only follow observation of a significant $F$ value in the ANOVA and can serve to determine which group (or groups) differs significantly from which other group (or groups).

There are two alternative methods of calculation. The selection of the proper one is based on whether the number of data ($N$) are equal or unequal in the groups.

***Groups with Equal Number of Data ($N_1 = N_2$)*** Two sets of calculations must be carried out: first, the determination of the difference between the means of pairs of groups and, second, the preparation of a probability rate against which each difference in means is compared (as shown in the first of the two examples in this section).

The means (averages) are determined (or taken from the ANOVA calculation) and ranked in either decreasing or increasing order. If two means are the same, they take up two equal positions (thus, for four means we could have ranks of 1, 2, 2, and 4 rather than 1, 2, 3, and 4). The groups are then taken in pairs and the differences between the means ($\bar{x}_1 - \bar{x}_2$), expressed as positive numbers, are calculated. Usually, each pair consists of a test group and the control group though multiple test groups may be compared if so desired. The relative rank of the two groups being compared must be considered. If a test group is ranked 2 and the control group is ranked 1, then we say that there are two places between them, while if the test group were ranked 3, then there would be three places between it and the control.

To establish the probability table, the SEM must be calculated as presented earlier or as

$$\sqrt{\frac{\text{error mean square}}{N}} = \sqrt{\frac{\text{mean square within group}}{N}}$$

where $N$ is the number of animals or replications per dose level. The mean square within groups ($MS_{wg}$) can be calculated from the information given in the ANOVA procedure (refer to the earlier section on ANOVA). The SEM is then multiplied by a series of table values (Harter, 1960; Beyer, 1976a,b) to set up a probability table. The table values used for the calculations are chosen according to the probability levels (note that the tables have sections for 0.05, 0.01, and 0.001 levels) and the number of means apart for the groups being compared and the number of "error" df. The error df is the number of df within the groups. This last figure is determined from the ANOVA calculation and can be taken from ANOVA output. For some values of df, the table values are not given and should thus be interpolated.

***Groups with Unequal Numbers of Data ($N_1 \neq N_2$)*** This procedure is very similar to that discussed above. As before, the means are ranked and the differences between the means are determined ($\bar{x}_1 - \bar{x}_2$). Next, weighing values

($a_{ij}$ values) are calculated for the pairs of groups being compared in accordance with

$$a_u = \sqrt{\frac{2N_iN_j}{(N_i + N_j)}} = \sqrt{\frac{2N_1N_2}{(N_1 + N_2)}}$$

This weighting value for each pair of groups is multiplied by $\bar{x}_1 - \bar{x}_2$ for each value to arrive at a $t$ value. It is the $t$ that will later be compared to a probability table.

The probability table is set up as before except that, instead of multiplying the appropriate table values by SEM, SEM[2] is used. This is equal to $\sqrt{\mathrm{MS}_{\mathrm{wg}}}$.

For the desired comparison of two groups at a time, either the $(\bar{x}_1 - \bar{x}_2)$ value (if $N_1 = N_2$) is compared to the appropriate probability table. or if $N_1 \neq N_2$, the $\bar{x}_1 - 1 - \bar{x}_2$ value is used. Each comparison must be made according to the number of places between the means. If the table value is larger at the 0.05 level, the two groups are not considered to be statistically different. If the table value is smaller, the groups are different and the comparison is repeated at lower levels of significance. Thus, the degree of significance may be determined. We might have significant differences at 0.05 but not at 0.01, in which case the probability would be represented at $0.05 > p > 0.01$.

**ASSUMPTIONS AND LIMITATIONS**

1. Duncan's assures a set $\alpha$ level or type I error rate for all tests when means are separated by no more than ordered step increases. Preserving this $\alpha$ level means that the test is less sensitive than some others, such as the Student–Newman–Keuls. The test is inherently conservative and not resistant or robust.

## 7.7 Scheffé's Multiple Comparisons

Scheffé's is another post hoc comparison method for groups of continuous and randomly distributed data. It also normally involves three or more groups (Scheffé, 1959; Harris, 1975). It is widely considered a more powerful significance test than Duncan's.

Each post hoc comparison is tested by comparing an obtained test value ($F_{\mathrm{contr}}$) with the appropriate critical $F$ value at the selected level of significance (the table $F$ value multiplied by $K - 1$ for an $F$ with $K - 1$ and $N - K$ degrees of freedom[2]), where $F_{\mathrm{contr}}$ is computed as follows:

(a) Compute the mean for each sample (group).
(b) Denote the residual mean square by $\mathrm{MS}_{\mathrm{wg}}$.
(c) Compute the test statistic as

$$F_{\text{contr}} = \frac{\left(C_1\overline{X}_1 + C_2\overline{X}_2 + \cdots + C_k\overline{X}_k^{12}\right)}{(K-1)\text{MS}_{\text{wg}}\left(C_1^2/n_1 + \cdots + C_k^2/n_k\right)}$$

where $C_k$ is the comparison number such that the sum of $C_1, C_2, \ldots, C_k = 0$.

**ASSUMPTIONS AND LIMITATIONS**

1. The Scheffé procedure is robust to moderate violations of the normality and homogeneity of variance assumptions.
2. It is not formulated on the basis of groups with equal numbers (as one of Duncan's procedures is), and if $N_1 \neq N_2$ there is no separate weighting procedure.
3. It tests all linear contrasts among the population means (the other three methods confine themselves to pairwise comparison, except they use a Bonferroni-type correlation procedure).
4. The Scheffé procedure is powerful because of its robustness, yet it is very conservative. Type I error (the false-positive rate) is held constant at the selected test level for each comparison.

### 7.8 Dunnett's *t* Test

Dunnett's *t* test (Dunnett, 1955, 1964) has as its starting point the assumption that what is desired is a comparison of each of several means with one other mean and only one other mean; in other words, one wishes to compare each and every treatment group with the control group but not compare treatment groups with each other. The problem here is that, in toxicology, one is frequently interested in comparing treatment groups with other treatment groups. However, if one wants to only compare treatment groups with a control group, Dunnett's is a useful approach. In a study with $K$ groups (one of them being the control) we will wish to make $K - 1$ comparisons. In such a situation, we want to have a $P$ level for the entire set of $K - 1$ decisions (not for each individual decision). The Dunnett's distribution is predicated on this assumption. The parameters for utilizing a Dunnett's table, such as found in his original article, are $K$ (as above) and the number of degrees of freedom for the within-group mean square ($\text{MS}_{\text{wg}}$). The test value is calculated as

$$t = \frac{|T_j - T_i|}{\sqrt{2\text{MS}_{\text{wg}/n}}}$$

where $n$ is the number of observations in each of the groups. The within-group mean square ($\text{MS}_{\text{wg}}$) is as we have defined it previously; $T_j$ is the control group mean and $T_i$ is the mean of, in order, each successive test group observation. Note that one uses the absolute value of the positive number resulting from subtracting $T_i$ from $T_j$. This is to ensure a positive number for our final $t$.

**ASSUMPTIONS AND LIMITATIONS**

1. Dunnett's seeks to ensure that the type 1 error rate will be fixed at the desired level by incorporating correction factors into the design of the test value table.
2. Treated group sizes must be approximately equal.

### 7.9 Williams *t* Test

The Williams *t* test (Williams, 1971, 1972) is popular, although its use is quite limited in toxicology. It is designed to detect the highest level (in a set of dose/exposure levels) at which there is no significant effect. It assumes that the response of interest (such as change in body weights) occurs at higher levels, but not at lower levels, and that the responses are monotonically ordered so that $X_0 \leq X_1 \leq \ldots \leq X_k$. This is, however, frequently not the case. The Williams technique handles the occurrence of such discontinuities in a response series by replacing the offending value and the value immediately preceding it with weighted average values. The test also is adversely affected by any mortality at high dose levels. Such mortalities "impose a severe penalty, reducing the power of detecting an effect not only at level *K* but also at all lower doses" (Williams, 1972, p. 529). Accordingly, it is not generally applicable in toxicology studies.

### 7.10 Analysis of Covariance

Analysis of covariance (ANCOVA) is a method for comparing sets of data which consist of two variables (treatment and effect, with our effect variable being called the "variate") when a third variable (called the "covariate") exists which can be measured but not controlled and which has a definite effect on the variable of interest. In other words, it provides an indirect type of statistical control, allowing us to increase the precision of a study and to remove a potential source of bias. One common example of this is in the analysis of organ weights in toxicity studies. Our true interest here is the effect of our dose or exposure level on the specific organ weights, but most organ weights also increase (in the young, growing animals most commonly used in such studies) in proportion to increases in animal body weight. As we are not here interested in the effect of this covariate (body weight), we measure it to allow for adjustment. We must be careful before using ANCOVA, however, to ensure that the underlying nature of the correspondence between the variate and covariate is such that we can rely on it as a tool for adjustments (Anderson et al., 1980; Kotz and Johnson, 1982).

Calculation is performed in two steps. The first is a type of linear regression between the variate $Y$ and the covariate $X$.

This regression, performed as described under the linear regression section, gives us the model

$$Y = a_1 + BX + e$$

which in turn allows us to define adjusted means ($\overline{Y}$ and $X$) such that $\overline{Y}_{1a} = \overline{Y}_1 - (\overline{x}_1 - X^*)$.

If we consider the case where $K$ treatments are being compared such that $K = 1, 2, \ldots, K$ and we let $X_{ik}$ and $Y_{ik}$ represent the predictor and predicted values for each individual $i$ in group $k$, we can let $X_k$ and $Y_k$ be the means. Then, we define the between-group (for treatment) sum of squares and cross products as

$$T_{xx} = \sum_{k-1}^{K} n_k \left(\overline{X}_K - \overline{X}\right)^2$$

$$T_{yy} = \sum_{k-1}^{K} n_k \left(\overline{Y}_K - \overline{Y}\right)^2$$

$$T_{xy} = \sum_{k-1}^{K} n_k \left(\overline{X}_k - \overline{X}\right)\left(\overline{Y}_k - \overline{Y}_k - \overline{Y}\right)$$

In a like manner, within-group sums of squares and cross products are calculated as

$$\sum xx = \sum_{k=1}^{k} \sum_{i} (X_{ik} - X_k)^2$$

$$\sum yy = \sum_{K=1}^{k} \sum_{i} (Y_{ik} - Y_k)^2$$

$$\sum xy = \sum_{k=1}^{k} \sum_{i}^{n} (X_{ik} - X_k)(Y_{ik} - Y_k)$$

where $i$ indicates the sum from all the individuals within each group and $n$ is the total number of subjects minus number of groups:

$$S_{xx} = T_{xx} + \Sigma_{xx} \qquad S_{yy} = T_{yy} + \Sigma_{xx} \qquad S_{xy} = T_{xy} + \Sigma_{xy}$$

With these in hand, we can then calculate the residual mean squares of treatments ($\mathrm{St}^2$) and error ($\mathrm{Se}^2$):

$$\mathrm{St}^2 = \frac{T_{yy} - \dfrac{S_{xy}^2}{S_{xx}} + \dfrac{\Sigma_{xy}^2}{\Sigma_{xx}}}{lc - 1} \qquad \mathrm{Se}^2 = \frac{\Sigma_{yy} - \Sigma_y^2 / \Sigma_{xx}}{f - 1}$$

*The slight difference in the results for the two approaches is due to rounding errors. It is not important biologically.

These can be used to calculate an $F$ statistic to test the null hypothesis that all treatment effects are equal:

$$F = \frac{\mathrm{St}^2}{\mathrm{Se}^2}$$

The estimated regression coefficient of $Y$ or $X$ is

$$B = \frac{\Sigma_{xy}}{\Sigma_{xx}}$$

The estimated standard error for the adjusted difference between two groups is given by

$$\mathrm{Sd} = \mathrm{Se}\frac{1}{n_j} + \frac{1}{n_j} + \frac{(X_i - X_j)^2}{\Sigma_{xx}}$$

where $n_0$ and $n_1$ are the sample sizes of the two groups. A test of the null hypothesis that the adjusted differences between the groups is zero is provided by

$$t = \frac{Y_1 - Y_0 - B(X_1 - X_0)}{\mathrm{Sd}}$$

The test value for $t$ is then looked up in the $t$ table with $f - 1$ degrees of freedom. Computation is markedly simplified if all the groups are of equal size.

**ASSUMPTIONS AND LIMITATIONS**

1. The underlying assumptions for ANCOVA are fairly rigid and restrictive. The assumptions include:
   a. That the slopes of the regression lines of $Y$ and $X$ are equal from group to group. This can be examined visually or formally (i.e., by a test). If this condition is not met, ANCOVA cannot be used.
   b. That the relationship between $X$ and $y$ is linear.
   c. That the covariate $X$ is measured without error. The power of the test declines as error increases.
   d. That there are no unmeasured confounding variables.
   e. That the errors inherent in each variable are independent of each other. Lack of independence effectively (but to an immeasurable degree) reduces sample size.
   f. That the variances of the errors in groups are equivalent between groups.
   g. That the measured data which form the groups are normally distributed. ANCOVA is generally robust to departures from normality.
2. Of the seven assumptions above, the most serious are the first four.

## 7.11 Modeling

The mathematical modeling of biological systems, restricted even to the field of toxicology, is an extremely large and vigorously growing area. Broadly speaking, modeling is the principal conceptual tool by which toxicology seeks to develop as a mechanistic science. In an iterative process, models are developed or proposed, tested by experiment, reformulated, and so on, in a continuous cycle. Such a cycle could also be described as two related types of modeling: explanatory (where the concept is formed) and correlative (where data are organized and relationships derived). An excellent introduction to the broader field of modeling of biological systems can be found in Gold (1977).

In toxicology, modeling is of prime interest in seeking to relate a treatment variable with an effect variable and, from the resulting model, predict effects at exact points where no experiment has been done (but in the range where we have performed experiments, such as "determining" $LD_{50}$), to estimate how good our prediction is, and, occasionally, to simply determine if a pattern of effects is related to a pattern of treatment.

For use in prediction, the techniques of linear regression, probit/logit analysis (a special case of linear regression), moving averages (an efficient approximation method), and nonlinear regression (for doses where data cannot be made to fit a linear pattern) are presented. For evaluating the predictive value of these models, both the correlation coefficient (for parametric data) and Kendall's rank correlation (for nonparametric data) are given. And finally, the concept of trend analysis is introduced and a method presented.

When we are trying to establish a pattern between several data points (whether this pattern is in the form of a line or a curve), what we are doing is interpolating. It is possible for any given set of points to produce an infinite set of lines or curves which pass near (for lines) or through (for curves) the data points. In most cases, we cannot actually know the "real" pattern. Se we apply a basic principle of science—Occam's razor. We use the simplest explanation (or, in this case, model) which fits the facts (or data). A line is, of course, the simplest pattern to deal with and describe, so fitting the best line (linear regression) is the most common form of model in toxicology.

## 7.12 Linear Regression

Foremost among the methods for interpolating within a known data relationship is regression—the fitting of a line or curve to a set of known data points on a graph and the interpolation ("estimation") of this line or curve in areas where we have no data points. The simplest of these regression models is that of linear regression (valid when increasing the value of one variable changes the value of the related variable in a linear fashion, either positively

or negatively). This is the case we will explore here using the method of least squares.

Given that we have two sets of variables, $x$ (say milligrams per kilogram of test material administered) and $y$ (say percentage of animals so dosed that die), what is required is solving for $a$ and $b$ in the equation $Y_i = a + bx_i$ [where the uppercase $Y_i$ is the fitted value of $y_i$ at $x_i$ and we wish to minimize $(y_i - Y_i)^2$]. So we solve the equations

$$b = \frac{\sum x_1 y_1 - n\bar{x}\bar{y}}{\sum x_1^2 - n\bar{x}^2} \qquad a = \bar{y} - b\bar{x}$$

where $a$ is the $y$ intercept, $b$ is the slope of the time, and $n$ is the number of data points.

Note that in actuality dose–response relationships are often not linear and instead we must use either a transform (to linearlize the data) or a nonlinear regression method (Gallant, 1975).

Note also that we can use the correlation test statistic (to be described in the correlation coefficient section, 7.15) to determine if the regression is significant (and therefore valid at a defined level of certainty. A more specific test for significance would be the linear regression ANOVA (Pollard, 1977). To do so we start by developing the appropriate ANOVA table.

Finally, we might wish to determine the confidence intervals for our regression line; that is, given a regression line with calculated values for $Y_i$ given $x_i$, within what limits may we be certain (with say a 95% probability) what the real value of $Y_i$ is?

If we denote the residual mean square in the ANOVA by $s^2$, the 95% confidence limits for $a$ (denoted by $A$, the notation for the true—as opposed to the estimated—value for this parameter) are calculated as

$$t_{n-2} = \frac{a - A}{\sqrt{s^2\left(\sum x^2\right)/\left(n\sum x_1^2 - n^2\bar{x}^2\right)}}$$

$$\frac{9.2 - A}{\sqrt{8.8(51)/[4(51) - (16)(10.562)]}} = \frac{9.2 - A}{\sqrt{448/35.008}}$$

$$= \frac{9.2 - A}{3.58} = -4.303$$

$$9.2 - A = -15.405$$

$$A = 9.2 - 15.405$$

**ASSUMPTIONS AND LIMITATIONS**

1. All the regression methods are for interpolation, not extrapolation. That is, they are valid only in the range that we have data—the experimental region. Not beyond.
2. The method assumes that the data are independent and normally distributed, and it is sensitive to outliers. The $x$-axis (or horizontal) component plays an extremely important part in developing the least-squares fit. All points have equal weight in determining the height of a regression line, but extreme $x$-axis values unduly influence the slope of the line.
3. A good fit between a line and a set of data (i.e., a strong correlation between treatment and response variables) does not imply any casual relationship.
4. It is assumed that the treatment variable can be measured without error, that each data point is independent, that variances are equivalent, and that a linear relationship does not exist between the variables.
5. There are many excellent texts on regression, which is a powerful technique. See, for example, Draper and Smith (1981) and Montgomery and Smith (1983), which are not overly rigorous mathematically.

## 7.13 Probit/Log Transforms and Regression

As we noted in the preceding section, dose–response problems (among the most common interpolation problems encountered in toxicology) rarely are straightforward enough to make a valid linear regression directly from the raw data. The most common valid interpolation methods are based upon probability ("probit") and logarithmic ("log") value scales, with percentage responses (death, tumor incidence, etc.) being expressed on the probit scale while doses ($Y_i$) are expressed on the log scale. There are two strategies for such an approach. The first is based on transforming the data to these scales, then doing a weighted linear regression on the transformed data (if one does not have access to a computer or a high-powered programmable calculator, the only practical strategy is not to assign weights). The second requires the use of algorithms (approximate calculation techniques) for the probit value and regression process and is extremely burdensome to perform manually.

Our approach to the first strategy requires that we construct a table with the pairs of values of $x_i$ and $y_i$ listed in order of increasing values of $Y_i$ (percentage response). Beside each of these columns a set of blank columns should be left so that the transformed values may be listed. We then simply add the columns described in the linear regression procedure. Log and probit values may be taken from any of a number of sets of tables (such as provided in Appendix I) and the rest of the table is then developed from these transformed $x_i'$ and $y_i'$ values (denoted as $x_i'$ and $y_i'$). A standard linear regression is then performed.

The second strategy we discussed has been broached by a number of authors (Bliss, 1935; Finney, 1977; Litchfield and Wilcoxon, 1949; Prentice, 1976). All of these methods, however, are computationally cumbersome. It is possible to approximate the necessary iterative process using the algorithms developed by Abramowitz and Stegun (1964) but even this merely reduces the complexity to a point where the procedure may be readily programmed on a small computer or programmable calculator.

**ASSUMPTIONS AND LIMITATIONS**

1. The probit distribution is derived form a common error function, with the midpoint (50% point) moved to a score of 5.00.
2. The underlying frequency distribution becomes asymptotic as it approaches the extremes of the range. That is, in the range of 16–84%, the corresponding probit values change gradually—the curve is relatively linear. But beyond this range, they change ever more rapidly as they approach either 0 or 100%. In fact, there are no values for either of these numbers.
3. A normally distributed population is assumed, and the results are sensitive to outliers.

## 7.14 Nonlinear Regression

More often than not in toxicology we find that our data demonstrate a relationship between two variables (such as age and body weight) which is not linear. That is, a change in one variable (say age) does not produce a directly proportional change in the other (e.g., body weight). But some form of relationship between the variables is apparent. If understanding such a relationship and being able to predict unknown points is of value, we have a pair of options available to us. The first, which was discussed and reviewed earlier, is to use one or more transformations to linearize our data and then to make use of linear regression. This approach, though most commonly used, has a number of drawbacks. Not all data can be suitably transformed; sometimes the transformations necessary to linearize the data require a cumbersome series of calculations, and the resulting linear regression is not always sufficient to account for the differences among sample values—there are significant deviations around the linear regression line (i.e., a line may still not give us a good fit to the data or do an adequate job of representing the relationship between the data). In such cases, we have available a second option—the fitting of data to some nonlinear function such as some form of the curve. This is, in general form, nonlinear regression and may involve fitting data to an infinite number of possible functions. But most often we are interested in fitting curves to a polynomial function of the general form

$$Y = a + bx + cx^2 + dx^2 + \cdots$$

where $x$ is the independent variable. As the number of powers of $x$ increases, the curve becomes increasingly complex and will be able to fit a given set of data increasingly well.

Generally in toxicology, however, if we plot the log of a response (such as body weight) versus a linear scale of our dose or stimulus, we get one of four types of nonlinear curves (Snedecor and Cochran, 1980):

1. Exponential growth, where $\log Y = A(Bx)$, such as the growth curve for the log phase of a bacterial culture.
2. Exponential decay, where $\log Y = A(B^{-x})$, such as a radioactive decay curve.
3. Asymptotic regression, where $\log Y = A - B(p^x)$, such as a first-order reaction curve.
4. Logistic growth curve, where $\log Y = A/(1 + Bp^x)$, such as a population growth curve.

In all these cases, $A$ and $B$ are constant while $p$ is a log transform. These curves are illustrated in Figure 7.

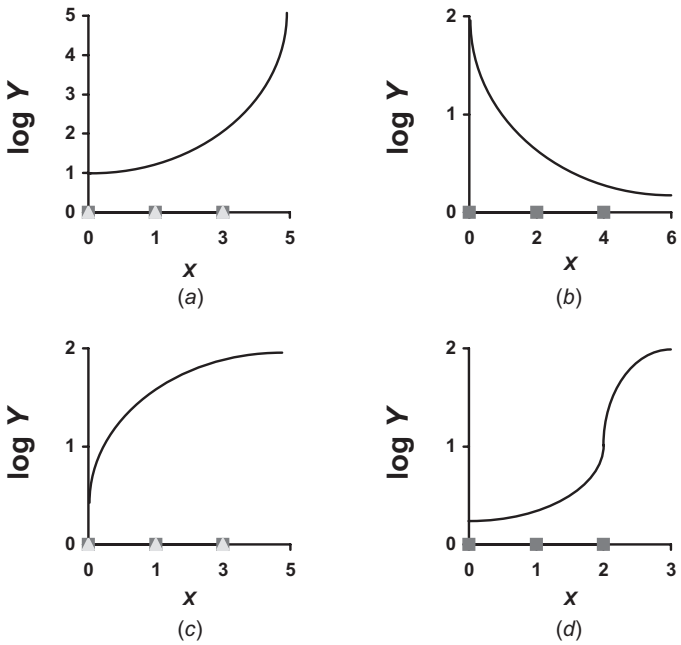


***Figure 7*** *Common Curvilinear Curves: (a) exponential growth law, $\log Y = A(B^x)$; (b) exponential decay law, $\log Y = A(B^{-x})$; (c) asymptotic regression, $\log Y = A - B(\psi^x)$; (d) logistic growth law, $\log Y = A/(1 + B\psi^x)$.*

All four types of curves are fit by iterative processes—that is, best guess numbers are initially chosen for each of the constants and, after a fit is attempted, the constants are modified to improve the fit. This process is repeated until an acceptable fit has been generated. Analysis of variance or covariance can be used to objectively evaluate the acceptability of it. Needless to say, the use of a computer generally accelerates such a curve-fitting process.

**ASSUMPTIONS AND LIMITATIONS**

1. The principle of using least squares may still be applicable in fitting the best curve if the assumptions of normality, independence, and reasonably error-free measurement of response are valid.
2. Growth curves are best modeled using a nonlinear method.

## 7.15 Correlation Coefficient

The correlation procedure is used to determine the degree of linear correlation (direct relationship) between two groups of continuous (and normally distributed) variables; it will indicate whether there is any statistical relationship between the variables in the two groups. For example, we may wish to determine if the liver weights of dogs on a feeding study are correlated with their body weights. Thus, we will record the body and liver weights at the time of sacrifice and then calculate the correlation coefficient between these pairs of values to determine if there is some relationship.

A formula for calculating the linear correlation coefficient ($r_{xy}$) is as follows:

$$r_{xy} = \frac{N\sum XY - (\Sigma X)(\Sigma Y)}{\sqrt{N\Sigma X^2 - (\Sigma X)^2}\sqrt{N\Sigma Y^2 - (\Sigma Y)^2}}$$

where $X$ is each value for one variable (such as the dog body weights in the above example), $Y$ is the matching value for the second variable (the liver weights), and $N$ is the number of pairs of $X$ and $Y$. Once we have obtained $r_{xy}$, it is possible to calculate $t_r$, which can be used for more precise examination of the degree of significant linear relationship between the two groups. This value is calculated as follows:

$$t_r = \frac{r_{zy}\sqrt{N-2}}{\sqrt{1-r_{zy}^2}}$$

This calculation is also equivalent to $r$ = sample covariance/($S_xS_y$), as was seen earlier under ANCOVA.

The value obtained for $r_{xy}$ can be compared to table values (Snedecor and Cochran, 1980) for the number of pairs of data involved minus 2. If the $r_{xy}$ is

smaller (at the selected test probability level, such as 0.05), the correlation is not significantly different from zero (no correlation). If $r_{xy}$ is larger than the table value, there is a positive statistical relationship between the groups. Comparisons are then made at lower levels of probability to determine the degree of relationship (note that if $r_{xy}$ is either 1.0 or –1.0, there is complete correlation between the groups). If $r_{xy}$ is a negative number and the absolute is greater than the table value, there is an inverse relationship between the groups; that is, a change in one group is associated with a change in the opposite direction in the second group of variables.

Since the comparison of $r_{xy}$ with the table values may be considered a somewhat weak test, it is perhaps more meaningful to compare the $t_r$ value with values in a $t$-distribution table for $N - 2$ degrees of freedom (df), as is done for the Student's $t$ test. This will give a more exact determination of the degree of statistical correlation between the two groups.

Note that this method examines only possible linear relationships between sets of continuous, normally distributed data.

**ASSUMPTIONS AND LIMITATIONS**

1. A strong correlation does not imply that a treatment causes an effect.
2. The distances of data points from the regression line are the portions of the data not "explained" by the model. These are called residuals. Poor correlation coefficients imply high residuals, which may be due to many small contributions (variations of data from the regression line) or a few large ones. Extreme values (outliers) greatly reduce correlation.
3. $X$ and $Y$ are assumed to be independent.
4. Feinstein (1979) has provided a fine discussion of the difference between correlation (or association of variables) and causation.

## 7.16 Kendall's Coefficient of Rank Correlation

Kendall's rank correlation, represented by $\tau$ (tau), should be used to evaluate the degree of association between two sets of data when the nature of the data is such that the relationship may not be linear. Most commonly, this is when the data are not continuous and/or normally distributed. An example of such a case is when we are trying to determine if there is a relationship between the length of hydra and their survival time in a test medium in hours. Both of our variables here are discontinuous, yet we suspect a relationship exists. Another common use is in comparing the subjective scoring done by two different observers.

Tau is calculated as $\tau = N/n(n - 1)$ where $n$ is the sample size and $N$ is the count of ranks, calculated as $N = 4({}^nC_i) - n(n - 1)$, with the computing of ${}^nC_i$ being demonstrated in the example.

If a second variable $Y_2$ is exactly correlated with the first variable $Y_1$, then the variates $Y_2$ should be in the same order as the $Y_1$ variates. However, if the correlation is less than exact, the order of the variates $Y_2$ will not correspond entirely to that of $Y$. The quantity $N$ measures how well the second variable corresponds to the order of the first. It has maximum value of $n(n - 1)$ and a minimum value of $-n(n - 1)$.

A table of data is set up with each of the two variables being ranked separately. Tied ranks are assigned as demonstrated earlier under the Kruskall–Wallis test. From this point, disregard the original variates and deal only with the ranks. Place the ranks of one of the two variables in rank order (from lowest to highest) paired with the rank values assigned for the other variable. If one (but not the other) variable has tied ranks, order the pairs by the variables without ties (Sokal and Rohlf, 1994).

The resulting value of tau will range from –1 to +1, as does the familiar parametric correlation coefficient $r$.

**ASSUMPTION AND LIMITATION**

1. A very robust estimator which does not assume normality, linearity, or minimal error of measurement.

## 7.17 Trend Analysis

Trend analysis is a collection of techniques that have been "discovered" by toxicology since the mid-1970s (Tarone, 1975). The actual methodology dates back to the mid-1950s (Cox and Stuart, 1955).

Trend analysis methods are a variation on the theme of regression testing. In the broadest sense, the methods are used to determine whether a sequence of observations taken over an ordered range of values of a variable (most commonly time) exhibit some form of pattern of change (e.g., an upward trend) associated with another variable of interest (in toxicology, some form or measure of dosage).

Trend corresponds to sustained and systematic variations over a long period of time. It is associated with the structural causes of the phenomenon in question, for example, population growth, technological progress, new ways of organization, or capital accumulation.

The identification of trend has always posed a serious statistical problem. The problem is not one of mathematical or analytical complexity but of conceptual complexity. This problem exists because the trend as well as the remaining components of a time series are latent (nonobservable) variables and, therefore, assumptions must be made on their behavioral pattern. The trend is generally thought of as a smooth and slow movement over a long term. The concept of "long" in this connection is relative and what is identified as trend for a given series span might well be part of a long cycle once the series is considerably augmented. Often, a long cycle is treated as a trend because

the length of the observed time series is shorter than one complete face of this type of cycle.

The ways in which data are collected in toxicology studies frequently serve to complicate trend analysis, as the length of time for the phenomena underlying a trend to express themselves is frequently artificially censored.

To avoid the complexity of the problem posed by a statistically vague definition, statisticians have resorted to two simple solutions: One consists of estimating trend and cyclical fluctuations together, calling this combined movement a *trend cycle*; the other consists of defining the trend in terms of the series length, denoting it as the longest nonperiodic movement.

Within the large class of models identified for trend, we can distinguish two main categories: deterministic trends and stochastic trends.

Deterministic trend models are based on the assumption that the trend of a time series can be approximated closely by simple mathematical functions of time over the entire span of the series. The most common representation of a deterministic trend is by means of polynomials or of transcendental functions. The time series from which the trend is to be identified is assumed to be generated by a nonstationary process where the nonstationarity results from a deterministic trend. A classical model is the regression or error model (Anderson, 1971) where the observed series is treated as the sum of a systematic part or trend and a random or irregular part. This model can be written as

$$Z_t = Y_t + U_t'$$

where $U_t$ is a purely random process, that is, $U_t$ ~i.i.d. (0, 2/$u$) (independent and identically distributed with expected value zero and variance 2/$u$).

Trend tests are generally described as $k$-sample tests of the null hypothesis of identical distribution against an alternative of linear order; that is, if sample $I$ has distribution function $F_i$, $i = 1$, then the null hypothesis

$$H_0 : F_1 = F_2 - \cdots = F_k$$

is tested against the alternative

$$H_1 : F_1 \geq F_2 \geq \cdots = F_k$$

(or its reverse), where at least one of the inequalities is strict. These tests can be thought of as special cases of tests of regression or correlation in which association is sought between the observations and its ordered sample index. They are also related to ANOVA except that the tests are tailored to be powerful against the subset of alternatives $H_1$, instead of the more general set $\{F_1 \neq F_j, \text{some } i \neq j\}$.

Different tests arise from requiring power against specific elements or subsets of this rather extensive set of alternatives.

The most popular trend test in toxicology is currently that presented by Tarone in 1975 because it is that used by the NCI in the analysis of carcinogenicity data. A simple, but efficient alternative is the Cox and Stuart (1955) test, which is a modification of the sign test. For each point at which we have a measure (such as the indidence of animals observed with tumors) we form a pair of observations—one from each of the groups we wish to compare. In a traditional NCI bioassay this would mean pairing control with low dose and low dose with high dose (to explore a dose-related trend) or each time period observation in a dose group (except the first) with its predecessor (to evaluate time-related trend). When the second observation in a pair exceeds the earlier observation, we record a plus sign for that pair. When the first observation is greater than the second, we record a minus sign for that pair. A preponderance of plus signs suggests a downward trend while an excess of minus signs suggests an upward trend. A formal test at a preselected confidence level can then be performed.

More formally put, after having defined what trend we want to test for, we first match pairs as $(X_1 - X_1 + c)$, $(X_2, X_2 + c)$, … $(X_{n'-c}, X_{n'})$, where $c = n'/2$ when $n'$ is even and $c = (n' + 1)/2$ when $n'$ is odd (where $n'$ is the number of observations in a set). The hypothesis is then tested by comparing the resulting number of excess positive or negative signs against a sign test table such as are found in Beyer (1976a,b).

We can, of course, combine a number of observations to allow ourselves to actively test for a set of trends, such as the existence of a trend of increasing difference between two groups of animals over a period of time.

**ASSUMPTIONS AND LIMITATIONS**

1. Trend tests seek to evaluate whether there is a monotonic tendency in response to a change in treatment. That is, the dose–response direction is absolute—as dose goes up, the incidence of tumors increases. Thus the test loses power rapidly in response to the occurrences of "reversals"—for example, a low-dose group with a decreased tumor incidence. There are methods (Dykstra and Robertson, 1983) which "smooth the bumps" of reversals in long data series. In toxicology, however, most data series are short (i.e., there are only a few dose levels).

Tarone's trend test is most powerful at detecting dose-related trends when tumor onset hazard functions are proportional to each other. For more power against other dose-related group differences, weighted versions of the statistic are also available (Breslow, 1984; Crowley and Breslow, 1984).

In 1985, the U.S. *Federal Register* recommended that the analysis of tumor incidence data is carried out with a Cochran–Armitage (Armitage, 1955; Cochran, 1954) trend test. The test statistic of the Cochran–Armitage test is defined as

$$T_{\mathrm{CA}} = \sqrt{\frac{N}{(N-r)r}}, \frac{\sum_{i=0}^{k}[R_1-(n_1/N)r]d_1}{\sqrt{\sum_{i=0}^{k}(n_i/N)d_i^2-\left(\sum_{i=0}^{k}(n_i/N)d_1\right)^2}}$$

with dose scores $d_i$. Armitage's test statistic is the square of this term ($T_{\mathrm{CA}}^2$). As one-sided tests are carried out for an increase of tumor rates, the square is not considered. Instead, the above-mentioned test statistic, which is presented by Portier and Hoel (1984), is used. This test statistic is asymptotically standard normally distributed. The Cochran–Armitage test is asymptotically efficient for all monotone alternatives (Tarone, 1975), but this result only holds asymptotically. And tumors are rare events, so the binominal proportions are small. In this situation approximations may become unreliable.

Therefore, exact tests which can be performed using two different approaches—conditional and unconditional—are considered. In the first case, the total number of tumors $r$ is regarded as fixed. As a result the null distribution of the test statistic is independent of the common probability $p$. The exact conditional null distribution is a multivariate hypergeometric distribution.

The unconditional model treats the sum of all tumors as a random variable. Then the exact unconditional null distribution is a multivariate binomial distribution. The distribution depends on the unknown probability.

## 8 METHODS FOR REDUCTION OF DIMENSIONALITY

Techniques for the reduction of dimensionality are those that simplify the understanding of data, either visually or numerically, while causing only minimal reductions in the amount of information present. These techniques operate primarily by pooling or combining groups of variables into single variables but may also entail the identification and elimination of low-information-content (or irrelevant) variables.

Descriptive statistics (calculations of means, standard deviations, etc.) are the simplest and most familiar form of reduction of dimensionality. Here we first need to address classification, which provides the general conceptual tools for identifying and quantifying similarities and differences between groups of things which have more than a single linear scale of measurement in common (e.g., which have both been determined to have or lack a number of enzyme activities). Then we will consider two collections of methodologies which combine graphic and computational methods, multidimensional/nonmetric scaling, and cluster analysis. Multidimensional scaling (MDS) is a set of techniques for quantitatively analyzing similarities, dissimilarities, and distances between data in a displaylike manner. Nonmetric scaling is an analogous set of methods for displaying and relating data when measurements are nonquantitative (the data are described by attributes or ranks). Cluster analysis is a

collection of graphic and numerical methodologies for classifying things based on the relationships between the values of the variables that they share.

The final pair of methods for reduction of dimensionality which will be tackled in this chapter are Fourier analysis and life table analysis. Fourier analysis seeks to identify cyclic patterns in data and can analyze the patterns or the residuals after the patterns are taken out. Life table analysis techniques are directed to identifying and quantitating the time course of risks (such as death or the occurrence of tumors).

## 8.1 Classification

Classification is both a basic concept and a collection of techniques which are necessary prerequisites for further analysis of data when the members of a set of data are (or can be) each described by several variables. At least some degree of classification (which is broadly defined as the dividing of the members of a group into smaller groups in accordance with a set of decision rules) is necessary prior to any data collection. Whether formally or informally, an investigator has to decide which things are similar enough to be counted as the same and develop rules for governing collection procedures. Such rules can be as simple as "measure and record body weights only of live animals on study" or as complex as that demonstrated by the expanded weighting classification procedure demonstrated below. Such a classification also demonstrates that the selection of which variables to measure will determine the final classification of data.

EXPANDED WEIGHTING PROCEDURE

| | | |
|---|---|---|
| I. | Is animal of desired species? | Yes/no |
| II. | Is animal member of study group? | Yes/no |
| III. | Is animal alive? | Yes/no |
| IV. | Which group does animal belong to? | |
| | A. Control | |
| | B. Low dose | |
| | C. Intermediate dose | |
| | D. High dose | |
| V. | What sex is animal? | Male/female |
| VI. | Is the measured weight in acceptable range? | Yes/no |

Classifications of data have two purposes (Hartigan, 1983; Gordon, 1981): data simplification (also called a descriptive function) and prediction. Simplification is necessary because there is a limit to both the volume and complexity of data that the human mind can comprehend and deal with conceptually. Classification allows us to attach a label (or name) to each group of data, to summarize the data (i.e., assign individual elements of data to groups and to characterize the population of the group), and to define the relationships between groups (i.e., develop a taxonomy).

Prediction, meanwhile, is the use of summaries of data and knowledge of the relationships between groups to develop hypotheses as to what will happen when further data are collected (as when more animals or people are exposed to an agent under defined conditions) and as to the mechanisms which cause such relationships to develop. Indeed, classification is the prime device for the discovery of mechanisms in all of science. A classic example of this was Darwin's realization that there were reasons (the mechanisms of evolution) behind the differences and similarities in species which had caused Linaeus to earlier develop his initial modern classification scheme (or taxonomy) for animals.

To develop a classification, one first sets bounds wide enough to encompass the entire range of data to be considered that is not unnecessarily wide. This is typically done by selecting some global variables (variables every piece of data have in common) and limiting the range of each so that it just encompasses all the cases on hand. Then one selects a set of local variables (characteristics which only some of the cases have, say the occurrence of certain tumor types, enzyme activity levels, or dietary preferences) and which thus serve to differentiate between groups. Data are then collected, and a system for measuring differences and similarities is developed. Such measurements are based on some form of measurement of distance between two cases ($x$ and $y$) in terms of each single-variable scale. If the variable is a continuous one, then the simplest measure of distance between two pieces of data is the Euclidean distance $d(x, y)$ defined as

$$d(x, y) = \sqrt{(x_i - y_i)^2}$$

For categorical or discontinuous data, the simplest distance measure is the matching distance, defined as

$$d(x, y) = \text{number of times } x_i \neq y_i$$

After we have developed a table of such distance measurements for each of the local variables, some weighting factor is assigned to each variable. A weighting factor seeks to give greater importance to those variables which are believed to have more relevance or predictive value. The weighted variables are then used to assign each piece of data to a group. The actual act of developing numerically based classifications and assigning data members to them is the realm of cluster analysis and will be discussed later in this chapter. Classification of biological data based on qualitative factors has been well discussed. Glass (1975) and Schaper et al. (1985) do an excellent job of introducing the entire field and mathematical concepts.

Relevant examples of the use of classification techniques range from the simple to the complex. Schaper et al. (1985) developed and used a very simple classification of response methodology to identify those airborne chemicals which alter the normal respiratory response induced by $CO_2$. At the other end of the spectrum, Kowalski and Bender (1972) developed a more mathemati-

cally based system to classify chemical data (a methodology they termed *pattern recognition*).

## 8.2 Statistical Graphics

The use of graphics in one form or another in statistics is the single most effective and robust statistical tool and at the same time one of the most poorly understood and improperly used.

Graphs are used for one of four major purposes. Each of the four is a variation on the central theme of making complex data easier to understand and use. These four major functions are exploration, analysis, communication and display of data, and graphical aids. Exploration (which may be simply summarizing data or trying to expose relationships between variables) is determining the characteristics of data sets and deciding on one or more appropriate forms of further analysis, such as the scatterplot. Analysis is the use of graphs to formally evaluate some aspect of the data, such as whether there are outliers present or if an underlying assumption of a population distribution is fulfilled. As long ago as 1960 (Anderson), some 18 graphical methods for analyzing multivariate data relationships were developed and proposed. Table 7 presents a summary of major graphical techniques that are available.

Communication and display of data are the most commonly used functions of statistical graphics in toxicology, whether used for internal reports, presentations at meetings, or formal publications in the literature. In communicating data, graphs should not be used to duplicate data that are presented in tables but rather should show important trends and/or relationships in the data. Though such communication is most common in a quantitative compilation of actual data, it can also be used to summarize and present the results of statistical analysis. The fourth and final function of graphics is one that is largely becoming outdated as microcomputers become more widely available. Graphical aids to calculation include nomograms [the classic example in toxicology of a nomogram is that presented by Litchfield and Wilcoxon (1949) for determining median effective doses] and extrapolating and interpolating data graphically based on plotted data.

There are many forms of statistical graphics (a partial list, classified by function, is presented in Table 7), and a number of these (such as scatterplots and histograms) can be used for each of a number of possible functions. Most of these plots are based on a Cartesian system (i.e., they use a set of rectangular coordinates), and our review of construction and use will focus on these forms of graphs.

Construction of a rectangular graph of any form starts with the selection of the appropriate form of graph followed by the laying out of the coordinates (or axes). Even graphs which are going to encompass multivariate data (i.e., more than two variables) generally have as their starting point two major coordinates. The vertical axis, or ordinate (also called the $Y$ axis), is used to present an independent variable. Each of these axes is scaled in the units of

**TABLE 7 Forms of Statistical Graphics (by Function)**

| Data Summary | Two Variables | Three or More Variables |
|---|---|---|
| | *Exploration* | |
| Box and whisker plot | Autocorrelation plot | Biplot |
| Histogram[a] | Cross-correlation plot | Cluster trees[a] |
| Dot array diagram | Scatterplot[a] | Labeled scatterplot[a] |
| Frequency polygon | Sequence plot | Glyphs and metroglyphs |
| Ogive | | Face plots |
| Stem-and-leaf diagram | | Fourier plots |
| | | Similarity and preference maps |
| | | Multidimensional scaling displays |
| | | Weathervane plot |
| **Distribution Assessment** | **Model Evaluation and Assumption Verification** | **Decision Making** |
| | *Analysis* | |
| Probability plot | Average versus standard deviation | Control chart |
| *Q–Q* plot | Component-plus-residual plot | Custom chart |
| *P–P* plot | Partial-residual plot | Half-normal plot |
| Hanging histogram | Residual plots | Ridge trace |
| Rootagram | | Youden plot |
| Poissonness plot | | |
| **Quantitative Graphics** | **Summary of Statistical Analyses** | **Graphical Aids** |
| | *Communication and Display of Data* | |
| Line chart | Means plot | Confidence limits |
| Pictogram | Sliding reference distribution | Graph paper |
| Pie chart | Notched box plot | Power curves |
| Contour plot | Factor space/response | Nomographs |
| Stereogram | Interaction plot | Sample size curves |
| Color map | Contour plot | Trilinear coordinates |
| Histogram | Predicted response plot | |
| | Confidence region plot | |

[a]Reviewed in the text of this chapter.

measure which will most clearly present the trends of interest in the data. The range covered by the scale of each axis is selected to cover the entire region for which data are presented. The actual demarking of the measurement scale along an axis should allow for easy and accurate assessment of the coordinates of any data point yet should not be cluttered.

Actual data points should be presented by symbols which present the appropriate indicators of location, and if they represent summaries of data from a normal data population, it would be appropriate to present a symbol for the mean and some indication of the variability (or error) associated with

that population, commonly by using "error bars" which present the standard deviation (or standard error) from the mean. If, however, the data are not normal or continuous, it would be more appropriate to indicate location by the median and present the range or semiquartile distance for variability estimates. The symbols which are used to present data points can also be used to present a significant amount of additional information. At the simplest level a set of clearly distinct symbols (circles, triangles, squares, etc.) is very commonly used to provide a third dimension of data (most commonly treatment group). But by clever use of symbols, all sorts of additional information can be presented. Using a method such as Chernoff's (1973) faces, in which faces are used as symbols of the data points (and various aspects of the faces present additional data, such as the presence or absence of eyes denoting presence or absence of a secondary pathological condition), it is possible to present a large number of different variables on a single graph.

The three other forms of graphs that are commonly used are histograms, pie charts, and contour plots.

Histograms are graphs of simple frequency distribution. Commonly, the abscissa is the variable of interest (such as life span or litter size) and is generally shown as classes or intervals or measurements (such as age ranges of 0–10 weeks, 10–20 weeks, etc.). The ordinate, meanwhile, is the incidence or frequency of observations. The result is a set of vertical bars, each of which represents the incidence of a particular set of observations. Measures of error or variability about each incidence are reflected by some form of error bar on top of or in the frequency bars, as shown in Figure 8. The size of class intervals may be unequal (in effect, one can combine or pool several small class intervals), but it is proper in such cases to vary the width of the bars to indicate differences in interval size.

Pie charts are the only common form of quantitative graphic technique which is not rectangular. Rather, the figure is presented as a circle out of which several "slices" are delimited. The only major use of the pie chart is in presenting a breakdown of the components of a group. Typically the entire set of data under consideration (such as total body weight) constitutes the pie while each slice represents a percentage of the whole (such as the percentages represented by each of several organs). The total number of slices in a pie should be small for the presentation to be effective. Variability or error can be readily presented by having a subslice of each sector shaded and labeled accordingly.

Finally, there is the contour plot, which is used to depict the relationships in a three-variable, continuous data system. That is, a contour plot visually portrays each contour as a locus of the values of two variables associated with a constant value of the third variable. An example would be a relief map that gives both latitude and longitude of constant altitude using contour lines.

The most common misuse of graphs is to either conceal or exaggerate the extent of the difference by using inappropriately scaled or ranged axis. Tufte (1983) has termed a statistic for evaluating the appropriateness of scale size, the *lie factor*, calculated as the ratio of the shown effect size to the range of potential change or effect. An acceptable range for the lie factor is from 0.95

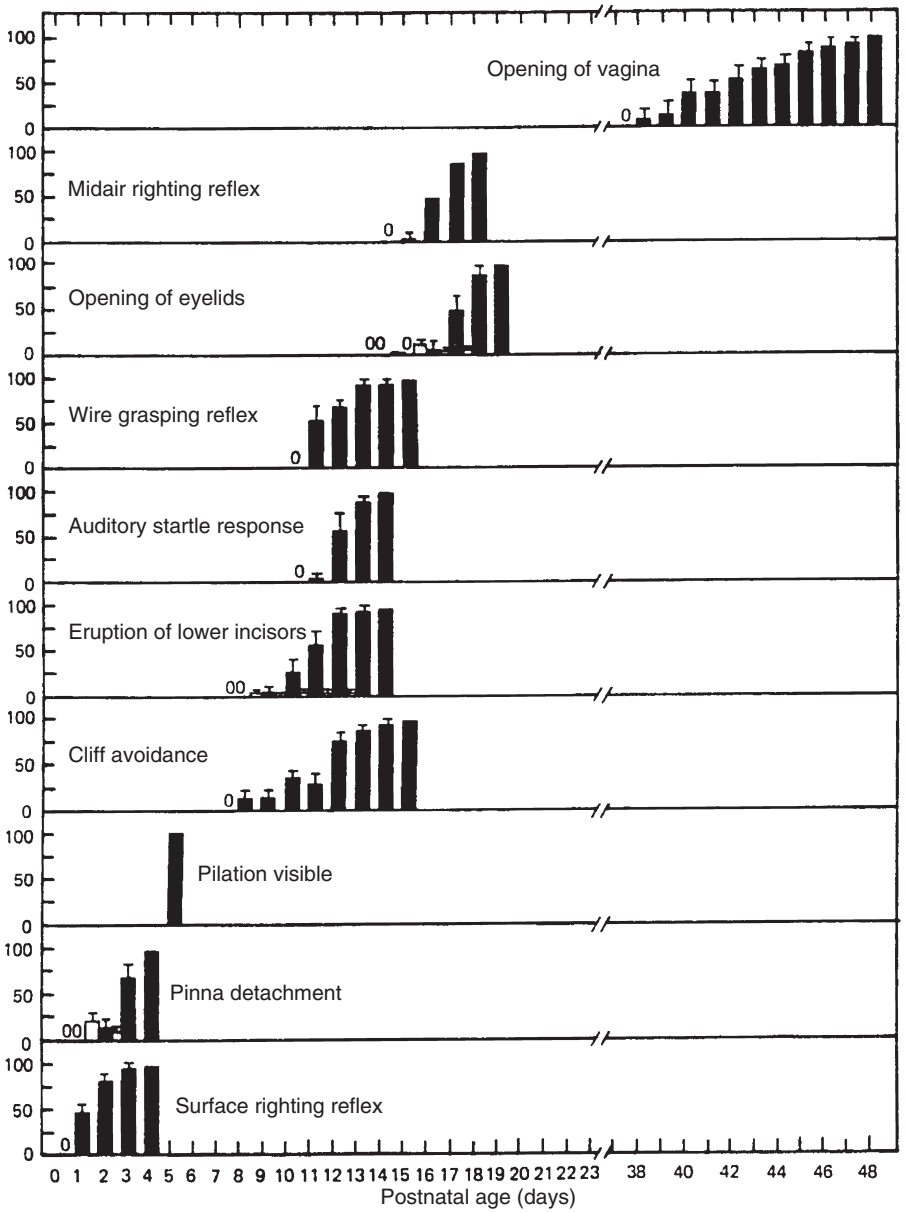


***Figure 8*** *Acquisitions of Post-Natal Development Landmarks in Rats.*

to 1.05. Less means the size of an effect is being understated, more that the effect is being exaggerated.

There are a number of excellent references available for those who would like to pursue statistical graphics more. Anscombe (1973) presents an excellent short overview, while others (Tufte, 1983, 1990, 1997; Schmid, 1983; Young, 1985) provide a wealth of information.

## 8.3 Multidimensional and Nonmetric Scaling

Multidimensional scaling (MDS) is a collection of analysis methods for data sets which have three or more variables making up each data point. MDS displays the relationships of three- or more-dimensional extension of the methods of statistical graphics.

MDS presents the structure of a set of objects from data that approximate the distances between pairs of the objects. The data, called similarities, dissimilarities, distances, or proximities, must be in such a form that the degree of similarities and differences between the pairs of the objects (each of which represents a real-life data point) can be measured and handled as a distance (remember the discussion of measures of distances under classifications). Similarity is a matter of degree—small differences between objects cause them to be similar (a high degree of similarity) while large differences cause them to be considered dissimilar (a small degree of similarity).

In addition to the traditional human conceptual or subjective judgments or similarity, data can be an "objective" similarity measure (the difference in weight between a pair of animals) or an index calculated from multivariate data (the proportion of agreement in the results of a number of carcinogenicity studies). However, the data must always represent the degree of similarity of pairs of objects.

Each object or data point is represented by a point in a multidimensional space. These plots or projected points are arranged in this space so that the distances between pairs of points have the strongest possible relation to the degree of similarity among the pairs of objects. That is, two similar objects are represented by two points that are close together, and two dissimilar objects are represented by a pair of points that are far apart. The space is usually a two- or three-dimensional Euclidean space but may be non-Euclidean and may have more dimensions.

MDS is a general term which includes a number of different types of techniques. However, all seek to allow geometric analysis of multivariate data. The forms of MDS can be classified (Young, 1985) according to the nature of the similarities in the data. It can be qualitative (nonmetric) or quantitative (metric MDS). The types can also be classified by the number of variables involved and by the nature of the model used, for example, classical MDS (there is only one data matrix and no weighting factors are used on the data), replicated MDS (more than one matrix and no weighting), and weighted MDS (more than one matrix and at least some of the data being weighted).

MDS can be used in toxicology to analyze the similarities and differences between effects produced by different agents in an attempt to use an understanding of the mechanism underlying the actions of one agent to determine the mechanisms of the other agents. Actual algorithms and a good intermediate-level presentation of MDS can be found in Davison (1983).

Nonmetric scaling is a set of graphic techniques closely related to MDS and is definitely useful for the reduction of dimensionality. Its major objective is to arrange a set of objects (each object, for our purposes, consisting of a number of related observations) graphically in a few dimensions while retaining the maximum possible fidelity to the original relationships between members (i.e., values which are most different are portrayed as most distant). It is not a linear technique, it does not preserve linear relationships (i.e., $A$ is not shown as twice as far from $C$ as $B$, even though its "value difference" may be twice as much). The spacings (interpoint distances) are kept such that if the distance of the original scale between members $A$ and $B$ is greater than that between $C$ and $D$, the distances on the model scale shall likewise be greater between $A$ and $B$ than between $C$ and $D$. Figure 5, presented earlier, uses a form of this technique in adding a third dimension by using letters to present degrees of effect on the skin.

This technique functions by taking observed measures of similarity or dissimilarity between every pair of $M$ objects, then finding a representation of the objects as points in Euclidean space that the interpoint distances in some sense "match" the observed similarities or dissimilarities by means of weighting constants.

## 8.4 Cluster Analysis

Cluster analysis is a quantitative form of classification. It serves to help develop decision rules and then use these rules to assign a heterogeneous collection of objects to a series of sets. This is almost entirely an applied methodology (as opposed to theoretical). The final result of cluster analysis is one of several forms of graphic displays and a methodology (set of decision classifying rules) for the assignment of new members into the classifications.

The classification procedures used are based on either density of population or distance between members. These methods can serve to generate a basis for the classification of large numbers of dissimilar variables such as behavioral observations and compounds with distinct but related structures and mechanisms (Gad, 1984; Gad et al., 1985) or to separate tumor patterns caused by treatment from those caused by old age (Hammond et al., 1978).

There are five types of clustering techniques (Everitt, 1980; Romesburg, 1984):

(a) *Hierarchical Techniques* Classes are subclassified into groups, with the process being repeated at several levels to produce a tree which gives sufficient definition to groups.

(b) *Optimizing Techniques* Clusters are formed by optimization of a clustering criterion. The resulting classes are mutually exclusive; the objects are partitioned clearly into sets.
(c) *Density or Mode-Seeking Techniques* Clusters are identified and formed by locating regions in a graphic representation which contains concentrations of data points.
(d) *Clumping Techniques* A variation of density-seeking techniques in which assignment to a cluster is weighted on some variables so that clusters may overlap in graphic projections.
(e) *Others* Methods which do not clearly fall into classes (a)–(d).

Romesburg (1984) provides an excellent step-by-step guide to cluster analysis.

## 8.5 Fourier (Time) Analysis

Fourier analysis (Bloomfield, 1976) is most frequently a univariate method used for either simplifying data (which is the basis for its inclusion in this chapter) or modeling. It can, however, also be a multivariate technique for data analysis.

In a sense, it is like trend analysis; it looks at the relationship of sets of data from a different perspective. In the case of Fourier analysis, the approach is by resolving the time dimension variable in the data set. At the most simple level, it assumes that many events are periodic in nature, and if we can remove the variation in other variables because of this periodicity (by using Fourier transforms), we can better analyze the remaining variation from other variables. The complications to this are (a) there may be several overlying cyclic time-based periodicities and (b) we may be interested in the time cycle events for their own sake.

Fourier analysis allows one to identify, quantitate, and (if we wish) remove the time-based cycles in data (with their amplitudes, phases, and frequencies) by use of the Fourier transform:

$$nJ_i = x_i \exp(-iw_i t)$$

where $n$ = length
$J$ = discrete Fourier transform for that case
$x$ = actual data
$i$ = increment in series
$w$ = frequency
$t$ = time

## 8.6 Life Tables

Chronic in vivo toxicity studies are generally the most complex and expensive studies conducted by a toxicologist. Answers to a number of questions are

sought in such a study—notably if a material results in a significant increase in mortality or in the incidence of tumors in those animals exposed to it. But we are also interested in the time course of these adverse effects (or risks). The classic approach to assessing these age-specific hazard rates is by the use of life tables (also called survivorship tables).

It may readily be seen that during any selected period of time ($t_i$) we have a number of risks competing to affect an animal. There are risks of (a) "natural death," (b) death induced by a direct or indirect action of the test compound, and (c) death due to such occurrences of interest of tumors (Hammond et al., 1978; Salsburg, 1980). And we are indeed interested in determining if (and when) the last two of these risks become significantly different than the "natural" risks (defined as what is seen to happen in the control group). Life table methods enable us to make such determinations as the duration of survival (or time until tumors develop) and the probability of survival (or of developing a tumor) during any period of time.

We start by deciding the interval length ($t_i$) we wish to examine within the study. The information we gain becomes more exact as the interval is shortened. But as interval length is decreased, the number of intervals increases and calculations become more cumbersome and less indicative of time-related trends because random fluctuations become more apparent. For a two-year or lifetime rodent study, an interval length or a month is commonly employed. Some life table methods, such as the Kaplan–Meyer, have each new event (such as a death) define the start of a new interval.

Having established the interval length we can tabulate our data (Cutler and Ederer, 1958). We start by establishing the following columns in each table (a separate table being established for each group of animals—i.e., by sex and dose level):

(a) Interval of time selected ($t_i$)
(b) Number of animals in group that entered that interval of study alive ($l_i$)
(c) Number of animals withdrawn from study during interval (such as those taken for interim sacrifice or that may have been killed by technician error) ($\omega_i$)
(d) Number of animals that died during interval ($d_i$)
(e) Number of animals at risk during interval, $l_i = l_i - \frac{1}{2}\omega_I$, or number on study at start of interval minus one-half number withdrawn during interval
(f) Proportion of animals that died, $=D_i = d_i/l_i$
(g) Cumulative probability of animal surviving until end of that interval of study, $P_i = 1 - D_i$, or 1 minus number of animals that died during interval divided by number of animals at risk
(h) Number of animals dying until that interval ($M_i$)
(i) Animals found to have died during interval ($m_i$)

(j) Probability of dying during interval of study $c_i = 1 - (M_i + m_i/l_i)$, or total number of animals dead until that interval plus animals discovered to have died during that interval divided by number of animals at risk through end of that interval
(k) Cumulative proportion surviving, $p_i$, equivalent to cumulative product of interval probabilities of survival (i.e., $p_i = p_1 \cdot p_2 \cdot p_3 \ldots p_x$)
(l) Cumulative probability of dying, $C_i$, equal to cumulative product of interval probabilities to that point (i.e., $C_i = c_1 \cdot c_2 \cdot c_3 \ldots c_i$).

With such tables established for each group in a study, we may now proceed to test the hypothesis that each of the treated groups has a significantly shorter duration of survival or that each of the treated groups died more quickly (note that plots of total animals dead and total animals surviving will give one an appreciation of the data but can lead to no statistical conclusions).

There are a multiplicity of methods for testing significance in life tables, with (as is often the case) the power of the tests increasing as does the difficulty of computation (Salsburg, 1980; Cox, 1972; Haseman, 1977; Tarone, 1975).

We begin our method of statistical comparison of survival at any point in the study by determining the standard error of the $K$-interval survival rate as (Garrett, 1947)

$$S_K = P_k \sqrt{\sum_1^k \left( \frac{D_i}{l'_x - d_x} \right)}$$

We may also determine the effective sample size ($l_1$) in accordance with

$$l_1 = \frac{P(1-P)}{S^2}$$

We may now compute the standard error of difference for any two groups (1 and 2) as

$$S_D = \sqrt{S_1^2 + S_2^2}$$

The difference in survival probabilities for the two groups is then calculated as

$$P_D = P_1 - P_2$$

We can then calculate a test statistic as

$$t' = \frac{P_D}{S_D}$$

This is then compared to the $z$-distribution table. If $t' > z$ at the desired probability level, it is significant at that level. With increasing recognition of the effects of time (both as age and length of exposure to unmeasured background risks), life table analysis has become a mainstay in chronic toxicology. An example is the reassessment of the $ED_{01}$ study (SOT, 1981) which radically changed interpretation of the results and understanding of underlying methods when adjustment for time on study was made.

The increased importance and interest in the analysis of survival data has not been restricted to toxicology but rather has encompassed all the life sciences. Those with further interest should consult Lee (1980) or Elandt-Johnson and Johnson (1980), both general in their approach to the subject.

## 9 META-ANALYSIS

Meta-analysis (meaning "analysis among") is being used increasingly in biomedical research to try to obtain a qualitative or quantitative synthesis of the research literature on a particular issue. The technique is usually applied to the synthesis of several separate but comparable studies.

### 9.1 Selection of Studies to Be Analyzed

The issue of study selection is perhaps the most troublesome issue for those doing meta-analysis. Several questions need to be addressed.

1. Should studies be limited to those which are published? It is well known that negative studies that report little or no benefit from following a particular course of action are less likely to be published than are positive studies. Therefore, the published literature may be biased toward studies with positive results, and a synthesis of these studies would give a biased estimate of the impact of pursuing some courses of action. Unpublished studies, however, may be of lower quality than the published studies, and poor research methods often produce an underestimate of impact. Moreover, the unpublished studies may be difficult to discover.

2. Should studies be limited to those which appear in peer-reviewed publications? Peer review is considered the primary method for quality control in scientific publishing. Some investigators recommend that only those studies which are published in peer-reviewed publications be considered in meta-analysis. Although this may seem an attractive option, it might produce an even more highly biased selection of studies.

3. Should studies be limited to those which meet additional quality control criteria? If investigators impose an additional set of criteria before including a study in meta-analysis, this may further improve the average quality of the studies used, but it introduces still greater concerns about selection bias. More-

over, different investigators might use different criteria for a "valid" study and therefore select a different group of studies for meta-analysis.

4. Should studies be limited to randomized controlled studies? This is a variant of the above question concerning quality control. At one time, rigid quality standards were more likely to be met by randomized controlled studies than by observational studies. Increasingly, however, observational methods have been used to evaluate certain kinds of effects, particularly those that are uncommon. A larger issue may well be that of combining data from studies performed in different laboratories and, even more so, using different strains of a single animal species.

5. Should studies be limited to those using identical methods? For practical purposes, this would mean using only separately published studies from the same lab in a limited time frame for which the methods were the same for all and the similarity of methods was monitored. This criterion is very difficult to achieve.

### 9.2 Pooled (Quantitative) Analysis

Usually, the main purpose of meta-analysis is quantitative. The goal is to develop better overall estimates of the degree of benefit achieved by specific exposure and dosing techniques based on the combining (pooling) of estimates found in the existing studies of the interventions. This type of meta-analysis is sometimes called a pooled analysis (Gerbarg and Horwitz, 1988) because the analysts pool the observations of many studies and then calculate parameters such as risk ratios or odds ratios from the pooled data.

Because of the many decisions regarding inclusion or exclusion of studies, different meta-analyses might reach very different conclusions on the same topic. Even after the studies are chosen, there are many other methodologic issues in choosing how to combine means and variances (e.g., what weighting methods should be used). Pooled analysis should report both relative risks and risk reductions as well as absolute risks and risk reductions (Sinclair and Bracken, 1994).

### 9.3 Methodologic (Qualitative) Analysis

Sometimes the question to be answered is not how much toxicity is induced by the use of a particular exposure but whether there is any biologically significant toxicity at all. In this case, a qualitative meta-analysis may be done in which the quality of the research is scored according to a list of objective criteria. The meta-analyst then examines the methodologically superior studies to determine whether or not the question of toxicity is answered consistently by them. This qualitative approach has been called methodologic analysis (Gerbarg and Horwitz, 1988) or quality scores analysis (Greenland, 1994). In some cases, the methodologically strongest studies agree with one another and

disagree with the weaker studies, which may or may not be consistent with one another.

## 10 BAYESIAN INFERENCE

It is useful to know the sensitivity and specificity of a test. Once a researcher decides to use a certain test, two important questions require answers: If the test results are positive, what is the probability that the researcher has the condition of interest? If the test results are negative, what is the probability that the patient does not have the disease? Bayes's theorem provides a way to answer these questions.

Bayes's theorem, which was first described centuries ago by the English clergyman after whom it is named, is one of the most imposing statistical formulas in biomedical sciences (Lindley, 1971). Put in symbols more meaningful for researchers such as pathologists, the formula is as follows:

$$P(D|T^{+}) = \frac{p(T^{+}|D^{+})\,p(D^{+})}{[p(T^{+}|D^{+})\,p(D^{+})] + [p(T^{+}|D^{-})\,p(D^{-})]}$$

where $p$ denotes probability, $D^{+}$ means that the animal has the effect in question, $D^{-}$ means that the animal does not have the effect, $T^{+}$ means that a certain diagnostic test for the effect is positive, $T^{-}$ means that the test is negative, and the vertical line (|) means "conditional upon" what immediately follows.

Most researchers, even those who can deal with sensitivity, specificity, and predictive values, throw in the towel when it comes to Bayes's theorem. This is odd because a close look at the above equation reveals that Bayes's theorem is merely the formula for the positive predictive value (Box and Tiao, 1973).

The numerator of Bayes's theorem merely describes cell $a$ (the true-positive results). The probability of being in cell $a$ is equal to the prevalence times the sensitivity, where $p(D^{+})$ is the prevalence (the probability of being in the effected column) and where $p(T^{+}|D^{+})$ is the sensitivity (the probability of being in the top row *given the fact of being in the effected column*). The denominator of Bayes's theorem consists of two terms, the first of which once again describes cell $a$ (the true-positive results) and the second of which describes cell $b$ (the false-positive error rate, or $p(T^{+}|D^{-})$, and is multiplied by the prevalence of nonaffected animals, or $p(D^{-})$. The true-positive results ($a$) divided by the true-positive plus false-positive results ($a + b$) gives $a/(a + b)$, which is the positive predictive value.

In genetics, an even simpler appearing formula for Bayes's theorem is sometimes used. The numerator is the same but the denominator is $p(T^{+})$. This makes sense because the denominator in $a/(a + b)$ is equal to all those that have positive test results, whether they are true-positive or false-positive results.

## 10.1 Bayes's Theorem and Evaluation of Safety Assessment Studies

In a population with a low prevalence of a particular toxicity, most of the positive results in a screening program for that lesion or effect would be falsely positive. Although this does not automatically invalidate a study or assessment program, it raises some concerns about cost-effectiveness, and these can be explored using Bayes's theorem (Racine et al., 1986).

A program employing an immunochemical-stain-based test to screen tissues for a specific effect will be discussed as an example. This test uses small amounts of antibody tissues for a specific effect, and the presence of an immunologically bound stain is considered a positive result. If the sensitivity and specificity of the test and the prevalence of biochemical effect are known, Bayes's theorem can be used to predict what proportion of the tissues with positive test results will have true-positive results (actually be affected).

Example 3 shows how the calculations are made. If the test has a sensitivity of 96% and if the true prevalence is 1%, only 13.9% of tissues with a positive test result are predicted to actually be affected.

Pathologist and toxicologists can quickly develop a table that lists different levels of test sensitivity, test specificity, and effect prevalence and shows how these levels affect the proportion of positive results that are likely to be true-positive results. Although this calculation is fairly straightforward and is extremely useful, it has seldom been used in the early stages of planning for large studies or safety assessment programs.

## 10.2 Bayes's Theorem and Individual Animal Evaluation

Suppose a pathologist is uncertain about an animal's cause of death and obtains a positive test result for a certain pathology. Even if the pathologist knows the sensitivity and specificity of the test, that does not solve the problem, because to calculate the positive predictive value, it is necessary to know the prevalence of the particular tissue/effect that the test is designed to detect. The prevalence is thought of as the expected prevalence in the population from which the animal comes. The actual prevalence is usually not known, but often a reasonable estimate can be made.

Say, for example, a pathologist evaluates a male primate that was observed to have easy fatigability and has signs of kidney stones but has no other symptoms or signs of parathyroid disease on physical examination. The pathologist considers the probability of hyperparathyroidism and decides that it is now perhaps 2% (reflecting that in 100 such primates probably only 2 of them would have the disease). This probability is called the prior probability, reflecting the fact that it is estimated prior to the performance of laboratory tests and is based on the estimated prevalence of a particular pathology among primates with similar signs and symptoms. Although the pathologist believes that the probability of hyperparathyroidism is low, he or she considers the

**Example 3** Use of Bayes's Theorem or 2 × 2 Table to Determine Positive Predictive Value of Hypothetical Tuberculin Screening Program

***Part 1.*** Beginning Data

| | |
|---|---|
| Sensitivity of immunological stain | 96% = 0.96 |
| False-negative error rate of test | 4% = 0.04 |
| Specificity of test | 94% = 0.94 |
| False-positive error rate of test | 6% = 0.06] |
| Prevalence of effect in tissues | 1% = 0.01 |

***Part 2.*** Use of Bayes's Theorem

$$p(D^+|T^+ = \frac{p(T^+|D^+p(D^+)}{[p(T^+|D^+)p(D^+)]+[PT^+|D^-)p(D^-)]}$$
$$= \frac{(\text{sensitivity})(\text{prevalence})}{(\text{sensitivity})(\text{prevalence})+(\text{false}-\text{postive error rate})(1-\text{prevalence})}$$
$$= \frac{(0.96)(0.01)}{(0.96)(0.01)+(0.06)(0.99)} = \frac{0.0096}{0.0096+0.0594} = \frac{0.0096}{0.0690}$$
$$= 0.139 - \mathbf{13.9\%}$$

***Part 3.*** Use of 2 × 2 Table with Numbers Based on Assumption That 10,000 Tissues are in Study

| | True Disease Status | | | | | |
|---|---|---|---|---|---|---|
| | Diseased | | Nondeseased | | Total | |
| | Number | Percentage | Number | Percentage | Number | Percentage |
| Test result | | | | | | |
| Positive | 96 | 96 | 594 | 6 | 690 | 7 |
| Negative | 4 | 4 | 9,306 | 94 | 9,310 | 93 |
| Total | 100 | 100 | 9,900 | 100 | 10,000 | 100 |

Positive predictive value = 96/690 = 0.139 = **13.9%**.

results of the serum calcium test to "rule" out the diagnosis. Somewhat to the pathologist's surprise, the results of the test were positive, with an elevated level of 12.2 mg dL$^{-1}$. He or she could order more special tests or stains for parathyroid disease, but some test results might come back positive and some negative.

Under the circumstances, Bayes's theorem could be used to make a second estimate of probability, which is called the posterior probability, reflecting the fact that it is made after the test results are known. Calculation of the posterior probability is based on the sensitivity and specificity of the test that was performed, which in this case was the serum calcium test, and on the prior prob-

ability, which in this case was 2%. If the serum calcium test had a 90% sensitivity and a 95% specificity, that means it had a false-positve error rate of 5% (specificity plus the false-positive error rate equals 100%). When this information is used in the Bayes equation, as shown in Example 4, the result is a posterior probability of 27%. This means that the patient is now in a group of primates with a significant possibility of parathyroid disease. In Example 4, note that the result is the same (i.e., 27%) when a 2 × 2 table is used. This is true because, as discussed above, the probability based on the Bayes theorem is identical to the positive predictive value.

In light of the 27% posterior probability, the pathologist decides to order a parathyroid hormone radioimmunoassay, even though this test is expensive. If the radioimmunoassay had a sensitivity of 95% and a specificity of 98% and the results turned out to be positive, the Bayes theorem could again be used to calculate the probability of parathyroid disease. This time, however, the

**Example 4** Use of Bayes's Theorem or 2 × 2 Table to Determine Posterior Probability and Positive Predictive Value

***Part 1.*** Beginning Data

| | |
|---|---|
| Sensitivity of first test | 90% = 0.90 |
| Specificity of first test | 95% = 0.95 |
| Prior probability of disease | 2% = 0.02 |

***Part 2.*** Use of Bayes's Theorem

$$\boldsymbol{p}(D^+|T^+) = \frac{p(T^+|D^+)\,p(D^+)}{[p(T^+|D^+)\,p(D^+)] + [p(T^+|D^-)\,p(D^-]}$$

$$= \frac{(0.90)(0.02)}{(0.90)(0.02) + (0.05)(0.98)}$$

$$= \frac{0.018}{0.018 + 0.049} = \frac{0.018}{0.067} = 0.269 = 27\%$$

***Part 3.*** Use of 2 × 2 Table

| | True Disease Status | | | | | |
|---|---|---|---|---|---|---|
| | Diseased | | Nondiseased | | Total | |
| | Number | Percentage | Number | Percentage | Number | Percentage |
| Test result | | | | | | |
| Positive | 18 | 90 | 49 | 5 | 67 | 6.7 |
| Negative | 2 | 10 | 931 | 95 | 933 | 93.3 |
| Total | 20 | 100 | 980 | 100 | 1000 | 100.0 |

Positive predictive value = 18/67 = 0.269 = 27%.

posterior probability for the first test (27%) would be used as the prior probability for the second test. The result of the calculation, as shown in Example 5, is a new probability of 94%. Thus, the primate in all probability did have hyperparathyroidism.

Why did the posterior probability increase so much the second time? One reason was that the prior probability was considerably higher in the second calculation than in the first (27% vs. 2%), based on the fact that the first test yielded positive results. Another reason was that the specificity of the second test was quite high (98%), which markedly reduced the false-positive error rate and therefore increased the positive predictive value.

**Example 5** Use of Bayes's Theorem or 2 × 2 Table to Determine Second Posterior Probability and Second Positive Predictive Value

***Part 1.*** Beginning Data

| | |
|---|---|
| Sensitivity of first test | 95% = 0.95 |
| Specificity of first test | 98% = 0.98 |
| Prior probability of disease | 27% = 0.27 |

***Part 2.*** Use of Bayes's Theorem

$$\boldsymbol{p}(D^+|T^+) = \frac{p(T^+|D^+)\,p(D^+)}{[p(T^+|D^+)\,p(D^+)] + [p(T^+|D^-)\,p(D^-)]}$$

$$= \frac{(0.95)(0.27)}{(0.95)(0.27) + (0.02)(0.73)}$$

$$= \frac{0.257}{0.257 + 0.0146} = \frac{0.257}{0.272} = 0.9449^* = 94\%$$

***Part 3.*** Use of 2 × 2 Table

| | True Disease Status | | | | | |
|---|---|---|---|---|---|---|
| | Diseased | | Nondiseased | | Total | |
| | Number | Percentage | Number | Percentage | Number | Percentage |
| Test result | | | | | | |
| Positive | 256 | 95 | 15 | 2 | 271 | 27.1 |
| Negative | 13 | 5 | 716 | 98 | 729 | 72.9 |
| Total | 269 | 100 | 731 | 100 | 1000 | 100.0 |

Positive predictive value = 256/271 = 0.9446* = **94%**.

*The slight difference in the results for the two approaches is due to rounding errors. It is not important biologically.

## 11 DATA ANALYSIS APPLICATIONS IN SAFETY ASSESSMENT STUDIES

Having reviewed basic principles and provided a set of methods for statistical handling of data, the remainder of this chapter will address the practical aspects and difficulties encountered in working safety assessment.

There are now common practices in the analysis of safety data, though they are not necessarily the best. These are discussed in the remainder of this chapter, which seeks to review statistical methods on a use-by-use basis and to provide a foundation for the selection of alternatives in specific situations. Some of the newer available methodologies (meta-analysis and Bayesian approachs) should be kept in mind, however.

### 11.1 Body and Organ Weights

Among the sets of data commonly collected in studies where animals are dosed with (or exposed to) a chemical are body weight and the weights of selected organs. In fact, body weight is frequently the most sensitive indication of an adverse effect. How to best analyze this and in what form to analyze the organ weight data (as absolute weights, weight changes, or percentages of body weight) have been the subject of a number of articles (Jackson, 1962; Weil, 1962; Weil and Gad, 1980).

Both absolute body weights and rates of body weight change (calculated as changes from a baseline measurement value which is traditionally the animal's weight immediately prior to the first dosing with or exposure to test material) are almost universally best analyzed by ANOVA followed, if called for, by a post hoc test. Even if the groups were randomized properly at the beginning of a study (no group being significantly different in mean body weight from any other group, and all animals in all groups within two standard deviations of the overall mean body weight), there is an advantage to performing the computationally slightly more cumbersome (compared to absolute body weight) changes in body weight analysis. The advantage is an increase in sensitivity because the adjustment of starting points (the setting of initial weights as a "zero" value) acts to reduce the amount of initial variability. In this case, Bartlett's test is first performed to ensure homogeneity of variance and the appropriate sequence of analysis follows.

With smaller sample sizes, the normality of the data becomes increasingly uncertain and nonparametric methods such as Kruskal–Wallis may be more appropriate (Zar, 1974).

The analysis of relative (to body weight) organ weights is a valuable tool for identifying possible target organs (Lee, 1999; Bickis, 1990). How to perform this analysis is still a matter of some disagreement, however. Weil (1962) presented evidence that organ weight data expressed as percentages of body weight should be analyzed separately for each sex. Furthermore, since the conclusions from organ weight data of males differed so often from those of

females, data from animals of each sex should be used in this measurement. Others (Grubbs, 1969; Weil, 1973; Boyd and Knight, 1963; Boyd, 1972) have discussed in detail other factors which influence organ weights and must be taken into account.

The two competing approaches to analyzing relative organ weights call for either

1. calculating organ weights as a percentage of total body weight (at the time of necropsy) and analyzing the results by ANOVA or
2. analyzing results by ANCOVA, with body weights as the covariates, as discussed previously by the author (Weil and Gad, 1980).

A number of considerations should be kept in mind when these questions are addressed. First, one must keep a firm grasp on the difference between biological significance and statistical significance. In this particular case, we are especially interested in examining organ weights when an organ weight change is not proportional to changes in whole-body weights. Second, we are now required to detect smaller and smaller changes while still retaining a similar sensitivity (i.e., $p < 0.05$ level).

There are several devices to attain the desired increase in power. One is to use larger and larger sample sizes (number of animals) and the other is to utilize the most powerful test we can. However, the use of even currently employed numbers of animals is being vigorously questioned and the power of statistical tests must therefore now assume an increased importance in our considerations.

The biological rationale behind analyzing both absolute body weight and the organ weight–body weight ratio (this latter as opposed to a covariance analysis of organ weights) is that in the majority of cases, except for the brain, the organs of interest in the body change weight (except in extreme cases of obesity or starvation) in proportion to total body weight. We are particularly interested in detecting cases where this is not so. Analysis of actual data from several hundred studies (unpublished data) has shown no significant difference in rates of weight change of target organs (other than the brain) compared to total body weight for healthy animals in those species commonly used for repeated-dose studies (rats, mice, rabbits, and dogs). Furthermore, it should be noted that ANOVA is of questionable validity in analyzing body weight and related organ weight changes, since a primary assumption is the independence of treatment—that the relationship of the two variables is the same for all treatments (Ridgemen, 1975). Plainly, in toxicology this is not true.

In cases where the differences between the error mean squares are much greater, the $F$ ratios will diverge in precision from the result of the efficiency of covariance adjustment. These cases are where either sample sizes are much larger or the differences between means themselves are much larger. This latter case is one which does not occur in the designs under discussion in any manner that would leave ANOVA as a valid approach because group means

start out being very similar and cannot diverge markedly unless there is a treatment effect. As we have discussed earlier, a treatment effect invalidates a prime underpinning assumption of ANOVA.

## 11.2 Clinical Chemistry

A number of clinical chemistry parameters are commonly determined on the blood and urine collected from animals in chronic, subchronic, and occasionally acute toxicity studies. In the past (and still in some places), the accepted practice has been to evaluate these data using univariate–parametric methods (primarily *t* tests and/or ANOVA). However, this can be shown to not be the best approach on a number of grounds.

First, such biochemical parameters are rarely independent of each other. Neither is our interest often focused on just one of parameters. Rather, there are batteries of parameters associated with toxic actions at particular target organs. For example, increases in creatinine phosphokinase (CPK), γ-hydroxybutyrate dehydrogenase (γ-HBDH), and lactate dehydrogenase (LDH) occurring together are strongly indicative of myocardial damage. In such cases, we are interested in not just a significant increase in one of these but in all three. Detailed coverage of the interpretation of such clinical laboratory tests can be found in other references (Martin et al., 1975; Harris, 1978; Gad and Chengelis, 1992; Loeb and Quimby, 1999) or elsewhere in this text.

Similarly, the serum electrolytes (sodium, potassium, and calcium) interact with each other; a decrease in one is frequently tied, for instance, to an increase in one of the others. Furthermore, the nature of the data (in the case of some parameters), either because of the biological nature of the parameter or the way in which it is measured, frequently is either not normally distributed (particularly because of being markedly skewed) or not continuous in nature. This can be seen in some of the reference data for experimental animals in Mitruka and Rawnsley (1977) or Weil (1982) in, for example, creatinine, sodium, potassium, chloride, calcium, and blood urea nitrogen.

## 11.3 Hematology

Much of what we said about clinical chemistry parameters is also true for the hematological measurements made in toxicology studies. Which test to perform should be evaluated by use of a decision tree until one becomes confident as to the most appropriate methods. Keep in mind that sets of values and (in some cases) population distribution vary not only between species but also between the commonly used strains of species and that "control" or "standard" values will "drift" over the course of only a few years.

Again, the majority of these parameters are interrelated and highly dependent on the method used to determine them. Red blood cell (RBC) count, platelet counts, and mean corpuscular volume (MCV) may be determined using a device such as a Coulter counter to take direct measurements, and the

resulting data are usually stable for parametric methods. The hematocrit, however, may actually be a value calculated from the RBC count and MCV values and, if so, is dependent on them. If the hematocrit is measured directly, instead of being calculated from the RBC count and MCV, it may be compared by parametric methods.

Hemoglobin is directly measured and is an independent and continuous variable. However, and probably because at any one time a number of forms and conformations (oxyhemoglobin, deoxyhemoglobin, methemoglobin, etc.) of hemoglobin are actually present, the distribution seen is not typically a normal one but rather may be a multimodal one. Here a nonparametric technique such as the Wilcoxon or multiple rank sum is called for.

Consideration of the white blood cell (WBC) and differential counts leads to another problem. The WBC count is, typically, a normal population amenable to parametric analysis, but differential counts are normally determined by counting, manually, one or more sets of 100 cells each. The resulting relative percentages of neutrophils are then reported as either percentages or are multiplied by the total WBC count with the resulting "count" being reported as the "absolute" differential WBC. Such data, particularly in the case of eosinophils (where the distribution does not approach normality), should usually be analyzed by nonparametric methods. It is widely believed that "relative" (%) differential data should not be reported because they are likely to be misleading.

Lastly, it should always be kept in mind that it is rare for a change in any single hematological parameter to be meaningful. Rather, because these parameters are so interrelated, patterns of changes in parameters should be expected if a real effect is present, and analysis and interpretation of results should focus on such patterns of changes. Classification analysis techniques often provide the basis for a useful approach to such problems.

## 11.4 Histopathological Lesion Incidence

The last 20 years have seen increasing emphasis placed on histopathological examination of tissues collected from animals in subchronic and chronic toxicity studies. While it is not true that only those lesions which occur at a statistically significantly increased rate in treated/exposed animals are of concern (for there are cases where a lesion may be of such a rare type that the occurrence of only one or a few such in treated animals "raises a flag"), it is true that, in most cases, a statistical evaluation is the only way to determine if what we see in treated animals is significantly worse than what has been seen in control animals. And although cancer is not our only concern, this category of lesions is that of greatest interest.

Typically, comparison of incidences of any one type of lesion between controls and treated animals are made using the multiple $2 \times 2$ chi-square test or Fisher's exact test with a modification of the numbers of animals as the denominators. Too often, experimenters exclude from consideration all those animals

(in both groups) which died prior to the first animals being found with a lesion at that site. The special case of carcinogenicity bioassays will be discussed in detail in Section 11.5.

An option which should be kept in mind is that, frequently, a pathologist can not only identify a lesion as present but also grade those present as to severity. This represents a significant increase in the information content of the data which should not be given up by performing an analysis based only on the perceived quantal nature (present/absent) of the data. Quantal data, analyzed by chi-square or Fisher's exact tests, are a subset (the $2 \times 2$ case) of categorical or contingency table data. In this case it also becomes ranked (or "ordinal") data—the categories are naturally ordered (e.g., no effect < mild lesion < moderate lesion < severe lesion). This gives a $2 \times R$ table if there are only one treatment and one control group or an $N \times R$ ("multiway") table if there are three or more groups of animals.

The traditional method of analyzing multiple, cross-classified data has been to collapse the $N \times R$ contingency table over all but two of the variables and to follow this with the computation of some measure of association between these variables. For an $N$-dimensional table this results in $N(N-1)/2$ separate analyses. The result is crude, "giving away" information and even (by inappropriate pooling of data) yielding a faulty understanding of the meaning of data. Though computationally more laborious, a multiway ($N \times R$ table) analysis should be utilized.

## 11.5 Carcinogenesis

In the experimental evaluation of substances for carcinogenesis based on experimental results in a nonhuman species at some relatively high dose or exposure level, an attempt is made to predict the occurrence and level of tumorogenesis in humans at much lower levels. An entire chapter could be devoted to examining the assumptions involved in this undertaking and review of the aspects of design and interpretation of animal carcinogenicity studies, and to special cases such as skin painting carcinogenicity studies (Wilson, 1982). Such is beyond the scope of this effort. The reader is referred to Gad (1998) for such an examination.

The single most important statistical consideration in the design of carcinogenicity bioassays in the past was based on the point of view that what was being observed and evaluated was a simple quantal response (cancer occurred or it did not) and that a sufficient number of animals needed to be used to have reasonable expectations of detecting such an effect. Though the single fact of whether or not the simple incidence of neoplastic tumors is increased due to an agent of concern is of interest, a much more complex model must now be considered. The time to tumor, patterns of tumor incidence, effects on survival rate, and age at first tumor all must now be included in a model.

The rationale behind this assumption is that, though humans may be exposed at very low levels, detecting the resulting small increase (over back-

ground) in the incidence of tumors would require the use of an impractically large number of test animals per group. This point was illustrated by Table 1, where, for instance, while only 46 animals (per group) are needed to show a 10% increase over a zero background (i.e., a rarely occurring tumor type), 770,000 animals (per group) would be needed to detect a tenth of a percent increase above a 5% background. As we increase dose, however, the incidence of tumors (the response) will also increase until it reaches the point where a modest increase (say 10% over a reasonably small background level (say 1%) could be detected using an acceptably small sized group of test animals (in Table 8 we see that 51 animals would be needed for this example case). There are, however, at least two real limitations to the highest dose level. First, the test rodent population must have a sufficient survival rate after receiving a lifetime (or two years) of regular doses to allow for meaningful statistical analysis. Second, we really want the metabolism and mechanism of action of the chemical at the highest level tested to be the same as at the low levels where human exposure would occur. Unfortunately, toxicologists usually must select the high dose level based only on the information provided by a subchronic or range-finding study (usually 90 days in length), but selection of either too low or too high a dose will make the study invalid for detection of carcinogenicity and may seriously impair the use of the results for risk assessment.

There are several solutions to this problem. One of these has been the rather simplistic approach of the NTP Bioassay Program, which is to conduct a three-month range-finding study with sufficient dose levels to establish a level which significantly (10%) decreases the rate of body weight gain. This dose is defined as the maximum tolerated dose (MTD) and is selected as the highest dose. Two other levels, generally one-half MTD and one-quarter MTD, are selected for testing at the intermediate and low dose levels. In many earlier NCI studies, only one other level was used.

**TABLE 8 Average Number of Animals Needed to Detect Significant Increase in Incidence of Event (Tumors, Anomalies, etc.) Over Background Incidence (Control) at Several Expected Incidence Levels Using Fisher Exact Probability Test ($p$ = 0.05)**

| | Expected Increase in Incidence, % | | | | | |
|---|---|---|---|---|---|---|
| Background Incidence, % | 0.01 | 0.1 | 1 | 3 | 5 | 10 |
| 0 | 46,000,000[a] | 460,000 | 4,600 | 511 | 164 | 46 |
| 0.01 | 46,000,000 | 460,000 | 4,600 | 511 | 164 | 46 |
| 0.1 | 47,000,000 | 470,000 | 4,700 | 520 | 168 | 47 |
| 1 | 51,000,000 | 510,000 | 5,100 | 570 | 204 | 51 |
| 5 | 77,000,000 | 770,000 | 7,700 | 856 | 304 | 77 |
| 10 | 100,000,000 | 1,000,000 | 10,000 | 1,100 | 400 | 100 |
| 20 | 148,000,000 | 1,480,000 | 14,800 | 1,644 | 592 | 148 |
| 25 | 160,000,000 | 1,600,000 | 16,000 | 1,840 | 664 | 166 |

[a]Number of animals needed in each group—controls as well as treated.

The dose range-finding study is necessary in most cases, but the suppression of body weight gain is a scientifically questionable benchmark when dealing with establishment of safety factors. Physiological, pharmacological, or metabolic markers generally serve as better indicators of systemic response than body weight. A series of well-defined acute and subchronic studies designed to determine the "chronicity factor" and to study onset of pathology can be more predictive for dose setting than body weight suppression.

Also, the NTP's MTD may well be at a level where the metabolic mechanisms for handling a compound at real-life exposure levels have been saturated or overwhelmed, bringing into play entirely artifactual metabolic and physiological mechanisms (Gehring and Blau, 1977). The regulatory response to questioning the appropriateness of the MTD as a high dose level (Haseman, 1985) has been to acknowledge that occasionally an excessively high dose is selected but to counter by saying that using lower doses would seriously decrease the sensitivity of detection.

## REFERENCES

Abramowitz, M., and Stegun, I. A. (1964). *Handbook of Mathematical Functions*. National Bureau of Standards, Washington, DC, pp. 925–964.

Anderson, E. (1960). A semigraphical method for the analysis of complex problems. *Technomet* 2:387–391.

Anderson, S., Auquier, A., Hauck, W. W., Oakes, D., VAndaele, W., and Weisburg, H. I. (1980). *Statistical Methods for Comparative Studies*. Wiley, New York.

Anderson, T. W. (1971). *The Statistical Analysis of Time Series*. Wiley, New York.

Anscombe, F. J. (1973). Graphics in statistical analysis. *Am. Statist.* 27:17–21.

Antonello, J. M., Clark, R. L., and Heyse, J. F. (1993). Application of Tahey trend test procedures to assess developmental and reproductive toxicity. I. Measurement data. *Fund. Appl. Toxicol.* 21:52–58.

Armitage, P. (1955). Tests for linear trends in proportions and frequencies. *Biometrics* 11:375–386.

Beyer, W. H. (1976a). *Handbook of Tables for Probability and Statistics*. Chemical Rubber Co., Boca Raton, FL.

Beyer, W. H. (1976b). *Handbook of Tables for Probability and Statistics*. CRC Press, Boca Raton, FL.

Bickis, M. G. (1990). Experimental design. In (D. L. Arnold, H. C. Grice, and D. R. Krewski (Eds.), *Handbook of in Vivo Toxicity Testing*. Academic, San Diego, CA, pp. 113–166.

Bliss, C. I. (1935). The calculation of the dosage-mortality curve. *Ann. Appl. Biol.* 22:134–167.

Bloomfield, P. (1976). *Fourier Analysis of Time Series: An Introduction*. Wiley, New York.

Box, G. E. P., and Tiao, G. C. (1973). *Bayesian Inference in Statistical Analysis*. Addison-Wesley, Reading, MA.

Boyd, E. M. (1972). *Predictive Toxicometrics*. Williams & Wilkins, Baltimore, MD.

Boyd, E. M., and Knight, L. M. (1963). Postmortem shifts in the weight and water levels of body organs. *Tox. Appl. Pharm.* 5:119–128.

Breslow, N. (1984). Comparison of survival curves. In M. F. Buse, M. J. Staguet, and R. F. Sylvester (Eds.), *Cancer Clinical Trials: Methods and Practice*. Oxford University Press, Oxford, pp. 381–406.

Chambers, J. M., Cleveland, W. S., Kleiner, B., and Tukey, P. A. (1983). *Graphical Methods for Data Analysis*. Wadsworth, Belmont.

Chernoff, H. (1973). The use of faces to represent points in K-dimensional space graphically. *J. Am. Stat. Assoc.* 68:361–368.

Cleveland, W. S. (1985). *The Elements of Graphing Data*. Wadsworth Advanced Books, Monterey, CA.

Cleveland, W. S., and McGill, R. (1984). Graphical perception: Theory, experimentation, and application to the development of graphical methods. *J. Am. Stat. Assoc.* 79:531–554.

Cochran, W. F. (1954). Some models for strengthening the common $x^2$ tests. *Biometrics* 10:417–451.

Cochran, W. G., and Cox, G. M. (1975). *Experimental Designs*. Wiley, New York.

Conover, J. W., and Inman, R. L. (1981). Rank transformation as a bridge between parametric and nonparametric statistics. *Am. Stat.* 35:124–129.

Cox, D. R. (1972). Regression models and life-tables. *J. Roy. Stat. Soc.* 34B:187–220.

Cox, D. R., and Stuart, A. (1955). Some quick tests for trend in location and dispersion. *Biometrics* 42:80–95.

Crowley, J., and Breslow, N. (1984). Statistical analysis of survival data. *Annu. Rev. Public Health* 5:385–411.

Cutler, S. J., and Ederer, F. (1958). Maximum utilization of the life table method in analyzing survival. *J. Chron. Dis.* 8:699–712.

Davison, M. L. (1983). *Multidimensional Scaling*. Wiley, New York.

Diamond, W. J. (1981). *Practical Experimental Designs*. Lifetime Learning Publications, Belmont, CA.

Diem, K., and Lentner, C. (1975). *Documenta Geigy Scientific Tables*. Geigy, New York, pp. 158–159.

Draper, N. R., and Smith, H. (1981). *Applied Regression Analysis*. Wiley, New York.

Duncan, D. B. (1955). Multiple range and multiple *F* tests. *Biometrics* 11:1–42.

Dunnett, C. W. (1955). A multiple comparison procedure for comparing several treatments with a control. *J. Am. Stat. Assoc.* 50:1096–1121.

Dunnett, C. W. (1964). New tables for multiple comparison with a control. *Biometrics* 16:671–685.

Dykstra, R. L., and Robertson, T. (1983). On testing monotone tendencies. *J. Am. Stat. Assoc.* 78:342–350.

Elandt-Johnson, R. C., and Johnson, N. L. (1980). *Survival Models and Data Analysis*. Wiley, New York.

Engelman, L., Hartigan, J. A. (1969). Percentage points of a test for clusters. *J. Am. Stat. Assoc.* 64:1647–1648.

Everitt, B. (1980). *Cluster Analysis*. Halsted, New York.

Everitt, B. S., Hand, D. J. (1981). *Finite Mixture Distributions*. Chapman and Hall, New York.

*Federal Register* (1985). No. 50, Vol. 50.

Federer, W. T. (1955). *Experimental Design*. Macmillan, New York.

Feinstein, A. R. (1979). Scientific standards vs. statistical associations and biological logic in the analysis of causation. *Clin. Pharmacol. Ther.* 25:481–492.

Finney, D. J., Latscha, R., Bennet, B. M., and Hsu, P. (1963). *Tables for Testing Significance in a 2 × 2 Contingency Table*. Cambridge University Press, Cambridge.

Finney, D. K. (1977). *Probit Analysis*, 3rd ed. Cambridge University Press, Cambridge.

Gad, S. C. (1984). Statistical analysis of behavioral toxicology data and studies. *Arch. Toxicol. Suppl.* 5:256–266.

Gad, S. C. (1998). *Statistics and Experimental Design for Toxicologists*, 3rd ed. CRC Press, Boca Raton, FL.

Gad, S. C., and Chengelis, C. P. (1992). *Animal Models in Toxicology*. Marcel Dekker, New York.

Gad, S. C., Reilly, C., Siino, K. M., and Gavigan, F. A. (1985). Thirteen cationic ionophores: Neurobehavioral and membrane effects. *Drug Chem. Toxicol.* 8(6):451–468.

Gad, S. C., and Taulbee, S. M. (1996). *Handbook of Data Recording, Maintenance and Management for the Biomedical Sciences*. CRC Press, Boca Raton, FL.

Gallant, A. R. (1975). Nonlinear regression. *Am. Stat.* 29:73–81.

Garrett, H. E. (1947). *Statistics in Psychology and Education*. Longmans, Green, New York, pp. 215–218.

Gehring, P. J., and Blau, G. E. (1977). Mechanisms of carcinogenicity: Dose response. *J. Environ. Path. Toxicol.* 1:163–179.

Gerbarg, Z. B., and Horwitz, R. I. (1988). Resolving conflicting clinical trials: Guidelines for meta-analysis. *J. Clin. Epidemiol.* 41:503–509.

Ghent, A. W. (1972). A method for exact testing of 2 × 2, 2 × 3, 3 × 3 and other contingency tables, employing binomiate coefficients. *Am. Midland Naturalist* 88:15–27.

Glass, L. (1975). Classification of biological networks by their qualitative dynamics. *J. Theor. Biol.* 54:85–107.

Gold, H. J. (1977). *Mathematical Modeling of Biological System—An Introductory Guidebook*. Wiley, New York.

Gordon, A. D. (1981). *Classification*. Chapman and Hall, New York.

Greenland, S. (1994). Invited commentary: A critical look at some popular meta-analytic methods. *Am. J. Epidemiol.* 140:290–296.

Grubbs, F. E. (1969). Procedure for detecting outlying observations in samples. *Technometrics* 11:1–21.

Hammond, E. C., Garfinkel, L., and Lew, E. A. (1978). Longevity, selective mortality, and competitive risks in relation to chemical carcinogenesis. *Environ. Res.* 16: 153–173.

Harris, E. K. (1978). Review of statistical methods of analysis of series of biochemical test results. *Ann. Biol. Clin.* 36:194–197.

Harris, R. J. (1975). *A Primer of Multivariate Statistics*. Academic, New York, pp. 96–101.

Harter, A. L. (1960). Critical values for Duncan's new multiple range test. *Biometrics* 16:671–685.

Hartigan, J. A. (1983). Classification. In S. Katz and N. L. Johnson (Eds.), *Encyclopedia of Statistical Sciences*, Vol 2. Wiley, New York.

Haseman, J. K. (1977). Response to use of statistics when examining life time studies in rodents to detect carcinogenicity. *J. Toxicol. Environ. Health* 3:633–636.

Haseman, J. K. (1985). Issues in carcinogenicity testing: Dose selection. *Fund. Appl. Toxicol.* 5:66–78.

Hicks, C. R. (1982). *Fundamental Concepts in the Design of Experiments*. Holt, Rinehart, and Winston, New York.

Hoaglin, D. C., Mosteller, F., and Tukey, J. W. (1983). *Understanding Robust and Explanatory Data Analysis*. Wiley, New York.

Hollander, M., and Wolfe, D. A. (1973). *Nonparametric Statistical Methods*. Wiley, New York, pp. 124–129.

Jackson, B. (1962). Statistical analysis of body weight data. *Toxicol. Appl. Pharmacol.* 4:432–443.

Kotz, S., and Johnson, N. L. (1982). *Encyclopedia of Statistical Sciences*, Vol. 1. Wiley, New York, pp. 61–69.

Kowalski, B. R., and Bender, C. F. (1972). Pattern recognition, a powerful approach to interpreting chemical data. *J. Am. Chem. Soc.* 94:5632–5639.

Kraemer, H. C., and Thiemann, G. (1987). *How Many Subjects? Statistical Power Analysis in Research*. Sage Publications, Newbury Park, CA.

Lee, E. T. (1980). *Statistical Methods for Survival Data Analysis*. Lifetime Learning Publications, Belmont, MA.

Lee, P. N., and Lovell, D. (1999). Statistics for Toxicology. In B. Ballantyne, T. Marrs, and T. Syversen (Eds.), *General and Applied Toxicology*, 2nd eds. Grove's Dictionaries, New York, pp. 291–302.

Lindley, S. V. (1971). *Bayesian Statistics: A Review*. SIAM, Philadelphia.

Litchfield, J. T., and Wilcoxon, F. (1949). A simplified method of evaluating dose effect experiments. *J. Pharmacol. Exp. Ther.* 96:99–113.

Loeb, W. F., and Quimby, F. W. (1999). *The Clinical Chemistry of Laboratory Animals*, 2nd ed. Taylor & Francis, Philadelphia, PA.

Marriott, F. H. C. (1991). *The Dictionary of Statistical Terms*. Longman Scientific & Technical, Essex, England.

Martin, H. F., Gudzinowicz, B. J., and Fanger, H. (1975). *Normal Values in Clinical Chemistry*. Marcel Dekker, New York.

Mendell, N. R., Finch, S. J., and Thode, H. C. Jr. (1993). Where is the likelihood ratio test powerful for detecting two component normal mixtures? *Biometrics* 49:907–915.

Mitruka, B. M., and Rawnsley, H. M. (1977). *Clinical Biochemical and Hematological Reference Values in Normal Animals*. Masson, New York.

Montgomery, D. C., and Smith, E. A. (1983). *Introduction to Linear Regression Analysis*. Wiley, New York.

Myers, J. L. (1972). *Fundamentals of Experimental Designs*. Allyn and Bacon, Boston.

Peto, R., and Pike, M. C. (1973). Conservatism of approximation; $(0 - E)^2/E$ in the log rank test for survival data on tumour incidence data. *Biometrics* 29:579–584.

Peto, R., Pike, M. C., Armitage, P., Breslow, N. E., Cox. D. R., Howard, S. V., Kantel, N., McPherson, K., Peto, J., and Smith, P. G. (1977). Design and analysis of randomized clinical trials requiring prolonged observations of each patient, II Analyses and examples. *Br. J. Cancer* 35:1–39.

Peto, R., Pike, M., Day, N., Gray, R., Lee, P., Parish, S., Peto, J., Richards, S., and Wahrendorf, J. (1980). Guidelines for simple, sensitive significance tests for carcinogenic effects in long-term animal experiments. In *IARC Monographs on the Evaluation of the Carcinogenic Risk of Chemicals to Humans, Supplement 2, Long-Term and Short-Term Screening Assays for Carcinogens: A Critical Appraisal*. International Agency for Research in Cancer, Lyon, pp. 311–346.

Portier, C., and Hoel, D. (1984). Type I error of trend tests in proportions and the design of cancer screens. *Commun. Stat. Theory Meth.* A13:1–14.

Prentice, R. L. (1976). A generalization of the probit and logit methods for dose response curves. *Biometrics* 32:761–768.

Racine, A., Grieve, A. P., and Fluhler, H. (1986). Bayesian methods in practice: Experiences in the pharmaceutical industry. *Appl. Stat.* 35:93–150.

Ridgemen, W. J. (1975). *Experimentation in Biology*. Wiley, New York, pp. 214–215.

Roe, F. J. (1989). Relevance for man of the effects of lactose, polzols and other carbohydrates on calcium metabolism. *Human Toxicol.* 8(2):87–98.

Romesburg, H. C. (1984). *Cluster Analysis for Researchers*, Vol. 43. Lifetime Learning Publications, Belmont, MA, pp. 45–58.

Salsburg, D. (1980). The effects of life-time feeding studies on patterns of senile lesions in mice and rats. *Drug Chem. Tox.* 3:1–33.

Schaper, M., Thompson, R. D., and Alarie, Y. (1985). A method to classify airborne chemicals which alter the normal ventilatory response induced by $CO_2$. *Toxicol. Appl. Pharmacol.* 79:332–341.

Scheffé, H. (1959). *The Analysis of Variance*. Wiley, New York.

Schmid, C. F. (1983). *Statistical Graphics*. Wiley, New York.

Siegel, S. (1956). *Nonparametric Statistics for the Behavioral Sciences*. McGraw-Hill, New York.

Sinclair, J. C., and Bracken, M. B. (1994). Clinically useful measures of effect in binary analyses of randomized trials. *J. Clin. Epidemiol.* 47:881–889.

Snedecor, G. W., and Cochran, W. G. (1994). *Statistical Methods*, 7th ed. Iowa State University Press, Ames.

Sokal, R. R., and Rohlf, F. J. (1994). *Biometry*, 3rd ed. W. H. Freeman, San Francisco.

SOTSOT $ED_{01}$ Task Force (1981). Reexamination of the $ED_{01}$ study—Adjusting for time on study. *Fund. Appl. Toxicol.* 1:8–123.

Tarone, R. E. (1975). Tests for trend in life table analysis. *Biometrika* 62:679–682.

Tufte, E. R. (1983). *The Visual Display of Quantitative Information*. Graphics Press, Cheshire, CT.

Tufte, E. R. (1990). *Envisioning Information*. Graphics Press, Cheshire, CT.

Tufte, E. R. (1997). *Visual Explanations*. Graphics Press, Cheshire, CT.

Tukey, J. W. (1977). *Exploratory Data Analysis*. Addison-Wesley, Reading, MA.

Tukey, J. W., Ciminera, J. L., and Heyes, J. F. (1985). Testing the statistical certainty of a response to increasing doses of a drug. *Biometrics* 41:295–301.

Velleman, P. F., and Hoaglin, D. C. (1981). *Applications, Basics and Computing of Exploratory Data Analysis*. Duxbury, Boston.

Weil, C. S. (1962). Applications of methods of statistical analysis to efficient repeated-dose toxicological tests. I. General considerations and problems involved. Sex differences in rat liver and kidney weights. *Toxicol. Appl. Pharmacol.* 4:561–571.

Weil, C. S. (1973). Experimental design and interpretation of data from prolonged toxicity studies. In *Proc. 5th Int. Congr. Pharmacol.*, Vol. 2. Beacon Press, San Francisco, CA, pp. 4–12.

Weil, C. S. (1982). Statistical analysis and normality of selected hematologic and clinical chemistry measurements used in toxicologic studies. *Arch. Toxicol.* 5(Suppl.): 237–253.

Weil, C. S., and Gad, S. C. (1980). Applications of methods of statistical analysis to efficient repeated-dose toxicologic tests. 2. Methods for analysis of body, liver and kidney weight data. *Toxicol. Appl. Pharmacol.* 52:214–226.

Williams, D. A. (1971). A test for differences between treatment means when several dose levels are compared with a zero dose control. *Biomet.* 27:103–117.

Williams, D. A. (1972). The comparison of several dose levels with a zero dose control. *Biomet.* 28:519–531.

Wilson, J. S., and Holland, L. M. (1982). The effect of application frequency on epidermal carcinogenesis assays. *Toxicology* 24:45–53.

Young, F. W. (1985). Multidimensional scaling. In S. Katz and N. L. Johnson (Eds.), *Encyclopedia of Statistical Sciences*, Vol. 5. Wiley, New York, pp. 649–659.

Zar, J. H. (1974). *Biostatistical Analysis*. Prentice-Hall, Englewood Cliffs, NJ, p. 50.

# *Combination Products: Drugs and Devices*

## 1 INTRODUCTION

Recent years have seen a vast increase in the number of new therapeutic products which are not purely drug, device, or biological, but rather a combination of two or more of these. Classical examples are implanted drug delivery systems (whose primary function is drug delivery) and drug-impregnated devices (in which drug delivery is an adjunct to the device function). Congress first acknowledged the need for specific regulation of such combination products in the 1990 Safe Medical Device Act.

## 2 HISTORICAL BACKGROUND

The history of this category includes a variety of product types dating at least from the perfection of the hypodermic needle (1855). There are many modern examples of implanted delivery systems, such as the insulin pump (1980). One fundamental driving force for delivery systems has been the growth of new pharmaceutical products, especially since the dramatic expansion of drug research after 1945.

That research has led to the synthesis and testing of millions of compounds for pharmacological and antimicrobial properties. Indeed, today much of that development is performed in automated computer-controlled systems, leading

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

to an even greater acceleration of the process. The continued emergence of a stream of novel and more complex combination products has blurred any distinguishing lines of regulatory authority and has complicated product designation and regulation. The issue of products combing a device and a drug, such as an asthma inhaler, has received considerable scrutiny over the past several years. But products combining a device and a biological, such as organ replacement or assist devices, have received less attention. Recent trends, however, suggest that device and biological combination products are quickly moving into the spotlight. A 1998 survey conducted by the U.S. Food and Drug Administration (FDA) identified hardware and tissue-engineered combination products as a rapidly growing trend in medical device technology (Herman et al., 1998).

Even less than drug and device combinations, device and biological products—which include, among other things, cellular and tissue implants, infused or encapsulated cells, artificial and replacement organs, heart valves and pumps, and cardiac, neural, and neuromuscular stimulation devices—do not fit neatly into existing regulatory paradigms (see Table 3 for a longer list of such products). For example, as part of the question of regulation, the FDA must take into account the possibility of tissue contamination and other hazards involved in using animal-derived tissues.

What has resulted to date is still a developing regulatory process. The written guidelines are fixed, but the day-to-day process is in flux (Merrill, 1994; March, 1998; Segal, 1999; Gopalaswamy and Gopalaswamy, 2008). It starts with determination of a principal mode of action (PMOA) which governs which center of the FDA will have the primary regulatory oversight. More recently, the FDA [Center for Drug Evaluation and Research (CDER), 2005] has promulgated a guidance on the nonclinical safety evaluation of drug–drug combinations. Such combinations are now becoming more frequent, most commonly by the 505(b)(2) route for approval. If the combination is of two approved and marketed products, the primary safety concern is for international (pharmacokinetics and pharmacodynamics). As such, the primary nonclinical safety studies that must be conducted are systemic toxicity studies of the fixed combination itself with toxicokinetic sampling components. Genetic toxicity and safety pharmacology studies of the combination are usually not required.

## 3 FUTURE TRENDS

Table 1 presents anticipated developments in the device combination product category between now and 2020 which lead to new clinical products. Three types of developments are generally expected. First, additional products designed for implanted delivery of insulin and other drugs. These include implanted pumps, possibly intelligent devices with improved biosensors to monitor concentrations in body fluids and make dynamic adjustments in delivery rates. Also there is the likely development of new polymeric timed-release

**TABLE 1 Likelihood New Combination Device Drug Technologies**

| | |
|---|---|
| Biosensors | Biosensors (for glucose, implantation, and systemic infection markers), genetic diagnostics, laser diagnosis and treatment, minimally invasive devices |
| Blood vessel prosthetics | Genetic therapy, tissue-engineered vessels, nerves, and devices |
| Bone prosthetics/growth | Artificial organs, tissue-engineered devices |
| Cardiac stimulation | Intelligent devices, microminiaturized devices |
| Cartilage prosthetics | Tissue-engineered device |
| Computer-aided clinical labs | Computer-aided diagnosis, networks of devices |
| Drug-impregnated devices | Device/drug/biological products |
| Endoscopy | Minimally invasive devices, telemedicine, virtual reality diagnostics |
| Genetics—cancer | Genetic diagnostics, genetic therapy |
| Hearing aids | Intelligent devices, microminiaturized devices, nonimplanted sensory aids |
| Heart pumps | Artificial organs |
| Heart valves | Artificial organs, tissue-engineered devices, device/drug/biological products |
| Home diagnostics | Home/self-monitoring and diagnosis |
| Image contrast agents | Medical imaging |
| Imaging: functional, content | Medical imaging, minimally invasive devices, networks of devices |
| Implanted drug | Biosensors, device/drug/biological products, delivery systems, home/self-therapy, intelligent devices, robotic devices |
| Integrated patient medical info systems | Computer-aided diagnosis, networks of devices, telemedicine |
| Kidney prosthetics | Artificial organs, home/self-therapy, tissue-engineered devices |
| Laser surgery | Laser diagnosis and treatment |
| Liver prosthetics | Artificial organs, tissue-engineered devices |
| Minimum invasive cardiology | Minimally invasive devices, vascular surgery |
| Minimum invasive neurosurgery | Minimally invasive devices |
| Magnetic resonance imaging | Greater resolution imaging |
| Nanotechnology | Microminiaturized devices |
| Nerve regeneration | Tissue-engineered devices |
| Neural stimulation | Artificial organs, electrical stimulation, intelligent devices |
| Neuromuscular stimulation | Electrical stimulation, home/self-therapy |
| Ocular prosthetics | Artificial organs, electrical stimulation, intelligent devices |
| Pancreas prosthetics | Artificial organs, tissue-engineered devices |
| Patient smart cards | Computer-aided diagnosis, networks of devices, telemedicine |
| Positron emission tomography (PET) imaging | Combined PET and computer-aided tomography imaging |
| Robotic surgery | Microminiaturized devices, robotic devices |
| Skin prosthetics | Tissue-engineered devices |
| Telemedicine—home use | Home/self-monitoring and testing, diagnosis, telemedicine |
| Telemedicine—radiology | Telemedicine |

devices which could improve the delivery of long-acting pharmaceuticals at optimized locations and rates.

Second, new developments in drug-impregnated devices are expected. Examples included new types of cardiac implants with antithrombogenic and anti-infective drugs as well as orthopedic implants with bacteriostatic coatings.

Finally, under development are new developments in drug delivery systems to simplify reliable use by unsophisticated patients in home settings, including the increasing elderly population. Examples included nasal and inhalation products.

Device regulation designation is by PMOA, which is generally straightforward but can become less clear as precedents accumulate and technology becomes more complex.

Although both extracorporeal and peritoneal dialysis systems are regulated as devices, dialysate concentrate for use with the former is a device but prepackaged dialysate for use with the latter is a drug. Sometimes consistency was elusive even when there was no combination, but just a single product. For example, in vitro diagnostics for detecting antibodies to HIV are regulated as biologicals when they are used for screening the blood supply but as medical devices when used for diagnostic or other screening purposes. When the FDA decides quickly and unequivocally on the regulatory status of a product, whether it was deemed a single product or was in combination with another product, there was relatively little opportunity for objection to the agency's decisions about how to regulate combination products and products whose status was uncertain. In the case of blood devices, the European Union (EU) has affirmed this process (Anonymous, 2000).

In the Safe Medical Devices Act of 1990 (SMDA), Congress took these issues in hand and amended the Federal Food, Drug and Cosmetic Act (FDCA) to make it easier for the FDA to regulate combination products in a rational fashion. The new provisions altered the substantive provisions of the FDCA only in minor respects. The main thrust of the new law was managerial, directing the FDA to make decisions about which center would have "primary jurisdiction" over a combination product based on the agency's understanding of the primary mode of action of the product.

For these products, the center's jurisdiction turns on the PMOA. If the primary mode of action is that of a drug, then the CDER has primary jurisdiction; if it is that of a device, jurisdiction is with the Center for Devices and Radiological Health (CDRH); if that of a biological product, the Center for Biologics Evaluation and Research (CBER) has this jurisdiction. As the statute prescribed, the regulations go on to state that the center with primary jurisdiction may consult with other agency components.

Although neither the statute nor the regulations explain what "primary jurisdiction" means, it seems clear that the FDA intends it to mean that the center that has primary jurisdiction will review the combination product and ordinarily give it just one approval, that is, a new drug application (NDA),

PMA (Product Marketing Application) or biological license application (BLA) as appropriate. Section 3.4(b) makes it clear, however, that the FDA's designation of one agency component as having primary jurisdiction does not preclude, in appropriate cases, the requirement for separate application, for example, a 510(k) and a BLA. When separate applications are required, both can be reviewed by the lead center, but "exceptional" cases may involve a second application to be reviewed by a different center. To facilitate this, the agency published new delegations giving officials in each of the three centers the authority to clear devices and to approve devices, drugs, biologicals, or any combination of two or more of them (FDA, 1991).

Contemporaneous with publication of the new regulations, the FDA made public three new intercenter agreements between CDRH and CBER, CDRH and CDER, and CDER and CBER. They describe the allocations of responsibility for numerous categories of specific products, both combination and noncombination. According to the regulations, these intercenter agreements are not binding; they are intended to "provide useful guidance to the public" and, as a practical matter, to FDA staff as well.

The intercenter agreements are a treasure trove of information. In addition to explicit guidance about which center has the lead with respect to particular products and whether one center or two will work on particular issues, they contain information and hints about whether the FDA believes it can regulate certain products at all, and if so, how (Pilot and Waldeermann 1998; Adams et al., 1997).

The regulations and intercenter agreements, however, do not answer every question, and the regulations recognize a role for the sponsor in cases of uncertainty. When the identity of the center with primary jurisdiction is unclear or in dispute or a sponsor believes its combination product is not covered by the intercenter agreements, a sponsor can request a designation from the FDA's product jurisdiction officer. A sponsor "should" file a request for designation with the product jurisdiction officer before submitting its application for marketing approval or an investigational notice. In practice, though, disputes or lack of clarity may not become evident until well into the review process, and it seems likely that the FDA would, if necessary, entertain requests for designation submitted at a later time.

Section 3.7(c) of the regulations lists the information to be included in the request, all of which must fit on 15 pages or less, including the identity of the sponsor, detailed information on the product, where the developmental work stands, the product's known modes of action and its primary mode of action, and, importantly, the sponsor's recommendation for which center should have primary jurisdiction and the reasons for the recommendation.

The FDA promises to check the request for designation for completeness within 5 working days of receipt and to issue a letter of designation within 60 days of receipt of a complete request. If the FDA does not meet the 60-day time limit, then the sponsor's recommendation for the appropriate lead center is honored.

The agency's letter of designation can be changed only with the sponsor's written consent or, if the sponsor does not consent, "to protect the public health or for other compelling reasons" (Gopalaswamy and Gopalaswamy, 2008). A sponsor must be given prior notice of any proposed nonconsensual change and must be given an opportunity to object in writing and at a "timely" meeting with the product jurisdiction officer and appropriate center officials.

The CDRH is designated the center for major policy development and for the promulgation and interpretation of procedural regulations for medical devices under the act. The CDRH regulates all medical devices inclusive of radiation-related device that are not assigned categorically or specifically to CDER. In addition, CDRH will independently administer the following activities (references to sections are the provisions of the act):

1. A. Small business assistance programs under Section 10 of the amendments [see Public Law (PL) 94–295]. Both CDER and CDRH will identify any unique problems relating to medical device regulation for small business.
   B. Registration and listing under Section 510 including some CDER-administered device applications. The CDRH will receive printouts and other assistance as requested.
   C. Color additives under Section 706, with review by CDER, as appropriate.
   D. Good Manufacturing Practices (GMPs) Advisory Committee. Under Section 520(f)(3), CDER will regularly receive notices of all meetings, with participation by CDER, as appropriate.
   E. Medical device reporting. The manufacturers, distributors, importers, and users of all devices, including those regulated by CDER, shall report to CDRH under Section 519 of the act as required. The CDRH will provide monthly reports and special reports as needed to CDER for investigation and follow-up of those medical devices regulated by CDER.

Table 2 presents the primary product responsibilities of CDER and CBER.

## 4 DEVICE PROGRAMS THAT CDER AND CDRH EACH WILL ADMINISTER

Both CDER and CDRH will administer and, as appropriate, enforce the following activities for medical devices assigned to their respective centers (references to sections are the provisions of the act):

1. A. Surveillance and compliance actions involving general controls violations, such as misbranded or adulterated devices under Sections 301, 501, and 502
   B. Warning letters, seizures, injunctions, and prosecutions under Sections 302, 303, and 304

**TABLE 2 Product Class Review Responsibilities**

| |
|---|
| *Center for Drug Evaluation and Review* |
| Natural products purified from plant or mineral sources |
| Products produced from solid tissue sources (excluding procoagulants, venoms, blood products, etc.) |
| Antibiotics, regardless of method of manufacture |
| Certain substances produced by fermentation |
| Disaccharidase inhibitors |
| HMG–CoA inhibitors |
| Synthetic chemicals |
| Traditional chemical synthesis |
| Synthesized mononuclear or polynuclear products including antisense chemicals |
| Hormone products |
| *Center for Biologics Evaluation and Review* |
| Vaccines, regardless of manufacturing method |
| In vivo diagnostic allergenic products |
| Human blood products |
| Protein, peptide, and/or carbohydrate products produced by cell culture (other than antibiotics and hormones) |
| Immunoglobulin products |
| Products containing intact cells or microorganisms |
| Proteins secreted into fluids by transgenic animals |
| Animal venoms |
| Synthetic allergens |
| Blood banking and infusion adjuncts |

C. Civil penalties under Section 303(f) and administrative restraint under Section 304(g)
D. Nonregulatory activities, such as educational programs directed at users, participation in voluntary standards organizations, and so on
E. Promulgation of performance standards and applications of special controls under Section 514
F. Premarket notification, investigational device exemptions including humanitarian exemptions, premarket approval, product development protocols, classification, device tracking, petitions for reclassification, postmarket surveillance under Sections 510(k), 513, 515, 519, 520(g) and (m), and 522, and the advisory committees necessary to support these activities
G. Banned devices under Section 516
H. FDA-requested and firm-initiated recalls whether under Section 518 or another authority and other Section 518 remedies such as recall orders
I. Exemptions, variances, and applications of current GMP (CGMP) regulations under Section 520(f)
J. Governmentwide quality assurance program
K. Requests for export approval under Sections 801(e) and 802

**TABLE 3 Examples of Existing Device–Drug Combination Products**

| | | |
|---|---|---|
| Cardiac output catheter | Heparin | As device in U.K. |
| Extracorporeal sets | Heparin | As device in U.K. |
| Viscose/rayon dressings | Povidone iodine | As drug in U.K. |
| Cardiovascular oxygenator | Heparin | As drug in U.K. defoamer reservoir |
| Paste bandages | Clioquinol, coal tar, calamine | As drug in U.K. (if they have antiinfective ancillary action) |
| Medicated tulle dressings | Chlorhexidine | As drug in U.K. |
| Antimicrobial drape | Iodophore | As device in U.K. |
| Antiseptic wipes | Chlorhexidine, centrimide, alcohol | As drug in U.K. |
| Cardiovascular guidewires | Heparin | As device in Spain |
| Guidewires | Heparin | As device in Spain, Switzerland, U.K. |
| Antibiotic bone cement | Antibiotic (e.g., gentamicin sulfate), colistin sulfomethate, sodium, erythromycin | As drug (but soon to be regulated as device) |
| Extracorporeal cardiotomy reservoirs and filters | Heparin | As devices in Spain, Benelux, Italy |
| Extracorporeal venous reservoirs and filters | Heparin | As devices in Spain, Benelux, Italy |
| Bacteriostatic urological catheters | Silver | As devices in three Benelux countries |
| Antiseptic island dressing | Chlorhexidine digluconate | As device in Italy |
| Spermicidal condoms | Nonoxynol-9 | As device in Germany |
| Pacemaker lead with porous tip (seulte) | Dexamethasone | As device |
| Pacemaker lead with protector mannitol capsule (Sweet Tip) | Mannitol | As device |
| Biomedicus centrifugal pump | Heparin | Not applicable |
| Peripheral vascular cannulae | Heparin | Not applicable |
| Surgical gauzes or nonwoven fabrics impregnated with iodophore | Iodophore | Not applicable |
| Surgical gauzes or nonwoven fabrics impregnated with alignates and Clioquinol | Clioquionol (NaCa alignates, clauden powder) | As device in Germany |
| Vascular prosthesis | Collagen, albumen | As devices in U.K. |

## 5 COORDINATION

The centers will coordinate their activities in order to assure that manufacturers do not have to independently secure authorization to market their product from both centers unless this requirement is specified in Section VII.

## 6 SUBMISSIONS

Submissions should be made to the appropriate center, as specified herein, at the following addresses: Food and Drug Administration, Center for Drug Evaluation and Research (CDER), Central Document Room (CDR), 5901-B Ammendale Road, Beltsville, Maryland 20705-1266 or Food and Drug Administration, Center for Devices and Radiological Health, Document Mail Center (HFZ-401), 9200 Corporate Blvd., Rockville, MD 20850.

For submissions involving medical devices and/or drugs that are not clearly addressed in this agreement, sponsors are referred to the product jurisdiction regulations (21 CFR Part 3). These regulations have been promulgated to facilitate the determination of regulatory jurisdiction but do not exclude the possibility for a collaborative review between the centers.

## 7 CENTER JURISDICTION

The following subsections provide details concerning status, market approval authority, special label/regulatory considerations, investigational options, and intercenter consultations for the categories of products specified. Section VII provides the general criteria that CDRH and CDER will apply in reaching decisions as to which center will regulate a product.

A. 1. (a) Device with primary purpose of delivering or aiding in the delivery of a drug that is distributed without a drug (i.e., unfilled)

*Examples*

Devices that calculate drug dosages

Drug delivery pump and/or catheter infusion pump for implantation iontophoresis device

Medical or surgical kit (e.g., tray) with reference in instructions for use with specific drug (e.g., local anesthetic)

Nebulizer

Small-particle aerosol generator (SPAG) for administering drug to ventilated patient

Splitter block for mixing nitrous oxide and oxygen

Syringe, jet injector, storage and dispensing equipment

*Status* Device and drug as separate entities

*Market Approval Authority* CDRH and CDER, respectively, unless the intended use of the two products, through labeling, creates a combination product

*Special Label/Regulatory Considerations* The following specific procedures will apply depending on the status of the drug delivery device and drugs that will be delivered with the device:

(i) It may be determined during the design or conduct of clinical trials for a new drug that it is not possible to develop adequate performance specifications data on those characteristics of the device that are required for the safe and effective use of the drug. If this is the case, then drug labeling cannot be written to contain information that makes it possible for the user to substitute a generic, marketed device for the device used during development to use with the marketed drug. In these situations, CDER will be the lead center for regulation of the device under the device authorities.

(ii) For a device intended for use with a category of drugs that are on the market, CDRH will be the lead center for regulation for the device under the device authorities. The effects of the device use on drug stability must be addressed in the device submission, when relevant. An additional showing of clinical effectiveness of the drug when delivered by the specific device will generally not be required. The device and drug labeling must be mutually conforming with respect to indication, general mode of delivery (e.g., topical, IV), and drug dosage/schedule equivalents.

(iii) For a drug delivery device and drug that are developed for marketing to be used together as a system, a lead center will be designated to be the contact point with the manufacturer(s). If a drug has been developed and marketed and the development and study of device technology predominate, the principal mode of action will be deemed to be that of the device, and CDRH would have the lead. If a device has been developed and marketed and the development and study of drug predominate, then, correspondingly, CDER would have the lead. If neither the drug nor the device is on the market, the lead center will be determined on a case-by-case basis.

*Investigation Options* IDE (Investigational Device Exemption) or IND as appropriate

*Intercenter Consultation* CDER, when the lead center, will consult with CDRH if CDER determines that a specific device is required as part of the NDA process. CDRH as lead center will consult with CDER if the device is intended for use with a marketed drug and the device creates a significant change in the intended use, mode of delivery (e.g., topical, IV), or dose/schedule of the drug.

1. Device with primary purpose of delivering or aiding in the delivery of a drug and distributed containing a drug (i.e., "pre-filled delivery system")

*Examples*
Nebulizer
Oxygen tank for therapy and over-the-counter (OTC) emergency use
Prefilled syringe

Transdermal patch

*Status* Combination product

*Market Approval Authority* CDER using drug authorities and device authorities as necessary

*Special Label/Regulatory Considerations* None

*Investigation Options* IND

*Intercenter Consultations* Optional

2. Device incorporating a drug component with the combination product having the primary intended purpose of fulfilling a device function

   *Examples*
   Bone cement containing antimicrobial agent
   Cardiac pacemaker lead with steroid-coated tip
   Condom, diaphragm, or cervical cap with contraceptive or antimicrobial agent (including virucidal) agent
   Dental device with fluoride
   Dental wood wedge with hemostatic agent
   Percutaneous cuff (e.g., for a catheter or orthopedic pin) coated/impregnated with antimicrobial agent
   Skin closure or bandage with antimicrobial agent
   Surgical or barrier drape with antimicrobial agent
   Tissue graft with antimicrobial or other drug agent
   Urinary and vascular catheter coated/impregnated with antimicrobial agent
   Wound dressing with antimicrobial agent

*Status* Combination product

*Market Approval Authority* CDRH using device authorities

*Special Label/Regulatory Considerations* These products have a drug component that is present to augment the safety and/or efficacy of the device.

*Investigation Options* IDE

*Intercenter Consultation* Required if a drug or the chemical form of the drug has not been legally marketed in the United States as a human drug for the intended effect

3. Drug incorporating a device component with the combination product having the primary intended purpose of fulfilling a drug function

   *Examples*
   Skin-prep pads with antimicrobial agent
   Surgical scrub brush with antimicrobial agent

*Status* Combination product

*Market Approval Authority* CDER using drug authorities and, as necessary, device authorities

*Special Label/Regulatory Considerations* Marketing of such a device requires a submission of an NDA with safety and efficacy data on the drug component or that it meet monograph specifications as generally

recognized as safe (GRAS) and generally recognized as effective (GRAE). Drug requirements (e.g., CGMPs, registration and listing, experience reporting) apply to products.

*Investigation Options* IND

*Intercenter Consultation* Optional

4. (a) Device used in the production of a drug either to deliver directly to a patient or for use in the producing medical facility (excluding use in a registered drug manufacturing facility)

*Examples*
Oxygen concentrators (home or hospital)
Oxygen generator (chemical)
Ozone generator

*Status* Device

*Market Approval Authority* CDER, applying both drug and device authorities

*Special Label/Regulatory Consideration* May also require an NDA if the drug produced is a new drug. Device requirements (e.g., CGMPs, registration and listing, experience reporting) will apply to products.

*Investigation Options* IDA or NDA as appropriate

*Intercenter Consultation* Optional

(b) Drug/device combination product intended to process a drug into a finished package form

*Examples*
Device that uses drug concentrates to prepare large-volume parenterals
Oxygen concentrator (hospital) output used to fill oxygen tanks for use within that medical facility

*Status* Combination product

*Market Approval Authority* CDER, applying both drug and device authorities

*Special Label/Regulatory Considerations* Respective drug and device requirements (e.g., CGMPs, registration and listing, experience reporting) will apply.

*Investigation Options* IDE or IND as appropriate

*Intercenter Consultation* Optional but will be routinely obtained

B. Device used concomitantly with a drug to directly activate or to augment drug effectiveness.

*Examples*
Biliary lithotriptor used in conjunction with dissolution agent
Cancer hyperthermia used in conjunction with chemotherapy

Current generator used in conjunction with an implanted silver electrode (drug) that produces silver ions for an antimicrobial purpose

Materials for blocking blood flow temporarily to restrict chemotherapy drug to the intended site of action

Ultraviolet and/or laser activation of oxsoralen for psoriasis or cutaneous T-cell lymphoma

*Status* Device and drug as separate entities

*Market Approval Authority* CDRH and CDER, respectively

*Special Label/Regulatory Considerations* The device and drug labeling must be mutually conforming with respect to indications, general mode of delivery (e.g., topical, IV), and drug dosage/schedule equivalence. A lead center will be designated to be the contact point with the manufacturer. If a drug has been developed and approved for another use and development and study of device technology predominate, then CDRH would have the lead. If a device has been developed and marketed for another use and development and study of drug action predominate, then CDER would have the lead. If neither the drug nor the device is on the market, the lead center will be determined on a case-by-case basis. If the labeling of the drug and device create a combination product, as defined in the combination product regulations, then the designation of the lead center for both applications will be based upon a determination of the product's primary mode of action.

*Investigation Options* IDE or IND as appropriate.

*Intercenter Consultations* Required

2. Device kits labeled for use with drugs that include both device(s) and drug(s) as separate entities in one package with the overall primary intended purpose of the kit fulfilling a device function

   *Example*

   Medical or surgical kit (e.g., tray) with drug component

*Status* Combination product

*Market Approval Authority* CDRH, using device authorities, is responsible for the kit if the manufacturer is repackaging a market drug. Responsibility for overall packaging resides with CDRH. CDER will be consulted as necessary on the use of drug authorities for the repackaged drug component.

*Special Label/Regulatory Consideration* Device requirements (e.g., CGMPs, registration and listing, experience reporting) apply to kits. Device manufacturers must assure that manufacturing steps do not adversely affect drug components of the kit. If the manufacturing steps do affect the marketed drug (e.g., the kit is sterilized by irradiation), an ANDA or NDA would also be required with CDRH as the lead center.

*Investigation Options* IDA or IND as appropriate

*Intercenter Consultation* Optional if ANDA or NDA not required

C. Liquids, gases, or solids intended for use as devices (e.g., implanted, or components, parts, or accessories to devices)

*Examples*
Dye for tissues used in conjunction with laser surgery to enhance absorption of laser light in target tissue
Gas mixtures for pulmonary function-testing devices
Gases used to provide "physical effects"
Hemodialysis fluids
Hemostatic devices and dressings
Injectable silicon, collagen, and Teflon
Liquids functioning through physical action applied to the body to cool or freeze tissues for therapeutic purposes
Liquids intended to inflate, flush, or moisten (lubricate) indwelling device (in or on the body)
Lubricants and lubricating jellies
Ophthalmic solutions for contact lenses
Organ/tissue transport and/or perfusion fluid with antimicrobial or other drug agent, that is, preservation solutions
Powders for lubricating surgical gloves
Sodium hyaluronate or hyaluronic acid for use as a surgical aid
Solution for use with dental "chemical drill"
Spray-on dressings not containing a drug component

*Status* Device

*Market Approval Authority* CDRH

*Special Label/Regulatory Considerations* None

*Investigation Options* IDE

*Intercenter Consultation* Required if the device has direct contact with the body and the drug or the chemical form of the drug has not been legally marketed as a human drug

D. Products regulated as drugs

*Examples*
Irrigation solutions
Purified water or saline in prefilled nebulizers for use in inhalation therapy
Skin protectants (intended for use on intact skin)
Sun screens
Topical/internal analgesic-antipyretic

*Status* Drug

*Market Approval Authority* CDER

*Special Label/Regulatory Considerations* None

*Investigation Options* IND

*Intercenter Consultations* Optional

E. Ad hoc jurisdictional decisions

*Examples*

| | Status | Center |
|---|---|---|
| Motility marker constructed of radiopaque plastic | Device | CDRH |
| Brachytherapy capsules, needles, etc., that are radioactive and may be removed from the body after radiation therapy has been administered | Device | CDRH |
| Skin markers | Device | CDRH |

*Status* Device or drug

*Market Approval Authority* CDRH or CDER as indicated

*Special Label/Regulatory Considerations* None

*Investigation Options* IDE or IND as appropriate

*Intercenter Consultation* Required to assure agreement on drug/device status

## 8 GENERAL CRITERIA AFFECTING DRUG/DEVICE DETERMINATION

The following represent the general criteria that will apply in making device/drug determinations:

A. *Device Criteria*

1. A liquid, powder, or other similar formulation intended only to serve as a component, part, or accessory to a device with a primary mode of action that is physical in nature will be regulated as a device by CDRH.
2. A product that has the physical attributes described in 201(h) (e.g., instrument, apparatus) of the act and does not achieve its primary intended purpose through chemical action within or on the body or by being metabolized will be regulated as a device by CDRH.
3. The phrase "within or on the body" as used in 201(h) of the act does not include extra corporeal systems or the solutions used in conjunction with such equipment. Such equipment and solutions will be regulated as devices by CDRH.
4. An implant, including an injectable material, placed in the body for primarily a structural purpose even though such an implant may be absorbed or metabolized by the body after it has achieved its primary purpose will be regulated as a device by CDRH.
5. A device containing a drug substance as a component with the primary purpose of the combination being to fulfill a device function is a combination product and will be regulated as a device by CDRH.

6. A device (e.g., machine or equipment) marketed to the user, pharmacy, or licensed practitioner that produces a drug will be regulated as a device or combination product by CDER. This does not include equipment marketed to a registered drug manufacturer.
7. A device whose labeling or promotional materials make reference to a specific drug or generic class of drugs unless it is prefilled with a drug ordinarily remains a device regulated by CDRH. It may, however, also be subject to the combination products regulation.

B. *Drug Criteria*

1. A liquid, powder, tablet, or other similar formulation that achieves its primary intended purpose through chemical action within or on the body or by being metabolized, unless it meets one of the specified device criteria, will be regulated as a drug by CDER.
2. A device that serves as a container for a drug or a device that is a drug delivery system attached to the drug container where the drug is present in the container is a combination product that will be regulated as a drug by CDER.
3. A device containing a drug substance as a component with the primary purpose of the combination product being to fulfill a drug purpose is a combination product and will be regulated as a drug by CDER.
4. A drug whose labeling or promotional materials make reference to a specific device or generic class of devices ordinarily remains a drug regulated by CDER. It may, however, also be subject to the combination products regulation.

## REFERENCES

Adams, D. G., Cooper, R. M., and Kahan, J. S. (1997). *Fundamentals of Law and Regulation: An In-depth Look at Therapeutic Products*. FDLI, Washington, DC.

Anonymous (2000). EU nations agree that directive will cover human blood devices. *Eur. Drug Device Report* 10(14):1–21.

Center for Drug Evaluation and Research (CDER) (1994). Premarket notifications: 510 (K) regulatory requirements for medical devices. *HHS FDA* 1990:4158.

Center for Drug Evaluation and Research (CDER) (2005). Nonclinical safety evaluation of drug or biologic combinations. FDA, Washington, DC.

Chapekar, M. S. (1996). Regulatory concerns in the development of biologic-biomaterials combination. *J. Biomed. Mater. Res.* 33:199–203.

Gopalaswamy, S., and Gopalaswamy, V. (2008). *Combination Products: Regulatory Challenges and Successful Product Development*. CRC Press, Boca Raton, FL.

Herman, W. A., Marlowe, D. F., and Harney, R. (1998). Future trends in medical device technology. *Biomedical Market Newsletter* 8:19–20.

March, E. (1998). Combination Products: Who's Regulating What and How—Part I. *Pharm. Eng.*, July/Aug., pp. 34–38.

Merrill, R. (1994). Regulation of drugs and devices: An evolution. *Health Affairs* 13:47–69.

Pilot, L., and Waldeermann, D. (1998). Food and Drug Modernization Act of 1997: Medical Device Provisions. *Food Drug Law J.* 53:267–295.

Segal, S. A. (1999). Device and biologic combination products—Understanding the evolving regulation. *Med. Dev. Diagn. Ind.*, Jan., pp. 180–184.

U.S. Food and Drug Administration (FDA) (1991). Assignment of agency component for review of premarket applications. Final Rule, 56 *Fed. Reg.* 58,754 (Nov. 21).

# *Qualification of Impurities, Degradants, Residual Solvents, and Leachables in Pharmaceuticals*

Impurities (from either materials intended to be part of the process of making a drug or formulating it), degradants (unintentionally formed by unintended reactions after the drug substance or product is produced and prone to increase in quantity over time due to instability of the product under conditions of storage), and residual solvents (purposely added to the synthesis product to facilitate synthesis, formulation, or dosage form production—always liquids of some degree of volatility) can become part of a drug product or substance in multiple ways. But the extent of their presence is now strictly governed by a series of International Conference on Harmonisation (ICH), U.S. Food and Drug Administration (FDA), and European Medicines Agency (EMEA) guidelines. These ICH guidelines call for these materials to be present at levels no greater than in product specifications and (because they serve no functional purpose in the drug) are both to be kept to a practical minimum and must be qualified for safety at the highest specification levels under the assumption of maximum potential patent use of the drug (and, therefore, exposure to the

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

unintended substances). Contaminants coming from the vessels and machinery used to manufacture a drug but not intended to be present and leachables/extractants are materials which transfer to a drug product from packaging or delivery systems. Undesired chemical substances which end up in pharmaceutical products may have a range of sources, but all of these must be evaluated for potential risks to patients. Depending on the source of the undesired substance, it is categorized and regulated somewhat differently. Categories of such substances include impurities, degradants, residual solvents, contaminants, leachables, and extractables. Each of these will be considered in this chapter.

Biologicals and biotechnology products have both similar and unique process impurity issues. The process needs to ensure that there are no residual cellular components in the biological product. Biotechnology processing needs to avoid causing structural deformities to the protein. In all cases, the process must be scrutinized closely. Checking for impurities at various steps throughout the manufacturing phase may help to pinpoint where the impurities are produced.

## 1 IMPURITIES

The ICH (2003a) "Guidance for Industry, Q3A: Impurities in New Drug Substances," is intended to "provide guidance for registration applications on the content and qualification of impurities in new drug substances produced by chemical syntheses and not previously registered in a region or member state" (p. 1). A new drug substance is not the final marketed product but the active ingredient used in the marketed product. Impurities in new drug substances are addressed from both a chemistry and safety perspective.

The guidance is not intended to apply to new drug substances during the clinical research stage of development (though such drugs in development must have consideration of meeting these requirements at the time of marketing approval) but rather addresses safety concerns associated with such substances during development. Nor does it cover natural product or biological process–produced drugs or extraneous contaminants that should not occur in new drug substances and are more appropriately addressed as good manufacturing practice (GMP) issues. The guidance further describes the circumstances in which impurities need to be reported, identified, and qualified.

The rationale for the reporting and control, identification, and qualification of impurities is discussed in the guidance. Organic impurities need to be summarized based on the actual and potential impurities most likely to arise during the synthesis, purification, and storage of a new drug substance. This discussion can be limited to those impurities that might reasonably be expected based on knowledge of the chemical reactions and conditions involved.

Studies conducted to characterize the structure of impurities present in a new drug substance at a level greater than the identification threshold (Table 1) should be described and any impurity from any batch or degradation

**TABLE 1 Thresholds for Action on Impurities in Drug Product**

| Maximum Daily Dose[a] | Reporting Threshold[b,c] | Identification Threshold[c] | Qualification Threshold[c] |
|---|---|---|---|
| ≤2 g day$^{-1}$ | 0.05% | 0.10% or 1.0 mg day$^{-1}$ intake (whichever is lower) | 0.15% or 1.0 mg day$^{-1}$ intake (whichever is lower) |
| >2 g day$^{-1}$ | 0.03% | 0.05% | 0.05% |

[a]The amount of drug substance administered per day.
[b]Higher reporting thresholds should be scientifically justified.
[c]Lower thresholds can be appropriate if the impurity is unusually toxic.

product from stability studies should be identified. If identification of an impurity or degradant is not feasible, a summary of the laboratory studies demonstrating the unsuccessful effort should be included in the application. If an impurity is pharmacologically or toxicologically active, identification of the compound should be conducted even if the impurity level is below the identification threshold.

The guidance also states:

> Qualification is the process of acquiring and evaluating data that establishes the biological safety of an individual impurity or a given impurity profile at the level(s) specified. The applicant should provide a rationale for establishing impurity acceptance criteria that includes safety considerations. The level of any impurity that is present in a new drug substance that has been adequately tested in safety and/or clinical studies would be considered qualified. Impurities that are also significant metabolites present in animal and/or human studies are generally considered qualified. A level of a qualified impurity higher than that present in a new drug substance can also be justified based on an analysis of the actual amount of impurity administered in previous relevant safety studies. If data are unavailable to qualify the proposed acceptance criterion of an impurity, safety studies to obtain such data can be appropriate when the usual qualification thresholds are exceeded.

ICH (2003b) Q3B(R) describes considerations for the qualification of impurities when thresholds are exceeded. If the level of impurity cannot be decreased to below the threshold or if adequate data is not available in the scientific literature to justify safety, then additional safety testing should be considered. The studies considered appropriate to qualify an impurity will depend on a number of factors, including the patient population, daily dose, and route and duration of administration. Toxicology studies are discussed briefly later in this chapter and in more detail in other chapters in this volume. Such studies can be conducted on the new drug substance containing the impurities to be controlled, although studies using isolated impurities can sometimes be appropriate.

ICH Q3A states that "safety assessment studies to qualify an impurity should compare the new drug substance containing a representative amount

of the new impurity with previously qualified material. Safety assessment studies using a sample of the isolated impurity can also be considered." The latter is especially important to consider for genetic toxicology studies and the importance of testing the isolated impurity is discussed in more detail at the end of this chapter.

Therefore, according to the guidance, if the maximum daily dose of the drug is less than 2 g $day^{-1}$ and the impurity intake is more than 0.15% or 1.0 mg $day^{-1}$, the qualification threshold has been reached, meaning safety studies will need to be performed. Lower thresholds can be appropriate if the impurity is unusually toxic. In addition, the impurity will need to be reported and identified. These studies include general and genetic toxicology studies, and possibly other specific toxicology endpoints, as appropriate. Discussion of specific toxicity testing with the relevant FDA division is recommended.

If considered desirable, a minimum screen (e.g., genotoxic potential) should be conducted. A study to detect point mutations and one to detect chromosomal aberrations, both in vitro, are considered an appropriate minimum screen.

Qualification studies for impurities are essentially bridging studies. If general toxicity studies are desirable, one or more studies should be designed to allow comparison of unqualified to qualified material. The study duration should be based on available relevant information and performed in the species most likely to maximize the potential to detect the toxicity of a degradation product. On a case-by-case basis, single-dose studies can be appropriate, especially for single-dose drugs. In general, a minimum duration of 14 days and a maximum duration of 90 days would be considered appropriate.

The genetic toxicology studies can include a minimum screen (a study to detect point mutations and one to detect chromosome aberrations, both in vitro). The general toxicology studies should include one or more studies designed to allow comparison of unqualified to qualified material. The study duration should be based on available relevant information and performed in the species most likely to maximize the potential to detect the toxicity of an impurity. On a case-by-case basis, single-dose studies can be appropriate, especially for single-dose drugs. In general, a minimum duration of 14 days and a maximum duration of 90 days would be considered appropriate.

Inorganic impurities are normally detected and quantified using pharmacopeial or other appropriate procedures. The need for inclusion or exclusion of inorganic impurities in a new drug substance specification should be discussed. Acceptance criteria should be based on pharmacopeia standards or known safety data. The control of residues of the solvents used in the manufacturing process for a new drug substance should be discussed and presented according to ICH (2003c) Q3C.

A registration application should include documented evidence that the analytical procedures are validated and suitable for the detection and quantification of impurities. Organic impurity levels can be measured by a variety of techniques, including those that compare an analytical response for an

**TABLE 2 Threshold for Degradation Products in New Drug Products**

| Maximum Daily Dose[a] | Threshold[b,c] |
|---|---|
| *Reporting Thresholds* | |
| ≤1 g | 0.1% |
| >1 g | 0.05% |
| *Identification Thresholds* | |
| <1 mg | 1.0% or 5 μg TDI, whichever is lower |
| 1–10 mg | 0.5% or 20 μg TDI, whichever is lower |
| >10 mg–2 g | 0.2% or 2 mg TDI, whichever is lower |
| >2 mg | 0.10% |
| *Qualification Thresholds* | |
| <10 mg | 1.0% or 50 μg TDI, whichever is lower |
| 10–100 mg | 0.5% or 200 μg TDI, whichever is lower |
| >100 mg–2 g | 0.2% or 3 mg TDI, whichever is lower |
| >2 g | 0.15% |

[a]The amount of drug substance administered per day.
[b]Thresholds for degradation products are expressed either as a percentage of the drug substance or as total daily intake (TDI) of the degradation product. Lower thresholds can be appropriate if the degradation product is unusually toxic.
[c]Higher thresholds should be scientifically justified.

impurity to that of an appropriate reference standard or to the response of the new drug substance itself. Differences in the analytical procedures used during development and those proposed for the commercial product should be discussed in the registration application. Analytical results should be provided in an application for all batches of a new drug substance used for clinical, safety, and stability testing as well as for batches representative of the proposed commercial process. The application should also contain a table that links the specific new drug substance batch to each safety study and each clinical study in which the new drug substance has been used. Any impurity at a level greater than the reporting threshold (Table 1) and total impurities observed in these batches of the new drug substance should be reported with the analytical procedures indicated. Table 2 is an illustration of reporting impurity results for identification and qualification in an application.

The guidance also states that when analytical procedures change results provided in the application should be linked to the procedure used, with appropriate validation information provided, including representative chromatograms of representative batches. The applicant should ensure that complete impurity profiles (e.g., chromatograms) of individual batches are available, if requested.

The ICH Q3A guidance also states that the specification for a new drug substance should include a list of impurities. Individual impurities with specific

acceptance criteria included in the specification for a new drug substance are referred to as specified impurities. Specified impurities can be identified or unidentified. A rationale for the inclusion or exclusion of impurities in a specification should be presented.

> According to ICH (2000): Acceptance criteria should be set no higher than the level that can be justified by safety data and should be consistent with the level achievable by the manufacturing process and the analytical capability. Where there is no safety concern, impurity acceptance criteria should be based on data generated on batches of a new drug substance manufactured by the proposed commercial process, allowing sufficient latitude to deal with normal manufacturing and analytical variation and the stability characteristics of the new drug substance. Although normal manufacturing variations are expected, significant variation in batch-to-batch impurity levels can indicate that the manufacturing process of the new drug substance is not adequately controlled and validated.

ICH "Q3B(R): Impurities in New Drug Products" was published November 2003 and is intended to provide guidance for registration applications on the content and qualification of impurities in new drug products produced from chemically synthesized new drug substances not previously registered in a region or member state. A new drug product is a finished dosage form, for example, a tablet, capsule, or solution, that contains a drug substance generally but not necessarily in association with one or more other ingredients. The Q3B(R) guidance complements the ICH guidance "Q3A: Impurities in New Drug Substances," which should be consulted for basic principles along with ICH "Q3C: Impurities: Residual Solvents" when appropriate.

Q3A addresses only those impurities in new drug products classified as degradation products of the drug substance or reaction products of the drug substance with an excipient and/or immediate container closure system (collectively referred to as *degradation products*). Generally, impurities present in a new drug substance need not be monitored or specified in a new drug product unless they are also degradation products. This guidance does not address impurities arising from excipients present in a new drug product or extracted or leached from the container closure system. This guidance also does not apply to new drug products used during the clinical research stages of development. It also does not cover the same types of products as in ICH Q3A(R): biological/biotechnological, peptides, oligonucleotides, radiopharmaceuticals, fermentation products and associated semisynthetic products, herbal products, and crude products of animal or plant origin. Also excluded from this guidance are extraneous contaminants that should not occur in new drug products and are more appropriately addressed as GMP issues, polymorphic forms, and enantiomeric impurities.

Qualification of an impurity for a new drug substance has similar concerns as Q3A. The main differences are the reporting, identification, and qualification thresholds (Table 2). The thresholds are basically higher than they were in Q3A; however, there are more categories for dosages. If the qualification

**TABLE 3 Illustration of Reporting Degradation Product Results for Identification and Qualification in an Application**

| Raw Result (%) | Reported Result (%) | Total Daily Intake (TDI) of Degradation Product | Action: Identification Threshold | Action: Qualification Threshold |
|---|---|---|---|---|
| | | *50 mg Maximum Daily Dose*[a] | | |
| 0.04 | Not reported | 20 | None | None |
| 0.2143 | 0.2 | 100 | None | None |
| 0.349 | 0.3 | 150 | Yes | None |
| 0.550 | 0.6 | 300 | Yes | Yes |
| | | *1.9 g Maximum Daily Dose*[b] | | |
| 0.049 | Not reported | 1 | None | None |
| 0.079 | 0.08 | 2 | None | None |
| 0.183 | 0.18 | 3 | Yes | None[c,d] |
| 0.192 | 0.19 | 4 | Yes | Yes[c] |

[a]Reporting threshold 0.1%; TDI rounded result in-micrograms; identification threshold 0.2%; qualification threshold, TDI equivalent to 0.4%.
[b]Reporting threshold 0.05%; TDI rounded result in millingrams; identification threshold 2 mg TDI (equivalent to 0.11%); qualification threshold 3 mg TDI (equivalent to 0.16%).
[c]After identification, if the response factor is determined to differ significantly from the original assumptions, it can be appropriate to remeasure the actual amount of the degradation product present and reevaluate against the qualification threshold.
[d]Although the reported result of 0.18% exceeds the calculated threshold value of 0.16%, in this case the action is acceptable since the TDI (when rounded) does not exceed 3 mg. Chromatograms with peaks labeled (or equivalent data if other analytical procedures are used) from representative batches, including chromatograms from analytical procedure validation studies and from long-term and accelerated stability studies, should be provided. The applicant should ensure that complete degradation product profiles (e.g., chromatograms) of individual batches are available, if requested.

thresholds given in Table 3 are exceeded and data are unavailable to qualify the proposed acceptance criterion of a degradation product, additional studies to obtain such data may be appropriate.

U.S. FDA (CDER) "Guidance for Industry, NDAs: Impurities in Drug Substances" was published in February 2000. The guidance refers applicants to ICH "Q3A: Impurities in New Drug Substances" when seeking guidance on identification, qualification, and reporting of impurities in drug substances that are not considered new drug substances. Q3A was developed by the ICH to provide guidance on the information that should be provided in a new drug application (NDA) in support of impurities in new drug substances that are produced by chemical syntheses. The FDA believes that such guidance on identification, qualification, and reporting of impurities should also be considered when evaluating impurities in drug substances produced by chemical syntheses that are not considered new drug substances. ICH Q3A defines a new drug substance (also referred to as a new molecular entity or new chemical entity) as a designated therapeutic moiety that has not been

**TABLE 4 Amount of Residual Solvent per Day at Maximum Clinical Dose (mg day$^{-1}$) of Drug**

| Residual Solvent | Concentration in Drug Product | Potential Maximum Clinical exposure (mg day$^{-1}$) | Reporting Threshold (0.05%) | Qualification Threshold (0.5%) |
|---|---|---|---|---|
| Name | In ppm | | In mg day$^{-1}$ | in mg day$^{-1}$ |

[a]Based on the determined highest allowable specification level of solvent.

previously registered in a region or member state. The definition also states that a new drug substance may be a complex, a simple ester, or a salt of a previously approved drug substance.

## 2 RESIDUAL SOLVENTS

ICH Q3C is intended to provide guidance for recommending acceptable amounts for residual solvents in pharmaceuticals for the safety of the patient. The guidance recommends use of less toxic solvents and describes levels considered to be toxicologically acceptable for some residual solvents. A complete list of the solvents included in this guidance is provided in a companion document entitled "ICH Q3C—Tables and List" which can be found at the ICH or FDA website. The list is not exhaustive, and other solvents may be used and later added to the list.

Residual solvents in pharmaceuticals are defined here as organic volatile chemicals that are used or produced in the manufacture of drug substances or excipients or in the preparation of drug products. The solvents are not completely removed by practical manufacturing techniques. Appropriate selection of the solvent for the synthesis of drug substance may enhance the yield or determine characteristics such as crystal form, purity, and solubility. Therefore, the solvent may sometimes be a critical parameter in the synthetic process. This guidance does not address solvents deliberately used as excipients nor does it address solvates. However, the content of solvents in such products should be evaluated and justified.

As there are no therapeutic benefits from residual solvents, all residual solvents should be removed to the extent possible to meet product specifications, EMPs, or other quality-based requirements. Drug products should contain no higher levels of residual solvents than can be supported by safety data. Some solvents that are known to cause unacceptable toxicities (carcinogens), such as benzene and carbon tetrachloride (class 1, see Table 1 in ICH, 1997), should be avoided in the production of drug substances, excipients, or drug products unless their use can be strongly justified in a risk–benefit assessment. Some solvents associated with less severe toxicity (nongenotoxic animal carcinogens or possible causative agents of other irreversible toxicity such as neurotoxicity or teratogenicity), such as acetonitrile and chlorobenzene (class 2), should be limited in order to protect patients from potential adverse effects. Ideally, less toxic solvents, such as acetic acid and acetone (class 3), should be

used where practical. Table 4 presents a format for a residual solvent qualification table.

This guidance does not apply to potential new drug substances, excipients, or drug products used during the clinical research stages of development nor does it apply to previously existing marketed drug products.

The guidance applies to all dosage forms and routes of administration. Higher levels of residual solvents may be acceptable in certain cases such as short-term (30 days or less) or topical application. Justification for these levels should be made on a case-by-case basis and discussed with the appropriate FDA division.

The limits of residual solvents may include a value for the permitted daily exposure (PDE), which is the maximum acceptable intake per day of residual solvent in pharmaceutical products. These limits vary depending on the class.

For solvents where quantities are limited to set values in pharmaceutical products because of their inherent toxicity, class 2 (Table 2) should be consulted. PDEs are given to the nearest 0.1 mg day$^{-1}$, and concentrations are given to the nearest 10 ppm.

For solvents with low toxic potential, solvents in class 3 (Table 3) may be regarded as less toxic and of lower risk to human health. Class 3 includes no solvent known as a human health hazard at levels normally accepted in pharmaceuticals. However, there are no long-term toxicity or carcinogenicity studies for many of the solvents in class 3. Available data indicate that they are less toxic in acute or short-term studies and negative in genotoxicity studies. It is considered that amounts of these residual solvents of 50 mg day$^{-1}$ or less (corresponding to 5000 ppm or 0.5% under option 1) would be acceptable without justification. Higher amounts may also be acceptable provided they are realistic in relation to manufacturing capability and GMPs.

For solvents for which no adequate toxicological data were found, the solvents listed (Table 4, ICH, 1997) may also be of interest to manufacturers of excipients, drug substances, or drug products. However, no adequate toxicological data on which to base a PDE were found. Manufacturers should supply justification for residual levels of these solvents in pharmaceutical products.

The Gaylor–Kodell (1980) method of risk assessment is appropriate for class 1 carcinogenic solvents. Only in cases where reliable carcinogenicity data are available should extrapolation by the use of mathematical models be applied to setting exposure limits. Exposure limits for class 1 solvents could be determined with the use of a large safety factor (i.e., 10,000–100,000) with respect to the no-observable-effect level (NOEL). Detection and quantitation of these solvents should be by state-of-the-art analytical techniques.

Acceptable exposure levels in this guidance for class 2 solvents were established by calculation of PDE values according to the procedures for setting exposure limits in pharmaceuticals (*Pharmacopeial Forum*, Nov.–Dec. 1989) and the method adopted by IPCS (International Program on Chemical Safety) for Assessing Human Health Risk of Chemicals [EHC (environmental health criteria) 170, World Health Organization (WHO), 1994]. These methods are similar to those used by the U.S. EPA [IRIS (Integrated Risk Information System)] and the U.S. FDA (*Red Book*) and others. The method is outlined here to give a

better understanding of the origin of the PDE values. It is not necessary to perform these calculations in order to use the PDE values tabulated in Section 4 of this document.

The PDE is derived from the NOEL or the lowest observable effect level (LOEL) in the most relevant animal study as follows:

$$\text{PDE} = \frac{\text{NOEL} \times \text{weight adjustment}}{F_1 \times F_2 \times F_3 \times F_4 \times F_5} \tag{1}$$

The PDE is derived preferably from a NOEL. If no NOEL is obtained, the LOEL may be used. Modifying factors proposed here for relating the data to humans are the same kind of *uncertainty factors* used in EHC 170 (WHO, Geneva, 1994) and *modifying factors* or *safety factors* in *Pharmacopeial Forum*. The assumption of 100% percent systemic exposure is used in all calculations regardless of route of administration.

The modifying factors are as follows:

- $F_1$ = Factor to account for extrapolation between species
  $F_1$ = 5 for extrapolation from rats to humans
  $F_1$ = 12 for extrapolation from mice to humans
  $F_1$ = 2 for extrapolation from dogs to humans
  $F_1$ = 2.5 for extrapolation from rabbits to humans
  $F_1$ = 3 for extrapolation from monkeys to humans
  $F_1$ = 10 for extrapolation from other animals to humans

Factor $F_1$ takes into account the comparative surface area–body weight ratios for the species concerned and for humans. Surface area ($S$) is calculated as:

$$S = kM \times 0.67 \tag{2}$$

Where $M$ is body mass and the constant $k$ has been taken to be 10. The body weights used in the equation are those shown below.

- $F_2$ = Factor of 10 to account for variability between individuals. A factor of 10 is generally given for all organic solvents and 10 is used consistently in this guidance:
- $F_3$ = Variable factor to account for toxicity studies of short-term exposure:
  $F_3$ = 1 for studies that last at least one half-lifetime (1 year for rodents or rabbits; 7 years for cats, dogs, and monkeys)
  $F_3$ = 1 for reproductive studies in which whole period of organogenesis is covered
  $F_3$ = 2 for 6-month study in rodents or 3.5-year study in nonrodents
  $F_3$ = 5 for 3-month study in rodents or 2-year study in nonrodents
  $F_3$ = 10 for studies of shorter duration

In all cases, the higher factor has been used for study durations between the time points (e.g., a factor of 2 for a 9-month rodent study).

- $F_4$ = Factor that may be applied in cases of severe toxicity (e.g., nongenotoxic carcinogenicity, neurotoxicity, or teratogenicity). In studies of reproductive toxicity, the following factors are used:

  $F_4$ = 1 for fetal toxicity associated with maternal toxicity

  $F_4$ = 5 for fetal toxicity without maternal toxicity

  $F_4$ = 5 for a teratogenic effect with maternal toxicity

  $F_4$ = 10 for teratogenic effect without maternal toxicity
- $F_5$ = Variable factor that may be applied if NOEL was not established. When only an LOEL is available, a factor of up to 10 could be used depending on the severity of the toxicity.

The weight adjustment assumes an arbitrary adult human body weight for either sex of 50 kg. This relatively low weight provides an additional safety factor against the standard weights of 60 or 70 kg that are often used in this type of calculation. It is recognized that some adult patients weigh less than 50 kg; these patients are considered to be accommodated by the built-in safety factors used to determine a PDE. If the solvent was present in a formulation specifically intended for pediatric use, an adjustment for a lower body weight would be appropriate.

As an example of the application of this equation, consider a toxicity study of acetonitrile in mice that is summarized in *Pharmeuropa*, Vol. 9, No. 1, Supplement, April 1997, page S24. The NOEL is calculated to be 50.7 mg kg$^{-1}$ day$^{-1}$. The PDE for acetonitrile in this study is calculated as

$$\text{PDE} = \frac{50.7\,\text{mg kg}^{-1}\,\text{day}^{-1} \times 50\,\text{kg}}{12 \times 10 \times 5 \times 1 \times 1} = 4.22\,\text{mg day}^{-1}$$

In this example,

$F_1$ = 12 to account for extrapolation from mice to humans

$F_2$ = 10 to account for differences between individual humans

$F_3$ = 5 because duration of study was only 13 weeks

$F_4$ = 1 because no severe toxicity was encountered

$F_5$ = 1 because NOEL was determined

## 3 EXTRACTABLES AND LEACHABLES

Leachables are chemical entities, either organic or inorganic, that migrate from pharmaceutical container closure system components into a drug product

formulation. Since patients can be exposed to leachables during normal use of a drug product, leachables are of potential safety concern. Extractables are compounds that are forced out of container closure system materials and components under laboratory experimental conditions. All extractables from a given pharmaceutical container closure system and its components are, therefore, potential leachables in a drug product incorporating the same container closure system components. Regulatory concern for regarding leachables and extractables is directly related to the potential for contamination and/or interaction of the drug product formulation with the container closer system, with the greatest concern focused on orally inhaled and nasal drug products (OINDP), which include metered-dose inhalers (MDIs), dry-powder inhalers (DPIs), inhalation solutions, suspensions and sprays, and nasal sprays (Norwood et al., 2007).

> Controlled extraction studies are an extremely important part of the pharmaceutical development process for OINDP and should be performed on critical components as identified by the manufacturer and regulatory authority. As stated in the PQRI L&E Recommendations: A controlled extraction study is a laboratory investigation into the qualitative and quantitative nature of extractables profiles of critical components of an OINDP container closure system. The purpose of a controlled extraction study is to systemically and rationally identify and quantify potential leachables, i.e., extractables, to the extent practicable, and within certain defined analytical threshold parameters.

Controlled extraction studies result in extractables profiles of OINDP components. Extractables profiles contain information which allows the identification, to the extent possible, and quantitation of individual extractables from a given component and therefore are an early indication of potential leachables of concern. Controlled extraction studies generally establish a basis for the development and validation of routine quality control methods for drug product leachables and, finally, allow for the correlation of extractables and leachables profiles. Although information on component composition from suppliers is very useful, helping to inform component selection and guide controlled extraction studies, such knowledge does not provide a complete extractables profile and therefore does not alleviate the requirement for controlled extraction studies no matter how "complete" the information might appear to be.

It is therefore critical that controlled extraction studies be performed properly and thoroughly. Specific expectations for "proper and thorough" controlled extraction studies will ultimately depend on the nature of the OINDP being developed. However, the PQRI L&E Working Group was able to establish some general best practice recommendations for OINDP controlled extraction studies based on the data that the group generated by conducting its own controlled extraction studies on the elastomer and polymer test articles. These recommendations are:

- Controlled extraction studies should employ vigorous extraction with multiple solvents of varying polarity.
- Controlled extraction studies should incorporate multiple extraction techniques.
- Controlled extraction studies should include careful sample preparation based on knowledge of analytical techniques used.
- Controlled extraction studies should employ multiple analytical techniques.
- Controlled extraction studies should include a defined and systematic process for identification of individual extractables.
- Controlled extraction study "definitive" extraction techniques and methods should be optimized.
- During the controlled extraction studies, sponsors should revisit supplier information describing component information.
- Controlled extraction studies should be guided by analytical evaluation thresholds (AETs) that are based on an accepted safety concern threshold.
- Qualitative and quantitative extractables profiles should be discussed with and reviewed by toxicologists so that any potential safety concerns regarding individual extractables, that is, potential leachables, are identified early in the development process.
- Polynuclear aromatics (PNAs), *N*-nitrosamines, and 2-mercaptobenzothiazole (MBT) are "special-case" compounds requiring evaluation by specific analytical techniques and technology-defined thresholds.

The characterization and control of leachables and extractables represent possibly the most significant challenges facing a pharmaceutical development team responsible for the development, registration, and manufacture of an OINDP. Indeed, detecting, identifying, and quantifying organic leachables is a formidable task. In contrast to drug substance or excipient-related impurities, organic leachables can represent a diversity of chemical structures and compound classes and are potentially present at widely varying concentrations in any particular OINDP. Additionally, the information available to a pharmaceutical development team on container closure system component composition and processing, which is provided by the component supplier, is often incomplete. In some cases, the supplier may provide no information. Thus, when an extractables study is first undertaken, the development team may only have a limited idea of what to look for and what extraction techniques and analytical methods to use for identification and assessment of potential leachables.

## 4 RESIDUAL METALS AND METAL CATALYSTS

In early 2008, the EMEA promulgated a standard for metals as impurities in pharmaceuticals. They organized metals of concern into categories, as presented in Table 5.

**TABLE 5 Class Exposure and Concentration Limits for Individual Metal Catalysts and Metal Reagents**

| | Oral Exposure | | Parenteral Exposure | | Inhalation Exposure[a] |
|---|---|---|---|---|---|
| Classification | PDE ($\mu g\,day^{-1}$) | Concentration (ppm) | PDE ($\mu g\,day^{-1}$) | Concentration (ppm) | PDE ($ng\,day^{-1}$) |
| Class 1A: Pt, Pd | 100 | 10 | 10 | 1 | Pt: 70[a] |
| Class 1B: Ir, Rh, Ru, Os | 100[b] | 10[b] | 10[b] | 1[b] | |
| Class 1C: Mo, Ni, Cr, V; metals of significant safety concern | 250 | 25 | 25 | 2.5 | Ni: 100 Cr (VI): 10 |
| Class 2: Cu, Mu, metals with low safety concern | 2,500 | 250 | 250 | 25 | |
| Class 3: Fe, Zu; metals with minimal safety concern | 13,000 | 1,300 | 1,300 | 130 | |

[a]Pt as hexachloroplatinic acid.
[b]Subclass limit: the total amount of listed metals should not exceed the indicated limit.

If synthetic processes of pharmaceutical substances are known or suspected to lead to the presence of metal residues due to the use of a specific metal catalyst or metal reagent, a concentration limit and validated test for residues of each specific metal should be set. All concentration limits should be realistic in relation to analytical precision, manufacturing capability, and reasonable variation in the manufacturing process. Since the use of metal catalysts or metal reagents during synthesis is restricted to validated and controlled chemical reactions, limitation of their residues in pharmaceutical substances itself will normally be sufficient. A limit for a metal residue in the pharmaceutical substance may however be replaced by a limit for that metal residue in the final medicinal product, as described below.

For pharmaceutical products administered via the oral, parenteral, or inhalation route of administration, two options are available when setting a concentration limit for a metal residue:

*Option 1:* For each metal, the concentration limit in parts per million (ppm) as stated in Table 5 can be used. The concentration limits in Table 1 have been calculated using equation (3) below by assuming a daily dose of 10 g of the drug product:

$$\text{Concentration ppm} = \frac{\text{PDE } \mu g\,day^{-1}}{\text{daily dose } g\,day^{-1}} \tag{3}$$

If all pharmaceutical substances in a drug product meet the option 1 concentration limit for all metals potentially present, then all these substances may be used in any proportion in the drug product as long as the daily dose of the drug product does not exceed 10 g day$^{-1}$. When the daily dose of the drug product is greater than 10 g day$^{-1}$, option 2 should be applied.

*Option 2a:* The PDE in terms of micrograms per day as stated in Table 1 can be used together with the actual daily dose of a pharmaceutical substance in the drug product to calculate the concentration of residual metal allowed in that pharmaceutical substance.

*Option 2b:* Alternatively, it is not considered necessary for each pharmaceutical substance to comply with the limits given in option 1 or the calculated limits using option 2a.

The PDE in terms of micrograms per day as stated in Table 1 can also be used with the known maximum daily dose of the drug product to determine the concentration of a metal residue originating from any of the pharmaceutical substances in the drug product (not the substance). This approach is considered acceptable provided that it has been demonstrated that the metal residue has been reduced to the practical minimum in every substance. This approach implies that the maximum levels of a metal in certain substances may be higher than the option 1 or option 2a limit but that this should then be compensated for by lower maximum levels in the other substances.

For pharmaceutical products applied via other routes of administration, the concentration limits should be set in consideration of the route of administration.

Without proper justification, parenteral limits/PDEs should be used for pharmaceutical substances that are administered by other routes of administration, including inhalation. Oral limits/PDEs may be applied if the absorption by other routes of administration is not likely to exceed the absorption following oral administration. For example, for cutaneous administration, oral concentration limits/PDEs are considered acceptable.

Platinum salts have been shown to be allergenic, with hexachloroplatinic acid being clearly the most allergenic (Malo, 2005). Consequently a specific limit for inhalation exposure for this molecule has been set at 70 ng day$^{-1}$ (see monograph). Chromium VI and nickel, when inhaled, have been associated with carcinogenicity. Therefore specific limits for inhalation exposure have been set for chromium VI at 10 ng day$^{-1}$ and for nickel at 100 ng day$^{-1}$ (see respective monographs).

For pharmaceutical products used for short-term and for life-saving indications, as the PDEs and concentration limits mentioned in this guideline are based on chronic use, higher PDEs and concentration limits may be acceptable in cases of short-term use (30 days or less). For instance, this may be applicable to contrasting agents, antidotes, or products for diagnostic use. This may however only be applied if neither an option 1 nor an option 2 limit is feasible.

Specific risk–benefit considerations, such as for compounds used for life-saving indications, may also warrant the use of higher limits. Justifications should be made on a case-by-case basis.

## REFERENCES

Ball, D., Blanchard, J., Jacobson-Kram, D., McClellan, R., McGovern, T., Norwood, D. L., Vogel, M., Wolff, R., and Nagao, L. M. (2007). Development of safety qualification thresholds and their use in drug product evaluation. *Toxicol. Sci.* 97(2):226–236.

Code of Federal Regulations, Title 21, Vol. 5 (April 1, 2005). 21 CFR 314.3(b); Title 21, Food and Drugs, Chapter I—Food and Drug Administration Department of Health and Human Services, Subchapter D—Drugs for human use. Part 314, applications for FDA approval to market a new drug, subpart A—General provisions, Section 314.3 Definitions. Revised April 1, 2005. Available: www.fda.gov.

European Medicines Agency (EMEA) (2004). Guidelines on the limits of genotoxic impurities. EMA, London.

European Medicines Agency (EMEA) (2008). Guideline on the specification limits for residues of metal catalysts or metal reagents. EMA, London.

Gaylor, D. W., and Kodell, R. L. (1980). Linear interpolation algorithm for low dose assessment of toxic substance. *J. Environ. Pathol. Toxicol.* 4:305.

International Conference on Harmonisation (ICH) (1995, Mar.). Guideline for industry, Q2A—Text on validation of analytical procedures. Available: http://www.fda.gov/cder/guidance/index.htm.

International Conference on Harmonisation (ICH) (1996a, Nov.). Guidance for industry, QiB—Photostability testing of new drug substances and products. Available: http://www.fda.gov/cder/guidance/index.htm.

International Conference on Harmonisation (ICH) (1996b, Nov.). Guidance for industry, Q2B—Validation of analytical procedures: Methodology. Available: http://www.fda.gov/cder/guidance/index.htm.

International Conference on Harmonisation (ICH) (1997, Dec.). Guidance for industry, Q3C—Impurities: Residual solvents. Available: http://www.fda.gov/cder/guidancelindex.htm.

International Conference on Harmonisation (ICH) (1999). Guidance for industry, Q6B—Specifications: Test procedures and acceptance criteria for bio-technological/biological products. Available: http://www.fda.gov/cder/guidance/index.htm.

International Conference on Harmonisation (ICH) (2000, Dec.). Guidance for industry, Q6A—Specifications: Test procedures and acceptance criteria for new drug substances and new drug products: Chemical substances. Available: http://www.fda.gov/cder/guidance/index.htm.

International Conference on Harmonisation (ICH) (2001, Aug.). Guidance for industry, QIA—Stability testing of new drug substances and products. Available: http://www.fda.gov/cder/guidance/index.htm.

International Conference on Harmonisation (ICH) (2003a, Feb.). Guidance for industry, Q3A(R)—Impurities in new drug substances. Available: http://www.fda.gov/cder/guidance/index.htm.

International Conference on Harmonisation (ICH) (2003b, Nov.). Guidance for industry, Q3B(R)—Impurities in new drug products. Available: http://www.fda.gov/cder/guidance/index.htm.

International Conference on Harmonisation (ICH) (2003c, Nov.). Guidance for industry, Q3C—Tables and list (2003c, Nov.). Available: http://www.fda.gov/cder/guidance/index.htm.

Malo, J.-L. (2005). Occupational rhinitis and asthma due to metal salts. *Allergy* 60:138–139.

Norwood, D. L. (2007). Understanding the challenges of extractables and leachables for the pharmaceutical industry—Safety and regulatory environment for pharmaceuticals. *Am. Pharm. Rev.* 10(2):32–39.

Norwood, D. L., Paskiet, D., Roberto, M., Feinberg, T., Schroeder, A., Poochikian, G., Wang, Q., Deng, T. J., DeGrazio, F., Munos, M. K., and Nagao, L. M. (2007). Best practices for extractables and leachables in orally inhaled and nasal drug products: An overview of the PQRI recommendations. *Pharm. Res.* 25:727–738.

Osterberg, R. E. (2005a). Extractables and leachables in drug products. *Am. Coll. Toxicol. Newslett.* 25(1):2–3.

Osterberg, R. E. (2005b). Potential toxicity of extractables and leachables in drug products. *Am. Pharm. Rev.* 8(2).

U.S. Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research (CDER) (1998, Oct.). Guidance for industry, metered dose inhaler (MDI) and dry powder inhaler (DPI), drug products chemistry, manufacturing, and controls documentation. Available: http://www.fda.gov/cder/guidance/index.htm.

U.S. Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research (CDER) (2000, Feb.). Guidance for industry, NDAs: Impurities in drug substances. Available: http://www.fda.gov/cder/guidance/3622fnl.htm.

U.S. Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research (CDER) (2005, Aug.). Guidance for industry, ANDAs: Impurities in drug products. CDER, Washington, DC.

U.S. Food and Drug Administration (FDA) (2005): http://www.fda.gov/cder/guidance/cGMPs/packaging.htm.

# *Selected Regulatory and Toxicological Acronyms*

| | |
|---|---|
| 510(k) | Premarket notification for change in a device |
| AALAS | American Association Laboratory Animal Science |
| AAMI | Association for the Advancement of Medical Instrumentation |
| ABT | American Board of Toxicology |
| ACGIH | American Conference of Governmental Industrial Hygienists |
| ACT | American College of Toxicology |
| ADE | Adverse drug event (of drug substances) |
| ADI | Allowable daily intake |
| AIDS | Acquired immunodeficiency syndrome |
| AIMD | Active implantable medical device |
| ANSI | American National Standards Institute |
| APHIS | Animal and Plant Health Inspection Service |
| ASTM | American Society for Testing and Materials |
| CAS | Chemical Abstract Service |
| CBER | Center for Biologics Evaluation and Research (FDA) |
| CDER | Center for Drug Evaluation and Research (FDA) |
| CDRH | Center for Devices and Radiological Health (FDA) |
| CFAN | Center for Food and Nutrition (FDA) |
| CFR | Code of Federal Regulations |
| CIIT | Chemical Industries Institute of Toxicology |
| CPMP | Committee on Proprietary Medicinal Products (U.K.) |

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

| | |
|---|---|
| CSE | Control standard endotoxin |
| CSM | Committee on Safety of Medicines (U.K.) |
| CTC | Clinical trial certificate (U.K.) |
| CTX | Clinical trial certificate exemption (U.K.) |
| CVM | Center for Veterinary Medicine (FDA) |
| DART | Development and reproduction toxicology |
| DHHS | Department of Health and Human Services |
| DIA | Drug Information Associates |
| DMF | Drug (or device) master file |
| DSHEA | Dietrary Supplement Health and Education Act |
| EEC | European Economic Community |
| EFPIA | European Federation of Pharmaceutical Industries Association |
| EM | Electron microscopy |
| EPA | Environmental Protection Agency |
| EU | European Union |
| FCA | Freund's complete adjuvant |
| FDA | Food and Drug Administration |
| FDCA | Food, Drug and Cosmetic Act |
| FDLI | Food and Drug Law Institute |
| FIFRA | Federal Insecticides, Fungicides and Rodenticides Act |
| GCP | Good clinical practices |
| GLP | Good laboratory practices |
| GMP | Good manufacturing practices |
| GPMT | Guinea pig maximization test |
| HEW | Department of Health, Education and Welfare (no longer existent) |
| HIMA | Health Industry Manufacturer's Association |
| HSDB | Hazardous Substances Data Bank |
| IARC | International Agency for Research on Cancer |
| ICH | International Conference on Harmonisation |
| ID | Intradermal |
| IDE | Investigational device exemption |
| IND(A) | Investigational new drug application |
| INN | International nonproprietary names |
| IP | Intraperitoneal |
| IRAG | Interagency Regulatory Alternatives Group |
| IRB | Institutional review board |
| IRLG | Interagency Regulatory Liaison Group |
| ISO | International Standards Organization |
| IUD | Intrauterine device |
| IV | Intravenous |
| JECFA | Joint Expert Committee for Food Additives |
| JMAFF | Japanese Ministry of Agriculture, Forestry, and Fishery |
| JPMA | Japanese Pharmaceutical Manufacturers Association |
| LA | Licensing Authority (U.K.) |

| | |
|---|---|
| LAL | *Limulus* amebocyte lysate |
| $LD_{50}$ | Lethal dose 50: dose calculated to kill 50% of subject population, median lethal dose |
| LOEL | Lowest observed effect level |
| MAA | Marketing authorization application (EEC) |
| MCA | Medicines Control Agency |
| MD | Medical device |
| MedDRA | Medical Dictionary for Regulatory Activities |
| MHW | Ministry of Health & Welfare (Japan) |
| MID | Maximum implantable dose |
| MOE | Margin of exposure |
| MOU | Memorandum of understanding |
| MRL | Maximum residue limits |
| MSDS | Material safety data sheet |
| MTD | Maximum tolerated dose |
| NAS | National Academy of Science |
| NCTR | National Center for Toxicological Research |
| NDA | New drug application |
| NIH | National Institutes of Health |
| NIOSH | National Institute of Occupational Safety and Health |
| NK | Natural killer |
| NLM | National Library of Medicine |
| NOEL | No-observable-effect level |
| NTP | National Toxicology Program |
| ODE | Office of Device Evaluation |
| OECD | Organization for Economic Co-operation and Development |
| PDI | Primary dermal irritancy |
| PDN | Product development notification |
| PEL | Permissible exposure limit |
| PhRMA | Pharmaceutical Research and Manufacturers Association |
| PL | Produce license (U.K.) |
| PLA | Produce license application |
| PMA | Premarket approval application |
| PO | Per os (orally) |
| PTC | Points to consider |
| QAU | Quality assurance unit |
| RAC | Recombinant DNA Advisory Committee |
| RCRA | Resources Conservation and Recovery Act |
| RTECS | Registry of Toxic Effects of Chemical Substances |
| SARA | Superfund/Amendments and Reauthorization Act |
| SC | Subcutaneous |
| SCE | Sister chromatic exchange |
| SNUR | Significant new use regulations |
| SOP | Standard operating procedure |
| SOT | Society of Toxicology |

| | |
|---|---|
| SRM | Standard Reference Materials (Japan) |
| STEL | Short-term exposure limit |
| TLV | Threshold limit value |
| USAN | United States Adopted Name Council |
| USDA | United States Department of Agriculture |
| USP | United States Pharmacopeia |
| VAERS | Vaccine adverse-event reporting system |
| VSD | Vaccine safety data link |
| WHO | World Health Organization |

# *Definition of Terms and Lexicon of Clinical Observations in Nonclinical (Animal) Studies*

*MOVEMENT*

**Anesthetized** The absence of or reduced response to external stimuli accompanied with a loss of righting reflex.

**Ataxia** Incoordination of muscular action involving locomotion, including loss of coordination and unsteady gait.

**Catalepsy** Condition characterized by waxy rigidity of muscles such that animal tends to remain in any position in which it is placed.

**Hyperactivity** Abnormally high level of motor activity.

**Hypersensitivity** Abnormally strong reaction to external stimuli such as noise or touch.

**Lethargy** State of deep and prolonged depression stupor from which it is possible to be aroused followed by immediate relapse.

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

**Low carriage** Animal's torso is carried very close to the ground during movement.

**Paralysis** Inhibition or loss of motor function; may be characterized by affected portion of body.

**Prostrate** Animal assumes a recumbent position due to loss of strength or exhaustion and may slow intermittent uncoordinated movements.

**Righting reflex** Ability of an animal, when placed on its back, to regain a position on all fours.

**Unsteady gait** Erratic manner or style of walking.

*RESPIRATION*

**Audible respiration** Abnormal respiratory sound heard while listening to breathing of animal (e.g., wheezing and rales).

**Bradypnea** Abnormal slowness of respiration rate.

**Cheyne–Stokes respiration** Breathing characterized by rhythmic waning and waxing of depth of respiration, with regularly recurring periods of apnea: seen especially in coma resulting from affection of nervous centers.

**Dyspnea** "Shortness of breath"; difficult or labored breathing.

**Gasping** Spasmodic breathing with the mouth open or laborious respiration with breath caught convulsively.

**Hyperpnea** Deep and rapid breathing.

**Hypopnea** Shallow and slow breathing.

**Irregular respiration** No definite cycle or rate of breathing.

**Labored respiration** Forced or difficult, usually irregular breathing.

**Tachypnea** Excessive rapidity of respiration rate.

*CONDITION OF SKIN AND FUR*

**Alopecia** Deficiency of hair (baldness).

**Cyanosis** Visible skin and/or mucous membranes turn dusky blue due to lack of oxygenation of blood.

**Necrosis** Actual tissue destruction, masses of dead/destroyed tissue.

*UROGENITAL REGION*

**Anuria** Absence of or sharp decline in urine excretion.

**Diarrhea** Abnormal frequency and liquidity of fecal discharge.

**Polyuria** Abnormally sharp increase in amount of urine excretion.

*CONVULSIONS AND TREMORS*

**Clonic** Often seen as a "paddling" motion of forelegs of animal.

**Convulsions** Transient, self-sustaining electrical dysrhythmias which have tendency to recur. Convulsions are generally associated with a finite period of unconsciousness and have a muscular involvement manifested as disorganized limb movements.

**Fasciculation** Rapid, often continuous contraction of bundle of skeletal muscle fibers which does not produce purposeful movement (twitching).

**Tonic** Muscular contraction, keeping limbs in fixed position, generally extended to rear.

**Torsion** Postural incoordination or rolling. Generally associated with vestibular (ear canal) system.

**Tremor** Fine oscillating muscular movements which may or may not be rhythmic.

*CONDITION OF EYES*

**Blepharospasm** Twitching or spasmodic contraction of orbicularis oculi muscle.

**Chemosis** Edema of conjunctiva(e)—conjunctival tissue responds to noxious stimuli by swelling.

**Chromodacryorrhea** Response of reddish conjunctival exudate; no blood cells present in exudate (i.e., not true "bloddy tears").

**Conjunctivitis** Inflammation of conjunctiva (mucous membrane which lines eyelids and is reflected into eyeball).

**Exophthalmos** Abnormal protrusion of eyeball from orbit.

**Lacrimation** Secretion of tears.

**Miosis** Constriction of pupil.

**Mydriasis** Dilation of pupil.

**Nystagmus** Abnormal involuntary movement of eyes. It may be rotational or horizontal or vertical plane.

**Ocular exudate** Secretion (usually transparent and yellow) directly from eye.

**Opacity** Loss of transparency of eyeball.

**Pinpoint pupils** Ultimate state of miosis.

**Ptosis** Refers to dropping of upper eyelid, thought to be due to impaired conduction in third cranial nerve.

*MISCELLANEOUS*

**Analgesia** Absence of (or reduced response to) painful stimuli.

**Hunched posture** Drawing-in of both ends of body and extremities with sharp arching of back.

**Kyphosis** Humpback—abnormal curvature and dorsal prominence of vertebrae column.

**Nasal discharge** Fluid secretion from nostrils.

**Piloerection** Body hair stands on end; dilation of pupils usually accompanies piloerection.

**Salivation** Excessive secretion of saliva from mouth.

**Straub tail** Condition, especially in mice, in which animal carries its tail in erect (vertical or nearly vertical) position. This sign is commonly associated with chemicals (e.g., morphine) that bind to opiate receptors.

*REFLEXES*

**Corneal reflex** Closure of eyelids in response to corneal touch (e.g., with soft brush bristle).

**Grip strength (or screen grip)** Measure of grip strength of forelimbs or hindlimbs; may be evaluated quantitatively or by subjective estimate or impairment (rodents only).

**Pinna reflex** Twitch of outer ear in response to gentle touch.

**Preyer's reflex (auditory startle response)** Involuntary movement of outer ears produced by auditory stimulus (especially in rats).

**Pupillary reflex** Contraction of pupil in response to light stimulation of retina.

**Righting reflex** Ability to land on (when dropped) or regain normal stance on all four limbs.

**Startle reflex** Response to sharp sound, touch, or other startling stimulus; response may range from "absent," to "normal," to "hyperreactive," including exaggerated jerking, jumping, frantic attempts to escape, and even convulsion.

# *Notable Regulatory Internet Addresses*

| Organization or Publication | Web Address (URL) | Sample Main Topics |
|---|---|---|
| ABPI | http://www.abpi.org.uk/ | |
| Adverse Reactions Bulletin | http://www.thomsonscience.co | |
| Agency for Toxic Substances and Disease Registry | www.atsdr.cdc.gov | |
| Association of Clinical Biochemists | http://www.leeds.ac.uk/acb/ | Items of general medical interest and an assay finder to help researcher find methods or labs to measure a wide variety of hormones, metals, enzymes, and drugs in body fluids |
| Australian Therapeutic Goods Administration | http://www.health.gov.au/tga | Medical devices; GMP codes; Parliamentary Secretary's Working; status document; party on complementary medicines; medical releases; publications; site map; related sites |

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

| Organization or Publication | Web Address (URL) | Sample Main Topics |
|---|---|---|
| BioMedNet | http://www.cursci.co.uk/BioMedNet/biomed.html/ or http://www.BioMedNet.com | World Wide Web club for biological and medical community (free membership) |
| Canadian Health Protection Board | http://www.hwc.ca/hpb | |
| Canadian Health Protection Branch | http://www.hc-sc.gc.ca/hpb | Medical devices; Chemical hazards; food; product safety; science advisory board; diseases; radiation protection; drugs; HPB transition policy, planning, and coordination |
| Centre for Medicines Research | http://www.cmr.org/ | |
| ChemInfo | www.indiana.edu/~cheminfo/ca_csti.html | SirCH: Chemical Safety or Toxicology Information |
| Clinical Pharmacology Drug Monograph Service | http://www.cponline.gsm.com | |
| Clinician's Computer-Assisted Guide to the Choice of Instruments for Quality of Life Assessment in Medicine | http://www.glamm.com/ql/guide.htm | Contains hypertext with references to QoL measurements divided into (a) general diseases, (b) specific diseases and therapies, (c) health organizations, and (d) bibliography |
| ClinWeb | http://www.ohsu.edu/clinweb | Oregon Health Sciences University |
| CNN Interactive (Health) | http://www.cnn.com/HEALTH/index.html | Up-to-date information on health issues including drug safety concerns and withdrawals |
| Code of Federal Regulations | http://www.access.gpo.gov/nara/cfr/index.html or http://www.access.gpo.gov/su_docs/aces/aces140.html | For proposed rules and regulations |
| | http://www.access.gpo.gov/nara/cfr/cfr-table-search.html | NARA code sections |
| Committee on Safety of Medicines (CSM) | http://www.open.gov.uk/mca/csmhome.htm | |
| Cornell Legal Library | http://www.law.cornell.edu | Code of Federal Regulations; Supreme Court Decisions; U.S. Code; Circuit Courts of Appeal |
| Current problems in pharmacovigilance | http://www.opwn.gov.uk/mca/mcahome.htm | |

| Organization or Publication | Web Address (URL) | Sample Main Topics |
|---|---|---|
| Cutaneous drug reactions | http://triz.dermatology.uiowa.edu/home.html | |
| DIA home page | http://www.diahome.org | Home page of Drug Information Association |
| Doctor's Guide to the Internet | http://www.psigroup.com | |
| Documents for Clinical Research | http://www.ams.med.unigoettingen.de/~rhilger/Document.html | *Declaration of Helsinki*, other documents and collection of related sites |
| Druginfonet | http://www.druginfonet.com | |
| EC DGXIII Telecommunications | http://www.ispo.cec.be/ | Information |
| EMBASE | http://www.healthgate.com/healthGate/price/embase.html | |
| European Protection Agency (EPA) | www.epa.gov | |
| Eudra Net: Network Services for the European Union Pharmaceutical Regulatory Sector | http://www.eudra.org | Includes information on European Agency for the Evaluation of Medicinal Products. |
| EMEA | http://www.eudra.org/emea.html | |
| Europa | http://www.cec.lu | Official website of European Union |
| European Agency for the Evaluation of Medicinal Products | http://www.eudra.org/en_home.htm | What's new; documents forum; Other sites |
| European Sites | http://www.eucomed.be/eucomed/links/links.htm | European institutions; related sites |
| European Pharmacovigilance Research Group | http://www.ncl.ac.uk/~neprg/ | |
| Food and Drug Administration (FDA) | www.fda.gov | Foods; human drugs; biologics; animal drugs; cosmetics; medical devices/radiological health |
| FDA CBER: Center for Biologics Evaluation and Research | http://www.fda.gov/cber | |
| CBER What's New | http://www.fda.gov/cber/whatsnew.htm | |
| FDA CDER: Center for Drug Evaluation and Research | http://www.fda.gov/cder | |
| FDA Adverse Events Database | http://www.fda.gov/cder/adr | |
| CDER What's New | http://www.fda.gov/cder/whatsnew.htm | |

| Organization or Publication | Web Address (URL) | Sample Main Topics |
|---|---|---|
| FDA CDRH | www.fda.gov/cdrh/index.html | Home page |
| Search site | www.fda.gov/cdrh/search.html | Search CDRH site |
| Comment | www.fda.gov/cdrh/comment4.html | Comment on CDRH site |
| Device advice | www.fda.gov/cdrh/devadvice/32.html | |
| PDF reader | www.fda.gov/cdrh/acrobat.html | |
| FDA CFSAN: Center for Food Safety and Applied Nutrition | http://vm.cfsan.fda.gov | |
| FDA Center for Toxicological Research | http://www.fda.gov/nctr | |
| FDA CVM: Center for Veterinary Medicine | http://www.fda.gov/cvm | |
| FDA—Bioengineered food | http:www.fda.gov/oc/biotech/default.htm | |
| FDA—Breast Implants | http:www.fda.gov/cdrh/breastimplants/index.html | |
| FDA—Cosmetics | http://vm.cfsan.fda.gov//~lrd/cosmetm.html | |
| FDA—Dietary supplements | http:vm.cfsan.fda.gov/~dms/supplmt.html | |
| FDA's Electronic Freedom of Information Act | http:www.fda.gov/foi/foia2.htm | |
| FDA—Field Operations | www.fda.gov/ora/ | What's new; import program; inspectional, science and compliance references; federal/state relations |
| Common Technical Document for the Registration of Pharmaceuticals for Human use: 08-24-00 | http://www.fda.gov/cder/guidance/4022dfts.htm | |
| Design Controls | www.fda.gov/ora/inspect_ref/qsreq/dcrpgd.html | Design control report and guidance text |
| | www.fda.gov/ora/inspect_ref/igs/elec_med_dev/emcl.html | Guide to inspections of electromagnetic compatibility aspects of medical device quality systems text |
| Guide to Inspections of Quality Systems | www.fda.gov/ora/inspect_ref/igs/qsit/qsitguide.htm | QSIT inspection handbook text |
| Guide to Inspections of Quality Systems | www.fda.gov/ora/inspect_ref/igs/qsit/QSITGUIDE.PDF | PDF version of QSIT inspection handbook text |
| Photosafety Testing 07-05-00 | http://www.fda.gov/cder/guidance/3281dft.htm | |
| Skin Irritation and Sensitization Testing of Generic Transdermal Drug Products 06:01:00 | http://www.fda.gov/cder/guidance/2887fnl.htm | |

| Organization or Publication | Web Address (URL) | Sample Main Topics |
|---|---|---|
| FDA—MedWatch | http://www.fda.gov/medwatch/ | FDA drug adverse-event reporting system |
| FDA—Tampons | http://www.fda.gov/oc/opacpm/topicindexes/tampons.html | |
| Food and Drug Law Institute | http://www.fdli.org | Special interest; publications; multimedia; order products; academic programs; directory of lawyers and consultants; contact us |
| Health Industry and Manufacturers Association (HIMA) | http://www.himanet.com | About HIMA; newsletter; HIMA calendar; industry resources; business opportunities; FDA/EPA/OSHA; reimbursement/payment; global year 2000; government relations; public relations; small company; diagnostics |
| Health on the Net | http://www.hon.ch | |
| Health information on Internet | http://www.wellcome.ac.uk.healthinfo/ | New bimonthly newsletter from the Wellcome Trust and the RSM |
| Hyppos Project | http://ifinet.it/hypposnet | Information in Italian and English about Hyppos Project, which has led to development of QoL tool for measurement of hypertensive patients in Italy. It contains a description of the project, the tool, publications about the development of the tool and its application, plus general references to QoL and hypertension. |
| International Classification of Disease (ICD)-10 | http://www.cihi.ca.newinit/scope.htm | |
| International Conference on Harmonisation (ICH) 3 Home Page | http://cc.umin.u-tokyo.ac.jp/ich/ich3.html | Official ICH website with documents (needs password) |
| ICH documents | http://www.pharmweb.net/pwmirror/pw9/ifpma/ich1/html | |
| International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human use | http://www.ifpma.org/ich1.html | |

| Organization or Publication | Web Address (URL) | Sample Main Topics |
|---|---|---|
| International Federation of Pharmaceutical Manufacturers | http://www.ifpharma.com | ICH documents and postings; international pharmaceutical issues |
| International Regulatory Monitor (Monitor) | http://www.gonsi.com/pubs | Editorial portion of newsletter |
| International Society of Pharmacoepidemiology | http://www.pharmacoepi.org | |
| Internet Grateful Med | www.igm.nlm.nih.gov | |
| InterPharma | http://www.interpharma.co.uk | The latter are vast sites with links to other databases for pharmaceutical support sites—http://www.MedsiteNavigator.com |
| *JAMA* | http://www.amahyphen;assn.org/jama | Gives many other useful U.S. sites |
| Japanese Ministry of Health and Welfare | http://www.mhw.go.jp/english/index.html | Organization; Y2K problem; statistics; white paper; related sites |
| Library of Congress | http://thomas.loc.gov | Searchable database of federal legislation, Congressional Record, and committee information |
| Market and Exploitation of Research | http://www.cordis.lu | |
| Medical Device Link | http://www.devicelink.com | News; consultants; bookstore; links; discussion; magazines (MDDI; MPMN; IVD Technology) |
| Medicines Control Agency (MCA) | http://www.opengov.uk/mcahome.htm | |
| Medical Matrix | http://www.medmatrix.org | |
| Medical Research Council | http://nimr.mcr.ac.uk/MRC/ | |
| MEDLINE (free) | http://www.ncbi.nlm.nih.gov/PubMed or http://www.medmatrix.org/Spages/medline.asp | List of free sites |
| MEDLINE | http://www.medmatrix.org/SPAges/medline.asp or http://www.medsitenavigator.com/medline/medline.html | Metasite with full and changing MEDLINE search engines; list of free sites |
| Medscape | http://www.medscape.com | |
| Multilingual glossary of medical terms | http://allserv.rug.ac.be/~rvdstich/eugloss/welcome.html | |
| National Archives and Public Records Administration | http://www.access.gpo.gov/su_docs/aces/aces140.html | Code of Federal Regulations; *Federal Register*; laws; U.S. Congress information |

| Organization or Publication | Web Address (URL) | Sample Main Topics |
|---|---|---|
| National Institutes of Health (U.S.) | http://www.nih.gov | |
| National Library Network | www.toxnet.nlm.nih.gov | TOXNET: Toxicology Data Network, a cluster of databases on toxicology, hazardous chemicals, and related areas |
| National Toxicology Program | http://ntp-server.niehs.nih.gov/ | |
| New Quality System (QS) Regulation | www.fda.gov/bbs/topics/ANSWERS/ANS00763.html | FDA talk paper announcing GMP final rule text |
| Organised Medical Network Information | http://www.omni.ac.uk | |
| Pharmaceutical and Medical Safety Bureau—Japan | http://www.mhlw.go.jp/english | |
| PharminfoNet | http://www.pharminfo.com or http://www.pharminfo.com/phrmlink.html | Independent assessment of therapeutics and advances in new drug development |
| Pharmweb | http://www.pharmweb.net | Information resource for pharmaceutical and health-related information |
| Quality of Life | http://www.glamm.com/ql/guide.htm | Choice of instrument |
| Quality of Life Assessment in Medicine | http://www.glamm.com/q1/ursl.htm | Contains hypertext with references to QoL measurements divided into (a) assessment tools, (b) reference organizations and groups, (c) diseases, symptoms, and specific populations, (d) the top-10 journals that publish articles of interest to QoL assessment in medicine, (e) methodology, and (f) bibliographical research. |
| Regulatory Affairs Professionals Society (RAPS) | http://www.raps.org | Certificates; resource center; publications; chapters; related links; contacting RAPS |
| Reuters Health Information Services | http://www.reutershealth.com | |
| SCRIP: World Pharmaceutical News | http://www.pjbpubs.co.uk/scrip | |
| SNOMED | http://snomed.org | Systemised Nomenclature of Human and Veterinary Medicines |

| Organization or Publication | Web Address (URL) | Sample Main Topics |
|---|---|---|
| Swedish Medical Products Agency | http://www.mpa.Se | |
| U.S. Department of Agriculture (USDA) | http://www.usda.gov | |
| Food Safety | http://www.foodsafety.gov/ | |
| USDA FMS: Farm Service Agency | http://www.fsa.usda.gov/pas/default.asp | |
| USDA FSA: Food and Nutrition Service | http://www.fns.usda.gov/fns/ | |
| USDA FSIS: Food Safety and Inspection Service | http://www.usda.gov/fsis | |
| U.S. Department of Commerce | http://204.193.246.62 | Bureau of Export Administration; International Trade Association; patent and trademark; National Institute of Standards and Technology |
| U.S. Pharmacopeia | www.usp.org/prn | |
| University of Pittsburgh | www.pitt.edu | |
| World Health Organization | http://www.who.int | Governance; health topics; information sources; reports; directorgeneral; about WHO; International Digest of Health; legislation (http://www.who.int/pub/dig.html) |
| WHO Collaborating Centre for International Drug Monitoring | http://www.who.ch/ or http://www.who.pharmasoft.se | |

# *Glossary of Terms Used in Clinical Evaluation of Therapeutic Agents*

**Abnormality** Sign, symptom, or laboratory result not characteristic of normal individuals.

**Adverse event** Unwanted effects that occur and are detected in populations. The term is used whether there is or is not any attribution to a medicine or other cause. Adverse events may be known parts of a disease that are observed to occur within a period of observation, and they may be analyzed to test for their frequency in a given population or trial. This is done to determine if there is an unexpectedly increased frequency resulting from nondisease factors such as medicine treatment. The term adverse event or adverse experience is used to encompass adverse reactions plus any injury, toxicity, or hypersensitivity that may be medicine related as well as any medical events that are apparently unrelated to medicine that occur during the study (e.g., surgery, illness, and trauma). See definition of *adverse reaction*.

**Adverse experience** See *adverse event*.

**Adverse reaction** Unwanted effect(s) (i.e., physical and psychological symptoms and signs) resulting from treatment. A less rigid definition of adverse reaction includes the previous definition plus any undesirable effect or problem that is present during the period of treatment and may or may not be a well-known or obvious complication of the disease itself. Thus, many

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

common personality, physical, psychological, and behavioral characteristics that are observed in medicine studies are sometimes characterized as adverse reactions even if they were present during baseline.

Synonyms of adverse reactions generally include adverse medical effects, untoward effects, side effects, adverse drug experiences, and adverse drug reactions. Specific distinctions among some of these terms may be defined operationally. For example, the term adverse reaction is used to denote those signs and symptoms at least possibly related to a medicine, whereas the term adverse experiences is used to include nonmedicine-related medical problems in a trial such as those emanating from trauma or concurrent illness. Distinctions among side effects, adverse events, and adverse reactions are illustrated in the definitions of the two former terms.

**Bias** (1) Point of view that prevents impartial judgment on issues relating to that point of view. Clinical trials attempt to control this through double blinding. (2) Any tendency for a value to deviate in one direction from the true value. Statisticians attempt to prevent this type of bias by various techniques, including randomization.

**Clinical significance** Quality of a study's outcome that convinces physicians to modify or maintain their current practice of medicine. The greater the clinical significance, the greater is the influence on the practice of medicine. The assessment of clinical significance is usually based on the magnitude of the effect observed, the quality of the study that yielded the data, and the probability that the effect is a true one. Although this operational definition is presented from the physician's perspective, the term could operationally be defined from the patient's perspective. Patients are primarily concerned with results that will lead to an improved quality of life or a lengthening of their life. In addition, clinical significance may be applied to either positive data or efficacy or negative safety data such as for adverse reactions. Synonyms include *clinical importance*, *clinical relevance*, and *clinical meaningfulness*.

**Clinical studies** Class of all scientific approaches to evaluate medical disease preventions, diagnostic techniques, and treatments. Investigational and marketed prescription medicine evaluations plus over-the-counter medicines are included.

**Clinical trials** Subset of those clinical studies that evaluates investigational medicines in phases I, II, and III. Phase IV evaluations of marketed medicines in formal clinical trials using the same or similar types of protocols to those used in phases I and III are also referred to as clinical trials.

**Compliance** (1) Adherence of patients to following medical advice and prescriptions. Primarily applied to taking medicine as directed, but also applies to following advice on diet, exercise, or other aspects of patient's life. (2) Adherence of investigators to following a protocol and related administrative and regulatory responsibilities. (3) Adherence of sponsors to following

regulatory, legal, and other responsibilities and requirements relating to clinical trial.

**Compound** Chemical synthesized or prepared from natural sources that is evaluated for its biological activities in preclinical tests.

**Development of medicines** Term *development* as applied to medicines is used in several different contexts, even within the pharmaceutical industry. This often leads to confusion and misunderstanding. No single definition is preferred, but the particular meaning intended should be made clear by all people using the term. Three operational definitions are presented, from the broadest to the narrowest:

1. All stages and processes involved in discovering, evaluating, and formulating a new medicine until it reaches the market (i.e., commercial sale).
2. All stages involving the evaluation and formulation of a new medicine (after the medicine has been discovered and has gone through preclinical testing) until it reaches the market.
3. Those stages after the preclinical discovery and evaluation that involve technical development. These processes include formulation work, stability testing, scaling up the compound for larger scale synthesis, and providing analytical support. Clinical trials are not included in this definition.

**Disease** Disorders (e.g., anxiety disorders, seizure disorders), conditions (e.g., obesity, menopause), syndromes, specific illnesses, and other medical problems that are an acquired morbid change in a tissue, organ, or organism. Synonyms are *illness* and *sickness*.

**Dosage regimen** (1) Number of doses per given time period (usually days), (2) time that elapses between doses (e.g., dose to be given every 6 h) or the time that the doses are to be given (e.g., dose to be given at 8 AM, noon, and 4 PM each day), or (3) quantity of a medicine (e.g., numbe of tablets, capsules) that are given at each specific time of dosing.

**Efficacy** Relative concept referring to the ability of a medicine to elicit a beneficial clinical effect. This may be measured or evaluated using objective or subjective parameters and in terms ranging from global impressions to highly precise measurements. Efficacy is assessed at one or more levels of organization (e.g., subcellular, cellular, tissue, organ, whole body) and may be extrapolated to other levels.

**Endpoint** Indicator measured in a patient or biological sample to assess safety, efficacy, or another trial objective. Some endpoints are derived from primary endpoints (e.g., cardiac output is derived from stroke volume and heart rate). Synonyms include outcome, variable, parameter, marker, and measure. See surrogate endpoint in the text. Also defined as the final trial objective by some authors.

**Incidence rate** Rate of occurrence of new cases of a disease, adverse reaction, or other event in a given population at risk (e.g., the incidence of disease $X$ is $Y$ patients per year per 100,000 population).

**Interpretation** Process whereby one determines the clinical meaning or significance of data after the relevant statistical analyses have been performed. These processes often involve developing an explanation of the data that are being evaluated.

**Medicine** When a compound or substance is tested for biological and clinical activity in humans, it is considered to be a medicine. Some individuals prefer to define a medicine as a compound that has demonstrated clinically useful properties in patients. This definition, however, would restrict the term to use sometime during or after phase II. Others use the term loosely and apply it to compounds with biological properties during the preclinical period that suggest medical usefulness in humans. The author has adopted the first definition for use in this book.

**Patient** Used almost exclusively throughout this book in preference to *subject* or *volunteer*. Patient is used to cover those cases in which the term *volunteer* would be appropriate.

**Pharmacodynamics** Processes of responses resulting from treatment with a medicine or compound. The processes include pharmacological, biochemical, physiological, and therapeutic effects. The pharmacodynamics of a response to treatment are presented with the scientific and/or clinical language of the disciplines involved in detecting, measuring, and describing the effects.

**Pharmacokinetics** Processes of absorption, distribution, metabolism, and excretion of compounds and medicines.

**Phases of clinical trials and medicine development** Four phases of clinical trials and medicine development exist and are defined below. Each of these definitions is a functional one and the terms are not defined on a strict chronological bases. An investigational medicine is often evaluated in two or more phases simultaneously in different clinical trials. Also, some clinical trials may overlap two different phases.

**Phase I** Initial safety trials on a new medicine, usually conducted in normal volunteers. An attempt is made to establish the dose range tolerated by volunteers for single and for multiple doses. Phase I trials are sometimes conducted in severely ill patients (e.g., in the field of cancer) or in less ill patients when pharmacokinetic issues are addressed (e.g., metabolism of a new antiepileptic medicine in stable epileptic patients whose microsomal liver enzymes have been induced by other antiepileptic medicines). Pharmacokinetic trials are usually considered phase I trials regardless of when they are conducted during a medicine's development.

**Phase IIa** Pilot clinical trials to evaluate efficacy (and safety) in selected populations of patients with the disease or condition to be treated, diagnosed, or prevented. Objectives may focus on dose–response, type of patient, frequency of dosing, or numerous other characteristics of safety and efficacy.

**Phase IIb** Well-controlled trials to evaluate efficacy (and safety) in patients with the disease or condition to be treated, diagnosed, or prevented. These clinical trials usually represent the most rigorous demonstration of a medicine's efficacy. Sometimes referred to as pivotal trials.

**Phase IIIa** Trials conducted after efficacy of the medicine is demonstrated but prior to regulatory submission of a new drug application (NDA) or other dossier. These clinical trials are conducted in patient populations for which the medicine is eventually intended. Phase IIIa clinical trials generate additional data on both safety and efficacy in relatively large numbers of patients in both controlled and uncontrolled trials. Clinical trials are also conducted in special groups of patients (e.g., renal failure patients) or under special conditions dictated by the nature of the medicine and disease. These trials often provide much of the information needed for the packaging insert and labeling of the medicine.

**Phase IIIb** Clinical trials conducted after regulatory submission of an NDA or other dossier but prior to the medicine's approval and launch. These trials may supplement earlier trials, complete earlier trials, or be directed toward new types of trials (e.g., quality of life, marketing) or phase IV evaluations. This is the period between submission and approval of a regulatory dossier for marketing authorization.

**Phase IV** Studies or trials conducted after a medicine is marketed to provide additional details about the medicine's efficacy or safety profile. Different formulations, dosages, durations of treatment, medicine interactions, and other medicine comparisons may be evaluated. New age groups, races, and other types of patients can be studied. Detection and definition of previously unknown or inadequately quantified adverse reactions and related risk factors are an important aspect of many phase IV studies. If a marketed medicine is to be evaluated for another (i.e., new) indication, then those clinical trials are considered phase II clinical trials. The term *postmarketing surveillance* is frequently used to describe those clinical studies in phase IV (i.e., the period following marketing) that are primarily observational or nonexperimental in nature to distinguish them from well-controlled phase IV clinical trials or marketing studies.

**Prevalence** Total number of people in a population that are effected with a particular disease at a given time. This term is expressed as the rate of all cases (e.g., the prevalence of disease $X$ is $Y$ patients per 100,000 population) at a given point or period of time.

**Research (on medicines)** Numerous definitions of research are used both in the literature and among scientists. In the broadest sense, research in the pharmaceutical industry includes all processes of medicine discovery, preclinical and clinical evaluation, and technical development. In a more restricted sense, research concentrates on the preclinical discovery phase, where the basic characteristics of a new medicine are determined. Once a decision is reached to study the medicine in humans to evaluate its

therapeutic potential, the compound passes from the research to the development phase.

**Research and development** When research and development are used together, it refers to the broadest definition for research (see above). Some people use the term research colloquially to include most or all of the scientific and medical areas (discovery, evaluation, and development) covered by the single term research and development. *Medicine development* has several definitions and, in its broadest definition, is exactly the same as the broad definition of research.

**Risk** Measure of (1) the probability of occurrence of harm to human health or (2) the severity of harm that may occur. Such a measure includes judgment of the acceptability of risk. Assessment of safety involves judgment, and there are numerous perspectives (e.g., patients, physicians, company, regulatory authorities) used for judging it.

**Safety** Relative concept referring to the freedom from harm or damage resulting from adverse reactions or physical, psychological, or behavioral abnormalities that occur as a result of medicine or nonmedicine use. Safety is usually measured with one or more of the following: physical examination (e.g., vital signs, neurological, ophthalmological, general physical), laboratory evaluations of biological samples (e.g., hematology, clinical chemistry, urinalysis), special tests and procedures (e.g., electrocardiogram, pulmonary function tests), psychiatric tests and evaluations, and determination of clinical signs and symptoms.

**Serious adverse reactions** Multiple definitions are possible and no single one is correct in all situations. In general usage referring to patients in clinical trials, a serious adverse reaction may be (1) any bad adverse reaction that is observed, (2) any bad adverse reaction that one does not expect to observe, (3) any bad adverse reaction that one does not expect to observe and is not in the label, or (4) any bad adverse reaction that has not been reported with standard therapy. Definitions also may be based on the degree to which an adverse reaction compromises a patient's function or requires treatment.

**Side effect** Any effect other than the primary intended effect(s) resulting from medicine or nonmedicine treatment or intervention. Side effects may be negative (i.e., an adverse reaction), neutral, or positive (i.e., a beneficial effect) for the patient. This term therefore includes all adverse reactions plus other effects of treatment. See *adverse reaction*.

**Site** Place where a clinical trial is conducted. A physician who has offices and sees patients in three separate locations is viewed as having one site. A physician who is on the staff of four hospitals could be viewed as having one or four sites, depending on how similar or different the patient populations are and whether the data from these four locations will be pooled and considered a single site. For example, a single physician who enrolls groups of patients at a university hospital, private clinic, community hospital, and

Veterans Administration Hospital should generally be viewed as having four sites, since the patient populations would be expected to differ at each site.

**Statistical significance** Probability that an event or difference occurred by chance alone. Thus, it is a measure of whether a difference is likely to be real, but it does not indicate whether the difference is small or large, important or trivial. The level of statistical significance depends on the number of patients studied or observations made as well as the magnitude of differences observed.

**Therapeutic window** Difference between the minimum and maximum doses that may be given patients to obtain an adequate clinical response and avoid intolerable toxic effects. The greater the value calculated for the therapeutic window, the greater a medicine's margin of safety. Synonyms are *therapeutic ratio* and *therapeutic index*.

**Volunteer** Normal individual who participates in a clinical trial for reasons other than medical need and who does not receive any direct medical benefit from participating in the trial.

# *Common Vehicles for Nonclinical Evaluation of Therapeutic Agents*

**TABLE 1 Acacia**

| | Route | Duration | Dose | Comments |
|---|---|---|---|---|
| Rat | *Oral* | 30 days | 500 mg/kg | Well tolerated |
| | | 90 days | 10 mL/kg | As 20% of formulation; well tolerated |
| Primate | *Oral* | 90 days | 100 mg/kg | Well tolerated, but with some reduction in food intake |

**TABLE 2 Acetate Sodium**

| | Route | Duration | Dose | Comments |
|---|---|---|---|---|
| Rat | *Intravenous* | 1 month | 1 mL/kg | Well tolerated as 5 mM solution in saline |

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

**TABLE 3 Acetic Acid**

| | Route | Duration | Dose | Comments |
|---|---|---|---|---|
| Rat | *Oral* | 90 days | 5 mL/kg | Well tolerated (gavage), 3% solution |
| | | 1 month | 10 mL/kg | Well tolerated, 20% solution |
| | *Intravenous* | 1 month | As pH buffer | Well tolerated |
| Mouse | *Oral* | 90 days | 5 mL/kg | Well tolerated (gavage), 3% solution |

**TABLE 4 Acetone (2-Propanone)**

| | Route | Duration | Dose | Comments |
|---|---|---|---|---|
| Rat | *Oral* | 2 weeks | 5 mL/kg | Higher doses cause acidosis; transitory neurobehavorial effects at this dose |
| | *Dermal* | 30 days | 5 mL/kg | Well tolerated |
| Mouse | *Oral* | 2 weeks | 3 mL/kg | Higher doses cause acidosis; transitory neurobehavorial effects at this dose |
| | *Dermal* | 2 years | 0.5 mL | Well tolerated |
| Guinea pig | *Dermal* | 1 month | 1 mL | Well tolerated |
| Rabbit | *Dermal* | 90 days | 1 mL | Defatting of application site |

**TABLE 5 Alginic Acid**

| | Route | Duration | Dose | Comments |
|---|---|---|---|---|
| Rat | *Intraperitoneal* | 1 month | 100 mg/kg | Well tolerated |

**TABLE 6 Anecortave Acetate**

| | Route | Duration | Dose | Comments |
|---|---|---|---|---|
| Rat | *Subcutaneous* | 4 doses | 2 mL/kg | Well tolerated |

**TABLE 7 Benzoic Acid**

| | Route | Duration | Dose | Comments |
|---|---|---|---|---|
| Rat | *Oral* | N/A | 100 mg | Well tolerated |

**TABLE 8 β-Cyclodextrin**

| | Route | Duration | Dose | Comments |
|---|---|---|---|---|
| Rat | *Oral* | 12 months | 500 g/kg | Hepatitis, nephrosis, acute tubular necrosis at dose levels above 20 g/kg |
| | *Intravenous* | — | — | Tubular hypertrophy at doses above 100 mg/kg/day at 3 months or longer |
| Primate | *Oral* | 12 months | — | Tubular hypertrophy at doses above 100 mg/kg/day at 3 months or longer |

**TABLE 9 Canola Oil**

| | Route | Duration | Dose | Comments |
|---|---|---|---|---|
| Dog | *Oral* | 1 month | 2 mL/kg | Well tolerated |

**TABLE 10 Capryol 90**

| | Route | Duration | Dose | Comments |
|---|---|---|---|---|
| Dog | *Oral* | 28 days | 1000 mg/kg | Well tolerated |
| | | | 2500 mg/kg | Well tolerated |
| Rat | *Oral* | Acute | | Well tolerated; $LD_{50} > 5$ g/kg |
| | | 28 days | 500, 1500, 2500 mg/kg | NOAEL of 2500 mg/kg |
| | | 7 days | 300, 1000, 2500 mg/kg | Well tolerated |
| Rabbit | *Cutaneous* | Acute | No dilution | Mildly irritant |
| | *Ocular* | Acute | No dilution | Moderately irritant |

**TABLE 11 Captisol**

| | Route | Duration | Dose | Comments |
|---|---|---|---|---|
| Rat | *Oral* | 1 month | 10 mL/kg | 12% solution, well tolerated |
| | *Intravenous* | 1 month | 4 mL/kg | 12% solution, well tolerated |
| Primate | *Oral* | 9 months | 1 g/kg | 10% solution, well tolerated |
| | *Subcutaneous* | 12 months with 3 weekly administrations | 120 mg/kg | Well tolerated |
| Mouse | *Oral* | 1 month | 500 mg/kg | 10% solution, well tolerated |
| | *Subcutaneous* | 6 months | 1200 mg/kg | NOAEL |
| | | 90 days | 1200 mg/kg | NOEL |

**TABLE 12 Carboxymethyl Cellulose (CMC)**

| | Route | Duration | Dose | Comments |
|---|---|---|---|---|
| Primate | *Oral* | 30 days | 5% in water | Well tolerated |
| | *Subcutaneous* | Acute | 10 mL/kg | Well tolerated |
| Rat | *Oral* | 1 year | 5% in water | Well tolerated |

**TABLE 13 Carboxymethyl Cellulose Calcium**

| | Route | Duration | Dose | Comments |
|---|---|---|---|---|
| Dog | *Oral* | 90 days | 1 mL/kg | Well tolerated; 1% solution |

**TABLE 14 Carboxymethyl Cellulose Sodium**

| | Route | Duration | Dose | Comments |
|---|---|---|---|---|
| Rabbit | *Oral* | 1 month | 0.5 mL/kg | Well tolerated; 1% solution |

**TABLE 15 Cetyl Alcohol**

| | Route | Duration | Dose | Comments |
|---|---|---|---|---|
| Mouse | *Intraperitoneal* | 1 month | 100 mg/kg | Well tolerated |

**TABLE 16 Citrate Buffer**

| | Route | Duration | Dose | Comments |
|---|---|---|---|---|
| Dog | *Intravenous* | 8 doses | 30 mL/kg/day | Well tolerated |
| | *Subcutaneous* | 30 days | 10 mL/kg/day | Well tolerated |
| Rat | *Oral* | 2 weeks | 15 mL/kg | Well tolerated (50 mM) |
| | | | 10 mL/kg | Well tolerated (50 mM) |

**TABLE 17 Citric Acid Buffer**

| | Route | Duration | Dose | Comments |
|---|---|---|---|---|
| Rat | *Oral* | 2 weeks | 15 mL/kg | Well tolerated (50 mM) |
| | | | 10 mL/kg | Well tolerated (50 mM) |

**TABLE 18 Collagen Matrix**

| | Route | Duration | Dose | Comments |
|---|---|---|---|---|
| Primate | *Implantation in humerus bone* | 6 months | Two strips/site (humerus right and left) | Well tolerated |
| Rabbit | *Implantation* | 6 months | Single application, 5 mL/kg | Well tolerated |

**TABLE 19 Corn Oil**

| | Route | Duration | Dose | Comments |
|---|---|---|---|---|
| Dog | *Oral* | 1 month | 3.0 mL/kg | Well tolerated |
| Rat | *Oral* | 20 doses | 5 mL/kg | Well tolerated |
| | | 1 dose | 10 mL/kg | Well tolerated |
| Mouse | *Oral* | 1 month | 2.5 mL/kg | Well tolerated |
| Rabbit | *Oral* | 1 month | 1 mL/kg | Well tolerated |
| Chick embryo | *Oral* | | | |
| | *Injection into egg* | Once | 0.1 μL/g | Less mortality than 1.0 μL/g egg |
| | | | 1 μL/g | Increase in mortality, decreased activity during righting reflex, running time, visual discrimination, and olfactory aversion test |

**TABLE 20 Cremophore EL**

| | Route | Duration | Dose | Comments |
|---|---|---|---|---|
| Dog | *Intravenous* | 1 month | 2 mL/kg | Well tolerated |
| Rat | *Oral* | 1 month | 100 mg/kg | Well tolerated |

**TABLE 21 Cyclohexane**

| | Route | Duration | Dose | Comments |
|---|---|---|---|---|
| Rat | *Oral* | 4 weeks | 5 mL/kg/day | Clinical signs: intermittent convulsive after dosing, piloerection round back and emaciated appearance |
| | *Dermal* | 30 days | 1 mL/kg/day | Well tolerated |
| Rabbit | *Oral* | 30 days | 0.5 mL/kg/day | Well tolerated |

**TABLE 22 D-Glucose anhydrous 30%–PEG 70% (V/V)**

| | Route | Duration | Dose | Comments |
|---|---|---|---|---|
| Dog | *Oral* | 2 weeks | 0.32 mL/kg | Well tolerated |
| | *Intravenous* | 2 weeks | Bolus 0.24–0.33 mL/kg infusion, 0.08–0.11 mL/kg/h | Well tolerated |
| Rat | *Intravenous* | 3 weeks | Bolus 0.8–1.07 mL/kg infusion, 0.266–0.356 mL/kg intravenous injection (into tail vein) followed by an intravenous injection for 6 h | Well tolerated |

**TABLE 23 Dextrose (0.5%)**

| | Route | Duration | Dose | Comments |
|---|---|---|---|---|
| Dog | *Intravenous* | 90 days | 150 mL/h | Well tolerated |
| Rat | *Intravenous* | 1 dose | 1.4 mL per animal | Well tolerated |

**TABLE 24 Diethyleneglycol-Monoethylether**

| | Route | Duration | Dose | Comments |
|---|---|---|---|---|
| Primate | *Intravenous* | 1 month | 0.355 mL/kg | Well tolerated |

**TABLE 25 Dimethylsulfoxide**

| | Route | Duration | Dose | Comments |
|---|---|---|---|---|
| Dog | *Intravenous* | 1 month | 1.25 mL/kg | Well tolerated |
| Rat | *Oral* | 7 days | 5 mL/kg | Well tolerated |
| | | 4 weeks | 5 mL/kg | Well tolerated |
| | *Intravenous* | 1 month | 200 mg/kg | Well tolerated |
| | *Intraperitoneal* | 1 month | 5 mL/kg | Well tolerated |
| Guinea pig | *Intravenous* | 1 month | 0.1 mL/kg | Well tolerated |
| Primate | *Oral* | Efficacity | 3 mL/kg/day | Well tolerated |
| Mouse | *Oral* | 4 weeks | 5 mL/kg | Well tolerated |
| | *Intraperitoneal* | 1 month | 100 mg/kg | Well tolerated |
| | | 3 days | 10 mL/kg | Well tolerated |
| Rabbit | *Subcutaneous* | 1 month | 1 mL/kg | Erythema, inflammation |

**TABLE 26 Dulbecco's Modified PBS**

| | Route | Duration | Dose | Comments |
|---|---|---|---|---|
| Rat | *Oral* | 4 weeks | 0.1, 0.8 and 1.2 mg/kg | Well tolerated |
| | *Intravenous* | 1 month | 1 mL/kg/day | Well tolerated |

**TABLE 27 Ethanol**

| | Route | Duration | Dose | Comments |
|---|---|---|---|---|
| Dog | *Oral* | 6 months | 400 mL/kg | Hepatopathy, myopathy, CNS changes |
| | | 90 days | 5 mL/kg | 5% solution; well tolerated |
| | | 1 month | 5 mL/kg | 7.5% solution; well tolerated |
| | *Intravenous* | Once | 1 mL/kg | 30% solution; CNS depression, ataxia |
| Rat | *Oral* | | 5 mL/kg | Depression |
| | | 1 month | 175 g/kg | Depression, decreased RBCs |
| | | 12 months | 1000 mg/kg | Fatty liver |
| | | 7 days | 10 mL/kg | 10% solution; well tolerated |
| | | 4 weeks | 2 mL/kg | 70% solution; hypokinesia, dyspnea, regurgitation, distended lungs/ileum, and swollen abdomen |
| | | 90 days | 8 mL/kg | 10% solution; well tolerated |
| | *Intravenous* | 12 months | 250 g/kg | Nephrosis, ATN, bladder changes, weight loss |
| Primate | *Oral* | 9 months | 250 mg/kg | Behavioral changes |
| Mouse | *Oral* | 6 months | 2500 mg/kg | Well tolerated |
| | | 1 month | 2.5 mL/kg | 5% solution; well tolerated |
| | *Intraperitoneal* | Acute | 5 mL/kg | 5% solution; well tolerated |
| | *Cutaneous* | 13 weeks | 100 μL/animal/day | 70% solution; well tolerated |

**TABLE 28 Gelucire 44/14**

| | Route | Duration | Dose | Comments |
|---|---|---|---|---|
| Rabbit | *Cutaneous* | Acute | 0.5 mL | Not irritant |
| | *Ocular* | Acute | 0.1 mL | Slight irritant |
| Rat | *Oral* | 28 days | 600, 1500, 2400 mg/kg/day | NOEL: 2400 mg/kg/day |
| | | 7 days | 600, 1500, 2400 mg/kg/day | NOEL: 2400 mg/kg/day |
| | | Acute | No dilution | $LD_{50}$: >2004 mg/kg/day |
| Dog | *Oral* | 3 months | 400, 1000, 2500 mg/kg/day | NOAEL: >2500 mg/kg/day |
| | | 14 days | 400, 1000, 2500 mg/kg/day | |

**TABLE 29 Glucose**

| | Route | Duration[a] | Dose | Comments |
|---|---|---|---|---|
| Dog | *Oral* | ADME | 2/10 mL/kg/day | 5% solution; well tolerated |
| Rat | *Oral* | 26 weeks | 0.71–8.6 mL/kg | 10% solution; well tolerated |
| | | Prelim | 5 mL/kg | 5% solution; well tolerated |
| | *Subcutaneous* | 2 weeks | 0.75 mL/kg | 5% solution; well tolerated |
| Primate | *Oral* | 13 weeks | 0.78–9.3 mL/kg | 10% solution; well tolerated |
| | | ADME card. vas. | 5 mL/kg | 5% solution; well tolerated |

[a]ADME: Absorption, distribution, metabolism, and excretion.

**TABLE 30 Glycerol**

| | Route | Duration | Dose | Comments |
|---|---|---|---|---|
| Rat | *Oral* | | 1000 mg/kg | Well tolerated |
| | | 1 month | 15 g/kg | Reduced adrenal weights |
| | | 1 month | 1000 mg/kg | Well tolerated |
| | *Subcutaneous* | | 10 mg/kg | Well tolerated |
| Guinea pig | *Oral* | 1 month | 500 mg/kg | Well tolerated |
| Mouse | *Oral* | 90 days | 500 mg/kg | Depression and reduced respiration |
| | *Intravenous* | 1 month | 100 mg/kg | Well tolerated |
| | *Subcutaneous* | Acute | 10 mg/kg | Well tolerated |
| | *Intraperitoneal* | 1 month | 250 mg/kg | Well tolerated |
| Rabbit | *Intravenous* | | 10 mg/kg | Well tolerated |

**TABLE 31 Gum Tragacanth**

| | Route | Duration | Dose | Comments |
|---|---|---|---|---|
| Mouse | *Oral* | 2 weeks | 10 mL/kg | 0.5% solution; Well tolerated |

**TABLE 32 Hydroxypropyl Betacyclodextrin**

| | Route | Duration | Dose | Comments |
|---|---|---|---|---|
| Dog | *Intravenous* | 1 month | 10 mL/kg | 40% solution; well tolerated |
| Rat | *Intravenous* | 1 month | 10 mL/kg | 40% solution; well tolerated |

**TABLE 33 Hydroxypropyl Cellulose**

| | Route | Duration | Dose | Comments |
|---|---|---|---|---|
| Rat | *Oral* | 90 days | 1000 mg/kg | Well tolerated |

**TABLE 34 Hydroxypropyl Methylcellulose**

| | Route | Duration | Dose | Comments |
|---|---|---|---|---|
| Dog | *Intraperitoneal* | 28 days | 200 mg/kg | Well tolerated |
| Rat | *Oral* | One dose | 10 mL/kg | 0.2%; well tolerated |
| | | One dose | 10 mL/kg | 0.5%; well tolerated |
| | | | 10 mL/kg | 0.5%; well tolerated |
| | *Intraperitoneal* | One dose | 5 mL/kg | 0.5%; well tolerated |
| Mouse | *Oral* | 10 doses | 10 mL/kg | 0.2%; well tolerated |
| | | One dose | 10 mL/kg | 0.5%; well tolerated |
| | *Intraperitoneal* | Acute | 50 mg/kg | Well tolerated |
| | | Acute | 5 mL/kg | 0.5%; well tolerated |

**TABLE 35 Isopropyl Alcohol**

| | Route | Duration | Dose | Comments |
|---|---|---|---|---|
| Rabbit | *Dermal* | 1 month | 1000 mg/kg | Well tolerated |

**TABLE 36 Isopropyl Myristate**

| | Route | Duration | Dose | Comments |
|---|---|---|---|---|
| Rabbit | *Dermal* | 1 month | 500 mg/kg | Well tolerated |

**TABLE 37 Labrafil MI944**

| | Route | Duration | Dose | Comments |
|---|---|---|---|---|
| Dog | *Oral* | 1 month | 2 mg/kg | Well tolerated |

**TABLE 38 Labrasol**

| | Route | Duration | Dose | Comments |
|---|---|---|---|---|
| Rat | *Oral* | Acute | 20, 22.4, 25.1, 28.21, and 31.60 g/kg | $LD_{50}$ = 22 g/kg; nontoxic |
| | | ADME | 10 and 150 mg/kg/day | |
| | | Segment II: embryofetal development | 1000, 2000, or 3000 mg/kg/day | NOEL: 3000 mg/kg/day with no indication of teratogenicity |
| | | 14 days | 100, 300, 1000 and 3000 mg/kg/day | NOAEL: 3000 mg/kg/day |
| | | 6 months | 300, 1000 and 3000 mg/kg/day | NOEL: 300 mg/kg/day; NOAEL: 3000 mg/kg/day |
| | *Intravenous* | 28 days | 10 mg/kg/day | |
| | *Cutaneous* | Patch test | 0.02 mL/animal | Well tolerated |
| | | Acute | | Very well tolerated |
| | *Ocular* | | | Slight irritant |
| Dog | *Oral* | 13 weeks | 0, 300, 1000, and 3000 mg/kg/day | NOEL: 1000 mg/kg/day; NOAEL: 3000 mg/kg/day |
| | | 14 days | 100, 300, 1000, and 3000 mg/kg/day | In high-dose group, moderate suppurative inflammation of lungs; no adverse effects on survival and clinical observations |
| | | 3 months | 300, 1000 and 3000 mg/kg/day | NOEL: 1000 mg/kg/day; NOAEL: 3000 mg/kg/day |
| Rabbit | *Cutaneous* | Patch test | 0.5 mL | Well tolerated |
| | *Ocular* | Acute | 0.1 mL | Slight irritant |

**TABLE 39 Lactose**

| | Route | Duration | Dose | Comments |
|---|---|---|---|---|
| Primate | *Inhalation* | 2 weeks | 1 L/min/animal | Well tolerated |

**TABLE 40 Lanolin**

| | Route | Duration | Dose | Comments |
|---|---|---|---|---|
| Rabbit | *Dermal* | 90 days | 1000 mg/kg | Well tolerated |

**TABLE 41 L-Ascorbic Acid**

| | Route | Duration | Dose | Comments |
|---|---|---|---|---|
| Rat | *Oral* | 90 days | 500 mg/kg | Hematological changes, weight loss |

**TABLE 42 Lauroglycol**

| | Route | Duration | Dose | Comments |
|---|---|---|---|---|
| Rabbit | *Cutaneous* | Acute | No dilution | Moderately irritant |
| | *Ocular* | Acute | No dilution | Slightly irritant |
| Rat | *Oral* | Acute | | $LD_{50}$: >2003 mg/kg/day |

**TABLE 43 Maltitol Solution**

| | Route | Duration | Dose | Comments |
|---|---|---|---|---|
| Rat | *Intraperitoneal* | 1 month | 500 mg/kg | Well tolerated |

**TABLE 44 Maltol**

| | Route | Duration | Dose | Comments |
|---|---|---|---|---|
| Guinea pig | *Oral* | 1 month | 75 mg/kg | Well tolerated |
| Rabbit | *Oral* | 1 month | 100 mg/kg | Well tolerated |

**TABLE 45 Mannitol**

| | Route | Duration | Dose | Comments |
|---|---|---|---|---|
| Primate | *Oral* | 2 months | 10 mL/kg | Well tolerated |

**TABLE 46 Methyl Cellulose**

| | Route | Duration | Dose | Comments |
|---|---|---|---|---|
| Rat | *Oral* | 1 month | 10 mL/kg | 0.5%; well tolerated |
| | | 1 month | 5 mL/kg | 0.5%; well tolerated |
| | | 14 doses | 10 mL/kg | 1%; well tolerated |
| | | 1 dose | 10 mL/kg | 2%; well tolerated |
| Guinea pig | *Oral* | 12 doses | 4 mL/kg | 0.5%; well tolerated |
| Primate | *Oral* | 1 month | 10 mL/kg | 0.1%; well tolerated |
| | | 1 month | 10 mL/kg | 0.5%; well tolerated |
| | | 2 weeks | 5 mL/kg | 0.5%; well tolerated |
| | | 1 month | 5 mL/kg | 1%; well tolerated |
| Mouse | *Oral* | 90 days | 10 mL/kg | 0.5%; well tolerated |
| Rabbit | *Oral* | 1 month | 4 mL/kg | 0.5%; well tolerated |
| Dog | *Oral* | 2 weeks | 10 mL/kg | 0.5%; well tolerated |

**TABLE 47 Mineral Oil**

| | Route | Duration | Dose | Comments |
|---|---|---|---|---|
| Rat | *Oral* | 1 month | 5 mL/kg | Well tolerated |
| Mouse | *Oral* | 1 month | 250 mg/kg | Well tolerated |
| Dog | *Oral* | 1 month | 2.5 mL/kg | Well tolerated |

**TABLE 48 Phosphate-Buffered Saline**

| | Route | Duration | Dose | Comments |
|---|---|---|---|---|
| Rat | *Oral* | 1 month | 10 mL/kg | Well tolerated |
| | *Intravenous* | 1 dose | 1 mL/kg | Well tolerated |
| | *Subcutaneous* | 1 month | 1 mL/kg | Well tolerated |
| | *Slow bolus* | 11 doses | 1 mL/kg | Well tolerated |
| Mouse | *Subcutaneous* | 6 months | 10 mL/kg | Well tolerated |
| Primate | *Oral* | 2 weeks | 10 mL/kg | Well tolerated |
| | | 2 weeks | 1.6 mL/kg | Well tolerated |
| | *Subcutaneous* | 1 week | 0.2 mL/kg | Well tolerated |
| | | 9 months | 1 mL/kg | Well tolerated |

**TABLE 49 Peanut Oil**

| | Route | Duration | Dose | Comments |
|---|---|---|---|---|
| Rat | *Oral* | 1 month | 10 g/kg | Well tolerated |
| | | 12 months | 10 g/kg | Well tolerated |
| | | 90 days | 10 g/kg | Well tolerated |

**TABLE 50 PEG 300**

| | Route | Duration | Dose | Comments |
|---|---|---|---|---|
| Guinea pig | *Intravenous* | 1 month | 1 mL/kg | Well tolerated |
| Mouse | *Oral* | ADME | 10 mL/kg/day | Well tolerated |
| Rabbit | *Oral* | 1 month | 500 mg/kg | Well tolerated |

**TABLE 51 PEG 400**

| | Route | Duration | Dose | Comments |
|---|---|---|---|---|
| Guinea pig | *Oral* | 1 month | 1000 mg/kg | Well tolerated |
| Mouse | *Oral* | 4 weeks | 10 mL/kg/day | Well tolerated |
| | *Intraperitoneal* | 1 month | 500 mg/kg | Well tolerated |
| | | 3 days | 10 mL/kg | 35%; well tolerated |
| | | 1 month | 2.5 mL/kg | 40%; well tolerated |
| Rat | *Oral* | 10 doses | 1.67 mg/kg | Well tolerated |
| | | 1 dose | 2 mL/kg | Well tolerated |
| | | 1 dose | 5 mL/kg | Well tolerated |
| | | 4 weeks | 5 mL/kg/day | Well tolerated |
| | | 1 month | 5 mL/kg | Well tolerated |
| | *Intravenous* | 1 dose | 0.5 mL/kg | Well tolerated |
| | | 4 weeks | 0.5 mL/kg | Well tolerated |
| | *Intraperitoneal* | 1 month | 5 mL/kg | 35%; well tolerated |
| | *Cutaneous* | 13 weeks | 2.5 mL/kg | Well tolerated |
| | | 104 weeks | 2.5 mL/kg | Well tolerated |
| Minipig | *Cutaneous* | 2 weeks | 2.5 mL/kg | Well tolerated |

**TABLE 52 Petrolatum**

| | Route | Duration | Dose | Comments |
|---|---|---|---|---|
| Rabbit | *Dermal* | 1 month | 1 mg/kg | Well tolerated |

**TABLE 53 Poloxamer**

| | Route | Duration | Dose | Comments |
|---|---|---|---|---|
| Rat | *Oral* | 1 month | 10 mL/kg | 7.5%; well tolerated |
| Mouse | *Oral* | 1 month | 10 mL/kg | Well tolerated |

**TABLE 54 Povidone**

| | Route | Duration | Dose | Comments |
|---|---|---|---|---|
| Rat | *Intramuscular* | 90 days | 1 mL/kg | 1%; well tolerated |

**TABLE 55 Propylene Glycol**

| | Route | Duration | Dose | Comments |
|---|---|---|---|---|
| Rat | *Oral* | 1 month | 2.5 mL/kg | Well tolerated |
| | | 2 weeks | 2 mL/kg | Well tolerated |
| | *Subcutaneous* | 4 weeks | 2.5 mL/kg | Well tolerated |
| Minipig | *Cutaneous* | 26 weeks | 2.5 mL/kg | Well tolerated |
| Mouse | *Oral* | 1 month | 10 mL/kg | 50%; well tolerated |
| | *Intraperitoneal* | 1 month | 2.5 mL/kg | 40%; well tolerated |
| Dog | *Oral* | 1 month | 2.5 mL/kg | Well tolerated |
| | | 13 weeks | 2 mL/kg | Well tolerated |

**TABLE 56 Rameb 7.5%**

| | Route | Duration | Dose | Comments |
|---|---|---|---|---|
| Primate | *Intranasal* | 1 month, 3 doses per day | 82.8 mg/mL (with treatment), 74.7 mg/mL (as placebo) | Well tolerated |

**TABLE 57 Sesame Oil**

| | Route | Duration | Dose | Comments |
|---|---|---|---|---|
| Rat | *Oral* | 1 month | 1 mL/kg | Well tolerated |
| Mouse | *Oral* | 1 month | 0.25 mL/kg | Well tolerated |
| Rabbit | *Oral* | 1 month | 0.5 mL/kg | Well tolerated |
| Dog | *Oral* | 1 month | 1 mL/kg | Well tolerated |

**TABLE 58 Sodium Acetate Trihydrate Buffer**

| | Route | Duration | Dose | Comments |
|---|---|---|---|---|
| Primate | *Intravenous* | 2 weeks | 1 mL/kg | Well tolerated |

**TABLE 59 Sodium Chloride 0.9%**

| | Route | Duration | Dose | Comments |
|---|---|---|---|---|
| Rat | *Intravenous* | 7 doses | 1 mL/kg | Well tolerated |
| | | 1 dose | 1 mL/kg | Well tolerated |
| | | 14 doses | 10 mL/kg | Well tolerated |
| | | 3 doses | 4 mL/kg | Well tolerated |
| | *Subcutaneous* | 1 dose | 0.1–0.4 mL | Well tolerated |
| | | 1 month | 4 mL/kg | Well tolerated |
| | | 56 doses | 2 mL/kg | Well tolerated |
| | *Slow bolus* | 1 dose | 1 mL/kg | Well tolerated |
| | | 1 dose | 10 mL/kg | Well tolerated |
| Mouse | *Intravenous* | 1 dose | 10 mL/kg | Well tolerated |
| | *Subcutaneous* | 1 dose | 10 mL/kg | Well tolerated |
| Rabbit | *Intravenous* | 1 dose | 0.1 mL/kg | Well tolerated |
| | *Intraperitoneal* | 1 dose | 0.1 mL/kg | Well tolerated |
| Dog | *Oral* | 1 dose | 0.282 mL/kg | Well tolerated |
| | *Intravenous* | 1 dose | 10 mL/kg | Well tolerated |
| | | 1 dose | 2 mL/kg | Well tolerated |
| | | 2 weeks | 5 mL/kg | Well tolerated |
| | *Subcutaneous* | 1 month | 0.025 mL | Well tolerated |
| | *Slow bolus* | 1 dose | 0.3 mL/kg | Well tolerated |
| Primate | *Subcutaneous* | 1 month | 0.67 mL/kg | Well tolerated |
| | | 56 doses | 0.5 mL/kg | Well tolerated |

**TABLE 60 Sodium Phosphate**

| | Route | Duration | Dose | Comments |
|---|---|---|---|---|
| Dog | *Oral* | 2 weeks | 10 mL/kg | 70 mM; well tolerated |
| Rat | *Oral* | 2 weeks | 10 mL/kg | 70 mM; well tolerated |

**TABLE 61 Tartaric Acid**

| | Route | Duration | Dose | Comments |
|---|---|---|---|---|
| Rat | *Oral* | 39 weeks | 0.5 mL/kg | Well tolerated |
| | | 2 weeks | 3 mL/kg | Well tolerated |
| Rabbit | *Oral* | Preliminary segment II | 3 mL/kg/day | Well tolerated |
| | | Segment II | 3 mL/kg/day | Well tolerated |

**TABLE 62 Transcutol**

| | Route | Duration | Dose | Comments |
|---|---|---|---|---|
| Cat | *Intravenous* | 1 month | 2 mL/kg | Well tolerated |
| Rabbit | *Dermal* | Skin irritation | 0.5 mL over 2-$cm^2$ area | 50%; nonirritant |
| | | 28 days | 0, 300, 1000, 3000 mg/kg/day | Undiluted; NOEL >1000 mg/kg/day |
| | *Ocular* | Eye irritation | 0.1 mL | 30%; slight irritation |
| | | Eye irritation | 0.1 mL | Undiluted; slight irritation |
| Rat | *Oral* | 90 days | 0, 0.25, 1, and 5% | NOEL is 1% |
| | | Acute | 5.0 g/kg | $LD_{50}$ > 5000 mg/kg |
| | | Fertility and embryotoxicity range-finding study | 500, 1000, 2000, 4000 mg/kg/day | NOEL > 500 mg/kg/day |
| Mouse | *Oral* | Acute | — | 6.6 g/kg tested toxic |
| | *Oral* | Chronic (12 months) | — | NOEL: 850–1000 mg/kg |
| Dog | *Oral* | 90 days | — | NOAEL: 1000 mg/kg |

**TABLE 63 Trisodium Citrate Dihydrate**

| | Route | Duration | Dose | Comments |
|---|---|---|---|---|
| Dog | *Oral* | 52 weeks | 10 mL/kg/day | Well tolerated |
| Rat | *Oral* | Segment III | 10 mL/kg/day | Well tolerated |
| | | 39 weeks | 10 mL/kg/day | Well tolerated |
| | | 4 weeks | 10 mL/kg/day | Well tolerated |
| Mouse | *Intravenous* | 13 weeks | 10 mL/kg/day | Well tolerated |

**TABLE 64 Tween 20**

| | Route | Duration | Dose | Comments |
|---|---|---|---|---|
| Rat | *Oral* | 1 month | 250 mg/kg | Well tolerated |
| | | 90 days | 500 mg/kg | Diarrhea |
| Mouse | *Oral* | 1 month | 10 mg/kg | Well tolerated |

**TABLE 65 Tween 80**

| | Route | Duration | Dose | Comments |
|---|---|---|---|---|
| Dog | *Oral* | 90 days | 5 mL/kg | As 1% of formulation; well tolerated |
| Rat | *Oral* | | 350 mg/kg | Well tolerated |
| | | 4 weeks | 5 mL/kg | 1%; well tolerated |
| | | 7 days | 10 mL/kg | 1%; well tolerated |
| | *Intravenous* | | 100 mg/kg | Well tolerated |
| Mouse | *Intraperitoneal* | 1 month | 10 mL/kg | 2%; well tolerated |
| | *Intranasal* | 3 days | 10 μL/nostril | 2%; well tolerated |
| Primate | *Oral* | 90 days | 5 mL/kg | 1%; well tolerated |

**TABLE 66 Xylitol**

| | Route | Duration | Dose | Comments |
|---|---|---|---|---|
| Primate | *Intranasal* | 1 month | 1200 μL/day for control and high concentration; 200 and 400 μL/day for respectively low and intermediate dose level | Well tolerated |

**TABLE 67 Summary of Vehicle Information**

| Excipient/Vehicle | Data in Table Number | CAS Number | Chemical Name | Formula | Key Toxicity Review Articles[a] | Animal Species Evaluated |
|---|---|---|---|---|---|---|
| 2-hydroxypropyl-β-cyclodextrin | NA | 128446-35-5 | | | Gould, S. (2005) | Rat, primate, mouse, rabbit, dog |
| Acacia | 3 | 9000-01-5 | Acaciae gummi | | TOXSYS; Anderson (1986); Bachmann, E. (1978) | Rat, primate |
| Acetate, sodium | 4 | 127-09-3 | Acetic acid sodium salt | $C_2H_3NaO_2$ | TOXSYS | Rat |
| Acetic acid | 5 | 64-19-7 | Ethanolic acid | $C_2H_4O_2$ | Cragg, S.T. (2001); Schonwald, S. (2004), "Irrigating Solutions" | Rat, mouse |
| Acetone | 6 | 67-64-1 | 2-Propanone | $CH_3COCH_3$ | Meditext, 2006 | Rat, mouse, guinea pig, rabbit |
| Acetylmethylamine in water | NA | 79-16-3 | N-Methylacetamide | $C_3H_7NO$ | NA | |
| Alginic acid | 7 | 9005-32-7 | Norgine | $(C_6H_3O_6)_n$ | JECFA, 49th (1997) | Rat |
| Anecortave acetate | 8 | 7753-60-8 | | $C_{23}H_{30}O_5$ | Jockovich (2006); Talsma (2004) | Rat |
| Benzoic acid | 9 | 65-85-0 | Benzoic acid | $C_7H_6O_2$ | TOXSYS; Nair, B. (2001); David, R. M., et al. (2001) | Rat |
| β-Cyclodextrin | 10 | 7585-39-9 | β-Dextrin | $C_{42}H_{70}O_{35}$ | TOXSYS; Toyoda, K. (1997); Waner, T. (1995); Marttin, E. (1998); Challa, R. (2005) | Rat, primate |
| BHT | NA | 128-37-0 | Butylated hydroxytoluene | $C_{15}H_{24}O$ | Lanigan, R. S. (2002); Nakagawa, Y. (1984); Briggs, D. (1989) | |

**TABLE 67** ***Continued***

| Excipient/Vehicle | Data in Table Number | CAS Number | Chemical Name | Formula | Key Toxicity Review Articles[a] | Animal Species Evaluated |
|---|---|---|---|---|---|---|
| Canola oil | 11 | 120962-03-0 | Canbra oil | | Evangelista, C. M. (2004) | Dog |
| Capryol 90 | 12 | 31565-12-5 | Propylene glycol monocaprylate | $C_{11}H_{22}O_3$ | Li (2005); Cho (2004) | Rat, dog, rabbit |
| Captisol | 13 | 182410-00-0 | β-Cyclodextrin sulfobutyl ether, sodium salt (CDSBE) | $C_{42}H_{70}$-*n*-$O_{35}$* $(C_4H_8SO_3Na)_n$ | TOXSYS | Rat, primate, mouse |
| Carboxymethylcellulose (CMC) | 14 | 9000-11-7 | Acetic acid; 2,3,4,5,6-pentahydroxyhexanal | $C_8H_{16}O_8$ | Mehlman, M. A. (2001), "Carboxymethylcellulose" | Primate, rat |
| Carboxymethylcellulose calcium | 15 | 9050-04-8 | Calcium CMC | | TOXSYS | Dog |
| Carboxymethylcellulose sodium | 16 | 9004-32-4 | Carmellose sodium | | Bar, A. (1995); Freeman, C. (2003); Cavender, F. L. (2001), "Sodium Carboxymethyl Cellulose"; Bachmann, E. (1978) | Rabbit |
| Cavasol W7 | NA | 128446-35-5 | 2-Hydroxypropylcycloheptaamylose | | Wacker Fine Chemicals (2006) | |
| Cetyl alcohol | 17 | 36653-82-4 | Hexadecan-1-ol | $C_{16}H_{34}O_1$ | TOXSYS; Bevans, C. (2001), p. 494. | Mouse |
| Citrate buffer | 18 | 77-92-9 | Sodium citrate–citric acid buffer | | Schonwald, S. (2004), "Acidifying and Alkalinizing Agents" | Dog, rat |
| Citric acid buffer | 19 | 77-92-9 | | $C_6H_8O_7$*$H_2O$ | Cragg, S. T. (2001), "Citric Acid" | Rat |
| CMC with dimethicone | NA | 9004-32-4<br>9006-65-9 | Carboxymethylcellulose sodium<br>Trimethyl-trimethylsilyloxysilane | $C_6H_{18}OSi_2$ | NA | |
| Coconut oil | NA | 8001-31-8 | NA | | Shadnia, S. (2005); National Toxicology Program (2001) | |
| Collagen matrix | 20 | 9007-34-5 | Collagen human | | McCarthy, D. M. (2002); Clark, D. P. (1989) | Primate, rabbit |
| Corn oil | 21 | 8001-30-7 | Corn germ oil, glyceridic | | TOXSYS; Wu, B. (2004); Dupont, J. (1990); DeWitt, J. C. (2005) | Dog, rat, mouse, rabbit, chick embryo |
| Cremophore EL | 22 | 61791-12-6 | Polyoxyl castor oil | | TOXSYS; Gelderblom, H. (2001); Ramadan, L. A. (2001); Lorenz, W. (1977) | Dog, rat |

| | | | | | | |
|---|---|---|---|---|---|---|
| Cyclohexane | 23 | 110-82-7 | Hexahydrobenzene; hexamethylene; hexanaphthene | $C_6H_{12}$ | Kreckmann, K. H. (2000); Malley, L. A. (2000); Gad, S. E. (2005), "Cyclohexane" | Rat, rabbit |
| DAM PEG (polyethylene glycol) | 24 | | | | NA | Dog, rat |
| Dextrose | 25 | 50-99-7 | D-Glucose, anhydrous; dextrosol | $C_6H_{12}O_6$ | Buard, A. (2003) | Dog, rat |
| Diethyleneglycolmonoethylether | 26 | 111-90-0 | | | Hardin, B. D. (1983); Hardin, B. D. (1984) | Primate |
| DMSO | 27 | 67-68-5 | Dimethylsulfoxide | $C_2H_6OS$ | TOXSYS; White, C. W. (1983); Augustine, K. A. (2000); Ali, B. H. (2001); Schonwald, S. (2004), "Irrigating Solutions" | Dog, rat, guinea pig, primate, mouse, rabbit |
| Dulbecco's modified PBS | 28 | | | | NA | Rat |
| EDTA | NA | 60-00-4 | Ethylenediaminetetraacetic acid | $C_{10}H_{16}N_2O_8$ | TOXSYS; Heimbach, J. (2000); Lanigan, R. S. (2002); Cavender, F. L. (2001), "Ethylenediamine Tetraacetic Acid" | |
| Ethanol | 29 | 64-17/5 | Ethyl alcohol | $C_2H_6O$ | TOXSYS; Church, A. S. (1997); Bevan, C. (2001), "Ethanol" | Dog, rat, primate, mouse |
| Gelucire 44/14 | 30 | 121548-04-7 | PEG-32 glyceryl laurate | | Yan (2005); Kawakami (2004) | Rabbit, rat, dog |
| Gelucire 50/13 | NA | 121548-05-8 | G-50-13 | | Sharma (2006) | |
| Glucose | 31 | 50-99-7 | Dextrose | $C_6H_{12}O_6$ | Robertson (2003) | Dog, rat, primate |
| Glycerol | 32 | 56-81-5 | Glycerine | $C_3H_8O_3$ | TOXSYS; Anderson, R. C.; Final report (2004) Glycerols | Rat, guinea pig, mouse, rabbit |
| Gum tragacanth | 33 | 9000-65-1 | | | Anderson, D. M. (1989); Bachmann, E. (1978); Hagiwara, A. (1991) | Mouse |

**TABLE 67** ***Continued***

| Excipient/Vehicle | Data in Table Number | CAS Number | Chemical Name | Formula | Key Toxicity Review Articles[a] | Animal Species Evaluated |
|---|---|---|---|---|---|---|
| Gum xanthane | NA | 11138-66-2 | | $(C_{35}H_{49}O_{29})_n$ | NA | |
| Hydroxypropyl β-cyclodextrin | 34 | 94035-02-6 | | $C_{42}H_{70}$–*n*-$O_{35}$ | Gerloczy, A. (1994) | Dog, rat |
| Hydroxypropyl cellulose | 35 | 9004-64-2 | Methocel | | TOXSYS; Cavender, F. L. (2001), "Hydroxypropyl Cellulose" | Rat |
| Hydroxypropyl methylcellulose | 36 | 9004-65-3 | Benecel MHPC, Hypromellose | | TOXSYS; Geerling, G. (2001); Maki, K. C. (2000); Rosen, P. A. (1987); Mehlman, M. A. (2001), "Hydroxypropyl Methylcellulose"; Obara, S. (1992) | Dog, rat, mouse |
| Isopropyl alcohol | 37 | 67-63-0 | *sec*-Propyl alcohol | $C_3H_8O$ | TOXSYS; Tyl, R. W. (1994); Bevan, C. (2001), "Isopropanol" ; Sivilotti, M. L. A. (2004); Burleigh-Flayer, H. (1998); Church, A. S. (1997); Allen, B. (1998) | Rabbit |
| Isopropyl myristate | 38 | 110-27-0 | Crodamol IPM | $C_{17}H_{34}O_2$ | TOXSYS; Komatsu, H. (1979); Campbell, R. L. (1981) | Rabbit |
| Labrafil M1944 | 39 | 62563-68-2 | Labrafil | | TOXSYS; Beckwith-Hall, B. M. (2002) | Dog |
| Labrasol | 40 | 85536-07-8 | Polyglycolyzed Glycerides | | Hu, Z. (2001) | Rat, dog, rabbit |
| Lactose | 41 | 63-42-3 (anhy) | *O*-β-D-Galactopyranosyl-(1->4)-α - D-glucopyranose | $C_{12}H_{22}O_{11}$ (anhy) | TOXSYS; Baldrick, P. (1997); Ahmad, S. K. (2004) | Primate |
| Lanolin | 42 | 8006-54-0 | Lanolin | | Kligman, A. M. (1983) | Rabbit |
| L-Ascorbic acid | 43 | 50-81-7 | Cevatine, Cevex, Cevital | $C_6H_8O_6$ | Bendich, A. (1990); Dykes, M. H. (1975); Temple, B. R. (2004) | Rat |
| Lauroglycol | 44 | 27194-74-7 | Lauric acid, monoester with propane-1,2-diol | | Liu (2006) | Rabbit, rat |

| | | | | | | |
|---|---|---|---|---|---|---|
| Maltitol solution | 45 | 9053-46-7 | Liquid Maltitol | $C_{12}H_{24}O_{11} + C_6H_{14}O_6$ | Walker, R. (1978); Modderman, J. P. (1993) | Rat |
| Maltol | 46 | 118-71-8 | 3-Hydroxy-2-methyl-4*H*-pyran-4-one | $C_6H_6O_3$ | Hironishi, M. (1996); Murakami, K. (2006) | Guinea pig, rabbit |
| Mannitol | 47 | 69-65-8 | D-Mannitol | $C_6H_{14}O_6$ | TOXSYS; Horvath, I. P. (1982); Lina, B. A. (1996) | Primate |
| Methane sulfonic acid | NA | 75-75-2 | Methylsulfonic acid | $CH_4O_3S$ | TOXSYS; Shertzer, H. G. (2001) | |
| Methyl cellulose | 48 | 9004-67-5 | Cellulose methyl ester | | TOXSYS; Mehlman, M. A. (2001), "Methylcellulose"; Bachmann, E. (1978); Sellers, R. S. (2005) | Rat, guinea pig, primate, mouse, rabbit, dog |
| Miglyol 810 | NA | 85409-09-2 | Caprylic, capric triglycerides | | TOXSYS; Traul, K. A. (2000); Sellers, R. S. (2005) | |
| Mineral oil | 49 | 8012-95-1 | Liquid paraffin | | TOXSYS; Dalbey, W. E. (2003); Trimmer, G. W. (2004); Nash, J. F. (1996) | Rat, mouse, dog |
| Neobee 1053 | NA | 73398-61-5 | Medium–chain triglycerides | | NA | |
| *N*-Methylpyrrolidone (Pharmasolv) | NA | 872-50-4 | 1-Methyl-2-pyrrolidinone | $C_5H_9NO$ | TOXSYS; Solomon, G. M. (1996); Trochimowicz, H. J. (2001); Lee, K. P. (1987) | |
| PBS (phosphate-buffered saline) | 50 | | | | NA | Rat, primate, mouse |
| Peanut oil | 51 | 8002-03-7 | Arachis oil, Fletcher's | | TOXSYS; "Final Report on the Safety Assessment of Peanut …" (2001) | Rat |

**TABLE 67** ***Continued***

| Excipient/Vehicle | Data in Table Number | CAS Number | Chemical Name | Formula | Key Toxicity Review Articles[a] | Animal Species Evaluated |
|---|---|---|---|---|---|---|
| PEG 300 | 52 | 25322-68-3 | Polyethylene glycol #300 | $H_0CH_2(CH_2OCH_2)_mCH_2OH$ | TOXSYS; Cavender, F. L. (2001), "Polyethylene Glycols"; Patel, M. (2005) | Guinea pig, mouse, rabbit |
| PEG 400 | 53 | 25322-68-3 | Polyethylene glycol #400 | $(C_2H_4O)_nH_2O$ | TOXSYS; Hermansky, S. J. (1995); Patel, M. (2005) | Rat, minipig, guinea pig, mouse |
| Petrolatum | 54 | 8009-03-8 | Yellow soft paraffin | | TOXSYS | Rabbit |
| Poloxamer | 55 | 9003-11-6 | Lutrol | $HO(C_2H_4O)$-$a$($C_3H_6O$)-$b$-H | Frim, D. M. (2004); TOXSYS; Grindel, J. M. (2002) | Rat, mouse |
| Polysorbate 80 | NA | 9005-65-6 | Polyoxyethylene (20) sorbitan monooleate | | TOXSYS; (1992) NTP Toxicology and Carcinogenesis Studies | |
| Povidone | 56 | 9080-59-5 | 2-Methoxy-6-methyl-phenol | $C_8H_{10}O_2$ | Beji, S. (2006) | Rat |
| Propylene glycol | 57 | 57-55-6 | 1,2-Dihydroxypropane | $C_3H_8O_2$ | TOXSYS; (1999) Final Report on the Safety Assessment of Propylene Glycol …; Cavender, F. L. (2001), "Propylene Glycol" | Rat, minipig, mouse, dog |
| Rameb 7.5% | 58 | | Randomly methylated-β-cyclodextrins | | Challa, R. (2005) | Primate |
| Sesame oil | 59 | 8008-74-0 | Sesame oil | | TOXSYS; Farber, T. M. (1976); Genovese, R. F. (1999) | Rat, mouse, rabbit, dog |
| Sodium acetate trihydrate buffer | 60 | 6131-90-4 | | | NA | Primate |
| Sodium chloride | 61 | 7647-14-5 | Salt, Halite | NaCl | Meneely, G. R. (1953); Meneely, G. R. (1958); Teitelbaum (2001); Caraccio, T. R. (2004), "Over the Counter Products" | Dog, rat, primate, mouse, rabbit |

| | | | | | | |
|---|---|---|---|---|---|---|
| Sodium phosphate | 62 | 7558-80-7 | | | Moore, G. L. (1988); Pierce, S. W. (2001) | Dog, rat |
| Solutol HS15/purified water | NA | 70142-34-6 | Polyethylene glycol-15-hydroxystearate | | Coon (1991); Ruchatz (1998) | |
| Succinate, sodium | NA | 150-90-3 | Succinic acid sodium salt | $C_4H_4Na_2O_4$ | TOXSYS; MSDS; Cragg, S. T. (2001), "Succinic Acid" | |
| Tartaric acid | 63 | 87-69-4 | *d*-Tartaric acid; 2,3-dihydroxybutanedioic acid | $HOOC(CH_2O)_2COOH$ | Sourkes, T. L. (1950); Cragg, S. T. (2001), "Tartaric Acid" | Rat, rabbit |
| Transcutol | 64 | 111-90-0 | 2-(2-Ethoxyethoxy)ethanol | $C_6H_{14}O_3$ | Liu, Z. (2006) | Cat, rabbit, rat |
| Trisodium citrate dihydrate | 65 | 6132043 | | $C_6H_5Na_3O_7 \cdot 2H_2O$ | NA | Dog, rat, mouse |
| Tween 20 | 66 | 9005-64-5 | Polysorbate 20 NF | | TOXSYS; Bartsch, W. (1976) | Rat, mouse |
| Tween 80 | 67 | 9005-65-6 | Armotan pmo-20, Tween(R) 80 | | TOXSYS; Daher, C. F. (2003); Fisherman, E. W. (1974); Sellers, R. S. (2005) | Rat, primate, mouse, dog |
| Xylitol | 68 | 87-99-0 | Xylite | $C_5H_{12}O_5$ | TOXSYS; Takahashi, K. (1999) | Primate |

*Note*: NA, not currently available.

[a]See Gad, S. C., Cassidy, C., Aubert, N., Spainhour, B., and Robbe, H. (2006). Nonclinical vehicle use in studies by multiple routes in multiple species. *Int. J. Toxicol.* 25:499–522.

# *Global Directory of Contract Toxicology Laboratories*

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

| Laboratory | Location | Phone Number | Website | Additional Services[a] | Rat | Rabbits | Dog | Carcinogenicity | DART | Inhalation | Primate | Pig | IV Infusion | Genotoxicity | Devices | Metabolism | Analytical | Special Studies[b] |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ABC Labs. | Missouri, IA; North Ireland; United Kingdom | (573) 474-8579, 44 (0) 2870 320 639 | www.abclabs.com | Bioanalysis, methods development radioactivity | X | | | | | | | | | | | X | X | |
| ADMETRx | Kalamazoo | (269) 372-3272 | www.admetrx.com | Yes | | | | | | | | | | | | | | |
| Applied Preclinical Services | Oxford, NJ | (908) 637-4427 | NO | | X | | X | | | X | | X | | | | X | | |
| Apptec/Wuxi | St. Paul, MN; Philadelphia, PA; Atlanta, GA | (888) 794-0077 (800) 622-8820 (888) 847-6633 | www.apptec-usa.com | Biotech, biocomp./ med. devices, microbiology | X | X | X | | | | | X | | X | X | | X | |
| Aptuit | Ledbury, India, United Kingdom and United States | 44 (0) 131 451 2560, (877) 988-2100 | www.aptuit.com/ | Yes, clinical, development, + | X | X | X | X | X | | | X | X | X | X | X | X | SP |
| Austrian Research Center | Seibersdorf, Austria | 43(0) 50550-0 | www.arcs.ac.at/ | | X | X | | | X | | X | | | | | | | SP |
| Azopharma & Aniclin | Miramar, FL; Oxford, NJ | (954) 604-9879 | www.Aniclin.com | Implant/surgical | X | X | X | X | X | X | X | X | X | | | X | X | X |
| BAS | West Lafayette, IN | (800) 845-4246 | www.bioanalytical.com | Yes | X | X | X | X | X | | | X | X | X | X | X | X | |
| Battelle | Columbus, OH | (800) 201-2011 | www.battelle.org | Yes | X | X | X | X | X | X | X | X | X | | | X | X | |
| | Richland, WA | | | | X | | X | X | | X | | | | | | | | |
| BEC Labs | Toledo, OH | (419) 693-5307, (888) BEC-LABS | www.beclabs.com | Yes | | | | | | | | | | | X | | | |
| Bioagri | Piracicaba, Campinas, and Sao Paulo, BRAZIL | +55-19-3429-7720 | www.bioagri.com.br | | X | X | X | | | | | | | X | | | X | |
| Biocon | Farmington, MN | (651) 460-3330 | | | X | X | | | | | | | | | | | | |
| Bio-Life | Neillsville, WI | (715) 743-3171 | No | Yes | | | | | | | | | | | X | | | WL |

| | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Bio-Quant | San Diego, CA | (858) 450-0048 | www.bio-quant.com | Tumor models, disease models | X | | X | | | | | X | | | | X | | |
| BioReliance | Rockville, MD | (800) 738-1000 | www.bioreliance.com | Full-range genotoxicity,[c] transgenic mouse CA models, p53 | X | X | | | | | | | X | X | | | X | |
| Biotechnics | Hillsborough, NC | (919) 245-3114 | www.biotechnics-inc.com | Yes | X | X | X | | | | | | | X | | | | |
| BRI | Vancouver, BC | (604) 432-9237 | www.bripharm.com | Yes, pharmaceutical, development, + | X | | | | | | | | | X | | X | X | |
| BRIDGE | Beijing, China; Gaithersburg, MD | +650-859-3853 | www. bridgepharmaceuticals. com | Regulatory, consulting | X | | X | | | | X | | X | X | | X | X | |
| BRT | Raleigh, NC | (919) 851-4499 | | *Specialty immunotox | X | X | | | | | | | | | X | | | |
| BTC | Irvine, CA | (949) 660-3185 | www.biologicaltestcenter. com | Yes, pediatrics, + | X | X | X | | | | | X | X | X | X | X | X | |
| Calvert | Olyphant, PA | (570) 586-2411 | www.calvertlabs.com | Yes | X | X | X | X | X | | X | X | X | | X | X | X | SP |
| CEDRA | Austin, TX | (512) 834-7766 | www.cedracorp.com | Yes | | | | | | | | | | | | | X | |
| Celsis | St. Louis, MO | (314) 487-6776 | www.celsislabs.com | Yes, stability, methods | X | | | X | X | X | | | | X | X | X | X | X |
| Central Toxicology Labs. (Sygenta) | Alderley Park, Macclesfield SK104TJ United Kingdom | +44 16255 15852, +44 16255 17314 (fax) | www.syngenta.com | Yes | X | X | X | X | X | X | X | | X | X | | X | X | SP |
| Cerep | Redmond, WA | (425) 895-8666 | www.cerep.com | Chemistry, in vitro | | | | | | | | | | | | | | SP |
| CeeTox Inc. | Kalamazoo, MI | 269.353.5555 | www.ceetox.com | Chemistry, in vivo, cell culture | | | | | | | | | | | | X | X | X |
| Chantest | Cleveland, OH | (216) 332-1665, (216) 332-1706 (fax) | www.chantest.com | HERG, ion channels | | | | | | | | | | | | | | |
| Charles River | Worchester, MA | (978) 658-6000 | www.criver.com | Tumor models, implant studies | X | X | X | | | | X | X | X | | X | X | X | SP |
| | Sparks, NV | (775) 331-2201 | | Primates, immunotox[c] | X | | | | | | X | | X | | | | | |
| | Horsham, PA[d] | (215) 443-8710 | | Phototoxicity,[d] ferret emesis[d] | | | | | X | | | | | | | | | |
| | Spencerville, OH | (419) 647-4196 | | | X | X | X | X | | | | | | | | | | |

| Laboratory | Location | Phone Number | Website | Additional Services[a] | Rat | Rabbits | Dog | Carcinogenicity | DART | Inhalation | Primate | Pig | IV Infusion | Genotoxicity | Devices | Metabolism | Analytical | Special Studies[b] |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Tranent, Scotland (Inveresk) | 44 (0) 1875 614555 | | Clinical, ferret emesis | X | X | X | X | X | X | X | X | | X | | X | X | SP |
| | Montreal, Quebec (CTBR) | (514) 630-8200 | | Clinical, ferret emesis | X | X | X | X | X | X | X | X | X | | | X | X | SP |
| Cirion | Canada | (450) 688-6445 | www.cirion.ca | Yes, clinical | X | X | X | | | | X | X | | X | X | X | | |
| CIT—Centre Int. de Toxicologie | Cedex, France | +33 2 32 292626, +33 2 32 678705 (fax) | www.citox.com | | X | X | X | X | X | | X | X | X | X | | X | X | SP |
| Clintox Bioservices | Hyderabad, India | 91(40) 55153222 | www.clintoxbio.com | Phototox, hamsters | X | X | | | | | | | | | | X | | SP |
| Commonwealth Biotechnologies, Inc. | Richmond, VA | (804) 648-3820 | www.cbi-biotech.com | Genomics | | X | | | | | | | | | | | | |
| Comparative Biosciences | Sunnyvale, CA | (408) 504-8871, (408) 738-9278 | www.compbio.com | Special discovery,[c] models | X | X | | | | | | | X | | X | X | X | |
| CorDynamics | Chicago, IL | (312) 421-8876 | www.cordynamics.com | CV safety only | | | | | | | | | | | | | | |
| Covance | Vienna, VA | (888) COVANCE, (888) 541-LABS | www.covance.com | Clinical, consulting | X | X | X | X | X | | X | X | X | X | X | X | | SP |
| | Madison, WI | | | | | | | | | | | | | | | X | X | SP |
| | Harrogete, United Kingdom | | | | | | | | | | | | | | | | | |
| | Muenster, FRG | | | | | | | | | | | | | | | | | |
| CPTC | Fairfield, NJ | (973) 808-7111 | www.cptclabs.com | Yes, clinical + | X | X | | | | | | | | | X | | X | |
| CXR | Dundee, United Kingdom | 44-1382-432163 | www.cxrbiosciences.com | Humanized mouse models | | | | | | | | | | | | | | |
| Dermtech | San Diego, CA | (800) 808-2774 | www.dermtechintl.com | Yes, consulting | | | | | | X | X | | | | | | | |
| Enviro-Bio-Tech-Ltd. | Bernville, PA | (610) 488-7664 | No | Chicken, cattle, food animals | | | | | | | | | | | | | X | WL |
| Experimur | Chicago, IL | (312) 842-9700 | http://www.experimur.com | PK | X | X | X | X | X | | X | X | X | | | | | |
| Exygen | State College, PA | (800) 281-3219, (800) 272-1019 (fax) | www.exygen.com | Methods, development | | | | | | | | | | | | X | X | |
| Fraunhofer ITEM | Nikolai-Fuchs Str., D-30625 Hanover, Germany | +49 511 5353 0, +49 511 5353 155 (fax) | www.item.fraunhofer.de | | X | X | X | X | X | X | | | | X | | X | X | |

| | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Frontier Biosciences | China | U.S. Office, (301) 251-0231 | www.frontierbsi.com | | X | X | X | X | X | | X | | | X | | X | | SP |
| Genelogic | Gaithersburg, MD | (301) 330-3733, (800) 436-3564 | www.genelogic.com | Yes | X | X | X | X | X | | X | X | X | | | X | X | SP |
| Gibraltar Laboratories, Inc. | Fairfield, NJ | (973) 227-6882, (973) 227-0812 (fax) | www.gibraltarlabsinc.com | Microbiology | X | X | | | | | | | | X | X | | | |
| Gwathmey | Cambridge, MA | (617) 491-0022, (617) 492-5545 (fax) | www.gwathmey.com | Consulting, biology | X | X | | | | | | | | | X | X | | |
| Harlan | Rehovot, Israel | 972 (0) 8-9409451 | None | Bioproducts, breeder | X | | X | | | | | | | | | X | X | SP |
| Haskell Laboratory | Wilmington, DE | (306) 366-5200 | | Yes | | | | | | | | | | X | | | | |
| Huntingdon | East Millstone, NJ | (732) 873-2550 | www.huntingdon.com | Discovery support | X | X | X | X | X | X | X | X | X | | | X | X | SP |
| | Cambridgeshire, United Kingdom | 44 (0) 1480 892000 | | Radioactivity | X | X | | X[e] | X | X | | | | X | | | X | |
| Hurley Consulting Associates Ltd. | One Main Street, Chatham, NJ, 07928 | 1-770-529-1154, 1-770-529-3783 (fax) | www.hurleyconsulting.com | Statistical analysis, study design, preclinical assessments, data management, QA, regulatory affairs, pharmacovigilances, training, etc | | | | | | | | | | | | X | X | |
| Ibex | Logan, UT | (435) 752-4448 | www.ibexpreclinical.com | Cardio telem. | X | | | X | X | | | | | | X | X | | |
| IITRI | Chicago, IL; McLean, VA | (312) 567-4000, (703) 918-4480 | www.iitri.com | Yes | X | | X | X | X | X | X | | | | | | | |
| ILS | RTP, NC | (919) 544-5857 | www.ils-limited.co.uk | Data analysis, ecotoxicology | X | | | X | X | | | | | X | | X | X | SP |
| INA (Int. Non-clinical Assessment) | Nagano-ken, Japan | +81-265-72-6616, +63-49-541-2294 (fax) | www.ina-research.co.jp/en/ | Pediatric tox | X | | X | X | X | | X | | | X | | | | |
| Industrial Toxicology Research Center (ITRC) | Lucknow Nadu, India | +91-522-2211547, +91-522-2228227 (fax) | www.itrcindia.org/index.html | Irritation, behavioural | | | | X | X | | | | | X | | | | |
| Intox | Pune, India | +91 22923633 | www.intoxlab.com/html/profile.htm | Irritation, Farm animals | X | X | | X | X | | | | | X | | | | |

| Laboratory | Location | Phone Number | Website | Additional Services[a] | Rat | Rabbits | Dog | Carcinogenicity | DART | Inhalation | Primate | Pig | IV Infusion | Genotoxicity | Devices | Metabolism | Analytical | Special Studies[b] |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Inveresk | Tranent, Scotland | 44 (0) 1875 614555 | www.inveresk.com | Clinical, ferret | X | X | X | X | X | X | | X | | X | | X | X | SP |
| | Montreal, Quebec | (514) 630-8200 | | emesis | X | X | X | X | X | X | X | X | X | | | X | X | SP |
| Irvine Analytical Labs, Inc. | Irvine, CA | (877) 445-6554 | www.ialab.com | Yes | | | | | | X | | | | | | | X | |
| Int. Inst. of Biotech and Toxicology | Tamil Nadu, India | +91-4111-274246 +91-4111-274455 | | Biotechnology, entomology, agronomy | | | | | | | | | | | | | X | |
| Int. Hungary Ltd. | Veszprem,Hungary | +36-1-4314840 fax +36-88-545-301 | www.trc.hu | Yes, toxicology | X | X | X | X | | X | | | X | | | X | | |
| IPS Therapeutis | Sherbrooke, Quebec | 819-564-5391, 819-820-6831 (fax) | www.ipstherapeutique.com/ | hERG testing, CV, safety pharm | | | | | | | | | | | | | | |
| ITL Labs Ltd. | Delhi, India | 91-11-27915654 | www.indiamart.com/itl/ | | | | | | | | | | | | | | X | |
| ITR | Montreal, Quebec | (514) 457-7400 | www.itrlab.com | Affiliated labs | X | X | X | X | X | X | X | | X | X | | X | X | SP |
| ITRI | Albuquerque, NM | (408) 428-9988 | www.itri.com | Yes | X | | | X | | X | | | | | | | X | |
| Jai Research Foundation International | Gujarat, India | + 44 1684 891519 | www.jrfonline.com | | X | X | X | X | X | | | | | X | | X | X | |
| JSW Research | Austria | | www.jswresearch.com | Yes | X | | | X | X | | X | | | X | | X | X | |
| KIT (Korean Institute of Technology) | Korea | +82-42-610-8105 | www.kitox.re.kr/eng_about.html | Yes | X | X | X | X | | | X | X | X | X | | | | |
| LABS | Laval, Canada | (450) 973-2240 | www.preclin.com | Yes | X | X | X | X | | | X | | X | | | X | | |
| | Veszprem, HU | (416) 815-0700 | www.trc.hu | | X | X | X | X | | X | | | X | | | X | | |
| Liberty Research | Waverly, NY | (607) 565-8131 (p) (607) 565-7420 (f) | NO | Cats, ferrets | | | X | | | | | | | | | | | |
| Lovelace | Albuquerque, NM | (505) 348-9456 | www.lrri.org | Clinical | | | | | | X | | | | | | | | |
| LPT | Hamburg, GE | 40-70-20-20 40-70-20-22-99 | www.LPT-pharm-tox.de | | X | X | X | X | X | X | X | X | X | X | X | X | X | SP |

| | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| LyChron | Mountain View, CA | 650-938-3675 | www.lychron.com | Pig device implant studies | | | | | | | | X | | | | | | |
| Maccine | Singapore | +65-6778-7305 | www.maccine.com | Yes | X | | | | | | | | X | | | X | X | |
| MB Research | Spinnerstown, PA | (215) 536-4110 | www.mbresearch.com | In vitro and alternative models | X | X | X | | | | | | | | X | | | |
| MDS Pharma Services | Lincoln, NE; L'Arbresle, France; Geneva, Switzerland | (514) 333-0033 | www.mdsps.com | Clinical, receptor screening | X | X | X | X | X | | X | | | X | | X | X | SP |
| | | | | | | | | | | | | | | X | | X | X | SP |
| | | | | | | | | | | | | | | | X | | | |
| | | | | | | | | | | | | | | | X | | | |
| Medicilon/MPI | Shanghai, China | 8621-51-320228 | www.medicilon.com | | X | | X | | | | X | | | X | | X | X | |
| MicaGenix | Greenfield, IN | (888)-779-5500 | www.micagenix.com | Histology, image analysis, Non-GLP mice and rats | | | | | | | | | | | | X | | |
| Micro Test | Agawam, MA | (800) 631-1680 | www.microtestlabs.com | Yes | | | | | | | | | | | X | | | |
| Midwest Bio Research | Evanston, IL | 847-682-6141(p) 847-556-6560(f) | www.midwestbioresearch.com | Consultation | | | | | | | | | | X | | X | | |
| MPI | Mattawan, MI | (269) 668-3336 | www.mpiresearch.com | Yes, radioactivity | X | X | X | X | X | | X | X | X | | X | X | X | SP |
| MSI | Minato-ku, Tokyo, Japan | 81 3 3454 7571 | www.ankaken.co.jp | Yes | | | | X | | X | X | | | | | X | | |
| MWRI | Kansas City, MO | (816) 753-7600 | www.mriresearch.org | Yes | X | | | X | | | | | | | | | | |
| nanoTox | Austin, TX; The Netherlands | (877) 804-2828 | www.nanotox.com | Yes, risk assessment of nanomaterials; particle sizing, nanodose formulations and analysis | | | | | | | | | | | | | X | X |
| NAMSA | Toledo, OH | (419) 666-9455 | www.namsa.com | Yes | | | | | | | | | | X | X | | | |
| | Irvine, CA | (949) 951-3110 | | | | | | | | | | | | X | X | | | |
| | Kennesaw, GA | (770) 427-3101 | | | | | | | | | | | | | | | | |
| Nelson Labs | Salt Lake City, UT | (800) 826-2088 | www.nelsonlabs.com | Yes | | | | | | | | | | | X | | X | |
| Next Century, INC | Newark, DE | (302) 453-0571 | www.nextcenturyinc.com | Yes | X | X | X | X | | X | X | X | X | | X | | | |
| Northern Biomedical Research | Muskegon, MI | (231) 759-2333 | | Yes, surgery, large animals | X | X | X | | | | X | X | X | | | X | | |
| Northview | Northbrook, IL | (847) 564-8181 | www.northviewlabs.com | Yes | X | X | X | X | X | | | | | X | X | X | X | |
| | Hercules, CA | (510) 694-9000 | | | | | | | | | | | | | | | | |
| | Spartan, SC | (864) 574-7728 | | | | | | | | | | | | | | | | |

| Laboratory | Location | Phone Number | Website | Additional Services[a] | Rat | Rabbits | Dog | Carcinogenicity | DART | Inhalation | Primate | Pig | IV Infusion | Genotoxicity | Devices | Metabolism | Analytical | Special Studies[b] |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| NOTOX | s'Hertogenbosch, The Netherlands | 31 73 640 67 00, 31 73 640 67 99 (fax) | www.notox.nl | Yes, birds, + | X | X | X | X | X | | | X | X | X | X | X | | WL |
| Nucro-Technics | Ontario, Canada | (416) 438-6727, (416) 438-3463 (fax) | www.nucro-technics.com | Microbiology | X | X | X | X | | | | | | X | X | | X | |
| Pacific Biolabs | Hercules, CA | (510) 694-9000 | http://www.pacificbiolabs.com/home.htm | Yes | X | X | X | X | X | | | | | X | X | X | X | |
| Pathology Experts | Basel, Switzerland Rye, NY | 44-61-422-09-29 (888) 473-4554 | www.pathexperts.com | Histology and histopathology | | | | | | | | | | | | | | |
| Perry Scientific | San Diego, CA | (858) 560-9000, (858) 467-9608 (fax) | www.perryscientific.com | Yes; large animals | X | X | X | X | X | | | X | X | | X | X | X | SP |
| Porsolt & Partners | Boulougne-Billancourt, France | +33146109990 | http://www.porsolt.com/ | Yes | X | X | X | | | | X | | X | | | X | X | |
| Product Safety Labs | Dayton, NJ | (732) 438-5100, (732) 254-9200 | www.productsafetylabs.com | Environmental, ferret emesis studies, radioactivity | X | X | | X | X | X | | | X | X | X | | X | WL |
| Provident Preclinical, Inc | Doylestown, PA | 215-348-3868, 215-348-5081 (fax) | www.ppicro.com | Yes | X | X | X | X | | X | X | X | | | X | | | |
| Quest | Newark, DE | (302) 369-5601 | www.questpharm.com | Yes, radioactivity | X | X | X | | X | | X | | | X | | X | X | |
| Rallis Research Center | Bangalore, India | +91-80-28394959, +91-80-28394015 | www.rallis.co.in | Yes | X | X | | X | | | | | | X | | X | X | |
| RCC Ltd. | Zelgliweg 1 CH-4452 Itegen, Switzerland | (410) 385-1666, US sales office, +41 61 975 1111 | www.rcc.ch | Consulting, regulatory | X | X | X | X | X | X | | X | X | X | X | X | X | SP |
| Ricerca | Concord, OH | (888) 763-4797 | www.ricerca.com | Synthesis,[a] product development, radioactivity | X | X | X | X | X | X | | | | | | X | X | |
| RTC S.p.A. | Via Tito Speri 12, I-40 Pomezia, Rome, Italy | +39 06 910 951, +39 06 910 5737 (fax) | www.rtc.it | Yes, consulting, regulatory | X | X | X | | X | | | | | | | | | WL |

| | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| RTI | RTP, NC | (919) 541-6000 | www.rti.org | Yes | X | | | X | X | | | | | | | X | X | |
| SafePharm | Derby, United Kingdom | 00 44 (0) 1332792896 | www.safepharm.com | Yes, (fish+) | X | | | X | X | X | | | | X | X | X | X | **SP** |
| SCANTOX | Hestehavevej 36a, DK-4623 Lille Skensved, Denmark | +45 56 82 11 00, +45 56 82 12 02 (fax) | www.scantox.com | | X | X | X | X | X | X | | X | X | X | X | | | |
| Sequani Limited | Bromyard Road, Herefordshire HR8 1LH, Ledbury, United Kingdom | +44 1531 634121, +44 1531 634753 (fax) | www.sequani.com | | X | X | X | X | X | | | X | X | X | X | | | |
| SGS Life Sciences Division | Fairfield, NJ | (800) 747-8782 | www.ustesting.sgsna.com | Implant, combustion, fish, industrial, consumer products | X | X | | X | | | | | | X | X | | X | **SP, WL** |
| Sinclair | PO Box 658 | (573) 387-4400 | www.sinclairresearch.com | Radiolabeled; cats; other | X | | | X | | | | X | X | | X | X | | |
| Sitek | Rockville, MD | (301) 926-4900 | www.siteklabs.com | Yes | X | | | | | | | | | X | X | X | | |
| Skeletech | Bothell, WA | (425) 424-2663 | www.skeletech.com | Bone,[c] CNS | X | X | X | | | | X | X | X | | X | | | |
| SNBL | Kagoshima, Japan | 81 (0) 99-294-2600, 81 (0) 99-294-3196 (fax) | www.snbl.com | Yes, clinical, +, Irritation | X | X | X | X | X | X | X | X | X | X | | | | **SP** |
| SNBL USA Ltd. | 6605 Merrill Creek Parkway, Everett, WA 98203 | (425) 407 0121, (425) 407 8601 (fax) | http://www.snblusa.com/ | Formulation | X | | | | X | | X | | X | | | | X | **SP** |
| Southern RI | Birmingham, AL | (888) 322-1166 | www.southernreaserch.org | Yes, formulation, development, tumor models | X | X | X | X | X | | X | X | X | X | X | X | X | **SP** |
| SP1-bio | France | 33 (0) 169 53 14 00 | www.spibio.com | Clinical | X | | | | | | | | | | | X | X | |
| Spring Valley Lab. | Woodbine, MD | (800) 864-1839 | www.springvalleylabs.com | Ferrets | X | X | | X | X | | | | | | X | | | |
| SRI | Palo Alto, CA | (650) 859-2000, (866) 451-5998 | www.sri.com | Formulation, CTM | X | X | X | X | | | | X | | X | X | X | X | |
| Stillmeadow Inc. | Sugarland, TX | (281) 240-8828 | www.stillmeadow.com | Companion animal studies | X | X | X | | | X | | X | | | | | X | |
| STS duo Tek | Rush, NY, Henrietta, NY | (716) 321-1130, (800) 836-4850 | www.stsduotek.com | Yes, ocular | X | X | | | | | | | X | X | X | | | |
| Tandem Labs | Ewing, NJ | (609) 434-0044 | www.tandemlabs.com | Yes | | | | | | | | | | | | | X | |

| Laboratory | Location | Phone Number | Website | Additional Services[a] | Rat | Rabbits | Dog | Carcinogenicity | DART | Inhalation | Primate | Pig | IV Infusion | Genotoxicity | Devices | Metabolism | Analytical | Special Studies[b] |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| TNO Pharma | Utrechtseweg 48, NL-3700AJ Zeist, 3704 HE, The Netherlands | +31 30 694 4806, +31 30 694 4845 (fax) | www.voeding.tno.nl | Yes, packaging, food and nutrition, clinical | X | X | X | X | X | X | X | X | X | X | X | X | X | **SP, WL** |
| Torrent Pharmaceuticals Ltd. | Gujarat, India | | www.torrentpharma.com | | | | | X | X | | | | | X | | | | |
| TOXIKON Corp. | 15 Wiggins Ave., Bedford, MA 01730 | (781) 275 3330, (781) 271-1136 (fax) | www.toxikon.com | Yes | X | X | X | X | X | | | X | X | X | X | X | | **WL** |
| UIC Tox Research Lab | Chicago, IL | (312) 996-2123 | www.uic.edu/labs/tox | Yes, radioactivity | X | X | X | | X | X | X | | X | | | X | X | |
| Valley Biosystems | West Sacramento, CA | 916-374-2735 | www.valleybiosystems.com | | | | | | | | X | | | | | | | |
| VIMTA | Hyderabad, India | +91-40-2726 4141 | www.vimta.com | Clinical | | | | | X | | | | | X | | | X | |
| Vivo Bio Tech, Ltd. | Hyderabad, India | 92-40-55784714, 91-40-55776112 (fax) | www.vivobio.com | Immunodeficient, Mice | X | X | | | X | | | | | X | | X | | |
| WIL Research Labs, Inc. | 1407 George Rd., Ashland, OH 44805 | (419) 289 8700, (419) 289 3650 (fax) | www.wilresearch.com | Yes, radioactivity | X | X | X | X | X | X | X | X | X | | | X | X | **SP, WL** |
| WuXi Pharmatech | 288 FuTe ZhongLu WaiGaoQiao, Shanghai 200131, P.R. China | | http://www.pharmatechs.com | Rodent tox. (oral, IV, and PK) | X | | | | | | | | X | X | | | | |
| XenoBiotics | Plainsboro, NJ | (609) 799-2295, (609) 799-7497 | www.xbl.com | Clinical, radioactivity | X | X | X | | | | | | | | | X | | |
| Xenotech | Lenexa, KS | (913) 438-7450 | www.xenotech.com | Yes | | | | | | | | | | | | X | | |

[a]A "yes" or "+" in this column indicates that there are more services available than just those listed.

[b]SP in this column indicates safety pharmacology, WL wildlife/environment testing is available and EN that environmental testing in available.

[c]Limited availability studies.

[d]Near unique capabilities.

[e]Hamster carcinogenicity is available.

# Biopharmaceuticals: Definition and Regulation

LINCOLN TSANG, PhD, FRSC, FIBiol, FRPharmS, and NATHAN CORTEZ

**Contents**

1 Introduction
2 United States
  2.1 How the United States Defines Biologics and Biopharmaceuticals
  2.2 Legal Foundations for Regulating US Biopharmaceuticals
  2.3 Legal Requirements for US Biopharmaceuticals
3 European Union
  3.1 How EU Law Defines a Biological Medicinal Product
  3.2 Legal Foundation for Regulation of Biological Medicinal Product
4 Japan
5 Conclusion
  References

## 1 INTRODUCTION

Compared with other types of pharmaceutical products, products derived from a biological source or a biotechnological process are structurally complex and involve manufacturing processes that require tight control to ensure their safety, quality, and efficacy. Biological products, because of their sheer size, are orders of magnitude more complicated than small-molecule drugs. This can be seen by a comparison of molecular weight, which can be used as a measure of the size of a given product. Moreover the product arising from the manufacturing process is often not a pure, homogeneous mixture. Rather, various forms of these molecules are usually present in the final product.

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

In scientific terms, conventional biological products such as blood-derived clotting products, vaccines, and those derived from high technology such as those employing a recombinant DNA technology are characterized as biological products. Because of these differences in respect of the product characteristics and manufacturing process, the regulatory oversight of biological products is distinguishable from conventional pharmaceutical products based on small molecules. This chapter addresses legal framework governing biological products principally in the United States and in the European Union. The regulatory landscape in Japan is briefly described particularly in relation to the recent changes to Japan's Pharmaceutical Affairs Law.

## 2 UNITED STATES

The United States has one of the most active and sophisticated systems in the world for ensuring the safety and effectiveness of biopharmaceuticals. To understand this system, it is important to understand (1) how the United States defines biopharmaceuticals and biologics, (2) the legal foundations for regulating these products, and (3) the rules that apply during various stages, including research, development, approval, and marketing. This section also highlights how the United States regulates biologics in relation to drugs.

### 2.1 How the United States Defines Biologics and Biopharmaceuticals

US law does not have a single, simple definition for *biologics* or *biopharmaceuticals*. The Food and Drug Administration (FDA) recognizes that most biologic products "are complex mixtures that are not easily identified or characterized" [1]. Traditionally *biologics* are substances that are derived from living organisms, such as humans, animals, plants, and microorganisms [2]. Today *biologics* include these substances as well as those produced by biotechnology [2]. A federal statute defines *biological product* as a virus, therapeutic serum, toxin, antitoxin, vaccine, blood, blood component or derivative, allergenic product, or analogous product, or arsphenamine or derivative of arsphenamine (or any other trivalent organic arsenic compound) that is "applicable to the prevention, treatment, or cure of a disease or condition of human beings" [3]. The corresponding federal regulation uses similar language, but clarifies several key terms [4]:

1. A *virus* is interpreted to be a product containing the minute living cause of an infectious disease and includes filterable viruses, bacteria, rickettsia, fungi, and protozoa, among other things.
2. A *therapeutic serum* is a product obtained from blood by removing the clot or clot components and the blood cells.

3. A *toxin* is a product containing a soluble substance poisonous to laboratory animals or to human in doses of one milliliter or less (or equivalent in weight) of the product, and having the property, following the injection of nonfatal doses into an animal, of causing to be produced therein another soluble substance that specifically neutralizes the poisonous substance and that is demonstrable in the serum of the animal thus immunized.
4. An *antitoxin* is a product containing the soluble substance in serum or other body fluid of an immunized animal that specifically neutralizes the toxin against which the animal is immune.

The regulation also clarifies how additional products may be biologics if they are "analogous" to certain categories of products listed in the definition. A product is a biologic if it is analogous to the following [5]:

1. A *virus*, if prepared from or with a virus or agent actually or potentially infectious, without regard to the degree of virulence or toxicogenicity of the specific strain used.
2. A *therapeutic serum*, if composed of whole blood or plasma or containing some organic constituent or product other than a hormone or an amino acid, derived from whole blood, plasma, or a serum.
3. A *toxin* or *antitoxin*, if intended, regardless of its source of origin, to be applicable to the prevention, treatment, or cure of diseases or injuries of human through a specific immune process.

Although these definitions seem to be relatively concrete, biological products come in many forms, including drugs, devices, and "combination" products [6]. The FDA regulates biopharmaceuticals as both drugs and biologics because they meet both definitions. US law, as described above, defines *biological products* by referring to several categories of tangible products. In contrast, the law defines *drugs* by their functions [7]. The term *drug* means "articles intended for use in the diagnosis, cure, mitigation, treatment, or prevention of disease in man" and "articles (other than food) intended to affect the structure or any function of the body of man" [8]. Thus the definitions of *drugs* and *biologics* are not mutually exclusive, which allows the FDA to regulate some products as both.

## 2.2 Legal Foundations for Regulating US Biopharmaceuticals

To understand how biopharmaceuticals are regulated in the United States, it is helpful to understand the underlying legal bases for regulation, how these laws have evolved, and how regulatory responsibility for biologics has shifted. Currently the Public Health Service Act authorizes the FDA to ensure the safety, purity, and potency of biologics. The FDA approves biologics for mar-

keting under section 351 of the Act [9]. The FDA also regulates biopharmaceuticals as drugs under the Federal Food, Drug, and Cosmetic Act. Thus the FDA now delegates responsibility for regulating biopharmaceuticals to two centers within the agency: the Center for Drug Evaluation and Research (CDER) and the Center for Biologics Evaluation and Research (CBER). Regulation under the Public Health Service Act precludes the manufacture of generic, or "follow-on" biologicals and "biosimilars."

The foundations for this regulatory system were set in 1902 with the Biologics Control Act, the first legislation to regulate a specific class of drugs [7]. The Biologics Control Act was a response to tragedies in St. Louis, Missouri, and Camden, New Jersey, in which several people died after taking diphtheria and small pox vaccines [10]. The purpose of the Act was to authorize the regulation of certain biologics, require manufacturers to obtain licensing, and authorize the government to inspect manufacturing facilities [7]. The Act prohibited companies from selling or transporting biologics that were either not manufactured at facilities licensed and inspected by the government or not labeled with the manufacturer's name and an expiration date [7].

Since the 1902 Act, the laws and regulations for biologics have steadily evolved, and responsibility for regulating biological products has shifted several times. In 1903, the federal government issued the first biologics regulations, administered by the Hygienic Laboratory in the Public Health and Marine Hospital Service. The regulations required manufacturers to annually renew their licenses and make their facilities available for unannounced inspections. In 1919, the regulations were amended to require manufacturers to report changes in manufacturing methods, equipment, and personnel. The regulations also required manufacturers to maintain manufacturing records and submit certain product samples for government inspection and approval [7].

These initial laws and regulations laid the foundation for the current biologics regulatory scheme. From the beginning the United States regulated biologics and drugs differently. The government did not regulate nonbiologic drugs until it passed the Pure Food and Drugs Act in 1906, which did not address biologics or the 1902 Biologics Control Act [7]. In fact Congress did not formally recognize the difference between drugs and biologics until after it passed the 1938 Federal Food, Drug, and Cosmetic Act (FDCA) [12]. In 1944, Congress reenacted the 1902 Biologics Control Act and recodified the Public Health Service Act. A major issue was the definitional overlap between drugs and biologics [12].

Between 1902 and 1972, regulatory responsibility for biologics transferred several times, ultimately settling with FDA, as shown by this brief timeline of the relevant transfers:

1930 The Hygienic Laboratory within the Public Health and Marine Hospital Service is redesignated as the National Institutes of Health (NIH).

| | |
|---|---|
| 1937 | The NIH is reorganized, and responsibility for biologics is transferred to the Division of Biologics Control. In 1944 it is renamed the Laboratory of Biologics Control. |
| 1948 | The Laboratory of Biologics Control is integrated into the NIH's National Microbiological Institute, which later becomes the Institute of Allergy and Infectious Diseases. |
| 1955 | Responsibility for biologics is transferred to the new Division of Biologics Standards, a new independent entity within the NIH. |
| 1972 | The Division of Biologics Standards is transferred from the NIH to the FDA, becoming the Bureau of Biologics. |
| 1982 | The Bureau of Biologics is merged with the Bureau of Drugs to form the National Center for Drugs and Biologics (NCDB). |
| 1983 | The biologics component of the NCDB is renamed the Office of Biologics Research and Review, within the Center for Drugs and Biologics (CDB). |
| 1988 | CDB split into two centers, the Center for Biologics Evaluation and Research (CBER), and the Center for Drug Evaluation and Research (CDER). |
| 2003 | Transfer of therapeutic biological products from CBER to CDER. |

The steady stream of reorganizations in many ways reflects the difficulty of both categorizing and regulating biologics. The FDA continues to struggle with these responsibilities. For instance, since the FDA created CBER in 1988, the agency has both overhauled the way it approves biologics, and once again shifted responsibility for certain biologics. First, the FDA established a single approval application, the Biological License Application (BLA) through the Food and Drug Modernization Act of 1997 (FDAMA), the most comprehensive rewrite of food and drug laws since 1938. Second, in 2003, the FDA shifted responsibility for therapeutic biologics from CBER to CDER, given CDER's role in regulating therapeutic drugs. CDER's new responsibilities include a wide array of biological products, including monoclonal antibodies for in vivo use, therapeutic proteins, and immunomodulators [10]. CBER retained authority over traditional biologic products such as vaccines, allergenic extracts, antitoxins, blood, and blood products, as well as products composed of human, bacterial, or animal cells [10].

### 2.3 Legal Requirements for US Biopharmaceuticals

The regulation of biologics continues to evolve. The transcendent growth of biotechnology research, spurred by the Human Genome Project, almost ensures that biologic regulations will require further tinkering to accommodate new products. The following is a brief synopsis of relevant US laws and regulations at various stages, including research, development, approval, and

marketing. Where relevant, we highlight where the rules for biologics differ from drugs.

***Research and Development*** The United States heavily regulates the research and development of biologics. At the preclinical stage, FDA requires companies to comply with regulations on good laboratory practices (GLPs) at 21 CFR part 58. The GLP regulations seek to ensure the quality and integrity of preclinical safety data submitted to the FDA. GLPs apply to nonclinical (preclinical) laboratory studies intended to support research or marketing applications, and address a broad range of topics, including personnel, facilities, and equipment. Ideally preclinical studies to support safety are subject to GLPs and should be supported by a statement that the study was conducted in compliance with the good laboratory practice regulations in 21 CFR part 58, or if the study was not conducted in compliance with those regulations, a brief statement of the reason for the noncompliance (21 CFR 312.23 (8) (iii)).

At the clinical stage, FDA sets minimum standards for clinical trials through several regulations and guidance documents, collectively known as good clinical practices (GCPs). GCPs are designed to ensure the quality and integrity of data submitted to FDA and protect the rights of human subjects. GCPs govern key personnel involved in clinical trials—particularly sponsors, and investigators—and address several important areas, including informed consent, institutional review boards (IRBs), and investigational new drug (IND) requirements.

Informed consent is governed by both federal and state law [14]. These laws generally require that before participating in clinical trials, human research subjects state in writing that they understand the risks of the trial and are participating voluntarily. Each informed consent document must contain several elements required by FDA regulations [15].

IRBs are also governed by federal and state law. FDA regulations require IRBs to provide initial and continuing review of clinical trials [16]. IRBs must ensure that investigators and sponsors protect the study subjects, obtain adequate informed consent, and adhere to other safeguards and reporting requirements [16]. Moreover FDA regulations require IRBs members to meet specific membership criteria [17].

Investigational biologics are subject to the FDA's investigational new drug (IND) requirements [18]. The IND application is the first formal submission to FDA, and the application must be submitted before initiating any clinical studies [7]. It is not a request for commercial marketing approval; rather, it is a request to be exempt from the federal statute that prohibits shipping "unapproved" drugs across state lines. Thus an IND permit allows the product to be shipped during investigational studies. The purpose of the IND requirement is to assure the FDA that the safety and rights of subjects will be protected in all phases of the investigation, and that the quality of the studies are adequate to permit the FDA to evaluate the product's safety and effectiveness [13].

***Approval*** The FDA approves biologics for marketing through the biological license application (BLA), which requires the applicant to show that the product is safe, pure, and potent [19]. The BLA submission is typically the culmination of years of research and development, through which the company submits preclinical and clinical data, physiochemical information, biological activity results, and manufacturing information [7]. Previously the FDA approved biologics through two license applications, the product license application (PLA) and the establishment license application (ELA). In 1996, CBER consolidated these applications into a single BLA for certain products, and in 1997, Congress extended the BLA to all biological products.

Although the BLA process differs in some ways from the new drug approval (NDA) application process for nonbiologic drugs, the required showing of safety and efficacy is similar, if not identical, between drugs and biologics [20]. While the FDA requires biologics to be "safe, pure, and potent," the agency interprets this language as requiring the same type of evidence in NDAs for nonbiologic drugs [20]. Nevertheless, there are differences between the BLA and NDA that reflect CBER's historical emphasis on manufacturing and process control. For instance, the FDA requires BLA applicants to submit detailed information on manufacturing processes so that the FDA can determine whether the manufacturer can produce a product consistent with current good manufacturing practices (cGMPs) and the manufacturing specifications listed in the BLA. The manufacturer's facility is also a major factor—its construction, design, layout, validation processes, and environmental monitoring must meet FDA standards.

After approval, biologics manufacturers must comply with the FDA's cGMP regulations [21]. These regulations govern the manufacturer's use of raw materials, buildings and facilities, production and process controls, packaging and labeling, laboratory controls, stability testing, expiration dates, production records, and the company's overall quality system. Although the same cGMP regulations apply to drugs and biologics, manufacturing biologics can be quite different. Physically and chemically, biologics act differently than drugs [11]. They are less defined, less pure, less stable, and degrade in more complex ways than most drugs [11]. Their potency also depends greatly on the underlying organisms from which they are produced [11]. Thus, if a manufacturer makes relatively minor changes to the manufacturing process of a biologic, the FDA may require the manufacture to demonstrate through new clinical studies that the process produces the same results as the original clinical studies [11].

***Marketing and Postapproval Requirements*** Once the FDA approves a biopharmaceutical for marketing, the agency applies a different set of regulatory standards. The main postapproval requirements govern: (1) adverse event reporting, (2) manufacturing under cGMPs, (3) lot release testing, (4) general reporting, and (5) postmarketing studies.

- The FDA's adverse event reporting system does not differ significantly between drugs and biologics. However, the FDA did not have a comprehensive adverse event reporting system for biologics until 1994 [22]. Biologics manufacturers can use two reporting systems: MedWatch and the Vaccine Adverse Event Reporting System (VAERS). MedWatch is administered by the FDA and covers drugs, biologics, medical devices, and special nutritional products. VAERS is jointly administered by the FDA and the Centers for Disease Control and Prevention (CDC), and covers adverse events following immunizations. FDA regulations require manufacturers to report serious, unexpected adverse events within 15 days. Less serious reports can be submitted in periodic follow-up, or distribution reports.
- The FDA's cGMP regulations specify minimum standards for manufacturing facilities and their production controls. These regulations generally apply to both drugs and biologics, but the FDA has additional cGMP-related regulations that focus on biologics [23]. CBER has also tailored cGMP requirements for "specified biotechnology and synthetic biological products" to be as similar to drug requirements as possible.
- The FDA's lot release regulations allow the agency to require manufacturers to submit samples of any licensed biological products for testing [24]. Manufacturers must submit to CBER representative samples of each lot, a lot release protocol, and a summary of the test results. Lots may not be released until CBER authorizes an "official release." However, CBER does not require lot release in all circumstances.
- The FDA requires manufacturers to report certain changes in the product, production process, quality controls, equipment, facilities, personnel, or labeling that are established in the approved license application [25]. The manufacturer must demonstrate that the change does not adversely affect the identity, strength, quality, purity, or potency of the product that may affect the product's safety or effectiveness. FDA regulations and guidance categorize each change as "minor," "moderate," or "major" based on the risk to the product's quality, safety, and effectiveness. The FDA must give prior approval before the manufacturer can implement "major" changes. "Moderate" changes must be reported to the FDA within 30 days. Minor changes must be reported annually.
- The FDA may require, at the time of product approval, that the manufacture agree to conduct additional testing on its biological product, called phase 4 studies. These postmarketing studies may further evaluate the product's safety, efficacy, or manufacturing methods. Sponsors that agreed to conduct phase 4 studies as part of their BLA approval must update the FDA annually.

## 3 EUROPEAN UNION

### 3.1 How EU Law Defines a Biological Medicinal Product

In the European Union the regulation of biological products is subject to continuing review taking account of the evolving science and technology. Directive 87/22/EEC (now repealed) provided the first time in EU law the legal definition of a medicinal product developed by a biotechnological process. The following processes were considered as biotechnological: recombinant DNA technology, controlled expression of genes coding for biologically active proteins in prokaryotes and eukaryotes including transformed mammalian cells, hybridoma, and monoclonal antibody methods. This definition remains unchanged since 1987, and it is now used for defining a biotechnological medicinal product as set out in the Annex to Regulation (EC) 726/2004 [26], which repealed Regulation (EC) 2309/93 [27] governing the European centralized procedure.

The definition of a process based on biotechnology is sufficiently broad to capture a wide arrange of medicinal products, such as recombinant proteins and gene-based therapeutics, and prophylactics, such as gene transfer medicinal products and DNA vaccines. Medicinal products manufactured by biotechnological processes as defined in the Annex to Regulation (EC) 726/2004 must be authorized centrally pursuant to article 3 of the Regulation.

In June 2003 the European Commission adopted a new Annex I to Directive 2001/83/EC [28] on the EU code relating to medicinal products for human use. This new Annex was adopted in the form of Commission Directive 2003/63/EC [29]. The new Annex was adopted for implementation of the International Conference on Harmonization (ICH) Common Technical Document (CTD) format. Annex I sets out the particulars and documents accompanying an application for marketing authorization irrespective of the EU procedure used for obtaining a marketing authorization. Directive 2003/63/EC defines a biological medicinal product, and this definition consists of two essential elements. First, the active substance is a biological substance. A biological substance is a substance that is produced by or extracted from a biological source. Any one of the following source is considered as a biological source: microorganisms, organs and tissues of either plant or animal origin, cells or fluids (including blood or plasma) of human or animal origin, and biotechnological cell constructs utilizing cell substrates. If the product is produced from primary cells such as certain prophylactic vaccines, the product is considered a biological medicinal product. Second, the product requires for its characterization and the determination of its quality a combination of physicochemical-biological testing together with the production process and its control.

The Commission has indicated that the following are considered as biological medicinal products: immunological medicinal products and medicinal

products derived from human blood and human plasma. EU law defines an immunological medicinal product as any medicinal product consisting of vaccines, toxins, serums, or allergen products. Vaccines, toxins, and serums cover, in particular, agents used to produce active or passive immunity, and to diagnose the state of immunity. An allergen product means any medicinal product that is intended to identify or induce a specific acquired alteration in the immunological response to an allergizing agent.

Medicinal products derived from human blood or human plasma means those based on blood constituents that are prepared industrially by public or private establishments, such as albumin, coagulation factors, and immunoglobulins of human origin. This definition reflects the way plasma derived medicinal products are manufactured in the European Union. This class of products may be produced by privately owned industry or by public organizations that are owned by the member state.

### 3.2 Legal Foundation for Regulation of Biological Medicinal Product

The regulatory framework governing biological medicinal products is based on the European Community Treaty, which aims at the free movement of goods within the European Union. Although the legal base is built on the principle of free trade of medicinal products within the European Union, the essential aim of any rules governing the production, distribution, and use of medicinal products must be firmly based on protection of public health. Recital 3 of Directive 2001/83/EC notes that the objective of public health protection must be attained by means that do not hinder the development of the pharmaceutical industry or trade in medicinal products within the European Union.

The EU regulatory system is based on cooperation among the competent authorities of the member states (including the member states of the European Economic Area, e.g., Norway, Liechtenstein, and Iceland) and various relevant European institutions such as the European Commission and the European Medicines Agency (formerly called the European Agency for the Evaluation of Medicinal Products). The European Medicines Agency (EMEA) was formally established in 1995 by virtue of Regulation (EC) 2309/93, which is now replaced by Regulation (EC) 726/2004. The EMEA's role is narrowly defined in the Regulation as a body responsible for coordinating the existing scientific resources put at its disposal by member states for the evaluation, supervision, and pharmacovigilance of medicinal products. In practice, the scientific work is carried out by the member states through the EMEA's advisory committees and working parties.

The Committee for Medicinal Products for Human Use (CHMP) is one of the main committees responsible for preparing the opinion of the EMEA on any question relating to the assessment of medicinal products for human use. Pursuant to Regulation (EC) 141/2000 [30] the Committee for Orphan Medici-

nal Products (COMP) was established to provide scientific opinion on whether a medicinal product meets the criteria under EU law for it to be designated as an orphan medicinal product.

The sector-specific rules governing medicinal products are set out in various legal instruments and administrative guidance which follow the following hierarchy [31].

- A Regulation is directly applicable and binding in its entirety on in all member states. Therefore it does not require a period of transposition into the domestic laws of the member states.
- A Directive is directly effective that requires it to be transposed into domestic laws in order to give effect to the Directive. Under EU law, member states are only required to implement the Directive with respect to its objectives and EU law does not control the manner and form of how a Directive is transposed into the national laws.
- A Decision is binding in its entirety upon persons to whom it is addressed.
- Opinion is not legally binding.

In addition the Commission has issued a number of technical and administrative guidelines such as those set out in various volumes of the Notice to Applicants in order to explain how EU law can workably put into practice. The EMEA has developed a body of scientific guidelines regarding the technical requirements for addressing issues relating to safety, quality and efficacy. Although guidelines are not legally binding, the European Courts have increasingly relied on such documents as an aid in interpretation of the legal requirements.

***Research and Development*** Clinical development in the European Union is regulated by Directive 2001/20/EC [32], which is commonly known as the Clinical Trials Directive. This Directive regulates all stages of clinical development in the European Union, including Phase I clinical studies involving healthy volunteers. The competent authorities of the member states are responsible for assessing the applications for clinical trial authorization. In assessing whether an approval should be granted, the competent authorities are required to ensure that conduct of the clinical trials comply with the principles of good clinical practice and the relevant ethical principles. Reference is made to the principles set out in the ICH E6 Guideline on Clinical Practice and the applicable version of the Declaration of Helsinki. Under the Clinical Trials Directive, competent authorities are required to make a determination of an application for clinical trial authorisation within 60 days from the date of submission. However, the Directive permits the member states to extend the statutory time limit for certain investigational medicinal products such as gene therapy, xenotransplantation, and products that are derived from biological source.

Compliance with the requirements of Directive 2001/20/EC is important. This is because Annex I to Directive 2001/83/EC (as amended) expressly requires that for the purpose of obtaining a marketing authorization, all clinical trials conducted within the European Union to comply with Directive 2001/20/EC. For clinical trials conducted outside the European Union and the data of which are used in support of an application for a marketing authorization, such clinical trials must be designed, implemented, and reported on the basis of principles that are equivalent to the provisions of Directive 2001/20/EC and carried out in accordance with the ethical principles that are reflected in the Declaration of Helsinki.

EU law expressly requires nonclinical (pharmacotoxicological) studies to be carried out in conformity with the provisions related to good laboratory practice set out in Directive 2004/10/EC [33] and Directive 2004/9/EC [34] on the inspection and verification of laboratory practice.

***Approval*** Approval process for biological medicinal products is the same as other chemically synthesized small molecules. The legal test is firmly based upon an assessment of risk–benefit balance. However, in assessing risk–benefit balance of a biological medicinal product, EU law requires the applicant to provide certain additional information. If the medicinal product contains a new biological active substance, the applicant must comply with the requirements set out in article 8(3) of Directive 2001/83/EC by providing results of the pharmaceutical and nonclinical testing, and clinical trials.

Given the safety and efficacy of a biological medicinal product is determined largely by the starting material used and the process, EU law requires applicants to describe and document the origin and history of starting materials. Starting materials mean any substance of biological origin such as microorganism, organs, and tissues of either plant or animal origin; cells or fluids of human or animal origin; and biotechnological cell constructs, including the cell substrates whether or not they are recombinant. Moreover, for medicinal products that are manufactured based on a cell bank system, the cell characteristics must be shown to have remained unchanged at the passage level used for the production and beyond. It is also a requirement to test all materials used in the process for adventitious agents, including animal spongiform encephalopathy agents. If the presence of potentially pathogenic adventitious agents is inevitable, the corresponding material must only be used when further processing is demonstrated to be capable of eliminating and/or inactivating such adventitious agents. The capability of the process must be validated. In comparison with chemically synthesized products, greater emphasis is placed on the in-process controls to ensure batch to batch consistency.

With respect to preclinical testing, EU law expressly states the testing program must be adapted for individual products. It is for the applicant for a marketing authorization to justify the testing program to elucidate the preclinical safety and biological activity of the product. EU law states that in establishing the testing program, the following must be taken into account:

- All tests requiring repeated administration of the product must be designed to take account of the possible induction of, and interference by antibodies.
- Examination of reproductive function, of embryo-fetal and perinatal toxicity of mutagenic potential and of carcinogenic potential must be considered. However, constituents other than the active substance(s) are incriminated, validation of their removal may replace the study.
- The toxicology and pharmacokinetics of an excipient used for the first time in the pharmaceutical field must be investigated.
- Where there is a possibility of significant degradation during storage of the medicinal product, the toxicology of degradation products must be considered.

In general, according to EU law, applicants are expected to carry out controlled clinical trials, randomized and as appropriate against a placebo and an established medicinal product (an active comparator) of proved therapeutic value. However, applicants may justify use of other trial design. The treatment of the control groups will vary from case to case and also will depend on ethical consideration and therapeutic area. In some cases it may be more justified to compare the efficacy of a new medicinal product with that of an established medicinal product of proved therapeutic value rather than with a placebo.

In the new European pharmaceutical legislation, a new regulatory path has been created for approval of a similar biological medicinal product, which is commonly known as a biosimilar medicinal product. The definition of a similar biological medicinal product as set out in article 10(4) of Directive 2001/83 (as amended by Directive 2004/27/EC) [35] to mean a biological medicinal product that is similar to a reference product does not mean the conditions in the definition of generic medicinal products owing to, in particular, differences relating to raw materials or differences in manufacturing processes of the biological medicinal product and the reference biological medicinal product. In such cases the results of appropriate nonclinical tests and/or clinical trials relating to these conditions must be provided. The EMEA has developed has now developed a series of technical guidelines to address the type and quantity of supplementary data to be provided [36].

***Marketing and Postapproval Requirements*** Regardless of whether a medicinal product is considered a conventional pharmaceutical product or a biological product, after grant of an approval, the holder of the marketing authorization is required to monitor the continuing risk–benefit balance of the product. In relation to the method of manufacture, in addition to ensuring compliance with good manufacturing practice in accordance with Directive 2003/94/EC [37] the authorization holder must take account of scientific and technical progress and introduce any changes that may be required to enable

the product to be manufactured and controlled by means of generally accepted scientific methods. Immunological products such as vaccines and products derived from human blood or plasma may be subject to official batch release testing at the request of a competent authority.

In the new pharmaceutical legislation, greater emphasis is placed on pharmacovigilance and risk management. Indeed, at pre-approval, applicants are required to submit to the competent authority a detailed description of the pharmacovigilance and of the risk amangement system that the applicant will introduce. In general, it is a requirement for the marketing authorization holder to record all suspected serious adverse reactions and to report them promptly to the competent authority within a defined time frame as set out in EU law. For new products, the marketing authorization holder is required to provide periodic, updated safety report at least every six months during the first two years following the initial placing on the market and once a year for the following two years. Thereafter the reports must be submitted at three yearly intervals or immediately on request by the competent authority. The Commission has developed a guidance document published in volume 9 of the Rules governing medicinal products in the European Union, which is currently being revised. This guidance takes account of various guidelines promulgated under the International Conference on Harmonization.

In order to establish and maintain a pharmacovigilance system, the holder of the marketing authorization is required to have permanently and continuously at his or her disposal an appropriately qualified person. This qualified person is personally responsible for ensuring that information about all adverse reactions that are reported to the personnel of the company and to medical representatives is collected and collated. This qualified person is also responsible for ensuring that all suspected serious adverse reactions are reported to the competent authority concerned.

## 4 JAPAN

On 25 July 2002, the Japanese House of Representatives passed the revised Pharmaceutical Affairs Law (PAL), which dates back to 1943. Although amendments have been made in the 1940s, 1960s, and 1970s, certain parts of the legislation required updating to take account of changes in science and technology, and the need for the liberalization of the Japanese market. Companies can now outsource the manufacturing process, allowing pharmaceutical companies to market their products in Japan without operating their own production facilities. Changes made to the PAL have also fueled growth in the clinical trial sector in Japan. Notwithstanding the revision of the PAL, the basic purpose of the law remains intact in that it is designed to protect public health by ensuring the safety, quality, and efficacy of medical products in Japan.

The revised PAL of 2002 aims at addressing the following challenges:

- The need to strengthen the safety measures related to medical devices
- The need to strengthen regulatory control over products based on biotechnology and genomic technology
- The need to strengthen the postmarketing safety monitoring and take account of the efforts in international harmonization for the technical assessment of pharmaceutical products

A new regulatory agency, the Pharmaceuticals and Medical Devices Agency (PMDA), has been created with the executive function of overseeing regulation of pharmaceutical products and medical devices. Similar to the system adopted by the United States and the European Union, the regulatory control of medicinal products in Japan is through a system of approval/licensing for certain regulated activities such as conduct of clinical trials, manufacture, marketing, distribution, sale, and supply of specific pharmaceutical products. The Japanese Pharmacopoeia is an integral part of the regulatory framework as it sets out the quality standards for certain pharmaceutical products or substances.

Japan is a party to the tripartite ICH process. Therefore all the adopted guidelines have been implemented as the basic technical standards for the evaluation of safety, quality, and efficacy for pharmaceutical and biotechnological products. In addition the requirements for good laboratory practices for conducting nonclinical safety testing of pharmaceutical products have been applied since the 1980s in the form of administrative instruction. The requirements for conducting clinical trials in accordance with good clinical practice have been implemented since 1990. Such standards have now been enforced through various ministerial ordinances.

The PAL sets out the broad legislative framework for regulating medical products. However, the Ministry for Health, Labor, and Welfare (MHLW) has the authority to issue ordinances and notifications setting out the details for regulating certain product types, such as pharmaceutical products, medical devices, in vitro diagnostic reagents, cosmetic and quasi-drugs. For example, Ministerial Ordinance No. 136, 2004 sets out standards for quality assurance for drugs, quasi-drugs, cosmetics, and medical devices. This Ordinance seemingly applies to drugs based on cells or tissues. The basic structure of this ordinance reflects the principles of good manufacturing practice where a focus is placed on the quality management system, quality control, personnel, training, documentation, and self-inspection.

Given that global trade and international harmonization are key to the development of a sustainable life sciences industry, closer international cooperation is key to tackling technical barriers to trade in medicines. In addition to the ICH process, increasingly regulatory authorities have entered into agreements to enable them to exchange confidential information about approval and safety of medicines.

In February 2007 the European Commission and the EMEA signed a confidentiality agreement with the MHLW and PMDA to enable both parties to

exchange confidential information relating to all legislation and guidance documents, postapproval pharmacovigilance, scientific advice, orphan designation, good clinical practices inspections, and so forth [39]. The US FDA has established a similar confidentiality arrangement with the European Commission and the EMEA.

## 5 CONCLUSION

This chapter introduces the legal and regulatory aspects pertaining to biological products in the United States and in the European Union. The regulatory laws in these two jurisdictions distinguish between conventional pharmaceutical products based on small molecules and biological products. While the legal test for regulatory approval is firmly based on an assessment of risk–benefit balance, the approach to such an assessment is distinctly different with respect to biological products. This is exemplified by the publication of a recent scientific review commissioned by the health ministers following serious adverse reactions that occurred in a first-in-human clinical trial involving a monoclonal antibody TGN 1412 at Northwick Park Hospital in London (March 2006). Six healthy volunteers experienced severe systemic adverse reactions after administration with the biological product. The adverse reactions were characterized as associated with cytokine release. The report emphasizes the importance of performing appropriately conducted preclinical studies in identifying the safe starting dose in humans. The report appears to accept that the conventional approach based upon NOAEL (no observed adverse effect level), a concept that is generally applicable to chemically synthesized small molecules, may not be appropriate. Instead, the principle based on minimal anticipated biological effect level (MABEL) is a good model for defining the safe starting dose, taking account of the novelty of the agent, its biological potency, its mechanism of action, the degree of species-specificity of the agent, the dose–response curves in vitro and in vivo. This concept is also articulated in the FDA guidance in consideration of lowering the starting dose based on a variety of factors that include the pharmacologically active dose (PAD) [38].

## REFERENCES

1. FDA, Center for Biologics Evaluation and Research (CBER). Frequently Asked Questions. http://www.fda.gov/cber/faq.htm
2. CBER. Science and the Regulation of Biological Products: From a Rich History to a Changing Future 6 (2001). http://www.fda.gov/cber/inside/centscireg.htm
3. Public Health Service Act §351(i), 42 USC §262(i).
4. 21 CFR §600.3(h).
5. Id. at §600.3(h)(5).

6. CBER. Frequently Asked Questions; FDA, Selected Guidance Documents Applicable to Combination Products. http://www.fda.gov/oc/combination/guidance.html
7. Mathieu M, ed. Biologics Development: A Regulatory Overview. 2004. See also Pub. L. No. 57-244, 32 Stat. 728, 1902. (An Act to Regulate the Sale of Viruses, Serums, Toxins, and Analogous Products in the District of Columbia, to Regulate Interstate Traffic in Said Articles, and for Other Purposes).
8. Federal Food, Drug, and Cosmetic Act §321(g)(1), 21 USC §201(g)(1).
9. Public Health Service Act §351, 42 USC §262.
10. Kracov DA. FDA's role in regulating biologics. In Pines WJ, ed. *FDA: A Century of Consumer Protection*. 2006, 195.
11. Buday PV. Fundamentals of United States biological product regulations. *Regul Affairs* 1991;3:223.
12. Korweck EL. Human biological drug regulation: Past, present, and beyond the year. *Food Drug LJ* 1995;50:123.
13. 21 CFR §312.22.
14. For example, 21 CFR part 50; 45 CFR part 46.
15. 21 CFR §50.25.
16. See 21 CFR part 56.
17. 21 CFR §56.107.
18. FDCA §505, 21 USC §355; 21 CFR part 312.
19. 21 CFR §601.2.
20. Beers DO. FDA regulation of biologic drugs. In *Generic and Innovator Drugs: A Guide to FDA Approval Requirements §13.02*, 6th ed. 2004.
21. 21 CFR parts 210, 211, and 600.
22. 21 CFR §600.80.
23. For example, 21 CFR parts 600, 601, 606, 607, 610, 630, 640, 660, 680.
24. 21 CFR §610.2(a).
25. 21 CFR §§610.12, 610.14.
26. Regulation (EC) No 726/2004 of the European Parliament and of the Council of 31 March 2004 laying down Community procedures for the authorization and supervision of medicinal products for human and veterinary use and establishing a European Medicines Agency. OJ L136, 31.03.2004, p1.
27. Council Regulation (EEC) No 2309/93 of 22 July 1993 laying down Community procedures for the authorization and supervision of medicinal products for human and veterinary use and establishing a European Agency for the Evaluation of Medicinal Products. OJ L214, 24.08.1993, p1.
28. Directive 2001/83/EC of the European Parliament and of the Council of 6 November 2001 on the Community code relating to medicinal products for human use. OJ L311, 28.11.2001, p67.
29. Commission Directive 2003/63/EC of 25 June 2003 amending Directive 2001/83/EC of the European Parliament and of the Council on the Community code relating to medicinal products for human use. OJ L159, 27.06.2003, p46.
30. Regulation (EC) No 141/2000 of the European Parliament and of the Council of 16 December 1999 on orphan medicinal products. OJ L18, 22.01.2000, p1.

31. Article 249 of the Treaty Establishing the European Community.
32. Directive 2001/20/EC of the European Parliament and of the Council of 4 April 2001 on the approximation of the laws, regulations and administrative provisions of the member states relating to the implementation of good clinical practice in the conduct of clinical trials on medicinal products for human use. OJ L121, 01.05.2001, p34.
33. Directive 2004/10/EC of the European Parliament and of the Council of 11 February 2004 on the harmonization of laws, regulations, and administrative provisions relating to the application of the principles of good laboratory practice and the verification of their applications for tests on chemical substances (codified version). OJ L50/44, 20.02.2004, p44.
34. Directive 2004/9/EC of the European Parliament and of the Council of 11 February 2004 on the inspection and verification of good laboratory practice (GLP) (codified version). OJ L50, 20.02.2004, p28.
35. Directive 2004/27/EC of the European Parliament and of the Council of 31 March 2004 amending Directive 2001/83/EC on the Community code relating to medicinal products for human use (text with EEA relevance). OJ L136, 30.04.2004, p34.
36. Tsang L, Beers D. Follow-on biological products: the regulatory minefield. In *Global Counsel Handbooks Life Sciences 2004/05*. Practical Law Company.
37. Commission Directive 2003/94/EC of 8 October 2003 laying down the principles and guidelines of good manufacturing practice in respect of medicinal products for human use and investigational medicinal products for human use (text with EEA relevance). OJ L262, 14.10.2003, p22.
38. FDA Guidance for Industry Estimating the Maximum Safe Starting Dose in Initial Clinical Trials for Therapeutics in Adult Healthy Volunteers (July 2005).
39. Press Release, Brussels 5 February 2007. Closer ties on medicines safety between EU and Japan (IP/07/135).

# Methods of Production of Biopharmaceutical Products and Assessment of Environmental Impact

PATRICIA D. WILLIAMS, PhD

**Contents**

1 Introduction
2 Host Systems for Biopharmaceutical Production
3 Guidelines for the Production and Acceptance of Recombinant Proteins in Biologic Systems
   3.1 Manufacturing Criteria for Cellular Systems
   3.2 Manufacturing Criteria for Transgenic Animals
4 Criteria for Release of Recombinant Protein Products
   4.1 Physiochemical Properties
   4.2 Biologic Activity
   4.3 Immunochemical Properties
   4.4 Purity and Potential Impurities
   4.5 Product Content
   4.6 Products Made Transgenically
5 Environmental Assessment and Categorical Exclusions
6 Demonstration of Comparability When a Manufacturing System Is Changed
7 Application of Comparability Criteria to a New Manufacturing Technology
8 Summary and Conclusions
References

## 1 INTRODUCTION

The genomics revolution over the past 20 years and subsequent sequencing of the human genome have provided opportunities for the pharmaceutical

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

industry in terms of drug targeting and identification of novel therapeutics. Advances in molecular biology and genetics have also played significant roles in the pharmaceutical development of these novel products. Specifically, the biomanufacture of protein therapeutics has relied on genetically engineered production systems. Biomanufacturing has evolved out of heterologous gene expression technology of the 1980s and 1990s and continues to evolve. From bacterial hosts to yeast to mammalian cells to even transgenic plants and animals, novel approaches have been developed to increase the capacity, efficiency, and safety of biopharmaceuticals.

In parallel, protein purification and characterization techniques have also markedly improved, increasing the yields and speeding analysis time for biologically active proteins. With this developing technology, biopharmaceutical production has been able to diversify from bacterial cells to yeast to mammalian cells and now into transgenic "bioreactors" (e.g., goats, cattle). Through an understanding of these novel production systems, the advantages of this new technology in contributing to the successful production of novel biopharmaceutical products can be realized.

At present, over 20 therapeutics produced via recombinant technology are currently on the market [1]. These products range from monoclonal antibodies to interferons (IFNs) to growth factors to human insulin. While in many cases relatively small-scale production in bacterial, yeast, or cell culture systems are amenable for proteins required in microgram quantities, new products are being developed requiring significantly higher doses and long-term administration. Thus the need for higher capacity manufacturing systems has arisen. In particular, protein production in transgenic animals has become the method of scale-up or a necessary choice for some protein therapeutics. Production in milk of transgenic sheep, goats, or cattle may yield between 5 and 30 g/L of recombinant protein, which is far beyond the production capacity of mammalian culture systems [2].

In 1969 the National Environmental Policy Act (NEPA) was enacted. This act requires that all federal agencies assess the environmental impacts of their actions and ensure that the interested and affected public is informed of environmental analyses [3]. The Food and Drug Administration (FDA) is required under the NEPA to consider the environmental impact of approving drug and biologic applications as an integral part of its regulatory process. The FDA's regulations in 21 CFR part 25 specify that environmental assessments (EAs) must be submitted as part of certain new drug applications (NDAs), abbreviated applications, applications from marketing approval of a biological products (BLAs), supplements to such applications, investigational new drug applications (INDs), and for various other actions (see 21 CFR 25.20), unless the action qualifies for categorical exclusion [3].

The issue of product comparability when manufacturing processes or systems are changed, in particular, poses an important challenge to the pharmaceutical and biotechnology industry. For example, how should a sponsor demonstrate that proteins secreted into the milk of transgenic cattle or goats

are comparable with those produced in *E. coli* or CHO cell systems? What methodologies should be employed to demonstrate product comparability? The need for analytical studies, in vitro testing, preclinical animal studies, or new clinical data will depend on the significance of the change in the system and the potential effect on the biologic product.

The systems available for production of biopharmaceuticals, and the approaches to introducing new technologies into the drug development process are discussed in the sections that follow.

## 2 HOST SYSTEMS FOR BIOPHARMACEUTICAL PRODUCTION

An overview of the manufacturing systems employed for the production of protein therapeutics is presented in Table 1. The advantages and disadvantages of the various manufacturing systems, as well as examples of therapeutic proteins currently marketed or under development utilizing these systems, are also presented.

One of the earliest and simplest expression systems for biopharmaceutical production is the eukaryotic bacteria, *E. coli*. The biology of *E. coli* is well understood, and bacteria are simple to grow and genetically stable. *E. coli* can accumulate extremely high concentrations of exogenous proteins in their cytoplasm (up to 20% of their total cellular protein), and can translocate proteins from the cytoplasm to the periplasm [4]. Although *E. coli* has proved to be a suitable host for a number of therapeutically useful proteins including IFNs, growth factors, and tissue plasminogen activator (tPA), *E. coli* cannot perform many of the same posttranslational modifications achievable in mammalian cells, and hence their utility is limited. Moreover many proteins expressed in *E. coli* will form insoluble inclusion bodies, and purification may be complicated [5].

Yeast strains such as *Saccharomyces cerevesia* and *Pichia pastoris* have also proved useful for expression of biotherapeutics. Yeast are rapidly growing eukaryotic cells, and they perform many of the same secondary protein modifications as mammalian cells. Yeast are inexpensive to grow and maintain, and can also secrete recombinant proteins directly into the culture media. This makes yeast very suitable for large-scale production of recombinant human proteins. *Pichia pastoris* are specifically capable of expressing correctly folded, secreted proteins, including those containing high levels of disulfide bonding. Secreted proteins can often be easily and efficiently purified from the fermentation media, often with only a single chromatographic step. For example, recombinant insulin was among the first proteins to be produced in large quantities in yeast [6]. While yeast do serve as excellent hosts for expression of recombinant proteins, they still do not perform all functions similar to mammalian cells, and can be extremely sensitive to temperature, aeration, and methanol concentration [7]. Marketed recombinant proteins produced in yeast include human insulin (Novolin®), granulocyte macrophage colony stimulating

**TABLE 1 Expression systems for biopharmaceutical production**

| System | Advantages | Disadvantages | Examples |
|---|---|---|---|
| Bacterial (*E. coli*) | • High growth rate<br>• Simple nutritional demands<br>• Genetic consistency<br>• Relative ease of purification<br>• High protein yield<br>• Well-defined biology<br>• Approved products | • Inability to properly fold, glycosylate, phosphorylate, and acylate proteins<br>• Difficulty in isolating proteins from inclusion bodies<br>• Rapid doubling time potential for mutation | • **Betaseron® (interferon β-1b)**<br>• **Trastuzumab® (monoclonal antibody)**<br>• **Nutropin® (human growth hormone)**<br>• **Numega® (interleukin-11)** |
| *Saccharomyces Cerevisiae* (yeast) | • Eukaryotic<br>• Rapid growth, low cost<br>• Ease of genetic manipulation<br>• Protease-deficient strains<br>• Secreted or intracellular protein products<br>• Approved products | • Improper posttranslation modification<br>• Temperature, $O_2$ sensitivity<br>• Differences in codon usage from mammalian cells<br>• Improper or excessive glycosylation | • **Engerix-B® (HBV vaccine)**<br>• **Novolin-B® (insulin)**<br>• **Humulin® (insulin)**<br>• **Leukine® (granulocyte macrophage colony stimulating factor)** |
| *Pichia pastoris* (yeast) | • Rapid growth, low cost<br>• Ease of genetic modification<br>• Protein folding capabilities<br>• Direct protein secretion into media<br>• Relative ease of purification | • Potential codon bias<br>• Posttranslation modification<br>• Sensitivity to aeration, methanol concentration, and temperature | • Influenza neuraminidase<br>• Botulinum neurotoxin A binding domain<br>• CC49 single-chain variable fragments<br>• Bovine tumor necrosis factor receptor I<br>• Insulin |
| Mammalian cell | • Many cell types available<br>• Posttranslational modification<br>• Growth in selectable markers<br>• Both suspension and adherent cultures<br>• Inducible expression<br>• Cytoplasmic, membrane, and secreted proteins produced<br>• Proper RNA splicing<br>• Approved products | • Requirement for adherent growth<br>• Cost of scale-up<br>• Time to develop well-characterized stable cell lines<br>• Lower levels of expression<br>• Limited passage number<br>• Low yields relative to cost of production<br>• Potential under sialylation of products | • **Retavase® (tPA)**<br>• **Epogen® (Erythropoeitin)**<br>• **Benefix™ (coagulation factor IX)**<br>• **Avonex™ (interferon β-1a)**<br>• **Pulmozyme® (DNAse I)**<br>• **Reopro® (monoclonal antibody)** |

| | | | |
|---|---|---|---|
| Baculovirus | • Eukaryotic—similar protein processing, modifications to mammalian cell culture<br>• Ease of suspension culture<br>• Soluble intracellular proteins<br>• Large viral genome can accommodate large proteins<br>• Noninfectious to mammalian cells | • Insect cell system—potentially different processing<br>• Inefficient splicing of large products<br>• Nonmammalian system<br>• Requires characterization of both cells and virus stocks<br>• Distinct glycosylation patterns | • HIV-1 gp160 IIIB<br>• Carcinoembryonic antigen<br>• Influenza vaccine<br>• Hemagglutinin<br>• Gp 160 (HIV protein) |
| Transgenic plant systems | • Low cost of raw materials<br>• Able to assembly of multimeric proteins<br>• Time to scale-up<br>• Safety (plants are not hosts for human pathogens)<br>• Efficient production of secretory IgAs (SigAs) in tobacco | • Concerns with posttranslational modifications including acylation, glycosylation, phosphorylation, and plant proteases<br>• Susceptibility to drought, elements<br>• Incompatibility of mammalian proteins in plant cells<br>• Presence and characterization of plant impurities in final products<br>• Use of pesticides/herbicides | • CaroRx™ (antistreptococcus)<br>• IgG (phase 1)<br>• *E. coli* LT (phase 1) |
| Transgenic animal systems | • Substantially-increased yields<br>• Decreased long-term costs<br>• Ease of purification from milk<br>• Variety of production species (goat, sheep, cattle)<br>• Mammary specific expression<br>• Ability to utilize both genomic and cDNA clones for effective expression | • Time to full scale-up<br>• Animal to animal variation<br>• Variations in feeding and milk production<br>• Controlling for animal illness | • Alpha-1-antitrypsin (sheep)<br>• Fibrinogen (sheep)<br>• Protein C (porcine)<br>• Antithrombin III (goat)<br>• Monoclonal antibody (goat) |

*Note*: **Products in bold are on the market [1]**

factor (Leukine®), and hepatitis B vaccine *(Engerix®)* [1]. Other products currently under study produced in yeast include influenza neuraminidase, botulinum neurotoxin A binding domain, and monoclonal antibody variable fragments [8].

Recombinant proteins can also be expressed in mammalian cell systems including Chinese hamster ovary (CHO) and baby hamster kidney (BHK) cells. While the initial bioreactor setup for these mammalian systems can be more costly than for either bacterial or yeast systems, mammalian cells can still produce large quantities of posttranslationally modified, viable recombinant product [2]. Indeed the majority of biopharmaceutical products marketed today are produced in mammalian cells, primarily CHO cells (Table 1). Mammalian cell approaches are often required when protein function is heavily dependent on posttranslational modifications, including glycosylation, phosphorylation, or acylation. Many protein biopharmaceutical products including DNAse I (Pulmozyme®), tissue plasminogen activator (Retavase®), erythropoietin (Epogen®), and FSH (Follistem®) can only be produced in their native structure in mammalian cells. Many mammalian cell systems are also highly inducible, meaning that expression of the desired product can be tightly controlled and highly stimulated. Methotrexate, for example, was used to select for high levels of dihydrofolate reductase (DHFR) expression in CHO cells producing recombinant tPA [2]. Among other advantages of mammalian cell systems are the numbers of distinct, specialized cell types available, growth of some cell types in suspension culture, and the ability of host cells to sequester proteins into specific cell compartments or secrete proteins directly into the culture media [9].

Despite these many advantages mammalian cells do not possess all the necessary enzymatic machinery to properly posttranslationally modify every recombinant protein. In many cases mammalian cell lines lack the enzymes necessary for proper sialylation of proteins [2]. While it is possible to stably introduce these sialyltransferases into the host cell, expression of these enzymes normally decreases with time and adds an extra variable to an already complicated cellular process. Another challenge with mammalian cells is that they often require a solid matrix for adherent growth. While many cells can be grown in suspension, others such as CHO must utilize large-scale spinner culture or microbead technology to achieve maximum yields. In many cases expression and growth of mammalian cells is not indefinite. Most cells exhibit a limited passage number, decrease their expression of recombinant protein, and cease to grow. Thus well-developed master cell banks and working cell banks must be constantly maintained in the event of contamination or loss of the culture [2]. Additionally mammalian cell culture is significantly more costly than bacterial culture with respect to growth media components such as serum and antibiotics.

Recombinant human proteins may also be expressed in baculovirus, such as *Spodoptera frugiperd*a (SF9) cells. These viruses possess large genomes capable of accommodating extremely large exogenous DNA sequences.

Numerous shuttle vectors are available that facilitate convenient transfer from bacteria or mammalian cells. Proteins such as HIV Gp 160, carcinoembryonic antigen, and a form of influenza vaccine have successfully been expressed in baculovirus systems [9]. Potential disadvantages of this system include inappropriate posttranslational modifications, use of insect rather than mammalian cells, and characterization of both the baculovirus and SF9 insect cells used.

In the future, neoplastic cells including spontaneous or virus-transformed cells, or other immortalized cell lines, may also serve as a preferred cell substrate for recombinant protein production. The use of malignant cells as host systems for recombinant proteins represents a natural extension of using neoplastic cells to produce purified biologicals, including interferons and monoclonal antibodies [10]. Obvious concerns with these agents as host systems for vaccine production include contamination with viable tumor cells or other adventitious agents. Contamination from residual DNA or biologically active proteins is also a possibility. Finally tumors are by nature heterogeneous, often consisting of numerous cell types exhibiting several distinct phenotypes and genotypes. Controlling for this heterogeneity represents an additional challenge for development of vaccines from neoplastic cell systems [10].

Transgenic plants might also be used for production of recombinant proteins. While many of these systems are still early in the development stage, plants offer very robust and high-capacity system for biopharmaceutical production. Since plants cannot always properly modify proteins as mammalian cells, their utility may be limited. However, early studies suggest that plants may serve as useful hosts for production of vaccines and antibodies. For example, HepB surface antigen has been successfully expressed in potatoes [11], and clinical trials are underway with secretory antibodies (SIgAs), such as CaroRx™, developed in plants [12].

Expression in transgenic animals, particularly dairy cattle, sheep, and goats, is a viable option for achieving large-scale recombinant protein production. Unlike bacterial, yeast, or mammalian cell systems, transgenic animal systems can express recombinant proteins in the milk at quantities up to 30 g/L. In addition to vastly increased capacity, milk provides a relatively simple matrix for purification [2]. Monoclonal antibodies, for instance, have successfully been expressed in transgenic mice and goats [13]. Furthermore transgenic animals are also capable of producing complex proteins folded with many of the posttranslational modifications such as glycosylation, amidation, and gamma-carboxylation [14].

As with any other biological system, transgenic animals are still susceptible to those adventitious agents capable of affecting bacterial, mammalian cell, or yeast systems. Moreover, unlike cells and bacteria, a somewhat greater degree of intraindividual genetic variation may be observed with a transgenic herd. Variability in food intake and/or milk production also represent points to consider in the development of transgenic animal populations. Characteristics of these various expression systems and examples of recombinant proteins produced using these technologies are described in Table 1.

## 3 GUIDELINES FOR THE PRODUCTION AND ACCEPTANCE OF RECOMBINANT PROTEINS IN BIOLOGIC SYSTEMS

As is the case for any new chemical entity, the FDA, the CPMP, and the International Conference on Harmonisation of Technical Requirements for Pharmaceuticals (ICH) have developed standards and guidelines for the production and characterization of recombinant biological products. In the subsequent section we will discuss some of the standards for manufacturing, release and finally, comparability of biologics produced in conventional versus transgenic systems. For reference, one may consult the ICH Tripartate Guidelines on Test Procedures and Acceptance Criteria for Biotechnological/Biological Products [15], and the FDA Points to Consider in the Manufacturing and Testing of Therapeutic Products for Human Use Derived from Transgenic Animals [16], the CPMP's Guideline on Comparability of Medicinal Products Containing Biotechnology-Derived Proteins as Active Substance [17], and the Points to Consider in the Characterization of Cell Lines Used to Produce Biologicals [18].

The FDA and ICH both convey messages that biologic manufacturers must thoroughly characterize all aspects of their manufacturing process, including the starting raw materials, strains of microbials or animals used, growth conditions, methods of introducing the novel genes into the host system, methods of isolation of the product, product characterizations including all analytical methods utilized, batch acceptance and rejection criteria, storage, expiration, and shipping. Regardless of the source of the recombinant protein, these elements must be addressed to ensure the safety and efficacy of the final product.

### 3.1 Manufacturing Criteria for Cellular Systems

For cellular systems (bacterial, mammalian cell-based, etc.) manufacturers must describe the origin of the cell lines used, the general characteristics, and genotype (if known) of the master and working cell lines, including any known genetic markers and the source tissue (if known). Growth conditions, including all components of the culture media, antibiotics, and/or growth factors should also be described. Production, characterization, and storage of both the master and working cell banks should be discussed. Master cell banks should be characterized for morphology, species of origin, split ratio, functionality, and identity. Working cell banks should be routinely characterized for contamination, sterility, and presence of viruses. Composition and source of the culture media and any other additional factors (serum, antibiotics) should be described.

Manufacturers should include reasonable detail about the source of the gene to be expressed, including, if known, the complete DNA sequence and any regulatory elements, the method of introduction of the recombinant gene into the host cell, and methods for identification of the transformed host.

Manufacturers should also describe all reagents utilized, selection criteria, methods for screening for recombinant protein and monitoring for stability or maintenance of gene expression (periodic testing). Production conditions should also be described, including the length of collection, the total batch size (cell number or volume), and frequency of recovery. Isolation procedures, including centrifugation techniques, chromatographic, or other biochemical techniques required for product purification, should also be described and validated for each biologically produced material. Finally, stability of the product should be determined, and storage conditions for the purified material and standards for acceptance or rejection should also be well defined [19].

### 3.2 Manufacturing Criteria for Transgenic Animals

The FDA and EU have also established specific guidelines for the manufacture and testing of therapeutic products derived from transgenic animals [16]. The FDA has specified its expectations on the use of transgenic animals for production of biopharmaceuticals, including relevant scientific points that should be addressed in any subsequent regulatory submission. Guidelines for biologic production in transgenic animals essentially follow those for cell-based systems.

Similar to cell-based systems, it is expected that sponsors provide detailed information on their genetic construct, including description of the native protein, the genetic structure of the transgene, any cloning and purification methods used, and any vectors such as yeast artificial chromosomes (YACs). The sponsor's strategy to create the final transgene construct should also be described. Specific DNA elements within the transgene, including enhancers, promoters, repressors, or other control regions, should be discussed, particularly if they are to have a planned effect on expression of the transgene.

The manufacturer should further describe production of the initial founder animals, including the history of the animals donating the gametes. Like master cell banks, transgenic production animals should have detailed veterinary records. The method of introduction of the transgene into the recipient animal(s) should be described, as well as characterization of methods used to assess transgene expression. Again, acceptance criteria should be established for presence of the transgene and to confirm expression of the transgene within desired limits in the transgenic animals. Also any deleterious effect of transgene expression on the animal's health should be duly noted.

Next, the transgenic animals must be monitored for stability of the transgene. This should continue through several rounds of germ-line passage using DNA analytical methods such as Southern blots. Then, stability of expression of the transgene must be established. While the transgene may be present in the transgenic animal, expression must be maintained. Unlike cell lines, transgenic animals cannot be stored indefinitely. Thus approaches must be developed to ensure that the transgenic product remains available for an extended period of time. This can be likened to the use of master cell banks (MCBs)

and working cell banks (WCBs). These might be termed Master Transgenic Bank and Master Working Transgenic Bank, respectively.

Breeding techniques and animal husbandry methods should be described, including addition or elimination of animals from the production herd. Animal feed and housing should be thoroughly described. Standards for registered facilities may be found in the *NIH Guide for the Care and Use of Laboratory Animals* [20].

Last, purification and characterization of the transgenic product should be adequately described. Procedures such as milking, exsanguination or extirpation of tissues should be detailed so as to "maximize safety, sterility, potency, and purity of the product." Although collection of the biologic material may not occur under sterile conditions, the area should be as clean as possible. The actual production lot should be well defined, and those tests found to be sensitive to potential changes in the product should be incorporated into the lot release protocol. Naturally, products derived from transgenic animals (milk, blood urine, semen, etc.) will have a unique set of concerns compared to a product derived from asceptic cell culture. As is the case for the ICH guidelines, the FDA expects that any biopharmaceutical product derived from transgenic animals be thoroughly characterized with respect to safety, purity, and potency. "The manufacturer should describe all tests performed for in-process control and final product acceptability." This is particularly important in transgenics, since unlike a relatively homogeneous cell population, a transgenic herd may consist of numerous unique individuals with material harvested at different times of the year in different temperatures. However, it is not very different from using blood donations from unique human individuals for composition of the starting pool for purification of plasma products.

## 4 CRITERIA FOR RELEASE OF RECOMBINANT PROTEIN PRODUCTS

The ICH expectations are that "acceptance criteria" or limits of acceptability will be established prior to final product characterization, regardless of the host cell system. More important, with respect to system, the final product should be compared to appropriate reference standards (if available), and ideally, with the naturally occurring protein. Products should be thoroughly characterized with respect to five distinct criteria (ICH Guideline, Specification: Test Procedures and Acceptance Criteria for Biotechnological/Biological Products [15]):

- Physiochemical properties
- Characterization of biologic activity
- Immunochemical properties
- Characterization of purity and potential impurities
- Yield (quantity)

In 1999 the ICH recommended the adoption of a series of guidelines defining the test procedures and acceptance criteria for biotechnological and biological products. This document provides general principles for setting a uniform set of international specifications for biotechnology products. The principles apply to "all proteins and polypeptides, their derivatives, and products of which they are components." It is assumed that these proteins are produced from recombinant or nonrecombinant cell-culture expression systems or transgenic animal systems, and that they will be highly purified and characterized using appropriate analytical procedures.

### 4.1 Physiochemical Properties

Any protein product, regardless of the host cell system, must be extensively characterized. This involves describing not only the primary amino acid structure but also information regarding any higher order structure (if known), which may be acquired by various biochemical and analytical means including, but not limited to, a mass spectrum, amino acid analysis and SDS-polyacrylamide gel electrophoretic (SDS-PAGE) analysis. Other characterizations might include isoelectric focusing, anion exchange or gel filtration chromatography, HPLC analysis, fluorescence spectroscopy, tryptic mapping, or even circular dichroism spectra.

The manufacturer must also characterize the degree of heterogeneity in the product with respect to glycosylation or other posttranslational modifications using standard biochemical analyses: SDS-PAGE, Western blotting, and RP-HPLC. Again, standards should be developed for such secondary modifications and limits set for acceptance or rejection of a particular lot of material. If a consistent degree of heterogeneity is maintained and is deemed acceptable for biological activity, then extensive evaluations of efficacy and safety may not be required.

### 4.2 Biologic Activity

As for any new lot of material, assessments of biological activity should be performed to demonstrate the efficacy of the product. This may be achieved though the use of animal-based biological assays, cell-based assays that evaluate the material at the cellular level, and biochemical activity assays such as enzymatic action. Standards should be properly defined. Potency and efficacy for individual lots of material can therefore be defined.

### 4.3 Immunochemical Properties

The immunochemical properties of the biopharmaceutical should also be fully characterized. This may include binding assays of the antibody to purified antigens, using defined regions of the antigen. Additionally the target antigen,

and if possible, the specific epitope should also be biochemically defined, particularly if immunochemical activity defines the primary biologic action.

### 4.4 Purity and Potential Impurities

Manufacturers should be able to demonstrate purity of the material. Post-translational modifications do not necessarily represent impurities, provided that they can be shown to be consistent and without effect on the biological activity. Because of the complex nature of biological systems, it is understood that certain impurities may be present. In this case they should be characterized and shown to be without biological effect. Acceptance criteria should be equal to or exceed results obtained in preclinical studies. Contamination should be strictly controlled using well-defined in-process acceptance criteria. Product stability should be likewise characterized based on real time studies conducted with representative material [19].

### 4.5 Product Content

Protein or peptide content should be well defined by an appropriate physiochemical test. If the quantity of product can be well correlated to a specific biological activity (standard curve), then an elaborate physiochemical test does not need to be performed.

### 4.6 Products Made Transgenically

With respect to the final product, transgenic recombinant proteins are essentially no different from proteins produced in bacterial, yeast, or mammalian cell systems. Donor DNA must be entirely characterized, as well as the methods for introduction of the recombinant. Production conditions should be well defined as will be methods for product harvest, purification, and storage. The product must be characterized for its physical and biologic properties. Many of the primary concerns with the use of transgenic animals in fact have already been considered during the development of purified human proteins. A number of marketed biologics derived from human plasma or serum, such as IFN α-N3 (Alferon N®) or Coagulation Factor IX (Mononine®) are prepared from pooled donor plasma, representing diverse donors in distinct environments. The plasma or serum obtained must then be extensively tested for purity and viral contamination. In a similar manner, transgenic recombinant products will routinely be pooled from multiple individuals and examined for adventitious agents. Following sample collection, postharvest techniques should again be similar to those used for bacterial or mammalian cell systems. Table 2 highlights the various aspects of the manufacturing process and release criteria for biopharmaceutical products. It becomes apparent that transgenic animal products must meet the same criteria as those required for other host systems.

**TABLE 2 Biologic manufacturing of recombinant proteins**

| | Cell Substrate | | |
|---|---|---|---|
| Element of System | Bacterial | Mammalian Cell | Transgenic Animal |
| **DNA expression construct** | • Sequence<br>• Expression vector<br>• Regulatory Elements<br>• Method of Production<br>• Purification<br>• Purity standards | Same | Same |
| **Host cell/animal system**<br>• History of Cell Line<br>• Master and Working Cell Banks<br>• Environmental Growth Conditions | • Monitor contamination | • Monitor contamination | • Animal pedigree<br>• Monitor for illness |
| **Method of gene transfer** | • Transformation | • Transfection | • Embryo manipulation |
| **Methods of identifying or characterizing recombinant proteins** | • Growth selection | • Antibiotic resistance | • Southern blots<br>• PCR |
| **Production conditions (environmental)** | • Controlled | • Controlled | • Less controlled<br>• Distinct individuals |
| **Method of product recovery** | • Batch fermentation | • Batch Fermentation | • Pool multiple individuals |
| **Product stability** | • Product specific | • Product specific | • Product specific |
| **Product storage** | • Product specific | • Product specific | • Product specific |
| **Criteria for release/ acceptance** | • Physiochemical<br>• Biologic activity<br>• Immunochemical<br>• Purity/impurities<br>• Yield | Same | Same |

## 5 ENVIRONMENTAL ASSESSMENT AND CATEGORICAL EXCLUSIONS

Any submission requiring action by the FDA must include either an environmental assessment (EA) or a claim of categorical exclusion from the requirement for an EA (21 CFR 51.15(a)).When an EA is submitted, it is evaluated

by the FDA and either an environmental impact statement (EIS) or a finding of no significant impact (FONSI) is filed by the Agency (21 CFR25.15(b)).

The following actions are generally categorically excluded from needing an EA or an EIS:

- If use of the active moiety does not increase
- If use of the active moiety increased, but the concentration of the substance at the point of entry into the aquatic environment (estimated introduction concentration; EIC) will be below 1 part per billion
- If the substance occurs naturally in the environment and the action does not alter significantly the concentration or distribution of the substance, its metabolites, or degradation products in the environment
- Action on an IND

The following assumptions are made in determining the EIC.

1. All of the drug product is used and enters the publicly owned treatment works,
2. Drug product usage occurs throughout the United States in a distribution that is proportional to the population, and amount of wastewater generated.
3. No metabolism occurs

A protein or nucleic acid comprised of naturally occurring amino acids or nucleosides, but having a sequence different form that of a naturally occurring substance, will normally qualify as a naturally occurring substance after considering metabolism.

The estimated introduction concentration (EIC) =

Kilogram (kg) of drug/biologic produced per year ÷ 365 days × $10^9$ micrograms per kg × liters of water per day entering the POTW (publicly owned treatment works; $1.2 \times 10^{11}$ liters/day) (1996 Needs Survey, Report to Congress at www.epa.gov/own)

Refinements to the calculations can be made to correct for metabolism, pharmacological activity of metabolites, nonproportional distribution of release, and other factors.

The maximum expected environmental concentration (MEEC) is the concentration that organisms would be exposed to in the environment. It is derived from the EIC after taking into account any spatial or temporal accumulation or depletion factors, such as dilution, degradation, or bioaccumulation. If a drug is rapidly inactivated or degraded under environmental conditions, then it is generally not necessary to institute further testing. If not, then tiered environmental (ecotoxicity) testing is required.

INDs generally involve relatively small amounts of drug and treatment of a limited number of recipients, so the environmental exposure is usually low. INDs are evaluated on a case-by-case basis to determine if extraordinary circumstances exist. Extraordinary circumstances include situations where there is potential for serious harm to the environment at the expected level of exposure, lasting effects on ecological community dynamics, or an adverse effect on an endangered specie or habitat of an endangered specie. The determination of extraordinary circumstances can be based on information from the agency, the sponsor, published sources, and other sources. Most commonly INDs will be submitted with a request for categorical exclusion from EA.

An example of an EA for a typical IND is provided below:

> [Company name] certifies that [drug or biologic name] is intended for use in clinical studies in which waste will be controlled and the amount of waste expected to enter the environment is expected to be nontoxic. Therefore, in accordance with 21 CFR 25.31, we request categorical exclusion from providing an environmental assessment.

## 6 DEMONSTRATION OF COMPARABILITY WHEN A MANUFACTURING SYSTEM IS CHANGED

In its Guidance on Demonstration of Comparability of Human Biologic Products, including Biotechnology Derived Products [18], the FDA formally elaborated its position on the steps manufacturers must take to demonstrate comparability following manufacturing changes (see Chapter Demonstration of Comparability of a Licensed Product after a Manufacturing Change). The FDA's position is for sponsors to discuss any potential changes with them prior to implementation to prevent unnecessary duplication of resources. These guidelines are applicable to any type of manufacturing change including changes from bacterial to mammalian cell to transgenic expression systems. While this list is not exhaustive, some of these potential changes include the following:

- Changes in DNA vector
- Changes in host cell system
- Changes in fermentation/culture process
- Equipment changes
- Changes in raw materials
- Changes in purification techniques
- Changes in storage conditions

Sponsors may make changes in manufacturing processes for a particular product for a number of reasons, including improvement of quality, yield, or simply manufacturing efficiency. Such changes may be frequent, and manufacturing changes have been successfully made during or even after the

completion of clinical studies. In these situations comparability data have provided assurance that the product would continue to be safe, pure, and effective. For example, in the past such changes have included conversion from pilot plant to full-scale production, a move of the production facilities from one location to another, and implementation of changes in different stages of the process such as fermentation, purification, and formulation. For changes made prior to product approval, the sponsor must fully describe the changes in any license application or investigational new drug application (IND). The criteria for establishing product comparability include the following:

- Biochemical characterization
- In vitro and in vivo bioactivity
- Pharmacokinetics
- Safety

In terms of product identity and biochemical characterization, when manufacturers transition from one system of production to another, qualified product standards are used to judge product identity and purity. Biologic activity testing either in vitro or in vivo are also frequently standardized and can be consistently performed on protein products derived from any production system whether it be bacterial, mammalian, or intact organism. In most cases, however, mimicking the biological activity in the clinic following a change in the manufacturing process would not be required provided that cell or biochemical assays can be shown to correlate to clinical response.

Regardless of the cell substrate utilized for biologic production, consistency in both upstream and downstream production processes will also play an important role in validating any novel biological system. Detailed knowledge of the stepwise manufacturing paradigm and elaboration of these methodologies may compensate to a large extent for changes in starting materials. Nevertheless, manufacturers should make available to the FDA extensive chemical, physical, and biological comparisons with side-by-side analyses of the old and new lots of material. Most important, tests should include those routinely used for release of the bulk drug substance and final drug products in addition to those tests aimed at evaluating the impact of the change on the product. Basically the regulatory position suggests that sponsors follow their normal procedures when implementing a manufacturing change, namely demonstrating comparability. The tests should be sensitive to detect any alteration induced by the manufacturing change [21]:

> *These tests should include tests routinely done on all production lots, those initially used to fully characterize product structure and identity and establish product consistency from one production lot to another . . .*

An example of changes in manufacturing system and how comparability was assessed from a regulatory perspective involves interferon β-1a (IFNβ-1a).

Recombinant IFNβ-1a is produced in CHO cells, and the primary amino acid sequence of the recombinant protein produced by these cells was identical to the naturally occurring IFNβ. IFNβ-1a is a single, glycosylated polypeptide 166 amino acids in length. An initial batch of drug product, termed BG9015 was used in phase 3 clinical trials. The sponsor subsequently developed a new CHO cell line carrying the IFNβ-1a gene and designated the product isolated from these cells BG9216. These particular CHO cells harboring the IFNβ-1a gene were adapted for the suspension culture. Data supporting the use of this cell line were submitted to the FDA, and it was determined that the specific activity of BG9216 was greater than BG9015 and contained an additional peak in the peptide map. Pharmacokinetic studies also demonstrated that BG9216 was not equivalent to BG9015. With these data the FDA determined that BG9216 was not comparable to BG9015.

The sponsor then developed yet another IFNβ-1a cell line, and the product produced by these cells was designated BG9418. This product was extensively characterized and compared side by side with BG9015. Biological, biochemical, and biophysical analyses demonstrated that the two molecules were comparable. Biological activities of each molecule were similar using several different antiviral and antiproliferative assays. Chromatography of peptides derived by proteolysis of the two proteins was nearly identical. Carbohydrate analysis was also performed, and both forms exhibited similar patterns. In this case the sponsor further performed clinical pharmacokinetic studies and showed that the two forms behaved identically. Based on these data, the FDA determined that drugs BG9015 and BG9418 were comparable and that data obtained during the study of BG9015 would support the licensing of the BG9418 molecule [22].

Change to an entirely new host cell system can also have either minor or significant effects on recombinant protein production. In a comparison of five distinct eukaryotic cell expression systems, Geisse et al. [9] assessed the expression of human leukemia inhibitory factor (hu-LIF) in five of the most commonly used cell lines. The yields and quality of protein product were assessed in CHO, Sp2/0, MEL, COS, and baculovirus-infected insect cells. Although recombinant, biologically active product was produced in every case, yields and glycosylation patterns varied widely among the systems. To date, such significant changes have yet to be implemented into a biologic manufacturing paradigm.

## 7 APPLICATION OF COMPARABILITY CRITERIA TO A NEW MANUFACTURING TECHNOLOGY

The potential impact induced by changes in manufacturing systems cannot be assumed. Thus major changes in the host system may or may not impact a product's identity or activity. Applying sound scientific principles of safety and

efficacy will permit manufacturers and regulators to objectively assess the impact of manufacturing changes.

Changes can occur in the manufacturing system at any point in the drug development cycle, and the extent of comparability testing may be related to the timing of such changes. Regardless of the stage of development, the FDA encourages sponsors to consult with them prior to instituting any major change in the manufacturing process. For biological products that the FDA has approved, the sponsor should submit information about manufacturing changes pursuant to 21 CFR 601.12 or 21 CFR 314.70(g) along with any FDA guidance on the changes described. The European Agency for the Evaluation of Medicinal Products (Committee for Proprietary Medicinal Products) has issued similar comparability statements in its year-end 2003 publication [17].

In general, the later in the development process a change takes place, the more studies may be expected to be conducted. Thus it is advantageous for biologic manufacturers to select an overall manufacturing approach as soon as possible. Similarly manufacturers will likely need to demonstrate some form of comparability each time major change takes place in the development process (see Figure 1.). Naturally, the best time to make significant changes to a biomanufacturing system is during the early development phase. If this is the case, few or potentially no comparability studies may be needed. When manufacturing changes are instituted during the preclinical (pre-IND) phase, limited batch comparison studies may be required to demonstrate comparability, but formal bioequivalence testing, including animal toxicology or pharma-

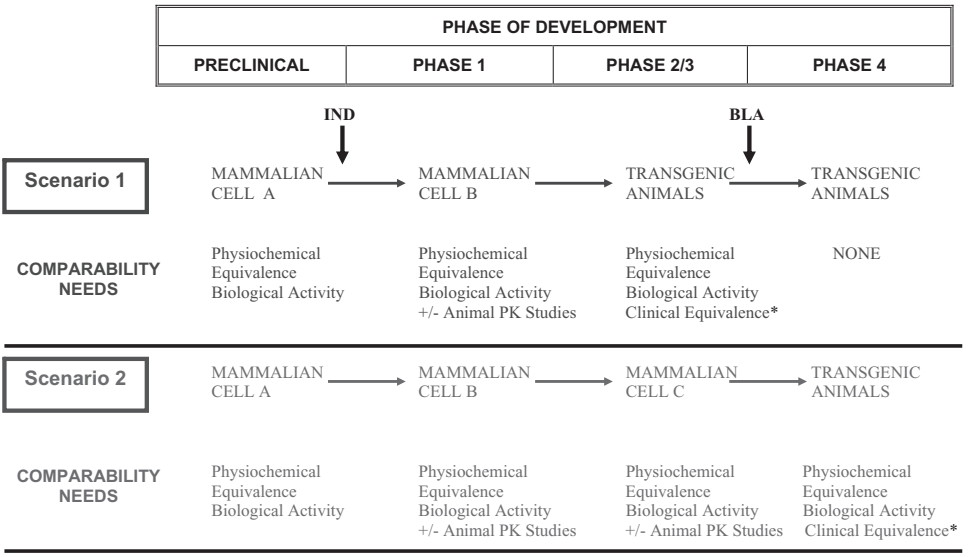


*Pharmacodynamic equivalence as demonstrated by validated surrogate markers or primary clinical endpoints.

**Figure 1** Comparability needs in relationship to development stage.

cokinetic studies, may not be required. The sponsor, in its application, should provide a rationale for the types of comparability testing performed based on the nature of the change in the manufacturing system.

In most cases changes in a manufacturing paradigm are unlikely to impose drastic changes in product quality. However, even if purity and identity can be unequivocally established following a change in manufacturing system, animal pharmacokinetic or pharmacodynamic studies may be warranted. While a subtle change in product quality may have little or no impact on the toxicological profile of a biologic, changes in glycosylation, in protein folding, or in the tertiary structure could impact the pharmacodynamics of a biologic. Ultimately, if clinical studies are warranted, product innovators will frequently have at their disposal surrogate markers of efficacy and/or toxicity. These markers should allow more limited studies with defined surrogate endpoints in lieu of costly pivotal clinical trials. Identification and validation of these surrogate markers in Phase 2 and 3 studies, respectively, will be an increasingly important step in the development of any novel biologic.

When a manufacturing change occurs during the clinical phase (between the IND and BLA or NDA), the sponsor will need to demonstrate that batches used are chemically, biologically, and perhaps clinically equivalent. Again, pharmacodynamic equivalence should be verified using a validated animal model. Clinical pharmacokinetic studies might be warranted if chemical, biochemical, and animal pharmacodynamic studies cannot demonstrate comparability with the earlier product. This is demonstrated in scenario 1 in Figure 1. As was the case for IFN β-1a discussed above, the sponsor needed to demonstrate that the product produced in the new CHO cell line was biologically equivalent to the earlier version used in an initial clinical study.

Finally, manufacturing changes might also be introduced in the post-approval phase (Figure 1, scenario 2), namely after the initial, approved version of the product has already been on the market. Sponsors might pursue limited pharmacodynamic studies provided that chemical and biochemical equivalence can be established. In such a case that physiochemical or biologic comparability cannot be demonstrated following a manufacturing change during the postmarketing phase, clinical batch-to-batch comparability must be demonstrated.

## 8 SUMMARY AND CONCLUSIONS

Numerous manufacturing systems are presently available for large-scale expression of recombinant proteins. Bacteria, yeast, mammalian cells, and transgenic animals have all been successfully utilized for large-scale production of recombinant biologics. In most cases the recombinant product is a cloned version of a previously purified human or animal protein; thus reference standards are available for comparison. The FDA and the European Regulatory Agency have since developed standards for both the production

and characterization of protein therapeutics produced using recombinant technology. These standards address the identity, purity, stability, and biological activity for recombinant proteins. Moreover, since it is expected that novel technology will facilitate conversion from one production method to another, both the FDA and the European CPMP have provided guidance on demonstrating comparability following changes to a manufacturing system. These guidances have provided a common pathway that can be followed for incorporating and evaluating changes within a biologic manufacturing system.

Any submission requiring action by the FDA must include either an environmental assessment (EA) or a claim of categorical exclusion from the requirement for an EA. INDs are frequently submitted with a request for categorical exclusion from EA, since they generally involve relatively small amounts of drug and treatment of a limited number of recipients such that the environmental exposure is usually low.

The criteria for demonstrating comparability are dependent on both the final product and its stage of development, rather than on perceived degree of change, especially as one may change host systems radically. While recombinant proteins in general are often more complex than small molecule drugs, the approach and standards for product comparability are similar. Structure, identity, purity, and biologic activity still remain the final endpoints of analysis, regardless of whether a novel drug is a small molecule or large, recombinant protein. While the process and its consistency and predictability are very important, the final drug product is the entity from which safety and efficacy will be evaluated. It is the responsibility of drug manufacturers to demonstrate product safety and efficacy of the final drug product throughout the development cycle where changes in process and manufacturing systems are common. With adherence to scientific principles and regulatory guidance, the development and commercialization of novel protein therapeutics utilizing manufacturing systems that meet the production needs of new lifesaving therapeutics should be realized.

## REFERENCES

1. PDR, *Physician's Desk Reference*. 2006. Thomson PDR, Montvale, NJ, USA.
2. Werner RG, Noe W, Kopp K, Schluter M. Appropriate mammalian expression systems for biopharmaceuticals. *Arzneimittelforschung* 1998;48(8):870–80.
3. Guidance for Industry Environmental Assessment of Human Drugs and Biologics Applications (1998).
4. Pines O, Inouye M. Expression and secretion of proteins in *E. coli*. *Mol Biotechnol* 1999;12(1):25–34.
5. Janne J, Hyttinen J-M, Peura T, Tolvanen M, Alhonen L, Halmekyto M. Transgenic animals as bioproducers of therapeutic proteins. *Ann Med* 1992;24:273–80.
6. Ladisch MR, Kohlmann KL. Recombinant human insulin. *Biotechnol Prog* 1998; 8(6):79–81.

7. Eckart MR, Bussineau CM. Quality and authenticity of heterologous proteins synthesized in yeast. *Curr Opin Biotech* 1996;7:525–30.
8. Pennell CA, Eldin P. In vitro production of recombinant antibody fragments in Pichia pastoris. *Res Immunol* 1998;149(6):599–603.
9. Geisse S, Gram H, Kleuser B, Kocher HP. Eukaryotic expression systems: A comparison. *Protein Express Purific* 1996;8:271–82.
10. Lewis AM, Krause P, Peden K. A Defined Risks Approach to the Regulatory Assessment of the Use of Neoplastic Cells as Substrates for Viral Vaccine Manufacture. US FDA, 1999.
11. Mason HS, Lam DM, Arntzen CJ. Expression of hepatitis B surface antigen in transgenic plants. *Proc Natl Acad Sci USA* 1992;89(24):11745–9.
12. Larrick JW, Yu L, Chen J, Jaiswal S, Wycoff K. Production of antibodies in transgenic plants. *Res Immunol* 1998;149(6):603–8.
13. Bruggemann M, Neuberger MS. Strategies for expressing human antibody repertoires in transgenic mice. *Immunol Today* 1996;17(8):391–7.
14. Clark AJ. The mammary gland as a bioreactor: Expression, processing, and production of recombinant proteins. *J Mammary Gland Biol Neoplasia* 1998;3(3): 337–50.
15. Specifications: Test procedures and Acceptance Criteria for Biotechnological/Biological Products, ICH Harmonized Tripartite Guideline, 1999.
16. US FDA, CBER, Points to Consider in the Manufacture and Testing of Therapeutic Products for Human Use Derived from Transgenic Animals. 1995.
17. Committee for Proprietary Medicinal Products (CPMP). Guidance on Comparability of Medicinal Products Containing Biotechnology-Derived Proteins as Active Substance. EMEA, 2003.
18. US FDA. Points to Consider in the Characterization of Cell Lines Used to Produce Biologicals. 1993.
19. US FDA. ICH Q5C Quality of Biotechnological Products: Stability Testing of Biotechnological/Biological Products. 1996.
20. Office for the Protection from Research Risks. Guide for the Care and Use of Laboratory Animals. Washington: National Academy Press, 1996.
21. Cavagnaro JA. Establishment of criteria for determining comparability of "well-characterized" proteins. *Dev Biol Stand* 1998;96:79–81.
22. US FDA. Summary Basis of Approval for Interferon beta-1a. PLA#95-0979, 1995.

# The Principles of ICH S6 and the Case-by-Case Approach

JOY A. CAVAGNARO, PhD DABT, RAC

**Contents**

1 History of Preclinical Safety Evaluation
2 ICH S6: A New Approach to Preclinical Safety Evaluation
3 Principles versus Practices
4 Implementation of ICH S6 Guidance
  4.1 Use of Animal Models
  4.2 Selection of Relevant Species
  4.3 Cross-Reactivity
  4.4 Analogous Products
  4.5 Use of Relevant Test Systems
5 Future of Preclinical Safety Evaluation and ICH S6
6 Conclusion
  References

## 1 HISTORY OF PRECLINICAL SAFETY EVALUATION

The conduct of toxicology studies in laboratory animals has been driven by experience, historical precedence, and governmental requirements, and the results of these studies usually, and reasonably, have led to restrictions on the use, or method of use, of the chemicals concerned [1]. The primary objective of pharmaceutical preclinical safety evaluation is to provide information essential for the initiation of clinical trials. Scientific rationale and controlled reproducible data are used to show that the initial human risk is so low as to be ethically and practically acceptable in relation to the medical value of the information to be obtained from humans. Preclinical safety studies performed throughout the course of product development facilitate and may guide work

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

**TABLE 1 Objectives of preclinical safety evaluation**

- To permit initiation of clinical trials and to support subsequent clinical investigations
- To recommend safe starting dose, dose escalation scheme, regimen, and route of administration
- To identify potential target organs of toxicity
- To identify parameters to monitor in the clinical trial (e.g., biomarkers of safety and activity)
- To discern the mechanism of activity/toxicity and reversibility or delay of effects
- To identify "at-risk" human populations by thorough definition of study subject inclusion/exclusion criteria
- To provide safety data to support product labeling claims
- To provide data to support potential product liability concerns
- To provide critical information to support termination of a potentially unsuccessful development program in a timely manner

in healthy volunteers and patients by confirming the acceptability of the risk–benefit ratio of the proposed clinical development and how to minimize any foreseeable risks (see Table 1). Efforts devoted to improving the predictive value of preclinical studies are critical to achieving these objectives.

In the early 1950s guidelines and standard approaches for general toxicity testing were introduced in the form of acute subchronic and chronic toxicity studies to help predict human risks. The increase in the number of new drugs, chemicals, and environmental pollutants led to the introduction of additional toxicological tests to screen for specific mutagenic, teratogenic, and other reproductive and/or carcinogenic activities to better determine toxicity to specific organ systems. This testing became standardized. During this time these approaches for toxicity testing were considered essential in ensuring adequate coverage of potential risks and in promoting consistency and improving the quality of data for review. They also provided guidance to industry on testing strategies that would generally be acceptable for evaluation by regulatory authorities.

In the late 1960s the FDA Goldenthal guidelines described, in general, the types of preclinical studies that could be used in support of the several phases of clinical investigation as well as the approval of a new drug application. Admittedly the guidelines were not intended to be used as protocols, but merely as guides [2]. The agency stated that it was in full agreement with the critics of current practices—and that the toxicology assessment of new drugs should keep pace with the more sophisticated technology, consistent with the objective of being as critical and comprehensive as possible. It was acknowledged that development of new methods in drug evaluation was proceeding at a rapid rate. The agency correctly cautioned that general acceptance of new procedures would be predicated on their applicability and predictive value [2].

When originally promulgated, the 1968 guidelines were intended to provide a framework of guidance on which a testing program for a specific compound

could be built. In practical usage the term "guideline" has come to connote a list of necessary tests to meet minimum regulatory requirements. Although guideline connotes various meanings to different individuals, to the toxicologist it has traditionally meant a comprehensive set of rules to follow for the testing of new drugs for registration with regulatory authorities. In the extreme it is a "check box" approach to the toxicity testing of new compounds for safety assessment. Thus lists of specific tests have come to be seen as absolute requirements for product registration, and not doing a test would mean not following the guidelines. However, comprehensive guidelines can only be based on what is known or can be anticipated [3].

Similar approaches were adopted by major regulatory authorities in other countries, including the United Kingdom, France, Germany, Italy, and Scandinavia and subsequently the European Union. The official agencies and the pharmaceutical industries' own experts have published recommendations for a wide range of types of toxicity tests to provide comprehensive data in stardardized form for assessment.

## 2 ICH S6: A NEW APPROACH TO PRECLINICAL SAFETY EVALUATION

Biopharmaceuticals have brought new challenges to toxicologists. In the early 1980s neither industry toxicologists nor regulatory scientists were sure of what constituted an appropriate toxicological assessment program for biopharmaceuticals. It was fortuitous that interferon, an extremely species-specific protein was one of the first human biopharmaceutical products. Despite performing a "traditional" toxicity package, including assessment in multiple species, toxicologists found that the animal studies did not predict the common adverse events observed in humans [4].

In 1986 the Biotechnology Working Party was established in Europe to focus on specific issues related to the development of biotechnology-derived pharmaceuticals. In July 1986 a Satellite Symposium to the IV International Congress of Toxicology was held in Tokyo, Japan. Approximately 135 scientists representing virtually all major countries attended the symposium. Among the attendees were government regulatory scientists, university scientists, and industrial scientists and research managers, all with an interest in the development of new biotechnology products [5]. A few statements made during this meeting are worth noting as they introduced an approach to preclinical safety evaluation that would distinguish the practice of biopharmaceutical preclinical development for the ensuing two decades.

> The availability of modern biotechnology products opens a wide range of exciting therapeutic possibilities. . . . For many areas of potential toxicity, no satisfactory safety tests are available, and most of the existing models are not yet properly validated. Nevertheless, since progress can only be made through

> accumulation of experience, it appears reasonable to suggest that a considered effort be made to develop and to use scientifically sound experimental systems, and to refrain from following the beaten track of routine toxicity testing. [4]

> There is no place for detailed programs of rigidly pre-defined tests to be applied automatically to all products. Instead, useful information will only be obtained if the need for any experimentation at all is decided according to the specific properties and planned uses of each compound. Then, the nature of the studies should be adapted to those individual circumstances. Toxicity testing in this area is most like a series of pharmacological explorations and should not be expected to follow conventional rigid guidelines. [6]

> In part our [FDA] experience with the endocrine drug products (the peptides) has influenced our approach for other proteins. We recognize that we do not know enough about the pharmacology of many of the new immunomodifiers and many of the other protein products. I believe the case-by-case approach is the most sensible course at this time I'm sure that as we learn more about the pharmacology of a class of proteins, for example, that our approach will be modified for that class. [7]

> Consistent with the case-by-case approach, some toxicological studies will follow a traditional approach, while others may deviate. Instead of focusing on procedures (length of studies, number of groups, number of animals, number of species, etc.) it would be best to focus on the goal behind such testing in order to identify and understand the potential toxicity of the agent. [8]

Based on experiences gained over the intervening decade, a concept paper was proposed by the FDA in 1995, under the auspices of the International Conference on Harmonization of Technical Requirements of Pharmaceuticals (ICH), for a new safety topic specifically relating to the preclinical safety evaluation of biotechnology-derived products. In February 1997 the Thirteenth CMR International Workshop provided an opportunity for international experts to discuss experiences and difficulties encountered in designing scientifically based preclinical safety evaluation programs for biopharmaceuticals. This two-day meeting brought together toxicologists and clinicians, from 32 pharmaceutical and biotechnology companies and regulators and regulatory advisors from the European Medicines Evaluation Agency and 9 countries: Denmark, France, Germany, Italy, Japan, the Netherlands, Sweden, the United Kingdom and the United States [9]. Recommendations arising from the CMR Workshop were taken into consideration by the Expert Working Group for the final drafting of ICH S6 guideline and agreement was reached at ICH 4 in Brussels in July 1997 [10].

At the ICH 4 meeting ICH M3 was also finalized, acknowledging that that there had been marked changes in the kinds of therapeutic products being developed (e.g., biotechnology-derived products) and that existing paradigms for safety evaluation might not always be appropriate or relevant. As such the safety evaluation of biotechnology-derived products should be considered on a case-by-case basis, as described in ICH S6 [11].

The principles of case-by-case assessment were also suggested in a 1998 paper by DeGeorge et al. entitled "Regulatory Considerations for Preclinical Development of Anticancer Drugs." The authors acknowledged that basic research continues to provide information about new cellular mechanisms central to malignancy and often leads to drugs that attempt to exploit those mechanisms. The optimal development of a new class of drug may differ from successful approaches used in the development of older well-established classes. The authors commented that new biological endpoints and new methods in toxicology may be discovered that cannot be anticipated. The recommendations in the manuscript emphasized the concerns to be addressed and the importance of avoiding excessively restrictive and specific guidelines so as not to impede the development of innovative therapies for clinical use [12].

## 3 PRINCIPLES VERSUS PRACTICES

The basic principles of toxicology are applicable across product classes. It is the specific attributes of the product that have the greatest influence on the successful practice of biopharmaceutical toxicology (see Table 2). This focus on product attributes has defined the case-by-case approach. Table 3 provides a further definition of the case-by-case approach (see Table 3).

Pharmaceuticals and biopharmaceuticals can be viewed as a product continuum based on size and complexity in molecular structure. However, as products have evolved, there has been a blurring of product attributes. Small molecules have become larger as the result of alternative scaffolding technologies (e.g., forming protein conjugates and fusion proteins) in order to improve exposure characteristics and dosing regimens. Large molecules have become smaller (e.g., antibody fragments and protein mimetics) in order to improve distribution and decrease potential immunogenicity. There are small molecules in development that exhibit unique species specificity making the traditional test species, rat and/or the dog, less relevant for safety assessment. Novel delivery technologies are also enabling alternative routes of delivery for biopharmaceuticals, such as by the oral and inhalation routes. Some products such as oligonucleotide drugs (ODNs) may have combined product attributes. For example, ODNs are synthetically derived but have complex chemical profiles and are catabolized in ways similar to those followed by certain biopharmaceuticals. Although toxicity assessments are designed to address hybridization-independent effects, some ODNs can also exhibit species specificity where analogous sequences may be needed to assess hybridization-dependent effects, namely toxicity related to exaggerated pharmacology.

## 4 IMPLEMENTATION OF ICH S6 GUIDANCE

The optimal preclinical development of a biopharmaceutical product has benefited from experience. This is because the strategy for designing programs is

**TABLE 2 Comparative product attributes**

| Product Attribute | Pharmaceutical | Biopharmaceutical |
|---|---|---|
| Manufacture | Chemical synthesis | Biological synthesis (cell culture, transgenic plants, transgenic animals) |
| Composition | Commonly organic chemical | Protein, carbohydrate, DNA, virus, bacteria, cell |
| Structure | Well-defined, linear | Complex, tertiary structure |
| Size | Generally <1 kDa | Generally >30 kDa up to 800 kDa |
| Purity | Homogeneous single-entity high chemical purity (except racemic mixtures) | Heterogenous mixture (microheterogeneity, aggregates) |
| Impurities | Easy to qualify; toxicity testing may be required | Difficult to qualify; toxicity generally not an issue but may affect immunogenicity |
| Product characterization | Specifications defined early in development, generally little change; usually one bioanalytical method (mostly LC/MS-MS) | Broadly defined at initial stages of developed and refined; several bioanalytical assays (e.g., HPLC, SDS-PAGE, MALDI-TOF) |
| Potency | Not determined | Required; generally in vivo but for some products in vitro acceptable. |
| Stability | Stable and not heat sensitive | Less stable and sensitive to heat and shear |
| Route of administration | Oral, topical, inhalation | Parenteral or targeted (IV, SC, IM, intracardiac, intrathecal, intraarticular, etc.) |
| Pharmacokinetics | Half-life usually minutes to hours | Half-life usually days to weeks |
| ADME | Rapid entry through blood capillaries, distribution to many organs/tissues, metabolized to active and non-active metabolites | Distribution limited based on size to plasma a/o extracellular fluid, degraded (catabolized) to endogenous amino acids |
| Drug–drug interactions | Influenced by metabolism (enzyme induction or inhibition)—can be significant | Interactions related to additive or synergistic pharmacological activity not metabolism |

**TABLE 2** ***Continued***

| Product Attribute | Pharmaceutical | Biopharmaceutical |
|---|---|---|
| Dose response | Linear—establishment of maximum tolerated dose (MTD) | Can be bell-shaped—establishment of optimum biological dose (OBD) |
| Targets | Intracellular or ligand–receptor | Cell–matrix, cell–cell, or ligand–receptor |
| Pharmacological activity | Active in standard screening and functional assays | Not active in standard screening assays; novel models created to address activity (e.g., transgenic animals, homologous test material) |
| Species specificity | Species independent; preclinical assessments for generally toxicity generally performed in one rodent (generally rat) and one nonrodent (generally dog) | Species specific; nonhuman primates often the only relevant species or design of specific animal models |
| Toxicological effects | Unpredictable; can be related to metabolites or unrelated to mechanism of action | Predictable based on known mechanism of action "exaggerated pharmacology"; animal models of disease often used to assess safety in addition to activity |
| Immunogenicity | Rare; allergic or hypersensitivity reactions may occur | Common; may affect PK or result in immune-mediated adverse events other than allergic and/or hypersensitivity responses |
| Dosing interval | Daily | Intermittent |
| Dose formulation | Complex | Simple |
| Demonstration of equivalence | Bioequivalence (generic equivalents) | Comparability within a development program (more difficult to define biogeneric) |

*Source*: Adapted from Cavagnaro (2002), Baumann (2006), and Horvath (personal communication).

influenced not only by the amount of existing knowledge concerning the pharmacology and toxicology for the specific product and product class but also by the expanding knowledge base of a number of product classes. A number of articles have been published over the years highlighting the challenges of

**TABLE 3 Case-by-case approach to preclinical safety evaluation**

| Is Not … | Is … |
|---|---|
| • Consistent with the "traditional" practices for small molecules | • Consistent with the principles for safety evaluation of small molecules |
| • Standardized approach | • Product or product-class specific |
| • Unique to biopharmaceuticals | • Science-based |
| • A minimalist approach | • Questions-based |
| • An opportunity to "get a better label" | • Experiential-based |
| • Just what a sponsor wants to do | • Targeted |
| • Practiced by all developers of biopharmaceuticals | • Flexible |
| • Embraced by all regulators | • Fair |
| • Easy to predict "if acceptable" to regulatory authorities | • More difficult to predict "if acceptable" to regulatory authorities |

preclinical development of biopharmaceuticals [13–20] as well as the valuable experiences in implementing the case-by-case approach [21–30]. The following is a discussion of a few key areas.

### 4.1 Use of Animal Models

A guiding principle in the design of preclinical safety studies is to parallel as closely as possible the clinical conditions of exposure. In accordance with the principle, much attention is paid to the dosing regimen with respect to the route of administration, duration of exposure, and dosing interval. With respect to mirroring the characteristics of the patient population to be exposed, normal animals appropriately parallel the typical phase 1 population in normal subjects (healthy volunteers). Where toxicologists deviate from the principle of correlation of clinical conditions of exposure is in the evaluation of potential toxicity in patients in the later safety and efficacy trials in phase 2 and beyond. The deviation from this principle relates primarily to the physiological state of the clinical populations involved as they are no longer healthy volunteers but rather individuals who have a specific disease and/or are very ill. Thus a relevant question for the toxicologist is to ask whether toxicology studies in normal animals adequately assess the risks in sick people. The answer to this question has even greater significance for biopharmaceuticals as first in human (FIH) phase 1 trials are often conducted in subjects with disease for ethical and practical reasons.

Animal models of disease play a critical role in the drug discovery process and are important in the lead candidate selection process as well. Categories of animal disease models include spontaneous disease, induced models (e.g., chemically, immunologically), xenograft models, infection models, and genetically modified models (e.g., transgenic knockouts (KOs) or knock-ins (KIs), humanized animals (e.g., expressing the human protein or receptor). The sub-

sequent evaluation and demonstration of biological activity of test articles in these animal models is often pivotal for further progression of these agents in the clinic.

The principle of estimating a therapeutic index prior to clinical trials typically involves determining the no observable adverse effect level (NOAEL) and comparing that to the projected human dose. In providing the estimate, the efficacious dose is typically obtained from in vitro data with human cells or tissues and in vivo preclinical pharmacology studies that involve animal disease models. Not infrequently the species used to estimate the toxic level is different from the species used to estimate an efficacious level. Thus the therapeutic index is not a true ratio as the units (species and/or conditions) are often different. On the other hand, if one were to obtain information relating to toxicity as well as efficacy from studies employing animal models of disease, a direct estimate of therapeutic index could be made provided that appropriate models had been characterized or validated in the relevant species.

The decision on the use animal models of disease for assessing safety is based on a consideration of a number of factors, not the least of which is whether an animal model of disease is available for the intended disease. There are also ethical and welfare considerations as in any proposal to use animals in scientific work. Animal models of human disease may not mimic all aspects of disease or be more sensitive. However, as long as there is a good understanding of the human disease as well as acknowledgment of specific limitations of the model, studies should allow for better predictions of risk in the intended disease populations (see Table 4).

**TABLE 4 Use of animal models of disease**

| Potential Advantages | Potential Disadvantages |
|---|---|
| • Ethical considerations in protection of humans<br>• May be useful to screen compounds for lead candidate<br>• Potential for increased sensitivity<br>• Disease condition may parallel target population<br>• Early opportunity to define biomarkers of safety and activity more relevant to intended population<br>• Direct estimate of therapeutic index feasible<br>• Opportunity to understand the pathogenesis of treatment related findings<br>• Support human safety and efficacy when human clinical trial is not feasible or ethical (“animal rule”) | • Ethical considerations in use of animals<br>• Relative paucity of background data<br>• Inherent variability<br>• May only represent one aspect of the disease<br>• Excess mortality may confound data interpretation<br>• Increased sensitivity may not be relevant<br>• May be limited in size of group<br>• Potential limitations in study design<br>• Potential interference of disease process with safety evaluation |

A variety of transgenic animals have been used over the years to help characterize the activity and safety of biopharmaceuticals. Knockout animals have been used to assess a worst-case scenario for maximum inhibition (e.g., MAbs blocking a receptor and suppression of endogenous cytokines or proteins); knock-in animals have been used to describe a worst-case scenario for overexpression (e.g., consequences of growth factor induced proliferation) and humanized transgenic animals (e.g., expressing the human receptor). In the latter case the clinical material can be directly assessed rather than relying on the species specific homologue or on indirect extrapolation from the actions of the clinical (human) therapy acting on targets in another species. In all cases, however, the models lack historical control data. There may be additional concerns with respect to potential differences related to epitope density, localization/compartmentalization, signal transduction pathways, regulation, unknown compensatory mechanisms, host defense mechanisms, or natural life history. Thus as a caution, models may either overestimate or underestimate identification and the magnitude of of a hazard.

### 4.2 Selection of Relevant Species

Traditionally toxicologists have used at least one rodent and nonrodent species for multidose toxicity studies. The use of two species is important for assessing potential variability of metabolism, for products with extensive distribution, and in cases where a relevant species has not been defined. The rat and dog are selected in most cases, usually on an empirical basis [2] without an open-minded consideration of whether alternate species might be better in terms of biochemistry and metabolism [1].

In 1987 Zbinden identified three main areas of concern for biopharmaceuticals: (1) the toxicity per se (intrinsic toxicity), (2) the adverse effects related to the pharmacodynamic properties (exaggerated pharmacology), and (3) the undesirable responses mainly due to indirect biological responses not related to pharmacodynamic properties [4]. Since contaminants are usually present only in small amounts (e.g., host cell proteins, contamination with viruses, DNA, endotoxin, antibiotics), their intrinsic toxicity is considered of minor significance and best addressed through rigorous quality control. Intrinsic toxicity is defined as toxicity unrelated to a product's pharmacodynamic properties, such as the clinical syndrome, flu-like symptoms, fever, nausea, and malaise, observed in humans following interferon administration, or fluid retention and interstitial pulmonary edema following IL 2 administration. The information gained from clinical observations are used to guide toxicological experiments in animals to gain a better understanding of the intrinsic toxic potential and hopefully define more predictive animal models. However, standard toxicological test procedures are poorly suited to a priori identify intrinsic effects. Toxicity that occurs as a consequence of the desired pharmacodynamic effect can often be predicted especially when the biological characteristics are

known. Biological toxicity due to indirect biological responses unrelated to pharmacodynamic properties includes toxicity related to structure such as the allergic and sensitizing potential of biopharmaceuticals [4]. See also immunogenicity discussed below.

ICH S6 states that safety evaluation programs should include relevant species demonstrating pharmacological activity. As such the mechanism of toxicity is defined by exaggerated pharmacology. Importantly, relevant pharmacological species can also be used to assess biological toxicity and in some cases intrinsic toxicity. In some cases it may be necessary to consider animal models of disease to evaluate intrinsic toxicity. For small molecules where toxicity may be the result of metabolites or due to extensive distribution to nontarget tissues, a species lacking the target may be relevant as the mechanism of toxicity may be unrelated to exaggerated pharmacology. In such cases the most sensitive species becomes the most relevant species with respect to dose extrapolation. A variety of techniques (e.g., immunochemical or functional tests) can be used to identify a relevant species.

There has been the misconception that only one species is expected for assessing general toxicity of biopharmaceuticals. However, the language in ICH S6 explicitly states that safety evaluation programs should normally include two relevant species, but in certain justifiable cases, one relevant species may suffice (e.g., when only one relevant species can be identified or where the biological activity of the biopharmaceutical is well understood). The guidance intentionally did not specify use of the "most relevant" in order to avoid the routine consideration of use of higher primate species (e.g., greatest homology of a protein or a receptor with chimpanzees or baboons). Demonstration of binding of the biopharmaceutical to a receptor or other target may not be sufficient to define a relevant species as not all molecules that bind actually induce activity. Therefore primary consideration should be given to receptor-mediated activity. Determining biological activity is based on an understanding of in vitro receptor occupancy, affinity, distribution, and in vitro and in vivo pharmacological effects. Importantly, toxicity studies in nonrelevant species are discouraged.

The ICH reproductive toxicology guidance considers the acceptability of a single species if it can be shown by means of pharmacological and toxicological data that the species selected is a relevant model for humans. ICH further states that there is little value in using a second species if it does not show the same similarities [31].

### 4.3 Cross-Reactivity

Tissue cross-reactivity studies for monoclonal antibody products were originally intended to explore off-target tissue binding. Studies have thus been used to screen candidates to avoid off-target binding. As stated in ICH S6, an animal species that does not express the desired epitope may be of some relevance for assessing toxicity if comparable unintentional tissue cross-reactivity to

humans is demonstrated. For example, if cross-reactivity was observed in human tissues, assessments of potential binding to nonhuman primate tissue is recommended in order to determine if similar binding is observed in order to justify the use of a nonhuman primate. If similar binding is not observed, the data could be used to justify use of rodents especially in cases where the epitope is not present in normal animals (e.g., tumor antigens). Tissue cross reactivity, however, has inappropriately expanded to testing a variety of animal species in order to select a relevant species for assessing general toxicity. As discussed above, selection of a relevant species is more correctly justified by pharmacological-based receptor or other specific target studies. Similar off-target binding to nonhuman primate nonreproductive tissues justifying use in general toxicity studies does not justify use of the nonhuman primate in reproductive or developmental toxicity studies. The design and relevance of tissue cross-reactivity studies for bispecific and trispecific antibodies, conjugated antibodies, and fusion proteins may differ from monoclonal antibody applications.

Unintentional cross-reactivity may also be suggested by Basic Local Alignment Search Tool (BLAST) that are routinely performed electronically to identify potential binding to nontarget tissue or to determine if there is unintentional cross-reactivity with an endogenous molecule such as following administration of a monoclonal antibody or a vaccine [32]. CEREP assay platforms routinely performed for small molecules for determining off-target binding are not used for biopharmaceuticals.

### 4.4 Analogous Products

For biopharmaceuticals that are unique in their species specificity (i.e., reacting only with monkeys, chimpanzees, and/or humans) use of the homologous protein or more broadly defined analogous products that recognize the ortholog of the original human target, have been considered in order to assess general and/or specific toxicity. The term "surrogate molecule" has been used incorrectly especially if one considers the judicious use of the term when referring to surrogate endpoints in clinical development. The acceptability of a surrogate endpoint requires extensive validation.

While analogous products are often used in early discovery and development to provide data in support for proof of concept, there is concern regarding their use for assessing safety unless they have been shown to be comparable to the clinical material (see Table 5). The use of analogous products is most often considered for assessing reproductive/development toxicity based on the inherent limitations of nonhuman primate models for characterizing reproductive risks. Analogous products are also considered for assessing carcinogenic risk for products intended for chronic use. Interestingly both of these preclinical risks are communicated in product labels. Since the manufacturing process is different than that of the clinical material, the product may contain

**TABLE 5 Considerations for use of analogous products**

| Potential Advantages | Potential Disadvantages |
|---|---|
| • Useful for understanding early pharmacological properties of the drug<br>• Useful for developing early proof of concept<br>• Provides the ability "to do" toxicology studies<br>• Opportunity to use a lower order species (rodents vs. NHP)<br>• May be only feasible approach if no usable species responding to the clinical biopharmaceutical is available (e.g., no terminable studies feasible in chimpanzees, carcinogenicity testing not feasible in NHPs) | • Could have different manufacturing process and thus difference in range of process related impurities/contaminants<br>• Could have different PK<br>• Could have different pharmacological activity and/or feedback control (toxicity profile may be different if based on exaggerated pharmacology)<br>• May be different for monoclonal antibody isotype (with associated activity)<br>• Could have different intrinsic toxicity unrelated to pharmacological activity<br>• Inability to assess relative immunogenicity of the clinical product<br>• May be of questionable relevance in following up mechanism of toxicity of adverse events in clinic |

differences in impurity contaminant profiles as well as stability. More important, however, there can be significant differences in potency and pharmacological activity.

In cases where analogous products are not used to establish proof of concept for a development candidate, such information would need to be developed prior to the conduct of toxicology studies. If an analogous product is not sufficiently comparable to the clinical product but utilized in order to satisfy the assessment of general toxicity in two species, it is unclear how to interpret the relevance of any adverse findings with the analogous product especially in cases where similar findings are not observed with the clinical material in a relevant species. The cost, production capabilities, and the additional resources required for a parallel product development program of an analogous product may also be prohibitive.

### 4.5 Use of Relevant Test Systems

***ADME*** Since metabolism and formation of active metabolites are not a concern for unmodified biopharmaceuticals, mass balance studies are uninformative. Tissue concentration of radioactivity using radioactive proteins is also difficult to interpret due to unstable radiolabel linkage, rapid in vivo catabolism, and recycling of radiolabeled amino acids into non–drug-related proteins/peptides.

PK-mediated interactions are unlikely except for the rare examples of certain hormones that may be displaced by various chemicals form the tight binding of specific transport proteins, and the depressant action of certain interferons and other cytokines on hepatic P450s.

***In vitro Safety Pharmacology*** Specific guidance on safety pharmacology, ICH S7B, was published after ICH S6. This guidance specifically excluded biopharmaceuticals, acknowledging that, since biopharmaceuticals achieve highly specific receptor targeting, it has usually been sufficient to evaluate safety pharmacology endpoints as a part of toxicology and /or pharmacodynamic studies [33]. Nevertheless, assessment of in vitro cardiotoxicity, such as in vitro IKr assays (human ether a-go-go related gene, hERG), has been performed for some biopharmaceuticals even though mechanistically biopharmaceuticals are not likely to enter cells and bind a site within the potassium channel due to their molecular size. With the exception of some highly selective peptide toxins, there are no data to support binding to proteins on the cell surface mediating subsequent effects on ion channels.

***Genotoxicity*** Biopharmaceuticals do not have the same distribution properties as small molecules, and they are therefore not expected to pass through cell and nuclear membranes to interact with DNA. Experience has confirmed that the standard battery of genotoxicity assays is not relevant for products that do not directly interfere with DNA or mitosis to induce gene mutations, chromosome aberrations, or DNA damage. While studies may be applicable for protein conjugates with a chemical organic linker, consideration is warranted if there is precedence of use with the linker or if there is no evidence of degradation of the protein conjugate. Additionally, unlike small molecules where there may be a cause for concern for testing for genotoxic impurities, process-related impurities associated with biopharmaceuticals include residual host cell proteins, fermentation components, column leachables, and detergents rather than organic chemicals.

***Immunogenicity*** The injection of human proteins in sufficient quantity into animals should be expected to elicit an immunological response. Interestingly some analogous products have also induced immune responses in the derivative species. The presence of neutralizing antibody can change the PK/PD profile and thus impact exposure margins and estimates of toxicity. In early studies with biopharmaceuticals the development of antibodies in a toxicology study was considered a reason to stop studies; however, we now know that we can "dose through" in animals similar to dosing practices in humans. While the presence of antibodies in animals is generally not predictive for humans, the information has helped in defining relative immunogenicity and identifying potential consequences of an immune response such as neoantigenicity, autoantigenicity, immune complex deposition, complement activation, and antibodies crossing the placenta.

Clinically relevant antibodies include clearing antibodies, sustaining antibodies, neutralizing antibodies, and antibodies that cross-react with endogenous proteins and other molecules.

A reliable antibody screening assay capable of detecting high- and low-affinity antibodies is recommended. The sampling time for antibody assessment is also important, in allowing long washout periods in some cases in order to ensure that components in the serum sample do not prevent the assay from detecting a true positive (e.g., interference by the biopharmaceutical due to a long half-life).

Although assays for cell-mediated immunity are more complex and demanding, the possible need for them should always be considered because they may represent the source of a "toxic" action while accompanying antibodies are only a signal that an immune response has occurred.

The possibility of induced autoimmune responses should also be considered as a cause of damage to nuclei, cytoplasm, and cells and tissues as well as the stimulus for antibody formation.

***Dose: Selection, Schedule, Duration, and Response*** According to ICH S6, dose levels should be selected to provide information on a dose response including a toxic dose and NOAEL [10]. While the NOAEL is generally considered for purposes of recommending initial doses for first in human studies, it was recognized that for certain classes of drugs or biologics (e.g., vasodilator, anticoagulants, monoclonal antibodies, or growth factors) where toxicity may arise from exaggerated pharmacology that the pharmacologically active dose (PAD) in these cases may be a more sensitive indicator of potential toxicity than the NOAEL and might therefore warrant lowering the minimum recommended starting dose (MRSD) [34]. More recently it has been suggested that the starting dose for the FIH for certain classes of biopharmaceuticals should be below the MABEL (minimal anticipated biological effect level) as predicted from all the available preclinical data (in vitro in humans and in vivo in animals) [35,36]. Importantly a number of factors have to be considered to optimize dose extrapolation (see Table 6). It is also important to consider if these initial trials are performed in normal subjects or subjects with disease.

Ideally the design of the toxicology studies should mimic the clinical trial. However, because of more rapid clearance in test animals, more frequent dosing than proposed for the clinical trial may be needed to model for the equivalent exposure. Importantly, when longer intervals are anticipated, such as once a month or once every six months, a more accelerated or contracted dosing scheme may be needed as longer intervals generally serve as immunizing regimen (e.g., similar to the intervals used for vaccines intended to induce an immune response). It is also important to understand whether dosing strategies induce tolerance, as was the case for IL 12 [37].

While identifying a maximum tolerated dose is preferred when designing toxicity studies in some cases, such a dose has been difficult to achieve for

**TABLE 6 Considerations for optimizing dose extrapolation across species**

- Receptor binding affinity
- Receptor saturation
- Potency to the target
- Upregulation of soluble receptors
- Tissue expression
- Stimulation of biological cascades by agonists at immune cell receptors
- Knowledge of other downstream signaling
- Potential for biological amplification
- Delayed induction of secondary mediator release
- Temporal dissociation of pharmacological effects from plasma levels due to persistent activity in nonplasma compartments
- Redundancy
- Homeostasis

*Note*: Specific considerations based on product class.

biopharmaceuticals. The use of a limit dose has been acceptable as the high dose for pharmaceuticals if an MTD cannot be achieved (e.g., 1 g/kg/day [31] and 1.5 g/kg/day ICH S1CR [38]). For biopharmaceuticals a limit dose is often defined as a maximum feasible dose (MFD) based on the physiological properties of the test material or dosage form allied to route of administration that may impose practical limitations in the amount that can be administered. A maximum absorbed dose (MAD) or pharmacodynamic maximum response dose (PMRD) based on saturation of absorption has also generally been acceptable for pharmaceuticals. However, in cases where for biopharmaceuticals with inherently low toxicity a maximum sensible dose (MSD) (A. Pilaro, personal communication) may need to be considered, a good understanding of the mechanism of action of the product, as well as of the intended clinical subject (i.e., normal volunteer vs. patient), is needed to support the adequacy of proposed multiples (e.g., 10–20X) for the high dose. In cases where lower multiples are justified, a more conservative estimate of FIH dose could be considered. Ironically it can be more difficult to select a FIH dose for a "safe drug."

## 5 FUTURE OF PRECLINICAL SAFETY EVALUATION AND ICH S6

The speed at which clinical development is expected to take place to bring novel therapies to market is challenging not only for small start-up companies but also for large established global companies. The inability to better assess and predict product safety can lead to failures during clinical development and, occasionally, after marketing [39]. Emerging technologies and novel therapies created the need and also the opportunity for considering novel approaches to toxicological testing in efforts to improve the predictive value of the data from preclinical studies for clinical decision making.

Many novel therapies could be considered "high risk" if only based on their uniqueness and lack of precedence. The introduction of novel therapies into the clinic has been facilitated by the cooperation between industry and regulatory scientists, and an adherence to sound scientific principles, common sense, and an approach based on flexibility. The "case-by-case" approach is dependent on acceptance by both regulators and industry that the interpretation of the data has to reflect best scientific practice and that no study in experimental animals can predict with certainty the outcome when a drug is given to humans [40].

In 2007 ICH S6 was proposed for updating ("maintenance") under the auspices of the ICH. Since the optimal design of preclinical programs is experiential based, the revisions would reflect accumulated experiences as well as advancements in science over the past decade—"a knowledge transfer." There is a concern, however, that revising this document may result in formalizing the emerging increase in studies performed for biopharmceuticals. The recent increase in the number and types of studies over the last few years has been attributed in part to both industry practice and regulatory expectations at better "aligning" preclinical development programs of biopharmaceuticals with pharmaceutical without consideration of the specific product attributes. As previously mentioned, flexibility in program design may also be needed for novel pharmaceuticals that share product attributes similar to biopharmaceuticals such as species specificity. The revised guidance could reduce the current flexibility in programs if new programs are expected to follow the "case studies" discussed in the guidance. The specific examples may not be applicable to either current or future novel products.

On the other hand, a revision may be an opportunity to redirect the emerging trend of increasing studies and increasing regional guidance. It could discourage the application of ICH guidance documents where biopharmaceuticals are specifically excluded and provide a better understanding of current regulatory expectations. Specific sections could be clarified to ensure optimal interpretation, for example, of species selection, immunogenicity assessment, duration of chronic studies, the number of species needed to assess chronic toxicity, the use of analogous products including interpretation of findings when an analogous product is not sufficiently comparable especially when toxicity differs significantly from results obtained with the clinical product, reproductive toxicity assessment strategies, carcinogenicity assessment strategies including whether "enhanced" chronic studies could inform the carcinogenicity section of the label. Other sections could be added, for example, on dose selection to support the first in human dose based on pharmacological activity, developmental immunotoxicity testing for immunomodulatory products, and information regarding risk management and communication.

Alternatively, ICH S6 could serve as an umbrella document. More specific guidance regarding the types of studies and areas of concern could then be provided as addendums or annexes with respect to product class, for example,

monoclonal antibodies, protein conjugates, growth factors, soluble receptors, oligonucleotide drugs.

## 6 CONCLUSION

Supported by advances in biotechnology and other enabling technologies, the practice of toxicology has shifted form a ritualistic standards-based approach to a rational science-based approach, which is best defined by studies designed to ask specific questions with an understanding of what need to be known and the ability to make acceptable assumptions. Studies are data driven and practical to obtain maximum information. Designs are modified, based on additional information. Identification of limitations and knowledge gaps are acknowledged and identified, and effort to develop new models to replace outdated models embracing novel technologies is an ongoing process.

Development principles for preclinical safety assessment have been and will continue to be a dynamic process that is strongly controlled by the expanding knowledge and improvements in product design. The full investigation of the potential usefulness of biopharmaceuticals requires the development of reliable animal model systems that allow assessment of toxicity and provide pharmacokinetic data that can be successfully scaled to humans in order to reduce risk factors before clinical testing. The design of relevant preclinical safety evaluation programs is consistent with global initiatives to facilitate and to improve clinical development programs. In the coming years stakeholders will be facing the issue of how to implement preclinical development programs for biopharmaceuticals and pharmaceuticals that better anticipate adverse effects, including development of new test systems that produce reliable results faster and at lower cost. Hopefully, preclinical evaluation programs will evolve and mature concurrently with more novel products and will focus on improving the predictive value of preclinical safety testing, targeting the toxicity testing to provide information to ensure that the correct data are collected from the most appropriate studies.

## ACKNOWLEDGMENT

The author would like to thank Dr. Richard Lewis and Dr. Tony Dayan for critically reviewing the manuscript.

## REFERENCES

1. Olson H, Betton G, Robinson D, Thomas K, Monro A, Kolaja G, Lilly P, Sanders J, Sipes G, Bracken W, Dorato M, Van Deun K, Smith P, Berger B, Heller A. Con-

cordance of the toxicity of pharmaceuticals in humans and animals. *Regul Toxicol Pharmacol* 2000;32:56–67.

2. Goldenthal EI. Current views on safety evaluation of drugs. FDA Papers May 1968.
3. Hayes TJ, Cavagnaro JA. Progress and challenges in the preclinical assessment of cytokines. *Toxicol Letts* 1992;64/65:291–7.
4. Zbinden G. Biotechnology Products Intended for Human Use, Toxicological Targets and Research Strategies. In CE Grahm (ed) *Preclinical Safety of Biotechnology Products Intended for Human Use: Clinical and Biological Research*, Vol. 235, Alan R. Liss, New York, 1987, pp 143–59.
5. Giss HE. Foreword. In CE Grahm (ed) *Preclinical Safety of Biotechnology Products Intended for Human Use: Clinical and Biological Research*, Vol. 235, Alan R. Liss, New York, 1987, pp xiii–xv.
6. Dayan AD. Rationality and regulatory requirements—A view from Britain. In CE Grahm (ed) *Preclinical Safety of Biotechnology Products Intended for Human Use: Clinical and Biological Research*, Vol. 235, Alan R. Liss, New York, 1987, pp 89–106.
7. Galbraith WM. Symposium discussion. In CE Grahm (ed) *Preclinical Safety of Biotechnology Products Intended for Human Use: Clinical and Biological Research*, Vol. 235, Alan R. Liss, New York, 1987, pp 189–206.
8. Finkle BS. Genetically engineered drugs: Toxicology with a difference. In CE Grahm (ed), *Preclinical Safety of Biotechnology Products Intended for Human Use: Clinical and Biological Research*, Vol. 235, Alan R. Liss, New York, 1987, pp 161–7.
9. Griffith SA, Lumley CE. Non-clinical safety studies for biotechnologically-derived pharmaceuticals: Conclusions for an International workshop. *Hum Exper Toxicol* 1998;17:63–83.
10. ICH S6 Preclinical Safety Evaluation of Biotechnology-Derived Pharmaceuticals. http://www.fda.gov/cder/guidance/index.htm.
11. ICH M3 Nonclinical Safety Studies for the Conduct of Human Clinical Trials for Pharmaceuticals. http://www.fda.gov/cder/guidance/index.htm.
12. DeGeorge JJ, Ahn C-H, Andrews PA, Brower ME, Giorgio DW, Goheer MA, Lee-ham DY, McGuinn WD, Schimdt W, Sun CJ, Tripathi SC. Regulatory considerations for preclinical development of anticancer drugs. *Cancer Chemother Pharmacol* 1998;41:173–85.
13. Dayan AD. Testing biotechnology recombinant DNA (rDNA) products. *BIRA J* 1990;9:7–10.
14. Zbinden G. Safety evaluation of biotechnology products. *Drug Safety* 1990;5(Suppl. 1):58–64.
15. Zbinden G. Predictive value of animal studies in toxicology. *Regul Toxicol Pharmacol* 1991;14:167–77.
16. Bass R, Kleeburg U, Schroder H, et al. Current Guidelines for the Preclinical Safety Assessment of Therapeutic Proteins. *Toxicol Letts* 1992;64/65:339–47.
17. Cavagnaro JA. Science-based approach to preclinical safety evaluation of biotechnology products. *Pharmaceutl Eng* 1992;12:32–3.

18. Claude JR. Difficulties in conceiving and applying guidelines for the safety evaluation of biotechnologically-produced drugs: Some examples. *Toxicol Letts* 1992; 64/65:349–55.
19. Terrell TG, Green JD. Issues with biotechnology products in toxicologic pathology. *Toxicol Pathol* 1994;22:187–93.
20. Cavagnaro JA. Preclinical safety assessment of biological products. In Mathieu M, ed. *Biologics Development: A Regulatory Overview*. Waltham, MA: Parexel International, 1993;23–40.
21. Sims J. Assessment of biotechnology products for therapeutic use. *Toxicol Letts* 2001;120:59–66.
22. Dayan AD. Safety evaluation of biological and biotechnology-derived medicines. *Toxicology* 1995;105:59–68.
23. Thomas JA. Recent developments and perspectives of biotechnology-derived products. *Toxicology* 1995;105:7–22.
24. Serabian MA, Pilaro AM. Safety assessment of biotechnology-derived pharmaceuticals: ICH and beyond. *Toxicol Pathol* 1999;27:27–31.
25. Ryan AM, Terrell TG. Biotechnology and its products. In *Handbook of Toxicologic Pathology*, 2nd ed. San Diego: Academic Press, 2002;479–500.
26. Verdier F. Preclinical safety assessment of vaccines in biotechnology and safety assessment. In Thomas JA, Fuchs RL, eds. *Biotechnology and Safety Assessment*. 3rd ed. Academic Press, San Diego, 2002;397–412.
27. Brennan FR, Shaw L, Wing MG, Robinson C. Preclinical safety testing of biotechnology-derived pharmaceuticals. *Mol Biotechnol* 2004;27:59–74.
28. Nakazawa T, Kai S, Kawai M, Maki E, Sagami F, Onodera H, Kitajima S, Inoue T. "Points to Consider" regarding safety assessment of biotechnology-derived pharmaceuticals in non-clinical studies (English Translation). *J Toxicol Sci* 2004;29: 497–504.
29. Cavagnaro JA. Predicting safety for novel therapies. *Regul Affairs J* 2006;(June): 1–4.
30. Snodin DJ, Ryle PR. Understanding and applying regulatory guidance on the non-clinical development of biotechnology-derived pharmaceuticals. *Biodrugs* 2006;1: 25–52.
31. ICH S5A Detection of Toxicity to Reproduction for Medicinal Products, September 1994.
32. BLAST. http://www.ncbi.nlm.nih.gov/Education/blasttutorial.html.
33. ICH S7B Nonclinical Evaluation of the Potential for Delayed Ventricular Repolarization (QT Interval Prolongation) by Human Pharmaceuticals, October 2005. http://www.fda.gov/cder/guidance/index.htm.
34. Estimating the Maximum Safe Starting Dose in Initial Clinical Trials for Therapeutics in Adult Healthy Volunteers, July 2005.
35. Early Stage Clinical Trial Taskforce, Joint ABPI/BIA Report, July 2, 2006. http://www.abpi.org.uk/information/pdfs/BIAABPI_taskforce2.pdf.
36. Expert Scientific Group on Phase One Clinical Trials, Final Report ("Duff Report"), November 30, 2006. http://www.dh.gov.uk/assetRoot/04/14/10/43/04141043.pdf.

37. Leonard JP, Sherman MT, Fisher GL, Buchanan LJ, Larsen G, Atkins MB, Sosman JA, Dutcher JP, Vogelzang NJ, Ryan JL. Effects of single dose interleukin-12 exposure on interleukin-12 associated toxicity and interleukin-γ production. *Blood* 1997;90:2541–8.
38. ICH S1C(R) Guidance on Dose Selection for Carcinogenicity Studies of Pharmaceuticals: Addendum on a Limited Dose and Related Notes, December 1997 http://www.fda.gov/cder/guidance/index.htm.
39. FDA Challenge and Opportunity on the Critical Path to New Medical Products, March 2004. http://www.fda.gov/oc/initiatives/criticalpath/whitepaper.html.
40. Academy of Medical Sciences Forum Safer Medicines Report. Pre-clinical Toxicology Working Group Report. November 2005.

# Implementation of ICH S6: EU Perspective

PETER R. RYLE, PhD, DipRCPath (Tox), FRCPath, and DAVID J. SNODIN, PhD, FRSC, MChemA, MSc

**Contents**

1 Introduction
2 European Union
3 EU Drug Regulatory Framework
  3.1 Pharmaceutical Legislation and Regulation
  3.2 European Medicines Agency and National Competent Authorities
  3.3 Marketing Authorization Applications (MAAs)
  3.4 Clinical Trial Authorizations (CTAs)
  3.5 Scientific Advice
4 The European Union and ICH
5 EU-Specific Regulatory Guidance
6 Application of the ICH S6 Guideline in the European Union
  6.1 General Comments on Application of the Guidance
  6.2 "High-Risk" Compounds and Clinical Trials
  6.3 Primate Supply and Use
  6.4 Duration of Chronic Toxicity Studies
  6.5 Comparability Testing of Biopharmaceutical Products
7 Concluding Remarks
  References

## 1 INTRODUCTION

The development and regulatory approval of biopharmaceuticals in the European Union (EU) has lagged slightly behind that in the United States. But

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

over the last decade this class of drug has assumed increasing importance in the development pipelines of EU-based pharmaceutical companies. In addition, many US and Japanese companies conduct development of their products in Europe for the EU market. Some countries in the European Union have acquired a reputation for innovation in the biopharmaceutical field, with various clusters of start-up and small biotechnology companies becoming established, particularly in the United Kingdom (mainly around the university cities of Oxford and Cambridge) and Germany. More recently France, the Netherlands, the Scandinavian countries (Sweden, Denmark, Norway, Finland), and smaller countries such as Belgium and Ireland can claim to have flourishing biotechnology industries, producing potential biopharmaceutical development candidates. Although Switzerland is not a member of the European Union, its location and the presence of both long-established large multinational pharmaceutical, as well as some smaller biotechnology start-up companies, make it a major contributor to the biopharmaceutical sector in Europe.

Examples of biopharmaceutical products that have received approval in the European Union are shown in Table 1. In addition to the approval of novel biopharmaceutical products, the first EU approval of a "generic" or "biosimilar" biopharmaceutical product (Omnitrope®) was granted in 2006 [1]. In general, most major new biopharmaceutical products that are approved by the US regulatory authority (FDA) are approved in the European Union at about the same time, or shortly after FDA approval for marketing. There are relatively few biopharmaceutical products that undergo region-specific development and marketing, largely due to the high cost of development of these products, and reflecting the fact that the drugs are often for areas of unmet medical need. In addition the EU population (slightly under 500 million) makes it one of the most attractive potential markets for new pharmaceutical products. Therefore the trend of an increasing proportion of new drugs in development or receiving EU marketing approval that are biotechnology derived probably reflects the situation in the United States and other regions. In 2004, 8 of 19 new therapeutics approved for use in the United Kingdom were biopharmaceuticals.

In terms of the application and implementation of ICH guidelines in the European Union, the approach from a technical and scientific viewpoint can be regarded as similar to that in the other major ICH regions (the United States and Japan). The major difference relates to the regulatory administrative structure within the European Union, as will be discussed below. Sponsors developing biopharmaceuticals for the US and Japanese markets will deal with a single regulatory agency for those regions throughout the development cycle. The EU situation is rather more complex, such that early in the development cycle the sponsor may interact with one or two EU national authorities to enable phase 1 and 2 trials to be initiated. As the development cycle proceeds, interactions with more national authorities and the central European Medicines Agency (EMEA) are likely to take place, working toward approval of the product by the central agency, through the so-called centralized procedure.

**TABLE 1 Examples of biopharmaceuticals approved in the European Union**

| Product Name | Action | Company | Therapeutic Use | Year Approved |
|---|---|---|---|---|
| Actrapid | Insulin analogue | Novo Nordisk | Diabetes mellitus | 2002 |
| Avastin | Anti-VEGF monoclonal Ab | Roche | Colo-rectal carcinoma | 2005 |
| Avonex | Interferon-α | Biogen IDEC | Multiple sclerosis | 1997 |
| Betaferon | Interferon-β | Schering | Multiple sclerosis | 2000 |
| Cerezyme | β-cerebrosidase | Genzyme | Gaucher's disease | 1997 |
| Dynepo | Epoetin | Shire | Anaemia | 2002 |
| Enbrel | Anti-TNFα fusion protein | Wyeth | Rheumatoid arthritis/psoriasis | 2000 |
| Erbitux | Anti-EGF monoclonal Ab | Merck | Colo-rectal carcinoma | 2004 |
| Herceptin | Anti-HER2 monoclonal Ab | Roche | Breast cancer | 2000 |
| Humalog | Lispro insulin analogue | Eli Lilly | Diabetes mellitus | 1996 |
| Humira | Anti-TNFα monoclonal Ab | Abbott | Rheumatoid arthritis | 2003 |
| Intron A | Interferon-α | Schering-Plough | Hepatitis/lymphoid tumors | 2000 |
| Kineret | IL-1 antagonist | Amgen | Rheumatoid arthritis | 2002 |
| Liprolog | Insulin analogue | Eli Lilly | Diabetes mellitus | 2001 |
| Mabcampath | Anti-CD52 monoclonal Ab | Genzyme | Chronic leukaemia | 2001 |
| Mabthera | Anti-CD20 monoclonal Ab | Roche | Rheumatoid arthritis/lymphoma | 1998 |
| Neulasta | PEG-G-CSF | Amgen | Neutropenia | 2002 |
| NovoSeven | Factor VIIa | Novo Nordisk | Haemophilia | 1996 |
| Osigraft | Osteogenic Protein-1 | Howmedica | Bone non-union | 2001 |
| Pegasys | PEG-interferon alpha | Roche | Chronic hepatitis | 2002 |
| Raptiva | Anti-CD11a monoclonal Ab | Serono | Psoriasis | 2004 |
| Remicade | Anti-TNFα monoclonal Ab | Centocor BV | Rheumatoid arthritis/Crohn's | 1999 |
| Simulect | Anti-CD25 monoclonal Ab | Novartis | Immune suppression | 1998 |
| Tysabri | Anti-α4 integrin monoclonal Ab | Elan | Multiple sclerosis | 2006 |
| Xolair | Anti-IgE monoclonal Ab | Novartis | Asthma | 2005 |

This chapter will therefore concentrate mostly on the major differences in regulatory structure between the European Union and other regions, which has bearing on obtaining regulatory guidance and scientific advice during the development of a biopharmaceutical for the European market. Such interactions with the regulatory agencies are crucial to the application of the case-by-case approach to preclinical development of biopharmaceuticals advocated by the ICH S6 guideline.

## 2 EUROPEAN UNION

The European Union per se was created in November 1993 when the Treaty on European Union came into effect. At that time 12 European countries had already joined the European Community (EC), formed in 1967 as a merger of the European Coal and Steel Community (ECSC, founded in 1952), the European Economic Community (EEC, founded in 1958), and the European Atomic Energy Community. The 12 countries that were part of the European Union in 1993 were Belgium, France, Italy, Luxembourg, the Netherlands, West Germany, Denmark, Ireland, the United Kingdom, Greece, Portugal, and Spain.

Since 1993 other countries have joined the European Union, with Austria, Finland, and Sweden joining in 1995, Cyprus, Czech Republic, Estonia, Hungary, Latvia, Lithuania, Malta, Poland, Slovakia, and Slovenia joining in 2004, and Bulgaria and Romania joining in January 2007. At the present time, there are 27 countries or "member states" in the European Union. At the time of writing, Turkey and Croatia are in "accession" negotiations to join the European Union, with several other countries, such as Macedonia, Albania, Montenegro, and Serbia, as potential future member states.

The European Union has a number of high-level institutions, the most important being as follows:

*The Council of Ministers.* The Council is the main decision-making body of the European Union. The ministers of the member states meet within the Council, and depending on the issue on the agenda, each country will be represented by the minister responsible for that subject (foreign affairs, finance, social affairs, transport, agriculture, etc.). The presidency of the Council is held for six months by each member state on a rotational basis.

*The European Parliament.* Following elections held in June 2004, the European Parliament had 732 members elected in the 25 member states of the European Union. Each member is elected for a five-year term. Most of the time Parliament and its members are based in Brussels where its specialist committees meet to scrutinize proposals for new EU laws.

*The European Commission.* Based in Brussels, the Commission consists of 27 commissioners, one from each member state, appointed for a five-year term by member states' governments. The Commission, which acts as the

European Union's civil service, and comprises approximately 14,000 officials, has the right of initiative, that is, to draw up proposals for Union legislation. The Commission negotiates on behalf of the member states in multilateral and bilateral trade matters and in the drawing up of association and membership agreements with nonmember countries.

*The European Court of Justice.* The European Court of Justice, consisting of 15 judges (appointed by the Member States) and 9 Advocates-General, is based in Luxembourg. It is responsible for arbitrating in disputes relating to the interpretation and application of EU legislation.

## 3 EU DRUG REGULATORY FRAMEWORK

### 3.1 Pharmaceutical Legislation and Regulation

Pharmaceutical regulation in the European Union is applied centrally and nationally in individual member states. In general, after joining the European Union, a member state is required to harmonize or revoke its existing legislation and to incorporate any new provisions agreed centrally into its national legislation. European Union pharmaceutical legislation dates back to the 1960s; Directive 65/65/EEC remained the basis of European Union rules for many years. This has been superseded by Directive 2001/83/EC (e.g., as amended by 2004/27/EC) relating to medicinal products for human use, which essentially consolidated previous relevant directives into one text. As explained in the preamble to Directive 2001/83/EC, the key aim of EU pharmaceutical legislation is the protection of public health, provided that this is achieved by means that will not hinder drug development. A committee system ("comitology") is often used in the European Union as a vital part of the adoption and implementation of legal statutes. Most EU legislation is passed by the Commission under powers delegated by the Council of Ministers, and in such cases there is no formal involvement of the general public, national parliaments, or the European Parliament. One example in drug regulation is Annex I to Directive 2001/83/EC, an annex containing a variety of definitions and concepts that forms the basis of technical requirements involved in drug regulation. An updated Annex 1 (Directive 2003/63) is now in operation [2].

Other important legal documents are as follows:

- Directive 2001/20/EC, which mainly concerns approval of clinical trials in the European Union.
- Council Regulation 726/2004/EC (updates and expands the previous regulation 239/1993/EC), which mainly concerns the duties of the European Medicines Agency (EMEA), establishment of the Committee for Medicinal Products for Human Use (CHMP), and the scope and mechanisms of the centralized procedure for marketing authorization applications (MAAs) and pharmacovigilance issues.

All the above-mentioned legal treatises apply to the European Economic Area (EEA; this comprises the 27 EU member states plus Liechtenstein, Iceland, and Norway—Switzerland chose by referendum not to join the EEA, but has certain Swiss–EU bilateral agreements). Switzerland acts as the representative of the European Free Trade Association, which is an observer in the ICH (International Conference on Harmonization) process, accepts ICH guidelines, but is not part of the European Economic Area. There is a legal obligation for the Swiss competent authority, Swiss Agency for Therapeutic Products (Swissmedic), to take account of decisions/authorizations in other territories that have equivalent medicinal product control.

### 3.2 European Medicines Agency and National Competent Authorities

Each member state has a national medicines agency, which is usually located in the capital city (but not, e.g., in Germany, the Netherlands, and Sweden). These agencies are listed in Table 2. Some national agencies combine the

**TABLE 2 Regulatory bodies (competent authorities) for human medicines in Europe**

| Country | Agency | Web Site |
|---|---|---|
| Austria | Bundesamt für Sicherheit im Gesundheitswesen | www.ages.at |
| Belgium | Directoraat generaal Geneesmiddelen<br>Direction générale Médicaments | www.afigp.fgov.be |
| Denmark | Lægemiddelstyrelsen | www.dkma.dk |
| Europe | European Medicines Agency (EMEA) | www.emea.europa.eu |
| Finland | Lääkelaitos | www.nam.fi |
| France | Agence Française de Sécurité Sanitaire des Produits de Santé | www.afssaps.sante.fr |
| Germany | Bundesministerium für Gesundheit und Soziale Sicherung<br>Bundesinstitut für Arzneimittel und Medizinprodukte<br>Paul-Ehrlich-Institut | www.bmgs.bund.de<br>www.bfarm.de/de/index.php<br>www.pei.de |
| Germany | Zentralstelle der Länder für Gesundheitsschutz bei Arzneimitteln und Medizinprodukten | www.zlg.nrw.de |
| Greece | National Organization for Medicines | www.eof.gr |
| Iceland | Lyfjastofnun | www.lyfjastofnun.is |
| Ireland | Irish Medicines Board | www.imb.ie |
| Italy | Ministero della Salute | www.ministerosalute.it |
| Liechtenstein | Liechtensteinische Landesverwaltung<br>Amt für Lebensmittelkontrolle und Veterinärwesen<br>Kontrollstelle für Arzneimittel | www.llv.li |

**TABLE 2** ***Continued***

| Country | Agency | Web Site |
|---|---|---|
| Luxembourg | Ministère de la Santé Division de la Pharmacie et des Médicaments | www.etat.lu/MS |
| Netherlands | Staatstoezicht op de volksgezondheid Inspectie voor de Gezondheidszorg | www.igz.nl |
| Netherlands | College ter Beoordeling van Geneesmiddelen (CBG) | http://www.cbg-meb.nl |
| Norway | Statens Legemiddelverk | www.legemiddelverket.no |
| Portugal | Instituto Nacional da Farmácia e do Medicamento | www.infarmed.pt |
| Spain | Agencia española del medicamento | www.agemed.es |
| Sweden | Läkemedelsverket | www.mpa.se |
| United Kingdom | Medicines and Healthcare products Regulatory Agency | www.mhra.gov.uk |
| Bulgaria | Bulgarian Drug Agency | www.bda.bg |
| Cyprus | Ministry of Health | www.pio.gov.cy |
| Czech Repbulic | State Institut for Drug Control | www.sukl.cz www.uskvbl.cz |
| Estonia | State Agency of Medicines | www.sam.ee |
| Hungary | National Institute of Pharmacy | www.ogyi.hu |
| Latvia | Food and Veterinary Service | zaale.vza.gov.lv |
| Lithuania | State Medicines Control Agency | www.vvkt.lt |
| Malta | Medicines Authority | www.medicinesauthority.gov.mt |
| Poland | Office for Medicinal Products | http://www.urpl.gov.pl |
| Romania | National Medicines Agency | www.anm.ro/home.html |
| Slovak Republic | State Institute for Drug Control | www.sukl.sk |
| Slovenia | Agency of the Republic of Slovenia for Medicinal Products and Medical Devices | http://www2.gov.si/mz/mz-splet.nsf |
| Switzerland | Swissmedic, Schweizerisches Heilmittelinstitut | www.swissmedic.ch |

regulation of medicines and medical devices (e.g., Medicines and Healthcare products Regulatory Agency, called MHRA, in the United Kingdom) while others regulate both human and veterinary medicines. The European Medicines Agency (EMEA) is a pan-European regulatory agency, responsible for both human and veterinary medicines, located in Canary Wharf on the east side of London. It was established on February 1, 1995, and its staff numbers have grown year by year; the EMEA budget for 2007 makes provision for up to 441 staff. The EMEA functions in a different manner to the US Food and Drug Administration. Its role is mainly in policy and coordination while

technical expertise is provided by staff from the national agencies [3]. There is a network of approximately 3500 European experts that are drawn mainly from the national agencies and academia [4] whose main roles relate to undertaking assessment of application dossiers, providing scientific advice, developing new guidelines, and sitting on various committees, working groups, and advisory groups. The mission statement and main tasks and responsibilities of the EMEA are shown in Table 3. More details on committees, working groups, and scientific advisory groups are shown in Table 4.

**TABLE 3 EMEA: Mission statement and main responsibilities**

| *Mission statement* |
|---|
| In the context of a continuing globalization, to protect and promote public and animal health:<br>• By developing efficient and transparent procedures to allow rapid access by users to safe and effective innovative medicines and to generic and nonprescription medicines through a single European marketing authorization<br>• By controlling the safety of medicines for humans and animals, in particular, through a pharmacovigilance network and the establishment of safe limits for residues in food-producing animals<br>• By facilitating innovation and stimulating research, hence contributing to the competitiveness of EU based pharmaceutical industry<br>• By mobilizing and coordinating scientific resources from throughout the European Union to provide high-quality evaluation of medicinal products, to advise on research and development programs, to perform inspections for ensuring fundamental GXP provisions are consistently achieved, and to provide useful and clear information to users and health-care professionals |
| *Main tasks and responsibilities* |
| • Scientific advice to member states and Community institutions on quality, safety, and efficacy of human and veterinary medicinal products<br>• Centralized (and to a lesser extent decentralized) authorization procedures: administration of these procedures to achieve a single evaluation and marketing authorization for medicinal products<br>• Pharmacovigilance and inspection: organization of procedures for effective surveillance (and possibly withdrawal) of medicinal products in the European Union, and to reinforce national inspection activities<br>• Advice to companies on drug development (scientific advice and protocol assistance)<br>• Committees, working parties, and scientific advisory groups<br>• Guidance documents<br>• EPARs (European Public Assessment Reports) and Withdrawal Public Assessment Reports<br>• List of European experts |

*Note*: GXP means "good clinical practice" (GCP), "good manufacturing practice" (GMP), and "good laboratory practice" (GLP) collectively.

**TABLE 4 EMEA: Committees, working groups, and scientific advisory groups on human medicines**

| *Committees* |
|---|
| • Management Board<br>• Committee on Human Medicinal Products (CHMP)<br>• Committee on Orphan Medicinal Products (COMP)<br>• Committee on Herbal Medicinal Products (HMPC) |
| *Working parties (WPs)* |
| • Safety Working Party (SWP)<br>• Quality Working Party (QWP)<br>• Efficacy Working Party (EWP)<br>• Other WPs: Pharmacogenetics (PgWP), Pharmacovigilance (PhVWP), Biologics (BWP), Gene Therapy (GTWP), Vaccines (VWP), Blood Products (BPWP), Cell-Based Products (CPWP), Scientific Advice (SAWP)<br>• Temporary WPs: Pediatric (PEG), Similar Biological (Biosimilar) Medicinal Products (BMWP) |
| *Scientific advisory groups (SAGs)* |
| • The role of the SAGs is to provide, on request from the committee concerned, an independent recommendation on scientific and technical matters relating to products under evaluation or any other scientific issues relevant to the work of the respective committees. While views expressed by SAGs are taken into account, the ultimate responsibility for the final opinions rests with the respective scientific committee.<br>• SAGs created to date: Oncology, Diagnostics, Anti-Infectives, Diabetes/Endocrinology, Cardiovascular Issues, Central Nervous System, HIV/Viral Diseases |

### 3.3 Marketing Authorization Applications (MAAs)

Four routes are available for obtaining a marketing authorization (MA) for a human medicinal product in the European Economic Area: independent national procedure, mutual recognition procedure (MRP), decentralized procedure (DCP), and centralized procedure (CP) [5–7].

***Independent National Procedure*** If a company wishes to market a product in one country only and there are no legal obligations to use a route other than the national one (which would apply if the type of drug or therapeutic area were within the scope of the centralized procedure), then an application can be made to one health authority, leading to a marketing authorization in that country alone.

***Mutual Recognition Procedure (MRP)*** This procedure begins with a national application to one member state (MS), and when a marketing

authorization has been obtained the applicant applies for "mutual recognition" of this initial authorization by the reference member state (RMS) in some or all of the other European economic area countries, called concerned member states (CMSs).

***Decentralized Procedure (DCP)*** The decentralized procedure was established as an application route in late 2005. This procedure is essentially a combination of the national and the mutual recognition procedures. The applicant chooses a reference member state to undertake the initial assessment. On completion of the RMS assessment, the concerned member states can put forward additional questions to those raised by the reference member state. If the questions are answered by the applicant to the satisfaction of the reference member state and concerned member states, the drug can be authorized. The benefit of the decentralized procedure is that all concerned member states are provided with the application dossier from the start of the procedure.

***Centralized Procedure (CP)*** This is the procedure of most interest for biopharmaceuticals, as this is the mandatory route for review and approval of such drugs in the European Union. In the centralized procedure a single application is submitted to the European Medicines Agency (EMEA). A variety of presubmission activities, starting six months before the intended start date of the centralized procedure, are required [8]. Two initial assessments by a "Rapporteur" and "Co-rapporteur" national authorities (one from each of two member states chosen by the EMEA) are made, leading to Day 80 Critical Assessment Reports. A consolidated list of questions (LoQ) is provided to the applicant at Day 120 when there is a clock stop, normally of three months, to allow the preparation and submission of responses. Following satisfactory negotiation of other steps in the procedure, the Committee on Human Medicinal Products will recommend authorization at Day 210, with authorization by the Commission at Day 277.

The centralized procedure is mandatory for the following cases:

- Medicinal products developed by means of one of the following biotechnological processes: recombinant DNA technology, controlled expression of genes coding for biologically active proteins in prokaryotes and eukaryotes including transformed mammalian cells, hybridoma and monoclonal antibody methods.
- New chemical entities for the following indications: acquired immunodeficiency syndrome (AIDS), cancer, neurodegenerative diseases, diabetes. Two more categories (autoimmune diseases and other immune dysfunctions, and viral diseases), will be included with effect beginning May 20, 2008.
- Orphan medicinal products.
- Biosimilar products.

The current CP application fee is 232,000 Euros for a single strength associated with one pharmaceutical form. Although the fees are high and the Committee on Human Medicinal Products review is extremely thorough, if successful the centralized procedure leads to an authorization that is valid in all European Economic Area countries, and for this reason is attractive to companies that have a presence in only one or a few European countries.

### 3.4 Clinical Trial Authorizations (CTAs)

Clinical trials are regulated by individual member states in the European Union. An applicant (or sponsor) submits data on the investigational medicinal product (IMP), and details of the proposed clinical trial, to the competent authority in the country in which the trial is to be carried out. The ethics committee responsible for the site where the trial is to take place also needs to give approval.

Clinical trials and clinical trial authorizations in the European Union are controlled under the Clinical Trial Directive, 2001/20/EC [9], and all member states are bound by its requirements. Under the provisions of the Directive, a clinical trial is an investigation in human subjects that is intended to discover or verify the clinical, pharmacological, and/or other pharmacodynamic effects of one or more medicinal products, identify any adverse reactions or study the absorption, distribution, metabolism, and excretion, with the object of ascertaining the safety and/or efficacy of those products. This definition includes pharmacokinetic studies.

All submissions need to include a completed application form and, prior to submitting the application, a reference number (EudraCT number) must be obtained [10]. In addition the applicant needs to supply additional documents (depending on the type of trial and investigational medicinal product), but normally including:

- Investigator's brochure (IB)
- Clinical trial protocol
- Investigational medicinal product dossier (IMPD)
- Subject information leaflet
- Informed consent

Detailed guidance on the preparation of the Investigational Medicinal Product Dossier is available [11]. In terms of preclinical aspects, the key considerations are:

- The investigational medicinal product dossier can be stand-alone or be constructed by cross-reference to relevant sections of the IB (applies to preclinical and clinical data only).
- Summaries of preclinical data should be provided, preferably using tabular formats. The headings for the preclinical part should follow those for

written summaries in the Common Technical Document (CTD) module 2.6. In contrast to the situation with the US FDA, full study reports and individual data (line listings) for the nonclinical studies are not required.

- An overall risk assessment should be included. This is intended to be a brief integrated summary that analyzes the preclinical and clinical data in relation to the potential risks and benefits of the proposed trial. Safety margins should be expressed on the basis of relative systemic exposure rather than applied dose.

The CTA approval process should take no longer than 60 days. Approval times are often shorter (e.g., 14–21 days for a healthy volunteer trial in the United Kingdom) and vary among member states.

### 3.5 Scientific Advice

The case-by-case approach that is needed for the effective preclinical development of biopharmaceuticals requires close collaboration and agreement between the sponsor and the regulatory agency at all stages of development. In the United States, such collaboration and scientific advice is often achieved through interaction with the FDA at pre-IND, end of phase 2, or pre-NDA meetings with the Agency.

In the European Union, scientific advice can be sought from national agencies at any stage of drug development [12,13]. Details are provided on agency Web sites (e.g., MHRA) [14], and fees are charged depending on the nature of the advice requested (one or more of quality, preclinical, clinical, regulatory, pharmacovigilance plans). Scientific advice (called "protocol assistance" in the case of orphan drugs) can also be obtained centrally from the EMEA. The concepts are similar to those for national scientific advice, but the procedures are considerably more complex (sometimes involving joint sessions with FDA), of longer duration, and more costly in terms of applicant fees [15]. EU guidance emphasizes that it is not the role of the Committee on Human Medicinal Products (CHMP) to substitute the industry's responsibility in the development of their products, and that any advice given is not legally binding with respect to any future marketing authorization application for the product concerned, although in the Notice to Applicants it is a requirement to include a copy of any formal scientific advice in the application dossier.

In terms of the particular questions asked as part of EMEA scientific advice, questions that can be answered by consulting a particular guideline are discouraged, although a question based on interpretation of a guideline would probably be acceptable. There is no direct equivalent of the US FDA pre-IND or end-of-phase-2 meeting in Europe. Since clinical trials are regulated on a national basis, and the review times for clinical trial submissions are quite short, most companies would submit an application for a CTA without holding a meeting beforehand, unless the treatment in question was poten-

tially controversial (e.g., involving a monoclonal antibody, gene therapy, or cell therapy). Once a program has negotiated phase 1, obtaining scientific advice either nationally and/or centrally would be strongly considered by most companies.

## 4 THE EUROPEAN UNION AND ICH

The European Union was and still is a major contributor to ICH, through the input of the European Commission (which represents the 27 member states) and the European pharmaceutical industry (European Federation of Pharmaceutical Industries and Associations, EFPIA) to the process. European technical and scientific support for the ICH process was provided by the Committee for Medicinal Products for Human Use (CHMP, formerly CPMP) of the European Medicines Agency (EMEA), on behalf of the European Commission. Therefore EU-based regulatory and industry professionals were actively involved in the development of the ICH guidelines, including the S6 guideline that covers preclinical testing of biotechnology-derived pharmaceuticals. The content and principles of the ICH S6 guideline have been covered in detail in other chapters, so it will not be addressed further here. Obviously the key message from the ICH S6 guidance was the need for a case-by-case approach to the preclinical testing of biopharmaceuticals, taking into account factors such as species specificity and immunogenicity of the products.

Prior to the implementation of the ICH S6 [16] guideline, an EU guideline covering the preclinical testing of biotechnology-derived drugs [17] was adopted in 1988. The ICH S6 guideline was approved by the CPMP of the EU regulatory agency (EMEA) in September 1997 and came into operation in member states in March 1998. The implementation of the ICH S6 guideline effectively replaced the previous 1988 guidance document. At the time of adoption of the ICH S6 guideline, there was only one other document in place that provided some preclinical guidance on biopharmaceuticals, a guideline covering the production and quality control of monoclonal antibodies [18]. This guidance recommended the use of tissue cross reactivity studies that are usually performed on monoclonal antibody products. A more recent initiative to update the EU monoclonal antibody guidance indicated that the update should concentrate on quality aspects only, since preclinical testing of these products was adequately addressed by adoption of the ICH S6 guidance [19].

## 5 EU-SPECIFIC REGULATORY GUIDANCE

Since the adoption of the ICH S6 guideline, and with the growing number of biopharmaceuticals entering development, the EMEA have developed a number of specific guidelines in relation to the preclincial testing of biopharmaceuticals. The current draft and approved EU-specific preclinical guidance documents that apply to biopharmaceuticals are listed in Table 5. In addition

**TABLE 5 EU-specific preclinical guidelines on biopharmaceuticals**

| Relevant Product Type | Guideline | Date |
|---|---|---|
| Vaccines (excluding DNA or viral vector vaccines) | CPMP/SWP/465/95: Preclinical pharmacological and toxicological testing of vaccines [31] | 1997 |
| Gene therapy, DNA vaccines, genetically modified tissue or cell-based products | CPMP/BWP/3088/99: Quality, preclinical, and clinical aspects of gene transfer products [32] | 2001 |
| Insulin analogues | CPMP/SWP/372/01: Nonclinical assessment of carcinogenic potential of insulin analogues [33] | 2001 |
| Insulin analogues | CPMP/SWP/2600/01: Need for assessment of reproductive toxicity of insulin analogues [34] | 2002 |
| Biotechnology-derived proteins | CPMP/3097/02: Comparability of medicinal products containing biotechnology-derived proteins as drug substance—preclinical and clinical issues [35] | 2003 |
| Gene therapy, DNA vaccines, genetically modified tissue or cell-based products. | EMEA/273974/05: Draft Annex on nonclinical testing for inadvertent germ-line transmission of gene transfer vectors [36] | 2005 |
| Vaccine adjuvants | CHMP/VEG/134716/2004: Guideline on adjuvants in vaccines for human use [37] | 2005 |
| Biotechnology-derived proteins | EMEA/CHMP/BMWP/101695/2006: Guideline on comparability of biotechnology-derived medicinal products after a change in the manufacturing process [38] (will replace CPMP/3097/02) | 2007 |

*Note*: The term guideline is used here to refer to testing guidelines, "Notes for guidance," as well as "Points to consider" documents.

to use of the guiding principles laid out in ICH, there are EU-specific preclincial guidelines relating to gene transfer medicinal products, vaccines, vaccine adjuvants, and insulin analogues. There has also been much recent activity with respect to evolution of guidelines for addressing comparability of biotechnology products after production process changes, and the development of biosimilar products, that contain some guidance on the preclinical data requirements. EU-specific guidelines relating to "biosimilar" products are listed in Table 6, and these include some guidance directed toward some specific "generic" forms of biopharmaceutical product that are likely to be coming off-patent in the coming years. Such products include recombinant forms of human insulin, growth hormone, and erythropoeitein.

**TABLE 6 EU preclinical guidelines on biosimilar medicinal products**

| Guideline | Type | Start Date |
|---|---|---|
| CHMP/437/04: Similar biological medicinal products [39] | Regulatory guideline | 2005 |
| EMEA/CHMP/42832/05: Similar biological medicinal products containing biotechnology-derived proteins as active substance—preclinical and clinical issues [40] | Testing guideline | 2006 |
| EMEA/CHMP/94526/05: Annex guideline—preclinical and clinical issues on similar medicinal products containing recombinant erythropoietins [41] | Testing guideline | 2006 |
| EMEA/CHMP/94528/05: Annex guideline—preclinical and clinical issues on similar medicinal products containing somatropin [42] | Testing guideline | 2006 |
| EMEA/CHMP/94529/05: Annex guideline—preclinical and clinical issues on similar medicinal products containing recombinant granulocyte-colony stimulating factor [43] | Testing guideline | 2006 |
| EMEA/CHMP/32775/05: Annex guideline—preclinical and clinical issues on similar medicinal products containing recombinant human insulin [44] | Testing guideline | 2006 |

In addition to the EU guidance documents that relate specifically to biopharmaceuticals, there are also a few other EU general preclinical guidelines that may need to be referred to in the planning of toxicity studies on biopharmaceuticals. For example, there is a guidance relating to toxicokinetic sampling of control animals in toxicity studies, in order to check for cross contamination with test substance [20]. An indication of potential evolution of new or modified guidance on nonclinical issues in the European Union can be gained from reference to the "work plan" for the Safety Working Party (SWP) on the EMEA Web site [21].

## 6 APPLICATION OF THE ICH S6 GUIDELINE IN THE EUROPEAN UNION

In addition to the unique regulatory structure in the European Union and the specific guidelines referred to above, there are a few other comments and issues relating to the interpretation and application of the ICH guidance.

### 6.1 General Comments on Application of the Guidance

All the national regulatory agencies, and the central agency (EMEA) in the European Union recognize and are generally familiar with the principles laid

out in the ICH S6 guideline. In view of the considerable expertise that has been gained in the various agencies through review of preclinical programs on the large number of approved biopharmaceuticals in the EU region, the level of scientific expertise and quality of advice given to applicants can be regarded as equivalent to that in the other major territories covered by ICH (the United States and Japan). In general, there is a good understanding of the need for a case-by-case approach to preclinical testing of biopharmaceutical products, and most EU preclinical assessors are anxious to avoid unnecessary animal studies that will add little to the human risk assessment. The strategies employed for preclinical testing of biopharmaceuticals have been generally similar to the flexible approaches that have been a hallmark of the Center for Biologics Evaluation and Research (CBER) at the US FDA. While some observers have commented that since the recent switch of assessment of many biotechnology-derived drugs from CBER to the drugs division of FDA (CDER—Center for Drugs Evaluation and Research), a more rigid small-molecule approach has been applied by the US agency, not strictly in the spirit of the ICH S6 guideline [22], a similar trend is not evident among the EU agencies. The EU approach is generally harmonized across all types of biopharmaceutical products, and it seems to be firmly anchored around the principles laid out in the ICH S6 guideline, as evidenced by recent publications on safety assessment of these products by regulators in Germany [26], as well as comments in the report from UK industry bodies (Association of British Pharmaceutical Industry, ABPI and BioIndustry Association, BIA) issued following the TGN1412 incident in 2006 [23] (see later). These publications reinforce the concept that preclinical studies on biopharmaceuticals should only be performed in species that show target homology with humans, and where the relevant human pharmacology/pharmacodynamics can be demonstrated. This approach is reinforced by the fact that human toxicity caused by biologicals normally results from "on-target" effects (i.e., exaggerated pharmacology), while the opposite tends to be the case for small-molecule drugs. There does not appear to be any drift toward a more conventional two-species (rodent/nonrodent) approach to toxicity testing for biopharmaceuticals in the European Union, regardless of species relevance, that has become evident recently in the United States, since CDER took over assessment of many of these products from CBER [21].

In Germany, where potentially two different agencies (PEI and BfArM) may review biopharmaceutical products (predominantly PEI, which has responsibility for review of blood/tissue products, vaccines, gene/cellular therapies and antibody products, although some biopharmaceuticals are reviewed by BfArM), the between-agency approach is consistent and ICH S6 compliant.

The major difference between application of the ICH S6 guideline in the European Union, compared to the United States and Japan, is the rather more complex interaction with the regulatory agencies in relation to agreeing on the preclinical testing strategy for biopharmaceuticals. While sponsor compa-

nies in the United States and Japan will work with a single agency (FDA or MOHW) during the development cycle, in the European Union several national agencies, and then the EMEA, may be consulted for scientific advice. Prior to initiation of the first human study on a biopharmaceutical in the European Union, the sponsor company will file a CTA with the relevant national authority for the country where the trial will be conducted, and may seek scientific advice from that authority prior to filing. For novel biopharmaceuticals, some sponsor companies also adopt a strategy of approaching a selection of national authorities in the EU countries that have had the opportunity to amass significant experience in the assessment of biopharmaceutical products (e.g., United Kingdom, Germany, France, the Netherlands, and Sweden) for scientific advice, fairly early in the development cycle, to obtain a general impression of likely EU data requirements. This process may then be further supported by a request for scientific advice from the EMEA, before submission of the MAA. Prior to and following MAA submission, further interaction between the sponsor and the appointed Rapporteur authority may take place to resolve any outstanding issues in the data submission. It may be a point of concern that due to the rather flexible case-by-case approach to preclinical studies advocated by ICH S6, there could be differences of opinion about data requirements between national authorities, but it is the role of the EMEA and the Rapporteur authorities to act as "moderators" in this regard. There is ample opportunity, particularly in the preclinical overview (section 2.4 of the Common Technical Document or CTD), for sponsor companies to defend their preclinical program and to justify any apparent data gaps. Major disputes during late development, or at the MAA, regarding the adequacy of the preclinical programme for a biopharmaceutical in the European Union appear to occur rarely, which seems to indicate that the scientific advice process and sponsor/agency relationships work quite well.

### 6.2 "High-Risk" Compounds and Clinical Trials

The severe, unexpected toxicity, not predicted by a 28-day toxicity study in the monkey, observed in a phase 1 study with a CD28 agonist monoclonal antibody (TGN1412) in healthy volunteers conducted in London in March 2006, has had repercussions regarding the extrapolation from preclinical data to starting doses in human phase 1 studies for some biopharmaceuticals, as well as some so-called high-risk small-molecule drugs. Following the TGN1412 phase 1 study incident, an Expert Scientific Group (ESG) was established in the United Kingdom to investigate why the adverse events seen in the volunteers had not been predicted by the preclinical safety studies in animals and to make recommendations about the future conduct of phase 1 clinical trials in the United Kingdom. The ESG, led by Professor Gordon Duff, published its final report in November 2006 [24]. As a result of the review process following the TGN1412 trial, the ESG made 22 recommendations regarding the preclinical and early clinical development of compounds that might pose a

"high risk" to volunteers or patients in early clinical studies. In the subsequent CHMP guideline that has been issued, relating to mitigation of risk for such studies [25], it is stated that all compounds entering the clinic should be assessed for certain "factors of risk," namely:

- Novelty of the mode of action, including possibility of pleiotropic effects or triggering of biological cascades (e.g., cytokine release).
- Nature of the target in humans, and level of available knowledge of the target.
- Relevance of animal species and models used for preclinical studies.

For drugs that might be identified to pose a "high risk" based on assessment of these criteria, prior to conduct of a phase 1 study (or possibly other types of trial) in the United Kingdom, a data package (basically the CTA package, supplemented by specific responses to some standard questions; see Table 7), is submitted to the UK authority (MHRA) for consideration by an Expert Advisory Group (EAG). Prior to submission of the CTA package, it is possible to request an opinion from MHRA as to whether the trial might be regarded as "high-risk," thereby warranting referral of the CTA package to the EAG

**TABLE 7 Standard questions requiring response prior to conduct of phase 1 trials with "high-risk" pharmaceuticals in the United Kingdom**

| Question | Information/Data Requested |
|---|---|
| 1 | A discussion of the function of the target in human. |
| 2 | A discussion of the ability of the subject to maintain a normal physiological response to challenge in the presence of the investigational product. |
| 3 | A rationale for the transition from preclinical to human testing, particularly with regard to highly species-specific molecules. |
| 4 | A discussion of the potential for on-target and off-target effects and how these will be handled in the clinic. |
| 5 | A discussion of the doses used in the relevant animal species (particularly with regard to the use in the animal model of the starting dose to be used in human). |
| 6 | A rationale for the starting dose in human (e.g., including receptor occupancy). |
| 7 | A rationale for the study population (particularly for the use of healthy volunteers). |
| 8 | A rationale for the administration schedule for the initial and subsequence cohorts. This should include the time interval between dose administered to individual subjects. |
| 9 | A rationale for the dose escalation particularly with regard to potential adverse effects. |
| 10 | The proposed trial site, including the facilities available. |

for review. The scope and remit of assessments performed by the EAG is as follows:

- First Time in Man (FTIM) studies with new compounds acting (directly or indirectly) via the immune system with a novel target or a novel mechanism of action or having a secondary potential effect on the immune system via a mechanism of action which currently is not well characterized.
- FTIM studies with novel compounds acting via a possible or likely species specific mechanism.
- Any FTIM studies which are otherwise seen as requiring expert advice.
- Other clinical trials involving classes of compound where MHRA may wish to seek external expert advice.
- Provide expert advice on whether a product's mechanism of action is novel and comes within the scope of the EAG.
- Provide MHRA with expert advice on pre-meeting scientific advice documentation for within scope compounds.
- Other clinical trials where MHRA may wish to seek advice or where there is a difficult risk/benefit balance.
- Other clinical trials involving products where a new class safety issue has been identified.

EAG review of a CTA package referred to the EAG by MHRA takes 40–45 days, after which time the CTA can be submitted to MHRA provided the EAG gives a positive opinion. This new process for assessment of CTAs for drugs that might pose a "high risk" in early human trials has already been implemented by the UK regulatory authority (MHRA). As mentioned above, following the TGN1412 event, and due to political pressure, the Safety Working Party (SWP) of EMEA has issued an EU guidance document with regard to the identification and mitigation of risk for first-in-man human clinical trials for investigational medicinal products [25].

At a practical level, the TGN1412 incident is likely to have an influence on the preclinical testing of certain biopharmaceutical products (particularly monoclonal antibodies and other immunomodulatory products), in the following respects:

- There will need to be clearer understanding of the relevance of the species used in preclinical studies to humans. This is likely to involve generation of relative receptor expression/density data between animals/humans and relative potency between animals/humans using suitable cell/tissue preparations, to be taken into account in interspecies dose scaling. Although the ICH S6 guideline makes reference to these issues, they will now assume greater importance and a more thorough examination as a result of the TGN1412 incident.

- Additional preclinical data to address potential for specific adverse events, such as cytokine release syndrome, may be required in some cases.
- The approach to setting the starting dose for "high-risk" human clinical trials may no longer be based on extrapolation from the no-observable-adverse-effect-level (NOAEL) in the animal toxicity studies, but rather on the "minimal anticipated biological effect level" (MABEL), derived from in vitro potency data, combined with predictive modeling of the kinetic behavior of the drug in man and/or estimates of potential receptor occupancy at the human starting dose. Examples of various approaches to calculation of MABEL values are given in the Final Report of the ESG on Phase One Clinical Trials [24], as well as in the ABPI/BIA report on early stage clinical trials [23]. An overview of the TGN1412 incident and the implications for conduct of phase 1 trials in Germany has been published by assessors from one of the German authorities, the Paul-Ehrlich Institute (PEI) [26]. Some other EU countries have also modified their assessments and procedures for approval of first-in-man clinical trials following the TGN1412 incident (e.g., the French agency, AFSSAPS, issued guidance on conduct of human phase 1 studies in July 2006, including recommendations for starting dose selection and dose escalation, although this guidance no longer seems to be available in English on their website).

### 6.3 Primate Supply and Use

Due to the high degree of species specificity of action of many biopharmaceuticals, and potential immunogenicity in lower species, primates are quite extensively used in preclinical safety testing of these products. There is a slight misconception in some regions outside the European Union that there are tight controls and some difficulties regarding the conduct of preclinical studies in primates in Europe. The reality of the situation is as follows:

- Breeding facilities for primates for laboratory use are limited within the European Union, with the exception of marmosets.
- There are well-established supply channels for import of purpose-bred monkeys (mainly cynomolgus monkeys, and some rhesus monkeys) into the European Union from breeding colonies in Eastern Asia or Mauritius.
- Monkeys used for drug development are almost exclusively "purpose bred," the use of wild-caught animals actually not being allowed in many member states.
- The use of primates in research has to be fully justified, and is subject to special review in some member states. However, provided that scientific justification can be provided, preclinical studies on primates are permitted, and there are a number of major contract research organizations

(CROs) throughout the EU that perform primate toxicity studies on a routine basis.
- Studies using great apes (e.g., chimpanzees) are generally not permitted in many member states, and there are no readily available colonies of great apes for research use in the European Union.

The protection of experimental animals in the European Union is covered by Directive 86/609/EEC [27]. This Directive is currently undergoing a process of review and update, a public consultation exercise on the proposed revisions having recently been undertaken. There are some proposed changes to the Directive that could impact on primate use, and potentially on biopharmaceutical development in the European Union:

- A proposal to restrict primate use to F2 animals (and subsequent generations) of wild-caught breeding stock. At present, many EU laboratories still use F1 generation animals, and a switch to F2 animals would lead to supply problems in the coming years, while breeders set aside a significant number of the F1 animals normally supplied to the European Union, for breeding purposes.
- An EU-wide ban on the use of great apes, with very limited exceptions. In reality this change may have little impact, since statistics show that virtually no great apes have been used for research purposes in the European Union in recent years (only six animals in 1999, and zero in 2002). However, for monoclonal antibodies that only show cross reactivity in humans and chimpanzees, it is still fairly common for a limited safety study in chimpanzees to be conducted to support the phase 1 human studies. Such studies would have to be conducted outside of the European Union, although it should be pointed out that it is unlikely that an EU national regulatory authority would expect chimpanzee data to be submitted as part of a CTA for a human-specific monoclonal antibody product. Data from literature sources, in vitro data, as well as data from any relevant transgenic mouse or rodent homologue antibody models would probably be used in the risk assessment in these circumstances.

There are efforts to reduce primate use in the European Union, which could impact on the use of monkeys for biopharmaceutical development programs in the future. For example, the National Centre for Replacement, Refinement and Reduction of Animals in Research (NC3Rs, based in the Union Kingdom) has recently hosted a workshop and published proposals in relation to the reduction in primate use in monoclonal antibody development [28]. They are advocating that so-called alternative approaches to safety testing of monoclonal antibodies, using transgenic rodents that express the human receptor for the antibody, or using rodent homologue versions of the human antibody, should be employed more routinely so as to reduce the numbers of primates

used. This initiative, while well-intentioned, possibly fails to understand the technical difficulties and interpretative challenges associated with the alternative approaches, and it remains to be seen whether such approaches would be accepted in place of primate toxicity studies by regulatory agencies in all ICH regions. In addition, while such approaches have been used in a few instances where human antibodies show no cross reactivity with conventional laboratory animal species (e.g., the anti-CD11a antibody, Efalizumab [29]), there has been no validation of the alternative approaches for an antibody that does cross react with monkeys, to assess whether the transgenic or rodent homologue models provide equivalent, relevant safety data compared to those obtained from a standard primate toxicity study. This debate will probably continue for some time, but for the present, sponsors based outside the European Union probably just need to be aware that there are groups working to reduce primate use in Europe.

### 6.4 Duration of Chronic Toxicity Studies

There has been some confusion regarding the required duration of chronic toxicity studies on biopharmaceutical products. The ICH S6 guideline states that studies of six months duration are usually adequate for biotechnology-derived drugs. In contrast, some regulators (including some assessors at CDER in the US agency) have referred to the ICH M3 guidance [30] on this topic, and suggested that 9 or 12 month toxicity studies may be necessary for chronic use biopharmaceuticals, using a small-molecule drug approach to the preclinical studies. In the European Union such confusion does not seem to exist, such that all agencies accept the ICH S6 guidance that a six-month animal study is generally adequate to address chronic toxicity of biopharmaceutical products.

### 6.5 Comparability Testing of Biopharmaceutical Products

It is reasonable to comment that the state of the art of comparability testing, as well as approaches used for assessment of biosimilar or follow-on biologics may actually be slightly more advanced in the European Union than in some other regions, due the effort that has been expended in the development of a number of guidelines in this area. These guidelines include specific documents on particular product types (e.g., insulins, growth hormone) that are likely to be developed as biosimilars in the coming years.

## 7 CONCLUDING REMARKS

ICH S6 has been uniformly implemented across the EU member states, and considerable experience of the application of the guiding principles in this document has been gained both locally (in many of the national authorities)

as well as centrally (within the EMEA). There is no evidence that there are major differences in the interpretation of the guideline within the European Union, compared to the other major ICH territories (the United States and Japan). The major difference in application of the guidance stems from the rather more complex regulatory structure and process for obtaining scientific advice, compared to the single-agency model that applies in the United States and Japan. Therefore much attention has been paid in this chapter to an explanation of the unique regulatory structure and the options for obtaining drug approvals within the European Union.

Safety evaluation of novel biopharmaceuticals has come into the spotlight in the European Union over the last year, as a result of the serious adverse events observed in the TGN1412 phase 1 clinical trial. This has already led to new initiatives with regard to extrapolation from preclinical animal data to early human trials, which will place more emphasis on a full understanding of species differences in pharmacology of biotechnology-derived drugs, as well as novel approaches to deriving starting doses for first-in-human studies for some of these molecules. In the short term there is likely to be a more cautious approach to setting of starting dosages for phase 1 studies on immunomodulatory biopharmaceuticals in the European Union, although there is no evidence that the TGN1412 incident has led to increased requirements for preclinical animal studies. In contrast, there has probably been an increased interest in additional in vitro studies to assist in pharmacokinetic/pharmacodynamic correlations between animals and humans so as to better understand the relevance of the animal safety data to humans. These activities are basically an extension of some of the principles originally laid out in the ICH S6 guideline, and should serve to further enhance the relevance and value of preclinical studies on biopharmaceuticals.

## REFERENCES

1. European Medicines Agency Press Release: European Medicines Agency adopts first positive opinion for a similar biological medicinal product. 27 January 2006. http://www.emea.europa.eu/pdfs/human/press/pr/3179706en.pdf
2. Commission Directive 2003/63/EC. Official Journal of the European Union. 27 June 2003. http://ec.europa.eu/enterprise/pharmaceuticals/eudralex/vol-1/dir_2003_63/dir_2003_63_en.pdf
3. Harman RJ. The EMEA–Drug regulation at a supranational level. *Pharm J* 2002;269:752–6.
4. European Experts. European Medicines Agency Web site. http://www.emea.europa.eu/htms/aboutus/experts.htm
5. Pignatti F, Boone H, Moulon I. Overview of the European regulatory system. *J Ambul Care Manage* 2004;27:89–97.
6. Bertele' V, Li Bassi L. A critique to the European regulatory system. *J Ambul Care Manage* 2004;27:98–104.

7. Ainsworth MA. New drugs and European procedures of approval. The European Agency for Evaluation of Medicinal Products' role. *Ugeskr Laeger* 2003;165: 1648–9.
8. EMEA pre-submission flowchart. European Medicines Agency Web site. http://www.emea.europa.eu/htms/human/presub/q31.htm
9. Directive 2001/20/EC of the European Parliament and of the Council of 4 April 2001 on the approximation of the laws, regulations and administrative provisions of the member states relating to the implementation of good clinical practice in the conduct of clinical trials on medicinal products for human use. http://europa.eu.int/eur-lex/lex/LexUriServ/LexUriServ.do?uri=CELEX:32001L0020:EN:HTML
10. European Clinical Trials Database. EudraCT supporting documentation. http://eudract.emea.eu.int/document.html
11. Detailed guidance for the request of authorization of a clinical trial on a medicinal product for human use to the competent authorities, notification of substantial amendments and declaration of the end of the trial. October 2005. http://ec.europa.eu/enterprise/pharmaceuticals/eudralex/vol-10/11_ca_14-2005.pdf
12. Kock M, Thomsen MK. Contact and dialogue between drug companies and the global regulatory authorities throughout research and development phase. *Ugeskr Laeger* 2003;165:1649–52.
13. Dejas-Eckertz P, Shaffner G. Scientific advice by the nationally competent authority and the EMEA on the conduct of clinical trials. *Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz* 2005;48:423–8.
14. MHRA Web site. Scientific advice for license applicants. http://www.mhra.gov.uk/home/idcplg?IdcService=SS_GET_PAGE&nodeId=121
15. European Medicines Agency. New framework for scientific advice and protocol assistance. EMEA/267/187/2005 Rev. 1, 26 April 2006. http://www.emea.europa.eu/pdfs/human/sciadvice/26718705en.pdf
16. ICH Harmonized Tripartite Guideline S6: Preclinical Safety Evaluation of Biotechnology-Derived Pharmaceuticals. July 1997. http://www.ich.org/LOB/media/MEDIA503.pdf
17. Pre-clinical biological safety testing on medicinal products derived from biotechnology (3BS13a). European Commission. http://pharmacos.eudra.org/F2/eudralex/vol-3/pdfs-en/3bs13aen.pdf
18. Guideline on Production and Quality Control of Monoclonal Antibodies. Document III/5271/94 in Volume 3AB4a of the Rules Governing Medicinal products in the European Union. http://www.tga.gov.au/docs/pdf/euguide/vol3a/3ab4aen.pdf
19. Concept paper on the need to revise the guideline on production and quality control of monoclonal antibodies. CHMP/BWP/64/04, 21 October 2004. http://www.emea.eu.int/pdfs/human/bwp/006404en.pdf
20. European Medicines Agency—Committee for Proprietary Medicinal Products. Guideline on the Evaluation of Control Samples In Nonclincial Safety Studies: Checking for Contamination with the Test Substance. CPMP/SWP/1094/04. 17 March 2005. http://www.emea.europa.eu/pdfs/human/swp/109404en.pdf
21. European Medicines Agency. Work plan for the Safety Working Party. EMEA/154420/2006 corr. January 2007. http://www.emea.europa.eu/pdfs/human/swp/15242006en.pdf

22. Schwieterman WD. Regulating biopharmaceuticals under CDER versus CBER: an insider's perspective. *Drug Disc Today* 2006;11:945–51.
23. Joint ABPI/BIA Report: Early stage clinical trial taskforce. 24 July 2006. http://www.abpi.org.uk/information/pdfs/BIAABPI_taskforce2.pdf
24. Expert Scientific Group on Phase One Clinical Trials. Final report. 30 November 2006. http://www.dh.gov.uk/en/Publicationsandstatistics/PublicationsPublicationsPolicyAndGuidance/DH_063117
25 European Medicines Agency—Committee for Medicinal Products for Human Use (CHMP). Guideline on Strategies to Identify and Mitigate Risks for First-in-Man Human Clinical Trials with Investigational Medicinal Products. EMEA/CHMP/SWP/2836707/2007. 19 July 2007. http://www.emea.europa.eu/pdfs/human/swp/2836707enfin.pdf
26. Liedert B, Bassus S, Schneider C, Kalinke U, Löwer J. Safety of phase 1 clinical trials with monoclonal antibodies in Germany—The regulatory requirements viewed in the aftermath of the TGN1412 disaster. *Int J Clin Pharmacol Ther* 2007;45:1–9.
27. Council Directive of 24 November 1986 on the approximation of laws, regulations and administrative provisions of the member states regarding the protection of animals used for experimental and other scientific procedures (86/609/EEC). http://ec.europa.eu/food/fs/aw/aw_legislation/scientific/86-609-eec_en.pdf
28. Chapman K, Pullen N, Graham M, Ragan I. Preclinical safety testing of monoclonal antibodies: the significance of species relevance. *Nature Drug Disc* 2004;6:120–6.
29. Clarke J, Leach W, Pippig S, et al. Evaluation of a surrogate antibody for preclinical safety testing of an anti-CD11a monoclonal antibody. *Reg Toxicol Pharmacol* 2004;40:219–26.
30. ICH Harmonized Tripartite Guideline M3: Nonclinical Safety Studies for the Conduct of Human Clinical Trials for Pharmaceuticals. November 2000. http://www.ich.org/LOB/media/MEDIA506.pdf
31. European Medicines Agency—Committee for Proprietary Medicinal Products. Preclinical pharmacological and toxicological testing of vaccines. CPMP/SWP/465/95. London, 17 December 1997. http://www.emea.eu.int/pdfs/human/swp/046595en.pdf
32. European Medicines Agency—Committee for Proprietary Medicinal Products. Note for Guidance on the Quality, Preclinical and Clinical Aspects of Gene Transfer Medicinal Products. CPMP/BWP/3088/99. London, 24 April 2001. http://www.emea.eu.int/pdfs/human/bwp/308899en.pdf
33. European Medicines Agency—Committee for Proprietary Medicinal Products. Points to Consider on the Non-clinical Assessment of the Carcinogenic Potential of Insulin Analogues. CPMP/SWP/372/01. London, 15 November 2001. http://www.emea.eu.int/pdfs/human/swp/037201en.pdf
34. European Medicines Agency—Committee for Proprietary Medicinal Products. Points to Consider on the Need for Assessment of Reproductive Toxicity of Human Insulin Analogues. CPMP/SWP/2600/01 final. London, 1 March 2002. http://www.emea.eu.int/pdfs/human/swp/260001en.pdf
35. European Medicines Agency—Committee for Proprietary Medicinal Products. Guideline on the Comparability of Medicinal Products Containing Biotechnology-

Derived Proteins as Active Substance. Nonclinical and Clinical Issues. 17 December 2003. http://www.emea.europa.eu/pdfs/human/ewp/309702en.pdf

36. European Medicines Agency—Committee for Proprietary Medicinal Products. Note for Guidance on the Quality, Preclinical and Clinical Aspects of Gene Tansfer Medicinal Products. Draft annex on the non-clinical testing for inadvertent germline transmission of gene transfer vectors. 17 November 2005. http://www.emea.europa.eu/pdfs/human/genetherapy/27397405en.pdf
37. European Medicines Agency—Committee for Proprietary MedicInal Products. GuidelIne on Adjuvants in Vaccines for Human Use. EMEA/CHMP/VEG/134716/2004. London, 20 January 2005. http://www.emea.eu.int/pdfs/human/vwp/13471604en.pdf
38. European Medicines Agency—Committee for Proprietary Medicinal Products. Guideline on Comparability of Biotechnology-Derived Medicinal Products after a Change in the Manufacturing Process. EMEA/CHMP/BMWP/101695/2006. 24 January 2007. http://www.emea.europa.eu/pdfs/human/biosimilar/10169506en.pdf
39. European Medicines Agency—Committee for Medicinal Products for human use. Guideline on Similar Biological Medicinal Products. CHMP/437/04, 30 October 2005. http://www.emea.europa.eu/pdfs/human/biosimilar/043704en.pdf
40. European Medicines Agency—Committee for Medicinal Products for Human Use. Similar Biological Medicinal Products Containing Biotechnology-Derived Proteins as Active Substance: Nonclinical and Clinical Issues. EMEA/CHMP/42832/05. 22 February 2006. http://www.emea.europa.eu/pdfs/human/biosimilar/4283205en.pdf
41. European Medicines Agency—Committee for Medicinal Products for Human Use. Annex Guideline: Nonclinical and Clinical Issues on Similar Medicinal Products Containing Recombinant Erythropoietins. EMEA/CHMP/94526/05. 22 March 2006. http://www.emea.europa.eu/pdfs/human/biosimilar/9452605en.pdf
42. European Medicines Agency—Committee for Medicinal Products for Human Use. Annex Guideline: Nonclinical and Clinical Issues on Similar Medicinal Products Containing Somatropin. EMEA/CHMP/94528/05. 22 February 2006. http://www.emea.europa.eu/pdfs/human/biosimilar/9452805en.pdf
43. European Medicines Agency—Committee for Medicinal Products for Human Use. Annex Guideline: Nonclinical and Clinical Issues on Similar Medicinal Products Containing Recombinant Granulocyte-Colony Stimulating Factor. EMEA/CHMP/94529/05. 22 February 2006. http://www.emea.europa.eu/pdfs/human/biosimilar/3132905en.pdf
44. European Medicines Agency—Committee for Medicinal Products for Human Use. Annex Guideline: Preclinical and Clinical Issues on Similar Medicinal Products Containing Recombinant Human Insulin. EMEA/CHMP/32775/05. 22 February 2006. http://www.emea.europa.eu/pdfs/human/biosimilar/3277505en.pdf

# Implementation of ICH S6: Japanese Perspective

TAKAHIRO NAKAZAWA, PhD

**Contents**

1 Introduction
2 Analysis of Data from the JPMA Questionnaire Survey Regarding ICH S6 Implementation in Japanese Pharmaceutical Companies
  2.1 Understanding the General Principles
  2.2 Japanese Practices in Preclinical Studies
3 Points-to-Consider Document in Japan
  3.1 Classification of Biopharmaceuticals
  3.2 Individual Considerations
  3.3 Items for Future Update of Japanese Points-to-Consider Document
4 Conclusion
References

## 1 INTRODUCTION

Biotechnology-derived pharmaceuticals (biopharmaceuticals) appeared for the first time in the 1980s for medical treatment of diseases, such as diabetes mellitus and hypophysical dwarfism. Since then the number and types of biopharmaceuticals have climbed and continue to dramatically increase. One reason for the increase is the evolution of recombinant manufacturing of biopharmaceuticals. This has provided sufficient amounts of proteins for development and clinical use, whereas, for example, the amount of insulin or growth hormones extracted from animal or human tissues had been limited. Another

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

benefit of recombinant technology is the production of proteins with primary amino acid sequences that are identical to those of endogenous human proteins. This reduces immunogenicity, which was a problem with the use of animal proteins in humans. Long-term use of bovine or porcine insulin induces the production of an antibovine insulin or antiporcine insulin antibody that can sometimes decrease its efficacy, change the pharmacokinetics, and cause immunological adverse effects. Human proteins are expected to have sufficiently low or no immunogenicity to be used for the long term in humans. Human insulin produced by recombinant DNA technology was shown to induce the production of the antibody at much lower level in humans than bovine or pork insulin. Although immunogenicity is reduced, long-term administration of human insulin also induces the production of antibody in humans in some patients. Similarly to human insulin, the immunogenicity of biopharmaceuticals in humans remains an issue.

From a preclinical point of view, human protein biopharmaceuticals may be immunogenic in animals. In the case where a human protein is highly immunogenic in animals, the safety evaluation may have technical limitations. Over a decade ago the following concerns/questions were raised about the scientific justifications on the safety assessment of biopharmaceuticals in preclinical studies: If low or no toxicity is observed at high doses of a biopharmaceutical in an animal that does not respond to the biological activity of the biopharmaceutical, can it simply be concluded that the failure of demonstrating toxicity in animals would mean low or absence of toxicity in humans? How is the potential toxicity of a biopharmaceutical interpreted if toxicity decreases after repeated administration due to the production of neutralizing immunogenicity? Should developers conduct a battery of genotoxicity studies for a biopharmaceutical? A human protein biopharmaceutical is positive in an antigenicity study [1], but is this meaningful? Should ADME studies be conducted with radiolabeled biopharmaceuticals?

To answer these questions, guidelines and/or Points-to-Consider documents were issued in the European Union, the United States, and Japan. However, differences in approaches for safety evaluation of biopharmaceuticals among those regions were identified, indicating the need for harmonizing these approaches among the three regions. In 1997 the three regions reached an agreement concerning preclinical safety evaluation of biotechnology-derived pharmaceuticals (ICH S6). On the basis of this ICH S6 guideline, Notification 326 was issued as ICH Step 5 by the Ministry of Health and Welfare (MHW) in 2000 [2]. Subsequently scientists from the National Institute of Health Sciences (NIHS), the Pharmaceuticals and Medical Devices Evaluation Center (currently, Pharmaceuticals and Medical Devices Agency, called PMDA), and Japan Pharmaceutical Manufacturers Association (JPMA) collaborated to publish a Japanese Points-to-Consider document regarding the safety assessment of biopharmaceuticals in preclinical studies in 2002 [3]. An English translation was published in 2004 [4].

## 2 ANALYSIS OF DATA FROM THE JPMA QUESTIONNAIRE SURVEY REGARDING ICH S6 IMPLEMENTATION IN JAPANESE PHARMACEUTICAL COMPANIES

A JPMA working group conducted a survey in 2001 to examine how the ICH S6 guideline was being implemented by the 83 pharmaceutical companies that belonged to JPMA at that time [5]. Responses to general principles were obtained from 54 out of 83 pharmaceutical companies. Moreover 34 biopharmaceuticals from 25 pharmaceutical companies were examined with respect to each company's specific points of considerations in their preclinical studies. In this section the key survey results are summarized along with the interpretation of the JPMA working group.

### 2.1 Understanding the General Principles

***Animal Species Selection*** Almost all pharmaceutical companies select animal species for preclinical studies on the basis of the results of the responsiveness of a test animal to the biological activity of a biopharmaceutical and its production of neutralizing antibodies. Such considerations are specifically applicable to biopharmaceuticals but not to new chemical entities (NCEs). This survey finding suggests that most pharmaceutical companies in Japan understand and implement the animal species selection in good accordance with ICH S6.

ICH S6 allows the use of only one species for subsequent long-term studies when the toxicity profiles of two species are comparable in a short-term study. JPMA's questionnaire included a question on whether the toxicity profiles are concluded to be similar when no toxicity is observed in the two species that are responsive to the biological activity of a biopharmaceutical. The most common response was that this decision is made on a case-by-case basis considering pharmacological and pharmacokinetic data. The second most common was that the toxicity profiles are considered similar when the highest dose for testing is scientifically justified (as discussed in the next section), closely followed by the third most common answer that profiles are not concluded to be similar since the comparison of toxicological changes cannot be done. There may not be a correct single answer. Because in many cases only the biological changes due to exaggerated pharmacological effects are observed, the JPMA working team concluded that it may be practical to consider these toxicity profiles as similar and to use only one species in the subsequent long-term studies as long as the pharmacologically responsive animal and the dose selection are appropriate.

***Highest Dose Selection*** It is sometimes difficult to set the highest dose for safety programs of a biopharmaceutical when low or no toxicity is observed in animals. The JPMA questionnaire survey revealed that most pharmaceutical

companies set the highest dose of a biopharmaceutical using the maximum feasible dose, which is the highest concentration that can be prepared based on the solubility of the biopharmaceutical, and the largest administration volume that can be administered to animals. The second most common response was a dose that is 10 to 100 times higher than the expected clinical dose, and the third was to use a dose several to 10 times higher than the clinical dose. However, the JPMA working team did not agree that the maximum feasible dose or a large multiple of the clinical dose is scientifically appropriate. More important is to determine the adverse biological changes induced by any exaggerated pharmacological effects at doses relevant to the expected clinical dose rather than toxicological changes induced by unrealistic doses. Moreover animal data gathered at maximum feasible doses or at doses that are 100 times higher than the clinical dose cannot ensure the safety of biopharmaceuticals in humans without the consideration of the difference in responsiveness to the biological activity between the test animals and humans and the intended clinical indication. Therefore the JPMA working team concluded that the test doses should be determined on the basis of the responsiveness of animal species to the biological activity of a biopharmaceutical and the clinical usage conditions.

## 2.2 Japanese Practices in Preclinical Studies

***Types of Biopharmaceuticals*** Survey data on preclinical safety assessment programs were analyzed for 34 biopharmaceuticals cases. The numbers of antibodies, human proteins, and human protein analogues either in the development or marketed as of 2001 in Japan were 13, 12, and 6, respectively. The remainder were bioconjugates, DNA-derived vaccines, and human T cell epitopes. Thus antibodies and human proteins are the two major biopharmaceuticals.

***In vivo Studies*** The in vivo preclinical studies conducted for 34 biopharmaceuticals are shown in Table 1. Single-dose and repeated-dose toxicity studies up to three months were conducted in most cases using rodents and nonrodents. Repeated-dose toxicity studies longer than three months were conducted for a third of the 34 cases. Nonrodents were used more often than rodents in considering the biological responsiveness of test animals to the biopharmaceuticals. Levels of antibodies were measured in most repeated-dose toxicity studies (31 out of 34 cases). The alteration of pharmacokinetics or pharmacodynamics by antibodies was observed in 13 out of 31 cases, which did not result in the discontinuation of the studies nor required significant changes in the study design. The JPMA working group agreed with the continuation of such studies, since neutralizing antibodies do not always interfere with the outcome of a toxicological response. However, the production of neutralizing antibodies precludes long-term studies, such as carcinogenicity studies. Reproductive toxicity studies were conducted for 11 out of 34 cases.

**TABLE 1 In vivo preclinical studies conducted for 34 biopharmaceuticals**

| In vivo Preclinical Studies | Cases |
|---|---|
| Rodent single-dose toxicity study | 24 |
| Nonrodent single-dose toxicity study | 21 |
| Rodent repeat-dose toxicity study (not longer than 3 months) | 24 |
| Nonrodent repeat-dose toxicity study (not longer than 3 months) | 29 |
| Rodent repeat-dose toxicity study (longer than 3 months) | 8 |
| Nonrodent repeated dose toxicity study (longer than 3 months) | 13 |
| Antigenicity studies | 12 |
| Immunogenicity studies (excepting antigenicity studies) | 11 |
| Reproductive toxicity study (segment 1) | 13 |
| Rodent reproductive/developmental toxicity study (segment 2) | 14 |
| Nonrodent reproductive/developmental toxicity study (segment 2) | 17 |
| Reproductive/developmental toxicity study (segment 3) | 11 |
| Genotoxicity studies | 16 |
| Carcinogenicity study (long-term) | 2 |
| Carcinogenicity study (short- to mid-term) | 1 |
| Local irritation studies | 24 |
| General pharmacology studies (category A) | 9 |
| General pharmacology studies (category B) | 3 |
| Safety pharmacology studies (core battery) | 9 |
| Safety pharmacology studies (follow-up or additional studies) | 5 |
| Other | 6 |

Antigenicity studies (active sensitization test in guinea pigs, PCA in rabbits or guinea pigs, passive hemagglutination reaction using sensitized rabbit serum, investigation of IgE-type antibody production capacity, degree of covalent bonding with proteins, polymerization of the drug or cross-antigenicity) defined by a book entitled *Drug Approval and Licensing Procedures in Japan* [1] were conducted for 11 out of 34 cases. The antigenicity studies were conducted in most cases before the ICH S6 notification. It was concluded by the ICH S6 expert working group that antigenicity studies conducted with biopharmaceuticals are not scientifically justified. A description of antigenicity studies was taken out of the 2006 edition of the book. The in vivo genotoxicity studies were done for 16 biopharmaceuticals; almost all the cases were prior to the ICH S6 notification. A few carcinogenicity studies were conducted for 34 biopharmaceuticals examined (e.g., recombinant hormones).

On the questionnaire was also a question asking what in vivo preclinical studies commonly conducted for NCEs were intentionally omitted for the 34 biopharmaceuticals (Table 2). Rodent repeated-dose toxicity, antigenicity, reproductive toxicity, in vivo genotoxicity, and carcinogenicity studies were in this category for a third to half of the 34 cases. The main reason for not conducting rodent repeated-dose toxicity studies was the lack of rodent responsiveness to the biological activities of the biopharmaceuticals. The reasons for not conducting reproductive toxicity studies were related to the

**TABLE 2 In vivo preclinical studies intentionally not conducted for 34 biopharmaceuticals**

| In vivo Preclinical Studies | Cases |
|---|---|
| Rodent single-dose toxicity study | 9 |
| Nonrodent single-dose toxicity study | 8 |
| Rodent repeat-dose toxicity study (not longer than 3 months) | 11 |
| Nonrodent repeat-dose toxicity study (not longer than 3 months) | 5 |
| Rodent repeat-dose toxicity study (longer than 3 months) | 10 |
| Nonrodent repeat-dose toxicity study (longer than 3 months) | 6 |
| Antigenicity studies | 15 |
| Immunogenicity studies (excepting antigenicity studies) | 4 |
| Reproductive/developmental toxicity study (segment 1) | 11 |
| Rodent reproductivedevelopmental toxicity study (segment 2) | 11 |
| Nonrodent reproductive/developmental toxicity study (segment 2) | 8 |
| Reproductive/developmental toxicity study (segment 3) | 11 |
| In vivo genotoxicity studies | 11 |
| Carcinogenicity study (long-term) | 15 |
| Carcinogenicity study (short- or mid-term) | 6 |
| Local irritation studies | 3 |
| General pharmacology studies (category A) | 5 |
| General pharmacology studies (category B) | 4 |
| Safety pharmacology studies (core battery) | 6 |
| Safety pharmacology studies (follow-up or additional studies) | 5 |
| Other | 0 |

administration period in a clinical setting, the patient population, nonresponsiveness of animal species to biological activity, and the existence of extensive safety data for humans on natural proteins. The main reason for not conducting antigenicity and in vivo genotoxicity studies was that those studies are not required by ICH S6. The main reasons for carcinogenicity studies were the lack of biological responsiveness and the production of neutralizing antibodies.

***In vitro Studies*** Table 3 shows the in vitro preclinical studies conducted for the 34 biopharmaceuticals. Most were in vitro genotoxicity studies that are not required by the ICH S6 guideline. This is not the case of poor understanding of the ICH S6. In vitro genotoxicity studies are usually conducted at the early stage of development. Almost all the in vitro genotoxicity studies examined by the JPMA survey in 2001 were conducted before the ICH S6 notification. The cross-reactivity studies were used to understand interspecies reactivity to a biopharmaceutical, especially in case of antibodies.

***ADME Studies*** Absorption, distribution, metabolism, and excretion studies were conducted for 24, 21, 17, and 19 out of the 34 biopharmaceuticals, respectively. No radiolabeled proteins were used for 20, 2, 1, and 13 out of the 34

**TABLE 3 In vitro preclinical studies conducted for 34 human proteins**

| In vitro Preclinical Studies | Cases |
|---|---|
| Ames tests | 14 |
| Chromosomal aberration test | 10 |
| Mouse lymphoma TK test | 3 |
| Cross reactivity study | 9 |
| In vitro general pharmacology studies (category A) | 4 |
| In vitro general pharmacology studies (category B) | 3 |
| In vitro safety pharmacology studies | 3 |
| Other | 3 |

biopharmaceuticals, respectively. The types of radiolabeled proteins in the remaining studies included conjugates of $^{125}$I-, $^{14}$C-, and $^{3}$H-. When using radiolabeled proteins, it is important to demonstrate that a radiolabeled test material maintains its activity and biological properties equivalent to those of the unlabeled material. In most cases this was done before using the radiolabeled protein. Other concerns are stoichiometric radiolabeling, the loss of radiolabel, recycling of a radiolabeled amino acid into non–drug-related protein, and disruption of stability. It is well known that there are technical limitations to the use of radiolabeled proteins for ADME studies. For example, when using $^{125}$I-radiolabeled proteins in distribution studies, the formation of inorganic iodine by deiodinization in vivo should be considered. Free $^{125}$I thus produced is accumulated in the thyroid gland: therefore a $^{125}$I-radiolabeled biopharmaceutical seems to distribute mainly in the thyroid gland. However, the JPMA working group agreed that with a good understanding of the technical limitations, ADME studies using radiolabeled proteins may provide some useful information for the planning of human study.

## 3 POINTS-TO-CONSIDER DOCUMENT IN JAPAN

Scientists from NIHS, PMDA, and JPMA collaborated to publish a Points-to-Consider document regarding the safety assessment of biopharmaceuticals in preclinical studies in 2002 [3]. The collaboration team intended to clarify their interpretation of the ICH S6 guideline and share recent Japanese practices on this matter. However, it was written in Japanese. Thus the collaboration team made an English translation of the document and also collected comments on the contents from experts in the United States and the European. The experts agreed to most ideas presented in the Japanese Points-to-Consider document. They also suggested more clarification of some other ideas. In light of these comments, the collaboration team revised and published the English translation of the document, such that the nonnative Japanese could correctly understand the contents [4]. In this section, I summarize the key points of the document, as they may be of some help to scientists in the pharmaceutical

industries and regulatory reviewers in countries other than Japan. However, it is important to note that the Points-to-Consider document is not a regulatory requirement in Japan. There may even be a gap between the contents of the document and some individual cases. Therefore the application of the contents to individual cases should be made on a scientifically justified case-by-case manner.

### 3.1 Classification of Biopharmaceuticals

A significant contribution of the Japanese Points-to-Consider document is the classification of biopharmaceuticals. When ICH S6 was written, the types of biopharmaceuticals whose information was available for discussion were mainly human proteins/peptides and diagnostic antibodies. Several new types of biopharmaceutical have been developed since the notification of the ICH S6. However, the knowledge on and experiences in the development of the new types of biopharmaceutical have not been shared. Thus the classification and considerations for many of the newer types is proposed in the Japanese Points-to-Consider document. Discussed next are the key points of the document [4].

Biopharmaceuticals covered by the ICH S6 guideline include protein and peptide products consisting of natural amino acids. The top portion of Table 4 shows the subcategories of biopharmaceuticals. Antibodies were initially considered to be included under the protein subcategory classification, as they consist of amino acids, but ended up being a separate category because the biological activities of antibodies differs substantially from those of other protein products. In recent years the development of human protein analogues has sporadically been observed with the intent of improving efficacy. Therefore approaches to the development of these analogues are also described. The ICH S6 covers the safety evaluation of biopharmaceuticals by taking into account the type and clinical applications. The considerations for each type of biopharmaceutical are described below. The safety of impurities and degradation products in biopharmaceuticals needs to be comprehensively assessed with respect to their quality and bioactivity.

***Proteins*** When a human protein is used at a concentration in blood exceeding the physiological level, studies for safety evaluation should be designed with reference to many of the considerations mentioned in the ICH S6 guideline. Moreover an entirely different physiological secretion pattern in humans should be considered. Changes in concentration in blood are considered more significant than the concentration itself for some classes of proteins. Biopharmaceuticals intended for use in sustained-release formulation show changes in blood concentration that diverge much from the physiological secretion pattern. Therefore, when the changes in the blood concentration of an exogenous human protein in blood differ from the physiological secretion patterns of an endogenous protein, attention should be paid to the potential changes in the physiological action.

**TABLE 4 Type of biopharmaceuticals/related biological medicines and scope of the ICH S6 guideline**

| |
|---|
| PROTEINS *(covered by ICH S6)* |
| Human-type protein |
| Nonhuman protein |
| Human-type protein analog consisting of natural amino acid |
| Human-type protein analog containing nonnatural amino acid |
| Bioconjugate of human-type protein and other protein |
| Bioconjugate of human-type protein and organic linker |
| PEPTIDES *(covered by ICH S6)* |
| Human-type peptide |
| Nonhuman-type peptide |
| Human-type peptide analogue consisting of natural amino acid |
| Human-type peptide analogue containing nonnatural amino acid |
| ANTIBODIES *(covered by ICH S6)* |
| Monoclonal antibodies/chimera antibodies |
| Immunoconjugates |
| PEPTIDE MIMICS *(not covered by ICH S6 but its basic principles can be used as reference)* |
| NCEs having a selective affinity to human peptide receptors |
| OLIGONUCLEOTIDE MEDICINES *(not covered by ICH S6 but its basic principles can be used as reference)* |
| Anti-sense compounds |
| RNAi |
| Aptamer |
| OTHER BIOLOGICAL MEDICINES *(not covered by ICH S6 and safety evaluation conforms to other standards)* |
| Antibiotics |
| Allergen extracts |
| Vitamins |
| Viral vaccines, etc. |

For animal proteins or human-type protein analogues consisting of natural amino acids (i.e., human-type protein analogues with the original human protein amino acids substituted with other natural amino acids, added natural amino acids, or deleted amino acids) potential differences in the potency and quality of biological activity between these pharmaceuticals and the original human proteins should be considered. For example, in the case of a human-type protein analogue in which the substituted site is in receptor recognition sites, its biological activity may be enhanced or diminished, and even a new biological activity may occur. Moreover, depending on the type and site of the

amino acid replaced, a new antigen determinant (epitope) may be expressed that results in changes of immunogenicity.

For human-type protein analogues with nonnatural amino acids, in addition to the considerations above, attention should be paid to the potential biological activity and pharmacokinetic behavior in the fragment containing the site in which this protein has been metabolized. For example, no genotoxicity studies are required for proteins that cannot pass through cell membranes, whereas the applicability of these studies should be discussed on a case-by-case basis for fragments containing nonnatural amino acids. No metabolism studies are required for proteins that are degraded into only amino acids, whereas metabolism studies may provide useful information on proteins containing nonnatural amino acids.

Two types of bioconjugates may exist. A bioconjugate of a human-type protein and another protein may have the combined biological activity of both proteins, and their effects on the body may be altered due to their interaction. Therefore conducting safety evaluation in pharmacological studies should be considered. On the other hand, a bioconjugate of a human-type protein and an organic linker can be studied similarly to human-type protein analogues containing nonnatural amino acids.

***Peptides*** Peptides, similar to proteins, consist of amino acids, although their molecular weights are lower than those of proteins. Therefore the considerations above for proteins are also applicable to peptides. Antibody formation, which is a key issue in animal experiments on human-type proteins, generally depends on molecular weight (i.e., the probability of antibody formation is low if the molecular weight is low). The guideline covers not only biotechnologically produced peptides but also chemically synthesized peptides.

***Antibodies*** Antibodies are usually targeted to specific receptors, particularly monoclonal antibodies. Many of these antibodies are inherently species-specific. It is important for the developer to verify species specificity in order to justify the use (or non-use) of a particular animal species in safety studies. In the cases where an appropriate animal model is not available, the use of homologous antibodies for animals or the use of relevant transgenic animals expressing human antigens should be considered. In addition, when an IgG antibody is used in possibly pregnant or lactating women, and on the basis of the intended indication, reproductive toxicity should be investigated because of the potential of the antibody to be transferred to the placenta or milk. Immunoconjugates of antibodies, conjugated with either other proteins or organic linkers, should be handled the same way as for the bioconjugates described above.

***Peptide Mimics and Oligonucleotide Medicines*** The middle section of Table 4 shows new types of pharmaceuticals not classified as biopharmaceuticals, although they may have a selective pharmacological action similar to that

of biopharmaceuticals. Their safety evaluation in animals is sometimes difficult to conduct. The ICH S6 guideline does not cover these pharmaceuticals, but its basic principles can be used as reference. Peptide mimics, antisense compounds, RNAi, and aptamer fall in this category. (Refer to the Points-to-Consider paper [4] for references.)

***Other Biological Medicines*** The conventional biologics shown in the bottom section of Table 4 are not covered by the ICH S6 guideline. Preclinical safety evaluation based on relevant standards is necessary for this group of pharmaceuticals. (Refer to the Points-to-Consider paper [4] for references.)

### 3.2 Individual Considerations

Additional useful ideas included in the Japanese Points-to-Consider document are summarized that require further clarification.

***Highest Dose Selection*** The highest dose for safety assessment programs in the preclinical studies should be selected by considering the intended maximum clinical exposure of a biopharmaceutical on the basis of its AUC. There is generally no need to investigate biopharmaceuticals at exposures much higher than the intended clinical exposure, unlike the case of NCEs. Another consideration is that the highest dose is as the dose at which the pharmacodynamic response has reached the plateau (pharmacodynamic maximum dose).

***Toxicological Effects and Pharmacological Action*** In some cases only exaggerated pharmacological effects may be observed in the toxicological studies of the biopharmaceuticals. Sometimes these effects are difficult to distinguish from a compound-related toxicity. However, if the effect is related to the mechanism (predictable) and is reversible, it should not be considered as an adverse effect.

In the event of lethality observed in a toxicity study it is prudent to determine whether the death is due to toxicity or an exaggerated pharmacological effect considering the clinical application. For example, death due to hemorrhage is sometimes observed in healthy animals after administration of biopharmaceuticals with an anticoagulation activity. Likewise death due to hypoglycemia can occur in healthy animals after administration of insulin. To attribute these death cases to toxicity has little value for determining human safety in clinical practice. Because these changes are observed only in healthy animals, and because biopharmaceuticals are prescribed for the normalization of abnormal functions in patients (e.g., hypercoagulopathy or hyperglycemia) through their pharmacological activity, one can easily assume that hemorrhage or hypoglycemia due to excessive expression of the pharmacological actions may occur.

***Metabolism Study*** The degradation of a protein to peptides and an amino acid moiety is commonly expected as a representative metabolic pattern. Therefore conventional biotransformation studies are not needed for biopharmaceuticals consisting of natural amino acids. However, metabolism studies of biopharmaceuticals containing nonnatural amino acid may provide useful information. In such cases radiolabeled proteins should be prepared to trace the pharmacokinetic behavior of the nonnatural amino acid fragments.

***In vitro Electrophysiological Study*** In vitro electrophysiological studies are generally not applicable to biopharmaceuticals. This is because NCEs act on each cellular channel after passing through the cell membrane, whereas biopharmaceuticals are not expected to act similarly because they cannot pass through the cell membrane.

***Single-Dose Toxicity Study*** An objective of single-dose studies is to define the relationship of dose with systemic and/or local toxicity. For biopharmaceuticals repeatedly administered clinically, the data from single-dose toxicity studies can be used to select doses for repeated-dose toxicity studies. Repeated-dose toxicity studies of biopharmaceuticals that have minimal toxicity can be performed without conducting single-dose toxicity studies under GLP conditions. Therefore single-dose toxicity studies in two animal species are not considered to be as necessary for these biopharmaceuticals as they are for NCEs. When a single-dose toxicity study is necessary, single-dose toxicity can be evaluated as a component of safety pharmacology or primary pharmacodynamic studies using animal models. When the doses set for the repeated-dose toxicity study are reasonable, the initial administration data obtained from the repeated-dose toxicity study can be used as data for single-dose toxicity studies. Since biopharmaceuticals need not be examined at high doses such as the approximate lethal dose, conducting single-dose toxicity studies merely to obtain information on the potential of toxic substances among others would be meaningless. Single-dose toxicity studies in nonrodents only should be considered in cases where rodents are not considered a relevant species.

***Repeat-Dose Toxicity Studies*** Typically toxicity studies are performed in two animal species. However, for toxicity studies for which there is only one relevant animal species, these studies may be performed using one animal species. When two animal species show the same toxicity profile in short-term studies, only one animal species may be used in long-term studies. Comparison of toxicity profiles means comparing the type and severity of any toxicity observed. However, biopharmaceuticals with a low toxicity may display no toxicity at high doses in some cases. It such cases it may still be important to select one species for assessment of chronic toxicity. More important, in cases where toxicity has not been clearly demonstrated justification of human

safe human doses would be based on an understanding of the therapeutic index (i.e., minimum effective biological dose and the maximum feasible dose).

***Reproductive and Developmental Toxicity Studies*** The requirement for reproductive and developmental toxicity studies depends on the clinical indication and intended patient population. For example, when (1) no relevant animal species exists, (2) a biopharmaceutical is not used for pregnant women or women of child-bearing potential (3) there is a structurally comparable natural biopharmaceutical for which there is much experience in clinical practice, or (4) a biopharmaceutical is indicated for patients with minimal child-bearing potential and indicated for those with serious diseases, reproductive, and developmental toxicity studies could be obviated.

When standard reproductive and developmental toxicity studies are not feasible due to problems of neutralizing antibody formation, although it is deemed necessary, the study design and dosing schedule may be modified on the basis of factors related to species specificity, immunogenicity, biological activity, and/or elimination half-life. For example, a reproductive and developmental toxicity study with a shorter periodic dosing than the whole-period dosing shown in the toxicology guideline for NCE can be meaningful. In addition alternative studies using relevant transgenic animals, or homologous proteins, should be considered. However, as reproductive performance in transgenic animals has not yet been clarified, careful selection of a study system is required.

The points to consider on the need for assessment of reproduction toxicity of human insulin analogues have been published [6].

***Carcinogenicity Studies*** When the following points are confirmed, generally, carcinogenicity studies are not necessary even when a biopharmaceutical is used for a long period:

1. It is used for substitution therapies at the physiological level.
2. It has no physiological activity that differs from that of endogenous substances.
3. Its biological action is not significantly stronger than that of endogenous substances.
4. It has no potential to induce tumor cell division (in the case of growth promoters).
5. It neither locally retains nor accumulates at a high concentration for a long period of time.
6. It does not have a sustained pharmacological action.
7. In repeated-dose toxicity studies when a dosing duration adequate for evaluation is attained, no preneoplastic lesions are observed.

8. The results of genotoxicity studies are negative in the case that the studies are relevant and have been conducted (e.g., bioconjugate with organic chemical linker).

Assessment of the carcinogenic/tumorigenic potential should be considered for some biopharmaceuticals because of the dosing duration, the relationship of target diseases with cancer, the biological activity of a product, the presence or absence of immunosuppressive action, in vitro data, and so forth. In the cases where a product is biologically active and nonimmunogenic in rodents and other studies have not provided sufficient information to allow an assessment of carcinogenic potential, then the use of one rodent species should be considered. Careful consideration should be given to the selection of dose. The use of a combination of pharmacokinetic and pharmacodynamic endpoints, with consideration of comparative receptor characteristics and intended human exposures, represents the most scientific approach to defining the appropriate dose. The rationale for the selection of dose should be provided. Points-to-consider on the preclinical assessment of the carcinogenic potential of insulin analogues have been published [7].

***Genotoxicity Studies*** It is generally not applicable to routinely implement the genotoxicity studies of biopharmaceuticals as required for NCEs. Proteins and peptides are not expected to interact directly with DNA or other chromosomal materials by passing through the cell membrane. On the other hand, ICH S6 describes "With some biopharmaceuticals, there is a potential concern about accumulation of spontaneously mutated cells (e.g., via selectively facilitating a predominating factor of proliferation) leading to carcinogenicity, alternative in vivo or in vitro models to address such concerns may have to be developed and evaluated." When in vitro or in vivo data suggest a potential biopharmaceuticals' ability to strongly stimulate cell proliferation, carcinogenicity studies should be considered. In the case of human-type proteins or peptides, it would be helpful to assess the necessity of conducting further studies to compare the physiological concentration of a biopharmaceutical in blood or tissue with that at which an enhanced activity of the biopharmaceutical on cell proliferation is observed. Human protein analogues should be evaluated for potential difference in activity from natural human-type proteins.

Genotoxicity studies should be considered for bioconjugates having an organic chemical linker molecule or proteins with nonnatural amino acids. The requirement for genotoxicity studies depends on whether a biopharmaceutical is a natural protein or an analogue such as a bioconjugate. Genotoxicity studies are not required for natural proteins, because they are not expected to interact directly with DNA or other chromosomal materials after passing through the cell membrane and because natural proteins are degraded into only natural amino acids. In the case of protein analogues containing nonnatural amino acids, it may be necessary to assess their genotoxicity under metabolic activa-

tion conditions, since there might be a possibility that chemical compounds having unknown activity can be formed by metabolic degradation. However, genotoxicity studies can be obviated even by demonstrating the inability of such a protein analogue and the fragments into cells using radiolabeled biopharmaceuticals or by other scientific evidence for their lack of genotoxicity.

### 3.3 Items for Future Update of Japanese Points-to-Consider Document

***New Types of Biopharmaceuticals*** There is accumulated knowledge and experience on new types of biopharmaceuticals since the issuance of the Japanese Points-to-Consider document in 2002 [3]. One of new types of biopharmaceuticals are the bioconjugates, namely PEGylated proteins. PEGylation is a useful method of improving the therapeutic potential of proteins by changing pharmacokinetics and in vivo pharmacodynamics. The preclinical information from a number of PEGylated proteins is helpful for the further clarification of safety assessment programs (particularly, metabolism, in vitro electrophysiological and genotoxicity studies) for bioconjugates. Another new category of biopharmaceuticals is therapeutic antibodies. The section on antibodies in the Japanese Points-to-Consider document should be revised because there are examples of therapeutic antibodies that have much more potent biological activity than those developed in the 1990s. These antibodies act strongly on biological systems of the body and sometimes induce marked biological changes. Furthermore some of the therapeutic antibodies acting on the immune system may induce cytokine release that results in a "cytokine release syndrome" in humans. This syndrome is one of the most serious adverse events caused by biopharmaceuticals in humans. Therefore intensive investigations on the mechanism underlying this syndrome and preclinical studies of the prediction of cytokine release syndrome are needed for those biopharmaceuticals with potent pharmacology toward activation of the immune system.

***New Technologies and Assays*** The use of transgenic animals for the safety assessment of biopharmaceuticals has become easier and less costly over the years. Transgenic animals can provide useful information when no relevant animal species are available, although there are limitations in terms of the historical background and how data will be used in the evaluation of the margin of safety. The use of homologous proteins is also a useful alternative when the biological activity of a biopharmaceutical is not properly studied in animals. Homologous proteins should be designed to produce similar biological activities in animals used for the safety assessment to those of biopharmaceuticals that will be used in humans.

In vitro electrophysiological studies have been introduced in ICH S7B [8] for the evaluation of the potential of QT prolongation of ECG for NCEs. Those studies are usually not applicable for biopharmaceuticals. This is because NCEs act on each type of cellular channel after passing through the cell

membranes, whereas biopharmaceuticals consisting of natural amino acids are not expected to act similarly because they cannot pass through the cell membrane. However, we may need to confirm whether this argument is relevant for bioconjugates and proteins containing nonnatural amino acids, since there might be a possibility that unknown chemical compounds can be formed by metabolic degradation. Likewise we should revisit the justification for not conducting genotoxicity studies for bioconjugates and proteins containing nonnatural amino acids. Preclinical studies of prediction of cytokine release syndrome are required for some categories of therapeutic antibodies acting on the immune system, as described in the previous section. There would be a species difference in responsiveness in terms of cytokine release induced by humanized antibodies. Therefore an in vitro cytokine release assay using human blood may also need to be considered.

***Timing of Preclinical Studies*** Neither ICH S6 [2] nor ICH M3 [9] defines the timing of preclinical studies to support biopharmaceutical clinical development. Considering limitations in the prediction of adverse effects in humans using preclinical models, it may be most important to clarify what preclinical information is useful for the first trial in humans. One should note that no toxicity is often observed even at a high dose of a biopharmaceutical in animals that are not biologically responsive to the biopharmaceutical. Preclinical programs that support the first dosing in humans may need to be optimized on a case-by-case basis. In particular, the necessity for in vitro electrophysiology, genotoxicity, and cytokine release studies should be justified by taking into consideration the types of biopharmaceutical and available information on the class of biopharmaceutical tested. There are some differences in preclinical studies required for phases 2 and 3 of clinical trials and registration between biopharmaceuticals and NCEs. For example, the durations of nonrodent repeat dosing studies are six and nine months for biopharmaceuticals and NCEs, respectively. ICH S6 allows the use of only one species for subsequent long-term studies if the toxicity profiles of a biopharmaceutical in two species are comparable in a short-term study. Moreover only one animal species may be used for a carcinogenicity study for a biopharmaceutical when the study is required. These different approaches described in ICH S6 are useful for most biopharmaceuticals. Therefore the preclinical programs after conducting the first trial in humans should be optimized on a case-by-case basis rather than determined by a common study package.

## 4 CONCLUSION

The safety assessment of biopharmaceuticals in preclinical studies has been improved in Japan with the implementation of the ICH S6. The analysis of data from a questionnaire survey conducted by JPMA suggests that the ICH S6 was well understood and adequately implemented in Japan. The Japanese

Points-to-Consider document helps industry scientists and regulatory reviewers understand the ICH S6 guideline. In particular, it is helpful for the clarification of case-by-case approaches to preclinical programs depending on the biopharmaceutical type. However, further updates of the Japanese Points-to-Consider document may be needed as newer types of biopharmaceuticals, technologies, and assays have developed.

## ACKNOWLEDGMENT

I deeply appreciate valuable comments from Drs. T. Inoue and S. Kitajima from NIHS, H. Onodera from PMDA, F. Sagami from JPMA, and M. Kawai and S. H. Heidel from Eli Lilly.

## REFERENCES

1. Antigenicity tests. *Drug Approval and Licensing Procedures in Japan 1992*. Yakugyo Jiho, Tokyo, 1992, pp. 128.
2. Ministry of Health and Welfare. Preclinical Safety Evaluation of Biotechnology-Derived Pharmaceuticals, 2000. Notification No. 326.
3. Pharmaceutical Non-clinical Investigation Group. "Points-to-consider" regarding safety assessment of biotechnology-derived pharmaceuticals in non-clinical studies. In *Handbook on Non-clinical Guidelines for Pharmaceuticals 2002*. Tokyo: Yakujinipposya, 2002, pp 83–94.
4. Nakazawa T, Kai S, Kawai M, Maki E, Sagami F, Onodera H, Kitajima S, Inoue T. "Points-to-consider" regarding safety assessment of biotechnology-derived pharmaceuticals in non-clinical studies (English translation). *J Toxicol Sci* 2004; 29:497–504.
5. Kamita Y, Shiga T, Koga T, Nomura A, Shibata M, Maeda Y, Horigome H, Kurokawa M, Suzuki I, Niwano Y, Nakazawa T, Matsuzawa T. Points-to-consider on non-clinical safety studies on biotechnology-derived pharmaceuticals—Analysis of questionnaire survey data by JPMA. *Iyakuhin kenkyu* 2003;34:358–67.
6. CPMP/SWP/2600/01. Points-to-Consider on the need for Assessment of Reproduction Toxicity of Human Insulin Analogues, 2002. www.emea.eu.int
7. CPMP/372/01. Points-to-Consider on the Non-clinical Assessment of the Carcinogenic Potential of Insulin Analogues, 2001. www.emea.eu.int
8. ICH S7B Step 4. Non-clinical Evaluation of the Potential for Delayed Ventricular Depolarization (QT Interval Prolongation) by Human Pharmaceuticals, 2005.
9. Ministry of Health and Welfare. Non-clinical Safety Studies for the Conduct of Human Clinical Trials for Pharmaceuticals (revision), 2000. Notification No. 1831.

# Implementation of ICH S6: US Perspective

MARY ELLEN COSENZA, PhD, MS, DABT, RAC

**Contents**

1 Introduction
2 Historical Perspective
3 ICH S6 Guidance
  3.1 General Principles
  3.2 Animal Species Selection
  3.3 Administration/Dose Selection
  3.4 Immunogenicity
4 ICH S6 Specific Considerations
  4.1 Safety Pharmacology/Single-Dose Studies
  4.2 Exposure
  4.3 Repeat-Dose Studies
  4.4 Immunotoxicity Studies
  4.5 Reproductive/Developmental Studies
  4.6 Genetoxicity Studies
  4.7 Carcinogenicity
  4.8 Local Tolerance
5 Summary
  References

## 1 INTRODUCTION

Several years ago an ICH guidance (ICH S6 Preclinical Safety Evaluation of Biotechnology-Derived Pharmaceuticals) was written to address the preclinical development and safety issues of products derived from biotechnology. ICH S6 defined biotechnology-derived pharmaceuticals as "products derived

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

from characterized cells through the use of a variety of expression systems including bacteria, yeast, insect, plant, and mammalian cells" [1]. The scope of the guidance included proteins, peptides, derivatives of these or products of which they are components. It also states that these principles may apply to "recombinant DNA protein vaccines, chemically synthesized peptides, plasma derived products, endogenous proteins extracted from human tissues, and oligonucleotide drugs," but does not cover "antibiotics, allergenic extracts, heparin, vitamins, cellular blood components, conventional bacterial or viral vaccines, DNA vaccines, or cellular and gene therapies."

Regulatory agencies throughout the world define biopharmaceuticals differently, and in part this has affected how they are regulated. In the United States most, but not all, biopharmaceutical products have been developed in accordance with ICH S6 since its publication in 1997. However, some product classes specified in the guidance at that time fell under the FDA definition of hormones and chemically synthesized products (e.g., oligonucleotides) and were thus considered "drugs." These "drugs" were reviewed by regulatory scientists in the Center for Drug Evaluation and Research (CDER) and approved as new drug applications (NDAs). Therapeutic proteins, including antibodies, were regulated as "biologics" and were reviewed by regulatory scientists in the Center for Biologics Evaluation and Research (CBER) and approved as biologics license applications (BLAs). In 2003, the divisions of CBER responsible for review of ICH S6-specified, biotechnology-derived pharmaceuticals merged with CDER. This merger changed the FDA center responsible for primary review of these products. The products, however, were, and continue to be, approved as BLAs [2]. The "traditional biologicals" (i.e., vaccines and blood products and the novel cellular, tissue, and gene therapy products including tissue engineered products) remained in CBER (see Table 1). In June 2003 the CBER staff (pharm/tox experts and medical officers) in the Office of Therapeutics were transferred to CDER, and resided in a new Office of Drug Evaluation, ODE VI. More recently this office was eliminated, and the various "biologics" reviewers were reassigned by therapeutic or disease area to be consistent with how pharmaceutical products are reviewed. Thus the review of biopharmaceuticals now takes place in the same review division as traditional drug therapeutics for similar indications.

Even before final approval there was a rapid adoption of the principles outlined by in the ICH S6 guideline both by industry and by the FDA. The main tenet of the ICH S6 guidance is to create a case-by-case, science-driven approach to biotechnology preclinical product development. That is to say, each molecule should be evaluated for both its physical and pharmacological properties. Therefore a product's pharmacological attributes needs to be understood in addition to its clinical use before embarking on a preclinical safety assessment development plan. The overarching goal of a preclinical development plan is to provide information for designing and conducting clinical trials ("the principles"). A case-by-case rational, science-based approach ("the practice") was defined, and has proved to be, the most appropriate way to develop biotechnology-derived products [3]. The approach

**TABLE 1 Regulation of biotechnology products**

| |
|---|
| *Products formally regulated by CBER, now by CDER* |
| • Proteins and modified proteins |
| • Cytokines |
| • Growth factors |
| • Ligands and receptors |
| • Antibodies |
| *Products regulated by CDER* |
| • Hormones |
| • Chemically synthesized peptides |
| • Oligonucleotides |
| *Products still regulated by CBER* |
| • Gene therapies |
| • Cellular therapies |
| • Engineered tissue products |
| • Vaccines |
| • Blood and blood products |
| • Antitoxins |

defined in S6 is appropriate for any pharmaceutical, regardless of its chemical or biological nature, that demonstrates species specificity, has an extended half-life, or requires a novel route of delivery.

## 2 HISTORICAL PERSPECTIVE

Over the past two decades many biotechnology-derived products have been approved in the United States. A selected list of these products is provided in Table 2. The products include recombinant endogenous-replacement proteins, cytokines, monoclonal antibodies, and fusion molecules. Other chapters in this book give more detailed "product-class-specific" descriptions of the preclinical development programs for many of these molecules.

In the early days of biotechnology product development, the focus was on "quality" issues [4] or process-related impurities. The concerns at that time were for carryover of other cellular proteins and DNA and for contamination with endotoxins, chemicals, and viruses. Of course, these concerns still exist, but methods for purification and assays for evaluation of clearance have alleviated the need for the safety assessment scientist to focus on contaminants; instead they are now asked to focus on the pharmacological activity of the molecules. An ICH guidance (Q6B Specifications: Test Procedures and Acceptance Criteria for Biotechnological/Biological Products) addresses the specific issues related to the manufacturing process [6]. Other product-related issues such as impurities do need to be considered by the safety assessment scientist, for

**TABLE 2 Examples of marketed biotechnology products**

| Year of Approval | Approved Biologics | Trade Name |
|---|---|---|
| 1982 | Insulin | Humulin® |
| 1985 | Growth Hormone | Protropin® |
| 1986 | Interferon-alpha | Roferon®, Intron A® |
| 1986 | Muromonab CD3 | Orthoclone OKT3® |
| 1987 | TPA (Alteplase) | Activase® |
| 1989 | Epoetin-alfa | Epogen® |
| 1990 | Interferon-gamma | Actimmune® |
| 1991 | Filgrastim (G-CSF) | Neupogen® |
| | Sargramostim (GM-CSF) | Leukine® |
| 1992 | Interleukin-2 | Proleukin® |
| | Antihemophilic factor | Recombinate®rAHF |
| 1993 | Interferon-beta | Betaseron® |
| | Dornase Alfa | Pulmozyme® |
| 1994 | Imiglucerase | Cerezyme® |
| | Abciximab | ReoPro® |
| 1997 | Rituximab | Rituxan® |
| | Daclizumab | Zenpax® |
| | Oprelvekin | Neumega® |
| | Becaplermin gel | Regranex® |
| 1998 | Traztuzumab | Herceptin® |
| | Infliximab | Remicade® |
| | Basiliximab | Simulect® |
| | Palivizumab | Synagis® |
| 2000 | Tenecteplase | TNKase® |
| 2001 | Pegfilgrastim | Neulasta® |
| | Darbepoetin | Aranesp® |
| | Alemtuzumab | Campath® |
| | Drotecogin alpha | Xigris® |
| 2002 | Interferon beta 1-a | Rebif® |
| | Ibritumomab tiuxetan | Zevalin® |
| 2003 | Alefacept | Amevive® |
| | Tositumomab | Bexxar® |
| | Efalizumab | Raptiva® |
| | Omalizumab | Xolair® |
| 2004 | Bevacizumab | Avastin® |
| | Natalizumab | Tysabi® |
| | Palifermin | Kepivance® |
| | Technetium 99m Tc fanolesomab | NeutroSpec® |
| 2005 | Abatacept | Orencia® |

*Sources*: www.fda.gov; and www.bio.org.

example, genetic variants, aggregate forms, chemical linkers, and differences in glycosylation patterns.

In 1997 the FDA issued an addendum to the original Points-to-Consider Document for Developing Monoclonal Antibodies. This document addressed

issues covering all aspects of drug development for these molecules. Items specific to safety assessment included the requirement for tissue cross-reactivity studies. These studies allow for evaluation of binding in tissues from different species with the idea that this could help justify species selection. Over the years other more specific methods have generally been used to select appropriate species for safety assessment studies. Binding is no longer considered sufficient to justify the selection of a species. Other proof of biological activity is generally desired as well. The next section reviews specific sections of the ICH S6 document and discuss how this guidance has been implemented in the United States.

## 3 ICH S6 GUIDANCE

### 3.1 General Principles

The General Principles section acknowledges that "conventional approaches to toxicity testing of pharmaceuticals may not be appropriate for biopharmaceuticals." This has led to the proposal for case-by-case, science-driven drug development. Some of the unique challenges that the biopharmaceutical molecules face include species specificity, immunogenicity, unique routes of administration, and intermittent dosing schedules. The addition of pharmacological parameters to standard GLP toxicology studies, or the addition of toxicology parameters to pharmacology studies, allows for the judicious use of animals. The challenge of conducting some of these assays under GLPs was recognized in this guidance and by the FDA. Over the years more and more of these specialty assays (e.g., immunotoxicity assessment in primates using flow cytometry) have been validated and are being conducted in compliance with GLPs.

### 3.2 Animal Species Selection

One of the greatest challenges in the preclinical development of biotechnology-derived molecules is that of species specificity. Unlike traditional "small molecules," such a molecule cannot be assumed to be active in the standard species used for toxicity testing. The lack of pharmacological activity in a species can then lead to "no effects" in that species. With a few exceptions, where there were potential effects of contaminants, these studies had little purpose and a questionable use of animals. Justification for using only one species can be based on the lack of a second species with biological activity or when the biological activity of the molecule or target toxicity can be adequately defined in just one species.

An example where the original species was not relevant is that of the early work performed on the interferons. The studies were performed in nonrelevant species and generated misleading information [7]. Several studies were conducted in rodents with recombinant human interferons with little evidence of

toxicity. This did not predict what was to occur in humans. Activity in monkey studies was more predictive, but there was the additional challenge of relative differences in the activity level between humans and nonhuman primates.

The Animal Species Selection section of the ICH document also refers to the use of homologous proteins and transgenic animals that express the human receptor. One example of a development program that relied on surrogates for safety assessment is that of infliximab [8]. Many of the challenges of these models are acknowledged in this section. Animal models of disease are also discussed and can be used with strong scientific rationale.

### 3.3 Administration/Dose Selection

Unlike most traditional pharmaceutical small molecules, biologics are not given once or twice a day, orally, in a pill or capsule. They are almost all dosed via a parenteral route and are not given (or "taken") daily. Some of these therapeutics are given in hospital settings or in clinicians' offices. Several are now self-administered on an approximately weekly basis. The dose regimen for the safety assessment studies should reflect the dosing regimen plan for the human studies. The dosing interval in the toxicology studies may be different depending on the half-life of the molecule in animals versus humans. Also the development of antibodies in the animals may alter the pharmacokinetics, and modifying the dosing interval has been shown to reduce the incidence of antibodies.

Dose selection can be a challenge with biopharmaceuticals. Often the dose-limiting toxicity is related to the pharmacology (often referred to as exaggerated pharmacology), and it can be difficult to establish a margin of safety. The slope of the dose–response curve between the intended level of effect and an effect that leads to toxicity (even if it is the same effect, e.g., an increase in hematocrit with erythropoietin-like molecules) may be very steep. The ability to produce formulations of proteins that allow for "high" doses can be difficult, especially if the protein formulation is viscous. Doses can be limited by practicality as well (dosing volume limits, maximal stable concentrations, etc.). There is always the desire to study a dose without effect, but this can be an additional challenge for molecules that are active at doses in the μg/kg level.

Other factors that need to be considered in dose selection are related to pharmacokinetics and drug metabolism. One area generally not deemed relevant for biologics is metabolism studies. Mass balance studies have not proved useful when performed with proteins. Radiolabeled studies have limited usefulness as these labeled molecules undergo rapid metabolism and the label is often unstable. Protein biologics undergo proteolytic degradation into smaller peptides and then into individual amino acids. Distribution is viewed more from the perspective of target distribution, except for gene and cell therapies and viral or bacterial vectors.

Recovery periods are very important considerations for studies with biopharmaceuticals. For many biopharmaceuticals, the pharmacodynamic effects

are prolonged either pharmacologically, through secondary "cascading" effects through various pathways, and/or as a consequence of a long half-life. It is important to have a recovery/washout period of the test agent in order to properly test for antibody levels. The longer recovery can greatly increase the ability of the assays to detect antibodies to the drug. Active test article levels in blood samples can interfere with the ability to measure true antibody levels.

### 3.4 Immunogenicity

It is generally well accepted that immunogenicity is not well predicted across species. It has also been well documented that many biological compounds intended for human use are immunogenic in animals [9]. ICH S6 states: "Most biotechnology-derived pharmaceuticals intended for humans are immunogenic in animals." Traditional antigenicity studies or guinea pig anaphylaxis studies are not useful for predicting immunogenicity in humans and are now generally recognized as not being appropriate studies for biologics. When these studies were conducted years ago, at the request of some regulators, they were generally positive and led to adverse effects in animals. Since there is little to no predictive value in these studies, and they were not considered appropriate, such studies have not been conducted since publication of ICH S6.

All preclinical (and clinical) studies with biopharmaceuticals should include measurements of total incidence of antibodies and a further characterization as to whether these antibodies are neutralizing. Assays to determine antibodies have become more sophisticated over the years, and the newest technologies allow detection at lower levels than was achievable with traditional ELISAs. Clinically relevant antibodies include those that are clearing, sustaining, neutralizing, and/or cross-react with endogenous proteins [10]. It is important to screen for the presence and development of antibodies to the test article throughout development. Since the consequence of these antibodies may range from no clinically significant effects to serious safety effects, assays need to be developed to determine if antibodies that appear are able to block the biological activity of the test compound. This may occur either by direct binding to an epitope with activity or by binding to a site in close proximity, resulting in steric hindrance to the active site.

Immunogenicity is a substantial complication for preclinical safety assessment studies. Antibodies can invalidate the animal model species. Antibody production alone, however, should not necessarily prohibit the conduct of these studies. The effect on pharmacokinetics and pharmcodynamics needs to be measured and evaluated. The potential consequences of the antibodies on endogenous molecules also needs to be evaluated. Secondary effects, such as antibody deposition, should be measured. The lack of ability to predict absolute human immunogenicity does not preclude the use of animals to assess the relative potential for an immune response.

## 4 ICH S6 SPECIFIC CONSIDERATIONS

### 4.1 Safety Pharmacology/Single-Dose Studies

Safety pharmacology parameters (e.g., cardiovascular) can be incorporated into standard toxicology studies, including single-dose studies if that is most appropriate. Molecules that cross-react in traditional species (rodents and dogs) allow for more flexibility in designing these studies. Molecules that only have activity in nonhuman primates can be more challenging. Knowledge of the intended molecular target and its presence on these target organs can aid in the decision as to whether specific studies will need to be conducted. Primates can be telemeterized for cardiovascular assessments. In recent years many CNS parameters have been validated in primates as well. In vitro assessments are generally not relevant for biologics for many of the same technical reasons well recognized with in vitro genotoxicity assays.

### 4.2 Exposure

Toxicokinetics must be assessed in preclinical toxicology studies, as with traditional pharmaceuticals. These data are necessary to prove exposure (which may differ with route) and to monitor the potential effects of antibodies on the exposure levels over time. Perhaps the greatest difference between small-molecule pharmaceuticals and large-molecule biopharmaceuticals is the potential for immunogenicity, which can greatly affect pharmacokinetics as previously discussed.

### 4.3 Repeat-Dose Studies

The ICH S6 document states: "This duration of animal dosing has generally been 1–3 months for most biotechnology derived pharmaceuticals." And then: "for chronic indications, studies of 6 months duration have generally been appropriate, although in some cases shorter or long durations have supported marketing authorizations." Most important, it states: "the duration of long-term toxicity studies should be scientifically justified." The six-month paradigm was based on the idea that all or most toxicity would be elicited during this time frame for molecules that are highly targeted and for which their toxicity is mainly based on exaggerated pharmacology. Although it may not always be as straightforward as this, six months has proved to be sufficient for the elucidation of chronic toxicity for almost all biologics that have been developed. Requests for longer term toxicity studies (9 or 12 months) have rarely added useful information to the safety assessment of these molecules [11]. This does not mean that six months should be a default without rigorous scientific evaluation. If there are indications of unexpected toxicities, or other scientific reasons to expect delayed effects, then longer term studies should be considered. Without these concerns six months should continue to be sufficient. If

there are indications of carcinogenic potential, then those questions should be dealt with via more specific studies. The issues of species specificity or the use of homologous proteins makes these questions more complicated and are another reason to consider the need to conduct longer term studies carefully. Longer term dosing may also increase the possibility of immunogenicity.

### 4.4 Immunotoxicity Studies

Since many biotechnology products are designed to modulate the immune system, basic immunotoxicity parameters have traditionally been evaluated as part of standard toxicity studies. Over time more and more immunotoxicity specific parameters have been validated in toxicology species, including primates. These parameters (humoral and cell-mediated) should be measured as appropriate. Other specific issues for biopharmaceuticals include immunogenicity, as previously discussed.

### 4.5 Reproductive/Developmental Studies

When a biopharmaceutical product cross-reacts with traditional reproductive toxicology species, these studies should be conducted as appropriate for the intended clinical population. When conducting these studies in rodents and rabbits, care should be taken to select the dosing regimen and to consider the impact of immunogenicity. The timing of dosing in these studies might need to be more frequent to ensure that there is adequate exposure during the pivotal stages of gestation. The exposure period for the segment 2 studies in rodents and rabbits is usually short enough to avoid a strong immunogenic response, but samples should be taken to measure test article levels and antibodies in the dose–range-finding studies.

When a molecule does not cross-react with rodents or rabbits, then a decision must be made as to whether to conduct these studies in primates or to make a rodent surrogate. There are several factors that need to be considered. For example, human IgGs are known to cross the placental barrier, and their response to teratogens has been shown to be similar to humans. Compared to rodents, nonhuman primates have a small number of offspring (usually just one offspring per mother), they are very expensive, and their gestation period is 150 days. In addition nonhuman primates have a low conception rate and a high spontaneous abortion rate, and the historical database for primate reproductive studies is not that large. The FDA has been more accepting of primate studies than agencies in other geographic regions. (See Chapters Reproductive Toxicity Testing for Biopharmaceuticals and Reproductive/Developmental Toxicity Assessment of Biopharmaceuticals in Nonhuman Primates for discussions of reproductive toxicity studies for biopharmaceuticals.) The considerations above must also be balanced against the challenges of developing and validating a surrogate to use in rodents or rabbits.

### 4.6 Genotoxicity Studies

It is generally accepted now that genotoxicity studies are not applicable for biopharmaceuticals, unless there is a "chemical" linker or toxic conjugate to the molecule. This is also clearly stated in the FDA Points-to-Consider Document for Monoclonal Antibodies [5].

### 4.7 Carcinogenicity

Carcinogenicity studies are generally considered inappropriate for biopharmaceuticals. Both the ICH S6 document and the ICH S1A (The Need for Long-term Rodent Carcinogenicity Studies of Pharmaceuticals) support this position [12]. ICH S6 states that need for evaluation of carcinogenic potential of these molecules depends on "duration of clinical dosing, patient population, and/or biological activity." Most of the early biotechnology molecules developed were for severe clinical indications and/or addressed unmet medical needs.

Currently (December 2007) there are no published traditional carcinogenicity studies for ICH S6 specified biotechnology-derived pharmaceuticals. Conducting a traditional two-year bioassay with these molecules is extremely challenging due to species specificity, immunogenicity, and the challenges of using homologous molecules. When there are reasons to be concerned about related issues such as potentially enhancing the growth of existing tumors through the intended pharmacological activity, then more specific studies have been considered. These studies have included (among others) assessments of the presence or absence of the drug's receptor on relevant tumor cells, effects of the drug on in vitro tumor cell growth rate, and effects of the drug on the growth rate of tumor cell xenografts in mice. Measurements of cell proliferation have also been added in some cases to repeat-dose animal studies as per ICH S6 when there is cause for concern.

### 4.8 Local Tolerance

Local tolerance assessments are usually incorporated into repeat-dose toxicity studies. Specific assessments include clinical observations (e.g., Draize scoring) and macroscopic and microscopic evaluations of the injection site. These types of studies may also be used to test formulation changes during the course of clinical development.

## 5 SUMMARY

The safety evaluation of biotechnology products has been an evolving process over the past two decades. ICH S6 did much to help give guidance to preclinical scientists (both in industry and government) on how to approach these development plans. In the future the basic questions for these products will remain the same, but the challenges will be greater. Global strategies for

development of biopharmaceuticals should be more science driven and problem focused. If we (industry scientists) want the reviewers (regulatory scientists) at the FDA (and at all agencies worldwide) to stick to the scientific principles addressed in ICH S6, then the safety assessment industry scientist must be able to use science (e.g., target liability) to justify development plans as well. Defaulting to ICH S6 without scientific justification is as imprudent as conducting traditional small-molecule pharmaceutical studies on biopharmaceuticals inappropriately.

With the changes in the global regulatory environment there are likely to be additional challenges. The overarching goal should still be to evaluate preclinically the potential effects of these molecules and to provide guidance to the clinicians. Safety assessments approaches should have a strong scientific rationale and use the appropriate animal species judiciously.

## REFERENCES


1. ICH S61. Preclinical Safety Evaluation of Biotechnology-Derived Pharmaceuticals. July 1997.
2. Schwieterman WD. Regulating biopharmaceuticals under CDER versus CBER: An Insider's perspective. *Drug Discov Today* 2006;11:19–20.
3. Cavagnaro JA. Preclinical safety evaluation of biotechnology-derived pharmaceuticals. *Nat Rev* 2002;1:469.
4. Dayan AD. Safety Evaluation of biological and biotechnology-derived medicines. *Toxicology* 1995;105:59.
5. FDA, CBER. Points to Consider in the Manufacture and Testing of Monoclonal Antibody Products for Human Use, 1997.
6. ICH Q6B. Specifications: Test Procedures and Acceptance Criteria for Biotechnological/Biological Products, August 1999.
7. Serabian MA, Pilaro AM. Safety assessment of biotechnology-derived pharmaceuticals: ICH and beyond. *Toxicol Pathol* 1999;27:2.
8. Treacy G. Using an analogous monoclonal antibody to evaluate the reproductive and chronic toxicity potential for a humanized anti-TNFα monoclonal antibody. *Hum Exper Toxicol* 2000;19:226.
9. Shankar G, Shores E, Wagner C, Mire-Sluis A. Scientific and regulatory considerations on the immunogenicity of biologics. *TRENDS Biotech* 2006;24:274–80.
10. Koren E, Zuckerman LA, Mire-Sluis AR. Immune responses to therapeutic proteins in humans—clinical significance, assessment and prediction. *Curr Pharmaceut Biotech* 2002;3:349.
11 Clarke J, Hurst C, Martin P, et al. Duration of chronic toxicity studies for biotechnology-derived pharmaceuticals: is 6 months still appropriate? *Regul Toxicol Pharmacol* 2008;50:2–22.
12. ICH S1A. The Need for Long-term Rodent Carcinogenicity Studies of Pharmaceuticals, March 1996.

# Comparison of Preclinical Development Programs for Small Molecules (Drugs/Pharmaceuticals) and Large Molecules (Biologics/Biopharmaceuticals): Studies, Timing, Materials, and Costs

CHRISTOPHER HORVATH, DVM, MS, DACVP

**Contents**

1 Introduction
2 Objective and Methods
3 Distinguishing Biologics from Drugs
  3.1 Properties of Drugs and Biologics
  3.2 Influence of Drug and Biologic Properties on Development Program Design
  3.3 Applicable Regulatory Guidelines for Drugs and Biologics
4 Comparative Preclinical Development Programs for a Hypothetical Drug and Biologic
  4.1 General Assumptions
  4.2 Therapeutic Target, Mechanism of Action, and Clinical Indication
  4.3 Clinical Development Plan to Be Supported
  4.4 Specific Assumptions
  4.5 "Traditional" Preclinical Development Plans to Support Clinical Development by Phase
  4.6 "Alternative" or Novel Preclinical Development Plans for Biologics
5 Messages for Preclinical Development Programs for Biologics
6 Summary
References

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

## 1 INTRODUCTION

Successful and efficient development of a new pharmaceutical requires the planning of an integrated development program that coordinates the trilogy of product manufacture—chemistry, manufacturing, and controls (CMC), preclinical studies (distribution, metabolism, and pharmacokinetic [DMPK], pharmacology and toxicology), and clinical trials—within the framework of the regulatory development strategy. Preclinical (often referred to as nonclinical) safety studies must support each successive phase of clinical development, as well as any significant changes to the method(s) of manufacturing, formulating, or administering the pharmaceutical. A desirable goal for the preclinical portion of the integrated development program should be to enable each phase of clinical development and each significant change in CMC methodology without ever becoming rate-limiting to the pharmaceutical development time line. At face value, this would seem to be a relatively simple task if one adheres to available guidance documents: International Conference on Harmonization (ICH) M3 [1] outlines the necessary types of studies and their timing relative to phase 1, 2, or 3 clinical trials, additional ICH guidelines (S1A through S8) [2–13] discuss selected types of safety studies in detail, ICH S6 [14] discusses special considerations for biopharmaceuticals (biotechnology-derived pharmaceuticals), and ICH Q5E [15] addresses comparability testing for biopharmaceuticals. However, the preclinical program must also evaluate potential changes in pharmacology, DMPK, immunogenicity, and/or safety properties as evolutionary changes are made to the process or scale of manufacture, the formulation, and the route or method of administration of the pharmaceutical.

This chapter had its origins in a discussion that occurred during a June 2004 meeting between Food and Drug Administration (FDA) pharmacology and toxicology reviewers and BioSafe, a committee of experts experienced in preclinical development of biotechnology-derived products organized under the auspices of the Biotechnology Industry Organization (BIO). The intent of that meeting was to review and discuss some of the inherent differences among preclinical development programs for biologics and drugs, particularly the complexities of reproductive toxicology testing of biologics. It was agreed that, consistent with ICH S6, preclinical programs for biopharmaceuticals should be designed on a case-by-case basis to address *relevant* safety issues, and that novel approaches might be necessary to accomplish this. For example, the preclinical program might require the use of nontraditional study designs, species, or endpoints or even the invention and use of a homologous (surrogate) test article. In these discussions it was difficult to reach agreement on what constitutes a *relevant* preclinical program, as the relevance of each preclinical program might depend on both the results of the component studies and their acceptance by the assigned FDA reviewers. Thus a conundrum was recognized: How does one design a program that might of necessity be unlike any other yet be reasonably likely to provide relevant and acceptable answers regarding safety?

For members of the pharmaceutical industry engaged in the biologics development this question raises practical concerns as well. While the resources and time needed for a traditional drug development program are well known, those for a nontraditional biologics development program are less well known, but might reasonably be expected to be greater. In addition, while the acceptability of a traditional drug development program is well established, the acceptability of each nontraditional biologics development program must be established. The degree of acceptability will be dependent on the degree of relevance, which can only be "known" retrospectively, or after completion of the program. Attempts to gain the a priori approval from FDA reviewers for nontraditional preclinical development programs are not always successful. Therefore nontraditional preclinical development programs may be run at risk of not attaining the necessary relevance and acceptability. For members of a pharmaceutical company's management this dilemma signals greater risk and a lower probability of success, and thus creates pressures to "de-risk" the preclinical development program for biologics.

From a purely pragmatic point of view, most pharmaceutical companies are interested in trying to determine the costs, resources, and time required to complete various different possible development programs for a product. For drugs or biologics, if multiple development pathways are possible and scientifically sound, it may be appropriate to pursue the path that is least costly or most rapid. For most drugs, the development path is relatively straightforward and well-trodden. For most biologics, finding an appropriate development path is often an exercise in trailblazing. The uncertainty of this process, as well as the inherent difference in some of the issues to be addressed for biologics, has the potential to result in more risky, costly, and time-consuming development programs.

To attempt to investigate some of the issues described above, I volunteered to try to assemble a comparative evaluation of the studies, timing, materials, and costs associated with preclinical development of a drug instead of a biologic. I based the comparison on experiences I have had with the preclinical programs for both a drug and a biologic (a monoclonal antibody) being developed for very closely related therapeutic targets in the same clinical indications. The resulting comparison was first presented at a February 2005 BioSafe/FDA reviewer workshop[1] and subsequently at the June 2005 BIO convention.[2]

## 2 OBJECTIVE AND METHODS

The objective of this chapter is to compare to the studies, materials, and costs associated with hypothetical preclinical development programs intended to

[1]Preclinical development of biologics and biotech derived pharmaceuticals: Principles and practices. University of Maryland Conference Center, Rockville, February 1, 2005.

[2]Key considerations in designing preclinical safety evaluation programs for biopharmaceuticals. Program designs and material needs: Investing for success. BIO 2005, Philadelphia, June 20, 2005.

support the clinical development of a small molecule (drug) and a large molecule (biologic).

To accomplish this objective, a detailed clinical development program was designed, from investigational new drug application (IND) through new drug or biologic licensing application (NDA or BLA). The respective preclinical development programs to support each filing were planned for a drug and biologic. Individual GLP- and ICH-compliant studies for both preclinical programs were then designed and submitted to different preclinical contract research organizations (CROs) for price estimates, to allow determination of the average price for each study. Although this comparison was conducted in 2004 and the study prices are likely no longer accurate, the relative prices of these study should remain comparable over time (see Chapter Preclinical Safety Study Design Templates and Estimated Costs updated listing of cost estimate for preclinical study designs). For an estimation of the amount of each test article needed for each study, the study designs included the doses to be evaluated, based on information available on each product at that time, and on common industry practices for setting doses.

## 3 DISTINGUISHING BIOLOGICS FROM DRUGS

For the purposes of this discussion, drugs are chemically synthesized pharmaceuticals, often described as "small molecules" or "new chemical entities"; biologics are "large molecules" or "biopharmaceuticals."

### 3.1 Properties of Drugs and Biologics

In addition to great differences in biophysical characteristics, the properties of biologics are vastly different than those of drugs. Key differences in the properties of drugs and biologics are summarized in Chapter The Principles of ICH S6 and the Case-by-Case Approach. It should be noted that some drugs have properties of biologics, and vice versa.

### 3.2 Influence of Drug and Biologic Properties on Development Program Design

The inherent differences in the properties of drugs and biologics strongly influence the design of their respective development programs. Some key considerations for the design of preclinical development programs for drugs and biologics are summarized in Table 1. Detailed discussion of each of these considerations is beyond the scope of this chapter. However, the magnitude of these differences can be illustrated by discussing one of the key considerations: selection of relevant specie(s) for preclinical testing.

As specified in ICH M3, general toxicology testing (e.g., single- and repeated-dose studies) should be conducted in two species: a rodent and a nonrodent. For drugs, the "default" choices are rats and dogs. The selection of these species (or occasionally other species) as the relevant species for preclinical studies is generally justified on the basis of comparative in vitro metabolism testing and demonstration of the extent to which the species-specific metabolism of the

**TABLE 1 Considerations in the design of preclinical safety evaluation programs for drugs and biologics**

| | Drugs | Biologics |
|---|---|---|
| Species specificity and activity | Often have pharmacologic activity in multiple species | Generally have pharmacologic activity restricted to a limited number of species |
| Selection of relevant rodent and nonrodent species for preclinical testing | Generally based on shared in vitro metabolite profile with humans. Pharmacologic activity is a secondary consideration. Rodent and nonrodent species are easily identified. | Generally based on shared pharmacologic activity. Identification of rodent species is generally difficult, if not impossible. Identification of nonrodent species is occasionally difficult. |
| Route of administration | Seldom changed during development | Often changed during development (e.g., IV to SC) |
| Manufacturing process and scale | May change manufacturing scale but seldom change process | Often change cell clone (e.g., higher producing), process, and/or scale (e.g., 500 to 2000 L) |
| Formulation | Seldom change formulation | Often change formulation (e.g., liquid to lyophilized) |
| Support to product changes during development | Generally established via biochemical specifications, occasionally with in vivo bioequivalence (PK or PK/PD) studies | Generally established via in vivo comparability testing. Often need to assess pharmacologic activity (PD) and immunogenicity in addition to PK. |
| Rate-limiting portion of development | Generally rate-limited by time needed for toxicology studies | Generally rate-limited by time needed for manufacturing material |
| Preclinical studies | Generally follow traditional, resource-defined designs | Often follow nontraditional (alternative), resource-intensive designs |
| Unique preclinical issues | Uncommon, though changing as drugs with properties of biologics are being developed (e.g., species-restricted pharmacologic activity) | Lack of relevant species compromises toxicity testing (e.g., limit to 1 species, unable to evaluate reproductive toxicity). Exposure-limiting immunogenicity compromises chronic studies |
| Availability of competent testing facilities | Readily available | Can be difficult to identify testing facilities with experience with nontraditional species, end points, study designs, routes of administration, assays, etc. |
| Acceptability of preclinical program to FDA reviewers | Predictable; often no need to get feedback from FDA reviewers prior to conduct or submission | Unpredictable. Generally a compelling need to get feedback from FDA reviewers prior to submission, and prior to conduct if possible. |

parent compound and the resulting metabolite profile is similar to that for humans. The intent is to identify any *active* metabolite(s) that might contribute to pharmacologic responses and to study in animals any *inactive* metabolite(s) that might contribute to toxicologic effects in humans. Pharmacologic activity in the chosen species, though desirable and often present, is not essential for toxicology testing. Because the focus of traditional toxicology testing is to discover and characterize unexpected and/or "off-target" toxicity, this approach to selection of relevant species is considered scientifically sound, even if the drug is not the pharmacologic activity in the chosen species. Thus the selection of relevant species for toxicity testing of drugs is generally based on metabolism and not on the pharmacologic mechanism of action.

In contrast, as specified in ICH S6, the primary consideration for the selection of a relevant species for toxicity testing of biologics is pharmacologic activity. This is because biologics are generally catabolized (degraded) rather than metabolized (chemically altered), and therefore generation of active metabolites with pharmacologic effects or inactive metabolites with toxicologic effects is uncommon. Thus off-target (nonpharmacologically mediated) toxicity is unlikely for biologics. This fundamental difference in the disposition of drugs and biologics means that the observed toxicity of biologics often represents "superpharmacology." It also means that an animal species in which there is no pharmacologic activity is generally not relevant for studying the safety of biologics. In recognition of this point, ICH S6 made allowances for use of only a single, relevant species in toxicity testing when such a decision could be scientifically justified.

### 3.3 Applicable Regulatory Guidelines for Drugs and Biologics

The ICH guidances pertaining to preclinical development [1–13] are generally applicable to both drugs and biologics, except as noted, with additional guidances specific to biotechnology-derived products [14, 15]. Careful reading and interpretation of all guidances is necessary to facilitate planning of the preclinical programs.

In addition to the ICH guidances discussed above, additional specific guidances related to pharmacology/toxicology evaluations to support clinical development should be considered based upon product class and/or indication (specific considerations for medical imaging agents, therapeutic radiopharmaceuticals, and pediatric drug products, as wells as for photo-safety testing, safety testing of drug metabolites, etc.).

## 4 COMPARATIVE PRECLINICAL DEVELOPMENT PROGRAMS FOR A HYPOTHETICAL DRUG AND BIOLOGIC

For this hypothetical comparison of the preclinical development programs for a drug and a biologic it was necessary to make some general assumptions about

the preclinical development programs, to define the therapeutic target, mechanism of action and clinical indication, to define the clinical development program, and then to make some specific assumptions for the drug and biologic.

### 4.1 General Assumptions

The general assumptions about the preclinical development programs were as follows:

- The comparison would begin at the point of nomination of the compound for clinical development; that is, the first series of studies would be IND-enabling.
- The respective preclinical development programs would represent typical, but idealized programs, with the focus being on safety only (e.g., no consideration of pharmacology or DMPK studies).
- All studies would be outsourced to contract research organizations (CROs) for study conduct, including any bioanalytical or other assays.
- All studies would comply with GLP regulations, as well as other appropriate guidelines (ICH M3, S6, etc.).
- All animals would be observed until test article had cleared (estimated to be five half-lives).
- All regulatory filings would be on final study reports.
- The manufacture of the respective test articles used in the studies would be outsourced to contract manufacturing organizations (CMOs) and sequential use of research-grade/non-GMP and GMP material would be allowed, if appropriate.
- The clinical development plan to be supported would be idealized, with minimal delays between phases and without substantive changes to the test article (e.g., no changes to manufacturing process or formulation) or the route of administration.
- No "run-at-risk" strategies would be allowed, such as concurrently validating the PK assay while collecting the samples to be assayed.

These assumptions were made to try to keep the number of variables to a minimum to allow this comparison, although hypothetical, to be as realistic as possible. Last, the key metrics to be compared would include the numbers and types of studies and their durations and costs, the quantity of material (test article) needed, the time required for manufacture and its costs, and the timing of the studies relative to clinical development.

### 4.2 Therapeutic Target, Mechanism of Action, and Clinical Indication

Designing an appropriate preclinical development program requires a complete understanding of the therapeutic target, mechanism of action, clinical

indication, and disease biology. This information is essential in choosing the relevant specie(s) and endpoints to be assessed. In addition it is necessary to have an understanding of the intended clinical route of administration, projected therapeutic concentration, projected therapeutic dose, projected dosing regimen, and the intended duration of dosing and exposure. These attributes are often summarized in the target product profile (TPP) that is generated by a project team during the development candidate nomination process. Detailed discussion of the methods for establishing this information is beyond the scope of this discussion.

For this hypothetical comparison, however, most of this information was known. The therapeutic targets were G protein-coupled receptors that were expressed on the cell surface of T cells, monocytes, and tissue macrophages. Through ligand-binding interactions these receptors were implicated in leukocyte trafficking, particularly to sites of active inflammation in autoimmune disease settings such as might be present in rheumatoid arthritis or multiple sclerosis. The intended mechanism of action was blockade of the receptor and thus prevention of ligand binding and subsequent leukocyte trafficking. The intended pharmacologic effect was downregulation of inflammation, or immune modulation. Data from animal models of disease induced in homozygous receptor knockout mice supported the targets, clinical indications, mechanism of action, and immunomodulatory effects of target knockout.

### 4.3 Clinical Development Plan to Be Supported

The proposed clinical development plan is summarized in Table 2 and illustrated in Figure 1. Based on the therapeutic target(s) and mechanism(s) of

**TABLE 2 Proposed clinical development program**

| Phase | 1a | 1b | 2a | 2b | 3 |
|---|---|---|---|---|---|
| Subjects | NHV | NHV | Patients | Patients | Patients |
| Women enrolled? | No | No | Post-menopausal | WCBP, DBP | WCBP |
| Regimen | SAD | MAD | MAD | MD | MD |
| Duration of dosing | Once | 2 weeks | 1 month | 3 months | ≥6 months |
| Phase duration (enrollment to report) | 6 months | 8 months | 18 months | 2 years | 3 years |

*Note*: DBP = double-barrier protection, MAD = multiple (repeated) ascending dose, WCBP = women of child-bearing potential, MD = multiple (repeated) dose, NHV = normal human volunteers, SAD = single ascending dose.

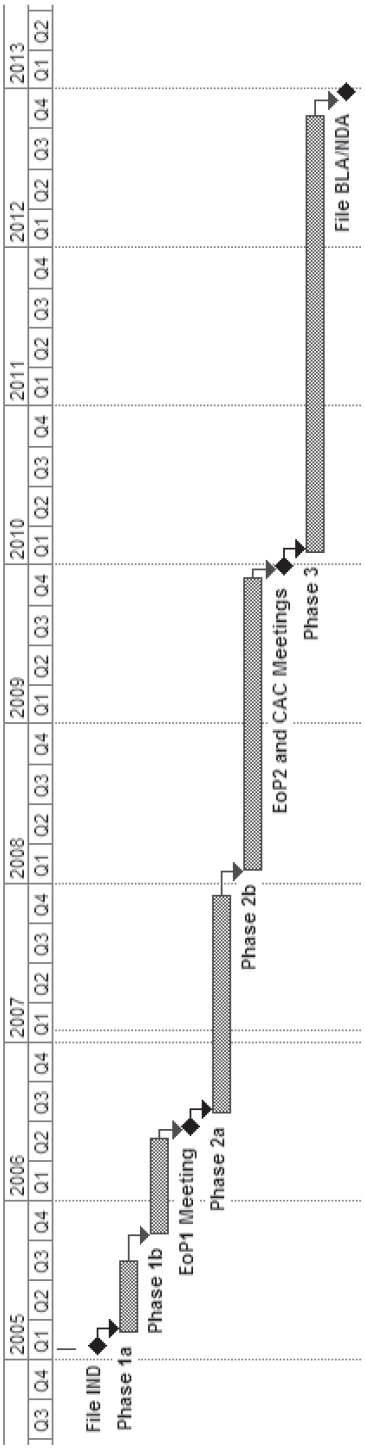


Starting date was assumed to be 18 January 2005.

**Figure 1** Proposed clinical development program.

action, the clinical indication to be supported is chronic administration in a non–life-threatening autoimmune and/or inflammatory disease. The plan calls for sequential conduct of phase 1a, 1b, 2a, 2b, and 3 clinical trials that enroll normal human volunteers (NVH) in phase 1 and patients in phases 2 and 3.

Dosing progresses from single ascending doses (SAD), to multiple ascending doses (MAD), to multiple doses (MD), as the duration of dosing increases from a single administration to dosing over 2-week, 1-month, 3-month, and ≥6-month periods. The duration of each clinical trial (from first patient in to clinical report) increases successively from 6 months to 3 years, with the duration of the entire clinical program, from IND through BLA filing, being 8 years. Milestones are incorporated to allow end of phase 1 (EoP1), end of phase 2 (EoP2), and the Carcinogenicity Assessment Committee (CAC) meetings. This clinical development program is idealized, as minimal time is allowed for the transition from completion of one phase to initiation of the next phase. As such, this clinical plan will put considerable pressure on the preclinical program to deliver the required support in advance of the need for it.

### 4.4 Specific Assumptions

The specific assumptions about the drug were as follows:

- The doses to be tested in humans in various studies would be 1, 3, 10, 30, and 100 mg/kg, administered orally once daily. The top viable dose would be 100 mg/kg.
- The relevant species were monkeys (in which the drug was pharmacologically active and metabolism was comparable to humans), rabbits, and rats (in which there was no pharmacologic activity but comparable metabolism).
- The route of administration would be oral, with daily dosing at doses of 3, 10, 30, 100, 300, and 1000 mg/kg utilized in various studies.
- The no observable adverse effect (dose) level (NOAEL) would be 100 mg/kg, with a maximum tolerated dose (MTD) of 1000 mg/kg.
- The cost of manufacturing test article for preclinical use would range from $10,000 to $70,000 per kilogram, with a cost of $25,000 per kilogram at the commercial scale.

The specific assumptions for the biologic, a monoclonal antibody (mAb), were as follows:

- The doses to be tested in humans in various studies would be 0.1, 0.3, 1, 3, and 10 mg/kg, administered intravenously once monthly. The top viable dose would be 10 mg/kg.
- The relevant species was the monkey, in which the biologic was pharmacologically active, with comparable affinity and activity to human test

systems in vitro. The mAb was not active in any other species, including mice, rats, guinea pigs, rabbits, dogs, and pigs.

- The route of administration would be intravenous, with weekly dosing at doses of 10, 30, and 100 mg/kg utilized in various studies.
- The NOAEL would be 100 mg/kg, without an MTD being established.
- The mAb would be immunogenic in monkeys, with approximately 50% of the animals developing anti-mAb antibodies, and approximately 25% of the animals developing neutralizing and/or clearing antibodies.
- The cost of manufacturing test article for preclinical use would range from $3000 to $30,000 per gram, with a cost of $1000 per gram at the commercial scale.

Obviously some of these specific assumptions could be questioned, and the associated costs might be very different for different products or within different companies. At the time, however, these were reasonable assumptions for the two products being compared, and these specific assumptions allowed calculation of the amount, time, and cost associated with manufacture of the test articles.

### 4.5 "Traditional" Preclinical Development Plans to Support Clinical Development by Phase

The first comparison to be made evaluates "traditional" preclinical development plans to support each phase of clinical development. While the term "traditional" might be thought to apply only to drugs, it is also true that certain preclinical programs for biologics have been established as well as accepted, and therefore could be regarded as traditional. This initial comparison assumes direct, uninterrupted progression of the product along a relatively simple development pathway. For the biologic this means that a relevant species has been identified and there are no planned changes to the cell line, manufacturing process, formulation, or route of administration during development. The need for "nontraditional" or "alternative" development pathways to address these issues will be discussed later.

***Essential Components Required for Starting Preclinical Development*** Prior to the design and initiation of a preclinical development program, there are a number of essential components that should be in place, as outlined in Table 3. These components include important information on the manufacture and characterization of the product, as well as the reagents and assays necessary for evaluation of the PK and PD properties and, for biologics, the immunogenicity and tissue cross-reactivity. In addition, most preclinical development programs would not be started without identification of the relevant species and some preliminary non–GLP-compliant studies to establish some general parameters, such as those associated with "discovery toxicology." These preliminary studies, generally conducted to facilitate nomination of a

**TABLE 3 Essential components required for starting preclinical development**

| | Criteria | Drug | Biologic (mAb) |
|---|---|---|---|
| Product information | Material selected | Compound and salt form | Initial cell line, growth conditions |
| | Production process known | Yes | Yes |
| | Initial material characterization | Impurities, solvents | LPS, viral, host cell proteins |
| | Pilot scale material produced in-house | ≤1 kg | ≤20 g |
| | Other criteria | NA | Potency (activity) assay |
| Assays and reagents | Dose retain analysis | HPLC | ELISA |
| | PK | LC/MS/MS | ELISA |
| | PD (receptor-ligand binding inhibition) | FACS | FACS |
| | Immunogenicity (PAHA) | NA | ELISA |
| | Tissue cross-reactivity | NA | IHC |
| Preclinical studies | Relevant species selected | Based on shared in vitro metabolite profile with humans. Rodent and nonrodent species identified. | Based on shared pharmacologic activity. Identification of rodent species is generally difficult, if not impossible. Identification of nonrodent specie(s) is occasionally difficult. |
| | Genetic toxicology screened | Mini-Ames<br>In vitro micronucleus | NA |
| | Discovery toxicology has set doses | Non-GLP 14-day rodent, non-GLP 14-day nonrodent | Non-GLP single-dose PK/PD/toxicity in relevant species |
| | Other criteria | Non-GLP receptor screening | Non-GLP tissue cross-reactivity |

*Note*: ELISA = enzyme-linked immunosorbent assay, FACS = fluorescence-activated cell sorting, HPLC = high-performance liquid chromatography, IHC = immunohistochemistry, LC/MS = liquid chromatography/MS = mass spectrometry, LPS = lipopolysaccharide (endotoxin), NA = not applicable, PAHA = (nonhuman) primate anti-human antibodies.

candidate for development, are intended to ensure that preclinical development has the appropriate preliminary information and/or tools to assess safety and that there are few surprises during development.

***Preclinical Support to Phase 1*** The studies, time, materials, and costs associated with preclinical support to phase 1 are summarized in Table 4 and illustrated in Figure 2. For the drug, the studies necessary to support phase 1 would typically consist of in vitro and in vivo genetic toxicity studies, dedicated safety pharmacology studies evaluating cardiovascular, central nervous system (CNS) and respiratory function,[3] and general toxicology studies of 1-month (28-day) dosing duration (with a 2-week recovery period) in a rodent and nonrodent species.[4] For this particular drug, a receptor antagonist, a study would also be conducted to assess species cross-reactivity in vitro by screening receptor binding on target cells from different species.[5]

Except for the general toxicology studies, phase 1-supporting studies are independent in nature and can be conduct in parallel. The general toxicology studies ideally would not be initiated until after the PK and PD assays are validated, by which time there may be some useful toxicity information from the safety pharmacology studies. This IND-enabling preclinical program for a drug requires approximately 10 months and costs around $1.2 M, including the roughly $40 K cost and 2 months required for manufacturing approximately 1.3 kg of non–GMP-compliant material. For a drug, most sponsors would elect to use GMP-compliant material, with the attendant delay in the ability to start.

For the mAb, the studies necessary to support phase 1 include tissue cross-reactivity and general toxicology studies; in accordance with ICH S6, dedicated safety pharmacology and genetic toxicology studies are not required. The general toxicology studies can be conducted in a single, pharmacologically relevant species and include both a single-dose[6] and 1-month repeated-dose study, with longer recovery periods (one and three months, respectively) to allow for the slower clearance of the mAb. Relative to those for a drug, these studies can be much more complicated and expensive, as they would routinely incorporate assay for immunogenicity (in addition to PK and PD) and assessment of safety pharmacology (cardiovascular, respiratory and CNS)

[3]Prior to beginning development, many sponsors would also have conducted in vitro safety pharmacology studies (e.g., hERG inhibition) to assess cardiovascular concerns, as well as single-dose and 5- to 7-day repeated-dose non–GLP-compliant studies to assess toxicity and help set doses for the GLP-compliant studies (see Table 3).

[4]A 14-day dosing duration would also be acceptable to support phase 1 single-dose to 2-week studies, although many sponsors are electing to conduct 28-day studies to support up to 1-month trials.

[5]Some sponsors would also screen related targets, if available, for recognition by the drug, which might be associated with secondary pharmacologic activity.

[6]A single-dose study is generally useful to fully characterize the clearance of a compound, such as this mAb, with a long half-life.

**TABLE 4 Preclinical support to phase 1**

| | | Drug | | | | Biologic | | |
|---|---|---|---|---|---|---|---|---|
| Phase 1 | | Time (Mo) | Material (g) | Cost ($K) | | Time (Mo) | Material (g) | Cost ($K) |
| Assay validation | PK (LC/MS/MS) | 4 | 1 | 15 | PK (ELISA) | 10 | 1 | 30 |
| | PD (LBI; FACS) | 6 | | 50 | PD (LBI; FACS) | 6 | | 50 |
| | | | | | Immunogenicity (ELISA) | 10 | | 30 |
| | | | | | Tissue cross-reactivity (IHC) | 2 | | 20 |
| Cross-reactivity | Receptor screening | 2 | 1 | 15 | Tissue cross-reactivity | 2 | 1 | 80 |
| Genetic toxicology | Mutagenicity (Ames) | 2.5 | 2 | 6 | | | | |
| | Chromosomal aberration in vitro (CHO) | 2.5 | 2 | 22 | | | | |
| | Chromosomal aberration in vivo (RMN) | 2.5 | 2 | 30 | | | | |
| Safety pharmacology | Cardiovascular, monkey | 2.5 | 25 | 150 | | | | |
| | Respiratory, rat | 2.5 | 5 | 38 | | | | |
| | CNS, rat | 2.5 | 5 | 20 | | | | |
| General toxicology | 1 mo (+2 wk rec) rat | 4 | 250 | 200 | Single-dose (+1 mo rec) monkey | 4 | 3 | 150 |
| | 1 mo (+2 wk rec) monkey | 4 | 1000 | 600 | 1 mo (+2 mo rec) monkey | 7 | 30 | 600 |
| Studies | | 10 | — | 1146 | | 17 | — | 960 |
| Material | | 2 | **1293** | 40 | | 6 | **35** | 1600 |
| Total | | **10** | — | **1186** | | **17** | — | **2560** |

*Note*: CHO = Chinese hamster ovary cells, ELISA = enzyme-linked immunosorbent assay, FACS = fluorescent-activated cell sorting or flow cytometry; IHC = immunohistochemistry, LBI = ligand-binding inhibition, rec = recovery, RMN = rat micronucleus.

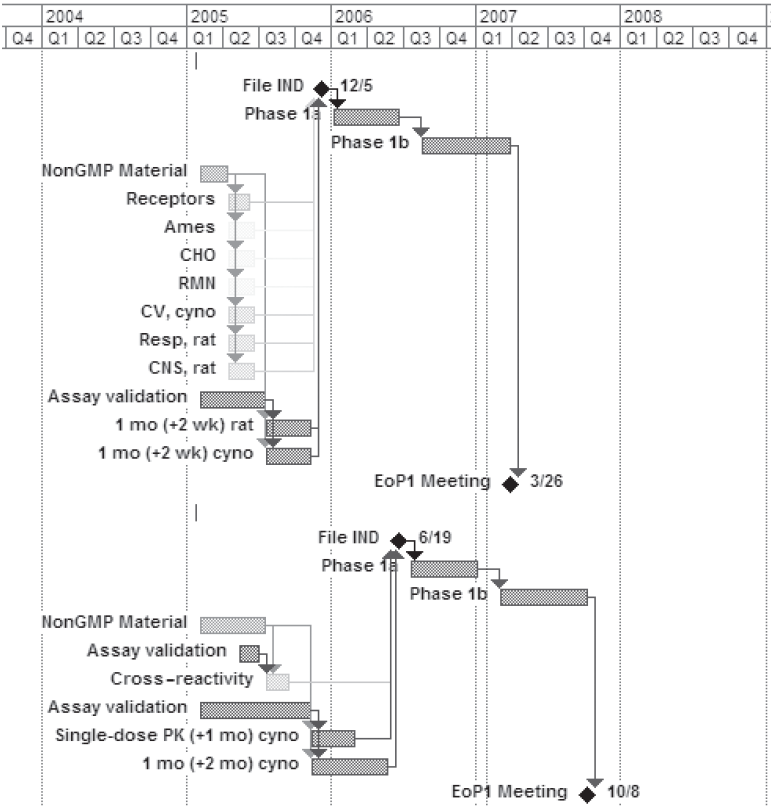


Preclinical programs to support a drug (top) or biologic (bottom). For comparison purposes, the starting date was assumed to be 18 January 2005.

**Figure 2** Preclinical support to phase 1.

parameters. This preclinical IND-enabling program for a mAb requires approximately 17 months and costs $2.6 M, including the $1.6 M cost and 6 months required to manufacture approximately 35 g of non–GMP-compliant material. Thus, relative to the program for the drug, the program for the mAb requires fewer studies and less material, but will take approximately 7 months longer and require $1.4 M more in total costs. Both the added cost and time are primarily related to the manufacture of material, although the longer time required for validation of the biological assays and clearance of the test article in the recovery periods also contribute to this.

Both of these IND-enabling preclinical programs could reasonably be expected to be acceptable to regulatory agencies, so there may be little need to try to obtain a preliminary agreement as to their acceptability from regulatory reviewers. That is, it would not be essential to arrange a pre-IND meeting (to present preliminary results) or even an earlier (pre-pre-IND) meeting (to

discuss planned studies) to assess whether the appropriate studies were done to support the phase 1 trial.[7]

***Preclinical Support to Phase 2*** The studies, time, materials, and costs associated with preclinical support to phase 2 are summarized in Table 5 and illustrated in Figure 3. For the drug, the studies necessary to support phase 2 would consist of rodent and nonrodent 3-month repeated-dose general toxicology studies and reproductive and developmental toxicology studies. The latter would include embryo/fetal (segment 2) studies in pregnant female rats and rabbits, and fertility (segment 1) studies in males and female rats. Range-finding (pilot) studies to establish maternal toxicity prior to segment 2 studies would be necessary as well. The use of rabbits and rats, in which pharmacologic activity is absent, for developmental toxicology can be justified on the basis of metabolic profiles comparable to humans in vitro. Such a decision is not unusual for drugs, where toxicology unrelated to the mechanism of action is a primary concern. A significant advantage to the use of rats and rabbits is the short gestation periods and less expensive animal costs. This preclinical program to support phase 2 for the drug takes approximately 9 months and costs $1.8 M, including the nearly $125 K cost and 2 months required for manufacturing approximately 5 kg of GMP-compliant material.

For the mAb, the studies necessary to support phase 2 would consist of a monkey 3-month repeated-dose general toxicology study (with a 2-month recovery period), segment 2 studies in pregnant female monkeys and segment 1 studies in male and female monkeys.[8] Although the duration of dosing in the latter studies is relatively short (1 to 3 months), these studies can be very difficult and time-consuming to conduct as they typically last at least 12 months. This time is required to identify females in estrus and to breed and enroll pregnant female monkeys into the studies. This preclinical program to support phase 2 for the mAb requires approximately 17 months and costs $4.3 M, including the $1.2 M cost and 5 months required for manufacturing approximately 400 g of GMP-compliant material. Thus, the program for the mAb is approximately 8 months longer, and $2.5 M more expensive than the program for the drug, primarily due to the use of monkeys (with long gestation periods) for reproductive and developmental toxicology.

As was the case for support for phase 1, these development programs to support phase 2 follow ICH guidelines and are therefore not controversial. They are unlikely to require advanced indication of their acceptability to regulatory agencies.

[7]This discussion understandably cannot address whether the *results* of the preclinical studies that were conducted would support the proposed clinical trials (e.g., the proposed first-in-human dose). For that reason, or for other reasons, a development project team might choose to request a pre-IND meeting or an earlier meeting.

[8]It should be noted that there are very few CROs that offer reproductive toxicity testing in nonhuman primates. To date, most of the reproductive toxicology studies in monkeys have been segment 2 (teratology) studies. Segment 1 (fertility) studies would be difficult to conduct.

**TABLE 5 Preclinical support to phase 2**

| | | Drug | | | | Biologic | | |
|---|---|---|---|---|---|---|---|---|
| Phase 2 | | Time (Mo) | Material (g) | Cost ($K) | | Time (Mo) | Material (g) | Cost ($K) |
| General toxicology | 3 mo (+1 mo rec) rat | 7 | 700 | 253 | | | | |
| | 3 mo (+1 mo rec) monkey | 7 | 3000 | 650 | 3 mo (+2 mo rec) monkey | 9 | 100 | 650 |
| Reproductive and developmental toxicology | Segment 2 RF, rat | 2.5 | 40 | 43 | | | | |
| | Segment 2 RF, rabbit | 2.5 | 500 | 57 | | | | |
| | Segment 2, rat | 3.5 | 100 | 128 | | | | |
| | Segment 2, rabbit | 3.5 | 500 | 185 | Segment 2, monkey (1 mo) | 12 | 50 | 950 |
| | Segment 1, rat, male | 6 | 50 | 193 | Segment 1, monkey, male (3 mo) | 12 | 120 | 710 |
| | Segment 1, rat, female | 6 | 50 | 119 | Segment 1, monkey, female (3 mo) | 12 | 120 | 742 |
| Studies | | 7 | — | 1628 | | 12 | — | 3052 |
| Material | | 2 | **4940** | 125 | | 5 | **390** | 1200 |
| Total | | **9** | — | **1753** | | **17** | — | **4252** |

*Note*: Rec = recovery, RF = range-finder.

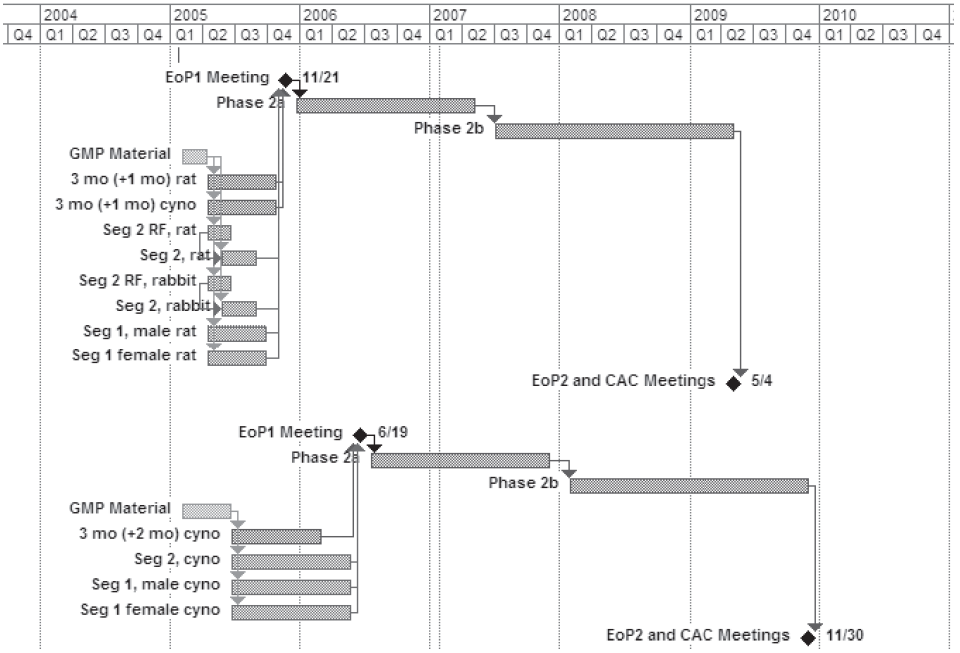


Preclinical programs to support a drug (top) or biologic (bottom). For comparison purposes, starting date was assumed to be 18 January 2005.

**Figure 3** Preclinical support to phase 2.

***Preclinical Support to Phase 3 and Registration*** The studies, time, materials, and costs associated with preclinical support to phase 3 and registration are summarized in Table 6 and illustrated in Figure 4.

For the drug, the studies necessary to support phase 3 include chronic toxicology in rats and monkeys (6 and 12 months, respectively) and perinatal/postnatal (segment 3) toxicology studies in pregnant female rats. These studies take approximately 22.5 months and $1.6 M, including the 2.5 months and $125 K to manufacture 5 kg of material. To support registration of the drug, carcinogenicity studies must be conducted in two rodent species. This example uses a 2-year bioassay in rats and a 6-month study in transgenic mice. The cost to support registration is approximately 38 months and $2.9 M, including 2 months to manufacture 3 kg of material for $75 K.

For the mAb, the studies necessary to support phase 3 include a 9-month chronic toxicology study[9] and a segment 3 reproductive toxicology

[9]Although ICH S6 indicates that chronic toxicology studies of 6 months duration may be sufficient, many Sponsors have been recently requested to conduct nine-month studies.

**TABLE 6 Preclinical support to phase 3 and registration**

| | | Drug | | | | Biologic | | |
|---|---|---|---|---|---|---|---|---|
| Phase 3 | | Time (Mo) | Material (g) | Cost ($K) | | Time (Mo) | Material (g) | Cost ($K) |
| Reproductive and developmental toxicology | Segment 3, rat | 9 | 250 | 350 | Segment 3, monkey (3 mo rec + 3 mo) | 18 | 120 | 1500 |
| Chronic toxicology | 6 mo (+2 mo rec) rat | 11 | 1250 | 355 | | | | |
| | 12 mo (+2 mo rec) monkey | 20 | 3600 | 770 | 9 mo (+3 mo rec) monkey | 16 | 400 | 750 |
| Studies | | 20 | — | 1475 | | 18 | — | 2250 |
| material | | 2.5 | **5100** | 125 | | 5 | **520** | 1800 |
| Total | | 22.5 | — | **1600** | | **23** | — | **4050** |
| Registration | | | | | | | | |
| Carcinogenicity | 2 yr rat | 36 | 3000 | 1700 | | | | |
| | 5 d/1 mo RF, Tg mouse | 4 | 10 | 225 | | | | |
| | 6 mo Tg mouse | 18 | 15 | 900 | | | | |
| Studies | | 36 | — | 2825 | | 18 | — | 2250 |
| Material | | 2 | **3025** | 75 | | 5 | **520** | 1800 |
| Total | | 38 | | **2900** | | **23** | — | **4050** |

*Note*: Rec = recovery, RF = range-finder, Tg = transgenic.

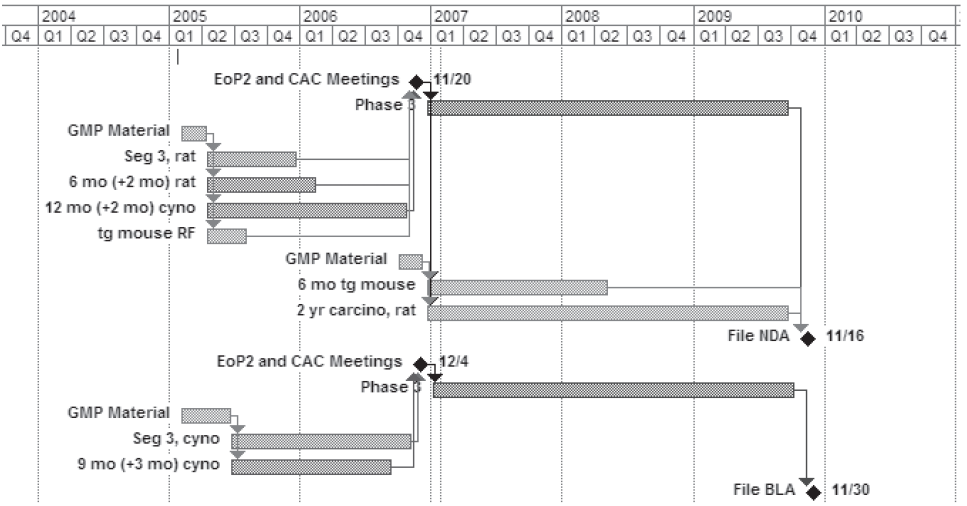


Preclinical programs to support a drug (top) or biologic (bottom). For comparison purposes, starting date was assumed to be 18 January 2005.

**Figure 4** Preclinical support to phase 3 and registration.

study[10] in monkeys. These studies take approximately 23 months and $4.1 M, including 5 months and $1.8 M to manufacture approximately 500 g of material. In accordance with ICH S6, carcinogenicity studies would not be conducted for a biologic if no relevant and appropriate species can be identified. While monkeys are a relevant species for this mAb, they are not an appropriate species for carcinogenicity testing. Therefore, after chronic toxicity testing, no additional studies are necessary to support registration. Once again, relative to the drug, the cost of this preclinical support is greater for the mAb, primarily due to the cost of its manufacture. However, the time needed is about 15 months shorter due to the lack of carcinogenicity testing for this biologic.

Again, these preclinical safety evaluation programs to support Phase 3 are designed to follow ICH guidelines and are relatively likely to be acceptable to regulatory agencies.

***The Integrated Preclinical Development Plans*** The total time, materials and costs associated with preclinical support for phase 1 through registration are summarized in Table 7 and the integrated preclinical support to clinical development is illustrated in Figure 5.

[10]It should be noted that conducting true segment 3 studies in monkeys are not feasible, as this would require several years to follow the first filial (F1) generation to reproductive maturity. However, some sponsors developing immunomodulatory biologics have been evaluating the immune system function in F1 monkeys to address the most relevant portion of segment 3 concerns for their product. To do this, some sponsors elect to combine immune function assessment of the offspring from a segment 2 study when using monkeys for reproductive toxicology testing.

**TABLE 7 Total time, materials, and costs for preclinical support for phase 1 through registration**

| | Drug | | | Biologic | | |
|---|---|---|---|---|---|---|
| | Time (Mo) | Material (kg) | Cost ($M) | Time (Mo) | Material (kg) | Cost ($M) |
| Phase 1 | 10 | 1.293 | 1.186 | 17 | 0.035 | 2.560 |
| Phase 2 | 9 | 4.940 | 1.700 | 17 | 0.390 | 4.252 |
| Phase 3 | 22.5 | 5.100 | 1.600 | 23 | 0.520 | 4.050 |
| Registration | 38 | 3.025 | 2.900 | — | — | — |
| Material | 8.5 | 15 | 0.4 | 16 | 1 | 4.6 |
| Studies | 73 | — | 7.0 | 47 | — | 6.3 |
| Total | 105 | ~15 | 7.4 | 112 | ~1 | 10.9 |

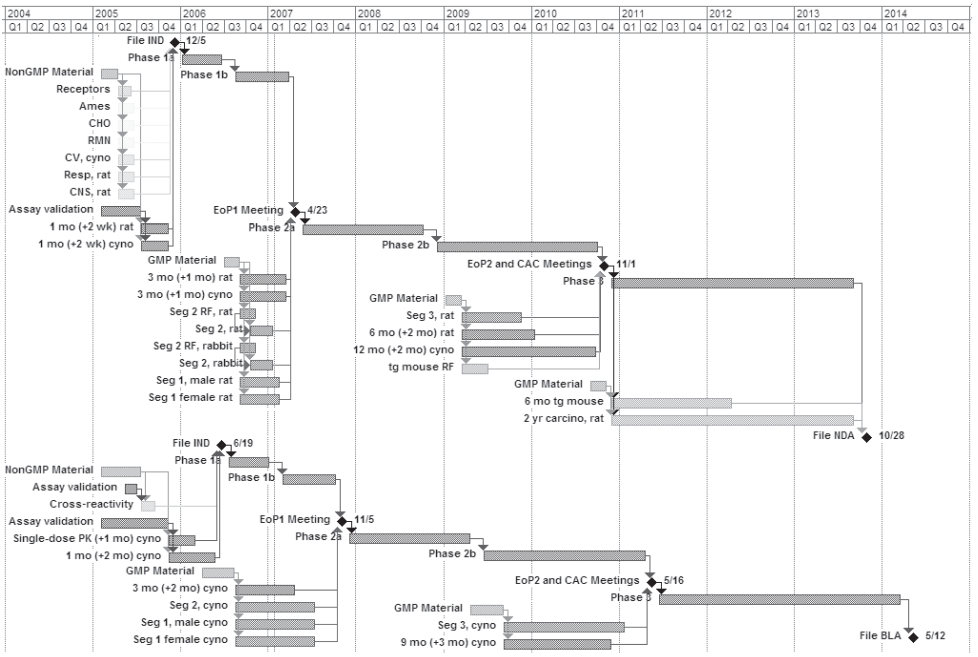


Preclinical programs to support a drug (top) or biologic (bottom). For comparison purposes, starting date was assumed to be 18 January 2005.

**Figure 5** Integrated preclinical support to clinical development.

For the drug, the studies conducted require 73 months in total, at a cost of around $7.0 M. The amount of material required is 15 kg, manufactured over 8.5 months at a cost of $0.4 M. Of the $7.4 M total cost, manufacturing represents only 5.4%. In contrast, the studies for the mAb required only 47 months and $6.3 M. Only around 1 kg of material is required for the mAb, with about 16 months required to manufacture it, at a cost of around $4.6 M. Manufacturing represents 42% of the $10.9 M total costs. Thus, relative to the drug, material supply for preclinical development of the mAb takes twice as long and is 10 times as expensive, even though only 1/15th as much material is required. The studies for the mAb can be completed in almost half the time, but cost approximately the same. Overall, this comparison demonstrates that it is more expensive to provide comparable preclinical support to clinical development for the mAb than the drug.

If both programs were started at the same time and progressed as scheduled, it would take 105 months (8 years, 9 months) until filing the NDA for the drug and 112 months (9 years, 4 months) until filing the BLA for the mAb. The additional 7 months required to bring the mAb to registration can be traced directly to the additional 7 months required to enable phase 1, which was related primarily to a longer time to manufacture and the longer recovery period necessary to observe clearance of the mAb in the one-month repeated-dose study. This illustration reveals an important difference between drugs and biologics: the rate-limiting portion of preclinical development for a drug is the time required to conduct the toxicology studies, while for a biologic it is the time required to manufacture the material.

***Critical Decisions to Be Made*** The preceding examples have demonstrated that the time required for manufacturing the appropriate material for preclinical testing is an important contributor to the overall time line for preclinical development; more so for biologics. When planning these programs, there are critical decisions to be made related to reserving the necessary manufacturing capacity. Whether manufacturing is to be done internally or outsourced to a contract manufacturing organization (CMO), the lead time necessary to secure manufacturing time via contract is generally 3 to 6 months for a drug and 6 to 12 months for a biologic.[11]

The differences in lead time are related to the inherent differences in the complexity of manufacturing and to the relative number of CMOs or internal capacity available for each type of product. It is generally easier, quicker, and less expensive to manufacture drugs, and easier to identify and reserve available manufacturing capacity for drugs. Similar lead times for advanced reservation of CRO capacity are often needed for certain types of preclinical studies, such as those using monkeys or requiring complicated technical procedures.

[11]For a drug, the lead time necessary for reserving manufacturing capacity remains 3 to 6 months throughout development, while over time it declines from 6 to 12 months to 6 to 9 months for a biologic.

A careful review of Figure 5 reveals that, because of the 3- to 6-month or 6- to 9-month lead time required for a drug or biologic, respectively, these critical decisions must generally be made prior to knowing whether the contractual commitment is justified. They therefore involve an element of risk. For example, a development project team must commit to manufacturing a drug or biologic for preclinical support to phase 1 as much as 5 or 18 months, respectively, ahead of the time when the material is actually needed for dosing. In many companies this would mean that a commitment to manufacture development material to facilitate preclinical support to phase 1 must be made before the product is officially accepted into development. This requirement for making critical decisions in advance of supportive results recurs in each phase of clinical development.

Some sponsors elect to pursue a minimal risk development strategy, with success of each phase of clinical development being demonstrated before preclinical or manufacturing support for the next phase is initiated. Sometimes such a strategy is dictated by financial constraints. This stage-gated approach imposes significant delays to the clinical development time line and ensures that the time until registration will be prolonged. The only way to prevent delays in clinical development that might arise from the preclinical program would be to manufacture the preclinical material and conduct the preclinical studies at risk. That is, complete all preclinical support to the next phase prior to knowing whether the current phase of clinical development will be successful, thereby rewarding the risk taken.

It is important to recognize that the magnitude of the resources, time, and money at stake for the manufacture of biologics is greater than for drugs, and that the manufacturing capacity is more limited. Sponsors may elect to accept certain risks to decrease the likelihood that there will be later delays. For example, sufficient quantities of a drug could be manufactured to enable multiple phases of preclinical and clinical development with minimal financial risk because manufacturing a drug is relatively inexpensive and quick. Furthermore, because of the inherent physicochemical properties of drugs, which lend themselves to early definition of the manufacturing specifications, there is also minimal risk that the material will not be representative of ("noncomparable" to) the "final" material, which is used in phase 3 trials. Relatively simple bioequivalence (comparative PK/PD) studies can be conducted in animals or humans to document this. This strategy is generally not viable for a biologic.

The most risk-filled phase of preclinical development, especially for a biologic, is that associated with the support to phase 3, which should be completed prior to the end of phase 2. Support to phase 3 is the most critical portion of preclinical development, as these are the pivotal studies that, with the carcinogenicity studies run concurrently with phase 3, will support registration. It is therefore essential that the material used to support phase 3 is representative of final product intended for use in humans.

When possible, GMP-compliant material from the clinical trial lots might be used, but waiting for this material may delay the start of the chronic toxicology studies. A drug manufactured by the final process and/or representative of the

final process material is generally available for phase 3 support. A biologic manufactured by the final process (final cell line, scale, process, etc.) is seldom available when it is time to initiate the studies required for phase 3 support. Thus for a biologic, simple bioequivalence studies may be replaced by more complicated comparability studies that assess PK/PD parameters and immunogenicity at a minimum, and perhaps safety and efficacy if warranted by observed differences (see Chapter Demonstration of Comparability of a Licensed Product after a Manufacturing Change). The point is that each project team must consider many different critical decisions in designing the integrated development plans. Each of these decisions will carry with it an element of risk with regard to financial, scientific or regulatory aspects and to the development time line.

### 4.6 "Alternative" or Novel Preclinical Development Plans for Biologics

So far this exercise has focused on a "traditional" and uncomplicated development program for a biologic. To facilitate this, some assumptions were made, such as that a relevant species was identified and there were no planned changes to the cell line, manufacturing process, formulation, or route of administration during development. These assumptions, however, are not typical for most biologics. In reality the nature of many biologics is that they often require "nontraditional" or "alternative" development pathways. This is the essence of the "case-by-case" or science-driven program design espoused by ICH S6.

For example, it may be difficult to identify a relevant species, such as when a mAb does not cross-react with any species other than humans or cross-reactivity is limited to chimpanzees, a protected species that is difficult to access and has significant limitations in their use. Or the biologic may recognize the target in species other humans, but the target is expressed in different tissues or on different cells or has different functions than in humans. In these cases one may choose to develop a homologous (or surrogate) biologic for which a pharmacologically relevant (generally rodent) species can be identified.

Biologics are also different than drugs in that the product characteristics may differ as changes are made to the cell line or manufacturing process. These changes could affect PK, PD, immunogenicity, efficacy, and/or safety and typically require comparability studies to assess the potential differences. Likewise it is not uncommon for biologics to be tested initially in their simplest forms, such as an aqueous solution administered intravenously (IV), to attempt to get an indication of the viability of the product for the intended clinical indication. Later in development it might be prefered that the biologic be formulated for delayed release after subcutaneous (SC) administration. These changes could also affect the in vivo properties of the product and typically require "bridging" studies to assess potential differences.

Thus the second set of comparisons to be made evaluates "alternative" preclinical development plans to support the lack of a relevant species and planned changes to the cell line, manufacturing process, formulation, or route of administration during development.

***Supporting Development without a Relevant Species*** The simplest definition of a relevant species might be "one that is biologically comparable to humans and in which the test article has comparable pharmacologic and toxicologic properties." It is, however, sometimes much easier to describe a relevant species than to identify one. For the GPCRs that were the targets in this example, demonstrating biological comparability might involve comparison of the amino acid sequences, cellular and tissue expression in normal and disease states, and in vitro biological functions for humans and several other species, with the intent being to identify the most comparable species (see Chapters Selection of Relevant Species, Tissue Cross-Reactivity Studies for Monoclonal Antibodies: Predictive Value and Use for Selection of Relevant Animal Species for Toxicity Testingn, Physiologic IgG Biodistribution, Transport, and Clearance: Implications for Monoclonal Antibody Products, The Role of Pharmacokinetics and Pharmacodynamics in Selecting a Relevant Species and Use of Animal Models of Disease in the Preclinical Safety Evaluation of Biopharmaceuticals).

A robust preclinical safety program may devote as much or more effort to demonstrating biological comparability of the chosen species as to evaluating the toxicity of the product in that species. Prior to undertaking this effort, most project teams would evaluate the extent to which the product recognized or was active against the target in candidate species. Simple receptor binding affinity or ligand binding inhibition studies are often used for this purpose (see Chapter Selection of Relevant Species). If an acceptable level of cross-recognition or "cross-reactivity" is present in the tested species, then additional efforts would be justified to determine the relative activity and potency of the product in that species in vitro, or to evaluate pharmacologic activity of the product in vivo, such as in an animal model of disease (see Chapter Use of Animal Models of Disease in the Preclinical Safety Evaluation of Biopharmaceuticals). Despite these efforts to demonstrate comparable biology and pharmacology as the key components for identification of a relevant species for safety testing, it will only be after testing in humans that the degree of relevance of the selected species will become known.

Perhaps all biologics would be pharmacologically active in chimpanzees, our closest nonhuman primate relative. Most biologics would also be active in one or more monkey species, particularly of the *Macaca* genus (e.g., *M. mulatta* or *M. fascicularis*, the rhesus and cynomolgus monkeys, respectively). However, some biologics, particularly mAbs, have such restricted, species-specific cross-reactivity that they recognize the target only in humans and chimpanzees and thus are active only in these species. Although chimpanzees have been used for preclinical testing programs, their protected status, relatively limited numbers, lack of suitability for invasive measurements (e.g., histopathology) or chronic, reproductive or developmental toxicology studies, and past use for testing a variety of biologic agents (which can be associated with interfering immunogenic responses) makes chimpanzees not practical for most programs. Furthermore in some countries there is a ban on the use of chimpanzees to support clinical development. For these reasons, some biotechnology companies will no longer nominate as a development candidate any biologic that has cross-reactivity restricted to chimpanzees only.

While most biologics companies will first evaluate cynomolgus monkeys as a potential relevant species (because of their relative availability), that is not to say that monkeys are always the most relevant species or that other species are not relevant. The degree of relevance of the selected species must be demonstrated for each program.

When a relevant species cannot be identified for preclinical testing of a biologic, ICH S6 supports the use of (1) transgenic animals expressing the human target (in which the biologic might be tested, provided that the target was appropriately functional), (2) knockout or transgenic animals to evaluate the potential effects of target inhibition or replacement therapy, respectively, or (3) the use of "homologous proteins," which are often referred to as "surrogate" biologics. Any of these choices requires a commitment to thoroughly characterize the surrogate product and (surrogate) testing system. That is, the choice of testing system (species, strain, genotype, etc.) must be demonstrated to be relevant. For transgenic or knockout animals, this might mean genotyping individual animals, as well as fully describing the resulting phenotype. For a surrogate product, this might mean characterizing the pharmacologic activity of the surrogate in the test species, as well as the biological function of the target in this species. For example, for a non–cross-reactive mAb directed against a human target, a relevant surrogate mAb might be one directed against the same target in a rodent species, provided that the target functions comparably in the rodent species. Any of these alternative choices, if intended to support preclinical safety testing, must also comply with appropriate regulatory guidelines. For example, the material used might best be manufactured in compliance with GMP guidelines and the studies might best be GLP compliant. In these surrogate studies all of the usual endpoints should be assessed, such as PK, PD, and immunogenicity. In essence this means that the decision to adopt a surrogate program to address safety issues in a pharmacologically relevant species is a decision to double the amount of preclinical (and CMC) work to be done. However, adoption of a surrogate may allow testing of all aspects of toxicology, potentially even through carcinogenicity testing. Some would argue that, because these alternative methods do not directly evaluate the development product, some degree of relevance is lost. The counterargument to this point of view might be that if the biologic is not active in an animal species, there is no choice but to develop an alternative preclinical development program.

Figure 6 illustrated the hypothetical studies and time lines associated with preclinical support for a mAb that has cross-reactivity (pharmacologic activity) restricted to humans and chimpanzees. In this example, the strategy to support clinical development consists of conducting single-dose and 1- and 3-month repeated-dose studies in chimpanzees with the development mAb and use of a surrogate rat–anti-mouse mAb in a complete toxicology program in mice. Note that the portions of the program conducted in chimpanzees or mice would be essentially identical to what would be done in monkeys or rats if either were a relevant species. Once adopted, the surrogate mAb program

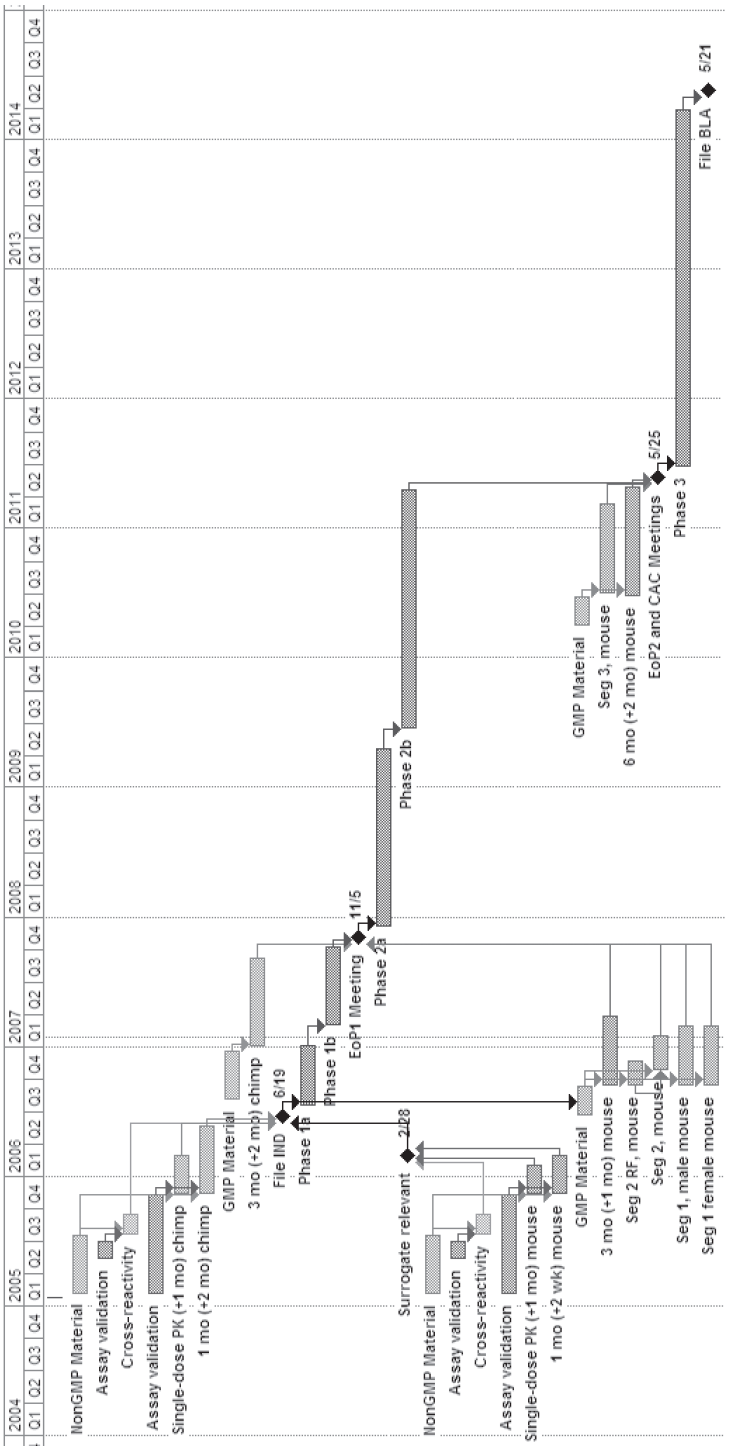


For comparison purposes, starting date was assumed to be 18 January 2005.

**Figure 6** Supporting development without a relevant species.

cannot be abbreviated; to the extent possible, appropriate studies must be conducted to demonstrate that the surrogate is relevant. In this case the results of the chimpanzee studies with the development mAb and the mouse studies with the surrogate mAb could be compared.

As an aside, many biologics companies would choose one or the other of these alternative strategies, not both. In some cases it might not be logistically or technically feasible to conduct chimp studies. In some cases it might not be technically possible or pharmacologically appropriate to use a surrogate. In no case does ICH S6 recommend that a surrogate be developed for the purpose of allowing testing in a second (rodent) species if a single relevant species is identified.

This example of an integrated development plan assumes that the surrogate mAb cell line was already established, so that no time was required for establishing it. In this case the time until BLA filing is essentially unchanged from the previous example in which monkeys were used, but the total cost of the program is about $8.1 M, which is $2.8 M less expensive. This savings is primarily related to the use of mice rather than monkeys. However, the lower cost does not include the internal costs associated with the concurrent (parallel) development of both the surrogate and development mAb cell lines. If the surrogate mAb cell line was not yet established at the time it was decided to pursue this strategy, an additional roughly 9 to 12 months would be necessary, with the BLA then filed about 1 year later.

As with a traditional approach there are numerous critical decisions to be made during development. One of the most important considerations is the extent to which the data generated by use of a surrogate is likely to be viewed as acceptable to regulatory authorities. From the project team's point of view, it is unwise to commit resources to the conduct of studies that would not be acceptable.[12] From a regulator's point of view, acceptability will be assessed both on the quality of the studies (appropriateness, design, and compliance) and their eventual results (demonstration of apparent relevance). While a given study may be understood to be irrelevant a priori (e.g., one that uses a species in which there is no pharmacologic activity), it is not true that all studies that appear to be relevant will prove to be so. That is, the true relevance of any study can only be known *after* human data are available for comparison. It becomes therefore important for an alternative preclinical program to attempt to establish the degree of *apparent* relevance of a study. But it is equally important to try to establish the degree of apparent regulatory *acceptability* of alternative development programs. That means that there are several times during development, generally prior to large resource commitments, when discussions should be held with regulatory authorities as to the acceptability of the proposed alternative development programs.

[12]In this context, acceptability refers to the scientific rationale and the study designs, rather than to the results, which can only be assessed for acceptability retrospectively.

If one reviews Figure 6, it is apparent that discussions with regulatory authorities might be held at two critical times. The first is around 12 months prior to beginning manufacture of the surrogate, when the commitment to reserve manufacturing capacity is needed, and when the acceptability of the alternative strategy should be discussed. This discussion might address whether GMP-compliant material, GLP-compliant studies and validated assays would be expected. Without an agreement as to the apparent acceptability of the proposed program, the program may be run at risk of failure (nonacceptability) from a regulatory point of view. The second discussion should be held several months prior to the time of IND submission, when the apparent relevance of the surrogate should be demonstrable based on the results of the studies conducted to date. At this time, if the cumulative data from the surrogate studies suggest that the surrogate approach is not relevant, then this alternative approach is not viable. For example, if our hypothetical mAb binds to a cell surface receptor and inhibits activation of human T cells in vitro and chimpanzee T cells in vivo, but the rodent surrogate mAb binds to mouse T cells and inhibits T cell activation in vitro, yet induces T cell depletion in vivo, the surrogate would not be relevant because it does not share the pharmacologic properties of the development mAb. Perhaps numerous efforts to re-engineer the surrogate mAb to a nondepleting mAb would be unsuccessful also. Or a nondepleting surrogate might be discovered, but the cell line might not reach necessary productivity levels, leading to failure for technical reasons. Thus, even with a preliminary agreement that an alternative approach might be relevant and acceptable, there is still considerable risk that the alternative approach will eventually not be successful in supporting the clinical development program.

***Supporting Changes to Cell Line, Manufacturing Process, Formulation, or Route of Administration*** Biologics generally evolve over the lifespan of product development, with changes along the way to cell line, manufacturing process, formulation, and/or the method or route of administration. Many programs in fact undergo changes in each of these categories. Because each of these changes has the potential to affect one or more properties of a biologic, which may in turn affect the PK, PD, immunogenicity, safety, or efficacy, it has been said for biologics that "the process defines the product." Therefore the preclinical development program must also support all significant changes made to the product or its manner of use.

For the mAb in this comparison, the initial product is from an early cell line that is formulated as a liquid, stored frozen, and administered by IV injection. This is sufficient for initial safety, PK/PD, and pharmacologic activity assessments in phase 1 but is not an acceptable product profile for commercialization. To be competitive in the intended clinical indication, the product needs to be stored as a lyophilized powder (to increase shelf life and concentration), to be reconstituted and injected SC as needed. The acceptable volume for SC injection dictates a much more concentrated formulation. Commercialization

also dictates a highly productive cell line is in place prior to phase 3. Each of the significant evolutionary changes to be made during development may require some degree of "bridging" or comparability testing, to demonstrate that the most recent version of the product maintains the qualities of the previous version (see Chapter Demonstration of Comparability of a Licensed Product after a Manufacturing Change). The challenge to the project team will be to decide when it is most efficient to test the effects of which changes.

Figure 7 illustrates the hypothetical studies and time lines associated with preclinical support for a mAb that undergoes changes to the route of administration (IV to SC) prior to phase 2, formulation (frozen liquid to lyophilized powder) midway through phase 2, and cell line (high-producing new clone) prior to phase 3. The initial studies included IV single-dose and 1-month repeated-dose studies in monkeys. To support the planned change from IV to SC administration for phase 2, one might conduct: a single-dose SC PK/PD study for comparison to the single-dose IV results and, if they are acceptable, then conduct the 3-month repeated-dose safety study with SC dosing. These studies should demonstrate that the SC route is tolerated and provides comparable PK/PD properties to the IV route. Likewise, to support switching from frozen liquid to lyophilized material during phase 2, one might conduct a single-dose PK/PD comparability study of the frozen and lyophilized materials. Last, one would ideally like to defer evaluating chronic and reproductive toxicity testing until after establishment of the final, high-producing cell line intended for commercialization and manufacture of the pivotal phase 3 clinical trial material. If the material intended to be used in the pivotal 9-month chronic toxicity testing study is derived from a new cell line and has not previously been tested in monkeys, one might consider first conducting a 1-month repeated-dose study to assess safety and comparability. Although not discussed here, one might also conduct clinical comparability studies in humans to support the change from IV to SC prior to phase 1 and any changes occurring between phases 2 and 3.

As was true for the alternative program to support the lack of a relevant species, the additional support required for changes to the route of administration, formulation, and cell line adds time and money to the preclinical program. In this case about 12 more months and an additional $2.7 M are required to enable BLA filing.

From Figure 7 it becomes apparent that discussions with regulatory authorities should be held at several critical times. While meetings to discuss clinical comparability protocols are an accepted practice, it is less common, but potentially just as important, to do the same for preclinical programs. The first meeting(s) would be to present the proposed changes and to assess the acceptability of the planned comparability testing.[13] Subsequent meetings might occur soon after completion of the animal comparability/safety testing

[13]In vivo comparability testing may not be required if the in vitro or biochemical characterization of the biologic is sufficient to support comparability of the pre- and post-change products.

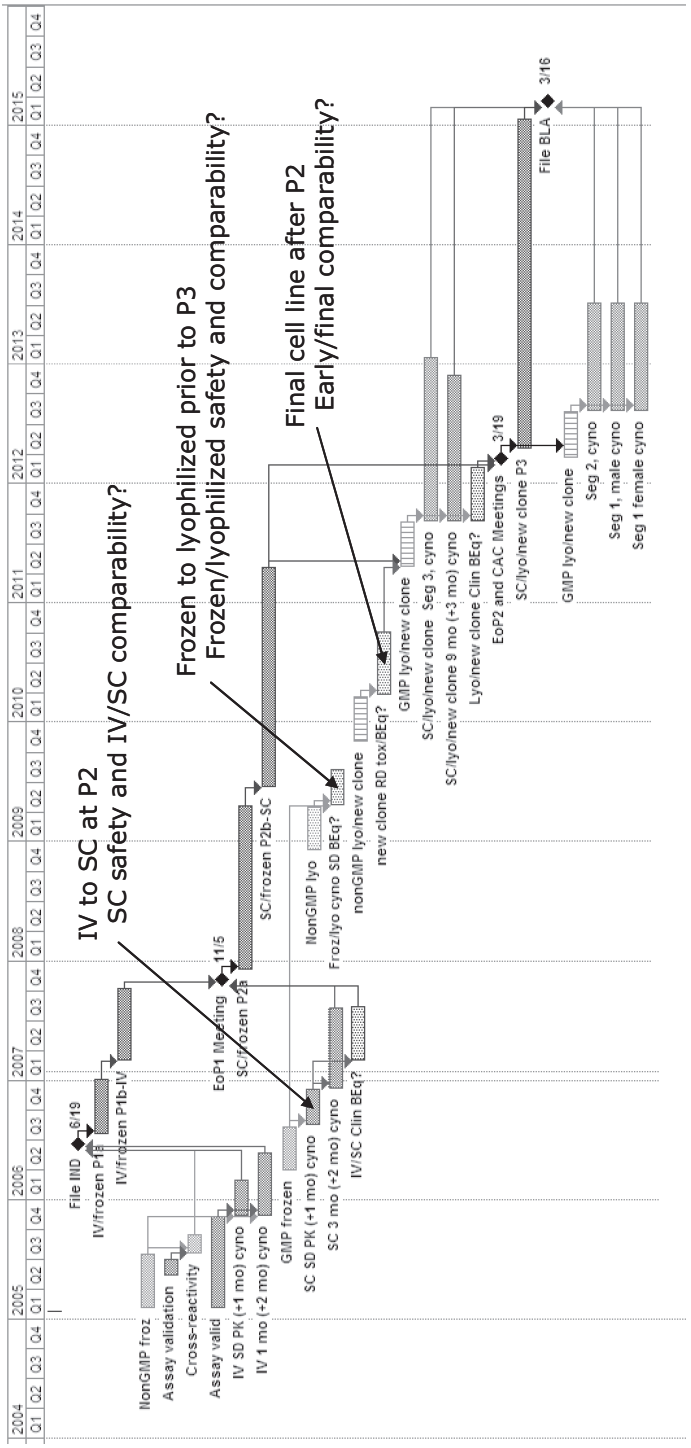


For comparison purposes, starting date was assumed to be 18 January 2005.

**Figure 7** Supporting changes to route of administration, formulation, and cell line.

(to determine whether comparability was established) and prior to human comparability testing. The point to be made is that an understanding of the acceptability of the proposed changes and the effect of those changes on the product profile is needed prior to continuing to the next phase of development. Without regulatory agreement to the plan, there will be a risk that the method of comparability testing will not be acceptable, and therefore not useful. Even with an agreement, there remains a risk that comparability will not be demonstrated, in which case a significant portion of the supportive data derived from the earlier material will not be appropriate for support of the later material. Failure to demonstrate comparability means that some studies may need to be repeated with the new material.

***The Challenge of Biologics Preclinical Development*** The examples above demonstrated that alternative approaches to preclinical development, intended to address lack of a relevant species or to support changes to the product or its manner of use, are associated with significantly more time required for development and often with greater expense. The delays and costs are related to the need to dually develop the surrogate and the product, or to perform comparability assessments of earlier and later versions of the product. In reality, many biologics programs must address issues with both appropriate safety testing and support of product changes. In this example, if it was necessary to develop a surrogate mAb and to support changes to the product and its use, the total time until BLA filing might be as long as 10 years, 10 months, and require around $11 M.

These "nontraditional" concerns are generally not encountered with drugs. For this reason it is often feasible for a drug program to design, conduct, and submit a complete preclinical development program that has a defined budget and time line and is of relatively certain regulatory acceptability. Consultations with regulatory authorities are therefore not as essential for a drug (with the exception of dose-setting for carcinogenicity testing). In contrast, the challenge with biologics is in designing preclinical development programs that address the relevant safety issues but, if nontraditional, are acceptable to regulatory authorities. Gaining some assurance of this acceptability for a biologic requires more frequent and earlier discussions with regulatory authorities.

## 5 MESSAGES FOR PRECLINICAL DEVELOPMENT PROGRAMS FOR BIOLOGICS

The fundamental responsibility of preclinical development is to enable clinical development by consistently providing the relevant target biology, pharmacology, PK/PD, immunogenicity, efficacy, or safety information, as well the required regulatory support, in advance of when it is needed. The goal is for preclinical development to never be rate-limiting for clinical development. The hypotheti-

cal comparisons and examples that are presented here serve to illustrate what a challenge this can be for some programs. From this exercise, several key messages for preclinical development programs for biologics emerge:

- *Plan strategically.* The entire integrated development program should be planned early, to identify when each clinical phase will require the supportive preclinical information. Critical decision points must be identified and these decisions made in a timely fashion. A strategy must be planned to address potential issues, such as lack of a relevant species or changes to the product or manner of use.
- *Design relevant studies to answer relevant questions.* The preclinical development program should consist of relevant studies designed to answer relevant concerns about pharmacology, PK/PD, efficacy, and safety, as well as comparability. No study should be conducted if it is not potentially capable of adding useful information to the knowledge of the product.
- *De-risk high-risk plans.* Whenever possible, discussions should be had with regulatory authorities to determine the acceptability of the preclinical program or studies for support of clinical development. Conducting alternative programs without some assurance as to their potential acceptability is a run-at-risk proposition. While gaining preliminary approval of the development program will not guarantee that it will eventually be acceptable, it will at least determine that it is not unacceptable. Only the program results will determine the eventual acceptability of the program.
- *Add value to clinical development.* The greatest value that preclinical development can offer is to enlighten or inform upon clinical development. This is more than just meeting regulatory requirements without becoming rate limiting. It includes, through careful selection of relevant species and incorporation of relevant endpoints, the ability to inform upon the likely PK/PD, pharmacology, efficacy and safety results to be seen in human clinical trials.

## 6 SUMMARY

The objective of this chapter was to compare to the studies, materials, and costs associated with hypothetical preclinical development programs intended to support the clinical development of a drug and a biologic. To accomplish this objective, an integrated development program was planned, from IND through NDA or BLA, and the respective preclinical development programs to support each filing were then designed and compared. This exercise demonstrated that relative to preclinical programs for drugs, there are some important differences in preclinical programs for biologics:

- Biologics often require nontraditional or alternative testing strategies due to limited species cross-reactivity (pharmacologic activity).
- Biologics often require additional studies to support changes to the product or its manner of use and to demonstrate comparability.
- Biologics often require investment of additional time, effort, and money to support a given phase of clinical development.
- Biologics often have greater risks of failing for technical, scientific, and/or regulatory reasons.
- Biologics often require significantly different timing for key development decisions to facilitate clinical development and manage risk.

Success for biologics will be achieved by de-risking the integrated development program with early, detailed, long-range strategic planning, frequent review of progress against plan, and early and frequent interactions with regulatory reviewers to assess acceptability of the plan and the results.

## ACKNOWLEDGMENTS

Learning the trade of preclinical development for biologics is a long and challenging, but fortunately shared, process. There seem to be no training programs that address the many nuances and differences from traditional drug development paradigms, the latter of which can be accessed in numerous texts. The results of development programs for biologics are generally not published, with the exception of summary information available after approval. However, the most valuable lessons to be learned often reside in the programs that do not reach approval. Thus much of what we can learn of biologics development comes from our colleagues. While I have had the opportunity to work on programs for a wide variety of biologics, I have been especially fortunate to have had the benefit of the council of others who have had more and different experiences than me. I am particularly grateful for the contributions of Lauren Black, Page Bouchard, Joy Cavagnaro, Mary Ellen Cosenza, James Green, Marie Green, Mark Milton, and George Treacy to my ongoing education.

## REFERENCES

1. ICH M3(R1). Maintenance of the ICH Guideline on Non-clinical Safety Studies for the Conduct of Human Clinical Trials for Pharmaceuticals. International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use, 9 November 2000.
2. ICH S1A. Guideline on the Need for Carcinogenicity Studies of Pharmaceuticals. International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use, 29 November 1995.

3. ICH S1B. Testing for Carcinogenicity of Pharmaceuticals. International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use, 16 July 1997.
4. ICH S1C(R1). Dose Selection for Carcinogenicity Studies of Pharmaceuticals and Limit Dose. International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use, 17 July 1997, November 2005.
5. ICH S2A. Guidance on Specific Aspects of Regulatory Genotoxicity Tests for Pharmaceuticals. International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use, 19 July 1995.
6. ICH S2B. Genotoxicity: A Standard Battery for Genotoxicity Testing of Pharmaceuticals. International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use, 16 July 1997.
7. ICH S3A. Note for Guidance on Toxicokinetics: The Assessment of Systemic Exposure in Toxicity Studies. International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use, 27 October 1994.
8. ICH S3B. Pharmacokinetics: Guidance for Repeated Dose Tissue Distribution Studies. International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use, 27 October 1994.
9. ICH S4. Duration of Chronic Toxicity Testing in Animals (Rodent and Nonrodent Toxicity Testing). International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use, 2 September 1998.
10. ICH S5(R2). Detection of Toxicity to Reproduction for Medicinal Products and Toxicity to Male Fertility. International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use, November 2005.
11. ICH S7A. Safety Pharmacology Studies for Human Pharmaceuticals. International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use, 8 November 2000.
12. ICH S7B. The Non-clinical Evaluation of the Potential for Delayed Ventricular Repolarization (QT Interval Prolongation) by Human Pharmaceuticals. International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use, 12 May 2005.
13. ICH S8. Immunotoxicity Studies for Human Pharmaceuticals. International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use, 15 September 2005.
14. ICH S6. Preclinical Safety Evaluation of Biotechnology-derived Pharmaceuticals. International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use, 16 July 1997.
15. ICH Q5E. Comparability of Biotechnological/Biological Products Subject to Changes in Their Manufacturing Process. International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use, 18 November 2004.

# Demonstration of Comparability of a Licensed Product after a Manufacturing Change

RICHARD M. LEWIS, PhD

**Contents**

1 Introduction
2 Definition of Terms
  2.1 Comparability
  2.2 Equivalence
  2.3 Generic Drugs and Bioequivalence
  2.4 Biogenerics
3 Regulatory Guidance for Determining Comparability
  3.1 The FDA
  3.2 The EMEA
  3.3 The ICH
4 General Principles and Practices
  4.1 Comparability May Not Be Required
  4.2 Manufacturing Changes
  4.3 Physicochemical Characterization
  4.4 Preclinical Considerations
  4.5 Clinical Considerations
  4.6 Specific Considerations Regarding Immunogenicity
5 Additional Applications of Comparability
6 Regulatory Submissions after US Marketing Approval
7 Conclusion
  References

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

## 1 INTRODUCTION

Biopharmaceuticals have traditionally been defined by their manufacturing processes. Over the course of a product's life cycle a biopharmaceutical manufacturer often faces the question of whether to make changes to the manufacturing process and when to implement such changes. The need to make a change may be driven by a variety of reasons, including contractual obligations, manufacturing capacity, improving purity, improving safety, increasing product yield, or optimizing the ultimate cost of goods of manufacture. There is also regulatory pressure to maintain current good manufacturing practices (cGMPs). The FDA has recently focused on maintaining cGMPs in their "GMPs for the twenty-first century" initiative [1]. Central to this proposal are the concepts of quality systems [2] and the process analytical technologies (PAT) initiative [3]. The PAT guidance is intended to describe a regulatory framework that will ensure the voluntary development and implementation of innovative pharmaceutical manufacturing and quality assurance [4]. The framework is intended to assist manufacturers in developing and validating new efficient tools to maintain product quality while controlling the manufacturing process. These concepts will need to be considered, among other decisions, when process changes are implemented. Since quality concepts need to be considered prior to implementing a process change, the most difficult decision for a product developer is when to implement manufacturing changes. Some of the issues to consider will be addressed in this chapter.

The initial goal in making a change is the demonstration that the original product has not changed in "identity, strength, quality, purity, and potency" and that previously developed safety and efficacy data will apply to the product processed using the new method. If any differences are observed, they must then be evaluated for their impact on the current product and relevance to the previously obtained data. Importantly, even minor changes must be considered as potentially altering the above-mentioned product attributes. This chapter will provide an overview of the key issues and considerations for implementing a manufacturing change. The process that a manufacturer will go through to determine these characteristics has been referred to as "comparability."

## 2 DEFINITION OF TERMS

### 2.1 Comparability

The concept of comparability was first introduced by the FDA in 1996 in its guidance, FDA Guidance Concerning Demonstration of Comparability of Human Biological Products, Including Therapeutic Biotechnology-derived Products [5]. The term *comparability* was specifically chosen (1) to recognize the possible lack of identity after a manufacturing change and (2) to distin-

guish the concept from other specific pharmaceutical regulatory terms: equivalence or bioequivalence. A successful demonstration of comparability between two biopharmaceutical materials does not mean that the products are identical in every way. Minor differences in the product may be identified, but the products can still be viewed as comparable. Similarly each batch or lot may not be exact but meet the boundaries or specifications that have been established for the process [6]. The challenge for developers, manufacturers, and regulators is to identify those characteristics that define the identity, strength, quality, purity, and potency. Although the term comparability is used in the FDA guidance, it is applied primarily to the changes in manufacturing; the FDA has also used comparability principles in other ways.

### 2.2 Equivalence

Equivalence is meant to be a demonstration of a specified statistical confidence that two compounds can achieve the same characteristics. It often has been used to describe the same outcome or endpoint frequency in clinical trials. Equivalence is not a comparative evaluation of similarity between different manufacturing methods of the same product or necessarily a comparison of extremely similar products. Equivalence also is a difficult statistical assurance to reach in a clinical trial, and therefore it is an extreme standard for comparison of manufacturing processes.

### 2.3 Generic Drugs and Bioequivalence

Bioequivalence is distinct from both equivalence and comparability and it is defined in the regulations for generic drugs. Generic drugs are approved on the basis of identity of strength, dosage form, safety, quality, performance characteristics, and route of administration. They must have the same indication for use and demonstrate bioequivalence. A generic drug is considered bioequivalent to an innovator if it exhibits the same rate and extent of absorption of the active or therapeutic ingredient, although inactive ingredients may vary, and becomes available at the site of drug action. Acceptable absorption parameters are those between 80% and 125% of those obtained with the proprietary agent under the same testing conditions. These mechanisms for approval are meant to address products that are identical to drugs, distinct from biologics, and have an innovator that has been approved as a drug. Generics are regulated under the Food Drug and Cosmetic (FD&C) Act, sections 505(b)(2) and 505(j). Under section 505(b)(2) of the FD&C Act, the FDA may rely on data for approval of a new drug application that is not developed by the applicant. Some or all of the clinical information can be provided by the literature or by references to a past FDA finding of safety and effectiveness for approved drugs. Generic drugs in the United States can be approved based on a demonstration of bioequivalence of pharmacokinetic data. Those products can be substituted (are interchangeable). The conditions

which are accepted for bioequivalence are defined in the regulation but those for comparability are generally negotiated with the FDA and based on the best scientific information that is available. The decision about comparability is determined on a case-by-case basis initially using multiple analytical tests to characterize the chemical and physical attributes of the process materials together with considerations of the complexity of the product and the scope of the manufacturing change.

## 2.4 Biogenerics

Biological products that provide for generic medications were not part of the Hatch–Waxman amendments to the FD&C Act. In part due to historical precedence and in part due to scientific concerns of the FDA, biologics do not have regulatory mechanisms to provide for the approval of "generic" compounds for a previously approved biological drug. As a result of the large degree of interest globally in providing the same advantages to biologic drug development as the Hatch–Waxman Act provided to pharmaceuticals, the concept of biogenerics is being discussed widely in industry and regulatory venues as well as congressional and public settings.

European health authorities have developed regulations that allow a form of generic biologics for recombinant proteins and monoclonal antibodies. The Committee for Medicinal Products for Human Use (CHMP), of the European Medicines Agency (EMEA), issued a series of guidelines that address their expectations for applications as Marketing Authorization for Similar Biological Medicinal Products [7]. To support this new approach CHMP has issued additional guidelines: (1) regarding the demonstration of quality, Similar Biological Medicinal Products Containing Biotechnology-Derived Proteins as Active Substance: Quality Issues [8], and (2) for assessment of applications containing clinical and nonclinical data, Similar Biological Medicinal Products containing Biotechnology-Derived Proteins as Active Substance: Non-clinical and Clinical Issues [9]. The development of so-called biosimilars is dependent on the principles of comparability and thus the above-mentioned guidelines refer to the EMEA's previously issued Guideline on Comparability of Medicinal Products Containing Biotechnology-Derived Proteins as Active Substance: Quality Issues [10] and also Guideline on Comparability of Medicinal Products Containing Biotechnology-Derived Proteins as Active Substance: Non-clinical and Clinical Issues [11]. These guidelines, the concept of biosimilars, and the use of comparability principles add to the importance of, and confusion over, the term comparability. In the described CHMP comparability guidelines, comparability can be applied to either a manufacturing change by a single manufacturer or to the regulatory application of a biologic claiming similarity to a previously approved product after patent protection has ended. Thus the principle of comparability is central to the development of the so-called biosimilars. In fact, the first two biosimilar products, both for human DNA-recombinant growth hormone, were approved under these guidelines in 2006 [12].

It is beyond the scope of this chapter to discuss biogenerics and biosimilars, also referred to as follow-on biologics (in the United States), subsequent entry biologics (in Canada), and multisource biological products. The fate of a regulatory mechanism for approval of follow-on biologics in the United States is still undetermined, and the debate will likely continue as biopharmaceuticals take on an increasing share of the drug market and thus an increasing importance to the health care market [13].

## 3 REGULATORY GUIDANCE FOR DETERMINING COMPARABILITY

### 3.1 The FDA

The FDA guidance describes the possible evaluations that may be needed to demonstrate comparability to the US regulatory authorities. The document is broad enough to address changes in manufacturing at premarketing or after licensing, although it primarily addresses changes during development after some clinical studies have been completed. The comparability exercise may include a combination of analytical testing, biological assays (in vitro or in vivo), assessment of pharmacokinetics and/or pharmacodynamics and toxicity in animals, and clinical testing (clinical pharmacology, safety, or efficacy). The usual progression of complexity is from analytical to animal studies to human pharmacokinetics and/or pharmacodynamics to clinical safety and efficacy studies. Analytical testing is regarded as the most precise measure of a molecule's attributes and thus serves as the first tier for comparability determination. If product differences are observed in analytical testing, then additional preclinical studies or a clinical pharmacokinetic or limited pharmacodynamic study may be warranted. If differences are observed after additional limited clinical testing, or if there is insufficient product knowledge of the impact of differences on safety, purity, or potency, then data from larger clinical studies may be needed.

As previously mentioned, product manufacturing changes are made for many reasons: to increase product supply through process optimization such as to increase yield, to increase manufacturing capacity by scale-up or duplication and addition of viral reduction methods, or to increase compliance with cGMPs. The precise manufacturing process of all drug products must be described to regulatory agencies. In the United States, changes in manufacturing must be reported in either an active IND or to the BLA depending on stage of product life cycle. (See Section 6 for various types of license supplements that can be used to report changes.)

The ultimate goal of an investigational product is the demonstration of safety and efficacy and in the case of a licensed product that it has suitably provided that demonstration. Thus the primary concern after a manufacturing change is if that change will affect the safety or efficacy (safety, identity, purity, or potency) that has been demonstrated previously. The guidance points out

that changes that appear minor may have major effects and major changes may exert minor effects. For this reason it is important to select those assays that can identify the characteristics that contribute to safety and efficacy.

### 3.2 The EMEA

The EMEA has also issued guidance on Comparability, as mentioned above, on quality issues [10] and for nonclinical and clinical issues [11]. In these guidelines the concept of comparability can be applied to either a manufacturing change by one manufacturer or to the application of a biologic claiming similarity to a previously approved product after patent protection has ended. This guidance offers more detail than either the ICH document or the FDA guidance. A complete section of the EMEA guideline addresses aspects of immunogenicity testing. It discusses factors that can contribute to changes in immunogenicity: measurement of antibody responses, antibody testing strategies, validation of antibody assays, and timing of sampling. The document insists that immunogenicity "be considered when a claim of comparability is made, especially when repeated administration is proposed."

### 3.3 The ICH

The International Conference on Harmonization has noted the importance of comparability and itself has issued guidance [14]. The FDA has accepted the ICH guidance, and it is a part of the material that offers advice in developing comparability data. In many respects it follows the previously published FDA guidance; however, with the benefit of an additional decade of experience with biopharmaceutical manufacturing processes and clinical validation, this document provides greater details for determining comparability.

The ICH published The Comparability of Biotechnological/Biological Products Subject to Changes in Their Manufacturing Process [14] in 2005. In comparison to the initial 1996 FDA guidance, this guidance addresses similar concerns with less emphasis on the preclinical and clinical aspects but it is similarly intended to address manufacturing changes for a product by a single manufacturer. It addresses changes that are made either during product development or manufacturing changes that are meant to improve the production of a marketed product. One principle, that is common to both documents and should be applied in all parts of the comparability exercise is that material before and after the manufacturing change should be compared in every analytical, biological, and stability evaluation as well as in nonclinical and clinical methods if deemed necessary.

By comparison to the EMEA document, the ICH guideline does not address immunogenicity in the same detail. It does suggest that immunochemical properties be a component of the characterization and considerations for immunogenicity should be a part of the planning of nonclinical and clinical

**TABLE 1 Comparability guidance documents**

| Agency | Guidance | Purpose | Reference |
|---|---|---|---|
| EMEA | Guideline on similar biological medicinal products | Information requirements for the marketing authorization applications | 5 |
| EMEA | Similar biological medicinal products containing biotechnology-derived proteins as active substance: quality issues | Quality issues in demonstrating comparability of similar biological medicinal products containing recombinant DNA-derived proteins (proteins and peptides, their derivatives and products) | 6 |
| EMEA | Similar biological medicinal products containing biotechnology-derived proteins as active substance: nonclinical and clinical issues | Nonclinical and clinical requirements for demonstrating comparability of similar biological medicinal products | 7 |
| EMEA | Guideline on comparability of medicinal products containing biotechnology-derived proteins as active substance: quality issues | Quality of the comparability exercise for either change in a manufacturing process or to demonstrate similarity in a marketed product | 8 |
| EMEA | Guideline on comparability of medicinal products containing biotechnology-derived proteins as active substance: nonclinical and clinical issues | Quality of the comparability exercise for either change in a manufacturing process or to demonstrate similarity in a marketed product | 9 |
| ICH/U.S.FDA | Comparability of biotechnology-derived, ICH Q5E | Comparability of biotechnological/biological products for changes made in the manufacturing process | 10 |
| U.S. FDA | FDA guidance concerning demonstration of comparability of human biological products, including therapeutic biotechnology-derived products | Comparability testing for manufacturing changes made prior to product and after product approval for a single manufacturer | 4 |

studies, in particular, when considering various aspects of the knowledge of the product use.

## 4 GENERAL PRINCIPLES AND PRACTICES

The scope of a comparability program should usually be established in relation to the significance of the change that is being evaluated, the stage of development, and the clinical indication. Each program should take into consideration how the change may affect the finished product as well as attempt to answer other questions: How much of a "difference" is acceptable? Is lack of "sameness" an issue for the particular stage of development? Are the pivotal safety and efficacy trials in progress? Do blood concentrations correlate with toxicity and/or efficacy [15]?

With a marketed product it is more important to provide assurance of comparability prior to committing to making the change. Even more than with a product in development, it is important to have confidence that the current manufacturing method produces a product on which there is confidence that there can be a reliance on previous data, including in vitro, preclinical, and clinical information, for the safety and efficacy of the current methods. Without complete confidence in the comparability, it will be difficult to evaluate trends in safety reporting.

In carrying out all aspects of the comparability exercise, it is important to consider previous knowledge of the product and, in particular, the relationship between the characteristics of the product and the effects documented by available preclinical and clinical experience. Other important considerations include recognizing the contribution of the particular production step that is changed, the potential impact of the change, and the ability to measure any predicted change. In evaluating the change, knowledge of the key physiological characteristics of the active product is central to selecting the proper traits to be tested and which manufacturing stages are relevant. In any case, the drug substance and drug product should both be included as part of the comparison. Also it is important to demonstrate a consistency of manufacturing, which is usually done with a limited number of manufacturing batches. The effect on stability should be measured and compared with the previous material [16].

Determining the potency of the final product and comparing the characteristics of drug substance and drug product are generally expected. However, in some cases the most sensitive step in the process to detect changes may be in intermediate fractions. In such cases key intermediate materials should be compared.

The effects that changes will have on the product at critical control points, in-process controls, and downstream steps must be considered and evaluated whenever possible. Samples from intermediate process steps may provide material that can be better evaluated for possible impurities or, in some cases, infectious agents. The milieu in which material is applied to purification

matrices may affect separation parameters. As such, it is not only the degree of purity, specific activity, or concentration but also the consistency of the material in which it is isolated. The presence or absence of proteolytic enzymes or carriers may affect the chemical nature of the product. The FDA has recently emphasized process analytical technologies (PAT), and it should be recognized that upstream manufacturing changes can have important effects on in-process measurements [3]. A change in the purity profile, as well as the impurity profile, of one step has the potential to change the effectiveness of a subsequent manufacturing step. It is always important to note that the resulting material from one manufacturing step is the starting material for the following part of the process.

While it is likely that minor changes will result in a comparable product, a series of small changes over time can amount to a significant change in the product. For this reason manufacturers should evaluate the potential for gradual "drift" of the product or, taken together, there may be an interaction that is unforeseen and may be a reason for doing comparability studies. Also the changes may have some interactive function that cannot be predicted and will prompt the need for comparability studies.

It is very important that a side-by-side comparison be performed using the product from the previous process with the product from the new process. Sufficient "old" product and intermediates must therefore be available (retained) to allow for the comparability studies. During the initial characterization of the product, numerous assays are selected to ensure in-process consistency, including essential characteristics of key intermediates. It is assumed that these assays describe the important characteristics of the product. For this reason most of these assays will be employed in demonstrating comparability.

Biological assays, in particular potency assays, are meant to demonstrate a relationship between the product and the desired biological effect. The correlation between the activity measured by the potency assay and the resulting clinical effect is an important criterion in both dosing and the determination of efficacy. International or other accepted reference standards, when available, should be incorporated in the assay.

In theory, comparability is the provision of data so that there is sufficient confidence in product made by a new manufacturing method that one can rely on the data previously developed using the former manufacturing method. Although it is not necessarily a stepwise process, in practice comparison data are generated first in vitro, then in preclinical studies and finally in clinical trials. The need for a step up in each case generally is based on the data obtained in the current tier.

### 4.1 Comparability May Not Be Required

If the manufacturing change is carried out prior to any preclinical or clinical testing, there is little need to compare the products made by the two methods

except for development purposes. The primary reason for determining comparability is to have a basis for reliance on previous data. The bar for establishing comparability will therefore likely be higher for changes made during or after phase 3 clinical studies.

### 4.2 Manufacturing Changes

The types of changes that a manufacturer might face are varied and diverse, from minor to major; to list some: transition from in vivo to in vitro production for a monoclonal antibody, changes in DNA vector, a new master cell bank, use of a different cell substrate (host cell system), changes in raw materials, different culture media and culturing conditions, different serum to serum-free tissue culture, changes in a purification step, changes in storage conditions, and changes in equipment and facilities including pilot versus full-scale product or a new production line. There may also be changes in the purity and impurity profiles, deliberate molecular modifications to the protein, or changes in the final formulation including changes in excipients, container closure, as well as a change from liquid to lyophilized product.

Some of the manufacturing changes that can alter biochemical structure include changes in cell substrate, raw materials, bioreactor conditions and purity [17]. These changes may lead to posttranslational modfications in glycosylation, may increase process- or product-related impurities that might be immunopotentiating, or may include degradation products that lead to aggregation. It is for this last reason that the FDA insists on evaluating the monomer pattern of the final product. A number of manufacturing changes have been shown to lead to aggregation. In one instance, merely the change in the container-closure system provided an increase in leached materials which induced aggregation [18]. Because aggregation can induce immunogenicity, alter pharmacokinetic (PK) characteristics, and induce hypersensitivity (e.g., intravenous immunoglobulin, IGIV), the evaluation of monomer patterns is usually a required evaluation in any comparability exercise.

There is a close relationship between the comparability of product made in different processes and in the applied current good manufacturing practices (cGMPs). The FDA has recognized that degree of GMP does not have to be as stringent for products in early phases of development [19,20]. It should also be noted that the better a product is characterized, the easier it is to demonstrate that a new process produces a comparable product.

### 4.3 Physicochemical Characterization

The availability of sensitive analytical tests to adequately characterize a product and to measure the predicted possible changes is central to a comparability assessment. Obviously tests should be chosen that will be the most likely to detect important changes. ICH Q6B offers advice on physicochemical, biological activity, and immunochemical properties, purity (impurity), con-

taminants and quantity [21]. The assay standards and statistical assurance of assay validity are also addressed as well as the application of testing to in-process material, drug substance, and drug product.

The importance of determining biological activity should be emphasized. The association of this activity with clinical effect is an important component of product characterization, in particular, potency and stability testing. Before embarking on animal or human trials to evaluate comparability, it is important to evaluate the data gathered in the quality testing and determine the need for additional studies. If any differences are observed, the nature of those differences will determine the need for animal or clinical studies. The association of the changes with known safety and therapeutic potential will help to contribute to the decision. In addition the dosing regimen, the therapeutic window, and the previous experience with the product will determine the need for in-life trials.

Often analytical testing and biological characterization are sufficient to provide evidence of comparability. One estimate of frequency of the use of clinical data for comparability demonstrations was 1% [22]. However, more recently the FDA has begun to request more clinical PK data for defining comparability of products in clinical development.

### 4.4 Preclinical Considerations

Changes in three-dimensional structure can ultimately affect pharmacokinetic profile, receptor affinity, and immunogenicity. Thus comparability programs often focus on those measures that can best be used to determine the retention of a molecular identity. While in most cases analytical testing and biological characterization are sufficient to provide evidence of comparability, various in vitro studies or in vivo animal studies may need to be considered. Animal pharmacokinetics may also be needed even in the absence of demonstrated differences in analytical testing of the functional assays for the product. This is because analytical testing may be insensitive to changes affecting pharmacokinetics and in vitro functional tests may not reflect the time-dependent aspects of distribution [5].

Assay performance criteria for biopharmaceuticals are often highly variable; therefore strict statistical criteria that attempt to rigorously establish traditional in vivo bioequivalence may not always be appropriate. In some cases an assessment of rate and extent of absorption as indicated by the maximum concentration ($C_{max}$), time of maximum concentration ($T_{max}$) and area under the curve ($AUC$) may be needed. In other cases complicating factors related to binding proteins, endogenous concentration, and unusual concentration-time profiles may need to be considered [15]. In cases where complications may arise from immune response to heterologous proteins, cross-over designs are inappropriate.

The extent of additional toxicology studies depends on where in the phase of clinical development or the life cycle of product the change is made, the

previously developed safety profile and on the magnitude of the manufacturing change [15]. Additional animal studies or "bridging studies" may be needed where the product has a narrow therapeutic index or where specific safety concerns are present, such as when the process change raises concerns about possible toxic impurities or adventitious agents that cannot be assessed with analytical testing. Bridging studies may also be indicated if there is a significant change in the final formulation or route of administration.

### 4.5 Clinical Considerations

ICH Q5E is emphatic in stating that both preclinical and clinical data may not be necessary if the manufacturer can provide assurance of comparability through the previously described analytical program. However, when these data are insufficient, additional evidence may be required, and this is determined on a case-by-case basis. Various studies may be recommended based on the results of the analytical studies, knowledge of the product, and clinical information, in particular, regarding dosing indication, therapeutic index, and previous clinical experience. The studies may be PK, pharacodynamics (PD), efficacy, safety, immunogenicity, or phase 4 studies. Like all other parts of the program, these studies should be based on a direct comparison [15]. Similarly the 1996 FDA guidance states that usually the reason for the analytical studies is to avoid additional clinical trials. When needed, the additional human studies will be to evaluate changes that may affect pharmacodynamics and pharmacokinetics [5].

The EMEA, CHMP, offers more detail in their advice; however, the specific need for clinical data, like the ICH and FDA guidances, will depend on the degree of comparability measured in analytical methods. A wide variety of conditions for drug use and the availability of a surrogate marker may affect how the clinical comparison is performed. If a surrogate marker is available, PK/PD studies may be sufficient, and the degree of knowledge of the product along with the degree of difference will dictate the need for additional clinical data. It will be important, especially for marketed products, to follow closely the safety profile of the product using well-designed pharmacovigiliance methods to be assured that no additional safety concerns were introduced with the change.

When manufacturing changes are made prior to approval, comparability usually is demonstrated through analytical and sometimes animal studies before proceeding to the next phase of development. If late in development (e.g., during phase 3), it might be necessary or valuable to demonstrate comparability as part of the trial(s).

It is important to note that in order to be able to adequately compare product made from two different manufacturing schemes, material must be available from both. There has been more than one manufacturer who changed process without retaining sufficient material of the former method to do adequate comparisons. As discussed above, it should be noted that the final product is not always the best or only material to use for comparative studies.

The concentration and nature of the impurities may be important in manufacturing steps further downstream.

### 4.6 Specific Considerations Regarding Immunogenicity

How does immunogenicity figure into comparability? Changes in manufacturing can undoubtedly give rise to changes in immunogenicity. The known association of particular molecular changes with changes in immunogenicity, such as secondary and tertiary conformational changes, may indicate the need for some assessment of the potential for an immune reaction to the new material if these changes are identified. Other considerations include the nature of the product itself, the dose schedule and route, the duration of therapy, and the immune status of the intended patient population. Given the nature of biological products, it is often difficult to determine the immunogenic potential of a particular product much less to be able to compare between similar products.

Immunogenicity concerns are based on a number of potential safety concerns. Adverse reactions can be based on the formation of immune complexes that can give rise to renal toxicity, complement activation, and, as recently reported, the induction of autologous antibodies that cross-react with the patient's own endogenous protein [23–28].

One example of a plasma-derived coagulation factor VIII showed a marked increase in inhibitor formation (antibodies to factor VIII that block activity) [29,30]. The question of immunogenicity is one of the most difficult questions to answer without clinical data. However, relative immunogenicity as measured by frequency of the development of antibodies or the relative magnitude of a reaction can be determined in animals, even though it is understood that treating animals with human proteins usually will result in the development of an immune reaction. The FDA has accepted one approach to demonstrate a lack of neo-antigenicity. This is the adsorption of antibodies against the novel material with the previous material. If all antibody activity can be adsorbed by the previous version, it suggests a lack of neo-antigenicity. The use of transgenic and knockout mice may also provide a means to evaluate neo-antigenicity. All evaluations of immunogenicity are dependent on the reliability, sensitivity, and specificity of the assays used.

Other chapters in this volume specifically address immunogenicity and preclinical models and may also be relevant in the context of comparability (Chapters General Toxicity Testing and Immunotoxicity Testing for Biopharmaceuticals and Immunogenicity of Therapeutic Proteins and the Assessment of Risk). The FDA has allowed changes in production without a clinical evaluation of immunogenicity.

## 5 ADDITIONAL APPLICATIONS OF COMPARABILITY

The FDA has applied the concept of comparability in other unique ways in order to address specific regulatory needs. It has recently discussed the concept

of comparability in its draft "Guidance for Industry: Minimally Manipulated, Unrelated, Allogeneic Placental/Umbilical Cord Blood Intended for Hematopoietic Reconstitution in Patients with Hematological Malignancies" [31]. Generally, comparability is used to demonstrate that there can be reliance that a novel process produces a safe and effective product based on data gathered from a previous method. (The "new product" relies on data gathered using the "old product.") In the umbilical cord blood guidance, the FDA has shown a willingness to use current data for licensure (i.e., as a demonstration of safety and efficacy) and to apply comparability principles to ensure safety and efficacy of product made under a previous method. In other words, rather than use comparability to make products available from a newer process, the FDA suggests using it to make products available from previous, unlicensed methods. (Data on the "new product" is used for confidence in the "old product.") This approach is important to ensure the availability of products for transplantation to individuals with very limited possibilities for matching donors. Because of the rapid changes in the field of umbilical cord blood transplantation, it is likely that many tests used today are not the same as those in a past decade nor validated in the same ways when these important products were first being collected and stored. Although much of the testing may not be identical, the cellular products may be the same or at least "comparable." With the assurance of comparability older products can be considered licensed.

In this guidance the FDA offers some direction for the demonstration of comparability. They request that the manufacturer provide evidence that the methods, facilities, and controls that were used to manufacture previous products conformed to cGMPs and to other applicable regulatory requirements. In addition they request the submission of validation summaries, as well as product characteristics such as total nucleated cell count, viable CD34 cell count, and number of colony-forming units. Stability data and information from the scientific literature can also be used. Clinical outcomes can be part of the comparability demonstration.

Upon approval of the application and the comparability data, all products could be made available under the license. Some of the concepts addressed for umbilical cord hematopoietic stem cells may be important in future comparability demonstrations in hematopoietic and other cell therapies.

## 6 REGULATORY SUBMISSIONS AFTER US MARKETING APPROVAL

For manufacturing changes that are made after marketing approval, the manufacturer must submit data demonstrating that the change does not affect the product. For licensed biologics, the submission will be as a biologics license supplement (BLS) and will fall in one of a few categories [32]. A prior approval supplement describes changes that have a substantial potential to have an

adverse effect on the identity, strength, quality, purity, or potency of the product. The changes should not be implemented until the supplement has been reviewed and approval given. In changes that are thought to have a moderate impact on the product, a changes being effected in 30 days (CBE30) supplement, or a changes being effected (CBE) is submitted, which enables implementation within 30 days or immediately, respectively. Data documenting the change must be submitted and will be reviewed by the FDA. Other changes are reported in an annual report. A decreased time to distribution of product after the change in CMC can be facilitated if the manufacturer has the foresight to identify those changes that may be made in the future and/or a series of changes that may be implemented. It is also possible to decrease the reporting burden. This can be accomplished with the approval of a "comparability protocol."

> A comparability protocol is a well-defined, detailed, written plan for assessing the effect of specific CMC changes in the identity, strength, quality, purity, and potency of a specific drug product as these factors relate to the safety and effectiveness of the product. A comparability protocol describes the changes that are covered under the protocol and specifies the tests and studies that will be performed, including the analytical procedures that will be used, and acceptance criteria that will be achieved to demonstrate that specified CMC changes do not adversely affect the product [32].

It is important to note that the submission of a comparability protocol is a prospective approach. That is, the protocol with tests and acceptance criteria, including specifications, must be submitted even before the testing has begun. It is a protocol alone. It is more important in this effort to select the most appropriate analytical methods to be used. There are a number of advantages of this approach. Once the protocol is approved by the FDA, as a prior approval supplement, it can lower the type of biologics license supplement category. For example if the manufacturing change would need to be reported as a prior approval supplement, the approval of a comparability protocol could reduce that category to a changes being effected in 30 days (CBE30). Similarly a CBE30 could be reduced to reporting through an annual report. In addition the FDA, having approved a specific protocol, is less likely to ask questions that might delay final implementation.

The FDA has provided two separate guidances for submitting comparability protocols. One guidance is provided for therapeutic recombinant DNA-derived protein products, naturally derived protein products, plasma derivatives, vaccines, allergenics, and therapeutic DNA plasmids. This guidance also applies to new drug applications (NDAs), abbreviated new drug applications (ANDAs), or supplements to these applications for protein drug products, and peptide products that cannot be fully characterized (e.g., complex mixture of small peptides) [33]. Guidance is also available that addresses changes for comparability protocols that would be submitted in NDAs, ANDAs, or supplements to these applications, except for applications for protein products [34].

## 7 CONCLUSION

The demonstration of comparability has been addressed nationally and internationally by several published regulatory guidance documents. It should be noted that the better a product is characterized, the easier it is to demonstrate that a new process produces a comparable product. Importantly, the most relevant knowledge base for the physicochemical and biological characteristics of the product rests with the manufacturer. It is this specific knowledge that is paramount in developing the essential tests to use in the comparability exercise. As changes in manufacturing are inevitable, it is critical to retain material from certain product batches in order to be able to compare materials from the process before and after manufacturing changes, using in-process materials at critical control points and at manufacturing steps that will be the most sensitive for molecular changes that could offer the most risk. A manufacturer considering a process change should evaluate the overall process and determine the need and reason for making the change. The change should be viewed as part of research and development until there are satisfactory results from the comparability exercise. Once the tests are identified, samples collected from both processes, analysis performed and conclusions drawn, it will be possible to implement a well-documented manufacturing improvement.

## REFERENCES

1. Pharmaceutical cGMPs for the 21st Century—A Risk-Based Approach Final Report, 2004. www.fda.gov/cder/gmp/gmp2004/GMP_finalreport2004.htm
2. Guidance for Industry, Quality Systems Approach to Pharmaceutical CGMP Regulations, 2006. http://www.fda.gov/cder/guidance/7260fnl.htm
3. Process Analytical Technologies Initiative, 2008. http://www.fda.gov/cder/OPS/PAT.htm
4. Guidance for Industry PAT—A Framework for Innovative Pharmaceutical Development, Manufacturing, and Quality Assurance, 2004. http://www.fda.gov/cder/guidance/6419fnl.pdf
5. FDA Guidance Concerning Demonstration of Comparability of Human Biological Products, Including Therapeutic Biotechnology-derived Products, 1996. http://www.fda.gov/cder/Guidance/compare.htm
6. Gerrard T. Statement to the Committee on Oversight and Government Reform, Safe and Affordable Biotech Drugs—The Need for a Generic Pathway, 2007.
7. Guideline on Similar Biological Medicinal Products, 2005. http://www.emea.europa.eu/pdfs/human/biosimilar/043704en.pdf
8. Similar Biological Medicinal Products Containing Biotechnology-Derived Proteins as Active Substance: Quality Issues, 2006. http://www.emea.europa.eu/pdfs/human/biosimilar/4934805en.pdf

9. Similar Biological Medicinal Products Containing Biotechnology-Derived Proteins as Active Substance: Non-clinical and Clinical Issues, 2006. http://www.emea.europa.eu/pdfs/human/biosimilar/4283205en.pdf
10. Guideline on Comparability of Medicinal Products Containing Biotechnology-Derived Proteins as Active Substance: Quality Issues, 2003. http://www.emea.europa.eu/pdfs/human/bwp/320700en.pdf
11. Guideline on Comparability of Medicinal Products Containing Biotechnology-Derived Proteins as Active Substance: Non-clinical and Clinical Issues, 2004. http://www.emea.eu.int/pdfs/human/ewp/309702en.pdf
12. Annual report of the European Medicines Agency, 2006. http://www.emea.europa.eu/pdfs/general/direct/emeaar/EMEA_Annual_Report_2006_full.pdf
13. Tsang L, Beers D. Follow-on Biological Products: The Regulatory Minefield. *Life Sciences, Global Counsel Handbooks* 2004;105.
14. Comparability of Biotechnological/Biological Products Subject to Changes in Their Manufacturing Process, Q5E 2005. http://www.fda.gov/cder/guidance/6677fnl.htm
15. Mordenti J, Cavagnaro J, Green J. Design of biological equivalence programs for therapeutic biotechnology products in clinical development: A perspective. *Phar Res* 1996;13:1427.
16. Guidance for Industry Q1A(R2). Stability Testing of New Drug Substances and Products, 2003. http://www.fda.gov/cder/guidance/5635fnl.htm
17. Swit M. Reg Aff. Focus, Hurdles on the Scientific Path to a Biogeneric Approval, 2005;10:43.
18. Rosenberg A, Worobec A. Risk-based approach to immunogenicity concerns of therapeutic protein products. Part 3: Effects of manufacturing changes in immunogenicity and the utility of animal immunogenicity studies. *BioPharm Int* 2005; 17:34–42.
19. Guidance for Industry INDs—Approaches to Complying with CGMP during Phase 1, 2006. http://www.fda.gov/CDER/guidance/6164dft.htm
20. Guidance for Industry, Investigators, and Reviewers Exploratory IND Studies, 2006. http://www.fda.gov/cder/guidance/7086fnl.htm
21. Specifications: Test Procedures and Acceptance criteria for Biotechnological/Biological Products, Q6B, 1999. http://www.ich.org/LOB/media/MEDIA432.pdf
22. Comparability Studies for Human Plasma-Derived Therapeutics, Workshop transcript, 2002. http://www.emea.europa.eu/pdfs/general/direct/emeaar/EMEA_Annual_Report_2006_full.pdfhttp://www.fda.gov/cber/minutes/plasma053002.htm
23. Bunn HF. Drug-induced autoimmune red-cell aplasia. *N Engl J Med* 2002;346: 522–3.
24. Casadevall N, Nataf J, Viron B, Kolta A, Kiladjian JJ, Martin-Dupont P, Michaud P, Papo T, Ugo V, Teyssandier I, Varet B, Mayeux P. Pure red-cell aplasia and antierythropoietin antibodies in patients treated with recombinant erythropoietin. *N Engl J Med* 2002;346:469–75.
25. Gershon SK, Luksenburg H, Cote TR, Braun MM. Pure red-cell aplasia and recombinant erythropoietin. *N Engl J Med* 2002;346:1584–6.
26. Zehnder JL, Leung LL. Development of antibodies to thrombin and factor V with recurrent bleeding in a patient exposed to topical bovine thrombin 1990;76: 2011–6.

27. Rapaport SI, Zivelin A, Minow RA, et al. Clinical significance of antibodies to bovine and human thrombin and factor V after surgical use of bovine thrombin. *Am J Clin Pathol* 1992;97:84–91.
28. Lawson JH, Lynn KA, Vanmatre RM, et al. Antihuman factor V antibodies after use of relatively pure bovine thrombin. *Ann Thorac Surg* 2005;79:1037–8.
29. Josic D, Buchacher A, Kannicht C, et al. Degradation products of factor VIII which can lead to increased immunogenicity. *Vox Sang* 1995;77:90.
30. Rosendaial F, Nieuwenhuis H, van den Berg H, et al. A sudden increase in factor VIII inhibiter development in multitransfused hemophilia A patients in the Netherlands. Dutch Hemophilia study group. *Blood* 1993;81:2180–6.
31. Draft Guidance for Industry: Minimally Manipulated, Unrelated, Allogeneic Placental/Umbilical Cord Blood Intended for Hematopoietic Reconstitution in Patients with Hematological Malignancies, 2006. http://www.fda.gov/cber/gdlns/cordbld.pdf
32. Changes to an Approved Application for Specified Biotechnology and Specified Synthetic Biological Products, 1997. http://www.fda.gov/cber/gdlns/chbiosyn.pdf
33. Guidance for Industry Comparability Protocols—Protein Drug Products and Biological Products–Chemistry, Manufacturing, and Controls Information, 2003. http://www.fda.gov/cber/gdlns/protcmc.pdf.
34. Guidance for Industry Comparability Protocols—Chemistry, Manufacturing, and Controls Information, 2003. http://www.fda.gov/cber/gdlns/cmprprot.pdf

# Selection of Relevant Species

MEENA SUBRAMANYAM, PhD, NICOLA RINALDI, PhD,
ELISABETH MERTSCHING, PhD, and DAVID HUTTO, PhD, DVM

**Contents**

1 Introduction
  1.1 Species Selection: Biologics versus Small Molecule Therapies
  1.2 General Considerations for Relevant Species Selection
2 In vivo Pharmacodynamic Effects
3 Ex vivo Biological Effect
  3.1 Fixed Endpoint Assays
  3.2 Signaling Assays
4 Presence and Distribution of Receptor in Predicted Tissues
5 Ex vivo / In vitro Interaction of the Biopharmaceutical with Known Target
  5.1 Binding Assays: Flow Cytometry Based Methods
  5.2 In vitro Binding Affinity Determination
6 Transcript Profiling
7 In silico Analysis
  7.1 Sequence Homology of Target Protein in Various Species
  7.2 Utilization of Microarray Data
8 Conclusions
References

## 1 INTRODUCTION

The goal of biopharmaceutical development is to maximize therapeutic benefit while minimizing the risk of treatment-related toxicity. To mimic putative interpatient treatment differences in test article responsiveness, it is important

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

not only to select a relevant species for conducting safety assessment studies but also to understand the rank order pharmacologic sensitivity of the relevant common laboratory animal species. The ICH S6 guidance defines a relevant species as "one in which the test material is pharmacologically active due to the expression of the receptor or an epitope (in the case of monoclonal antibodies)." The guidance discourages conduct of studies in nonrelevant species because of concerns about generation of misleading information. Animal studies are critical to demonstrate cross-reactivity of the biopharmaceutical not only with target tissues but with nontarget tissues as well, which in turn facilitates risk assessment in humans. Appropriate selection of a pharmacologically relevant species that is sensitive to administration of the biopharmaceutical in question will enable identification of factors that most reproducibly affect the therapeutic index. A thorough evidence-based evaluation of the species selection criteria must therefore be performed prior to conducting toxicity studies. The feasibility of conducting such studies in two relevant species as per current regulatory requirement should also be evaluated.

This chapter will discuss various experimental approaches used to select the relevant species for conduct of toxicology studies for biopharmaceuticals, as well as highlight advances made in scientific approaches and technologies to facilitate this process. Methods discussed include the traditional immunohistochemistry and tissue cross-reactivity studies, flow cytometry, protein sequencing, and functional in vitro assays, as well as newer approaches such as utilization of microarray databases for genomic mRNA expression data and use of transcript profiling studies as an adjunct to functional assays, to understand similarity in pharmacological responsiveness between animals and humans.

## 1.1 Species Selection: Biologics versus Small Molecule Therapies

The rationale and experimental means by which relevant and appropriate species are selected for preclinical safety evaluations are different when comparing biopharmaceuticals with small molecular, chemically synthesized therapeutics. Unlike protein therapeutics that remain in the extracellular space, small molecular entities may be widely distributed within the biophase, may accumulate intracellularly, and may be associated with "off-target" toxicities, that is, toxicities not associated with or attributable to the targeted receptor or biochemical pathway. Moreover these molecules are frequently subject to metabolic biotransformation to other chemical entities either through enzymatically mediated chemical reactions (phase 1 metabolism) or conjugation with other biomolecules (phase 2 metabolism) that may alter the distribution, excretion and, importantly, toxicity of the metabolite. Biodistributive and metabolic effects on a given small molecule may vary widely among animal species, including humans. Therefore the relevance of candidate animal species for preclinical safety evaluation of small molecular therapeutics is based primarily on comparisons of the so-called metabolite profile of a particular chemical drug in a range of the common rodent and nonrodent laboratory animals. The question being asked in that instance is, which species most closely resem-

bles humans with regard to the identity, number, and quantity of metabolites produced when the chemical therapeutic is added to an ex vivo or in vivo system containing the necessary metabolic components? In contrast, protein-based therapeutics remain in the extracellular space and are not metabolized, and generally induce toxicities via their known mechanism of action, so-called exaggerated pharmacology. Thus selection of species for preclinical safety evaluation of biopharmaceuticals is based on demonstration of the pharmacologic effect in that animal species. This can be done in a variety of ways, as discussed in detail below.

Even if a species is considered relevant based on pharmacological activity of the test article, the many limitations of using animals to predict toxicity of biopharmaceuticals in humans should be recognized. These limitations include variability in the expression pattern of the target, inherent differences in protein processing and clearing mechanisms in animals, differences in immune system development and phenotypes of immune cells, as well as the potential immunogenicity of biopharmaceuticals. By designing appropriate studies and experiments to identify the relevant species for conduct of toxicology studies, the risk of missing a major safety signal can be mitigated. The approaches and examples described in this chapter to illustrate the use of specific methodologies in the selection of a relevant species were selected from a diverse pool of experiments performed during the nonclinical development of either immunoglobulin fusion protein therapeutics or humanized monoclonal antibody therapeutics.

## 1.2 General Considerations for Relevant Species Selection

Numerous experimental approaches can be employed to test the hypothesis that a given laboratory animal species is relevant to humans with respect to a biopharmaceutical and its known target and mechanism of action. While the number of available evaluation methods can be expected to increase with the discovery and deployment of new technologies, the current commonly used methods comprise a relatively short list that can be rank ordered based on the ability of these methods to generate scientifically compelling data. Figure 1 depicts a proposed impact-based rank ordering of common methods. Many of these methods will be discussed in greater detail in subsequent sections of this chapter.

In general, it is reasonable to conclude that the data sets considered to be most compelling are those derived from relatively intact biological systems (whole animal- or cellular-based systems) while less compelling data could be expected to be produced through the analysis of isolated molecules (in vitro assays or in silico analyses). Further, even though a particular method may provide superior and desired data regarding species relevance, all therapeutics may not be amenable for evaluation using a given method. For example, not all biopharmaceuticals elicit measurable changes in a qualified pharmacodynamic marker, nor can a given biopharmaceutical be expected to have the inherent biochemical properties needed to perform as a useful assay reagent.

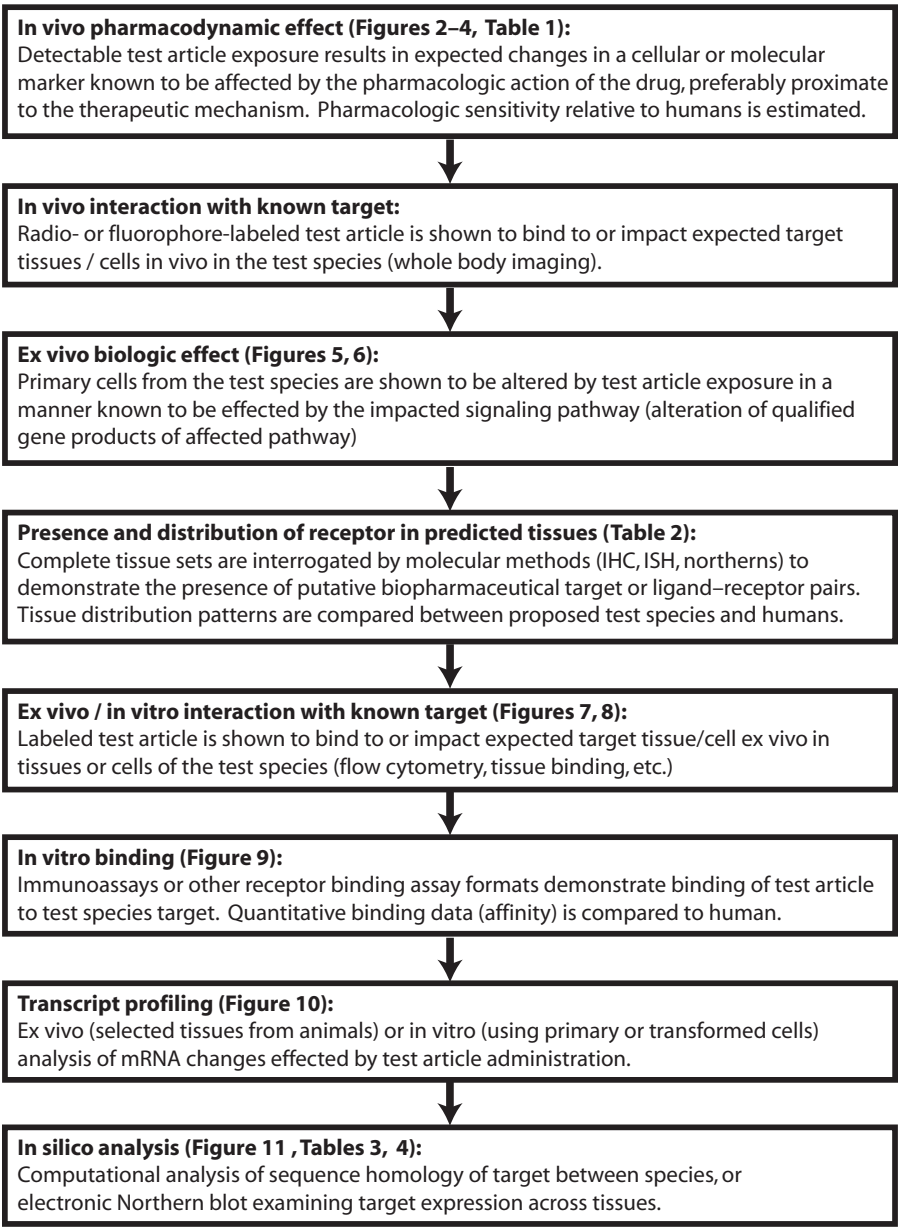


**Figure 1** Proposed rank ordering of methods informing species selection for safety assessment of biopharmaceuticals. Various methods used for selecting pharmacologically relevant species for toxicological studies of biopharmaceuticals are presented, ordered (*top to bottom*) by the extent to which the data might impact the decision on which species to use. In cases where the methods are further discussed in this chapter, the relevant figure/table numbers are provided. These types of analyses may also be used for creating data packages for small molecules, although not typically for species selection.

As in all other areas of scientific inquiry, if one can interpret collected data, there is often value added in the collection of relevant data beyond that considered sufficiently compelling. It is also within the realm of possibility that the described methods for assessing species relevance may provide conflicting data. In such an instance, preference would be given to data derived from the higher order, intact biopharmaceutical systems.

The remainder of this chapter discusses the various methods in the order presented in Figure 1. However, this may not necessarily be the order in which the experiments are performed, as in silico and in vitro assays for an initial examination of species relevance may be easier and less costly to perform, and therefore worth completing prior to the more time-consuming and expensive in vivo assessments.

## 2 IN VIVO PHARMACODYNAMIC EFFECTS

The strongest line of evidence for selecting a pharmacologically responsive species will be the demonstration of similar pharmacodynamic effect across species, utilizing molecular markers reflective of the test article response. In cases where there is a known molecular marker of the test article effect in humans, illustration of a similar change in that marker in nonclinical species under consideration lends confidence to the selection of that species. This is particularly true if the marker is related to the mechanism of action (MOA). For novel molecules, markers of effect on the pathway could serve to demonstrate the desired effect (or lack thereof) of the novel molecule in species under evaluation for relevance. Apart from serving as reporters of the biopharmaceutical's activity to aid in species selection, pharmacodynamic markers also may assist in optimizing dose and frequency of administration, and in the understanding of the relationship between exposure (pharmacokinetics) and efficacy or safety.

One example in which clinical markers of exposure to a specific cytokine have been well studied is the upregulation of interferon-responsive genes and proteins upon administration of interferon-β (IFNβ). Clinical studies have demonstrated that serum neopterin levels (among others) are induced three- to fivefold by administration of interferon-β preparations, with a peak level reached at 24 to 48 hours after treatment [1,2,3]. A similar upregulation, in terms of both fold induction and timing of the induction, has also been observed in rhesus macaques (Figure 2). Neopterin, a catabolic product of guanosine triphosphate, is synthesized by macrophages upon stimulation by interferons, and is a marker of activation of cellular immunity. Based on the in vivo pharmacodynamic activity data, the rhesus monkey is considered an appropriate pharmacologically responsive species for toxicological evaluation of IFNβ.

Another molecular marker reflective of the pharmacological mode of the biopharmaceutical's action, and one that translated from rodent models to nonhuman primates and humans, was the reduction in the level of circulating

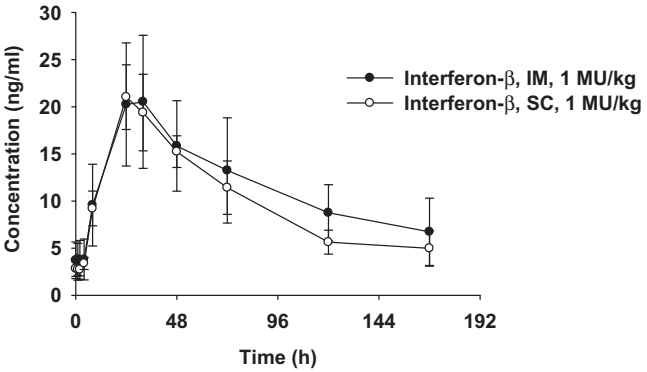


**Figure 2** Induction of neopterin by administration of interferon-β to rhesus macaques. Mean neopterin concentration after a single dose administration of interferon-β in rhesus monkeys. Concentration over time profile plotted from animals treated IM (*closed circles*) or SC (*open circles*). MU = megaunits (=5 mcg/kg)

peripheral blood lymphocytes after administration of an immunoglobulin fusion protein. The fusion protein was designed to bind to a cell surface receptor expressed on lymphocytes and co-engage $Fc_\gamma$ receptors expressed on antigen-presenting cells. The proposed mode of action of the therapeutic was the lysis of test article-bound target cells through the mediation of antibody-dependent cellular cytotoxicity like reaction by $Fc_\gamma$ receptor-expressing cells.

In a chronic toxicity study conducted in 36 naive cynomolgus monkeys, the test article was administered weekly via intravenous injection for 12 months. The study comprised one control group and two test article-treated groups. Peripheral blood from the study animals was subject to flow cytometry analysis to determine the lymphocyte levels (counts/μl) in dosed animals compared to baseline. The percent change from pre-dose baseline and range for absolute lymphocytes and T cell subsets at week 52 are shown in Table 1. The absolute lymphocyte counts as well as counts of T cell subsets including $CD2^+$, $CD3^+$, $CD4^+$, and $CD8^+$ T cells were reduced in a dose-dependent manner following the administration of the test article, confirming test article exposure as well as pharmacologic activity.

The effect of a single course of weekly test article or placebo treatment was then studied in the patient population for a period of 12 weeks. The data on circulating lymphocyte levels were aggregated for the treatment group and compared to the placebo. Both groups had comparable total lymphocyte counts at baseline. There was approximately 39% reduction in total lymphocyte counts in the test article treated group over the course of treatment, although total lymphocyte count remained above the lower limit of normal throughout this time. The placebo group showed a stable profile over time (Figure 3).

**TABLE 1 Mean reductions (range) in lymphocytes and T cell subsets in cynomolgus monkeys**

| Dose Group | Total Lymphs (%) | $CD2^+$ T Cells (%) | $CD3^+$ T Cells (%) | $CD4^+$ T Cells (%) | $CD8^+$ T Cells (%) |
|---|---|---|---|---|---|
| Saline | +34 (–6 to +187) | +38 (–3 to +163) | +43 (–12 to +176) | +42 (–15 to +175) | +48 (–8 to +171) |
| Test article dose level-1 | –21 (–69 to +35) | –27 (–80 to +69) | –35 (–82 to +66) | –40 (–85 to +52) | –28 (–80 to +88) |
| Test article dose level-2 | –46 (–66 to –15) | –63 (–79 to –41) | –75 (–94 to –31) | –85 (–98 to –34) | –64 (–89 to –27) |

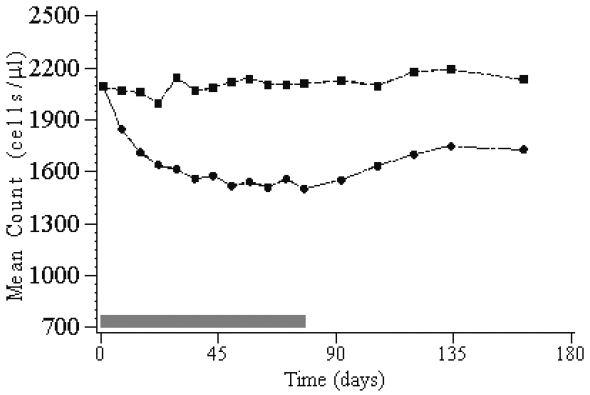


**Figure 3** Mean total lymphocyte count after one course of treatment. Peripheral blood was collected from individuals administered test article (circles) or placebo (squares) on a weekly basis, and subject to flow cytometry analysis to determine mean lymphocyte counts. A standard panel of fluorochrome-conjugated antibodies was used to identify the various lymphocyte sub populations. The solid bar indicates the dosing interval.

Figure 4*a* and 4*b* provides a detailed analysis of the test article's effect on the CD4 and CD8 T lymphocyte populations, respectively. The CD4 counts showed a 47% reduction from baseline in the treatment group (circles, Figure 4*a*) and the CD8 counts showed a 53% reduction from baseline in the treatment group (circles, Figure 4*b*). The placebo group (squares, Figure 4*a* and 4*b*) showed a stable profile over time.

Overall, the pharmacodynamic effects of the test article exposure were qualitatively consistent between the cynomolgus monkeys and humans, confirming the similarity in the pharmacological activity of the test article.

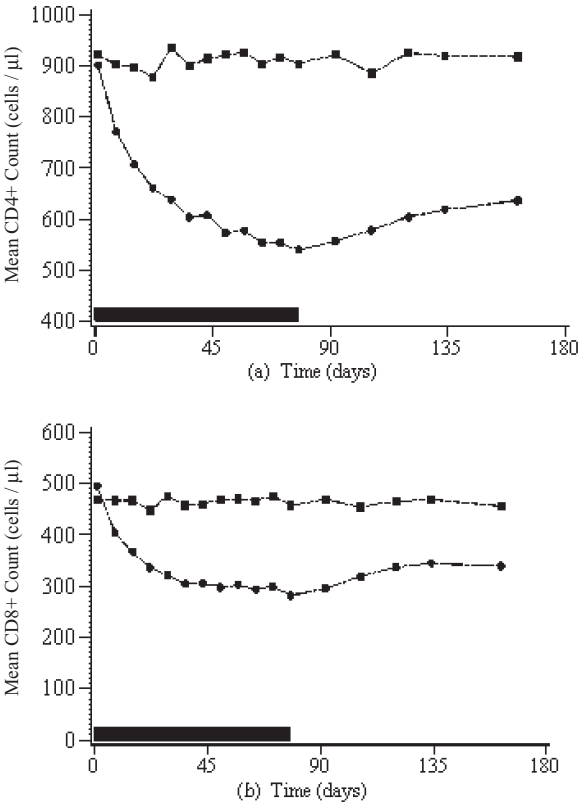


**Figure 4** Mean T cell count after one course of treatment. Peripheral blood was collected from individuals administered test article (*circles*) or placebo (*squares*) on a weekly basis and stained for the presence of CD4 (*a*) or CD8 (*b*) lymphocytes using specific antibodies conjugated to fluorochromes. Samples were then analyzed on the flow cytometer to determine mean relative counts of each subpopulation of cells. The solid bar indicates the dosing interval.

## 3 EX VIVO BIOLOGICAL EFFECT

The practice of demonstrating functional activity of the therapeutic in the selected species is critical, as binding to the desired target does not always translate to functional receptor activation. Since the toxicity observed with biopharmaceuticals is most likely due to exaggerated pharmacology, it is necessary to confirm that the downstream effects observed upon binding of the test article to the target proteins are similar between humans and the selected species. For example, if the test article stimulates activation of subsets of T lymphocytes in humans with a specific activation profile, it is important that this function is conserved in the species selected for performing toxicology

studies. Similarly a test article such as an immunoglobulin fusion protein may mediate antibody-dependent cellular cytotoxicity (ADCC) or complement-dependent cytotoxicity (CDC). Such activities can impact the overall toxicology assessments as they may be integrally linked to the MOA of the therapeutic. For well-characterized targets, information on their function in different species may be readily available in the literature. If the target protein is novel, however, its function will have to be characterized in pharmacology studies to determine the key biochemical drivers for test article activity.

### 3.1 Fixed Endpoint Assays

Endpoint assays such as proliferation or cytotoxicity assays are routinely used for functional assessments. For these assessments, primary cells, transformed cells, or cells transfected with the target receptor are exposed to range of concentrations of the test article. Proliferation or cytoxicity is then measured using a variety of methods such as crystal violet vital dye staining, MTT/MTS incorporation, or a luminescence readout like ATP lite. In addition, assays that analyze phosphorylation of specific transcription factors, or release of specific cytokines and chemokines, are also common. Figure 5 illustrates the measure of functional consequences of receptor–test article interaction by quantifying cytokine release. Cells from the species under evaluation were cultured in the presence of serial dilutions of the test article or control reagents, and supernatants harvested for determination of cytokine levels by ELISA (i.e.,

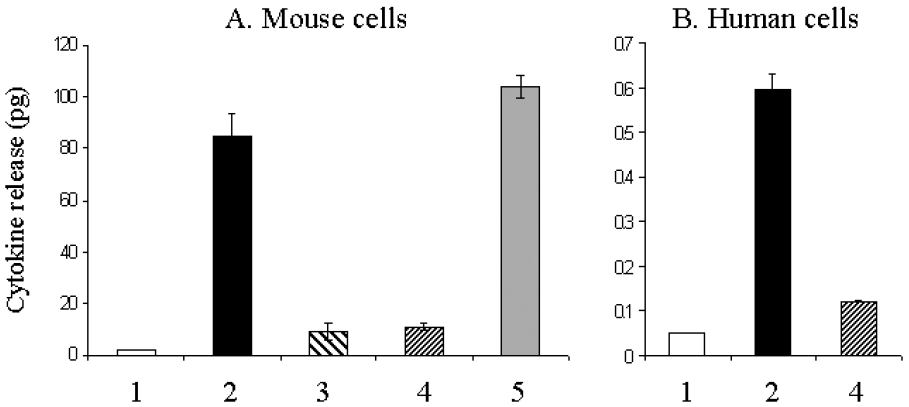


**Figure 5** Inhibition of cytokine release in mouse and human cells. Mouse cells (*a*) and human cells (*b*) were stimulated in vitro with an antireceptor antibody and cytokine release was measured in the supernatant (in pg per million cells). Cytokine release was quantified in the supernatants of unactivated cells (bar 1), activated cells without biopharmaceutical Y (bar 2), and activated cells incubated with biopharmaceutical Y at 5 μg/ml (bar 3) or 10 μg/ml (bar 4). As a negative control, cells were exposed to an irrelevant fusion protein (bar 5).

enzyme-linked immunosorbent assay). Human and mouse cells were activated by cross-linking the receptor of interest using a specific antibody in the presence or absence of the test article, biopharmaceutical Y. Spontaneous cytokine release from unactivated cells was very low (Figure 5*a* and 5*b*, bar 1). In the absence of biopharmaceutical Y, stimulation of mouse and human cells resulted in the release of an expected cytokine (Figure 5*a* and 5*b*, bar 2). In the presence of biopharmaceutical Y, the production of the cytokine was strongly inhibited in both human and mouse cells (Figure 5*a*, bar 3 and 4; Figure 5*b*, bar 4). Release of the cytokine, however, was not inhibited when an irrelevant protein was used (Figure 5*a*, bar 5). Although the amounts of cytokine released by the human and mouse cells were substantially different, the inhibitory effect of biopharmaceutical Y was similar between the cell types. These results indicated that the test article was equally effective in blocking cytokine release mediated by target receptor activation in vitro in mouse and human cells.

### 3.2 Signaling Assays

Another approach to studying the similarities and differences in the pharmacology of the test article in species of interest involves the characterization of the signaling cascade downstream of the protein of interest. Signaling experiments, albeit more challenging, have the advantage of providing more clarity on the putative MOA of the test article. Binding of the test article to the target receptor may result in activation of kinases or other key transcription factors. In the example shown in Figure 6, the ability of biopharmaceutical Y to block signaling mediated through the target receptor was studied. Activation of the

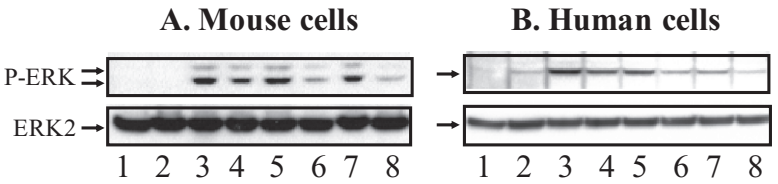


**Figure 6** Effect of a fusion protein on ERK phosphorylation in activated mouse and human cells. Mouse bone marrow cells (*a*) and human cord blood cells (*b*) were activated with antireceptor Ab for 5, 10, and 15 minutes, lysed, and the total proteins loaded on a SDS-Page gel. After transfer to a membrane, ERK phosphorylation was detected using antibodies specific for the phosphorylated (active, P-ERK) form of ERK (upper blots). As a control for the amount of ERK in the samples, antibodies specific for total ERK (activated and nonactivated forms) were used (lower blots). Unstimulated cells untreated (lane 1) and treated with biopharmaceutical Y (lane 2) were used as controls. Cells were activated for 5 minutes (lanes 3, 4), 10 minutes (lanes 5, 6), or 15 minutes (lanes 7, 8) in the absence (lanes 3, 5, 7) or presence (lanes 4, 6, 8) of biopharmaceutical Y.

kinase protein ERK by phosphorylation was determined after stimulation of mouse and human cells. Presence of phosphorylated ERK and total ERK (phosphorylated and unphosphorylated) was detected using specific antibodies. As shown below, ERK phosphorylation was low in unstimulated mouse and human cells (Figure 6*a* and 6*b*, lane 1). Upon activation with antireceptor antibodies in the absence of biopharmaceutical Y, ERK phosphorylation was induced (Figure 6*a* and 6*b*, lanes 3, 5, 7). Pre-incubation of cells with the test article decreased activation-mediated phosphorylation of ERK at all time points in mouse and human cells (Figure 6*a* and 6*b*, lanes 4, 6, 8 compared to lanes 3, 5, 7 respectively).

Collectively, the results from the functional experiments illustrated in Figures 5 and 6 suggested that from a pharmacological perspective, the mouse was a relevant species for evaluating the toxicity of the test article.

## 4 PRESENCE AND DISTRIBUTION OF RECEPTOR IN PREDICTED TISSUES

In the area of relevant species identification, the objectives of immunohistochemistry and tissue-binding studies (often referred to as tissue cross-reactivity studies) are to evaluate the relevance of a given species for use in toxicity studies with a biopharmaceutical and to identify expected and unexpected tissue binding of therapeutics in human and animal tissues.

Immunohistochemical methods can be used to evaluate the tissue distribution of the biopharmaceutical's target(s) or the tissue distribution of other relevant molecular components of a targeted biochemical pathway. A comparison of the distribution of these molecules between possible test species and humans often provides valuable information on the similarities or differences of tissue expression of molecules targeted by the test article between the queried species. The general expectation is to perform tissue cross-reactivity studies prior to human exposure to the new therapeutic. The purpose of these studies is to demonstrate, to the extent possible, what tissues and cells the intact biopharmaceutical binds to. Since monoclonal antibodies and other molecules containing an Fc region or additional binding sites may bind to more than one target, these experiments report localization of the test article to a tissue or cell irrespective of the known or expected binding mechanism. For therapeutic antibodies, the FDA recommends a comprehensive tissue cross-reactivity evaluation as defined in Points to Consider in the Manufacture and Testing of Monoclonal Antibody Products for Human Use (Docket No. 94D-0259). This document recommends conducting immunohistochemistry analysis at two antibody dilutions across triplicate specimens of approximately 32 frozen human tissue types (collected from three unrelated donors) for the therapeutic antibody and its isotype control. An analogous study is conducted in parallel in tissues from various animal species (two or three unrelated donors) for additional justification of the relevance of the toxicity species.

While this has been standard practice for a number of years, product developers have recently started questioning the relevance and utility of these types of studies for predicting human safety.

When test articles are chimeric, humanized, or fully human antibodies, three different staining methods can be used. These include avidin-biotin complex (when the test article is biotinylated), direct labeling of test article with fluorochromes such as fluorescein isothiocyanate (FITC) or rhodamine, or precomplexing the unlabeled test article with a labeled antihuman IgG. The latter method is most often used, as it allows secondary signal amplification by a variety of methods and, by precomplexing, avoids nonspecific binding of the secondary antihuman antibody to human antibodies that are present in the evaluated human tissue sections. It is also performed in instances where conjugation of biotin or a fluorochrome onto the test article could alter its affinity for the protein of interest. Even though this additional step often increases the backgound staining, the indirect approach has the advantage of amplifying the signal, and is therefore valuable when expression of the target protein is expected to be relatively low in tissues of interest.

For all test articles an isotype control antibody labeled in a similar fashion must be included to demonstrate specificity of staining observed with the test article. As with any assay procedure, both positive (containing cell types that express the target at high levels) and negative control tissues should be stained to assess extent of nonspecific binding under standardized immunohistochemical conditions. In general, the positive control may be a tissue element or cell line transfected with or known to express the target antigen, sepharose or agarose beads coated with the target antigen, or the target antigen spotted and cross-linked onto UV-resin slides. During optimization of the staining procedure the reactions of the test article with the positive control but not the negative control, as well as the lack of reactivity of the isotype control antibody, should be confirmed, to demonstrate the sensitivity, specificity, and reproducibility of the assay. A board-certified anatomical pathologist usually performs analysis of the immunolabeled tissues using a scoring system that assesses the identity and relative proportion of cells staining within a given immunopositive tissue, the relative intensity of staining, and the location (cytoplasm vs. cell surface) of staining. The slides are evaluated for integrity of the tissue sample as well as for specificity of the staining seen in the tissues based on knowledge of expression of the target antigen in question. For the reasons mentioned above, a biopharmaceutical known to have acceptable potency and binding affinity to its target may not perform well in a tissue-binding study. There are numerous examples of highly potent monoclonal antibodies that have been completely negative in tissue-binding studies, due to their poor performance as immunohistochemical reagents.

An example of results from a tissue cross-reactivity study comparing binding of a monoclonal antibody therapeutic to human, cynomolgus monkey, and mouse tissues is shown in Table 2. A cell line that did not express the target was used as negative control tissue; the same cell line transfected with the

**TABLE 2 Cross-species tissue cross-reactivity study of a monoclonal antibody**

| Tissue | Cell Type | Human | Cynomolgus Monkey | Mouse |
|---|---|---|---|---|
| Urinary bladder | Urothelium | 2 (3–4) | 3 (3) | NS |
| Ureter | Urothelium | 2 (3–4) | 3 (2–4) | NA |
| Tonsil | Mucosal epithelium | 3 (2–3) | 3 (2–3) | NA |
| Uterus-cervix | Mucosa | 3 (1–3) | 1 (2) | NA |
| Eye | Corneal epithelium | 1 (1) | NS | 1 (1)[a] |
| Breast | Glandular epithelium | 3 (1–3) | 3 (1–3) | NS |
| Fallopian tube | Tubular epithelium | 2 (1–2) | 1 (3) | NS |
| Kidney | Tubular epithelium–cortex | 3 (3) | 1 (3)[b] | 3 (1–2) |
| Lung | Alveolar epithelium | 3 (2–3) | 2 (3) | NS |
| Pancreas | Ductular epithelium | 3 (2) | 2 (1–2) | 2 (1) |
| Prostate | Glandular epithelium | 2 (1–2) | 1 (2) | 3 (2)[c] |
| Thyroid | Follicular epithelium | 3 (2–3) | 2 (2) | NS |
| Uterus-endometrium | Endometrial mucosa | 2 (2) | NS | 3 (1–3)[d] |
| Adrenal | Cortical epithelium | 3 (1–2) | 2 (1) | NS |
| Pituitary | Adenohypophysis epithelium | 3 (2) | NS | NS |
| Liver | Sinusoidal mesenchymal cells | 3 (2–4) | 3 (2–3) | NS |

*Note*: Cryosections of normal human, cynomolgus monkey, and mouse tissues listed (in addition to others) were obtained and stained with biotinylated test article. Data are shown for the optimal antibody concentration; the second concentration used was 5× higher (data not shown). The cell types that demonstrated staining are listed, along with the number (out of three) of sections that showed staining, with the staining intensity range in parentheses. 1 = Light staining or occasional cells stained—minimal. 2 = Light-medium staining and/or small numbers of cells/types of cells labeled—mild. 3 = Moderate staining and/or medium numbers of cells/types of cells labeled—moderate. 4 = Dark staining and/or large numbers of cells/types of cells labeled—marked. NS: No staining seen. NA: Tissue section not obtained.

[a]Lens fiber stained, not corneal epithelium.

[b]Tubular epithelium staining in papilla, not cortex.

[c]Tubular epithelium and intertubal stroma showed staining.

[d]Smooth muscle cells were also stained.

target was used as a positive tissue. Tissues were sectioned, frozen, and treated with avidin, biotin, and nonspecific IgG to block endogenous binding sites. Two concentrations of biotinylated test article (or control antibody) were applied to the tissue sections, followed by streptavidin HRP and substrate. After drying, the slides were read by a pathologist. Adequacy of tissues for staining was validated using an anti-CD31 (platelet endothelial cell adhesion molecule) antibody as positive control, and an isotype matched irrelevant antibody as a negative control. Semiquantitative results are shown in Table 2. Qualitatively, the pathologist determined that based on the staining intensity and localization of staining, the cynomolgus monkey, and human tissue samples showed a high degree of similarity in binding the test article, whereas only a few of the murine tissues were stained. Therefore a murine version of the antibody was

used for mouse efficacy and toxicity studies, but the humanized antibody was used in the toxicology studies in the cynomolgus monkey.

## 5 EX VIVO / IN VITRO INTERACTION OF THE BIOPHARMACEUTICAL WITH KNOWN TARGET

To demonstrate functional pharmacological similarities between humans and the animal species under evaluation, biological assay systems are routinely utilized. In some instances primary cells expressing the target are utilized in these functional experiments. In most cases, however, cell lines containing the appropriate signaling machinery are transfected with the related cell-surface receptor from the test species. Specificity of interaction of test article to target is demonstrated initially through binding assays and then extended to other qualified endpoints such as phosphorylation or nuclear translocation of relevant proteins in the signaling cascade.

### 5.1 Binding Assays: Flow Cytometry Based Methods

Flow cytometry based assays are routinely utilized to measure binding of biotherapeutics to various target cells (peripheral blood lymphocytes, bone marrow derived cells) from multiple species. This analysis can provide insight into the types of cells that the biopharmaceutical can bind, as well as the relative number of receptors on the cells. In instances where the target acts by dimerizing or trimerizing with signaling partners, it is important to determine whether the affinity of binding and the binding partners are similar between species. Flow-based methods require appropriate positive and negative controls, as well as an isotype control or irrelevant protein control to rule out nonspecific binding. Besides primary cells, binding assays can also be performed with cells transfected with proteins of interest or with cultured commercially available cells from various species, provided that the target receptor is expressed on the cell types.

The flow cytometry experiments described below outline a series of experiments conducted with primary cells collected from various species to determine the ability of a human biopharmaceutical (X) to bind in vitro to the target protein in these species. Figure 7 illustrates a flow cytometry assay conducted using murine bone marrow derived cells expressing the murine homologue of the biopharmaceutical's target receptor. The ligand for the target human receptor was used as a positive control in the assay, as an antibody to directly stain the murine homologue was not available. Fluorochrome conjugated secondary antibodies specific for the test article were used as detection agents. As shown in Figure 7*a*, biopharmaceutical X did not bind to the receptor expressed on mouse bone marrow derived cells. The human ligand also did not bind to the mouse receptor. Figure 7*b* illustrates an independent experiment performed using the murine version of biopharmaceuti-

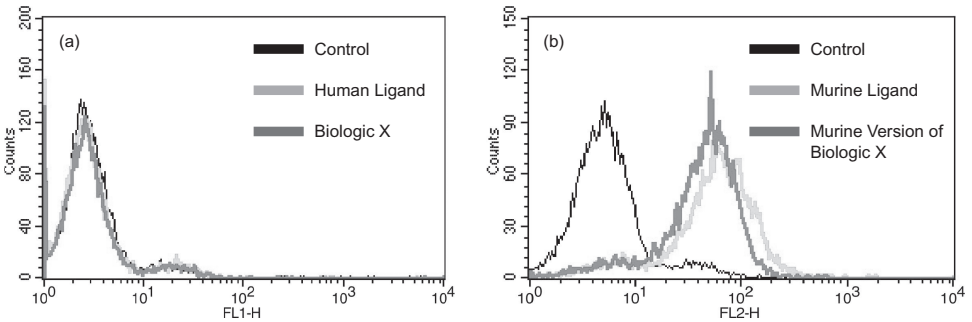


**Figure 7** Evaluation of biopharmaceutical X binding to mouse bone marrow derived cells. (*a*) Biopharmaceutical X binding to mouse bone marrow-derived cells was evaluated using flow cytometry. Mouse cells were incubated with biopharmaceutical X or the human ligand. Secondary antibodies that cross-reacted with both biopharmaceutical X and the natural human ligand were used to detect binding. Cells were stained with the secondary antibodies only, to detect non-specific binding (control). (*b*) Bone marrow derived mouse cells were stained as described above with a murine version of biopharmaceutical X and the murine ligand for the receptor. Secondary antibodies that cross-reacted with the murine version of biopharmaceutical X and the murine ligand were used to detect binding.

cal X and the natural ligand for the mouse receptor as the positive control. As expected, the murine version of biopharmaceutical X and the murine ligand bound to mouse cells, thereby confirming the expression of the target receptor. These results indicated that biopharmaceutical X did not cross-react with murine receptors and therefore the mouse was not a suitable species for conduct of toxicology studies.

Receptor-ligand binding assays with biopharmaceutical X were then performed on cynomolgus monkey peripheral blood cells. Cells were incubated first with the test article, and then with antibodies specific for the target receptor. Binding of biopharmaceutical X did not interfere with binding of the antibody, as the two molecules interacted with distinct epitopes on the target protein. As shown in Figure 8 (left panel), presence of cells in the upper-right quadrant in the plot demonstrated that all the cynomolgus cells that stained with the antireceptor antibody (*X*-axis) had also bound the human ligand (*Y*-axis). Figure 8 (right panel) shows the binding pattern of biopharmaceutical X to the cynomolgus cells. Presence of a large population of cells in the upper-right quadrant indicated that the test article bound to the same subset of cells that stained positively with the antireceptor antibody. The test article, however, also exhibited nonspecific binding to cells that did not show reactivity with the ligand or the antireceptor antibody.

The in vitro binding data shown in Figures 7 and 8 supported the selection of the cynomolgus monkey over mice as a relevant species for conduct of toxicology studies for biopharmaceutical X.

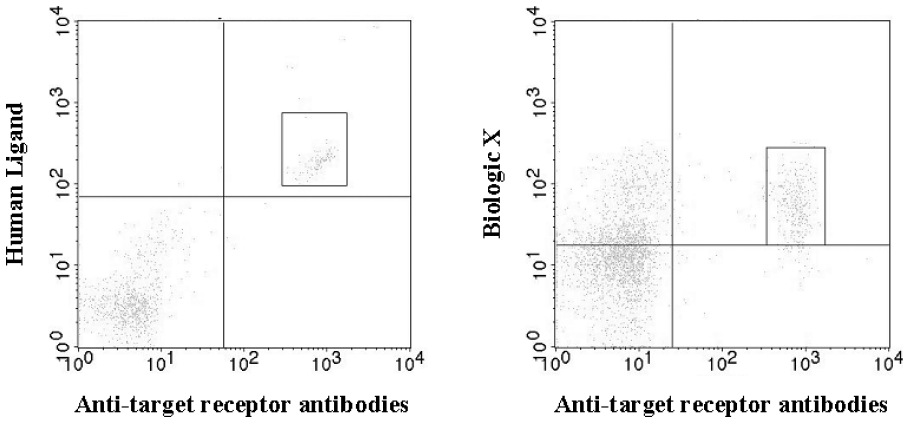


**Figure 8** Evaluation of binding of biopharmaceutical X to cynomolgus monkey peripheral blood cells. Cynomolgus monkey cells were stained with fluorochrome-labeled biopharmaceutical X or the natural ligand for the human receptor. Expression of the target receptor on the cynomolgus cells was detected using antibodies specific for the target receptor.

### 5.2 In vitro Binding Affinity Determination

Besides qualitative determination of binding to the target receptor, more specific dose–response studies and binding affinity measurements can be performed to confirm selection of appropriate species. This determination may be necessary for agonist therapeutics for which the affinity and avidity of binding determine the strength of the signal induced. In the example illustrated in Figure 9, the affinity of binding of a humanized monoclonal antibody (MAb) therapeutic to cultured fibroblasts from humans and cynomolgus monkeys was compared. Various concentrations of the humanized MAb were added to the cells. Background binding was determined using an isotype control antibody.

The functional affinities of the humanized antibody for the target protein on human and cynomolgous cells were computed to be 0.05 and 0.03 nM, respectively. These results indicated that the humanized MAb bound with similar affinity to the target receptor on human and cynomolgus cells.

## 6 TRANSCRIPT PROFILING

Another approach, taking advantage of recent technological developments, is to determine the similarity of the pharmacological responsiveness between species through the conduct of a transcript-profiling study. This will typically be performed using either primary cells or transformed cell lines, with primary cells being preferable due to the greater relevance to the in vivo scenario. Cells

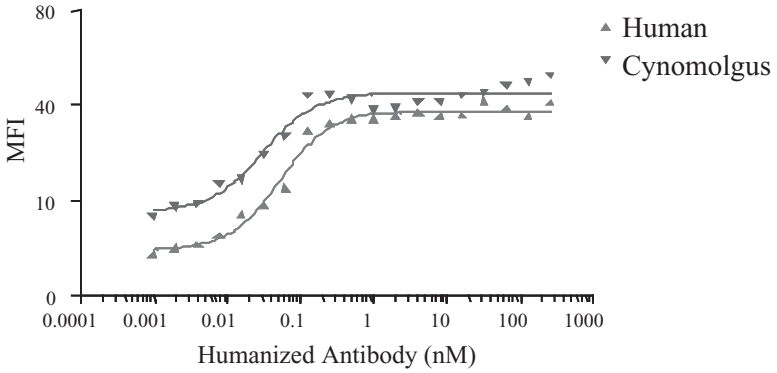


**Figure 9** Affinity of binding of a humanized antibody to cultured human and cynomolgus fibroblast cells. Cultured human (triangles) and cynomolgus fibroblast (upside-down triangles) cells were removed from the flasks by trypsin and washed with PBS. One million cells were incubated with various concentrations of humanized Ab, at 4°C for 2 hours. The cells were washed, and binding was detected with PE-conjugated goat antihuman IgG using a FACSCalibur flow cytometer. The binding of another humanized Ab of the same isotype as the test article, but with irrelevant specificity, was used as an isotype control (data not shown). The mean fluorescence intensities (MFI) for each sample were plotted against the antibody concentration.

from the likely target organ(s) from different species can be exposed to the test article, then mRNA purified from the cells, and gene expression studies performed using microarrays. Commercially available microarrays exist for many of the species commonly used for toxicological studies, including mice, rats, dogs, and rhesus and cynomolgus monkeys. If pharmacological markers are already known from prior research, up- or downregulation of those markers can be confirmed in the additional species being assessed. In addition the overall overlap in gene expression response can be compared between species (taking into account the overlap between genes assayed on the microarray). Ideally the most relevant species would show similarly high up- or downregulation of prior known pharmacological markers, and of additional biologically relevant genes that are significantly changed. Further a high overall gene expression response similarity indicating that analogous pathways are affected across species is desirable. A gene ontology analysis can be performed for each species, where the functions of up- or downregulated genes are evaluated to determine whether, in general, the same biological functions are affected. The changed genes can also be classified as belonging to known biological pathways, to determine whether there are any pathways that are affected in one species and not others.

In the preclinical development of biopharmaceuticals, restricted species cross-reactivity may limit the species in which toxicology evaluations can be

performed. In higher order species such as chimpanzees, very limited safety data can be derived as organ and tissue collection is not feasible for detailed analysis and sufficient numbers of animals may not be available for chronic toxicity testing. In order to support extended clinical dosing, and to provide a more comprehensive safety assessment, a surrogate (or homolog) approach can be used. A surrogate of the therapeutic that interacts with rodent or other nonhuman primate species can be utilized in the conduct of toxicology studies, as long as similarity in in vitro and in vivo pharmacology is demonstrated. In the example described below, a surrogate of a humanized MAb therapeutic was evaluated for pharmacological similarity, due to restricted species cross-reactivity of the therapeutic. A transcript profiling experiment, besides additional studies, was performed, to demonstrate the similarity in the in vitro pharmacological effect of the biopharmaceutical and the surrogate antibody.

Transcript profiling was used to determine whether similar transcripts were induced or repressed when three cell lines (colon carcinoma cell line, primary endothelial cells, and primary fibroblasts) were treated with the MAb (biopharmaceutical Z) or the surrogate antibody. Cells were grown to confluence, and treated with a fixed dose of cytokine plus vehicle, or cytokine plus biopharmaceutical Z, surrogate antibody, nonspecific antibody control, or positive control. The cytokine was used in order to amplify the activation response through the target receptor. Cells were lysed after two hours of treatment, and transcript profiling was performed using the human HG-U133 Plus 2.0 Affymetrix® GeneChips®.

Initial analysis confirmed that the cell lines were responsive to treatment, as the positive control caused up- or downregulation of genes to approximately the same extent across all three cell lines. Biopharmaceutical Z and the surrogate antibody caused specific changes (after the effect of the nonspecific control was removed) to the same extent in the three cell lines, although the effects were not nearly as significant as with the positive control. In addition biopharmaceutical Z was known to upregulate expression of several cytokines; this upregulation was confirmed in the three cell lines examined, and a similar magnitude of upregulation was seen using the surrogate antibody. Next genes that were selected as meeting statistical cutoffs in the biopharmaceutical Z treated samples, and in the surrogate treated samples, were compared between the two treatments (Figure 10). These graphs illustrate that the overall magnitude of intensity and ratios were highly similar between the two treatments. In addition treatment with biopharmaceutical Z (and the surrogate antibody, data not shown) showed similar up- or downregulation as treatment with the positive control, although selected transcripts were upregulated to a greater degree by the positive control. Finally, gene ontology analysis found very similar biological functions affected by biopharmaceutical Z and the surrogate antibody, and differences were overwhelmingly due to transcripts in one treatment or the other barely failing to meet statistical cutoffs.

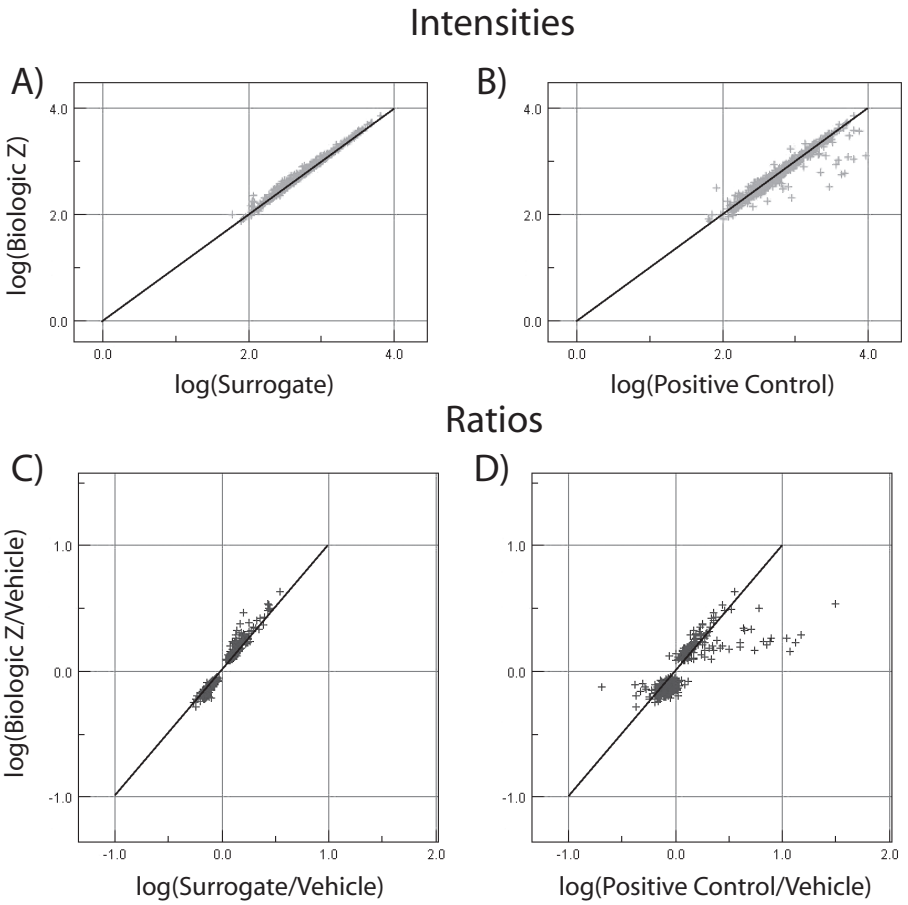


**Figure 10** Comparison of signal intensities and ratios between treatments. Signal intensities (*a, b*) and ratios (*c, d*), comparing transcripts meeting significance cutoffs for biopharmaceutical Z with the surrogate antibody (*a, c*) and the positive control (*b, d*). Identical intensities or ratios will fall on the 1 : 1 diagonal line. The overall pattern of signal intensities and ratios was highly similar between biopharmaceutical Z and the surrogate antibody, with marginally higher intensities/ratios following treatment with biopharmaceutical Z. The positive control upregulated a set of transcripts to a higher degree than biopharmaceutical Z; these fall to the right of the 1:1 diagonal line.

## 7 IN SILICO ANALYSIS

### 7.1 Sequence Homology of Target Protein in Various Species

Another question to address is the likelihood of the biopharmaceutical candidate binding to the target protein in the animal species being examined for relevance. A convenient and commonly used approach to investigate this is to

determine the sequence homology of the target protein between humans and the selected species. When a protein shares a high degree of homology across species, the probability of the test article binding to the protein across species is high. The implicit assumption when sequences are highly similar across species is that there is a shared commonality of function such as similar pharmacological function. Reciprocally, protein sequences in other species that share little or low homology to their human counterpart are less likely to bind a biopharmaceutical candidate targeted to the human protein. However, like all experimental methods, demonstration of sequence identity or similarity between species has limitations, as the relative importance of each amino acid in the sequence is not equivalent and is complicated to assess. For example, amino acids directly involved in the binding of the test article must be conserved across species, as well as those involved in the maintenance of the secondary and tertiary structure of the protein. This method will therefore provide a definitive answer only for proteins with sequences that are totally homologous to the human sequence, and only for those proteins that function as monomers. In all other cases determination of percent sequence homology will facilitate identification of those species that have the highest probability of binding the test article. Follow on in vitro binding assays will be needed to confirm these results. Figure 11 illustrates the protein sequence homology of $Fc_{\varepsilon}RI$-alpha chain across a range of species and Table 3 summarizes the percent sequence homology across species.

The homology of the macaque's (cynomolgus and rhesus) protein sequences to the human sequence was over 90%. In contrast, dog, rat, and mouse sequences shared low homology with the human $Fc_{\varepsilon}RI$-alpha protein. Based on these results, the receptor of nonhuman primates would be more likely to bind a biopharmaceutical designed to interact with the human $Fc_{\varepsilon}RI$-alpha chain. This conclusion, however, would have to be confirmed in additional in vitro binding assays.

### 7.2 Utilization of Microarray Data

In addition to the standard methods for determining the pattern of target expression, the availability of genomic mRNA expression data allows for in silico experiments. Since the advent of microarray technology in the mid-1990s [4,5,6], hundreds of thousands of mRNA samples have been tested on microarrays. Data have been generated from normal and diseased tissues of many species. Many of the datasets associated with published literature are publicly available; in addition private companies provide such data for a fee. The relative level of expression of genes in the targeted pathway, or of the gene targeted by a monoclonal antibody, can then be assessed across multiple tissue types, in diseased tissues, or in disease models. The expression pattern in potential species for preclinical studies can be compared with the expression pattern in human tissues. A metric to select the most relevant species might be difficult to construct; however, species that are clearly not relevant based on lack of

```
           1        10        20        30        40        50        60        70        80        90       100
           |--------+---------+---------+---------+---------+---------+---------+---------+---------+---------|
     Human MAPAMESPTLLCVALLFFAPDGVLAVPQKPKVSLNPPWNRIFKGENVTLTCNGNNFFEVSST-KWFHNGSLSEETNSSLNIVNAKFEDSGEYKCQHQQVN
      Cyno MAPAMESPTLLCVALLFFAPDGVLAVPQKPTVSLNPPWNRIFKGENVTLTCNGSNFFEVSSM-KWFHNGSLSEVANSSWNIVNADFEDSGEYKCQHQQFD
    Rhesus MAPAMESPTLLCVALLFFAPDGVLAVPQKPTVSLNPPWNRIFKGENVTLTCNGSNFFEVSSM-KWFHNGSLSEVANSSLNIVNADFEDSGEYKCQHQQFD
       Dog MPASMGGPALLWLALLLSSPGVMSSDTLKPTVSMNPPWNTILKDDSVTLTCTGNNSLEVDSA-VWLHNNTTWQETTSRLDINKAQIQDSGEYRCRENRSI
       Rat   MDTGGSARLCLALVLISLGVMLTATQKSVVSLDPPWIRILTGDKVTLICNGNNSSQMNST-KWIHNDSISNVKSSHWVIVSATIQDSGKYICQKQGFY
     Mouse   MVTGRSAQLCLALLFMSLDVILTATEKSVLTLDPPWIRIFTGEKVTLSCYGNNHLQMNSTTKWIHNGTVSEVNSSHLVIVSATVQDSGKYICQKQGLF
 Consensus m...mg.palLclALlf.spdv.l..tqKp.vs$#PPWnrIfkg#.VTLtCnGnN..#v.St.kW.HNgs.s#v..S.l.Iv.A..#DSGeY.Cq.#...

           101      110       120       130       140       150       160       170       180       190       200
           |--------+---------+---------+---------+---------+---------+---------+---------+---------+---------|
     Human ESEPVYLEVFSDWLLLQASAEVVMEGQPLFLRCHGWRNWDVYKVIYYKDGEALKYWYENHNISITNATVEDSGTYYCTGKVWQLDYESEPLNITVIKAP-
      Cyno DSEPVHLEVFSDWLLLQASAEVVMEGQPLFLRCHSWRNWDVYKVIYYKDGEALKYWYENHNISITNATVEDSGTYYCTGKLWQLDCESEPLNITVIKAQ-
    Rhesus DSEPVHLEVFSDWLLLQASAEVVMEGQPLFLRCHSWRNWDVYKVIYYKDGEALKYWYENHNISITNATVEDSGTYYCTGKLWQLDCESEPLNITVIKAQ-
       Dog LSDPVYLTVFTEWLILQASANVVMEGESFLIRCHSWKNLRLTKVTYYKDGIPIRYWYENFNISISNVTTKNSGNYSCSGQIQQKGYTSKVLNIIVKKEPT
       Rat KSKPVYLNVMQEWLLLQSSADVVLDNGSFDIRCRSWKKWKVHKVIYYKDDIAFKYSYDSNNISIRKATFNDSGSYHCTGYLNKVECKSDKFSIAVVKDY-
     Mouse KSKPVYLNVTQDWLLLQTSADMILVHGSFDIRCHGWKNWNVRKVIYYRNDHAFNYSYESP-VSIREATLNDSGTYHCKGYLRQVEYESDKFRIAVVKAY-
 Consensus .S.PVyL.Vf.#WLlLQaSA#v!$eg.sf.iRChsWknw.v.KViYYk#g.a.kYwY#n.n!SI.naT..#SGtY.CtG.l.q..yeS..lnI.V.Ka..

           201      210       220       230       240       250       261
           |--------+---------+---------+---------+---------+---------+|
     Human -REKY-WLQFFIPLLVVILFAVDTGLFISTQQQVTFLLKIKRTRKGFRLLNPHPKPNPKNN
      Cyno -HDKY-WLQFLIPLLVAILFAVDTGLFISTQQQVTFLLKIKRTRKGFKLLNPHPKPNPKSN
    Rhesus -HDKY-WLQFLIPLLVAILFAVDTGLFISTQQQVTFLLKIKRTRKGFKLLNPHPKPNPKSN
       Dog KQNKYSGLQFLIPL-VVILFAVDTGLFISTKQQLTYLLQIKRTRKNKK---PEPGKN
       Rat -TIEYRWLQLIFPSLAVILFAVDTGLWFSTHKQFESILKIQKTGKGKKKG
     Mouse -KCKYYWLQLIFPLLVAILFAVDTGLLLSTEEQFKSVLEIQKTGKYKKVETELLT
 Consensus ...kY.wLQf.iPllvvILFAVDTGLfiST.qQ.t.lLkIkrTrKgkk...p.p..n....
```

**Figure 11** Alignment FcεRI-alpha chain protein sequences across species. Human, mouse, rat, dog, cynomolgus macaque, and rhesus macaque protein sequences coding for the alpha chain of the IgE receptor ($Fc_{\varepsilon}RI$) were aligned and compared. Residues that are identical in all six species are highlighted in red. Residues common to more than half the sequences are listed in blue, and residues found to be identical in half the sequences or less are shown in black.

**TABLE 3 FcεRI-alpha protein sequence homology across species**

| Species | FcεRI-alpha Chain Percent Sequence Homology |
|---|---|
| Mouse | 51 |
| Rat | 49 |
| Dog | 55 |
| Cynomolgus macaque | 91 |
| Rhesus macaque | 92 |

expression of a target or pathway in the appropriate organs can be excluded from further consideration.

This technology is well established, but caveats to its use remain. One drawback to overcome is the lack of data on most species. Currently databases of human, mouse, and rat tissues are extensive, but data from additional species are sparse. In addition only selected strains of mice and rats have been profiled. This might require use of those strains for toxicology studies or confirmation of results in the desired strains through additional transcript profiling or quantitative PCR experiments. Another caveat is that microarray results are not quantitative, and cannot be reliably compared in terms of absolute amounts across species because of differences in the sequences printed on the microarray, and differences in hybridization kinetics, among others. Also quantitation at the RNA level does not always correlate with quantification at the protein level. For a truly quantitative comparison, additional experiments would need to be performed. Finally, the computational methods for determining the most similar patterns of expression require further exploration as there is considerable debate on the optimal statistical analysis methods that can be utilized to identify meaningful differences in expression patterns. Moreover reliability of results generated with this approach are dependent on the purity of the tissue sample and quality of the RNA extracted, as well as the probes selected for analysis. Contamination with blood or other cell types is sometimes difficult to avoid and may alter the conclusions. For all the reasons mentioned above, information derived from such databases needs to be confirmed with other techniques.

For example, the expression of EGF was assessed using the Gene Logic® Ascenta® database. The "eNorthern" (electronic Northern blot, to examine mRNA expression across tissue types) function was used to find which tissues expressed EGF, and at approximately what level. The data were collected by Gene Logic® using the Affymetrix® GeneChip® platform, where oligonucleotides spotted on a microarray probe the expression of tens of thousands of mRNAs. Two metrics result from these experiments: a present (absent) call indicating whether the mRNA was detected in the sample, and a signal measurement suggestive of the number of copies of that mRNA present. The results of the assessment of EGF expression are shown in Table 4. EGF was

**TABLE 4 Assessment of expression of EGF across species and strains**

| | Human | | Mouse (129S3) | | Mouse (C57BL6) | | Mouse (DBA) | | Rat (Wistar) | | Rat (Sprague–Daweley) | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Organ | Present | Median Signal | Present | Median Signal | Present | Median Signal | Present | Median Signal | Present | Median Signal | Present | Median Signal |
| Bladder | 0% | 18 | 14% | 5 | 0% | 4 | 14% | 8 | 0% | 12 | 0% | 10 |
| Bones | 50% | 23 | 0% | 37 | 0% | 28 | 13% | 29 | NA | NA | NA | NA |
| Brain | 15%[c] | 19[c] | 0% | 7 | 0% | 10 | 0% | 4 | 6% | 5 | NA | NA |
| Breast | 79% | 54 | 40% | 68 | 14% | 47 | 63% | 80 | NA | NA | NA | NA |
| Colon | 15% | 14 | 0% | 5 | 0% | 32 | 0% | 32 | 83% | 23 | NA | NA |
| Esophagus | 14% | 8 | 100% | 236 | 100% | 286 | 100% | 173 | 0% | 12 | NA | NA |
| Heart | 39% | 36 | 17% | 33 | 0% | 24 | 0% | 10 | 0% | 10 | 0% | 9 |
| Kidney | 94% | 368 | 100%[a] | 2623[a] | 100%[a] | 3616[a] | 100%[a] | 4071[a] | 100%[b] | 474[b] | 100%[a] | 540[a] |
| Liver | 50% | 18 | 13% | 5 | 0% | 6 | 0% | 3 | 22% | 13 | NA | NA |
| Lung | 22% | 14 | 0% | 8 | 0% | 6 | 0% | 5 | 0% | 11 | NA | NA |
| Lymph node | 20% | 12 | 0% | 151 | 0% | 21 | 67% | 114 | NA | NA | NA | NA |
| Ovary | 5% | 10 | 0% | 9 | NA | NA | 0% | 3 | 0% | 12 | NA | NA |
| Pancreas | 96% | 179 | 100% | 488 | 33% | 343 | 100% | 464 | NA | NA | NA | NA |
| Prostate | 47% | 22 | 80% | 326 | 100% | 741 | 100% | 924 | NA | NA | NA | NA |
| Spleen | 12% | 15 | 0% | 11 | 0% | 5 | 0% | 11 | NA | NA | NA | NA |
| Stomach | 4% | 12 | 0% | 25 | 29% | 51 | 0% | 35 | 13% | 11 | NA | NA |
| Testis | 0% | 7 | 0% | 23 | 0% | 12 | 86% | 127 | 0% | 17 | NA | NA |
| Thymus | 3% | 8 | 0% | 5 | 0% | 5 | 0% | 11 | NA | NA | NA | NA |
| Uterus | 14% | 12 | 0% | 5 | 0% | 5 | 0% | 3 | NA | NA | NA | NA |

*Note*: Results from eNortherns of EGF in the Gene Logic® Ascenta® database are shown. The probe sets selected were 206254_at (human), 102774_at (mouse), and U04842_at (rat). The data were exported, and tissues with data across multiple species selected. NA indicates that the given tissue was not profiled in that species/strain. Shown are the percentages of tissues expressing EGF (% present) and the median of the signal for each tissue type (arbitrary units).

[a]Average of left and right kidney samples.

[b]Left kidney only.

[c]Average over multiple brain sections.

consistently present across 94% of kidney samples and 96% of pancreas in the human samples tested, with relatively high signal intensities in both. It was also present in 100% of kidney samples across both mice and rats. EGF was present in only 33% of pancreas samples from the mouse C57BL6 strain, suggesting a difference from humans and possibly that this strain would be less appropriate for use in testing an anti-EGF therapy. The pancreas was not assessed in the rats. Another significant difference across species was that EGF was expressed in the esophagus in the mouse strains but not in the human samples. This may suggest a potential difference in toxicity between the two species.

## 8 CONCLUSIONS

Biopharmaceuticals pose unique challenges to clinical development, wherein exaggerated pharmacology becomes the main driver for toxicities observed. This increases the importance of defining a pharmacologically relevant species for preclinical safety assessment. A variety of experimental approaches and technologies are available to assist in the selection of the relevant species for conduct of toxicology studies or, as stated previously, to test the proposed null hypothesis that a given laboratory animal species is not relevant to humans with respect to a biopharmaceutical and its known target and mechanism of action. These methods for analysis, however, should be selected prudently based on the test article and its pharmacology. An evidence-based approach is needed to rule in or rule out species with the understanding that experimental data from various methods may not always be aligned to rule in the species as biologically and pharmacologically relevant. Appropriate design of these experiments, with adequate controls, will ultimately aid in proper interpretation of the data that will form the basis for the selection of the species for conduct of toxicology studies. The diligence paid during the species selection exercise will facilitate the minimization of risks and, more important, create better awareness of risks, during clinical development of biopharmaceuticals.

## ACKNOWLEDGMENTS

Experimental data discussed in this chapter were derived from Biogen Idec sponsored studies. The authors wish to acknowledge the contributions of many scientists in the Research and Development organization at Biogen Idec in the design and execution of these studies.

Study protocols involving use of animals were approved by the Institutional Animal Care and Use Committee at the institution where the study was performed. Further information is available upon request.

## REFERENCES


1. Williams GJ, Witt PL. Comparative study of the pharmacodynamic and pharmacologic effects of Betaseron and AVONEX. *J Interferon Cytokine Res* 1998;18(11): 967–75.
2. Sturzebecher S, Maibauer R, Heuner A, Beckmann K, Aufdembrinke B. Pharmacodynamic comparison of single doses of IFN-beta1a and IFN-beta1b in healthy volunteers. *J Interferon Cytokine Res* 1999;19(11):1257–64.
3. Bagnato F, Pozzilli C, Scagnolari C, Bellomi F, Pasqualetti P, Gasperini C, Millefiorini E, Galgani S, Spadaro M, Antonelli G. A one-year study on the pharmacodynamic profile of interferon-beta1a in MS. *Neurology* 2002;58(9):1409–11.
4. Pease AC, Solas D, Sullivan EJ, Cronin MT, Holmes CP, Fodor SP. Light-generated oligonucleotide arrays for rapid DNA sequence analysis. *Proc Natl Acad Sci USA* 1994;91(11):5022–6.
5. Lipshutz RJ, Morris D, Chee M, Hubbell E, Kozal MJ, Shah N, Shen N, Yang R, Fodor SP. Using oligonucleotide probe arrays to access genetic diversity. *Biotechniques* 1995;19(3):442–7. Review.
6. Schena M, Shalon D, Davis RW, Brown PO. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* 1995;270(5235): 467–70.

# Tissue Cross-Reactivity Studies for Monoclonal Antibodies: Predictive Value and Use for Selection of Relevant Animal Species for Toxicity Testing

WILLIAM C. HALL, VMD, PhD, DACVP, SHARI A. PRICE-SCHIAVI, PhD, DABT, JOAN WICKS, DVM, PhD, DACVP, and JENNIFER L. ROJKO, DVM, PhD, DACVP

**Contents**

1 Introduction
2 Antibody Types and Their Evolution
3 Biological Effects of Antibodies
4 Tissue Cross-Reactivity Procedures
  4.1 The Tissues
  4.2 The Controls
  4.3 The Staining Method
5 Interpretation of Tissue Cross-Reactivity Findings
  5.1 CDR-Mediated "Specific" Staining
  5.2 Non-CDR-Mediated "Nonspecific" Staining
  5.3 Tissue Adequacy and Staining Adequacy
  5.4 Background Staining
  5.5 Scoring
  5.6 Types of Tissue Cross-Reactivity Staining Patterns
  5.7 Context
  5.8 Significance of Tissue Cross-Reactivity Findings
  5.9 Animal Tissue Cross-Reactivity Testing
6 Biological Relevance of Human Tissue Cross-Reactivity Studies
  6.1 Valence
  6.2 ADCC
  6.3 Immune Complex Disease
7 Conclusions
  References

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

## 1 INTRODUCTION

With the advent of hybridoma technology where murine myeloma cells could be programmed to produce antibody targeted to a specific epitope [1], a new era was ushered in for biomedical research that resulted in a legion of applications for these novel molecules. The first monoclonal antibody approved for human use was OKT3, which was initially used for rescue therapy for acute renal graft rejection by suppressing CD3+ T cells [2], and later expanded to prevent acute graft rejection for multiple organs in early postoperative periods [3–5]. Prior to the development of OKT3 murine monoclonal antibody, antithymocyte globulin of equine or lagomorph origin was similarly used to accomplish immunosuppression [6]. Both worked well, but the specificity of the monoclonal antibody proved more potent as an immunosuppressive agent than did the more variable polyclonal anti-bodies used previously. The serum half-life of OKT3 murine monoclonal antibody was only about 50 hours, so activity was short-lived. Part of the reason for the shortened half-life was the development of antibodies (human anti-mouse antibody [HAMA]) against the foreign OKT3 mouse IgG [6]. Subsequently, technologies were developed to provide antibodies that would not be immunogenic in an effort to prolong the circulation time (i.e., duration of treatment) of the molecule, improve efficacy, and reduce unwanted side effects. This was begun by the substitution of human Ig molecule sequences in the murine moiety to reduce antigenicity, followed by humanization of the Ig by site directed mutagenesis to further reduce antigenicity. Finally, fully human antibodies were developed using transgenic mice that expressed human immunoglobulins.

Because of the wave of technology, the regulatory agencies needed to ensure the safety of these novel molecules in some manner that differed from traditional drug safety. Hence, the FDA composed a Draft Points to Consider in the Manufacture and Testing of Monoclonal Antibody Products for Human Use in 1994, which was modified in 1997. The purpose of the Points to Consider and the subsequent International Conference on Harmonization Guidelines for Preclinical Safety Evaluation of Biotechnology-Derived Pharmaceuticals (ICH guidelines, S6) was to establish guidelines for industry to ensure the safety of these novel molecules. Nothing was clad in stone when these documents were put into place, and each molecule was and still is considered unique, both in the target of the antibody as well as in the antibody structure and biologic characteristics themselves.

Among other things, regulatory guidance addresses determination of the structural integrity, specificity (quantifying affinity, avidity, and immunoreactivity), and potency of the antibody. Once these data are obtained for an antibody, a tissue cross-reactivity study on normal human tissues should be conducted in order to determine if binding of the molecule to intended and unintended targets is observed by immunohistochemistry.

Monoclonal antibodies generally bind to an epitope consisting of 5 or 6 amino acids, and characterization of the complementarity determining region

(CDR) of the molecule is very important to identify the amino acids recognized. Since there are 20 amino acids, the theoretical probability of the antibody having full cross-reactivity with a linear epitope identical to that recognized by the CDR region is 1 in $20^6$ or 1 in 64,000,000. At first glance there appears to be a low probability of cross-reactivity with the proteins in human tissues, but considering the large number of proteins throughout the cells of the body, there is actually a high probability of a cross-reactivity occurring with any given monoclonal antibody. The number indicated above does not include partial binding of an epitope with the CDR, nor conformational epitopes, both of which could greatly increase that probability. For example, roughly 5% of over 800 murine monoclonal antibodies directed against viral antigens from herpes simplex, cytomegalovirus, Epstein–Barr, vaccinia virus, myxoviruses, paramyxoviruses, arenaviruses, flaviviruses, orthoviruses, rhabdoviruses, coronaviruses, and human retroviruses cross-reacted with normal mouse tissues [7]. These represent viral antigens, but monoclonal antibodies directed against bacterial antigens react similarly. Moreover, almost all monoclonal antibodies directed against tumor antigens cross-react with the counterpart antigen found on the normal cell of origin of that tumor. If one accesses the human genome database with a BLAST search (http://www.ncbi.nlm.nih.gov/BLAST), and searches for the occurrence of 6 arginines in sequence in a protein, approximately 384 "hits" are generated (at least today), indicating that this sequence is present in a number of various human proteins. Thus, the probability of exposing a cross-reactivity in human tissues with a monoclonal antibody is high.

What does a tissue cross-reactivity study actually measure and why is it useful? Cross-reactivity assays measure what cells or noncellular tissue elements express the target epitope recognized by the CDR region of the monoclonal antibody and the location of that antigen on, within, or around various cells. Cross-reactivity assays also recognize off-target (cross-reactive) epitopes in tissue. Under optimized conditions cross-reactivity assays should not measure Fc binding of the monoclonal antibody, which is hopefully blocked by the addition of serum-containing immunoglobulins that bind to the Fc receptors in the tissues. Immunohistochemical cross-reactivity assays are useful because they provide information regarding target epitope distribution, off-target epitope distribution, and in vitro species comparisons (e.g., human versus monkey or rodent profiles) that assist in species selection for in vivo (preclinical) safety studies.

Let us digress a bit to discuss in general (specifics will follow) the tissue cross-reactivity study. Human tissues are obtained from autopsy or surgical specimens and snap frozen. Cryosections are prepared and fixed in a manner wherein antigens are not lost or altered. The samples are then blocked with serum, treated to block endogenous reaction enzymes (peroxidases, alkaline phosphatases, etc.) and other substances (endogenous biotin) that may interfere with interpretation of the reaction, run through a series of approximately a dozen more wash and reagent steps, desiccated, counterstained, and

coverslipped for interpretation. Tissue binding (target or off-target) is then evaluated microscopically. The procedures are lengthy and must be highly controlled to ensure accuracy of all tissue elements stained. These conditions definitely are not exactly what a monoclonal antibody would experience following injection into a human. The goal is to ensure the accuracy of the staining and minimize background. In other words, the goal is to provide an assay and not replicate the conditions obtained in vivo.

Since most of the monoclonal antibodies currently under development are human or humanized, the goal of the assay is to detect the bound monoclonal antibody and not the background endogenous immunoglobulins in the tissues. This means labeling the antibody or utilizing a secondary antibody that recognizes only the monoclonal antibody that is labeled (or unlabeled) so that it can be differentiated from the endogenous Ig.

As a preclinical tool, the tissue cross-reactivity assay is one of the first assays that should be performed to provide guidance to the investigator for the applicability of the particular monoclonal antibody clone for in vivo use. Careful scrutiny at this stage could save years of additional preclinical investigation of an unwanted cross-reactivity. In addition, the immunohistochemistry procedures are of great assistance in finding a relevant animal model for the in vivo preclinical studies. Thus, the purpose of this chapter is not to detail the immunohistochemical methods used in a tissue cross-reactivity study, but rather to outline important considerations in the conduct of the tissue cross-reactivity study and provide best practices guidelines for interpreting the information gained from this assay relative to human safety. In addition, the use of this assay along with others can be used to determine a relevant species (one that expresses the target or off-target cross-reactive epitopes) for preclinical safety studies.

## 2 ANTIBODY TYPES AND THEIR EVOLUTION

In humans, there are five isotypes of antibodies, IgG, IgA, IgD, IgE, and IgM, which are defined by the structures of their heavy chains and their abilities to form multimers (Figure 1) [8]. IgG is the most abundant isotype present in serum with average serum concentrations ranging from 0.5 to 9 mg/ml depending on the IgG subtype. This is followed by IgA (3 mg/ml), IgM (1.5 mg/ml), IgE (0.05 mg/ml), and IgD (trace). Each antibody isotype has unique functions. Critical functions of IgG include opsonization, complement activation, antibody-dependent cell-mediated cytotoxicity (ADCC), passive immunity, and regulation of B cells. Both IgM and IgD act as antigen receptors on naive B cells, and soluble, multimeric forms of IgM are involved in complement activation. IgA is involved in mucosal and passive neonatal immunity, while IgE is involved in immediate hypersensitivity [8].

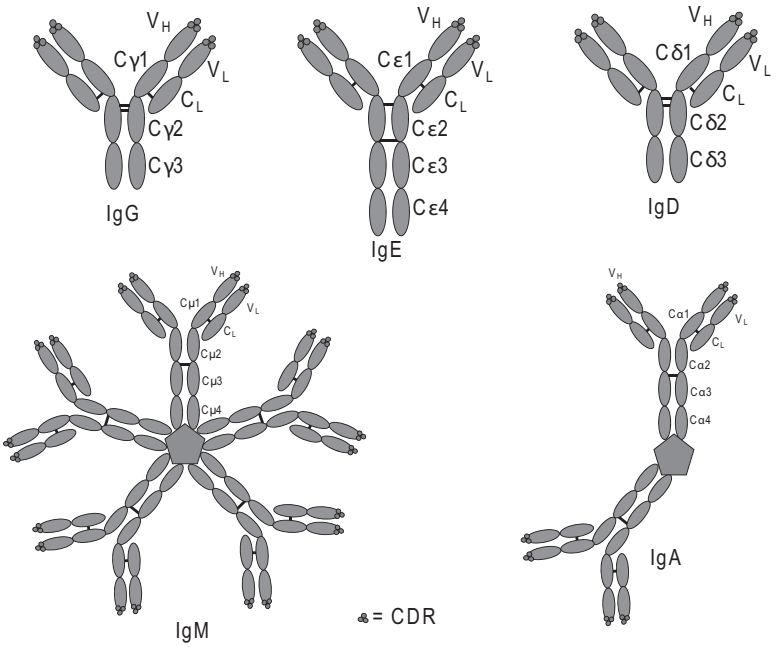


**Figure 1** Isotypes of human antibodies.

Because of their relatively large size and other physicochemical characteristics, antibodies generally have prolonged serum half-lives compared to smaller circulating proteins with averages as follows: IgG—23 days; IgA—6 days; IgM—5 days; IgD—3 days; and IgE—2 days [8]. Antibody fragments such as Fabs and $F(ab')_2$ have significantly shorter half-lives that range from minutes to hours [9]. In general, the smaller the antibody fragment, the shorter the serum half-life, as the smaller antibody fragments are cleared through the kidneys more quickly than larger fragments or whole antibodies (discussed in more detail in Chapter Physiologic IgG Biodistribution, Transport, and Clearance: Implications for Monoclonal Antibody Products). In addition, the ability of the antibody to reach its target antigen is affected by size and affinity or avidity (discussed in more detail in Chapter Physiologic IgG Biodistribution, Transport, and Clearance: Implications for Monoclonal Antibody Products). The choice of whole antibody or antibody fragment for a therapeutic candidate thus depends on several factors including the location of the target antigen, the desired activity of the antibody or fragment, affinity/avidity of antigen binding, and desired serum concentrations of the therapeutic candidate.

Most therapeutic antibodies on the market or in development are IgG or some fragment or derivative thereof. IgG is divided into four subtypes: IgG1, IgG2, IgG3, and IgG4, with IgG1 being the most abundant and IgG4 being least abundant in serum (Table 1). As with the main isotype groups, each subtype of IgG has unique properties and functions (Figure 2). The choice of

**TABLE 1 Biological properties of human IgG isotypes**

| | $IgG_1$ | $IgG_2$ | $IgG_3$ | $IgG_4$ |
|---|---|---|---|---|
| Serum concentration (%) | 65 | 23 | 8 | 4 |
| Fc binding | +++ | + | +++ | ± |
| Placental passage | ++ | ± | ++ | ++ |
| ADCC | +++ | + | +++ | ± |
| CF | +++ | ++ | ++++ | ± |

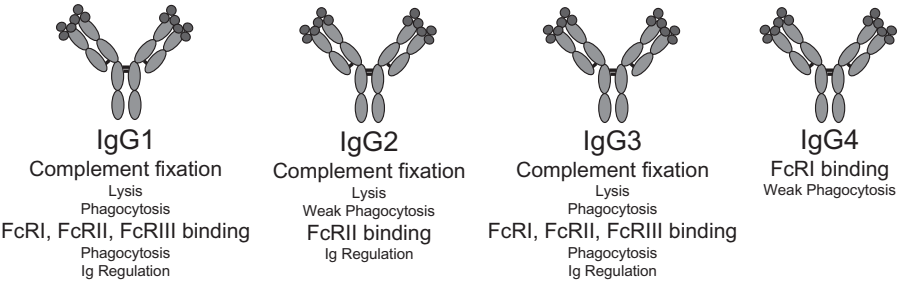


**Figure 2** Isotype subgroups of human IgG molecules.

isotype subgroup is dependent upon the desired activities for a given therapeutic candidate. There are several ways to generate these antibody-based therapeutics. Briefly, mouse monoclonal antibodies can be generated by standard methods after immunization of mice (Figure 3*a*). When the clone of choice has been selected, the variable regions can be cloned using routine molecular biology strategies. The resulting variable region fragments can then be cloned into an expression vector encoding human antibody light and heavy chain regions to generate a mouse/human chimeric antibody (approximately 30% mouse and 70% human). These chimeric antibodies can be expressed by and purified from stably transfected cell lines such as CHO or NSO cells. To generate humanized antibodies, the chimeric versions can be mutagenized in the variable regions to make the variable framework more similar to human antibody sequences, which may be selected from human antibody sequences found in antibody databases such as Kabat and others. The resulting humanized antibody (approximately 3% mouse and 97% human) can then be expressed and purified.

There are two main methods for generating fully human antibodies (Figure 3*b*). In the first, transgenic mice that express fully human IgG can be immunized with target antigen and the resulting antibody of choice can be cloned and expressed as described above. In the second, phage display libraries expressing fully human antibody Fabs can be screened or "panned" using the antigen of interest and the clone of choice can be cloned into and expressed

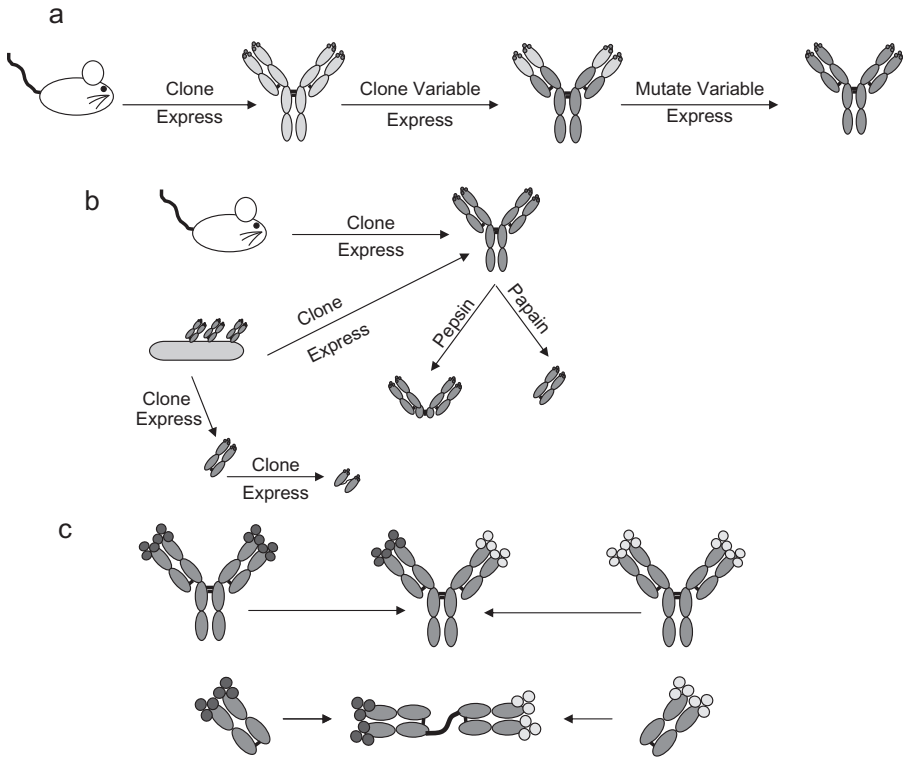


**Figure 3** Generation of antibodies and antibody fragments. (*a*) Mouse monoclonal antibodies are produced after immunization and production of hybridomas. Chimeric antibodies are produced by cloning the variable regions of a mouse antibody onto a human antibody framework. Humanized antibodies are produced by site directed mutagenesis of chimeric antibodies to make the variable regions of the mouse antibody more closely resemble a human antibody. (*b*) Fully human antibodies are produced by immunization of transgenic mice that express fully human antibodies or by screening human antibody phage display libraries. (*c*) Antibody fragments are produced by enzymatic cleavage of human antibodies or screening phage display libraries. Bispecific or multivalent antibody fragments can be made using standard molecular biology techniques.

from a human antibody expression vector. Bispecific antibodies, which recognize two different antigens, can be generated in a similar way from two different "parental" antibodies (Figure 3*c*).

Antibody fragments can be generated by papain digestion of full antibodies to yield Fab fragments ($V_H$, CH1, $V_L$, $C_L$) or pepsin to yield $F(ab')_2$ fragments (Figure 3*b*). In these cases the resulting antibody fragment must be purified from the byproduct Fc region or fragments. As this extra purification step can complicate (and add additional cost) the development of antibody fragments,

it is more common to screen phage display libraries that display antibody Fabs or other antigen binding fragments and then to clone the candidate of interest into a desired format and expression vector. This way it is easy to produce other types of antibody fragments such as Fv's or scFv's or bispecific and multivalent fragments.

Regardless of the type of resulting product (whole antibody or fragment, monospecific or bispecific), tissue cross-reactivity testing should be performed prior to toxicity studies to establish patterns of on- and off-target tissue binding and to confirm the relevance of the selected species for toxicity testing. For tissue cross-reactivity testing of any type of antibody-based therapeutic candidate, a similar irrelevant negative control antibody or fragment should be produced. In the case of whole antibodies, a human IgG of the same isotype (IgG1, IgG2, etc.) should be used, and for antibody fragments (which do not contain Fc or heavy chain regions) a similar fragment directed against an irrelevant epitope or the same non–cross-reactive epitope from a different species should be used. Ideally the negative control whole antibody would be identical to the test article except for an inactive or different CDR region.

## 3 BIOLOGICAL EFFECTS OF ANTIBODIES

As mentioned briefly above, antibodies have multiple, complex functions. In the body, the main effector functions of antibodies include pathogen/toxin neutralization, activation of phagocytosis and antibody-dependent cellular cytotoxicity (ADCC), and complement activation. Therapeutic approaches using antibodies or antibody fragments take advantage of these natural activities.

Pathogen or toxin neutralization occurs when an antibody CDR region binds to a pathogen or toxin and sterically hinders or blocks binding of the pathogen or toxin to its respective cellular receptor. As neutralizing activity can be mediated by the antibody CDR regions, it can be performed by antibodies of any isotype or even antibody fragments, if one ignores the role that the Fc or other non-CDR regions may play in the antibody activity. This neutralizing activity is utilized in passive vaccination and in therapeutic antibodies directed against specific pathogens or toxins such as Synagis®, which prevents respiratory syncytial virus binding to its cellular receptor by binding to and blocking RSV F protein (reviewed in [10]). In addition, this neutralization strategy is commonly used for other therapeutic strategies such as those for inflammatory and cardiovascular disease where the antibody or antibody fragment is used to block binding of an inflammatory cytokine or pro-thrombotic protein to its receptor either by binding to the ligand or its receptor. For example, Infliximab®, a chimeric anti–TNFα antibody, is used to neutralize TNFα activity in various inflammatory diseases [11]. In cardiovascular condi-

tions, ReoPro®, an anti-IIb/IIIa Fab, blocks the platelet glycoprotein IIb/IIIa receptors, thus inhibiting fibrinogen binding, platelet cross-linking and thrombus formation [12].

Antibodies also can mediate phagocytosis and ADCC via their Fc receptor binding regions. Antibody bound to target antigen may bind through the Fc region of the antibody to Fc receptors expressed on phagocytic cells such as macrophages and neutrophils (Fc receptors are discussed in more detail in Chapter Physiologic IgG Biodistribution, Transport, and Clearance: Implications for Monoclonal Antibody Products). Upon Fc receptor binding, signal transduction pathways are activated that result in engulfment and degradation of the bound particle as well as release of inflammatory mediators. Likewise, for ADCC, binding of antigen-bound antibody to Fc receptors on effector cells such as NK cells results in activation of the cells to release cytokines and cytolytic granules. In general, both phagocytic and cytotoxic functions are more efficient when multiple antibodies are bound to cell surface antigens resulting in higher avidity binding to the Fc receptors on the effector cells. If the desired effect of a therapeutic antibody is phagocytic or cytolytic activity, the isotype subgroup chosen should be one that binds to Fc receptors expressed on phagocytic or cytolytic cells (IgG1 or IgG3). Herceptin® and Rituximab®, therapeutic antibodies directed against ErbB-2 and CD20, respectively, have been reported to produce antitumor effects through Fc-receptor mediated mechanisms (reviewed in [13]).

For complement-dependent cytotoxicity, a complex cascade of protein binding and cleavage occurs that culminates in final complement mediated effector functions of phagocytosis, recruitment and activation of leukocytes, and osmotic lysis. It has been reported that Rituximab® (anti-CD20) utilizes a complement-dependent mechanism for tumor cell destruction [14]. If the desired effect of a therapeutic antibody is complement-dependent activity, the isotype subgroup chosen should be one that binds to complement proteins (IgG1 or IgG3).

Besides these classical effector mechanisms, other antibody-based mechanisms of action may be utilized, including targeted delivery of drugs, inflammatory mediators, or toxins; induction of apoptosis by modulation of critical signaling pathways; and direct linkage of target and effector cells by bispecific or multivalent antibody fragments.

## 4 TISSUE CROSS-REACTIVITY PROCEDURES

As mentioned earlier, the purpose of this review is not to detail the actual methods used for tissue cross-reactivity studies but rather to highlight the many important factors that must be taken under consideration during the conduct, evaluation, and regulatory review and of the study. Specific details for tissue collection, generation of control materials, and staining procedures themselves are readily available in the literature.

### 4.1 The Tissues

***Human Tissues*** According to the Points-to-Consider and other regulatory documents, tissues from at least three unrelated human donors should be evaluated in a tissue cross-reactivity study in order to screen for individual variation. In a 3-donor, 37-tissue cross-reactivity study, up to 111 different donors could theoretically be represented. To the extent possible, both males and females should be equally represented.

Tissue banks are the most common source of human tissues. While surgical biopsy accessions are preferred, they are limited in availability and often unavailable for many tissues (e.g., brain and other vital organs). Therefore the majority of tissue samples used in cross-reactivity studies are those acquired at autopsy. For autopsy accessions all efforts must be made to minimize the time interval between death and tissue collection. Information provided with specimens generally includes age, gender, and usually race and some clinical history and/or cause of death. As suggested in the Points-to-Consider document, the tissues used in a standard tissue cross-reactivity study are acquired from adults (>18 years of age). Pediatric tissues are often very difficult to obtain and are usually not used in a standard tissue cross-reactivity study unless there is a clear pediatric indication for the test article.

***Animal Tissues*** The ICH S6 guidelines and the FDA's Points to Consider in the Manufacture and Testing of Monoclonal Antibody Products for Human Use (US FDA [CBER], 28 February, 1997) state that for biological products, preclinical testing must be accomplished in a relevant species. A relevant species can be determined by immunohistochemistry, flow cytometry, or pharmacologic procedures. For tissue cross-reactivity (immunohistochemistry) studies in animals, at least two donors are recommended, with equal gender representation as appropriate. Since animals such as inbred rodent strains have relatively limited heterogeneity, staining of more than two donors is not considered as important as in humans, nonhuman primates, or other outbred strains, where the amount of genetic variability is much more extensive. The degree of individual variability in nonhuman primates still is not considered to be as great as in humans, but there is a growing trend toward staining three donors for animal tissues.

Animal tissues are collected at necropsy of humanely euthanized, purpose-bred laboratory animals immediately following death. Some nonhuman primates used for tissue acquisition may be wild caught. The animals used are preferably matched as closely as possible to that of the strain or origin of the animals to be used for preclinical toxicity testing.

***Considerations in Comparing Cross-Reactivity Results between Humans and Animals*** Several factors must be taken under consideration when comparing tissue cross-reactivity results between humans and animals. First, conditions surrounding acquisition of animal tissues are clearly different from those

of human tissues. Thus, the quality of the animal tissues is much easier to control because the tissues are taken at necropsy and are frozen within a short time following death. For human tissues the relative age of the donors used is likely to be greater than that of the animal donors, and therefore normal aging changes may not be as prominent in the animal tissues. Further the cause of death of animals used as tissue donors is euthanasia, in contrast to humans where the cause of death is due to a variety of causes including trauma, myocardial infarction, or stroke (any of which may have overt or subtle systemic effects). Combined with the relatively young ages of the animal donors, it can be assumed that underlying disease processes such as diabetes, high blood pressure, or cardiovascular disease are not present in animal tissues. Finally, the physiological processes that occur during natural death of human tissue donors might influence antigen expression and the resulting observed staining pattern of a particular test article. These are considerations that will be addressed further in the interpretation of the relevance of the species for the tissue cross-reactivity study.

***Tissue Microarrays*** Tissue microarrays containing tissue core samples of all the required tissue types in the FDA Points-to-Consider document are available. While tissue microarrays can be useful for preliminary investigation of reactivity of a test article, they have several disadvantages for optimal evaluation of antibody cross-reactivity in tissues. The tissue cores (usually ≤2 mm each) represent only a very small sample of the tissue and generally do not present a wide panorama of the variety of tissue elements present in a larger sample (epithelium, vascular and/or intrinsic smooth muscle, mesothelium, etc.). For example, a 2 mm core sample collected from the medulla of the kidney might not contain glomeruli or proximal tubular epithelium. In the kidney there are indeed multiple types of tubules, all with a variety of antigen expression patterns depending on their functional roles and location in the kidney. Even for a particular cell type (e.g., epithelium in the internal os versus external os of the cervix), there may be heterogeneity in antigen expression patterns, and unless expression is widespread or diffuse in distribution, reactivity is more easily missed than it would be in a larger section of tissue. Finally, a key part of the evaluation of tissue cross-reactivity is examining the relationship between the cells of an organ and the stromal tissues. This is enhanced by size of the specimen.

***Preparation of Tissues*** Tissues are quick-frozen in OCT embedding media, and 4 to 6 μm cryosections are placed on slides and allowed to dry, or stored at 4 °C for a few days or at –80 °C for longer terms prior to fixation. Prolonged retention of tissue sections is usually to be avoided even at –80 °C as antigen deterioration can occur even under these conditions. Blood smears are prepared fresh and allowed to dry prior to fixation. Tissue preparation (sectioning, drying, fixation) is governed by the requirements needed to preserve the

epitope on the positive control samples in preliminary studies to establish the correct conditions for the assay.

### 4.2 The Controls

***Isotype Control Antibody*** The isotype (or negative) control antibody should be an antibody of the same isotype subclass as the test material (IgG1, IgG2, etc.). The isotype control stained slide controls for staining due to binding to Fc receptors or similar non–CDR-mediated events. However, even with an isotype-matched control, nonspecific staining (non–CDR-mediated) can be observed with the test material but not with the negative control. Ideally the ideal negative control antibody is identical to the test antibody with a different CDR region. This type of control helps normalize or minimize variables such as glycosylation differences and allotypic variation of the antibody backbone that might contribute to non–CDR-mediated staining.

***Assay Control*** The assay control involves omission of the primary antibody (test article or negative control antibody) during the staining reaction and permits determination of staining by the secondary or tertiary antibodies or other components of the reaction process. This control slide may or may not be included based on the immunohistochemical method chosen.

***Tissue Staining Control*** Along with morphology, a tissue's ability to be stained for a common antigen provides evidence that the tissue is stainable and thus of sufficient quality to be used in a tissue cross-reactivity study. In the case of human tissues where the time between death and collection of tissues may be more prolonged compared to that for animal tissues, this control helps verify that proteins within the tissue have maintained their integrity. Depending on the lab, various staining control antibodies may be used including CD31, transferrin receptor, or $\beta_2$-microglobulin. The objective is to choose an antibody that recognizes an epitope that is expressed in all or most tissues. This indicates suitability of the tissues used in the cross-reactivity study but does not guarantee that the epitope recognized by the test article survived tissue collection, storage, and processing.

***Positive and Negative Tissue Controls*** In order to control the staining run and to determine if the method that has been chosen is adequately working, a positive tissue control is used. In addition a negative control tissue is used to ensure the specificity of the test article. A tissue with well-documented expression of the target antigen is the ideal positive tissue control because it is treated identically to the other tissues of the study. However, often a target antigen is not normally expressed by tissues, or on occasion may be washed from the tissues during the immunohistochemistry procedures. In these cases a neoplastic or other type of diseased tissue is used (e.g., rheumatoid arthritis or inflammatory bowel disease) or another type of positive control

must be used as outlined below. A suitable negative tissue control consists of a tissue or tissue element that does not express the target antigen.

If acceptable positive control tissues are not available, target antigen-expressing cell lines may be used. In some cases there may be cell lines that normally express the target antigen. If such cell lines are not available, cell lines transfected with the target antigen may be used. Target antigen expression and the ability to detect expression of that antigen by immunohistochemistry may be verified through the use of a commercial antibody known to react with the target antigen. When using cell lines, an appropriate negative control cell line should also be evaluated. This cell line should be as similar to that of the positive control cell line as possible and should be grown under similar conditions (e.g., suspension versus monolayer, serum versus serum-free). If transfected cells are chosen as the positive control cell line, then nontransfected or mock transfected cells should be used as the negative control cell line. Often epitope density cannot be established for the cell line, but if available, high-density and low-density cell lines should be chosen to provide an internal sensitivity control for the assay.

When neither tissue nor cell lines are available, another alternative is purified target antigen. This type of control might be necessary when the test article is directed against a bacterial toxin and there is no cell line available that expresses the whole toxin or the targeted subunit. In these cases the negative "tissue" control would consist of a protein that is known to be nonreactive with the antibody being tested.

### 4.3 The Staining Method

In 1942 Coons et al. described an immunofluorescence technique for detecting cellular antigens in tissue sections [15]. This method utilized a fluorescent label bound to the primary antibody to localize the target antigen. In this case the fluorescent label was the detection method. The use of an "immunoenzyme" approach to detect binding of the primary antibody was introduced a quarter of a century later [16] with the introduction of peroxidase-labeled antibodies.

Immunohistochemical staining can be direct or indirect. Direct immunohistochemical staining methods utilize only a primary antibody, which may be conjugated to horseradish peroxidase, biotin, alkaline phosphatase, or other chromogens. In the case of biotin-labeled primary antibodies, avidin or streptavidin linked to peroxidase binds to the biotin allowing detection of reactivity of the test antibody with the tissue. Indirect immunohistochemical staining methods utilize secondary, tertiary, or even quaternary antibodies, any of which may be linked either to biotin or enzyme (e.g., peroxidase).

The ultimate choice of the method to be used in a tissue cross-reactivity study hinges on the fine balance between giving the best signal with the least amount of background or nonspecific staining. Background staining is associated with the detection methods used and endogenous or exogenous tissue

pigments. While little can be done to affect the presence of tissue pigments, various blocks can be used to reduce staining associated with the detection methods. Detection enzymes such as horseradish peroxidase or alkaline phosphatase have counterparts endogenous to various tissues. For example, myeloperoxidase is present in mast cell and neutrophil granules. Myeloperoxidase utilizes the same substrate that is used during the chromogen development phase of the immunohistochemistry staining procedure. Several blocks are available that can be applied to the tissues prior to application of the horseradish peroxidase-labeled streptavidin or antibody that will substantially reduce the activity of the endogenous peroxidase enzyme. Other blocks are available for endogenous biotin present in tissue, thus preventing detection of endogenous biotin.

Protein blocks are applied to the tissues prior to application of the primary antibody to reduce nonspecific staining. Nonspecific staining is non–CDR-mediated staining. Protein blocks often contain serum, which contain immunoglobulins that bind to the Fc receptors in the tissues, blocking Fc attachment to those tissue receptors. Other proteins such as casein or BSA may be added to block nonspecific protein–protein interactions.

The following factors should be considered when choosing a staining method for a tissue cross-reactivity study: (1) test article forms that are available, (2) test article affinity, (3) proposed plasma level for clinical studies, (4) epitope stability, and (5) ability to scale up the method to stain a large number of tissues. These various issues are addressed below.

***Test Article Forms*** For preclinical testing (including tissue cross-reactivity) of antibody-based therapeutics, it is best to use the actual therapeutic candidate (or the closest derivative) to obtain the most accurate preclinical data about that therapeutic candidate. Most monoclonal antibody-based therapeutic candidates are fully human, humanized, or chimeric antibodies or fragments. This poses some challenge to tissue cross-reactivity studies on human and nonhuman primate tissues, as standard indirect immunohistochemical staining using an antihuman IgG secondary antibody yields excessive background staining of endogenous human IgG present in the tissue sections. For this reason antibodies or antibody fragments can be labeled in various ways to facilitate detection of binding to human or animal tissues. Labeling is particularly useful for facilitating detection of tissue binding of antibody fragments that do not contain an Fc region. However, any time an antibody or antibody fragment is labeled, it is critical to ensure that the label itself does not interfere with CDR-mediated binding to target epitope. At a minimum, characterization of the labeled antibody using techniques such as BiaCore, ELISA, and/or flow cytometry should be performed to evaluate the labeled antibody's ability to bind to target epitope and to determine any effect on epitope binding affinity and/or specificity. Although there are many different ways to label antibodies, one of the most common labels used for antibodies in tissue cross-reactivity studies is biotin.

Biotin is a member of the B vitamin family that is normally involved in various processes including biosynthesis of fatty acids, metabolism of branched amino acids, and gluconeogenesis. It binds with high affinity to avidin and streptavidin ($KD = 10^{-15}$ M). This small, bicyclic compound can be easily conjugated to macromolecules without disrupting their function. Biotin esters are most often used for antibody labeling. These compounds react with primary amino groups ($-NH_2$) present on lysines and at the N-terminus of the protein to form stable amide bonds. Other types of biotin reagents are available if biotinylation of groups other than primary amines is desired. After the labeling reaction, excess biotin and any other reaction by-products should be removed by column chromatography, dialysis, or desalting columns. Usually antibodies are labeled with three to six biotins per molecule, although this number can be different depending on how the reaction is set up and on how many available lysines are present for labeling. Excess biotin molecules can interfere with the CDR binding to the antigen. When an antibody or antibody fragment is labeled with biotin for cross-reactivity testing, a similarly labeled negative control antibody or antibody fragment should also be prepared.

Using biotinylated antibodies for immunohistochemical staining takes advantage of the high-affinity binding of biotin to avidin or streptavidin. For this reason biotinylated antibodies (either the primary antibody or secondary antibody) are easily detected using enzyme-conjugated (horseradish peroxidase or alkaline phosphatase) avidin or streptavidin. This type of procedure is useful when staining human or nonhuman primate tissues with chimeric, humanized, or human antibodies as it eliminates the need for an antihuman IgG secondary antibody. In addition this type of labeling facilitates staining with antibody fragments that do not contain an Fc region. Two important considerations when using biotinylated reagents for immunohistochemistry are use of a similarly labeled negative control antibody and adequate blocking of endogenous biotin in the tissue sections. Blocking of endogenous biotin is achieved by sequential incubation of the sections with avidin and biotin. In addition an assay control in which the primary antibody is omitted from the staining reaction should be performed to evaluate the level of endogenous biotin (to ensure adequate blocking) and other pigments as well as any nonspecific staining by secondary or detection reagents. A word of caution about the use of biotin is that its presence on some antibody molecules can result in non–CDR-mediated interactions forming between the tissues and the biotin resulting in nonspecific staining of the tissues. Depending on where the biotin molecules attach, biotinylation also can directly alter the test article CDR or trigger conformational changes in the test article antibody molecule that affects CDR-mediated binding. A concern is that the biotinylated negative control antibody may not behave in the same manner as the biotinylated test article.

As mentioned earlier, there are other ways to label antibodies for use in tissue cross-reactivity studies. Detailed discussion of various labels and staining methods associated with them are beyond the scope of this review. However,

whatever the label chosen, careful consideration should be given to its effect on the test article and to the specifics of the staining method (controls, buffers, etc.) that must be used with that type of label.

In cases where the test article is derived from the same species as the tissue being stained and labeling is not feasible due to interference with the CDR, a precomplexing method may be attempted. This method involves precomplexing of the primary antibody and the anti-IgG secondary antibody prior to application to the tissue and has been described for both mouse and human antibodies [17–19]. Alternatively, if available, anti-idiotypic antibodies (antibodies that are directed specifically against the hypervariable region of the test article) can be used to detect test article binding to the tissues. Care should be taken to ensure that the anti-idiotypic antibodies do not recognize endogenous IgG or modulate test article binding directly.

***Test Article Affinity*** Different immunohistochemical staining methods have different levels of theoretical amplification (Table 2). For antibodies with relatively low affinities, it may be desirable or necessary to use a method with a high level of amplification, such as a tertiary indirect in order to see reactivity with the target antigen. For antibodies with relatively high affinity for the target antigen or for highly expressed antigens, the level of signal amplification may not be as critical for detection of staining.

***Proposed Plasma Level for Clinical Studies*** The tissue cross-reactivity study is an immunohistochemistry assay, which, because of its nature, does not replicate the conditions obtained in vivo. Moreover, one is looking at tissue sections where the cells are cleaved and all portions of the tissues and cells (membrane, cytoplasm, and nucleus) and surrounding milieu are exposed to the same concentration of the antibody, which is different from the intravascular and perivascular concentrations observed in vivo (see below). Since buffers, protein types, and electrolytes also differ in vivo compared to the immunohistochemistry conditions, one cannot expect to attain high concentra-

**TABLE 2 Theoretical amplification and maximum test article concentration for various immunohistochemical procedures**

| Test Article Form | Method | Theoretical Amplification | Theoretical Maximum Test Article Concentration |
|---|---|---|---|
| Unconjugated | Secondary indirect | ++ to +++ | 10–50 μg/ml |
| Unconjugated | Precomplex | + to +++ | 10–20 μg/ml |
| Biotinylated | Direct | + to +++ | 10–300 μg/ml |
| HRP-conjugated | Direct | + to ++ | 10–300 μg/ml |
| Other labels | Indirect | ++ to ++++ | 10–100 μg/ml |
| Anti-Id | Indirect | ++ to +++ | 10–100 μg/ml |

tions of the monoclonal antibodies in the in vitro assay compared to in vivo because of the background staining associated with high concentrations. It is good to attempt these concentrations, but for all practical purposes it is unrealistic for many, if not most, test articles. A less sensitive procedure providing less background could be chosen, but that would defeat the goal of the tissue cross-reactivity study to determine the possible cellular and tissue location of the antigen recognized by the antibody. The tissue cross-reactivity study is an in vitro experiment utilized to determine the location and density of an antigen. That said, an attempt should be made at least in preliminary studies to use the proposed clinical concentrations, but if excess background results, the higher concentrations would do little to assist in interpretation and should not be utilized. The data from the preliminary studies can be used to support the concentrations used.

***Epitope Stability*** How and whether to fix tissues to preserve morphology and maintain the reactive epitope is always a question. The secondary and tertiary structures of an epitope are modified by excessive fixation and sometimes drying, which can alter staining of the fixed tissues. Further, since protein structure is based on a number of factors, including hydrogen bonding between amino acid side chains, factors such as pH and salt will play a role in the protein structure and, in turn, in the antigen–antibody reaction if that reaction is dependent on a secondary, tertiary, or quaternary structure. While tissue collected and placed into formalin is not a consideration for tissue cross-reactivity studies, minor fixation of cryosections with standard fixatives aid in maintaining tissue morphology. Further, some antigens are preserved better with such minor fixation. Thus, different fixatives such as acetone, methanol, neutral buffered formalin, or paraformaldehyde may be considered. Each fixative has a different mechanism of action. Acetone is an organic solvent, methanol is dehydrating, and aldehydes cross-link amino acids. The effect of these fixatives on target and off-target antigen preservation should be considered during the method development.

***Concentration Selection*** During method development the primary antibody is titrated across a range of concentrations in order to determine both an ideal concentration (the lowest test article concentration that produces maximum binding to the target antigen) and a higher concentration of antibody that allows detection of reactivity that is of relatively lower affinity. The choice of this higher concentration is usually a multiple of the ideal concentration that approaches the proposed clinical concentration as determined by the levels of background staining. These two concentrations are commonly within a concentration range of 10 to 20-fold. If a study is conducted using lower-fold difference in high and low staining concentrations (e.g., threefold), appropriate justification should be provided. Rarely, three concentrations of antibody may be chosen if the range between the ideal concentration and the proposed plasma concentration is large (>50 to 100-fold).

## 5 INTERPRETATION OF TISSUE CROSS-REACTIVITY FINDINGS

### 5.1 CDR-Mediated "Specific" Staining

To interpret potential cross-reactivity findings correctly, the pathologist must consider the nature of the epitope-test article CDR interaction, the type of test article and staining method as well as the histologic characteristics of the tissue being stained. First, staining of a tissue component could indicate binding to the target epitope (e.g., a leukocyte CD marker) mediated by the test article CDR, or it could indicate CDR-independent binding (nonspecific sticking or Fc receptor mediated binding). Just because a test article binds to an unexpected tissue component does not mean the binding is nonspecific. It could mean that the target CD marker is expressed at that site. For example, CD4 was originally recognized as specific for helper T cells but was later found to be present on parathyroid epithelium [20,21] as well as monocytes and macrophages. Thus, finding CD4 on helper T cells was judged an expected, on-target, CDR-mediated staining, while finding CD4 on parathyroid epithelium was initially judged an unexpected, off-target, CDR-mediated staining. In contrast, the T10B9 anti-CD3 monoclonal antibody reacts with a monomorphic membrane CD3/T cell receptor complex epitope but does not react with cytoplasmic CD3 in human T lymphocytes and does not recognize membrane CD3/T cell receptor complex or cytoplasmic CD3 in rhesus monkey or cotton-top tamarin T lymphocytes [22]. However, T10B9 also reacts with cytoplasmic filaments in human, rhesus monkey, and cotton-top tamarin epithelial cells that do not express CD3/T cell receptor complex.[1] Thus finding T10B9 cross-reactive, non–CD3/non–T cell receptor complex-epitopes on human and nonhuman primate epithelial cells was judged an off-target, CDR-mediated staining.

### 5.2 Non–CDR-Mediated "Nonspecific" Staining

Whether expected or unexpected and on- or off-target, CDR-mediated binding is very different from non–CDR-mediated binding. Non–CDR-mediated binding can be due to binding of the Fc portion of the test article to Fc receptors on monocyte/macrophages or other cell types (Figure 4; reviewed in Chapter Physiologic IgG Biodistribution, Transport, and Clearance: Implications for Monoclonal Antibody Products). Non–CDR-mediated binding can also be due to nonspecific stickiness, mediated by protein–protein, protein–nucleoprotein interactions, carbohydrate–carbohydrate (lectin-like) interactions, or van der Waals interactions. Thus, except for epitope : CDR matching, the types of physicochemical interactions that contribute to specific CDR binding are often similar to the types of physicochemical interactions that mediate nonspecific (non–CDR-mediated) binding. Additionally certain tissues or tissue components appear to be stickier than others. For example, vascular

[1]Interestingly this epithelial cytoplasmic filament cross-reactivity did not appear to lead to any toxic effects in clinical trials [23,24], reviewed in [25].

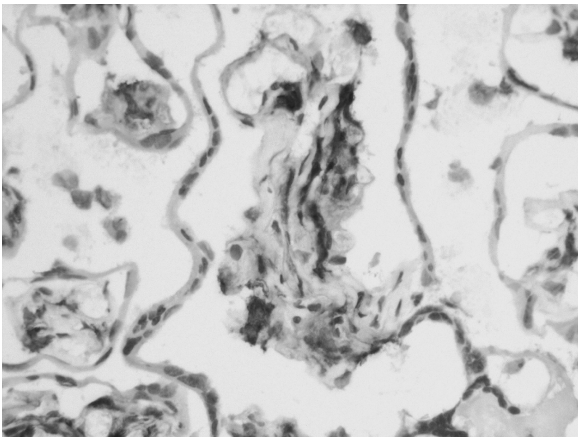

**Figure 4** FcγR1 in human placenta. CD64 staining of Hofbauer cells and perivascular monocyte/macrophages and dendritic cells. No staining of placental trophoblast epithelium or endothelium.

and intrinsic smooth muscle cells are often stickier than endothelial cells and/or perivascular collagen, as is mucin in the gastrointestinal tract and cervix. The stickiness of a tissue (and the propensity to bind antibody or protein reagents nonspecifically) also generally increases as the time between death and tissue collection increases. Additionally, as patients die, their peripheral tissues become poorly perfused and increasingly depleted of oxygen (anoxia). Anoxic endothelial cells contract and the spaces between adjacent endothelial cells widen, leading to protein leakage into the tissues. These serum proteins might coat (protect) the sticky sites on vascular smooth muscle cells. Alternatively, if the test article is directed against a serum protein (e.g., complement), the cross-reactivity staining patterns might appear greater than they really are due to reactivity with serum leaked into perivascular or more distal tissues.

### 5.3 Tissue Adequacy and Staining Adequacy

Interpretation of cryosections depends on tissue adequacy and staining adequacy. Tissue adequacy depends on the overall size and quality of each piece as well as whether the essential tissue components are adequately represented. For example, a piece of human uterus should have surface and glandular endometrium with endometrial stromal cells, muscular tunics, and, if possible, serosa (mesothelium). To ensure tissue adequacy, individual tissue cryosections should be used. Slides with multiple tissues (especially those with multiple small pieces—multitissue arrays) generally should not be used.[2] These

[2]Once a target or off-target reactivity is identified, a multitissue array might be useful in screening. For example, multitissue arrays might be used to identify how many livers are positive for a particular hepatocyte epitope.

small pieces often do not include all representative tissue components. Moreover in many cross-reactivity studies the key test article-stained cells might be evident in low numbers or in specialized portions of the tissue. For example, a tumor angiogenesis epitope might be present on normal endothelial cells in ovary that undergoes angiogenesis as part of postovulation corpus luteum formation during regular menstrual cycling. These test article-stained cells might only represent 1% of the endothelial cells of an ovary but are critical to identify as potential targets for that test article.

Staining adequacy is addressed by staining replicate cryosections from each tissue for an antigen known to be expressed on components common to all tissues (e.g., CD31, $\beta_2$-microglobulin, transferrin receptor), as discussed previously. This ensures that the tissue can be stained using immunohistochemical methods and thus is a suitable sample for the tissue cross-reactivity study.

## 5.4 Background Staining

Interpretation of cryosections depends on the level and type of background staining. Background staining can result from the staining procedure chosen for that test article or from endogenous or exogenous tissue pigments. Most immunohistochemical procedures used for tissue cross-reactivity are ultimately based on peroxidase cleavage of substrate–chromogen complexes to allow deposition of chromogen on tissue. If the staining procedure is avidin–biotin complex (ABC) based (e.g., biotinylated primary antibody), the tissue cryosections must be treated with specialty blocks to reduce endogenous biotin staining. These blocks generally are not required in staining of paraffin tissues unless antigen retrieval procedures are used. Certain tissue components (e.g., salivary gland duct epithelium) have higher endogenous biotin, which may compromise interpretation of that tissue component. Depending on the chromogen used in tissue cross-reactivity procedure, the background tissue pigments might have the same color or quality of staining as the chromogen. For example, the brown of diaminobenzidine is similar to the brown of melanin seen frequently in skin and eye sections. Other commonly encountered pigments include endogenous erythrocyte-derived pigments (reddish-yellow hemoglobin, yellowish-brown hemosiderin or hematoidin) or exogenous pigments (e.g., carbon pigments in macrophages in lungs from smokers or donors from polluted environments, india ink used to mark surgically obtained specimens).

## 5.5 Scoring

Scoring of test article staining in a tissue cross-reactivity study is a judgment call best made by a pathologist with broadly based experience in interpretation of frozen sections and comprehensive understanding of histology and immunohistochemistry. The actual scoring should include information regarding the cell or tissue component type, histologic location and pattern, subcellular location, and number or frequency of stained tissue components as well as the

staining intensity. As a typical cross-reactivity study contains three donors per tissue, statistical analysis of the findings is generally not appropriate.

### 5.6 Types of Tissue Cross-Reactivity Staining Patterns

We have reviewed information from 161 typical cross-reactivity studies conducted by PAI-CRL.[3] These studies were selected from more than approximately 800 studies conducted between 1990 and 2006. The information is presented as percentages of total, but it is important to note that the percentages may exceed 100% as not all the studies had identical study design. For example, several studies included more than one test article, but the calculations were based on a per study basis. The specific data are proprietary, and only generalities are presented in the summary tables. The types and forms of the test articles and their epitopes are listed in Table 3, and the frequency of unexpected cross-reactivity is listed in Table 4.

The types and frequencies of cross-reactivity findings are listed by general category in Table 5. The most frequent unexpected cross-reactivities were recognized in neural, epithelial, or contractile filament components. Narrow-

**TABLE 3 Test articles examined in 161 PAI-CRL tissue cross-reactivity studies**

| | |
|---|---|
| Types of test article | 33% Human IgG1, 16% humanized IgG1<br>7% Human IgG2, ≤1% human IgG3, 6% human IgG4<br>2% Human Fab<br>10% Chimeric, 10% mouse IgG1, 2% mouse IgG2a<br>6% Fusion molecules or conjugates |
| Forms of test article | 47% Unconjugated<br>15% Biotinylated<br>34% Fluoresceinated<br>1% HRP-conjugated<br>≤1% Other |
| Epitopes | 29% CD markers<br>25% Cytokine/chemokine receptors<br>7% Integrins<br>7% Infectious or toxic agents<br>9% Other |

**TABLE 4 Frequency of unexpected cross-reactivity in 161 PAI-CRL tissue cross-reactivity studies**

| | |
|---|---|
| % Unexpected cross-reactivity | None (69%)<br>Broad spectrum (6%)[a]<br>Narrow spectrum (21%)[b] |

[a]Multiple elements in multiple tissues.
[b]One to two elements in a few tissues, or one element in multiple tissues.

[3]Charles River Pathology Assoicates.

**TABLE 5 Types and frequencies of unexpected tissue cross-reactivity findings in 161 PAI-CRL studies**

| Unexpected Cross-Reactivity | Broad Narrow-Spectrum? | % of Total Studies |
|---|---|---|
| Neural tissue components | Broad | 4% |
| | Narrow | 8% |
| | | Axons, nerve endings—2% |
| | | Neuropil—3% |
| | | Schwann cells—2% |
| | | Glial cells—1% |
| Endocrine tissue components | Broad | 2% |
| | Narrow | 4% |
| | | Pituitary epithelium—2% |
| | | Parathyroid epithelium—2% |
| Gonadal tissue components | Broad | 1% |
| | Narrow | 4% |
| | | Seminiferous tubule cells—2% |
| | | Gonadal fibroblasts—1% |
| | | Oocytes—1% |
| Kidney tissue components | Broad | Not observed |
| | Narrow | Glomerular tuft cells—1% |
| Mucosal epithelium | Broad | 4% |
| | Narrow | 8% |
| | | Stratified squamous epithelium—5% |
| | | Basal lamina—1% |
| | | Transitional cell epithelium (urinary)—1% |
| | | Fallopian tube epithelium—1% |
| Gland or duct epithelium | Broad | 2% |
| | Narrow | 7% |
| | | Apocrine gland epithelium—4% |
| | | Salivary duct or acinar epithelium—2% |
| | | Sebaceous duct epithelium—1% |
| | | Pancreatic duct or acinar epithelium—1% |
| Cytoplasmic (contractile) filaments | Broad | 7% |
| | Narrow | 2% |
| | | Myoepithelium—1% |
| | | Cardiac myocytes—1% |
| | | Skeletal myocytes—1% |
| | | Myofibroblasts—1% |
| Endothelium | Broad | 1% |
| | Narrow | 1% |
| Other | Broad and/or narrow | Adipose (fat) cells—1% |
| | | Mesothelium—1% |
| | | Nuclei—1% |
| | | Dendritic cells—1% |

spectrum cross-reactivity (1–3 elements in a few tissues or 1 element in multiple tissues) was recognized more frequently than broad-spectrum cross-reactivity, except for contractile filaments in which broad-spectrum (multiple elements in multiple tissues) staining was recognized more often.

Examples of broad-spectrum and narrow-spectrum cross-reactivities are illustrated in Figures 5 and 6. Examples of unexpected staining of neural tissue elements are illustrated in Figures 7 and 8. Figure 8 also illustrates the very important point that large tissue sections are necessary for adequate evaluation of potential cross-reactivity. This particular cross-reactivity was seen in glial cells, primarily in dorsal but not ventral white

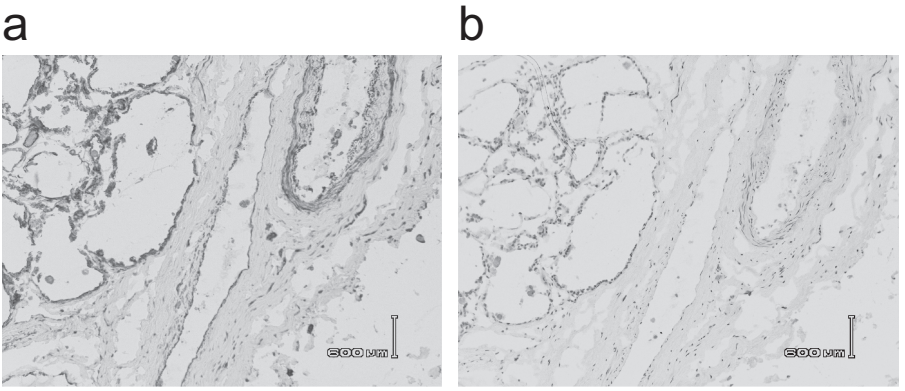


**Figure 5** (*a*) Unexpected broad-spectrum cross-reactivity with epithelium, endothelium and selected vascular smooth myocytes. No staining of interstitial (stromal) cells, collagen, or nuclei. (*b*) No staining in replicate sections stained by negative control antibody at similar staining concentration.

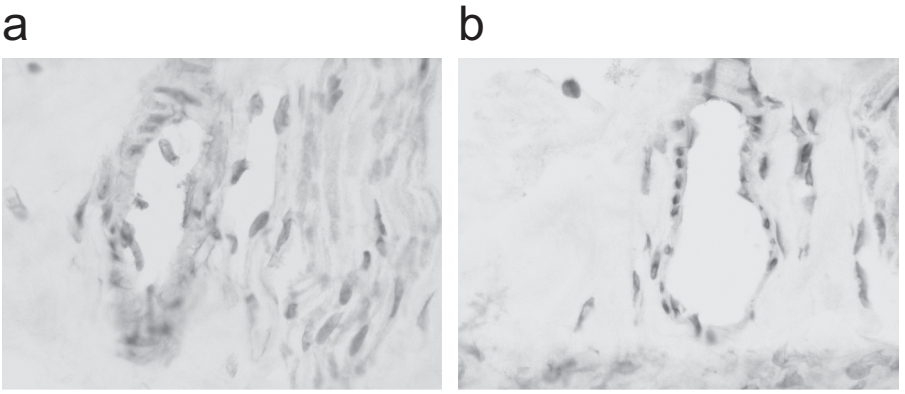


**Figure 6** (*a*) Unexpected narrow-spectrum cross-reactivity with contractile filaments in vascular smooth myocytes. No staining of interstitial (stromal) cells, collagen, nuclei, or adjacent peripheral nerve. (*b*) No staining in replicate sections stained by negative control antibody at similar staining concentration.

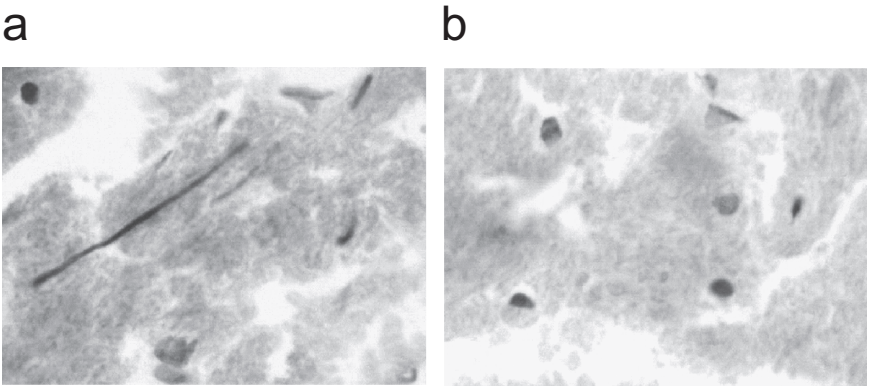


**Figure 7** (*a*) Unexpected narrow-spectrum cross-reactivity with axons in human brain. No staining of neural tissue components. (*b*) No staining in replicate sections stained by negative control antibody at similar staining concentration.

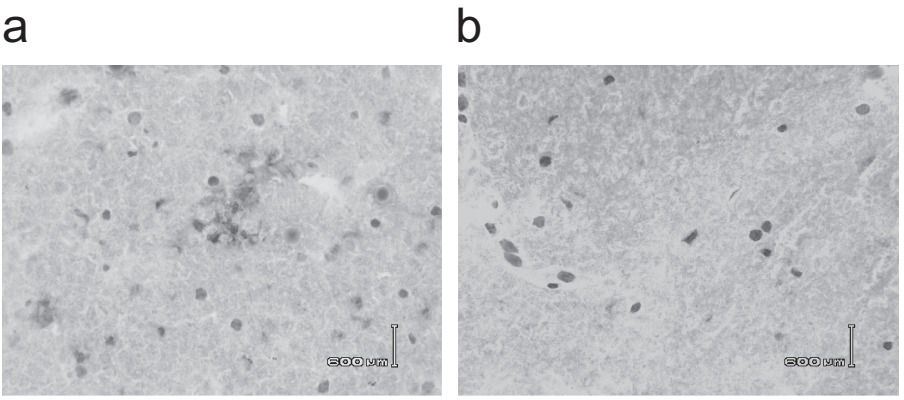


**Figure 8** (*a*) Unexpected narrow-spectrum cross-reactivity with glial cells in dorsal tracts of human spinal cord. Clustering of glial cells might indicate a very small glial scar. No staining of neural tissue components. (*b*) Greatly reduced cross-reactivity with glial cells in ventral tracts of human spinal cord.

matter tracts in the spinal cord. This difference (and potentially this cross-reactivity) would not have been recognized had smaller or incomplete tissue pieces been used. Figure 9 illustrates cross-reactivity with contractile filaments (cross-striations and intercalated discs) in heart muscle cells.

### 5.7 Context

Interpretation of test article staining should always be done in context of the published literature, knowledge of histology and physiologic processes, and common sense. We have previously seen staining of variably sized cytoplasmic inclusions in liver Kupffer cells when staining with antibodies directed against leukocyte markers. The staining did not represent true cross-reactivity with

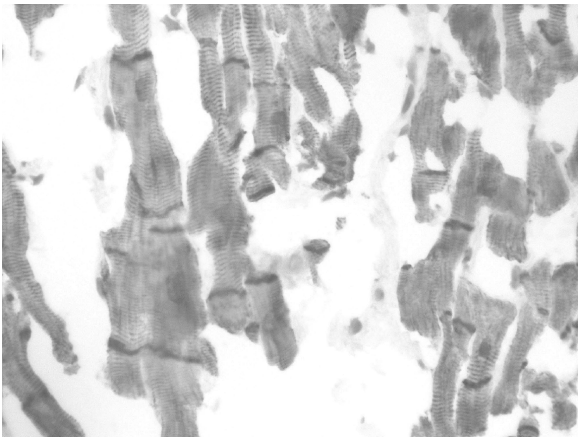

**Figure 9** Unexpected cross-reactivity with contractile filaments (cross-striations, intercalated discs) in cardiac myocytes. No staining of sarcolemmal cells, interstitial cells or endothelium.

Kupffer cell components but rather represented reactivity with leukocyte fragments that had been phagocytized by the Kupffer cells. Similar staining might be expected at other sites of leukocyte fragment clearance such as monocyte/macrophages lining the cords of Billroth in spleen or subcapsular or medullary sinusoids in lymph node. Likewise, in staining human tissues for the presence of an integrin highly expressed on platelets, distinct particles in Kupffer cells that resembled platelets also stained. The initial thought was that Kupffer cells themselves were staining. To unravel this, a commercial antibody against glycoprotein 1b, an antigen expressed only on platelets, was purchased and applied to the tissues. The stained antigens were thus identified as platelets within the Kupffer cells, indicating that either platelets were phagocytized by the Kupffer cells before death or phagocytosis of platelets occurred as a postmortem event during the interval between death and autopsy.

## 5.8 Significance of Tissue Cross-Reactivity Findings

As discussed above, the FDA and other international regulatory authorities strongly recommend tissue cross-reactivity studies as part of any monoclonal antibody test article approval/registration package. Additionally, tissue cross-reactivity studies provide an immunohistochemical window through which to examine patterns of CDR-mediated and non–CDR-mediated staining. This window can reveal potential efficacy or toxicity targets as well as novel information regarding potential sites of expression of target epitope or related but off-target cross-reactive epitopes. However, the presence of tissue cross-reactivity does not equate to toxicity. Likewise, the absence of tissue cross-reactivity does not necessarily ensure safe administration of monoclonal antibody test articles. Rather, tissue cross-reactivity just reveals potential sites to monitor for test article binding. When the actual monoclonal antibody

drug is administered to humans or animals, it becomes incorporated into the circulating immunoglobulin pool and is distributed, cleared, and/or salvaged by physiologic (and in heterologous species, immunologic) processes (Chapter 11). Whether monoclonal antibody test articles will bind to these or other (FcR, nonspecific, or non–CDR-mediated) sites will depend on the route of test article administration, extent of tissue distribution following administration, test article binding affinity, and the nature of, and potential access to, the cross-reactive subcellular sites and/or extracellular components.

The potential hierarchy of cross-reactivity findings according to site of test article staining is presented in Table 6. Test article cross-reactivity with membrane targets generally has been considered to be of greatest potential biological relevance or toxicity because access to this compartment is theoretically greater. Membrane binding to high-density epitopes also has been considered more likely to trigger events such as complement fixation and cell destruction, alterations in cell signaling, orADCC than antibody binding to other cellular locations (see above).

The information in Table 6 is in fact oversimplified. Access to target and potential off-target epitopes is first and foremost affected by antibody distribution. The physiologic distribution of monoclonal antibody test articles (Chapter 11) suggests that the drug will be presented sequentially to intravascular components (blood cells, plasma components), endothelium, vessel wall, perivascular interstitial fluid and extracellular fibrils, cell membrane, cell cytoplasm (if receptor-mediated or nonspecific uptake occurs), nuclei (if nuclear uptake occurs). Additionally some of the administered antibody will be cleared, compartmentalized intracellularly, and/or salvaged by physiologic processes conducted by endothelium, monocytes/macrophages, and epithelium. Last, normal cell turnover might provide nuclear, membrane, or cytoplasmic fragments that could be present as possible binding targets in blood or tissue.

## 5.9 Animal Tissue Cross-Reactivity Testing

Animal tissue cross-reactivity studies can provide information regarding whether the animal species under consideration might provide relevant models for safety/toxicity testing. Cross-reactivity is one of a number of methods for determining a relevant species; others include flow cytometry and assessing expected pharmacologic events to a monoclonal antibody. However, the animal tissue cross-reactivity staining patterns alone are not sufficient to ensure or deny the relevance of toxicity testing in that species. The example cited earlier in this chapter of the differences in T10B9 staining of membrane CD3/TCR complex molecules on human T lymphocytes and cytoplasmic filaments in human and primate epithelial cells is useful [22]. While toxicity testing in the primate might not reveal anything about the potential for toxicity following T10B9 binding to T lymphocyte membranes, it might reveal information regarding access and/or toxicity to subcellular epithelial compartments. Finally, the appropriate use of a relevant species to identify any adverse events or to

**TABLE 6 Potential for toxicity for cross-reactivity by cytologic or histologic location**

| Cytologic Location | Homologue(s) | Example(s) of Homologues | Potential for Toxicity for Unexpected Cross-Reactivity |
|---|---|---|---|
| Membrane | CD markers, receptors | CD3 in medullary thymocytes | Probably highest regulatory concern |
| Cytoplasm | Cytokines, chemokines, rare CD markers | CD3 in cortical thymoctyes | Probably limited in vivo access |
| Cytoplasmic filaments | Cytoskeleton, functional internal borders | Intermediate or contractile filaments | Probably limited in vivo access |
| Cytoplasmic granules | Organized subcellular sites (e.g., endosomes, lysosomes, ER/Golgi) | Pituitary hormones | Probably limited in vivo access |
| Nucleus | Nucleus, nucleolus | Steroid hormone receptors | Immune-mediated disease? |
| Soluble | Serum, interstitial fluid, urine, CSF, other body fluids | Cytokines, chemokines, complement factors, shed tumor antigens | Little to unknown |
| Extracellular structures | Basal lamina, basement membrane, interstitial fibers | Laminin, elastin, collagen, intercellular substance | Unknown? Immune-mediated disease? |

make a final determination as to whether antibody binding actually occurs in vivo is important. Many of the cross-reacting antigens are cytoplasmic, and appropriately designed toxicity studies can help unravel their biological significance. Reviewing the animal tissues subsequent to injection of the monoclonal antibody into them for changes related to the antigens, and also demonstration of the injected monoclonal antibody binding to suspect tissues in vivo, can add a great measure to the safety of the product.

## 6 BIOLOGICAL RELEVANCE OF HUMAN TISSUE CROSS-REACTIVITY STUDIES

Several types of hypersensitivity reactions can occur in the host receiving monoclonal antibody therapy, and the tissue cross-reactivity study might be predictive of some of these hypersensitivity reactions. Theoretically, adverse reactions result in tissue destruction because of the activation of complement or release

of perforin and/or other inflammatory mediators by effector cells (K cells, neutrophils, macrophages) bound to the antibody by Fc receptors [26]. One example of a possible type II hypersensitivity reaction is described in Sections 6.2. In practice, type II hypersensitivity reactions are more likely elicited by anti-idiotypic responses to monoclonal antibody infusion than to the infusion itself. However, the anti-CD20 monoclonal antibody rituximab has been linked to cytokine release. This adverse reaction is speculated to be mediated by cross-linking of the antibody to target cell receptors, Fc engagement, and complement activation [27]. Other reactions, such as type I (IgE mediated hypersensitivity), types III (immune complex) and IV (cell mediated hypersensitivity), might also play a role in host disease. For example, administration of infliximab has been associated with type I, II, and III hypersensitivity reactions due to development of anti-idiotype antibodies [28]. Local injection site reactions to monoclonal antitumor necrosis factor-α antibodies (e.g., adalimumab) are reported to elicit type IV hypersensitivity reactions [29]. There are other considerations that assist in this evaluation of the biological relevance of human tissue cross-reactivity studies, such as the valence of the antibody, the subclass affecting Fc binding and cytokine release from effector cells, concentration and distribution of the antibody molecule in the tissues, and location of the stained cross-reacting antigen to determine its accessibility by the antibody.

In the tissue cross-reactivity studies, the antibody has equal access to all tissues and all cell components (membrane, cytosol, nucleus) of the tissues on the section. This is not true in vivo where access to the tissues is governed by passive diffusion of the antibody to the tissue. Moreover, unless uptake by tissues is receptor-mediated, cell membranes preclude entrance of an antibody into the cells. In addition there are blood–brain, blood–nerve, blood–eye, and blood–testis barriers characterized by specialized endothelium that reduce movement of immunoglobulin into these protected spaces. Thus, some tissues have relatively little access by antibodies compared to others. Likewise, antigens within cells have little chance of access to the antibody compared to cell membrane or transmembrane antigens.

Antibodies injected intravenously are more accessible to tissues associated with the vasculature (endothelium, blood leukocytes, platelets, and erythrocytes) compared with deeper, more complex tissues that are poorly perfused. Thus, the location of the cells in the body relative to vascular perfusion and diffusion of antibodies are important aspects of interpretation of the cross-reactivity studies. For example, ReoPro® (abciximab, c7E3 Fab) is an antibody Fab directed against glycoprotein IIb/IIIa on the platelet surface. This receptor is responsible for interacting with fibrinogen and von Willebrand factor that result in platelet aggregation and thrombosis. Blockage of these receptors prevents platelet aggregation and thrombosis [30].

### 6.1 Valence

ReoPro® a Fab monovalent, and this precludes cross-linking of platelets that would occur if the antibody were bivalent. The half-life of ReoPro® is approxi-

mately 10 minutes with a secondary half-life of 30 minutes in humans, in part because of the size of the molecule but, more important, because of the intravascular location of the platelets as well as the high affinity of the antibody for the platelet receptor. In clinical trials ReoPro® has proved to be an effective agent in preventing platelet aggregation and thrombosis. Additionally, a portion of the gpIIb/IIIa molecule is a $\beta_3$-integrin receptor (e.g., vitronectin receptor, $\alpha_v\beta_3$) expressed by different cell types, including neovascular endothelium and smooth muscle. The binding of ReoPro® to neovascular endothelium reportedly reduces attachment of the endothelial cells to the basal lamina, which is an important aspect of its usefulness as a therapeutic for coronary angioplasty [30–31]. A bivalent antibody produced against the same antigen in mice resulted in platelet loss [32].

## 6.2 ADCC

ADCC is a reaction that can occur with IgG1 and IgG3 antibodies where the antibody binds to the antigen in question and killer cells (K cells and CD8+ K cells, NK-cells) then bind to the Fc portion of the antibody by the FcγRIII or other receptors on the effector cells [26].

An IgG1 chimeric (mouse/human) antibody was developed against an antigen expressed on a large number of human colon carcinomas [33] and a cross-reactivity study was performed. The antibody bound to endothelial cells of the gray matter of the cerebrum (Figure 10). The same reactivity occurred in the cynomolgus monkey, and a toxicity study was conducted in this relevant species. In order to eliminate the possibility of cell repair and resolution of any lesions that might develop with this antibody, a small number of the

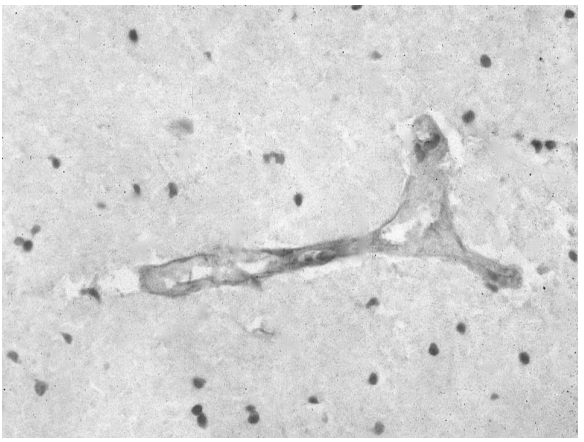

**Figure 10** Staining of endothelium of human cerebral gray matter with chimeric antibody to a colon carcinoma antigen. The same endothelial staining pattern occurred in cynomolgus monkey cerebral gray matter. ABC Immunoperoxidase.

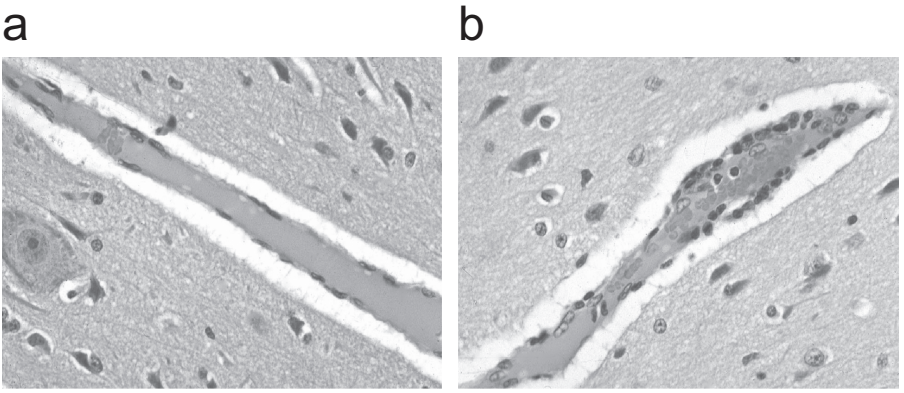


**Figure 11** Vessels in the cerebral gray matter of control (*a*) and dosed (*b*) cynomolgus monkey 72 hours following IV injection of an antibody directed against colon carcinoma. Note the leukocytes adherent to the endothelium characteristic of ADCC.

animals were sacrificed 72 hours following IV injection of the antibody. The pathology report was negative for any effects and a phase 1 trial was initiated. Of the six persons injected with the antibody, three developed dizziness and short-term memory loss, which subsided upon withdrawal of the drug. Additional review of the monkey brain tissues revealed the presence of mononuclear cells along the endothelial surface of small blood vessels in the cerebral cortical gray matter (Figure 11*a* and 11*b*). The cell infiltrates were not further characterized by staining for CD16 or any of the CD markers. The reaction nevertheless was consistent with ADCC, indicating the value of an appropriate model and careful interpretation of both the pathology and the cross-reactivity study.

### 6.3 Immune Complex Disease

In order to induce immune complex disease following immunization with a foreign protein (bovine gamma globulin), one generally needs to inject high levels of the foreign protein to produce the disease [32]. For symptoms and lesions of immune complex disease to occur, many immune complexes are needed. This usually means that the amount of protein injected must exceed the levels of antibodies produced in order to attain a level of antigen excess. For a nonhuman primate immunized against bovine gamma globulin, approximately 4.9 mg/kg is necessary to produce antibody equivalence for development of chronic immune complex disease [32–34]. For immune complex disease to occur, two conditions must be met: (1) the protein must be recognized by the host as foreign and (2) the levels of foreign protein must be high enough to exceed the level of antibody generated against them. The majority of monoclonal antibodies are human or humanized and therefore less likely than rodent antibodies to elicit anti-antibody responses, unless these are anti-

idiotypic responses directed against the therapeutic antibody CDR. Moreover therapeutic levels of most monoclonal antibodies are much lower, resulting in a dose of less than 100 μg/ml of plasma or approximately 3 mg/kg body weight. Effective monoclonal antibody concentrations usually are much less than 100 μg/ml, on the order of 1 to 20 μg/ml.

The story is different for an antigen targeted by the monoclonal antibody (site-directed immune complex disease). Immune complex disease can result by reactivity of the host against the bound antibody. Again, little reactivity is observed against the monoclonal antibody having host specific characteristics (human, humanized, and in most cases, chimeric antibodies), but this differs with a xenogenic immunoglobulin. In a number of studies we have assessed where the toxicity model was other than the homologous antibody host species, site-directed immune complex disease has been observed. Generally, serum complement levels are reduced during the toxicity portion of the study, and there is the deposition of host Ig and C3 at the sites of antibody localization accompanied by infiltrates of various inflammatory cells.

Sometimes, development of immune disease that targets the antigen is a welcomed event. Some anti-tumor monoclonal antibodies (e.g., monoclonal antibody 17-1A directed against the colon carcinoma antigen) are expected to elicit an anti-idiotypic host response. The anti-idiotypic antibody structure mimics the tumor antigen and can act as a vaccine resulting in an anti-anti-idiotypic response against the antigen [38–40]. Response can be either antibody mediated or T cell mediated.

## 7 CONCLUSIONS

Tissue cross-reactivity studies, although burdensome, provide a rational in vitro assay to determine the range and intensity of distribution of potential epitopes reactive with a monoclonal antibody test article prior to its administration to humans. In addition, cross-reactivity studies provide a useful tool to identify animal species for safety assessment. The cross-reactivity profiles of different species can be compared to the profiles obtained in human tissues. The predictive value of the assay lies in incorporating the characteristics of the monoclonal antibody (isotype, subtype, and other molecular modifications) with the biological activity of the molecule itself, and the potential in vivo distribution of it.

## REFERENCES

1. Kohler G, Milstein C. Continuous cultures of fused cells secreting antibody of predefined specificity. *Nature* 1975;256:495–7.
2. Cosimi AB, Colvin RB, Burton RC, Rubin RH, Goldstein G, Kung PC, Hansen WP, Delmonico FL, Russell PS. Use of monoclonal antibodies to T-cell subsets for

immunologic monitoring and treatment in recipients of renal allografts. *N Engl J Med* 1981;305:308–14.

3. Bristow MR, Gilbert EM, O'Connell JB, Renlund DG, Watson FS, Hammond E, Lee RG, Menlove R. OKT3 monoclonal antibody in heart transplantation. *Am J Kidney Dis* 1988;11:135–40.
4. Gay WA, Jr., O'Connell JG, Burton NA, Karwande SV, Renlund DG, Bristow MR. OKT3 monoclonal antibody in cardiac transplantation. Experience with 102 patients. *Ann Surg* 1988;208:287–90.
5. Norman DJ, Shield CF 3rd , Barry J, Bennett WM, Henell K, Kimball J, Funnell B, Hubert B. Early use of OKT3 monoclonal antibody in renal transplantation to prevent rejection. *Am J Kidney Dis* 1988;11:107–10.
6. Smith SL. Immunosuppressive therapies in organ transplantation. *Medscape* 2002. (http://www.medscape.com/viewarticle/437182).
7. Oldstone MB. Molecular mimicry and immune-mediated diseases. *Faseb J* 1998; 12:1255–65.
8. Abbas AK, Lichtman AH, Polischuk J. *Cellular and Molecular Immunology*, 4th ed. Philadelphia: WB Saunders.
9. Reff ME, Hariharan K, Braslawsky G. Future of monoclonal antibodies in the treatment of hematologic malignancies. *Cancer Control* 2002;9(2):152–66.
10. Le Calvez H, Yu M, Fang F. Biochemical prevention and treatment of viral infections—a new paradigm in medicine for infectious diseases. *Virol J* 2004;1:12.
11. Kirman I, Whelan RL, Nielsen OH. Infliximab: mechanism of action beyond TNF-alpha neutralization in inflammatory bowel disease. *Eur J Gastroenterol Hepatol* 2004;16:639–41.
12. Tcheng JE, Kereiakes DJ, Lincoff AM, George BS, Kleiman NS, Sane DC, Cines DB, Jordan RE, Mascelli MA, Langrall MA, Damaraju L, Schantz A, Effron MB, Braden GA. Abciximab readministration: results of the ReoPro Readministration Registry. *Circulation* 2001;104:870–5.
13. Brekke OH, Sandlie I. Therapeutic antibodies for human diseases at the dawn of the twenty-first century. *Nat Rev Drug Discov* 2003;2:52–62.
14. Harjunpaa A, Junnikkala S, Meri S. Rituximab (anti-CD20) therapy of B-cell lymphomas: direct complement killing is superior to cellular effector mechanisms. *Scand J Immunol* 2000;51:634–41.
15. Coons AH, Creech HJ, Jones RN, Berliner E. The demonstration of pneumococcal antigen in tissues by the use of fluorescent antibody. *J Immunol* 1942;45:159–63.
16. Avremeas S. Coupling of enzymes to proteins with glutaraldehyde: use of the conjugates for the detection of antigens and antibodies. *Immunohistochemistry* 1971;6:394.
17. Fung K, Messing A, Lee VM-Y, Trojanowski JQ. A novel modification of the avidin-biotin-complex method for immuniohistochemical studies of transgenic mice with murine monoclonal antibodies. *J Histochem Cytochem* 1992;40:1319–28.
18. Hierck BP, Iperen LV, Groot ACG-D, Poelmann RE. Modified indirect immunodetection allows study of murine tissue with mouse monoclonal antibodies. *J Histochem Cytochem* 1994;42:1499–1502.

19. Tuson JR, Pascoe EW, Jacob D. A novel immunohistochemical technique for demonstration of specific binding of human monoclonal antibodies to human cryostat sections. *J Histochem Cytochem* 1990;38:923.
20. Faure GC, Tang JQ, Mathieu P, Bene MC. A CD4-like molecule can be expressed in vivo in human parathyroid. *J Clin Endocrinol Metab* 1990;71:656–60.
21. Hellman P, Karlsson-Parra A, Klareskog L, Ridefelt P, Bjerneroth G, Rastad J, Akerstrom G, Juhlin C. Expression and function of a CD4-like molecule in parathyroid tissue. *Surgery* 1996;120:985–92.
22. Campana D, Thompson JS, Amlot P, Brown S, Janossy G. The cytoplasmic expression of CD3 antigens in normal and malignant cells of the T lymphoid lineage. *J Immunol* 1987;138:648–55.
23. Waid TH, Lucas BA, Amlot P, Janossy G, Yacoub M, Cammisuli S, Jezek D, Rhoades J, Brown S, Thompson JS. T10B9.1A-31 anti-T-cell monoclonal antibody: preclinical studies and clinical treatment of solid organ allograft rejection. *Am J Kidney Dis* 1989;14:61–70.
24. Waid TH, Thompson JS, McKeown JW, Brown SA, Sekela ME. Induction immunotherapy in heart transplantation with T10B9.1A-31: a phase I study. *J Heart Lung Transplant* 1997;16:913–16.
25. Bunin N, Aplenc R, Leahey A, Magira E, Grupp S, Pierson G, Monos D. Outcomes of transplantation with partial T-cell depletion of matched or mismatched unrelated or partially matched related donor bone marrow in children and adolescents with leukemias. *Bone Marrow Transplant* 2005;35:151–8.
26. Male D. Hypersensitivity type II. In Roitt I, Brostoff J, Male D (eds), *Immunology*. Mosby International, London, 1998, pp 319–28.
27. Winkler U, Jensen M, Manzke O, Schulz H, Diehl V, Engert A. Cytokine-release syndrome in patients with B-cell chronic lymphocytic leukemia and high lymphocyte counts after treatment with an anti-CD20 monoclonal antibody (rituximab, IDEC-C2B8). *Blood* 1999;94(7):2217–24.
28. Lee SJ, Kavanaugh A. Adverse reactions to biologic agents: focus on autoimmune disease therapies. *J Allergy Clin Immunol* 2005;116(4):900–5.
29. Zeltser R, Valle L, Tanck C, Holyst MM, Ritchlin C, Gaspari AA. Clinical, histological, and immunophenotypic characteristics of injection site reactions associated with etanercept: a recombinant tumor necrosis factor alpha receptor: Fc fusion protein. *Arch Dermatol* 2001;137(7):893–9.
30. Lefkovits J, Plow EF, Topol EJ. Platelet glycoprotein IIb/IIIa receptors in cardiovascular medicine. *N Engl J Med* 1995;332:1553–9.
31. Romagnoli E, Burzotta F, Trani C, Biondi-Zoccai GG, Giannico F, Crea F. Rationale for intracoronary administration of abciximab. *J Thromb Thrombolysis* 2007; 23(1):57–63.
32. Teeling JL, Jansen-Hendriks T, Kuijpers TW, de Haas M, van De Winkel JG, Hack CE, Bleeker WK. Therapeutic efficacy of intravenous immunoglobulin preparations depends on the immunoglobulin G dimers: studies in experimental immune thrombocytopenia. *Blood* 2001;98:1095–9.
33. Hurwitz E, Adler R, Shouval D, Takahashi H, Wands JR, Sela M. Immunotargeting of daunomycin to localized and metastatic human colon adenocarcinoma in athymic mice. *Cancer Immunol Immunother* 1992;35:186–92.

34. Noble B, Brentjens JR. Experimental serum sickness. *Meth Enzymol* 1988;162: 484–501.
35. Birmingham DJ, Hebert LA, Cosio FG, VanAman ME. Immune complex erythrocyte complement receptor interactions in vivo during induction of glomerulonephritis in nonhuman primates. *J Lab Clin Med* 1990;116:242–52.
36. Hebert LA, Birmingham DJ, Shen XP, Cosio FG, Fryczkowski A. Rate of antigen entry into the circulation in experimental versus naturally occurring immune complex glomerulonephritis. *J Am Soc Nephrol* 1994;5:S70–5.
37. Hebert LA, Cosio FG, Birmingham DJ, Mahan JD, Sharma HM, Smead WL, Goel R. Experimental immune complex-mediated glomerulonephritis in the nonhuman primate. *Kidney Int* 1991;39:44–56.
38. Fagerberg J, Frodin JE, Ragnhammar P, Steinitz M, Wigzell H, Mellstedt H. Induction of an immune network cascade in cancer patients treated with monoclonal antibodies (ab1). II. Is induction of anti-idiotype reactive T cells (T3) of importance for tumor response to mAb therapy? *Cancer Immunol Immunother* 1994;38: 149–59.
39. Fagerberg J, Frodin JE, Wigzell H, Mellstedt H. Induction of an immune network cascade in cancer patients treated with monoclonal antibodies (ab1). I. May induction of ab1-reactive T cells and anti-anti-idiotypic antibodies (ab3) lead to tumor regression after mAb therapy? *Cancer Immunol Immunother* 1993;37:264–70.
40. Mellstedt H, Frodin JE, Biberfeld P, Fagerberg J, Giscombe R, Hernandez A, Masucci G, Li SL, Steinitz M. Patients treated with a monoclonal antibody (ab1) to the colorectal carcinoma antigen 17–1A develop a cellular response (DTH) to the internal image of the antigen (ab2). *Int J Cancer* 1991;48:344–9.

# Physiologic IgG Biodistribution, Transport, and Clearance: Implications for Monoclonal Antibody Products

JENNIFER L. ROJKO, DVM, PhD, DACVP, and SHARI PRICE-SCHIAVI, PhD, DABT

**Contents**

1 Introduction
2 Biodistribution of Endogenous IgG
  2.1 Egress from Intravascular Fluid Compartment into Interstitial Fluid Compartment
  2.2 Tissue-specific Barriers to IgG Transport/Uptake
  2.3 IgG within the Interstitial Fluid Compartment
3 Immunoglobulin Clearance
  3.1 Lymphatic Clearance
  3.2 Mononuclear Phagocyte System and Endothelial Cell Clearance of IgG–$Fc_\gamma$ Receptors
  3.3 Effect of IgG Structure on Plasma IgG Clearance and Tissue Catabolism
4 Specialized Tissue-Specific, IgG Transport or Clearance Mechanisms
  4.1 FcRn
  4.2 pIgR on Epithelium and Other Cell Types
  4.3 Kidney
  4.4 Breast (Mammary Gland)
  4.5 Ovary
  4.6 Placenta and Uterus
  4.7 Seminal Fluid
  4.8 Lung
  4.9 Skin
  4.10 Liver
  4.11 Salivary Gland
  4.12 Colon

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

4.13 Thyroid
4.14 Cytosolic Transport in Epithelium, Mesothelium, and Other Tissues
4.15 Erythrocyte CR1 (CD35) Clearance of IgG-Containing Immune Complexes
5 Conclusion
References

## 1 INTRODUCTION

Following administration, monoclonal antibodies become part of the endogenous IgG pool for distribution, transport and clearance unless these physiologic processes are altered by antibody CDR–epitope interactions or host anti-mononclonal antibody immune responses. This review focuses on physiologic (endogenous) IgG biodistribution, clearance and transport processes. Most of the information on these processes was derived from human clinical or laboratory studies. When available, information on exogenously administered monoclonal antibody test articles (data obtained from experimental studies in mice, monkeys or other species or clinical studies in humans) is included. The information is assumed to apply generally to humans and non-human primates (particularly cynomolgus monkeys), unless otherwise noted. Monoclonal antibody pharmacokinetics and pharmacodynamics were reviewed by Lobo and colleagues [1].

This review is intended to provide helpful information regarding potential monoclonal antibody IgG test article (non–CDR-mediated) binding to specific tissues, cell types or other tissue elements, and subcellular locations following exogenous intravenous administration in animal model test systems. Particulars specific to other routes of administration were reviewed by Lobo et al. [1].

Concerning plasma IgG characteristics in humans, endogenous human IgG (150 kd) has a plasma half-life greater than 21 days in humans [2]. The plasma level ranges from 8 to 16 mg/ml in healthy volunteers. Eighty percent of the total Ig in plasma is IgG with 65% of total IgG being IgG1 (Figure 1) [2]. With these levels of total IgG, administration of a therapeutic antibody, even at relatively high doses, is not expected to change the total level of plasma IgG or specifically IgG1 significantly. Therefore, it is safe to assume that exogenously administered antibodies will be distributed, transported, and cleared through normal physiologic processes [1].

Similar to that of endogenous IgG, the plasma half-life for human IgG1, IgG2, and IgG4 therapeutic antibodies can be up to three to four weeks in humans. This long plasma half-life has been ascribed to the protection from catabolism conferred by the neonatal Fc receptor (FcRn) as part of its role in regulating IgG homeostasis [3–5] (see below). In beta-2-microglobulin-

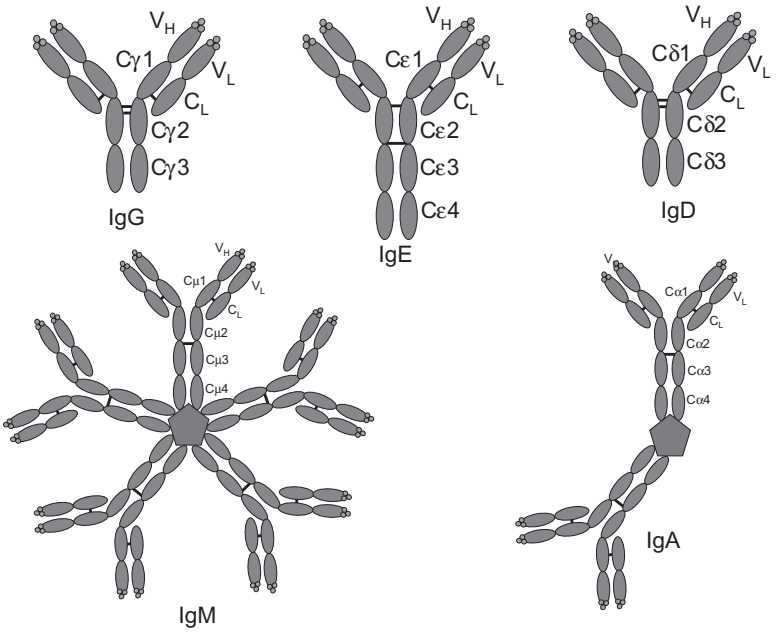


**Figure 1** Antibody isotypes present in humans.

deficient mice, marked reductions in plasma IgG half-life (beta-phase [4]) are concomitant with reduced FcRn expression. On the other hand, the plasma half-life of human IgG3 is shorter due to a single amino acid difference (arginine substitution for histidine at position 435) in the FcRn binding domain [6] (see discussions of FcRn and isotype below) that results in less FcRn binding.

Fab (50 kd) and scFv (27 kd) antibody fragments have shorter plasma half-lives (0.5 to 21 hours) because of more rapid glomerular filtration and clearance [7]. Plasma clearance, particularly of the xenogenic (usually mouse) portions of humanized and/or chimeric monoclonal antibodies, might be accelerated by development of anti-mouse (or other species), anti-isotypic, anti-idiotypic, or anti-allotypic antibodies that foster immune complex formation or reticuloendothelial clearance via several different FcR forms or elicit allergic or anaphylactic responses [7,8].

## 2 BIODISTRIBUTION OF ENDOGENOUS IgG

Among the benefits of intravascular administration of antibody-based therapies are the rapid intravascular biodistribution and high-level bioavailability. Several factors influence the biodistribution of immunoglobulins within the

vascular and tissue compartments. Intravascular fluid biodistribution of endogenous (or exogenous) IgG to any human tissue depends on the cardiac output (liver 28%, kidney 23%, brain 14%, skin 9%, skeletal muscle 16%, heart muscle 5%, rest of body 6% [9]) and vascularity of that tissue.[1] Within each tissue, IgG biodistribution is affected by the locations of capillary beds and postcapillary venules. Tissue compartment IgG levels are increased by processes promoting egress from vessels into tissue (convection, transcellular transport, receptor-mediated transport, and diffusion) and decreased when IgG is removed by catabolism or lymphatic drainage (convection, diffusion of fluid-phase antibodies, and migration of IgG or immune complexes associated with mononuclear or dendritic cells [1]). Further details of pharmacokinetic characteristics of exogenously administered monoclonal antibodies including CDR-mediated, epitope interaction-dependent influences on biodistribution and epitope-independent, plasma concentration-dependent clearance mechanisms were reviewed by Lobo et al. [1].

### 2.1 Egress from Intravascular Fluid Compartment into Interstitial Fluid Compartment

In liver, lung, kidney glomeruli, and spleen, the high vascularity and high cardiac output (first-pass effect) increases the potential for high uptake of antibody by that tissue. In less vascularized tissues (e.g., lymph node, remainder of kidney[2]), higher plasma concentrations are needed to generate higher concentration gradients to facilitate diffusion to more distant targets [13,14]. The rate of diffusion into interstitium is regulated by plasma concentration, plasma–interstitium concentration gradient, and outward diffusion into lymphatics, and this has been estimated as 5% to 10% of the rate of diffusion in aqueous solutions [1,15,16].

To reach the perivascular interstitium, immunoglobulins and other molecules must breach the endothelial cell barrier. Simionescu and Anthohe [17] note that an adult human contains more than $10^{13}$ endothelial cells which cover 7000 square meters and mass approximately 1 kg.[3] Egress to interstitium is affected by type of endothelium (continuous, fenestrated, or sinusoidal), endothelial uptake and transport processes (receptor-dependent or receptor-independent uptake, endocytosis and transcytosis[4]), paracellular (between or

[1]Cardiac output to rhesus monkey tissues was described by Forsyth et al. [10].

[2]The capillaries of the nonglomerular renal cortex (e.g., peritubular capillaries) represent a second capillary bed that derives from ramifications of the efferent glomerular arteriole. The vascular sequence is renal artery—arcuate arteries/arcuate arterioles—cortical arterioles—afferent (glomerular) arteriole—glomerular capillaries—efferent (glomerular) arteriole—peritubular capillaries [11,12].

[3]Pries and Kuebler [18] indicated a total surface area of 350 square meters and a mass of 100 g.

4Simionescu et al. [19] defined transcytosis as the polarized transport of plasma proteins across cells via caveolae or endosomal transport vesicles to the subjacent tissue or bidirectionally across other polarized cells such as epithelial cells.

alongside cells) transport, and tissue-specific barriers to macromolecule transport (e.g., blood–brain, blood–eye, or blood–nerve barriers; see below). Leakage of plasma proteins is greater from postcapillary venules than from capillaries as interendothelial pores are generally larger (50–60 angstroms compared with 18–20 angstroms [20]).

In addition molecules have been identified on continuous endothelium that facilitate binding of albumin to endothelium, increase capillary permeability, and theoretically can increase paracellular transport or transcytosis of albumin-bound molecules. An example of this is the gp60 albumin-binding glycoprotein molecule identified on continuous endothelium in multiple tissues (heart, lung, skeletal muscle, adipose, peritoneum, and intestinal smooth muscle) but absent from continuous endothelium in cerebral cortex, and sinusoidal or fenestrated endothelium in liver, adrenal, pancreas, and intestinal lamina propria [21,22].

Antibody fragments (either administered directly or generated by catabolism in the test species) distribute differently than whole IgG molecules. Monoclonal antibody fragments diffuse into interstitial fluid compartments faster and/or more completely than whole monoclonal IgG molecules in tissues with fenestrated or continuous vascular endothelium. However, antibody fragments and whole IgG molecules have comparable rates of diffusion in tissues with sinusoidal vascular endothelium (e.g., liver and spleen) in tumor-bearing mice [23]. The faster transfer across fenestrated or continuous capillary endothelium has been ascribed to the lower molecular weight (and increased capacity for diffusion) of the antibody fragments [23]. Antibody size-independent processes also limit IgG transfer from vessels to interstitium and include interstitial fluid/matrix pressure and cellular macromolecular exclusion [15]. Once within a tissue, exposure of potential target tissue elements to exogenously administered monoclonal antibodies can be affected by distance from blood vessels. Highly sticky monoclonal antibodies might associate with the first targets they encounter and will not distribute well to more distant targets. This might also rapidly decrease apparent plasma concentration and reduce concentration gradient across interstitium or parenchyma, further reducing distribution to more distant tissue elements.

Leukocyte-associated IgG egress from plasma is potentially mediated by binding of endogenous IgG or exogenously administered monoclonal antibodies to $Fc_{\gamma}RI$ or $Fc_{\gamma}RIII$ on circulating neutrophils and monocytes [2] or $Fc_{\gamma}RIIa$ on platelets [24,25]. Human peripheral blood monocytes also bear MHC class I-related FcRn that can transport IgG from plasma to tissue fluids, dendritic cells, or intestinal macrophages [26].

Transendothelial IgG transport may be mediated by binding of endogenous IgG or exogenously administered monoclonal antibodies to $Fc_{\gamma}RIIa,b$ on endothelium (e.g., $Fc_{\gamma}RIIa$ on skin microvessels [27]); $Fc_{\gamma}RIIb$ on placenta villous (fetal) endothelium [28,29]; or FcRn on placental syncytiotrophoblasts and fetal endothelium [28]. Transcellular transport is further described below under FcRn-mediated transport, $Fc_{\gamma}RIIa$-mediated transport, and specialized skin microvessel transport mechanisms.

Interendothelial (paracellular sieving) IgG transport is generally not efficient across continuous endothelium during homeostasis, but it can show increased efficiency in normally fenestrated, incomplete (kidney glomerulus) or sinusoidal endothelium (liver, spleen) or in inflamed vessels. Interendothelial paracellular sieving is generally restricted to the transport of water, ions, and other small molecules that can traverse the 6 nm gap found in about 30% of postcapillary venules [19]. Interepithelial paracellular IgG sieving may also occur. Paracellular sieving can increase early in inflammation when adjacent endothelial cells or epithelial cells retract and increase the paracellular space available [30].

Last, monoclonal antibodies or antibody fragments with specificity for endothelial transport systems (e.g., endosomal pathways) have been proposed as targeting devices to enhance therapeutic drug delivery or overcome tissue barriers [31].

## 2.2 Tissue-specific Barriers to IgG Transport/Uptake

In health, the blood–brain barrier is formed by tightly joined, nonfenestrated capillary endothelium with contributions by supporting perivascular astrocytic end-feet [32–34] and generally is thought to limit brain entry of endogenous IgG (or exogenously administered monoclonal antibody IgG test article) from plasma. However, endogenous IgG and other large hydrophilic molecules (including amyloid-$\beta_{1\text{-}40/42}$) can undergo brain–blood efflux [35]. This efflux is thought to be regulated by the adenosine triphosphate (ATP)-binding cassette transporter P-glycoprotein (P-gp) located within endothelial caveolae along with membrane and transport-related moieties cholesterol, caveolin-1, and caveolin-2 [36]. Additionally brain capillary endothelium expresses transferrin receptors; specific anti-transferrin receptor antibody-linked liposomes can be transferred into capillary endothelium but apparently do not enter brain parenchyma [37]. Whether the blood–brain movement of endogenous IgG or exogenously administered monoclonal antibodies is affected by $Fc_{\gamma}RI$-positive perivascular cells (macrophages, microglia) is not completely understood [38–44].

The blood–brain barrier can be overcome in certain circumstances. For example, in rats, uptake of mouse anti-*Cryptococcus neoformans* monoclonal antibodies into brain or cerebrospinal fluid is very limited after intravenous administration, but good uptake occurs after intrathecal administration [45]. Additionally, the blood–brain barrier can be overcome when molecules that are highly expressed by brain capillary endothelial cell membranes (e.g., transferrin) are targeted by specific antibodies as reported by Pardridge et al. [46] using a monoclonal anti-transferrin antibody in rats. In that study, the anti-transferrin antibody was cleared biexponentially, first by rapid, saturatable, transient hepatic extraction and later by slow continuous brain endothelial removal compared to the monoexponential clearance of isotype-matched

negative control antibody by the liver [46].[5] Further non–CDR-mediated binding can theoretically occur to FcRn or other receptors on brain endothelium. In rats, confocal microscopy experiments demonstrated FcRn expressed by Glut1 transporter-coexpressing capillary endothelium at the blood–brain barrier as well as by choroid plexus epithelium [47].

The blood–brain barrier generally protects the retina of the eye, but the barrier is interrupted at the prelaminar optic nerve head [48]. The aqueous humor of the eye is accessible to plasma macromolecules and contains albumin, with lesser amounts of transferrin, IgA, and IgG [49]. Although avascular, the cornea receives nutrients via percolation from the limbal vasculature (i.e., scleral venous plexus and associated arterioles). Repeated intravenous infusion of rabbit polyclonal anti-HSA IgG into nonimmunized rabbits indicated slowly increasing centripetal diffusion into rabbit cornea (approximately 1% per day), with corneal anti-HSA IgG reaching the avascular central portions of the cornea and equilibrating at approximately 70% of serum anti-HSA IgG at approximately 70 days; diffusion and equilibration were independent of electrostatic charge [50]. Last, while the blood–brain barrier can protect the brain parenchyma, it cannot protect the meningeal tissue. Intravascular polyclonal immunoglobulin therapy has been associated with a single case of aseptic meningitis, but the pathogenesis was not detailed [51].

Peripheral nerves and spinal nerve roots (including endoneurium) are not protected by the blood–brain barrier but are afforded some protection by a blood–nerve barrier consisting of continuous endothelium with tight junctions and surrounding pericytes [52]. Experiments using Ig-coated erythrocytes have suggested that Schwann cells have $Fc_{\gamma}RIII$ (and perhaps low-affinity $Fc_{\gamma}RII$) but not $Fc_{\gamma}RI$, both of which readily bind IgG1 and IgG3 but bind only small amounts of IgG4 and fail to bind IgG2 [53,54].

Other blood–tissue barriers (e.g., blood–thymus, blood–testis) are less well defined [55,56]. An air–blood barrier also has been described that affects the systemic uptake of aerosolized molecules [57]. Therapeutants that target glycoprotein adhesion molecules on endothelium might surmount the blood–brain or other tissue barriers [58].

### 2.3 IgG within the Interstitial Fluid Compartment

Once egress from plasma is accomplished, IgG is the most abundant immunoglobulin in tissue and is particularly prominent in the interstitial fluid compartment [2]. In healthy adult human volunteers, dermal interstitial fluid endogenous IgG levels were demonstrated to be 30% of plasma IgG levels [59]. In mice, IgG transport from capillary beds into interstitium is most frequent in skin and muscle with equivalent or less transport in adipose tissue and liver and much less transport in other tissues [60–62] consistent with the

[5]Minor amounts were cleared by kidney, but virtually no clearance was achieved in heart or lung [46].

major sites of IgG catabolism originally reported by Waldmann and Strober [63]. Information regarding potential differences in catabolism between monkeys, humans, and rodents has been reported [62,64].

## 3 IMMUNOGLOBULIN CLEARANCE

Clearance of endogenous and exogenous immunoglobulin takes place via a number of mechanisms. Among these are phagocytosis by mononuclear phagocytic cells (reticuloendothelial clearance), lymphatic clearance, and other tissue-specific and/or specialized mechanisms (discussed in more detail below).

### 3.1 Lymphatic Clearance

Drainage of interstitial fluid IgG into lymphatics depends on the concentration gradient and occurs passively via convection [1]. The leg lymph fluid of humans contains less than 25% of the amount of IgG found in plasma [65]. Other specialized lymphatic channels drain peritoneal, pleural, or pericardial cavities. Lymphatic drainage enters lymph nodes via the subcapsular sinus, and the lymph fluid IgG is available for reticuloendothelial clearance of IgG (subcapsular sinus, perivascular or parenchymal macrophages, perivascular or parenchymal dendritic cells, sinusoidal endothelium) as detailed next.

### 3.2 Mononuclear Phagocyte System and Endothelial Cell Clearance of IgG–Fc$_\gamma$ Receptors

Mononuclear phagocytes (macrophages, Kupffer cells, spleen or lymph node) and sinusoidal or other specialized endothelium remove IgG from plasma, most frequently as IgG-containing immune complexes, IgG aggregates, or IgG bound to cell-associated epitopes [66–69]. The initial binding event often occurs between the Fc portion of the IgG molecule and specific Fc receptors (Fc$_\gamma$R) on reticuloendothelial cells (Figure 2).

#### *Fc Receptors*

*Fc$_\gamma$RI* High-affinity Fc$_\gamma$RI (CD64) are found constitutively on monocytes, macrophages, dendritic cells, mast cells, platelets, and microglia and mediate internalization of antigen–antibody complexes into antigen-presenting cells for processing, antigen presentation and/or clearance [2,38,42,67,68,70]. In addition binding of antibody to Fc$_\gamma$R through the Fc receptor binding domain of IgG heavy chains can elicit several cell-mediated effector functions. In monocyte-macrophages, Fc$_\gamma$RI binding can trigger phagocytosis, antibody-dependent, cell-mediated cytotoxicity or other functions, including respiratory burst and cytokine/chemokine release. γ-interferon and β-interferon have

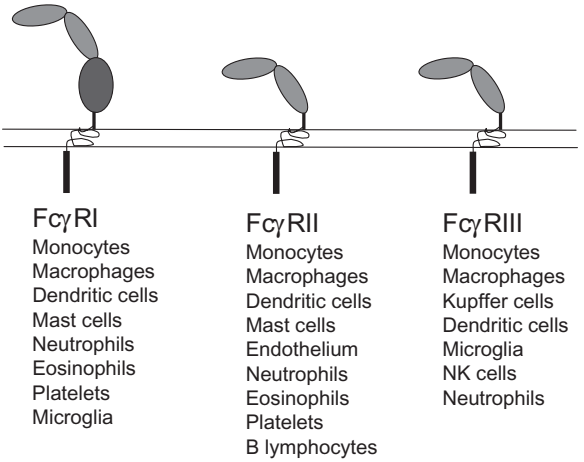


**Figure 2** Fc receptors and the cell types upon which they are expressed.

opposing actions on monocyte-macrophage $Fc_{\gamma}RI$ and function in vivo [71]. Immune complexes (e.g., monoclonal antibody–*Cryptococcus* neoformans) engage human and murine microglial $Fc_{\gamma}RI$ (and $Fc_{\gamma}RIII$) to initiate microglial activation and chemokine release [42]. In vitro, $Fc_{\gamma}RI$ are induced on human glomerular mesangial cells following treatment with IL-10 or γ-interferon but not IL-3, transforming growth factor-beta1 and granulocyte-macrophage colony-stimulating factor [72].

*$Fc_{\gamma}RII$* Low-affinity $Fc_{\gamma}RII$ (CD32) exists as two isoforms: $Fc_{\gamma}RIIa$ (CD32a) and $Fc_{\gamma}RIIb$ (CD32b) on B lymphocytes, monocytes, macrophages, dendritic cells, and mast cells. $Fc_{\gamma}RII$ (CD32) is also evident on endothelium with $Fc_{\gamma}RIIb$ (CD32b) described on placental chorionic villous endothelium where FcRI (CD64) and $Fc_{\gamma}RIIa$ (CD32a) are absent and where CD32b likely functions to transport IgG and/or scavenge immune complexes [29,67,68]. CD32a co-localizes with, and is negatively regulated by, CD31 (PECAM-1) in platelet plasma membranes [25]. $Fc_{\gamma}RIIa$ (CD32a) also has been implicated in clearance of antibody-coated erythrocytes in vivo as treatment of $Fc_{\gamma}RIIa$-transgenic mice with antihuman CD32a monoclonal antibody. MDE-8 downregulates $Fc_{\gamma}RIIa$ (and partially downregulates $Fc_{\gamma}RI$) and MDE-8 prevents immune-mediated hemolytic anemia [73]. Notably, $Fc_{\gamma}RIIa$-mediated phagocytosis of anti-RhD antibody-coated erythrocytes is favored by human IgG3 [74].

$Fc_{\gamma}RIIa$ (CD32a) also has been demonstrated on natural killer cells [75], primary human brain microglial cells [42], and on small vessel endothelium in the papillary (superficial and periadnexal) dermis of the skin which do not express CD64 ($Fc_{\gamma}RI$) or CD16 ($Fc_{\gamma}RIII$) and in which $Fc_{\gamma}RIIa$ engagement triggers Ca influxes and endosomal uptake of IgG [27]. Furthermore, although isolated human vein endothelial cells are normally negative for $Fc_{\gamma}RIIa$

(CD32a), they can express the molecules following activation by proteolytic enzymes (e.g., thrombin) or cytokines (e.g., tumor necrosis factor, γ-interferon [76]). Coggleshall has also proposed a role for $Fc_{\gamma}RIIb$ (CD32b) and associated phosphatases in the regulation of inhibitory responses of myeloid (particularly neutrophil and macrophage lineage) cells [77]. Binding of IgG1 and IgG2b but not IgM or IgG2a to the surface of monocyte precursors (e.g., HL60 cells) might influence monocyte differentiation at the expense of neutrophil differentiation [78].

*$Fc_{\gamma}RIII$* $Fc_{\gamma}RIII$ (CD16) is observed on human monocytes, macrophages, Kupffer cells, dendritic cells, and microglia, and its presence likely is important in the clearance of circulating immune complexes [42]. Two isoforms have been identified: transmembrane $Fc_{\gamma}RIIIa$ (natural killer cells, macrophages, subset of T cells) and glycosylphosphatidylinositol-anchored $Fc_{\gamma}RIIIb$ (neutrophils, eosinophils [69,79]). In mice, the receptors for CD32, CD16, and CD14 are often closely associated on the membranes of monocyte cell lines [80]. Notably, CD16 and CD32 are closely associated in immature B lymphocytes, whereas mature B lymphocyes lose CD16 and only express CD32 [81].

Reticuloendothelial clearance of exogenously administered monoclonal IgG antibodies assumes binding to FcR on the reticuloendothelial (mononuclear phagocyte system) cells. High-dose intravenous IgGs might add themselves to circulating immune complexes to foster clearance of the complexes via macrophage $Fc_{\gamma}RIII$ receptors [82]. The rate of reticuloendothelial clearance of anti-RLD antibody-erythrocyte immune complexes in healthy human controls has been estimated by experimental half-life calculations ($t_{1/2} = 100 \pm 7.5$ min [83]).

In monkeys, human(ized) test article administration can elicit primate antihuman antibodies (PAHA). PAHA binding to the human(ized) IgG test articles leads to formation of circulating immune complexes and subsequent clearance by reticuloendothelial cells or trapping by dendritic cells (e.g., follicular dendritic cells in spleen, lymph node, or mucosal-associated lymphoid tissues [67,68]). Alternatively, the test article might bind to its epitope (or a cross-reactive epitope) on circulating cells or intravascular protein, generating soluble or cell-associated immune complexes for reticuloendothelial clearance (e.g., depletion of CD45-positive leukocytes in monkeys following administration of monoclonal antihuman CD45 [13]).

***Molecular Determinants of Binding of Human, Mouse, and Rabbit IgG Isotype/Subclass to FcR*** Plasma levels, biodistribution, and clearance of endogenous IgG and exogenously administered IgG are affected by species, isotype, and subclass. FcR binding is conferred by the hinge region of the heavy chain ($C_H2$) of human IgG. Amino acids at both ends of the hinge region determine affinity of isotype-specific hierarchical binding of monomeric human and mouse IgGs to high-affinity human $Fc_{\gamma}RI$ with preference for HuIgG1, HuIgG3, and MsIgG2a > HuIgG4 >> HuIgG2 [84–88]. MsIgG1 and MsIgG2b do not bind human $Fc_{\gamma}RI$ [85]. Each HuIgG and MsIgG2a molecule has two

binding sites for $Fc_{\gamma}RI$ with one within each heavy chain $C_H2$ domain. In HuIgG3, a unique first exon and replacement of a long flexible spacer for the more rigid spacer encoded by repetitive exons 2–4 in HuIgG1 confer segmental flexibility but reduce $Fc_{\gamma}RI$ binding [89]. In HuIgG4, which retains a relatively rigid hinge, critical changes in the hinge-link region that decrease $Fc_{\gamma}RI$ binding likely include leucine to phenylalanine at position 234 and serine to proline at position 331 [86]. In HuIgG2, the reduced affinity is mediated by changes in leucine (positions 234, 235), glycine (position 237), and threonine (position 337 [87,88]). Proteolysis shows the binding of rabbit IgGs to human $Fc_{\gamma}RI$ to be mediated by rabbit IgG $C_H3$ domain with little role for the $C_H2$ domain (but molecular analysis was not done) [85]. Binding of human IgGs to $Fc_{\gamma}RIII$ is also determined by the $C_H2$ domain [90].

Differences in Fc receptor γ-chain transmembrane domains (particularly polar or charged residues among the C-terminal 11 amino acids) influence cell membrane receptor expression and function [91]. The charged residues might protect $Fc_{\gamma}RI$ and $Fc_{\gamma}RIIIa$ receptors during membrane-associated receptor recycling in proteasomes associated with the endoplasmic reticulum or other intracellular structures [91].

FcRn binding is affected by amino acid changes in the $C_H2$ and $C_H3$ domains of IgG molecules, including the arginine substitution for histidine at position 435 characteristic of HuIgG3 but not HuIgG1 or HuIgG2 [6]. Determinants of MsIgG1 FcRn binding include $C_H2 : C_H3$ residues isoleucine 253, histidine 310, glutamine 311, histidine 433, and asparagine 434 [4]. While determinants of HuIgG1 FcRn binding include isoleucine 253, histidine 310, and histidine 435, histidine 433 is not involved [92]. $Fc_{\gamma}RI$ stoichiometry allows binding of two IgGs (one per $C_H2$ domain); FcRn stoichiometry allows binding of only one IgG. Once FcRn binds to the $C_H2 : C_H3$ interface of the first IgG molecule, binding of the second IgG is prohibited as has been demonstrated for mouse antibodies [4,93,94].

***$Fc_{\gamma}R$ Polymorphisms*** FcR polymorphisms are correlated with susceptibility to some infectious and immune-mediated diseases, and they also might influence monoclonal antibody transport, clearance, function, and beneficial (antitumor) or adverse (cytokine-eliciting) responses [95–99]. In rats, genetic influences such as allotype have been shown to affect the function of fibroblast-expressed $Fc_{\gamma}RIIa$ [100]. In humans, eight genes on the long arm of chromosome 1 code for $Fc_{\gamma}RI$, $Fc_{\gamma}RII$, and $Fc_{\gamma}RIII$ genes. Genetic polymorphisms in $Fc_{\gamma}RIIa$ and $Fc_{\gamma}RIIIa$ appear to influence IgG binding to FcR, and they might influence therapeutic responses. For example, the clinical response of human patients to the humanized anti-CD20 monoclonal antibody rituximab correlates with natural killer cell $Fc_{\gamma}RIIIa$ polymorphisms and potentially with rituximab-armed NK antibody-dependent, cell-mediated cytotoxic responses [97,98]. Patients homozygous for valine at position 158 of the FCGR3A gene have better clinical responses than patients homozygous for phenylalanine at position 158 or heterozygotes [98]. Interestingly genetic

polymorphisms in human monocyte $Fc_{\gamma}RIIa$ affect in vitro T cell proliferation induced by chimeric mouse-human IgG2 anti-CD3 monoclonal antibodies [95], and they might also influence cytokine release following monoclonal antibody administration.

### 3.3 Effect of IgG Structure on Plasma IgG Clearance and Tissue Catabolism

The rate of IgG clearance is proportional to the plasma IgG concentration, with faster clearance rates for higher plasma IgG concentrations [1,3]. Isotype, subclass, and amino acid changes in FcR-binding regions affect clearance [93,101] as does overall carbohydrate structure and composition [102–104]. Although less than 3% of endogenous IgG consists of carbohydrate, the carbohydrate structure and composition influences rapid liver clearance and/or catabolism but does not affect the slower IgG catabolism typical of skin and muscle [104]. Some of these effects are thought to be mediated by hepatic removal of IgG with galactosyl moieties by hepatocyte asialoglycoprotein receptors [102,104]. Hepatic and/or splenic removal of mannose-containing IgGs is mediated by mannose receptors on hepatic Kupffer cells and hepatic and/or splenic sinusoidal endothelium, macrophages, and/or dendritic cells [105–107]. Other tissues with the potential for IgG catabolism and relatively rapid IgG loss across mucosal surfaces include kidney and lung [104]. The effect of IgG isotype and subclass on FcR-mediated transport and the effect of carbohydrate moieties on IgG binding to Fc receptors (FcR) on brain microglia and other cells are discussed further below.

The carbohydrate content of IgG might affect IgG binding to FcR as deglycosylated IgG antibodies bind brain microglial FcR with reduced affinity [44], although an earlier study was unable to demonstrate differences between intact and aglycosylated antibody with regard to FcR binding and signal transport [108]. Certainly different methods of production result in different glycosylation patterns (e.g., mammalian cell lines—CHO, NSO, etc.—yeast, transgenic animals), but it remains unclear what effect differences in glycosylation patterns (i.e., what carbohydrate moieties are present, degree of glycosylation, and order of carbohydrate moieties) have on antibody binding.

## 4 SPECIALIZED TISSUE-SPECIFIC, IgG TRANSPORT OR CLEARANCE MECHANISMS

Once an antibody has egressed into the interstitial space, its movement is somewhat more complicated than simple diffusion around the interstitial space (until it binds to epitope) or flow into lymphatics for mononuclear phagocytic clearance. There are a number of mechanisms by which antibodies (that are not bound via CDR to target or a cross-reactive epitope) are transported across and to a variety of cell types. Below are discussed several specialized and/or tissue-specific mechanisms of IgG transport and/or clearance.

### 4.1 FcRn

Endogenous IgG is present in luminal secretions of human and monkey gastrointestinal, oral, respiratory, and urogenital mucosal epithelium (e.g., intestine, pharynx, lung, kidney, uterus), which is normally very resistant to passive diffusion of macromolecules as part of innate barrier function. These mucosal epithelia have developed specialized mechanisms, including transport organelles (vesicles, endosomes, etc.), that mediate transcytosis of macromolecules, including immunoglobulins [109,110]. The most important route probably is MHC class I related Fc receptor (FcRn) mediated IgG transport across and within multiple epithelia, including human (adult and neonatal) small intestine [26,111–115]. FcRn consists of a 3-domain α-subunit noncovalently bound to $\beta_2$-microglobulin [26,110,113]. IgG salvage, sorting and/or recycling occurs when cystosolic endosomes take up IgG at the plasma membrane. Polarized bidirectional transport has been described for human intestinal epithelium with FcRn and its associated endosomes localized in the apical cytoplasm in a punctate pattern (just below the apical membrane or zonula occludens tight junctions) with the apical pathway of transcytosis more efficient than the basolateral pathway [109]. Similarly FcRn is located apically in human (and mouse) placental syncytiotrophoblasts, where it mediates IgG uptake from the maternal circulation [28], and in human renal proximal tubule cells, where it mediates IgG salvage from the urinary filtrate [116]. The pH dependence of FcRn binding [26,109,117] is attributed to FcRn presence/function within acidic recycling and/or sorting endosomes [118–120], some of which become multivesicular bodies or are degraded in lysosomes and others of which release IgG to the environment either by dispersal into the plasma membrane or by sustained release from endosomes as they approach the plasma membrane from the peripheral cytoplasm [119,120]. Binding to the FcRn within the acidic endosomes presumably protects the sorted IgG from catabolism, allowing release of the salvaged IgG to plasma, interstitium, dendritic, and/or target cells at neutral pH. The process of FcRn-mediated transcytosis or transcellular transport thus consists of three phases: endocytosis, intracellular transport, and exocytosis, all of which might be pertinent to therapeutic monoclonal antibody development and administration [8].

The Fc portion ($Fc_\gamma$) of IgG is essential for FcRn-mediated transcytosis. Monoclonal antibodies have been designed with $Fc_\gamma$ modifications that affect the affinity or pH of binding to FcRn and that can reduce FcRn-mediated salvage, and hasten clearance of other (endogenous or exogenously administered) IgG antibodies [121]. Such antibodies might be developed as therapeutic candidates to avert autoantibody-mediated damage.

FcRn was originally known as the Brambell receptor, based on multiple studies conducted by F. W. Brambell involving maternal transmission of antibodies across placenta and neonatal intestine in rabbits [3,5,122,123]. Transport was saturatable with increases in plasma IgG leading to reductions in plasma half-life ($t_{1/2}$) according to the formula

$$t_{1/2} = \frac{C \ln 2}{a(C-b)},$$

where $a = +0.34$, $b = 2.5$, and $C$ = total IgG concentration (mg/ml).

In adult mice, FcRn is expressed on endothelium in many tissues as well as on hepatocytes, where it serves a protective, anticatabolic, function by compartmentalizing, salvaging, and recycling IgG, but it does not mediate biliary clearance [4,62,124]. In mice, FcRn is preferentially expressed in cytoplasmic vesicles in small vessel endothelium (e.g., small arterioles, venules, and capillaries, liver sinusoids) skin, muscle, and liver, but it is not expressed by larger vessel endothelium (e.g., liver central veins or portal vessels in liver [62]) consistent with primary sites of IgG transport and catabolism in rodents [64]. FcRn is expressed on human placental endothelium [125] and also assumed to be expressed on widespread endothelium in adult humans but only secondary and not primary sources were reviewed herein [1,126]. FcRn is also expressed by monocytes, macrophages (including about one-fourth of intestinal macrophages), and dendritic cells; binds immune complexes; and is postulated to function in the acidic pH that characterizes inflammation or the tumor interstitium [26,127]. FcRn is preferentially expressed intracellularly by the majority of peripheral blood monocytes, with lower levels associated with the cell membrane [26]. The heavy chain of FcRn in humans is noncovalently associated with membrane molecules such as MHC class I molecule $\beta_2$-microglobulin expressed on endothelium, epithelium, leukocytes, placental syncytiotrophoblasts, dendritic cells, and macrophages as well as CD68 (macrosialin) and Nel-Macro on intestinal macrophages [26,110,128].

FcRn-mediated protection for endogenous IgG and exogenously administered human monoclonal antibodies is plasma concentration-dependent and epitope-independent, as has been reviewed recently [1]. However, Junghans et al. [129] suggest that human FcRn (in its Brambell receptor protective role) may differentially transport and catabolize antigen–antibody complexes formed between immune endogenous IgG and exogenously administered human and chimeric monoclonal antibodies via differential catabolism of antigen–antibody complexes.

FcRn binding is not affected by carbohydrate residues on IgG [130]. Last, at the time of this review there is no information regarding FcRn expression on surface lining cells (e.g., mesothelial cells, meningeal cells, or synoviocytes) available in the literature. Passive (nontranscytotic) mesothelial cell transport of IgG is discussed further below.

### 4.2 pIgR on Epithelium and Other Cell Types

The polymeric IgA receptor/secretory component (pIgA-R/SC, also designated pIgR) mediated transport represents another, less common and less utilized, form of specialized endosomal epithelial transcytosis. pIgR are found

primarily on mucosal epithelium, including biliary epithelium, and hepatocytes in humans and rodents [131]. Immunofluorescence has demonstrated the secretory component (SC) on human intestinal and respiratory goblet cells, other mucosal epithelia (gall bladder, renal tubules), and gland and/or duct epithelia (pancreas, sweat glands [132,133]). In rodents, hepatocytes produce a secretory component, allowing dimeric IgA to attach to hepatocyte sinusoidal membrane pIgR and be translocated via cytoplasmic vesicles to the biliary canalicular membrane for excretion into bile [134,135].

Unlike FcRn-mediated transcytosis of IgG, pIgR is unidirectional and transports dimeric IgA (approximately 3 g per day in an adult human) more frequently than pentameric IgM. IgG transport occurs when monomeric IgG is complexed with polymeric IgA [136]. Information is not available as to whether pentameric IgM or IgM fragments can complex with monomeric IgG and facilitate transport IgG-IgM complex transport via pIgR. However, following monoclonal IgG antibody administration, heterologous (or less often homologous) IgM antimonoclonal IgG antibody responses can be detected. Thus, while pIgR-mediated transport of monoclonal IgG antibodies is likely to be infrequent in the homologous species (e.g., human antibodies in humans), it appears to occur more often as immune complexes in heterologous species used for efficacy or toxicity testing. The presence of pIgR on hepatocytes as well as mucosal surfaces might foster rapid liver clearance of exogenously administered IgG antibodies if antihuman IgM antibodies were present prior to, or as a result of, monoclonal IgG antibody administration in the heterologous species. Alternatively, both homologous and heterologous immunoglobulins can be taken up by hepatocyte asialoglycoprotein receptors (discussed with liver) and enter lysosomes for degradation [135].

Unlike FcRn-mediated transcytosis, the initial pIgR-Ig binding occurs at the cell surface rather than within endosomes, and usually leads to receptor degradation or loss to the apical environment rather than recycling [61,109]. Internalization occurs via a nocodozole-sensitive, microtubule-associated process [137,138].

pIgR is thought to be most prevalent in embryogenesis with less utilization in perinatal or adult human tissues. Immunoglobulins, secretory component, and the associated J chain are described in multiple embryonic tissues, including gastrointestinal, respiratory and urogenital tracts, gonads, endocrine tissues, mesothelial-lined surfaces, and functional blood–tissue barriers. As gestation progresses, the tissues remodel, and the distances between many mucosal and glandular epithelium increase while the distribution and function of pIgR/SC-mediated transport decrease [139]. pIgR is absent from adult human glomerular mesangial cells [140] but has been demonstrated on cultured human endometrial cells [141]. In adult mice, pIgR mediates Ig transport, principally of IgA and IgM, from the basolateral surface to the apical surface and lumen of intestine and bile canaliculi and ducts where it contributes to mucosal defense [142]. High concentrations of IgA and IgG are found in human and rat bile [143,144] and have been attributed to Ig transport by pIgR and FcRn,

respectively [142,145], although nonspecific macromolecular transport (paracellular sieving [145]) might contribute when the liver is inflamed and IgG is produced locally.

Most of the mechanistic work on pIgR has focused on IgA transport rather than IgG transport. In polarized epithelial cells, receptors that mediate transcytosis (e.g., pIgR) are separated from other basolateral receptors (e.g., transferrin receptor) following delivery to recycling endosomes, although nearly 65% are re-transported rapidly to the basolateral surface via early endosomes [146]. About 35% of transferrin–transferrin receptor complexes and dimeric IgA–pIgR complexes enter into submembranous early endosomes and later co-localize into perinuclear recycling endosomes for a slower separatory phase [146]. The net transport of IgA is affected by four overlapping processes: surface clearance, endocytosis, basolateral recycling, and transcytosis [146].

pIgRs also are involved in disease pathogenesis. Pneumococci directly interact with pIgR to gain access to, or traverse, human respiratory epithelial cells [147,148]. Binding of Epstein–Barr virus (EBV) to EBV-specific IgA triggers internalization of the EBV–IgA complex by nasal epithelial cells and might initiate persistent nasal infection [149,150] cytomegalovirus (CMV) to CMV-specific IgA triggers internalization of the CMV–IgA complex.

### 4.3 Kidney

The glomerular capillary endothelial bed receives 23% of the cardiac output and serves as the primary blood filter. The glomerular capillary endothelium is fenestrated with discontinuous tight junctions and/or gap junctions (Kuhn et al., 1975) and glomerular endothelial cells express low affinity FcR distinct from $Fc_{\gamma}RI$, $Fc_{\gamma}RII$, and $Fc_{\gamma}RIII$ [151] now recognized as FcRn (reviewed in Lobo et al. [1]). This engenders direct contact between the underlying or intervening mesangial cells and the intravascular fluid within the glomerular capillary lumens, particularly at the glomerular angles [11,12,152,153], and both endothelium and mesangial cells probably participate in Ig recycling. The gap junctions of the mesangial cells [152] might facilitate this transport. Furthermore IgG molecules or their fragments can cross via the large pores in the glomerular basement membrane in healthy individuals [154]. Therefore the passage of exogenously administered monoclonal IgG test articles across the glomerular capillary endothelium into the mesangium might be expected, independent of antibody CDR specificity. Fab and scFv fragments have shorter plasma half-lives (0.5 to 21 hours) due to quicker glomerular filtration and clearance [7].

The mesangium also participates in removal of Ig (IgM, IgA > IgG) or immune complexes from blood via FcR (including $Fc_{\gamma}RI$ and $Fc_{\gamma}RIIIa$ on activated mesangium [72,155]) in humans and likely mediates nonspecific transport of macromolecules including Ig, possibly via integrin-mediated asso-

ciation with FcR-positive migrating leukocytes [153,156–158]. Glomerular epithelial cells (which also are in contact with mesangial cells as well as with the urinary filtrate in the urinary space) might complete the cycle of IgG transport and recycling by neonatal Fc receptor (FcRn)-mediated transcytosis [116] (see FcRn above). Alternatively, subepithelial deposition of immune complexes might occur subsequent to removal of immune complexes by attachment to FcRn on glomerular visceral epithelial cells (podocytes [116]). Last, exogenously administered monoclonal antibodies have been shown to delay the renal clearance of soluble epitopes from plasma (e.g., anti-Tac delays clearance of shed Tac [IL-2 receptor-alpha {IL-2Rα}] in mice [159]). Differential renal catabolism of the Tac–anti-Tac immune complex and the uncomplexed anti-Tac monoclonal antibody might be influenced by FcRn-mediated sorting/recycling in intracellular vesicles [129,159] (see FcRn below). There apparently is no contribution by pIgR or the asialoglycoprotein receptor as these structures are absent from adult human glomerular mesangial cells [140].

Renal proximal tubular epithelium (e.g., convoluted tubules) reclaim protein, including immunoglobulin, from the urinary filtrate via sorting/recycling mechanisms in intracellular vesicles. The apical (lumen-facing) surface of these cells features a specialized brush border which expresses FcRn closely associated with $\beta_2$-microglobulin [110,116]. Studies with cultured human renal proximal tubular epithelial cells have demonstrated bidirectional IgG transport and salvage by fully functional FcRn [110] as discussed further in the next section.

### 4.4 Breast (Mammary Gland)

Breast epithelium is responsible for the physiologic secretion of IgG into milk and, to a lesser extent, into nonlactating breast secretions. Although the biodistribution of exogenously administered monoclonal antibodies into breast milk or nonlactating breast secretion is not well understood, bovine IgG (and/or fragments thereof) is found in breast milk after dietary exposure [160]. This indicates the potential for transport of exogenously administered monoclonal antibodies (and/or antibody fragments) into breast milk or nonlactating breast secretion by physiologic mechanisms. FcRn have also been demonstrated on neoplastic human breast epithelial cells in tumors and lymph node metastases as well as morphologically normal adjacent epithelium but not tumor or adjacent endothelium [161].

### 4.5 Ovary

IgG is found in ovarian (intra)follicular fluid or corpus lutea secondary to ovarian (ovulation) hemorrhage or macrophage-mediated transport. Macrophages have been localized to ovarian stroma, to thecal layers of ovarian follicles, and within early, mid-, late, and regressing corpora lutea [162].

### 4.6 Placenta and Uterus

Different FcRs have different roles to play in maternal IgG transport across placental barriers in humans. $Fc_{\gamma}RI$, $Fc_{\gamma}RIIa$, $Fc_{\gamma}RIIb$, and $Fc_{\gamma}RIII$ are expressed at the surface of Hofbauer cells (placental macrophages), $Fc_{\gamma}RIIb$ and FcRn are expressed by chorionic (fetal) endothelium, and $Fc_{\gamma}RIII$ and FcRn are expressed by trophoblasts (fetal chorionic epithelium [28,29]). FcRn expression by syncytiotrophoblasts likely mediates a significant portion of maternal IgG transport in humans and IgG; $\beta_2$-microglobulin, and FcRn co-localize in plasma membrane granules or vesicles [28,29,128,163,164]. IgA and IgM do not cross placental barriers and are absent from fetal circulation. In contrast, fetal accumulation of maternal IgG (particularly IgG1) begins by human gestation week 16. Steady transplacental IgG increases lead to equivalent levels of maternally derived IgG1 in maternal and fetal circulation by gestation week 26, and greater IgG1 levels in fetal compared to maternal circulation at delivery. IgG3 and IgG4 levels are usually equivalent in maternal and fetal circulation at term while IgG2 levels remain greater in maternal circulation compared to fetal circulation [28,165]. The preference for IgG1 transport has also been shown using in vitro models of placental transport [166,167] and is reflected in the preference of high-affinity $Fc_{\gamma}RI$ for HuIgG1, HuIgG3 > HuIgG4 >> HuIgG2 [84–88]. The volume of IgG transported is consistent with in vitro calculations of placental binding site capacity ($10^{15}$ sites/mg) and affinity ($10^{6}$ M) using paired fetal side-oriented and maternal side-oriented placental membrane vesicles [168]. Last, the discovery of FcRn by Brambell was based on rabbit studies in which whole IgG and Fc fragments easily crossed from mother to fetus, whereas F(ab′)2 fragments failed to cross [3,5,122,123].

Functional pIgR has been described on a human endometrial cell line [141], but its role in maternal–fetal IgG transport or transcytosis in the nonpregnant uterus has not been described.

Immunoglobulin profiles of human cervical mucus indicate approximately twice as much IgG (30 mg/dl) as IgA (15 mg/dl) overall; however, there are both biphasic and menstrual cycle influences on Ig levels [169]. Different from blood plasma, the IgA2 subclass predominates in female genital tract secretions with lesser amounts of IgGA1 [170].

### 4.7 Seminal Fluid

IgG is more concentrated in seminal fluid than IgA, and only low levels of IgM are evident. As in blood plasma, more IgA1 than IgA2 is evident in seminal fluid.

### 4.8 Lung

Normal pulmonary (bronchoalveolar) lavage fluid contains IgG at lower concentrations (approximately 1%) than plasma. IgA is selectively concentrated

relative to plasma proteins in bronchoalveolar lavage fluid, as is indicative of epithelial transcytosis rather than just leakage from vessels [171].

By immunoelectron microscopy, endogenous IgG has been localized to epithelial membranes, tubular myelin, and granular material within alveoli [172,173]. The major barrier to bidirectional air : blood exchange is composed of type I and type II alveolar epithelial cells (distal airways) that form zonulae occludens (tight junctions). Paracellular (between cells) diffusion of macromolecules is considered very slight and probably insignificant in healthy (uninflamed) lung regardless of direction (uptake or egress). Endothelium and interstitium contribute little to the blood–air–blood barrier. Alveolar protein (including IgG) transport is reduced by increasing molecular size and may be affected by protein catabolism (apical surface peptidases on alveolar lining epithelium [173]).

The rate of protein clearance has been estimated as 10% of the rate of fluid clearance from alveoli [173]. IgG clearance is probably mediated by FcRn transcytosis in distal type I alveolar epithelium and more proximal bronchial epithelium. Type I alveolar epithelium and bronchial epithelium contain the necessary subcellular structures for FcRn-mediated transcytosis: vesicles, membrane invaginations, caveolae, and clathrin-coated pits [173,174]. FcRn mRNA is expressed in lung although the cell types and locations have not yet been determined [112]. Moreover, primary alveolar epithelial monolayer cell cultures express functional FcRn [173]. pIgA–R/SC transcytosis is thought to contribute little to distal (alveolar) airway IgG transport but might mediate more proximal (bronchial or bronchiolar) IgA transport [173]. Uptake of an aerosolized IgG Fc-erythropoietin fusion molecule and subsequent erythropoietin-induced reticulocytosis has been demonstrated in human and nonhuman primates [175].

### 4.9 Skin

FcγRIIa (CD32a) is expressed in small-vessel endothelium in the papillary (superficial and periadnexal) dermis [27]. Binding of IgG or immune complexes to isolated human dermal microvessel endothelial cells initiates rapid receptor cross-linking, intracellular Ca fluxes, and endosomal internalization of the IgG–$Fc_{\gamma}$RIIa complexes [27]. FcRn is expressed principally in basal and suprabasal human keratinocytes in normal epidermis [176] and is also expressed in mouse skin [62]. In healthy adult human volunteers, dermal interstitial fluid endogenous IgG levels were demonstrated to be 30% of plasma IgG levels [59].

### 4.10 Liver

FcRn protein is expressed by rat hepatocytes, particularly at the canalicular surface [145] and might facilitate serum-to-bile IgG transport or prevent

catabolism by hepatocytes [124,145]. The liver also has other specialized mechanisms for clearance and/or transport of glycoproteins and immunoglobulins–secretory component and the asialoglycoprotein receptor (ASGP-R). In the rat, asialoglycoproteins are taken up by liver hepatocytes for lysosomal degradation via the asialoglycoprotein receptor, while IgA, which is the predomiant immunoglobulin involved in immunological protection at mucous membrane surfaces, is cleared from plasma and transported through hepatocytes and into the bile via secretory component [135,177,178]. In humans, there is less hepatic transport of IgA because secretory component is expressed by biliary epithelium and there is more local synthesis of IgA within hepatobiliary tissues [177,178]. Humans are therefore unable to clear IgA–antigen complexes via the hepatic transport pathway, and this might contribute to development of IgA–mediated immune complex diseases [178]. For example, alcoholic liver disease is associated with deposition of IgA along the liver sinusoids in the majority of alcoholics [178]. Although there is considered to be minimal uptake of asialoglycoproteins by secretory component or IgA by ASGP-R, there is some evidence that heterologous immunoglobulin, specifically human IgA administered to rats, can bind to both receptors [135]. Many therapeutic glycoproteins including antibodies may further contain terminal GlcNAc and/or Gal residues, which facilitate rapid clearance as a result of binding to mannose receptor and/or ASGP-R in the liver [179]. As described above, the liver is involved in the first-pass effect because of to its high vascularity and high cardiac output, so, it is probable that ASGP-R and/or secretory component plays a role in therapeutic (or exogenous) antibody transport or clearance.

### 4.11 Salivary Gland

Saliva contains IgA (0.6–49.2 mg/dL), IgG (0.03–4.86 mg/dl), small amounts of IgM (0.04–2.6 mg/dl), and secretory component [180]. pIgR expression has been described for human parotid salivary gland and is responsible for epithelial cell transport of IgA and secretory component into saliva [180,181]. The overwhelming majority (92–99%) of transported IgA is synthesized locally, while approximately half of transported IgG derives from serum with consistent accumulations in salivary interstitium [182]. IgA transport into saliva is increased by chewing [181]. The means of IgG transport into saliva has not been described; however, several authors have proposed local IgG production and nonspecific (transudative or paracellular sieving) of mucosal or serum IgG leaked from inflamed vessels into saliva or other oral fluids in reponse to viral or bacterial infections [180,183–185]. IgA, IgM, and IgG transport into saliva is reduced by some types of stimulation [180]. Increases in salivary IgG and IgM are also seen in patients with immunoglobulin-producing plasmacytomas or paraproteinemias [180,186].

The timing of IgG appearance in saliva (and nasal secretions) is delayed compared to its appearance in serum. Following subcutaneous administration

of tetanus toxoid, peak anti-toxin IgG serum levels preceded peak anti-toxin IgG salivary and nasal secretion levels by two weeks. The levels of interstitial anti-toxin IgG were not measured [187].

### 4.12 Colon

In normal healthy volunteers, immunoglobulins are secreted into the colon lumen at the following rates: IgG (48 mg/day), polymeric IgA (220 mg/day), monomeric IgA (12 mg/day), and IgM (24 mg/day). For comparison, values for nonimmunoglobulin proteins were also calculated (e.g., albumin at 150 mg/day) by Prigent-Delecourt and colleagues [188].

### 4.13 Thyroid

Zimmer and colleagues [189,190] have reported that IgG is transported via epithelial transcytosis from the pericapillary basolateral aspect to the apical membrane of the follicular epithelium. Within the thyroid follicular epithelium, IgG has been localized to endoplasmic reticulum and Golgi by way of immunogold electron microscopy. IgGs also have been localized to the thyroid follicular lumen and at the apical membrane of the thyroid follicular epithelial cells [189,190]. Moreover normal thyroid follicular epithelium expresses FcRn while Graves's disease thyroid follicular epithelium expresses $Fc_{\gamma}RIIb$ as well as FcRn [191]. These molecules are thought to mediate IgG transport from the pericapillary basolateral aspect of the follicular cells to the apical membrane, although direct basal to apical movement has not been clearly demonstrated [191]. IgG transport is accelerated in the immune-mediated thyroid disease (e.g., Graves's disease), particularly in more susceptible females. The gender-dependence of susceptibility may be affected by androgen-mediated inhibition of $Fc_{\gamma}RIIb$ epithelial expression and transcytosis [191]. These associations might be particularly important in light of the localization of HLA-restricted class I and class II molecules and processed thyroglobulin to the same subcellular sites. This might facilitate antigen presentation or antigen–antibody complex formation leading to thyroid follicular epithelial cell damage [191]. In humans, pIgA-R/SC is also expressed by fetal thyroid, but it has not been described in adult thyroid [139] and has not been accorded a role in immune-mediated thyroid disease.

Thyroid clearance of exogenously administered monoclonal antibodies is reported to be affected by the concentration and injection volume of the administered drug. Following intraperitoneal administration of irrelevant CDR-containing mouse IgG2a monoclonal antibody to rats, thyroid uptake of radiolabeled mouse IgG2a monoclonal antibody has been demonstrated for higher doses and more concentrated solutions while lower doses or less concentrated solutions remain within blood vessels and do not penetrate the thyroid parenchyma [192].

### 4.14 Cytosolic Transport in Epithelium, Mesothelium, and Other Tissues

In epithelium, cytosolic or non–receptor-mediated transport is inefficient and often leads to lysosomal degradation of the macromolecule [109]. However, macromolecules, including IgG, are readily transported across the mesothelial surface of the peritoneum. The mesothelial lining cells are flat cells with pinocytotic vesicles but no caveolae comparable to endothelial cells [193]. Gap junctions separate interdigitated peritoneal mesothelial cells. IgG transport is known to be bi-directional and passive in rats [194–198] and also occurs in chronic peritoneal dialysis in humans [199].

There was no information regarding FcRn expression on mesothelial cells, meningeal cells, or synoviocytes available in the literature.

### 4.15 Erythrocyte CR1 (CD35) Clearance of IgG-Containing Immune Complexes

In primates, additional mechanisms exist for clearance of immune complexes. Erythrocyte CR1 (CD35) bind immune complexes via C3b and C4b ligand sites. Once bound to erythrocyte CR1, the immune complexes are transported to the liver where they are cleared by liver macrophages (Kupffer cells). Mouse antibody (MsIgG1-, MsIgG2a-, MsIgG2b-, and MsIgA-) containing immune complexes bind rapidly to erythrocyte CR1 and peak magnitude binding is high. MsIgG3- and MsIgM-containing immune complexes bind slowly to erythrocyte CR1, and the peak magnitude binding is low. Release rates are rapid for MsIgG2a- and MsIgG2b-containing immune complexes, intermediate for MsIgG1-containing immune complexes, relatively slow for MsIgA- and MsIgM-containing immune complexes, and very slow for MsIgG3-containing immune complexes [200]. Similarly, Kavai and colleagues [201] demonstrated reduced incorporation of C3b-iC3b into IgM-containing immune complexes compared to IgG-containing immume complexes. In general, high-affinity mouse antibodies activate complement better than, but bind to erythrocyte CR1 less well than, low-affinity mouse antibodies [202].

## 5 CONCLUSION

The movement of IgG through tissue is more complex than its circulation through the vasculature, its binding (or not binding) to target or off-target epitope, and its clearance. As described above, antibodies can move out of the vasculature and be carried through various cell types through interactions with a number of different molecules that are expressed in many different tissue types. As such, many tissue and/or cellular compartments are available to therapeutic antibodies—vascular, interstitial, cell membrane, and cytoplasmic. During the course of therapeutic antibody development, tissue cross-reactivity testing is required

(see Chapter Tissue Cross-Reactivity Studies for Monoclonal Antibodies: Predictive Value and Use for Selection of Relevant Animal Species for Toxicity Testing). In many cases the test article (therapeutic candidate) stains only expected tissue elements (cell types and subcellular locations). However, some test articles display tissue cross-reactivity with unexpected cell types and/or subcellular locations. Although tissue cross-reactivity results are not necessarily considered to be predictive of toxicity, they do yield information as to potential tissue sites that warrant closer scrutiny and/or additional evaluation during preclinical toxicity studies or early phase clinical trials. As a follow-up to tissue cross-reactivity testing, tissues submitted for histopathological evaluation during the toxicity studies can be stained using specialized techniques for detection of the test article (therapeutic candidate) to take a snapshot of its biodistribution at any given time point during the toxicity study. We have conducted approximately 20 of these studies in cynomolgus monkeys, rhesus monkeys, and marmosets. However, because of the proprietary nature of the work, specific details cannot be provided. In general, the test article (therapeutic candidate) was localized either at sites of expected or tissue cross-reactive staining as predicted by the tissue cross-reactivity study or at sites consistent with physiological IgG biodistribution, transport, and clearance. For most studies of this kind, the test article (therapeutic candidate) was localized at a variable subset of sites consistent with both the tissue cross-reactivity study results and the normal physiologic processes described above. Similar distribution patterns were recognized for exogenously administered antibody fragments in mice [62]. Specifically, $F(ab')_2$ fragments prepared from preimmune serum were cleared by renal filtration, whereas $F(ab')_2$ fragments prepared from anti-FcRn serum tended to localize in skin and muscle (and to a lesser extent, adipose tissue and liver) consistent with the distribution of FcRn [62]. These types of studies highlight the complexity of biodistribution, transport, and clearance of exogenous IgG. How these factors affect potential toxic effects of any therapeutic antibody is still unclear for many marketed antibody-based therapeutics and therapeutic candidates. However, greater understanding of IgG biodistribution, transport, and clearance will facilitate more efficient development of safer, more effective antibody-based therapeutics.

## REFERENCES


1. Lobo ED, Hansen RJ, Balthasar JP. Antibody pharmacokinetics and pharmacodynamics. *J Pharm Sci* 2004;93(11):2645–68.
2. Roitt I. *Essential Immunology*. Oxford: Blackwell Scientific, 1994.
3. Brambell FW, Hemmings WA, Morris IG. A theoretical model of gamma-globulin catabolism. *Nature* 1964;203:1352–4.
4. Ghetie V, Hubbard JG, Kim JK, Tsen MF, Lee Y, Ward ES. Abnormally short serum half-lives of IgG in beta 2-microglobulin-deficient mice. *Eur J Immunol* 1996;26(3):690–6.

5. Junghans RP. Finally! The Brambell receptor (FcRB). Mediator of transmission of immunity and protection from catabolism for IgG. *Immunol Res* 1997;16(1): 29–57.
6. Ghetie V, Ward ES. Transcytosis and catabolism of antibody. *Immunol Res* 2002; 25(2):97–113.
7. Roskos LK, Davis CG, Schwab GM. The clinical pharmacology of therapeutic monoclonal antibodies. *Drug Dev Res* 2004;61:108–20.
8. Peterson E, Owens SM, Henry RL. Monoclonal antibody form and function: manufacturing the right antibodies for treating drug abuse. *AAPS J* 2006;8(2): E383–90.
9. Bard P. *Medical Physiology*, 11th ed. St. Louis: Mosby, 1961.
10. Forsyth RP, Nies AS, Wyler F, Neutze J, Melmon KL. Normal distribution of cardiac output in the unanesthetized, restrined rhesus monkey. *J Appl Physiol* 1968;25(6):736–41.
11. Venkatachalam MA, Kriz W. Anatomy. In Heppinstall RH, ed. *Pathology of the Kidney*, 4th ed. Boston: Little, Brown, 1992;1–92.
12. Tisher CC, Madsen KM. Anatomy of the kidney. In Brenner MD, ed. *Brenner and Rector's The Kidney*, 5th ed. Philadephia: WB Saunders, 1996;3–71.
13. Matthews DC, Appelbaum FR, Eary JF, et al. Radiolabeled anti-CD45 monoclonal antibodies target lymphohematopoietic tissue in the macaque. *Blood* 1991;78(7):1864–74.
14. Matthews DC, Appelbaum FR, Eary JF, et al. Phase I study of (131)I-anti-CD45 antibody plus cyclophosphamide and total body irradiation for advanced acute leukemia and myelodysplastic syndrome. *Blood* 1999;94(4):1237–47.
15. el-Kareh AW, Braunstein SL, Secomb TW. Effect of cell arrangement and interstitial volume fraction on the diffusivity of monoclonal antibodies in tissue. *Biophys J* 1993;64(5):1638–46.
16. Rippe B, Haraldsson B. Transport of macromolecules across microvascular walls: the two-pore theory. *Physiol Rev* 1994;74(1):163–219.
17. Simionescu M, Antohe F. Functional ultrastructure of the vascular endothelium: changes in various pathologies. *Handb Exp Pharmacol* 2006;(176 Pt 1):41–69.
18. Pries AR, Kuebler WM. Normal endothelium. *Handb Exp Pharmacol* 2006; 176(Pt 1):1–40.
19. Simionescu M, Gafencu A, Antohe F. Transcytosis of plasma macromolecules in endothelial cells: a cell biological survey. *Microsc Res Tech* 2002;57(5):269–88.
20. Palade GE, Simionescu M, Simionescu N. Structural aspects of the permeability of the microvascular endothelium. *Acta Physiol Scand* Suppl 1979;463:11–32.
21. Schnitzer JE. gp60 is an albumin-binding glycoprotein expressed by continuous endothelium involved in albumin transcytosis. *Am J Physiol* 1992;262(1 Pt 2): H246–54.
22. Tiruppathi C, Song W, Bergenfeldt M, Sass P, Malik AB. Gp60 activation mediates albumin transcytosis in endothelial cells by tyrosine kinase-dependent pathway. *J Biol Chem* 1997;272(41):25968–75.
23. Demignot S, Pimm MV, Baldwin RW. Comparison of biodistribution of 791T/36 monoclonal antibody and its Fab/c fragment in BALB/c mice and nude mice bearing human tumor xenografts. *Cancer Res* 1990;50(10):2936–42.

24. Tomiyama Y, Kunicki TJ, Zipf TF, Ford SB, Aster RH. Response of human platelets to activating monoclonal antibodies: importance of Fc gamma RII (CD32) phenotype and level of expression. *Blood* 1992;80(9):2261–8.
25. Thai IM, Ashman LK, Harbour SN, Hogarth PM, Jackson DE. Physical proximity and functional interplay of PECAM-1 with the Fc receptor Fc gamma RIIa on the platelet plasma membrane. *Blood* 2003;102(10):3637–45.
26. Zhu X, Meng G, Dickinson BL, et al. MHC class I-related neonatal Fc receptor for IgG is functionally expressed in monocytes, intestinal macrophages, and dendritic cells. *J Immunol* 2001;166(5):3266–76.
27. Groger M, Fischer GF, Wolff K, Petzelbauer P. Immune complexes from vasculitis patients bind to endothelial Fc receptors independent of the allelic polymorphism of FcgammaRIIa. *J Invest Dermatol* 1999;113(1):56–60.
28. Saji F, Samejima Y, Kamiura S, Koyama M. Dynamics of immunoglobulins at the feto-maternal interface. *Rev Reprod* 1999;4(2):81–9.
29. Lyden TW, Robinson JM, Tridandapani S, et al. The Fc receptor for IgG expressed in the villus endothelium of human placenta is Fc gamma RIIb2. *J Immunol* 2001; 166(6):3882–9.
30. Minshall RD, Malik AB. Transport across the endothelium: regulation of endothelial permeability. *Handb Exp Pharmacol* 2006;176(Pt 1):107–44.
31. Ding BS, Dziubla T, Shuvaev VV, Muro S, Muzykantov VR. Advanced drug delivery systems that target the vascular endothelium. *Mol Interv* 2006;6(2):98–112.
32. Benarroch EE. Neuron-astrocyte interactions: partnership for normal function and disease in the central nervous system. *Mayo Clin Proc* 2005;80(10):1326–38.
33. Abbott NJ, Ronnback L, Hansson E. Astrocyte-endothelial interactions at the blood-brain barrier. *Nat Rev Neurosci* 2006;7(1):41–53.
34. Kim JH, Kim JH, Park JA, et al. Blood-neural barrier: intercellular communication at glio-vascular interface. *J Biochem Mol Biol* 2006;39(4):339–45.
35. Terasaki T, Ohtsuki S. Brain-to-blood transporters for endogenous substrates and xenobiotics at the blood-brain barrier: an overview of biology and methodology. *NeuroRx* 2005;2(1):63–72.
36. Jodoin J, Demeule M, Fenart L, et al. P-glycoprotein in blood-brain barrier endothelial cells: interaction and oligomerization with caveolins. *J Neurochem* 2003; 87(4):1010–23.
37. Gosk S, Vermehren C, Storm G, Moos T. Targeting anti-transferrin receptor antibody (OX26) and OX26-conjugated liposomes to brain capillary endothelial cells using in situ perfusion. *J Cereb Blood Flow Metab* 2004;24(11):1193–204.
38. Ulvestad E, Williams K, Bjerkvig R, Tiekotter K, Antel J, Matre R. Human microglial cells have phenotypic and functional characteristics in common with both macrophages and dendritic antigen-presenting cells. *J Leukoc Biol* 1994; 56(6):732–40.
39. Carson MJ, Reilly CR, Sutcliffe JG, Lo D. Mature microglia resemble immature antigen-presenting cells. *Glia* 1998;22(1):72–85.
40. Webster SD, Park M, Fonseca MI, Tenner AJ. Structural and functional evidence for microglial expression of C1qR(P), the C1q receptor that enhances phagocytosis. *J Leukoc Biol* 2000;67(1):109–16.

41. Webster SD, Galvan MD, Ferran E, Garzon-Rodriguez W, Glabe CG, Tenner AJ. Antibody-mediated phagocytosis of the amyloid beta-peptide in microglia is differentially modulated by C1q. *J Immunol* 2001;166(12):7496–503.
42. Song X, Shapiro S, Goldman DL, Casadevall A, Scharff M, Lee SC. Fcgamma receptor I- and III-mediated macrophage inflammatory protein 1alpha induction in primary human and murine microglia. *Infect Immun* 2002;70(9):5177–84.
43. Koenigsknecht-Talboo J, Landreth GE. Microglial phagocytosis induced by fibrillar beta-amyloid and IgGs are differentially regulated by proinflammatory cytokines. *J Neurosci* 2005;25(36):8240–9.
44. Rebe S, Solomon B. Deglycosylation of anti-beta amyloid antibodies inhibits microglia activation in BV-2 cellular model. *Am J Alzheimers Dis Other Demen* 2005;20(5):303–13.
45. Goldman DL, Casadevall A, Zuckier LS. Pharmacokinetics and biodistribution of a monoclonal antibody to Cryptococcus neoformans capsular polysaccharide antigen in a rat model of cryptococcal meningitis: implications for passive immunotherapy. *J Med Vet Mycol* 1997;35(4):271–8.
46. Pardridge WM, Buciak JL, Friden PM. Selective transport of an anti-transferrin receptor antibody through the blood–brain barrier in vivo. *J Pharmacol Exp Ther* 1991;259(1):66–70.
47. Schlachetzki F, Zhu C, Pardridge WM. Expression of the neonatal Fc receptor (FcRn) at the blood-brain barrier. *J Neurochem* 2002;81(1):203–6.
48. Hofman P, Hoyng P, vanderWerf F, Vrensen GF, Schlingemann RO. Lack of blood-brain barrier properties in microvessels of the prelaminar optic nerve head. *Invest Ophthalmol Vis Sci* 2001;42(5):895–901.
49. Bours J. The protein distribution of bovine, human and rabbit aqueous humour and the difference in composition before and after disruption of the blood/aqueous humour barrier. *Lens Eye Toxic Res* 1990;7(3–4):491–503.
50. Verhagen C, Breeboart AC, Kijlstra A. Diffusion of immunoglobulin G from the vascular compartment into the normal rabbit cornea. *Invest Ophthalmol Vis Sci* 1990;31(8):1519–25.
51. Marie I, Herve F, Lahaxe L, Robaday S, Gerardin E, Levesque H. Intravenous immunoglobulin-associated cranial pachymeningitis. *J Intern Med* 2006;260(2): 164–7.
52. Hadden RD, Gregson NA, Gold R, Smith KJ, Hughes RA. Accumulation of immunoglobulin across the "blood-nerve barrier" in spinal roots in adoptive transfer experimental autoimmune neuritis. *Neuropathol Appl Neurobiol* 2002;28(6): 489–97.
53. Vedeler CA. Demonstration of Fc gamma receptors on human peripheral nerve fibres. *J Neuroimmunol* 1987;15(2):207–16.
54. Vedeler CA, Matre R, Kristoffersen EK, Ulvestad E. IgG Fc receptor heterogeneity in human peripheral nerves. *Acta Neurol Scand* 1991;84(3):177–80.
55. Houston LL, Nowinski RC, Bernstein ID. Specific in vivo localization of monoclonal antibodies directed against the Thy 1.1 antigen. *J Immunol* 1980;125(2): 837–43.
56. Badger CC, Krohn KA, Shulman H, Flournoy N, Bernstein ID. Experimental radioimmunotherapy of murine lymphoma with 131I-labeled anti-T-cell antibodies. *Cancer Res* 1986;46(12 Pt 1):6223–8.

57. Simionescu D, Simionescu M. Differentiated distribution of the cell surface charge on the alveolar-capillary unit. Characteristic paucity of anionic sites on the air-blood barrier. *Microvasc Res* 1983;25(1):85–100.
58. Lossinsky AS, Shivers RR. Structural pathways for macromolecular and cellular transport across the blood–brain barrier during inflammatory conditions. Review. *Histol Histopathol* 2004;19(2):535–64.
59. Svedman C, Yu BB, Ryan TJ, Svensson H. Plasma proteins in a standardised skin mini-erosion (I): permeability changes as a function of time. *BMC Dermatol* 2002;2:3.
60. Moldoveanu Z, Epps JM, Thorpe SR, Mestecky J. The sites of catabolism of murine monomeric IgA. *J Immunol* 1988;141(1):208–13.
61. Mestecky J, Moro I, Underdown BJ. Mucosal Immunology. Ogra PL, Mestecky J, Lamm ME, Strober W, Bienenstock J, McGhee JR, eds. *Mucosal Immunoglobulins*, 2nd ed. 1998. London: Academic Press, 1998;133–52.
62. Borvak J, Richardson J, Medesan C, et al. Functional expression of the MHC class I-related receptor, FcRn, in endothelial cells of mice. *Int Immunol* 1998; 10(9):1289–98.
63. Waldmann TA, Strober W. Metabolism of immunoglobulins. *Prog Allergy* 1969; 13:1–110.
64. Henderson LA, Baynes JW, Thorpe SR. Identification of the sites of IgG catabolism in the rat. *Arch Biochem Biophys* 1982;215(1):1–11.
65. Olszewski WL, Engeset A. Capillary transport of immunoglobulins and complement proteins to the interstitial fluid and lymph. *Arch Immunol Ther Exp (Warsz)* 1978;26(1–6):57–65.
66. Roccatello D, Picciotto G, Ropolo R, et al. Kinetics and fate of IgA-IgG aggregates as a model of naturally occurring immune complexes in IgA nephropathy. *Lab Invest* 1992;66(1):86–95.
67. Fanger NA, Wardwell K, Shen L, Tedder TF, Guyre PM. Type I (CD64) and type II (CD32) Fc gamma receptor-mediated phagocytosis by human blood dendritic cells. *J Immunol* 1996;157(2):541–8.
68. Fanger NA, Voigtlaender D, Liu C, et al. Characterization of expression, cytokine regulation, and effector function of the high affinity IgG receptor Fc gamma RI (CD64) expressed on human blood dendritic cells. *J Immunol* 1997;158(7): 3090–8.
69. Ravetch JV. Fc receptors. *Curr Opin Immunol* 1997;9(1):121–5.
70. Guyre CA, Keler T, Swink SL, Vitale LA, Graziano RF, Fanger MW. Receptor modulation by Fc gamma RI-specific fusion proteins is dependent on receptor number and modified by IgG. *J Immunol* 2001;167(11):6303–11.
71. Van WJ, Lipinski P, Abadie A, et al. Antagonistic action of IFN-beta and IFN-gamma on high affinity Fc gamma receptor expression in healthy controls and multiple sclerosis patients. *J Immunol* 1998;161(3):1568–74.
72. Uciechowski P, Schwarz M, Gessner JE, Schmidt RE, Resch K, Radeke HH. IFN-gamma induces the high-affinity Fc receptor I for IgG (CD64) on human glomerular mesangial cells. *Eur J Immunol* 1998;28(9):2928–35.
73. van Royen-Kerkhof A, Sanders EA, Walraven V, et al. A novel human CD32 mAb blocks experimental immune haemolytic anaemia in FcgammaRIIA transgenic mice. *Br J Haematol* 2005;130(1):130–7.

74. Wiener E, Dellow RA, Mawas F, Rodeck CH. Role of Fc gamma RIIa (CD32) in IgG anti-RhD-mediated red cell phagocytosis in vitro. *Transfus Med* 1996;6(3): 235–41.

75. Metes D, Ernst LK, Chambers WH, Sulica A, Herberman RB, Morel PA. Expression of functional CD32 molecules on human NK cells is determined by an allelic polymorphism of the FcgammaRIIC gene. *Blood* 1998;91(7):2369–80.

76. Favaloro EJ. Differential expression of surface antigens on activated endothelium. *Immunol Cell Biol* 1993;71(Pt 6):571–81.

77. Coggeshall KM. Regulation of signal transduction by the Fc gamma receptor family members and their involvement in autoimmunity. *Curr Dir Autoimmun* 2002;5:1–29.

78. Micouin A, Rouillard D, Bauvois B. Induction of macrophagic differentiation and cytokine secretion by IgG1 molecules in human normal monocytes and myelogenous leukemia cells. *Leukemia* 1997;11(4):552–60.

79. Ravetch JV, Perussia B. Alternative membrane forms of Fc gamma RIII(CD16) on human natural killer cells and neutrophils. Cell type-specific expression of two genes that differ in single nucleotide substitutions. *J Exp Med* 1989;170(2): 481–97.

80. Hisaka H, Kataoka M, Higuchi Y, Matsuura K, Yamamoto S. Close localization of mouse CD14 and CD32/16 in the cell surface of monocytic cell lines. *Pathobiology* 1999;67(2):92–8.

81. de AB, Mueller AL, Verbeek S, Sandor M, Lynch RG. A regulatory role for Fcgamma receptors CD16 and CD32 in the development of murine B cells. *Blood* 1998;92(8):2823–9.

82. Delire M. Different modes of action of high-dose immunoglobulins in rheumatoid arthritis. *Acta Univ Carol [Med] (Praha)* 1994;40(1–4):95–9.

83. Muller C, Traindl O, Schwarz M, et al. Increased Fc-receptor-mediated clearance via reticuloendothelial system in patients with nephrotic syndrome. *Nephron* 1991;58(2):150–4.

84. Anderson CL, Abraham GN. Characterization of the Fc receptor for IgG on a human macrophage cell line, U937. *J Immunol* 1980;125(6):2735–41.

85. McCool D, Birshtein BK, Painter RH. Structural requirements of immunoglobulin G for binding to the Fc gamma receptors of the human tumor cell lines U937, HL-60, ML-1, and K562. *J Immunol* 1985;135(3):1975–80.

86. Canfield SM, Morrison SL. The binding affinity of human IgG for its high affinity Fc receptor is determined by multiple amino acids in the CH2 domain and is modulated by the hinge region. *J Exp Med* 1991;173(6):1483–91.

87. Chappel MS, Isenman DE, Everett M, Xu YY, Dorrington KJ, Klein MH. Identification of the Fc gamma receptor class I binding site in human IgG through the use of recombinant IgG1/IgG2 hybrid and point-mutated antibodies. *Proc Natl Acad Sci USA* 1991;88(20):9036–40.

88. Chappel MS, Isenman DE, Oomen R, Xu YY, Klein MH. Identification of a secondary Fc gamma RI binding site within a genetically engineered human IgG antibody. *J Biol Chem* 1993;268(33):25124–31.

89. Tan LK, Shopes RJ, Oi VT, Morrison SL. Influence of the hinge region on complement activation, C1q binding, and segmental flexibility in chimeric human immunoglobulins. *Proc Natl Acad Sci USA* 1990;87(1):162–6.

90. Morgan A, Jones ND, Nesbitt AM, Chaplin L, Bodmer MW, Emtage JS. The N-terminal end of the CH2 domain of chimeric human IgG1 anti-HLA-DR is necessary for C1q, Fc gamma RI and Fc gamma RIII binding. *Immunology* 1995; 86(2):319–24.

91. Kim MK, Huang ZY, Hwang PH, et al. Fcgamma receptor transmembrane domains: role in cell surface expression, gamma chain interaction, and phagocytosis. *Blood* 2003;101(11):4479–84.

92. Kim JK, Firan M, Radu CG, Kim CH, Ghetie V, Ward ES. Mapping the site on human IgG for binding of the MHC class I-related receptor, FcRn. *Eur J Immunol* 1999;29(9):2819–25.

93. Kim JK, Tsen MF, Ghetie V, Ward ES. Identifying amino acid residues that influence plasma clearance of murine IgG1 fragments by site-directed mutagenesis. *Eur J Immunol* 1994;24(3):542–8.

94. Popov S, Hubbard JG, Kim J, Ober B, Ghetie V, Ward ES. The stoichiometry and affinity of the interaction of murine Fc fragments with the MHC class I-related receptor, FcRn. *Mol Immunol* 1996;33(6):521–30.

95. Parren PW, Warmerdam PA, Boeije LC, Capel PJ, van de Winkel JG, Aarden LA. Characterization of IgG FcR-mediated proliferation Of human T cells induced by mouse and Human anti-CD3 monoclonal antibodies. Identification of a functional polymorphism to human IgG2 anti-CD3. *J Immunol* 1992;148(3):695–701.

96. Wu J, Edberg JC, Redecha PB, et al. A novel polymorphism of FcgammaRIIIa (CD16) alters receptor function and predisposes to autoimmune disease. *J Clin Invest* 1997;100(5):1059–70.

97. Koene HR, Kleijer M, Algra J, Roos D, von dem Borne AE, De HM. Fc gammaRIIIa-158V/F polymorphism influences the binding of IgG by natural killer cell Fc gammaRIIIa, independently of the Fc gammaRIIIa-48L/R/H phenotype. *Blood* 1997;90(3):1109–14.

98. Cartron G, Dacheux L, Salles G, et al. Therapeutic activity of humanized anti-CD20 monoclonal antibody and polymorphism in IgG Fc receptor Fc gamma RIIIa gene. *Blood* 2002;99(3):754–8.

99. Su K, Wu J, Edberg JC, McKenzie SE, Kimberly RP. Genomic organization of classical human low-affinity Fc gamma receptor genes. *Genes Immun* 2002;3 Suppl 1:S51–6.

100. Haagen IA, Geerars AJ, Clark MR, van de Winkel JG. Interaction of human monocyte Fc gamma receptors with rat IgG2b: a new indicator for the Fc gamma RIIa (R-H131) polymorphism. *J Immunol* 1995;154(4):1852–60.

101. Medesan C, Matesoi D, Radu C, Ghetie V, Ward ES. Delineation of the amino acid residues involved in transcytosis and catabolism of mouse IgG1. *J Immunol* 1997;158(5):2211–7.

102. Mattes MJ. Biodistribution of antibodies after intraperitoneal or intravenous injection and effect of carbohydrate modifications. *J Natl Cancer Inst* 1987; 79(4):855–63.

103. Wright A, Morrison SL. Effect of altered CH2–associated carbohydrate structure on the functional properties and in vivo fate of chimeric mouse-human immunoglobulin G1. *J Exp Med* 1994;180(3):1087–96.

104. Wright A, Sato Y, Okada T, Chang K, Endo T, Morrison S. In vivo trafficking and catabolism of IgG1 antibodies with Fc associated carbohydrates of differing structure. *Glycobiology* 2000;10(12):1347–55.

105. Lennartz MR, Cole FS, Shepherd VL, Wileman TE, Stahl PD. Isolation and characterization of a mannose-specific endocytosis receptor from human placenta. *J Biol Chem* 1987;262(21):9942–4.

106. Stahl PD. The mannose receptor and other macrophage lectins. *Curr Opin Immunol* 1992;4(1):49–52.

107. Dong X, Storkus WJ, Salter RD. Binding and uptake of agalactosyl IgG by mannose receptor on macrophages and dendritic cells. *J Immunol* 1999;163(10):5427–34.

108. Groenink J, Spijker J, van den Herik-Oudijk IE, et al. On the interaction between agalactosyl IgG and Fc gamma receptors. *Eur J Immunol* 1996;26(6):1404–7.

109. Dickinson BL, Badizadegan K, Wu Z, et al. Bidirectional FcRn-dependent IgG transport in a polarized human intestinal epithelial cell line. *J Clin Invest* 1999; 104(7):903–11.

110. Kobayashi N, Suzuki Y, Tsuge T, Okumura K, Ra C, Tomino Y. FcRn-mediated transcytosis of immunoglobulin G in human renal proximal tubular epithelial cells. *Am J Physiol Renal Physiol* 2002;282(2):F358–65.

111. Simister NE, Mostov KE. Cloning and expression of the neonatal rat intestinal Fc receptor, a major histocompatibility complex class I antigen homolog. *Cold Spring Harb Symp Quant Biol* 1989;54(Pt 1):571–80.

112. Simister NE, Mostov KE. An Fc receptor structurally related to MHC class I antigens. *Nature* 1989;337(6203):184–7.

113. Junghans RP, Anderson CL. The protection receptor for IgG catabolism is the beta2-microglobulin-containing neonatal intestinal transport receptor. *Proc Natl Acad Sci USA* 1996;93(11):5512–6.

114. Story CM, Mikulska JE, Simister NE. A major histocompatibility complex class I-like Fc receptor cloned from human placenta: possible role in transfer of immunoglobulin G from mother to fetus. *J Exp Med* 1994;180(6):2377–81.

115. Mikulska JE, Pablo L, Canel J, Simister NE. Cloning and analysis of the gene encoding the human neonatal Fc receptor. *Eur J Immunogenet* 2000;27(4): 231–40.

116. Haymann JP, Levraud JP, Bouet S, et al. Characterization and localization of the neonatal Fc receptor in adult human kidney. *J Am Soc Nephrol* 2000; 11(4):632–9.

117. Zhu X, Peng J, Raychowdhury R, Nakajima A, Lencer WI, Blumberg RS. The heavy chain of neonatal Fc receptor for IgG is sequestered in endoplasmic reticulum by forming oligomers in the absence of beta2-microglobulin association. *Biochem J* 2002;367(Pt 3):703–14.

118. Teter K, Chandy G, Quinones B, Pereyra K, Machen T, Moore HP. Cellubrevin-targeted fluorescence uncovers heterogeneity in the recycling endosomes. *J Biol Chem* 1998;273(31):19625–33.

119. Ober RJ, Martinez C, Lai X, Zhou J, Ward ES. Exocytosis of IgG as mediated by the receptor, FcRn: an analysis at the single-molecule level. *Proc Natl Acad Sci USA* 2004;101(30):11076–81.
120. Ober RJ, Martinez C, Vaccaro C, Zhou J, Ward ES. Visualizing the site and dynamics of IgG salvage by the MHC class I-related receptor, FcRn. *J Immunol* 2004;172(4):2021–9.
121. Vaccaro C, Zhou J, Ober RJ, Ward ES. Engineering the Fc region of immunoglobulin G to modulate in vivo antibody levels. *Nat Biotechnol* 2005;23(10):1283–8.
122. Brambell FW, Hemmings WA, Oakley CL, Porter RR. The relative transmission of the fractions of papain hydrolyzed homologous gamma-globulin from the uterine cavity to the foetal circulation in the rabbit. *Proc R Soc Lond B Biol Sci* 1960;151:478–82.
123. Brambell FW. The transmission of immune globulins from the mother to the foetal and newborn young. *Proc Nutr Soc* 1969;28(1):35–41.
124. Telleman P, Junghans RP. The role of the Brambell receptor (FcRB) in liver: protection of endocytosed immunoglobulin G (IgG) from catabolism in hepatocytes rather than transport of IgG to bile. *Immunology* 2000;100(2):245–51.
125. Antohe F, Radulescu L, Gafencu A, Ghetie V, Simionescu M. Expression of functionally active FcRn and the differentiated bidirectional transport of IgG in human placental endothelial cells. *Hum Immunol* 2001;62(2):93–105.
126. Stevenson GT, Anderson VA, Leong WS. Engineered antibody for treating lymphoma. *Recent Results Cancer Res* 2002;159:104–12.
127. Tannock IF, Rotin D. Acid pH in tumors and its potential for therapeutic exploitation. *Cancer Res* 1989;49(16):4373–84.
128. Kristoffersen EK, Matre R. Co-localization of beta 2-microglobulin and IgG in human placental syncytiotrophoblasts. *Eur J Immunol* 1996;26(2):505–7.
129. Junghans RP, Carrasquillo JA, Waldmann TA. Impact of antigenemia on the bioactivity of infused anti-Tac antibody: implications for dose selection in antibody immunotherapies. *Proc Natl Acad Sci USA* 1998;95(4):1752–7.
130. Hobbs SM, Jackson LE, Hoadley J. Interaction of aglycosyl immunoglobulins with the IgG Fc transport receptor from neonatal rat gut: comparison of deglycosylation by tunicamycin treatment and genetic engineering. *Mol Immunol* 1992; 29(7–8):949–56.
131. Perez JH, Wight DG, Wyatt JI, Van SM, Mullock BM, Luzio JP. The polymeric immunoglobulin A receptor is present on hepatocytes in human liver. *Immunology* 1989;68(4):474–8.
132. Tourville DR, Adler RH, Bienenstock J, Tomasi TB, Jr. The human secretory immunoglobulin system: immunohistoligical localization of gamma A, secretory "piece," and lactoferrin in normal human tissues. *J Exp Med* 1969;129(2):411–29.
133. Tourville DR, Tomasi TB, Jr. Selective transport of gamma A. *Proc Soc Exp Biol Med* 1969;132(2):473–7.
134. Takahashi I, Nakane PK, Brown WR. Ultrastructural events in the translocation of polymeric IgA by rat hepatocytes. *J Immunol* 1982;128(3):1181–7.
135. Schiff JM, Fisher MM, Jones AL, Underdown BJ. Human IgA as a heterovalent ligand: switching from the asialoglycoprotein receptor to secretory component during transport across the rat hepatocyte. *J Cell Biol* 1986;102(3):920–31.

136. Kaetzel CS, Robinson JK, Lamm ME. Epithelial transcytosis of monomeric IgA and IgG cross-linked through antigen to polymeric IgA. A role for monomeric antibodies in the mucosal immune system. *J Immunol* 1994;152(1):72–6.

137. Hunziker W, Mellman I. Expression of macrophage-lymphocyte Fc receptors in Madin-Darby canine kidney cells: polarity and transcytosis differ for isoforms with or without coated pit localization domains. *J Cell Biol* 1989;109(6 Pt 2): 3291–302.

138. Hunziker W, Male P, Mellman I. Differential microtubule requirements for transcytosis in MDCK cells. *EMBO J* 1990;9(11):3515–25.

139. Gurevich P, Zusman I, Moldavsky M, et al. Secretory immune system in human intrauterine development: immunopathomorphological analysis of the role of secretory component (pIgR/SC) in immunoglobulin transport (review). *Int J Mol Med* 2003;12(3):289–97.

140. Leung JC, Tsang AW, Chan DT, Lai KN. Absence of CD89, polymeric immunoglobulin receptor, and asialoglycoprotein receptor on human mesangial cells. *J Am Soc Nephrol* 2000;11(2):241–9.

141. Ball JM, Moldoveanu Z, Melsen LR, et al. A polarized human endometrial cell line that binds and transports polymeric IgA. *In Vitro Cell Dev Biol Anim* 1995; 31(3):196–206.

142. Underdown BJ, Schiff JM. Immunoglobulin A: strategic defense initiative at the mucosal surface. *Annu Rev Immunol* 1986;4:389–417.

143. Manning RJ, Walker PG, Carter L, Barrington PJ, Jackson GD. Studies on the origins of biliary immunoglobulins in rats. *Gastroenterology* 1984;87(1):173–9.

144. Mullock BM, Shaw LJ, Fitzharris B, et al. Sources of proteins in human bile. *Gut* 1985;26(5):500–9.

145. Blumberg RS, Koss T, Story CM, et al. A major histocompatibility complex class I-related Fc receptor for IgG on rat hepatocytes. *J Clin Invest* 1995;95(5): 2397–402.

146. Sheff DR, Daro EA, Hull M, Mellman I. The receptor recycling pathway contains two distinct populations of early endosomes with different sorting functions. *J Cell Biol* 1999;145(1):123–39.

147. Zhang JR, Mostov KE, Lamm ME, et al. The polymeric immunoglobulin receptor translocates pneumococci across human nasopharyngeal epithelial cells. *Cell* 2000; 102(6):827–37.

148. Elm C, Braathen R, Bergmann S, et al. Ectodomains 3 and 4 of human polymeric Immunoglobulin receptor (hpIgR) mediate invasion of *Streptococcus pneumoniae* into the epithelium. *J Biol Chem* 2004;279(8):6296–304.

149. Sixbey JW, Yao QY. Immunoglobulin A-induced shift of Epstein-Barr virus tissue tropism. *Science* 1992;255(5051):1578–80.

150. Gan YJ, Chodosh J, Morgan A, Sixbey JW. Epithelial cell polarization is a determinant in the infectious outcome of immunoglobulin A-mediated entry by Epstein-Barr virus. *J Virol* 1997;71(1):519–26.

151. Aarli A, Matre R, Thunold S. IgG Fc receptors on epithelial cells of distal tubuli and on endothelial cells in human kidney. *Int Arch Allergy Appl Immunol* 1991; 95(1):64–9.

152. Kuhn K, Stolte H, Reale E. The fine structure of the kidney of the hagfish (*Myxine glutinosa L.*): A thin section and freeze-fracture study. *Cell Tissue Res* 1975; 164(2):201–13.

153. Latta H. An approach to the structure and function of the glomerular mesangium. *J Am Soc Nephrol* 1992;2(10 Suppl):S65–73.

154. Tencer J, Frick IM, Oquist BW, Alm P, Rippe B. Size-selectivity of the glomerular barrier to high molecular weight proteins: upper size limitations of shunt pathways. *Kidney Int* 1998;53(3):709–15.

155. Radeke HH, Gessner JE, Uciechowski P, Magert HJ, Schmidt RE, Resch K. Intrinsic human glomerular mesangial cells can express receptors for IgG complexes (hFc gamma RIII-A) and the associated Fc epsilon RI gamma-chain. *J Immunol* 1994;153(3):1281–92.

156. Gomez-Guerrero C, Hernandez-Vargas P, Lopez-Franco O, Ortiz-Munoz G, Egido J. Mesangial cells and glomerular inflammation: from the pathogenesis to novel therapeutic approaches. *Curr Drug Targets Inflamm Allergy* 2005;4(3):341–51.

157. Mauer SM, Fish AJ, Blau EB, Michael AF. The glomerular mesangium. I. Kinetic studies of macromolecular uptake in normal and nephrotic rats. *J Clin Invest* 1972; 51(5):1092–101.

158. Thompson EM, Evans DJ. Association of mesangial IgM with IgM deposits in the macula densa: an indication of non-specific macromolecule transport rather than immune reactant? *Nephrol Dial Transplant* 2001;16(9):1910–3.

159. Junghans RP, Waldmann TA. Metabolism of Tac (IL2Ralpha): physiology of cell surface shedding and renal catabolism, and suppression of catabolism by antibody binding. *J Exp Med* 1996;183(4):1587–602.

160. Clyne PS, Kulczycki A, Jr. Human breast milk contains bovine IgG. Relationship to infant colic? *Pediatrics* 1991;87(4):439–44.

161. Cianga P, Cianga C, Cozma L, Ward ES, Carasevici E. The MHC class I related Fc receptor, FcRn, is expressed in the epithelial cells of the human mammary gland. *Hum Immunol* 2003;64(12):1152–9.

162. Takaya R, Fukaya T, Sasano H, Suzuki T, Tamura M, Yajima A. Macrophages in normal cycling human ovaries; immunohistochemical localization and characterization. *Hum Reprod* 1997;12(7):1508–12.

163. Leach JL, Sedmak DD, Osborne JM, Rahill B, Lairmore MD, Anderson CL. Isolation from human placenta of the IgG transporter, FcRn, and localization to the syncytiotrophoblast: implications for maternal-fetal antibody transport. *J Immunol* 1996;157(8):3317–22.

164. Kristoffersen EK. Human placental Fc gamma-binding proteins in the maternofetal transfer of IgG. *APMIS* Suppl 1996;64:5–36.

165. Einhorn MS, Granoff DM, Nahm MH, Quinn A, Shackelford PG. Concentrations of antibodies in paired maternal and infant sera: relationship to IgG subclass. *J Pediatr* 1987;111(5):783–8.

166. Malek A, Sager R, Schneider H. Transport of proteins across the human placenta. Am *J Reprod Immunol* 1998;40(5):347–51.

167. Malek A. Ex vivo human placenta models: transport of immunoglobulin G and its subclasses. *Vaccine* 2003;21(24):3362–4.

168. Eaton BM, Oakey MP. Image analysis of protein profiles from paired microvillous and basal syncytiotrophoblast plasma membranes from term human placenta and characterization of IgG binding to membrane vesicles. *Placenta* 1997;18(7): 569–76.

169. Kutteh WH, Moldoveanu Z, Mestecky J. Mucosal immunity in the female reproductive tract: correlation of immunoglobulins, cytokines, and reproductive hormones in human cervical mucus around the time of ovulation. *AIDS Res Hum Retroviruses* 1998;14 Suppl 1:S51–5.

170. Moldoveanu Z, Huang WQ, Kulhavy R, Pate MS, Mestecky J. Human male genital tract secretions: both mucosal and systemic immune compartments contribute to the humoral immunity. *J Immunol* 2005;175(6):4127–36.

171. Delacroix DL, Marchandise FX, Francis C, Sibille Y. Alpha-2-macroglobulin, monomeric and polymeric immunoglobulin A, and immunoglobulin M in bronchoalveolar lavage. *Am Rev Respir Dis* 1985;132(4):829–35.

172. Bignon J, Jaurand MC, Pinchon MC, Sapin C, Warnet JM. Immunoelectron microscopic and immunochemical demonstrations of serum proteins in the alveolar lining material of the rat lung. *Am Rev Respir Dis* 1976;113(2):109–20.

173. Kim KJ, Malik AB. Protein transport across the lung epithelial barrier. *Am J Physiol Lung Cell Mol Physiol* 2003;284(2):L247–59.

174. Spiekermann GM, Finn PW, Ward ES, et al. Receptor-mediated immunoglobulin G transport across mucosal barriers in adult life: functional expression of FcRn in the mammalian lung. *J Exp Med* 2002;196(3):303–10.

175. Dumont JA, Bitonti AJ, Clark D, Evans S, Pickford M, Newman SP. Delivery of an erythropoietin-Fc fusion protein by inhalation in humans through an immunoglobulin transport pathway. *J Aerosol Med* 2005;18(3):294–303.

176. Cauza K, Hinterhuber G, Dingelmaier-Hovorka R, et al. Expression of FcRn, the MHC class I-related receptor for IgG, in human keratinocytes. *J Invest Dermatol* 2005;124(1):132–9.

177. Brown WR, Kloppel TM. The role of the liver in translocation of IgA into the gastrointestinal tract. *Immunol Invest* 1989;18(1–4):269–85.

178. Brown WR, Kloppel TM. The liver and IgA: immunological, cell biological and clinical implications. *Hepatology* 1989;9(5):763–84.

179. Raju TS, Briggs JB, Chamow SM, Winkler ME, Jones AJ. Glycoengineering of therapeutic glycoproteins: in vitro galactosylation and sialylation of glycoproteins with terminal N-acetylglucosamine and galactose residues. *Biochemistry* 2001; 40(30):8868–76.

180. Hurlimann J. Immunoglobulin synthesis and transport by human salivary glands. Immunological mechanisms of the mucous membranes. *Curr Top Pathol* 1971; 55:69–108.

181. Proctor GB, Carpenter GH. Chewing stimulates secretion of human salivary secretory immunoglobulin A. *J Dent Res* 2001;80(3):909–13.

182. Strober W, Blaese RM, Waldmann TA. The origin of salivary IgA. *J Lab Clin Med* 1970;75(5):856–62.

183. Malamud D. Oral diagnostic testing for detecting human immunodeficiency virus-1 antibodies: a technology whose time has come. *Am J Med* 1997;102(4A):9–14.

184. Takahashi K, Mooney J, Frandsen EV, Kinane DF. IgG and IgA subclass mRNA-bearing plasma cells in periodontitis gingival tissue and immunoglobulin levels in the gingival crevicular fluid. *Clin Exp Immunol* 1997;107(1):158–65.
185. Hochman N, Zakay-Rones Z, Shohat H, et al. Antibodies to cytomegalo and Epstein-Barr viruses in human saliva and gingival fluid. *New Microbiol* 1998; 21(2):131–9.
186. Brandtzaeg P. Human secretory immunoglobulins. II. Salivary secretions from individuals with selectively excessive or defective synthesis of serum immunoglobulins. *Clin Exp Immunol* 1971;8(1):69–85.
187. Newcomb RW, Ishizaka K, DeVald BL. Human IgG and IgA diphtheria antitoxins in serum, nasal fluids and saliva. *J Immunol* 1969;103(2):215–24.
188. Prigent-Delecourt L, Coffin B, Colombel JF, Dehennin JP, Vaerman JP, Rambaud JC. Secretion of immunoglobulins and plasma proteins from the colonic mucosa: an in vivo study in man. *Clin Exp Immunol* 1995;99(2):221–5.
189. Zimmer KP, Scheumann GF, Bramswig J, Bocker W, Harms E, Schmid KW. Ultrastructural localization of IgG and TPO in autoimmune thyrocytes referring to the transcytosis of IgG and the antigen presentation of TPO. *Histochem Cell Biol* 1997;107(2):115–20.
190. Zimmer KP, Schmid KW, Bocker W, et al. Transcytosis of IgG from the basolateral to the apical membrane of human thyrocytes in autoimmune thyroid disease. *Curr Top Pathol* 1997;91:117–28.
191. Estienne V, Duthoit C, Reichert M, et al. Androgen-dependent expression of FcgammaRIIB2 by thyrocytes from patients with autoimmune Graves' disease: a possible molecular clue for sex dependence of autoimmune disease. *FASEB J* 2002;16(9):1087–92.
192. Barrett JS, Wagner JG, Fisher SJ, Wahl RL. Effect of intraperitoneal injection volume and antibody protein dose on the pharmacokinetics of intraperitoneally administered IgG2a kappa murine monoclonal antibody in the rat. *Cancer Res* 1991;51(13):3434–44.
193. Feriani M, Biasioli S, Chiaramonte S, et al. Anatomical bases of peritoneal permeability: a reappraisal. Anatomy of peritoneum. *Int J Artif Organs* 1982;5(6): 345–8.
194. Flessner MF, Dedrick RL, Reynolds JC. Bidirectional peritoneal transport of immunoglobulin in rats: tissue concentration profiles. *Am J Physiol* 1992;263 (1 Pt 2):F15–23.
195. Flessner MF, Dedrick RL, Reynolds JC. Bidirectional peritoneal transport of immunoglobulin in rats: compartmental kinetics. *Am J Physiol* 1992;262 (2 Pt 2):F275–87.
196. Flessner MF, Dedrick RL. Monoclonal antibody delivery to intraperitoneal tumors in rats: effects of route of administration and intraperitoneal solution osmolality. *Cancer Res* 1994;54(16):4376–84.
197. Flessner MF. The role of extracellular matrix in transperitoneal transport of water and solutes. *Perit Dial Int* 2001;21 Suppl 3:S24–9.
198. Rosengren BI, Carlsson O, Venturoli D, al RO, Rippe B. Transvascular passage of macromolecules into the peritoneal cavity of normo- and hypothermic rats in vivo: active or passive transport? *J Vasc Res* 2004;41(2):123–30.

199. Krediet RT, Koomen GC, Vlug A, et al. IgG subclasses in CAPD patients. *Perit Dial Int* 1996;16(3):288–94.
200. Yokoyama I, Waxman F. Isotypic and clonal variations in the interactions between model monoclonal immune complexes and the human erythrocyte CR1 receptor. *Mol Immunol* 1992;29(7–8):935–47.
201. Kavai M, Rasmussen JM, Baatrup G, Zsindely A, Svehag SE. Inefficient binding of IgM immune complexes to erythrocyte C3b-C4b receptors (CR1) and weak incorporation of C3b-iC3b into the complexes. *Scand J Immunol* 1988;28(1): 123–8.
202. Marzocchi-Machado CM, Polizello AC, Azzolini AE, Lucisano-Valim YM. The influence of antibody functional affinity on the effector functions involved in the clearance of circulating immune complexes anti-BSA IgG/BSA. *Immunol Invest* 1999;28(2–3):89–101.

# The Role of Pharmacokinetics and Pharmacodynamics in Selecting a Relevant Species

MARTIN D. GREEN, PhD, and MELANIE HARTSOUGH, PhD

**Contents**

1 Introduction
2 Pharmacodynamics Properties
  2.1 Species Specificity
  2.2 Binding Properties
3 Pharmacokinetics
  3.1 Disposition
  3.2 Immunogenicity
4 Conclusion
  Refer ences

## 1 INTRODUCTION

A relevant animal species is one in which the biopharmaceutical product, as a function of orthologous receptors or antigens, causes a similar biological response and mediates similar effector pathways to that occurring in humans [1]. The appropriate selection and use of a relevant animal species in preclinical toxicity testing is dependent on pharmacokinetic (PK) and pharmacodynamic (PD) measures, and these measures are fundamental for the extrapolation of the preclinical toxicity data to the clinical situation. PD properties, starting with receptors, form the scientific basis for the selection of an animal species for preclinical safety assessment of a biopharmaceutical. Translating the PD findings from laboratory animals to humans is often accomplished through an

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

understanding and extrapolation of PK relationships between these models and the clinical setting. In particular, PD responsiveness typically yields evidence of the intensity and dynamic range of a response, whereas PK parameters serve as the means to convert the PD relationships gained from the preclinical models, such as those relating to plasma levels, dose, and dosing regimens, into the practical issues involved in a clinical study (e.g., selection of an initial dose, dose escalation scheme, and patient monitoring scheme).

Although the selection of a relevant species is clearly recognized as a defining principle for the safety assessment of biopharmaceuticals, this concept is not generally instrumental in the safety assessment of drugs that are typically of significantly lower molecular weight.[1] However, a limited application of this concept may be applied to drugs of smaller molecular weight. In this approach animal models for use in preclinical toxicity evaluations [2] and carcinogenicity studies [3,4] are selected for toxicity assessment that possess clinically meaningful metabolites due to similarities in the pathways for cytochrome 450 enzymes ($P_{450}$). In general, evaluating the data for a drug given to animals that produces metabolites that are common to humans allows for a more accurate interpretation of the toxicity profile and a more representative risk–benefit assessment. As biopharmaceuticals do not undergo metabolism by $P_{450}$, this concept is not applicable to large molecular weight products. Nevertheless, a common perspective between biopharmaceuticals and drugs is the need to establish a predictive preclinical model for the evaluation of human safety and identification of clinically relevant outcomes.

## 2 PHARMACODYNAMIC PROPERTIES

For most biopharmaceuticals, exaggerated pharmacological responses are a major contributor to the profile of adverse clinical events. Thus, the use of a relevant animal species to determine the toxicity profile of a biopharmaceutical is essential to the process of assessing overall safety. In some cases, the frequency of potential adverse events is relatively common and readily predictable from the pharmacodynamics of a product. For example, tissue plasminogen activator (tPA) has a restricted pharmacological activity, which is the ability to lyse clots and induce fibrinolysis. These effects apply across a number of species, albeit with different relative potency, and they subsequently lead to an expected increased incidence of bleeding. In other instances, the exact nature of predicting a specific adverse event from the PD properties of a product is significantly less reliable. Such is the case for biopharmaceuticals that inhibit TNF function. Although a general immune suppression may be anticipated with these products, the immune suppression unexpectedly leads

[1]Drugs are defined as products made by synthetic chemical processes and have lower molecular weights than biopharmaceutical products, which are made by cells and have molecular weights often greater than 1000 daltons.

to an increased incidence of tuberculosis in humans. Thus, the utility of the toxicology findings in animals depends on the selection of an appropriate animal model, with an understanding of its limitations, the frequency of the adverse occurrence, and the degree of impact on safety, and all these are dependent on the ability to extrapolate preclinical findings in a comprehensive manner. Although many important clinically adverse events may often be linked to information related to exaggerated pharmacological responses and are logical extrapolations from the preclinical safety data, not all important adverse events lend themselves to be assessed in preclinical models. For example, traztuzumab, which is used to treat breast cancer, causes a cardiomyopathy that resembles the effects of anthracyclines. This adverse event is not linked to its pharmacological actions mediated by its binding to the HER2 receptor, its only known site of action.

Without a proper understanding of the PK and PD responsiveness of an animal model elected for preclinical safety evaluation, it is difficult to distinguish between true positive or negative findings and outcomes. In the early 1990s with FIAU (fialuridine) and more recently with TGN 1412, an agonist monoclonal antibody (mAb) acting on CD28, preclinical studies yielded a false sense of anticipated clinical safety, due to a lack of corresponding PD responsiveness between animal models used for preclinical safety assessment and subjects of the clinical investigations. For FIAU, an antiviral nucleoside, commonly used animal models such as the mice or dogs failed to demonstrate the liver toxicity that was later observed in patients with viral hepatitis [5]. Rats did reveal some evidence of potential liver toxicity but at doses approximately 2000-fold greater than those demonstrated to be toxic to patients. Although unusual as an animal model, the woodchuck was later identified as a relevant model of human hepatitis B infections that was similarly responsive in terms of liver toxicity and sensitivity to doses of FIAU [6]. Similarly, TGN 1412 was investigated in nonhuman primates that were relatively unresponsive in terms of symptoms exhibited and thus yielded a false sense of understanding in terms of potential toxicities to humans [7]. The simple realization that the animal models used were not pharmacodynamically responsive should have led to a more conservative approach to the design of the clinical study with respect to the number of patients put at risk in a given cohort, clinical monitoring, dose escalation scheme and duration of treatment.

### 2.1 Species Specificity

Sometimes PD properties are common across species, thus allowing for pertinent toxicity data to be collected from a variety of species. An example of this is granulocyte colony stimulating factor (G-CSF), a cytokine produced by various cells including monocytes, fibroblasts, and endothelial cells that regulates the production of neutrophils within the bone marrow, influences neutrophil proliferation and differentiation, and contributes to increased phagocytic function, respiratory burst, and antibody-dependent killing. Keller

et al. [8] reported that recombinant, human G-CSF administered to mice, rats, hamsters, dogs, and nonhuman primates produced the expected neutrophil effects at doses ≥1 μg/kg/day in all of the species. In other instances the PD properties restrict the applicability of animal species commonly used for safety testing. For example, in contrast to G-CSF, recombinant human granulocyte-macrophage colony stimulating factor (rhGM-CSF) exerts strong effects in nonhuman primates (including cynomolgus and rhesus monkeys) but is pharmacologically inactive in rodents. In the rhesus monkey, rhGM-CSF increased the proliferation and differentiation of myeloid precursors, resulting in a 20-fold increase in mature granulocytes counts in blood and in an activation of granulocytes and monocytes. At high doses it caused inflammation of the serosal surfaces of the heart and liver and at the site of injection [9]. Moreover secondary to a significant increase in granulocyte count, cardiovascular adverse events such as thrombosis also occurred. These effects are similar to that observed in humans. Given the PD agreement in responsiveness to humans, it was found that the rhesus monkey, rather than rodents, was predictive of clinical outcomes.

Yet a potentially more complicated problem is one of properly interpreting and extrapolating the effects of some biopharmaceuticals across closely related animal species that are pharmacologically responsive, in order to determine the species most clinically predictive. This problem is exemplified by interleukin-2 (IL-2), a cytokine used clinically to treat metastatic, renal cell carcinoma. IL-2 is produced by activated T cells and mediates a variety of effects, such as stimulation of lymphocyte activation and proliferation, enhancement of lymphocyte-directed cytotoxicity, and induction of interferon-γ secretion. It is pharmacologically active in rodents, rabbits, sheep, and humans [10]. Similar PD effects to those in humans are observed in mice, but in rats a related, but different, set of consequences occurs, even though the same underlying causative mechanisms exist. Moreover, the relative potency of human, recombinant IL-2 is approximately 20-fold less in mice and 70-fold less in rats when compared to humans. These differences are also reflected in the relevancy of the toxicity profiles of the two rodent species to that observed in humans. For example, the most prominent clinical toxicity of IL-2 results from an increased vascular permeability, producing extravasation of fluid and subsequent pulmonary edema and pleural effusions. Ultimately, a condition collectively referred to as vascular or capillary leak syndrome, characterized by the loss of fluid and proteins out of the circulating blood volume and into surrounding tissues, is produced. This set of conditions leads to extremely low blood pressure, multiple organ failure, and shock. In the setting of oncology, the clinical use of IL-2 is characterized by weight gain due to retained fluid, pulmonary edema, pleural effusion, and ascites. Other clinically significant side effects include fever, nausea, eosinophilia, and thrombocytopenia. In rats given an intraperitoneal injection of IL-2, pleural effusion was inconsistently observed and minor in intensity. In addition, while lung weights were increased due to cellular infiltration, primarily eosinophilic in nature, frank evidence of pulmonary edema was

not observed. Intravenous administration of IL-2 to rats also failed to yield similar effects as that in humans, perhaps due to a decreased systemic persistence. In contrast, mice given IL-2 presented with similar toxicities as humans, severe pulmonary edema and pleural effusion and lymphocytic cell infiltration [11,12,13].

Some biopharmaceuticals exhibit a high degree of species selectivity that may restrict their biological activity to nonhuman primates (e.g., cynomolgus and rhesus monkeys) for preclinical safety. Hart et al. [14] reported the use of an in vitro IL-4-dependent, T cell proliferation assay to identify relevant species for testing the toxicity of an anti-IL-4 mAb (a humanized version of the murine parent mAb 3B9). The findings of the assay revealed that the antihuman IL-4 antibody inhibited the T cell response to recombinant cynomolgus monkey IL-4, an orthologous molecule with an amino acid sequence identity of 93% to the clinical version. No reactivity was found with mouse, rat, cow, goat, or horse IL-4.

Among the most highly selective cytokine biopharmaceuticals are the interferons. Interferons are members of a large family of related proteins that may be divided into two categories: type I and type II interferons. Type I interferons (i.e., α- and β-interferon) possess antiviral and anti-proliferative properties, whereas type II interferons (i.e., γ-interferon) have immunostimulatory activity. Several animal species were examined for their responsiveness to interferons, and with the exception of nonhuman primates, all tested animal species were found to be unresponsive [15].

Increasingly, the selectivity and specificity of biopharmaceuticals for human receptors often limit the range of responsive animal models (as described above for interferons and IL-4 mAb) or require the creation and use of different strategies. Alternate approaches to the problem of obtaining useful information include the use of homologous (also referred to as analogous or surrogate) proteins for the biopharmaceutical, the creation of transformed animal models such as transgenic animals, and the use of models of disease. Preclinical studies based on these alternative approaches should display a parallel and hopefully correct relative PD relationship in order to yield predictive information.

In rare instances the number of pharmacologically responsive preclinical species is very narrow, and only a single species such as the chimpanzee may be available to appropriately assess preclinical PD and PK endpoints. Since toxicology testing cannot be performed in chimpanzees due to their protective status, alternative methods are usually considered for standard safety assessments. An example of this is described by Clarke et al. [16] for efalizumab, a humanized mAb directed against the α-chain component of lymphocyte function associated antigen1 (LFA-1), also known as CD11a, that binds only to chimpanzee and human forms of the molecule. Efalizumab is used clinically for the treatment of psoriasis, and the PD basis for treating psoriasis is the antagonization of T lymphocyte infiltration into psoriatic lesions by blocking the T cell surface molecule LFA-1, thereby preventing interaction with the

endothelial cell surface protein ICAM-1. To provide an adequate safety assessment, muM17, a chimeric rat anti-mouse CD11a mAb, was created and evaluated. This analogous molecule bound with similar specificity and affinity as efalizumab to the human CD11a and was found to have similar pharmacological activities in mice as efalizumab in humans. In order to bridge the preclinical safety information obtained from muM17 to the human setting, a PK/PD model was developed in mice [17]. This model was useful in selecting a clinically relevant dose based on the PD rather than the PK profile because of differing capacities of the CD11a binding between mice and humans and its influence on pharmacokinetics.

In some cases animal models of disease that mimic the anticipated clinical situation are needed to provide a set of relevant safety data, since the pathophysiological background is essential in providing the appropriate receptor or effecter mechanism. For example, erythropoietin, a glycoprotein with a molecular weight of 30 kDa, stimulates division and differentiation of committed erythroid progenitor cells and is not species restricted in terms of its primary biological activity. It induces the release of reticulocytes from the bone marrow into the bloodstream where they mature into erythrocytes and increase hematocrit. Too rapid a response to this growth factor will increase the viscosity of blood, thereby increasing vascular resistance and likelihood of a stroke. Erythropoietin did not demonstrate any cardiovascular toxicities in the general toxicity studies or in the extensive cardiovascular safety pharmacology studies performed in healthy animals [18], perhaps due to inherent compensating mechanisms. In contrast to a lack of findings in normal animals, a difference in cardiovascular regulatory response was observed in either spontaneously hypertensive rats or in isolated tissues obtained from these animals [19,20]. When erythropoietin was used in patients with chronic renal failure, hypertension was observed as a major clinical toxicity.

## 2.2 Binding Properties

Binding properties of a biopharmaceutical are important determinants in establishing the relevance of a species for safety assessment. An empirical assessment of binding is often necessary to establish the extent and location of binding across species. For example, in vitro studies of adalimumab, a fully human recombinant mAb directed against human TNFα, have revealed it to be a potent inhibitor of TNFα-related actions for humans, cynomolgus monkeys, and dogs, whereas no binding to the rat TNFα and only weak inhibition of the murine TNFα were observed [21]. Another example is humanized mAb alemtuzumab. This antibody recognizes CD52 and cross-reacts with epitopes from cynomolgus monkeys, rhesus monkeys, and baboons. However, in rhesus monkey and baboons as well as some individual cynomolgus monkeys, CD52 expression is found on erythrocytes, in contrast to humans in which it is absent [22]. Moreover, the binding affinity of alemtuzumab to cynomolgus CD52 is approximately 16-fold less than that to the human epitope. Hence,

the binding to erythrocytes in these preclinical models could yield a false positive signal of hemolytic anemia, relative to humans.

The role of the various subcomponents of a receptor or epitope may contribute to the overall complexity of biopharmaceutical binding, pharmacokinetics, and pharmacological responsiveness and, in fact, is often an unknown factor in PK and PD differences between species. For example, the IL-2 receptor is a trimeric receptor composed of α, β, and γ chains, and the assembly of different groupings of these subcomponents determines the pharmacological relationship to pharmacokinetics. The trimeric complex binds IL-2 with very high affinity (Kd = $10^{-11}$ M), when compared to IL-2's affinity for the IL-2Rβγ complex (Kd = $10^{-9}$ M) or IL-2Rα subunit (Kd = $10^{-8}$ M). Moreover, the IL-2Rα chain is not involved with intracellular signaling, whereas the IL-2Rβ and IL-2Rγ subunits are critically important for pharmacological activity. Thus the therapeutic activity of IL-2 is actually more complex than just cytokine receptor binding. It also requires the recruitment of antigen-activated T cells and CD56 bright NK cells, an event dependent on the expression of all three receptors subunits as a trimeric complex with high affinity. In contrast, the IL-2-induced toxicity is associated with the lower affinity complexes, in particular, the binding of IL-2 to the IL-2Rβγ complexes located on CD56 dim NK cells. Hence the relative affinity dictates the range of IL-2-dependent effects, as picomolar concentrations are associated with the therapeutic aspects of the pharmacological activity of IL-2 and nanomolar levels with toxicity [23,24].

In rare cases intra-species variation may be important when assessing overall clinical predictability. Similar to the differences in $P_{450s}$ that may be observed within humans and animals, the binding properties may vary within a species (even within a relevant species), which may confound the overall interpretation of the toxicology study. A good example of this is provided by Klingbeil and Hsu [25]. The authors describe the safety testing of Hu1D10, a humanized mAb directed against the posttranslational form of HLA-DR expressed on normal B cells as well as B cell leukemias and lymphomas. Klingbeil and Hsu reported that Hu1D10 administration to some monkeys resulted in severe acute adverse effects (e.g., respiratory suppression, increased heart rate, and urticaria) that in some cases required life-sustaining intervention, while no such adverse effects were observed in other Hu1D10-treated monkeys or the control monkeys. The extreme differences in effect were associated with the variation of HLA-DR expression, in that those animals that were antigen-positive exhibited toxicities while those that were antigen-negative did not.

## 3 PHARMACOKINETICS

Pharmacokinetic studies are often used to establish the extent of exposure to a biopharmaceutical in a preclinical study. This information not only validates dosing but also provides a means of extrapolating exposure across species. It

therefore can be used as the basis for selecting an initial, safe starting dose for clinical use, determining therapeutic targets based on plasma levels associated with pharmacological effects, and providing an upper limit to clinical dosing based on PK exposure.

In general, extrapolation of PK data across species is dependent on the molecular weight of the biopharmaceutical, the involvement of receptor-mediated uptake, and the disposition and the route of administration [26]. A clinical dose may be extrapolated from preclinical data on either a body weight or surface area basis [27]. The appropriate method of extrapolations depends on the underlying factors governing disposition. Surface area based extrapolations assume a proportionate first-order dependent process, which may have active components such as receptor-mediated uptake mechanisms or more passive ones such as glomerular filtration. Generally, larger molecular weight biopharmaceuticals of greater than 70 kDa will scale by body weight, and lower molecular weight products will more often be applicable to scaling based on body surface area conversion. It is not uncommon for the saturation of various active processes to result in a nonlinearity in PK between plasma levels and dose, thus complicating extrapolation and interpretation when a wide range of doses are studied. This range of doses often includes and extends beyond a range with clinical utility. Thus, it is important to consider doses that are clinically relevant to ensure that clearance across species is highly predictive, whether it is a function of body weight or surface area.

### 3.1 Disposition

The PK of biopharmaceuticals are governed by factors such as absorption, distribution, catabolism, and excretion (similar principles to drugs) as well as other factors such as binding proteins and antibodies directed against the biopharmaceutical. Unlike drugs that readily diffuse, biopharmaceuticals, due to their molecular weight and shape, do not diffuse but are initially confined to the circulating vasculature; however, with time they distribute to the extravascular space by various factors, including bulk flow and convection. Catabolism of biopharmaceuticals is achieved by the same proteolytic processes that break down endogenous proteins, occurs in many sites in vivo, and usually involves receptor-mediated endocytosis followed by proteolysis in lysosomes. In addition to the intracellular mechanisms, catabolism may also occur at extravascular sites, such as the mediation of degradation by local proteolytic activity after subcutaneous injection, a process that influences overall bioavailability [28,29]. The route of absorption from these extravascular sites is influenced by molecular size and is partitioned between direct absorption into the blood stream and lymphatic system. Similarly, the mechanisms of elimination of a biopharmaceutical from blood are also dictated by molecular size. For instance, biopharmaceuticals of greater than 70 kDa do not undergo renal glomerular filtration and degradation; instead, the liver plays a predominant role in their catabolism and PK behavior. In contrast, the smaller the biophar-

maceutical the more likely the kidney, not the liver, will play a major role in elimination. Those biopharmaceuticals that undergo renal filtration are often readsorbed and/or degraded by various renal structures, including the brush border and tubules. While these disposition mechanisms are similar across species, it is important to understand that the certain aspects will vary across species due to physical differences. For example, upon administration of a biopharmaceutical into a tissue site, the ratio of the volume injected to the total volume of the organ may vary across species and result in different proportions of anatomical structures and cellular populations being exposed. Thus, it is important to understand what mechanisms are expected to influence absorption, distribution, catabolism, and excretion in humans to help elucidate the relevancy of the species with regard to extrapolation of PK endpoints.

Besides the influence of physical anatomical features such as the glomerular filtration, the PK behavior of biopharmaceuticals may be significantly affected by receptor-mediated processes and other uptake processes that outweigh more passive aspects of disposition. An example is the glycosylation patterns of a biopharmaceutical. The terminal glycoforms such as sialic acid and mannose residues play a significant role in regulating catabolism and influence several PK parameters, including systemic exposure, as measured by AUC and half-life. Glycosylation profiles may vary across species, which may result in different product clearance rates due to the disparate rates of uptake by, for example, the asialoglycoprotein or mannose/GlcNAc receptor-mediated pathways [30]. It is therefore important to consider these PK-related mechanisms when selecting a relevant species for preclinical testing.

In contrast to PD, in which the validity of the study for the assessment of safety often depends on specific receptors or antigens, the validity of PK studies are not as highly dependent on the presence of specific receptors or other mechanisms that are clinically homologous. For example, the disposition of tPA is strongly dependent on blood flow patterns to various organs, such as the liver, as well as on the presence of various receptors involved in uptake. It has a molecular weight of 65 kDa and is composed of a serine protease domain at the carboxyl terminal end joined to other domains typical of plasma proteases. Also contained in tPA is a fibronectin-type finger region, a growth factor domain and two kringle domains. About 8% of the total molecular mass consists of glycoforms, which play a significant role in the in vivo disposition. Included in the glycoforms is a high mannose type moiety. tPA circulates as both an unbound moiety and a complex consisting of plasminogen inhibitor (PAI-1) and other protease inhibitors such as plasminogen inhibitors 1 and 2, C1-esterase inhibitor, α2-macroglobulin and α- antiplasmin [31–34]. The PK of tPA have been studied in several different species, including mice, rats, rabbits, dogs, and monkeys [35–40]. A generally applicable and consistent linear PK pattern has emerged from these studies across varies species that is a good approximation of that found clinically. After infusion of tPA, plasma levels decrease in a rapid initial pattern that is characterized by a half-life of 1 to 3 minutes (α-phase). The initial phase is dominant with respect to

exposure and includes approximately 70% of the total AUC. A second phase of disposition follows with a half-life of 10 to 40 minutes (β-phase) and finally a third phase with a half-life of 1 to 2 hours (γ-phase). The degree of compartmentalization is dependent on the quantitative ability and reliability of the assay used to measure tPA. The final phase of disposition only accounts for approximately 7% to 10% of the total AUC. Total plasma clearance of tPA closely approximates hepatic plasma flow in all species and ranges from 16 to 23 ml/min/kg; the initial volume of distribution corresponds to plasma volume of 46 to 91 ml/kg. A highly significant correlation was observed between tPA clearance and body weight in animal species and humans, and this provides the basis for predicting human PKs from animal data [41].

In the liver, tPA is taken up in various cellular populations and illustrates the varied nature of the distribution of a biopharmaceutical within an organ. A saturable, receptor-mediated endocytosis of tPA fundamentally occurs followed by lysosomal degradation when tPA is taken up by parenchymal, endothelial, and Kupffer cells in the liver [42]. In the rat, endothelial cells, as compared to parenchymal cells, accounted for a 20-fold higher amount of radioactive activity on a per milligram of protein basis [43,44]. Also Kupffer cells exhibited a 4-fold higher level of uptake, as compared to parenchymal cells. When the cellular mass of cells comprising the liver was considered, endothelial cells accounted for 55%, parenchymal cells 40%, and Kupffer cells 6% of the total uptake of tPA in the liver. The hepatic receptor-mediated uptake by these cells is regulated by two distinct systems. A receptor on endothelial cells that recognizes high-mannose glycoform is responsible for approximately one-half of the total [45]. Additionally a receptor that recognizes protein structure rather than glycoforms seems to be operant and attributes to the remaining disposition of the biopharmaceutical [46–48].

Other pathways exist for the interactions of biopharmaceuticals with receptors on cell surfaces, including receptor binding and shedding of the resulting complex into the circulation and receptor binding that is neither shed nor internalized but stable on the cell surface. Monoclonal antibodies undergo a recycling process via the FcRn receptor [49], as well as nonspecific uptake and the Fcγ receptor's binding. These mechanisms are generally found in various species, but their specificity and relative activity may vary and may need to be considered when selecting an appropriate animal model.

### 3.2 Immunogenicity

A major factor in extrapolating pharmacokinetics across species is the development of antiproduct antibodies, which, if this occurs, may confound interpretation of the PK data. Thus, the clinical relevance of PK and PD information based on preclinical data can only be understood to the extent that the nature, magnitude, and timing within the course of dosing is relative to the development of antiproduct antibodies. Furthermore, characteristics (e.g., precision, accuracy, and specificity) of the assays necessary to quantitate and characterize

the antiproduct antibodies and the time of collecting samples in the preclinical study are important. Due to the potential assay interference from the presence of high blood concentrations of the biopharmaceutical, sufficient time must elapse between the last administration of dose and the collection of plasma samples to obtain accurate data. Conversely, with too much time allowed to pass after the last dose of the biopharmaceutic, the immune response may be undetectable. Nevertheless, a comparison of early and late levels at comparable points of time after dosing of the biopharmaceutic may provide a robust signal of immunogenicity. In preclinical studies, immunogenicity will vary among species and additionally among the strains of animals that may be used in safety studies, as the immune response to the biopharmaceutical may differ as the expression of major histocompatibility complexes varies among strains.

Although immunogenicity may arise in the course of exposure of biopharmaceuticals, preclinical PK studies may be successfully conducted by monitoring the development of antiproduct antibodies when this factor is considered in the analysis of the data. For example, preclinical studies conducted using lenercept provided useful PK information for extrapolation to humans [50]. Lenercept is a fusion protein that combines the extracellular domain of two human p55 tumor necrosis factor receptors with the hinge and constant domain C2 and C3 sequences of the human immunoglobulin G1 heavy chain. The biopharmaceutical has a molecular weight of 120 kDa and possesses eight potential sites of glycosylation. Lenercept was investigated for its potential application to various clinical conditions including rheumatoid arthritis. In the manuscript by Richter et al., the pharmacokinetics of lenercept were studied in RoRo rats, Himalayan rabbits, beagle dogs, and cynomolgus monkeys [50]. Several doses of lenercept was administered intravenous by bolus or short infusion; rats were given 0.2 or 5 mg/kg, rabbits 5 mg/kg, monkeys 4.0 or 5.0 mg/kg, and dogs 0.11 and 0.14 mg/kg. Groups sizes were relatively small and ranged from eight to two subjects per species, and blood samples were analyzed for lenercept levels with an ELISA. Neutralizing antibodies to the biopharmaceutical were also measured during the course of the investigation. Various PK endpoints were collected from the different species studied and examined for their fit in an allometric scaling procedure. Sample collection times for PK assessment extended into the period that the development of antiproduct antibodies were observed. The minimal sampling time was 192 hours for rats and a maximum of 437 hours for dogs after injection. A triphasic PK profile was observed, in monkeys, dogs, and rabbits, but not rats, given a dose of 5 mg/kg intravenously. Following an initial, rapid decrease in levels, a second, slower disposition phase occurred, followed by a third phase of rapidly decreasing levels between days 6 to 10. The third phase reflected an immune response directed toward lenercept that showed intra-subject variability. When the effect of antibodies to lenercept in the third phase are subtracted from the data set, the plasma levels were found to be superimposable and linear to dose across species, thus allowing for the allometric scaling of the PK parameters

to the clinical situation. In fact, predicted human clearance and the volume of distribution at steady state based on the preclinical data were found to be in agreement with the clinical studies performed [51]. More likely the clearance of lenercept is a result of the uptake mediated by glycosylation and Fc binding and is not reflective of glomerular filtration as the molecular size exceeds 70 kDa.

## 4 CONCLUSION

Biopharmaceuticals represent a broad but discrete class of large molecular weight therapeutic entities that are characterized by their specific pharmacological activities and distinctive pharmacokinetics. The selection of an appropriate animal model is dependent on a combination of PD and PK factors. As described in this chapter, it is essential to understand the relationship of the basic pharmacology of a biopharmaceutical (signaling, receptor presence, binding properties, etc.) and the associated PK properties to that expected in humans, in order to select animal species that will have the most predictive value in safety assessments.

## REFERENCES

1. ICH S6: Preclinical Safety Evaluation of Biotechnology-Derived Pharmaceuticals. http://www.fda.gov/cder/guidance/1859fnl.pdf
2. Obach RS, Baxter JG, Liston TE, Silber BM, Jones BC, Macintyre F, Rance DJ, Wastall P. The prediction of human pharmacokinetic parameters from preclinical and in vitro metabolism data. *J Pharmacol Exp Therapeut* 1977;283:46–58.
3. Hengstler JG, Van Der Burg B, Steinberg P, Oesch F. Interspecies differences in cancer susceptibility and toxicity. *Drug Metab Rev* 1999;31:917–70.
4. Morton DM. Importance of species selection in drug toxicity testing. *Toxicol Letts* 1998;102–3, 545–50.
5. McKenzie R, Freed MW, Sallie R, Conjeevaram H, Bisceglie AM, Yoon P, Savarese B, Kleiner D, Tsokos M, Luciano C, Pruett T, Stotka J, Straus SE, Hoofnagle JH. Hepatic failure and lactic acidosis due to Fialuridine (FIAU), an investigational nucleoside analogue for chronic hepatitis B. *N Eng J Med* 1995;333:1099–105.
6. Tennant BC, Baldwin BH, Graham LA, Ascenzi MA, Hornbuckle WE, Rowland PH, Tochkov IA, Yeager AE, Erb HN, Colacino JM, Lopez C, Engelhardt JA, Bowsher RR, Richardson FC, Lewis W, Cote PJ, Korba BE, Gerin JL. Antiviral activity and toxicity of fialuridine in the woodchuck model of hepatitis B virus infection. *Hepatology* 1998;28:179–91.
7. Expert Scientific Group on Phase One Clinical Studies. Final Report. The Stationery Office. Her Majesty's Stationery Office, St. Clements House, 2–16 Colegate, Norwich NR3 1BQ.
8. Keller P, Smalling R. Granulocyte colony stimulating factor: animal studies for risk assessment. *Int Rev Exp Pathol* 1993;34(pt A):173–88.

9. Robison RL, Myers LA. Preclinical safety assessment of recombinant human GM-CSF in rhesus monkeys. *Int Rev Exp Pathol* 1993;34A:149, 172.
10. Anderson TD, Hayes TJ. Toxicity of human recombinant interleukin-2 in rats. *Lab Invest* 1989;60:331–46.
11. Anderson TD, Hayes TJ, Gately MK, Bontempo JM, Stern LL, Truit GA. Toxicity of human recombinant interleukin-2 in the mouse is mediated by interleukin-activated lymphocytes: separation of efficacy and toxicity by selective lymphocyte subset depletion. *Lab Invest* 1988;59:598–612.
12. Gately MK, Anderson TD, Hayes TJ. Roles of asialo GM1-positive lymphoid cells in mediating the toxic effects of recombinant interleukin-2 in mice. *J Immunol* 1988;141:189–200.
13. Rosenstein M, Ettinghausen SE, Rosenberg SA. Extravasation of intravascular fluid mediated by the systemic administration of recombinant interleukin-2. *J Immunol* 1986;137:1735–42.
14. Hart TK, Blackburn MN, Brigham-Burke M, Dede K, Al-Mahdi P, Zia-Amirhoseine P, Cook RM. Preclinical efficacy and safety of pascolizumab (SB 240683): a humanized anti-interleukin-4 antibody with therapeutic potential in asthma. *Clin Exp Immunol* 2002;130:93–100.
15. Terrell TG, Green JD. Comparative pathology of recombinant murine interferon-γ in mice and recombinant human interferon-γ in cynomolgus monkeys. *Exp Pathol* 1993;34B: 73, 101.
16. Clarke J, Leach W, Pippig S, Joshi A, Wu BM, House R, Beyer J. Evaluation of a Surrogate Antibody for Preclinical Safety Testing of an AntiCD11a mAb. *Regul Toxicol Pharmacol* 2004;40:219–26.
17. Wu B, Joshi A, Ren S, Ng C. The application of mechanism-based PK/PD modeling in pharmacodynamic-based dose selection of muM17: a surrogate monoclonal antibody for Efalizumab. *J Pharmaceut Sci* 2006;95:1258–68.
18. Dempster AM. Pharmacological testing of recombinant human erythropoietin: implications for other biotechnology products. *Drug Dev Res* 1995;35:173–8.
19. Muntzel M, Hannedouche T, Lacour B, Drueke TB. Effect of erythropoietin on hematocrit and blood pressure in normotensive and hypertensive rats. *J Am Soc Nephrol* 1992;3:182–7.
20. Tsukade H, Ishimitsu T, Ogawa Y, Sugimoto T, Yagi S. Direct vassopressor effects of erythropoietin in genetically hypertensive rats. *Life Sci* 1992;52:1425–34.
21. Humira. Scientific Discussion. European Public Assessment Report. European Medicines Agency. http://www.emea.europa.eu/humandocs/Humans/EPAR/humira/humira.htm
22. MabCampath. Scientific Discussion. European Public Assessment Report. European Medicines Agency. http://www.emea.europa.eu/humandocs/Humans/EPAR/mabcampath/mabcampath.htm
23. Rao BM, Driver I, Lauffenburger DA, Wittrup KD. Interleukin 2 (IL-2) variants engineered for increased IL-2 receptor alpha-subunit affinity exhibit increased potency arising from a cell surface ligand reservoir effect. *Mol Pharmacol* 2004; 66:864–9.

24. Rao BM, Girvin AT, Ciardelli T, Lauffenburger DA, Wittrup KD. Interleukin-2 mutants with enhanced alpha-receptor subunit binding affinity. *Protein Eng* 2003; 16:1081–7.
25. Kingbeil C, Hsu DH. Pharmacology and safety assessment of humanized monoclonal antibodies for therapeutic use. *Toxicol Pathol* 1999;27:1–3.
26. Mordenti J, Chen S, Moore J, Ferraiolo B. Interspecies scaling of clearance data for five recombinant proteins. *Pharmaceut Res* 1991;8:1351–9.
27. Mordenti J, Green JD. The role of PKs and PDs in the development of therapeutic proteins. In Rescigno A, Thakur A (eds), *New Trends in Pharmacokinetics*. New York: Plenum Press, 1991, pp 411, 424.
28. Supersaxo A, Hein WR, Steffen H. Effect of molecular weight on the lymphatic absorption of water-soluble compounds following subcutaneous administration. *Pharmaceut Res* 1990;7:167–9.
29. Supersaxo A, Hein WR, Gallanti H, Steffen H. Recombinant human interferon alpha-2a: delivery to lymphoid tissue by selected modes of application. *Pharmaceut Res* 1988;5:472–6.
30. Park EI, Yiling M, Unverzagt C, Gabius H-J, Baenziger JU. The asialoglycoprotein receptor clears glycoconjugates terminating with sialic acida2,6GalNAc. *Proc Nal Acad Sci USA* 2005;102:17125–9.
31. Higgins DL, Bennett WF. Tissue plasminogen activator: the biochemistry and pharmacology of variants produced by mutagenesis. *Ann Rev Pharmacol Toxicol* 1990; 30:91–121.
32. Lucore CL, Sobel BE. Interactions of tissue-type plasminogen activator with plasma inhibitors and their pharmacologic implications. *Circulation* 1988; 77:660–9.
33. Sprengers ED, Kluft C. Plasminogen activator inhibitors. *Blood* 1987;69:381–7.
34. Haggroth L, Mattsson C, Friberg J. Inhibition of the human tissue plasminogen activator in plasma from different species. *Thromb Res* 1984;33:583–94.
35. Tanswell P, Heinzel G, Greischel A, Krause J. Nonlinear pharmacokinetics of tissue-type plasminogen activator in three animal species and isolated perfused rat liver. *J Pharmacol Exp Ther* 1990;255:318–24.
36. Mohler MA, Refino CJ, Chen SA, Chen AB, Hotchkiss AJ. D-phe-pro-arg-chloromethylketone: its potential use in inhibiting the formation of in vitro artifacts in blood collected during tissue-type plasminogen activator thrombolytic therapy. *Thromb Hemostasis* 1986;56:160–4.
37. Krause J, Seydel W, Heinzel G, Tanswell P. Different receptors mediate the hepatic catabolism of tissue-type plasminogen activator and urokinase. *Biochem J* 1990; 267:647–52.
38. Fuchs HE, Berger HB, Pizzo SV. Catabolism of human tissue plasminogen activator in mice. *Blood* 1985;65:539–44.
39. Fong KL, Crysler CS, Mico BA, Boyle KE, Kopia GA, Kopaciewicz L, Lynn RE. Dose-dependent pharmacokinetics of recombinant tissue-type plasminogen activator in anesthetized dogs following intravenous infusion. *Drug Metab Dis* 1988; 16:201–6.
40. Korninger C, Stassen JM, Collen D. Turnover of human extrinsic (tissue-type) plasminogen activator in rabbits. *Thromb Haemostasis* 1981;46:658–61.

41. Mordenti J. Man versus beast: pharmacokinetic scaling in mammals. *J Pharmaceut Sci* 1986;75:1028–40.
42. Rijken DC, Otter M, Kuiper J, van Berkel TJ. Receptor-mediated endocytosis of tissue-type plasminogen activator (t-pa) by liver cells. *Thromb Res Suppl* 1990;10:63–71.
43. Smedsrod B, Einarsson M, Peroft H. Tissue plasminogen activator is endocytosed by mannose and galactose receptors of rat liver cells. *Thromb Haemostasis* 1988; 59:480–4.
44. Kuiper J, Otter M, Rijken DC, van Berkel TJ. Characterization of the interaction in vivo of tissue-type plasminogen activator with liver cells. *J Biol Chem* 1988; 263:18220–4.
45. Smedsrod B, Malmgren M, Ericsson J, Laurent TC. Morphological studies on endocytosis of chondroitin sulphate proteoglycan by rat liver endothelial cells. *Cell Tiss Res* 1988;253:39–45.
46. Hotchkiss A, Refino CJ, Leonard CK, O'Connor JV, Crowley C, McCabe J, Tate K, Nakamura G, Powers D, Levinson A. The influence of carbohydrate structure on the clearance of recombinant tissue-type plasminogen activator. *Thromb Haemostasis* 1988;60:255–61.
47. Bakhit C, Lewis D, Billings R, Malfroy B. Cellular catabolism of recombinant tissue-type plasminogen activator: identification and characterization of a novel high affinity uptake system on rat hepatocytes. *J Biol Chem* 1987;262:8716–20.
48. Krause J, Seydel W, Heinzel G, Tanswell P. Different receptors mediate the hepatic catabolism of tissue-type plasminogen activator and urokinase. *Biochem J* 1990; 267:647–52.
49. Datta-Mannan A, Witcher DR, Tnag Y, Watkins J, Wroblewski VJ. Monoclonal antibody clearance. Impact of modulating the interaction of IgG with the neonatal Fc receptor. *J Biol Chem* 2007;282(3):1709–17. http://www.jbc.org/cgi/doi/10.10741/jbc.M607161200
50. Richter WF, Gallati H, Schiller CD. Animal pharmacokinetics of the tumor necrosis factor receptor-immunoglobulin fusion protein lenercept and their extrapolation to humans. *Drug Meta Disp* 1999;27(1):21–5.
51. Kneer J, Dumont E, Birnbock H, Kusano T, Gallati H, Lesslauer W. A new TNF-neutralizing agent—lenercept—(TNFR55-IgG1, Ro 45–2081): pharmacokinetic/dynamic data over a 100-fold dose-range in healthy volunteers, and rheumatoid arthritis patients. *Rheumatol Eur* 1996;25 (Suppl 1) 52.

# Use of Animal Models of Disease in the Preclinical Safety Evaluation of Biopharmaceuticals

JOHAN TE KOPPELE, PhD, and RENGER WITKAMP, PhD

**Contents**

1 Introduction
2 Animal Models for Disease
3 Humanized Models for Translational Research
4 From Pharmacology to Toxicology—What Is Predictable?
5 Why Test Safety in Disease Models?
6 Disease Models to Evaluate Undesirable Disease Progression
7 Disease Models to Evaluate Abnormal Kinetic Behavior of the Compound
8 Immunogenicity Issues
9 Concluding Remarks
References

## 1 INTRODUCTION

The objective of the ICH S6 guidance for Preclinical Safety Evaluation of Biotechnology Derived Pharmaceuticals (biopharmaceuticals) [1] is a preclinical safety evaluation program that is based on relevant (animal) models. ICH S6 requires safety testing in a species in which the biopharmaceutical is pharmacologically active. In the absence of a relevant species, the use of transgenic animals expressing the human receptor or the use of homologous proteins should be considered. Animal models for human diseases may also be used. Although these models are mostly used to evaluate pharmacological

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

features of the product, they may be of value in evaluating the safety of biopharmaceuticals:

- To improve prediction of the toxicity or therapeutic index.
- To support the mechanistic interpretation of safety studies (as an alternative or a complement to toxicity studies in normal animals).
- To evaluate the undesirable disease progression.
- To predict a different kinetic behavior of the compound in healthy and diseased animals.

The FDA Critical Path Opportunities Report [2] recognizes that there is a need for "better animal disease or tissue injury models that could provide more accurate predictions of the toxicity of drugs, devices, and biological products that are used in ill or injured patients. Use of such models could also enhance our understanding of the potential toxic effects of compounds associated with many types of medical devices." It is important to note that the regulatory authorities will always require solid scientific justification for the use of these animal models for disease to support safety.

## 2 ANIMAL MODELS FOR DISEASE

Disease models include spontaneous models, induced models (e.g., treated with a compound, a microorganism or surgically), gene knockout(s), and transgenic animals (Table 1). Most such models are based on the effect of a certain stimulus that induces a pathological condition. Examples include the streptozotocin-mediated destruction of beta cells leading to type 1 diabetes [3], surgical dissection of the cruciate ligament resulting in osteoarthritis [4], or type II collagen-specific activation of the immune system, leading to arthritis [5]. A "stimulus leading to pathology" is essentially present in the spontaneous models. "Spontaneous pathologies" are often the result of the anatomical and/or metabolic makeup of a genetically distinct animal strain. For instance, aberrant hip architecture results in hip osteoarthritis in the German Shepherd dog [6]. Selective breeding resulted in the obese Zucker rat (fa/fa) [7,8]. This strain is considered a spontaneous model for obesity and type 2 diabetes because of its hyperglycemia, hyperinsulinemia, insulin-resistance, hypertriglyceridemia, and hypercholesterolemia. Alternatively, elimination of a gene may result in a loss of homeostasis, leading to disease. A particularly well-known general example is the loss of p53 tumor suppressor gene function in the majority of human cancers. In mice, the elimination of Apo lipoprotein E or low-density lipoprotein receptor (ApoE knockout and LDL receptor knockout) results in hyperlipidemic hypercholesterolemia and atherosclerosis [9,10,11].

Anti-sense oligonucleotides and, more recently, siRNA present exciting opportunities to eliminate specific genes or gene transcripts. There is widespread appreciation of siRNA not only for target validation and as a potential

**TABLE 1 Categories of animal models for disease, illustrated with some examples**

| Category | Example | Reference |
|---|---|---|
| Spontaneous models | Spontaneous osteoarthritis: Dunkin Hartley guinea pigs, Shepherd dog | [36] [6] |
| *Mostly resulting from genetic, metabolic, developmental, or anatomical features of specific species and/or animal strains* | Diabetes/obesitas: zucker rat | [7,8] |
| | Diabetes model in the cat | [37] |
| Induced models: Surgically | Porcine wound healing models | [38] |
| | Stenosis in the pig | [39] |
| | Osteoarthritis: Cruciate ligament transection in dogs | [4] |
| | Occlusion (e.g., to induce renal failure) | [33] |
| | Fractures (various species) | [40] |
| Induced models: Chemically | Chemically induced tumors: Mammary tumors, DMBA-rat model | [41] |
| | Urinary bladder tumors (*N*-butyl-N-(4-hydroxybutyl)nitrosamine-induced rat model) | [42] |
| | Diabetes: Streptozotocin, alloxan | [3] |
| | Chemically induced liver failure | [34] |
| | Parkinson's disease models (MPTP) | [43] |
| Induced models: Immunological | Arthritis: Collagen-induced arthritis | [5] |
| | Experimental autoimmune Encephalomyelitis (preclinical model for multiple sclerosis) | [44] |
| | Inflammatory bowel disease/Crohn's disease | [45] |
| | Adoptive transfer of antibodies to the self-antigen, or immune cells to naive animals | [46,47] |
| Infection models (bacterial, viral, fungal, protozoan, parasitic, etc.) | *Listeria monocytogenes* in mice | [29] |
| | Pneumococcal infection model | [48] |
| | HIV-infection models | [49] |
| | Fungal infection models | [50] |
| Xenograft models | Tumor xenografts, various models | [51,52] |
| | Psoriasis xenografts on nude mice | [15] |
| Transgenic models: Knockout, and knock-in | Hyperlipidemia/hypercholesterolemia: ApoE knockout, LDL receptor knockout, ApoE-variant transgenes | [9,10] |
| | Diabetes/obesitas: Ob/ob mouse, db/db mouse | [7,53] |

novel therapeutic approach but also as a tool for disease models, combining time- and tissue-controlled knockout with high transcript specificity.

There has been explosive growth in the number of disease models in recent decades, especially in the field of the knockouts and transgenic rodents. A description of the most frequently used models alone would take a separate volume, and even that would be outdated within no time. Information on the selection of models and background data can easily be found on the Internet. The US National Center for Research Resources (NCRR) provides overviews and links [12]. In addition the main providers of laboratory animals have very useful information on their Web sites. Readers looking for overviews on animal models per disease may find useful information in the *Drug Discovery Today: Disease Models* review journal (http://www.drugdiscoverytoday.com).

In our opinion there are a few important general considerations regarding the use of animal models for disease in this respect:

- A transgenic mouse is not automatically a model for disease. Most diseases are multifactorial. Aberrations and pathological imbalances are not limited to one mechanism, especially in the advanced stages of disease when more mechanisms are offset. In view of the complexity of biological systems it is increasingly recognized that the balanced modulation of multiple targets provides a better therapeutic effect, and a more favorable side-effect profile than the modulation of a single target.
- A disease model should be properly characterized by its phenotype, and enough background data should be available. This also applies for the more than 100,000 knockout mice that will become available over the next few years from large governmental programs (Canada, China, European Union, United States). In addition sufficient historical or reference data on the effects of chemical and biological compounds should be available.
- The model should ideally be available for comparison and be accessible to more than one party. For commercial models, options for noncommercial institutions to perform fundamental or mechanistic studies are recommended.

## 3 HUMANIZED MODELS FOR TRANSLATIONAL RESEARCH

Novel approaches for the design and development of in vivo models to predict drug safety in patients will become more and more relevant. In order to optimize prediction for the human situation, the models should be clearly defined, controllable, and as "humanized" as possible, for instance, using strategies based on novel technologies like siRNA-mediated gene knock-down and on systems enabling the tight and reliable control of human (trans) gene expression in a time- and tissue-specific manner. The so-called translational research to predict human safety (and efficacy) requires continual information exchange

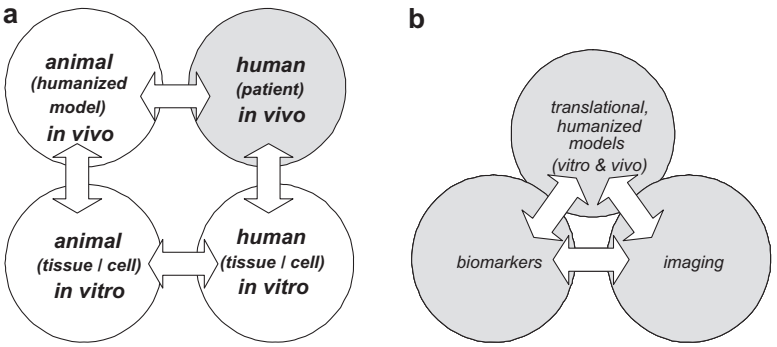


**Figure 1** (*a*) Translational research strategies to predict human safety and efficacy based on in vitro studies (in animal and human cells) and animal studies. (*b*) Use of biomarkers and imaging techniques to increase the predictability of drug R&D.

between in vitro studies (animal and human cells) and animal studies (Figure 1*a*). New approaches for the design and development of safety assessment using human in vitro systems are becoming increasingly important for both biopharmaceutials and pharmaceuticals. Such tests are likely to comprise human stem-cell cultures, or co-cultures, combined with animal disease models. In addition to humanized translational test models, biomarkers and imaging are key to enhancing the predictability of drug research and development (Figure 1*b*). Ideally biomarkers are identical in animals and humans and measurable in vivo and in vitro.

In general, humanized models that aim for improved predictability for the human situation require the insertion of human features in animal-derived test models. Basically, this can be achieved at the gene, cell or even tissue level (Table 2). Insertion of human genes within in vitro test systems is relatively simple with the technology presently at hand. In vivo gene transfer is more complicated. Nowadays transgenic disease animals are considered essential to predict drug efficacy in humans. The ideal transgenic animals are those in which multiple genes can be overexpressed (or silenced) one by one, controllable per tissue/organ, and more important, controllable over time (on/off). This way local manipulation of human gene expression can be useful; overexpression can be achieved with locally administered or delivered vectors, and siRNA technology is full of promise as a suppressor of gene activity.

Human cells, tissues, or organs implanted or transplanted in animals present a relatively strong and valuable tool, provided that immune rejection of the human material can be avoided. Immune-deficient animals like mice may well be used to develop organ-like structures (teratomas) from transplanted stem cells [13] or with combinations of human cells or human tissue (e.g., synoviocytes and cartilage [14], and activated mononuclear cells and transplanted skin [15]).

**TABLE 2 Types and examples of translational, humanized test models**

| | | Example | Reference |
|---|---|---|---|
| | *Transfer of human gene* | | |
| In vitro | Cell transfection | | |
| In vivo local | Local delivery gene delivery by electroporation or a locally administered vector | Interleukin-10 knockout and overexpression on neointima formation in hypercholesterolemic (ApoE*3-Leiden) mice | [54] |
| In vivo systemic | Knock-in transgenic animal, preferably with time- and tissue-specific control of gene expression | hCETP combined with hApoE*3-Leiden | [10] |
| | *Transfer of human cell/humanized cell* | | |
| In vitro | Cocultures of human and animal cells | Synovial fibroblast and articular cartilage | [14] |
| In vivo | Stem cell | Cancer growth in a human tissue microenvironment | [13] |
| | Immune competent cells | SCID mice transplanted with articular cartilage and synoviocytes | [14] |
| | *Transfer of human tissue* | | |
| In vivo | Tissue engineering | | |
| | Tissue/organ tranplantation | Transplantation of tissue on immune-deficient mice: (e.g., [normal] skin from psoriasis patient in combination with administration of activated immune cells) | [15] |

**CASE EXAMPLE: TRANSGENIC HUMANIZED MOUSE MODELS FOR HYPERLIPIDEMIA**

Cardiovascular disease models demonstrate aspects/features of humanized disease models. For hyperlipidemia-based atherosclerosis, there is a lack of a spontaneous in vivo mouse model because mice are resistant to (diet-induced) hyperlipidemia. Therefore a number of transgenic mice have been developed to better understand the pathogenesis in humans and obtain mouse models predictive of the human disease. Overexpression of ApoB, ApoE-variants, and knockout of ApoE, the LDL receptor or LPL (lipopro-

tein lipase), have enabled hyperlipidemic transgenic mice models to be developed. The spontaneous hypercholesterolemic ApoE knockout and the LDL receptor knockout are widely used as atherosclerosis models [9].

Given the multifactorial origin of atherosclerosis, combinations of knock-in and knockout transgenics have been produced. For instance, combining the ApoE*3Leiden transgenic, having a diet-controllable cholesterol profile, with the human CETP gene (cholesterol ester transfer protein) resulted in a transgenic model with human-like LDL as well as HDL profiles that are responsive to treatment with statins and fibrates (as in the human situation). [10] Figure 2

The humanized mouse model is used for safety assessment of drugs that affect lipid profiles as a side effect (and thereby possibly influence the risk of atherosclerosis and, subsequently, myocardial infarction). A typical example is the disturbed lipid profile and increased risk of myocardial infarction experienced by AIDS patients following the long-term use of HIV-1-protease inhibitors [16,17].

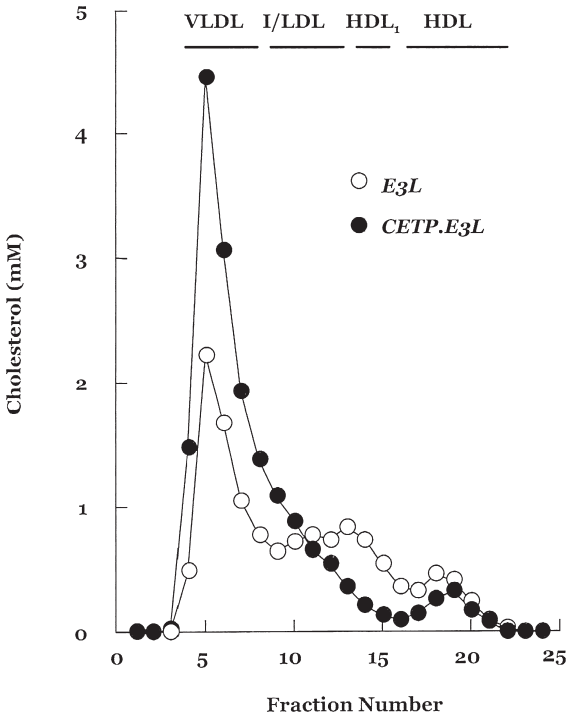


**Figure 2** Effect of inserting a human CETP gene on cholesterol lipoproteins in E3L mice fed a Western-type diet containing 0.25% cholesterol. The E3L.hCETP mouse has shown to respond to statins, fibrates, and the new CETP inhibitor torcetrapib in a similar way as humans do, as manifested by decreased apoB-containing lipoproteins and increased HDL levels.

## 4 FROM PHARMACOLOGY TO TOXICOLOGY—WHAT IS PREDICTABLE?

"Exaggerated pharmacology" often poses an important biopharmaceutical safety issue. Most adverse effects of biopharmaceuticals are a direct consequence of their primary pharmacological action. Off-target effects, apart from hypersensitivity developing during repeated dosing, are rare. For many biopharmaceuticals the immune system is the intended target of the therapy, and the immunotoxicity observed is often exaggerated pharmacology. Biopharmaceuticals that are highly selective in their molecular target are likely to exert mechanism-based side effects if the target is expressed in/on other nontarget cell types. However, although biopharmaceuticals are generally very selective for their primary target, their effects can sometimes be highly pleiotropic. A recent well-known example is calamity with TGN1412, which produced a sudden and rapid release of pro-inflammatory cytokines in healthy volunteers, with severe consequences [18]. Therefore the safety assessment of a biopharmaceutical is critically dependent on an in-depth understanding of the pharmacology and physiology in order to anticipate the risks associated with the potentially intense effects on the target system and interaction with other biological systems. Another important difference between biopharmaceuticals and low molecular weight compounds is the dose–response relationship. Biopharmaceutical products may exhibit unusual dose–response curves that are nonlinear (saturation of target receptors at low dose), bell-shaped, and very steep with an apparent "on/off switch" [19,20]. Taken together, this supports the idea of combining pharmacology and safety studies in the case of biopharmaceuticals. If needed, exaggerated pharmacological effects can also be studied using the animal homologues of the human target. However, as also stated in the S6 guidance [1], the production process, range of impurities/contaminants, pharmacokinetics, and exact pharmacological mechanism(s) may differ between the homologous form and the product intended for clinical use.

## 5 WHY TEST SAFETY IN DISEASE MODELS?

It is well known that the response to a drug can differ between diseased and nondiseased individuals. For example, the expression of both pro- and anti-inflammatory mediators changes in many diseases and conditions such as rheumatic diseases, myocardial infarction, angina, aging, and obesity [21]. This may have an impact on toxicity and pharmacokinetics, and in particular, when the drug interacts with these mediators. In addition the pharmacokinetics of a drug can be affected by disease in general, a topic that is discussed later in this chapter.

Certain classes of drugs exhibit side effects that are characteristic of their group of compounds. In some cases, side effects are inherent to the pharma-

cological effect (exaggerated pharmacology, mechanism-based side effect). In other cases, side effects can differ among drugs belonging to a particular therapeutic class. For example, both infliximab (anti-TNFα monoclonal antibody) and etanercept (TNFα receptor p75Fc fusion protein) share an increased likelihood of opportunistic infections. However, there are indications that tuberculosis and some other granulomatous infections occur more frequently among patients treated with infliximab than among those treated with etanercept [22].

Based on specific disease knowledge, optimization of drug selection involves their common side effect (e.g., dyslipidemia, lipodystrophy of HIV protease inhibitors, [17]). In our laboratory we frequently use transgenic mice carrying the human apolipoprotein E3-Leiden variant gene ([11], see also the case example). These mice exhibit hyperlipoproteinemia and human-like lipoprotein profiles and develop atherosclerosis with all the characteristics of human pathology. In addition to investigating the effects of lipid-lowering drugs or diets, we are increasingly seeing requests to study indirect or side effects. Published examples are the side effect of dexamethasone as an inhibitor of venous graft thickening [23] and the impairment of plasma lipid profiles by Ritonavir [17]. In addition the model has proved its value for investigating the diabetogenic properties of certain drugs.

Despite the obvious demand for animal disease models in evaluating the safety of drugs [1,2], it is difficult to find examples for recently approved biopharmaceuticals or compounds under development. In the European Public Assessment Reports (EPARs) [24] very few registrations describe the use of animal disease models specifically for evaluating the safety of biopharmaceuticals. One illustrative example from the EPAR database is bevacizumab. It is a monoclonal antibody that binds to the human vascular endothelial growth factor (VEGF) to reduce the vascularization of tumors, thereby inhibiting tumor growth. Inasmuch as (VEGF-mediated) vascularization is essential to healing wounds, the concern was raised that the anti-VEGF activity of bevacizumab might delay the wound healing process in tumor patients undergoing biopsy procedures or surgery. A linear incision model (to mimic surgical incision) and a circular wound model (to mimic an ulcerative lesion) were used in rabbits to demonstrate that bevacizumab causes a reversible, dose-related delay in wound healing. Another group of growth factor therapeutics EGFR (epidermal growth factor receptor) inhibitors is used to treat colorectal and non–small-cell lung cancer. On limited occasions unique and dramatic dermatological side effects (skin rash) occurred because of the inhibition of EGFR-signaling pathways in the skin [25]. These two examples illustrate the mechanism-based side effects of the growth modulation factor or their receptors.

With the broadening of the therapeutic applications of biopharmaceuticals, new safety issues can be anticipated in the future. Since the late nineties, rituximab (anti-CD20 monoclonal antibody) has been used to treat non–Hodgkin's

lymphoma. Recently studies on rituximab for the treatment of autoimmune diseases (renal disease and transplantation) have been initiated [26]. Yet another example are the studies performed with recombinant factor VIIa in animal models for trauma to show that rFVIIa was not associated with the systemic activation of coagulation or the microthrombosis of end organs [27].

## 6 DISEASE MODELS TO EVALUATE UNDESIRABLE DISEASE PROGRESSION

Immunosuppression is regarded as a potentially serious adverse effect of certain classes of biopharmaceuticals and may result in increased infection rates [22,28] or increased rates in primary or secondary malignancies. The ability to target and neutralize macrophage-derived inflammatory cytokines and pathways has proved to be one of the most important advances in the treatment of rheumatoid arthritis, Crohn's disease, and several other systemic inflammatory diseases. However, data that have emerged following the approval of infliximab, etanercept, and adalimumab show an increased likelihood of opportunistic infections caused by intracellular organisms including tuberculosis (*Mycobacterium tuberculosis*), *Listeria monocytogenes, Mycobacterium avium intracellulare*, and some fungal species [22]. There are indications that tuberculosis and some other granulomatous infections are likely to occur more frequently among patients treated with infliximab (monoclonal antibodies) than among those treated with etanercept (soluble TNF receptors) [22]. Animal models of tuberculosis have demonstrated the importance of TNFα in controlling and containing intracellular pathogens. An attractive mouse model to study the effects on host resistance to intracellular bacteria and potential differences between compounds is the *L. monocytogenes* model in mice [29]. In our laboratory this model has proved its value in studying the side effects of anti-TNF compounds as manifest by a difference in the *L. monocytogenes* clearance from the spleen. In addition to infections, an increase of malignancies is regarded as a potential side effect of biopharmaceuticals that interact with cytokine functioning. There is indeed some evidence of a dose-dependent increased risk of malignancies, in particular non–Hodgkin's lymphoma, in patients with rheumatoid arthritis treated with anti-TNF antibody therapy [30]. Associations with solid tumors seem to be less clear. Published reports on the effect of biopharmaceuticals on the relative tumor incidence in animal models could not be found. This is probably due to slow development of tumors and the development of antibodies in animals during treatment. In case of etanercept, mice and rats developed neutralizing antibodies within two to three weeks following initiation of twice-weekly administration. Therefore the CHMP (EMEA) concluded that there are probably no meaningful animal studies that can further evaluate the theoretical risk of increased malignancies resulting from chronic TNF inactiva-

tion. Instead, the company will conduct long-term surveillance for tumors in humans [24].

## 7 DISEASE MODELS TO EVALUATE ABNORMAL KINETIC BEHAVIOR OF THE COMPOUND

There are many examples of the abnormal kinetic behavior of drugs during disease. Obviously these effects are dependent on the nature of the compound and the pathophysiological condition. For example, generalized inflammation affects many physiological processes, including organ and peripheral blood flow, protein binding, and enzyme activity. Most data published so far are for low molecular compounds where the main focus has been on biotransformation and drug transporters, which is generally not relevant for biopharmaceuticals. During inflammation the expression of both pro- and anti-inflammatory mediators are changed. This can have pharmacokinetic consequences. Clearance of highly bound and efficiently eliminated drugs may be reduced in the presence of inflammation, resulting in increased circulating drug concentrations [21].

Sepsis, septic shock, and other critical conditions can have profound effects on the kinetics of compounds [31,32]. Drug absorption following intramuscular, subcutaneous, transdermal, and oral administration may decline due to a reduced perfusion of muscles, skin, and splanchnic organs. Compromised tissue perfusion may also affect drug distribution, resulting in a reduced distribution volume. On the other hand, the increase in capillary permeability and interstitial edema during sepsis and septic shock may enhance drug distribution. Changes in plasma protein binding, body water, tissue mass, and pH may also affect drug distribution. The kidneys are an important excretion pathway for many drugs. Renal failure, which often accompanies sepsis and septic shock, will result in the accumulation of both the parent drug and its metabolites. Studies with animal models for renal failure [33], hepatic failure [34], or sepsis may be indicated especially for those compounds that are intended for use in critically ill patients and/or those in intensive care settings.

Other diseases and conditions that can have an effect on the kinetics include cardiac failure, diabetes, and (severe) obesity [35]. Dermal disease and lesions can affect the absorption (unintended) of a compound through the skin. Therefore it may be useful to at least include safety endpoints when testing drugs for psoriasis and dermatitis in animal models.

## 8 IMMUNOGENICITY ISSUES

The immunogenicity issue is discussed in detail elsewhere in this book (see Chapter Immunogenicity of Therapeutic Proteins and the Assessment of Risk). However, in relation to the topics addressed above, the following issues are of specific interest here:

- It is important to know whether the pharmacokinetic behavior of the compound is affected by an immune response.
- The disease severity or progression should not be directly related to immunogenic properties of the compound in the test species. Immunogenicity may cause an effect similar to that observed with adjuvants. In preclinical animals models for rheumatoid arthritis, multiple sclerosis, and inflammatory bowel disease, activation of the immune system with adjuvant is applied to evoke an autoimmune response (e.g., against type II collagen in collagen-induced arthritis, or against myelin-peptides in experimental autoimmune encephalomyelitis as a model for multiple sclerosis).
- Vice versa, the disease may affect the immunogenicity of biopharmaceuticals. For instance, the activated immune system in diseases like rheumatoid arthritis, multiple sclerosis, and inflammatory bowel disease is likely to produce a faster and/or stronger response to biopharmaceuticals. Inasmuch as this applies to patients, the same issue may have relevance for safety testing; that is, application of a relevant inflammatory disease model may give greater insight and better predictability than healthy animals do.

## 9 CONCLUDING REMARKS

The application of animal disease models fits in well with the rational, science-based, case-by-case approach for preclinical safety evaluation of biopharmaceuticals [20]. Drug-related (class effects), disease-related, and therapeutic mechanism-based toxicity require the integration of safety and efficacy research, including kinetics and (if relevant) metabolism. The use of animal models of disease to assess in vivo activity together with safety is expected to provide a better understanding of the therapeutic index, and therefore improved clinical-dose selection. This is especially relevant for biopharmaceuticals inasmuch as their side effects often involve exaggerated pharmacology. Compared to new chemical entities, safety evaluation and pharmacology are less distinct from each other in the case of biopharmaceuticals. So far there are a limited number of examples of approved biopharmaceuticals that have been assessed for safety in animal models of disease. However, we expect a rapid increase in these figures when suitable models become available in greater numbers. It appears that many developers are including animal models of disease, if available, in current programs to answer specific questions.

Continued development and better application of humanized in vitro models (stem cells, cell lines, or ex vivo tissue, including transgenesis) and in vivo models (transgenesis and transplantation) that closely mimic the human situation should ultimately improve the predictive value of preclinical studies to human safety.

## REFERENCES


1. ICH S6 Guidance. http://www.ich.org/
2. FDA Critical Path Opportunities Report. http://www.fda.gov/oc/initiatives/criticalpath/reports/opp_report.pdf
3. Rees DA, Alcolado JC. Animal models of diabetes mellitus. *Diab Med* 2005;22: 359–70.
4. Brandt KD. Transection of the anterior cruciate ligament in the dog: a model of osteoarthritis. *Sem Arthritis Rheum* 1991;21:22–32.
5. Van den Berg WB. Lessons from animal models of arthritis. *Cur Rheum Rep* 2002;4:232–9.
6. Smith GK, Mayhew PD, Kapatkin AS, McKelvie PJ, Shofer FS, Gregor TP. Evaluation of risk factors for degenerative joint disease associated with hip dysplasia in German Shepherd Dogs, Golden Retrievers, Labrador Retrievers, and Rottweilers. *J Am Vet Med Assoc* 2001;219:1719–24.
7. Chen D, Wang MW. Development and application of rodent models for type 2 diabetes. *Diab Obes Metab* 2005;7:307–17.
8. Mathe D. Dyslipidemia and diabetes: animal models. *Diab Metab* 1995;2:106–11.
9. Huang Y. Transgenic and gene-targeted mice in the study of hyperlipidemia. In Xu Q, ed. *A Handbook of Mouse Models of Cardiovascular Disease*. New York: Wiley, 2006;33–41.
10. Westerterp M, van der Hoogt CC, de Haan W, Offerman EH, Dallinga-Thie GM, Jukema JW, Havekes LM, Rensen PC. Cholesteryl ester transfer protein decreases high-density lipoprotein and severely aggravates atherosclerosis in APOE*3-Leiden mice. *Arterioscl Thromb Vas Biol* 2006;26:2552–9.
11. van den Maagdenberg AM, Hofker MH, Krimpenfort PJ, De Bruijn I, Van Vlijmen B, Van der Boom H, Havekes LM, Frants RR. Transgenic mice carrying the apolipoprotein E3-Leiden gene exhibit hyperlipoproteinemia. *J Biol Chem* 1993; 268:10540–5.
12. National Center for Research Resources (NCRR). http://www.ncrr.nih.gov/comparative_med.asp
13. Tzukerman M, Skorecki KL. A novel experimental platform for investigating cancer growth and anti-cancer therapy in a human tissue microenvironment derived from human embryonic stem cells. *Meth Mol Biol* 2006;331:329–46.
14. Muller-Ladner U, Gay S. The SCID mouse—a novel experimental model for gene therapy in human rheumatoid arthritis. *Drugs Today (Barcelona)* 1999;35:379–88.
15. Caspary F, Elliott G, Nave BT, Verzaal P, Rohrbach M, Das PK, Nagelkerken L, Nieland JD. A new therapeutic approach to treat psoriasis by inhibition of fatty acid oxidation by Etomoxir. *Brit J Dermatol* 2005;153:937–44.
16. Carr A, Samaras K, Chisholm DJ, Cooper DA. Pathogenesis of HIV-1-protease inhibitor-associated peripheral lipodystrophy, hyperlipidaemia, and insulin resistance. *Lancet* 1998;351:1881–3.
17. den Boer MA, Berbee JF, Reiss P, van der Valk M, Voshol PJ, Kuipers F, Havekes LM, Rensen PC, Romijn JA. Ritonavir impairs lipoprotein lipase-mediated lipolysis and decreases uptake of fatty acids in adipose tissue. *Arterioscl Thromb Vas Biol* 2006;26:124–9.

18. Suntharalingam G, Perry MR, Ward S, Brett SJ, Castello-Cortes A, Brunner MD, Panoskaltsis N. Cytokine storm in a phase 1 trial of the anti-CD28 monoclonal antibody TGN1412. *N Engl J Med* 2006;355:1018–28.
19. Thomas PT. Nonclinical evaluation of therapeutic cytokines: immunotoxicologic issues. *Toxicology* 2002;174:27–35.
20. Cavagnaro JA. Preclinical safety evaluation of biotechnology-derived pharmaceuticals. *Nat Rev Drug Discov* 2002;1:469–75.
21. Kulmatycki KM, Jamali F. Drug disease interactions: role of inflammatory mediators in disease and variability in drug response. *J Pharm Pharmaceut Sci* 2005;16: 602–25.
22. Winthrop KL. Risk and prevention of tuberculosis and other serious opportunistic infections associated with the inhibition of tumor necrosis factor. *Nat Clin Pract Rheumatol* 2006;2:602–10.
23. Schepers A, Pires NM, Eefting D, de Vries MR, van Bockel JH, Quax PH. Short-term dexamethasone treatment inhibits vein graft thickening in hypercholesterolemic ApoE3Leiden transgenic mice. *J Vas Surg* 2006;43:809–15.
24. European Public Assessment Reports (EPARs) database. http://www.emea.europa.eu/htms/human/epar/eparintro.htm
25. Agero AL, Dusza SW, Benvenuto-Andrade C, Busam KJ, Myskowski P, Halpern AC. Dermatologic side effects associated with the epidermal growth factor receptor inhibitors. *J Am Acad Dermatol* 2006;55:57–70.
26. Salama AD, Pusey CD. Drug insight: rituximab in renal disease and transplantation. *Nat Clin Pract Nephrol* 2006;2:221–30.
27. Schreiber MA. Preclinical trauma studies of recombinant factor VIIa. *Crit Care* 2005;9:S25–8.
28. Giles JT, Bathon JM. Serious infections associated with anticytokine therapies in the rheumatic diseases. *J Intens Care Med* 2004;19:320–34.
29. Edelson BT, Unanue ER. Immunity to *Listeria* infection. *Curr Opin Immunol* 2000;12:425–31.
30. Bongartz T, Sutton AJ, Sweeting MJ, Buchan I, Matteson EL, Montori V. Anti-TNF antibody therapy in rheumatoid arthritis and the risk of serious infections and malignancies: systematic review and meta-analysis of rare harmful effects in randomized controlled trials. *JAMA* 2006;295:2275–85.
31. van Dalen R, Vree TB. Pharmacokinetics of antibiotics in critically ill patients. *Intens Care Med* 1990;16(Suppl 3):S235–8.
32. De Paepe P, Belpaire FM, Buylaert WA. Pharmacokinetic and pharmacodynamic considerations when treating patients with sepsis and septic shock. *Clin Pharmacokinet* 2002;41:1135–51.
33. Lieberthal W, Nigam SK. Acute renal failure. II. Experimental models of acute renal failure: imperfect but indispensable. *Am J Physiol Ren Physiol* 2000;278: F1–12.
34. Bélanger M, Butterworth RF. Acute liver failure: a critical appraisal of available animal models. *Metab Brain Dis* 2005;20:409–23.
35. Cheymol G. Effects of obesity on pharmacokinetics implications for drug therapy. *Clin Pharmacokinet* 2000;39:215–31.

36. Jimenez PA, Glasson SS, Trubetskoy OV, Haimes HB. Spontaneous osteoarthritis in Dunkin Hartley guinea pigs: histologic, radiologic, and biochemical changes. *Lab Animal Sci* 1997;47:598–601.
37. O'Brien TD. Pathogenesis of feline diabetes mellitus. *Mol Cell Endocrinol* 2002;197(1–2):213–9.
38. Sullivan TP, Eaglstein WH, Davis SC, Mertz P. The pig as a model for human wound healing. *Wound Repair Regener* 2001;9:66–76.
39. Kelly BS, Heffelfinger SC, Whiting JF, et al. Aggressive venous neointimal hyperplasia in a pig model of arteriovenous graft stenosis. *Kidney Int* 2002;62: 2272–80.
40. Nunamaker DM. Experimental models of fracture repair. *Clin Ortho Rel Res* 1998;355S:S56–65.
41. Morse MA, Baird WM, Carlson GP. Distribution, covalent binding, and DNA adduct formation of 7,12-dimethylbenz(a)anthracene in SENCAR and BALB/c mice following topical and oral administration. *Cancer Res* 1987;47:4571–5.
42. Grubbs CJ, Moon RC, Squire RA, Farrow GM, Stinson SF, Goodman DG, Brown CC, Sporn MB. 13-cis-retinoic acid: inhibition of bladder carcinogenesis induced in rats by *N*-butyl-*N*-(4-hydroxybutyl)nitrosamine. *Science* 1977;198:743–4.
43. Yang SC, Markey SP, Bankiewicz KS, London WT, Lunn G. Recommended safe practices for using the neurotoxin MPTP in animal experiments, *Lab Animal Sci* 1988;38:563–7.
44. Nagelkerken L, Blauw B, Tielemans M. IL-4 abrogates the inhibitory effect of IL-10 on the development of experimental allergic encephalomyelitis in SJL mice. *Int Immunol* 1997;9:1243–51.
45. Hoffmann JC, Pawlowski NN, Kuhl AA, Hohne W, Zeitz M. Animal models of inflammatory bowel disease: an overview. *Pathobiology* 2002–2003;70:121–30.
46. Kim HY, Kim S, Chung DH. FcgammaRIII engagement provides activating signals to NKT cells in antibody-induced joint inflammation. *J Clin Invest* 2006;116: 2484–92.
47. Rao P, Segal BM. Experimental autoimmune encephalomyelitis. *Meth Mol Med* 2004;102:363–75.
48. Orihuela CJ, Tuomanen EI. Models of pneumococcal disease. *Drug Discov Today Dis Mod* 2006;3:69–75.
49. Stoddart CA, Reye RA. Models of HIV-1 disease: a review of current status. *Drug Discov Today Dis Mod* 2006;3:113–9.
50. Steinbach WJ, Zaas AK. Newer animal models of Aspergillus and Candida infections. *Drug Discov Today Dis Mod* 2004;1:87–93.
51. Sausville EA, Burger AM. Contributions of human tumor xenografts to anticancer drug development. *Cancer Res* 2006;66:3351–4.
52. Golas JM, Lucas J, Etienne C, Golas J, Discafani C, Sridharan L, Boghaert E, Arndt K, Ye F, Boschelli DH, Li F, Titsch C, Huselton C, Chaudhary I, Boschelli F. FSKI-606, a Src/Abl inhibitor with in vivo activity in colon tumor xenograft models. *Cancer Res* 2005;65:5358–64.
53. Allen TJ, Cooper ME, Lan HY. Use of genetic mouse models in the study of diabetic nephropathy. *Curr Atheroscl Rep* 2004;6:197–202.

54. Eefting D, Schepers A, De Vries MR, Pires NM, Grimbergen JM, Lagerweij T, Nagelkerken LM, Monraats PS, Jukema JW, van Bockel JH, Quax PH. The effect of interleukin-10 knock-out and overexpression on neointima formation in hypercholesterolemic APOE*3-Leiden mice. *Atherosclerosis* 2007;193:335–42.

# Safety Pharmacology: Similarities and Differences between Small Molecules and Novel Biopharmaceuticals

EDWARD W. BERNTON, MD

**Contents**

1 Introduction
2 Safety Pharmacology Regulatory Requirements for Small Molecules and Biopharmaceuticals
  2.1 The "Core Battery" Approach for Small-Molecule Safety Pharmacology
  2.2 Safety Pharmacology for Biopharmaceuticals
3 Testing for QT Liability and Torsadogenic Potential of New Drugs
  3.1 Background and History
  3.2 Preclinical Evaluation for QT Liability
  3.3 Clinical Evaluation of QT Liability—When Is It Applicable to Biopharmaceuticals?
4 Adding Predictive Safety Biomarkers into Standard Toxicology Protocols for Biopharmaceuticals
  4.1 CNS Assessments
  4.2 Inflammation and Immune Activation or Suppression
  4.3 Cell Proliferation or Apoptosis in Tissue Samples
  4.4 End-organ Toxicities
5 An Integrated Approach to Preclinical and Clinical Immunotoxicity Assessments
  5.1 Regulatory Guidances
  5.2 New Techniques for Evaluating Immunosuppressive Activity
6 Novel Pharmacodynamic Measures Using Flow Cytometry and Blood Mononuclear Cells as a Surrogate for Tissue Responses
7 Low Predictive Value of Certain Toxicity Studies in Nonhuman Primates
8 Summary
References

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

## 1 INTRODUCTION

Safety pharmacology studies were given regulatory definition in the 2001 ICH S7A guidance [1]. They were differentiated from primary and secondary pharmacodynamic studies used for exploration of structure–activity relationships and lead selection and validation, which generally are not directly supportive of safety of the proposed human uses and dose ranges. Safety pharmacology is defined as "those studies that investigate the potential undesirable pharmacodynamic effects of a substance . . . in relation to exposure in the therapeutic range and above." The emphasis for small molecule drugs is on (1) identification of unsuspected "off-target" activities independent of the primary pharmacodynamic effect, (2) further investigation of adverse effects suspected from either clinical or preclinical studies, or (3) detailed investigation of adverse effects that are commonly associated with the specific therapeutic class of the drug candidate (i.e., potassium ion channel cardiac effects with neuroleptics). The safety evaluation of new biopharmaceuticals is often made more difficult by the higher percentage of novel drug targets, compared to small molecules, especially for monoclonal antibodies and soluble receptors. Target specificity is usually very high. For these products, safety pharmacology tends to be a more focused investigation of highly specific drug–target interactions, and an in-depth characterization of pharmacodynamic effects.

## 2 SAFETY PHARMACOLOGY REGULATORY REQUIREMENTS FOR SMALL MOLECULES AND BIOPHARMACEUTICALS

### 2.1 The "Core Battery" Approach for Small-Molecule Safety Pharmacology

The ICH S7A guidance recommends that the safety pharmacology studies be designed with specific consideration of the pharmacologic activities of the test article. It recognizes, however, that in early development, sufficient information may not be available to design such studies. In this case a more general approach to safety pharmacology is recommended, referred to as the "core battery."

The core safety pharmacology is designed to detect organ-based toxicities, with the emphasis on cardiovascular, central nervous, and respiratory systems, where even transient dysfunction secondary to drug exposure can be life-threatening. Other organ systems, such as the gastrointestinal, renal, or hematopoetic symptoms can be transiently impaired by adverse pharmacodynamic effects without causing irreversible harm.

A core battery is normally hierarchical in structure. Screening tests are suggested in section 2.7 of the guidance for CNS, cardiovascular, and respira-

tory systems. Examples are given of possible follow-up studies providing more detailed analysis of pharmacologic effects. For example, blood pressure, heart rate, and ECG, ideally by telemetry in an unrestrained animal, can provide initial cardiovascular safety assessment. If abnormalities are detected, then full analysis via cardiac catheterization might be performed to study cardiac output, ventricular contractility, vascular resistance, and interactions with endogenous mediators such as vagolytic or vagotonic agents, catecholamines, and so forth. In vitro testing such as the hERG screen for potential effects on the potassium channel cardiac repolarization can also be performed as safety pharmacology, if not performed as screens during the lead selection process. Typical CNS core batteries might be the Irwin behavioral battery in mice or rats, which can be expanded to include locomotor activity, rotatating rod tests of coordination, analgesia screens, PTZ seizure threshold, and/or more extensive characterization of dose–response curves if indicated.

The core safety pharmacology battery can also be performed with a small molecule's major metabolite, particularly if that metabolite is unique to human metabolism and the toxicology species did not have significant exposure to it.

All these studies should be performed in compliance with good laboratory practice (GLP) where feasible. The screening battery is frequently performed with full GLP. More complex or specialized studies should be well-designed, properly controlled, and carefully documented "when not in compliance." When safety pharmacology endpoints are incorporated into toxicology studies, not an uncommon situation, they should be conducted under GLP.

Certain information from commonly used lead selection screens may also suggest specific safety pharmacology screens for small molecules, such as the results of standardized receptor binding screens, cross-species plasma binding profiles, in vitro metabolism profiles, and in vitro studies on cell phenotype or function. It should be pointed out that standard single- and repeat-dose toxicology protocols and the usual in-life observations, clinical laboratory parameters, and histopathology are relatively insensitive means for detection of nonfatal abnormalities in the organ function of central nervous, respiratory, and cardiovascular systems, and in particular the immune system. Nonetheless, in general experience it is relatively uncommon for the routine safety pharmacology battery to disclose dose-limiting toxicities that can alter the design of human trials if these were not observed in primary and secondary pharmacology studies or in GLP toxicology studies. Additionally diseased patient populations may have increased sensitivity to pharmacodynamic effects of the test article (i.e., cancer patients or tumor-bearing test animals to circulatory effects of TNF-alpha and other cytokines, potentiation of hematotoxicity, nephrotoxicity, cardiotoxicity by other concurrent cancer therapies, safety in renal- or hepatic-impaired populations) not reflected in standard safety pharmacology.

## 2.2 Safety Pharmacology for Biopharmaceuticals

For biotechnology-derived pharmaceuticals, including proteins, polypeptides, virally vectored vaccines, or gene-therapy therapeutics, and oligonucleotides, safety pharmacology tends to be a more focused investigation of highly specific drug–target interactions. Interspecies variation in metabolism and metabolites, clearance, and protein-binding becomes less of an issue, but the frequent species specificity of both the target interaction (i.e., monoclonal antibodies, certain growth factors, and cytokines) and postreceptor signaling and effects, particularly for immunomodulators, makes selection of appropriate in vivo models more difficult.

Recognizing these issues, the S7A guidance states, "For biotechnology-derived products that achieve highly specific receptor targeting it is often sufficient to evaluate the safety pharmacology endpoints as part of toxicology and/or pharmacodynamic studies; therefore the safety pharmacology studies can be reduced or eliminated for these products." But because of the challenge that many biopharmaceuticals present, some quite species-specific in their binding interactions, for predictive toxicology and safety pharmacology, in 1997 the ICH published a consensus guidance for such preclinical safety evaluations. This S6 guidance, "Preclinical Safety Evaluation of Biotechnology-Derived Pharmaceuticals" [2], together with the ICH S8 guidance "Immunotoxicity Studies for Human Pharmaceuticals" [3] provides some further information relevant to safety pharmacology for biopharmaceuticals, many of which have immunomodulatory activities.

The safety evaluation of new biopharmaceuticals is made more difficult by the rising percentage of novel drug targets, compared to small molecules, especially for monoclonal antibodies. Often the function and even the distribution of these novel targets is incompletely characterized or understood. Thus the thorough delineation of pharmacologic effects and dose-limiting human toxicities can require an extremely thoughtful, individualized, and product-specific research program.

For biopharmaceuticals such as growth factors and cytokines with agonistic properties, transgenic knock-in mice which overexpress the murine homologue of the candidate molecule can provide useful information, particularly regarding chronic drug exposures and potential developmental effects of the candidate. Alternately, mice can be dosed with an active murine homologue of recombinant DNA origin, if the human protein lacks full activity in this test species. In the case of such surrogates for monoclonal antibody products with restricted species cross-reactivity, the model should be carefully validated by in vitro studies of mouse target binding affinities and pharmacodynamic activity, as described by Clarke et al. [4]. For soluble receptors and monoclonal antibody therapeutics designed to interfere with signaling via a specific receptor or ligand, siRNA can be used to silence the therapeutic target, both in vitro and in preclinical models, allowing investigation of the pharmacologic consequences in both healthy animals and models of disease [5].

Alternatively, a knockout transgenic mouse strain can be similarly utilized to understand the safety issues involved in target inhibition or neutralization. For example, the phenotype of CXCR5 and integrin beta 6 knockout mice helped characterize the pharmacodynamic and pathologic consequences of human monoclonal antibodies blocking both these receptors. CXCR5 null mice had defects in delayed-hypersensitivity responses but resistance to a common preclinical inflammatory bowel disease model. Integrin beta 6 null mice demonstrated chronic inflammation in the lung and skin but resistance to bleomycin-induced pulmonary fibrosis [6–8]. Various humanized transgenic mouse strains also can be used to investigate toxicity and pharmacodynamic responses of specific biopharmaceuticals interacting specifically with human immune-cell targets [9]. In preclinical disease models all of these strategies may be part of early target validation work, but such models also have great utility at a later point when utilized with a greater emphasis on safety pharmacology endpoints. As Green and Black state in their useful review of safety assessment for immunomodulatory biopharmaceuticals [10], "In this field of development, preclinical models often need to reflect recent technology innovations; therefore, these models are not always 'validated' in the conventional sense. Experience to date suggests that improved methods and approaches are needed as these agents are developed for use in lower or moderate risk patient populations."

In evaluating the suitability of animal test species for safety pharmacology, it is essential to confirm that the binding affinity of the product and its target is similar in the tissues of the test species and in human tissues. Similarly the distribution of the target of monoclonal antibody therapies should compare in tissue cross-reactivity studies in both human tissues and those from the test species. Primary cell cultures or cell lines derived from mammalian cells can be used to examine the effects of the biopharmaceutical on cell phenotype, signaling pathways, cytokine secretion, and proliferation, and these may provide useful information on the relevance or lack of sensitivity of a proposed in vivo safety species. The S6 guidance specifically states, "toxicity studies in nonrelevant species may be misleading and are discouraged." Examples of the validation of pharmacodynamic activity of the biopharmaceutic in toxicology and safety species are provided in publications by Hart et al. that review the preclinical studies for monoclonal antihuman IL-4 and IL-5 antibodies [11,12].

The species-specificity of activity for some biopharmaceuticals has immediate implications when one considers the dependence of the standard "core" and "extended" safety pharmacology batteries on rodent models, or in some cases dogs and minipigs, or when absolutely required, nonhuman primates [13]. For biopharmaceuticals with highly species-specific drug–target interactions, opportunities for traditional safety pharmacology screens can be highly constrained, and if primate studies are required, both expensive and time-consuming. Thus safety pharmacology studies for pharmaceuticals tend to be limited to those clearly required to further evaluate findings in standard toxicology

studies or else clearly required by considerations such as mechanism of action or drug class.

The pharmacokinetic evaluation of biopharmaceuticals is generally simplified by the usual metabolism of products to small peptides and to amino acids, and thus classical biotransformation and metabolism studies are rarely necessary. Routine studies to assess mass balance are not useful. However, both single- and multiple-dose toxicokinetic data are essential in safety pharmacology asessments, and these can be complicated by two factors: (1) biphasic clearance with a saturable, initial, receptor-dependent clearance phase, which may cause nonlinearity in dose–exposure relationships and doseresponses [14] and (2) antibody production against an antigenic biopharmaceutical that can alter clearance or activity in more chronic repeat-dose safety studies in the preclinical model.

In general, both dose–exposure and exposure–response relationships tend to be less linear and predictable for many biopharmaceuticals than is seen with most small molecule drugs. This can involve saturable receptor-binding, upregulation of soluble receptors that alter clearance, stimulation of biologic cascades by agonists at immune cell receptors, and temporal disassociation of pharmacologic effects from plasma levels due to persistence and activity in nonplasma compartments or the delayed induction of secondary mediator release. All such factors make the prediction of both estimates of exposure and adverse effect profiles in humans generally less reliable than that based on allometric scaling from toxicokinetic studies of small molecules, and argue for conservative human starting doses and cautious dose escalation in the initial clinical studies of novel biopharmaceuticals.

The induction of antibody formation against biopharmaceuticals in animal test species is not highly predictive for the potential of antibody formation in humans. Humans may also develop antibodies to human or humanized proteins that may not necessarily abrogate a therapeutic response. However, in repeat-dose safety animal studies with a biopharmaceutical, it is imperative that the sponsor (developer) establish that any antibody response detected does not neutralize the biopharmceutical or significantly alter its clearance or the animal's exposure. Product sponsors should modify the interpretation of safety results to account for any such attenuation of exposure. In general, biopharmaceuticals given by an intravenous route can be expected to be less antigenic than those administered subcutaneously.

## 3 TESTING FOR QT LIABILITY AND TORSADOGENIC POTENTIAL OF NEW DRUGS

### 3.1 Background and History

Relationships between the whimsically named human ether-a-go-go related gene (hERG), drug-induced QT interval prolongation, and the torsades de

pointes (TdP) arrhythmia are now a serious and constant theme in drug safety evaluation. hERG encodes the inward rectifying potassium channel (i(Kr)). Mutations in five cardiac potassium rectifier channel genes, including KCNQ1, HERG, SCN5A, KCNE1, and KCNE2, constitute the principal cause of inherited long-QT syndrome (LQTS). Typically each family carries its own private mutation, and the disease manifests with varying phenotype and incomplete penetrance, even within particular families. Individuals with a high-penatrance channel gene mutation have a prolonged QT interval on their electrocardiogram and have a propensity toward serious arrhythmias, including the often fatal TdP.

In the 1990s a pattern of sudden death and occurrence of TdP was noticed in patients taking a variety of commonly prescribed drugs, most notably terfenadine, astimazole, and cisipride. This was usually associated with a prolonged QT interval on the electrocardiogram. Further investigation found that these drugs could cause QT prolongation and subsequent arrhythmia, usually at elevated blood levels or in susceptible individuals. Virtually all drugs known to cause TdP block the rapidly activating component of the delayed rectifier potassium current (I(Kr)). Arrhythmias are more likely to occur if drug-induced QTc prolongation coexists with other risk factors, such as individual susceptibility, presence of congenital long-QT syndromes, heart failure, bradycardia, electrolyte imbalance, drug overdose of a QTc prolonging drug, female sex, restraint, old age, hepatic or renal impairment, or slow metabolizer status.

In particular, the interaction of ketoconazole, which inhibits CYP3A4 metabolism of terfenadine and cisipride, was found to be associated with higher risk of TdP. These discoveries led to withdrawal certain drugs from the market and black-box labeling of risk of QT prolongation and torsades for others. As a class effect, most class Ia and III antiarrhythmic drugs prolong QT, but unlike antihistamines or antireflux therapies, these are used to treat serious cardiac disease and titrated to effect with careful patient and ECG observation. Therefore these drugs have remained on the market. Other noncardiac drugs are now marketed with label warnings regarding QT liability. These include chloroquin, clarithromycin, domperidone, droperidol, erythromycin, haloperidol, methadone, pentamidine, and thioridazine. Patients with known long-QT syndrome are cautioned to avoid using them.

Nevertheless, the FDA considered a risk of fatal cardiac arrhythmias, in even a minuscule percentage of patients, to be a toxicity that needed to be exhaustively evaluated prior to marketing authorization of a new small molecule drug. In 2005 the agency published Guidance E14, "Clinical Evaluation of QT/QTc Interval Prolongation and Proarrythmic Potential for Nonantiarrhythmic Drugs." This guidance outlined the "definitive QT study" as a mandatory clinical evaluation for new drugs, as well as other practices for analysis of arrhythmagenic potential in phase 3 study data. The guidance established as a drug toxicity the drug effect on the QT interval as a surrogate for risk of TdP.

### 3.2 Preclinical Evaluation for QT Liability

In 2005 the FDA published Guidance S7B, "Nonclinical Evaluation of the Potential for Delayed Ventricular Repolarization by Human Pharmaceuticals," considered an extension of the S7A safety pharmacology guidance. The S7B guidance states, "the objectives of studies are to: (1) identify the potential of a test substance and its metabolites to delay ventricular repolarization, and (2) relate the extent of delayed ventricular repolarization to the concentrations of a test substance and its metabolites. The study results can be used to elucidate the mechanism of action and, when considered with other information, estimate risk for delayed ventricular repolarization and QT interval prolongation in humans."

The testing should include in vitro assays that evaluate drug effects on current through a native or expressed potassium rectifier channel protein, such as that encoded by hERG or measurement of ion currents in drug-exposed isolated animal or human myocytes or cultured cell lines. It should also include in vivo electrophysiologic studies of drug effects on the relationship between QT and RR intervals, using dog, monkey, swine, ferret, or rabbit. Use of mice or rats is not recommended, since the primary ion currents controlling repolarization are different than in humans. The guidance suggests that the in vivo assay can be integrated into the cardiovascular safety pharmacology study. The fact that high-sensitivity in vitro QT assays have relatively low specificity has led to the development of additional models for electrophysiologic studies, such as the arterially perfused rabbit ventricular wedge preparation [15,16].

In vitro studies need to be well-designed, with known QT-prolonging or TdP-associated drugs used as positive controls, and careful characterization of the test item dose–response. A drug's IC50 for hERG inhibition or for other measures of the effect on cardiac repolarization is usually compared to the effective therapeutic plasma concentration of the unbound drug (ETPC(unbound)). If the IC50/ETPC ratio is greater than 50 to 100, this generally predicts a lack QT liability of the drug when dosed in patients, and negative results in a definitive QT interval clinical study. However, lower values can be poorly predictive. For example, drugs withdrawn from the market due to TdP or drugs with a measurable incidence of TdP in humans have hERG IC50/ETPC ratios ranging from 0.1 to 31. For the majority of drugs with no reports of TdP in humans, the ratio is greater than 30 [17].

### 3.3 Clinical Evaluation of QT Liability—When Is It Applicable to Biopharmaceuticals?

The E14 guidance on clinical evaluation of the QT/QTc interval prolongation states: "an adequate premarketing investigation of the safety of a new phar-

maceutical agent should include rigorous characterization of its effects on the QT/QTc interval." This evaluation should be performed at both the steady-state therapeutic exposure, and at three- to fourfold higher exposures, to model inadvertent overdose or impaired drug clearance. Sometimes preliminary clinical studies are required to establish the maximum tolerated steady-state exposure to be used in the definitive QT study. Studies must be double-blinded and placebo-controlled, and incorporate a positive control, such as moxifloxin, to establish the sensitivity of the trial to detect a 10 ms average effect on QTc. (QT interval corrected for heart rate). Study designs can involve a multiperiod crossover or parallel groups. When possible these studies are done in healthy volunteers. Careful attention must be given to replicate sets of ECGs to minimize inherent variability and QT hysteresis in QTc measurements, to provide an adequate ECG baseline to correct for circadian and meal effects, and to obtain adequate sampling of blood levels to document relationships of ECG changes with drug exposure. Usually a commercial "core ECG lab" is utilized to conduct the ECG analysis. All ECG tracings and the measurements made to determine the QT interval must be archived in a format suitable for submission to the FDA, using a digital file format.

Typically these studies cost from $1.5 to $3 million to conduct. Based on the perceived probability of a significant QT interval effect (based on preclinical studies, class effects, and ECG data from phase 1 studies) a decision must be made whether to conduct the definitive QT study prior to proceeding with a proof-of-concept study in patients, or whether to delay this study until the proof of concept (POC) has been demonstrated. Of note, for many biopharmaceutical products it would not be possible nor ethical to dose to steady state in healthy volunteers, to dose at two to four times anticipated therapeutic exposures, or to use a crossover design with reasonable washout periods. Thus a QT study performed with a biopharmaceutical may need to vary from the usual design and the E14 guidance, and may present great challenges for subject or patient recruitment.

Historically, therapy with biopharmaceuticals have not been felt to associated with risk of TdP. It was felt that the specificity of receptor interactions common to cytokines, growth factors, and monoclonal antibodies made high-affinity off-target interactions with ion channel proteins extremely unlikely. A theoretical exception to this might be certain neuropeptides, some of which are known to modulate certain neuronal ion channels. To this author's knowledge, no biologic therapy has been associated with TdP with the exception of vasopressin [18], where the arrhythmia was secondary to vasopressin-induced bradycardia. A search of the National Library of Medicine bibliographic database crossing either the term "torsades" or "arrhythmia" with the terms "interleukin," "monoclonal antibody," "TNF-alpha," "erythropoietin," "colony-stimulating factor," or "cytokine therapy" revealed only one report, which described the association of interleukin-11 therapy for thrombocytopenia in

elderly patients with atrial fibrillation. However, in vitro electrophysiologic studies showed no direct cardiac membrane effects of therapeutic concentrations of IL-11 [19].

This is not to say that cardiotoxicity is not seen with biopharmaceuticals. Cardiomyopathy is now a well-recognized complication of trastuzumab and and has been reported with bevacizumab treatment, in particular in combination with other cytotoxic cancer therapies [20]. Myocarditis and pericarditis are a well-documented complications of vaccinia immunization [21], and could also complicate use of a pox-virus vector for other therapeutics. In 1995 Genetics Institute suspended phase 2 cancer trials of Interleukin-12 for serious toxicities including cardiac arrhythmia. However, such toxicities are best detected by incorporation of biomarkers for myocardial damage such as troponin-T into preclinical and early clinical studies, and continual ECG monitoring for arrhythmia in preclinical and early clinical studies, *not* by in vitro explorations of electrophysiology.

With the transfer of most biopharmaceutical INDs from CBER to CDER in 2003, there has been an increased tendency to apply the small-molecule paradigm for evaluation of QT liability to biopharmaceutical product candidates, and to request information on hERG assays or plans for definitive clinical QT studies. This does not seem reasonable based on the postmarketing safety data for biopharmaceuticals, nor on scientific grounds as discussed above. If these investigations become routinely required, they will only add significant time and costs to the process of biopharmaceutical product evaluation and have little ultimate impact on patient safety.

While the E14 guidance does not exempt biopharmaceuticals from the requirement for "rigorous characterization" of QT effects, in the past both scientific and practical considerations limited its application to these therapies. While there are no reports of TdP secondary to a biopharmaceutical, and currently a leading cardiac core lab provider knew of no definitive QT studies submitted for biologic products, it is clear that such a study will be required for at least one new biopharmaceutical, an erythropoietin. Regarding regulatory strategy, sponsors of new biopharmaceuticals must either inquire early (at a pre-IND meeting) if a definitive QT study will be required for marketing approval, and risk a positive answer in the absence of clear FDA policy, or wait until an end-of-phase-2 meeting, and plan such studies, if needed, in parallel with pivotal phase 3 trials. Given the extremely low probability of biopharmaceutical product development being limited by QT interval effects, this second strategy would appear more reasonable. While in the past the various CDER divisions frequently handled QTc questions with consults to the Cardiorenal Division, now the Interdiscplinary Review Team (IRT), provides this expertise for the primary review divisions, upon request.

## 4 ADDING PREDICTIVE SAFETY BIOMARKERS INTO STANDARD TOXICOLOGY PROTOCOLS FOR BIOPHARMACEUTICALS

The usual GLP 30- or 60-day repeat-dose toxicology study with a recovery group offers an opportunity to perform a more systematic investigation of the more subtle pharmacodynamic or toxicologic effects of biopharmaceuticals than those endpoints usually incorporated into such protocols. Some of these demand tissue samples, but many involve noninvasive biomarkers that can be carried forward into early phase human studies. These might include CNS assessments, inflammation and immune activation or suppression, cell proliferation or apoptosis in tissue samples, and end-organ toxicities.

### 4.1 CNS Assessments

An Irwin test at one or several study time points, for rodent studies, or a 24 hour period of actigraphy in a cage designed to measure spontaneous locomotor activity. In early clinical studies, clinical actigraphy with wrist-worn recording devices can detect not only hypo- or hyperactivity but changes in sleep latency, sleep fragmentation, and daytime sleep, often sensitive indicators of CNS side effects (www.ambulatory-monitoring.com/references.html).

### 4.2 Inflammation and Immune Activation or Suppression

While cytokine release may be localized both in tissue distribution and in time, secondary measurements can reflect cytokine release or resulting immune cell activation at time points with other scheduled safety labs. One of the most useful, with a long history of clinical use, is C-reactive protein (CRP), for which EIA kits are available for species including rodents, rabbits, and dogs. CRP is a serum acute-phase reactant, and levels can increase by over 2 logs driven mainly by release of IL-1, IL-6, and TNF-alpha. The peak increase usually lags cytokine release by 24 to 48 hours and can remain above baseline for several days to a week once the acute phase response is stimulated.

Neopterin is a metabolite of guanosine triphosphate that is produced and secreted by mainly by monocyte/macrophage/dendritic cells. Secretion is greatly upregulated when these cells are activated by cytokines, principally gamma-interferon, and serum blood levels are elevated with infections due to virus or intracellular pathogens, organ rejection, and other conditions associated with activation of mononuclear phagocytes. EIA methods to measure neopterin are readily available. The analyte is conserved across humans and all common toxicology species. Both CRP and neopterin can be more sensitive than clinical signs of immune activation (fever, weight loss, hypoactivity) in both NHP models and human studies. If CRP or neopterin prove to be dose-

dependent biomarkers for an on-target pharmacodynamic effect, they can easily be incorporated as analytes into early-phase human trials. An example of the use of these markers in a FIH healthy volunteer single-dose safety and PK study is given in Table 1.

These biomarkers are most useful when they can be carried forward into phase 1 studies of novel products, where, particularly in healthy volunteer studies, a lack of therapeutic endpoints or surrogates (e.g., as viral load in HVB patients), makes pharmacodynamic information difficult to obtain. The first-in-human study results given in Table 1 demonstrate the utility of two biomarkers, CRP and neopterin, when carried forward into a first-in-human study of a novel toll-like receptor ligand in healthy volunteers. These biomarkers, while nonspecific, were a more sensitive indicator of biologic activity than other classical endpoints such as adverse events or routine safety laboratory monitoring. When biomarkers are incorporated into preclinical studies, as dose-ranging pharmacology studies, GLP toxicology studies, or "phase 0" preclinical studies, they can often provide robust information in the early phase of clinical development at little additional cost. CRP can be measured with commercially available kits in virtually any toxicology species.

Surface activation markers on circulating monocytes and T lymphocytes are also sensitive markers of immune activation secondary to either biopharmaceuticals or pathologic disease states. For lymphocytes, CD69, CD25, HLADR, and transferin receptor represent early-, late-, and proliferative-stage markers of activation. Lymphocyte CD69 can be upregulated as early as 4 to 8 hours after ex vivo stimulation of peripheral blood. Monocyte early activation also results in increased CD69 expression, as well as increased expression of CD14 receptors, CD11b, and HLADR, depending on the timing and stimulus. Human flow cytometry reagents can be used for blood samples from nonhuman primates, and reagents are also available for mouse and, to some extent, for other toxicology species. When immune cells are direct targets of immunostimulatory biopharmaceuticals, these activation markers have great utility (along with measurements of cytokine release) for in vitro studies to compare dose–response relationships between human cells and cells from toxicology species. When immune cells are targets of intentionally or potentially immunosuppressive therapeutics, these markers can also be used after stimulation and activation of cells to examine the dose–response relationships involved in inhibition of activation responses. An excellent example is the in vitro stimulation of monocytes with bacterial lipopolysacharide, using supernatant TNF-alpha accumulation as an endpoint. For development of anti-inflammatory PDE4 inhibitors, a lipid-A antagonist, and a p38 MAP kinase inhibitor, among others, this system was used not only for in vitro dose–response information but carried forward into phase 1 healthy volunteer studies as a pharmacodynamic marker (using ex vivo LPS stimulation of whole blood with TNF measurements). This allowed PK/PD correlations in the first-in-human trials [22–24].

**TABLE 1 A first-in-human single-dose, dose-escalation study of biologic X, a putative TLR ligand and immune-response modifier in healthy volunteers**

| Subject Number/ Sex (Ranked by $C_{max}$) | $C_{max}$, (ng/ml) | Adverse Events, 1–5 Days Post Dose | Elevated CRP Post Dose (ULN 3 mg/L)[a] | Elevated Neopterin (ULN 10 nM/L)[a] |
|---|---|---|---|---|
| 103/M | 750 | Back pain, mild | | |
| 107/F | 756 | | | |
| 106/M | 1,253 | | | |
| 116/M | 1,344 | | 17.0/16.3 | 16/9.8 |
| 104/F | 1,657 | | | |
| 103/M | 1,659 | | | |
| 105/M | 1,677 | | | |
| 110/M | 1,783 | | | |
| 111/M | 1,976 | | | |
| 108/F | 2,008 | | | |
| 114/M | 2,359 | | | |
| 125/F | 2,667 | Fever (38.0), mild nausea, Mild myalgia | 9.2/8.4 | 28.2/21.4 |
| 120/M | 2,704 | | | |
| 117/M | 2,794 | | | |
| 124/M | 2,951 | | 3.8/2.8 | |
| 119/M | 3,872 | | 26.2/24.8 | 10.9/5.8 |
| 126/F | 4,689 | Mild somnolence, headache, anorexia | | |
| 127/M | 5,318 | | | |
| 128/M | 5,413 | | 4.9/2.6 | 11.6/7.6 |
| 130/M | 6,537 | | 19.7/19.5 | 47.2/43.3 |
| 132/M | 6,771 | | | |
| 133/F | 8,992 | Fever 37.8, chills, fatigue, chest discomfort, myalgia | 12.0/11.6 | 17.4/11.9 |
| 135/F | 9,107 | Fever, myalgia | 4.2/3.8 | 19.4/12.8 |
| 134/F | 9,444 | | | |
| 115/M | 15,108 | Fever (38.8) × 12 h, myalgia, chills | 34.0/33.5 | 28.9/21.9 |

*Note*: Dose cohorts of placebo (not shown) and 20, 50, 100, and 200 units of biopharmaceutical X, a putative TLR ligand and immunomodulator, dosed in healthy volunteers. Elevations of C-reactive protein and neopterin 48 hrs after dosing correlate with exposure and with dose emergent AEs. These nonspecific biomarkers appear more sensitive than symptoms or fever to detect biologic activity or reactogenicity.

[a]Value/change from baseline.

### 4.3 Cell Proliferation or Apoptosis in Tissue Samples

If segments of tissues harvested for routine histology from toxicology studies are snap frozen, a variety of informative endpoints can be examined in specific tissues. After pulsing of animals with the nucleotide analogue bromodeoxyuridine (BrdU), proliferative indexes can be calculated with fluorescent anti-BrdU antibodies, or by using antibodies to proliferating cell nuclear antigen (PCNA)-cyclin. Indexes of apoptotic cells can be calculated either by TUNNEL assay, UTP-nick translation, or fluorescent Annexin V binding to detect exposed phosphotidyl serine on cell membranes. Newer reagents such as fluorine-containing annexin analogues will allow in vivo imaging of apoptosis by micro PET techniques [25–27].

### 4.4 End-organ Toxicities

Serum creatinine and BUN, the most common indicators of renal function used in both clinical and preclinical safety laboratory panels, are relatively insensitive markers of injury, particularly for the renal tubules. Urinary measurements of alanine aminopeptidase and *N*-acetyl-beta-D-glucosaminidase and kidney injury molecule-1 (KIM-1) can provide much more sensitivity when nephrotoxicity is a potential safety concern [28,29]. These are also suitable for safety monitoring in early-phase human trials if preclinical studies validate such use to monitor product nephrotoxicity.

In a similar manner, troponin-T can greatly increase the sensitivity of safety labs to detect subclinical cardiac myotoxicity. Rat, dog, and pig troponin-T can all be determined using the automated Roche clinical diagnostic system for human Troponin-T [30]. This would have been of obvious relevance to the safety pharmacology of Herceptin®, where the association with cardiomyopathy was only established after widespread clinical use.

Likewise gene expression analysis utilizing DNA microarrays has established the new field of toxicogenomics, providing preclinical research with new predictive tools for organ toxicicity, including hepatotoxicity [31]. Of note, inflammation and immune activation has been shown to greatly potentiate the hepatotoxicity of certain small molecules which demonstrated sporadic hepatotoxicity in the postmarketing period, such as ranitidine. This may have implications for potential hepatotoxic interactions between small molecules and certain immunomodulatory biopharmaceuticals [32].

## 5 AN INTEGRATED APPROACH TO PRECLINICAL AND CLINICAL IMMUNOTOXICITY ASSESSMENTS

Immune system receptors and soluble signal proteins represent targets of a large group of biopharmaceuticals. *This means that immunotoxicity, defined as the unintended suppression or enhancement of immune function, is an impor-*

*tant area of safety pharmacology for many biopharmaceuticals.* It is complicated by the fact, which will be discussed below, that for a variety of reasons nonprimate species are not always suitable for safety pharmacology, and even nonhuman primates such as macaque monkeys and chimpanzees may have significant differences in their immune physiology and pharmacologic responses compared to humans.

### 5.1 Regulatory Guidances

The ICH S8 document, approved in 2005, provides general guidance on immunotoxicity studies for human pharmaceuticals, although the emphasis is on evaluation of small molecules for immunosuppressive effects. ICH S8 suggests that standard toxicity studies (STS) be reviewed for any evidence of immunotoxicity. Of special note would be findings from hematology, albumin/globulin ratios from clinical chemistry, weights of thymus and spleen, and histology of lymphoid organs (which may include Peyers patch, BALT, and bone marrow, when appropriate). The standard study design recommended for immunotoxicity assessments is a 28-day exposure in rodents, with a high dose that "should be above the NOAEL but below a level inducing changes secondary to stress (mediated by increased corticosteroid release)." Careful attention should be paid to the histopathology methods, as reviewed by the European Society of Toxicological Pathology in a 2005 publication [33].

The selection of immunotoxicity study endpoints should be based on any cause for concern raised by a "weight of evidence" review of various factors: for example, findings from STS, pharmacological properties of the drug, intended patient population, structural similarities to known immunomodulators, drug disposition, and/or clinical information. This S8 guideline allows for more flexible approaches; there is no given set of rules. Goals of more detailed immunotoxicology studies should be to characterize a relationship between exposure and altered immune endpoints and to sufficiently characterize the immunotoxicity profile to allow monitoring and management of such risks in human trials.

Additional possible preclinical immunotoxicity endpoints mentioned in the S8 guidance are as follows:

- The T cell dependent antibody response to immunization with SRBC or KLH.
- Immunophenotyping either by immunohistochemistry in lymphoid tissues or flow cytometry in blood samples to enumerate changes in specific leukocyte populations.
- NK cell activity assays.
- Host resistance studies, usually using murine challenge with pathogens such as influenza virus, CMV, *Listeria*, streptococcus, or tumor host resistance models with subcutaneous (sc). injection of tumor cell lines.

- In vitro or ex vivo macrophage/neutrophil functional assays including ADCC, phagocytosis, oxidative burst, and chemotaxis.
- Assays to measure cell-mediated immunity, with either in vivo sensitization and intradermal challenge (delayed-type hypersensitivity) or in vitro measurements of T cell responses to antigens or mitogens.

It should be noted that these recommended studies are all designed to detect suppression of normal immune responses, not inappropriate activation or dysregulation occurring in any compartment of the immune system. This represents one of the serious limitations of the S8 guidance. The suggested test systems also give little consideration to the difficulty of selecting a relevant test species for biopharmaceuticals that may be highly specific for a primate or human target.

The S8 guidance suggests that more comprehensive immunotoxicity studies (beyond the STS) should be completed "before the exposure of a large population of patients," which is usually phase 3. However, this suggestion is often not applicable to biopharmaceuticals. For many biopharmaceuticals the initial clinical development decision is whether a single-dose safety and pharmacokinetic study can be safely accomplished in healthy volunteers. Data are needed to guide dose selection and protocol design for these studies, or to determine whether the risk profile of the product requires first-in-human PK studies to be conducted in patients, who potentially can realize some current or future benefit. Such initial clinical pharmacology patient studies are frequently difficult to recruit, given that single-dose or short-term exposure to drug is not expected to result in near-term therapeutic benefit, only in safety and pharmacokinetic data that may enable future efficacy studies. Patients prefer participation in phase 2 or 3 trials offering more immediate potential benefit and provision of longer term medical care for their illness. Additionally initial PK and safety trials in patients may provide a less sensitive and accurate safety signal, complicated by variable comorbidities of concurrent illness and the frequent simultaneous treatment of disease with other therapeutics.

For these reasons, determining the nature, predictability, and reversibility of immune system effects become a critical factor in weighing the risk–benefit ration of an phase 1 trial in healthy volunteers, as well as the determining a safe starting dose, safe maximum dose, and methods for monitoring potential immunotoxicity in these subjects. Over the past decade single-dose PK and safety trials of biopharmaceuticals in healthy volunteers have often preceded initial patient studies, even for products with known immunomodulatory effects at the anticipated therapeutic chronic dose exposure.

While dozens of such healthy volunteer studies have been safely conducted, the catastrophic syndrome of T cell activation and depletion, cytokine release, and multi-organ failure seen in the 2006 Tegenero 1412 phase I study [34] reminded clinical investigators, sponsors, and regulatory agencies of the very

real risks to healthy volunteers posed by inadequately and incompletely characterized biopharmaceuticals. Thus for many biopharmaceuticals, the preclinical investigations of potential immunotoxicity will occur prior to the first human trial, to inform the decision of whether the risk profile is suitable for a healthy volunteer trial. This is also becoming the case for many small-molecule noncytotoxic therapies for cancer. While traditional cytotoxic and potentially mutagenic cancer therapies were first studied clinically in cancer patients with progressive disease who were refractory to standard therapies, and doses then escalated to the maximum tolerated (and highly toxic) dose, it is now not uncommon for the PK, tolerability, and safety of a noncytotoxic cancer drug to first be investigated in healthy volunteers. This was the case for the oral kinase inhibitors gefitinib and erlotinib, where healthy volunteers were also used to characterize drug–drug interactions and it is becoming more common with biopharmaceuticals as well [35–37]. In such cases, just as with biopharmaceuticals, careful preclinical investigation of potential immunotoxicity is required.

### 5.2 New Techniques for Evaluating Immunosuppressive Activity

The assessment of immunosuppressive drug effects by ex vivo measurement of stimulated lymphocyte function in peripheral blood is becoming more informative. Ex vivo assays have been used to characterize the in vitro and in vivo interactions of immunosuppressive drugs such as cyclosporine, tacrolimus, mycophenolate mofetil, and FK778. These assays can be utilized to characterize pharmacodynamics in preclinical species, including the rat, carried forward into clinical pharmacology studies, and even utilized, in some cases, for therapeutic drug monitoring in patients. While often applied to small-molecule immunomodulators, some of these techniques are also applicable to biopharmaceuticals, preclinically and in phase 1 trials, especially for studying their interactions with approved small-molecule immunomodulators [38–40].

New techniques such as the Cylex photoluminescent detection of CD3 lymphocyte activation by mitogen or antigen, approved by the FDA for monitoring immunosuppression in transplant patients, are convenient, species nonspecific, and well suited for repeated measurements in preclinical and clinical studies of biopharmaceuticals [41]. This assay utilizes whole blood.

## 6 NOVEL PHARMACODYNAMIC MEASURES USING FLOW CYTOMETRY AND BLOOD MONONUCLEAR CELLS AS A SURROGATE FOR TISSUE RESPONSES

Over the past several years new techniques have vastly expanded the potential of flow cytometry to provide pharmacodynamic endpoints for a variety of

pharmacotherapies, in a manner noninvasive enough to be carried forward from preclinical exposure–response studies into early-phase clinical pharmacology studies.

Intracellular signaling can now be visualized in peripheral blood lymphocytes with phosphospecific antibodies to MEK, ERK, p38 MAPK, STAT, JNK/SAPK, and other signal transducers. This allows pharmacodynamic monitoring of signal transduction agonists or inhibitors (especially when coupled with ex vivo T cell activation by phorbal esters or other agents, which can be performed in whole blood) [42–46].

Another interesting example is the ability of flow cytometric techniques to monitor histone acetylation levels of peripheral blood cells, and provide pharmacodynamic analysis of histone deacetylase inhibitors being utilized in oncology patient trials [47,48].

Currently flow-cytometric methods are also under development to examine effects of various inhibitors of PARP (poly(ADP-ribose)polymerase) that are in preclinical development or clinical trials and use native or stimulated PARP activity in peripheral blood mononuclear cells.

## 7 LOW PREDICTIVE VALUE OF CERTAIN TOXICITY STUDIES IN NONHUMAN PRIMATES

For biopharmaceuticals that are agonists at activation receptors on human immune cells (e.g., anti-CD3 or anti-CD4) or antagonists at inhibitory receptors, even safety studies in nonhuman primates may not accurately predict a product's pharmacodynamic effects and the exposure–response relationship in humans. For example, no clinical evidence of autoimmunity was seen in cynomolgus monkey studies of CTLA-4 blockade [49]. However, autoimmune phenomena, including colitis, dermatitis, vitiligo, panhypopituitarism, and thyroiditis, were among dose-limiting toxicities in trials of this product in melanoma patients [50]. In general, trials of various OKT3 or anti-CD3 antibodies in nonhuman primates have also underestimated the immune activation and cytokine-release syndromes seen in first-in-human trials.

The July 2006 report of the Joint ABPI/BIA Early Stage Clinical Trial Taskforce [51] provides useful guidance on assessing the adequacy of preclinical safety characterization and the safety of proposed human starting doses for novel biopharmaceuticals. The report states:

> Since toxicity usually arises from exaggerated pharmacological effects, sometimes combined with a narrow therapeutic range and/or a steep dose–response relationship from the NOAEL to toxicity, the pharmacologically active dose is often a more sensitive indicator of potential toxicity for biotechnology-derived pharmaceuticals (compared to the preclinical NOAEL adjusted by a safety

> margin). The safety assessment should take due account of the dose–, concentration– and time–response relationships, receptor occupancy, relative potency in animals versus humans and, unless otherwise justified, the starting dose in the first-in-human study should be below the MABEL as predicted from all available preclinical data (i.e., in vitro data in humans and animals and in vivo data in animals).

The MABEL is the "minimum anticipated biologically effective level." In the case of the Tegenero monoclonal anti-CD28 "superagonist" antibody, in vitro incubation of the product with human T cells caused release of large amounts of TNF-alpha, gamma-interferon, and IL-2. Parallel in vitro studies with cynomolgus monkey T cells were not performed. These studies would likely have demonstrated a different pharmacodynamic profile in the safety species. In vivo studies in cynomolgus monkeys showed very limited release of cytokines after TGN1412 dosing [52]. This was used to justify a healthy volunteer FIH trial dose using 1/10 the NOAEL (5 mg/kg) as a maximum dose, and 1/500 of the claimed NOAEL as the starting dose (0.1 mg/kg) Despite this large safety margin versus the monkey NOAEL, life-threatening toxicities due to cytokine-release syndrome and profound T cell depletion were seen in the first human cohort that was dosed [53].

The ABPI/BIA taskforce recommendation of basing the first human dose on the MABEL, rather than solely on the NOAEL in a toxicology species, would have led to a lower and less hazardous human starting dose. These recommendations should guide design of future preclinical safety programs and future FIH clinical protocols, particularly for products with novel targets and mechanisms and pharmacodynamics and with restricted species specificity.

In a preclinical study of the immunosuppressive drug FK778, suppression of lymphocyte proliferation was studied with whole blood from four cynomolgus monkeys and four healthy human volunteers. This study showed the IC50 for FK778 to be lower in human than in monkey blood [54]. These and similar data were used to design the FIH trials, and this example provided guidance for extrapolations from in vivo preclinical safety studies.

With regards to glycosylated protein biopharmaceuticals, including monoclonal antibodies, it may be important to consider another distinction between protein therapeutics made in human and nonhuman mammalian cells, and between the human and nonhuman primate immune systems. One significant difference between humans and the Macaque monkeys typically used in safety studies, and between plasma-derived human immunoglobulins and recombinant antibodies produced in nonhuman mammalian cell lines, is the pattern of sialic acid usage. Most mammalian glycoproteins, including antibodies, display two major sialic acids, *N*-acetylneuraminic acid (Neu5Ac) and *N*-glycolylneuraminic acid (Neu5Gc). However, humans lack Neu5Gc on glycoproteins, because of a mutation in CMP-Neu5Ac hydroxylase, which occurred after evolutionary divergence from great apes [55]. Indeed this difference

accounts for the resistance of monkeys and chimpanzees to human malarial infections [56].

SIGLEC family receptors, which recognize these sialic acids, serve to down-regulate macrophage activation induced via FcR interactions, as well as TCR-mediated lymphocyte activation, by signaling through intracellular tyrosine-based inhibitory motifs (ITIMs). Antibodies produced in CHO cells may have sialic acid residues that are not recognized efficiently by human SIGLECs, and thus may be poor ligands for these important inhibitory regulatory receptors [57]. The pattern of glycosylation and sialic acid usage, particularly in the Fc region, strongly affect the effector functions of antibodies [58]. For example, nonfucosylated MAbs produced in mutated CHO cells are far more potent inducers of ADDC in human cell systems than their homologues, which are highly fucosylated in normal CHO cells [59].

Additionally a recent article by Nguyyen et al. reports that human T cells give a much stronger proliferative response than chimpanzee T cells when activated via the TCR, although responses are similar after stimulation by the mitogen, PHA. This is correlated with the evolutionary loss in humans of the inhibitory Siglec-5 receptor on lymphocytes. Transfection of human T cells with chimpanzee Siglec-5 eliminated this increased responsiveness to TCR activation [60]. These data should help refute those that argue that more predictive preclinical safety studies of human biopharmaceuticals might, in the future, be performed using chimpanzees instead of monkeys. Indeed in terms of sialic acid and SIGLEC receptor usage, new world Aotus monkeys (unlike the chimpanzee) may more closely model human pharmacologic responses to immunomodulators [61].

However, nonhuman primates do overcome one significant limitation of nonprimate safety species, which is the lack of gamma-delta T cell populations. Part of the innate immune surveillance system, gamma-delta T cells are critical for clearance of intracellular bacterial pathogens, and they respond to low molecular weight nonpeptidic phospho-antigens in an MHC-unrestricted fashion. These cells express CD28, and other co-stimulatory receptors and can rapidly produce large amounts of TNF-alpha and gamma-interferon when activated. They are well represented in the T cell repertoire of nonhuman primates, and they appear to respond similarly to phosphoantigens [62]. Nevertheless, it is unknown whether differential expression of inhibitory SIGLEC receptors in monkeys compared to humans may alter gamma-delta T cell reactivity to potential biopharmaceuticals, which could target or stimulate these cells.

## 8 SUMMARY

The burden is on the clinical investigator, the biopharmaceutical sponsor, and the regulatory reviewer to be certain that the risk of exaggerated pharmaco-

dynamic effects, off-target toxicities, and other harmful outcomes in first-in-human trials are reduced to the maximum extent possible, especially for healthy volunteer studies, which are now increasingly common for biopharmaceuticals. Novel therapeutics require a careful comparison of the pharmacodynamic effects in human cells and tissues and those of the safety model species, thoughtful and individualized design of safety pharmacology and immunotoxicity studies, and careful attention to selection of a safe starting dose and to pharmacodynamic monitoring in clinical protocols. Clinical investigators and institutional review boards should examine data carefully with this mandate in mind, and not rely exclusively on conclusions of the regulatory review process for first-in-human trials.

## REFERENCES


1. ICH Guidance S7A.
2. ICH Guidance S6. Preclinical Safety Evaluation of Biotechnology-Derived Pharmaceuticals.
3. ICH Guidance S8. Immunotoxicity Studies for Human Pharmaceuticals.
4. Clarke J, Leach W, Pippig S, et al. Evaluation of a surrogate antibody for preclinical safety testing of an anti CD-11a monoclonal antibody. *Regul Toxicol Pharmacol* 2004;40(3):219–26.
5. Leung R, Whittaker P. RNA interference: from gene silencing to gene-specific therapeutics. *Pharmacol Ther*. 2005;107(2):222–39.
6. Dufour JH, Dziejman M, Liu MT, et al. IFN-γ-inducible protein 10 (IP-10;CXCL10)-deficient mice reveal a role for IP-10 in effector T cell generation and trafficking. *J Immunol* 2002;168:3195–3204.
7. Huang X, Wu J, Zhu W, Pytela R, Sheppard D. Expression of the human integrin beta6 subunit in alveolar type II cells and bronchiolar epithelial cells reverses lung inflammation in beta6 knockout mice. *Am J Respir Cell Mol Biol* 1998;19(4): 636–42.
8. Sheppard D. Integrin-mediated activation of transforming growth factor-beta(1) in pulmonary fibrosis. *Chest*. 2001;120(1 Suppl):49S–53S.
9. Herzyk D, Gore E, Polsky R, et al. Immunomodulatory effects of anti-CD4 antibody in host resistance against infections and tumors in human CD4 transgenic mice. *Infect Immun* 2001;69(2):1032–43.
10. Green J, Black, L. Overview of preclinical safety assessment for immunomodulatory biopharmaceuticals. *Hum Exp Toxicol* 200 19(4):208–12.
11. Hart TK, Blackburn MN, Brigham-Burke M, et al. Preclinical efficacy and safety of pascolizumab (SB 240683): a humanized anti-interleukin-4 antibody with therapeutic potential in asthma. *Clin Exp Immunol* 2002;130(1):93–100.
12. Hart TK, Cook RM, Zia-Amirhosseini P, et al. Preclinical efficacy and safety of mepolizumab (SB-240563), a humanized monoclonal antibody to IL-5, in cynomolgus monkeys. *J Allergy Clin Immunol* 2001;108(2):250–7.

13. *www.mdsps.com/over/MarketingMaterials/SafetyPharmacologyBrochure.pdf*.
14. Meijer R, Koopmans R, ten Berge I, Schellekens P. Pharmacokinetics of murine anti-human CD3 antibodies in man are determined by the disappearance of target antigen. *J Pharmcol Exp Ther* 2002;300(1):346–53.
15. Lawrence C, Pollard C, Hammond T, and Valentin J. Nonclinical proarrythmia models: predicting torsades de pointes. *J Pharmacol Toxicol Meth* 2005;52(1): 46–56.
16. Finlayson K, Witchel H, McCulloch J, and Sharkey J. Acquired QT interval prolongation and HERG: implications for drug discovery and development. *Eur J Pharmacol* 2004;500(1–3):129–42.
17. Redfern W, Carlsson L, Davis A, et al. Relationships between preclinical cardiac electrophysiology, clinical QT interval prlongation, and torsades de points for a broad range of drug: evidence for a provisional safety margin in drug development. *Cardiovasc Res* 2003;58(1):32–45.
18. Kupferschmidt H, Meier CH, Sulzer M, et al. Clinico-pharmacological case (2). Bradycardia and ventricular tachycardia of the torsades de pointes type as a side effect of vasopressin: 3 case reports. *Schweiz Rundsch Med Prax* 1996;85(11): 340–3.
19. Sartiani L, De Paoli P, Lonardo G, et al. Does recombinant human interleukin-11 exert direct electrophysiologic effects on. single human atrial myocytes? *J Cardiovasc Pharmacol* 2002;39(3):425–34.
20. Schneider JW, Chang AY, Garratt A. Trastuzumab cardiotoxicity: speculations regarding pathophysiology and targets. for further study. *Semin Oncol* 2002;29(3 Suppl 11):22–8.
21. Eckart RE, Love SS, Atwood JE, et al. Incidence and follow-up of inflammatory cardiac complications after smallpox vaccination. *J Am Coll Cardiol* 2004;44(1): 201–5.
22. Parasrampuria D, de Boer P, Desai-Krietger D, et al. Single-dose pharmacokinetics and pharmacodynamics of RWJ 67657, a specific p28 MAP kinase inhbitor: a first-in-human study, *J Clin Pharmacol* 2003;43(4):406–13.
23. Wong Y, Rossignol D, Rose, J, et al. Safety, pharmacokinetics, and pharmacodynamics of E5564, a lipid A antagonist, during an ascending single-dose clinical study. *J Clin Pharmacol* 2003;34(7):745–42.
24. Timmer W, Leclerc V, Birraux G, et al. The new PDE4 inhibitor roflumilast is efficacious in exercise-induced asthma and leads to suppression of LPS-stimulated TNF-alpha ex vivo. *J Clin Pharmacol* 2002;42(3):297–303.
25. Valenti MT, Azzarello G, Vinante O, et al. Differentiation, proliferation and apoptosis levels in human leiomyoma and leiomyosarcoma. *J Cancer Res Clin Oncol* 1998;124(2):93–105. *J Immunol Meth* 1995;184(1):39–51.
26. Vermes I, Haanen C, Steffens-Nakken H, et al. A novel assay for apoptosis: flow cytometric detection of phosphatidylserine expression on early apoptotic cells using fluorescein labelled Annexin V. *J Immunol Meth* 1995;184(1):39–51.
27. Damianovich M, Ziv I, Heyman SN, Rosen S, et al. ApoSense: a novel technology for functional molecular imaging of cell death inmodels of acute renal tubular necrosis. *Eur J Nucl Med Mol Imaging* 2006;33(3):281–91.

28. Verplanke AJ, Herber RF, de Wit R. Comparison of renal function parameters in the assessment Of cis-platin induced nephrotoxicity. *Nephron* 1994;66(3):267–72.
29. Ichimura T, Hung CC, Yang SA, et al. Kidney injury molecule-1: a tissue and urinary biomarker for nephrotoxicant-induced renal injury. *Am J Renal Physiol* 2004;286(3): F552–63.
30. Herman EH, Zhang J, Lipshultz SE, et al. Correlation between serum levels of cardiac troponin-T and the severity of the chronic cardiomyopathy induced by doxorubicin. *J Clin Oncol* 1999;17(7):2237–43.
31. Fielden MR, Pearson C, Brennan R, Kolaja KL. Preclinical drug safety analysis by chemogenomic profiling in the liver. *J Pharmacogenomics* 2005;5(3):161–71.
32. Waring JF, Liguori MJ, Luyendyk JP, Microarray analysis of lipopolysaccharide potentiation of trovafloxacin-induced liver injury in rats suggests a role for proinflammatory chemokines and neutrophils. *J Pharmacol Exp Ther* 2006;316(3): 1080–7.
33. Ruehl-Fehlert C, Bradley A, George C, et al. Harmonization of immunotoxicity guidelines in the ICH process—pathology considerations from the guideline committee of the European Society of Toxicological Pathology (ESTP). *Exp Toxicol Pathol* 2005;57(1):1–5.
34. Suntharalingam G, Perry MR, Ward S, et al. Cytokine storm in a phase 1 trial of the anti-CD28 monoclonal antibody TGN1412. *N Engl J Med* 2006;355(10): 1018–28.
35. Swaisland HC, Smith RP, Laight A, et al. Single-dose clinical pharmacokinetic studies of gefitinib. *Clin Pharmacokinet* 2005;44(11):1165–77.
36. Swaisland HC, Ranson M, Smith RP, et al. Pharmacokinetic drug interactions of gefitinib with rifampicin, itraconazole and metoprolol. *Clin Pharmacokinet* 2005;44(10):1067–81.
37. Frohna P, Lu J, Eppler S, et al. Evaluation of the absolute oral bioavailability and bioequivalence oferlotinib, an inhibitor of the epidermal growth factor receptor tyrosine kinase, in a randomized, crossover study in healthy subjects. *J Clin Pharmacol* 2006;46(3):282–90.
38. Barten MJ, Rahmel A, Garbade J, et al. C0h/C2h monitoring of the pharmacodynamics of cyclosporin plus mycophenolate mofetil in human heart transplant recipients. *Transplant Proc* 2005;37(2):1360–1.
39. Barten MJ, Streit F, Boeger M, et al. Synergistic effects of sirolimus with cyclosporine and tacrolimus: analysis of immunosuppression on lymphocyte proliferation and activation in rat whole blood. *Transplantation* 2004;77(8):1154–62.
40. Barten MJ, Dhein S, Chang H, et al. Assessment of immunosuppressive drug interactions: inhibition of lymphocyte function in peripheral human blood. *J Immunol Meth* 2003;283(1–2):99–114.
41. Zeevi A, Britz JA, Bentlejewski CA Monitoring immune function during tacrolimus tapering in small bowel transplant recipients. *Transpl Immunol* 2005;15(1): 17–24.
42. Perez O, Krutzik P, Nolan G. Flow cytometric analysis of kinase signaling cascades. *Meth Mol Biol* 2004;263:67–94.

43. Krutzik P, Hale M, Nolan G. Characterization of the murine immunological signaling network with phosphospecific flow cytometry. *J Immunol* 2005;175(4): 2366–73.

44. Krutzik PO, Clutter MR, Nolan GP. Coordinate analysis of murine immune cell surface markers and intracellularphosphoproteins by flow cytometry. *J Immunol* 2005;175(4):2357–65.

45. Grammer A, Fischer R, Lee O, et al. Flow cytometric assessment of the signaling status of human B lymphocytes from normal and autoimmune individuals. *Arthritis Res Ther* 2004;6(1):28–38.

46. Chow S, Hedley D, Grom P, Magari R, Jacobberger JW, Shankey TV. Whole blood fixation and permeabilization protocol with red blood cell lysis for flow cytometry of intracellular phosphorylated epitopes in leukocyte subpopulations. *Cytometry A* 2005;67(1):4–17.

47. Ronzoni S, Faretta M, Ballarini M, et al. New method to detect histone acetylation levels by flow cytometry. *Cytometry* 2005;66(1):52–61.

48. Chung EJ, Lee S, Sausville EA, et al. Histone deacetylase inhibitor pharmacodynamic analysis by multiparameter flowcytometry. *Ann Clin Lab Sci* 2005;35(4): 397–406.

49. Keler T, Halk E, Vitale L, et al. Activity and safety of CTLA-4 blockade combined with vaccines in cynomolgus macaques. *J Immunol* 2003;171(11):6251–9.

50. Ribas A, Camacho LH, Lopez-Berestein, et al. Antitumor activity in melanoma and anti-self responses in a phase I trial with the anti-cytotoxic T lymphocyte-associated antigen 4 monoclonal antibody. *J Clin Oncol* 2005;23(35):8968–77.

51. *http://www.abpi.org.uk/information/pdfs/BIAABPI_taskforce2.pdf*.

52. Tegenero. 1412 Investigators Brochure, 19 Dec 2005.

53. Suntharalingam G, Perry MR, Ward S, et al. Cytokine storm in a phase 1 trial of the anti-CD28 monoclonal antibody TGN1412. *N Engl J Med* 2006;355(10):1018–28.

54. Birsan T, Dambrin C, Klupp J, et al. Effects of FK778 on immune functions in vitro in whole blood from non-human primates and healthy human volunteers. *Transpl Immunol* 2003;11(2)164–7.

55. Gagneux P, Cheriyan M, Hurtado-Ziola N, et al. Human-specific regulation of alpha 2-6-linked sialic acids. *J Biol Chem* 2003;278(48):48245–50.

56. Martin MJ, Rayner JC, Gagneux P, Barnwell JW, Varki A. Evolution of human-chimpanzee differences in malaria susceptibility:relationship to human genetic loss of N-glycolylneuraminic acid. *Proc Natl Acad Sci USA* 2005;102(36):12819–24.

57. Scallon BJ, Tam SH, McCarthy SG, Cai AN, Raju TS. Higher levels of sialylated Fc glycans in immunoglobulin G molecules Can adversely impact functionality. *Mol Immunol* 2007;44(7):1524–34.

58. Hodoniczky J, Zheng YZ, James DC. Control of recombinant monoclonal antibody effector functions by Fc N-glycanremodeling in vitro. *Biotechnol Prog* 2005;21(6): 1644–52.

59. Niwa R, Hatanaka S, Shoji-Hosaka E, et al. Enhancement of the antibody-dependent cellular cytotoxicity of low-fucose IgG1 is independent of FcgammaRIIIa functional polymorphism. *Clin Cancer Res* 2004;10(18 Pt 1):6248–55.

60. Nguyen DH, Hurtado-Ziola N, Gagneux P, Varki A. Loss of Siglec expression on T lymphocytes during human evolution. *Proc Natl Acad Sci USA* 2006;103(20): 7765–70.
61. McConkey EH, Varki A. Genomics. Thoughts on the future of great ape research. *Proc Natl Acad Sci USA* 2005;102(36):12819–24.
62. Cairo C, Propp N, Hebbeler AM, et al. The Vgamma2/Vdelta2 T-cell repertoire in Macaca fascicularis: functional responses to phosphoantigen stimulation by the Vgamma2/Jgamma1.2 subset. *Immunology* 2005;115(2):197–205.

# Genetic Toxicology Testing of Biopharmaceuticals

DAVID JACOBSON-KRAM, PhD, DABT, and HANAN GHANTOUS, PhD, DABT

The potential for carcinogenic risk should be considered for all new pharmaceuticals especially those drugs that are intended for chronic use. In general, results from carcinogenicity testing are only available at the time of submission of a new drug application (NDA) or a biological license application (BLA). As a result hundreds or even thousands of individuals may be exposed during the drug development process without an understanding of potential for carcinogenicity. Phase 1 studies for new drugs are often performed in healthy volunteers for whom there is no obvious benefit. So risks for this population, in particular, should be exceedingly low.

Because carcinogenicity results are not available during the course of clinical trials, the FDA relies on the results of genotoxicity studies as a surrogate for potential cancer risk. The standard battery of assays recommended for small molecules as described in the ICH S2B guidance (A Standard Battery for Genotoxicity Testing of Pharmaceuticals) [1] was, however, not designed to include assessments of biopharmaceuticals. This is because in the absence of specific transport mechanisms, it was considered unlikely that large peptides and protein molecules would enter the cell and interact with cellular DNA or other chromosomal material to pose genotoxic risk. It is possible that large molecules enter a cell in a nonspecific manner, for example, through endocytosis. In fact externally applied restriction endonucleases have been shown to induce chromosomal aberrations in cultured cells [2,3].

In general, genotoxicity standard assays (e.g., bacterial reverse mutation assay [Ames test], in vitro chromosomal aberration assay, mouse lymphoma gene mutation assay, and rodent micronucleus assay) may not be suitable assays because the test cells do not contain the appropriate receptors to transport the product (i.e., not a relevant species) or because the biopharmaceutical

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

has an indirect mechanism that is not expressed in isolated in vitro systems. False positives have also been observed in the standard Ames test because of the presence of growth-promoting constituents in the test sample such as histidine or its precursors. In addition the genetic toxicity assays often rely on cytotoxicity to ensure assay validity, an effect difficult to reach with most biopharmaceuticals. Conducting these assays with large quantities of protein can sometimes confound the interpretation of the results.

General testing for genotoxic impurities may also not be appropriate, since the impurities and contaminants that may be contained in a biopharmaceutical product include residual host cell proteins and nucleic material, fermentation components, manufacturing process components such as column leachables and detergents, bacteria and viral particles, they do not include organic chemicals typically found in small-molecule manufacturing. However, genotoxicity testing of protein- or immunoconjugates (products containing organic chemical linkers) might be appropriate, particularly when a residual organic linker is found in the product because of the instability of the conjugate during storage or upon dilution in serum.

If genotoxicity appears to be a potential issue with a biologic product, it is uncertain whether the standard ICH battery would be adequate to assess this potential hazard. In a publication by Gocke et al. in 1999 [4], genotoxicity data for biologics were obtained from several sponsors. In summary, 4 of 78 biologic compounds assessed for genotoxicity (using a total of 177 assays) demonstrated genotoxic effects. These four were proteins: One was an antibody fragment conjugated to an organic chemical linker, and the positive effect was concluded to be produced by the linker. The second protein was positive in the mouse lymphoma test but not in the Ames test and the micronucleus test. However, no information was provided describing the compound, so no conclusion can be drawn. The third compound was a natural lipase, and the chromosomal aberrations observed with this product are most likely a result of the endonucleases released upon destruction of lysosomal membranes. This suggests that the protein was able to enter the cell in a nonspecific fashion. The fourth protein, glucagon, showed very weak and inconsistent positive signals with both endogenous and recombinant glucagon, and the effect is most likely attributed to its physiological activity.

In addition, and as discussed in the Draft guidance for biopharmaceuticals [5], a positive result in a genotoxicity assay of a biopharmaceutical may actually be the result of an exaggerated pharmacological response and not a direct DNA modification by the product, as is the case for lipase, glucagon, erythropoietin, and DNAse. DNAse was predictably positive in an in vitro clastogenicity assay only following electroporation. Recombinant and native erythropoietin increase the frequency of micronucleated polychromatic erythrocytes in bone marrow of mice but produce negative results in the bacterial reverse mutation tests and do not induce chromosomal aberrations in vitro in Chinese hamster ovary (CHO) cells or human peripheral blood lymphocytes . The induction of micronuclei was shown to be due to the acceler-

ated proliferation and differentiation of erythrocytes and promotion of early release of polychromatic erythrocytes [6,7,8], an exaggerated effect of its mechanism of action. Therefore, in cases where it is deemed necessary to assess the ability of the biopharmaceutical to modify genetic material, careful attention should be given to the study design and to the interpretation of the results.

Twenty-seven out of 44 FDA-approved biopharmaceuticals have been tested in a battery of genotoxicity assays. Eighty-five different assays performed yielded negative results. The most commonly performed assays were the Ames test, the chromosomal aberration assay in human lymphocytes, the mouse lymphoma gene mutation assay, and the mammalian in vivo erythrocyte micronucleus test. Examples of the range of biopharmaceutical products tested include, dornase alfa (deoxyribonuclease I-DNAse), trastuzumab (mAb to human epidermal growth factor receptor 2), alteplase (tissue plasminogen activator), infliximab (mAb to the human tumor necrosis factor α).

In determining whether genotoxicity testing is appropriate and necessary, a variety of factors should be considered (see Figure 1 [5]):

- The properties of the biopharmaceutical.
- Whether the product is produced by fermentation or chemically synthesized.
- Presence of residual organic chemicals or small molecules used in modification of a protein or inclusion of transporter sequences that localize the biopharmaceutical to the nucleus during manufacturing.
- The stability of the product should be considered, whether the product is a protein- or immunoconjugate.
- The presence of a "free" organic linker in the product.
- The pharmacology and mechanism of action of the product.
- If the product alters DNA or RNA integrity, such as DNAses and ribozymes, the feasibility of the product and by-products interacting with nuclear material in intact cells under normal physiological conditions (i.e., in the absence of cell permeability enhancing agents and techniques) should be taken into consideration before performing standard genotoxicity tests.

The standard battery of genotoxicity tests or selected tests might be appropriate to determine the toxicity of conjugates or organic/chemical linkers in the molecule (e.g., Mylotarg). In these cases careful attention should be given to the study design of tests used to assess the potential for gene mutation, chromosome aberrations, and/or DNA damage.

Biopharmaceuticals may be nongenotoxic but have tumorigenic potential. For example, cytokines and growth factors may directly lead to cell proliferation as a function of drug receptor activation or may secondarily modulate growth as a result of production and activation of a complex cascade of cytokines and moieties (e.g., nitrous oxide). While the concerns for these types of

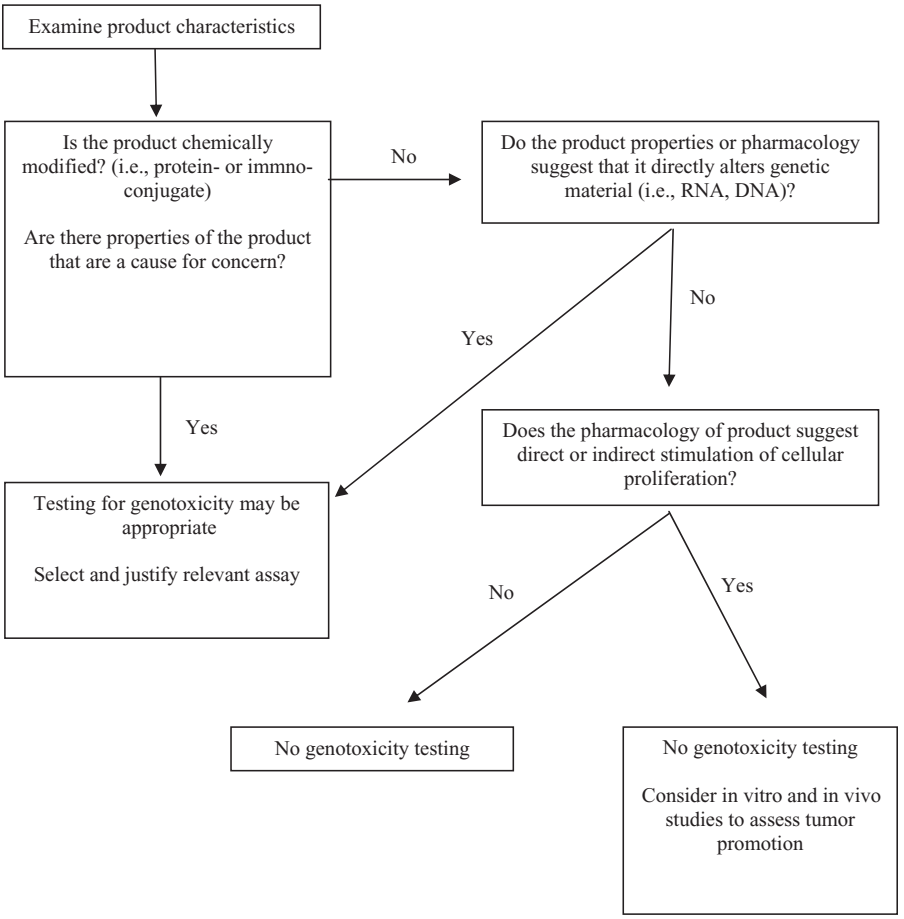


**Figure 1** Genetic toxicity decision tree for biopharmaceuticals.

products in promoting tumor growth are valid, they cannot be addressed with genotoxicity assays. Appropriate in vitro and in vivo assays should, however, be considered (e.g., assessing the ability of the growth factor to stimulate tumor growth or alter the sensitivity of tumor cells to chemotherapy using in vitro and/or in vivo tumor xenograft models).

## REFERENCES

1. ICH S2B. A Standard Battery for Genotoxicity Testing of Pharmaceuticals, 1997.
2. Obe G, Jonas R, Schmidt S. The restriction endonuclease Alu I induces chromosomal aberrations in human peripheral blood lymphocytes in vitro. *Mutat Res* 1986;163:271–5.

3. Vasudev V, Obe G. Evidence for a recptor-mediated endocytosis of Alu I in Chinese hamster ovary cells. *Mutat Res* 1988;197:109–16.
4. Gocke E, Albertini S, Brendel-Schwaab S, Muller L, Suter W, et al. Genotoxicity testing of biotechnology-derived products. Report of a GUM task force. *Mutat Res* 1999;436:137–56.
5. FDA Draft Guidance for Industry, Nonclinical Safety Evaluation of Biotechnology-Derived Pharmaceuticals (in preparation).
6. Yajima N, Kurata Y, Imai E, Sawai T, Takeshita Y. Genotoxicty of genetic recombinant erythropoietin in a novel test system. *Mutagenesis* 1993;8:231–6.
7. Yajima N, Kurata Y, Sawai T, Takeshita Y. Comparative studies in induction of micronuclei by three genetically recombinant and urinary human erythropoietins. *Mutagenesis* 2003;8:237–41.
8. Yajima N, Kurata Y, Sawai T, Takeshita Y. Induction of micronucleated erythrocytes by recombinant human erythropoietin. *Mutagenesis* 2003;8:221–9.

# General Toxicity Testing and Immunotoxicity Testing for Biopharmaceuticals

JEANINE L. BUSSIERE, PhD, DABT

**Contents**

1 Introduction
  1.1 Molecule Is Pharmacologically Active in Rodents and Nonrodents
  1.2 Molecule Is Pharmacologically Active in Nonrodents Only
  1.3 Molecule Is Pharmacologically Active in Humans and Chimpanzees Only
2 Immunotoxicity Testing
3 Conclusions
  References

## 1 INTRODUCTION

For general toxicity testing of biopharmaceuticals, the first step is identifying a "relevant" species (see Chapters Selection of Relevant Species and The Role of Pharmacokinetics and Pharmacodynamics in Selecting a Relevant Species). The choice of relevant species will determine the strategy for testing the general toxicity of the molecule. If the molecule is pharmacologically active in both rodent and nonrodent species, then the toxicology studies should be conducted in two species [1]. Often biopharmaceuticals are more restricted in their species cross-reactivity, and may only be pharmacologically active in nonhuman primates. In this case general toxicity testing in one species is sufficient, but the challenge is to conduct specialty studies in this relevant species such as repro-

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

ductive toxicity, immunotoxicity, or safety pharmacology. If the biopharmaceutical is only active in chimpanzees, then alternate strategies for assessing safety need to be considered. The strategies could include use of a surrogate molecule (i.e., the homologous protein in the animal species, or a monoclonal antibody that cross-reacts with the animal receptor/ligand) or use of a transgenic or knockout mouse. These various options will be discussed in this chapter.

### 1.1 Molecule Is Pharmacologically Active in Rodents and Nonrodents

If the biopharmaceutical is pharmacologically active in rodents and nonrodents, a more standard toxicology program can be used. In general, repeat-dose studies in both rodents and nonrodents are conducted to support first-in-human (FIH) trials, along with an assessment of safety pharmacology. The duration of repeat-dose studies is similar to those studies supporting the development of small-molecule pharmaceuticals and should follow the ICH M3 guidance [2]. In addition, if the biopharmaceutical is a monoclonal antibody, then an in vitro tissue cross-reactivity study is conducted with a panel of human tissues as well as tissues from the species used in the toxicology studies (discussed in Chapter Tissue Cross-Reactivity Studies for Monoclonal Antibodies: Predictive Value and Use for Selection of Relevant Animal Species for Toxicity Testing). Ideally the route and frequency of administration should mimic the clinical regimen as closely as possible. Cardiovascular safety pharmacology can be assessed in a stand-alone study in nonrodents, or these endpoints can be incorporated into the repeat-dose study. Respiratory and CNS activity can be assessed in separate studies in rats or incorporated into the repeat-dose study. Between the FIH study and the biological license application (BLA), additional toxicology studies will include subchronic and chronic studies (the duration of these studies are outlined in the ICH M3 and S6 guidances). Typically a duration of six months is acceptable for the chronic study [3]. In addition reproductive toxicology studies are conducted in rats (fertility, embryo-fetal development, peri/postnatal development) and rabbits (embryo-fetal development). For certain indications (i.e., oncology) reproductive testing in one species is generally sufficient.

Immunogenicity is an important issue to consider for biopharmaceuticals as it can limit toxicity testing. An immune response to the biopharmaceutical is expected because the test agents are human proteins administered to animals, and the response can limit the duration of the toxicity studies. Antibodies to a test agent can neutralize its activity and thus reduce exposure. In addition antidrug antibodies can potentially cause toxicity such as deposition of immune complexes in the kidney. This issue is discussed further in Chapter 10.

Mainly because of species specificity and immunogenicity, carcinogenicity studies are generally not appropriate for biopharmaceuticals. However, in cases of immune suppressants or growth factors where there is cause for concern, assessment of carcinogenicity/tumorigenicity has been expected. If the biopharmaceutical is pharmacologically active in rodents, a carcinogenicity

study can be conducted if needed, and if immunogenicity does not prevent exposure. In species-specific immune suppressants, the nonchronic toxicity study has been extended to 9 and 12 months to address this concern [3].

Immunotoxicity testing as outlined in the ICH S8 document [4] is not intended to cover biopharmaceuticals. However, many of the same assays can be used to assess immunologic effects of biopharmaceuticals [5,6]. These assays can help assess immunopharmacology as well as potential immunotoxicity as discussed later in this chapter.

### 1.2 Molecule Is Pharmacologically Active in Nonrodents Only

A typical toxicology program for a biopharmaceutical may be limited to repeat-dose studies in nonhuman primates (NHP) only (if that is the only species in which the drug is pharmacologically active), a safety pharmacology study, reproductive toxicology studies, and a tissue cross-reactivity study (for monoclonal antibodies only). Local tolerance of the test article can be evaluated in the repeat-dose toxicology study (histopathological evaluation of the injection site) or can be assessed in a stand-alone rabbit local tolerance test. Although the biopharmaceutical may not cross-react with the rabbit target, this test allows for the evaluation of the irritant potential of the clinical candidate.

The NHP has played an important role in the development of biopharmaceuticals by facilitating the general safety assessment and evaluation of these products in specialty disease areas. For example, aging primates are used to study geriatric diseases and osteoporosis, and for many ocular indications including macular degeneration, and retinopathy, or for testing retinal implants. Nonhuman primate models are also being developed to better understand the effects of drugs on the reproductive system and the immune system in order to better mimic effects that may be seen in humans. Several books have been published describing the various models in NHP and their use in drug development [7,8].

Nonhuman primate toxicity tests of biopharmaceuticals most often use macaques—primarily cynomolgus monkeys (*Macaca fasicularis*), although rhesus monkeys (*Macaca mulatta*) are also sometimes used. A large historical database exists for both rhesus and cynomolgus monkeys, and many contract research organizations (CROs) have experience with both strains. Some advantages of Macaques are that blood volume is not as limited as in rodents and many of the blood/serum-based markers of toxicity in NHP can then be used in clinical trials, allowing for direct comparison of the toxic effects of the drug in the preclinical studies with the effects seen in human patients.

Marmosets (*Callithrix jacchus*) may also be used and their small size (350–500 g) is both an advantage and a disadvantage. The advantages can include the small amount of test material needed and the relatively small amount of space required for suitable housing. The main disadvantage of small size is the low volume of blood that may be obtained. Also many biological therapeutics cross-react with cynomolgus or rhesus targets but not marmoset. Fewer CROs have experience with the marmoset and the historical database is more

limited. However, certain CROs do have experience with toxicity testing in marmosets [9].

The age and size of the monkey are important considerations in the toxicity testing of biopharmaceuticals. Generally, animals should not be smaller than 2 kg as this limits the blood volume available for sampling. Younger animals are also less able to handle the stress of the various procedures on the study and may be more susceptible to GI effects (i.e., diarrhea), leading to confounding toxicities unrelated to the test article. In addition smaller animals are likely sexually immature and may respond to the drug differently than an adult animal. Evaluating effects on reproductive organs in the general toxicity studies is also problematic in these immature animals due to extensive variability or a lack of/exaggerated response. The appropriate age of the animals may further depend on the biological activity of the compound, and the age of the expected patient population. Most CROs have historical data ranges for clinical pathology parameters from animals of various age ranges as well as from various sources. Maintaining the same strain and source of animals throughout the drug development program, and not switching because of animal availability, is very important. Animals from different sources can vary significantly in hematology or immunophenotyping parameters, which can make interpretation across studies difficult.

Neonatal or juvenile monkey studies are difficult to conduct, but these studies may be necessary depending on the intended patient population. If a juvenile monkey study is to be conducted to support use of the therapeutic in pediatric patients, the appropriate age in the cynomolgus monkey, that closely matches the intended patient population, must be carefully considered. Unfortunately, the appropriate age may vary depending on the target organ of the therapeutic, and target organs of potential toxicity. For example, the age of a monkey that is appropriate to mimic neurologic developmental parameters in humans may be different than that to mimic immunologic or reproductive developmental parameters [10]. Neonatal studies can only be conducted by CROs that have breeding capabilities on site, as it is difficult to ship animals less than six months of age because of the stress of shipping, yet few CROs have a large enough population of young animals the same age to use for a toxicology study. For the development of new therapies for geriatric diseases (prostate disease, ocular pathology, osteoporosis, diabetes, Alzheimer's disease, etc.) it is important to understand how age-related disease and pathology develop. The cynomolgus monkey has been used for this type of testing and some CROs have special groups of older animals (generally 13 years of age and older). Ovarectomized cynomolgus monkeys are the most well-established model for osteoporosis [11], and ocular toxicity testing is also well-established in NHP [12–15].

The viral status of the monkeys is also important to test, especially prior to longer term studies (≥3 months duration). Many biological therapeutics are immune modulators, and latent viral infections can manifest in the

chronic studies and confound the interpretation. Choosing B-virus negative animals is most important when an immunomodulatory compound is being tested.

Cardiovascular and CNS effects are often evaluated as part of the general toxicity program. ECGs can be measured in the repeated-dose toxicology studies although these are usually conducted in anesthetized animals. Stand-alone safety pharmacology studies (mainly cardiovascular/respiratory safety pharmacology studies) can be conducted in alert, telemetered animals. These are typically nonterminal single-dose studies, however, and additional data from anesthetized animals in repeated-dose studies can be valuable to look at potential multiple-dose effects. CNS safety pharmacology studies have not been validated in the NHP, but a small number of observations can be added to the cardiovascular/respiratory study. In addition, if the biopharmaceutical is a monoclonal antibody, an in vitro tissue cross-reactivity study is conducted with a panel of human and NHP tissues to address nonspecific binding of the monoclonal antibody.

Because the pharmacological action of biopharmaceuticals may occur at very low doses, a no observable effect level (NOEL) may not be seen in the repeated-dose toxicology studies. Testing the biological therapeutic at doses lower than the clinical range to try to achieve a NOEL does not add value to the clinical program. The no observable adverse effect level (NOAEL) is typically used in the safety studies. It can be difficult to determine what findings in the toxicology study are due to exaggerated pharmacology and when these findings become adverse and represent toxicity. Several considerations can be used to determine if these effects should be considered adverse, including a combined analysis of the biological and statistical effects, the presence of a dose–response relationship, whether the findings are seen in both sexes, whether the findings are outside the historical control range, and if related histopathological correlates exist. In addition, the clinical indication, the monitorability and reversibility of the effect, and the risk–benefit analysis for the patient population need to be considered. An adverse effect may be considered to be a change that may impair performance and generally have a detrimental effect on growth, development or life span, so it should be an effect that would be unacceptable if it occurred in a human clinical trial [16]. In addition, because the pharmacological activity of biopharmaceuticals may be very different in the disease state for which the test agent is being developed than in the healthy animals employed in the toxicology studies, adverse findings or exaggerated pharmacological effects may not be seen in the toxicology studies. If little to no toxicity is seen, it may not be possible to define a maximum tolerated dose (MTD). In such cases the use of a maximum feasible dose (MFD) or the use of reasonable multiples over the clinical doses has been sufficient to demonstrate safety. What constitutes "reasonable multiples" will depend on the clinical indication (life threatening vs. non–life threatening), the patient population (consideration of special populations i.e., children,

elderly, women of child-bearing potential), chronic versus acute treatment, concomitant medications, alternative therapies, and so forth.

### 1.3 Molecule Is Pharmacologically Active in Humans and Chimpanzees Only

In certain cases the human protein has been so species-specific that it will only cross-react with human and chimpanzees (*Pan troglodytes*). Although safety studies can be conducted in chimpanzees, many limitations exist (no histopathology can be conducted, only small animal numbers can be used, limited CROs can conduct the studies and/or countries may have restrictions in use, limited historical control data exist, etc.) such that these studies are not generally conducted. In cases where the chimpanzee is the only relevant nonhuman species, alternative strategies could include testing a surrogate molecule (the rodent protein in rodents, a monoclonal antibody that cross reacts with the cynomolgus epitope, etc.) or using transgenic or knockout mice that overexpress or have a deletion of the targeted protein.

***Surrogate Molecules*** One alternative to using nonhuman primates (or if the biopharmaceutical does not cross-react with nonhuman primates) is to use a surrogate (homologous, analogous) molecule. This would be the homologous protein (i.e., the murine protein for use in mice) or, for a monoclonal antibody therapeutic, an antibody that cross-reacts with the rodent or cynomolgus monkey target. This allows for the safety testing of the pharmacologic activity of the test agent, but it does not allow for testing of the clinical candidate itself. The surrogate molecule, however, should resemble the clinical candidate as much as possible with regards to the production process, range of impurities/contaminants, pharmacokinetics, binding affinity, and pharmacological mechanism. Making and characterizing a second molecule along with the clinical candidate results in the use of a great deal of additional resources. Additional assays must be developed to detect the product and antibodies that might form to the product. However, these efforts can allow for a greater understanding of the potential toxicities of the therapeutic candidate. Some toxicology programs have utilized this approach, for example, to assess reproductive and developmental toxicities with a murine surrogate of interferon-gamma [17] and with a monoclonal antibody surrogate [18,19].

***Transgenic/Knockout Animals*** Knockout and transgenic mice are rapidly gaining acceptance as routine tools for mechanistic research, and they offer considerable promise for generating specific models of toxicological importance. Gene-targeted or "knockout" animals have been created using molecular and cellular genetic engineering techniques to specifically lack an endogenous gene [20]. Knockout mice have been used to assess drug specific-

ity, to investigate mechanisms of toxicity, and to screen for mutagenic and carcinogenic activities of therapeutic candidates. Similarly the impact of novel therapeutic candidates can be estimated in knockout mice; generation of viable and fertile animals with null mutations for a potential target protein implies that pharmacological inhibition of the molecule in vivo will elicit no major adverse effects. Furthermore the apparent lack of an in vivo phenotype could be used in conjunction with substantial evidence of in vitro efficacy to support the selection of a likely NOAEL for use in preclinical pharmacology and toxicology studies. Knockout and transgenic mice, however, are often structurally normal even if functional abnormalities are apparent, while many engineered mice appear to lack both structural and functional defects. Subtle phenotypes (functional and/or structural changes resulting from the genetic engineering event) sometimes may be unmasked using pharmacological challenges or other physiological stressors [21,22].

Particular emphasis in future pharmacology and toxicology studies will likely be directed toward conditional knockout mice (to evaluate the impact of chemically mediated inhibition of a particular gene product at the relevant stage of life) and "humanized" knock-in animals (in which the endogenous mouse gene is replaced with the homologous human gene to examine its role in disease or drug metabolism). Humanized mice are of particular importance as these animals can be employed to evaluate the efficacy and toxicity of human proteins that are not pharmacologically active in normal rodents or that induce a neutralizing antibody response that limits long-term exposure. One particular criticism is that humanized mice manufacture one or a few human proteins of interest, but other proteins that interact with the human molecule are still of mouse origin. The physiological effect of human–mouse protein interactions may differ slightly—or substantially—from that of the normal human–human protein association. With the increasing number of species-specific biopharmaceuticals in development, it becomes even more important to demonstrate that the knockout mice are a viable alternative to testing in NHP and are relevant to the findings seen in humans.

## 2 IMMUNOTOXICITY TESTING

Immunotoxicity testing guidelines exist for small molecules where the toxicology is largely unpredictable and rodent species are typically used [4]. Despite the lack of a specific guidance on immunotoxicity evaluation until now, most biopharmaceuticals have assessed the immunotoxic potential of the biopharmaceutical as a part of general single- and/or repeat-dose toxicity studies [5,23].

The ICH guidance on immunotoxicity testing suggests a weight-of-evidence decision making approach. The general principles that apply to this guideline are as follows:

1. All new human pharmaceuticals should be evaluated for the potential to produce immunotoxicity.
2. Methods include standard toxicity studies and additional immunotoxicity studies conducted as appropriate. Whether additional immunotoxicity studies are appropriate should be determined by a weight of evidence review.

Data from standard toxicity tests could indicate an adverse effect on the immune system, including changes in hematology, alterations in immune organ weights or histology, changes in serum globulins, increased incidence of infections, or increased occurrence of tumors. If the weight of evidence suggests that additional testing is needed, then it is suggested that an immune function study be conducted, such as the T-dependent antibody response (TDAR) and immunophenotyping of leukocyte populations. For small molecule drugs, these studies are generally conducted in rodents in a 28-day study.

Although the ICH S8 guidance does not specifically apply to biopharmaceuticals, the same principals for understanding immunotoxicity can be applied. Additionally, for human biopharmaceuticals, the immune system is often the intended target of the therapy and the immunotoxicity observed may be exaggerated pharmacology. In this case NHP are generally used and the immune tests have been selected based on the known immunomodulatory properties of the drug. These assays have also been used as pharmacodynamic markers of drug activity or efficacy for the immune modulators. It is important to distinguish among immunopharmacology, where the immune system is the target organ of the therapeutic effect; immunotoxicity, where nontarget immune effects such as autoimmunity or immunosuppression may be observed; and immunogenicity, which represents an immune response to the drug.

Specialized immunotoxicity tests typically are not routinely conducted unless an effect on the immune system was seen in the general toxicity studies, or there is a known pharmacologic effect of the test agent on the immune system. Additional immunotoxicity testing are conducted if there is cause for concern. The current regulatory guidelines recommend that immunopathology be used as the initial screen to detect immunotoxicity, since standard hematology and histopathology are often sufficient to detect immune system alterations. Immunopathology can include total and differential white blood cell counts, and evaluation of the histopathology of lymphoid organs such as the thymus, spleen, lymph nodes, gut-associated lymphoid tissue (GALT), and the bone marrow. In addition more detailed measurements of any change in size and cellularity of immune cells, germinal center development, cortex–medulla ratio of the thymus, and immunohistochemistry of the lymphoid organs may be included. Again, these parameters are examined as part of general toxicity studies for biopharmaceuticals.

Several important factors should be considered when including immunotoxicity testing into GLP (good laboratory practice) toxicology studies, especially if they are conducted in nonhuman primates. These include (1) whether

the assays have been validated, (2) whether to use the main study animals or a satellite group, and (3) the timing of these tests within the context of the GLP toxicology study. The advantages of using the "main" study animals instead of "satellite" animals for immunotoxicity testing are reduced animal use and correlation of any immunotoxicity findings with other toxicities seen in those same animals. The disadvantage of using main study animals is that the additional manipulations for immune testing (e.g., injection of an antigen for determining antibody response) may influence the toxicity or immunogenicity of the therapeutic agent. It is very important to include several baseline measurements because of the variability seen between animals, and even in the same animal over time. Because of the small number of NHP per group, it is important to reduce the variability in the assays as much as possible with regard to antigen source, technique, and so forth.

Flow cytometry is often included in a GLP toxicology study of immunomodulators to evaluate changes in lymphocyte subsets, including T cells ($CD4^+$, $CD8^+$), B cells ($CD20^+$), NK cells ($CD16^+$), and monocytes ($CD14^+$). These assays are typically conducted using peripheral blood, which allows for repeated sampling over time within the same animal. However, immunophenotyping can also be conducted on tissues to determine if there are effects on lymphocyte trafficking, although time points are limited to study termination unless serial biopsies can be performed (i.e., on lymph nodes). Serial biopsies may be difficult because they cannot be performed by all laboratories, and potential infections or other effects on the animals can affect data interpretation. Flow cytometry has also been used for more functional endpoints of immune competence, including lymphocyte activation, cytokine release, phagocytosis, apoptosis, oxidative burst, and natural killer (NK) cell activity. Any of these endpoints can be added if the mechanism of action of the drug suggests involvement of a particular function or type of immune cells.

In NHP the assay most commonly used to assess immune competence is the T cell dependent antibody response (TDAR). The ability to mount an antigen-specific antibody response requires a fully functioning immune system of T cells, B cells, antigen-presenting cells, cytokine production, and so forth. Animals are generally immunized with keyhole limpet hemocyanin (KLH) or tetanus toxoid (TT), and circulating antigen-specific antibody levels are measured by enzyme-linked immunosorbent assay (ELISA) or other methods. Immunization with KLH or TT should occur prior to drug treatment to assess the effects on the secondary antibody response (i.e., first immunization sc on day-7 and second immunization 14 days later), and the other antigen can be injected after two weeks of treatment to determine the effect on the primary immune response 7 to 10 days later. This immunization regimen allows for the assessment of both the primary and the secondary T cell dependent antibody response within the 28-day GLP toxicology study. For studies of longer duration, a booster immunization can be given at a later time point to assess the effect on the memory response, or to see if an altered response returns to normal during the recovery period.

Other immune parameters can be measured in the NHP such as cytokine measurements and delayed type hypersensitivity measurements, although these are less well characterized. Many human ELISA kits for cytokines can be used to measure cytokine levels in the NHP, but it is very important to determine if the reagents in these kits do truly cross-react with NHP cytokines. Whereas many of the human reagents do cross-react, exceptions exist, and the reagents need to be tested prior to use on a toxicology study.

Although immunomodulation can be assessed in the NHP, the assays are less well characterized than those used in the rodent. One issue is the lack of consistent protocols, and the timing of incorporating these assays into standard GLP toxicology studies varies. More historical control data are needed, and many assays have not been tested with an immunomodulatory control to confirm the level of sensitivity of the assay for detecting a mild/moderate immune modulator (both immunoenhancing and immunosuppressive activity). Currently there is an ILSI/HESI initiative to pool these types of data from numerous companies to improve our understanding of these assays in the NHP. Greater variability is inherent to NHP than in-bred rodents, and the animal number per group is generally much smaller than in rodent studies. It is therefore critical to find ways of reducing the variability in the assay to allow for more meaningful data interpretation. Strategies to improve study designs can include decreasing the inter-animal variability (using animals from the same source and of similar ages, decreasing stress during the study, increasing the number of baseline samples, etc.) and decreasing assay variability (standardizing the antigen source, assay technique, timing, etc.).

We are now at the stage at which assays of immunomodulation can be conducted in nonhuman primates, but unfortunately, we lack sufficient knowledge on the data generated across all the various developmental programs on biopharmaceuticals to date regarding which assays are the most useful in predicting immunomodulatory effects in humans. Assay methods thus need to be standardized so that we can truly compare the data to make that determination. Comparing data from the NHP with the immunotoxicity data in rodents would be useful to evaluate whether the NHP is more predictive of the human response. Additionally regulatory agencies should continue to treat the immunotoxicity testing of biological therapeutics on a case-by-case basis. However, immune testing in NHP for biopharmaceuticals goes beyond the estimation of immunotoxicity. Immune testing can be very valuable for understanding the pharmacology of an immune modulator and can help establish pharmacodynamic markers that can then be used in clinical trials. Combining all of the available data in NHP will allow for an improvement in the models and a better understanding of the value of these data. In addition differences have been seen in immune parameters (especially immunophenotyping) between cynomolgus monkeys from different geographical locations. It is therefore very important to keep the same source of animals for toxicology studies throughout the drug development program.

Recent adverse events reported with immunomodulatory monoclonal antibodies, such as Tysabri and TGN1412, highlight the need to improve the predictivity of immune effects in humans. During the clinical testing of a novel superagonist anti-CD28 monoclonal antibody, TGN1412, six healthy male volunteers suffered from a systemic inflammatory response characterized as a "cytokine storm" and became critically ill [24], a response that had not been predicted from the preclinical studies in NHP [25]. Despite the use of a battery of murine, NHP studies, and even ex vivo human cell assays [26], the immunological models used in TGN1412 preclinical testing were of insufficient predictive power to anticipate the serious adverse events in humans [27]. Cytokine release syndrome (CRS) has been seen with other monoclonal antibodies such as OKT3, a murine anti-CD3 monoclonal antibody [28]. A humanized anti-CD3 antibody, HuM291, with a modified Fc 28 domain was tested to see if the CRS could be eliminated [29]. Both antibodies were tested in chimpanzees, and there was no clinical evidence of CRS, although substantial cytokine secretion was detected. Clinical testing of HuM291 also showed that humans still experienced mild to moderate CRS. This suggests that even chimpanzees may not be the best model for assessing the clinical effects of cytokine release (or the side effects associated with anti-CD3 therapy in humans may not necessarily result directly from cytokine secretion) [30]. Evidence in vitro has also implicated CD16 and LFA-1 on NK cells in CRS [31].

Natalizumab (Tysabri), an anti-α4 integrin monoclonal antibody approved for the treatment of multiple sclerosis, was recently withdrawn from the market temporarily due to cases of progressive multifocal leukoencephalopathy, a demyelinating disease of the central nervous system associated with immunosuppression [32]. These cases highlight our incomplete understanding of the immune system and the translation of preclinical results to humans.

What could be done to better assess immunomodulatory effects in NHP that will be predictive of the outcome in humans? One important consideration comes back to the "relevance" of the animal species. Not only should the binding affinity and functionality at the target be considered, but the relative potency as well can be a critical factor. Understanding the dose-concentration–response relationships for receptor occupancy, receptor modulation and functionality, and target expression and distribution in comparison to normal human subjects and in disease state can help us to better assess risk to patients.

## 3 CONCLUSIONS

Toxicity testing of biopharmaceuticals is highly dependent on species specificity. Every program could be different in the type of studies that are needed to support safety assessment (i.e., case-by-case approach). If the biopharmaceutical cross-reacts with the rodent target, then general toxicity tests can be conducted in two species similar to what is done with pharmaceuticals. If the

molecule is only pharmacologically active in nonrodents (usually nonhuman primates), than a more focused or targeted toxicology program (one species) is conducted. If the molecule is not active in nonhuman primates, then alternative strategies for conducting safety studies such as testing with surrogate (homologous animal protein) molecules, or using transgenic/knockout mice, need to be considered. These strategies have certain caveats that can be used to better understand the potential toxicities of the therapeutic candidate. Although immunotoxicity testing guidelines are not established for biopharmaceuticals, many of the same assays and methods have been used routinely to understand potential effects on the immune system (whether unintended or pharmacologic activity). Further work is needed to understand potential immune effects of biopharmaceuticals, especially in nonhuman primates, to improve our predictivity of safety in humans.

## REFERENCES


1. International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use. S6. Preclinical Safety Evaluation of Biotechnology-Derived Pharmaceuticals, 1997.
2. International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use. M3. Maintenance of the ICH Guideline on Non-clinical Safety Studies for the Conduct of Human Clinical Trials for Pharmaceuticals, 2000.
3. Clarke J, Hurst C, Martin P, Vahle J, Ponce R, Mounho B, Heidel S, Andrews L, Reynolds T, Cavagnaro J. Duration of chronic toxicity studies for biotechnology-derived pharmaceuticals: Is six months still appropriate? *Regul Toxicol Pharmacol* 2008;50:2–22.
4. International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use. S8. Immunotoxicity studies for human pharmaceuticals, 2005.
5. Cavagnaro J. Immunotoxicity assessment of biotechnology products: a regulatory point of view. *Toxicology* 1995;105:1–6.
6. Cavagnaro JA, Mielach FA, Myers MJ. Perspectives on the immuntoxicological evaluations of therapeutic products: assessment of safety. *Meth Immunotoxicol* 1995;1:37–9.
7. Korte R, Weinbauer GF. Towards *New Horizons in Primate Toxicology*. Muenster: Waxmann, 2000;1–257.
8. Korte R, Vogel F, Weinbauer GF. *Primate Models in Pharmaceutical Drug Development*. Muenster: Waxmann, 2002;1–188.
9. Zuhlke U, Korte S, Niggermann B, Fuchs A, Muller W. The common marmoset (Callithrix jacchus) as a model in biotechnology. In Korte R, Vogel F, Weinbauer GF, eds. *Primate Models in Pharmaceutical Drug Development*. Muenster: Waxmann, 2002;119–25.
10. Mann DR, Fraser HM. The neonatal period: a critical interval in male primate development. *J Endocrinol* 1996;149:191–7.

11. Jerome C. Osteoporosis and aging: the nonhuman primate model. In Korte R, Vogel F, Weinbauer GF (eds), *Primate Models in Pharmaceutical Drug Development*. Waxmann Munster, 2002;85–90.
12. Dayhaw-Barker P. The eye as a unique target for toxic and phototoxic effects. In Korte R, Weinbauer GF, eds. *Towards New Horizons in Primate Toxicology*. Muenster: Waxmann, 2000;145–58.
13. Goralczyk R. Histological aspects of primate ocular toxicity with special emphasis on canthaxanthin-induced retinopathy in the cynomolgus monkey model. In Korte R, Weinbauer GF, eds. *Towards New Horizons in Primate Toxicology*. Muenster: Waxmann, 2000;159–74.
14. Niggerman B. Ocular toxicity investigations in primates and options for improvements. In Korte R, Weinbauer GF, eds. *Towards New Horizons in Primate Toxicology*. Muenster: Waxmann, 2000;189–201.
15. Zrenner E. Objective functional evaluation: electroretinography (ERG), electro-oculography (EOG), visually evoked cortical potentials (VEP) and multifocal ERG. In Korte R, Weinbauer GF, eds. *Towards New Horizons in Primate Toxicology*. Muenster: Waxmann, 2000;175–88.
16. Dorato MA, Engelhardt JA. The no-observed-adverse-effect-level (NOAEL) in drug safety evaluation: Use, issues and definition(s). *Regul Toxicol Pharmacol*, 2005;42:265–74.
17. Bussiere JL, Hardy LM, Hoberman AM, Foss JA, Christian MS. Reproductive effects of chronic administration of murine interferon-gamma. *Reprod Toxicol*, 1996;10(5):379–91.
18. Clarke J, Leach W, Pippig S, Joshi A, Wu B, House R, Beyer J. Evaluation of a surrogate antibody for preclinical safety testing of an anti-CD11a monoclonal antibody. *Regul Toxicol Pharmacol* 2004;40:219–26.
19. Treacy G. Using an analogous monoclonal antibody to evaluate the reproductive and chronic toxicity potential for a humanized anti-TNF-alpha monoclonal antibody. *Hum Exp Toxicol* 2000;19(4):226–8.
20. Bolon B, Galbreath E, Sargent L, Weiss J. Genetic engineering and molecular technology. In Krinke G, ed. *The Laboratory Rat*. London: Academic Press, 2000; 603–34.
21. Doetschman T. Interpretation of phenotype in genetically engineered mice. *Lab Animal Sci* 1999;49:137–43.
22. Bolon B, Galbreath EJ. Use of genetically engineered mice in drug discovery and development: wielding Occam's razor to prune the product portfolio. *Int J Toxicol* 2002;21:55–64.
23. Brennan FR, Shaw L, Wing MG, Robinson C. Preclinical safety testing of biotechnology-derived pharmaceuticals. *Mol Biotechnol* 2004;27:59–74.
24. Ganesh Suntharalingam G, Perry MR, Ward S, Brett SJ, Castello-Cortes A, Brunner MD, Panoskaltsis N. Cytokine Storm in a Phase 1 Trial of the Anti-CD28 Monoclonal Antibody TGN1412. *N Engl J Med* 2006;355:1–11.
25. Hopkin M. Can super-antibody drugs be tamed? *Nature* 2006;440:855–6.
26. Wing MG, Waldmann H, Isaacs J, Compston DAS, Hale G. Ex-vivo whole blood cultures for predicting cytokine-release syndrome: dependence on target antigen and antibody isotype. *Ther Immunol* 1995;2:183–90.

27. Schneider CK, Kalinke U, Löwer J. TGN1412—a regulator's perspective. *Nat Biotechnol* 2006;24:493–6.
28. Abramowicz D, Schandene L, Goldman M, et al. Release of tumor necrosis factor, interleukin-2, and gamma-interferon in serum after injection of OKT3 monoclonal antibody in kidney transplant recipients. *Transplantation* 1989;47:606.
29. Hsu D, Shi JD, Homola M, Rowell TJ, Moran J, Levitt D, Druilet B, Chinn J, Bullock C, Klingbeil C. A humanized anti-CD3 antibody, HuM291, with low mitogenic activity, mediates complete and reversible T-cell depletion in chimpanzees. *Transplantation* 1999;68:545–54.
30. Norman DJ, Vincenti F, DeMattos AM, Barry JM, Levitt DJ, Wedel NI, Maia M, Light SE. Phase 1 trial of HuM291, a humanised anti-CD3 antibody, in patients receiving renal allografts from living donors, *Transplantation* 2000;70:1707–12.
31. Wing MG, Moreau T, Greenwood J, Smith RM, Hale G, Isaacs J, Waldmann H, Lachmann PJ, Compston A. Mechanism of first-dose cytokine-release syndrome by CAMPATH 1-H: involvement of CD16 (FcγRIII) and CD11a/CD18 (LFA-1) on NK cells. *J Clin Invest* 1996;98:2819–26.
32. Sheriden C. Third Tysabri adverse case hits drug class. *Nat Rev* 2005;4:357–8.

# Reproductive Toxicity Testing for Biopharmaceuticals

PAULINE L. MARTIN, PhD

**Contents**

1 Introduction and Background
2 Testing of Small Molecular Weight Drugs versus Large Molecular Weight Biopharmaceuticals
3 Reproductive Testing of Biopharmaceuticals That Show Limited Species Cross-Reactivity
4 Reproductive Toxicity Testing of Monoclonal Antibodies
5 Reproductive Toxicity Testing of Nontraditional Biopharmaceuticals
6 Discussion
References

## 1 INTRODUCTION AND BACKGROUND

The importance of evaluating the effects of pharmaceutical agents on the development of the embryo and fetus has been appreciated for at least 40 years following the thalidomide tragedy of 1957 to 1961. Since that time reproductive toxicity testing strategies have been developed for the evaluation of the reproductive and developmental toxicity of chemicals and medical products [1]. In 1966 the US Food and Drug Administration (FDA) published guidelines for Reproduction Studies for Safety Evaluation of Drugs for Human Use. The testing strategies were subsequently refined, and in 1994 the expert working group of the International Conference of Harmonization (ICH) of technical requirements for registration of pharmaceuticals for human use issued a final guideline for industry Detection of Toxicity to Reproduction

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

for Medicinal Products—ICH S5A [2]. According to the ICH S5A guidance document, for most medicinal products a three-study design consisting of fertility and early embryonic development (preferred species rat), embryofetal development (preferred species rat and rabbit), and prenatal and postnatal development (preferred species rat) is considered to be adequate. The S5A guidance also states that other strategies or combinations of studies and study designs might be as valid or more valid as the three-study design depending on the circumstances. In addition S5A guidance states "if it can be shown by means of kinetics, pharmacological, and toxicological data that the species selected is a relevant model for the human, a single species can be sufficient. There is little value in using a second species if it does not show the same similarities to humans."

For issues unique to the nonclinical safety testing of therapeutic biological products the ICH S6 guidance, Preclinical Safety Evaluation of Biotechnology-Derived Pharmaceuticals, was developed and finalized in 1997 [3]. The S6 guidance describes reproductive toxicity testing briefly and states that the specific study design and dosing schedule may be modified based on issues related to species specificity, immunogenicity, biological activity, and/or a long elimination half-life. Importantly, one of the main principles of S6 is use of relevant species for toxicity assessments. S6 also states that there may be extensive public information available regarding potential reproductive and/or developmental effects of a particular class of compounds (e.g., interferons) where the only relevant species is the nonhuman primate. In such cases mechanistic studies indicating that similar effects are likely to be caused by a new but related molecule may obviate the need for formal reproductive/developmental toxicity studies. Both S5A and S6 guidances emphasize the need for flexibility in reproductive toxicity testing strategies.

For FDA product labeling, a pregnancy category is designated for each pharmaceutical based on the risk to the fetus. The categories are A, B, C, D, and X with A having proved safety in humans and X having proved adverse effects in humans (Table 1). Thalidomide and isotretinoin are category X drugs. Most pharmaceuticals fall into either category C or B. In the absence of well-controlled studies in humans, a pharmaceutical will be designated category C if animal studies have shown an adverse effect or if no animal studies have been conducted. Since the human risk is unknown for category C designated drugs, they should be used during pregnancy only when needed and if the potential benefit to the patient justifies the potential risk to the fetus. In general, for a category B designation, animal studies would need to reveal no evidence of harm to the fetus.

In the following sections the testing strategies that have been applied in support of the approved biopharmaceuticals are reviewed and discussed in terms of the relevance to the evaluations of human safety and the lessons that have been leaned for future reproductive developmental toxicity testing. All information provided in the following narrative, unless otherwise referenced, have been extracted from the FDA summary basis of approval information www.fda.gov, the European Medicines Agency (EMEA) centrally authorized

**TABLE 1 FDA pregnancy category designations for product labeling**

| Category | Description |
|---|---|
| A | Adequate, well-controlled studies in pregnant women have not shown an increased risk of fetal abnormalities |
| B | Animal studies have revealed no evidence of harm to fetus; however, there are no adequate and well-controlled studies in pregnant women<br>or<br>Animal studies have shown an adverse effect, but adequate and well-controlled studies in pregnant women have failed to demonstrate a risk to fetus. |
| C | Animal studies have shown an adverse effect, and there are no adequate, well-controlled studies in pregnant women<br>or<br>No animal studies have been conducted, and there are no adequate, well-controlled studies in pregnant women |
| D | Studies, adequate, well-controlled, or observational, in pregnant women have demonstrated a risk to fetus. However, benefits of therapy may outweigh the potential risk |
| X | Studies, adequate, well-controlled, or observational, in animals or pregnant women have demonstrated positive evidence of fetal abnormalities; the use of the product is contraindicated in women who are or may become pregnant |

product reviews (EPARs) www.emea.europa.eu, and the *Physicians Desk Reference* (PDR).

## 2 TESTING OF SMALL MOLECULAR WEIGHT DRUGS VERSUS LARGE MOLECULAR WEIGHT BIOPHARMACEUTICALS

In general, the reproductive and developmental testing strategies have not differed greatly for small molecular weight drugs and large molecular weight biopharmaceuticals. Small molecular weight drugs are generally not species specific and can therefore be tested using standard well-validated study designs in rodents and rabbits for which extensive databases are now available. Some biopharmaceuticals are also not species specific and can therefore be tested using the same testing strategies as those used for small molecules, with minor modifications to the study designs based on the route of administration and dosing regimen based on a longer half-life of the biopharmaceutical. The non–species-specific biopharmaceutics include recombinant forms of human insulin and human growth factor, erythropoietin and erythropoetin analogues, granulocyte colony stimulating factor, and various enzymes.

Although similar testing strategies can be applied to small molecular weight drugs and large molecular weight biopharmaceuticals, there are some notable differences between small molecules and large molecular weight biopharmaceuticals that need to be taken into consideration. Small molecular weight drugs can freely diffuse across membranes. The degree of diffusion across

membranes depends on the physical properties of the molecule [4,5]. Therefore small molecular weight drugs can cross the placenta and produce direct exposure of the embryo/fetus to the drug. They can be secreted in the milk and be absorbed across the neonatal gut and can diffuse across the blood–testis barrier and have direct effects on spermatazoa. Therefore, because of their small size and their ability to diffuse across membranes, small molecules have the potential to affect many stages of reproduction and development.

Large molecular weight proteins and peptides, on the other hand, because of their large size, do not readily diffuse across membranes, and therefore any potential effects are more likely to be secondary effects of maternal or parental toxicity rather than direct effects on the fetus or the spermatozoa. For example, administrations of insulin, insulin analogues, or glucagon-like peptide 1 to pregnant animals, at doses that exceed the clinical dose, produce maternal hypoglycemia. Since the intended use of these hormones is to normalize blood glucose, the effects seen in the toxicology animals can be considered to be exaggerated pharmacological effects. Sustained maternal hypoglycemia leads to fetal malnutrition resulting in pre- and postimplantation losses and skeletal and visceral abnormalities [6,7]. The effects seen on the fetuses with the insulin analogues do not differ from that of regular human insulin when administered at suprapharmacological doses and do not differ from that of oral blood-glucose lowering drugs [8]. Similarly the erythropoeitic agents, at suprapharmacological doses, produce increases in maternal hematocrit and an increase in blood viscosity leading to impaired blood flow and oxygen delivery to various tissues in the developing fetus. As a consequence fetal growth retardation and decreased ossification are seen.

Therefore for large molecular weight biopharmaceuticals that have restricted biodistribution and have biological effects that are indistinguishable from that of the endogenous hormone, the effects on the fetus can be predicted by an understanding of the biology in the adult. The adverse effects on the fetus are not seen when maternal blood parameters are maintained within normal ranges. Because of the secondary effects seen on the fetus at suprapharmacological doses all the pharmacologically active molecules carry a pregnancy category of C except for insulin lispro, which failed to show maternal or fetal toxicity in rat and rabbit developmental studies and was designated as category B. It is, however, unlikely that insulin lispro is less toxic than the other insulin analogues since its pharmacological effects in animals are similar to that of normal human insulin. Therefore, for biopharmaceuticals that have poor penetration across biological membranes and have a well-defined biology in the adult animal, a single relevant species assessment of embryofetal development may be adequate. Some of the newer biotechnology derived therapeutics have been approved with reproductive/developmental toxicity assessments in a single species, rats (e.g., agalsidase beta and laronidase). No reproductive toxicity studies were performed for the platelet-derived growth factor (PDGF) whose intended route of administration is topical and showed limited systemic absorption.

The toxicity of small molecular weight drugs, however, cannot necessarily be predicted based solely on an understanding of the biology of the parent molecule in the maternal animals. Many small molecular weight drugs are susceptible to metabolic degradation, which may result in the formation of metabolites with different toxicity, biology, and distribution from that of the parent molecule. Since metabolic processes can differ across species, a two-species assessment of embryofetal development may be appropriate for most small-molecule drugs. For protein therapeutics, however, the expected consequence of metabolism is the degradation to small peptides and individual amino acids. Therefore for protein molecules that consist of only naturally occurring amino acids, the generation of toxic metabolites is unlikely and a single species assessment may be appropriate.

## 3 REPRODUCTIVE TESTING OF BIOPHARMACEUTICALS THAT SHOW LIMITED SPECIES CROSS-REACTIVITY

Some human specific proteins do not cross-react with rats and rabbits, and therefore alternate species need to be considered for the reproductive toxicity testing of these agents. The type I interferons are examples of this class of proteins. The type I interferons show a high degree of species specificity such that the macaque, and in some cases only the rhesus macaque, is a suitable biologically responsive species for toxicology testing. In addition the type I interferons are highly immunogenic in macaques. After about two weeks of dosing the dams develop neutralizing antibodies toward the human interferon that reduces maternal exposure and eliminates the pharmacological response. Therefore the design of the embryofetal development studies for these molecules has, in some cases, involved modification of study designs so that critical periods in development could be evaluated. The embryofetal development studies for these molecules showed an increased incidence of abortions. The abortions occurred early during gestation and were associated with decreased serum progesterone [9,10]. Since all the type I interferons have similar pharmacology, they all produced a similar early abortifacient effect. Treatment with the interferons was not associated with a tertatogenic effect in macaques or in patients with multiple sclerosis [11].

Because the type I interferons produced no adverse effects on embryofetal development and the early abortifacient effect could be predicted based on a measurement of female endocrine changes, the need to conduct additional macaque embryofetal development studies for type I interferons in nonhuman primates was considered unnecessary. Therefore, for the second-generation pegylated versions of the alpha interferons, only hormonal analysis was conducted in macaques. This is the only known example of where mechanistic studies have been accepted in place of developmental studies in accordance with ICH S6 guidance.

## 4 REPRODUCTIVE TOXICITY TESTING OF MONOCLONAL ANTIBODIES

The monoclonal antibodies represent a unique class of protein molecules with regard to developmental toxicity testing. Although antibodies are very large molecular weight proteins (~150 kDa), they are known to cross the placenta from mother to fetus [12]. The transfer of antibodies from mother to fetus occurs predominantly in the second and third trimesters in humans and in nonhuman primates and involves a specific transport mechanism. Because fetal exposure to therapeutic antibodies is likely if women are treated with therapeutic antibodies during pregnancy, the precaution sections of the product labels for each of the approved antibody products carries wording to this effect. The transport mechanism requires binding of the constant portion (Fc) of the antibody to a receptor (FcRn) on the placenta. Binding of antibodies to FcRn on endothelial cells is also involved in the antibody salvage pathway that results in serum persistence of antibodies [13,14]. Therefore all proteins that contain an intact Fc and can bind to FcRn will be transported across the placenta. This includes monoclonal antibodies and engineered proteins that contain the Fc antibody fragment in order to increase serum persistence of proteins, peptides, or drugs. Fab antibody fragments, such as abciximab, are not transported across the human placenta [15].

The approaches taken in the evaluation of the reproductive toxicity testing for monoclonal antibodies has varied greatly depending on the species cross-reactivity and immunogenicity of the antibody. The testing strategies employed and the results from these studies are summarized in Table 2.

For two of the approved monoclonal antibodies, infliximab and efalizumab, cross-reactivity was seen only to human and chimpanzee antigens. The chimpanzee is a protected species and therefore not appropriate for evaluating developmental toxicity. Since these monoclonal antibodies were being developed for non–life-threatening diseases, Crohn's disease, rheumatoid arthritis, and psoriasis, and included women of childbearing potential, an alternate strategy needed to be developed. In both cases surrogate monoclonal antibodies were developed that bound to the murine version of the antigen. With these surrogate antibodies all stages of reproduction, including embryofetal development, could be evaluated in the mouse [16,17]. However, the use of a surrogate (homologous) monoclonal antibody does not test the product intended for clinical use, and therefore these types of studies evaluate the potential risks associated with inhibition of a particular pathway rather than risks associated with administration of the clinical product. One alternative approach that has been employed is to develop a transgenic mouse that expresses the human antigen, thereby allowing the testing of the clinical product [18]. The disadvantages of this approach are that the biology in the transgenic mouse may not be the same as that in humans and the mice will likely develop an immune response to the human protein.

For most of the other monoclonal antibodies, species cross-reactivity has been limited to nonhuman primates. For these molecules the need to conduct reproductive and developmental studies has to be carefully considered on a case-by-case basis. S5A states that nonhuman primates are best used when the objective of the study is to characterize a relatively certain reproductive toxicant, rather than detect a hazard. The nonhuman primate reproductive toxicity studies are not powered to detect infrequent events.

For the antibodies intended to treat immune mediate diseases—adalimumab, natalizumab, and omilizumab—a macaque embryofetal development study was conducted. In these studies pregnant macaques were treated for a period including the period of organogenesis (gestation days 20–50, first trimester), and fetuses were examined at the end of the second trimester (gestation day-100). For the approved monoclonal antibodies only natalizumab and bevacizumab have shown cross-reactivity to species other than nonhuman primates. For natalizumab, a two-species embryofetal development assessment was conducted (macaque and guinea pig), whereas for bevacizumab, which was intended to treat cancer, only a single relevant species assessment was conducted (rabbit).

Fusion proteins that consist of recombinant forms of human proteins fused to the Fc portion of human IgG tend to be less species specific than monoclonal antibodies. Etanercept and abatacept showed cross-reactivity to rodents and rabbits, and therefore a full reproductive and developmental toxicity package could be obtained in rodents and rabbits for these agents. For alefacept species cross-reactivity was limited to the macaque.

For some monoclonal antibodies the epitope may only be present in the intended disease population. For example, the epitope may only be unregulated during disease, or the monoclonal antibody may target epitopes found only on viruses or bacteria. In such cases a single relevant species may be selected to assess off-target toxicity based on nonspecific binding to normal tissues. In these cases, however, unless there is nonspecific binding to reproductive organs, the nonhuman primate would not be a relevant species for assessing reproductive toxicity.

Reviews of the results from the reproductive toxicity testing for the approved monoclonal antibody biopharmaceuticals have shown very few incidences of harm to fetuses. The only notable effects that have been observed are hematologic changes in macaque fetuses exposed to natalizumab, skeletal abnormalities in rabbit fetuses exposed to bevacizumab, and immunological deficits in mice exposed to the efalizumab murine surrogate. In each of these examples the effects seen in the fetuses were predicable based on effects seen in the adult animals.

Natalizumab is a monoclonal antibody against the human α4 integrins (α4β1 and α4β7). The α4 integrins are expressed on leukocytes and hematopoeitic progenitor cells and mediate homing and adhesive functions. In disease states natalizumab prevents the migration of leucocytes into inflamed tissues and thereby suppresses inflammation [19]. In adult cynomolgus macaques the

**TABLE 2 Summary of reproductive toxicity testing at time of approval for monoclonal antibodies and Fc-containing fusion proteins**

| Product (Approval Date) | Mechanism of Action | Indication | Reproductive/ Developmental Toxicity Test Species |
|---|---|---|---|
| Zenapax daclizumab (1997) | Anti-CD25 (IL-2R) (mAb-IgG1) | Acute organ rejection | Not done |
| Simulect basiliximab (1998) | Anti-CD25 (IL-2R) (mAb-IgG1) | Acute organ rejection | Cynomolgus, macaque (EFD) |
| Remicade infliximab (1998) | Anti-TNF (mAb-IgG1) | Crohn's disease, Rheumatoid arthritis | Mouse—murine homologue (F, EFD, PPD) |
| Humira adalimumab (2002) | Anti-TNF (mAb-IgG1) | Rheumatoid arthritis | Cynomolgus monkey (EFD) |
| Raptiva efalizumab (2003) | Anti-CD11a (T cells) (mAb-IgG1) | Psoriasis | Mouse—murine homologue (F, EFD, PPD) |
| Tysabri natalizumab (2004) | Anti-α4 integrin (mAb-IgG4) | Multiple sclerosis | Cynomolgus monkey (EFD), Guinea pig (F, EFD, PPD) |
| Xolair omalizumab (2003) | Anti-IgE (mAb-IgG1) | Asthma | Cynomolgus monkey (F, EFD) |
| Rituxan rituximab (1997) | Anti-CD20 (B cells) (mAb-IgG1) | Refractory non-Hodgkin's lymphoma | Not done |
| Herceptin trastuzumab (1998) | Anti-HER2 (mAb-IgG1) | Metastatic breast cancer | Cynomolgus monkey (F, EFD) |

| Effects in Adult Animals in Chronic Toxicity Studies | Doses Tested in EFD Study (NOAEL or highest dose tested) | Effects on Fertility, Pregnancy and Development | Clinical Dose | Pregnancy Category |
|---|---|---|---|---|
| Information not available | Not done | Unknown | 1 mg/kg | C |
| *Macaque* None | 5 mg/kg (GD20-50) | No harm to the fetus | 20 mg (~0.3 mg/kg) | B |
| *Mouse* None | 40 mg/kg (GD6-18) | ↓ Male fertility? No harm to the fetus | 3–5 mg/kg | B |
| *Macaque* ↓ thymus weight, ↓ spelnic follicular centers | 100 mg/kg (GD20-97) | No harm to fetus | 40 mg (~0.6 mg/kg) | B |
| *Mouse* ↓ humoral immune response, ↑ WBC, ↓ cellularity lymph nodes | 30 mg/kg | No harm to the fetus, ↓ humoral immune response, ↓ cellularity lymph nodes | 0.7 mg/kg | C |
| *Macaque* ↑ WBC, ↑ reticulocytes, ↑ spleen weight | 30 mg/kg (GD20-70) | ↓ Female fertility, mild anemia ↑ spleen weight, ↓ liver and thymus weight, no teratogenicity | 300 mg (~4.3 mg/kg) | C |
| *Macaque* None | 75 mg/kg (GD20-50) | No harm to the fetus | 150–375 mg (~2–5 mg/kg) | B |
| ↓ B cells | N/A | Unknown | 375 mg/m$^2$ (~10 mg/kg) | C |
| None | 50 mg/kg (GD20-50) | No harm to the fetus | 2 mg/kg | B |

**TABLE 2** ***Continued***

| Product (Approval Date) | Mechanism of Action | Indication | Reproductive/ Developmental Toxicity Test Species |
|---|---|---|---|
| Campath alemtuzumab (2001) | Anti-CD52 (T and B cells) (mAb-IgG1) | Chronic lymphocytic leukemia | Not done |
| Erbitux cetuximab (2004) | Anti-EGF (mAb-IgG1) | Metastatic colorectal carcinoma | Not done |
| Avastin bevacizumab (2004) | Anti-VEGF (mAb-IgG1) | Metastatic colorectal cancer | Cynomolgus monkey (F) and rabbit (EFD) |
| Enbrel etanercept (1998) | TNFR:Fc IgG1 Anti-TNF | Rheumatoid arthritis, JRA, PsA | Rat (F, EFD, PPD) and rabbit (EFD) |
| Amevive alefacept (2003) | LFA3:Fc IgG1Anti-CD2 (T cells) | Psoriasis | Cynomolgus monkey (EFD, PPD) |
| Orencia abatacept (2006) | CTLA4:Fc IgG1(T-cells) | Rheumatoid arthritis | Mouse (EFD), rat (F, EFD, PPD), rabbit (EFD) |
| Reopro abciximab (1997) | Anti-GPIIb/IIIa Fab platelet inhibitor | Cardiac ischemic complications —PTCA | Not done |

*Source*: All information included in this table and the accompanying text have been extracted from the US FDA summary basis of approval information www.fda.gov, the European Medicines Agency (EMEA) centrally authorized product reviews (EPARs) www.emea.europa.eu and the Physicians Desk Reference (PDR).

*Note*: F = fertility, EFD = embryofetal development, and PPD = pre and postnatal development.

only notable toxicological findings following natalizumab treatment were increased circulating leukocytes, increased reticulocytes, and increased spleen weights. These effects are most likely due to the primary pharmacological effect of natalizumab of decreased adhesion of leukocytes and hematopoetic progenitors. Treatment of pregnant macaques with natalizumab resulted in fetal effects consisting of mild anemia, reduced platelet counts, increased

| Effects in Adult Animals in Chronic Toxicity Studies | Doses Tested in EFD Study (NOAEL or highest dose tested) | Effects on Fertility, Pregnancy and Development | Clinical Dose | Pregnancy Category |
|---|---|---|---|---|
| Information not available | N/A | Unknown | 3-30 mg (~0.04–0.4 mg/kg) | C |
| *Macaque* Skin/epidermal toxicity | N/A | Unknown | 400 mg/m$^2$ (~10 mg/kg) | C |
| *Macaque* ↓ Menstrual cycles, physeal dysplasia in adolescent males, ↓ body weight gain | ≤10 mg/kg (GD6-18) | Impaired fertility ↓ fetal body weight skeletal alterations | 5 mg/kg | C |
| None (15 mg/kg) | 40 mg/kg (GD6-18) | No harm to fetus | 50 mg (~0.7 mg/kg) | B |
| ↓ T cells | 5 mg/kg (GD20-birth) | No harm to the fetus | 7.5 or 15 mg (0.1–0.2 mg/kg) | B |
| ↓ Serum IgG, ↓ T and B cells activation ↓spelnic B cells | 200 mg/kg (GD6-18) | No harm to the fetus, ↑ T-dependent Ab response, thyroid inflammation | 10 mg/kg | C |
| | Not done | Not transported across human placenta in vitro | 0.25 mg/kg | C |

spleen weight, and reduced liver and thymus weight—associated with increased splenic extramedullary hematopoesis, thymic atrophy, and decreased hepatic hematopoiesis. These findings are generally consistent with the known pharmacological effects of natalizumab of decreased adhesion of hematopoetic progenitors and leukocytes and were completely reversible upon clearance of drug from the blood.

Bevacizumab is a monoclonal antibody against human vascular endothelial growth factor (VEGF). By inhibiting VEGF, bevacizumab inhibits tumor angiogenesis and thereby inhibits tumor growth. Toxicology studies conducted in young adult cynomolgus macaques showed bevacizumab treatment-related physeal dysplasia in immature animals with open growth plates and in mature females with reduced endometrial proliferation, an absence of corpora lutea, lower uterine weights, and decreased menstrual cycles [20]. In the six-month chronic toxicity study, decreased body weight gains were observed that were most likely due to decreased growth resulting from reduced angiogenesis in the bone growth plates. These effects are due to inhibition of VEGF that is required for normal physiological angiogenesis in the bone growth plates and in the ovaries. Bevacizumab showed cross-reactivity to cynomolgus macaque VEGF and to a lesser extent to rabbit VEGF. Because bevacizumab bound to rabbit VEGF, the rabbit was selected as an appropriate species for evaluating the developmental effects of bevacizumab. The lower sensitivity of the rabbit to bevacizumab was compensated for by increasing the dose levels and dose frequency relative to human dosing. Decreased fetal and maternal body weights were seen with an increase in the number of fetal resorptions and an increased incidence of specific gross and skeletal fetal alterations. These effects seen in the rabbit fetuses could be predicted based on the known biology of VEGF and the observations in the young adult animals.

Efalizumab is an antibody against the CD11a subunit of LFA-1 (lymphocyte function-associated antigen-1, a leukocyte cell surface protein); it inhibits the binding of LFA-1 to ICAM-1, 2, and 3 (intercellular adhesion molecules 1, 2, and 3). Efalizumab prevents the binding of LFA-1 on activated T lymphocytes to ICAM-1 on endothelial cells. Eflaizamab thereby inhibits immune-mediated processes involved in psoriasis. As mentioned previously, efalizumab does not bind to LFA-1 in any animal species suitable for conducting reproductive toxicity testing. Therefore the chronic toxicity studies and the reproductive toxicity studies were conducted with a murine surrogate of the human antibody referred to as muM17 [17]. In adult mice muM17 treatment-related effects consisted of increased circulating leukocytes, increased spleen weight, and decreased cellularity of the lymph nodes. These effects are most likely due to altered trafficking of white cells secondary to reduced adhesion. MuM17-treated adult animals showed a reduced primary antibody response to neoantigen challenge. In the developmental studies decreased antibody response was observed in the pups that was reversible by 25 weeks of age. Therefore the effects seen in the pups are consistent with the effects seen in the adults, and these effects are predictable based on the known biology of CD11a inhibition.

The other monoclonal antibodies shown in Table 2 produced either no toxicity or minimal biological effects in adult animals and no harmful effects to the fetus. One interesting observation is that even for those molecules that did show an effect in the fetuses, the effects that were observed were minor compared to effects that might be predicted based on observations in genetically deficient mice. Mice genetically altered to lack the $\alpha 4$ integrin are not

viable. These knockout animals show cardiac and placental defects that are embryolethal [21]. Cardiac defects and embryo lethality is also seen in rats treated with certain small molecules inhibitors of α4 integrins [22]. However, the developmental studies conducted in macaques showed no cardiac toxicity, and fetuses and neonates were essentially normal except for the reversible hematologic changes. Mice genetically altered to lack VEGF are not viable because VEGF is essential to embryonic vasculogenesis [23]. However, in the rabbits treated with bevacizamab during pregnancy, most of the fetuses did survive and showed predominantly skeletal alterations. Mice genetically altered to lack tumor necrosis factor are viable but show disorganized lymphoid architecture and a reduced humoral immune response [24]. However, developmental studies conducted with two anti-TNFα antibodies and one TNFα receptor-IgG Fc fusion protein, infliximab surrogate (cV1q), adalimumab, and etanercept have failed to show any detrimental effect on the fetal or neonatal immune system.

One reason why monoclonal antibodies produce fewer developmental toxicological effects than would be predicted from the genetically deficient rodents and fewer effects than seen with certain small molecular weight drugs may be that embryonic exposure to monoclonal antibodies during the critical period of organogenesis is minimal in primates. In humans and nonhuman primates the transfer of antibodies increases throughout the second and third trimesters [25–27]. By the time of the cesarean sections in the nonhuman primate embryofetal development studies (gestation day-100, end of the second trimester), monoclonal antibody concentrations in the fetuses have reached only a fraction of the maternal levels (~25%). By the time of birth in macaques (gestation day-165), antibody levels in the neonates are similar to that of the mothers [28]. Therefore detection of monoclonal antibody in fetal blood at gestation day-100 does not imply that the embryo was exposed to the antibody during the period of organogenesis (GD 20–50). It is likely that only the dams were significantly exposed to the monoclonal antibody during the period of organogenesis. Therefore for monoclonal antibodies there is a very low likelihood of teratogenic effects occurring due to direct embryonic exposure during organogenesis. However, if the monoclonal antibody has a biological or toxicological effect on development, it is more likely to have an effect on growth and maturation during the fetal period where antibody exposure is high than on organogenesis where exposure is low.

Because the period of organogenesis for rodents and rabbit (50% and 38%, respectively) occupies a greater proportion of the total gestational period than in nonhuman primates (15%) or humans (13%), the likelihood of placental transfer occurring during organogenesis is higher in rodents and rabbits than in nonhuman primates or humans. Therefore, if rodents or a rabbits are selected as the toxicology species for developmental studies, they may overestimate the risk for humans. This has been demonstrated for a lymphotoxin β-receptor IgG1-Fc fusion protein that showed a lack of lymph node development in the F1 generation when administered to pregnant mice [29] but normal lymph node development when administered to pregnant macaques [30]. Therefore

study design and choice of species are critically important for monoclonal antibodies.

Historically for the monoclonal antibodies intended to treat life-threatening diseases, such as oncology and organ rejection, reproductive toxicity studies have not routinely been conducted for all molecules prior to registration. The absence of this information resulted in the products being designated as pregnancy category C—namely the risk to the fetus is unknown. However, more recently approved oncology products have included embryofetal development toxicity studies either in cynomolgus macaques (trastuzumab) or in the rabbit (bevacizumab). With trastuzumab, which was developed to treat metastatic breast cancer and therefore included women of child-bearing potential, no adverse developmental effects were seen, allowing the product to carry a pregnancy category B. With bevacizumab the skeletal and visceral malformations that were seen resulted in the product remaining as a category C.

Cetuximab was approved in 2004 for the treatment of metastatic colorectal cancer. At the time of registration no developmental toxicology studies were conducted, consistent with that of many of the other oncology agents, and the product was designated as pregnancy category C. However, a macaque embryofetal development study was requested as a postmarketing commitment. Cetuximab is an antibody against epidermal growth factor receptor (EGFR). In adult macaques cetuximab produces skin toxicity due to the primary pharmacological action of EGF inhibition. Mice genetically altered to lack EGFR survive for up to eight days after birth and suffer from impaired epithelial development in several organs, including skin, lung, and gastrointestinal tract [31]. Therefore, based on the observations seen with the other monoclonal antibodies in macaques, it would be expected that the effects on the macaque fetus would be less than that seen in the genetically deficient animals and more consistent with that seen in the adult animals. Because epidermal toxicity can be expected in the fetuses, the product will likely retain its pregnancy category C designation.

One notable exception to this overall generalization for antibody therapeutics, that maternal biology predicts fetal biology (toxicity), is alefacept. Alefacept treatment produces profound depletions of T cells in adult animals due to its primary mechanism of action. However, in fetuses from macaques treated with alefacept during organogenesis and throughout pregnancy, no adverse effects on the immune system were seen even though fetal exposure was demonstrated. Therefore alefacept carries a pregnancy category of B.

## 5 REPRODUCTIVE TOXICITY TESTING OF NONTRADITIONAL BIOPHARMACEUTICALS

Most of the case studies described above are for recombinant forms of endogenous human proteins and human antibodies developed by traditional recombinant DNA technology. Newer technologies now have the ability to engineer

molecules with specific desired properties that may deviate from naturally occurring sequences. For these agents the principles outlined in the S6 guidance document may not always apply. In addition conjugation of proteins with drugs in order to improve the targeting of the drug to certain sites in the body or to improve the serum half-life may result in molecules that have properties of both large molecular weight biopharmaceuticals and small molecular weight drugs. One example is that of toxin-conjugated antibodies where the antibody molecules functions to target a cytotoxic agent to a tumor cell. The scope of the toxicity testing for these agents may depend on the novelty of the toxin in addition to the species cross-reactivity of the antibody. Conjugation of biologically inert molecules such as polyethylene glycol, in order to increase serum half-life of the biotherapeutic or to reduce immunogenicty, has shown no adverse impact on the toxicity profile relative to the unconjugated versions. For the polyethyene glycol modified version of biopharmaceuticals a reduced reproductive and developmental toxicology package may be appropriate if the biological effects of the therapeutic can be shown not to be altered by the conjugation.

Oligonucleotides are a class of biopharmaceuticals that do not fall under the traditional biotechnology-derived pharmaceuticals as described in ICH S6 guidance. However, many of the principles of ICH S6 do apply to the oligonucelotides because consideration of species specificity does apply. While reproductive toxicology studies were performed with the clinical material for the approved oligonucleotide pegaptanib, many classes of oligonucleotides are specific for binding to human mRNA sequences and therefore are inactive in rodents. In these instances the generation of surrogate oligonucleotides toward the relevant rodent mRNA is the only meaningful way to establish the safety of these molecules [32].

## 6 DISCUSSION

Review of the reproductive/developmental toxicity studies that have been conducted in support of the approved biopharmaceuticals have shown that historically a flexible case-by-case approach has been applied. For biopharmaceuticals that cross-react only with nonhuman primates a limited reproductive toxicity testing strategy has been employed based on the patient population, the indication, and the proposed clinical use. The studies that have been performed have been sufficient to inform the patient populations of potential risk to the fetus.

Based on the very limited database of approved biopharmaceuticals, a few generalizations can be made. For pharmacologically active proteins that produce exaggerated pharmacology in the mothers, the toxicological effects seen in the fetuses are secondary effects of maternal toxicity rather than direct teratogenic effects. Large molecular weight protein cross the placenta very poorly, and therefore fetal exposure is minimal. For these agents the aim in the clinic is to maintain physiological parameters within normal limits and

thereby avoid harm to the fetus. Because the fetal effects are secondary to exaggerated pharmacology in the mothers, the fetal effects can be predicted based on an understanding of the pharmacology in the adult animals. Therefore for molecules of a similar pharmacological class an argument could be made that mechanistic studies in adult animals could obviate the need to conduct formal developmental studies, as suggested by the S6 guidance document. In this way the number of animal studies need for the evaluation of developmental toxicity could be reduced. This approach is particularly important when the only relevant species is the nonhuman primate. However, the approach has been successfully applied only to the type I interferons.

The second generalization that can be made is that for monoclonal antibodies that produce biological effects in adult animals, similar biological effects are likely to occur in the F1 generation if the fetuses are exposed to the monoclonal antibody during the second and third trimesters and/or postnatally. Therefore, since a biological effect in the fetus is considered adverse, the product will most likely be labeled as category C. Monoclonal antibodies are unlikely to produce direct teratogenic effect in primates since exposure during the embryonic period is limited. Therefore, for these agents the need to conduct embryofetal development studies, especially when the only relevant species is the nonhuman primate, needs to be carefully considered.

It should be emphasized that the generalization made above are based on a very limited database of mostly traditional recombinant human proteins and monoclonal antibodies. However, these generalizations may not apply to second-generation molecules that contain components other than naturally occurring amino acids. Also the generalization that large molecular weight molecules, other than antibodies, do not cross the placenta may not be applicable to all molecules and may depend on the dose that is administered relative to clinical doses. If high enough doses are administered to pregnant animals, sufficient penetration may occur to produce a toxicological effect in the fetuses even if the fetal exposure relative to the maternal exposure is very low. Pegaptanib, an oligonucleotide-polyethylene glycol conjugate VEGF inhibitor with a molecular weight of 50 kDa was shown at high doses to cross the placenta in mice in sufficient amounts to produce developmental defects in the fetus similar to those seen in rabbits with bevacizumab. However, possibly because the developmental effects occurred only at exposure levels that were vastly in excess of clinical exposures, the developmental effects are not described in the product label, which carries a pregnancy category of B. Smaller molecular weight oligonucleotides, when tested at lower doses, were shown not to cross the placenta in significant amounts in mice [32].

For biopharmaceuticals for oncology indications primate reproductive toxicity studies have not historically been conducted. The absence of a negative developmental toxicity study results in the product carrying a pregnancy category C designation that indicates the risk to the fetus as unknown, and therefore the use of the drug during pregnancy should be avoided unless the benefit to the patient outweights the risk to the fetus. For oncology indications the

biopharmaceutical is frequently administered in combination with known teratogenic chemotherapeutic agents, for example, irinotecan, 5-fluorouracil or paclitaxel (pregnancy category D). Therefore a potential risk to the fetus exists for most drug–biopharmaceutical combinations. More recently, however, there has been a progression toward a requirement, or an expectation, that embrofetal development studies will be conducted for all biopharmaceuticals regardless of the indication or the species cross-reactivity. It is important to justify the use of animals in such cases for characterizing human risk.

For biopharmaceuticals that require developmental toxicity testing the design of the developmental studies should be adapted according to the type of molecule under investigation. The standard developmental protocols that have been developed for the evaluation of small molecular weight drugs may not be appropriate for all biopharmaceuticals. This is particularly important when the only relevant species is the nonhuman primate. If a primate developmental study is conducted, the design of the study should be based on good scientific judgment, as emphasized in the regulatory guidances, and not on a perceived regulatory requirement. The standard macaque embryofetal development study design has been adapted from the rodent and rabbit study designs in which pregnant animals are treated during the period of organogenesis and fetal examinations are conducted at a time when adequate anatomical evaluations can be conducted [33,34,10]. However, the macaque study design may need to be modified depending on the nature of the molecule being tested and the known biology of the molecule.

One important point about embryofetal development studies that needs to be emphasized is that traditionally an embryofetal development study examines only gross anatomical effects in the fetuses following exposure of the mothers to the drug during organogenesis. Histopathological examination of fetal tissues is not routinely conducted. Therefore for biopharmaceuticals that produce exposure during the fetal period, such as the antibodies, and therefore may have effects on the maturation of the immune system, for example, these effects would not be detected in a standard embryofetal development study unless additional immunological endpoints are incorporated. For rodent studies, functional and immunological endpoints are incorporated into the pre- and postnatal development studies. However, when the only relevant species is the nonhuman primate and only a single developmental study is planned, the study design should be modified to include additional clinically relevant endpoints and to cover the periods of greatest maternal and fetal exposure. These considerations, however, are no different than those for evaluating general toxicity where immune endpoints are routinely included in single- and/or repeat-dose toxicity studies to characterize toxicity (see Chapter General Toxicity Testing and Immunotoxicity Testing for Biopharmaceuticals).

The macaque embryofetal development study design has been shown to be appropriate for detecting the teratogenic effects of the small molecular weight drugs thalidomide and isotretinoin [35,36]. For large molecular weight biopharmaceuticals that do not cross the placenta the standard macaque

embryofetal study design will only be able to detect maternal effects that may be just as readily detectable in nonpregnant females. Even for monoclonal antibodies that are known to cross the placenta, the design of the standard embryofetal development study may not be optimal for fully evaluating risk to the F1 generation. Although antibodies are transferred across the placental from mother to fetus, the transport is not consistent throughout pregnancy. In humans and nonhuman primates very little antibody exposure occurs during the first trimester. Antibodies are transported from mother to fetus during the latter part of pregnancy mostly during the third trimester [12]. It is unlikely that the embryo will be significantly exposed to an antibody during the dosing period if the dosing is restricted only to the first trimester. Therefore utilizing the "standard" study design for monoclonal antibodies examines maternal effects and at most the effect of minimal fetal exposure during a limited portion of the second trimester.

Although a three-segment approach to reproductive toxicity testing is routinely conducted for small molecular weight drugs and biopharmaceuticals that cross-react with rodents, the need to conduct a three-segment evaluation in primates for primate-specific biopharmaceuticals needs to be carefully considered. For nonhuman primates a single-study design that involves examination of neonates from dams dosed throughout pregnancy and lactation may be a more appropriate study design than conducting separate embryofetal development and postnatal development studies. A combined study could reduce the number of nonhumans primates and provide more meaningful information. It is not clear what additional information is gained from terminating macaque pregnancies midgestation. Important points for consideration for primate reproductive toxicity testing are the ethical consideration of using primates versus rodents and the duration of the nonhuman primate studies. An embryofetal development study in macaques can take approximately one year to complete, and a combined embryofetal development and postnatal development study can take up to two years to complete. Therefore, if these studies are considered to be essential for evaluating human risk, the timing of the studies relative to clinical testing may need to be adjusted to ensure timely access of therapeutics to patients with debilitating diseases.

Another question that needs to be considered is whether fertility studies conducted in nonhuman primates provide sufficient useful safety information for the patients to justify the use of the animals? The one aspect of the nonhuman primate fertility studies that cannot accurately be assessed is fertility. Nonhuman primates have a naturally low fertility rate [37] and high spontaneous abortion rate [38] such that the number of animals that would be required to demonstrate a meaningful effect on fertility would be too large to be practical or ethical. Therefore the "fertility" studies in nonhuman primates focus on evaluation of hormone levels and semen analysis. This limited information provides minimal safety information over and above the standard toxicology endpoints, and therefore the value of conducting these studies should be carefully considered.

In conclusion, as defined in the ICH S6 guidance, the reproductive/developmental toxicity testing for biopharmaceuticals needs to be justified on a case-by-case basis. No single testing strategy can be applicable to all types of molecules. The reproductive/toxicity testing strategy should be based on good science and a thorough understanding of the molecular class and the pharmacological and toxicological properties of the molecule.

## REFERENCES


1. Collins TFX. History and evolution of reproductive and developmental toxicology guidelines. *Curr Pharmaceut Des* 2006;12:1449–65.
2. Guidance for Industry. Detection of Toxicity to Reproduction for Medicinal Products. ICH-S5A, 1994. www.fda.gov
3. Guidance for Industry. Preclinical Safety Evaluation of Biotechnology-Derived Pharmaceuticals. ICH-S6, 1997. www.fda.gov
4. Pacifici GM, Nottoli R. Placental transfer of drugs administered to the mother. *Clin Pharmacokinet Concepts* 1995;28:235–69.
5. Van der Aa EM, Peereboom-Stegeman JHJC, Noordhoek J, Gribnau FWJ, Russel FGM. Mechanisms of drug transfer across the placenta. *Pharm World Sci* 1997; 20:139–48.
6. Hofmann T, Horstmann G, Stammberger I. Evaluation of the reproductive toxicity and embryotoxcity of insulin glargine (LANTUS) in rats and rabbits. *Intl J Toxicol* 2002;21:181–9.
7. Tanigawa K, Kawaguchi M, Tanaka O, Kato Y. Sketelatal malformations in rat off spring: long-term effects of maternal-insulin induced hypoglycemia during organogenesis. *Diabetes* 1991;40:1115–21.
8. Baeder C. Embryotoxicological/teratological investigation, including effects on postnatal development, of the new oral antidiabetic glimepride after oral administration to rats and rabbits. *Clin Reprod* 1993;27:45–60.
9. Trown PW, Wills RJ, Kamm JJ. The preclinical development of Roferon®-A. *Cancer* 1986;57:1648–56.
10. Henck JW, Hilbish KG, Serabian MA, Cavagnaro JA, Hendrickx AG, Agnish ND, Kung AHC, Mordenti J. Reproductive toxicity testing of therapeutic biotechnology agents. *Teratology* 1996;53:185–95.
11. Sandberg-Wolheim M, Frank D, Goodwin TM, Giesser B, Lopez-Bresnahan M, Stam-Moraga M, Chang P, Francis GS. Pregnancy outcomes during treatment with interferon beta-1a in patients with multiple sclerosis. *Neurology* 2005;65:802–6.
12. Simister NE. Placental transport of immunoglobulin G. *Vaccine* 2003;21:3365–9.
13. Firan M, Bawdon R, Radu C, Ober RJ, Eaken D, Antohe F, Ghetie V, Ward ES. The MHC class I-related receptor, FcRn, plays an essential role in the maternofetal transfer of γ-globulin in humans. *Intl Immunol* 2001;13:993–1002.
14. Roopenian DC, Christianson GJ, Sproule TJ, Brown AC, Akilesh S, Jung N, Petkova S, Avanessian L, Choi EY, Shaffer DJ, Eden PA, Anderson CL. The MHC class I-like IgG receptor controls perinatal IgG transport, IgG homeostasis, and the fate of IgG-Fc-coupled drugs. *J Immunol* 2003;170:3528–33.

15. Miller RK, Mace K, Polliotti B, DeRita R, Hall W, Treacy G. Marginal transfer of ReoPro™ (Abciximab) compared with immunoglobulin G (F105), inulin and water in the perfused human placenta in vitro. *Placenta* 2003;24:727–38.
16. Treacy G. Using an analogous monoclonal antibody to evaluate the reproductive and chronic toxicity potential for a humanized anti-TNFα monoclonal antibody. *Hum Exp Toxicol* 2000;19:226–8.
17. Clark J, Leach W, Pippig S, Joshi A, Wu B, House R, Beyer J. Evaluation of a surrogate antibody for preclinical safety testing of an anti-CD11a monoclonal antibody. *Regul Toxicol Pharmacol* 2004;40:219–26.
18. Bugelski PJ, Herzyk DJ, Harmsen AG, Gore EV, Williams DM, Maleeff BE, Badger AM, Truneh A, O'Brien SR, Macia RA, Wier PJ, Morgan DG, Hart TK. Preclinical development of keliximab, a primatized anti-CD4 monoclonal antibody, in human CD4 transgenic mice: characterization of the model and safety studies. *Hum Exp Toxicol* 2000;19:230–43.
19. O'Connor P. Natalizumab and the role of alpha 4-integrin antagonism in the treatment of multiple sclerosis. *Expert Opin Biol Ther* 2007;7:123–36.
20. Ryan AM, Eppler DB, Hagler KE, Bruner RH, Thomford P, Hall J, Shopp RL, O'Neill CA. Preclinical safety evaluation of rhuMabVEGF, an antiangiogenic humanized monoclonal antibody. *Toxicol Pathol* 1999;27:78–86.
21. Yang JT, Rayburn H, Hynes RO. Cell adhesion events by α-4 integrins are essential in placental and cardiac development. *Development* 1995;121:549–60.
22. Crofts F, Pino M, DeLise B, Guittin P, Barbellion S, Brunel P, Potdevin S, Bergamann B, Hofmann T, Lerman S, Clark RL. Different embryo-fetal toxicity effects for three VLA-4 antagonists. *Birth Defects Res* 2004;71:55–68.
23. Ferrar N, Carver-Moore K, Chen H, Dowd M, Lu L, O'Shea KS, Powell-Braxton L, Hillan KJ, Moore MW. Heterozygous embryonic lethality induced by targeted inactivation of the VEGF gene. *Nature* 1996;380:439–42.
24. Pasparakis M, Alexopoulou L, Episkopou V, Kollias G. Immune and inflammatory responses in TNFα-deficient mice: A critical requirement for TNFα in the formation of primary B cell follicles, follicular dendritic cell networks and germinal centers, and in the maturation of the humoral immune response. *J Exp Med* 1996; 184:1397–411.
25. Malek A, Sager R, Kuhn P, Nicholaides KH, Schneider H. Evaluation of maternofetal transport of immunoglobulins during human pregnancy. *Am J Reprod Immunol* 1996;36:248–55.
26. Fujimoto K, Terao K, Cho F, Honjo S. The placental transfer of IgG in the cynomolgus monkey. *Jap J Med Sci Biol* 1983;36:171–6.
27. Coe CL, Kemnitz JW, Scneider ML. Vulnerability of placental antibody transfer and fetal complement synthesis to disturbance of the pregnant monkey. *J Med Primatol* 1993;22:294–300.
28. Coe CL, Lubach GR, Izard KM. Progressive improvement in the transfer of maternal antibody across the order primates. *Am J Primatol* 1994;32:51–5.
29. Rennert PD, Browning JL, Hochman PS. Selective disruption of lymphotoxin ligands reveals a novel set of mucosal lymph nodes and unique effects on lymph node cellular organization. *Intl Immunol* 1997;11:1627–39.
30. Martin PL. *Primate Models in Pharmaceutical Drug Development*. Muenster: Waxmann, 2002;105–18.

31. Miettinen PJ, Berger JE, Meneses J, Phung Y, Pedersen RA, Werb Z, Derynck R. Epithelial immaturity and multiorgan failure in mice lacking epidermal growth factor receptor. *Nature* 1995;376:337–41.
32. Henry SP, Denny KH, Templin MV, Yu RZ, Levin AA. Effects of human and murine antisense oligonucleotide inhibitors of ICAM-1 on reproductive performance, fetal development, and post-natal development in mice. *Birth Defects Res* 2004;71:359–67.
33. Weinbauer GF. *Primate Models in Pharmaceutical Drug Development*, Waxmann Verlag GmbH, 2002;49–66.
34. Buse E, Habermann G, Osterburg I, Korte R, Weinbauer GF. Reproductive/developmental toxicity and immunotoxicity assessment In the nonhuman primate model. *Toxicology* 2003;185:221–7.
35. Hendrickx AG, Makori N, Peterson P. Nonhuman primates: their role in assessing developmental effects of immunomodulatory agents. *Hum Exp Toxicol* 2000; 19:219–25.
36. Hendrickx AG. The sensitive period and malformation syndrome produced **by** thalidomide in crab-eating monkey (*Macaca fascicularis*). *J Med Primatol* 1973;2:267–76.
37. Meyer JK, Fitzsimmons D, Hastings TF, Chellman GJ. Methods for the prediction of breeding success in male cynomolgus monkeys (*Macaca fascicularis*) used for reproductive toxicology studies. *J Am Associ Lab Animal Sci* 2006;45:31–6.
38. Hendrie TA, Peterson PE, Short JJ, Tarantal AF, Rothgarn E, Hendrie MI, Hendrickx AG. Frequency of prenatal loss in a macaque breeding colony. *Am J Primatol* 1996;40:41–53.

# Reproductive/Developmental Toxicity Assessment of Biopharmaceuticals in Nonhuman Primates

GERHARD F. WEINBAUER, PhD, WERNER FRINGS, PhD, ANTJE FUCHS, PhD, MICHAEL NIEHAUS, PhD, and INGRID OSTERBURG

**Contents**

1 Introduction
2 Which Nonhuman Primate Model?
3 Fertility and Early Embryonic Development (Segment 4.1.1)
4 Prenatal and Postnatal Development (Segment 4.1.2)
5 Embryofetal Development (Segment 4.1.3)
6 Reproductive Failure and Relevance for Developmental Toxicity Studies
7 Prenatal and Postnatal Development of the Immune System
8 Conclusion
References

## 1 INTRODUCTION

Assessment of developmental and reproductive toxicity (DART) is commonly done in rodents (mice and rats) and in rabbits (mainly developmental toxicity). These evaluations use standardized experimental designs described in detail in the respective ICH guidelines (Table 1). The entire spectrum of DART testing comprises three segments:

- Fertility and early embryonic development (segment 4.1.1)
- Prenatal and postnatal development (segment 4.1.2)
- Embryofetal development (segment 4.1.3)

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

**TABLE 1 Guidances for the conduct of developmental and reproductive toxicity studies and their effective dates**

| Guidance | Date |
|---|---|
| Detection of toxicity to reproduction for medicinal products (ICH S5A) | March 1994 |
| Reproductive toxicology: male fertility studies (ICH S5B) | November 2000 |
| Nonclinical studies for the safety evaluation of pharmaceutical excipients | May 2005 |
| Safety testing of drug metabolites | June 2005 |
| Considerations for developmental toxicity studies for preventive and therapeutic vaccines for infectious disease indications | February 2006 |
| Nonclinical safety evaluation of pediatric drug products | February 2006 |
| Nonclinical safety evaluation of drug and biological combinations | March 2006 |
| Immunotoxicity studies for human pharmaceuticals (ICH S8) | May 2006 |

Beyond these three detailed developmental studies, a number of FDA guidelines can have a bearing on the need for the conduct of DART studies (Table 1).

Rodents and rabbits represent the standard DART species for small molecules and also for testing of biopharmaceuticals to which standardized designs can be applied. However, the development of biopharmaceuticals frequently necessitates the use of nonhuman primate models (NHP). A major reason for the requirement for NHP is species-specific cross-reactivity. Many therapeutic antibodies are highly specific and may not even recognize rodent, rabbit, or dog tissues. According to a recent survey "for the majority of MAbs (monoclonal antibodies) on the market cynomolgus monkeys have been used as the toxicology species and viewed by the regulators as the most relevant species" (National Centre for the Replacement, Refinement and Reduction of Animals in Research 2006, http://www.nc3rs.org.uk). Confined cross-reactivity might also render the marmoset as the relevant species. Another reason for using NHP is immunogenicity of the test item. Unlike for small molecules, large molecules bear the danger of eliciting an immune response and antibody formation. It can also happen that neutralizing antibodies to the test item are being produced, in which case the test item is rendered biologically ineffective or with diminished activity. For humanized antibodies (containing human sequences/fragments) and large "human" molecules, it is generally assumed that NHP are more appropriate for prediciting immunogenicity than rodent models. However, it must be borne in mind that there are cases where the NHP is not predictive at all and that the assumption above cannot be generalized [1–3].

Nonetheless, the need for DART studies using NHP is a frequent encounter with biopharmaceuticals, and rather than being an alternate approach, NHPs became an essential species for DART. Generally, NHPs offer several advantages over rodents and rabbits with regard to DART because of similarity to

human, for example, with regard to endocrinology of testicular and ovarian function, endocrinology of early pregnancy, placental morphology and physiology, timing of implantation, and rates of embryonic development, and also because of similar responses to known human teratogens, for example, thalidomide and vitamin A [4]. The purpose of this contribution is to provide a state-of-the-art account on the feasibility and limitations of using NHPs for DART evaluation in the context of preclinical development of biopharmaceuticals.

## 2 WHICH NONHUMAN PRIMATE MODEL?

Reproductive and developmental physiologies have been described for a number of NHPs including Old World monkeys and New World monkeys. Notwithstanding this, among NHPs, the cynomolgus monkey model is now predominantly used for general and reproductive/developmental toxicity studies, followed by marmosets and rhesus monkeys. At our site the overall use of different NHPs in toxicology approximates 80% for cynomolgus monkey, 12% for marmoset, and 8% for rhesus monkey.

Cynomolgus monkeys *(Macaca fascicularis)* are sexually active and fertile throughout the entire year and are not considered to express distinct and significant reproductive seasonality. Under feral conditions, pregnant females were discovered throughout the entire year [5], and menstrual cyclicity was unrelated to seasonal environment [6,7]. Our own observations confirm the absence of ovarian seasonality in the cynomolgus monkey both under laboratory conditions and under natural light conditions [8]. For the cynomolgus monkey model well-established experimental designs for reproductive and developmental toxicity testing, and immunotoxicity assessment, are available [9,10]. The reproductive physiology and endocrinology have been extensively studied, and numerous publications demonstrate the relevance of the cynomolgus monkey as a preclinical model for human reproduction and development [11,12]. At present the cynomolgus monkey is the model of choice for DART assessment. Another advantage of the cynomolgus monkey resides in the fact that many assays for immune system evaluation are available for this species (see Chapter General Toxicity Testing and Immunotoxicity Testing for Biopharmaceuticals).

Rhesus monkey*s (Macaca mulatta)* are another species used for DART evaluation [4]. However, this species displays pronounced seasonality of reproductive activity for both sexes [13,14]. In this primate, reproductive functions are entirely shut off or severely diminished for approximately half of the year. In environments with distinct seasons sexual activity and gonadal activity are present roughly throughout October until March, but this may vary substantially for individual animals. It is crucial to consider that the annual rhythmicity of reproductive cycles can persist over years in captivity and under indoor artificial light pattern conditions [15]. Hence in indoor circumstances seasonality of reproduction may be uncoupled from the outdoor season, and eventually an animal may display periodic reproductive activation in unexpected months.

During the out-of-season periods, reproductive hormone secretion and gonadal activity are at a complete halt. Although the basic reproductive and developmental physiology appear similar to human in the rhesus monkey and many endpoint parameters are established, the distinct reproductive seasonality requires special timing and has practical implications for widespread use of this model for DART studies. With regard to immune system evaluation it can be assumed that most of the available tests for cynomolgus monkey will also be applicable to rhesus monkeys.

Marmosets *(Callithrix jacchus)* reside in the canopy of tropical forests and do not show reproductive seasonality of fertility or breeding. The basic reproductive and developmental features of the marmoset model are well described [16–19]. In general, reproductive physiology and endocrinology of the marmoset is substantially different from that of human and Old World monkeys [18] and the clinical relevance of the marmoset for reproductive toxicity assessment is unclear [20]. Among these differences are lack of menstrual bleeding, multiple ovulation (2–4 ova/cycle), twin/triplet pregnancies, common chorion/anastomoses leading to hematopoietic XX/XY chimerism, and the absence of key male fertility genes (e.g., DAZ = deleted-in-azoospermia gene) [17, 21–25]. Marmosets and tamarins seem to be the only anthropoid primate species that regularly exhibit multiple ovulation [26]. On the other hand, the sensitivity of marmosets to thalidomide derivates is well documented [27]. A totally unexpected difference between marmosets and Old World primates including human resides in the fact that marmosets lack luteinizing hormone (LH) [28]. Recent work suggests that this could even be a common feature of New World monkeys [25]. According to recent concepts the pituitary produces and releases chorionic gonadotropin (CG) instead of LH and CG stimulates gonadal steroid hormone production. Despite these differences a number of parameters are available for reproductive and developmental toxicity testing in this species (Tables 2, 3, 4 and 6). Aside from the different reproductive physiology, evaluation of immunotoxicity endpoints is somewhat limited in the marmoset model. Whereas standard techniques such as histology and hematology as well as immunophenotyping are available, further specific approaches to immune system evaluation are scarce. These tests comprise T cell dependent antibody responses (e.g., KLH) and limited determinations of immunoglobulins and cytokines. Analysis of NK cell activity is hampered by the limited blood volume available from marmosets.

## 3 FERTILITY AND EARLY EMBRYONIC DEVELOPMENT (SEGMENT 4.1.1)

Studies of fertility and embryonic development are designed to identify toxic effects resulting from treatment from before mating (both males and females) and continuing through mating and implantation. Unlike in rodents, the phase of conception until implantation is not evaluated in NHPs. These studies also

**TABLE 2 Parameters for evaluation of female fertility in the nonhuman primate model**

| | Parameter |
|---|---|
| *Oogenesis* | Histology |
| *Ovarian cyclicity* | *Endocrinology* or vaginal smears |
| Endometrial changes | Biopsies, ultrasound |
| Ultrasound examination | Follicular growth, ovarian cysts |
| *Endocrinology*[a] | |
| *Fertility* | Mating |

*Note*: Parameters refer to cynomolgus monkey. Italicized parameters are also available for the marmoset model. Since marmosets lack LH, the analysis of LH bioactivity reflects CG activity/levels.
[a]LH = luteinizing hormone, FSH = follicle-stimulating hormone, *CG* = *chorionic gonadotropin*, *prolactin*, *estrogens*, *progesterone*, *androgens*, inhibin A, inhibin B.

**TABLE 3 Parameters for evaluation of male fertility in the nonhuman primate model**

| | Parameter |
|---|---|
| *Spermatogenesis* | Histology, DSP, spermatogenic stages, flow cytometry |
| *Sperm maturation* | Epididymis |
| *Semen* | |
| *Testicular size* | *Caliper or ultrasound* |
| Testis biopsy | |
| *Endocrinology*[a] | |
| Prostate status | Prostate volume, uroflow |
| *Fertility* | Mating |

*Note*: Parameters refer to cynomolgus monkey. Italicized parameters are also available for the marmoset model. Since marmosets lack LH, the analysis of LH bioactivity reflects CG activity/levels.
[a]LH = luteinizing hormone, FSH = follicle-stimulating hormone, *CG* = *chorionic gonadotropin*, prolactin, *estrogens*, DHT = dihydrotestosterone, androstenedione, *progesterone*, *androgens*, inhibin A, inhibin B, SHBG = sex hormone binding globulin. DSP = daily sperm production.

assess potential changes in the estrous cycle of females and possible functional effects that may not be detected in acute toxicity testing by histological examinations of the male reproductive organs. Whereas the guidelines recommend physical mating for testing fertility—albeit feasible—this is not being done routinely in NHP mainly for two reasons: (1) Default litter size in macaques is one and checking for implantation sites is obsolete. Twin pregnancies/births are extremely rare in macaques with an overall twin live births incidence around 0.1% [7,29]. We encountered four twin pregnancies among more than 3600 pregnancies. (2) Spontaneous fertility rates in macaques are clearly below 100% (typically 40–70%) and pre-implantation loss in macaques is probably around 25% [30]. This scenario necessitates either large group sizes or

**TABLE 4 Parameters for evaluation of postnatal development in the nonhuman primate model**

| | Parameter |
|---|---|
| TK/antibody analysis | Maternal milk/blood, infant blood |
| Mother–infant interaction | Video recording and analysis |
| Functional and behavioral test battery | >20 items |
| *Growth and clinical signs* | |
| Immune system | *Antibodies* |
| | *Immunoglobulins* |
| | *Immunophenotyping* |
| | NK cell activity[a] |
| | Lymphocyte proliferation |
| | *TDAR* and DTH[b] |
| | CD immunohistochemistry of lymphatic organs |
| | *Cytokines* |
| *Hematology/clinical chemistry* | |
| *Organ weights* | |
| *Histopathology* | |
| *Optional* | ECG, ocular investigations, skeletal growth (DEXA), blood gas analysis |

*Note*: In case of continued maternal dosing, milk can be collected for further analysis. Parameters refer to cynomolgus monkey. Pre-/postnatal studies are considered feasible in the marmoset model (italicized parameters should be accessible in the marmoset model).
[a]NK = natural killer cell.
[b]TDAR = T cell dependent antibody response, DTH (=delayed-type hypersensitivity test) is problematic in NHP.

pronounced toxicity in order to be able to detect statistically significant effects on fertility via mating success. Compared to the sophisticated clinical parameters available (Tables 2 and 3), fertility test by mating is rather insensitive and is therefore not recommended except for specific indications that the test item might affect mating behavior. The request for physical mating in NHP fertility studies may be related to differential interpretation of ICH guidelines [31].

A large array of parameters is available for macaques for the evaluation of effects on male and female fertility (Tables 2 and 3). Most of these parameters correspond to clinical endpoints used by gynecologists and andrologists for diagnosis of infertility. For the marmoset the number of established fertility parameters is less comprehensive than for macaques (Tables 2 and 3). However, those parameters that are available for marmosets are considered sufficient to detect clear-cut effects on male and female fertility. Most important, no assays for the determination of inhibins are available for marmosets. This is unfortunate as inhibin B does represent a good marker for testicular toxicity [32] and inhibins are useful as markers of ovarian function [33].

## 4 PRENATAL AND POSTNATAL DEVELOPMENT (SEGMENT 4.1.2)

The set of prenatal and postnatal studies is designed to detect adverse effects on the conceptus and the offspring following exposure of the female from implantation through weaning. In the offspring, functional parameters such as behavior, maturation (puberty), and fertility are also tested. For the latter, the F1 generation is being mated to produce the F2 generation. Different from rodents and for feasibility considerations, the F1 infants are not observed throughout puberty and are not being mated for production of F2 in the NHP. Sexual maturity in cynomolgus monkeys occurs around 2.5 to 5 years of age. If F1 and F2 generation would be produced for cynomolgus monkeys according to the prevailing guidelines, such studies would require many monkeys and an estimated 5 to 8 years duration. Marmosets attain breeding maturity by 1.5 to 2 years of age. Yet, if marmosets would be used, a guideline-based study design would still require approximately 4 years.

Selection and enrollment of animals follow the same procedures used in an embryofetal development study. Dosing commences on gestational day (GD) 20 and lasts until termination of pregnancy. Duration of pregnancy in cynomolgus monkeys is around 160 days. For 167 control pregnancies the gestational duration ranged between 134 days and 184 days. It is possible to continue maternal dosing after delivery in order to evaluate transfer of test item to the newborn via breast milk. Transient antibody development of infants from mothers dosed with test item during lactation has been observed (confidential data not shown). A study design overview and a list of available parameters for infant evaluation are given in Figure 1 and Table 4.

A comprehensive test battery for behavioral and functional development is available for the cynomolgus monkey infant (Table 5). Many of these tests use the Brazleton-based human neonatal behavioral assessment scale, and the tests have been modified and adopted for use in NHP [34,35]. For mother–infant interaction, a video recording is evaluated quantitatively using a special computer program. Mother–infant pairs are videotaped in a home cage situation at 3 to 4 months of age. The tapes are scored for separations and reunions of the mother and infant, as well as which party initiates the separation/reunion and the reaction of the other party [36].

Whereas the experimental design is quite standardized regarding maternal aspects, the duration of postnatal infant assessments is rather variable between studies, for example, from 7 days to 720 days in our experience. Evidently there is a need for standardization of the timing and type of test batteries for infants and of the time period the infant should be raised. For behavioral tests, a postnatal observation period of 9 months appears essential since these tests can only by applied from the age of 6 months onward. For evaluation of immune system development, a period of 6 months appears mandatory since from that age onward blood volume is sufficient to conduct several tests, such as the T cell dependent antibody response (TDAR), NK cell

**TABLE 5 Available test battery for assessment of behavioral and functional development in the cynomolgus monkey infant**

| |
|---|
| Respiration rate |
| Muscle tonus[a] |
| Elicited state[a] |
| Dorsiflexion[a] |
| Grasp support[a] |
| Righting reflex[a] |
| Prone progression[a] |
| Clasp support[a] |
| Visual following[a] |
| Lip smack orient[a] |
| Sucking[a] |
| Rooting[a] |
| Snout reflex[a] |
| Glabellar tap[a] |
| Nystagmus[a] |
| Moro reflex[a] |
| Grip strength[b] |
| Pupillary reflex[c] |
| Learning ability |
| Mother-infant interaction (video-based) |

*Note*: If a learning test is included in the test battery, the infants should be raised for at least 9 months as this test is used first at the age of 6 months.
[a]Performed on days 1 and 7 postpartum.
[b]Performed on day 28 postpartum.
[c]Performed on days 1 and 7 postpartum—if negative the test is repeated on day 14 postpartum or later.

activity test, lymphocyte proliferation test, and immunophenotyping. In the case of TDAR, a clear response was elicited at the age of 6 months and followed by an increased response to the second KLH injection at the age of 9 months (Figure 2). Finally, consensus is also needed as to whether or not to terminate the infant for full histopathological examination at termination of the study.

## 5 EMBRYOFETAL DEVELOPMENT (SEGMENT 4.1.3)

These studies are conducted in order to detect adverse effects on the pregnant female and the development of the embryo and fetus exposed in utero. Basically the period between implantation and closure of the hard palate is being investigated. In the NHP, however, only the period between postimplantation

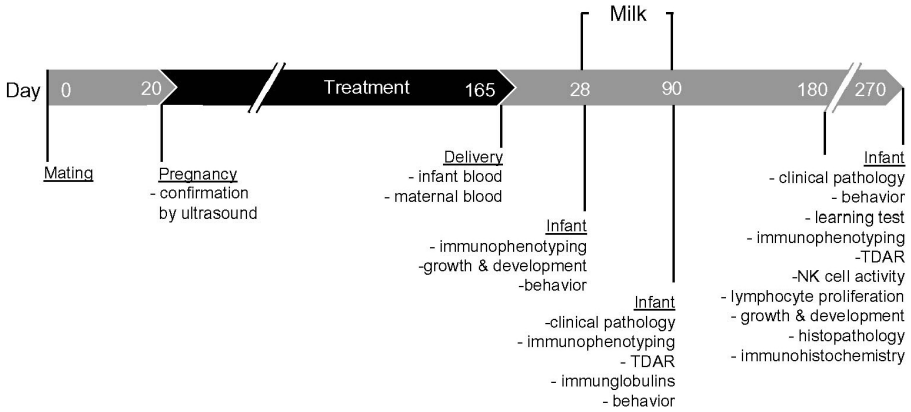


**Figure 1** Study design for the pre- and postnatal development evaluation in the cynomolgus monkey. Maternal treatment is either discontinued at term or is continued into the lactation period in order to investigate transfer of test item to the infant. Transfer of biopharmaceuticals via milk has been observed in this species. The duration of the postnatal observation period is variable and not yet standardized. In our laboratory the longest postnatal observation period covered 720 days. A variety of parameters is available for testing infant development (see also Table 4 for details). For biopharmaceuticals, a minimum period of 6 months appears appropriate. If the behavioral test battery comprises learning tests that are recommended from the age of 6 months onward, infants should be observed for at least 9 months.

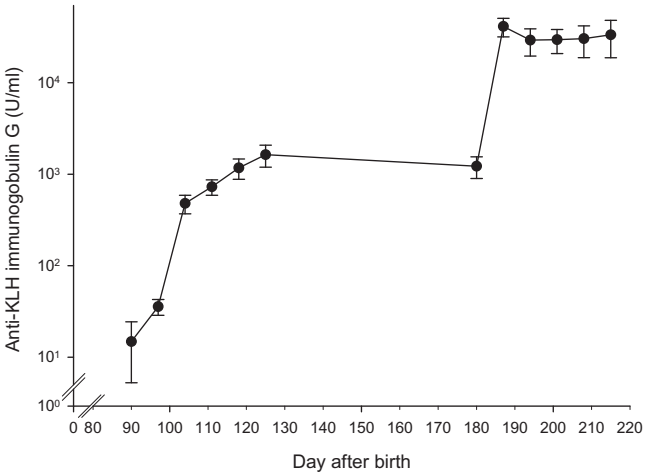


**Figure 2** T-cell dependent antibody response in immature cynomolgus monkeys at various postnatal ages. Keyhole limpet hemocyanin (KLH) was administered intradermally at 100 μg in Freunds's incomplete adjuvant. Note that a clear antibody response is present at about 3 months of age followed by an increased response after second immunization at about 6 months of age. Data represent mean values ± SEM of 5 to 6 animals (males and females combined).

and closure of the hard palate is covered. Animals are then dosed throughout the entire phase of organogenesis and the fetus is removed by cesarean section for examination.

The standard design and parameters for an embryofetal development study in cynomolgus monkeys are detailed in Figure 3 and Table 6. It is well established that the cynomolgus monkey is a relevant model for detection of effects on embryofetal development [37,38]. Timed matings are achieved by vaginal smear-based monitoring of the ovarian cycle and pairing around the suspected time point of ovulation followed by ultrasound examination for pregnancy. Since ovarian cycles cannot be synchronized, every female animal—once pregnant—is on its own time line and it may take some time—depending on the colony size of regularly cycling female animals—until all animals are pregnant and enrolled into the study. Test items are normally administered during GD 20 to 50 followed by cesarean section on GD 100 ± 1. If biopharmaceuticals with defined targets in the immune system are being tested, it may be advisable to prolong the period of dosing. For example, fetal thymic $CD3^+$ cells appear

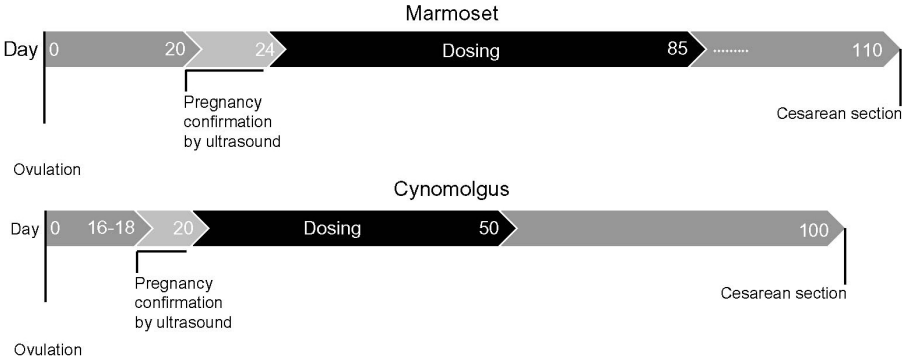


**Figure 3** Comparison of study design for embryofetal development evaluation in the marmoset and in the cynomolgus monkey. Following mating, pregnancy confirmation is achieved by ultrasound in both species. Double-sided arrows indicate the length of the dosing period. Note that the dosing period starts later (day 25 vs. day 20) and last longer (until day 85/longer [dotted line] vs. day 50) in marmosets compared to the cynomolgus monkey. Palate fusion occurs during GD 75 to 80 in the marmoset monkey and during GD 45 to 50 in the cynomolgus monkey. Cesarean sections are performed routinely on GD 110 ± 1 in the marmoset model and on GD 100 ± 1 in the cynomolgus monkey. Duration of pregnancy is somewhat shorter in the marmoset (approx. 140 to 150 days) compared to the cynomolgus monkey (approx. 155 to 165 days). It is important to recognize that for the marmoset, the mated pair must be housed together throughout the entire study duration, whereas for the cynomolgus monkey, the male is removed upon mating and the female is housed singly. This fact has implications for the study conduct with biopharmaceuticals in that male marmoset blood should also be analyzed for the presence of test item upon termination of the study.

**TABLE 6 Parameters for evaluation of embryofetal development in the nonhuman primate model**

| Maternal Parameters | Fetal Parameters |
|---|---|
| *TK analysis* | *Fetal examination* |
| *Hematology/clinical chemistry* | *Organ weights* |
| *Immunophenotyping* | *Histopathology*/CD immunocytochemistry of lymphatic organs |
| *Immunoglobulins* | *Immunoglobulins* |
| *Antibodies* | *Morphometry* |
| | Cardiac assessment |
| | *Amniotic fluid* |
| | Umbilical cord blood flow |
| | Umbilical cord blood for TK/blood gas analysis |

*Note*: Parameters refer to cynomolgus monkey. Italicized parameters are also available for the marmoset model. Parameters for a placental transfer study in cynomolgus monkey and marmosets comprise TK/PK samples, such as maternal blood before (animal not sedated) or at (animal sedated) cesarean section, blood from umbilical cord, and collection of amniotic fluid.

on GD 60 [39] and immunoreactive B cells are only present from GD 85 onward [10]. Studies with interferons or cytokines used a dosing phase of GD 20 to 70 or GD 20 to 80 [4]. In our laboratory we extended the dosing period from GD 20 up to GD 90 in embryofetal development studies using specific biopharmaceuticals.

A study plan for assessing embryofetal development in the marmoset is available [40]. Earlier work demonstrated the sensitivity of marmosets to thalidomide derivates [27]. Since marmosets do not have menstrual bleedings, cycle monitoring and timed matings were the pattern of circulating progesterone concentrations [41]. Embryofetal development is slower compared to macaques or human [42] and—following pregnancy confirmation by ultrasound—the dosing period ranged from GD 25 up to GD 109. Cesarean sections were performed on gestational day 110 ± 1, and fetuses were removed for further examination. Figure 3 compares the basic design of an embryofetal development study in the cynomolgus monkey and the marmoset. The number of live fetuses at cesarean section ranged from one to four with 62% twins and 24% triplets. Moreover a compound with teratogenic activity induced malformations at a rate of 99% (confidential data not shown). A major concern with such marmoset studies is the lack of a sizable reference database. On the other hand, the presence of the same malformation in twin fetuses from a test item exposed mother animal would—in all likelihood—be considered as evidence for test item related embryofetal toxicity without the need to consult a reference database. Since, however, the available experience using the marmoset approach is very limited, judgments and assumptions may change with increasing usage and experience in that nonhuman primate species (Table 7).

**TABLE 7 Presumed advantages and disadvantages of the marmoset model for embryofetal development evaluation**

| Advantages | Disadvantages |
|---|---|
| Body weight | Limited reference database |
| Litter size | Male presence required throughout study |
| Mating success | Animal supply |
| | Limited blood volume |

*Note*: As the experience using this approach is currently very limited, our judgments, and assumptions may need to be modified with increasing usage of this NHP.

If there is concern whether the biopharmaceutical is able to cross the placental barrier, a placental transfer study can be undertaken both in cynomolgus monkey or marmoset. Animals receive a single high dose of test item on GD 100 (cynomolgus) or GD 110 (marmoset) followed by cesarean section at a time point determined by the half-life of the test item. Parameters to be studied comprise TK/PK samples such as maternal blood before (animal not sedated) or at (animal sedated) cesarean section, blood from umbilical cord, and collection of amniotic fluid for analysis of test item concentrations. The qualitative pattern of immunoglobulin G transfer across the placenta and relative to gestational duration appears comparable between human and cynomolgus monkey [10,43,44]. Immunoglobulin G levels increase during the last trimester with a prebirth peak followed by a postnatal decline [10].

## 6 REPRODUCTIVE FAILURE AND RELEVANCE FOR DEVELOPMENTAL TOXICITY STUDIES

Reproductive failure is significant among primates. A review of this aspect in nine macaque species concluded that overall reproductive failure comprises—on average—16.3% abortions, 9.9% stillbirths, 21.9% neonatal deaths, and 15.2% infant deaths [45]. More recently Hendrie et al. [7] reviewed the incidence of prenatal loss in macaque breeding colonies. Throughout a 8 to 9 years observation period, the annual prenatal loss rates varied between 13% to 23% in the rhesus monkey (seasonal breeder), 10% to 50% in the bonnet monkey (seasonal breeder), and 8.6% to 28% in the cynomolgus monkey (non-seasonal breeder). Small [45] described six studies in the cynomolgus monkeys in which abortion rates ranged from 11.7% to 30.2%.

The cumulative prenatal loss rate for cynomolgus monkey until GD 150 across a 9-year period (1984 to 1993) was 14.8 % (65 losses in 439 pregnancies) [7] (Table 8). In our experience, cumulative prenatal loss rate in control animals was 12.9% (40 losses in 311 timed pregnancies) until GD 150 in time-mated toxicity studies across a 10-years period (1994–2004). The vast majority of prenatal losses occurred prior to GD 100. Interestingly the frequency of

**TABLE 8 Cumulative prenatal loss rates in timed-mated cynomolgus monkeys**

| Pregnancies | Losses | Percentage | Period | Data Source |
|---|---|---|---|---|
| 439 | 65 | 14.8 | 1984–1993 | Hendrie 1996 [7] |
| 311 | 40 | 12.9 | 1994–2004 | Covance[a] |
| 330 | 29 | 8.8 | 1994–2004 | SNBL[b] |
| 217 | 30 | 13.8 | — | CRL[c] |
| 138 | 13 | 9.4 | 2004–2006 | Covance[d] |

[a]Combined data from control animals from studies with cesarean section on gestational days 100 or 150. Animals were dosed during variable periods of gestation (range: days 16 to 150) using various dose routes (oral, subcutaneous, intramuscular, intravenous).
[b]Shin Nippon Biomedical Laboratories November 1, 2004, brochure.
[c]Charles River Laboratories—data until pregnancy term (Reproductive Toxicology Historical Data CD, Society of Toxicology 2006, file 7HCD-Nevada from 27.02.2006).
[d]Data from untreated pregnant animals until gestational day 70.

pregnancy failure did not relate to variable housing or management conditions such as handling, shipping, relocation, parity, and indoor time-mated versus outdoor random-mated [7]. In a cohort of 138 pregnant animals in whom dosing started after GD 70, we encountered 13 (9.4%) losses until GD 70 (Table 8).

It is important to recognize, however, that cumulative prenatal loss rates do not reflect variability between studies and single experimental groups. Despite reasonably low cumulative prenatal loss rates, abortion rates within a study can vary considerably. For separate controlled mating studies, prenatal loss rates ranging from 0% to 20% have been reported for the cynomolgus monkey [46–49]. A retrospective analysis of 70 embryofetal development studies from our laboratory suggests that—in statistical terms—prenatal loss rates of 10% to 30% occurred in 50% of studies and prenatal loss rates of 0% or more than 45% occurred in 5% of studies, respectively. Clearly, for cynomolgus monkeys—like other primate species—prenatal loss is a physiological reality and should be taken into account when planning for group size in developmental toxicity studies. With regard to potential untoward effects of a test item on embryofetal development—aside from the prenatal loss rate—one should also consider the timing of prenatal loss and whether a dose-dependency is prevalent.

The incidence of prenatal loss in the marmoset was 17%, and losses were detected between GD 36 and GD 102 [40]. Prenatal loss consistently affected all embryos/fetuses per female. Heger and colleagues reported a 20% to 35% postimplantation loss in the marmoset [50].

As already described earlier, NHP also experience significant loss at term and shortly thereafter [45]. Hence in a pre-/postnatal toxicity study overall reproductive failure (prenatal loss, stillbirth, neonatal and infant death) can amount to 40% or even 50% losses in single experimental groups. For developmental toxicity studies naive animals are frequently requested. Female animals with proven fertility and breeding experience might improve this

situation, in particular, with regard to their response to birth and thereafter. Interestingly, though, reproductive success appeared comparable between primiparous and multiparous cynomolgus monkeys [51] and rhesus monkeys/bonnet monkeys [7].

European guidelines are forthcoming that will recommend social housing of nonhuman primates whenever possible. Whether pair-housing or even group-housing will be feasible for pregnant nonhuman primates in toxicity studies remains questionable. Reports are available suggesting an increased prenatal loss and reproductive failure in pair-housed versus single-housed pregnant cynomolgus monkeys. Similarly production of viable offspring is more successful during individual housing compared to being exposed to groups of cage-mates.

## 7 PRENATAL AND POSTNATAL DEVELOPMENT OF THE IMMUNE SYSTEM

The pre- and postnatal development of the immune system in the cynomolgus monkey was studied using various approaches including detailed histological and immunohistochemical techniques [10,39]. Relative to pregnancy duration, the timing of crucial events—such as hematopoietic stem cell formation; cell migration to fetal thymus, liver, and bone marrow formation; preparedness for immune function and formation of memory cells—is strikingly similar between human and cynomolgus monkey but substantially different from mouse. It is evident that the prenatal and postnatal development and maturation of the immune system are largely comparable between human and cynomolgus monkey but are clearly different from rodents. The major difference relates to the fact, that in primates immune system maturation is generally more advanced relative to gestational and postnatal age compared to rodents.

Similar to humans [52,53], the CD4–CD8 ratio decreases during the postnatal phase in macaques. We observed a CD4–CD8 ratio reduction, on average, from 2.2 at one month of age to 1.3 at 12 months of age in the cynomolgus monkey and—in another study—a ratio decrease from 2.2 to 1.6 (Figure 4). Baroncelli et al. reported a ratio decrease from approx. 3.1 to 1.5 within 12 months in the same species [54]. A similar pattern was obtained for the rhesus monkey with a CD4–CD8 ratio decline from 3.5 in neonates to 0.7 in adult animals [55].

Data from KLH-vaccinated cynomolgus monkey infants show an evident TDAR in animals of 3 months of age and a pronunced secondary antibody response to a booster immunization approximately 3 months thereafter (Figure 2). Since TDAR requires functional antigen-presenting cells, T cells, and B cells (including the switch to antibody-producing plasma cells), it is supposed to be the most relevant functional assay for a general assessment of immunosuppression. Besides others, another functional assay to be considered for use in juvenile monkeys is a functional test of natural killer (NK) cell activity.

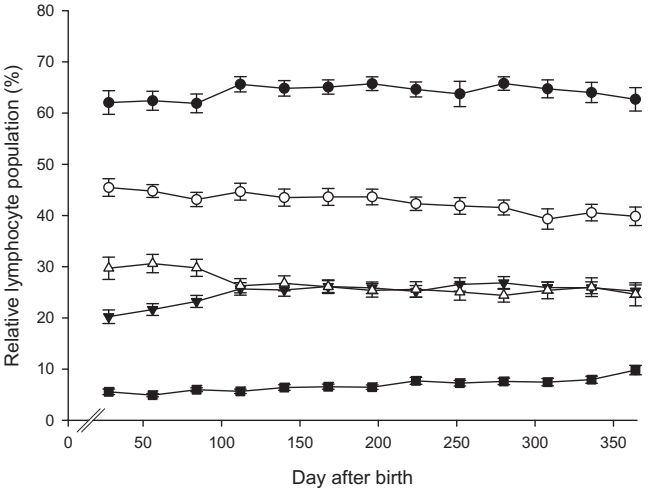


**Figure 4** Postnatal distribution of lymphocyte subsets by immunophentyping analysis in immature cynomolgus monkeys. Typically CD4–CD8 ratio declines during postnatal development, (e.g., from 2.2 to 1.6 within 12 months in this data set). Others reported a ratio decrease of 3.1 to 1.5 within 12 months for the same species [54]. Data represent mean values ± SEM of 26 to 32 animals (males and females combined). ● $CD3^+$, ○ $CD4^+$, △ $CD20^+$, ▼ $CD8^+$, ■ $CD16^+$.

However, it is known that the NK cell activity is low in neonatals of different species and that the ability to be activated by the functional stimulator interleukin IL-2 is reduced [56,57]. Therefore, and since a relatively high blood volume is required, this assay is recommended from the age of 6 months onward in the cynomolgus monkey.

## 8 CONCLUSION

DART evaluation of biopharmaceuticals frequently necessitates the use of NHP owing to species specificity. Cynomolgus monkey currently represents the most frequently used NHP. Fertility studies and embryofetal development studies are also feasible in the marmoset. A broad array of immune function tests is available for adult and infant cynomolgus monkeys. Primates experience significant prenatal losses and also losses around term, and this should be considered when planning group sizes. Whereas the study designs for male/female fertility and embryofetal development are established, a need for standardization prevails for the design duration of the postnatal phase in pre-/postnatal studies. The pre- and postnatal development of the immune system is similar between human and cynomolgus monkey but clearly different from rodents such as mouse. For biopharmaceuticals and immune system evaluation, a minimum postnatal observation period of 6 months is recommended.

If the behavioral test battery includes a learning test, a minimum of 9 months for postnatal evaluation is appropriate.

## ACKNOWLEDGMENTS

We are greatly indebted to Shiela Srivastav and Jan Sternberg for their help during preparation of this chapter.

## REFERENCES

1. Frings W, Cavagnaro JA. Predicting clinical immunogenicity: Intended or unintended. In Weinbauer G, Buse E, Muller W, Vogel F, eds. *New Developments and Challenges in Primate Toxicology*. Muenster: Waxmann, 2005;9–21.
2. Bugelski PJ, Treacy G. Predictive power of preclinical studies in animals for the immunogenicity of recombinant therapeutic proteins in humans. *Curr Opini Mol Ther* 2004;6(1):10–6.
3. Kropshofer H. Preclinical immunogenicity risk assessment of biotherapeutic proteins. In Weinbauer G, Vogel F, eds. *Novel Approaches towards Primate Toxicology*. Muenster: Waxmann, 2006;91–102.
4. Henck JW, Hilbish KG, Serabian MA, et al. Reproductive toxicity testing of therapeutic biotechnology agents. *Teratology* 1996;53(3):185–95.
5. Kavanagh M, Laursen E. Breeding seasonality among long-tailed macaques, Macaca fascicularis, in Peninsula Malaysia. *Int J Primatol* 1984;5:17–29.
6. Dang DC. Absence of seasonal variation in the length of the menstrual cycle and the fertility of the crab-eating macaque (*Macaca fascicularis*) raised under natural daylight ratio. *An Biolog Biochim Biophys* 1977;17:1–7.
7. Hendrie T, Peterson P, Short J, et al. Frequency of prenatal loss in a macaque breeding colony. *Am J Primatol*. 1996;40:41–53.
8. Niehoff M, Bergmann M, Müller W, Weinbauer GF. Effects of group housing on reproductive parameters. In Weinbauer G, Buse E, Muller W, Vogel F, eds. *New Developments and Challenges in Primate Toxicology*. Muenster: Waxmann, 2004;137–44.
9. Buse E, Habermann G, Osterburg I, Korte R, Weinbauer GF. Reproductive/developmental toxicity and immunotoxicity assessment in the nonhuman primate model. *Toxicology*. 2003;185(3):221–7.
10. Buse E. Development of the immune system in the cynomolgus monkey: the appropriate model in human target toxicology. *J Immunotoxicol* 2005;2:211–6.
11. ILAR. Nonhuman primate and other animal models of women's health. *Inst Lab Animal Res*. 2004;45:83–219.
12. Weinbauer G, Korte R. Reproduction in nonhuman primates—A Model Sytem for Human Reproductive Physiology and Toxicology. Muenster: Waxmann, 1999.
13. Herndon JG, Bein ML, Nordmeyer DL, Turner JJ. Seasonal testicular function in male rhesus monkeys. *Hormones Beh* 1996;30(3):266–71.

14. Ghosh D, Sengupta J. Patterns of ovulation and pre-implantation embry development, during the breeding season in rhesus monkeys kept under semi-natural conditions. *Acta Endocrinol* 1992;127:168–73.
15. Wickings EJ, Nieschlag E. Suppression of spermatogenesis over two years in rhesus monkeys actively immunized with follicle-stimulating hormone. *Fertility Sterility* 1980;34(3):269–74.
16. Luetjens CM, Weinbauer GF, Wistuba J. Primate spermatogenesis: new insights into comparative testicular organisation, spermatogenic efficiency and endocrine control. *Biol Rev Camb Phil Soc* 2005;80(3):475–88.
17. Hearn JP, Lunn SF. The endocrinology of the ovarian cycle and of pregnancy in the marmoset monkey, *Callithrix jaccus*. *J Endocrinol* 1975;65(3):1P.
18. Abbott DH, Barnett DK, Colman RJ, Yamamoto ME, Schultz-Darken NJ. Aspects of common marmoset basic biology and life history important for biomedical research. *Comp Med* 2003;53(4):339–50.
19. Li LH, Donald JM, Golub MS. Review on testicular development, structure, function, and regulation in common marmoset. *Birth Defects Res B Dev Reprod Toxicol*. 2005;74(5):450–69.
20. Zuhlke U, Weinbauer G. The common marmoset (*Callithrix jacchus*) as a model in toxicology. *Toxicol Pathol* 2003;31(Suppl):123–7.
21. Haig D. What is a marmoset? *Am J Primatol* 1999;49(4):285–96.
22. Gilchrist RB, Wicherek M, Heistermann M, Nayudu PL, Hodges JK. Changes in follicle-stimulating hormone and follicle populations during the ovarian cycle of the common marmoset. *Biol Reprod* 2001;64(1):127–35.
23. Gromoll J, Weinbauer GF, Skaletsky H, et al. The Old World monkey DAZDeleted in AZoospermia) gene yields insights into the evolution of the DAZ gene cluster on the human Y chromosome. *Hum Mol Genet* 1999;8(11):2017–24.
24. Jaquish CE, Tardif SD, Toal RL, Carson RL. Patterns of prenatal survival in the common marmoset (*Callithrix jacchus*). *J Med Primatol* 1996;25(1):57–63.
25. Muller T, Simoni M, Pekel E, et al. Chorionic gonadotrophin beta subunit mRNA but not luteinising hormone beta subunit mRNA is expressed in the pituitary of the common marmoset (*Callithrix jacchus*). *J Mol Endocrinol* 2004;32(1):115–28.
26. Tardif SD, Smucny DA, Abbott DH, Mansfield K, Schultz-Darken N, Yamamoto ME. Reproduction in captive common marmosets (*Callithrix jacchus*). *Comp Med* 2003;53(4):364–8.
27. Merker HJ, Heger W, Sames K, Sturje H, Neubert D. Embryotoxic effects of thalidomide-derivatives in the non-human primate *Callithrix jacchus*. I. Effects of 3-(1,3-dihydro-1-oxo-2H-isoindol-2-yl)-2,6-dioxopiperidine (EM12) on skeletal development. *Arch Toxicol* 1988;61(3):165–79.
28. Muller T, Gromoll J, Simoni M. Absence of exon 10 of the human luteinizing hormone (LH) receptor impairs LH, but not human chorionic gonadotropin action. *J Clin Endocrinol Metab* 2003;88(5):2242–9.
29. Resuello RG. Cynomolgus monkey twins. *Vet Rec*. 1987;121(7):155.
30. Binkerd PE, Tarantal AF, Hendrickx AG. Embryonic/fetal loss and spontaneous malformations In nonhuman primates. In Neubert D, Merker HJ, Hendrickx AG, eds. *Non-human Primates—Developmental Biology and Toxicology*. Wien-Berlin: Ueberreuter Wissenschaft, 1988;115–27.

31. Mineshima H, Endo Y, Ogasawara H, et al. Impact of globalization under the ICH guidelines on the conduct of reproductive toxicity studies—report on current status in Japan, Europe and the US by questionnaire survey. *J Toxicol Sci* 2004; 29(3):201–15.
32. Foppiani L, Schlatt S, Simoni M, Weinbauer GF, Hacker-Klom U, Nieschlag E. Inhibin B is a more sensitive marker of spermatogenetic damage than FSH in the irradiated non-human primate model. *J Endocrinol* 1999;162(3):393–400.
33. Fraser HM, Groome NP, McNeilly AS. Follicle-stimulating hormone-inhibin B interactions during the follicular phase of the primate menstrual cycle revealed by gonadotropin-releasing hormone antagonist and antiestrogen treatment. *J Clin Endocrinol Metab* 1999;84(4):1365–9.
34. Golub MS, Gershwin ME, Hurley LS, Hendrickx AG, Saito WY. Studies of marginal zinc deprivation in rhesus monkeys: infant behaviour. *Am J Clin Nutrit*. 1985; 42(6):1229–39.
35. Golub MS, Sassenrath EN, Chapman LF. Mother-infant interaction in rhesus monkeys treated clinically with delta-9-terahydrocannabiol. *Child Dev* 1981; 52:389–92.
36. Maestripieri D, Jovanovic T, Gouzoules H. Crying and infant abuse in rhesus monkeys. *Child Dev*. 2000;71(2):301–9.
37. Hendrickx AG, Binkerd P. Nonhuman primates and teratological research. *J Med Primatol* 1990;19:81–108.
38. Hendrickx AG, Tzimas G, Korte R, Hummler H. Retinoid teratogenicity in the macaque: verification of dosing regimen. *J Med Primatol* 1998;27(6):310–8.
39. Buse E, Habermann G, Vogel F. Thymus development in *Macaca fascicularis* (cynomolgus monkey): an approach for toxicology and embryology. *J Mol Histol* 2006;37(3–4):161–70.
40. Fuchs A, Weinbauer G. Feasibilty of embryofetal development studies in the marmoset (*Callithrix jacchus*). In Weinbauer G, Vogel F, eds. *Novel Approaches towards Primate Toxicology*. Muenster: Waxmann 2006;110–22.
41. Harlow CR, Gems S, Hodges KH, Hearn JP. The relationship between plasma progesterone and the timing of ovulation and early embryonic development in the marmoset monkey (Callithrix jacchus). *J Zool London* 1983;201:273–82.
42. Merker HJ, Sames K, Csato W, Heger W, Neubert D. The embryology of *Callithrix jacchus*. In Neubert D, Merker HJ, Hendrickx AG, eds. *Non-human Primates—Developmental Biology and Toxicology*. Wien-Berlin: Ueberreuter Wissenschaft, 1988;219–42.
43. Simister NE. Placental transport of immunoglobulin G. *Vaccine* 2003;21(24): 3365–9.
44. Fujimoto K, Terao K, Cho F, Honjo S. The Placental transfer of IgG in the cynomolgus monkey. *Jap J Med Sci Biol* 1983;36:171–76.
45. Small MF. Reproductive failure in macaques. *Am J Primatol* 1982;2(2):137–47.
46. Cukierski MA, Prahalada S, Zacchei AG, et al. Embryotoxicity studies of norfloxacin in cynomolgus monkeys: I. Teratology studies and norfloxacin plasma concentration in pregnant and nonpregnant monkeys. *Teratology* 1989; 39(1):39–52.

47. Tarantal AF, O'Brien WD, Hendrickx AG. Evaluation of the bioeffects of prenatal ultrasound exposure in the cynomolgus macaque (*Macaca fascicularis*): III. Developmental and hematologic studies. *Teratology* 1993;47(2):159–70.
48. Carlock L, Oneda S, Bussiere J. An assessment of the effects of human soluble IL-4 receptor on reproduction and neonatal development when administered intravenously to pregnant cynomolgus monkeys. *Toxicologist* 2003;72:327.
49. Hardy L, Ortega S, Rogers B, Sinicropi D, Clarke J, Oneda S. Recombinant human nerve growth factor is not teratogenic when administered subcutaneously to cynomolgus monkeys. *Toxicologist* 1998;42(1–S):259.
50. Heger W, Merker HJ, Neubert D. Frequency of prenatal loss in marmoset monkeys (*Callithrix jacchus*). In Neubert D, Merker HJ, Neubert D, eds. *Non-human Primates—Developmental Biology and Toxicology*. Wien-Berlin: Uebereuter Wissenschaft, 1988;120–40.
51. Boot R, Leussink AB, Vlug RF. Influence of housing conditions on pregnancy outcome in cynomolgus monkeys (*Macaca fascicularis*). *Lab Animals* 1985; 19(1):42–7.
52. Denny T, Yogev R, Gelman R, et al. Lymphocyte subsets in healthy children during the first 5 years of life. *J Am Med Assoc* 1992;267(11):1484–8.
53. Heldrup J, Kalm O, Prellner K. Blood T and B lymphocyte subpopulations in healthy infants and children. *Acta Paediat* 1992;81(2):125–32.
54. Baroncelli S, Panzini G, Geraci A, et al. Longitudinal characterization of CD4, CD8 T-cell subsets and of haematological parameters in healthy newborns of cynomolgus monkeys. *Vet Immunol Immunopathol* 1997;59(1–2):141–50.
55. DeMaria MA, Casto M, O'Connell M, Johnson RP, Rosenzweig M. Characterization of lymphocyte subsets in rhesus macaques during the first year of life. *Eur J Haematol* 2000;65(4):245–57.
56. Hodge S, GHodge G, Flower R, Han P. Cord blood leucocyte expression of functionally significant molecules involved in the regulation of cellular immunity. *Scand J Immunol* 2001;53(1):72–8.
57. Frings W, Weinbauer G. Flow-ytometry-based evaluation of lymphocyte subsets and natural killer cell activity in developing and adult cynomolgus monkeys. *Toxicologist* 2004;78(1–S):430.

# Preclinical Evaluation of Cancer Hazard and Risk of Biopharmaceuticals

JOY A. CAVAGNARO, PhD, DABT, RAC

**Contents**

1 Background
2 Human Experience
3 Determining the Need for Preclinical Carcinogenicity Assessment
  3.1 Peptide Hormones and Growth Factors
  3.2 Immunomodulatory Biopharmaceuticals
  3.3 Cell Line Used to Produce Biopharmaceuticals
  3.4 Cellular-Based Therapies
  3.5 Gene Therapies
4 Additional Considerations for Biopharmaceuticals
5 Preclinical Approaches for Assessing Carcinogenic Potential
  5.1 Rodent Bioassay
  5.2 Short-term Carcinogenicity Assays
  5.3 Proliferation Indexes from Repeat-Dose Chronic Toxicity Studies
  5.4 Tumor Host-Resistance Models
  5.5 Transgenic Animals
  5.6 In vitro Cell Proliferation
6 Current Experience in Carcinogenicity Assessments of Biopharmaceuticals
  6.1 Immunomodulators
  6.2 Peptide Hormones
  6.3 Growth Factors
  6.4 Cytokines
  6.5 Enzyme Replacement Therapies
  6.6 Other

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

7 Rationale for Not Performing the Rodent Bioassay
8 Management and Communication of Carcinogenic Risk
References

*I am not so worried about a positive rodent bioassay because we usually explain it away; I am more worried about a negative one*
*—Professor Gerhard Zbinden (personal communication)*

## 1 BACKGROUND

Carcinogenesis is the term used to denote the development of neoplasia that can be induced experimentally by exposure to exogenous agents or can occur spontaneously without intentional or active intervention [1]. Neoplasias result following a complex biological process consisting of multiple cellular mutations followed by the selective growth of mutated cells, and the eventual progression of cells to malignancy [1,2]. While most of the known carcinogenic agents act directly on the cell and damage somatic cell DNA as part of the multi-step process, changes in the rate of cell division and cell death are also be critical to the process of malignant transformation. The diverse modes of action (MOA) of carcinogenic agents in relation to their effects on specific stages in the natural history of cancer development allow for greater congruence of many theories on carcinogenesis [1].

Currently the basic experimental paradigm used to identify potential human carcinogens and estimate their potency is the rodent bioassay [3,4,5]. This assay is nearly universally applied to new molecular entities, and positive results have an important impact on the future of a drug, including its approvability. When needed, carcinogenicity testing is usually performed during a clinical development program once activity and safety have been demonstrated in the intended patient population. A positive tumor response in the rodent bioassay is a strong stimulus to toxicologists to attempt to fully understand the reasons for the response [6]. Mechanistic studies may be suggested to support the claim that effects are not relevant or not of great concern to humans under the conditions of clinical use. Even when rodent carcinogenicity findings do not prevent approval, the results can restrict marketing or otherwise reduce their perceived competitive utility based on the potential risks communicated in the label [7].

In the early 1960s the National Cancer Institute (NCI) developed procedures to formalize the process for the evaluation of the human risk of cancer from chemical exposures [8]. This approach was based on decades of research that demonstrated that chemicals that were known to cause cancer in humans could also induce cancer in laboratory animals. The bioassay was originally envisioned as a cancer screen useful for identifying agents that would be examined in human epidemiology studies, assuming that relatively few compounds would induce tumors in animals. Despite the fact that the approach

by the NCI was only intended as an initial screening tool for carcinogenic activity the use of two rodent species exposed for two years as the primary means to identify potential human hazards was in widespread use by the early 1970s. In 1975 the approach was developed into a recommendation that formed the basis of regulatory guidance [9]. Rodent bioassays were subsequently expected for evaluating the safety of long-term chemical exposure and for the marketing of chronically administered drugs or those expected to have persistent effects. As such, the rodent bioassay became the "gold standard" for carcinogenicity assessment. The bioassay also evolved to occupy a primary role in the identification of agents considered "possible" or "probable" human carcinogens by the International Agency for Research on Cancer (IARC), and those "reasonably anticipated to be human carcinogens" by the National Toxicology Program (NTP) Report on Carcinogens [6].

Contemporary quantitative risk-assessment methodology for chemicals estimates cancer risks using slope factors derived from the rodent bioassay. These cancer-slope factors are based on, and intended to represent, the potency of the chemical under conditions of continuous lifetime exposure. However, in many situations for drugs used for chronic indications, exposure is not continuous over a lifetime but intermittent over short durations. When estimating risk from less-than-lifetime exposure, the assumption is usually made that cancer risk from exposure at a given rate decreases proportionally with decreasing exposure duration, and adjustments are made accordingly in the risk calculation. Although this approach represents a means to develop cancer risk estimates for limited exposure duration scenarios, its validity has never been well established [10]. One example that calls into question this paradigm is the recognized age differences in susceptibility to particular agents. If a carcinogen affects an animal or human primarily in a particular life stage, a short-term exposure during that stage may be very effective in producing cancer, while the same short-term exposure during a different life stage may be ineffective. When compared to continuous exposure over a lifetime, the apparent potency of the short-term exposure may be greater or less than that of continuous exposure, depending on when the short-term exposure occurs relative to the period of greatest susceptibility to the particular carcinogen [11].

Although sophisticated mathematical models are often relied upon for quantitative risk assessments, assessments have also been made on more qualitative evaluation of animal test data. In part this is because mathematical models are often based on assumptions on mutation rates and not all carcinogens are primary mutagens.

Over the past decade other concerns have been raised about the continued use of rodent data and its relevance for predicting human risk [11–16]. For example, more than half of the chemicals evaluated in the rodent bioassay have tested positive in one or more rodent species. One explanation has been that the large doses that are tested (i.e., maximum tolerated doses) may overwhelm the body's natural detoxification mechanisms. Cyclosporin A (CyA), a pharmaceutical used clinically as an immunosuppressant, is nongenotoxic (or

equivalent in an in vitro mammalian cell assay) and negative in the rodent bioassay but in humans is associated with an increase in the development of B cell lymphomas and squamous cell carcinomas, particularly of the cervix [17]. CyA and other immunosuppressants such as anti-lymphocyte globulin (ALG) probably act by their therapeutic role of depressing immunity resulting in a loss of normal body defenses against spontaneous cancer-causing agents, oncogenic viral agents, and so forth. There is also increasing evidence that some chemicals produce cancer in rodents through species-specific mechanisms that are irrelevant to human physiology and thus may not necessarily predict human hazard. Examples include urinary bladder tumors in rats exposed to high doses of saccharin or melamine, male rat kidney tumors induced by compounds like d-limonene that produce a characteristic male rat hydrocarbon nephropathy, rat mammary tumors produced in response to β-adrenergic blocking agents, and mesovarial leiomyomas in rats exposure to β-agonists like salbutamol. Endocrine tumors in rats including thyroid tumors or gastric ECL-cell carcinoid tumors have also been shown to be rodent specific as well as peroxisome proliferators that cause rodent liver tumors [16,18]. These examples are where specific biochemical mechanisms have been elucidated.

An increased knowledge about carcinogenic differences in the mode of action and processes between rodents and humans coincides with an emerging view that cancer arises from fewer pathways then scientists once assumed. Consideration has therefore been given to refocusing cancer hazard identification to the specific aspects of the tumor response, including toxicity exhibited in an organ, the history of response of the tumor type in that organ to other substances, as well as any evidence of altered oncogene or tumor suppressor gene expression. Measures of cell proliferation, evidence of altered hormonal stimulation, structure activity relationships (SAR), comparative exposure and metabolism data, pharmacodynamic and mechanism of action information, are therefore being considered in addition to other alternative short-term bioassays. The relevance of the rodent tumors to human cancer hazard is also being considered through the application of a "mode of action" analysis [19]. The mode of action is sufficient evidence to draw a reasonable working conclusion regarding the influence of the test articles on key processes. The mode of action concept permits information on precursor events to be evaluated and incorporated in the risk-assessment process.

Gottman and colleagues compared 121 replicated rodent carcinogenicity assays from the NCI/NTP and the literature to estimate the reproducibility of the rodent carcinogenicity assays for SAR studies and risk assessment [20]. The results indicated a concordance of 57% between the overall rodent carcinogenicity classifications from the replicate bioassays. This value did not improve substantially when additional biologic information (species, sex, strain, target organs) was considered, suggesting that rodent carcinogenicity assays are much less reproducible than previously expected. Based on these findings the authors concluded that it is not only difficult to identify carcinogens in general, but also to identify reliably powerful multi-species and multi-

organ carcinogens. Risk assessment of nongenotoxic carcinogens of importance for human safety from a statistical point of view is even more difficult due to the high spontaneous tumor background in a two-year study in rodents against which positive or negative effects can be missed.

Since drug-induced cancers are rare in human, the results of routine carcinogenicity testing may be misleading as an indication of risk to humans. Therefore the results need careful interpretation. Because of the increasing concerns regarding the relevance and reproducibility of the rodent bioassay, it has been suggested that the bioassay may not be needed in the future to predict carcinogenic risk of pharmaceuticals if sufficient information on the specific agent can be derived from other studies [18,21]. High-throughput screens based on genomic, proteomic, and metabolomic cancer biomarkers are increasingly being used to prioritize chemicals for long-term testing. Based on the major advances in computer technology, chemoinformatics, and predictive toxicology, the accumulated results of rodent carcinogenicity studies in public databases and regulatory files can be used to improve the scientific bases of regulatory and product development decisions. It has been proposed that over time with increased experience and confidence in carcinogenicity predictive software, it may be possible to reduce carcinogenicity testing for compounds that have molecular structures that are highly represented in well-defined carcinogenicity databases [7]. The additional information for risk assessment would include whether the test article is a direct-acting DNA mutagen, induces liver enzymes, causes hyperplasia or toxicity in particular organs, causes cell proliferation, is cytotoxic, immunosuppressive or causes hormonal perturbations [16].

In 2001 the International Programme on Chemical Safety (IPCS) Conceptual Framework for Evaluation (animal) Mode of Action for Chemical Carcinogens provided a framework for a generic approach to the principles commonly used for evaluating mode of action. It outlined a list of elements to be considered in analyzing whether available data support a particular mode of action. In 2006 this "Harmonization Project" completed work to extend the 2001 Framework to address the issue of human relevance. The IPCS Framework for Analyzing the Relevance of a Cancer Mode of Action for Humans was published in a special issue of *Critical Reviews in Toxicology*, Volume 36(10), November–December 2006. Ultimately it will be published as a WHO document. A technical working group focused on cancer hazard identification, operating under the umbrella of the Health and Environmental Sciences Institute (HESI) in Washington, DC, has recently been formed to explore the hypothesis that carcinogenic signals can be identified in short-term studies exclusive of the rodent bioassay. Another group, the Predictive Safety Testing Consortium (PSTC) carcinogenicity working group, is focused on identifying signatures of nongenotoxic carcinogenicity. The PSTC was formed in the United States in 2006 under the auspices of the Critical Path Institute (C-Path) to enable pharmaceutical companies to pool their monetary and data resources to develop biomarkers for preclinical safety testing. A similar initiative is ongoing under the European Commission's Sixth Framework

Programme (FP6). The FP6 is the framework for the EU activity in the field of science, research, and innovation. The main objective of the FP6 is to contribute to the creation of the European Research Area (ERA) by improving integration and coordination of research in Europe. One of seven priority thematic areas is Life Science, Genomics and Biotechnology for Health.

The ultimate question that must be answered during clinical development of a biopharmaceutical is, Does the proposed use of the biopharmaceutical carry a carcinogenic risk for humans in the intended patient population under the conditions of intended clinical use? The answer should consider whether the risk is real or theoretical based on knowledge of the particular product or product class. In selected instances human experience in long-term disease and genetic disorders may reveal the nature and magnitude of the clinical risk, such as the known propensity of patients with certain types of congenital immune disorders to develop lymphomas. The availability of appropriate preclinical models and study designs must also be considered not only to identify the hazard but also to support rationalization of the potential risk. A follow-up question could be asked as to whether animal data would alter the perception of risk. Determining optimal management of the potential risk through patient monitoring and follow-up and/or restrictions in use, as well as providing effective communication of the potential risk through informed consent and ultimately in the product label, are all critical components to the processes of human carcinogenic risk assessment and risk management.

## 2 HUMAN EXPERIENCE

Despite the high percentage of positive rodent carcinogens there are a relatively few pharmaceuticals demonstrated to be carcinogenic in humans [22]. The known human carcinogens have been genotoxic, immunosuppressive, hormonally active, or belong to a class of reactive dermatologicals (e.g., arsenicals) [17]. Unfortunately, it is very difficult to attribute cancer cases to a particular product unless the cancer occurs in a high percentage of individuals, the relationship of the cancer incidence has been studied rigorously (e.g., the associations of PUVA therapy and various skin neoplasms years later), or the product induces a rare form of cancer, as was the case for DES [23].

The time to occurrence of human malignancy varies widely. Most estimates suggest years of exposure are required. For many of the well-defined cancers arising in susceptible hosts, as well as for patients with known carcinogen exposures, the period before development of a malignant tumor is often measured in decades. The complex accumulation of factors influencing the development of neoplasms include viral infections, chronic antigenic stimulation, direct or synergistic carcinogenic effects of certain immunosuppressive agents, sunlight exposure, and other exogenous carcinogenic stimuli in addition to an individual's genetic susceptibility to certain cancers.

Some malignancies occur in association with a wide variety of immune suppression conditions such as inherited diseases of the immune system, acquired immunodeficiency disorders, disorders of immune dysregulation, and iatrogenic immunosuppression (e.g., patients receiving varying degrees of immunosuppressant therapies). These malignancies may occur within a few years. Regardless of how the immunodeficiency is produced, the malignancy is generally associated with an increased risk of certain cancers, namely B cell lymphomas, usually associated with Epstein-Barr virus (EBV), squamous cell carcinomas associated with human papilloma virus (HPV), particularly of the cervix, and Kaposi's sarcoma associated with herpes virus 8 (HHV-8) in patients with AIDS [24]. These tumors are predominantly associated with viral infections that cannot be kept under control because of the immunodeficiency. In such cases it is unlikely that the immunosuppressive therapies themselves are directly carcinogenic; rather, the carcinogenic stimulus is more likely due to induction of immunosuppression that leads to the specific viral-associated tumors [24].

The precise role of immunological surveillance in tumorigenesis is not well defined for the majority of malignancies. The occurrence of a unique spectrum of malignancies in immunosuppressed individuals suggests either that immune surveillance is only important in certain tumors or that the duration needed to see an increased incidence of many more common tumors (e.g., colorectal, breast, lung, or prostate carcinomas) is not reached. Suppression of T cell mediated immunity has, however, been unequivocally associated with an increased incidence of certain malignancies. In patients with profound defects in T cell immunity the time to tumor detection is often shorter than for cancers induced by other mechanisms.

The risk of developing lymphoproliferative disorders following immunosuppressive therapy is also associated with the duration and intensity of immunosuppression [25]. In early postmarketing surveillance following the approval of the first approved therapeutic monoclonal antibody, Orthoclone OKT®3 (muromonab CD3), a murine monoclonal antibody directed against the CD3 antigen on human T cells, the incidence of lymphomas was disproportionately higher in heart and lung transplant recipients when compared to renal transplant recipients, presumably due to the more intensive regimen (dose and duration) in the first two indications. Unfortunately, with transplant patients it is often difficult to evaluate the contribution of any specific agent because many are used in combination regimens for treatment of rejection. See below and Chapter Immunomodulatory Biopharmaceuticals and Risk of Neoplasia for a discussion of concerns with immunodulatory biopharmaceuticals.

## 3 DETERMINING THE NEED FOR PRECLINICAL CARCINOGENICITY ASSESSMENT

The need for carcinogenicity assessment of a biopharmaceutical is not different than that for a conventional pharmaceutical. The need is based on an

intended duration of (i.e., ≥6 months of continual dosing or dosing frequently in an intermittent manner for a chronic or recurrent condition), cause for concern, indication and patient population including the nature of the disease and its effect on life span and overall patient health, route and exposure, and extent of systemic exposure when certain delivery systems may result in prolonged exposures. The need is also established if there is cause for concern based on the product class or if there is evidence of preneoplastic lesions in repeat-dose chronic studies. For pharmaceuticals an additional concern is long-term retention of parent compound or metabolite(s), which may result in local tissue reactions or other pathological responses (See Table 1) [26]. In most cases the rat and/or mouse have been deemed relevant species for assessment of conventional carcinogenic risk of conventional pharmaceuticals based on regulatory precedent.

In the early days of the production of human proteins by recombinant DNA technology using *E. coli* bacteria, there was considerable and reasonable concern about the presence of bacterial impurities that may be carried through the purification processes. Genotoxicity studies were widely used to ensure that these impurities were not mutagenic or clastogenic. However, with the development of improved methods of synthesis and purification, together with highly sophisticated and sensitive analytical techniques for detection of possible impurities, the need for genotoxicity studies has been greatly reduced. Despite this important progress, toxicologists in industry have continued to perform genotoxicity studies and/or the studies have been requested by regulatory agencies even though the data may not have apparent scientific relevance (see Chapter Genetic Toxicology Testing of Biopharmaceuticals). Ironically, regardless of their lack of relevance, the absence of positive findings in the various genotoxicity assays have often been used as "scientific" justification for not needing to assess carcinogenic potential in vivo in animals

### 3.1 Peptide Hormones and Growth Factors

In the case of biopharmaceuticals the particular cause for concern for carcinogenic risk is raised by the action of peptide hormones and growth factors. These may directly promote cellular growth through direct activation of the drug receptor or through other factors involved in growth regulation [27]. Thus for growth factors, where one of the putative mechanisms of concern would be increased proliferation, an evaluation of this endpoint in the appropriate target tissue in in vivo studies in a relevant species is important in order to determine potential concern for tumorigenic risk. Although unidentified, target or nontarget tissues may also be a cause for concern.

### 3.2 Immunomodulatory Biopharmaceuticals

For biopharmaceuticals intentionally developed to target critical immune-signaling molecules, current in vitro and in vivo methods confirm

**TABLE 1 Factors to consider for carcinogenicity testing of pharmaceuticals**

| Factors | Considerations When Rodent Bioassays Performed | Considerations When Rodent Bioassays Not Needed |
|---|---|---|
| Duration and exposure | • Expected clinical use continuous for 6 months<br>• Expected to be used frequently in an intermittent manner in the treatment of chronic or recurrent conditions<br>• For certain delivery systems which may result in prolonged exposures | • Used infrequently or for a short duration of exposure (e.g., anesthetics and radiolabeled imaging agents) if no cause for concern |
| Cause for concern | • Previous demonstration of carcinogenic (oncogenic) potential in the product class that is considered relevant to humans<br>• Structure-activity relationship suggesting carcinogenic risk<br>• Evidence of preneoplastic lesions in repeated-dose toxicity studies<br>• Long-term tissue retention of parent compound or metabolite(s) resulting in local tissue reactions or other pathophysiological responses | |
| Genotoxicity | • Recommended ICH test battery<br>• A single positive result in any assay for genotoxicity does not necessarily mean that the test compound poses a genotoxic hazard to humans (also concern for nongenotoxic carcinogens), although in vivo positive may be considered differently than in vitro positive | • Unequivocally genotoxic compounds, in the absence of other data, that are presumed to be transspecies carcinogens, implying a hazard to humans (chronic toxicity of up to 1 year may be needed if drug is intended to be used chronically to detect early tumorigenic effects) |
| Indication and patient population | • Expected to be completed before application for marketing approval (not needed prior to conduct of large-scale clinical trials unless cause for concern)<br>• For pharmaceuticals that are intended for adjuvant therapy in tumor free patients or for prolonged use in noncancer indications (may be additional conditions for compounds especially designed for use in children) | • May be conducted postapproval if drug is intended for life-threatening or severely debilitating disease, especially where no satisfactory alternative therapy exists<br>• Oncolytic agents intended for treatment of advanced disease (if treatment is successful, there may be later concerns regarding secondary cancers) |

**TABLE 1 Factors to consider for carcinogenicity testing of pharmaceuticals (*Continued*)**

| Factors | Considerations When Rodent Bioassays Performed | Considerations When Rodent Bioassays Not Needed |
|---|---|---|
| Route and exposure | • Route should be the same as the intended clinical route when feasible<br>• Alternative route is acceptable if similar metabolism and systemic as well as target organ exposure (e.g., lung for inhalation agents) can be demonstrated (exposure may be derived from pharmacokinetic data) | |
| Extent of systemic exposure | • Pharmaceuticals applied topically (e.g., dermal and ocular routes of administration) if knowledge is already available from systemic exposure further local testing may not be necessary unless concentration or duration are very prolonged<br>• Cause for concern for photocarcinogenic potential (dermal application generally in mice) | • Pharmaceuticals showing poor systemic exposure from topical route<br>• Pharmaceuticals administered by ocular route unless cause for concern or unless there is significant systemic exposure<br>• For different salts, acids, or bases of the same therapeutic moiety, where prior carcinogenicity studies are available (provided no significant changes in pharmacokinetics, pharmacodynamics, or toxicity) (When changes in exposure and consequent toxicity are noted, additional bridging studies may be used to determine whether additional carcinogenicity studies are needed)<br>• For esters and complex derivatives, similar data would be valuable in assessing the need for an additional carcinogenicity study (determined on a case-by-case basis) |

| | | |
|---|---|---|
| Endogenous peptides and protein substances and their analogues | • Should be considered for endogenous peptides or proteins if indicated by treatment duration, clinical indication, or patient population (providing neutralizing antibodies are not elicited to such an extent in repeated-dose studies as to invalidate the results)<br>— for products where there are significant differences in biological effects to the natural counterpart(s)<br>— for products where modifications lead to significant changes in structure compared to the natural counterpart<br>— for products resulting in humans in a significant increase over the existing local or systemic concentration (i.e., pharmacological levels; concern is increased if principal action is to increase cell proliferation or to prolong cell survival; combination therapy with known carcinogenic stimulus) | • Endogenous substances given essentially as replacement therapy (i.e., physiological levels), particularly where there is previous clinical experience with similar products (e.g., animal insulins, pituitary-derived growth hormone, and calcitonin). |
| Need for additional testing | • Further research may be needed, investigating the mode of action (to confirm the presence or the lack of carcinogenic potential for humans)<br>• Mechanistic studies are useful to evaluate the relevance of tumor findings in animals for human safety | |

*Source*: Adapted ICH S1A, The Need for Long-term Rodent Carcinogenicity Studies of Pharmaceuticals (http://www.fda.gov/cder/guidance/index.htm).

immunosuppressant activity, rather than reveal a clinically useful designation of safe, "pharmacologically immunosuppressive," or toxic "adversely immunosuppressive." Concerns following chronic treatment of immunosuppressive agents include the potential impairment leading to opportunistic infections and/or lymphoproliferative disorders. The pharmacological activity of therapeutic immunosuppressive drugs is thought to make them highly likely to act as tumor promoters/co-carcinogens, even in the absence of genotoxic activity. For example, potent immunosuppressive agents may increase the risk of tumor promotion by impairing the endogenous tumor surveillance mechanism. As far as we know, cellular and humoral immunosuppression probably carry similar risks. Thus it is assumed that this class of drug would represent a cancer hazard in the absence of confirmatory standard bioassay [28] (see Chapter Immunomodulatory Biopharmaceuticals and Risk of Neoplasia).

### 3.3 Cell Lines Used to Produce Biopharmceuticals

The international quality guidance document ICH Q5D requires tumorigenicity testing for products for which live cells cannot be excluded or products that have minimal downstream purification (e.g., conventional live virus vaccines) [29]. The guidance states that tumorigenicity testing of proteins is generally not needed provided that residual host cell DNA limits are established and met. These limits were based on careful consideration of the amount of DNA involved and evidence about the likelihood that it might transmit a viable oncogenic agent. The above-mentioned guidance does not require tumorigenicity testing for cell lines that have been previously shown to be tumorigenic, but the sponsor may want to evaluate the previous results in light of current proposed indications. However, if the cell line is to be licensed as a nontumorigenic cell line, it has to be demonstrated by a negative result in tumorigenicity assays. Newer previously uncharacterized diploid cell lines should be evaluated for their tumorigenic potential by using cells from the master cell bank (MCB). The assessments should include comparison between the cell line under evaluation and a positive (tumorigenic) cell line used as a positive control. Animal models for assessing tumorigenicity include athymic mice (Nu/Nu), suckling mice or hamsters treated with antithymocyte serum, and/or irradiated and thymectomized mice reconstituted (T-B+) with bone marrow of healthy mice [29]. Additional considerations may also be applied to cells from nonmammalian sources used to produce therapeutic proteins.

### 3.4 Cellular-Based Therapies

Tumorigenicity studies are also needed for certain cell therapy products based on the specific product attributes and clinical indications. Somatic cell therapy includes cells from various sources such as adult cells derived from the individual being treated (autologous), cells donated from one human to another (allogeneic), and cells from an animal source and used for human treatment (xenogeneic). Stem cell therapy includes cells derived from a variety of adult

human tissue, fetal tissue, and embryonic tissue. The two characteristics that distinguish stem cells for other types of cells are (1) stem cells can self-renew and (2) stem cells can differentiate into other types of cells. These specific stem cell proliferative properties raise concerns for the potential tumorigenicity of such products. Tumorigenicity studies are thus needed for human derived embryonic stems cells and may also be needed for fetal derived and adult tissue derived multipotent stem cells produced by extended culture. Embryonic stem cells have been shown to produce a benign tumor called a teratoma in immunosuppressed mice. Teratomas typically contain a mixture of many differentiated or partially differentiated cell types.

When performed, the clinically relevant route of administration should be used and appropriate animal species, including appropriate study design and duration, should be justified 6 months up to lifetime observation depending on the particular cell type. Positive controls are important for a number of reasons, including the potential technical errors related to tumor cell implantation, specimen processing, and analysis in order to ensure that findings from tumorigenicity study are interpretable. Importantly, unlike for other pharmaceuticals or biopharmaceuticals, the results of the tumorigenicity study must be available to support initiation of the initial clinical trial (see Chapter Considerations in Design of Preclinical Safety Evaluation Programs to Support Human Cell-Based Therapies).

### 3.5 Gene Therapies

The concern regarding carcinogenicity of gene therapies is based on the theoretical potential for insertional mutagenesis of the vector and/or a direct effect of the expressed transgene. Concerns for potential insertional mutagenesis are also raised with DNA vaccines (see Chapter Preclinical Safety Evaluation of Viral Vaccines).

Reports of gene therapy associated leukemia related to insertional activation of proto-oncogenes by retroviral vectors have occurred and have raised serious concerns in the field of gene therapy [30]. Biodistribution studies are performed preclinically to determine whether there is vector integration and persistence of vector in tissues. Evaluation of tumorigenic potential takes into account vector-specific and patient-specific issues, including whether the vector is integrating or nonintegrating, as well as pediatric versus adult population and whether the disease being treated is serious or life threatening. As with most biopharmaceuticals the standard rodent bioassay is generally not appropriate as rodents may not be susceptible to the viral vector, daily administration of vector is not feasible, and/or host immune response to vector and/or transgene will likely limit toxicity and effects on tumor development. Alternative animal models will likely be more appropriate for specific integrating gene therapy vectors, such as genetically modified mice or neonatal mice including animal models of the intended disease state [31]. Importantly, a long-term follow-up is required of all clinical trial participants receiving gene therapy products in order to properly monitor for long-term adverse effects of gene

therapy treatments and mitigate their potential impact on study participants [32]. (See also Chapters Preclinical Safety Evaluation of Viral Vaccines and Preclinical Safety Evaluation of Biopharmaceuticals.)

## 4 ADDITIONAL CONSIDERATIONS FOR BIOPHARMACEUTICALS

Clinical indications for biopharmaceuticals have evolved from uses as replacement therapies and therapies for life-threatening diseases to uses as lifetime therapies for chronic diseases. Because of their short half-lives, conventional pharmaceuticals used to treat chronic disease are generally administered daily unless specifically designed for long-acting release. Most biopharmaceuticals, on the other hand, are slowly eliminated and thus dosed intermittently. However, in some cases, such as a humanized monoclonal antibody, a single administration may lead to extended duration of exposure prolonging a biological response for days to months. In the extreme, hypothetically, a single administration of a gene therapy to correct a gene defect could result in lifetime reconstitution of the genetic defect and potential cure of the disease.

The international safety guidance ICH S6 offers a discussion about assessing carcinogenic potential, including the variety of approaches that should be considered. The utility of a single rodent species should be considered in those cases where a product is biologically active and nonimmunogenic in rodents and when other studies have not provided sufficient information to allow assessment of carcinogenic potential. In such cases careful consideration should be given to the selection of doses including the rationale for dose selection. The use of a combination of pharmacokinetic and pharmacodynamic endpoints with consideration of comparative receptor characteristics and intended human exposures represents the most scientifically based approach for defining appropriate doses [27].

The standard rodent bioassay is often inappropriate in large part due to species specificity and/or immunogenicity of human proteins in rodents. Most animals develop an immune response to foreign proteins. This cross-species immunogenicity of a human protein in a foreign host is an important component in the challenge of deciding how best to evaluate carcinogenicity in accepted preclinical animal models [33]. The development of neutralizing antibodies in various animal species has generally been the determining factor, limiting the duration of repeat-dose toxicity testing of recombinant derived human proteins due to potential alteration of PK/PD and influence on exposure. In some cases it has been advisable to continue dosing beyond documentation of the first presence of antibodies especially when the antibody response is not correlated with dose. In addition, when antibody responses are also observed in humans, treatment through the antibody response has been considered for repeat-dose animal toxicity assessment to mimic the human dosing regimen. The risk of anaphylaxis and immune complex disease in the test species should be recognized, since occurrence will also limit the feasibility of long-term testing. Antibody development will also be a problem for long-term

repeat-dosing or lifetime exposure. For example, when renal function becomes impaired by age, immune complex deposition could be a more important issue. While nonhuman primates may be a relevant species for repeat-dose toxicity studies with biopharmaceuticals, dosing for a major part of their lifetime (~20 years or more) is impractical and no longer feasible.

The production and use of homologous proteins (also referred to as "analogous" or "surrogate" molecules) are useful for "proof of concept" decision-making and for providing an early understanding of mechanism of action during the development of human proteins. In a limited number of cases they have also been useful for safety assessment, for example, in the assessment of reproductive and chronic toxicity, and in particular, for the safety of humanized monoclonal antibodies [34,35]. In some cases, however, it may not be possible to derive and/or manufacture the homologous protein; for instance, the biology may be different among the various species despite a high sequence homology of the protein or receptor, the mechanism of action of the receptor may differ, or the homologous protein may also be immunogenic following repeat dosing. The challenge when considering chronic dosing with homologous proteins is to ensure their comparability with the human protein product with respect to purity, potency, and stability and also in manufacturing process. For example, the protein should be comparable with respect to production specifications, impurities and contaminants, formulation, stability, as well as ensuring appropriate level of GMP compliance [36]. A case can be made to assess carcinogenic potential with the homologous product if it can be determined that the product and pharmacological activity are comparable. In such cases as the clinical material is not being evaluated, a single rodent species should be acceptable for carcinogenicity assessment.

Products that may have the potential to stimulate growth or induce proliferation or clonal expansion of cell types, in particular, transformed cells, all processes that may eventually lead to neoplasia should be evaluated with respect to receptor expression in various malignant and normal human cells that are relevant to the patient population under study [27]. In such cases normal human cell lines and multiple human cancer cell lines expressing the relevant receptor, as well as primary cells derived from human tumor explants, should be used for in vitro assessment. When in vitro data demonstrate enhanced growth, further studies in relevant in vivo xenograft animal models with receptor expressing tumor cell lines may be needed. In addition incorporation of sensitive indexes of cellular proliferation in long-term repeat-dose toxicity studies may provide useful information.

## 5 PRECLINICAL APPROACHES FOR ASSESSING CARCINOGENIC POTENTIAL

The considerations above have been used to scientifically justify a relevant approach on a case-by-case basis for currently approved and marketed biopharmaceuticals. Table 2 outlines the various approaches currently used to

**TABLE 2 Preclinical approaches for assessing carcinogenic potential**

| Model | Advantages | Limitations |
|---|---|---|
| | *Validated/accepted models* | |
| Two-year rodent bioassay | • Standardized<br>• Large historical database | • Costly<br>• Time-consuming<br>• High animal numbers ($N$ = 50 to 100/sex/group)<br>• Variability exists between species and strains<br>• May not detect weak carcinogens or promoters under standard protocol<br>• Difficulty separating spontaneous geriatric findings from induced changes<br>• Use of homologous protein is needed for species-specific proteins |
| Nonhuman primate bioassay | • Used to assess oral contraceptives | • Duration ≥20 years<br>• Difficulties in interpretation due to low animal numbers<br>• No longer feasible on ethical and welfare grounds |
| Tumorigenicity assessment in immunocompromised rodents | • Most commonly used for assessing tumorigenic potential of cell lines to produce biopharmaceuticals<br>• Used to assess tumorigenic potential of categories of cellular products via intended route of administration<br>• May allow use of clinical product | • Lack of available positive controls for recombinant proteins and monoclonal antibodies<br>• Optimal study design, including adequate number of animals and study duration not established<br>• Lack of knowledge on which genetic defect and which strain of rodent is most appropriate |

| | | |
|---|---|---|
| *Characterized/supportive models* | | |
| Initiation-promotion models in Rodent species<br>• Mouse skin tumor assay<br>• Rat liver focus assay | • Evaluates the two-stage concept of carcinogenesis (initiation and promotion)<br>• Both systems are capable of evaluating tumor promotion<br>• Reduced cost and time<br>• Used to study mechanisms of carcinogenesis | • Both systems are limited to one target tissue<br>• Not validated for biopharmaceuticals<br>• Use of homologous protein is needed for species-specific proteins<br>• Experimental difficulties |
| Syrian hamster embryo (SHE) cell transformation assay (In vitro) (SHE cells are used to evaluate the potential of a wide range of agents to induce morphological transformations. Following carcinogen exposure (1–7 days), SHE cells display a multistage pattern of neoplastic progression similar to that observed in vivo) | • Evaluates potential to induce morphological transformation<br>• Displays multistage pattern of progression similar to in vivo carcinogenesis<br>• Useful for prioritizing genotoxic chemicals for further development<br>• Good concordance with rodent bioassay | • Unable to discriminate between genotoxic and nongenotoxic rodent carcinogens<br>• Not useful in determining which nongentoxic rodent carcinogens are going to be human noncarcinogens<br>• Limited if any data with biopharmaceuticals<br>• Use of homologous protein is needed for species-specific proteins |
| Cell Transformation Assays (in vitro) (e.g., Balb/c 3T3, C3H10T1/2, Bhas 42 cells) | • Reduced cost and time<br>• Screen for nongenotoxic carcinogens | • Experimental difficulties<br>• Variation in sensitivity among laboratories<br>• Refinement in protocol needed regarding test concentrations and re-examination when equivocal data are obtained<br>• Limited if any data with biopharmaceuticals |

**TABLE 2 Preclinical approaches for assessing carcinogenic potential (*Continued*)**

| Model | Advantages | Limitations |
|---|---|---|
| *Alternative models (transgenic animals) "short term assays"/supportive models (activity often limited to one or a few tissues)* | | |
| −p53 +/− (expresses heterozygous inactivation of the p53 tumor suppressor gene, which is critical to cell cycle control and DNA repair) | • Reduced cost<br>• Reduced time and number of animals<br>• Semivalidated (biopharmaceuticals were not included in industry sponsored validation efforts) | • Defined inbred line and may be unlike random outbred species<br>• Not suitable for evaluating nongenotoxic carcinogens<br>• Better standardization of pathology evaluations is needed (e.g., distinguishing hyperplasia from adenoma)<br>• Equivocal responses occur<br>• Doubtful relevance for growth factors or immunomodulatory/immunosuppressive proteins<br>• Limited data available with biopharmaceuticals<br>• May require rodent homologue |
| Tg-AC (skin painting model; expresses a mutation of the v-ras proto-oncogene, which has been demonstrated to detect human carcinogens) | • Detects both genotoxic and nongenotoxic carcinogens<br>• Semivalidated (biopharmaceuticals were not included in industry sponsored validation efforts) | • Defined inbred line and may be unlike random outbred species<br>• No historical use of IV or SC route of administration (primarily topical and oral)<br>• Limitations with respect to extrapolation of dose and organ specificity<br>• Better standardization of pathology evaluations is needed (e.g., distinguishing hyperplasia from benign neoplasms)<br>• Equivocal responses occur<br>• Limited if any data available with biopharmaceuticals<br>• May require rodent homologue |

| | | |
|---|---|---|
| Tg-Hras2 (carries human cH-ras proto-oncogenes driven by their own promoter region, which has been demonstrated to detect human carcinogens) | • Reduced cost<br>• Reduced time and number of animals<br>• Semivalidated<br>• Detects spectrum of weakly to strongly genotoxic compounds<br>• Appears to have greater sensitivity than nontransgenic mice to compounds that are human carcinogens but are Ames negative<br>• Has be used to assess recombinant human growth factor (rhKFG) (biopharmaceuticals were not included in industry sponsored validation efforts) | • Defined inbred line and may be unlike random outbred species<br>• Not suitable for evaluating nongenotoxic carcinogens<br>• Better standardization of pathology evaluations is needed<br>• Dependent on specific mechanism(s)<br>• May not be susceptible to relevant tumor type<br>• Equivocal responses occur<br>• Limited historical database on background incidence<br>• Limited if any data available with biopharmaceuticals<br>• May require rodent homologue |
| XPA –/– (deletion across exons 3 and 4 of both XPA alleles, rendering cells totally defective in nucleotide excision DNA repair; animals comparable to human xeroderma pigmentosum) | • Reduced cost<br>• Reduced time and number of animals<br>• Semivalidated<br>• Sensitive to genotoxic carcinogens that are reactive to the nucleotide excision repair pathway | • Defined inbred line and may be unlike random outbred species<br>• Not suitable for evaluating nongenotoxic carcinogens<br>• Better standardization of pathology evaluations is needed (e.g., distinguishing hyperplasia from adenoma)<br>• Dependent on specific mechanism(s)<br>• May not be susceptible to relevant tumor type<br>• Equivocal responses occur<br>• Limited historical database on background incidence<br>• Limited if any data available with biopharmaceuticals<br>• May require rodent homologue |

**TABLE 2 Preclinical approaches for assessing carcinogenic potential (*Continued*)**

| Model | Advantages | Limitations |
|---|---|---|
| Eμ-pim-1 (expresses the pim-1 proto-oncogene, which has been shown to increase the susceptibility to T and B cell lymphomas) | • Sensitive assay for carcinogens that target the lymphoid system | • Defined inbred line and may be unlike random outbred species<br>• Clear results only for those carcinogens that are genotoxic<br>• Limited validation<br>• Better standardization of pathology evaluations is needed (e.g., distinguishing hyperplasia from adenoma)<br>• Equivocal responses occur<br>• No data available for biopharmaceuticals<br>• May require rodent homologue |
| | *Alternative models (other)/ supportive models* | |
| Neonatal rodent (utilizes the increased tumorigenic sensitivity of the new-born rodent. Model consists of a 28-day range-finding experiment and a 1 year-long tumorigenicity study. Test article is administered on days 8 and 15. | • Sensitive biological assay<br>• Historical database (>40 years)<br>• Detects moderately to strongly genotoxic carcinogens<br>• Distinguishes between definite genotoxicity and weak to nongenotoxic potential<br>• Multiple routes of administration have been used<br>• May be able to use clinical product | • Uncertainty about its range of predictiveness of nongenotoxic carcinogens<br>• Uncertainty about its range of predictiveness for genotoxic carcinogens<br>• Limited historical database on background incidence<br>• Inadequate data are available with biopharmaceuticals |
| | *Experimental/ developmental models* | |
| Spontaneous/chemically induced tumors (increase in number, change in tumor types or earlier appearance of tumors) | • Potential opportunity to address potency<br>• Decreased concern if no findings<br>• May provide some information with respect to risk and level of concern | • Clinical product may be immunogenic<br>• Limited historical database on background incidence<br>• Variability in tumor induction rate<br>• Limited data available with biopharmaceuticals<br>• May require rodent homologue |

| | | |
|---|---|---|
| Tumor host resistance in rodents (e.g., B16F10 melanoma) | • Used to assess immunotoxicity potential of test articles<br>• May provide some information with respect to risk and level of concern<br>• Has been used to assess impact of immunosuppressive biopharmaceuticals on host resistance to neoplasia | • Limited to disease promotion/progression<br>• May require rodent homologue |
| Tumor xenograft in immunosuppressed mice | • Routinely performed for assessing efficacy of cancer drugs<br>• May provide some information with respect to risk and level of concern<br>• Has been used to evaluate the impact of growth factors on tumor growth and metastasis where the human tumor cells express the receptor for the growth factor | • Some difficulties in transferring standardized assays used to determine activity to new assays for safety assessment<br>• Variability in models (nude, SCID, SCID-Beige, etc.)<br>• Immunosuppression can either increase or decrease tumor growth |
| In vitro/ex vivo mitogenicity in human cell lines or animal cell lines (expressing receptor or transfected with human receptor) | • Test system sensitive for mitogenic and/or metabolic effects of test article<br>• Provide data in nontumor as well as tumor tissues (containing aberrant receptors)<br>• May predict in vivo effects under certain circumstances<br>• Increased mitogenicity justifies request of in vivo data<br>•"doable"<br>• Can provide some information as to relative potency—cause for concern | • Limited data with respect to correlation of findings with in vivo effects<br>• May be difficult to develop an in vivo model to validate findings |

**TABLE 2 Preclinical approaches for assessing carcinogenic potential (*Continued*)**

| Model | Advantages | Limitations |
|---|---|---|
| Transgenic knockout (chronic inhibition of target) | • Useful for assessing worst-case scenario | • May overestimate response—effects may differ from those of therapeutic agent<br>• Difficult to extrapolate a positive response<br>• Test article is not administered therefore no information with respect to dose and exposure |
| Transgenic knock-in (overexpression of protein) | • Useful for assessing worst-case scenario | • May overestimate response—effects may differ from those of therapeutic agent<br>• Difficult to extrapolate a positive response<br>• Test article is not administered therefore no information with respect to dose and exposure |
| Transgenic "humanized" mouse (human receptor) | • Allows testing of the clinical product | • May overestimate response<br>• Difficult to extrapolate a positive response<br>• May not have appropriate physiological regulation of gene expression<br>• May not have appropriate physiological response to test article |
| Neonatal tolerance induction in rodents | • Option of dosing animals very early in life to produce animals tolerant to subsequent chronic administration of human proteins (potential to conduct rodent bioassay)<br>• Use of clinical product | • No data on positive controls<br>• Limited if any data on biopharmaceuticals |
| Chronic toxicity with proliferation endpoints | • Opportunity to determine carcinogenic potential earlier in development using sensitive biomarkers<br>• May provide some information with respect to risk and level of concern | • Limited experience<br>• Increased cell proliferation has been identified as a risk factor for carcinogenesis but is not a defining endpoint<br>• May not be sufficient to use this data for human risk assessment in absence of tumor endpoint |

assess the carcinogenic potential of pharmaceuticals and highlights the potential advantages and limitations of the specific models for biopharmaceuticals. Additional discussion on the various approaches used to assess the carcinogenic potential of pharmaceuticals is provided below.

### 5.1 Rodent Bioassay

As previously discussed, the rodent bioassay has recently been under scrutiny for its relevance for identifying the carcinogenic hazard of pharmaceutical products. The assays are the most costly (up to 800–1000 animals/test (500/species); approximately 2 million US dollars each for a rat and mouse study) and time-consuming (2 years of treatment; an additional 1–2 years for histopathological analysis and report writing) of the preclinical regulatory requirements [7]. It is especially important that the results provide value for predicting human risk. Recently a retrospective evaluation was made on 60 systemic two-year studies of pharmaceuticals reviewed in the Center for Drug Evaluation and Research (CDER) FDA between January 2002 and December 2004. Based on an evaluation of the CDER/FDA database, the authors concluded that additional information and data, beyond what are already being collected, are needed to better predict or obviate the need for two-year carcinogenicity studies. The authors further acknowledged that many of the mechanisms of carcinogenesis found with pharmaceuticals may not be amenable to early prediction at this time as many of the mechanisms identified to date for pharmaceuticals appear to be nongenotoxic in nature and thus may require the prolonged treatment to be expressed [37,38]. However, these mechanisms may not be associated with human tumor risk. Nevertheless, based on this initial review, the FDA does not currently support the idea that short-term studies accurately predict the potential neoplastic findings in long-term assays of pharmaceuticals [16].

### 5.2 Short-term Carcinogenicity Assays

As the International Conference on Harmonization (ICH) Expert Working Group on Safety has conducted discussions involving how best to assess potential human cancer risk of pharmaceuticals, important questions were raised regarding the added value of a second rodent species. Following review of various databases created and evaluated by the participants, it was suggested that under certain circumstances data from alternative short-term assays may prove of equal or greater value to the bioassay in a second rodent species, and those assays were presented as possible alternatives in the guidance document [39,40].

In 1996 a collaborative effort was subsequently initiated under the auspices of the Health and Environmental Sciences Institute (HESI) branch of the International Life Sciences Institute (ILSI) with the specific purpose of facilitating a focused systematic evaluation of several of the new alternative models proposed within the ICH guidance. Participation in this research collaboration was global in scope encompassing Europe, Japan, and the United States and

included representatives from academe, industry, and governmental agencies. The effort was not undertaken to determine if the results of alternative assays correlated with the results obtained in the rodent bioassay, but rather, whether the data from these assay could add value to the process of human risk assessment. The research involved input from over 50 industrial, governmental (United States, Denmark, Netherlands, and Japan), and academic laboratories representing a financial commitment of approximately 33 million US dollars [41]. The focus of the ICH discussions was pharmaceuticals. Importantly, no biopharmaceuticals were included in the collaborative study. Therefore the validity of extrapolating the results to biopharmaceuticals is currently not known [42].

The research program was overseen by a Steering Committee of scientists drawn from academe and pharmaceutical companies. The models under consideration included the p53+/–, ras H2+/–, Tg.AC, Xpa–/–, Xpa–/–/p53+/– transgenic animal models, the neonatal mouse model, and the in vitro Syrian hamster embryo (SHE) assay. The protocols used were based on the existing knowledge of each model. Positive control chemicals were used to demonstrate that each testing laboratory could undertake and report a positive assay for the model under test. A number of articles have been published commenting on the utility of the alternative short-term bioassays for assessing the carcinogenic potential of pharmaceuticals [43–52].

The overall conclusion of the ILSI effort was that none of the models used was suitable as a definitive determinant of potential cancer risk, but they could serve as useful hazard identification models similar to the rodent bioassay. For those chemical compounds found positive in the rodent bioassay yet not considered carcinogenic in humans, there were fewer positive results in transgenic mouse models; that is, the models exhibit a better correlation with actual human data than the rodent bioassay. The lack of correlation between organ specificity and potential human carcinogenicity limits the information that can be gained with respect to mechanism of toxicity, especially for pharmaceuticals that are nongenotoxic. Additionally, because the positive responses were frequently only seen at the highest doses, the models are more limited for quantitative risk assessment and determination of a dose response. Since the neonatal mouse appeared only to detect chemicals that were moderate to strongly genotoxic carcinogens, the model was considered to be of limited usefulness. The data also suggested that the SHE assay was unable to discriminate between genotoxic and nongenotoxic chemicals or between rodent carcinogens versus noncarcinogens and thus would likely be less useful for predicting human carcinogenicity versus noncarcinogenicity.

The report concluded that a reasonable weight of evidence on the evaluation of potential risk to humans could be achieved if information from the short-term bioassay is used in conjunction with information from other sources such as the rat bioassay, the Ames assay, the chemistry of the test article, and the repeat-dose studies evaluating intermediate markers of malignancy from knowledge of the pharmacological activities of the test substance and from

additional mechanistic research [41]. Consistent with the earlier recommendation [28], the usefulness of animal models in evaluating strongly immunosuppressive chemicals was questioned for providing any added value due to the expected result of viral tumors associated with the induction of immunosuppression. Further research was recommended to identify models with greater relevance to mechanisms of toxicity in humans as well as short-term in vivo models to evaluate nongenotoxic carcinogenesis and tumor promoters. These research efforts may provide models for identifying carcinogenic hazard that are more relevant for biopharmaceuticals.

### 5.3 Proliferation Indexes from Repeat-Dose Chronic Toxicity Studies

The ICH S6 guidance states that incorporation of sensitive indices of cellular proliferation in long-term repeated-dose toxicity studies may provide useful information [27]. Tumor cell proliferation is a key characteristic of stepwise neoplastic promotion and progression and is often detectable prior to the appearance of overt neoplastic changes. Detectable histological features related to tumor cell proliferation include Ki67expression and increased mitotic rate and growth phase of affected cells, sometimes with consequential changes in the morphology of affected tissues. The Ki67 protein is expressed in all phases of the cell cycle except $G_0$. It has therefore been suggested as a more sensitive biomarker for cellular proliferation than visualization of mitoses [53]. Immunochemical staining for Ki67 or proliferating cell nuclear antigen (PCNA) can be performed on formalin fixed tissues. However, these staining methodologies, which detect significant increases in proliferation, may not be able to assess small changes in proliferation as the rate of cell death could still have an effect on tumor formation, especially in tissues that have a low basal proliferative rate. While an increase in proliferation rate can result in an increase in mutation rate, the death of a cell sustaining a mutation will prevent progression to neoplasia. Thus a large increase in proliferation rate may sometime be of little or no consequence when the rate of cell death is equal or higher, especially when the cells affected undergo terminal differentiation. Conversely, no change in proliferation but a reduced rate of cell death can also have a significant effect on tumor formation. Thus, while information on proliferation potential can be of some value with respect to hazard identification, the correlation between proliferation and a tumor endpoint is not inevitable and needs to be evaluated in each case. It is difficult to prospectively evaluate all potential targets. This is difficult to do even when the target tissue is identified with a dedicated evaluation.

### 5.4 Tumor Host-Resistance Models

Host-resistance assays offer a way to determine the relative significance of a compound-related decrease in immune function, commonly one of such

magnitude as to be associated with a change in the histological appearances of the lymphoid [54]. Implanted tumor cell host-resistance models have also been utilized in hazard identification and risk assessment in immunotoxicology [55]. Immunomodulators can be assessed for their potential to suppress host defense against experimental metastases and growth factors, and other agents can be assessed for their potential to directly accelerate the tumorigenic response via an immunological or other process. At the very least, such models may be able to demonstrate the relative potency of the compound in showing whether the compound is an immunomodulator or a strong immunosuppressant compound. The results obtained would then be of value in carcinogenic risk communication. Because animal host-resistance models can be comparable to and may reflect human disease, their greatest utility is in providing a perspective on how compound-related changes in immune function impact on host resistance to disease [54].

### 5.5 Transgenic Animals

In addition to spontaneous mutant strains of mice and rats, transgenic mice (knock-ins and knockouts) are frequently used as models to understand and confirm the biological activity of biopharmaceuticals. They have also been used to assess aspects of toxicity [56]. Humanized mice have been developed to express target cells with the human receptor to determine binding and its consequences of an agonist or antagonist. In such cases the clinical product binds to the intended receptor and can therefore be used to assess toxicity. Knock-ins have also been used to model a worst-case scenario of overexpression of a protein, for example, of an endogenous growth factor. Knockouts have been used to model a possible worst-case scenario of blocking a receptor, for example, a monoclonal antibody directed against a specific receptor. In these cases the clinical product is not evaluated, but the consequences of the extreme exaggerated pharmacological response of knocking in or knocking out a key signaling molecule that results in subsequent tumor formation may suggest a higher risk and subsequently require additional characterization. A negative response can also be used to better characterize and communicate potential risk.

In all cases where transgenic mice are used it is important to understand the limitations of the specific model in the context of the pathology demonstrated in order to optimize extrapolation of the results. Model attributes that may influence interpretation include epitope distribution, density, localization/compartmentalization, turnover, expression, function, regulation, comparability of signal transduction pathways/regulation, and natural life history. Divergent and even discrepant experimental results have been obtained in transgenic mouse models, depending on the transgene and the promoter used due to variation in promoter regulation and expression phenotype. The background genotype of the strain used may also be important, and variation in

the sites of insertion of the transgene into the host genome may be difficult to control [42].

### 5.6 In vitro Cell Proliferation

Increased cell proliferation has been identified as a risk factor for carcinogenesis; therefore in vitro proliferation assays are generally useful. An understanding of receptor expression in normal and malignant cell lines is an important component to hazard identification. Mechanisms whereby cell growth might become disordered include (1) abnormalities of growth factor production, (2) abnormalities of growth factor receptors, (3) disturbance of postreceptor signal transmission and normal control of activation and reduced production of a growth inhibitory factor, and (4) sensitivity to such factors.

The risk of increased stimulation by a growth factor is generally decreased if the growth factor acts on committed cells with a limited life span and if the stimulated cells readily undergo terminal differentiation. The use of a variety of in vitro evaluations, including in vitro cell proliferation, proved to be important in characterizing the level of concern during the re-evaluation of the safety and dosing of erythropoiesis-stimulating agents (ESAs) in cancer patients with respect to increase mortality and/or tumor promotion (see Table 3) [58].

**TABLE 3 Preclinical evidence of carcinogenic risk with erythropoiesis stimulating agents**

| Finding | Comment |
|---|---|
| Epo receptor (EpoR) is not an oncogene. The EpoR gene is not significantly amplified or over expressed in solid tumors and over expression of constitutively activated mutant forms of EpoR does not transform cells | If Epo-induced signaling could drive proliferation of cancer cells this would increase cause for concern |
| EpoR hyperactivating mutations result in polycythemia and are not a feature of malignancy. In addition in clinical conditions in which Epo is overexpressed (e.g., Chuvash polycythemia) or in which EpoR signaling is not controlled (EpoR truncations), polycythemia results but with no increase in tumors | The fact that EPO overexpression in humans is not associated with malignancy is reassuring |
| EpoR gene is transcribed in most tissues and cell lines at low to moderate levels. Levels of EpoR are rarely elevated in tumors and cell lines above that observed in the normal tissue of tumor origin | Molecules that deliver proliferative signals are frequently expressed at high levels by at least some tumors |

**TABLE 3 Preclinical evidence of carcinogenic risk with erythropoiesis stimulating agents (*Continued*)**

| Finding | Comment |
|---|---|
| EpoR protein synthesis does not necessarily correlate with cell surface expression or signaling of the EpoR Less than 1% of EpoR normally gets to the surface of the cell because of inefficient processing, protein degradation, requirements for limiting accessory molecules for trafficking to the surface, requirements for limiting accessory molecules for intracellular signaling, and short cell-surface half-life | In addition to low-level expression of EpoR on the cell surface, EPO does not bind to tissues other than those expressing EpoR |
| Studies of the direct role of Epo–EpoR in signaling, proliferation, migration, and survival of cancer cells have not yielded conclusive results | Effects of such treatments have been in models despite the fact that most have used levels of rHuEpo (>10 U/ml) that are unattainable in patients |
| All rodent tumor models (23 independent studies) have demonstrated that ESAs do not enhance tumor growth | ESAs have actually shown increase sensitivity of tumor cells to radiation and chemotherapy |
| ESAs do not mediated any consistent adverse effect on tumor angiogenesis in rodent tumor models | Increased tumor oxygenation actually reduces hypoxia-regulated VEGF levels and consequently tumor angiogenesis |
| Data do not support a meaningful effect of ESAs on mobilization of endothelial progenitor cells | Consistent with lack of an affect on tumor vascularization |
| Proliferative response to nonhematological cells is not altered in vitro or in vivo | Reassurance for lack of off-target effect |
| No tumorigenic or unexpected mitogenic response in any tissue type has been noted in 6-month chronic toxicity studies | In vivo support of in vitro data |

*Source*: Adapted 2007 ODAC Meeting, Information Package Darbepoetin alfa (BLA#103951) and Epoetin alfa (BLA#103234)

## 6 CURRENT EXPERIENCE IN CARCINOGENICITY ASSESSMENTS OF BIOPHARMACEUTICALS

Examples of product class carcinogenicity hazard identifications and assessments and ultimate risk communications for biopharmaceuticals approved in the United States for chronic use or based on potential cause for concern are provided in Table 4a (products without carcinogenicity assessment) and Table 4b (products with carcinogenicity assessment). The data are derived from publicly available regulatory assessments and product labels. Specific examples are discussed below.

### 6.1 Immunomodulators

A variety of immumodulatory biopharmaceuticals have been approved spanning various degrees of immune modulation including intended marked immune suppression. A specific discussion on assessing carcinogenic risk of immunomodulatory biopharmaceuticals is presented in Chapter Immunomodulatory Biopharmaceuticals and Risk of Neoplasia. It is of interest to review the individual product labels for their ultimate communication of carcinogenic risk. This important communication varies in a number of respects: for each product as well as products within a class, whether or not preclinical studies were performed, and the impact of these studies with specific findings on the label.

***Abatacept*** Abatacept is a T cell activation inhibitor. It is a fusion protein composed of the extracellular domain of T lymphocyte associated antigen 4 linked to the portion of IgG (human heavy chain fragment). T cell activation is implicated in the pathogenesis of rheumatoid arthritis. Abatacept is indicated for treating RA patients. The mouse carcinogenicity study for abatacept (CTLA4Ig) was conducted because of concerns related to prolonged immunosuppression. There were no "preneoplastic" lesions in the 6-month chronic toxicity studies. The treated mice in the carcinogenicity study developed mammary tumors and lymphomas (increased incidence after 6 months) at a higher incidence than controls, and a no effect dose level could be determined. The tumors were shown likely to be related to tumorigenic mouse retroviruses (MTV and MTLV). A chronic 12-month toxicity study in monkeys, however, did not replicate the effect, even though monkeys in the study had been previously exposed to an endogenous lymphocryptovirus. These viruses share a tropism for B lymphocytes and have a propensity to oncogenicity. A significant increase in the incidence of lymphoma, mammary tumors, or other tumors was not observed in the clinical studies. It could be argued then that the mouse was more sensitive or that the mouse confirmed the anticipated risk, namely prolonged immune suppression, which results in an increased incidence of immune suppression-related tumors under certain circumstances.

The human experience was not unlike that of the anti-TNFs with respect to an approximate 3- to 3.5-fold higher rate of lymphoma in the intended population (patients with RA) who are at higher risk for the development of lymphoma or lung cancer, particularly those with highly active disease. Similar is the statement that the role of the particular product and the development of human malignancies is not known or fully understood. There is also the added precaution to lessen the putative risk in patients to avoid use with other immunosuppressive agents.

***Anakinra*** Anakinra is likewise indicated for the treatment of patients with rheumatoid arthritis. Anakinra is the recombinant form of human IL-1 receptor antagonist (IL-1ra) and is identical to the naturally occurring, nonglycosylated form of the protein, with an additional N-terminal methionine residue. Binding of IL-1ra to the IL-1R1 receptor does not initiate IL-1 mediated cell signaling,

**TABLE 4a Examples of communication of human carcinogenic risk without preclinical assessment of carcinogenic risk**

| Biopharmaceutical | Class | Mechanism of Action | Preclinical Studies | Human Experience/Product Label |
|---|---|---|---|---|
| | | | *Immunomodulators* | |
| Etanercept<br>Enbrel® | • Recombinant DNA-derived dimeric fusion protein (anti-TNF)<br>• Extracellular ligand binding portion of the human tumor necrosis f actor receptor linked to the Fc portion of $IgG_1$ | • Binds tumor necrosis factor α (TNF-α) and blocks interaction with cell surface tumor necrosis receptors | • No rodent bioassay<br>• Not genotoxic—in vitro and in vivo mutagenesis studies | • In the control and open-labeled portions of the clinical trials a threefold higher incidence of lymphomas was observed compared to the general population. While patients with RA or psoriasis, particularly those with active disease may be at high risk (up to sevenfold) for the development of lymphoma, the potential role of TNF-blocking therapy in the development of malignancies is not known<br>*Warnings*<br>• Malignancies<br>• Use of Enbrel in patients receiving concurrent cyclophosphamide therapy is not recommended |
| Infliximab<br>Remicade® | • Chimeric $IgG_{1\kappa}$ monoclonal antibody (anti-TNF) | • Binds TNF-α | • No rodent bioassay<br>• Not genotoxic (Ames, in vivo MN)<br>• No findings in 6-month study with murine homologue (up to 8× human dose)<br>• Significance to human risk unknown | • In the controlled portions of clinical trials of some TNF-blocking agents including Remicade, more malignancies (excluding lymphoma and nonmelanoma skin cancers) have been observed in patients receiving the TNF-blocker compared with control patients<br>• The potential role of TNF-blocking therapy in the development of malignancies is not known. Rates in clinical trials of Remicade cannot be compared to rates in clinical trials of other TNF-blockers and may not predict rates observed in a broader patient population<br>• Patients with Chron's disease, rheumatoid arthritis, or plaque psoriasis, particularly patients with highly active disease and/or chronic exposure to immunosuppressant therapies, may be at a higher risk (up to several fold) than the general population of lymphoma, even in the absence of TNF-blocking therapy<br>• Caution should be exercised in considering Remicade treatment in patients with a history of malignancy or in continuing treatment in patients who develop malignancy while receiving Remicade |

| | | | | |
|---|---|---|---|---|
| | | | | *Black box warning*<br>• Rare postmarketing cases of hepatosplenic T cell lymphoma reported. This rare type of T cell lymphoma has a very aggressive disease course and is usually fatal within 2 years of diagnosis<br>• All lymphomas occurred in patients on concomitant treatment with azothioprine or 6-mercaptopurine<br>• Causal relationship to hematosplenic T cell lymphoma remains unclear<br>• Malignancies including non-Hodgkin's lymphoma and Hodgkin's disease have been reported in patients following approval |
| Adalimumab<br>Humira® | • Fully human $IgG_1$ monoclonal antibody (anti-TNF) | • Neutralizes the biological activity of TNF-α by binding with high affinity to the soluble transmembrane form of TNFα and inhibits binding of TNFα with its receptors | • No rodent bioassay<br>• Not genotoxic (Ames or in vivo MN) | *Adverse reactions*<br>• Neoplasia: adenoma, carcinomas such as breast, gastrointestinal, skin, urogenital and others; lymphoma and melanoma<br>*Warnings and precautions*<br>• In the controlled portions of clinical trials of some TNF-blocking agents, including Humira, more cases of malignancy have been observed among patients receiving TNF-blockers, compared to control patients<br>• Lymphoma rates were observed at approximately 3.5-fold higher than the expected general population<br>• Rates in clinical trials for Humira cannot be compared to rates of clinical trials with other TNF-blockers and may not predict the rates observed in a broader patient population<br>• Patients with rheumatoid arthritis, particularly those with higher active disease, are at a higher risk for the development of lymphoma |

**TABLE 4a Examples of communication of human carcinogenic risk without preclinical assessment of carcinogenic risk (*Continued*)**

| Biopharmaceutical | Class | Mechanism of Action | Preclinical Studies | Human Experience/Product Label |
|---|---|---|---|---|
| Daclizumab Zenapax® | • Humanized $IgG_1$ monoclonal antibody (anti-CD25) | • Binds specifically to the alpha subunit of IL-2 receptor (TAC or CD 25)<br>• Inhibits IL-2 mediated activation of lymphocytes, a critical pathway in the cellular immune response involved with allograft rejection | • No rodent bioassay<br>• Not genotoxic (Ames or in vitro chromosomal aberrations assay (V79)) | • Addition of Zenapax did not increase the number of posttransplant lymphomas up to 3 years posttransplant<br>*Warnings*<br>• While the incidence of lymphoproliferative disorders in the limited clinical trial experience was no higher in patients treated with Zenapax compared with placebo-treated patients, patients with immunosuppressive therapy are at higher risk for developing lymphoproliferative disorders and should be monitored accordingly |
| Basiliximab Simulect® | • Recombinant chimeric (murine/human) monoclonal antibody ($IgG_{1\kappa}$) (anti-CD25) | • Binds specifically to an blocking interleukin-2 receptor α-chain (CD25) on surface of activated T cells<br>• Functions as an IL-2 receptor antagonist | • No rodent bioassay<br>• Not genotoxic (Ames, V79 Chinese hamster cells) | • Overall incidence of malignancies among all patients in the controlled studies was not significantly different between Simulect and placebo-treated groups<br>*Warnings*<br>• While the incidence of lymphoproliferative disorders was not higher n the SIMULECT treated patients than in placebo-treated patients, patients on immunosuppressive therapies are at increased risk for developing these complications and should be monitored accordingly |
| Murumonab-CD3 Orthoclone OKT3® | • Murine monoclonal antibody to human CD3 T cells (anti-CD3) | • Early activation of T cells and premature release of cytokines (e.g., IL-6 and IL-10) followed by blocking of T cell function | • No rodent bioassay | *Warnings*<br>• Serious and sometimes fatal neoplasias in patients<br>• High risk of EBV related B cell neoplasms, skin cancer, melanoma, leukemia, breast cancer, adenocarcinoma<br>• No apparent increase with drug combinations<br>• Monitoring required |

| | | | | |
|---|---|---|---|---|
| | | *Peptide hormones* | | |
| Somatropin<br>Genotropin® | Recombinant somatropin (rDNA origin) | • Identical to human growth hormone of pituitary origin<br>• Stimulates linear growth in patients who lack endogenous growth hormone also normalizes concentrations of IGF-1(insulin-like growth factor-1/ somatamedin C) | • No rodent bioassay<br>• Not genotoxic (Ames, gene mutations in mammalian cells grown in vitro (mouse L5178Y cells), in vivo chromosomal damage (BM cells in rats) | • Leukemia has been reported in a small number of pediatric patients treated with growth hormone, including growth hormone of pituitary origin and recombinant somatropin. The relationship, if any, between leukemia and growth hormone therapy is uncertain<br>*Contraindications*<br>• In general, somatropin is contraindicated in the presence of active malignancy. Any preexisting malignancy should be inactive and its treatment complete prior to instituting therapy with somatropin. Somatropin should be discontinued if there is evidence of recurrent activity. Since growth hormone deficiency may be an early sign of the presence of a pituitary tumor (or, rarely, other brain tumors), the presence of such tumors should be ruled out prior to initiation of treatment. Somatropin should not be used in patients with any evidence of progression or recurrence of an underlying intracranial tumor<br>*Precautions*<br>• Neoplasm has been reported in patients treated with somatropin after their first neoplasm. Intracranial tumors, in particular meingiomas, in patients treated with radiation to the head for their first neoplasm were the most common of these second neoplasms. In adults, it is unknown whether there is any relationship between somatropin replacement therapy and CNS tumor recurrence<br>• Patients should be monitored carefully for any malignant transformation of skin lesions |

**TABLE 4a Examples of communication of human carcinogenic risk without preclinical assessment of carcinogenic risk (*Continued*)**

| Biopharmaceutical | Class | Mechanism of Action | Preclinical Studies | Human Experience/Product Label |
|---|---|---|---|---|
| Somatropin Humatrope® | Recombinant somatropin (rDNA origin) | • Identical to human growth hormone of pituitary origin<br>• Stimulates linear growth in patients who lack endogenous growth hormone also normalizes concentrations of IGF-1(insulin-like growth factor-1/ somatamedin C) | • No rodent bioassay<br>• No evidence of Humatrope-induced mutagenicity | • Leukemia has been reported in a small number of pediatric patients who have been treated with growth hormone, including growth hormone of pituitary origin as well as recombinant DNA origin (somatrem and somatropin). The relationship, if any, between leukemia and growth hormone therapy is unknown<br>• There are also reports of association between acromegaly and colon cancer (not in label)<br>*Contraindications*<br>• In general, somatropin is contraindicated in the presence of active malignancy. Any preexisting malignancy should be inactive and its treatment complete prior to instituting therapy with somaptorpin. Somatropin should be discontinued if there is evidence of recurrent activity. Since growth hormone deficiency may be an early sign of the presence of a pituitary tumor (or, rarely, other brain tumors), the presence of such tumors should be ruled out prior to initiation of treatment. Somatropin should not be used in patients with any evidence of progression or recurrence of any underlying intracranial tumor |

| | | | | |
|---|---|---|---|---|
| Somatropin Nutropin® | Recombinant somatropin (rDNA origin) | • Identical to human growth hormone of pituitary origin<br>• Stimulates linear growth in patients who lack endogenous growth hormone also normalizes concentrations of IGF-1(insulin-like growth factor-1/ somatamedin C) | • No rodent assay<br>• No genotoxicity studies | *Precautions*<br>• Patients with preexisting tumors of growth hormone deficiency secondary to an intracranial lesion should be examined routinely for progression or recurrence of the underlying disease process. In pediatric patients, clinical literature has revealed no relationship between somatropin replacement therapy and central nervous system (CNS) tumor recurrence or new extracranial tumors. However, in childhood cancer survivors, an increased risk of a second neoplasm has been reported in patients treated with somatropin after their first neoplasm. Intracranial tumors, in particular meingiomas, in patients treated with radiation to the head for their first neoplasm, were most common of these second neoplasms. In adults, it is unknown whether there is any relationship between somatropin replacement therapy and CNS tumor recurrence<br>• Patients should be monitored carefully for any malignant transformation of skin lesions<br>• Leukemia has ben reported in a small number of GHD patients treated with GH. It is uncertain whether this increased risk is related to the pathology of GH deficiency itself, GH therapy, or other associated treatments such as radiation therapy for intracranial tumors. On the basis of current evidence, experts cannot conclude that GH therapy is responsible for these occurrences. The risk to GHD, CRI, or Turner syndrome patients, if any, remains to be established<br>• There are also reports of association between acromegaly and colon cancer (not in label) |

**TABLE 4a Examples of communication of human carcinogenic risk without preclinical assessment of carcinogenic risk (*Continued*)**

| Biopharmaceutical | Class | Mechanism of Action | Preclinical Studies | Human Experience/Product Label |
|---|---|---|---|---|
| | | | | *Contraindications*<br>• In general, somatropin is contraindicated in the presence of active malignancy. Any preexisting malignancy should be inactive and its treatment complete prior to instituting therapy with somatropin. Somatropin should be discontinued if there is any evidence of recurrent activity. Since growth hormone deficiency may be an early sign of the presence of a pituitary tumor (or, rarely, other brain tumors), the presence of such tumors should be ruled out prior to initiation of treatment. Somatropin should not be used in patients with any evidence of progression or recurrence of an underlying intracranial tumor<br>*Precautions*<br>• Patients with preexisting tumors or growth hormone deficiency secondary to an intracranial lesion should be examined routinely for progression or recurrence of the underlying disease process. In pediatric patients, clinical literature has revealed no relationship between somatropin replacement therapy and central nervous system (CNS) turmor recurrence or new extrracranial tumors. However, in childhood cancer survivors, an increased risk of a second neoplasm has been reported in patients treated with somatropin after their first neoplasm. Intracranial tumors, in particular meningiomas, in patients treated with radiation to the head for their first neoplasm, were the most common of these second neoplasms. In adults, it is unknown whether there is any relationship between somatropin replacement therapy and CNS tumor recurrence<br>• Patients should be monitored carefully for any malignant transformation of skin lesions |

| | | | | |
|---|---|---|---|---|
| Somatropin Omnitrope™ (first approved biosimilar) | Recombinant somatropin (rDNA origin) | • Identical to human growth hormone of pituitary origin<br>• stimulates linear growth in patients who lack endogenous growth hormone also normalizes concentrations of IGF-1(insulin-like growth factor-1/ somatamedin C) | • No rodent assay<br>• No genotoxicity studies | • Leukemia has been reported in a small number of pediatric patients who have been treated with growth hormone, including growth hormone of pituitary origin and recombinant GH. The relationship, if any, between leukemia and growth hormone therapy is uncertain<br>*Contraindications*<br>• Omnitrope should not be used when there is any evidence of neoplastic activity. Intracranial lesions must be inactive and antitumor therapy complete prior to the institution of therapy. Omnitrope should be discontinued if there is evidence of tumor growth<br>*Precautions*<br>• Patients should be monitored carefully for any malignant transformation of skin lesions |
| Somatropin Valtropin® (approved as a biosimilar) | Recombinant somatropin (rDNA origin) | • Identical to human growth hormone of pituitary origin<br>• Stimulates linear growth in patients who lack endogenous growth hormone also normalizes concentrations of IGF-1(insulin-like growth factor-1/ somatamedin C) | • No rodent bioassay<br>• Not genotoxic (Ames, in vitro Chinese hamster ovary and Chinese hamster lung cell chromosomal aberration assay, and the in vivo mouse micronucleus assay | • In published literature, leukemia has been reported in a small number of pediatric GHD patients treated with somatropin. It is uncertain whether this increased risk is related to the pathology of GHD, somatropin therapy, or other associated treatments such as radiation therapy for intracranial tumors. So far, epidemiological data have failed to confirm the hypothesis of a relationship between somatropin therapy and leukemia |

**TABLE 4a Examples of communication of human carcinogenic risk without preclinical assessment of carcinogenic risk (*Continued*)**

| Biopharmaceutical | Class | Mechanism of Action | Preclinical Studies | Human Experience/Product Label |
|---|---|---|---|---|
| | | | | *Contraindications*<br>• In general, somatropin is contraindicated in the presence of active malignancy. Any preexisting malignancy should be inactive and its treatment complete prior to instituting therapy with somatropin. Somatropin should be discontinued if there is evidence of recurrent activity. Since growth hormone deficiency may be an early sign of the presence of a pituitary tumor (or, rarely, other brain tumors), the presence of such tumors should be ruled out prior to initiation of treatment. Somatropin should not be used in patients with any evidence of progression or recurrence of an underlying intracranial tumor<br>*Precautions*<br>• Patients with preexisting tumors or growth hormone deficiency secondary to an intracranial lesion should be examined routinely for progression or recurrence of the underlying disease process. In pediatric patients, clinical literature has revealed no relationship between somatropin replacement therapy and central nervous system (CNS) tumor recurrence or new extracranial tumors. However, in childhood cancer survivors, an increased risk of a second neoplasm has been reported in patients treated with somatropin after their first neoplasm. Intracranial tumors, in particular meningiomas, in patients treated with radiation to the head for their first neoplasm, were the most common of these second neoplasms. In adults, it is unknown whether there is any relationship between somatropin replacement therapy and CNS tumor recurrence<br>• Patients should be monitored carefully for any malignant transformation of skin lesions |

| | | | | |
|---|---|---|---|---|
| Insulin detemir Levemir® | • Recombinant human insulin analogue (rDNA origin) Long-acting basal insulin analogue | • Regulation of glucose metabolism | • No rodent bioassay<br>• Not genotoxic (Ames, human lymphocyte peripheral blood lymphocyte chromosome aberration test, in vivo mouse micronucleus test)<br>• Relative mitogenicity less than human insulin | • No malignancies reported in Adverse Reactions in clinical trials<br>• No warnings or precautions regarding carcinogenic risk |
| | | | *Growth factors* | |
| Filgrastim Neupogen® | • Recombinant derived human growth factor (rhG-CSF) | • Acts on cells of the hematopoietic cells by binding to specific cell surface receptors and stimulating proliferation, differentiation commitment, and some end-cell functional activation | • No rodent biosasay<br>• Not mutagenic (Ames) | *Precautions*<br>• Potential effect on malignant cells<br>• The possibility that Neupogen can act as a growth factor for any tumor type cannot be excluded<br>• Neupogen should not be administered in the period 24 hours before the administration of chemotherapy |
| Sargramstim Leukine® | • Recombinant derived human growth factor (rh GM-CSF) | • Stimulates proliferation and differentiation of hematopoietic progenitor cells | • No rodent bioassay<br>• Leukine increases the cytotoxicity of monocytes toward certain neoplastic cell lines and activates polymorphonuclear neutrophils to inhibit the growth of tumor cells | *Precautions*<br>• The possibility that Leukine can act as a growth factor for any tumor type, particularly myeloid malignancies, cannot be excluded<br>• Because of the possibility of tumor growth potentiation, precaution should be exercised when using this drug in any malignancy with myeloid characteristics. Should disease progression be detected during Leukine treatment, therapy should be discontinued |
| Epoietin alpha Epogen® | • Recombinant derived humanized growth factor | • Stimulates red blood cell proliferation | • No rodent bioassay<br>• Not genotoxic (Ames, in vitro chromosomal aberrations in mammalian cells, in vivo mouse MN, in vitro gene mutation of HGPRT locus) | *Warnings*<br>• Increased mortality and/or tumor progression |

**TABLE 4a Examples of communication of human carcinogenic risk without preclinical assessment of carcinogenic risk (*Continued*)**

| Biopharmaceutical | Class | Mechanism of Action | Preclinical Studies | Human Experience/Product Label |
|---|---|---|---|---|
| Becaplermin Regranex® (topical) | • Recombinant-derived human growth factor (rh PDGFβ) | • Promotes the chemotactic recruitment and proliferation of cells involved in wound repair and enhancing the formation of granulation tissues | • No rodent bioassay<br>• Not genotoxic (Ames, in vitro chromosomal aberrations, in vitro mammalian point mutations, DNA damage/repair, in vivo rodent MN and BM) | *Precautions*<br>• Regranex administered every other day for 13 days displayed histological changes indicative of accelerated bone remodeling consisting of periosteal hyperplasia<br>*Contraindications*<br>• Known neoplasm(s) at the site(s) of application |
| | | | *Cytokines* | |
| Interferon alpha-2a Roferon® | • Recombinant-derived human cytokine (type 1 interferon) | • Direct antiproliferative activity against tumor cells, inhibition of virus replication, and modulation of the host immune response play important roles in antitumor antiviral activity | • No rodent bioassay<br>• Not genotoxic (Ames, in vitro cytogenetic assay in human PBLs)<br>• A chromosomal defect following the addition of human leukocyte interferon to lymphocyte cultures from a patient suffering form lymphoproliferative disease has been reported. Other studies have failed to detect chromosomal abnormalities following treatment of lymphocyte cultures form healthy donors | • No malignancies reported in Adverse Reactions in clinical trials<br>• No warnings or precautions regarding carcinogenic risk |

| | | | | |
|---|---|---|---|---|
| Pegylated interferon alfa-2a Pegasys® | • Covalent conjugate of recombinant alfa-2a interferon with monomethoxy polyethylene glycol (PEG) (type 1 interferon) | • Interferons bind to specific receptors on the cell surface, initiating intracellular signaling via complex cascades of protein–protein interactions leading to rapid activation of gene transcription<br>• Interferon-stimulated genes modulate many biological effects including inhibition of viral replication in infected cells, inhibition of cellular proliferation and immunomodulation<br>• Clinical relevance of these in vitro activities is not known | • No rodent bioassay<br>• Not genotoxic (Ames, in vitro chromosomal aberrations) | • No malignancies reported in Adverse Reactions in clinical trials<br>• No warnings or precautions regarding carcinogenic risk |
| Peginterferon PegIntron™ | • Covalent conjugate of recombinant alfa-2b interferon with monomethoxyl polyethylene glycol (PEG) (type 1 interferon) | • See PEGASYS<br>• Also upregulation of the TH1-helper cell subset in vitro<br>• Clinical relevance of these findings are unknown | | • No malignancies reported in Adverse Reactions in clinical trials<br>• No warnings or precautions regarding carcinogenic risk |

**TABLE 4a Examples of communication of human carcinogenic risk without preclinical assessment of carcinogenic risk (*Continued*)**

| Biopharmaceutical | Class | Mechanism of Action | Preclinical Studies | Human Experience/Product Label |
|---|---|---|---|---|
| Interferon beta-1a Rebif® | • Recombinant derived interferon beta-1a (type 1 interferon) | • Binding of interferon beta to its receptor initiates a complex cascade of intracellular events that leads to the expression of numerous interferon-induced gene products and markers, including 2′, 5′-oligoadenylate synthetase, beta2-microglobulin, and neopterin, which may mediate some of the biological activities | • No rodent bioassay<br>• Not genotoxic (Ames, in vitro cytogenetic assay in human lymphocytes) | • No malignancies reported in Adverse Reactions in clinical trials<br>• No warnings or precautions regarding carcinogenic risk |
| Interferon beta-1a Avonex® | • Recombinant derived interferon beta-1a (type 1 interferon) | Modulates antiviral, antiproliferative and immunomodulatory activities | • No rodent bioassay<br>• Not genotoxic (Ames, in vitro cytogenetic assay in human PBLs, does not directly bind DNA) | • No malignancies reported in Adverse Reactions in clinical trials<br>• No warnings or precautions regarding carcinogenic risk |

| | | | | |
|---|---|---|---|---|
| Interferon gamma-1b Actimmune® | • Recombinant derived interferon gamma-1b (type 2 interferon) | • Interferon gamma binds to different cell surface receptors than alpha and beta<br>• Specific effect of interferon gamma include enhancement of oxidative metabolism of macrophages, antibody-dependent cellular cytotoxicity, activation of natural killer cells, and expression of Fc and major histocompatibility antigens | • No rodent bioassay<br>• Not genotoxic (Ames, in vivo mouse MN and mouse bone marrow) | • No malignancies reported in Adverse Reactions in clinical trials<br>• No warnings or precautions regarding carcinogenic risk |
| Interferon alfacon-1 Infergen® | • Consensus interferon (type 1 interferon) | • While all alpha interferon have similar biological effects, not all activities shared by each and in many cases the extent of activity various substantially for each interferon subtype | • No rodent bioassay<br>• Not genotoxic (Ames, in vitro Chromosomal Aberrations in Human PBLs) | • No malignancies reported in Adverse Reactions in clinical trials<br>• No warnings or precautions regarding carcinogenic risk |

**TABLE 4a Examples of communication of human carcinogenic risk without preclinical assessment of carcinogenic risk (*Continued*)**

| Biopharmaceutical | Class | Mechanism of Action | Preclinical Studies | Human Experience/Product Label |
|---|---|---|---|---|
| | | *Enzyme replacement therapies (ERTs)* | | |
| Imigulcerase Cerezyme® | • Recombinant human lysosomal glycoprotein enzyme β-glucocerebrosidase | • Catalyzes the hydrolysis of glycoplipid glucerebroside to glucose and ceremide | • No rodent bioassay<br>• No genotoxicity studies | • No malignancies reported in Adverse Reactions in clinical trials<br>• Studies have not be conducted in either animals or humans to assess the potential effects on carcinogenesis or mutagenesis<br>• No warnings or precautions regarding carcinogenic risk |
| Agalsidase Fabryzyme® | • Recombinant human lysosomal α-galactosidase A enzyme | • Catalyzes hydrolysis of glubtriasyl ceramide and other α-galactyl-terminated neutral glycosphingolipids to ceremide dihexoside and galactose | • No rodent bioassay<br>• No genotoxicity studies | • No malignancies reported in Adverse Reactions in clinical trials<br>• There are no animal or human studies to assess carcinogenic or mutagenic potential<br>• No warnings or precautions regarding carcinogenic risk |
| Laronidase Aldurazyme® | • Recombinant polymorphic variation of human α-L-iduronidase lysosomal hydrolase | • Catalyzes the hydrolysis of terminal α-L-iduronic acid residues of dermatan sulfate and heparin sulfate | • No rodent bioassay<br>• No genotoxicity studies | • No malignancies reported in Adverse Reactions in clinical trials<br>• Studies to assess mutagenic and carcinogenic potential have not been conducted<br>• No warnings or precautions regarding carcinogenic risk |

| | | | | |
|---|---|---|---|---|
| | | | *Other* | |
| Anakinra Kineret™ | • Recombinant, nonglycosylated human interleukin-1 receptor antagonist (rhIL-1ra) | • Blocking activity of IL-1 by competitively inhibiting IL-1 binding to the interleukin-1 type I receptor (IL-1RI) | • No rodent bioassay<br>• Not genotoxic in standard in vitro and in vivo battery<br>• An assessment of carcinogenic risk was requested by European authorities: additional in vitro mitogenesis/cell proliferation data provided; literature search revealed no published data on increased tumorigenicity in transgenic or knockout animals | *Malignancies*<br>• Lymphoma incidence rate was 3.6 fold higher than rate for general population<br>• Thirty-seven malignancies other than lymphoma were observed. Of these, the most common were breast, respiratory system, and digestive system<br>• While patients with RA, particularly those with active disease, may be at higher risk (up to several-fold) for the development of lymphoma, the role of IL-1 blockers in the development of malignancy is not known<br>*Precautions*<br>Immunosuppression<br>• Impact of Kineret on the development of malignancies is not known |
| Omalizumab Xolair® | • Recombinant DNA-derived humanized IgG1k monoclonal antibody (anti-IgE) | • Inhibits binding of IgE to high affinity IgE receptor on surface of mast cells and basophils | • No rodent bioassay<br>• Not mutagenic (Ames) | *Warnings*<br>• Malignant neoplasms observed in 0.5% XOLAIR-treated compared to 0.2% in controls in clinical studies of asthma and other allergic disorders<br>• Varity of types of tumors<br>• Majority of patients were observed for <1 year<br>• Impact of longer exposures or use in patients at higher risk for malignancy is not known |
| Pegaptanib Macugen® (ocular) | • Covalently conjugated oligonucelotide | • Selected vascular endothelial growth factor antagonist | • No rodent bioassay<br>• Not genotoxic (Ames, in vitro chromosomal aberration, cell transformation) *E. coli* strain positive | • The safety and efficacy of Macugen beyond 2 years has not been demonstrated<br>• No malignancies noted in AE profile |

*Source*: Information based on product label or FDA Pharmacology/Toxicology Review for FDA Approved Drug Products (http://www.accessdata.fda.gov/scripts/cder/drugsatfda/); supportive information EPAR's for authorized medicinal products for human use (http://www.emea.europa.eu/htms/human/epar/a.htm).

**TABLE 4b Examples of communication of human carcinogenic risk *with* preclinical assessment of carcinogenic risk**

| Biopharmaceutical | Class | Mechanism of Action | Preclinical Studies | Human Experience/Product Label |
|---|---|---|---|---|
| *Immunomdulators* | | | | |
| Abatacept Orencia™ | • Recombinant soluble fusion protein<br>• Consists of the extracellular domain of human cytotoxic T lymphocyte associated antigen 4 (CTLA-4) linked to the modified Fc (hinge, CH2 and CH3) portion of human immunoglobulin G1 (IgG1) produced in a mammalian cell expression system (anti-CTLA-4) | • Selective co-stimulation modulator, inhibits T cell activation by binding to CD80 and CD86, thereby blocking the interaction of CD28<br>• Activation provides a co-stimulatory signal necessary for full activation of T cells | • In a mouse carcinogenicity study, weekly sc injections up to 84 weeks was associated with increased incidence of malignant lymphomas (0.8 to 3× MRHD) and mammary gland tumors in mid and high doses in females<br>• Mice from this study were infected with MuLV and MMTV. These viruses are associated with an increased incidence of lymphomas and mammary gland tumors respectively, in immunsuppressed mice<br>• In a one-year toxicity study in cynomolgus monkeys, at weekly doses up to 9× MRHD no morphologic changes were observed, despite the presence of a virus (lymphocryptovirus) known to cause these lesions in immunosuppressed monkeys within the time frame of this study<br>• Relevance of this finding to the clinical use of Orencia is unknown | • In the placebo-controlled portion of the clinical trial the rate observed for lymphoma was approximately 3.5× higher than expected in the general population<br>• Patients with RA, particularly those with highly active disease, are at higher risk for the development of lymphoma. Other malignancies were also observed<br>• The role of Orencia in the development of malignancies in humans is unknown<br>*Precautions*<br>Immunosupression<br>• Possibility exists for drugs inhibiting T cell activation, including Orencia to affect malignancies, since T cells mediate cellular immune responses. Impact of treatment with Orencia on the development and cause of malignancies is not fully understood |

| | | | | |
|---|---|---|---|---|
| | | | • Not genotoxic (Ames, CHO/HGPRT forward point mutation assay, in vitro chromosomal aberrations in human PBLs) | • Orencia should not be administered with TNF agonists<br>• Orencia is not recommended for use concomitantly with anakinra |
| Natalizumab Tysabri® | Recombinant humanized $IgG_{4\kappa}$ antibody (anti-α4 integrin) | Selective adhesion molecule inhibitor produced in murine myeloma cells, binds to α4-submunit of human integrin which is highly expressed on the surface of all leukocytes, with exception of neutrophils and blocks the interaction with vascular cell adhesion molecule-1 (ICAM-1) | • No rodent bioassay<br>• Tumor promotion potential was evaluated by first identifying human tumor cell lines that bind natalizumab in vitro<br>• Tumor promotion potential was then evaluated by assessing tumor growth in athymic nude mice after tumor transplantation (sc) followed by treatment with natializumab. Natalizumab did not show any tumor promoting potential on two α4 expression tumors (leukemia, melanoma) in a nude mouse xenograft model<br>• Not genotoxic (Ames, in vitro chromosomal aberrations in human PBLs) | • No malignancies reported in Adverse Reactions in clinical trials<br>• No warnings or precautions regarding carcinogenic risk |
| Alefacept Amevive® | • Recombinant dimeric fusion protein<br>• Consists of intracellular CD2-binding portion of the human leukocyte function antigen-3 linked of the Fc portion of human IgG1 (anti-LFA-3/CD2) | • Interferes with lymphocyte activation by specifically binding to the lymphocyte antigen CD2, and inhibiting LFA-3/CD2 interaction | • No rodent bioassay<br>• Not genotoxic ( in vitro and in vivo studies) | *Warnings*<br>• Animals (nonhuman primates) developed B cell hyperplasia and one animal developed lymphoma<br>• Role of AMEVIVE in the development of lymphoid malignancy and hyperplasia |

**TABLE 4b Examples of communication of human carcinogenic risk *with* preclinical assessment of carcinogenic risk (*Continued*)**

| Biopharmaceutical | Class | Mechanism of Action | Preclinical Studies | Human Experience/Product Label |
|---|---|---|---|---|
| | | | | observed in nonhuman primates and the relationship to humans is unknown<br>• Immunodeficiency-associated lymphocyte disorders (plasmacytic hyperplasia, polymorphic proliferation and B cell lymphoma) occurs inpatients that have congenital acquired immunodeficiencies including those resulting form immunosuppressive therapies<br>*Precautions*<br>• Amevive may increase the risk of malignancies<br>• Amevive should not be administered to patients with a history of systemic malignancy<br>• Caution should be exercised when considering use in patients at high risk for malignancy<br>• Advise discontinuation if patient develops a malignancy |

| | | | | |
|---|---|---|---|---|
| *Peptide hormones* | | | | |
| Insulin glulisine Apidra® | • Recombinant human insulin analogue (rDNA origin). Rapid action insulin analogue | • Regulation of glucose metabolism | • No rodent bioassay<br>• 12-month chronic rat study (up to 20× MRHD). There was a non–dose-dependent higher incidence of mammary gland tumors in female rats administered insulin glulisine compared to untreated controls. Incidence of mammary tumors for insulin glulisine was compared to untreated controls. Incidence of insulin glulisine and regular human insulin was similar. Relevance of these finding to humans is unknown<br>• Not genotoxic (Ames, in vitro mammalian chromosome aberration test in V79 Chinese hamster cells, in vivo mammalian erythrocyte micronucleus test in rats)<br>• relative mitogenicity similar to human insulin | • No malignancies reported in Adverse Reactions in clinical trials<br>• No warnings or precautions regarding carcinogenic risk |
| Insulin Exubera® (inhaled) | • Recombinant human insulin (rDNA origin) | • Regulation of glucose metabolism | • No rodent bioassay<br>• Six-month chronic study (rats) (up to 8.3× MRHD based on body surface area) and 6-month chronic study (cynomolgus monkeys) (up to 1.4× MHRD based on body | • No malignancies reported in Adverse Reactions in clinical trials<br>• No warnings or precautions regarding carcinogenic risk |

**TABLE 4b Examples of communication of human carcinogenic risk *with* preclinical assessment of carcinogenic risk (*Continued*)**

| Biopharmaceutical | Class | Mechanism of Action | Preclinical Studies | Human Experience/Product Label |
|---|---|---|---|---|
| | | | surface area)—both were maximum tolerated doses based on hypoglycemia. No effect evident on cell proliferation indexes in alveolar or bronchiolar area in the lung in either species<br>• Not mutagenic (Ames) | |
| Insulin glargine Lantus® | • Recombinant human insulin analogue (rDNA origin) | • Regulation of glucose metabolism | • Standard 2-year bioassays performed in rats (up to 10× MRHD based on body surface area) and mice (up to 5× MHRD based on body surface area). Findings in female mice were not conclusive due to excessive mortality in all dose groups during the study. Histiocytomas found at injections sites in male rats (statistically significant) and male mice (not statistically significant) in acid vehicle containing groups. These tumors were not found in female animals, in saline control, or insulin comparator groups using a different vehicle<br>• Relevance of these finding to humans is unknown | • No malignancies reported in Adverse Reactions in clinical trials<br>• No warnings or precautions regarding carcinogenic risk |

| | | | | |
|---|---|---|---|---|
| | | | • Not genotoxic (Ames and HGPRT tests, and in tests for detection of chromosomal aberrations (cytogenetics in vitro in V79 cells and in vivo in Chinese hamsters)<br>• Relative mitogenicity greater than human insulin | |
| Insulin aspart Novolog® | • Recombinant human insulin analogue (rDNA origin) Long-acting basal insulin analogue | • Regulation of glucose metabolism | • No rodent bioassay<br>• 52-week chronic repeat dose study in rats (up to 32× MRHD based on body surface area). At highest dose Novolog increased the incidence of mammary tumors in females when compared to untreated controls. The incidence of mammary tumors was not significantly different than for regular human insulin. Relevance of these findings to humans is not known<br>• Not genotoxic (Ames, mouse lymphoma cell forward gene mutation test, human peripheral blood lymphocyte chromosome aberration test, in vivo micronculeus test in mice, ex vivo UDS test in rat liver hepatocytes<br>• Relative mitogenicity less than or equal to human insulin | • No malignancies reported in Adverse Reactions in clinical trials<br>• No warnings or precautions regarding carcinogenic risk |

**TABLE 4b Examples of communication of human carcinogenic risk *with* preclinical assessment of carcinogenic risk (*Continued*)**

| Biopharmaceutical | Class | Mechanism of Action | Preclinical Studies | Human Experience/Product Label |
|---|---|---|---|---|
| Teriparatide Forteo™ | • Recombinant human parathyroid hormone (1–34) Identical sequence to the 34N-terminal amino acids of the 84-amino acid human parathyroid hormone | • Primary regulator of calcium and phosphate metabolism in bone and kidney<br>• Physiological actions of PTH include regulation of bone metabolism, renal tubular reabsorption of calcium and phosphate, and intestinal calcium absorption | • Two carcinogenicity bioassays were conducted in rats (3–60× MRHD). Treatment resulted in a marked dose-related increase in the incidence of osteosarcoma, a rare malignant bone tumor, in both male and female rats. Osteosarcomas were observed at all doses and the incidence reached 40–50% in the high-dose groups. Teriparatide also caused a dose-related increase in osteoblastoma and osteoma in both sexes. Bone tumors in rats occurred in association with a large increase in bone mass and focal osteoblast hyperplasia<br>• Second 2-year study was carried out in order to determine the effect of treatment duration and animal age on the development of bone tumors. Female rats were treated for different periods between 2 and 26 months of age (at 3× to 20× MRHD based on body surface area). The | *Black box warning*<br>• In male and female rats, teriparatide caused an increase in the incidence of osteosarcoma (a malignant bone tumor) that was dependent on dose and treatment duration. Effect was observed at systemic exposures to teriparatide ranging from 3 to 60 times the exposure in humans given a 20 mcg dose. Because of the uncertain relevance of the rat osteosarcomas finding to humans, teriparatide should be prescribed only to patients for whom the potential benefits are considered to outweigh the potential risk. Teriparatide should not be prescribed for patients who are an increased baseline risk for osteosacrcoma (including those with Paget's disease of bone or unexplained |

| | | | | |
|---|---|---|---|---|
| | | | study showed that the occurrence of osteosarcoma, osteoblastoma and osteoma was dependent on dose and duration of exposure. Bone tumors were observed when immature 2-month old rats were treated with 20× MRHD for 24 months or with 3× and 20× MRHD for 6 months. Bone tumors were also observed when mature 6-month-old rats were treated with 20× MRHD for 6 months or 20 months. Tumors were not detected when mature 6-month old rats were treated with 3× MRHD for 6 or 20 months. Results did not demonstrate a difference in susceptibility to bone tumor formation, associated with teriparatide treatment, between mature and immature rats. Relevance of these rat findings to humans is uncertain | elevations of alkaline phosphatase, open epiphyses, or prior radiation therapy involving the skeleton)<br>*Warnings*<br>• Patients with bone metastases or a history of skeletal malignancies should be excluded from treatment with Forteo<br>*Precautions*<br>• Safety and efficacy of Forteo have not been evaluated beyond 2 years of treatment. Consequently use of the drug for more than 2 years is not recommended<br>*Information for patients*<br>• Patients should be made aware that FORTEO caused osteosarcomas in rats and that the clinical relevance of these findings is unknown |
| *Growth factors* | | | | |
| Mecasermin Increlex™ | • Recombinant derived human insulin-like growth factor-1 (rhIGF-1) | • IGF-1 is the principal hormonal mediator of statural growth. In target tissues the type 1 | • Two-year rat chronic bioassay in rodents was performed<br>• Increased incidence of adrenal medullary hyperplasia and pheochromocytoma observed | *Contraindications*<br>• Increlex is contraindicated in the presence of active or suspected neoplasia and therefore should be |

**TABLE 4b Examples of communication of human carcinogenic risk *with* preclinical assessment of carcinogenic risk (*Continued*)**

| Biopharmaceutical | Class | Mechanism of Action | Preclinical Studies | Human Experience/Product Label |
|---|---|---|---|---|
| | | IGF-1 receptor, which is homologous to the insulin receptor is activated by IGF 1 leading to intracellular signaling which stimulates multiple processes leading to statural growth | in male rates at ≥1× MRHD (males) and ≥0.3× (females) the MRHD. Increased incidence of kertoacanthoma in the skin in rats ≥4× MRHD (males) and 7× MRHD (females). Increased incidence of mammary gland carcinoma in males and females at 7× MRHD. Based on excess mortality secondary to IGF-1 induced hypoglycemia in skin and mammary tumor findings only observed at greater than human MTD<br>• *No interpretation provided as to relevance of findings above to humans*<br>• Not clastogenic (in vitro chromosomal aberration assay and in vivo mouse MN) | discontinued if evidence of neoplasia develops<br>• Increlex is not intended for use with chronic treatment of pharmacological doses of anti-inflammatory steroids |
| Peg-filgrastim Neulasta™ | • Covalent conjugate of recombinant-derived human growth factor (rh G-CSF and monomethoxypolyethylene glycol) | • Acts on cells of the hematopoietic cells by binding to specific cell surface receptors and stimulating proliferation, differentiation commitment, and | • No rodent bioassay<br>• No genotoxicity studies were performed<br>• In a toxicity study of 6 months duration in rats given sc once weekly up to 1000 mc/kg (approx 23× MRHD) no precancerous or cancer lesions | *Precautions*<br>• G-CSF receptor through which pegfilgrastim and filgrastim act has been found on tumor cell lines including some myeloid, T lymphoid, lung, head and neck and bladder tumor cell lines |

| | | | | |
|---|---|---|---|---|
| | | some end-cell functional activation | | • Possibility that pegfilgrastim can act as a growth factor on any tumor type cannot be excluded |
| Darpopoetin alpha Aranesp™ | • Recombinant human derived erythropoeisis stimulating protein closely related to erythropoietin | • Stimulates erythropoietin by same mechanism of action as endogenous erythropoietin | • No rodent bioassay<br>• No tumorigenic or unexpected mitogenic response observed in any tissue type<br>• No binding to human tissues other than those expressing the erythropoietin receptor<br>• No alteration of proliferative response of non-hematological cells in vitro or in vivo<br>• Not genotoxic (Ames, in vitro chromosomal aberration, in vivo mouse MN) | *Warnings*<br>• Erythrocyte-stimulating agents increase mortality and/or tumor progression |
| Palifermin Kepivance® | • Recombinant derived human growth factor (rhKFG) | | • No rodent bioassay<br>• Not clastogenic or mutagenic effects were observed but such studies are not informative for biological products<br>• Tg.rasH2 model completed postapproval. The study is mentioned in the Summary of Product Characteristics (SPC) and the postapproval procedural steps summary | *Precautions*<br>• Kepivance has been shown to enhance the growth of human epithelial tumor cell lines in vitro (15× MRHD) and to increase the rate of tumor cell line growth in a human carcinoma xenograft model (25–67× MRHD) |

**TABLE 4b Examples of communication of human carcinogenic risk *with* preclinical assessment of carcinogenic risk (*Continued*)**

| Biopharmaceutical | Class | Mechanism of Action | Preclinical Studies | Human Experience/Product Label |
|---|---|---|---|---|
| | | | | • In mice and rats, Kepivance enhanced proliferation of epithelial cells as measured by Ki67 immunohistochemical staining and BrDu uptake and demonstrated and increase in tissue thickness of the tongue, buccal mucosa and GI tract<br>• Safety and efficacy of Kepivance has not been established in patients with non hematologic malignancies<br>• Effects of Kepivance on stimulation of KGF-receptors-expressing non-hematopoietic tumors in patients are not known |
| Oprelvekin Neumega® | • Recombinant *E. coli* derived human growth factor (rhIL-11) | • Primary hematopoietic activity is stimulation of megakaryocytopoisis and thrombopoiesis | • No rodent bioassay<br>• In vitro Neumega did not stimulate growth of tumor colony-forming cells harvested from patients with a variety of malignancies<br>• Not-genotoxic (in vitro studies) | • Safety and efficacy of chronic administration has not been established<br>• Treatment should be discontinued at least two days before staring the next planned cycle of chemotherapy |

| | | | | |
|---|---|---|---|---|
| *Cytokines* | | | | |
| Interferon beta-1b Betaseron® | • Recombinant interferon beta-1b (Type I interferon) | • Binding of human interferon-1b to specific cell receptors induces the expression of a number of interferon-induced gene products (e.g., 2′,5′-oligoandenylate synthetase, protein kinase, and indoleamine 2,3-dioxygenase that are believed to be mediators of the biological actions | • No rodent bioassay<br>• Not genotoxic (Ames, Balb/3T3 mouse embryo transformation assay) | • No malignancies reported in Adverse Reactions in clinical trials<br>• No warnings or precautions regarding carcinogenic risk |
| *Other* | | | | |
| Dornase alpha Pulmozyme® | • Recombinant enzyme (rhDNAse) | • Selectively cleaves DNA released by degenerating leucocytes that accumulate in response to infection in CF patients | • Two-year chronic bioassay was conducted in rats up to 28× human dose (no increase in benign or malignant neoplasms) | • No malignancies reported in Adverse Reactions in clinical trials<br>• No warnings or precautions regarding carcinogenic risk |

*Source*: Information based on product label or FDA Pharmacology/Toxicology review for FDA Approved Drug Products (http://www.accessdata.fda.gov/scripts/cder/drugsatfda/); supportive information EPAR's for authorized medicinal products for human use (http://www.emea.europa.eu/htms/human/epar/a.htm).

and therefore effectively competitively inhibits the biologic activity of interleukin-1. There was no discussion of the immunotoxic or carcinogenic potential of IL-1ra in the original published FDA pharm/tox review [58]. A comment was made that genotoxicity assays had been done, but they were not considered to be of particular value because they were designed to detect mutagenic effects of small-molecule drugs, chemicals, and environmental agents that cause direct damage to DNA molecules. As such they were regarded as inappropriate for protein biotherapeutics. An assessment of carcinogenic risk was requested by the European authorities [59]. In the scientific discussion of anakinra it was stated that the carcinogenic potential was evaluated with an emphasis on risk of direct tumor production, risk of tumor stimulation, and indirect effects on tumor growth. The risk of tumor production was considered unlikely due to lack of findings in the genotoxicity assays (see comment above regarding relevance) and lack of tumors found in the 6-month rat study. The risk of tumor stimulation was also considered to be low as binding of anakinra to IL-1 receptors did not cause signal induction and data were also presented indicating that anakinra did not appear to directly stimulate mitogenesis or cell proliferation. With respect to indirect effects on tumor growth, data were presented to support lack of immunosuppression in the rat and monkey toxicity studies. A literature search revealed no published data on increased tumorigenicity in transgenic or knockout mice with an analogous defect. In addition, information was provided based on seven years of experience by the director of the laboratory that produced IL-1ra overexpressing mice and IL-1ra mutant knockout, stating that that no increased carcinogenic risk in the mice was observed.

Again, the higher incidence rate of lymphomas in RA patients compared to the general population was communicated as a warning in the label as well as by the fact that RA patients, particularly those with active disease, may be at higher risk for the development of lymphoma, but the role of IL-blockers in the development of malignancy was not known.

***Muromonab CD3*** Orthoclone OKT® 3, the first US approved monoclonal antibody, is a murine monoclonal antibody directed against the CD3 antigen on human T cells. It was approved by the US FDA for the treatment of acute renal allograft rejection in June 1986. No formal preclinical carcinogenicity assessment was performed because the muromonab CD3 did not cross-react in rodents. Nevertheless, because of the expected risk of lymphoproliferative disorders following immunosuppressive treatment, the Center for Biologics Evaluation and Research, FDA required a commitment at the time of licensure for the manufacturer to perform specific postmarketing surveys on the incidence of lymphomas in humans in addition to routine postmarketing surveillance to determine the number of malignancies following Orthoclone OKT®3 treatment. Ultimately the risk of serious and sometimes fatal neoplasias in patients was communicated as a warning in the product label in context of the diseases being treated as well as concomitant treatments including other immunosuppressive agents.

## 6.2 Peptide Hormones

***Growth Hormones*** In 1988 Watanabe et al. reported leukemia in patients receiving growth hormone (GH) therapy [60]. Over the next several years there were additional reports suggesting a GH-related increase in the incidence of leukemia. In 1997 Allen et al. reported that the incidence of leukemia in GH-treated patients without risk factors for leukemia was comparable to that in the general population of age-matched children using data form the National Cooperative Growth Study in the United States and Canada [61]. Nishi et al. followed with a report in 1999 based on data collected from more than 32,000 GH-deficient patients from 1975 through 1997. The authors concluded that the incidence of leukemia in GH-treated patients without risk factors is no greater than in the general population aged 0 to 15 years, and that a possible increased occurrence of leukemia with GH treatment appears to be limited to patients with risk factors (e.g., Fanconi's anemia, neoplasia, previous radiation, or chemotherapy). There was also no suggestion that longer periods of GH treatment were associated with more frequent development of leukemia [62].

Preclinical carcinogenicity assessment of growth hormones has not been performed for currently marketed growth hormones. In an effort to better understand the potential carcinogenic risk of growth hormone, rodent bioassays were recently performed using recombinant rat and mouse growth hormones (human homologues) in chronic rat and mouse bioassays [63]. It was difficult to compare exposure, but the levels at the high doses were higher than the basal or peak level in humans. The authors concluded that the lack of carcinogenic effects following chronic administration of GH in these two bioassays lends greater weight of evidence that high circulating levels of GH would not be associated with greater risk of tumors in subjects receiving GH replacement therapy. Unfortunately, the authors did not provide a justification for using two rodent species, especially since it was noted that there was a limited quantity of rmGH available for the mouse bioassay; therefore the high dose was lower than the dose needed to induce weight gain in mice in a 5-week study.

The current labels for growth hormones read "leukemia has been reported in a small number of pediatric patients who have been treated with growth hormone, including growth hormone of pituitary origin and recombinant somatropin. The relationship, if any, between leukemia and growth hormone is uncertain. Carcinogenicity studies have not been performed for the currently approved recombinant growth hormones."

The data using the homologous products did not show carcinogenic effects in rodents and thus can be used to support the position that there is a lack of convincing evidence for an increased risk of cancer in children receiving recombinant GH therapy. However, it is unclear whether the data are sufficient to modify the current labels for communicating carcinogenic risk. Furthermore, since the clinical data already seemed to support the lack of effect

in GH-treated patients without risk factors, it could have been of interest to have included conditions in, for example, the rat bioassay that addressed the effect of previous chemotherapy or irradiation to better understand, in a well-controlled rodent animal model, the potential impact of known human risk factors on tumor formation.

***Insulin Analogues*** Insulin, a polypeptide hormone produced by pancreatic islet β cells, was discovered in 1921. It was initially isolated from bovine/porcine glands up to the mid-1970s. Native human insulin has, in addition to its metabolic actions, a weak mitogenic effect. This effect has become important for the safety assessment of modified insulins, since structural modification of the insulin molecule could increase the mitogenic potency, possibly resulting in growth simulation of preexisting neoplasms.

Recombinant "normal" insulin has been produced since the 1980s. In the 1990s recombinant insulin analogues became available. The basic assumption that insulin was "safe" and not carcinogenic was not tested. However, results of carcinogenicity testing with the analogue AspB10 showed tumors in animals. In subsequent in vitro studies AspB10 insulin was found to possess a mitogenic activity that exceeded that of normal human insulin. It has now been shown that normal insulin can cause tumors in rats within one year of dosing at very high doses, implying a threshold for tumorigenicity.

In view of the life long exposure and large patient populations, insulins with increased mitogenic effect in relation to the unmodified human insulin in current use were thus considered to constitute a major public health concern. Therefore a thorough assessment of carcinogenic potential was recommended for all new modified insulins in 2001 [64].

In addition to the standard battery of genotoxicity studies, new insulin analogues are generally evaluated for in vitro and in vivo mitogenicity compared to normal insulin. The duration of the rodent study is based on an understanding of relative mitogenicity compared to normal insulin. Six-month repeat-dose studies have been sufficient if mitogenicity potential is low, and longer term studies are recommended if mitogenicity potential is high. In the later cases both normal insulin and ASP10 are generally included as study controls. Assessment of risk is based on cross-species pharmacology comparisons and the margin of safety compared to the intended clinical exposure.

To date, no malignancies have been reported as adverse reactions in clinical trials with insulin analogues. There are also no warnings or precautions regarding carcinogenic risk in the product labels for insulin analogues. The relevance of the findings in the rat bioassay to humans for insulin glargine, an insulin analogue with greater mitogenicity than insulin, was stated as unknown.

***Insulin-like Growth Factor*** Human Insulin like growth factor (rhIGF-1) is identical to endogenous human IGF-1. The actions of IGF-1 as a "somatomedin" are obligate for growth hormone to be able to stimulate bone and body growth. In addition to mediating many of the activities of GH, IGF-1 also has

"insulin-like" metabolic activities that support statural growth. The hormonal, cellular, and molecular mechanisms mediating the actions of rhIGF-1 on growth and metabolism are well characterized. IGF-1 concentrations are elevated in acromegaly, and there have been reports confirming that agromegaly predisposes to colonic neoplasia. While it seems probable that GH/IGF-1 may in some way be involved in the pathogenesis, it is unlikely to be an initiating carcinogen; other intraluminal local environmental influences may be important [65].

The package insert for rIGF-1 contains no interpretation or statements with respect to significance provided for the findings observed in a two-year chronic bioassay in rats with rhIGF-1 that showed an increase in tumors. Presumably this is because it was reasoned that the skin and mammary tumor findings were observed only at doses greater than the human maximum tolerated dose. Nevertheless, clinical dosing is contraindicated in the presence of active and suspected neoplasia, and discontinuation of dosing is recommended if neoplasia develops.

***PTH and PTH Analogues*** Parathyroid hormone and its analogues represent a new class of agents with anabolic effects on the skeleton. In two-year carcinogenicity studies of PTH (1-34) in two strains of rats and one strain of mice the animals developed osteosarcomas when given PTH and related peptides from weaning to 18 months of age. Many of the tumors were discovered by direct palpation and were often metastatic at the time of discovery, suggesting that they had been present for a long time. While there was a dose-related incidence in osteosarcomas, in some cases tumors occurred in animals at exposures equivalent to those commonly used in the clinical studies.

The relevance of the animal findings to the clinical administration is not currently known. The FDA subsequently developed a guidance to clarify the Agency's thinking regarding the impact of these preclinical findings on PTH drug development programs for the treatment and/or prevention of osteoporosis [66]. As a result of the concern about carcinogenicity, studies to evaluate carcinogenic potential should be performed for PTH and related peptides, although these studies could entail unique design features.

In an effort to improve the benefit-to-risk ratio of PTH in the context of the uncertain relevance of the findings in rodents, it was strongly recommended that participation in clinical studies be limited to adults with severe osteoporosis who have completed bone maturation. It was further advised that any case of osteosarcoma (or other bone tumor) be immediately reported and long-term follow-up be conducted for patients treated with PTH. Importantly, subjects in clinical trials of PTH and PTH analogues should be informed about the occurrence of osteosarcomas in rodents.

The current label for PTH(1-34) , the first approved PTH analogue, reports dose-related increase in osteosarcomas, osteoblastomas, and osteomas in two-rat carcinogenicity studies as well as results from a second two-year study in rats designed to determine the effect of treatment duration and animal

development of bone tumors. The study showed that the occurrence of osteosarcomas, osteoblastomas, and osteomas was dependent on dose and duration of exposure. However, tumors were not detected when mature 6-month old rats were treated with 3× the human dose for 6 or 20 months. Again, the label states that the relevance of the rat findings to humans is uncertain. The Medication Guide highlights the findings in a section entitled "What is the most important information I should know about Foreto™. The guide states that osteosarcomas in humans is a very serious but very rare cancer that occurs in about four out of every million older adults each year and that it is not known if humans treated with Forteo™ also have a higher chance of getting osteosarcoma.

Similar to concerns regarding the relative potency of insulin analogues, the relative carcinogenic potential of PTH and analogues will need to be compared to PTH (1-34). The actual details of the study design may need to be discussed with respective regulatory authorities. However, since osteosarcomas are due to exaggerated pharmacology, it is likely that the same regulatory and labeling standards will be applicable to all PTH and analogue products regardless of potency [67]. The question remains as to whether a two-year bioassay will be needed to communicate carcinogenic risk. Notably there is an apparent lack of osteosarcomas in humans with very long term PTH secreting tumors and in pseudohypothyroidism.

## 6.3 Growth Factors

Most growth factors are highly conserved molecules and demonstrate biological activity across species. With the exception of GM-CSF, which appears to be uniquely specific, most growth factors have been demonstrated to have activity in rodent species. There have been no rodent carcinogenicity studies reported to date and no findings of concern in short-term or chronic toxicity studies for approved growth factors: PDGF (topical), G-CSF, GM-CSF, KGF, or EPO. In addition no rodent bioassays have been reported for other growth factors that have been in clinical development, including NGF, BDNF, CNTF, EGF, TGF-α, and TGF-β. Growth factors have at least a theoretical risk for tumor promotion of cells with receptors that respond to the growth factor. These products are unlikely to be used for trivial indications; however, rodent carcinogenicity studies are unlikely to provide any useful information. The potential for tumorigenicity will likely best be determined in humans under the clinical conditions of use.

***Becaplermin*** Becaplermin is produced by recombinant DNA technology by insertion of the gene for the β chain of the platelet-derived growth factor (PDGF) into the yeast, *Saccharomyces cerevisiae*. Becaplermin has biological activity similar to that of endogenous platelet-derived growth factor, which includes promoting the chemotactic recruitment and proliferation of cells involved in wound repair and in enhancing the formation of granulation tissue.

Becaplermin is indicated for the topical treatment of lower extremity diabetic neuropathic ulcers that extend into the subcutaneous tissue or beyond and have an adequate blood supply.

In reviewing the data on becaplermin, the CPMP decided that there were no suitable tests that the applicant could employ to produce unequivocal evidence that becaplermin is or is not carcinogenic since no data were available in humans. The CPMP agreed that the risk of skin tumors or other malignancies was limited, although a theoretical concern remained since the active substance was a growth factor. The contraindications for use of the product considered known neoplasm at or near the site of application and urged caution in the use of the product in patients with known malignancies. Postmarketing follow-up to obtain additional long-term clinical safety data with particular emphasis on the incidence of tumor formation was also recommended [68].

The possibility that growth factors can act as growth stimulants for some tumor type will likely never be excluded and therefore should be communicated in product labels. However, it may be possible that the concern may be reduced through in vitro or in vivo short-term studies to better understand and characterize the potential risk.

***Palifermin*** Palifermin, a human keratinocyte growth factor (KGF), produced by recombinant DNA technology in *E. coli*., targets epithelial cells by binding to specific cell-surface receptors, stimulating proliferation, differentiation, and upregulation of cytoprotective mechanisms. The first 23 N-terminal amino acids of KGF were deleted to improve protein stability. The key findings in toxicology studies performed in rats and monkeys were generally attributable to the pharmacological activity of palifermin, specifically, proliferation of epithelial tissue.

The effect of rHuKGF on the proliferation rate of 41 tumor lines was assessed as well as the effect on the growth of subcutaneous receptor positive human tumor xenografts in athymic mice. One tumor showed a statistically significant positive. The overall data to predict potential for tumor promotion in humans were deemed questionable.

The current product label, under precautions, states that the effects of Kepivance™ on stimulation of KGF receptor expressing, nonhematopoietic tumors are not known. Kepivance™ has been shown to enhance the growth of epithelial tumor cell lines in vitro and to increase the rate of tumor cell growth in a human carcinoma xenograft model. The label advises that patients should be informed of the evidence of tumor growth and stimulation in cell cultures and in animal models of nonhematopoietic human tumors. Of note, the mutagenicity sections states that although no clastogenic or mutagenic effects were observed, such studies are generally not informative for biological products.

Although not required for approval, a short-term bioassay, Tg-Hras2 model, was conducted postapproval and is mentioned in the Summary of Product Characteristics (SPC) and the postapproval procedural steps summary. There were no findings suggestive of tumor growth promotion in this study.

### 6.4 Cytokines

Interferons are uniquely species specific and are not active in rodents. In addition the class-specific product attributes, including the mechanisms of action of the respective subtypes, reduces cause for concern for carcinogenic risk. No malignancies have been reported under adverse reactions in clinical trials, and there are no warnings or precautions regarding carcinogenic risk in the product labels or adverse reports reported postmarketing.

### 6.5 Enzyme Replacement Therapies

A variety of enzyme replacement therapies have been approved including those listed in Table 4a. Despite the fact that these products are used chronically, no genotoxicity or carcinogenicity studies have been performed presumably based on specific product class attributes.

### 6.6 Other

***Omalizumab*** Omalizumab is a recombinant humanized monoclonal antibody that binds to human IgE that is indicated for use in adults and adolescents with moderate to severe persistent asthma who are inadequately controlled with inhaled corticosteroids and have a positive skin test or in vitro reactivity to a perennial aeroallergen. Omalizumab is not pharmacologically active in rodents. Studies in wild-type mice showed a different cell distribution of IgE affinity receptor (FcεR1) compared to humans and humanized (FcεR1) transgenic mice were not available during development. Tumor host resistance models (e.g., B16 melanoma) could have been used; however, this would have required development of a humanized FcεR1 transgenic model with appropriate FcεR1 cell distribution as well as an appropriate homologous anti-mouse IgE mAb. In addition to the considerations above the chemical structure of omalizumab did not represent a carcinogenic risk, and there was no clinical epidemiological data supporting the role of IgE in immunosurveillance. Therefore no additional studies were performed to assess carcinogenic risk. In the clinical trials of less then a year study duration, a slightly higher increase in malignant neoplasms compared to controls was reported. The label reads that the impact of longer exposures or use in patients at higher risk for malignancy is not known.

***DNase*** A rodent bioassay was conducted in rats based on the intended patient population (in large part children) as well as the relevancy and feasibility of the rodent to assess toxicity. No adverse findings were observed.

## 7 RATIONALE FOR NOT PERFORMING THE RODENT BIOASSAY

Some of the shortcomings of the rodent bioassay have been discussed above. There is an increasing body of evidence suggesting that the data necessary to

identify potential human hazards may be obtained without the current reliance on the results of rodent bioassay(s) [17,18]. There are also various proposals suggesting what might constitute a relevant assessment of carcinogenic risk of pharmaceuticals. One proposed assessment includes an evaluation of a number of possible characteristics. The components of the assessment include in vitro and in vivo genotoxicity studies, a six-month rodent (or non-rodent study depending on the biologic activity of the molecule), additional pharmacologic assays to further assess pharmacodynamic or immunomodulatory effects and define dose–response relationships, pharmacokinetic studies to describe systemic exposure and dose–response behavior, and an alternative short-term carcinogenicity assay. A rodent bioassay would only be needed if all other data are inconclusive [18]. Another proposal includes conduct of a 13-week repeat dose study focusing on four endpoints: genotoxicity, immunosuppression, estrogenicity, and increased cell proliferation. These data would be combined with structure-activity models of chemical effects [17]. Short-term assessments of these effects are expected to provide predictive signals that would allow scientists to determine the likelihood of cancer in rodents, as well as the human relevance of the associated pathways. It is also important to consider human experience of spontaneous, often genetic diseases and diseases in humans or animals that may mimic potential effects of the biopharmaceutical with greater potency and specificity and that show what happens in the target species.

It is important to emphasize that while some scientists concur, in principle, that alternative assays may provide information for the integrative assessment of carcinogenic risk, they also admit that the assays may not provide a definitive answer. As previously discussed, the FDA is still considering the totality of the available data, including the results of rodent bioassay in most cases for identifying carcinogenic hazard of pharmaceuticals [16]. However, in the United States there is the opportunity to submit a waiver of carcinogenicity testing based on supported scientific rationale considering the specific product attributes, all available preclinical data and the intended patient population (see Table 5).

## 8 MANAGEMENT AND COMMUNICATING OF CARCINOGENIC RISK

*Do animal data alter the perception of carcinogenic risk? Are the preclinical data adequate to support the clinical use of the product in patients? To what extent can we accept lack of experimental data before a product is used to treat humans? Should a study be done if the relevance of the test system is questionable? Can risks be communicated without "doing a study"? What and whose risk is it not "doing a study"?*

The goal of preclinical safety testing, especially carcinogenicity studies, is to protect patients by helping to prevent toxic exposures or at least

**TABLE 5 Considerations supporting waiver of carcinogenicity studies for Fuzeon® (enfuviritide, T-20)**

| Sponsor Arguments | FDA Response | FDA Recommendation |
|---|---|---|
| *Concerns related to product class* | | *Reasons to support conducting carcinogenicity testing of T-20* |
| • T-20 is a novel retroviral agent with respect to biological and chemical characteristics and therefore there is no concern based on carcinogenicity of other retroviral agents<br>• Unique mechanism of action of T-20 involves a specific protein–protein interaction with the extracellular portion of the transmembrane glycoprotein gp41 of the HIV virus, thus blocking viral fusion with the T cell membrane<br>• Polypeptides are not known to be carcinogenic | • Insofar as T-20 is a novel retroviral agent (there are no other products in its class), the argument is irrelevant whether other products in its class have not demonstrated carcinogenic activity<br>• Question of whether there was evidence for T-20 tissue (protein-protein) interactions (that might suggest that T-20 stimulates cell division, or blocks cell apoptosis). (No data are available that looked at T-20-tissue interactions except that NIH had conducted a study of T-20 that produced unambiguous or difficult-to-interpret data)<br>• There are polypeptides that are known to be carcinogenic. All those that were located are hormones or ones that demonstrated hormonal activity | • In some cases T-20 administration was shown to cause chronic irritation and inflammation at injection sites when administered by subcutaneous injection or infusion. In some cases, injection of T-20 caused hard masses at injection sites, suggesting that T-20 was not well-absorbed<br>• Administration of T-20 is expected to exceed 6 months in some patients<br>• T-20 is a new molecular substance |
| *Structure-activity relationships* | | |
| • T-20 is expected to be metabolized into smaller peptides and amino acids. As such, there are no classical structural alerts suggesting a carcinogenic risk | | |

**TABLE 5** ***Continued***

| Sponsor Arguments | FDA Response | FDA Recommendation |
|---|---|---|
| *Concerns from repeated-dose toxicity studies* | | |
| • In repeat-dose toxicity studies of T-20 (6 months in rats, 9 months in monkeys), hyperplasia and preneoplastic lesions have not been observed | • In a 28-day repeat-dose intravenous study of T-20 in monkeys, splenic hyperplasia was observed in most of the animals (all of the high-dose animals), and the severity of splenic hyperplasia was greater in T-20-injected animals than in control animals. One animal in the high-dose group had a spleen that was greater than 3% of its body weight. This hyperplasia likely resulted from an immune reaction to T-20. However, we do not believe it is correct to state that there is no evidence of treatment-related hyperplasia | |
| *Long-term retention of parent compound or metabolite* | | |
| • Pharmacokinetic studies in rats after intravenous injection revealed a 2.4-hour half-life and low distribution (26.3 ml) suggesting that there is a little long-term retention of T-20 in rodents | • Studies where T-20 was administered by subcutaneous injection to rats or monkeys indicated poor absorption at the injection site in some instances, which in turn suggest long-term retention of T-20 is possible under some circumstances<br>• Studies of the rates of elimination of radioactivity in feces, urine, and air after a single intravenous injection of tritiated T-20 in the rat suggest catabolism of T-20 amino acids and reincorporation of radiolabel into body tissue proteins. Highest amount of radiolabel appeared in skeletal muscle 48 hours after T-20 administration. No systemic degeneration of muscle tissue was noted after chronic T-20 administration | |

**TABLE 5 Considerations supporting waiver of carcinogenicity studies for Fuzeon® (enfuviritide, T-20) (*Continued*)**

| Sponsor Arguments | FDA Response | FDA Recommendation |
|---|---|---|
| *Genotoxicity* | | *Strongest arguments for NOT conducting carcinogenicity testing of T-20* |
| • There was no evidence of genotoxicity in the standard genotoxicity test battery bacteria reverse mutation (Ames) tests, forward gene mutations in mammalian (Chinese hamster ovary AS42) cells, or mouse bone marrow MN test | • Not all carcinogens are genotoxic | • T-20 did not induce tumor in 6-month repent-dose studies in rats and 9-month repeat-dose studies in monkeys<br>• T-20 is a polypeptide comprised of naturally occurring L-amino acids. Some data indicate that T-20 is metabolized to amino acids and they, in turn, are incorporated into body tissues<br>• T-20 was not mutagenic in the standard battery of mutagenicity studies<br>• The only proteins that are carcinogenic in animals are proteins that demonstrate hormonal action. T-20 does not have hormonal action |
| *Compounds producing chronic irritation/inflammation* | | |
| • Repeated twice-daily injection via sc administration of T-20 to rats or monkeys resulted in various degrees of inflammation and irritation at the injection sites. In the 6-month rat study, the incidence and severity of lesions was comparable to that observed in vehicle-treated animals. Evidence of an injection site reaction in primates was not common until week 17. Injection site reactions have been observed in humans, which vary in severity and rarely cause treatment discontinuation<br>• Patients are instructed to rotate sites of injection to avoid chronic irritation/ inflammation at a single site | • Early studies of T-20 (e.g., rat 6-month repeat-dose study) were inconclusive as to whether injection site inflammation was solely a result of the procedure of the drug or some combination of the two. In retrospective review of data, it appears that inflammation is, at least in part, a reaction to the drug. It is irrelevant that microscopic evidence of an injection site reaction in primates was not common until week 17. In the same (9-month primate) study, increased eosinophils (compared with controls) were measured at 3 months (earliest postdosing hematological measurement), and inflammatory response were evident in the spleens (increased germinal centers and lymphocytic proliferation) and thymuses (cortical lymphocytic depletion) at 3 months. In the 28-day, repeat-dose primate study antibodies to T-20 were detected at day 7. Thus inflammation, while not clearly evident macroscopically until week 17 in the 9-month study, was microscopically evident earlier on | |

**TABLE 5** ***Continued***

| Sponsor Arguments | FDA Response | FDA Recommendation |
|---|---|---|
| | • Observations that injection site reactions in T-20 patients rarely cause treatment discontinuation, and that "patients are instructed to rotate sites of injection, thus avoiding chronic irritation/ inflammation at a singe site" are irrelevant to the carcinogenic risk assessment but important to the potential carcinogenic risk assessment | |
| *Patient population* | | *Waiver granted* |
| • T-20 is intended for patients with advanced HIV infection, and will be tested in patients who have experienced all three classes of marketed antiviral drugs. These patients have limited treatment options<br>• It is expected that patients may continue treatment of T-20 beyond 6 months, the treatment duration limit beyond which carcinogenicity studies for conventional pharmaceutical agents are generally needed. But, HIV patients have limited treatment options and the benefit–risk ratio for T-20 is considered extremely high, regardless of the outcome of animal carcinogenicity studies | • It is true that HIV infection is a life-threatening disease and that patients with advanced systemic HIV infection face limited treatment options. But, there are treatment options available to some patients. Since the advent of HAART regimens and the consequential decrease in mortality and increase in longevity of these individual, carcinogenic risk from anti-HIV drugs is a concern, especially for children. ICH guidance S1A states that where life expectancy in the indicated population is short (i.e., less than 2–3 years) long-term carcinogenicity studies may not be required. Many HIF-infected individuals are not living longer than 3 years, so this argument does not apply to all HIV-infected individuals who might receive T-20<br>• Treatment with T-20 beyond 6 months is a reason for conducting carcinogenicity testing. It is no longer true that carcinogenic risk is not a concern in the treatment of HIV-infected individual | • Most compelling reason to be concerned about carcinogenic potential of T-20 is that the subcutaneous administration of T-20 causes chronic inflammation. This adverse effect in humans can be monitored in humans<br>• Rats have limited body surface area for subcutaneous injections sites and the procedure harms the rats<br>• It is reasonable to grant the requested waiver unless there is a compelling reason to think that T-20 might be carcinogenic<br>• Other evidence suggests that T-20 is not a carcinogen |

**TABLE 5 Considerations supporting waiver of carcinogenicity studies for Fuzeon® (enfuviritide, T-20) (*Continued*)**

| Sponsor Arguments | FDA Response | FDA Recommendation |
|---|---|---|
| *Relevance of carcinogenicity testing of polypeptides* | | |
| • Standard carcinogenicity bioassays may be in appropriate for compounds such as large molecular weight proteins and polypeptides | • Large molecular weight proteins are not mentioned in the ICH guidance documents referenced above | |

*Source*: Adapted NDA No. 21–281 Pharmacology/Toxicology Review—William H. Taylor (http://www.accessdata.fda.gov/scripts/cder/drugsatfda/).

warn patients of the risks so that appropriate risk–benefit decisions can be made for the product and ultimately each patient. Despite the experimental uncertainties and limitations, appropriate communication of carcinogenic risk must be addressed. The initial guidance on the rodent bioassay for identifying carcinogenic hazard represented the best thinking of the time on how to accomplish the critical task of prospectively identifying chemicals that pose a carcinogenic risk to humans [18]. In the 40 years since this guidance was published, these recommendations continue to form the foundation of cancer risk assessment despite significant progress in our understanding of the biology of the carcinogenic response in animals and humans. However, the availability of large and extensive databases currently in development by various groups should support the development of a more rational, science-based approach to this process.

The specific product attributes of pharmaceuticals and biopharmaceuticals are blurring. Pharmaceuticals are increasing in molecular weight, including a variety of alternative scaffolds in efforts to extend half-lives and allow better efficacy and more convenient dosing regimens, while biopharmaceuticals are becoming smaller, focusing on more specific molecular interactions in efforts to improve efficacy and reduce immunogenicity and cost of goods. As is true for conventional pharmaceuticals, the cause for concern for tumorigenicity of biopharmaceuticals is heightened based on knowledge and plausibility of particular mode of action.

Special issues of concern following chronic treatment of immunomodulatory pharmaceuticals and biopharmaceuticals include the potential for immune impairment leading to opportunistic infections and/or lymphoproliferative disorders. Mitogenicity is a concern for exogenously administered biopharmaceuticals such as hormones and growth factors and may also be a concern for pharmaceuticals designed to stimulate their endogenous production. For DNA-based therapies, integration and the consequences of insertional muta-

**TABLE 6 Factors to consider for carcinogenicity testing of biopharmaceuticals**[a]

| For Rodent Bioassay | Against Rodent Bioassay |
|---|---|
| Chronic use | Species-specific protein (rodent lacks receptor or epitope, receptor or epitope, is only upregulated in disease, product induces immune response in rodents) |
| Cause for concern based on product class (e.g., growth factor, immunomodulatory protein) | Relative risk is considered no greater than other products in the class (class label) |
| Studies have routinely been performed for pharmaceuticals intended for chronic use | Studies have not been previously performed with approved products in class |
| Feasibility is not an issue based on availability of the homologous test article determined to be comparable to the clinical product (e.g., recombinant protein or monoclonal antibody) | Homologous test article[b] may not be relevant (different pharmacological activity, different physiological regulation, product is otherwise not comparable, e.g., product quality issues) |
| Effecting the intended pathway may have expected consequences | Absence of findings in rodents would not necessarily preclude precaution or warning in label of potential human risk |
| Sponsor has an obligation in the label to provide an assessment of the potential carcinogenic effect of its product. (This obligation may be satisfied as a postmarketing commitment) | Alternative assays can be used to identify potential risk—together with thorough understanding of nature and biological properties of the product obtained from studies or literature references. Such information should be provided in the label |
| Regulatory expectation | Increased animal use with little predictive value for human risk (based on one or more of the above considerations); raising regulatory hurdles for similar products in the future |

*Note*: [a]The testing is to provide meaningful information for human risk assessment and product labeling that informs physicians and patients about risks.
[b]The animal version of the human biopharmaceutical (e.g., rodent version of a recombinant protein or a monoclonal antibody).

genesis is a cause for concern, and for certain cell therapies based on specific product attributes there is an increased concern for the potential to cause tumors. Table 6 summarizes the factors for consideration of carcinogenicity assessment of biopharmaceuticals.

Bucher and Portier have suggested that "it is important for toxicologists to reevaluate our collective understanding of adverse biological responses in short-term in vivo and in vitro assays. That in spite of the recognized limitations, our understanding of carcinogenicity has advanced to the point that data from these assays could support decisions that are as protective of the public

health as are current approaches relying on the results of the rodent bioassay "[6]. MacDonald concurs suggesting that with careful consideration from academic, regulatory, and industry scientists, and an appropriately rigorous process for decision making, a significant enhancement over the current process of carcinogenicity assessment can be envisioned. Further he notes that a prospective, comprehensive evaluation of the biologic effects of chemicals could better utilize our current understanding of the carcinogenic processes and could ultimately yield a more efficient and effective assessment of potential human risk [18].

Importantly, appropriate labeling is critical to communicating potential human risk and also when appropriate should be supported with a risk mitigation plan. As most assessments of carcinogenic risk are made on individual products, additional consideration may also be needed when patients are exposed to multiple products for treating the intended disease. The need for a prospectively defined pharmacovigilance plan is a final imperative for ensuring the availability of safe medicines [36].

## ACKNOWLEDGMENTS

I would like to thank Drs. Richard Lewis, Michael McClain, James McDonald, and Professor Anthony Dayan for their critical reading of this chapter and thoughtful comments.

## REFERENCES

1. Pitot HC, Dragan YP. Facts and theories concerning the mechanisms of carcinogenesis. *FASEB J* 1991;5:2280–6.
2. Yuspa SH. Overview of carcinogenesis: past, present and future. *Carcinogenesis* 2000;21:341–4.
3. Contrera JF, Jacobs AC, DeGeorge JJ. Carcinogenicity testing and the evaluation of regulatory requirements for pharmaceuticals. *Regul Toxicol Pharmacol* 1997; 25:30–45.
4. Jena GB, Kaul CL, Ramarao P. Regulatory requirements and ICH guidelines on carcinogenicity testing of pharmaceuticals: a review on current status. *Ind J Pharmacol* 2005;37:209–22.
5. Williams GM, Iatropoulos MJ. Principles of testing for carcinogenic activity. In Hayes AW, ed. *Principles and Methods of Toxicology*, 4th ed. Philadelphia: Taylor and Francis, 2001;1–42.
6. Bucher JR, Portier C. Human carcinogenic risk evaluation. Part V: The National Toxicology Program vision of assessing the human carcinogenic hazard of chemicals. *Toxicol Sci* 2004;82:363–6.
7. Contrera JF, MacLaughlin P, Hall LL, Kier LB. QSAR modeling of carcinogenic risk using discrimiant analysis and topological molecular descriptors. *Curr Drug Discov Technol* 2005;2:55–67.

8. Boorman GA, Maronpot RR, Eustis SL. Rodent carcinogenicity bioassay: past, present and future. *Toxicol Pathol* 1994;22(2):105–11.
9. Sontag JM, Page NP, Saffiotti U. Guidelines for Carcinogen Bioassay in Small Rodents. National Cancer Institute, NCI-CG-TR-1. DHEW Publ. No. (NIH) 76-801. Washington: US GPO.
10. Halmes NC, Roberts SM, Tolson JK, Portier CJ. Reevaluating cancer risk estimates for short-term exposure scenarios. *Toxicol Sci* 2000;58:32–42.
11. Crump KS, Howec RB. The multistage model with a time-dependent dose pattern: applications to carcinogenic risk assessment. *Risk Anal* 1984;4:163–76.
12. Weisburger JH, Williams GM. Carcinogen testing: current problems and new approaches. *Sci* 1981;214:401–7.
13. Huff J, Haseman J, Rall D. Scientific concepts, value and significance of chemical carcinogenesis studies. *Ann Rev Pharmacol Toxicol* 1991;31:621–52.
14. Benigni R, Zito R. The second National Toxicology Program comparative exercise on the prediction of rodent carcinogenicity: definitive results. *Muta Res* 2004; 566:49–63.
15. Gold LS, Slone TH, Ames BN. What animal cancer tests tell us about human cancer risk? Overview of analyses of the carcinogenic potency database. *Drug Metab Rev* 1998;30:359–404.
16. Jacobs A. Prediction of 2-year carcinogenicity study results for pharmaceutical products: How are we doing? *Toxicol Sci* 2005;88:18–23.
17. Cohen SM. Alternative models for carcinogenicity testing: weight of evidence evaluation across models. *Toxicol Pathol* 2001;29:183–90.
18. MacDonald JS. Human carcinogenic risk evaluation. Part IV: Assessment of human risk of cancer from chemical exposure using a global weight-of-evidence approach. *Toxicol Sci* 2004;82:3–8.
19. Meek ME, Bucher JR, Cohen SM, Dellarco V, Hill RN, Lehman-Mckeeman LD, Longfellow DG, Pastoor T, Seed J, Patton DE. A Framework for human relevance of information on carcinogenic modes of action. *Crit Rev Toxicol* 2004; 33(6):591–653.
20. Gottmann E, Kramer S, Pfahringer B, Helma C. Data quality in predictive toxicology: reproducibility of rodent carcinogenicity experiments. *Environ Health Persp* 2001;109:509–14.
21. Cohen SM. Human carcinogenic risk evaluation: an alternative approach to the two-year rodent bioassay. *Toxicol Sci* 2004;80:225–9.
22. IARC. Monographs on the Evaluation of Carcinogenic Risk to Humans. Vol. 50. Pharmaceutical Drugs. International Agency for Research on Cancer, Lyons, France, 1990.
23. Jacobs A, Jacobson-Kram D. Human carcinogenic risk evaluation. Part III. Assessing the cancer hazard and risk in human drug development. *Toxicol Sci* 2004; 81:260–2.
24. Cohen SM. Infection, cell proliferation and malignancy. In Parsonnet J, Henning S, eds. *Microbes and Malignancy: Infection as a Cause of Cancer*. New York: Oxford University Press, 1999:89–106.

25. Hutsch T, Kapp A, Spergel J. Immunomodulation and safety of topical calcineurin inhibitors for the treatment of atopic dermatitis. *Dermatology* 2005;211:174–87.
26. ICH S1A. International Conference on Harmonization ICH S1A (1995). Need for Carcinogenicity Studies of Pharmaceuticals, 1995. http://www.fda.gov/cder/guidance/index.htm
27. ICH S6. International Conference on Harmonization ICH S6. Preclinical Safety Evaluation of Biotechnology-derived Pharmaceuticals, 1997. http://www.fda.gov/cder/guidance/index.htm
28. Hastings KL. Assessment of immunosuppressant drug carcinogenicity: standard and alternative models. *Hum Exp Toxicol* 2000;19:261–5.
29. ICH Q5D Q5D. Quality of Biotechnological/Biological Products: Derivation and Characterization of Cell Substrates Used for Production of Biotechnological/Biological Products; Availability, 1998. http://www.fda.gov/cder/guidance/index.htm
30. Hacein-Bey-Abina S, Von Kalle C, Schmidt M, McCormack MP, Wulffraat N, Leboulch P, Lim A, Osborne CS, et al. LMO2-associated clonal T cell proliferation in two patients after gene therapy for SCID-X1. *Science* 2003;302(5644):415–9.
31. Shou Y, Ma Z, Lu T, Sorrentino BP. Unique risk factors for insertional mutagenesis in a mouse model of XSCID gene therapy. *Proc Nat Acad Sci USA* 2006;103(31):11730–5.
32. Guidance for Industry. Gene Therapy Clinical Trials-Observing Subjects for Delayed Adverse Events, November 2006. http://www.fda.gov/cber/gdlns/gtclin.pdf
33. Claude JR. Difficulties in conceiving and applying guidelines for the safety evaluation of biotechnology-produced drugs: some examples. *Toxicol Lett* 1992;64–64:349–55.
34. Clark J, Leach W, Pippig S, Joshi A, Wu B, House R, Beyer J. Evaluation of a surrogate antibody for preclinical safety tesing of an anti-CD11a monoclonal antibody. *Regul Pharmacol* 2004;40:219–26.
35. Treacy G. Using an analogous monoclonal antibody to evaluate the reproductive and chronic toxicity potential for a humanized anti-TNFalpha monoclonal antibody. *Hum Exp Toxicol* 2000;19:226–8.
36. Cavagnaro JA, Spindler P. Methods for predicting tumorigenicity of immunomodulatory pharmaceuticals. *Hum Exp Toxicol* 2000;19:213–5.
37. Melnick RL, Kohn MC, Portier CJ. Implications for risk assessment of suggested nongenotoxic mechanisms of chemical carcinogenesis. *Environ Health Persp* 1996;104(Suppl 1):123–34.
38. Lima BS, Van der Laan JW. Mechanisms of nongenotoxic carcinogenesis and assessment of the human hazard. *Regul Toxicol Pharmacol* 2000;32:135–43.
39. Van Oosterhout JP, Van der Laan JW, De Waal EJ, Olejniczak K, Hilgenfeld M, Schmidt V, et al. The utility of two rodent species in carcinogenic risk assessment of pharmaceuticals in Europe. *Regul Toxicol Pharmacol* 1997;25:6–17.
40. ICH S1B. International Conference on Harmonization ICH S1B. Testing for Carcinogenicity of Pharmaceuticals, 1997. http://www.fda.gov/cder/guidance/index.htm

41. Cohen SM, Robinson D, MacDonald J. Alternative models for carcinogenicity testing. *Toxicol Sci* 2001;64:14–19.
42. Rosenblum IY, Dayan AD. Carcinogenicity testing of IL-10: principles and practicalities. *Hum Exp Toxicol* 2002;21:347–58.
43. Contrera JF, DeGeorge JJ. In vivo transgenic biossays and assessment of the carcinogenic potential of pharmaceuticals. *Environ Health Persp* 1998;106:71–80.
44. Tenant RW, Spalding J, French JE. Evaluation of transgenic mouse bioassays for identifying carcinogens and noncarcinogens. *Mutat Res* 1999;265:119–27.
45. Gulzeian D, Jacobson-Kram D, McCullough CB, Olson H, Recio L, Robinson D, Storer R, Tennant R, Ward JM, Neumann DA. Use of transgenic animals for carcinogenicity testing: considerations and implications for risk assessment. *Toxicol Pathol* 2000;28:482–99.
46. Popp JA. Criteria for evaluation of studies in transgenic models. *Toxicol Pathol* 2001;29:20–3.
47. Ashby J. Expectations for transgenic rodent cancer bioassay models. *Toxicol Pathol* 2001;29:177–82.
48. MacDonald J, French JE, Gerson RJ, Goodman J, Inoue T, Jacobs A, Kasper P, Keller D, Lavin A, Long G, et al. The utility of genetically modified mouse assays for identifying human carcinogens: a basic understanding and path forward. *Toxicol Sci* 2004;77:188–94.
49. Flammang TJ, Von Tungeln LS, Kadlubar FF, Fu PP. Neonatal mouse assay for tumorigenicity: alternative to the chronic rodent bioassay. *Regul Toxicol Pharmcol* 1997;26:230–40.
50. Omen GS. Assessment of human cancer risk: challenges for alternative approaches. *Toxicol Pathol* 2001;29:5–12.
51. Goodman JI. A perspective on current and future uses of alternative models for carcinogenicity testing. *Toxicol Pathol* 2001;29:173–6.
52. Mauthe RJ, Gibson DP, Bunch RT, Custer L. The Syrian hamster embryo (SHE) cell transformation assay: review of the methods and results. *Toxicol Pathol* 2001; 29(Suppl):138–46.
53. Brown DC, Gatter KC. Ki67 protein: the immaculate deception? *Histopathology* 2002;40:2–11.
54. Wierda D. Can host resistance assays be used to evaluate the immunotoxicity of pharmaceuticals? *Hum Exp Toxicol* 2000;19:244–5.
55. Luster MI, et al. Risk assessment in immunotoxicology. II. Relationships between immune and host resistance tests. *Fund App Toxicol* 1993;21:71–82.
56. Bugelski PJ, Herzyk DJ, Rehm S, Harmsen AG, Gore EV, Williams DM, Maleeff BE, Badger AM, Truneh A, O'Brien SR, Macia RA, Wier PJ, Morgan DG, Hart TK. Preclinical development of Keliximab, a primatized anti-CD4 monoclonal antibody, in human transgenic CD4 mice: characterization of the model and safety studies. *Hum Exp Toxicol* 2000;19:230–43.
57. FDA Pharmacology/Toxicology review (anakinra) Kineret. http://www.fda.gov/cder/foi/nda/2001/103950-0_Kineret_Pharmr.PDF12
58. Sinclair AM, Todd MD, Forsythe MD, Knox SJ, Elliott S, Begley CG. Expression and function of erythropoietin receptors in tumors: implications for the use of erythropoiesis-stimulating agents in cancer patients. *Cancer* 2007;110:477–88.

59. EPAR Kineret. http://www.emea.europa.eu/humandocs/Humans/EPAR/kineret/kineret.htm12
60. Watanabe S, Tsunematsu Y, Komiyama A, Fujimoto J. Leukemia in patients treated with growth hormone (Letter). *Lancet* 1988;1:1159–60.
61. Allen DB. Safety of human growth hormone therapy: current topics. *J Pedia* 1996; 128:S8–13.
62. Nishi Y, Tanaka T, Takano K, Fujieda K, Igarashi Y, Hanew K, Hirano T, Yokoya S, Tachibana K, Saito T, Watanabe S. Recent Status in the Occurrence of Leukemia in Growth Hormone-Treated Patients in Japan. *J Clin Endocrinol Metab* 1999; 84:1961–5.
63. Farris GM, Miller GK, Wollenberg GK, Molon-Noblot S, Chan C, Prahalada S. Recombinant rat and mouse growth hormones: risk assessment of carcinogenic potential in 2-year bioassays in rats and mice. *Toxicol Sci* 2007;97:548–61.
64. Points to Consider on the Non-clinical Assessment of the Carcinogenic Potential of Insulin Analogues (CPMP/SWP/372/01), November 2001.
65. Jenkins PJ, Besser GM, Farclough PD. Colorectal neoplasia in acromegaly. *Gut* 1999;44:585–7.
66. Guidance for Industry. Development of Parathyroid Hormone for the Prevention and Treatment of Osteoporosis, May 2000.
67. Hodsman AB, Bauer DC, Dempster DW, Dian L, Hanley DA, Harris ST, Kendler DL, McClung MR, Miller PD, Olszynski WP, Orwoll E, Yuen CK. Parathyroid hormone and teriparatide for the treatment of osteoporosis: a review of the evidence and suggested guidelines for its use. *Endocrine Rev* 2005;26:688–703.
68. EPAR Regranex. http://www.emea.europa.eu/humandocs/PDFs/EPAR/Regranex/028799en6.pdf

# Immunogenicity of Therapeutic Proteins and the Assessment of Risk

HUUB SCHELLEKENS, MD, PhD, and WIM JISKOOT, PhD

**Contents**

1 Introduction
2 Definition of Risk
3 Grading the Possible Consequences of Immunogenicity
4 Predicting the Probability of an Immune Response
  4.1 Breaking B Cell Tolerance
  4.2 Classical Activation of the Immune System by Foreign Proteins
  4.3 Mixed Type Immune Reactions
5 Other Factors Influencing the Antibody Response
  5.1 Length of Treatment
  5.2 Route of Administration
  5.3 Biological Activities of the Product
  5.4 Presence of Impurities
  5.5 Product Modifications
  5.6 Patient Characteristics
  5.7 Impact of Concomitant Therapy
6 Assay Strategies
  6.1 Types of Assays
  6.2 Assay Sensitivity
  6.3 Measuring Neutralizing Antibodies
  6.4 Timing of Blood Sampling
  6.5 Evaluation of Impurities
7 Studying the Consequences of Immunogenicity
  References

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

## 1 INTRODUCTION

Nearly all therapeutic proteins induce an immune response, although the incidence differs [1,2]. Some proteins induce an immune response only very rarely, and others in the majority of patients. The antibodies may have a variety of consequences. Low level of binding antibodies have, in general, no effects, although they may sometimes decrease the half-life and biological activity. High levels of neutralizing antibodies may interfere with the efficacy of the therapeutic proteins. The antibodies may also neutralize endogenous proteins, which may have serious biological consequences. Antibodies may also modulate the side effects of the therapeutic proteins. With some products like interferon alpha-2 the reduction of side effects may be the first sign of an immunogenic response. With monoclonal antibodies the induction of antibodies may increase the symptoms of the side effects.

The immunogenic response in itself may also lead to complications such as skin reactions, anaphylaxis-like reaction, transfusion reactions, and serum sickness. These complications may also be the results of an immunogenic response to impurities and contaminants. Both the EMEA and the FDA expect the evaluation of possible biological and clinical consequences and the assay strategy of immunogenicity to be based on a risk based approach. The elements for a risk analysis and management are discussed in this chapter, and the factors influencing the probability of an antibody response and the consequences will be categorized.

## 2 DEFINITION OF RISK

The definition of risk is the probability times the consequences (Figure 1). So the probability of an immune response is not synonymous with its risk. A high risk can be associated with a relative high probability, but also with a low probability if the consequences are severe. For example, the probability of an immune response to epoetin is rather low, but one of the consequences, antibody-induced severe anemia, is severe [3]. This makes the risk of immunogenicity of epoetins relatively high.

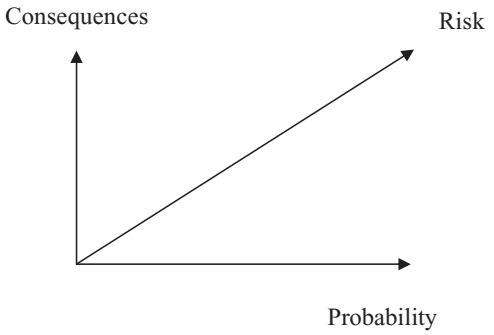


**Figure 1** Risk as a combination of probability and consequences.

**TABLE 1 Important considerations when assessing the possible consequences of immunogenicity**

| |
|---|
| 1. Absence of alternative treatments |
| 2. Similarity with important endogenous factors |
| 3. To be administered in high doses |
| 4. Presence of allergenic structures |

## 3 GRADING THE POSSIBLE CONSEQUENCES OF IMMUNOGENICITY

Table 1 lists important considerations when assessing the consequences of immunogenicity. A severe consequence of immunogenicity is the loss of efficacy of the product, if there is no alternative treatment and the product has a clinically important effect. Antibodies to factor VIII that inhibit its activity in hemophilia patients, for which there are no alternatives, are clearly severe, whereas inhibiting antibodies to Botox® used for cosmetic reasons will be considered low risk even without alternative treatment options. Loss of efficacy by the immune response may occur with any product independent whether the product is human (or derived from a human gene), nonhuman, or an artificial protein. Another severe consequence is the cross-neutralization of a biologically important endogenous factor by antibodies induced by the therapeutic protein. This can only occur if the protein drug is (partly) similar with the endogenous factor. An example is the megakaryocytic growth and differentiation factor (MGDF) that was under development for the treatment of thrombocytopenia. This truncated pegylated molecule induced antibodies in volunteers and cancer patients that cross-reacted with endogenous thrombopoetin, resulting in severe thrombocytopenia needing platelet transfusions [4].

Side effects such as anaphylaxis-like reactions and serum sickness are mainly caused by the formation of immune complexes seen with monoclonal antibodies administered in high doses. (See Chapter Assessment of Autoimmunity and Hypersensitivity.)

## 4 PREDICTING THE PROBABILITY OF AN IMMUNE RESPONSE

There are two mechanisms by which therapeutic proteins induce antibodies: the classical activation of the immune system by foreign proteins and the breaking of B cell tolerance by human proteins. The two mechanisms differ in time of onset and response level as we have described extensively previously [2]. Also the immunological mechanisms behind the two types of immune activation differ fundamentally and therefore also the characteristics of the product that are involved in induction of antibodies.

### 4.1 Breaking B Cell Tolerance

By definition, we are immune tolerant to products that are copies of endogenous proteins such as interferons, colony stimulating factors, and epoetins. The

products not necessarily need to be exact copies of the natural proteins to share this immune tolerance. When human therapeutic proteins induce antibodies, they are breaking B cell tolerance, which starts with the activation of autoreactive B cells. The probability of breaking B cell tolerance depends on several factors (Table 2). Presenting the self-epitopes in an array form is a very potent way to activate B cells [6,7]. This explains why aggregates of human proteins are the most important factor in induction of antibodies [8,9,10]. These aggregates may not be immediately present in the product but may appear during storage, making stability and formulation an important issue in predicting the immunogenicity. Some factors influence the formation of aggregates, such as the presence of oxidized therapeutic protein and the reduced solubility by the lack of glycosylation by expression of a glycoprotein in a prokaryotic host [11].

There are only a few studies in experimental model systems on the properties of the aggregates that break B cell tolerance, indicating that only multiple order aggregates (>trimers) are involved [8]. There are different methods to identify the presence of aggregates such as size exclusion chromatography and SDS-PAGE and Western blotting. However, these methods may miss immunogenic aggregates. Other methods such as analytical ultracentrifugation, field-flow fractionation, and light-scattering techniques have shown to be able to detect aggregates that go undetected by standard assays.

The only biological test to study the capacity of a protein product to break B cell tolerance are mice made transgenic for the specific protein [12,13]. These mice are immune tolerant, and there is a good correlation of an immune response in these mice and in patients. Although these models have helped to identify the factors important for breaking B cell tolerance and also have been useful in improving the formulation of products, there is not yet enough experience to use them as absolute predictors of immunogenicity of human proteins.

## 4.2 Classical Activation of the Immune System by Foreign Proteins

The classical activation of the immune system is driven by the presence of non–self-epitopes. If the therapeutic protein is of nonhuman origin like microbially derived streptokinase and asparaginase or bovine adenosine deamidase,

**TABLE 2 Most important factors increasing the probability of an immune response to human proteins**

| |
|---|
| 1. Aggregates |
| 2. Chronic treatment |
| 3. Subcutaneous administration |
| 4. Impurities |
| 5. Breaking tolerance in immune tolerant mice |

a strong immune response is highly likely, especially after subcutaneous administration (Table 3). If the product is a modified human protein, prediction becomes more difficult. The level of divergence of the human amino acid sequence is not very informative because the consequence of the sequence deviation is highly dependent on the type of amino acid change and the location of the change in the molecule. In some cases a single amino acid change has been reported to induce a complete new epitope [14]. In others, extensive amino acid exchanges have no effect on immunogenicity.

Also in the prediction of the introduction of new epitopes by protein modification immune-tolerant transgenic mice can be used. Mice transgenic for human insulin and tissue plasminogen activator have been used to identify new epitopes in products with amino acid modifications [14,15].

Antisera have also been used to monitor reduction in immunogenicity of a protein. A reduced binding with the modified protein by serum raised by the original product was considered to show reduced immunogenicity [16]. However, binding of antibodies is an antigenicity, which is a physical process. An antigenicity is different from an immunogenicity, the capacity to induce an immune response. Although in some cases these properties are related, in many cases proteins with a high antigenicity for neutralizing antibodies, proved to be non-immunogenic.

Also a number of in vitro stimulation and binding tests and computational models are promoted as predictors of immunogenicity [17]. However, all these tests have their limitations. T cell proliferation assays have the drawback that many products are capable of either inducing some level of T cell activation or inhibit cell proliferation, leading to false positive and negative effects. The computational algorithms that are claimed to predict T cell epitopes only give limited information on the interaction of the proteins with the immune system and underdetect epitopes. Although these assays and algorithms may help predict to some extent which parts of the proteins will be involved in immunogenicity, their limitations are also evident when they are used to reduce immunogenicity. There is hardly any convincing evidence of a clinically relevant reduction of antibody induction.

Glycoproteins produced in nonhuman cells such as plant cells and yeast cells may have a modified glycan structure. Although natural antibodies exist that react with nonhuman glycan structures [18], there is no example of an immune reaction that was mounted to modified glycan structure of a therapeutic protein.

**TABLE 3 Most important factors increasing the probability of an immune response to foreign proteins**

| |
|---|
| 1. Level of non-self |
| 2. Subcutaneous administration |

### 4.3 Mixed Type Immune Reactions

In some cases therapeutic proteins induce antibodies both by breaking B cell tolerance and the vaccine-like activation of the immune system, depending on the level of immune tolerance of the patients. An example is the immune response induced by factor VIII in hemophilia patients [19]. The antibody response is related to the type of defect in the factor VIII gene in these patients. If the genetic defect leads to an inability to produce factor VIII, the patients lack immune tolerance, and they respond as to a foreign protein. In patients producing a nonfunctional but immunologically normal factor VIII, antibodies are produced by breaking B cell tolerance. It is likely, although not studied, that chimeric proteins consisting of both a human and nonhuman part such as chimeric and humanized monoclonal antibodies also will produce a mixed type of response.

## 5 OTHER FACTORS INFLUENCING THE ANTIBODY RESPONSE

### 5.1 Length of Treatment

Foreign proteins like streptokinase and asparaginase induce antibodies often after a single injection. Breaking B cell tolerance by human protein takes in general more than 6 months of chronic treatment. This illustrates the fundamental difference between a vaccine-like response to foreign proteins and breaking of immune tolerance against human proteins.

### 5.2 Route of Administration

The route of administration influences the likelihood of an antibody response independent of the mechanism of induction. The probability of an immune response is the highest with subcutaneous administration, less probable after intramuscular administration and intravenous administration is the least immunogenic route. There are no studies comparing parenteral and nonparenteral routes of administration. However, as both mucosal tissues and the skin are immune competent organs designed to keep invaders out of the body, intranasal, pulmonary, and transdermal administration of therapeutic proteins may increase the risk of an immune response as compared to parenteral routes.

### 5.3 Biological Activities of the Product

The biological activities of the product are influencing the immune response. An immune stimulating therapeutic protein is more likely to induce antibodies than an immune suppressive protein. Monoclonal antibodies targeted to cell bound epitopes are more likely to induce an immune response than monoclonal antibodies with a target in solution [20]. Also the Fc bound activities of monoclonal antibodies have an influence. A monoclonal antibody with a modified Fc part has been reported to be less immunogenic than the unmodified monoclonal antibody [21].

### 5.4 Presence of Impurities

Impurities may influence immunogenicity. The immunogenicity of products as human growth hormone, insulin, and interferon alpha-2 have declined over the years due to improved downstream processing and formulation, reducing the level of impurities. There are studies showing that the induction of antibodies by oxidized protein cross-reacted, with the unmodified product and host cell derived endotoxin acting as an adjuvant [22]. The probability of an immune response therefore increases with the level of impurities. Besides their possible role as adjuvant, impurities may also induce an immune response to themselves. Moreover they may be the cause of skin reactions, allergies and other side effects.

### 5.5 Product Modifications

Product modifications that are intended to enhance half-life potentially also increase the exposition of the protein to the immune system and may increase immunogenicity. In addition the modification may reduce biological activity necessitating more protein for the same biological effect.

Pegylation is claimed to reduce the immunogenicity of therapeutic proteins by shielding [23]. There is evidence that pegylation reduces the immunogenicity of nonhuman proteins like bovine adenosine deamidase and asparginase. Whether pegylation also reduces the capacity of human proteins to break B cell tolerance is less clear. There are reports of high immunogenicity of pegylated human proteins such as PEG-rhuMDGF, but the immunogenicity of unpegylated MDGF products is unknown.

### 5.6 Patient Characteristics

Gender, age, and ethnic background have all been reported to influence the incidence of antibody response to specific therapeutic proteins. However, the only patient characteristic that consistently has been identified for a number of different products is the disease that the patients suffer from. Cancer patients are less likely to produce antibodies to therapeutic protein than other patients. The most widely accepted explanation for this difference is the immune-compromised state of cancer patients, both by the disease as by anticancer treatment. Also the median survival of patients on treatment by therapeutic proteins may be too short to develop an antibody response. In any case, cancer reduces the probability of an antibody response to a protein considerably.

### 5.7 Impact of Concomitant Therapy

As the experience in cancer patients shows, immune suppressive therapy reduces the probability to develop an immune response to proteins. In addition immune suppressive drugs such as methotrexate are often used in

conjunction with monoclonal antibodies and other protein drugs that further serves to reduce the immune reactions.

## 6 ASSAY STRATEGIES

A specifically designed assay strategy is an essential part of every risk analysis and management program for a new therapeutic protein. When validating the assays and also when evaluating the data on immunogenicity and their clinical impact, it is important that assays are not standardized and international reference and standard preparations are in general not available [25]. This makes it difficult to compare studies with similar products. That is, the same product studied in comparable patient populations may show a wide variety of antibody incidence. Also the different analyses differ in their criteria for considering a patient antibody positive. In some studies all patients with at least a single positive binding assay result were included in the antibody positive groups. Because it is unlikely that a transient level of binding antibodies has any biological effect, this type of analysis leads to an underestimation of the effects of antibodies.

The more logical approach is to categorize patients depending on the AUC of the neutralizing antibody levels. This type of analysis also allows study of the dose–response relation between antibody levels and clinical effects. In some cases the unpredictable disease progression, the lack of good efficacy markers and the extended effect after treatment has stopped, may complicate the evaluation of the clinical effects of immunogenicity.

Because of the lack of adequately designed studies and standardized assays for many products the consequences of their immunogenicity is unclear. So assuming a product to be non-immunogenic based on literature is impossible. In other words, every new protein should be assumed to be immunogenic unless proved otherwise with a well-designed and validated assay strategy.

Recently a number of papers have appeared written mainly by authors in the US biotechnology industry, and they are the basis of a growing consensus on the principles of immunogenicity testing [26,27,28]. One single assay is not considered sufficient to evaluate the immunogenicity of a new protein drug. Most antibody assay strategies are based on different stages approach: the first stage being a screening assay to identify the antibody positive sera and the second stage being the assay to evaluate whether the antibodies are neutralizing. This may be followed by assays to establish titer, affinity, and isotype of the antibodies.

### 6.1 Types of Assays

The screening assay is, in general, a binding assay, mostly an ELISA-type assay, or a radio-immune-precipitation method. Standard ELISA-type immunoassays are not always considered appropriate, however, for measuring binding

antibodies because the circulating therapeutic protein may interfere with the assay. So a bridging assay, which in the ELISA plate is coated with the protein drug and anti-drug antibodies are detected with a labeled version of the drug, has been advocated as the best type of screening assay. A bridging assay detects all types of antibodies independent of type and species of origin. The bridging assay may miss low-affinity antibodies. Therefore, for the early immune response, the use of surface plasmon resonance technology as BIAcore is advocated rather than the ELISA type of assay methodology. However, the sensitivity of ELISAs and RIAs for high-affinity antibodies is generally higher than the BIAcore assays.

Screening assays are designed for optimal sensitivity to avoid false negatives. For new proteins defining an absolute sensitivity is impossible because of lack of positive sera. An alternative approach is to set the cut point for the assay at the 5% false positive level using a panel of normal human sera and/or untreated patient sera representative of the groups to be treated.

### 6.2 Assay Sensitivity

To get an idea of the sensitivity of the assay in quantitative terms, affinity purified animal sera are used. In general, these are obtained by immunization of rabbits by the therapeutic product. The rabbit serum diluted in normal human serum can also serve as a positive control or to provide a standard curve. Because the screening assays are tuned for high sensitivity and a predefined false positive rate, all initial positive sera need to be confirmed. A confirmation can be achieved by another binding assay based on another principle than the original assay. So, if an ELISA type of assay is used, the confirmation can be a radio-immune precipitation assay or a BIAcore assay. An initial positive can also be confirmed by doing a displacement step before doing the original assay. Product is added to the serum and should result in a predefined relative reduction in signal, such as 25% or 50%, or an absolute reduction based on a low positive control. Only the confirmed positives are labeled as positives and are then further analyzed. Confirmed positive sera should be tested for neutralizing antibodies that may interfere with the biological and clinical activity. Assays for neutralizing activity are based on the inhibition of the biological effect of the protein in vitro. Because every product has its own specific biological effect, assays for neutralizing activity need to be designed individually.

### 6.3 Measuring Neutralizing Antibodies

The assay for neutralizing antibodies is, in general, a modification of the potency assay of the therapeutic protein product. The potency assay is in most cases an in vitro cell based assay. A predefined amount of product is added to the serum and the reduction of activity is evaluated in the bioassay. Sometimes the assay is performed in two steps. An assay with the lowest

dilution of serum and a minimal amount of protein is used to screen for neutralizing activity and, if any neutralization is identified, a full titration of the sera is performed.

An important issue in the neutralization assay is the amount of therapeutic protein product to add. This dose should be based on the dose–response curve of the protein in the assay. The reduction should be in the linear part of the curve and be significant enough to be reliably testable in the assay. The titer of the neutralization is often expressed in units. The highest dilution of the serum with a significant inhibition is defined to contain one neutralizing unit. Sometimes the activity is expressed in arbitrary units using an animal antiserum as reference. Further characterization of the antibodies may include evaluation of Ig isotype and affinity. Although alternative methods may be employed, the BIAcore assay has become the standard for these types of analyses.

### 6.4 Timing of Blood Sampling

Another important aspect when studying the immunogenicity of therapeutic proteins concerns the timing of the blood sampling. With products such as monoclonal antibodies that have a relative long half-life, the circulating product may interfere with the detection of induced antibodies. Sampling sera up to weeks after the last injection is then necessary to avoid the interference of circulating protein drug.

### 6.5 Evaluation of Impurities

In general, the evaluation of the immunogenicity of impurities is restricted to binding antibodies. Only in the case where the biological activity of the impurity is known and its inactivation is biologically relevant, a neutralization assay may be necessary.

The most important source of impurities is the host cells. And in nearly all biological products traces of host cells can be identified. Not all of them will raise antibodies. Based on the literature about the immunogenicity of a protein, it is general safe to assume that if the level of contaminant is below 0.5 μg per dose, the induction of antibodies is unlikely. For a proper binding antibody assay, the material resulting from a complete extraction and purification process of a mock production run of the host cells without the inserted gene (0 cells) should be used. This procedure should results in material containing the impurities as present in the product. The impurities that are co-purified with the therapeutic protein will also be present in the whole cell extract, but the presence of other substances can introduce interference in the binding assay. The 0-cell product can also be used in the confirmatory displacement assay and to immunize rabbits to obtain a positive control serum. The same material can also be used in an immunoblot or BIAcore assay, which can serve as confirmatory assays.

## 7 STUDYING THE CONSEQUENCES OF IMMUNOGENICITY

To evaluate the biological consequences of antibodies, patients with positive responses should be monitored for their impact on the pharmacokinetic and pharmacodynamic effect of the therapeutic protein. Also the effect on adverse events and possible neutralization of the endogenous protein, in cases where human proteins are administered at pharmacological doses, should be monitored. When the consequences of the immunogenicity are known, the risk can be established.

## REFERENCES


1. Schellekens H. Immunogenicity of therapeutic proteins: clinical implications and future prospects. *Clin Ther* 2002;24(11):1720–40.
2. Schellekens H. Bioequivalence and the immunogenicity of biopharmaceuticals. *Nat Rev Drug Discov* 2002;1(6):457–62.
3. Casadevall N, Nataf J, Viron B, Kolta A, Kiladjian JJ, Martin-Dupont P, Michaud P, Papo T, Ugo V, Teyssandier I, Varet B, Mayeux P. Pure red cell aplasia and antierythropoietin antibodies in patients treated with recombinant erythropoietin. *New Engl J Med* 2002;346:469–75.
4. Vadhan-Raj S. Clinical experience with recombinant human thrombopoietin in chemotherapy-induced thrombocytopenia. *Semin Hematol* 2000;37(2 Suppl 4): 28–34.
5. Klastersky J. Adverse effects of the humanized antibodies used as cancer therapeutics. *Curr Opin Oncol* 2006;18(4):316–20.
6. Bachmann MF, Rohrer UH, Kundig TM, Burki K, Hengartner H, Zinkernagel RM. The influence of antigen organization on B cell responsiveness. *Science* 1993; 262:1448–51.
7. Chakerian B, Lenz P, Lowy DR, Schiller JT. Determinants of autoantibody induction by conjugated papillomavirus virus-like particles. *J Immunol* 2002;169: 6120–6.
8. Hermeling S, Schellekens H, Maas C, Gebbink MF, Crommelin DJ, Jiskoot W. Antibody response to aggregated human interferon alpha2b in wild-type and transgenic immune tolerant mice depends on type and level of aggregation. *J Pharm Sci* 2006;95(5):1084–96.
9. Moore WV, Leppert P. Role of aggregated human growth hormone (hGH) in development of antibodies to hGH. *J Clin Endocrinol Metab* 1980;51:691–7.
10. Braun KL, Labow MA, Alsenz J. Protein aggregates seem to play a key role among the parameters influencing the antigenicity of interferon alpha (IFN-alpha) in normal and transgenic mice. *Pharm Res* 1997;14:1472–8.
11. Runkel L, Meier W, Pepinsky RB, Karpusas M, Whitty A, Kimball K, Brickelmaier M, Muldowney C, Jones W, Goelz SE. Structural and functional differences between glycosylated and non-glycosylated forms of human interferon-beta (IFN-beta). *Pharm Res* 1998;15:641–9.

12. Hermeling S, Jiskoot W, Crommelin D, Bornaes C, Schellekens H. Development of a transgenic mouse model immune tolerant for human interferon Beta. *Pharm Res* 2005;22(6):847–51. Epub 2005 Jun 8.
13. Hermeling S, Crommelin DJ, Schellekens H, Jiskoot W. Structure-immunogenicity relationships of therapeutic proteins. *Pharm Res* 2004;21(6):897–903.
14. Ottesen NP, Jami J, Weilguny D, Duhrkop M, Bucchini D, Havelund S, Fogh JM. The potential immunogenicity of human insulin and insulin analogues evaluated in a transgenic mouse model. *Diabetologia* 1994;37:1178–85.
15. Stewart HPG, Hollingshead PG, Pitts SL, Chang R, Martin LE, Oakley H. Transgenic mice as a model to test the immunogenicity of proteins altered by site-specific mutagenesis. *Mol Biol Med* 1989;6:275–81.
16. Elliott S, Chang D, Delorme E, Dunn C, Egrie J, Giffin J, Lorenzini T, Talbot C, Hesterberg L. Isolation and characterization of conformation sensitive antierythropoietin monoclonal antibodies: effect of disulfide bonds and carbohydrate on recombinant human erythropoietin structure. *Blood* 1996;87:2714–22.
17. Flower DR, Doytchinova IA. Immunoinformatics and the prediction of immunogenicity. *Appl Bioinformatics* 2002;1(4):167–76.
18. Buzas EI, Gyorgy B, Pasztoi M, Jelinek I, Falus A, Gabius HJ. Carbohydrate recognition systems in autoimmunity. *Autoimmunity* 2006;39(8):691–704.
19. Lee CA, Lillicrap D, Astermark J. Inhibitor development in hemophiliacs: the roles of genetic versus environmental factors. *Semin Thromb Hemost* 2006;32(Suppl 2):10–4.
20. Pendley C, Schantz A, Wagner C. Immunogenicity of therapeutic monoclonal antibodies. *Curr Op in Mol Ther* 2003;5:172–9.
21. Presta LG. Engineering of therapeutic antibodies to minimize immunogenicity and optimize function. *Adv Drug Deliv Rev* 2006;58(5–6):640–56. Epub 2006 May 23.
22. Milner RDG, Flodh H. In *Immunological Aspects of Human Growth Hormone: Proceedings of a Workshop*, London, 12 November 1985. Medical Education Services, Oxford (1986).
23. Veronese FM, Pasut G. PEGylation, successful approach to drug delivery. *Drug Discov Today* 2005;10(21):1451–8.
24. Rutgeerts P, Van Assche G, Vermeire S. Optimizing anti-TNF treatment in inflammatory bowel disease. *Gastroenterology* 2004;126(6):1593–610.
25. Schellekens H, Ryff JC, van der Meide PH. Assays for antibodies to human interferon-alpha: the need for standardization. *J Interferon Cytokine Res* 1997; 17(Suppl 1):S5–8.
26. Neyer L, Hiller J, Gish K, Keller S, Caras I. Confirming human antibody responses to a therapeutic monoclonal antibody usinga statistical approach. *J Immunol Meth* 2006;315(1–2):80–7. Epub 2006 Aug 7.
27. Geng D, Shankar G, Schantz A, Rajadhyaksha M, Davis H, Wagner C. Validation of immunoassays used to assess immunogenicity to therapeutic monoclonal antibodies. *J Pharm Biomed Anal* 2005;39(3–4):364–75.
28. Mire-Sluis AR, Barrett YC, Devanarayan V, et al. Recommendations for the design and optimization of immunoassays used in the detection of host antibodies against biotechnology products. *J Immunol Meth* 2004;289:1–16.

# Assessment of Autoimmunity and Hypersensitivity

JACQUES DESCOTES, MD, PharmD, PhD, and THIERRY VIAL, MD

**Contents**

1 Introduction
2 Overview of the Clinical Experience
  2.1 Autoimmunity
  2.2 Hypersensitivity
3 Regulatory Guidance
4 Preclinical Evaluation of the Potential of Biopharmaceuticals for Autoimmunity
5 Preclinical Evaluation of the Potential of Biopharmaceuticals for Hypersensitivity Reactions
6 Conclusion
References

## 1 INTRODUCTION

The immune system is an extremely complex network of closely interacting regulatory and effector cells and molecules, whose primary function is the discrimination of self from non-self in order to maintain homeostasis in living organisms. The renewal, differentiation, and activation of immunocompetent cells ensure an adequate level of immune responsiveness under the control of many mechanisms with either redundant or conflicting outcome. It is important to bear in mind, however, that not all immune responses are beneficial to the host. Some may result in autoimmunity when directed against self-constituents of the body, or in hypersensitivity when directed against "innocent" foreign antigens (allergens) [11]. Biopharmaceuticals, either via

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

their (immuno)-pharmacological mechanism of action, or their chemical structure (foreign proteins), most often have the potential for inducing autoimmunity and/or hypersensitivity reactions in treated human subjects. Therefore autoimmunity and hypersensitivity are critical issues to be considered during the preclinical as well as the clinical safety evaluation of most, if not all biopharmaceuticals.

## 2 OVERVIEW OF THE CLINICAL EXPERIENCE

There is a growing body of clinical evidence demonstrating that autoimmunity and hypersensitivity are relatively common and potentially severe adverse effects of biopharmaceuticals [26,37,41].

### 2.1 Autoimmunity

Autoimmunity is still largely a mystery. It consists of specific immune responses involving antibodies and T cells directed against self-antigens. Importantly, physiological autoimmune responses are relatively common as most healthy individuals normally produce autoantibodies. Autoimmune diseases occur when autoreactive T cells and/or autoantibodies cause tissue damage resulting in overt pathological conditions [8]. They are relatively common diseases as 5% to 8% of the US population have been estimated to be affected. Autoimmune diseases can target any organ system, and occur at nearly any age. Approximately 80 diseases have been described. They are usually classified into organ-specific and systemic diseases, depending on the affected organ(s), even though this classification does not necessarily reflect a common pathophysiology. A number of predisposing factors are strongly thought to be involved, especially genetic predisposition and environmental exposures, including drug therapy.

Autoimmune diseases have been reported to be more frequent in human subjects treated with several recombinant cytokines [38]. For instance, increased titers or the new occurrence of autoantibodies have been observed in hepatitis C patients treated with the recombinant interferons-alpha (IFNα). Quite a few clinical case reports describe the development of organ-specific as well as systemic autoimmune diseases including systemic lupus erythematosus, insulin-dependent type I diabetes mellitus, autoimmune thrombocytopenia, autoimmune hemolytic anemia, myasthenia gravis, and autoimmune thyroiditis in patients under IFNα therapy. Although the mechanism involved is not fully elucidated, the available data support the pathogenic potential of IFNα in autoimmunity [31]. In contrast, autoimmune effects associated with IFNβ therapy are thought to be of lesser concern based on the current clinical evidence [38]. Thyroid autoimmunity in contrast to other autoimmune diseases is frequent in patients treated with recombinant interleukin-2 (rIL-2). Thus, among 281 previously euthyroid cancer patients treated with rIL-2, up to 41%

developed thyroid dysfunction [16]. However, rIL-2 therapy rather rarely exacerbates other latent autoimmune diseases, including diabetes mellitus, myasthenia gravis, and rheumatoid arthritis [38]. Autoimmunity has also been reported to be an adverse event of antitumor necrosis factor-alpha (TNFα) therapy involving either the monoclonal antibody infliximab or the fusion protein etanercept [2]. Although autoimmune diseases, such as rheumatoid arthritis or Crohn's disease are their primary therapeutic targets, autoimmune complications, such as systemic lupus erythematosus have been described [9], and this emphasizes the complex interplay of the cytokine and immune cell networks in the development of autoimmunity [21].

## 2.2 Hypersensitivity

Because the word allergy has so often been used misleadingly, hypersensitivity is the recommended term to refer to those adverse drug events mediated by inadvertent immune responses that develop in a small number of human subjects [14]. Typically hypersensitivity reactions involve either specific antibodies or T lymphocytes directed against a given drug antigen. They can affect nearly every organ or tissue of the body, but one organ or tissue is often a predominant target. They are the consequence of the exquisite capacity of the immune system to recognize structural elements (epitopes) of non-self molecules, and mount specific responses involving the immunological memory [11].

Since the vast majority of biopharmaceuticals are large foreign molecules, they have the potential to act as direct immunogens; that is to say, they can directly both sensitize the treated host and then trigger an adverse clinical reaction upon a subsequent rechallenge. Virtually all therapeutic proteins, including human therapeutic proteins, humanized or human monoclonal antibodies, are immunogenic to some extent [15]. The most common finding in relation to the immunogenicity of biopharmaceuticals is the development of specific antibodies whose clinical consequences range from no apparent adverse effects to potentially life-threatening complications. Indeed specific antibodies can be neutralizing and result in altered pharmacokinetics or reduced clinical response, or they can evoke hypersensitivity reactions such as immune complex-associated diseases and acute hypersensitivity reactions [12]. The incidence and characteristics of these antibodies are dependent on several criteria, which include the product, the dose and duration of treatment, the frequency of dosing, the route of administration, the immune status, and/or genetic profile of the patient, the disease type, and the functional activities of the protein. For instance, the rate of anti-IFNα antibody formation ranges from zero to more than 50% of patients across clinical studies. Although no obvious clinical consequence are associated with anti-IFNα antibodies in most patients, changes in pharmacokinetics due to neutralizing antibodies are sometimes noted, but reduced clinical response has rarely been documented. Anti-IFNα antibodies have so far not been associated with immune complex-associated diseases nor hypersensitivity reactions, and no IgE-mediated reactions to

IFNα have ever been reported to the best of our knowledge [38]. Similarly anti-IFNβ antibodies have commonly been detected in the sera of treated patients. Neutralizing antibodies seem to be more frequent with IFNβ than IFNα. The former may be associated with a higher relapse rate or shortened time to first relapse in patients with multiple sclerosis [33], and worsening of the disease has been suggested to be correlated with persistent high titers of neutralizing antibodies [19]. Nevertheless, extremely rare hypersensitivity reactions to IFNβ have so far been reported. The clinical experience gained with recombinant IL-2 confirms that only a relatively small fraction of specific antibodies are neutralizing and that ensuing hypersensitivity reactions are exceptional [38].

Predicting the consequences of specific antibodies to therapeutic proteins is in fact extremely difficult, as exemplified by the recombinant hemophilia factors VIII and IX. While specific antibodies to factor VIII have been detected in the sera of up to 30% to 50% of treated patients, resulting in transient inhibition and hemorrhage risk, neutralizing antibodies to factor IX have been detected in only 1% to 3%; these antibodies were associated with severe anaphylactic reactions in approximately 50% of the affected patients [18]. Specific antibodies against therapeutic monoclonal antibodies are also commonly reported. For instance, antibodies to the anti-TNFα agent infliximab, a chimeric monoclonal $IgG_1$ antibody, have been detected in the sera of 10% to 60% of patients and are not uncommonly associated with reduced efficacy. Importantly, anti-infliximab antibodies have been associated with an increased incidence of infusion reactions [4]. Although the majority of infusion reactions are not IgE-mediated, true anaphylactic reactions have been described [6].

To date, the most severe immune-mediated adverse event associated with specific antibodies to a therapeutic protein is erythropoietin-associated pure red cell aplasia [17]. Hundreds of cases have been described with a majority involving patients with chronic kidney disease receiving subcutaneous injections of erythropoietin-α. There is a major concern that therapeutic proteins used as replacement therapy in case of deficiencies affecting a natural protein of the body can trigger a specific immune response, the most likely manifestation of which is the production of neutralizing antibodies as described above. By inhibiting the biologic effects of both the recombinant and natural protein, neutralizing antibodies can result in severe complications as exemplified by pure red cell aplasia in patients where neutralizing antibodies eradicated any residual release of erythropoietin, the hormone required for erythrocyte production. Because erythropoietin-specific antibodies are involved, but the immune response was initially mounted against the potentially "slightly different," recombinant erythropoietin, used as a therapeutic agent compared to endogenous erythropoietin, pure red cell aplasia can be defined as a hypersensitivity reaction as well as an autoimmune reaction.

Finally, it should be emphasized that drug-induced hypersensitivity reactions can be either immune mediated (e.g., anaphylaxis) or non–immune mediated [43]. Infusion reactions associated with a number of therapeutic

monoclonal antibodies are a typical example of non–immune-mediated hypersensitivity reactions. Indeed the majority of reported infusion reactions are not associated with the detection of specific IgE. A nonspecific activation of the complement cascade is an attractive hypothesis as demonstrated with rituximab, a chimeric anti-CD20 monoclonal antibody [34].

## 3 REGULATORY GUIDANCE

Not much progress has been made in the area of predictive methods for autoimmunity and hypersensitivity induced by biopharmaceuticals since Gerhardt Zbinden's review published in 1990 [44]. Not surprisingly, regulatory guidance is very scarce at the present time. There are no comprehensive guidelines on the approaches recommended or required for predicting hypersensitivity during product development, and absolutely none on the prediction of autoimmunity. This presumably reflects the current lack of adequately standardized and validated animal models and assays that can be used during preclinical safety evaluation to predict either risk potential [3,10].

Although the ICH S6 guideline on the preclinical safety evaluation of biotechnology-derived pharmaceuticals clearly identified immunogenicity as a critical safety issue for biopharmaceuticals, and states that "the expression of surface antigens on target cells may be altered, which has implications for autoimmune potential," the document states "testing strategies may require screening studies followed by mechanistic studies to clarifly this issues." No specific recommendation was provided regarding practical modalities that could be helpful to predict either autoimmunity or hypersensitivity reactions associated with biopharmaceuticals basedon the lack of adequately standardized and validated animal models and assays, as mentioned above. This may also be the reason that ICH S8 Guideline on Immuntoxicology Studies for Human Pharmaceuticals (2006) specifically excludes guidance on drug-induced hypersensitivity and autoimmunity.

## 4 PRECLINICAL EVALUATION OF THE POTENTIAL OF BIOPHARMACEUTICALS FOR AUTOIMMUNITY

Our currently poor understanding of the mechanisms leading to autoimmunity and autoimmune diseases following drug therapy is a major hurdle. To date, no animal model or assay can reliably predict the potential, of biopharmaceuticals or pharmaceuticals, for inducing autoimmunity reactions in human subjects.

The search for autoantibodies in the sera of treated animals during standard toxicity testing cannot be recommended as it has proved to be usually negative or inconclusive [36]. Autoantibodies have sometimes been detected in the sera of animals treated with various compounds (either pharmaceuticals or

chemicals) suspected to induce autoimmunity in animals and/or human subjects. However, the design of such studies was overtly inconsistent, and results were obtained in animal strains that are not conventionally used in "regulatory toxicity" studies, such as Brown Norway (BN) rats or C57BL/6, BALB/c, and AKR mice.

The popliteal lymph node assay (PLNA) has long been proposed as a tool to investigate and predict drug and chemically induced autoimmunity [29]. In the direct PLNA, mice or rats are injected with the test article subcutaneously into one footpad and with the vehicle into the contralateral footpad. Popliteal lymph nodes are typically removed on day +7, then weighted and a weight index is calculated. The assay is considered to be positive when the weight index is $\geq 2$. The increase in weight of the treated popliteal lymph node has been assumed to reflect a pseudo-graft versus host reaction mimicking a systemic autoimmune response. Although over 100 compounds have been tested, there is no firm evidence at the present time that the direct PLNA is a reliable predictor of drug-induced autoimmunity. Importantly, no biopharmaceutical has seemingly ever been tested in this assay. The early direct PLNA procedure was subsequently refined, but the databases using any of these refinements are always quite small. The latest procedure—the modified PLNA—has been proposed as a predictive tool for drug-induced immunostimulation or sensitization [23]. Although many similarities can be found in the pathophysiological mechanisms of sensitization and autoimmunity, it is unclear to what extent, if any the modified PLNA can be useful to predict drug-induced autoimmunity.

In the clinic, autoimmunity associated with biopharmaceuticals is usually reflected by more frequent autoimmune diseases as mentioned previously. Although the mechanism(s) involved is not elucidated, an immunopharmacological effect that would reveal an underlying predisposition toward autoimmunity or a latent autoimmune disease is an attractive hypothesis. A number of rodent strains including (NZB/NZW) $F_1$, non-obese diabetic (NOD), and MRL-*lpr/lpr* mice, BN and bio-breeding (BB) rats have long been shown to develop more frequent autoimmune diseases spontaneously [30]. There is some evidence that immunomodulatory biopharmaceuticals may accelerate or exacerbate the natural course of the spontaneous disease. Thus IFN$\alpha$ [20] and the interferon inducer tilorone [39] were shown to accelerate lupus-associated mortality in (NZB/NZW) $F_1$ mice, whereas IFN$\beta$ was found to accelerate autoimmune type I diabetes in NOD mice [1]. However, major discrepancies were also noted between responses in these animal models and the clinical data. Although autoimmune type I diabetes has been reported to be an adverse effect of IFN$\alpha$ in human subjects [38], IFN$\alpha$ was found to inhibit the development of diabetes in autoimmune-prone NOD mice [32]. Similarly rIL-2, which is associated with more frequent autoimmune diseases in treated patients [38], did not exacerbate the lupus disease in (NZB/NZW) $F_1$ mice [28]. Our limited understanding of the pathophysiology of the spontaneous animal disease as well as uncertainties regarding the immunopharmacological mechanism of the

tested compound that may be dissimilar in animals compared to humans account for these discrepancies. In any case, this does not lend much support to the routine use of these models for predicting the potential of biopharmaceuticals to increase the incidence of autoimmune diseases in human subjects.

## 5 PRECLINICAL EVALUATION OF THE POTENTIAL OF BIOPHARMACEUTICALS FOR HYPERSENSITIVITY REACTIONS

Preclinical immunogenicity testing is most often limited to monitoring antibody formation in rodents and nonhuman primates, even though immune responses to therapeutic proteins in conventional animal models poorly predict immunogenicity in humans [5,42]. Indeed immunogenicity is detected in animal studies even if the protein is minimally immunogenic in humans. For instance, specific antibodies have been detected in less than 5% of patients treated with the anti-TNFα fusion protein etanercept, whereas antibodies are present in 80% to 100% of animals in preclinical toxicity studies [7]. Similarities between animal and human antibody responses, however, have also been reported. For instance, the immunogenicity of three recombinant and pituitary human growth hormones was notably similar in rhesus monkeys and in humans [45]. However, such similarities are far too much inconsistent to be considered as a reliable predictor of risk for hypersensitivity reactions in humans.

A variety of assays have been designed and validated to measure specific antibodies in the sera of treated animals. In addition efforts have been paid to minimizing the immunogenicity of therapeutic proteins. As these issues are addressed comprehensively in another chapter of this volume, the focus here will be on predicting problems that may ensue from the presence of such antibodies, namely hypersensitivity reactions. Presumably because allergic reactions have long been considered to be nonreproducible in animal models, limited efforts have been paid to designing predictive animal models until recently. Unexpectedly, the consequence is that no adequately standardized and validated model is available at the present time.

Anaphylactic models in guinea pigs have long been used [35]. In systemic anaphylaxis models guinea pigs are sensitized by one or several subcutaneous or intradermal injections of the test article sometimes mixed with an adjuvant, such as aluminum hydroxide, and then after a rest period of variable duration (usually 1–3 weeks), the animals are challenged intravenously with a non–acutely toxic dose. Clinical signs, primarily difficulties in breathing, develop within minutes after the intravenous injection and death may ensue rapidly. In cutaneous passive anaphylaxis, the pooled serum of sensitized guinea pigs is injected intradermally to naive animals and the test article mixed with a dye, such as Evan's blue, is injected intravenously. When specific antibodies are present in the pooled serum, the local antigen–antibody reaction is evidenced by a blue spot arising within a few hours. The diameter of the blue spot can

be used as a quantitation of antibody titers. Both systemic and cutaneous passive anaphylaxis models have been designed in other species as well, such as mice. Despite their extensive use for the past 40 years, these models have several major limitations. They are unsuitable for predicting hypersensitivity reactions to small molecules, namely the vast majority of pharmaceuticals, because they have to bind covalently to a carrier protein to become immunogenic (haptenization), and no procedure is available to reproduce this spontaneous phenomenon experimentally. Although large molecules, such as proteins, can act as direct immunogens and induce anaphylaxis in sensitized animals, proteins of human origin that are not immunogenic in humans consistently induce severe anaphylaxis in guinea pigs, and thus anaphylaxis guinea pig models are not at all relevant to predict the potential of human or humanized biopharmaceuticals for inducing hypersensitivity reactions in humans. The same conclusion applies whatever the animal species.

Because predicting the risk of biopharmaceuticals for inducing hypersensitivity reactions is so crucial, efforts are currently being paid to designing animal models and assays that could prove more predictive. Genetically engineered mice are promising models [22]. They include either knockout (KO) or transgenic mice. In KO mice the deletion of a gene expressing a cytokine or a chemokine considered to be pivotal in hypersensitivity responses may provide a more predictive model. KO mouse models are increasingly used in contact hypersensitivity research [40]. Transgenic mice overexpress a gene of interest, including human genes. Transgenic mice expressing the human insulin gene in pancreatic β cells were found unable to produce human insulin-specific antibodies and helpful to evaluate the potential immunogenicity of several insulin analogues [25]. Similarly a mouse model transgenic for the human IFNβ gene could be used to evaluate the immunogenicity of novel IFNβ analogues [13]. Although genetically engineered mouse models are widely thought to be promising, a number of hurdles have to be overcome before they can be recommended for routine use in the prediction of the potential of biopharmaceuticals to induce hypersensitivity reactions. Overall, these models are poorly standardized. No historical database is available, so interpreting results obtained with any new compound can be tricky. Finally, these models have been suggested to be helpful to evaluate immunogenicity, but it is not known whether these or other models can predict the risk of hypersensitivity reactions. It is indeed important to bear in mind that the clinical experience clearly showed that immunogenicity evidenced by specific antibodies in the sera of treated patients does not necessarily result in clinically patent hypersensitivity reactions.

## 6 CONCLUSION

Overall, autoimmunity and hypersensitivity reactions associated with drug therapy are of major concern in drug development and beyond [24,27]. Such

adverse events have been described in association with biopharmaceuticals. This overview of current animal models and assays demonstrates that predicting this risk during preclinical safety evaluation is hardly possible for the time being. Current research efforts in the field of hypersensitivity may result in more predictive models within a reasonable time frame. In contrast, significant progress in the field of autoimmunity is likely to take much more time.

Today, whatever the findings during animal studies of novel biopharmaceuticals, it is impossible to provide any reassurance regarding the safety of such compounds in human subjects as far as autoimmunity and hypersensitivity reactions are concerned. Therefore biopharmaceuticals should be considered high-risk compounds. Adequate follow-up of treated patients and risk management are absolutely essential. Assessing the potential of biopharmaceuticals for inducing autoimmunity and hypersensitivity reactions should not be restricted to clinical trials. Indeed, because pure red cell aplasia associated with recombinant erythropoietins was identified only after several years of clinical use, long-term postmarketing surveillance (pharmacovigilance) is crucial.

## REFERENCES


1. Alba A, Puertas MC, Carrillo J, Planas R, Ampudia R, Pastor X, Bosch F, Pujol-Borrell R, Verdaguer J, Vives-Pi M. IFN beta accelerates autoimmune type 1 diabetes in nonobese diabetic mice and breaks the tolerance to beta cells in nondiabetes-prone mice. *J Immunol* 2004;173:6667–75.
2. Atzeni F, Turiel M, Capsoni F, Doria A, Meroni P, Sarzi-Puttini P. Autoimmunity and anti-TNF-alpha agents. *Ann NY Acad Sci* 2005;1051:559–69.
3. Bala S, Weaver J, Hastings KL. Clinical relevance of preclinical testing for allergic side effects. *Toxicology* 2005;209:195–200.
4. Baert F, Noman M, Vermeire S, Van Assche G, D'Haens G, Carbonez A, Rutgeerts P. Influence of immunogenicity on the long-term efficacy of infliximab in Crohn's disease. *N Engl J Med* 2003;348:601–8.
5. Bugelski PJ, Treacy G. Predictive power of preclinical studies in animals for the immunogenicity of recombinant therapeutic proteins in humans. *Curr Opin Mol Ther* 2004;6:10–16.
6. Cheifetz A, Mayer L. Monoclonal antibodies, immunogenicity, and associated infusion reactions. *Mount Sinai J Med* 2005;72:250–6.
7. Chirino AJ, Ary ML, Marshall SA. Minimizing the immunogenicity of protein therapeutics. *Drug Discov Today* 2004;9:82–90.
8. Davidson A, Diamond B. Autoimmune diseases. *N Engl J Med* 2001;345:340–50.
9. De Bandt M, Sibilia J, Le Loet X, Prouzeau S, Fautrel B, Marcelli C, Boucquillard E, Siame JL, Mariette X, Club Rhumatismes et Inflammation. Systemic lupus erythematosus induced by anti-tumour necrosis factor alpha therapy: a French national survey. *Arthritis Res Ther* 2005;7:R545–51.
10. Descotes J. Autoimmunity and toxicity testing. *Toxicol Letts* 2000;112–3:461–5.

11. Descotes J, Vial T. Immune response in human pathology: hypersensitivity and auto-immunity. In Nijkamp FP, Parnham M, eds. *Textbook of Immunopharmacology*, 2nd ed. Basel: Birkhäuser Verlag, 2005;117–29.
12. Frost H. Antibody-mediated side effects of recombinant proteins. *Toxicology* 2005;209:155–60.
13. Hermeling S, Jiskoot W, Crommelin D, Bornaes C, Schellekens H. Development of a transgenic mouse model immune tolerant for human interferon beta. *Pharmaceut Res* 2005;22:847–51.
14. Johansson SG, Bieber T, Dahl R, Friedmann PS, Lanier BQ, Lockey RF, Motala C, Ortega Martell JA, Platts-Mills TA, Ring J, Thien F, Van Cauwenberge P, Williams HC. Revised nomenclature for allergy for global use: report of the Nomenclature Review Committee of the World Allergy Organization, October 2003. *J Allergy Clin Immunol* 2004;113:832–6.
15. Koren E, Zuckerman LA, Mire-Sluis AR. Immune responses to therapeutic proteins in humans—clinical significance, assessment and prediction. *Curr Pharmaceut Biotechnol* 2002;3:349–60.
16. Krouse RS, Royal RE, Heywood G, Weintraub BD, White DE, Steinberg SM, Rosenberg SA, Schwartzentruber DJ. Thyroid dysfunction in 281 patients with metastatic melanoma of renal carcinoma treated with interleukin-2 alone. *J Immunother* 1995;18:272–8.
17. Lim LC. Acquired red cell aplasia in association with the use of recombinant erythropoietin in chronic renal failure. *Hematology* 2005;10:255–9.
18. Lusher JM. Inhibitor antibodies to factor VIII and factor IX: management. *Sem Thromb Hemost* 2000;26:179–88.
19. Malucchi S, Sala A, Gilli F, Bottero R, Di Sapio A, Capobianco M, Bertolotto A. Neutralizing antibodies reduce the efficacy of beta IFN during treatment of multiple sclerosis. *Neurology* 2004;62:2031–7.
20. Mathian A, Weinberg A, Gallegos M, Banchereau J, Koutouzov S. IFN-alpha induces early lethal lupus in preautoimmune (New Zealand Black x New Zealand White) F1 but not in BALB/c mice. *J Immunol* 2005;174:2499–506.
21. Miossec P. Cytokines and autoimmune diseases: from the control of autoimmune diseases with anti-cytokine treatment to the induction of autoimmunity with cytokine treatment. In House RV, Descotes J, eds. *Cytokines in Human Health: Immunotoxicology, Pathology and Human Health*. Totowa NJ: Humana Press, 2007; 233–58.
22. Moser R, Quesniaux V, Ryffel B. Use of transgenic animals to investigate drug hypersensitivity. *Toxicology* 2001;158:75–83.
23. Nierkens S, Pieters R. Murine models of drug hypersensitivity. *Curr Opin Allergy Clin Immunol* 2005;5:331–5.
24. Olsen NJ. Drug-induced autoimmunity. *Best Pract Res Clin Rheumatol* 2004;18:677–88.
25. Ottesen JL, Nilsson P, Jami J, Weilguny D, Duhrkop M, Bucchini D, Havelund S, Fogh JM. The potential immunogenicity of human insulin and insulin analogues evaluated in a transgenic mouse model. *Diabetologia* 1994;37:1178–85.
26. Pichler WJ. Adverse side-effects to biological agents. *Allergy* 2006;61:912–20.
27. Ratajczak HV. Drug-induced hypersensitivity: role in drug development. *Toxicol Rev* 2004;23:265–80.

28. Ravel G, Christ M, Ruat C, Burnett R, Descotes J. Effect of murine recombinant IL-2 on the course of lupus-like disease in (NZBxNZW) F1 female mice. *Immunopharmacol Immunotoxicol* 2002;24:409–21.
29. Ravel G, Descotes J. Popliteal lymph node assay: facts and perspectives. *J Appl Toxicol* 2005;25:451–8.
30. Ravel G, Descotes J. Animal models of autoimmune diseases and their use in safety evaluation studies. *Persp Exp Clin Immunotoxicol* 2007;1:30–9.
31. Selmi C, Lleo A, Zuin M, Podda M, Rossaro L, Gershwin ME. Interferon alpha and its contribution to autoimmunity. *Curr Opin Invest Drugs* 2006;7:451–6.
32. Sobel DO, Ahvazi B. Alpha-interferon inhibits the development of diabetes in NOD mice. *Diabetes* 1998;47:1867–72.
33. Sorensen PS, Ross C, Clemmesen KM, Bendtzen K, Frederiksen JL, Jensen K, Kristensen O, Petersen T, Rasmussen S, Ravnborg M, Stenager E, Koch-Henriksen N. Clinical importance of neutralising antibodies against interferon beta in patients with relapsing-remitting multiple sclerosis. *Lancet* 2003;362:1184–91.
34. Van der Kolk LE, Grillo-Lopez AJ, Baars JW, Hack CE, Van Oers MH. Complement activation plays a key role in the side-effects of rituximab treatment. *Brit J Haematol* 2001;115:807–11.
35. Verdier F, Chazal I, Descotes J. Anaphylaxis models in the guinea-pig. *Toxicology* 1994;93:55–61.
36. Verdier F, Patriarca C, Descotes J. Autoantibodies in conventional toxicity testing. *Toxicology* 1997;119:51–8.
37. Vial T, Choquet-Kastylevsky G, Descotes J. Adverse effects of immunotherapeutics involving the immune system. *Toxicology* 2002;174:3–11.
38. Vial T, Descotes J. Clinical adverse effects of cytokines on the immune system. In House RV, Descotes J. eds. *Cytokines in Human Health: Immunotoxicology, Pathology and Human Health*. Totowa NJ: Humana Press, 2007;319–48.
39. Walker SE. Accelerated mortality in young NZB/NZW mice treated with the interferon inducer tilorone hydrochloride. *Clin Immunol Immunopathol* 1977;8: 204–12.
40. Wang B, Esche C, Mamelak A, Freed I, Watanabe H, Sauder DN. Cytokine knockouts in contact hypersensitivity research. *Cytokine Growth Factor Rev* 2003;14: 381–9.
41. Weber RW. Adverse reactions to biological modifiers. *Curr Opin Allergy Clin Immunol* 2004;4:277–83.
42. Wierda D, Smith HW, Zwickl CM. Immunogenicity of biopharmaceuticals in laboratory animals. *Toxicology* 2001;158:71–4.
43. Yates AB, DeShazo RD. Allergic and nonallergic drug reactions. *South Med J* 2003;96:1080–7.
44. Zbinden G. Safety evaluation of biotechnology products. *Drug Safety* 1990;5 (Suppl 1):58–64.
45. Zwickl CM, Cocke KS, Tamura RN, Holzhausen LM, Brophy GT, Bick PH, Wierda D. Comparison of the immunogenicity of recombinant and pituitary human growth hormone in rhesus monkeys. *Fund Appl Toxicol* 1991;16:275–87.

# Current Practices in the Preclinical Safety Assessment of Peptides

SHAWN M. HEIDEL, DVM, PhD, and TODD J. PAGE, PhD

**Contents**

1 Introduction
2 Species Selection
3 Genetic Toxicology
4 Safety Pharmacology
5 Chronic Toxicity Testing
6 Carcinogenicity Testing
7 Recombinant versus Synthetic Peptides
8 Conclusions
References

## 1 INTRODUCTION

Peptides constitute the earliest class of biopharmaceutical medicines and, perhaps, the most studied to date. Insulin has been an effective diabetes therapy for more than 75 years in the United States. Insulin has gone through an evolution that began with its isolation from animal pancreata and continues with numerous analogues being manufactured with state-of-the-art bacterial, mammalian, and yeast recombinant technology [1]. There has been an explosion of peptides in drug development, and there are currently dozens of marketed therapeutic peptides [2]. This chapter reviews the preclinical development experience across this class of important biopharmaceuticals, highlighting what has gone before and some considerations for the future.

A specific definition of what constitutes a peptide is difficult to find. Most people involved in developing biopharmaceuticals would define a peptide as

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

a small biopharmaceutical or a traditional biopharmaceutical, but would then elaborate on development issues for peptides manufactured by recombinant rather than synthetic chemical processes. For the purposes of this chapter, and the peptides surveyed in Table 1, peptides are defined as those biopharmaceuticals with 5 to 100 amino acids. A distinction will also be made with regard to recombinant versus chemical manufacturing and will be discussed with respect to available guidance, or lack thereof, and experience.

The road map for developing peptides is by association to existing biopharmaceutical and/or new chemical entity guidance documents rather than a guidance that is specific for peptides. The International Conference on Harmonization (ICH) has published a guidance specific for preclinical development of biopharmaceuticals (ICH S6), which is appropriate for most peptides. However, other safety guidance, including regional guidance, may need to be considered dependent on the manufacturing process, the inclusion of unnatural amino acids, or the intended patient population. For instance, there is regional guidance for insulin and parathyroid hormone in Europe and the United States, respectively [3,4]. Regulatory approaches to biopharmaceuticals in Japan are covered in separate chapter (Chapter Implementation of ICH S6: Japanese Perspective). A review of approved peptide products revealed preclinical studies or parameters that are a blend of typical biopharmaceutical and new chemical entity (NCE) programs (Table 1). Therefore there is a blend of studies recommended for peptides based on guidances developed for biopharmaceuticals that includes ICH S6 and those for NCEs (e.g., ICH S1-S5, S7-S8, and M3) (see www.ich.org and www.fda.gov).

## 2 SPECIES SELECTION

Species selection for most peptides has taken into account the pharmacological activity of the peptide in the test species, which is the primary concern for species selection of biopharmaceuticals. Species selection for NCEs is based mostly on in vitro metabolite comparisons between the preclinical species and humans with an assumption of comparable cross-species biologic activity. In contrast to NCEs, biopharmaceuticals are degraded to naturally occurring amino acids rather than metabolized by p450 enzymes to xenobiotic metabolites. Therefore in vitro metabolism studies are not done early in development for peptides. Rather, in vitro or in vivo pharmacology studies are conducted to demonstrate species relevance prior to preclinical toxicology studies. For example, measuring a reduction in glucose following insulin administration would be a demonstration of pharmacological activity. In many cases pharmacology is not as simple as measuring a decrease in glucose levels. When the mechanism of action is more complex, in vitro cell based assays can be used to demonstrate pharmacology. Using a larger, protein biologic to provide an example, interference with a cytokine-induced cell proliferation assay can be used to demonstrate cross-species pharmacology. In this model an antibody

**TABLE 1 Overview of studies used to support peptide development**

| Peptide | Class | Safety Pharmacology[a]/Chronic Toxicity | Carcinogenicity Testing | Genetic Toxicity | Method of Manufacture | Approval Date |
|---|---|---|---|---|---|---|
| Bivalrudin (Angiomax, Angiox) [10,11] | Cardiac | Safety pharmacology studies appear to have been done separately<br>1 month (rat)<br>1 month (cyno)[c] | No | Standard battery[b] | Synthetic | FDA-2000<br>EMEA-2004 |
| Cetrorelix (Cetrotide) [26,27] | Hormone | Safety pharmacology studies appear to have been done separately<br>6 month (rat)<br>6 month (dog) | No | Standard battery | Synthetic | FDA-2000<br>EMEA-1999 |
| Eptifibatide (Integrelin) [12,13] | Cardiac | Separate safety pharmacology studies<br>One month (rat)<br>One month (cyno) | No | Standard battery | Synthetic | FDA-1998<br>EMEA-1999 |
| Enfuvirtide (Fuzeon) [15,16] | Antiviral | Safety pharmacology included in toxicology studies<br>6 month (rat)<br>9 month (cyno) | No | Standard battery | Synthetic | FDA-2003<br>EMEA-2003 |
| Nesiritide (Natrecor) [28] | Cardiac | Cardiac safety pharmacology separate<br>2 week (rat)<br>2 week (monkey) | No | Ames only | Recombinant | FDA-2001 |

**TABLE 1** ***Continued***

| Peptide | Class | Safety Pharmacology[a]/Chronic Toxicity | Carcinogenicity Testing | Genetic Toxicity | Method of Manufacture | Approval Date |
|---|---|---|---|---|---|---|
| Ziconotide (Prialt) [29,30] | Pain | Non-GLP safety pharmacology studies, GLP toxicology studies used to support non-GLP studies<br>14 day (monkey)<br>28 day (rat)<br>42 day (dog) | No | Standard battery | Synthetic | FDA-2004<br>EMEA-2005 |
| Exenatide (Byetta) [31,32] | Hormone | Separate CNS study (mice), separate CV study (monkey)<br>182 day (mouse)<br>91 day (rat)<br>273 day (monkey) | 104 week (rat), benign thyroid C cell adenoma<br>104 week mouse | Standard battery | Synthetic | FDA-2005<br>EMEA-2006 |
| Leuprolide (Lupron-Depot) [33] | Hormone | ND | 2 year (rats, mice), benign pituitary hyperplasia and adenoma (rat, 24 months, 100× clinical dose) | ND | Synthetic | FDA 1997 |
| Calcitonin-Salmon (Fortical, Forcaltonin) [34,35] | Hormone | ND | 1 year (SD and Fisher rats) nonfunctioning pituitary adenomas (130–160× clinical dose) | Standard battery | Recombinant | FDA-2005<br>EMEA-1999 |
| Octreotide (Sandostatin) [36] | Hormone | ND | 99 weeks (mice)<br>116 weeks (rats) | Were performed, specifics not given | Synthetic | FDA-1988 |

| | | | | | | |
|---|---|---|---|---|---|---|
| rhPTH(1-34) (Forteo, Forsteo) [37,38] | Hormone | Separate safety pharmacology studies in rat and dog, studies in cyno were included in chronic tox<br>6 month (rat)<br>12 month (monkey) | 2 year (rat)<br>Observed malignant metastatic osteosarcoma. | Standard battery | Recombinant | FDA-2002<br>EMEA-2003 |
| Insulin lispro (Humalog) [39] | Insulin analogue | Safety pharmacology studies appear to have been done separately.<br>12 month (rat)<br>12 month (dog) | No<br>in vitro studies to assess DNA synthesis stimulation using HepG2 cells | Standard battery | Recombinant | FDA-1996<br>EMEA-1996 |
| Insulin glulisine (Apidra) [7,8] | Insulin analogue | Separate CV and respiratory studies<br>12 month (rat)<br>6 month (dog) | No, but the 12-month chronic study was designed to investigate the carcinogenic potential of the compound | Standard battery | Recombinant | FDA-2004<br>EMEA-2004 |
| Insulin glargine (Lantus) [40,41] | Insulin analogue | Separate safety pharmacolgy. CNS, CV, pulmonary, renal<br>6 month (rat)<br>6 month (dog) | 2 year (mouse)<br>2 year rat | Standard battery | Recombinant | FDA-2000<br>EMEA-2000 |
| Insulin detemir (Levemir) [6,20] | Insulin analogue | 6 month (rat)<br>12 month (dog) | No<br>in vitro mitogenic potential (CHO-K1, MCF-7, human osteosarcoma, L6-hIR) | Standard battery | Recombinant | FDA-2005<br>EMEA-2004 |
| Insulin aspart (NovoLog, NovoRapid) [42,43] | Insulin analogue | 12 month (rat)<br>12 month (dog) | No<br>Although mammary tumors were observed in the 12 month (rat) study | Standard battery | Recombinant | FDA-2000<br>EMEA-1999 |

*Source*: All data for this table was taken from Label, FDA Summary Basis for Approval, and EPAR (European Public Assessment Report) documents.

[a]Safety Pharmacology ICHS7.

[b]Genetic Toxicity Standard Battery ICHS2A.

[c]Cynomolgus monkey (cyno).

to the human cytokine is assessed for its ability to bind to the intended test species cytokine homologue by measuring inhibition of cytokine-induced cell proliferation. This demonstration of pharmacology is usually sufficient to justify the use of the test species for preclinical safety evaluation of the anti-human cytokine antibody.

Toxicology testing for peptides is typically completed in two species. In the peptide development programs that were reviewed, rats were the most common rodent and nonrodents were about equally split between monkeys and dogs. All of the 12 peptides where data were available were tested in at least two species. In some cases three species were evaluated (e.g., exenatide, ziconatide) (Table 1). This is consistent with ICH S6, which indicates two species should be used, if appropriate. However, many biopharmaceuticals are evaluated only in nonhuman primates because of the lack of pharmacology or the development of a significant immune response in rodents. If a peptide is pharmacologically active in only a single species, then it is appropriate to use only that species for toxicology testing. Conducting studies in a species that lacks pharmacology would not provide useful information, and may in fact complicate the extrapolation of the animal data to humans. For example, an absence of findings in a species that is not pharmacologically responsive may be misinterpreted as a lack of potential safety issues, even though high-dose pharmacology was not evaluated. Like other biopharmaceuticals, pharmacology in the test species should be the key determinant of species selection for peptides primarily because peptides are not metabolized like NCEs.

## 3 GENETIC TOXICOLOGY

Genetic toxicology studies were completed for almost all peptides in the survey, which is a requirement for NCEs (per ICH S2) but not necessary and/or appropriate for most biopharmaceuticals (per ICH S6) (see Chapter The Principles of ICH S6 and the Case-by-Case Approach). The main reason that these studies are not needed for biopharmaceuticals is the inability of biopharmaceuticals to cause direct DNA damage. Most biopharmaceuticals do not cross cell or nuclear membranes to access DNA. Moreover direct DNA damage is caused by chemically reactive molecules that bind or associate with DNA, which is not expected for peptides. While scientific rationale and current guidance documents clearly indicate that genetic toxicology tests are not required for biologics, they have been completed for most approved peptides. The likely reason for this paradox is that most tests were completed prior to the implementation of ICH S6 in 1997. However, these tests have been completed for some more recently approved peptides. The rationale for this is not clear.

A pivotal publication of 78 biopharmaceuticals provides data supporting the lack of positive genotoxicity with peptides. The publication concluded, "genotoxicity testing is generally inappropriate and unnecessary" [5]. Additional support for this position comes from the regulatory agencies; in a 2004 review

the EMEA stated, "Although not applicable to biotechnology-derived pharmaceuticals, insulin detemir was tested for gene mutations in bacteria and for chromosome aberrations in vitro and in vivo. As expected, there was no evidence of genotoxic potential in any of these tests" [6]. If impurities from the manufacturing process have not been previously evaluated in genetic toxicology studies, they may need additional testing. In those cases it is appropriate to test only the impurity rather than the peptide. This does not imply that genetic toxicology studies are necessary for all synthetically produced peptides, only that it needs to be evaluated with consideration given to impurities from the manufacturing process rather than the peptide. The lack of genetic toxicity risk of the peptide is the same regardless of whether the peptide is produced by recombinant or synthetic methods. Based on good scientific rationale and data, genetic toxicology studies should not be routinely conducted for peptides.

Genetic toxicology testing of peptides containing "nonnatural" amino acids has been controversial. One conceivable genetic risk would be incorporation of these unique amino acids into proteins. Even if this were possible, the likelihood of incorporation into a protein critical for DNA replication or repair seems remote. In addition incorporation of a nonnatural amino acid into host protein(s) in and of itself would not be sufficient for concern. Malfunction of the protein to increase inherited mutations would also be required. Another conceivable risk of the nonnatural amino acid is the potential metabolism of the unique amino acid to a genotoxic metabolite. Although these rationale may provide a scientific justification for not performing genetic toxicity testing, in many cases studies may still be needed to support regulatory concerns like those in Japan (see Chapter Implementation of ICH S6: Japanese Perspective). Given the potential scientific and regulatory concerns, it may be prudent to complete genetic toxicology testing of nonnatural amino acids.

Some biopharmaceuticals may cause cellular proliferation that could increase the likelihood of inheriting a spontaneous mutation. However, genetic toxicology studies are not designed to evaluate these concerns. The potential for inducing a mitogenic response can be addressed in other toxicology studies.

## 4 SAFETY PHARMACOLOGY

Separate safety pharmacology studies were completed for many peptides in the survey. Separate safety pharmacology studies are usually performed for NCEs but not biopharmaceuticals. Since rodents are typically used for stand-alone central nervous system and respiratory studies, the usefulness of these studies for NCEs is based on the assumption that rodents are pharmacologically responsive or have some of the same metabolites as would be expected in humans. The lack of concern for metabolism and the significant reduction or absence of pharmacology in rodents with most biopharmaceuticals eliminates the relevance of these studies. In contrast to the larger

biopharmaceuticals, many of which are species specific, peptides are frequently pharmacologically active in rodents, and therefore useful data can be generated in separate safety pharmacology studies. However, as with large biopharmaceuticals, an absence of pharmacological activity of the peptide in rodents should similarly preclude conduct of these studies. When rodent studies are not appropriate, evaluations are usually done within a repeated dose toxicology study in nonrodents. For example, quantitative electrocardiograms (ECGs), detailed neurologic and behavioral evaluations, and qualitative measurements of respiratory rate and depth assess effects on the cardiovascular, central nervous system, and respiratory systems, respectively. Pharmacological activity should determine the appropriate species for safety pharmacology studies, which is consistent with ICH S6 as well as other ICH guidance.

Findings in safety pharmacology evaluations of peptides are generally consistent with anticipated pharmacology. Because the pharmacology of most peptides would not be predicted to affect safety pharmacology, no findings were observed in most studies. Other findings, such as decreased blood pressure, heart rate, breathing rate, and increased QTc for insulins were due to hypoglycemia and considered exaggerated pharmacology based on predicted pharmacology [7,8].

## 5 CHRONIC TOXICITY TESTING

The duration of chronic toxicity testing for peptides is driven primarily by their intended indication and duration of treatment. Of the peptides studied, the duration of the pivotal repeat dose toxicology studies, ranges from a low of two weeks, for neseritide, to a high of one year for many of the insulin analogues. For the 9 chronic use peptides included in our analysis, chronic toxicology studies of 6, 9, or 12 month durations were completed for 3, 2, or 4 of them, respectively. This is more consistent with guidance for NCEs, which typically have 9- or 12-month durations due to ICH S4 guidance. Most other chronic use biopharmaceuticals have completed chronic studies of 6-month duration, which is consistent with ICH S6 guidance [9].

Eptifibatide and bivalrudin, two peptides tested for 28 days in the chronic toxicity studies, provide interesting case studies [10–13]. Bivalrudin is a linear 20 amino acid peptide that is derived from hirudin, an anticoagulant found in the saliva of leeches, and is intended for use during invasive intravascular procedures. Eptifibatide is a cyclic heptapeptide that is an RGD (arginine, glycine, aspartic acid) mimetic that acts as an inhibitor of platelet aggregation and is indicated for the treatment of acute coronary syndrome. As a class of drugs, RGD mimetics are based on a sequence (RGD) found in fibronection that mediates its interactions with the integrins [14]. These observations illustrate that it is not the structure (linear vs. cyclic) or the origin (nonhuman vs. human) that determines the length of the chronic toxicology studies for peptides but the intended indication and duration of treatment.

One of the most striking observations that can be made of the peptides as a class is their remarkable safety profile. Of all of the peptides currently on the market that were investigated for this analysis (Table 1), there was only one observation in the chronic toxicology studies that was not linked to exaggerated pharmacology. Admittedly, our analysis is biased by the fact that only approved compounds were included; however, this data set is still very instructive. The observation in question was from enfuvirtide, a 36 amino acid linear peptide derived from the HIV-1 glycoprotein gp41 [15,16].

Enfuvirtide interferes with viral attachment and entry into target cells by disrupting the interaction of viral gp120 with target cell receptors (CD4, CXCR4, CCR5). In the repeat-dose toxicology studies thymic lymphocyte depletion was observed in a 9-month cynomolgus study. However, this finding was not considered to be adverse as it was attributed to exaggerated pharmacology. Additionally, microgranuloma and inflammatory loci were observed in the lungs of rats after 28 days of treatment. Presumably, these findings were related to the affinity of the peptide for the formyl peptide receptor (FPR), which has been demonstrated to mediate an inflammatory response at the site of bacterial infection via the recruitment of phagocytic cells. The pharmacological basis for these observations is being investigated by the manufacturer postapproval.

As our analysis demonstrates, in almost all cases the toxic effects of peptides can be predicted from their inherent pharmacological properties. It would follow then that one of the most important steps in the successful development of a peptide is selection of the appropriate pharmacologically responsive species for testing. Moreover, since the pharmacological effects of biologics are usually predictable and observed in studies of 6-month duration or less, the utility of longer studies is not readily apparent.

## 6 CARCINOGENICITY TESTING

The pharmacology of the peptide, duration of treatment, and patient population should be taken into consideration in designing and/or evaluating the need for carcinogenicity testing. As previously discussed, peptides are not expected to be genotoxic and therefore are not considered complete carcinogens [17]. The main concern for peptides is the potential for mitogenesis or tumor promotion.

ICH S6 indicates that carcinogenicity studies are not generally needed for endogenous substances given essentially as replacement therapy, particularly where there is previous clinical experience with similar products. The issue of carcinogenicity testing for peptides is complicated by the fact that most approved peptide drugs are hormone or hormone analogues. Supraphysiological doses of a hormone or biopharmaceutical with hormonal properties would be anticipated to cause hyperplasia, and therefore there would be a cause for concern for neoplasia.

Human parathyroid hormone (1-34) is an example of a peptide that is administered at supraphysiological doses compared to naturally secreted human parathyroid hormone (PTH). Substantial increases in bone mass, consistent with the pharmacology of PTH 1-34, were observed in a two-year carcinogenicity study in rats. In addition bone tumors were observed at all doses [18]. The data from this initial study suggested that the tumors resulted from the long duration of treatment coupled with the exaggerated pharmacological response in rats to daily treatment with PTH 1-34. Therefore a second study was conducted to more fully characterize the dose–response and duration of treatment on the formation of bone tumors. This study identified a no-effect level and quantified the effect on bone mass using computed tomography and histomorphometry [19]. These studies provided data indicating important differences between the pharmacological response in rats compared to human clinical data, which suggests the increased incidence of bone neoplasia in rats is likely not predictive of an increased risk of bone cancer in adult osteoporotic humans treated with PTH 1-34 for a limited duration. Moreover these studies demonstrate the importance of understanding the species differences for interpreting human risk.

In cases where exaggerated pharmacology might be predicted to give rise to positive findings in carcinogenesis studies, it can be very helpful, when available, to include the natural human analogue that the peptide drug is derived from for comparison purposes. This will assist in determining the relative risk of the novel peptide analogue compared to the normal analogue. This approach has been important for the development of many insulin analogues because of the propensity for mammary gland hyperplasia. For example, there was an observation of increased proliferation in the mammary gland of rats treated with suprapharmacological doses of insulin detemir [6,20]. When the study was repeated using suprapharmacological doses of human insulin as a comparator, it was found that the increases in mammary gland proliferation were similar between human insulin and insulin detemir. Therefore the risk assessment for tumors with this insulin analogue was the same as that for normal insulin, demonstrating the usefulness of including the normal human orthologue as a comparator.

In vitro studies are frequently useful for evaluating the mitogenic properties of peptides. The European guidance for insulin analogues promotes in vitro testing to determine the mitogenicity of new analogues compared to normal human insulin and AspB10, an analogue with increased mitogenicity and tumorigenic properties [21,22]. The absence of an increase in mitogenicity would decrease the cause for concern and likely eliminate the need for a two-year carcinogenicity study with the novel analogue. In cases where a peptide is not pharmacologically active in rodents, in vitro mitogenic assays may be critical for the assessment for tumor promotion. Another useful test for determining the ability to promote tumors is to evaluate multiple tumor types for the receptor target. If tumors do not contain the receptor, then there would be no risk of tumor promotion. If tumors contain the receptor, it would be important to evaluate the ability of the peptide to promote mitogenesis

in comparison with the normal human orthologue, as was done with insulin analogues. Comparable mitogenicity would indicate limited risk with novel analogues given at physiological levels as replacement therapy.

The Carcinogenicity Assessment Committee (Enfuvirtide, Summary Basis for Approval, Pharmacology, 2003, http://www.fda.gov/cder/foi/nda/2003/021481_fuzeon_review.htm) notes in the FDA approval documents for enfuviritide highlight considerations for carcinogenicity testing. The strongest arguments not to test were (1) the lack of structural alerts in a peptide, (2) the absence of hyperplasia in the chronic toxicity studies, and (3) the polypeptides with known carcinogenic properties that are hormones or demonstrate hormonal activity [16]. The arguments to test were (1) chronic inflammation was observed at the injection site, (2) the indication was chronic use, and (3) the product is a new therapeutic. The most compelling reason for testing was thought to be chronic inflammation, but this is easily monitored and would be managed in the clinic. Other points to consider are whether the peptide binds a cell-surface receptor to elicit signal transduction. The lack of signal transduction or immunosuppression would decrease the cause for concern for carcinogenicity. One theoretical example would be a peptide that binds a plasma-derived cytokine with little consequence for immunity. The final FDA decision was to waive the need for carcinogenicity testing with enfuviritide.

Even therapies that are potent immunosuppressive agents present challenges for carcinogenicity assessment. This is mainly due to their pharmacology, which makes it highly likely that they will act as tumor promoters. Therefore it could be assumed that these therapies would present a carcinogenic risk without a two-year study. In addition completing a full two-year study in rodents at clinically relevant doses has proved difficult because of rodent sensitivity to the pharmacological effects, which results in low survival. The pharmacological properties coupled with difficulties in completing lifetime studies in rodents has led to the suggestion that alternative models to two-year carcinogenicity studies might be more appropriate for risk assessment of immunosuppressive drugs [23].

A threshold for the carcinogenicity of hormones is widely accepted by regulatory agencies [24,25]. Therefore using historical practices for peptide hormones is not appropriate for establishing broad guidelines on the need for carcionogenicity testing for all peptides. While it is difficult to understand a scientific rationale for carcinogenesis testing for classes of peptides that do not have mitogenic properties (e.g., antiviral, cardiac peptides), in other cases it would seem to be warranted because the downstream pharmacology may be mitogenic (GnRH analogues) as demonstrated by multiple tumors in 2-year rodent carcinogenicity studies with analogues, such as Leu prolide [33]. For these reasons the need for carcinogenicity testing needs to carefully evaluated, primarily taking into account the expected pharmacology of the molecule.

Carcinogenicity testing of chronic use peptides has been evaluated in a number of ways, ranging from in vitro assays coupled with an evaluation of hyperplasia in repeat-dose studies (insulin detemir) to full two-year studies

(i.e., PTH 1-34). The type of carcinogenicity assessment is dependent on the accumulated database for the therapeutic class, including data from positive and negative controls that can be used to gauge risk.

## 7 RECOMBINANT VERSUS SYNTHETIC PEPTIDES

The differences between the manufacture of recombinant and synthetic peptides typically should not result in a different battery of toxicology studies. Pharmacology should be the primary determinant of the appropriate studies and thus the preclinical development programs would be expected to be similar regardless of the method of manufacture. A notable exception would be the level of process-related impurities that might be introduced during synthetic production. In those cases the impurities may need to be qualified in genetic toxicology tests and/or other tests. In the absence of novel impurities, the development considerations should be the same for either manufacturing method and should primarily focus on potential toxicities associated with exaggerated pharmacology.

## 8 CONCLUSIONS

Peptides are a unique class of biopharmaceuticals and have been developed with a battery of toxicology studies that is a blend of those traditionally used for NCEs and biopharmaceuticals. Historically peptides were developed much like NCEs, probably owing to the experience of the sponsor and the various regulatory agencies, which was predominantly NCE focused. The pharmacology of the peptide should be the primary consideration for toxicology testing. This is mainly due to the lack of potentially reactive metabolites or competition for metabolic clearance that are important drivers of NCE testing. Exaggerated pharmacology is the main concern for peptides and thus should be the main determinant in selecting the appropriate battery of toxicology studies.

The development of peptides will likely continue to appear somewhat different from larger biopharmaceuticals. This is mostly driven by the rodent being a pharmacologically responsive species for many peptides. Therefore many peptides will continue to have repeat-dose toxicology studies, separate safety pharmacology studies, and carcinogenicity assessments in rodents. The rationale for not performing similar studies with larger biopharmaceuticals is often based on the absence of pharmacology in rodents and/or the induction of an immune response that eliminates exposure needed for a meaningful evaluation. Pharmacology in rodents has led to the majority of the differences in toxicology studies for peptides compared to other biopharmaceuticals.

Many marketed peptides are chronic use hormones that have undergone carcinogenicity testing. The potential for superpharmacologic hyperplasia is a common concern for hormones and emphasizes the need for assessment of

carcinogenic potential. This has led to full two-year carcinogenicity assessments for some peptide hormones. Recent advances in the understanding of specific pharmacological classes, such as insulin analogues, have provided in vitro tools to assess the risk of mitogenicity without completing a two-year study. As knowledge and technical capabilities progress, new methods for carcinogenicity assessments will arise and should be considered. Many other biopharmaceuticals do not have specific pharmacology requiring an extensive two-year carcinogenicity bioassay. Examples are those peptides that do not bind cell-surface receptors or are used only for acute treatment. The most appropriate assessment for carcinogenic potential of peptides should be justified and may not include a two-year study in rodents.

As classes of peptides have evolved, it has become clear that sponsors and regulatory agencies consider pharmacological activity as the primary concern and therefore follow the principles outlined in the ICH S6 guidance. This approach is science-driven based on the lack of potentially reactive metabolites and the emphasis on exaggerated pharmacology as the primary concerns for improving the predictive value of safety evaluation programs. For new peptide classes in development, as with any new medicine, the key is to develop the appropriate hypotheses that dictate the appropriate scientific considerations and studies.

## REFERENCES


1. Chance RE, Frank BH. Research, development, production, and safety of biosynthetic human insulin. *Diabet Care* 1993;16(Suppl 3):133–42.
2. Marx V. Watching Peptide Drugs Grow Up. *Chem Eng News* 2005;83(11):17–24.
3. EMEA. Points to consider document on the non-clinical assessment of the carcinogenic potential of insulin analogues, 2001. http://www.emea.europa.eu/pdfs/human/swp/037201en.pdf
4. FDA. Development of Parathyroid Hormone for the Prevention and Treatment of Osteoporosis, 2000. http://www.fda.gov/cder/guidance/3789dft.htm
5. Gocke E, Albertini S, Brendler-Schwaab S, Muller L, Suter W, Wurgler FE. Genotoxicity testing of biotechnology-derived products: report of a GUM task force. Gesellschaft fur Umweltmutationsforschung. *Mutat Res* 1999;436(2): 137–56.
6. Insulin Detemir. Summary Basis for Approval, EPAR, 2006. http://www.emea.europa.eu/humandocs/PDFs/EPAR/levemir/093604en6.pdf
7. Insulin Glulisine. Summary Basis for Approval, Pharmacology Review, 2004. http://www.fda.gov/cder/foi/nda/2004/21-629_Apidra.htm
8. Insulin Glulisine. Summary Basis for Approval, EPAR, 2006. http://www.emea.europa.eu/humandocs/PDFs/EPAR/apidra/121804en6.pdf
9. Clarke J, Hurst C, Martin P, Vahle J, Ponce R, Mounho B, Heidel S, Andrews L, Reynolds T, Cavagnaro J. Duration of chronic toxicity studies for biotechnology-derived pharmaceuticals: Is six months still appropriate? *Regul Toxicol Pharmacol* 2008;50:2–22.

10. Bivalrudin. Summary Basis for Approval, Pharmacology Review, 2000. http://www.fda.gov/cder/foi/nda/2000/20873_Angiomax.htm
11. Bivalrudin. Summary Basis for Approval EPAR, 2004. http://www.emea.europa.eu/humandocs/PDFs/EPAR/angiox/103304en6.pdf
12. Eptifibatide. Summary Basis for Approval, Pharmacology Review, 1998. http://www.fda.gov/cder/foi/nda/98/20718_Integrilin.htm
13. Eptifibatide. Summary Basis of Approval, EPAR, 2004. http://www.emea.europa.eu/humandocs/PDFs/EPAR/Integrilin/110099en6.pdf
14. Humphries MJ, Olden K, Yamada KM. A synthetic peptide from fibronectin inhibits experimental metastasis of murine melanoma cells. *Science* 1986; 233(4762):467–70.
15. Enfuvirtide. Summary Basis for Approval, Pharmacology Review, 2003. http://www.fda.gov/cder/foi/nda/2003/021481_fuzeon_review.htm
16. Enfuvirtide. Summary Basis for Approval, EPAR, 2004. http://www.emea.europa.eu/humandocs/PDFs/EPAR/fuzeon/169503en6.pdf
17. Pitot HC, Dragan YP, *Casarett and Doull's Toxicology*, 6 ed. Klaassen C, ed. 2001; 266–78.
18. Vahle JL, Sato M, Long GG, Young JK, Francis PC, Engelhardt JA, Westmore MS, Linda Y, Nold JB. Skeletal changes in rats given daily subcutaneous injections of recombinant human parathyroid hormone (1–34) for 2 years and relevance to human safety. *Toxicol Pathol* 2002;30(3):312–21.
19. Vahle JL, Long GG, Sandusky G, Westmore MS, Ma YL, Sato M. Bone neoplasms in F344 rats given teriparatide [rhPTH(1–34)] are dependent on duration of treatment and dose. *Toxicol Pathol* 2004;32(4):426–38.
20. Insulin Detemir. Summary Basis for Approval, Pharmacology Review, 2005. http://www.fda.gov/cder/foi/nda/2005/021-536_LevemirTOC.htm
21. Dideriksen LH, Jorgensin LN, Drejer K. Carcinogenic effect on female rats after 12 months administration of the insulin analogue B10 Asp. *Diabetes* 1992;41:143A.
22. Milazzo G, Sciacca L, Papa V, Goldfine ID, Vigneri R. ASPB10 insulin induction of increased mitogenic responses and phenotypic changes in human breast epithelial cells: evidence for enhanced interactions with the insulin-like growth factor-I receptor. *Mol Carcinog* 1997;18(1):19–25.
23. Hastings KL. Assessment of immunosuppressant drug carcinogenicity: standard and alternative animal models. *Hum Exp Toxicol* 2000;19(4):261–5.
24. Jacobs A, Jacobson-Kram D. Human carcinogenic risk evaluation. Part III: Assessing cancer hazard and risk in human drug development. *Toxicol Sci* 2004; 81(2):260–2.
25. Silva Lima B, Van der Laan JW. Mechanisms of nongenotoxic carcinogenesis and assessment of the human hazard. *Regul Toxicol Pharmacol* 2000;32(2);135–43.
26. Cetrorelix. Summary Basis for Approval, Pharmacology Review, 2000. http://www.fda.gov/cder/foi/nda/2000/21-197_Cetrotide.htm
27. Cetrorelix. Summary Basis for Approval, EPAR, 2003. http://www.emea.europa.eu/humandocs/PDFs/EPAR/Cetrotide/297998en6.pdf
28. Nesiritide. Summary Basis for Approval, Pharmacology Review, 2001. http://www.fda.gov/cder/foi/nda/2001/20-920_Natrecor.htm

29. Ziconatide. Summary Basis for Approval, Pharmacology Review, 2004. http://www.fda.gov/cder/foi/nda/2004/21-060_Prialt.htm
30. Ziconatide. Summary Basis for Approval, EPAR, 2006. http://www.emea.europa.eu/humandocs/PDFs/EPAR/Prialt/14122704en6.pdf
31. Exenatide. Summary Basis for Approval, Pharmacology, 2005. http://www.fda.gov/cder/foi/nda/2005/021773_ByettaTOC.htm
32. Exenatide. Summary Basis for Approval, EPAR, 2006. http://www.emea.europa.eu/humandocs/PDFs/EPAR/byetta/H-698-en6.pdf
33. Leuprolide. Label, 1993. http://www.fda.gov/cder/foi/label/2004/20263slr024_lupron_lbl.pdf
34. Calcitonin. Summary Basis of Approval, EPAR, 1999. http://www.emea.europa.eu/humandocs/PDFs/EPAR/Forcaltonin/238798en6.pdf
35. Calcitonin. Label, 2005. http://www.fda.gov/cder/foi/label/2005/021406lbl.pdf
36. Octreotide. Label, 2005. http://www.fda.gov/cder/foi/label/2005/019667s050lbl.pdf
37. Forsteo. Summary Basis For Approval, EPAR, 2004. http://www.emea.europa.eu/humandocs/PDFs/EPAR/forsteo/659802en6.pdf
38. Forteo. Summary Basis For Approval, Pharmacology Review, 2002. http://www.fda.gov/cder/foi/nda/2002/21-318_FORTEO_Pharmr_P1.pdf
39. Insulin Lispro. Summary Basis for Approval, EPAR, 2004. http://www.emea.europa.eu/humandocs/PDFs/EPAR/Humalog/060195en6.pdf
40. Insulin Glargine. Summary Basis for Approval, Pharmacology Review, 2000. http://www.fda.gov/cder/foi/nda/2000/21081_Lantus.htm
41. Insulin Glargine. Summary Basis for Approval, EPAR, 2003. http://www.emea.europa.eu/humandocs/PDFs/EPAR/Lantus/061500en6.pdf
42. Insulin Aspart. Summary Basis for Approval, Pharmacology Review, 2000. http://www.fda.gov/cder/foi/nda/2000/20-986_NovoLog.htm
43. Insulin Aspart. Summary Basis for Approval, EPAR, 2004. http://www.emea.europa.eu/humandocs/PDFs/EPAR/Novorapid/272799en6.pdf

# Enzyme Replacement Therapies

LAURA ANDREWS, PhD, DABT, and
MICHAEL O'CALLAGHAN, DVM, PhD, MRCVS

**Contents**

1 Introduction
2 Historical Background
  2.1 Pathophysiology of LSDs
  2.2 Therapies for LSDs
  2.3 LSDs and Models
  2.4 Delivering ERTs
  2.5 Defeating the Pathophysiology of LSDs
3 Current Technologies
4 Regulatory Guidance
  4.1 Cerezyme®
  4.2 Fabrazyme®
  4.3 Aldurazyme®
  4.4 Myozyme®
  4.5 Acid Sphingomyelinase
5 Conclusion
  References

## 1 INTRODUCTION

Enzyme replacement therapy (ERT) is a therapeutic approach in which the specific enzyme that is absent or inactive in affected individuals is replaced with a functional enzyme molecule. Pancreatic enzyme preparations of porcine or bovine origin have been available in the United States for treatment of exocrine pancreatic insufficiency (EPI) in children and adults with cystic fibrosis and chronic pancreatitis since before the enactment of the Federal Food, Drug and Cosmetic Act of 1938 (ref FDA guidance on EIP April 2004). A

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

modified version of adenosine deaminase (ADA), polyethylene modified ADA (PEG-ADA), was approved in 1990 as an enzyme replacement therapy for immunodeficiency due to ADA deficiency. This chapter addresses the issues facing development of enzyme replacement treatments for the specific group of mutational conditions termed lysosmal storage diseases (LSDs). First, some of the pathophysiological features of LSDs will be described, with a discussion of the obstacles to delivering therapeutic proteins to diseased lysosomes. The patchy past history of attempts to treat these long considered "untreatable" conditions will be covered, before moving to the most recent decade of significant advances in enzyme replacement therapy (ERT) that has opened up a new era of hope for LSD sufferers. Some of the key advances that have led to successful commercial production of enzymes for ERT will be addressed. Particular attention will be given to production scale up and the challenges of providing in vitro and in vivo verification of product consistency through the period from preclinical studies to mature commercial production. The changing regulatory framework that has governed the development of ERTs with reference to obtaining worldwide licensure, a necessity for participating companies if these therapies are to be commercially viable, will be covered. We will address some of the unique issues posed by LSDs, in particular, the emerging recognition that substrate load in animal models and patients significantly complicates the translation of preclinical safety (and other) findings into a human risk–benefit equation. In addition we will attempt to illustrate some of the most pertinent challenges and developments that have led to the recent expansion of therapeutic options for many LSD patients who once faced a uniformly dire prognosis on diagnosis of their disease. Finally, we will address the current status of ERTs as they pertain to the LSDs as a group, noting alternative and emerging ancillary therapies such as chaperones and gene therapy. We will also discuss two major remaining obstacles: notably the problem of how to gain access to the central nervous system for the more than 50% of LSDs with neurologic manifestations, and how the biopharmaceutical industry and regulatory agencies are to address the ever-diminishing pool of patients with rarer LSDs now that effective treatments are available for the relatively more populous conditions, Gaucher, Fabry, MPS I and II, and Pompe.

## 2 HISTORICAL BACKGROUND

LSDs are deceptively classified as a single group of diseases, based simply on the common finding of accumulated macromolecular material in the lysosomes of various tissues. While there are some similarities among certain LSDs such as the mucopolysaccharidoses (MPSs1-7) most of the named LSDs are highly dissimilar by pathophysiology, organ involvement, and therapeutic challenge. Until the advent of recombinant gene technology these diseases were essentially untreatable; however, hints of potentially successful approaches were provided by pioneering work in bone marrow transplantation, and more recently, delivery of tissue-extracted enzymes such as Ceredase® [1].

## 2.1 Pathophysiology of LSDs

LSDs comprise a collection of approximately 50 rare conditions characterized by abnormal and varying accumulation of metabolic substrates in the lysosomes of many cell types and tissues. Substrate accumulation is the result of impairment (total absence, partial absence, and structural distortion) of specific lysosomal hydrolases essential for degradation of macromolecular cellular components. Individual LSDs are the result of fortuitous pairings of random parental mutations in the genes for a particular lysosomal enzyme. Clinical manifestations of the individual LSDs vary widely as a result of the cell types most affected by abnormal accumulation of substrate. Typically many cell types throughout the body accumulate the misdigested lysosomal material, but the degree of accumulation is not always the determining factor in whether a particular organ or system is seriously impaired. For example, substantial accumulation of substrate in neurons is common to many LSDs, but this accumulation may not be manifest clinically in some (Pompe disease, MPS IV and VI), whereas others present with severe and rapidly progressive neurologic impairment (MPS I, II, III, and VII) [2]. Similarly in other organs the degree of dysfunction is not correlated with degree of accumulation in particular cells, but rather the impairment that the substrate imparts on the particular cell type. For example, in the kidneys of Fabry patients accumulation of globotriaocylceramide in podocytes is uniformly intense, but with only moderate consequences for glomerular filtration. The severe renal (and CNS) pathology in Fabry disease rather arises from relatively modest substrate accumulation in the endothelium, resulting in significant pro-inflammatory upregulation, with numerous downstream pathologic consequences [3].

Most LSDs have a short rapid course, beginning in early childhood, progressing to premature death within a few months or years. A few, such as Fabry disease, have more prolonged courses characterized by predominantly subclinical progression of the underlying pathology (other signs of the disease may, however, be present), only progressing to significant clinical decline in the third or fourth decade with death usually occurring in middle age. Several LSDs (e.g., MPS-I, Pompe) are also well recognized for having distinct and different subsets of clinical phenotypes, varying by clinical onset and progression. These range from severe, rapid-onset infantile forms with early death to more slowly progressive or late-onset adult forms of the disease. This variability of phenotype in such small patient populations complicates the task of defining clinical trial cohorts. More than half of the LSDs have severe neurologic phenotypes, some with devastatingly rapid progression. The difficulty of effectively delivering therapies across the blood–brain barrier remains a significant obstacle for these conditions.

Etiologies for most of the LSDs have been well documented, some for almost a century. The unique variations of phenotype, age at onset, clinical progression, and severity of the individual LSDs, depend on the degree of enzymatic deprivation, determined principally by whether the mutational mix

results in complete absence of the lysosomal enzyme (null) or variable partial reduction or dysfunction of the protein (misfolding, missing sections of the enzyme, etc.). As a result therapeutic approaches to the LSDs, while possessing some common themes, must allow for individually tailored strategies.

## 2.2 Therapies for LSDs

The first successful efforts to treat LSDs came in the form of allogeneic bone marrow transplants aimed at providing at least residual levels of circulating enzyme from the grafted cells of normal, closely matched donors (often a sibling or family member)[2]. While bone marrow transplants remain an option for some LSDs still lacking viable therapies, they rarely achieve the necessary blood levels of enzyme and exposure for full- or long-term reversal of the disease phenotype.

The first major breakthrough in providing significant and sustained therapy for any of the LSDs came in the form of Ceredase® (alglucerase), an ERT for Gaucher disease developed by Genzyme and approved for commercialization in 1991. Ceredase® was the native human enzyme extracted from large numbers of human placentas and purified on an unprecedented manufacturing scale. While this product proved to be highly effective in treating the predominant macrophage-based pathology of Gaucher disease, the long-term logistics and commercial challenge of maintaining this type of product was daunting. However, Ceredase® established the therapeutic credibility of ERT, such that alternative manufacture of lysosmal enzymes by bioreactor technology became an immediate and potentially viable alternative [1]. As a result Ceredase® was quickly surplanted by a follow-on bioreactor product Cerezyme® in 1997, although a few patients are still maintained on Ceredase® treatment as a result of immunological intolerance to the recombinant enzyme.

In the mid-1980s rapid expansion of recombinant protein production techniques ushered in an explosion of protein and antibody-based therapies for many unmet diseases [4] However, commercial development of ERTs for rare diseases was not possible without one other simultaneous development—passing of the US Orphan Drug Act in 1983, later emulated by similar legislation in Japan (1993) and Europe (1999). By providing exclusive marketing rights for a period of several years to the company first able to obtain approval for a viable therapy (defined as diseases with less than 200,000 affected individuals in the US Orphan Drug Act), there was an incentive to engage in the long and risky business of developing therapies for rare diseases. Absence of direct competition also permitted pricing practices and revenue generation that could not have arisen in the open market. As a result there are now eight approved ERTs (Ceredase®, Cerezyme®, Fabrazyme®, Replagal®, Aldurazyme®, Naglazyme®, Myozyme®, and Elaprase®), with several others in various stages of development. This has been an important advance, the basis for which, bioreactor-produced enzymes, will be reviewed below.

### 2.3 LSDs and Models

What is evident immediately on attempting to develop a therapy for any particular LSD is the scant (and for many LSDs, virtually absent) information on disease pathophysiology. Beyond recognition of the enzyme deficiency responsible for the disease and the biochemistry of the accumulating substrate, much of the literature is in the form of case reports on individuals or small patient sets, often with minimal characterization of disease pathology. The paradox is that a more complete body of literature only develops after an effective treatment becomes available (e.g., Gaucher and Fabry diseases). Considerable initial research is therefore needed by any company seeking a successful therapeutic for a specific LSD.

Experience gained from developing the different ERTs (and in some cases their predecessors) has highlighted the individual complexities of providing drug exposure to the relevant tissues for a particular LSD. To characterize enzyme kinetics, drug exposure, inter- and intracellular transport, and substrate elimination, numerous molecular biology tools and in vivo strategies have been developed, specific to LSD biology. An obstacle has largely been the mismatch between animal models (mostly knockout or transgenic mice, or naturally occurring mutations in various species) and the human counterpart of the LSD in question. For Gaucher disease, at the time of developing Ceredase®, and later Cerezyme®, no suitable animal model was available since the murine knockout was embryonically lethal. For Fabry disease, there was a viable knockout mouse, however, alternative metabolic pathways in key cell types (endothelium, podocytes, cardiomyocytes) meant that only proof-of-principle pharmacodynamic data could be obtained from affected tissues (liver, vascular smooth muscle), thus severely limiting the value of the model. For Pompe disease, the opposite was true: there were two very suitable models, a naturally occurring mutation in quail (with significant functional impairment) and a knockout mouse with many of the pathologic features of the disease, thus allowing for significant preclinical investigation and development of a relatively complete preclinical dossier. Other development programs such as Aldurazyme® for MPS I relied on data from a naturally occurring dog model of the disease with the inherent difficulty of small data sets and inability of developing an all encompassing preclinical program. The generation of a nonlethal mouse model for Niemann Pick disease has allowed for detailed and relevant investigation into the safety of ERT in a model that has substrate accumulation.

### 2.4 Delivering ERTs

In healthy individuals lysosomal enzymes are transported from the endoplasmic reticulum to the lysosome via mannose-6-phosphate (M6P) mediated vesicles or other receptor-mediated transporters [5]. The phosphate-labeled enzyme in the case of M6P-mediated transport is released to the lysosome by

clipping the phosphate and recyling the M6P vesicle to the endoplasmic reticulum. Some vesicles are also cycled to the cell membrane as part of a mechanism for mopping up enzyme that may have escaped into the interstitial space [6]. ERT is dependent on this latter mechanism of plasma membrane receptor cycling for capture of therapeutically delivered enzyme. The efficiency of enzyme uptake in the target tissue of specific LSDs depends on a complex interplay of receptor mediated mechanisms. This interplay begins with the delivery of the enzyme to the intravascular space by infusion. The unique relative combinations of exposed N-glycosylated sugars (mannose, mannose-6-phosphate, sialic acid, galactose) on the glycosylation sites of the enzyme is crucial to tissue targeting of the enzyme. It is the combination of relative organ blood flow, absolute receptor density, and relative receptor mix that ultimately determines exposure, uptake, and potency for any particular cell type/tissue. Organs such as the liver with multiple, densely populated receptor systems and high blood flow take up the bulk of delivered enzyme, an advantage in a disease such as Gaucher in which the liver is a key player. The opposite can be said for Pompe disease in which the skeletal muscle is poorly endowed with the relevant M6P receptor, thereby suffering from relatively poor exposure.

Since base glycosylation pattern and secondary decoration of the glycosylation sites can be manipulated by bioreactor conditioning or posttranslational chemistry respectively, manufacturers can significantly influence the therapeutic behavior of their products. Such flexibility, however, also produces the potential for unpredictable product variability and the need for very tight controls on bioreactor conditions and verification of product specifications. Delivered enzymes are therefore a controlled mix of enzyme species decorated with varying combinations of mannose, mannose-6-phosphate, sialic acid, and galactose moieties (among others). For example, based on the seven glycosylation sites of Myozyme®, and the combination and permutations of the oligosaccharide species potentially able to occupy these sites, several hundred thousand different versions of the enzyme are possible products of the individual cells in a bioreactor.

## 2.5 Defeating the Pathophysiology of LSDs

The final link in the therapeutic chain is delivery of the enzyme to the lysosome, enzyme activation, and processing of the substrate. Unlike small-molecule therapeutics, the delivery and eventual fate of the enzyme is a one-way street—once delivered to the lysosme, little if any recycling (or re-diffusion) occurs. This is a particular pharmacokinetic model that perplexes some regulators as well as some developers more used to small-molecule therapeutics. What occurs in the lysosome in many ERTs remains only partially understood. In the case of Myozyme®, for example, there is considerable processing (clipping and re-arrangement of the enzyme) into two successive, and active, multi-subunit complexes of the enzyme over several days [7]. Further the manner in which the substrate is processed, or in some cases fails

to be processed (e.g., GL-3 in cardiomyocytes of Fabry patients, glycogen in the skeletal muscle of the most severely affected Pompe infants), is only partially understood. Considerable ongoing effort is focused on improving the yield of substrate removal, particularly when these failures negatively impact the best outcomes of ERT [8].

## 3 CURRENT TECHNOLOGIES

Since the early 1980s production of most therapeutic human proteins was performed in bioreactors containing well-characterized bacteria or yeasts [4]. More recently immortalized mammalian cell lines such as CHO (Chinese hamster ovary) and some human cell lines have increasingly been employed, particularly for more complex proteins [4]. Mammalian cell lines, while more demanding to produce and maintain, are necessary when correct glycosylation and appropriate folding behavior are required of the protein. Such is the case for most lysosomal enzymes, although new methods of engineering yeasts to produce humanized glycosylation patterns by selective knockouts and knock-ins of specific genes has recently provided a potentially more economic bioreactor alternative [9]. Cells in the bioreactor have been genetically modified to produce and secrete the desired human protein. Harvesting of the secreted protein is conducted either at the end of each culture run by processing the entire contents of the bioreactor before starting again (batch mode) or in the case of bioreactors equipped with cell microcarriers, in perfusion mode, by siphoning off culture medium in a continuous process. While a detailed description of bioreactor management is not the subject of this chapter, several features of the manufacturing process have a direct impact on preclinical testing. One is that significant control can be exerted over the composition of the protein, and more particularly the glycosylation pattern, by manipulating growth and culture conditions at several points in the manufacturing process. The second is that each scale-up to bigger reactors to meet patient demand changes the culture conditions and proportionality, resulting in a potentially different mix of enzyme species extracted from the bioreactor. These changes in composition, while relatively minor in most cases, can significantly alter the pharmacokinetics and potency of the product and are closely monitored by regulatory agencies. To place these challenges in context, we include a brief description of the bioreactor steps that most commonly impact product characteristics.

In general, the first step in producing bioreactor enzymes involves selecting the most appropriate cDNA from which to generate the protein. The gene for the human protein (enzyme in the case of ERT) is then inserted into a suitable vector (usually a plasmid), and transfected into the cell line intended for use in the bioreactor. Numerous transfected cell lines are tested for replication efficiency (clonal selection), based on activity/expression levels in high-throughput, plate format screens. During clonal selection, manipulation of the

culture system with methotrexate or other gene expression modifiers may be employed to force expression in favor of a particular glycosylation pattern. A small number of candidates may be chosen from this process for benchtop flask culturing, from which the first more complete analysis of protein product can be performed by assays such as SDS-PAGE. At this stage the final candidate clone is selected to move forward into early production runs.

During the next stage a series of parallel activities accompanies progressive expansion of culturing into benchtop bioreactors (8 L), thence up to roughly 30 L systems. These include progressive purification steps and the development of more sophisticated assays to characterize the protein, including antibody-based assays (ELISAs, Westerns). The purpose at this stage is to produce sufficient purified enzyme to begin in vivo testing and to more accurately characterize the enzyme, particularly the glycosylation pattern and other characteristics that could determine in vivo behavior. Further modification of the enzyme product can be introduced in this midstage. One method is to introduce back into the cell culture media chemicals such as kifunensin that induce remodeling of the oligosaccharide side chains by promoting sialidases to clip certain sugars. The other way to modify the decorating oligosaccharides is to treat the purified product posttranslationally with enzymes specific for their activity on sialic acid, galactose, mannose, or glucosamine residues. This way the oligosaccharide complexes can be clipped to expose a different and more beneficial selection of signaling sugars. Each of these steps requires an array of assays to characterize each step and the resulting product. One further potential modification is open to the manufacturer. This involves taking the final glycoslyated product through a series of biochemical steps to restructure and adorn the oligosaccharides with the desired species (e.g., bisphosphorylation of a particular group of mannose terminations).

Bioreactor scale-up remains a considerable challenge for producers of ERTs. One reason is that pivotal trials for LSDs are conducted on small numbers of patients, using quantities of enzyme that can be produced in smaller bioreactors, before the need to scale up quantities of product (and bioreactor size) to meet commercial demand. This is a conundrum that places considerable pressure on sponsors. The pitfall is that different scales may change the product profile and pharmacokinetics characteristics sufficiently that regulators require duplication of preclinical safety studies and even clinical trials after the pivotal phase 3 trials have been completed. This places an undue burden on sponsors seeking registration for these rare conditions.

## 4 REGULATORY GUIDANCE

Little regulatory guidance was in place during development of the early recombinant human enzymes. Ceredase® and Cerezyme® were developed in the absence of the ICH S6 document, and therefore safety packages for these enzymes were a combination of traditional safety studies as well as specific

studies to address the unique properties of a recombinant human protein. Studies for the recombinant human enzymes were designed to establish a safety margin above the clinical dosing regimen and to identify any possible effects of the administered protein. Since these enzymes are naturally occurring human proteins, the likelihood of side effects was low. However, understanding the pharmacokinetics, biodistribution, and exposure was key to establishing the safety profile of the enzymes.

With acceptance and implementation of the ICH S6 guidance document, more consistent development of the recombinant human proteins was possible. Similar studies were performed to assess the short- and long-term safety of administered proteins. Challenges induced by immunogenicity and the ability to dose normal animals for sufficient duration became the biggest issue. The utility of long-term dosing in rodents became questionable due to the anticipated generation of antibodies to the recombinant human protein. A major consideration was the relevance of the animal model, and which study would enable prediction of safety outcomes in patients. As more and more recombinant enzymes reached the development phase, the best studies and relevant endpoints were increasingly challenged. It was necessary to conduct studies for appropriate lengths of time, but also to deliver the protein in a manner similar to the clinical program. For ERTs this meant hour-long infusions in an animal considered likely to display the appropriate safety signals without exhibiting the hypersensitivity associated with administration of a human protein. Despite the lack of utility and predictability of many of these safety studies, they were still considered necessary for the safety assessment of products such as Fabrazyme® and Aldurazyme®.

Several recombinant human proteins have been approved over the last 10 years for use as long-term ERT in patients. Because of the specificity of the individual enzymes for particular lysososmal storage products, demonstration of a clear pharmacodynamic effect can only be performed in genetically modified animal models lacking the enzyme of interest (or, in rare naturally occurring disease models). Many genetically induced animal models are embryonic lethal or have alternative compensatory mechanisms such that the pathology of the disease is not consistent with the human disorder. In addition, as a result of significant antibody responses in some knockout mouse models, administration of recombinant human proteins may preclude meaningful interpretation. Long-term studies of the human protein are often not achievable in these animal models. Unfortunately, toxicology studies in normal animals are limited by the presence of native enzyme and the absence of pathologic substrate accumulation, both features that we now recognize can significantly impact safety of the product. While development programs for some of the enzyme replacement therapies were conducted prior to the ICH S6 document, many were nonetheless consistent with the guidance. The challenges and relevance of some of the development programs are described below.

It is important to remember that as a class of proteins, ERTs are expected to be relatively safe because of their specific targeting via mannose or mannose

6 phosphate receptors, their endogenous protein nature, and their relatively little activity at a physiologic pH. Since these are compelling arguments, safety evaluations for some early programs were very brief and limited to only a single species. Further, until recently, all of the safety studies conducted to support clinical used normal animals. In light of recent developments (see below) the relevance of this approach, and the relevant species, need to be reconsidered for development work.

### 4.1 Cerezyme®

Imiglucerase (Cerezyme®) was approved in 1997 for use as long-term enzyme replacement therapy in patients with confirmed diagnoses of non-neuropathic (type 1) or chronic neuropathic (type 3) Gaucher disease. Gaucher disease, also called gluocosylceramide lipidosis or β-glucocerebrosidase (GCR) deficiency, is the most common of the sphingolipidosis or lipid storage diseases, and is inherited in an autosomal recessive manner. Imiglucerase catalyses the breakdown of glucocerebrosidase, thereby preventing the buildup glucocerebroside. Imiglucerase is the recombinant form of alglucerase. Alglucerase (Ceredase®), derived from human placental tissue, was the first mannose-terminated-B-glucocerebrosidase.

The clinical features of Gaucher disease include anemia and thromoboctyopenia due to splenic sequestration, and bone marrow displacement by accumulating Gaucher cells. Hepatomegaly, spenomegaly, and osteonecrosis are common symptoms in type 1 and 3 Gaucher disease. The neuropathic forms also manifest neurological abnormalities such as seizures, dementia, spasticity, ataxia, and loss of intellectual function. Imiglucerase is given by infusion over 2 to 3 hours every two weeks, starting at a specified dose that can be lowered as therapy (and clinical response to treatment) progresses.

The efficacy of imiglucerase was not studied in an in vivo preclinical model because there is was no satisfactory animal model of Gaucher disease. Receptor binding was characterized in vitro using rat alveolar macrophages. Data from preclinical studies showed similar pharmacodynamic actions of imiglucerase and alglucerase. Three-month repeated intravenous dose toxicity studies were performed in rats and monkeys to support clinical trials and approval. Anti-imiglucerase antibodies were detected in all dose groups in rats. However, the onset of the antibody response and the number of animals responding was dose dependent. Interestingly, the anti-imiglucerase antibody response was more prominent in monkeys than in rats. The most common clinical side effect is hypersensitivity. Because humans can develop antibodies that may affect efficacy, it is recommended that patients be monitored for any allergic reaction during their first year of treatment.

There are some interesting and important features of the development strategy adopted for this product which differ from more conventional approaches. It seems reasonable to assume that a nature-identical human protein produced by rDNA technology will be inherently safe if administered

to humans. Therefore it seems reasonable to conclude that risk assessment of proteins that are not nature-identical should focus on differences from the natural product, since such differences might affect important processes such as primary biological action, tissue distribution, storage, metabolism, clearance, and immune response.

Imiglucerase differs from the native glycoprotein in two important respects. First, there is a single amino acid substitution in the primary sequence, which arose in the cloning process. Second, the glycosyl moiety of the protein is deliberately modified to increase the affinity of rGCR for the mannose receptors of macrophages. The philosophy used in the risk assessment program was, first, to compare the rDNA product with the natural protein and, second, to investigate directly the preclinical safety of imiglucerase. It is important to recognize that the pivotal 13-week toxicity study in the primate was performed with material from a 2000 L batch, which is effectively the clinical marketed finished product. This lot was a prevalidation lot manufactured by the same process, as the 2000 L validated process, and although full characterization was not conducted, it did meet all release specifications and therefore should be indicative of the clinical/market material at the 2000 L scale. The data from this animal study were therefore central to interpretation of potential human risk. Because no satisfactory animal model of Gaucher disease existed, it was not possible to predict the efficacy of Cerezyme® from animal studies.

The development program was designed to elucidate the fundamental biological properties and action of imiglucerase. However, in comparison to the preclinical toxicology plan, which is generally presented for low molecular weight pharmaceuticals, there were several obvious differences.

Broadly, there was no study of clastogenicity, only one species was used in the evaluation of acute toxicity, and no studies of potential effects on reproductive function or oncogenic potential were performed. In contrast, detailed studies of the biologically important receptor binding and macrophage uptake were presented together with detailed assessments of the kinetic processes defining receptor binding and plasma clearance.

Macrophage mannose receptor binding is a particularly important issue in the assessment of rGCR as an enzyme replacement therapy because macrophages are the predominant cell type that accumulates lipid in Gaucher patients. In addition demonstrating that the recombinant protein behaved virtually identically to the naturally occurring protein would provide reassurance in the risk assessment process.

A series of pharmacokinetic and ADME studies was performed with imiglucerase in order to assess clearance and targeting of imiglucerase in vivo. Because there was no suitable animal model for Gaucher disease, these studies were performed in normal animals.

Three toxicology studies were conducted: an acute study in rats and two 13-week studies in rats and monkeys. Most findings in these studies could be attributed to infusion of large amounts of exogenous protein that directly or indirectly (through an immune response) effected changes. As would be

expected, repeated administration of Cerezyme® to rats and monkeys elicited a dose-dependent antibody response in more than 50% of the animals exposed to the test article. Since this antibody response against imiglucerase was demonstrated in the rat and monkey, it was considered that additional longer term toxicology studies were not justified as potential target organ exposure to the test article could not be quantified. In view of minimal, subchronic toxicity findings this was considered an acceptable safety package for marketing approval. An Ames test was also performed.

### 4.2 Fabrazyme®

In 2003 agalsidase beta (Fabrazyme®) was approved for use in patients. Agalsidase beta is a recombinant human α-Galactosidase A enzyme with the same amino acid sequence as the native enzyme. In Fabry disease an X-linked inherited deficiency in α-Galactosidase (αGAL) leads to progressive lysosomal accumulation of globotriasylceramide (GL-3) in most tissues of the body. As a recombinant equivalent of naturally occurring α-Galactosidase, recombinant human α-Galactosidase (r-hαGAL) was also expected to reduce GL-3 accumulation in tissue, and thereby ultimately to alleviate the clinical consequences of GL-3 accumulation. Pharmacodynamic studies were designed to evaluate the potential for r-hαGAL to clear GL-3 substrate from tissues of an αGAL SV129 knockout mouse model. The use of the mouse model is justified as it provides the only biochemical model for Fabry Disease.

A safety pharmacology study in dogs assessed cardiovascular and respiratory toxicity. Neurological toxicity was not assessed as part of the safety pharmacology study since r-hαGal is a large protein and unlikely to cross the blood–brain barrier. In accordance with the International Conference on Harmonization (ICH) S7 guideline, Safety Pharmacology Studies for Human Pharmaceutics, secondary pharmacology studies are not warranted as available information from toxicology studies and clinical studies, including 24 months of safety information from patients in the phase 3 trial who have continued on treatment but have not identified a cause for concern such that secondary pharmacology studies would be required.

Biodistribution studies were performed in Balb/C mice, the αGAL SV129 knockout mouse model and Sprague–Dawley rats to assess the uptake and stability of r-hαGAL in target organs. Additionally pharmacokinetic analyses were performed across three species (rat, dog, and monkey) with either single-dose or repeat-dose administration of r-hαGAL. Accumulation studies were conducted in rats and monkeys.

In the toxicity studies the maximum administered dose was identified as the NOAEL. A traditional lethal dose study was not performed. Instead, the relationship between dose levels and toxicity were evaluated. These studies demonstrated that r-hαGAL had extremely low toxicity. It was unlikely that a clear toxic dose could have been identified for r-hαGAL given that r-hαGAL cannot be sufficiently concentrated to deliver a lethal dose and test animals

will not tolerate the volume required to deliver a less concentrated lethal dose. This approach is consistent with the International Conference on Harmonization (ICH) S6 guideline, Preclinical Safety Evaluation of Biotechnology Derived Pharmaceuticals, which states that for some classes of products with little to no toxicity, it may not be possible to define a specific maximum dose, so consideration should be given to the volume that can safely and humanely be administered to the test animal.

Both single-dose and repeat-dose toxicity studies were conducted in two species, rat and dog, and rat and monkey, respectively. In developing plans for the toxicity studies, the Sprague–Dawley rat was selected as a relevant species to test. A 27-week repeat-dose toxicity study in SD rats was conducted to evaluate the long-term safety profile of r-hαGAL. Given the lifelong use of r-hαGAL in Fabry patients, it was recognized that data from a second species at high doses would be useful. A second long-term safety study was therefore conducted in cynomolgus monkeys. The ICH guideline provides that in principle, the dosing duration for the toxicity study may be six months even in the case of a test substance for which the expected period of clinical use exceeds six months. In addition long-term administration of a compound that is known to be immunogenic in an animal model is unlikely to give relevant safety information. Due to the immunogenic nature of the recombinant protein in the nonhuman primate, dosing for a period of nine months was unlikely to provide any additional information over dosing to six months. Conduct of a long-term infusion study was the most technically achievable in the nonhuman primate.

Additionally antigenic studies were performed in two species (rat and monkey) and an embryofetal development study was conducted in rats. Sperm morphology evaluation, conducted as part of the 25-week repeat-dose toxicology study in monkeys was inconclusive due to the immature age of the monkeys. A segment 1 study in rats was also conducted. Perinatal development studies have not been conducted, and the recommendation that r-hαGAL should not be used during pregnancy, unless clearly necessary, is adequately reflected in the proposed labeling. Given that r-hαGAL is a nature-identical human protein, and based on the therapeutic indication and patient population, studies to assess effects of r-hαGAL on mutagenic and carcinogenic potential were not considered necessary as part of the development program.

There were no significant treatment-related findings in any of these studies. The only finding consistent with administration of test article was a hypersensitivity response seen in the rat study. As predicted, antibodies to agalsidase beta were detected in both chronic studies, but the antibodies did not have an impact on safety assessments or pharmacokinetics of the therapeutic.

In the clinical setting, infusion reactions occurred in many patients treated with agalsidase beta, and some of these reactions were severe. Infusion reactions occurred in some patients after receiving antipyretics, antihistamines, and oral steroids.

### 4.3 Aldurazyme®

Laronidase (α-l-iduronidase) is intended for the treatment of mucopolysaccharidosis I (MPSI). The MPS disorders are caused by deficiencies of specific lysosomal enzymes required for the catabolism of glycosaminoglycans (GAGs). Accumulation of GAG substrates occurs in a variety of tissues and is dependent on the location of the affected substrates and their rate of turnover. Deficiency of α-l-iduronidase results in accumulation of dermatan and heparan sulfate in many tissues, particularly connective tissue, leading to chronic progressive loss of joint function, disordered skeletal growth, organomegaly, and neurologic impairment in more affected cases.

Canine MPS I was discovered in a Plott hound that presented with corneal clouding [10]. Studies by Shull and Neufeld showed that the dogs were deficient in α-l-iduronidase [11]. Being null, MPS I dogs are genetically similar to the most severe form of MPS I in humans, but clinically they more closely resemble moderately affected patients. These animals provide a valuable biochemical/clinical model for MPS I disease. Since they have no confounding residual enzyme activity, they accumulate GAGs in relevant tissues, and their clinical phenotype closely resembles the human disease.

MPS I dogs were used to evaluate the biodistribution of rh-iduronidase, pharmacodynamic reduction of tissue GAGs and in vivo efficacy. Pharmacokinetics and the effects of various doses and regimens of rh-iduronidase were also evaluated in this model. Heterozygote dogs from the MPS I colony, referred to as normal carriers, were used to derive normal tissue levels of α-l-iduronidase activity and GAG storage.

The pharmacokinetics of rh-iduronidase were characterized in two studies, along with five pharmacodynamic studies in which pharmacokinetics, biodistribution, or both, were determined. The pharmacodynamic studies were conducted in the MPS I dog evaluating reduction of tissue and urine GAG levels. These studies ranged from short-term (5–12 days) to long-term studies lasting up to 74 weeks. Multiple-dose levels were tested to determine the effect of dose on levels of α-l-iduronidase activity and GAG in tissues and urine. The effects of dosing regimen, including once and three times weekly, as well as continuous infusion, were also tested. In the long-term studies, assessments were made of the effects of treatment on the clinical symptoms of the disease. Two additional toxicokinetic studies were conducted as part of an acute single-dose intravenous (IV) infusion toxicity study in normal dogs and as part of a 26-week IV infusion toxicity study in cynomolgus monkeys.

There were no significant treatment-related findings in any of these studies. The only test article related finding was a hypersensitivity response seen in a few of the animals treated. Antibodies to laronidase were detected in the chronic studies, but the antibodies did not have an impact on the safety assessment or pharmacokinetics of the therapeutic. In addition there were no findings at six months that had not been noted at the three-month time points.

In the clinical studies the most common adverse events observed with laronidase treatment were upper respiratory tract infection, rash, and injection site reaction. The most common adverse reactions requiring intervention were infusion-related reactions (IRRs), particularly flushing. Those requiring intervention were offset by slowing the infusion rate, temporarily stopping the infusion, and/or administering additional antipyretics and/or antihistamines.

### 4.4 Myozyme®

In 2006 recombinant alglucosidase alfa (Myozyme®) received approval in the United States and Europe for the treatment of Pompe disease in all patients. While the development program of alglucosidase alfa was in consideration of the ICH S6 document, and with further consideration to conduct studies in the most relevant species with the greatest predictability, additional pressure was placed on this program for the establishment of safety in animal studies.

The GAA knockout mice used for the nonclinical program were generated by Raben et al. [12,13]. The model was produced by inserting a neomycin-resistance (neo) gene, into exon 6 of the murine GAA gene, resulting in disruption of the coding region. Mice homozygous for this disruption ($6^{neo}/6^{neo}$) lack enzyme activity, accumulate lysosomal glycogen, and "recapitulate(s) critical features of both the infantile and the adult forms of the disease at a pace suitable for the evaluation of enzyme or gene replacement" [12]. By 3 weeks of age, homozygous $6^{neo}/6^{neo}$ mice begin to accumulate glycogen in cardiac and skeletal muscle lysosomes due to the lack of GAA enzyme activity. However, they grow normally, reach adulthood, and remain fertile. By 8 to 9 months of age, obvious muscle wasting and a weak and waddling gait develops, and, by 18 to 19 months of age, they have severe progressive muscle-wasting resembling advanced late-onset Pompe disease. At this age, they have pronounced glycogen accumulation in multiple organs (including skeletal muscle, diaphragm, heart, and brain), as well as cardiomyopathy, hypotonia, severe motor disability, and profound muscle weakness and wasting [12].

Glycogen accumulation and extent of muscle pathology in GAA knockout mice is not as striking as that noted in infantile-onset Pompe patients. Based on life span $6^{neo}/6^{neo}$ GAA mice more closely resemble late-onset Pompe disease in humans. Despite these differences GAA knockout mice provided a valuable tool for preclinical studies. Since they lack GAA enzyme activity, they can be used to evaluate the pharmacokinetics and biodistribution of GAA activity following administration of clinically relevant doses of rhGAA. Moreover, since the heart and skeletal muscle of GAA knockout mice contain accumulated glycogen, they can also be used to evaluate the pharmacodynamic effects of rhGAA dose and dosing regimens on glycogen depletion. Since most pharmacodynamic studies were conducted on asymptomatic mice less than 6 months of age, we were unable to determine if rhGAA could reverse or stabilize the clinical phenotype.

While useful for evaluating substrate reduction and tissue distribution, the knockout mouse also presented specific challenges to the preclinical development program. After repeated administration of rhGAA, a predictable hypersensitivity response to the recombinant human protein was observed. Frequently this reaction resulted in morbidity and mortality associated with test article administration. This hypersensitivity reaction responded to diphenhydramine prior to, and as necessary, during dosing; however, such intervention complicated interpretation of toxicity studies.

In vivo, single- and repeat-dose pharmacodynamic studies were conducted in GAA knockout mice to evaluate the efficacy of rhGAA. Most of these studies were designed to assess depletion of glycogen from target tissues, while others assessed the time course of glycogen depletion and re-accumulation following administration of rhGAA. Tissue glycogen content was measured by biochemical assay and histomorphometric analysis. In addition a safety pharmacology study was performed in beagle dogs.

A series of pharmacokinetic studies were performed to assess absorption of rhGAA across single- and repeat-dose administration, species (mouse, rat, dog, and monkey), dose, formulation, and process scale. Process scale-up to larger reactors to accommodate the high dose and rapid expansion of patient numbers has placed an especially demanding pharmacokinetic burden on this program, even postapproval. Biodistribution studies were performed in GAA knockout mice to assess the uptake and tissue residence time of rhGAA in target organs.

Toxicity studies were designed to establish a maximum tolerated dose and safety profile in mice, rats, dogs, and cynomolgus monkeys. Two single-dose acute toxicity studies in rats and dogs, two repeat-dose subchronic toxicity studies in rats, one repeat-dose subchronic toxicity study in mice, and two repeat-dose chronic toxicity studies in cynomolgus monkeys were performed to evaluate the safety of rhGAA administered intravenously (IV). Furthermore three reproductive toxicity studies were conducted in mice to assess the effect of rhGAA on fertility, early embryonic development and embryofetal development.

### 4.5 Acid Sphingomyelinase

Niemann–Pick disease is caused by partial or complete absence of the lysosomal enzyme acid sphingomyelinase, resulting in lysosmal accumulation of sphingomyelin in numerous tissues, but particularly in cells of the macrophage lineage and in neurons. Clinically patients suffer from progressive hepatosplenomegaly and associated pathology, and many eventually succumbing to pulmonary complications and pneumonia in late childhood to middle age (Niemann–Pick type B). Recent characterization of Niemann–Pick type B populations suggests significant pathophysiology related to high cholesterol profiles, resulting in many early deaths from vascular disease. In the most

severe classification of the disease (Niemann–Pick type A) rapidly progressive neurologic disease claims individuals in early childhood.

An appropriate knockout mouse model for Niemann–Pick disease is available. This ASMKO mouse was produced by inserting a neomycin-resistant gene (neo) into exon 2 of the *ASM* gene resulting in disruption of the coding region [14]. Mice homozygous for this disruption lack acid sphingomyelinase activity and develop a clinical condition resembling Niemann–Pick A disease. Born normal, these mice develop trembling and ataxia by 8 weeks of age, progressing through lethargy to severe ataxia by 4 months and death between 6 and 8 months of age. Progressive accumulation of sphingomyelin occurs in macrophage lineage cells, the lungs, and the CNS in a manner closely resembling the human disease. This model has allowed for significant preclinical development efforts and disease-related characterization.

As discussed previously, relevant animal models and predictability of safety have become significant issues. Recent experience with developing ERT for Niemann–Pick disease has highlighted the potential risk of relying on standard toxicity models and not identifying the most relevant test species.

To establish the safety profile for recombinant human acid sphingomyelinase (rhASM), several animal models were evaluated including normal rats and dogs. Findings from three single- and repeat-dose toxicity studies resulted in a no-observable-effect level (NOEL) of the highest doses tested. However, in parallel efforts to maximize efficacy of rhASM in different organ systems, unexpected deaths occurred in ASMKO mice at several dose levels, all well below the previously determined NOEL. Signs of toxicity included lethargy, hypothermia, and death, symptomatic of clinical shock. Subsequent characterization of this response established that the toxicity is directly related to substrate load, but the load can be managed pharmacodynamically by initially dosing at low levels to deplete the most labile substrate. Doses can then be increased to the therapeutic levels necessary to maintain control of substrate in all tissues. Appropriate adjustments to the clinical dosing regimen have been incorporated in our phase 1 protocol to manage this risk. The fact that toxicity occurred only in knockout animals indicates that for LSDs (and perhaps other diseases with substrate accumulation) future toxicological studies may need to be conducted in animal models of the disease to ensure that the most sensitive species has been included.

## 5 CONCLUSION

Recombinant enzymes manufactured from mammalian cell-based bioreactors have established a solid treatment paradigm for several lysosomal storage diseases, a class of rare diseases until recently considered untreatable. While small clinical populations, variable animal models, and the need to upscale manufacturing during and after approval have provided formidable regulatory challenges, the fact that eight separate enzyme therapies have emerged in little

over a decade is remarkable. Credit must go to the companies that have committed to this risky field, and also to the foresight of legislatures that established critical orphan drug laws and to the regulatory bodies in many countries open to creative approaches by sponsors. The examples presented of successful preclinical programs demonstrate the importance of flexibility based on unique aspects of the individual diseases and the availability of relevant animal models to address specific safety concerns, consistent with the ICHS6 science-based approach.

New challenges resulting from this decade of cumulated experience are, however, placing increasing pressure on maintaining this momentum. Among the most pressing are (1) the troubling antibody responses to recombinant human enzymes in patients—nearly impossible to investigate effectively in animal models; (2) the recent discovery that the substrate load (in some LSDs) may impart unique toxicity risks that cannot be explored in normal animals; and (3) increasing evidence that despite predictable dose–responses in relevant models, some patients (and some tissues) are not pharmacodynamically (and clinically) responsive even at elevated doses. These, among others, continue to impose increasingly complex preclinical and CMC regulatory obligations at a time when increasingly rare LSDs invariably dictate more restrained investment. When combined with the absence of a clear pathway to accessing the CNS, the prospect for the next decade of ERT development is less certain. Based on recent small-molecule approaches to substrate inhibition and chaperone-assisted refolding for certain mutations, the next wave of development for LSDs, including those with neurologic involvement, may come through small-molecule therapies (possibly in combination with proteins). Gene therapy with long-lived vectors, and in combination with tissue-specific delivery, remains a viable but complex option. On the positive side are the increasingly creative animal models and laboratory technologies available to support these more complex paradigms—there is no doubt that they will be needed!

## REFERENCES

1. Pastores GM, Sibille AR, Grabowski GA. Enzyme therapy in Gaucher disease type 1: dosage efficacy and adverse effects in 33 patients treated for 6 to 24 months. *Blood* 1993;82(2):408–16.
2. Watts W, Gibbs D, eds. *Lysosomal Storage Diseases: Biochemcial and Clinical Aspects*. London: Taylor and Francis, 1986.
3. Thurberg BL, Rennke H, Colvin RB, Dikman S, Gordon RE, Collins AB, Desnick RJ, O'Callaghan M. Globotriaosylceramide accumulation in the Fabry kidney is cleared from multiple cell types after enzyme replacement therapy. *Kidney Int* 2002;62(6):1933–46.
4. Alcamo E. *DNA Technology: The Awesome Skill*. San Diego: Academic Press, 1996.
5. Kornfeld S. Structure and function of the mannose 6-phosphate/insulinlike growth factor II receptors. *Annu Rev Biochem* 1992;61:307–30.

6. Griffiths G, Hoflack B, Simons K, Mellman I, Kornfeld S. The mannose 6 phosphate receptor and the biogenesis of lysosomes. *Cell* 1988;52(3):329–41.
7. Moreland RJ, Jin X, Zhang XK, Decker RW, Albee KL, Lee KL, Cauthron RD, Brewer K, Edmunds T, Canfield WM. Lysosomal acid alpha-glucosidase consists of four different peptides processed from a single chain precursor. *J Biol Chem* 2005;280(8):6780–91.
8. Fukuda T, Ewan L, Bauer M, Mattaliano RJ, Zaal K, Ralston E, Plotz PH, Raben N. Dysfunction of endocytic and autophagic pathways in a lysosomal storage disease. *Ann Neurol* 2006;59(4):700–8.
9. Hamilton SR, Davidson RC, Sethuraman N, Nett JH, Jiang Y, Rios S, Bobrowicz P, Stadheim TA, Li H, Choi BK, Hopkins D, Wischnewski H, Roser J, Mitchell T, Strawbridge RR, Hoopes J, Wildt S, Gerngross TU. Humanization of yeast to produce complex terminally sialylated glycoproteins. *Science* 2006;313(5792): 1441–3.
10. Shull RM, Munger RJ, Spellacy E, Hall CW, Constantopoulos G, Neufeld EF. Canine alpha-L-iduronidase deficiency. A model of mucopolysaccharidosis I. *Am J Pathol* 1982;109(2):244–8.
11. Spellacy E, Shull RM, Constantopoulos G, Neufeld EF. A canine model of human alpha-l-iduronidase deficiency. *Proc Natl Acad Sci USA* 1983;80(19):6091–5.
12. Raben N, Nagaraju K, Lee E, Kessler P, Byrne B, Lee L, LaMarca M, King C, Ward J, Sauer B, Plotz P. Targeted disruption of the acid alpha-glucosidase gene in mice causes an illness with critical features of both infantile and adult human glycogen storage disease type II. *J Biol Chem* 1998;273(30):19086–92.
13. Raben N, Nagaraju K, Lee E, Plotz P. Modulation of disease severity in mice with targeted disruption of the acid alpha-glucosidase gene. *Neuromuscul Disord* 2000; 10(4–5):283–91.
14. Horinouchi K, Erlich S, Perl DP, Ferlinz K, Bisgaier CL, Sandhoff K, Desnick RJ, Stewart CL, Schuchman EH. Acid sphingomyelinase deficient mice: a model of types A and B Niemann-Pick disease. *Nat Genet* 1995;10(3):288–93.

# Toxicology of Oligonucleotide Therapeutics and Understanding the Relevance of the Toxicities

ARTHUR A. LEVIN, PhD, DABT, and SCOTT P. HENRY, PhD, DABT

**Contents**

1 Introduction
2 Chemistry
  2.1 Aptamers
  2.2 Immunostimulatory Oligonucleotides
  2.3 Single-Stranded Oligonucleotides Used for Antisense Activity
  2.4 Double-Stranded RNAs–siRNA
3 Pharmacokinetics
  3.1 Assessing the Pharmacokinetics of Oligonucleotide Drugs
  3.2 A Brief Summary of the Pharmacokinetics of Oligonucleotide Drugs
4 Toxicity of Oligonucleotide Therapeutics
  4.1 Overview
  4.2 Genotoxicity
  4.3 Safety Pharmacology
  4.4 Acute Toxicity
  4.5 Subchronic and Chronic Toxicity
  4.6 Reproductive and Endocrine
  4.7 Immunotoxicity
  4.8 Carcinogenicity
5 Summary
References

## 1 INTRODUCTION

This chapter will describe the existing practices for assessing the toxicity of oligonucleotide therapeutics and the rationale for these practices. The only

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

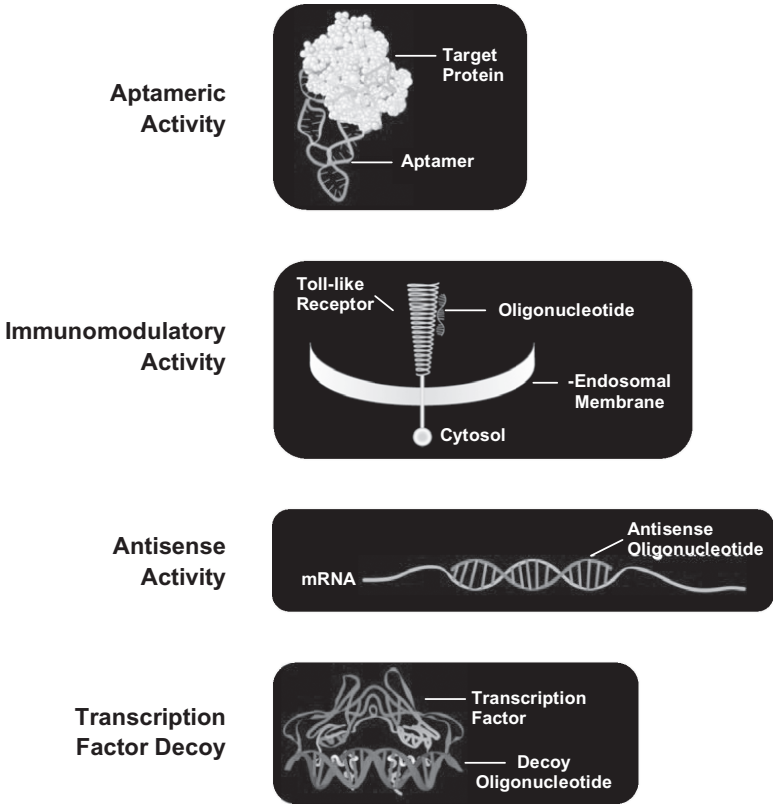


**Figure 1** Cartoon depicting the mechanism of action of oligonucleotide molecules for aptameric interactions with proteins, interactions with specific receptors of the innate immunity, target mRNAs through hybridization, and transcriptional activators as transcription decoys. Note siRNA works through a hybridization dependent mechanism as depicted for antisense.

way to discuss the practices is to put them into the context of the existing body of knowledge on the toxicity of these novel therapeutic entities. In this regard the authors are indebted to those institutions that have shared their scientific findings in the open literature allowing the science behind the application of therapeutic oligonucleotides to be shared and advanced.

There are at least four categories of oligonucleotide therapeutics: single-stranded oligonucleotides that work through RNAse H or steric effects, double-stranded RNAs that work through the dicer pathway, immunostimulatory oligonucleotides that work by stimulating innate immunity, and aptamers that work by binding to proteins and modifying their function (Figure 1). (A fifth group, transcription factor decoys, is not going to be covered in this review.) These are a diverse group of agents, with widely differing chemistries.

Because of such diversity the testing strategies that work for some compounds may not be applicable to all, or in some cases there are insufficient published data to know whether a strategy works.

## 2 CHEMISTRY

### 2.1 Aptamers

Aptamers are nonnaturally occurring nucleic acid derived molecules that bind to protein and peptide targets with high affinity and specificity. A typical therapeutic aptamer comprises 15 to 40 nucleotides, folds into a unique secondary and/or tertiary structure that allows specific target recognition, and contains a substantial number of nucleotides with modifications (e.g., 2′-methoxy) that prevent degradation by serum nucleases. Aptamers are often derivatized site specifically with high molecular weight polyethylene glycol (PEG) to increase their effective size, decrease renal filtration, and increase circulating half-life in vivo. Most commonly aptamers are used to target extracellular protein interactions. That way the mechanisms of action of aptamers are more comparable to monoclonal antibodies or traditional drugs than to antisense ODNs or siRNAs that act by hybridization to mRNAs intracellularly.

### 2.2 Immunostimulatory Oligonucleotides

Immunostimulatory CpG oligodeoxynucleotides are approximately 20 to 30 nucleotides in length, with phosphorothioate or occasionally phosphodiester linkages. Phosphorothioate linkages replace the more labile phosphodiester linkages that increase the resistance of the oligonucleotide to nucleases. The distinguishing features of these oligonucleotides are their palindromic hexamer motifs that stimulate receptors in the innate immune system through interaction with TLR9 receptor. The compounds that Coley Pharmaceuticals, Dynavax, and others are currently developing share a great deal of similarity with the early first generation phosphorothioate oligodeoxynucleotides used as antisense agents: that is to say, they are roughly 6000 to 7000 Da molecular weight with around 20 to 30 negative charges. Because each phosphorothioate linkage is a chiral center, there are $2^{N-1}$ stereoisomers possible (where $N$ = number of nucleotides). The phosphorothioate linkages tend to increase protein binding compared to the phosphodiester linkages of endogenous nucleic acids. The increase in protein binding has important effects on distribution and toxicity of these drugs (reviewed in [1]).

Newer classes of immunostimulatory oligonucleotides that are composed of RNA and bind different receptors are now being identified. Single- and double-stranded RNA have the potential to be proinflammatory, but the receptor and sequence-dependence are distinct from CpG oligonucleotides. These RNA oligonucleotides tend to bind to TLR3, 7, and 8 as well as the

more recently characterized soluble cytoplasmic receptors such as RIG-1 (see below). Like antisense DNA, the proinflammatory effects of RNA are a property to potentially be exploited for immunomdulatory effects or minimized in the case of siRNA.

## 2.3 Single-Stranded Oligonucleotides Used for Antisense Activity

For simplicity we will concentrate on two classes of single-stranded antisense oligonucleotides (ASOs) . The first are those with phosphorothioate backbones. These compounds can have both deoxy ribose and ribose sugars. Oligonucleotides that employ RNase H to cleave the target mRNA require at least a portion of the molecule to be DNA-like because the RNase H recognizes a RNA/DNA duplex. These compounds are approximately 20 nucleotides in length with 19 negative charges and $2^{19}$ stereoisomers. These compounds differ from endogenous DNA (or RNA) by the substitution of a sulfur for a nonbridging oxygen in each phosophodiester linkage. In addition some oligonucleotides are modified at the 2′ position of the ribose with methoxyethyl (second-generation ASOs) or methoxy groups. This modification increases affinity for the mRNA target and slightly alters protein binding. Most important, it markedly reduces the susceptibility of the oligonucleotide to exonuclease cleavage, and methoxyethyl modifications tend to slightly diminish the binding of the phosphorothioate oligonucleotide to proteins. Even this small change in protein binding has significant consequences to the pharmacokinetic and toxicologic properties (reviewed in [1]).

The second class of single-stranded ASO are the morpholinos. These compounds have the typical heterocyclic bases attached to a substituted morpholino ring system. These are linked together with dimethylamino phosphinyldeoxy functions. These compounds bind and act through steric action on the mRNA. The morpholinos differ from immunostimulatory oligonucleotides and from RNase H-utilizing oligonucleotides because they are uncharged. The absence of a charge diminishes protein binding significantly, which has important implications for toxicities and for distribution. Binding to plasma proteins has the beneficial effect of protecting the oligonucleotide from glomerular filtration, and enhancing binding to cells and cell uptake. Without charges, the morpholinos are more susceptible to rapid excretion in urine. On the other hand, the minimal protein binding of morpholinos reduces their potential to have toxicities associated with protein binding, for example, changes in activated partial thromboplastin time (aPTT) [2].

## 2.4 Double-Stranded RNAs–siRNA

These oligonucleotides differ from the other classes of nucleotides described above in that they are duplexes rather than single stranded. Because this class of nucleotides is relatively new and because there are two strands that can be modified chemically, it is more difficult to generalize about the chemistry of

this class. Because unmodified RNA is so labile, some chemical modifications or protection from nucleases are needed to reduce degradation in plasma and tissues in order to be effective in vivo. Some siRNAs utilize phosphorothioate linkages, and others do not. Sugar modification such as 2′ O-methyl and 2′ flouro have also been used to improve durability and affinity for targets, often in complex repeating patterns in either the sense or antisense strand of the duplex (reviewed in [3]). What is critical to this chapter is that the duplexed oligonucleotide has different physical properties than a single-stranded nucleic acid and a molecular weight that is twofold greater. The increase in molecular weight and the duplex nature significantly affect the cell uptake of these compounds. Unlike single-stranded antisense oligonucleotides, these double-stranded oligonucleotides do not readily penetrate cells. This has led some researchers to conjugate them with small lipophilic molecules like cholesterol: a modification that may have significant consequences for the toxicity profile. Liposomal formulations are the other approach that has been taken to improve cell uptake with siRNA therapeutics. The use of complex formulations sets siRNA apart from the other classes of oligonucleotide therapeutics.

## 3 PHARMACOKINETICS

### 3.1 Assessing the Pharmacokinetics of Oligonucleotide Drugs

On the surface the assessment of the pharmacokinetics of oligonucleotide therapeutics is not entirely different than the assessment of the pharmacokinetics of traditional drugs. The simplest assessment of oligonucleotide pharmacokinetics can be obtained by using radiolabeled compounds with nonexchangeable tritium on one or more of the heterocycles [4,5]. Mass balance experiments with compounds so labeled provided the first data on terminal elimination half-lives and organ distribution. In addition, depending on site of the radiolabel, it is possible to follow the labeled heterocycle through catabolism to the end product $CO_2$. Phosphorothioate oligonucleotides can be labeled on the sulfur atom, but these can be labile and the resultant products are low molecular weight entities found in urine, possibly in the form of sulfate, thiophosphate, or low molecular weight thiated oligonucleotide fragments.

With the advent of sensitive bioanalytical assays like capillary gel electrophoresis, hybridization-dependent ELISA [6] and LC/MS [7–9], it is now possible to perform toxicokinetic analyses as part of toxicity studies either in satellite groups or in study animals without the use of radiolabel [10]. These techniques provide a basis of assessing the exposure of target organs in toxicology studies and allow for the determination of critical concentrations for the production of morphologic or functional changes. These relationships can be used to design dose and schedule for clinical trials [11].

One strategy being used for a full assessment of the pharmaco/toxicokinetics of oligonucleotide therapeutics is to perform a mass balance study with

radiolabeled material followed by quantitation with combustion of tissues and scintillation counting or with whole body autoradiography. This type of study provides an understanding of distribution and routes of clearance. It is possible to also collect information on the metabolic fate of the oligonucleotide, but more often metabolite identification and toxicokinetics can be performed on samples from animals dosed with nonradiolabeled material using HPLC, CGE, or LC/MS [10].

If a novel oligonucleotide chemistry or route of administration is being tested for the first time in a species, it is advantageous to understand the tissue kinetics to design dose regimens for toxicity studies. However, in most cases it is possible to estimate what the kinetics of a novel oligonucleotide sequence are based on data from other sequences with related chemistry. In those cases the pharmacokinetics studies can be performed in conjunction with subchronic studies. A strategy employed by our group uses a satellite group in large animal studies to assess tissue levels after a single dose, after a loading regimen, and then at specified intervals after the completion of the loading regimen. This design allows us to understand first-dose kinetics, as well as multiple-dose kinetics and finally drug clearance from plasma and tissues. Properly spaced sampling allows for the estimation of tissue half-lives in large animals with relatively small numbers of animals. In the animals in the toxicology study groups, we also analyze plasma and tissues samples for drug and metabolite levels at the end of the treatment and recovery periods. It is these results that allow us to relate tissue/plasma concentrations of drug and metabolites with histopathologic or functional changes.

### 3.2 A Brief Summary of the Pharmacokinetics of Oligonucleotide Drugs

***Aptameric Oligonucleotides*** The pharmacokinetics of aptameric oligonucleotides is highly dependent on the chemical modifications. In most cases these compounds are modified to have long residence time in plasma by the addition of pendants to the backbone structure consisting of polyethylene glycol. With these pendants plasma residence time can be substantially prolonged, often to a time scale of hours or days. The pharmacokinetics of a particular drug can be tailored by the modifications to yield an optimal clearance rate. For this class of drugs protein binding is essential for both prolonging half-life as well the actual activity of the drug. Longer half-lives in plasma are beneficial for aptamers that have intracellular sites of activity as they seem to increase tissue uptake [12,13]. Like PS ODNs aptamers accumulate in renal proximal tubular cells, and in resident macrophages in various organs, and are visible when stained with hematoxylin.

***Phosphorothioate Oligonucleotides*** Single-stranded phosphorothioate ODNs (PS ODN), either immunostimulatory or ASO, administered parenterally appear rapidly in plasma where they are bound to hydrophilic sites on

plasma proteins, and are thus protected from glomerular filtration. These binding sites are not the hydrophilic sites where lipophilic drugs bind, and there is little competition for plasma protein binding by ODNs [14]. Plasma kinetics are polyphasic with a short distribution phase that is on the order of hours, and then a terminal elimination phase that has half-lives on the order of days or weeks (reviewed in [11,15,16]). The initial distribution phase may also be influenced by plasma metabolism. Depending on the chemistry of the PS ODN, there can be metabolism by nucleases in plasma, or if the PS ODN is protected from exonuclease metabolism with 2′ modifications, then there is little or no metabolism in plasma.

The initial distribution phase is primarily related to binding to and distribution into tissues, with liver, kidney, lymph nodes, and spleen being the sites of highest binding and uptake. For PS ODN little renal clearance occurs because of the binding of the PS ODN to plasma proteins. However, at very high doses plasma protein binding can be saturated and spillage into urine is more likely (reviewed in [17]). Uptake into tissues is also saturable, and therefore the distribution kinetics are dose dependent, with clearance rates slowing and AUC increasing with increasing doses [16,18–20]. Once bound to cells, PS ODNs transit into cells by moving down concentration gradients from extracellular compartments to intracellular compartments, probably by shuttling from one protein binding site to another, across cell membranes. PS ODNs in cells bind to available targets, but the majority of the PS ODN in cells is probably bound to intracellular proteins (reviewed in [1]).

Within cells, tissues, and plasma, ubiquitous nucleases metabolize (or catabolize) PS ODN drugs, and not the cytochrome P450 enzymes that typically metabolize low molecular weight drugs. PS ODNs therefore do not compete with traditional drugs for metabolic processes, reducing the potential for drug–drug interactions. Whole body elimination is the result of metabolism in tissues, and a re-equilibration of metabolites and parent drug out of tissues and into circulation, where they are ultimately excreted in the urine. These processes are very similar in laboratory animals and humans. In fact the processes are similar enough that for a given chemistry (e.g., PS ODNs or second-generation PS ODNs) doses scale from species to species on the basis of body weight, not surface area, allowing for good extrapolations from laboratory animals to humans [21,22]. In these extrapolations, mice are outliers compared to rats, rabbits, dogs, monkeys, and humans. This later point is key for assessing the relevance of findings in animal studies to humans, and for the design of early clinical trials. For a given sequence and route of administration it is possible to predict plasma levels in human subjects directly from data in monkeys with a high level of confidence in PS ODNs, and to some extent it is possible to generalize from sequence to sequence, albeit with less precision than within a single sequence.

After subcutaneous or intravenous administration, PS ODNs tend to accumulate in kidney, liver, spleen, bone marrow, and lymph nodes (reviewed in [1]). While the kidney generally accumulates the greatest concentration of

oligonucleotide, the liver with its greater mass accumulates a greater total burden. Approximately 20% to 30% of the total dose accumulates in liver. At the cellular level oligonucleotide can be found in all cell types in the liver, with the phagocytically active Kupffer cells accumulating it in phagolysomes. When these cells are stained with hematoxylin, the oligonucleotides appear as basophilic granules. Similarly the phagocytically active proximal tubular cells also tend to accumulate oligonucleotide in basophilic granules [23,24]. As dose is increased, liver and kidney concentrations tend to saturate and that leads to greater distribution to the spleen, bone marrow, and other lower avidity organs [18].

***Single-Stranded Morpholino Oligomers*** The phosphorodiamindate morpholino oligomers (PMOs), because of their low level of protein binding, have very different kinetics than all of the other classes of oligonucleotide therapeutics. Limited protein binding makes these compounds more rapidly excreted in urine than other oligonucleotides, and initial clearance is therefore less dependent on tissue uptake than it is on glomerular filtration. Clearances for PMOs from plasma in rats is between 1 and 33 ml/min with a volume of distribution of 0.4 to 56 L/kg [2]. The plasma elimination half-life is variable, with sequence and tissue half-lives in liver and kidney reported to be as long as 7 to 14 days. Metabolism does not appear to be a key factor in terminal elimination of PMOs, with greater than 95% of the dose recovered intact. Because of the lack of charge, sequence plays a much greater role in determining the pharmacokinetics of the oligomers than it does for the charged PS ODNs or siRNAs. In those molecules the polyanionic nature tends to dominate over the chemical characteristics associated with differences in sequence. For PMOs then there is less predictability from drug to drug, which is a key disadvantage relative to PS ODN.

In the clinic a single slow bolus injection of 90 mg of AVI 4126 had a multiphasic distribution plasma kinetics with half-lives of approximately 1 and 11 hours, but no data were reported on elimination half-life. The volume of distribution was roughly 250 ml/kg, somewhat different than those seen in rats. Metabolites were not detected in plasma [25].

***Double-Stranded RNA-Based Therapeutics*** The pharmacokinetics of unmodified double-stranded RNAs is largely a product of rapid degradation, urinary filtration, and tissue uptake. The role of enzymatic degradation can be minimized through modifications to the backbone, often 2′ F or 2′ O-methyl groups [3]. Unmodified double-stranded RNA and double-stranded RNA with one of the two strands a full diester and the other strand a full phosphorothioate (PO/PS duplex), distribute rapidly to the liver and kidney after injection, peaking at 5 minutes, with radiolabel appearing in bladder urine over the same time frame. Oligonucleotides labeled with $^{125}$I on the 3′-termini administered by intravenous injection and radiolabel can be found in the kidney and liver as soon as 1 hour and as late as 72 hours (the first and last observations). The

highest concentrations is observed in kidney. The more stable PO/PS duplex accumulated slightly more in spleen, lung, and heart, and slightly less to liver and kidney [26].

## 4 TOXICITY OF OLIGONUCLEOTIDE THERAPEUTICS

### 4.1 Overview

All the oligonucleotide therapeutic agents discussed in the review are synthesized chemically and none are produced in biological reactions. As a result these compounds have been regulated not as biologics, but as "drugs." Oligonucleotide-based therapeutics are biotechnology-dervied products and could be covered by ICH S6. Some aspects of oligonucleotide therapeutic agents are much more akin to biologics. For example, the chemical characterization of these compounds often shows that they have complex profiles more similar to a biologic agent than a traditional drug [27]. Additionally their metabolism is more akin to catabolism in that they are reduced to nucleotides by nucleases much the same as biologics are catabolized to amino acids. However, because oligonucleotides were synthetically derived, they have been tested in nonclinical assays like small molecules, with complete toxicity characterizations in genotoxicity, safety pharmacology, genotoxicity, subchronic, and, when appropriate for the indication, chronic toxicity and carcinogenicity assays.

When trying to understand the toxicity of oligonucleotide therapeutics, it is useful to classify the sources of toxicity. An oligonucleotide therapeutic can produce toxicity either through a hybridization-dependent or hybridization-independent mechanism.

***Hybridization-Dependent Toxicities*** Toxicity can be induced through hybridization in two ways. First, through exaggerated pharmacology: hybridization reduces the expression of the target mRNA and produces toxicity related to the intended gene product. This type of toxicity is possible, but is usually discovered in preclinical pharmacology studies. The easiest way to avoid this type of toxicity is to carefully select pharmacology (gene) targets that are not critical for cell survival. As is obvious but bears stating, the only way to induce exaggerated pharmacology in a toxicology study is to have an active oligonucleotide. Because mRNA sequences are often species specific, that means that it may be necessary to test different sequences in the toxicology models in addition to, or in place of, the sequences that will be used in human clinical trials. While it is possible to select oligonucleotide sequences that have activity across species, it often happens that specific surrogate (homologous) oligonucleotides are needed for one or more of the animal species used in toxicity studies. A broader discussion of surrogate drugs appears below.

The second mechanism of hybridization-dependent toxicity is related to the reduction in expression of an unintended target by an antisense mechanism.

These so-called off-target effects are uncommon, on the basis of the low probability of having a perfect match to a 20-mer sequence two places in the genome. The probability of having an off-target effect can be further reduced by performing the appropriate blast searches for matches in other known genes in the expressome and selecting sequences for drugs that are unique.

There may be marked differences on the potential for the different classes of oligonucleotide drugs to have hybridization-dependent off-target effects. Aptamers are generally decorated with PEG groups, so they may have reduced affinity for complementary sequences and, more important, a reduced propensity for entering cells. On the other hand, because aptamers are designed to interact with specific proteins, it is possible for off-target effects with aptamers to be related to off-target protein binding.

The potential for off-target effects with siRNAs is just becoming recognized. The molecular basis for siRNA activity makes off-target phenomena more likely to occur with siRNA than with oligonucleotides that work through RNase H. There are three primary reasons for the greater potential for hybridization-dependent off-target effects [28]. First, in an siRNA two different sequences are being administered, the "sense" and "antisense" strands. Second, there is greater tolerance of mismatches with this mechanism. Only a portion of the 3′ region of the antisense strand of the siRNA needs to be a perfect match. The so-called seed region matches allow for many more potential binding sites than would be allowed if full complementarity (of the 20-mer) was required for the cleavage of a target mRNA [28]. Third, it has been reported that siRNA can induce changes in the methylation state of genomic DNA and thereby alter gene expression (either silence or activate off-target genes through an interaction with the RNA-induced initiation of transcriptional gene silencing—the RITS complex) [29]. Thus, because siRNA can produce more than just specific reductions in target protein expression, this mechanism brings with it the greater potential for hybridization-dependent off-target effects than RNaseH mediated antisense activity. However, this issue has been identified early in the understanding of siRNA mechanisms, and therefore it can be addressed as these drugs advance in toxicity studies and clinical trials.

***Hybridization-Independent Toxicities*** Many, if not most, of the toxicities that have been observed with PS ODN and other oligonucleotide classes are related to the hybridization-independent effects. These class effects are related to the chemistry, and many are known to be related to the interaction of oligonucleotides with proteins (reviewed in [17,30,31] [1]). Toxicities such as the prolongation of aPTT, the activation of complement, and immunostimulation are all examples of oligonucleotide protein interactions that are independent of hybridization. Thus reducing protein binding also reduces the likelihood of some of these toxicities, but that comes at a cost. Reduced protein binding

reduces uptake into cells and increases the fraction of the drug that gets filtered by the kidney (and it may increase renal concentrations). So protein binding is a factor that needs to be understood, not just avoided.

Some hybridization-independent toxicities are sequence-dependent. The most obvious example of this is CpG oligonucleotides that have specific sequence motifs that make them more immunostimulatory through interactions with Toll-like receptors or protein components of innate immunity (reviewed in [32]). While immunostimulatory effects are the basis of the pharmacologic activity for a whole class of oligonucleotides, for single-stranded antisense compounds and compounds that work via antisense or through the siRNA pathways, stimulating innate immunity is considered an undesirable effect [33,34]. The CpG motifs have been well characterized as motifs that control hybridization-independent effects of oligonucleotides. Whether there are other motifs that have there own unique "toxic" or pharmacologic effects remains to be determined.

In our experience with PS ODNs, there may be quantitative variation in potency for some toxicities, but there are a great number of qualitative similarities in toxicity profiles. A number of the toxicities may be considered common to the chemical class. The responses from one PS ODN sequence to the next are similar within a single species. We and others have found that there are striking differences in the responses of rodents and nonhuman primates.

***Surrogate Oligonucleotide Drugs*** Because of species-dependent sequence differences, siRNA and ASOs may not hybridize with the mRNA for the target gene in the animal species used in toxicity studies. With oligonucleotide therapeutics (excluding aptamers) it is possible to design and synthesize species-specific compounds that are perfectly complementary to the mRNA for the target gene in a specific species. Not only is it possible, but from what we know about the synthetic process and what is understood about synthesis-related impurities, we can predict that species-specific oligonucleotides will have similar chemistry manufacturing and controls as the sequence used in clinical trials. These animal-specific drugs (surrogates) can be tested alongside the human sequence in toxicity studies to establish if there are biologic effects (exaggerated pharmacology) associated with downregulating the target gene. This concept has been part of the toxicity profiling of antisense drugs for the last decade [35,36]. One of the reasons that this approach is viable is that the class-related toxicity profiles and the pharmacokinetics are so similar from sequence to sequence. Thus, if there is a unique toxicity profile in the surrogate drug, it most likely arises from the effects on the expression of the target. For PS ODNs, this differentiation is useful because there are a number of class-related toxicities like the propensity for immunostimulation and mononuclear cell infiltrates in rodents that are not clinically relevant. Using surrogates allows one to distinguish between these "nonrelevant" toxicities and toxicities

related to reductions in target expression. In the toxicity assessment of low molecular weight drugs or some biopharmaceuticals, investigators simply increase the dose if the pharmacologic effects of a drug are weaker in animals than in humans. For siRNA and ASOs, simply overdosing will not produce an antisense effect through hybridization. So unlike low molecular weight drugs, overdosing is not a practical solution for overcoming low activity. There are, of course, exceptions, and these need to be considered. For example, if there are one or two mismatches near the termini of an antisense or siRNA construct, it is possible that there would be an antisense effect. The degree of activity depends on the placement of the mismatches and is sometimes target specific. In these cases we would propose that there is no need for using a surrogate drug sequence to characterize exaggerated pharmacology, but including a nonactive control group might be considered for comparison purposes to distinguish class-related effects from exaggerated pharmacology.

In practice, there are times when surrogates need not be used simply because the target does not exist in the toxicity study species (or has a completely different function) or the target is related to an infectious agent not normally present in the toxicity study species. Aptamers may represent a unique challenge in this regard. It may not always be possible to develop species-independent aptamer, and because aptamer design is more empirical than antisense and siRNA, the choice of a surrogate is not clear, nor is there any assurance that an aptamer surrogate would have the same molecular mechanism. Thus the surrogate approach is not as rational for aptamers.

The scope of the toxicity studies when surrogate molecules have been used includes pharmacology studies, subchronic and chronic studies, reproductive toxicity studies, immunotoxicity studies, and even carcinogenicity studies. Short-term assays like safety pharmacology studies are too short for there to be an antisense effect, so we have not used the surrogate approach in these assays. However, when safety pharmacology endpoints are included in subchronic or chronic studies, surrogates are assessed.

When designing a toxicity study with a surrogate drug, for practical purposes, it is necessary to decide whether to run a full dose-response curve for the surrogate or for the human sequence. We have elected to use a single dose of the surrogate, one that is known to produce significant (if not maximal) pharmacologic activity. This dose is often not equivalent to the high dose in the study because the high dose is generally associated with marked nonspecific effects. One could argue for a full dose–response analysis for the surrogate, but for practical purposes and humane reasons a well-selected single-dose group should be sufficient unless there is marked toxicity associated with the reduction in the target protein.

### 4.2 Genotoxicity

***Assessing the Genotoxicity*** Most of the published data on the genotoxicity of oligonucleotide therapeutic agents is for the PS ODNs. PS ODNs have

been routinely examined in the standard battery of genetic toxicology assays that include the Ames bacterial mutagenicity assay, the CHO cell chromosomal aberrations assay, the L5178Y/TK (+/–) mouse lymphoma mammalian gene mutation assay, and the mouse micronucleus assay. Because these compounds are highly soluble and have low cytotoxicity, most in vitro assays have been run at or near the limit concentration of 5000 μg/ml. While there is little published data on the other classes, because they have been accepted for clinical trials, we assume that the findings with the other classes of oligonucleotide therapeutics are not markedly genotoxic.

***Genotoxic Potential of Oligonucleotide Therapeutics*** Published data from genotoxicity assays for a number of different PS ODNs indicate that there is little or no genotoxic potential. Additionally ISIS has tested more than 10 unique sequences, all routinely negative (Table 1).

Representative example of these assays have been thoroughly reviewed and published for PS ODN [37]. Included in these assays are both the demonstration of cytotoxicity and the demonstration of cellular uptake of the test compounds, such that there is clear evidence for exposure of cells to oligonucleotide and any metabolites that might be formed.

There was concern that metabolism of these compounds at the high concentrations used in genotoxicity assays would produce changes in the nucleotide pools. This was unfounded as the compounds have been routinely negative for the 2′-methoxyethyl modified PS ODNs even with the presence of modified nucleosides. However, in the testing of several 2′-MOE ASOs, no genetic toxicity has been observed. It turns out that the 2′-O-methoxyethyl modified nucleotides are highly resistant to exonuclease cleavage, so little 2′-MOE modified nucleotide is released.

Because of the uniform negativity of in vitro genotoxicity assays, EMEA has published guidance on the genetic toxicity potential for antisense oligonucleotides [38]. The document has two primary concerns: (1) the potential for base mispairing due to liberation of metabolites, resulting in point mutations in newly synthesized DNA, and (2) site-specific mutations resulting from triplex formation with DNA [39]. The traditional gene mutation and clastogenicity assays are considered of suitable specificity and sensitivity to address the first potential issue. However, based on the negative response in these assays, in light of the documentation of exposure and metabolism, the EMEA concluded that PS ODN were not likely to pose a genotoxic hazard by this mechanism. As a result the report suggested that in vitro testing of novel antisense inhibitors in these assays was not necessary. The FDA has not accepted similar arguments, so in practice, for drug registration in the United States, it is still necessary to do the in vitro tests.

In vivo genotoxicity studies or clastogenicity studies are not mentioned in the EMEA guidance, and PS ODNs have been routinely tested in the mouse micronucleus assay. Again, even when exposure to the bone marrow is well documented, the assays are routinely negative. Other assays like unscheduled

**TABLE 1 Summary of genetic toxicity assays**

| | | Bacterial Mutagenesis | | | In vitro Assays | | | In vivo Assay |
|---|---|---|---|---|---|---|---|---|
| | Isis Compound Number | Salmonella | E. coli | Cytogenetics (Chromosomal Aberation) | Mouse Lymphoma Mammalian Mutagenesis | CHO/HGPRT Mammalian Mutagenesis | Rat Hepatocyte UDS | Mouse Micronucleus |
| First generation | 2,105 | Neg | — | Neg | — | Neg | Neg | Neg |
| | 2,302 | Neg | Neg | Neg | Neg | — | — | Neg |
| | 2,922 | Neg | Neg | Neg | Neg | — | — | Neg |
| | 3,521 | Neg | Neg | Neg | — | — | — | Neg |
| | 5,132 | Neg | Neg | Neg | — | — | — | Neg |
| Second generation | 107,248 | Neg | Neg | — | Neg | — | — | — |
| | 112,989 | Neg | Neg | — | Neg | — | — | — |
| | 104,838 | Neg | Neg | — | Neg | — | — | Neg |
| | 113,715 | Neg | Neg | — | Neg | — | — | — |
| | 301,012 | Neg | Neg | — | Neg | — | — | Neg |
| | 325,568 | Neg | Neg | — | Neg | — | — | — |

DNA synthesis assays have been run occasionally for PS ODNs and have been negative.

The PS ODNs are metabolized to nucleotides that are identical to endogenous nucleotides. It is possible, and perhaps even likely, that as oligonucleotide therapeutics move further away from building oligonucleotide therapeutics from the endogenous nucleotides to building the oligomers from synthetically modified nucleotides and heterocycles (bases), and that mutagenic components will be accidentally or intentionally included. The result could be a drug or metabolic product that is mutagenic. For that reason when new chemistries are included in oligonucleotide therapeutics, it is important that they be tested in the genotoxicity assays that are appropriate.

### 4.3 Safety Pharmacology

***Assessing the Safety Pharmacology of Oligonucleotide Therapeutic Agents*** The battery of safety pharmacology studies for oligonucleotide compounds can be ascertained from the ICH guidance on the topic.

> The purpose of the safety pharmacology core battery is to investigate the effects of the test substance on vital functions. In this regard, the cardiovascular, respiratory and central nervous systems are usually considered the vital organ systems that should be studied in the core battery. In some instances, based on scientific rationale, the core battery may need to be supplemented or may not need to be implemented.
>
> The exclusion of certain test(s) or exploration(s) of certain organs, systems or functions however should be scientifically justified.

The safety pharmacology of oligonucleotide therapeutics should include application of the generally recognized guidance where there is scientific rationale. Clearly, understanding the effects of oligonucleotides on vital organ functions is an important part of all toxicologic assessments. What is less clear is whether there is a requirement for tests for which there is little scientific rationale. For example, oligonucleotide therapeutics administered by subcutaneous, intravenous, or oral routes do not cross the blood–brain barrier, and there is little or no accumulation in the brain. As such, these compounds might be expected to have no CNS effects. Under conditions where there is little or no exposure, is there a sufficient rationale to justify the study? Other batteries of safety pharmacology studies are more easily justified scientifically.

There are other complex issues with regard to the assessment of safety pharmacology studies. With the oligonucleotide therapeutics that modulate the translation of mRNA to proteins, like antisense/siRNA, there is a lag between administration of the drug and the pharmacologic activity that is mediated by reduction in protein levels. This lag is related to the mechanism of action and how long it takes for a reduction in protein synthesis to be reflected in reduced protein levels. Because of this lag it is probably better to

perform as much of the safety pharmacology battery as part of repeated-dose toxicity studies as possible, and not as acute studies.

***Cardiovascular Safety*** Effects on vascular tone have received the most attention, as a result of toxicities that were identified very early in the development of the first systemically administered PS ODNs. We showed that there was correlation between complement activation and the reported alterations in blood pressure and heart rate in monkeys treated with high doses of PS ODN [40–42]. The observation appears to be unique or at least more prominent in monkeys, and that has actually driven the need to characterize the toxicity of oligonucleotide therapeutics in monkeys as the nonrodent species. (Note that complement activation has not proved to be a significant issue in the clinic, although in clinical studies we strive to avoid plasma levels that might be associated anaphylactoid responses in monkeys.)

Because the monkey is the most sensitive species for the effects on complement, we have opted to perform cardiovascular safety studies as part of subchronic study protocols using implanted telemetry units in selected monkeys (two/sex in at least two different treatment groups, including the high dose) to measure ECG, mean arterial pressure, heart rate, and body temperature. These parameters are then recorded at frequent intervals for a 24-hour period prior to treatment to establish circadian fluctuations and normal response to various stimuli encountered during the day. This technique enables us to assess both acute alterations related to complement activation after single doses and chronic safety after repeated administration.

The in vitro hERG assay has become the standard for testing for effects on QT prolongation. Various PS ODNs have been tested against the human ether-a-go-go channels, and under the conditions of the assays there is no apparent affect up to 1050 μg/ml reviewed in [43]). These concentrations are many-fold higher than the concentrations attained in hearts in animals in toxicity studies. Thus, based on the very large margins of safety, there is little concern for QT prolongation in patients. This is not surprising, considering the chemical nature of the oligonucleotides as polyanionic molecules being approximately 7000 Da. Most of the compounds that have typically been associated with alteration of ion channel function have been small compounds capable of interacting with the ion channel. We would therefore recommend that the hERG assay be eliminated from the battery of safety pharmacology studies.

***Renal Safety Pharmacology*** Oligonucleotides accumulate in the kidney, particularly in the proximal convoluted tubular epithelial cells. At high doses in monkeys the kidney is a target organ for toxicity. Thus assessment of renal function for oligonucleotide therapeutics is essential. The concerns for renal toxicity are largely attributed to the concentration of oligonucleotide in tissue over time, and therefore it is probably inappropriate to examine renal function in the context of the typical single-dose experiments used in safety pharmacology assays. It is more relevant to examine renal function as part of the repeat-

dose toxicology studies. For renal safety, monkeys, with their pharmacokinetics similarities to humans, are considered by the authors to be the most relevant species. Evaluation of renal function relies primarily on routine serum chemistry, urinalysis, and histopathology. In addition quantitative urine total protein and creatinine ratios can be employed. Protein–creatinine ratios are a well-accepted biomarker in human clinical trials. Other measures of tubular function have occasionally been examined such as renal blood flow, glomerular filtration, urine, glucose, urine amino acids, measurement of specific low molecular weight proteins in urine, and enzyme markers of epithelial cell damage in urine. These assays are designed to assess tubular transport or the viability of the proximal tubular epithelium. Markers like α-glutathione-*S*-transferase, *N*-acetylglucose aminidase, and retinol-binding protein levels have all been applied to nonclinical and clinical assessment of renal function of oligonucleotide therapeutics. Because of difficulties in obtaining clean urine samples for these analyses, cystocentisis samples are preferable, but cage pan collections have been used. In general, no changes in these markers have been observed even when kidneys are stressed by glucose load or under conditions requiring the concentration of urine. The most reliable measure of renal changes has been total urine protein or protein–creatinine ratios.

***Pulmonary and Gastrointestinal Safety Pharmacology*** Pulmonary and gastrointestinal function studies have been performed for PS ODNs and are uniformly negative in the typical rodent assays. In one experiment a collection of PS ODN sequences were administered to rats, and CNS, pulmonary, and cardiovascular function were evaluated. Included in this group of compounds were both active and inactive compounds for the rat. The results show that the compounds are uniformly negative in this series of short-term safety pharmacology studies, suggesting that as a class these compounds do not affect these parameters (Isis Pharmaceuticals, unpublished data).

For pulmonary safety pharmacology studies, some laboratories are collecting the pulmonary safety data as part of subchronic or chronic toxicity studies by collecting data on respiratory rates and blood gases. Using subchronic treatment may be both more relevant and may reduce the numbers of additional studies and animals to perform the assessment. Although the preponderance of the data suggest that with the current PS ODNs, including 2′-MOE modified compounds, there are no acute pharmacologic effects for the class. Obviously collecting additional data will provide a rationale for only doing scientifically relevant assays.

***CNS Safety Pharmacology*** CNS safety pharmacology studies do not appear to be necessary based on our understanding of the biodistribution of PS ODNs. These drugs, because of their polyanionic nature, do not cross the blood–brain barrier; therefore, after administration by intravenous or subcutaneous injections, there is little or no accumulation in the brain. Some have argued that there are low accumulations in brain from analyses of whole brain

homogenates and therefore it is not unreasonable to perform CNS safely pharmacology. However, quantitative whole body autoradiographs suggest that the neuronal tissue is not accumulating oligonucleotide.

For ASOs injected or infused directly into the CNS, the situation is quite different Neurons are known to take up PS ODNs, and antisense activity in CNS can be shown after local injections or intraventricular infusions. For PS ODNs that are administered locally to the CNS, safety pharmacology studies should be performed. For systemically administered PS ODNs, the scientific rationale is weak.

### 4.4 Acute Toxicity

Oligonucleotide therapeutics have a relatively low order of acute toxicity, with the notable exception of the polyanionic PS ODNs, which have been known to activate complement through the alternative pathway in monkeys. Because these compounds do not cross the blood–brain barrier nor accumulate in cardiac muscle, and because the uptake of these compounds is dependent on saturable processes, very high doses of oligonucleotides tend not to produce central effects or acute organ failure, thus lowering the potential for acute lethality.

***Assessing Acute Toxicity*** Assessment of acute toxicity can be performed in rodents as a stand-alone study, as part of acute toxicity screens, or as part of the dosing ranging process for the mouse micronucleus assay. Acute studies in nonrodent species have not generally been performed for PS ODNs because of the well-understood activation of a complement cascade that can result in an anaphylactoid-like response. This response will be similar for each sequence. For phosphorothioate oligodeoxynucleoitdes without 2′ modifications, there is a remarkably reproducible threshold plasma concentration for the drug that is associated with activation of the complement. The same threshold was observed for multiple sequences. PS ODNs that are partially modified with 2′-methoxyethyl groups share properties, though the threshold is greater than for unmodified PS ODNs (reviewed in [31]). Because of this reproducible toxicity and the well-understood nature of the complement activation, acute toxicity studies for PS ODNs in primates are difficult to justify.

If one uses a novel chemistry, a pyramiding dose escalation type of study in primates may provide useful information for dose selection. However, the class similarities in the toxicity profile, short of using a novel chemistry, suggest that such study would be of limited value for most PS ODN therapeutics.

***Acute Toxicities*** The estimated lethal doses in mice of various PS ODNs with and without 2′-MOE modifications are all approaching 1 g/kg, indicating a low level of acute toxicity in mice (Table 2). In monkeys, a high-dose administration of PS ODNs results in the inhibition of factor H of the alternative complement cascade. Factor H is a endogenous inhibitor of the

**TABLE 2 Estimated dose required to produce 50% mortality in mice injected intravenously**

| | Sequence | LD50 (mg/kg) |
|---|---|---|
| First generation | ISIS 2105 | 890 |
| | ISIS 2922 | 720 |
| | ISIS 2302 | >1000 |
| | ISIS 3521 | 500 |
| | ISIS 5132 | >1000 |
| Second generation | ISIS 104838 | 2000 |
| | ISIS 301012 | >2000 |

complement cascade, and removing this "brake" from the system may promote complement activation and the anaphylactoid-like response that has been so well characterized [40–42]. When plasma concentrations of unmodified PS ODNs exceed 50 to 75 μg/ml, there is activation of the cascade in monkeys. The end result is the release of anaphylatoxins C3a and C5a. In vitro assays with human plasma indicate that humans may be less susceptible to the activation of the complement cascade, and it appears that in human plasma the cascade may be more susceptible to inhibitory effects of PS ODN (reviewed in [31]). The complement activation is directly responsible for the cardiovascular collapse observed acutely in nonhuman primates. This relationship was definitively demonstrated using a specific complement inhibitor CAB2 [42]. Because of this clear mechanistic relationship, acute toxicity studies in monkeys with PS ODNs simply reproduce what is already known about complement activation and add little to our understanding of the toxicity of a particular sequence.

Another effect associated directly with plasma concentrations is related to a transient concentration-related inhibition of the tenase complex by PS ODNs [44–46]. The inhibition results in a prolongation of aPTT, but has little or no effect on PT. Only one arm of the clotting cascade is affected, and the effects are transient. We have demonstrated in both clinical trials and in nonclinical trials that there is less than a 1 second increase in aPTT per μg/ml. Typical concentrations of PS ODNs after intravenous infusion of 3 mg/kg in monkeys or humans are in the range of 10 to 20 μg/ml, suggesting that at $C_{max}$ there might be a transient increase in aPTT of 10 to 20 seconds [47,48]. This inhibitory effect is directly proportional to plasma concentrations so that, as plasma is cleared of PS ODN, the inhibition reverses. With a distribution half-life of less than 1 hour, this effect is transient and is almost completely reversed within 3 hours. When oligonucleotides are administered by the subcutaneous route, $C_{max}$ is blunted and typical peak plasma concentrations after a dose of 3 mg/kg are in the range of 3 to 5 μg/ml.

Thus for subcutaneously administered PS ODNs the prolongation is not clinically significant. Even in monkeys treated with high doses of PS ODN

there has been no indication of hemorrhage, though at very high doses (~50 mg/kg) bruising has been observed [49]. In our clinical experience dosing more the 3000 patients, no significant adverse effects have ever been reported related to prolongation of aPTT.

Some oligonucleotide therapeutics (aptamers) have been designed specifically to inhibit clotting, and for those compounds the anticoagulant effects are more pronounced, as intended. The doses used in toxicity studies are designed to characterize the superpharmacology of these agents.

Because PMOs are not charged and do not contain the phosphorothioate linkages, they do not have the same potential for these acute effects. However, they will have their own unique toxicities. To date, PMOs have not been published in any detailed form.

The double-stranded siRNA chemistries that are being employed include some of the modifications (e.g., phosphorothioate) that result in protein binding and are therefore charged and capable of binding to plasma proteins. Acute toxicity data for these compounds have not been reported at this time. Because they consist of two 20-mer strands, they still have the potential to produce similar kinds of toxicities as other PS ODNs.

By encapsulating oligonucleotides in liposomes, it is possible to reduce the acute effects of siRNA compounds. This was demonstrated for a PS ODN in a "stealth" liposome formulation [50]. These types of delivery systems are being increasingly employed with siRNA. Oligonucleotide formulations can reduce the propensity to produce the typical PS ODN effects, but it will be necessary to characterize their acute effects.

### 4.5 Subchronic and Chronic Toxicity

The protocols for performing subchronic and chronic studies with oligonucleotide therapeutics are not dissimilar to those used for low molecular weight drugs, with a few exceptions. Much of what is done today to characterize the toxicity of this class of drugs is relatively unchanged from the Points-to-Consider documents on oligonucleotide therapies, by Black and others [35,36]. Additional understanding and experience gained in the past decade make it appropriate to reconsider some of these points.

***Assessing Subchronic and Chronic Toxicity*** Subchronic and chronic toxicity studies for oligonucleotides have been performed primarily in rodents (often mice) and monkeys. Most other small-molecule drugs are characterized in rats and dogs. The choice of species is based on a number of factors. For small molecules, species are often selected on the basis of similarities in metabolism, pharmacokinetics, or frankly, tradition. For oligonucleotides, the catabolic pathways for degradation of the oligonucleotides are very similar from species to species. Because metabolism was not an issue, the ability of the test species to mimic responses in humans became the driving force for the selection of nonhuman primates. Selection of a relevant species to mimic the human

response is also the driving force for species selection for biopharmaceuticals. Mice were generally used as the rodent species for early studies on the toxicity of oligonucleotides because many of the pharmacology models are in mice, and therefore surrogate molecules were readily available. In hindsight, this was a fortuitous decision (vide infra). Monkeys were selected originally because they were thought to be more representative of humans than dogs. Again in hindsight, that was a good decision because the pharmacokinetics of oligonucleotide drugs appear to be very similar in monkeys and humans. In fact, at similar doses on a milligram per kilogram basis, plasma levels of a given sequence administered by the same route of administration are nearly identical in monkeys and humans. Clearance, volume of distribution, initial distribution half-lives and terminal elimination half-lives are all similar between monkeys and humans. Compared to rodents, the toxicity profiles in monkeys may be more predictive of those in humans.

The selection of mice over rats was fortuitous because rats treated with PS ODNs develop a low-grade low molecular weight proteinuria. It is dose-dependent (usually observed at doses ≥10 mg/kg) and independent of sequence. Most important, proteinuria is *not* seen in mice, monkeys, or man. Proteinuria may be related to decreases in tubular reabsorption of low molecular weight proteins. This effect may be attributed to a simple competitive interaction for a particular receptor or receptors on tubular epithelium between the oligonucleotide and low molecular weight proteins. Competition for a receptor may explain why these changes are readily reversible and are not associated with other changes in kidney function. Why rats are more sensitive is not known, but may be related to the greater susceptibility of rats to development of renal abnormalities (i.e., proteinuria). Male rats are notoriously sensitive to kidney insult as evidence by the aging rat nephropathy that develops in this gender and species, and rats show unusual sensitivity to agents, like the nephropathy associated with puromycin treatment. Rats also normally excrete far more protein in their urine than monkeys or humans. Understanding of the relative species specificity of this effect and the potential mechanistic basis would not have been possible before a relatively large database of safety information was available in mice, monkeys, and humans to interpret the species specificity.

Whether dogs can be used to characterize the toxicity of oligonucleotide therapeutics is an open question. In preliminary studies, pharmacokinetics of PS ODNS in the dog appear similar to monkey and human. The dog may be a suitable model for toxicity studies, but our experience is limited and the limited genomic information available for dogs makes designing surrogate molecules more challenging. It is possible to show reductions in target protein expression in dogs with oligonucleotide sequences that are homologus to multiple species. One interesting observation is that dogs did not appear to have the same complement activation response with PS ODNs, seen in primates (Isis, unpublished observations). Thus single-dose studies in monkeys need to be performed to characterize the potential of a new sequence to

activate the complement cascade. With the acute toxicity associated with complement activation characterized in the most sensitive species (monkey), separate subchronic studies could be performed in dogs. Dog studies have been used to successfully support INDs for locally and orally administered oligonucleotide therapeutics. This strategy would provide a complete characterization of toxicity.

### Subchronic and Chronic Effects

*Basophilic Granules* One of the hallmarks of exposure to oligonucleotide therapeutic agents is the appearance of basophilic granules in cells and tissues that accumulate oligonucleotides. These basophilic granules are often found in phagocytically active cells, like macrophages, Kupffer cells, and epithelial cells and the proximal convoluted tubule of the kidney (Figure 2). Hematoxylin stains synthetic oligonucleotides like it stains endogenous oligonucleotides, and the accumulation of oligonucleotides in phagolysomes results in this distinct punctuate appearance. We have confirmed that the basophilic material consists of oligonucleotide using immunohistochemistry and fluorescently labeled oligonucleotides [23,24]. Electron microscopic examination of the tissues with granules reveals that these granules are membrane bound as would be expected if they are the product of phagocytosis or endocytosis. These granules are common to all of the oligonucleotide therapeutic classes that are taken up into cells.

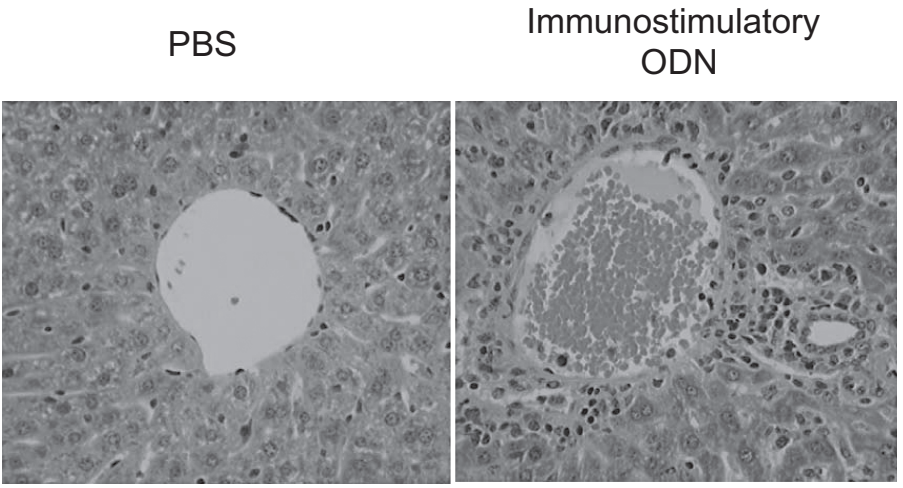


**Figure 2** Photomicrograph of the renal cortex of a monkey treated with 40 mg/kg/wk for 5 weeks with a typical second generation 20-mer oligonucleotide modified with methoxyethyl groups on the 5 residues at each terminus. Stained with hematoxylin and eosin. The arrows point toward basophilic granules in the cells of the proximal convoluted tubules that are abundant throughout the renal cortex at this dose.

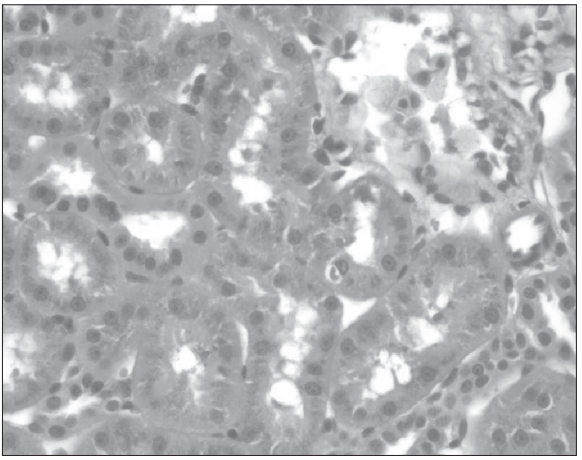

**Figure 3** Photomicrograph of sections of liver from female CD-1 mice treated by subcutaneous injection with either saline (PBS) or an immunostimulatory oligodeoxynucleoitde at 4 mg/kg ever other day for 1 week. Lymphohystiocytic infiltrates are apparent adjacent to the central vein in the section from the ODN-treated mouse.

*Lymphohistiocytic Infiltrates* The other hallmark of exposure that is common in rodents treated with PS ODNs are lymphohistiocytic infiltrates in numerous organs (Figure 3). These mononuclear cell infiltrates have been described in liver, kidney, salivary glands, pancreas, uterus, urinary bladder, as well as other tissues. These infiltrates will resolve after treatment is ceased, but the clearance of them is prolonged (reviewed in [43]). They have been described as perivascular in some tissues with mononuclear cells appearing to migrate from the vasculature. They are not a prominent feature in other species.

Histocytosis is often also present in lymph nodes and spleen. This effect is related to the pro-inflammatory effects of PS ODNs. These effects are primarily observed in rodents and are associated with cytokine and chemokine release. The uncharged PMOs have also reported increases in lymph node macrophages [2].

The CpG oligonucleotides selected to be highly immunostimulatory are many-fold more potent at producing these effects than are some of the sequences selected for their antisense activity. It was in fact shown early on that by excluding CpG motifs and by replacing cytosines with 5-methyl cytosines, the proinflammatory potential of the sequence can be minimized. The CpG oligonucleotides work through Toll-like 9 receptors of innate immunity, but non-CpG oligonucleotides may work through other as yet unidentified receptors [51].

Mononuclear cell infiltrates have been occasionally observed in the high-dose groups in monkey studies; primarily at the site of subcutaneous injection, the severity and the incidence is much lower than in rodents. In rodents,

depending on the sequence and chemistry, infiltrates have been observed after doses as low as 1 mg/kg and perhaps lower for oligonucleotides designed to be immunostimulatory. The rodents are much more sensitive to these effects than primates and that sensitivity is unrelated to pharmacokinetic differences. In primate studies, tissue concentrations of PS ODNs often exceed those observed in rodents. Rodents appear to overpredict for other species, including humans, presumably.

The siRNAs also produce pro-inflammatory responses and release of cyokines like interferons through Toll-like receptor 3, 7, and or 8 [33,34,52]. More siRNA and single-stranded RNAs with 5′ phosphates interact with the retinoic acid inducible protein (RIG-1) in the cytosol and stimulate innate immunity as part of antiviral mechanisms [53,54]. For this activity there is no currently identified sequence motif; rather, it is the presence of exogenous RNA that is apparently being recognized as foreign (viral). How this stimulation of innate immunity will be modulated or exacerbated by chemical modifications remains to be seen. For now, siRNAs like the single-stranded oligonucleotide therapeutics have the potential for proinflammatory effects.

*Cardiovascular* In rodents, lymphohistiocytic infiltrates have been observed in the heart and in the perivascular space in various organs. Tissue macrophages in the heart can contain basophilic granules. In monkeys, no changes in ECG, heart rate, or blood pressure have been observed with numerous oligonucleotides. Under conditions when complement is activated, changes in blood pressure and cardiovascular collapse due to hypotension have been observed [55], but we have shown this to be related to complement activation rather than a direct effect on the cardiovascular system. (Of course, other mechanisms for hypotension could occur with other chemistries. For example, if an oligonucleotide formulation or its metabolic products chelates calcium, reductions in ionizable calcium could also produce a hypotensive crisis.) For most PS ODNs the most likely cause of hypotension is complement activation as demonstrated with complement inhibitors.

*Hematologic* Dosing mice with high doses of PS ODNs can result in slight reductions in red cell parameters, probably the result of increased destruction of cells because of the slight splenomegaly seen at higher doses. The reduction in red cell parameters does not appear to be related to synthesis as the morphology of marrow is generally unremarkable in mice. With highly immunostimulatory oligonucleotides, when treatment causes significant splenomegaly, we have often observed reductions in platelets that follow the time course of the splenomegaly, beginning with the first increases in spleen weights and reversing only when the splenic enlargement reverses. Production of platelets in these mice appears normal as there is no diminution of megakaryocyte number or size. The reduction in platelets is nevertheless often accompanied

by extramedullary hematopoiesis featuring megakaryocytes, suggesting that production is still active.

In monkeys treated with some sequences of PS ODNs at doses more than 20 mg/kg/wk, reductions in platelets have been observed. Reductions in platelets are not common to all sequences; rather, the reduction seems to be related to specific sequences and not others. This effect is unlikely to be related to pharmacology because it has been observed for compounds with diverse targets. These sequence-dependent reductions in platelets have generally been observed in high-dose monkeys, and some reductions in platelets have occasionally been observed in human subjects treated with the same sequences. In both monkeys and in humans the reduction in platelets appears to be reversible over the course of weeks or months depending on the half-life of the sequence. This reduction in platelets is distinct from transient sequestration of platelets that has been observed during constant intravenous infusions of PS ODNs: an effect probably related to the polyanionic nature and one that reverses within minutes or hours as the plasma is cleared of oligonucleotide.

*Immunologic* As discussed above, the most obvious immunologic effects in mice are the lymphohistiocytic infiltrates commonly observed in the parenchyma of numerous organs. These collections of histiocytes can sometimes be observed perivascularly and have been at times found in clusters, giving the appearance of microgranulomas. As would be expected, the increase in mononuclear cells in tissues is sometimes associated with increases in circulating monocytes, and plasma cytokine and chemokines concentrations.

The uptake of oligonucleotide drugs by the tissue macrophages causes their enlargement and gives them the appearance of being activated, though we have not determined if specific markers of activation are displayed on these engorged cells. Studies characterizing the immunotoxicity of oligonucleotides do not provide any evidence of macrophage dysfunction in mice. The engorgement of tissue macrophages occurs in monkeys as well as mice, but again, no indications of altered macrophage function have been recorded. Specific studies to address the consequences of loading macrophages with oligonucleotide need to be performed with each of the classes of oligonucleotide drugs.

Administration of oligos to rodents can induce the production of immunoglobulins in a nonspecific way [56,57]. These increases in immunoglobulins are not directed specifically toward the oligonucleotide, but rather, they represent a polyclonal expansion of B cells probably secondary to direct mitogenic effects on B cells associated with PS ODNs. In general, there is little or no evidence that oligonucleotides currently being used in clinical trials are antigenic. It is possible to reduce the proinflammatory effects of oligonucleotides by avoiding the stimulatory motifs and by chemical modifications. Modifications like the inclusion of 5-methyl cytosine and the addition of 2′ alkoxy groups can reduce the mitogenic potential of PS ODNs [58].

*Hepatic* Liver is the organ that contains the second highest oligonucleotide concentration in all species studied [11,70]. Liver is the organ that accumulates the greatest quantity of oligonucleotide. Concentrations in renal cortex may be higher, but the greater size of the liver makes it the organ that accumulates the greatest quantity of oligonucleotide for PS ODNs. PS ODNs are present in all the cell types within liver [23,24,59], but aptamers and siRNA may have more limited distribution. The presence of oligonucleotide in hepatocytes makes those cells good targets for antisense activity. Antisense activity has also been demonstrated in Kupffer cells [60,61]. Much of the toxicity for PS ODNs has already been reviewed [31]. In mice treated with PS ODNs, the mononuclear infiltrates and the associated single-cell necrosis result in slight increases in serum transaminases. These are reversible after the removal from treatment, and there is ample literature suggesting that the more potent the proinflammatory effects of a sequence the greater likelihood that there will be increases in transaminases [31,58,62,63]. The rat is somewhat less sensitive to the hepatic effects of PS ODNs than mouse. In monkey, increases in transaminases are not generally observed, and there are few hepatic effects at clinically relevant doses and even at doses 10-fold greater than those in clinical trials. This includes dose regimens for several different 2′-MOE ASO that ranged up to 140 mg/kg/wk for 5 weeks, or 80 mg/kg/wk for 13 weeks (Isis Pharmaceuticals, Inc., unpublished observations).

Consistent with the monkey data, PS ODN and 2′-MOE ASO have been largely absent of hepatotoxicity in clinical trials [47,48,64,65]. The only notable exception was the report of mild and transient increases in ALT of patients treated with anti-HIV PS ODN, GEM 91 [66]. This was also likely attributable to the proinflammatory effects that were present for this early generation of ASO comprised of 27 nucleotide residues along with CpG motiffs.

*Renal* Kidney is the organ that contains the highest oligonucleotide concentration after oligonucleotide administration, and it is a target organ for toxicity in monkeys and rats. Oligonucleotide in kidney accumulates as a result of reuptake of solutes by the kidney: a normal function of the kidney. Oligonucleotide is filtered at the glomerulus and is readily reabsorbed by the proximal tubular epithelium [67]. The oligonucleotide is then taken up into these cells by endocytosis, like other solutes [68]. The majority of oligonucleotide in proximal tubular epithelium resides in membrane-bound endosomes and lysosomes, generally oriented toward the luminal (apical) surface.

At high concentrations of oligonucleotide there are histologic changes in the proximal tubular cells [49,69]. The preferential uptake of oligonucleotide by proximal tubular epithelium and processing into endosomes and lysosomes is similar in mice, rats, and monkeys, though mice take up less oligonucleotide into kidney than other species.

In the absence of changes in cell morphology, these basophilic granules are considered not toxicologically significant. Often associated with basophilic granules are cytoplasmic vacuoles in the tubular epithelial cells [70]. The size

and incidence of these vacuolar changes can be variable depending on the fixative, processing, and other unknown variables. The vacuoles are typically single-to-multiple large clear cytoplasmic vacuoles, often with small amounts of basophilic material in them. In rare instances when the concentrations of oligonucleotide are high enough, there can be swelling and rupture of cells, but again this is not solely determined by concentration but rather by a combination of factors like fixation conditions and preservation. Vacuolation, and the mild degenerative changes, are reversible after treatment is withdrawn. The vacuoles are not associated changes in renal tubular function [70].

The vacuoles result from the localized high concentrations of hydroscopic material in phagolysosomes. Oligonucleotides are highly water soluble molecules that could easily be extracted if they are not completely cross-linked by fixative in tissues. Because of these properties it is also possible that high concentrations of oligonucleotide contained in membrane-bound subcellular compartments would be osmotically active upon fixation, particularly in slower fixation processes such as immersion in formalin. The vacuoles in the proximal tubular cells of monkeys treated with oligonucleotides differ from classical tubular vacuolar degeneration. Vacuolar degeneration is characterized by generalized organelle swelling, but this is *not* manifested in the tubules in oligonucleotide-treated monkeys.

Other changes in renal morphology associated with oligonucleotide treatment include dose- and concentration-dependent degenerative changes in the proximal tubular epithelium and regenerative changes. Degeneration is characterized as minimal reductions in the height of the brush border and height of the proximal tubular cells. Tubular cell regeneration is characterized by more active-looking nuclei. As concentrations increase, focal tubular epithelial cell degeneration, and finally, frank epithelial cell degeneration, can occur. The concentrations that produce these effects are highly dependent on the chemistry of the oligonucleotide that is accumulating in the tubule. For unmodified PS ODNs, concentrations above 2000 to 3000 μg/g of tissue can produce frank tubular cell degeneration [71]. For oligonucleotides that are modified with 2′-MOE, the concentrations required to produce these changes in morphology are nearly twice as high.

It is only at these very high renal concentrations that some functional changes begin to occur. The dose required to produce these changes are at or exceed 40 mg/kg/wk. At this dose the only functional changes noted occasionally in monkeys are slightly increased urinary protein levels [70,71]. The effect of renal accumulations of oligonucleotide accumulations in granules and vacuoles on renal function was studied in monkeys treated with a 2′-MOE ASO administered at 40 mg/kg/wk (administered on days 1, 3, 5, 7, 14, 21, and 28). This regimen produced a typical pattern of vacuoles and minimal proximal tubular degeneration after a month of treatment [70]. A battery of renal functional assessment was performed including BUN, creatinine, creatinine clearance, amino acid secretion, GFR, and inulin clearance. In addition renal functional stress tests were performed to look at the how well treated monkeys

could concentrate urine, and respond to a glucose load. The only change associated with these histologic changes at the higher doses was a low incidence of mild increases in urinary protein secretion. Protein-creatinine ratios increased from approximately 0.2 at baseline to just above the upper limit of normal, 1.2, and the increase was tubular in origin [70] (unpublished, Isis Pharmaceuticals, Inc.).

There were no other functional changes observed. No changes in mean serum creatinine or BUN and no changes in electrolytes were measured. Effective renal plasma flow rate and glomerular filtration rate were unchanged in treated monkeys, and there were no changes in excretion of substrates for tubular reabsorption, including glucose, amino acids, or β2-microglobulin [70]. Typical markers of tubular damage like *N*-acetyl glucosaminidase (NAG) were unaffected. Even with this detailed assessment of renal and tubular epithelial cell function, no tubular functional changes were induced by treatment with 40 mg/kg/wk of a second-generation oligonucleotide. There were no indications of changes in glomerular function, nor were there changes in an ultrastructural assessment of morphology.

With the information that is available from these types of studies, it is possible to define concentration–response relationships for each of the morphologic changes observed in the kidneys of monkeys. A qualitative plot of these relationships demonstrates that concentrations predicted for therapeutic doses of oligonucleotide produce renal cortex concentrations well below those associated with any significant renal morphologic findings (Figure 4). The existing data suggest that renal effects are more a function of concentration than they

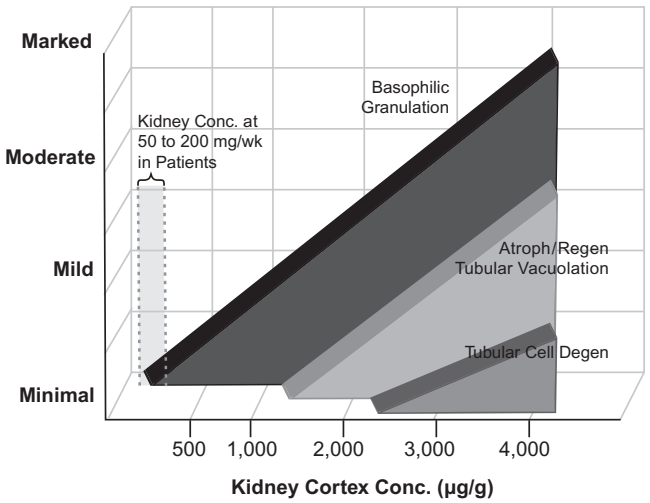


**Figure 4** Schematic representation of the relationship between renal cortex concentration and morphologic changes present in the kidneys of monkeys dosed with second-generation antisense oligonucleotides.

are of duration of exposure, and renal morphologic changes do not appear to progress in monkeys treated from one to three months. We are collecting data on longer term exposures.

The atrophic and regenerative changes, as well as the vacuolar changes, are all reversible after treatment is removed. The rate of reversal is related to the half-life of the drug, so that unmodified phosphorothioate oligonucleotides with their shorter half-lives reverse faster than the compounds modified to have prolonged half-lives. A thorough understanding of the dose–response and concentration–response relationships, markers like urinary protein-to-creatinine ratios to monitor for renal functional changes, and the reversibility of the changes are all important factors that need to be characterized in the toxicity program for any systemically administered oligonucleotide therapeutic.

### 4.6 Reproductive and Endocrine

Reproductive toxicity studies with PS ODNs have been performed with both human-specific sequences as well as species-specific sequences. Like chronic studies, this combination of human drug and surrogate provides insight into whether any effects observed are class related or target related.

Fertility, fetal development, and development and reproductive function of off-spring receiving PS ODNs have been assessed in mice, rats, and rabbits. To our knowledge, no reproductive toxicity studies have been performed in primates treated with oligonucleotides. Traditional embryonic and fetal exposure protocols have been used for these studies. In these studies we have included assessments of placental and embryonic exposure to the drugs. Even at doses that produce maternal toxicity, there have been no definitive indications of direct effects on in utero development. This is most likely the result of the absence of significant exposure to the conceptus. Placental concentrations of oligonucleotide are very low, and little or no oligonucleotide can be found in fetal tissues even when the oligonucleotide is administered by constant infusion over the course of development [72]. Fetal kidney concentrations are less than 1% of those of the maternal kidney and most fetal tissues contain oligonucleotide concentrations below the limits of detection or quantitation for capillary gel electrophoresis. The low levels of placental uptake and transfer reduce the exposure and the potential effects of oligonucleotide therapeutics in the developing embryo and fetus.

Fertility assessments performed on male rodents treated with PS ODNs have also been uniformly negative. This finding is consistent with the pharmacokinetics of this class. The highly charged and water-soluble oligonucleotide therapeutics are effectively excluded from the seminiferous tubules by the blood–testes barrier. The interstitial cells of the testes can accumulate oligonucleotide, but detailed autoradiography and immunohistochemistry suggest that there is none in the seminiferous tubules or in the developing sperm.

The female reproductive organs like the uterus and ovaries are not in privileged sites in that they will accumulate oligonucleotides after treatment, and

they are sites where mononuclear cell infiltrates can be seen in rodents treated with high doses of PS ODNs. High-dose treatment of mice with at least one sequence has produced reductions in reproductive performance, most likely secondary to local inflammation in the ovaries or the uterus. This effect on ovaries was associated with a dose-dependent decrease in the fertility index in mice at doses at or over 20 mg/kg/wk. These same proinflammatory effects have also been shown to induce premature labor and abortion in rodents treated with maternal toxic doses of PS ODNs. This phenomenon has been observed in both mice and rabbits treated with PS ODNs at maternal toxic doses. For this reason it is unlikely that this phenomenon is clinically relevant.

There are several examples where investigators have used antisense inhibitors to study the effects on reproductive function. The most notable was the use of a 2′-*O*-methyl modified oligonucleotide targeted to VEGF. In this study antisense treatment produced a 50% inhibition of VEGF mRNA and was associated with developmental abnormalities similar to that observed in the VEGF-deficient mice [73]. These effects were attributed to the inhibition of VEGF because of both the similarity of the phenotype and the lack of effect on VEGF expression or fetal development of a 5-base mismatched control. While these effects are clearly related to the targeted pharmacology, it does raise important points about the accessibility of fetus to antisense inhibitors during development. It is possible that the exposure of the vascular endothelium to the VEGF inhibitor in placenta is sensitive to antisense inhibition at least at certain points in development, or it is possible these effects were related to inhibition of maternal VEGF systemically during an important phase of development. In any case, the assessment of fetal effects as well as extra-fetal effects of novel therapeutic agents is critical, and careful assessment of oligonucleotide therapeutic agents should be performed with active agents when possible, despite the absence of significant transfer to the developing conceptus.

## 4.7 Immunotoxicity

The most obvious effect on the immune system with oligonucleotide administration is the tendency to stimulate a pro-inflammatory reaction in some species (reviewed in [31,74]). Whether this is a true immunotoxicity is one of the questions that needs to be addressed for oligonucleotide therapeutics.

***Assessing the Immunotoxicity*** Currently the assessment of the immunotoxicity of oligonucleotide therapeutics is performed using the typical tiered set of rodent assays. We have assessed both the human oligonucleotide and the surrogate "rodent-specific" sequences. Unfortunately, these batteries of tests are much more relevant to the assessment of immunosuppressive drugs and are not particularly well adapted to immunostimulatory compounds like oligonucleotides. While there have been discussions of how to monitor and assess immune-stimulating drugs, there are still no well-accepted guidelines.

***Immunotoxicity of Oligonucleotide Therapeutics*** The pro-inflammatory activity of the ASOs produces a constellation of effects, which includes splenomegaly, lymphoid hyperplasia, and multi-organ lymphohistiocytic cell infiltrate driven by the production of pro-inflammatory cytokines [57,63,75]. Incorporation of immunostimulatory sequences like unmethylated CG dinucleotides mimic bacterial DNA [76]. Unmethylated CG dinucleotides interact with receptors of the innate immune system such as TLR9 to produce a shift toward T-helper type 1 immunity [77,78]. Oligonucleotides with these motifs modulate the immune system. We have recently reported on a direct evaluation for effects on immune function.

In these studies ISIS 2302 (a PS ODN inhibitor of human ICAM-1 with no CG motifs) was administered to normal mice, and the effects of the drug on both the humoral and cellular immune functions were assessed by a battery of commonly used tests. As expected, treatment with high doses produced an increase in splenic weight that was accounted by increased splenocyte numbers, including increases in B cell numbers, which appeared to cause a shift in the population leading to decreased relative numbers of T cells but no decrease in absolute T cell numbers. There were no meaningful changes in antibody response in a sheep red blood cell assay, and the response to mitogens was not diminished. Consistent with an increase in total B cells, previously published information on this class of compound has documented an increase in total IgG and/or IgM [79,80]. No antibodies specific to dsDNA or ISIS 2302 were detected in sera. Oligonucleotides have been shown to elicit their proinflammatory effects directly on cells of the innate immune system producing a polyclonal increase in B cells, and therefore specific antibody responses are absent [32,74]. This is a key finding for the technology. Polyclonal expansion including IgG and IgM secretion is expected in rodents (and perhaps other species), and this occurs in the absence of antibodies directed toward the oligonucleotide and with no greater specificity to DNA than would be expected on the basis of chance.

Cellular responses, as measured by a mixed-lymphocyte reaction, cytotoxic T lymphocyte response, and NK cell activity, were all undiminished, and if anything, there was a slight increase in CTL and NK responses. As would be expected by the histologic profile and the known increases in cytokine and chemokine production associated with the administration of PS ODNs in rodents, in this series of experiments there was no diminution in immune response. Administering a mouse-specific ICAM-1 inhibitor produced reductions in mixed lymphocyte reactions. This inhibition was expected as this is one of the desired pharmacologic effects of reducing ICAM-1 expression.

***Relevance of Immunotoxicity Findings to Humans*** Monkeys, unlike rodents, do not display the same constellation of effects associated with the proinflammatory effects of PS ODNs. Human responses are thought to be more like the monkey. The human response differs from the monkey response in a few ways. First, at high doses of PS ODNs (>5 mg/kg) some subjects

experience fever chills and rigors, that are readily reversible and treatable with nonsteroidal anti-inflammatory drugs. This particular toxicity has been dose limiting in some studies. These same doses have been associated with elevations in cytokine levels [65]. This febrile response has not been recorded in monkeys, though some increases in cytokines have been observed in monkeys. The second difference is that in humans, there is a characteristic local response to subcutaneous injections. This response is a relatively mild local erythematous area typically with blanching in the central region. After a subcutaneous injection of over 100 mg in one milliliter, these reactions are approximately 2 to 3 cm in diameter, are reddened but not painful, not pruritic, and not raised. They generally resolve in three to seven days. Slight erythema has been described in monkeys receiving subcutaneous injections; these reactions differ from those seen in human subjects in that they do not feature the central blanching and annular appearance.

### 4.8 Carcinogenicity

Oligonucleotides are not genotoxic in traditional in vitro and in vivo assays, and as such, they are not thought to be classical carcinogens. No reports of carcinogenesis with oligonucleotides are in the literature. Lifetime exposure of rodents to PS ODNs, with their propensity to cause mitogenesis and proliferation of stroma and mononuclear cells, is expected to have some potential for inducing hyperplastic or even neoplastic changes. Neoplasias have been observed in one study with a PS ODN administered by subcutaneous injections on alternate days for two years to mice (Isis, unpublished observations). The systemic pro-inflammatory effects combined with the local irritation produced by injections resulted in neoplastic transformation of some of the infiltrating cells. The histiocytic sarcomas that were observed were dose related. A less immunostimulatory PS ODN was administered in parallel, and it did not produce that same neoplastic change. If the treatment was intermittent, allowing for the injection sites to heal between dosing cycles, there also was no neoplastic change. These later results clearly support the PS ODN treatment having promoter-like effects, *not* initiator effects. If the neoplastic changes are associated with repeated administration to highly inflamed sites and local and systemic mononuclear infiltrates, then it is likely that this positive bioassay for a PS ODN does not have clear clinical relevance.

## 5 SUMMARY

The toxicology of oligonucleotides was at one time thought to be a real limitation to the technologies. To date, careful assessment of toxicity in the types of studies described here have demonstrated that it is likely that there may be attractive therapeutic indexes, particularly with the second-generation antisense oligonucleotides. The other classes of oligonucleotide therapeutics are

also moving forward, and all of them will have distinctive hurdles to overcome. With strong scientific commitment to the assessment of toxicity and understanding mechanisms of toxicity, it is likely that these classes with continue to advance through preclinical studies and clinical trials as well.

## ACKNOWLEDGMENTS

The authors would like to acknowledge the editorial comments and discussions with Dr. Christopher Horvath of Archemix. In addition they would like to acknowledge Lori Cooper for her help in the preparation of this manuscript and Robert Saunders for his expertise with EndNote.

## REFERENCES

1. Levin AA, Yu RZ, Geary RS. Basic PrInciples of the PharmacokInetics Of antisense Oligonucleotide drugs. In Crooke ST, ed. *Antisense Drug Technology*. New York: Dekker, 2007.
2. Iversen PL. Morpholinos. In Crooke ST, ed. *Antisense Drug Technology*. New York: Dekker, 2007.
3. Manoharan M. Harnessing RNA interference pathways for therapeutics utilizing chemistry: chemically modified siRNAs, drug delivery systems and antagomirs. In Crooke ST, ed. *Antisense Drug Technology*. New York: Dekker, in press.
4. Cossum PA, Sasmor H, Dellinger D, Truong L, Cummins L, Owens SR, Markham PM, Shea JP, Crooke ST. Disposition of the 14C-labeled phosphorothioate oligonucleotide ISIS 2105 after intravenous administration to rats. *J Pharmacol Exp Therapeut* 1993;267(3):1181–90.
5. Cossum PA, Troung L, Owens SR, Markham PM, Shea JP, Crooke ST. Pharmacokinetics of a $^{14}$C-labeled phosphorothioate oligonucleotide, ISIS 2105, after intradermal administration to rats. *J Pharmacol Exp Therapeut* 1994;269(1):89–94.
6. Yu RZ, Baer B, Chappel A, Geary RS, Chueng E, Levin AA. Development of an ultrasensitive noncompetitive hybridization-ligation enzyme-linked immunosorbent assay for the determination of phosphorothioate oligodeoxynucleotide in plasma. *Anal Biochem* 2002;304(1):19–25.
7. Gaus HJ, Owens SR, Winniman M, Cooper S, Cummins LL. On-line HPLC electrospray mass spectrometry of phosphorothioate oligonucleotide metabolites. *Anal Chem* 1997;69(3):313–9.
8. Griffey RH, Greig MJ, Gaus HJ, Liu K, Monteith D, Winniman M, Cummins LL. Characterization of oligonucleotide metabolism in vivo via liquid chromatography/electrospray tandem mass spectrometry with a quadrople ion trap mass spectrometer. *J Mass Spec* 1997;32:305–13.
9. Murphy AT, Brown-Augsburger P, Yu RZ, Geary RS, Thibodeaux S, Ackermann BL. Development of an ion-pair reverse-phase liquid chromatographic/tandem mass spectrometry method for the determination of an 18-mer phosphorothioate oligonucleotide in mouse liver tissue. *Eur J Mass Spec* 2005;11(2):209–15.

10. Yu RZ, Geary RS, Levin AA. Application of novel quantitative bioanalytical methods for pharmacokinetic and pharmacokinetic/pharmacodynamic assesments of antisense oligonucleoutides. *Curr Opin Drug Discov Dev* 2004;7(2):195–203.
11. Yu RZ, Geary RS, Levin AA. Pharmacokinetics and pharmacodynamics of antisense oligonucleotides. In Meyers RA, ed. *Encyclopedia of Molecular Cell Biology and Molecular Medicine*, Weinheim, Germany: Wiley-VCH, 2007.
12. Boomer RM, Lewis SD, Healy JM, Kurz M, Wilson C, McCauley TG. Conjugation to polyethylene glycol polymer promotes aptamer biodistribution to healthy and inflamed tissues. *Oligonucleotides* 2005;15(3):183–95.
13. Healy JM, Lewis SD, Kurz M, Boomer RM, Thompson KM, Wilson C, McCauley TG. Pharmacokinetics and biodistribution of novel aptamer compositions. *Pharmaceut Res* 2004;21(12):2234–46.
14. Watanabe TA, Geary RS, Levin AA. Plasma protein binding of an antisense oligonucleotide targeting human icam-1 (ISIS 2302). *Oligonucleotides* 2006;16(2): 169–80.
15. Geary RS, Yu RZ, Siwkowski A. Pharmacokinetic/pharmacodynamic properties of phosphorothioate 2′-*O*-(2-methoxyethyl) modified antisense oligonucleotides in animals and man. In Crooke ST, ed. *Antisense Drug Technology*. New York: Dekker, 2007.
16. Geary RS, Yu RZ, Leeds JM, Ushiro-Watanabe T, Henry SP, Levin AA, Templin MV. Pharmacokinetic properties in animals. In Crooke ST, ed. *Antisense Drug Technology: Principles, Strategies, and Applications*. New York: Dekker, 2001; 119–54.
17. Levin AA, Geary RS, Leeds JM, Monteith DK, Yu RZ, Templin MV, Henry SP. The pharmacokinetics and toxicity of phosphorothioate oligonucleotides. In Thomas JA, ed. *Biotechnology and Safety Assessment*. Philadelphia: Taylor and Francis, 1998;151–75.
18. Phillips JA, Craig SJ, Bayley D, Christian RA, Geary RS, Nicklin PL. Pharmacokinetics, metabolism and elimination of a 20-mer phosphorothioate oligodeoxynucleotide (CGP 69846a) after intravenous and subcutaneous administration. *Biochem Pharmacol* 1997;54(6):657–68.
19. Geary RS, Ushiro-Watanabe T, Truong L, Freier SM, Lesnik EA, Sioufi NB, Sasmor H, Manoharan M, Levin AA. Pharmacokinetic properties of 2′-*O*-(2-methoxyethyl)-modified oligonucleotide analogs in rats. *J Pharmacol Exp Therapeut* 2001;296(3):890–7.
20. Geary RS, Yu RZ, Levin AA. Pharmacokinetics of phosphorothioate antisense oligodeoxynucleotides. *Curr Opin Invest Drugs* 2001;2(4):562–73.
21. Geary RS, Holmlund J, Kwoh TJ, Yu RZ, Dorr FA. Dosing of antisense oligodeoxynucleotides (ODNS) by ideal body weight, rather than actual body weight, may be preferred. Paper presented at Proceedings of the 11th NCI-EORTC-AACR Symposium, 2000.
22. CDER. Guidance for industry—estimating the maximum safe starting dose in initial clinical trials for therapeutics in adult healthy volunteers. In Center for Drug Evaluation and Research (U.S. Department of Health and Human Services), 2005; 1–27.
23. Butler M, Stecker K, Bennett CF. Cellular distribution of phosphorothioate oligodeoxynucleotides in normal rodent tissues. *Lab Invest* 1997;77(4):379–88.

24. Butler M, Stecker K, Bennett CF. Histological localization of phosphorothioate oligodeoxynucleotides in normal rodent tissue. *Nucleosides Nucleotides* 1997; 16(7–9):1761–4.
25. Devi GR, Beer TM, Corless CL, Arora V, Weller DL, Iversen PL. In vivo bioavailability and pharmacokinetics of a c-myc antisense phosphorodiamidate morpholino oligomer, AVI-4126, in solid tumors. *Clin Cancer Res* 2005;11(10):3930–8.
26. Braasch DA, Paroo Z, Constantinescu A, Ren G, Oz OK, Mason RP, Corey DR. Biodistribution of phosphodiester and phosphorothioate siRNA. *Bioorg Med Chem Letts* 2004;14(5):1139–43.
27. Capaldi DC, Scozzari AN. Manufacturing and analytical processes. In Crooke ST, ed. *Antisense Drug Technology*. New York: Dekker, 2007.
28. Jackson AL, Burchard J, Schelter J, Chau BN, Cleary M, Lim L, Linsley PS. Widespread siRNA "off-target" transcript silencing mediated by seed region sequence complementarity. *RNA* 2006;12:1179–87.
29. Verdel A, Jia S, Gerber S, Sugiyama T, Gygi S, Grewal SI, Moazed D. RNAI-mediated targeting of heterochromatin by the RITS complex. *Science* 2004; 303(5658):672–6.
30. Levin AA, Monteith DK, Leeds JM, Nicklin PL, Geary RS, Butler M, Templin MV, Henry SP. Toxicity of oligodeoxynucleotide therapeutic agents. In Crooke ST, ed. *Antisense Research and Application*. Heidelberg: Springer, 1998;169–215.
31. Levin AA, Henry SP, Monteith D, Templin M. Toxicity of antisense oligonucleotides. In Crooke ST, ed. *Antisense Drug Technology*. New York: Dekker, 2001; 201–67.
32. Krieg AM. CpG motifs in bacterial DNA and their immune effects. *Ann Rev Immunol* 2002;20:709–60.
33. Schlee M, Hornung V, Hartmann G. SiRNA and isRNA: two edges of one sword. *Mol Ther* 2006;14(4):463–70.
34. Kariko K, Bhuyan P, Capodici J, Weissman D. Small interfering RNAs mediate sequence-independent gene suppression and induce immune activation by signaling through Toll-like receptor 3. *J Immunol* 2004;172(11):6545–9.
35. Black LE, DeGeorge JJ, Cavagnaro JA, Jordan A, Ahn C-H. Regulatory considerations for evaluating the pharmacology and toxicology of antisense drugs. *Antisense Res Dev* 1993;3:399–404.
36. Black LE, Farrelly JG, Cavagnaro JA, Ahn C-H, DeGeorge JJ, Taylor AS, DeFelice AF, Jordan A. Regulatory considerations of oligonucleotide drugs: updated recommendations for pharmacology and toxicology studies. *Antisense Res Dev* 1994;4:299–301.
37. Henry SP, Monteith DK, Matson JE, Mathison BH, Loveday KS, Winegar RA, Lee PS, Riccio ES, Bakke JP, Levin AA. Assessment of the genotoxic potential of ISIS 2302: a phosphorothioate oligodeoxynucleotide. *Mutagenesis* 2002;17(3):201–9.
38. EMEA. CHNP SWP Reflection Paper on the Assesment of the Genotoxic Potential of Antisense Oligodeoxynucleotides. EMEA/CHNP/SWP/199726/2004 (2005).
39. Wang G, Seidman MM, Glazer PM. Mutagenesis in mammalian cells induced by triple helix formation and transcription-coupled repair. *Science* 1996;271(5250): 802–5.

40. Henry SP, Giclas PC, Leeds J, Pangburn M, Auletta C, Levin AA, Kornbrust DJ. Activation of the alternative pathway of complement by a phosphorothioate oligonucleotide: potential mechanism of action. *J Pharmacol Exp Therapeut* 1997; 281(2):810–6.

41. Henry SP, Jagels M, Hugli T, Giclas P. Species and pathway-specific activation of complement by a phosphorothioate oligonucleotide. Paper presented at Immunopharmacology.

42. Henry SP, Beattie G, Yeh G, Chappel A, Giclas PC, Mortari A, Jagels MA, Kornbrust DJ, Levin AA. Complement activation is responsible for acute toxicities in rhesus monkeys treated with a phosphorothioate oligodeoxynucleotide. *Intl J Immunopharmacol* 2002;2(12):1657–66.

43. Henry SP, Kim T-W, Kramer-Stickland K, Zanardi TA, Fey RA. Toxicologic properties of 2′-methoxyethyl chimeric antisense inhibitors in animals and man. In Crooke ST, ed. *Antisense Drug Technology*. New York: Dekker, in press.

44. Henry SP, Novotny W, Leeds J, Auletta C, Kornbrust DJ. Inhibition of coagulation by a phosphorothioate oligonucleotide. *Antisense Nucl Acid Drug Dev* 1997;7: 503–10.

45. Sheehan JP, Lan H-C. Phosphorothioate oligonucleotides inhibit the intrinsic tenase complex. *Blood* 1998;92(5):1617–25.

46. Sheehan JP, Thao PM. Phosphorothioate oligonucleotides inhibit the intrinsic tenase complex by an allosteric mechanism. *Biochemistry* 2001;40(16):4980–9.

47. Leeds JM, Mant TG, Amin D, Kisner DL, Zuckerman JE, Geary RS, Levin AA, Shanahan WR, Jr., Glover JM. Phase I safety and pharmacokinetic profile of an intercellular adhesion molecule-1 antisense oligodeoxynucleotide (ISIS 2302). *J Pharmacol Exp Therapeut* 1997;282(3):1173–80.

48. Sewell LK, Geary RS, Baker BF, Glover JM, Mant TGK, Yu RZ, Tami JA, Dorr AF. Phase i trial of isis 104838, a 2′-methoxyethyl modified antisense oligonucleotide targeting tumor necrosis factor-alpha. *J Pharmacol Exp Therapeut* 2002; 303(3):1334–43.

49. Henry SP, Bolte H, Auletta C, Kornbrust DJ. Evaluation of the toxicity of isis 2302, a phosphorothioate oligonucleotide, in a 4-week study in cynomolgus monkeys. *Toxicology* 1997;120:145–55.

50. Yu RZ, Geary RS, Leeds JM, Ushiro-Watanabe T, Fitchett JR, Matson JE, Mehta R, Hardee GR, Templin MV, Huang K, et al. Pharmacokinetics and tissue disposition in monkeys of an antisense oligonucleotide inhibitor of ha-ras encapsulated in stealth liposomes. *Pharmaceut Res* 1999;16(8):1309–15.

51. Senn JJ, Burel S, Henry SP. Non-CpG containing antisense 2′ MOE oligonucleotides activate a proinflammatory response independent of TLR-9 or myd88. *J Pharmacol Exp Therapeut* 2005;314:972–9.

52. Reynolds A, Anderson EM, Vermeulen A, Fedorov Y, Robinson K, Leake D, Karpilow J, Marshall WS, Khvorova A. Induction of the interferon response by siRNA is cell type- and duplex length-dependent. *RNA* 2006:12(6):988–93.

53. Hornung V, Ellegast J, Kim S, Brzozka K, Jung A, Kato H, Poeck H, Akira S, Conzelmann KK, Schlee M, et al. 5′-triphosphate RNA is the ligand for RIG-1. *Science* 2006;314:935–6.

54. Pichlmair A, Schulz O, Tan CP, Naslund TI, Liljestrom P, Weber F, Reis e Sousa C. RIG-1-mediated antiviral responses to single-stranded RNA bearing 5′-phosphates. *Science* 2006;314(5801):997–1001.
55. Henry SP, Geary RS, Yu R, Levin AA. Drug properties of second-generation antisense oligonucleotides: How do they measure up to their predecessors? *Curr Opin Invest Drugs* 2001;2(10):1444–9.
56. Liang H, Nishioka Y, Reich CF, Pisetsky DS, Lipsky PE. Activation of human B cells by phosphorothioate oligodeoxynucleotides. *J Clin Invest* 1996;98(5): 1119–29.
57. Klinman DM, AE-Kyung Y, Beaucage SL, Conover J, Krieg AM. CpG motifs present in bacterial DNA rapidly induce lymphocytes to secrete interleukin 6, interleukin 12, and interferon γ. *Proc Nat Acad Sci USA* 1996;93:2879–83.
58. Henry SP, Stecker K, Brooks D, Monteith D, Conklin B, Bennett CF. Chemically modified oligonucleotides exhibit decreased immune stimulation in mice. *J Pharmacol Exp Therapeut* 2000;292(2):468–79.
59. Graham MJ, Crooke ST, Monteith DK, Cooper SR, Lemonidis KM, Stecker KK, Martin MJ, Crooke RM. In vivo distribution and metabolism of a phosphorothioate oligonucleotide within rat liver after intravenous administration. *J Pharmacol Exp Therapeut* 1998;286(1):447–58.
60. Ponnappa BC, Israel Y. Targeting kupffer cells with antisense oligonucleotides. *Frontiers Biosci* 2002;7:e223–33.
61. Ponnappa BC, Israel Y, Aini M, Zhou F, Russ R, Cao QN, Hu Y, Rubin R. Inhibition of tumor necrosis factor alpha secretion and prevention of liver injury in ethanol-fed rats by antisense oligonucleotides. *Biochem Pharmacol* 2005;69(4): 569–77.
62. Henry SP, Zuckerman JE, Rojko J, Hall WC, Harman RJ, Kitchen D, Crooke ST. Toxicological properties of several novel oligonucleotide analogs in mice. *Anticancer Drug Des* 1997;12:1–14.
63. Zhao Q, Temsamani J, Zhou R-Z, Agrawal S. Pattern and kinetics of cytokine production following administration of phosphorothioate oligonucleotides in mice. *Antisense Nucl Acid Drug Dev* 1997;7:495–502.
64. Kastelein JJP, Wedel MK, Baker BF, Su J, Bradley JD, Yu RZ, Chuang E, Graham MJ, Crooke RM. Potent reduction of apolipoprotein B and low-density lipoprotein cholesterol by short-term administration of an antisense inhibitor of apolipoprotein B. *Circulation* 2006;114:1729–35.
65. Kwoh JT. An overview of the clinical safety experience of first and second generation antisense oligonucleotides. In Crooke ST, ed. *Antisense Drug Technology*. New York: Dekker, 2007.
66. Schechter PJ, Martin RR. Safety and tolerance of phosphorothioates in humans. In Crooke ST, ed. *Antisense Research and Applications*. Berlin, Heidelberg: Springer-Verlag, 1998;231–41.
67. Oberbauer R, Schreiner GF, Meyer TW. Renal uptake of an 18-mer phosphorothioate oligonucleotide. *Kidney Int* 1995;48:1226–32.
68. Rappaport J, Hanss B, Kopp JB, Copeland TD, Bruggeman LA, Coffman TM, Klotman PE. Transport of phosphorothioate oligonucleotides in kidney: implications for molecular therapy. *Kidney Int* 1995;47:1462–9.

69. Monteith DK, Levin AA. Synthetic oligonucleotides: The development of antisense therapeutics. *Toxicol Pathol* 1999;27(1):8–13.
70. Henry SP, Johnson M, Zanardi TA, Fey R, Auyeung D, Lappin PB, LevIn AA. Renal uptake and tolerability of a 2′-*O*-methoxyethyl modified antisense oligonucleotide (ISIS 113715) in monkey. *Toxicol Sci* (in press).
71. Monteith DK, Horner MJ, Gillett NA, Butler M, Geary RS, Burckin T, Ushiro-Watanabe T, Levin AA. Evaluation of the renal effects of an antisense phosphorothioate oligodeoxynucleotide in monkeys. *Toxicol Pathol* 1999;27(3):307–17.
72. Soucy NV, Riley JP, Templin MV, Geary R, de Peyster A, Levin AA. Maternal and fetal distribution of a phosphorothioate oligonucleotide in rats after intravenous infusion. *Birth Defects Res B Dev Reprod Toxicol* 2006;77(1):22–8.
73. Driver SE, Robinson GS, Flanagan J, Shen W, Smith LEH, Thomas DW, Roberts PC. Oligonucleotide-based inhibition of embryonic gene expression. *Nat Biotechnol* 1999;17:1184–7.
74. Krieg AM. The role of CpG motifs in innate immunity. *Curr Opin Immunol* 2000; 12(1):35–43.
75. Sparwasser T, Hultner L, Koch ES, Luz A, Lipford GB, Wagner H. Immunostimulatory CpG-oligodeoxynucleotides cause extramedullary murine hemopoiesis. *J Immunol* 1999;162:2368–74.
76. Klinman DM. Immunotherapeutic uses of CpG oligodeoxynucleotides. *Nat Rev Immunol* 2004;4(4):249–8.
77. Takeshita F, Leifer CA, Gursel I, Ishii KJ, Takeshita S, Gursel M, Klinman DM. Cutting edge: role of Toll-like receptor 9 in CpG DNA-induced activation of human cells. *J Immunol* 2001;167:3555–8.
78. Krieg AM. Antitumor applications of stimulating Toll-like receptor 9 with CpG oligodeoxynucleotides. *Curr Oncol Rep* 2004;6(2):88–95.
79. Branda RF, Moore AL, Hong R, Mccormack JJ, Zon G, Cunningham-Rundles C. B-cell proliferation and differentiation in common variable immunodeficiency patients produced by an antisense oligomer to the rev gene of hiv-1. *Clin Immunol Immunopathol* 1996;79(2):115–21.
80. Branda RF, Moore AL, Lafayette AR, Mathews L, Hong R, Zon G, Brown T, McCormack JJ. Amplification of antibody production by phosphorothioate oligodeoxynucleotides. *J Lab Clin Med* 1996;128(3):329–38.

# Preclinical Safety Evaluation of Biological Oncology Drugs

THERESA REYNOLDS, BA, DABT

**Contents**

1 Background
2 Therapeutic Targets and Molecular Classes
3 Considerations for Preclinical Safety Assessment of Oncology Products
3.1 Clinical Application
3.2 Drug Attributes
3.3 Material Production
4 Application of Standard Toxicology Models
5 Summary and Conclusion
References

## 1 BACKGROUND

The development of biopharmaceuticals in oncology began in 1986 with the introduction of IFN-α-2a (Roferon A) and IFN-α-2b (Intron A) for the treatment of hairy cell leukemia, a subtype of chronic lymphoid leukemia that accounts for approximately 2% of all leukemia cases in the United States [1]. These initial marketing approvals were followed by labels for AIDS-related Kaposi's sarcoma (1988), and hepatitis C (1995). The introduction of protein-derived therapeutics to small but significant areas of unmet medical need followed by expansion to new indications has evolved in parallel with our understanding of underlying biological mechanisms of disease. Since the introduction of the interferons to the practice of oncology, the pace of marketing approvals for protein therapeutics has continued to accelerate, and new products for cancer treatment have kept pace (Figure 1). In the 20 years since

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

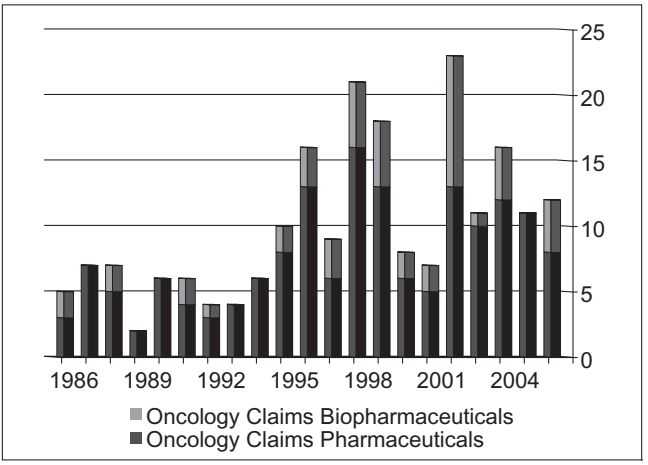


**Figure 1** US FDA approvals in oncology from 1986 to 2006 (2006 data reported through July 24, 2006).

the introduction of IFN-α, a number of new and novel therapies have emerged from the biotechnology sector that have made a significant impact on disease outcomes and patient care. The proven benefit and future promise of oncology treatments derived from biotechnology is reflected in the proliferation of new biopharmaceuticals that are under study in clinical and preclinical settings.

## 2 THERAPEUTIC TARGETS AND MOLECULAR CLASSES

As of 2006 the US FDA has granted marketing approval to 22 biotechnology-derived pharmaceuticals with varying applications to oncology practice (Table 1). Some common mechanistic themes are as follows:

- Expansion of specific cell populations to replace those lost to cytotoxic chemotherapy (erythropoietin, G-CSF, GM-CSF, IL-11), or to enhance immunocompetence (IL-2, IFN-α).
- Targeted destruction of malignant cells via specific receptor binding by antibodies (anti-HER2, anti-EGF, anti-CD20, anti-CD52).
- Use of specific receptors to target delivery of a toxin to malignant cells (immunoconjugates, fusion proteins, antibody drug conjugates).
- Alteration of the local microenvironment to starve cells of essential nutrients (anti-VEGF, L-asparaginase enzymes).

Biotechnology has made significant contributions to supportive care in oncology, developing recombinant growth factors to replenish cell populations vulnerable to the dose-limiting neutropenia, anemia, and thrombocytopenia

**TABLE 1 Biopharmaceuticals approved by US FDA for use in oncology from 1986 to 2006**

| Generic Name | Trade Name | Description | Type | Approval Date |
|---|---|---|---|---|
| Aldesleukin | Proleukin | rh IL-2 | Growth factor | 5/5/1992 |
| Alemtuzumab | Campath | hu Anti-CD52 | MAb | 5/7/2001 |
| Asparaginase | Elspar | rh L-asparaginase | Enzyme | 8/1/2002 |
| Bevacizumab | Avastin | hu Anti-VEGF | MAb | 2/26/2004 |
| Darbepoetin alpha | Aranesp | rh Novel erythropoetic stimulating protein | Growth factor | 9/17/2001 |
| Denileukin diftitox | Ontak | r Diphtheria toxin + IL-2 | Fusion protein | 2/5/1999 |
| Epoetin alpha | Epogen | rh Erythropoetin | Growth factor | 7/26/1999 |
| Cetuximab | Erbitux | Chimeric (hu/mu) anti-epithelial growth factor receptor | MAb | 2/12/2004 |
| Filgrastim | Neupogen | Recombinant human G-CSF | Growth factor | 2/20/1991 |
| Gemtuzumab Ozogamicin | Mylotarg | Humanized anti-CD33 + calicheamicin | ADC | 5/17/2000 |
| IFN-α-2a | Roferon A | IFNα-2a | Cytokine | 6/4/1986 |
| IFN-α-2b | Intron A | IFNα-2b | Cytokine | 6/4/1986 |
| Ibritumomab tiuxetan | Zevalin | mu Anti-CD20 + Y-90 & In-111 | Immunoconjugate | 2/19/2002 |
| Palifermin | Kepavance | rh Keratinocyte growth factor | Growth factor | 12/15/2004 |
| Oprelvekin | Neumega | rh IL-11 | Growth factor | 11/25/1997 |
| Pegaspargase | Oncaspar | Pegylated r L-asparaginase | Enzyme | 2/1/1994 |
| Pegfilgrastim | Neulasta | Pegylated rh G-CSF | Growth factor | 1/31/2002 |
| Rasburicase | Elitek | r Urate oxidase | Enzyme | 7/12/2002 |
| Rituximab | Rituxan | Chimeric (hu/mu) anti-CD20 | Chimeric Mab | 11/26/1997 |
| Sargramostim | Leukine, Prokine | rh GM-CSF | Growth factor | 11/7/1996 |
| Tositumomab | Bexxar | Chimeric (hu/mu) anti-CD20 + I-131 | Immunoconjugate | 6/27/2003 |
| Trastuzumab | Herceptin | hu Anti-HER2 | MAb | 9/25/1998 |

*Source*: See US FDA [3]. 2006 data reported through July 24, 2006.

*Note*: ADC: antibody drug conjugatehu; hu: humanized; IFNα: interferon alpha; MAb: monoclonal antibody; Mu: murine; r: recombinant; Rh: recombinant human.

associated with cytotoxic chemotherapy, and an enzyme (rasburicase) to deplete plasma uric acid secondary to therapeutic-induced tumor lysis.

Although biotechnology-derived imaging agents and vaccines are part of the oncology armamentarium, they will not be discussed in detail in this chapter as they have less in common with the molecular entities listed above. The imaging agents are single use and are not designed for therapeutic benefit, while vaccines have their own specialized regulatory path (see Chapter Preclinical Safety Evaluation of Viral Vaccines).

## 3 CONSIDERATIONS FOR PRECLINICAL SAFETY ASSESSMENT OF ONCOLOGY PRODUCTS

Although the molecular structures and therapeutic indications vary, there is a consistent approach to characterization of the preclinical safety profile that requires consideration of a number of factors:

- *Clinical application*: Duration of treatment, current standard of care, and patient population.
- *Drug attributes*: Mechanism of action, species cross-reactivity, availability of pharmacodynamic markers for evidence of activity, long drug half-life, potential for drug accumulation, and immunogenic potential.
- *Material production*: Evolution of production methods and appropriate bridging studies.

These assessments are discussed broadly in other chapters. This section will address their application as it pertains to preclinical safety for oncology indications.

### 3.1 Clinical Application

Phase 1 studies with biopharmaceuticals for oncology indications have typically been conducted in cancer patients rather than in healthy volunteers, due in large part to product specific concerns including the potential for an anti-drug antibody response to treatment that could preclude future therapy, or expression of a tumor-specific antigen that is not present in nontumor-bearing individuals thus potentially impacting pharmacologic activity and/or pharmacokinetic parameters. With the early enrollment of patients in the course of clinical development comes the hope of an earlier indication of therapeutic benefit, resulting in multiple-dose phase 1 trials designed to assess disease activity in terms of progression, stable disease, or patient benefit. Phase 1 clinical trials of three-months' exposure (with an option to extend treatment in patients who benefit) are increasingly common and assist in recruiting patients to experimental therapies that may alter the course of their disease. However,

because patients cannot be subjected to an experimental therapy when an established standard of care (SOC) with proven efficacy exists, patients enrolled into phase 1 programs are usually those whose disease has not been responsive to standard therapy. While phase 1 studies are generally single-agent studies of the experimental therapeutic, pivotal clinical trials to support product approval must demonstrate benefit over the existing standard of care. As a result pivotal clinical studies typically compare patients treated according to SOC with those given SOC plus experimental drug.

The earliest impact of the intended clinical application on the preclinical safety program is on the duration of the IND-enabling multiple-dose toxicity study. As is true with other therapeutic indications, the duration of the supporting preclinical study must be equal to the duration of the planned clinical study. Because oncology trials allow for extended (and often indefinite) treatment of patients who benefit, the key component to the determination of the duration of the preclinical safety study is *the point at which the clinical protocol specifies that determination of patient benefit will be made in order to extend treatment*. The preclinical study must support the "protocol-specified" duration of treatment in the clinical protocol. In the case of a 3-month treatment period for phase 1, a preclinical safety study with a three-month treatment period would be advisable. Extension of treatment for patients with responsive disease can occur following negotiation with regulatory authorities, and this is a familiar attribute of clinical trials in oncology. At a March 2006 Oncology Drugs Advisory Committee Meeting, the FDA acknowledged the value of expediting development of novel oncology products, and noted that most sponsors start out with 1-month toxicology studies in support of phase 1 trials while seeking regulatory authorization for patient treatment in excess of the treatment duration supported by the toxicology program. In the absence of 3-month animal safety data, the FDA requested sponsors to limit patient treatment, and requested 3-month toxicology studies prior to allowing patient treatment for extended durations. Continuation of clinical dosing beyond the nonclinical testing period has been approved based on the acceptability of toxicities (e.g., reversibility, degree of potential harm) and the ability to monitor anticipated toxicity endpoints. Furthermore potential patient risk–benefit must be weighed against the toxicologic profile characterized in animal studies when sponsors seek regulatory authorization for indefinite extension of patient treatment. [4,5].

While phase 1 trials often include patients of different treatment backgrounds and, depending on the mechanism of action of the experimental therapeutic, may include many different tumor types (e.g., an "all-comers" scenario), pivotal clinical trials are more tightly controlled in order to maximize the opportunity to detect an efficacy signal. As mentioned previously, the earliest pivotal clinical trials compare SOC alone and SOC in combination with experimental therapy to determine whether the experimental therapeutic provides additional patient benefit. Once this incremental benefit is demonstrated, additional clinical trials may be conducted to determine whether the experimental therapy is beneficial in earlier stage disease (e.g., the adjuvant

**TABLE 2 Single-dose pharmacokinetic/safety interaction study for Herceptin®**

| Group | Test Materials[a] | Dose (mg/kg) |
|---|---|---|
| 1 | Trastuzumab | 1.5 |
| 2 | Paclitaxel | 4.0 |
| 3 | Doxorubicin | 1.5 |
| 4 | Doxorubicin + cyclophosphamide | 1.5 + 15 |
| 5 | Trastuzumab + paclitaxel | 1.5 + 4.0 |
| 6 | Trastuzumab + doxorubicin | 1.5 + 1.5 |
| 7 | Trastuzumab + doxorubicin + cyclophosphamide | 1.5 + 1.5 + 15 |

*Source*: See [6].

[a]Multiple test materials (Groups 4-7) were administered in the order shown. Group size: Three female monkeys/group.

setting). Assessment of the preclinical safety of combination therapy requires that likely mechanisms of interaction, (e.g., biological or pharmacokinetic data) be taken into consideration.

Whereas small molecules employ p450 activation/inhibition profiles to elucidate metabolic interactions/antagonisms, there is no comparable tool to model the potential interactions of multiple biologics or between biologics and cytotoxics. To assess potential interactions between a protein therapeutic and cytotoxic chemotherapy SOC, one approach is to conduct a single-dose PK interaction study to assess the impact of the protein therapeutic on the PK of the SOC cytotoxics in a pharmacologically relevant species (Table 2; [7,8]). The objective of these studies is to determine whether the protein therapeutic has an effect on the PK, particularly the peak concentration ($C_{max}$) and/or time of peak concentration ($T_{max}$) that might result in potentiation of the toxicity of the cytotoxic, or reduced efficacy associated with decreased exposure.

Cause for concern related to additive or synergistic mechanisms of activity or toxicity should be addressed in a scientifically appropriate manner in consultation with regulatory authorities as necessary. Because of the broad distribution of potential combinations and mechanisms of activity/toxicity, and the complexity of species-specificity, there is not a one size fits all approach that can be applied. Current FDA regulatory guidance in this area is limited to combination products, with the intent to co-package and/or co-market as a defined treatment [9].

### 3.2 Drug Attributes

The impact of the drug attributes of biopharmaceuticals on preclinical safety assessment programs for oncology products is not unique to this therapeutic area. As is true for all biopharmaceuticals, species cross-reactivity must be determined prior to selection of an appropriate animal model for safety evaluation. The nature of cross-reactivity can be based on a combination of phar-

macologic activity, receptor homology, and target tissue binding, as appropriate. The value of highly human-specific and tumor-specific cell targeting in cancer treatment is unequivocal, but the impact of highly specific targeted human therapies on the selection of appropriate animal models is significant. Often the more targeted the therapy, the more limited is the choice of animal model; in many cases species cross-reactivity is limited to nonhuman primates (NHPs). For biopharmaceuticals that target a receptor (or biological process) that is expressed (or active) at very low levels in normal human tissue, and detected at even lower levels in normal NHP tissue, there is likely to be little opportunity to assess toxicities associated with exaggerated pharmacologic activity. The availability of pharmacodynamic markers for evidence of activity in preclinical safety studies is a useful tool in differentiating pharmacologic activity from suprapharmacologic toxicity and providing a meaningful therapeutic index. For example, potential clinical toxicities such as cytokine release associated with tumor lysis syndrome will not be evident in preclinical safety studies. In the case of highly selective biopharmaceuticals an apparent lack of toxicity in preclinical models may not be predictive of subsequent clinical experience.

Although biopharmaceuticals must be administered parenterally, their relatively large size, from roughly 15 kd for proteins such as recombinant human IL-2 (Proleukin®) to roughly 150 kd for monoclonal antibodies such as Herceptin®, confers a long half-life. The pharmacokinetic properties of biopharmaceuticals must be taken into account when determining dose and frequency of administration to reduce the potential for drug accumulation and potential toxicities secondary to suprapharmacologic exposure in vast excess of clinical relevance, and/or activation of endogenous scavenging mechanisms associated with massive protein overload. Either scenario can complicate interpretation of preclinical safety signals and extrapolation of appropriate clinical doses.

The immunogenic potential of biopharmceuticals is assessed in preclinical safety studies through careful monitoring of anti-therapeutic antibodies during the course of the treatment and recovery periods. An appropriate recovery period must be included for drug washout so that anti-drug antibodies can be measured in the absence of drug interference. Although it is generally agreed that there is little relevance to clinical safety of an observed anti-drug antibody response in preclinical safety studies, the generation of anti-drug antibodies can antagonize drug effect or result in untoward immune-mediated events that can confound the interpretation of the data.

### 3.3 Material Production

During the course of a drug development program, improvements in production methods, including cell line yield, process improvements, and formulation modification evolve to fit the scale and scope of the clinical trials, and ultimately the marketplace. For each change, appropriate bridging studies are conducted. Depending on the nature of the change, and the data generated during

comparability testing, further comparative studies in animals or even humans may be warranted. This topic is discussed in Part III (see Chapter Demonstration of Comparability of a Licensed Product after a Manufacturing Change).

## 4 APPLICATION OF STANDARD TOXICOLOGY MODELS

Many of the standard toxicology models apply to oncology programs, and there are some additional points worthy of consideration. Systemic toxicity studies have been discussed previously in the context of the duration of preclinical studies supporting clinical development, and these studies are essential for characterizing the safety profile of a biopharmaceutical. Given that most clinical development plans include multiple dosing in phase 1, the utility of single-dose animal studies for anything other than dose ranging or overt tolerability is questionable and should be supported scientifically. If the scientific hypothesis for clinical efficacy requires consistent exposure to drug over time, then single-dose animal studies as part of an IND-enabling program provide little information regarding potential human risks associated with clinical administration. As discussed previously, the treatment period for multiple-dose systemic toxicity studies must equal the protocol-specified treatment period of clinical trials. An exception to this exists for chronic toxicity studies conducted to support chronic treatment in patients; these studies are generally conducted in support of licensing authorizations but may be conducted earlier in the clinical development life cycle as needed to support clinical trials of six months' duration or greater. In general, chronic toxicity studies of six months are sufficient to support chronic treatment in oncology, which is consistent with ICH S6. In a "white paper" authored by a working group of BioSafe (a subcommittee of the Regulatory Affairs Committee of the Biotechnology Industry Organization, BIO) an analysis of publicly available data to determine the extent to which data from chronic studies was predictive of human response and whether the six-month duration was appropriate. The authors concluded that the six-month duration was generally appropriate to support chronic clinical dosing, although there may be specific circumstances that might warrant longer term studies [10].

With regard to safety pharmacology endpoints it is standard practice to include cardiovascular and respiratory endpoints as part of the multiple-dose systemic toxicity studies. For biopharmaceuticals lacking cross-reactivity in rodents, it is possible to conduct limited neurobehavioral assessments of peripheral and central nervous system function in nonhuman primates, but with the caveat that unlike humans or even dogs, these wild animals are most likely to mask any physical deficit that would place them at a selective disadvantage in their natural habitat. In addition to incorporating safety pharmacology endpoints into systemic toxicity studies, these studies can be further maximized by assessing local tolerance at the sites of drug administration, thereby eliminating the need for a separate study of local tolerance in rabbits.

The standard battery of genetic toxicity studies is considered irrelevant for biopharmaceuticals per ICH S6. However, exceptions are made in cases where cause for concern exists, such as the presence of an organic linker molecule in a not yet approved conjugated protein product, or where new and appropriate testing systems are developed to elucidate genotoxic risk. For biopharmaceuticals that received marketing approval prior to the finalization of ICH S6 in 1997, it is not unusual to note the presence of a full battery of genetic toxicity studies as a component of the approval package. Likewise the rodent bioassay for carcinogenic potential is generally considered to be inappropriate for biopharmaceuticals due to species-specificity and immunogenic potential of human/humanized protein products in rodents.

Reproductive toxicity studies may be conducted, depending on the mechanism of action of the biopharmaceutical. In cases where NHPs represent the suitable animal model for safety testing, a full battery of male and female fertility, embryonic and developmental toxicity, and well as late-stage gestational, parturition, placental and lactational drug transfer and infant developmental studies have been applied in the preclinical safety evaluation of protein therapeutics. These studies are lengthy and complex, and unlike systemic toxicity studies where all animals begin and end treatment at the same time, female animals are enrolled on study as their pregnancy and/or hormonal status permits, creating a rolling study start and estimated times of study completion. A novel study design for assessing the potential embryonic and developmental toxicity in rabbits that were expected to mount an anti-drug antibody response to treatment has been used to successfully identify toxicities associated with drug treatment (Table 3). Rabbits had been demonstrated to be a cross-reactive species to the experimental therapeutic (tenecteplase), and it was anticipated that dose administration throughout the period of organogenesis would result in the formation of anti-drug antibodies that would impact exposure. In that model the period of organogenesis (gestation days 6–18) is

**TABLE 3 Embryofetal toxicity study using divided dosing**

| Group | Dosing Days |
|---|---|
| Control | DG 6–18 |
| Low | DG 6–10 |
| Mid | DG 6–10 |
| High | DG 6–10 |
| Low | DG 11–14 |
| Mid | DG 11–14 |
| High | DG 11–14 |
| Low | DG 15–18 |
| Mid | DG 15–18 |
| High | DG 15–18 |

*Source*: See [11].

*Note*: DG: day of gestation; 18 rabbits/group.

divided into discrete dosing intervals, and separate groups of animals receive drug for one of the three 4- to 5-day intervals. This study satisfied ICH tripartite guideline stages C and D of the reproductive process in a nonrodent species.

Constructing a scientific rationale for the preclinical safety testing strategy in support of a clinical development plan for oncology products is a key component of drug development. The scientific rationale should take into consideration a number of factors, including the clinical patient population and duration of exposure, species-specificity, available and appropriate in vitro and in vivo models, maximizing the value of each systemic toxicity study for responsible animal use, immunogenicity and impact on exposure, and an awareness of drug production changes that can impact the ability to rely on the foundation of previously conducted preclinical safety studies. This is consistent with the thinking that should be applied to the creation of a preclinical safety plan regardless of the therapeutic indication for which a biopharmaceutical is being studied.

## 5 SUMMARY AND CONCLUSION

Despite meaningful clinical advances over the past two decades, a great deal of unmet medical need remains in oncology. Our increasing understanding of the biological mechanisms of cancer has enabled the development of highly specific, targeted therapies that are rewriting the cancer descriptors from "tissue of origin," as in breast lung, colon, and prostate cancer, to the underlying mechanisms of pathology, as in HER2 overexpression. Hybrid molecular entities such as antibody–drug conjugates, fusion proteins, and one-armed antibodies, once considered novel and unusual are becoming more prevalent as the physical attributes of each biopharmaceutical are engineered to meet a particular therapeutic need—a classic case of the basic engineering principle of "form follows function (see Chapters Strategy Considerations for Developing the Preclinical Safety Testing Programs for Protein Scaffold Therapeutics and Preclinical Safety Evaluation of Immunotoxins). As our understanding of common pathologic processes has evolved, biopharmaceuticals that were initially developed as oncology treatments have provided benefit in other areas of unmet medical need. For example, since the introduction of the IFN-α in 1986 for the treatment of a small, but significant subset of chronic lymphoid leukemias, a critical role for IFN-α therapy in hepatitis was established. Rituxan®, a chimeric monoclonal antibody to the CD-20 receptor of B cells, was initially approved for the treatment of non-Hodgkin's lymphoma, and is now approved for the treatment of patients with rheumatoid arthritis.

Additional safety studies may be required to support the movement into new indications, and the complex "designer" biopharmaceuticals will continue to require creative approaches to safety assessment. In either case, the requirement for, and design and execution of, those studies should be driven by sound scientific rationale. Ultimately, the objective of preclinical safety evaluation is well articulated in ICH S6:

- Identification of an initial safe starting dose and dose escalation scheme for clinical trials.
- Identification of potential target organs of toxicity and reversibility of effect.
- Identification of safety parameters for clinical monitoring.

These objectives are best fulfilled through thoughtful planning and significant cross-functional generation of clinical development assumptions by toxicologists, clinicians, researchers, pharmacokineticists, bioanalytical assay scientists, and the scientists and engineers responsible for making and formulating the biopharmaceutical drug product. The design and implementation of scientifically appropriate in vitro and in vivo studies that are integrated into the support of a well-articulated clinical plan will best allow us to expeditiously move promising new therapies forward to patients in need while judiciously applying the human, animal and financial resources required.

## REFERENCES


1. Saven A, Ellison DJ, Piro, LD. *Clinical Oncology*. New York: Churchill Livingstone Inc. 1995;2023–34.
2. www.accessdata.fda.gov/scripts/cder/onctools/yearlistclaim.cfm
3. www.accessdata.fda.gov/scripts/cder/onctools/yearlistclaim.cfm
4. Pilaro A. Nonclinical perspective on initiating phase 1 studies for biological oncology products—Case studies. Oncology Drugs Advisory Committee, March 13, 2006.
5. Green M. Pre-clinical requirements for phase 1 studies—biological oncology products. Oncology Drugs Advisory Committee, March 13, 2006.
6. Reynolds T, Baughman S, Schofield C, Palazzolo M, Dalgard D, Thomas DA. A 3-week interaction study of rhuMAb HER2 with doxorubicin, cyclophosphamide and paclitaxel in rhesus monkeys. *Toxicologist* 1996;30(1):36.
7. Modi NB, Lin YS, Reynolds T, Shaheen A, Christian BC. Pharmacokinetics of Xubix, an orally active IIbIIIa antagonist, in the presence of heparin, aspirin and rt-PA in beagles. *J Cardiovas Pharmacol* 1998;32:397–405.
8. Gaudreault J, Shiu V, Bricarello A, Christian BJ, Zuch CL, Mounho B. Concomitant administration of bevacizumab, irinotecan, 5-fluorouracil, and leucovorin: nonclincial safety and pharmacokinetics. *Intl J Toxicol* 2005;24:357–63.
9. 21 CFR part 3.2(e), Docket No. 2004N-0194, Federal Register, Vol 70, No. 164, 8/25/05.
10. Clarke J, Hurst C, Martin P, Vahle J, Ponce R, Mounho B, Heidel S, Andrews, L, Reynolds T, Cavagnaro J. Duration of chronic toxicity studies for biotechnology-derived pharmaceuticals. *Regul Toxicol Pharmcol* 2008;50:2–22.
11. Gross MC, Leach W, Prince W, Hoberman A, Barnett J, O'Neill C. Embryotoxic and teratogenic evaluation of TNK-tPA in rabbits: a study around antigenicity issues. *Toxicologist* 2000;54(1):397.

# Preclinical Safety Evaluation of Monoclonal Antibodies

GEORGE TREACY, MS, and PAULINE MARTIN, PhD

**Contents**

1 Introduction
2 Fc Binding of Monoclonal Antibodies
3 Fully Murine Monoclonal Antibodies
4 Chimeric Monoclonal Antibodies
5 Humanized Monoclonal Antibodies
6 Fully Human Monoclonal Antibodies
7 Preclinical Development Programs with Monoclonal Antibodies
8 Future of Antibody Development
9 Summary
References

## 1 INTRODUCTION

Monoclonal antibodies (mAbs) are populations of identical, monospecific antibodies with defined specificity and affinity for a target antigen. The high degree of specificity for their target is the property that makes monoclonal antibodies so valuable clinically. As such, they can be used not only therapeutically to treat and prevent disease but also to diagnose a wide variety of diseases, and to detect drugs, abnormal proteins, viruses, and bacteria. Monoclonal antibodies are produced by the daughter cells of a single antibody-producing lymphocyte, often using an immortal hybridoma cell line grown in vitro. Monoclonal antibodies can also be constructed synthetically and produced as engineered recombinant proteins. In both cases cells can be grown continually in culture to produce large amounts of protein.

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

Since the approval of the first therapeutic monoclonal antibody in 1986, considerable advances have been made in antibody technology [1,2]. With the improvement in antibody technology an increase in the clinical success of this group of molecules has been seen [3].

Monoclonal antibodies exhibit the same general structural and functional characteristics as naturally acquired antibodies, and they can be of any antibody class (IgG, IgM, IgD, IgE, or IgA) or isotype within a class. The most commonly developed monoclonal antibodies are of the IgG class, and they are tetrameric proteins consisting of two identical heavy chains and two identical light chains (Figures 1 and 2). Each IgG molecule contains two antigen combining sites formed by the N-terminal regions of the heavy and light chains, which determine the antigen recognition and binding properties of the monoclonal antibody. The specificity of a given monoclonal antibody is determined by the precise amino acid sequences of the heavy and light protein chains in the antigen-combining region (complementary-determining region, CDR); the

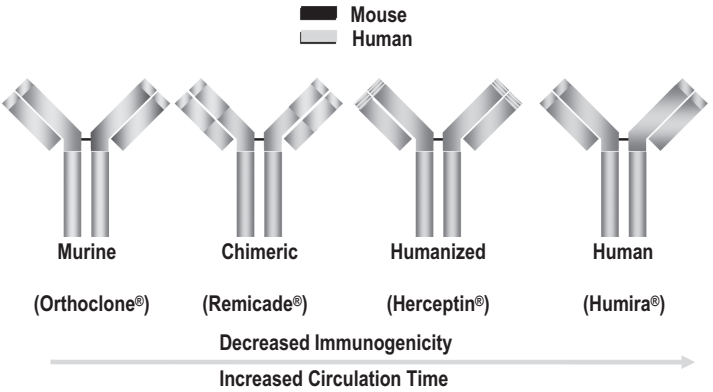


**Figure 1** Structure of a monoclonal antibody and its interaction with antigen. See color insert.

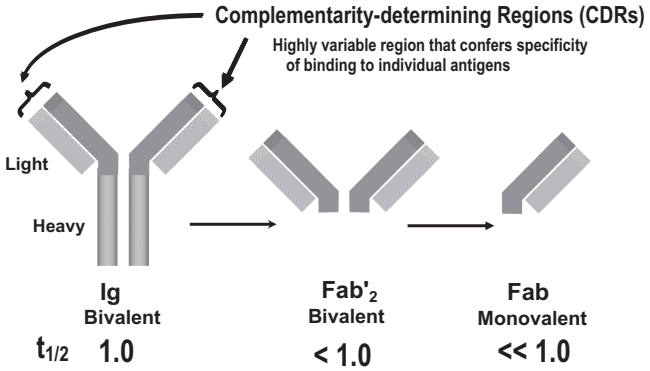


**Figure 2** Examples of various classes of monoclonal antibodies. See color insert.

sequences of this region are therefore somewhat variable among monoclonal antibodies of different specificities. The antigen-combining region is also known as the variable region because of this sequence variation. The remainder of the molecule is relatively constant in sequence among different monoclonal antibodies, and it functions to allow interactions with other immune system components. This region of the monoclonal antibody determines the class and isotype of the antibody, and is known as the constant region or Fc region.

Monoclonal antibodies can be grouped according to their potential use: (1) binding to a cell surface target, with recruitment of immune response and target cell lysis; (2) binding to a cell surface receptor causing apoptosis; (3) cross-linking to a cell-killing reagent (e.g. immunotoxin); (4) binding to a target to block an interaction (antagonist); (5) binding to a receptor to stimulate a downstream process (agonist); and (6) catalysis (catalytic antibodies).

Many monoclonal antibodies have now been approved for the diagnosis or treatment of various diseases. The target antigens include soluble factors such as tumor necrosis factor (TNFα), vascular endothelial growth factor (VEGF) and IgE, cell surface antigens on lymphocytes (CD3, CD20, CD25, CD52, CD11a, CD49d), antigens upregulated on tumor cells (HER2, EGFR), and viral specific antigens RSV (see Chapter Preclinical Safety Evaluation of Biological Oncology Drugs). Antibodies have also been used to deliver toxins or radioactivity to tumor cells and imaging agents to tumors (see Chapter Preclinical Safety Evaluation of Immunotoxins). In addition to a high degree of specificity and affinity for the target antigen, monoclonal antibodies exhibit in vivo pharmacokinetic properties similar to those of naturally acquired antibodies.

## 2 FC BINDING OF MONOCLONAL ANTIBODIES

Besides the binding of the variable region (Fab) of the monoclonal antibody to the target antigen, some monoclonal antibodies also require binding of the constant region to receptors on effectors cells. These receptors, referred to as Fcγ receptors (FcγR), are expressed on immune cells and consist of three distinct classes FcγRI (CD64), FcγRII (CD32), and FcγRIII (CD16). Binding to FcγR can lead to antigen-dependent cellular cytotoxicity (ADCC), phagocytosis, or other cellular effects. ADCC can be beneficial when the target antigen is present on a tumor cell. However, when the target antigen is present on lymphocytes, ADCC may be detrimental. For example anti-CD3 and anti-CD4 antibodies can lead to T cell depletion. Also binding of monoclonal antibodies to immune cell antigens may lead to cytokine release syndrome that may result in side effects of fever, nausea, and bronchospasm. IgG1 antibodies have greater potential effector function than IgG2 and IgG4 antibodies. However, IgG1 antibodies can be modified to have reduced or enhanced effector function, depending on the desired therapeutic effects [4,5].

Effector function can also be eliminated by removing the Fc portion of the molecule to produce a Fab molecule (e.g., abciximab). However, Fab molecules

have very rapid elimination from the circulation relative to full-length molecules, and for certain clinical indications this may not be desirable. The reason for the shorter half-life of Fab molecules relative to full-length antibodies is that binding of the Fc portion of the antibody to another type of Fc receptor found on endothelial cells (FcRn) leads to antibody internalization and recycling, resulting in enhanced serum persistence. Therefore molecules that show poor binding to FcRn are rapidly eliminated from the serum.

## 3 FULLY MURINE MONOCLONAL ANTIBODIES

The first therapeutic monoclonal antibodies developed were fully murine sequence antibodies. These antibodies were developed by immunizing mice with a human antigen. Antigen-specific B cells were isolated from the immunized mice and were fused to an immortal cell line to generate monoclonal hybridomas secreting fully murine mAbs. (Figure 3). The first approved monoclonal antibody was muromomab-CD3, which is a murine monoclonal antibody against human CD3 on T lymphocytes. This antibody was developed for the treatment of acute transplant rejection. The main disadvantage of using murine monoclonal antibodies to treat human diseases is that because the antibodies contain murine sequences, they can be highly immunogenic in humans. Humans treated with muromomab-CD3 develop human anti-murine antibodies (HAMA) that result in reduced exposure to the antibody and a

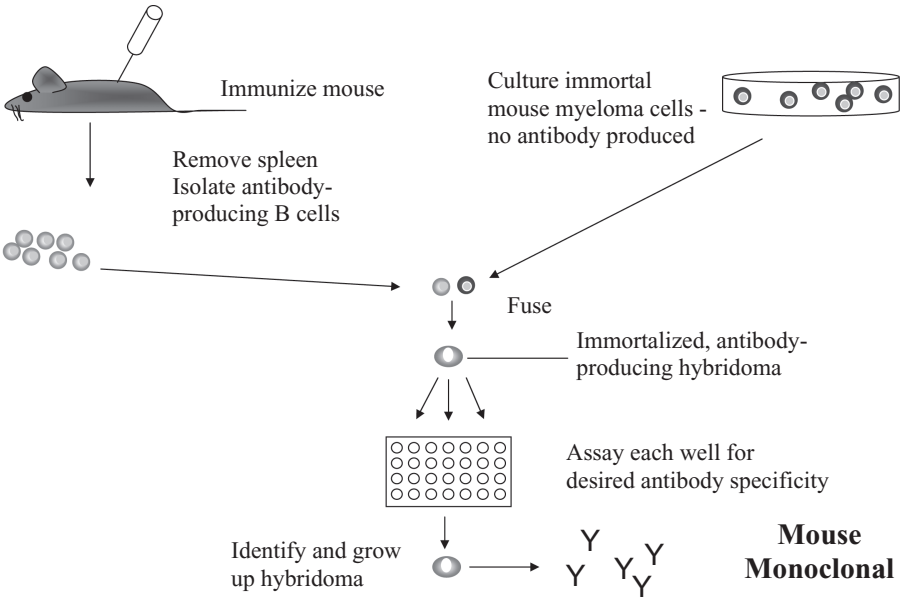


**Figure 3** Production of murine monoclonal antibodies.

loss of efficacy even in the presence of concomitant immune suppression. Muromomab and other antibodies that bind to T cell receptors and activate the T cell can also cause cytokine release, which produces a systemic inflammatory response characterized by hypotension, pyrexia, and rigors [6]. The cytokine release can cause life-threatening pulmonary edema. Deaths due to cytokine release syndrome have been reported with OKT3. Cytokine release can be anticipated then for monoclonal antibodies that activate T cells and can usually be treated appropriately.

In addition to the immunogenic potential of murine monoclonal antibodies, another disadvantage of murine antibodies is their rapid elimination from the circulation relative to human antibodies. Muromomab has an elimation half-life of about 18 hours, whereas human antibodies have half-lives of weeks. Human antibodies have long serum persistence because they are recycled by endothelial cells. The Fc portion of murine monoclonal antibodies binds very poorly to human FcRn, and therefore murine antibodies are not recycled resulting in their poor serum persistence [7]. The converse, however, is not true. Human Fc binds well to murine FcRn, and therefore human antibodies have long serum persistence in mice. The lack of FcRn binding in combination with a HAMA response results in rapid clearance from the serum.

Because of these limitations murine antibodies have had very limited clinical use. The other murine antibodies include two radiolabeled versions of anti-CD20 antibodies that have been developed for the treatment of non-Hodgkins lymphoma and an anti-PSMA antibody-imaging agent for visualizing prostate tumors (Table 1). These antibodies are intended for short-term use and are administered at low doses. For these antibodies the short serum half-life and the potential to develop HAMA are not considered a limitation for therapy. Interestingly, since ibritumomab tiuxetan and tositumomab-$I^{131}$ are B cell depleting therapies, their mechanism of action reduces the immune response [8].

**TABLE 1 Summary of approved fully murine monoclonal antibodies**

| Generic Name | Trade Name | Sponsor Company | Mechanism | Antibody Isotype | Indication | Approval Date |
|---|---|---|---|---|---|---|
| Muromonab-CD3 | Orthoclone® OKT3 | Ortho Biotech | Anti-CD3 | IgG2a | Transplant rejection | 1986 |
| Abciximab | ReoPro® | Centocor | Anti-GPIIb/IIa | Fab | Cardiovascular | 1997 |
| Capromab pendetide | ProstaScint® | Cytogen Corp | Anti-PSMA | IgG1 kappa | Oncology imaging | 1996 |
| Ibritumomab tiuxetan | Zevalin® | Biogen Idec | Anti-CD20-Yttrium 90 | IgG1 kappa | Oncology | 2002 |
| Tositumomab-$I^{131}$ | Bexxar® | Corixa | Anti-CD20 Iodine 131 | IgG2a Lambda | Oncology | 2003 |

**TABLE 2 Summary of approved chimeric monoclonal antibodies**

| Generic Name | Trade Name | Sponsor Company | Type | Antibody Isotype | Indication | Approval Date |
|---|---|---|---|---|---|---|
| Rituximab | Rituxan® | Genentech | Anti-CD20 | IgG1 kappa | Oncology | 1997 |
| Basiliximab | Simulect® | Novartis | Anti-CD25 | IgG1 kappa | Transplant rejection | 1998 |
| Infliximab | Remicade® | Centocor | Anti-TNFα | IgG1 kappa | Immune-mediated diseases | 1998 |
| Cetuximab | Erbitux® | Imclone | Anti-EGFR | IgG1 kappa | Oncology | 2004 |

## 4 CHIMERIC MONOCLONAL ANTIBODIES

Because of the limitations seen with the fully murine monoclonal antibodies the next generation of antibodies were chimeric antibodies (Table 2, Figure 2). Chimeric antibodies are constructed from variable regions derived from a murine source and constant regions derived from a human source. The chimeric antibodies show less of an immunogenic response than the fully murine antibodies, but human anti-chimeric antibody (HACA) rates can still be high with these antibodies. The chimeric antibodies nevertheless have relative long serum half-lives (9.5 days for infliximab) relative to the fully murine antibodies because the Fc portion is human and can therefore bind to human FcRn in a manner similar to that of native antibodies.

## 5 HUMANIZED MONOCLONAL ANTIBODIES

Humanized antibodies are constructed with only antigen binding regions (CDRs) derived from a mouse, and the remainder of the variable region derived from a human source. The humanized monoclonal antibodies represent the largest group of the currently approved monoclonal antibodies (Table 3, Figure 2). These antibodies have been highly successful in the clinic, and few serious immune-mediated adverse reactions have been observed. Two of the approved humanized monoclonal antibodies are IgG4 antibodies, gentuzumab ozogamicin and natalizumab. The remainder of the humanized antibodies are IgG1 antibodies. IgG4 antibodies have the potential advantage over IgG1 antibodies in certain situations because they lack effector function.

## 6 FULLY HUMAN MONOCLONAL ANTIBODIES

As of 2006 only two fully human monoclonal antibodies had been approved (Table 4). Adalimumab is a fully human antibody against human TNFα that

**TABLE 3 Summary of approved humanized monoclonal antibodies**

| Generic Name | Trade Name | Sponsor Company | Mechanism | Antibody Isotype | Indication | Approval Date |
|---|---|---|---|---|---|---|
| Daclizumab | Zenapax® | Hoffman-La Roche | Anti-CD25 | IgG1 | Transplant rejection | 1997 |
| Palivizumab | Synagis® | MedImmune | Anti-RSV | IgG1 kappa | Anti-viral | 1998 |
| Trastuzumab | Herceptin® | Genentech | Anti-HER2 | IgG1 kappa | Oncology | 1998 |
| Gemtuzumab ozogamicin | Mylotarg® | Wyeth-Ayerst | Anti-CD33 immunotoxin | IgG4 kappa | Oncology | 2000 |
| Alemtuzumab | Campath® | Millennium/LEX | Anti-CD52 | IgG1 kappa | Oncology | 2001 |
| Omalizumab | Xolair® | Genentech | Anti-IgE | IgG1 kappa | Asthma | 2003 |
| Efalizumab | Raptiva® | Genentech | Anti-CD11a | IgG1 kappa | Psoriasis | 2003 |
| Bevacizumab | Avastin® | Genentech | Anti-VEGF | IgG1 | Oncology | 2004 |
| Natalizumab | Tysabri® | Biogen idec | Anti-α4-integrin | IgG4 | Multiple sclerosis | 2004 |

**TABLE 4 Summary of approved fully human monoclonal antibodies**

| Generic Name | Trade Name | Sponsor Company | Mechanism | Antibody Isotype | Indication | Approval Date |
|---|---|---|---|---|---|---|
| Adalimumab | Humira® | Abbott | Anti-TNFα | IgG1 kappa | Immune-mediated diseases | 2002 |
| Panitumumab | Vectibix™ | Amgen | Anti-EGFR | IgG2 kappa | Oncology | 2006 |

was developed from phage display technology. However, even though adalimumab is a fully human antibody, it has shown a significant human antihuman antibody (HAHA) response in patients [8]. Therefore fully human antibodies may not necessarily be an advantage from an immunogenicity perspective.

## 7 PRECLINICAL DEVELOPMENT PROGRAMS WITH MONOCLONAL ANTIBODIES

The design of preclinical safety evaluation of monoclonal antibodies is based on the class of monoclonal, the target epitope, and the clinical indication. Safety evaluation programs are then developed based on the availability of relevant species and specific attributes of the product on a "case-by-case" basis. Prior to issuance of regulatory guidance documents by the FDA and ICH [9,10] safety evaluations have included the use of nonrelevant assays (e.g., gentoxcity assays) and nonrelevant species. The nonclinical development programs have also included the use of murine homologues (surrogate

monoclonal antibodies) when warranted. See Table 5 for examples of development programs for some nononcology indications.

The property of monoclonal antibodies that makes them highly desirable as therapeutics, namely the high degree of specificity for their target, also makes then a challenge from a preclinical development perspective. The earliest monoclonal antibodies were so specific for binding to their human target protein that they did not bind to the analogous target protein in any animal species except for chimpanzees. The chimpanzee is not an appropriate species for toxicology testing because the amount of nonclinical safety information that can be acquired from the chimpanzee is severally limited. Infliximab and efalizumab are examples of monoclonal antibodies developed to treat chronic non–life-threatening indications that bound only to their target in humans and chimpanzees. The preclinical safety program using the human monoclonal antibody was therefore restricted to limited nonterminal safety studies in chimpanzees and to in vitro tissue cross reactivity to a panel of human tissues.

In vitro tissue cross-reactivity studies are a key component of the safety assessment of monoclonal antibodies. These studies are required by regulatory authorities, and they need to be conducted on a panel of human tissues [9,10,11]. The purpose of these studies is to identify potential binding of the monoclonal antibody to nontarget tissues. A comparison of the in vitro cross-reactivity to human tissues and to tissues from one or more animal species may be performed to help determine species relevance for subsequent toxicology evaluation. In some cases tissue cross-reactivity alone has been used as a criterion for species selection for toxicology. Demonstration of a similar binding profile in human and animal tissues alone is, however, not usually considered to be an adequate means for selecting a toxicology species. The monoclonal antibody should be shown not only to bind to the analogous protein in the animal but also to produce a similar biological response in the human and in the animal. However, for some monoclonal antibodies that target tumor-associated antigens or virus (e.g., RSV) there will be no biological assay to show bioactivity between human and animal.

Because of the limited safety information provided by the chimpanzee studies and the in vitro human tissue cross-reactivity study, the need to develop an alternate strategy to evaluate safety for the human/chimpanzee specific monoclonal antibodies was appreciated. The ICH S6 guidance document for the preclinical safety testing of biotechnology-derived pharmaceuticals includes a provision for using a surrogate monoclonal antibody in such cases [9]. In both the infliximab and efalizumab examples anti-murine monoclonal antibodies were developed that could be tested in the mouse [12,13]. With these surrogate molecules chronic toxicity studies as well as reproductive toxicity studies were conducted. However, even though the monoclonal antibody being tested in these studies was not the same monoclonal antibody that was being evaluated in clinical trials, this alternate approach is considered to be acceptable. The preclinical safety program for infliximab was initiated

**TABLE 5 Example preclinical safety evaluation programs with monoclonal antibodies**

| Product Name | Species Specificity | Initial Clinical Indication | Types of Studies to Support Approval |
|---|---|---|---|
| Infliximab (anti-TNFα) Remicade® | Chimpanzee specific | Crohn's disease | Tissue cross-reactivity: human<br>Genetic toxicity: Ames,[a] in vitro chromosomal abberations human lymphocyte (CAHL),[a] mouse micronucleus (MM)[a]<br>Toxicity: single dose—chimpanzee, rat;[b] repeat dose—chimpanzee, rat,[b] mouse[c]<br>Fertility: mouse[c]<br>Embryofetal development: mouse[c]<br>Pre- and postnatal development: mouse[c] |
| Efalizumab (anti-CD11a) Raptiva® | Chimpanzee specific | Psoriais | Tissue cross-reactivity: human, chimpanzee, mouse[c]<br>Toxicity: single dose—chimpanzee; mouse;[c] repeat dose—chimpanzee, mouse[c]<br>Immunotoxicity: chimpanzee, mouse[c]<br>Fertility: mouse[c]<br>Embryofetal development: mouse[c]<br>Pre- and postnatal development: mouse[c] |
| Adalimumab (anti-TNFα) Humira® | Reacts with cynomolgus macaque, no cross-reactivity with rodents | Rheumatoid arthritis | Tissue cross-reactivity: human, chimpanzee, cynomolgus, marmoset, rhesus, baboon, rat<br>Genetic toxicity: Ames[a], MN[a]<br>Toxicity: single dose—mouse[b] and rat;[b] repeat dose—mouse[b] and cynomolgus[b]<br>Embryofetal development: cynomolgus[b] |

**TABLE 5** ***Continued***

| Product Name | Species Specificity | Initial Clinical Indication | Types of Studies to Support Approval |
|---|---|---|---|
| Omalizumab (anti-IgE) Xolair® | Reacts with cynomolgus macaque, no cross-reactivity to rodents | Asthma | Tissue cross-reactivity: human, cynomolgus<br>Genetic toxicity: Ames[a]<br>Toxicity: single dose—mouse,[b] cynomolgus; repeat dose—cynomolgus<br>Fertility: cynomolgus<br>Embryofetal development: cynomolgus |
| Palivizumab (anti-RSV) Synagis® | Reacts with antigen on RSV<br>No cross-reactivity to human or animal antigens | Prophylaxis of serious lower respiratory tract disease caused by RSV | Tissue cross-reactivity: human, cynomolgus<br>Toxicity: single dose—cynomolgus[b] and rabbit;[b] repeat dose—rat[b]<br>Virus challenge model: cotton rat |

*Note*: All information included in this table and the accompanying text have been extracted from the US FDA summary basis of approval information (www.fda.gov) the European Medicines Agency (EMEA) centrally authorized product reviews (EPARs; www.emea.europa.eu) and the *Physicians Desk Reference* (PDR).

[a]Nonrelevant test system.

[b]Nonrelevant species.

[c]Use of murine homologue (surrogate antibody).

before the issuance of the ICH S6 guidance document, and consequently some studies, now considered to be unnecessary, were conducted with infliximab. For example, gentoxicity studies are now considered not to be appropriate for biopharmaceuticals and studies in biologically nonrelevant species are not considered to be appropriate. The only example where studies in a nonrelevant species may be appropriate is where similar nontarget binding of the monoclonal antibody is demonstrated to human tissues and to tissues in a biologically nonrelevant species.

Omalizumab and adalimumab are examples of monoclonal antibodies, developed for chronic non–life-threatening indication, that showed cross-reactivity to cynomolgus macaques as well as to humans. This broader species cross-reactivity allowed for a more thorough preclinical safety evaluation of the human monoclonal antibody developed for human use in the cynomolgus macaque. In the case of omalizumab, fertility studies and developmental toxicitiy studies were conducted in macaques in addition to the chronic toxicity studies. For adalimumab, no reproductive and developmental toxicity studies were conducted. The value of conducting fertility and developmental studies in macaques with monoclonal antibodies is described in Chapter Reproductive Toxicity Testing for Biopharmaceuticals.

The last example shown in Table 5 is palivizumab, a monoclonal antibody against an antigen expressed on a virus (RSV). Since this antigen does not exists in any animal species, including humans, all toxicology species are considered equally nonrelevant. The only relevant species is one that is infected with the virus. This example provides an interesting case where animal disease models could be considered more relevant than normal animals. Nevertheless, for palivizumab single-dose acute toxicity studies were performed in rats, rabbits, and macaques.

## 8 FUTURE OF ANTIBODY DEVELOPMENT

Recombinant DNA technologies can be used to manipulate monoclonal antibody sequences to produce monoclonal antibodies with improved properties, such as higher affinity, increased functional activity, and reduced immunogenicity for in vivo applications. Human sequence antibodies from transgenic mice, phage display, ribosome display, yeast display libraries, monoclonal antibody fragments, single-chain Fv, single-domain fragments, diabodies, minibodies (scFv fused to CH3 domain of mAbs), bispecific and multivalent mAb toxin-conjugates, and immunocytokines [2]. It is now possible to engineer monoclonal antibodies to have desired properties such as optimal serum half-life, reduced immunogenicity, and reduced or enhanced ADCC and CDC activities. It is also possible to engineer some monoclonal antibodies that will cross-react with multiple species, including rodents. The development of these monoclonal antibodies could allow more extensive toxicology evaluation than could be conducted with the species restricted monoclonal antibodies, assuming they are not immunogenic in the animals.

## 9 SUMMARY

Monoclonal antibodies have proved to be value therapeutic and diagnostic agents. In general, the clinical safety of monoclonal antibodies has been good. The high degree of specificity of monoclonal antibodies for their targets has contributed to the high success rate in the clinic. However, their high degree of specificity has also restricted the nonclinical safety testing that can be conducted with these molecules. The nonclinical development programs for monoclonal antibodies have been developed on a case-by-case basis depending on the clinical indication, the intended clinical use, and an understanding of the biology of the targeted antigen. In the future the advancement in monoclonal antibody technology may lead to monoclonal antibodies that do not behave like natural antibodies and may be less species restricted. In these cases the nonclinical safety testing may differ from that of the currently approved monoclonal antibodies.

## REFERENCES

1. Weiner LM. Fully human therapeutic monoclonal antibodies. *J Immunother* 2006; 29:1–9.
2. Prezta LG. Engineering of therapeutic antibodies to minimize immunogenicity and optimize function. *Adv Drug Deliv Rev* 2006;58:640–56.
3. Reichert JM, Rosensweig CJ, Faden LB, Dewitz MC. Monoclonal antibody successes in the clinic. *Nat Med* 2005;23:1073–8.
4. Armour MR, Clark AG, Hadley LM, Williamson LM: Recombinant human IgG molecules lacking Fc gamma receptor I binding and monocyte triggering activities, *Eur J Immunol* 1999;29:2613–24.
5. Sheilds RL, Namenuk AK, Hong K, Meng J, Rae J, Briggs J, Xie D, Lai J, Stadlen A, Li B, Presta LG. High resolution mapping of the binding site on human IgG1 for FcγRI, FcγRII, FcγRIII and FcRn and design of IgG1 variants with improved binding to the Fcγ. *J Biol Chem* 2001;276:6591–604.
6. Sgro C. Side effects of a monoclonal antibody, muromomab CD3/orthoclone OKT3. *Toxicology* 1995;105:23–9.
7. Ober RJ, Radu CG, Ghetie V, Ward ES. Differences in promiscuity for antibody-FcRn interactions across species: implications for therapeutic antibodies. *Intl Immunol* 2001;13:1551–9.
8. Hwang WYK, Foote J. Immunogenicity of engineered antibodies. *Methods* 2005; 36:3–10.
9. United States-Food and Drug Administration (CBER). Points to Consider in the Manufacture and Testing of Monoclonal Antibody Products for Human Use, 1997. http://www.fda.gov/cber/gdlns/ptc_mab.pdf
10. ICH Harmonized Tripartite Guideline. S6 Preclinical Safety Evaluation of Biotechnology-Derived Pharmaceuticals, 1997. http://www.ich.org/pdfICH/s6.pdf

11. European Medicines Evaluation Agency (EMEA) Guideline. Production and Quality Control of Monoclonal Antibodies, Directive 75/318/EEC, 1995. http://www.q-one.com/guidance/emea.htm
12. Treacy G. Using an Analogous Monoclonal Antibody to Evaluate the Reproductive and Chronic Toxicity Potential for a Humanized Anti-TNFα Monoclonal Antibody. *Hum Exp Toxicol* 2000;19:226–9.
13. Clark J, Leach W, Pippig S, Joshi A, Wu B, House R, Beyer J. Evaluation of a surrogate antibody for preclinical safety testing of an anti-CD11a monoclonal antibody. *Regul Toxicol Pharmacol* 2004;40:219–26.

# Immunomodulatory Biopharmaceuticals and Risk of Neoplasia

PETER J. BUGELSKI, PhD, FRCPath, CLIFFORD SACHS, PhD, DABT, JOEL CORNACOFF, DVM, PhD, DABT, PAULINE MARTIN, PhD, and GEORGE TREACY, MS

**Contents**

1 Introduction
2 Clinical Experience with Immunosuppressive Agents and IMBP
3 Immunomodulatory Biopharmaceuticals are Unlikely to be Complete Carcinogens
4 Immunomodulatory Biopharmaceuticals are Unlikely to be Conventional Tumor Promoters
5 Immunomodulatory Biopharmaceuticals as Atypical Tumor Promoters
6 Immunomodulatory Biopharmaceuticals as Potential Tumor Progressors
7 Critical Review of Model Systems for Identification and Characterization of Atypical Promoters and Tumor Progressors
7.1 In vitro Mechanism-Based Systems
7.2 In vivo Systems
8 Classification of the Hazard Posed by IMBP and Risk Assessment
9 Conclusions
References

## 1 INTRODUCTION

A number of immunomodulatory biopharmaceuticals (IMBPs), such as peptides, recombinant proteins, soluble receptors, and monoclonal antibodies, have been approved or are currently in development to treat chronic inflammatory diseases. These agents may be administered daily or, if long-acting,

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

intermittently. For IMBP, two immunologic characteristics are important: (1) immunogenicity due to the response to a foreign protein and (2) the intended pharmacologic effects. These two properties may interact and facilitate achieving adequate exposure in toxicity studies. That is, the higher doses employed in toxicity studies may result in suppressed immunogenicity and thus possibly revealing clinically relevant, immunosuppressant activity or dose correlates [1]. Special issues or concerns following chronic treatment with IMBPs include the potential for immune impairment leading to opportunistic infections and/or lymphoproliferative disorders [2]. As is true for conventional pharmaceuticals, in addition to duration of therapy, the cause for concern for tumorigenicity (malignancy) may be heightened based on knowledge and plausibility of the mechanism of drug action, especially if an IMBP is administered with overtly immunosuppressive drugs. However, despite years of experience with testing environmental contaminants, food additives, and pharmaceuticals, predicting and quantitating an increased risk of neoplasia in patients receiving therapeutic doses of IMBPs remains a daunting task. Moreover, because the risk of neoplasia is a multifactorial issue, no "one size fits all" approach is likely to have much value.

The discussion in this chapter will focus on well-characterized, high-purity protein biopharmaceuticals derived from conventional recombinant expression systems that modulate immune function. Specifically, the points we raise are relevant to monoclonal antibodies, fusion proteins, and soluble receptors directed toward cytokines, cells, and receptors that regulate immune function. Also the discussion will include biopharmaceuticals that downmodulate immune function (e.g., a monoclonal antibody that neutralize a cytokine) or the "naked" antibody intended to selectively eliminate a specific immune cell population (e.g., an anti-CD4 antibody), that is, "naked" as the term is used to discriminate among nonconjugated antibodies and immunotoxins or radioimmunotherapeutics. Excluded from consideration will be biopharmaceuticals that cause sustained, polyclonal expansion of lymphoid cells, express growth factor-like effects on non-immune cells, or bear chemical modifications that can introduce genotoxic moieties. Nevertheless, our discussion will be germane to monoclonal antibodies that are intended to direct an immune response toward a non-immune cellular target, such as a tumor antigen. As with most issues concerning the safety of biopharmaceuticals, for any specific agent a science-driven, case-by-case approach that is consistent with the principles set forth in ICH S6 and is developed in the context of our growing experience with biopharmaceuticals in general and our knowledge base on that specific biopharmaceutical will be required for a proper risk assessment.

Implicit to our recommendation for IMBPs is that viral expression system derived DNA or excipient derived contaminant carcinogens are no longer an issue. Although, for example, host cell DNA can be a risk factor for neoplasia [3], the current stringent biochemical characterization, purification, and virus inactivation steps taken with all well-characterized biopharmaceuticals adequately mitigate any potential risk from these sources.

## 2 CLINICAL EXPERIENCE WITH IMMUNOSUPPRESSIVE AGENTS AND IMBP

It is well understood that cancer is no single disease. Each cell or tissue of origin likely has its own cause of neoplasia and, in turn, expresses a different disease. What is relevant for one tumor type, such as breast cancer, may have little relevance for another, such as lymphoma. Similarly not all agents that increase the risk of neoplasia are likely to act through the same mechanism or cause the same tumors. To set the stage for discussing predicting risk of neoplasia for patients receiving IMBPs, we should understand what neoplasms are relevant for those patients. Because clinical data on IMBPs are limited, to do this, we will turn to the more extensive data on small-molecule pharmaceuticals (xenobiotics).

For a number of overt, "broad spectrum" immunosuppressive xenobiotics (e.g., azathioprine) there is sufficient clinical experience to indicate the types of neoplasms for which there is an increased risk. These tumor types are listed in Table 1. Also listed are the tumors that occur in the unfortunate "experiment of nature," namely patients infected with human immunodeficiency virus type 1 (HIV-1) and the tumors that may occur at higher incidence with more selective yet strong immunosuppressants (e.g., cyclosporin, sirolimus, and tacrolimus). Compared to the broad spectrum immunosuppressive agents listed above, most IMBPs express a highly selective regulatory influence on the immune system; modulating the activity of host defense systems rather than mediating frank immunosuppression.

As shown by Table 1, the list of tumor types relevant to immunosuppression is surprisingly short. Although there are isolated case reports of a number of other tumor types, such as lung cancer [4,5], in patients receiving broad spectrum immunosuppressive therapies or in HIV-infected patients, the tumor types listed in Table 1 account for the vast majority of neoplasms associated with immunosuppression and IMBPs. Notably the major tumor types afflicting humans—colon, breast, and prostate—are not included because, as will become clear from our discussion, the increased risk of neoplasia posed by immunosuppression and especially IMBPs cannot be generalizabled to all neoplasms. The strategy we adopt for hazard identification and risk assessment for IMBPs therefore does not address the increased incidence of all possible tumor types. For all tumor types listed either as viral or radiation based in their etiology (Table 2), the strategy we adopt for hazard identification and risk assessment for IMBPs will not address the potential mechanisms of carcinogenesis.

## 3 IMMUNOMODULATORY BIOPHARMACEUTICALS ARE UNLIKELY TO BE COMPLETE CARCINOGENS

As noted above, the increased risk of neoplasia due to immunosuppression or IMBPs is limited to tumor types where a defined viral or radiation-based

**TABLE 1 Human neoplasms associated with immunosuppression**

| Tumor Type | Immunosuppressant with Increased Incidence | Reference | Putative Carcinogen(s) | Reference |
|---|---|---|---|---|
| Non-Hodgkin's lymphoma | Cyclosporin<br>Azathioprine and 6-mercaptopurine<br>Sirolimus<br>Human immunodeficiency virus (HIV) infection | Mougel et al. [40], Kwon and Farrell [42]<br>Ibanez et al. [45], Schulz et al. [46] | Epstein-Barr virus (EBV)<br>Human T cell leukemia virus-I (HTLV-I)<br>Human Herpes Virus 8 (HHV-8) | Ahmed and Heslop [41], Jarrett [43]<br>Malnati et al. [44] |
| Hodgkin's lymphoma | Sirolimus | Martin–Gomez et al. [47] | Epstein-Barr virus (EBV) | Ahmed and Heslop [41] |
| Kaposi's sarcoma | Cyclosporin<br>HIV infection | Catatani et al. [48], Schulz et al. [46] | HHV-8 | Chang et al. [49] |
| Non-melanoma skin cancers | Azathioprine<br>Cyclosporin<br>Prednisone and azathioprine<br>Tacrolimus<br>Infliximab (anti-TNFα)<br>Efalizumab (anti-CD11a)<br>HIV infection | Moloney et al. [50], Tiu et al. [52], Fortina et al. [53]<br>Euvrard et al. [55], Esser et al. [56]<br>Leonardi et al. [57], Schulz et al. [46] | Human papilloma viruses (HPV)<br>Ultraviolet (UV) radiation | Purdie et al. [51]<br>Madan et al. [54] |
| Hepatocarcinoma | Azathioprine | Vivarelli et al. [58] | Hepatitis virus B and C | Robinson [59], Koike et al. [60] |

**TABLE 2 Carcinogenic mechanisms for the neoplasms relevant to IMBPs**

| Carcinogen | Putative Mechanism of Carcinogenesis | Reference |
|---|---|---|
| Epstein–Barr virus (EBV) | Expression of EBV nuclear antigens 1–6 | Li and Minarovits [61] |
| Human T cell leukemia virus-I (HTLV-I) | p40 Tax-mediated transcription of viral promoter, deregulation of cell cycle and genomic instability | Mahieux and Gessain [62] |
| Human herpes virus 8 (HHV-8) | Activation of oncogenes and expression of a virally encoded VEGF receptor | Flaitz and Hicks [63] |
| Human papilloma viruses | Genomic instability due to E6/E7 oncogenes | Snijders et al. [64] |
| Hepatitis virus B and C | Activation of oncogenes and suppression of growth control | Feitelson [65] |
| Ultraviolet radiation | DNA damage | Bachelor and Bowden [17] |

process is at play. Because these processes are sufficient to cause neoplasia and so tumors can occur without immunosuppression, it is evident that immunosuppression, per se, is neither necessary nor sufficient to be carcinogenic. Were this not so, immunosuppression would be expected to be associated with a wider variety of tumor cell types and a wider variety of sites of origin. Moreover, if immunosuppression or imune modulation were carcinogenic, one would expect a higher incidence of tumors. For example, some of the constituents of tobacco smoke are well-recognized carcinogens. In a recent study of 8622 smokers of more than 30 cigarettes a day, the incidence of lung cancer in was 3.4% [6]. In contrast, the incidence of squamous cell cancer of the head and neck in a series of 1515 liver transplant patients receiving potent immunosuppressive treatment was only 0.86% [7]. Similarly the incidence of nonmelanoma skin cancer in a series of 15,789 rheumatoid arthritics receiving prednisone or TNF inhibitors was only 0.02% [8].

Finally, several of the diseases likely to be treated with IMBPs are associated with an increased incidence of the tumor types whose incidence is increased by immunosuppression even in the absence of immunosuppression. For example, studies of arthritis [9] and polymyalgia rheumatica/giant cell arteritis [10] found an increased incidence of lymphoma over population controls but failed to find evidence of an increased risk with immunosuppressive therapy. Taken together, these findings suggest that IMBPs are unlikely to be complete carcinogens, so we need to look to other steps in carcinogenesis where they may act.

## 4 IMMUNOMODULATORY BIOPHARMACEUTICALS ARE UNLIKELY TO BE CONVENTIONAL TUMOR PROMOTERS

Carcinogenesis can be considered in three steps: initiation, promotion, and progression [11]. In this model, initiation is believed to be the result of heritable changes (mutations) in the genome of a cell. Initiators (or their metabolites) must gain access to the genome and through chemical or radiochemical interactions with DNA result in mutations. This is exemplified by the classic "complete" carcinogens, such as 3-methylcholanthrene, and it is also probably the case with azathioprine as this potent immunosuppressant has been shown to be genotoxic [12]. Because IMBPs are non-DNA reactive proteins of generally greater than 30,000 kDa molecular weight, there is no evidence that they can gain access to the genome to interact with DNA. Moreover the metabolites of IMBPs are oligopeptides and ordinary amino acids, and they are thus very unlikely to be initiators. In sum, it is unlikely that IMBP act at tumor initiators.

In the process of neoplastic transformation, it is considered unlikely that a single mutation can result in the multiplicity of phenotypic changes associated with malignancy [13,14]. For full malignant transformation to occur, a cell must survive in its native milieu, accumulate a number of mutations, and expand its numbers as a nascent neoplastic clone. At this stage the nascent clone can be considered neoplastic, but it is not yet malignant. For a malignancy to occur, the nascent clone must continue to accumulate transformed phenotypes so as to include the ability to invade, disseminate, and escape destruction outside its native milieu (i.e., in the interstitial space, the lymphatic system, or the blood stream). Collectively these steps can be considered as tumor promoting.

In the classic model of initiation and promotion, conventional tumor promoters (CTP) inhibit the intracellular mechanisms that can eliminate the nascent clone, facilitate accumulation of the mutations necessary for full malignant transformation, or disrupt intercellular signaling [15]. A classic example of a CTP is 12-*O*-tetradecanoylphorbol-13-acetate (TPA, also referred to as phorbol myristate acetate, PMA). The mechanisms proposed for CTP that are relevant for the tumor types associated with immunosuppression, such as nonmelanoma skin tumors and lymphomas, are listed in Table 3. The table is limited to mechanisms of promotion and does not include tumor initiation mechanisms, namely those that directly or indirectly damage DNA (e.g., free radicals mediated by metabolism of ethanol) [16] or are mechanisms of neovasularization (e.g., angiogenesis in response to UV-A and B) [17]. The table also excludes mechanisms of CTP that have only been studied in the context of irrelevant tumors, such as TCDD in hepatocarcinogenesis [18].

As shown in Table 3, the mechanisms for CTP believed to be relevant in skin tumors and lymphomas act on the initiated cell, and notably, there are five mediated within the cytoplasm. These mechanisms are unlikely to be relevant to IMBPs because, as proteins, IMBPs (or their oligopeptide metabolites) cannot gain access to DNA in the nucleus. The remaining putative mechanisms of CTP—disruption of gap junctions, activation of protein kinase

**TABLE 3 Mechanisms of conventional tumor promotion in neoplasms relevant to IMBPs**

| Putative Mechanism of Promotion | Tumor Type | Site of Action | Reference |
|---|---|---|---|
| Inhibition of protein phosphatases | Skin tumors | Cytoplasm | Slaga et al. [15] |
| Hypermethylation of tumor suppressor genes | Skin and lymphoma | Cytoplasm | Van Doorn et al. [66], Pini et al. [67] |
| Loss of pRb tumor suppressor activity | Skin and lymphoma | Cytoplasm | Flaitz and Hicks [63] |
| Loss of p53 tumor suppressor activity | Lymphoma | Cytoplasm | Fesus et al. [68] |
| Hypomethylation of genomic DNA | Skin and lymphoma | Cytoplasm | Bachman et al. [69], Pini et al. [67] |
| Loss of gap junction mediated intercellular communication | Skin tumors | Cytoplasm or cell surface | Trosko and Tai [11] |
| Activation of protein kinase C | Skin tumors | Cytoplasm or cell surface | Slaga et al. [15] |
| Hyperplasia secondary to inflammation | Skin tumors | Cytoplasm or cell surface | Marks et al. [70], Bachelor and Bowden [17] |

C, and inflammation—can be mediated from the cell surface. However, as highly selective downregulators of the immune function, IMBPs are unlikely to mediate tumor promotion via these mechanisms.

## 5 IMMUNOMODULATORY BIOPHARMACEUTICALS AS ATYPICAL TUMOR PROMOTERS

In an alternate form of tumor promotion, referred to as atypical tumor promoters (ATPs), elimination of the nascent clones may be blocked by inhibiting apoptosis [19], by failure to eliminate nascent transformed cells, or by failure to control an active infection (or activation of a latent viral infection [20]) of a carcinogenic virus. Host defense mechanisms are primarily responsible for regulating extrinsic apoptosis, cytotoxicity, and control of viral infections. If an IMBP were to mediate tumor promotion, it would be acting through such a mechanism.

There are a number of host defense effector mechanisms, both humoral and cellular, that, when inhibited, have the potential to mediate atypical tumor promotion. These are listed in Table 4. As shown in the table, essentially all the effector cells of the immune system, with the exception of granulocytes, have the potential to eliminate nascent transformed cells or influence the

**TABLE 4 Host defense effectors where inhibition by biopharmaceutical may mediate atypical tumor promotion**

| Defense Mechanism | Putative Mechanism of Atypical Tumor Promotion | Reference |
|---|---|---|
| Natural killer cells | Failure to control virus infection. Failure to eliminate nascent transformed cells and clones in situ. Failure to eliminate cells in lymphatics and in blood stream | Barao and Ascensao [71] |
| Macrophages | Failure to control virus infection. Failure to eliminate nascent transformed cells and clones in situ. Failure to eliminate nascent colonization. | Salek-Ardakani et al. [72], Andreesen et al. [73] |
| B lymphocytes | Failure to control virus infection. Failure to present tumor or viral-antigens to T cells. | Capello et al. [74] Guinana et al. [75] |
| Dendritic cells | Failure to control virus infection. Failure to present tumor or viral antigens to T cells. | Lund et al. [76], Qu et al. [77] |
| T lymphocytes | Failure to control virus infection. Failure to eliminate virally infected cells. Failure to eliminate nascent transformed cells and clones in situ. Failure to eliminate cells in lymphatics. Failure to eliminate nascent colonization. | Maini et al. [78], Yu and Fu [79] |
| Lymphokines | Failure to control virus infection. Failure to mediate extrinsic apoptosis. Failure to mediate cytotoxicity. | See Table 5 |

course of viral infections. Thus IMBPs directed toward the effector cells have the potential to act as ATPs. The role of granulocytes in eliminating tumor cells is more controversial. Although there is some work that suggests that granulocytes can kill tumor cells [21], there is also work that suggests that granulocytes (via production of free radicals) can act as tumor initiators or promoters [22].

Cytokines and interleukins are proteins produced by a variety of immune and non-immune cells that bind receptors on their target cells and elicit a variety of responses. As of 2008, 35 interleukins have been described [23]. The principle source of interleukins are immune cells (macrophages and lymphocytes), but they are also produced by a number of immune cells, e.g., Interleukin-6 (IL-6) in a variety of cell types [24], and they are sometimes referred to by the more general term cytokines. Cytokines, because of their ability to affect the activity of a large number of cells and cell types, can amplify an

**TABLE 5 Lymphokines/cytokines with antitumor activity whose inhibition may mediate atypical tumor promotion**

| Cytokine | Reference |
|---|---|
| Tumor necrosis factor-α | Hori et al. [80] |
| Defensins | Kagan et al. [81] |
| Interferons | Brandacher et al. [82]<br>Rees [83]<br>Chada et al. [84] |
| CXC and C-C chemokinesIP-10, MCP-3, MIG and SDF-1α) | Chada et al. [84] |
| IL-1α and β | Veltri and Smith [85] |
| IL-2 | Kuhn and Dou [26] |
| IL-3 | Hansen [86] |
| IL-4 | Maini et al. [87] |
| IL-6 | Maini et al. [87]<br>Kurebayashi [88] |
| IL-7 | Appasamy [89] |
| IL-12 | Chada et al. [84] |
| IL18 | Pages et al. [90] |
| IL-21 | Roda et al. [91] |
| IL-23 | Oniki et al. [92] |
| IL-24 | Chada et al. [84] |
| IL-27 | Oniki et al. [92] |
| IL-28 α and β | Zitzmann et al. [25] |
| IL-29 | Zitzmann et al. [25] |
| IL-32 | Goda et al. [93] |

immune response, and thus they may play a central role in host defense against neoplasia. Antitumor activity has been ascribed to a majority of cytokines (Table 5). Antitumor activity can be either direct (e.g., IL-29 inducing apoptosis of tumor cells [25]) or indirect (e.g., by activating cytotoxic immune cells or by IL-2 activating cytotoxic T cells [26]). Thus IMBPs directed toward cytokines by anticytokine monoclonal antibodies or soluble receptors have the potential to act as ATPs.

## 6 IMMUNOMODULATORY BIOPHARMACEUTICALS AS POTENTIAL TUMOR PROGRESSORS

To be clinically significant, the nascent transformed clone(s) must not only survive but grow locally (so as to damage or destroy the adjacent normal tissue) or disseminate and grow at distant sites as metastases. These steps can be considered tumor progression [11]. At the cellular level the mechanisms by which host defense can inhibit the ability of a malignant clone to grow locally or colonize a distant site have already been described in Table 4. An additional step in tumor progression is the ability to call up a new

**TABLE 6 Host defense effector mechanisms whose inhibition may inhibit tumor promotion and/or progression**

| Antibodies against Host Defense Effector | Potential Effect of Inhibition on Tumor Progressor Activity | Reference |
|---|---|---|
| Anti-TNFα | Down regulation of tumor-promoting genes | Devoogdt et al. [94] |
| Anti-IL-4 | Shift Th1/Th2 balance favoring tumor immunity | Becker [95] |
| Anti-IL-6 | Shift Th1/Th2 balance favoring tumor immunity | Becker [95] |
| Anti-IL-8 | Inhibition of angiogenesis | Melnikova and Bar-Eli [96] |
| Anti-IL-10 | Shift Th1/Th2 balance favoring tumor immunity | Frumento et al. [97] |
| Anti-IL-26 | Inhibition of activation of STAT 1 and 3 in tumor cells | Hor et al. [98] |
| Anti-IL-31 | Inhibition of Jak1, Jak2, STAT1, –3, –5 Pi3 signaling pathways in tumor cells | Diveu et al. [99] |
| Anti-regulatory T cells (e.g., anti-CD25) | Shift Th1/Th2 balance favoring tumor immunity | Frumento et al. [97] |
| Anti-cytotoxic T cell antigen-4 (CTLA-4) antibodies | Rejection of established tumors and enhanced tumor immunity in mice | Leach et al. [100] |
| Anti-B cell antibodies | Decrease in tumor-promoting IgGs | Barbera-Guillem et al. [101] |

blood supply (angiogenesis). Without angiogenesis, although a neoplastic clone may survive and even expand as a diffuse cellular infiltrate, metastasis as discrete space occupying or tissue destructive lesions cannot occur [27]. In this context host defense mechanisms can be a double-edged sword, and in contrast to their conventional role in inhibiting tumor growth, some can act as tumor progressors. In this case inhibition of host defense may have a paradoxical beneficial effect. These are listed in Table 6.

## 7 CRITICAL REVIEW OF MODEL SYSTEMS FOR IDENTIFICATION AND CHARACTERIZATION OF ATYPICAL PROMOTERS AND TUMOR PROGRESSORS

As mentioned above, the mechanisms by which IMBPs are likely to act as atypical tumor promoters are all centered on their desired, albeit exaggerated,

pharmacology. Put another way, if IMBPs are acting as tumor promoters, it is likely due to a "toxic" effect on the immune system, meaning immunotoxicity. There are numerous approaches to the study of immunotoxicity, but it is beyond the scope of this chapter to discuss the pros and cons of the various purely mechanistic approaches to the study of IMBP as putative tumor promoters. The reader is therefore directed to Chapter 16, which reviews immunotoxicity for treatment of in vitro model systems that can be used to identify and characterize the specific pharmacology of IMBPs that can lead to tumor promotion. Here the discussion will focus on systems that directly address tumor promotion.

### 7.1 In vitro Mechanism-Based Systems

As shown in Table 7, a variety of in vitro systems have been used to study both CTP and ATP (in some cases the target of an IMBP has been studied and not the IMBP). One of the principal advantages for all these systems is that they can utilize human cells. (It should be noted that not all the systems cited in Table 7 use human cells, but likely depend on the "equivalent" nonhuman systems.) It is widely accepted (and codified in ICH6) that to have scientific relevance and meaning, the test article (in this case the IMBP) must be pharmacologically active in the test system. As reviewed by Green and Black [1], most IMBPs directed toward human targets are inactive in rodent systems. For this reason IMBPs are tested in vivo almost exclusively in nonhuman primates. Thus the IMBPs in development can be tested in the systems like those listed in Table 7, and there is no need for a homologous (also sometimes referred to "analogous" or "surrogate") IMBPs. (The issues with surrogate systems will be discussed below.)

The principal disadvantage of the systems listed in Table 7 is due to their very sophistication: they test for effects on a specific gene or phenotype, and generally do not test multiple mechanisms simultaneously. This concern could be partially mitigated by conducting a battery of such tests. However, because of the multiplicity of mechanisms involved in tumor promotion and progression, it is unlikely that an adequate battery of tests would be practical to design and validate to fully define the potential hazard posed by a specific IMBP. Another disadvantage inherent to the types of tests listed in Table 7 is the difficulty in translating the hazard identified by in vitro findings into a risk assessment for patients receiving therapeutic doses of an IMBP. No conceivable in vitro system could ever replicate the complexity of the intact immune network or the influences of pharmacokinetics. Thus, while in vitro approaches can identify and partially characterize the hazard posed by IMBPs, they have limited value in the final risk assessment for patients.

### 7.2 In vivo Systems

A number of in vivo systems have been used to evaluate CTP and ATP. These are listed in Tables 8, 9, and 10. The lists, although largely

**TABLE 7 In vitro model systems for identification and characterization of atypical tumor promoters and tumor progression**

| In vitro Systems | Example for Conventional Tumor Promoters | Reference | Example for Atypical Tumor Promoters or Their Targets | Reference |
|---|---|---|---|---|
| Expression of specific tumor promoter genes | Enhanced expression of secretory leukocyte protease inhibitor by TPA | Schlingemann et al. [102] | Enhanced expression of secretory leukocyte protease inhibitor by TNFα | Devoogdt et al. [94] |
| Expression of viral tumor promoter genes | Enhanced expression of Epstein–Barr early antigen | Ohmori et al. [103] | Upregulation of HPV oncoproteins by a variety of cytokines | Lembo et al. [104] |
| Adhesion | Inhibition of integrin-mediated adhesion | Chang and Chambers [105] | Antitumor antibodies inhibit adhesion | Blumenthal et al. [106] |
| Invasion | Enhanced invasion of model extracellular matrix | Fridman et al. [107] | Antitumor antibodies inhibit invasiveness | Blumenthal et al. [106] |
| Cell–ell communication | Inhibition of gap junction communication | Noguchi et al. [108] | Inhibition of gap junction communication by IL-1 | Hu and Xie [109] |
| Intracellular signaling | Increased protein kinase C activity or decreased protein phosphatase activity | Slaga et al. [15] | Anti-CD3 activates mitogen-activated kinase Erk2 | Von Willebrand et al. [110] |
| Apoptosis | Inhibition of apoptosis | Fesus et al. [68] | Anti-IL-2 antibodies inhibit Fas mediated apoptosis | Algeciras–Schimnich [111] |

complete in terms of model systems, are only exemplary and are not intended to be an exhaustive review of the entire literature on CTP or ATP in such systems.

Listed in Table 8 are what can be termed mechanistic approaches. Genetic defect and knockout mice where the gene or gene product has been shown to be involved in carcinogenesis (e.g., p53 heterozygous knockout mice) have been used rather extensively with complete carcinogens and some CTP, and these systems show some promise for replacing the standard two-year bioassay in mice [28]. Use of mice has, however, received much less attention for tumor promoters, and we have found no examples for use of mice genetically deficient in a gene shown to be involved in carcinogenesis with IMBP. In part, this observation likely reflects the lack of cross-reactivity of most IMBPs with murine targets due to species specificity.

In another form, the gene defect or knockout is directed toward immune cells or cytokines that are the targets for IMBPs. What are now classic examples are CD4 deficient "nude" mice, and severe combined immunodeficient SCID mice [29]. These systems have been shown to reveal the roles that various cells and cytokines may play in neoplasia and are thus proving to be an invaluable research tool. For hazard identification for IMBP, such systems can provide an important "proof of concept" that the target of the IMBP may play a role in emergence of spontaneous neoplasms. However, because there is no treatment per se, there can be no "dose response," and because the deficit is present throughout development and life, its role in risk assessment is currently limited. In the future, as technology advances and conditional knockouts where the gene product is downregulated in a controlled fashion become available, this situation will likely change. Although phenotype issues in KO mice may suggest potential toxicology issues, they often don't predict in vivo toxicity. In addition to the gene of interest, the phenotype of KO mice is influenced by traits that are unique to the background strain and by unpredictable interactions arising from strain-specific traits and elimination of the gene of interest that can lead to unexpected results. Even when high doses of an anti-cytokine monoclonal antibody (mAb) are administered, drug properties of disposition, pharmacodynamics, and pharmacokinetics are essential determinants of toxicity that cannot be evaluated in KO mice.

The remaining models listed in Table 8 are not measures of tumorigenesis but rather measures of host defense against neoplasia and/or tumor progression. The principal advantage of these systems is that they are in vivo models and, if they utilize murine cells, can be run in intact mice. The principal disadvantages are similar to those described for the in vitro systems listed in Table 7. Additionally, and of great importance, these systems are rodent based and no nonhuman primate form of the assay has been defined. Because most IMBPs are human or primate specific, most murine models cannot be run in a meaningful fashion with the actual IMBP, and a murine homologue (analogue) IMBP must be used. As will be discuss below, the reliance on surrogate IMBPs limits the usefulness of such systems.

**TABLE 8 Critical review of mechanism based in vivo model systems for identification and characterization of atypical tumor promoters and tumor progression**

| Model System | Example for Conventional Tumor Promoters | Reference | Example for Atypical Tumor Promoters | Reference | Advantages | Disadvantages |
|---|---|---|---|---|---|---|
| Genetic defects and Knockout systems | Estradiol increased the incidence of ethylnitrosourea induced uterine tumors in p53 knockout mice. | Gray et al. [112] | Interferon-γ deficient mice are more susceptible to methylcholanthrene-induced tumorigenesis. | Blankenstein and Qin [113] | May be already available and well studied.<br>Can be created for most cytokines or effector cells.<br>May provide strong proof of concept. | May alter expression of "bystander" genes.<br>Compensatory effects may mask deficit.<br>"All or nothing" effect may overestimate hazard. |
| Experimental metastasis | Exposure to B16F1 cells to PMA in vitro and subsequent IV injection increases lung metastases. | La Porta and Comolli [114] | Metastases increased with B16 melanoma in human CD4 transgenic mice treated with anti-asialoGM1, anti-thy-1, or anti-CD4. | Herzyk et al. [31] | Test system is readily available.<br>Database is limited. | Usually requires murine surrogate. |
| Immunization/ rejection | Treatment of tumor-bearing mice with PMA and subsequent transplantation results in tumor growth in immunized mice. | Kadhim et al. [115] | Anti-IL-12 (but not anti-TNFα) antibodies inhibits secondary cytotoxic lymphocyte response to β-gal transfected ES cells. | Mahnke and Schirrmacher [116] | Test system is readily available.<br>Database is limited. | Usually requires murine surrogate. |

Occasionally, however, the actual IMBP will be active in a murine system. Abatacept, a soluble fusion protein that consists of the extracellular domain of human cytotoxic T lymphocyte-associated antigen 4 (CTLA-4) linked to the modified Fc (hinge, CH2, and CH3 domains) portion of human immunoglobulin G1 (IgG1), is a unique example of a IMBP that was evaluated in a standard two-year carcinogenicity studies [30]. Unlike most IMBPs, abatacept is pharmacologically active in mice. In a mouse carcinogenicity study, following weekly subcutaneous injections of 20, 65, or 200 mg/kg of abatacept for up to 84 and 88 weeks in males and females, abatacept was associated with increases in the incidence of malignant lymphomas (at all doses) and mammary gland tumors in females at doses ≥65 mg/kg. Importantly, the mice from this study were infected with murine leukemia and mouse mammary tumor viruses, which increase incidence of lymphomas and mammary gland tumors, respectively, in immunosuppressed mice. Although in the cumulative abatacept clinical trials the rate observed for lymphoma is approximately 3.5-fold higher than expected in an age- and gender-matched general population, the RA population is known to have overall a higher incidence of lymphoma than the general population. Thus the relevance of the carcinogenicity findings to the clinical observation of increased lymphoma in patients receiving abatacept is unknown.

An other notable example is the work described by Herzyk et al. [31] where the effects of a Primatized® antihuman CD4 monoclonal antibody on experimental metastases with B16 melanoma cells was studied, and an increase in the number of lung colonies was found. Enabling this work, a murine CD4 knockout mouse reconstituted with human CD4 had been described in the literature and was available for license by the sponsor [32]. Moreover, in these mice, murine CD4 was faithfully replaced on T cells by human CD4, and the human protein mediated its physiologic function as an accessory binding protein in cellular and humoral immunity.

The work by Herzyk et al. [31], however, appears to be unique. To date, there have been no other reports of a human knock-in system being used in risk assessment of neoplasia, and similarly the literature on the use of tumor immunization/rejection models to evaluate the effects on IMBP on neoplasia is very scant. Because of the paucity of data, and because the anti-CD4 antibody studied by Herzyk et al. was never marketed and thus never used widely in patients, it is currently impossible to judge the predictive power of these systems for an increased risk of neoplasia in patients receiving IMBPs.

If a homologue IMBP must be used, a number of issues need to be addressed prior to embarking on studies to support registration. Some of these are listed in Table 9. And, while most of these issues will not be insurmountable, addressing them will not be trivial and will take significant time and effort. Moreover in some cases it may not be practical to conduct studies with homologue IMBP that comply with good laboratory practices. Given these issues, and the limited relevance of data from homologous systems to patients receiving

**TABLE 9 Potential issues with the applicability of homologous IMBPs**

| Attribute | Potential Issue |
|---|---|
| Role of target in immune function | Therapeutic target may not subserve the same function in rodents and humans. |
| Immunogenicity | Rodent homologue may be immunogenic in rodents. |
| Efficacy | Epitope bound by the rodent homologue mAb may not express the same functional activity as the human IMBP. |
| Fc functionality | Rodent homologue may not express the same Fc functionality as the human IMBP. |
| Stability | Rodent homologue may not have sufficient stability to allow testing. |
| Formulation | May be impossible to formulate the rodent homologue at a high enough concentration to allow testing. |
| Purity | May be impractical to produce the homologue with sufficient purity to allow testing. |
| Influence of expression system | Glycosylation pattern of the homologue may not reflect that of the human IMBP influencing exposure. |

therapeutic doses of an IMBP, the value of homologous systems is questionable.

Listed in Table 10 are the standard approaches to evaluating the carcinogenic potential of chemicals. As shown in the table, there are examples of the use of these systems for CTP and cyclosporin and, in the case of the two-year bioassay, an IMBP (abatacept, a cytotoxic lymphocyte antigen-4 (CTLA-4)-Ig fusion protein). Also as shown in the table, each model system has advantages and disadvantages. The primary drawback shared by these systems is that they rely on what can be referred to as "adventitious" or "environmental" initiation; that is, unlike the neoplasms clinically relevant to IMBPs where a viral or radiation initiator is likely, there is no defined tumor initiator. Issues with sometimes ill-defined background rates of tumorigenesis, irrelevant tumor sites, and strain and gender differences also pose very serious problems in experimental design and interpretation. In addition, as described by Hastings [33], rodents typically have been especially responsive to the pharmacological/toxicological effects of immunosuppressants, making it difficult to conduct lifetime bioassays at doses reasonably equivalent to those that would be used clinically. Taken together, these issues make false negative findings likely (i.e., by failure to show that an IMBP is an ATP). When coupled with the specific disadvantages listed in Table 10, the scientific rationale for the transgenic systems, neonatal mouse and two-year bioassay is quite poor.

Listed in Table 11 are a number of co-carcinogenesis models that have been used with CTP and in some cases APT. As with the models listed in

**TABLE 10 Critical review of in vivo model systems relying on serendipitous initiation for identification and characterization of atypical tumor promoters and tumor progression**

| Model System | Example for Conventional Tumor Promoters | Reference | Example for Atypical Tumor Promoters | Reference | Advantages | Disadvantages |
|---|---|---|---|---|---|---|
| Transgenic systems | RasH2 mice show increased tumors with a variety of nongenotoxic carcinogens | Yamamoto et al. [117] | RasH2 mice show increased tumors with cyclosporin | Yamamoto et al. [117] | Some test systems available but expensive<br>Custom systems very expensive and can take years to obtain in adequate numbers<br>Limited but appreciable data on chemical carcinogens and promoters | Usually requires murine surrogate<br>May not be susceptible to relevant tumor type (e.g., lymphoma)<br>Dependent on specific mechanism(s)<br>Limited background tumor database makes ensuring statistical power problematic<br>Fail to detect immunosuppressive drugs |

**TABLE 10** ***Continued***

| Model System | Example for Conventional Tumor Promoters | Reference | Example for Atypical Tumor Promoters | Reference | Advantages | Disadvantages |
|---|---|---|---|---|---|---|
| Neonatal mouse | TPA increases skin tumors in two-stage model | Furstenberger [118] | Cyclosporin failed to increase tumors | McClain et al. [119] | Test system readily available<br>Not dependent on specific mechanism<br>Dose-dependent effects | Usually requires murine surrogate<br>Neither designed nor powered to detect promoters<br>Limited background tumor database makes ensuring statistical power problematic<br>Immunogenicity of test article (even murine monoclonals) may confound assay |
| Two-year bioassay | Tumor promoter chlorophene increased tumors in mice but not rats | Yamarik [120] | Increased lymphoid neoplasms with cyclosporin<br>Abatacept (CTLA-4-Fc fusion protein) increased lymphomas in mice | Ryffel [121]<br>Anon. [30] | Test system is readily available<br>Large background database<br>Not dependent on specific mechanism<br>Dose-dependent effects | Usually requires murine surrogate<br>Neither designed nor powered to detect promoters<br>Immunogenicity of test article (even murine monoclonals) may confound assay |

**TABLE 11 Critical review of in vivo initiation-promotion model systems for identification and characterization of atypical tumor promoters and tumor progression**

| Model System | Example for Conventional Tumor Promoters | Reference | Example for Atypical Tumor Promoters | Reference | Advantages | Disadvantages |
|---|---|---|---|---|---|---|
| Chemical carcinogenesis | Promotion of chemically induced skin tumors by phorbolesters | Marks et al. [70] | Azathioprine shortened tumor latency period when administered prior to exposure to 3-methylcholanthrene | De Jong et al. [122] | Test system readily available<br>Large database on chemical carcinogens and promoters | Usually requires murine surrogate<br>May express irrelevant tumor types |
| Radiation carcinogenesis | Promotion of ionizing radiation initiation by phorbolester in mouse skin | Bowden et al. [123] | Cyclosporin increases incidence of lymphomas in irradiated mice | Yabu et al. [124] | Test system readily available<br>Limited database | Usually requires murine surrogate<br>May express irrelevant tumor types |

**TABLE 11** ***Continued***

| Model System | Example for Conventional Tumor Promoters | Reference | Example for Atypical Tumor Promoters | Reference | Advantages | Disadvantages |
|---|---|---|---|---|---|---|
| Photocarcinogenesis | Increase in skin tumors induced by low-dose UV by phorbolesters | Yamawaki et al. [125] | No examples found | | Test system readily available<br>Relies on relevant tumor initiator | Usually requires murine surrogate<br>May not express all relevant tumor types (e.g., lymphoma)<br>Limited database |
| Viral carcinogenesis | Cyclosporin increases incidence of lymphomas in murine herpes virus 68 infected mice | Sunil–Chandra et al. [39] | Abatacept (CTLA-4-Fc fusion protein) increased lymphomas in mice infected with murine leukemia virus | Anon. [30] | Test system readily available<br>Can be designed to rely on a relevant tumor initiator | Usually requires murine surrogate<br>May not express all relevant tumor types (e.g., skin tumors)<br>Murine virus may not be relevant to human (e.g., murine mammary tumor virus, MMTV)<br>Limited database |

Table 8 and 9, in most cases the human directed IMBP will not be active in a rodent test system and the use of a surrogate IMBP is required. As described above, this limits their relevance for most IMBPs. Moreover, although these systems may be likely to demonstrate the ATP activity of the homologous IMBP and thus demonstrating that a hazard exists, one is still faced with the difficulty in extrapolating to humans exposed to the human virus (or relevant dose of UV) and receiving therapeutic doses of the actual IMBP. Again, as with all rodent models, while hazard identification may be more or less straightforward, risk assessment is likely to be problematic. An interesting example of testing the actual IMBP in primates is the finding that alphacept (an LFA-3-Fc fusion protein) leads to reactivitation of Epstein–Barr virus infection in primates, which ultimately leads to lymphomas. However, like the models listed in Table 10, this was a serendipitous finding rather than the result of a deliberate experiment.

## 8 CLASSIFICATION OF THE HAZARD POSED BY IMBP AND RISK ASSESSMENT

In the preceding sections the case that IMBPs are associated with a limited range of neoplasms, that these neoplasms are by and large associated with a recognized tumor inititiator (i.e., viruses or ultraviolet radiation), that IMBPs are likely acting as atypical tumor promoters, rather than complete carcinogens, and that atypical tumor promotion is likely to be an inherent immunotoxicologic consequence of the intended pharmacologic activity of the IMBP has been laid out. What remains is how to evaluate the risk posed by IMBPs to patients and how that risk can be quantitated and communicated.

IMBPs intended for use in preventing allograft rejection in transplant patients or as disease modifiers in rheumatoid arthritis or psoriasis are likely to be administered chronically. Thus it is normally expected that sponsors would have conducted preclinical studies to determine the carcinogenic potential of candidate compounds. For small molecule pharmaceuticals, this would mean that rodent carcinogenicity bioassays would be performed under most circumstances. However, immunosuppressant drugs, in general, and IMBPs, in particular, present unique challenges with respect to the issue of carcinogenicity bioassays.

As described above, the pharmacological activity of IMBPs (even in the absence of genotoxic activity) may allow them to act as atypical tumor promoters. Thus, even in the absence of confirmatory standard rodent bioassay data, it can be assumed that this class of drug would pose an increased risk of neoplasia to patients receiving chronic dosing. As suggested by Hastings [33] and Cohen [34], rather than embarking on what would be a very expensive (in terms of both monetary and animal resources) and—given the nonvalidated nature of chronic testing of ATP and the probability of false negative or

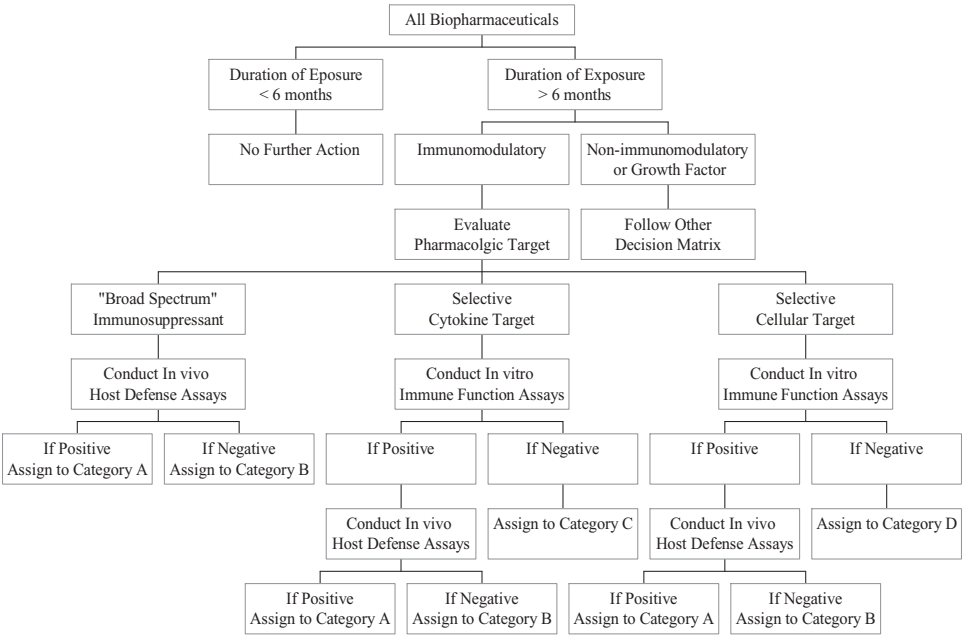


**Figure 1** Decision matrix for evaluation and classification of the hazard of IMBP acting as atypical tumor promoters.

spurious findings—a possibly fruitless standard approach, alternative models might be more appropriate for risk assessment with this class of drug.

One such alternative approach is to stipulate that IMBPs pose a hazard, then to characterize the hazard experimentally by classifying the nature of the hazard, and finally, to make a risk assessment based on that classification. A scheme for characterizing and classifying the hazard posed by IMBPs is shown in decision tree form in Figure 1. It should be noted that the purpose of classifying the risk posed by IMBPs is not to diminish recognition of the hazard. As reviewed by Melnick et al. [35] and Perera [36], some "nongenotoxic carcinogens" are in fact genotoxic, some nongenotoxic pharmacologic activities can result in oxidative damage to DNA, and for receptor mediated effects, a threshold tumor response is not "inevitable." Thus, a classification approach may not be meaningful for all agents. Rather, the purpose is to ensure that the hazard is recognized and that the relative risk can be communicated to the physicians prescribing and the patients receiving IMBPs.

As seen in Figure 1, the first step in characterizing the hazard posed by an IMBPs is to determine the duration of its intended use. For short-term exposures, namely less than six months, in keeping with long-established guidelines, no further action is required. The next step is to determine if, based on its known pharmacologic activity, the agent in question fits the definition of an

IMBP. If not, another decision matrix is required. This circumstance will not be discussed further in this chapter.

Having established that the agent is an IMBP that will be used for longer than six months, the next step is to evaluate the agent for suspicious structures or pharmacologic activity. This should include one or more in silico or "on paper" evaluations for known genotoxic or carcinogenic moieties [37], hormone-like effects [38], growth factor-like effects. If the agent has any of these characteristics, another decision matrix is then required. Again, this circumstance will not be discussed further in this chapter. However, for the vast majority of IMBPs, there will be no pharmacologic evidence of growth factor-like activity and no structural evidence of a genotoxic hazard. Thus there will be no scientific rationale for conducting genotoxicity testing on an IMBP.

The next step is to sort the IMBPs on the basis of their intended pharmacologic effects. If an IMBP is broadly active against multiple cell types or inhibits the activity of multiple cytokines, it can be considered a "broad spectrum" IMBP. Although this is the case with some pharmaceuticals such as cyclosporin, it will likely be a rare finding with an IMBP. Should this circumstance arise, it is likely that the agent will be active in in vitro immune function assays. And, rather than conduct in vitro assays whose outcome is almost forgone, the agent should be tested in in vivo host defense assays. These could include a viral challenge model (e.g., murine gamma herpes virus 68 as described by Sunil-Chandra et al. [39] or an experimental metastases model as described by Herzyk et al. [31]). Data generated from knockout systems may also be useful at this juncture. It must be noted, however, that the conduct of such models is not a trivial undertaking, and in most cases reliance will have to be placed on surrogates with their inherent limitations (see above). That being said, in vivo host defense models can play a pivotal role in the final classification of the hazard and the risk assessment for an IMBP and the outcome of the host defense assays can be used in classifying the IMBP as class A or B.

For IMBPs with more narrowly defined mechanisms of pharmacologic action, be they directed toward immune effector cells or cytokines, in vitro immune function assays may be a useful intermediate step. In many cases in vitro assays can be conducted with human cells, avoiding the need for a surrogate IMBP. Also, during the discovery phase for most IMBP some type of in vitro immune function assays will likely have been conducted. Moreover, for most drug targets a literature on the role of the target cell or cytokine in the immune function will likely exist. If in vitro data relevant to host defense already exists, and the IMBP or its target is known to modulate functions important in host defense, there is little rational for repeating in vitro assays, and one can go directly to in vivo host defense assays. It is expected that this would be the usual case, and depending on the outcome of the host defense assays, the IMBP would be assigned to class A or B.

Under more unusual circumstances an IMBP, when tested in a limited number of in vitro assays, would fail to show evidence that it can adversely affect host defense against neoplasia. These circumstances may pose a dilemma.

How much in vitro testing would be required to assign an IMBP to class C or D? There is likely no simple answer to this question. It may be answered on a case-by-case basis, as determined by the nature of the target for the IMBP and its intended clinical use. Alternatively, it may be more practical to go directly to in vivo host defense assays and assign the IMBP to class A or B.

Having classified the IMBP based on its effects on immune function, the next and final step would be to make a risk assessment and to communicate the risk to physicians and to the patients receiving therapeutic doses of the IMBP. This can be accomplished through the existing mechanisms of the Investigator's Brochure, Informed Consent documents, and ultimately, the product label.

## 9 CONCLUSIONS

In conclusion, IMBPs by virtue of their intended pharmacologic activity may act as atypical tumor promoters and thus pose a hazard for developing certain tumor types, especially lymphomas and skin cancer. Although conducting preclinical studies to identify this hazard is more or less straightforward, determining the relative risk to patients is and is likely to remain problematic. Traditional approaches that are used with pharmaceuticals are unlikely to be of much value. One alternative option may be to classify and label IMBPs based on their mechanism of action, for example, IMBPs that have an expected potential for increasing infection or lymphoproliferative disorders. This statement of the potential consequences would not be supported by lifetime animal data. One could also consider defining relative immunomodulation/immune suppression based on results of immunotoxicity assays. A product class label could later be defined based on a better understanding of the mechanism of action of the particular product and/or human data. In any event, the need for a prospective pharmacovigilance program to monitor carcinogenic risk is imperative for the clinical development program of IMBPs, including subsequent postmarketing commitments for products, especially if no animal studies can be performed.

## REFERENCES

1. Green JD, Black L. Overview. Status of preclinical safety assessment for immunomodulatory biopharmaceuticals. *Hum Exp Toxicol* 2000;19:208–12.
2. Cavagnaro JA, Spindler P. Methods for predicting tumorigenicity of immunomodulatory biopharmaceuticals. *Hum Exp Toxicol* 2000;19:213–5.
3. Dorant PM, Classen IJ, van Kreyl CF, Van Steenis G, Wester PW. Risk assessment on the carcinogenic potential of hybridoma cell DNA: implications for residual contaminating cellular DNA in biological products. *Biologicals* 1997;25:381–90.

4. Pham SM, Kormos RL, Landreneau RJ, Kawai A, Gonzalez-Cancel I, Hardesty RL, Hattler BG, Griffith BP. Solid tumors after heart transplantation: lethality of lung cancer. *Ann Thoracic Surg* 1995;60:1623–6.
5. Engels EA, Brock MV, Chen J, Hooker CM, Gillison M, Moore RD. Elevated incidence of lung cancer among HIV-infected individuals. *J Clin Oncol* 2006;24: 1383–8.
6. Haiman CA, Stram DO, Wilkens LR, Pike MC, Kolonel LN, Henderson BE, Le Marchand L. Ethnic and racial differences in the smoking-related risk of lung cancer. *N Engl J Med* 2006;354:333–42.
7. Scheifele C, Reichart PA, Hippleer-Benscheidt M, Neuhaus P, Neuhaus R. Incidence of oral, pharyngeal and laryngeal squamous cell carcinomas among 1515 patients after liver transplantation. *Oral Oncol* 2005;41:670–6.
8. Chakravarty EF, Michaud K, Wolfe F. Skin cancer, rheumatoid arthritis and tumor necrosis factor inhibitors. *J Rheumatol* 2005;32:2130–215.
9. Franklin J, Lunt M, Bunn D, Symmons D, Silman A. Incidence of lymphoma in a large primary care derived cohort of cases of inflammatory arthritis. *Ann Rheumat Dis* 2006;65:617–22.
10. Askling J, Klareskog L, Hjalgrim H, Baecklund E, Bjorkholm M, Ekbom A. Do steroids increase lymphoma risk? A case-control study of lymphoma risk in polymyalgia rheumatica/giant cell arteritis. *Ann Rheumat Dis* 2005;64:1765–8.
11. Trosko JE, Tai MH. Adult stem cell theory of the multi-stage, multi-mechanism theory of carcinogenesis: role of inflammation on the promotion of initiated stem cells. *Contribut Microbiol* 2006;13:45–65.
12. Smith CC, Archer GE, Forster EJ, Lambert TR, Rees RW, Lynch AM. Analysis of gene mutations and clastogenicity following short-term treatment with azathioprine in MutaMouse. *Environ Mol Mutagen* 1999;34:131–9.
13. Boukamp P. Non-melanoma skin cancer: What drives tumor development and progression? *Carcinogenesis* 2005;26:1657–67.
14. Pettit SJ, Seymour K, O'Flahaerty E, Kirby JA. Immune selection in neoplasia: towards a microevolutionary model of cancer development. *Brit J Cancer* 2000;82:1900–6.
15. Slaga TJ, DiGiovanni J, Weinberg LD, Budnova IV. Skin carcinogenesis; characteristics, mechanisms and prevention. *Prog Clin Biol Res* 1995;391:1–20.
16. Mufti SI. Alcohol acts to promote incidence of tumors. *Cancer Detect Prev* 1992;16:157–62.
17. Bachelor MA, Bowden GT. UVA-mediated activation of signaling pathways involved in skin tumor promotion and progression. *Sem Cancer Biol* 2004;14: 131–8.
18. Schrenk D, Schmits H-J, Bohnenberger S, Wagner B, Worner W. Tumor promoters as inhibitors of apoptosis in rat hepatocytes. *Toxicol Letts* 2004;149:43–50.
19. Peter ME, Legembre P, Barnhart BC. Does CD95 have tumor promoting activities? *Biochim Biophys Acta* 2005;1755:26–36.
20. van den Bosch CA. Is endemic Burkitt's lymphoma an alliance between three infections and a tumor promoter? *Lancet Oncol* 2004;5:738–6.

21. Di Carlo E, Forni G, Lollini P, Colombo MP, Modesti A, Musiani P. The intriguing role of polymorphonuclear neutrophils in antitumor reactions. *Blood* 2001;97: 339–45.
22. de Visser KE, Eichten A, Coussens LM. Paradoxical roles of the immune system during cancer development. *Nat Rev Cancer* 2006;6:24–37.
23. Collison LW, Workman CJ, Kuo TT, Boyd K, Wang Y, Vignali KM, Cross R, Sehy D, Blumberg RS, Vignali DA. The inhibitory cytokine IL-35 contributes to regulatory T-cell function. *Nature* 2007;450(7169):566–9.
24. Seghal PB. Interleukin-6: molecular pathophysiology. *J Investig Dermatol* 1990; 94(Suppl):2S–6S.
25. Zitzmann K, Brand S, Baehs S, Goke B, Meinecke J, Spottl G, Meyer H, Auernhammer CJ. Novel interferon-lambdas induce antiproliferative effects in neuroendocrine tumor cells. *Biochem Biophys Res Commun* 2006;344:1334–41.
26. Kuhn DJ, Dou QP. The role of interleukin-2 receptor alpha in cancer. *Frontiers Biosci* 2005;10:1462–74.
27. Herbst RS. Therapeutic options to target angiogenesis in human malignancies. *Expert Opin Emerg Drugs* 2006;11:635–50.
28. Mitsumori K. Evaluation on carcinogenicity of chemicals using transgenic mice. *Toxicology* 2002;181–2, 241–4.
29. Kelland LR. Of mice and men: values and liabilities of the athymic nude mouse model in anticancer drug development. *Eur J Cancer* 2004;40:827–36.
30. Anonamous. Orencia® (abatacept) package insert. Bristol-Myers Squibb Company, Princeton, NJ, 2007.
31. Herzyk DJ, Gore ER, Polsky R, Nadwodny KL, Maier CC, Liu S, Hart TK, Harmsen AG, Bugelski PJ. Immunomodulatory effects of anti-CD4 antibody in host resistance against infections and tumors in human CD4 transgenic mice. *Infect Immunity* 2001;69:1032–43.
32. Bugelski PJ, Herzyk DJ, Rehm S, Harmsen AG, Gore EV, Williams DM, Maleeff BE, Badger AM, Truneh A, O'Brien SR, Macia RA, Wier PJ, Morgan DG, Hart TK. Preclinical development of keliximab, a Primatized® anti-CD4 monoclonal antibody, in human CD4 transgenic mice: characterization of the model and safety studies. *Hum Exp Toxicol* 2000;19:230–43.
33. Hastings KL. Assessment of immunosuppressant drug carcinogenicity: standard and alternative models. *Hum Exp Toxicol* 2000;19:261–5.
34. Cohen S. Human carcinogenic risk evaluation: an alternative approach to the two-year bioassay. *Toxicol Sci* 2004;80:225–9.
35. Melnick RL, Kohn MC, Portier CJ. Implications for risk assessment of suggested nongenotoxic mechanisms of chemical carcinogenesis. *Environ Health Persp* 1996; 104(Suppl 1):123–34.
36. Perera FP. Perspectives on the risk assessment for nongenotoxic carcinogens and tumor promoters. *Environ Health Persp* 1991;94:231–5.
37. Matthews EJ, Kruhlak NL, Cimino MC, Benzm RD, Contrera JF. An analysis of genetic toxicity, reproductive and developmental toxicity, and carcinogenicity data: II. Identification of genotoxicants, reprotoxicants, and carcinogens using in silico methods. *Regul Toxicol Pharmacol* 2006;44:97–110.
38. Devillers J, Marchand-Geneste N, Carpy A, Porcher JM. SAR and QSAR modeling of endocrine disruptors. *SAR & QSAR Environ Res* 2006;17:393–412.

39. Sunil-Chandra NP, Arno J, Fazakerley J, Nash AA. Lymphoproliferative disease in mice infected with gammaherpesvirus 68. *Am J Pathol* 1994;145:818–26.
40. Mougel F, Dalle S, Balme B, Houot R, Thomas L. Aggressive CD30 large cell lymphoma after cyclosporine given for putative atopic dermatitis. *Dermatology* 2006;213:239–41.
41. Ahmed N, Heslop HE. Viral lymphomagenesis. *Curr Opin Hematol* 2006;13: 254–95.
42. Kwon JH, Farrell RJ. The risk of lymphoma in the treatment of inflammatory bowel disease with immunosuppressive agents. *Crit Rev Oncol Hematol* 2005;56:169–78.
43. Jarrett RF. Viruses and lymphoma/leukemia. *J Pathol* 2006;208:176–86.
44. Malnati MS, Dagna L, Ponzoni M, Lusso P. Human herpesvirus 8 (HHV-8/KSHV) and hematologic malignancies. *Rev Clin Exp Hematol* 2003;7:375–405.
45. Ibanez JP, Monteverde ML, Goldber J, Diaz MA, Turconi A. Sirolimus in pediatric renal transplantation. *Transplant Proc* 2005;37:682–4.
46. Schulz TF, Boshoff CH, Weiss RA. HIV infection and neoplasia. *Lancet* 1996; 348:587–91.
47. Martin-Gomez MA, Pena M, Cabello M, Burgos D, Gutierrez C, Sola E, Acedo C, Bailen A, Gonzalez-Molina M. Posttransplant lymphoproliferative disease: a series of 23 cases. *Transplant Proc* 2006;38:2448–50.
48. Cattani P, Capuano M, Graffeo R, Ricci R, Cerimele F, Cerimele D, Nanni G, Fadda G. Kaposi's sarcoma associated with previous human herpesvirus 8 infection in kidney transplant recipients. *J Clin Microbiol* 2001;39:506–8.
49. Chang Y, Cesarman E, Pessin MS, Lee F, Culpepper J, Knowles DM, Moore PS. Identification of herpesvirus like sequences in AIDS-associated Kaposi's sarcoma. *Science* 1994;266:1865–9.
50. Moloney FJ, Comber H, O'Lorcain P, O'Kelly P, Conlon PJ, Murphy GM. A population based study of skin cancer incidence and prevalence in renal transplant recipients. *Brit J Dermatol* 2006;154:498–504.
51. Purdie KJ, Surentheran T, Sterling JC, Bell L, McGregor JM, Proby CM, Harwood CA, Breuer J. Human papillomavirus gene expression in cutaneous squamous cell carcinomas from immunocompromised and immunocompetant individuals. *J Investig Dermatol* 2005;125:98–107.
52. Tiu J, Li H, Rassekh C, van der Sloot P, Kovach R, Zhang P. Molecular basis of posttransplant squamous cell carcinoma: the potential role of cyclosporine A in carcinogenesis. *Laryngoscope* 2006;116:762–9.
53. Fortina AB, Piaserico S, Caforio AL, Abeni D, Alaibac M, Angelini A, Iliceto S, Peserico A. Immunosuppressive level and other risk factors for basal cell carcinoma and squamous cell carcinoma in hear transplant recipients. *Arch Dermatol* 2004;140:1079–85.
54. Madan V, Hoban P, Strange RC, Fryer AA, Lear JT. Genetics and risk factors for basal cell carcinoma. *Brit J Dermatol* 2006;154(Suppl 1):5–7.
55. Euvrard S, Ulrich C, Lefrancois N. Immunosuppressants and skin cancer in transplant patients: focus on rapamycin. *Dermatol Surg* 2004;30:628–33.
56. Esser AC, Abrill A, Fayne S, Doyle JA. Acute development of multiple keratoacanthomas and squamous cell carcinomas after treatment with infliximab. *J Am Acad Dermatol* 2004;50(5 Suppl):S75–7.

57. Leonardi CL, Toth D, Cather JC, Langley RG, Werther W, Compton P, Kwon P, Wetherill G, Curtin F, Menter A. A review of malignancies observed during efalizumab (Raptiva) clinical trials for plaque psoriasis. *Dermatology* 2006;213:204–14.
58. Vivarelli M, Cucchetti A, Piscaglia F, La Barba G, Bolondi L, Cavallari A, Pinna AD. Analysis of risk factors for tumor recurrence after liver transplantation for hepatocellular carcinoma: key role of immunosuppression. *Liver Transplant* 2005;11:497–503.
59. Robinson WS. The role of hepatitis B virus in the development of primary hepatocellular carcinoma: Part I. *J Gastroenterol Hepatol* 1992;7:622–38.
60. Koike K, Moriya K, Kimura S. Role of hepatitis C virus in the development of hepatocellular carcinoma: transgenic approach to viral hepatocarcinogenesis. *J Gastroenterol Hepatol* 2002;17:394–400.
61. Li H, Minarovits J. Host-cell dependent expression of latent Epstein-Barr virus genomes: regulation by DNA methylation. *Adv Cancer Res* 2003;89:133–56.
62. Mahieux R, Gessain A. HTLV-1 associated adult T-cell leukemia/lymphoma. *Rev Clin Exp Hematol* 2003;7:336–61.
63. Flaitz CM, Hicks MJ. Molecular piracy: the link to viral carcinogenesis. *Oral Oncol* 1998;34:448–53.
64. Snjiders PJ, Steenbergen RD, Heideman DA, Meijer CJ. HPV-mediated cervical carcinogenesis: concepts and clinical implications. *J Pathol* 2006;208:152–64.
65. Feitelson MA. Parallel epigenetic and genetic changes in the pathogenesis of hepatitis virus-associated hepatocellular carcinoma. *Cancer Letts* 2006;239:10–20.
66. van Doorn R, Gruis NA, Willemze R, van der Velden PA, Tensen CP. Aberrant DNA methylation in cutaneous malignancies. *Sem Oncol* 2005;32:479–87.
67. Pini JT, Frachina M, Taylor JM, Kay PH. Evidence that general genomic hypomethylation and focal hypermethylation are two independent molecular events of non-Hodgkin's lymphoma. *Oncol Res* 2004;14:399–405.
68. Fesus L, Szondy Z, Uray I. Probing the molecular program of apoptosis by cancer chemopreventive agents. *J Cell Biochem* 1995;22(Suppl):151–61.
69. Bachman AN, Curtin GM, Doolittle DJ, Goodman JI. Altered methylation in gene-specific and GC-rich regions of DNA is progressive and nonrandom during promotion of skin tumorigenesis. *Toxicol Sci* 2006;1:406–18.
70. Marks F, Fürstenberger G, Heinzelmann T, Müller-Decker K. Mechanisms in tumor promotions: guidance for risk assessment and cancer chemoprevention. *Toxicol Letts* 1995;82/83:907–17.
71. Barao I, Ascensao JL. Human natural killer cells. *Arch Immunol Ther Exp* 1998; 46:213–29.
72. Salek-Ardakani S, Arrand JR, Mackett M. Epstein-Barr virus encoded interleukin-10 inhibits HLA-class I, ICAM-1, and B7 expression on human monocytes: implications for immune evasion by EBV. *Virology* 2002;304:342–51.
73. Andreesen R, Hennemann B, Krause SW. Adoptive immunotherapy of cancer using monocyte-derived macrophages: rationale, current status, and perspectives. *J Leukocyte Biol* 1998;64:419–26.
74. Capello D, Rossi D, Gaidano G. Post-transplant lymphoproliferative disorders: molecular basis of disease histogenesis and pathogenesis. *Hematol Oncol* 2005; 23:61–7.

75. Guinan EC, Gribben JG, Boussiotis VA, Freeman GJ, Nadler LM. Pivotal role of the B7:CD28 pathway in transplantation tolerance and tumor immunity. *Blood* 1994;84:3261–82.

76. Lund JM, Linehan MM, Iijima N, Iwasaki A. Cutting edge: plasmacytoid dendritic cells provide innate immune protection against mucosal viral infection in situ. *J Immunol* 2006;177:7510–4.

77. Qu M, Muller HK, Woods GM. Chemical carcinogens and antigens contribute to cutaneous tumor promotion by depleting epidermal Langerhans cells. *Carcinogenesis* 1997;18:1277–9.

78. Maini MK, Boni C, Ogg GS, King AS, Reignat S, Lee CK, Larrubia JR, Webster GJ, McMichael AJ, Ferrari C, Williams R, Vergani D, Bertoletti A. Direct ex vivo analysis of hepatitis B virus-specific CD8(+) T cells associated with control of infection. *Gastroenterology* 1999;117:1386–99.

79. Yu P, Fu YX. Tumor-infiltrating T lymphocytes: friends or foes? *Lab Investig* 2006;86:231–45.

80. Hori K, Ehrke MJ, Mace K, Maccubbin D, Doyle MJ, Otsuka Y, Mihich E. Effect of recombinant human tumor necrosis factor on the induction of murine macrophage tumoricidal activity. *Cancer Res* 1987;47:2793–8.

81. Kagan BL, Ganz T, Lehrer RI. Defensins: a family of antimicrobial and cytotoxic peptides. *Toxicology* 1994;87:131–49.

82. Brandacher G, Winkler C, Schroecksnadel K, Margreiter R, Fuchs D. Antitumoral activity of interferon-gamma involved in impaired immune function in cancer patients. *Curr Drug Metab* 2006;7:599–612.

83. Rees RC. MHC restricted and non-restricted killer lymphocytes. *Blood Rev* 1990; 4:204–10.

84. Chada S, Ramesh R, Mhashilkar AM. Cytokine- and chemokine-based gene therapy for cancer. *Curr Opin Mol Ther* 2003;5:463–74.

85. Veltri S, Smith JW Jr. Interleukin 1 trials in cancer patients: a review of the toxicity, antitumor and hematopoietic effects. *Stem Cells* 1996;14:164–76.

86. Hansen F. Hemopoietic growth and inhibitory factors in treatment of malignancies. A review. *Acta Oncol* 1995;34:453–68.

87. Maini A, Morse PD, Wang CY, Jones RF, Haas GP. New developments in the use of cytokines for cancer therapy. *Anticancer Res* 1997;17:3803–8.

88. Kurebayashi J. Regulation of interleukin-6 secretion from breast cancer cells and its clinical implications. *Breast Cancer* 2000;7:124–9.

89. Appasamy PM. Biological and clinical implications of interleukin-7 and lymphopoiesis. *Cytokines Cell Mol Ther* 1999;5:25–39.

90. Pages F, Berger A, Lebel-Binay S, Zinzindohoue F, Danel C, Piqueras B, Carriere O, Thiounn N, Cugnenc PH, Fridman WH. Proinflammatory and antitumor properties of interleukin-18 in the gastrointestinal tract. *Immunol Letts* 2000;75: 9–14.

91. Roda JM, Parihar R, Lehman A, Mani A, Tridandapani S, Carson WE 3rd. Interleukin-21 enhances NK cell activation in response to antibody-coated targets. *J Immunol* 2006;177:120–9.

92. Oniki S, Nagai H, Horikawa T, Furukawa J, Belladonna ML, Yoshimoto T, Hara I, Nishigori C. Interleukin-23 and interleukin-27 exert quite different antitumor and vaccine effects on poorly immunogenic melanoma. *Cancer Res* 2006;66: 6395–404.
93. Goda C, Kanaji T, Kanaji S, Tanaka G, Arima K, Ohno S, Izuhara K. Involvement of IL-32 in activation-induced cell death in T cells. *Intl Immunol* 2006;18:233–40.
94. Devoogdt N, Revets H, Kindt A, Liu YQ, De Baetselier P, Ghassabeh GH. The tumor-promoting effects of TNF-α involves the induction of secretory leukocyte protease inhibitor. *J Immunol* 2006;177:8046–52.
95. Becker Y. Molecular immunological approaches to biotherapy of human cancers—a review, hypothesis and implications. *Anticancer Res* 2006;26:1113–34.
96. Melnikova VO, Bar-Eli M. Bioimmunotherapy for melanoma using fully human antibodies targeting MCAM/MUC18 and IL-8. *Pigmented Cell Res* 2006;19: 395–405.
97. Frumento G, Piazza T, Di Carlo E, Ferrini S. Targeting tumor-related immunosuppression for cancer immunotherapy. *Endocrine, Metab Immune Disord Drug Targ* 2006;6:233–7.
98. Hor S, Pirzer H, Dumoutier L, Bauer F, Wittmann S, Sticht H, Renauld JC, de Waal Malefyt R, Fickenscher H. The T-cell lymphokine interleukin-26 targets epithelial cells through the interleukin-20 receptor 1 and interleukin-10 receptor 2 chains. *J Biol Chem* 2004;279:33343–51.
99. Diveu C, Lak-Hal AH, Froger J, Ravon E, Grimaud L, Barbier F, Hermann J, Gascan H, Chevalier S. Predominant expression of the long isoform of GP130-like (GPL) receptor is required for interleukin-31 signaling. *Eur Cytokine Net* 2004; 15:291–302.
100. Leach DR, Krummel MF, Allison JP. Enhancement of antitumor immunity by CTLA-4 blockade. *Science* 1996;271:1734–6.
101. Barbera-Guillem E, May KF Jr, Nyhus JK, Nelson MB. Promotion of tumor invasion by cooperation of granulocytes and macrophages activated by anti-tumor antibodies. *Neoplasia* 1999;1:453–60.
102. Schlingemann J, Hess J, Wrobel G, Breitenbach U, Gebhardt C, Steinlein P, Kramer H, Furstenberger G, Hahn M, Angel P, Lichter P. Profile of gene expression induced by the tumour promoter TPA in murine epithelial cells. *Intl J Cancer* 2003;104:699–708.
103. Ohmori K, Miyazaki K, Umeda M. Detection of tumor promoters by early antigen expression of EB virus in Raji cell using fluorescence miniplate-reader. *Cancer Letts* 1998;132:51–59.
104. Lembo D, Donalisio M, De Andrea M, Cornaglia M, Scutera S, Musso T, Landolfo S. A cell-based high-throughput assay for screening inhibitors of human papillomavirus-16 long control region activity. *FASEB J* 2006;20:148–50.
105. Chang PL, Chambers AF. Transforming JB6 cell exhibit enhanced integrin-mediated adhesion to osteopontin. *J Cell Biochem* 2000;78:8–23.
106. Blumenthal RD, Hansen HJ, Goldenberg DM. Inhibition of adhesion, invasion, and metastasis by antibodies targeting CEACAM6 (NCA-90) and CEACAM5 (Carcinoembryonic Antigen). *Cancer Res* 2005;65:8809–17.

107. Fridman R, Lacal JC, Reich R, Bonfil DR, Ahn CH. Differential effects of phorbol ester on the in vitro invasiveness of malignant and non-malignant human fibroblast cells. *J Cell Physiol* 1990;142:55–60.
108. Noguchi M, Nomata K, Watanabe J, Kanetake H, Saito Y. Changes in gap junction intercellular communication in renal tubular epithelial cell in vitro treated with renal carcinogens. *Cancer Letts* 1998;122:77–84.
109. Hu VW, Xie HQ. Interleukin-1 alpha suppresses gap junction-mediated intercellular communication in human endothelial cells. *Exp Cell Res* 1994;213: 218–23.
110. Von Willebrand M, Jascur T, Bonnefoy-Berard N, Yano H, Altman A, Matsuda Y, Mustelin T. Inhibition of phosphatidylinositol 3-kinase blocks T cell antigen receptor/CD3-induced activation of the mitogen-activated kinase Erk2. *Eur J Biochem* 1996;235:828–35.
111. Algeciras-Schimnich A, Griffith TS, Lynch DH, Paya CV. Cell cycle-dependent regulation of FLIP levels and susceptibility to Fas-mediated apoptosis. *J Immunol* 1999;162:5205–11.
112. Gray LE, Ostby J, Farrell J, Rehnberg G, Linder R, Cooper R, Goldman J, Slott V, Laskey J. A dose–response analysis of methoxychlor-induced alterations of reproductive development and function in the rat. *Fund Appl Toxicol* 1989; 12:92–108.
113. Blankenstein T, Qin Z. Chemical carcinogens as foreign bodies and some pitfalls regarding cancer immune surveillance. *Adv Cancer Res* 2003;90:179–207.
114. La Porta CA, Comolli R. Angiogenic capacity and lung-colonizing potential in vivo is increased in weakly metastatic B16F1 cells and decreased in highly metastatic BL6 cells by phorbol esters. *Clin Exp Metast* 1998;16:399–405.
115. Kadhim S, Burns BF, Birnboim HC. In vivo induction of tumor variants by phorbol 12-myristate 13-acetate. *Cancer Letts* 1987;38:209–14.
116. Mahnke YD, Schirrmacher V. Characteristics of a potent tumor vaccine-induced secondary anti-tumor T cell response. *Intl J Oncol* 2004;24:1427–34.
117. Yamamoto S, Urano K, Nomura T. Validation of transgenic mice harboring the human prototype c-Ha-ras gene as a bioassay model for rapid carcinogenicity testing. *Toxicology Letts* 1998;102–3:473–8.
118. Furstenberger G, Schweizer J, Marks F. Development of phorbol ester responsiveness in neonatal mouse epidermis: correlation between hyperplastic response and sensitivity to first-stage tumor promotion. *Carcinogenesis* 1985;6:289–94.
119. McClain RM, Keller D, Casciano D, Fu P, MacDonald J, Popp J, Sagartz J. Neonatal mouse model: review of methods and results. *Toxicol Pathol* 2001;29(Suppl): 128–37.
120. Yamarik TA. Safety assessment of dichlorophene and chlorophene. *Intl J Toxicol* 2004;23(Suppl 1):1–27.
121. Ryffel B. The carcinogenicity of ciclosporin. *Toxicology* 1992;73:1–22.
122. De Jong M, Coppee MC, De Halleux F, Deckers C, Maisin H. Immunosuppression and carcinogenesis: effects of azathioprine on induction of sarcomas by 3-methylcholanthrene. *Neoplasma* 1977;24:139–46.
123. Bowden GT, Jaffe D, Andrews K. Biological and molecular aspects of radiation carcinogenesis in mouse skin. *Rad Res* 1990;121:235–41.

124. Yabu K, Warty VS, Gorelik E, Shinozuka H. Cyclosporine promotes the induction of thymic lymphomas in C57BL/6 mice irradiated by a single dose of gamma-radiation. *Carcinogenesis* 1991;12:43–6.
125. Yamawaki M, Katiyar SK, Anderson CY, Tubesing KA, Mukhtar H, Elmets CA. Genetic variation in low-dose UV-induced suppression of contact hypersensitivity and in the skin photocarcinogenesis response. *J Investig Dermatol* 1997;109: 716–21.

# Strategy Considerations for Developing the Preclinical Safety Testing Programs for Protein Scaffold Therapeutics

STANLEY A. ROBERTS, PhD, DABT, GARY WOODNUTT, PhD, and
CURT W. BRADSHAW, PhD

**Contents**

1 Introduction
2 Safety Assessment Program Strategy Development
3 Choice of Animals
4 Duration and Types of Studies
5 Dosage Selection
6 Metabolism and Pharmacokinetics
7 Biomarkers and Indicators of Toxicity
8 Conclusions
References

## 1 INTRODUCTION

A number and variety of conjugated peptides/proteins are currently being developed as potential therapeutic agents. The objective of this chapter is to discuss some of the key issues that contribute to the design of the preclinical safety assessment program for an emerging group of biopharmaceuticals that we have designated as protein scaffold therapeutics. These new compounds are diverse and include several different protein scaffolds such as monoclonal antibodies, transferrin, and albumin. Small pharmacologically active molecules

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

(i.e., pharmacophores) are attached to the scaffolds so that these therapeutics can have many potential modes of intended pharmacologic action. The first-generation antibody-based scaffolds were immunotoxin conjugates directed toward cell surface receptors that acted as delivery systems for cytotoxic radionuclides (e.g., Bexxar® and Zevalin®) or chemical toxins (Mylotarg®). The antibody-directed specificity (e.g., CD33 for Mylotarg®, CD20 for Bexxar® and Zevalin®) allowed the cytotoxic activity to be preferentially directed toward the target cells. In the case of Mylotarg® the conjugated molecule is internalized via a receptor, after which the cytotoxin is released intracellularly to kill the targeted cell. Several of these conjugated therapeutics have proved to be clinically successful, with only relatively minor side effect profiles related to either the proteinaceous nature of the drug (e.g., antigenicity or hypersensitivity) or to the pharmacology of the conjugate (e.g., anemia, myelosuppression, tumor lysis syndrome). Collectively these data also suggest that administration of a protein conjugate per se does not increase the potential for systemic toxicity and in fact can provide significant improvement over the pharmacophore alone.

The targeted delivery of cytotoxins to neoplastic cells represented a significant step forward in improving the safety and efficacy of cancer therapy. Recently a new class of protein scaffold therapeutics has been developed where the role of the macromolecular scaffold (antibody, $F_c$, albumin, transferrin, etc.) is to improve the pharmacokinetics by providing protection for the active moiety (small molecule or peptide) from metabolic or urinary clearance. For these compounds the pharmacophore that is attached to the protein scaffold provides the desired pharmacologic target activity. The pharmacophore can either be produced as a fusion protein or it can be chemically linked to a specific site on the scaffold. An interesting feature of this latter technology is the specific stoichiometry for the macromolecule and pharmacophore combination, which allows for a precise characterization of the final drug product.

For this chapter we will specifically focus on protein scaffold therapeutic agents where the primary objective is to improve the pharmacokinetics of the pharmacophore. Other chapters in this book will specifically discuss immunotoxin conjugates [1]. The various protein scaffold therapeutics that are currently being advanced through late-stage discovery and/or clinical development are detailed in Table 1. As previously discussed, the biological activity of these agents is derived from the pharmacophore through which the pharmacological activity (e.g., inhibition or activation) is achieved. For example, GLP-1 (glucagon-like-peptide 1) analogues have been attached to albumin or transferrin and angiopoietin-2 binding peptides have been appended to $F_c$ domains of antibodies. Interestingly the pharmacological potency of the conjugated peptides is often lower after the protein scaffold therapeutic has been created. Presumably this reduction in activity may be the result of steric hindrance of the pharmacophore in binding to the target or from alterations in the conformation of the peptide.

**TABLE 1 Protein scaffold therapeutics under development**

| Scaffold | Company | Products | Pharmacophore | Clinical Status | Noteworthy Details |
|---|---|---|---|---|---|
| Albumin | Human genome sciences | Albuferon | Interferon-alpha | Phase 3 | Yeast |
| | | Albugon | Glucagon-like peptide-1 | Phase 1 | |
| | | Albuleukin | Interleukin-2 | Phase 1 (no longer active) | |
| | | Albutropin | Human growth hormone | Phase 1 (no longer active) | |
| | Conjuchem | CJC-1134 | Exendin-4 | Phase 2 | Albumin from yeast with chemical conjugation |
| | | CJC-1295 | Growth hormone | Phase1 (clinical hold) | |
| | | CJC-1131 | Glucagon-like peptide-1 | Phase 2 (no longer active) | |
| | | CJC-1008 | Dynorphin | Phase 2 (no longer active) | |
| Fc | Amgen | Enbrel | Tumor necrosis factor receptor | Launched | Peptibodies (bacterial expression) |
| | | AMG-531 | Thrombopoietin mimetic | Phase 3 | |
| | | AMG-386 | Angiopoietin antagonist | Phase 1/2 | |
| | | AMG-623 | BAFF antagonist | Phase 1 | |
| | | AMG-403 | Nerve growth factor | Phase 1 | |
| | Genentech | BR3-Fc | BR3 receptor | Phase 1 | |
| | EMD Lexigen | Fc-epo | Erythropoietin | Preclinical | |
| Antibody | CovX | CVX-045 | Thrombospondin-1 mimetic | Phase 1 | Mammalian expression |
| | | CVX-060 | Angiopoietin-2 antagonist | Phase 1 | |
| | | CVX-096 | GLP-1 receptor agonist | Preclinical | |
| Transferrin | BioRexis | Transferrin-GLP-1 | Glucagon-like peptide-1 | Phase 1 | Yeast expression |
| | | Transferrin-interferon beta | Interferon beta | Preclinical | |

The protein scaffolds are generally considered to be pharmacologically inert. The primary function for these scaffolds is to provide biological stability to prevent the pharmacophores from being rapidly cleared from the body. Scaffolds based on the $F_c$ domain also remove the potential for ligand binding to the variable domain which may reduce the need for conducting certain studies (i.e., tissue cross-reactivity) during the preclinical safety assessment of these new agents. Other antibody scaffolds have $F_{ab}$ domains that are specifically created to interact with synthetic linkers and not receptors or other ligands. However, the potential still does exist for some $F_c$ specific function for these types of molecules. Nevertheless, not all of these agents are antibody based; albumin and transferrin are examples of other scaffolds that are currently under evaluation.

The normal physiological roles for albumin include maintaining the osmolarity of the blood and aiding in the transport of fatty acids and hormones. While it can be presumed that these functions would not be appreciably altered by the fusion process that creates the biopharmaceutical agent, the potential implications on safety are not known. However, the protein load provided by dosing an albumin conjugate would be minor in comparison to the overall albumin pool in the body, which could reduce the potential safety implications. On the other hand, the possible interaction of a transferrin fusion protein with its specific receptor could theoretically alter the internalization process and the subsequent endosomal acidification release of transported iron. The impact of this type of conjugate on the distribution and availability of iron throughout the body has not been reported. Thus the potential does exist for some protein scaffold therapeutics to have safety implications that would require specialized testing (i.e., monitoring of normal physiological activities) to accurately assess the preclinical safety of the new therapeutic agent.

Consistent with the case-by-case paradigm, the potential direct or secondary effect(s) of any scaffold should be carefully considered and monitored if judged necessary. An additional issue with these natural scaffolds is that each of these fusion proteins could be considered "foreign" by the immunological surveillance network. While the complex structure of these macromolecules may disguise the active pharmacophore moiety to some extent, it is still possible for immunogenicity and hypersensitivity reactions to result from this alteration of the scaffold. The extent to which this occurs will, in all likelihood, be dependent on the individual molecule and will not necessarily be entirely predictable for a particular scaffold. The inability to predict animal or human immunogenicity of these agents has been already been demonstrated by the diversity of antigenicity responses to various human protein therapeutics [2,3].

## 2 SAFETY ASSESSMENT PROGRAM STRATEGY DEVELOPMENT

The strategies to develop the preclinical safety assessment programs for biotechnology-derived therapeutics have often been characterized as case by

case. Historically this customization has resulted from the need for flexible strategy designs from a variety of different challenges that have included rapid metabolism, inappropriate tissue distribution, poor chemical stability, and absence of the target or lack of efficacy in the animal species that were planned for use in the pharmacology or toxicology studies. In contrast, the preclinical toxicology and drug metabolism studies necessary for the regulatory approval of new drug candidates for small molecules (NCEs) have been standardized for decades [4], in part, the result of the perceived chemical and biological similarity of these molecules. Since 1991 these developmental strategies and study designs have been the subject of International Conference on Harmonization (ICH) activities [5]. Guidances have been issued for a number of toxicological, pharmacokinetic, and drug metabolism studies that are conducted for small molecule NCEs. In contrast, only the ICH S6 guidance for biotechnologically related molecules has been published [6].

As previously discussed, the common goal for protein scaffold therapeutics is to extend the duration of pharmacologic action of the pharmacophore by increasing the circulating half-life. The development strategies required for regulatory submission and approval of these fusion molecules could represent a combination of the studies conducted for traditional small molecule NCEs and biopharmaceuticals. These programs may be expected, however, to be relatively similar to biopharmaceuticals for the pharmacological, ADME/PK (absorption, distribution, metabolism, excretion/pharmacokinetics), and toxicological characteristics of the various components of the molecule.

## 3 CHOICE OF ANIMALS

A number of considerations can alter the design of the preclinical safety program for new pharmaceuticals whether they are traditional small-molecule NCEs or biotechnology derived. One of the earliest choices that can be made is the selection of animal species for the toxicology program. This decision can often be made a number of months prior to the filing of the IND. For the development of small-molecule NCEs the general criteria for selecting the rodent and nonrodent toxicology species include the presence of the pharmacological target, demonstration of efficacy, and a similarity to humans (in vitro) for ADME and PK characteristics. In an effort to develop more successful small-molecule drug candidates, many pharmaceutical companies now expend considerable effort in identifying new drug candidates for which the ADME and PK characteristics have been optimized prior to selection of a lead candidate for development. The predictability of this research over the last few years has become increasingly successful thanks to tremendous advances in analytical methodologies, pharmacokinetic modeling capabilities, and the increased availability of animal and human tissues [7]. As a result a considerable knowledge base of ADME and PK characteristics will have often been

established for small-molecule NCEs in several animal species by the time that the IND is filed.

Based on the primary objective of improving the pharmacokinetic profile, a new protein scaffold therapeutic candidate will also have generally been selected for development following optimization of ADME and PK characteristics. Other critical candidate selection issues may also include the potential immunogenicity of the therapeutic molecule, the possible induction of autoimmunity, and the comparability between animals and humans for tissue cross-reactivity. The immunogenicity potential of a new therapeutic agent can be viewed as a potential mechanism for increased clearance, which could result in the dramatically reduced exposure to the drug during a multiple-dose toxicity study (e.g., subchronic studies of one to six months in length). This type of response could effectively disqualify an animal species from longer term toxicology testing because of the reduced drug exposure. It is also possible for the expression of antidrug antibodies to have other profound toxicologic effects that can theoretically include deposition of immune complexes in the kidney.

Considerable efforts are often expended on the quantification and characterization of the ability of the toxicology test species to build an antibody response to biopharmaceuticals. While this may help explain some effects seen in animal toxicology studies, these antigenicity results are often extended to an attempt to predict the response in humans. However, the immunogenicity of human proteins in animals is not a perfect predictor for the production of human antihuman antibodies in clinical studies [2,8]. While altered physiological states in humans, including disease-induced immunosuppression, may play a role in this lack of correlation, it does remain that antigenicity in humans is a poorly understood process. Therefore the observation of rodent antihuman antibodies or primate antihuman antibodies may have more relevance to the interpretation of the results from the animal toxicology studies than the potential to predict immunogenicity in humans.

The choice of animals for the toxicology testing of protein scaffold therapeutics therefore may depend on a potentially complex case-by-case decision-making process. All relevant information related to the expected (or known) efficacy against the target, ADME/PK characteristics, immunogenicity, and toxicity of these molecules must be utilized in the design of the toxicology program. This informational database should include all that is known about the chemistry and biological actions of the macromolecular scaffold and the pharmacophore. For successful regulatory development, it is expected that all new candidate drugs use both a rodent and nonrodent species in the preclinical safety programs [4,9]. For small NCE drug candidates, the most commonly used rodent species is the rat, with the beagle dog being the most frequently used nonrodent [9,10]. These two species are usually chosen based on a combination of factors that include ease of use, widespread availability, economics, historical database of toxicological and pathological parameters, historical predictability of animal to human toxicities [10], and as previously indicated,

their pharmacological appropriateness and ADME/PK comparability to humans.

Since several of the scaffolds for protein scaffold therapeutics are based on endogenous proteins, the potential for antigenicity is anticipated to be low in humans [8]. However, since these scaffolds are based on human molecules, they are expected to be immunogenic in animals particularly in rodents. While rats are commonly used in most safety assessment programs, they often develop antibodies to human-derived proteins [8]. Therefore their use in a preclinical safety assessment program, as previously discussed, could be seriously compromised by the reduction in systemic drug exposure. Nevertheless, rats are always given serious consideration as a species for use in preclinical safety assessment because of their well-described historical database, relatively modest size (and reduced need for bulk drug) and their ease of use and care.

At the beginning of a new drug development program, it will probably be necessary to conduct some multiple-dose studies in rats to evaluate the immunogenic potential of the new molecule. If the result is positive, then the option to use another rodent species could include mice as the choice for subchronic or chronic toxicity studies. Although mice are infrequently used as the second species in preclinical safety assessment programs, they are often utilized in the efficacy studies for biotechnology-derived molecules. While animal husbandry issues may be somewhat challenging for some strains of mice, their small size can make them a potentially valuable species for safety assessment. As with rats the potential would seem to be high for mice to develop antibodies to a human-based protein. It would therefore be valuable to conduct the appropriate studies prior to including mice in a safety assessment program. On the other hand, if neither rats nor mice are considered appropriate, it may then be justified that only nonrodents be used in the toxicology program.

When considering the potential challenges with using rodents for protein scaffold therapeutics, it is clear that the preferred species will probably be nonhuman primates. The primary species currently used is the cynomolgous monkey, although rhesus monkeys have been used in a number of toxicology programs. Either species would seem appropriate provided that they respond pharmacologically to the drug. Protein scaffold therapeutics should also have the potential for reduced immunogenicity in primates, since the protein scaffolds and pharmacophores are directly derived from, or are similar to, those found in humans. The dramatically improved access to healthy captivity-raised cynomolgus and rhesus monkeys over the past two decades have supported the dramatic upturn in their use for biological molecule testing. However, due to continuing concerns about the contagious diseases carried by nonhuman primate species and the high bulk drug product requirements, the search has continued for primate toxicology species that carry fewer dangerous contagions, are easier to handle, and have lower body weights. Marmosets have been of interest in this regard but a commercially viable breeding program has not yet been established that could meet the needs for the worldwide development

of even a few new biological products. Their relative lack of availability has not allowed a comprehensive data base to be developed that could be used to evaluate their more widespread use in toxicology development programs. It should also be noted that dogs are seldom used in the toxicity testing for biopharmaceuticals for a variety reasons that include their perceived propensity to develop antihuman antibodies and high susceptibility to anaphylaxis. In addition there is a scarcity of immunologic reagents available for dogs, and their large body weight would require significantly higher quantities of the active pharmaceutical ingredient.

The species selection process for targeted biopharmaceuticals such as monoclonal antibodies and protein scaffold therapeutics may use tissue cross-reactivity studies as a critical component in the decision-making process. These studies can be conducted using a variety of tissues from humans and the toxicology program animal species [11]. The objective of these studies is to determine the potential presence of target-based binding and to gain at least a semiquantitative estimation of the degree of binding in these tissues. Additionally the results observed in the animals are compared to those found in human tissues. Based on the knowledge and experience of the researchers, these results may be extended to hypothesize upon the potential mechanism(s) of pharmacological and/or toxicological action of the agent. Whether this binding results from a direct and specific interaction with an intended receptor/epitope or against unintended targets may need to be determined by subsequent experiments. It is important to remember that these studies can often produce results for which the human relevancy is not well understood, so substantial challenges can exist for projecting the safety of the new molecule in humans when the importance of these studies is too heavily weighted [11].

Tissue cross-reactivity experiments should be initiated with caution, as they may add only minimal scientific value to the safety assessment process but may create considerable challenges. Clearly, the choice of the primary animal species for protein scaffold therapeutics should be one that expresses the appropriate receptor/epitope. This could be established by a variety of efficacy-based experiments both in vivo and/or in vitro. The tissue cross-reactivity studies of a new therapeutic agent would then be conducted in the preferred toxicology species, with the results being compared to the binding seen in normal human tissues. One of the primary challenges with these studies can be that the pharmacologic target may, or may not, be found in the animal models of human disease or in the normal animals that would be used for toxicology studies. For a species to be considered appropriate in a toxicology program, the presence of target-directed binding in animals would be expected. In an appropriately qualified species, off-target binding could be interpreted as being suggestive of the potential for toxicity. However, apparent off-target binding should not be used as an unequivocal indicator of toxicity. Instead, it may be beneficial to further explore the potential nature of this binding in various animal models so that this observation can be placed in its proper perspective.

It is also possible that some targets may only be expressed in humans, and then any binding observed in animals would be off-target related. In these circumstances animal toxicology studies could, and should, still be conducted. However, this type of program would not be ideal as it would only be able to provide a generalized toxicity assessment before the human clinical studies are conducted. Depending on the nature of any effects seen in animals (and their reversibility), it should still be possible to conduct human phase 1 trials provided that these studies are carefully and thoughtfully designed. Some important factors that could improve the clinical study design for this scenario might include a low starting dose, dosing only one patient at a time, and having ready access to emergency medical equipment and staff [12].

If an appropriate species cannot be identified based on the lack of the target or poor efficacy, it has also been proposed that toxicology studies could be conducted using homologous molecules. This type of experiment has proved quite useful for anti-sense molecules [13]. However appealing this approach might be, it should not be universally applied to all types of biopharmaceuticals. Because of the high potential for this type of study to produce misleading or poorly understood results, it would seem that this strategy should only be used if the pharmacology of the homologous molecule is very well understood. While endogenous pharmacophores may be highly conserved across various species, the biological control mechanisms could be substantially different from one animal species to another. This could include scenarios where the pharmacology of the agent is similar but the balance of signaling and control pathways would be substantially different. Alternatively, it is possible that an individual pharmacophore could have completely different biochemical or physiological roles in different animal species. It would seem that unless the overall biological control of the pharmacophore system is well understood, or can be determined without exhausting time and resources, these studies could prove to be of very little practical value.

In summary, the choice of animal species to use for the toxicological testing of protein scaffold therapeutics will continue to be a complex decision-making process. In general, rats and monkeys will probably continue to be the initial species to be considered for the safety assessment program for a protein scaffold therapeutic provided that they possess the appropriate pharmacological target. However, because of potential immunogenicity issues in rodents, the primary species for testing will continue to be the nonhuman primate. The choice of a second species for toxicology testing, or an alternative strategy such as homologous pharmacophores, must be given careful consideration as an incorrect choice could provide the potential to significantly hinder or delay the safety assessment process.

## 4 DURATION AND TYPES OF STUDIES

One of the practical challenges in developing a new therapeutic (pharmaceutical or biopharmaceutical) is to conduct the appropriate length and type of

toxicology study that will support an ongoing clinical program. Numerous cost pressures exist for all new pharmaceutical agents so that only those candidate drugs with the highest potential for success should proceed into and through development. Some of these economic factors also produce challenges to time lines for the conduct and reporting of toxicology studies. The first animal toxicology studies for the development of new pharmaceuticals are usually acute and/or short-term dose-range finding studies. After these initial studies are completed, the IND-enabling studies are initiated to support the regulatory submissions. The four week (i.e., one-month) study is usually the first conducted, and it is used to support the phase 1 or 2 human clinical studies for which the duration of continuous dosing must be equal to or less then four weeks. If standard development programs are conducted, the four-week animal toxicology study is then followed by studies that are three and six months in length.

As a new drug advances through the development process it is necessary for the animal toxicology studies to be equivalent, or longer, in duration than the human clinical studies. This strategy is generally referred to as the one-for-one dosing duration concept, and it can be a challenge for an aggressive human clinical research program that can easily outpace the ongoing animal toxicology studies. This is particularly true for the extension of dosing for human patients in phases 1 and 2. Because of the lengthy time required to order animals and conduct these toxicology studies, the proper planning and communication with clinical research teams and project management can ensure that these human clinical trials are not delayed. If a new agent is promising, and appropriate resources (e.g., bulk pharmaceutical product) can be identified, it may be desirable to accelerate the conduct of the three- or six-month toxicology studies by conducting them in parallel with the one-month study. From an operational viewpoint, the two studies can be conducted using a single protocol with the one-month study being an "interim" necropsy. Each of these studies incorporates the appropriate number of animals as if these studies were conducted separately. Therefore while these studies may provide some minimal cost savings from an administrative view, this strategy does require some "front loading" of resources, which includes the often limited supplies of the bulk active pharmaceutical ingredient. This strategy also incurs more initial risk, so it can be challenging for smaller companies or for any whose resources are limited or carefully controlled. In a recent Oncologic Drugs Advisory Committee meeting, a senior Food and Drug Administration official indicated that three-month toxicology studies would be sufficient to support a clinical program of new drugs prior to approval provided there are no overriding safety concerns [14]. This strategy could provide some advantages for promising new drugs, including protein scaffold therapeutics, that are being developed for critical diseases for which there are no currently available cures.

Another important consideration is whether developmental and reproductive toxicity studies and carcinogenicity studies should be conducted for protein scaffold therapeutics. As discussed in ICH S6, the decision to conduct

these studies should be based on the nature of the product, its biological/pharmacological actions, the clinical indication, and the intended patient population [6]. In addition the challenges elicited by the potential immunogenicity of the new drug in animals would be an important consideration. Because antidrug antibodies have the potential to decrease the circulating drug levels during the multiple day-dosing paradigms required for development or reproductive safety assessment studies, this can seriously complicate the interpretation of these studies; in fact it would question their value in the overall risk assessment process. Additionally the studies should not be required unless there are specific biological, pharmacological, or toxicological alerts that are suggestive of toxicity to the fetus, neonate, or mother.

Other factors that could impact the potential conduct of these studies may include the potential for class labeling. Recently the antidiabetic agent Byetta® was designated as a class C pregnancy hazard [15]. While the mechanisms for these effects on fetal development have not been reported, it is possible that there will ramifications for future therapeutic agents in this class. As with all types of safety assessment studies, the scientific rationale for the developmental and reproductive toxicity program should be carefully considered for protein scaffold therapeutics. Appropriate studies can then be conducted so that direct discussions with regulatory agency scientists can be held to achieve consensus.

The conduct of most in vitro genetic toxicology studies for biopharmaceutical molecules would generally not be necessary as indicated in ICH S6 [6]. Simply based on their size and physicochemical properties, it would not be expected that protein scaffold therapeutics would interact with genetic or chromosomal materials. Unless there are specific structural alerts, cellular or organ-specific toxicities, or other factors that increase concern, genetic toxicology tests should not, per se, be required for protein scaffold therapeutics. These same criteria would also apply for the animal carcinogenicity testing of protein scaffold therapeutics.

## 5 DOSAGE SELECTION

The dosage selection process can be challenging for any type of toxicology program, whether it be for small-molecule NCEs or biopharmaceutical derived molecules. The primary goals of such studies include identification of target organs, the no effect dosage level (NOEL), no adverse effect level (NOAEL), and the maximally tolerated dosage (MTD). For orally administered small-molecule NCEs it is often necessary to demonstrate that absorption and distribution are sufficiently adequate to allow the characterization of the toxicological potential of the molecule. In cases where the absorption is poor, it is often necessary to administer very high dosages to achieve reasonable safety multiples for systemic exposure (i.e., blood AUC levels). Most biopharmaceuticals will not be administered orally, so simple questions of absorption

are not relevant. However, more challenging questions related to the disposition of a biopharmaceutical may come into play especially when novel parenteral dosage forms or routes are used. That is, the molecule could encounter bioavailability challenges from the site of injection due to differential dissolution or solubility characteristics that could be exacerbated by high drug concentrations in the formulation.

Ultimately the fundamental question for all classes of molecules may eventually become "how much is enough"? That is, for any relatively nontoxic molecule it may not be possible to achieve toxicologically significant evidence of effects, including toxicity, even after very high and even heroic dosages. However, simple physicochemical characteristics such as test article solubility can limit the amount of active agent that can be administered via parenteral administration. It is also possible that a poorly soluble test article may require the need to exceed the maximum volume of vehicle that can be safely administered during the conduct of an animal toxicology study. While this limitation can be at least partially overcome by multiple daily administrations, or even continuous daily infusion, it is still possible to reach a point at which no higher dosages of the drug or vehicle can be given. For low toxicity small-molecule drugs, a suggested limit of 25-fold over the clinical human exposure has traditionally been used. The magnitude of this safety factor would certainly seem reasonable as it would ensure that the potential of significant toxicities would be explored in animals. If any effects were to be observed, they could then be used to help establish the maximally suggested starting dosage in humans.

Because of the pharmacological specificity that is designed into protein scaffold therapeutics, it can be anticipated that any toxicities would be minimal and/or would probably be related to exaggerated pharmacology. It is possible that any observed biological effects may be directly related to the desired effects of the compound, or they may result from secondary effects. Pharmacodynamic effects may even play a greater role in establishing the toxicology dosages for protein scaffold therapeutics than what has been used for small-molecule NCEs. A potential complication may be that the normal animals used for toxicology experiments may only minimally express the necessary targets integral for the pharmacological mechanism of action. In these cases it may be necessary to require that arbitrary safety factors be reached to establish the upper limits of guided therapeutic toxicology studies. Whether this should be established via a body surface area normalized dosage or via anticipated systemic exposure (i.e., AUCs), it would seem reasonable to continue to use a 25-fold safety factor as the maximum limit. It is ironic to note that the "safer" (i.e., less overtly toxic) agents may need to be tested at substantially higher concentrations or dosages than compounds that are "less safe."

In conclusion, the ability to define the dosages of a protein scaffold therapeutic to be used in toxicology studies may prove to be difficult. Because the potential exists for these compounds to produce minimal toxicity, the dose selection process may rely more heavily on exaggerated pharmacodynamic effects than traditional small molecules have historically used.

## 6 METABOLISM AND PHARMACOKINETICS

To file an IND for a small-molecule NCE drug candidate, it is expected that the ADME (absorption, distribution, metabolism, and excretion) and PK characteristics would be generally understood prior to the human phase 1 study. At this early stage of drug development, these studies would be conducted in the animal species that were used for the pharmacology and toxicology studies. Typically the pharmacokinetics of small molecules would be determined in the relevant rodent and nonrodent species by the administration of nonradiolabeled drugs via both the intravenous and oral (if intended) routes of administration. These studies would allow the comprehensive determination of many PK characteristics including half-life ($T_{1/2}$), volume of distribution ($V_d$), time to maximum blood levels ($T_{max}$), bioavailability ($F$), exposure (AUC), and potentially the route of excretion. Additional data could also have been gathered on the metabolic pathways of these new compounds using $^{14}$C- or $^{3}$H-radiolabels, animal and human tissues, and new techniques in qualitative and quantitative mass spectrometry.

For biopharmaceuticals, the pharmacokinetic and drug metabolism expectations for filing an IND are much less extensive than for small molecules. The pharmacokinetic studies should utilize the same route of administration and include the animal species used in the pharmacology and toxicology studies. The basic pharmacokinetic characteristics from the intravenous route should include; alpha and beta $T_{1/2}$, $C_{max}$, AUC, and $V_d$. If other parenteral routes of administration are used (i.e., subcutaneous, intramuscular), it would also be important to determine the $T_{max}$ and bioavailability. For biopharmaceuticals it is possible that various formulations would need to be investigated to define the effect on PK parameters. This could also necessitate the conduct of pharmacokinetic bridging studies in the selected toxicology species to ensure equivalent exposure that would result from sustained or delayed release formulations.

If all of the components of a protein scaffold therapeutic are naturally occurring or endogenous (e.g., proteins), then no classic drug metabolism, disposition, biotransformation or excretion studies would generally be required. That is, the metabolism, distribution, and excretion of these materials would be well understood and would not require further experimentation. Similarly, if some of these fundamental components (e.g., amino acids) are not naturally occurring but the metabolic disposition has been well-described, then no additional ADME studies should be necessary. However, this decision would depend on the chemical, physiological, and potential toxicological nature of the component. Because of their macromolecular nature, protein scaffold therapeutics would generate myriad metabolites, and the technical complexities involved in their analysis would require substantial resources with minimal scientific gain. Therefore, if the parent molecules are stable in the presence of serum or the appropriate tissue matrix, then the need to characterize the potentially innumerable metabolites would generally not warrant

the potentially considerable expenditure of scientific or labor resources required. The impact of nonnaturally occurring components on the safety assessment of protein scaffold therapeutics would require a careful assessment of the scientific needs. For instance, if the candidate molecule contained nonnatural amino acid components but the parent was stable in the presence of blood or tissues, then the need for traditional ADME studies would seem to be unnecessary. However, structural alerts and/or significant instability of the molecule with the subsequent potential release of these components would necessitate the need for a more detailed examination of ADME characteristics, particularly as the development of the compound proceeded. An important consideration in determining the necessity for additional studies being conducted would be the actual dose of these nonnatural components. Since these macromolecules will be quite large, it is probable that any single nonnatural amino acid or other component would be dosed at a vanishingly low amount that would be without biological consequence.

## 7 BIOMARKERS AND INDICATORS OF TOXICITY

The ability to successfully detect and quantify toxicity in animal studies is a critical goal for the safety assessment of pharmaceuticals and biopharmaceuticals. Generalized or target organ specific toxicities in animals have typically been evaluated by a variety of different parameters, including clinical observations, body weight, food consumption, hematology, clinical chemistry, urinalysis, tissue histopathology, specific organ system functionality, and target-specific biochemical/organ markers. Based on safety concerns and resource utilization issues, it is advantageous to identify an undesirable drug as early as possible in a discovery or development program. The ability to predict or detect toxicities has dramatically changed over the past decade as advancements in analytical chemistry, molecular pathology, genomics, and proteomics have become widely available. However, the continuing need in drug discovery and development remains to minimize the false negative or false positive drugs so that resources can be successfully directed to the best candidate compounds. Therefore expanding the use of new technologies should be encouraged to better screen for animal toxicities and to identify undesirable candidate therapeutics early in the discovery phase before regulatory-based studies are initiated. As the screening and predictive capabilities of these technologies are improved for animals, it would obviously be valuable to develop the ability to extrapolate these data to better prediction of human toxicities.

Protein scaffold therapeutics represent a wide variety of chemical families and pharmacological classes. As such, it is difficult to propose a "one size fits all" guiding strategy for assessing their safety. The monitoring of traditional parameters or markers of toxicity and pathology should continue to be used as the primary screening tools for assessing the safety of protein scaffold therapeutics. However, because of their nature, protein scaffold therapeutics may produce few classic manifestations of toxicity except for exaggerated

pharmacology. This will necessitate the need to establish the dosages for the toxicology studies based on pharmacological effects that would then require monitoring of the appropriate pharmacodynamic marker(s). These data would be used to support or validate the dose-limiting effects in the actual toxicology study. Depending on their nature and composition, protein scaffold therapeutics should also be examined for specific scaffold based toxicity. For example, monoclonal antibody scaffolds might require an assessment of effector function in vitro and/or an appraisal of all of the appropriate target tissues during a complete histopathological assessment in an animal toxicology study. Other scaffolds such as albumin or transferrin molecules would focus on an examination for disturbances of fluid or electrolyte balance or alterations in iron storage or metabolism, respectively.

It may be inappropriate or misleading to conduct some standard toxicology assays for protein scaffold therapeutics. For instance, the hERG assay is a common screening assay that is used to determine the potential for small-molecule NCEs to prolong the QT interval. However, this assay is sensitive to changes in protein and buffer concentrations [16]. Since many protein scaffold therapeutics are high molecular weight protein fusion or conjugate molecules, it would be difficult to adequately control for the addition of these large quantities of protein in this assay. It would probably be of greater value to monitor instead the cardiovascular system by conducting ECGs and doing blood pressure measurements during the nonrodent (e.g., primate) good laboratory practice toxicity study. It is also possible to monitor a number of additional physiological systems in nonhuman primates during a toxicology study, including the central nervous system (clinical signs), autonomic control (body temperature), and lung function (respiration rate) as would provide a generalized assessment of the safety pharmacology of this molecule.

The choice of assays or biomarkers that are used for assessing the safety of protein scaffold therapeutics should only be made after careful and thoughtful review. Specialized testing should be conducted for the scaffold or pharmacophore only after careful consideration has been given to the chemical structures and/or potential physiologic or pharmacologic actions.

## 8 CONCLUSIONS

In summary, protein scaffold therapeutics represent a new class of drugs that has been designed to improve the pharmacokinetic profile of smaller and more rapidly metabolized pharmacophores. As a result of the unique chemical and pharmacological nature of these therapeutic agents the case-by-case customized design of toxicology and ADME/PK studies will continue to be an evolving area of experimental design. The development strategies of each new drug will require a careful individual evaluation so that the appropriate experiments can be conducted. As the knowledge base increases for individual therapeutic agents and/or various compound classes, the preclinical program may require frequent adjustments to improve the risk assessment process for

the human patient. One example is that some compound or pharmacological classes may require similar preclinical study strategies for regulatory approval, which would include monitoring common markers and conducting similar studies and assays. At the other extreme, most compounds for the foreseeable future will probably require a case-by-case assessment that will best be designed by a fully integrated collaboration with regulatory scientists. It would therefore be encouraged for each sponsor company to have frequent and detailed communications with scientists at the regulatory agencies to ensure the successful development of new therapeutic agents.

## REFERENCES


1. Brown J, Entwistle J, Glover N, MacDonald GC. Preclinical safety evaluation of immunotoxins. In Cavagnaro JA, ed. *Preclinical Safety Evaluation of Biopharmaceuticals*, New York: Wiley, 2006;649–67.
2. Armin T, Carter G. Immunogenicity issues with therapeutic proteins. *Curr Drug Discov* 2004;4:20–9.
3. Shellekens H. Immunogenicity of therapeutic proteins. *Nephrol Dial Transplant* 2003;18:1257–9.
4. Goldenthal EI. Current views on safety evaluation of drugs. *FDA papers* 1968; 13–18.
5. Office Website for ICH. http://www.ich.org
6. ICH S6 Preclinical Safety Evaluation of Biotechnology-Derived Pharmaceuticals. http://fda.gov/cder/guidance/index.htm
7. Granneman RG, Marsh KM, Roberts SA. 2005, Unpublished results.
8. Bugelski PJ, Treacy G. Predictive power of preclinical studies in animals for the immunogenicity of recombinant therapeutic proteins in humans. *Curr Opin Mol Ther* 2004;6:10–16.
9. Dorato M, Vodicnik M. *The Toxicological Assessment of Pharmaceutical and Biotechnology Products*. New York: Taylor and Francis, 2001;243–83.
10. Olson H, Betton G, Robinson D, Thomas K, Monro A, Kolaja G, Lilly P, Sanders J, Sipes G, Bracken W, Dorato M, Von Deun K, Smith P, Berger B, Heller A. Concordance of the toxicity of pharmaceuticals in humans. *Regul Toxicol Pharmacol* 2000;32:56–67.
11. Hall W, Price-Schiavi SA, Wicks J, Rojko JL. Design and analysis of tissue cross reactivity studies. In Cavagnaro JA, ed. *Preclinical Safety Evaluation of Biopharmaceuticals*. New York: Wiley, 2006.
12. Expert Scientific Group on Phase One Clinical Trials Final Report 30 November 2006. www.tsoshop.co.uk
13. Levine A. The toxicology of oligonucleotides therapeutics and understanding the relevance of the toxicities. In Cavagnaro JA, ed. *Preclinical Safety Evaluation of Biopharmaceuticals*. New York: Wiley, 2006;357–74.
14. ODAC March 13, 2006 meeting minutes. http://www.fda.gov/ohrms/dockets/AC/06/minutes/2006-4203M1.pdf
15. Byetta® Product Label. http://www.fda.gov/cder/foi/label/2005/021773lbl.pdf
16. Gintant G. *Personal Communication*, Abbott, Abbott Park IL, 2006.

# Preclinical Safety Evaluation of Immunotoxins

JENNIFER G. BROWN, PhD, JOYCELYN ENTWISTLE, PhD, NICK GLOVER, PhD, and GLEN C. MACDONALD, PhD

**Contents**

1 Introduction
2 Immunotoxin Development
  2.1 Target Antigens
  2.2 Toxins
3 Rational Design of Immunotoxins
  3.1 Immunogenicity and Toxicity
  3.2 VB4-845
  3.3 Mechanism of Action
4 Immunotoxin Evaluation Using the ICH S6 Guidelines
  4.1 Specificity, Cytotoxicity, and Serum Stability
  4.2 Selection of Animal Model and Route of Administration
  4.3 Safety Assessment
5 Progress of Immunotoxin Therapeutics
6 Summary
  References

## 1 INTRODUCTION

Chemotherapy represents the most common approach in the treatment of cancer. However, the overall clinical success of chemotherapeutics is often limited by drug resistance and nonselective targeting resulting in dose-limiting toxicities. The preference for highly potent anticancer molecules that specifically target tumor cells while demonstrating minimal toxicity toward normal

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

tissue guides the rational design of the next generation of anticancer therapeutics.

Cancer immunotherapy became possible with the identification of tumor-associated antigens as well as the discovery of monoclonal antibodies (MAbs) that have the intrinsic properties of high affinity and specificity for their target antigen [38]. Antibodies exert their biological effector functions (e.g., antibody-dependant cellular toxicity (ADCC) or complement-dependant cellular toxicity (CDC)) once the antibody has bound to its target antigen. Despite the unique specificity of MAbs, in many cases the clinical benefit has been marginal prompting the search for more potent tumor-selective drugs that possess a different mechanism of action. The coupling of a cytotoxin to a MAb (or Mab fragment) has resulted in the generation of immunotoxins that derive their unique specificity from the antibody and impart a potent cell death signal to the targeted cells. Immunotoxins are emerging as important therapeutic agents for the treatment of a number of carcinomas and haematologic cancers [28]. In addition they are also being investigated for other diseases such as HIV [39,53], graft versus host disease [62], and autoimmune diseases [63].

This chapter will outline the development of immunotoxins and describe the preclinical development required for advancing VB4-845, an anti-EpCAM targeting scFv linked to a truncated form of *Pseudomonas* exotoxin A(252-608), into the clinic.

## 2 IMMUNOTOXIN DEVELOPMENT

The potency of an immunotoxin is dependent on the biochemical properties of both the antibody and toxin moieties. Of paramount importance are the characteristics of the target antigen and the antibody affinity for that antigen, rate of internalization into the cell, as well as the efficiency of the intracellular processing and the type of toxin.

### 2.1 Target Antigens

The development of a successful immunotoxin is clearly dependent on the choice of the antigen [7]. Prerequisites would be an antigen having a medium to high density of expression and a relatively homogeneous distribution on the tumor cell surface. As most toxins exert their mechanism of action in the cytosol by catalytically disrupting protein synthesis, an internalizing antigen is a necessity. Antigens that are shed from the cell surface, such as the carcino-embryonic antigen (CEA) and the non-Hodgkin's lymphoma (NHL) idiotype, do not represent optimal targets as free antigen would be competing with antigens displayed on cancer cells for the immunotoxin [64]. If shedding antigens are to be targeted, higher doses of drug would be required to remove the competing antigen from the circulation [36,45].

A number of solid tumor targets that are preferentially expressed on cancer cells have been identified such as the epithelial cell adhesion molecule

(EpCAM), HER2/neu (c-erbB-2), EGFR, cytokine receptors, mesothelin, as well as the carbohydrate-associated antigens such as LewisY that are highly expressed in many epithelial tumors [52]. Similarly many differentiation antigens have been identified for hemapoietic malignancies, including CD19, CD22, CD25, CD30, CD33, and CD56 [28].

### 2.2 Toxins

Certain plants, fungi, and bacteria produce pathogenic peptide toxins that are able to kill mammalian cells. Examples of these include ricin, gelonin, saporin, bryodin, pokeweed antiviral protein, and bouganin derived from plants, the fungal-derived toxins such as restrictocin and mitogillin, and the bacterially derived toxins diptheria toxin (DT) and *Pseudomonas* exotoxin A (ETA/PE). Although many of these toxins, as well as several other more exotic varieties, have been used for the construction of immunotoxins, the ones most commonly used have been ricin [32,33] from castor bean (*Ricinus communis*), diptheria toxin [23,61] from *Corynebacteriun diphtheria*, and *Pseudomonas* exotoxin A [46,47] from *Pseudomonas aeruginosa*. Toxins such as DT and ETA are extremely potent in killing tumor cells, and they exert their effect in the cytosol by interrupting protein synthesis, resulting in cell death [27].

## 3 RATIONAL DESIGN OF IMMUNOTOXINS

In general, first-generation immunotoxins consisted of chemically conjugating toxins to full-length antibodies. However, many of these toxins, such as ricin, naturally contain a cell-binding domain that targets normal tissue and proved unsuccessful in both animal models and in the clinic due to their nonspecific binding to normal tissue. In one case the attachment of ETA to the intact anti-TAC antibody that binds to CD25 on T cells and T cell malignancies resulted in severe liver toxicity in a phase 1 trial [16,46].

The design of immunotoxins has become more sophisticated, and efforts have been made to engineer safer, more efficacious molecules through a better understanding of the biochemistry of the toxins and antibodies as well as the physiological limitations surrounding effective delivery. For example, immunotoxins have been constructed with the cell-binding domain removed [25,55], thereby reducing toxicity and ensuring targeting through the antibody moiety. In addition genetically engineered fusion constructs containing antibody fragments with increased stability have been generated to enhance tumor penetration and to maximize serum half-life.

### 3.1 Immunogenicity and Toxicity

Despite the recent advances in immunotoxin design, there remain two major obstacles still to be resolved, namely immunogenicity and toxicity. Although the humanized or fully human antibody moiety of an immunotoxin has limited

recognition by the immune system, toxins are highly immunogenic and rapidly elicit an immune response upon administration to patients. Further many people have been immunized against diptheria and already possess neutralizing antibodies against the toxin. Up to 20% of the general population possesses anti-*Pseudomonas* antibodies as a consequence of *Pseudomonas* infections, and this number can be as high as 80% in long-term hospitals patients (VBI, unpublished data). Similarly patients who have been exposed to castor oil may have developed anti-ricin antibodies. Although considerable success has been achieved in the treatment of patients with leukemias and lymphomas [48] due in part to the immunosuppressive nature of these diseases, this is not the case for immunocompetent patients with solid tumors who may rapidly develop antibodies precluding repeat systemic administration of the immunotoxin [20,26,44].

Another serious side effect of immunotoxin treatment and a consideration in designing a preclinical program for immunotoxins is vascular leak syndrome (VLS). VLS is a nonspecific, non–antigen-related toxicity characterized by fluid leakage from the capillaries into the tissue resulting in low blood pressure and reduced blood flow to internal organs. Major symptoms are low blood pressure, edema, and low levels of albumin. VLS symptoms have been observed in many immunotoxin trials, and although these symptoms are generally manageable, reports of vascular collapse have been reported with certain ricin-based immuntoxins [14]. In order to abrogate vascular damage, several ricin A chain constructs, with mutations in the VLS-associated motif, have been engineered and are currently being evaluated [3,59].

### 3.2 VB4-845

VB4-845 is a recombinant fusion protein consisting of a tumor-targeting humanized single-chain antibody fragment, 4D5MOCB, specific for epithelial cell adhesion molecule (EpCAM) linked to a truncated form of *Pseudomonas* exotoxin A (ETA) that lacks the cell-binding domain, ETA(252-608). EpCAM is a cell surface marker that is highly expressed on carcinoma cells of epithelial origin, but has limited expression on normal cells [37,51,65]. Once inside the cell, ETA is a potent inhibitor of protein synthesis that induces cell death [15]. VB4-845 is a single 70 kDa protein, produced in E104 *E. coli* cells and is being developed for intratumoral injection for patients suffering from squamous cell carcinoma of the head and neck (SSCHN).

As with other immunotoxins, immunogenicity and toxicity are two of the major challenges that limit the use of VB4-845 in the clinic. Direct administration of VB4-845 into the tumor offers a number of advantages over systemic delivery. The intratumoral route provides a higher concentration of drug to the tumor than could be achieved by intravenous injection. Moreover, since the interstitial protein concentration is minimal relative to the circulation, in particular for protein in excess of 60 kDa, the concentration of antibodies preexisting or generated over the course treatment would be low relative to

the administered dose, thereby minimizing the neutralizing effects of an anti-immunotoxin response. As well, since immunotoxin uptake by the tumor should be maximized due to intratumoral injection, this route of delivery is predicted to limit the interaction of the immunotoxin with nontarget tissue and thus minimize the likelihood of dose-limiting toxicity.

### 3.3 Mechanism of Action

It has been wellestablished that ETA irreversibly inhibits protein synthesis in mammalian cells by adenosine di-phosphate (ADP)-ribosylation of elongation factor 2 [29,49]. To demonstrate that the activity of VB4-845 is consistent with that of ETA by the inhibition of protein synthesis, the uptake of [3H]leucine was measured in EpCAM positive and EpCAM negative tumor cell lines, following the addition of VB4-845 to the cell cultures. VB4-845 inhibited protein synthesis in EpCAM positive SW2 cells with an $IC_{50}$ of 0.01 pM but not in the EpCAM negative control cell line RL over the range of concentrations tested (0.0001–100 pM) [12]. The activity of VB4-845 was shown to be due to the inhibition of protein synthesis and is consistent with the mechanism of action of ETA (Figure 1). Results also indicate that EpCAM is required on the cell surface for treatment with VB4-845 to result in pharmacological activity (VBI, unpublished data).

To examine the requirements to take an immunotoxin into the clinic, the preclinical developmental strategy of VB4-845 is described to examine the challenges encountered to progress this antibody from bench to clinic by following the ICH S6 guidelines.

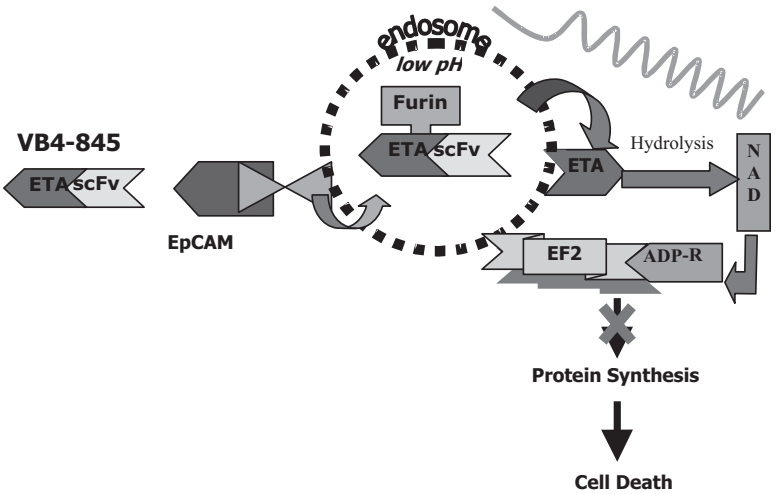


**Figure 1** Mechanism of action of VB4-845. See color insert.

## 4 IMMUNOTOXIN EVALUATION USING THE ICH S6 GUIDELINES

Each immunotoxin is evaluated in a series of stages. These include assessment of specificity, cytotoxicity, and serum stability; selection of a relevant animal species and route of drug administration; and finally safety evaluation.

### 4.1 Specificity, Cytotoxicity, and Serum Stability

Initially immunotoxins are tested in vitro for specificity, cytotoxicity, and serum stability. The specificity of VB4-845 for EpCAM expressing tumor cells was demonstrated by flow cytometry and cell growth analyses of various epithelial-derived tumor and normal cell lines originating from different tissue types. VB4-845 showed strong cell surface reactivity to EpCAM positive tumor cell lines such as Cal 27, a squamous cell carcinoma of the tongue, but not to the colon-derived, EpCAM negative cell line, Colo-320. The specificity of VB4-845 was further highlighted by assessing cytotoxicity against a panel of cell lines exhibiting varying degrees of EpCAM expression. As predicted by the binding data, all EpCAM positive tumor cell lines were sensitive to killing by VB4-845, with $IC_{50}$ values ranging from 0.001 to 1.84 pM. Normal cell lines, on the other hand, either were not affected by VB4-845 ($IC_{50} > 500$ pM) or were anywhere between 50- to 1000-fold less sensitive to VB4-845 than their counterpart tumor cell lines. For example, VB4-845 displayed potent activity against the bladder tumor cell line, TCCSUP ($IC_{50} < 0.005$ pM) but displayed no activity against the normal bladder cell line, HMVEC-bd ($IC_{50} > 500$ pM). This differentiation was also demonstrated in EpCAM positive and negative efficacy models [12].

Because immunotoxins are specifically targeted therapies, toxicities due to binding to receptors/antigens present on normal tissues occur at a much lower frequency than toxicities experienced with other anticancer drugs such as chemotherapy. According to FDA specifications [13], all antibodies are required to be tested for immunoreactivity against a human tissue panel of 33 normal frozen human tissues using immunohistochemistry (IHC). The in vitro immunoreactivity binding pattern of VB4-845 was membrane-associated in epithelial tissues, as is consistent with the detection of EpCAM expression by other antibodies [4]. Although normal epithelial cells express EpCAM, it is generally restricted to the basolateral portion of the cell [42], and therefore minimal in vivo binding is expected by VB4-845. To further ensure minimal patient toxicity, VB4-845 will be directly administered into SCCHN tumors. It should be noted that although tumor-reactive antibodies preferentially target cancer cells, they often cross-react with certain normal tissues that have limited expression of the target antigen, giving rise to significant toxicities. This does not preclude the use of such molecules as therapeutics. However, it underscores the requirement for testing the immunotoxin in a suitable animal model to determine its degree of interaction with critical normal tissues in the body.

Serum stability is another important attribute of immunotoxins as each immunotoxin must be sufficiently stable to progress from the blood into the tissues and be capable of penetrating into a tumor in a sufficient concentration to result in tumor cell death [8,12]. Examination at 37 °C showed that VB4-845 was a stable immunotoxin [12].

Although extensive in vitro testing is performed on each immunotoxin to assess its cytotoxicity and specificity, it is animal model studies that ultimately determine the serum half-life, serum stability, as well as the efficacy of an immunotoxin. A compilation of the in vitro and in vivo data enables a decision to be made on the potential of each immunotoxin to be an effective and safe therapeutic and to determine whether human clinical trials are warranted.

### 4.2 Selection of Animal Model and Route of Administration

As part of evaluating a new biopharmaceutical, the drug should be administered in a relevant animal species in which it is pharmacologically active. For an immunotoxin the drug should bind to the intended receptor or epitope expressed in the animal model. However, this result may not always be possible. In the case of VB4-845, cross-reactivity using IHC was examined in several animal species commonly used for toxicology studies: mouse, rat, dog, cynomolgus and rhesus monkeys, as well as the chimpanzee. No cross-reactivity was observed in any tissues of animals normally used for toxicology studies. Some cross-reactivity was observed in chimpanzee tissue, but because the pattern was not the same as that observed in human tissues, it was not considered to be a relevant species.

The use of nonrelevant animal species is discouraged as the results of such studies may be misleading (per ICH S6). Therefore the next option, when available, is to substitute a relevant transgenic animal model for a pharmacologically relevant species. Transgenic animal models for human EpCAM have been developed, but they either have a different tissue expression pattern from that seen in humans [41,43] or the model has not been validated [40], making them unsuitable for the evaluation of the safety of anti-EpCAM immunotherapeutics.

While safety evaluation programs require studies to be conducted in two species to characterize drug toxicity, should no viable option exist for a relevant species to conduct toxicology studies, the FDA suggests that toxicity be assessed in a single species [13]. Although no animal model system was available to examine possible binding of VB4-845 to EpCAM receptors, numerous studies have examined ETA-conjugated immunotoxins. It is well documented that the intravenous administration of ETA immunotoxins to rats results in symptoms that resemble VLS as seen in human immunotoxin trials [20,57]. Thus the choice of the Sprague–Dawley rat for toxicological testing was made for the well-known effects of immunotoxins in this animal model.

The route and treatment regimen in toxicology studies are expected to be as close as possible to that in the clinic. Single-dose studies aid in selecting the

route of administration for a toxicology study if the intended human route of administration can not be mimicked in the animal. For VB4-845, the chosen route of administration in the clinic was intratumoral. While intratumoral administration was successfully demonstrated in a preclinical efficacy mouse model [12], this mode is not a viable option in toxicology studies. Therefore intradermal drug administration was used as a surrogate route of administration. Intravenous dosing, although not intended in the clinic, was used in the toxicology study with the rat as a comparison to examine the possible "worst case" effect of systemic exposure.

### 4.3 Safety Assessment

***Single-Dose Toxicology Studies*** Drug dosage in an animal model should be conducted so that a dose–response relationship may be examined. This may range from a no observed effect level (NOEL) and no observed adverse effect level (NOAEL) all the way to a maximum tolerated dose (MTD) of the drug. The dose response in a single-dose study will assist to determine the dose levels to be selected for a repeated-dose toxicology study. These levels will also aid in determining the first-dose level of the drug in humans as well as the therapeutic index and the margin of safety when dosing humans. When considering dosage in test animals, the volume must also be considered. Ethical maximum volumes have been determined for different species, which may mean that if there is a limitation to the drug concentration, a maximum feasible dose will be determined instead of an MTD.

Clinical signs noted in the dose-ranging study conducted in Sprague–Dawley rats administered VB4-845 locally (intradermal, ID) and systemically (intravenous, IV) were related to injection site lesions that exhibited a dose-dependent effect. There were no other findings in animals locally dosed. Animals that were systemically dosed had an increase in red blood cell parameters, total red blood cell counts, hemoglobin, and hematocrit and a decrease in albumin, total protein, and albumin–globulin ratio. While all these findings were dose dependent, these variations were within the normal physiological range.

Single-dose studies also help to determine whether dosing modifications are required based on the bioavailability or pharmacokinetics of the drug in the test species. Route of administration may also be modified if the drug has limited bioavailability by the chosen clinical route, or else treatment regimen may be changed to compensate for high clearance or low drug exposure.

***Repeated-Dose Toxicology Studies*** The repeated-dose toxicity study is expected to be a toxicology study under GLP conditions [17], which examines clinical signs, hematology, clinical biochemistry, urinalysis, and bone marrow to evaluate the effect of the drug during and after administration. Additional animal groups are required so that drug effect can be examined immediately after dosing in one group of animals and then later in another group of animals

to determine whether any observed effects are reversible or in some cases whether the drug effects are delayed. Safety pharmacology, toxicokinetics, and immunogenicity may also be examined in this study for immunotoxins.

Based on the injection site reactions in the single-dose studies, it was anticipated that there would be difficulties dosing the Sprague–Dawley rat via the ID route. Therefore a subcutaneous (SC) route of administration was used as another representative route of local administration. Injection site reactions (slight erythema, edema, superficial necrosis, ulcerations, and scab formation) attributable to VB4-845 that were dose-related and noted at or above 5.0 μg/kg in most cases had resolved by the end of the observation period. Dose-dependent but transient changes were noted in hematology, coagulation parameters, and serum chemistry and were most likely due to acute tissue injury and inflammation at the injection sites. Liver enzyme levels were elevated in rats at upper-dose levels but returned to the normal physiological range by the end of the observation period. Systemic administration of VB4-845 (77.8 μg/kg) in rats resulted in microvascular injury and pulmonary edema, with subsequent hypoxia; these findings were consistent with VLS previously observed in rats exposed to ETA-based immunotoxins [22,56].

Examination of the various parameters may identify markers that vary with dose and may be used to examine drug effects in human studies. For example, although few effects were attributable to the local administration of VB4-845 in Sprague–Dawley rats, liver enzyme levels were elevated in rats at upper-dose levels. This provides a marker to follow drug response in clinical patients. As there was no evidence of toxicity following local administration at any of the dose levels tested, no NOAEL was attained at the highest dose tested (77.8 μg/kg), and it can be assumed that higher doses may be possible. In contrast, the animals treated systemically with VB4-845 experienced lethal toxicity at the same dose (77.8 μg/kg), thus illustrating a lower NOAEL for this route of delivery.

While animal studies are used for safety assessment prior to administration to humans, preclinical testing may not always predict human effect. For example, continuous-infusion therapy with 260F9 monoclonal antibody-recombinant ricin A chain resulted in severe neurotoxic effects in humans that were not demonstrated in monkey toxicology studies [21]. In such cases where suitable animal models are not available for safety testing it is important to consider the application of an appropriate safety factor to provide a margin of safety for protection of humans receiving the initial clinical dose [11,18].

***Immunogenicity*** Immunogenicity is a significant complicating factor surrounding the administration of immunotoxins to humans [19]. Immunogenicity can arise from either the antibody or toxin portions if they are foreign proteins. The shift from murine to humanized or human antibodies has reduced the immune response due to the antibody portion [10,35]. However, immune responses are still expected because the toxins employed are either of bacterial or plant origin and are thus inherently highly immunogenic.

The immune responses to VB4-845 by both the intended local route and systemic route of administration were investigated. Not surprisingly, a dose-dependent anti-drug antibody response was observed to both the antibody and toxin portions of the construct and a similar level antibody titer was induced regardless of sex. The VB4-845 titer following local administration was significantly higher than that observed after systemic administration at the same dose level, a phenomenon that has been demonstrated to be due to the use of different routes of delivery [50].

Although relatively high serum antibody titers were produced against VB4-845, this is not expected to have a negative impact on drug administration. Local administration of drug is expected to ensure a high local drug concentration in the tumor before coming into contact with anti-VB4-845 antibodies in the circulation. This point was illustrated with scFv(FRP5)-ETA, where a patient had a complete clinical response to intratumoral treatment (second administration) even though an immune-response was generated that completely neutralized the immunotoxin [2].

***Drug Exposure Evaluation*** Immunotoxin exposure is largely dependent on the antibody portion used for targeting [10]. While whole IgGs may have half-lives up to 36 hours [5], scFv single-chain fragments can have a half-life as short as an hour or less. Although larger antibody portions are more stable and may prolong drug exposure, a larger sized molecule has more difficulty in tumor penetration [1]. A smaller sized antibody portion may not have as long a half-life, but it may permit better access to tumor cells and is able to leave the circulatory system more quickly, thereby reducing the exposure time of the endothelia to the toxin, and perhaps decrease VLS toxicity. $F_{ab}$ antibody fragments or pegylation of the molecule increases the protein size, making it more stable but still small enough for cell entry.

As no relevant species exists, biodistribution studies were conducted in xenograft mice bearing EpCAM positive and negative tumors. This study confirmed that VB4-845 was retained in EpCAM positive tumors [12]. Although there was some detection in EpCAM negative tumors, this was most likely due to increased tumor vascularization. The biodistribution study also indicated that other organs may be targeted. However, toxicology studies did not result in any toxicity findings, indicating that the accumulation of radioactivity did not reflect binding or internalization of the immunotoxin within these tissues [12].

Toxicokinetic (TK) analysis of VB4-845 indicated that there was no gender difference. Nevertheless, different profiles were generated based on the mode of administration. Local administration resulted in a maximum VB4-845 plasma concentration of 50 ng/ml 4 h after administration (Figure 2). The subcutaneous dose appeared to be incompletely absorbed into the circulation, resulting in low bioavailability (13%) upon sampling (Table 1). Systemic administration resulted in a maximum VB4-845 plasma concentration of 1000 ng/ml after 10 minutes. The disappearance profile following the systemic

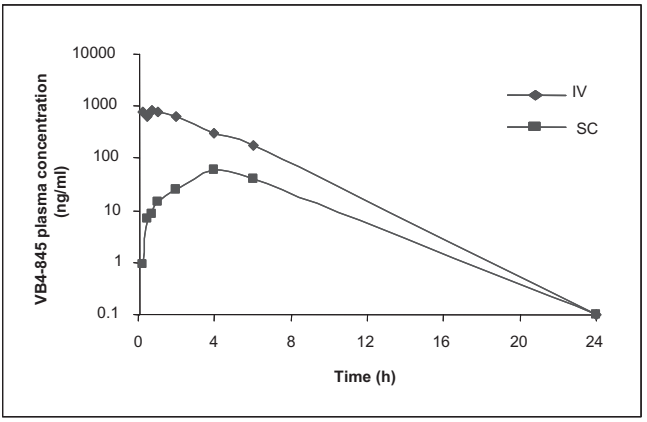


**Figure 2** Mean plasma concentration-time curves following either systemic (intravenous) or local (subcutaneous) administration of VB4-845 on day 1. VB4-845 was not detected at 24 hours in either mode of drug administration.

**TABLE 1 Toxicokinetic profile of VB4-845 administered SC or IV in Sprague–Dawley rats**

| Route | $C_{max}$ (ng/ml) | $AUC_{0\text{-inf}}$ (ng·h/ml) | k ($h^{-1}$) | Ka ($h^{-1}$) | $t_{1/2}$ (h) | CL (ml/kg/h) | $V_d$ (ml/kg) | %F |
|---|---|---|---|---|---|---|---|---|
| SC | 50 at 4 h | 421 | NC | 0.31 | NC | NC | NC | 13 |
| IV | 1000 at 10 min | 3242 | 0.3 | NC | 2.3 | 24 | 80 | NC |

Abbreviations: $AUC_{0\text{-inf}}$ = area under the curve, CL = clearance, $C_{max}$ = maximum plasma concentration, %F = bioavailability, k or Ka = elimination constant, NC = not calculated, $t_{1/2}$ = half-life, $V_d$ = volume of distribution. The pooled data set from male and female rats was used for calculations.

dose was consistent with a one-compartment distributional model and first-order elimination with a half-life of 2.3 hours.

Comparison of plasma concentrations between days 1 and 7 showed no difference with local injections. However, concentrations following systemic injections were much lower on day 7 as compared to day 1. A dramatic enlargement of the distribution space, or possibly an unspecified bioadaptation is suggested, or more likely, the observed immune response depleted the blood plasma concentration. As mentioned earlier, an immune response raised against immunotoxins can affect the TK profile of a drug. The antibodies raised against the response bind to the antibody portion of the immunotoxin, causing it to be cleared at a quicker rate. This can decrease the half-life of the drug and, in most cases, is the stopping point for re-administration of the drug. Low bioavailability after local administration will not be an issue as VB4-845 is directly targeted to the tumor but rather indicates that there will be low systemic exposure to the immunotoxin.

Toxic effects of metabolites from drug degradation do not need to be monitored for immunotoxins (per IHC S6). As a recombinant protein, immunotoxins entering the human body are quickly degraded to small peptides and amino acids in the blood by proteases that specifically target foreign proteins and are cleared by the kidney.

***Safety Pharmacology*** Safety pharmacology is important to assess the drug's effect on physiological functions through in vitro and in vivo assessment of central nervous, cardiovascular, and respiratory systems. These studies are required for biotechnology-derived biopharmaceuticals and assist to establish the type of monitoring parameters that may be required in clinical studies. However, because immunotoxins are specifically targeted drugs, these studies do not need to be conducted as individual studies but may be combined with toxicology studies that evaluate safety pharmacology endpoints.

Clinical signs observed in VB4-845 toxicology studies allowed for an in vivo assessment of the central nervous system and functional ability of the dosed rats. Animals dosed locally exhibited normal behavior, whereas those dosed systemically exhibited neurological symptoms (wobbly gait) and respiratory difficulties (dyspnea). Full safety pharmacology studies have not been conducted with VB4-845 because they are not required prior to the first administration in humans for anticancer drugs examined in end-stage cancer patients (per ICH S7A). VB4-845 has progressed to pivotal clinical trials, and therefore safety pharmacology studies examining cardiovascular and respiratory response are planned.

***Immunotoxicity Studies*** Immunotoxicology studies are required for pharmaceuticals that may affect the immune system through suppression, enhancement, or sensitivity (per ICH S8). In general, immunotoxins do not have an effect on the immune system, and standard testing batteries are not recommended for biotechnology-derived pharmaceuticals. However, examination of hematology results together with detailed histopathological assessment of immune organs from single- or repeated-dose toxicology studies are routinely performed with biopharmaceuticals, and these evaluations provide initial information on whether the immune system has been affected.

***Reproductive Performance and Developmental Toxicology Studies*** The assessment of reproductive performance and developmental toxicity is dependent on the clinical indication and patient population. The majority of immunotoxins in development are currently used for the treatment of cancer, a patient population that tends to be an older, nonreproductive population. As well, potential reproductive concerns for the patients in ongoing clinical trials are usually addressed in the clinical trial protocols.

Although VB4-845 will never be administered systemically and does not bind to human placenta, studies are planned as a step for marketing registration. Because a relevant species has not been identified, and the rat is unsuit-

able for use because it cannot be dosed systemically, embryofetal development will be examined in time-mated female rabbits to determine placenta binding and potential for transplacental passage. A formal study will not be conducted unless fetal effects are observed.

***Genotoxicity Studies*** While the majority of the principles of the ICH S6 apply to immunotoxins, because they are a biotechnology-derived biopharmaceutical, certain tests such as genotoxicity and carcinogenicity studies do not directly apply.

Examination of genotoxicity of pharmaceuticals is required to assess the interaction of the drug with DNA. These studies are generally not applicable to immunotoxins. Unlike chemotherapeutics that cause cell death through DNA interaction, immunotoxins mediate cell death by preventing protein synthesis. However, immunotoxins use a linker to connect the toxin to the antibody that may need to be examined if it is an organic linker and has the ability to bind DNA (per ICH S6). The majority of immunotoxins use either a nonreducible thioether linker for intact toxins or a disulfide bond for A chains and ribosome-inactivating proteins and do not interact with DNA.

***Carcinogenicity Studies*** Immunotoxins, in general, do not have the ability to transform cells or promote the growth of transformed cells. Therefore carcinogenicity bioassays are considered inappropriate. This is in contrast to toxins fused to growth factors that may promote tumor growth under certain circumstances if, for example, a less than toxic dose were administered. Should the immunotoxin interact with growth factors or cytokines, in vitro studies may be required to examine whether growth is promoted in transformed cells.

## 5 PROGRESS OF IMMUNOTOXIN THERAPEUTICS

Cancers of hematologic origin are more accessible to immunotoxin therapy and thus more amenable to treatment than solid tumors. One of the most promising immunotoxins, currently in development is BL22 used for the treatment of hairy cell leukaemia. BL22 is comprised of an anti-CD22 dsFv linked to truncated PE, and it has been evaluated in a phase 1 clinical trial of patients with B cell malignancies [31] and ongoing in phase 2 clinical trials [30]. Of the 32 patients treated in the phase 1 study, 16 hairy cell leukaemia patients responded with 11 patients having a complete remission and 2 having a partial remission. Neutralizing antibodies were only observed in 4 of the 16 patients. Although VLS was observed in some patients, the dose-limiting toxicity was the cytokine release syndrome. A variety of immuntoxins have undergone clinical testing, and the testing has been comprehensively reviewed by Kreitman [28] and Schaede and Reiter [54].

Currently two immunoconjugates have been approved by the FDA for clinical use in the treatment of cancer. Mylotarg® (Gemtuzumab ozogamicin,

Wyeth-Ayerst Laboratories) is a CD33 antibody conjugated to a calicheamicin, a cytotoxic antibiotic and Ontak® (denileukin diftitox, Ligand Pharmaceuticals, Inc.) is an IL-2 cytokine conjugated to the toxin DT. Both drugs target hematological cancers. While Mylotarg® is not conjugated to a toxin, it provides a good example of the progress of an antibody immunoconjugate from preclinical development to regulatory approval for market.

The targeted antigen, CD33, of Mylotarg® is not expressed in any other species besides humans and large primates; therefore a cross-reactive species was not available for Mylotarg® [9]. Instead, repeated-dose toxicology studies were conducted in Sprague–Dawley rats and cynomolgus monkeys that indicated hepato, renal, and hematopoietic toxicities due to drug administration. Mylotarg® was more immunogenic in rats than monkeys and had a slow clearance rate in both species, three and seven days, respectively. Safety pharmacology conducted in dogs resulted in minor changes in blood pressure and cardiac output that were noted at 16-fold above human starting dose with changes in ECG and heart rate noted at 52-fold above human dose. Reproductive toxicology was not conducted; however, histopathological changes were noted in rat testes and atrophy in the mammary gland. Developmental toxicology studies conducted in rats showed dose-related development effects with decreases in fetal weights, increases in embryo/fetal mortality, and fetal digital malformations, with reabsorptions at the highest dose. The cytotoxic antibiotic used in Mylotarg®, calicheamicin, kills cells by interacting with DNA and causing breakage. Therefore genotoxicity studies were conducted using an in vivo mouse micronucleus assay and confirmed that it was clastogenic. Mylotarg® was determined to have an acceptable toxicity profile in clinical trials and few patients developed antibodies to the drug. Infusion reactions were observed that sometimes occur with monoclonal antibody infusions. Preclinical studies predicted the liver toxicity observed in patients, which for the most part was transient and reversible [9]. The preclinical studies together with the response rate in clinical trials were satisfactory for approval of Mylotarg® for CD33 positive acute myeloid leukemia patients in first relapse who are 60 years of age or older. Since then, additional studies indicate further use of Mylotarg® with combination therapy [58] as well as other population groups [6].

## 6 SUMMARY

Immunotoxins continue to be actively investigated as viable alternatives to conventional therapies for a variety of diseases. An array of different recombinant, antibody formats are now available for use in immunotoxins. While these design changes have improved the overall in vitro and preclinical in vivo efficacy of immunotoxins, increased potency does not address either of the two major concerns for drugs of this type: immunogenicity and toxicity. As such, immunotoxins in their current form may have limited application other than to those disease conditions either where the patients are immunocompro-

mised, as in the case for leukemia, or where the drug can be delivered directly, as was demonstrated for VB4-845 in the treatment of SCCHN and transitional cell carcinoma of the bladder. In order to expand the utility of immunotoxins to achieve a comparable safety profile, design considerations will be required to minimize immunogenicity and toxicity. To this end, pegylation [34,60] or T and/or B cell epitope depletion [24] from the toxin portion of the immunotoxin may be an alternative means to minimize an immune response while the discovery of newer toxins with better safety profiles may minimize nonspecific toxicities. In addition increasing the safety profile of immunotoxins through the discovery of new and more selective tumor targets will only serve to broaden their clinical use in the treatment against cancer.

## REFERENCES


1. Adams GP, Schier R. Generating improved single-chain Fv molecules for tumor targeting. *J Immunol Meth* 1999;231:249–60.
2. Azemar M, Djahansouzi S. Regression of cutaneous tumor lesions in patients intratumorally injected with a recombinant single-chain antibody-toxin targeted to ErbB2/HER2. *Breast Cancer Res Treat* 2003;82:155–64.
3. Baluna R, Coleman E. The effect of a monoclonal antibody coupled to ricin a chain-derived peptides on endothelial cells in vitro: insights into toxin-mediated vascular damage. *Exp Cell Res* 2000;258:417–24.
4. Balzar M, Winter MJ, de Boer CJ, Litvinov SV. The biology of the 17-1A antigen (Ep-CAM). *J Mol Med* 1999;77:699–712.
5. Behr T, Becker W, Hannappel E, Goldenberg DM, Wolf F. Targeting of liver metastases of colorectal cancer with IgG, F(ab′)2, and Fab′ anti-carcinoembryonic antigen antibodies labeled with 99mTc: the role of metabolism and kinetics. *Cancer Res* 1995;55:5777s–85s.
6. Brethon B, Auvrignon A, Galambrun C, Yakouben K, Leblanc T, Bertrand Y, Leverger G, Baruchel A. Efficacy and tolerability of gemtuzumab ozogamicin (anti-CD33 monoclonal antibody, CMA-676, Mylotarg) in children with relapsed/refractory myeloid leukemia. *BMC Cancer* 2006;6:172.
7. Brinkmann U. Recombinant immunotoxins: protein engineering for cancer therapy. *Mol Med Today* 1996;439–46.
8. Brinkmann U. Recombinant antibody fragments and immunotoxin fusions for cancer therapy. *In vivo* 2000;14:21–8.
9. Bross PF, Beitz J, Chen G. Approval summary: gemtuzumab ozogamicin in relapsed acute myeloid leukemia. *Clin Cancer Res* 2001;7:1490–6.
10. Colcher D, Goel A, Pavlinkova G, Beresford GW, Booth B, Batra SK. Effects of genetic engineering on the pharmacokinetics of antibodies. *Quart J of Nucl Med* 1999;43:132–9.
11. DeGeorge JJ, Ahn CH, Andrews PA, Brower ME, Giorgio DW, Goheer MA, Lee-Ham DY, McGuinn WD, Schmidt W, Sun CJ, Tripathi SC. Regulatory considerations for preclinical development of anticancer drugs. *Cancer Chemother Pharmacol* 1998;41:173–85.

12. Di Paolo C, Willuda J, Kubetzko S, Lauffer I, Tschudi D, Waibel R, Pluckthun A, Stahel RA, Zangemeister-Wittke U. A recombinant immunotoxin derived from a humanized epithelial cell adhesion molecule-specific single-chain antibody fragment has potent and selective antitumor activity. *Clin Cancer Res* 2003;9:2837–48.
13. FDA. Points to consider in the manufacture and testing of monoclonal antibody products for human use. *J-Immunother* 1997;20:214–43.
14. Fidias P, Grossbard ML, Lynch T. A phase II study of the immunotoxin N901-blocked ricin in small-cell lung cancer. *Clin Lung Cancer* 2002;3:219–22.
15. FitzGerald D, Pastan I. Pseudomonas exotoxin: recombinant conjugates as therapeutic agents. *Biochem Soc Transact* 1992;20:731–4.
16. FitzGerald DJ, Waldmann TA, Willingham MC, Pastan I. Pseudomonas exotoxin-anti-TAC. cell-specific immunotoxin active against cells expressing the human T cell growth factor receptor. *J Clin Invest* 1984;74:966–71.
17. Food and Drug Administration. Good Laboratory Practice for Nonclinical Laboratory Studes. 21 Code Federal Regulation 1996;58.
18. Food and Drug Administration. Estimating the Safe Starting Dose in Clinical Trial for Therapeutics in Adult Healthy Volunteers, 2002.
19. Frankel AE. Reducing the immune response to immunotoxin. *Clin Cancer Res* 2004;10:13–15.
20. Frankel AE, Kreitman RJ, Sausville EA. Targeted toxins. *Clin Cancer Res* 2000; 6:334.
21. Gould BJ, Borowitz MJ, Groves ES, Carter PW, Anthony D, Weiner LM, Frankel AE. Phase I study of an anti-breast cancer immunotoxin by continuous infusion: report of a targeted toxic effect not predicted by animal studies. *J Natl Cancer Inst* 1989;81:775–81.
22. Haggerty HG, Warner WA. BR96 sFv-PE40 Immunotoxin: Nonclinical safety assessment. *Toxicol Pathol* 1999;27:87–94.
23. Herschman HR, Simpson DL, Cawley DB. Toxic ligand conjugates as tools in the study of receptor-ligand interactions. *J Cell Biochem* 1982;20:163–76.
24. Jones TD, Phillips WJ, Smith BJ, Bamford CA, Nayee PD, Baglin TP, Gaston JS, Baker MP. Identification and removal of a promiscuous CD4+ T cell epitope from the C1 domain of factor VIII. *J Thromb Haemost* 2005;3:991–1000.
25. Kondo K, FitzGerald DJ, Chaudhary V. Activity of immunotoxins constructed with modified pseudomonas exotoxin a lacking the cell recognition domain. *J Biol Chem* 1988;263:9470–5.
26. Kreitman RJ. Immunotoxins in cancer therapy. *Curr Opin Immunol* 1999; 11:570–8.
27. Kreitman RJ. Recombinant toxins for the treatment of cancer. *Curr Opin Mol Ther* 2003;5:44–51.
28. Kreitman RJ. Immunotoxins for targeted cancer therapy. *AAPS J* 2006; 8:E532–51.
29. Kreitman RJ, Pastan I. Accumulation of a recombinant immunotoxin in a tumor in vivo: fewer than 1000 molecules per cell are sufficient for complete responses. *Cancer Res* 1998;58:968–75.
30. Kreitman RJ, Pastan I. BL22 and lymphoid malignancies. *Best Pract Res Clin Haematol* 2006;19:685–99.

31. Kreitman RJ, Wilson WH, Bergeron K, Raggio M, Stetler-Stevenson M, FitzGerald DJ, Pastan I. Efficacy of the anti-CD22 recombinant immunotoxin BL22 in chemotherapy-resistant hairy-cell leukemia. *N Engl J Med* 2001;345:241–7.
32. Krolick KA, Uhr JW, Slavin S, Vitetta ES. In vivo therapy of a murine B cell tumor (BCL1) using antibody-ricin A chain immunotoxins. *J Exp Med* 1982;155: 1797–809.
33. Krolick KA, Villemez C, Isakson P, Uhr JW, Vitetta ES. Selective killing of normal or neoplastic B cells by antibodies coupled to the A chain of ricin. *Proc Natl Acad Sci USA* 1980;77:5419–23.
34. Kubetzko S, Balic E, Waibel R, Zangemeister-Wittke U, Pluckthun A. PEGylation and multimerization of the anti-p185HER-2 single chain Fv fragment 4D5: effects on tumor targeting. *J Biol Chem* 2006;281:35186–201.
35. Kuus-Reichel K, Grauer L. Will immunogenicity limit the use, efficacy and future development of therapeutic monoclonal antibodies. *Clin Diagnost Lab Immunol* 1994;365–72.
36. Levy R, Miller RA. Therapy of lymphoma directed at idiotypes. *J Natl Cancer Inst Monogr* 1990;61–8.
37. Litvinov SV, Velders MP, Bakker HA, Fleuren GJ, Warnaar SO. Ep-CAM: a human epithelial antigen is a homophilic cell-cell adhesion molecule. *J Cell Biol* 1994; 125:437–46.
38. Maynard J, Georgiou G. Antibody Engineering. *Annu Rev Biomed Eng* 2000; 2:339–76.
39. McHugh L, Hu S, Lee BK, Santora K, Kennedy PE, Berger EA, Pastan I, Hamer DH. Increased affinity and stability of an anti-HIV-1 envelope immunotoxin by structure-based mutagenesis. *J Biol Chem* 2002;277:34383–90.
40. McLaughlin PM, Harmsen MC, Dokter WH, Kroesen BJ, van der MH, Brinker MG, Hollema H, Ruiters MH, Buys CH, de Leij LF. The epithelial glycoprotein 2 (EGP-2) promoter-driven epithelial-specific expression of EGP-2 in transgenic mice: a new model to study carcinoma-directed immunotherapy. *Cancer Res* 2001; 61:4105–11.
41. McLaughlin PM, Kroesen BJ, Dokter WH, van der MH, de Groot M, Brinker MG, Kok K, Ruiters MH, Buys CH, De Leij LF. An EGP-2/Ep-CAM-expressing transgenic rat model to evaluate antibody-mediated immunotherapy. *Cancer Immunol Immunother* 1999;48:303–11.
42. Momburg F, Moldenhauer G. Immunohistochemical study of the expression of a Mr 34,000 human epithelium-specific surface glycoprotein in normal and malignant tissues. *Cancer Res* 1987;47:2883–91.
43. Mosolits S, Campbell F. Targeting human Ep-Cam in transgenic mice by anti-idiotype and antigen based vaccines. *Int J Cancer* 2004;112:669–77.
44. Pai LH, Wittes R, Setser A, Willingham MC, Pastan I. Treatment of advanced solid tumors with immunotoxin LMB-1: An antibody linked to Pseudomonas exotoxin. *Nat Med* 1996;2:350–3.
45. Parker BA, Vassos AB, Halpern SE, Miller RA, Hupf H, Amox DG, Simoni JL, Starr RJ, Green MR, Royston I. Radioimmunotherapy of human B-cell lymphoma with 90Y-conjugated antiidiotype monoclonal antibody. *Cancer Res* 1990;50: 1022s–8s.

46. Pastan I. Immunotoxins containing pseudomonas exotoxin A: a short history. *Cancer Immunol Immunother* 2003;52:338–41.
47. Pastan I, FitzGerald D. Pseudomonas exotoxin: chimeric toxins. *J Biol Chem* 1989; 264:15157–60.
48. Pastan I, Hassan R, FitzGerald DJ, Kreitman RJ. Immunotoxin Treatment of Cancer. *Annu Rev Med* 2007;58:221–37.
49. Perentesis JP, Miller SP, Bodley JW. Protein toxin inhibitors of protein synthesis. *Biofactors* 1992;3:173–84.
50. Porter S. Human immune response to recombinant human proteins. *J Pharm Sci* 2001;90:1–11.
51. Proca DM, Niemann TH, Porcell AI, DeYoung BR. MOC31 immunoreactivity in primary and metastatic carcinoma of the liver: report of findings and review of other utilized markers. *Appl Immunohistochem Mol Morphol* 2000;8:120–5.
52. Reiter Y. Recombinant immunotoxins in targeted cancer cell therapy. *Adv Cancer Res* 2001;81:93–124.
53. Saavedra-Lozano J, Cao Y, Callison J, Sarode R, Sodora D, Edgar J, Hatfield J, Picker L, Peterson D, Ramilo O, Vitetta ES. An anti-CD45RO immunotoxin kills HIV-latently infected cells from individuals on HAART with little effect on CD8 memory. *Proc Natl Acad Sci USA* 2004;101:2494–9.
54. Schaedel O, Reiter Y. Antibodies and their fragments as anti-cancer agents. *Curr Pharm Des* 2006;12:363–78.
55. Siegall CB, Chaudhary V. Functional analysis of donaims II, Ib and III of pseudomonas exotoxin. *J Biol Chem* 1989;264:14256–61.
56. Siegall CB, Liggett D. Characterization of vascular leak syndrome induced by the toxin component of pseudomonas exotoxin-based immunotoxin and its protential inhibition with nonsteroidal anti-inflammatory drugs. *Clin Cancer Res* 1997;3: 339–45.
57. Siegall CB, Liggitt D. Prevention of immunotoxin-mediated vascular leak syndrome in rats with retention of antitiumor activity. *Proc Natl Acad Sci USA* 1994; 91:9514–18.
58. Sievers EL, Larson RA, Stadtmauer EA, Estey E, Lowenberg B, Dombret H, Karanes C, Theobald M, Bennett JM, Sherman ML, Berger MS, Eten CB, Loken MR, van Dongen JJ, Bernstein ID, Appelbaum FR. Efficacy and safety of gemtuzumab ozogamicin in patients with CD33-positive acute myeloid leukemia in first relapse. *J Clin Oncol* 2001;19:3244–54.
59. Smallshaw JE, Ghetie V, Rizo J, Fulton JRJ, Trahan L, Ghetie MA, Vitetta E. Genetic engineering of an immunotoxin to eliminate pulmonary vascular leak in mice. *Nat Biotechnol* 2003;21:387–91.
60. Tsutsumi Y, Onda M, Nagata S, Lee B, Kreitman RJ, Pastan I. Site-specific chemical modification with polyethylene glycol of recombinant immunotoxin anti-Tac(Fv)-PE38 (LMB-2) improves antitumor activity and reduces animal toxicity and immunogenicity. *Proc Natl Acad Sci USA* 2000;97:8548–53.
61. Uchida T, Pappenheimer AM, Jr, Harper AA. Diphtheria toxin and related proteins. II. Kinetic studies on intoxication of HeLa cells by diphtheria toxin and related proteins. *J Biol Chem* 1973;248:3845–50.

62. van Oosterhout YV, Van EL, Schattenberg AV, Tax WJ, Ruiter DJ, Spits H, Nagengast FM, Masereeuw R, Evers S, de WT, Preijers FW. A combination of anti-CD3 and anti-CD7 ricin A-immunotoxins for the in vivo treatment of acute graft versus host disease. *Blood* 2000;95:3693–701.
63. van Vuuren AJ, Van Roon JA, Walraven V, Stuij I, Harmsen MC, McLaughlin PM, Van de Winkel JG, Thepen T. CD64-directed immunotoxin inhibits arthritis in a novel CD64 transgenic rat model. *J Immunol* 2006;176:5833–8.
64. White CA, Weaver RL, Grillo-Lopez AJ. Antibody-targeted immunotherapy for treatment of malignancy. *Annu Rev Med* 2001;52:125–45.
65. Willuda J, Honegger A, Waibel R. High thermal stability is essential for tumor targeting of antibody fragments: engineering of a humanized anti-epithelial glycoprotein-2 (epithelial cell adhesion molecule) single-chain Fv fragment. *Cancer Res* 1999;59:5758–67.

# Preclinical Safety Evaluation of Blood Products

RICHARD M. LEWIS, PhD

**Contents**

1 Introduction
2 Hemoglobin-Based Oxygen Carriers
3 Therapies for Diseases of Genetically Deficient Proteins
4 Immune Globulins
5 Coagulation Factors for Replacement Therapy
6 Immunogenicity
7 Recombinant Biotechnology-Derived Coagulation Factors
8 Gene Therapy for the Treatment of Coagulation Disorders
9 Conclusion
References

## 1 INTRODUCTION

The historical development of blood products, in particular, plasma derivatives, reflects the development of "traditional" biological products. The early products were derived from donated blood. Their development through the 1960s and '70s consisted primarily of refining purification techniques. Because they are necessarily sourced from human materials, the issue of disease transmission has always been a concern. Along with improvements in purification came better donor selection, better donor testing, and manufacturing methods designed to separate and inactivate infectious agents. Later the implementation of manufacturing steps introduced specifically to reduce possible viral contaminants became important. The biotechnology revolution using cell culture methods allowed the consistent production of large amounts of

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

products without the concern of human-derived infectious agents. The field has further progressed with advancing science to the consideration of gene therapy applications for many of the disorders that were previously treated using replacement therapy. Although current products under development primarily address diseases associated with single gene mutations or a single protein's lack of activity, the possibility of curing individuals using a biological therapy seems possible.

The Office of Blood Research and Review (OBRR) in CBER has a long history of review and approval of biological products. Many of the earliest blood products were not evaluated using what is now considered to be conventional preclinical and toxicological evaluations. This is in part because the Bureau of Biologics and its predecessor, Division of Biologics Control (transferred from NIH to FDA in 1972) [1] regulated products under a different mandate and in part because the blood products are often meant to replace or supplement levels of existing biologic factors. Albumin, fibrinogen, and gamma globulin were first purified from plasma using ethanol fractionation described by Cohn in 1940 [2]. It has been the assumption that the supplementing of normal plasma derivatives, with material isolated from plasma, does not cause classical toxicity. That is not to say that these products were not evaluated for harmful effects but that evaluation was most often linked to determination of effects of inherent activity or activity associated with changes in molecular structure. Changes such as cleavage of proteins in the product could result in generation of enzymatic activity. For example, such activity might subsequently result in lack of stability, increase in pro-thrombotic potential, or other effects caused by active fragments. An important case of such an incident was the association of hypotension with the administration of purified protein derivative, a product composed primarily of albumin [3]. Currently such molecular changes are evaluated and, as in all biological products, the issue of immunogenicity is also addressed and will be discussed.

The use of animals for the preclinical evaluation of blood derivatives primarily encompassed evaluation for activity [4], and later for viral contamination, but little from the perspective of actual toxicological endpoints. Many of the hemophiliacs who were treated with early versions of antihemophilic factor (AHF) and Factor IX became infected with hepatitis [5]. This provided the major reason for developing viral inactivation methods for AHF concentrates. The hepatitis agents were referred to as non-A, non-B hepatitis (NANB). The preclinical demonstration that the active virus had been inactivated required the use of chimpanzees, which were injected with the AHF concentrate. During the early 1980s plasma fractionators used chimpanzee studies to demonstrate the effectiveness for the reduction of HBV and NANB hepatitis infectivity [6]. Ultimately, previously untreated patients were evaluated in clinical trials that demonstrated the utility of a number of viral reduction methods for what was by then known as hepatitis C [7].

More recently some of the focus of preclinical evaluation of blood products has shifted to toxicological evaluations. This has come with the development

of recombinant materials or biochemically modified molecules to address specific disease states. In the case of the hemoglobin-derived oxygen carriers, with extensive toxic characteristics, many aspects of preclinical assessment have been employed as will be discussed below.

Conventional preclinical studies are those using principles similar to those for toxicological assays of traditional drug products and used to determine such parameters as absorption, distribution, metabolism and excretion (ADME), no adverse effect level (NOAEL), maximum tolerated dose (MTD) levels, and determination of relevant animal species. These studies are generally conducted in concentrations based on multiples of the proposed human dose. Evaluation endpoints usually include mortality, clinical signs, body weight, food consumption, clinical chemistry, hematology, gross pathology, and histopathology.

The FDA refers to a few official documents for guidance in accomplishing a valid preclinical development program. These include International Conference on Harmonization (ICH) accepted guidances as well as US FDA guidance documents:

- Guidance for Industry: S6 Preclinical Safety Evaluation of Biotechnology-Derived Pharmaceuticals (1997)
- Guidance for Industry: M3 Nonclinical Safety Studies for the Conduct of Human Clinical Trials for Pharmaceuticals (1997)
- Guidance for Industry and Reviewers: Estimating the Safe Starting Dose in Clinical Trials for Therapeutics in Adult Healthy Volunteers

Additional biological product specific advice is provided in other documents:

- Points to Consider in the Manufacture and Testing of Monoclonal Antibody Products for Human Use (1997)
- Guidance for Industry: Guidance for Human Somatic Cell Therapy and Gene Therapy (1998)
- Guidance for Industry and Reviewers: Estimating the Safe Starting Dose in Clinical Trials for Therapeutics in Adult Healthy Volunteers

The ICH S6 document should be used in conjunction with the ICH M3 guidance, which provides the appropriate timing and duration of general toxicology studies as they relate to the proposed clinical trials. The principles and use of these various guidance documents are discussed in depth elsewhere in this volume. The intent in this chapter is to describe some of the preclinical programs that are publicly known and to offer suggestions for the development of plans for various groups of products regulated in the Office of Blood and Blood Products, the CBER, and the FDA.

## 2 HEMOGLOBIN-BASED OXYGEN CARRIERS

Although often referred to as blood substitutes, the major intent of hemoglobin-derived products has been for use as oxygen carriers. The hemoglobin-based oxygen carriers (HBOCs) represent a unique, and what has proved to be a difficult, class of blood product. The majority of products in this class have used the oxygen-binding characteristic of hemoglobin while attempting to avoid the toxicities associated with the molecule when not presented in the red cell. Early clinical studies using stroma-free human hemoglobin showed renal dysfunction, hypertension and severe abdominal pain [8]. This class of products demonstrated extreme toxicity in initial studies and changes to formulation and molecular structure have been made in attempts to address the toxicity [9]. One of the primary challenges to the developers of novel hemoglobin products is this demonstration of lack of toxicity.

Because there are a variety of methods that attempt to address the toxicities of hemoglobin [10], each novel material requires investigations into all of the various toxicities. Since it is known that hemoglobin causes renal, neural, and cardiac toxicity, a novel modification that is intended to remove or reduce these toxicities must be tested to address the effects on each of these target organs.

Although the HBOCs have attempted to find utility in different clinical settings where the risk–benefit profiles are different than for use as "blood substitutes," many aspects of the toxicities must still be addressed. For all of these reasons the most extensive requirements for preclinical safety information for therapeutics regulated as blood products have been for the hemoglobin-derived products. Because the hemoglobin-based oxygen carriers are viewed as a class, the FDA has taken the approach that each hemoglobin therapeutic candidate should address the issues of the others. The FDA draft guidance lists numerous toxicities that should be addressed [11]. These include vasoactivity, gastrointestinal toxicity, neurotoxicity, and cardiac toxicity as well as activity in vascular tone, oxidative stress, synergy with endotoxin, and enzymatic indications of pancreatic and liver toxicity. The guidance is very prescriptive regarding physicochemical characterization, estimation of in vitro activity, as well as some of the animal models that should be used to address particular toxicities. As in most preclinical safety tests, doses sufficient to observe toxicity are suggested, and at least one study in large animals should be used. The FDA draft guidance also encourages the use of animal models of disease to address many of the possible toxicities for a "complete safety profile." For example, a fully instrumented animal model study resembling the proposed clinical use should be used to measure cardiac, renal, and pulmonary function. Also recommended is an extensive list of laboratory data to be collected during these studies, such as effects on microvasculature, enzymes, histological and physical tests to evaluate renal function, and a collection of clinical chemistry parameters. One example of a preclinical program included studies on cardiovascular effects (3 species), genotoxicity (3 in vitro models), gastrointestinal effects

(3 species), hemorrhagic shock (3 species), hemostasis (1 specie), toxicology (5 studies in 3 species), oxygen delivery (1 specie), PK (1 specie), renal effects (1 specie), and reproductive toxicology (1 specie).

## 3 THERAPIES FOR DISEASES OF GENETICALLY DEFICIENT PROTEINS

As mentioned above, one of the primary uses of plasma-derived products is to replace missing proteins, in particular, genetic deficiencies such as hemophilia or immune deficiency. These products are administered intravenously. The development program for most of these products has not included extensive preclinical animal evaluation. Often animal models have been used to demonstrate product effectiveness prior to clinical trials. However, other more recent development programs have included toxicology assessments. For example, in the case of one particular non-enzyme plasma-derivative product deficient in some individuals, the toxicology program that was recommended included single- and multiple-dose toxicity studies. These were to be evaluated in rodents. The single-dose study incorporated a determination of the maximum tolerated dose as well as clinical observations and complete histopathology.

Another program recently recommended for a product with novel modifications included single-dose studies in a rodent and a nonrodent as well as another study in a higher mammal as a means to enter phase 1 trials. However, multiple-dose animal studies would be necessary and could probably be done in parallel to the single-dose human trials.

Although coagulation factor deficiency such as factor VIII and Factor IX is found only in males, any product that might be administered to females should include reproductive toxicology studies and follow the ICH S5A outline for segment 2 evaluation. Because deficient products are normally proteins, fertility and perinatal reproductive toxicology studies can be seen scientifically as extremely unlikely to provide any information. Teratology studies might reveal effects on organogenesis that occur as a result of complement or coagulation system changes or immunologic mediators, so such studies are recommended.

Homologous recombinant proteins have had to undergo a more extensive evaluation preclinically. For example, a recombinant factor VIII product was evaluated in acute and subacute studies in mice, rats, rabbits, dogs, and nonhuman primates over a range of concentrations and dose frequency [12].

## 4 IMMUNE GLOBULINS

Although there is no specific FDA document that provides guidance for preclinical studies for immune globulins, there is a draft guidance for safety,

efficacy, and pharmacokinetic clinical trials to support licensure of intravenous immune globulin [13]. The guidance recommends that in clinical trails safety data be collected to include capture of adverse effects (AEs) regardless of determination of association with the product (temporal association is primary consideration). The characteristics of the particular infusion associated with an AE, as well as statistical consideration for evaluation of all AEs, should be included in any marketing application.

The efficacy of the product should be determined using the subject's bacterial infection rates and the pharmacokinetics of product, including the relationship of PK and infection. The ability of many of these products to block infection has been demonstrated.

In the future the FDA may develop guidance for preclinical toxicology studies but one has not been published at this time. The guidance recommending design of clinical studies can be used to model animal studies to support the various clinical trials.

A recent presentation by an FDA reviewer noted that for a number of immune globulin products surveyed, no preclinical animal studies were presented in the INDs [14]. The immune globulins surveyed included immune globulin (intravenous), as well as hyperimmune globulins both intravenous and intramuscular forms. This situation is likely due to the long history of these products and the similarities in their manufacturing.

Regarding the use of hyperimmune globulins such as used for treatment of bacterial, viral, or toxin diseases, the safety data from other immune globulins from the same manufacturer provide most of the information and assurance of safety. If one manufacturer has an approved immune globulin, the material isolated for a hyperimmune most likely will be manufactured in the same way except that a larger portion of the source plasma will have the specific immune globulin. As such, there is a high degree of confidence in the safety of the material from the previous nonspecific immune globulin.

In contrast, a literature report of a new IGIV discusses limited animal studies in preparation for human trials [15]. The development of this product, produced by Baxter, has used a series of preclinical animal testing methods to address concerns for the use of a plasma derivative. Although it would not be considered a "classical" toxicology program, acute toxicity was evaluated along with the application of a number of models in order to gain insight into the in vivo characteristics of the product. An in vivo assay was used to show protection from *Klebsiella pneumoniae* and *Streptococcus pneumoniae*. Bronchospastic activity in anesthetized guinea pigs was chosen to evaluate the potential for anaphylactoid reactions after rapid arterial injection, and the potential for changes in blood pressure was evaluated in the hypotensive rat model. Dog studies were used to evaluate possible cardiovascular, respiratory, and coagulation reactions. Thrombogenicity was measured in rabbits. Pharmacokinetic characteristics were studied in rats. Acute toxicity studies in mice and rats were used to determine the NOAEL.

## 5 COAGULATION FACTORS FOR REPLACEMENT THERAPY

As mentioned above, preclinical pharmacology and toxicology for plasma derivatives has not included preclinical assessments commonly employed for small-molecule drug products. Most preclinical development involves comparative pharmacokinetics and the use of animal models has been an important component of this evaluation. The coagulation factors have shown the potential for efficacy in animal models such as the von Willebrand swine and hemophilic dog [16,17].

These animal models of disease should be considered early on as potentially providing some of the data necessary for initiation of human studies. Besides utility as proof of concept, they can add to understanding dose response as well as help evaluate some safety endpoints. New products resulting from improved manufacturing can be compared with previously produced material using pharmacokinetic parameters.

In one "second-generation" recombinant factor VIII product, ReFacto®, hemophilic dogs were used to demonstrate utility for bleeding time correction, association with von Willebrand's factor, and evaluating pharmacokinetics [18]. A plasma-derived factor VIII was used for comparison. In another study that used a dog model, a recombinant von Willebrand factor was evaluated [19]. Pharmacokinetic data were collected as well as pharmacodynamic measurements of platelet aggregation support (ristocetin cofactor activity) and cuticle wound blood flow. An important component of these studies was the suitability of the model. These models were chosen because of the biochemical deficiency of the particular factors and the parallel clinical syndromes. Such in vivo data can help in determining activity and dosing when such a product is first used in human trials. The Refacto® molecule was also studied in rats and monkeys to determine its no observed adverse effect level, that was more than 10 times normal circulating levels. The major toxicity observed was the development of antibodies to the molecule that blocked activity and resulted in an acquired hemophilia syndrome. Similar findings were demonstrated when plasma-derived material was injected into monkeys [20].

## 6 IMMUNOGENICITY

Current regulatory concerns with protein therapies include the possibility of induction of autologous antibodies that interact with the therapeutic agent or, worse, with the subject's own residual analogous factor [21–23]. With the use of some bovine thrombin products, there has in fact been an association with coagulopathies and autologous antibodies [24–26]. Therefore the immunogenic potential needs to be addressed in the development program of every new product as well as manufacturing changes or with changes in excipient. The concern of immunogenicity has been expressed by US and European

authorities and the Australian authorities have accepted the EMEA position [27–29]. Even prior to the development of recombinant affector proteins, an important consideration for the safety of the subjects using the plasma-derived coagulation factor therapies was the development of "inhibitors" to coagulation enzymes [30–32]. The development of these inhibitors, so named because of their ability to block activity and not initially known to be antibodies, was a serious event. The induction of this immune response was tragic for hemophiliacs because it drastically changed the ability to treat bleeding episodes or to provide prophylactic therapy [33].

Because of the unknown component of biotechnologically derived factors, evaluation of the development of such inhibitors was an important concern in their development [34,35]. In some cases, the hematopoietic compounds induced autologous antibodies, not only toward the biological drug but also toward the patient's own biological factor, leaving them with severe anemia [17–19].

Major regulatory agencies have developed guidance to specifically address immunogenicety in preclinical studies, if possible. Certainly sponsors should collect and present data on immunogenicity from clinical trials [36]. The FDA has particularly addressed the issue of antibodies to clotting factors at a public meeting [12]. The discussion at that meeting suggested possible future use of mouse models for inhibitor testing, including transgenic mice and hemophilia mice. In other venues the FDA has encouraged the development of in vitro immunological response testing [37]. Other methods for establishing an estimate of immunogenicity have included the use of immunochemical (ELISA) assays to compare response and to evaluate cross-reactivity for products that have changed formulation or manufacturing. The FDA has accepted the use of such immunochemical evaluations in some cases as part of product characterization. However, to date, the best method of detecting changes in immunogenicity of novel products or improved manufacturing is a clinical study with previously treated patients. Certainly some estimation of antigenicity should be done preclinically. Immunogenicity has also been a concern for gene therapy studies as mentioned below (also see Chapters 16 and 20 in this volume addressing immunogenicity).

## 7 RECOMBINANT BIOTECHNOLOGY-DERIVED COAGULATION FACTORS

The first therapeutic recombinant DNA-derived coagulation protein licensed by the FDA was factor VIII for treatment of hemophilia A in 1992 [38]. This step forward was a landmark in hemophilia therapy. This biotechnology process reduced the theoretical risk of human-derived viruses and seemed to provide for an unlimited market supply although other human and animal proteins were often used in the manufacturing and formulation of many recombinant

products. Plasma-derived factors are still manufactured for a number of reasons, including cost and availability, but the use of recombinant factors is recommended by the Medical and Scientific Advisory Council of the Hemophilia Foundation [39]. The current virus reduction methods for plasma-derived concentrates are considered extremely safe and have virtually eliminated product-derived infections of HIV and hepatitis [40]. Additional manufacturing methods are also effective in reducing the chances of a parvovirus infection. Nevertheless, the use of recombinant coagulation factors has revolutionized not only manufacturing but patient therapy [41–43].

As discussed above, animal models have played an important part in characterizing recombinant molecules. Nevertheless, more classical toxicological evaluations are being applied in the preclinical setting. Research continues on the recombinant factor VIII and Factor IX using molecular techniques to improve the molecules as therapeutics. Modifications that increase activity, increase half-life, and increase expression are being investigated. In fact modifications that would make these genes better candidates for gene therapy are also being evaluated [44–46].

## 8 GENE THERAPY FOR THE TREATMENT OF COAGULATION DISORDERS

The concept of a cure for hemophilia has been a goal of scientists and practitioners for many years and studies are currently underway to address this goal [47]. As in the history of these products, the use of animal models as a proof of concept has been central to preclinical testing [48,49]. The specific aspects of the toxicology evaluation of gene therapy, in general, is addressed in Chapter 32, but some discussion of these animal models as they relate to hemophilia is pertinent here. As in all current studies, the issue of immunogenicity is an important consideration whose evaluation should be included in gene therapy product studies. In animal models, some have demonstrated the induction of an immune reaction to the expressed protein even after delivery of the gene by adenovirus [50]. Increased expression, decreased toxicity from viral elements, and decreased immunogenicity of the expressed proteins had been a formidable goal and has shown some progress as evaluated in animal models [51–54].

## 9 CONCLUSION

The preclinical assessment of blood products has been based on studies designed to answer specific questions on product-specific attributes with a consideration of the intended patient population. In many cases relevant animal models of the disease have been used to assess safety in addition to providing proof-of-concept information to support clinical development.

## REFERENCES


1. Bren L. The Road to the Biotech Revolution: Highlights of 100 Years of Biologics Regulation, 2006. FDA Consumer. http://www.fda.gov/fdac/106_toc.html
2. Cohn EJ, Strong LE, Hughes WI, et al. Preparation and properties of serum and plasma proteins. IV. A system for the separation into fractions of the protein and lipoprotein components of biological tissue and fluids. *J Am Chem Soc* 1946; 68:459–75.
3. Alving BM, Hojima Y, Pisano JJ, Mason BL, Buckingham RE, Mozen MM, et al. Hypotension associated with prekallikrein activator (Hageman factor fragments) in plasma protein fraction. *N Engl J Med* 1978;299:66–70.
4. Graham JB, Buckwalter JA, Hartley LJ, Brinkhous KM. Canine hemophilia. Observations of the course, the clotting anomaly, and the effects of blood transfusions. *J Exp Med* 1949;90:97–111.
5. Mannucci PM. AIDS, hepatitis and hemophilia in the 1980s: memoirs from an insider. *J Thromb Haemost* 2003;1:2065–9.
6. Leveton L, Sox HC, Stoto MA. HIV and the blood supply: an analysis of crisis decision-making. Executive summary. *Transfusion* 1996;36:919–27.
7. Brettler DB, Alter HJ, Dienstag JL, Forsberg AD, Levine PH. Prevalence of hepatitis C virus antibody in a cohort of hemophilia patients. *Blood* 1990;76: 254–6.
8. Savitsky J, Doczi J, Black J, Arnold J. A clinical safety trial of stroma-free hemoglobin. *Clin Pharmacol Ther* 1978;23:73–80.
9. Stowell CP. What ever happened to blood substitutes? *Transfusion* 2004; 44:1403–4.
10. Buehler PW, Alayash AI, Toxicities of hemoglobin solutions: in search of in vitro and in vivo model systems. *Transfusion* 2004;44:1516–30.
11. Guidance for Industry. Criteria for Safety and Efficacy Evaluation of Oxygen Therapeutics as Red Blood Cell Substitutes. Draft Guidance. http://www.fda.gov/cber/gdlns/oxytherbld.htm
12. FDA, CBER, Workshop on Factor VIII Inhibitors, 2003. http://www.fda.gov/cber/minutes/fctrviii112103t.htm
13. Guidance for Industry. Safety, Efficacy, and Pharmacokinetic Studies to Support Marketing of Immune Globulin Intravenous (Human) as Replacement Therapy for Primary Humoral Immunodeficiency. Draft Guidance. http://www.fda.gov/cber/gdlns/igivimmuno.pdf
14. Preclinical Data to Support Human Studies of Immune Globulin Products. Annual FDA Science Forum, April 2006. http://www.accessdata.fda.gov/scripts/oc/sciencefforum/sf2006/Search/preview.cfm?abstract_id=952&backto=author
15. Teschner W, Butterweck HA, Auer W, Muchitsch EM, Weber A, Schwarz H-P. Biochemical and Pre-clinical Characterization of a New10% Liquid Triple Virally Reduced Human Intravenous Immune Globulin (IGIV, 10% TVR). http://www.bo-conf.com/ppb05/PPB05_reports.pdf
16. Giles AR. Use of an animal model to evaluate hemophila therapies, 184. *Prog Clin Biol Res* 1984;150:265–76.

17. Giles AR, Tinlin S, Hoogendoorn H, Fournel MA, Ng P, Pancham N. In vivo characterization of recombinant factor VIII in a canine model of hemophilia a (factor VIII deficiency). *Blood* 1988;72:335–9.
18. Brinkhous K, Sandberg H, Widlund L, Read M, Nichols T, Sigman J, et al. Preclinical pharmacology of albumin-free b-domain deleted recombinant factor VIII. *Sem Thromb Hemostasis* 2002;28:269–72.
19. Schwarz HP, Dorner F, Mitterer A, Mundt W, Schlokat U, Pichler L. Turecek PL. Preclinical evaluation of recombinant von Willebrand factor in a canine model of von Willebrand disease. *Wien Klin Wochenschr* 1999;111:181–91.
20. Summary Basis for Approval, 2000. Antihemophilic Factor (recombinant) Refacto®. http://www.fda.gov/cber/sba/ahfgen030600S.pdf
21. Bunn HF. Drug-induced autoimmune red-cell aplasia. *N Engl J Med* 2002;346: 522–3.
22. Casadevall N, Nataf J, Viron B, Kolta A, Kiladjian JJ, Martin-Dupont P, Michaud P, Papo T, Ugo V, Teyssandier I, Varet B, Mayeux P. Pure red-cell aplasia and antierythropoietin antibodies in patients treated with recombinant erythropoietin. *N Engl J Med* 2002;346:469–75.
23. Gershon SK, Luksenburg H, Cote TR, Braun MM. Pure red-cell aplasia and recombinant erythropoietin. *N Engl J Med* 2002;346:1584–6.
24. Zehnder JL, Leung LL. Development of antibodies to thrombin and factor V with recurrent bleeding in a patient exposed to topical bovine thrombin. *Blood* 1990;76:2011–6.
25. Rapaport SI, Zivelin A, Minow RA, et al. Clinical significance of antibodies to bovine and human thrombin and factor V after surgical use of bovine thrombin. *Am J Clin Pathol* 1992;97:84–91.
26. Lawson JH, Lynn KA, Vanmatre RM, et al. Antihuman Factor V Antibodies after use of relatively pure bovine thrombin. *Ann Thorac Surg* 2005;79:1037–8.
27. Guideline on Comparability of Medicinal Products containing Biotechnology-Derived Proteins as Active Substance: Non-clinical and Clinical Issues. http://www.tga.gov.au/docs/pdf/euguide/emea/309702en.pdf
28. Committee for Medicinal Products for Human Use (CHMP), 2003. Guideline on Comparability of Medicinal Products Containing Biotechnology-Derived Proteins as active Substance: Quality Issues. http://www.emea.eu.int/pdfs/human/bwp/320700en.pdf
29. Committee for Medicinal Products for Human Use (CHMP), 2006. Concept Paper on Guideline on Immunogenicity assessment of Therapeutic Proteins. http://www.emea.eu.int/pdfs/human/biosimilar/24651105en.pdf
30. Schwarzinger I, Pabinger I, Korninger C, et al. Incidence of inhibitors in patients with severe and moderate hemophilia A treated with factor VIII concentrates. *Am J Hematol* 1987;24(3):241–5.
31. Ehrenforth S, Kreuz W, Scharrer I, et al. Incidence of development of factor VIII and factor IX inhibitors in hemophiliacs. *Lancet* 1992;339(8793):594–8.
32. Peerlinck K, Arnout J, Gilles JG, Saint-Remy JM, Vermylen J. A higher than expected incidence of factor VIII inhibitors in multitransfused haemophilia A patients treated with an intermediate purity pasteurized factor VIII concentrate. *Thromb Haemost* 1993;69:115–18.

33. Key N. Inhibitors in congenital coagulation disorders. *Br J Haem* 2004;127: 79–391.
34. Lusher JM, Arkin S, Abildgaard CF, et al. Recombinant factor VIII for the treatment of previously untreated patients with hemophilia A: safety, efficacy, and development of inhibitors. *N Engl J Med* 1993;328(7):453–9.
35. Scharrer I, Bray GL, Neutzling O. Incidence of inhibitors in haemophilia A patients—a review of recent studies of recombinant and plasma-derived factor VIII concentrates. *Haemophilia* 1999;5(3):145–54.
36. Committee for Medicinal products for Human Use (CHMP), 2006. Concept paper on Guideline on Immunogenicity assessment of Therapeutic Proteins. Doc. Ref. EMEA/CHMP/BMWP/246511/2005. http://www.emea.eu.int/pdfs/human/biosimilar/24651105en.pdf
37. Comparability Studies for Human Plasma-Derived Therapeutics, 2002. http://www.fda.gov/cber/minutes/plasma053002.htm
38. Lusher J, Abildgaard C, Arkin S, Mannucci PM, Zimmermann R, Schwartz L, Hurst D. Human recombinant DNA-derived antihemophilic factor in the treatment of previously untreated patients with hemophilia A: final report on a hallmark clinical investigation. *J Thromb Haemost* 2004;2:574–83.
39. Medical and Scientific Advisory Council of the Hemophilia Foundation, Recommendation #177. http://www.hemophilia.org/NHFWeb/MainPgs/MainNHF.aspx?menuid=57&contentid=693
40. Tabor E. The epidemiology of virus transmission by plasma derivatives: clinical studies verifying the lack of transmission of hepatitis B and C viruses and HIV type 1. *Transfusion* 1999;39:1160–8.
41. Kurachi K. Recombinant antihemophilic factors. *Biotechnology* 1991;19:177–95.
42. Limentani SA, Roth DA, Furie BC, Furie B. Recombinant blood clotting proteins for hemophilia therapy. *Semin Throm Hemost* 1993;19:62–72.
43. Lusher JM. Recombinant clotting factor concentrates. *Baillieres Clin Haematol* 1996;9:291–303.
44. Saenko E, Ananyeva N, Shima N, et al. The future of recombinant coagulation factors. *J Thromb Haem* 2003;1:922–30.
45. Pipe S. Coagulation factors with improved properties for hemophilia gene therapy. *Semin Thromb Hemost* 2004;30:227–37.
46. Pipe S. The promise and challenges of bioengineered recombinant clotting factors. *J Thromb Haemost* 2005;3:1692–701.
47. High K. Clinical gene transfer studies for hemophilia B. *Semin Thromb Hemost* 2004;30:257–67.
48. VandenDriessche T, Collen D, Chuah M. Gene therapy for the hemophilias. *J Thromb Haem* 2003;1:1550–8.
49. Scallan C, Lillicrap D, Jiang H, et al. Sustained phenotypic correction of canine hemophilia A using an adeno-associated viral vector. *Blood* 2003;102:2031–7.
50. Chuah M, Schiedner G, Thorrez L, et al. Therapeutic factor VIII levels and negligible toxicity in mouse and dog models of hemophilia A following gene therapy with high-capacity adenoviral vectors. *Blood* 2003;101:1734–43.

51. Thorrez L, VandenDriessche L, Collen D, Chuah M. Preclinical gene therapy studies for hemophilia using adenoviral vectors. *Semin Thromb Hemost* 2004; 30:173–83.
52. Rawle F, Lillicrap D. Preclinical animal models for hemophilia gene therapy: predictive value and limitations. *Semin Thromb Hemost* 2004;30:205–13.
53. Chuah M, Collen D, Vandendriessche T. Preclinical and clinical gene therapy for haemophilia. *Haemophilia* 2004;10(s4):119–25.
54. Nathwani A, Davidoff A, Tuddenham E. Prospects for gene therapy of haemophilia. *Haemophilia* 2004;10:309–18.

# Preclinical Safety Evaluation of Viral Vaccines

A. MARGUERITE DEMPSTER, PhD, DABT, and RICHARD HAWORTH, FRCPath, DPhil

**Contents**

1 Introduction
2 Background
3 Current Technologies
  3.1 Recombinant Protein Vaccines
  3.2 Virus-like Particles
  3.3 DNA Vaccines
  3.4 Virus-Based Vectors
  3.5 Mixed Modality Paradigms
4 Adjuvants and Immunomodulators
5 Considerations for Preclinical Safety Evaluation
  5.1 Quality Control Testing of Material Intended for GLP Studies
  5.2 Testing for Viral Contamination
  5.3 Immunogenicity Evaluation within Toxicity Studies
  5.4 Species Selection
  5.5 Design of Safety Assessment Studies
6 Case Studies for Potential HIV Therapies
  6.1 Synthetic Peptide
  6.2 DNA Plasmid
  6.3 Mixed Modality
References

## 1 INTRODUCTION

For nearly a century vaccines have proved to be one of the most important and effective medical treatments. Because of successful vaccination the

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

smallpox virus has been virtually eradicated. In addition to smallpox, vaccines continue to prevent millions of deaths worldwide and protect children from a variety of serious diseases such as diphtheria, pertussis, measles, tetanus, rubella, and mumps [1]. As effective as currently marketed vaccines are, diseases caused by other viruses such as human immunodeficiency virus type 1 (HIV-1), human papillomavirus (HPV), and hepatitis C virus (HCV) are not presently controlled by vaccination, and they clearly represent a significant unmet medical need. For example, according to the United Nations program on AIDS, approximately 40 million people worldwide were living with HIV by the end of 2006, approximately 4 million were newly infected and 3 million had lost their lives to AIDS [2]. It is believed that at least 20 million people have died of AIDS since the first cases of AIDS were identified in 1981 [2,3].

The purpose of this chapter is to review and discuss the preclinical safety evaluation strategy for vaccine approaches to the prophylaxis and treatment of viral diseases. This chapter will discuss the newer approaches to vaccination and will include recombinant proteins, peptides, polysaccharides, DNA plasmids, and viral vectors with and without adjuvants. It is outside the scope of this chapter to discuss whole cells expressing immunogens, live attenuated viruses, bacteria, or parasites.

## 2 BACKGROUND

Billions of doses of vaccines were safely administered over the last century to millions of infants, children, and adults [1,4]. Vaccine-associated adverse events were nevertheless identified in a minority of patients. Many of the adverse events were directly related to the vaccine; others as in the case of MMR have not been substantiated with data but still cause considerable public unease [1]. One documented case is that of vaccination with a formalin-inactivated respiratory syncytial virus (RSV) that, when given prophylactically, caused enhanced disease after exposure to the virus and resulted in the death of some children [1,4,5]. An increased incidence of Guillain–Barré syndrome appeared to be associated with a swine flu vaccine that was tested in humans in the late 1970s [1,6]. Guillain-Barré syndrome has also been associated with other vaccines including those for polio, measles, and hepatitis B [6].

Successful vaccination requires both aspects of adaptive (or acquired) immunity: specificity and memory. The primary purpose of a vaccine is to prime the host's immune system with an infectious agent that has been modified so that it is no longer infectious but still immunogenic enough that the host can mount an immune response before the agent produces its adverse effects [7]. Prior to advances in molecular technology that allowed for genetic manipulation of the virus and/or toxoids, the main types of vaccines developed were live attenuated vaccines, inactivated vaccines, and chemically altered exotoxins [7]. The virulence of live attenuated vaccines (e.g., oral polio virus and cholera) was eliminated by various methods, including treatment with heat

and chemicals or enzymes, or through multiple passages of the virus in culture [1,7]. Although effective and safe in most individuals, these vaccines were not recommended for people with impaired immune systems, and in rare cases reversion to virulence was noted. For example, after the introduction of oral polio virus sporadic cases of vaccine-associated paralytic polio were observed. It was thought that the few attenuating mutations present in one vaccine strain caused the live polio virus to revert to virulence [1]. However, it was a rare occurrence, and given the risk of polio and the perceived benefit, the vaccine continued to be used until it was replaced by an inactivated or killed vaccine. Although the killed vaccines were associated with fewer adverse effects, they were less effective [7]. The third major type of vaccine at that time was formalin-inactivated toxoids (e.g., pertussis, diphtheria, and tetanus). Since these vaccines, such as those against whooping cough, developed in the 1940s demonstrated good efficacy and safety, they were combined with diphtheria and tetanus toxoids. Similar to the single vaccines, these vaccines exhibited a good safety profile. However, some adverse reactions such as seizures and infantile spasms, observed in the 1970s, caused sufficient public concern that the use of these vaccines declined dramatically [1,8,9]. Not surprisingly, their decline in use was associated with an increased incidence in those diseases [1].

The development of recombinant DNA, large-scale cell culture technologies, and advances in synthetic DNA and protein chemistry introduced a wide range of potential pharmaceutical products including cytokines, receptor agonists/antagonists, hormones, growth factors, monoclonal antibodies, and gene therapy products. These advances enabled the development of the next generation of vaccines: recombinant protein, DNA plasmid, and recombinant viral vectors.

## 3 CURRENT TECHNOLOGIES

### 3.1 Recombinant Protein Vaccines

Improvements in genetic engineering facilitated the identification, construction, and production of novel recombinant antigens [1]. This approach was particularly successful for HBV as it reduced the risk that the formulation could be contaminated with residual viral particles [6,10]. Instead of purifying HBV from urine of infected individuals, the segments of the HBV that encode for hepatitis B surface antigen (HBsAg) were inserted into a yeast plasmid using DNA recombinant technology. At first the administration of HBV vaccine was restricted to individuals at high risk of exposure to HBV, but as the vaccine proved to be efficacious and safe, its use increased to the point that it became common practice in some countries to vaccinate all infants. Although reactions were infrequent in adults and rare in infants and children, some reactions were considered significant. A review of all the safety reports with HBV vaccine found that the vaccine-associated reactions included

immediate (anaphylaxis and urticaria), delayed (e.g., systemic lupus erythematosus, glomerulonephritis), hematologic, ophthalmic, and neurologic toxicities [11]. Given the temporal relationship to vaccination, the similarity between the serious reactions to the vaccine and the extrahepatic effects of infection with HBV and the potential induction of soluble antigen–antibody complexes indicated that although rare, these reactions were related to vaccination. As the benefit of HBV vaccine far outweighs the potential risks, this vaccine continues to be used [11].

While recombinant protein technology eliminated some of the potential risks associated with vaccine therapy (i.e., reversion to virulence), these vaccines still require adjuvants that can produce adverse local reactions and hypersensitivity on occasion. These newer vaccines will share some of the other potential risks of vaccines such as enhancement of disease or infection, antigenic competition, and cross-reactive antibodies [6]. Toxicity due to molecular mimicry in which antibodies to the viral antigen cross-react with other tissues in the body has been a regulatory concern for many years [4,6,11,12]. Cross-reactive antibodies were a particular issue for the development of vaccines against groups A and B Streptococcus (GAS, GBS), Lyme disease and group B meningococcus [4,13–15]. In one example, rheumatic heart disease developed after vaccination to GAS. It was found that the antigen presented by GAS was similar enough to a host's cardiac antigen that the antibodies attacked the cardiac valve [12].

### 3.2 Virus-like Particles

Structural proteins derived from viruses such as HBV have the ability to spontaneously assemble into particles called virus-like particles (VLPs). These VLPs are similar to virions and consist of highly repetitive and ordered structures, but unlike recombinant protein vaccines, VLPs can elicit both humoral and cellular responses [16]. In addition these particles are easily purified and can be produced in large quantities [16]. Several VLPs are in clinical trials and have shown to be effective. This strategy has been particularly successful for HPV, and one product has been recently approved by the FDA for prophylactic use [17].

### 3.3 DNA Vaccines

Interest in the potential for DNA vaccines was increased after a number of publications reported that bacterial plasmids could produce immune responses following administration to vertebrates [18]. DNA vaccines consist of a bacterial DNA plasmid that contains a gene encoding the antigen of interest, a strong viral promoter, and a terminator sequence to ensure that the gene will be expressed in mammalian cells [18–20]. When injected by various routes of administration (e.g., intramuscularly and intradermally), DNA vaccines are transcribed, translated, and the encoded protein is presented by an antigen-

presenting cell (APC) in the context of a self major histocompatibility complex (MHC) [21,22]. DNA vaccines have been found to generate all types of immunity including cytotoxic T cell (CTL) responses, T helper cells, and antibodies [20]. It is thought that DNA vaccines can elicit a CTL response while recombinant proteins cannot because the antigen can be delivered into the cell, allowing the antigen to enter the intracellular processing pathway and resulting in the presentation of its relevant peptides on MHC class I molecules to stimulate a CTL response. Recombinant protein vaccines on the other hand, tend to be taken up by the endolysosomal system where it degrades into peptides and associates with MHC class II molecules that stimulate T helper cells rather then cytolytic T cells [20].

There are potential safety issues with DNA vaccines, some similar to other vaccine strategies and some unique to DNA vaccines. These concerns include induction of autoimmunity, potential integration of plasmid DNA into the host genome, induction of immunological tolerance, immunotoxicity, and altered immune responsiveness to other vaccines and infection [21,23,24].

***Induction of Autoimmunity*** The bacterial elements of a DNA plasmid can contribute to its immunogenic potency. It is well known that oligonucleotides having the sequence purine-purine-CG-pyrimidine, in which the CpG sequence is unmethylated, can activate antigen-presenting cells in vitro and exert immune stimulating effects in vivo [25–27]. It has also been shown that these CpG motifs can stimulate polyclonal T, B, and NK cells that can release immunomodulatory cytokines such as IL-2, TNF$\alpha$, $\gamma$-interferon, and IL-6 [21,26,28,29]. The safety concern is that the promotion of immune activation could lead to anti-DNA antibodies, which may then contribute to and/or accelerate the development of autoimmunity [21, 23]. Additionally bacterial, but not mammalian, DNA can stimulate production of IgG anti-DNA autoantibodies that can lead to glomerulonephritis in mice [21,30,31]. To assess this risk, the effect of vaccination of several DNA vaccines was investigated in normal Balb/c mice and mice prone to lupus (NZB/W). A threefold increase in the number of B cells secreting IgG antibodies against mammalian DNA was induced when Balb/c mice were repeatedly vaccinated whereas the serum anti-DNA antibody titers only rose 35% to 60% [21]. No effect on the onset or severity of autoimmune disease was observed in either the normal or lupus-prone mice, and therefore these findings indicated that this modest level of autoantibody production was not sufficient to induce autoimmunity [21]. These findings are consistent with other reports where no autoimmune-mediated pathology or development of systemic autoimmune response was observed in normal animals after DNA vaccination [22,23,32,33]. Furthermore other studies performed in a number of species, including rodents, rabbits, and nonhuman primates, have not reported any anti-DNA antibodies [22,23,33,34,35,36].

***Integration into Host Genome*** DNA plasmid needs to enter the cell to produce an immune response. Therefore a safety concern is whether

integration of the plasmid DNA into the host genome occurs, and if so, will the risk of genetic instability or cell growth dysregulation, mutagenesis, and carcinogenicity increase [23,24]. Results from biodistribution and persistence studies have typically shown that although plasmid DNA can be detected in many organs shortly after injection, DNA plasmid could only be detected at the injection site several weeks after injection. [23]. Long-term persistence may facilitate the integration, and in one study, when DNA plasmid was injected intramuscularly in mice, it was still present in the muscle 19 months later [23,37]. However, in this and other studies, no integration was observed; the plasmids were predominantly extrachromosomal [23,37,38,39]. Nevertheless, in one study investigators using a newly developed PCR assay to assess integration identified four independent integration events upon plasmid injection followed by electroporation of the injected muscle in vivo. Electroporation of the injection site was considered to have increased plasmid DNA delivery into cells and increased plasmid integration frequency [23,40]. Biodistribution studies are also useful for assessing for distribution to the gonads, but again studies have found that although plasmid can be detected in the gonads shortly after injection, they do not persist [23].

***Immunological Tolerance*** Most preclinical studies (for both efficacy and safety) utilize healthy adult animals. However, there is a concern that the risk for certain clinical populations with reduced immune function, including infants, children, and elderly, is not adequately assessed [21]. Evidence suggests that recognition of foreign determinants is acquired at distinct stages of maturation ranging from early gestation until days or weeks after birth. Since the protein encoded by a DNA vaccine is produced endogenously and expressed in the context of self MHC, the potential exists for the neonatal immune system to recognize the vaccine antigen not as foreign but as "self," resulting in tolerance [21]. Experiments with DNA plasmids in neonatal mice showed long-lasting neonatal tolerance (persisting for more than nine months) when newborn Balb/c mice were injected with a DNA plasmid [21]. However, in other experiments treatment with DNA plasmids did not induce tolerance in neonatal mice [23]. Experiments in aged mice demonstrated a diminished immune-response relative to younger animals but cytokine and antibody production increased following repeated administrations suggesting that age-related changes could be overcome with additional dosing [21,23].

A number of DNA vaccines have progressed into clinical trials for the prevention and/or treatment of HIV, malaria, cytomegalovirus (CMV) and hepatitis B and C [23,24]. So far the DNA vaccines have been well tolerated in clinical trials [23] and have produced both humoral and cellular responses in some trials [41], but overall the potency has been disappointing [20]. Although DNA is typically injected intramuscularly, alternative delivery systems have been evaluated. One such system that has been tested clinically for hepatitis B involves coating plasmid DNA onto gold beads, which are then propelled into the epidermis using a needle-free delivery system [20,42,43].

One such delivery system known as PMED (particle-mediated epidermis delivery) was the first to show that DNA vaccines could produce a humoral response in human trials [44,45]. Also this method was associated with an increased response in patients that had not responded to the HBV protein vaccine [20,45,46].

### 3.4 Virus-Based Vectors

In gene therapy where the goal is to replace a malfunctioning gene with a normal gene viral vectors were initially considered to be an efficient method of ensuring integration into the genome. However, early clinical trials proved to be largely unsuccessful, primarily due to immune reactions to the vector [47,48], and in one case an immune reaction is believed to be the cause of the death of a volunteer in a clinical trial for ornithine transcarbamylase (OTC). The high dose of the vector, which was administered by direct infusion into an hepatic artery, apparently produced a cytokine cascade that led to disseminated intravascular coagulation, acute respiratory distress, and multi-organ failure [49]. Second, it has been shown that the production of neutralizing antibodies, even at moderate doses of vector, decrease the uptake of adenoviral vectors by cells, including antigen-producing cells [47,50]. Therefore formation of neutralizing antibodies can clearly restrict the number of repeat doses given in a clinical trial. However, what is considered to be a negative for gene therapy may be seen as a positive for vaccine therapy, since viral vectors can produce both innate and adaptive immune responses [47,48]. A number of vectors are based on viruses, including variola, vaccinia, avipoxivirus, adenovirus, and alphaviruses [51].

Initially adenoviruses were chosen for several reasons: (1) they cause mild disease in humans, (2) their genome has been well characterized (3) they are relatively easy to manipulate genetically, and (4) they exhibit a broad tropism transducing different cell types [47]. Adenovirus are double-stranded DNA viruses with a genome of approximately 34 to 43 kb; they are species specific, and they include a number of distinct serotypes. There are about 51 known human serotypes and 27 simian serotypes including 7 from chimpanzees [47]. Adenoviruses are made replication defective by removing the E1 gene, but there are vectors with other genes such as E3 and E4 that can also be removed [47].

A number of experiments in mice, rabbits, and nonhuman primates have demonstrated the potential utility of adenoviral vectors in producing immune responses against various viral pathogens, as reviewed by Tastis [47]. While a number of preclinical models have demonstrated the potential utility of these vectors, preexisting immunity against human adenoviruses will still likely curtail the number of clinical administrations of the vector. Moreover neutralizing antibodies against the vector can reduce the potential efficacy upon repeated dose [47]. Investigators have started to evaluate other species, and there are several examples of chimpanzee-derived adenoviruses that have

been evaluated in preclinical models. A vector produced from the chimpanzee AdC68 virus shares almost 90% homology with the human adenovirus 4 [52] and, similar to the HuAd5 virus, enters the cell via the coxsackie adenovirus receptor (CAR), so this vector should exhibit similar tropism to humans [47]. Although greater than 80% of chimpanzee sera will be positive for neutralizing antibodies, humans tested so far do not have neutralizing antibodies for AdC68 [52,53].

As with any vaccine technology the adenoviral vectors share some of the same potential safety issues already discussed. Like DNA vaccines, adenoviral vectors may integrate into the genome and may produce unacceptable immunotoxicity. In addition there is a risk that a viral vector could combine with an endogenous virus to form a chimera. Administration of a viral vector could produce both an innate immune reaction (at high doses) and an antigen-specific response (at lower doses). However, for vaccine treatment, the vector will typically be administered intramuscularly or intradermally so that the systemic exposure is less than for intravenous administration, which is a common route of administration for gene therapy. Figure 1 illustrates these two responses.

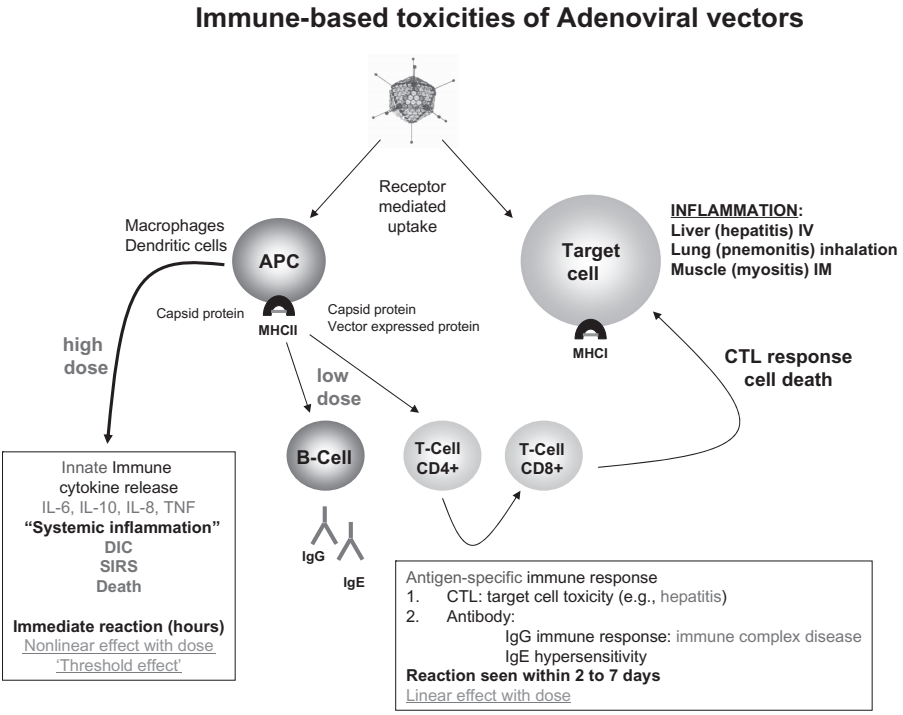


**Figure 1** Potential immune-responses following administration of a viral vector. DIC: disseminated intravascular coagulation; SIRS: systemic inflammatory response syndrome. Image courtesy of Sarah Taplin.

The risk of adenoviral vector-induced oncogenicity is another potential safety concern as human adenoviruses are known to be oncogenic in rats. However, in humans it is considered to be a low risk, for extensive analysis has failed to find human adenovirus DNA associated with human tumors [54]. Therefore the use of in vivo carcinogenicity assays for adenoviruses may be uninformative given the lack of any association of adenoviruses with human malignancy. Also the known oncogenic component (E1 region) of the nonhuman primate (NHP) adenoviral vectors has been deleted from the product. However, other potential oncogenes may be present, and therefore it may be necessary to assess the oncogenicity of the vector.

### 3.5 Mixed Modality Paradigms

Most of the technologies discussed above have been used as single entities. However, because the potency of DNA vaccines in clinical trials has been disappointing and because neutralizing antibodies to adenoviral vectors may restrict the number of administrations in the clinic, investigators have developed a regimen in which a priming dose of a DNA vaccine is followed by a boost using a different modality such as recombinant protein or recombinant viral vector. This prime/boost regimen is currently being evaluated in a number of animal models [18,20]. In one study rhesus macaques were immunized using a heterologous prime/boost regimen of different adenoviral vectors using a combination of human group C Ad6 or Ad5 vectors with a rare human serotype belonging to subgroup D AD24, or two chimpanzee adenoviruses, C32 and C33. These data showed that the prime/boost regimen increased the HCV T cell responses two- to threefold and that boosting could be delayed for over two years after priming, and thus indicating that there is long-term maintenance of resting memory cells [55]. Another study assessed DNA prime followed by a protein boost. In this study mice were immunized with HIV1084 *env*-coding DNA and then boosted with homologous 1084i gp160. Results showed that although immunization with gp160 produced high-titer antibodies, it required two inoculations, whereas in mice primed with DNA, antibodies were produced after a single boost [56].

## 4 ADJUVANTS AND IMMUNOMODULATORS

Advances in immunology have delivered more specific targeted types of vaccine, such as subunit, peptide, and DNA plasmids rather than inactivated whole organisms. These new types of vaccine represent the result of intensive research into epitope mapping, T and B cell receptor binding, and antigen selection. Unfortunately, the refinement has resulted in a reduction in immunogenicity [57]. Nonliving vaccines generally have proved ineffective at inducing potent cell-mediated immune (CMI) responses, particularly of the Th1 type. Although live vaccines can induce an effective CTL response, they also

risk causing disease in immunosuppressed patients who in chronic diseases, for example, HIV and HCV, often form the target patient population. In addition some viruses fail to grow in culture (e.g., HCV) making the development of attenuated viruses difficult. Achieving effective cellular immune-responses while using rationally designed recombinant, peptide, or DNA vaccines is the holy grail of modern vaccine development.

To supplement immunogenicity, adjuvants are required. Adjuvants can be defined as "agents that act nonspecifically to increase the specific immune-response or responses to an antigen." Therefore a particular adjuvant should work with a wide range of antigens. While a great deal of effort is paid to antigen design and mechanism of action, traditionally there has been less effort spent on discovering new adjuvants and understanding their mechanisms of action.

A wide range of unrelated substances can act as adjuvants, including surfactants, oil emulsions, mineral gels, and bacteria-derived substances. Aluminium adjuvants have been in widespread use for over 70 years [58]. Although the exact mechanism of action is still not understood, it is clear that physical association of the antigen with aluminium is required. Usually aluminium hydroxide or aluminium phosphate gels are used. These adjuvants have historically been used as part of bacterial or viral vaccines to help stimulate neutralizing antibodies. However, alum is a weak adjuvant for antibody induction to recombinant protein vaccines, and it induces a Th2 rather than a Th1 response. Alum is also not effective at producing mucosal IgA antibody responses, but it does upregulate costimulatory signals on monocytes and produces a release of IL-4. These adjuvants are not effective at stimulating CMI responses, which are now understood as being important in clearing persistent infections, such as hepatitis C, tuberculosis, and HIV. In these cases a Th1 response consisting of IFN$\gamma$ production and B cell IgG2a secretion is considered more effective than a Th2 response consisting of IL-4 and IL-5 production and B cell production of IgG1 and IgE. In the laboratory this problem can be overcome by using Freunds complete antigen (FCA), which induces Th1 responses. However, this adjuvant produces injection site necrosis and disseminated granulomatous inflammation that precludes its use in the clinic. Inflammation at the site of vaccination results in the release of cytokines (e.g., TNF$\alpha$) that induce dendritic cell activation and stimulate the CMI response. Despite the recognized link between local tissue reaction and adjuvant efficacy, there is a need for novel adjuvants that have the efficacy of FCA but the safety profile of aluminium adjuvants. A list of different types of adjuvant which have been tested for use in infectious disease vaccines is provided in Table 2.

Adjuvants can be classified into two main groups based on their mechanism of action. The first group are particulate materials that act as vaccine delivery systems and target associated antigens into APC. These include emulsions, microparticles, iscoms, and liposomes. The second group are immunostimulatory and are derived mainly from pathogens. These include lipopolysaccharide

**TABLE 1 Selection of regulatory guidelines addressing preclinical safety testing**

| Vaccine Class | Guideline | Date |
|---|---|---|
| All vaccines | European Medicines Evaluation Agency (EMEA) notes for guidance of preclinical pharmacological and toxicological testing of vaccines | 1997 |
| Combination vaccines | EMEA notes for guidance on pharmaceutical and biological aspects of combined vaccines | 1998 |
| Viral vector and DNA vaccines | EMEA note for guidance on the quality, preclinical, and clinical aspects of gene transfer medicinal products | 2001 |
| Cell-based vaccines | EMEA points to consider on the manufacture and quality control of human somatic cell therapy medicinal products | 2001 |
| Smallpox vaccines | EMEA note for guidance on the development of vaccinia virus-based vaccines against smallpox | 2002 |
| Influenza vaccines | EMEA points to consider on the development of live attenuated influenza vaccines | 2003 |
| Adjuvanted vaccines | EMEA guideline on adjuvants in vaccines for human use | 2005 |
| Recombinant protein/ peptide vaccines | US Food and Drug Administration (FDA) points to consider in the production and testing of new drugs and biologics produced by recombinant DNA technology | 1985 |
| Viral vector and cell-based vaccines | FDA guidance for industry: guidance for human somatic cell therapy and gene therapy | 1998 |
| Vaccines for pregnant women and women of child-bearing potential | FDA guidance for industry, considerations for reproductive toxicity studies for preventative vaccines for infectious disease indications | 2000 |
| DNA vaccines | FDA guidance for industry, considerations for plasmid DNA vaccines for infectious disease indications | 2007 |
| All vaccines | World Health Organisation (WHO) guidelines on nonclinical evaluation of vaccines | 2003 |
| Includes vaccines but more relevant to other biologics | International Conference on Harmonization (ICH) S6: preclinical safety evaluation of biotechnology-derived pharmaceuticals | 1997 |
| All vaccines | ICH S7A: Safety pharmacology studies for human pharmaceuticals | 2000 |

*Source*: Adapted from LeBron et al. 2005 [62].

(LPS), monophosphoryl lipid A, and CpG DNA. These substances contain pathogen-associated molecular patterns (PAMPs) that are recognized by and stimulate cells of the immune system. Many of these substances, for example, LPS, have been discovered to act as ligands for Toll-like receptors (TLR). These receptors are expressed on APC, recognize highly conserved motifs on bacteria and viruses, and mediate APC activation. Vaccine developers are now focusing

**TABLE 2 Adjuvants that have been tested for stimulating the immune response to vaccines**

| Class of Adjuvant | Examples |
|---|---|
| Mineral Salts | Aluminium hydroxide,* aluminium phosphate,* calcium phosphate* |
| Immunostimulatory adjuvants | Saponins (e.g., QS21), MDP derivatives, bacterial DNA (CpG oligos), LPS, MPL and synthetic derivatives, lipopeptides, cytokines (e.g., GM-CSF, IL-2, IL-12) |
| Lipid particles | Liposomes, virosomes,* iscoms, cochleates, emulsions (e.g., Freunds adjuvant, SAF, MF59*) |
| Particulate adjuvants | Poloxamer particles, virus-like particles, PLG microparticles |
| Mucosal adjuvants | Cholera toxin (CT), mutant toxin (e.g., LTK63, LTR72), heat labile enterotoxin (LT), microparticles, polymerized liposomes, chitosan |

*Source*: Adapted from Singh and O'Hagan. [57].

*Note*: All of these adjuvants have been evaluated in clinical trials with the exception of cochleates and polymerized liposomes. Only those adjuvants marked * are currently licensed as adjuvants in approved vaccine products.

on the rational design and selection of adjuvants that act on TLR. A useful way to consider the role of adjuvants in recombinant vaccines is that because the final product resembles infection closely enough to initiate an effective immune-response, they ensure the biological consequences of the vaccine.

The acceptable level of toxicity for an adjuvant will depend on its application. For adjuvants intended for use in therapeutic vaccines the level will be higher than for those intended for prophylactic vaccines in healthy individuals. This is particularly true of therapeutic vaccines intended for treatment of cancer patients or life-threatening infectious disease.

Preclinical evaluation of vaccine adjuvants is normally carried out within toxicity studies that also evaluate the whole vaccine. The injection site is evaluated macroscopically and scored for local reaction (e.g., erythema, edema, skin sloughing) at intervals following dosing (e.g., 1, 2, 3 and 14 days). Photographs of the skin sites may be useful for internal use within companies but are not required by regulatory authorities. In addition all of the other traditional endpoints (e.g., haematology, pathology) are assessed. Novel adjuvants need to be evaluated for toxicity in the absence of other vaccine components. In most cases for biotechnology-derived adjuvants, such as DNA plasmid encoding a human cytokine, this can be achieved by adding an additional group of animals to the toxicity studies evaluating the final product. However, a novel chemical-based adjuvant should be treated as a new chemical entity and therefore requires a classical toxicological programme as outlined in the EMEA guideline on adjuvants (Table 1). This will consist of acute and repeat-dose toxicity studies in two species, pharmacokinetic (PK) and tissue distribution studies, genotoxicity testing and testing for hypersensitivity, anaphylaxis and pyrogenicity. Immunological assays are often included as part of the toxicity

assessment and are helpful to demonstrate the benefit of the adjuvant. In some studies different potential adjuvant–antigen combinations can be compared for both immunogenicity and toxicity [59].

No adjuvant is licensed as a medicinal product in its own right, but only as a component of a particular vaccine. Therefore preclinical and toxicology studies need to be designed on a case-by-case basis to evaluate the safety profile of the adjuvant and adjuvant/ vaccine combination [60]. Evaluation in preclinical studies is important for identifying the optimum composition and formulation process and also for allowing development of tests for quality control [61]. Data from these studies also helps plan protocols for subsequent clinical trials from which safety and efficacy in humans can be evaluated.

## 5 CONSIDERATIONS FOR PRECLINICAL SAFETY EVALUATION

The wide range of different types of biotechnology-derived vaccines has resulted in the publication of a diverse range of regulatory guidelines. These take the form of a guidance note or points to consider. Rapid changes in technology have required a similarly rapid response from regulators on new requirements for assessing safety (Table 1). The relationship between vaccine companies and regulators has been constructive, and input into guidance from all sources has been solicited. The expectation is that toxicity studies will be conducted in animal species on all new vaccines prior to first administration to humans. Exceptions to this rule include combined vaccines containing known antigens. A conservative approach is merited because most vaccines are intended for prophylactic use in healthy humans. The risk–benefit evaluation requires that new vaccines have minimal risk of toxicity. Therapeutic vaccines that are intended for use in patients, frequently with life-threatening diseases such as HIV or hepatitis C infection, have a different risk–benefit evaluation. Thus an effective new vaccine for these diseases might be approved even with a low level of acceptable adverse effects.

Reading the published guidelines represents an important first step in designing a suitable package of toxicity studies. However, early communication with the regulatory authorities is encouraged, and this represents a valuable opportunity for companies to receive feedback on preclinical and clinical plans. The benefit of this dynamic regulatory environment is that it permits companies to design a preclinical development program that does not rely wholly on the published guidelines but rather on scientific justification for each study and endpoint measured. Preclinical safety studies should be performed in compliance with good laboratory practice (GLP) regulations.

### 5.1 Quality Control Testing of Material Intended for GLP Studies

There is a requirement that the test material used in GLP preclinical studies should resemble the same quality of product that will be tested in the clinic.

Therefore quality control (QC) testing of vaccines normally includes the following assays, which must be passed prior to material being released for use in preclinical toxicology studies: sterility, endotoxin, general safety, identity, mass, potency, purity, and stability [62]. These assays should be performed on the final product using the clinical formulation.

Ideally the test material used in the toxicology studies should come from the vaccine lot, produced according to cGMP, that is intended for the FIH clinical trial. However, frequently this is not possible, and non-GMP material may be used subject to demonstrating the comparability of this vaccine lot to the intended clinical lot. For those vaccines that incorporate antigens or adjuvants that are biologically active in limited species, it may be necessary to perform some or all of the testing using homologues. For example, a plasmid expressed cytokine may form part of the adjuvant for a novel vaccine. This expressed cytokine may only be biologically active in humans and chimpanzees. In this situation there will be an ethical case or biological case (e.g., similarity to human skin for dermally applied vaccine) for selecting a lower species and using a homologue form of the cytokine. So suitable QA testing must also be applied to the plasmid encoding the homologue protein prior to its use in preclinical studies. However, there is a regulatory expectation that the final clinical product will also be tested. The formulation should be identical to the intended clinical formulation. Stability of the vaccine is determined before and after the GLP study to ensure adequate stability during storage prior to dosing.

### 5.2 Testing for Viral Contamination

Vaccines produced using eukaryotic cells have the potential to be contaminated with either endogenous or exogenous viruses resulting from accidental introduction at some stage of the process. A battery of analytical and safety tests is applied to the cell banks (master and working), virus seeds (master and working), and vaccine batches to address this concern. These tests are performed in vitro, in vivo, and by laboratory analysis. The range of tests required will depend on species and tissue source of the cells, history of the cells and original virus isolate, and the culture media ingredients.

### 5.3 Immunogenicity Evaluation within Toxicity Studies

Evaluating the immune response to the vaccine at the same time as evaluating toxicity is recommended in a 2003 WHO guideline (Table 1). This makes good sense because any toxicity observed with vaccines is frequently related to the immune response, and correlation of these two endpoints provides a fuller explanation of the test results. Assays to measure the immune response include those which measure antigen-specific antibody responses (e.g., ELISA) and cell-mediated responses (e.g., γ-interferon—ELISPOT). Developing these

assays can be a challenge for certain species. Therefore the choice of relevant animal model for toxicity testing ideally should be the same as the animal model selected for immunogenicity or efficacy testing. Recent FDA guidance state that preclinical studies to specifically assess if DNA vaccination causes autoimmune disease are no longer required (Table 1). This follows experience that any anti-DNA antibodies produced following vaccination are usually at such a low level as to be insufficient to cause disease in normal animals or accelerate disease in autoimmune-prone mice. Similarly the absence of an immune-response against cells expressing the vaccine-encoded antigen (including muscle cells and dendritic cells) suggests that an autoimmune response directed against tissues in which such cells reside is unlikely. It is also recommended that prior to use of a DNA vaccine in children or newborns, the vaccine be first tested for safety and immunogenicity in adults and appropriate animal models be used to evaluate the potential of the vaccine to induce neonatal tolerance. Preclinical testing to evaluate the risk of hypersensitivity is problematic because this type of response is rare and idiosyncratic in the clinic. Conventional endpoints evaluated in the repeat-dose toxicity study provide a screen for the development of hypersensitivity in the animal studies. However, the low number of animals used and the important genetic variability inherent in the human population are important factors in weakening the predictive power of current preclinical studies.

### 5.4 Species Selection

Immunologists will have selected the most appropriate species for establishing immunogenicity of the vaccine. Frequently they will have used animal models of, for example, tolerance or in the case of cancer targets, xenograft models. Usually naive animals that respond to all constituents of the vaccine will also have been used and assays developed for use in this species. For a live viral vaccine, infection and protection studies might have been performed in a species susceptible to infection with the wild-type virus. Regulatory guidelines allow for a single species to be used for the toxicology assessment. The key requirement is that the species chosen should develop a similar immune response to that expected for humans postvaccination [63]. In practice, either rodents or rabbits are often used. However, if elements of the vaccine are inactive in these species, such as plasmid encoding human cytokines, then either an alternative sensitive species, such as a nonhuman primate or a nonhuman homologue of the cytokine, should be used. Which is the correct approach? This is an ambiguous regulatory area for those developing vaccines. On the one hand, there are sound ethical reasons for avoiding the use of nonhuman primates in safety testing, and their use must be justified in each case. On the other hand, there is an expectation by the regulatory authorities that the clinical product should be evaluated in a sensitive species. In the case of a plasmid encoded human cytokine that is only active in nonhuman primates this presents a dilemma. Animal use ethical committees will scrutinize a

preclinical study and ask why a homologue of the cytokine cannot be used enabling the use of, for example, rodents instead. However, there remains uncertainty as to whether a safety package developed using a nonhuman homologue would be currently acceptable to regulators for use in risk assessment prior to human studies. This is an area where improved regulatory guidance is clearly required. An example where a murine homologue was used in preclinical toxicity studies is provided by Vical Inc., which tested murine GM-CSF in a repeat-dose toxicity study in mice [32]. However, it is not known if this safety package was acceptable in the absence of testing the clinical product, plasmid DNA encoding multivalent malaria antigens and human GM-CSF, in a nonhuman primate. With increasing species specificity of vaccines incorporating human cytokines, it seems likely that under the current regulatory environment there will be increasing use of nonhuman primates in safety assessment. In any case, studies should be carefully designed to optimize the information obtained with as few animals as possible, especially if nonhuman primates are used.

Phylogenetically, macaques are closely related to humans and, as such, are often used to evaluate vaccine safety and immunogenicity [64]. Vaccine responses in rhesus macaques, as opposed to rodents, have been found to be better predictors of human responses to several malaria vaccine candidates [59]. The rhesus macaque has substantial class II and to a lesser extent class I MHC homology with humans [65], and it has proved useful in selecting malaria vaccine formulations with improved clinical immunogenicity and efficacy [66]. In addition the nonhuman primate allows cutaneous immunologic endpoints, such as delayed-type hypersensitivity (DTH) responsiveness to vaccine-related antigens to be measured. This is an important consideration because for vaccines the possible adverse effects "immunotoxicity" are generally related to the immunological response rather than to any inherent toxicity of the components within the final product.

## 5.5 Design of Safety Assessment Studies

For each vaccine candidate a tailored package of safety assessment studies should be performed. The package will reflect the type of vaccine. As a minimum these should include repeat-dose toxicity, local tolerance, and reproductive and development toxicity. Local tolerance is evaluated by visual assessment of the dosing sites, and the skin reaction is scored according to the Draize method for erythema and oedema formation [67]. Epithelial sloughing can also be scored. Plasmid DNA or adenoviral vectors require studies of biodistribution and integration (Table 1), whereas live viral vaccines require virulence and potentially neurovirulence testing. Additional separate studies might include single-dose toxicity and safety pharmacology. In many cases these endpoints can be obtained from the repeat-dose study. This represents an efficient use of animals and enables comparisons among different endpoints for the same animal.

***Carcinogenicity and Genotoxicity Tests*** Carcinogenicity studies and genotoxicity assays are generally not considered appropriate for testing vaccines and are therefore not required. Exceptions to this arise if a new chemical-type adjuvant is included in the formulation, if the vaccine includes components that may be considered a risk for mutagenicity and/or tumorigenicity by virtue of their novel nature (e.g., new adenovirus-derived vectors) or if the production process for the vaccine has used a cell line with a potential risk. Since adjuvants are intended for infrequent use at low dose, the EMEA considers the direct risk of tumor induction by these as negligible and does not require long-term carcinogenicity tests (Table 1). Viral vectors may also have the potential to recombine with wild-type viruses, revert to virulence, and cause oncogenic mutations or suppress immune function [68]. Biodistribution studies should be conducted to establish disposition of viral DNA, and separate assays should evaluate if the viral vector is itself pathogenic. If a DNA vaccine contains a vector with extensive homology to the human genome or known oncogenic potential, then close attention must be paid to the results of the biodistribution study, and if positive results are obtained, an integration assay may be needed.

***Single-Dose Study*** Although single-dose studies are advised in the EMEA guideline, they are generally not particularly useful for products intended for repeat dosing in the clinic. Many of the endpoints, such as clinical signs, body weight, and macroscopic observations of the dosing site can be obtained following the first dose during a repeat-dose study. However, if they are needed, for example, for a vaccine designed for single dosing in the clinic, a 14-day period following dosing to study termination is a normal period that enables development of the primary immune response.

***Repeat-Dose Study*** The route of administration, dose, and dosing intervals used in the preclinical safety assessment studies should reflect the intended clinical study as closely as possible. If a particular dosing device is to be used in the clinic, then this device should also be used in preclinical studies. In contrast to the development of small molecules, administering the full clinical dose is acceptable for vaccine candidates, generally in a nonrodent species, rather than large multiples of the clinical dose. In some cases, for example, in small species, it may not be possible to administer the entire clinical dose or use the clinical dosing device. In these situations agreement prior to study start should be reached with regulators.

If immunogenicity studies demonstrate that a similar immune response is obtained at intervals shorter than the proposed clinical dosing interval (e.g., 4–6 weeks), then a truncated dosing interval (2–3 weeks) can be used in the repeat dose safety studies. The FDA recommends that one more dose than the clinical dosing regimen ($N$) be used in animal studies. This is termed the $N + 1$ rule [61]. The number of animals used per group should be sufficient to detect

a treatment-related change. In practice, the number of animals used depends on the species selected for practical and ethical reasons. For example, a rat study might have 10 animals per sex per group. In contrast, a nonhuman primate study might have 3 per sex per group. There are no fixed guidelines for this and group sizes should be justified.

Repeat-dose toxicity studies include a wide range of antemortem and postmortem evaluations. Antemortem evaluations usually include mortality, clinical signs, body weights, food consumption, ophthalmology, urinalysis, serum biochemistry, haematology, coagulation, and blood for immunogenicity assays. Necropsies typically take place at two time points following the final vaccine dose. The first time point at 2 to 7 days evaluates acute changes, both at the final dosing site and systemically. An advantage of waiting until 7 days is that it enables development of an immune response to the final dose which can be assayed. In addition any pathology related to the peak immune response is more likely to be detected at this time point. In contrast, a necropsy at 2 days enables histological evaluation of the acute response at the injection site and systemically. In most cases a judgment can be made on the relative importance of these different aspects to the safety assessment. The second time point at 2 to 4 weeks evaluates delayed toxicity and recovery of any treatment-related effects observed at the earlier time point. At necropsy, gross examination of all major tissues, organ weights, and a complete list of tissues is collected. Selection of tissues for microscopic examination depends to some extent on the type and route of vaccination. However, in all cases major organs and the site of vaccination should be evaluated.

Treatment-related effects are generally related to the development of an immune response and consist of local skin or muscle inflammation, enlarged draining lymph nodes, and potentially increased spleen weight. Effects are also frequently seen in haematology parameters. Since they relate to the expected immune response, these changes are generally not considered adverse. In certain cases additional endpoints to these studies should be added. For example, when testing certain vectors that target the liver such as recombinant adenovirus serotype 5, it may be of use to evaluate effects on hepatic enzymes, because P450 enzymes have been demonstrated to be reduced following dosing and this could affect drug interactions [69].

***Safety Pharmacology Studies*** Safety pharmacology studies test for treatment-related changes in key physiological functions, such as respiratory, central nervous, and cardiovascular systems. WHO guidance suggests that if there are components of the vaccine (e.g., toxoids) that may influence these body systems, then safety pharmacology studies should be performed. However, in those cases where the pharmacology is well characterized and the systemic exposure is low, there may not be a requirement to perform these tests (Table 1; ICH S7A). A case-by-case approach is usually followed, and if it is considered that these evaluations are required, then frequently the endpoints can be incorporated into the repeat-dose study. This approach is particularly valuable when a large animal model is being used.

***Reproductive Toxicology Studies*** Most vaccines are destined to be used in a population that includes women of child-bearing potential. Therefore most vaccines need to be tested for adverse effects on reproduction and development [70]. The general approach taken is that the benefit of vaccination among pregnant women usually outweighs the risk for potential adverse effects on the mother or developing offspring when the risk for disease exposure is high, infection poses a special risk to mother or fetus, and the vaccine is unlikely to cause harm [71]. For example, live virus vaccines are not recommended for use in pregnant women because of the risk of viral transmission to the fetus. There is a continuing medical need for maternal immunization with vaccines that could prevent serious disease in the newborn, for example, group B streptococcal disease.

Currently, unless the vaccine is specifically indicated for maternal immunization, no data are collected regarding the vaccine's safety in pregnant women prior to licensing. Pregnant women are normally excluded from clinical trials, and therefore risk assessment for adverse effects on development is based on developmental toxicity studies in animals. The potential for adverse effects on fertility can also be assessed by careful examination of the reproductive tract tissues in the repeat-dose toxicity studies. Such preclinical data are used to inform the vaccine labeling regarding its use during pregnancy. FDA draft guidance on reproductive toxicity studies for prophylactic vaccines requires that the following be assessed: antibody production in the pregnant animal, antibody transfer from the pregnant female to the fetus, and the persistence and effects of the antibody response in the newborn. The FDA has issued guidance on study design and recommends an embryofetal development protocol that tests effects on the pregnant/lactating female, embryonic and fetal development, and pre- and postnatal development up to weaning (Table 1). For embryofetal toxicity studies it is recommended to include a group with immunizations prior to mating in order to achieve antibody exposure during the entire embryofetal period [63]. Vaccination should also take place during pregnancy. Postmating immunization is recommended for females that are not submitted to caesarean for fetal examination. Pups from these immunized dams are necropsied at the end of the lactation period.

The FDA is proposing to amend its labeling regulations and require that a summary assessment of the risks of using a product during pregnancy and lactation be included. This summary would enable a wider discussion of the animal and any human data available.

***Biodistribution and Integration Assays*** An additional specialized GLP study is normally required to evaluate the dissemination and persistence of DNA plasmid and viral vector-based vaccines. These are termed biodistribution studies, and they track the levels of vector in a panel of tissues at varying times following dosing of vaccine on a single occasion by the clinical route. Mouse, rabbit, and minipig have been used for this purpose [72]. Necropsies are performed at staggered timepoints (e.g., 1 day, 14 day, 28 day, 56 day, and 140 day) and need to be carefully carried out following special precautions to

avoid cross-contamination of residual plasmid between tissues [73]. A recommended panel of tissues to be collected as frozen specimens is now provided in the recent FDA guidelines for DNA vaccines and includes blood, heart, brain, liver, kidney, bone marrow, ovaries, testes, lung, draining lymph nodes, spleen, and muscle or skin at the site of administration and subcutis at the injections site (Table 1).

Briefly, DNA is extracted from these tissues, and the copy number of vector per microgram of genomic DNA is determined by a sensitive assay, for example, quantitive PCR (Q-PCR). For many of these types of products, wide dissemination to many organs is expected via the blood. This results in low levels of vector being detected at early time points (e.g., days 1–2) in many tissues, including blood, heart, liver, and reproductive organs [74]. However, at later time points (e.g., day 30) these tissues are generally negative as judged by a particular threshold of vector concentration (e.g., 100 copies/μg genomic DNA). In this case there is not considered to be any safety concern. However, persistence above this threshold level at the last time point of the study (e.g., day 140) in any tissue including administration site raises the possibility that the vector will have integrated into genomic DNA at significant levels and thus caused insertional mutagenesis. An alternative explanation is that the vector is persisting episomally and does not represent a risk of mutagenesis. To distinguish between these two explanations, companies are requested to test for integration if plasmid persists in any tissue at levels exceeding 30,000 copies/μg genomic DNA by study termination. Developing a robust and sensitive assay has proved difficult, and there are a number of imperfect options from which to select. In principle, the simplest assay consists of gel purification of genomic DNA using size exclusion in an effort to remove low molecular weight DNA, which contains nonintegrated vector [39]. To confirm a positive result in this assay that is due to integrated vector and not contamination with free vector, an additional assay, the repeat-anchored integration-capture (RAIC) PCR, can be used to confirm genomic insertion sites [40]. To enable risk assessment, the copy number of plasmid persistence should be compared to the spontaneous mutation rate. Literature and regulatory agency experience is developing in this area, and there is now an improved understanding of the biological significance of persistent plasmid DNA.

***Virulence and Potency Testing*** Live attenuated viral vaccines are considered to have a theoretical risk of reversion to virulence that needs to be tested. This is done by monitoring genetic or phenotypic markers of attenuation or virulence and can include in vitro and in vivo potency tests [62].

## 6 CASE STUDIES FOR POTENTIAL HIV THERAPIES

While many infectious diseases are controlled by vaccination, there are other diseases where novel and safe vaccines are urgently required. The obvious

example is HIV where no proven vaccination exists. Vaccination for HIV is a difficult prospect for several reasons: (1) it was never considered to be safe enough to trial a live-attenuated virus, (2) effective vaccination requires other approaches than envelope-only constructs, (3) because of the genetic drift of the virus it is unlikely that single variants will be effective, (4) transmission of infection tends to occur from cell to cell contact rather than by free virus particles, and finally (5) there are no relevant animal models [75]. A number of potential vaccine strategies have been evaluated both preclinically and clinically and are discussed below.

### 6.1 Synthetic Peptide

An HIV-1 p17-based synthetic peptide (HGB-30) coupled to a carrier protein (KLH, keyhole limpet hemocyanin) has been administered with alum as an adjuvant [75]. The potential toxicity of HGB30-KLH/alum (HGB-30) was assessed in a number of species. In mice, guinea pigs, and goats the safety evaluation was limited to clinical observations. In chimpanzees, in addition to clinical observations, clinical pathology was also assessed, and in rabbits, dogs and cynomolgus monkeys, autopsies were conducted in addition to the other parameters mentioned above. In general, animals were immunized subcutaneously at doses ranging from 20 to 1000 μg/kg HGB-30 at 2- to 8-week intervals for a total of 2 immunizations. No toxicity or mortality was observed in any of the species. In some cases when the vaccine was administered in Freund's adjuvant, abscesses and/or hypersensitivity reactions occurred and were considered to be related to this adjuvant, particularly as no skin reactions were observed when alum was substituted for Freund's adjuvant. A longer term study (up to 1 year) was carried out in cynomolgus monkeys and chimpanzees. No toxicity was seen following a total of 6 to 8 immunizations. The immune response to the vaccine was also evaluated, and antibodies to HGB-30 were observed in all species by 4 weeks after a booster immunization. T cell proliferative responses to HGB-30 in mice were variable but did reach statistical significance in at least one experiment and CTL responses were inconsistent.

The phase 1 clinical trial for HGB-30 consisted of 18 seronegative healthy volunteers (6/group) who were administered 10, 20, or 50 μg of the vaccine at 0, 6, and 14 weeks. Physical examinations and blood analyses were performed, and all subjects were followed for 1 year. Initial results indicated that the vaccine was not associated with any adverse effects. Similarly to that observed in mice and chimpanzees, antibodies to HGP-30 were observed in volunteers. Since T cell proliferative responses are considered to be optimal with fresh cells, extensive studies were carried out on one volunteer's peripheral blood lymphocytes. Increased T cell proliferation was observed. The T cell proliferation was maximal 6 weeks after the third immunization but then declined at 6 months after the third immunization.

### 6.2 DNA Plasmid

Another strategy for HIV vaccination is the use of DNA plasmid encoding various HIV-1 genes. Different plasmids ranging from the simplest encoding for the Clade B gag gene only, to a combination of six plasmids including Clade B env gene, Clade B gag-pol-nef genes, Clade C gag-pol-nef fusion protein, env protein for Clade C, gag-pol-nef fusion protein for Clade A, and env protein for Clade A were evaluated in repeat-dose toxicity and biodistribution studies [12,76]. These clades represent different HIV subtypes and are prevalent in different areas. For example, Clade C is the most prevalent subtype and predominates in sub-Saharan Africa and Asia [56]. One of the HIV DNA plasmids also contained a cytokine adjuvant (IL-2). In addition plasmids for Ebola, the severe acute respiratory syndrome (SARS) virus, and the West Nile virus were also evaluated in this report.

The repeat-dose toxicity and biodistribution studies were designed to support a phase 1 clinical trial in human subjects. The DNA plasmid was intramuscularly injected using a device called the Biojector 2000 at doses of up to 8 mg/dose DNA plasmid at 3 intervals (0, 1, and 2 months). The Biojector is a needle-free injection system for intramuscular (IM) and subcutaneous (SC) injections of liquid medications including vaccines. In general, rabbits were injected (using the Biojector) with either 3 (to mimic the clinical plan) or 4 immunizations ($N + 1$ in line with current FDA regulatory standards). The dose levels were either 4 or 8 mg, and the rabbits were dosed with one of the different HIV DNA plasmids every 4 weeks. The standards parameters (clinical observations, body weight and food consumption determination, ophthalmoscopy, clinical pathology, organ weight measurements, and gross and microscopic pathologic evaluations) were included in the repeat-dose toxicity study. Animals were necropsied approximately 2 weeks after the final immunization. Treatment-related findings were limited to the injection site and consisted of mild to moderate irritation or inflammation and were recoverable. All other changes were considered to be incidental and not related to treatment.

The biodistribution study evaluated the same DNA plasmids. In one series of studies, 6- to 7-week-old Hsd:ICR (CD-1) mice were administered 100 μg DNA plasmid vaccine by either the IM or intravenous (IV) route. Animals were dosed once and then necropsied (5/sex/group for treated and 1/sex/group for control) on study days 8 and 50. A number of organs including gonads, liver, heart, brain, kidney, lymph nodes, injection site (muscle), lung, bone marrow, and heart were collected for Q-PCR analysis. In a second series of studies, 12- to 13-week old New Zealand white rabbits (same number/group as with the mouse study) were dosed with 2 mg of DNA plasmid delivered IM by the Biojector device. Animals were terminated approximately 1 week and 1 and 2 months after a single injection. A number of organs (similar to the mouse study but including adrenal glands and thymus) were collected for Q-PCR analysis. In a third study, 6- to 7-week-old CD-1 mice were inoculated

with 100 μg DNA plasmid intramuscularly, and 28 days later injected muscles were harvested and assessed for possible integration of the DNA plasmid. Results showed that for the most part, plasmids remained at or near the injection site and did not distribute widely throughout the body. Sporadic low-level signals (hundreds to at most thousands of copies) were detected in other tissues in some rabbits, whereas tens of thousands to millions of copies/μg DNA were seen at the injection site. The signal at the injection site steadily decreased 1 and 2 months postdose to a few hundred copies/μg DNA at the 2-month time point. This distribution pattern was observed for all of the different DNA plasmids (including Ebola, SARs, and West Nile virus) tested regardless of backbone, promoters, or gene inserts. Finally no integration was noted in the integration study. These findings show that DNA plasmids are well tolerated, do not distribute widely, and do not appear to integrate into the genome.

### 6.3 Mixed Modality

Investigators are currently assessing the potential safety and efficacy of a mixed modality treatment regimen (i.e., combining a DNA vaccine with a viral vector). In this example preclinical toxicity and biodistribution studies were conducted in mice using a DNA vaccine in combination with a modified vaccinia virus Ankara based HIV vaccine to support a phase 1 trial in healthy HIV-1-uninfected volunteers [77,78].

In this example the DNA vaccine (pTHr·HIVA) is based on a novel direct gene transfer vector pTH [77], and the second component, MVA-HIVA, is based on a modified vaccinia virus Ankara. Both of these components contain most of the HIV-1 clade A *gag* protein coupled to conserved HIV-1 clade A CTL epitopes arranged in a polypeptide string that served as the immunogen [77].

Balb/c mice about 6 to 7 weeks old were selected. Because the CTL epitope is recognized by this strain, Balb/c mice are considered a biologically relevant strain and species. A CTL response was confirmed in all mice following administration. Four groups of mice were assessed for toxicity (10/sex/group) and persistence (10/sex/group for treated mice and 3/sex/group for vehicle group). Mice were administered the vehicle, 2 doses (50 μg/dose) of DNA plasmid intramuscularly on days 29 and 43, 2 doses ($10^6$ pfu/dose) MVA intradermally on days 29 and 43, or a combination of DNA (on days 1 and 14, same dose) followed by 2 doses of MVA on days 29 and 43 at the same dose. The dose levels selected represented a safety margin of 1750-fold and 70-fold over the proposed clinical dose for DNA and MVA, respectively, based on body mass. All mice for the toxicity evaluation were necropsied on day 49/50. Parameters tested included mortality, clinical observations, local tolerance, ophthalmoscopy, body weight and food consumption, clinical pathology, organ weights, and gross and microscopic evaluation. In the persistence study, one vehicle-treated/sex and three vaccine-treated/sex were evaluated on days 46 and 78.

Remaining animals in the persistence study were retained for possible further assessment.

Treatment with DNA alone, MVA alone, and in combination was well tolerated in mice. Some of the local reactions observed were considered to be related to the injection procedure rather than to the vaccine components. Other findings were considered to be incidental in nature and not related to the vaccine treatment. Persistence was assessed by PCR, and results showed that positive signals were observed at the injection sites on both days 46 and 78. Equivocal positive results were observed in some mice (including vehicle controls) in a few other tissues, and these were most likely due to contamination and the high sensitivity of the PCR method. Since there was no persistence in any other tissue except for the injection site, no integration assays were carried out.

In summary, administration of DNA plasmid alone, MVA alone, and in combination revealed no adverse findings, and furthermore, with the exception of the injection sites, no persistence was noted in any organ including the gonads. For safety testing in the clinic, healthy HIV-1-uninfected volunteers (male and female) were recruited into one of three trials [78]. Women of childbearing potential were included in the trial if they were not pregnant (confirmed by urine test), were not lactating, and were willing to use an approved method of contraception from screening until four months after final vaccination. The contraceptive regimen was required for male volunteers as well.

Twenty-six volunteers were recruited. In the first trial, 18 volunteers received either a 100 μg or 500 μg dose of plasmid THr.HIVA DNA intramuscularly on two occasions (days 0 and 21). In the second trial, 8 volunteers were administered two intradermal injections on days 0 and 21 with $5 \times 10^7$ pfu MVA-HIVA, and then in the third trial, 9 volunteers from the first trial were boosted with two doses of MVA-HIVA (21 days apart) 9 to 14 months after their last vaccination with DNA. All volunteers were followed for up to six months to two years after the last vaccination.

In all three trials, the candidate vaccines were well tolerated alone and in combination. Clinical signs were limited to reactions (soreness, redness, etc.) at the intradermal injection site, and no adverse events were considered to be related to vaccine treatment.

## REFERENCES

1. O'Hagan DT, Rappuoli R. The safety of vaccines. *Drug Discov Today* 2004;9: 846–54.
2. Anonymous. United Nations Programme on HIV/AIDS. 2004 Report on the Global AIDS Epidemic, 2006. http://www.unaids.org
3. Gomez-Roman VR, Florese RH, Peng B, Montefiori DC, Kalyanaraman VS, Venzon D, Srivastava I, Barnett SW, Robert-Guroff M. An adenovirus-based HIV subtype B prime/boost vaccine regimen elicits antibodies mediating broad antibody-

dependent cellular cytotoxicity against non-subtype B HIV strains. *J Acquir Immune Defic Syndr* 2006;43:270–7.

4. Brennan FR, Dougan G. Non-clinical safety evaluation of novel vaccines and adjuvants: new products, new strategies. *Vaccine* 2005;23:3210–22.
5. Piedra PA. Clinical experience with respiratory syncytial virus vaccines. *Pediatr Infect Dis J* 2003;22:S94–9.
6. Bussiere JL, McCormick GC, Green JD. Preclinical safety assessment considerations in vaccine development. *Pharm Biotechnol* 1995;6: 61–79.
7. Perrie Y. Vaccines: an overview and update. *Pharmaceut* 2006;276:209–12.
8. Strom J. Further experience of reactions, especially of a cerebral nature, in conjunction with triple vaccination: a study based on vaccinations in Sweden 1959–65. *Br Med J* 1967;4:320–3.
9. Stewart GT. Toxicity of pertussis vaccine. *Lancet* 1977;310:1130.
10. Dreesman GR, Hollinger FB, Sanchez Y, Oefinger P, Melnick JL. Immunization of chimpanzees with hepatitis B virus-derived polypeptides, *Infect Immun* 1981;32: 62–7.
11. Grotto I, Mandel Y, Ephros M, Ashkenazi I, Shemer J. Major adverse reactions to yeast-derived hepatitis B vaccines—a review. *Vaccine* 1998;16:329–34.
12. Sheets RL, Stein J, Manetz TS, Andrews C, Bailer R, Rathmann J, Gomez PL. Toxicological safety evaluation of DNA plasmid vaccines against HIV-1, ebola, severe acute respiratory syndrome, or West Nile virus is similar despite differing plasmid backbones or gene-inserts. *Toxicol Scis* 2006;91:620–30.
13. Hayrinen J, Pelkonen S, Finne J. Structural similarity of the type-specific group B streptococcal polysaccharides and the carbohydrate units of tissue glycoproteins: evaluation of possible cross-reactivity. *Vaccine* 1989;7:217–24.
14. Hemmer B, Gran B, Zhao Y, Marques A, Pascal J, Tzou A, Kondo T, Cortese I, Bielekova B, Straus SE, McFarland HF, Houghten R, Simon R, Pinilla C, Martin R. Identification of candidate T-cell epitopes and molecular mimics in chronic Lyme disease. *Nat Med* 1999;5:1375–82.
15. Coquillat D, Bruge J, Danve B, Latour M, Hurpin C, Schulz D, Durbec P, Rougon G. Activity and cross-reactivity of antibodies induced in mice by immunization with a group B meningococcal conjugate. *Infect Immun* 2001;69:7130–9.
16. Boisgerault F, Moron G, Leclerc C. Virus-like particles: a new family of delivery systems. *Expert Rev Vaccines* 2002;1:101–9.
17. Anonymous. Gardasil product description, 2006. http://www.fda.gov/cber/products/hpvmer060806.htm
18. Donnelly JJ, Ulmer JB, Shiver JW, Liu MA. DNA vaccines. *Annu Rev Immunol* 1997;15:617–48.
19. Snodin DJ, Ryle PR. Understanding and applying regulatory guidance on the nonclinical development of biotechnology-derived pharmaceuticals. *Bio Drugs* 2006;20:25–52.
20. Liu MA. DNA vaccines: a review. *J Intern Med* 2003;253:402–10.
21. Klinman DM, Takeno M, Ichino M, Gu M, Yamshchikov G, Mor G, Conover J. DNA vaccines: safety and efficacy issues. *Springer Semin Immunopathol* 1997;19: 245–56.

22. Parker SE, Borellini F, Wenk ML, Hobart P, Hoffman SL, Hedstrom R, Le T, Norman JA. Plasmid DNA malaria vaccine: tissue distribution and safety studies in mice and rabbits, *Hum Gene Ther* 1999;10:741–58.
23. Schalk JA, Mooi FR, Berbers GA, van Aerts LA, Ovelgonne H, Kimman TG. Preclinical and clinical safety studies on DNA vaccines. *Hum Vaccines* 2006;2:45–53.
24. Smith HA, Klinman DM. The regulation of DNA vaccines. *Curr Opin Biotechnol* 2001;12:299–303.
25. Donnelly J, Berry K, Ulmer JB. Technical and regulatory hurdles for DNA vaccines. *Int J Parasitol* 2003;33:457–67.
26. Krieg AM, Yi AK, Matson S, Waldschmidt TJ, Bishop GA, Teasdale R, Koretzky GA, Klinman DM. CpG motifs in bacterial DNA trigger direct B-cell activation. *Nature* 1995;374:546–9.
27. Davis HL, Weeratna R, Waldschmidt TJ, Tygrett L, Schorr J, Krieg AM. CpG DNA is a potent enhancer of specific immunity in mice immunized with recombinant hepatitis B surface antigen. *J Immunol* 1998;160:870–6.
28. Halpern MD, Kurlander RJ, Pisetsky DS. Bacterial DNA induces murine interferon-gamma production by stimulation of interleukin-12 and tumor necrosis factor-alpha. *Cell Immunol* 1996;167:72–8.
29. Klinman DM, Yi AK, Beaucage SL, Conover J, Krieg AM. CpG motifs present in bacteria DNA rapidly induce lymphocytes to secrete interleukin 6, interleukin 12, and interferon gamma, *Proc Natl Acad Sci USA* 1996;93:2879–83.
30. Gilkeson GS, Ruiz P, Howell D, Lefkowith JB, Pisetsky DS. Induction of immune-mediated glomerulonephritis in normal mice immunized with bacterial DNA. *Clin Immunol Immunopathol* 1993;68: 283–92.
31. Gilkeson GS, Pippen AM, Pisetsky DS. Induction of cross-reactive anti-dsDNA antibodies in preautoimmune NZB/NZW mice by immunization with bacterial DNA. *J Clin Invest* 1995;95:1398–1402.
32. Parker SE, Monteith D, Horton H, Hof R, Hernandez P, Vilalta A, Hartikka J, Hobart P, Bentley CE, Chang A, Hedstrom R, Rogers WO, Kumar S, Hoffman SL, Norman JA. Safety of a GM-CSF adjuvant-plasmid DNA malaria vaccine. *Gene Ther* 2001;8:1011–23.
33. Choi SM, Lee DS, Son MK, Sohn YS, Kang KK, Kim CY, Kim BM, Kim WB. Safety evaluation of GX-12: A new DNA vaccine for HIV infection in rodents. *Drug Chem Toxicol* 2003;26:271–84.
34. Xiang ZQ, Spitalnik SL, Cheng J, Erikson J, Wojczyk B, Ertl HC. Immune responses to nucleic acid vaccines to rabies virus. *Virology* 1995;209:569–79.
35. Tuomela M, Malm M, Wallen M, Stanescu I, Krohn K, Peterson P. Biodistribution and general safety of a naked DNA plasmid, GTU-MultiHIV, in a rat, using a quantitative PCR method. *Vaccine* 2005;23:890–6.
36. Liu MA, McClements W, Ulmer JB, Shiver J, Donnelly J. Immunization of non-human primates with DNA vaccines. *Vaccine* 1997;15:909–12.
37. Wolff JA, Ludtke JJ, Acsadi G, Williams P, Jani A. Long-term persistence of plasmid DNA and foreign gene expression in mouse muscle. *Hum Mol Genet* 1992;1: 363–9.
38. Ledwith BJ, Manam S, Troilo PJ, Barnum AB, Pauley CJ, Griffiths TG, Harper LB, Schock HB, Zhang H, Faris JE, Way PA, Beare CM, Bagdon WJ, Nichols WW.

Plasmid DNA vaccines: assay for integration into host genomic DNA. *Dev Biol* 2000;104:33–43.

39. Ledwith BJ, Manam S, Troilo PJ, Barnum AB, Pauley CJ, Griffiths TG, Harper LB, Beare CM, Bagdon WJ, Nichols WW. Plasmid DNA vaccines: investigation of integration into host cellular DNA following intramuscular injection in mice. *Intervirology* 2000;43:258–72.
40. Wang Z, Troilo PJ, Wang X, Griffiths TG, Pacchione SJ, Barnum AB, Harper LB, Pauley CJ, Niu Z, Denisova L, Follmer TT, Rizzuto G, Ciliberto G, Fattori E, Monica NL, Manam S, Ledwith BJ. Detection of integration of plasmid DNA into host genomic DNA following intramuscular injection and electroporation. *Gene Ther* 2004;11:711–21.
41. Liu MA, Ulmer JB. Human clinical trials of plasmid DNA vaccines. *Adv Genet* 2005;55:25–40.
42. Dean HJ. Epidermal delivery of protein and DNA vaccines. *Expert Opin Drug Deliv* 2005;2:227–36.
43. Montgomery DL, Ulmer JB, Donnelly JJ, Liu MA. DNA vaccines. *Pharmacol Therapeut* 1997;74:195–205.
44. Fuller DH, Loudon P, Schmaljohn C. Preclinical and clinical progress of particle-mediated DNA vaccines for infectious diseases. *Methods* 2006;40: 86–97.
45. Roy MJ, Wu MS, Barr LJ, Fuller JT, Tussey LG, Speller S, Culp J, Burkholder JK, Swain WF, Dixon RM, Widera G, Vessey R, King A, Ogg G, Gallimore A, Haynes JR, Heydenburg FD. Induction of antigen-specific CD8+ T cells, T helper cells, and protective levels of antibody in humans by particle-mediated administration of a hepatitis B virus DNA vaccine. *Vaccine* 2000;19:764–78.
46. Swain WE, Heydenburg FD, Wu MS, Barr LJ, Fuller JT, Culp J, Burkholder J, Dixon RM, Widera G, Vessey R, Roy MJ. Tolerability and immune responses in humans to a PowderJect DNA vaccine for hepatitis B. *Dev Biol* 2000;104:115–9.
47. Tatsis N, Ertl HC. Adenoviruses as vaccine vectors. *Mol Ther* 2004;10:616–29.
48. Basak SK, Kiertscher SM, Harui A, Roth MD. Modifying adenoviral vectors for use as gene-based cancer vaccines. *Viral Immunol* 2004;17:182–96.
49. Anonymous. Assessment of adenoviral vector safety and toxicity: report of the National Institutes of Health Recombinant DNA Advisory Committee. *Hum Gene Ther* 2002;13:3–13.
50. Fitzgerald JC, Gao GP, Reyes-Sandoval A, Pavlakis GN, Xiang ZQ, Wlazlo AP, Giles-Davis W, Wilson JM, Ertl HC. A simian replication-defective adenoviral recombinant vaccine to HIV-1 gag. *J Immunol* 2003;170:1416–22.
51. Polo JM, Dubensky TW, Jr. Virus-based vectors for human vaccine applications. *Drug Discov Today* 2002;7:719–27.
52. Zhou D, Cun A, Li Y, Xiang Z, Ertl HC. A chimpanzee-origin adenovirus vector expressing the rabies virus glycoprotein as an oral vaccine against inhalation infection with rabies virus. *Mol Ther* 2006;14:662–72.
53. Farina SF, Gao GP, Xiang ZQ, Rux JJ, Burnett RM, Alvira MR, Marsh J, Ertl HC, Wilson JM. Replication-defective vector based on a chimpanzee adenovirus. *J Virol* 2001;75:11603–13.
54. Berk AJ. *Adenoviruses In Encyclopaedia of Genetics*. San Diego: Academic Press, 2006;15–9.

55. Fattori E, Zampaglione I, Arcuri M, Meola A, Ercole BB, Cirillo A, Folgori A, Bett A, Cappelletti M, Sporeno E, Cortese R, Nicosia A, Colloca S. Efficient immunization of rhesus macaques with an HCV candidate vaccine by heterologous priming-boosting with novel adenoviral vectors based on different serotypes. *Gene Ther* 2006;13:1088–96.
56. Rasmussen RA, Ong H, Kittel C, Ruprecht CR, Ferrantelli F, Hu SL, Polacino P, McKenna J, Moon J, Travis B, Ruprecht RM. DNA prime/protein boost immunization against HIV clade C: safety and immunogenicity in mice. *Vaccine* 2006;24: 2324–32.
57. Singh M, O'Hagan DT. Recent advances in vaccine adjuvants. *Pharmaceut Res* 2002;19:715–28.
58. Brewer JM. (How) do aluminium adjuvants work? *Immunology Let* 2006;102: 10–15.
59. Pichyangkul S, Gettayacamin M, Miller RS, Lyon JA, Angov E, Tongtawe P, Ruble DL, Heppner DG, Jr., Kester KE, Ballou WR, Diggs CL, Voss G, Cohen JD, Walsh DS. Pre-clinical evaluation of the malaria vaccine candidate P. falciparum MSP1(42) formulated with novel adjuvants or with alum. *Vaccine* 2004;22: 3831–40.
60. Sesardic D. Regulatory considerations on new adjuvants and delivery systems. *Vaccine* 2006;24(Suppl 2):S2–S7.
61. Goldenthal KL, Cavagnaro JA, Alving CR, Vogel FR. Safety evaluation of vaccine adjuvants. National Cooperative Vaccine Development Working Group. *AIDS Res Hum Retroviruses* 1993;9:S45–S49.
62. Lebron JA, Wolf JJ, Kaplanski CV, Ledwith BJ. Ensuring the quality, potency and safety of vaccines during preclinical development. *Expert Rev Vaccines* 2005;4: 855–66.
63. Verdier F, Morgan L. Predictive value of pre-clinical work for vaccine safety assessment. *Vaccine* 2001;20:S21–S23.
64. Kennedy RC, Shearer MH, Hildebrand W. Nonhuman primate models to evaluate vaccine safety and immunogenicity. *Vaccine* 1997;15:903–8.
65. Boyson JE, Shufflebotham C, Cadavid LF, Urvater JA, Knapp LA, Hughes AL, Watkins DI. The MHC class I genes of the rhesus monkey: different evolutionary histories of MHC class I and II genes in primates. *J Immunol* 1996;156:4656–65.
66. Garcon N, Heppner DG, Cohen J. Development of RTS,S/AS02: a purified subunit-based malaria vaccine candidate formulated with a novel adjuvant. *Expert Rev Vaccines* 2003;2:231–8.
67. Anonymous. OECD guideline for testing of chemicals No. 404, 2006. http://www.oecd.org/
68. Chabicovsky M, Ryle P. Non-clinical development of cancer vaccines: regulatory considerations, *Regul Toxicol Pharmacol* 2006;44:226–37.
69. Callahan SM, Ming X, Lu SK, Brunner LJ, Croyle MA. Considerations for use of recombinant adenoviral vectors: Dose effect on hepatic cytochromes P450. *J Pharmacol Exp Therapeut* 2005;312:492–501.
70. Verdier F, Barrow PC, Burge J. Reproductive toxicity testing of vaccines. *Toxicology* 2003;185:213–9.
71. Gruber MF. Maternal immunization: US FDA regulatory considerations. *Vaccine* 2003;21:3487–91.

72. Leamy VL, Martin T, Mahajan R, Vilalta A, Rusalov D, Hartikka J, Bozoukova V, Hall KD, Morrow J, Rolland AP, Kaslow DC, Lalor PA. Comparison of rabbit and mouse models for persistence analysis of plasmid-based vaccines. *Hum Vaccines* 2006;2:113–8.
73. Haworth R, Pilling AM. The PCR assay in the preclinical safety evaluation of nucleic acid medicines. *Hum Exp Toxicol* 2000;19:267–76.
74. Vilalta A, Mahajan RK, Hartikka J, Rusalov D, Martin T, Bozoukova V, Leamy V, Hall K, Lalor P, Rolland A, Kaslow DC. I. Poloxamer-formulated plasmid DNA-based human cytomegalovirus vaccine: evaluation of plasmid DNA biodistribution/persistence and integration. *Hum Gene Ther* 2005;16:1143–50.
75. Naylor PH, Sztein MB, Wada S, Maurer S, Holterman D, Kirkley JE, Naylor CW, Zook BC, Hitzelberg RA, Gibbs CJ, Jr. Preclinical and clinical studies on immunogenicity and safety of the HIV-1 p17-based synthetic peptide AIDS vaccine—HGP-30-KLH. *Int J Immunopharmacol* 1991;13(Suppl 1):117–27.
76. Sheets RL, Stein J, Manetz TS, Duffy C, Nason M, Andrews C, Kong WP, Nabel GJ, Gomez PL. Biodistribution of DNA plasmid vaccines against HIV-1, ebola, severe acute respiratory syndrome, or West Nile virus is similar, without integration, despite differing plasmid backbones or gene inserts. *Toxicol Scis* 2006;91:610–9.
77. Hanke T, McMichael AJ, Samuel RV, Powell LA, McLoughlin L, Crome SJ, Edlin A. Lack of toxicity and persistence in the mouse associated with administration of candidate DNA- and modified vaccinia virus Ankara (MVA)-based HIV vaccines for Kenya. *Vaccine* 2002;21:108–14.
78. Cebere I, Dorrell L, McShane H, Simmons A, McCormack S, Schmidt C, Smith C, Brooks M, Roberts JE, Darwin SC, Fast PE, Conlon C, Rowland-Jones S, McMichael AJ, Hanke T. Phase I clinical trial safety of DNA- and modified virus Ankara-vectored human immunodeficiency virus type 1 (HIV-1) vaccines administered alone and in a prime-boost regime to healthy HIV-1-uninfected volunteers. *Vaccine* 2006;24:417–25.

# Preclinical Safety Evaluation of Biopharmaceuticals

MERCEDES A. SERABIAN, MS, DABT, and YING HUANG, PhD

**Contents**

1 Introduction
2 Definition of Gene Therapy
3 Mode of Gene Transfer
4 Distinctions between Preventive (Prophylactic) Viral and Plasmid DNA Vaccines and Therapeutic Gene Therapy Products
5 Historical Background of Gene Therapy Products
6 Worldwide Regulation of Gene Therapy Products
7 Complimentary Roles of the Food and Drug Administration (FDA) and Recombinant DNA Advisory Committee (RAC) in Gene Therapy
8 Vectors Used in Gene Therapy Products
  8.1 Nonviral Vectors
  8.2 Replication-Deficient Viral Vectors
  8.3 Other Viral Vectors
  8.4 Replication-Competent Oncolytic Vectors
9 Use of Adjuvants with Therapeutic Gene Therapy Products
10 Preclinical Studies for Gene Therapy Products
  10.1 Demonstrating Proof of Concept
  10.2 Toxicology Assessment
  10.3 Overall Toxicology Study Designs
  10.4 General Considerations
  10.5 Good Laboratory Practice (GLP)
  10.6 Non-GLP Studies
  10.7 Biodistribution Considerations
  10.8 Later Stage Product Development
11 Conclusion
References

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

## 1 INTRODUCTION

New discoveries in the field of gene therapy are occurring at a rapid pace, resulting in the identification of genes that contribute in some way to a particular disease or abnormal medical condition. Although marketed therapies may exist for many of these diseases, such as hemophilia or some forms of cancer, these pharmaceuticals can only treat the symptoms of the disease but cannot cure it. The hope of those working in the field of gene therapy is that the products that result from these discoveries will ultimately result in effective treatments, even cures, for many diseases or conditions that arise due to defective genes. The goal is the development of novel vectors engineered to express the "good" genes and delivered to the appropriate targets in the human body, where they will hopefully correct the medical condition for the person's lifetime or for a short period of time, as needed.

There is a scientific process by which to achieve this goal. This pathway involves the application of in vitro and in vivo preclinical studies (Figure 1) to generate data to support use of these novel therapeutic agents in humans. Preclinical studies for any experimental pharmaceutical agent, including gene therapy products, are conducted to provide better scientific insight into the potential pharmacologic and toxiologic effects that may result from the administration of the agent in humans. These studies are critical for helping to define the risk–benefit ratio of the experimental agent of interest in the desired clinical population, both prior to initiation of a clinical trial and throughout the development program for the pharmaceutical agent. Early-phase clinical trial design

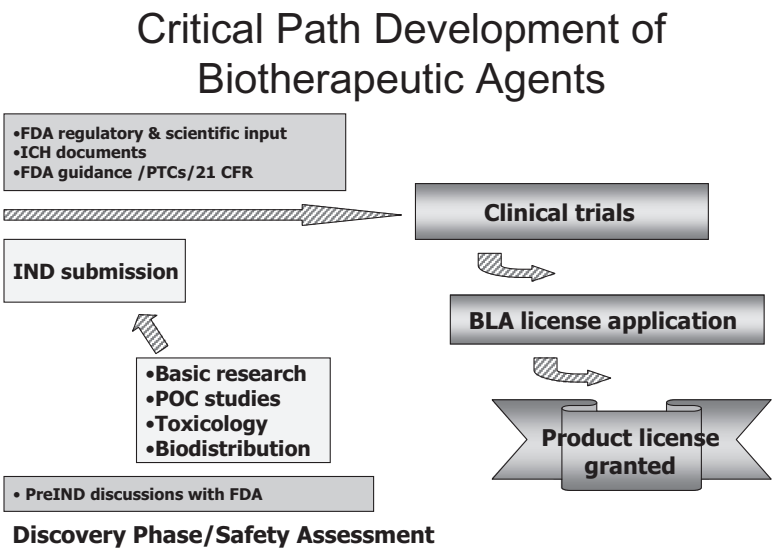


**Figure 1** Critical path development of gene therapy products in the United States.

can be constrained by inadequate preclinical testing of the investigational product. This chapter provides a perspective of the preclinical assessment of gene therapy products intended for clinical use in therapeutic indications.

## 2 DEFINITION OF GENE THERAPY

Gene therapy is a medical intervention based on the modification of the genetic material (DNA or RNA) of living cells. This intervention can involve the ex vivo modification of autologous, allogeneic, or xenogeneic cells, using replication-deficient viral vectors or nonviral vectors (e.g., plasmid DNA) for subsequent administration into humans, or the intervention may involve the direct in vivo administration of the gene therapy product into humans [1]. According to the scientific literature, experimental human gene therapy is currently being tested for the treatment of various genetic disorders, among these (1) diseases due to inborn errors in a single defective gene, such as severe combined immunodeficiency (SCID), hemophilia, or muscular dystrophy; (2) polygenic diseases, such as diabetes mellitus or coronary heart disease; and (3) acquired genetic diseases, such as cancer or rheumatoid arthritis. Theoretical approaches to correcting faulty genes include (1) insertion of a normal gene into a nonspecific location on the genome to replace a nonfunctional gene, (2) exchanging an abnormal gene for a normal gene through homologous recombination, (3) repair of the abnormal gene through selective reverse mutation that returns the gene to its normal function, or (4) the regulation of a particular gene (i.e., turning "on" or "off").

Only oncolytic viruses that carry a heterologous transgene are defined as gene therapy products. Other oncolytic viruses, such as the naturally attenuated and genetically modified viruses, are classified as viral therapy; these viruses will be briefly discussed in this chapter. The desired property for all oncolytic viruses is the potential to lyse the target tumor cells, although some oncolytic viruses may display other biological effects due to the expressed transgene. All oncolytic viruses are designed or modified to selectively replicate in tumor cells, but with an attenuated viral life cycle in normal cells. This mechanism is usually achieved through the mutation of the viral coding genes that are critical for viral replication in normal cells, but spared in tumor cells, or through the control of viral early gene expression by using tumor-specific promoters. The cellular targets for oncolytic viruses can be cell surface molecules and antigens that are recognized by these viruses for viral entry as well as intracellular components that are responsible for the host interaction with the viral components involved in the viral life cycle [2].

## 3 MODE OF GENE TRANSFER

Two basic modes of gene transfer currently exist for gene therapy products, namely ex vivo or in vivo gene transfer. Ex vivo gene transfer is achieved by

administering cells that are transduced with vectors that express the desired transgenes or tumor-associated antigens. A wide variety of somatic cells and hematopoietic cells can be transduced with viral vectors such as adenoviral and retroviral vectors. Certain types of somatic cells can be pulsed with naked DNA plasmid or transduced via a lipid/ligand DNA complex. Somatic cells can also be transduced with microorganism-based vectors, such as bacterial or yeast vectors. The transduced cells can then be administered via a route designed to deliver the gene therapy to the anatomic location where gene expression and the related biological activity are desired. In vivo gene transfer is the direct delivery of biological vectors and vector-based vaccines to the animal or patient. Direct injection can be in situ, which constitutes local delivery such as intratumoral injection, subcutanous injection, and others, or in vivo, which is direct systemic delivery (usually intravenous administration) into the individual.

## 4 DISTINCTIONS BETWEEN PREVENTIVE (PROPHYLACTIC) VIRAL AND PLASMID DNA VACCINES AND THERAPEUTIC GENE THERAPY PRODUCTS

Gene transfer products are used as therapeutic agents and as preventive vaccines. A preventive vaccine, consisting of a live-attenuated microorganism or virus, or plasmid DNA-based, is administered to elicit an immune response(s) that results in the prevention of infectious diseases in humans. On the other hand, a therapeutic vaccination with a gene therapy product is the transfer of a molecule with the intent to generate or intensify a host immune response to an antigen that is related to a disease already present in the patient. For therapeutic vaccines the immune response to the viral particles (i.e., neutralizing antibodies) should be avoided in order to achieve the desired effect in humans that is also safe. Although many of the scientific principles in this chapter can also be applied to viral or plasmid DNA vaccines used to prevent infectious diseases, the focus of this chapter is on the therapeutic gene therapy products that are being developed to treat established diseases (infectious or malignant).

## 5 HISTORICAL BACKGROUND OF GENE THERAPY PRODUCTS

The first gene therapy clinical trials in the United States began in the late 1980s with the ex vivo transduction of tumor-infiltrating lymphocytes (TIL) with a retroviral vector that carried the marker gene NeoR, administered to melanoma patients [3,4]. Another trial quickly followed in 1990, with the injection of autologous lymphocytes that were ex vivo transduced with a retroviral vector expressing the ADA gene into a child with ADA-deficient SCID [5]. The first trial involving the direct injection of a DNA plasmid occurred in 1992 with administration of plasmid expressing HLA-B7 into HLA-B7 positive

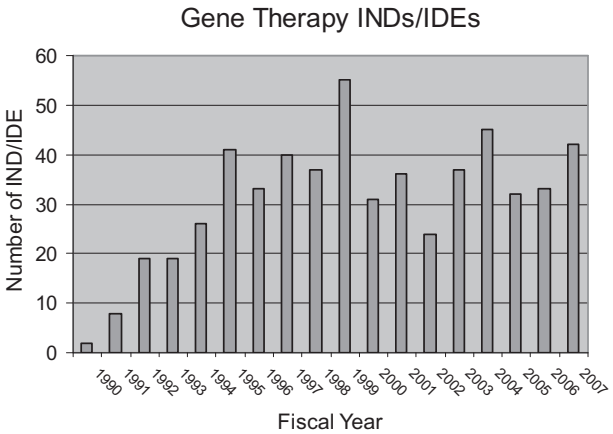


**Figure 2** Number of applications submitted to the FDA for the conduct of clinical trials using gene therapy products for therapeutic indications (fiscal year encompasses October of the previous year to October of the next year).

tumors [6]. The first clinical trial using adenovirus vector occurred in cystic fibrosis patients in 1993 [7], and was extended to the use of adeno-associated vectors (AAV) in this patient population in 1995 [8]. This was followed by accelerated numbers of clinical trials with a variety of vectors, peaking in 1999. However, with the death of Jesse Gelsinger, a teenager with a mild ornithine transcarbamylase (OTC) deficiency, in September 1999 following the intrahepatic infusion of an adenoviral vector expressing a normal OTC gene and the subsequent regulatory and public scrutiny, the number of new gene therapy clinical trials initiated in the United States was notably reduced (Figure 2).

Following this incident, numerous public meetings were held in the United States to discuss inadequacies in the scientific, ethical, and regulatory oversight of clinical trials involving gene therapy. As a result in the United States these study protocols now have more regulatory oversight by FDA/CBER, as well as wider assurance of public exposure and scientific and ethical discussions regarding this area of medical research through the Office of Biotechnology Activities (OBA) at NIH. Many of these research protocols are discussed in a public forum by the NIH Recombinant DNA Advisory Committee (RAC; http://www4.od.nih.gov/oba/RAC/RAC_FAQS.htm). The gene therapy field began to recover when reports of some of the early clinical trials, such as the use of an AAV vector expressing factor IX for hemophilia [9] and an adenoviral oncolytic vector for the treatment of head and neck cancer [10] showed some preliminary evidence of activity. However, many gene therapy trials have ended at an early stage, primarily due to the lack of a lasting biological effect with few products reaching phase 3 clinical trials in the United States. With the exception of one gene therapy product, an adenoviral vector expressing the p53 gene (rAd-*p53*; Gendicine) that was approved in China in 2003 for the

treatment of head and neck cancer [11] and a recombinant oncolytic adenovirus (H101; Oncorine) that was approved in China in 2005 also for the treatment of head and neck cancer [12], there are currently no marketed products in the world.

Elsewhere the success of the administration of autologous hematopoietic cells transduced ex vivo with retrovirus expressing the γc gene for the treatment of children with SCID-X1 disease was reported in 2000 in a clinical trial in France [13]. Unfortunately, several years after administration of the gene therapy product, several children developed a leukemia-like syndrome that was attributed to the combination of the integrating retrovirus and the γc gene itself [14]. As in the aftermath of the Gelsinger death, many public discussionswere held in the United States, both by the NIH (http://www4.od.nih.gov/oba/RAC/SSMar05/index.htm) and by the FDA (February 28, 2003—http://www.fda.gov/ohrms/dockets/ac/cber03.html#BiologicalResponseModifiers; and March 4,2005—http://www.fda.gov/ohrms/dockets/ac//05/transcripts/2005-4093T2_02.htm). Although these events conclusively demonstrated the biological complexity of the gene therapies and the real potential for adverse effects in the clinical population, it did not halt the progression of other clinical trials that used other gene therapy approaches, and in fact catalyzed a move forward with the engineering and testing of potentially safer vectors based on the knowledge that was gained from analysis of these adverse events.

## 6 WORLDWIDE REGULATION OF GENE THERAPY PRODUCTS

As evidenced by the marketing approval of the gene therapy product Gendicine in China, clinical trials of gene therapy products is extending worldwide. Many regulatory bodies are now responsible for evaluating the safety and effectiveness of gene therapy products. In Europe, the European Medicines Agency (EMEA) plays a vital role in the marketing procedure for medicines in the European Union through the Committee for Medicinal Products for Human Use (CHMP). The EMEA/CHMP coordinates the evaluation and supervision of medicinal products via both "centralized (community)" and "decentralized (mutual-recognition)" mechanisms. One of the CHMP working parties, the Gene Therapy Working Party (GTWP), is responsible for providing recommendations to the CHMP on all matters that relate to gene therapy, including the development of regulatory guidance (http://www.emea.europa.eu/htms/human/genetherapy/genetherapy.htm).

Other European agencies that play a role in the regulation of gene therapy products in member states include the French Agency for the Sanitary Safety of Health Products (AFSSAPS; http://agmed.sante.gouv.fr/), the Gene Therapy Advisory Committee (GTAC) in the United Kingdom (www.doh.gov.uk/genetics/gtac/index.htm), and the Paul Erlich Institute in Germany (www.pei.de/themen/gentherapie/gene_therapy_reg.htm).

Regulatory bodies in Asia that evaluate gene therapy products include the Ministry of Health, Labour and Welfare (MHLW) (www.mhlw.go.jp/english) and the Pharmaceuticals and Medical Devices Agency (www.pmda.go.jp/index-e.html) in Japan, the Korean Food and Drug Administration (KFDA) in South Korea (www.kfda.go.kr), and the State Food and Drug Administration (SFDA) in China (www.sfda.gov.cn/eng).

The International Conference on Harmonisation (ICH) has officially recognized the rapidly progressing area of gene therapy through the creation of the Gene Therapy Discussion Group (GTDG), which is made up of regulatory and industry representatives from the United States, European Union, and Japan, as well as experts from Health Canada, WHO, and the European Free Trade Association (EFTA). This group is charged with the monitoring of emerging scientific issues related to gene therapy and communicating principles that may benefit the harmonization of the regulation of gene therapy products through public ICH workshops, and a publicly accessible ICH gene therapy web page containing ICH press statements, GTDG meeting minutes, ICH documents related to gene therapy and links to other major gene therapy public workshops. (http://www.ich.org/cache/html/1386-272-1.htm). In addition the World Health Organization (WHO) has also published guidelines for the nonclinical safety of DNA vaccines (http://www.who.int/).

## 7 COMPLIMENTARY ROLES OF THE FOOD AND DRUG ADMINISTRATION (FDA) AND RECOMBINANT DNA ADVISORY COMMITTEE (RAC) IN GENE THERAPY

The FDA and the National Institutes of Health (NIH) in the United States have complimentary responsibilities with respect to the regulation of human gene therapy. The NIH, through the Office of Biotechnology Activities (OBA), is mandated by Congress to review all clinical trials conducted using gene therapy (1) if the sponsor of the research or the institution where the trial will be conducted receives any money from NIH for DNA research and/or (2) if the institution that develops the product receives funding from NIH for recombinant DNA research. The DNA Recombinant Advisory Committee (RAC), under the purview of OBA, consists of a review panel of up to 21 experts and serves an advisory role to the NIH director and to OBA. The RAC holds regular public meetings to discuss proposed gene therapy trials and issues, focusing on the scientific, safety, and ethical issues involved. Material submitted to the RAC are available for public view [15]. The RAC has no regulatory oversight; this function is held exclusively by the FDA.

The FDA is mandated by US statue as the regulatory body that is responsible for the oversight of pharmaceuticals (including biologics) [16,17]. FDA's primary role is to ensure that manufacturers produce high-quality, safe, and efficacious biologics, drugs, and devices that are properly studied in human subjects prior to approval for commercial marketing. The FDA Center for

Biologics Evaluation and Research (CBER) is responsible for the regulatory review of gene therapy products. All clinical trials conducted in the United States that involve the administration of investigative gene therapy products must be reviewed by CBER. Sponsors of experimental gene therapy products must file a "Notice of Claimed Investigational Exemption for a New Drug" (IND) with the FDA. This regulatory body is responsible for determining whether a particular gene therapy product may be administered to humans as defined by the proposed clinical trial protocol [18]. FDA requires the submission of preclinical data to support the planned clinical trial (21 CFR 312.23), which allow the FDA reviewers to conduct independent reviews to assess the safety of each respective gene therapy product. Unlike the RAC review, the FDA's review is not conducted in a public forum and all information submitted to FDA is confidential (http://www.fda.gov/cber/gene.htm).

## 8 VECTORS USED IN GENE THERAPY PRODUCTS

Gene therapy is one of the most rapidly evolving areas of clinical research. There are a large and increasing number of different vector types that are currently being studied in animals and/or in humans. Many of these vectors are engineered to carry specific genetic material into the cells or tissues of individuals with a myriad of diseases, while others are designed to kill cancer cells by their pathogenic nature alone. The fundamental challenges to generating an effective gene therapy vector for treating diseases are accomplishing efficient gene delivery, obtaining biologically relevant levels of transgene expression in the desired target cell(s), and maintaining expression of the transgene without interfering with host gene expression.

Presented below are summary descriptions of gene therapy vector types that are most representative of the products administered in clinical trials.

### 8.1 Nonviral Vectors

A nonviral vector consists of "naked" plasmid (plasmid DNA alone) or plasmid DNA formulated with other entities designed to assist cell entry of the DNA complex via various mechanisms such as endocytosis; frequently these formulations include an artificial lipid sphere with an aqueous core (liposomes), basic proteins, receptor ligands, or polymers. Generally, nonviral vectors have shown safety advantages over viral vectors because of the reduced cytotoxicity, pathogenicity, and potential for insertional mutagenesis. Potential safety concerns for nonviral vectors include (1) DNA tissue biodistribution and any subsequent adverse biological effects, (2) the potential for insertional mutagenesis mediated by the elements incorporated in the plasmid backbone sequence, (3) clinical toxicities from the formulated plasmid DNA product, (4) toxicities due to the delivery procedure of the plasmid DNA, and (5) toxicities

**Nonviral Vectors—Plasmid DNA**

Advantages
- Cells do not have to be dividing for transduction
- Episomal replication; low risk of insertional mutagenesis
- Can be directly transferred to host
- Can accommodate large genes

Disadvantages
- Transduction efficiency usually low—need lipid or other transfer technique (i.e., electroporation)
- Duration of expression is low (degradation of gene in cytoplasm
- Potential interaction with normal cells
- Expression of foreign proteins—concern for autoimmunity

**Figure 3** Characteristics of plasmid DNA nonviral vectors.

resulting from the expressed protein, such as autoimmunity. While the potential for insertional mutagenesis for plasmid vectors is a safety concern, it remains theoretical as there have been no reports of the occurrence of such an event as a result of the administration of nonviral gene therapy products in animals or humans. This is likely due to the mechanism of episomal replication mediated by the plasmid vector backbone, which contains elements that confer the two essential requirements for extrachromosomal stability: replication of the DNA during the cell cycle and segregation to daughter cells during mitosis [19]. Some characteristics of plasmid DNA as vectors are presented in Figure 3.

Plasmid DNA vectors are delivered mainly via direct injection, as the injected DNA can be taken up by a variety of targeted somatic cells, such as skeletal muscle and cardiac muscle. Despite direct injection the injected plasmid DNA retains the ability to distribute via hematogenous spread to nontarget tissues. This DNA biodistribution profile is usually assessed in preclinical studies. Persistence of the injected plasmid DNA at the injection site in rodents for greater than one year following intramuscular injection without chromosome integration has been reported [20]. Several physical methods have been developed to enhance the transfection of plasmid DNA delivered to local tissue. For example, electroporation, which is commonly used, involves the application of controlled electric pulses via a medical device delivered to the DNA injection site following plasmid administration in order to enhance the transfection of plasmid DNA and to contain the DNA at the injection site. The electroporation procedure itself can cause local toxicity such as muscle degeneration, thus the safety of the complete delivery procedure (i.e., device settings) should be assessed in animals.

The plasmid DNA is provided in a vehicle formulation to enable direct injection of the product in vivo. This vehicle may be as simple as phosphate-

buffered saline, or it can be more complex; for example, cationic lipid–DNA complexes have been widely used in clinical trials. The potential toxicities of these formulations depend on the physical and chemical properties of the liposome(s) and on the potential interaction of these lipids with the plasmid DNA. Yew and Scheule reported that the body's response to cationic lipid–DNA complexes is highly dependent on both the dose level and the route of administration. High-dose levels of a cationic lipid–DNA complex administered intravenously in mice can activate the innate immune response, resulting in the induction of proinflammatory cytokines and immune cell activation, with lethality as the worst-case outcome. These investigators found that some of these reactions appeared to be related to the structure of the plasmid DNA that was formulated with the lipid [21]. Safety assessment is dependent on the type of lipid(s) that constitute the final DNA complex, the stability of the complex, the route of administration of the complex, and the dose levels administered.

## 8.2 Replication-Deficient Viral Vectors

***Adenoviral Vectors*** Adenoviral vectors (Ad) are one of the most extensively used viral vectors for ex vivo and in vivo gene transfer that have been studied for use in a diversity of diseases, including cancer. This is mainly due to the advantages these double-stranded DNA viruses have over other viral vectors, such as their high infectivity in certain species including humans, lack of direct virally mediated pathogenesis in their natural hosts including humans, and transduction of a wide spectrum of both proliferating and quiescent cell types. In addition the adenoviral genome remains episomal in the transduced cell, and thus does not have a propensity to integrate, greatly reducing the concerns of insertional mutagenesis. The replication-deficient state of frequently used Ad vectors is established by the deletion of critical viral regulatory genes such as E1 and E3. The deletion of these viral early genes also allows the vectors to accommodate significant lengths of heterologous transgene and promoter sequences to achieve the desired transgene expression. Several generations of replication-deficient Ad vectors have been developed, and the manufacturing process for some of them has been successfully scaled up for clinical use. The "first-generation" Ad vectors have the deletion of the early genes: E1-deleted or E1 and E3-deleted. The "second-generation" Ad vectors have the deletion of E2 and/or E4 genes, in addition to the deletion of E1 or E1 and E3. Helper-dependent Ad vectors have most of the Ad coding sequences deleted. These gutless vectors can be classified as "third-generation" Ad vectors.

Potential safety concerns for Ad vectors include (1) DNA tissue biodistribution and any subsequent adverse biological effects and (2) significant immune response and inflammatory response to the vector. Some characteristics of Ad vectors are presented in Figure 4. The use of nonhuman Ad vectors to circumvent the drawbacks of using human Ad vectors, such as an immune

**Adenoviral Vectors**

Advantages

- Cells do not have to be dividing for transduction
- Able to target cells bearing capsid receptors
- Wide range of tissue and host infectivity
- Episomal replication; low risk of insertional mutagenesis

Disadvantages

- Inflammatory immune response: capsid proteins, viral genes expressed by target cells
- Potential for recombination/complementation

**Figure 4** Characteristics of adenoviral vectors.

response and an inflammatory response to the protein capsid that coats the virus and/or to the viral genes that are expressed by target cells, is emerging [22]. The safety and viral biodistribution for each type of Ad vector should be established prior to use in humans.

***Adeno-associated Viral Vectors (AAV)*** Wild-type AAV is a member of the *dependovirus* genus of the family *Parvoviridae*, and is a nonpathogenic human DNA virus. Multiple serotypes of these single-stranded DNA viruses have been identified, and neutralizing antibodies against the AAV capsid proteins are frequently detected in human serum. Recombinant AAV vectors (rAAV) contain no viral-coding sequence. In particular, they do not express Rep proteins that play a key role in the DNA replication, site-specific integration, and cellular growth inhibitory effects of AAV. Thus the transgene and the virus capsid are the only source of foreign antigens. The documented spread of rAAV to the gonadal tissues of various animal species and the presence of rAAV DNA in semen, peripheral blood, and sputum of human subjects injected with rAAV2 vectors raises the concern regarding the risk of rAAV dissemination to various human tissues, especially the reproductive organs [23].

Although AAV remains episomal in the transduced cell, thus reducing the concerns of insertional mutagenesis, the potential to randomly integrate into host DNA exists. There have been no reports of insertional mutagenesis and resulting tumorigenicity from human trials using rAAV-based gene therapy. Although tumors reported in a murine disease model were attributed to rAAV-based gene transfer, tumors in humans attributable to injected AAV vectors have not occurred [24]. The observed persistence of rAAV DNA sequences in human biological samples also raises concerns regarding the potential for horizontal transmission from the subjects injected with the vector and their family members and health care providers.

## AAV Vectors

Advantages
- Cells do not have to be dividing for transduction
- No known pathology of wild-type AAV
- Able to target multiple cell types
- Episomal replication; low risk of insertional mutagenesis
- Long-term gene expression

Disadvantages
- Limitations in accommodating the size of heterogeneous sequence
- Potential for recombination and/or complementation
- Potential for immune response to capsid protein
- Unable to halt gene expression if toxicity occurs
- Potential for dissemination and germline transfer

**Figure 5** Characteristics of adenoviral-associated viral (AAV) vectors.

Potential safety concerns for AAV vectors include (1) DNA tissue biodistribution and any subsequent adverse biological effects, (2) the potential for insertional mutagenesis and any subsequent adverse biological effects, and (3) the immune response to the vector. Some characteristics of AAV vectors are presented in Figure 5.

***Retroviral and Lentiviral Vectors*** Recombinant retroviral vectors (*gammretrovirus* genus, *Retroviridae* family) can only transduce actively dividing cells such as hematopoietic cells and stem cells. Although engineered to be replication deficient, retroviral vectors can provide long-term expression because they efficiently and stably integrate into the host genome after the infection of host cells. However, safety concerns regarding the pathogenic potential of retroviral vectors exist: (1) the production of a replication competent retrovirus (RCR) and (2) the potential for insertional mutagenesis, resulting in oncogene activation. An instance where vector contaminated with RCR was injected into monkeys resulted in the development of a T cell lymphoma, presumably due to oncogene activation [25]. The recent findings in the clinical trial using ex vivo retrovirus transduced hematopoietic cells for the treatment of X-linked SCID, have provided the proof that (1) retroviral vector-derived DNA integrated into the cellular DNA of several subjects and (2) this integration resulted in several oncogenic events [26; NIH Recombinant DNA Advisory Committee Meeting, March 11, 2008]. Much scientific discussion on vector engineering to decrease the probability of recombination and insertional mutagenesis is ongoing. Considerations include (1) the use of self-inactivating (SIN) vectors that delete most of the U3 region of the 3′LTR (long terminal repeat) because this site harbors the major transcriptional functions of the retroviral genome, (2) development of vectors from non-oncogenic

**Retroviral Vectors**

Advantages
- Able to transduce dividing cells
- Stable, long-term gene integration into host genome
- High levels of gene expression

Disadvantages
- Target cells have to be actively dividing; low percentage of cells transduced
- Potential for contamination by RCR
- Potential for insertional mutagenesis
- Potential for germline transfer

**Figure 6** Characteristics of retroviral vectors.

retroviruses (e.g., lentivirus), and (3) the incorporation of insulator sequences [27]. Additional potential safety concerns for retroviral vectors include (1) DNA tissue biodistribution and any subsequent adverse biological effects and (2) the potential for germline integration following in vivo gene transfer. Some characteristics of retroviral vectors are presented in Figure 6.

Lentiviruses (*lentivirus* genus, *Retroviridae* family) can originate from five serogroups: primates, sheep and goats, horses, cats, and cattle. The replication-defective, attenuated lentiviral vectors that have been administered to humans thus far are HIV based. The use of HIV as a basis for lentiviral vectors is attractive because it infects dividing, nondividing, and terminally differentiated cells; it can accommodate long sequences that are stably expressed due to integration into the cell chromosome; and it appears to be less immunogenic. With retroviral vectors, the product used to "treat" many diseases (e.g., severe combined immunodeficiency) is the ex vivo manipulation of stem cells because actively dividing cells are a prerequisite for successful retroviral vector transduction. However, this process is not necessary for lentiviral vectors; they can be administered by direct in vivo injection into humans. Lentiviral vectors carrying a specific transgene have been associated in the scientific literature with possible use in many diseases, such as metabolic diseases, cancer, and neurological diseases, among many others. For example, injection of a lentiviral vector expressing glial cell-derived neurotrophic factor (GDNF) into the striatum and substantia nigra of a rhesus monkey Parkinson disease model resulted in improvement of neurological function and prevention of nigrostriatal degeneration [28].

The lentiviral vectors have safety concerns similar to the retroviral vector-based gene therapy products: (1) the production of a replication competent lentivirus (RCL) during manufacturing and (2) the potential for insertional mutagenesis, resulting in oncogene activation. One concern unique to the lentiviral vectors is the possibility that the vector can be mobilized in vivo,

**Lentiviral (HIV-based) Vectors**

Advantages
- Infects dividing, nondividing, terminally differentiated cells
- Highly attenuated, SIN constructs
- Inserts into host genome like HIV; stable integration and long-term gene expression
- Less immunogenic

Disadvantages
- Generation of replication-competent virus (RCL)
- Viral mobilization in vivo; infection of human subject
- Unable to halt gene expression if toxicity occurs
- Potential for insertional mutagenesis (Lentiviruses have similar, site-specific integration to HIV)
- Potential for germline transfer

**Figure 7** Characteristics of lentiviral vectors.

resulting in infection of the injected subject with newly recombinant or wild-type HIV. Various safety modifications are being considered by researchers, including (1) production of the vector using a packaging plasmid that does not contain HIV genes with the potential for facilitating viral recombination and replication, (2) vectors that are self-inactivating (SIN), and (3) testing for specific internal promoters that regulate gene expression but do not introduce the potential for mediating undesired gene expression such as oncogenes. In addition the use of other lentiviruses, such as equine infectious anemia virus (EIAV), is being considered as a strategy to circumvent the issue of host infection. Additional potential safety concerns for these vectors include (1) DNA tissue biodistribution and any subsequent adverse biological effects and (2) the potential for germline integration. Some characteristics of lentiviral vectors are presented in Figure 7.

### 8.3 Other Viral Vectors

Other types of replication-deficient viral vectors that have been used in the gene therapy field include *Herpes simplex* viral (HSV) vectors that (1) are able to transduce nondividing cells and (2) are highly infective for neurologic tissue and vaccinia vectors. Vaccinia vectors in turn (1) are able to transduce nondividing cells and (2) have the ability to efficiently infect many types of cells. The primary safety concerns for HSV vectors are the potential for tropism to the CNS and the potential for latency and reactivation. Vaccinia vectors contain the same backbone as the smallpox vaccine, thus the available safety databases for vaccinia administration in humans consist primarily of preventive vaccination in a healthy population. Principal safety concerns with the use of vaccinia vectors include (1) their ability to replicate in humans and possibly

the production of a replication competent vaccinia (RCV), (2) the potential for germline integration, and (3) the potential for toxicity in immune-compromised populations such as cancer patients. Any safety concerns regarding the route of administration and dosing regimens for therapeutic vaccinia vectors should be addressed before the vector is used in humans.

### 8.4 Replication-Competent Oncolytic Vectors

The oncolytic viruses include adenovirus, measles, reovirus, vesicular stomatitis virus (VSV), HSV, poxvirus, and vaccinia. Specific examples include (1) ONYX-015, which is an adenoviral oncolytic virus, administered to patients with liver metastases of colorectal cancer and pancreatic cancer [29], (2) Reolysin, which is an oncolytic reovirus administered to patients with glioma [30], and (3) MV-CEA, which is an oncolytic measles virus expressing carcinoembryonic antigen, administered to patients with ovarian cancer [31]. Some oncolytic viruses are wild type and are apparently not pathogenic in humans, such as the Newcastle disease virus (NDV), which is an RNA avian paramyxovirus. PV701, a naturally attenuated, replication-competent strain of NDV, has been administered to patients with advanced solid tumors [32]. The applicability of oncolytic viruses as a therapy for clinical oncology trials is due to their potential selectivity: the ability to kill tumor cells but not normal cells. However, the level of attenuation of viral replication in normal cells is limited for most oncolytic vectors.

The desired properties of these viruses include (1) stable and efficient viral replication in tumor cells in vivo, (2) lateral spread of virus to surrounding tumor cells (i.e., "bystander effect"), (3) tumor-selective killing, also called "controlled" tropism, and (4) avoidance of early detection/"neutralization" by the immune system. Potential safety concerns for this vector class include (1) the potential for viral tropism to normal human cells, (2) increased viral spread and replication in nontarget tissues in immunosuppressed individuals, (3) synergistic potential for viral replication in nontarget tissues when administered in combination with radiation, chemotherapy, prodrugs, and other agents, and (4) any toxicities due to an added transgene. One major challenge for safety assessment of these products is the selection of an animal species that reflects the biology of the respective oncolytic virus in humans (e.g., the use of transgenic mice expressing one of the receptors, CD46, for measles virus) and the availability of adequate tumor-bearing models that can be used to assess antitumor activity.

## 9 USE OF ADJUVANTS WITH THERAPEUTIC GENE THERAPY PRODUCTS

Similar to prophylactic vaccines, many of the therapeutic gene therapy products are administered in conjunction with adjuvants in order to enhance

the immune system to respond in a vigorous manner to the injected antigen. The adjuvants can include aluminum salts (aluminum hydroxide, aluminum phosphate, and potassium aluminum sulfate), QS-21 (a saponin extract from the *Quillaja saponaria* [soapbark] tree), bacterial adjuvants, liposomes, and the newer MF59 (an oil-in-water emulsion manufactured by Chiron) and ISCOMATRIX(r) (a saponin-based adjuvant manufactured by CSL Limited). Recombinant proteins such as granulocyte-monocyte colony stimulating factor (GM-CSF), interleukin-2 (IL-2) and interleukin-12 (IL-12), which also serve to stimulate the immune system, have been administered in combination with many therapeutic gene therapy products. The adjuvant list is continually growing as a result of the development of new therapeutic and preventive vaccines that may require novel adjuvants.

The use of an adjuvant concurrent with a gene therapy product could result in toxicity due to the molecular mimicry between a vaccine antigen and a self antigen [33]. By their intended action of increasing the immune response, adjuvants administered with a gene therapy product can potentially exacerbate some toxicities because of an inappropriate immune response. In addition the adjuvants can be toxic, depending on their intrinsic chemical properties. Safety concerns regarding the adjuvant alone can include (1) the potential for proinflammatory effects of immunomodulatory adjuvants like cytokines, (2) the induction of hypersensitivity and autoimmunity, and (3) the intrinsic toxicity of these compounds [34]. General strategies to assess the safety of therapeutic gene therapy vaccines administered with adjuvants should include evaluation of the intended clinical formulation (gene therapy product plus adjuvant) when administered via the intended clinical immunization regimen using a relevant animal species. A major challenge for gene therapy vaccine–adjuvant combination is safety assessment using animal species that display a response similar to the human host response to both the vaccine and the adjuvant. This consideration is more complex in diseased individuals, who may be more vulnerable to any potential risks of therapeutic vaccination compared to normal human populations. For a general scientific discussion and toxicology testing principles for adjuvants, the reader is referred to the adjuvant guideline published by the EMEA [35].

## 10 PRECLINICAL STUDIES FOR GENE THERAPY PRODUCTS

Preclinical studies for any experimental pharmaceutical agent are conducted to provide better scientific insight into the potential pharmacologic and toxiologic effects that may result from the administration of the agent in humans. These studies are critical as they help define the risk–benefit ratio of the experimental agent of interest in the desired clinical population, both prior to initiation of a clinical trial and throughout the development program for the pharmaceutical agent.

### 10.1 Demonstrating Proof of Concept

The initial definitive first step in the movement from the "bench" to the "bedside" is the supportive scientific justification of the use of a particular agent to substantiate an appropriate rationale for use in a specific disease indication (Figure 1). This rationale, termed proof-of-concept (POC) studies, focuses on assessing the bioactivity and pharmacology profile of a particular gene therapy product to understand the ability of that product to induce the desired biologic effect via both in vitro (e.g., the trophic effect of an expressed transgene on a specific receptor or cell population) and in vivo (e.g., the use of an animal disease model) systems in order to enable advancement to first use of the product in humans. These POC studies are the initial translation step, as they help establish the basis for conducting the desired clinical trial through the generation of data to show (1) the biological feasibility of the use of the gene therapy product in the target disease, (2) the level of gene expression resulting from administration of the product (the "pharmacokinetic" effect), and (3) the extent of the morphological and functional correction (the "pharmacodynamic" effect) that results from administration of the product. In addition these POC studies help define a pharmacologically active dose range, with establishment of an optimal biological dose (OBD) and a minimally effective dose (MED), determine a potentially optimal route for product administration, and provide guidance regarding a possible dosing regimen/schedule for the first clinical trial. This pharmacological effect may translate to correction of a genetic defect, such as with an inborn error of metabolism; the amelioration or slowed progression of a disease, such as cancer; or lessened severity/frequency of some of the symptoms of a particular disease, such as Parkinson's disease.

For gene therapy products, the selection of the vector system is put to the test initially in the POC studies. These studies provide the first in vitro assessment of the choice of the viral or nonviral vector, the cell population to be used (if applicable), as well as the selection of a therapeutic transgene (if applicable) to treat a specific disease.

***Consideration of Relevant Animal Species*** One very important goal of the POC studies for a gene therapy product is to establish at least one biologically relevant animal species (if one exists). This selection process can entail assessment of the permissiveness/susceptibility of various animal species to replication by viral vectors, the pharmacological responsiveness to the transgene, the sensitivity of the species to the actions of the administered population of cells, and the comparative physiology of the animal species to humans. Relevancy determination may not stop at the species level as, for example, the responsiveness of particular strains of mice to the gene therapy product of interest may also need to be evaluated. For instance, Mazzolini et al. showed that following systemic injection of AdCMVmIL-12, heterogeneous toxicity was evident in different strains of mice depending on the levels of gene trans-

duction achieved in the livers of these various strains [36]. For some gene therapy products, the relevant species may not be considered a "standard" laboratory animal species for preclinical testing, such as the applicability of cotton rats in the evaluation of the activity and safety of adenovirus [37] or the use of Syrian hamsters to assess oncolytic adenovirus [38]. The expressed transgene also needs to be considered when determining a relevant animal species, as many of the recombinant human protein products have been shown to be biologically active in a restricted number of species, such as human interferon activity in nonhuman primates. For some POC studies that are conducted in rodents the analogous animal transgene (i.e., murine interferon) is used. In these instances additional characterization of the human product and the animal product would be conducted to determine the similarities and differences between the human and animal investigational materials. There are many other examples in the scientific literature, both for the vectors and for the expressed transgenes.

The injection of cells ex vivo transduced to express a specific transgene, such as autologous $CD8^+$ cells transduced with lentiviral vector to express an anti-sense HIV envelope (env) gene [39] or hematopoietic stem cells transduced with retrovirus to express the FAA gene [40], highlights another aspect of the biological relevancy of animals. The injection of human cells into immune competent animals cannot be accomplished due to the immune response of the animal to the foreign cells that will occur. Therefore, for the conduct of POC studies using human transduced cells, the use of immune-incompetent animals, such as genetically immunodeficient rodents (SCID mice, nude rats) or chemically immunosuppressed rodents and large animals (e.g., using cyclosporine), is necessary in order to prevent or slow the rejection of the cells. Another approach that has been used is the administration of the animal cellular analogue, which has been transduced with the clinical vector to express the human transgene into immune competent animals. POC studies have also been conducted using the animal cellular analogue, transduced with the clinical vector encoding the animal transgene, notably in instances where a rodent species has been used in which the human protein was not biologically active. One such example is human autologous bone marrow derived stromal cells (hBMSCs) transduced with plasmid expressing human IL-12 (hIL-12) for peritumoral injection into patients with glioblastoma. Some POC studies were performed using tumor-bearing immune-competent rodents injected with rBMSCs transduced with plasmid expressing mIL-12 [41]. In this situation characterization of the human product—cells and transgene—and the animal product would be conducted to determine significant differences. The degree of understanding of the relationship between an animal cell and its human correlate is an important factor in determining the strength of the extrapolations from findings in animals to human risk assessment.

Another factor that may enter into the selection paradigm for the most appropriate species for evaluation of a specific gene therapy product is the route/method of product administration. For example, if the delivery of the

gene therapy product incorporates a system that is more complicated than a needle and syringe (i.e., catheter-pump device delivery system or intracoronary infusion by a cardiac catheter), then the use of a pharmacologically responsive large animal species (if one exists) can also be considered in order to adequately evaluate the activity of the gene therapy product when dispensed according to the anticipated delivery procedure for humans. At times, however, a large animal species may be necessary if an untested delivery device is contemplated for clinical use, in order to evaluate the resulting activity, such as vector spread and resulting transduction pattern in the targeted tissue due to the administration procedure itself.

Knowledge of a product's mechanism of action that is acquired in preclinical trials can be beneficial in the design and conduct of the clinical trials. Ideally the pharmacological response of the species to the gene therapy product will mimic the response seen in humans—which will only be verified following initiation of a clinical trial.

***Animal Models of Disease*** For many of the diseases that are under consideration for treatment with a gene therapy product, a corresponding animal model of disease exists, which to some extent, may reflect the pathophysiology of the disorder. Figure 8 depicts some examples of animal models of disease that have been used to study the pharmacological activity of various gene therapy products intended for treatment of a specific human disorder. Existing models include spontaneous disease models such as hemophilic dogs,

**Animal Models of Disease to Study Potential Gene Therapies**

| Animal Model | Disease |
|---|---|
| CLAD dogs | Leukocyte adhesion deficiency |
| *Gusb* mice | MPS VII; Sly disease |
| *mdx* mice | Muscular dystrophy |
| Canine X-linked MD (GRMD) | Duchennes MD |
| db/db mice | Diabetes |
| Factor IX-deficient dog | Hemophilia B |
| PDAPP Tg mice | Alzheimers disease |
| Aged marmosets | Alzheimers disease |
| *rcd1* dogs | Retinitis pigmentosa |
| $W/W^v$ mice | Fanconi's anemia |
| $Hbb^{th3/+}$ mice | ß-thalassemia |
| Murine tumor xenografts | Cancer |
| Spontaneous tumors in companion animals | Cancer |

**Figure 8** Examples of animal models of disease that have been used to study potential gene therapy agents.

nonspontaneous disease models (chemically, surgically, immunologically induced) such as the MPTP nonhuman primate model of Parkinson's disease, and genetically modified disease models (transgenics, knockouts, knock-ins) such as the transgenic murine HER-2 model for breast cancer [42].

POC studies using an animal model of disease can then be designed in a manner that reasonably mimics the planned clinical protocol, including the gene therapy product administration schedule and the parameters (both activity and safety) monitored. For example, in a porcine myocardial infarct model, product administration considerations could include the timing of product delivery (using the clinical delivery system) relative to the appearance of the infarct, the number and location of the injections of product into the heart and the product concentration, volume, and rate of administration. The parameters monitored could include functional assessments such as ECGs, echocardiography, and/or SPECT and morphological endpoints such as hematology, serum biochemistry, and cardiac markers (i.e., troponin, LDH, and CPK/CPK-MB) and a comprehensive microscopic examination of the heart. A preclinical study designed in such a manner, using several vector dose levels, would provide insight into the determination of an optimal biological dose level, as well as a no-effect dose level, which would then help to guide the rationale for selection of the dose levels for the subsequent toxicology studies, as well as for the human trial.

While the use of preclinical disease models offers many advantages, such as the potential for identification of biomarkers for clinical trials and a direct estimation of the therapeutic index, there are also disadvantages to this approach. The disadvantages can include a paucity of historical/baseline data and inherent variability of the model, the technical feasibility of using a particular disease model, the onset/severity of the injury in the animals as this may be different from humans, and the fact that the model may only emulate select aspects of the human pathophysiology of the disease [43]. It is important to use behavioral models that have adequate sensitivity and specificity to allow for correlation with morphological and biochemical findings, as in the area of spinal cord injury.

## 10.2 Toxicology Assessment

According to the paradigm depicted in Figure 1, the next step on the critical pathway to use of a gene therapy product in a human disease population is the evaluation of the safety of the product of interest through the conduct of toxicology studies. These studies are intended to provide an acceptable risk–benefit ratio to the patients that will receive a particular gene therapy product. While the preclinical studies focus on goals that are similar for small-molecule agents or for traditional biotechnology-derived products (e.g., monoclonal antibodies, cytokines, growth factors), gene therapy products are complex and preclude a standard design of preclinical studies, which is consistent with the

"case-by-case" approach to the safety assessment of biotechnology-derived pharmaceuticals.

***Relevant Animal Species/Models*** The evaluation of the potential toxicity of each gene therapy product should build on the previously conducted POC studies. Important factors for continued product development such as the animal species, the route and procedure for product administration, the potentially effective dose range, and the dosing regimen will have been established from the POC results, providing the rationale for proceeding into preclinical toxicology studies. The collection of toxicity endpoints in the disease models that are used to assess pharmacologic activity is scientifically justified because, if a relevant disease model exists and is readily accessible, the contribution of the disease-related changes in physiology and/or the underlying pathology to any toxicities observed with the use of a particular vector and/or the expressed transgene can be determined. It is important to determine whether administration of the gene therapy product results in exacerbation of the existing disease/disease symptoms or even the induction of a new disease. Caution should be extended, however, in the interpretation of the toxicities that are observed in animal models of disease for the reasons previously stated regarding the POC evaluation in these models. A preclinical study designed to evaluate both activity and toxicology endpoints in a biologically responsive animal model of disease has also been called a "hybrid" study.

Normal, pharmacologically relevant animal species are also often used in toxicology studies to evaluate the safety of gene therapy products. As discussed for the POC studies, many factors of the gene therapy product of interest are considered in the selection of the most appropriate animal species, including vector class, cell population (if applicable), and expressed transgene. The species used should ideally be sensitive not just to infection of the vector but also to the pathological consequences of infection that may be induced by the replication competent virus related to the vector. In some cases lower animal species may be transformed into "humanized" models that express the human target receptor(s) that may not be present or be only partially present in the animals. For example, the use of CD155 transgenic mice, which are engineered to express the CD155 receptor—which is the target for poliovirus [44] or Ifnar (KO) × CD46 transgenic mice that express human CD46 with human-like tissue specificity for the measles virus infection [45]—could be considered biologically applicable animal models for use in evaluation of the potential toxicity of the respective gene therapy product.

The utilization of relevant animal species/models that can mimic the human physiology as closely as possible is critical in order to provide an activity and a safety profile for the gene therapy product of interest. This way measurable boundaries can be obtained for an acceptable risk–benefit ratio that will allow for the administration of the gene therapy product into the desired human population (i.e., bench to bedside translation; Figure 9). There should be a

justifiable use of the animal resources employed to assess the pharmacology and safety of the gene therapy product.

## 10.3 Overall Toxicology Study Designs

The toxicology assessment of a gene therapy product should be comprehensive enough to identify, characterize, and quantify the potential local and systemic toxicities (Figure 10). Outcomes measured are acute/chronic toxicities, reversibility of toxicities, delayed toxicities, and any dose–response effects on these various outcomes. The approach used to ascertain a safety profile for a gene therapy product is novel; therefore the preclinical studies that are designed and conducted for these products are unique. For example, broad safety concerns for both the ex vivo and the in vivo administration of a gene therapy product originate from multiple factors, including (1) the potential for adverse reactions such as immunogenicity, to the ex vivo transduced cells, (2) vector and transgene toxicities, and (3) the potential risks of the delivery procedure. Direct injection of transduced cells or vector to vital organs such as the brain and heart generate questions regarding not only the potential toxicity of the gene therapy product but also the risks associated with the delivery procedure and the delivery device. While certain aspects of the overall scientific principles of internationally recognized guidelines such as those sponsored by the International Conference on Harmonisation (ICH), notably the S6 document, "Preclinical Safety Evaluation of Biotechnology-Derived Pharmaceuticals" [46], can be used, traditional toxicology programs are of minimal value in the determination of safety for gene therapy products. Many times the preclinical testing methodology is developed specifically for the gene therapy product of interest with the incorporation of novel technologies. Guidances for various gene therapy products released by different scientific and regulatory bodies can be utilized, as appropriate, for evaluation of a particular gene therapy product.

***Ex vivo Transduced Cells*** The toxicological assessment of the cellular component of ex vivo transduced cells may include endpoints that are similar to those evaluated for somatic cellular therapies. Depending on the cell population that is transduced, the following considerations should be incorporated for safety assessment: local environmental influence on cell survival, differentiation/phenotype expression, cell migration, host immune response, local and systemic reactions, and tumorigenicity. In addition the potential for induction of an immune response to the transduced cells can result in unwanted toxicities. For example, the expression of surface antigens on the transduced cells may be altered, potentially leading to an autoimmune response. Design of the preclinical studies should include assessment of the induction of autoimmunity or other immune responses. The considerations for the preclinical toxicology assessment of ex vivo transduced cells follow the same pathway as for gene therapy products that are administered directly into the recipient.

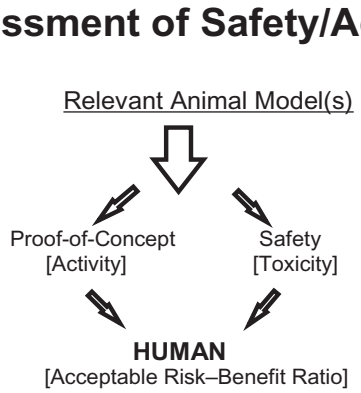


**Figure 9** Translational studies from preclinical to clinical trials.

**Toxicology Assessment for Gene Therapy Products**

Evaluate single/repeat exposure to the vector product (V)
- Toxicities related to the vector
- Biodistribution of the vector in blood and tissues

Evaluate the safety of gene expression (T)
- Persistence, level of expression in vivo
- Identify target tissues/functional endpoints
- Delayed toxicities/reversibility of toxicities

Evaluate V + T
- Characterize general toxicities
- Identify specific toxicities
- Characterize dose/exposure

**Figure 10** Overall paradigm for the assessment of the safety of a gene therapy product.

The development of clonal T cell proliferation in several children with X-SCID postadministration of autologous hematopoietic cells transduced ex vivo with retrovirus expressing the γc gene raised the awareness of the gene therapy scientific community regarding the potential safety concerns of an integrating vector from theoretical to actual [26,47]. The FDA has issued a draft guidance document that discusses the potential risks of delayed adverse events following exposure to gene therapy products as a consequence of persistent biological activity of the genetic material or other components of the products used to carry the genetic material, which could result in adverse effects on normal cell function. As listed in the guidance, factors that are likely to increase the risk of delayed adverse events in humans include persistence of the viral vector, integration of genetic material into the host genome, prolonged expression of the transgene, and altered expression of the host's genes [48].

### *In vivo Administration of a Gene Therapy Product*

*Vector Considerations* For the design of the most clinically applicable toxicology studies the route of administration, and the dosing schedule, knowledge of the planned clinical use of the gene therapy product regarding the dose level range, is important. Most gene therapy products are given either as a single administration or as a minimal number of repeat administrations over a relatively limited time interval. The route of administration and the dosing schedule for each gene therapy product in animals should mimic the intended clinical scenario as closely as possible. The presence and persistence of the vector in target and nontarget tissues, termed biodistribution, can help to provide insight toward extrapolation of the safety and rationale for the timing of multiple injections in humans. This evaluation, discussed in more detail later in this chapter, is analogous to the determination of a pharmacokinetic profile for a biological pharmaceutical in the instance of a recombinant human protein.

In parallel with an understanding of the injected vector's biodistribution profile should come an understanding of its potential toxicities. Adverse effects in animals can be dependent on the vector class and the route of administration. For example, the inflammatory reactions in response to adenovirus capsid proteins are well documented in the scientific literature. Intravenous administration of adenovirus in normal C57Bl/6 mice resulted in an acute inflammatory response in the liver [49], and direct instillation into the lungs of cotton rats resulted in a dose-related inflammatory reaction in the bronchioles [50]. In addition the toxicities observed in an animal model of disease given the same vector type can be potentially different. For example, the altered biodistribution of adenoviral vectors following intravascular administration in cirrhotic rats resulted in pulmonary hemorrhagic edema due to the presence of pulmonary intravascular macrophages resulting in notable vector biodistribution to the lungs [51].

It is well documented that humans generate neutralizing antibodies to select adenoviral and AAV vector serotypes (e.g., Ad5 and AAV2) following administration. These neutralising antibodies can prevent vector distribution to target tissue, accelerate clearance of the virus, and decrease transgene expression/persistence of expression [52]. In addition to reducing the kinetic and activity profile of the gene therapy product, induction of autoimmunity or other adverse immune effects may occur. The potential for stimulating an immune response (humoral and/or cellular) to the nucleic acid, the viral protein, or other parts of the vector construct is always a possibility and thus should be appropriately assessed.

The nature of the vector under investigation should be considered in the design of preclinical studies. For example, for preclinical evaluations of vectors that have a documented potential to integrate, such as gammaretrovirus and lentivirus, or possess the potential for latency and reactivation, such as the

Herpes virus, there could be used instead a tiered approach consisting of an assessment of the vector persistence followed by vector integration, latency, and/or reactivation, depending on the vector type.

*Transgene Considerations* The safety of the expressed transgene should be evaluated in appropriately designed preclinical studies as well. In most cases, the same animal species that is biologically sensitive to the vector of interest is also responsive to the translated protein. However, in some instances, the transgene product, such as human interleukin-12 or interferon, is biologically active only in humans and nonhuman primates, potentially necessitating the use of monkeys in toxicology studies. In other instances where published toxicology data exist for a transgene product that is active only in humans or nonhuman primates, another species analogue, such as murine IL-12, may suffice for the transgene product [53].

Similar to the recombinant human proteins, the potential for an immune response (cellular or humoral) to the expressed transgene exists. For example, delivery of transgenes that encode various endogenous enzymes, receptors, and structural proteins may elicit antibodies against both the transgene and the endogenous components expressed in normal cells and tissues, resulting in an autoimmune response. Similarly clinical delivery of transgenes that express fusion or chimera proteins can theoretically be immunogenic because of to their foreign nature. Thus the potential development of an immune response (humoral or cell mediated) for each gene therapy product should be investigated, as appropriate.

While persistent transgene expression may be a desired endpoint for some gene therapy products, toxicity can be an undesired outcome of transgene overexpression and/or a delayed abnormal immune response. Prolonged expression of transgenes such as growth factors, growth factor receptors, or immunomodulating agents may be associated with long-term risks due to from unregulated cell growth and malignant transformation, autoimmune reactions to self-antigens, or other unanticipated adverse effects. Long-term preclinical studies should be considered to evaluate these concerns.

The relationship between the dose level of the gene therapy product administered and the level and persistence of transgene expression in specific tissues and any toxicities that are observed should be determined. These data are important to ascertain, as they will affect the clinical monitoring and planned dose escalation in the clinical trials.

## 10.4 General Considerations

The design of the preclinical toxicology studies incorporates the basic tenets used in toxicology study designs for biological and small-molecule pharmaceuticals. The design should allow for the inclusion of appropriate control groups (may be more than one group, e.g., untreated controls, sham controls,

null vector controls, vehicle controls), several dose groups, adequate numbers of animals per gender in each group for each sacrifice time point, and a clinically relevant route of administration. The controls could consist of the vehicle formulation, a sham control if the administration procedure is unique or an unknown delivery device is used, or an empty vector if a novel vector system is used. Several groups of animals that will be given the gene therapy product at dose levels that bracket and exceed the intended dose level range in humans, if possible, should be included. The multiples of the human dose required to determine an adequate safety margin can vary with the vector class, the "equivalency" of the injection route used in animals to the clinical route (i.e., intratumoral in human vs. subcutaneous or intramuscular in animals), and the relevance of the animal species/model to humans. The duration of each study conducted is dependent on the product, the dosing scheme, and the subject population of interest. The biodistribution and persistence profile of the vector and the expression profile of the transgene often guide the duration of the toxicology studies and the time intervals selected for sacrifice of animals.

The inclusion of safety endpoints such as clinical signs, physical exams, body weights, food consumption/appetite, serum chemistry, hematology, coagulation, urinalysis, organ weights, gross pathology, histopathology, and anti-vector/anti-transgene antibody measurements are considered routine for preclinical toxicology studies for gene therapy products. Other endpoints specific for the assessment of potential adverse effects due to the class of vector (e.g., neuropathology for Herpes virus), due to the transgene product (e.g., undesired cell proliferation due to a growth factor), or due to the proposed clinical population (e.g., diabetes or cancer) can include ECGs and other cardiac parameters, ophthalmology exams, immunotoxicity assessment, behavioral assessment, neurotoxicity evaluation, specialized histopathology (i.e., immunohistochemistry, special stains), as well as other parameters.

The traditional carcinogenicity study as described in the ICH S1B guideline and the standard battery of genotoxicity studies as described in the ICH S2 guidelines are generally not applicable to gene therapy products [54–56]. As previously discussed, the concern for insertional mutagenesis of the vector is an important potential safety issue that should be addressed, and the tumorigenic potential of an expressed transgene of concern should be evaluated in adequately designed preclinical studies.

## 10.5 Good Laboratory Practice (GLP)

In the United States preclinical toxicology studies performed in support of the use of gene therapy products in clinical trials should be conducted in compliance with good laboratory practice (GLP) as per the US regulation, 21 CFR part 58. Preclinical in vivo pharmacology studies (i.e., proof of concept) and in vitro pharmacology studies are not expected to be conducted under GLP. In the early stages of the development program for a gene therapy product the preclinical studies are not always conducted in full compliance with GLP. The

FDA has agreed that these preclinical studies can be conducted at a facility that is capable of complying with a prospectively designed study protocol. Non-GLP studies are acceptable as long as they are performed in accordance with a prospectively designed protocol and the data are of sufficient quality and integrity to support the proposed clinical trial. The study report that is submitted to FDA should identify any areas that deviate from the prospectively designed protocol and the potential impact of these deviations on study integrity. Many times initial preclinical toxicology data for these products are collected in proof-of-concept studies conducted using an animal model of disease, which may require unique animal care issues; thus complete compliance with GLP may be difficult. The expectation is that pivotal toxicology studies submitted to support a marketing application will be conducted under GLP.

## 10.6 Non-GLP Studies

In addition to the GLP toxicology studies and well-controlled non-GLP studies conducted for various gene therapy products, published information in peer-reviewed scientific journals, as well as from public presentations and discussions, can be consulted to support the safety of the gene therapy product of interest. For example, all information regarding a specific gene therapy product that is submitted to the NIH Recombinant DNA Advisory Committee (RAC), located in the Office of Biotechnology Activities (OBA), is considered to be in the public domain. Thus the toxicology information for one vector serotype in a class (e.g., adeno-associated viruses) may help in the safety evaluation of another serotype in the same vector class. If extensive preclinical or clinical safety data for the same or similar vector type exist, toxicity assessment of the gene therapy product of interest may be less extensive, depending on such factors as the dose levels, dosing regimen, route of administration, and expression cassette that were previously used. Depending on the extent of the change to a particular vector that has been administered to humans, in vitro or small in vivo bridging studies may suffice.

The data generated from the toxicology studies should allow a determination of a no-observable-adverse-effect-level, or NOAEL, and should characterize the relationship of the documented toxicities to the dose of the gene therapy product that is given. The extrapolation of this NOAEL to the clinical arena is accomplished in several ways, including body weight, body surface area, and organ weight/volume. An understanding of the association of the dose levels employed in the toxicology studies to the resulting toxicities, as well as the effect of administration route and dosing schedule on any toxicities observed in the animals, is the desired goal.

## 10.7 Biodistribution Considerations

While the determination of the overall safety of an experimental agent is a desired common objective for all pharmaceutical agents, the approach used

to achieve this goal for a gene therapy product is distinct from the approach used for small-molecule products, as well as for the traditional biologic therapies (e.g., monoclonal antibodies and recombinant cytokines). A specific difference is reflected in the determination of the pharmacokinetic profile of a gene therapy product. While standard absorption, disposition, metabolism, and excretion (ADME) studies are conducted for small-molecule products and pharmacokinetic assessment is applied for biologics, the localization or biodistribution of the vector is assessed in animals. Following administration via (ideally) the planned clinical route, the determination of the dissemination of the vector to the intended therapeutic site(s), as well as to nontarget tissues, provides an important part of the overall safety profile of each gene therapy product (Figure 11). Such data should guide the toxicology further study designs, with respect to dose levels, dosing schedule, and route of administration. The biodistribution study should include appropriate control animals and groups of animals given dose levels of the gene therapy product that overlap the dose levels used in the toxicology studies. Multiple sacrifice intervals are also included in order to capture a comprehensive kinetic profile to show peak vector levels and to characterize the persistence of vector signal in certain tissues and biological fluids (i.e., blood, semen) over time.

Biodistribution analysis is conducted at the molecular level. The current gold standard is a quantitative polymerase chain reaction (Q-PCR) assay that detects the number of vector copies per microgram of genomic DNA. Biological fluids and tissue samples are carefully harvested (to avoid cross-contamination) from control and gene therapy product-injected animals at

**Vector Biodistribution to Intended Therapeutic Site/Unintended Site(S)**

Presence of vector sequence via Q-PCR analysis
- Dissemination of vector to nontarget tissues
- Dissemination of vector to the germline
- Vector persistence and clearance profile

Study conducted prior to phase 1 trials when:
- New class of vector, with no/little experience (e.g., capsid modification)
- Change in formulation (e.g., lipid carrier)
- Change in ROA with an established vector
- Change in dosing schedule
- New transgene; potential toxicities unknown
- Potential for the transgene to induce toxicity if aberrantly expressed in nontarget tissue

**Figure 11** Considerations for the assessment of vector biodistribution.

several time points postadministration. Usually a default panel of tissues is collected and analyzed, including blood, injection site(s), gonads, brain, liver, kidneys, lung, heart, and spleen. Depending on the vector type and the transgene expressed, as well as the route of administration, other tissues should be analyzed (i.e., draining lymph nodes and contralateral sites of injection, bone marrow, and eyes) [48].

Another very important outcome of this analysis is the determination of vector presence in the germline (Figure 5). For every biodistribution study conducted, samples of gonadal tissue (testes or ovaries) from each animal should be harvested and analyzed. If a significant, persistent positive signal is detected using the state-of-the-art methodology, then a more extensive testing paradigm may be necessary. Further studies to determine the cell type that is positive for vector presence (i.e., germ cells, stromal cells, leukocytes) should be the next step. If the outcome continues to implicate inadvertent germline transfer, then the conduct of reproductive and developmental toxicology studies may be needed in order to assess the potential for vertical transmission of the gene sequences. The ICH S5(R2) guideline should be consulted for the overall design of these studies [57]. The risk of inadvertent germline transmission and the subsequent potential for integration are currently the major focus of the ICH GTDG. This group has released a consideration paper on this issue entitled "General Principles to Address the Risk of Inadvertent Germline Integration of Gene Therapy Vectors" (for more information, refer to the ICH Web site for details, http://www.ich.org/cache/html/1386-272-1.htm).

The presence of a vector sequence in nontarget tissues at significant levels should suggest the need for further analysis of the tissue to determine the levels of transgene expressed. These data, coupled with results of other safety endpoints, such as clinical pathology and histopathology evaluation, will determine whether vector presence or gene expression is correlated with any detrimental effects on the tissue.

Instances in which the biodistribution studies are conducted prior to use in humans can include the following: when the gene therapy product belongs to a new vector class, when there is little or no published data on the vector or the transgene of interest, if there is a change in the vector backbone for an established vector, if there is a change in formulation, if there is a change in the route of vector administration, or if there is a known or suspected potential for the transgene to induce toxicity that may be aberrantly expressed in nontarget tissues. For example, Cearly and Wolfe showed that injection of AAV 7, 8, 9, and rh10 vectors containing the same genome into the same area of rodent brains resulted in distinctly different transduction patterns due to the different capsid proteins. These differences translate into potential effects on (1) product activity if each vector displays unique capabilities at transducing different substructures of the brain and on (2) product toxicity if overexpression occurs in unwanted areas of the CNS to which the vector has disseminated [58].

If biodistribution studies are not performed prior to initiation of clinical trials, these studies should be incorporated in the development program of the gene therapy product so that data are generated prior to beginning any large pivotal studies in humans.

### 10.8 Later Stage Product Development

As the gene therapy product development program continues into the later phase clinical trials and toward the ultimate goal of a marketed product (Figure 1), additional consideration has to be given to the components of a product label (i.e., the package insert) that require data from preclinical studies, on mutagenicity, carcinogenicity, and reproductive toxicity. The intent of the label is to convey information on the safety and effectiveness of the approved product to the patient population receiving the respective product and to communicate any associated or perceived risk to both the physician and to the patient. Sponsors of gene therapy products should consult with the applicable regulatory agencies early in the product development program to ensure that the timing and design of the necessary preclinical studies are in place in accordance with the regulators' expectations.

## 11 CONCLUSION

Included on the Web site of the American Society of Gene Therapy (ASGT) is a section entitled "Brief History of Gene Therapy." This synopsis is a snapshot of the notable advances and setbacks in this field since 1966. Although these biologics are complex, rapid advances have been made with the discovery of new and novel viruses, as well as with the re-engineering of existing viruses. Without preclinical studies that are appropriately designed to provide insight on dose/activity and dose/toxicity relationships for the gene therapy products that are to be used in predetermined disease settings, early phase clinical trial designs are limited. Careful and thorough evaluation of the biological plausibility and safety of a gene therapy product is an essential first step on the critical pathway to a successful clinical product development program [59,60]. Unique and comprehensive preclinical testing approaches using biologically relevant animal species/models are critical elements in risk assessment of such novel products. The preclinical study designs will need to consider the vector system, the expressed transgene, the transduced cell type (as applicable), and the clinical indication to adequately assess the potential risks of each gene therapy product. Since risk assessment leads to risk communication, the preclinical data provide an effective means of risk communication to the physicians that manage the clinical trials, as well as to the individuals consenting to receive these experimental therapies. Ultimately any such risk communication will transfer to the marketed label for the approved gene therapy product.

## ACKNOWLEDGMENTS

The authors wish to acknowledge the helpful contributions of many colleagues in the Office of Cellular, Tissue, and Gene Therapies. We are especially grateful to Drs. Richard McFarland, Stephanie Simek, and Daniel Takefman for their thorough and helpful review of the manuscript during its preparation.

## REFERENCES

1. *FDA Guidance for Industry: Guidance for Human Somatic Cell Therapy and Gene Therapy*, March 1998. http://www.fda.gov/cber/gdlns/somgene.pdf
2. Bell J, Lichty B, Stojdl D. Getting oncolytic virus therapies off the ground. *Cancer Cell* 2003;4:7–11.
3. GeMCRIS, Gene Transfer Protocol Reports. www.gemcris.od.nih.gov
4. Rosenberg SA, Anderson WF, Blaese M, Hwu P, Yannelli JR, Yang JC, Topalian SL, Schwatzentruber DJ, Weber JS, Ettinghausen SE, Parkinson DN, White DE. The development of gene therapy for the treatment of cancer. *Ann Surg* 1993;218:455–63.
5. Blaese RM, Culver KW, Miller AD, Carter CS, Fleisher T, Clerici, M, Shearer G, Chang L, Chiang Y, Tolstoshev. P, Greenblatt JJ, Rosenberg SA, Klein H, Berger M, Mullen CA, Ramsey WJ, Muul L, Morgan RA, Anderson WF. T-lymphocyte-directed gene therapy for ADA-SCID: initial trial results after 4 years. *Science* 1995; 270:475–80.
6. Nabel GJ, Nabel EG, Yang ZY, Fox BA, Plautz GE, Gao X, Huang L, Shu S, Gordon D, Chang AE. Direct gene transfer with DNA-liposome complexes in melanoma: expression, biologic activity, and lack of toxicity in humans. *Proc Na Acad Sci USA* 1993;23:11307–11.
7. Crystal RG, Jaffe A, Brody S, Mastrangeli A, McElvaney NG, Rosenfeld M, Chu CS, Danel C, Hay J, Eissa T. A phase 1 study in cystic fibrosis patients of the safety, toxicity, and biological efficacy of a single administration of a replication deficient recombinant adenovirus carrying the cDNA of the normal cystic fibrosis transmembrane conductance regulator gene in the lung. *Hum Gene Ther* 1995;6:643–66.
8. Flotte T, Carter B, Conrad C, Guggino W, Reynolds T, Rosenstein B, Taylor G, Walden S, Wetzel R. A phase I study of an adeno-associated virus-CFTR gene vector in adult CF patients with mild lung disease. *Hum Gene Ther* 1996;7: 1145–59.
9. Kay MA, Manno CS, Ragni MV, Larson PJ, Couto LB, McClelland A, Glader B, Chew AJ, Tai SJ, Herzog RW, Arruda V, Johnson F, Scallan C, Skarsgard E, Flake AW, High KA. Evidence for gene transfer and expression of factor IX in hemophilia patients treated with an AAV vector. *Nat Genets* 2000;24:257–61.
10. Khuri FR, Nemunaitis J, Ganly I, Arseneau J, Tannock IF, Romel L, Gore M, Ironside J, MacDougall RH, Heise C, Randlev B, Gillenwater AM, Bruso P, Kaye SB, Hong WK, Kirn DH. A controlled trial of intratumoral ONYX-015, a selectively-replicating adenovirus, in combination with cisplatin and 5-fluorouracil in patients with recurrent head and neck cancer. *Nat Med* 2000;8:879–85.

11. Peng Z. Current status of Gendicine in China: recombinant human Ad-p53 agent for treatment of cancers. *Hum Gene Ther* 2005;16:1016–27.
12. Guo J, Xin H. Chinese gene therapy: splicing out the west? *Science* 2006;314: 1232–5.
13. Cavazzana-Calvo M, Hacein-Bey S, de Saint Basile G, Gross F, Yvon E, Nusbaum P, Selz F, Hue C, Certain S, Casanova JL, Bousso P, Deist FL, Fischer, A. Gene therapy of severe combined immunodeficiency (SCID)-X1 disease. *Science* 2000;288:669–72.
14. Thrasher AJ, Hacein-Bey-Abina S, Gaspar HB, Blanche S, Davies EG, Parsley K, Gilmour K, King D, Howe S, Sinclair J, Hue C, Carlier F, von KalLe C, de Saint BasiLe G, Le Deist F, Fischer A, Cavazzana-Calvo M. Failure of SCID-X1 gene therapy in older patients. *Blood* 2005;105:4255–7.
15. National Institutes of Health. Amended Charter: Recombinant DNA Advisory Committee. Updated April 13, 2005. http://www4.od.nih.gov/oba/rac/RACCharter2005.pdf
16. The Public Health Service Act. Title 42. Approved on July 1, 1944. http://www.fda.gov/opacom/laws/phsvcact/phsvcact.htm
17. Federal Food, Drug, and Cosmetic Act, as amended through December 31, 2004. http://www.fda.gov/opacom/laws/fdcact/fdctoc.htm
18. Kessler DA, Siegel JP, Noguchi PD, Zoon KC, Feiden KL, Woodcock J. Regulation of somatic cell-therapy and gene therapy by the Food and Drug Administration. *N Engl J Med*, 1993;329:1169–73.
19. Glover DJ, Lipps HJ, Jans DA. Towards safe, non-viral therapeutic gene expression in humans. *Nat Rev Genet* 2005;6:299–310.
20. Wolff JA, Budker V. The mechanism of naked DNA uptake and expression. *Adv Genet* 2005;54:3–20.
21. Yew NS, Scheule RK. Toxicity of cationic lipid-DNA complexes. *Adv Gene* 2005; 53:189–214.
22. Bangari DS, Mittal SK. Current strategies and future directions for eluding adenoviral vector immunity. *Curr Gene Ther* 2006;6:215–26.
23. Tenenbaum L, Lehtonen E, Monahan PE. Evaluation of risks related to the use of adeno-associated virus-based vectors. *Curr Gene Ther* 2003;3:545–65.
24. Donsante A, Vogler C, Muzyczka N, Crawford JM, Barker J, Flotte T, Campbell-Thompson M, Daly T, Sands MS. Observed incidence of tumorigenesis in long-term rodent studies of rAAV vectors. *Gene Ther* 2001;8:1343–6.
25. Donahue RE, Kessler SW, Bodine D, McDonagh K, Dunbar C, Goodman S, Agricola B, Byrne E, Raffeld M, Moen R, Bacher J, Zsebo KM, Nienhuis AW. Helper virus induced T cell lymphoma in nonhuman primates after retroviral mediated gene transfer. *J Exp Med* 1992;176:1125–35.
26. Hacein-Bey-Abina S, Von Kalle C, Schmidt M, McCormack MP, Wulffraat N, Leboulch P, Lim A, Osborne CS, Pawliuk R, Morillon E, Sorensen R, Forster A, Fraser P, Cohen JI, de Saint Basile D, Alexander I, Wintergerst U, Frebourg T, Aurias A, Stoppa-Lyonnet D, Romana S, Radford-Weiss I, Gross F, Valensi F, Delabesse E, Macintyre E, Sigaux F, Soulier J, Leiva LE, Wissler M, Prinz C, Rabbitts TH, Le Deist F, Fischer A, Cavazzana-Calvo, M. LMO2-associated clonal

T-cell proliferation in two patients after gene therapy for SCID-XI. *Science* 2003;302:415–9.

27. Anson D. The use of retroviral vectors for gene therapy—What are the risks? A review of retroviral pathogenesis and its relevance to retroviral vector-mediated gene delivery. *Genet Vaccines Ther* 2004;2:9–22.
28. Loewen N, Poeschla EM. Lentiviral vectors. *Adv Biochem Eng/Biotech* 2005;99: 169–91.
29. Biederer C, Ries S, Brandts CH, McCormick F. Replication-selective viruses for cancer therapy. *J Mol Med* 2002;80:163–75.
30. Stoeckel J, Hay JG. Drug evaluation: Reolysin-wild type reovirus as a cancer therapeutic. *Curr Opin Mol Therapeu* 2006;3:249–60.
31. Hasegawa K, Pham L, O'Connor MK, Federspiel MJ, Russell SJ, Peng KW. Dual therapy of ovarian cancer using measles viruses expressing carcinoembryonic antigen and sodium iodide symporter. *Clin Cancer Res* 2006;6:1868–75.
32. Pecora AL, Rizvi N, Cohen GI, Meropol NJ, Sterman D, Marshall JL, Goldberg S, Gross P, O'Neil JD, Groene WS, Roberts MS, Rabin H, Bamat MK, Lorence RM. Phase I trial of intravenous administration of PV701, an oncolytic virus, in patients with advanced solid cancers. *J Clin Oncol* 2002;20:2251–66.
33. Schijns VE, Tangrås A. Vaccine adjuvant technology: from theoretical mechanisms to practical approaches. *Dev Biol* 2005;121:127–34.
34. Brennan FR, Dougan G. Non-clinical safety evaluation of novel vaccines and adjuvants: new products, new strategies. *Vaccine* 2005;23:3210–22.
35. EMEA Guideline on Adjuvants In Vaccines for Human Use, January 2005. www.emea.eu.int/pdfs/human/vwp/13471604en.pdf
36. Mazzolini G, Narvaiza I, Perez-Diez A, Rodriguez-Calvillo M, Qian C, Sangro B, Ruiz J, Prieto J, Melero I. Genetic heterogeneity in the toxicity fo systemic adenoviral gene transfer of interleukin-12. *Gene Ther* 2001;8:259–67.
37. Prince GA, Porter DD, Jenson AB, Horswood RL, Chanock RM, Ginsberg HS. Pathogenesis of adenovirus type 5 pneumonia in cotton rats (Sigmodon hispidus). *J Virol* 1993;67:101–11.
38. Thomas MA, Spencer JF, La Regina MC, Dhar D, Tollefson AE, Toth K, Wold WS. Syrian hamster as a permissive immunocompetent animal model for the study of oncolytic adenovirus vectors. *Cancer Res* 2006;3:1270–6.
39. Manilla P, Rebello T, Afable C, Lu X, Slepushkin V, Humeau LM, Schonely K, Ni Y, Binder GK, Levine BL, MacGregor RR, June CH, Dropulic, B. Regulatory considerations for novel gene therapy products: a review of the process leading to the first clinical lentiviral vector. *Hum Gene Ther* 2005;1:17–25.
40. Williams DA, Croop J, Kelly P. Gene therapy in the treatment of Fanconi anemia, a progressive bone marrow failure syndrome. *Curr Opin Mol Therapeut* 2005;5: 461–6.
41. December 3–4, 2003, Recombinant DNAAdvisory Committee (RAC) meeting. http://www4.od.nih.gov/oba/rac/meeting.htm
42. Disis ML, Scholler N, Dahlin A, Pullman J, Knutson KL, Hellstrom KE, Hellstrom, I. Plasmid-based vaccines encoding rat neu and immune stimulatory molecules can elicit rat neu-specific immunity. *Mol Cancer Ther* 2003;10:995–1002.

43. Cavagnaro JA. Preclinical safety evaluation of biotechnology-derived pharmaceuticals. *Nat Rev* 2002;1:469–75.

44. Dragunsky EM, Ivanov AP, Abe S, Potapova SG, Enterline JC, Hashizume S, Chumakov KM. Further development of a new transgenic mouse test for the evaluation of the immunogenicity and protective properties of inactivated poliovirus vaccine. *J Infect Dis* 2006;194:804–60.

45. Peng KW, Frenzke M, Myers R, Soeffker D, Harvey M, Greiner S, Galanis E, Cattaneo R, Federspiel MJ, Russell SJ. Biodistribution of oncolytic measles virus after intraperitoneal administration into Ifnar(tm)-CD46Ge transgenic mice. *Hum Gene Ther* 2003;16:1565–77.

46. International Conference on Harmonisation (July 1997). *S6: Preclinical Safety Evaluation of Biotechnology-Derived Pharmaceuticals*. http://www.ich.org/cache/compo/502-272-1.htm

47. Couzin J, Kaiser J. As Gelsinger case ends, gene therapy suffers another blow. *Science* 2005;307:1028.

48. [Draft] *FDA Guidance for Industry: Gene Therapy Clinical Trials—Observing Participants for Delayed Adverse Events*, August 2005. http://www.fda.gov/cber/gdlns/gtclin.pdf

49. Zhang Y, Chirmule N, Gao GP, Qian R, Croyle M, Joshi B, Tazelaar J, Wilson JM. Acute cytokine response to systemic adenoviral vectors is mediated by dendritic cells and macrophages. *Mol Ther* 2001;5:697–707.

50. Bauer SR, Pilaro AM, Weiss KD, Testing of adenoviral vector gene transfer products: FDA expectations. In *Adenoviral Vectors for Gene Therapy*. Curiel DT, Douglas JT. eds. San Diego: Academic Press, 2002; 615–53.

51. Smith JS, Tian J, Lozier JN, Byrnes AP. Severe pulmonary pathology after intravenous administration of vectors in cirrhotic rats. *Mol Ther* 2004;6:932–41.

52. Jiang H, Couto LB, Patarroyo-White S, Liu T, Nagy D, Vargas JA, Zhou S, Scallan CD, Sommer J, Vijay S, Warren D, Mingozzi F, High KA, Pierce GF. Effects of transient immunosuppression on adeno associated virus-mediated liver-directed gene transfer in rhesus macaques and implications for human gene therapy. *Blood*, prepublished online, July 25, 2006; DOI 10.1182.

53. Sung MW, Chen SH, Thung SN, Zhang DY, Huang TG, Mandeli JP, Woo SL. Intratumoral delivery of adenovirus-mediated interleukin-12 gene in mice with metastatic cancer in the liver. *Hum Gene Ther* 2002;6:731–43.

54. International Conference on Harmonisation (July 1997). *S1B: Testing for Carcinogenicity of Pharmaceuticals*. http://www.ich.org/cache/compo/276-254-1.htm

55. International Conference on Harmonisation (July 1995). *S2A: Guidance on Specific Aspects of Regulatory Genotoxicity Tests for Pharmaceuticals*. http://www.ich.org/cache/compo/276-254-1.htm

56. International Conference on Harmonisation (July 1997). *S2B: Genotoxicity: A Standard Battery for Genotoxicity Testing of Pharmaceuticals*. http://www.ich.org/cache/compo/276-254-1.htm

57. International Conference on Harmonisation (Nov, 1995; amended Nov, 2000). *S5(R2): Detection of Toxicity to Reproduction for Medicinal Products and Toxicity to Male Fertility*. http://www.ich.org/cache/compo/276-254-1.htm.

58. Cearley CN, Wolfe JH. Transduction characteristics of adeno-associated virus vectors expressing cap serotypes 7, 8, 9, and rh10 in the mouse brain. *Mol Ther* 2006;13:528–37.
59. Leiden JM. Human gene therapy: the good, the bad, and the ugly. *Circul Res* 2000;86:923–5.
60. Borellini F, Ostrove J. The transfer of technology from the laboratory to the clinic: In process controls and final product testing. In Meager A, ed. *Gene Therapy Technologies, Applications and Regulations*. New York: Wiley, 1999;359–73.

# Considerations in Design of Preclinical Safety Evaluation Programs to Support Human Cell-Based Therapies

JOY A. CAVAGNARO, PhD, DABT, RAC

**Contents**

1 Introduction
2 Regulation of Cell-Based Therapies in the United States and Europe
3 Types of Cell-Based Therapies
  3.1 Ex vivo Activated Mononuclear Leukocytes
  3.2 Living Autologous Cells Manipulated Ex vivo Intended for Structural Repair or Reconstruction (MAS Cells)
  3.3 Adult Stem Cells (Somatic Stem Cells)
  3.4 Hemapotopoietic Stem Cells
  3.5 Endothelial Progenitor Cells (EPCs)
  3.6 Mesenchymal Stem Cells (MSCs)
  3.7 Human Embryonic Stem Cells (hESC)
  3.8 Xenogeneic Cells
4 Case-by-Case Approach Based on Product Attributes and Indication
5 Principles of Preclinical Safety Evaluation for Stem Cell Based Therapies
6 Pharmacology
7 Safety Pharmacology
8 Exposure Assessment
9 General Toxicity
10 Reproductive and Developmental Toxicity
11 Genotoxicity
12 Tumorigenicity
13 Local Tolerance
14 Expectations for pre-preIND Filing in the United States

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

15 The Transition into Clinical Trials
16 Conclusions
Appendix Pharmacology/Toxicology Considerations Regarding Somatic Cell Products
References

## 1 INTRODUCTION

The sixteenth-century German-Swiss physician, Philippus Aureolus Theophrastus Bombastus von Hohenheim ("aka Paracelsus") may be credited with the earliest references to cell therapy. Although Paracelsus is probably best recognized as the "father of toxicology" for his pioneering work in systematizing the study of effects of chemicals and drugs, he spoke of healing of the heart with heart, lung with lung, spleen with spleen—indeed as "like heals like" in his 1536 book *Der grossen Wundartzney* [1]. Four centuries later, the Swiss physician. Paul Niehans would, by chance, become known as "the father of cell therapy" after treating a patient, whose parathyroid was accidentally removed by a fellow physician, with a homogenate of the parathyroid gland of an ox. Niehans claimed that he could cure illnesses through injection of live cells extracted from healthy organs. He believed that adding new tissue stimulated rejuvenation and recovery (see Table 1). In the 1960s the use of cells and tissue extracts from young or fetal animals, "fresh cell therapy," was still popular in Germany, and this practice lasted until the late 1990s when it was finally banned following increases in severe immune-mediated allergic diseases resulting from repeated injections of heterologous antigenic materials.

Today, with the exception of bone marrow for hematopoietic reconstitution, therapeutic cellular transplantation is an emerging technology. In recent years novel approaches in the potential restoration of function through cellular transplantation have included the use of fetal human or xenogeneic neural tissue for Parkinson's disease, ectopically implanted pancreatic islets for diabetes, Schwann cells and olfactory ensheathing glia for spinal cord injury, encapsulated chromaffin cells for pain, and various types of stem cells for the treatment of diabetes, cardiac disease, and central nervous system injuries or disease [2]. There have also been trials of encapsulated cells to provide enzymes that either remove "toxic" products or provide activation of prodrugs to therapeutics, usually anticancer derivatives.

Today the field of "regenerative medicine" refers to the broad range of disciplines adopted by groups working toward a common goal of replacing or regenerating damaged or diseased cell populations or tissues. There are approximately 270 cell types in the human body [3]. Understanding the biology as well therapeutic potentials of the many cell types is an important reason for pursing research on embryonic stem cells. Theoretically embryonic stem cells represent a source of cells capable of development into normal

**TABLE 1 Selected milestones in the field of cell-based therapies**

| | |
|---|---|
| 1931 | First injection of living cells (parathyroid cells from an ox) |
| 1937 | Implantation of cerebral cells |
| 1949 | First injection of lyophilized cells |
| 1956 | Successful engraftment of bone marrow cells in mice following irradiation |
| 1963 | Self-renewing cells discovered in mouse bone marrow |
| 1964 | Single cell in teratocarcinoma isolated and remained undifferentiated in culture |
| 1968 | First successful bone marrow transplantation between two siblings |
| 1978 | Hematopoietic stem cells discovered in human cord blood |
| 1981 | Murine embryonic stem cells derived from mouse inner cell mass |
| 1983 | Fetal cells reported to repair spinal cord injury |
| 1988 | Isolation of murine hematopoietic stem cells |
| 1988 | Fetal pancreatic cells used in diabetic patients |
| 1991 | Fetus-to-fetus cell therapy |
| 1992 | Neural stem cells cultured in vitro as neurospheres |
| 1995 | Law to prohibit federal funds for research where human embryos are created or destroyed |
| 1997 | Cloning of Dolly the sheep by somatic cell nuclear transfer (SCNT) |
| 1997 | Isolation of putative progenitor endothelial cells |
| 1997 | Fresh cell therapy banned in Germany |
| 1998 | First human embryonic stem cell line derived (hESC) |
| 1998 | First established culture of human embryonic germ cells (hEGC) |
| 1998 | Bone marrow derived dendritic cells first described |
| 1999 | Cardiac tissue formation after bone marrow transplantation |
| 2000 | Development of neurons from bone marrow |
| 2001 | Cloning of first early (4 to 6 cell) human embryos for purpose of generating hESCs |
| 2001 | hESCs differentiated into early-stage neuronal cells |
| 2001 | Muscle tissue formation after bone marrow transplantation |
| 2001 | Stem cells from adipose tissue produce cartilage, bone, muscle |
| 2001 | Restriction of US federal funding for hESC research (use of approved cells only) |
| 2003 | Children's primary teeth as new source of stem cells |
| 2005 | Cord blood derived embryonic-like stem cell derived from umbilical cord blood |
| 2006 | hESC lines derived from single blastomeres |
| 2007 | Discovery of new type of stem cell in amniotic fluid |
| 2007 | hESC-based progenitor cells produce multiple nerve growth factors |

differentiated cells that, when properly engineered can replace diseased tissues and restore normal function. In addition to selection of the most appropriate cell type, several different modes of transfer may be used to deliver cells to optimize therapeutic benefit. Cells may be directly infused, surgically implanted, or contained within an encapsulating membrane. Cells can be delivered through catheters, implanted in devices, or delivered through extracorporeal devices. In some cases genes can also be introduced into the cells to improve

the efficacy and/or safety of the various types of cell-based therapies. The combination of cell and gene therapy using both nonviral and viral gene transfer vectors is beyond the scope of this chapter (see Chapter Preclinical Safety Evaluation of Biopharmaceuticals).

Over 500 companies are currently involved in cell therapy technology, and of these 121 are involved in stem cell therapies [4]. Clinical proof of principle has been demonstrated across several diseases for transplanted cells including multipotent fully differentiated cells in culture, cells for in vivo terminal differentiation, mixtures of multipotent and differentiated cells, and pluripotent cells for restoring function. No single source of cells is best or even suitable for all disease indications. In addition more than one source of cells may be used for the same indication; for example, adult bone marrow derived cells, endothelial progenitor cells, resident cardiac cells, and embryonic stem cells are all being studied for cardiac repair [3,6–8]. In diabetes sources of new β islet cells have included fetal pancreas, ex vivo expansion of endogenous progenitors β cell from adult pancreas, xenotransplants, umbilical cord blood, adult bone marrow, and embryonic stem cell lines [9,10] (see Table 2).

Most applications for cell-based therapies are still in the experimental research and clinical trial stages. As novel cell-based therapies move from discovery research to the clinic, developers and regulators will face new and continually evolving issues and uncertainties involving long-term efficacy and safety. The success of cell-based therapy will depend on optimal selection, manufacture, and survival of the implanted cells. However, because of the heterogeneity in cell types and scope of indications, the preclinical development, strategies, clinical development, and evaluation requirements are expected

**TABLE 2 No single source of cells is suitable for all indications**

| Intervention | Indication | Types of Cells |
|---|---|---|
| Hematopoietic, immune replacement | Cancer<br>Autoimmune disorders<br>Immunodeficiencies | Lymphocyte-based therapies, autologous tumor cell vaccines, blood, bone marrow, UCB stem cells, HSCs, hMSCs, etc. |
| Metabolic replacement. support | Ophthalmic disorders<br>Diabetes<br>Lysosomal storage diseases<br>Renal/liver failure | Human retinal stem cells, tissue-derived stem cells, islet cells, bioartificial kidney, bioartificial liver, hepatocytes, etc. |
| Tissue repair, regeneration | Cardiovascular disorders<br>Neurological/ neurodegenerative Disorders<br>Bone and joint disorders<br>Wound healing[a] | hMSCs, cardiomyocytes, tissue-derived stem cells including olfactory cells neural cells, osteoblasts, chondrocytes, keratinocytes, dermal fibroblasts, hESCs, etc. |

[a]The FDA has approved several cell therapy products as devices for skin replacements; for the treatment of acute (e.g., burns) and chronic (e.g., diabetic ulcers) inflammatory skin disorders.

to be optimized according to product attributes and clinical indication. As is true for other novel biopharmaceuticals, the diversity of potential products coupled with uncertainties about unique product specific attributes creates a challenge for toxicologists responsible for predicting and characterizing their safety and estimating potential risks. However, knowledge gained from traditional biopharmaceuticals developed and approved throughout the 1980s and 1990s can be applied to novel cellular therapies. The challenges in moving cell therapy to the clinic is to understand what aspects of the results obtained with animal models can be easily translated to therapy and where caution must be exercised in extrapolating animal studies to ultimate human treatment [5].

## 2 REGULATION OF CELL-BASED THERAPIES IN THE UNITED STATES AND EUROPE

The FDA regulates cellular products as biological products. The Center for Biologics Evaluation and Research (CBER) is designated as the agency with primary jurisdiction for the pre-market review and regulation. Cellular products intended for use as somatic cell therapy are subject to regulation pursuant to the PHS Act (42 USC 251) and also fall within the definition of drugs (21 USC 312 (g)). Somatic cell therapy products are defined as autologous (i.e., self), allogeneic (i.e., intraspecies), or xenogeneic (i.e., interspecies) cells that have been propagated, expanded, selected, pharmacologically treated, or otherwise altered in biological characteristics ex vivo to be administered to humans and applicable to the prevention, treatment, cure, diagnosis, or mitigation of disease or injuries [11]. Specific examples of cellular products regulated by FDA as biological cellular products include (1) autologous or allogeneic lymphocytes activated and expanded ex vivo including lymphokine-activated killer cells (LAK) and tumor infiltrating lymphocytes (TIL cells); (2) encapsulated autologous, allogeneic, or xenogeneic cells or cultured cell lines intended to secrete a bioactive factor or factors (e.g., insulin, growth hormone, neurotransmitters); (3) autologous or allogeneic somatic cells including genetically modified cells (e.g., hepatocytes, myocytes, fibroblasts, bone marrow- or blood-derived hematopoietic stem cells, lymphocytes that have been genetically modified; (4) cultured cell lines; and (5) autologous or allogeneic bone marrow transplants using expanded or activated bone marrow cells [12]. Transplantation of minimally manipulated bone marrow and vascularized whole organs is regulated by the Health Resource Services Administration (HRSA), and allogeneic unrelated bone marrow transfer falls under the authority of the National Bone Marrow Donor Program standards under HRSA.

In accordance with the statutory provisions, cellular products must be in compliance with CFR part 312 regardless of whether the finished product is shipped across state lines; this is due to the use of ancillary products used in manufacture. Ancillary products (e.g., bioreactors and cell culturing systems; components of culture media; agents used to purge, select, or stimulate specific cell populations) are not intended to be present in final products, but they may

have an impact on the safety, purity, or potency of the products. At the investigational stage these products must be in compliance with part 312. Therefore clinical trials are to be conducted under INDs. Although products regulated by FDA as biological products must also meet drug or device requirements, the agency does not require duplicate pre-market approvals.

Prior to 1993 the FDA did not require pre-marketing approval for many types of transplantation, including bone marrow transplants. To the extent that some bone marrow transplantation products are subject to extensive manipulation prior to transplantation, they are treated the same as somatic cell therapy and gene therapy products. FDA defined "manipulation" as the ex vivo propagation, expansion, selection, or pharmacological treatment of cells, or other alteration of their biological characteristics (11). Over the past two decades the FDA has published a number of guidance documents outlining various expectations for ensuring the safety of cellular therapies (13) (see Table 3).

The first guidance document relating to regulatory expectations for a cellular therapy product was entitled Points to Consider in the Collection, Processing, and Testing of Ex vivo Activated Mononuclear Leukocytes for Administration to Humans, and published in August 1989. In February 1996, FDA published a guidance for manipulated autologous structural (MAS) cells, which are autologous cells manipulated and then returned to the body for structural repair or reconstruction. Manipulation required for MAS cell products is more than that required for autologous bone marrow, where the source of material is harvested but otherwise undergoes minimal manipulation. Therefore MAS cell products were considered to fall within the definition of cellular therapy products. One year later the FDA published A Proposed Approach to the Regulation of Cellular and Tissue-Based Products.

The regulatory framework of the Proposed Approach document is a tiered risk-based approach: Products thought to present greater risk receive more regulatory oversight, and require more extensive control in manufacturing and clinical studies as well as more product characterization. The approach focuses on three general areas: (1) prevention of use of contaminated tissues or cells with the potential for transmitting infectious disease; (2) prevention of improper handling or processing that might contaminate or damage tissue or cells, and (3) assurance that clinical safety and effectiveness is demonstrated for tissue or cells that are highly processed, are used for other than their normal function, are combined with nontissue components, or are used for metabolic purposes. The manufacturing issues for cells lie in the preparation of the cell cultures and the assurance that cells are free of microbial contamination, such as mycoplasma (if cultured), or otherwise sterile. Appropriate screening/testing of donor tissue for communicable disease is essential as such criteria are necessary for accepting donor source materials to initiate production.

Consistency in lot-to-lot preparation of cells is paramount in the demonstration of manufacturing control. Critical manufacturing controls include

**TABLE 3 Key FDA regulatory guidance documents relating to cell-based therapies**

| |
|---|
| Points to Consider in the Collection, Processing, and Testing of Ex vivo Activated Mononuclear Leukocytes for Administration to Humans (August 1989) |
| Guidance on Applications fro Products Comprised of Living Autologous Cells Manipulated Ex vivo and Intended for Structural Repair and Reconstruction (May 1996). |
| FDA Proposed Approach to Regulation of Cellular and Tissue-Based Products (February 1997) |
| Guidance for Industry: Guidance for Human Somatic Cell Therapy and Gene Therapy (March 1998) |
| PHS Guideline on Infectious Disease Issues in Xenotransplantation. OMB Control No.0910-0456 (January 2001) |
| Guidance for Industry: Source Animal, Product, Preclinical, and Clinical Issues Concerning the Use of Xenotransplantation Products in Humans (final guidance) (April 2003) |
| Guidance for Reviewers: Instructions and Template for Chemistry, Manufacturing, and Control (CMC) Reviewers of Human Somatic Cell Therapy Investigational New Drug Applications (INDs) (draft guidance) (August 2003) |
| Guidance for FDA Review Staff and Sponsors: Content and Review of Chemistry, Manufacturing, and Control (CMC) Information for Human Gene Therapy Investigational New Drug Applications (INDs) (draft guidance) (November 2004) |
| Current Good Tissue Practices for Human Cell, Tissue, and Cellular Tissue-Based Product Establishments; Inspection and Enforcement (November 2004) |
| Guidance for Industry: Minimally Manipulated, Unrelated, Allogeneic Placental/ Umbilical Cord Blood Intended for Hematopoietic Reconstitution in Patients with Hematologic Malignancies. (draft guidance) (December 2006) |
| Guidance for Industry: Cell Selection Devices for Point of Care Production of Minimally Manipulated Autologous Peripheral Blood Stem Cells (PBSCs) (July 2007) |
| Guidance for Industry: Preparation of IDEs and INDs for Products Intended to Repair or Replace Knee Cartilage (draft guidance) (July 2007) |
| Guidance for Industry: Eligibility Determination for Donors of Human Cells, Tissues, and Cellular and Tissue-Based Products (HCT/Ps) (August 2007) |
| Guidance for FDA Reviewers and Sponsors: Content and Review of Chemistry, Manufacturing, and Control (CMC) Information for Human Somatic Cell Therapy Investigational New Drug Applications (INDs) (April 2008) |

*Note*: See http://www.fda.gov/cber/guidelines.htm.

standardization and optimization of reagents and processing criteria. Product characterization and development of acceptance criteria is also important, including (1) establishing specific characteristics to ensure product integrity, (2) identifying product parameters that anticipate adverse events, (3) developing analytical approaches for evaluating proposed acceptance criteria for in-process intermediates and final cellular product, and (4) controlling purity and impurity profiles of the final cellular product. Impurities arising from the scale-up of manufacturing such as host-cell contaminants, endotoxin contamination, and residual DNA levels are all potential safety concerns. Stability of the

protein sequence and posttranslational modifications are additional concerns for cell-based gene therapy products. Use of bovine serum is acceptable in the United States provided on that the source of the serum is demonstrated to be from herds reared for the entirety of their lives in certified BSE-free countries, although concerns have been expressed for use of bovine materials. Karyotypic analysis for determining genetic stability is also important for cells cultured for a long time.

In 2001 FDA put in place a comprehensive system consolidating the regulation of human cells, tissues, and cellular and tissue-based products (HCT/Ps) into one regulatory program. The definition of human cells, tissues, and cellular and tissue-based products (HCT/Ps) includes a broad range of cell- and tissue-based therapies and is intended to cover HCT/Ps at all stages of their manufacture, from recovery through distribution. The final new part 1271 regulation of chapter Assessment of Autoimmunity and Hypersensitivity is made up of six subparts. Subpart part A includes general provisions pertaining to the scope and applicability of part 1271, definitions and subpart B includes registration and listing procedures. The donor-suitability proposed rule contains subpart C of part 1271, and the GTP proposed rule contains subparts D, E, and F (see Table 3).

In Europe, the Council of Ministers approved the EC Regulation on Advance Therapy Medicinal Products in May 2007. Directive 2001/83/EC covers genetic medicinal products and somatic cell therapy medicinal products. Additional requirements are laid down in Regulations (EC)No726/2004 and a series of other European Commission Directives. The Advanced Therapy Medicinal Product (ATMP) regulations are amendments to Directive 2004/23/EC that provide quality and safety rules for the donation, testing, processing, preservation, storage, and distribution of human cells and tissues. All member states support the Regulation, which was implemented across the European Union at the end of 2007. The regulation governs production, distribution, and use of advanced therapies, including stem cell therapies for treating or preventing disease in humans. The Regulation also brings all innovative therapies into a single legislative framework and sets up a centralized marketing authorization procedure for such treatments.

The United Kingdom was first in Europe to introduce legislation covering human embryo research in 1990 following passage of the Human Fertilisation and Embryology (HFE) Act. The HFE Act included provisions in connection with human embryos and any subsequent development of such embryos prohibiting certain practices with embryos and gametes and the establishment of a Human Fertilisation and Embryology Authority. Currently in the United Kingdom the regulatory framework for human stem cell research is split into two, governing storage and use of human embryos and storage and use of human tissues and somatic cells. In July 2007 new regulations further implementing the EU Tissues and Cells Directive came into force in the United Kingdom. The Human Tissue (Quality and Safety) Regulations 2007 and the Human Fertilisation and Embryology (Quality and Safety) Regulations 2007 amend the current regulations governing the licensing of facilities in carrying

out work in relation to human tissues and cells intended to be used in human applications. The Regulations 2007 set out requirements for obtaining a license from the relevant authority, either the Human Fertilisation and Embryology Authority (HEFA) or the Human Tissue Authority (HTA) in order to carry out the various acts covered by the Directive in relation to human derived tissue and cells that are intended for use in human applications. In September 2007, following considerable debate that will likely continue among experts over the next few years, the HFEA decided to permit the creation of hybrid or chimera embryos (human–animal) for research purposes.

## 3 TYPES OF CELL-BASED THERAPIES

### 3.1 Ex vivo Activated Mononuclear Leukocytes

Adoptive immunotherapy is a general term describing the transfer of immunocompetent cells to a tumor-bearing host. The major research challenge in advancing adoptive immunotherapies has been to develop immune cells with specific antitumor reactivity that can be generated in large enough quantities for transfer to tumor-bearing patients. Examples of passive nonspecific immunotherapy include transfer of lymphokine-activated killer (LAK) cells or tumor-infiltrating lymphocytes (TIL), activated in vitro by interleukin 2 (IL 2) or other lymphokines. Examples of passive specific immunotherapy include transfer of specific immune cells, such as NK cells, draining lymph node cells or "antigen-loaded" pulsed dendritic cells. The pulsed dendritic cells function as potent immunostimulators of native T cells. A key safety issue for these various cellular therapies is related to potential contamination with autologous malignant cells and the selection of immunoresistant malignant cells by the ex vivo activation procedure. The potential for cells cultured through many passages to transform to an exogenous growth factor-independent phenotype should also be addressed.

Initial studies of systemic administration of autologous LAK cells and rIL2 in patients with advanced cancer were reported by Rosenberg and others in the early to mid-1980s [14–16]. The regimen was based on animal models in which the systemic administration of LAK cells plus rIL2 mediated the regression of established pulmonary and hepatic metastases from a variety of murine tumors in several strains of mice. Patients showed some benefit; however, LAK cells had to be generated form lymphocytes obtained through multiple leukaphereses [17,18]. Methods have subsequently been developed to shorten the culture period and increase the proliferation of cells to improve the cell yield.

Infusion of tumor-infiltrating lymphocytes (TIL) isolated from tumor samples and expanded in vitro with rIL2 resulted in a more efficient tumor reduction in animal models [19] and showed activity in clinical studies [20]. A significant impediment to generating therapeutic TILs resides in the inherent

limitations of the immunogenicity of the tumor from which the TILs are derived. In animal models using poorly immunogenic tumors, the therapeutic efficacy of TILs is limited. Tumors genetically engineered to produce certain cytokines have been found to contain TILs with enhanced in vitro and in vivo antitumor activity. Studies have shown that efficacy has been improved with TILs that are characterized by less time in tissue culture, longer telomeres, and a less differentiated phenotype.

In 1996 Celluzi et al. reported that active immunization strategies using antigen-pulsed dendritic cells were useful for inducing tumor-specific immune responses [21]. The authors showed that the major histocompatability complex class I–presented peptide antigen pulsed onto dendritic antigen-presenting cells (APCs) induced protective immunity in mice to lethal challenge by a tumor transfected with the antigen gene. The immunity was antigen specific, requiring expression of the antigen gene by the tumor target, and its elimination by in vivo depletion of CD8-T lymphocytes [21]. The first therapeutic tumor antigen vaccine evaluated by the FDA was Sipuleucel-T (Provenge(r)), a peptide-pulsed dendritic tumor cell vaccine manufactured by Dendreon for treatment of men with asymptomatic metastatic androgen independent prostate cancer (AIPC). Dosing is based on a minimal number of total nucleated cells in the starting material, which contains T, B, and NK cells and monocyte populations and a minimum number of CD45+ cells in the final product. The proposed mechanism of action of Sipuleucel-T is presentation of prostatic acid phosphatase (PAP) antigen by activated APCs to stimulate tumor-specific T cells in the patient [22]. Following a favorable advisory committee review in March 2007, the FDA issued an "approvable letter" in May requesting additional data from the sponsor in order to support a final approval.

In July 2007, a US biotech firm, Northwest Biotherapeutics, received approval form the Swiss Institute of Public Health to market to selected centers in Switzerland the first brain cancer vaccine, DCVax-Brain®, an autologous patient-specific dendritic cell based vaccine.

### 3.2 Living Autologous Cells Manipulated Ex vivo Intended for Structural Repair or Reconstruction (MAS Cells)

MAS cells are derived from a patient's tissues. The cells are manipulated ex vivo and then implanted into the same patient with the intent of providing repair or reconstruction of a structure rather than via systemic action of the cells. The manipulation of MAS cells involves dissociation of the selected tissue into individual cells, which are then propagated and expanded into large numbers of cells using tissue culture methods. Since MAS cells are implanted within an enclosed space, they are considered to be more similar to tissue and device products. Examples of MAS cells include chondrocytes for repair of focal cartilage defects, autologous fat cells for cosmetic augmentation, and autologous keratinocytes for dermal wound healing. In 1997 FDA approved the first MAS cellular therapy product under an accelerated approval

procedure, Carticel® manufactured by Genzyme Tissue Repair, which involves the autologous transfer of healthy chondrocytes into damaged knee joints following ex vivo cell culture and proliferation.

### 3.3 Adult Stem Cells (Somatic Stem Cells)

Adult stem cells are cells found in a number of differentiated tissues that can self-renew and differentiate (with certain limitations) to give rise to the specialized cell types of the tissue. Adult stem cells may also be able to give rise to specialized cell types of a completely different tissue, a phenomenon known as transdifferentiation or plasticity. However, the presence of pluripotent adult stem cells remains a subject of scientific debate.

Identification of stem cells from different sources has broadened the prospects for clinical applications. The types of stem cells that can be isolated from adult bone marrow include hematopoietic stem cells, endothelial progenitors, and mesenchymal stem cells. The adult brain contains neural stem cells that can generate the brains three major cell types of the brain, namely nerve cells (neurons) and nonneural cells, astrocytes, and oligodendrocytes. Epithelial stem cells in the lining of the digestive tract can give rise to a number of cell types, including absorptive cells, goblet cells, and enteroendocrine cells. Skin stem cells in the basal layer of the epidermis can give rise to keratinocytes, and follicular stem cells at the base of the hair follicle can give rise to both the hair follicle and the epidermis.

There are two characteristics that distinguish stem cells from other types of cells: stem cells can self renew, and they can differentiate into other types of cells. Stem cells renew themselves for long periods through cell division, and given certain conditions, they can be induced to become cells with special functions.

### 3.4 Hemapotopoietic Stem Cells

Hematopoietic stem cells (HSCs) can be derived from bone marrow, umbilical cord blood, placenta, or mobilized peripheral blood. The hemapoietic tissues may have cells with long- and short-term regeneration capacities and committed multipotent, oligopotent, and unipotent progenitors. HSCs give rise to all the blood cell types including myeloid and lymphoid lineages as wells as some dendritic cells. There is a growing body of evidence suggesting that HSCs are plastic and that, at least under certain circumstances, HSCs may participate in the generation of tissues other than those of the blood system. HSCs are phenotypically identified by their small size and certain surface markers. Murine HSCs have been demonstrated to be negative for lineage markers (Lin-), including myeloid and B and T cell lineages, and positive for c-kit, Sca-1, and CD34 [17]. The cell markers used to characterize the commonly accepted type of human HSCs include $CD34^{+}$, $CD59^{+}$, $Thy/CD90^{+}$, $CD38^{lo/-}$, $C\text{-}kit^{-/lo}$, and

lin- [23]. However, not all HSCs are covered by these combinations; for example, some cells lack CD34 [24,25].

### 3.5 Endothelial Progenitor Cells (EPCs)

Adult endothelial progenitor cells were first described by Asahara and colleagues showing that a purified subpopulation of CD34-expressing cells isolated from the blood of adult mice could differentiate into endothelial like cells in vitro [26]. Endothelial progenitor cells derived from bone marrow have the ability to differentiate into endothelial cells (cells that line blood vessels). These progenitor cells have been reported to express CD34, CD133, CD31, VE-cadherin, VEGFR2, and c-Kit. CD133 and CD34 are often used as markers for selection and purification because CD133 is not found on mature endothelial cells, and CD34 is not found on the undifferentiated stem cells from which endothelial progenitor cells derive [27,28].

Endothelial progenitor cells have been shown to participate in a number of regenerative activities. They contribute to pathologic angiogenesis such as that found in retinopathy and tumor growth. There is evidence that circulating EPCs play a role in the repair of damaged blood vessels after a myocardial infarction. Higher levels of circulating endothelial progenitor cells detected in the bloodstream predict better outcomes and fewer repeat heart attacks [29]. Recently tissue-resident stem cells have been isolated from the heart, and they are capable of differentiating to the endothelial lineage [30]. There is also increasing evidence suggesting that there are additional bone marrow derived cell populations (e.g., myeloid cells and mesenchymal cells) and non–bone marrow derived cells that can give rise to endothelial cells [28]. Although the role of EPCs in neovascularization has been convincingly shown by several groups, their precise function in neovascularization remains undefined.

### 3.6 Mesenchymal Stem Cells (MSCs)

Mesenchymal stem cells represent a rare population of bone marrow derived cells that do not express CD34 or CD1 Under defined in vitro and in vivo conditions mesenchymal cells can differentiate into a variety of cell types, including osteoblasts, chondrocytes, adipocytes, myocytes, and stromal fibroblasts [31,32,33]. Human MSCs have high expansion potential, genetic stability, and reproducible characteristics as has been proved in many laboratories. They have the ability to migrate to sites of tissue injury and to modify the response of immune cells. The later characteristic may support the development of allogeneic stem cell approaches when autologous stem cells are not readily available [34]. Because MSCs can produce important growth factors and cytokines, they may also be able to exert their effect without direct participation in long-term tissue repair. MSCs have not been shown to be capable of regenerating or maintaining a whole tissue compartment as has been shown for

hematopoietic stem cells (HSCs). Although MSCs can engraft and differentiate to some extent in host tissue, they do not have an ability comparable with the de novo tissue-forming ability of embryonic stem cells in embryoid bodies.

### 3.7 Human Embryonic Stem Cells (hESC)

Human embryonic stem cells (hESCs) were first derived from the inner cell mass (ICM) from the blastocyst stage (~100–200 cells) of embryos generated by in vitro fertilization [35,36]. In humans the blastocyst is an early-stage embryo, approximately 4 to 5 days old. The blastocyst can be formed by means of either in vitro fertilization or somatic cell nuclear transfer, in which the nucleus of a somatic cell is combined with an enucleated oocyte. Methods have been developed to derive hESCs from the late morula stage (30–40 cells) and from arrested embryos (16–24 cells incapable of further development), and more recently from single blastomeres isolated from 8-cell embryos (37).

Human embryonic stem cells are defined by the presence of several transcription factors and cell surface proteins. The transcription factors oct-4, nanog, and sox2 ensure the suppression of genes that lead to differentiation and the maintenance of pluripotency [38]. The cell surface proteins most commonly used to identify hESCs are the glycolipids SSEA3 and SSEA4 and the keratin sulfate antigens Tra-1-60 and Tra-1-81 [39]. The characteristic features of hESCs include their proliferative and self-renewing properties and their ability to differentiate into derivatives of all three primary germ cell layers: ectoderm, endoderm, and mesoderm. hESCs maintain pluripotency through multiple cell divisions. Although hESCs can form all somatic tissues, they cannot form all of the other "extraembryonic" tissues necessary for complete development, such as the placenta and membranes, so they cannot give rise to a complete new individual [40]. The only other cells with proven pluripotency similar to that of hESCs are embryonic germ cells (hEGCs) derived from primordial germ cells [40]. Biologically hEGCs have many properties in common with hESCs [41]. In humans, hEGCs were first established in culture in 1998 from tissue derived from an aborted fetus [42].

As of October 2006, 500 hESC lines had been reported in the literature (including posters and meeting abstracts). Of these, 455 were normal, no-disease-related cell lines and 45 were disease related [43]. In response to concerns regarding the ethics of deriving hESCs from early-stage embryos, the US federal government initiated a moratorium on the use of federal funds to derive and/or work on hESCs after August 9, 2001 [44]. At this time, a total of 71 hESC lines were identified as being acceptable for study using US government funds. Only 15 of the 27 hESCs lines that are eligible are fully characterized and available for research to scientists around the world [43]. All hESCs derived after August 9, 2001, are considered by the US government to be ineligible for NIH funding. As of October 2006, 384 of these lines were derived

from normal embryos. Of these, 73 have been fully characterized, karyotyped, shown to express hESC "stemness" markers, demonstrated to differentiate into derivatives of all three germ cell layers both in vitro and in vivo, and fully described with respect to derivation and propagation methods [43].

As long as embryonic stem cells in culture are grown under certain conditions, they can remain undifferentiated (unspecialized). But if cells are allowed to clump together to form embryoid bodies they begin to differentiate spontaneously into specific cell types (muscle cells, nerve cells, etc.). To generate cultures of specific types of differentiated cells, various growth factors are used in the culture media [45].

The scope of current stem cells research is larger than providing cells for transplantation. Cultured pluripotent stem cells are also being used in toxicity testing, identifying drug targets, studying cell differentiation, and understanding the prevention and treatment of birth defects.

### 3.8 Xenogeneic Cells

Xenogeneic cell therapy refers to live cells or tissues from a nonhuman animal source administered to a human recipient in vivo or ex-vivo. The main interest for using xenogeneic cells and tissues is the insufficient supply of human cells or organs. While attempts have been made to implant xenogeneic cells or tissues into patients, these attempts have often failed due to immunological incompatibility. From the public health point of view, the risk of cross-species transmission of infectious agents remains the most serious obstacle for using xenogeneic cells in humans.

## 4 CASE-BY-CASE APPROACH BASED ON PRODUCT ATTRIBUTES AND INDICATION

Ideally cell-based therapy products should (1) proliferate sufficiently to generate desired quantity of cells, (2) differentiate into desired cell type(s), (3) retain normal chromosomal complement, (4) survive in recipient after transplantation, (5) integrate into the surrounding tissue after transplantation, (6) exhibit controlled proliferation, (7) function appropriately for the duration of the recipient's life, and (8) avoid harm to the recipient. The source of cells in addition to the clinical indication influence the type and level of concern regarding safety [46] (see Table 4). While some concerns can be generalized across various cell types—such as about the migratory potential of transplanted cells in vivo and the possibility of infectious contamination—concerns regarding genetic instability, inherent or culture-induced tumorigenic potential, and the potential immunogenicity of host versus graft and graft versus host are more specific to the particular cell types. The remainder of this chapter will focus on the preclinical development considerations related to adult and embryonic stem cell based therapies to enable clinical trials.

**TABLE 4 Product attributes based on source of cell-based therapies**

| Product Attributes | Primary Cells (Allogeneic) | Adult Stem Cells (Autologous) | Adult Stem Cells (Allogeneic) | Embryonic Stem Cells (Allogeneic) |
|---|---|---|---|---|
| • Accessibility | • Easy | • Easy | • Easy | • Difficult |
| • Proliferative capacity | • None | • Limited (multipotent); reprogram to become more "stem-like"? | • Limited (multipotent); reprogram to become more "stem-like"? | • Unlimited (pluripotent) |
| • Differentiation capacity | • None | • Restricted (decreases with age of donor as well as time in culture; may not differentiate into cells that are needed) | • Restricted (decreases with age of donor as well as time in culture; may not differentiate into cells that are needed) | • Unlimited (reliable selection procedures needed to produce homologous cell populations) |
| • Genetic manipulation | • None | • None | • More difficult than ESCs | • Can be manipulated to incorporate foreign nucleic acid sequences in a stable manner |
| • Quality (reproducibility) | • No | • Yes | • Yes, not derived as single clones | • Yes, single clone |
| • Immunologic risk (transplant rejection) | • Possible | • Low (culturing ex vivo could lead to allogeneic epitopes) | • Possible | • Possible |
| • Safety (transmission of infectious agent(s)) | • High risk | • Low risk | • Moderate risk | • Moderate risk |
| • Safety (tumorigenicity) | • No concern | • No concern | • Increased concern based upon length of culture period | • Moderate to high concern |
| • Economic feasibility/advantage | • High cost | • High cost | • Moderate cost | • Potentially lower cost |
| • Ethical/legal concerns | • None | • None | • Fewer | • Many |

## 5 PRINCIPLES OF PRECLINICAL SAFETY EVALUATION FOR STEM CELL BASED THERAPIES

Stem cells are a dynamic biological entity and are intended to persist in the transplant environment for prolonged periods. They respond to their environment and undergo changes that are dependent on extrinsic signals as well as their intrinsic cellular properties and are thus more complex than defined metabolic pathways of drugs. However, the principles of preclinical safety evaluation for stem cell based therapies are similar to those of small-molecule drugs and closely parallel biopharmaceutical products. Studies are expected to define a safe starting dose and dose escalation scheme for clinical trials, to provide information to support the route of administration and the application schedule, to identify target organs of toxicity and parameters to use in monitoring clinical trials and the extent of follow up, and to determine populations that may be at greater risk for toxicity [47–52].

Study designs should take into consideration the type of cells that will be evaluated, selection of a relevant animal model, including a model that mimics the intended clinical population or disease state, and the treatment regimen, including route of administration. In most cases animal models of disease will be the most relevant model to assess not only proof of concept but also safety.

Two approaches have been used for predicting how stem cells will behave after transplantation into humans. Either human cells are transplanted into animals (generally rodents) with or without immune suppression or labeled, or unlabeled analogous cells are transplanted into the respective animal species. Both approaches have their limitations since transplantation is regulated by a species-specific network of biologically active molecules such as growth factors, cytokines, cell adhesion molecules and receptors, and extracellular matrices that may differ among species and thus influence engraftment [5].

Cells derived from humans at the same stage of development with the same overall characteristics may also differ significantly, as humans show a high frequency of allelic variability, which may be greater than the transplant models that frequently use inbred strains. In addition human cells tested in animal models may behave differently than human cells transplanted into humans because of provoked immune responses and/or intrinsic differences in cell properties or host environments for which limited information is available [5]. Although the overall properties may seem similar when analogous products are compared, there may be significant biological differences in cells isolated at similar stages of development in terms of developmental patterns, environmental cues, and growth factor dependence in culture as well as expression of specific markers for more restricted precursors that makes isolation and purification of subsets of progenitors more difficult [5].

When the data from analogous cell products are used to guide clinical trials, in-depth comparison between the animal and human cells is needed. The

concerns with determining product comparability are not unlike those raised when using homologous or analogous proteins [49]. Potential processing, formulation, and storage may be different; cellular morphology, phenotype, and the level of engraftment and persistence may differ; and the cells may have different function or regulation (e.g., species-specific signaling may be present). Limited characterization of the animal product may also introduce uncertainty for extrapolation (e.g., potentially different impurities/contaminants). Ensuring comparability is even more important when assessing toxicity, as extrapolation of an initial safe starting dose is generally based on safety rather than activity in animals.

## 6 PHARMACOLOGY

Before initiating clinical studies, it is essential to demonstrate that cell preparations possess well-defined biological activity. Studies should be adequate to demonstrate the intended pharmacological activity in relevant in vitro and in vivo models. Initially in vitro studies, addressing cell and tissue morphology, proliferation, phenotype, heterogeneity, and the level of differentiation may be used to support pharmacological activity [52]. Relevant animal model selection is based on the biological and structural goal of the study, the disease under study, the animal's anatomy and physiology and size, availability, housing requirements, cost and technical feasibility, as well as historical experience with the model.

Animal transplant models of human disease support a rationale for conducting the initial clinical trials in patients. These studies provide information concerning the feasibility of the treatment and explore the establishment of a dose–response relationship between the cellular product and activity outcomes. The studies are also designed to understand the basic mechanisms of structure and function, host–implant interface, and systemic effects after transplantation. Such observations are particularly important for products that contain cells that are not terminally differentiated at the time of transplantation. Since most human diseases do not occur spontaneously in animals, animal models of human disease that are generally created by chemical, surgical, or immunological methods are often imperfect [5]. In all cases the abilities and limitations of each animal model including level of characterization of the model should be considered for extrapolating findings to humans.

Ideally biomarkers of activity should be identified at various times over the course of the study to support the pharmacodynamic activity (e.g., normalization of insulin, improvement in beta cell function as measured by C-peptide level, or control of glucose following transplantation of β pancreatic islet cells); improvement of motor coordination in mice with spinal cord damage following transplant of neurons; or repair of heart function (e.g., functional measures such as LV ejection fraction, pressure volume loops, ventricular pressure and heart wall thickness). Such markers may also be useful in subsequent clinical

**TABLE 5 In vivo preclinical demonstration of POC: Stem-cell based cellular therapies**

- Identify which cell(s) will be used.
- Characterize cell population(s) using multiparametric analytical testing.
- Establish comparability of cellular products if an analogous product is being evaluated.
- Perform studies in animal transplant model of human disease (consider cell biology, anatomy, pathophysiology, biomechanics, etc; immunosuppression regimen, identify limitations).
- Consider reasonable group size such as 10/group or 5/sex/group for rodents; total of 3 to 5/group for large animals (consider attrition based on surgical procedures, concomitant treatments, historical information on the model, etc.; consider incorporation of various controls including placebo, sham, and positive controls).
- Explore dose–response relationship (determine minimal or optimal cell number, maximum feasible dose).
- Assess feasibility of intervention (timing of lesion by chemical or surgical insult, location of cell transplant, route of administration, biocompatibility of delivery device, etc.).
- Determine biomarkers of activity.
- Define functional activity (desired effect), and justify timing and frequency of assessments (assessors should be blinded to treatment groups).
- Assess early response and durability of response (justify timing).
- Identify possible incorporate safety endpoints.
- Establish comparability of clinical material over the course of clinical development.
- Consider more than one species to enhance knowledge of cell placement and dose extrapolation.

trials. If the intended use is to restore function of deficient cells (tissue regeneration) functional tests should be implemented to demonstrate that function is restored. If the intended use is adoptive immunotherapy or vaccination in cancer patients, immune assays to assess immunological effects should be used [52]. Studies should also be designed to determine the optimal amount of cells needed to achieve the desired effects and the extent that immunosuppressive agents act synergistically in promotion of graft survival. (see Table 5)

## 7 SAFETY PHARMACOLOGY

Cells by themselves or by secreting pharmacologically active substances may have effects on the CNS, cardiac, respiratory, renal, or GI systems. Safety pharmacology should therefore be considered on a case-by-case basis depending on the specific characteristics of the cell-based product [52]. In general, specific assessments are made as part of the toxicology assessments rather than as stand-alone studies consistent with the assessments made with protein-based biopharmaceuticals [50]. The fundamental physiological differences (e.g., total blood volume, pulmonary capillary surface area, and volume) should

be considered when extrapolating across species as well as specific attributes of the cells including viability and potential cell aggregates. Flow rate is particularly important when administering cells by the intravenous route. Assessment of CNS effects will be more relevant to therapeutic interventions to the nervous system.

## 8 EXPOSURE ASSESSMENT

Although the practices for conventional PK/ADME studies are not relevant for cell-based therapies, the principles of understanding exposure are relevant to determining optimal dose and understanding cell localization and migration. Various noninvasive (nonterminal) imaging modalities including positron emission tomography (PET), single-photon emission tomography (SPECT), or magnetic resonance imaging (MRI), fluorescence, and bioluminescence imaging provide visualization of grafted cells, evaluation of targeting of transplants to specific areas, and monitoring the regenerative process changes after cell therapy. Ensuring accurate cell delivery to the correct physiological site is fundamental to determining the activity and safety of cell-based products. Therefore studies are conducted to determine spatiotemporal information about cellular trafficking patterns, tissue distribution (biodistribution), and persistence. It is also important to characterize the dynamics of cell proliferation, growth, alterations in phenotype, and the kinetics of cell death. If there is no effect after cell transplantation, it is important to determine if that is due to incorrect cell function or to improper cellular targeting. In addition the unintended migration of the cells including release of bioactive products to "ectopic sites" may result in adverse effects.

Various strategies are used for cell labeling and tissue visualization. Ideally labeling techniques should not perturb the target cell, be able to quantify the cell number, have minimum transfer to nontarget cells, and be compatible with histochemical methods used for morphological analysis. Examples of labeling strategies include radioisotopes, fluorescent dyes, and 5-bromo-2-deoxyuridine (BrdU). BrdU binds to DNA during its synthesis and labels the nucleus. It can be identified in both frozen and paraffin-embedded sections by either immunostaining as brown staining (peroxidase) or red staining (alkaline phosphatase) after treatment with an appropriate anti-BrdU antibody [53]. Chloromethylbenzamido-DiI-derived (cm-DiI), a red autofluorescent dye that binds to the cell membrane, can be identified in frozen sections [54].

While these techniques are sensitive, they can be hampered by cytotoxic side effects of the labeling procedures and can be limited by dilution and loss of the marker over time as a result of cell division [55]. Thus, while useful for measuring the efficiency of acute delivery, these techniques may offer minimal long-term information regarding cell survival, proliferation, and differentiation [56].

Use of viral transduction methods allows stable integration of reporter genes—such as green (GFP) or yellow (YFP) fluorescent proteins, lacZ,

alkaline phosphatase, or firefly luciferase—permitting the evaluation of viable cells by fluorescence microscopy and flow cytometry. Reporter gene imaging can be used to track long-term cell survival, migration, division, and differentiations, and this method provides information on location and numbers of cells [57,58]. Cells can also be labeled via in situ hybridization for Y chromosome showing survival of male donor cells into female recipients [59]. Human cells can further be tracked in animal tissue by qPCR for human specific sequences. The disadvantages of reporter gene, Y chromosome, and qPCR methods is the requirement of a biopsy or animal sacrifice for the performance of the cell tracking in tissues [60].

Magnetic resonance imaging (MRI) can be used to track cells labeled with contrast materials. The two classes of contrast material that are generally used as MRI imaging labels are gadolinium chelates and iron oxide nanoparticles [61]. Iron oxide particles are a class of superparamagnetic MRI contrast agents. The particles range from tens of nanometers in diameter, termed ultrasmall superparamagnetic iron oxide (USPIO), to 100 nm (superparamanetic iron oxide [SPIO]), to some larger than 1 μM, known as micrometer-sized iron oxide particles (MPIOs). These compounds consist of magnetite (iron oxide) cores that are coated with dextran or siloxanes encapsulated by a polymer or further modified to facilitate internalization. They can confirm cell delivery and monitor short-term migration into surrounding tissues and can also be used in larger animals and humans [61–64]. MRI methods are, however, unable to distinguish between viable and nonviable cells. Nuclear labeling techniques, such as $^{18}$F, can theoretically be used in combination with tracking methods to evaluate the anatomy of the transplant in relation to its function [60,65].

Ultimately multiple imaging modalities as well as newer more sensitive approaches to track live cells may be needed for complete functional monitoring of the various cellular repair processes.

## 9 GENERAL TOXICITY

Although guidance documents have been issued by regulatory agencies to provide assistance in the preclinical safety of cellular products [51,52], toxicologists should recognize that the ultimate study designs including appropriate animal models may be product specific (see Table 6). Ideally toxicology studies designed to support product safety are fully GLP compliant.

However evaluation of safety in animal models of disease may be difficult to conduct with strict adhere to current good laboratory practices (cGLPs) due to novel routes of administration, limited animal numbers, uniqueness of the animal model of disease, and/or novel surgical procedures. In such cases the principles of the regulation can still be followed, and where deviations occur, they should be evaluated for their impact on the expected clinical application and discussed. An in-depth discussion should be provided to support a regulatory submission.

**TABLE 6 In vivo preclinical safety evaluation: Stem-cell based cellular therapies**

*Exposure assessment*

- Determine cellular fate/ biodistribution and persistence (cell migration/trafficking, engraftment, differentiation, cell phenotype, fusion, tissue integration, proliferation, survival). Use of noninvasive (nonterminal) imaging modalities and special histological stains (terminal). Determine presence of ectopic foci. Justify multiple time points (e.g., 1 week, 1 month, 3 months, 6 months or 12 months).

*General toxicity assessment*

- Select one or more available relevant models (e.g., age of animal, relevance to human disease).
- Select test article (human cells or comparable analogous cells).
- Determine whether immunosuppressive agents are needed and their affect on activity.
- Justify number of animals/sex/group (e.g., 10/group or 5/sex/group for rodents; total of 3 to 5/group for large animals per time point).
- Justify dose levels (include minimum effective dose, multiples of proposed human dose; consider negative and positive controls).
- Justify route and timing of administration (if device is used, assess biocompatibility with cells).
- Determine general health status.
- Assess daily food consumption.
- Assess weekly body weights.
- Assess clinical laboratory parameters (hematology, coagulation, serum chemistry).
- Assess local tolerance (infusion toxicity, implant site reaction, local microenvironment, release of paracrine factors).
- Assess host immune response (inflammatory response in target/nontarget tissue, auto-antibodies, sensitized cells against normal tissues/organs).
- Assess morphologic response in target/nontarget tissue.
- Identify functional and other laboratory endpoints (assessors should be blinded to treatment groups).
- Identify target tissues.
- Characterize toxicities (delayed toxicity/ reversibility of toxicity).
- Assess macroscopic pathology (gross observations) at justified sacrifice intervals.
- Assess microscopic pathology (histopathology) at justified sacrifice intervals; analyze multiple sections from site of transplantation. Consider specific IHC staining in addition to standard staining.

*Tumorigenicity assessment*

*Established cell lines, embryonic pluripotent stem cells, adult tissue derived multipotent stem cells-extended culture*

- Determine one relevant animal model (chemically immunosuppressed rodent, genetically immunocompromised/immunodeficient rodent species).
- Confirm cellular engraftment in animal model via intended clinical route of administration.
- Justify number of animals/sex/group (e.g., minimum of 20/group; 10/sex/group).
- Justify dose level(s) (consider maximum feasible dose, positive and negative controls).
- Justify study duration (minimum of 6 months up to 12 months).
- Assess microscopic pathology (interpretation of findings—frequency of tumor formation, inappropriate proliferation without malignant formation, etc.).

When preclinical studies are conducted "in house" or are reported in peer-reviewed journals and intended to support clinical trials, they should be well-controlled and designed to answer specific toxicological questions. The data should also be available in sufficient detail to allow an independent review of the studies. Ideally study designs include not only efficacy but toxicological endpoints such as defined clinical laboratory parameters, macroscopic, and microscopic evaluation of tissues.

As is true for all biopharmaceuticals, toxicity studies should be performed in relevant animal models. For cellular therapies these models are often in animal models intended to mimic the human disease. When possible the intended human cells are utilized for assessments with or without low-dose immunosuppressants (e.g., 10 mg/kg, i.p. dose of cyclosporine A in rats). The immunosuppressive drug is generally administered prior to the cell dose and extended for a specified period after the transplant. In cases where it is not feasible to use the intended clinical material, largely due to the inability of the cells to engraft sufficiently into the host, analogous cells can be used to assess preclinical safety and activity. In such cases it is important to understand the potential impact of any differences between the analogous product and the clinical product in order to improve extrapolation of safe and active cell doses to humans.

In many cases, only one relevant animal model of disease may be available to assess activity and/or safety; in other cases, more than one model may be available and sufficiently characterized to allow for meaningful extrapolation. Two species, including larger animal species, can be used to assist extrapolation of dose to humans and/or to test the safety of cell-based product with the intended clinical delivery device. In the latter case, it is important to assess not only biocompatibility of the cells with the delivery device but the delivery reproducibility, namely the quantitative recovery of cells after injection based on the formulation and concentration to be used.

Usually three dose levels are assessed, including the minimal effective dose and a maximum feasible dose. It is also good to look at previous preclinical and clinical experience with similar products for estimating clinical dose. Doses are extrapolated based on body weight, body surface area, or target organ volume, depending on the route of administration. The rate of delivery is especially important for cells administered intravenously. Other considerations include cell concentration and delivery volume, the number of injections, and cell viability. Cell viability is generally equal or greater than 70%.

Multiple time points are included in study designs to assess acute and late effects as well as the reversibility of any adverse effects. The route of administration should reflect the intended clinical use. Repeated-dose toxicity studies are only relevant if the clinical use includes multiple dosing. For example, in the case of β pancreatic islet cells, toxicity studies should mimic the clinical scenario of retransplantation assessment of acute toxicities and cumulative effects that would preclude retransplantation.

The potential for autoimmunity should be considered when cells are used for immunotherapy purposes [52]. The basic local alignment search tool

(BLAST) is generally used to identify the potential cross-reactivity with endogenous molecules which could signal an autoimmune concern [66].

## 10 REPRODUCTIVE AND DEVELOPMENTAL TOXICITY

The need for reproductive and developmental toxicity studies is dependent on the product, the clinical indication, and the intended patient population [50,52]. Consideration is based on the nature of any expressed products and/or inappropriate biodistribution. Effects to the reproductive system that were identified in exposure and general toxicity assessments that suggest a cause for concern must be addressed in these more specific studies. If studies are necessary, study designs will likely need to be altered to accommodate selection of relevant animal model, dose selection, and dosing frequency.

## 11 GENOTOXICITY

Similar to protein-based biopharmaceuticals, the standard battery of genotoxicity studies is not considered to be relevant for cell-based therapies unless there is a specific cause for concern regarding the nature of any expressed products that would indicate a potential interaction directly with DNA or other chromosomal material [50,52]. The conduct of genotoxicity studies for assessing the genotoxic potential of process related contaminants is also not considered to be appropriate [50].

## 12 TUMORIGENICITY

One of the greatest concerns about using undifferentiated stem cells is their potential to develop tumors [67,68]. The investigation of immortalization, malignant transformation, and tumorigenicity, however, are unlikely to be addressed by conventional rodent bioassays for carcinogenicity or the alternative short-term transgenic mouse assays (see Table 7).

Human stem cell lines are actually defined by their formation of benign tumors (teratomas) following injection of cells into an ectopic site in immunocompromised mice. The teratomas are comprised of differentiated tissue originating from all three embryonic germ layers, thus confirming the ability of the cell lines to give rise to all cell types in the body [45]. To date, tumor formation has not been reported in animal or human studies with stem cell therapies when passaged for standard periods such as 6 to 8 weeks. However, there has been at least one report of late-passaged mesenchymal stem cells (16–20 weeks) developing chromosomal abnormalities and spontaneous cancerous transformation in vitro [69].

**TABLE 7 Dose extrapolation for FIH dose: Stem-cell based cellular therapies**

| *Points to consider* |
|---|
| • Comparability to clinical product if analogous product is used (processing, cellular morphology, phenotype, species-specific signaling, impurities, contaminants, formulation and storage) |
| • Response to cytokines, chemokines, growth factors, adhesion molecules, and their receptors |
| • Inherent limitation of animal model compared to human disease (e.g., etiology, progression, phenotype, pathology; acute vs. chronic disease) |
| • Influence of host environment of human recipient (age, disease, medical history) |
| • Allelic variability in humans (compared to inbred animal strains) |
| • Transplant and surgical technique impacting diffusion and growth of cells |
| • Timing of transplantation |
| • Physiological and biological in the response of the transplanted cells to the local environment |
| • Immune response |
| • Dose administration (concentration, volume, rate of delivery, location of injection, number of injections, cell stability, delivery device) |
| • Scaling factor (e.g., body weight, body surface area, volume of target organ) |
| • Cross species validation (if more than one animal model is available) |

The need for assessment of tumorigenic risk should therefore be considered based on the cell type (including cell source and "stemness" of the cells) as well as the degree of post collection ex vivo manipulation. Analysis may include assessments of proliferative capacity, dependence on exogenous stimuli, and response to apoptosis stimuli or genomic modification.

When needed, a tumorigenicity study should be performed with cells that are at the limit or beyond routine cell culturing. Generally, tumorigenicity is evaluated in genetically immunodeficient or chemically immunosuppressed rodents (mice or rats), 20 animals per group, via the intended clinical route of administration for at least 6 months duration. Although in some cases, based on specific product attributes, cells may be minimally detectable at 6 or 12 months even in the presence of immunosuppressive regimes. Ideally the study should be designed to include both negative and positive control groups. A pilot study is often performed to confirm cell engraftment. One dose group may be sufficient, if it is the maximum feasible dose. Tissues found to contain cells or expressed products during the biodistribution studies should also be analyzed with special emphasis during the tumorigenicity study [52].

## 13 LOCAL TOLERANCE

One of the goals of preclinical evaluation is to improve efficacy and safety by optimizing both the method of cell delivery and the site of cell delivery

(transplantation site). Thus many novel routes of delivery have been employed based on specific product attributes and intended clinical indication. For example, various sites have been successful for transplantation of β islet cell-based therapies in experimental animal models, including intraportal, kidney subcapsule, intrasplenic, intrathymic, intraperitoneal, anterior chamber of the eye, subcutaneous, intra-abdominal, and omental pouch. Other unique routes of dosing include autologous adult olfactory neuroepithelial-derived neuronal progenitor cells for spinal cord repair delivered via the intralesional route, allogeneic adult bone marrow derived MSC for acute myocardial infarction via the intravenous route, and human fetal-derived progenitor cells for treatment of myelodegenerative CNS disease via the intracerebral route.

As a result of these unique delivery routes, there is an increased potential for toxicity to tissues at or near the injection site. Assessment of local tolerance and tissue compatibility of the clinical formulation is usually evaluated as part of the animal model of efficacy and/or general toxicity study, thus obviating the need for separate local tolerance studies. There may be additional concerns in cases where analogous products are used including potential differences in formulation, stability, and so forth.

## 14 EXPECTATIONS FOR PRE-PREIND FILING IN THE UNITED STATES

The Office of Cell Tissue and Gene Therapy (CBER/FDA) has historically encouraged sponsors that are developing products in an emerging field to engage in early dialogue to better understand the current regulatory concerns and expectations. The information necessary to engage in early discussions with the agency includes (1) a summary description of the intended clinical product; (2) a summary outline (e.g., one page) of the proposed clinical trial, to include the subject population, possible dose levels, dosing regimen, route of administration, dosing procedure (devices, if intended for delivery of the product), and parameters that will be assessed; (3) a comprehensive summary of all preclinical (in vitro and in vivo) activity/proof-of-concept and toxicity studies conducted and the results obtained, including relevant publications; (4) a detailed discussion, with protocol outlines, regarding what additional preclinical studies (toxicity/safety and proof of concept) are planned to adequately address the safety of the product; and (5) specific questions relating to pharmacology and toxicology issues/assessments. While the comments made during these early discussions with FDA are nonbonding, they are intended to help in preparing the subsequent pre-IND submission. The agency, however, discourages the conduct of pivotal preclinical studies if there are significant questions related to the proposed approach. The general issues and specific considerations are provided in the Appendix. In the EU, the scientific advice

procedures in various countries and at the EMEA level are also useful in guiding sponsors developing novel therapies.

## 15 THE TRANSITION INTO CLINICAL TRIALS

As novel cell-based therapies move from preclinical studies into the clinic, researchers and regulators face new and evolving issues and uncertainties involving long-term safety and efficacy. Justification for proceeding with clinical trials is based on establishment of a rational scientific basis, demonstrated proof of concept, and an adequate demonstration of safety in relevant animal model(s). In comparison to pharmaceuticals, cell therapies require an increased coordination and logistical consideration to implement the clinical trial. It is obviously important to be vigilant against ex vivo contamination of the cell-based products.

The first in human trial (FIH) for cell-based therapies is in patients often with no available alternative therapies rather than in normal volunteers. These are also populations where clinical effects are less easy to manifest themselves and where adverse events may be difficult to distinguish from disease-associated events. The surgical procedures and use of immunosuppressants may also contribute to the adverse event profile. Therefore it is even more important to understand the nature and extent of the pathological process underlying the disease symptoms of the patients enrolled in the early clinical trials.

The timing of the transplant as well as location of the transplant are important considerations in the trial design. Dose extrapolation algorhythms differ from conventional pharmaceuticals [52]. In many cases the highest dose in animals may be limited to a maximum feasible dose. Extrapolations need to take into consideration the cell concentration and number of cell injections and expected cell migration due to potential differences in anatomy and physiology. Fewer dose groups are generally employed (e.g., 3 to 4) to define an optimal biological dose rather than a maximum tolerated dose. A review of published clinical trials with similar products and indications is therefore useful to support initial dose and dose escalation schemes. If an analogous product has been used to support preclinical evaluations, comparability assessment is needed to reduce the uncertainty of extrapolations.

Although it may be difficult to definitively assess location and function at a cellular level, a variety of nonivasive technologies are currently being explored to track cells and to observe changes in metabolic activity in specific tissue areas, such as monitoring glucose metabolism using PET in stroke patients. The potential for immune reactions against implanted cells is also closely monitored and may depend not only on the source of cells but on the extent of tissue damage or bleeding at the implantation site(s).

Other aspects of trial design including randomization, acceptability of sham operation in a placebo arm, appropriate statistical approaches, as well

as stopping rules and duration of follow-up. The duration of follow-up to monitor for complications and any delayed effects including, in some cases, detection of possible malignant change or chromosomal abnormalities will differ based on the specific product or product class and indication.

## 16 CONCLUSIONS

During the past two decades there have been rapid and exciting advancements in cell therapy research. Today stem cell-based therapy is a therapeutic possibility. A number of cell-based therapy products have transitioned into clinical development for a range of applications including genetic diseases, cancer, cardiac disorders, diabetes, diseases of the nervous system, bone and joints, and wounds of the skin and soft tissues. While a single source of cells may not best or even suitable for all indications, some cells may have the potential to be definitive therapies for replacement, restoration of function, or reconstruction of various tissues. Current manufacturing technologies also have the potential to provide an unlimited number of defined and standardized cells for transplantation, but rigorous characterization is needed to address biological heterogeneity, variability, and incomplete cellular definition.

Cell-based products present similarities as well as complex issues not encountered with "traditional biologicals." Innovative uses of animal models have contributed to advances toward clinically applicable treatment options based on studies designed to answer specific questions. Novel models and experimental paradigms are best considered as a case-by-case approach applied to each specific product. Noninvasive imaging methods play a critical role in cell therapy by identifying how many cells reach the desired target and, if they survive, divide and differentiate to fulfill their intended function.

Successful translation of cell-based therapies into clinical benefit, however, still faces many challenge, including developing more efficient methods for differentiation into the desired cell type(s), maintaining genetic stability during long-term culture, ensuring the absence of potentially tumorigenic cells, optimizing the implantation medium to reduce infusion toxicity, and optimizing the surgical delivery procedures (see Table 8). Process and analytical methods as well as definition of the final product will likely evolve throughout the clinical trials, and therefore assessing safety will be an ongoing iterative process during clinical development.

Regulatory and ethical issues are often inextricably linked with novel therapies. Since cell-based therapy is a rapidly evolving field for both researchers and regulators, it is essential that each potential new cellular therapy be considered individually and that researchers maintain frequent and personal contact with regulatory agencies. Developers and regulators need to identify the issues that must be considered in order to best evaluate the risks and the benefits.

**TABLE 8 Current challenges of stem-cell based cellular therapies**

- Understanding cell specialization (biological signals to differentiate cells into specific cell types required for normal organ function)
- Understanding cell cycle control (regulation/control of cell division/differentiation)
- Establishing specific cell characteristics to ensure product integrity
- Ensuring availability of clinical grade ancillary products (cytokines, growth factors, transcription factors)
- Improving cell scalability
- Optimizing cell preservation techniques
- Ensuring long term genetic stability of cells
- Optimizing cell delivery
- Improving evaluation and understanding of cell–host interactions for improving graft retention and survival
- Understanding potential risks if undifferentiated or partially differentiated stem cells are needed to show benefit
- Identifying exogenous controls if inappropriate function of cells or unregulated cell growth (malignant cells)
- Minimizing ethical and political constraints while encouraging proper control of the sources and procedures

## APPENDIX PHARMACOLOGY/TOXICOLOGY CONSIDERATIONS REGARDING SOMATIC CELL PRODUCTS

### General Questions to Ask Regarding the Completed and Planned Preclinical Studies

- Do toxicological data derived from preclinical models provide information for the clinical management of potential toxicities?
- How predictive will be the preclinical activity/toxicity data for the human response?
- What will be the impact of the preclinical data on clinical development?

### Product Used in Preclinical Studies:

Data should be provided to show the comparability of the product lot used in the POC studies, the toxicology studies, and the cell trafficking studies to the intended clinical lot. When possible, the intended clinical product should be used to conduct the pivotal preclinical toxicology studies. Immune tolerance to human cells (genetically or chemically immunosuppressed animals) allows use of the intended clinical product. In cases where analogous animal cells are used, comparability to the human product should be provided.

### Product Delivery Issues

Ideally preclinical safety evaluation studies should be designed based on the intended clinical route of administration using the intended clinical delivery device (if applicable).

### Animal Species/Models

Studies should be designed in a relevant model. Consideration should be given to the usefulness/ability of existing models to (1) mimic the disease/human population (activity/efficacy) and (2) predict safety (toxicity) in the context of the similarity in anatomy and pathophysiology. A discussion of whether the disease status of the animal has an impact on pharmacologic/toxicologic activity and whether the investigative therapy has an impact on the disease status of the animal should be provided.

While toxicology studies can be conducted in healthy animals, considerations should be given to hybrid pharmacology-toxicology study designs when possible for incorporation of both activity and toxicity endpoints. In any case, adequate numbers dose groups should be used, and time points should be justified. In most cases both genders should be used.

### Cellular Fate Post Transplantation

The fate of the cells should be determined, including comprehensive histopathological evaluation of cell trafficking/migration, engraftment/survival posttranslation (e.g., differentiation, plasticity, transdifferentiation and/or fusion), as well as unwanted expression or alteration of phenotypes.

### Tumorigenicity Potential

Tumorigenicity should be considered based on product attributes and intended clinical indication. If performed, the clinically relevant route of administered should be used, and the appropriate animal species, including the appropriate study design and duration, should be justified.

## ACKNOWLEDGMENTS

The author would like to thank Mercedes Serabian, MS, Dr. Richard Lewis and Professor Anthony Dayan for their critical reading of this chapter and their thoughtful comments.

## REFERENCES

1. Kuettner H. Verimpfung an stelle der transplantation hochwertiger organe. *Zentralblatt fuür Chirugie* 1912;1:390–7.
2. Sagen J. Cellular therapies for spinal cord injury: What will the FDA need to approve moving from the laboratory to the human? *J Rehab Res Dev* 2003;40: 71–80.
3. LeBlanc K, Pittenger MH. Mesenchymal stem cells: progress toward promise. *Cytotherapy* 2005;7:36–45.
4. Cell Therapy—Technologies, Markets and Companies. Strategic Report Cell Therapy Technologies, Markets and Companies. Jain PharmaBiotech, 2007;1–690

(http://pharmabiotech.ch/reports/celltherapy/) http://www.piribo.com/publications/drug_delivery/cell_therapy.html
5. Ginis I, Rao MS. Toward cell replacement therapy: promises and caveats. *Exp Neurol* 2003;184:61–77.
6. Honold J, Addmus B, Lehman R, Zeiher AM, Dimmeler S. Stem cell therapy of cardiac disease: An update. *Nephrol Dial Transplant* 2004;19:1673–7.
7. Lewis RM, Gordon DJ, Poole-Wilson PA, Borer JS, Zannad F. Similarities and differences in design considerations for cell therapy and pharmacologic cardiovascular trials. *Cardiology* 2008;110:73–80.
8. Boyle AJ, Schulman SP, Hare JM. Is stem cell therapy ready for patients? Stem cell therapy for cardiac repair ready for the next step. *Circulation* 2006;114:339–52.
9. Skyler JS. Cellular therapy for type 1 diabetes. Has the time come? *JAMA* 2007; 297:1599–600.
10. Burns CJ, Persaud SJ, Jones PM. Stem cell therapy for diabetes: So we need to make beta calls? *J Endocrinol* 2004;183:437–43.
11. Application of Current Statutory Authority to Human Somatic Cell Therapy Products and Gene Therapy Products. Federal Register 1993;58(197):53248–51.
12. Brady RP, Newberry MS, Girard VW. The Food and Drug Administration's Statutory and Regulatory Authority to Regulate Human Pluripotent Stem Cells. In Ethical Issues in Human Stem Cell Research, vol. 1. Report and Recommendations of the National Bioethics Advisory Commission, Appendix D, pp. 93–8.
13. Halme DG, Kessler DA. FDA regulation of stem-cell-based therapies. *N Engl J Med* 2006;355:1730–5.
14. Grimm EA, Mazumder A, Zhang HZ, Rosenberg SA. Lymphokine-activated killer cell phenomenon: lysis of natural killer resistant fresh solid tumor cells by interleukin-2 activated autologous human peripheral blood lymphocytes. *J Exp Med* 1982;155:1823–41.
15. Rosenberg S. Lymphokine activated killer cells: a new approach to the immunotherapy of cancer. *J Cancer Inst* 1985;75:595–603.
16. Lafreniere R, Rosenberg SA. Successful immunotherapy of murine experimental hepatic metastases with lymphokine-activated killer ceils and recombinant interleukin-2. *Cancer Res* 1985;45:3735–41.
17. Rosenberg SA, Lotze MT, Muul LM, Leitman S, Chang AE, Ettinghausen SE, Matory YL, Skibber JM, Shiloni E, Vetto JT, et al. Observations on the systemic administration of autologous lymphokine-activated killer cells and recombinant interleukin-2 to patients with metastatic cancer. *N Engl J Med* 1985;313:1485–92.
18. Bar MH, Snzol M, Atkins MB, et al. Metastatic malignant melanoma treated with combined bolus and continuous infusion interleukin-2 and lymphokine activated killer cells. *J Clin Oncol* 1990;8:1138–47.
19. Rosenberg SA, Spiess P, Lafreniere R. A new approach to the adoptive immunotherapy of cancer with tumor infiltrating lymphocytes. *Science* 1986;233:1318–21.
20. Kradin FL, Lazarus DS, Dubinett SM, et al. Tumor infiltrating lymphocytes and interleukin-2 in treatment of advanced cancer. *Lancet* 1989;2:577–80.
21. Celluzzi CM, Mayordomo JI, Storkus WJ, Lotze MT, Faolo LD, Jr. Peptide-pulsed dendritic cells induce antigen-specific CTL-mediated protective tumor immunity. *J Exp Med* 1996;183:283–7.

22. Sipuleucel-T, Dendreon (BLA-STN 125197) Indicated for the Treatment of Men with Asymptomatic Metatatic Hormone Refractory Prostate Cancer. Cellular, Tissue and Gene Therapies Advisory Committee Meeting March 29, 2007. http://www.fda.gov/ohrms/dockets/ac/cber07.htm#CellularTissueGeneTherapies
23. Berger JA, Spoo A, Dwenger A, Burger M, Behringer D. CSCR4 chemokine receptors (CD184) and alpha4beta1 integrins mediate spontaneous migration of human CD34+ progenitors and acute myeloid leukaemia cells beneath marrow stromal cells (pseudoemperipolesis). *Br J Haemat* 2003;122:579–89.
24. Bhatia M, Bonnet D, Murdoch B, Gan OI, Dick JE. A newly discovered class of hematopoitic cells with SCID-repopulating activity. *Nat Med* 1998;4:1038–45.
25. Gou Y, Lubbert M, Engelhardt M. CD34-Hematopoietic stem cells: current concepts and controversies. *Stem Cells* 2003;21:15–20.
26. Asahara T, Murohara T, Sullivan A, Silver M, Van der Zec R, Li T, Witzenbichler B, Schatteman G, Isner JM. Isolation of putative progenitor endothelial cells for angiogenesis. *Science* 1997;275:964–7.
27. Hristov M, Erl W, Weber PC. Endothelial progenitor cells: isolation and characterization. *Trends Cardiovas Med* 2003;13:201–6.
28. Urbich C, Dimmeler S. Endothelial progenitor cells. Characterization and role in vascular biology. Circulation *Res* 2004;95:343–53.
29. Werner N, Kosiol S, Schiegl T, Ahlers P, Walenta K, Link A, Bohm M, Nickenig G. Circulating endothelial progenitor cells and cardiovascular outcomes. *N Engl J Med* 2005;353:999–1007.
30. Beltrami AP, Barlucchi L, Torella D, Baker M, Limana F, Chimenti S, Kassahra H, Rota M, Musso E, Urbanek K, Leri A, Kajstura J, Nadal-Ginard B, Anversa P. Adult cardiac stem cells are multipotent and support myocardial regeneration. *Cell* 2003; 114:763–76.
31. Caplan AI. Mesenchymal stem cells. *J Orthopedic Res* 9:641–50.
32. Prockop DJ. Marrow stromal cells as stem cells for nonhematopoietic tissues. *Science* 1997;276:71–4.
33. Pittenger MF, Mackay AM, Beck SC, et al. Multilineaege potential of human mesenchymal stem cells for nonhematopoietic tissues. *Science* 1999;284:143–7.
34. Le Blanc K, Pittenger MF. Mesenchymal stem cells: progress toward promise. *Cytotherapy* 2005;7:36–45.
35. Thomson JA, Itskovitz-Eldor J, Shapiro SS, Waknitz MA, Swiergiel JJ, Marshall VS, Jones JM. Embryonic stem cell lines derived from human blastocysts. *Science* 1998; 282:1145–7.
36. Reubinoff BE, Pera MF, Fong CY, Trounson A, Bongso A. Embryonic stem cell lines from human blastocysts: somatic differentiation in vitro. *Nat Biotech* 2000; 18:399–404.
37. Klimanskaya I, Chung Y, Becker S, Lu S-J, Lanza R. Human embryonic stem cell lines derived from single blastomeres. *Nature* 2006;444:481–4.
38. Boyer LA, Lee TI, Cole MF, et al. Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell* 2005;122:947–56.
39. Adewumi O, Aflatoonian B, Ahrlund-Richter L, et al. Characterization of human embryonic stem cells by the International Stem Cell Initiative. *Nat Biotech* 2007; 25:803–16.

40. de Wert G, Mummery C. Human embryonic stem cells: research, ethics and policy. *Hum Reprod* 2003;18:672–82.
41. Schamblott MJ, Axelman J, Littlefield JW, Blumenthal PK, Huggins GR, Cui Y, Cheng L, Gearhart JD. Human embryonic germ cell derivatives express a broad range of developmentally distinct markers and proliferate extensively in vitro. *Proc Nat Acad Sci USA* 2001;98:113–8.
42. Shamblott MJ, Axelman J, Wang S, Bugg EM, Littlefield JW, Donovan PJ, Blumenthal PK, Huggins GR, Gearhart JD. Derivation of pluripotent stem cells form cultured human primordial germ cells. *Proc Nat Acad Sci USA* 1998;95:13726–31.
43. Strulovici Y, Leopold PL, O'Connor TP, Pergolizzi RG, Crystal RG. Human embryonic stem cells and gene therapy. *Mol Ther* 2007;15:850–66.
44. National Institutes of Health. Information on eligibility criteria for federal funding of research on human embryonic stem cells. Stem Cell Information, 2001. http://stemcells.nih.gov/policy
45. Hentze H, Graichen R, Colman A. Cell therapy and the safety of embryonic stem cell-derived grafts. Elsevier (www.sciencedirect.com), 2006. http://www.aseanbiotechnology.info/Abstract/21020806.pdf
46. Schletter J. Regulatory requirements for stem cell-based therapies. *Regul Affairs J* 2003;14.
47. Vogel G. Read or not? Human ES cells head toward the clinic. *Science* 2005; 308:1534–8.
48. Parson A. The long journey from stem cells to medical product. *Cell* 2006; 125:9–11.
49. Cavagnaro JA. Preclinical safety evaluation of biotechnology-derived pharmaceuticals. *Nat Rev Drug Discov* 2002;1:469–75.
50. ICH S6 Preclinical Safety Evaluation of Biotechnology-derived Pharmaceuticals. http://www.fda.gov/cder/guidance/index.htm
51. Guidance for Industry. Guidance for Human Somatic Cell Therapy and Gene Therapy, March 1998. http://www.fda.gov/cber/guidelines.htm
52. CHMP Guideline on Human Cell-Based Medicinal Products EMEA/CHMP/410869/206, January 2007. http://www.emea.eu.int
53. Horan PK, Melnicoff MJ, Jensen BD, Slezak SE. Fluorescent cell labeling for in vivo and in vitro cell tracking. Methods in *Cell Biol* 1990;33:469–90.
54. Fukuhara S, Tomita S, Nakatani T, et al. Comparison of cell labeling procedures for bone marrow cell transplantation to treat heart failure: long-term quantitative analysis. *Transplant Proc* 2002;34:2718–21.
55. Lyons AB. Analyzing cell division in vivo and in vitro using flow cytometric measurement of CFSE dye dilution. *J Immunol Meth* 2000;243:147–54.
56. Edinger M, Cao Y-A, Verneris MR, Bachman MH, Contag CH, Negrin RS. Revealing lymphoma growth and the efficacy of immune cell therapies using in vivo bioluminescence imaging. *Blood* 2003;1010:640–8.
57. Tohill MP, Mann DJ, Mantovani CM, Wiberg M, Terenghi G. Green fluorescent protein is a stable morphological marker for Schwann cell transplants in bioengineered nerve conduits. *Tissue Eng* 2004;10:1359–67.
58. Brazelton TR, Blau HM. Optimizing techniques for tracking transplanted stem cells in vivo. *Stem Cells* 2005;23:1251–65.

59. Shen LH, Li Y, Chen J, Cui Y, Zhang C, Kapke A, Lu M, Savant-Bhonsale S, Chopp M. One-year follow-up after bone marrow stromal cell treatment in middle-aged female rats with stroke. *Stroke* 2007;38:2150–6.
60. Boersma HH, Tromp SC, Hofstra L. Stem cell tracking: reversing the silence of the lamb. *J Nucl Med* 2005;46:200–3.
61. Rogers WJ, Meyer CH, Kramer CM. Technology insight: in vivo cell tracking by use of MRI. *Nat Clin Pract Cardiovas Med* 2006;3:554–62.
62. Jendelova P, Herynek V, Urdzikova L, et al. Magnetic resonance tracking of transplanted bone marrow and embryonic stem cells labeled by iron oxide nanoparitcles in rat brain and spinal cord. *J Neurosci Res* 2004;76:232–43.
63. Bulte JWM, Zhang S-C, van Gelderen P, Herynek V, Jordan EK, Duncan ID, Frank JA. Neurotransplantation of magnetically labeled oligodendrocyte progenitors: magnetic resonance tracking of cell migration and myelination. *Proc Nat Acad Sci USA* 1999;96:15256–61.
64. Barnett BP, Arepally A, Karmarkar PV, Qian D, Gilson WD, Walcak P, Howland V, Lawler L, Lauzon C, Stuber M, Kraitchman DL, Bulte JWM. Magnetic resonance-guided, real-time targeted delivery and imaging of magnetocapsules immunoprotecting pancreatic islet cells. *Nat Med* 2007;13:986–91.
65. Ma B, Hankenson K, Dennis J, Caplan A, Goldstein S, Kilbourn M. A simple method for stem cell labeling with fluorine 18. *Nucl Med Biol* 2005;32:701–5.
66. NIH BLAST. http://0-www.ncbi.nlm.nih.gov.catalog.llu.edu/BLAST/
67. Vogel G. Ready or not? Human ES cells head toward the clinic. *Science* 2005; 308:1534–8.
68. Brickman JM, Burdon TG. Pluripotency and tumorigenicity. *Nat Genet* 2002;32: 557–8.
69. Rubio D, G-C J, Martin MC, de la Fuente R, Cigudosa JC, Lloyd AC, Bernad A. Spontaneous human adult stem cell transformation. *Cancer Res* 2005;65:3035–9.

# Preclinical Safety Evaluation of Biopharmaceuticals: Combination Products (Biologic/Device)

BRUCE BABBITT, PhD, and BARRY SALL

**Contents**

1 Introduction
2 Definition of Combination Products
3 General Points Applicable to All Device-Biologic Combination Products
  3.1 Functionality of the Biologic
  3.2 Functionality of the Device
  3.3 Properties of the Combined Product
  3.4 Regulatory Issues
4 Model Combination Product: Antibody-Coated Stent
5 Preclinical Testing Considerations
  5.1 Biologic Orientation on the Device
  5.2 Potential Elution (Biologic) Species
6 Pharmacological/Toxicological Assessment
  6.1 Definition of New Terms
  6.2 Choice of Animal Species/Pharmacology Studies
  6.3 Toxicology Testing
7 Device Considerations
8 Regulatory Considerations
  References

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

## 1 INTRODUCTION

Combination products include a wide variety of products that merge medical devices, pharmaceuticals, and/or biologics. Combinations can be physical, such as a drug-coated coronary stent [1,2] where a drug is applied to the implantable device, or they can involve specialized delivery devices for either pharmaceuticals or biologics where contact between the device and biologic is brief. Such delivery devices can include intravascular catheters, pen injectors, or even nebulizers and inhalers. In some of these cases the linkage between the biologic and the device may be through labeling, rather than physical contact. As these examples indicate, the range of combination products is very wide, involving a nearly limitless number of technical, clinical, and regulatory challenges. New technologies such as tissue-engineering (see Chapter Tissue Engineered Products: Preclinical Development of Neo-Organs) and nanotechnology are already in development, and they will add to this complexity as they move closer to the clinic.

The preclinical testing program for a biologic-device combination product should adequately address safety issues associated with device components and those that involve the biologic. However, these two topics can not be considered in isolation. The device component and associated manufacturing processes have the potential to affect the biologic, and vice versa. This interaction forces those responsible for developing the preclinical testing program to utilize a multidisciplinary approach [3]. One of the many challenges involved in this effort is the fact that the technical people are usually familiar with either device or biologic testing, but not both. Combining ICH mandated testing requirements with ISO 10993 testing requirements and adding other specialized testing in a cost-effective and timely manner is the challenge that this chapter will examine.

This chapter discusses the key issues that should be considered when preparing a preclinical development plan for a combination product that includes components generally regulated as medical devices and biologics. It is intended to provide the reader with examples of major elements of a preclinical development plan along with a discussion of the rationale for selecting particular options. Key issues that should be addressed for any biologic-device combination product are discussed in general terms. Then a strategy for a specific hypothetical product is explored in detail.

## 2 DEFINITION OF COMBINATION PRODUCTS

The FDA defines combination products in 21 CFR 3. Four types of combination products are described:

| TYPES OF COMBINATION PRODUCTS | BIOLOGIC/DEVICE EXAMPLE |
|---|---|
| A product comprised of two or more regulated components; i.e., drug/device, biologic/device, drug/biologic, or drug/device/biologic, that are physically, chemically, or otherwise combined or mixed and produced as a single entity | Antibody coated coronary artery stent |
| Two or more separate products packaged together in a single package or as a unit and comprised of drug and device products, device and biological products, or biological and drug products | BMP putty and spinal cage |
| A drug, device, or biological product packaged separately that according to its investigational plan or proposed labeling is intended for use only with an approved individually specified drug, device, or biological product where both are required to achieve the intended use, indication, or effect and where upon approval of the proposed product the labeling of the approved product would need to be changed; e.g., to reflect a change in intended use, dosage form, strength, route of administration, or significant change in dose | A novel biologic administered using a 510(k) cleared pen injector |
| Any investigational drug, device, or biological product packaged separately that according to its proposed labeling is for use only with another individually specified investigational drug, device, or biological product where both are required to achieve the intended use, indication, or effect | Gene therapy agent and delivery catheter |

Combinations of two biologics or two devices are not considered combination products in this context.

Once a Sponsor determines that their product fits one of the definitions of a combination product, the sponsor then must determine the lead Center within the FDA that will assume regulatory responsibility for the product. This determination is made by considering the primary mode of action (PMOA) [4] of the product. In simple terms, if the main function of the product is a device function, then it is likely that the Center for Devices and Radiological Health (CDRH) will have primary responsibility for the product. On the other hand, either the Center for Biologics Evaluation (CBER) or the Center for Drug Evaluation and Research (CDER) will be assigned primary review responsibility based on whether the PMOA of the product is dependent on

the pharmacological effects of a synthetic drug or a well-characterized biologic (CDER) versus other more complex biologics (CBER). Responsibility means that the Center managing the regulatory review uses its regulatory mechanisms to authorize the clinical trial and, ultimately, commercialization of the product. If CDRH is responsible for the product, it would follow the Investigational Device Exemption (IDE) and PreMarket Approval (PMA) pathways. If CBER is the lead Center, an Investigational New Drug Application (IND) and Biologic License Application (BLA) would be necessary, whereas if CDER has primary responsibility, then an IND and New Drug Application (NDA) would be the regulatory pathway. In many cases, precedents created with other similar products will provide clear direction as to the responsible Center. In other cases, jurisdiction is less clear such that the Sponsor may choose to submit a Request for Designation (RFD) [5] to the Office of Combination Products (OCP). The RFD contains a description of the product, available data on its use and mechanism of action, as well as a suggested product designation. OCP will review the RFD and inform the Sponsor of their decision. Some redacted decision letters are available at the OCP Web site (http://www.fda.gov/oc/combination/rfd.html). Once the responsible Center is identified, the Sponsor can then begin to assess preclinical data requirements and determine the scheduling for the necessary studies.

## 3 GENERAL POINTS APPLICABLE TO ALL DEVICE-BIOLOGIC COMBINATION PRODUCTS

Developers of combination products should rigorously evaluate their product in order to construct a preclinical testing plan that is scientifically and medically sound and take into account business objectives. Often biologic-device combination product development efforts involve collaborations between a biotech company and a more traditional device company. Each organization brings its own technical and business expertise and expectations to the project. Biologic developers are accustomed to designing comprehensive, multi-species preclinical testing programs that by themselves may take two or more years to complete. This is in large contrast to many device companies whose entire product life cycle, from development through end-of-commercial sales, typically takes little more than two years. Furthermore, preclinical testing for biologic products follows CBER/CDER and ICH guidances that are quite different from the medical device ISO 10993 standard. Therefore, the preclinical development team for a biologic-device combination product needs to address "cultural" as well as technical issues. The following sections describe in general terms the broad issues that should be considered when designing the preclinical development program for a biologic-device combination product.

### 3.1 Functionality of the Biologic

Whether a biologic is directly physically associated with the surface of a device (via some type of chemical cross-linking or simple adsorption), or alternatively,

embedded in some type of matrix that is used to coat the device, it is very likely that the pharmacology/toxicology profile of the biologic will be altered when compared to administration of the biologic as a single agent.

If the physical association of the biologic with the device causes even small changes in the structure of the molecule (either while bound to the device or upon release into the local environment), then it is possible to decrease or even eliminate its intrinsic receptor binding activity and downstream physiological effects. Alternatively, the possibility exists to increase the overall binding affinity and resultant pharmacological and/or toxicological effects of the biologic, in vivo, if multivalent binding (e.g., of an antibody to its cognate antigen) results from the presence of high-density clusters of the molecule on the surface of the device.

A major challenge for both developers and regulators of biologic/device combination products is the potential need to re-define certain key parameters, for example, "dose"/exposure, distribution, and persistence, that have historically formed the basis for a comprehensive preclinical characterization of the pharmacology/toxicology profile of stand-alone biologics. As is the case with most biological drug products, while the basic principles of toxicological assessment are much the same, actual practices may vary from the traditional drug approach. It is likely in certain cases that new animal models or analytical methods will be needed in order to adequately define potential human toxicities and/or safe starting doses and an appropriate dose escalation scheme.

### 3.2 Functionality of the Device

If the device component of the biologic/device combination product has already been cleared or approved as a stand-alone product, some aspects of device functionality will be known. However, the effect of the biologic will need to be investigated. In the case where the device and biologic are physically associated with each other, the process used to join the two components can affect device performance in a variety of ways. Easily determined parameters such as dimensions and clearances must be confirmed. More subtle issues arise when physical performance and durability are assessed. Additional manufacturing procedures associated with the antibody coating could, at least in theory, alter the properties of the entire device. In some cases, test methods that have been successfully employed for stand-alone devices may need to be modified to accurately assess all relevant parameters that must be considered for the combination product.

### 3.3 Properties of the Combined Product

While it is important to evaluate the component parts of the combination product, the fully integrated finished product(s) must be evaluated as well.

Devices that have long histories of reliability as stand-alone products may behave differently when associated with a biologic. Conversely, when a biologic is associated with a device, a variety of chemical, physical, or conformational modifications could occur and alter its biologic activity in some manner. In the preclinical development phase, proof-of-concept (POC) testing is used to make a preliminary efficacy determination and is often designed to generate safety data.

Combining the device and biologic will almost certainly affect stability. Packaging systems may frequently be altered in significant ways. The form of the biologic may also change from, for example, a lyophilized powder to a form bound to a solid surface or entrapped in a matrix. Furthermore, the manufacturing methods used to combine the two components may significantly affect the biologic's stability. In any case, stability studies should be conducted with product that has undergone the full manufacturing, packaging, and sterilization processes.

Sterilization methods should be carefully considered. Many medical devices undergo terminal ethylene oxide (EtO) or radiation sterilization. Both of these processes can alter a biologic component, either through chemical modification or radiolysis. Whatever sterilization method or bioburden limit is selected, it must balance the competing needs of product safety and product integrity. Preclinical testing will need to utilize a product that has undergone all processing steps, including sterilization and packaging, to ensure that preclinical testing uses test articles that have been exposed to all the product stresses encountered in the commercial manufacturing process.

### 3.4 Regulatory Issues

The overall regulatory approach for a combination product is governed by several considerations. First, one must consider which Center has primary jurisdiction for the product and if the other Center will participate in a collaborative or consultative role. Once these parameters are defined, the Sponsor will know the type of submissions necessary to gain approval for the clinical trial(s), and ultimately, for the marketing application. Second, if either or both of the components (biologic and device) are already approved in the case of the biologic or cleared in the case of the device, the Sponsor should obtain authorization to cross-reference data in those submissions and determine which data are relevant to the combination product. As discussed in the following specific example (see below), not all data from component parts are relevant to the combination product. Several FDA guidance documents [6,7] should be reviewed when preparing the initial regulatory evaluation. Last, regulatory staff should contact the FDA, either the responsible Center or OCP, early in the process and confirm that their plans are consistent with FDA expectations. Such contacts, while important, are not a substitute for a formal pre-IDE or a type B pre-IND meeting [8,9] (CDRH types of meetings), but they do address key questions early in the development process.

## 4 MODEL COMBINATION PRODUCT: ANTIBODY-COATED STENT

In order to present clearly and specifically the unique preclinical testing challenges associated with the early-stage development of biologic/device combination products, an antibody-coated stent will be used as a model. For this particular case study the antibody chosen for manufacturing has either not yet been approved for use in humans or, if it has, not for the newly targeted indications. In addition, the combination product utilizes a previously approved stent and delivery system.

The rationale for selection of a monoclonal antibody (mAb) as the biologic component of our model biologic-device combination product is based primarily on the following two major factors: (1) there seems to be a high degree of interest in these next-generation type of products due to the potential for very high affinity interactions with specific targets (antigens), in vivo, with the potential for a much more favorable risk–benefit ratio of the therapy as compared to the use of more conventional "nonspecific"drugs, and (2) antibodies represent a very well-characterized class of biologics from a manufacturing, preclinical, clinical, and regulatory standpoint.

Another reason for focusing on antibody-coated stents throughout the remainder of this chapter is that discussion of preclinical strategy covering the very broad range of biologic-device combination products currently under development would, by definition, require both the writers and readers to have a solid working knowledge of the basic science aspects of a broad array of biologics (living cells, genes, etc.). (See Chapters Preclinical Safety Evaluation of Biopharmaceuticals and Considerations in Design of Preclinical Safety Evaluation Programs to Support Human Cell-Based Therapies for a discussion of specific considerations for developing gene and cellular therapies, respectively.)

## 5 PRECLINICAL TESTING CONSIDERATIONS

One of the first decisions to be made in order to formulate an adequate preclinical testing strategy for biologic-device combination products is to determine a relevant form(s) of the product to test in selected animal species for assessment of potential pharmacological and toxicological effects in humans. In general, one can approach an analysis of the potential in vivo effects of a device-biologic combination product from the following two mechanistically distinct standpoints: (1) local effects arising from the direct interaction of living cells and/or soluble factors with the device or the antibodies bound to the surface of the device, and (2) local or systemic effects arising from the interaction of living cells or soluble factors with any of a variety of antibody-containing species capable of eluting off of the surface of the implanted device. In addition, it may be necessary to consider evaluation of the mAb alone in addition to assessing the combination product, based on intrinsic attributes of the mAb as well as the availability of relevant preclinical and clinical safety and efficacy data.

Finally, a major challenge for both developers and regulators of biologic-device combination products is the potential need to reconsider the way that certain key parameters typically associated with preclinical assessment, such as dose in the case of the biologic and durability in the case of the device, are defined, measured, and analyzed.

## 5.1 Biologic Orientation on the Device

In antibody-coated stents it is possible for the heterogeneity in the overall orientation of the antibody molecule on the surface of the device to lead to different pharmacological and/or toxicological effects, in vivo. The degree of heterogeneity is likely to be strongly influenced both by the intrinsic nature of the antibody (primary, secondary, and tertiary structure) and by the particular method (adsorption, chemical cross-linking, etc.) selected for binding the selected antibody to the device.

For example, if a high percentage of the device-associated antibody molecules are attached "backward" to the surface of the device—that is, with their antigen binding domains (Fab) facing inward and their $F_c$ domains facing outward—then there is a relatively high likelihood of local interactions of the immobilized antibodies with circulating immune cells expressing high levels of $F_c$ receptors (Figure 1). It is the responsibility of the manufacturing team

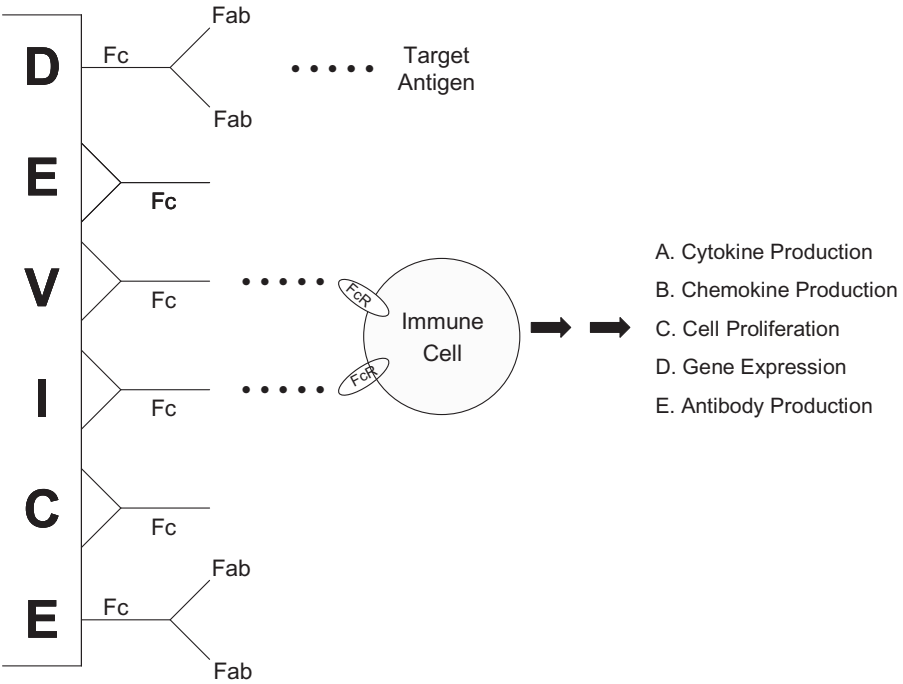


**Figure 1** Antibody orientation on the surface of the device.

to fully characterize and control (minimize) this potential variability in the orientation of the biologic prior to the start of preclinical testing.

Once immune cells become bound to the surface of the antibody-coated stent via multivalent interactions with exposed $F_c$ domains, there is the potential for a variety of immunomodulatory effects (cytokine release, inhibition of normal $F_c$ receptor mediated events, etc.) that could lead to the generation of adverse events, in vivo.

As a worst-case scenario the possibility exists to break natural tolerance to human $F_c$ and lead to the formation of anti-$F_c$ antibodies, in vivo, due to the "presentation" of such a high local concentration of $F_c$ domain epitopes to various components of the cellular immune system. As a result, the generation and function of endogenous immunoglobulins could be negatively affected by the presence of these circulating antibodies (immunosuppression).

In addition, although the typical antibody-coated device will likely be designed to allow rapid elution of the drug into the local microenvironment over a relatively short period of time (hours to days), the short-term ability of the device to interact with target antigens present in the local milieu will be negatively affected by the presence of improperly oriented antibody molecules on the device surface.

### 5.2 Potential Elution (Biologic) Species

It is straightforward that the final implantable combination product, as a finished product, needs to be assessed for both potential device-related and biologic-related pharmacological effects and potential toxicities in one or more animal species. However, once the biologic component of the combination product starts to elute off of the device (again, focusing on antibody-coated stents as a model, and assuming that the intention of the product is to "deliver" the biologic to the local microenvironment), attention must be paid to the changing nature of the implanted biologic/device and the appearance over time of various antibody-containing species (cells and/or soluble factors) first in the microenvironment around the device, and thereafter in the general circulation.

In terms of antibody-coated stents, the following is a list of potential antibody-containing species that could, in theory, be eluted directly from the surface of the implanted biologic/device combination product, in vivo (Figure 2):

- Monomeric antibody molecules (native conformation)
- Monomeric antibody (altered conformation)
- Aggregated antibodies (native or altered conformation)
- Degradation products (antibody derived peptides)
- Antibody–antigen complexes
- Antibody–cell complexes

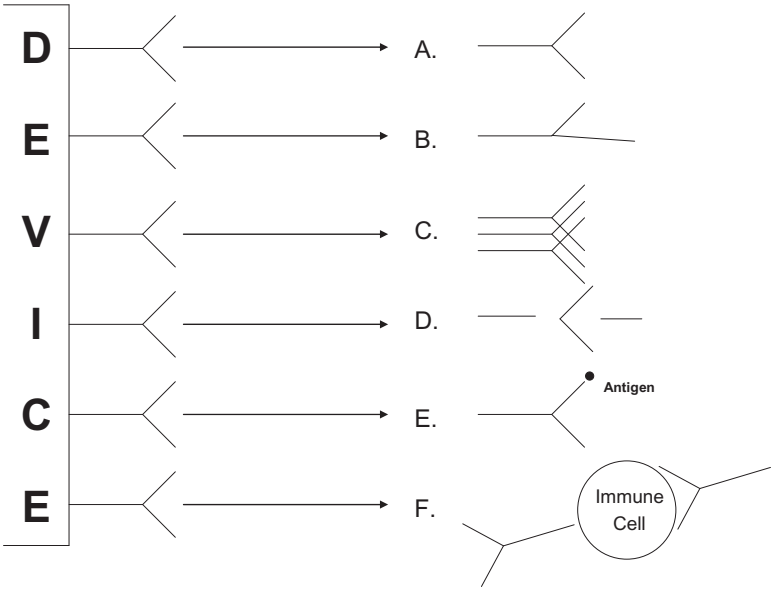


**Figure 2** Antibody-containing elution species.

Again, the combination device would be optimized during the manufacturing process to ensure a consistent antibody coupling and control for the various species. In order to determine which, if any, of the antibody-containing elution species have the potential to exert either pharmacological or toxic effects in humans, the various species could be generated and tested, ex vivo, via extended exposure of the combination product to various biological fluids (e.g., whole blood, serum, plasma), and then to assess potential effects of the "supernatant" from these experiments in at least one animal species. If any meaningful effects are uncovered, then studies could be designed to determine which molecular entity is responsible for the observed effects.

This experimental approach has limitations in that it ignores the potential critical interactions of immune cells and other living cells, in vivo, with the antibody-coated surface of the device. These cellular "interactions" can occur either via direct binding to the device surface of cells expressing receptors (e.g., cognate antigen) capable of recognizing various antibody domains, or alternatively, via secretion of cellular enzymes capable of directly modifying these antibodies. As such, further in vivo studies would be needed to characterize potential toxicities.

## 6 PHARMACOLOGICAL / TOXICOLOGICAL ASSESSMENT

Based on the inherent complexity of biologic/device combination products, together with the assumption that the potential in vivo adverse effects of the

combination is not simply equal to the sum of what is known regarding their component parts, it is clear that the design of an adequate program for the preclinical characterization of a biologic-device combination product will need to be approached on a case-by-case basis. In theory, the potential adverse effects of the combination product can be derived from in vivo interactions with the implanted biologic device, as designed, the biologic device modified by local physiological activities, or any of a broad array of molecular (or cellular) species eluting off of the surface of the device over time capable of exerting either local or systemic effects (see Section 5.2 above).

Another key factor in defining an adequate preclinical testing program for these products is whether a biopharmaceutical company chooses to use a biologic (antibody in this case) that has already received FDA approval for use in humans, even though it is likely not for the same indication targeted by the stent manufacturer. Use of an approved biologic with the device may provide information as to what to expect under a worst-case scenario, such as all of the biologic being released immediately into the systemic circulation. However, it is still very important to understand what impact attaching the mAb to the surface of a given medical device and implanting the combination product into a patient (route of administration) has on potential toxicity. For example, release of the mAb directly into the coronary artery might be expected to result in a different PK/PD profile than a direct IV or IM injection of the biologic alone. In addition, if the mAb has not previously been administered to the same patient population that was previously treated with the stand-alone medical device, potential differences between the different patient populations must also be very carefully considered.

### 6.1 Definition of New Terms

A major challenge for both developers and regulators of biologic-device combination products is the potential need to reconsider how to measure certain key parameters typically associated with the preclinical assessment of stand-alone medical devices and biologics with the goal toward extrapolation of the information to humans.

These parameters include, but are not limited to, establishment of a minimal effective dose (MED), a no observable dose (NOEL), a no observable adverse effect dose (NOAEL), and a maximum tolerated dose (MTD). From the standpoint of our antibody-coated stent model, and antibody "dose," it is possible that the product (device) labeling could limit the dose of mAb a patient is allowed to receive to how much protein is able to be coated onto just one stent. Alternatively, since stents are made in various lengths in order to match the length of the lesion, the longer the stent, the more antibodies that can be delivered.

Issues related to PK, distribution, persistence, and stability will also need to be considered. Optimizing methods to assess local effects as well as potential changes in the immunogenicity of the biologic component of the

combination product may also be needed. Careful consideration must be given to identification of the most relevant animal models, including use of an animal model of disease to better assess toxicity.

Quantitative assessment of these key parameters could require the development of new analytical methods compared to the relatively well established methods currently in use for characterization of stand alone medical devices and biologics.

## 6.2 Choice of Animal Species/Pharmacology Studies

Given that the various animal species and experimental protocols typically used for preclinical testing of conventional devices versus biologics are quite different, a major challenge for developers of biologic-device combination products is to determine the best way to combine the essential elements of these studies into a manageable number of predictive models for their particular product. A very basic aspect of this challenge is that, in general, animal species used for preclinical testing of implantable devices (especially proof-of-concept (POC) type of studies) tend to consist of larger animals due to their greater compatibility with surgical procedures. For example, porcine models have been used extensively for testing of drug-coated stents [1,2]. For biologics such as mAbs the most important product attribute for defining a relevant preclinical species is whether the target antigen or epitope is present in sufficient amounts in the chosen animals. In some cases a species may be selected for its similarity to humans based on its potential off-target binding to tissues as demonstrated by tissue cross-reactivity studies.

It is likely that manufacturers of devices (e.g., stents) will strongly prefer testing their antibody-coated products in the established porcine models for POC type of studies in order to make critical go or no go decisions regarding choice of antibody besides proceeding into toxicology studies and then human testing. If so, developers will need to screen their antibodies for their ability to bind to porcine tissues in order to perform POC studies without any complications (e.g., immunogenicity due to species-specific effects) that would compromise their ability to evaluate the results of the study. In certain cases companies might be faced with having to use immunosuppressed animals or a porcine homologue of the antibody (surrogate antibody) in order to carry out "gold standard" POC studies, Alternatively, development of a different animal model that will bind the antibody and allow implantation of the stent may be needed as well. A practical challenge in this area might be that contract research organizations experienced in performing the POC type studies with devices may not have experience with biological products and associated assays, and vice versa.

From a pharmacological standpoint, monomeric antibodies in their native conformation (whether free, cell-bound, or still attached to the device) are most likely to maintain their intrinsic ability to bind stably to their target antigens, in vivo, leading to the desired therapeutic effect. However, while

device-associated or locally released mAbs would be expected to exert mostly local pharmacological or toxicological effects, cell-bound antibodies have the potential to exert more systemic effects as a result of the trafficking of immune cells to or from the site of the implanted device. In addition, it is possible that target antigens bound to antibodies on the surface of various immune cell subsets would not be cleared via the same pathways typically used to clear more conventional antibody–antigen complexes. These potential concerns would need to be addressed in the preclinical safety evaluation program for the biologic-device combination product.

### 6.3 Toxicology Testing

***General Biologic/Device Considerations*** The following is a list of potential unique considerations for assessing the safety of biologic-device combination products:

1. As mentioned above, the types of animal species typically used to assess the toxicology profile of the mAb may not be the same species typically chosen for ISO 10993 device safety testing.
2. Toxicology testing for devices typically focuses greatly on local effects of the product, while there is always a major focus on potential systemic effects with biologics.
3. The issue of species-specific interactions is unique to biologics and complicates the design, performance, and interpretation of biologic-device pharm/tox study results.
4. Patients have much shorter exposures to most biologics (e.g., months, intermittent) compared to conventional devices (e.g., chronic, lifetime).
5. Compared to the majority of biologics, which are typically fully catabolized in vivo into amino acids over the short-term (hours, days), devices are expected to change very little in vivo even after long-term exposure to most biological fluids.
6. The potential for immune responses against biologics (hypersensitivity, anti-drug antibodies, immune complexes, etc.) is something that most stent (alone or as drug-coated stents) manufacturers have not had to consider in prior development programs.
7. ISO 10993 testing is intended to evaluate materials such as metals and polymers that essentially do not interact directly with the human immune system. However, the presence of an antibody on the device surface introduces the potential for cross-species reactions that may not be indicative of performance in humans. These interactions may affect some or all ISO 10993 in vivo tests. For example, the duration of chronic toxicity testing may be limited based on induction of immune responses or carcinogenicity studies may not be appropriate based on the specific product attributes of the antibody.

8. The biologic could be attached to the surface of the device using a broad array of materials and coupling techniques. Therefore, in addition to the conventional device-based biocompatibility testing for coating materials, close attention must be paid to potential in vivo effects derived from the presence of biologic-device attachment factor complexes or derivatives that could be generated following implantation of the device in humans.

***Considerations Unique to Antibody-Coated Stents*** From a toxicological standpoint, the release of aggregated antibodies (native or altered conformation) from the surface of an antibody-coated stent into the microenvironment or systemic circulation has the potential to induce an array of immunomodulatory effects via direct interaction with several key components of the human cellular immune system (dendritic cells, macrophages, neutrophils, B cells, NK cells, etc.). For example, aggregated antibodies have the potential to bind with high affinity and cross-link $F_c$ receptors on the surface of various immune cells resulting in (1) overproduction of proinflammatory cytokines and chemokines, (2) increase in the level of tyrosine phosphorylation of intracellular proteins, (3) induction of new gene expression, (4) increase in phagocytic activity, (5) increase in antigen presentation, (6) increase in cell cycle progression (cell proliferation), and (7) increase in the production and release of reactive oxygen species. These predominantly immunostimulatory effects can often lead to toxic reactions (adverse events) in humans.

In addition, since it is now well established that aggregated proteins are much more immunogenic in animals and in humans than native monomeric versions of the same proteins, the potential exists for aggregated antibodies released into the circulation from the surface of antibody-coated stents to induce human antihuman antibody (HAHA) responses with the theoretical potential to cross-react against endogenous related immunoglobulins. The local release from the surface of the implanted biologic/device of pre-formed antibody–antigen complexes, or alternatively, much larger immune cell–antibody complexes, followed by deposition of these complexes into the nearby vasculature could lead to inflammation and subsequent damage to blood vessels, tissues, and joints in close proximity to the combination product. It is less likely, but still possible, that these immune complexes could also exert negative effects, in vivo, via systemic access to the kidneys, lungs, skin, and other organs.

Since all of the potential immunotoxicological effects mentioned above related to the use of antibody-coated stents are typically highly concentration dependent, it is possible that the use of only very small amounts of protein on the surface of the device to exert a desired localized therapeutic effect, will mitigate, but certainly not eliminate these potential toxicities. Nevertheless, the purpose of the preclinical toxicological evaluation will be to address the potential impact of these theoretical concerns.

## 7 DEVICE CONSIDERATIONS

Assessment of biologic/device combination product functionality needs to include all of the testing typically performed on bare metal stents (BMS). In the case of our antibody-coated stent model the BMS have been previously approved by CDRH, allowing some of the BMS data to be cross-referenced and not repeated when the effect of the antibody coating on the functionality of the device is considered negligible. These judgments must be made on a case-by-case basis and should be confirmed with the Office for Device Evaluation.

The finished, packaged and sterilized product must be considered as a whole. In accordance with the Design Control provisions of the Quality System Regulation (21 CFR 820.30) medical devices undergo a formal risk analysis throughout product development. Such an analysis identifies potential risks inherent in the design, materials, and method of use of the device ranks the risks and ensures that unacceptable risks are mitigated. The original risk analysis for the BMS must be re-evaluated to consider the addition of the antibody coating. The risk analysis should consider the physical effects of the coating, packaging modifications, or manufacturing process modifications as well as the physiological and immunological effects discussed above. Considerable effort will be directed to the physiological and immunological effects, but developers should not lose sight of some of the more practical concerns, such as deliverability of the stent and the effect of packaging modifications on handling of entire system. Experts with a good understanding of these technical areas must be brought into the risk analysis process. Human factor issues should also be addressed, especially if there have been any modifications to the instructions for use for the delivery system.

## 8 REGULATORY CONSIDERATIONS

As discussed, successful development of a combination product will require a thorough understanding of the requirements for the individual components as well as an understanding of what questions need to be asked based on the specific product attributes when the individual components are combined. It will be important to determine early in development which FDA Center will have primary responsibility for the combination product as that will likely influence the development strategy. As the products will likely be unique with little regulatory precedent, it will be important for Sponsors to communicate with the FDA, again early in the development process, to ensure that the program addresses all the important issues necessary for introducing the product into humans for the first time.

## REFERENCES

1. J&J Cypher SSE. http://www.fda.gov/cdrh/PDF2/P020026b.PDF
2. BSC Taxus SSE. http://www.fda.gov/cdrh/pdf3/P030025b.pdf
3. Getting Started with a Combination Product. Part I: Overcoming the Regulatory Challenges in This Ever-Expanding Field Yields Rewards. http://www.devicelink.com/mddi/archive/03/03/018.html
   Getting Started with a Combination Product. Part II: European Regulations. The European Union adds further wrinkles to the combination product market. http://www.devicelink.com/mddi/archive/03/04/019.html
4. Definition of Primary Mode of Action of a Combination Product. http://www.fda.gov/OHRMS/DOCKETS/98fr/05-16527.htm
5. Guidance for Industry and FDA Staff: How to Write a Request for Designation (RFD). http://www.fda.gov/oc/combination/Guidance-How%20to%20Write%20an%20RFD.pdf
6. Guidance for Industry and FDA Staff: Early Development Considerations for Innovative Combination Products. http://www.fda.gov/oc/combination/innovative.pdf
7. Guidance for Industry and FDA Current Good Manufacturing Practice for Combination Products. http://www.fda.gov/oc/combination/OCLove1dft.pdf
8. Early Collaboration Meetings under the FDA Modernization Act (FDAMA); Final Guidance for Industry and for CDRH Staff. http://www.fda.gov/cdrh/ode/guidance/310.pdf
9. Guidance for Industry. Formal Meetings with Sponsors and Applicants for PDUFA Products. http://www.fda.gov/cber/gdlns/mtpdufa.htm

# Tissue Engineered Products: Preclinical Development of Neo-Organs

TIMOTHY A. BERTRAM, DVM, PhD, and MANUEL JAYO, DVM, PhD

**Contents**

1 Introduction
2 Medical Needs and Challenges
3 Regulatory History/Development
4 Establishing a Regulatory Pathway for Neo-Organ
5 Development of Neo-Organ Products
6 Neo-Organ Products
  6.1 Development Strategies
  6.2 Raw Material Testing
  6.3 Final Neo-Organ Product Testing
7 Transition to Clinical Testing
  7.1 Defining and Testing a Prototype
  7.2 Production of Neo-Organs for Preclinical Testing
8 Bringing Neo-Organ Platforms to Clinical Testing
9 Conclusions
  References

## 1 INTRODUCTION

In the early 1930s during his tenure at Rockefeller University, Charles Lindbergh, the aviator, first contemplated preclinical development of manufactured human organs (neo-organs) for implantation into the body [1]. However, it took 50 years of scientific research and technology advancements

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

for the idea of augmenting or replacing diseased organs with laboratory-grown components to emerge into the realm of possibility. Today technologies have matured to the point that a diseased or missing organ can be regenerated de novo with a tissue-engineered neo-organ.

The term *tissue engineering* was coined at a National Science Foundation workshop in 1987 to mean "the application of principles and methods of engineering and life sciences toward fundamental understanding of structure-function relationships in normal and pathological mammalian tissues and the development of biological substitutes to restore, maintain or improve tissue function" [2]. Tissue engineering draws on specialized expertise from two traditional disciplines: engineering and the life sciences. The combination of these technologies forms a foundation upon which the commercial development of neo-organs is possible.

Preclinical development of neo-organs faces complex scientific questions and regulatory hurdles. The term *neo-organ product* will be used to indicate any product composed of synthetic or natural biodegradable materials, with or without living cells and/or cellular products, implanted in the body to incorporate, replace, and/or regenerate a damaged tissue or organ. Today ex vivo development of partial and complete neo-organs by the emerging regenerative medical industry fulfills a significant unmet medical need for patients who have partial or complete organ or tissue loss. In this chapter we will consider development challenges and available solution strategies for bringing these transformational technologies to patients in need.

## 2 MEDICAL NEEDS AND CHALLENGES

Currently a patient's diseased organ replacement requires the transplantation of whole organs (e.g., kidneys, livers, hearts, lungs, and pancreases) donated or taken from another body system or living being (human–autologous/allogeneic or animal–xenogeneic). Regardless of source, organs are in finite supply, and complications such as chronic metabolic alterations, immunologic rejection, and product shelflife have rendered this therapeutic approach both costly and risky. Furthermore patients who receive nonautologous organ transplants require life-long chemical immunosuppression or face partial or complete organ rejection.

Recent scientific advances involving tissue engineering technologies and their use in regenerative medicine have helped us better understand host tissue responses and basic needs during healing, repair, and regeneration. With these advances comes a need to establish product development pathways that fit into existing and emerging regulatory guidelines while remaining commercially viable. Production of a neo-organ requires defined raw materials and a reproducible manufacturing process whereby the appropriate constituents (i.e., cells and biomaterials) can be cultured and combined ex vivo to form an implantable neo-organ product. Following implantation, both partial to

complete neo-organs must become vascularized by host tissues, integrate, heal, grow, respond to physiological needs or injury, and otherwise function at a level sufficient to support the patient's life.

As mentioned above, development of regenerative medical products for organ augmentation and/or replacement is dependent on the availability of appropriate raw materials. Important factors to consider include the source, quality, and handling by which the raw materials are processed into an implantable product. The life-sustaining role of the organ being augmented or replaced must also be considered when sourcing raw materials. As a rule, the manufacturing challenges presented by simple organs (e.g., skin) will be fundamentally different than those presented by vital organs (viscous or solid organs, e.g., heart, kidney, liver, pancreas). Nonetheless, there are fundamental aspects shared by regenerative medical products for which generalizations can be made regarding development approaches and potential regulatory pathways.

Tissue engineering has progressed to the point where neo-organ products are ready now for wholesale manufacturing and commercialization. In response to these advances, the United States Food and Drug Administration (FDA) established the Office of Combination Products (OCP) as the entry point into the regulatory review process for companies seeking approval for combination products, as these products are expected to increase as technological advances continue to merge therapeutic products and blur the historical lines of sepapration between FDA's medical product centers. In the case of neo-organ products composed of a device (e.g., scaffold), biological (e.g., cells), and/or biological active (e.g., pharmaceutical), one or two product centers could be involved.

## 3 REGULATORY HISTORY/DEVELOPMENT

The OCP was established in 2002 under the Medical Device User Fee and Modernization Act (MDUFMA) to address the regulatory complexity for combination products. The MDUFMA gave OCP broad responsibilities, covering the regulatory life cycle of drug-device, drug-biologic, and device-biologic combination products—categories that encompass most neo-organ products being developed today. However, the OCP assigns primary regulatory responsibility for, and oversight of, combination products on a case-by-case basis to one of three product centers based on the product's primary mode of action.

The three product centers that can lead the regulatory review of a neo-organ product include Center for Devices and Radiological Health (CDHR), Center for Biologics Evaluation and Research (CBER), and the Center for Drug Evaluation Research (CDER). CDRH is responsible for regulating those products that are primarily medical devices. A medical device is defined as "An instrument, apparatus, implement, machine, contrivance, implant, or other similar or related article, including a component, part, or accessory,

which is: (1) recognized in the official National Formulary, or the United States Pharmacopoeia, or any supplement to them; (2) intended for use in the diagnosis of disease or other conditions, or in the cure, mitigation, treatment, or prevention of disease, in man or other animals; or (3) intended to affect the structure or any function of the body of man or other animals, and which does not achieve any of [its] primary intended purposes through chemical action within or on the body of man or other animals and which is not dependent upon being metabolized for the achievement of any of its primary intended purposes" [3]. CDER's primary objective is to ensure that all prescription and over-the-counter (OTC) medications are safe and effective. In contrast, CBER is responsible for regulating human tissue, cells, and/or their products under CFR 21, 1270, and 1271 and biologics under 351 of PHS Act 610 of CFR 21. Because the final product produced by tissue engineering and regenerative medicine technologies frequently involves the combination of two or more FDA-regulated components, one or more of the above-mentioned FDA Centers may be assigned to regulate a neo-organ product.

For example, the scaffold or biomaterial component of a neo-organ would normally fall under CDRH jurisdiction, while biological components (i.e., cells) would normally fall under CBER's authority if these components were being used alone. Similarly a biologically active pharmaceutical product would normally be reviewed by CDER. However, review of a combination product can require expertise from multiple centers and the lead center is assigned based on the component that the FDA decides is delivering the primary mode of action.

## 4 ESTABLISHING A REGULATORY PATHWAY FOR NEO-ORGAN

Certain regenerative products may fit into existing regulatory pathways for drugs, devices, or biologicals. Since these singular pathways are already well established, the focus of this discussion will be on neo-organ products composed of a combination of materials (biologics, drugs, and/or devices). In recent years substantial clarification about appropriate regulatory pathways for evaluating combination product neo-organs has occurred. Presently the United States occupies a lead position in establishing regulatory pathways for combination products. The OCP is providing tissue engineering and regenerative medicine industries with guidance as the regulatory authority (21 CFR 3) [4] for combination product development.

A combination product is composed of any combination of a drug and a device; a biological product and a device; a drug and a biological product; or a drug, device, and a biological product. Under 21 CFR 3.2 [4] a combination product is defined to include the following:

1. A product comprised of two or more regulated components (i.e., drug-device, biologic-device, drug-biologic, or drug-device/biologic) that are

physically, chemically, or otherwise combined or mixed and produced as a single entity.

2. Two or more separate products packaged together in a single package or as a unit and comprised of drug and device products, device and biological products, or biological and drug products.
3. A drug, device, or biological product packaged separately that according to its investigational plan or proposed labeling is intended for use only with an approved individually specified drug, device, or biological product where both are required to achieve the intended use, indication, or effect and where, upon approval of the proposed product, the labeling of the approved product would need to be changed (e.g., to reflect a change in intended use, dosage form, strength, route of administration, or significant change in dose).
4. Any investigational drug, device, or biological product packaged separately that according to its proposed labeling is for use only with another individually specified investigational drug, device, or biological product where both are required to achieve the intended use, indication, or effect. [7]

A neo-organ product represents a combination product that is composed of two or more regulated components, under 21 CFR 3.2, that may be mixed together to form a single entity, or that are separate but cannot achieve their intended use without being used together. A sponsor seeking to obtain regulatory guidance for a combination product prepares a Request for Designation (RFD) document [5], laying out key information requested by the FDA (Table 1).

The RFD document presents the sponsor's recommendation and rationale for how the combination product should be regulated. The FDA's decision about regulatory pathway rests on its evaluation of the primary mode of action, a judgment that focuses on the primary therapeutic process whereby the neo-organ activates the therapeutic outcome—usually a choice between scaffold and cellular or constitutive components. If a product is sufficiently related to a product already regulated by a particular center or one that has

**TABLE 1 Request for designation information needed for the food and drug administration**

| |
|---|
| Name of product |
| Composition of product |
| Primary mode of action |
| Method of manufacture |
| Related products currently regulated by the FDA |
| Duration of product use by the patient |
| Science supporting product development |
| Primary route of administration |

a defined pathway, FDA's requirements for preclinical support of clinical trials may mirror the precedent product's regulatory pathway. The FDA has 60 days from the date of RFD submission to render its decision. Once a pathway is identified, the sponsor can engage the lead reviewing center's personnel for the optimal study plan to support their first clinical trials.

Specific guidance on engaging the OCP and establishing communications with the FDA can be found at the FDA's Web site [6]. Interacting with OCP prior to establishing a preclinical development approach can assist in linking to the proper regulatory authority and necessary regulatory guidelines. For some products the primary mode of action is not readily apparent, and the designation of lead center may be based on the most relevant therapeutic activity, intended therapeutic use, or the most relevant safety and efficacy questions. The guidelines of the designated lead regulatory center will provide the framework for a potential regulatory pathway for entry into the clinic and ultimate product registration. A current assignment algorithm can be obtained at the FDA's Web site [4].

## 5 DEVELOPMENT OF NEO-ORGAN PRODUCTS

The first tissue-engineered and regenerative medical products were device products for skin conditions, regulated by CDRH, and based on culturing skin cells in simple layers. The earliest clinical applications of tissue engineering revolved around the use of essentially flat materials designed to stimulate wound care. The cellular component was considered more as a physical barrier than a biologically active component of the product itself. Dermagraft®, a product originally manufactured by Advanced Tissue Sciences, Inc. that is now manufactured by Smith & Nephew, used skin cells isolated from neonatal foreskins seeded onto a polymeric scaffold. Apligraf®, a product from Organogenesis Inc., used similar technology to seed and culture cells on collagen-based scaffolds. Tissue-engineered skin substitutes dominated the regulatory and development pipeline in the late 1990s. The FDA regulated these products as class III devices. As these products moved from the laboratory to the clinic, development issues such as cell source (allogeneic or autologous), scaffold materials (natural or synthetic), and bioreactor design encountered additional safety challenges, including immunocompatibility responses in patients. As these challenges were solved, further commercial demands for scale-up and cost-of-goods considerations had to be reconciled against the magnitude of incremental medical benefit over existing approaches (i.e., cadaveric skin).

Medtronic, Inc. markets one of the first commercial products that combined a device and a biologically active compound, bone morphogenic protein (BMP). Medtronic's INFUSE® Bone Graft/LT-CAGE® lumbar tapered fusion device contains a collagen sponge infused with the BMP peptide housed within a metal cage. This product was developed as a combination product in which CDRH maintained a leadership role for regulatory guidance and

consultation with CBER and CDER occurred for complete regulatory oversight. The combination product was approved for treatment of lower spine disc disease via the FDA's premarket approval (PMA) process. The individual components were each regulated by the appropriate pathway: the recombinant human BMP2 (manufactured in accordance with biologics regulations), the collagen sponge to carry the protein (manufactured under device regulations), the sterile water to resuspend the protein (manufactured by a supplier in compliance with drug regulations), and the metal cage device (manufactured under device guidelines) housing the sponge plus protein that is surgically implanted between the patient's vertebrae. Later CDRH approved INFUSE® Bone Graft, consisting of the collagen, sterile water, and protein components (no metal cage) for use in open fractures of the tibia. The product is used in addition to standard treatment protocols, which may or may not include using an intramedullary nail to stabilize the fracture.

Many of these early successes in the tissue engineering field were gained in relatively simple two-dimensional (2D) skin tissue or 3D bone fracture structures with limited vital organ function required immediately after implantation. Organ regeneration goes beyond traditional tissue engineering, which focuses on using in vitro and ex vivo methods to replace as much of the tissue as possible before implantation. Tissue engineering techniques are applied to regenerative medical products, but these products also focus on utilizing the natural healing processes of the body to recapitulate the organ being augmented or replaced.

One example of a regenerative medicine product in development is Tengion, Inc.'s focus to replace complex organs while developing the basic biological understanding necessary to ensure integration into the patient's own body. Tengion is currently bringing forward advanced neo-organ technology designed to augment or replace failing internal organs. Tengion's product used preclinical and clinical development pathways offered by the OCP. The regulatory lead center was assigned to CBER based on the FDA's assessment of primary mode of action (cells–biologics) with collaboration from CDRH to cover the device (scaffold) components.

Tengion's technology platform brings functionality and vascularization of a complex organ by using autologous and homologous progenitor cells, isolated and cultured ex vivo, which are then seeded onto a degradable biomaterial whose shape and biomechanical properties are optimized to the body organ being augmented or replaced. The cell-seeded neo-organ construct provides the minimal raw materials necessary for tissue regeneration and the target organ's physiologic function at the time of surgical implantation. Following implantation, the product's scaffold is degraded, bioabsorbed, and replaced by organized stroma within the patient. Final and complete parenchymal organ regeneration is patient specific.

Using the neo-organ construct as a template, the body regenerates healthy tissue, restoring function to the patient's failing organ. By providing the patient with an autologous-homologous neo-organ for tissue regeneration, many of

the necessary co-therapies used in traditional donor transplantation techniques are avoided, such as immunosupression. Deriving the biologically active cellular component from the patient avoids the traditional issue of limited organ donor supply.

In considering the development pathways traveled by these early transformational products, we can begin to see a framework that combines established technologies with emerging regulatory pathways and defines preclinical and clinical development strategies for products that address previously untreatable diseases. Development of therapeutic modalities to treat and potentially cure such maladies as diabetes, heart disease, renal failure, and other diseases caused by malfunctioning, damaged, or failing tissues is now a reality with regenerative medical therapies envisioned by the US Department of Health and Human Services [7] as "tissues for life" (Executive Summary, para. 2).

## 6 NEO-ORGAN PRODUCTS

Since most neo-organs will be regulated as combination products, a tiered format of analyses allows for each component to be evaluated as well as the final combination product. The neo-organ preclinical development program can be based on a foundation of current risk analysis principles combined with clinically relevant functional endpoints to establish structure–function relationships before entry into human clinical trials. Ultimately the scientific approach employed will include established anatomical, surgical, clinical, cell biological, physiological, and material sciences principles appropriate for the specific neo-organ product under development [8–14].

Neo-organ safety studies will include both acute and chronic studies, since the product will develop into an organ that remains with the patient for a lifetime. Importantly, long-term studies provide evidence for at least two outcomes vital for all products within this category: product durability and the neo-organ's capacity for becoming responsive to each individual patient's need (i.e., capacity or size). Regenerative processes are dynamic and can continue for several months. Therefore short-term studies are focused primarily on characterization of acute host tissue biological responses, biomaterial degradation, neo-tissue integration, and neo-organ functional failure or comorbidity, Longer term studies focus on evaluating the capability of the neo-organ to integrate and mold within the patient's organ systems or body size over time, neo-organ functionality, and the absence of chronic inflammatory or neoplastic processes.

### 6.1 Development Strategies

Planning for clinical testing involves defining the unmet medical need (disease characterization), determining how the intended use of the neo-organ product addresses the need, and defining processes that yield reproducible production of a safe and efficacious product for placement into patients. Developing a

scientifically based definition of what constitutes a successful outcome (i.e., primary clinical endpoint) is a prerequisite for entering clinical trials. Applying existing regulatory guidelines for product testing and manufacture will be required during preclinical testing prior to its use in a human subject.

Defining the optimal regulatory pathway is done in partnership with the FDA, most likely through the OCP. With this information in hand, a sponsor can initiate preclinical testing, from which considerations for clinical testing emerge. A scientifically based program produces a foundation for exploratory clinical testing where final product characteristics are defined, production processes are standardized, and justifications that indicate readiness for entering into the clinical testing phase are anticipated. Table 2 presents an overview of a prototypical neo-organ product development process.

Transitioning into an initial exploratory clinical evaluation rests upon safety and functional understanding built during preclinical evaluation. The preclinical program establishes assessment objectives related to specific product characteristics and the supporting production processes. Translational medicine study results define the extent to which preclinical information demonstrates

**TABLE 2 Overview of a potential testing program to support clinical entry of a prototypical neo-organ product**

| |
|---|
| *Process manufacturing control* |
| • Define product production and early manufacturing processes |
| • Establish cell, tissue, and biomaterial sourcing and qualify vendors |
| • Validate in-process and final release testing schemes |
| • Characterize adventitious agents and impurities for each raw material |
| • Define lot-to-lot consistency criteria |
| • Validate quality control procedures |
| *Preclinical studies* |
| • Complete in vitro and in vivo safety testing |
| • Define toxicity testing of raw materials |
| • Evaluate biomaterial biocompatibility |
| • Establish immunogenic and inflammatory responses to raw materials |
| • Develop rationale for animal and in vitro models to test product function |
| • Define endpoints for establishing individual component and/or final product safety |
| • Define endpoints for establishing organ or tissue functionality |
| • Define endpoints for establishing organ or tissue durability |
| *Clinical trials* |
| • Develop rationale for safety and clinical benefit (risk–benefit analysis) |
| • Design exploratory and confirmatory trials |
| • Select patient population and define inclusion/exclusion criteria |
| • Identify investigational comparators and control treatments |
| • Establish primary and secondary study endpoints |
| • Consider options for data analysis and potential labeling claims |

**TABLE 3 Preclinical data collection prior to clinical study initiation**

| | In vitro | In vivo |
|---|---|---|
| Raw materials supply | | |
| Cells[a] | + | ± |
| Scaffold | + | ± |
| Manufacturing process controls: In-process and potency | | |
| Cellular processing[a] | + | − |
| Biomaterial processing | + | − |
| Final combination product[a] | + | + |
| Preclinical development studies | | |
| Safety and efficacy | + | + |
| Endpoint selection | + | + |
| Translation into clinical design | ± | + |

[a]Products comprised of scaffold material and biologically active substances (e.g., pharmaceutical agents) do not require evaluation of cellular components.

the desired clinical outcome [15,16]. Relevant clinical endpoints need to be incorporated, wherever possible, into the animal studies. As such the endpoints need to be identified before the design of animal studies is finalized. Table 3 provides an overview of the preclinical data that can be generated prior to exploratory clinical trial initiation.

Regulatory considerations impact the types of data required and production process technologies that must be in place prior to initiating clinical trials. There are three considerations that have significant impact on the preclinical development plan: the extent of cellular manipulation, relationship between cell source and application, and scaffold characteristics. Table 4 summarizes some regulatory considerations for cellular and scaffold components.

## 6.2 Raw Material Testing

***Cells*** Cellular components are living "raw" materials. They can be derived from the patient (autologous), other donors (allogenic), or animals (xenogenic). An additional consideration is the relationship between the cell source and the target tissue of the neo-organ product. Homologous cells are derived from the organ to be augmented or replaced. Heterologous cells are derived from a tissue or organ other than the augmentation or replacement target. Cell quality standards have been extensively reviewed and defined. Major objectives are controlling introduction of infectious agents and preventing contamination during neo-organ manufacturing. Contamination concerns include cross-contamination between patients and environmental contamination from the facility and equipment. Infectious agents may come into production with the cell source or be exogenously introduced via materials that come in contact with neo-organ materials during production (e.g., bovine-derived material).

**TABLE 4 Regulatory considerations for the development of a neo-organ product**

| | Description | Impact |
|---|---|---|
| Manipulation of cells | For structural and nonstructural tissues, manipulation is minimal if it involves centrifugation, separation, cutting, grinding and shaping, sterilization, lyophilizing, or freezing (e.g., cells are removed and reintroduced in a single procedure).<br>Manipulation is not minimal if cells are expanded during culture or growth factors are used to activate cells to divide or differentiate.<br>Defined in 21 CFR 1271.3(f); see also [27]. | More extensive regulatory requirements applied to neo-organ products when cellular manipulation is more than minimal. |
| Cell source and application | Homologous use is interpreted as the augmentation tissue using cells of the same cellular origin. Examples include applying bone cells to skeletal defects and using acellular dermis as a urethral sling.<br>Nonhomologous use examples include using cartilage to treat bladder incontinence or hematopoietic cells to treat cardiac defects.<br>Defined in 21 CFR 1271.3(c); see also [27]. | Nonhomologous use triggers additional requirements for entering clinical trials [28]. |
| Scaffold characterization | Final scaffold composition and design determine whether the neo-organ product is characterized as a device, a biologic, or a combination product.<br>Defined in Quality Systems Regulations (QSRs) in 21 CFR 820 [22]; see also [29]. | Devices are held to the QSRs in 21 CFR 820 [22], biologics are required to comply with good manufacturing processes (GMPs) [29] and combination products are often required to comply with both sets of regulations and guidelines. |

Neo-organs designed to regenerate an entire organ must contain a sufficient number of viable cells of the proper type(s) at the time of implantation to support and/or promote de novo growth of multiple tissue layers resulting in the development of a functional, persistent organ. Therefore a cell source must exist that can expand to the required numbers in vitro or in vivo. Scientific efforts are focused on establishing cellular sources (e.g., stem cell and progenitor cells) that can be expanded and guided to differentiate into the required cell type. Following implantation, these cells must retain functional and structural consistency with the tissue and organ being replaced. Each manipulation along a differentiation process carries a potential safety risk. The total number of manipulations required to align cell source and tissue target must be determined. Also assessments to determine persistence of differentiated phenotype and function must be defined and tested in preclinical studies.

Many stem cells can be expanded and guided along a differentiation pathway to achieve a terminal functional state. Preclinical development of stem cells for neo-organ products must also demonstrate that the selected cellular raw material can safely restore original tissue structure and function (Table 5). Additionally, because stem cells have been manipulated into a new phenotype and incorporated into a neo-organ, preclinical studies must determine that safe maturation of a multilayered tissue occurs ex vivo or in vivo, depending on the product's primary mode of action. Finally, the safety of augmentation or regeneration processes must be assessed along with determining whether the implanted cells persist in their new (or final) phenotype and/or remain localized to the target organ site for the duration of the organ's life expectancy.

Differentiation stability and phenotype/localization persistence must also be determined for cells obtained from heterologous sources such as umbilical cord tissue and blood (UCB) [17] or adipose tissue (AT) [18]. Endothelial progenitor cells are isolated with UCBs from the blood component, adipose tissues and Wharton's jelly-derived myofibroblasts from tissue biopsies [18] contain cells capable of becoming multiple cell lines: adipogenic, osteogenic, chondrogenic, and myogenic [18]. When using these or similar heterologous

**TABLE 5 Cell source and potential safety risk**

| Source | Ex vivo Manipulation | Potency | Safety Risks[a] |
|---|---|---|---|
| Embryonic stem cell | Extensive | Pluripotent | +++ |
| Bone marrow | High | Multipotent | ++ |
| Umbilical cord | High | Multipotent | ++ |
| Adipose tissue | High | Multipotent | ++ |
| Adult tissue progenitor | Minimal to moderate | Unipotent | + |

*Note*: + to +++ = progressively increasing levels of risk.
[a]Safety risks estimation is proposed and based on (1) biological controls for cellular stability, immunogenicity, tissue growth differentiation (i.e., risk of dysplasia or neoplasia), and (2) extent of ex vivo manipulation required to produce a neo-organ product.

cell sources, the preclinical program must evaluate purity, homogeneity, and the potentials for dysplasia, metaplasia, and neoplasia. However, one benefit of using adult stem cells from bone marrow (BM) or AT is the cells' ability to express growth factors and drive angiogenesis. This benefit must be weighed against the potential for the emergence of an undefined functional fate and possible cell expansion limits imposed by the need to maintain required phenotype characteristics, lineage compositions, and/or immunological behavior [19].

Within cell lineage differentiation pathways, the inherent potency of the starting material may be restricted based on original source. Bone marrow derived cells have the ability to differentiate into eosinophils, erythrocytes, megakaryocytes, osteoclasts, and lymphoid cells [20]. However, these hematopoietic lineages have limited potential to become a distinct tissue such as liver. Mesenchymal stem cells, on the other hand, can differentiate into a broader array of distinct tissues such osteocytes, chondrocytes, adipocytes, myocytes, and bone marrow stromal cells [20]. This plurality of outcomes requires identification of circumstances under which an aberrant outcome is manifested.

Regardless of cell source, the preclinical program demonstrates that the chosen raw cellular materials are capable of regenerating into a functional new organ. The cellular component must also be shown to function in accordance with the requirements of the native tissues and organ being replaced or augmented. Committed adult progenitor cells obtained from the same tissue being replaced (homologous tissue source) represent the lowest tier of safety concern, especially when they are returned to their original location to recapitulate homologous organ function.

Currently there is scientific interest around manipulating less differentiated cellular material (e.g., embryonic stem cells) to become differentiated and functional cells that may form tissues and, ultimately, functioning organs. The greater the pluripotency of the starting cell population, the higher the potential for becoming a different cell type (dysplasia) and/or exhibiting uncontrolled growth (neoplasia). Neo-organ products using less differentiated cells or cells from an allogenic source carry potential risks of immune response, rejection, and/or altered homestatic controls, including neoplasia. Any of these responses can lead to abnormal cellular and tissue growth. Driving undifferentiated cells to a steady state by ex vivo manipulation during the manufacture of neo-organs can introduce high concentrations of growth factors or involve manipulation techniques whose effects may persist postimplantation. Potential consequences of these ex vivo manipulation techniques include unresponsive or overresponsive cell growth, altered differentiation, and abnormal tissue growth, organic structure, and function.

Preclinical development programs are required to establish modeling systems that adequately evaluate and/or detect potential risks associated with cellular differentiation, tissue growth, and organogenesis—all necessary components of a neo-organ regenerative process. Sufficient control studies are

needed to evaluate in vivo proliferation and differentiation of a neo-organ's cellular components. Additionally tests must be designed to distinguish normal differentiation and growth from adverse and uncontrolled tissue growth patterns. Moreover, if a patient receiving a neo-organ product requires medication with immunocompromising drugs, the neo-organ's capacity to integrate with native tissue could be affected. As such, postimplantation studies under a variety of co-therapeutic conditions may need to be incorporated into the preclinical development strategy.

The goal of these studies would be to demonstrate postimplantation neo-organ function and maintenance of the desired cellular differentiation and phenotypic state. Cellular components of neo-organ products may have the potential to undergo excessive proliferation (risk of hyperplasia), altered cellular differentiation (risk of dystrophy), or respond abnormally to homeostatic control of tissue growth (risk of dysplasia). Certain categories of risks that a preclinical program must evaluate can be predicted by the combination of cell source and target location. Table 6 illustrates this principle and lists some of the assays that should be incorporated into preclinical evaluation of neo-organ products depending on which combination applies. In all cases preclinical programs should be designed to provide evidence that the implanted neo-organ product produces stable organic function and proper therapeutic outcome by integrating into the body's systems and maintaining life.

For neo-organ product development, simply demonstrating that the cellular constituent is stable is insufficient. Cellular components of neo-organ products are expected to follow normal cell cycle and differentiation pathways. They are also expected to become part of a functioning tissue. Furthermore the neo-organ product is expected to develop into a life-sustaining organ. Achieving this ultimate level of de novo tissue regeneration and organ function generally requires the use of a structural component that forms a scaffold or temporary biocompatible frame upon which cellular components can achieve the desired integration and maturation. Testing of the scaffold or device components used in neo-organ products is well-established from a long history of traditional device product development.

***Scaffold*** When cells are placed onto a scaffold, scientific and regulatory considerations focus on ensuring that both the raw materials comprising the scaffold and its three-dimensional characteristics are biocompatible [21]. Biocompatibility testing involves evaluation of the scaffold's potential cytotoxicity at several levels. The first level is biocompatibility with cells placed onto the scaffold. The second level is potential toxicity to the recipient's tissues postimplantation. The third level is the consequences of a patient's immune and inflammatory responses to the neo-organ product following implantation. Biocompatibility must continue throughout the in vivo regeneration process; therefore biocompatibility testing should parallel preclinical evaluation of the scaffold's biomechanical properties during new tissue or organ growth.

**TABLE 6 Risk-based assay selection for the evaluation of cellular constituent in neo-organ products**

| | Cell Source/Neo-Organ Product | | | | |
|---|---|---|---|---|---|
| | Autologous Homologous | Autologous Heterologous | Allogeneic Homologous | Allogeneic Heterologous | Embryonic Homologous/Heterologous |
| | | | *In vitro* | | |
| Karylogy | + | + | + | + | ++ |
| Phenotype | + | + | + | + | +++ |
| Cell cycle | + | + | + | + | +++ |
| Metabolic | + | + | + | + | ++ |
| Cytokine | + | ++ | + | ++ | ++ |
| Endocrine | + | ++ | + | ++ | ++ |
| | | | *In vivo* | | |
| Morphology | + | + | + | + | ++ |
| Function | + | ++ | + | ++ | +++ |
| Tissue Integration | + | ++ | ++ | ++ | ++ |
| Organ Physiology | + | + | + | + | ++ |
| Systems biology | + | ++ | + | ++ | +++ |
| | | | *Special* | | |
| Immunotoxicity | – | – | ++ | ++ | ++ |
| Dystrophy | – | + | – | ++ | +++ |
| Chorista | + | – | ++ | – | ++ |
| Hamartia | – | + | – | ++ | +++ |
| Anaplasia | – | + | – | ++ | +++ |
| Neoplasia | – | + | – | ++ | +++ |

*Note*: Estimated risk is proposed based on considerations outlined in the text. – = limited or no risk; + to +++ = progressively increasing levels of risk and extent of study required for each area to be assayed. Chorista is defined as a focus of tissue that is histologically normal but not normally found in the organ or structure in which it is located. Hamartia is defined as a formation defect characterized by the abnormal arrangement or combination of tissues normally present in a specific area.

Synthetic, natural or semi-synthetic composite materials are readily available from various commercial sources, but the quality control of a material varies substantially between medical and research grades. Many aspects of device testing have been outlined in 21 CFR 820 [22]; however, during the preclinical phase of testing the most relevant of these guidelines are biomaterial composition, process validation, and design controls.

Typically the design input phase is a continuum beginning with feasibility and formal input requirements and ending with early physical design activities. Engineering input on final prototype specifications follows the initial design input phase and establishes the design reviews and qualification. For a combination neo-organ product, defining quality for the chemical polymer (e.g., polyglycolic acid or PGA) or natural material (e.g., collagen), including any residues introduced during machine processing (e.g., mineral oil), can require quality systems regulations (QSR) integration into a product that would otherwise be regulated as a biologic. Since most neo-organ products are combination products, testing of scaffold, cells, and the cell-seeded scaffold (i.e., construct) are required to ensure that in exploratory clinical trials, the product is sterile, potent, fit for use, and composed of the appropriate raw materials to function properly following in vivo placement.

Scaffolds can be made from natural or synthetic materials. Such materials fall under the category of biomaterials. A biomaterial can be considered a single element or compound, which is a composite or mixture of elements, and is synthesized or derived to be used in the body to preserve, restore, or augment the structure or function of the body. Examples of natural materials for scaffold construction are: extracellular matrix, collagen, fibrin, and polysaccharides (e.g., chitosan or glycosaminoglycans). Natural materials, unless they are obtained from the patient who receives the neo-organ implant, will cause an immunogenic response. This is not always the case with synthetic materials.

Commonly used synthetic materials include polylactic acid (PLA), polyglycolic acid (PGA), polycaprolactone (PCL), and poly(lactic-co-glycolic acid) (PLGA). PLGA copolymer has been approved for use by the FDA in a number of surgical products and is commonly used by the biomedical industry. Two benefits of PLGA are its biodegradability and biocompatibity. Biodegradability is important because it allows the scaffold to be absorbed by the surrounding tissues without surgical removal (e.g., PGA undergoes spontaneous hydrolysis in body tissues). Biocompatibility is important as it supports the appropriate cellular activity and facilitates molecular and mechanical signaling systems. These characteristics help optimize tissue regeneration without eliciting undesirable cellular effects or inducing undesirable local or systemic responses in the host tissue following implantation.

For newly developed, as yet unapproved, novel biomaterial scaffolds, the regulatory preclinical guidelines developed for device materials evaluation are particularly useful in guiding the pre-clinical program of a neo-organ product. In general, the following preclinical tests should be included for such scaffold materials: acute, subchronic, and chronic toxicity; irritation to

skin, eyes, and mucosal surfaces; sensitization; hemocompatibility; genotoxicity; carcinogenicity; and the materials' effect on reproduction and embryonic development. Last, target organ toxicity evaluations, such as neurotoxicity and/or immunotoxicity assays, may be necessary as part of the preclinical development. For listings of evaluation tests to consider, see Tables 7 and 8.

Collectively these guidelines serve as points to consider for preclinical development program of any neo-organ composed of a biocompatible biomaterial-based scaffold. The ultimate focus of biomaterial and scaffold testing is to evaluate its in vivo behavior following implantation. Characterizing the degradation profile ensures that the scaffold's supportive properties are maintained long enough to allow the regenerating tissue to acquire functional and structural integrity. Defining scaffold breakdown products identifies biochemical factors that may impact reparative, inflammatory, immunologic, and regenerative processes following product implantation. Measuring biomechanical properties such as stress–strain relationships and other characteristics ensures that the scaffold portion of the combination product will perform properly after implantation into the body.

### 6.3 Final Neo-Organ Product Testing

***Model System Selection*** Safety and function evaluation of a neo-organ product is conducted in animals, and these findings are foundational to the design of the first clinical trial protocol. Translational studies are the basis for safely transitioning a potential neo-organ product into clinical testing. Since the neo-organ and its components can invoke multiple homeostatic (e.g., metabolic), defense (e.g., immune), and healing (e.g., inflammation) pathways, animal studies provide an approach to understand the inherent function of the neo-organ product.

Preclinical studies can be conducted in large (e.g., dog, goat, pig) or small (e.g., rat, mouse) animal models for the purpose of translating safety and function to the clinical trial design. Selection of the correct animal model should be based on clinical endpoint similarities between the animal model's pathophysiological, physiological, and structural characteristics and those observed in the clinical indication. Exploratory clinical trials for most medical products utilize normal human volunteers as the first line of clinical testing. However, some neo-organ products cannot be tested outside of the intended clinical population. As such, the animal model employed in translational medicine should resemble the human condition as closely as possible—immune status, inflammatory response, and healing pathways. Medical treatment approaches used to treat the human condition (e.g., surgical procedure) and monitoring methods to follow a clinical benefit or risk (e.g., imaging) should also be mimicked as closely as possible in the animal model.

Surgical aspects of implanting the neo-organ product may enter into the preclinical development paradigm. If new surgical procedures are required, or the new technology uses approved products in an off-label manner, an assess-

**TABLE 7 Initial evaluation tests for consideration**

| Device Categories | | | Biological Effect | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Body Contact (see 4.1) | | Contact Duration A Limited (24 h) B Prolonged (24 h to 30 days) C Permanent (>30 days) | Cytotoxicity | Sensitization | Irritation or Intracutaneous Reactivity | System Toxicity (Acute) | Subchronic toxicity (Subacute Toxicity) | Genotoxicity | Implantation | Haemocompatibility |
| Surface devices | Skin | A | × | × | × | — | — | — | — | — |
| | | B | × | × | × | — | — | — | — | — |
| | | C | × | × | × | — | — | — | — | — |
| | Mucosal membrane | A | × | × | × | — | — | — | — | — |
| | | B | × | × | × | 0 | 0 | — | 0 | — |
| | | C | × | × | × | 0 | × | × | 0 | — |
| | Breached or compromised surfaces | A | × | × | × | 0 | — | — | — | — |
| | | B | × | × | × | 0 | 0 | — | 0 | — |
| | | C | × | × | × | 0 | × | × | 0 | — |
| External communicating devices | Blood path, indirect | A | × | × | × | × | — | — | — | × |
| | | B | × | × | × | × | 0 | — | — | × |
| | | C | × | × | 0 | × | × | × | 0 | × |
| | Tissue/bone dentin communicating + | A | × | × | × | 0 | — | — | — | — |
| | | B | × | × | 0 | 0 | 0 | × | × | — |
| | | C | × | × | 0 | 0 | 0 | × | × | — |
| | Circulating blood | A | × | × | × | × | — | 0 | — | × |
| | | B | × | × | × | × | 0 | × | 0 | × |
| | | C | × | × | × | × | × | × | 0 | × |
| Implant devices | Tissue bone | A | × | × | × | 0 | — | — | — | — |
| | | B | × | × | 0 | 0 | 0 | × | × | — |
| | | C | × | × | 0 | 0 | 0 | × | × | — |
| | Blood | A | × | × | × | × | — | — | × | × |
| | | B | × | × | × | × | 0 | × | × | × |
| | | C | × | × | × | × | × | × | × | × |

*Source*: Reproduced from [22].
*Note*: X = ISO evaluation tests for consideration; 0 = additional tests which may be applicable; note + tissue includes tissue fluids and subcutaneous spaces; note ∧ for all devices used extracorporeal circuits.

**TABLE 8 Supplementary evaluation tests for consideration**

| Device Categories | | | Biological Effect | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Body Contact (see 4.1) | | Contact duration A Limited (–24 h) B Prolonged (24 h to 30 days) C Permanent (>30 days) | Cytotoxicity | Sensitization | Irritation or Intracutaneous Reactivity | System Toxicity (Acute) | Subchronic toxicity (Subacute Toxicity) | Genotoxicity | Implantation | Haemocompatibility |
| Surface devices | Skin | A | — | — | — | — | — | — | — | — |
| | | B | — | — | — | — | — | — | — | — |
| | | C | — | — | — | — | — | — | — | — |
| | Mucosal membrane | A | — | — | — | — | — | — | — | — |
| | | B | — | — | — | — | — | — | — | — |
| | | C | 0 | — | — | — | — | — | — | — |
| | Breached or compromised surfaces | A | — | — | — | — | — | — | — | — |
| | | B | — | — | — | — | — | — | — | — |
| | | C | 0 | — | — | — | — | — | — | — |
| External communicating devices | Blood path, indirect | A | — | — | — | — | — | — | — | — |
| | | B | — | — | — | — | — | — | — | — |
| | | C | × | × | — | — | — | — | — | — |
| | Tissue/bone dentin communicating | A | — | — | — | — | — | — | — | — |
| | | B | — | — | — | — | — | — | — | — |
| | | C | 0 | × | — | — | — | — | — | — |
| | Circulating blood | A | — | — | — | — | — | — | — | — |
| | | B | — | — | — | — | — | — | — | — |
| | | C | × | × | — | — | — | — | — | — |
| Implant devices | Tissue bone | A | — | — | — | — | — | — | — | — |
| | | B | — | — | — | — | — | — | — | — |
| | | C | × | × | — | — | — | — | — | — |
| | Blood | A | — | — | — | — | — | — | — | — |
| | | B | — | — | — | — | — | — | — | — |
| | | C | × | × | — | — | — | — | — | — |

*Source*: Reproduced from [22].

ment of new safety risks may be needed that are not directly related to the product itself. The number of possible new approaches or new applications of current surgical products is sufficiently large that each needs to be considered on a case-by-case basis in close collaboration with the FDA or regulatory

agency responsible for ensuring safety. Such considerations are beyond the scope of this chapter.

The biological activity of a neo-organ product may preclude the use of commonly used animal models for safety testing (e.g., rats and dogs). Safety evaluation programs should include the use of species relevant to both safety testing and providing proof-of-concept evaluation of the neo-organ's ability to function and sustain homeostasis. A relevant species is one in which the test product is biologically active in regenerating or replacing the native organ.

Safety evaluation programs have historically used two relevant species. However, in certain cases one relevant species may suffice when biological activity (e.g., organ function) or technical feasibility (e.g., surgical implantation) considerations render the use of another species irrelevant. Even under circumstances where two species may be necessary for short-term toxicity studies, it may be possible to justify using only one species to evaluate durability, growth potential, and full regenerative outcome in subsequent long-term toxicity studies. Long-term studies may need to extend for months or even years to demonstrate that the neo-organ can function in a safe and appropriate manner.

For neo-organ products, safety studies in nonrelevant species may actually provide misleading results. Lower order species have a greater potential for regeneration, allowing a product to appear to have more regenerative capacity than would be manifested in humans. New developments in animal models of human diseases may bring these advanced technologies to greater use in studying the replacement and augmentation of diseased organs (e.g., transgenic animal models reflecting human disease pathogenesis). These models may provide further insight, not only in determining the mechanism of regeneration but also in safety evaluation (e.g., evaluation of undesirable promotion of disease progression). Scientific justification for safety evaluations using these animal models of human disease should be provided.

Since a neo-organ must function at the moment of implantation, animal studies provide an understanding of how to evaluate initial body responses as well as longer term outcomes reflecting the desired benefit—an augmented or replaced tissue or organ. Since most neo-organs remain with the patient for a lifetime following surgical implantation, translational study duration should extend to a time when final clinical outcome is achieved. Considerable thought should be given to the duration of preclinical studies. Since the implications of these therapies relates to patient longevity, the duration of these studies may extend from months to years. Nonetheless, the final outcome is frequently achieved in a shorter time period of time, so the potential to conduct shorter duration studies based on final patient outcome may present a rational solution to testing clinical utility in the shortest possible time while ensuring a high outcome benefit–risk ratio.

***Endpoint Selection*** As with other medical products, preclinical development can be conducted using selected toxicological endpoints (e.g., in-life, serum chemistry, hematological, urinalysis, and target and sentinel organ his-

tological evaluation) based on the target organ being replaced or augmented. Final organ function will influence the appropriate endpoints to monitor during in-life and postmortem analysis. Neo-organ safety evaluation would also focus on the mechanism of product function, extent of neo-organ integration into the body, neo-organ structure–function relationships, appropriate cellular, local and target tissue, and systems biology aspects.

In-life observations, serum chemistry, hematology, urinalysis, and body weight gain coupled with imaging and other clinically relevant procedures can be used to monitor the progression of neo-organ incorporation into the body and assumption of a functional role in homeostasis. Abnormal tissues, areas of adhesion, and any macroscopically or functionally abnormal tissues should be collected and evaluated postmortem. Histopathological evaluation of target organ tissues, potential target organs, and peripheral tissues should be evaluated postmortem to evaluate adverse effects and document the neo-organ's function(s). Immunohistochemical evaluation may also be conducted to demonstrate cell phenotype, tissue morphogenesis, and neo-organ characteristics.

Cellular analysis will depend on cell source, with higher risk cell sources requiring more extensive in vitro, in vivo, and ex vivo analysis. Genetic stability and cell phenotypic analysis are also focus areas in preclinical studies, since they provide evidence of appropriate regeneration and function of the cellular component in the neo-organ product.

Understanding which endpoints are available and appropriate for first clinical trials is achieved through translation of preclinical development study results. Standards are provided for the proper safety evaluation of neo-organ products whether they are regulated as a device [23] or a biological product [24,25]. Although the optimal testing strategy will typically be product specific [23], some basic guidelines for testing device-like neo-organ products can be found in the ISO10993-1 guidance document [26]. These guidelines also provide insight into possible testing strategies to evaluate the final product after all components are combined (Table 9).

Final neo-organ products that are cell based or neo-organs in which the primary mode of action is mediated through cellular constituents of a scaffold-cell combination product require an assortment of studies. Scaffold and cellular components are evaluated through appropriate endpoint selection. Endpoints selected should have direct translation to the clinical testing phase,

**TABLE 9 Test categories described in ISO10993-1 (USP 71)**

| In vitro Assays | In vivo Assays |
|---|---|
| Cytotoxicity | Irritation |
| Pyrogenicity | Sensitization |
| Hemocompatability | Acute systemic toxicity |
| Genotoxicity/genetic tests | Subchronic toxicity |
| | Local tolerance |

**TABLE 10 General preclinical development testing paradigms for a cell-based neo-organ product**

| Cellular Component | Scaffold | Combination |
|---|---|---|
| Phenotype characterization<br>Genetic stability | Early stage (acute)<br>Late stage (chronic)<br>Biocompatability<br>Biomechanical properties<br>Degradation profile | Early stage (acute toxicity)<br>Late stage (chronic toxicity) |

emphasizing the importance of having matching endpoints for both clinical and laboratory situations.

***Experimental Design*** Experimental design for both in vitro and in vivo translational studies is influenced by the endpoints that will be evaluated in clinical studies. The experimental design should also recognize that new endpoints may need to be evaluated in clinical trials. An example of a preclinical development program for a combined cell-based neo-organ product is presented in Table 10.

Specific testing approaches are not absolutely defined. However, testing approaches and their scope for a specific neo-organ product can often be predicted by comparison with development strategies used for related component technology platforms. As more combination products are approved for clinical use, the choice of related technology platforms for comparison will broaden.

Transitioning a neo-organ product from preclinical development to a scientifically based clinical evaluation involves several steps. First to be demonstrated is the ability of the neo-organ to invoke a bodily response that has therapeutic benefit. Second, a controlled and reproducible manufacturing process must be in place. Finally, the safety of each component and the final product (a safe and functioning neo-organ) must be demonstrated. This transition to clinical evaluation is typically the first point of regulatory authority and governance body interactions, and an area where there are established procedural approaches and controls to review and clarify the experimental design of preclinical studies.

## 7 TRANSITION TO CLINICAL TESTING

### 7.1 Defining and Testing a Prototype

Prior to initiating a clinical testing program, a neo-organ product's specific characteristics must be defined to the point that the product can be repeatedly and reproducibly manufactured for in vitro and in vivo testing as defined above. Product characteristics should be sufficiently stable to allow for

data-driven demonstration of their clinical utility. Once a prototype is defined, its characteristics can be evaluated in a series of tests to define the limits of the initial design criteria. Durability testing establishes failure points, limits of neo-organ application, and evaluates the achievement of design criteria.

Anticipating the clinical conditions, complications, and untoward events that may arise during clinical testing also establishes a prototype product's potential for clinical utility. A neo-organ product is seldom introduced as a final functioning organ, since even highly engineered neo-organ products require scaffold degradation, neo-vascularization, tissue integration, and functional maturation stages to be complete prior to healing. Therefore characterization of pharmacological responsiveness and electrophysiological parameters, along with determining phenotypic and structural features of the emerging neo-tissue and/or neo-organ following implantation, are key data for supporting claims about the product's ultimate clinical benefit.

Human testing of specific design elements of a final neo-organ prototype is the culmination of a series of biological, physical, and chemical evaluations obtained during the preclinical prototyping phase. As defined above, these evaluations span sourcing and control of raw materials, assembly processes (i.e., in-process testing), and release criteria. Additionally the product's shelf life, shipping conditions (temperature, humidity, nutrients, etc.), stability, sterility, and method of use are established in the preclinical testing phase. Translational studies provide anticipated protocols for unique surgical procedures, recovery times, and clinical management practices during and after implantation once a final prototype product with the fully embodied characteristics is available.

Using previously tested technology platforms can accelerate neo-organ product entry into clinical testing. Leveraging prior approval of scaffold materials, cell processing methods, culture media components, or transport containers can greatly reduce the number of variables that need to be tested in product prototyping and preclinical testing phases. Additionally availability of historical data for employed technologies can assist in the development of final prototype testing strategies and early clinical trial designs.

## 7.2 Production of Neo-Organs for Preclinical Testing

Production process definition for neo-organ products in the preclinical testing stage of development takes on increasing importance relative to traditional noncombination devices, drugs, and biologics. Clearly, established product characteristics, standard operating procedures, and the use of clinical-grade production in late-stage preclinical testing can accelerate the deployment of a facility to manufacture final preclinical prototype products and extend to the manufacture of GMP-grade products for initial human trials. Neo-organ technologies can vary substantially, but small GMP facilities can meet neo-organ production needs for both late-stage preclinical testing and early clinical trials

**TABLE 11 Regulated practices for consideration when taking a neo-organ to clinical testing**

| | Regulation |
|---|---|
| Good tissue practices | 21 CFR 1271 |
| Good manufacturing practices | 21 CFR 210 and 211 |
| Good laboratory practices | 21 CFR 58 |
| Good clinical practices | 21 CFR 50 |
| Quality systems regulations | 21 CFR 820[a] |

[a]Replaced cGMPs for neo-organ products regulated as devices.

for some products. Facility design considerations are outside of the scope of this chapter, but a GMP-qualified facility that can provide the required clean room processing, shipping and receiving procedures, and HVAC systems for airflow maintenance should be identified as part of the preclinical development phase of neo-organ product-development strategies.

Production of a neo-organ product for preclinical testing can be regulated by a number of different guidelines depending on the combination product composition (Table 11). Good tissue practices and good manufacturing practices established in the preclinical testing phase may go through an IND/ BLA under 21 CFR 312/601 or IDE/PMA 21 CFR 812/814. Preclinical testing programs using the final neo-organ product or process may require that the test product be manufactured in a manner consistent with the conditions required for making the human neo-organ.

When designing the preclinical testing program for a neo-organ product, consideration is given to not only the clinical testing phase but also the bioprocessing activities required to manufacture a neo-organ product. Testing of any neo-organ product requires defining the bioprocess approach that will be used in early clinical testing. Without this planning, bioprocesses used in the preclinical testing may be dissimilar to those used in the clinical trials. Differences in the process may impact final neo-organ product characteristics to such a degree that release and in-process testing results established in the preclinical stage of development no longer apply to the product being tested in the clinical trials. Such an outcome may render the preclinical test results invalid or inappropriate for the product intended for clinical testing.

## 8 BRINGING NEO-ORGAN PLATFORMS TO CLINICAL TESTING

As discussed above, technology platforms that intend to recapitulate a tissue (e.g., skeletal muscle, bone, cardiac muscle) or an organ may address a range of unmet medical needs from simple cosmetic defects (tissue-focused technologies) to life-threatening maladies (organ and organ system replacement). Bringing a neo-organ technology to clinical testing may rest on the scope of

unmet medical need and availability of alternative therapies. The array of available alternative therapies influences the early testing strategy of a particular product by determining comparable products to be evaluated, selection of animal models, appropriate endpoints, and amount of preclinical information needed to enter into clinical testing. Ultimately the risk–benefit analysis between safety and functionality of the prototype product and other available products has a direct influence on the ability to test the product in human trials.

Tissue-focused technologies, such as bone and tendon repair, may move through preclinical development into clinical trials via routes that have been established by previous successes (e.g., Depuy's Restore®). If animal models and alternative therapeutic approaches are established, comparing the benefit of a proposed product to an existing therapy may be an one approach to potential clinical testing. Ultimately comparing the benefit of a complex tissue or neo-organ versus the "gold standard" commercial product or surgical therapy is the foundational rationale for evaluating potential human use.

Neo-organ technologies offer the promise of offsetting the worldwide organ shortage. Despite this great promise the pathway to clinical testing for a product that replaces an entire organ is the least clearly defined. Although definitive clinical benefit, discernable endpoints, established animal models, and delivery mechanisms may exist, the procedures for "connecting" the neo-organ product to other parts of the body pose substantial development hurdles and require extensive preclinical testing programs.

The complexity of whole organ replacement by neo-organ products spans defining what is actually being replaced through defining what ancillary products may be needed if all organ functions are not recapitulated by the neo-organ product. Traditional therapeutic approaches have generally focused on one pathway or target (e.g., pharmaceutical) or possibly two therapeutic benefits such as structural and functional restoration (e.g., cartilage repair products). However, products that replace an entire organ (e.g., kidney) or body part (e.g., limbs) will need to consider broad functional testing of both exocrine/excretory, metabolic, structural, and endocrine functions before clinical testing can be considered.

The transition to clinical testing of more complex neo-organ products will have commensurate preclinical testing requirements to demonstrate not only the functionality of each component being replaced or augmented but also the biological responsiveness of the integrated organ to native homeostatic mechanisms (e.g., integration with blood pressure or glucose control). Matters such as building percutaneous conduits and controlling skin infections and/or biofilms may be substantial development hurdles for the use of products outside the body. For products intended for use inside the body, clinical testing may hinge on finding solutions for vascular connections, waste product release pathways (e.g., urinary tract and GI), and clinical monitoring of neo-organ development. Similarly the length of time it takes to achieve clinical outcome may need to be determined before clinical testing can be considered.

Biosensors and integration of biosensors with neo-organ products are becoming a reality. Moving into clinical testing with these products requires definition of "recovery pathways" in the event of product failure. Adjunct therapies to be applied concurrently with neo-product use if all organ functions are not replaced with the neo-organ itself need to be defined. An understanding of neo-organ durability/longevity and replacement methods in the event that the neo-organ "wears out" needs to be developed. Finally, developing an understanding of product failure frequency and rate is important for proper clinical management.

Despite these hurdles, the lure of replacing an entire organ is considerable. The benefit to society of replacing a kidney or pancreas would be unimaginable. As technologies for in vitro organ growth advance and regenerative templates for entire organs are pioneered (e.g., through such technologies as organ printing), the potential to replace, regenerate, repair, and restore entire organ systems is being considered. Regenerative medical approaches may yield solutions for devastating human conditions including congenital agenesis, cancer, degenerative disorders, and infectious diseases. However, entry into clinical testing with such products has not yet been defined.

## 9 CONCLUSIONS

It is inevitable that, in the future, neo-organ products will represent an important class of treatments for patients needing neo-organ and tissue replacements, especially for many diseases in which available therapeutic options offer suboptimal outcomes. These products have the potential to satisfy significant unmet medical needs with an almost unimaginable benefit—a cure, not just a treatment. Currently regenerative medical products hold out the hope that downside risk is limited, since they reduce or even eliminate the potential for rejection without the onerous cost and difficulties associated with immunosuppressive therapy. Autologous-homologous products represent the best example. These products may offer unmatched risk–benefit profiles with the potential for rapid approval and introduction into the appropriate patient populations. Anticipated results include reductions in health care costs and substantial patient benefits, particularly when there are no medically acceptable alternatives.

The path to clinical entry has been partially paved for some of these breakthroughs, especially those that can emerge from application of established development strategies. It is conceivable that neo-organ products can be brought to market more rapidly and efficiently than traditional medical products (e.g., pharmaceuticals) and offer unknown advantages by their customized and individually tailored nature. Furthermore logistical advantages occur with patient-specific neo-organs, so well-defined studies (rather than time-consuming and costly large-scale preclinical studies to define unknown risks), smaller trial sizes (due to the customized nature of the products), and postregistration

follow-up (these products, once implanted, become part of the patient) can all be considered in the development of this transformational technology.

## ACKNOWLEDGMENTS

The authors thank Belinda Wagner for her guidance and assistance in helping us to present this new material in a clear and concise fashion. We would also like to thank David Watson for gathering information needed to compile this chapter.

## REFERENCES

1. Nelson JW. The lone eagle as medical researcher. *Navy Med* 2003;94:18–22.
2. Viola J, Bhavya L, Grad O. Emergence and evolution of a shared concept. In *The Emergence of Tissue Engineering as a Research Field*. Washington, DC: National Science Foundation, 2003. from, http://www.nsf.gov/pubs/2004/nsf0450/emergence.htm
3. FDA. Is the product a medical device? *Device Advice*, 1998. http://www.fda.gov/cdrh/devadvice/312.html#contents
4. FDA. 21 CFR Part 3: Product Jurisdiction, 2006. http://www.fda.gov/oc/ombudsman/part3&5.htm
5. FDA. Guidance for Industry and FDA Staff: How to Write a Request for Designation, 2005. http://www.fda.gov/oc/combination/howtowrite.html
6. FDA. Office of Combination Products, 2006. http://www.fda.gov/oc/combination/
7. US Department of Health and Human Services. 2020: A New Vision—A Future for Regenerative Medicine, 2006. http://www.hhs.gov/reference/newfuture.shtml
8. Anderson JM. Biological responses to materials. *Ann Rev Mater Res* 2001;31:81–110.
9. ASTM F981-99. Standard Practice for Assessment of Compatibility of Biomaterials for Surgical Implants with Respect to Effect of Materials on Muscle and Bone. 1999.
10. Cotran RS, Kumar V, Collins T. *Robbin's Pathologic Basis of Disease*, 6th ed. Philadelphia: WB Saunders, 1999.
11. Lewis RW, et al. Recognition of adverse and nonadverse effects in toxicity studies. *Toxicol Pathol* 2002;30:66–74.
12. Long G. Recommendations to guide determining cause of death in toxicity studies. *Toxicol Pathol* 2004;32:269–70.
13. Ratner BD, et al. *Biomaterials Science: An Introduction to Materials in Medicine*. San Diego: Elsevier Science/Academic Press, 1996.
14. Rubin E, et al. *Rubin's Pathology*. 4th ed. Philadelphia: Lippincott/Williams Wilkins, 2005.
15. Preti RA. Bringing safe and effective cell therapies to the bedside. *Nat Biotechnol* 2005;23:801–4.

16. Weber DJ. Navigating FDA regulations for human cells and tissues. *BioProcess Int* 2004;2:22–7.
17. Schmidt D. et al. Engineering of biologically active living heart valve leaflets using human umbilical cord-derived progenitor cells. *Tissue Eng* 2006;12:3209–21.
18. Zuk PA, et al. Multilineage cells from human adipose tissue: implications for cell-based therapies. *Tissue Eng* 2001;7:211–28.
19. Haider H. Bone marrow cells for cardiac regeneration and repair: current status and issues. *Expert Rev Cardiovasc Ther* 2006;4:557–68.
20. Ikada Y. Challenges in tissue engineering. *J R Soc Interface* 2006;3:589–601.
21. FDA. Required Biocompatibility Training and Toxicology Profiles for Evaluation of Medical Devices: May 1, 1995 (G95-1). http://www.fda.gov/cdrh/g951.html
22. FDA. 21 CFR part 820: Quality System Regulation, 2006, http://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfcfr/CFRSearch.cfm?CFRPart=1271
23. FDA. Proposed Approach to Regulation of Cellular and Tissue-Based Products, 1997. http://www.fda.gov/cber/gdlns/celltissue.txt
24. FDA. Biological Products Regulated under Section 351 of the Public Health Service Act; Implementation of Biologics License; Elimination of Establishment License and Product License, 1999. http://www.fda.gov/cber/rules/elapla.pdf
25. FDA. Guidance for Industry FDA Staff: Minimal Manipulation of Structural Tissue Jurisdictional Update, 2006. http://www.fda.gov/oc/combination/manipulation.html
26. ISO 10993-5. Biological Evaluation of Medical Devices. Part 5: Tests for In vitro Cytotoxicity. 1999.
27. FDA. 21 CFR part 1271: Human Cells, Tissues, and Cellular and Tissue-Based Products, 2006. http://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfcfr/CFRSearch.cfm?CFRPart=1271
28. Halme DG, Kessler DA. FDA regulation of stem-cell-based therapies. *N Engl J Med* 2006;355:1730–5.
29. FDA. Medical Device Quality Systems Manual: A Small Entitiy Compliance Guide, 1997. http://www.fda.gov/cdrh/dsma/gmpman.html

# GLP Requirements and Current Practices

TANYA SCHARTON-KERSTEN

**Contents**

1 Background for GLP Regulations
2 History and Scope of Good Laboratory Practice (GLP) Regulations
3 Common Elements of GLP Regulations
4 GLP Regulations and Compliance in the United States
5 Global GLP Regulations and Compliance
6 Comparison of FDA and OECD GLP Regulations
7 Pragmatic Approach to GLP Compliance
8 Site Selection for Preclinical Studies
9 Conclusions
References

## 1 BACKGROUND FOR GLP REGULATIONS

Regulations and guidelines for the conduct of preclinical testing of drugs and biologics were introduced by US and other governments to ensure the quality and integrity of preclinical data submitted in support of clinical trial applications (IND, CTA, etc.) and/or marketing approvals (BLA, NDA, MAA, etc.). The good laboratory practices (GLP) regulations and guidelines foster mutual acceptance of test data in the international community, serve to minimize redundant testing, minimize animal use, and maximize protection of human health. The purpose of this chapter is to provide guidance on compliance with GLP provisions in the conduct of preclinical (nonclinical) laboratory studies of biopharmaceuticals. Application of the US promulgated GLP provisions

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

(21 CFR 58) is the primary focus, with more limited reference to global GLP regulations and compliance.

Within the United States, the Food and Drug Administration (FDA) regulates the conduct of preclinical laboratory studies under part 58, Good Laboratory Practice for Nonclinical Laboratory Studies, of title 21 of the Code of Federal Regulations (21 CFR 58). The most current version of the GLP regulations may be accessed through the US National Archives and Records Administration Web site [1]. A similar but distinct set of GLP regulations is used by the US Environmental Protection Agency (EPA) to govern laboratory studies of pesticides [2] and other toxic chemical substances [3].

The GLP provisions promulgated by the FDA specifically apply to preclinical safety data generated in test systems both in vitro and in vivo. In section 58.3(d) of the GLP regulations, the FDA defines the term nonclinical laboratory study as "in vivo or in vitro experiments in which test articles are studied prospectively in test systems under laboratory conditions to determine their safety." Examples include, but are not limited to, physical-chemical testing, toxicity studies, mutagenicity studies, tissue residue depletion studies, and analytical and clinical chemistry testing. Basic exploratory or proof-of-principle studies carried out to determine whether a test article has any potential utility or to determine physical or chemical characteristics of a test article are excluded from the US GLP regulations. However, complete exclusion of many elements of GLP preclinical animal studies has become relatively unusual as both industry and the regulatory community recognize the ethical, scientific and long-term cost saving benefits associated with application of GLP provisions. Animal use is reduced when study designs are rigorously reviewed. Careful execution of studies also minimizes the need to repeat testing, thereby further reducing animal use and costs associated with test redundancy. Rigorous documentation during study conduct data collection facilitates communication of study results in both scientific publication and to regulatory agencies.

From the FDA perspective it is expected that all preclinical safety studies supporting the safety of research or marketing application (IND, CTD, BLA, NDA, etc.) will be performed in a manner consistent with GLP provisions and that Sponsors justify any nonconformance (21 CFR 312.23(a)(8)(iii)). Nonetheless, it is neither a regulatory requirement nor often feasible to conduct all preclinical studies within a drug (biologic) development program in 100% GLP compliance. The lack (or perceived lack) of an independent quality unit in small companies or academic organizations can prohibit 100% GLP compliance for internally conducted studies. Alternatively, technical resources for specialized studies, particularly those where both pharmacological and safety testing are combined; limited capacity, particularly for biologics testing; and prohibitive costs can impact outsourcing of studies to GLP knowledgeable contractors.

There are many misconceptions surrounding GLP principles. For example, sponsors often believe that the GLP provisions may be applied to the testing of samples from clinical studies. However, the introductions to the FDA GLPs

in 21 CFR part 58 and to the Organization for Economic Cooperation and Development (OECD) Principles of GLP state that applicable studies do not include studies utilizing human subjects or clinical studies or field trials in animals. The FDA and other regulatory authorities do require a demonstration of the quality of test data from clinical studies. However, this is by means of compliance with good clinical practice (GCP) and International Conference for Harmonization (ICH) provisions as well as conformance with CLIA (Clinical Laboratories Improvement Act). There is also a common but inappropriate use of the term GLP to refer to the quality of test material used in preclinical studies. Test material that is not fully GMP compliant should be stated as such. The term GLP cannot be used to indicate less than GMP material. Other common terminologies used to designate non-GMP material include, "research grade" or "pre-GMP" material.[1]

Application of GLP principles to testing of biopharmaceuticals, including cell therapies and infectious agent vaccines and therapies, can be difficult. Biological test materials often require special handling and/or less conventional animal models for preclinical safety testing. For example, development of vaccines and therapies for counterterrorism applications and for infectious diseases such as human immunodeficiency virus associated with the acquired immune deficiency syndrome (AIDS) is fostering novel vaccines approaches and therapies. Handling of the related test materials, and even the test systems, can require adherence to defined biosafety level (BSL) containment provisions [4]. Similarly the toxicity testing may require conduct in species other than rodents and dogs, which have been the conventional models for pharmaceuticals. As a consequence test facility options for biological materials are relatively limited when compared to that available for drug compounds. To accommodate the required studies, sponsors are often faced with either long wait times at test facilities, building GLP compliance capacity within their own organizations, bringing necessary scientific, permitting, and general biologic expertise to conventional test facilities, and/or educating university and smaller private laboratories on GLP provisions. In all cases it has become increasingly important for the sponsor to have a pragmatic understanding of domestic and global GLP provisions in order to select the test laboratory (qualification) and to ensure that study performance and reported data meet acceptable standards (monitoring) for both scientific validity and regulatory compliance.

The chapter aims to (1) provide an overview of US and global GLP regulations and (2) discuss pragmatic application of GLP elements to preclinical studies, particularly those where conventional 100% GLP compliance may not be feasible.

[1]While one may insert such terms in a study report, and CROs may accept such descriptions, it should be noted that the FDA and other regulatory agencies expect that the definitive safety studies be conducted with test article that is comparable to the product proposed for the initial clinical studies. Any differences, compliance (GMP/nonGMP) or other (manufacturing methods, product attributes), must be justified within the relevant regulatory submissions (e.g., IND or BLA/CTD).

## 2 HISTORY AND SCOPE OF GOOD LABORATORY PRACTICE (GLP) REGULATIONS

The Federal Food Drug and Cosmetics Act (FD&C) Act and Public Health Service (PHS) Act (biologics only) require that sponsors of FDA-regulated products submit evidence of their product's safety and efficacy in research (IND, IDE) and/or marketing applications (BLA, NDA, PMA). These products include human drugs and biological products, and human medical devices. FDA uses these data to answer questions regarding the toxicity profile of the test article, the observed no adverse effect dose level (NOAEL) in the test system, the risks associated with clinical studies involving humans or animals, the potential teratogenic/reproductive, genotoxic, carcinogenic, or other adverse effects of the article, and the level of use that can be approved.

The GLP regulations (21 CFR 58) were promulgated in the mid-1970s based on inspections of animal laboratories revealing unsuitable industry practices within studies submitted to the FDA [5–8]. The quality and integrity from such studies were not considered adequate to ensure product safety in accord with the FD&C Act, PHS Act, and other applicable laws. As a result of these findings the FDA promulgated the good laboratory practice (GLP) regulations. The proposed GLP regulations were published in the Federal Register of November 1976 [9]. The regulations establish standards for the conduct and reporting of animal laboratory studies and are intended to ensure the quality and integrity of safety data submitted to FDA.

The final rule for GLP regulations was published in the Federal Register of December 22, 1978 (43 FR 59986), and it became effective on July 20, 1979 [10]. Revisions to the cGLP regulations, to reduce regulatory and paperwork burdens, were proposed in the Federal Register of October 29, 1984 (49 FR 43530), with the final amendment published in the Federal Register of September 4, 1987 (52 FR 33768) [11,12]. Two ancillary documents published by the FDA—Guidance for Industry: Good Laboratory Regulations Management Briefings [13] and Good Laboratory Practices: Questions and Answers [14]—provide pragmatic information on application of the GLP regulations and the FDA's interpretations thereof. Answers to many common questions and conceptions, may be found within these documents. The FDA/ORA Bioresearch Monitoring Information Web page contains numerous US based GLP resources [15].

In May 1981 the member countries for the Organization for Economic Cooperation and Development (OECD) approved the OECD Principles of Good Laboratory Practice. The GLP Web site of the OECD contains detailed reference information and links for global GLP compliance [16]. Revised OECD GLP principles were adopted by the Council on November 26, 1997. The complete OECD GLP regulations are comprised of a series of guidance documents with the first document, No. 1 OECD Principles of Good Laboratory Practice, providing the overarching guidance for GLP compliance among

the member countries [17–30].[2] The OECD GLP provisions are the most commonly referenced global standards for GLP compliance.

Within the European Union, the European Commission (EC) has codified the GLP requirements for medicinal products in the Introduction and General Principles chapter of Directive 2003/63/EC [31]. Within this document it is stated that nonclinical (pharmacotoxicological) studies must be carried out in conformity with the provisions related to GLP laid down in Council Directives 87/18/EEC[3] on the harmonization of regulations and administrative provisions relating to the application of the principles of GLP and the verification of their application for tests in chemical substances; and 88/320/EEC[4] (corrigenda 1, 2, and 3) on the inspection and verification of GLP [32,33]. The OECD principles have been adopted by the European Union and published, in their revised form, in the appendix to Directive 2004/10/EC [32]. The GLP Web site of the EC contains detailed reference information and links for member state GLP compliance [34].

Establishment of global GLP standards as well as mutual recognition agreements permit sharing of data among regulatory authorities and reduces the need for redundant testing by product sponsors. Data generated in an OECD member country, in accordance with OECD test guidelines and the Principles of Good Laboratory Practice, are accepted by regulatory agencies of other OECD member countries. The US FDA has also established specific Memorandum of Understanding (MOU) for GLP compliance with several European and non-European countries. The European Union has concluded Mutual Recognition Agreements for GLP provisions within its member states and with Israel, Japan, and Switzerland. By means of the Treaty of the European Economic Area (EEA) of December 13, 1993, the European Regulations and Directives also apply to Iceland, Liechtenstein, and Norway. The Web sites of the various national GLP compliance and monitoring bodies may be consulted to verify mutual data recognition [35].

## 3 COMMON ELEMENTS OF GLP REGULATIONS

Both the OECD and US FDA GLP regulations are intended to establish minimal standards for conducting and reporting preclinical laboratory studies. Study organization, personnel, control of test and reference materials, the test system, data collection, reporting, communication, and storage are all covered within the domestic and global GLPs. The required period(s) of storage for raw data, documentation, protocols, specimens, and interim and

[2]The OECD member countries are Australia, Austria, Belgium, Canada, the Czech Republic, Denmark, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Italy, Japan, Korea, Luxembourg, Mexico, the Netherlands, New Zealand, Norway, Poland, Portugal, the Slovak Republic, Spain, Sweden, Switzerland, Turkey, the United Kingdom, and the United States.
[3]Directive 2004/10/EC replaces Directive 87/18/EEC.
[4]Directive 2004/9/EC has replaced Directive 88/320/EEC as of March 11, 2004.

final reports is defined within the GLP regulations (e.g., 21 CFR 58.190 and 195).

The US FDA GLP regulations (21 CFR 58) are comprised of 11 subparts (A–K), beginning with a definition of the scope and terminology (subpart A, General Provisions), followed by minimal requirements for the following:

- Organization and personnel (subpart B)
- Facilities (subpart C)
- Equipment (subpart D)
- Test facilities operation (subpart E)
- Test and control articles (subpart F)
- Protocol for conduct of a nonclinical laboratory study (subpart G)
- Records and reports (subpart J)

The FDA regulations end with a discussion of procedures for Disqualification of Test Facilities (subpart K).

A critical feature of the GLP regulation is the integral role of an independent quality assurance unit (QAU). Per 21 CFR 58.35 the quality assurance unit, an entity that is entirely separate and independent of the personnel engaged in the direction and conduct of the study, is responsible for monitoring each study to ensure management that the facilities, equipment, personnel, methods, practices records, and controls are in conformance with the GLP provisions. Specific responsibilities include maintenance of a master schedule, maintenance of copies of all protocols, inspection of each laboratory study with documented communication of any problems to management and the study director, submission of periodic status reports on each study to management and the study director, determination that no deviations from approved protocols or SOPs were made without proper authorization and documentation, review of the final report to ensure accuracy of the methods and reported data, preparation and signature of a statement to be included in the final report with the date of inspections and findings reported to management, and maintenance of records for review by authorized FDA employees during inspections. In many cases it is either the lack of internal resources to perform these activities or a lack of understanding of the nature of the QAU responsibilities that preclude a sponsor from performing GLP regulated studies in-house. Moreover the need for integrated quality assurance within *all* GLP studies primarily precludes even the best planned and documented studies from claiming full GLP adherence if the responsibilities are not considered and implemented prior to study initiation.

Similarly the OECD GLP regulations (GLP Principles No. 1 in the series) begin with the definition of scope and terminology (section I, Scope and Definition of Terms), followed by 10 subparts (section II) providing minimal requirements for the following:

- Test facility organization and personnel (subpart 1)
- Quality assurance programme (subpart 2)
- Facilities (subpart 3)
- Apparatus, material, and reagents (subpart 4)
- Test systems (subpart 5)
- Test and reference items (subpart 6)
- Standard operating procedures (subpart 7)
- Performance of the study (subpart 8)
- Reporting of study results (subpart 9)
- Storage and retention of records and materials (subpart 10)

Additional guidance is found in the remaining series of documents (Nos. 2–14) of the OECD GLP series, which covers aspects such as management of multisite studies [29] and responsibilities of the study director [25] and the sponsor [27].

## 4 GLP REGULATIONS AND COMPLIANCE IN THE UNITED STATES

The FDA relies on documented adherence to GLP requirements by preclinical laboratories in judging the acceptability of safety data submitted in support of research and/or marketing permits. The FDA monitors compliance with the GLP requirements through a program of regular facility inspections and targeted data audits at private, government, and public laboratories. A precedent for the current inspection program was established in December 1976 when the FDA began a pilot inspection program of representative testing programs to ascertain how well toxicology laboratories (sponsor, contract, and university facilities) were adhering to the proposed GLP regulations. The results of the inspection program were published in OPE Study 42, Results of the Nonclinical Toxicology Laboratory Good Laboratory Practices Pilot Compliance Program [36].

Laboratory compliance with GLP requirements is still monitored by the FDA through a facility inspection and data audit program according to criteria described in chapter 48 of its Compliance Program Guidance Manual [6]. The FDA conducts surveillance and directed inspections. Surveillance inspections are periodic, routine determinations of a laboratory's compliance with GLP regulations. These inspections include a facility inspection and audits of on-going and/or recently completed studies. Directed inspections are conducted to achieve a specific purpose, such as verifying the reliability, integrity, and compliance of critical safety studies being reviewed in support of pending applications or investigating issues involving potentially unreliable safety data and/or violative conditions brought to the FDA's attention. At

the end of each inspection the facility is issued an Establishment Inspection Report (EIR). Inspectional observations may also be listed on a form FDA-483. Findings listed on the 483 represent inspectional observations that have not been detected and corrected by the firm through its internal procedures and are either not considered minor in nature or are more than one-time occurrences that may impact on the firm's operations, study conduct, or data integrity. All findings, regardless of reporting on the 483, are listed on the EIR and are discussed with the firm's management [6]. Depending on the nature of the findings a warning letter may or may not follow the issuance of the 483. A warning letter differs from a 483 in that the warning is issued from the FDA's district office after higher level FDA officials have concluded that the 483 and EIR findings warrant further formal notification to the firm where, the FDA believes, serious violations exist. In most cases a firm will prepare a formal reply to FDA within one to two weeks of receiving either a 483 or warning letter. The warning letter serves to prompt voluntary corrective action at the test facility or firm and to establish a history of warning(s) should further regulatory action be required at a later date (i.e., injunction or prosecution). For reviews of the history and present use of the FDA-483 as well as the procedures for firms responding to form 483 and warning letters, the reader is referred to articles by two former compliance officials [37,38]. Receipt of warning letters by GLP laboratories is relatively uncommon, whereas many larger and longstanding test laboratories have received a 483s at one time in their tenure.

The FDA Web site contains a list of the testing facilities that have been inspected since October 1, 1989 [39]. The list provides the current name and former (if known) names of the testing facility, the dates the facility has been inspected, and the classification of the inspection results. Active toxicology laboratories, university laboratories, foreign laboratories, and inactive laboratories are listed. The list is updated quarterly. The most common outcome of inspections include NAI (no action indicated), VAI (voluntary action indicated), and OAI (official action indicated). For inspections with an OAI classification, regulatory and administrative actions will be recommended. Warning letters may or may not be issued with an OAI classification.

Results of FDA inspections are available through a variety of sources. The official classification of a facility's inspection is reported on the FDA's compliance Web site [39], which contains the results from current and all previous inspections. Alternatively, many contract laboratories are quite willing to provide a regulatory inspection history for the past five years and/or the actual EIRs or 483s with their response letters to FDA upon request. Finally, the EIRs may be requested through the Freedom of Information Act (FOIA) [40]. The latter option, while a formal possibility, can be a very lengthy process. Unlike the European GLP monitoring programmes, the US FDA does not issue certificates or other documents of compliance. Warning letters are immediately available, upon issuance, on the FDA Web site [41].

## 5 GLOBAL GLP REGULATIONS AND COMPLIANCE

From a compliance perspective, Directive 2004/9/EC [33] lays down the obligation of the member states to designate the authorities responsible for GLP inspections in their territory. The Directive requires that the OECD Revised Guides for Compliance Monitoring Procedures for GLP and the OECD Guidance for the Conduct of Test Facility Inspections and Study Audits must be followed during laboratory inspections and study audits. Directive 2004/10/EC [32] requires member states to take all measures necessary to ensure that laboratories carrying out safety studies on chemical products (cosmetics, industrial chemicals, medicinal products, food additives, animal feed additives, pesticides) comply with the OECD Principles of Good Laboratory Practice.

Test facilities within the European Union are required to notify their national GLP monitoring authority as soon as they claim to conduct studies in accordance with the OECD GLP principles. All laboratories claiming GLP compliance are entered into a periodic national compliance program to monitor GLP adherence. The core of the compliance-monitoring program is a periodical inspection scheme, with an intended frequency of one inspection every two years. As circumstances require, the inspection schedule may be intensified. If a test facility is considered to operate in compliance with the OECD GLP principles, an Endorsement of Compliance will be issued, as indicated in Directive 2004/9/EC. This endorsement only states that the facility was operating in compliance at the time of the inspection. It may be used for inclusion in study reports to indicate that the test facility has been inspected for GLP compliance with good results. The endorsement of compliance also states which areas of expertise are operating in compliance with GLP provisions.

Foreign laboratories may apply for inspection by various European monitoring programs. The monitoring authorities are generally willing to accommodate such requests if (1) the facilities submit evidence that data produced by them are or will soon be submitted to the regulatory (receiving) authorities, (2) there is no recognized monitoring authority in the country, and (3) the requesting facility is willing to bear the full costs of the inspection. The GLP Web site for the European Union [34] contains information on the application of GLP Directives in the European Union by the member states including the legislative adoption or "transposition" of the GLP directives, contact information for the national authorities on GLP, and public lists of inspected test facilities.

The Japanese drug regulatory body is the Ministry of Health, Labor and Welfare (MHLW). The standards concerning the nonclinical safety testing of pharmaceuticals (good laboratory practices) have been in place since 1982 in the form of administrative instruction. Such standards have been enforced since April 1997 as ministerial ordinances. One of the Ministry's 11 bureaus, the Organization for Pharmaceutical Safety and Research (OPSR/KIKO or Drug Organization) is responsible for GLP compliance monitoring. Thus, as

in the United States and Europe, pivotal safety studies must comply with good laboratory practice (GLP) standards. For those studies to which GLP applies, the preclinical testing facilities will be subject to GLP compliance inspections that occur after an application is submitted. Based on the findings of the KIKO, a facility will receive one of three ratings: (1) class A for compliance with GLP, (2) class B indicating some improvements are possible, but the effects of noncompliance on the data's reliability are not considered significant (improvements are encouraged), or (3) class C where there is noncompliance with GLP. The MHLW will accept the test results from facilities receiving a class A or B rating. Mutual acceptance of GLP inspection results and data from several countries is provided through the bilateral agreements. Japan is one of the three (tripartite) members of the International Conference on Harmonization (ICH) and a member state of the OECD. An overview of the competent authorities for GLP compliance and regulations within Japan, as well as the applicable laws, regulations, and administrative provisions stipulating the principles of GLP verification and confirmation, may be found in Sectoral Annex A of the January 1, 2002, Japan-EC Mutual Recognition Agreement [42].

## 6 COMPARISON OF FDA AND OECD GLP REGULATIONS

The FDA has published a reference table with a side-by-side comparison of OECD, 21 CFR 58 (FDA) and US EPA GLP provisions [43]. As stated throughout the chapter, the objectives of the GLP principles are similar, but the methods of implementing GLP provisions, the programs for ensuring GLP compliance, and the scope of the provisions vary considerably between regions. With respect to scope, the GLP provisions of the OECD are generally applicable to all nonclinical testing of chemicals, whereas the 21 CFR 58 regulations pertain to food and color additives, animal food additives, human and animal drugs, medical devices for human use, and biological products and electronic products (21 CFR 58.1).

Another important difference between the US FDA and OECD regulations is that the former is, in a formal sense, limited to collection of toxicological data, whereas the OECD regulations are applied to both toxicological and pharmacological data. Sponsors considering submission to non-US agencies should strongly consider collection of both data sets in compliance with relevant GLP regulations. In practice, current industry practice even in the United States is to collect both toxicological and pharmacological data in compliance with applicable aspects of the GLP regulations, such as under a protocol, with a final report, with a study director, and with some QA oversight, whenever possible.

The programs for assuring GLP compliance have already been described above in Sections 4 and 5. In summary, the European Union guided by OECD requires registration of all facilities claiming conduct of studies under GLP in a program for GLP compliance. The identified laboratories are inspected by

regional authorities on a periodic basis, and they are issued certificates of endorsement that support GLP compliance in technical disciplines within the European Union. In contrast, the US FDA relies on documentation of GLP compliance by the sponsor. The FDA periodically inspects some but not all facilities claiming GLP compliance and study data thereof. The FDA does not issue certificates or other documents of compliance. Establishment inspection reports and/or FDA 483s are often available from the laboratories and are always available by request through the Freedom of Information Act.

## 7 PRAGMATIC APPROACH TO GLP COMPLIANCE

As noted earlier in the chapter, it is not always possible or feasible to conduct studies in 100% compliance with GLPs. From a regulatory perspective the sponsor is responsible for describing and justifying areas of noncompliance. Section 21 CFR 312.23(a)(8)(iii), which pertains to IND content and format states: "For each nonclinical laboratory study subject to the good laboratory practice regulation under part 58, a statement that the study was conducted in compliance with the good laboratory practice in part 58, or if the study was not conducted in compliance with the regulation, a brief statement of the reason for non-compliance." In the case of novel biopharmaceuticals or pharmaceuticals, it may not always be possible to begin the clinical development program with preclinical testing run under full GLP compliant conditions. For example, the sponsor may lack a quality assurance unit internally, or external laboratories may not have animal models suitable for conducting the primary study. Ultimately the decision on site selection for preclinical studies is based on a matrix of considerations, and rational departures from full GLP compliance may be necessary.

The lack of full GLP compliance, particularly for preclinical studies collected early in a development program, does not inhibit data from being submitted and reviewed by regulatory agencies, provided that the sponsor adequately describes and justifies areas of noncompliance. The FDA in its 1981 Question and Answer Document [14] clearly indicates that even safety supporting preclinical studies with areas of noncompliance may be acceptable provided that the quality and integrity of the study has not been impacted [14]:

Q: Does FDA reject nonclinical laboratory studies that have not been conducted in full compliance with GLPs?

A: Not necessarily. The GLP compliance program provides guidance on this issue. For FDA to reject a study it is necessary to find that there were deviations from the GLPs and that these deviations were of such a nature as to compromise the quality and integrity of the study covered by the agency inspection.

**TABLE 1 Examples of GLP (non)compliance statements that may be included in preclinical study reports or within FDA submissions**

| |
|---|
| This study will not be/was not conducted in full compliance with part 58 of 21 CFR. However, the study will be/was conducted according to the study protocol and standard operating procedures of _____. There is no formal involvement of a quality assurance unit |
| Currently acceptable practices of good animal husbandry will be followed, e.g., Guide for the Care and Use of Laboratory Animals: DHHS Publication No. (NRC) 86-23, revised 1996. |
| This study will not be conducted in full compliance with part 58 of 21 CFR. However, the study will be conducted according to the study protocol and standard operating procedures of _____. There is no formal involvement of a quality assurance unit. |
| Currently acceptable practices of good animal husbandry will be followed, e.g., Guide for the Care and Use of Laboratory Animals: DHHS Publication No. (NRC) 86-23, revised 1996. |
| This study was not performed in strict compliance with the US FDA good practice regulations (GLPs) as amended (21 CFR, part 58). GLPs were followed, except that no in-life critical phase audits were conducted. Efforts were made to operate with documented procedures, standard operating procedures, and the study protocol. No circumstances occurred during the study that might have affected the quality or integrity of the data. The raw data contained in this report were subjected to a final audit. |

*Source*: Adapted Cavagnaro [44].

Nonetheless, a movement toward full compliance is expected during the evolution of the development program. During the interim, reports from such studies should state the reasons for noncompliance and the sponsor must justify the noncompliance within submissions to the FDA. Recently a framework for the staged application GLP provisions to preclinical testing of novel biologics has been presented in the context of gene therapy product development [44]. Sections of the presentation have been adapted to provide detailed examples of compliance statements that might be used to describe studies not conducted in full compliance with GLP provisions (Table 1) and to describe regulatory expectations for GLP compliance of pharmacology, biodistribution, and toxicology studies (Table 2). The latter table also indicates elements of the GLP provisions that can be applied such as use of a study plan, documented test methods, assignment of a single point of control, and inclusion of quality assurance oversight.

## 8 SITE SELECTION FOR PRECLINICAL STUDIES

Preclinical studies may either be conducted in house (within a sponsor's own laboratories) or outsourced to a contract laboratory or grantee. When a

**TABLE 2 Framework for application of GLP provisions to novel biologics: Example for gene transfer research**

| Type of Study | Level of Compliance | Example Language or Comments |
|---|---|---|
| Pharmacology—in vitro or in vivo activity and/or mechanism | By convention non GLP | • Use of laboratory procedures<br>• Novel in vitro test systems/animal models<br>• Use of research grade vector preparation and possibly more than one lot<br>• Limited information on product characteristics |
| Proof of concept—in vivo assessment of efficacy and safety in the same model | In accordance with the principles of GLP . . . | • Study protocol<br>• Animal model of disease (unvalidated)<br>• Research grade or early GMP preparation<br>• Limited or no quality assurance oversight |
| Biodistribution studies—PCR testing, etc. | In compliance with GLPs with the exception of . . . | • Study protocol may be a part of the toxicology study<br>• Methodology/standardization of assays evolving (level of sensitivity improving)<br>• Pre-GMP test material<br>• Laboratory procedures; draft SOPs<br>• QA review of the final report only |
| Pilot toxicology studies—selection of relevant species | In compliance with GLPs with the exception of . . . | • Study protocol/outline<br>• Limited numbers; not statistically relevant<br>• Research/representative test material not GMP/clinical<br>• Limited stability data for test article |
| Pivotal toxicology studies—support route of administration, dose, regimen, patient population selection, etc. | In full compliance with GLP provisions | • Study protocol and amendments<br>• SOPs<br>• Test material clinical GMP or pre-validation lot<br>• Stability and product characterization data<br>• QA audit all phases |

*Source*: Adapted Cavagnaro [44].

decision is made to outsource, it is critical for the sponsor to remember that GLP compliance and scientific integrity are ultimately the sponsor's responsibility. When the service of a consulting laboratory, contractor, or grantee is used, the final responsibility for the quality of the study still resides with the sponsor. It is the sponsor's obligation to notify the other party that the service is part of a preclinical laboratory study that must be conducted in compliance with GLP provisions.

Once the a decision has been taken to outsource a preclinical study, contractor selection based on quality and technical standards is a key element of successful study conduct. Generally, one begins with identification of laboratories that are capable of conducting the proposed scope of work, followed by quality audits to ensure that the test facility has can conduct the study(ies) in a GLP compliant manner. The FDA inspection Web site is often a good resource for identifying potential contractors, particularly when used in combination with referrals from industry representatives conducting similar studies. Whenever possible, the screening of contract research organizations (CROs) should go beyond business development representatives to potential study directors or operational heads to ensure that technical needs, containment requirements (BSL1-4), and/or desired animal models are accessible. To this end it is advisable to communicate as many details as possible regarding the type of test article, animal species/test system, duration of study, and so forth, to ensure that the information provided is accurate. When the list of potential contractors has been narrowed down to two or three, a more formal qualification process should be initiated. Generally, this begins with a confidentiality agreement (CDA) followed by a qualification exercise, consisting of both an offsite and onsite review of the facility's quality and technical expertise.

An example of a contractor selection plan for a preclinical study requiring GLP compliance is provided in the tables below. Tabular (checklist) type of examples are provided for:

- Activities to conduct prior to scheduling the audit (Table 3),
- Scheduling the audit (Table 4),

**TABLE 3 Example checklist of test facility selection and inspection tools prior to scheduling audit**

| Activity | Complete |
|---|---|
| • Execute confidentiality agreement | |
| • Communicate a detailed scope of work | |
| • Identify one or more potential study directors | |
| • Identify the primary test facility and all offsite (secondary or subcontract) facilities that could be utilized[a] | |
| • Collect physical addresses of all potential facility areas for study | |

[a]Subcontract and offsite areas must be identified at this stage.

**TABLE 4 Example checklist of scheduling audit one to four weeks prior to visit**

| Activity | Complete |
|---|---|
| • Set audit schedule based on scope of work and physical sites proposed for study conduct[a] | |
| • Provide proposed agenda and previsit document request (below) | |
| • Schedule travel and communicate travel schedule to contract laboratory | |

[a]Six to 8 hours for primary facility and additional time, including transportation, for offsite or satellite facilities.

**TABLE 5 Example checklist of previsit document request two weeks prior to visit**

| Activity | Received |
|---|---|
| • Company prospectus; if available | |
| • Regulatory inspection history[a] | |
| • Index of Standard Operating Procedures for relevant functional areas | |
| • Organizational Diagram for relevant functional areas | |
| • CV of proposed study director(s) and principle investigator(s) | |
| • Test item receipt instructions and example intake form | |
| • SOP(s) for protocol deviation and amendments | |
| • SOP(s) for in-phase study inspections | |

[a]EU GLP compliance certificate or US inspection history with most recent Establishment Inspection Report.

- Previsit document requests (Table 5),
- Audit agenda (Table 6)
- Structure of an audit report (Table 7)

Issues that are of particular relevance for biopharmaceuticals include the following:

- Time constraints imposed by transport of test material to the test site that may involve permits from either the CDC or USDA, or other national requirements when test material or the contract laboratory is located outside of the United States.
- Identification of appropriate test article receipt and storage conditions at the test facility with consideration of any required facility departures from conventional procedures for test material receipt, handling, and storage.
- Review of standard facility procedures for test material preparation for dosing and/or delivery. Test materials for biopharmaceuticals tend to be very limited in quantity and handled in much smaller quantities than that used for pharmaceuticals. Many test facilities are not used to handling

**TABLE 6 Example checklist of audit agenda one week prior to visit**

| Activity | Time Frame |
|---|---|
| *Opening meeting/introductions* | 5–10 minutes |
| Suggested attendees: proposed study director, QA representative, client services representative, management representative | |
| *Sponsor overview and proposed scope of work* | 15 minutes |
| *Brief contractor overview of company and services* | 15 minutes |
| *Discussion of GLP inspection history; compliance issues* | 15 minutes |
| *Facility tour and document request during tour* | 2–4 hours |
| • Test and control article receipt/storage/archival/distribution<br>• Test article formulation<br>• Test article analysis<br>• Animal husbandry support areas<br>• Animal husbandry and treatment areas<br>• Necropsy<br>• Clinical pathology<br>• Histopathology<br>• Other relevant laboratory areas<br>• Computer systems/backup<br>• Document archival<br>• Tissue and slide archival | |
| *Record review* | 1–1.5 hour |
| Training records, environmental records, AAALAC or equivalent certification, feed and water monitoring data, animal health records, etc. | |
| *Review assay validation(s) or qualifications, as needed.* | |
| *Review procedures for project management within the contract facility* | |
| SOPs | 1–1.5 hour |
| Close out | 15–30 minutes |

small volumes and do not routinely use the disposable and small quantity equipment (e.g., micropipets, plasticware) that are second nature in cell and molecular biology laboratories.

- Availability of stability data to support the operational conditions that will need to be used by the contract test facility to execute the sponsor's proposed activities. The test facility may require several hours or days to dose a series of treatment groups. Often times vaccines may only be stable

**TABLE 7 Example of the structure of an audit report**

| |
|---|
| • Logistical details |
| Date of visit, sponsor and contractor participants with contact information, addresses of all facilities visited |
| • Overview of findings (critical, major, minor) |
| • Facilit(ies) description |
| • GLP inspection information |
| • Detailed narrative of functional areas reviewed and all findings (critical, major, minor) with classification rationale within relevant narrative |
| • List of SOPs reviewed with comments as needed |
| • List of non-SOP documents with comments as needed |
| • Signature page for auditor(s) and date of finalization of report |
| *Optional* |
| • Recommendations for corrective and preventative actions |
| • Space for contractor responses to be included in body of report |

at room temperature for several hours. This information needs to be communicated early to the test facility such that appropriate adjustments to test material handling can be considered and incorporated into the protocol or supporting study specific documents.

The purpose of the qualification/audit is for the sponsor to investigate the standard practices of the contractor. Quality as well as technical capabilities should be evaluated. One should determine where the test article will actually be stored, the department that will conduct the in-life portion, make an assessment of procedures (SOPs) for dosing, identify differences between the sponsor's general practices and the facilities operating procedures (protocol/study-specific procedures (SSPs), and assess the process of interim data exchange and the method of reporting of SOP and protocol deviations to the study director and the sponsor.

Protocol development begins the formal process of establishing study details. Study-specific procedures may be used to append important technical details specified by the sponsor. Most contractors will hold a prestudy initiation meeting, which is an opportunity for the sponsor to meet study staff via teleconference and to provide to critical study staff an overview of the study objectives, last minute details on special handling, and so forth. Study monitoring, particularly of the first day of dosing and at critical necropsy time points, can be very valuable for ensuring appropriate study conduct and in establishing a sponsor's quality and technical expectations. As the data become available, technical and quality review of raw data, including querying any delays in reporting and/or any discrepancies with data in real time, often saves considerable time during the reporting stage.

## 9 CONCLUSIONS

Conduct of preclinical testing in compliance with GLP regulations has become an established part of clinical development. US and global regulatory agencies require that sponsors describe and justify all areas of noncompliance for safety (United States/Europe/Japan) and pharmacological (Europe) supporting studies. For novel biopharmaceuticals and pharmaceuticals it is not always possible to achieve 100% GLP compliance due to the technical requirements and/or unique animal models that may be used to assess safety. Limited resources early in product development particularly impact adherence to GLP compliance. As a product progresses in development, the regulatory expectation is that the GLP knowledge of the sponsor and full compliance of the studies will also evolve. The chapter provides a starting point for US and global GLP compliance and a pragmatic approach to contractor selection for studies intended for full or partial GLP compliance.

In conclusion, the GLP provisions are a regulatory tool to ensure the integrity and quality of the preclinical study data submitted in support of a product's safety for use in humans. GLP regulations pertain to preclinical safety testing within the United States (21 CFR 58) and to both safety and efficacy (pharmacological) testing globally (OECD GLP provisions). Compliance with GLP principles is ultimately the responsibility of the sponsor of the research or marketing application, regardless of the laboratory (in house, contractor, or grantee) that performs the studies for a given product. GLP compliance itself should not be interpreted as inferring scientific competency or a facility's compliance with other regulatory standards. From this perspective the GLP provisions are only one facet of the regulatory expectation for preclinical testing. Other standards include, but are not limited to, 9 CFR test parameters: virus safety testing [45], ICH guidances for analytical methods and safety testing [46], relevant FDA guidance for Industry [47], standards for animal care and use [48], and import/export permits for the test materials from CDC [49] and USDA [50]. Finally, guidance from the World Health Organization on GLP regulations may also be considered, particularly for global development of novel biopharmaceuticals [51].

## REFERENCES

1. United States National Archives and Records Administration1. www.gpoaccess.gov/cfr/index.html
2. US EPA. Toxic Substances Control, Good Laboratory Practice standards, Code of Federal Regulations, Title 40, Part 792, effective December 29, 1983. http://www.epa.gov/compliance/monitoring/programs/fifra/glp.html
3. US EPA. Pesticide Programs, Good Laboratory Practice Standards, Code of Federal Regulations, Title 40, Part 160, effective May 2, 1984. http://www.epa.gov/compliance/monitoring/programs/fifra/glp.html

4. US Department of Health and Human Services, Centers for Disease Control and Prevention, and National Institutes of Health. *Biosafety in Microbiological and Biomedical Laboratories (BMBL)*, 4th ed. Washington: US Government Printing Office, 1999. http://www.cdc.gov/OD/OHS/biosfty/bmbl4/bmbl4toc.htm
5. Marshall E. The murky world of toxicity testing. *Science* 1983;220:1130–2.
6. US FDA. Compliance Program Guidance Manual, Program 7348.808, Chapter 48, Bioresearch Monitoring, Good Laboratory Practice, Completion Date 09/03/93. http://www.fda.gov/ora/compliance_ref/bimo/7348_808/48-808.pdf
7. Anderson MA. *GLP Quality Audit Manual*, 3rd ed, Boca Raton: CRC Press, 2000.
8. RAPS. Chapter 18. Good Laboratory Practices and Good Clinical Practices. In *2005 Fundamentals of US Regulatory Affairs*. Rockville: Regulatory Affairs Professional Society, 2005.
9. US FDA. Good Laboratory Practice (GLP) Regulations, Proposed Rule. Federal Register, November 19, 1976. Vol. 41:51206–51230. http://www.fda.gov/ora/compliance_ref/bimo/GLP/76proposedrule.pdf
10. US FDA. Good Laboratory Practice (GLP) Regulations, Final Rule. Federal Register, December 22, 1978. Vol. 43:59985–60025. http://www.fda.gov/ora/compliance_ref/bimo/GLP/78fr-glpfinalrule.pdf
11. US FDA. Good Laboratory Practice (GLP) Regulations, Proposed Rule. Federal Register. October 29, 1984. Vol. 49:43530–43537. http://www.fda.gov/ora/compliance_ref/bimo/GLP/84proposedrule.pdf
12. US FDA. Good Laboratory Practice (GLP) Regulations (Amendment), Final Rule. Federal Register. September 4, 1987. Vol. 52:33768–33782. http://www.fda.gov/ora/compliance_ref/bimo/GLP/87fr-glpamendment.pdf
13. US FDA. Guidance for Industry. Good Laboratory Practice Regulations Management Briefings. Post Conference Report. August 1979 (Minor editorial and formatting changes made November 1998). http://www.fda.gov/ora/compliance_ref/bimo/q_as.htm
14. US FDA. Good Laboratory Practice Regulations, Questions and Answers. June 1981 (Minor editorial and formatting changes made December 1999, September 2000). http://www.fda.gov/ora/compliance_ref/bimo/GLP/qna.htm
15. US FDA ORA/FDA/ORA Bioresearch Monitoring Information Page (general) http://www.fda.gov/ora/compliance_ref/bimo/default.htm.
16. GLP Web site of the OECD. http://www.oecd.org/department/0,2688,en_2649_34381_1_1_1_1_1,00.html
17. OECD. No. 1: Principles on Good Laboratory Practice, revised 1997. http://www.olis.oecd.org/olis/1998doc.nsf/LinkTo/env-mc-chem(98)17
18. OECD. No. 2: Revised Guides for Compliance Monitoring Procedures for Good Laboratory Practice. http://www.olis.oecd.org/olis/1995doc.nsf/LinkTo/ocde-gd(95)66
19. OECD. No. 3: Revised Guidance for the Conduct of Laboratory Inspections and Study Audit. http://www.olis.oecd.org/olis/1995doc.nsf/LinkTo/ocde-gd(95)67
20. OECD. No. 9: Guidance for the Preparation of GLP Inspection Reports. http://www.olis.oecd.org/olis/1995doc.nsf/LinkTo/ocde-gd(95)114

21. OECD. No. 4: Quality Assurance and GLP, revised 1999. http://www.olis.oecd.org/olis/1999doc.nsf/LinkTo/env-jm-mono(99)20
22. OECD. No. 5: Compliance of Laboratory Suppliers with GLP Principles, revised 1999. http://www.olis.oecd.org/olis/1999doc.nsf/LinkTo/env-jm-mono(99)21
23. OECD. No. 6: The Application of the GLP Principles to Field Studies, revised 1999. http://www.olis.oecd.org/olis/1999doc.nsf/LinkTo/env-jm-mono(99)23
24. OECD. No. 7: The Application of the GLP Principles to Short&Term Studies, revised 1999. http://www.olis.oecd.org/olis/1999doc.nsf/LinkTo/env-jm-mono(99)24
25. OECD. No. 8: The Role and Responsibilties of the Study Director in GLP Studies, revised 1999. http://www.olis.oecd.org/olis/1995doc.nsf/LinkTo/ocde-gd(95)115
26. OECD. No. 10: The Application of the Principles of GLP to Computerised Systems, 1995. http://www.olis.oecd.org/olis/1995doc.nsf/LinkTo/ocde-gd(95)115
27. OECD. No. 11: The Role and Responsibility of the Sponsor in the Application of the Principles of GLP. http://www.olis.oecd.org/olis/1998doc.nsf/LinkTo/env-mc-chem(98)16
28. OECD. No. 12: Requesting and Carrying Out Inspections and Study Audits in Another Country. http://www.olis.oecd.org/olis/2000doc.nsf/LinkTo/env-jm-mono(2000)3
29. OECD. No. 13: The Application of the OECD Principles of GLP to the Organisation and Management of Multi-site Studies. http://www.olis.oecd.org/olis/2002doc.nsf/LinkTo/env-jm-mono(2002)9
30. OECD. No. 14: The Application of the Principles of GLP to In vitro Studies. http://appli1.oecd.org/olis/2004doc.nsf/linkto/env-jm-mono(2004)26
31. Directive 2003/63/EC. Commission Directive 2003/63/EC of 25 June 2003 Amending Directive 2001/83/EC of the European Parliament and of the Council on the Community Code Relating to Medicinal Products for Human Use. http://europa.eu.int/eur-lex/lex/LexUriServ/LexUriServ.do?uri=CELEX:32003L0063:EN:HTML
32. Directive 2004/10/EC of the European Parliament and of the Council of 11 February 2004 on the Harmonization of Laws, Regulations and Administrative Provisions Relating to the Application of the Principles of Good Laboratory Practice and the Verification of Their Applications for Tests on Chemical Substances. http://ec.europa.eu/enterprise/chemicals/legislation/glp/directives_en.htm
33. Directive 2004/9/EC of the European Parliament and of the Council of 11 February 2004 on the Inspection and Verification of Good Laboratory Practice (GLP). http://ec.europa.eu/enterprise/chemicals/legislation/glp/directives_en.htm
34. GLP Web site of the European Commission. http://ec.europa.eu/enterprise/chemicals/legislation/glp/index_en.htm
35. US FDA. International cooperative agreements. http://www.fda.gov/oia/
36. US FDA. OPE Study 42. Results of the Nonclinical Toxicology Laboratory Good Laboratory Practices Pilot Compliance Program. September 1977.
37. Chesney DL. The FDA-483: its history and present use. *BioPharm* 1996;9(4):30–2.
38. Chesney DL, Kelly AE. Responding to FDA-483's and warning letters. *BioPharm* 1998;11(10):26–9. http://www.ispe.org/page.ww?section=Technical+Articles&name=Responding+to+483's+and+Warning+Letters

39. US FDA. ORA/FDA/ORA Bioresearch Monitoring Information Page (GLP specific). http://www.fda.gov/ora/compliance_ref/bimo/glp/default.htm
40. US FDA. Freedom of Information Requests. http://www.fda.gov/opacom/backgrounders/foiahand.html
41. US FDA. Warning Letters. http://www.fda.gov/foi/warning.htm
42. Japan. EU–Japan Mutual Recognition Agreement, 1 January 2002. Sectoral Annex on Good Laboratory Practice (GLP) for Chemicals. http://www.mofa.go.jp/region/europe/eu/agreement-glp.pdf
43. US FDA. Comparison Chart of FDA, Environmental Protection Agency (EPA) and Organization for Economic Co-Operation and Development (OECD) Good Laboratory Practices, June 2004. http://www.fda.gov/ora/compliance_ref/bimo/comparison_chart/
44. Cavagnaro J. Regulatory oversight of gene transfer products: good science and good sense. Session II: Preclinical Development of Gene Therapy Vectors from Petri Dish to Patient. Presentation 5th Annual American Society of Gene Therapy Meeting, Boston, MA.
45. 9 CFR 113 APHIS regulations. http://www.access.gpo.gov/nara/cfr/waisidx_04/9cfr113_04.html
46. International Conference for Harmonization (ICH) Quality and Safety Guidelines. http://www.ich.org/cache/compo/276-254-1.html
47. CBER Guidances/Guidelines/Points to Consider. http://www.fda.gov/cber/guidelines.htm
48. Institute of Laboratory Animal Research, Commission on Life Sciences, National Research Council. Guide for Care and Use of Laboratory Animals, NIH Publication No. 85-23, revised 1985.
49. Centers for Disease Control and Prevention. Etiologic Agent Import Program. http://www.cdc.gov/od/eaipp/
50. USDA Animal and Plant Health Inspection Service. Import and Export regulations and instructions. http://permanent.access.gpo.gov/lps3025/index-33.htm
51. WHO Handbook for Good Laboratory Practice (GLP): Quality Practices for Regulated Non-clinical Research and Development. http://www.who.int/tdr/publications/publications/pdf/glp-handbook.pdf

# Preclinical Safety Study Design Templates and Estimated Costs

GARY W. WOLFE, PhD, DABT

**Contents**

1 Introduction
2 Study Outlines and Cost Estimates
- 2.1 Dose Escalation Study in Rodents
- 2.2 Dose Escalation Study in Nonrodents
- 2.3 Acute Toxicology Study in Rodents
- 2.4 Acute Toxicology Study in Nonrodents
- 2.5 Four-Week Toxicology Study in Rodents
- 2.6 Four-Week Toxicology Study in Nonrodents
- 2.7 Three-Month Toxicology Study in Rodents
- 2.8 Three-Month Toxicology Study in Nonrodents
- 2.9 Six-Month Toxicology Study in Rodents
- 2.10 Nine-Month Toxicology Study in Nonrodents
- 2.11 Carcinogenicity Study in Sprague–Dawley Rats
- 2.12 Carcinogenicity Study in Mice
- 2.13 Carcinogenicity Testing using the Transgenic Mouse (p53+/–, rasH2 or Tg.AC)
- 2.14 Immunotoxicity Study in Sprague–Dawley Rats
- 2.15 Immunotoxicity Study in Monkeys
- 2.16 Vaccine Study in Rabbits
- 2.17 Phototoxicity Study in Rodents
- 2.18 Biodistribution Study in Sprague–Dawley Rats
- 2.19 Study of Fertility and Early Embryonic Development to Implantation in Sprague–Dawley Rats (Segment 1)
- 2.20 Study for Effects on Embryofetal Development and Toxicokinetics in Rats (Segment 2)
- 2.21 Study for Effects on Embryofetal Development and Toxicokinetics in New Zealand White Rabbits (Segment 2)

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

2.22 Study for Effects on Pre- and Postnatal Development, Including Maternal Function in Sprague–Dawley Rats (Segment 3)
2.23 Evaluation of Cardiovascular (Hemodynamic) Function in the Conscious Telemetered Nonrodent
2.24 Neuropharmacology Study in Sprague–Dawley Rats
2.25 Pulmonary Assessment in the Conscious Rat
2.26 Evaluation of Cardiovascular Function Using Cloned hERG Channels Expressed in Human Embryonic Kidney (HEK293) Cells
2.27 Evaluation of Cardiovascular Function Using Action Potential Parameters in Isolated Dog Purkinje Fibers
2.28 *Salmonella typhimurium/Escherichia coli* Plate Incorporation Mutation Assay in the Presence and Absence of Induced Rat Liver S-9
2.29 Chemical Induction of Micronucleated Polychromatic Erythrocytes in Mouse Bone Marrow Cells
2.30 Lymphoma Mutagenesis in the L5178Y TK+/– Mouse with Colony Size Evaluation in the Presence and Absences of Induced Rat Liver S-9 with a Confirmatory Study
2.31 In vitro Chromosomal Aberration
2.32 Tissue Cross-Reactivity
Appendix A Normal Human Tissues Used in Cross-Reactivity Testing
Appendix B Clinical Pathology Parameters
Appendix C Organ Weights
Appendix D Tissue Preservation
Appendix E Draize Scores
Appendix F Analytical / Bioanalytical Costs
Appendix G Special Stains
Appendix H References
Appendix I Estimated Costs for GLP Compliant Preclinical Study Designs to Support Clinical Development

## 1 INTRODUCTION

Preclinical studies are conducted to support the use of the test article in humans in the clinic. The preclinical studies are designed to identify target organs, and biomarkers and to set starting dose levels, recommend dose-escalation schemes and target organs of toxicity. In designing the preclinical study it is important to design the study to answer specific questions concerning the test article while maximizing the use of the animals. An example would be in the dose range studies to design studies using a minimum number of animals to answer the question "What is the toxicity following a single dose in that species and what doses should be used in acute and repeat dose studies in the

rodent?" Studies to support safety should comply with GLPs possibly except for some of the initial single dose studies. It is also important to have the preclinical plan with preclinical study designs reviewed by the regulatory agencies as part of a pre-IND of IND process.

One important aspect in the conduct of the preclinical studies is the identification of a qualified contract research organization (CRO) when studies are not performed internally. These CROs may be identified via the internet websites, literature from national and local toxicology meetings, and regulatory documents such as 483's. In addition to checking references with colleagues, the CRO should be visited and procedures, facilities, staff, and previous experience should be evaluated before entrusting your preclinical program with the CRO (see Chapter Survey of Preclinical Toxicology Programs for Approved Biopharmaceuticals).

Study designs are provided for the following study types:

*Dose escalation studies in rodent and nonrodent* These studies are generally the first time the test article is administered to a living system. They provide an initial indication of toxicity, organs involved, and doses that can be used in longer term studies. Many times these studies are not conducted in full compliance with GLPs.

*Acute toxicology studies in rodent and nonrodent* These studies are used to identify doses causing no adverse effects and doses causing significant life-threatening effects. In addition to toxicity these studies frequently include toxicokinetics to identify the disposition of the test article in the rodent and nonrodent. These studies also provide additional support for the dose levels to be used in longer term studies.

*Four-week toxicology studies in rodent and nonrodent* The four-week studies are designed for subchronic exposure of rodents and nonrodents to the test article. These studies also look at reversibility of any toxicity observed. Many times these are the pivotal studies used to support the first in human dosing. Toxicokinetic assessments are generally included in repeat-dose toxicity studies. When testing biopharmaceuticals, studies also include assessment and characterization of immune response (immunogenicity).

*Three-month toxicology studies in rodent and nonrodent* These studies have a similar design to the four-week studies but support repeated human dosing up to three months.

*Six or nine-month toxicology study in rodent and nonrodent* These studies have a similar design to the 4- and 13-week studies but support chronic repeated human dosing beyond 6 months.

*Carcinogenicity study in mice and rats* These studies are designed to identify a tumorigenic potential and to assess the relevant risk in humans. These studies are conducted for pharmaceuticals that are generally administered over the life of a human and are generally not conducted for biopharmaceuticals (see Chapters Preclinical Evaluation of Cancer

Hazard and Risk of Biopharmaceuticals and Immunomodulatory Biopharmaceuticals and Risk of Neoplasia).

*Immunotoxicity studies in rats and nonhuman primates* Immunotoxicity of pharmaceuticals and biopharmaceuticals includes hematology, organ weights, and pathology of immune-sensitive organs from the repeat-dose studies (4- and 13-week studies). For biopharmaceuticals these tests may need to be extended to T cell dependent antibody responses, imunophenotyping, natural killer cell assays, host resistance studies, and macrophage/neutrophil function. A potential design for a natural killer cell assay is presented in these designs.

*Vaccine study in rabbits* Vaccines are evaluated in rabbits following the dosage regimen that is intended for humans. In many cases these studies include the same dose as is used in humans. In cases where achieving a full human dose cannot be assessed in rodents, nonrodents are needed in particular to determine the local reactogenicity of a full human dose (Draize Score).

*Biodistribution study in rats for plasmids, cellular and gene therapies* Biodistribution and in many cases integration, depending on the type of vaccine, are evaluated in rats following single or multiple doses using qPCR. In many cases these studies can be combined with repeat-dose studies.

*Reproduction studies in rats and rabbits* Reproductive effects are evaluated on the gametes, in the developing organism (teratology), and in the progeny. These studies are generally conducted for all pharmaceuticals, but on a case-by-case basis for biopharmaceuticals.

*Safety pharmacology studies in rodent and nonrodent* These studies are designed to identify undesirable pharmacological effects in several systems primarily the cardiovascular, respiratory, and CNS systems. Other studies may be included based on the pharmacological activity of the test article. In many cases these studies are not required for biopharmaceuticals and/or may be included in the 4- and 13-week studies and 6- and 9-month studies.

*Genetic toxicology studies* These studies designed to identify potential point mutations, and/or clastogenic damage/chromosomal alterations are required for pharmaceuticals but generally not required for biopharmaceuticals, since they are not applicable for proteins.

*Tissue cross-reactivity (TCR) studies* Tissue cross-reactivity studies with human tissues (or cells if applicable) are conducted prior to Phase 1 to search for cross-reactions with the intended target and/or nontarget tissue. In special cases of bispecific antibodies, each parent antibody is evaluated individually in addition to testing the bispecific product. Human cells or tissues are surveyed immunocytochemically or immunohistochemically with appropriate controls. Animal species are also surveyed to determine relevant species for toxicology studies.

*Cost estimates* Cost estimates are provided for planning and budgetary purposes, and they vary widely based on a number of factors with each individual CRO. The estimates are based generally on average prices for the study designs presented; but the final price may vary for specific circumstances:

- Final design. Numbers of animals, blood collection intervals and time points, clinical pathology intervals, and tissues examined.
- Special study requests. CTD tables, additional QA inspections, peer review, or accelerated report submission.
- Percentage of each study conducted in house. Cost increases due to specialty sections (e.g., pathology, clinical pathology) being subcontracted by the CRO.
- Discounted individual studies. A program having a large number of studies.
- Unutilized capacity of CRO. Price decreases to fill CRO's available space.
- Flexibility in initiation and reporting. Proper planning tends to decrease the CRO's costs.

In summary, a well-designed preclinical program that addresses questions of toxicity conducted internally or at a well-qualified CRO is necessary for a satisfactory preclinical safety evaluation program to support successful clinical development.

## 2 STUDY OUTLINES AND COST ESTIMATES

### 2.1 Dose Escalation Study in Rodents

| | |
|---|---|
| Purpose | What is the toxicity following a single dose to the nonrodent?<br>What doses should be used in acute and repeat dose studies in the rodent? |
| Timeline initiation report | 4 to 6 weeks following authorization<br>4 to 8 weeks following termination |
| Estimated cost | $15,000 to $35,000 |
| Regulatory guidelines and compliance | No specific guidelines and FDA GLP optional |
| Total animals | 36 (18/sex) rats (or mice) |
| Number/sex/group | 2 rats/sex/group with 4 groups in dose escalation phase<br>5 rats/sex/group with 2 groups in MTD phase |

| | |
|---|---|
| Dosing route | Generally a parenteral route (IV, IM, or sc) for biopharmaceuticals<br>Oral route usually for pharmaceuticals |
| Dosing frequency | Day 1 and every 72 hours thereafter 2 animals/sex dosed for a total of 4 escalating doses to determine MTD<br>5 rats/sex/group dosed with the vehicle and at the MTD either as a single dose or multiple doses up to 14 days |
| Formulations | Usually biopharmaceuticals dosed as received if clinical formulation is available<br>Pharmaceuticals formulated by the CRO and require formulation analysis for concentration, homogeneity, and stability |
| Clinical signs | Twice daily |
| Clinical observations | 0.5, 1, and 2 hours after each dose and daily in the MTD phase |
| Body weights | Before and after each dose in the dose escalation phase and prior to initiation, weekly, and at termination in the MTD phase |
| Food consumption | Weekly in the MTD phase |
| Clinical pathology (optional) | Not required |
| Terminal necropsy | 2–3 days after dosing in the dose escalation phase<br>14 days after dosing in the MTD phase |
| Organ weights (optional) | See Appendix C |
| Tissue preservation (optional) | See Appendix D |
| Histopathology (optional) | MTD groups gross lesions only or all tissues |
| Report | One copy of the draft final report<br>One bound and unbound copy of the final and an electronic copy in Word and PDF |
| Archive | Yes for at least 1 year at CRO |

## 2.2 Dose Escalation Study in Nonrodents

| | |
|---|---|
| Purpose | What is the toxicity following a single dose to the nonrodent?<br>What doses should be used in acute and repeat-dose studies in the nonrodent? |

| | |
|---|---|
| Timeline | |
| initiation | 4 to 8 weeks following authorization |
| report | 4 to 8 weeks following termination |
| Estimated cost | $50,000 to $75,000 |
| Regulatory guidelines and compliance: | No specific guidelines and FDA GLP optional |
| Total Animals | 12 monkeys or dogs (6/sex) |
| Number/sex/group | 2 monkeys or dogs/dose with up to 4 doses/animal |
| | 2 monkeys or dogs/group with 2 groups in MTD phase |
| Dosing route | Generally a parenteral route (IV, IM, or sc) for biopharmaceuticals |
| | Oral route usually for pharmaceuticals |
| Dosing frequency | Day 1 and every 72 hours thereafter the same 2 animals/sex dosed for up to 3 escalating doses to determine MTD |
| | 2 animals/sex/group dosed with vehicle and at the MTD either as a single dose or multiple doses up to 14 days |
| Formulations | Usually biopharmaceuticals dosed as received if clinical formulation is available |
| | Pharmaceuticals are formulated by the CRO and require formulation analysis for concentration, homogeneity, and stability |
| Clinical signs | Twice daily |
| Clinical observations | 0.5, 1, and 2 hours after each dose each dose and daily in the MTD phase |
| Body weights | Before and after each dose in the dose escalation phase and prior to initiation, weekly, and at termination in the MTD phase |
| Food consumption | Weekly in the MTD phase |
| Clinical pathology (optional) | Not required |
| Terminal necropsy | 2 to 3 days after dosing in the dose escalation phase |
| | 14 days after dosing in the MTD phase |
| Organ weights (optional) | See Appendix C |
| Tissue preservation (optional) | See Appendix D |
| Histopathology (optional) | MTD groups gross lesions only or all tissues |
| Report | One copy of the draft final report |
| | One bound and unbound copy of the final and an electronic copy in Word and PDF |
| Archive | Yes for at least 1 year at CRO |

### 2.3 Acute Toxicology Study in Rodents

| | |
|---|---|
| Purpose | What are the toxicities and toxicokinetics of a test article after a single dose?<br>What doses should be used in acute and repeat dose studies in the rodent? |
| Timeline | |
| initiation | 4 to 6 weeks following authorization |
| report | 6 to 10 weeks following termination |
| Estimated cost | $20,000 to $40,000 |
| Regulatory guidance and compliance | Single–dose acute toxicity testing for pharmaceuticals (August 1996) and FDA GLP |
| Total Animals | 40 without toxicokinetics<br>76 with toxicokinetics |
| Number/sex/group—Main | 5/sex/group in groups 1–4 rats (or mice) |
| Number/sex/group—TK (optional) | 6/sex/group in groups 2–4 (assumes <1 ml of blood/rat/time point and <0.1 ml of blood/mouse/time pont)<br>Additional volumes will require additional animals |
| Dosing route frequency | Generally a parenteral route (IV, IM, or sc) for biopharmaceuticals<br>Oral route for pharmaceuticals |
| Dosing frequency | Single dose at 4 dose levels (control, low, mid, and high dose) |
| Formulations | Usually biopharmaceuticals dosed as received if clinical formulation is available<br>Pharmaceuticals are formulated by the CRO and require formulation analysis for concentration, homogeneity, and stability |
| Clinical signs | Twice daily |
| Clinical observations | Weekly |
| Body weights | Weekly and prior to necropsy |
| Food consumption | Weekly |
| Toxicokinetics | Day 1 TK animals are bled at 6 time points (3/sex/group) after dosing |
| Antibodies (for biopharmaceuticals only) | Prior to study initiation and at termination |
| Clinical pathology (optional) | Day 14 |
| Hematology | See Appendix B |
| Clinical chemistry | See Appendix B |
| Coagulation | See Appendix B |
| Urinalysis | See Appendix B |
| Terminal necropsy | Day 14—full necropsy |

| | |
|---|---|
| Organ weights (optional) | See Appendix C |
| Bone marrow smears | Yes from sternum |
| Tissue preservation | See Appendix D |
| Histopathology (optional) | Groups 1 and 4 all tissues<br>Groups 2 and 3—target tissues and gross lesions |
| Report | One copy of the draft final report<br>One bound and unbound copy of the final and an electronic copy in Word and PDF |
| Archive | Yes for at least 1 year at CRO |

### 2.4 Acute Toxicology Study in Nonrodents

| | |
|---|---|
| Purpose | What are the toxicities and toxicokinetics of a test article after a single dose when administered to monkeys or dogs?<br>What doses should be used in acute and repeat-dose studies in the nonrodent? |
| Timeline | |
| initiation | 4 to 8 weeks following authorization |
| report | 6 to 10 weeks following termination |
| Estimated cost | $100,000 to $120,000 (dogs)<br>$150,000 to $200,000 (monkeys, IV bolus)<br>$175,000 to $225,000 (monkey 4-hour infusion) |
| Regulatory guidelines and compliance | Single–dose acute toxicity testing for pharmaceuticals (August 1996) FDA GLP |
| Total animals | 16 monkeys or dogs |
| Number/sex/group | 2 monkeys or dogs/sex/group |
| Dosing route | Generally a parenteral route (IV, IM, or sc) for biopharmaceuticals<br>Oral route for pharmaceuticals |
| Dosing frequency | Single dose at 4 dose levels (control, low, mid, and highdose) |
| Formulations | Usually biopharmaceuticals dosed as received if clinical formulation is available<br>Pharmaceuticals are formulated by the CRO and require formulation analysis for concentration, homogeneity, and stability |
| Clinical signs | Twice daily |
| Clinical observations | Twice daily |
| Body weights | Weekly and prior to necropsy |
| Food consumption | Daily |
| Toxicokinetics | Day 1 TK animals are bled at 6 time points after dosing |

| | |
|---|---|
| Antibodies (for biopharmaceuticals only) | Prior to study initiation, and on day 14 |
| Clinical pathology (optional) | Day 14 |
| Hematology | See Appendix B |
| Clinical chemistry | See Appendix B |
| Coagulation | See Appendix B |
| Urinalysis | See Appendix B |
| Terminal necropsy | Day 14—full necropsy |
| Organ weights (optional) | See Appendix C |
| Tissue preservation (optional) | See Appendix D |
| Histopathology (optional) | Groups 1 and 4 all tissues<br>Groups 2 and 3—target tissues and gross lesions<br>Recommended—All tissues in all groups |
| Report | One copy of the draft final report<br>One bound and unbound copy of the final and an electronic copy in Word and PDF |
| Archive | Yes for at least 1 year at CRO |

## 2.5 Four-Week Toxicology Study in Rodents

| | |
|---|---|
| Purpose | What are the subchronic toxicities and toxicokinetics of a test article when administered for 4 weeks?<br>What is the reversibility, persistence, or delayed occurrence of any effect after a 4-week treatment-free period? |
| Timeline | |
| initiation | 4 to 8 weeks following authorization |
| report | 10 to 12 weeks following termination |
| Estimated price | $200,000 to $250,000 |
| Regulatory compliance | FDA GLP |
| Total animals | 88 male and 88 female rats (or mice) |
| Number/sex/group—Main | 15/sex/group in groups 1–4 |
| Number/sex/group—Recovery | 5/sex/group in groups 1 and 4 |
| Number/sex/group—TK | 6/sex/group in groups 2–4 (assumes <1 ml of blood/rat/timepoint and <0.1 ml of blood/mouse/time pont)<br>Additional volumes will require additional animals |
| Dosing route | Generally a parenteral route (IV, IM, or sc) for biopharmaceuticals<br>Oral route for pharmaceuticals |

| | |
|---|---|
| Dosing frequency | Dosed daily for 4 weeks for pharmaceuticals, but for biopharmaceuticals frequency based on TK data or proposed clinical regimen |
| Formulations | Usually biopharmaceuticals dosed as received if clinical formulation is available<br>Pharmaceuticals are formulated by the CRO and require formulation analysis for concentration, homogeneity, and stability |
| Clinical signs | Twice daily |
| Clinical observations | Weekly |
| Body weights | Weekly and prior to necropsy |
| Food consumption | Weekly |
| Ophthalmology | Prior to initiation and during weeks 4, and 8 |
| Toxicokinetics—optional if conducted in an earlier study | Days 1 and 28 TK animals (3/sex/group) are bled at 6 time points after dosing |
| Antibodies (for biopharmaceuticals only) | Prior to study initiation, and weeks 4 and 8 |
| Neutralizing antibodies (for biopharmaceuticals only) | Prior to study initiation, and weeks 4 and 8 |
| Clinical pathology | During weeks 4 ad 8 |
| Hematology | See Appendix B |
| Clinical chemistry | See Appendix B |
| Coagulation | See Appendix B |
| Urinalysis | See Appendix B |
| Terminal necropsy | Week 4—full necropsy in 15/sex/group in groups 1–4 |
| Recovery necropsy | Week 8—full necropsy in 5/sex/group in groups 1 and 4 |
| Organ weights | See Appendix C |
| Bone marrow smears | Yes from sternum |
| Tissue preservation | See Appendix D |
| Histopathology | Groups 1 and 4 all tissues (terminal and recovery)<br>Groups 2 and 3, and recovery—target tissues and gross lesions |
| Report | One copy of the draft final report<br>One bound and unbound copy of the final and an electronic copy in Word and PDF |
| Archive | Yes for at least 1 year at CRO |

## 2.6 Four-Week Toxicology Study in Nonrodents

| | |
|---|---|
| Purpose | What are the subchronic toxicities and toxicokinetics of a test article when administered to monkeys/dogs for 4 weeks?<br>What is the reversibility, persistence, or delayed occurrence of any effect after a 4-week treatment-free period? |
| Timeline | |
| initiation | 6 to 8 weeks following authorization |
| report | 10 to 12 weeks following termination |
| Estimated cost | $400,000 to $450,000 (dogs)<br>$500,00 to $550,000 (monkey IV bolus)<br>$550,00 to $650,000 (monkey 4-hour infusion) |
| Regulatory guidelines | FDA GLP |
| Total animals | 32 monkeys or 40 dogs |
| Number/sex/group—main | 3 monkeys or 4 dogs/sex/group in groups 1–4 |
| Number/sex/group—recovery | 2/sex/group in groups 1–4 |
| Dosing route | Generally a parenteral route (IV, IM, or sc) for biopharmaceuticals<br>Oral route for pharmaceuticals |
| Dosing frequency | Dosed daily for pharmaceuticals but for biopharmaceuticals frequency based on TK data or proposed clinical regimen |
| Formulations | Usually biopharmaceuticals dosed as received if clinical formulation is available.<br>Pharmaceuticals are formulated by the CRO and require formulation analysis for concentration, homogeneity, and stability |
| Clinical signs | Twice daily |
| Clinical observations | Weekly |
| Body weights | Weekly and prior to necropsy |
| Food consumption | Weekly |
| Ophthalmology | Prior to initiation and during weeks 4 and 8 |
| Toxicokinetics—optional if conducted in earlier studies | Day 1, and at termination, animals are bled at least at 6 time points after dosing |
| Antibodies (for biopharmaceuticals only) | Prior to study initiation, and weeks 4 and 8 |
| Neutralizing antibodies (for biopharmaceuticals only) | Prior to study initiation, and weeks 4, and 8 |

| | |
|---|---|
| Clinical pathology | Prior to study initiation<br>During weeks 4 and 8 in all surviving animals |
| Hematology | See Appendix B |
| Clinical chemistry | See Appendix B |
| Coagulation | See Appendix B |
| Urinalysis | See Appendix B |
| Terminal necropsy | Week 4—full necropsy in all main study animals |
| Recovery necropsy | Week 8—full necropsy in recovery animals |
| Organ weights | See Appendix C |
| Bone marrow smears | Yes from sternum |
| Tissue preservation | See Appendix D |
| Histopathology | Groups 1 and 4 all tissues Groups 2 and 3—target tissues and gross lesions<br>Recommended—all tissues in all groups |
| Report | One copy of the draft final report<br>One bound and unbound copy of the final and an electronic copy in Word and PDF |
| Archive | Yes for at least 1 year at CRO |

### 2.7 Three-Month Toxicology Study in Rodents

| | |
|---|---|
| Purpose | What are the subchronic toxicities and toxicokinetics of a test article when administered to rats for 3 months?<br>What is the reversibility, persistence, or delayed occurrence of any effect after the 3-month period? |
| Timeline | |
| initiation | 6 to 8 weeks following authorization |
| report | 10 to 12 weeks following termination |
| Estimated cost | $250,000 to $300,000 |
| Regulatory compliance | FDA GLP |
| Total animals | 176 |
| Number/sex/group—main | 15/sex/group in groups 1–4 rats (or mice) |
| Number/sex/group—recovery | 5/sex/group in groups 1 and 4 |
| Number/sex/group—TK | 6/sex/group in groups 2–4 (assumes <1 ml of blood/rat/timepoint and <0.1 ml of blood/mouse/time point)<br>Additional volumes will require additional animals |
| Dosing route | Generally a parenteral route (IV, IM, or sc) for biopharmaceuticals<br>Oral route for pharmaceuticals |

| | |
|---|---|
| Dosing frequency | Dosed daily for 13 weeks for pharmaceuticals but for biopharmaceuticals frequency based on TK data or proposed clinical regimen |
| Formulations | Usually biopharmaceuticals dosed as received if clinical formulation is available |
| | Pharmaceuticals are formulated by the CRO and require formulation analysis for concentration, homogeneity, and stability. |
| Clinical signs | Twice daily |
| Clinical observations | Weekly |
| Body weights | Weekly and prior to necropsy |
| Food consumption | Weekly |
| Ophthalmology | Prior to initiation, and during weeks 13 and 17 |
| Toxicokinetics—optional if conducted in a previous study | Days 1 and 90 TK animals (3/sex/group) bled at 6 time points after dosing |
| Antibodies (for biopharmaceuticals only) | Prior to study initiation, and weeks 4, 13, and 17 |
| Neutralizing antibodies (for biopharmaceuticals only) | Prior to study initiation, and weeks 4, 13, and 17 |
| Clinical pathology | During weeks 13 and 17 in all main study animals |
| Hematology | See Appendix B |
| Clinical chemistry | See Appendix B |
| Coagulation | See Appendix B |
| Urinalysis | See Appendix B |
| Terminal necropsy | Week 13—full necropsy in 15/sex/group in groups 1–4 |
| Recovery necropsy | Week 17—full necropsy in 5/sex/group in groups 1 and 4 |
| Organ weights | See Appendix C |
| Bone marrow smears | Yes from sternum |
| Tissue preservation | See Appendix D |
| Histopathology | Groups 1 and 4 all tissues (main and recovery) |
| | Groups 2 and 3, and recovery—target tissues and gross lesions |
| Report | One copy of the draft final report |
| | One bound and unbound copy of the final and an electronic copy in Word and PDF |
| Archive | Yes for at least 1 year at CRO |

## 2.8 Three-Month Toxicology Study in Nonrodents

| | |
|---|---|
| Purpose | What are the subchronic toxicities and toxicokinetics of a test article when administered to monkeys or dogs for 3 months?<br>What is the reversibility, persistence, or delayed occurrence of any effect after the 3-month period? |
| Timeline | |
| initiation | 4 to 8 weeks following authorization |
| report | 10 to 12 weeks following termination |
| Estimated cost | $450,000 to $500,000 |
| Regulatory compliance | FDA GLP |
| Total animals | 32 monkeys or 40 dogs |
| Number/sex/group—main | 3 monkeys or 4 dogs/sex/group in groups 1–4 |
| Number/sex/group—recovery | 2/sex/group in groups 1–4 |
| Dosing route | Generally a parenteral route (IV, IM, or sc) for biopharmaceuticals<br>Oral route for pharmaceuticals |
| Dosing frequency | Dosed daily for 13 weeks for pharmaceuticals but for biopharmaceuticals frequency based on TK data or proposed clinical regimen |
| Formulations | Usually biopharmaceuticals dosed as received if clinical formulation is available<br>Pharmaceuticals are formulated by the CRO and require formulation analysis for concentration, homogeneity, and stability |
| Clinical signs | Twice daily |
| Clinical observations | Weekly |
| Body weights | Weekly and prior to necropsy |
| Food consumption | Weekly |
| Ophthalmology | Prior to initiation, and during weeks 13 and 17 |
| Toxicokinetics—optional if conducted in earlier studies | Days 1 and 90 all animals bled at least at 6 time points after dosing |
| Antibodies (for biopharmaceuticals only) | Prior to study initiation, and weeks 4, 13, and 20 |
| Neutralizing antibodies (for biopharmaceuticals only) | Prior to study initiation, and weeks 4, 13, and 20 (for biopharmaceuticals only) |

| | |
|---|---|
| Clinical pathology | Prior to study initiation<br>During weeks 13 and 17 in all surviving animals |
| Hematology | See Appendix B |
| Clinical chemistry | See Appendix B |
| Coagulation | See Appendix B |
| Urinalysis | See Appendix B |
| Terminal necropsy | Week 13—full necropsy in all main study animals |
| Recovery necropsy | Week 17—full necropsy in recovery animals |
| Organ weights | See Appendix C |
| Bone marrow smears | Yes from sternum |
| Tissue preservation | See Appendix D |
| Histopathology | Groups 1 and 4 all tissues (main and recovery)<br>Groups 2 and 3—target tissues and gross lesions<br>Recommended—all tissues in all groups |
| Report | One copy of the draft final report<br>One bound and unbound copy of the final and an electronic copy in Word and PDF |
| Archive | Yes for at least 1 year at CRO |

### 2.9 Six-Month Toxicology Study in Rodents

| | |
|---|---|
| Purpose | What are the toxicities and toxicokinetics of a test article when administered for 6 months? |
| Timeline | |
| initiation | 6 to 10 weeks following authorization |
| report | 10 to 12 weeks following termination |
| Estimated cost | $350,000 to $550,000 |
| Regulatory guidelines and compliance | S4A duration of chronic toxicity testing in animals (rodent and nonrodent testing) and FDA GLP |
| Total animals | 156 Sprague–Dawley rats (or mice) |
| Number/sex/group—main | 15/sex/group in groups 1–4 |
| Number/sex/group—TK | 6/sex/group in groups 2–4 (assumes <1 ml of blood/rat/time point and <0.1 ml of blood/mouse/time point)<br>Additional volumes will require additional animals |
| Dosing route | Generally a parenteral route (IV, IM, or sc) for biopharmaceuticals<br>Oral route for pharmaceuticals |

| | |
|---|---|
| Dosing frequency | Dosed daily for 26 weeks for pharmaceuticals but for biopharmaceuticals frequency based on TK data or proposed clinical regimen |
| Formulations | Usually biopharmaceuticals dosed as received if clinical formulation is available<br>Pharmaceuticals are formulated by the CRO and require formulation analysis for concentration, homogeneity, and stability |
| Clinical signs | Twice daily |
| Clinical observations | Weekly |
| Body weights | Weekly and prior to necropsy |
| Food consumption | Weekly |
| Ophthalmology | Prior to initiation, and during week 26 |
| Toxicokinetics—optional if conducted in earlier studies | Day 1, and at termination, animals (3/sex/time point) bled at least at 6 time points after dosing |
| Antibodies (for biopharmaceuticals only) | Prior to study initiation, and weeks 4, 13, and 26 |
| Neutralizing antibodies (for biopharmaceuticals only) | Prior to study initiation, and weeks 4, 13, 26, 39, and 43 |
| Clinical pathology | At termination |
| Hematology | See Appendix B |
| Clinical chemistry | See Appendix B |
| Coagulation | See Appendix B |
| Urinalysis | See Appendix B |
| Terminal necropsy | Week 26—full necropsy |
| Organ weights | See Appendix C |
| Bone marrow smears | Yes from sternum |
| Tissue preservation | See Appendix D |
| Histopathology | Groups 1 and 4 all tissues<br>Groups 2 and 3—target tissues and gross lesions |
| Report | One copy of the draft final report<br>One bound and unbound copy of the final and an electronic copy in Word and PDF |
| Archive | Yes for at least 1 year at CRO |

### 2.10 Nine-Month Toxicology Study in Nonrodents

| | |
|---|---|
| Purpose | What are the chronic toxicities and toxicokinetics of a test article when administered to monkeys or dogs for 9 months? |

| | |
|---|---|
| Timeline | |
| initiation | 8 to 12 weeks following authorization |
| report | 10 to 12 weeks following termination |
| Estimated cost | $450,000 to $650,000 |
| Regulatory guidelines and compliance | S4A duration of chronic toxicity testing in animals (rodent and nonrodent testing) and FDA GLP |
| Total animals | 32 beagle dogs<br>24 cynomolgus monkeys |
| Number/sex/group—main | 4 dogs/sex/group in groups 1–4<br>3 monkeys/sex/group in groups 1–4 |
| Dosing route | Generally a parenteral route (IV, IM, or sc) for biopharmaceuticals<br>Oral route for pharmaceuticals |
| Dosing frequency | Dosed daily for 39 weeks for pharmaceuticals but for biopharmaceuticals frequency based on TK data or proposed clinical regimen<br>For biopharmaceuticals 6-month duration generally appropriate |
| Formulations | Usually biopharmaceuticals dosed as received if clinical formulation is available<br>Pharmaceuticals are formulated by the CRO and require formulation analysis for concentration, homogeneity, and stability |
| Clinical signs | Twice daily |
| Clinical observations | Weekly |
| Body weights | Weekly and prior to necropsy |
| Food consumption | Weekly |
| Ophthalmology | Prior to initiation, and during Weeks 39 |
| Electrocardiograms | Prior to initiation, and at termination |
| Toxicokinetics—optional if conducted in earlier studies | Day 1, and at termination, animals are bled at least at 6 time points after dosing |
| Antibodies (for biopharmaceuticals only) | Prior to study initiation, and weeks 4, 13, 26, and 39 |
| Neutralizing antibodies (for biopharmaceuticals only) | Prior to study initiation, and weeks 4, 13, 26, and 39 (for biopharmaceuticals only) |
| Clinical pathology | Prior to study initiation<br>During weeks 4, 13, 26, and 39 in all surviving animals |
| Hematology | See Appendix B |
| Clinical chemistry | See Appendix B |
| Coagulation | See Appendix B |
| Urinalysis | See Appendix B |

| | |
|---|---|
| Terminal necropsy | Week 39—full necropsy in all surviving animals |
| Organ weights | See appendix C |
| Bone marrow smears | Yes from sternum |
| Tissue preservation | See Appendix D |
| Histopathology | Groups 1 and 4 all tissues<br>Groups 2 and 3—target tissues and gross lesions<br>Recommended—all tissues in all groups |
| Report | One copy of the draft final report<br>One bound and unbound copy of the final and an electronic copy in Word and PDF |
| Archive | Yes for at least 1 year at CRO |

### 2.11 Carcinogenicity Study in Sprague–Dawley Rats

| | |
|---|---|
| Purpose | What is the carcinogenic potential of a test article when administered to Sprague–Dawley Rats for at least 104 weeks? |
| Timeline | |
| initiation | 8 to 10 weeks following authorization |
| report | 26 to 30 weeks following termination |
| Estimated cost | $1.0 to 1.5 M |
| Regulatory guidance and compliance | S1A need for long-term rodent carcinogenicity studies of pharmaceuticals and S1B testing for carcinogenicity of pharmaceuticals and FDA GLP |
| Total animals | 480 Sprague–Dawley rats |
| Number/sex/group—Main | 60/sex/group in groups 1–4 |
| Dosing route | Generally a parenteral route (IV, IM, or sc) for biopharmaceuticals<br>Oral route for pharmaceuticals |
| Dosing frequency | Dosed daily for 104 weeks for pharmaceuticals but for biopharmaceuticals frequency based on TK data or proposed clinical regimen |
| Formulations | Usually biopharmaceuticals dosed as received if clinical formulation is available<br>Pharmaceuticals are formulated by the CRO and require formulation analysis for concentration, homogeneity, and stability |
| Clinical signs | Twice daily |
| Clinical observations | Weekly with special attention to tumor development |

| | |
|---|---|
| Body weights | Weekly for the first 13 weeks, and every 4 weeks thereafter |
| Food consumption | Weekly for the first 13 weeks, and every 4 weeks thereafter |
| Blood smears | At 12 and 18 months and prior to necropsy a blood smears obtained from all animals<br>Smears are read in the high-dose and control only unless indicated by toxic effects |
| Terminal necropsy | Week 104—full necropsy in all surviving animals |
| Organ weights | See Appendix C |
| Bone marrow smears | Sternum at termination |
| Tissue preservation | See Appendix D |
| Histopathology | Minimum—all tissues in groups 1 and 4 and target tissues, tumors, and gross lesions from groups 2 and 3<br>Recommended—all tissues in all groups |
| Report | One copy of the draft final report<br>One bound and unbound copy of the final and an electronic copy in Word and PDF |
| Archive | Yes at least 1 year at the CRO |

### 2.12 Carcinogenicity Study in Mice

| | |
|---|---|
| Purpose | What is the carcinogenic potential of a test article when administered to mice for at least 104 weeks? |
| Timeline | |
| initiation | 8 to 10 weeks following authorization |
| report | 26 to 30 weeks following termination |
| Estimated cost | $1.0 to 1.5 M |
| Regulatory guidelines and compliance | S1A need for long-term rodent carcinogenicity studies of pharmaceuticals and S1B testing for carcinogenicity of pharmaceuticals and FDA GLP |
| Total animals | 480 mice |
| Number/sex/group | 60/sex/group in groups 1–4 |
| Dosing route | Generally a parenteral route (IV, IM, or sc) for biopharmaceuticals<br>Oral route for pharmaceuticals |
| Dosing frequency | Dosed daily for 104 weeks for pharmaceuticals but for biopharmaceuticals frequency based on TK data or proposed clinical regimen |

| | |
|---|---|
| Formulations | Usually biopharmaceuticals dosed as received if clinical formulation is available<br>Pharmaceuticals are formulated by the CRO and require formulation analysis for concentration, homogeneity, and stability |
| Clinical signs | Twice daily |
| Clinical observations | Weekly |
| Body weights | Weekly for the first 13 weeks, and every 4 weeks thereafter |
| Food consumption | Weekly for the first 13 weeks, and every 4 weeks thereafter |
| Blood smears | At 12 and 18 months and prior to necropsy a blood smear obtained from all animals<br>Smears are read in the high-dose and control only unless indicated by toxic effects |
| Terminal necropsy | Week 104—full necropsy in all surviving animals |
| Organ weights | See Appendix B |
| Bone marrow smears | Sternum at termination |
| Tissue preservation | See Appendix D |
| Histopathology | Minimum—all tissues in groups 1 and 4 and target tissues, tumors, and gross lesions from groups 2 and 3<br>Recommended—all tissues in all groups |
| Report | One copy of the draft final report<br>One bound and unbound copy of the final and an electronic copy in Word and PDF |
| Archive | Yes at least 1 year at CRO |

### 2.13 Carcinogenicity Testing using the Transgenic Mouse (p53+/–, rasH2, or Tg.AC)

| | |
|---|---|
| Purpose | What is the carcinogenic potential of a test article when administered to a transgenic mouse for 26 weeks? |
| Timeline | |
| initiation | 12 to 16 weeks following authorization |
| report | 13 to 17 weeks following termination |
| Estimated cost | |
| Pilot | $250,000 to $350,000 |
| Main | $650,000 to $750,000 |

| | |
|---|---|
| Regulatory guidelines and compliance | S1A need for long-term rodent carcinogenicity studies of pharmaceuticals and S1B testing for carcinogenicity of pharmaceuticals and FDA GLP |
| Total animals | |
| Pilot | 80 mice |
| Main | 200 mice |
| Animal model strain | p53+/– mice for genotoxic drugs, rasH2 for nongenotoxic drugs, or Tg.AC mice for dermal administration |
| Number/sex/group | Pilot 10/sex/group in groups 1–4<br>Main 25/sex/group in groups 1–4 |
| Dosing route | Generally a parenteral route (IV, IM, or sc) for biopharmaceuticals<br>Parenteral or oral route for pharmaceuticals<br>Dermal route for Tg.AC mice |
| Dosing frequency | Generally dosed daily for 4 (pilot) or 26 weeks (main) for pharmaceuticals and biopharmaceuticals |
| Formulations | Usually biopharmaceuticals dosed as received if clinical formulation is available<br>Pharmaceuticals are formulated by the CRO and require formulation analysis for concentration, homogeneity, and stability |
| Clinical signs | Twice daily |
| Clinical observations | Weekly |
| Body weights | Weekly |
| Food consumption | Weekly |
| Clinical pathology (optional) | Hematology, coagulation, and clinical chemistry at termination |
| Terminal necropsy | Pilot study (week 4) or main study (week 26)—full necropsy in all surviving animals |
| Organ weights | See Appendix B |
| Tissue preservation | See Appendix D |
| Histopathology | Minimum—all tissues in groups 1 and 4 and target tissues, tumors, and gross lesions from groups 2 and 3<br>Recommended—all tissues in all groups |
| Report | One copy of the draft final report<br>One bound and unbound copy of the final and an electronic copy in Word and PDF |
| Archive | Yes at least 1 year at CRO |

## 2.14 Immunogenicity Study in Sprague–Dawley Rats

| | |
|---|---|
| Purpose | What is the T cell dependent antibody response to KLH (keyhole limpet hemocyanin) in the rat?<br>(Note: this design may be incorporated with other short-term rat studies) |
| Timeline | |
| initiation | 4 to 8 weeks following authorization |
| report | 10 to 12 weeks following termination |
| Estimated price | $200,000 to $250,000 |
| Regulatory compliance | FDA GLP |
| Total animals | 24 male and 24 female Sprague–Dawley rats |
| Number/sex/group—main | 6/sex/group in groups 1–4 |
| Dosing route | Generally a parenteral route (IV, IM, or sc) for biopharmaceuticals<br>Oral route for pharmaceuticals |
| Dosing frequency | Dosed daily for 2 to 4 weeks for pharmaceuticals but for biopharmaceuticals frequency based on TK data or proposed clinical regimen |
| Formulations | Usually biopharmaceuticals dosed as received if clinical formulation is available<br>Pharmaceuticals are formulated by the CRO and require formulation analysis for concentration, homogeneity, and stability |
| Clinical signs | Twice daily |
| Clinical observations | Weekly |
| Body weights | Weekly and prior to necropsy |
| Food consumption | Weekly |
| KLH administration | 1 and 2 weeks prior to termination, KLH is administered |
| Blood collection | Serial collections at one or more time points |
| Terminal necropsy | Spleen collected on ice |
| Organ weights | Spleen |
| Assay | Titers for anti-KLH antibodies |
| Report | One copy of the draft final report<br>One bound and unbound copy of the final and an electronic copy in Word and PDF |
| Archive | Yes for at least 1 year at CRO |

## 2.15 Immunotoxicity Study in Monkeys

| | |
|---|---|
| Purpose | What is the T cell dependent antibody response to KLH (keyhole limpet hemocyanin) in the monkey?<br>(Note: this design may be incorporated with other short-term monkey studies) |
| Timeline | |
| initiation | 6 to 8 weeks following authorization |
| report | 10 to 12 weeks following termination |
| Estimated cost | $200,000 to $300,000 |
| Regulatory guidelines | FDA GLP |
| Total animals | 16 monkeys |
| Number/sex/group | 2 monkeys /sex/group in groups 1–4 |
| Dosing route | Generally a parenteral route (IV, IM, or sc) for biopharmaceuticals<br>Oral route for pharmaceuticals |
| Dosing frequency | Dosed daily for pharmaceuticals but for biopharmaceuticals frequency based on TK data or proposed clinical regimen |
| Formulations | Usually biopharmaceuticals dosed as received if clinical formulation is available<br>Pharmaceuticals are formulated by the CRO and require formulation analysis for concentration, homogeneity, and stability |
| Clinical signs | Twice daily |
| Clinical observations | Weekly |
| Body weights | Weekly and prior to necropsy |
| Food consumption | Weekly |
| KLH administration | 1 and 2 weeks prior to termination, KLH is administered |
| Blood collection | Serial collections at one or more time points |
| Terminal necropsy | Spleen collected on ice |
| Organ weights | Spleen |
| Assay | Titers for anti-KLH antibodies |
| Report | One copy of the draft final report<br>One bound and unbound copy of the final and an electronic copy in Word and PDF |
| Archive | Yes for at least 1 year at CRO |

## 2.16 Vaccine Study in Rabbits

| | |
|---|---|
| Purpose | What are the subchronic toxicities and toxicokinetics of a vaccine when administered to rabbits? |
| Timeline | |
| initiation | 6 to 8 weeks following authorization |
| report | 10 to 12 weeks following termination |
| Estimated cost | $400,000 to $450,000 |
| Regulatory guidelines | FDA GLP |
| Total animals | 72 rabbits |
| Number/sex/group | 9 rabbits/sex/group in groups 1–4 |
| Dosing route | Generally intramuscular or subcutaneous |
| Dosing frequency | Dosed based on clinical regimen, generally at 2 to 3 week intervals for 2 to 3 doses |
| Formulations | Usually vaccines are dosed as received if clinical formulation is available |
| Clinical signs | Twice daily |
| Clinical observations | Weekly |
| Body weights | Weekly and prior to necropsy |
| Food consumption | Weekly |
| Draize scoring | 0, 24, 48, 72, 96, and 120 hours after each dose |
| | See Appendix E for scoring |
| Clinical pathology | Prior to study initiation |
| | 2 to 3 days after each dose |
| Hematology | See Appendix B |
| Clinical chemistry | See Appendix B |
| Coagulation | See Appendix B |
| Urinalysis | See Appendix B |
| Terminal necropsy | 3 days after final dose |
| Recovery necropsy | 15 days after final dose |
| Organ weights | See Appendix C |
| Bone marrow smears | Yes from sternum |
| Tissue preservation | See Appendix D |
| Histopathology | Groups 1 and 4 all tissues |
| | Groups 2 and 3—target tissues and gross lesions |
| Report | One copy of the draft final report |
| | One bound and unbound copy of the final and an electronic copy in Word and PDF |
| Archive | Yes for at least 1 year at CRO |

## 2.17 Phototoxicity Study in Rodents

| | |
|---|---|
| Purpose | What is the potential phototoxic effect on the eyes and skin of rodents after a single dose followed by exposure to radiation from a xenon lamp (simulate sunlight)? |
| Timeline | |
| initiation | 4 to 6 weeks following authorization |
| report | 6 to 10 weeks following termination |
| Estimated cost | $40,000 to $50,000 |
| Regulatory guidance and compliance | Photosafety testing (May 2003) and FDA GLP |
| Total animals | 50 Long–Evans rats |
| Number/sex/group | 5/sex/group in groups 1–5 |
| Dosing route | Generally a parenteral route (IV, IM, or sc) for biopharmaceuticals<br>Oral route for pharmaceuticals |
| Dosing frequency | Single dose at 4 dose levels (control, low, mid, and highdose) plus positive control |
| Formulations | Usually biopharmaceuticals dosed as received if clinical formulation is available<br>Pharmaceuticals are formulated by the CRO and require formulation analysis for concentration, homogeneity, and stability |
| Clinical signs | Twice daily |
| Body weights | Daily and prior to necropsy |
| Exposure procedure | At appropriate intervals after dosing (e.g, 24 and 120 hours for eyes, and 4, 24, and 120 hours for skin) rats are anesthetized, restrained, and placed approximately one meter from the UVR source at the time of UVR exposure<br>An instrumental UVR response dose equivalent to approximately 0.5 minimal erythema dose (MED) is delivered to each rat |
| Clinical observations | Prior to exposure and 1 and 4 hours after dosing, continuously during anesthesia and exposure, and daily for 3 days |

| | |
|---|---|
| Skin observations | Prior to exposure, 30 minutes after completion of each UVR exposure, and on days 2, and 3 after exposure |
| Ophthalmological observations | Prior to exposure and at termination |
| Terminal necropsy | Three days after the final UVR exposure |
| Tissue preservation | Eyes and skin from all animals |
| Histopathology | Eyes and skin from control and high-dose animals |
| Report | One copy of the draft final report<br>One bound and unbound copy of the final and an electronic copy in Word and PDF |
| Archive | Yes for at least 1 year at CRO |

### 2.18 Biodistribution Study in Sprague–Dawley Rats

| | |
|---|---|
| Purpose | What is the biodistribution of a gene therapy product when administered to Sprague–Dawley rats after a single dose or after repeated dosing? |
| Timeline | |
| initiation | 4 to 8 weeks following authorization |
| report | 10 to 12 weeks following termination |
| Estimated price | $200,000 to $250,000 |
| Regulatory compliance | FDA GLP |
| Total animals | 24 male and 24 female Sprague–Dawley rats |
| Number/sex/group | 6/sex/group in groups 1–4 |
| Dosing route | Generally a parenteral route (IV, IM, or sc) for biopharmaceuticals |
| Dosing frequency | Either a single dose or dosing daily for 2 to 4 weeks for gene therapy products or based proposed clinical regimen |
| Formulations | Usually gene therapy products are dosed as received if clinical formulation is available. |
| Clinical signs | Twice daily |
| Clinical observations | Weekly |
| Body weights | Weekly and prior to necropsy |
| Food consumption | Weekly |
| Terminal necropsy | Week 4 |

| | |
|---|---|
| PCR analysis | The following tissues are removed in the order listed below with a fresh set of clean instruments for each organ. The tissues (at least 0.5 g sample for each tissue-weight not required for protocol fulfillment) are snap frozen in liquid nitrogen and stored at or below –70 °C:<br>1. Blood (X ≥ 0.6 ml of blood is collected into an EDTA tube and then transferred to a cryovial and snap frozen)<br>2. Ovaries/testes (paired)<br>3. Liver (left lateral lobe)<br>4. Thymus<br>5. Heart (apex)<br>6. Lung (right<br>7. Small intestine (representative sample as site it not specified<br>8. Kidney<br>9. Spleen (median)<br>10. Mesenteric lymph nodes<br>11. Representative skin and subcutis (at exposure site)<br>12. Representative thigh muscle (at exposure site)<br>13. Bone marrow<br>14. Brain<br>All tissues are processed for the presence of the gene therapy agent in the tissues using a GLP validated method for biodistribution. Skin and subcutis and thigh muscle are processed for the incorporation of gene therapy agent using a GLP validated method. Tissues are not be pooled for analysis. Analysis includes qPCR (quantitative polymerase chain reaction) for the presence of the gene therapy agent in the tissue. |
| Report | One copy of the draft final report<br>One bound and unbound copy of the final and an electronic copy in Word and PDF |
| Archive | Yes for at least 1 year at CRO |

### 2.19 Study of Fertility and Early Embryonic Development to Implantation in Sprague–Dawley Rats (Segment 1)

| | |
|---|---|
| Purpose | What are the potential effects of a test article on gonad function, mating behavior, implantation, and general fertility in male and female rats when administered through mating and implantation? |
| Timeline | |
| initiation | 4 to 6 weeks following authorization |
| report | 6 to 8 weeks following termination |
| Estimated cost | $125,000 to $175,000 |
| Regulatory guidance and compliance | ICH-S5A detection of toxicity to reproduction for medicinal products and FDA GLP |
| Total animals | 100 males and 100 females |
| Number/sex/group | 25/sex/group in groups 1–4, control, low dose, mid dose, and high dose, respectively |
| Dosing route | Generally a parenteral route (IV, IM, or sc) for biopharmaceuticals<br>Oral route for pharmaceuticals |
| Dosing frequency | Males dosed for 4 weeks prior to mating (up to 10 may be required), throughout the mating period, and through the day prior to termination<br>Females dosed 2 weeks prior to mating, throughout the mating period, and through GD 7<br>Frequency may differ for biopharmaceuticals based on TK data or proposed clinical regimen |
| Formulations | Usually biopharmaceuticals dosed as received if clinical formulation available<br>Pharmaceuticals are formulated by the CRO and require formulation analysis for concentration, homogeneity, and stability |
| Clinical signs | Twice daily |
| Clinical observations | Weekly |

| | |
|---|---|
| Body weights | Males weighed on the first day of treatment prior to dosing, twice weekly during treatment, and at termination<br>Females weighed on the first day of treatment prior to dosing, twice weekly during premating treatment phase, and twice weekly during mating<br>Confirmed-mated females are also weighed on days 0, 3, 7, 10, and 13 of presumed gestation |
| Estrous cycle determination | During the 2-week dosing premating phase, daily vaginal smears assessed for stage of estrus<br>Estrous cycle determination continued until confirmation of mating occurred or the mating phase ended |
| Mating | 1:1 with identification of both parents of a litter |
| Uterine examinations | Performed at any point after midpregnancy but generally on GD 13<br>Uterine contents are examined for corpora lutea, implantation sites, and live and dead conceptuses |
| Necropsy | Necropsy of males and unconfirmed females include an examination of the external features of the carcass; external body orifices; and cervical, abdominal, and thoracic viscera<br>Special attention given to the organs of the reproductive tract |
| Antibodies (for biopharmaceuticals only) | Prior to initiation and at termination |
| Sperm evaluation | Sperm count and motility in the males surviving to termination. |
| Tissue preservation | Testes, epididymides, ovaries, and uteri from all animals |
| Histopathology | Not conducted unless indicated by treatment-related effects |
| Report | One copy of the draft final report<br>One bound and unbound copy of the final and an electronic copy in Word and PDF |
| Archive | At the CRO for at least one year |

## 2.20 Study for Effects on Embryofetal Development and Toxicokinetics in Rats (Segment 2)

| | |
|---|---|
| Purpose | What is the maternal and embryofetal toxicity and teratogenic potential and toxicokinetics of a test article when administered daily to pregnant rats from implantation to closure of the hard palate? |
| Timeline | |
| initiation | 4 to 6 weeks following authorization |
| report | 8 to 12 weeks following termination |
| Estimated cost | $100,000 to $125,000 |
| Regulatory guidance and compliance | ICH-S5A detection of toxicity to reproduction for medicinal products and FDA GLP |
| Total females | 118 time-mated (≤GD3) females received |
| Number/group | 25/sex/group in groups 1–4, control, low dose, mid dose, and high dose, respectively |
| Number/group for TK | 6/sex/group in groups 1–4 |
| Dosing route and frequency | Generally a parenteral route (IV, IM, or sc) for biopharmaceuticals<br>Oral route for pharmaceuticals<br>Dose GD 6 to 17 for pharmaceuticals, but frequency for a biologic based on TK data or proposed clinical regimen |
| Formulations | Usually biopharmaceuticals dosed as received if clinical formulation available<br>Pharmaceuticals are formulated by the CRO and require formulation analysis for concentration, homogeneity, and stability |
| Clinical signs | Twice daily |
| Body weights | Main study animals weighed on GD 0, 4, 6, 8, 10, 12, 14, 16, 18, and 20<br>Toxicokinetic animals weighed on GD 0, 4, 6, 8, 10, 12, 14, 16, and 18 |
| Food consumption | Beginning on GD 4, food consumption measured at body weight intervals |
| Toxicokinetics | Blood samples obtained from 3 animals/time point on GD 6 and 17 of the treatment phase at approximately 0.5, 1, 2, 4, 8, and 24 hours postdose (±2 min) |

| | |
|---|---|
| Uterine examinations | Performed on GD 20<br>Uterine contents are examined for corpora lutea, number of live and dead fetuses, and gross lesions |
| Antibodies (for biopharmaceuticals only) | Prior to study initiation, and at termination |
| Fetal examinations | Each fetus (live or dead) sexed, weighed, and examined for external abnormalities<br>Approximately half of all fetuses from each litter processed for visceral examination<br>Heads preserved in Bouin's fixative, and evaluated by Wilson's technique<br>Internal organs of the thoracic and abdominal cavities of the fetuses examined in the fresh state using Staples's technique<br>Remaining fetuses eviscerated, processed for skeletal examination using the Alizarin Red S staining method, and evaluated<br>Findings are judged to be variations or malformations:<br>Malformations are developmental deviations that (1) are gross structural changes, (2) are incompatible with life, or (3) may affect the quality of life. Variations are structural deviations that are thought to have no effect on body conformity or the well-being of the animal. |
| Tissue preservation | All fetuses retained in Bouin's fixative or glycerin |
| Report | One copy of the draft final report<br>One bound and unbound copy of the final and an electronic copy in Word and PDF |
| Archive | At the CRO for at lest one year |

### 2.21 Study for Effects on Embryofetal Development and Toxicokinetics in New Zealand White Rabbits (Segment 2)

| | |
|---|---|
| Purpose | What are the maternal and embryofetal toxicity, teratogenic potential, and toxicokinetics of a test article when administered to pregnant rabbits from implantation to closure of the hard plate? |

| | |
|---|---|
| Timeline | |
| initiation | 4 to 6 weeks following authorization |
| report | 8 to 12 weeks following termination |
| Estimated cost | $125,000 to $150,000 |
| Regulatory guidance and compliance | ICH-S5A detection of toxicity to reproduction for medicinal products and FDA GLP |
| Total females | 92 Time-mated female Hra:(NZW)SPF rabbits |
| Number group | 20/group in groups 1–4 |
| Number/group for TK | 3/group in groups 1–4 |
| Dosing route | Generally a parenteral route (IV, IM, or sc) for biopharmaceuticals<br>Oral route for pharmaceuticals |
| Dosing frequency | Dose GD 7 to 20 for pharmaceuticals, but frequency for a biologic may differ based on TK data or proposed clinical regimen |
| Formulations | Usually biopharmaceuticals dosed as received if clinical formulation available<br>Pharmaceuticals are formulated by the CRO and require formulation analysis for concentration, homogeneity, and stability |
| Clinical signs | Twice daily |
| Body weights | Main study rabbits are weighed on GD 0, 4, 7, 9, 11, 13, 15, 18, 21, 24, 27, and 29<br>TK rabbits are weighed on GD 0, 4, 7, 9, 11, 13, 15, and 18 |
| Food consumption | Food consumption is measured for the main study rabbits at GD 4 to 5, 5 to 7, 7 to 9, 9 to 11, 11 to 13, 13 to 15, 15 to 17, 17 to 18, 18 to 20, 20 to 21, 21 to 22, 22 to 24, 24 to 26, 26 to 27, and 27 to 29 |
| Toxicokinetics | Blood samples obtained from 3 animals/time point on GD 7 and 20 of the treatment phase at approximately 0.5, 1, 2, 4, 8, and 24 hours postdose (±2 min) |
| Uterine examinations | Performed on GD 29<br>Uterine contents examined for corpora lutea, number of live and dead fetuses, and gross lesions |
| Antibodies (for biopharmaceuticals only) | Prior to study initiation, and at termination |

| | |
|---|---|
| Fetal examinations | Each live fetus weighed and examined for external abnormalities<br>Live fetuses killed and a midcoronal slice made in the head of each fetus to evaluate the contents of the cranium<br>Internal organs of the thoracic and abdominal cavities of all fetuses examined in the fresh state using Staples's technique, and the sex of each fetus determined<br>Viscera removed and discarded<br>Carcasses processed for skeletal examination using the Alizarin Red S staining method and evaluated<br>Findings judged to be variations or malformations<br>Malformations are developmental deviations that (1) are gross structural changes, (2) are incompatible with life, or (3) may affect the quality of life. Variations are structural deviations that are thought to have no effect on body conformity or the well-being of the animal. |
| Tissue preservation | All fetuses are retained in Bouin's fixative or glycerin |
| Report | One copy of the draft final report<br>One bound and unbound copy of the final and an electronic copy in Word and PDF |
| Archive | At the CRO for at last one year |

### 2.22 Study for Effects on Pre- and Postnatal Development, Including Maternal Function in Sprague–Dawley Rats (Segment 3)

| | |
|---|---|
| Purpose | What are the effects of the test article on pregnant and lactating dams and on the development of the conceptus and offspring following exposure of the dams from implantation through weaning? |
| Timeline | |
| initiation | 4 to 6 weeks following authorization |
| report | 8 to 10 weeks following termination |
| Estimated cost | $200,000 to $250,000 |
| Regulatory guidance and compliance | ICH-S5A detection of toxicity to reproduction for medicinal products and FDA GLP |

| | |
|---|---|
| Total animals | 100 $F_0$ time mated (≤GD 4) females received |
| Number/sex/group | 25/sex/group in groups 1–4 |
| Dosing route | Generally a parenteral route (IV, IM, or sc) for biopharmaceuticals<br>Oral route for pharmaceuticals |
| Dosing frequency | Implantation through the end of lactation, generally daily from GD 6 through PND 20<br>Frequency may differ for biological, based on TK data or proposed clinical regimen<br>$F_1$ generation not dosed |
| Formulations | Usually biopharmaceuticals dosed as received if clinical formulation available<br>Pharmaceuticals are formulated by the CRO and require formulation analysis for concentration, homogeneity, and stability |
| Clinical signs | Twice daily |
| Clinical observations | Weekly |
| $F_0$ Body weights | GD 0, 4, 6, 8, 10, 12, 14, 16, 18, and 20 and on PND 0, 4, 7, 10, 14, 17, and 21 |
| $F_0$ Food consumption | GD 4–6, 6–8, 8–10, 10–12, 12–14, 14–16, 16–18, and 18–20<br>Food consumption will not be measured during lactation |
| $F_1$ Litter observations | At birth—litter size (number alive and dead), and the sex, weight, and clinical observations of individual pups<br>At PND 4—litter size and sex, weight, and observations of individual pups prior to culling; litters with more than 8 pups culled to 4 males and 4 females<br>At PND 7, 14, 21—litter size and sex, weight, and observations of individual pups<br>General development—pinna unfolding (PND 1), surface righting (PND 4), hair growth (PND 7), incisor eruption (PND 7), eye opening (PND 11), and auditory startle (PND 21) |
| Weaning | 1 pup/sex/litter randomly selected for $F_1$ generation |

| | |
|---|---|
| Postweaning | Pup observation may include the following:<br>Vaginal opening—starting on PND 30<br>Cleavage of the balanopreputial gland—starting on PND 35<br>Locomotor activity—PND 22 ± 1 day<br>Water maze—PND 30 ± 2 |
| $F_1$ breeding | Following 7 weeks of postweaning one female and one male from the same groups are selected to cohabit; sibling mating avoided<br>Mating confirmed by presence of vaginal plug or sperm |
| $F_1$ body weights | GD 0, 7, 14, and 20 and on PND 0. |
| $F_2$ litter observations | At birth—litter size (number alive and dead), and the sex, weight, and clinical observations of individual pups<br>PND 0—all pups killed and preserved in 10% NBF |
| Necropsy | $F_0$ females<br>$F_1$ males and females |
| Antibodies (for biopharmaceuticals only) | Prior to study initiation, and at termination |
| Tissue preservation | Gross lesions |
| Histopathology | Not conducted unless indicated by treatment-related effects |
| Report | One copy of the draft final report<br>One bound and unbound copy of the final and an electronic copy in Word and PDF |
| Archive | At the CRO for at least one year |

### 2.23 Evaluation of Cardiovascular (Hemodynamic) Function in the Conscious Telemetered Nonrodent

| | |
|---|---|
| Purpose | What is the effect of the test article on the hemodynamic function in the conscious telemetered nonrodent? |
| Timeline | |
| initiation | 2 to 4 weeks following authorization |
| report | 6 to 8 weeks following termination |
| Estimated cost | $100,000 to $150,000 |
| Regulatory guidelines and compliance | ICH S7A safety pharmacology studies for human pharmaceuticals and FDA GLP |
| Total animals | 8 cynomolgus monkeys or beagle dogs |
| Number/sex | 4/sex/dose |

| | |
|---|---|
| Dosing route | Expected clinical route of administration when feasible<br>Generally a parenteral route (IV, IM, or sc) for biopharmaceuticals and oral route for pharmaceuticals |
| Dosing frequency | 4 animals/sex treated with 4 doses (control, low, mid, and high dose) with at least 7 days between doses |
| Formulations | Usually biopharmaceuticals dosed as received when clinical formulation available<br>Pharmaceuticals are formulated by the CRO and require formulation analysis for concentration, homogeneity, and stability |
| Clinical signs | Twice daily |
| Clinical observations | Prior to and following each dose |
| Body weights | Prior to each dose |
| Cardiovascular analysis | In instrumented dogs, blood pressure (diastolic, mean, and systolic) recorded every 15 minutes for the first 4 hours with hourly measurements thereafter for 24 hours and ECGs to include P duration, PR interval, QRS interval, R amplitude, and QT/QTc interval recorded every 15 to 30 minutes for the first 4 hours and hourly thereafter for 8 hours<br>Times may vary based on toxicokinetic parameters |
| Report | One copy of the draft final report<br>One bound and unbound copy of the final and an electronic copy in Word and PDF |
| Archive | At least one year at CRO |

### 2.24 Neuropharmacology Study in Sprague–Dawley Rats

| | |
|---|---|
| Purpose | What are the neuropharmacology effects (functional observational battery, spontaneous motor activity, and motor coordination) in Sprague–Dawley rats following a single dose? |
| Timeline | |
| initiation | 2 to 4 weeks following authorization |
| report | 6 to 8 weeks after termination |
| Estimated cost | $20,000 to $40,000 |

| | |
|---|---|
| Regulatory guidelines and compliance | ICH S7A safety pharmacology studies for human pharmaceuticals and FDA GLP |
| Total animals | 120 Sprague–Dawley rats |
| Number/sex/group | 10 males/group in groups 1–4 for each test (functional observational battery, spontaneous motor activity, and rotarod)<br>Tests may be conducted in 5 rats/sex/group |
| Dosing route | Expected clinical route of administration when feasible<br>Generally a parenteral route (IV, IM, or sc) for biopharmaceuticals and oral route for pharmaceuticals<br>Oral route usually for pharmaceuticals |
| Dosing frequency | Prior to each test |
| Formulations | Usually biopharmaceuticals dosed as received when clinical formulation available<br>Pharmaceuticals are formulated by the CRO and require formulation analysis for concentration, homogeneity, and stability |
| Functional observational battery | Rats dosed and placed in open field and observed for signs of pharmacological or toxicological activity at 15, 30, and 45 minutes, 1, 2, 3, 4, and 24 hours following treatment<br>Observation times may vary based on the toxicokinetics of the test article<br>Observances for the following will include:<br>Seizures/convulsions, awareness reaction, startle response, vocalization, irritability, decreased abdominal tone, increased secretion, body tremors, decreased grip strength, immobility, motor activity, ataxia, abnormal posture, stereotypy, excretion, decreased respiration, piloerection, loss of righting, pupil size, nociceptive (pain) response, corneal reflex, and pinnal reflex<br>Body temperatures taken on all animals at 60 minutes (±5 min) following dose administration |

| | |
|---|---|
| Rotarod | Rats are pre-trained and act as their own control to maintain themselves on the rotarod for one minute<br>Rats are tested on the rotarod at 30 and 60 minutes (±2 minutes) postdose<br>Observation times may vary based on the pharmacokinetics of the test article |
| Spontaneous motor activity | Rats are orally dosed, and individually placed in a photobeam activity system for 60 minutes postadministration of the test article/vehicle, to count spontaneous motor activity over a 20 minute period<br>Spontaneous motor activity is counted at the following time intervals (minutes): 0–5, 5–10, 10–15, 15–20<br>Observation times may vary based on the pharmacokinetics of the test article |
| Report | One copy of the draft final report<br>One bound and unbound copy of the final and an electronic copy in Word and PDF |
| Archive | At least one year at the CRO. |

### 2.25 Pulmonary Assessment in the Conscious Rat

| | |
|---|---|
| Purpose | What are the potential effects on pulmonary function in the conscious rat? |
| Timeline | |
| initiation | 2 to 4 weeks following authorization |
| report | 4 to 6 weeks after termination |
| Estimated cost | $25,000 to $45,000 |
| Regulatory guidelines and compliance | ICH S7A safety pharmacology studies for human pharmaceuticals and FDA GLP |
| Total animals | 16 male Sprague–Dawley rats |
| Number sex/group | 4 males/group in groups 1–4 |
| Dosing route and frequency | Generally a parenteral route (IV, IM, or sc) for biopharmaceuticals<br>Oral route usually for pharmaceuticals prior to each test |

| | |
|---|---|
| Methods | Pleural pressure transmitters surgically implanted using aseptic surgical techniquein each rat<br>Rats trained initially for two days in the head-out plethysmographic chamber for 15 minutes each<br>On day of dosing, each rat placed in its plethysmographic chamber and allowed to stabilize for 5 minutes<br>Following stabilization, the following respiratory parameters are measured for five minute for baseline levels:<br>Airway Resistance ($cmH_2O$/ml/s), Tidal Volume (ml/breath), minute volume (ml/min), and respiratory rate (breaths/min)<br>Rats then removed from the chambers and dosed with the test article or vehicle<br>30 minutes following the dosing, rats returned to the plethysmographic chamber and respiratory parameters measured continuously for 5 minutes<br>After 1, 2, and 4 hours, each animal returned to its designated plethysmographic chamber and respiratory parameters are measured continuously for 5 minutes following a 5-min stabilization period |
| Report | One copy of the draft final report<br>One bound and unbound copy of the final and an electronic copy in Word and PDF |
| Archive | At least one year at the CRO |

### 2.26 Evaluation of Cardiovascular Function Using Cloned hERG Channels Expressed in Human Embryonic Kidney (HEK293) Cells

| | |
|---|---|
| Purpose | What is the potential of the test article to prolong the human QT interval in vivo by assessing the rapidly inward rectifying potassium current (Kr) conducted by hERG (human ether-a-go-go related gene) channels stably expressed in a HEK293 cell line? |
| Timeline | |
| initiation | 2 to 4 weeks following authorization |
| report | 4 to 6 weeks following termination |
| Estimated cost | $15,000 to $20,000 |

| | |
|---|---|
| Regulatory guidelines and compliance | ICH S7A safety pharmacology studies for human pharmaceuticals and FDA GLP |
| Cells | HEK293 cells from *Homo sapien* embryonic kidney epithelial cell line transformed with adenovirus 5 DNA and maintained in cell culture incubators |
| Dosing route | In vitro |
| Formulations | Biopharmaceuticals or pharmaceuticals dissolved in physiological saline containing 0.1% DMSO to obtain at least 5 different concentrations |
| | Biopharmaceuticals and pharmaceuticals formulated by the CRO and require formulation analysis for concentration, homogeneity, and stability |
| Positive control | Cisapride, a potent blocker of hERG channels, expressed in HEK293 cells and clinically shown to result in QT prolongation and ventricular arrythmias |
| Study design | Cells harvested using trypsin/EDTA and plated onto glass cover slips prior to evaluation |
| | Following whole-cell patch clamp with demonstration of stable hERG current, control values taken followed by exposure of the cells to the vehicle, then each concentration of the test article from lowest to highest concentration and finally the positive control |
| Parameters | Peak tail currents from the last 30 s of the baseline period averaged and compared to 30 s of data recorded in the presence of each test article solution |
| | Current inhibition calculated as percent according to Inhibition (%) = $100 \times (I_{test}/I_{baseline})$, where $I_{test}$ is the peak tail current measured in the presence of the test solution and $I_{baseline}$ is the peak tail current measured prior to exposure to the test solution |
| Report | One copy of the draft final report |
| | One bound and unbound copy of the final and an electronic copy in Word and PDF |
| Archive | At least one year at CRO |

## 2.27 Evaluation of Cardiovascular Function Using Action Potential Parameters in Isolated Dog Purkinje Fibers

| | |
|---|---|
| Purpose | What is the potential of the test article to prolong the human QT interval in vivo by assessing the action potential parameters in the Beagle dog Purkinje fibers? |
| Timeline | |
| initiation | 2 to 4 weeks following authorization |
| report | 4 to 6 weeks following termination |
| Estimated cost | $20,000 to $25,000 |
| Regulatory guidelines and compliance | ICH S7A safety pharmacology studies for human pharmaceuticals and FDA GLP |
| Total animals | 4 beagle dogs |
| Dosing route | In vitro |
| Formulations | Biopharmaceuticals or pharmaceuticals dissolved in Tyrodes's solution to obtain at least 5 different concentrations |
| | Biopharmaceuticals and pharmaceuticals formulated by the CRO and require formulation analysis for concentration, homogeneity, and stability |
| Positive control | dl-soltalol, a β-adrenergic antagonist that prolongs cardiac action potential by selectively blocking the rapidly activating delayed rectifier potassium current (IKr) |
| Study design | 4 Purkinje fibers obtained from 4 different dogs isolated from dog hearts and placed in a recording chamber and perfused with Tyrode's solution at 35–38 °C |
| | Fibers electrically paced via bipolar silver electrodes and impaled with 3M KCl-filled glass microelectrodes |
| | Action potentials continuously monitored |
| | Control values taken, followed by exposure of fibers to vehicle, then each concentration of the test article from lowest to highest concentration and finally the positive control |

| | |
|---|---|
| Action potential parameters | Resting membrane potential (RMP)<br>Maximum rate of depolarization of the action potential upstroke ($V_{max}$)<br>Action potential amplitude (APA)<br>Overshoot (OS)<br>Action potential duration at 30%, 50%, and 90% ($APD_{30}$, $APD_{50}$, and $APD_{90}$) repolarization |
| Report | One copy of the draft final report<br>One bound and unbound copy of the final and an electronic copy in Word and PDF |
| Archive | At least one year at CRO |

### 2.28 *Salmonella typhimurium/Escherichia coli* Plate Incorporation Mutation Assay in the Presence and Absence of Induced Rat Liver S-9

| | |
|---|---|
| Purpose | What is the potential of a test article with and without metabolic activation to produce point mutations in bacteria strains with base substitutions and frameshift mutations in operons coding for synthesis of specific amino acids? |
| Timeline | |
| initiation | 2 to 4 weeks following authorization |
| report | 4 to 6 weeks following termination |
| Estimated cost | $10,000 to $15,000 |
| Regulatory guidelines/compliance | ICH S2A specific aspects of regulatory genotoxicity tests for pharmaceuticals, S2Bgenotoxicity: a standard battery for genotoxicity testing of pharmaceuticals and FDA GLPs |
| Metabolic activation | Standard rat liver S-9 prepared by inducing male Sprague–Dawley rats with Aroclor-1254 or Phenobarbital and/or β-naphthoflavone |
| Test system identification | *Salmonella typhimurium* strains—TA98, TA100, TA1535, TA1537<br>*Escherichia coli* strain—WP2 uvrA |
| Positive controls | 2-AA, 2-NF, 9-AA, NaAz, and MMS |
| Determination of solubility/miscibility | Conducted to determine the maximum achievable concentration of ANS 201 in the selected solvent (water, DMSO, acetone, or ethanol). |

| | |
|---|---|
| Dose selection | Performed with and without S-9 activation using tester strains TA100 and WP2 uvrA only to identify test article concentrations that produce 0–100% toxicity. |
| Mutation assay | Performed using the four *Salmonella typhimurium* strains (TA98, TA100, TA1535, TA1537) and the *Escherichia coli* strain WP2 uvrA strain bu plate incorporation method of treatment<br>Plates incubated at 37 °C for 36 to 72 hours and then scored for number of revertants per plate |
| Confirmatory assay | If mutation assay negative, the same strains and test article concentrations is tested using the preincubation method of treatment |
| Criteria of response | Negative, if all stains treated with test article had mean reversion frequencies that are less than twice those observed in the corresponding solvent control plates in TA98 and TA100 and less than three times in TA 1535, TA1537, and WP2 uvrA, and there is no evidence of a concentration–response<br>Positive, if either TA98 or TA 100 exhibit a mean reversion frequency that is at least double those observed in the corresponding control in at least one concentration, or if TA1535, TA1537, or WP2 uvrA exhibit a threefold increase in mean reversion frequency compared to the solvent control in at least one concentration and the response must be concentration-dependent |
| Report | One copy of the draft final report<br>One bound and unbound copy of the final and an electronic copy in Word and PDF |
| Archive | At least one year at the CRO |

## 2.29 Chemical Induction of Micronucleated Polychromatic Erythrocytes in Mouse Bone Marrow Cells

| | |
|---|---|
| Purpose | What is the potential of a test article to produce clastogenic genetic damage as measured by induced micronucleated polychromatic erythrocytes (MPCE) in bone marrow cells? |
| Timeline | |
| initiation | 2 to 4 weeks following authorization |
| report | 4 to 8 weeks following termination |
| Estimated cost | $15,000 to $25,000 |
| Regulatory guidelines/ compliance | ICH S2A specific aspects of regulatory genotoxicity tests for pharmaceuticals, S2Bgenotoxicity: a standard battery for genotoxicity testing of pharmaceuticals and FDA GLPs |
| Abbreviations | PCE—polychromatic erythrocytes<br>NCE—normochromatic erythrocytes<br>MPC—micronucleated polychromatic erythrocytes |
| Total mice | 90 CD-1 mice (or rats) |
| Number/sex/group | 10/sex/group in groups 1–4 (control, low, mid, and high dose) and 5/sex in the positive control group |
| Positive control | Cyclophosphamide (CP) at 80 mg/kg or trimethylenemelamine (TEM) at 1.0 mg/kg |
| Dosing route | Generally a parenteral route (IV, IM, or sc) for biopharmaceuticals<br>Oral route for CP or intraperitoneal for TEM |
| Range finding | Performed only if sufficient information is not available on the toxicity of the test article in mice. |
| Assay | Approximately 24, 48, or 72 hours after dosing, mice killed and bone marrow obtained from both femurs<br>Slide of a bone marrow suspension prepared, stained with Wright-Giemsa stain, dried, and scored blind<br>Number of PCE among the total erythrocytes (PCE + NCE) determined for each animal by counting at least 200 erythrocytes<br>Number of MPCE scored for 1000 or 2000 PCE per animal |

| | |
|---|---|
| Criteria of response | Positive, if there is a positive dose–response trend or statistically significant increase in the number of MPCE at one or more dose levels compared to control |
| Report | One copy of the draft final report<br>One bound and unbound copy of the final and an electronic copy in Word and PDF |
| Archive | At least one year at the CRO |

### 2.30 Lymphoma Mutagenesis in the L5178Y TK+/– Mouse with Colony Size Evaluation in the Presence and Absences of Induced Rat Liver S-9 with a Confirmatory Study

| | |
|---|---|
| Purpose | What is the potential of a test article with and without metabolic activation to induce forward mutations at the thymidine kinase locus of L5178Y TK+/– mouse lymphoma cells? |
| Timeline | |
| initiation | 2 to 4 weeks following authorization |
| report | 4 to 8 weeks following termination |
| Estimated cost | $15,000 to $25,000 |
| Regulatory guidelines/compliance | ICH S2A specific aspects of regulatory genotoxicity tests for pharmaceuticals, S2B genotoxicity: a standard battery for genotoxicity testing of pharmaceuticals and FDA GLPs |
| Metabolic activation | Standard rat liver S-9 prepared by inducing male Sprague–Dawley rats with Aroclor-1254 or Phenobarbital and/or β-naphthoflavone |
| Positive control | Hycanthone methanesulfonate (HYC)—induces mutations at the TK locus without metabolic activation<br>7,12-Dimethylbenz[α]anthracene (DMBA)—induces mutations at the TK locus with metabolic activation |
| Dose selection | Performed to identify the concentrations that produce 0–100% toxicity. |

| | |
|---|---|
| Mutation assay | Positive controls (HYC and DMBA) and five concentrations of the test article used with and with metabolic activation<br>Cells exposed to each dose, and the supernatant obtained, and incubated for 20 and 44 hours<br>After a 2-day expression period, the cultures cloned with the restrictive agent trifluorothymidine (allows the growth of TK–/– cells only) or vehicle control |
| Results | Mutation frequency (MF) calculated as mean of all mutants per plate ÷ mean of all colonies in corresponding plates<br>Induced MF also calculated |
| Criteria | Positive response is if at least one culture had an MF that is two times or greater than the average MF of the corresponding solvent control cultures, and the response is concentration dependent |
| Confirmatory assay | Performed without S-9 activation and exposure period extended to 24 hours to confirm mutation assay results |
| Report | One copy of the draft final report<br>One bound and unbound copy of the final and an electronic copy in Word and PDF |
| Archive | At least one year at the CRO |

### 2.31 In vitro Chromosomal Aberration

| | |
|---|---|
| Purpose | What is the potential of a test article with and without metabolic activation to induce structural chromosome alterations in Chinese hamster ovary (CHO) cells? |
| Timeline | |
| initiation | 2 to 4 weeks following authorization |
| report | 4 to 8 weeks following termination |
| Estimated cost | $15,000 to $25,000 |
| Regulatory guidelines/compliance | ICH S2A specific aspects of regulatory genotoxicity tests for pharmaceuticals, S2B genotoxicity: a standard battery for genotoxicity testing of pharmaceuticals and FDA GLPs |

| | |
|---|---|
| Metabolic activation | Standard rat liver S-9 prepared by inducing male Sprague–Dawley rats with Aroclor-1254 or Phenobarbital and/or β-naphthoflavone |
| Positive control | Methyl methanesulphonate, ethyl methanesulfonate, ethylnitrosurea, mitomycin, or 4-nitroquinone-*N*-oxide for the nonmetabolic activation and benzo(a)pyrene or cyclophosphamide for the metabolic activation |
| Dose selection | Performed as the loss of growth potential of cells induced by a 4-hour expression in growth medium |
| Assay | Approximately $10^6$ cells treated with at least three doses of the test article with and without the S9 fraction for two hours at 37 °C in growth medium |
| | Exposure period stopped by two washing the cells twice with saline |
| | Following the addition of fresh growth medium and BrdU (5-bromo-2′-deoxyuridine), cells incubated for 20 hours. |
| | Colcemid included in the medium the last three hours |
| | Slides are then prepared with 10% Giemsa for scoring of chromosomal aberrations |
| Results | Standard forms are used to score 100 to 200 cells for gaps, breaks, fragments, reunion figures, and numerical aberrations such as polyploidy cells |
| Criteria | Type of aberration, frequency, and dose–response considered when evaluating the potential of a test article to induce chromosomal aberrations |
| Report | One copy of the draft final report |
| | One bound and unbound copy of the final and an electronic copy in Word and PDF |
| Archive | At least one year at the CRO |

## 2.32 Tissue Cross-Reactivity

| | |
|---|---|
| Purpose | What is the potential for expected or unexpected tissue binding (or cross-reactivity) of antibodies (test articles) in human and animal tissues?<br>What is the relevance of a given species for use in toxicity studies with that antibody? |
| Timeline | |
| initiation | 2 to 4 weeks following authorization |
| report | 4 to 6 weeks following termination |
| Estimated cost | Human tissue cross-reactivity (3 donors) $45,000 to $50,000<br>Animal tissue cross-reactivity (2 donors) $65,000 to 70,000 |
| Regulatory guidelines/ compliance | No specific guidelines and FDA GLP optional |
| Staining methods | Avidin-biotin (ABC) for biotinylated test articles or<br>Tertiary antibody detection for FITC labeled test article or<br>Precomplexing with a labeled anti-human IgG for an unlabeled test article |
| Negative control | Tissue element or cell line that does not express the target antigen or<br>Beads coated with an irrelevant antigen or<br>Irrelevant antigen spotted and cross-linked to UV-resin slides |
| Assay control | Used to define background |
| Positive control | Tissue element or cell line known to express the target antigen or<br>Sepharose or agarose beads coated with target antigen target antigen spotted and cross-linked on UV-resin slides |
| Results | Specific reactions of the antibody with the positive control material and lack of specific reactivity with the negative control material demonstrate the sensitivity, specificity, and reproducibility of the assay<br>In a typical cross-reactivity assay, a staining method most appropriate for the antibody is developed<br>In a typical tissue cross-reactivity study (see Appendix J, cryosections of normal human (3 unrelated donors) and/or animal (2 unrelated donors) tissues are stained |

| | |
|---|---|
| Criteria | Slides initially evaluated to determine if the tissue is adequate and normal<br>Staining judged specific (CDR mediated) or nonspecific (non-CDR mediated) by comparison with control slides<br>Any specific staining judged to be an expected or unexpected reactivity based on the known expression of the target antigen<br>Any staining judged to be specific scored for intensity, frequency, and staining affinity |
| Report | One copy of the draft final report<br>One bound and unbound copy of the final and an electronic copy in Word and PDF |
| Archive | At least one year at the CRO |

## APPENDIX A NORMAL HUMAN TISSUES USED IN TISSUE CROSS-REACTIVITY TESTING

Adrenal
Bladder
Blood cells
Bone marrow
Breast
Cerebellum
Cerebral cortex
Colon
Endothelium
Eye
Fallopian tube
Gastrointestinal tract
Heart
Kidney (glomerulus, tubule)
Liver
Lung
Lymph node
Ovary
Pancreas
Parathyroid

Pituitary
Placenta
Prostate
Skin
Spinal cord
Spleen
Striated muscle
Testis
Thymus
Thyroid
Ureter
Uterus (cervix, endometrium)

## APPENDIX B CLINICAL PATHOLOGY PARAMETERS

### Hematology

Red blood cell (erythrocyte) count
Hemoglobin
Hematocrit
Mean corpuscular volume
Mean corpuscular hemoglobin
Mean corpuscular hemoglobin concentration
Platelet count
White blood cell (leukocyte) count
Differential blood cell count
Blood smear
Reticulocyte count

### Coagulation

Prothrombin time
Fibrogen (optional)
Activated partial thromboplastin time

### Clinical Chemistry

Glucose
Urea nitrogen
Creatinine
Total protein
Albumin
Globulin
Albumin/globulin ratio
Cholesterol
Total bilirubin
Alanine aminotransferase
Alkaline phosphatase
Gamma glutamyltransferase
Aspartate aminotransferase
Calcium
Inorganic phosphorus
Sodium
Potassium
Chloride

### Urinalysis

Bilirubin
Glucose
Microscopic examination of sediment
Protein
Volume
Appearance/color
Blood
Ketones
pH
Specific gravity
Urobilinogen

## APPENDIX C ORGAN WEIGHTS

At scheduled sacrifices, the following organs (when present) are weighed; paired organs are weighed together:

Adrenal (2)
Brain
Epididymis (2)
Heart
Kidney (2)
Liver
Lung
Ovary (2)
Pituitary gland
Prostate
Salivary gland (mandibular) (2)
Seminal vesicle
Spleen
Testis (2)
Thymus
Thyroid (2 lobes) with parathyroid
Uterus

Organ-to-body weight and organ-to-brain weight ratios are generally reported as percentages

## APPENDIX D TISSUE PRESERVATION

The following tissues (when present) from each animal are preserved in 10% neutral-buffered formalin, unless otherwise indicated below:

Adrenal (2)
Aorta
Brain
Cecum
Colon
Duodenum
Epididymis (2)[a]
Esophagus
Eye (2 with optic nerve)[a]
Femur with bone marrow (articular surface of the distal end)
Gallbladder (except rats)
Harderian gland[a]
Pancreas
Pituitary gland
Prostate
Rectum
Salivary gland [mandibular (2)]
Sciatic nerve
Seminal vesicle (except dogs)
Skeletal muscle (biceps femoris)
Skin
Spinal cord (cervical, thoracic and lumbar)
Spleen
Sternum with bone marrow
Stomach

Heart
Ileum
Injection site(s) if applicable
Jejunum
Kidney (2)
Lesions
Liver
Lung with large bronchi
Lymph node (mesenteric and/or mesenteric)
Mammary gland (females)
Ovary (2)
Testis (2)[a]
Thymus
Thyroid (2 lobes) with parathyroid
Tongue
Trachea
Urinary bladder
Uterus
Vagina

[a]Preserved in modified Davidson's fixative.

## APPENDIX E DRAIZE SCORES

The following scores are based on the Draize scale for scoring skin irritation:

### ERYTHEMA AND ESCHAR FORMATION

0 No erythema
1 Very slight erythema (barely perceptible)
2 Well-defined erythema
3 Moderate to severe erythema
4 Severe erythema (beet redness) to slight eschar formation (injuries in depth)

Maximum possible = 4

### EDEMA FORMATION

0 No edema
1 Very slight edema (barely perceptible)
2 Slight edema (edges of area well defined by definite raising)
3 Moderate edema (raised approximately 1 mm)
4 Severe edema (raised more than 1 mm and extending beyond area of exposure)

Maximum possible = 4

J. H. Draize, G. Woodard, and H. O. Calvery (1944). Methods for the study of irritation and toxicity of substances applied topically to the skin and mucous membranes. *J. Pharmacol. Exp. Ther.* 82: 377–390.

## APPENDIX F ANALYTICAL/BIOANALYTICAL COSTS

### Dose Analysis

| | |
|---|---|
| *Method validation* | $10,000 to $20,000 |

- Method transferred or developed using an aqueous vehicle
- Validation includes the proposed concentrations and storage periods to be used in the toxicology studies
- Stability determined for 24 hours and up to 8 days
- Report is included in the study report
- Analyses conducted following GLPs

| | |
|---|---|
| Study analysis | $5000 to $7500 |

### Bioanalytical Method Validation and Sample Analysis

| | |
|---|---|
| Method validation for a pharmaceutical (generally LS/MS/MS) or biopharmaceutical (ELISA) | $15,000 to $30,000 |

- Method transferred or developed using an serum or plasma
- Validation includes the proposed concentrations anticipated in the toxicology studies
- Long-term stability included
- Report is included in the study report
- Analyses conducted following GLPs
- Method validation for additional species is generally about 75% to 80% of initial species

| | |
|---|---|
| Individual sample analysis | $60 to $100 |

### Toxicokineic Blood Collection Cost per Time Interval

| | |
|---|---|
| Rodent | $10,000 to $20,000 |
| Nonrodent | $15,000 to $25,000 |

## APPENDIX G SPECIAL STAINS

| STAIN | AFFINITY |
|---|---|
| Alcian Blue | Mucopolysaccharides |
| Acid Fuchsin | Connective tissue |
| Dunn-Thompson | Hemoglobin |
| Hematoxylin and Eosin | Nuclei (hematoxylin) and cytoplasm (eosin) |

| STAIN | AFFINITY |
|---|---|
| Immunoperoxidase (Horseradish peroxidase) | Antibody–antigen interactions |
| Immunofluorescence (FITC, Rhodamine, or Texas Red) | Antibody–antigen interactions |
| Oil Red O (ORO) | Fat |
| PAS (Periodic Acid-Schiff) | Glycogen, mucin, mucoprotein, glycoprotein, fungi |
| Perl's Iron Stain | Iron |
| Silver Stain (Gomori Methenamine Silver Stain) | Fungi and *Pneumocystis carinii* |
| Trichrome | Collagen |
| Verhoeff'Elastic Stain | Elastic fibers |
| Wright-Giemsa | Peripheral blood smears |
| Wright Stain | Differentiation of blood cells |

## APPENDIX H REFERENCES

For additional details see ICH Guidance Documents at:
http://www.fda.gov/cder/guidance/index.htm

Single Dose Acute Toxicity Testing for Pharmaceuticals (1996)

| | |
|---|---|
| M3 Nonclinical Safety Studies for the Conduct of Human Clinical Trials for Pharmaceuticals | 11/1997 |
| S1C Dose Selection for Carcinogenicity Studies of Pharmaceuticals | 3/1995 |
| S1C(R) Guidance on Dose Selection for Carcinogenicity Studies of Pharmaceuticals: Addendum on a Limit Dose and Related Notes | 12/4/1997 |
| S2A Specific Aspects of Regulatory Genotoxicity Tests for Pharmaceuticals | 4/1996 |
| S2B Genotoxicity: A Standard Battery for Genotoxicity Testing of Pharmaceuticals | 11/21/1997 |
| S3A Toxicokinetics: The Assessment of Systemic Exposure in Toxicity Studies | 3/1995 |
| S3B Pharmacokinetics: Guidance for Repeated Dose Tissue Distribution Studies | 3/1995 |
| S4A Duration of Chronic Toxicity Testing in Animals (Rodent and Nonrodent Toxicity Testing) | Posted 6/25/99 |
| S5A Detection of Toxicity to Reproduction for Medicinal Products | 9/1994 |
| S5B Detection of Toxicity to Reproduction for Medicinal Products: Addendum on Toxicity to Male Fertility | 4/1996 |

| | |
|---|---|
| S6 Preclinical Safety Evaluation of Biotechnology-Derived Pharmaceuticals | 11/1997 |
| S7A Safety Pharmacology Studies for Human Pharmaceuticals | 7/2001 |
| S7B Nonclinical Evaluation of the Potential for Delayed Ventricular Repolarization (QT Interval Prolongation) by Human Pharmaceuticals | 10/19/2005 |
| S8 Immunotoxicity Studies for Human Pharmaceuticals | 4/12/2006 |

Points to Consider in the Manufacture and Testing of Monoclonal Antibody Products for Human Use (1997)

## APPENDIX I ESTIMATED COSTS FOR GLP COMPLIANT PRECLINICAL STUDY DESIGNS TO SUPPORT CLINICAL DEVELOPMENT

| Study Title | Guidelines | Timelines: Weeks to Initiation after Authorization | Timelines: Weeks for Report after Termination | Estimated cost (USD) |
|---|---|---|---|---|
| *Safety Pharmacology* | | | | |
| Evaluation of Cardiovascular Function Using Cloned hERG Channels Expressed in Human Embryonic (HEK293) Cells | ICH S7B The Nonclinical Evaluation of the Potential for Delayed Repolarization (QT Interval Prolongation) by Human Pharmaceuticals | 2–4 | 4–6 | 15,000–20,000 |
| Evaluation of Cardiovascular Function Using Action Potential Parameters in Isolated Purkinje Fibers | ICH S7B The Nonclinical Evaluation of the Potential for Delayed Repolarization (QT Interval Prolongation) by Human Pharmaceuticals | 2–4 | 4–6 | 20,000–25,000 |
| Evaluation of Cardiovascular (Hemodynamic) Function in the Conscious Telemetered—Nonrodent | ICH S7A Safety Pharmacology Studies for Human Pharmaceuticals and ICH S7B The Nonclinical Evaluation of the Potential for Delayed Repolarization (QT Interval Prolongation) by Human Pharmaceuticals | 2–4 | 6–8 | 100,000–150,000 |
| Neuropharmacology study in Sprague-Dawley Rats | ICH S7A Safety Pharmacology Studies for Human Pharmaceuticals | 2–4 | 6–8 | 20,000–40,000 |
| Pulmonary Assessment in the Conscious Rat | ICH S7A Safety Pharmacology Studies for Human Pharmaceuticals | 2–4 | 4–8 | 25,000–45,000 |
| Tissue Cross-Reactivity (TCR) | | | | |
| Human tissue cross-reactivity (3 donors) | NA | 2–4 | 4–6 | 45,000–50,000 |
| Animal tissue cross-reactivity (2 donors) | NA | 2–4 | 4–6 | 65,000–70,000 |

**APPENDIX I** ***Continued***

| Study Title | Guidelines | Timelines: Weeks to Initiation after Authorization | Timelines: Weeks for Report after Termination | Estimated cost (USD) |
|---|---|---|---|---|
| *General Toxicity* | | | | |
| Dose Escalation Study in Rodents[a] | NA | 4–6 | 4–8 | 15,000–35,000 |
| Dose Escalation Study in Nonrodents[b] | NA | 4–8 | 4–8 | 50,000–75,000 |
| Acute Toxicology Study in Rodents[a] | Single Dose Acute Toxicity Testing for Pharmaceuticals (August 1996) | 4–6 | 6–10 | 20,000–40,000 |
| Acute Toxicology Study in Nonrodents[b] | Single Dose Acute Toxicity Testing for Pharmaceuticals (August 1996) | 4–8 | 6–10 | 100,000–250,000 |
| Four-Week Toxicology Study in Rodents[a] | ICH M3 Timing of Nonclinical Studies for the Conduct of Human Clinical Trials for Pharmaceuticals | 4–8 | 10–12 | 200,000–250,000 |
| Four-Week Toxicology Study in Nonrodents[b] | ICH M3 Timing of Nonclinical Studies for the Conduct of Human Clinical Trials for Pharmaceuticals | 6–8 | 10–12 | 400,000–650,000 |
| Three-Month Toxicology Study in Rodents[a] | ICH M3 Timing of Nonclinical Studies for the Conduct of Human Clinical Trials for Pharmaceuticals | 6–8 | 10–12 | 250,000–300,000 |
| Three-Month Toxicology Study in Nonrodents[b] | ICH M3 Timing of Nonclinical Studies for the Conduct of Human Clinical Trials for Pharmaceuticals | 4–8 | 10–12 | 450,000–500,000 |

| | | | | |
|---|---|---|---|---|
| Six-Month Toxicology Study in Rodents[a] | ICH S4A Duration of Chronic Toxicity Testing in Animals (Rodent and Non-rodent Testing) | 6–10 | 10–12 | 350,000–550,000 |
| Nine-Month Toxicology Study in Nonrodents[a] | ICH S4A Duration of Chronic Toxicity Testing in Animals (Rodent and Non-rodent Testing) | 8–12 | 10–12 | 450,000–650,000 |
| *Genotoxicity* | | | | |
| *Salmonella typhimurium/ Escherichia coli* Plate Incorporation Mutation Assay in the Presence and Absence of Induced Rat Liver S-9 | ICH S2A Specific Aspects of Regulatory Genotoxicity Tests for Pharmaceuticals, ICH S2B Genotoxicity: A Standard Battery of Genotoxicity Testing of Pharmaceuticals | 2–4 | 4–6 | 10,000–15,000 |
| In vitro Chromosomal Aberration | ICH S2A Specific Aspects of Regulatory Genotoxicity Tests for Pharmaceuticals and ICH S2B Genotoxicity: A Standard Battery of Genotoxicity Testing of Pharmaceuticals | 2–4 | 4–8 | 15,000–25,000 |
| Lymphoma Mutagenesis in the L5178YTK +/– Mouse with Colony Size Evaluation in the Presence and Absence of Induced Rat Liver S-9 with a Confirmatory Study | ICH S2A Specific Aspects of Regulatory Genotoxicity Tests for Pharmaceuticals and ICH S2B Genotoxicity: A Standard Battery of Genotoxicity Testing of Pharmaceuticals and | 2–4 | 4–8 | 15,000–25,000 |
| Chemical Induction of Micronucleated Polychromatic Erythrocytes in Mouse Bone Marrow Cells | ICH S2A Specific Aspects of Regulatory Genotoxicity Tests for Pharmaceuticals and ICH S2B Genotoxicity: A Standard Battery of Genotoxicity Testing of Pharmaceuticals | 2–4 | 4–8 | 15,000–25,000 |
| *Carcinogenicity Assessment* | | | | |
| Carcinogenicity Study in Sprague-Dawley Rats | ICH S1A The Need for Long-term Rodent Carcinogenicity Studies of Pharmaceuticals and ICH S1B Testing for Carcinogenicity of Pharmaceuticals (2yr) | 8–10 | 26–30 | 1–1.5 M |

**APPENDIX I** ***Continued***

| Study Title | Guidelines | Timelines: Weeks to Initiation after Authorization | Timelines: Weeks for Report after Termination | Estimated cost (USD) |
|---|---|---|---|---|
| Carcinogenicity Study in Mice | ICH S1A The Need for Long-term Rodent Carcinogenicity Studies of Pharmaceuticals and ICH S1B Testing for Carcinogenicity of Pharmaceuticals (2 yr) | 8–10 | 26–30 | 1–1.5 M |
| Carcinogenicity Testing Using the Transgenic Mouse (p53+/–, ras H2, or Tg.AC) | ICH S1A The Need for Long-term Rodent Carcinogenicity Studies of Pharmaceuticals and ICH S1B Testing for Carcinogenicity of Pharmaceuticals (6 months) | 12–16 | 13–17 | 650,000–750,000 |
| *Reproductive/Developmental Toxicity Studies* | | | | |
| Study of Fertility and Early Embryonic Development to Implantation in Sprague-Dawley Rats (Segment 1) | ICH S5A Detection of Toxicity to Reproduction for Medicinal Products and FDA | 4–6 | 6–8 | 125,000–175,000 |
| Study for Effects on Embryofetal Development and Toxicokinetics in Rats (Segment 2) | ICH S5A Detection of Toxicity to Reproduction for Medicinal Products | 4–6 | 8–12 | 100,000–125,000 |
| Study for Effects on Embryofetal Development and Toxicokinetics in New Zealand White Rabbits (Segment 2) | ICH S5A Detection of Toxicity to Reproduction for Medicinal Products | 4–6 | 8–12 | 125,000–150,000 |

| | | | | |
|---|---|---|---|---|
| Study for Effects on Embryofetal Development and Toxicokinetics in the Monkey (Segment 2) | ICH S5A Detection of Toxicity to Reproduction for Medicinal Products | 39–52 | 13–26 | 750,000–1,250,000 |
| Study for Effects on Pre- and Postnatal Development, Including Maternal Function in Sprague-Dawley Rates (Segment 3) | ICH S5A Detection of Toxicity to Reproduction for Medicinal Products | 4–6 | 8–10 | 200,000–250,000 |
| Study for Effects on Pre- and Postnatal Development, Including Maternal Function in the Monkey (Segment 3) | ICH S5A Detection of Toxicity to Reproduction for Medicinal Products | 39–52 | 21–34 | 1,000,000–1,500,000 |
| Local Tolerance | | | | |
| Rabbit Ear Vein | | 2–4 | 4–6 | 1000–5000 |
| Other Toxicity Studies | | | | |
| Irritation and sensitization—rabbit, guinea pig | NA | 2–4 | 4–6 | 9000–15,000 |
| Hemolytic Potential | NA | 2–4 | 4–6 | 6000–10,000 |
| Photoxicity Study in Rodents[a] | NA | 4–6 | 6–10 | 40,000–50,000 |

[a]Rodent (mouse or rat).
[b]Nonrodent (dog or monkey).
[c] ICH Guidance Documents (www.ich.org).
[d]Single Dose Acute Toxicity Testing for Pharmaceuticals (August 1996).

# Practical Considerations in the Design of Preclinical Safety Assessments for Biopharmaceuticals

DAMON R. DEMADY, PhD

**Contents**

1 Introduction
2 Species Selection
3 Animal Use, Numbers, and Gender
4 Quality of the Test Article
5 Dose Selection and Administration
  5.1 Determination of No Observed Adverse Effect Level (NOAEL)
  5.2 Determination of the Maximum Tolerated Dose (MTD)
6 Pharmacokinetics (PK) and Toxicokinetics (TK)
7 Blood Sampling and Maximum Blood Volume
8 Assessment of Toxicity
  References

## 1 INTRODUCTION

While good scientific principles should always be applied to preclinical assessments of biopharmaceuticals, practical considerations for the implementation of said methods should be given equal consideration when designing and conducting these studies. Some of these considerations are the subject of this chapter and will be discussed briefly mainly through tabular illustration.

A number of recent FDA developments and initiatives have addressed novel approaches to biopharmaceutical product development and regulatory review, ranging from reorganization that has led to a fundamental shift of primary regulatory authority for therapeutic biologics from the Center for Biologics Evaluation and Research (CBER) to the Center for Drug Evaluation and

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

Research (CDER) through guidance documents and concept papers on key initiatives issued or under way. The result of the transfer of the biopharmaceuticals from CBER to CDER has resulted in a partial adaptation of CDER guidelines and preferences superimposed on prior CBER and ICH precedents as they relate to the contrasts between biologics and small- molecule drug development. More and more, regulatory agencies are requiring assessments for biologics that have increasingly similar features to what one would undertake in a small-molecule development program, although it is not clear in all cases that the additional studies that may be required provide important additional information to support clinical trials.

The primary goals of preclinical safety evaluation articulated in ICH S6 are namely (1) to identify an initial safe dose and subsequent dose escalation schemes in humans, (2) to identify potential target organs for toxicity and for the study of whether such toxicity is reversible, and (3) to identify safety parameters for clinical monitoring. These goals are generally accepted across all product classes.

Small-molecule pharmaceuticals typically require two species for general toxicology evaluation that have large historical databases within a CRO and/or the public literature. In contrast, biologics should be evaluated in at least one, preferably two, species in which the test article is biologically active, namely a receptor or other entity that specifically recognizes the test article and elicits a specific pharmacological response.

## 2 SPECIES SELECTION

The evaluations to determine relevant species for toxicity evaluation include various receptor binding assays and tissue cross-reactivity assessments, the later being routinely performed for monoclonal antibody based products (see Chapters Tissue Cross-Reactivity Studies for Monoclonal Antibodies: Predictive Value and Use for Selection of Relevant Animal Species for Toxicity Testing and Preclinical Safety Evaluation of Monoclonal Antibodies). Species specificity can be determined using properly controlled in vitro cell-based response assays, cloning of receptors and ligands to determine compatibility, receptor-based response assays, immunoassays, genomic-based assays, and traditional biochemical-based assays as well as in vivo assessments with validated endpoints and markers for specificity.

The ICH S6 guidance states that toxicity studies in nonrelevant species may be misleading and so are discouraged. When no relevant species exists, the use of relevant transgenic animals expressing the human receptor or the use of homologous proteins should be considered.

A pragmatic consideration for long-term toxicology studies of biopharmaceuticals is the development of antibodies against the test article in question. If these antibodies are neutralizing, then exposure to active drug over the long term (roughly longer than 3–4 weeks) is unlikely, negating the utility of the study to qualify the safety and characterize exposure in the species in question.

It is not uncommon for biopharmaceuticals not to be pharmacologically active in the standard toxicology species of rats, dogs, or pigs. While surrogate molecules active in rodent species can be acceptable, typically a nonhuman primate (NHP) is the relevant species to assess safety and characterize exaggerated pharmacology, but properly controlled studies should be conducted to determine responsive species in particular to justify the use of NHPs.

## 3 ANIMAL USE, NUMBERS, AND GENDER

Once a species is selected, the number of animals per group must be decided. The interrelationship between the probability of a given toxic effect in humans and the numbers of animals required in a toxicology study to ensure the same toxic effect can be observed in Table 1. Clearly, to pick up low probability events would be prohibitive in terms of animal usage. In efforts to maximize the chances of seeing an effect, it is generally accepted that toxicology studies use a lower number of animals per group number, and use a higher dose multiple or dose frequency (within reason, ethical guidelines, and test article-driven feasibility) to compensate for commonly underpowered studies especially NHP studies. This approach maximizes the chances of observing relevant test article-related toxicity. Particularly in the case of NHPs, higher dose multiples, more frequent monitoring, and sample collections incorporating noninvasive pharmacodynamic or toxicity evaluations and longer duration studies, are factors that can help compensate for smaller sample size.

On average, 10 to 15 rodents and 3 to 5 nonrodents per sex per group are used for repeat-dose toxicity studies. If possible, 5 rodents and 2 to 3 nonrodents per sex per group should be used for recovery arms. The reuse of nonrodent species is a common practice, particularly for pharmacokinetic studies and in studies in which telemetry is used. Early pilot non-GLP studies for

**TABLE 1 Relationship between animal numbers and observing human toxicity**

| Probability of Toxic Effects in Human | Animals in Experiments ($p$ = 0.95) |
|---|---|
| 100 | 1 |
| 80 | 2 |
| 60 | 4 |
| 50 | 5 |
| 40 | 8 |
| 20 | 14 |
| 10 | 29 |
| 5 | 59 |
| 2 | 149 |
| 1 | 299 |
| 0.1 | 2,995 |
| 0.01 | 29,956 |

dose selection purposes may also involve the reuse of animals. However, it may not be appropriate in all instances to reuse animals based on the level of invasiveness of the procedure, the degree of pain and distress, or the potential for underlying toxicity from a previous compound. When testing is performed on a novel compound or an innovative family class, or when a rare effect is expected, minimization of animal numbers per study group could be risky or counterproductive [1].

To investigate a possible gender effect, regulatory testing should be performed on both genders. One gender can be used for investigative and exploratory studies (e.g., mechanistic studies). Such an approach may also be possible in pilot studies when the absence of a gender effect has been confirmed previously or if the intended clinical indication is for use in a single gender. In safety and toxicity studies, the use of a control group is scientifically justified. Current guidelines recommend the use of control groups in some cases, and this may mean both positive and negative controls. In most toxicity studies, control groups are necessary to evaluate test article related effects. Individual animal variation among nonrodent animals remains an important consideration. The control group is critical when evaluating subtle clinical effects or histological lesions because such findings can be hidden by spontaneous anomalies and individual animal variability. In most toxicity studies a control group is required. However, in some cases, a control group may not be required (e.g., in pilot, exploratory, or dose-finding studies, or when there are adequate in-house historical data of the vehicle used in the formulation) [1].

## 4 QUALITY OF THE TEST ARTICLE

"Safety concerns may arise from the presence of impurities or contaminants. It is preferable to rely on purification processes to remove impurities and contaminants rather than to establish a preclinical testing program for their qualification. In all cases, the product should be sufficiently characterized to allow an appropriate design of preclinical safety studies. In general, the product that is used in the definitive pharmacology and toxicology studies should be comparable to the product proposed for the initial clinical studies." (ICH S6)

One of the primary issues repeatedly arising from the guideline and from the very nature of the complexities of process development, scale up, and eventual large-scale manufacture of biologics is what material is appropriate to use during the various stages of nonclinical development. As stated above, the material should be representative of the anticipated clinical material. This needs to be supported by data, namely identity, purity, stability, and potency. See Table 2 for suggested analytical tests for test article used in preclinical studies.

The material need not be made under fully GMP compliant conditions to be comparable or representative of the clinical lots. As with all material used for pivotal IND enabling studies, the characterization described in the certificate of analysis should be conducted under GLP conditions. Lots used for

**TABLE 2 Suggested analytical tests for test articles used in preclinical studies**

| | Dosage Form Sampled | | | | |
|---|---|---|---|---|---|
| Analytical Test Performed | Sterile Injectable | Pellet Implant | Solution or Ointment for Ophthalmic Administration | Inhalation Products | Oral Dosage Forms |
| Sterility | X | X | | X | |
| Identity | X | X | X | X | X |
| Assay (potency) | X | X | X | X | X |
| Purity | X | X | X | X | X |
| Endotoxin (LAL) | X | | | | |
| Particulates/aggregates | X | | | | |
| Release rate | | X | | | |
| pH | X | | X | X | X |
| Microbial limits | X | X | X | X | X |
| Contaminants (product related, process related) | X | X | X | X | X |
| Dissolution | | | | | X |
| Homogeneity | | | | X | X |

toxicology, pharmacology, and safety assessments should be less pure than clinical material if they are to be different than the clinical lot.

Bioburden testing is required for biologics used in pivotal preclinical studies to support clinical development and registration. Endotoxin levels are the primary concern in this regard. There are several major regulatory documents listed below (taken from Associates of Cape Cod Inc. Web site: *http://www.acciusa.com/*) that describe how drugs, devices, dialysate, water, and other substances are to be tested for endotoxin.

- Guideline on Validation of the Limulus Amebocyte Lysate Test as an End-Product
- Endotoxin Test for Human and Animal Parenteral Drugs, Biological Products, and Medical Devices, FDA, December, 1987.
- Interim Guidance for Human and Veterinary Drug Products and Biologicals, Kinetic LAL Techniques, FDA, July 1991.
- Bacterial Endotoxins Test, United States Pharmacopoeia
- Transfusion and Infusion Assemblies and Similar Medical Devices, United States Pharmacopoeia (USP)
- Bacterial Endotoxins Test, European Pharmacopoeia
- Bacterial Endotoxins Test, Japanese Pharmacopoeia XIV

The USP now recognizes two tests—the pyrogen test conducted with rabbits and the bacterial endotoxins test, also termed the limulus amebocyte lysate

(LAL) test. Additionally the agency has approved the use of the bacterial endotoxins test for many drug and device products.

The endotoxin limits (measured by endotoxin units, or EU) apply for drugs and devices administered to humans or animals are listed below:

- Limit for parenteral drugs: 5.0 EU/kg. Assume a 70 kg human adult, unless administering drugs to children.
- Limit for drugs being injected intrathecally: 0.2 EU/kg.

Vehicle selection is an important consideration in all animal investigations. The vehicles should offer optimal absorption and exposure of test article. Consideration should be given for time of absorption before exposure but should not influence the results obtained for the test article under investigation, and as such they should ideally be biologically inert, have no effect on the biophysical properties of the compound, and have no toxic effects on the animals. If a component of the vehicle has biological effects, the dose should be limited such that these effects are minimized or not produced. Simple vehicles used to administer test articles include aqueous isotonic solutions, buffered solutions, co-solvent systems, suspensions, and oils. When administering suspensions, the viscosity, pH, and osmolality of the material need to be considered. The use of co-solvent systems needs careful attention because the vehicles themselves have dose-limiting toxicity. Laboratories are encouraged to develop a strategy to facilitate selection of the most appropriate vehicle based on the animal study being performed and the properties of the substance under investigation. Table 3 lists some of the common vehicles used for test article administration that are generally considered to be nontoxic and nonirritating.

## 5 DOSE SELECTION AND ADMINISTRATION

Body weights and surface area are considered in dose selection (see Table 4). A variety of physiological parameters are also considered in dose considerations. (see Table 5 through 8)

Toxicology studies are designed to establish a no observable adverse effect level (NOAEL) and a maximum tolerated dose (MTD) or a maximum feasible dose (MFD). The later is determined in part by the dosage form and maximum volumes that can be administered in an animal via the intended route of administration. (see Tables 9 through 11)

### 5.1 Determination of No Observed Adverse Effect Level (NOAEL)

The common definition of NOAEL is “the highest experimental point that is without adverse effect” or “the highest dose used that produces no evidence of adverse effects using the most susceptible but appropriate species and the

**TABLE 3 Common vehicles for test article administration**

| | |
|---|---|
| **Oral** | Water |
| | Methyl cellulose (0.5–5% aqueous suspension) |
| | Carboxymethyl cellulose |
| | Hydroxypropylmethyl cellulose |
| **Dermal** | Physiological saline |
| | Water |
| | Ethanol |
| | Acetone |
| | Mineral oil |
| **Parenteral** | Physiological saline (sterile) |
| | Water (sterile) |

**TABLE 4 Equivalent surface area dosage conversion factors**

| | Mouse (20 g) | Rat (150 g) | Monkey (3 kg) | Dog (8 kg) | Minipig (15 kg) | Human (60 kg) |
|---|---|---|---|---|---|---|
| Mouse | 1 | 1/2 | 1/4 | 1/6 | 1/11 | 1/12 |
| Rat | 2 | 1 | 1/2 | 1/3 | 1/6.4 | 1/6 |
| Monkey | 4 | 2 | 1 | 3/5 | 1/2.7 | 1/3 |
| Dog | 6 | 4 | 3/2 | 1 | 1/1.8 | 1/2 |
| Minipig | 11 | 6.4 | 2.7 | 1.8 | 1 | 0.9 |
| Human | 12 | 7 | 3 | 2 | 1/1.1 | 1 |

Example: 50 mg/kg in the mouse to equivalent in monkeys assuming equal mg/m$^2$: 50 mg/kg * 1/4 = 13 mg/kg.

**TABLE 5 Comparative physiological reference values**

| Species | Average Life Span (years) | Body Weight (kg) | Food Consumption Factor (g/day) | Food Consumption Factor[a] | Water Consumption (ml/day) | Inhalation Rate (m$^3$/day) |
|---|---|---|---|---|---|---|
| Human | 70 | 70 | 2000 | 0.028 | 1400 | 20 |
| Mouse | 1.5–2 | 0.03 | 4 | 0.13 | 6 | 0.052 |
| Rat | 2 | 0.35 | 18 | 0.05 | 50 | 0.29 |
| Hamster | 2.4 | 0.14 | 12 | 0.083 | 27 | 0.13 |
| Guinea pig | 4.5 | 0.84 | 34 | 0.040 | 200 | 0.40 |
| Rabbit | 7.8 | 3.8 | 186 | 0.049 | 410 | 2 |
| Cat | 17 | 3 | 90 | 0.030 | 220 | 1.2 |
| Dog | 12 | 12.7 | 318 | 0.025 | 610 | 4.3 |
| Monkey (rhesus) | 18 | 8 | 320 | 0.040 | 530 | 5.4 |

*Source*: Modified from US EPA, Development of Statistical Distributions or Ranges of Standard Factors Used in Exposure Assessments, Office of Health and Environmental Assessments, EPA No. 600/8-85/010, NTIS, PB85-242667, 1985.

[a]Fraction of body weight consumed per day as food.

**TABLE 6 Comparative physiological parameters**

| | Mouse (0.02 kg) | Rat (0.25 kg) | Rabbit (2.5 kg) | Monkey (5 kg) | Dog (10 kg) | Human (70 kg) |
|---|---|---|---|---|---|---|
| Surface area (m2) | 0.008 | 0.023 | 0.17 | 0.32 | 0.51 | 1.85 |
| Mean life span potential (yr) | 2.7 | 4.7 | 8.0 | 22 | 20 | 93 |
| Total plasma protein (g/100 ml) | 6.2 | 6.7 | 5.7 | 8.8 | 9.0 | 7.4 |
| Plasma albumin (g/100 ml) | 3.3 | 3.2 | 3.9 | 4.9 | 2.6 | 4.2 |
| Hematocrit (%) | 45 | 46 | 36 | 41 | 42 | 44 |
| Total ventilation (L/min) | 0.03 | 0.12 | 0.80 | 1.7 | 1.50 | 8.0 |
| Respiratory rate (L/min) | 163 | 85 | 51 | 38 | 2 | 12 |
| Heart rate (beats/min) | 624 | 362 | 213 | 192 | 96 | 65 |
| Oxygen consumption (ml/hr/g body wt) | 1.6 | 0.85 | 0.5 | 0.4 | 0.34 | 0.20 |

*Source*: Adapted from B. Davies and T. Morris (1993). Physiological parameters in laboratory animals and humans. *Pharmaceutical Research* 10(7):1093–1095.

**TABLE 7 Comparative volumes of various organs and body fluids**

| | Mouse (0.02 kg) | Rat (0.25 kg) | Rabbit (2.5 kg) | Monkey (5 kg) | Dog (10 kg) | Human (70 kg) |
|---|---|---|---|---|---|---|
| Organ volumes (ml) | | | | | | |
| Brain | — | 1.2 | — | — | 72 | 1,450 |
| Liver | 1.3 | 19.6 | 100 | 135 | 500 | 1,700 |
| Kidneys | 0.35 | 3.7 | 15 | 30 | 60 | 300 |
| Heart | 0.095 | 1.2 | 6 | 17 | 120 | 300 |
| Spleen | 0.1 | 1.3 | 1 | — | 36 | 200 |
| Lungs | 0.1 | 2.1 | 17 | — | 120 | 1,200 |
| Gut | 1.5 | 11.3 | 120 | 230 | 500 | 1,700 |
| Muscle | 10.0 | 245 | 1,400 | 2,500 | 5,500 | 35,000 |
| Adipose | — | 10.0 | 120 | — | — | 10,000 |
| Skin | 3.0 | 40.0 | 110 | 500 | — | 7,800 |
| Blood | 1.7 | 13.5 | 165 | 375 | 900 | 5,200 |
| Total body water (ml) | 15.0 | 167 | 1,800 | 3,500 | 6,000 | 42,000 |
| Intracellular fluid (ml) | — | 93.0 | 1,200 | 2,400 | 3,300 | 24,000 |
| Extracellular fluid (ml) | — | 75.0 | 625 | 1,000 | 2,800 | 18,000 |
| Plasma volume (ml) | 1.0 | 7.8 | 110 | 225 | 500 | 3,000 |

*Source*: Adapted from B. Davies and T. Morris (1993). Physiological parameters in laboratory animals and humans. *Pharmaceutical Research* 10(7):1093–1095.

**TABLE 8 Comparative cardiovascular characteristics**

| | Energy Metabolism | | Cardiac Function | | | | | Arterial Blood Pressure (mmHg) | |
|---|---|---|---|---|---|---|---|---|---|
| Species | cal/kg/day | Cal/m$^2$/day | Heart Wt. (g/100 g) | Heart Rate (beats/min) | Stroke Vol (ml/beat) | Cardiac Output (L/min) | Cardiac Index (L/m$^2$/min) | Systolic | Diastolic |
| Rat | 120–140 | 760–905 | 0.24–0.58 | 250–400 | 1.3–2.0 | 0.015–0.079 | 1.6 | 88–184 | 58–145 |
| Rabbit | 47 | 810 | 0.19–0.36 | 123–330 | 1.3–3.8 | 0.25–0.75 | 1.7 | 95–130 | 60–90 |
| Monkey | 49 | 675 | 0.34–0.39 | 165–240 | 8.8 | 1.06 | — | 137–188 | 112–152 |
| Dog | 34–39 | 770–800 | 0.65–0.96 | 72–130 | 14–22 | 0.65–1.57 | 2.9 | 95–136 | 43–66 |
| Human | 23–26 | 790–910 | 0.45–0.65 | 41–108 | 62.8 | 5.6 | 3.3 | 92–150 | 53–90 |
| Pig | 14–17 | 1100–1360 | 0.25–0.40 | 55–86 | 39–43 | 5.4 | 4.8 | 144–185 | 98–120 |

*Source*: Adapted from M. A. Hollinger and M. J. Derelanko (2002). *Handbook of Toxicology*, 2nd ed. New York: CRC Press LLC, pp. 1275.

**TABLE 9 Suggested ranges of needle sizes for test article administration**

| | Gavage | IV | | IP | | SC | | IM | |
|---|---|---|---|---|---|---|---|---|---|
| Species | Recommended | Ideal | Range | Ideal | Range | Ideal | Range | Ideal | Range |
| Mouse | Premature infant feeding tube cut to 70 mm, marked at 38 mm. | 25 or 27 | 25–30 | 25 or 27 | 22–30 | 25 or 27 | 22–30 | 25 or 27 | 22–30 |
| Rat | 3-inch ball-tipped intubation needle | 25 | 25–30 | 25 | 22–30 | 25 | 25–30 | 25 | 22–30 |
| Rabbit | Number 18 French catheter, cut to 15 inches, marked at 12 inches | 21 | 21–22 | 21 | 18–23 | 25 | 22–25 | 25 | 22–30 |
| Dog | Kaslow stomach tube 12Fr>- 24 inches: Davol 32Fr>-intubation tube | 21 | 21–22 | — | — | 22 | 20–23 | 21 or 25 | 20–25 |
| Monkey | Number 8 French tube (nasogastric gavage) | 25 | 25–30 | — | — | 22 | 22–25 | 25 | 22–25 |

*Source*: Adapted from C. S. Auletta (2002), *Handbook of Toxicology*, 2nd ed. New York: CRC Press LLC, pp. 98.

**TABLE 10 Recommended test article administration volumes (considered maximal)**

| | Route and Volumes (ml/kg) | | | | | |
|---|---|---|---|---|---|---|
| Species | Oral | SC[a] | IP | IM[b] | IV (bolus) | IV (slow) |
| Mouse | 10 (50) | 10 (40) | 20 (80) | 0.05[c] (0.1) | 5 | (25) |
| Rat | 10 (40) | 5 (10) | 10 (20) | 0.1[c] (0.2) | 5 | (20) |
| Rabbit | 10 (15) | 1 (2) | 5 (20) | 0.25 (0.5) | 2 | (10) |
| Dog | 5 (15) | 1 (2) | 1 (20) | 0.25 (0.5) | 2.5 | (5) |
| Macaque | 5 (15) | 2 (5) | (10) | 0.25 (0.5) | 15 | (4) |
| Marmoset | 10 (15) | 2 (5) | (10) | 0.25 (0.5) | 2.5 | (10) |
| Minipig | 10 (15) | 1 (2) | 1 (20) | 0.25 (0.5) | 2.5 | (5) |

*Source*: Adapted from K. Diehl, R. Hull, D. Morton, R. Pfister, Y. Rabemampianina, D. Smith, J. M. Vidal, and C. van de Vorstenbosch (2001). A good practice guide to the administration of substances and removal of blood, including routes and volumes. *Journal of Applied Toxicology* 21: 15–23.

[a]Maximum of 2–3 SC sites per day.
[b]Maximum of two IM sites per day.
[c]Values in ml/site.

**TABLE 11 Repeated intravenous infusion: dose volumes/rates (and possible maximal volumes/rates)**

| Daily infusion period (hours) | Mouse | Rat | Rabbit | Dog | Macaque | Minipig |
|---|---|---|---|---|---|---|
| | | | *Total daily volume (ml/kg)* | | | |
| 4 | 20 | 20 | [a] | 20 | 15 | [a] |
| 24 | 96 (192) | 60 (96) | 24 (72) | 24 (96) | 60 | 24 |
| | | | *Rate (ml/kg * h)* | | | |
| 4 | [a] | 5 | [a] | 5 | 10 | [a] |
| 24 | 4 (8) | 2.5 (4) | 1 (3) | 1 (4) | 4 | 1 |

*Source*: Adapted from K. Diehl, R. Hull, D. Morton, R. Pfister, Y. Rabemampianina, D. Smith, J. M. Vidal, and C. van de Vorstenbosch (2001). A good practice guide to the administration of substances and removal of blood, including routes and volumes. *Journal of Applied Toxicology* 21: 15–23.

*Note*: In some cases two sets of values are shown. Those in parentheses are the possible maximal values.
[a]Wide variability of values, nonstandardized.

most sensitive indicator." The NOAEL must, by definition, be one of the experimental doses tested.

The NOAEL is used by regulatory agencies to establish acceptable doses. Safety factors are often used to adjust from the NOAEL to a dose that is

deemed safe for initial human trials, typically 10 to 100-fold. While there are some inherent weaknesses in the concept of the NOAEL (minimizing the dose–response, in general, in favor of a single point and the probabilistic concept that exposures that test fewer animals tend to produce larger NOAELs), it is still widely used in toxicological evaluations. Dorato et al. [2] discuss the working definitions of the NOAEL, factors to be taken into account when determining the NOAEL, as well as the shortcomings inherent in the methods employed for determining the NOAEL.

When planning nonclinical studies that are focused on establishing a NOAEL to support the next phase of clinical development, a few major considerations should be kept in mind:

1. Number of animals per dose group. The larger the $n$, the higher is the probability of identifying toxicity.
2. Number and range of doses administered. The more dose groups and the wider the range of doses, the more likely toxicity will be observed.
3. Endpoints. The more numerous the evaluations (organ weights, tissue histology, clinical pathology, etc.) the more likely toxicity will be observed.

### 5.2 Determination of the Maximum Tolerated Dose (MTD)

The MTD is defined generally as the highest dose of the test agent that is predicted not to alter the animals' longevity or growth because of noncancer effects. The MTD thus varies inversely with the toxicity of a chemical. The major endpoints factoring into the determination of a MTD are as follows:

1. *Mortality*. Generally a mortality rate of 10% to 15% of a dose group due to test article administration indicates that a MTD has been achieved at that dose.
2. *Body weight changes*. Generally, a 10% to 15% decrease in percent body weight gain (not absolute body weight) compared to control groups indicates that a MTD has been achieved at that dose.
3. *Histology*. The identification of target organs with effects clearly attributable to test article indicates that a MTD has been achieved at that dose.

The three parameters above are not the sole set of criteria for determining a MTD, but rather, they are the more significant indicators of toxicity to guide the investigator in characterizing the toxicity of a test article and implementing subsequent studies. The MTD thus determined applies only to that particular study type for the duration of the in-life portion. When planning a preclinical study with preexisting relevant data from a previous study, it is generally applicable to use the previously established MTD to set doses for the subsequent study by selecting a high dose lower than that of the MTD, and then decreasing doses

from there to round out mid and lower doses. This will maximize the probability of dose groups surviving the duration of the study in adequate numbers.

## 6 PHARMACOKINETICS (PK) AND TOXICOKINETICS (TK)

The S6 guidelines mention the difficulty of standard guidances for PK and TK studies for biologics. However, single- and repeat-dose PK studies along with tissue distribution studies can be useful. It should be noted as in the discussion concerning immunogenicity that the formation of neutralizing antibodies against the biologic can affect exposure by clearing the drug, thus significantly reducing or eliminating exposure as well as quenching the pharmacodynamic responses to drug action (e.g., ablation of neopterin response after neutralizing antibody formation against human interferon beta treatment in rhesus monkeys). Some cases, as is the case with interferons exposure, are best assessed through the measurement of pharmacodynamic biomarkers of activity.

Biologics are mostly administered parenterally: subcutaneously (SC), intramuscularly (IM), and intravenously (IV). Intravenous administration is the most pharmocokinetically reliable mode of administration, but IM or SC is preferred for biopharmaceuticals that will be administered chronically. Absorption of biopharmaceutical products is influenced by molecular weight but also by non-compound dependent factors like co-administration of albumin, depth of injection, blood flow at injection site, exercise, rubbing, and temperature [3].

The site of administration also influences their pharmacokinetic behavior. Following SC or IM administration, the bioavailability of biopharmaceuticals is often lower than with small molecules due to proteolysis during interstitial and lymphatic transit. Oral delivery of biopharmaceuticals is limited by barriers such as enzymatic and pH-dependent degradation in the gastrointestinal tract, low epithelial permeability, and instability under formulation conditions. Therefore biopharmaceuticals have an extremely low (<1%) and erratic bioavailability after oral administration [3].

Immunogenicity assays for investigating the frequency and consequences of antibody development against a protein therapeutic agent are typically based on an immunoassay technique (mostly ELISAs of various types). However, other assay formats are available such as radioimmunoprecipitation assay, surface plasmon resonance, and electrochemiluminescence [3]. Assays for measuring antibody response should be established in the early preclinical stage of development to estimate the value of the applied animal models (see Chapters General Toxicity Testing and Immunotoxicity Testing for Biopharmaceuticals and Immunogenicity of Therapeutic Proteins and the Assessment of Risk).

Because of the catabolism of proteins to (mostly) endogenous amino acids, classical biotransformation studies as performed for small molecules are not needed. Additionally, limitations of current analytical methods to detect and distinguish metabolites and the putative lack of pharmacological or toxicological activity of the metabolites, remain obstacles. Similarly mass balance studies

usually used to determine the excretion pathways of small molecules (and their metabolites) are not useful for biologics [3].

For radiolabeled proteins, placement of the radiolabel is critical to the success of the study in regards to tissue distribution and other parameters related to "absorption-distribution-metabolism-excretion" or ADME characterization.

## 7 BLOOD SAMPLING AND MAXIMUM BLOOD VOLUME

An important parameter to consider in the design of toxicology studies is the amount of blood that can be obtained for use, not only in pharmacokinetic/toxicokinetic assessments but for the monitoring of routine clinical pathology parameters (hematology, coagulation, clinical chemistry) as well as other pharmacodynamic markers and other endpoints that may be included specific to the product class. Tables 12 through 14 provide guidance with respect to the method of blood draw as well as maximal volumes and recoveries for various species.

**TABLE 12 Advantages and disadvantages of various methods of blood sampling**

| Route/Vein | General Anaesthesia | Tissue Damage[a] | Repeat Bleeds | Volume | Species |
|---|---|---|---|---|---|
| Jugular | No | Low | Yes | +++ | Rat, dog, rabbit |
| Cephalic | No | Low | Yes | +++ | Macaque, dog |
| Saphenous/lateral tarsal | No | Low | Yes | ++(+) | Mouse/rat, marmoset/macaque, dog |
| Marginal ear | No (local) | Low | Yes | +++ | Rabbit, minipig |
| Femoral | No | Low | Yes | +++ | Marmoset/macaque |
| Sublingual | Yes | Low | Yes | +++ | Rat |
| Lateral tail | No | Low | Yes | ++(+)+ | Rat, mouse/marmoset |
| Central ear artery | No (local) | Low | Yes | +++ | Rabbit |
| Cranial vena cava | No | Low | Yes | +++ | Minipig |
| Tail tip amputation (<1–3 mm) | Yes | Mod | Limited | + | Mouse/rat |
| Retrobulbar plexus | Yes | Moderate/high | Yes | +++ | Mouse/rat |
| Cardiac[b] | Yes | Moderate | No | +++ | Mouse/rat/rabbit |

*Source*: Adapted from K. Diehl, R. Hull, D. Morton, R. Pfister, Y. Rabemampianina, D. Smith, J. M. Vidal, and C. van de Vorstenbosch (2001). A good practice guide to the administration of substances and removal of blood, including routes and volumes. *Journal of Applied Toxicology* 21: 15–23.

[a]The potential for tissue damage is based on the likely incidence of it occurring and the severity of any sequelae, (e.g., inflammatory reaction or histological damage).

[b]Only carried out as a terminal procedure under general anaesthesia.

**TABLE 13 Comparative total blood volumes and recommended maximum blood sample volumes per body weight**

| Species (weight) | Blood Volume (ml) | 7.5% (ml) | 10% (ml) | 15% (ml) | 20% (ml) |
|---|---|---|---|---|---|
| Mouse (25 g) | 1.8 | 0.1 | 0.2 | 0.3 | 0.4 |
| Rat (250 g) | 16 | 1.2 | 1.6 | 2.4 | 3.2 |
| Rabbit (4 kg) | 224 | 17 | 22 | 34 | 45 |
| Dog (10 kg) | 850 | 64 | 85 | 127 | 170 |
| Macaque (rhesus) (4 kg) | 280 | 21 | 28 | 42 | 56 |
| Macaque (cynomolgus) (5 kg) | 325 | 24 | 32 | 49 | 65 |
| Minipig (15 kg) | 975 | 73 | 98 | 146 | 195 |

*Source*: Adapted from K. Diehl, R. Hull, D. Morton, R. Pfister, Y. Rabemampianina, D. Smith, J. M. Vidal, and C. van de Vorstenbosch (2001). A good practice guide to the administration of substances and removal of blood, including routes and volumes. *Journal of Applied Toxicology* 21: 15–23.

**TABLE 14 Limit volumes for blood sampling and recovery periods**

| Single Sampling | | Multiple sampling | |
|---|---|---|---|
| Circulatory Blood Volume Removed (%) | Approximate Recovery Period | Circulatory Blood Volume Removed in 24 h (%) | Approximate Recovery Period |
| 7.5 | 1 week | 7.5 | 1 week |
| 10 | 2 weeks | 10–15 | 2 weeks |
| 15 | 4 weeks | 20 | 3 weeks |

*Source*: Adapted from K. Diehl, R. Hull, D. Morton, R. Pfister, Y. Rabemampianina, D. Smith, J. M. Vidal, and C. van de Vorstenbosch (2001). A good practice guide to the administration of substances and removal of blood, including routes and volumes. *Journal of Applied Toxicology* 21: 15–23.

## 8 ASSESSMENT OF TOXICITY

In addition to measuring various "in-life" parameters, postmortem assessments provide useful information with respect to understanding potential toxicities as well as insight as to potential mechanisms of toxicity. A summary of comparable weights of various target organs is provided in Table 15. This table is also useful when considering dose for specific targeted deliveries (e.g., via intercranial, intercardiac, or inhalation routes). An understanding of both incidence and severity of a histopathological finding is important to the understanding of toxicity (see Tables 16 and 17). For example a low incidence of a finding graded as severe could be predictive of a response in a subset of patients in a clinical population. Likewise a dose-related increase in a finding of lower severity may be a predictor of a potential adverse clinical response. The recommendations of a safe starting dose and dose escalation scheme and

**TABLE 15 Comparative weights of various organs**

| | Mouse (0.025 kg) | Rat (0.25 kg) | Rabbit (2.5 kg) | Rhesus monkey (4 kg) | Dog (10 kg) | Human (70 kg) |
|---|---|---|---|---|---|---|
| Organ weights (g) | | | | | | |
| Brain | 0.4 | 2.0 | 14 | 90 | 80 | 1400 |
| Liver | 1.9 | 10.0 | 80 | 150 | 320 | 1800 |
| Kidneys | 0.3 | 2.0 | 15 | 25 | 50 | 300 |
| Heart | 0.08 | 1.0 | 5 | 19.0 | 80 | 340 |
| Spleen | 0.1 | 0.9 | 1 | 8 | 25 | 180 |
| Adrenals | 0.004 | 0.05 | 0.5 | 1.4 | 1 | 14 |
| Lung | 0.14 | 1.5 | 20 | 33 | 100 | 1000 |

**TABLE 16 Common abbreviations and codes used in histopathology assessments**

| Code | Finding or Observation |
|---|---|
| + (1) | Minimal grade lesion |
| ++(2) | Mild or slight grade lesion |
| +++(3) | Moderate grade lesion |
| ++++(4) | Marked or severe grade lesion |
| +++++(5) | Very severe or massive grade lesion |
| (No entry) | Lesion not present or organ/tissue not examined |
| + | Tissue examined microscopically |
| – | Organ/tissue present, no lesion in section |
| A | Autolysis precludes examination |
| B | Primary benign tumor |
| I | Incomplete section of organ/tissue or insufficient tissue for evaluation |
| M | Primary malignant tumor |
| M | Organ/tissue missing, not present in section |
| N | No section of organ/tissue |
| N | Normal, organ/tissue within normal limits |
| NCL | No corresponding lesion for gross finding |
| NE | Organ/tissue not examined |
| NSL | No significant lesion, organ/tissue within normal limits |
| P | Lesion present, not graded (e.g., cyst, anomaly) |
| U | Unremarkable organ/tissue, within normal limits |
| WNL | Organ/tissue within normal limits |
| X | Not remarkable organ/tissue, normal |
| X | Incidence of listed morphology, lesion present |

*Source*: Adapted from J. C. Peckham (2002). *Handbook of Toxicology*, 2nd ed. New York: CRC Press LLC, pp. 655.

**TABLE 17 Frequently used grading schemes (severity scores) for histopathology findings**

| Severity Degree | Proportion of Organ Affected | (A) Grade | Percentage of Organ Affected | (B) Grade | Percentage of Organ Affected | (C) Grade | Percentage of Organ Affected | (D) Grade | Percentage of Organ Affected | (E) Grade | Quantifiable Finding |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Minimal | Very small amount | 1 | (A1) < 1–25% (A2) < 1–15% | 1 | <1% | | | 1 | <1% | | |
| Slight | Very small to small amount | | | 2 | 1–25% | 1 | 1–25% | | | 1 | 1–4 foci |
| Mild | Small amount | 2 | (A1) 26–50% (A2) 16–35% | | | 2 | 26–50% | 2 | 1–30% | 2 | 5–8 foci |
| Moderate | Middle or median amount | 3 | (A1) 51–75% | 3 | 26–50% | 3 | 51–76% | 3 | 31–60% | 3 | 9–12 foci |
| Marked | Large amount | 4 | (A1) 76–100% (A2) 61–100% | | | | | | | 4 | >12 foci |
| Moderately severe | Large amount | | | 4 | 51–75% | | | | | | |
| Severe | Very large amount | | | 5 | 76–100% | 4 | 76–100% | 4 | 61–90% | | |
| Very severe or massive | Very large amount | | | | | | | 5 | 91–100% | | |

*Source*: Adapted from J. C. Peckham (2002). *Handbook of Toxicology*, 2nd ed. New York: CRC Press LLC, pp. 656.

clinical monitoring parameters are based on the totality of the toxicology and exposure data.

## ACKNOWLEDGMENTS

The author thanks Joshua Bender of the Pittsburgh Technology Council for his assistance in compiling the tables presented in this chapter.

## REFERENCES

International Symposium on Regulatory Testing and Animal Welfare. Recommendations on best scientific practices for safety evaluation using nonrodent species. *ILAR J* 2002;43(Suppl).

Dorato MA, Engelhardt JA. The no observed adverse effect level in drug safety evaluations: use, issues, and definitions. *J Appl Toxicol* 2005;21:15–23.

Baumann A. Early development of therapeutic biologics—pharmacokinetics. *Curr Drug Metab* 2006;7:15–21.

# Survey of Preclinical Toxicology Programs for Approved Biopharmaceuticals

ANITA MARIE O'CONNOR, PhD

**Contents**

1 Introduction
2 Tissue Cross-Reactivity
3 Single-Dose Toxicity
4 Repeat-Dose Toxicity
5 Genetic Toxicity
6 Carcinogenicity
7 Reproductive/Developmental Toxicity
8 Immunotoxicity
9 Conclusions
References

## 1 INTRODUCTION

This chapter provides a survey of US marketed biopharmaceuticals approved over the past 20 years. It is a simply a profile of the species and types of toxicology studies done for each drug for marketing approval. For a complete list of all preclinical studies for a biopharmaceutical, the reader is directed to the Web sites of the Food and Drug Administration (FDA) and the European Medicines Agency (EMEA).

The information found in Table 1 was obtained from a variety of sources: personal communications with the toxicologists who designed the programs, the FDA Web site, the EMEA Web site, and the package insert for the product

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

**TABLE 1 Toxicology studies to support biopharmaceutical approvals**

| Biopharmaceutical | Description ↔ Target | Indication | Toxicology Studies[a] |
|---|---|---|---|
| Abraxane™ (paclitaxel formulated in human serum albumin in a 1 : 9 ratio) (US) Approval: 2005 | Albumin bound form of paclitaxel ↔ causes apoptosis in cancer cells | Metastatic breast cancer | Single dose: rats, dogs, swine[b]<br>Repeat dose: athymic nude mice implanted with MX-1 tumors; other studies were complicated by immune reactions to human serum albumin<br>Fertility and early embryonic development: rats<br>Developmental: rats<br>Prenatal and postnatal development: none<br>Genetic toxicology[c]: Ames test, human lymphocyte chromosomal aberration assay, CHO/HGPRT gene mutation assay, mouse micronucleus assay<br>Carcinogenicity: none |
| Actimune® (interferon gamma-1b) (US) Approval: 1999 | Human recombinant interferon gamma-1b ↔ multiple targets within the lymphokine regulatory network | Infections associated with chronic granulomatous disease; malignant osteopetrosis | Fertility and early embryonic development: cynomolgus monkeys (male and female), mice (surrogate, male and female)<br>Developmental: primate[d]<br>Prenatal and postnatal development: primate<br>Genetic toxicology: Ames test, micronucleus assay in mice<br>Carcinogenicity: none |
| Activase® (alteplase) (US) Approval: 1996 | Recombinant human tissue plasminogen activator ↔ plasminogen with fibrin present | Restoration of function of central venous access devices as assessed by the ability to draw blood | Developmental: rats, rabbits<br>Genetic toxicology: Ames test, human lymphocyte chromosomal aberration assay<br>Carcinogenicity: short-term experiments in rodents (i.e., not standard 2-year study) |

| | | | |
|---|---|---|---|
| Adlurazyme® (laronidase) (US) Approval: 2003 | Laronidase ↔ enzyme replacement therapy | Mucopolysaccharidosis-1 | Single dose: rats, dogs<br>Repeat dose: dogs, cynomolgus monkeys<br>Fertility and early embryonic development: rats<br>Developmental: rats<br>Prenatal and postnatal development: none<br>Genetic toxicology: none<br>Carcinogenicity: none |
| Alferon (interferon alpha-N3) (US) Approval: 1989 | Interferon alpha-N3 derived from leukocytes ↔ multiple targets within the lymphokine regulatory network | Refractory or recurring external condylomata acuminata | Fertility and early embryonic development: none<br>Developmental: none<br>Prenatal and postnatal development: none<br>Genetic toxicology: none<br>Carcinogenicity: none |
| Amevive® (Alefacept) (US) Approval: 2003 | Dimeric glycosylated fusion protein ↔ binds to CD2 on T lymphocytes | Chronic plaque psoriasis | Single dose: rats (not relevant), baboons<br>Repeat dose: baboons, cynomolgus monkeys<br>Fertility and early embryonic development: none<br>Developmental: cynomolgus monkeys<br>Prenatal and postnatal development: cynomolgus monkeys<br>Genetic toxicology: in vitro and in vivo<br>Carcinogenicity: evaluated in a chronic cynomolgus monkey study<br>Other: flow cytometry and immunhistochemical analysis of lymphoid tissues for peripheral lymphocyte subsets; humoral immunity |

**TABLE 1** ***Continued***

| Biopharmaceutical | Description ↔ Target | Indication | Toxicology Studies[a] |
|---|---|---|---|
| Apidra®<br>(insulin glulisine)<br>(US) Approval: 2004 | Recombinant human insulin analogue (rapid-acting insulin) ↔ glucose metabolism | Diabetes | Single dose: mice, rats, dogs<br>Repeat dose: rats, dogs<br>Fertility and early embryonic development: rat (male and female)<br>Developmental: rats, rabbits<br>Prenatal and postnatal development: none<br>Genetic toxicology: Ames test, in vitro mammalian chromosome aberration test in V79 Chinese hamster cells, and in vivo mammalian erythrocyte micronucleus test in rats<br>Carcinogenicity: used the 12-month repeat dose study instead of a standard 2-year bioassay<br>Other: cardiovascular study in dogs; drug interaction study with NPH insulin; immunogenicity in rabbits |
| Apligraf®<br>(living human skin substitute)<br>[US(CDRH)]<br>Approval: 2000 | Living, bi-layered skin substitute containing type 1 bovine collagen Extracted and purified from bovine tendons and viable allogeneic human fibroblast and keratinocytes cells isolated from human infant foreskin | Leg and diabetic foot ulcers | In vitro immunology, and Athymic nude mice proof of concept studies |

| | | | |
|---|---|---|---|
| Aranesp™ (darbepoetin alpha) (US) Approval: 2001 | Hyperglycosylated recombinant human erythropoietin ↔ stimulation of erythropoiesis | Anemia associated with chronic renal failure; anemia due to nonmyeloid malignancies in patients treated with chemotherapy | Single dose: rats, dogs<br>Repeat dose: rats, dogs<br>Fertility and early embryonic development: rats<br>Developmental: rats, rabbits<br>Prenatal and postnatal development: rats<br>Genetic toxicology: in vitro bacterial and CHO cell assays, mouse micronucleus assay<br>Carcinogenicity: proliferation in vitro and in vivo<br>Tissue cross-reactivity: human tissues<br>Other: safety studies in a variety of species |
| Avastin™ (bevacizumab) (US) Approval: 2004 | Recombinant human monoclonal antibody ↔ binds to vascular endothelial growth factor | Metastatic colorectal cancer in combination with 5-fluorouracil based chemotherapy | Single dose: none<br>Repeat dose: cynomolgus monkeys<br>Fertility and early embryonic development: rabbits<br>Developmental: rabbits<br>Prenatal and postnatal development: none<br>Genetic toxicology: none<br>Carcinogenicity: none<br>Tissue cross-reactivity: rabbit, cynomolgus tissues<br>Other: wound healing, hemolytic potential, blood compatability, Cisplatin renal injury induced in rabbits, rabbit thrombosis model |
| Avonex® (interferon beta-1a) (US) Approval: 1996 | Recombinant human interferon beta-1a ↔ binds to multiple receptors | Multiple sclerosis | Single dose: rhesus monkeys<br>Repeat dose: rhesus monkeys<br>Fertility and early embryonic development: none<br>Developmental: none<br>Prenatal and postnatal development: none<br>Genetic toxicology: Ames test<br>Carcinogenicity: none |

**TABLE 1** ***Continued***

| Biopharmaceutical | Description ↔ Target | Indication | Toxicology Studies[a] |
|---|---|---|---|
| Betaseron®<br>(interferon beta-1b)<br>(US) Approval: 2003 | Human recombinant interferon beta-1b ↔ multiple targets within the lymphokine regulatory network | Multiple sclerosis | Single dose: African green monkeys (subacute)<br>Repeat dose: none (due to antibody development)<br>Fertility and early embryonic development: rhesus monkeys<br>Developmental: rhesus monkeys<br>Prenatal and postnatal development: none<br>Genetic toxicology: Ames test, in vitro transformation assay, human peripheral blood lymphocyte assay<br>Carcinogenicity: none |
| Bexxar®<br>(tositumomab and $I^{131}$ tositumomab)<br>(US) Approval: 2003 | Tositumomab is a murine IgG lambda monoclonal antibody ↔ targets the CD20 antigen on normal and malignant B cells $I^{131}$ tositumomab is tositumomab covalently linked to $I^{131}$ | CD20 positive follicular non-Hodgkin's lymphoma | Fertility and early embryonic development: none<br>Developmental: none<br>Prenatal and postnatal development: none<br>Genetic toxicology: none<br>Carcinogenicity: none |

| | | | |
|---|---|---|---|
| Campath® (alemtuzumab) (US) Approval: 2001 | Recombinant humanized monoclonal antibody ↔ directed against CD52, which is expressed on the surface of normal and malignant B and T lymphocytes, NK cells, monocytes, macrophages, and tissues of the male reproductive system | B cell chronic lymphocytic leukemia | Single dose: cynomolgus monkeys<br>Repeat dose: cynomolgus monkeys<br>Fertility and early embryonic development: none<br>Developmental: none<br>Prenatal and postnatal development: none<br>Genetic toxicology: none<br>Carcinogenicity: none<br>Tissue cross-reactivity: human tissues<br>Other: cardiovascular and respiratory safety studies in anesthetized cynomolgus monkeys |
| Carticel® (autologous cultured chondrocytes) (US) Approval: 1997 | Autologous cultured chrondocytes | Repair of clinically significant, symptomatic, cartilaginous defects of the femoral condyle (medial, lateral or trochlear) caused by acute or repetitive trauma | Other: proof of concept/safety studies in dogs and rabbits |
| CEA-Scan® (arcitumomab; technetium-99 labeled) (US) Approval: 1996 | Acritumomab labeled with technetium-99 ↔ radiodiagnostic agent | Imaging agent for colorectal cancer | Fertility and early embryonic development: none<br>Developmental: none<br>Prenatal and postnatal development: none<br>Genetic toxicology: none<br>Carcinogenicity: none |

**TABLE 1** ***Continued***

| Biopharmaceutical | Description ↔ Target | Indication | Toxicology Studies[a] |
|---|---|---|---|
| Cerezyme® (imiglucerase) (US) Approval: 1994 | Recombinant beta-glucocerebrosidase ↔ enzyme replacement therapy | Type I Gaucher's disease | Single dose: rats<br>Repeat dose: rats, cynomolgus monkeys<br>Fertility and early embryonic development: none<br>Developmental: none<br>Prenatal and postnatal development: none<br>Genetic toxicology: Ames test<br>Carcinogenicity: none |
| Elaprase (idursulfase) (US) Approval: 2006 | Recombinant human idursulfase ↔ enzyme replacement therapy | Hunter syndrome (Mucopolysaccharidosis II) | Single dose: rats, cynomolgus monkeys<br>Repeat dose: cynomolgus monkeys<br>Fertility and early embryonic development: male rats<br>Developmental: none<br>Prenatal and postnatal development: none<br>Genetic toxicology: none<br>Carcinogenicity: none |
| Elitek® (rasburicase) (US) Approval: 2002 | Recombinant urate oxidase ↔ catalyzes enzymatic oxidation of uric acid | Malignancy-associated or chemotherapy-induced hyperuricemia | Fertility and early embryonic development: rats (male and female)<br>Developmental: none<br>Prenatal and postnatal development: none<br>Genetic toxicology: Ames test, unscheduled DNA synthesis, chromosome analysis, mouse lymphoma and micronucleus tests<br>Carcinogenicity: none |

| | | | |
|---|---|---|---|
| Elspar (asparaginase) (US) Approval: 1994 | Recombinant L-asparaginase ↔ depletion of aspargine: a metabolic defect of asparagine synthesis in malignant cells which are dependent on an exogenous source of asparagine | Acute lymphocytic leukemia in combination with chemotherapy | Single dose: rabbits<br>Repeat dose: dogs, monkeys<br>Fertility and early embryonic development: none<br>Developmental: mice, rats, rabbits<br>Prenatal and postnatal development: none<br>Genetic toxicology: Ames test<br>Carcinogenicity: none |
| Enbrel® (etanercept) (US) Approval: 1998 | Fusion protein (TNFR:Fc) ↔ competitive inhibitor of TNF | Rheumatoid arthritis; Psoriatic arthritis; Ankylosing spondylitis; Psoriasis | Single dose: none<br>Repeat dose: cynomolgus monkeys, rats, mice, rabbits<br>Fertility and early embryonic development: none<br>Developmental: rats, rabbit<br>Prenatal and postnatal development: rats (abbreviated)<br>Genetic toxicology: battery of in vitro and in vivo tests<br>Carcinogenicity: none |
| Epicel® (cultured epidermal autografts) (US) Approval: unclear if ever formally approved by FDA marketed since 1987 | Autologous keratinocytes are co-cultured with irradiated murine cells to form cultured epidermal autografts | Permanent skin replacement product for patients with life-threatening burns | This product was on the market in 1987, before a regulatory framework was put in place by FDA. |

**TABLE 1** ***Continued***

| Biopharmaceutical | Description ↔ Target | Indication | Toxicology Studies[a] |
|---|---|---|---|
| Epogen®/Procrit® (epoetin alfa) (US) Approval: 1993 | Recombinant human erythropoietin ↔ committed erythroid progenitors in bone marrow by replacing or increasing erythropoietin | Anemia of chronic renal failure patients, zidovudine-treated HIV-infected patients and cancer patients on chemotherapy; reduction of allogeneic blood transfusion in surgery patients | Fertility and early embryonic development: rats<br>Developmental: rats<br>Prenatal and postnatal development: rats<br>Genetic toxicology: Ames test, chromosomal aberrations in mammalian cells, micronuclei in mice, gene mutation at the HGPRT locus<br>Carcinogenicity: none |
| Erbitux™ (cetuximab) (US) Approval: 2004 | Chimeric monoclonal antibody ↔ inhibits the EGF receptor | Metastatic colorectal cancer (in combination with irinotecan based chemotherapy) | Single dose: mice (not relevant), rats (not relevant)<br>Repeat dose: rats, cynomolgus monkeys<br>Fertility and early embryonic development: none<br>Developmental: none<br>Prenatal and postnatal development: none<br>Genetic toxicology: Ames test, micronucleus test in rats (not relevant)<br>Carcinogenicity: none<br>Tissue Cross-Reactivity: monkey, goat, rat, rabbit, and human tissues, hepatic tissues of numerous species |

| | | | |
|---|---|---|---|
| Exubera® (insulin) (US) Approval: 2006 | Human insulin inhalation powder ↔ glucose metabolism | Type 1 or type 2 diabetes | Single dose: rats, monkeys (nonpivotal exploration studies)<br>Repeat dose: rats, cynomolgus monkeys<br>Fertility and early embryonic development: none<br>Developmental: none<br>Prenatal and postnatal development: none<br>Genetic toxicology: Ames test<br>Carcinogenicity: none<br>Other: cardiovascular study in dogs |
| Fabrazyme® (agalsidase beta) (US) Approval: 2003 | Recombinant human alpha galactosidase ↔ enzyme replacement therapy | Fabry's disease | Single dose: rats<br>Repeat dose: rats<br>Fertility and early embryonic development: none<br>Developmental: none<br>Prenatal and postnatal development: none<br>Genetic toxicology: none<br>Carcinogenicity: none<br>Other: cardiovascular function in dogs |
| Gem 21S® (Growth factor enhanced matrix) [US(CDRH)] Approval: 2005 | Tissue growth factor, recombinant human platelet-derived growth factor, and a synthetic bone matrix | Periodontal bone defects | Single dose: mice<br>Repeat dose: none<br>Fertility and early embryonic development: none<br>Developmental: none<br>Prenatal and postnatal development: none<br>Genetic toxicology: Ames test<br>Carcinogenicity: none<br>Other: dermal sensitization in guinea pigs, in vitro cytotoxicity, intracutaneous reactivity in rabbits, muscle implantation in rabbits |

**TABLE 1** ***Continued***

| Biopharmaceutical | Description ↔ Target | Indication | Toxicology Studies[a] |
|---|---|---|---|
| Genotropin (somatropin) (US) Approval: 1995 | Recombinant human growth hormone ↔ hormone replacement therapy | Growth hormone deficiency; growth failure in children and adults | Fertility and early embryonic development: rats<br>Developmental: rats, rabbits<br>Prenatal and postnatal development: rats<br>Genetic toxicology: Ames test, gene mutations in mammalian cells grown in vitro (mouse L5178Y cells), and chromosomal damage in intact animals (bone marrow cells in rats)<br>Carcinogenicity: none |
| Glucagen® (glucagon) (US) Approval: 1998 | Recombinant human glucagon ↔ hormone replacement therapy | Hypoglycemic reactions in diabetic patients and as a diagnostic aid to reduce organ motility during radiologic examination of the gastrointestinal tract | Single dose: rats, mice<br>Repeat dose: rats, dogs<br>Fertility and early embryonic development: none<br>Developmental: rats, rabbit<br>Prenatal and postnatal development: none<br>Genetic toxicology: Ames test, human lymphocyte, micronucleus assay in mice<br>Carcinogenicity: none<br>Other: cardiovascular and respiratory function in dogs |
| Glucagon (glucagon) (US) Approval: 1998 | Recombinant human glucagon ↔ hormone replacement therapy | Hypoglycemic reactions in diabetic patients and as a diagnostic aid to reduce organ motility during radiologic examination of the gastrointestinal tract | Fertility and early embryonic development: rats<br>Genetic toxicology: Ames test<br>Other: approval was based mainly on a bioequivalence study comparing the recombinant product to the sponsor's animal sourced product |

| | | | |
|---|---|---|---|
| Herceptin® (trastuzumab) (US) Approval: 1998 | Recombinant humanized monoclonal antibody ↔ HER2 | Breast cancer | Single dose: mice, rhesus monkeys<br>Repeat dose: rhesus monkeys, cynomolgus monkeys<br>Fertility and early embryonic development: cynomolgus monkeys<br>Developmental: cynomolgus monkeys<br>Prenatal and postnatal development: cynomolgus monkeys<br>Genetic toxicology: Ames test, chromosome aberration assay in human peripheral lymphocytes, in vivo mouse micronucleus<br>Carcinogenicity: none<br>Tissue cross-reactivity: human and monkey tissues |
| Humatrope (somatropin) (US) Approval: 1987 | Recombinant human growth hormone ↔ hormone replacement therapy | Growth hormone deficiency; growth failure in children and adults | Fertility and early embryonic development: none<br>Developmental: none<br>Prenatal and postnatal development: none<br>Genetic toxicology: unclear<br>Carcinogenicity: none |
| Humira™ (adalimumab) (US) Approval: 2002 | Antitumor necrosis factor human monoclonal antibody ↔ binds to TNF-α | Rheumatoid arthritis; psoriatic arthritis; ankylosing spondylitis | Single dose: mice, rats<br>Repeat dose: mice, cynomolgus monkeys<br>Fertility and early embryonic development: none<br>Developmental: cynomolgus monkeys<br>Prenatal and postnatal development: none<br>Genetic toxicology: Ames test, in vivo mouse micronucleus<br>Carcinogenicity: none<br>Tissue cross-reactivity: human tissues |

**TABLE 1** ***Continued***

| Biopharmaceutical | Description ↔ Target | Indication | Toxicology Studies[a] |
|---|---|---|---|
| Infergen® (Interferon alfacon-1) (US) Approval: 1997 | Recombinant nonnaturally occurring type-1 interferon ↔ multiple targets within the lymphokine regulatory network | Hepatitis C | Single dose: golden Syrian hamsters, rhesus monkeys<br>Repeat dose: golden Syrian hamsters, rhesus monkeys<br>Fertility and early embryonic development: golden Syrian hamsters<br>Developmental: golden Syrian hamsters, rhesus monkeys, cynomolgus monkeys<br>Prenatal and postnatal development: none<br>Genetic toxicology: Ames test, cytogenicity assay in human lymphocytes<br>Carcinogenicity: none |
| Intron A® (interferon alpha-2b) (US) Approval: 1983 | Recombinant interferon alpha 2b ↔ multiple targets within the lymphokine regulatory network | Hairy cell leukemia, malignant melanoma, follicular lymphoma, condylomata acuminata, AIDS-related Kaposi's sarcoma hepatitis C, hepatitis B | Single dose: mice, rats, rhesus monkeys<br>Repeat dose: mice, rats, rabbits, cynomolgus monkeys<br>Fertility and early embryonic development: rats, rabbits<br>Developmental: rats, rabbits, rhesus monkeys<br>Prenatal and postnatal development: rats, rabbits<br>Genetic toxicology: mutagenicity studies<br>Carcinogenicity: none |

| | | | |
|---|---|---|---|
| Kepivance® (Palifermin) (US) Approval: 2004 | Human recombinant keratinocytes growth factor ↔ KGF receptor | Oral mucositis in cancer patients | Single dose: rats, rhesus monkeys<br>Repeat dose: athymic nude mice, rats, rhesus monkeys, cynomolgus monkeys<br>Fertility and early embryonic development: rats<br>Developmental: rabbits, rats<br>Prenatal and postnatal development: none<br>Genetic toxicology: Ames test, in vitro mammalian chromosome aberration assay, in vivo mouse micronucleus assay<br>Carcinogenicity: Tg.rasH2 transgenic mouse model<br>Other: antigenicity in guinea pigs, mice, and rats; tumor cell lines (effects of the product on proliferation); human tumor xenografts in athymic nude mice; hemolytic potential studies |
| Kineret™ (anakinra) (US) Approval: 2001 | Human interleukin-1 receptor antagonist ↔ multiple targets within the lymphokine regulatory network | Rheumatoid arthritis | Single dose: rats, cynomolgus monkeys<br>Repeat dose: rats, rhesus monkeys, cynomolgus monkeys<br>Fertility and early embryonic development: rats<br>Developmental: rabbits, rats<br>Prenatal and postnatal development: none<br>Genetic toxicology: none<br>Carcinogenicity: none<br>Other: antigenicity study in rhesus monkeys |
| Lantus® Insulin glargine (US) Approval: 2000 | Recombinant glargine insulin (insulin analogue; long-acting) ↔ glucose metabolism | Diabetes | Single dose: mice, rats, dogs<br>Repeat dose: rats, dogs<br>Fertility and early embryonic development: rats<br>Developmental: rats, rabbits<br>Prenatal and postnatal development: rats<br>Genetic toxicology: Ames test, in vitro mammalian chromosome aberration assay, in vivo cytogenetic test, HGPRT test with V79 cells<br>Carcinogenicity: 2-year study in mice, 2-year study in rats study<br>Other: immunotoxicity studies in rabbits and guinea pigs |

**TABLE 1** ***Continued***

| Biopharmaceutical | Description ↔ Target | Indication | Toxicology Studies[a] |
|---|---|---|---|
| Leukine® (sargramostim) (US) Approval: 1991 | Recombinant human granulocyte macrophage colony-stimulating factor ↔ stimulates the production of neutrophils, monocytes/ macrophages, myeloid-derived dendritic cells | Following induction chemotherapy in older patients with acute myelogenous leukemia, after BMT, and before and/or after PBSCT | No information in the label |
| Levemir (insulin) (US) Approval: 2005 | Recombinant insulin detemir (insulin analogue; long-acting basal) ↔ glucose metabolism | Diabetes type I (adults and pediatrics) and type II (adults) | Single dose: mice, rats<br>Repeat dose: rats, dogs<br>Fertility and early embryonic development: rats<br>Developmental: rats, rabbits<br>Prenatal and postnatal development: rats<br>Genetic toxicology: in vitro reverse mutation study in bacteria, human peripheral blood lymphocyte chromosome aberration test, in vivo mouse micronucleus test.<br>Carcinogenicity: none<br>Other: mitogenicity in CHO-K1 and human B10 osteosarcoma cells, human hepatoma cells and MCF-7 cells to include drug binding study to IGF-1 and insulin receptors; local tolerance in pigs; immunogenicity studies in rabbits |

| | | | |
|---|---|---|---|
| Lucentis™ (ranibizumab) (US) Approval: 2006 | Recombinant humanized monoclonal antibody fragment (Fab) ↔ binds to vascular endothelial growth factor (VEGF) | Neovascular (wet) age-related macular degeneration | Single dose: none<br>Repeat dose: cynomolgus monkeys<br>Fertility and early embryonic development: none<br>Developmental: none<br>Prenatal and postnatal development: none<br>Genetic toxicology: none<br>Carcinogenicity: none<br>Tissue cross-reactivity: human tissues<br>Special toxicology studies: intravitreal in rabbits and cynomolgus monkeys hemolytic potential, blood compatibility and vitreal fluid compatibility, IV verteporfin photodynamic therapy following laser-induced choroidal neovascularization in cynomolgus monkeys |
| Mylotarg™ (gemtuzumab ozogamicin) (US) Approval: 2000 | Recombinant humanized IgG4 kappa monoclonal antibody conjugated with calicheamicin ↔ chemotherapy agent, binds to CD33 on leukemic blasts and immature normal cells of myelomonocytic lineage | CD33 positive acute myeloid leukemia | Single dose: rats, cynomolgus monkeys, chimpanzees, mice, dogs<br>Repeat dose: rats, cynomolgus monkeys<br>Fertility and early embryonic development: rats (male and female)<br>Developmental: rats<br>Prenatal and postnatal development: none<br>Genetic toxicology: mouse in vivo micronucleus test<br>Carcinogenicity: none<br>Tissue cross-reactivity: human, cynomolgus monkey, and rat tissues<br>Other: cardiovascular safety in dogs, in vitro blood compatibility, derivatives of gamma calicheamicin on murine bone marrow hematopoietic colony formation in vitro; in vitro stability of hP67.6 conjugate in human, monkey, and rat plasma |

**TABLE 1** ***Continued***

| Biopharmaceutical | Description ↔ Target | Indication | Toxicology Studies[a] |
|---|---|---|---|
| Myozyme® (alglucosidase alfa) (US) Approval: 2006 | Alglucosidase alfa ↔ enzyme replacement therapy | Pompe disease | Single dose: rats, dogs<br>Repeat dose: rats, mice, monkeys<br>Fertility and early embryonic development: mice<br>Developmental: mice<br>Prenatal and postnatal development: none<br>Genetic toxicology: none<br>Carcinogenicity: none |
| Naglazyme™ (galsulfase) (US) Approval: 2005 | Recombinant human *N*-acetylgalactosamine 4-sulfase ↔ lysosomal hydrolase that cleaves the sulfate ester at the end of glycosaminglycans | Enzyme replacement therapy for mucopolysaccharidosis VI | Single dose: rats, dogs<br>Repeat dose: cynomolgus monkeys<br>Fertility and early embryonic development: rats<br>Developmental: rats<br>Prenatal and postnatal development: none<br>Genetic toxicology: none<br>Carcinogenicity: none |
| Neulasta™ (pegfilgrastim) (US) Approval: 2002 | Recombinant human pegylated recombinant methionyl granulocyte colony stimulating factor ↔ hematopoetic cells | Reduction of incidence of infection in non-myeloid cancer patients | Single dose: rats<br>Repeat dose: rats, cynomolgus monkeys<br>Fertility and early embryonic development: rats<br>Developmental: rats, rabbits<br>Prenatal and postnatal development: rats<br>Genetic toxicology: none<br>Carcinogenicity: none<br>Tissue cross-reactivity: none<br>Other: antigenicity in dogs and chimpanzees (these studies were done as pharmacology and PK studies) placental transfer in pregnant rats |

| | | | |
|---|---|---|---|
| Neumega (oprelvekin) (US) Approval: 1997 | Recombinant human interleukin-11 ↔ stimulates the proliferation of hematopoietic stem cells, megakaryocyte progenitor cells, lymphoid cells, and induces the differentiation of megakaryocytes into platelets | Prevention of chemotherapy-induced thrombocytopenia and reduction of the need for platelet transfusions in patients with nonmyeloid malignancies | Single dose: rats, cynomolgus monkeys<br>Repeat dose: rats, cynomolgus monkeys<br>Fertility and early embryonic development: rats<br>Developmental: rats<br>Prenatal and postnatal development: none<br>Genetic toxicology: human lymphocyte metaphase analysis, mammalian cell mutation assay, mouse micronucleus test<br>Carcinogenicity: none<br>Other: interaction study in cynomolgus monkeys using Il-11 and G-CSF or GM-CSF |
| Neupogen® (filgrastim) (US) Approval: 1991 | Recombinant human granulocyte colony stimulating factor (G-CSF) ↔ hematopoetic cells | Chemotherapy-induced neutropenia | Single dose: mice, rats, hamsters, cynomolgus monkeys<br>Repeat dose: hamster, rats, dogs, cynomolgus monkeys<br>Fertility and early embryonic development: rats<br>Developmental: rats, rabbits<br>Prenatal and postnatal development: none<br>Genetic toxicology: Ames test<br>Carcinogenicity: none |
| Norditropin® (somatropin) (US) Approval: 2000 | Human recombinant growth hormone ↔ connective tissue, mineral, protein, carbohydrate, and lipid metabolism | Growth failure due to lack of endogenous growth hormone in pediatrics | Single dose: mice<br>Repeat dose: rats<br>Fertility and early embryonic development: none<br>Developmental: none<br>Prenatal and postnatal development: none<br>Genetic toxicology: none<br>Carcinogenicity: none<br>Other: transgenic human somatropin mice tested for immungenicity |

**TABLE 1** ***Continued***

| Biopharmaceutical | Description ↔ Target | Indication | Toxicology Studies[a] |
|---|---|---|---|
| Novolog<br>(insulin aspart)<br>(US) Approval: 2000 | Rapid-acting insulin analogue ↔ glucose metabolism | Diabetes type I and II | Single dose: mice<br>Repeat dose: rats, dogs<br>Fertility and early embryonic development: rats (males and females)<br>Developmental: rats, rabbits<br>Prenatal and postnatal development: rats<br>Genetic toxicology: Ames tests, mouse lymphoma cell forward gene mutation test, human peripheral blood lymphocyte chromosome aberration test, in vivo micronucleus test in mice and ex vivo UDS test in rat liver hepatocytes<br>Carcinogenicity: 2 one-year studies in rats<br>Other: cardiovascular effects in pigs |
| Nutropin depot<br>(somatropin)<br>(US) Approval: 1999 | Sustained release growth hormone ↔ connective tissue, mineral, protein, carbohydrate, and lipid metabolism | Growth failure due to lack of endogenous growth hormone in pediatrics; replacement of growth hormone in adults; Turner Syndrome; nongrowth hormone deficient short stature; growth failure associated with renal failure | Single dose: none<br>Repeat dose: rhesus monkeys<br>Fertility and early embryonic development: none<br>Developmental: none<br>Prenatal and postnatal development: none<br>Genetic toxicology: none<br>Carcinogenicity: none<br>Other: immunogenicity in transgenic mice expressing rhGH |

| | | | |
|---|---|---|---|
| Ontak<br>(denileukin diftitox)<br>(US) Approval: 1999 | Cytotoxic fusion protein composed of diphtheria toxin fragments A and B and Il-2 ↔ cytocidal action of diphtheria toxin to cells that express the IL-2 receptor | Persistent or recurrent cutaneous T cell lymphoma whose malignant cells express the CD25 component of the IL-2 receptor | Fertility and early embryonic development: none<br>Developmental: none<br>Prenatal and postnatal development: none<br>Genetic toxicology: Ames test, chromosomal aberration assay<br>Carcinogenicity: none |
| Oncaspar®<br>(pegaspargase)<br>(US) Approval: 1994 | Pegylated (PEG) recombinant human L-asparaginase ↔ depletion of asparagine for leukemic cells | Acute lymphoblastic leukemia | Fertility and early embryonic development: none<br>Developmental: none<br>Prenatal and postnatal development: none<br>Genetic toxicology: Ames test<br>Carcinogenicity: none |
| Omnitrope<br>(somatropin)<br>(US) Approval: 2006 | Human recombinant growth hormone ↔ connective tissue, mineral, protein, carbohydrate, and lipid metabolism | Growth failure in pediatrics due to lack of endogenous growth hormone; replacement hormone therapy in adults | *(follow-on protein product)*<br>Fertility and early embryonic development: rats (males and females)<br>Developmental: rats, rabbits<br>Prenatal and postnatal development: rats<br>Genetic toxicology: none<br>Carcinogenicity: none |

**TABLE 1** ***Continued***

| Biopharmaceutical | Description ↔ Target | Indication | Toxicology Studies[a] |
|---|---|---|---|
| Orencia™ (abatacept) (US) Approval: 2005 | Soluble fusion protein that consists of the extracellular domain of human cytotoxic T lymphocyte associated antigen 4 (CTLA-4) linked to the modified Fc (hinge, CH2, and CH3 domains) portion of human immunoglobulin IgG ↔ binds to T lymphocyte antigens CD80 and CD86 thus blocking interaction with CD28 | Rheumatoid arthritis | Single dose: cynomolgus monkeys<br>Repeat dose: mice, rats, cynomolgus monkeys<br>Fertility and early embryonic development: rats<br>Developmental: rats, mice, rabbits<br>Prenatal and postnatal development: rats<br>Genetic toxicology: Ames reverse mutation, Chinese hamster ovary ↔ hypoxanthine guanine phosphoribosyl-transferase forward point mutation assay, cytogenetics in human lymphocytes<br>Carcinogenicity: 2-year assay (subcutaneous in mice) |
| Orthoclone OKT3® (muromomab-CD3) (US) Approval: 1992 | Murine monoclonal antibody ↔ binds to the CD3 antigen of T lymphocytes | Rejection of transplanted organs | Fertility and early embryonic development: none<br>Developmental: none<br>Prenatal and postnatal development: none<br>Genetic toxicology: unclear<br>Carcinogenicity: none |
| Osteocel® (bone matrix) (US) Approval: 2005 | Allogeneic bone matrix ↔ repair, replacement, and/or reconstruction of bone defects | Regeneration of bones in orthopedic indications | Does not require pre-market approval by the FDA because it meets the regulatory definition of human cells, tissue, and cellular and tissue-based products, or HCT/Ps. |

| | | | |
|---|---|---|---|
| Pegasys® (peginterferon alfa-2a) (US) Approval: 2002 | Covalent conjugate of interferon alfa-2a with a single branched bis-monomethoxy polyethylene glycol (PEG) chain ↔ multiple targets within the lymphokine regulatory network | Hepatitis C, hepatitis B | Single dose: cynomolgus monkeys<br>Repeat dose: cynomolgus monkeys<br>Fertility and early embryonic development: male rhesus monkeys, female cynomolgus monkeys<br>Developmental: rhesus monkeys<br>Prenatal and postnatal development: none<br>Genetic toxicology: Ames test, in vitro chromosomal aberration assay in human lymphocytes<br>Carcinogenicity: none |
| Peg-Intron™ (pegylated interferon alfa-2b) (US) Approval: 2001 | Covalent conjugate of recombinant interferon alfa-2b with monomethoxy polyethylene glycol (PEG) ↔ multiple targets within the lymphokine regulatory network | Hepatitis C | Single dose: mice, rats, cynomolgus monkeys, rhesus monkeys<br>Repeat dose: rats, cynomolgus monkeys<br>Fertility and early embryonic development: cynomolgus monkeys<br>Developmental: rhesus monkeys<br>Prenatal and postnatal development: none<br>Genetic toxicology: Ames test, chromosomal aberrations in human lymphocyte assay, mouse micronucleus test<br>Carcinogenicity: none |
| Proleukin, IL-2® (aldesleukin) (US) Approval: 1992 | Recombinant human interleukin-2 ↔ receptors on T cells and NK (natural killer) cells | Metastatic renal cell carcinoma, metastatic melanoma | Single dose: none<br>Repeat dose: rats<br>Fertility and early embryonic development: none<br>Developmental: rats<br>Prenatal and postnatal development: unclear<br>Genetic toxicology: none<br>Carcinogenicity: none |

**TABLE 1** ***Continued***

| Biopharmaceutical | Description ↔ Target | Indication | Toxicology Studies[a] |
|---|---|---|---|
| Pulmozyme® (dornase alfa) (US) Approval: 1993 | Recombinant human DNAase ↔ cleaves DNA in sputum of lungs of cystic fibrosis patients to reduce sputum viscosity | Respiratory complications in cystic fibrosis | Single dose: rats, monkeys (inhalation)<br>Repeat dose: rats, monkeys (inhalation)<br>Fertility and early embryonic development: rats<br>Developmental: rats, rabbits<br>Prenatal and postnatal development: rats, cynomolgus monkeys<br>Genetic toxicology: Ames tests, chromosomal aberrations in cultured human peripheral blood lymphocytes, in vivo mutagenicity test (mouse bone marrow)<br>Carcinogenicity: 2-year rat inhalation study<br>Other: intravenous and intratracheal studies in rats and monkeys; blood and serum compatability (human and monkey), ocular toxicity |
| RAPTIVA™ (efalizumab) (US) Approval: 2003 | Recombinant humanized monoclonal antibody ↔ binds to CD 11a on leukocytes | Moderate to severe plaque psoriasis | Single dose: chimpanzees, mice (surrogate antibody)<br>Repeat dose: mice (surrogate antibody), chimpanzees<br>Fertility and early embryonic development: mice (surrogate antibody<br>Developmental: mice (surrogate antibody)<br>Prenatal and postnatal development: mice (surrogate antibody)<br>Genetic toxicology: none<br>Carcinogenicity: none<br>Tissue cross-reactivity: human, chimpanzee tissues, mouse tissue (surrogate antibody)<br>Other: immunotoxicity in mice (surrogate antibody) |

| | | | |
|---|---|---|---|
| Remicade® (infliximab) (US) Approval: 1998 | Chimeric monoclonal antibody ↔ binds with high affinity to the soluble and transmembrane forms of TNFα and inhibits binding of TNFα with its receptors | Crohn's disease, rheumatoid arthritis, ankylosing spondylitis, psoriatic arthritis, ulcerative colitis | Single dose: rats (not relevant), chimpanzees<br>Repeat dose: rats (not relevant), chimpanzees, mice (surrogate)<br>Fertility and early embryonic development: mice (surrogate)<br>Developmental: mice (surrogate)<br>Genetic toxicology: Ames, mouse micronucleus, human lymphocyte<br>Prenatal and postnatal development:<br>Genetic toxicology: Ames test, mouse micronucleus, human lymphocyte<br>Carcinogenicity: TNFα knockout mice, anti-murine TNFα antibodies on tumor development, 6-month mouse (surrogate antibody)<br>Tissue cross-reactivity: human tissues |
| Reopro® (abciximab) (US) Approval: 1997 | Fab fragment of chimeric monoclonal antibody 7E3 ↔ glycoprotein IIb/IIIa and vitronectin receptors | Reduction of acute blood clot related complications; treatment of unstable angina | Fertility and early embryonic development: none<br>Developmental: none<br>Prenatal and postnatal development: none<br>Genetic toxicology: in vitro and in vivo tests<br>Carcinogenicity: none |

**TABLE 1** ***Continued***

| Biopharmaceutical | Description ↔ Target | Indication | Toxicology Studies[a] |
|---|---|---|---|
| Rebif®<br>(interferon beta 1-a)<br>(US) Approval: 2002 | Recombinant interferon beta 1-a ↔ multiple targets within the lymphokine regulatory network | Multiple sclerosis | Single dose: rats, mice, cynomolgus monkeys<br>Repeat dose: rats, cynomolgus monkeys<br>Fertility and early embryonic development: built into repeat-dose study in cynomolgus monkeys<br>Developmental: cynomolgus monkeys<br>Prenatal and postnatal development: none<br>Genetic toxicology: Ames test, "unscheduled DNA synthesis" in cultured HeLa cells, chromosome aberrations in human lymphocytes, micronucleus induction in bone marrow cells of mice, chromosome aberration in the Chinese hamster bone marrow<br>Carcinogenicity: none<br>Other: peripheral blood leukocytes derived from humans, mice, rats, rabbits, dogs, and cynomolgus monkeys in vitro |
| Retavase®<br>(reteplase)<br>(US) Approval: 1996 | Nonglycosylated deletion mutation of tissue plasminogen activator (tPA) containing the kringle 2 and the protease domains of human tPA ↔ catalyzes the cleavage of endogenous plasminogen to generate plasmin | Acute myocardial infarction | Single dose: rats, rabbits, cynomolgus monkeys<br>Repeat dose: rats, dogs, cynomolgus monkeys<br>Fertility and early embryonic development: rats<br>Developmental: rats, rabbits<br>Prenatal and postnatal development: unclear<br>Genetic toxicology: Ames test, chromosome aberration test, micronucleus tests in mice and rats<br>Carcinogenicity: none |

| | | | |
|---|---|---|---|
| Rituxan™ (rituximab) (US) Approval: 1997 | Chimeric murine monoclonal antibody ↔ binds to the CD20 antigen | CD20 positive B cell non-Hodgkin's lymphoma, arthritis | Single dose: mice, guinea pigs (both not relevant species), cynomolgus monkeys<br>Repeat dose: cynomolgus monkeys<br>Fertility and early embryonic development: none<br>Developmental: cynomolgus monkeys<br>Prenatal and postnatal development: cynomolgus monkeys<br>Genetic toxicology: none<br>Carcinogenicity: none<br>Tissue cross-reactivity: various human tissues |
| Roferon-A® (interferon alfa-2a) (US) Approval: 1984 | Interferon alfa-2a ↔ multiple targets within the lymphokine regulatory network | Hepatitis C, chronic myelogenous leukemia, hairy cell leukemia | Fertility and early embryonic development: rhesus monkeys<br>Developmental: rhesus monkeys<br>Prenatal and postnatal development: rhesus monkeys (late fetal development)<br>Genetic toxicology: Ames tests, chromosome damage in human lymphocytes<br>Carcinogenicity: none |
| Simulect® (basiliximab) (US) Approval: 1998 | Chimeric murine monoclonal antibody ↔ reacts with the CD25 antigen on T cells, inhibiting the binding of interleukin-2 (IL-2) to its receptor (IL-2R)-appears to bind only to activated T lymphocytes and to monocytes ↔ macrophages | Prevention of rejection episodes in kidney transplant recipients | Single dose: none<br>Repeat dose: rhesus monkeys<br>Fertility and early embryonic development: none<br>Developmental: cynomolgus monkeys<br>Prenatal and postnatal development: none<br>Genetic toxicology: Ames test, chromosomal aberration tests with V79 Chinese hamster cells<br>Carcinogenicity: none<br>Tissue cross-reactivity: human tissues |

**TABLE 1** ***Continued***

| Biopharmaceutical | Description ↔ Target | Indication | Toxicology Studies[a] |
|---|---|---|---|
| Synagis™ (palivizumab) (US) Approval: 1998 | Recombinant humanized monoclonal antibody ↔ epitope in the A antigenic site of the F protein of respiratory syncytial virus (RSV) | Prevention of respiratory tract disease caused by syncytial virus in pediatric patients | Single dose: cynomolgus monkeys, rabbits<br>Repeat dose: rats<br>Fertility and early embryonic development: none<br>Developmental: none<br>Prenatal and postnatal development: none<br>Genetic toxicology: none<br>Carcinogenicity: none<br>Tissue cross-reactivity: human tissues, cynomolgus monkeys tissues |
| Synvisc® (hylan G-F 20) [US(CDRH]) Approval: 1997 | Mixture of Hylan A and Hylan B polymers derived from chicken combs ↔ lubrication of knee joint | Lubrication of the knee joint and reduction of pain in an osteoarthritic knee | Single dose: mice, rabbits<br>Other: immunogenicity in owl monkeys, in vitro cytotoxicity in mouse fibroblast cell lines, hemocompatibility, pyrogenicity, tissue implantation in rabbits |
| Thyrogen® (thyrotropin alfa for injection) (US) Approval: 1998 | Recombinant human thyroid stimulating hormone ↔ TSH receptors | Adjunctive diagnostic used in screening of cancer patients with thyroid cancer without thyroid glands | Single dose: rats<br>Repeat dose: rats<br>Fertility and early embryonic development: none<br>Developmental: none<br>Prenatal and postnatal development: none<br>Genetic toxicology: Ames test<br>Carcinogenicity: none |
| Thymoglobulin® [anti-thymocyte globulin (rabbit)] (US) Approval: 1999 | Anti-thymocyte globulin (rabbit) ↔ selective depletion of T cells | Immunosuppressive polyclonal antibody used in conjunction with other immune suppression treatments to treat acute rejection in renal transplant | Fertility and early embryonic development: none<br>Developmental: none<br>Prenatal and postnatal development: none<br>Genetic toxicology: none<br>Carcinogenicity: none |

| | | | |
|---|---|---|---|
| Tnkase™ (tenecteplase) (US) Approval: 2000 | Tissue plasminogen activator ↔ binds to fibrin and converts plasminogen to plasmin | Decrease in mortality associated with myocardial infarction | Single dose: rats, rabbits, dogs<br>Repeat dose: rats, rabbits, dogs<br>Fertility and early embryonic development: none<br>Developmental: rabbit<br>Prenatal and postnatal development: none<br>Genetic toxicology: none<br>Carcinogenicity: none<br>Other: local intra-arterial tolerance in the rabbit, local paravenous tolerance in the rats; in vitro hemolytic potential and blood compatibility testing<br>Special toxicology studies: renal safety in dogs, cardiovascular safety in cynomolgus monkeys, cardiovascular and respiratory safety in rabbits |
| Tysabri® (natalizumab) (US) Approval: 2004 | Recombinant humanized IgG4 antibody ↔ binds to the α4β1 subunit of integrin | Multiple sclerosis | Single dose: guinea pigs, mice<br>Repeat dose: cynomolgus monkeys, albino mice, rhesus monkeys (drug interaction study with Avonex and Tysabri)<br>Fertility and early embryonic development: guinea pigs<br>Developmental: guinea pigs, cynomolgus monkeys<br>Prenatal and postnatal development: cynomolgus monkeys<br>Genetic toxicology: mouse lymphoma, human lymphocytes<br>Carcinogenicity: (see Other)<br>Tissue cross-reactivity: yes<br>Other: xenograft SCID mice, xenograft athymic nude mice, proliferative in vitro studies with tumor cells |

**TABLE 1** ***Continued***

| Biopharmaceutical | Description ↔ Target | Indication | Toxicology Studies[a] |
|---|---|---|---|
| V Ectibix™ (panitumumab) (US) Approval: 2006 | Fully human monoclonal antibody ↔ targets human epithelial growth factor receptor | Metastatic colorectal cancer | Single dose: none<br>Repeat dose: cynomolgus monkeys<br>Fertility and early embryonic development: cynomolgus monkeys<br>Developmental: cynomolgus monkeys<br>Prenatal and postnatal development: none<br>Genetic toxicology: none<br>Carcinogenicity: none<br>Tissue cross-reactivity: human, cynomolgus monkey, rodent, rabbit tissues<br>Other: safety pharmacology study in cynomolgus monkeys |
| Xigris™ (drotrecogin alfa) (US) Approval: 2001 | Recombinant activated protein C ↔ hydrolyzes coagulation factors Va and XIIa | Sepsis | Single dose: cynomolgus monkeys, rhesus monkeys (continuous infusion), mice<br>Repeat dose: cynomolgus monkeys<br>Fertility and early embryonic development: rabbits<br>Developmental: rats<br>Prenatal and postnatal development: none<br>Genetic toxicology: none<br>Carcinogenicity: none<br>Tissue cross-reactivity: none |

| | | | |
|---|---|---|---|
| Xolair® (omalizumab) (US) Approval: 2003 | Recombinant humanized monoclonal antibody ↔ targets IgE | Asthma | Single dose: mice (not relevant), monkeys<br>Repeat dose: cynomolgus monkeys<br>Fertility and early embryonic development: cynomolgus monkeys<br>Developmental: cynomolgus monkeys<br>Prenatal and postnatal development: cynomolgus monkeys<br>Genetic toxicology: Ames test<br>Carcinogenicity: none<br>Tissue cross-reactivity: human, cynomolgus tissues |
| Zenapax® (daclizumab) (US) Approval: 1997 | Recombinant humanized monoclonal antibody ↔ directed against the high-affinity interleukin-2 receptor | Prophylaxis of acute organ rejection in de novo allogeneic renal transplantation; used concomitantly with an immunosuppressive regimen, including cyclosporine and corticosteroids in patients who are not highly immunized | Single dose: mice, rabbit, cynomolgus monkeys<br>Repeat dose: cynomolgus monkeys<br>Fertility and early embryonic development: none<br>Developmental: none<br>Prenatal and postnatal development: none<br>Genetic toxicology: Ames test<br>Carcinogenicity: none<br>Other: renal and cardio allograft studies in cynomolgus monkeys |
| Zevalin™ (ibritumomab tiuxetan) (US) Approval: 2002 | Immunoconjugate murine anti-CD20 monoclonal antibody + $Y^{90}$ + $In^{111}$ ↔ reacts with normal and malignant B lymphocytes | B cell non-Hodgkin lymphomas | Single dose: cynomolgus monkeys<br>Repeat dose: cynomolgus monkeys<br>Fertility and early embryonic development: none<br>Developmental: none<br>Prenatal and postnatal development: none<br>Genetic toxicology: none<br>Carcinogenicity: none<br>Tissue cross-reactivity: B cell cross-reactivity (only) |

**TABLE 1** ***Continued***

| Biopharmaceutical | Description ↔ Target | Indication | Toxicology Studies[a] |
|---|---|---|---|
| Zorbtive (somatotropin) (US) Approval: 2003 | Recombinant human somatotropin ↔ receptors on a variety of cell types including myocytes, hepatocytes, adipocytes, lymphocytes, and hematopoietic cells. | Short bowel syndrome | Fertility and early embryonic development: unclear<br>Developmental: rats, rabbits<br>Prenatal and postnatal development: unclear<br>Genetic toxicology: none<br>Carcinogenicity: none |

*Source*: The main source of the information in this table is either the Food and Drug Administration (FDA) Web site or European Medicines Agency (EMEA) Web site. For biopharmaceuticals approved before 1995 limited information is available so the author used the most recent label (package insert) to derive the reproductive, carcinogeneticity, and genetic toxicology studies completed for marketing approval.

[a]For brevity, only the pivotal toxicology studies are given. Examples of studies not included are proof-of-concept, comparability, toxicokinetic, pharmacokinetic, pharmacodynamic, local tolerance, and miscellaneous in vitro and in vivo studies. Pharmacokinetic and pharmacodynamic experiments generally used the pivotal toxicology animal model(s) because of the species specificity of the biopharmaceutical.

[b]It should not be inferred from the table that only one toxicology study per species was conducted for each biopharmaceutical. For example, several repeat-dose studies of different durations could have been conducted in the same species.

[c]Genetic toxicology language is from package inserts (most recent label).

[d]The primate species is not clarified in the label (package insert) for Actimune®.

(i.e., the label). There is a dearth of public information for drugs that came on the market before 1995, but sufficient public records exist for biopharmaceuticals approved during the last 10 years. Some of the early drug labels are lacking in descriptions of the genetic, reproductive, and carcinogenicity studies, particularly the radiopharmaceuticals, and in these cases the table is left blank.

## 2 TISSUE CROSS-REACTIVITY

Tissue cross-reactivity studies are required by FDA for monoclonal antibody products to determine if the product binds to target and/or nontarget tissues. They are also performed for nonmonoclonal biopharmaceuticals if warranted. For example, ARANESP™ was tested in a human tissue panel ex vivo to determine if it bound to nontarget tissues or cross-reacted with related cytokine receptors. These studies are also used to explore known or potential clinical adverse safety events (i.e., mechanism of toxicity). For example, one patient in a Raptiva™ study developed unilateral hearing loss. This finding was further evaluated by cross-reactivity studies with human optic chiasm, acoustic nerve, and inner ear tissues.

## 3 SINGLE-DOSE TOXICITY

The classic single-dose study using highly exaggerated doses (10–1000 times the human dose) of drug is not always feasible for large molecular weight biopharmaceutical drugs. Biopharmaceuticals are traditionally not oral drugs. Thus there is a limit to the amount of drug that can be formulated for intravenous or subcutaneous administration. Moreover there is a limit to the volume of drug solution that can be infused into a preclinical animal model without causing toxicities resulting in a rapid expansion of the extra cellular fluid volume and/or infusion of large quantities of a foreign protein.[1] Also one of the goals of the single-dose study for both animal and human drugs is to determine the toxicities that could occur due to an overdose. This is less of a concern with biopharmaceutical drugs, since many must be administered in a clinic setting. For all of these reasons the single-dose toxicology study[2] is not as germane for biopharmaceuticals as for traditional small-molecule drugs. Examples of biopharmaceuticals for which no single dose studies were conducted are Lucentis™, Avastin™, Simulect™, and Vectibix™.

[1]Thus safety factors tend to be lower (e.g., 2–10× on average) for biopharmaceuticals compared to small molecules.

[2]The classic single-dose study is actually more useful as a single-dose pharmacokinetic study than a toxicity study.

## 4 REPEAT-DOSE TOXICITY

A more relevant study for biopharmaceuticals is the repeat-dose study where exaggerated clinical doses are given to an animal model using the intended clinical route of administration. For most biopharmaceuticals this is the intravenous route of administration. Interesting exceptions are Lucentis™ and Pulmozyme®, given intraocularly and by inhalation, respectively, in the pivotal repeat-dose studies.

A unique problem encountered in a biopharmaceutical repeat-dose study is the development of antibodies to the drug, specifically neutralizing antibodies that diminish or eliminate its pharmacological activity. For example, the highly species specific Avonex® (interferon beta 1-alpha) causes neutralizing antibody formation within two weeks. Thus the pivotal repeat dose study for this drug was only a 14-day study in rhesus monkeys. Repeat-dose studies for Abraxane™ were complicated by immune reactions to its human serum albumin component. Another example is the pivotal repeat dose study in cynomolgus monkeys for Campath®, whose results were confounded by antibody development. However, there are products for which it is possible to dose through antibody development (i.e., Rituxan).

Determining the species specificity of a biopharmaceutical is an important first step in designing a toxicology program. This is accomplished with in vitro assays and tissue cross-reactivity experiments. For example, in vitro binding studies are done to assess the binding of the product to receptors in various species. By this approach it was determined that only primate IgE receptor binds to Xolair® precluding use of any species lower than primate in the preclinical program. Tissue cross-reactivity studies with rabbit and cynomolgus tissues were done to justify the use of these two species as the pivotal study animal models for Avastin™. However, there are biopharmaceuticals for which there is no relevant species in which to test the drug, except for the chimpanzee. In these cases a surrogate product may be developed. Raptiva™, for example, is only pharmacologically active in chimpanzees, a species for which limited data can be collected in a toxicology study (i.e., no necropsies permitted). A mouse surrogate antibody was thus developed and tested for Raptiva™ in single-dose, repeat-dose, and reproductive toxicity studies. A mouse surrogate antibody was also developed for the early Remicade® studies. There are examples of biopharmaceuticals that are relatively nonspecies specific and multiple animal models can be used for toxicity testing (i.e., ORENCIA™). Since many biopharmaceuticals are not active in lower species, the cynomolgus monkey is frequently used in the safety assessment of biopharmaceuticals.

The safety pharmacology studies in which the respiratory, CNS, and cardiovascular systems are evaluated in a nonrodent model are not de facto for a biopharmaceutical. Quite frequently the safety endpoints are in the protocols for the pivotal repeat-dose studies, such as a cynomolgus study, obviating the need for separate safety pharmacology studies.

## 5 GENETIC TOXICITY

Genetic toxicity testing is not required per ICH guidelines (ICH S6 [1]). The existence of genetic toxicity tests for biopharmaceuticals in the accompanying table is due, in part, to drug development prior to publication of the ICH guideline. There are no instances of positive outcomes for genetic toxicity testing with biopharmaceuticals except for ABRAXANE™, which was (not surprisingly) positive in genetic toxicity tests due to the paclitaxel component of the drug, and MYLOTARG™, positive in a mouse micronucleus test. The positive result for MYLOTARG™ is consistent with the known ability of calicheamicin (a component of the drug) to cause double-stranded breaks in DNA.

## 6 CARCINOGENICITY

The traditional two-year rodent carcinogenicity bioassay is also not required per ICH guidelines (S6 [1]) and is rarely done for biopharmaceuticals. However, because many biopharmaceuticals are immunosuppressive, there are concerns that specific products (e.g., growth factors) may cause inappropriate proliferation leading to neoplasm, especially if the drug is given chronically to patients. In these cases alternative carcinogenicity studies may be considered (i.e., Tg.rasH2 transgenic mouse model for KEPIVANCE®). Carcinogenicity (or tumorigenicity potential) has also been evaluated in chronic repeat-dose toxicology studies in primates (i.e., AMEVIVE®, APIDRA®).

## 7 REPRODUCTIVE/DEVELOPMENTAL TOXICOLOGY

The cynomolgus monkey model has been utilized for species specific biopharmaceuticals (e.g., Herceptin). One advantage of cynomolgus monkeys is that they cycle monthly compared to rhesus monkeys, which only experience an estrus cycle twice a year. The standard model (rat) is also used if the rat is deemed a relevant species, as are surrogate molecules for use in the rat. The segment 1, 2, and 3 studies are sometimes combined into one study, depending on the needs of the development program. Another alternative is to deduce the toxic effects of a biopharmaceutical on the reproductive system from the anatomical pathology results of the pivotal repeat dose study. For example, in the cynomolgus repeat dose studies for AVASTIN™, there were dose-dependent adverse effects on uterine and ovarian weights and follicular development.

## 8 IMMUNOTOXICITY

Immunotoxicity testing in various animal models is common for biologics because of their immunomodulatory nature. Since most biopharmaceuticals

are humanized proteins, the minimum requirement is to define the antibody response to the biopharmaceutical. Other aspects of the humoral response can also be evaluated in repeat-dose studies. Immunohistochemical staining of lymphoid organs and lymphocyte phenotyping have also been components of repeat-dose studies. The appearance of product specific antibodies may not be associated with acute or target organ toxicity (i.e., ORENCIA™). More important, immune system responses to a human biopharmaceutical in an animal model are not necessarily predictive of the effects on the human immune system to the biopharmaceutical.

## 9 CONCLUSIONS

It can be deduced from Table 1 that most biopharmaceuticals approved to date are monoclonal antibodies and the most common indication is oncology. However, barring oncology, the biopharmaceuticals span an impressive range of indications. In addition to monoclonal antibodies, other prevalent types of biopharmaceuticals are fusion or conjugate proteins, growth factors, replacement enzymes, and peptides. Cellular and tissue therapies are rarer but becoming more prevalent.

Although the preclinical programs for the biopharmaceuticals in Table 1 are each unique and were developed with an experimental, case-by-case approach, they do have some practices that distinguish them from small molecule drugs. The most important practice is that the species chosen for the toxicology study generally had the greatest pharmacological activity when pretested in a variety of species by binding assays. It was not automatically assumed that a two-year carcinogenetic study was informative if the biopharmaceutical had proliferative properties; other approaches such as cell assay and short-term rodent models were used. Tissue-binding studies were always conducted for monoclonal antibodies; use of tissue cross-reactivity studies for other types of biopharmaceuticals were found to be informative on a case-by-case basis. It was not assumed that a single-dose toxicology study was required for the preclinical program; in many cases this study was supplanted by a thoughtfully designed repeat-dose study. Genetic toxicology studies were rarely needed for biopharmaceuticals. If conducted, these studies were always negative, except for drugs having a chemical component. The cynomolgus monkey model, developed for biopharmaceuticals, often supplanted the traditional rat reproductive toxicology study. Endpoints of the segment 1 and 3 studies were frequently built into the segment 2 study. Last, immunotoxicity testing in variety of forms is a critical component of the preclinical program for biopharmaceuticals because most biopharmaceuticals are currently believed to either trigger an immune response (humoral or cellular) or suppress the immune system.

## ACKNOWLEDGMENTS

*The author thanks Theresa Reynolds and Mary Ellen Cosenza for their insightful comments on preclinical strategies for biopharmaceuticals.*

## REFERENCES

1. ICH Guidance S6. *www.fda.gov/cder/guidance/1859fnl.pdf*

# Science and Judgment in Establishing a Safe Starting Dose for First-in-Human Trials of Biopharmaceuticals

JENNIFER VISICH, PhD, and RAFAEL PONCE, PhD, DABT

**Contents**

1 Introduction
  1.1 General Considerations for Approaching the First Human Dose
  1.2 Establishing the Safety Profile for the Biological Therapeutic
  1.3 Factors That Increase Concern in Transitioning from Preclinical to Clinical Dosing
2 Calculation of Starting Doses for Biologics
  2.1 Calculation of a Human Equivalent Dose for Protein Therapeutics
  2.2 Examples of Estimations of First-in-Human Doses for Protein Therapeutics
3 Dose Escalation and Proposed Top Doses
4 Beyond Phase 1
5 Conclusions
  References

## 1 INTRODUCTION

A significant milestone in the development of a new therapeutic is the first administration into humans. Whereas this is generally accompanied by a sense of excitement at the prospect of advancing a promising therapeutic toward treatment of patients, there is also a strong focus on the potential for harm, particularly harm that was not predicted by the available preclinical studies. This

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

central concept of benefit versus risk forms the primary basis for decisions used by the developing institution, the regulatory authorities, and the clinical community in establishing acceptable parameters for the first clinical trial for novel therapeutic agents, and the underlying lynchpins of this assessment are the initial dose and the subsequent dose escalation strategy. This chapter reviews the decision processes surrounding the determination of the initial dose of a novel biopharmaceutical agent to be administered to subjects in clinical trials.

### 1.1 General Considerations for Approaching the First Human Dose

It is not practical to consider development of specific regulatory guidance for each of the diverse biologic therapeutics under development to treat various diseases. Rather, current practice is to rely on a risk-based strategy to protect subject safety. It is within this framework that the regulatory agencies rely on current scientific knowledge, best practices within the industry, and a case-by-case approach that is tailored to each therapeutic agent and specific subject population to evaluate the initial clinical testing of a novel biopharmaceutical. Specific case examples, which are presented later in this chapter, will provide some insight into how these concepts are applied.

A number of regulatory guidance documents touch on various aspects of the transition from preclinical to clinical study of a novel biopharmaceutical. The primary guidance documents describing the conduct and use of preclinical safety evaluation in supporting the initial clinical trials are as follows:

- Preclinical Safety Evaluation of Biotechnology-Derived Pharmaceuticals (ICH S6) [1]
- Non-clinical Safety Studies for the Conduct of Human Clinical Trials for Pharmaceuticals (ICH M3) [2]
- Points to Consider in the Manufacture and Testing of Monoclonal Antibody Products for Human Use [3]
- Estimating the Maximum Safe Starting Dose in Initial Clinical Trials for Therapeutics in Adult Healthy Volunteers [4]
- Guideline on Strategies to Identify and Mitigate Risks for First-in-Human Clinical Trials with Investigational Medicinal Products [5]

In practice, the regulatory framework governing the first dose of a biopharmaceutical in humans is designed around several fundamental tenets: know the safety profile for the test agent, maintain a positive risk–benefit profile in each subject population, and beware of the unknown.

### 1.2 Establishing the Safety Profile for the Biological Therapeutic

As described in ICH S6 guidance, preclinical safety studies for biopharmaceuticals are designed to identify potential toxicities (and their reversibility)

associated with treatment under conditions that mimic the intended clinical use (i.e., dose route, duration, and frequency with representative drug product). In so doing, these studies inform the clinicians, ethical review committees and regulators regarding the nature of potential side effects and appropriate subject monitoring, and guide the initial dose selection and dose escalation strategy.

Because biopharmaceuticals are often designed to mimic an endogenous human molecule (i.e., protein, peptide, nucleotide or other derivative), the burden is on the developer to demonstrate that the species used in the safety studies are appropriate surrogate models for humans; see [1]. This requirement recognizes the potential for broad interspecies differences in DNA/protein sequence homology, binding affinity, target distribution, and underlying biological pathways or processes that ultimately affect the value of the animal model to predict the human response. The requirement for such species justification data in the preclinical evaluation of biopharmaceuticals is an important area of divergence from the preclinical assessment of small molecules. Thus it is not meaningful to interpret the lack of toxicity observed for biopharmaceutical in species that have no demonstrated binding, activity, or other data to justify their use as predictive models of potential human response.

As the basis for decisions about the first dosing in humans, the preclinical safety program needs to mimic the intended clinical regimen. Although these studies are commonly conducted in normal healthy animals, they may also require development and use of an animal model that mimics the health status or condition if it is deemed to possibly sensitize the subject to treatment. For example, we have conducted studies evaluating the safety of recombinant human Factor XIII in an animal model of extracorporeal blood circulation [6] and the safety of recombinant human thrombin in an animal model of hepatic resection [7].

Dose selection and administration route are likewise defined by the intended clinical use. Of particular relevance to the selection of the initial clinical starting dose is the dose range used in the preclinical study and the dose–response relationship for identified toxicities. Under current regulatory practice [4], dose selection in the phase 1 enabling regulatory safety study should establish a no adverse effect level (NOAEL) and a maximal tolerated dose (MTD). Thus the design of the phase 1 enabling preclinical safety study often relies on data from prior pilot animal studies that define the tolerated dose range. But tolerability data in the preclinical species alone is generally not sufficient to establish an intended clinical dosing regimen because of potential interspecies differences in receptor/target affinity, distribution, or potency, or pharmacokinetic differences in biodistribution or clearance. Whereas an accurate human pharmacokinetic profile for the biopharmaceutical can only be established upon clinical testing, ex vivo and in vitro testing of human and animal cells can be used to identify differences in receptor/target affinity, distribution, or pharmacological activity. Accounting for such interspecies differences in selecting the final doses of the preclinical study provides a rational strategy that is matched to a target clinical dose range.

## 1.3 Factors That Increase Concern in Transitioning from Preclinical to Clinical Dosing

We have acknowledged that regulatory and clinical practices rely on a risk-based approach to establish the starting dose and dose escalation strategy for the initial clinical evaluation of a novel biopharmaceutical. But what are those factors that raise concern when transitioning from preclinical to clinical dosing? As a rule, agents associated with a threshold dose–response relationship for toxicity that lack premonitory biomarkers of toxicity, or that cause irreversible toxicity will have a markedly more conservative clinical program compared with agents that have a shallow dose–response relationship for toxicity, have clearly established premonitory biomarkers, that undergo complete reversibility of the toxicity. For example, preclinical safety evaluation of recombinant human Factor XIII demonstrated a threshold dose–response relationship for toxicity associated with few premonitory signs [8]. Thus clinical development of this compound was associated with a low starting dose, cautious dose escalation strategy, and intensive monitoring for evidence of toxicity.

In addition first-in-class therapeutics, particularly those with potential for serious adverse reactions, warrant an extra level of caution based on the pharmacological mechanism of action to account for our lack of knowledge. Recent experience with an experimental monoclonal antibody, TGN1412 (produced by TeGenero AG, Germany), against CD28 has galvanized interest in the process surrounding first-in-human trials. In March 2006, six healthy volunteers received a single 0.1 mg/kg body weight intravenous infusion of TGN1412 [9]. Within 1 to 2 hours after infusion, all six patients reported multiple symptoms including headache, myalgia, nausea, vomiting, bowel urgency and diarrhea, fever, erythema, peripheral vasodilation, hypotension, and tachycardia. Between 16 and 20 hours after treatment, the subjects had neurological, pulmonary, and renal impairment; disseminated intravascular coagulation; and lymphopenia and monocytopenia. Treatment with corticosteroids and/or an IL-2R mAb was used to downregulate the cytokine response. It was ultimately surmised that treatment with TGN1412 led to a systemic inflammatory response syndrome attributed to the release of TNF-α, IFN-γ, and IL-2.

Preclinical studies of TGN1412 were conducted in rhesus (*Macaca mulatta*) and cynomolgus (*Macaca fascicularis*) monkeys and demonstrated no evidence of a similar cytokine release response. Dose selection was based on a 500-fold margin of safety against the NOAEL established in cynomolgus monkeys (50 mg/kg). Comparison of the human and cynomolgus monkey CD28 sequence revealed an identical amino acid sequence in the critical region associated with TGN1412 contact, and TGN1412 binds to T cells from various nonhuman primates and humans [10]. Preclinical safety studies demonstrated no clinical toxicity at doses up to 50 mg/kg in cynomolgus monkeys, although increased serum cytokines were observed, including IL-2, IL-5, and IL-6 [11]. Subsequent data from TeGenero AG revealed comparable in vitro cytokine induction and T cell activation and proliferation for human and

cynomolgus monkey peripheral blood mononuclear cells [10]. The basis for the marked interspecies difference in response to TGN1412 has not been established but may relate to evolutionary differences in T cell response to stimulation between humans and nonhuman primates [12] or T cell co-stimulation related to $F_c$-receptor binding [13–15], or could represent faulty assumptions regarding dose extrapolation [16].

The experience with TGN1412 has resulted in an industrywide introspection regarding our uncertainty and lack of knowledge regarding interspecies extrapolation of safety findings, mechanism of action for novel pharmacologic pathways, and conduct of first-in-human (FIH) trials. These have been summarized in a number of reviews and editorials (e.g., [17,18–20]). In addition the European Medicines Agency (EMEA) has recently released guidance on requirements for first-in-human clinical trials for potential high-risk therapeutics [5]. Under this guidance, high-risk therapeutics are defined by the mode of action, including novelty, knowledge of the proposed mode of action and underlying physiology, type of dose–response relationship, and relevance of animal models. Agents that have a pleiotropic mechanism with the potential for systemic effects (e.g., the immune system) or that bypass physiological control mechanisms (e.g., CD3 or CD28 (super-)) agonists are identified as examples. As in our prior example, we would include agents with the potential to directly stimulate plasma coagulation to this list.

The EMEA guidance presents a departure from NOAEL-based methods for establishing the starting dose for FIH trials when the therapeutic is considered to be a high-risk agent. Rather, for high-risk agents, the FIH dose should be based on the anticipated minimal anticipated biological effect level (MABEL) *in humans* with incorporation of appropriate safety factors. The assessment of whether a particular agent is considered to fall under the high-risk category is conducted on a case-by-case basis, as summarized in Table 1. The concept of a MABEL is also described in the FDA maximum recommended starting dose (MRSD) guidance by the inclusion of the term PAD (pharmacologically active dose) [4]. According to the FDA guidance, the PAD is to be considered for cases where it may be a more sensitive indicator of potential toxicity than the NOAEL (certain classes of drugs or biologics where toxicity may arise via exaggerated pharmacologic such as vasodilators, anticoagulants, monoclonal antibodies, or growth factors). As such, it is recommended that the PAD would warrant lowering the MRSD. The terms MABEL and PAD will be used interchangeably in the remainder of this chapter. This is a good reflection of how safety margins should be considered for therapeutic proteins.

## 2 CALCULATION OF STARTING DOSES FOR BIOLOGICS

The preceding section describes the kinds of preclinical safety studies performed to enable the characterization of both the pharmacodynamics and the

**TABLE 1 Characteristics of standard versus high-risk therapeutics**

| | Standard Therapeutic | High-Risk Therapeutic |
|---|---|---|
| Mode of action | • Follow-on agent or therapeutic target<br>• Well-understood mechanisms<br>• Target has limited physiological interactions<br>• Target has linear or sublinear dose–response relationship<br>• Easily monitored pharmacology/toxicity<br>• Toxicity is reversible | • Novel<br>• Poorly understood mechanism<br>• Pleiotropic or systemic activity<br>• Bypasses or overwhelms physiological controls<br>• Potential for amplification, supralinear, or threshold dose–response<br>• Lack of biomarkers of effect/toxicity<br>• Irreversible toxicity |
| Nature of target | • Good specificity of therapeutic to target or disease<br>• Target is noncritical, has redundancy | • Poor specificity of therapeutic to target or disease<br>• Inherent risk of targeting specific structures, nonredundant systems<br>• Critical biological effect |
| Relevance of animal models | • Good structural homology, target distribution, and signaling pathways<br>• Comparable pharmacological effects<br>• Animal models established and considered predictive | • Poor structural homology, target distribution, signaling pathways, and pharmacological effects<br>• Animal models are of limited relevance to study pharmacology and toxicology<br>• Disease state alters relevance of studies in normal animals |
| Basis for first-in-human dose | • NOAEL-based approach<br>• Healthy volunteers possible<br>• Utilize all available in vitro and in vivo data, including receptor binding and occupancy, concentration–response relationships, PK data in relevant species<br>• Apply appropriate safety factor | • MABEL-based approach<br>• Patient population likely<br>• Utilize all available in vitro and in vivo data, including receptor binding and occupancy, concentration–response relationships, PK data in relevant species<br>• Apply appropriate safety factor |

toxicities associated with a new biopharmaceutical. At this point the toxicologist must utilize the information gathered to estimate a scientifically justified, safe starting dose in humans. There is no fixed algorithm used to estimate a safe starting dose. The estimation is performed on a case-by-case basis and must be justified both with regard to science and risk–benefit considerations.

The first step in the process is to understand the nature of the toxicity and the dose at which there is no observed adverse effect (NOAEL) and as described in the preceding sections, in the case of high-risk therapeutics, if possible, the minimal anticipated biological effect level (MABEL) or PAD and/or the dose at which there is no effect at all (NOEL). The NOAEL, the MABEL or PAD, and the NOEL are important starting points at which to base safe first doses in humans. However, there are many situations where the mechanism of action of a protein therapeutic is not measurable in healthy animals or where there is measurable pharmacology resulting from every dose tested (no NOEL estimated; MABEL or PAD is estimated), or where a NOAEL is not determined because an MTD has not been reached. In these cases the most relevant dose level should be used for the estimation of the first dose. The decision on which dose level should be based on the population to be included in the phase 1 and the risk–benefit scenario for these subjects. If the first phase 1 is conducted in healthy subjects, the goal may be that the first dose does not elicit a pharmacology of any kind; thus a dose based off the NOEL or the MABEL (or PAD) is desirable. If the first phase 1 is in subjects with disease, there may be a higher tolerance for the first dose to cause some pharmacologic effect but minimal adverse events; thus a dose based on the NOAEL or the MABEL (or PAD) is satisfactory.

In addition to in vivo data, in vitro receptor affinity and density data should be included in the first-in-human estimations when applicable. These data can be used to calculate the predicted receptor occupancy over a range of concentrations of the biopharmaceutical. The concentrations that result in a minimal receptor occupancy (e.g., 5–10%) should be compared to the circulating concentrations that correspond to the in vivo NOAEL, NOEL, and MABEL (or PAD). For example, ReoPro (abciximab) is a $F_{ab}$ fragment designed to bind to the platelet GPIIb/IIIa receptor and inhibit platelet aggregation. Preclinical animal models demonstrated that with more than 80% receptor blockade arterial thrombosis was prevented. Thus dose extrapolation from animals to humans could be used to estimate clinical doses associated with an 80% receptor blockade (and that could be monitored via flow cytometric analyses) and to define a rational dose escalation strategy. Alternatively, Mehrishi et al. have estimated more than 80% CD28 receptor occupancy among the six subjects treated with TGN1412 using equilibrium-binding calculations based on reported affinity data and assumptions of the number of cells and receptors/cell [16].

The final two steps of calculating a starting dose include the conversion of the NOEL, MABEL (or PAD), or the NOAEL to a human equivalent dose (HED) and the application of a safety factor that reduces the estimation of the human starting dose to allow for sensitivity differences in humans that may result in adverse events. The dose conversion can be done either using a pharmacokinetic-based interspecies scaling or a weight or surface area dose normalization. The safety factor applied to the calculated HED can be based on several factors, including nature of the dose–toxicologic response

relationship, what is known about the safety of other proteins with similar mechanism of actions, and in vitro sensitivity comparisons of the preclinical species versus human.

### 2.1 Calculation of a Human Equivalent Dose for Protein Therapeutics

Once the appropriate dose levels (NOAEL, NOEL, MABEL, and/or PAD) have been determined using both in vivo and in vitro data, these levels should be scaled for human estimation. In most cases the preclinical species have received repeated doses of the biopharmaceutical, resulting in increased exposure over the intended human exposure of a single dose. While total exposure in preclinical species (e.g., weekly or monthly) can be leveraged to calculate safety margins over expected human exposures in studies following the initial single ascending dose study, the emphasis for first-in-human dose estimations is most commonly on a dose-to-dose comparison, not on total weekly or monthly doses.

***Pharmacokinetic-Based Interspecies Scaling*** Allometric scaling is commonly used for the prediction of pharmacokinetic parameters of small molecules in humans. Allometric scaling is based on the assumption that there are anatomical, physiological, and biochemical similarities among the species included in the scaling. The power equation $Y = aW^b$ in its logarithmic form, $\log Y = \log a + b \log W$, has been used to describe the relationship of body weight ($W$) to both physiological properties ($Y$) such as organ weight or cardiac output to pharmacokinetic properties such as clearance, volume of distribution, and half-life. In the case of pharmacokinetic scaling, the weights of the animals tested and their observed PK parameters of interest (most commonly clearance) are plotted on a log–log plot where the $y$-intercept ($a$) and the slope ($b$) are calculated and then applied to a calculation of the predicted parameter in humans. The resulting estimate for clearance in humans can be used to estimate the first dose by using the following equation [21]:

$$\text{Human dose (mg)} = \frac{\underset{(\text{mg}*\text{h/L})}{\text{AUC in animal}} \times \underset{(\text{L/h})}{\text{Predicted clearance in humans}}}{\text{Safety factor}}.$$

In addition pharmacokinetic modeling can be used to predict human exposures, but these models are also based on parameter estimates that are initially scaled from preclinical species.

In the case of biopharmaceuticals, while two species are recommended, often times there is only one relevant species included in general preclinical safety studies because of species specificity. In either case, one or two species data are not sufficient to estimate the needed allometric terms. Therefore allometric scaling is not commonly used for biopharmaceuticals. There have

been a number of publications by Mahmood et al. (e.g., [22,23,24]) on the addition of additional factors (maximum life-span factors, brain weights, etc.) to improve the allometric scaling of protein drugs, but these improved methods still require data from greater than two species for accurate estimations. Last, as stated above, one of the major assumptions is that the drug will be handled similarly from species to species. In the case of biopharmaceuticals it is quite common for the species tested to have different sensitivities to protein drugs due to receptor density, receptor affinities, and interspecies differences in receptor signaling. These species specific differences are overlooked in the basic allometric estimations.

***Weight and Body Surface Area Interspecies Scaling*** The simplest scaling method is a one-to-one conversion of the animal dose of interest to a human equivalent dose (HED) via weight based (mg/kg) or body surface area normalization (mg/m$^2$). Choice of weight based versus body surface area scaling is usually on a case-by-case basis and requires some knowledge about the NOAEL, NOEL, or MABEL (or PAD) in several preclinical species and that the source of the observed toxicity is based on an exposure parameter that is correlated with doses administered by weight. For biopharmaceuticals, body surface area normalization is more widely used as it leads to a more conservative estimation of the starting dose. The FDA MRSD guidance provides tables that enable the conversion of several species doses to HEDs based on the following calculation:

$$\text{HED} = \text{Animal dose (mg/kg)} \times \left( \frac{\text{Animal weight (kg)}}{\text{Human weight (kg)}} \right)^{0.33}.$$

The body surface area conversion factors (BSA-CFs) included in the FDA guidance are based on an average species body weight.

***Application of a Safety Factor*** Following the conversion of the NOAEL, MABEL (or PAD), or NOEL to a human equivalent dose (HED), the resulting dose is usually divided by a safety factor in order to calculate the proposed starting dose in humans. This safety factor provides a safety margin between a safe dose in preclinical species and the first humans receiving the therapeutic. The default safety factor for most first-in-human studies is 10-fold, however, the assignment of this factor should be science-driven.

## 2.2 Examples of Estimations of First-in-Human Doses for Protein Therapeutics

The examples below illustrate strategies that have been implemented successfully in the estimation of safe first-in-human doses of biopharmaceuticals. Each of the molecules is characterized in terms of the type of therapeutic, whether

it is a standard or high-risk therapeutic and why, the relevant toxicology specie(s) studied, its mechanism of action (MOA), mechanism of toxicity (MOT), the planned phase 1 study design and population, and the identified NOAEL, NOEL, and MABEL (or PAD), where applicable. The molecule characterization is followed by a summary of the estimation of the first-in-human dose using the algorithm supplied by the FDA MRSD to calculate the HED and applying a safety margin based on case-by-case considerations.

### *Example 1*

| | |
|---|---|
| Type of molecule | Fusion protein |
| Classification | Standard (well-understood MOA, easily monitored pharmacology) |
| Toxicology species | Mouse, monkey |
| MOA | Known, not novel, dose-responsive biomarkers present in two relevant toxicology species |
| MOT | Known, exaggerated pharmacology |
| Phase 1 study | Single, escalating-dose in healthy subjects |
| NOAEL | No AEs observed at any dose levels |
| NOEL | No NOEL determined in either species |
| MABEL/PAD | 5 mg/kg in mouse and monkey |

The MABEL/PAD in the mouse and the monkey were divided by the BSF-CFs to calculate the corresponding HEDs:

$$\text{HED (mg/kg)} = \frac{\text{MABEL}_{\text{mouse}}}{13} = 0.38\,\text{mg/kg},$$

$$\text{HED (mg/kg)} = \frac{\text{MABEL}_{\text{monkey}}}{3} = 1.67\,\text{mg/kg}.$$

The lowest resulting HED was chosen (0.3 mg/kg) and a 10-fold margin was applied to estimate the first-in-human dose of 0.03 mg/kg.

### *Example 2*

| | |
|---|---|
| Type of molecule | Cytokine |
| Classification | High-risk (poorly understood MOA and MOT) |
| Toxicology species | Monkey |
| MOA | Poorly understood, dose-responsive biomarkers in 1 toxicology species |
| MOT | Poorly understood, monitorable |
| Phase 1 study | Repeat, escalating dose in end stage disease population |
| NOAEL | NOAEL identified; 100 μg/kg |

| | |
|---|---|
| NOEL | No NOEL determined |
| MABEL/PAD | 10 μg/kg in monkey |
| Other | In vitro studies showed similar receptor affinity for human and monkey |

The HED was calculated using the MABEL/PAD observed in monkeys:

$$\text{HED} = \frac{\text{MABEL}_{\text{monkey}}}{3} = 3.3\,\mu\text{g/kg}.$$

Because the phase 1 study was planned in end-stage disease patients, the in vitro receptor affinity was similar between monkey and human, and the NOAEL was 10-fold higher than the MABEL/PAD; a 3-fold safety margin was applied to the MABEL/PAD-based HED resulting in a safe first-in-human dose of 1.0 μg/kg.

### *Example 3*

| | |
|---|---|
| Type of molecule | PEGylated cytokine |
| Classification | High-risk (toxicity in one species and not the other) |
| Toxicology species | Mouse, monkey |
| MOA | Known, not novel, dose-responsive biomarkers measured in monkey |
| MOT | Unknown, observed in mice only |
| Phase 1 study | Single escalating dose in healthy subjects |
| NOAEL | No NOAEL identified in mice, no AEs observed at any dose level in monkeys |
| NOEL | NOEL identified in monkey; 0.015 mg/kg |
| MABEL/PAD | Not determined |

The mouse data included histologic evidence of mild cardiac injury that increased in incidence with increasing dose (0.3, 3, and 30 mg/kg) but had no functional consequences. The monkey data showed no evidence of these histologic findings. Similar findings were noted in mouse studies performed with a marketed agent in the same class as the tested biopharmaceutical, but likewise not observed in monkeys or humans. The NOEL in monkeys was based on a biomarker thought to be closely linked to the mechanism of action of the biopharmaceutical. The HED was calculated as follows:

$$\text{HED} = \frac{\text{NOEL}_{\text{monkey}}}{3} = 0.005\,\text{mg/kg}.$$

A 10-fold safety margin was applied to the HED and the resulting safe starting dose was 0.5 μg/kg. In addition this starting dose resulted in a 46-fold safety margin over the lowest dose tested in the mouse. The phase 1 study was conducted in healthy subjects and included additional cardiac monitoring.

## 3 DOSE ESCALATION AND PROPOSED TOP DOSES

The examples above illustrate the kinds of considerations that should be included in estimations of first-in-human dosing of biopharmaceuticals. In addition to the starting dose, the proposed dose escalation and the margin of safety surrounding the proposed top dose should also be considered carefully. In terms of dose escalation, the nature of the dose–response curve and inadequate knowledge of the mechanism of action and/or toxicity are the most important determinants of the need for caution in designing an ascending dose schema in humans. Many escalations with highly potent biopharmaceuticals begin with a conservative starting dose and proceed with dose doubling for several cohorts before eventually switching to half-log increases once the chance of unexpected toxicities has lessened. The proposed top dose of a single ascending dose escalation is usually designed to be a moderate multiple of the estimated therapeutic dose of the biopharmaceutical in humans. At a minimum, it is important to design preclinical safety studies to include a dose that at least matches the proposed top dose in humans (when appropriately scaled), although some regulatory reviewers have asked for a 10-fold or higher multiple in the preclinical study. The rationale for such a design should depend on the risk–benefit profile and the feasibility of achieving such doses. One should be mindful that for many biopharmaceutical agents that work through specific receptor-mediated processes, one would not expect to see significant novel biological activity with doses that exceed saturation of the receptor other than nonspecific toxicity associated with excessive protein loading.

## 4 BEYOND PHASE 1

Once the phase 1 has been completed, next steps include the comparison of observed human exposure data (maximum concentration, $C_{max}$ and area under the concentration–time curve, AUC) to the exposures observed in the supportive preclinical studies, prediction of the human exposure in future clinical trials (based on dose level and frequency) and the calculation of the "actual" safety margins provided by the repeat-dose preclinical toxicology studies. The need for safety margins (based on preclinical safety and single-dose human data) to support phase 1b or phase 2 dosing is usually considered on a case-by-case basis.

## 5 CONCLUSIONS

Calculation of the starting dose in first-in-human trials with biopharmaceuticals requires a good understanding of the science behind both the mechanism of action and the mechanism of toxicity of the test article. The FDA has provided an algorithm in its MRSD guidance to calculate a human equivalent

dose based on the NOAEL in the most sensitive species tested. However, thoughtful design of the supportive preclinical safety studies includes consideration of the estimated therapeutic human dose level and dose ranges that will allow for the estimation of a NOAEL, and ideally a NOEL, and/or a MABEL (or PAD), if possible. Based on the risk classification of the biopharmaceutical, the most appropriate preclinical dose level (NOAEL, NOEL, or MABEL) should be used to calculate the HED, and the application of a safety factor should be based on data gathered during in vitro and in vivo preclinical development.

## REFERENCES


1. ICH, Topic S6. Note for Guidance on Preclinical Safety Evaluation of Biotechnology-Derived Pharmaceuticals, 1998 (CPMP/ICH/302/95).
2. ICH, Topic M3 (R1). Note for Guidance on Non-clinical Safety Studies for the Conduct of Human Clinical Trials for Pharmaceuticals, 2000 (CPMP/ICH/286/95).
3. FDA CBER, Points to Consider in the Manufacture and Testing of Monoclonal Antibody Products for Human Use, 1997.
4. FDA CDER, Guidance for Industry: Estimating the Maximum Safe Starting Dose in Initial Clinical Trials for Therapeutics in Adult Healthy Volunteers, 2005.
5. EMEA CPMP, Guideline on Strategies to Identify and Mitigate Risks for First-in-Human Clinical Trials with Investigational Medicinal Products (EMEA/CHMP/SWP/28367), 2007.
6. Ponce R, Armstrong K, Andrews K, Hensler J, Waggie K, Heffernan J, Reynolds T, Rogge M. Safety of recombinant human factor XIII in a cynomolgus monkey model of extracorporeal blood circulation. *Toxicol Pathol* 2005;33(6):702–10.
7. Heffernan JK, Ponce RA, Zuckerman LA, Volpone JP, Visich J, Giste EE, Jenkins N, Boster D, Pederson S, Knitter G, Palmer T, Wills M, Early RJ, Rogge MC. Preclinical safety of recombinant human thrombin. *Regul Toxicol Pharmacol* 2007;47(1):48–58.
8. Ponce RA, Visich JE, Heffernan JK, Lewis KB, Pederson S, Lebel E, Andrews-Jones L, Elliott G, Palmer TE, Rogge MC. Preclinical safety and pharmacokinetics of recombinant human factor XIII. *Toxicol Pathol* 2005;33(4):495–506.
9. Suntharalingam G, Perry MR, Ward S, Brett SJ, Castello-Cortes A, Brunner MD, Panoskaltsis N. Cytokine storm in a phase 1 trial of the anti-CD28 monoclonal antibody TGN1412. *N Engl J Med* 2006;355(10):1018–28.
10. Duff GC. Expert scientific group on phase 1 clinical trials (Interim report). 2006; 356 pages. www.dh.gov.uk/en/consultations/closedconsultations/DH_4139038.
11. Doi F, Kakizaki S, Takagi H, Murakami M, Sohara N, Otsuka T, Abe T, Mori M. Long-term outcome of interferon-alpha-induced autoimmune thyroid disorders in chronic hepatitis C. *Liver Int* 2005;25(2):242–6.
12. Nguyen DH, Hurtado-Ziola N, Gagneux P, Varki A. Loss of Siglec expression on T lymphocytes during human evolution. *Proc Natl Acad Sci USA* 2006;103(20): 7765–70.

13. Wise MP, Gallimore A, Godkin A. T-cell costimulation. *N Engl J Med* 2006; 355(24):2594–2595; author reply 2595.
14. Colaco CA. T-cell costimulation. *N Engl J Med* 2006;355(24):2595; author reply 2595.
15. Sharpe AH, Abbas AK. T-cell costimulation—biology, therapeutic potential, and challenges. *N Engl J Med* 2006;355(10):973–5.
16. Mehrishi JN, Szabo M, Bakacs T. Some aspects of the recombinantly expressed humanised superagonist anti-CD28 mAb, TGN1412 trial catastrophe lessons to safeguard mAbs and vaccine trials. *Vaccine* 2007;25(11):3517–23.
17. Liedert B, Bassus S, Schneider CK, Kalinke U, Lower J. Safety of phase I clinical trials with monoclonal antibodies in Germany—the regulatory requirements viewed in the aftermath of the TGN1412 disaster. *Int J Clin Pharmacol Ther* 2007;45(1):1–9.
18. Schneider CK, Kalinke U, Lower J. TGN1412—a regulator's perspective. *Nat Biotechnol* 2006;24(5):493–6.
19. Kenter MJ, Cohen AF. Establishing risk of human experimentation with drugs: lessons from TGN1412. *Lancet* 2006;368(9544):1387–91.
20. Hanke T. Lessons from TGN1412. *Lancet* 2006;368(9547):1569–70; author reply 1570.
21. Mordenti J. Dosage regimen design for pharmaceutical studies conducted in animals. *J Pharm Sci* 1986;75(9):852–7.
22. Mahmood I. Interspecies scaling of protein drugs: prediction of clearance from animals to humans. *J Pharm Sci* 2004;93(1):177–85.
23. Mahmood I. The correction factors do help in improving the prediction of human clearance from animal data. *J Pharm Sci* 2005;94(5):940–5; author reply 946–7.
24. Mahmood I, Green MD, Fisher JE. Selection of the first-time dose in humans: comparison of different approaches based on interspecies scaling of clearance. *J Clin Pharmacol* 2003;43(7):692–7.

# *Introduction to Clinical Trials*

**John Goffin**

*Department of Oncology, Juravinski Cancer Center, McMaster University Hamilton, Ontario, Canada*

**Contents**

1 Goals of Chapter
2 Goals of Clinical Trials and What Is at Stake
3 Introduction to Phase I–IV Clinical Trials
  3.1 Introduction to Phase I Trials
  3.2 Introduction to Phase II Trials
  3.3 Introduction to Phase III Trials
  3.4 Introduction to Phase IV Trials
4 Principles of Trials Development
  4.1 Big Picture, Small Picture
  4.2 Human Element
  4.3 Multidisciplinary Nature of Clinical Trials
  4.4 Know Your Audience, Know Your Market
5 Example in Drug Development
  References

## 1 GOALS OF CHAPTER

The purpose of this chapter is to consider the overall goals and requirements of conducting clinical trials. It is an opportunity to avoid pitfalls by viewing the larger picture. This chapter seeks to provoke consideration of key issues without duplicating the more detailed work of later chapters.

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

## 2 GOALS OF CLINICAL TRIALS AND WHAT IS AT STAKE

The ultimate goal of drug development is the creation of new, safe, and effective compounds for treating human disease. Clinical trials comprise the portion of this endeavor involving human subjects. While the basic tenants of scientific inquiry do not differ from preclinical research, the stakes are higher and the regulations more stringent.

The cost of conducting clinical trials can be measured in two ways: the human cost and the resource cost. The human cost is the cost from the patient's perspective. The patient suffers from a condition dire enough that experimental therapy is a consideration. He or she holds out hope for this therapy and trusts to the scientific skill and integrity of those conducting the trial. The patients expose themselves to an incompletely understood therapy and usually suffer some degree of toxicity in order to gain uncertain benefit. Prior to a drug being declared useful or not, hundreds or thousands of patients may be involved in trials related to the drug.

On another balance sheet, there is the impressive economic burden of drug development. The cost of successfully bringing a new drug to market is now in the range of $800 million [1]. The interval between the start of clinical testing and the submission of an application for regulatory approval of a new drug is estimated at 6 years [1]. Even so, fields such as oncology are seeing an increase in drugs under study [2]. Yet there are limits to the number of clinical centers able to conduct trials. More importantly, there is a limit to the number of patients that are eligible to participate in a given trial, either by reason of demographic factors, comorbidity, incompatible disease parameters, or willingness. These limitations suggest that investigators must be selective about which drugs they study in clinical trials.

While drug discovery still involves an element of happenstance, contemporary drug development is ever more focused on mechanisms specific to a given disease. Frequently, therefore, a disease population will have been targeted during preclinical development. It is up to the clinical trials process to assess whether the new agent is both safe and effective in this or other populations. Generally, the first concern is assessing drug toxicity and the related dosing and pharmacokinetics. Following this, some evidence of efficacy is sought. If it is found, efficacy must be confirmed in larger, randomized trials. Finally, postmarketing surveillance studies may be performed. These successive clinical trials are usually categorized by phase, and these phases will be introduced below.

## 3 INTRODUCTION TO PHASE I–IV CLINICAL TRIALS

### 3.1 Introduction to Phase I Trials

***Purpose*** New drugs are first introduced into human subjects in phase I trials. The primary goal of these first studies is to assess the safety of the agent and to determine an acceptable dose for further study. Related goals include the assessment of pharmacokinetics as well as pharmacodynamics. To study pharmacokinetics is to study how the body affects the drug: How is the drug absorbed? How is the drug distributed between body compartments? How is the drug metabolized and excreted? Pharmacodynamics is the relationship between drug exposure and drug effect. Here

we ask what normal physiological or disease processes are altered when a drug is administered at varying doses.

***Methods*** The method used is to some extent dictated by the drug and disease under consideration. In fields other than oncology, phase I trials are typically undertaken in healthy volunteers. Typically, increasing doses of a drug are employed in small successive cohorts of patients. Each cohort is assessed, and subsequent dose levels are only used if excessive toxicity (often termed dose-limiting toxicity) is not encountered. At each dose level, blood or other body fluid is taken for pharmacokinetic studies.

In oncology studies, the first and lowest dose level may be based upon animal toxicities (e.g., 10% of the dose that is lethal in 10% of mice ($LD_{10}$)) and dose increments are often based upon a modified Fibonacci sequence (1, 2, 3, 5, 8, 13, ...), a scheme that decreases the dose increment with each subsequent level. The notion is to limit patient exposure to dose-limiting toxicity through more cautious later stage dose increases. Alternative dosing schemes employ one patient per dose level or a continuously modified dosing increment based upon observed toxicities; the goal of such alternative methods is to increase phase I efficiency and limit the number of patients who receive too little or too much drug [3]. At some point, toxicity is deemed to be excessive, and the appropriate dose level is then established, typically at the dose just below this point of excessive toxicity.

Pharmacokinetics is the study of the drug absorption, transport, distribution, metabolism, and elimination; the goal is to improve drug delivery and efficacy. An understanding of the molecular target may have implications for drug exposure. For example, antimetabolites used against cancer are considered to be most effective in the DNA (deoxyribonucleic acid) synthesis phase (S-phase) of the cell cycle. To best inhibit tumor growth, it is considered optimal to maintain a constant or prolonged exposure of the cancer to drug such that most cells are caught as they transit through S-phase. Pharmacokinetic analysis can tell the investigator if such an exposure is occurring and may prompt alternative dose schedules in subsequent studies.

Pharmacodynamic assays—assays that assess the effect of the drug on normal physiology or disease—may be useful in assessing whether a drug is likely to have a clinical effect. In cardiology, for example, the effects of a new agent on subjects' blood pressure or electrocardiogram may be relevant [4]. In studies of new antibodies or other targeted therapies, a therapeutic effect may be seen without the dose-dependent toxicities expected with other agents (e.g., the antimetabolite methotrexate used in rheumatoid arthritis or cancer). Conducting assays that demonstrate molecular changes in the relevant target could serve as a proof of concept for the agent; this, in turn, could prevent the need for higher dose levels, levels that could induce toxicity and would increase the duration of the study.

***Results*** At the end of a phase I study, acute toxicities should be understood. Toxicities related to more long-term exposure may not be apparent until future studies are undertaken. In conjunction with the pharmacokinetic assays and any pharmacodynamic work, an assessment must be made as to whether further studies should be conducted, and, if so, at what dose. Pharmacokinetic analysis may suggest that changes in dose or dosing frequency are required. In instances where toxicity may be excessive at doses not expected or observed to have a useful biological

effect, further phase I studies may be designed to circumvent the toxicity. While preliminary activity against disease may be observed in phase I studies, the initial assessment of positive clinical outcomes is primarily the arena of phase II studies.

### 3.2 Introduction to Phase II Trials

***Purpose*** Phase II studies are conducted to assess the initial activity of an agent against disease. Further information is gathered about an agent's adverse effects, and additional pharmacokinetic or pharmacodynamic studies may be conducted.

***Methods*** Unlike phase I studies, which may employ many different doses of an agent, phase II trials typically employ one or occasionally a few dose levels. Larger cohorts of patients are exposed to the drug in order to observe one or more clinical endpoints. The measured endpoints will vary depending upon the drug and field of study. In trials of heart failure, for example, physiological parameters (e.g., ventricular remodeling) may be assessed in addition to clinical measures such as exercise tolerance [5, 6]. Vaccine studies typically assess safety and immune responses and may involve both treatment and control groups [7]. In oncology, tumor response (shrinkage) rates have traditionally been used as a measure of response, but newer targeted drugs have led to greater reliance upon endpoints such as stable disease rates. Prior to conducting the study, investigators should specify what minimal level of drug activity will be accepted as evidence to warrant subsequent investigation. Phase II studies should be designed as precursors to phase III studies.

Phase II studies may be single-arm assessments of drug activity; such studies have an implied comparator of prior trials or clinical experience. Alternatively, randomized studies may be conducted, comparing the experimental arm with either a placebo, a standard therapy control arm, another experimental arm, or different doses of the experimental arm itself. The randomized study, while of limited power, may improve drug development by increasing the likelihood of selecting the best drug or dose for further development [8]. When a standard treatment arm is used as a comparator, that arm may serve as a barometer for the severity or nature of the disease in the overall study cohort. Excellent or poor results in the experimental arm are interpreted in light of the control arm.

A more recent study type, the randomized discontinuation study, begins with a lead-in period in which all subjects receive the experimental arm. After a predetermined period, subjects are randomized between continuing the study drug and receiving a placebo or no therapy. The lead-in period eliminates noncompliant subjects and unresponsive disease, increasing the likelihood of differences being observed in the randomized portion of the study. The cost is in the greater number of patients required for the study due to drop-out in the initial nonrandomized period [9].

***Results*** As noted, the clinical endpoints vary widely based upon disease and agent type. If a drug effect was seen, it must be considered whether the effect was sufficiently interesting in light of existing therapies or other study arms. If a clinical effect was not seen, one must assess whether this could be explained by any biological surrogates or pharmacokinetic studies also undertaken. The clinical efficacy must

be assessed in the face of observed toxicities. More severe toxicities might be acceptable for lifesaving therapies but not for agents directed at minor ailments. At the end of the phase II study, the investigator should have an initial assessment of a new agent's impact on a disease as well as a better understanding of the toxicity profile.

Two important and frequently used statistical concepts should be introduced here. The first is power. In clinical terms, power is the probability that a study will find that a drug is effective when the drug truly *is* effective. Statistically, it may be described as Power = $1 - \beta$, where $\beta$ is the probability of a study finding a drug ineffective despite the truth being that the drug *is* effective—$\beta$ is therefore also called the $\beta$ error. A related term, the $\alpha$ error, represents the opposite mistake; it is the chance that a study will find a drug effective when in truth the drug is ineffective. By general agreement, the value of $\alpha$ is usually set at 0.05. Power increases with larger studies (i.e., more patients) and when more prespecified clinical events occur. Phases I and II trials typically employ small numbers of patients, which tends to increase error rates and limit statistical options. Nevertheless, statistics can inform us of the limitations of our knowledge. For example, if we observed 3 of 25 patients with cancer to have tumor responses, we could determine that—with 95% likelihood—the true response rate was from about 3–30% [10]. If we had hoped for better, we would need to carefully consider any next trial. Phases III and IV studies, described below, rely heavily on thoughtful consideration of $\alpha$ and $\beta$ errors.

### 3.3 Introduction to Phase III Trials

***Purpose*** Phase III studies are typically large randomized studies designed to demonstrate useful clinical activity in a specific disease setting. The process of randomizing patients between different treatment arms is fundamental to avoiding biased interpretations of outcomes.

***Methods*** The design of phase III studies is critical both in addressing a specific hypothesis and in the pragmatic sense of making a drug useful in clinical practice. Fundamentally, this means that an appropriate patient population must be selected, all treatments must be clinically relevant, and the expected improvement in outcome must be both clinically meaningful and statistically measurable. Eligibility criteria—those criteria that determine which patients may join the study—must define a population that is both adequately generalizable to include patients representative of the diseased cohort but also homogeneous enough to retain statistical power and to be applicable to a usefully recognizable disease group. For example, studies may be difficult to interpret when they include both early- and late-stage patients. If a study is positive, to which population is it best applied? If negative, might it be positive in one of the disease subpopulations if a study were done only in that group.

Treatment arms cannot ignore previously existing therapies. With respect to heart failure, a new drug must take into account that many patients will also be on ACE (angiotensin-converting enzyme) inhibiters, $\beta$-blockers, diuretics, antiplatelet agents, and possibly other medications. Excluding these medications may make the study uninterpretable in the real-world clinical context and, more importantly, it may be unethical.

The endpoint of a phase III study should be an accepted and clinically relevant one that is specified before the trial is conducted. For example, in many cancers, an improvement in response rate is not considered an adequate phase III endpoint, whereas improvements in survival or disease-free survival may be accepted. Secondary endpoints—quality of life, for example—may be employed but must be recognized as such at study completion.

A common difficulty with phase III studies is inadequate power. This is often due to an overly optimistic estimate of improvement in a clinical outcome, an estimate that may be a product of resource limitations. A lesser and potentially meaningful improvement may be missed if too few patients are accrued to the study or follow-up is too short.

***Results*** The primary and any secondary clinical outcomes must be assessed and interpreted as planned. In circumstances where the primary outcome is of borderline significance or where the primary and secondary clinical outcomes are disparate, explanations may be considered and used as hypotheses for future study. Post hoc analyses are frequently conducted but can only be hypothesis generating.

### 3.4 Introduction to Phase IV Trials

***Purpose*** Phase IV studies, sometimes called pharmacoepidemiologic studies, are those that are conducted after a drug has been approved for marketing. Such studies, often large, may assess a drug for uncommon toxicities that may be undetectable in smaller phases I–III studies, or they may establish the activity or tolerability of a drug in a particular population or practice setting.

Studies conducted to assess new methods of drug administration, combinations with other agents, or activity in other diseases—that is, studies seeking a new marketing indication—are better described and conducted as the phases I–III studies they represent. Similarly, a distinction can be made between trials seeking to answer a specific postmarketing question and those conducted solely to increase market share, so-called seeding trials. In the latter, there may be an incentive for the involved physicians to prescribe the drug in question and there may be no intent to publish the results [11, 12].

***Methods*** Phase IV studies may be conducted in several ways.

1. Descriptive studies, sometimes collections of drug toxicities captured over time, may identify new problems. These may range from case studies to series of patients collected by companies or regulatory bodies. Although resource intensive, large prospective cohort studies may also be conducted to capture infrequent adverse events.
2. Randomized studies may be used to compare an agent to other similar agents or to confirm earlier results.
3. Case–control studies or retrospective cohort studies can be conducted after data on a drug has accumulated. This would typically be done to assess for unusual side effects or associations of a drug with the development of a subsequent disease, such as malignancies or autoimmune sequelae.

4. Cross-sectional studies, although perhaps less useful, assess drug exposure and outcomes in a population at a specific time. Causality may be more difficult to assess if a sequential temporal relationship cannot be determined [12].

***Results*** The results of phase IV studies may be required to fulfill regulatory requirements after accelerated approval of a new drug. The additional numbers and prolonged follow-up provided by postmarketing studies may also be crucial in revealing important but infrequent toxicities. On occasion, these findings may lead to the withdrawal of a drug from the market, as, for example, after cardiovascular complications were associated with the anti-inflammatory drug rofecoxib [13, 14].

## 4 PRINCIPLES OF TRIALS DEVELOPMENT

### 4.1 Big Picture, Small Picture

***Overall Goal: Improved Patient Care*** The details involved in protocol design and regulatory requirements can be overwhelming. Remembering the fundamental goal of clinical research—improved patient care—can be an aid; study design and decision making should be influenced by the consideration of what is best for patients.

Patients seek relief from suffering. The investigator should therefore choose the most relevant endpoint for a given trial. Studies of rhinitis may reasonably examine patient reporting of nasal discharge and congestion [15], while studies of pancreatic cancer must consider an agent's impact on survival or more relevant measures of symptoms or quality of life. Research protocols must be designed with these parameters in mind. The outcome of interest must be described in sufficient detail that it may be easily replicated, a matter as important in assessing a study's value in support of regulatory approval as it is to an understanding of what benefit a drug may be to future patients. Any clinical trial must assess the toxicities associated with treatment. Known adverse effects must be clearly described and provisions made for the adjustment of treatment to mitigate such toxicities should they occur. Of course, for sufficiently severe toxicities, a warning system must be in place to inform patients, investigators, and the companies and agencies overseeing the study. The details of such reporting requirement may vary, but the act of sharing such information is sensible.

***Quality*** After careful protocol development comes the messy process of administering a protocol. Invariably, aspects of the protocol appear to be open to interpretation, and at some point there will be lapses in study conduct or paperwork. The maintenance of quality in a study means always trying to adhere to the letter and spirit of the protocol. It means that the responsible investigator must be available to arbitrate whether patients are actually eligible and whether protocol violations have occurred. It means that study coordinators must vigorously pursue the complete assessment of patients and the related documentation. Every effort must be made to follow patients to the completion of study. A poorly followed or documented study may be difficult to interpret and may not be acceptable to regulatory agencies or other entities overseeing the trial.

***Nothing in Isolation—The Bench and the Bedside*** The present era is one of exciting new agents, many directed at specific targets in the disease process. Even while such agents must undertake the staged clinical trials process, they may evoke interesting biological questions with implications for ongoing or future studies. The prospective collection, banking, and analysis of biological specimens may reveal subsets of patients for whom a new agent may have particular benefit.

For example, small-molecule tyrosine kinase inhibitors directed at the endothelial growth factor receptor (EGFR) have been investigated in patients with non–small cell lung cancer. Despite good preclinical data [16], clinical studies demonstrated more limited benefit, ultimately resulting in limitations of access to one such drug, gefitinib, previously approved by the Food and Drug Administration (FDA)under accelerated approval [17]. The investigation of tumor samples, however, revealed that some tumors had mutations in the tyrosine kinase domain of the *EGFR* gene, with corresponding protein changes and apparent improvements in clinical responses [18, 19]. Unfortunately, this finding was made posthumously for gefitinib, but the implications for future development of this class of drug are clear. When feasible, biological investigations and specimen preservation should continue during the clinical period of study.

### 4.2 Human Element

***Differences between Mice and Humans*** Despite the fact that 99% of mouse genes have human counterparts [20], several important issues separate the species. First, important differences in biology can mean significantly different drug metabolism and elimination, such that pharmacokinetics can only be generally predicted [21]. Second, human xenografts planted in mice may respond to drug therapy, but such responses are not consistently predictive of response phase II clinical studies [22]. This supports the necessity of clinical studies. Third, ethics dictates that both the goals and conduct of preclinical and clinical studies must differ. In animals, while the suffering and distress of animals is to be minimized [23], it is accepted that toxicities must be observed in other species to understand new agents and protect the humans that are subsequently exposed. By contrast, the very structure of trials in humans is one of careful staging to avoid excessive toxicity or any death. Earlier studies establish safety while later studies assess for useful clinical activity of a drug.

***Relevance of Ethics*** There are more and less obvious aspects of ethics involved in clinical drug development. We have fortunately recognized and codified the obvious, so, for example, it is universally recognized that withholding effective treatment for the sole purpose of observing natural disease history is unethical [24]. But there are less flagrant examples that affect study design.

The phase I study by its nature poses ethical conundrums. It is a study designed to assess toxicity and an acceptable dose for a drug, with clinical benefit being a secondary consideration. Thus, subjects put themselves at risk for uncertain benefit, and healthy volunteers stand no chance of clinical benefit. But the phase I trial is accepted for several reasons. First and foremost, if one accepts that our society wishes to continue to make progress against disease, it becomes an unavoidable necessity. A new drug must at some point be introduced into the human population.

This must be done in a careful and systematic fashion, but risk can only be minimized, not eliminated.

Second, patients who face the option of a phase I study are often those who have a disease without further standard therapeutic options. Although the chance of benefit for a given patient is likely to be very low, a chance for therapeutic success may be motivation enough [25], and altruism may play a smaller role in patient decision making than frequently thought [26]. Yet even when informed consent may be forthcoming, phase I studies are at greater risk than later phase studies for violating the principle of beneficence (i.e., offering insufficient benefit to justify risk) and for abusing the desperation of a vulnerable patient population at the expense of the ethical principle of justice [27].

Another challenging aspect of phase I studies is drug dosing. In oncology, it has been observed that benefit derived from new cytotoxic drugs occurs more frequently when doses are near the limit of acceptable toleration of side effects [28]. This means that patients who receive lower drug doses earlier in the study are less likely to have benefit, although they may also have less toxicity. Phase I dosing is therefore a balance between minimizing toxicity and maximizing any possible benefit for the greatest number of patients [25]. It is thus incumbent on investigators to carefully plan dosing increments during protocol development and assess side effects as the trial progresses.

Phase III studies, though more likely to confer benefit than phase I studies, still pose ethical challenges. One such difficulty is the decision about whether to stop a trial during interim analysis. A trial of hormone therapy (letrozole) after curative surgery for breast cancer was stopped at an interim analysis when the treated patients demonstrated lower rates of disease recurrence [29]. It may reasonably be asked whether such a study might better be continued blinded until longer follow-up was available or a survival difference was or was not found. While unquestionably it is better to avoid recurrence of breast cancer, the cost of adopting such therapy must be balanced against an incomplete study, other potentially better therapies, or trials that might be aborted by early adoption of the considered drug [30]. We are also accepting the financial cost of a new drug by its adoption. A society may reasonably consider for any therapy whether the gains so achieved are incurred at a reasonable cost in terms of other societal concerns. Such issues make it apparent that ethics is not a matter of nebulous constructs but an integral consideration for clinical trials.

***Quality of Life*** Another aspect of research that separates the clinical from the preclinical phase is the human interpretation of ailments. From pain to dyspnea, humans demonstrate a range of subjective degrees of discomfort from the insults of disease [31, 32].

Although less concrete and more difficult to assess than endpoints such as survival or hospital admissions, quality of life or symptom control data can be meaningful to patients and clinicians. In circumstances where endpoints such as survival are not readily demonstrated, such as in rheumatoid arthritis, measurements of quality of life, symptoms, and function are useful to assess drug efficacy [33]. Investigators should endeavor to use validated scales so that the results are less open to question. Still, quality of life measures have provided challenges. How often does one conduct measurements? How does one account for the inevitably missing data points [34]?

In the field of oncology, quality of life scales alone have yet to prove sufficient for drug approval by the FDA. In contrast, other simple and easily comprehensible measures of pain or composite endpoints that include pain have been accepted as a basis for drug marketing [35].

### 4.3 Multidisciplinary Nature of Clinical Trials

***Actors*** The manifold tasks and varied expertise required to conduct contemporary clinic trials necessitate the input and assistance of several groups. Prior to initiating a clinical trial, it must be assured that all the players are properly cued. Table 1 lists the persons and groups that typically must be available to conduct a trial, listed roughly in order of appearance but not importance.

Due to the diverse resources required to conduct clinical trials, it is not always practical for an organization to maintain capacity for every aspect of study conduct.

**TABLE 1 Entities Involved in Clinical Trials**

| Entity | Role |
|---|---|
| Principle investigator | While not all trials are conceived by the principle investigator, the principle investigator is responsible for the overall conduct of the trial. |
| Funding agency/ company | This may be a corporate, government, or charitable agency. In addition to funding, companies may supply drug. These bodies are frequently involved in receiving and disseminating reports of adverse events. |
| Statistician | Statisticians are involved in study design, interim analyses, and the final analysis. |
| Study coordinators | Study coordinators are involved in all aspects of trials: protocol and form creation, submission of the protocol to various review boards and government regulatory agencies, patient consent and registration, as well as data collection, cleaning, and summation. |
| Contract and financial administrators | These persons negotiate agreements between funding agencies and centers conducting the trial, aid in the creation of budgets, and distribute funds necessary to conduct the trial. |
| Scientific review committee | This body reviews the scientific merit of a clinical trial and may suggest improvements. |
| Health/safety committee | Although not involved in all studies, this group is responsible for ensuring that investigators adhere to regulations regarding infectious and hazardous substances. |
| Institutional review board/ethics committee | This body assesses whether the study meets the standards of respect for persons, beneficence, and justice and will prohibit substandard studies. |
| Data safety monitoring board | Created before the initiation of the trial, this body provides objective oversight of the study and may recommend early closure of a study for reasons of either significant early benefit or excessive toxicity. |
| Pharmacists | Pharmacists are responsible for research drug control and accounting. |
| Nursing staff | Drug administration and sample collection requires both nursing staff and physical space, sometimes including facilities for overnight visits. |
| Pharmacokinetics specialists | Pharmacokineticists are usually involved in phase I drug design and sample collection and analysis but may also be involved in later phase studies. |
| Outcomes assessments staff (e.g., radiologists) | Depending upon the outcomes being assessed, radiologists or other specialists may be required to interpret study data. In some instances, independent and blinded individuals or groups may be used to assess study data in a more objective fashion. |

For this reason, an industry of contract research organizations has arisen to provide research services not available from in-house sources. These organizations can provide services such as research ethics review, protocol preparation, study administration, regulatory consultation, and radiologic imaging support. They can offer the advantages of expertise and efficiency in trial conduct, with offsetting disadvantages of decreased control over details, the need to rely on the contract agency for quality, and the need for careful communication with respect to the hired agency's responsibilities and goals [36].

***Statisticians*** The early inclusion of an experienced statistician is advisable for most studies. In order to obtain a useful study result, a hypothesis must be generated and a statistical test must be chosen prior to study conduct. *Post hoc* statistical analyses can lead to new hypotheses for future research but cannot generate definitive answers [37].

A statistician can help to clarify the question under consideration. For example, when conducting a phase II study in heart failure, one may wish to assess the difference in exercise duration between two treatment arms [6]. Using the expected or minimally acceptable difference and the desired error rates, a statistician can advise on the number of patients that need to be recruited to the trial. Failure to determine this need may result in a futile, underpowered study or one which unnecessarily exposes excess patients to an experimental therapy.

In larger, phase III studies, the patient exposure and resource stakes are typically greater. As with our phase II example, realistic expected differences between endpoints must be considered. It must be decided whether the new therapy is likely to be superior, or whether the investigator wishes only to demonstrate that it is noninferior (although either less toxic, more convenient, or substantially cheaper), as the sample size will be larger in the latter case and the hypothesis test different. The ethical challenges of the interim analyses were previously mentioned, but the statistical challenges can also be substantial. One must estimate how many events are required in a population to sufficiently conduct the analysis, then employ a test that will assess the difference while accounting for repeated statistical testing. The goal is to avoid both false-positive studies and prolongation of a futile trial [38].

***Setting*** During study development, investigators must decide where the trial will be conducted: primarily among academic centers and cooperative organizations or in community centers, usually under the auspices of a pharmaceutical company and frequently organized by contract research organizations. In addition, a study will be domestic or international.

Traditionally, academic centers and organizations have conducted clinical trials, although this has been changing [39]. While the clinical trials infrastructure is more commonly in place in academic centers, community centers have demonstrated the ability to conduct clinical trials as effectively as academic centers [40–43], and organizations have formed that may efficiently recruit patients within such centers [39]. Community-center-based trials may have the advantage of a more generalizable patient population than that seen in academic centers [44, 45]. Limitations of community trials may include limited recruitment despite declared interest, a need for financial incentives, a need for easy documentation, and a lack

of perceived benefit to the physician [46] or managed care organization [47], although these characteristics are by no means exclusive to community trials [48]. The corporate control of data, use of for-hire ethics boards, and the greater dependence on financial incentives can leave some community trials more open to question [39]. Indeed, possibly as a result of publication bias and data control, publications of industry-sponsored work tend more often to report in favor of the experimental therapy [49]. For this reason, studies conducted by academic centers may offer superior credibility.

While the logistical and regulatory convenience of domestically conducted clinical trials is undisputed, there may be advantages to studies conducted on an international basis. Most evidently, the recruitment pool may be vastly increased, particularly when countries are included where nonexperimental options are relatively limited—a source also of some ethical debate [50]. Dollar costs may also be reduced when developing countries are involved [51]. The result of international studies may be more generalizable and more readily accepted by clinicians debating the applicability of a trial to their setting. While the international adoption of standards such as the Guideline for Good Clinical Practice aims to facilitate drug development by improving the acceptance of trial results by the regulatory bodies of differing countries [52], the actual conduct of such studies can still be challenging. Differing bureaucracies and approval methods for experimental studies can mean expensive or prolonged approval processes. In developing countries, the conduct of trials may require increased support for centers with little experience conducting clinical trials, and simplified and minimized information collection. Despite the sometimes difficult logistics, it is recommended that randomization remain centralized [53].

### 4.4 Know Your Audience, Know Your Market

***Who Is the Audience?*** When developing a clinical trial, one must take into consideration the interested parties. First and foremost, there is the patient, who must deem the trial safe and attractive. There is the clinical investigator (and institutional review board), who must find the trial to be of sufficient scientific and ethical merit to allow accrual. There are the regulators, who may need to approve the trial for it to proceed and who will eventually need to approve an agent for nonexperimental use. And finally, there is the market, really an amalgam of the wills of patients and clinicians as influenced by competing therapies. While the term market connotes a mercenary purpose, the consideration of a drug's market is both worthy of time and compatible with the goal of optimal patient care.

***Considering Current and Evolving Practice*** Clinical trials are not conducted in isolation. Rather, they become available to patients as an option alongside existing standard therapies. This imposes limitations on the experimental and control arms for a trial. For example, in many instances a patient commencing treatment for more than very mild rheumatoid arthritis (RA) would be a candidate for methotrexate [54]. Starting a patient purely on an experimental therapy could thus be deemed inadequate, and the experimental arm may need to employ both methotrexate and the experimental agent in combination. Similarly, it is considered unethical to unnecessarily delay treatment through the use of a placebo in the control arm of

a patient with RA [55]. Obviously, a trial that fails to consider these points is unlikely to be allowed to proceed, and even if approved may be unable to accrue.

Recognizing variations in clinical practice, a flexible treatment scheme has sometimes been adopted by trialists. In lieu of defining a specific control arm in a clinical trial, investigators may be allowed to choose the particular control or treatment arm that will be employed at their center [56, 57]. Such a trial is more likely to be attractive to a wider range of clinicians, as they may adopt local practices to the trial in question. This has obvious benefits for accrual and may enhance the generalizability of the study. On the other hand, such an open model may make it less clear what is being compared. For example, if the experimental arm contains several forms of therapy, a local investigator may not know if his or her standard regimen has been adequately compared to the experimental arm based on the primary analysis. While subset analyses may be performed, they are typically exploratory.

In addition to present practice, other concurrently operating clinical trials may impact on future practice and the ability to conduct a trial under development. First, a trial in the same population may compete for the finite pool of potential participants. Second, if a competing study is finished and found to be positive before the developing or ongoing study is complete, study completion may become impossible. The competing study may change the standard treatment landscape, alter investigator equipoise over the developing or ongoing study, and inhibit patient accrual. Patients will need to be informed of the evolving standard, and they may choose to avoid or withdraw from the trial.

Just as standards exist in clinical practice, methods of conducting clinical trials are largely standardized. While trial methodology is evolving, investigators and review boards may be uncomfortable with new methods. For example, in phase I studies in oncology, the common method of accrual is to admit cohorts of three to six patients at successive drug doses. Alternative methods, such as the accrual of one patient per dose level, or the continuous reassessment of the maximum tolerated dose using Bayesian methods have been advocated as potentially more efficient [3]. However, there is evidence that the implementation of new study methods is delayed, suggesting the discomfort of physicians or reviewing committees [58].

***Considering Endpoints*** The choice of endpoint depends upon both the disease under consideration and the phase of clinical development of the drug. In congestive heart failure (CHF), for example, past successes in improving clinical outcomes have made it difficult to further improve results and to detect such improvements in phase III trials [59]. For drug development, this means that having an early, phase II assessment of activity is important to determine whether a drug should go on to phase III study. Given that phase II trials are intended to be shorter and smaller than phase III trails, using longer term endpoints such as hospitalization or mortality is unlikely to be practical. Surrogate endpoints are therefore considered for these phase II trials. While clinical endpoints represent measures of disease important to patient well-being or survival, surrogate endpoints are alternative endpoints that represent disease biology or a secondary clinical outcome and are intended to shorten the investigative timeline. To be valid, surrogates must correlate well with improvements in important clinical endpoints. One example of such a surrogate is brain natriuretic peptide, a neurohormone that predicts left ventricular function and prognosis and that has also become a diagnostic test [60]. While there is disagree-

ment about which surrogates are useful in CHF [61], the patient exposure to experimental therapy and the cost required by phase III studies dictate that an effort be made to use phase II studies, and surrogate markers can serve a useful role. Phase III endpoints must be more clinically relevant, in part because surrogate endpoints are not entirely reliable. In CHF, therefore, mortality is still a preferred measure of efficacy, although hospitalization rates and other secondary measures may be considered [62].

Endpoints once deemed of limited clinical value may gain importance through greater experience. Improvements in disease-free survival, an endpoint less concrete than overall survival, have historically not been regarded as sufficient to merit a change in clinical practice in many areas of oncology. More recently, analysis of accumulated studies has suggested that 3-year disease-free survival is an accurate surrogate of 5-year survival when administering adjuvant chemotherapy to patients who have had curative surgery for colon cancer [63]. The use of oxaliplatin in the adjuvant colon cancer setting was approved by the FDA on the basis of a disease-free survival benefit, and there is the potential to use such surrogates to shorten drug development time [64].

## 5 EXAMPLE IN DRUG DEVELOPMENT

To further understand the clinical trial process, it is useful to consider an example. The field of oncology has seen an increase in the number of experimental agents directed at specific disease mechanisms. These targeted drugs are sometimes considered to have the ability to prevent tumor growth while not actually causing tumor shrinkage (tumor response), and may be termed cytostatic agents. Typically, new drugs are first studied in patients with advanced, metastatic disease, and tumor response has been employed as a surrogate for clinically important endpoints such as survival. The challenge in studying cytostatic drugs is that they may not induce tumor response and may be less effective in patients with greater burdens of disease. Hence, useful drugs may be missed if tumor response is relied upon to demonstrate activity [65].

Such were the considerations during the development of marimastat, a matrix metalloproteinase inhibitor. Matrix metalloproteinases are a family of proteins that degrade extracellular matrix and thus facilitate the migration and metastasis of tumor cells and facilitate vascular growth. Preclinical work suggested marimastat inhibited this process [66]. Except for the first study, performed in healthy volunteers [67], phase I studies suggested a dose-limiting arthritis [68, 69]. These studies indicated doses for further work and suggested that achievable plasma levels were likely sufficient to achieve target inhibition.

Few single-agent phase II studies were performed, and tumor responses were rare [70–72]. With the understanding that marimastat might not show typical responses in tumors, a large study was performed with various tumor types to assess a surrogate endpoint, a change in tumor markers [73]. With the exception of prostate-specific antigen, the tumor markers that were used are not sufficiently associated with clinical endpoints that they are usually accepted as surrogates [65]. While an impact on tumor markers was suggested by this and another study [74], there was no clear evidence of improvement in any clinical endpoint.

Acknowledging the difficulty in detecting activity in metastatic disease, Miller et al. conducted a randomized phase II study in the adjuvant breast cancer setting [75]. This trial encountered musculoskeletal toxicity that prevented drug administration from being sufficiently sustained to warrant further adjuvant study.

Phase II data could thus be regarded as tenuous, but optimism was such that phase III drug development proceeded. In fact, for both the lung cancer and gastric cancer trials, there was no phase II data to support phase III efforts [76, 77]; the study in gastric cancer was based in part on pathological changes noted in a phase I trial [78]. The results of phase III studies were almost universally disappointing [77, 79–81], although minimal activity was seen in gastric cancer [76]. Development of the drug ceased.

It is unfair to be overly critical of the participants in such a story, but certain issues may be usefully considered. First, phase I studies may demonstrate some aspects of a drug's toxicity, but only with more patients and longer term follow-up will toxicity become clear. This became more evident in the phase II study in the adjuvant breast cancer setting, and flushing out the toxicity profile is another argument for phase II studies beyond looking for initial clinical activity. A resource-intensive phase III study would likely have been aborted in the same adjuvant situation. Second, surrogate markers can be misleading [60, 82, 83]. To be considered true surrogate markers, they must be biologically relevant, show a consistent and proportional relationship between a change in the marker and a clinically meaningful endpoint, and this relationship should be demonstrable in repeated studies [60]. Most tumor markers do not satisfy these requirements, and thus their use was probably not justified. That said, even markers directly in the biological pathway of a drug are not a guarantee of adequate surrogacy, as redundant and alternative molecular pathways may dilute or eliminate the relationship of the surrogate to a clinical endpoint. Unfortunately, an adequate biological surrogate test had not been established for marimastat. Proceeding to phase III studies based on uncertain surrogate markers was thus a gamble.

How does one decide when to carry out phase III studies in oncology for cytostatic drugs? This is still an evolving field. In terms of using clinical outcomes, the use of stable disease is being used by default, although there is modest evidence of a relationship between this and the more concrete endpoint of survival [84–88]. As response and even stable disease may be difficult to demonstrate in advanced malignancy, biomarkers are likely to remain relevant. Measuring direct effects on tumor is likely ideal, but many tumors are not readily accessible for repeat biopsy after treatment. In this instance, one might pursue changes in biomarkers in accessible tissue such as skin. There is still the hazard, however, that skin changes may not be representative of tumor changes. In either case, unless a similar drug has established a true surrogate relationship for the biomarker in question, investigators are left to establish the relationship, a very difficult task during the limited number of trials undertaken with a developing drug. In the absence of a validated surrogate or true clinical evidence of activity, the preclinical or clinical biological data must be compelling to proceed with large randomized studies. If it is, investigators might consider whether it is better to study the drug in the setting of earlier disease, perhaps in the adjuvant setting. While the benefit of a cytostatic agent may be more evident in this setting, larger treatment groups and longer follow-up are typically required to detect the small improvements in outcome often seen in early disease.

## REFERENCES


1. DiMasi, J. A., Hansen, R. W., and Grabowski, H. G. (2003), The price of innovation: New estimates of drug development costs, *J. Health Econ.*, 22, 151–185.
2. Roberts, T. G., Jr., Lynch, T. J., Jr., and Chabner, B. A. (2003), The phase III trial in the era of targeted therapy: Unraveling the "go or no go" decision, *J. Clin. Oncol.*, 21, 3683–3695.
3. Eisenhauer, E. A., O'Dwyer, P. J., Christian, M., and Humphrey, J. S. (2000), Phase I clinical trial design in cancer drug development, *J. Clin. Oncol.*, 18, 684–692.
4. Kuhlmann, J. (1997), Drug research: From the idea to the product, *Int. J. Clin. Pharmacol. Ther.*, 35, 541–552.
5. Konstam, M. A. (2005), Reliability of ventricular remodeling as a surrogate for use in conjunction with clinical outcomes in heart failure, *Am. J. Cardiol.*, 96, 867–871.
6. Narang, R., Swedberg, K., and Cleland, J. G. (1996), What is the ideal study design for evaluation of treatment for heart failure? Insights from trials assessing the effect of ACE inhibitors on exercise capacity, *Eur. Heart J.*, 17, 120–134.
7. Farrington, P., and Miller, E. (2003), Clinical trials, *Methods Mol. Med.*, 87, 335–352.
8. Simon, R., Wittes, R. E., and Ellenberg, S. S. (1985), Randomized phase II clinical trials, *Cancer Treat. Rep.*, 69, 1375–1381.
9. Freidlin, B., and Simon, R. (2005), Evaluation of randomized discontinuation design, *J. Clin. Oncol.*, 23, 5094–5098.
10. Simon, R. (1987), How large should a phase II trial of a new drug be? *Cancer Treat. Rep.*, 71, 1079–1085.
11. Spilker, B. (1991), Marketing-oriented clinical studies, in *Guide to Clinical Trials*, Raven Press, New York, pp. 367–369.
12. Spilker, B. (1991), Classification and description of phase IV postmarketing study designs, in *Guide to Clinical Trials*, Raven Press, New York, pp. 44–58.
13. Bombardier, C., Laine, L., Reicin, A., Shapiro, D., Burgos-Vargas, R., Davis, B., Day, R., Ferraz, M. B., Hawkey, C. J., Hochberg, M. C., Kvien, T. K., and Schnitzer, T. J. (2000), Comparison of upper gastrointestinal toxicity of rofecoxib and naproxen in patients with rheumatoid arthritis. VIGOR Study Group, *N. Engl. J Med.*, 343, 1520–1528.
14. Bresalier, R. S., Sandler, R. S., Quan, H., Bolognese, J. A., Oxenius, B., Horgan, K., Lines, C., Riddell, R., Morton, D., Lanas, A., Konstam, M. A., and Baron, J. A. (2005), Cardiovascular events associated with rofecoxib in a colorectal adenoma chemoprevention trial, *N. Engl. J. Med.*, 352, 1092–1102.
15. U.S. Department of Health and Humans Services F.a.D.A.C.f.D.E.a.R. Guidance for Industry—Allergic Rhinitis: Clinical Development Programs for Drug Products, http://www.fda.gov/cder/guidance/2718dft.pdf, 2000, accessed Nov. 10, 2005.
16. Bunn, P. A., Jr., and Franklin, W. (2002), Epidermal growth factor receptor expression, signal pathway, and inhibitors in non-small cell lung cancer, *Semin. Oncol.*, 29, 38–44.
17. U.S. Food and Drug Administration, FDA Public Health Advisory—New Labelling and Distribution Program for Gefinitib (Iressa), http://www.fda.gov/cder/drug/advisory/iressa.htm, 2005, accessed Nov. 10, 2005.
18. Lynch, T. J., Bell, D. W., Sordella, R., Gurubhagavatula, S., Okimoto, R. A., Brannigan, B. W., Harris, P. L., Haserlat, S. M., Supko, J. G., Haluska, F. G., Louis, D. N., Christiani, D. C., Settleman, J., and Haber, D. A. (2004), Activating mutations in the epidermal growth factor receptor underlying responsiveness of non-small-cell lung cancer to gefitinib, *N. Engl. J. Med.*, 350, 2129–2139.

19. Paez, J. G., Janne, P. A., Lee, J. C., Tracy, S., Greulich, H., Gabriel, S., Herman, P., Kaye, F. J., Lindeman, N., Boggon, T. J., Naoki, K., Sasaki, H., Fujii, Y., Eck, M. J., Sellers, W. R., Johnson, B. E., and Meyerson, M. (2004), EGFR mutations in lung cancer: Correlation with clinical response to gefitinib therapy, *Science*, 304, 1497–1500.
20. Waterston, R. H., Lindblad-Toh, K., Birney, E., Rogers, J., et al. (2002), Initial sequencing and comparative analysis of the mouse genome, *Nature*, 420, 520–562.
21. Collins, J. M., Zaharko, D. S., Dedrick, R. L., and Chabner, B. A. (1986), Potential roles for preclinical pharmacology in phase I clinical trials, *Cancer Treat. Rep.*, 70, 73–80.
22. Voskoglou-Nomikos, T., Pater, J. L., and Seymour, L. (2003), Clinical predictive value of the in vitro cell line, human xenograft, and mouse allograft preclinical cancer models, *Clin. Cancer Res.*, 9, 4227–4239.
23. U.S. Department of Agriculture A.a.P.I.S, The Animal Welfare Act, http://www.aphis.usda.gov/lpa/pubs/awact.html, 2002.
24. World Medical Association, World Medical Association Declaration of Helsinki, http://www.wma.net/e/policy/b3.htm#note1, 2004.
25. Agrawal, M., and Emanuel, E. J. (2003), Ethics of phase 1 oncology studies: Reexamining the arguments and data, *JAMA*, 290, 1075–1082.
26. Joffe, S., Cook, E., Clark, J., and Weeks, J. (2003), Altruism among participants in cancer clinical trials, *Proc. Am. Soc. Clin. Oncol.*, 22, 523.
27. Kong, W. M. (2005), Legitimate requests and indecent proposals: Matters of justice in the ethical assessment of phase I trials involving competent patients, *J. Med. Ethics*, 31, 205–208.
28. Von Hoff, D. D., and Turner, J. (1991), Response rates, duration of response, and dose response effects in phase I studies of antineoplastics, *Invest. New Drugs*, 9, 115–122.
29. Goss, P. E., Ingle, J. N., Martino, S., Robert, N. J., Muss, H. B., Piccart, M. J., Castiglione, M., Tu, D., Shepherd, L. E., Pritchard, K. I., Livingston, R. B., Davidson, N. E., Norton, L., Perez, E. A., Abrams, J. S., Therasse, P., Palmer, M. J., and Pater, J. L. (2003), A randomized trial of letrozole in postmenopausal women after five years of tamoxifen therapy for early-stage breast cancer, *N. Engl. J. Med.*, 349, 1793–1802.
30. Bryant, J., and Wolmark, N. (2003), Letrozole after tamoxifen for breast cancer—what is the price of success? *N. Engl. J. Med.*, 349, 1855–1857.
31. Carter, L. E., McNeil, D. W., Vowles, K. E., Sorrell, J. T., Turk, C. L., Ries, B. J., and Hopko, D. R. (2002), Effects of emotion on pain reports, tolerance and physiology. *Pain Res. Manag.*, 7, 21–30.
32. Rietveld, S. (1998), Symptom perception in asthma: A multidisciplinary review, *J. Asthma*, 35, 137–146.
33. Kreutz, G. (1999), European regulatory aspects on new medicines targeted at treatment of rheumatoid arthritis, *Ann. Rheum. Dis.*, 58(Suppl 1), I92–I95.
34. Bernhard, J., Cella, D. F., Coates, A. S., Fallowfield, L., Ganz, P. A., Moinpour, C. M., Mosconi, P., Osoba, D., Simes, J., and Hurny, C. (1998), Missing quality of life data in cancer clinical trials: Serious problems and challenges, *Stat. Med.*, 17, 517–532.
35. Johnson, J. R., Williams, G., and Pazdur, R. (2003), End points and United States Food and Drug Administration approval of oncology drugs, *J. Clin. Oncol.*, 21, 1404–1411.
36. Welling, P., Lasagna, L., and Banakar, U. (1996), *The Drug Development Process—Increasing Efficiency and Cost-Effectiveness*, Marcel Dekker, New York, pp. 317–351.
37. Moye, L. A., and Deswal, A. (2002), The fragility of cardiovascular clinical trial results, *J. Card. Fail.*, 8, 247–253.

38. Eisenhauer, E., Bonetti, M., and Gelber, R. (2005), Principles of clinical trials, in Cavalli, F., Hansen, H., and Kaye, S., Eds., *Textbook of Medical Oncology*, Martin Dunitz, London, pp. 99–136.
39. Silversides, A. (2004), The tribulations of community-based trials, *CMAJ*, 170, 33.
40. Bressler, N. M., Hawkins, B. S., Bressler, S. B., Miskala, P. H., and Marsh, M. J. (2004), Clinical trial performance of community- vs university-based practices in the submacular surgery trials (SST): SST report no. 2, *Arch. Ophthalmol.*, 122, 857–863.
41. Hjorth, M., Holmberg, E., Rodjer, S., Taube, A., and Westin, J. (1995), Patient accrual and quality of participation in a multicentre study on myeloma: A comparison between major and minor participating centres, *Br. J. Haematol.*, 91, 109–115.
42. Koretz, M. M., Jackson, P. M., Torti, F. M., and Carter, S. K. (1983), A comparison of the quality of participation of community affiliates and that of universities in the Northern California Oncology Group, *J. Clin. Oncol.*, 1, 640–644.
43. Begg, C. B., Carbone, P. P., Elson, P. J., and Zelen, M. (1982), Participation of community hospitals in clinical trials: Analysis of five years of experience in the Eastern Cooperative Oncology Group, *N. Engl. J. Med.*, 306, 1076–1080.
44. Layde, P. M., Broste, S. K., Desbiens, N., Follen, M., Lynn, J., Reding, D., and Vidaillet, H. (1996), Generalizability of clinical studies conducted at tertiary care medical centers: A population-based analysis, *J. Clin. Epidemiol.*, 49, 835–841.
45. Sharpe, N. (2002), Clinical trials and the real world: Selection bias and generalisability of trial results, *Cardiovasc. Drugs Ther.*, 16, 75–77.
46. Pearl, A., Wright, S., Gamble, G., Doughty, R., and Sharpe, N. (2003), Randomised trials in general practice—a New Zealand experience in recruitment, *N. Z. Med. J.*, 116, U681.
47. Donahue, D. C., Lewis, B. E., Ockene, I. S., and Saperia, G. (1996), Research collaboration between an HMO and an academic medical center: Lessons learned. *Acad. Med.*, 71, 126–132.
48. Keinonen, T., Keranen, T., Klaukka, T., Saano, V., Ylitalo, P., and Enlund, H. (2003), Investigator barriers and preferences to conduct clinical drug trials in Finland: A qualitative study, *Pharm. World Sci.*, 25, 251–259.
49. Montaner, J. S., O'Shaughnessy, M. V., and Schechter, M. T. (2001), Industry-sponsored clinical research: A double-edged sword, *Lancet*, 358, 1893–1895.
50. Emanuel, E. J., Currie, X. E., and Herman, A. (2005), Undue inducement in clinical research in developing countries: Is it a worry? *Lancet*, 366, 336–340.
51. Hayasaka, E. (2005), Approaches vary for clinical trials in developing countries, *J. Natl. Cancer Inst.*, 97, 1401–1403.
52. International Conference on Harmonisation Steering Committee, Guideline for Good Clinical Practice, http://www.ich.org/MediaServer.jser?@_ID=482&@_MODE=GLB, 1996, accessed 2005.
53. Yusuf, S., Mehta, S. R., Diaz, R., Paolasso, E., Pais, P., Xavier, D., Xie, C., Ahmed, R. J., Khazmi, K., Zhu, J., and Liu, L. (2004), Challenges in the conduct of large simple trials of important generic questions in resource-poor settings: The CREATE and ECLA trial program evaluating GIK (glucose, insulin and potassium) and low-molecular-weight heparin in acute myocardial infarction, *Am. Heart J.*, 148, 1068–1078.
54. American College of Rheumatology Subcommittee on Rheumatoid Arthritis. Guidelines for the management of rheumatoid arthritis: 2002 Update (2002), *Arthritis Rheum.*, 46, 328–346.
55. Strand, V. (2004), Counterpoint from the trenches: A pragmatic approach to therapeutic trials in rheumatoid arthritis, *Arthritis Rheum.*, 50, 1344–1347.

56. Arriagada, R., Bergman, B., Dunant, A., Le Chevalier, T., Pignon, J. P., and Vansteenkiste, J. (2004), Cisplatin-based adjuvant chemotherapy in patients with completely resected non-small-cell lung cancer, *N. Engl. J. Med.*, 350, 351–360.

57. Neoptolemos, J. P., Dunn, J. A., Stocken, D. D., Almond, J., Link, K., Beger, H., Bassi, C., Falconi, M., Pederzoli, P., Dervenis, C., Fernandez-Cruz, L., Lacaine, F., Pap, A., Spooner, D., Kerr, D. J., Friess, H., and Buchler, M. W. (2001), Adjuvant chemoradiotherapy and chemotherapy in resectable pancreatic cancer: A randomised controlled trial, *Lancet*, 358, 1576–1585.

58. Dent, S. F., and Eisenhauer, E. A. (1996), Phase I trial design: Are new methodologies being put into practice? *Ann. Oncol.*, 7, 561–566.

59. Massie, B. M. (2003), The dilemma of drug development for heart failure: When is the time to initiate large clinical trials? *J. Card. Fail.*, 9, 347–349.

60. Anand, I. S., Florea, V. G., and Fisher, L. (2002), Surrogate end points in heart failure, *J. Am. Coll. Cardiol.*, 39, 1414–1421.

61. DeMets, D. L. (2000), Design of phase II trials in congestive heart failure, *Am. Heart J.*, 139, S207–S210.

62. Committee for Medicinal Products for Human Use, Note for Guidance on Clinical Investigation of Medicinal Products for the Treatment of Cardiac Failure, Addendum on Acute Heart Failur, http://www.emea.eu.int/pdfs/human/ewp/298603en.pdf, 2004.

63. Sargent, D. J., Wieand, H. S., Haller, D. G., Gray, R., Benedetti, J. K., Buyse, M., Labianca, R., Seitz, J. F., O'Callaghan, C. J., Francini, G., Grothey, A., O'Connell, M., Catalano, P. J., Blanke, C. D., Kerr, D., Green, E., Wolmark, N., Andre, T., Goldberg, R. M., and De Gramont, A. (2005), Disease-free survival versus overall survival as a primary end point for adjuvant colon cancer studies: Individual patient data from 20,898 patients on 18 randomized trials, *J. Clin. Oncol.*, 23, 8664–8670.

64. Andre, T., Boni, C., Mounedji-Boudiaf, L., Navarro, M., Tabernero, J., Hickish, T., Topham, C., Zaninelli, M., Clingan, P., Bridgewater, J., Tabah-Fisch, I., and De Gramont, A. (2004), Oxaliplatin, fluorouracil, and leucovorin as adjuvant treatment for colon cancer, *N. Engl. J. Med.*, 350, 2343–2351.

65. Gelmon, K. A., Eisenhauer, E. A., Harris, A. L., Ratain, M. J., and Workman, P. (1999), Anticancer agents targeting signaling molecules and cancer cell environment: Challenges for drug development? *J. Natl. Cancer Inst.*, 91, 1281–1287.

66. Hidalgo, M., and Eckhardt, S. G. (2001), Development of matrix metalloproteinase inhibitors in cancer therapy, *J. Natl. Cancer Inst.*, 93, 178–193.

67. Millar, A. W., Brown, P. D., Moore, J., Galloway, W. A., Cornish, A. G., Lenehan, T. J., and Lynch, K. P. (1998), Results of single and repeat dose studies of the oral matrix metalloproteinase inhibitor marimastat in healthy male volunteers, *Br. J. Clin. Pharmacol.*, 45, 21–26.

68. Rosemurgy, A., Harris, J., Langleben, A., Casper, E., Goode, S., and Rasmussen, H. (1999), Marimastat in patients with advanced pancreatic cancer: A dose-finding study, *Am. J. Clin. Oncol.*, 22, 247–252.

69. Wojtowicz-Praga, S., Torri, J., Johnson, M., Steen, V., Marshall, J., Ness, E., Dickson, R., Sale, M., Rasmussen, H. S., Chiodo, T. A., and Hawkins, M. J. (1998), Phase I trial of Marimastat, a novel matrix metalloproteinase inhibitor, administered orally to patients with advanced lung cancer, *J. Clin. Oncol.*, 16, 2150–2156.

70. Evans, J. D., Stark, A., Johnson, C. D., Daniel, F., Carmichael, J., Buckels, J., Imrie, C. W., Brown, P., and Neoptolemos, J. P. (2001), A phase II trial of marimastat in advanced pancreatic cancer, *Br. J. Cancer*, 85, 1865–1870.

71. Quirt, I., Bodurth, A., Lohmann, R., Rusthoven, J., Belanger, K., Young, V., Wainman, N., Stewar, W., and Eisenhauer, E. (2002), Phase II study of marimastat (BB-2516) in malignant melanoma: A clinical and tumor biopsy study of the National Cancer Institute of Canada Clinical Trials Group, *Invest. New Drugs*, 20, 431–437.
72. Rosenbaum, E., Zahurak, M., Sinibaldi, V., Carducci, M. A., Pili, R., Laufer, M., DeWeese, T. L., and Eisenberger, M. A. (2005), Marimastat in the treatment of patients with biochemically relapsed prostate cancer: A prospective randomized, double-blind, phase I/II trial, *Clin. Cancer Res.*, 11, 4437–4443.
73. Nemunaitis, J., Poole, C., Primrose, J., Rosemurgy, A., Malfetano, J., Brown, P., Berrington, A., Cornish, A., Lynch, K., Rasmussen, H., Kerr, D., Cox, D., and Millar, A. (1998), Combined analysis of studies of the effects of the matrix metalloproteinase inhibitor marimastat on serum tumor markers in advanced cancer: Selection of a biologically active and tolerable dose for longer-term studies, *Clin. Cancer Res.*, 4, 1101–1109.
74. Primrose, J. N., Bleiberg, H., Daniel, F., Van, Belle, S., Mansi, J. L., Seymour, M., Johnson, P. W., Neoptolemos, J. P., Baillet, M., Barker, K., Berrington, A., Brown, P. D., Millar, A. W., and Lynch, K. P. (1999), Marimastat in recurrent colorectal cancer: Exploratory evaluation of biological activity by measurement of carcinoembryonic antigen, *Br. J. Cancer*, 79, 509–514.
75. Miller, K. D., Gradishar, W., Schuchter, L., Sparano, J. A., Cobleigh, M., Robert, N., Rasmussen, H., and Sledge, G. W. (2002), A randomized phase II pilot trial of adjuvant marimastat in patients with early-stage breast cancer, *Ann. Oncol*, 13, 1220–1224.
76. Bramhall, S. R., Hallissey, M. T., Whiting, J., Scholefield, J., Tierney, G., Stuart, R. C., Hawkins, R. E., McCulloch, P., Maughan, T., Brown, P. D., Baillet, M., and Fielding, J. W. (2002), Marimastat as maintenance therapy for patients with advanced gastric cancer: A randomised trial, *Br. J. Cancer*, 86, 1864–1870.
77. Shepherd, F. A., Giaccone, G., Seymour, L., Debruyne, C., Bezjak, A., Hirsh, V., Smylie, M., Rubin, S., Martins, H., Lamont, A., Krzakowski, M., Sadura, A., and Zee, B. (2002), Prospective, randomized, double-blind, placebo-controlled trial of marimastat after response to first-line chemotherapy in patients with small-cell lung cancer: A trial of the National Cancer Institute of Canada—Clinical Trials Group and the European Organization for Research and Treatment of Cancer, *J. Clin. Oncol.*, 20, 4434–4439.
78. Tierney, G. M., Griffin, N. R., Stuart, R. C., Kasem, H., Lynch, K. P., Lury, J. T., Brown, P. D., Millar, A. W., Steele, R. J., and Parsons, S. L. (1999), A pilot study of the safety and effects of the matrix metalloproteinase inhibitor marimastat in gastric cancer, *Eur. J. Cancer*, 35, 563–568.
79. Bramhall, S. R., Rosemurgy, A., Brown, P. D., Bowry, C., and Buckels, J. A. (2001), Marimastat as first-line therapy for patients with unresectable pancreatic cancer: A randomized trial, *J. Clin. Oncol.*, 19, 3447–3455.
80. Bramhall, S. R., Schulz, J., Nemunaitis, J., Brown, P. D., Baillet, M., and Buckels, J. A. (2002), A double-blind placebo-controlled, randomised study comparing gemcitabine and marimastat with gemcitabine and placebo as first line therapy in patients with advanced pancreatic cancer, *Br. J. Cancer*, 87, 161–167.
81. King, J., Zhao, J., Clingan, P., and Morris, D. (2003), Randomised double blind placebo control study of adjuvant treatment with the metalloproteinase inhibitor, Marimastat in patients with inoperable colorectal hepatic metastases: Significant survival advantage in patients with musculoskeletal side-effects, *Anticancer Res.*, 23, 639–645.
82. Thompson, D. F. (2002), Surrogate end points, skepticism, and the CAST study, *Ann. Pharmacother.*, 36, 170–171.
83. Stadler, W. M., and Ratain, M. J. (2000), Development of target-based antineoplastic agents, *Invest. New Drugs*, 18, 7–16.

84. Cesano, A., Lane, S. R., Poulin, R., Ross, G., and Fields, S. Z. (1999), Stabilization of disease as a useful predictor of survival following second-line chemotherapy in small cell lung cancer and ovarian cancer patients, *Int. J. Oncol.*, 15, 1233–1238.
85. Howell, A., Mackintosh, J., Jones, M., Redford, J., Wagstaff, J., and Sellwood, R. A. (1988), The definition of the "no change" category in patients treated with endocrine therapy and chemotherapy for advanced carcinoma of the breast, *Eur. J. Cancer Clin. Oncol.*, 24, 1567–1572.
86. Murray, N., Coppin, C., Coldman, A., Pater, J., and Rapp, E. (1994), Drug delivery analysis of the Canadian multicenter trial in non-small-cell lung cancer, *J. Clin. Oncol.*, 12, 2333–2339.
87. Rapp, E., Pater, J. L., Willan, A., Cormier, Y., Murray, N., Evans, W. K., Hodson, D. I., Clark, D. A., Feld, R., and Arnold, A. M. (1988), Chemotherapy can prolong survival in patients with advanced non-small-cell lung cancer—report of a Canadian multicenter randomized trial, *J. Clin. Oncol.*, 6, 633–641.
88. Sargent, D. J., Wieand, H. S., Haller, D. G., Gray, R., Benedetti, J. K., Buyse, M., Labianca, R., Seitz, J. F., O'Callaghan, C. J., Francini, G., Grothey, A., O'Connell, M., Catalano, P. J., Blanke, C. D., Kerr, D., Green, E., Wolmark, N., Andre, T., Goldberg, R. M., and De Gramont, A. (2005), Disease-free survival versus overall survival as a primary end point for adjuvant colon cancer studies: Individual patient data from 20,898 patients on 18 randomized trials, *J. Clin. Oncol.*, 23, 8664–8670.

# *Regulatory Requirements for Investigational New Drug*

**Venkat Rao**
*National and Defense Programs, Defense Division, Alexandria, Virginia*

**Contents**

1 Introduction
2 Investigational New Drug Application Process
  1 Roadmap for Future IND Product Development
3 GLP Regulations in Nonclinical Investigations
4 Investigational New Drug cGMP Compliance Requirements
  4.1 cGMP for IND Phase I Clinical Trials
  4.2 cGMP for IND
  4.3 From cGMP to Quality Systems
5 Role of Orphan Drug Act in Investigational New Drug
6 Regulatory Requirements to Protect Human Subjects
7 Requirements for Oversight: IRB
  7.1 Composition of IRB
8 Requirements of Financial Disclosure
  8.1 Covered Clinical Studies
  8.2 Certification and Disclosure of Requirements
  8.3 Disclosure Statement Evaluation
9 Requirements for Good Tissue Practice Compliance
10 Requirements for IND Labeling

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

11 Monitoring of Investigational New Drug Research
11.1 Clinical Risk Assessment
11.2 Computerized Systems in Clinical Trials
11.3 Quality Assurance
12 Emerging Biosafety and Biosecurity Requirements
13 Conclusions
Appendix: Applicable and Relevant Regulations Covering IND
References
Bibliography

## 1 INTRODUCTION

Investigational new drug (IND) is a key phase within the drug development life cycle that requires considerable interactions between pharmaceutical companies and the Food and Drug Administration (FDA), the federal agency with principal responsibility to license new medical products for human use. As the IND approval authority it is the FDA's responsibility to protect public health by ensuring the safety, efficacy, and security of human and veterinary drugs and biological products. At the same time, the FDA's mission is to enable rapid advancement of pharmacological therapeutics development through technological innovations that make medicines more effective, safer, and affordable.

In the context of the IND process, these dual missions may seem contradictory and require a carefully balanced consideration of the risk–benefit of new medical products. For on the one hand, IND is the mechanism for new innovations in pharmacological therapeutics to meet the public health challenges; it is also the venue for introducing new risks with potentially devastating impact on public health and general environment. With the increasing number of new product development projects involving genetic engineering and recombinant DNA (deoxyribonucleic acid) technology in biological product development, the scope and volume of cell-based new biologics product development have expanded dramatically in the past two decades. For example, a new class of cell-based recombinant technology products generally known as the human cells, tissues, and cellular and tissue-based products (HCT/Ps) have created an entire array of investigational biologics. Similarly, new biodefense-oriented medical countermeasures such as vaccines, immunoglobulins, and monoclonal antibodies to protect and counter bioterrorism-related threats have potentials to introduce long-term human health and ecological risks.

Therefore, IND review and approval process will have to take into consideration not only the inherent benefits of introducing novel technological solutions for medical countermeasures during clinical trial, but consider the potential adverse impact to clinical trial subjects and the long-term public health and environmental consequences.

Nevertheless, medical product development is never complete without safety and efficacy data collected directly from studies on human subjects, which is why every drug or biologics product developer is bound to submit an IND application. This application process requires considerable forethought and preparation to ensure that the product under development is suitable for studies with human subjects.

Clinical investigators participating in the clinical studies and the study sponsors must document every facet of the study planning, data generation, and data management as long as the IND is in effect.

Regulatory affairs covering new medical product development are influenced by external factors driven by industry interests and new legislations promulgated by Congress aimed at improving the quality and safety of pharmaceuticals to protect and promote public health. Proactive efforts of the FDA in international forums and multilateral engagements with entities such as the International Conference on Harmonisation (ICH) and the Organization for Economic Cooperation and Development (OECD) brings scientific and technical discussions and harmonization of quality systems and safety guidelines for medical products registration.

In the past three decades a couple of major external events posed unique challenges to the drug development policies in general and IND-related regulatory affairs in particular.

First, the advent of AIDS (acquired immunodeficiency syndrome) in the 1980s drew unprecedented attention to the entire drug development and approval process. Changes in the new drug approval process was aimed at broadening involvement of patient community during the early phase of drug development activities and an overall pressure on the government agency to expedite the new drug approval process. Enormous efforts were made by those affected by the epidemic to gain "expanded access" to investigational drugs even before FDA granted a formal approval.

There simply was no mechanism in place within the regulatory affairs environment to circumvent the key milestones required in the new drug approval process and hesitation in the scientific circles to provide expanded access to highly toxic investigational drugs with potentially dangerous adverse effects. Combined efforts from the regulatory affairs and scientific community created the new National Institutes for Allergy and Infectious Diseases (NIAID) and funded the AIDS treatment research initiative known as Community Programs for Clinical Research on AIDS (CPCRA). A network of clinical study centers were created under the CPCRA to provide HIV-infected community access to IND products through participation in clinical trials. Expanded access to AIDS-related investigational drugs were made possible through IND treatment and parallel track protocols.

The complex nature of the HIV epidemic and highly toxic properties of some IND hindered broader access to possible pharmacological interventions for the affected community. However, the FDA finally issued regulations that broadly interpret use of IND for therapeutic purposes, which provided access to thousands of critically ill patients drugs not yet formally licensed for commercial use. NIAID created a "parallel track" program where selected INDs were made available to HIV-infected patients who could not participate in the clinical trial and have no other clinical alternatives. The parallel track policy was invoked to access IND for broader therapeutic uses, when available clinical evidence is less than adequate to apply the treatment IND clause to broaden access to investigational drugs for therapeutic uses.

Second, post-9/11 challenges to develop medical countermeasures to manage the threat of bioterrorism introduced unexpected demands on the new drug development regulations. The global threat environment called for expedited development of medical countermeasures such as vaccines, antidotes, and diagnostics to prevent

and protect the public against biological warfare agents. With very limited production capacity for vaccine products in the biotechnology industry community, and an assortment of experimental candidate products at very early stages in development, challenges were daunting at the starting point of this major national undertaking. Next, federal regulators and the biodefense product development community were faced with a major technological hurdle by way of potency testing for protective effects of experimental vaccines requiring live-agent tests, which completely eliminated the option of clinical trials for product efficacy determination. With clinical data only on the safety of the product, for most part regulatory decision to grant new biological licensure application would be based on efficacy data from experimental animal studies.

In 2002, the FDA published, a landmark "animal efficacy rule" amending the drug and biologics product regulations to allow appropriate studies in animals in cases where excessive toxicity would make human efficacy studies not ethical or feasible [1]. For example, the FDA approved ciprofloxacin for postexposure management of inhalational anthrax based on serum concentrations of the drug in human studies as a surrogate measure to serum concentrations associated with survival of animals exposed to aerosolized *Bacillus anthracis* spores. This data together with related data on the efficacy of ciprofloxacin in humans on other infections was the basis for the approval.

Use of the anthrax vaccine as an investigational new drug by the U.S. Department of Defense is yet another area where regulatory affairs relating to IND came head on with the national priority to rapidly develop and deploy medical countermeasures to protect military personnel and civilians from biological warfare agents. Controversies surrounding the military management of the unprecedented waiver to the FDA requirements of informed consent process and the role of institutional review board (IRB), the institutional level oversight authority over the conduct of the anthrax clinical trial, are yet to be resolved. As there are no published clinical studies on the efficacy or long-term safety of the anthrax vaccine, the protective value of the vaccine in humans is unknown.

With the best of intentions to protect the military and the public from the real threat of biological weapons, the wrenching decision by the military authorities to vaccinate uniformed personnel with an IND vaccine product through an unprecedented waiver of informed consent put the government in unchartered territory on the regulatory policy relating to IND development.

The following sections in this chapter will examine in detail the investigational new drug application process; the prevailing regulations relating to various aspects of IND investigation; latest requirements under good tissue practice for highly sophisticated biotechnology-derived products; various mechanisms, tools, and performance matrix to monitor IND clinical investigations; and emerging biosafety and biosecurity considerations in IND clinical investigations.

## 2 INVESTIGATIONAL NEW DRUG APPLICATION PROCESS

The IND is an embedded phase within the long drug development life cycle that could run anywhere from 10 to 13 years depending on the candidate product and study requirements. During the preclinical experimental investigations, a candidate

drug product with promising pharmacological potentials against a target disease or a physiological condition, and did not cause unacceptable damage to healthy tissue, is eligible to move into the IND phase. Through the IND process, the study sponsor requests permission from the FDA to begin clinical trials.

The IND is also a mechanism through which the pharmaceutical industry sponsor will obtain exemption to transport the IND product across state lines. Under the current federal law only licensed drugs are permitted for broader distribution and transportation across state lines. Thus, IND allows transport of investigational drug products for expanded clinical trials involving geographically dispersed multiple study locations.

The FDA grants three types of INDs: (a) investigator-initiated IND, (b) emergency use of IND, and (c) treatment IND. These INDs are granted either for commercial uses or for research use categories. A commercial uses category basically refers to an application from a pharmaceutical industry sponsor involved in medical product development for commerce. Whereas the research use category covers clinical investigations performed with the academic objective to better understand the clinical pathologies and effectiveness of medical countermeasure strategies. Research-oriented clinical investigations may involve either experimental products or an approved drug for new medical interventions.

*Investigator-Initiated IND* The application is submitted by a clinical investigator who is the study sponsor as well as the principal clinical investigator. An investigator-initiated IND may not have a commercial intent and primarily to investigate an unapproved medical countermeasure, or an approved product for a new indication not covered under the approved label. Investigators are required to file an IND application if the planned study with an approved drug involves a new patient population, which was not the basis in the clinical studies used in the earlier new drug licensure application.

*Emergency Use of IND* During a public health emergency the need for an experimental product may arise, but without sufficient time to complete the formal IND application process. Under such extraordinary circumstances the regulatory agency may authorize shipment of an investigational product for a specified use in advance of an IND. A public health official should make a request to the FDA under emergency situations to obtain the authorization for emergency uses.

*Treatment IND* The FDA issues a treatment IND for interim broader clinical uses if the experimental product shows promise in clinical trials for serious public health problems or immediately life-threatening conditions while the final clinical work is still under progress and the FDA review underway. For example, the 1980s AIDS epidemic introduced treatment INDs as part of a long-term effort to incorporate the concept of expanded access into the IND regulations. With very limited therapeutic options available to control and manage an expanding epidemic, use of investigational products for therapeutic purposes was made possible through this regulatory mechanism. In 2006, the FDA proposed new rules allowing seriously ill patients, and those with HIV/AIDS with limited or no treatment options, easier access to unapproved investigational products.

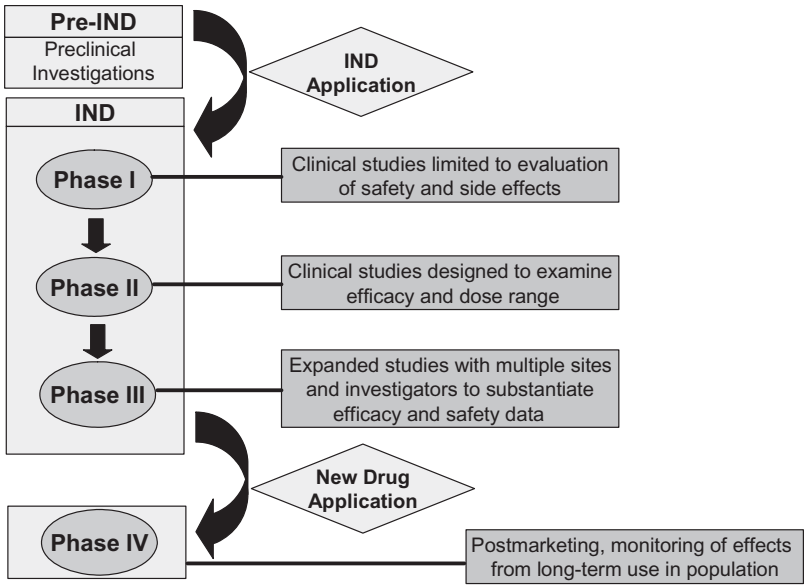


***EXHIBIT 1*** *Key development milestones in the investigational new drug development life cycle. Phase IV generally refers to a postmarketing surveillance of a newly approved medical product.*

Exhibit 1 illustrates the key milestones in the IND phase of the drug development life cycle. At the conclusion of pharmacological and toxicological investigations on candidate experimental products during pre-IND phase, the drug sponsor makes a key decision to proceed with the IND application for those candidates with promising pharmacological data on the efficacy and acceptable toxicological data on safety. Hence, the IND application must have a sound rationale and supportive experimental data covering the following three broad areas:

*Pharmacology and Toxicological Data* Data from studies on experimental animals or similar model systems allow an assessment on the safety of the product to begin testing in human subjects and promising pharmacological effectiveness together with the underling biochemical and molecular mechanism of action. This set of experimental data is collectively known as the pharmacokinetics and pharmacodynamics (PK/PD) and serves as a key benchmark on the bioavailability and mode of action.

*Clinical Protocol and Investigator Information* IND application must have a detailed protocol of the proposed clinical studies and risks to human subjects, if any, during the course of the clinical investigation. This section should also include sufficient information on the educational background and technical qualifications of the clinical investigators, who are generally practicing clinicians, and other medical and scientific professional with oversight responsibility. Also, included under this section is a certified undertaking by the principal investigator to obtain informed consent from all human subjects participating in the clinical trial and approved by the institutional review board (IRB) specially created to oversee the conduct of a proposed clinical trial.

*Product Manufacture Information* This section will provide a summary of the product composition, chemistry, formulation information, product stability

data, and manufacturer-related information, controls used for manufacture, and product quality assurance.

Phase I clinical studies focus on clinical pharmacology and short-term tolerance tests involving only a small number (not more than 50) healthy volunteers. The primary goal of phase I is the safety of the IND and a determination of dose ranges and appropriate routes of administration. Depending on the study design, phase I may involve limited pharmacokinetic analysis. Although in most cases phase I studies are performed on healthy individuals, there are instances patients are enrolled as human subjects during this phase when clinical investigation involves highly toxic products or targeted against life-threatening diseases such as AIDS or cancer.

Phase II clinical trials are designed to examine efficacy and refine the dose range from previous investigation [2]. These trials are longer in duration, taking 2 or more years, compared to phase I and involve a larger pool of human subjects and may involve more than one study location. Most phase II study designs are randomized, case-controlled investigations, where a group of patients receiving the IND drug, the "treatment group," is compared with a matched group of patients with comparable clinical profiles, case history, and factors such as age, sex, and other demographic background, the "control group," receives a placebo or standard therapy. Most phase II studies are double blinded by design in that both the patients and clinical investigators administering the study do not know the composition of the treatment and control groups and who is getting the IND product under study. The randomized, case-controlled study design greatly improves the validity of the clinical outcome data, and the double-blind design reduces errors in the study interpretation or other forms of bias.

Phase III trials are expanded investigations involving an even larger pool of human subjects running into thousands and involving multiple study locations and clinical investigators to address differences in responses due to demographic and geographic factors. Phase III trials are designed to further substantiate observations on safety and efficacy from previous clinical investigations, and a larger pool of human subjects allow the appearance of potentially less frequent adverse events not captured in smaller study populations.

At the conclusion of the phase III trial phase, the study sponsor would file applications with the FDA to obtain commercial licensure for the IND. A new drug application (NDA) is filed for a drug candidate, and for a biologics product this application is called biological licensure application (BLA). A product with an NDA/BLA is no longer an IND and is now commercially available for the public. According to industry reports, about 20% of the IND candidates make it through all phases of the research and are finally approved.

Phase IV could be either a postmarketing study to obtain additional information on the new drug in terms of rare adverse events, additional benefits, and optimal usage not captured during the clinical trials. Given a much broader population now exposed to the new medical product, early phase monitoring in the postmarketing period is generally considered phase IV. This phase is unique in that clinical investigation up until this point is a controlled study with a "treatment" group compared with a matched "control" group. Phase IV has no control group but merely tracks reports for adverse events or other product-related performance data. A phase IV data collection requirement is not mandatory for every approved medical product.

Depending on the potentials for adverse events from controlled clinical studies, the FDA may direct a sponsor to perform postmarketing surveillance for known and unexpected adverse events. Alternatively, a sponsor may proactively implement a phase IV plan to collect data on the potential benefits from the new drug to capture or increase market share for the new medical product.

Phase IV is sometimes differentiated from postmarketing surveillance when this phase actually involves a more structured design and clinical intervention within the terms of the product license. This differs from a postmarketing surveillance, which is mostly noninterventional and observational of the general population using a newly approved medical product. An interventional phase IV is carried out at hospitals or clinics to further consolidate the efficacy database of the newly approved product. This could be part of a risk management plan by the sponsor to obtain additional efficacy and safety data before general release of a new product.

The totality of findings from a well-designed, multiple clinical studies using targeted patient population with a demographic makeup similar to the general population forms the scientific rationale and data analytics to support submission of an NDA or a BLA to the FDA. It could take as many as 3 years for the FDA to review the nonclinical and clinical trials of data submitted in support of an NDA/BLA package before making a final decision on approval.

As part of this protracted engagement between the study sponsors and the FDA during the IND phase, several meetings are held as milestones for key decision points.

Exhibit 2 illustrates a work flow diagram for the industry meetings with the FDA during the IND phase and description of the meeting objectives and outcomes.

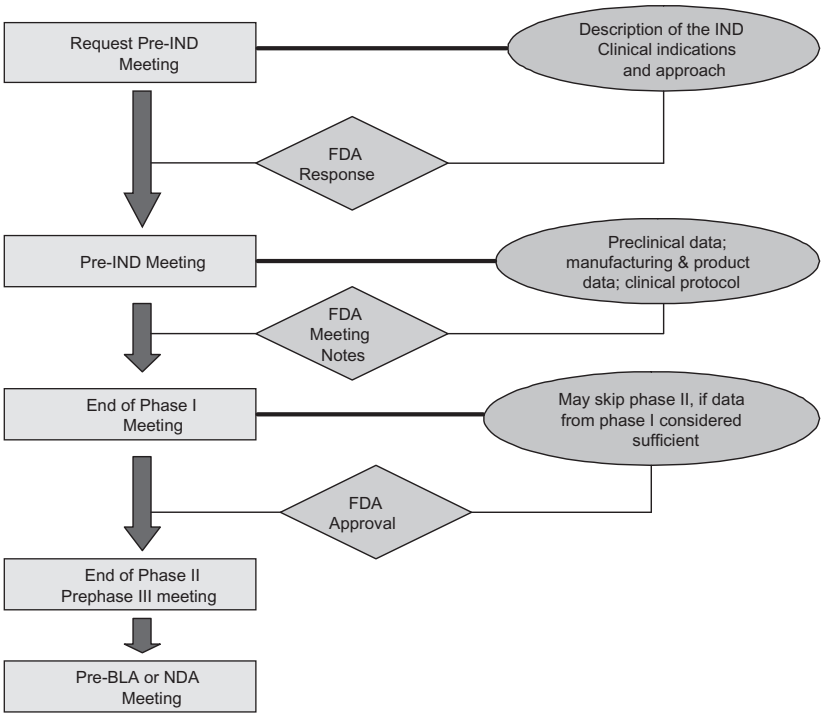


***EXHIBIT 2*** *Flow diagram of industry meeting with the Food and Drug Administration during various phases of the IND process.*

Product sponsor with plans to begin the IND phase must first submit a request for a pre-IND meeting with the FDA and include in that request a brief description of the experimental product, description of the clinical indication, and clinical study approach. Once the meeting date is established, the product sponsor must submit a pre-IND package (as required under 21 CFR 312.82) providing more detailed information on the preclinical experimental data, product manufacturing information, preliminary information on physiochemical characterization of the product and manufacturing specifications, and the outline of the proposed clinical protocol(s).

Exhibit 3 summarizes the general focus of the key IND meetings and the nature of information requirements for discussions. For example, during the pre-IND meeting, the general focus of the information requirements are chemistry and formulation-centric data, whereas the end of phase II (EOP-II) meeting covers study progress review, manufacturing, product stability, safety issues, process validation, and quality systems related information unique to drugs, biologics, and rDNA protein biotechnology drugs. The EOP-II meeting will review the efficacy study results, data gaps and deficiencies, and phase III study plans. Sponsors may update the FDA on potential problems identified and resolved, and information necessary in support of the marketing application.

The IND submissions are a complicated undertaking that should not only meet the technical content requirements, but also strictly comply with the content and format specifications (21 CFR 312.23).

Exhibit 4 is an example of the content requirements in a typical IND submission compliant with the regulatory requirements.

An investigator brochure must be included in the IND submission if the candidate product is supplied to clinical investigators who are not part of the study sponsor's organization. An investigator brochure together with the clinical trial protocol are the fundamental documents required in the IND submission. The investigator brochure must provide a description of the candidate product, a summary of the pharmacology and toxicology, summary information on safety and effectiveness, and a description of the risks of adverse events and recommended precautions or special monitoring.

Once the IND package is received from the sponsor-investigator, the FDA will assign an IND number. A regulatory project manager (RPM) for each submission handles all administrative matters related to processing the IND and serves as the regulatory point of contact for the study sponsor and/or investigator.

Exhibit 5 is a work flow diagram of the IND review and approval process. IND submissions could be made either as hardcopy document or electronic submission. An alternate mechanism to submitting IND is through the master file submission (21 CFR 314.420). A master file will contain product and manufacturing information but does not include clinical protocol. Permit holder for the master file could at a later period add the clinical protocol information when filing the IND. The master file submission format protects proprietary product and manufacturing information from other outside organizations participating in the clinical trial. Using a cross-reference filing format, authorized persons would then file relevant clinical information without access to other proprietary information. The FDA will access both IND and master file information to begin the review. The review process begins only when the IND files are populated in the master file.

Essentially, the IND review team for a drug or biologics product candidate consists of the (a) RPM, (b) product reviewer, (c) pharmacology/toxicology reviewer,

| IND Meetings | General Objectives | Drugs/Biotechnology Products/Conventional Biologics | |
|---|---|---|---|
| | | Safety | Dose/Formulation |
| **Pre-IND Meeting** | • Physical, chemical, and biological characteristics<br>• Manufacturers<br>• Sources and method of preparation<br>• Removal of toxic agents<br>• Quality controls<br>• Formulations<br>• Sterility<br>• Pharmacology<br>• Toxicology<br>• Stability | • Drugs from human sources<br>• Drugs from animal sources<br>• Biotechnology drugs<br>• Botanical drugs<br>• Reagents from animal or cell line sources | • Novel excipients<br>• Novel dosage forms<br>• Drug-device delivery systems |

| EOP-II | General Objectives | Drugs | rDNA Protein Biotechnology Drugs | Conventional Biologics[a] |
|---|---|---|---|---|
| **End-of-Phase II Meeting (EOP-II)** | • Physical, chemical, and biological characterization<br>• Process validation<br>• Environmental consideration<br>• Manufacturing considerations<br>• Facility-related issues<br>• Novel policy issues or concerns | Unique physicochemical and biological properties<br><br>Physicochemical characterization<br><br>Starting materiel designation<br><br>Qualification of impurities<br><br>Removal of adventitious agents<br><br>Approach to specifications<br><br>Sterilization process validation<br><br>Stability protocols<br><br>Environmental impact | Removal of adventitious agents<br><br>Approach to specifications<br><br>Sterilization process validation<br><br>Stability protocols<br><br>Environmental impact<br><br>Bioassay<br><br>Adequacy of cell bank characterization[b]<br><br>Removal of product and product-related impurities<br><br>Bioactivity of product-related substances | Removal of adventitious agents<br><br>Approach to specifications<br><br>Sterilization process validation<br><br>Stability protocols<br><br>Environmental impact<br><br>Coordination of facility design<br><br>Process validation consideration<br><br>Potency assay |

[a] Nonrecombinant vaccines and blood products.
[b] Would include, but not limited to, biochemical characterizations such as peptide map, amino acid sequence, disulfide linkages, higher order structure, glycosylation sites and structures, other posttranslational modifications and plans for completion, if still incomplete.

***EXHIBIT 3*** *Summary of the objectives of IND meetings and information requirements for drugs/biologics candidates.*

(d) clinical reviewer, and (e) statistical reviewer. During the first 30 days of the review period, the RPM will communicate with the study sponsor to obtain clarification on review comments from the IND review team and resolve any pending issues from the review team.

| General IND Submission Format and Content |
|---|
| Cover Sheet |
| Table of Contents |
| Introductory Statements and General Investigation Plan |
| Investigator's Brochure |
| Protocol for Each Planned Study |
| Chemistry, Manufacturing, and Control Information |
| Toxicology and Pharmacology Information |
| IRB Approved Consent Form |
| Previous Human Experience |
| Additional Information |

***EXHIBIT 4*** *Illustrative example of the information content requirements in a typical IND submission by sponsor of the medical product development.*

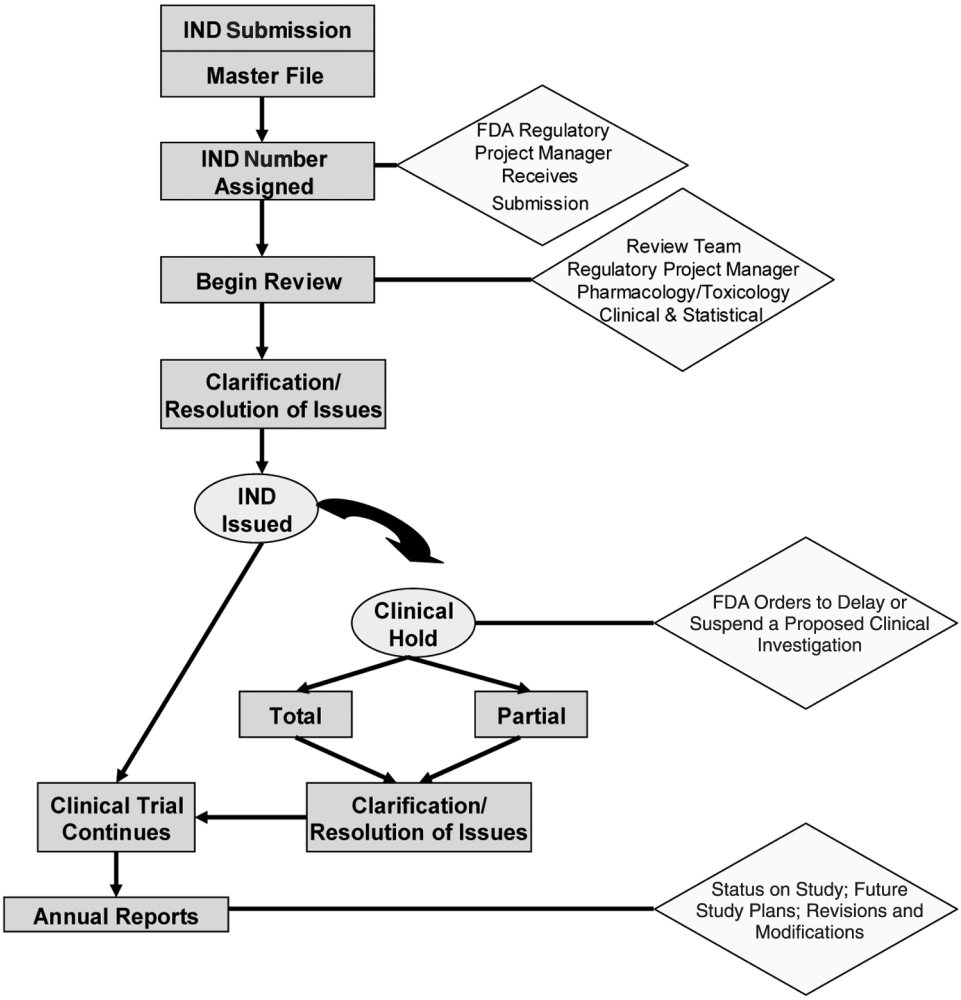


***EXHIBIT 5*** *Flow diagram of the IND review and approval process.*

Usually within 30 days after submission, the IND goes into effect and the decision is communicated via a letter as long as there are no holds placed on the IND.

During the IND review, regulators may decide to place a clinical hold. A clinical hold is an order issued by the FDA to delay a proposed clinical investigation or to suspend an ongoing investigation. When a complete clinical hold is placed on an IND, all clinical work comes to a halt. The FDA may place a partial clinical hold to delay or suspend part of the proposed clinical investigation, meaning that holds may be placed on some clinical protocols within the study while others are allowed to proceed without delay. The regulatory criteria guiding the FDA's decision to place a clinical hold may be one or a combination of risk factors such as: (a) human subjects would be exposed to unreasonable and significant risk; (b) the investigator brochure was misleading, or erroneous, or incomplete; (c) the clinical investigators are not qualified; or (d) IND does not contain sufficient information to assess the risks to subjects of the proposed study (21 CFR 312.42). A hold is placed on more advanced phase II/III trials if a proposed study design is deficient to meet the study objectives.

It is up to the study sponsor to provide the deficient information and clarifications to release the clinical hold placed on the trial. If the sponsor response addresses all the issues detailed in the clinical hold letter, the response is considered complete and the FDA is required to respond within 30 days of receipt of the submission. However, the 30-day clock does not apply to partial or incomplete responses to the clinical hold.

Clinical hold is a risk management mechanism available to the regulators to address potential deficiencies and safety-related issues during the IND phase of the product development. For example, the new generation of tissue-derived biologics products and vaccines pose considerable challenge in the areas of safety, quality, and product potency determination. To a large measure, clinical trials are designed to collecting data to addressing these sorts of issues during the product licensure application. As stated above, one of the more common reasons a clinical hold may be placed is when the FDA cannot determine from an IND clinical protocol the risks to human subjects during clinical investigations.

One of the alternative options to circumvent clinical trial requirements and facilitate product development in areas related to national security such as bioterrorism threat, or a major public health emergency such as pandemic influenza, is to approve medical treatment based on efficacy data from experimental animals. The regulation, known as "animal rule," provides a mechanism for the FDA to expedite the approval process as long as the efficacy data from preclinical investigations on experimental animals are well-designed and data from the studies considered sufficient and adequate by the expert review committee. Only safety data from a phase I clinical trial is required as supportive evidence in making the final determination.

## 1 Roadmap for Future IND Product Development

There is a growing concern in the pharmaceutical industry, scientific community, and among regulators on the decrease in the number of IND applications submitted over the past several years. Regulatory burden is cited as a major concern by the industry, although other factors related to extremely expensive up-front investments

required in new drug/biological development projects and the unpredictability of the product development activities. The product development community is beginning to employ the power of computational tools through bioinformatics to develop predicative tools on safety, effectiveness, and scale-up and manufacture of candidate therapeutics products.

Recognizing the importance of technology and regulatory compliance alignment in the development of novel therapies, the FDA unveiled a strategic document in 2004, the Critical Path Initiatives [3], which was aimed at identifying existing challenges in pharmaceuticals-oriented technology domains having a direct impact on the discovery and development of novel medicinal products. As part of this effort the Critical Path Opportunities List was compiled grouping program areas of emphasis under six categories covering mostly new product development strategies and technologies, development of new product development, testing and evaluation methods, and application of best business practices in product development. A brief summary of the six grouped critical path opportunities aimed at IND development activities are:

1. *Better evaluation tools such as biomarkers* for pregnancy, cardiovascular diseases, infectious diseases, cancer, neuropsychiatric diseases, and the like.
2. *Streamline clinical trials* in terms of innovative trial designs, application of best business practices in clinical data management, trial protocol development, and data analytics.
3. *Harnessing the power of bioinformatics* to identify new drugs and biologics candidates for development, identify safety biomarkers, adverse effects data mining, modeling and analysis, and failure analysis.
4. *Moving manufacturing toward twenty-first century* covering biotechnology-derived products, new generation vaccines, detection of contaminants in biologics during product development, scale-up and manufacturing, tissue engineering, product potency evaluation, and application of nanotechnology for therapeutics development.
5. *Development of products to meet public health needs* such as antimicrobial testing strategies, safety of blood and blood-derived products, and novel animal models in biodefense product development.
6. *Product development for at-risk population* with a specific focus on pediatrics. This would include better extrapolation of dose–response regimen in pediatric clinical trials, drug metabolism and therapeutic response, and new therapies for juvenile diabetes.

The Critical Path Initiative is unclear on market-driven factors critical to successful new drug development. First, biotechnology companies are involved in highly sophisticated genomic and proteomics-based product development and clinical gene therapy work in a domain with unclear regulatory implications and product liability issues due to potentials for long-term risks to public health. Commercially oriented research organizations are afraid of long-term liability lawsuits and have fewer incentives to take an active role in government-funded product development projects. This has to a large extent adversely reduced the available capacity to undertake a broad range of new product development projects.

Second, insurance coverage reform requires an equally forward leaning regulatory policy environment to encourage present shortage of providers of insurance to cover clinical trials. Whereas funded by the National Institutes of Health product development programs provide a modest coverage of the insurance cost incurred as part of the clinical trials, most of the cost is borne by the industry, hospitals, and academic institutions involved in medical product development related projects. At present, there are no defined mechanisms available within the existing regulations to the industry to directly address these issues with the regulatory agencies.

Finally, a roadmap for the future is incomplete if channeling of the R&D investment fails to reflect the requirements identified in the Critical Path Initiative. These industry investments in the R&D may not be consistent with the goals identified in the Critical Path Opportunities List. Industry investment in R&D priorities are defined and driven by market consideration. For example, considerable R&D resources are invested in protecting existing market shares, or to develop what are called "me too" drugs, with focus mostly in the formulary redesign and testing and not in breakthrough research in novel therapeutics development. Likewise, considerable product R&D and development is invested in derivative drugs taking the resources away from targets identified in the Critical Path Opportunities List.

Regulatory analysts are of the opinion that the national policy toward advanced therapeutics development requires a broader engagement of the R&D community, academia, and the industry, together with increased funding for processes development and assessments. Similarly, additional policy-level initiatives are required to simplify legal and financial barriers to clinical trials and additional resources to emerging genomics and proteomics-based product development.

## 3 GLP REGULATIONS IN NONCLINICAL INVESTIGATIONS

In the drug development process, preclinical investigation is a critical step during which drug molecules discovered from early phases of toxicological screening are subjected to a comprehensive animal testing before an IND can be filed. The regulatory requirements governing this process have imposed strict rules as to how a newly discovered chemical or biochemical molecule will be tested and evaluated prior to approval for testing in humans. The existing good laboratory practice (GLP) guidelines clearly lay out how preclinical experimentations is done in order to ensure the safety of the drug molecules, which then forms the basis for filing an IND application for approval to begin clinical trials (21 CFR 58).

Fundamentally, GLP is a quality system concerned with the organizational process and conditions under which nonclinical health and environmental safety studies are conducted. The organizational, personnel, and facility related GLP compliance requirements are written to ensure quality of data produced by nonclinical and clinical laboratories meet the best business practices and provide international acceptance to the quality of data generated in support of a new medical product licensure application.

The GLP regulations are part of the broad good laboratory practices for conducting preclinical experimental investigations in pharmaceutical product development for research or marketing permits. The scope of GLP regulations includes food and color additives, animal food additives, human and animal drugs, medical devices for

human use, biological products, and electronic products (biomedical devices). Compliance with GLP regulations is part of the requirements under the IND application for drugs and biologics (21 CFR312).

Existing GLP guidelines clearly define the parameters of preclinical experimental investigation to include *in vivo* and *in vitro* experiments in which test articles are studied prospectively in test systems under laboratory conditions to determine their safety. However, these tests do not include studies utilizing human subjects or clinical studies or field trials in animals. Also, preclinical experiments covered under this guideline does not include basic exploratory studies carried out to determine whether a test article has any potential utility or to determine physical or chemical characteristics of a test article (21 CFR58.3).

When product development companies outsource nonclinical studies to contract research facilities or academic research and development establishments, the GLP regulations require every entity participating in the nonclinical study included as part of the IND application and must comply with the provisions set forth under this regulation. As part of the enforcement mechanism the FDA may conduct facility inspection or authorize a designated third party with credentials to perform facility visit to ensure all laboratory records and study specimens related to the study are maintained and remain within the scope of the IND application under investigation.

Exhibit 6 schematically represents the key GLP requirements as part of preclinical investigations of an IND product, comprised of facility infrastructure components, business process, and personnel experience.

The GLP guidelines related to infrastructure cover experimental animal facility performing clinical research sites. Facility floor plans and material and process flows to attain operational isolation and prevent cross-contamination is one of the key criteria in the GLP compliance. Facility floor design and process flow plans are assisted by standard operating procedures to all study methods, operational-related

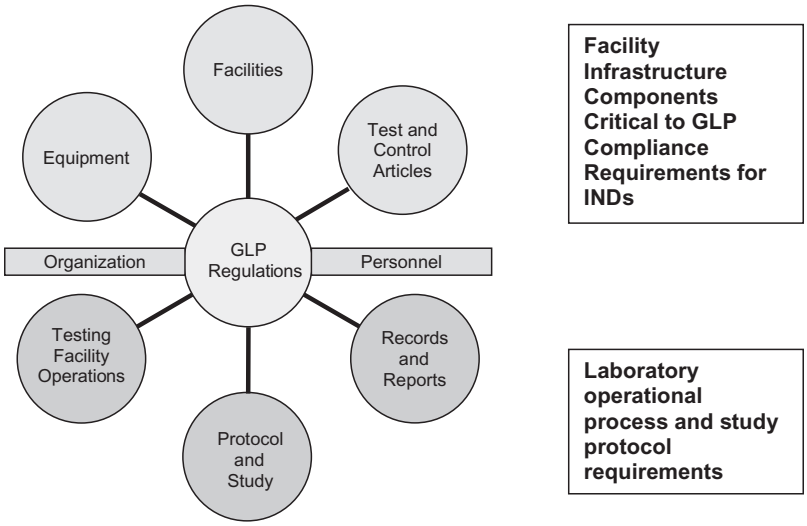


***EXHIBIT 6*** *Key Good Laboratory Practices requirements during the IND product development phase.*

issues such as cleaning and maintenance, equipment calibration and certifications, and routine inspections.

The GLP guidelines require that experimental product-related nonclinical studies and clinical investigations on the IND should have approved written protocols. These protocols should clearly outline the objectives of the study, proposed study methodology and details related to the test materials, procedures, analytical plans, and maintain supportive documentation.

The organization and personnel aspects relating to GLP cover strict reporting requirements for laboratory personnel and testing facility management in nonclinical laboratory studies. For example, individuals found at any time to have an illness that might adversely affect the quality of work and pose risk of infection to others in the facility must be reported immediately. The quality assurance unit within the facility has the responsibility to monitor studies and assure management that the facilities, equipment, personnel, methods, practice, records, and controls are in compliance with the GLP regulations. During an FDA inspection, the written procedures and internal quality assurance (QA) logs and records must be available for inspection.

A key international initiative aimed at instituting and harmonizing GLP is the Organization for Economic Cooperation and Development (OECD) guidelines, a multilateral organization with 30 member states and 70 other participating nations. The goal of OECD GLP guidelines is to ensure "data generated in the testing of chemicals in an OECD member country in accordance with OECD Test Guidelines and OECD principles of GLP are accepted in other member countries for purposes of assessment and other uses relating to the protection of man and the environment" [4].

The OECD GLP guidelines provide a framework for laboratories to plan, perform, monitor, record, report, and archive experimental laboratory studies. These guidelines would assist regulatory agencies within the member state, where the GLP-certified laboratory facility exists, inspect, and ensure compliance with the national GLP guidelines set in compliance with OECD. At the same time, data generated from a GLP-certified facility assures other international regulatory agencies within the OECD that the study results on pharmaceutical compounds could be relied upon as to the hazard and potential risks to users, consumers, and the general environment.

## 4 INVESTIGATIONAL NEW DRUG cGMP COMPLIANCE REQUIREMENTS

Although the U.S. federal government efforts to mandate the safety and purity of drugs goes as far back as 1902, when Congress decided to have biological products manufacturing facility licensed individually to protect the public from dangerously contaminated sera and vaccines, it was only in 1962 that the concept of "manufacturing controls" was introduced in the legislative statute, which was promulgated as "current good manufacturing practices" or simply cGMP.

cGMPs are essentially a family of systems consisting of policy procedures and written analytical documentation to guide a facility at the process levels on medical product manufacturing related activities. The goal of the cGMP is to ensure reli-

ability of a product manufactured at the facility through an established set of standards and processes for quality, purity, potency, composition, and identity claimed by the product sponsor. As a result, cGMP covers the entire gamut of the production systems, which includes plant and grounds, equipment and utensils, sanitation of building and facilities, quality assurance and quality control, production and process controls, warehousing, distribution and postdistribution process, and records access and archival system.

Compliance with cGMP is a fundamental requirement for a medical product development company whether involved in IND or routine manufacturing of licensed products. During the early years after cGMP promulgation, pharmaceutical industries experienced problems relating to potency, cross-contamination, sterility, and labeling related issues. As a result, the FDA initiated the Intensified Drug Inspection Program (IDIP) as an inspection mechanism to regulate the industry. If violations are found during an inspection, product(s) in the entire production line cannot be distributed until the industry demonstrates full compliance. For example, if during a routine GMP audit, some unknown particles were discovered in a production process, the plant will be ordered to temporarily shut down until the contaminants can be identified and removed from the production system. The facility management will perform a full-blown analysis involving sampling from the entire production lines, complete a battery of tests to identify the contaminants, and take measures to eliminate the problem. Only after a clear demonstration of these efforts would a resumption of normal production activity be allowed at the facility.

### 4.1 cGMP for IND During Phase I Clinical Trials

The 2006 FDA Guideline on the Preparation of IND Products (for human and animal uses) primarily addresses the regulatory compliance to cGMP regulations required under the Federal Food, Drug and Cosmetic Act (FD&C Act). These guidelines have no legally enforceable authority but are viewed as recommendations to address cGMP requirements in the production of INDs for phase I studies.

The earlier 1991 guideline addressed primarily the large-scale industrial manufacturing environment and not others such as small- or laboratory-level production of investigational new drugs. Also, the 1991 guideline did not clarify fully the FDA's programmatic expectation to adopt an incremental approach to institute manufacturing controls for the INDs. While addressing these issues, the 2006 guidelines represent the FDA's efforts to formally establish an approach guiding implementation of manufacturing controls in relation to IND products for phase I clinical trials. IND production settings covered in these guidelines include small-scale manufacturing at laboratory, batch productions for exploratory studies, and multiproduct and multibatch testing of IND products manufactured for phase I clinical investigations.

The 2006 FDA guidelines apply to all IND drug and biological products (including finished dosage forms used as placebos) for phase I clinical studies, which includes investigational recombinant and nonrecombinant therapeutic products, vaccines, allergenic products, *in vivo* diagnostics, plasma derivatives, blood and blood components, gene therapy products, and somatic cellular therapy products that are subject to cGMP requirements. However, these FDA guidlines do not apply

to (a) human cell or tissue products, (b) clinical trial of products subject to the device approval or clearance provisions within the existing regulations, (c) INDs manufactured for phase II/III clinical studies, and (d) approved products that are being used during phase I studies for other study endpoints such as route of exposure and new indications not covered in the approved label.

Exhibit 7 illustrates the key elements of a cGMP program for a facility involved in the manufacture of IND for clinical studies. The system components to establish

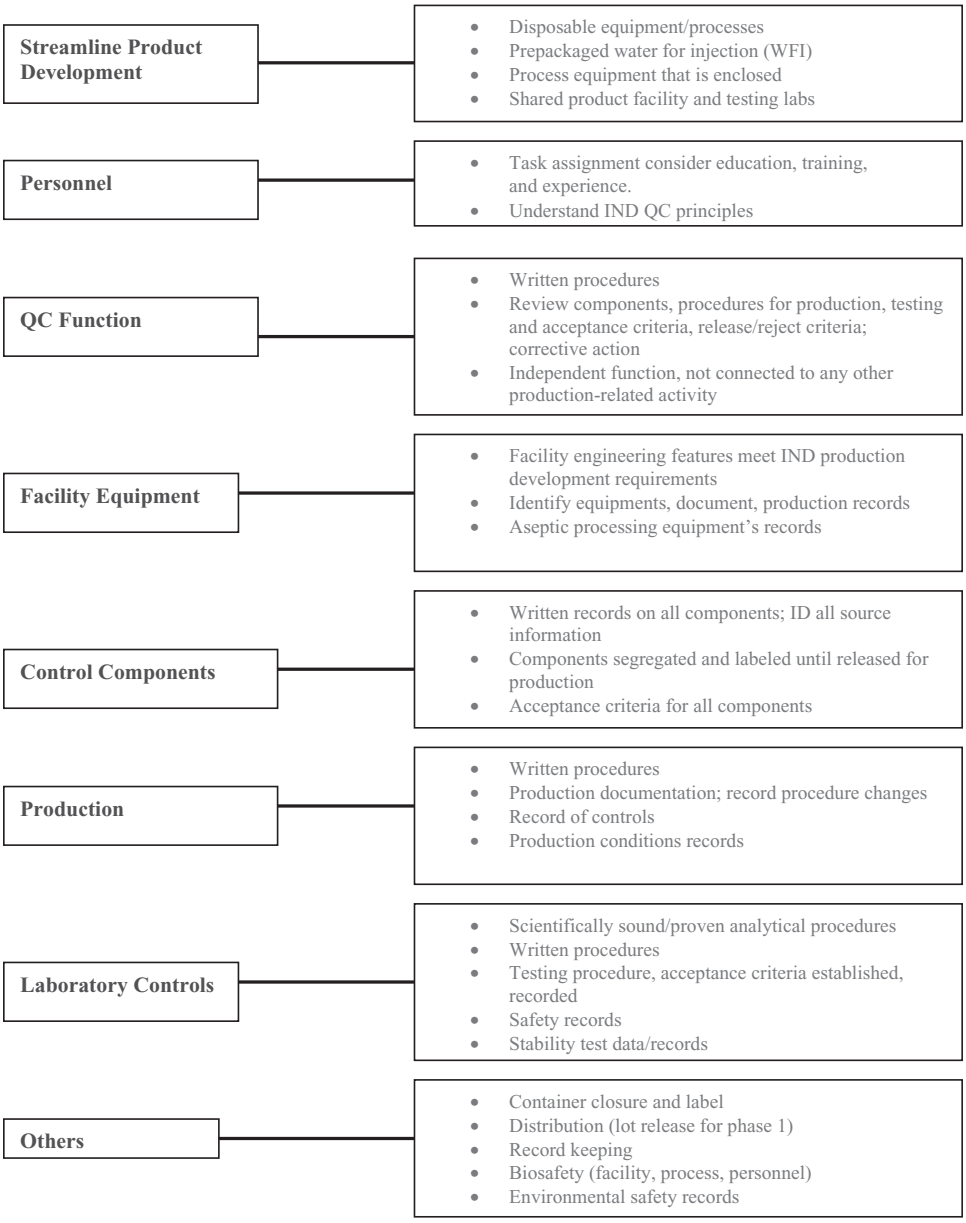


***EXHIBIT 7*** *FDA recommended approaches to complying with current good manufacturing practices during IND phase I investigations. (Based on FDA [5].)*

manufacturing controls for an IND are similar to routine cGMP programs at pharmaceutical facilities covering process elements, facility and personnel, production and laboratory controls, quality assurance, and control. In the case of investigational biological products requiring special precautions, biosafety and biosecurity-related issues are part of the overall facility and process-level compliance requirements.

With safety and quality of product as the primary focus, the guidelines for the production of IND for use in phase I clinical studies are centered on the establishment of cGMP-driven quality control process. The nature and extent of manufacturing controls needed to achieve the desired quality criteria differ not only between the IND products and commercial manufacture but also among the IND products manufactured for various phases of clinical studies. However, regulatory guidelines are yet to delineate cGMP requirements to these product development phases.

### 4.2 cGMP for IND

In its efforts intended to streamline the IND process, while at the same time ensure the safety and quality of drugs at the earliest stages in the development pipeline, the FDA excluded most of the phase I INDs from the cGMP regulations for human drugs, including biological products [5]. The FDA maintains regulatory oversight on the production of INDs under the general statutory cGMP authority and the requirements set forth under the IND application authority. The amendment of the cGMP regulation was considered as part of the FDA's overall efforts to guide the industry with a consistent framework to manage and establish controls on the manufacturing at the early product development stages. However, the FDA withdrew the final rule published in the 2006 *Federal Register* to evaluate comments received from the industry and published a modified rule in 2008. The final rule specifies that 21 CFR Part 211 no longer applies for IND, including those under the exploratory products category manufactured for use in phase I clinical trials.

The regulatory implications invoking the general statutory cGMP authority for IND means that overarching goals to ensure quality, purity, potency, and composition of the investigational product under clinical trial meet the general standards set forth under the cGMP. Hence, facilities manufacturing the IND for clinical studies must comply with all requirements that include the facility, equipment and utensils, sanitation of building and facilities, quality assurance and quality control, production and process controls, warehousing, distribution and postdistribution process, and records access and archival system.

### 4.3 From cGMP to Quality Systems

The concept of quality systems and its relevance to cGMP requirements were first identified under the FDA guidelines for finished medical devices intended for human use (21 CFR 820).

Exhibit 8 is a comparative summary of the quality systems applicable to cGMP requirements under the FDA guidelines and the International Organization for Standardization (ISO) 9001 requirements for medical devices intended for human use. The quality systems components under these guidelines cover organizational structure and control, management systems, quality audits, and personnel and training. The requirements for subcontractors involved in the product manufacturing

| cGMP Quality Systems Requirements[a] | ISO 9001 Requirements[b] |
|---|---|
| • Design plans to be reviewed, updated, and approved as the project evolves<br>• Design transfer control<br>• The effective dates of documents must be identified<br>• Label controls must be implemented<br>• Limitations or allowable tolerances must be available to personnel performing equipment adjustments<br>• Manufacturer maintenance of specific distribution records<br>• The manufacturer must follow specific evaluation, reporting, and recall procedures for defective products<br>• The manufacturer must mark confidential quality records<br>• Manufacturer maintenance of quality records for no less than 2 years from the date of release of the product for commercial distribution<br>• Manufacturer recording of specific data on records supporting the manufacture of product<br>• The development of a quality system record<br>• Manufacturer recording of specific data on servicing records | • Activities for consideration during quality planning<br>• Provision for contract review activities<br>• Design projects to be assigned to qualified personnel equipped with adequate resources<br>• Organizational and technical interfaces to be defined, documented, transmitted, and reviewed during the design process<br>• The verification of subcontracted products<br>• The control of customer-supplied products<br>• Control process for inspection, measuring, and test equipment |

[a]Based on 21CFR, Part 820; [b]ISO 9001 (2000).

***EXHIBIT 8*** *Comparative summary of the quality systems applicable to current good manufacturing practices within the FDA and ISO 9001 guidelines.*

activities are required to adhere to the quality assurance and quality control standards established by the sponsoring entity.

In 2004, the FDA introduced a transformational paradigm to cGMP through a quality systems-based approach for drugs and biologics products. The quality systems approach is based on the policy to encourage the pharmaceutical and biotechnology industries to adopt risk management principles. Regulatory shift to quality-systems-driven cGMP is in response to consistent problems with quality assurance and quality assurance operations in the medical products manufacturing sector. The FDA has declared quality systems approach as the mechanism to improve the predictability, consistency, integration, and overall effectiveness of its regulatory operation [3].

A risk-based approach to management controls, inspection, and oversight is one of the key pillars of the quality systems. Need for a risk-based regulatory approach was considered critical in the context of rapid advances in the process development and complex manufacturing processes for biotechnology-derived products. Process elements in these complex operations are at best ill defined, with limited manage-

ment controls and potentials for failures to meet quality assurance/quality control (QA/QC) targets. Other product-related quality systems considerations relate to safety issues such as contamination in the production systems, potentials for adverse human health, and environmental impact due to the biological agents and systems involved in the manufacture of specialized medical countermeasures such as vaccines against bioterrorism.

The risk-based approach essentially involves a scheme to prioritize the pharmaceutical facility inspection decision-making process. Through this prioritization scheme, the FDA will determine inspection of those facilities posing the greatest public health impact. The nature and extent of inspections at these facilities remains flexible and change with risk reduction strategies implemented at these facilities. A similar risk-based approach is used to assess product safety review as well. This includes product quality of INDs, preapproval chemistry, manufacturing controls, and postapproval supplement processes. The performance matrix to measure effectiveness of risk-based quality system implementation are (a) continuous improvement in product manufacturing, (b) increased product quality and process efficiency, and (c) availability of new medical products.

As part of the risk-based management initiative, the FDA established the Office of New Drug Chemistry (ONDC) within CDER, to establish a risk-based pharmaceutical quality assessment system with a focus on critical product quality attributes as it relates to safety and efficacy. Quality attributes such as product chemistry, formulation, manufacturing processes, and performance are targeted for process optimization and continued improvements.

Management controls are critical to addressing GMP issues such as the establishment of institutional-level policy and governance structures that communicate management intentions and priorities. Such a framework includes establishment of a systematic review of quality data trends on a regularly scheduled basis, resourcing plans and prioritization, and setting performance matrix and incentive systems. A risk-based decision framework is ideally suited to develop up front a decision scheme to prioritize these activities, set an internal monitoring to track status, and adjust the oversight requirement based on quality-systems-based performance matrix and internal audits. In particular, internal audits are a proactive mechanism to discover deviations from quality systems and correct deficiencies before a minor error escalates to a crisis level. Thus, management systems consistent with the risk-based framework for quality systems offers a win–win solution in terms of protecting the core business interests of the industry and the regulatory goals for cGMP compliance.

## 5 ROLE OF ORPHAN DRUG ACT IN INVESTIGATIONAL NEW DRUG

Orphan drugs belong to a FDA category designation for medical countermeasures intended for use in a rare disease or condition defined under Section 526 of the Food and Drug Act. Designation of an IND under the exclusive approval provisions of the orphan drug provides the manufacturer with treatment use of investigational orphan drugs. Orphan drug status also guarantees a 7-year period of exclusive marketing of the licensed drug. As of December 2008, the FDA has listed a total of 1951 pharmaceutical products under the orphan drug designation [6]. Orphan

drug products are expected to be used in treatment of over 6 million people annually in the United States.

In order for an IND product sponsor to avail the orphan drug status for the candidate product under development, sufficient documentary evidence should be prepared for submission demonstrating that:

1. The IND candidate is developed to treat a disease or condition for which the drug is intended to affect fewer than 200,000 people in the United States or, if the drug is a vaccine, diagnostic drug, or preventive drug, fewer than 200,000 dose-units are administered annually.
2. There is no reasonable expectation that the costs in research and development for the IND candidate will be recovered from sales of the IND drug even to more than 200,000 people, or dose-units, within the United States.

IND applicants seeking orphan drug provisions at the early investigational phase of the product development must write and obtain written recommendations from the FDA clarifying the nonclinical and clinical investigational data requirements to satisfy the regulatory provisions. In particular, the IND application must provide considerable details with authoritative references to description of the disease or condition for which the drug is proposed to be investigated and the proposed indication or indications for use for such disease conditions, and the basis for conclusion that the drug is for the disease or condition that is rare in the United States (21CFR 316.10).

As part of this application requirements, IND applicants must provide sufficient data backed up with analysis of available nonclinical and clinical data pertinent to the drug and the disease to be studied with supportive publications. This requirement is considered particularly advantageous for an IND product intended to treat a life-threatening or severely debilitating illness, especially when there are no other satisfactory alternative therapy available. IND orphan drug applicants meeting these requirements are expected to go through an expedited FDA review.

On the basis of background information submitted by the IND applicant seeking orphan drug status as part of the preclinical and clinical investigations, the FDA will determine (a) whether the disease or condition for which the drug is intended is rare, or not so, in the United States, and (b) whether there exists sufficient evidence and supportive rationale for permitting investigational use of the drug for the rare disease condition.

A product manufacturer could avail the orphan drug designation for a previously unapproved drug, as long as the supportive documents specify a rare disease or condition. Alternatively, a new orphan drug designation could be requested for a drug already in the market, as long as the new indications based on new research and development suggests its use in the treatment of a rare disease or condition. The current regulation allows even an approved drug, which does not have an orphan drug designation, to apply and receive such a status as long as the drug has indicated use like the orphan drug for the same rare disease or condition, but is shown to have superior clinical response or demonstrated safety by way of adverse drug reactions. More than one sponsor may receive orphan drug designation of the same drug for the same disease or condition, as long as each applicant files an independent request with the FDA seeking such a designation.

| IND Orphan "Same Drug" Category | "Same Drug" Regulatory Definition |
|---|---|
| **Small Drug Molecules** | An IND candidate that contains the same active moiety (small drug molecule) as a previously approved drug and is intended for the same use as the previously approved drug, even if the particular ester or salt (including a salt with hydrogen or coordination bonds) or other noncovalent derivative such as a complex, chelate, or clathrate has not been previously approved, except that if the subsequent drug can be shown to be clinically superior to the first drug, it will not be considered to be the same drug. |
| **Large Drug Molecule (Macromolecule)** An IND candidate that contains the same principal molecular structural features (but not necessarily all of the same structural features) and is intended for the same use as a previously approved drug, except that, if the subsequent drug can be shown to be clinically superior, it will not be considered to be the same drug. Criterion applied for various categories of macromolecules. | Two protein drugs would be considered the same if the only differences in structure between them were due to posttranslational events or infidelity of translation or transcription or were minor differences in amino acid sequence; other potentially important differences, such as different glycosylation patterns or different tertiary structures, would not cause the drugs to be considered different unless the differences were shown to be clinically superior. |
| | Two polysaccharide drugs would be considered the same if they had identical saccharide repeating units, even if the number of units were to vary and even if there were postpolymerization modifications, unless the subsequent drug could be shown to be clinically superior. |
| | Two polynucleotide drugs consisting of two or more distinct nucleotides would be considered the same if they had an identical sequence of purine and pyrimidine bases (or their derivatives) bound to an identical sugar backbone (ribose, deoxyribose, or modifications of these sugars), unless the subsequent drug were shown to be clinically superior. |
| | Closely related, complex partly definable drugs with similar therapeutic intent, such as two live viral vaccines for the same indication, would be considered the same unless the subsequent drug was shown to be clinically superior. |

***EXHIBIT 9*** *Summary of the definitions for IND orphan "same drug" categories for small and large drug molecules.*

Exhibit 9 summarizes the definitions for IND orphan "same drug" categories for small and large drug molecules. The decision criteria as to when an IND can be considered an orphan drug are driven by the inherent differences in the chemical composition of the moiety. For example, covalent or noncovalently modified derivatives of previously approved small drug molecules, which are mostly synthetic organic compounds, could be classified as an orphan drug as long as it is indicated for a rare disease or condition, or is shown to have a superior clinical response compared to the original compound.

The definition for what constitutes "same drug" is more complex for biologics, which are macromolecules such as proteins, complex polysaccharide drugs, and polynucleotide drugs. Here the differences are based not on the primary polymer sequence but on differences in the higher structure due to posttranslational modifications as in the case of proteins, or postpolymerization modification as in the case of polysaccharide- or polynucleotide-based drugs.

The provisions within the orphan drug approval process provide a market protection of 7 years during which time the FDA will not approve another sponsor's marketing application for the same drug. Once approved, the updated list of products with orphan drug status is published periodically in a report titled "Approved Drug Products with Therapeutic Equivalence Evaluations." The FDA also publishes a cumulative list of pharmaceutical products that have received orphan drug designation and a separate list of those that have received marketing approval. As of December 2008, there are a total of 1951 pharmaceutical products with orphan drug designation and 325 products with marketing approval.

## 6 REGULATORY REQUIREMENTS TO PROTECT HUMAN SUBJECTS

Protection of human subjects in a clinical trial is a critical requirement under existing regulations with the aim that participants in an investigational research are familiar with the study and are provided with a minimum amount of information to provide informed consent. Therefore, the scope of regulations covering protection of human subjects (21 CFR, Part 50) pertains mostly to compliance with the overarching regulatory goals protecting the rights and safety of subjects involved in clinical research investigations. The roles and responsibilities of the IRB (discussed in the following sections) covers additional specific obligations and commitments at the institutional levels to the standards of conduct by the investigators, sponsoring agencies, and the institutional authority to the overall safety and protection of clinical research subjects.

Exhibit 10 is a flowchart illustrating a general guideline for a clinical investigator to determine if a waiver or alteration of an informed consent requirement is practical or feasible to their proposed IND clinical investigations. The principal investigator leading the clinical study must first establish whether the proposed clinical study poses greater than minimal risk to the human subjects. If it is determined that the study poses more than minimal risk, then a waiver/alteration of the informed consent requirement is not allowed. However, if it is determined that the proposed study poses only a minimal risk, the principal investigator must establish the rationale that (a) the proposed waiver will not affect the rights and welfare of the human subjects, and (b) it is appropriate to provide patient information to the study subject later. The IRB should review these request waivers as displayed in the flowchart and communicate to the principal investigator the decision on waiver request.

Hence, informed consent constitutes a fundamental requirement in clinical research planning and management. Under the prevailing regulations, no clinical investigation involving human subjects could proceed without the investigator having first obtained the legally effective informed consent of the subject or the subject's legally authorized representative. In order for the human subject or his or

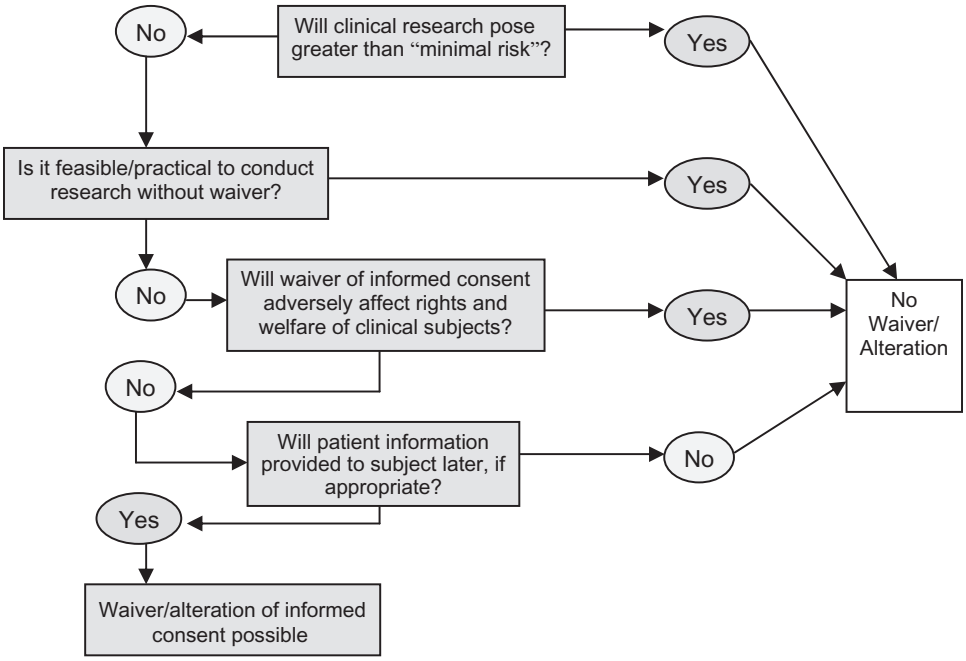


***EXHIBIT 10*** *Flowchart illustrating a general guideline for clinical investigator to determine if a waiver or alteration of an informed consent requirement is practical or feasible to their proposed IND clinical investigations.*

her authorized representative to make an informed decision, it is contingent upon the clinical investigating team to provide the information in a format easy to understand and disclose all relevant scientific, technical, and legally binding issues pertaining to the proposed study. The informed consent paperwork should not have any legal language that might waive any of the subjects' legal rights and release the investigator, sponsoring agency, and the performing institution from liability due to negligence.

Although informed consent is a fundamental prerequisite, there may be instances when a investigator working on an IND may not be able to obtain consent, such as (a) when the human subject is in a life-threatening situation and requires immediate administration of the investigational product, or (b) when the human subject is unable to communicate with, or could not obtain a legally effective consent form, and insufficient time to obtain consent from subject's legal representative, or (c) when no alternative methods or approved, or generally recognized therapy is available that has an equal or greater likelihood of saving the subject's life. Under a combination of circumstances listed here, the investigator and the physician conducting the clinical trial must certify in writing citing the reason(s) and substantiate with additional records as needed and submit to the IRB within 5 working days after the use of the IND test article.

The president of the United States may waive the informed consent requirement for the administration of an IND (including an antibiotic or a biological product) to the members of the U.S. armed forces in connection with a particular military

operation, in support of a specific protocol under an IND application by a Department of Defense sponsored clinical investigation (10 U.S.C. 1107[f]). Under this statute the president must first establish, at the request of the secretary of defense, that obtaining consent is not feasible, despite the best interests of the military member, and apply the standards and criteria set forth under the FDA regulations. A key consideration in the presidential request for waiver is the determination that the military operation presents substantial risk of exposure to chemical, biological, nuclear, or other form of exposure from the deployment that are likely to produce serious life-threatening injury or illness, or even death, and that there are no known or acceptable biodefense-related medical countermeasures currently available that could meet the potential health threat posed by the military operation.

As required under this statute, the Department of Defense will constitute an IRB with at least three members not affiliated to the Defense Department or the federal government to review and approve the proposed IND protocol without informed consent.

The flowchart in Exhibit 10 illustrates a general guideline for a clinical investigator to determine if a waiver or alteration of an informed consent requirement is practical or feasible to their proposed IND research. It is incumbent upon the investigator and the study sponsor to determine response to each of these questions in the decision tree and provide supportive documentation to the IRB as required under the regulations.

Clinical investigators must provide a clear statement of objective and explain the purpose of the research, expected duration of subject participation, description of the study protocol and actual procedures, and clearly identify procedures that are experimental in nature. The informed consent disclosures must provide sufficient information on the potential risks to health and well-being of the subject, while describing the anticipated benefits from participation in the study. The subjects must be aware of their rights and confidentiality of health information such as disease condition, medical treatments, and so forth.

The patient information and consent forms (PICF) must be approved by the IRB prior to use. A fully executed informed consent should be dated and signed by the subject or the subjects legally authorized representative at the time of the consent. All singed consent forms should be retained by the clinical investigator throughout the study period and a copy of the consent form provided to the subject.

A recent study by Beardsley et al. [7] investigated the patient knowledge and satisfaction regarding the informed consent process for cancer clinical trials and found that the lengths of PICFs, submitted to the IRB over the past few years, have increased with time.

Exhibit 11 illustrates the number of pages in the PICFs of 102 patients participating in 27 therapeutics clinical trials across 4 hospitals over the past 6 years on approved cancer clinical trials. A notable observation from this study was that although the number of pages in the PICF has grown dramatically from a mean of 7 pages (range 3–9) in 2000 compared to 11 pages (range 7–21) in 2005, important information for the patient was missing in several cases and that the patient understanding was inversely proportional to the page count of the PICFs. Evidently, the number of pages in the PICF does not correspond to the effectiveness in communicating the complex details of a clinical trial protocol.

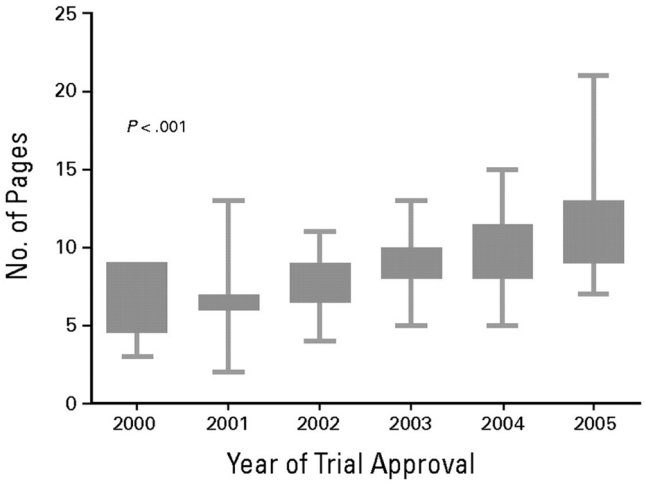


***EXHIBIT 11*** *Number of pages in the participant information and consent form in a sample of clinical studies performed during the past 5 years. (Source: Beardsley et al. [7].)*

## 7 REQUIREMENTS FOR OVERSIGHT: IRB

At the institutional level clinical trials require a thoroughly objective and unbiased oversight and review of proposed investigations by a duly appointed IRB. It is the responsibility of the IRB to conduct a technical review upfront of a proposed clinical investigation that supports applications for research involving INDs or marketing permits for products regulated by the FDA. The goal of an IRB review process is to ensure protection of the rights and welfare of human subjects proposed under a clinical investigation (21 CFR, Part 56).

No clinical investigation could begin without proper review and approval of an IRB, unless the FDA provides a formal waiver of any of the IRB requirements, including the requirement for review for specific research activities otherwise covered under these regulations. All clinical investigations involving INDs must meet the IRB requirements and obtain proper approvals prior to submission to the FDA for a formal review of the application. IND product applications with data generated from a clinical research activity conducted without proper review and documentation of an initial and continuing IRB review process may be rejected by the FDA from further consideration.

Under the existing regulations, some categories of clinical investigations may be exempt from a formal IRB review process. A couple of these exemptions apply to IND candidates; for example, investigational new products of test articles for emergency use, provided that such emergency use is reported to the IRB within 5 working days. However, subsequent use of the test article must proceed only after completion of a proper IRB review. There are other likely scenarios under which an investigational new product may be used during an emergency such as biological countermeasures products such as vaccines and therapeutics against bioterrorism-related biological agents. With much of the product development activity in the preclinical experimental research for new generation biodefense products, an emergency use of promising candidates may become necessary.

As another example, a waiver to current requirements under the formal IRB review process may be obtained if a clinical investigation for an investigational new drug for research purposes has commenced before July 27, 1981, and was subject to the requirements of IRB review under the then prevailing regulations, and provided that the investigation remained under the review of an IRB, and compliant with requirements in effect before July 27, 1981.

The IRB represents an institutional-level mechanism to approve, monitor, and report on regulatory aspects relating to clinical research set forth under FDA regulations and guidelines. As part of this effort IRB has:

1. Authority to review and approve or reject all clinical research activities.
2. Authority to seek additional information given to subject as part of informed consent to ensure that sufficient information is provided to protect the rights and welfare of subjects in the clinical research.
3. Verify all documentation relating to informed consent as required under the FDA regulations. Furthermore, IRB will determine whether a waiver is required to allow a legally authorized representative to sign a written consent form under certain conditions as long as it does not present more than minimal risk of harm to subjects.
4. Authority to ask and obtain from the principal investigator of an IND clinical research relevant information to clarify, supplement, or modify the research plan to meet the institutional and regulatory requirements.
5. Responsibility to notify clinical investigator and the institution in writing on the decision to approve or reject a proposed research activity or modification required to obtain the IRB approval.
6. Responsibility to perform reviews of research at appropriate intervals to ensure the clinical investigations are in accordance with the IRB approved protocols and plans.
7. Responsibility to inform study sponsor any exceptions to informed consent and duly document that the disclosure has occurred.
8. Responsibility to ensure that if children are some or all the subjects for a clinical study, the study plans are in compliance with appropriate regulations (under 21 CFR, Part 50).

Exhibit 12 illustrates key criteria for approval of a clinical research by the IRB. As the core of the approval process, IRB must establish that the clinical research proposal poses potentially minimal risk to human subjects, and the anticipated benefits associated outweigh potential risks associated with the clinical study.

### 7.1 Composition of IRB

The IRB consists of at least five members drawn from a multidisciplinary background in order to facilitate a proper review of the clinical research activities undertaken by the institution. From the standpoint of a review of a proposed clinical research activity involving an IND product, it is crucial that the IRB members be thoroughly familiar with the technical requirements, potential vulnerability of the

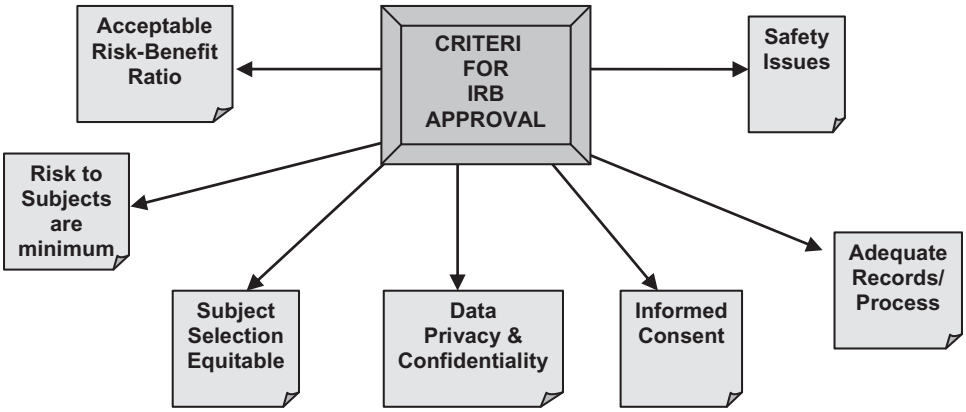


***EXHIBIT 12*** *Key criteria considered by the institutional review board as part of review and approval of a clinical research proposal.*

clinical subjects under the proposed investigation, the nature and extent of institutional commitments, other considerations including cultural background, sensitivity to such issues as community attitudes toward race and gender issues, and the safeguard of rights and welfare of human subjects. Vulnerability of clinical subjects to a clinical trial involving an IND would have to consider cases when children, pregnant women, prisoners, elderly subjects, or mentally and/or physically disabled persons are part of a clinical investigation.

Under existing regulations an IRB should be comprised of (a) qualified men and women drawn not entirely from one profession to minimize bias; (b) members representing both scientific and nonscientific disciplines relevant to bringing a balanced consideration of the overall institutional review process; (c) members from external community drawn either from professional societies and community organizations not affiliated with the institution; and (d) members without a conflict of interest, either as participants in the proposed study or as study sponsors.

The IRB is required to follow a systematic review process, follow written procedures, and document the review process (a) prior to the commencement of a clinical research activity, (b) continuing review of the research and reporting its findings and actions, and (c) document IRB review findings, approvals, and reporting to regulatory agencies as required.

Upon submission of all relevant documents of a proposed clinical study, an IRB is established to approve and oversee on a continuing basis to determine the nature, frequency, and extent of reviews required and the need for verifications from sources other than investigators that no material changes have occurred since previous IRB review. The IRB will establish written procedures for reporting changes in research protocols and ensure that changes in approved protocols were reviewed and approved prior to actual changes made to the study. A formal approval by the IRB requires that the review include at least one IRB member whose primary area of focus is in the nonscientific area, and that a majority on the IRB committee approves the proposed clinical research.

It is the responsibility of the IRB to provide formal reports to the appropriate institutional officials and the FDA of any untoward findings in the course of ongoing clinical investigations such as unanticipated problems involving risks to human

subjects, breaches in compliance to approved study protocols, noncompliance with regulations or requirements, or determinations of the IRB. If a study is suspended or terminated for any of these reasons, the decision must be communicated by the IRB immediately to the appropriate institutional authorities and the FDA.

## 8 REQUIREMENTS OF FINANCIAL DISCLOSURE

As part of the overall review of all clinical studies submitted in marketing applications, existing regulations require the study outcomes not tainted by potential sources of bias in clinical studies due to financial interests of clinical investigators. This review must ensure disclosure, if any, of the potential financial interest for clinical investigators participating in the clinical studies in the form of payment arrangements such as royalties, or proprietary interest in the product such as a patent, or equity interest with the study sponsor submitting the new product application.

The FDA is required to review financial disclosure information for all new human drugs and biological products and marketing applications and reclassification petitions for medical devices. IND applications may have certain exemptions to these requirements, depending on the nature of the clinical investigation, study design, and the type of data collected during these investigations.

This requirement applies to all sponsors submitting marketing applications to the FDA for approval of a drug, device, or biological product. It is the responsibility of the sponsor to ensure all financial disclosures in the form of disclosure statements and certifications are submitted for each clinical investigator involved in the clinical studies. Financial disclosure statements are required for even submissions of clinical study results from other investigations directly relevant to the application but obtained from investigators who are not part of the sponsor-funded clinical studies. As required under the law, regulators will use the clinical study technical information submitted by the sponsor together with disclosure statements and information collected during on-site inspections to determine the reliability of data and ensure no biases are introduced in the interpretation of the study results from clinical investigations toward a new product development, including INDs.

### 8.1 Covered Clinical Studies

As part of the financial disclosure requirements, the FDA defines what constitutes covered clinical studies (20 CFR, Part 54.2). A covered clinical study could be the complete study or parts of the study protocol dealing with the efficacy of a drug or device in humans submitted in a marketing application or reclassification petition. The key to this requirement is the determination that the study, either in total or parts, deals with the efficacy of a product and/or a study outcome making a singular but significant demonstration of safety. This would not include the phase I clinical investigation of INDs, which would generally involve dose-thresholds determination, clinical tolerance, pharmacokinetics, and general pharmacological studies except those involved in the determination of product efficacy.

Financial disclosures include clinical investigators directly involved in the study and those under subcontract to the principal clinical investigator working on a large

study at multiple locations. These disclosures also include spouses and children of clinical investigators so as to eliminate potential sources of bias in clinical studies.

### 8.2 Certification and Disclosure Requirements

Regulations require the sponsors of clinical studies to submit to the FDA a list of all clinical investigators who were involved in the clinical studies evaluating the efficacy and safety of the product, including the exemptions afforded to certain INDs. These submissions must provide complete and accurate details of financial emoluments in any form accrued by the investigators either directly or indirectly (through a subcontract) from the study sponsors. Clinical investigators with IND exemptions must provide the sponsors with sufficient accurate information to allow subsequent disclosure or certification.

In order to certify that a clinical investigator has no financial interest, the study sponsor should obtain from each clinical investigator directly or indirectly (through a subcontract) a copy of FDA Form 3454 declaring absence of any financial interests or arrangements, duly signed by the chief financial officer or other responsible corporate official of the sponsor. If the certification covers only parts of the clinical data in the application, those parts should be clearly identified as such and appended with a list of studies covered under the certification. It is the responsibility of the clinical investigator to update relevant changes to the financial relationship with the sponsor or its affiliated contractor during the course of the study.

### 8.3 Disclosure Statement Evaluation

Existing regulations provide a combination of evaluation strategies on the potential impact of any disclosed financial interest on the reliability of the study through (a) financial disclosure information and certifications furnished by the sponsor, (b) information collected from on-site inspection, and (c) effect of study design.

As part of the financial disclosure evaluation, regulations will consider potential for impact on study reliability based on the nature and extent of a disclosed financial interest, extent of downside financial benefits from an approved product, and steps taken by the sponsors to mitigate the potential biases on the study outcome. A key to this evaluation is the assessment of the overall study design, particularly large studies with multiple investigators scattered at different geographic locations and investigating on different study endpoints. Single and double-blind clinical trial designs, quantitatively verifiable endpoints at multiple locations involving different investigators, study design development, and actual study administration and data collection by unconnected investigating team working at different locations are some of the mitigating measures to minimize, or possibly eliminate, potentials for bias, should there be a financial interest among some investigators participating in the clinical study.

A key goal of the regulatory evaluation is to ensure reliability of clinical data submitted as part of a marketing application of human drug, biological product or device, and applicable INDs. If during the evaluation process it was revealed that financial interests on the part of a clinical investigator may have influenced the study outcome, regulators could respond by (a) initiating audits of data derived from the clinical investigation in question, (b) request submission of additional analysis of

data to reassess the interpretation of study results, (c) request the sponsor to conduct additional investigation through an independent third party, and (d) refuse to consider the study results as part of overall consideration for an agency action (21 CFR 54.5).

Study sponsors must retain all relevant financial records of clinical investigators involved in the study. These records must have all the changes in financial status submitted by the clinical investigator to the sponsor during the course of the investigation.

## 9 REQUIREMENTS FOR GOOD TISSUE PRACTICE COMPLIANCE

With the advent of recombinant technologies in biological product development, the scope and volume of cell-based new biologics product development have expanded dramatically in the past decade. This group of therapeutic products, generally known as the human cells, tissues, and cellular and tissue-based products (HCT/Ps) has opened a vast range of therapeutic products development using cell-based recombinant technologies. Although the first-generation cell-based therapeutic products were mostly blood and tissue transplantation, the scope of cell-based therapeutics has now expanded into other areas such as tissue repair and regeneration, modification of immune functions, facilitate tissue/organ regeneration, and gene replacement therapies. The potential for adverse effects and health and environmental risks associated with HCT/Ps presents a unique regulatory challenge.

Recognizing the need for an entirely new regulatory framework, the FDA proposed current good tissue practices (cGTP) rule together with the cGMP to prevent introduction, transmission, and spread of communicable diseases and mitigate potential environmental contamination from HCT/Ps infectious agents. This together with the requirements under cGMP for production of safe, pure, and efficacious products addresses both process controls at the manufacturing phases as well as the potential impact of releases into the general environment. As a result, the cGTP guidelines outline the methods used in the manufacture of HCT/Ps as well as recordkeeping and establishment of a quality program (21 CFR, Parts 16, 1270, and 1271).

All HCT/Ps are required to comply with the cGTP guidelines, and those candidate products, considered from initial assessment as having potentials for inducing adverse environmental impact, are required to comply with both the cGTP and cGMP requirements and are required to obtain premarket approval through the IND application process for biologics products.

The cGTP guidelines introduce additional regulatory compliance requirements in the IND clinical research investigations. Both sponsors and clinical study investigators are expected to comply with additional cGTP-driven requirements in terms of facility-level process controls, recordkeeping, and establishment of quality programs. Additional requirements of labeling, reporting, inspection, and enforcement apply under the cGTP to all HCT/Ps IND product development related activities. It is relevant to note that the cGTP requirements for HCT/Ps are regulated under the authority of the Public Health Services Act (PHS Act) and not as drugs, devices, and/or biological products. The regulatory community acknowledges the overall intent of cGTP is to improve protection of public health through good clinical care

and application of sound scientific principles to improve quality and reliability of IND product related data [8].

Facilities working on HCT/Ps are required to investigate any adverse reaction with the potential for communicable disease related to a HCT/P under development and report immediately for broader distribution. Under the cGTP guidelines, adverse reaction are defined as a noxious and unintended response to any HCT/P for which there is a reasonable possibility that the response may have been caused by the product or a causal relationship cannot be ruled out [3]. As set forth in the cGTP guidelines, it is the responsibility of the facility handling the HCT/Ps and sponsors to investigate any noxious or unintended adverse reaction and report. Regulatory agencies receiving these reports from multiple sources will look for a general pattern in the adverse events reporting to determine the nature of the emerging trend and seriousness in terms of public health impact. For example, several hospitals reporting an outbreak of methicillin-resistant *Staphylococcus aureus* (MRSA) following a procedure involving a therapeutic product from a human tissue may trace its origin to a single establishment with contamination with MRSA.

Facilities handling HCT/Ps are required to report to the FDA of any serious adverse reactions involving a communicable disease. The cGTP regulatory guidelines define reporting requirements for adverse reactions involving a communicable disease as those that are (a) fatal, (b) life-threatening, (c) result in permanent impairment of body function or permanent damage to body structure, or (d) necessitate medical or surgical intervention including hospitalization. This report must be filed with the FDA within 15 days of initial receipt of information. According to published records, this will be the first federal requirements for reporting of adverse reactions from transplanted HCT/Ps.

Perhaps, adverse event surveillance linked to facilities and operations involving HCT/P is the most critical element in the cGTP guidelines for it requires a mechanism to detect and identify adverse events by physicians and infection control practitioners in a clinical setting, and a reporting mechanism to risk managers at the facility level and onward to the regulatory agencies. The FDA has established the MedWatch program, the safety information and adverse event reporting program, to encourage and enable voluntary reporting, which is crucial for the overall program success. Although MedWatch handles issues involving commercially available pharmaceutical products, facilities performing HCT/P-related IND clinical investigational activities are required to use the same systems for reporting adverse events.

## 10 REQUIREMENS FOR IND LABELING

The drug label, also known as the package insert, is intended to provide an accurate and concise summary of all relevant information necessary for the user to determine the safety and efficacy for approved indications. Current regulations outline the overall requirements on content and format of labeling for human prescription drugs (21 CFR, Part 201.56). As required under this regulation, the label must provide information that is accurate and should not be written in a promotional tone, and most of all misleading in terms of the effectiveness, approved indications, and contraindications. As far as possible the label claims must be based on data

**XXX Research Center**
**YYY Address St, City, State 00000**
**Phone: 000-00-0000**

**Pt Name (or study ID number): Date (dispensed)**
**Dr. ZZZ (must be MD) Visit # (or way to track):**

**Drug name and strength or study acronym (include Manufacturer)**
**Take as directed. Bring bottle to each clinic visit. Do not discard when empty.**

**Caution:** New Drug.

(Limited by Federal Law to Investigational Use.)

***EXHIBIT 13*** *Sample investigational drug label for exclusive use during the clinical trial.*

derived from clinical studies and not based on experimental data from nonclinical investigational studies.

The labeling requirements set forth under the current regulations for all drugs and biological products apply to the IND as well.

Exhibit 13 illustrates a sample investigational drug label for exclusive use during the clinical trials. The label format and content should be in strict compliance with the regulatory requirements for package inserts with the exception that the label for the IND must clearly mention that under federal law the product is meant only for investigational uses and a note of caution that the content is a new drug. The name of the physician on the label must be a participating clinician in the IND clinical trial.

The IND labeling issue is particularly relevant in the "off-label" and investigational use of marketed drugs, biologics, and medical devices. Off-label use of an approved drug refers to a use, by a practicing physician, of a drug that is not included in the approved label for human prescription drugs.

A general topic of discussion in the clinical research and regulatory affairs community is whether an off-label use of an approved drug constitutes experimental investigation that normally requires formal review and approval by an IRB and strictly follows the informed consent process. Likewise, there is some confusion as to whether an off-label use of an approved drug requires formal submission of an IND application to FDA.

The requirements under good medical practice and the references within the existing drug approval process address these issues.

Good medical practices establish the principles and values that guide medical professionals in clinical care and service delivery. Although the good medical practices are essentially addressed to clinical practitioners, its broader goal is to let the public know what they can expect from medical services rendered at a clinical setting. This is particularly relevant in the context of off-label use of an approved drug for an indication not included in the approved label.

If a physician uses a product for an indication not on the approved label based on his or her clinical knowledge and patient history, it is extremely important that the use be based on sound scientific rationale and a thorough understanding of the product indications and potentials for adverse effects and maintain complete records

of all treatment procedures and therapeutic regime for off-label uses. However, under the current regulations there is no requirement of off-label uses for a formal submission of an IND application, investigational device exemption, or an IRB review. Instead, the clinical knowledge and expert judgment of the physician based on a detailed knowledge of the product and the patient clinical condition guide the off-label use of an approved drug or medical devices. Off-label uses within an institutional medical practice may require an IRB review or other existing institutional oversight and governance process.

It is important to note that the off-label use of an approved drug is not the same as the investigational use of an approved pharmaceutical product since investigational use indicates the use of an approved product in a clinical study protocol must comply with the FDA requirement under the IND application procedures (21 CFR 312). This compliance requirement is essential when the objective of the investigation is to develop additional information relating to product safety or efficacy, in which case an IND application and due process is required. However, this requirement is not considered essential if the proposed investigation meets all the following six conditions: (a) Proposed investigation will not lead to a new NDA application in support of a new indication for use or support any significant change in the approved labeling for the pharmaceutical product; (b) proposed investigation is not intended to support a significant change in advertising of the product; (c) proposed investigation does not involve a route of administration, or dosage level, or use in a subject population, or any other factor that could significantly increase the risks associated with the use of the pharmaceutical product; (d) proposed investigation requires a formal IRB review or any other related IND application filing requirements; (e) proposed investigation is an attempt to comply with a requirement concerning promotion and sale of a pharmaceutical product; and (f) proposed investigation does not invoke the exemption from informed consent requirements for emergency research (21 CFR 50.24).

Hence, the off-label use of an approved drug is guided by the good medical practice and is the practitioner's responsibility for prescribing a drug for uses not indicated in the approved label. The off-label uses in a clinical setting should be based on a sound scientific rationale and a thorough understanding of the pharmacological data determining the label content. Although the IND application requirements do not apply under these conditions, legal actions are more important in the off-label use of an approved drug. Physicians are at the risk of facing a malpractice lawsuit for negligent use of any drug whether or not the FDA has approved the use of that drug (both label and off-label). Therefore, labeling does not preclude a physician from applying his or her accumulated clinical knowledge and expert judgment in the determination of off-label uses.

In contrast to off-label use, most investigational use of an approved pharmaceutical product in a clinical study protocol must comply with the existing requirement under the IND application procedures.

## 11 MONITORING OF INVESTIGATIONAL NEW DRUG RESEARCH

From a regulatory compliance standpoint, monitoring IND research revolves around nonclinical and clinical research facilities and data generated from these

investigations. Apart from the core requirements of quality and integrity of investigational data submitted in support of an IND application, monitoring ought to protect the rights and welfare of the human subjects involved in clinical research. Recognizing the need for a comprehensive monitoring program, the FDA has instituted the Bioresearch Monitoring (BIMO) program to fundamentally ensure the quality and integrity of the data submitted to support IND applications, and simultaneously ensure protection of the rights and welfare of human subjects during this process.

The BIMO program covers both on-site inspections and data audits to monitor all aspects of the experimental and clinical research conducted in support of the FDA regulatory approval process. According to published reports, the BIMO program on average conducts over 1000 inspections annually and covering GLP audits of nonclinical testing labs, clinical investigators, sponsors/monitors, and the IRBs [6]. Implemented through four multicenter compliance programs, BIMO covers facilities and clinical research activities at both domestic and international locations.

Large medical institutions, both private clinical research facilities and academic teaching institutions with significant clinical research related activities, have established a detailed institutional-level monitoring of the IND-related clinical research studies, including standard operating procedures and GCP both at the institutional and study levels.

Exhibit 14 illustrates institutional-level GCP compliance monitoring elements during an IND clinical research activity. The goal of institutional-level monitoring is to oversee clinical research activities and ensure that these activities are conducted, recorded, and reported in compliance with the established protocol and standard operating procedures and good clinical procedures.

Monitoring is a continuous activity requiring reporting and regularly scheduled reviews by the IRB, which has the oversight responsibility. Depending on the clinical study design, a medical review committee may be established to specially monitor a clinical protocol. On the contrary, audit is a systematic and independent investigation conducted by an external team of all aspects of a clinical study and its overall compliance with the standard operating procedures, GCP, and applicable regulations.

Most often the goal of monitoring review is to improve performance, establish data integrity, protect human subjects, and establish compliance with internal procedures and regulatory requirements. Exhibit 14 lists typical monitoring elements related to GCP compliance in IND clinical research relating to process improvements, data integrity, protection of human subject, and compliance with GCP and applicable regulations. The responsibility for most of the GCP compliance rests with the principal investigator and the monitoring responsibility with the IRB.

### 11.1 Clinical Risk Assessment

A key element of the institutional-level monitoring of the IND research activity is the establishment of a risk assessment and risk management program. As part of the risk assessment process, it is important to distinguish between hazard and risk. Whereas hazard may be defined as any factor—internal or external to the clinical investigation—that could cause harm, risk is the measure of the probability that

| Compliance Element | Description | Lead Responsibility | Monitoring Responsibility |
|---|---|---|---|
| Regulatory documentation | Documentation and resources to track communications with FDA, work flow, other regulatory compliance documentation (EPA, OSHA, biosafety, etc). | Principal investigator | IRB/ regulatory compliance department |
| Documentation of roles and responsibilities | Filed by the principal investigator on the roles and responsibility of clinical research team members; educational qualifications; credentials; etc. | Principal investigator | IRB/ chief medical officer |
| Clinical research management system (CRMIS) | CIO establishes a customized CRMIS to track work flow and link database. | CIO/principal investigator | IRB |
| Patient outreach program | Oversee clinical subject recruitment process; resources for outreach and enrollment and retention; cost-effectiveness analysis. | Principal investigator | IRB/study operations support office |
| Informed consent | Electronic documentation to track all paperwork; quality assurance; educational resources to assist patient education on the study and informed consent. | Principal investigator | IRB |
| Inclusion/exclusion criteria documentation | Case report forms, detailed checklists; supportive documentation; links to database. | Principal investigator | IRB/medical review committee |
| Adverse effects reporting system | Serious adverse event (SAE), unexpected adverse event (UAE) reports in real time; and periodic; links to database. | Principal investigator | IRB/medical review committee/ FDA |
| Drug/biologics accountability | IND/IDE application templates; SOPs; sponsor IND documentation. | Principal investigator/ study sponsor | IRB/pharmacy department |

***EXHIBIT 14*** *Institutional level monitoring of IND of good clinical practice compliance during IND clinical investigations.*

harm will be caused by the hazard. Therefore, presence of hazard alone does not constitute establishment of risk to a research study. What is needed is a structured risk assessment process that takes into account all technical, operational, and information systems and processes related to an IND clinical investigation and a clearly worked out risk mitigation strategy.

Exhibit 15 is a notional illustration of various elements of a clinical risk assessment and risk mitigation process. It is important to note that the clinical team should identify potential hazards inherent in a trial and their associated risks, potential consequences, and a reasonable approach mitigating the risks.

Hazard identification and assessment should identify and sort hazard by its origin and potential impact such as:

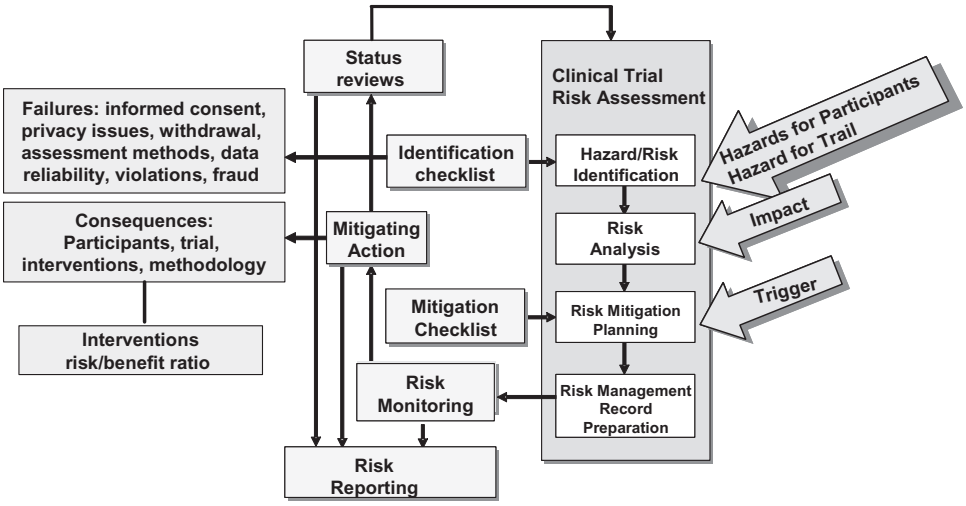


***EXHIBIT 15*** *Notional illustration of various clinical trial risk assessment and risk mitigation process as part of an institutional monitoring of IND research activities.*

1. Hazard for the participants arising due to failure to properly complete the informed consent process or information systems failure to protect the privacy of participants.
2. Hazard for participants due to the nature of intervention mechanism set up as part of the study design such as unexpected and/or expected adverse effects, assessment methods such as biopsy, radiation, and pharmaceutical adjuvants used in the IND formulation.
3. Hazard for the trial such as potentials for an incomplete study due to failures in human subjects recruitment and follow-up, violation of the inclusion/exclusion criteria, reliability of the study results, procedural errors, improper information assurance, quality systems failure, failure in adherence to study protocol, general fraud, and misrepresentation.

Risk management approaches must carefully weigh the consequences of the hazard, either to the study participant or the clinical trial, or both, before developing options for mitigating the risks. For example, establishment of a training program for informed consent process, privacy protection, and information assurance and quality systems would address several participant- and study-specific hazards. Similarly, establishment of systems to monitor and report adverse effects and systems to maintain awareness and ability to report adverse events, systematic review of the study design and clinical trial protocol to assess the statistical power and reliability, and a well-designed monitoring and reporting program to track and report study violations are some of the more commonly employed options to mitigate risks associated with IND clinical investigations.

## 11.2 Computerized Systems in Clinical Trials

Clinical research management systems are the indispensable platform to seamlessly integrate clinical research, IND activities and regulatory management. Clinical

research systems employ clinical-rules-based decision systems to help guide clinical practices that are key to clinical trials and establish a collaborative environment for information exchange, storage, retrieval, and analysis. Recognizing the importance of clinical information systems to IND, the BIMO program inspections and audits are centered on the data resident in the clinical research management systems to ensure highest standards of quality, reliability, and conformity with the regulations.

Information gathered during clinical studies should meet the established criteria of quality to remain compliant with requirements for electronic data and a mechanism to audit the system for data attribution, accuracy, and originality. Guidance is available to the industry on how these data quality requirements might be satisfied where computerized clinical research management systems are employed to generate, analyze, modify, archive, and transmit clinical data [2]. This guidance also addresses requirements of the Electronic Records/Electronic Signatures rule (21 CFR Part 11). The guidance may be applicable to source documents created in hardcopy and later entered into a computerized clinical research management system, or directly entered into a computerized system, or automatically captured by a computerized system.

The guidance to industry identifies standard operating procedures relevant to use of computerized systems such as:

1. *Data Entry:* To ensure data attributability through identification of individuals entering the data, passwords protection to limit and track access control, electronic signatures, audit trails, and date and time stamps.
2. *System Features:* Features that will facilitate collection of quality data such as consistent use of terminologies, data tags to facilitate inspection and review, ability to retrieve data, maintain collateral information relevant to data integrity, and system capability to reconstruct a study to backtrack how data were obtained and managed in support of an audit.
3. *Security:* To include physical and logical security. Physical security refers to internal safeguards built into the system and external safeguards to restrict access to authorized users. The system must have a robust feature to prevent unauthorized access. Logical security refers more to maintaining data integrity and to ensure that the information resident in the system is not altered, browsed, or transferred using external software applications.
4. *System Dependability:* To ensure that the system is in conformity with the study sponsors' established requirements for completeness, accuracy, and reliability. System documentation should be readily available for inspection during site visits. Sufficient documentation on software systems validation such as written design specifications, test plan, and test results and demonstration validating the system design specifications.
5. *System Controls:* To include software version controls, contingency plans in the event of a system failure, and a backup and recovery plan to retrieve electronic records.
6. *Training Records:* To include documentation on qualification of individuals managing the database systems and data entry activities, and training records for verification that suitable training was provided to individuals performing these functions.

7. *Records Inspection:* During a facility inspection all records submitted to the Agency may be audited for track changes, no matter how they were created or maintained.
8. *Electronic Signatures:* Electronic signatures are intended to be the legally binding equivalent of conventional handwritten signatures.

## 11.3 Quality Assurance

As part of an IND product development activity, it is the responsibility of the sponsor to ensure compliance with all pertinent regulations and quality standards. For example, product development methodologies will be compliant with section 505(b) of the Food, Drug, and Cosmetic Act, 21 U.S.C. § 355(b), methodology used for preparation of the drug substance, and the control testing used to monitor its identity, strength, quality, and purity, as required for IND submission.

Exhibit 16 illustrates the dramatic shift in the regulatory landscape of QA before and now, propelled due to an increasing emphasis on quality systems and the establishment of a robust QA/QC as part of the entire product development life cycle. In the last decade, the regulatory affairs related to QA during IND have shifted to a more activist role with explicit requirements for audits and greater contacts between the industry and the regulators on QA-related matters. The reporting requirements have increased transparency of the IND information collection processes and improved access to QA data.

From the industry standpoint, the IND product development and clinical trials landscape has changed significantly in the past two decades, propelling the need for a reform in regulatory compliance requirements in quality control and quality assurance. More and more clinical investigations now involve extended studies, involving multiple sites and having a large volume of clinical trial subjects at each site. Added

| Regulatory Landscape Before | Regulatory Landscape Now |
|---|---|
| No explicit requirement for QA audits | Audit required under the FDA International Conference on Harmonisation (ICH) GCP Consolidated Guidelines. Section 5.19 (Audit). |
| Mostly left to the sponsors to implement QA as an in-house program and considered a cornerstone of program management success. | Explicit requirement under the FDA International Conference on Harmonisation (ICH) GCP Consolidated Guidelines. Section 6.11 (Quality Control and Quality Assurance—Documents). Other QA guidelines are: European Union (EU) The Engage Guidelines; World Health Organization GCP Guidelines. |
| No explicit QA related reporting requirements | Greater contact with the industry on QA. |
| QA was a "black box" as far as FDA was considered and interpreted as a tool for success for sponsor's product development program | Increased transparency and reporting requirements on misconduct and activities leading to compromise on the safety and security of human subjects. |

***EXHIBIT 16*** *Comparison of the regulatory affairs role in the quality assurance and quality control during the IND phase of product development.*

to this is the new and expanded role now played by clinical research organizations and a vastly expanded pool of clinical investigators now taking part in IND-related clinical studies. Large pharmaceutical companies have reached out to global destinations for IND-related clinical trials bring in countries and medical research institutions that were traditionally not part of the product development pipelines. Access to highly sophisticated clinical research information systems and electronic connectivity have made possible information collection, sharing, and analysis in a widely distributed global environment. Finally, an increasing number of IND clinical studies are now designed that allow participation of sensitive human subjects based on age group and pre-existing clinical conditions. These are compelling enough reasons for the regulators to recognize the need for a compliance framework addressing quality control and assurance as an overall requirement throughout the product development process.

The quality assurance and product technical support departments of the sponsors are responsible for assuring that IND study participants meet the quality objectives and comply with the established guidelines. Participating institutions (hospitals, academic medical institutions, and clinical research organizations) are required to implement and follow written approved procedures to ensure all operations are performed in accordance with quality systems guidelines under the GCP and cGMP regulations and in-house policies.

All personnel supporting or engaged in the IND-related manufacture, testing, or quality assurance are required to comply with the established written approved procedures. In addition, these institutions must have a comprehensive quality program that controls all manufacturing including a complete document control system, documentation review and approval, auditing, and training.

As part of the overall document control process related to QA, study sponsors and participating institutions must maintain an approved document entry and tracking system that includes specifications, standard operating procedures, testing protocols, batch records, and test reports. Study sponsors may establish management oversight of the entire process with real-time oversight of critical production and testing activities, as well as audits.

Exhibit 17 illustrates a typical list of areas covered during QA audits of a biologics facility. The audit area covers floor areas for proposed activities, equipment identified for the projects, documentation, employee training, and facility audit/biosafety inspection records. Facilities used for the IND product development may be inspected to verify the availability of appropriate space, compliance with biosafety requirements, equipment and personnel to operate the following functional areas: (a) fermentation suite, including downstream processing, (b) purification suite separated into upstream and downstream purification rooms, (c) production support areas, (d) process and facility utility area, (e) waste treatment area, (f) warehousing, (g) QA/QC, (h) cell banking, (i) finished and in-process material storage, and (j) offices.

## 12 EMERGING BIOSAFETY AND BIOSECUIRTY REQUIREMENTS

Biotechnology and pharmaceutical companies with an array of biologics product development portfolios and clinical research organizations are beginning to place

| Facilities and Equipment | Qualification and Calibration |
|---|---|
| ▪ Production and control laboratory equipment<br>▪ Normal maintenance<br>▪ Preventative maintenance<br>▪ Maintenance and servicing of equipment | ▪ Equipment installation and operational qualification<br>▪ Equipment performance validation<br>▪ Cleaning validation<br>▪ Equipment calibration<br>▪ Computer software validation |
| **Sanitization** | **Production** |
| ▪ Analytical methods validation in support of cleaning<br>▪ Facility cleaning<br>▪ Equipment cleaning | ▪ Raw materials/supplies<br>▪ Sampling<br>▪ Quarantine<br>▪ Bulk manufacture<br>▪ Packing<br>▪ Rejected materials/supplies |
| **Storage and Distribution** | **Process Validation Quality Control** |
| ▪ Complaints<br>▪ Product recalls<br>▪ Self-inspection | ▪ Microbiological/environmental support<br>▪ Analytical support<br>▪ Raw material support |
| **Documentation (Technical)** | **Documentation (Information/Systems)** |
| ▪ Production records<br>▪ Standard operating procedures<br>▪ Standard manufacturing procedures<br>▪ Protocols<br>▪ Change control<br>▪ Equipment specifications<br>▪ Raw material specifications | ▪ Training procedures<br>▪ Computer program specifications<br>▪ Documentation control of process deviations<br>▪ Calibration and test documents<br>▪ Validation documents<br>▪ Purchase orders<br>▪ Authorization to ship |

***EXHIBIT 17*** *Typical list of areas of compliance investigated during audits of IND product development programs.*

considerable importance on laboratory biosecurity with a focus on improving security at microbiological research facilities, clinical laboratories, and ancillary laboratory services such as biological material storage and distribution facilities. A key element of this growing awareness requires a clear delineation of the concepts of biosafety and biosecurity in IND product development activities in the context of new regulations. Whereas biosafety refers to institutional-level measures to prevent and mitigate the accidental release of biologic agents and toxins, biosecurity refers to instructional measures that guard against the deliberate release of pathogens for malicious purposes (including bioterrorism). Thus far, existing U.S. and international regulations and guidelines have focused on biosafety rather than biosecurity.

In the aftermath of the 9/11 terrorist attacks, followed in the same year by a string of anthrax attacks in the United States, the Congress passed two significant pieces of legislation. First, the Uniting and Strengthening America by Providing Appropriate Tools Required to Intercept and Obstruct Terrorism (USA PATRIOT) Act of

2001 established criminal penalties for possession, shipping, and receiving of certain biological agents, known as select agents (SA) and toxins, if used as a weapon or for any reason not plausibly justified for prophylactic, protective, bona fide research, or other peaceful purposes.

Second, the Public Health Security and Bioterrorism Preparedness and Response Act (PHSBPRA) of 2002 greatly expanded controls over dangerous pathogens and toxins stored, used, and transferred between laboratory and ancillary facilities within the United States. These legislations establish the regulatory premise to introduce biosecurity practices at research laboratories handling dangerous etiologic agents and toxins as part of an overall national security program.

For the most part, biosafety and biosecurity practices are guided by institutional policy and governance set up for this purpose. The environmental health and safety (EH&S) division is most often charged with addressing all biosafety and environmental regulatory affairs as they apply to development, production, and testing of IND products. It is up to the management leadership in an organization to recognize and prioritize the importance of biosafety and biosecurity as fundamental requirements at all phases of product development activities. This requires a careful integration of EH&S-related guidelines and regulations throughout the product life-cycle development. For an effective EH&S program it is important that staff be highly qualified in the areas biosafety, environmental health, and risk assessment and having working knowledge of the regulations covering product development.

Exhibit 17 is a summary of typical areas of compliance investigated during an audit of an IND product development facility. Evidently, the audit areas are cGMP and quality systems related areas but with a focus on the safety and security of the product and processes with a goal of protecting the workers in the occupational setting and the general environment. As part of the safety and security-related audits, the EH&S team may review floor plans as they relate to potential for cross-contamination and release of potentially infectious materials in the working environment. Production process and laboratory equipment, air and water distribution systems, and sanitary systems inspected to ensure built-in engineering protection segregate and contain hazardous chemicals and biological materials from drinking water systems, heating, ventilating, and air-conditioning (HVAC) systems. Warehouse and receiving facility for nonclinical experimental facilities are where animal housing is maintained and bulk raw chemicals are stored. Finally, biocontainment facilities involving biosafety levels 2 or 3 (BSL-2 or BSL-3) are required for handling potentially dangerous etiologic agents and infectious materials for experimental purposes.

As required, the EH&S will perform a biosafety-related risk analysis, known as the maximum credible event (MCE) analysis to determine the potential for accidental release of dangerous biological agents into the occupational environment within the facility and general environment outside, and assess the facility-level crisis and consequence management capability and resources for effective containment of an accidental release during manufacturing, processing, storage, or animal testing operations.

EH&S staff perform preaward audits of subcontractor facilities to ensure compliance with all applicable and current federal, state, and local laws, codes, ordinances, and regulations, as well as all public health services safety and health provisions. These activities would include:

- Review of subcontractor's facility safety plan and complete site visit to verify floor plan details on the ground and review safety records, biosafety oversight committee meeting minutes, and standard operating procedures (SOPs).
- Review the type of hazardous chemicals regulated under the Resource Conversation and Recovery Act (RCRA, 1976) and perform hazard assessment as required.
- Conduct detailed site visits and proposed testing of production areas and associated support areas. All areas are assessed for the appropriateness of the biological safety level employed and compliance with the biosafety standard including general housekeeping, security, and general safety (fire, chemical, radiation, and electrical).
- Review all documentation required by federal, state, and local regulations and related in-house documentation. Documents that are reviewed may include, but not limited to, chemical hygiene plans, safety SOPs, biosafety/safety committee meeting minutes, HAZCOM plan, biological safety plan, training records, engineering control equipment maintenance and certification records, written respirator program, and medical surveillance program.
- Written recommendations are provided as part of the facility biosafety visit and a corrective action plan is requested.
- Monitor facility compliance with biosafety and environmental practices through annual inspections and frequent communications with the facility EH&S staff.

The EH&S staff from the sponsoring organization may be required to provide support and additional oversight over the entire spectrum of facility biosafety, environmental regulatory affairs, workers protection requirements for hazard analysis, and risk assessment for IND product-related development activities.

Requirements for biosafety practices and compliance with environmental and worker protection are unique for each project and could change as the nature of activities change. For example, the volume of a potentially infectious material stored at a facility could determine the biosafety requirement. Thus, a lower biosafety level for a storage operation (with no volume changes) may not apply for a production/testing-related operation where volume changes are likely. Likewise, a higher volume of a hazardous chemical stored at the facility could warrant application of RCRA regulations, whereas a lower volume would be excluded from the regulation.

Both sponsoring organization and participants in the product development activities can make no compromise on biosafety and environmental regulations compliance, designed to be used as screens upfront in facility selection process and monitored by designated staff from both organizations to ensure that the facility remains fully compliant with regulatory requirements throughout the IND activities.

## 13 CONCLUSIONS

The IND is a key phase within the drug development life cycle. The regulatory affairs relating to IND requires a balancing consideration of the inherent benefits

of introducing novel medical products for human use, while at the same time protect clinical trial subjects from potentially harmful test candidates during the preclinical product development phase.

The FDA regulations are the principal drivers of the regulatory affairs related to IND development. Existing FDA regulations relating to IND clinical trials such as cGMP, GLP, and GCP guide IND development for drugs and biologics, and new initiatives such as the GTP attempt to incorporate the expanded scope and volume of recombinant technology-based HCT/Ps and genomic therapies. The FDA has reinvigorated the scope of regulatory affairs related to clinical risk assessment, quality assurance, quality control, and use of computerized systems in clinical trials. Electronic IND submissions have eliminated the need for voluminous hardcopy submission, which require the entire study planning, preparation, execution, and management within the information systems domain.

Biotechnology companies are involved in highly sophisticated biologics products development based on genomics and functional proteomics with unclear regulatory implications, risks to clinical subjects, product liability, and potentials for long-term public health risks. Recognizing the importance of technology and regulatory compliance in the development of novel therapies, the FDA launched the Critical Path Initiatives covering product development strategies and technologies in these IND development areas.

Recently, the FDA initiated the BIMO to proactively ensure the quality and integrity of data submitted in support of IND applications and to protect the rights and welfare of clinical trial subjects. Through a combination of on-site inspections and data audits BIMO attempts to monitor all aspects of the experimental and clinical research conducted in support of the regulatory approval process.

In the aftermath of 9/11, research laboratories are beginning to place considerable importance to improving biosecurity at microbiological research facilities, clinical laboratories, and ancillary laboratory facilities to guard against misuse of pathogenic materials and select agents used in IND research and medical countermeasure products development activities. At the institutional level, integration of a robust EH&S program for biosafety and biosecurity as part of the product development life cycle is an essential regulatory requirement.

## APPENDIX: APPLICABLE AND RELEVANT REGULATIONS COVERING IND

### Key IND-Related Regulations

| | |
|---|---|
| 21CFR Part 312 | Investigational New Drug Application |
| 21CFR Part 312.82 | Early Consultation for Pre-Investigational New Drug Meeting |
| 21CFR Part 314 | INDA and NDA Applications for FDA Approvals to Market New Drug |
| 21CFR Part 314.42 | Revisions to Agency Requirements on IND Applications |
| 21CFR Part 314.420 | INDA Application, Master File Submission |
| 21CFR Part 316 | Orphan Drugs |
| 21CFR Part 316.10 | Content and Format of a Request for Written Recommendation to get Orphan Drug Status |

| | |
|---|---|
| 21CFR Part 58 | Good Laboratory Practice for Non-clinical Laboratory Studies |
| 21CFR Part 58.3 | Physical and Chemical Characteristics of Test Article in GLP guidelines |
| 21CFR Part 50 | Protection of Human Subjects |
| 21CFR Part 50.24 | Informed Consent Requirements for Emergency Research |
| 21CFR Part 56 | Institutional Review Boards |
| 21CFR Part 201.56 | Content and Format for Drug Labeling |
| 21CFR Part 54 | Financial Disclosure by Clinical Investigators |
| 21CFR Part 54.2 | Covered Clinical Studies for Financial Disclosures by Clinical Investigators |
| 21CFR Part 54.5 | Agency Evaluation of Financial Interests for Clinical Investigators |
| 21CFR Part 820 | cGMP Quality Systems Guidelines for Finished Medical Devices Intended for Human Use |
| 10 U.S.C. 1107 | United States Code—Notice of Use of an IND or a Drug Unapproved for its Applied Use |
| 21CFR Part 16 | Protection of Human Research Subjects |
| 21CFR Part 1270 | Current Good Tissue Practice Compliance |
| 21CFR Part 1271 | Current Good Tissue Practice Compliance |
| 21CVR Part 11 | Electronic Records/Electronic Signatures Rule |

### Other Relevant Regulations Applicable to IND Development

| | |
|---|---|
| FD&C | Federal Food, Drug and Cosmetic Act, 1938 |
| USA PATRIOT | Providing Appropriate Tools Required to Intercept and Obstruct Terrorism Act, 2001 |
| PHSBPRA | Public Health Security and Bioterrorism Preparedness and Response Act, 2002 |
| RCRA | Resource Conversation and Recovery Act, 1976 |
| BMBL | Biosafety in Microbiological and Biomedical Laboratories Guidelines (4th Edition) |
| OSHA | Occupational Safety and Health Act, 1970 |
| ISO9001 | International Organization for Standardization 9000 Series for Quality Management System |

## ACKNOWLEDGMENT

Author would like to acknowledge the interest and commitment of the National Defense Program of the Computer Sciences Corporation to support technical excellence, and Mr. Alvin Keith, the Business Unit Executive, for the support in the preparation of this chapter.

## REFERENCES

1. FDA (2002), New Drug and Biological Drug Products; Evidence Needed to Demonstrate Effectiveness of New Drugs When Human Efficacy Studies Are Not Ethical or Feasible,

Final Rule, 21 CFR Parts 314 and 601, Docket No. 98N-0237, Health and Human Services, Washington DC.

2. FDA (1999). Computerized Systems Used in Clinical Trials—Guidance to Industry. Bioresearch Monitoring Program, Food and Drug Administration, Rockville, MD.
3. FDA (2004), *Pharmaceutical cGMPs for the 21st Century—A Risk-Based Approach*, Department of Health and Human Services, U.S. Food and Drug Administration, Final Report. Washington, DC.
4. Organization for Economic Cooperation and Development (OECD) (1998), OECD Series on Principles of Good Laboratory Practice and Compliance Monitoring, ENV/MC/CHEM (98)17, 34.
5. FDA (2008). CGMP for Phase I Investigational Drugs—Guidance to Industry. Division of Drug Information, HFD-240, Center for Drug Evaluation and Research, Food and Drug Administration, Rockville, MD.
6. FDA (2008). Cumulative List of Products with Orphan Designation. The Office of Orphan Products Developments, Health and Human Services, Washington, DC.
7. Beardsley, E., Jefford, M., and Mileshkin, L. (2007), Longer consent forms for clinical trials compromise patient understanding: So why are they lengthening? *J. Clin. Oncol.*, 25(9), 13–14.
8. Burger, S. R. (2003), Current regulatory issues in cell and tissue therapy, *Cytotherapy*, 5(4), 289–298.

## BIBLIOGRAPHY

Bhattacharyya, S. (2006), in *Product Development, Preclinical Testing and Toxicity Studies*, CBER Presentation at the Drug Information Association Meeting, June 18–22, Philadelphia, PA.

Hirschfeld, S. (2006), in IND "202" Clinical Perspective. CBER Presentation at ISCT 6th Annual Somatic Cell Therapy Symposium, September 26–28, Bethesda, MD.

# *Preclinical Assessment of Safety in Human Subjects*

**Nancy Wintering and Andrew B. Newberg**
*Department of Radiology, University of Pennsylvania, Philadelphia, Pennsylvania*

**Contents**

1 Preparing Human Studies in Light of Preclinical Studies
2 Estimated Biodistribution and Pharmacokinetics in Humans
3 Initial IND Process
4 Overview of Sections of IND
5 IND Exemption Status
6 Institutional Review Board Issues
7 Study Design
8 Subject Selection
9 Safety Measures
10 Monitoring for Adverse Events
11 Preparing for Phase II Studies and Additional Safety Evaluation
12 Conclusions
References

## 1 PREPARING HUMAN STUDIES IN LIGHT OF PRECLINICAL STUDIES

The first study to assess for safety in human subjects begins with the initial development of the drug. The particular physiological target for the drug should be known (e.g., whether it is designed to bind to serotonin receptors or specific tumor receptors). This target is determined by the group or company designing the pharmaceuti-

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

cal, which develops a molecule that is based upon existing ones or based upon known physiological molecules. Thus, the initial safety assessment targets the most likely pharmacological effects of similar molecules. For example, a drug that is intended to bind to the serotonin reuptake sites would be expected to have a safety profile similar to that of other drugs that bind the serotonin reuptake sites. With this knowledge in mind, the initial approach toward evaluating the safety of the new drug would include an evaluation of all pharmacological effects of known related drugs.

In addition to the presumed effects of a new drug based on related drugs, extensive animal studies (as described elsewhere in this book) also provide important information regarding how to assess for safety in human subjects [1]. The Food and Drug Administration (FDA) guidelines for nonclinical safety studies required for the conduct of human trails was published in 1997 [2]. Initial data in mouse and rat models provides preliminary information regarding the physiological changes to blood counts, electrolytes, liver function, immune function, and cholesterol levels. The initial small-animal data also provides clues of potential clinical adverse effects such as disturbances in behavior, appetite, sleep, and activity levels. If such effects are observed, then special care must be taken to observe similar effects in human subjects, and the study must be designed accordingly to evaluate these effects and also to reduce the risk of adverse effects. Initial toxicology studies in animals typically require a small and large animal in addition to nonhuman primates. However, there may be reasons to forego such studies in the case of a drug that targets a process for which there is no appropriate animal model. For example, recent studies of radiopharmaceuticals designed to bind to amyloid plaque in patients with Alzheimer's disease did not require study in nonhuman primates since there is no good model. Therefore, the results are not very meaningful in evaluating the potential effect in human subjects.

Toxicology studies typically require the administration of several different doses, often at substantially higher doses than will eventually be given to human subjects. The drug is then given to the animals for a time period similar to that in which the drug will actually be given. The test article is then given with measurements made of the concentration in the serum of the animals throughout the study. At the end of the dosing regimen, animals are typically sacrificed in order to assess for drug-related changes. Of course, if any results of the animal toxicology studies suggest that the drug might, in fact, pose a substantial danger to human subjects, this might prevent its even being tested in human subjects.

Observations for mortality, morbidity, and the availability of food and water are conducted at regular intervals for all animals. Observations for clinical signs are usually conducted daily on main study animals. In addition, some type of functional observational battery (FOB) is conducted on surviving main study animals prior to initiation of dosing, throughout the trial, and then at completion. Body weights are measured and recorded daily for all animals. Food consumption is measured and recorded daily for main study animals. Ophthalmoscopic examinations are conducted on animals pretest and on main study animals at termination. Blood and urine samples for clinical pathology evaluations are collected from all main study animals at termination. Blood samples for determination of the plasma concentrations of the test article are collected at designated time points throughout the study.

Upon completion of the study, complete necropsy examinations are performed on all main study animals, organ weights are measured, and selected tissues are microscopically examined. A similar analysis is performed in a large-animal model such as dogs with similar evaluations of clinical and laboratory assessments in addition to necropsy evaluations. At the conclusion of these studies, if there are any alterations observed in the animals, reporting of these findings in the investigational new drug (IND) are required and also necessitates careful evaluation in human subjects. Animal studies provide valuable information to assist in the assessment of safety during the development of a test article for human studies. Thus, if there was a significant change observed in platelet function in dogs, then platelet function should be closely evaluated in humans. Often this requires evaluation for at least the duration of the effect observed in animals, and sometimes longer. Thus, if platelet effects were noticed to last up to 2 days in animals, this should be evaluated for at least 2 days and most likely longer in human subjects.

## 2 ESTIMATED BIODISTRIBUTION AND PHARMACOKINETICS IN HUMANS

Similar to toxicology studies, pharmacological evaluation also follows from animal studies and estimates of drug effects. Serum and tissue evaluation of drug concentrations and metabolism is necessary for determining the overall pharmacological effects and to evaluate potential adverse effects of the test article. Biodistribution evaluation in mice and rats and nonhuman primates can be very valuable in predicting the distribution in humans.

Some drugs are easier to have their biodistribution and pharmacology quantified. For example, many radiopharmaceutical products, which combine a radioactive atom with a molecule that follows some aspect of body physiology, can have their biodistribution evaluated through a series of imaging studies in which the radioactivity emitted from the product is detected with scans performed at a number of time points after administration of the drug. This allows for an evaluation of the distribution of the drug as well as the dosimetry that refers to the radioactive exposure an individual receives from the product. Nonradioactive drugs typically require an evaluation of serum concentration of the drug at various time points after administration. It may also be necessary to evaluate urine and fecal samples to determine drug concentrations and help evaluate the mechanism of excretion from the body.

The distribution and concentration of the drug in the body also helps with dose determination. If it is found that a drug is excreted more rapidly in human subjects than in animals, it might be appropriate to increase the dose administered. Drug escalation trials are frequently performed with chemotherapy agents and psychiatric drugs to evaluate both safety and efficacy. The dose of the drug is increased at regular intervals, usually after at least three subjects have had no significant adverse reactions or toxicity. If the drug is determined to be relatively safe, or has minor safety concerns, then the next step is to appropriately design the phase I safety evaluation in human subjects to evaluate the metabolic and pharmacologic actions of the study drug, side effects, and possible early evidence of the drug's effectiveness in humans as described in 21 CRF part 312.21. Preclinical studies to evaluate safety in humans are an essential step in the drug development process and will provide

valuable information for the investigational new drug application and the investigational drug brochure.

## 3 INITIAL IND PROCESS

In order to begin any clinical investigation in one or more humans that involves a test article, the study sponsor or sponsor-investigator is required to file an IND application with the FDA. For clarification the "sponsor" means the person who initiates the clinical investigation but who does not conduct the investigation; the "sponsor-investigator" is an individual who both initiates and conducts a clinical investigation either alone or with others. This distinction is important since there are legally binding responsibilities associated with each role. Guidance from the Center for Drug Evaluation and Research (CDER) and the Center for Biologics Evaluation and Research (CBER) at the FDA describes a number of important regulations and provides guidance regarding the IND process. This information must be consulted for the development of any new pharmaceutical products and the submission of an IND application in accordance with 21 CRF part 312.

Under current regulations, any use in the United States of a drug product not previously authorized for marketing in the United States requires submission of an IND to the FDA. The Code of Federal Regulations sections 21 CFR 312.22 and 312.23 contain the general principles underlying the IND submission and the general requirements for an IND's content and format.

The IND process has a number of components that are summarized below. These sections are also important due to the need for the FDA to determine what safety data is most important to evaluate in human subjects. Most pharmaceutical companies are very aware of the requirements for the IND process. For investigators in the academic or clinical setting, the IND is more formidable and more comprehensive than investigative protocols. However, it is generally recommended that investigators request a telephone conference with the appropriate FDA office prior to submitting an IND so that the process can run as smoothly as possible. In this conference call, specific issues regarding the evaluation of safety in humans can be discussed so that the final IND and study protocols are more likely to be acceptable to the FDA and also the institutional review board. Ultimately, the IND application is designed to assist and guide investigators (either private or public) to study the safety and pharmacology of new drugs.

Once an IND is filed with the FDA, there is a 30-day period in which the FDA is able to review the IND information, request additional information, and provide a nonobjection letter to enable the investigators to begin human studies, or put the study on hold. If additional information is required, this can sometimes be in the form of additional toxicology studies in animals including genotoxicity. The FDA might also require additions to be made to the study protocol for more extensive safety evaluation. This review might include additional types of measures or adding more time points for evaluation. Once these issues are resolved, and the protocol and IND application has been amended accordingly, the FDA may issue a nonobjection letter to begin a phase I human study. Timely and open communication with the FDA is encouraged to facilitate the IND process. If an investigator does not respond, an IND application may be put on hold. This is generally an unfavorable

outcome since the FDA is no longer under any time restriction to process the application, which can often substantially prolong the initiation of a study. However, the Institutional Review Board (IRB) approval of the amended protocol must always be obtained prior to the initiation of the study.

## 4 OVERVIEW OF SECTIONS OF IND

There are a number of sections required for an IND submission as described below and summarized from 21 CFR 312.23 for studies with a new drug or for a new indication of an approved drug:

1. Section 1 is the cover sheet (Form FDA-1571) containing the name, address, and telephone number of the sponsor, the date of the application, and the name of the investigational new drug. The phase or phases of the clinical investigation to be conducted is identified. By signing this form, the sponsor (or sponsor-investigator) makes a commitment not to begin clinical investigations until an IND covering the investigations is in effect. There is also a commitment that an Institutional Review Board that complies with the federal requirements will be responsible for the initial and continuing review and approval of each of the protocols in the proposed clinical investigation. The investigator also states a commitment to conduct the investigation in accordance with all other applicable regulatory requirements. The name and title of the person responsible for monitoring safety evaluation and also the conduct and progress of the clinical investigation is provided. The sponsor (or the sponsor-investigator) or the sponsor's authorized representative then signs the cover letter form.

2. *Table of Contents*

3.i. *Introductory Statement and General Investigational Plan* This section usually includes a brief summary of the pharmaceutical agent to be tested, its presumed pharmacological and clinical effects, and how this will be tested through this IND.

3.ii. *Summary of Previous Human Experience* This section includes currently performed human studies. For completely new pharmaceutical products, this will be very brief and simply state that no human studies have been performed. However, for pharmaceuticals that are now being tested for other indications, there may be a substantial amount of information regarding the safety of physiological effects of the drug. It is important to provide a history of the drug development that includes a comprehensive summary from the literature that includes study populations, sample size, adverse events, and a summary of study outcomes.

3.iii. *Withdrawal of Drug* This section reports any reasons why the drug, or comparable drugs that have been developed, has been withdrawn for medical or other reasons.

3.iv. *Overall Plan for Investigational Year*

a. Rationale for this study: The reason for the study design, including reasons for specific safety testing, and evaluation of the pharmacology profile and initial efficacy are described here.

b. Indications for this study: This section typically establishes why this study is needed and describes the overall goals for developing the investigational product.

c. General approach for evaluating this drug: This section describes whether the proposed study is for safety, biodistribution, pharmacology, dose escalation, or some other study design.

d. Clinical trials to be conducted during the first year: a comprehensive description of the anticipated activity of the first year of the study.

e. Estimated number of patients to be given the drug.

f. Risks: This section identifies the risks and severity and seriousness of the drug risks based upon the toxicological data found in animal studies and in prior studies conducted with humans with the drug or related drugs.

4. This section is reserved for the FDA.

5. *Investigator's Brochure* The investigator's brochure is required for clinical trials involving sites other than the one of the primary sponsor of the IND. Thus, a single site study of a new drug may not require an investigator's brochure. However, if multiple sites are involved, then an investigator's brochure is required to provide information about the drug and to ensure consistency in the conduct of the study at each site. The investigational brochure is developed in accordance with 21 CRF 312.55.

5.i. This section includes the drug substance and the structural formula of the drug.

5.ii. This section provides a summary of the toxicological and pharmacological effects of the drugs in animals and to humans if known.

5.iii. This section describes the pharmacokinetics and toxicogical effects of the drug in animals and humans if known.

5.iv. This section provides a summary of known information related to safety and effectiveness in humans from prior clinical studies.

5.v. This section includes risks and side effects on the basis of prior experience with the drug or with similar drugs and provides precautions for drug monitoring and safety requirements in the use of the investigational drug.

6.1. *Protocols*

6.1.a. *Objectives and Purpose* Describes an overview of the reasons and goals for the study—that is, to evaluate safety, biodistribution, pharmacology, efficacy, and the like.

6.1.b. *Personnel and Qualifications* This section lists and describes the individuals and the sites/organizations involved in executing the study providing evidence of their individual qualifications.

6.1.c. *Patient Selection and Number* A basis for the number and type of subjects to be recruited for the study are described. The source of recruitment must be mentioned particularly if it involves patient populations rather than controls. Justification of the number of subjects is also essential for ensuring that the appropriate numbers of subjects are studied without putting too many individuals at risk.

6.1.d. *Study Design* This section details the overall study design and is similar to what would be included in a full IRB submission including evaluative measures, "the type of control group that will be used and the methods to be used to minimize bias on the parts of the subjects, investigators and analysis" (21 CRF 312.23), and standardized procedures that will be used for data collection, data analysis, and

statistical analysis. It is important to state that safety is a primary component of the study and to design the protocol and standardized measures accordingly. Human studies must include safety as a critical evaluation measure.

6.1.e. *Determination of Dose, Maximum Dose, and Duration* The dose, route of administration, and maximum dose and individual dose exposure are described, usually based on animal data or prior human studies.

6.1.f. *Observations and Measures* In this section, the specific types of measures, both physiological and clinical (i.e., safety) are described. It is also important to describe what measures will be regarded as safe and which as adverse effects giving clear ranges for various measures.

6.1.g. *Clinical Procedures and Minimization of Risk* All clinical procedures are to be described with a focus on monitoring how risk will be minimized during the study. This might refer to existing animal or prior human experience and might also describe how adverse events will be managed.

7. *Chemistry, Manufacturing, and Control Information*

7.i. *Chemistry* This section describes the actual chemical composition of the drug including both active and inactive ingredients "to assure the proper identification, quality, purity and strength of the investigational drug." The molecular weight and structure is also typically provided. In IND phase I studies, the emphasis in this section is on the identification of the raw materials that comprise the new drug.

7.ii. *Manufacturing and Control*

a. *Drug Substance* This is usually an extensive section that details where and how the drug is to be produced. Drug production should be according to good manufacturing practices and includes a description of the facilities and chemical reactions required to produce the material. The standard operating procedures for production do not have to specifically be included, but this section essentially follows them, and they can be included in the IND appendices to clarify the production process of the drug. A comprehensive description of the facilities and the personnel trained to oversee the drug production processes are included. If the drug is manufactured off-site, additional drug standard operating procedures and accountability measures are documented. The scope of the investigation will guide the amount of information to be submitted with the initial IND application. However, as the drug development proceeds and the study moves from a limited clinical investigation to a larger drug production and subject enrollment, the manufacturing and control procedures will be reported accordingly.

b. *Drug Product* This lists all the components of the drug product including active and inactive components, possible alternatives, and materials that do not appear in the final product but are used during manufacturing. Acceptable limits and analytical methods for assuring product stability, purity, and sterility are also included.

c. *Placebo Product* In studies that use a placebo in a controlled clinical investigation, a brief description of the manufacture and control is provided.

d. *Labeling* This section provides copies of all labels and labeling to be provided to each investigational site.

e. *Environmental Analysis Requirements* Describes whether an analysis of environmental issues is required.

8. *Pharmacology and Toxicology Information*

8.i. *Pharmacology and Drug Disposition* This section describes in detail the pharmacological properties of the investigational product as currently understood. This might be based primarily on animal studies but can also include previous human experience. This section is divided into each of these sections for clarification of the information provided to the FDA. In particular, data on nonprimate and primate species can be included under separate subheadings. Information provided should include the known pharmacology and biodistribution in animals and the presumed pharmacology in human subjects.

8.ii. *Toxicology* Similar to the pharmacology studies, this section can be expanded into subsections relating data from nonprimate and primate species with regard to toxicology studies.

a. Results on laboratory and clinical data as well as postmortem organ analysis are usually provided. It is often preferable to have two nonprimate species and one nonhuman primate species included in the analysis.

b. Each toxicology study that supports the safety of the investigator and data will be reviewed.

8.iii. *Statements of Compliance or Noncompliance* These statements of compliance indicate that the studies were conducted in compliance with good laboratory practice or provide a statement that provides the reasons or explanation for noncompliance with good laboratory practice (part 58).

9. *Previous Human Experience with Investigational Drug* In human studies, this section can be extremely brief. For drugs that have been used in humans for other purposes and are now being used for new indications, there may be substantial human experience. This information can be extremely valuable in establishing safety, pharmacology, and toxicology profiles of the study drug or test article.

9.i. This section provides detailed information that is relevant if the drug has been studied or marketed in the United States or in other countries. Comprehensive descriptive information must be provided about any clinical studies that have been conducted with the drug. This information may include published studies. Other information that may be relevant to the proposed investigation can be referenced in a bibliography or literature review.

9.ii. This section is needed if the test article is a combination of drugs previously investigated or marketed.

9.iii. If the drug has been marketed outside of the United States, this information should be described here, including whether the drug has been withdrawn from marketing for potential safety reasons.

10. *Additional Information* This section can include miscellaneous aspects of drugs including those that might have drug dependence and abuse potential or drugs that have radioactive elements. In pediatric studies, plans to assess safety and effectiveness in this population are described. Additional information that would aid in the evaluation of the drug safety can be included.

## 5 IND EXEMPTION STATUS

With these requirements in mind, it is also important to evaluate whether an IND is actually necessary. Many times in the clinical setting a particular pharmaceutical

product is used in an off-label manner. However, such a use in a research study may potentially require an IND. For example, if a drug with the specific indication for hypertension is used to help heart failure and an investigator wishes to test the use of the drug for heart failure, an IND might be required. However, the FDA and CFR regulations allow for an investigator to use a pharmaceutical product without an IND if the following conditions are met.

To begin with, the drug which is to be used must already be lawfully marketed in the United States. The investigational study must not be intended to support a new indication for use or to support a significant change in the advertising of the product. Importantly, the study should not involve the use of a different route of administration or dosage level or use in a patient population that significantly increases the risks associated with the use of the drug product. As with all studies, the study must be submitted, reviewed, and approved by the IRB and be conducted in accordance with standard regulations.

If these conditions are met, then an IND may not be required and the study can be performed under an exempt status. The advantage of an IND exemption is that it allows academic researchers who do not have an interest in altering the indications or marketing of a particular pharmaceutical to conduct a clinical investigation. Frequently, an investigator or clinician is interested in the pharmaceutical for a clinical or purely research perspective. In such cases, they may not need to utilize the full IND process.

It should also be clearly stated that there is no such thing as an "IND exemption" in the sense that there is no letter or form that can be obtained from the FDA. The FDA can always be contacted if there is a question as to whether an IND is required. Some institutions have initiated programs through their research offices or IRBs to help investigators determine whether an IND is required for a specific study product and may even provide a letter attesting that an IND is not required. However, for any investigator working in conjunction with a sponsor or a pharmaceutical company, or for all completely new materials, an IND is required.

## 6 INSTITUTIONAL REVIEW BOARD ISSUES

IRB review is not substantially different for initial prospective studies of drugs in human subjects than in other nonsafety studies. However, the use of a new pharmaceutical agent typically requires more careful consideration due to the potential for a higher level of risk to patient safety and the additional institutional risk. It is important that the IRB can determine exactly how safety will be evaluated and how adverse effects will be managed. The specific descriptive language in consent forms can be difficult to develop since the drug has never been used in human subjects. Care must be taken to explain this lack of human experience, and it must be clear to the IRB and to the subject who gives consent to participate in the study that the principal aim of the study is to test the drug's overall safety. It is important to state that this is the first time this drug has been administered to human subjects.

The timing of IRB approval and submission of the IND to the FDA can sometimes be challenging. The FDA requires a 30-day period to review an IND. During that period, the FDA can request additional information regarding the preclinical

physiology, pharmacology, and toxicology status of the pharmaceutical. In addition, the FDA may also request changes to the study design and protocol that has been submitted to the IRB for review. Thus, while IRB approval can be obtained concurrently with FDA evaluation of the IND, sometimes there is a substantial amount of interaction with the regulatory entities that require changes to protocols depending on the recommendations of both the IRB and FDA. In other words, if the original study design will evaluate safety data such as serum electrolytes at time 0 and 4 hours after drug administration, the FDA may request an additional blood draw at 24 hours. Approval to add this additional blood draw would have to be obtained as an amendment to the IRB protocol and may also require an additional revision to the consent form. Sometimes the IRB will not provide approval of the study until the FDA provides its "nonobjection" letter. Thus, it is sometimes helpful to know the FDA response prior to full submission to the IRB.

Any revisions that an IRB requires in the protocol or consent must also be submitted to the FDA as amendments. The amendment process is much different, though. For an amendment to be considered active, all that is required is IRB approval and submission to the FDA. Once these two requirements are met, the amendment can be considered to be in effect and future subjects can be studied according to the amended protocol.

However, if the investigator is conducting the study with a sponsor, the sponsor will act as the liaison between the investigator and the FDA. In clinical trails, the study cannot be initiated until the FDA, the IRB, and the sponsor have granted approval in writing for the study to begin.

## 7 STUDY DESIGN

All clinical studies must be performed under the current good manufacturing practices and good clinical practices guidelines (these practices are described by the guidelines produced through the International Conference on Harmonisation [3]). Further, they must be conducted with the requirements of appropriate ethical conduct as set by the Nuremberg Code [4], the Declaration of Helsinki [5], the 1962 Kefauver–Harris Amendments to the Food and Drug Act, and the Belmont Report [6]. Many of these guidelines are formalized in the Code of Federal Regulations [7].

The focus of safety studies is primarily to assess for all possible adverse physiological and clinical effects of the drug over the time period that the individual is exposed to the drug. The time period of exposure depends on the dose, dosing regimen, and half-life of the drug. For radiopharmaceuticals that require administration only on a single day, subjects are usually evaluated for only 1 or 2 days. Prolonged follow-up is not usually necessary since it is unlikely for the drug to have some effect days after it is in the subject's body. On the other hand, drugs that are administered over many days require evaluation over a similarly long time period. It may even be necessary to evaluate subjects months or even a year later. Some of the issues related to study design are described below in more detail.

## 8 SUBJECT SELECTION

The subjects usually recruited for phase I studies of safety are healthy controls. This provides the best evaluation of safety in the human body. Rigorous prescreening consisting of serum and urinalysis along with comprehensive medical histories can assist in recruiting a healthy control group that is not being treated for chronic medical or psychiatric conditions. In phase I studies conducted with an IND, healthy controls should have normal physiological measures and clinical status, thus making abnormal changes related to the new drug easier to detect. It is also ideal that subjects are not taking any medications, including over-the-counter drugs, since these could potentially interfere with the evaluation of physiological effects of the new drug or may interfere with its pharmacology. Such drug interaction assessments are usually made after the initial safety assessment in controls.

There are a number of circumstances in which healthy controls are not appropriate in the initial safety assessment of a drug. When a drug is specific for a target population, measuring safety in healthy individuals may not accurately reflect the safety of the drug in the manner in which it will be used. For example, a drug that targets Alzheimer's disease (AD) pathophysiology (i.e., the development of amyloid plaque) may not make sense to study in healthy individuals since they do not have the plaque formation in the first place. Thus, the drug will bind very differently in healthy controls compared to AD patients. If the physiological half-life of the drug in controls is 12 hours, but in AD patients is 1 week, then the safety profiles in the different populations may also be vastly different.

Another reason that controls might not be appropriate would be if the drug is expected to have substantial risks but is developed for a population in which such risks may be warranted. The classic example here is of chemotherapy agents. These drugs, which often have substantial effects on blood counts or electrolytes, cannot be tested in healthy individuals since the drugs are too dangerous with no benefit either to them or to healthy controls. In these cases, the target population is studied, usually with a dose escalation design in which the first small group of three to five subjects is studied with a relatively low dose. If the lower dose does not result in substantial adverse effects, then the next group of subjects is studied at an incremental increase in dose. The drug dose continues to increase until specific safety criteria thresholds are reached. Thus, if hematological toxicity reaches level 3, then the drug is not given in any higher doses. After the initial safety assessment, then an efficacy trial can be performed at the highest, safe, dose.

One problem with studying safety in disease populations is that many of these individuals are regularly taking one or more drugs that may affect safety and pharmacology of the new drug. It is not always feasible or safe to withdraw all medication to wash out existing drugs prior to the initiation of a new drug regimen. For example, many patients with cancer take a variety of medications designed to help with their cancer, their overall health and mental well-being, or other related problems. Care must be taken to exclude patients with specific medications that are expected to interfere with the test drug. This may be determined by the evaluation of other related drugs or simply based on knowledge of the physiological processes associated with the different drugs.

**TABLE 1 Laboratory Values Assessed for Safety Studies**

| |
|---|
| Cell blood count with differential (CBC) |
| Coagulation values (prothrombin time and partial thromboplastin time) |
| Prolactin and cortisol levels |
| Electrolyte panel, including Blood Urea Nitrogen (BUN) and creatinine |
| Thyroid function tests, including Thyroid Stimulating Hormone (TSH) |
| Follicle Stimulating Hormone (FSH) and Leutinizing Hormone (LH) |
| Urinalysis, routine |
| Liver function tests |
| Cholesterol |
| Urine toxicology screen |
| Autoimmune screening panel |
| Pregnancy test |

## 9 SAFETY MEASURES

Safety measures usually include both laboratory as well as clinical assessment. Laboratory values are typically evaluated at baseline and then at regular intervals spanning the subject's exposure to the test drug. For single-dose drugs, the evaluation may last only a couple of days while for drugs given over many days, the laboratory values must range over a similar time period. Table 1 shows typical laboratory values that are assessed in such studies. The most commonly used values include complete blood count, electrolytes, and liver function, but additional studies should be evaluated as determined through the results of preclinical animal studies. Values are compared to determine if substantial changes are associated with the administration of the test drug. If abnormalities are observed, then changes should be followed until they resolve. Usually, clear guidelines are established prior to the study that indicate what will be considered a drug-related change. In many cases, this requires an assessment of change rather than absolute values since a patient may experience a 10% drop in platelet count but still be in the normal range. This may ultimately lead to the need for evaluation of platelet counts prior to receiving the drug and if the counts are already low, limiting the use of the drug may be necessary.

In addition to laboratory values, physiological measures of heart rate, blood pressure, respiratory rate, and oxygen saturation may be useful in evaluating the safety of the test drug. Electrocardiography (EEG) is also important, especially when the drug may have specific cardiac effects. As with laboratory values, such measures might be made across the period during which the drug is administered. For a single-use drug such as a radiopharmaceutical imaging agent, physiological measures might be performed from a period immediately prior to drug administration up to one hour after administration. Subsequent measures might be made several hours later, 1 day, or even 2 days later. As with laboratory values, specific ranges of change might be described to clearly delineate what will and will not be considered a safety issue with the test drug.

A clinical evaluation including physical exam and report from the patient are also necessary at predefined intervals. The physical exam often includes evaluation of the skin (especially for an allergic reaction), hair, eyes, mucosa, heart, lungs, abdomen, and extremities. In addition, a neurological exam of strength and sensa-

tion may be useful. In addition to the psychological measures and laboratory values, the subjective experience of a study participant is also valuable. An ongoing dialog or direct questioning with the study participant should include whether they have the onset of any specific symptoms such as headache, dizziness, lightheadedness, gastrointestinal discomfort, nausea, chest or abdominal pain, weakness, and shortness of breath, sleep disturbance, or changes in appetite or mood.

It should also be noted that these extensive safety evaluations might be lessened after the initial experience with the test drug. Thus, after the initial safety evaluation, a more streamlined approach to safety measurement might be needed. In the example above of a single-dose radiopharmaceutical, eventual monitoring might only be required up to the first 10 or 20 min after administration.

## 10 MONITORING FOR ADVERSE EVENTS

Adverse events and serious adverse events have a very specific definition. Such events must be carefully evaluated and assessed to determine if the test drug is, in fact, related to the event. *Unexpected adverse drug experience*s are defined as any adverse drug experience, the specificity or severity of which is not consistent with the current investigator brochure; or, if an investigator brochure is not required or available, the specificity or severity of which is not consistent with the risk information described in the general investigational plan or elsewhere in the current application. Such adverse events may be minor unexpected laboratory changes or other clinical responses that are not anticipated. It is important that these adverse events be regarded and documented very specifically. For example, hepatic necrosis would be unexpected (by virtue of greater severity) if the investigator brochure only referred to elevated hepatic enzymes or hepatitis. It should also be clear that the term "unexpected" refers to an adverse drug experience that has not been previously observed rather than from the perspective of such experience not being anticipated from the pharmacological properties of the drug.

*Serious adverse drug experiences* are defined as any adverse drug experience occurring at any dose that results in any one of the following outcomes: Death, a life-threatening adverse drug experience, inpatient hospitalization or prolongation of existing hospitalization, a persistent or significant disability/incapacity, or a congenital anomaly/birth defect. A serious adverse drug experience may also be defined based upon appropriate medical judgment, as an event that may jeopardize the patient or subject and require medical or surgical intervention to prevent one of the outcomes listed above. Examples of such medical events include allergic bronchospasm requiring intensive treatment in an emergency room or at home. It is extremely important for the investigator to anticipate which effects are associated with a patient's clinical condition and which effects can be associated with the study drug or the rigors of the study procedures.

Reporting of adverse events is critical to any preclinical safety study. The requirements include that the sponsor (investigator) notify the FDA and all participating investigators in a written safety report of: (a) any adverse experience associated with the use of the drug that is both serious and unexpected or (b) any finding from tests in laboratory animals that suggests a significant risk for human subjects including reports of mutagenicity, teratogenicity, or carcinogenicity. Each notification

should be made as soon as possible and in no event later than 15 calendar days after the sponsor's initial receipt of the information and should be prominently identified as an "IND safety report."

The sponsor must also notify the FDA by telephone or by facsimile transmission of any unexpected fatal or life-threatening experience associated with the use of the drug as soon as possible but no later than 7 calendar days after the sponsor's initial receipt of the information. Each telephone call or facsimile transmission to the FDA should be transmitted to the FDA new drug review division in the Center for Drug Evaluation and Research or the product review division in the Center for Biologics Evaluation and Research that has responsibility for review of the IND.

## 11 PREPARING FOR PHASE II STUDIES AND ADDITIONAL SAFETY EVALUATION

Once the safety, pharmacology, and toxicology data is available from the initial human studies, this information is reported to the FDA as part of the next step to begin phase II studies. The phase II study protocols can be submitted under the same IND as amendments. The data from the phase I studies is typically included as new information that helps support the protocol to study the use of the drug in more specific patient populations and also in an expanded sample size.

The phase II trials will typically continue to evaluate safety, but on a lesser scale, and with more specific evaluations. Thus, if in the phase I study, electrolytes, liver function, cholesterol, and complete blood counts were evaluated with only the liver function studies showing any change, the phase II study may only include evaluation of liver function tests. If no laboratory abnormalities are observed in the phase I study, then no additional laboratory values may be necessary. Whatever changes to the safety evaluation of the drug in the phase II studies, these will ultimately have to be acceptable to the FDA and the IRB review committees.

## 12 CONCLUSIONS

Overall, the process of moving studies from the animal preclinical studies to human use studies is complex, but following the regulatory guidelines for the FDA IND process, the IRB, good manufacturing practices, and good clinical practices is imperative and relatively straightforward. The Code of Federal Regulations specifies the requirements for these studies and can be obtained in book or online formats. By following these requirements, the ability to propose and execute phase I studies can be readily performed by both industry and academic sponsors and investigators.

## REFERENCES

1. Schacter, B. Z. (2006), *The New Medicines: How Drugs Are Created, Approved, Marketed, and Sold*, Praeger, Westport, CT.
2. FDA. Center for Drug Evaluation and Research (CDER) (1997), Guidance for industry: M3 nonclinical safety studies for the conduct of human clinical rails for pharmaceuticals; available at http://www.fda.gov/cder/guidance/1855fnl.pdf.

3. International Conference on Harmonisation (1994), *ICH harmonized tripartite guideline: Clinical safety data management; definitions and standards for expedited reporting: International conference on harmonization of technical requirements for registration of pharmaceuticals for human use*, Geneva, Switzerland.
4. Nuremberg Military Tribunal (1949), *Nuremberg Code*, U.S. Government Printing Office, Washington, DC.
5. World Medical Association (2002), *Ethical Principles for Medical Research Involving Human Subjects*, World Medical Association, Ferney-Voltaire, France.
6. National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research (1979), The Belmont Report: Ethical principles and guidelines for the protection of human subjects of research, Department of Health, Education, and Welfare; available at http://ohrp.osophs.dhhs.gov/humansubjects/guidance/Belmont.htm.
7. *Code of Federal Regulations* (CFR) (2006), U.S. Government Printing Office, Washington, DC.

## ADDITIONAL SOURCE

FDA. Center for Drug Evaluation and Research (CDER) (1995), CBER. Guidance for industry: Content and format of investigational new drug application for Phase I studies of drugs, including well characterized therapeutic biotechnology-derived products; available at http://www.fda.gov/cder/guidance/phase1.pdf.

# *Predicting Human Adverse Drug Reactions from Nonclinical Safety Studies*

**Jean-Pierre Valentin,[1] Marianne Keisu,[2] and Tim G. Hammond[1]**

[1]*Safety Assessment, AstraZeneca, Macclesfield, Cheshire, United Kingdom*
[2]*Patient Safety, AstraZeneca, Södertälje, Sweden*

**Contents**

1 Background
  1.1 Reasons for Drug Attrition
  1.2 Frequently Used Definitions
  1.3 Data Availability
2 Assessment of Predictive Value of Nonclinical Safety Testing to Humans by Organ Systems
  2.1 Cardiovascular System
  2.2 Nervous System
  2.3 Respiratory System
  2.4 Gastrointestinal System
  2.5 Hepatic System
  2.6 Renal and Urinary System
  2.7 Endocrine System
  2.8 Hemopoietic System
  2.9 Immunological System
  2.10 Skin
3 Special Considerations
  3.1 Biologic and Biotechnology-Derived Pharmaceutical, Biopharmaceuticals, and Biotech Drugs

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

3.2 Genotoxicity
3.3 Genital System and Teratology
3.4 Safety Biomarkers
4 Summary and Future Challenges
4.1 New Targets and New Approaches to Treat Diseases
4.2 Science and Technology
4.3 Regulatory Requirements
4.4 Training and Development
References

## 1 BACKGROUND

### 1.1 Reasons for Drug Attrition

The reasons for drug attrition have evolved over the years; over the last decade, however, lack of safety (combining both nonclinical and clinical) remains the major cause of attrition during clinical development, accounting for approximately 35–40% of all drug discontinuation (see Table 1) [1–3]. More worrying is the fact that over the last few years despite a widened/increased testing battery there is no clear trend toward a reduction of the attrition due to safety reasons.

In this section, a brief summary of the nature, frequency, and consequences of adverse drug reactions (ADRs) in two clinical situations is presented. They are ADRs experienced by healthy volunteers and patients participating in clinical studies with potential new medical entities (NMEs) and those experienced by patients prescribed licensed medicines. A review of these two situations points to areas of success with the current practices for nonclinical safety testing but also identifies areas where further research might lead to new or better nonclinical safety testing.

Prior to reviewing the literature, it is worth considering some frequently used definitions. Differences can be found in the literature in how the concepts are defined. We have decided here to use the definitions published by the International Conference on Harmonisation (ICH) [4, 5], as those concepts are used (direct or with some adaptations) by many pharmaceutical companies and are recognized by regulatory agencies.

**TABLE 1 Evolution of Reasons for Drug Discontinuation in Clinical Development**

| Reasons | Percentage [2] | Percentage [3] | Percentage [1] |
|---|---|---|---|
| Portfolio considerations | 23 | 22 | 20 |
| Clinical efficacy | 22 | 23 | 26 |
| Clinical safety | 12 | 20 | 14 |
| Toxicology[a] | 23 | 19 | 21 |
| Other[b] | 20 | 16 | 19 |

[a]Includes general/safety pharmacology.
[b]Includes reasons such as clinical pharmacokinetics/bioavailability, nonclinical efficacy, nonclinical pharmacokinetics/bioavailability, formulation, patent, legal, or commercial, regulatory. Overall, safety reasons accounted for up to ~40% of all drug discontinuation.

## 1.2 Frequently Used Definitions

An *adverse event* (AE) can be defined as any unfavorable and unintended sign (including an abnormal laboratory finding), symptom, or disease temporally associated with the use of a medicinal product, whether or not considered related to the medicinal product.

An *adverse drug reaction* (ADR) is an adverse event where a causal relationship between the AE and the medicinal product is at least a reasonable possibility.

*Serious adverse events/adverse drug reactions* are defined as those that might be significant enough that if related to the medicinal product lead to important changes in the way the medicinal product is developed or used. This is particularly true for reactions that in their most severe forms threaten life or function.

The *severity*, which is the quantification of the reactions/symptoms (mild, moderate, severe), is used to describe grades of discomfort. One often used definition has been suggested by Tangrea et al. [6]: (i) mild, slightly bothersome, relieved with symptomatic treatment; (ii) moderate, bothersome, interferes with daily activities, only partially relieved with symptomatic treatment; and (iii) severe, prevents regular activities, not relieved with symptomatic treatment. A single serious ADR is always significant and very often has a high impact, its impact depending on when in the development process it occurs, and on the perceived benefit–risk profile of the NME. It can lead to the discontinuation of a drug in development, a significant limitation in the use of a drug (precaution, contraindication), or even to the withdrawal of the drug from the marketplace. A nonserious ADR can be more or less severe in its intensity, and its impact will depend upon its frequency and intensity. Nonserious ADRs can lead to a high degree of noncompliance, if they are perceived as annoying even if the symptoms they cause are not medically serious.

Pharmacological classification can divide ADRs in humans into five types (Table 2). Type A reactions (approx. 75% of all ADRs belong to this category) result from an exaggeration of the drug's normal pharmacological action when given in the usual therapeutic dose; they are normally dose dependent. Conventional pharmacology studies, combining primary, secondary, and safety pharmacology studies, can therefore reasonably be expected to predict type A ADRs. Functional toxicological measurements may predict type C ADRs. Conventional toxicology studies address type D ADRs, whereas prediction of type B responses (traditionally deemed as not predictable) occurring only in "susceptible" individuals require a more extensive

**TABLE 2 Classification of Adverse Drug Reactions (ADRs) in Humans**

| | | |
|---|---|---|
| Type A | Dose-dependent; predictable from primary, secondary, and safety pharmacology | Main cause of ADRs (~75%), rarely lethal |
| Type B | Idiosyncratic response, not predictable, not dose related, usually serious | Responsible for ~25% of ADRs, but majority of lethal ones |
| Type C | Long-term adaptive changes | Commonly occurs with some classes of drug |
| Type D | Delayed effects, e.g., carcinogenicity, teratogenicity | Low incidence |
| Type E | Rebound effects following discontinuation of therapy | Commonly occurs with some classes of drug |

*Source*: Adapted from Redfern et al. [7].

nonclinical and clinical evaluation to identify individuals at risk. The increased possibility to identify genetic risk factors will hopefully lead to a better understanding of this type of reactions [8]. The type E ADRs are rarely investigated nonclinically using functional measurements unless there is cause for concern.

### 1.3 Data Availability

The first human studies (phase I and early phase II) are generally very safe as a carefully selected population limit the potential for adverse effects (Table 3). In fact, molecules with a significant potential to generate serious ADRs are probably never given to healthy volunteers and are only given to patients where a significant benefit might be expected (e.g., refractory cancer patients) only with great care. These studies are to be conducted under frequent and extensive surveillance for the emergence of potentially worrisome ADRs. While AEs occur in phase I/healthy volunteer studies, they are generally more related to the required changes in lifestyle (e.g., caffeine deprivation) and the experimental procedures (e.g., needle puncture) than to the drugs [9, 10]. Nonclinical safety testing probably contributes significantly to the maintenance of this good track record; this is supported by published reports showing that single-dose nonclinical safety studies could overall accurately predict the clinical outcome [11–13]. The common ADRs observed with a high incidence (10–30%) during this early phase is linked to the gastrointestinal (nausea) and central nervous (headache) systems. In addition, an ADR specific to the NME under investigation that is pharmacologically mediated can be detected at this stage.

Early phase II studies are usually so-called dose-finding studies in selected patients and give the first indication of what type A reactions are to be expected in the targeted patient population. During continued clinical development as the population of exposed patients increase both in numbers and diversity, an increasing number of patients report AEs, with a wide variation in the type, frequency, and severity. Nonserious ADRs are often mechanism, drug class, or disease related (Table 4). Such ADRs limit the utility of an NME by restricting its use to those patients who either do not experience or can tolerate the ADRs, and they do not usually pose a "safety" issue. Serious ADRs tend to be present only at low frequencies. Pharmacological mechanism related to serious ADRs can occur in sensitive individuals, those with unusual kinetics, and in the presence of kinetic or occasion-

**TABLE 3 Risks to Healthy Volunteers in Phase I Clinical Research Trials**

| Year | Number of Volunteers | Moderately Severe AE | Potentially Life-threatening AE | Deaths | References |
|---|---|---|---|---|---|
| 1965–77 | 29,162 | 58 (0.2%) | — | 0 | 14 |
| 1980 | — | — | — | 1 | 15 |
| 1983 | — | — | — | 1 | 16 |
| 1984 | — | — | — | 1 | 17 |
| 1986–87 | 8,163 | 45 (0.55%) | 3 (0.04%) | 0 | 18 |
| 1986–95 | 1,015 | 43 (3%) | 0 | 0 | 9 |
| 2000 | — | — | — | 1 | 19 |
| 2006 | 6 | — | 6 (100%) | 0 | 20 |

*Source*: Adapted from Redfern et al. [7].

**TABLE 4 Major Causes of Acute Functional Adverse Drug Reactions**

| Acute Adverse Drug Reaction | Example |
|---|---|
| Augmented ("supratherapeutic") effect of interaction with the primary molecular target | Pronounced bradycardia with a β blocker; pronounced hypotension with an angiotensin II receptor antagonist |
| Interaction with the primary molecular target present in nontarget tissues | Sedation caused by anti-histamines |
| Interactions with secondary molecular targets | Interactions with the hERG cardiac channel leading to QT interval prolongation (e.g., some antipsychotics and antihistamines drugs) |
| Nonspecific effects | |
| Pharmacologically active metabolites | Inhibition of the hERG channel transcription by desmethyamiodarone, metabolite of amiodarone |

*Source*: Adapted from Redfern et al. [7].

ally dynamic drug interactions. In principle, such ADRs might be predictable from safety pharmacology testing, although it should be acknowledged that safety pharmacology testing is usually conducted in young adult healthy animals under conditions that may be suboptimal to detect such effects. Occasional nonpharmacological (type B) serious ADRs occur; these can be induced by direct chemical toxicity, hypersensitivity or immunological mechanisms. Serious ADRs always limit the use of an NME by requiring warnings, precautions, and contraindications or preclude regulatory approval (depending on the frequency, seriousness, and perceived benefit–risk balance of the NME). In addition to preventing the development of NMEs likely to induce certain types of serious ADRs, a key contribution that can be made by nonclinical safety testing is in the elucidation of the mechanisms responsible for these ADRs. Once the mechanism responsible for the ADR is known, it becomes possible to prepare soundly argued precautions and contraindications.

When medicines are on the market, the actual incidence of serious ADRs is difficult to judge. The main source today for judging the safety profile of a drug on the market comes from spontaneous reports (from health professionals and in some countries patients) to health authorities. Numerous publications discuss the limitations of spontaneous reporting systems because of, for example, the underreporting of serious ADRs and the difficulty to assess the frequency of a certain ADR due to lack of exposure data in such systems [21–23]. However, based on these data and on studies looking at ADRs in a defined population such as hospitalized patients it is clear that serious ADRs occur with sufficient frequency to be a serious concern [24]. One authoritative review concludes that between 1 in 30 and 1 in 60 physician consultations are caused by ADRs (representing 1 in 30–40 patients [25]). The same review concludes that 4–6% of hospital admissions could be the result of ADRs. Although there is debate about the number of deaths caused by ADRs, the figure of around 100,000 deaths per year in the United States is often quoted [7, 24]. The majority of the above references are about serious ADRs that could be predicted or avoided, only a subset are idiosyncratic [26, 27].

The frequency of a serious ADR can be very low (e.g., 0.25–1.0 cases of rhabdomyolysis per 10,000 patients treated with a statin [28]; however, when millions of patients are under treatment, this can generate substantial morbidity. Serious ADRs may also be due to clinical error (e.g., misprescribing contraindicated drugs) or to

patient self-medication error—especially in the era of mass media communication and information. Independent of the cause of a certain serious ADR, it is important to investigate the pharmacological mechanisms driving these events. For example, the elucidation of the connection between drug-induced Torsades de Pointes, QT interval prolongation, and hERG potassium channel blockade has been considered as a major advance in this area, leading to avoidance of NMEs that have the potential to cause QT interval prolongation due to cardiac arrhythmias and sudden death by the rapid development of nonclinical in vitro screening assays of medium to high throughput capabilities [29].

To better understand the main causes for ADR-related drug withdrawals, medicines withdrawn from either the United States or worldwide markets were reviewed [25, 30]. The principal reasons, presented in Table 5, highlighted the fact that several of these toxicities fall into the remit of nonclinical safety testing, with the usual suspects ranking top of the list such as toxicities related to the cardiovascular, hepatic, and nervous systems. The prominence of arrhythmias in Stephens' review [25] probably reflects the recent interest in Torsades de Pointes type of arrhythmias.

Limitations of the data set are now considered. In most cases, hard evidence regarding the predictive value of nonclinical testing is not readily available in the public domain [12, 31–38]. When an NME has a serious negative effect in a nonclinical test, there may be limited information in the public domain, as the result may have precluded clinical development, and therefore the information was never published. If there has been no effect in a nonclinical test, and likewise no effect on the corresponding variable in humans, these negative data may have been deemed not to be of general interest. Thus there is a huge amount of unpublished information that might be of value for better understanding certain types of adverse effects. In addition, the scientific and medical community is left with examples of side effects in humans that were not identified during nonclinical assessment.

One publication attempted to explore the predictive value of safety/general pharmacology assays to humans [12]. Some significant correlations were reported. For example, decreased locomotor activity in rodents was positively correlated with dizziness and sleepiness in humans; decreased intestinal transit in rodents was correlated with constipation and anorexia in humans; decreased urinary and sodium excretions in the rat were correlated with edema in humans; decreased blood pressure in dogs was positively correlated to flushes, dizziness, headache, and malaise in humans; increased heart rate in dogs was correlated to palpitation; and increased blood flow in dogs was correlated to flushes and headache. Rather more bizarrely,

**TABLE 5 Evolution of Main Safety Reasons for Drug Withdrawal over Last 40 Years**

| Worldwide Withdrawal (121 medicines) [30] | Percentage | U.S. Withdrawal (95 medicines) [25] | Percentage |
|---|---|---|---|
| Hepatic toxicity | 26 | Cardiovascular toxicity incl. arrhythmias | 19 (12) |
| Hematological toxicity | 10 | Neuropsychiatric effects/abuse liability/ dependency | 12 |
| Cardiovascular toxicity | 9 | Hepatic toxicity | 9 |
| Dermatological effects | 6 | Bone marrow toxicity | 7 |
| Carcinogenicity issues | 6 | Allergic reactions | 6 |

the findings of analgesia, decreased body temperature, and anticonvulsive activity in rodents were each correlated with thirst in humans; the finding of pressure reflex to vagal stimulation was correlated with sleepiness, malaise, and thirst in humans. This indicates the limitations of such methods and the fact that correlations between events do not necessarily imply causal relationships.

There are, however, numerous examples of drugs that cause AEs in humans, which would be detectable in nonclinical safety evaluation assays. Some notable examples of individual drugs showing untoward effects in nonclinical studies that are correlated in a quantitative sense with AEs in humans have been reported [7]. For example, (a) the sedative effects of clonidine in various animal species and humans; (b) the propensity of cisapride to prolong ventricular repolarization; (c) the respiratory depressant effects of morphine; (d) the nephrotoxic effect of cyclosporine; and (e) the gastrointestinal effects of erythromycin. These examples illustrated the very good agreement of effects across all species tested and across a narrow range of doses, concentrations, or exposures.

However, there are areas that should be carefully considered to ensure optimization of the assays and ultimately increase the predictive value of nonclinical testing. These include, but are not limited to: (a) species differences in the expression or functionality of the molecular target mediating the adverse effects; (b) differences in pharmacokinetic properties between test species and humans; (c) sensitivity of the test system; (d) optimization of the test conditions; (e) appropriately statistically powered study designs; (f) appropriate timing of functional measurements in relation to the time of maximal effect, maximal exposure, and/or maximal tissue concentration; (g) delayed/chronic effects of parent drug and/or metabolites; (h) difficulty of detection in animals in standard nonclinical safety studies (e.g., arrhythmia, headache); and (i) assessment of a suboptimal surrogate endpoint that predict with some degree of confidence the clinical outcome (e.g., QT/QTc interval prolongation as a surrogate of TdP). Overall careful consideration should be paid to the sensitivity, specificity, and overall predictivity of nonclinical assays prior to using them for building an integrated risk assessment.

## 2 ASSESSMENT OF PREDICTIVE VALUE OF NONCLINICAL SAFETY TESTING TO HUMANS BY ORGAN SYSTEMS

### 2.1 Cardiovascular System

Over the last few years, data have been generated to assess the value of nonclinical tests to predict the potential of NMEs to prolong the duration of the QT interval of the electrocardiogram and ultimately the proarrhythmic potential of these drugs. The published data converged in that an integrated risk assessment based upon data on the potency against hERG, an in vivo repolarization assay, and if necessary an in vitro action potential assay are in a qualitative sense predictive of the clinical outcome [39–42]. These data have been further supported by publications suggesting that a 30-fold margin between the highest free plasma concentration of a drug in clinical use ($C_{max}$) and the concentration inhibiting by 50% the hERG current ($IC_{50}$) could be adequate to ensure an acceptable degree of safety from arrhythmogenesis with a low risk of obtaining false positives [43–45]. Most recently, Wallis et

al. [46] have reported that, based on a data set of 19 compounds, combining data from the hERG assay and the in vivo QT repolarization assay predicted in 90% of cases the clinical outcome of the "thorough QT/QTc study." In addition, the author suggested that a robust electrocardiographic monitoring during the phase I clinical trials combined with the preclinical data predicted in all cases the outcome of the thorough QT/QTc study. These data support the view that a dedicated thorough QT/QTc study as required per ICH E14 [47] does not add any additional value to the integrated QT risk assessment. These preliminary data would require to be strengthened with a larger data set. Such initiative is ongoing under the auspice of the ILSI-Health and Environmental Sciences Institute (HESI). Along the same lines, in a review of 25 anticancer drugs, Schein et al. [48] showed that based on general clinical assessment and pathology rather than on functional changes both primates and dogs predicted cardiovascular toxicities in 90% of cases in which the toxicities were observed in humans (Table 6). The authors suggested that physiological measurement monitoring could have further improved the overall predictivity to the clinical outcome. Furthermore, Olson et al. [13] showed good concordance between cardiovascular findings in dogs and humans. The translation was less robust between humans and rodents possibly due to the technical challenges associated with monitoring of cardiovascular function in rodents. Furthermore, based on a data set of 88 drugs, Igarashi et al. [12] showed good concordance between some pharmacological findings in animal models and their associated clinical adverse drug reactions (e.g., blood pressure reduction versus flushes, dizziness, headache, malaise; increase in heart rate versus palpitation and increase in blood flow versus flushes and headache) [12].

**TABLE 6 Species and Assays Predictive Value of Humans' Safety Endpoints**

| Safety Endpoint | Specie and/or Assay | Number of Compound | Sensitivity (%) | Specificity (%) | Predictivity (%) | References |
|---|---|---|---|---|---|---|
| Injection site | Dog | 25 | 67 | 53 | 56 | 48 |
| | Monkey | 23 | 59 | 67 | 71 | 48 |
| Integument | Dog | 25 | 43 | 56 | 52 | 48 |
| | Monkey | 23 | 50 | 77 | 76 | 48 |
| Cardiovascular | Species combined | 150 | | | 80 | 13 |
| | Dog | 25 | 70 | 60 | 64 | 48 |
| | Monkey | 23 | 50 | 54 | 57 | 48 |
| Respiratory | Dog | 25 | 80 | 20 | 32 | 48 |
| | Monkey | 23 | 59 | 38 | 47 | 48 |
| Nervous system | Species combined | 150 | | | 61 | 13 |
| | Rodent | 21 (7) | | | 33 | 55 |
| | Dog | 21 (7) | | | 29 | 55 |
| | Monkey | 21 (7) | | | 66 | 55 |
| Bone marrow | Rodent | 21 (13) | | | 75 | 55 |
| | Dog | 25 | 100 | 0 | 80 | 48 |
| | Dog | 21 (13) | | | 85 | 55 |
| | Monkey | 23 | 95 | 0 | 90 | 48 |
| | Monkey | 21 (13) | | | 100 | 55 |
| Lymphoid | Dog | 25 | 100 | 25 | 28 | 48 |
| | Monkey | 23 | 0 | 68 | 71 | 48 |

**TABLE 6** ***Continued***

| Safety Endpoint | Specie and/or Assay | Number of Compound | Sensitivity (%) | Specificity (%) | Predictivity (%) | References |
|---|---|---|---|---|---|---|
| Gastrointestinal | Species combined | 150 | | | 83 | 13 |
| | Rodent | 21 (13) | | | 82 | 55 |
| | Dog | 25 | 100 | 0 | 92 | 48 |
| | Dog | 21 (13) | | | 92 | 55 |
| | Monkey | 23 | 81 | 0 | 80 | 48 |
| | Monkey | 21 (13) | | | 86 | 55 |
| Hepatic | Species combined | 150 | | | 52 | 13 |
| | Rodent | 21 (6) | | | 83 | 55 |
| | Dog | 25 | 100 | 8 | 56 | 48 |
| | Dog | 21 (6) | | | 100 | 55 |
| | Monkey | 23 | 100 | 27 | 71 | 48 |
| | Monkey | 21 (6) | | | Not tested | 55 |
| Renal and urinary | Species combined | 150 | | | 68 | 13 |
| | Rodent | 21 (3) | | | 100 | 55 |
| | Dog | 25 | 80 | 7 | 36 | 48 |
| | Dog | 21 (3) | | | 100 | 55 |
| | Monkey | 23 | 90 | 21 | 52 | 48 |
| | Monkey | 21 (3) | | | 100 | 55 |
| Neuromuscular | Dog | 25 | 86 | 17 | 36 | 48 |
| | Monkey | 23 | 71 | 57 | 66 | 48 |
| QT interval prolongation | Human—hERG in vitro[a] | 19 | 82 | 75 | 79 | 46 |
| QT interval prolongation | Dog in vivo[a] | 19 | 83 | 86 | 85 | 46 |
| QT interval prolongation | Dog Purkinje fiber[a] | 19 | 20 | 100 | 50 | 46 |
| QT interval prolongation | Combine nonclinical assay[a] | 19 | 90 | 88 | 89 | 46 |
| Torsades de pointes | Rabbit—in vitro | 64 | 65 | 89 | 75 | 72 |
| Visual function | Zebrafish | 37 | 71 | 78 | 59, 73 | |
| Seizure | Zebrafish | 25 | 76 | 63 | 72 | 58 |

[a]Within twofold the free therapeutic plasma concentration. (Values in parentheses indicate) number of compounds showing toxicity in humans. The sensitivity was determined as the ratio of true positive (true positive/false negative). A high sensitivity reflects a low level of false positives. The specificity was determined as the ratio of true negative/true negative plus false positive. A high specificity reflects a low level of false negatives. The predicitvity was determined as the ratio of true positive plus true negative/the total number of compounds evaluated.

## 2.2 Nervous System

Most of the adverse drug effects relating to the nervous system impact on the quality of life rather than the risk to life (e.g., lethargy, anorexia, insomnia, personality changes, nausea). There are, however, some serious life-threatening adverse effects involving the nervous system (e.g., loss of consciousness and convulsions). Some of them reflect the fact that the central nervous system (CNS) controls the other two vital organ systems for that CNS impairment could be fatal (e.g., decreased respira-

tory drive leading to respiratory arrest; decreased sympathetic outflow leading to cardiovascular collapse). The nervous system adjusts the function of the other acutely vital organ systems according to current and long-term requirements of the organism. Therefore, drug effects on cardiovascular and respiratory functions can be mediated via a direct action within the CNS or via sensory nerve endings located in the cardiovascular and pulmonary systems. Some CNS adverse effects can be indirectly life threatening. For example, drowsiness, cognitive impairment, motor coordination, dizziness, involuntary movement, and visual or auditory disturbances can all affect driving performance; moreover, depression or personality changes can lead to suicidal tendencies. As an illustration, the number of deaths in the United States between 1984 and 1996 in patients receiving terfenadine was 396, a small proportion of which were attributed to sudden death resulting from Torsades de Pointes. This overall low incidence of fatalities is nevertheless a significant improvement over the first generation of antihistamine drugs that have been suspected to be responsible for significant fatalities in car accidents resulting from their sedative effects [49, 50]. General pharmacological tests for effects on the nervous system are usually observational studies of rodent general activity or multidimensional functional assays of motor activity [51–53]. For a series of 84 new drugs (excluding anticancer agents) studied in Japan, an evaluation of their capacity to predict adverse reactions in humans showed a general nonspecific correlation. For example, decreased locomotor activity in rodents was positively correlated with dizziness and sleepiness in humans [12]. A degree of overprediction was reported, particularly from studies that used high doses. Similarly, in the study of 45 miscellaneous drugs by Fletcher [54], high-dose effects such as ataxia and convulsions in animals did not occur in humans, and subjective effects such as dizziness, headache, dry mouth and sweating in humans were not predicted by animal studies. The correlation was stronger for other effects on the central nervous system.

Where effects on the central nervous system have been assessed in conventional toxicity studies using both clinical monitoring and histopathological examination of the brain and nervous tissue, a reasonable degree of concordance has been shown. Evaluation of the effects of up to 25 diverse anticancer drugs in dogs, monkeys, and humans showed a reasonable degree of concordance (nearly 40%) in neurological and neuromuscular toxic effects ([48]; Table 6). Dogs and monkeys had similar predictive values and high doses were needed to achieve the best correlation, whereas specific symptoms correlated poorly. The earlier study of 21 anticancer drugs by Owens indicated only a moderate correlation between neurotoxic effects in humans and animals [55]. The correlation was strong for alkylating agents but less so for other classes of drugs studied. Interestingly, the study of 150 miscellaneous drugs by Olson and colleagues [13, 56] showed that, overall, the nonrodent data were better correlated with adverse neurological effects in humans than the rodent data. A key neurological safety liability is the potential for drug-induced seizure (see Table 7); in this regard it is encouraging to note that the proconvulsive potential of marketed drugs was detected in a zebrafish larvae assay with a significant level of sensitivity, specificity, and overall predictivity [58] (Table 6).

So, while the data indicate poor prediction of subjective neurological effects, the information on the significant toxicities of anticancer drugs indicates that the conventional approach using histopathological examination detects potentially serious neurotoxic effects.

**TABLE 7 Incidence of Adverse Drug Reactions Related to Each of the Major Physiological Functions**

| Physiological System | Percentage of Compounds Reporting ADR with Incidence >3% (%) |
|---|---|
| Cardiovascular | 35 |
| Arrhythmias | 3 |
| Central & peripheral nervous system | 56 |
| Seizure | 3 |
| Gastrointestinal | 67 |
| Motility | 44 |
| Hepatic | 11 |
| Immune | 0 |
| Renal & urinary | 17 |
| Respiratory | 32 |

*Note*: Data extracted from BioPrint [57]. Based on a set of 1138 drugs annotated for human ADRs. Not all compound-ADR annotations include incidence data; figures have been corrected to account for missing incidence data.

Certain types of side effects to the CNS (e.g., agitation, hallucinations, headache) are reported with some medicines. Such effects are not easily identifiable in animal studies.

***Special Senses*** Relatively few instances of visual, auditory, or vestibular disturbances are reported in early clinical studies with new drugs. As such, there is a paucity of data comparing effects on these functions between laboratory animals and humans. However, ophthalmoscopic examination is usually performed in toxicity studies and is routinely accompanied by histopathological examination of the eye, prior to dosing of humans with a new drug. Consequently, it is probable that any agent that provokes severe ocular damage in animals after relatively short periods of dosing would not progress to clinical studies. Moreover, agents that have potent cataractogenic properties or that are severely toxic to the retina would be identified in relatively short, repeat-dose studies. Emerging data suggest that performing an optomotor assay in either the zebrafish [59] or in rodents [60] could detect drug-induced impairment of visual function with a reasonable sensitivity and specificity toward the clinical outcome. Specific tests of auditory function are seldom done routinely. But careful clinical observation of animal comportment probably eliminates agents that produce acute and severe auditory or vestibular damage.

## 2.3 Respiratory System

A large body of data has accumulated on experimental methodology for examination of the effects of environmental and occupational chemicals on the respiratory tract. This is because inhalation is a primary mode of human exposure to foreign materials [61]. However, for inhaled drugs, there is limited information available in the public domain; part of the difficulty is that any significant respiratory side effects would be considered as unacceptable and consequently not progressed into humans. So we are probably left only with the circumstance of having taken a compound into humans, seeing an adverse effect, and then attempting to find suitable animal

models that could have predicted it in order to screen against it in the future. Some companies working in this area have expressed concerns over the ability to predict potential for "cough" reactions or bronchospasm. The effects of drugs are evaluated preclinically in dedicated safety pharmacology studies [62]; until the implementation of the ICHS7A [62], this was usually done alongside cardiovascular assessment in anesthetized dogs. In the comparison of 104 investigational new drugs by Igarashi and colleagues [12] in which this approach was used, respiratory disturbance was not frequently reported in humans. However, when respiratory related ADRs were reported in humans, these were not predicted by safety pharmacology testing.

In toxicity studies, effects on respiration are usually evaluated by clinical observation and histopathology of lungs and air passages, although it is usually recognized that clinical observation are "inappropriate to assess drug effects on respiratory function" [62]. In the study of 45 drugs by Fletcher [54], both toxicology and pharmacology animal studies overpredicted respiratory effects in humans. Similarly, Schein and colleagues [48] noted that this form of screening in nonrodents predicted respiratory signs or respiratory pathology in four out of five cases, but with a high percentage of false positives (Table 6).

Nowadays respiratory function is usually assessed in rodents using whole-body plethysmography, which provides an indication of drug effect on ventilatory parameters. Dennis Murphy [63] has argued that drug effect on both ventilatory parameters and lung mechanics should be assessed prior to first administration to humans since changes in lung mechanic (e.g., bronchoconstriction) may remain undetected/unpredicted using whole-body plethysmography methods.

### 2.4 Gastrointestinal System

Clinical gastrointestinal adverse reactions account for ~18% of total ADRs, and 20–40% in hospitalised patients. There is a high degree of underreporting [64]. Diarrhea alone accounts for about 7% of all drug ADRs. For example, more than 700 drugs are implicated in causing diarrhea [65]. The majority are functional in nature (nausea, vomiting, dyspepsia, abdominal cramps, and diarrhea or constipation) and fewer are pathological (e.g., ulceration) or enhanced susceptibility to infection (e.g., pseudomembranous colitis) [66]. Overall, ~80% of gastrointestinal (GI) ADRs are predictable type A pharmacological reactions [67], that is, predictable from the primary and/or secondary pharmacological targets. From 1960 to 1999 two drugs withdrawn due to GI toxicity—Indoprofen and Pirprofen nonsteroidal anti-inflammatory drugs (NSAIDs) [30]. NSAID use in the United States alone is estimated to be responsible for over 100,000 hospitalisations and 17,000 deaths per year [68]. More recently, a prospective analysis of 18,820 United Kingdom patients; 17 deaths attributed to GI ADRs; most attributed to NSAID use [69].

The review of safety pharmacology studies performed in Japan on 88 noncancer drugs showed a good correlation between rodent intestinal transport and general adverse effects such as anorexia and constipation in humans [12]. In the review of conventional toxicology studies that included histopathology of the gastrointestinal tract, Olson and colleagues [13] showed good concordance between gastrointestinal effects in animals and humans, particularly for nonsteroidal anti-inflammatory drugs, anti-infective, and anticancer agents. In that review, large animal data were

a better predictor than data obtained from rodents. The data also showed good correlation between animal toxicology studies and humans for a diverse set of 45 drugs [47] and may be a link to the fact that large number of drugs are associated with gastrointestinal ADRs, thus increasing the sensitivity of detection level.

The rodent, dog, monkey, and human GI toxicity data also showed a strong correlation in the study of 21 anticancer drugs by Owens [55]. Surprisingly, in the study of 25 anticancer drugs by Schein [48], the dog was superior to the monkey as a predictor of adverse GI effects in humans (Table 6). For example, monkeys were remarkably resistant to vomiting, an adverse event that was observed in humans with 21 of the 25 compounds. Gastrointestinal tract toxicity was a significant contributor to the remarkably good quantitative correlation of toxicity across species based on dose/body surface area for the 18 anticancer drugs studies by Freireich and colleagues [70]. This is not surprising since oncology drugs tend to be used at maximum tolerated doses at which gastrointestinal side effects are quite common.

It has been suggested that the GI tract of dogs is highly physiologically similar to that of humans in terms of motility patterns, gastric emptying, and pH, particularly in the fasted state [71]. This observation, coupled with the ability to use a formulation similar to that used in humans, makes the canine GI tract a most relevant model.

### 2.5 Hepatic System

Hepatotoxicity is an important adverse drug effect and a relatively common reason for termination of the development of an NME [13, 73]. At present, drug-induced hepatic injury accounts for more than 50% of cases of acute liver failure in the United States [74]. In conventional nonclinical studies of toxicity, the cornerstone of the assessment of hepatotoxic potential is measurement of circulating liver enzymes and hepatic histopathology [75]. A review of 38 chemicals, 24 of which were drugs that produce hepatic toxicity in humans, showed a concordance of 80% with findings in conventional toxicity studies [76]. Hepatic toxicity was not underpredicted in the study of 25 anticancer drugs in dogs and monkeys that used conventional hepatic enzyme measurements and histopathology ([48]; Table 6). The study of anticancer drugs by Owens [55] showed a similar good correlation.

Conversely, the study of data on 150 drugs exhibiting human toxicities showed that the concordance between hepatotoxicity found in animal studies and that observed in clinical practice was little more than 50% [13]. This larger study undoubtedly included agents that developed idiosyncratic responses, which are not usually detected in early clinical trials because of their rarity. This is a significant problem—in recent years there have been notable examples of hepatic toxicity of a poorly understood, idiosyncratic nature that have caused the withdrawal of marketed drugs despite extensive and essentially negative nonclinical testing and large clinical trials. The thiazolidinedione troglitazone, an antidiabetic drug, was associated with serious hepatic injury in patients despite its lack of hepatic toxicity in preclinical studies [77]. Another example is bromfenac, a nonsteroidal anti-inflammatory drug [74]. Hepatic failure also occurred in clinical trials with the nucleoside analog fialuridine as a result of mitochondrial disturbance and steatosis. Despite long-term treatment of monkeys, dogs, and rats with fialuridine, the only hepatic effects observed were

increases in apoptosis and nuclear atypia in rats [77]. Ximelagatran was found to cause liver enzyme elevations such as alanine aminotransferase (ALT) in 7.9% of patients treated in long-term studies (>35 days). Routine nonclinical animal data did not show any liver enzyme elevations. Ximelagtran and its metabolites were tested in an extensive study program including cell viability, mitochondrial function, formation of reactive metabolites, and reactive oxygen species without finding an explanation to the mechanism behind the ALT elevation [78].

It is probable that most NMEs that produce severe hepatotoxicity in animals are not tested in humans, so that the true level of concordance is likely to remain obscure. However, overall, the data seem sufficiently robust to conclude that overt liver damage observed in animal toxicity studies indicates potential risk of hepatic toxicity in humans. This underlines the prudence of a critical histopathological examination of the liver tissue in nonclinical studies and careful patient monitoring in response to any hepatic alerts from animal studies.

### 2.6 Renal and Urinary System

Renal toxicity is assessed by conventional histopathology, measurement of blood urea and electrolytes, and examination of urine volume and the sediment it contains. Concordance in the database of 150 drugs reviewed by Olsen and colleagues [13] was fair. A good correlation was noted among the 21 anticancer drugs reviewed by Owens [55], with rodents and dogs performing equally well. However, in a review of 45 drugs, renal toxicity was correctly predicted by animal studies in 3 instances but overpredicted in 22 others [54]. Similarly, the study of 25 anticancer drugs in dogs or primates correctly predicted renal toxicity in 9 cases, underpredicted in 1, and overpredicted in 14 [48].

### 2.7 Endocrine System

Endocrine changes during nonclinical studies are routinely assessed only by histological examination of endocrine organs, unless there are particular reasons to suspect endocrine effects. Olson and colleagues [13] noted only moderate concordance (60%) between animals and humans. As might be expected from the way in which the endocrine system responds to stimuli, these effects were not common in humans and generally occurred after phase I studies. The review by Fletcher [54] indicated that endocrine findings in nonclinical studies significantly overpredict effects in humans. Endocrine effects, particularly those involving the adrenal gland, are commonly reported in toxicity studies [79]. These findings often represent adaptive alterations to repeated doses of drugs and usually manifest as changes in glandular weight and cellular atrophy or hypertrophy. These changes might not have significant implications for human safety in single-dose studies, but they characterize possible endocrine effects that need to be assessed in clinical trials.

### 2.8 Hemopoietic System

Hemopoiesis is routinely assessed by examination of peripheral blood, bone marrow smears, and histopathology of the blood-forming and lymphoid organs. Theus and Zbinden [80] reviewed prior industry practice for the assessment of coagulation in

1984 and found substantial deficiencies. The screening practice that they proposed for animal studies is similar to that used in humans and has now been almost universally adopted for pharmaceutical testing.

There is substantial data on the concordance of adverse effects on hemopoietic tissue due to anticancer and antimitotic drugs between animals and humans. The evidence indicates good correlation for both rodents and humans and dogs and humans for myelotoxicity, although the particular cell series affected sometimes differs [48]. Thrombocytopenias were correctly predicted for 13 of 18 anticancer drugs that produced this toxicity in humans. Moreover, in the series of 18 anticancer drugs studied by Freireich and colleagues, hemopoietic toxicity was one of the most significant contributors to the remarkably good quantitative correlation across species based on dose/body surface area [70].

A reasonable correlation between animals and humans was also noted for decreases in white blood cell counts in the study of 139 drugs by the Japanese Pharmaceutical Manufacturers Association [12]. Anticancer drugs and antibodies did predominate in this series, but the authors also detected a considerable number of false negatives and false positives in their data.

### 2.9 Immunological System

Specific tests of immune function are not routinely performed for conventional new drugs prior to their use in humans. An international collaborative study showed that examination of peripheral blood white cells, histological examination of thymus and spleen, and, in particular, careful histological examination of lymphoid tissue in the rat is a good primary method of identifying agents that are significant direct-acting immunotoxins [81]. New screens of immune function in animals are sometimes proposed for drug assessment [82, 83], but more sophisticated tests of immune function might be more appropriate and safely conducted in human studies. Coping with the potential impact of biotechnology-derived pharmaceuticals on immune status and immunogenicity is a special challenge for which careful attention to the principles of immunology is needed [84].

### 2.10 Skin

Of all tissues, skin shows the least concordance between effects in animal studies and human patients. A general lack of predictive reliability for skin reactions in humans has been noted in the reviews of anticancer and other drugs ([48, 54, 55]; Table 6). Adverse skin hypersensitivity effects have caused the development of a relatively large number of potential NMEs to be terminated [13, 73, 85].

## 3 SPECIAL CONSIDERATIONS

### 3.1 Biologic and Biotechnology-Derived Pharmaceutical, Biopharmaceutical, and Biotech Drugs

The definitions of biology-related therapeutic agents and examples of agents are presented in Table 8; there are a number of important differences between biologics

**TABLE 8 Definitions of Biology-Related Therapeutic Agents**

| Term | Definition | Examples |
|---|---|---|
| Biologic | A therapeutic agent derived from a biological source or produced using a biological process. For the purposes of this chapter, this will include drugs obtained by extraction from human or animal tissues and body fluids or produced by biotechnological means (recombinant or hybridoma technology).<br>It will include cell-, virus-, and bacteria-based products as well as DNA-based therapeutics. It will NOT include certain small-molecule agents produced by fermentation (e.g., antibiotics) or plant-derived products (botanicals). | Plasma-derived proteins (albumin, clotting factors, immunoglobulin preparations)<br>Tissue-derived proteins (animal insulins, human pituitary-derived growth hormone)<br>Vaccines (live, attenuated, or killed whole-cell preparations, recombinant peptide vaccines, DNA, and viral vector vaccines)<br>Recombinant peptides and proteins<br>Monoclonal antibodies<br>Products derived from transgenic animals or plants<br>Gene therapy and related tissue or cell therapy products<br>Antisense drugs |
| Biotechnology-derived pharmaceutical, biopharmaceutical, biotech drug[a] | Any drug produced by biotechnological means, such as recombinant or hybridoma technology, are generally regarded to include recombinant peptides or proteins, antibodies, genetically modified tissue, and cell-based products as well as DNA-based products. | Vaccines (as above)<br>Recombinant proteins or peptides and derivatives (e.g., erythropoietin, clotting factors, insulins, or insulin analogs)<br>Monoclonal antibodies and antibody fragments (e.g., Fabs)<br>Transgenic animal products (e.g., rh albumin)<br>Gene therapy products (as above) |

[a]These terms are basically interchangeable and broadly refer to the same type of therapeutic agents.

and conventional small molecules that need to be understood to design effective nonclinical and clinical programs for these products. They refer principally to the production process, molecular weight and composition, microheterogeneity, species specificity, pleiotropism, dose-level selection, metabolism, and catabolism [86, 87]. The available regulatory guidance on the nonclinical safety testing of biologics has recently been reviewed [88]. Although there are a number of product-class-specific guidelines that have been published over the years (for review see Ryle [87]), the most relevant general guidance on the nonclinical testing of biologics is the ICH S6 [89]. The guideline emphasizes the need for a flexible, case-by-case approach and appropriate species selection, with special attention paid to immunologically mediated effects in animals and their relevance to human patients. Some of the special considerations that need to be taken into account in the nonclinical testing of biologics include the source and quality of the test material (in particular batch-to-batch variability), species selection, immunogenicity (in particular after repeated dosing), dosing schedule, dose levels and study durations, immunomodulation, and the assessment of local tolerance. The serious adverse events observed with the CD28 agonist antibody TGN-1412 in a phase I study in early 2006 ([20]; Table 3) demonstrate how difficult it is to lay out any general examples of typical nonclinical

programs for biologics. Recent investigations on the mechanisms of the TGN-1412-mediated "cytokine storm" will enable the development of novel procedures to improve nonclinical safety testing of immunomodulatory therapeutics [90]. Meanwhile this serious adverse event has led to evolving regulatory expectations for this type of product [91].

### 3.2 Genotoxicity

There has been much contention about the relevance of genotoxicity assays to the testing of pharmaceutical agents since their introduction more than 30 years ago [92]. However, extensive study has led to a better understanding of the chemical determinants that provoke genotoxic effects through electrophilic attack of biological macromolecules [93]. As a consequence of this understanding, mutagenic activity is often simply avoided in the drug discovery process with the exception of certain classes of drugs aimed at treating cancer (for review see Guzzie-Peck [86]). Nevertheless, prior to first human exposure, in vitro tests for mutations and chromosomal damage are routinely carried out according to internationally agreed technical guidelines that are based on a large body of historical data for diverse chemicals. However, it can be difficult to assess human risk when unexpected or unexplained activity in these bacterial or mammalian cell tests occurs. Such activity usually precludes dosing to healthy volunteers at least until further work elucidates the mechanism of activity and characterizes any hazard. Subsequently, in vivo assays of bone marrow micronucleus, peripheral blood cytogenetics, and liver unscheduled DNA (deoxyribonucleic acid) synthesis in rodents are usually done. The technical performance of these tests has also been the subject of international collaborative studies. In silico approaches are now used routinely and tend to supersede in vitro testing of genotoxic potential.

An area of growing regulatory concern is the assessment of potentially genotoxic impurities of pharmaceuticals. The ICH and the European Medicine Agency (EMEA) have published guidance documents focusing on the safety evaluation of impurities in pharmaceutical drug substances and drug products [94–100]. The EMEA guidance is based on a threshold of toxicological concern (TTC) derived from animal carcinogenicity data using multiple worst-case assumptions to estimate a daily dose associated with a lifetime cancer risk of 1 in 100,000, a risk level considered acceptable for genotoxic impurities in human medicines. Based on these assumptions, presentation of the TTC as a single figure infers an unwarranted level of precision and supports the adoption of a more flexible approach by regulatory authorities when evaluating new drug products; a range within five fold of the TTC limit would be sensible. Furthermore, the limit is based on 70 years continuous daily exposure, a scenario that is uncommon for most medicines and irrelevant to the preregistration clinical development phase. To address this later point a staged TTC has been developed that proposes limits based on shorter durations of treatment (e.g., up to 1-year). Based on recent history, this approach has been acceptable by some authorities but not others, and it is imperative that steps are taken to reach a common agreement between the pharmaceutical industry and regulatory authorities worldwide in order that new medicines can continue to be developed and delivered to benefit patients in a safe and timely manner [101].

### 3.3 Genital System and Teratology

Reproductive changes are seldom reported in early clinical trial studies, largely due to the exclusion of women of child-bearing potential from these studies [54]. In a report from the Federal Institute for Occupational Safety and Health, several reviews as well as studies on individual compounds have been analyzed with respect to the suitability of different study designs and endpoints to detect effects on male reproduction in animal species [102]. However, only a few studies were available that characterize the human situation. A considerable inter- and intraindividual variability was noted with respect to key parameters associated with fertility in men (e.g., sperm count, motility, morphology, and volume). Interspecies extrapolation factors were derived from the most sensitive endpoint in laboratory animals. Despite the small database and limitations of the studies that prevented any robust conclusion, it was felt that humans were generally not more susceptible to reproductive toxicants than laboratory animals as was originally assumed. For the purpose of hazard identification a subacute study exploring concentrations that produce significant general toxicity might be sufficient. If effects were found, for the purpose of risk assessment the no-adverse-effect level has to be identified by testing sensitive endpoints. Although a subchronic study was felt preferable, a subacute study may be sufficient. Similar observations emerged from a collaborative study in Japan of 16 drugs, 12 of which were associated with infertility in humans; results showed that histopathological endpoint was the most sensitive method for preclinical detection of drugs with antifertility properties [103]. A recommendation from the Federal Institute for Occupational Safety and Health report was to develop and validate a rabbit model allowing sequential sperm analysis and better observation of behavior [102].

Teratology, the study of abnormal prenatal development and congenital malformations induced by exogenous chemical and physical agents is being primarily assessed using in vivo approaches. Although the interspecies concordance and extrapolation to humans is in large part unknown, mainly because teratogenic compounds are either not progressed into humans or progressed under very restricted usage. Although, animal studies in mammalians remain the gold standard, other alternative methods/assays are being explored. In this regard, it is worth mentioning three in vitro methods endorsed as scientifically validated by the European Centre for In Vitro Alternative Methods (ECVAM) in 2001 [104]: the embryonic stem cell test, the micromass test, and the whole-embryo culture. More recently, the use of the zebrafish opens some promising avenues [105].

### 3.4 Safety Biomarkers

The field of safety biomarkers (SBMs) is advancing rapidly. Our improved understanding of the molecular bases of organ toxicity suggests that monitoring specific molecular responses may provide improved prediction of human outcomes and in doing so provide "bridging SBMs" that may eliminate much of the current uncertainty in extrapolating from laboratory models to human outcome. Modern high-throughput technologies for proteins or endogenous metabolites offer a major opportunity to systematically identify sensitive and specific plasma or urine SBMs that could serve as an index of damage specific to each of the important

internal tissues and organs. SBMs can serve many decision-making purposes. Depending on where SBMs are used in the drug discovery/development process they demand different levels of validation and qualification. SBMs can be deployed for (i) target related toxicity, (ii) CD family-related toxicity, (iii) unexpected toxicity during GLP studies, and (iv) unexpected toxicity during clinical development. In accepting an SBM as qualified for regulatory decision-making, FDA has exhibited clearly that it will operate only within a broad scientific consensus. This means that:

- Data on an SBM from different investigative methods should be convergent and support a single hypothesis for its role in a certain organ toxicity.
- There are no or few data that are incompatible with the hypothesis.
- The data available in support of an SBM are persuasive to independent expert peer groups.
- The data should have originated from multiple and independent investigations in several laboratories.

A suitable way to reach this consensus is to share the interest in developing biomarkers in consortia formed by, e.g., pharma and diagnostic companies, academia, and regulators.

## 4 SUMMARY AND FUTURE CHALLENGES

### 1 New Targets and New Approaches to Treat Diseases

Advances in molecular biology and biotechnology allow for the identification of new molecular targets, leading to the discovery and development of newer pharmaceutical agents that act at these novel molecular sites in an attempt to ameliorate the disease condition. Moreover, new therapeutic approaches are being developed (e.g., gene therapy, biotech products) that offer new challenges to assessing their safety in humans. Inherent in the novelty of new targets and new approaches is the risk of unexpected and unwanted effects that may or may not be detected based on current scientific knowledge and with current techniques and assays.

One of the biggest challenges for the biotechnology and pharmaceutical companies in the twenty-first century will be to develop and deliver drugs that fit the individual patient's biology and pathophysiology ("personalized medicine") [106, 107]. In addition, new developments in therapeutic approaches, such as those involving biopharmaceuticals and gene therapy, offer promise in the treatment or prevention of diseases for which current approaches are ineffective.

At the same time health care costs are increasing dramatically in many countries, and aging populations often require treatment with multiple drugs. In addition increased drug development costs and recent high-profile issues relating to drug safety highlight the need for finding new medicines with acceptable safety profiles while minimizing development cost.

There is a need to identify/screen out drug candidates with poor safety profiles as early as possible in the drug discovery process, and nonclinical safety assessment functions have an important role to play in achieving this goal [108].

To increase the overall value of the nonclinical assays would require to better understand their predictive value for humans; to achieve this would require reviewing analysis data available in the public domain as well as proprietary information and sharing the outcome of such effort widely. This could ideally be achieved via consortia involving academic institutions, regulatory agencies, and the pharmaceutical industry. Such approaches have been recently initiated [109, 110].

## 2 Science and Technology

Nonclinical safety evaluation faces significant scientific challenges to keep pace, to adapt, and to incorporate new technologies in the evaluation of NMEs in nonclinical assays/models and identifying the effects that pose a risk to human volunteers and patients. Recent examples have included the use of electrophysiological techniques to evaluate the effects of NMEs on the hERG channel [111] and telemetry techniques to assess the effects of NMEs on the duration of the QT interval in unstressed animals (for review see McMahon et al. [29]), therefore enabling to establish integrating QT risk assessment reliable and predictive of the clinical outcome. The development and utilization of technologies and approaches that have a direct clinical correlate should also be encouraged; for example, the utilization of echocardiography to assess drug effect on ejection fraction can be applied in both a nonclinical and clinical setting. Furthermore, efforts should continue to construct databases relating to the predictive value of nonclinical assays to humans either through retrospective analysis or through purposely designed studies [13, 39–42, 57, 72].

### 4.3 Regulatory Requirements

Continued benefit–risk assessment is today at the center of drug development and drug life-cycle maintenance together with the requirement to submit risk management plans to regulatory agencies together with licensing applications [112, 113]. These requirements have helped pharmaceutical companies to develop cross disciplinary working to ensure identification, assessment, and better understanding of safety risks and to devise risk minimization activities. Wherever feasible conditional approvals based on fewer patients and with rigorous prospective safety follow-up should be considered. In its March 2004 Critical Path Report [114], the Food and Drug Administration (FDA) suggests that limited exploratory IND investigations in humans (phase 0) can be initiated with less, or different, nonclinical support, that is required for traditional IND studies because exploratory IND (e-IND) studies should present fewer potential risks than do traditional phase I studies that look for dose-limiting toxicities. The nonclinical program should be considered on a case-by-case basis depending on the specific objectives for a given e-IND.

### 4.4 Training and Development

Disciplines involved in nonclinical safety evaluation of NMEs face significant challenges of attracting, training, and certifying investigators to ensure the future of these disciplines. The paucity of training in certain biomedical scientific disciplines (toxicology, pathology, pharmacology, physiology) has had detrimental long-lasting effects such as (a) an impact on the development of intact animal models of human

function and disease; (b) an impact on skills to conceptualize biomedical hypothesis and experiments at the level of the intact animal; and (c) an impact on the process of nonclinical and clinical drug discovery and development. There is a clear need to ensure all parties involved in the training, education, and development of individuals working in these disciplines work together to ensure continuous supply of these key skills.

Developing organized, planned, and prospective methods for clinical safety data review to interpret and act on this data will require integrated databases, possibilities to search historical data, and possibilities to bring information together from multiple sources in a new way would enable assessing the concordance between preclinical and clinical safety data and ultimately refine the nonclinical safety testing strategies to select the safest candidate drugs.

## REFERENCES


1. Kola, I., and Landis, J. (2004), Can the pharmaceutical industry reduce attrition rates? *Nature Rev./Drug Disc.*, 3, 711–715.
2. Kennedy, T. (1997), Managing the drug discovery/development interface, *Drug Disc. Dev.*, 2, 436–444.
3. Lasser, K. E., Allen, P. D., Woolhandler, S. J., Himmelstein, D. U., Wolfe, S. M., and Bor, D. H. (2002), Timing of new black box warnings and withdrawals for prescription medications, *JAMA*, 287, 2215–2220.
4. Anon. (1996), ICH E2C(R1) Harmonised Tripartite Guideline, Clinical Safety Data Management: Periodic Safety Update Reports for Marketed Drugs. CPMP/ICH/288/95.
5. Anon. (2003), ICH E2D Harmonised Tripartite Guideline, Post-approval Safety Data Management: Definitions and Standards for Expedited Reporting. CPMP/ICH/3945/03.
6. Tangrea, J. A., Adrianza, M. E., and McAdams, M. (1991), A method for the detection and management of adverse events in clinical trials, *Drug Inf. J.*, 25, 63–80.
7. Redfern, W. S., Wakefield, I. D., Prior, H., Hammond, T. G., and Valentin, J. P. (2002), Safety pharmacology—A progressive approach, *Fund. Clin. Pharmacol.*, 16, 161–173.
8. Wilke, R. A., Lin, D. W., Roden, D. M., Watkins, P. B., Flockhart, D., Zineh, I., Giacomini, K. M., and Krauss, R. M. (2007), Identifying genetic risk factors for serious adverse drug reactions: Current progress and challenges, *Nature Rev./Drug Dis.*, 6, 904–916; available at: www.nature.com/reviews/drugdisc.
9. Sibille, M., Deigat, N., Janin, A., Kirkesseli, S., and Durand, D. V. (1998), Adverse events in phase-I studies: A report in 1015 healthy volunteers, *Eur. J. Clin. Pharmacol.*, 54, 13–20.
10. Rozenzweig, P., Brohier, S., and Zipfel, A. (1993), The placebo effect in healthy volunteers: Influence of experimental conditions on the adverse events profile during phase I studies, *Clin. Pharmacol. Thera.*, 54(5), 578–583.
11. Greaves, P., Williams, A., and Eve, M. (2004), First dose of potential new medicines to humans: How do animals help? *Nature Rev./Drug Disc.*, 3, 226–236.
12. Igarashi, T., Nakane, S., and Kitagawa, T. (1995), Predictability of clinical adverse reactions of drugs by general pharmacology studies, *J. Toxicol. Sci.*, 20, 77–92.

13. Olson, H., Betton, G., Robinson, D., Thomas, K., Monro, A., Kolaja, G., Lilly, P., Sanders, J., Sipes, G., Bracken, W., Dorato, M., Van Deun, K., Smith, P., Berger, B., and Heller, A. (2000), Concordance of the toxicity of pharmaceuticals in humans and in animals, *Regul. Toxicol. Pharmacol.*, 32(1), 56–67.
14. Zarafonetis, C. J., Riley, P. A., Willis, P. W., et al. (1978), *Clin. Pharmacol. Ther.*, 24, 127–132.
15. Kolata, G. B. (1980), The death of a research subject, *Hastings Center Report*, 10, 5–6.
16. Daragh, A., Kenny, M., Lambe, R., and Brick, I. (1985), Sudden death of a volunteer, *Lancet*, I 93–94.
17. Anon. (1985), Editorial. Death of a volunteer, *BMJ*, 290, 1369–1370.
18. Orme, M., Harry, J., Routledge, P., and Hobson, S. (1989), Healthy volunteer studies in Great Britain: The results of a survey into 12 months activity in this field, *Br. J. Clin. Pharmacol.*, 27, 125–133.
19. McCarthy, M. (2001), Healthy volunteer dies in US physiology study, *Lancet*, 357, 2114.
20. Anon. (2006), Expert Scientific Group on Phase I Clinical Trials. Interim report. Duff, G. W. (chairman); accessed November 4, 2006 at: http://www.dh.gov.uk/assetRoot/04/13/75/69/04137569.pdf.
21. Alvarez-Requejo, A., Carvaajal, A., Vega, T. L., and Bégaud, B. (1994), Undereporting of adverse drug reactions in a Spanish regional centre of pharmacovigilance, *Drug Inf. J., Abstr.* 249 (Suppl. 1), S104.
22. Bégaud, B., Martin, K., Haramburu, F., and Moore, N. (2002), Rates of spontaneous reporting of adverse drug reactions in France, *JAMA*, 288(13), 1588.
23. Brewer, T., and Colditz, G. A. (1999), Postmarketing surveillance and adverse drug reactions: Current perspectives and future needs, *JAMA*, 281, 824–829.
24. Lazarou, J., Pomeranz, B. H., and Corey, P. N. (1998), Incidence of adverse drug reactions in hospitalized patients: A meta-analysis of prospective studies, *JAMA*, 279, 1200–1205.
25. Stephens, M. D. B. (2004), Introduction, in Talbot, J., and Waller, P., Eds., *Stephens' detection of new adverse drug reactions*, 5th ed., John Wiley & Sons, Chichester, UK, pp. 1–91.
26. Avery, A. A., Taylor, R. L., Partidge, M., Neil, K., et al. (2001), Investigating preventable drug-related admissions to a medical admissions unit, *Pharmacoepidemiol. Drug. Saf.*, 10(S103), 243.
27. Bhalla, N., Duggan, C., and Dhillon, S. (2003), The incidence and nature of drug-related admission to hospital, *Phara. J.*, 270, 583–586.
28. McKenney, J. M. (2005), Pharmacologic options for aggressive low-density lipoprotein cholesterol lowering: Benefits versus risks, *Am. J. Cardiol.*, 96(4A), 60E–66E.
29. McMahon, N., Pollard, C., Hammond, T. G., and Valentin, J. P. (2007), Cardiovascular safety pharmacology, in Sietsema, W. K., and Schwen, R., Eds., *Nonclinical Drug Safety Assessment—Practical Considerations for Successful Registration*, FDA News, Washington, DC, 87–123.
30. Fung, M., Thornton, A., Mybeck, K., Wu, J. H., Hornbuckle, K., and Muniz, E. (2001), Evaluation of the characteristics of safety withdrawal of prescription drugs from worldwide pharmaceuticals markets—1960 to 1999, *Drug Inform. J.*, 35, 293–317.
31. Calabrese, E. J. (1984), Suitability of animal models for predictive toxicology: Theoretical and practical considerations, *Drug Metab. Rev.*, 15, 505–523.

32. Garratini, S. (1985), Toxic effects of chemicals: Difficulties in extrapolating data from animals to man, *Ann. Rev. Toxicol. Pharmacol.*, 16, 1–29.
33. Grieshaber, C. K., and Marsoni, S. (1986), Relation of preclinical toxicology to findings in early clinical trials, *Cancer Treat. Rep.*, 70, 65–72.
34. Lumley, C. E., and Walker, S. R. (1985), The value of chronic animal toxicology studies of pharmaceutical compounds: a retrospective analysis, *Fund. Appl. Toxicol.*, 5, 1007–1024.
35. Lumley, C. E., Parkinson, C., and Walker, S. R. (1992), An international appraisal of the minimal duration of chronic animal studies, *Human Exp. Toxicol.*, 11, 155–162.
36. Monro, A., and Mehta, D. (1996), Are single-dose toxicology studies in animals adequate to support single dose of a new drug in humans? *Clin. Pharmacol. Ther.*, 59, 258–264.
37. Oser, B. L. (1981), The rat as a model for human toxicological evaluation, *J. Toxicol. Environ. Health*, 8, 521–642.
38. Zbinden, G. (1994), Predictive value of animal studies in toxicology, *Regul. Tox. Pharm.*, 14, 167–177.
39. Ando, K., Hombo, T., Kanno, A., Ikeda, H., Imaizumi, M., Shimizu, N., Sakamoto, K., Kitani, S., Yamamoto, Y., Hizume, S., Nakai, K., Kitayama, T., and Yamamoto, K. (2005), QT PRODACT: in vivo QT assay with a conscious monkey for assessment of the potential for drug-induced QT interval prolongation, *J. Pharmacol. Sci.*, 99, 487–500.
40. Miyazaki, H., Watanabe, H., Kitayama, T., Nishida, M., Nishi, Y., Sekiya, K., Suganami, H., and Yamamoto, K. (2005), QT PRODACT: Sensitivity and specificity of the canine telemetry assay for detecting drug-induced QT interval prolongation, *J. Pharmacol. Sci.*, 99, 523–529.
41. Omata, T., Kasai, C., Hashimoto, M., Hombo, T., and Yamamoto, K. (2005), QT PRODACT: Comparison of non-clinical studies for drug-induced delay in ventricular repolarization and their role in safety evaluation in humans, *J. Pharmacol. Sci.*, 99, 531–541.
42. Sasaki, H., Shimizu, N., Suganami, H., and Yamamoto, K. (2005), QT PRODACT: Inter-facility variability in electrocardiographic and hemodynamic parameters in conscious dogs and monkeys, *J. Pharmacol. Sci.*, 99, 513–522.
43. De Bruin, M. L., Pettersson, M., Meyboom, R. H. B., Hoes, A. W., and Leujkens, H. G. M. (2005), Anti-hERG activity and the risk of drug-induced arrhythmias and sudden death, *Eur. Heart J.*, 26, 590–597.
44. Webster, R., Leischmann, D., and Walker, D. (2002), Towards a drug concentration effect relationship for QT prolongation and torsades de pointes, *Curr. Opin. Drug Disc. Develop.*, 5, 116–126.
45. Redfern, W. S., Carlsson, L., Davis, A. S., Lynch, W. G., MacKenzie, I., Palethorpe, S., Siegl, P. K. S., Strang, I., Sullivan, A. T., Wallis, R., Camm, A. J., and Hammond, T. G. (2003), Relationship between preclinical cardiac electrophysiology, clinical QT interval prolongation and torsade de pointes for a broad range of drugs: Evidence for a provisional safety margin in drug development, *Cardiovas. Res.*, 58, 32–45.
46. Wallis, R. (2007), QT-understanding the complexities of defining the translation of animal data to humans. 7th Annual Meeting of the Safety Pharmacology Society, Edinburgh, September 20–21.
47. Anon. (2005), ICH E14: The clinical evaluation of QT/QTc interval prolongation and proarrhythmic potential for non-antiarrhythmic drugs. London, 25 May 2005. CPMP/ICH/2/04. http://www.emea.eu.int/pdfs/human/ich/000204en.pdf.

48. Schein, P. S., Davis, R. D., Carter, S., Newman, J., Schein, D. R., and Rall, D. P. (1970), The evaluation of anticancer drugs in dogs and monkeys for the prediction of qualitative toxicities in man, *Clin. Pharmacol. Ther.*, 11, 3–40.
49. Cimbura, G., Lucas, D. M., Bennett, R. C., Warren, R. A., and Simpson, H. M. (1982), Incidence and toxicological aspects of drugs detected in 484 fatally injured drivers and pedestrians in Ontario, *J. Forensenic Sci.*, 27, 855–867.
50. Weiler, J. M., Bloomfield, J. R., Woodworth, G. G., Grant, A. R., Layton, T. A., Brown, T. L., McKenzie, D. R., Baker, T. W., and Watson, G. S. (2000), Effects of fexofenadine, diphenhydramine, and alcohol on driving performance. A randomised, placebo-controlled trial in the Iowa driving simulator, *Ann. Intern. Med.*, 132, 354–363.
51. Mattsson, J. L., Spencer, P. J., and Albee, R. R. (1996), A performance standard for clinical and functional observation battery examination of rats, *J. Am. Coll. Toxicol.*, 15, 239.
52. Irwin, S. (1968), Comprehensiive observational assessment: 1a. A systematic, quantitative procedure for assessing the behavioural and physiologic state of the mouse, *Psychopharmacologia (Berl.)*, 13, 222–257.
53. Haggerty, G. C. (1991), Strategies for and experience with neurotoxicity testing of new pharmaceuticals, *J. Am. Coll. Toxicol.*, 10, 677–687.
54. Fletcher, A. P. (1978), Drug safety tests and subsequent clinical experience, *J. Roy. Soc. Med.*, 71, 693–696.
55. Owens, A. H. (1962), Predicting anticancer drug effects in man from laboratory animal studies, *J. Chron. Dis.*, 15, 223–228.
56. Olson, H., Betton, G., Robinson, D., Thomas, K., Monro, A., Kolaja, G., Lilly, P., Sanders, J., Sipes, G., Bracken, W., Dorato, M., Van Deun, K., Smith, P., Berger, B., and Heller, A. (2000), Concordance of the toxicity of pharmaceuticals in humans and in animals, *Regul. Toxicol. Pharmacol.*, 32(1), 56–67.
57. Krejsa, C. M., Horvath, D., Rogalski, S. L., Penzotti, J. E., Mao, B., Barbosa, F., and Migeon, J. C. (2003), Predicting ADME properties and side effects: The BioPrint approach, *Curr. Op. Drug Dis. Dev.*, 6(4), 470–480.
58. Winter, M. J., Redfern, W. S., Hayfield, A. J., Owen, S. F., Valentin, J-P., and Hutchinson, T. H. (2008), Validation of a zebrafish larval locomotor activity screen for assessing the seizure-liability of early-stage development drugs, *J. Pharmacol. Toxicol. Methods.*, 57, 176–187.
59. Richards, F. R., Alderton, W. K., Kimber, G. M., Liu, Z., Strang, I., Redfern, W. S., Valentin, J-P., Winter, M. J., and Hutchinson, T. H. (2008), Validation of the use of WIK and TL strain zebrafish larvae for visual safety assessment, *J. Pharmacol. Toxicol. Methods.*, 58, 50–58.
60. Maung, K. P., Storey, S., McKay, J., Bigley, A., Heathcote, D., Elliott, K., Valentin, J-P., Hammond, T. G., and Redfern, W. S. (2008), Validation of an optometry system for measurement of visual acuity in Han Wistar rats, *J. Pharmacol. Toxicol. Methods.*, 58, 152.
61. Nemery, B., Dinsdale, D., and Verschoyle, R. D. (1987), Detecting and evaluating chemical-induced lung damage in experimental animals, *Bull. Eur. Physiopathol. Respir.*, 23, 501–528.
62. Anon. (2000), ICHS7A: Safety pharmacology studies for human pharmaceuticals. CPMP/ICH/539/00; accessed on November 16, 2000 at: http://www.emea.eu.int/pdfs/human/ich/053900en.pdf.
63. Murphy, D. J. (2005), Comprehensive non-clinical respiratory evaluation of promising new drugs, *Toxicol. Appl. Pharmacol.*, 1, 207(2 Suppl), 414–424.

64. Lewis, J. H. (1986), Gastrointestinal injury due to medicinal agents, *Am. J. Gastroenterol.*, 81, 819–834.
65. Chassany, O., Michaux, A., and Bergmann, J. F. (2000), Drug-induced diarrhoea, *Drug Saf.*, 22, 53–72.
66. Ghahremani, G. G. (1999), Gastrointestinal complications of drug therapy, *Abdom. Imaging*, 24, 1–2.
67. Gatenby, R. A. (1995), The radiology of drug-induced disorders in the gastrointestinal tract, *Semin. Roentgenol.*, 30, 62–76.
68. Whittle, B. J. (2003), Gastrointestinal effects of nonsteroidal anti-inflammatory drugs, *Fundam. Clin. Pharmacol.*, 17, 301–313.
69. Pirmohamed, M., James, S., Meakin, S., Green, C., Scott, A. K., Walley, T. J., Farrar, K., Park, B. K., and Breckenridge, A. M. (2004), Adverse drug reactions as cause of admission to hospital: Prospective analysis of 18,820 patients, *BMJ*, 329, 15–19.
70. Freireich, E. J., Gehen, E. A., Rall, D. P., Schmidt, L. H., and Skipper, H. E. (1966), Quantitative comparison of toxicity of anticancer agents in mouse, rat, hamster, dog, monkey and man, *Cancer Chemother. Rep.*, 50, 219–244.
71. Dressman, J. B. (1986), Comparison of canine and human gastrointestinal physiology, *Pharmcol. Res.*, 3, 123–131.
72. Lawrence, C. L., Bridgland-Taylor, M. H., Pollard, C. E., Hammond, T. G., and Valentin, J-P. (2006), A rabbit Langendorff heart proarrhythmia model: Predictive value for clinical identification of torsade de pointes, *Br. J. Pharmacol.*, 149(7), 845–860.
73. Lumley, C. (1990), in Walker, S. R., Ed., *Animal Toxicity Studies: Their Relevance for Man*, Quay, Lancaster, pp. 49–56.
74. Lee, W. M. (2003), Drug-induced hepatoxicity, *N. Engl. J. Med.*, 349, 474–485.
75. Amacher, D. E. (1998), Serum transaminase elevations as indicators of hepatic injury following administration of drugs, *Regul. Toxicol. Pharmacol.*, 27, 119–130.
76. Hayes, A. W. et al. (1982), Correlation of human hepatotoxicants with hepatic damage in animals, *Fund. Appl. Toxicol.*, 2, 55–66.
77. Schwartz, S., Raskin, P., Fonseca, V., and Graveline, J. F. (1998), Effect of troglitazone in insulin-treated patients with type II diabetes mellitus, *N. Engl. J. Med.*, 338, 861–866.
78. Kenne, K., Skanberg, I., Glinghammar, B., Berson, B., Pessayre, D., Flinois, J-P., Beaune, P., Edebert, I., Diaz Pohl, C., Carlsson, T., and Andersson, T. B. (2007), Prediction of drug induced liver injury in humans by using in vitro methods: The case of ximelagatran, *Toxicol. In Vitro*, 22(3), 730–746.
79. Ribelin, W. E. (1984), The effects of drugs and chemicals upon the structure of the adrenal gland, *Fund. Appl. Toxicol.*, 4, 105–119.
80. Theus, R., and Zbinden, G. (1984), Toxicological assessment of the hemostatic system, regulatory requirements, and industry practice, *Regul. Toxicol. Pharmacol.*, 4, 74–95.
81. Dayan, A. D. et al. (1998), Report of a validation study of assessment of direct immunotoxicology in the rat, *Toxicology*, 125, 183–201.
82. Dean, J. H. (1997), Issues with introducing new immunotoxicology methods into the safety assessment of pharmaceuticals, *Toxicology*, 119, 95–101.
83. Dean, J. H., Hinks, J. R., and Remander, B. (1998), Immunotoxicology assessment in the pharmaceutical industry, *Toxicol. Lett.*, 102–103, 247–255.
84. Cavagnaro, J. A. (2002), Preclinical safety evaluation of biotechnology-derived pharmaceuticals, *Nature Rev. Drug Disc.*, 1, 469–475.
85. Lichfield, J. T. (1961), Forecasting drug effects in man from studies in laboratory animals, *JAMA*, 177, 104–108.

86. Crommelin, D. J. A., Storm, G., Verrijk, R., de Leede, L., Jiskoot, W., and Hennink, W. E. (2003), Shifting paradigms: Biopharmaceuticals versus low molecular weight drugs, *Int. J. Pharm.*, 266, 3–16.

87. Ryle, P. R. (2007), Special considerations in the preclincial testing of biologics: The ICHS6 guideline, in Sietsema, W. K., and Schwen, R. Eds., *Nonclinical Drug Safety Assessment—Practical Considerations for Successful Registration*, FDA News, Washington, DC, pp. 301–330.

88. Snodin, D. J., and Ryle, P. R. (2006), Understanding and applying regulatory guidance on the nonclinical development of biotechnology-derived pharmaceuticals, *Biodrugs*, 20, 25–52.

89. Anon. (1997), Preclinical Safety Evaluation of Biotechnology-derived Pharmaceuticals. ICH Harmonised Tripartite Guideline S6. Geneva: International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use, CPMP/ICH/302/95.

90. Stebbings, R., Findlay, L., Edwards, C., Eastwood, D., Bird, C., North, D., Mistry, Y., Dilger, P., Liefooghe, E., Cludts, I., Fox, B., Tarrant, G., Robinson, J., Meager, T., Dolman, C., Thorpe, S. J., Bristow, A., Wadhwa, M., Thorpe, R., and Poole, S. (2007), "Cytokine Storm" in the Phase I Trial of Monoclonal Antibody TGN1412: Better Understanding the Causes to Improve PreClinical Testing of Immunotherapeutics, *J. Immunol.*, 179, 3325–3331.

91. Anon. (2007), Guideline on requirements for first time in man clinical trials for potential high-risk medicinal products, London, March 22, 2007. EMEA/CHMP/SWP/28367/2007.

92. Clive, D. (1985), Mutagenicity tests in drug development: Interpretation and significance of test results, *Regul. Toxicol. Pharmacol.*, 5, 79–100.

93. Benigni, R., and Zito, R. (2003), Designing safer drugs: (Q)SAR-based identification of mutagens and carcinogens, *Curr. Topics Med. Chem.*, 3, 1289–1300.

94. Guzzie-Peck, P. (2007), Genotoxicity testing and risk management, in Sietsema, W. K., and Schwen, R., Eds., Nonclinical Drug Safety Assessment—Practical Considerations for Successful Registration, FDA News, Washington, DC, pp. 197–272.

95. Anon. (2005), Guideline on the non-clincial investigation of the dependence potential of medicinal products, London, April 21, 2005. EMEA/CHMP/SWP/94227/2004.

96. Anon. (2006), EMEA Guideline on the Limits of Genotoxic Impurities. Committee for Medicinal Products for Human Use. The European Medicines Agency, London. CPMP/SWP/5199/02. EMEA/CHMP/QWP/251344/2006.

97. Anon. (2006), HMPC concept paper on the development of a guideline on the assessment of genotoxic constituents in herbal substances/preparations. Committee on Herbal Medicinal Products European Medicines Agency, EMEA/HMPC/413271/2006.

98. Anon. (2006), ICH Q3A(R2) Impurities in new drug substances, in *International Conference on Harmonisation Harmonised Tripartite Guideline*. Current Step 4 version 25; available at: http://www.ich.org/LOB/media/MEDIA422.pdf.

99. Anon. (2006), ICH Q3B(R2) Impurities in new drug products, in *International Conference on Harmonisation Harmonised Tripartite Guideline*. Current Step 4 version 2; available at: http://www.ich.org/LOB/media/MEDIA421.pdf.

100. Anon. (2005), ICH Q3C(R3) Impurities: Guideline for residual solvents, in *International Conference on Harmonisation Harmonised Tripartite Guideline*. Current Step 4 version.

101. Humfrey, C. D. N. (2007), Recent developments in the risk assessment of potentially genotoxic impurities in pharmaceutical drug substances, *Toxicol. Sci.*, 100(1), 24–28.

102. Mangelsdorf, I., and Buschmann, J. (2002), *Extrapolation from Results of Animal Studies to Humans for the Endpoint Male Fertility*. Project F1642, Federal Institute for Occupational Safety and Health, Dortmund, Germany.
103. Takayama, S., Akaike, M., Kawashima, K., Takahashi, M., and Kurokawa, Y. (1984), A collaborative study in Japan on optimal treatment period and parameters for detection of male fertility disorders induced by drugs in rats, *Regul. Toxicol. Pharmacol.*, 14, 266–292.
104. Bremer, S., Pellizzer, C., Adler, S., Paparella, M., and de Lange, J. (2002), Development of a testing strategy for detecting embryotoxic hazards of chemicals in vitro by using embryonic stem cell models, *ATLA*, 30(Suppl. 2), 107–109.
105. Gustafson, A. L., Weiser, T., Clemann, N., Hossaini, A., Janaitis, C., Bluemel, J., Delongeas, J. L., and Hill, A. (2008), *Validation of zebrafish as a model for screening teratogenicity*. Annual Meeting of the Society of Toxicology, Washington, D.C., March.
106. Frueh, F. W., and Gurwitz, D. (2004), From pharmacogenetics to personalized medicine: A vital need for educating health professionals and the community, *Pharmacogenomics*, 5, 571–579.
107. Ginsburg, G. S., and Angrist, M. (2006), The future may be closer than you think: A response from the Personalized Medicine Coalition to the Royal Society's report on personalized medicine, *Personlized Med.*, 3(2), 119–123.
108. Lesson, P. D., and Springthorpe, B. (2007), The influence of drug-like concepts on decision-making in medicinal chemistry, *Nature Rev. Drug Disc.*, 6(11), 881–890.
109. FDA (2007), FDA's Response to the Institute of Medicine's 2006 Report. The Future of Drug Safety—Promoting and Protecting the Health of the Public. U.S. Department of Health and Human Services Food and Drug Administraion (FDA), January.
110. Anon. (2006), The Innovative Medicines Initiative (IMI) Strategic Research Agenda. Creating Biomedical R&D Leadership for Europe to Benefit Patients and Society; available at: http://www.efpia.org/4_pos/SRA.pdg.
111. Bridgland-Taylor, M. H., Hargreave, A. C., Easter, A., Orme, A., Harmer, A., Henthorn, D. C., Ding, M., Davis, A., Small, B. G., Heapy, C. G., Abi-Gerges, N., Paulsson, F., Jacobson, I., Schroeder, K., Neagle, B., Alberston, N., Hammond, T. G., Sullivan, M., Sullivan, E., Valentin, J-P., and Pollard, C. E. (2006), Optimisation and validation of a medium-throughput electrophysiology-based hERG assay using IonWorks™ HT, *J. Pharmacol. Toxicol. Methods*, 54, 189–199.
112. Report of CIOMS Working Group VI. Management of Safety Information from Clinical Trials, 2005.
113. Anon. (2005), CHMP Guideline on Risk Mangement Systems for Medicinal Products for Human Use, EMEA/CHMP/96268/2005.
114. Anon. (2004), Innovation or stagnation, challenge and opportunity on the critical path to new medicinal products, March 2004.

# *History of Clinical Trial Development and the Pharmaceutical Industry*

**Jeffrey Peppercorn,[1] Thomas G. Roberts, Jr.,[2] and Tim G. Hammond[3]**

[1]*Division of Medical Oncology, Duke University, Durham, North Carolina*
[2]*Noonday Asset Management, L.P., Charlotte, North Carolina*
[3]*Department of Safety Pharmacology, AstraZeneca, Macclesfield, Cheshire, UK*

**Contents**

1 Introduction
2 From Heroic to Evidence-Based Medicine
3 Role of FDA and Rise of Pharmaceutical Industry
4 Protection of Human Research Subjects and Birth of Bioethics
5 Academia/Industry Collaboration
6 Base Study: Early-Stage Oncology Trials
7 Conclusion
References

## 1 INTRODUCTION

For over 2000 years medicine was practiced based on lessons passed down from teacher to student drawn from ancient beliefs regarding illness in the human body and the actions that must be taken to restore health. The principles of Hippocrates and the writings of Galen served as a guide for therapy. There is no doubt that observation of what worked and what did not work, and refinement of practice

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

based on experience, has also shaped medical practice from its earliest days, but for centuries, theory held sway over empiricism.

The situation changed with the systematic application of data collection and observation of outcomes and evolved into the modern practice of evidence-based medicine, driven primarily by the results of clinical trials. The history of clinical trials begins with the recognition of disease as a discrete condition that is similar enough from patient to patient to allow the rational application of the experimental method to determine which therapies work and which do not. It continues with recognition of the need to control for biases that can emerge in clinical experimentation resulting in the establishment of the gold standard of the double-blind randomized controlled trial. In addition, this history involves the recognition of the fact that using human subjects to gain scientific knowledge and help future patients requires regulation and procedures to ensure that the rights of these subjects as individuals are protected. Finally, this history involves the development of government agencies to support and oversee the development of clinical research and the emergence of the pharmaceutical industry as the dominant force in the development of new therapeutic agents. Each of these elements, and other aspects of this history, could be explored in a separate book, and each discipline from oncology to cardiology to psychiatry to surgery has unique stories and features that are beyond the scope of this chapter. We present here an overview of some of the key developments and ongoing issues that should serve as an introduction to the rich history of clinical trials and the rise of the pharmaceutical industry.

## 2 FROM HEROIC TO EVIDENCE-BASED MEDICINE

Medical practice from the time of Hippocrates (460–361 BC) was based on theory, and chief among these was the humoral theory that held that the human body was composed of four humors: yellow bile, black bile, phlegm, and blood. Disease was understood to arise from an imbalance of humors; the role of the physician was to bring the individual back into balance, by adding or removing substances from the body. In essence, this was the ultimate form of personalized medicine, where detection of a fever did not lead to a search for an underlying disorder but indicated an imbalance unique to the individual that needed to be corrected. Balance could be restored by vomiting, diarrhea, or sweating or taking in substances such as mercury or arsenic. One of the most dramatic means of restoring balance, and of demonstrating the knowledge and power of the clinician, was through bloodletting, or phlebotomy.

This practice, depicted on ancient Greek vases dating to the time of Hippocrates, and described extensively in the writings of Galen (AD 129–200), physician to Roman Emperor Marcus Aurelius, was a dominant form of medical therapy for over 2000 years until the mid-nineteenth century. Galen wrote several books on bloodletting and established the use of bloodletting as a form of heroic medicine, stating "the first and most important indications for phlebotomy are … the severity of the disease and the strength of the patient" [1]. Patients with severe disease might be bled to the point of syncope in an effort to restore health [2]. Though this may seem counterproductive based on our current understanding of illness, there are conditions such as heart failure or kidney failure where, in the absence of diuresis or

dialysis, phlebotomy may have helped some patients. Further, the impetus to "heroic medicine" where severe disease in an otherwise healthy patient is treated with a dramatic, but unproven therapy, still exists in medicine and society as evidenced by the use of bone marrow transplant for breast cancer as recently as the 1990s [3].

The centrality of phlebotomy to practice in Galen's time is indicated by his instructions to subject patients to bloodletting, not only when they were sick but also for disease prevention. He wrote in *On Treatment by Venesection*, "the time for phlebotomy is not only when severe disease is established, but also when it is likely to occur" [1]. Galen's teachings were translated into Arabic and preserved in the Arabic world while Europe was in the Dark Ages. This practice was later adopted by European monks and continued throughout the Renaissance and into the nineteenth century.

One of the chief American proponents of bloodletting, and the practice of heroic medicine, was Benjamin Rush (1745–1813). In the face of the yellow fever epidemic of 1793, when thousands were dying, Rush advocated not just bloodletting but copious bloodletting, regardless of the state of the patient [4]. Rush was a notable physician in Philadelphia (in addition being signatory to the Declaration of Independence and treasurer of the U.S. Mint), and when the epidemic struck he tried a variety of therapies including purging with calomel (mercury), blistering, wrapping patients in vinegar-soaked blankets, and cinchona bark before deciding that more drastic measures were needed [4]. He reportedly saw over 100 patients a day and was both practitioner and promoter of bloodletting, continuing to advocate the practice for widespread use through yellow fever epidemics in 1794 and 1797 [5]. Rush called on his students to "Venerate the Lancet. It is the Magna gratia Coeli, The great gift of Heaven" [4]. Bloodletting was perhaps most famously used in American history in the treatment of President George Washington. Despite four heroic bloodlettings totaling up to 2 liters of blood in less than 24 hours (about 40% of total volume for an average male), performed by colleagues of Benjamin Rush, Washington succumbed to what was likely bacterial epiglotitis and died on December 14, 1799 [6].

Thus, principles and theories established over 2000 years earlier continued to provide the basis of medical practice; while knowledge of biology and human physiology advanced, the rational use of therapy did not. This unfavorable situation would begin to change with the conduct of clinical experiments and, importantly, the distribution of information regarding new therapeutic interventions and methods of validation. Progress toward the age of clinical trials required both the willingness to try new things (i.e., experiment) and the appreciation of the need for systematic observation, recording, and analysis of outcomes.

One of the first documented clinical experiments was the trial of inoculation as a treatment for smallpox performed by Cotton Mather during the epidemic of 1721 (Table 1). Inoculation, the practice of introducing a small amount of infectious material into a healthy subject in order to convey a mild case of disease that would prevent development of a more severe life-threatening form, was practiced in China as early as 1000 bc. However, it was not accepted practice in 1721 when Mather convinced Dr. Zabdiel Boylston to inoculate Bostonians against a smallpox epidemic, and they decided to collect data. Their report of 2% mortality among inoculated patients versus 14.9% among naturally infected patients, represents one of the first know instances of clinical data collection to guide future clinical practice [7].

**TABLE 1 Timeline of Select Events in History of Clinical Research**

| | |
|---|---|
| 1721 | Mather smallpox inoculation in one of earliest documented clinical experiments |
| 1753 | Lind study of citrus for scurvy compares multiple concurrent treatments |
| 1798 | Jenner conducts smallpox vaccination trial |
| 1834 | Trousseau conducts first blinded placebo-controlled trials |
| 1836 | Louis pioneers systematic evaluation of therapy and numerical method. |
| 1846 | Morton publicly demonstrates value of ether for anesthesia |
| 1847 | Semmelweis studies role of hygiene in preventing maternal death with historical controls |
| 1870 | Lister conducts single-arm study of antisepsis with historical control |
| 1896 | Fibiger conducts first randomized control trial, of diphtheria serum for treatment of diptheria |
| 1938 | U.S. Food Drug and Cosmetics Act requires regulation of claims of new medications |
| 1946 | Brittish MRC conducts first modern randomized control trial, evaluates streptomycin for TB |
| 1949 | Nuremburg Code states importance of voluntary informed consent for research |
| 1954 | Salk polio vaccine trial, landmark large randomized placebo control trial |
| 1964 | Declaration of Helsinki stresses concern for the interests of research subjects |
| 1978 | Belmont report produces a practical guide to ethical research conduct |

In a testament to the slow dissemination of medical knowledge in the eighteenth century, the same study was essentially conducted by Edward Jenner in England in 1796. Jenner found that exposure to small amounts of cowpox could prevent both fulminant cowpox and the more lethal smallpox; he coined the term vaccination, based on *vaccinia*, Latin for cowpox [8].

It is possible to see the origin of clinical trials in these early clinical experiments. Though these efforts lacked the structure and methodology of anything resembling a modern clinical trial, the notion of studying or demonstrating the effectiveness of an intervention, rather than applying it based solely on theory, as was the case with phlebotomy, marked a profound shift in the approach to patient care. Other pioneering examples include the demonstration of the value of citrus (vitamin C) for scurvy by James Lind in 1747 [9], demonstration of ether (anesthesia) by William Morton in 1846 [10], and demonstration of the value of hygiene for prevention of childbed fever by Ignaz Phillip Semmelweis in 1847 [9].

Lind conducted a study among 12 Brittish sailors assigned to different dietary therapies for treatment of scurvy, which is estimated to have killed over 1 million people in the seventeenth and eighteenth centuries, and showed rapid recovery of sailors eating citrus. He published "Treatise on the Scurvy" in 1753 [9]. Morton demonstrated the value of ether to prevent operative pain to a room full of skeptical surgeons in the Massachusetts General Hospital Amphitheatre on October 16, 1846 [10]. His work was publicized by Henry J. Bigelow in the November 1846 *Boston Medical and Surgical Journal* [5]. Semmelweis conducted a natural experiment of sorts, demanding handwashing by physicians and medical students between work with cadavers and work on the delivery wards and demonstrated a drop in the rate of childbed fever and death from sepsis from 10% of mothers to less than 2% after imposition of this change in hygiene [5]. It is notable that widespread adoption of the practice of anesthesia for surgery and handwashing in obstetrics lagged many years behind these clinical demonstrations and publication highlighting the importance of an established method for demonstrating efficacy in medicine and the importance of community reliance on evidence to guide practice [9].

Though case series and public demonstrations are a form of clinical research, the true origin of systematic clinical research dates to the era of the Paris Clinical School

and the work of Pierre Charles Alexandre Louis. As noted above, bloodletting, whether by leech or lancet, was used widely though the nineteenth century based on the inherited wisdom for the time of Galen. The first evidence-based challenge to this practice emerged when Pierre Louis, a young physician at La Charite Hospital in Paris in the 1820s, decided to evaluate the impact of bloodletting compared to other treatments on outcomes in pneumonia. The very concept of such a study represents a shift in the understanding of disease. As historian of science Charles Rosenberg has noted in *The Therapeutic Revolution*, the movement toward recognition of diseases as discrete entities as opposed to personal states of imbalance was a necessary conceptual step for the study of the impact of therapeutic interventions on populations of patients, and the collection of data to guide therapy for future patients [11]. This subtle difference between therapies as means to restore balance versus therapies as targets for specific processes was critical for the development of evidence-based medicine. The development of the randomized clinical trial required both innovation in methodology and a transformation in this concept of disease. However, this step was accompanied by other changes in medicine, perhaps equally important for the development of clinical trials, such as movement of the place of illness from home to hospitals, allowing evaluation of large numbers of like cases, improvements in dissemination of knowledge, and a skepticism toward unproven therapies [11].

Pierre Louis evaluated 77 cases of pneumonia, stratified by whether they were bled on days 1–4 of illness or days 5–9. He found that 44% (18 of 41) of patients undergoing early bleeding died, compared to 25% (9 of 36) undergoing delayed bleeding [12]. That there was no comparison to a "no bleeding" control group speaks to the importance of bleeding as a therapy during this period. Describing what he called the "numerical method," Louis wrote about the need to compare treatments among comparable numbers of patients to determine which treatment should be used in the future [13]. This was a radical suggestion at the time and contradicted the established practice of medicine based on the physician's clinical judgment on a case-by-case basis. The move toward standardized evaluation, by no means robust in Louis' time, now seems obvious, yet echoes of the debates over evidence-based medicine versus clinical judgment remain today.

Louis' work was published in 1836, entitled *Researchers on the Effects of Blood Letting in Some Inflammatory Diseases* [14]. His findings sufficiently challenged the dogma that he conceded: "The results of my researches … are so little in accordance with general opinion, that it is not without a degree of hesitation I have decided to publish them." He published patient-specific data in detailed tables so that his work could be reviewed and analyzed by others. Even in the face of his own empirical evidence, he questioned whether a more aggressive "dosing" might have achieved better results. "Should we obtain more important results if, as is practiced in England, the first bleeding were carried to Syncope? The practice deserves a trial, but great success cannot, I think be anticipated; since many cases, the history of which I have drawn up, and which were fatal, were bled to a sufficient extent" [14]. In the face of this evidence, among the first well-documented clinical evaluations in which we can see the origins of modern clinical research [12], he reluctantly concluded that "bloodletting has had very little influence on the progress of pneumonitis, or erysipelas of the face, and angina tonsilarris, in the cases under my observation" [14].

Though Pierre Louis relied on large numbers of cases, rather than randomization, he had hit upon both the importance of empirical observation to determine the safety and efficacy of therapy and the need to control for confounding factors among patients undergoing different interventions. Further, he argued not only that such evaluation was valid, but that "therapeutics cannot advance without it" [14]. Louis relied on advances in mathematics and probability theory to develop a basis for clinical science and called for evaluation of a magnitude of harms and benefits of any intervention to determine its worth [15].

Of note, bloodletting declined based on the work of Louis and others but persisted for many decades. A Philadelphia physician reported in 1862 that among over 9500 cases, only one was treated with general bloodletting, 12 with cupping, and 3 with leeching [11].

The work of Pierre Louis inspired the great American clinician Oliver Wendell Holmes who argued in 1843 that clinical evidence must guide practice [16]. Physicians from Great Britain and the United States took up the scientific method of Pierre Louis and began to collect outcome data on common practice, essentially conducting retrospective observational evaluation of common therapeutic practices [13]. In the absence of such evidence for most common medical practices of that day, this dawning awareness of the need for clinical research initially led to therapeutic nihilism [11]. However, it also produced the necessary environment for development of advances in care.

When Joseph Lister pioneered antiseptic surgical techniques in 1865 and studied their impact in subsequent years, he evaluated them based on clinical evidence by comparing outcomes from cases prior to the technique with those after introduction of the technique, essentially conducting a single-arm study and using a historical control [13, 17]. Similar advances in the management of infectious disease were made through testing of new therapies and comparison with historical controls in studies of diphtheria.

The problem with historical controls is that improvements in outcome may be due to the intervention, due to other changes in practice, or due to changes in the patient population that may occur over time. Therefore a pioneering advance in the history of clinical trials was the use of concurrent controls, first conducted in the trial of diptheria serum by Danish physician Johannes Fibiger in 1896 [18].

Fibiger (1867–1928) alternately assigned new patients with diphtheria to standard therapy versus standard therapy plus diphtheria serum depending on the day of treatment between May 1896 and May 1897 [19]. The primary outcome was mortality, and secondary outcomes included croup and fever. He also evaluated toxicity, primarily development of serum sickness. Among 484 patients, 8/239 (3%) treated with serum died versus 30/245 (12%) controls [19]. Though no formal statistics were used, this trial demonstrated the importance of large numbers, randomization (future studies would improve over alternating days as a means of randomization), and concurrent controls. As Fibiger wrote: "In many cases a trustworthy verdict can only be reached when a large number of randomly selected patients are treated with the new remedy and at the same time, and equally large number of randomly selected patients are treated as usual" [18].

If Fibiger took the first steps toward pioneering the randomized control trial, the final steps in establishing this methodology were taken by the British Medical Research Council (MRC) in 1946 in what is typically termed the first true random-

ized control trial. Interestingly, at this juncture the history of clinical trials merges with the history of the pharmaceutical industry. The MRC conducted a trial of the promising new drug streptomycin for treatment of pulmonary tuberculosis, and one of the ethical rationales for randomization was that the drug, produced by the American pharmaceutical company Merck, was in short supply in Britain [20]. Widespread availability of the drug in America precluded a randomized trial, but in Britain, one of the only ways to obtain the drug was through participation in the MRC trial. The trial randomized 97 patients using random numbers in sealed envelopes to streptomycin versus control in a double-blinded fashion [20]. The trial demonstrated both the value of randomization and the value of streptomycin for tuberculosis. There was "considerable improvement" of chest X rays among 55% in the streptomycin group versus 8% of the controls, and 7% deaths in the streptomycin group versus 27% among controls [20].

Since the 1940s when the methodology of the randomized clinical trial was definitively established and embraced by the medical community, the pace of advancement in medical therapy has been remarkable. A clinical trial system for systematic assessment of novel interventions has been established, and government agencies, academic institutions, and private industry have been organized around the goal of discovering new treatments and testing them in a scientifically rigorous fashion. Other important steps along this pathway included the development of the trial system with phase I trials to test first in human interventions or combinations focused on safety, phase II studies to expand evaluation of safety in a select patient population and to evaluate efficacy, and phase III trials to provide a randomized controlled comparison between an experimental intervention and a standard therapy. There are of course variants of this trial structure, such as phase I/II trials, randomized phase II trials, pilot studies, phase IV trials (typically aftermarket trials to provide real-world safety and efficacy data), and emerging phase 0 trials (very small trials using low doses to study method of action) [21], and a full discussion of trial design is beyond the scope of this chapter. The conceptual steps described above paved the way for our modern system of clinical research.

Two other aspects of modern clinical trials that deserve note in this history, however, are the development of blinding and the placebo-controlled trial. Both of these are related to controlling for bias. When a research subject knows he or she is receiving an intervention that is supposed to work, it may affect subjective and even objective outcomes through psychological factors. Blinding means that the subject does not know what intervention he or she is receiving. Double blinding, used in some trials, means the physicians providing the intervention and assessing the outcome also do not know which intervention the subject is receiving. One common form of blinding involves use of a placebo. A placebo, based on the Latin word for "I shall please," is an inert substance that is not expected to have any direct therapeutic value but that can be used in a trial to make a subject believe he or she is receiving a therapy and help sort out psychological effects of being treated from true physiologic effects. In modern clinical trials, this is used in randomized studies when there is no appropriate standard of care, or when you want to compare adding a novel intervention to a standard intervention to the standard intervention alone.

The history of using placebo controls, or "blinded assessment," in clinical research dates to the late eighteenth century when blinded assessment was used to determine

if "mesmerism," a then popular form of therapy in which the practitioner directed a sort of psychic force (characterized as a form of animal magnetism) at a patient to treat illness [22]. Researching the history of this subject, Kaptchuk relates that the first experiment actually used blindfolds to prevent the subject from knowing if mesmerism was really being applied, and it was conducted in 1784 in the house of Ben Franklin. Later, blind assessment was combined with the use of placebo in trials intended to study the validity of homeopathy, a practice of treating disease with minute amounts of a substance linked to the disease (on the principle of "like cures like") [22]. These trials conducted in 1834, by the French physician Armand Trousseau (later famous for describing a syndrome of migratory blood clots associated with abdominal cancer), were the first blinded placebo-controlled trials, leading Trousseau to conclude that homeopathic remedies were no more active than placebo [22]. Use of placebo controls did not become a standard part of clinical research until later in the nineteenth century when they were used in a series of German studies examining the health effects of a variety of natural and nutritional remedies [22]. The use of randomization combined with blinding and placebo controls appears to have first emerged in the Michigan tuberculosis trials in 1926 in which a gold-based intravenous therapy was compared to intravenous water [22]. In 1954, this methodology was famously used to establish the safety and efficacy of the Salk polio vaccine, in a placebo-controlled trial involving almost 2 million subjects [23].

The degree to which medical therapy has evolved from reliance on theory, and distrust of empiricism, as recently as the time of Benjamin Rush, to reliance on evidence-based medicine and the randomized clinical trial, is exemplified by debate over whether treatment in a clinical trial itself now represents the standard of medical care in some settings [24, 25]. This view has been widely advocated in oncology, given both the poor outcomes with standard therapy for many diseases, and the clear advances demonstrated in clinical trials, particularly in treating childhood malignancy. The concept of a "trial effect" has been postulated, whereby treatment within a clinical trial conveys therapeutic benefit over and above the benefit of the experimental intervention itself [26]. Though it is not clear that such a trial effect holds up when comparisons of patients treated within clinical trials and patients treated outside of trials are subjected to the type of rigorous standards used to assess any intervention [27], it is notable that the clinical trial has emerged not only as a means to demonstrate the value of novel therapies, but as *the* recommended option for the care of some patients [24].

## 3 ROLE OF FDA AND RISE OF PHARMACEUTICAL INDUSTRY

No discussion of the history of clinical trials and drug development would be complete without mentioning the role of the Food and Drug Adminstration (FDA) and its analogous regulatory authorities throughout the world. Prior to the twentieth century, the U.S. government did little to regulate the marketing of therapeutic products; fraudulent claims were commonplace and went unpunished. Over the last 100 years, the FDA's role has evolved to encompass three fundamental assurances: safety, efficacy, and adequate and accurate labeling [28].

Most observers consider 1906 to be the birth of the modern FDA [29]. In that year, Congress passed the original *Pure Food and Drugs Act*, which was signed into

law by Theodore Roosevelt. The statute prohibited misbranded and adulterated foods, drinks, and drugs in interstate commerce. Specifically, the law required companies to disclose weights and measures of their products and to provide labels disclosing whether their products contained alcohol, morphine, opium, cocaine, heroin, eucaine, chloroform, cannabis, chloral hydrate, or acetanilide. Several forces provided the impetus for the law, including an exposé of the harmful preservatives used in the meat-packing industry as well as the realization of an increasing incidence of drug addiction from the use of "patent medicines." Congress originally designated the Bureau of Chemistry in the Department of Agriculture, founded in 1862, to enforce the law; it would not be until 24 years later, in 1930, that the agencies name would be changed to the FDA.

It is important to note that the original Pure Food and Drugs Act focused on ensuring accurate disclosure about product content, but the act did not prohibit firms from making false therapeutic claims as long as the firms provided accurate content labeling. Congress enacted the Sherley Amendment, in 1912, to address this loophole. The amendment specifically prohibited firms from the labeling of medicines with false therapeutic claims if the labels were intended to defraud the purchaser. That this legal standard was difficult to prove weakened the practical impact of the amendment. Radiothor, a radium-containing tonic that could be fatal with chronic ingestion, and Lash-Lure, an eyelash dash that blinded some women, are just two examples of worthless or harmful products widely marketed at the time. Efforts to enhance the laws stalled in Congress for years until the occurrence, in 1937, of the unfortunate Elixir Sulfanilamide disaster.

Sulfanilamide had been shown, by 1937, to have dramatic, curative activity against streptococcal infections. In an attempt to capitalize on a recognized demand for the agent in liquid form, especially among children, Tennessee-based S.E. Massengill Co. sought to manufacture an elixir formulation. Chemists at the company found that sulfanilamide would dissolve in diethylene glycol and that this solvent had an attractive mixture of fragrance and flavor. The company was not required to, nor did it chose to, carry out toxicity testing on the new formulation. Tragically, the highly toxic analog of antifreeze killed more than 100 people, many of whom were children. In the wake of the tragedy, Congress and the president moved quickly to enact the *Food, Drug, and Cosmetic Act*. The new law required, for the first time, that drugs be safe for their intended use. Specifically, the law required premarket approval for all new drugs; firms had to demonstrate proof that their drugs were safe prior to their marketing. Furthermore, the law gave the FDA authority to identify drugs that would require a prescription from a physician.

It was not until 1962, almost 25 years after the original Food, Drug, and Cosmetic Act, that Congress amended the 1938 act by codifying a requirement that drugs demonstrate adequate evidence of efficacy in addition to safety. Congress indicated that the FDA should require "full reports of investigations which have been made to show whether or not such drug is safe for use and whether such drug is effective in use" [30]. The new law was in many ways the result of hearings held by Senator Estes Kefauver. The senator had worked for years to try to reform the agency, but his efforts languished until the thalidomide tragedy, which played out mainly in Europe, provided the catalyst for Congressional action. Congress also gave the FDA control over the regulation of drug development trials, including the requirement

for informed consent, the regulation of drug advertising, and the power to establish and enforce good manufacturing practices. As part of the law Congress required the FDA to assess the efficacy of all drugs introduced since 1938.

Over the next 40 years, Congress enacted several additional laws governing the FDA, including the Orphan Drug Act and the Federal Advisory Committee Act. Mutiple sets of drug regulations were also developed during the latter half of the twentieth century to refine and interpret the laws. The 1990s, however, proved to be the next highly consequential period for the FDA's evolution. The AIDS crisis in the late 1980s and early 1990s provided the major impetus for reform. Activists targeted the FDA for failing to approve drugs quickly enough in the face of the crisis. Congress responded over the period of 1992–1997 by enacting two laws, the Prescription Drug User Fee Act of 1992 (PDUFA) and the FDA Modernization Act of 1997. Collectively, these laws have had a profound impact on the FDA's regulation of drugs intended to treat serious or life-threatening illnesses. Programs such as Fast Track Designation, Priority Review, and Accelerated Approval have each worked to expedite the development of drugs intended to treat serious illnesses such as cancer.

Of these three programs, the accelerated approval mechanism has had the largest impact [28]. The provisions under accelerate approval permit the FDA to approve agents to treat serious or life-threatening illnesses before the clinical benefit necessary to meet the standard for regular approval has been demonstrated. Specifically, the FDA can grant initial drug approval on the basis of a surrogate measure of clinical benefit (e.g., CD4 count or tumor shrinkage) if the treatment is intended to treat a serious or life-threatening illness and is reasonably likely to be superior to available therapies [31]. The FDA in turn receives agreement from the drug's sponsor to complete confirmatory trials in the postapproval period. If the postapproval trials fail to show clinical benefit, the FDA has a mechanism to remove the drug from the market in a timely manner. If history is any guide, the regulatory evolution of the FDA will continue into its next 100 years. Recent efforts have focused on enhancing the agency's ability to monitor the safety of drugs and devices in the postapproval period [32]. The agency has also become more active in collaborating with sponsors, academe, and clinical societies in order to improve the return on the nation's public and private investment in research and development. The FDA's Critical Path project is one such promising area of regulatory research. The program includes efforts to provide a structure for the inclusion of response or toxicity biomarkers into the drug development process, an area that is certain to receive additional attention in the coming years [33].

Amidst this regulatory framework, and in concert with the evolution toward evidence-based medicine, the commercial pharmaceutical industry has risen to become the dominant source in the development of new therapeutics. From the development of aspirin in 1897 to today's molecularly targeted cancer therapies, most diseases now have at least some symptomatic or curative medicinal treatments. In addition to development of the basic science that paved the way for rational development of new drugs, and the clinical science that created the foundation to determine their safety and efficacy, two pieces of legislation played a major role in paving the way for the modern pharmaceutical industry. Both pieces of legislation arose from a perceived need to provide a greater economic incentive for the development of new drugs. The Orphan Drug Act, passed in 1983, provided 7 years of

exclusive marketing and large tax credits to cover research and development costs for any drug designed to treat a rare disease. Similarly, the Drug Price Competition and Patent Term Restoration Act, passed in 1984, provided industry with extended patent protection to cover the significant time required for research and development and allowed generic versions of drugs to be produced and marketed without repeated expenditure for clinical trials of the generic formulation. Though other factors likely contributed, the passage of these laws in the early 1980s corresponded with an exponential rise in expenditure on research and development by the pharmaceutical industry.

In 2006, research and development (R&D) spending for the biopharmaceutical industry alone was as high as $55.2 billion [34]. Member companies of the Pharmaceutical Research and Manufacturers of America (PhRMA) spent approximately $2 billion on R&D in 1980; over the last quarter century, they have grown their R&D expenditures at an astonishing compound annual growth rate of more than 12%. This growth in private R&D spending has dramatically outstripped the growth in spending by the National Institutes of Health (NIH). In 2006, the NIH budget amounted to $28.6 billion, a little more than half of the amount spent by biopharmaceutical companies [35]. This growth reflects underlying secular changes in the process of drug development. Once the purview of a few fine chemical companies located in the Upper Rhine Valley in Switzerland and in the greater Philadelphia area, the effort to discover and develop new drugs has transformed into a commercially focused, capital-intensive enterprise fraught with risk and uncertainty and producing many more failures than successes. In fact, fewer than 1 in 10 drugs entering clinical trials will ever gain marketing approval from the FDA [36], and at least 1000 molecules are screened for each that enters clinical testing [34]. For some diseases such as cancer, the ratio of winners to losers is even more unfavorable [37]. Of the thousand of agents in some stage of clinical or preclinical development, as few as 20 new molecular entities will reach the FDA regulatory standard for approval. The high failure rate and demanding regulatory standard together drive the cost of new drug development, which by some estimates exceeds $800 million for each successful approval [38].

The effort to manage the investment in drug development has led to the formation of another thriving entity—the contract research organization (CRO). With more than 1000 CROs in operation and with industrywide revenue exceeding $17 billion, for-profit CROs have, in large part, taken over academia's historic role in organizing new drug development. Clinical researchers from the academe remain crucial to some aspects of development, but CROs now often take the lead in identifying study centers, recruiting patients, acquiring and monitoring data, and even performing relevant statistical analsyses. As of December 2007, there were more than 11,000 clinical trials actively recruiting patients for a treatment clinical trial according to clinicaltrials.gov. The work required to plan and execute all of these trials has overwhelmed the resources in both the academe and in the pharmaceutical industry, especially as the pharmaceutical industry completes substantial downsizing efforts. A series of high-profile incidents involving research subjects sustaining harm along with potential conflicts of interest have drawn criticism over the industry's increasing reliance on CROs. It is possible that at some time in the future, CROs will be subject to additional regulatory scrutiny and action; for now, their role in clinical drug development appears secure.

## 4 PROTECTION OF HUMAN RESEARCH SUBJECTS AND BIRTH OF BIOETHICS

The history of clinical trials and the pharmaceutical industry is closely tied to the history of bioethics, the concerns of which now shape the regulation of clinical research and pharmaceutical development. Though medical ethics traces its routes back at least to the teachings of Hippocrates and the dictum "primum non nocere" or "first do no harm," the rise of modern bioethics began in the twentieth century with the reaction to the atrocities of Nazi science. It is important to understand that the origin of modern bioethics as it concerns research ethics (the ethics of clinical research) emerged from a reaction to scandal, rather than an a priori guide to how to conduct clinical science. There is a strong and continued emphasis on protection of human subjects from research that at times appears to come into conflict with the need to recruit patients to clinical trials to advance care for future patients. By recognizing the origins of the concerns of human subjects' protections, and the continued need for emphasis on respect of research subjects as persons, it is possible to conduct clinical research efficiently, while securing the ethical foundation of such science.

Nazi science was characterized by use of human beings as research subjects without their consent, to explore the limits of human endurance, the effects of exposure to extremes and a variety of toxins, and characterized by complete disregard for the well-being of the subject [39]. The crimes of Nazi science were publicly exposed at the Nuremburg Medical Trial of 1946–1947 in which 23 German doctors were tried for experiments performed on victims in concentration camps from 1933 to 1945 [40]. The Nuremburg Code, published in 1949, was developed to not only ensure that such atrocities were never repeated but to advance the ethical conduct of all studies involving human subjects. The first principle established by Nuremburg was the requirement for voluntary informed consent on the part of the research subject [41]. This principle remains at the heart of human subjects protection and is a core component of all further codes of ethical research, though how to define and obtain meaningful informed consent remains a subject of debate. Additional principles related to the need for clinical research to be conducted in the interest of humanity, based on good preclinical science, designed to minimize harm to subjects, with consideration of the welfare of subjects and their right to decline participation ensured throughout the research [41]. The Nuremburg Code was a starting point, but further guidance was needed to establish broadly based guidelines for research ethics.

In the wake of involvement by physicians in Nazi experiments, the World Medical Association (WMA) established the Committee on Medical Ethics in 1952. Based on a series of meetings and debates, the WMA published the Declaration of Helsinki in 1964 [42]. The declaration emphasized the rights of research subjects to be informed of the potential risks involved in research and of the voluntary nature of research participation. This document recognized a distinction between research with therapeutic intent that might benefit the subject and research with scientific goals only, but emphasized that under all circumstances "the interest of science or society could never take precedence over considerations related to the well-being of the subject" [42]. Ethicists have debated the practicality of imposing such a principle on research that by definition involves unknown risks to subjects, but the

principle of respect for the rights and well-being of the subject that the Declaration establishes is clear.

No history of clinical trials would be complete without consideration of the significance of the Tuskeegee syphilis trial. This study was initiated in the 1930s by the U.S. Public Health Service to evaluate the impact of untreated syphilis in a cohort of African American men. Though there were few known effective therapies for syphilis at the time of initiation of the study, the deadly natural history of the disease had already been demonstrated in Scandinavian subjects [43]. In the context of that time, this question of the course of untreated syphilis in black patients clearly arose out of racism. Further, the subjects were tricked into participation in an observational study under the false premise that they would be receiving therapy. For example, subjects undergoing a lumbar puncture for research purposes were solicited to receive a "special treatment" from the nurse [43]. Though this study was initiated prior to the Nuremburg Code and Declaration of Helsinki, it was continued by the U.S. Public Health Service and published in major journals until 1972, when the scandal was exposed by the press [43].

Failure to consider the well-being of research subjects was also not exclusive to racist research. In 1966, H. Beecher, a professor of anesthesiology at Harvard Medical School wrote a landmark exposé of 22 clinical studies supported by top academic, government, and industrial institutions, and reported in leading medical journals, that violated the rights of the participants, frequently through failure to obtain informed consent [44]. In response to Beecher's publication, the U.S. National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research was established, and the report of the commission, the Belmont Report was released in 1978 [45]. The first principle, respect for persons, called for respect of subjects' autonomy and the need for special protections for those with decreased autonomy (e.g., in children, subjects with neuropsychiatric disorders, or imprisonment). The second principle, beneficence, called for proactively taking steps to promote the well-being of research subjects, in addition to seeking to avoid harm and respecting their wishes. Finally, the Belmont Report recognized the principle of justice, which called for research to be conducted among populations who could benefit from the research. The report emphasizes that all clinical research must be clearly distinguished for routine clinical practice, that voluntary informed consent is required, and that research must be carefully designed to maximize potential benefits and minimize potential harms to the research participants [45].

Many guidelines and national and international codes of research ethics have been developed and promoted since the publication of the Nuremburg Code. The core principles in 13 major codes and declarations were evaluated and analyzed recently by Emmanuel and colleagues; they elucidated seven ethical principles that should guide all clinical research [46]. These principles of value, scientific validity, fair subject selection, favorable risk–benefit ratio, independent review, informed consent, and respect for potential and enrolled subjects are proposed as universal requirements of clinical research, grounded in the underlying principle of respect for persons [46].

The need for regulation of clinical research should not obscure the fact that, for many interventions, testing of safety and efficacy within a clinical trial may be the most ethical way to provide a novel intervention to a patient for whom outcomes with standard therapy are inadequate. A prime example was high-dose chemother-

apy and bone marrow transplantation for breast cancer. Early-phase single-arm studies among patients with advanced disease demonstrated results that appeared substantially better than those seen among historical controls treated without transplant [47]. In fact, advocates for breast cancer patients and some clinicians questioned the ethics of conducting randomized trials for a "proven therapy" [48]. In a tawdry portion of the history of clinical research, the story was complicated by the falsification of data by a prominent researcher, Bezwoda, who presented fabricated randomized clinical trial data at major scientific meetings and in publication [49], further promoting the use of this intervention prior to confirmation in ongoing randomized trials, and in fact, likely slowing accrual to those trials [3]. In the end, well-designed randomized control trials demonstrated no improvement of outcomes for women with advanced breast cancer from high-dose chemotherapy and transplant compared to less aggressive and toxic forms of therapy [50].

## 5 ACADEMIA–INDUSTRY COLLABORATION

In addition to general concerns over human subjects protection in clinical research, there has been recent increased attention to the role of the pharmaceutical industry in clinical research. In brief, this concern stems from the fact that the pharmaceutical industry now plays the dominant role as the sponsor of clinical research and the fact that most elements of the pharmaceutical industry are for-profit. As noted above, in 1992, pharmaceutical industry investment in research exceeded the operating budget of the NIH for the first time. Increased funding appears to have translated into increased sponsorship of published clinical research. An increase in documented pharmaceutical sponsorship of clinical research over time has been noted in stroke trials [51], oncology trials [52, 53] and in randomly selected trials from five medical journals [54]. The majority of studies in many areas of medical research are now supported in some way by the pharmaceutical industry.

One interesting aspect of the association between pharmaceutical sponsorship and clinical research has been the observation that pharmaceutical industry involvement correlates with publication of positive clinical trials. This association was first noted by Davidson et al. in 1986 [55]. Davidson's initial observation was supported by a systematic review reported in 2003 by Bekelman et al. [56]. Pooled analysis of 37 studies, which collectively included 1140 clinical trials, demonstrated that industry sponsorship correlated with positive study outcomes or "pro-industry" conclusions with an odds ratio of 3.60 (95% confidence interval, 2.63–4.91). The reason for this association is unclear, but potential explanations include biased trial design, or interpretation of results, superior or safer selection of agents to take forward into later phase trials, or failure to publish negative studies [52]. These possibilities support the movement toward clinical trial registries and further evaluation of this association.

Concerns over potential bias in research sponsored by the pharmaceutical industry have heightened the tension over closer academia–industry collaborations. It is not surprising that as the pharmaceutical industry became the dominant funding source for clinical trials and the academic medical centers remain the dominant site for clinical research that academia and industry would by necessity become partners. Several studies have addressed conflicts of interest in academia–industry relationships. In 2000, Boyd and Bero reported that 7.6% of faculty investigators reported

financial ties with industry with the percent of involvement growing over time, likely representing an underestimate based on variable and voluntary reporting requirements [57]. In an effort to manage potential bias and conflicts of interest in academia–industry collaboration, the International Committee of Medical Journal Editors issues guidelines for clinical research, but as of 2002, it was not clear that these guidelines were widely implemented, leaving this an ongoing area of interest and evaluation [58].

Ultimately, well-regulated collaboration between academic centers and the pharmaceutical industry should continue to yield advances in therapy to the betterment of society. In fact, the model of academia–industry collaboration in oncology stands out as an example of what can be achieved when the resources and efficiencies of industry partner with the intellectual resources of the academic world, and increasingly with clinical care in both academic and community-based research centers.

## 6 CASE STUDY: EARLY-STAGE ONCOLOGY TRIALS

Over the past two decades, the growth of clinical research in oncology has exceeded the growth in all other areas of medicine. A great deal of scrutiny of the cancer-related research infrastructure has accompanied this growth, providing an excellent case study on which to highlight many of the issues raised in this chapter. As of early December 2007, there were 6214 clinical trials actively recruiting patients with cancer that were registered with the National Institutes of Health (NIH; clinicaltrials.gov). Most (80%) of these trials were early-phase, or developmental trials (phases I and II), with 1653 phase I trials and 3323 phase II trials. The number of drugs and biologics in clinical trials for the treatment of cancer is now greater than the combined total of the next two most represented therapeutic classes, anti-infectives and immunologics [36]. Cancer drug development has transformed from a low-budget, government-sponsored enterprise to a high-stakes multi-billion-dollar industry with hundreds of biotech and pharmaceutical companies seeking approval and adoption of their products [59].

The structure regulating clinical trials is complex, with multiple agencies and groups sharing responsibility and oversight. At the federal level, the FDA regulates industry-sponsored research, while the Office of Human Research Protection (OHRP) oversees human research sponsored by the Department of Health and Human Services (HHS). In the wake of highly publicized lapses in the oversight of clinical research and instances of questionable ethics of some investigators, some of which led to tragic outcomes, the OHRP was recently elevated from the NIH to the Office of the HHS Secretary. In addition, individual or centralized institutional review boards (IRBs), ethics committees, clinical investigators, and sponsors all add layers of oversight to clinical cancer trials to ensure that they are performed safety and ethically.

Phase I trials represent the first testing of an experimental agent in humans, acting as a point of translation of years of preclinical work into the clinic [37]. As already outlined, the major objectives during phase I are to characterize the agent's toxicity profile and to determine a schedule and dose appropriate for further testing. Phase I cancer trials differ from phase I trials in other areas of medicine in two important ways. First, phase I trials in other areas of medicine typically enroll

healthy participants, whereas phase I trials in oncology almost always enroll patients who have cancer and who have exhausted standard treatments. Second, investigators and patients seek to realize therapeutic benefit in phase I cancer trials, usually as a secondary endpoint. Unlike phase I trials in other areas of medicine, treating physicians almost always enroll patients in phase I cancer trials with therapeutic intent, and patients often expect to benefit [60, 61].

The ethical basis of phase I cancer trials has been questioned, in part because they involve potentially vulnerable cancer patients near the end of life [62]. Some ethicists have raised concern that patients who chose to participate may experience significant risks with little chance to benefit. Others have pointed out that patients who participate have unrealistic expectations about the probability of benefit or the goals of research, despite having gone through the informed consent process [63]. Still others have been concerned that cancer patients may have judgment that is clouded by their illness and therefore cannot make truly voluntary decisions.

Much of the debate has now evolved to focusing on the estimated risk–benefit ratio in phase I trials. Over the 5-year period from 1986 to 1991, three groups published response rates in meta-analyses of phase I clinical trials [64–66]. Rates of objective response (usually as defined by tumor shrinkage by greater than 50%) ranged from 4 to 6%. Toxic death rates were reported to be around 0.5%. A major limitation to the relevance of these studies is that they are outdated and do not include assessment of the risks and benefits associated with phase I trials of the newer targeted agents. Researchers at Harvard recently reported trends in the risks and benefits to patients participating in phase I clinical trials submitted for presentation at the American Society of Clinical Oncology over the period of 1991–2002 [37]. The overall respective response and toxic death rates of 3.8 and 0.54% were similar to those published in prior meta-analyses. What was striking, however, were the trends over time. The overall toxic death rate for 213 published studies from this sample decreased over time, from 1.1% over the first 4 years of the study (1991–1994) to 0.06% over the most recent 4-year period (1999–2002). After adjusting for characteristics of the investigational agents and the experimental trials, the odds of patients dying from an experimental treatment fell by more than 90% from 1991 to 2002. Interestingly, the odds of a patient dying from a targeted or biologic treatment were four times lower than the odds of dying from a traditional cytotoxic agent. Rates of objective response also declined over time but by proportionally much less.

There has been no explicitly articulated standard about what determines a socially acceptable risk–benefit ratio [67]. Agrawal and Emanuel recommended comparing risk–benefit ratios in phase I cancer trials to socially accepted determinations already used for cancer such as FDA approval standards. By this construct, the risk–benefit ratio of many phase I cancer trials may not be clearly worse than those used by the FDA in its approval of anticancer drugs or by medical oncologists in their treatment decisions. In the Harvard study [37], some phase I trials produced as high or higher rates of response than those used to support FDA approvals. FDA approvals for irinotecan in colon cancer and topotecan in ovarian cancer were based on phase II response rates of 10–15%. In comparison, the response rates among trials in the top decile of phase I trials analyzed by the Harvard group exceeded 13%. Response rates in phase I trials are also comparable to those in the third-line or greater treatment of many solid tumors. Making these considerations more complex is the

understanding that patients confronting death may have higher tolerances for risk than agencies overseeing their clinical research. George Zimmer, an English professor and phase I trial participant, wrote that "the enemy is not pain or even death, which will come for us in any eventuality. The enemy is cancer, and we want it defeated and destroyed. This is how I wanted to die—not a suicide and not passively accepting, but eagerly in the struggle" [68].

## 7 CONCLUSION

Clinical trials have evolved from a time when empiricism was viewed with derision by those who believed medical therapy should be based on theory and a priori understanding of health and disease to the foundation of evidence-based medical practice. Recognition of diseases as discrete entities, rather than personal states of imbalance, transition of health care from home to institutions such as hospitals, and the development of methods for valid comparisons of therapy and control of bias were all critical factors in development of the modern clinical trial. Clinical trials have been instrumental both in preventing the use of ineffective treatments from bloodletting, to bone marrow transplant, to breast cancer and in providing a platform for development of novel therapeutics that have revolutionized medicine and improved outcomes for patients. The progress in clinical trials has been built on the efforts of many diligent investigators and countless patients who have become research subjects. Along the way, we have learned to design better trials and to take better care of patients participating in these trials. Institutions and industry have emerged to regulate and promote clinical research and the development of complex relationships between government, industry, clinical research organization, academia, and community practices to advance medical knowledge safely and efficiently is ongoing. The optimal design of clinical trials remains contested in many areas, ethical questions regarding the quality and requirements of informed consent abound, and debates regarding the balance of evidence-based practice versus clinical wisdom continue.

We are entering an era of targeted therapy based on understanding of the molecular biology of disease and the ability to rationally design drugs for selected targets. In addition, we are finding that understanding the discrete disease entity alone is not sufficient for optimal treatment, and the specifics of the individual patient, from co-morbidities, physiologic status, and pharmacogenomic differences can play a major role in the outcome of any given therapy. Now more than ever before, clinical trials have the potential to transform medicine and improve health, and the dedication of researchers, and most importantly patients who are willing to participate in clinical research are needed to write the next chapter in the history of medicine.

## REFERENCES

1. Brain, P. (1986), *Galen on Bloodletting*, Cambridge University Press, Cambridge, United Kingdom.
2. Haller, J. S., Jr. (1986), Decline of bloodletting: A study in 19th-century ratiocinations, *South Med. J.*, 79(4), 469–475.

3. Antman, K. H., Rowlings, P. A., Vaughan, W. P., et al. (1997), High-dose chemotherapy with autologous hematopoietic stem-cell support for breast cancer in North America, *J. Clin. Oncol.*, 15(5), 1870–1879.
4. Kopperman, P. (2004), "Venerate the Lancet": Benjamin Rush's yellow fever therapy in context, *Bull. Hist. Med.*, 78, 539–574.
5. Garrison, F. (1929), *History of Medicine*, 4th ed., W.B. Saunders, Philadelphia.
6. Morens, D. M. (1999), Death of a president, *N. Engl. J. Med.*, 341(24), 1845–1849.
7. Best, M., Neuhauser, D., and Slavin, L. (2004), "Cotton Mather, you dog, dam you! I'l inoculate you with this; with a pox to you": smallpox inoculation, Boston, 1721, *Qual. Saf. Health Care*, 13(1), 82–83.
8. Gross, C. P., and Sepkowitz, K. A. (1998), The myth of the medical breakthrough: Smallpox, vaccination, and Jenner reconsidered, *Int. J. Infect. Dis.*, 3(1), 54–60.
9. Bender, G. (1966), *Great Moments in Medicine*, Northwood Institute Press, Detroit.
10. Campagna, J. A. (2005), The end of religious fatalism: Boston as the venue for the demonstration of ether for the intentional relief of pain, *Surgery*, 138(1), 46–55.
11. Vogel, M., and Rosenberg, C. (1979), *The Therapetuic Revolution*, University of Pennsylvania Press, Philadelphia.
12. Best, M., and Neuhauser, D. (2005), Pierre Charles Alexandre Louis: Master of the spirit of mathematical clinical science, *Qual. Saf. Health Care*, 14(6), 462–464.
13. Bull, J. P. (1959), The historical development of clinical therapeutic trials, *J. Chronic Dis.*, 10, 218–248.
14. Louis, P. (1836), *Rechearches on the Effects of Bloodletting in Some Inflammatory Diseases*, C.G. Putnam, Boston.
15. Pernick, M. (1983), The calculus of suffering in 19th-century surgery, *Hastings Center Report*, 13.
16. Holmes, O. W. (1843), The contagiousness of puerperal fever. *N. Engl. Quart. J. Med. Surg.*, 1, 503–530.
17. Tan, S. Y., and Tasaki, A. (2007), Joseph Lister (1827–1912): Father of antisepsis, *Singapore Med. J.*, 48(7), 605–606.
18. Fibiger, J. (1898), Om Sreumbehandlung af Difteri, *Hospitalstdende*, 4(6). In Hrobjartsson, A., Gotzsche, P. C., and Gluud, C. (1998), The controlled clinical trial turns 100 years: Fibiger's trial of serum treatment of diphtheria, *BMJ*, 317(7167), 1244.
19. Hrobjartsson, A., Gotzsche, P. C., and Gluud, C. (1998), The controlled clinical trial turns 100 years: Fibiger's trial of serum treatment of diphtheria, *BMJ*, 317(7167), 1243–1245.
20. Yoshioka, A. (1998), Use of randomisation in the Medical Research Council's clinical trial of streptomycin in pulmonary tuberculosis in the 1940s, *BMJ*, 317(7167), 1220–1223.
21. Anon. (2006), Drive for drugs leads to baby clinical trials, *Nature*, 440(7083), 406–407.
22. Kaptchuk, T. (1998), *Bull. Hist. Med.*, 72(3), 389–433.
23. Lambert, S. M., and Markel, H. (2000), Making history: Thomas Francis, Jr, MD, and the 1954 Salk Poliomyelitis Vaccine Field Trial, *Arch. Pediatr. Adolesc. Med.*, 154(5), 512–517.
24. Gelber, R. D., and Goldhirsch, A. (1988), Can a clinical trial be the treatment of choice for patients with cancer? *J. Natl. Cancer Inst.*, 80(12), 886–887.
25. Antman, K., Schnipper, L. E., and Frei, E., 3rd. (1988), The crisis in clinical cancer research. Third-party insurance and investigational therapy, *N. Engl. J. Med.*, 319(1), 46–48.
26. Braunholtz, D. A., Edwards, S. J., and Lilford, R. J. (2001), Are randomized clinical trials good for us (in the short term)? Evidence for a "trial effect," *J. Clin. Epidemiol.*, 54(3), 217–224.

27. Peppercorn, J. M., Weeks, J. C., Cook, E. F., et al. (2004), Comparison of outcomes in cancer patients treated within and outside clinical trials: Conceptual framework and structured review, *Lancet*, 363(9405), 263–270.
28. Roberts, T. (2006), Food and Drug Adminstation role in oncology product development, in Chabner, B. A., ed., *Cancer Chemotherapy and Biotherapy: Principles and Practice*, 4th ed., Lippincott Williams & Wilkins, Philadelphia, pp. 502–515.
29. Chabner, B. A., and Roberts, T. G. (2007), The FDA in 2006: Reasons for optimism, *Oncologist*, 12(3), 247–249.
30. Drug Amendments of 1962, Pub. L. No. 87–781, 76 Stat. 780 (1962), codified as amended at 21 U.S.C. §321.
31. Johnson, J. R., Williams, G., and Pazdur, R. (2003), End points and United States Food and Drug Administration approval of oncology drugs, *J. Clin. Oncol.*, 21(7), 1404–1411.
32. Hennessy, S., and Strom, B. L. (2007), PDUFA reauthorization—drug safety's golden moment of opportunity? *N. Engl. J. Med.*, 356(17), 1703–1704.
33. Woosley, R. L., and Cossman, J. (2007), Drug development and the FDA's Critical Path Initiative, *Clin. Pharmacol. Ther.*, 81(1), 129–133.
34. PRaMoA (2007), *Pharmaceutical Industry Profile 2007*, PhRMA, Washington, DC.
35. Loscalzo, J. (2006), The NIH budget and the future of biomedical research, *N. Engl. J. Med.*, 354(16), 1665–1667.
36. Mathieu, M. (2006), *Parexel's Pharmaceutical R&D Statistical Sourcebook 2006/2007*, Parexel International Corporation, Waltham, MA.
37. Roberts, T. G. Jr, Goulart, B. H., Squitieri, L., et al. (2004), Trends in the risks and benefits to patients with cancer participating in phase 1 clinical trials, *JAMA*, 292(17), 2130–2140.
38. DiMasi, J. A., Hansen, R. W., and Grabowski, H. G. (2003), The price of innovation: New estimates of drug development costs, *J. Health Econ.*, 22(2), 151–185.
39. Annas, G., and Grodin, M. (1992), *The Nazi Doctors and the Nuremberg Code: Human Rights in Human Experimentation*, Oxford University Press, New York.
40. Leaning, J. (1996), War crimes and medical science, *BMJ*, 313(7070), 1413–1415.
41. Anon. (1949), The Nuremberg Code, in *Trials of War Criminals before the Nuernberg Military Tribunals under Control Council Law No. 10*, U.S. Government Printing Office, Washington DC, pp. 181–182.
42. World Medical Association (1996), Declaration of Helsinki (1964), *BMJ*, 313, 1448–1449.
43. Brandt, A. M. (1978), Racism and research: The case of the Tuskegee Syphilis Study, *Hastings Cent. Rep.*, 8(6), 21–29.
44. Beecher, H. K. (1966), Ethics and clinical research, *N. Engl. J. Med.*, 274(24), 1354–1360.
45. National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research (1978), The Belmont Report: Appendix. Vol 1, U.S. Government Printing Office, Washington, DC, Chap. 9.
46. Emanuel, E. J., Wendler, D., and Grady, C. (2000), What makes clinical research ethical? *JAMA*, 283(20), 2701–2711.
47. Canellos, G. P. (1997), Selection bias in trials of transplantation for metastatic breast cancer: Have we picked the apple before it was ripe? *J. Clin. Oncol.*, 15(10), 3169–3170.
48. Mello, M. M., and Brennan, T. A. (2001), The controversy over high-dose chemotherapy with autologous bone marrow transplant for breast cancer, *Health Aff. (Millwood)*, 20(5), 101–117.

49. Weiss, R. B., Rifkin, R. M., Stewart, F. M., et al. (2000), High-dose chemotherapy for high-risk primary breast cancer: An on-site review of the Bezwoda study, *Lancet*, 355(9208), 999–1003.
50. Farquhar, C., Marjoribanks, J., Basser, R., et al. (2005), High dose chemotherapy and autologous bone marrow or stem cell transplantation versus conventional chemotherapy for women with metastatic breast cancer, *Cochrane Database Syst. Rev.*, 3, CD003142.
51. Dorman, P. J., Counsell, C., and Sandercock, P. (1999), Reports of randomized trials in acute stroke, 1955 to 1995. What proportions were commercially sponsored? *Stroke*, 30(10), 1995–1998.
52. Peppercorn, J., Blood, E., Winer, E., et al. (2007), Association between pharmaceutical involvement and outcomes in breast cancer clinical trials, *Cancer*, 109(7), 1239–1246.
53. Djulbegovic, B., Lacevic, M., Cantor, A., et al. (2000), The uncertainty principle and industry-sponsored research, *Lancet*, 356(9230), 635–638.
54. Buchkowsky, S. S., and Jewesson, P. J. (2004), Industry sponsorship and authorship of clinical trials over 20 years, *Ann. Pharmacother.*, 38(4), 579–585.
55. Davidson, R. A. (1986), Source of funding and outcome of clinical trials, *J. Gen. Intern. Med.*, 1(3), 155–158.
56. Bekelman, J. E., Li, Y., and Gross, C. P. (2003), Scope and impact of financial conflicts of interest in biomedical research: A systematic review, *JAMA*, 289(4), 454–465.
57. Boyd, E. A., and Bero, L. A. (2000), Assessing faculty financial relationships with industry: A case study, *JAMA*, 284(17), 2209–2214.
58. Schulman, K. A., Seils, D. M., Timbie, J. W., et al. (2002), A national survey of provisions in clinical-trial agreements between medical schools and industry sponsors, *N. Engl. J. Med.*, 347(17), 1335–1341.
59. Chabner, B. A., and Roberts, T. G., Jr (2005), Timeline: Chemotherapy and the war on cancer, *Natl. Rev. Cancer*, 5(1), 65–72.
60. Meropol, N. J., Weinfurt, K. P., Burnett, C. B., et al. (2003), Perceptions of patients and physicians regarding phase I cancer clinical trials: Implications for physician-patient communication, *J. Clin. Oncol.*, 21(13), 2589–2596.
61. Daugherty, C., Ratain, M. J., Grochowski, E., et al. (1995), Perceptions of cancer patients and their physicians involved in phase I trials, *J. Clin. Oncol.*, 13(5), 1062–1072.
62. Miller, M. (2000), Phase I cancer trials. A collusion of misunderstanding, *Hastings Cent. Rep.*, 30(4), 34–43.
63. Henderson, G. E., Churchill, L. R., Davis, A. M., et al. (2007), Clinical trials and medical care: Defining the therapeutic misconception, *PLoS Med.*, 4(11), e324.
64. Von Hoff, D. D., and Turner, J. (1991), Response rates, duration of response, and dose response effects in phase I studies of antineoplastics, *Invest. New Drugs*, 9(1), 115–122.
65. Decoster, G., Stein, G., and Holdener, E. E. (1990), Responses and toxic deaths in phase I clinical trials, *Ann. Oncol.*, 1(3), 175–181.
66. Estey, E., Hoth, D., Simon, R., et al. (1986), Therapeutic response in phase I trials of antineoplastic agents, *Cancer Treat. Rep.*, 70(9), 1105–1115.
67. Agrawal, M., and Emanuel, E. J. (2003), Ethics of phase 1 oncology studies: Reexamining the arguments and data, *JAMA*, 290(8), 1075–1082.
68. Daugherty, C. K., Siegler, M., Ratain, M. J., et al. (1997), Learning from our patients: One participant's impact on clinical trial research and informed consent, *Ann. Intern. Med.*, 126(11), 892–897.

# *Adaptive Research*

**Michael Rosenberg**
*Health Decisions, Inc., Durham, North Carolina*

**Contents**

1 Types of Adaptive Techniques: Strategic and Operational
2 Strategic Adaptations
  2.1 Drug Performance and Rising Dose Escalation Studies
  2.2 Adaptive Dose Finding
  2.3 Sample Size Reestimation\
  2.4 Adaptive Randomization
  2.5 Seamless Designs: Rolling One Phase into Next
  2.6 Other Strategic Adaptations
3 Operational Side of Adaptive Research\
  3.1 Enrollment
  3.2 Site Performance
  3.3 Adaptive Site Monitoring
4 Implementation Issues
  4.1 Data Capture and Validation
  4.2 Planning
  4.3 Process Optimization
  4.4 Decision Making
  4.5 Adaptive Data Monitoring
  4.6 Regulatory Considerations
5 Promise of Adaptive Methods
  References

Adaptive research denotes an approach to clinical trials that incorporates what is learned during the course of a study or development program into how it is completed, without compromising validity or integrity. Adaptive components need

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

not be confined to the frequently encountered but unduly narrow vision of enabling changes in a study's design, valuable and interesting as such changes are. Rather, adaptive methods may encompass potential changes in all program-related resources and activities, including changes in logistical, monitoring, and recruitment procedures, and sometimes even personnel and travel requirements. The goal of adaptive methods is making better and more timely decisions to allocate all study resources more efficiently, reduce costs and timelines, and better achieve informational goals compared to traditional study and program approaches.

Efficient management is particularly important in activities as complex as clinical research, which involves a range of activities that includes patient recruitment, randomization, supply chain logistics, and flow of information. Additional complexity commonly arises when pharmaceutical studies are conducted at multiple sites, often in different countries, cultures, and languages. Effective management of clinical trials requires continuous monitoring and measurement of numerous activities. The essence of adaptive research is to continuously measure progress in the many aspects of a complex study, learn from such measures, and, based on what is learned, act expeditiously to make changes to improve the remainder of the study and even an entire development program.

On a pragmatic level, adaptive studies not only require the ability to measure outcomes of interest continuously but also to make data and summarized information about those measurements available in a timely manner to different audiences according to study role. This is essential for effective study management. In a clinical context, this means not just continuously tracking trial data collected on case report forms but also generating performance metrics that enable refinements in operations. This learn-as-you-go approach contrasts with the traditional black-box methodology of clinical trials in which data and particularly operational indices are often lacking altogether or available too late to enable study personnel to respond: Clean data are generally not available until after a study is completed, and study performance metrics such as recruitment rate, reasons for screen failures, and the like are often lacking entirely.

Interest in adaptive methods has mounted as a result of the soaring cost of clinical research and numerous trial failures, including particularly costly and well-publicized failures of major late-stage trials. In addition to greater efficiency, adaptive methods provide a number of appealing advantages, such as a more nuanced view of product performance that may enable earlier strategic decisions about appropriate target populations, earlier warnings about ineffective trials, and a broader view of research programs as continuous, integrated activities rather than a staccato, linear series of separate trials with inevitable delays in between.

Adaptive methods seek to bring "gold-standard" trial methodology up to date by taking advantage of the many technological advances during the 60 years since the trials that established comparative clinical trials as a means of assessing a pharmaceutical product's performance in the late 1940s [1]. Particularly salient are markedly more efficient approaches to data capture and validation, the growing power of affordable desktop computers, and the ability of the Internet to move data throughout the world easily, quickly, and cheaply.

Advances in data capture and validation open possibilities for much earlier understanding of study performance and trends. Computational advances stimulated the development of statistical methodology enabling midcourse "looks" at study progress based on data collected to date. The most common example is

sequential analyses, which preserve design integrity but inform decisions that improve the course of the study. The simplest result of such an interim analysis is early stopping for futility. Although statistical methodology is beyond the scope of this chapter, both books and software have become available that focus on adaptive research. See, for example, Chow and Chang [2] and Hu and Rosenberger [3].

An additional benefit of adaptive approaches is enabling a more nuanced perspective into candidate performance, following the realization over time that a simple "yes–no" answer as to a drug's efficacy is likely an oversimplification. Adaptive trials maintain the same high standards of scientific integrity and reliability as the standard methodology while dramatically improving operational and economic efficiency and informational breadth, depth, and quality.

Adaptive research also allows clinical researchers to employ the same basic management principles as typical modern businesses, using real-time data and analysis to inform decisions that continually optimize operations. Figure 1 contrasts the continuous acquisition of knowledge in adaptive studies with a conventional trial that does not generate clean, meaningful data until much later. Adaptive methods require continuously updated operational performance metrics; the conventional approach often lacks the real-time data essential to make such metrics available to improve trial management.

Preserving study integrity—the ability to perform unbiased clinical evaluations—is paramount with all clinical evaluations, including adaptive ones. While commonplace in other industries, managing in response to changing data is fairly new in clinical research because excluding bias has been achieved primarily by denying access to data, including much of the data that would be useful for trial management. Like conventional trials, adaptive research still relies on techniques to exclude bias, including blinding. However, adaptive trials also employ additional planning and special operational procedures that prevent those performing the study from accessing unblinded results data. Only designated individuals have access to the information required to make decisions about specific adjustments during the trial. Firewalls

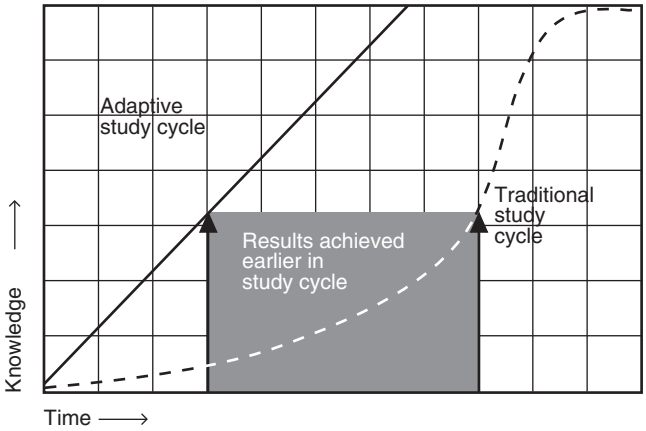


***FIGURE 1*** *With the traditional approach (dashed line), most knowledge is acquired at the end of the trial. An adaptive approach uses newer techniques to capture and analyze data continuously, to support both scientific inferences and performance measures—knowledge that managers can use to support decisions to tightly manage the trial. The difference between the two approaches represents the ignorance handicapping the managers of a traditional trial. (Figure copyright 2006, Health Decisions, Inc. Used by permission.)*

must be incorporated from the outset to ensure that decision makers cannot jeopardize, whether knowingly or not, the study's scientific integrity. Fortunately, sophisticated computer access control and data encryption techniques provide useful tools for controlling the dissemination of information and potential sources of bias.

Developing protocols for adaptive trials demands more attention than conventional planning because multiple scenarios must be considered and specific plans included for addressing each. For example, a study that involves midstudy elimination of dosing arms (pruning, a common technique in dose-finding trials) normally includes safeguards to ensure that investigators are blinded as to which dose is being eliminated. Maintaining the blind may require logistic changes such as different packaging of study supplies. Additional examples of measures to exclude bias are included in Section 4.

## 1 TYPES OF ADAPTIVE TECHNIQUES: STRATEGIC AND OPERATIONAL

Any study involves both a strategic design and operational plans. Adaptive techniques similarly fall into the two broad categories of strategic adaptations and operational components. Strategic adaptations refer to changes in the study's design, such as the number of subjects to be included or how patients are to be allocated to treatment arms during the study. Operational adaptations focus on how the study is run: recruitment, enrolling patients, improving data quality (measured by number of queries generated), assuring timely responses to data discrepancies, detecting site performance problems early, and efficient means of allocating resources such as field monitors. Although much recent discussion focuses on the strategic components of adaptive research, the operational components can be at least as beneficial to timelines and budgets. Moreover, operational adaptations have the advantage of not requiring regulatory approval. Operational adaptive methods enable far tighter trial management through continuous monitoring and refinement of the many elements involved, resulting in savings in time and expenses of 20% or more compared with traditional management techniques. In the light of ever increasing study budgets, the savings from operational adaptations have strong appeal.

There are several prerequisites for both strategic and operational components of adaptive research: the continuous, timely collection of data and generation of metadata (data *about* the data, such as performance metrics that might include rates of enrollment, screen failures, queries, and the like); the ability to rapidly turn a stream of raw data into meaningful information; and the ability to present information in different forms to meet the needs of study staff performing different functional roles. All these are essential to enabling specific informed actions to improve trial operations. Indeed, at many points during a trial, the measures of how the study is progressing may demand greater attention than the actual data collected as part of the clinical evaluation; the need for trial data may be intermittent (e.g., at an interim analysis), while the need for metadata to inform management decisions is continuous.

Efficient electronic data collection (EDC) is an absolute requirement for adaptive research. Unfortunately, not all EDC is efficient in practice. EDC for adaptive purposes must be able to capture data in electronic form shortly after it is generated and quickly transfer it to a central database. Many EDC systems currently used in

the pharmaceutical industry rely on hand keyboard entry ("Web-based" EDC systems), which result in unnecessary delays in data acquisition. Such delays typically range from several days to several weeks, reflecting the lack of enthusiasm among clinical personnel for performing the tedious chore of keyboard data entry. A second and more important shortcoming of many EDC systems is their inability to track metadata, or the metrics that enable effective study management. EDC systems that fail to provide an integrated means of addressing important dynamic components of adaptive trials such as supply management and patient randomization can induce as many problems as they solve. The fact that Web-based EDC systems have existed for more than 7 years and yet the majority of clinical trials continue to be done with paper and pen reflects the industry's experience that Web-based EDC is expensive, difficult to implement and maintain, and often fails to deliver bottom-line benefit. Alternatives to Web-based EDC for adaptive trials are included in Section 4.

While substantial improvements in efficiency may be gained through strategic adaptive methods, the optimal approach to clinical trials incorporates both strategic and operational elements. Combining both elements promises to go beyond improving the mechanics of individual studies to improve how studies and development programs are managed. Like the navigators of old who planned long voyages based on inaccurate maps and poor or nonexistent navigational equipment, planners of clinical trials today must start out with plans based on guesses. In clinical trials, the guesses are about key trial parameters such as the size of the treatment effect, variability of data, dropout rates, and even the appropriate range of dosages to test. Modern drug companies, like the early explorers, have often learned too late that their best planning efforts have landed them not at the intended destination with rich economic rewards but in unforeseen surroundings with bleak prospects. Adaptive techniques represent a GPS (Global Positioning System) for present-day clinical trials: We must still make initial guesses, but the capacity to make multiple midcourse corrections, on many levels, is an integral part of the study. Rather than relying on a series of discrete decision points, an adaptive approach substitutes a process of continuous assessment and response. As a result, study staff is no longer condemned to making guesses and then sticking with them until the study is done, only then discovering how each guess compares to reality.

Clinical development is evolving from the traditional model of discrete phases punctuated by pauses to a continuous process. The traditional model is one of defining safety and kinetics (phase I), then defining optimal doses (phase II), then pivotal studies required for regulatory approval (phase III). The emerging integrated model is largely made possible by an adaptive approach that renders the process continuous, with faster decision making along the way, the flexibility to shorten (or lengthen) certain components in response to data generated, and minimizing or eliminating the gap between studies. Indeed, this so-called learn–confirm model can extend beyond development to postmarketing, maximizing learning about a product throughout its life cycle, minimizing postmarket problems, and maximizing opportunities.

The following sections that discuss specific adaptive techniques are arranged according to the chronology of development, starting with the earliest clinical studies in a product's life cycle, efforts to define the safety envelope, and the balance between efficacy and safety. We then progress through larger studies that focus on

better definition, selection of dosing for large-scale confirmatory (pivotal) studies, and finally pivotal studies. An important difference in how study design is approached, especially in regard to adaptive techniques, between earlier studies to define safety and dosing and the pivotal studies that serve as the basis for marketing approval, is that regulatory bodies grant a greater degree of autonomy in the early (learning) evaluations (subject, of course, to the broad constraint of not jeopardizing the safety of subjects involved in evaluations). Thus, the learning phase currently offers greater opportunity to apply adaptive approaches. The use of adaptive methods in the learning phase allows earlier, more accurate determination of promising and less promising products, reduced development time and expense, and reduced wasting of resources on candidates that eventually fail to reach market.

## 2 STRATEGIC ADAPTATIONS

### 2.1 Drug Performance and Rising Dose Escalation Studies

The earliest phases of development involve confirmation of pharmacokinetic (PK) and pharmacodynamic (PD) assessments from animal models and determination of approximate dosing ranges for human administration. Such studies are usually performed in "normal" (disease-free) populations and may, to a limited degree, explore these parameters in individuals for whom the product is ultimately targeted. The keys for these early studies are rapid acquisition and assessment of data and timely decision making, minimizing the interval between when a dose is administered and when a decision can be made with respect to the next dosing.

For PK and PD studies as well as early studies that rely on these parameters (e.g., dose escalation), assessments are normally batched to reduce the expense involved in running samples. However, the balance between the reduced costs of batched assessments—always a choke point—and the potential economic gains from earlier availability of assessment results, may warrant a reexamination of traditional reliance on batching samples. Each assay and product under evaluation differs, but researchers should at least consider the alternative of obtaining data from individual tests earlier.

Rising-dose escalation studies illustrate the critical nature of rapid, accurate data in achieving high velocity in clinical studies and programs. The interval between dosing, assessment, and a decision about progressing to the next higher dose often consumes 6 weeks. Adaptive techniques, with their focus on the rapid collection and assessment of data, as well as rapid dissemination of information for appropriate action, may reduce this interval to days or even a single day. This can be accomplished by (1) exploiting the availability of data very quickly after drug administration and collection of safety data, (2) using a system that efficiently collects and summarizes data, and (3) disseminating and displaying the information via a universal mechanism such as the Web.

Figure 2 shows a system that meets these requirements and enables next-dose decisions to be made within 2 days of the last dose; this rate-limiting step is dictated by a 48-hour primary observation period after administration. The requirement for rapid data availability mandates use of a digital pen (Fig. 3) for data collection along with processes and software that enable the data for each individual to be

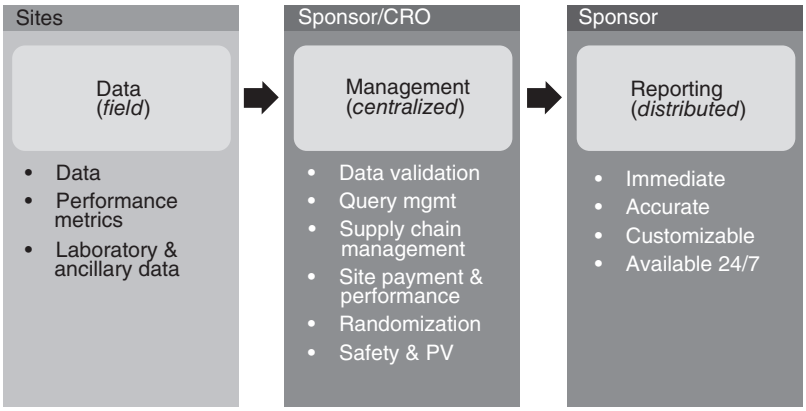


***FIGURE 2*** *Immediate, clean data requires the ability to collect, digest, and report a variety of data and performance metrics. While data collection systems are common, these may take extended periods to enter data. Most systems currently in use lack the "middleware" (central box) necessary for rapid data collection and validation as well as the ability to report data rapidly. In addition, few systems routinely generate performance metrics, such as indicators of recruitment progress and issues, essential to running a tight study that minimizes timelines. (Figure copyright 2007, Health Decisions, Inc. Used by permission.)*

***FIGURE 3*** *A digital pen used for data capture. This method electronically reads in data, eliminating the requirement for manual data entry. (Figure copyright 2006, Health Decisions, Inc. Used by permission.)*

displayed within minutes of when it is collected. Summaries that show data on all subjects are displayed over the Web, enabling timely access by decision makers regardless of location or time.

The benefits of this system are significant. First, the usual bottleneck the delays getting clean data is eliminated, making each subject and the collective experience of all subjects immediately available to appropriate members of the study team. A stream of raw data is transformed into meaningful information, summarized with drill-down capabilities, and presented in a manner that supports decisions about next steps. Second, decisions are more timely, nuanced and better informed because accumulated information can be tracked and analyzed along the way, rather than

waiting until the end to examine all data. Third, the ability to make this information immediately available to all decision makers regardless of location drives timelier and more efficient study management across all offices and sites associated with a study. Such an approach can easily cut four or more weeks from each decision cycle. With an early-phase study, this can translate to a reduction of 50% or more in study timeline.

## 2.2 Adaptive Dose Finding

Conventional dose-finding studies typically utilize a modest number of equal-sized treatment arms, generally three to four, often with a comparator arm. Each arm is administered a different dose for a given period, data are examined and, hopefully, justify a decision to proceed with the most promising doses into pivotal testing (the "confirm" part of the learn–confirm model). Planning for this approach involves selecting the most promising doses, then bracketing them; in effect, this builds on earlier dose escalation studies that gave a sense of the same information but with fewer subjects.

Dose-finding studies are conducted with the goal of winnowing dosing choices to a small enough number to allow proceeding to larger, more expensive pivotal trials. Compared to earlier studies, these involve more subjects receiving each dose, yielding more data, enabling greater understanding and greater certainty about efficacy and safety. Such studies are often planned ("powered") with enough subjects to achieve statistical differentiation between dosing arms for the efficacy outcome. Although reassuring to decision makers, this approach may not be necessary to achieve the objective of rejecting some of the dosing arms. Adaptive techniques make possible earlier decisions to reject less effective or less safe doses, thus conserving resources and time. In addition, the ability to examine data as it accumulates can lead to superior understanding of the data and contribute to better decisions about doses to carry forward.

Under an adaptive scenario, information is made available as data are generated and go/no-go decisions can be made not at some predetermined interval but whenever the information generated yields sufficient knowledge to justify the step. In many instances, less desirable treatment arms (outliers) will be apparent early on. A key benefit of the adaptive approach is the ability to accumulate adequate information early in the trial to justify eliminating less promising arms, allowing the concentration of resources (notably study subject) to arms that show greatest promise. In such early learning-phase trials, regulators generally allow sponsors the latitude to forego a rigidly defined decision point in favor of making decisions on dosing arms when the sponsors are comfortable that sufficient information has accumulated (sponsor's risk). Adaptive methods provide an opportunity for sponsors to make earlier, better informed judgments without pushing every arm to a sample size that provides statistically significant results. Rather, there is the flexibility to discontinue a treatment arm as soon as it becomes apparent that it is less desirable than others.

The adaptive approach changes the way dose-finding evaluations are performed. Importantly, since some arms can likely be eliminated early, there is the luxury of starting with a larger number of arms than would be possible with the traditional approach. For example, in a study of a neuroprotective agent administered a single time by IV (intravenous) soon after stroke, researchers established a procedure for

selecting 1 of 16 possible doses to be given each patient, with the dose in each instance selected based on data observed to that point in the trial. In the actual conduct of the trial, 15 different doses were tried [4].

The ability to test more arms at this stage means that earlier development often need not be as thorough. And because arms are cut off early, there is little or no extra expense in gathering data on a greater number of dosing arms and then focusing resources where they are most needed—in differentiating the final two or three dosing arms rather than in accumulating additional information about arms that data already show lack promise. Another benefit of earlier termination of less promising arms is that fewer patients are exposed to less efficacious and less safe doses. In addition, the ability to examine data as they are generated—where each new patient, visit, and evaluation adds to an existing storehouse of information, and trends may emerge from the changing data—often provides a far more nuanced perspective on product performance than a single cross-sectional view at the conclusion of an evaluation.

Figure 4 illustrates how an adaptive dose-finding study is conducted. If the informational goal is defined by 80 patients, then we begin with a number of dosing arms, in this case 8. Early on, data will likely make it apparent that many of these arms

| Time (mo) | 2 | 4 | 6 | 8 | 10 | 12 | 14 | 16 | TOTAL |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | cost/pt | $15,000 |
| ADAPTIVE | | | | | | | | | |
| | 10 | | | | | | | | 10 |
| | 10 | 12.5 | 12.5 | 12.5 | | | | | 48 |
| 50/mo | 10 | 12.5 | 12.5 | 12.5 | 16.7 | 16.7 | | | 81 |
| | 10 | 12.5 | 12.5 | 12.5 | 16.7 | 16.7 | | | 81 |
| comparator | 10 | 12.5 | 12.5 | 12.5 | 16.7 | 16.7 | | | 81 |
| total enrolled | 50 | 50 | 50 | 50 | 50 | 50 | | | 300 |
| #/arm/mo | 10 | 12.5 | 12.5 | 12.5 | 16.7 | 16.7 | | | |
| cost | | | | | | | | | $4,500,000 |
| NON-ADAPTIVE | | | | | | | | | |
| | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 80 |
| | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 80 |
| 50/mo | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 80 |
| | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 80 |
| comparator | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 80 |
| total enrolled | 50 | 50 | 50 | 50 | 50 | 50 | 50 | 50 | 400 |
| #/arm/mo | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | |
| | | | | | | | | TOTAL | $6,000,000 |

***FIGURE 4*** *Comparison of adaptive (top) and traditional dose-finding studies. Dropping less promising arms early on (pruning) allowed this study to achieve its informational goals four months sooner, at a cost $1.5 million less, by monitoring incoming data from the study's outset and focusing resources on the most promising arms. After the first observation period, one arm (T3) is terminated because of poor results for efficacy, safety or both. The remaining arms are continued to the next observation period, at which time an additional arm (T2) is similarly cut. The remaining dose arm and the comparator (P) may under certain circumstances then be rolled into the pivotal evaluation, utilizing data already gathered. (Figure copyright 2006, Health Decisions, Inc. Used by permission.)*

are less promising and can be cut off earlier than is customary. Although Figure 3 shows discrete, evenly spaced decision points, these are in practice more likely to be irregular, with the most obvious decisions coming early in the study. The result of this approach in this example shows how the desired number of patients is reached earlier and at a lesser cost than would be the case if we had taken a traditional approach of using, say, four dosing arms that enroll 80 patients each. If each patient costs $15,000, the adaptive study will be completed 25% faster (saving 4 months) and cheaper (saving $1.5 million). The cost savings are even greater in the light of the institutional costs of maintaining a company for an extra 4 months, a factor that is particularly compelling for developing companies where such institutional costs can be a key determinant of success and, in some cases, survival.

### 2.3 Sample Size Reestimation

The progression of a compound into large-scale testing demands major commitments of time and money. A key driver of both study duration and cost is the number of patients required for successful completion of the study. The appropriate sample size is determined based on the magnitude of difference in outcome measure between the product being evaluated and a comparator, acceptable levels of possible error (risk of false-positive or false-negative results), variance of data, and rates of subject compliance and drop-out. With luck, initial estimates of these factors may be off only modestly, but they will still be off. If the estimates err more than modestly, the consequences for sample size requirements can be enormous. For example, reducing the estimate of treatment effect by one-half can quadruple the sample size for a fixed sample test and will require the maximum sample size for a group sequential test [5].

Underestimation of sample size may result in a study that fails to detect a difference between the test drug and the comparator, even if one is present; overestimation of required sample size wastes time, money, and other resources. The severe penalty for underestimation means that, in practice, sample size estimates err on the high side, with the consequence that time and money will often be spent unnecessarily in overcompensating for the possibility of falling short of the sample size necessary for detecting a difference. It is therefore not surprising that much research is devoted to approaches to adjusting sample size during clinical trials and a variety of methods are in use [6].

The benefit of an adaptive approach is that it enables an informational goal to be met precisely, arriving at a defined goal of informational content rather than by surrogate measures, neither undershooting nor overshooting. In this context, reestimation means adjusting each of the initially estimated parameters based on data that has actually been observed in the study rather than continuing to rely on the earlier guesses made without benefit of such observations. Reestimation can be employed multiple times for multiple course corrections, although careful attention must be paid to statistical techniques to ensure that design integrity is preserved. Sample size reestimation can be viewed as an extension of well-accepted group sequential trial designs that have evolved over the past two decades.

The value of sample size reestimation is illustrated by a recent oncology study, where this technique allowed the study to be completed 9 months sooner than originally planned. An interim analysis demonstrated an effect size ($\delta$) considerably

stronger than originally anticipated. The savings from the resulting sample size reduction also extended beyond recruitment itself, eliminating the cost of treatment, supplies, monitoring, and follow-up for each patient that might have been included in excess of the number required to meet the study's informational goals. The oncology study's use of sample size reestimation contributed to a savings of $16 million in development costs. Even greater financial benefits flowed from a 9-month reduction in time to market.

### 2.4 Adaptive Randomization

Adaptive randomization can provide greater efficiency by altering the probability that a new subject entering the study will be allocated to a treatment arm that accumulating trial data shows to be more desirable and by ensuring that the desired statistical power is preserved [3]. Early on, when little is known about patient response to different treatments, a subject will be equally likely to be randomized into any treatment arm; during the course of the study, as outcome information accumulates, the randomization ratio can be continuously changed to favor the more beneficial or safe outcome (response-adaptive randomization), to balance covariates (risk factors that modify the probability of an outcome) across different treatment arms (covariate-adaptive randomization), or to balance an undesired deviation from the intended allocation ratio (treatment-adaptive randomization) [2]. There are also sophisticated combinations, such as covariate-adjusted response-adaptive randomization, a procedure that takes into account previous patient responses to treatments, previous patient covariates, and the covariates of the patient to be randomized [3].

In contrast, normal fixed allocation schedules ensure that each patient entering the study will have the same probability of allocation to each arm throughout the life of the study regardless of whether study data shows some arms to be unpromising by reason of lesser efficacy or safety. Continuing with fixed allocation ratios despite evidence of lesser efficacy or safety raises ethical concerns. When data shows an imbalance in covariates between treatment groups, continuing with a fixed allocation procedure can undermine the ability to draw valid inferences about differences in treatment effect. Moreover, continuing with a fixed allocation in the unfortunate event that an imbalance develops in the size of actual treatment groups can undermine statistical power and thus jeopardize the validity of the trial as a whole.

With response-adaptive randomization, allocation ratios are most commonly changed based on favorable outcome (assuming safety not to be an issue). There are different algorithms for achieving this goal. The most common is the randomized "play-the-winner" scheme, which assumes that the outcome for the previous patient is known before the next patient is assigned a treatment group. If the treatment for the preceding patient has a positive outcome (a success), an additional "ball" representing that patient's treatment group is added to the randomization pool; if the preceding patient has a negative outcome, no "ball" is added to the urn. If randomization were based on a bowl of black and white Ping-Pong balls, the study would begin with the same number of white and black balls. Over time, however, if one treatment is more beneficial than others, the randomization pool would progressively be seeded in favor of that treatment [2].

Response-adaptive randomization offers the additional advantage of collecting more data on patient response to the drug under test in the doses that are likely to reach market. This may result in greater understanding of the drug's behavior in the formulation that physicians will prescribe and patients will receive after approval. Physicians and patients will both benefit from superior prescribing information. When deciding whether to approve the drug, regulators may also benefit from the availability of more and better information on the doses that will be prescribed in the event of approval.

Covariate-adaptive randomization seeks to balance covariates across treatment groups by weighting the randomization procedure to increase the number of patients with certain covariates in the treatment group or groups in which these covariates have turned out to be underrepresented [2, 3]. (In practice, covariates can also be accounted for through mathematical techniques such as multivariable analysis.)

Treatment-adaptive randomization balances the number of patients assigned to each treatment group by using any of several weighting schemes based on adjusting the number of hypothetical balls in an urn in favor of the treatment group with lagging membership or creating the algorithmic equivalent of a coin that is biased in favor of that treatment group [2].

Adaptive randomization can be combined with a Bayesian approach to study design. This allows patients to be randomized to a specific arm in direct proportion to the degree of promise it shows relative to other arms. Thus, as experience accumulates and successful treatments are retained while unsuccessful ones dropped, the allocation ratio is progressively slanted in favor of the successful treatment. When a certain balance has been achieved, the study is declared to be completed. For example, if a study begins with a 1:1 randomization scheme but then adds each successful outcome back to the randomization pool over time and drops the unsuccessful outcomes (failures), the randomization pool becomes successively seeded in favor of the successful arm. A Bayesian approach allows a study to be defined as completed when a predetermined proportion is reached, say, 95% of patients are randomized into a particular arm, reflecting the experience of far greater success with that arm. The Bayesian approach is intuitively appealing in that it incorporates each new piece of information on success or failure of treatment into how the study proceeds from that point onwards. The Bayesian approach is also more broadly appealing in that it provides a way to minimize patient exposure to less successful arms on a continuous basis. In practice, however, Bayesian approaches may be complicated by the complex definitions of success and the requirement to consider safety, which is difficult to measure on a dichotomous scale.

The mechanics of adaptive randomization and a Bayesian approach require a centralized, real-time randomization system in addition to the other prerequisites for conducting adaptive studies, such as efficient data capture and cleaning, previously discussed. Besides supporting adaptive randomization, centralization and the immediacy of the electronic process also provide the ability to stop enrolling patients immediately when the target population size is reached, eliminating unnecessary effort and expense that are inevitable in less centralized systems. Even more important, the ability to stop patient enrollment instantly reduces the exposure of additional patients to less desirable treatments.

### 2.5 Seamless Designs: Rolling One Phase into Next

With growing appreciation of drug development as a continuous process, the notion of defining best doses and then rolling straight into pivotal studies utilizing the identified best doses becomes compelling. The strongest reasons for doing so are (1) minimizing the delay, often 12 months or more, between dose-finding studies and initiation of pivotal studies; (2) the efficiencies gained from going through the trial startup process a single time rather than twice; (3) a head start recruiting investigators and test subjects for the pivotal study; (4) the potential to adapt the population of the confirmatory phase based on data on responsive subgroups from the learning phase; and (5) the ability to combine data from the learning phase and the confirmatory phase in conducting a final analysis of trial data.

However, it is important to recognize as well that the break between phases can be an important time to analyze data and perfect the final design of a separate pivotal study. Study managers should weigh the possible need for an interval to allow greater analysis and more refined planning against the potentially huge savings in time and expense if phases can be combined. Additionally, managers should weigh the need to take advantage of an interim period for discussions with regulatory authorities to assure their acceptance of the plan as a basis for approval in the event of a successfully executed study.

A large study that is progressively refined from the phase II dose-finding stage and then continues into the phase III confirmatory study is among the most complex adaptive strategies but also the most rewarding. Executing this strategy is demanding because it requires a great deal of advance planning to anticipate and deal with different possible outcomes. In addition, the final statistical analysis may be extraordinarily complex.

The use of adaptive methods to combine phase II and phase III studies begins with establishing a number of dosing arms and pruning those down to the manageable two to three (including a comparator) for a confirmatory stage. Once the final doses have been identified, the study is then quickly expanded and, assuming a second pivotal study is warranted, it is implemented immediately (Fig. 3).

Simulations can be extremely useful in modeling possible outcomes and their ramifications. Adopting the strategy of rolling a phase II study into phase III requires confidence that all the methodologies and infrastructure essential for the adaptive approach are in place and functioning well—quick and accurate data capture, rapid data validation, the prompt generation of meaningful information, readiness for continuous decision making, and the capacity to manage logistics of the study, such as supply chain management, with great efficiency based on timely data. The benefits of rolling a phase II study into phase III are commensurate with the effort: This approach can easily reduce development time by a year or more and save many millions of development dollars.

Eliminating the gap between phases is only one of the benefits of rolling directly from a phase II study into phase III, conducting, in effect, one continuous study instead of two separate ones. This has the added advantage of incorporating in the phase III study the knowledge gained from relevant arms in the early dose-finding portions. There are also benefits for patient recruitment and data collection. Patients already screened in the dose-finding phase can be enrolled in the pivotal portion of the study, and existing patient data collected in the dose-finding phase can contrib-

ute to greater efficiency in the pivotal phase. However, it is important to be mindful that the protocol must not change from that specified at the initiation of the study; this places a premium on the challenging task of anticipating all outcomes when planning the study. Despite the challenges, the risks associated with combining phases II and III are low. The worst-case scenario for such an adaptive study, with no successful adaptations carried out, is the equivalent of conducting the same study conventionally, without the use of adaptive methods.

Combining data from the two phases requires procedures to control the type 1 error rate for the comparison of the test drug with the control. In addition, the final statistical analysis for the combined data from the learning and confirmatory phases will be more complex than usual [2]. It is also important to think through the plan for the seamless trial to ensure that the issues that might be analyzed between two separate studies are adequately analyzed in advance.

One of the greatest advantages of combining phases is that the entire startup process is handled once instead of twice. A single protocol can be developed, reviewed, and approved in a single process for both phases. Recruitment of investigators and patients is simplified because some of the investigators for the pivotal phase will already be familiar with the study. However, combined studies also require that many other investigators be set and ready to go when the final dosing decision is made. This raises a host of other issues such as making appropriate provisions for addressing an investigational review board (IRB) and making consent forms and study materials available in a timely manner. Considerably greater attention than usual is required to ensure that all requisite components mesh.

### 2.6 Other Strategic Adaptations

Another important class of adaptive techniques lays particular emphasis on extending the period of assessment and decision making. For example, it is possible to plan an adaptation that allows changing the test hypothesis from superiority to noninferiority, and to redesign multiple endpoints to update correlations or change the hierarchical order. It is also possible to establish a decision rule with criteria for determining whether to refocus the study on a subpopulation that is specified in advance. Noninferiority studies may be handled quite differently, especially in regard to hierarchical outcomes: These may, for example, declare success for an initial noninferior phase and then continue the study with the hope of generating data to support a superiority claim. The advantage of this approach is the potential for earlier marketing of the product. Obtaining data to support a noninferiority claim requires a more modest sample size and may yield results that allow the product to enter the market with a noninferior approval while the study continues to accumulate information that may support a superiority claim.

The range of techniques for strategic adaptations continues to evolve with the refinement of existing methods as well as the addition of new approaches and tools.

## 3 OPERATIONAL SIDE OF ADAPTIVE RESEARCH

The fundamental principles of adaptive research apply not only to the sophisticated and innovative techniques for strategic adaptations but also to many of the activities

common to virtually all clinical studies, whether or not they involve strategic adaptations. Implementing efficient data capture, rapid data cleaning, generation of a range of performance metrics, and readiness for informed decision making based on such information can produce dramatic gains in efficiency. For historical reasons, the pharmaceutical industry makes little use of such capabilities today. Nonetheless, there is no reason why the industry should not take advantage of what amount to the same management techniques already used by most contemporary businesses to bring clinical studies into the modern era.

Since the necessary changes for operational adaptations need not affect study or program design, they do not require regulatory approval. Operational adaptations can therefore be implemented immediately, and their benefits are at least as profound as those flowing from strategic adaptations. One clear illustration is a large, complex phase III evaluation of an Alzheimer's drug candidate, where efficient collection of data and performance metrics enabled the completion of patient enrollment in record time and the closing of the database within 2 weeks of the last patient visit. As a result, this study saved 1.6 years and $32 million in direct costs measured against the sponsor's internal projections of 5 years and $100 million [7, 2000; this article can be found online at http://www.healthdec.com/media/articles/AnAlzheimersDrugGoesonTrial.pdf].

As previously noted, the same capabilities involved in making such gains in efficiency are required for all types of adaptive studies. In a broader sense, such capabilities represent the application of principles of tight management to the complex realities of clinical studies. Other industries have shown the way. The principle of just-in-time inventory brought new efficiencies to the automobile industry; the same principle, managing operations based on continuous, real-time information about important business processes, has been widely adopted in manufacturing and other highly competitive industries. While pharmaceutical development is considerably more complex and knowledge based than manufacturing, intelligent management can apply the same general principles to clinical studies while preserving study integrity and validity through careful operational controls and information management, preserving blinding, randomization, and other hallmarks of clinical research. The need for specific measures to exclude the possibility of bias does not, in an age of sophisticated access control systems, require near total ignorance for all study personnel of all study operations until the very end, when it is too late to take advantage of data and performance metrics for effective trial management. The need to remedy major shortcomings in the efficiency of current development practices is evident; so is the availability of methods with the potential to solve current problems and make dramatic, rapid improvements in efficiency.

Surprisingly, one critical requirement for tight study management is frequently overlooked: the timely reporting of performance metrics. Effective management is impossible without timely, accurate information about performance; in clinical studies, achieving tight management through operational adaptations requires the same capabilities as strategic adaptations: rapid data collection from the field, rapid data cleaning, timely analysis and summarization, and, importantly, presentation of information in different forms meaningful to staff performing different functional roles. The study manager, for example, may be centrally interested in knowing why certain sites are enrolling faster than others and will therefore want to track frequency of screen failures and the distribution of different reasons for them. The

field monitor may wish to know how to help a site decrease its query rate, allowing her to spend less time in managing minutiae of the study and, thanks to more accurate data collection and the reduced incidence of queries, more time helping sites conduct the study efficiently and achieve database lock faster. The head of R&D may be most interested in the projected dates for completion of enrollment and database lock. In summary, good performance metrics enable greater understanding of study progress, far tighter control, more effective allocation of resources such as monitoring time, faster enrollment, and, in the larger scheme of things, shorter timelines and lower costs.

### 3.1 Enrollment

Closely monitoring the progress of enrollment and the incidence of reasons some patients cannot be enrolled allows continuous tuning of enrollment strategies and criteria. Comparison of site performance, especially in larger studies, generally reveals a wide range of performance. The study manager's job is often to use such information to dig deeper, determine the reasons some sites lag in enrollment while others excel, and intervene as necessary to improve enrollment across the entire study. In some cases, close scrutiny of enrollment performance at different sites may reveal difficulties repeatedly encountered despite best efforts. In that event, case study managers may confront a decision about taking measures to overcome suboptimal enrollment that is not due to operational inefficiencies, such as adding sites.

Comparison of site performance usually does reveal that certain sites are enrolling more effectively than the rest. The reasons may vary; the ability to track a variety of enrollment metrics in real time enables a manager to determine what the reasons are. Analysis begins with examining the frequency of screen failures and the reasons for them. Real-time access to performance metrics of the individual sites allows early identification of effective and ineffective recruitment activities for the particular study and its population. The lessons, both good and bad, should be shared immediately with all sites to enhance recruitment efforts studywide.

Detailed performance metrics on recruitment can also quickly determine whether any specific inclusion/exclusion criteria are disproportionately hindering recruitment of the desired study population. If this is the case, it is imperative to find out quickly. It may be appropriate to consider different recruitment strategies to obtain access to a more suitable population. Failing that, possible adjustments in inclusion/exclusion criteria may require consideration. Understanding the options depends on having timely, accurate enrollment data from investigative sites. The ability to manage enrollment effectively is impaired in direct proportion to the degree to which the availability of such data is lacking or delayed.

### 3.2 Site Performance

Much higher levels of efficiency are possible when study managers and monitors can track site performance continuously and in sufficient detail to identify and address problems. Most studies currently lack the ability to track such metrics. As a consequence study management becomes a passive affair left to the vagaries of the site monitor. Because monitors often lack training and experience managing

sites, there may be no one at all providing effective management. Furthermore, depending on any individual as an information and management filter invariably introduces subjectivity and uncertainty, no matter how well trained that individual may be.

Query rates and query response times on submitted data are easily measured and provide strong indications of site performance in general. Query metrics are important on several levels. Most importantly, the existence of numerous queries late in a study delays database lock and all downstream events, including analyses, progression, and submissions. In addition, instituting rapid data entry, validation, and return of queries produces faster, more meaningful feedback to sites on their performance, enabling each site to reduce errors, often quite significantly. By contrast, systems that rely on paper and pen, with double key data entry, generally take a month or more—often much more—to return queries to sites about data previously submitted. During the interval, each site, unaware of the problems in the data it collected, will continue to make the same errors, increasing the volume of errors and the time required to correct them and inevitably delaying database lock.

Key site performance metrics that should be continuously tracked include patients screened, screen failures and reasons, good clinical practice compliance issues, query rates, query resolution, the number and age of outstanding queries, and specific case report form (CRF) fields, forms, and validation ranges that are generating the most queries at each site. Number of adverse events, both serious and nonserious, should also be tracked.

Continuous attention to site performance indices has the additional benefit of allowing technology-enabled distributed management with reduced requirements for a monitor to go to each site to determine how things are going. Rather, the generation and tracking of such parameters enables all members of the study team to examine performance metrics. The metrics enable each member of the team to focus attention where it is most needed. The availability of detailed metrics also opens the door to performance-based site management, allowing study management to institute incentives for strong performance and disincentives for poor performance. Performance-based management stands in stark contrast with the present approach, with each site knowing it will be paid the same amount for each patient regardless of performance. Although this is currently the standard approach to "management" of investigational sites, it hardly deserves the name.

### 3.3 Adaptive Site Monitoring

Monitoring is one of the most expensive components of clinical research, typically accounting for one-third of study costs. Yet the objective of monitoring, while important, is fairly modest: To ensure that the data in the database is accurate. The high costs of monitoring investigational sites are often assumed to be inevitable sacrifices to the cause of ensuring data accuracy. An adaptive approach, however, can use the stream of timely information from the field to allow many functions now performed during site visits to be handled centrally, enabling far more rapid and standardized assessment of site performance based on performance metrics while also reducing costs.

The highest levels of monitoring efficiency are made possible by use of electronic data capture methods, such as the digital pen, that can serve as source documenta-

tion. Source document verification itself accounts for approximately 80% of monitoring time. The digital pen and similar devices, detailed in Section 4, are used by the clinician to write in a familiar way on a form laid out with a grid that enables software to match data entries to appropriate fields. Captured data is stored in the digital pen and transmitted to a central database, where the pen is inserted in a dock attached to a computer. With this approach, the digital data is not a transcription of the source data; it is the source data itself. It is unnecessary for the monitor to compare electronic data to an original piece of paper. Data first recorded in electronic form, including data values and an image of each completed form (reviewed if necessary to resolve any ambiguities), is checked and corrected promptly through queries to site personnel after both automated and human review. Use of EDC technology capable of serving as source data provides great cost savings in the source data verification process by eliminating the time and expense of comparing electronic data to stacks of printed material.

Much of the work that monitors currently do can be done much better and more accurately by systems that harness technology. Data capture that obviates the need for traditional source data verification is one example. Effective use of such technology in conjunction with adaptive techniques can reduce monitoring costs by two-thirds or more while enabling far faster and more effective verification of data accuracy.

Adaptive approaches to site monitoring reflect a fundamental change in the way studies are conducted. Increasingly, the tedious, repetitive manual tasks such as comparing two values on different pieces of paper will be replaced by electronic tools and automated processes. The evolution to such processes will be further improved with the adoption of electronic medical records, ultimately resulting in a monitoring process requiring no manual steps. As a consequence, the monitoring function will be changed from that of the box checker of today into that of a manager and consultant whose main job will be to assure site performance. The monitor will be able to focus on optimizing each site's performance against trial objectives rather than constantly comparing the content of one database field to the content of the same field on a paper CRF. Monitors can spend less time on minutiae and more time observing, analyzing, and reacting to study trends. The job of the monitor will increasingly shift to anticipating and addressing issues before they can develop into problems that require extensive cleanup.

Performance metrics can help monitors identify problems at a sufficiently specific level to suggest both causes and potential solutions. The number of visits needed at sites can be determined based on each site's performance metrics rather than equal numbers of visits at similar intervals for all sites. Sites with more and bigger problems can receive more site visits; sites with outstanding performance, fewer. For example, with the tracking of the number of unmonitored fields, the accumulation of a sufficient number can trigger a site visit. But while a good monitor might verify between 500 and 1000 fields per day, depending on the study, a less experienced monitor might do half that. The less experienced monitor could schedule a visit when, say 500 fields accumulate, while a more effective monitor could wait longer. Using an arbitrary, uniform schedule, the less experienced monitor might have a difficult time, while a more experienced monitor would waste time at the site and traveling.

## 4 IMPLEMENTATION ISSUES

An adaptive development program requires a different approach to study design, data capture, job functions, business processes, and planning. When first moving into adaptive research, the simplest approach is to start with baby steps while refining processes and increasing capabilities. Phase II adaptive dose finding and sample size reestimation are generally the most straightforward types of adaptive trials to undertake.

### 4.1 Data Capture and Validation

Adaptive research is critically dependent on the timely availability of accurate information. Information to effectively manage any given study may include specific decision criteria related to strategic adaptations, but it always involves continuous assessment of operational performance measures. Careful attention to operational issues is a prerequisite for achieving strategic adaptive goals. For example, slow availability of outcome information will delay and possibly cripple efforts at adaptive dose finding, and slow recruitment inevitably undermines any strategic adaptive decisions because a shortage of patients guarantees a shortage of data. Thus, the trial will be inefficient no matter how sophisticated or sound the planned adaptations.

Efficient data capture is a prerequisite for adaptive studies. In contrast to Web-based EDC and its attendant delays for data reentry, two newer forms of data capture have demonstrated superiority for adaptive studies. The first is an electronic pen incorporating an optical sensor that records keystrokes as data are entered on paper CRFs. After completion of CRFs, the pen is docked and the data transmitted to a central location, where keystrokes are converted to numbers and letters and the data recorded in the study database, along with an exact electronic image of each CRF, available for review by data management personnel in the case of ambiguous or missing data. Together with complementary systems that enable automated validation, the electronic pen provides the capability of having information from a patient visit—summarized and interpreted information, as opposed to raw data—on a sponsor's desktop before a patient even leaves the investigational site.

The second new form of data capture consists of fax-back systems. These offer another way to allow the simplicity of pen-and-paper data entry coupled with efficient software to read incoming forms. Fax-back tools, first introduced more than a decade ago, have improved markedly in recent years, adding the ability to read data electronically and store both the data and an image of each completed form in the study database.

In practice, digital pen systems have proven to be the most accurate and timely data-capture method, with query rates of approximately 1% (1 query issued per 100 fields) compared with approximately 5–10% for fax-back systems. Both methods are significant improvements over older pen-and-paper, double-key manual data entry, which often has query rates on the order of 20%. However, the most important determinant of query rates is not the data-capture method itself but the "middleware"—the software and/or systems that bring data into the trial system, validate it, and return queries. The main reason for the superior efficiency of the electronic pen is its linkage with the sophisticated middleware that enables queries to be returned

to investigational sites within minutes of their submitting data. The middleware not only markedly reduces the feedback time but also enables tracking of query rates by interviewer, question, and a multitude of other variables that help identify sources of error so that they can be immediately corrected. This is a boon for adaptive research since the ability to capture data, analyze it, and rapidly respond to performance metrics such as query rates is the essence of an adaptive approach.

A data-capture device such as the digital pen allows the clinician an intuitive, familiar means of recording and transmitting data to the study database, where it can be transformed electronically to information meaningful for individual study roles through "widgets" (Fig. 5) or the Web. And, as previously noted, if the CRF can be used as source data, the work of monitors is greatly diminished, allowing them to focus on management rather than rote tasks.

Advanced data management systems also incorporate sophisticated procedures that automate validation and query generation. These go well beyond the simple checks such as data ranges to incorporate algorithms that consider trends, consistency across visits as well as within a visit, and head off future problems by identifying studywide weaknesses as well as individual site or interviewer problems. These processes similarly help assure rapid availability of clean data as well as providing rapid feedback and management assistance that assure better site performance in the future.

## 4.2 Planning

Since potential adaptations and related decision criteria must be specified in advance, the effort required to plan an adaptive study is greater than for a conventional trial.

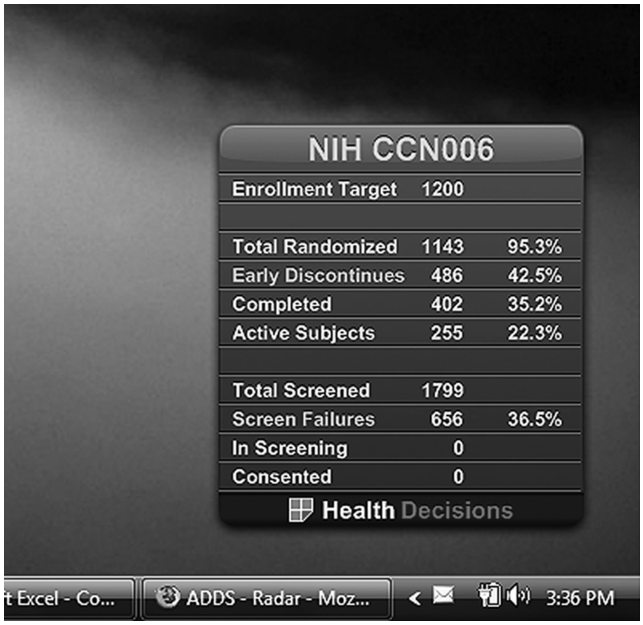


***FIGURE 5*** *Desktop "widget" that provides immediate project status, updated in real time. (Figure copyright 2006, Health Decisions, Inc. Used by permission.)*

Protocols employing adaptive techniques must specify through predetermined decision rules precisely what will occur in different circumstances and with different values emerging in trial data. This approach is a requisite element to ensure that scientific integrity is not compromised, knowingly or otherwise. Simulation during planning can effectively be used to explore a variety of "what if" scenarios to determine sample size adequate to control variance and other elements affecting statistical power. The simulations are also useful in analyzing decision rules for planned adaptations and operational issues such as ensuring adequate supplies in the event of different adaptive changes.

Planning of adaptive trials must include specific arrangements to exclude bias. This includes measures to eliminate the possibility of unauthorized parties inferring from adaptations actually made in the course of a trial the data that brought about the adaptations, effectively unblinding the study. Such back solving of study data can be prevented by measures such as limiting access to information about the specific rules governing planned adaptations and defining decision criteria in terms of ranges rather than precise values. The latitude allowed by regulatory authorities for blinding and other study elements are, however, substantially greater during early testing than during confirmatory studies.

When unblinded assessments are required (such as sample size reestimations or, in some instances, pruning treatment arms), a firewall must be in place to ensure that those running the study do not gain access to information that may affect how the study is conducted. For example, an adaptive trial protocol may require independent (internal or external, focusing on safety and/or efficacy) analyses that require access to unblinded data. Decisions may also involve a broader range of business issues such as whether to discontinue a study. While such decisions may be delicate, it is critical that safeguards be incorporated to prevent contamination of the ongoing study; this is equally true whether required by regulatory stipulations (confirmatory phase) or not (learning phase), as compromised data creates opacity in a situation that demands clarity. These situations are normally handled by compartmentalized groups who are not involved in the study itself, often by a steering committee that may have other groups (such as safety) reporting to it. The ultimate power to decide, subject to the specified decision criteria, whether to implement adaptive changes often rests with the steering committee.

Simulation can also be used not only to explore the likelihood of different scenarios that affect study design but also for optimizing business processes and defining job functions to support the development program's objectives. Although the effort required is greater, the thorough analysis of possible future scenarios provides an invaluable opportunity to moderate or eliminate potential choke points. Over time, the longer view and greater flexibility that are hallmarks of adaptive research can substantially improve allocation of a development program's financial resources and accelerate attainment of the program's goals.

### 4.3 Process Optimization

Despite the current variability in the conduct of clinical trials, almost all trials require repeatedly performing the same essential processes. These processes can be defined much more clearly and precisely than is customarily done. The analysis required to define a process precisely will often suggest ways to improve the process.

It is process improvement in combination with new technology that brings huge rewards. The gains from inserting expensive new technology into an inefficient process are minimal.

Optimizing processes for adaptive research requires minimizing the individual variations in trial processes, including variations among the approaches taken by different project managers. The organization of the trial should be implemented in a system rather than according to each project manager's idiosyncrasies. The test of whether the project manager's role is defined precisely enough is whether it would be easy to plug in a replacement.

Furthermore, optimizing a study or program based on analysis of data as it is collected is not just a matter of inserting more frequent interim analyses into existing processes. The trial processes must provide a continuous flow of up-to-date, clean data. To the extent that current processes render data unavailable—whether because it has not been entered, validated, passed on from person A to person B, or made accessible in a useful form—study managers may be prevented from making key decisions or deceived by incomplete data into making decisions that turn out to be erroneous.

### 4.4 Decision Making

In adaptive trials, learning and decision making no longer wait until the very end, with the sudden transformation from having no data and knowledge to having a deluge of data and the need to extract knowledge all at once. Progress must be considered weekly, emerging trends identified continuously, and decisions taken as necessary to optimize the trial.

The point of gaining a greater understanding of data at an earlier time is to inform decisions to optimize the course of a trial. Greater efficiency in an adaptive study is the result of active managers making timely decisions directing specific changes. Having information earlier achieves little if managers fail to understand it and, where appropriate, to take action.

### 4.5 Adaptive Data Monitoring

The challenges of adaptive data monitoring require careful attention in implementing strategic adaptations. Besides attention to the exclusion of bias through organizational arrangements such as independent data management committees and independent statistical services, there must also be careful attention to arrangements for timely, accurate analysis of data to facilitate all potential adaptive decisions. Statistical methods must be selected and approved, and analytical procedures and tools created and tested to avoid delay or confusion in the implementation of predefined decision rules [8].

### 4.6 Regulatory Considerations

Regulators have been receptive to proposed methods of conducting clinical trials more efficiently while preserving validity and integrity. Understandably, they have not granted blanket approval for clinical researchers to try all new methods for strategic adaptations. However, they have shown a willingness to consider adaptive

techniques. Regulators have also expressed greater receptiveness to certain adaptive techniques and have indicated some general preferences for the use of those techniques. As detailed guidance for conducting adaptive studies is awaited, it is possible to provide a brief summary of regulatory preferences.

1. Two of the easiest adaptive components to apply are sample size reestimation and pruning in dose-finding studies. Regulatory groups allow sponsors a great deal of latitude as to how they conduct early studies, and the issue of defining doses suitable for final (pivotal) testing is basically at the sponsor's own risk. Sample size reestimation can be used in this early stage as well as more powerfully in late (pivotal) testing; the benefits of sample size reestimation in the early phases are sufficiently powerful as to merit serious consideration for all such studies.

2. All possible adaptations in a trial must be specified in advance, as well as the criteria for deciding whether to carry out the adaptations. Adaptive studies should never be thought of as providing license to make any changes that happen to occur to trial managers along the way; it would be a grievous mistake to think that adaptive techniques provide a less stringent substitute for well-planned, carefully executed studies, or that regulators regard adaptive techniques so indulgently.

3. Simulations are often helpful in understanding possible outcomes and are of interest to regulators. Simulations allow assigning probabilities to different potential developments in the trial and analyzing the repercussions of each development. Newer software and powerful computers facilitate exploration of scenarios, often forcing consideration of study variables that might otherwise go unattended. Simulations increase understanding of the effects of specific adaptations on the conduct and results of the trial. Both the results of simulations and the software used to conduct simulations for trial planning should be submitted to regulators. It is advisable to request a special protocol assessment to review an adaptive design and, at that time, to provide simulations demonstrating that the type 1 error is preserved.

4. If decisions on adaptations require considering unblinded data, regulators may prefer that the data is analyzed and reviewed by independent statisticians and an independent data management committee rather than by the sponsor. It is important, however, to recognize and limit the power of outside bodies that may fail to fully appreciate business consequences of managerial decisions. Not surprisingly, academic groups are often less attuned to the business consequences of decisions made about clinical studies.

5. Criteria for decisions on adaptations should be specified and managed in such a way as to minimize the possibility of unauthorized parties inferring important blinded information from the specific adaptations that are carried out. For example, if sample size is to be readjusted based on the magnitude of the treatment effect, knowledge of the sample size chosen might enable unauthorized parties to back solve for the treatment effect. Such concerns can be allayed in at least two ways. First, the specific technique and criteria that will be used for sample size reestimation can be kept confidential. For example, there should be restricted access to knowledge about the type of sample size reestimation that is planned—whether reestimation will be based on the size of the treatment effect alone, or the size of the placebo effect, or the variability observed in trial data, or based on multiple parameters. Second, the decision criteria themselves should be specified in ranges;

thus, when the sample size is readjusted to a specific value, the most that can be inferred by those who learn of the adjustment is that a parameter fell somewhere within a range.

6. Regulators will want detailed information about the measures taken to ensure the exclusion of bias. There must not only be a "firewall" between those who are allowed access to specific information and those who are not, but the firewall must be one that regulators agree is likely to be effective.

## 5 PROMISE OF ADAPTIVE METHODS

Adaptive techniques represent a major advance in pharmaceutical development made possible by advances in communications and technology. In many aspects, adaptive techniques represent a step into the type of modern management that is the norm in other large industries. Modernizing trial management requires evolutionary changes built on time-honored principles—but also requires greater flexibility in an industry known for its conservatism and "not invented here" attitude. Adaptive methods have been developed over time, based on solid theoretical foundations, specifically to address many of the limitations and inefficiencies of the current approach to clinical research.

The changes required for the shift to adaptive methods are profound and extend beyond technology to encompass work processes and the functioning of individuals performing different study roles. Technological tools and related processes are meant to provide a stream of current information that sheds light on study progress and bottlenecks. The new technologies and processes do not replace managers. Rather, they enable managers truly to manage, by providing a wealth of current status information and communications that drive effective functioning studywide.

Committing to an adaptive development program offers the potential for great rewards. Because of these high rewards, adaptive research is often assumed to be accompanied by high risk. However, a properly conceived and managed adaptive program actually reduces risk. This is by design—reducing risk is a major objective of adaptive techniques. Indeed, utilizing adaptive techniques entails no risk of conducting less efficient studies—the worst that could happen in an adaptive trial is that no adaptive changes would be made, leaving study performance at the accustomed level. On the other hand, every adaptive change that is made saves time, reduces costs, improves the amount and quality of information produced, or provides some combination of these benefits. Indeed, the greater risk lies in *not* utilizing adaptive techniques. Entire programs are condemned to suboptimal decision making and the inevitable consequences: unnecessary expenditures, avoidable delays, and windows of opportunity slammed shut before new drugs can make their way through the process of clinical evaluation.

The existing conventional approach to clinical studies has spawned many horror stories that reflect the risks of the current approach to clinical research. Every experienced pharmaceutical researcher has a horror story to tell. Virtually all such stories include an account of a small initial problem that grows to disturbing proportions before it is recognized and addressed. Adaptive processes can reduce the instance of such horror stories and the risks they embody by providing earlier, better indicators of incipient problems.

The potential of adaptive research to improve operations industrywide, for companies large and small, is simply too great to ignore. The potential gains in efficiency—the savings in both time and money—are dramatic. Adaptive tools and techniques are here and ready. They address many of the specific drawbacks and weaknesses of traditional methods. With concentrated attention by the industry and regulators, adaptive methodology can make important contributions to a new and more productive era in clinical research.

## REFERENCES


1. Doll, R. (1998), Controlled trials: The 1948 watershed, *BMJ*, 317, 1217–1220.
2. Chow, S.-C., and Chang, M. (2007), *Adaptive Design Methods in Clinical Trials*, Chapman & Hall/CRC, Boca Raton, FL, pp. 47–67, 55–59, 58–60, 171.
3. Hu, F., and Rosenberger, W. (2006), *The Theory of Response-Adaptive Randomization in Clinical Trials*, J Wiley, Hoboken, NJ, pp. 4, 6–7.
4. Berry, D. A., Mueller, P., Grieve, A. P., Smith, M., Parke, T., Blazek, R., Mitchard, N., and Krams, M. (2000), Adaptive Bayesian designs for dose-ranging drug trials, in Gatsonis, C., Kass, R. E., Carlin, B., Carriquiry, A., Gelman, A., Verdinelli, I., and West, M., Eds., Case Studies in Bayesian Statistics V, Springer, Verlag, New York, pp. 99–181.
5. Jennison, C., and Turnbull, B. (2006), Efficient group sequential designs when there are several effect sizes under consideration, *Statist. Med.*, 35(6), 917–932.
6. Chuang-Stein, C., Anderson, K., Gallo, P., and Collins, S. (2006), Sample size reestimation: A review and recommendations, *Drug Info. J.*, 40(4), 475–484.
7. Schoenberger, C. (2000), An Alzheimer's drug goes on trial, *Forbes Mag.*, March 20, pp. 94–96.
8. Abrams, K., Myles, J., and Spiegelhalter, D. (2004), *Bayesian Approaches to Clinical Trials and Health-Care Evaluation*, Wiley, Chichester, UK, pp. 202–224.

# *Organization and Planning*

**Sheila Sprague[1] and Mohit Bhandari[2]**

*[1]Department of Clinical Epidemiology & Biostatistics and [2]Division of Orthopaedic Surgery, Department of Surgery, McMaster University, Hamilton, Ontario*

**Contents**

1 Protocol
- 1.1 Protocol Format
- 1.2 Protocol Amendment Procedure
- 1.3 Prestudy Requirements

2 Finance
- 2.1 Reviewing an Offer to Participate in Clinical Trial
- 2.2 Budget Considerations
- 2.3 Investigator Agreements
- 2.4 Sponsor Interactions
- 2.5 Regulatory Documentation (Essential Documents)

3 Patient Selection
- 3.1 Research Ethics Board Approval (Institutional Review Board Approval)
- 3.2 Feasibility
- 3.3 Sample Size
- 3.4 Recruitment Methods
- 3.5 Screening
- 3.6 Informed Consent
- 3.7 Patient Confidentiality
- 3.8 Intention to Treat Analysis

4 Treatment Schedules
- 4.1 Supplies
- 4.2 Adherence to Protocol
- 4.3 Follow-up
- 4.4 Monitoring Visits
- 4.5 Composite Endpoints

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

5 Patient Response Evaluation
5.1 Outcome Measures
5.2 Case Report Forms
5.3 Adverse Event Reporting
5.4 Data Queries
5.5 Adjudication
6 Design Considerations
6.1 Study Design (Observational Studies and Randomized Controlled Trials)
6.2 Trial Organization and Responsibilities
6.3 Data Management
6.4 Quality Control\
6.5 Study Close-out Activities
Bibliography

## 1 PROTOCOL

The study investigator is responsible for the review of all protocols and for the assessment of the feasibility of the clinical site to conduct the clinical research study. The investigator is also responsible for reviewing the protocol and providing input to the design so that the protocol can be properly conducted at the investigator's clinical site. The investigator is also responsible for ensuring that the research ethics board or the institutional review board approval is obtained before initiating the protocol.

Each sponsor's protocol received by a clinical site should describe the purpose, the objectives, the study design, the methods, and the planned analysis of a clinical study. The protocol must be sufficiently detailed so that the study can be conducted accurately and properly. Since the protocol often outlines the ethical, clinical, and regulatory responsibilities of both parties (the sponsor and the clinical site), it usually acts as a contract between the investigator and the sponsor.

Each protocol must be reviewed by the investigators, research coordinators, and other relevant clinical and research associates to ensure that the following takes place: (1) The study is safe; (2) the study is based on good medical science; (3) the study is ethically acceptable; (4) the objective of the study is clear and the study design, sample size, procedures, and statistical analysis will enable the study objectives to be met; (5) the protocol satisfies the clinical sites standard operating procedures (SOPs) and the good clinical practice (GCP) guidelines; (6) the protocol is financially and practically feasible for the clinical site and the investigator; (7) the investigator and clinical site have a sufficient number of potential study participants to meet the protocol enrolment goals; (8) the investigator has the sufficient staff to successfully conduct the trial; and (9) the investigator has sufficient resources and equipment to successfully complete the study.

### 1.1 Protocol Format

We will describe several protocol formats that are commonly used by industry-sponsored trials and by investigator-initiated trials (non-industry-funded trials).

***Industry-Sponsored Trials*** The first format is the one that our institution (McMaster University/Hamilton Health Sciences) recommends in their standard operating procedures (SOPs).

*Cover Page and Table of Contents* The protocol should begin with a cover page that outlines the status of the trial, the revision date and number, the trial's full title, the study number, and the sponsor. There should also be space for signatures and dates signed for the author, the sponsor approval, and the investigator approval, as the protocol often acts as a contract between the investigator and the sponsor. The sponsor contact names and telephone numbers should also be listed to respond to investigator queries or any adverse events. A detailed table of contents, including a list of all appendices, should follow the cover page.

*Introduction* The introduction contains a brief summary of the relevant background information on the study design and the protocol methodology. Sufficient background information should include the development of the study agent or device, a description of the disease process to be studied, and the current medical treatments available. The introduction should provide enough detail for the readers to clearly understand the rationale for the study. Relevant information regarding pharmacological and toxicological properties of the test article and previous efficacy and safety experience should be included.

*Study Objectives* The objectives of the study should be clearly stated, and there should be a description on how the objectives are related to the design of the study. Both primary and secondary objectives should be included.

A brief overview of the study design, indicating how the study objectives will be met, should be provided. This part of the protocol includes a description of the study type (i.e., randomized controlled trial, prospective cohort study, phase III, multicenter, double blind, placebo controlled), details on the specific treatment groups, information on the sample size in each group required for the total trial and at each clinical site, how the study participant identification numbers are assigned, and the type, sequence, and duration of the study periods. A brief description of the methods and procedures to be used during the study should also be included. There should also be a discussion on the study design that details the rationale for selecting the study design. Critical decision points and any atypical features of the study design should be described. The treatment schedule, the dosages for the study agents, and the follow-up schedule and follow-up requirements should be discussed.

*Study Participant Population* The next section of the protocol describes the study participant population. The study's inclusion criteria (the criteria each study participant must satisfy to participate in the study) are described. These criteria can include the following: age, sex, race, diagnosis, diagnostic test result requirements, concomitant medication requirements, severity of symptoms and signs of disease, and the ability and willingness to perform study requirements and to provide informed consent. The criteria must be sufficiently detailed to provide the investigative site with the information needed to recruit eligible study participants and to ensure a homogenous study population. The exclusion criteria describes the items that would eliminate a potential study participant from being enrolled in a study. Exclusion criteria can include, but are not limited to, the following: previous medical history, pregnancy, childbearing potential, current or past therapy, severity of disease, current medical conditions, a minimum time since the last clinical study, and substance abuse problems. There should be a discussion detailing the rationale for the

inclusion and exclusion criterion to obtain the homogenous study population required.

*Concomitant Medications* The listing of the required therapies including the dose, frequency, and duration should be detailed in this section. Required therapies are medications that are required in addition to the test treatment. The investigator must also determine whether the sponsor or the site is providing these medications and determine the accountability procedures required by the sponsor. There must also be a list of allowed medications and disallowed medications. Allowed medications are the medications that are permitted during the study, the dosage, the frequency, and the conditions under which they are permitted. Disallowed medications are medications that the study participant is not permitted to use during parts of or the entire study. The study participant and investigator are often told to document the allowed medications and the disallowed medications (if there is a protocol deviation) on the patient diary cards and on the case report forms.

*Study Plan and Methods* The study plan and the methods portion of the protocol details the plan of action, all procedures to be followed, and the methods to be used throughout the study. The activities for each follow-up visit are clearly outlined, as well as any specific criteria that must be met before the study participant moves to the next stage of the study. For difficult testing methodologies, a separate section may be included to describe in detail the methods of the testing. If the study participant must have specific results from this testing to enter, or to continue in the study, these criteria are described. Types of study activities that would be included in this section include: medical history, type of physical examination, blood or urine testing, electrocardiograms (ECGs), diagnostic testing such as pulmonary function tests, symptom measurement, dispensing and retrieval of medication, diary card exchange, study subject number assignment, adverse event review, and the like. Each visit should be detailed in separate sections in the protocol, such as 1 Visit 1, 2 Visit 2, and so forth. The investigator should ensure that this section provides clear methods, procedures, and timings of activities.

*Study Medication Supplies* The study medication supplies portion of the protocol describes the test treatments including placebo with formulation and dosage information. It also details the packaging, dispensing, retrieval, and accounting of these supplies. All ingredients and the formulation of the investigational test article and any placebos that are used in the study are described. The precise dosing required during the study is discussed. The method of packaging, labeling, and blinding is described. The method of assigning treatments to study subjects and the test article and study subject numbering system is detailed. If applicable, the method of blinding to make the test treatments indistinguishable should be explained. The method of test article coding, code storage, and code access is described. If a third party such as a pharmacist will be dispensing or blinding the test article, then specific instructions on how and when the blinding will occur will be detailed in this section.

The coding and study subject test article assignments for the study are detailed. Generally, all study staff and study participants are blinded to the identity of the test treatments. This section describes when a study subject treatment code assignment is given and if, and when, it will be changed. A description of where the test

article randomization code will be kept by both the investigator and the sponsor or delegate is required and the procedure for breaking this code, if necessary, should also be included. There should also be a section describing the study subject kit packaging, which should detail the administration and dosage of study medication, as well as how the elements of the kit are distributed. Generally, the medication is labeled visit 1, visit 2, and so forth for easier dispensing. Any specific instructions for the administration of the test article treatments, such as label removal, test spray, or how and when the test article is to be returned should be detailed in this section. Explicit details regarding the dosage of the test treatment are explained. Instructions for broken test article applicators or lost test articles are also described. If extra treatments are provided for emergencies, the circumstances regarding their use should be described. Explicit instructions for the storage of the test treatments at the investigative site are detailed, as well as if special conditions are required, such as refrigeration. Federal regulations and the International Conference on Harmonisation (ICH) guidelines require a complete accounting of all test materials received, disbursed, and returned.

*Discontinued-Study Participants* The research protocol should outline how investigators will handle study participant withdrawal, dropouts, or other discontinued participant subjects. It also details how to handle these situations within the study context and whether these study subjects are to be replaced.

*Adverse Events* Investigators should include a listing of expected adverse events from previous clinical studies with the test article. They may also include the policies and procedures for the reviewing and the reporting of adverse events that occur during the clinical study.

*Concomitant Illness* The protocol should decribed the procedure for documentation of any concomitant illness, which a study participant develops during the study. A concomitant illness is an illness, disorder, or any other medical condition that is not considered a consequence of the study medication being administered.

*Variables and Evaluations* The investigators should describe how clinical study efficacy and safety variables will be evaluated. The studies primary and secondary variables will be identified and discussed. Descriptions of evaluations are given here including how they are recorded, measured, or calculated.

*Statistical Analysis* It is important to provide details on how the study results will be analyzed and reported. Specifically, sample size determination, objectives or hypotheses being tested, levels of significance, statistical tests being used, and the methods used for missing data. The method of evaluation of the data for treatment failures, noncompliance, and study participant withdrawals is presented. If an interim analysis will be performed, the rationale and conditions are clearly described. A description of how the safety data and adverse events will be tabulated and evaluated is included.

*Informed Consent* This section of the protocol describes the procedures and responsibilities for obtaining informed consent for the study.

*Study Monitoring* A description of study monitoring policies and procedures, including the right for the sponsor or method center's representatives, federal government, and/or other regulatory authorities to verity and audit the study data, is presented in this section.

*Deviations from Protocol* The method for handling protocol deviations is described.

*Discontinuation of Study* This section describes the conditions under which the study would be discontinued.

*Confidentiality of Data* The importance of confidentiality of data and the fact that the data is owned by the sponsor or by the coordinating methods center are detailed.

*Publication of Study Data* The conditions for publication of study data are outlined.

*Research Ethics Board Approval* This section states that the investigator is required to obtain research ethics board (REB) or institutional review board (IRB) approval for the protocol, informed consent form, and any protocol amendments.

*Investigator's Statement* An outline of the investigator's responsibilities and agreements are included in the protocol.

***Investigator-Initiated Trials*** For investigator-initiated randomized controlled trials, the Canadian Institute of Health Research requires the following format.

*Need for Trial* The problem to be addressed and the principal research questions to be addressed are listed. There should also be a strong justification as to why a trial is needed now. Evidence from the literature, professional and consumer consensus, and pilot studies should be cited if available. References to any relevant systematic reviews should be provided and the need for the trial in the light of these reviews should be discussed. If you believe that no relevant previous trials have been done, provide details of the search strategy used for existing trials. A description of how the results of this trial will be used should be provided.

*Proposed Trial* This section of the protocol begins by stating the proposed trial design, including whether the trial is open-label, double or single blinded, and so forth. The planned trial interventions for both the experimental and control groups and the proposed practical arrangements for allocating participants to trial groups are discussed. For example, the randomization method and if stratification or minimization are to be used provide justification for using either methodology. Factors that will be stratified or minimized should be listed. The proposed methods for protecting against other sources of bias include blinding or masking. If blinding is not possible, please explain why and give details of alternative methods proposed or implications for interpretation of the trials results. The planned inclusion/exclusion criteria should be listed.

There should be a justification for the proposed duration of treatment period and the proposed frequency and duration of follow-up. The rationale for the proposed primary and secondary outcome measures and how the outcome measures will be measured at follow-up should be provided.

For the proposed sample size, include both the sample size for the control and treatment groups, a brief description of the power calculations detailing the outcome measures on which these have been based, and provide event rates, means, and medians, as appropriate. Provide a justification for the size of difference that the trial is powered to detect and mention whether the sample size calculation takes into account the anticipated rates of noncompliance and loss to follow-up.

A description of the planned recruitment rate, how recruitment will be organized, over what time period recruitment will take place, and what evidence there is that the planned recruitment rate is achievable should follow. There should be a discussion on any likely problems with compliance and loss to follow-up; the evidence that the compliance and loss to follow-up figures are based on should also be included.

The details of the planned analyses are described, including any planned subgroup analyses and the proposed frequency of analyses (including any interim analyses).

The investigators should specify whether the trial addresses any economic issues. This is not a requirement; however, it is important to justify the inclusion/exclusion of any health economic studies and give details of any study proposed.

The final section of the protocol on the proposed trial should provide an accurate budget, budget justification, and the length of the trial.

*Details of Study Team* A description of the study team must be provided that explains the overall trial management, including the role of each applicant proposed, the steering committee, the methods center, and whether a data safety and monitoring committee will be established and its composition. The role of international collaboration should be discussed, including the nature of and the need for any international collaboration. The proposed participating centers should be listed along with their experience with previous trials and estimated recruitment rates.

### 1.2 Protocol Amendment Procedure

Protocol amendments can be suggested by either the sponsor or the investigator and can be made after the protocol has been finalized. A draft protocol with the proposed amendments is prepared and reviewed internally and by the investigators before it is approved. The investigator site must obtain REB or IRB approval of the protocol amendment before it is implemented.

### 1.3 Prestudy Requirements

Before a study can be initiated at a clinical site, the following activities must be completed:

- Receive final protocol from sponsor or methods center.
- Receive amendments from sponsor or methods center.
- Finalize budget with sponsor or methods center.
- Distribute protocol amendments to relevant study team members.
- Receive investigator's brochure from sponsor or methods center.
- Read investigator's brochure.
- Complete qualified investigator undertaking, FDA 1572A, or investigator agreement, as required.
- Prepare informed consent form.
- Submit documentation to the REB or IRB.
- Receive REB approval and Health Canada REB application form (if applicable).
- Receive case report form (CRF) books from sponsor or methods center.
- Design source documents, as necessary.
- Send all required documentation to sponsor or methods center.
- Receive clinical supplies from sponsor or methods center.
- Inventory supplies and return documentation of receipt to sponsor or methods center.
- Plan subject recruitment strategy.
- Prepare regulatory documentation binder/file.
- File regulatory documentation received to date.
- Set up necessary local resource utilization for study (e.g., pharmacy, laboratory, etc.).
- Set up a contract with the sponsor or methods center.
- Conduct in-services with each involved department.

The investigator or delegate must send the following documents to the sponsor or methods center before the study can begin:

- Final signed protocol.
- Final signed amendments.
- Signed Qualified Investigator Undertaking Form, FDA form 1572, or investigator agreement.
- Current curriculum vitae (CVs) and medical licenses of principal investigator and subinvestigators.
- Signed financial disclosure document.
- Budget approval documentation.
- REB or IRB approval letter(s) approving protocol, consent form, advertisements, and any other relevant documents.
- Copy of REB or IRB approved consent form.
- Copy of current REB or IRB membership list or letter from REB or IRB.
- Copy of laboratory license, normal ranges, and CV of director, if required.

Finally, before each study is initiated, appropriate research staff members are informed of the requirements of the study and their role and responsibilities regarding the conduct of the study.

## 2 FINANCE

### 2.1 Reviewing an Offer to Participate in Clinical Trial

Participating in clinical trials requires a substantial commitment of both time and effort, and participation often continues for months or even years. There are usually clear financial incentives to participating in clinical trials; however, other incentives include a chance to collaborate with other clinical investigators and opportunities to improve knowledge about the disease and treatment being investigated. Another advantage, which may be offered in some trials, is the exposure to new investigative techniques or access to special equipment or facilities.

The scientific, practical, and financial implications need to be considered before agreeing to participate in a clinical trial. The first item to assess is the study question and the study methodology. It is important to ensure that it is a relevant question and that the study methodology is sound and will meet the goals of the trial. The eligibility criteria must be carefully assessed to ensure that your clinical site has sufficient eligible patients to be recruited for a clinical study. It is often necessary to perform a survey of potentially eligible patients over a 4-week period to provide an accurate estimate of recruitment rates.

It is important to assess the impact of the trial on your patients, as your primary obligation is to protect the welfare your patients. It is important to consider if the patients will be required to have any investigations or procedures that are not part of standard care, and whether these will be painful and will possibly put the patients at risk. You need to estimate how much time the patient will devote to the study and how much, if any, compensation the patient may receive.

The time required for the overall coordination, the patient enrolment and follow-up, and the data collection also needs to be carefully considered. It is best to have an experienced clinical research coordinator carefully review the protocol and case report forms to provide an accurate estimate on the time and resources required to successfully participate in a clinical trial. Tasks often take longer than anticipated, so it is often good to add in a little extra time to account for the unanticipated. It is also important to know whether the study will be published, who will write the manuscripts, and what the authorship policies are. After carefully weighing these items, you can make an informed decision to participate in a clinical trial.

### 2.2 Budget Considerations

The investigator should consider all potential costs and prepare an accurate budget for running the clinical trial at his or her clinical site. These costs include: (1) administrative assistant time, (2) research associate/nurse time, (3) supplies, (4) expenses incurred by the patient, (5) extra medical and hospital costs (i.e., pharmacy, radiology, laboratory tests), and (6) physician time. It is also important to include

departmental and institutional overhead costs in the budget and to have your institution's grants and contracts office, as well as the REB or IRB, review your budget.

### 2.3 Investigator Agreements

The investigator is responsible for the review or preparation and approval of the contracts with the sponsor, and the investigators institution's legal advisor or delegate is responsible for legal review of these contracts. When the sponsor provides the contact, the investigator and the institution's legal advisor or delegate reviews the contract to ensure that it contains the responsibilities assigned to the investigator's site, indemnification language, budget information, reporting requirements, and any other information required by the legal council. If the sponsor does not provide a contract, the investigator's site will provide an agreement to the sponsor that covers the same elements. When completed, the authorized institutional official for the investigator site and the investigator review all agreements. Usually three originals are prepared and signed: one for the investigator, one for the investigator's site, and one for the sponsor. It is important to allow sufficient time for both the sponsor and investigator's site to review and amend the contract or agreement. If possible, this process should begin at the same time as the ethics review.

### 2.4 Sponsor Interactions

The investigator is responsible for complete and proper study communication with the sponsor. A number of study visits are typically required by industry-sponsored trials. In investigator-initiated trials, which often have less funding available, fewer site visits are conducted.

The first visit is referred to as the *prestudy inspection visit*, and the purpose of this visit is for the sponsor to ensure that the clinical site is equipped to perform the study properly. The investigator should provide a tour of the facility and describe the patient population base and the methods that will be used to enrol participants into the study.

If the prestudy inspection visit is successful and the investigator agrees to participate in the trial, there is an *initiation visit*. The purpose of the initiation visit is for the sponsor to ensure that the investigator and the research associates have a correct understanding of the protocol activities and the methodologies. This is also an excellent opportunity for the investigator or research associates to ask any questions about the trial. In addition, the procedures for reporting adverse events, test article storage, and laboratory tests are reviewed.

Once the study is active and research participants are being enrolled, *ongoing study visits* occur. The frequency of these visits may vary, and the purposes are to ensure that the study protocol is being followed and that all documentation in the regulatory files is up-to-date. Adverse events and protocol deviations are also reviewed and outstanding issues are discussed.

After patient enrolment is complete, a *close-out visit* usually occurs. The purpose of this visit is to review study progress, to discuss how the test article is to be returned to the sponsor, and to determine how any corrections or outstanding issues will be resolved.

Appropriate research staff must communicate regularly with the sponsor, and all critical communication, including telephone calls and emails, must be documented. The investigator or delegate must notify the sponsor about the enrolment of the first research participant, recruitment progress, and if any adverse events occur. The investigator or research associate can also contact the sponsor if there is a question regarding a patient's eligibility status or about the protocol. All communication should be recorded in the regulatory binder.

### 2.5 Regulatory Documentation (Essential Documents)

Most countries take measures to regulate the development and marketing of medications and medical products. The investigator is ultimately responsible for obtaining, maintaining, and storing all required clinical study documentation at the site, although this task can be delegated to the research staff. An essential document is defined as any document that permits the evaluation of the conduct of a study and the quality of the data produced. Examples of essential documents include the study protocol, study manuals, and research ethics communications. When a study is being planned, it is important to collect and organize all regulatory documentation in an organized fashion. As the study progresses, copies of all documents are stored in the regulatory binder.

## 3 PATIENT SELECTION

### 3.1 Research Ethics Board Approval (Institutional Review Board Approval)

An ethics committee is an independent body of medical professionals and lay members. The responsibility of an ethics committee is to ensure the safety, well-being, and human rights of the research participants. Ethics committees review the protocol and consent forms to ensure that the trial is justified, safe, and that the patients are properly informed. All research involving human subjects should be referred to the local IRB or REB. The responsibilities of the IRB/REB are listed in Table 1.

### 3.2 Feasibility

Assessing the study's feasibility is closely related to reviewing an offer to participate in a clinical study. The eligibility criteria must be carefully assessed to ensure that

**TABLE 1 Responsibilities of the Ethics Committee**

| |
|---|
| To review an application for ethical approval of a research protocol in a reasonable time |
| To consider the qualifications of the investigator |
| To review each ongoing trial at an institution |
| To recommend modifications to the patient information and consent form, when appropriate |
| To review payments to trial participants |
| To determine that, when necessary, the trial protocol addresses ethical concerns such as consent by a patients legal representative and studies where prior consent is not possible |
| To perform duties in accordance with written operating procedures |
| To retain all relevant records for at least 3 years after completion of a trial |

your clinical site has sufficient eligible patients to be recruited for a clinical study. It is often necessary to perform a survey of potentially eligible patients over a 4-week period to provide an accurate estimate of recruitment rates. If you do not have a sufficient number of patients to make participation worthwhile, completing the trial successfully is not feasible.

As mentioned previously, the time required for the overall coordination, the patient enrolment and follow-up, and the data collection also needs to be carefully considered. It is important to be sure that you and your research staff have sufficient time and resources to successfully complete the trial requirements.

### 3.3 Sample Size

A sample size calculation before the conduct of a trial can help to estimate an appropriate sample size and to false-negative (β error) results. Before calculating the sample size, the investigator should clearly state primary and secondary outcome parameters. The primary outcome parameter is the one that the investigators consider to be the most important. Subsequently, any other measures are designated to be secondary outcome parameters. The initial distinction between primary and secondary outcome measures is important because the amount of outcome parameters impacts the threshold significance level that needs to be used to determine if the result is significant or not. A significance level of $p = 0.05$ is used by convention for the main outcome parameter. That means that a chance of 5% is being accepted to conclude that there is a significant difference between two groups, when in fact there is none (type I error or α error). For any additional secondary outcome parameters, adjustments of the significance level need to be made depending on the number of analyzed parameters.

The magnitude of the difference in the primary outcome parameter that the investigators consider clinically relevant should be the basis for the sample size calculation. Alternatively, this difference can be simply hypothesized. The sample size calculation will reveal how many participants per group are necessary in order to show if that difference truly exists or not.

In addition to the hypothesized difference in the primary outcome parameter and the significance level (usually $\alpha = 0.05$), the acceptable power of the study and the anticipated standard deviations of the primary outcome parameter in the two groups need to be established before proceeding with the sample size calculation. A study power of 0.8 is a conventionally accepted standard, which means that the investigators are willing to accept a 20% probability that there is no difference between two groups when a difference actually exists (β error). Any increase in study power and decrease of the α level of significance will result in a higher sample size requirement. The anticipated standard deviations in the two groups can be either determined by performing a preliminary pilot study or from data in the literature. If no data is available, they only can be estimated.

Even at best, a sample size calculation is based upon the best available "guestimate" of treatment difference between treatment groups. To improve the reliability of an a priori sample size calculation, investigators can conduct a pilot study of 20–50 patients to gain an estimate of the treatment effect in their proposed study population.

### 3.4 Recruitment Methods

Patient recruitment strategies vary depending on whether the study is investigating a chronic or an acute condition. For a chronic condition, research staff can screen pre-existing patient data, schedule appointments for potentially eligible patients, and identify patients who may be willing to participate in the study immediately after receiving approval from the ethics board. With a chronic condition, a known patient population can be recruited early and all at once. Patients with an acute condition can only be recruited when they present to the participating physicians; so enrolment will be staggered.

### 3.5 Screening

It is important to screen to identify potential patients who are eligible for participation in the clinical trial. Screening refers to identifying potentially eligible participants by applying the clinical trial's eligibility criteria. Areas that can be used for screening include hospital admission records, operating room schedules, and clinic appointment schedules. A screening log is used to maintain the screening information on each potential study participant that is screened, including if they met the inclusion and exclusion criteria, if informed consent was provided, and if they qualified and were entered into the study.

### 3.6 Informed Consent

Informed consent is the process by which a patient voluntarily confirms his or her willingness to participate in a clinical trial. Prior to giving consent, the study investigator or delegate must inform the patient of all aspects of the trial (Table 2).

**TABLE 2 What Patients Need to Know Before Participating in Clinical Research**

| |
|---|
| The purpose of the research |
| The trial treatment(s) and the probability for random assignment to each treatment |
| The trial procedures to be followed, including all invasive procedures |
| The subject's responsibilities |
| Any aspects of the trial that are experimental |
| The reasonably foreseeable risks and benefits |
| Any alternative treatment(s) that may be available |
| Any compensation available to the patient in the event of a trial-related injury |
| The anticipated payment or expenses, if any, to the patient for participating in the trial |
| The patient's participation is voluntary and the patient may refuse to participate or withdraw from the trial at any time without prejudice |
| Who will have access to their original medical records |
| The records identifying the patient will be kept confidential |
| If the results of the trial are published, the patient's identity will remain confidential |
| The patient will be informed if information becomes available that may be relevant to the patient's willingness to continue to participate in the trial |
| The person to contact for additional information on the trial |
| The foreseeable circumstances or reasons under which the patient's participation in the trial may be terminated |
| The expected duration of participation |
| The approximate number of patients involved in the trial |

**TABLE 3 Three of the Most Fundamental Rights**

| | |
|---|---|
| 1. | The patient's participation in the research trial is voluntary. |
| 2. | The patient has the right to refuse to participate or withdraw from the trial without providing a reason. Refusing to participate or withdrawing from the trial will not affect his/her subsequent medical care. |
| 3. | The patient will be informed of any new findings that may affect his/her willingness to continue participating in the trial. |

Informed consent must be documented in a written form, signed, and personally dated by the patient or by the patient's legally acceptable representative and by the person who conducted the informed consent discussion. Conduct the informed consent discussion in a quiet room and allow adequate time for questions. During the discussion, it is vital to communicate in nontechnical language and to take into account any language barriers.

Patients who participate in research trials have many rights and they need to be informed of these rights during the informed consent discussion. Three of the most fundamental rights are listed in Table 3.

### 3.7 Patient Confidentiality

The principal investigator and all study staff are responsible for maintaining patient confidentiality throughout the clinical trial. Confidentiality refers to the prevention of disclosing a research participant's identity and medical information to nonauthorized individuals. The study participant's involvement in a clinical trial must be kept private between the principal investigator, the appropriate study staff, and the primary care physician. The information is only to be shared when there is written permission by the study participant. All research participants' names and data obtained from medical records must be kept confidential. Patients are identified through a number that is assigned to them at study enrollment. When providing source documentation (medical records) to the sponsor, all personal identifiers must be removed. Finally, all data must be stored in a secure area, such as a locked cabinet or a password-protected computer in a locked office.

### 3.8 Intention to Treat Analysis

Intention to treat analysis is a method of data analysis on the method of the intention to treat a research participant (i.e., the planned treatment regimen) rather than the actual treatment regimen given. It has the consequence that participants allocated to a treatment group should be followed up, assessed, and analyzed as members of that group regardless of their compliance with therapy or the protocol, or whether they crossed over later to the other treatment group.

## 4 TREATMENT SCHEDULES

### 4.1 Supplies

The investigator or delegate is responsible for the accurate and complete accountability of the clinical supplies (i.e., drugs, surgical equipment, laboratory equipment,

etc.) used in a clinical trial and for proper storage of the supplies according to the sponsor's written instructions. The study protocol or the investigator's brochure will provide information about the study test article, its proper use, and the required storage conditions during the trial. The study staff is responsible for assigning study identification numbers and test article numbers to the supplies and maintaining the records of the use of supplies. The test articles are often blinded, and the blind must be maintained as instructed in the study protocol. A note in the file must be prepared to describe the situations when and how the blind is broken.

### 4.2 Adherence to Protocol

It is vital that all procedures and activities described in the study protocol are followed throughout the entire trial. Adhering to the protocol ensures reliable results, high-quality research, and efficient trial progress. Any deviations from the protocol need to be reported to the sponsor as soon as possible after they occur. Most trials have specific case report forms for recording and reporting protocol deviations.

### 4.3 Follow-up

It is the responsibility of the participating center to ensure that all patients are followed up as outlined in the study protocol. Follow-up visits should be scheduled within the time frames indicated in the protocol. The research participants must be made aware of all procedures that will be performed and the time each one will take (i.e., blood tests, radiographs, gait analysis). It is also necessary to communicate with any hospital departments that are assisting with the follow-up (i.e., laboratory medicine, radiology, physiotherapy). A plan of action should also be in place to prepare for any participants who are withdrawn or who do not attend their follow-up appointment.

### 4.4 Monitoring Visits

The purpose of a monitoring visit is to verify that the rights and well-being of patients are protected. The monitor will ensure that patients are being enrolled according to the protocol, and that the consent procedure has been carried out correctly and that the patients fulfill the eligibility criteria. In addition, the monitor will check the case report forms for legibility, accuracy, and completion of all data points. Some of these are then compared with the source documentation, such as notes in the medical charts, patient diaries, questionnaires, and test results, to ensure accuracy and completeness of reporting. The monitoring visits also allow time to discuss any questions about the study protocol or issues that have arisen that are not covered in the protocol.

The monitoring visits typically occur after the first patient has been enrolled and then every 2–6 weeks, depending on the rate of patient enrollment. To prepare for a monitoring visit, the research staff should check that all the administrative paperwork is in order, assemble the source documents, ensure that the case report forms are clean and all queries are resolved, and notify all departments involved in the study of the monitor visit as the monitor may need to check these areas.

### 4.5 Composite Endpoints

A composite endpoint is an endpoint that is defined in terms of two or more primary clinical endpoints at the patient level. The rationale for including a composite endpoint as opposed to a primary endpoint is that diseases often need multidimensional characterization, low event rates in component endpoints, may need to account for mortality, and the treatment effect on primary endpoints may be small. For example, a cardiovascular study could have three primary endpoints: cardiovascular death, stroke, and myocardial infraction. The composite endpoint would be time to the first occurrence of cardiovascular death, stroke, and myocardial infraction.

When event rates are low, the use of composite endpoints in clinical trials allows investigators to reduce sample size and the duration of follow-up. These advantages come at a price: The interpretation of the effect on the intervention is complicated, and the combined endpoint can be profoundly misleading. It is important to report the event rate for each component of the outcome when reporting on trials using composite endpoints.

## 5 PATIENT RESPONSE EVALUATION

### 5.1 Outcome Measures

It is important to document all trial-related activities to provide a lasting record of how the trial was conducted and of the data that was collected. Study outcomes may be based on safety, efficacy, or another trial objective and must be clearly defined in the protocol. Outcomes are typically recorded on the case report forms.

### 5.2 Case Report Forms

Data collected during a trial should be tailored to answer the hypotheses that are proposed within the protocol. To ensure that the correct data are collected, case report forms (CRFs) are designed. Patient data from source documents, such as the patient's medical records, are used to complete the CRFs. The completed forms are sent to the coordinating center responsible for the collection of the trial data. The coordinating center will then ensure that the final data set is complete, accurate, and complies with good study methodology.

The completion of CRFs will vary from study to study. All CRFs must be completed accurately and legibly and be reported in a timely manner. They must correctly reflect the data found in the source documents (medical records). All questions must be answered in an attempt to limit the number of data queries. Any changes or corrections must be dated and initialed and should not obscure the original entry. Finally, the CRFs must be completed in accordance to the trial-specific instructions. Table 4 describes data that should be included in a case report form.

### 5.3 Adverse Event Reporting

A serious adverse event (SAE) is any adverse occurrence or response to a drug/intervention, whether expected or not, that requires in-patient hospitalization or prolongation of existing hospitalization, that causes congenital malformation, that

**TABLE 4 Data Often Collected on Case Report Forms**

| |
|---|
| Patient identification (i.e., study ID and initials) |
| Patient demographic details (i.e., age, sex, height) |
| Adherence to protocol inclusion and exclusion |
| Baseline medical history |
| Diagnosis |
| Medication prior to procedure |
| Treatment details |
| Tracking of adverse events and other key outcomes |
| Discharge details |
| Follow-up visits |

results in persistent or significant disability or incapacity, that is life threatening, or that results in death. The principal investigator is ultimately responsible for ensuring that all serious adverse events are correctly reported to the sponsor and the REB/IRB in the required time frame (often 24 hours). Typically, the following information is required: type of event, whether it is expected or unexpected, a description of the SAE, action taken, the outcome, and whether the research participants remain on the study protocol. In addition, the local study investigator is asked whether the SAE should warrant continuation of the study, changes to the protocol, or revisions to the information or consent form.

An adverse event (AE) is any untoward event that significantly affects the research participant's well-being and does not fit the criteria of a SAE. It is often optional to report adverse events to the REB or IRB; however, it is necessary to report them to the sponsor.

Before the trial begins it is a good idea to familiarize yourself with the safety profile of the product and the SAE and AE reporting procedures so that you know what to do when an SAE or AE occurs. During the trial, it is important to inform patients of any potential AEs and encourage them to report all AEs. It is necessary to follow all patients who have experienced an SAE or AE until their condition is stable or resolved.

### 5.4 Data Queries

Although every effort is made to ensure the correct completion of case report forms, errors can and do occur. To guarantee data quality, a system must be in place to check and query all data. Data queries are a necessary part of a clinical trial, as they help ensure the quality of the data and the integrity of the trial. Data queries are generated by the data management team in response to missing, inconsistent, or illegible data, or to protocol deviations (Table 5). It is best to respond to data queries as quickly as possible. It is also sometimes necessary to discuss data queries with the study monitor or project manager if clarification is required.

### 5.5 Adjudication

A blinded, central adjudication of outcomes can be used in clinical trials as a way of reducing bias and random error in determining outcome events. This process may be especially important in clinical trials when the intervention cannot be blinded,

**TABLE 5 Purposes of Data Queries**

| |
|---|
| To clarify or confirm data |
| To request missing data |
| In special cases, to request additional data not specified in the CRFs |
| As a teaching medium for the correct completion of study documents |

such as in many surgical trials, or when the diagnosis of the primary outcome has low observer agreement. In addition to determining outcomes, central adjudication has also been used to assess eligibility of patients, protocol violations, and co-interventions.

There are many factors for clinical investigators to consider when deciding whether to use a central adjudication process for determining outcomes in a trial. Most importantly, the investigators must weigh the expected benefit of adjudication for accurate determination of outcomes against the substantial investment of resources involved and practicality undergoing this process. The investigators must also consider the potential for the adjudication process itself to bias the results of the trial.

To centrally adjudicate outcomes for a trial, it takes a considerable amount of administrative and expert time to collect the relevant information, prepare the information for review, review each case, and participate in consensus meetings. There may also be challenges involving the availability, validity, or usefulness of some documentation sources. If the treating physician is also required to make a judgment about whether an outcome occurred, it may be possible to compare the committee's judgment with the physician's to determine whether central adjudication is necessary.

Once the investigators or sponsors decide to centrally adjudicate outcomes, they must make several other decisions about the process, including who the adjudicators will be, what material must be evaluated, which judgments must be made, how to train the adjudicators, establishing a set of standardized decision rules, what the committee size will be, whether decisions will be made in pairs or by the full committee (if in pairs, whether to assign cases randomly), how to reach consensus on a decision, and how to monitor the accuracy of the process.

The agreement among adjudicators on a particular case can be affected by the number of decisions necessary, the number of choices for the outcome, and the complexity of the judgments. Disagreements can result from forgetting a decision rule, encountering a case for which there is no relevant decision rule, not having enough information, making an error, or the outcome is difficult to determine given all of the relevant information.

Outcomes assessment/adjudication committees review important endpoints reported by trial investigators to determine whether or not they meet protocol-specified criteria. Members of the adjudication committee may request radiographs, chart notes, operative reports, and other pertinent material to guide their decision making about a defined outcome. All attempts should be made to blind the committee to treatment allocation. This includes careful masking of all X rays and reports. Such committees are most desirable when the assessment of outcomes requires an element of judgment or subjectively (i.e., fracture healing), or when the intervention cannot be blinded. Ultimately, the adjudication committee members work together to limit bias in the outcomes assessment of a clinical trial.

## 6 DESIGN CONSIDERATIONS

### 6.1 Study Design (Observational Studies and Randomized Controlled Trials)

The types of study designs used in clinical research can be classified broadly according to whether the study focuses on describing the distributions or characteristics of a disease or if it focuses on elucidating its determinants (Table 6). Descriptive studies describe the distribution of a disease, particularly what type of people have the disease, in what locations, and when the disease occurred. Cross-sectional studies, case reports, and case series represent types of descriptive studies.

Case reports are an uncontrolled, descriptive study design involving an intervention and outcome with a detailed profile of one patient. Expansion of the individual case report to include multiple patients is a case series. Although descriptive studies are limited in their ability to make causal inferences about the relationship between risk factors and an outcome of interest, they are helpful in developing a hypothesis that can be tested using an analytic study design.

Analytic studies focus on determinants of a disease by testing a hypothesis with the ultimate goal of judging whether a particular exposure causes or prevents disease. Analytic design strategies are broken into two types: observational studies, such as case–control and cohort studies, and experimental studies, also called clinical trials. The difference between the two types of analytic studies is the role that the investigator plays in each of the studies. In the observational study, the investigator simply observes the natural course of events. In the trial, the investigator assigns the intervention or treatment.

One type of observational study is the case–control study, which starts with the identification of individuals who already have the outcome of interest, cases, and are compared with a suitable control group without the outcome event. The relationship between a particular intervention or prognostic factor and the outcome of interest is examined by comparing the number of individuals with each intervention or prognostic factor in the cases and controls. Case–control studies are described in greater detail later.

In the cohort study design, the cohort represents a group of people followed over time to see whether an outcome of interest develops. Ideally, this group meets a

**TABLE 6 Hierarchy of Evidence**

| Study Design | |
|---|---|
| Randomized controlled trials | Less bias |
| Controlled trials (e.g., controlled before–after studies) | |
| Case–control studies and cohort studies | |
| Cross-sectional studies | |
| Expert opinion, case reports, and case series | More bias |

level of certain predetermined criteria representative of a population of interest and is followed with well-defined outcome variables. Usually, this group is matched with a control population selected on the presence or absence of exposure to a factor of interest. The purpose of this type of study is to describe the occurrence of certain outcomes with time and to analyze associations between prognostic factors and those outcomes.

Randomized controlled trials classically are held as the standard against which all other designs should be measured. In a randomized controlled trial, patients are assigned to a treatment group or a control group. The control group usually receives an accepted treatment or no treatment at all, whereas the treatment group is assigned the intervention of interest. Randomized controlled trials are thought to represent the highest quality of evidence based on their methodological strengths of randomization of patient assignment and blinding of intervention and outcome.

### 6.2 Trial Organization and Responsibilities

We will provide an overview of the organization of a multicenter trial with emphasis on the trial committees, methods and coordinating center, and participating sites. Figure 1 details how a trial may be organized.

The complexity of a trial with multiple participating centers and hundreds of participating investigators requires key organizing committees to overlook the conduct of the trial, to assure patient safety, and to limit bias in outcomes assessment.

The steering committee is responsible for the overall design and conduct of the trial. The members of this committee may or may not be direct participants of the proposed trial. Often the committee consists of the principal investigators, a biostatistician, a trial methodologist, and other key individuals deemed important to the design and conduct of the study. The steering committee communicates with the trials coordinating center, data safety monitoring board, outcomes adjudication committee, and participating sites on a regular basis. At the completion of the trial, the steering committee maintains responsibility for the data analysis and manuscript preparation on behalf of all study investigators and participating sites.

All clinical trials require safety monitoring, but not all trials require monitoring by a formal committee external to the trial investigators. Data Safety and Monitoring Boards (DSMBs) or Data Monitoring Committees (DMCs) have generally been established for large, randomized multicenter studies that evaluate interventions intended to prolong life or reduce risk of a major adverse health outcome.

Outcomes Adjudication Committees review important endpoints reported by trial investigators to determine whether they meet protocol-specified criteria. Outcomes adjudication is important when the primary outcome of a study requires judgement and is prone to bias in its assessment, especially if the assessment cannot be blinded.

The foundation of a large multicenter surgical trial is the methods and coordinating center. All the day-to-day activities, including centralized randomization, data management, and overall coordination of the trial, occur at this site. The coordinating center can be a contract research organization that “monitors” the trial for a group of clinical investigators. Alternatively, it can be the site of a principal investigator.

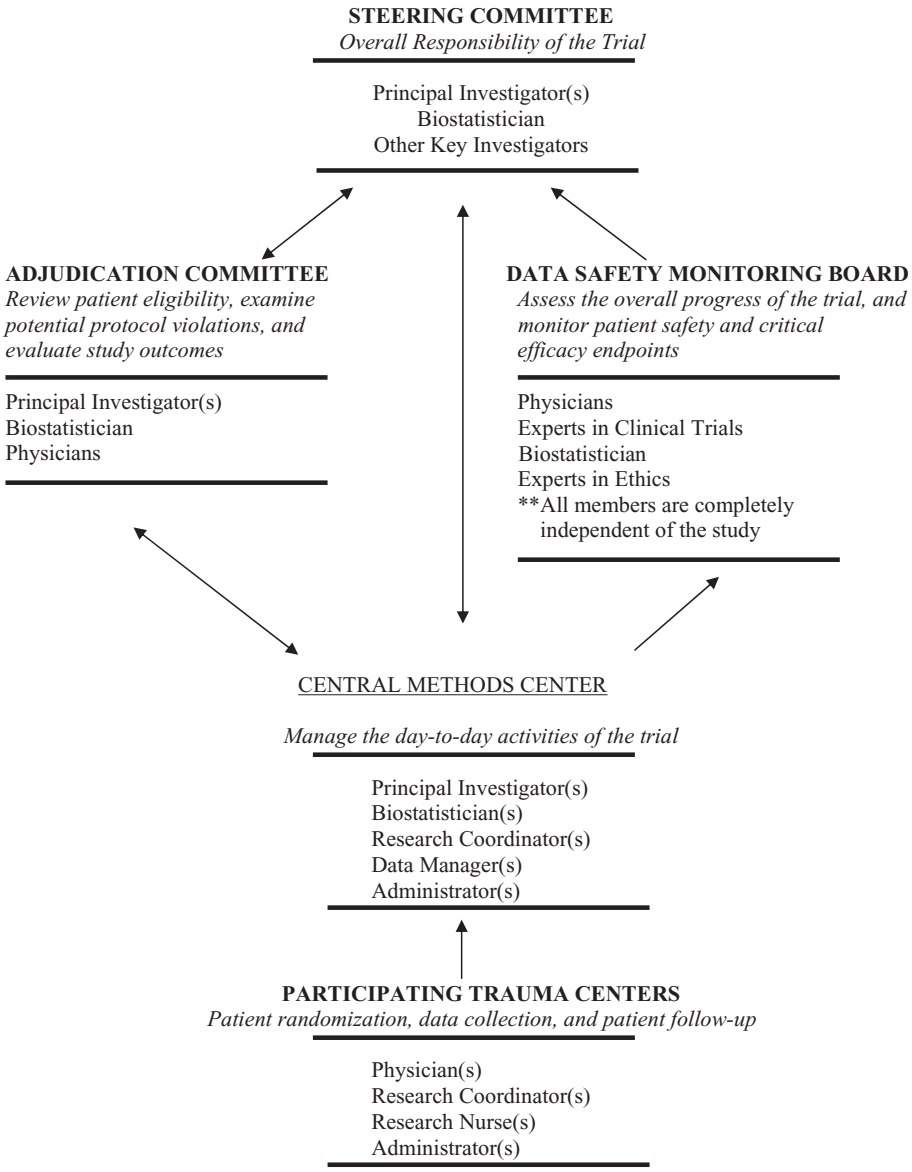


***FIGURE 1*** *Trial organization.*

Each participating site often has more than one investigator enrolling patients for a multicenter trial. In this situation, one investigator from each site is designated as the "site principal investigator (PI)." This site PI serves as the primary contact for the methods and coordinating center and acts on behalf of the participating center. Each participating center also has a dedicated research coordinator who manages the day-to-day issues of patient enrollment, data collection, and follow-up planning. The research coordinator and the site PI work together to ensure compliance, data quality, and effective communication with the methods center.

Participating centers typically identify patients through direct within-center referral or referral from other medical centers. Clinical centers should receive complete sets of all data forms prior to joining the study. The research coordinator for each participating center will ensure that all forms are complete and faxed to the methods center as soon as completed. Each site should be given a follow-up schedule that details the type of information and forms to be completed.

## 6.3 Data Management

Collection of data by the site investigator is often the first step in a complex procedure leading to a clean database and final study report generation. Failure to adhere to data collection timelines can significantly delay this procedure. Case report form data pages and additional information is logged, and then the data is entered or scanned into a trial database. Any data queries are faxed or sent to the center for completion or resolution and the database is subsequently amended. Data is usually double-verified to check for accuracy of the data entry.

## 6.4 Quality Control

Quality control methods are operation systems and processes established to ensure the quality of a clinical trial, the accuracy and integrity of data, and the compliance with regulations. The clinical investigator is responsible for ensuring that quality control methods are conducted for each study at their clinical site. It is important for sites to have standard operating procedures (SOPs) to help ensure quality and consistency for both general research procedures as well as for procedures for each clinical study. The purpose of the SOPs is to describe in detail who will do each research task and how each task will be completed. Most research institutions have SOPs available for review.

## 6.5 Study Close-out Activities

A number of activities need to be completed by the study staff before the close-out of a study. Table 7 lists several of these items. It is important that everything is organized and complete for the close-out visit.

**TABLE 7 Items to Be Completed for Close-out of Clinical Trial**

| |
|---|
| All monitoring visits |
| All case report forms |
| Any corrections on the case report forms |
| Test articles from research participants |
| Regulatory binder |
| Return all study material not used |
| Completion report to the REB or IRB |
| Arranges for secure storage of CRFs, source documents, and regulatory documents |

## BIBLIOGRAPHY

Bhandari, M., and Schemitsch, E. H. (2004a), Randomized trials: A brief history and modern perspective, *Tech. Orthop.*, 19, 54–56.

Bhandari, M., and Schemitsch, E. H. (2004b), Planning a randomized trial: An overview, *Tech. Orthop.*, 19, 66–71.

Bhandari, M., and Schemitsch, E. H. (2004c), Beyond the basics: The organization and coordination of multicenter trials, *Tech. Orthop.*, 19, 83–87.

Bhandari, M., and Tornetta, P. (2004), Issues in the hierarchy of study design, hypothesis testing, and presentation of results, *Tech. Orthop.*, 19, 57–65.

Bristol, D. R. (1989), Sample sizes for constructing confidence intervals and testing hypotheses, *Stat. Med.*, 8, 803–811.

Canadian Institute of Health Research, http://www.cihr-irsc.gc.ca/; accessed Dec. 2005.

Flather, M., Aston, H., and Stables, R. (2001), *Handbook of Clinical Trials*, ReMedica Publishing, London.

Food and Drug Administration (1998), Statistical principles for clinical trials, *Fed. Reg.*, 63(179), 49583–49598.

Food and Drug Administration (2001), Guidance for Clinical Trial Sponsors: On the establishment and operation of clinical trial data monitoring committees. Center for Biologics Evaluation and Research, November.

Guyatt, G. H., and Rennie, D., eds. (2001), *User's Guides to the Medical Literature: A Manual for Evidence-Based Clinical Practice*, American Medical Association Press, Chicago.

Hartz, A., and Marsh, J. L. (2003), Methodologic issues in observational studies, *Clin. Orthop.*, 413, 33–42.

Jackowski, D., and Guyatt, G. (2003), A guide to measurement, *Clin. Orthop.*, 413, 80–89.

McMaster University, Hamilton Health Sciences, and St. Joseph's Healthcare (2004), Standard Operating Procedures for Clinical Research, May.

Montori, V. M., Busse, J. W., Permanyer-Miralda, G., Ferreira, I., and Guyatt, G. H. (2005), How should clinicians interpret results reflecting the effect of an intervention on composite endpoints: Should I dump this lump? *ACP J. Club*, 143(3), A8.

Sprague, S., Hanson, B., and Bhandari, M. (2003), Informed consent: What your patients need to know about entering clinical research studies, *Can. J. Diag.*, 20(10), 29–31.

Zlowodzki, M., Bhandari, M., Driver, R., Obremskey, W. T., and Kregor, P. (2004a), Beyond the basics: Internet-based data management, *Tech. Orthop.*, 19, 88–93.

Zlowodzki, M., Bhandari, M., Brown, G., Cole, P., and Swiontkowski, M. F. (2004b), Planning a randomized trial: Determining the study sample size, *Tech. Orthop.*, 19, 72–76.

# *Process of Data Management*

**Nina Trocky[1] and Cynthia Brandt[2]**

[1]*The University of Maryland Baltimore School of Nursing*
[2]*Center for Medical Informatics, Yale University School of Medicine, New Haven, Conneticut*

**Contents**

1 Introduction
2 Data Management Study Plan: Overview and Definitions
3 Quality Plan
  3.1 Quality Control
  3.2 Quality Assurance
  3.3 Continual Quality Improvement
4 Data Management Team Structure
5 Case Report Form Design and Guidelines
  5.1 Overview and Definitions
  5.2 CRF Design and Development
  5.3 Measurement Basics
  5.4 Standards for Data
  5.5 Testing and Reviewing CRFs
  5.6 Summary: Well-Designed CRFs
6 Data Acquisition and Handling Guidelines
  6.1 Definitions and Overview
  6.2 Data Acquisition
  6.3 Data Handling
7 Summary
  References

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

## 1 INTRODUCTION

"Collecting data that are accurate, honest, reliable and credible is one of the most important and most difficult objectives of conducting clinical research" [1]. The process of data management is composed of various discrete and interrelated activities that are intrinsically linked in an individual and building in a successive fashion. Once the processes and activities are identified, they can then be implemented and managed. Once initiated and implemented, they can be measured and monitored.

There is a need to have a structured and orderly approach to manage these complex processes, and a data management plan may help provide this structure. These plans may be referred to by various names such as data management plan (DMP), data study plan (DSP), data management study file (DMSF), and others that may be specific to the workplace or environment. In general, these plans do provide direction to conform practice to regulate the complex processes involved. Although they all will be based upon regulatory and sponsor requirements as well as good clinical practice (GCP) guidelines, the actual implementation may vary due to interpretation differences. Successful trial management demands careful planning and deliberate execution.

For the purposes of this chapter we attempt to provide a conceptual framework and structure for the components that may be included in a data management study plan (DMSP). The DMSP will be individualized to each clinical study that is implemented. This chapter is organized by components that might be included in a DMSP. The main sections to be addressed in this chapter include the quality plan, data management team structure, case report form (CRF) design and guidelines, and data acquisition and handling guidelines.

## 2 DATA MANAGEMENT STUDY PLAN: OVERVIEW AND DEFINITIONS

The DMSP can provide the framework for how and where the processes and procedures are defined and outlined for individual protocols or studies. The goal of the plan is to help standardize data management practice and minimize practice variations, assure compliance with regulations and institutional policies and procedures, and facilitate consistency of data management while improving study performance.

The components of the DMSP define the road map and the specifics of what, how, who, when, and why and help define the activities at the institutional level. There will be specific elements to support the activities of data management such as data collection, entry, handling, and developing discrepancy resolution guidelines. The entire DMSP should be designed in concert with quality tools that support the incremental review, periodic measurement, and adjustments to process deviations. The quality plan is a key section of the DMSP, and quality concepts and specific quality actions are included at every section in the DMSP.

The end product of the DMSP is quality data that can support testing of the hypothesis and compliance with the regulations and policies. It is important to identify the activities and personnel involved in data management prospectively. Once the trial is initiated, the defined activities must be monitored and measured

against pre-established benchmarks and thresholds to minimize deviations to the plan. The structure of the DMSP must be generic and applicable to all protocols or studies implemented while the content or specific procedures and definitions will be individualized. The DMSP should be used in concert with institutional standard operating procedures (SOPs) and is influenced by all appropriate regulatory and institutional requirements and good clinical practice (GCP) guidelines. It may even reference a manual of operations. The DMSP should emphasize quality and consistency from the design of the protocol through publication. It provides guidelines for ensuring consistent data collection, accurate data transfer, subject safety, regulatory compliance, and safety monitoring and reporting. In general, a DMSP might include the following components:

- Protocol
- Scope of work and study contract
- Data management plan
- Data validation testing guidelines
- Data entry and tracking guidelines
- Data coding guidelines
- Data handling and transfer guidelines
- Quality control and quality assurance plan
- Data management team structure
- Serious adverse events reporting and reconciliation guidelines
- Training guidelines
- Data import and export guidelines
- Study report guidelines
- Database management system specification guidelines
- Annotated case report forms
- Data storage and archiving guidelines

This chapter will describe the four sections to include in the DMSP:

- Quality plan
- Data management team structure
- Case report form design and guidelines
- Data acquisition and handling guidelines

## 3 QUALITY PLAN

The definition for good clinical practice is: "A standard for the design, conduct, performance, monitoring, auditing, recording, analysis, and reporting of clinical trials that provides assurance that the data and reported results are credible and accurate, and that the rights, integrity, and confidentiality of trial subjects are protected" [2]. A quality plan is a key element in the DMSP, and quality processes need to be incorporated into all components of data management activities. The purpose

of a quality plan is to decrease the degree of uncertainty in the measurement and reporting of data as well as to increase the reliability and validity of data utilized to test the study hypothesis. It should describe the quality control and the quality assurance procedures to be utilized for the study activities including the percentage of subjects for whom source documents will be monitored. Three components of the quality plan deserve particular attention: quality control, quality assurance, and continual quality improvement. Together these three approaches will measure and monitor the characteristics and traits of the data and related processes that support the clinical data management operations. They build upon each other to complement and encompass more and more processes that can be measured and analyzed and if needed, corrected. The quality components are built upon the stabilizing elements of the standard operating procedure, training processes, and basic trial conduct such as data collection and analysis activities. It is this combination of interwoven and interdependent components that facilitate the proper planning and execution of data management activities, which facilitates optimizing results.

### 3.1 Quality Control

Guidance for Industry E6 Good Clinical Practice: Consolidated Guidance defines quality control (QC) as "The operational techniques and activities undertaken within the quality assurance system to verify that the requirements for quality of the trial-related activities have been fulfilled" [3]. Trial output is ultimately affected by the operational checks and balances throughout the data management processes that are performed at every step of the clinical trials process and applied to each stage of data handling that ultimately affect the trial output [1]. Identification earlier in the process will reduce errors by creating a robust checks and balance system. Implementing QC procedures will help to assure that trial procedures and processes have occurred as well as that the activities have been performed in accordance with some threshold or standard. A deviation or error may be noted soonest and adjusted when there is an established, ongoing measurement and data processing review. Since data management processes are performed by a variety of individuals, QC checks must be incorporated into the daily work flow of each data management team member. Quality control in clinical trials or clinical research should generally encompass all the individual procedures that involve protection of human subjects from research risk and affect reliability of the data in order to assure internal consistency of the trial.

Quality control checks may be thought of as snapshots that provide limited insight into the "health" of the data management activities. These snapshots will be used to decide whether to delve further in order to evaluate or assess data integrity components. QC activities may consist of written SOPs and work instructions to be followed such as the process to resolve data discrepancies or clarifications, data validation procedures, or personnel training requirements. QC checks can set predefined thresholds such as levels for acceptance testing for an electronic data capture tool.

The DMSP should include study-specific QC activities or steps for critical activities in the entire data management process from protocol and case report form development and subject accrual to data collection, archiving, and cleaning. QC steps can be performed in real time or midstream and some examples are:

1. Comparing programmed edit checks to the standard study-specific edit checks prior to sign off or implementation of an electronic data collection form for a study. An example could be an allowable data range that could vary across studies or populations such as age- or gender-related laboratory values.
2. Performing source document verification (visually comparing the original data source) of all the critically defined data fields just entered into a data collection instrument or computer system following data entry.
3. Following randomization of a new study subject, verifying the steps and the randomization assignment by double checking the work instruction or standard operating procedure.
4. Following the ordering of an investigational agent, a second qualified individual double checks the dosing based on the protocol investigational agent schema.
5. Reviewing and closing a univariate discrepancy as soon as the electronic database notifies the data entry operator of an expected value.

The quality assurance (QA) focus will take advantage of the cross-functional relationships of the data management team by directing efforts on searching for, identifying, and resolving patterns of weakness within one or more processes [1]. Building upon errors or inconsistencies identified at the operator level through the quality control checks, the quality assurance component of the quality plan can focus on measuring performance outcomes rather than just individual process variations.

### 3.2 Quality Assurance

Both QC and QA activities are intimately tied to data management. In contrast to QC, QA activities apply to actions that have been performed in the past or apply to past events. Guidance for Industry E6 Good Clinical Practice: Consolidated Guidance defines QA activities as "all those planned and systematic actions that are established to ensure that the trial is performed and the data are generated, documented (recorded), and reported in compliance with GCP and the applicable regulatory requirements" [4]. These activities can assure that predefined data quality requirements have been met and that compliance with certain standard operating procedures or specific data-handling guidelines are measured.

Two components of QA would be audits of performance and audits of systems. As the DMSP attempts to provide the structure for data management activities to occur in an organized and efficient order, QA audits will serve to measure process and outcomes of the data management activities. Performance audits would examine how closely an individual or a research site adheres to a procedure or a process. For example, a performance-focused audit surrounding data entry could include a second independent data entry operator (DEO) reviewing a sample of fields or CRFs completed by the initial DEO. Compliance with data entry guidelines or discrepancy resolution guidelines might be evaluated. An audit of a system could

involve a review of the procedure in order to identify a deficiency in the SOP versus a deficiency in the process of annotation.

These audits should be independent of all trial-related activities that should be used to examine and measure the degree of compliance to a certain procedure or measurement of a performance indicator such as queries generated per site or per study monitor.

Audit reports may address timeliness, completeness, compliance, and consistency of data as compared to the expected results. Additionally, audit reports may be used to communicate the need for corrective action plans to address deficiencies, inconsistencies, and other operational limitations that would adversely influence overall study conduct. A QA activity may include sampling a portion of data in order to assess overall confidence in the quality of a process, an outcome, or a specific tack. QA activities may also be employed to ensure QC procedures have been effective in identifying and correcting errors. Some examples of QA activities that might be included in a DMSP are:

1. Assessing the data management study plan for completeness (required sections have been completed and maintained current).
2. Auditing a sample of the personnel-training file against the role-based requirements.
3. Sampling a number of serious adverse events to determine if the events were reported as required by the study protocol.
4. Running an ad hoc query to length of time each research site takes to resolve open discrepancies.
5. Reviewing the cycle times for the last three studies built by the clinical programmers to determine if each was within the defined time period (e.g., 5 days).
6. Audit query resolution processes of each in-house clinical monitor to the specific SOP.

### 3.3 Continual Quality Improvement

Correcting variations in practice or compliance deficiencies must be addressed, but ongoing surveillance with the goal of improving processes is also important. The continuous quality improvement process is a natural companion to QC and QA tools. Adopted from the principles of total quality management (TQM), continuous quality improvement (CQI) serves as an additional tool to correct inefficient or ineffective processes. CQI is a method of evaluation that is composed of structure, process, and outcome evaluations that focus on improvement efforts [5]. CQI can help institutions enhance existing programs and improve effectiveness of processes because the variation in practice can be viewed from multiple angles. Effective CQI activities depend on the collaborative work environment, which is how data management operations are developed.

Drilling down to identifying a process to improve, even incrementally, can be achieved by incorporating the CQI methodology. The emphasis is on identifying the root cause(s) of the problem and then designing an appropriate intervention to eliminate or reduce the reasons for the problem. So, first a problem or issue must

be identified. Next a corrective action plan is developed. Then the plan is set in action. Finally, the results of the actions are monitored. One relatively convenient model to employ is called the FOCUS PDSA model [6]. The relatively simple steps may be utilized to quickly frame the problem, identify the reasons for the problem, and create an effective intervention. The steps are outlined below and will begin with an understanding and agreement on what the problem is and why it exists:

- **F**ind a process to improve or an problem to correct.
- **O**rganize the interdisciplinary team to discuss process or the problem.
- **C**larify what is known and gather supporting documentation.
- **U**nderstand why the process variation occurs or the problem exists.
- **S**elect a process improvement activity based on the above analysis.

Once the process or the problem is thoroughly understood and the remedial actions are defined, the next step is to continue through a process improvement plan as:

- *Plan* Create a timeline of resources, activities, training, and target dates to establish the completion date. Develop a data collection plan, tools for measuring outcomes, and define thresholds for determining when targets have been met.
- *Do* Implement interventions and collect data.
- *Study* Analyze the results of data collected and evaluate reasons for variation, if any.
- *Act* Act on what is learned and determine next steps. If the intervention was successful, determine if the current processes or procedure were not followed and so new procedures are not needed. Additional training may be required.

Conversely, practice variations may have resulted because procedure is antiquated or nonexistent. Finally, review of the PDCA system may be necessary as it was not robust enough to propel change. In this case it may be necessary to repeat the PDCA cycle.

Continual quality improvement builds upon the narrow or individually focused efforts of the QC checks and the broader procedure and process focused audits, organizational and system checks of the QA activities. The CQI model enhances the quality plan by directing collaborative energies toward processes not individuals. Data management is process driven. CQI efforts, likewise, focus on critical processes that require movement from one person to another and to another but from a systems perspective. CQI can be thought of as a shared effort that enables people to work together across organizational boundaries to improve shared processes [1].

A robust quality plan will force the data management team to evaluate their practices to measure how well they match their assumptions, formal guidelines, and predefined measurements of performance. In fact, it is the outputs from this plan that will form the team's quality profile. The determination and measurement of quality cannot rest in the hands of a select few. Rather it must be a shared responsibility and expectation among all participants on the data management team.

## 4 DATA MANAGEMENT TEAM STRUCTURE

It is important for data management team members to understand the scope of their work perfectly. The team members need to understand the deliverable or a specific work component (a portion of a larger deliverable) that they are expected to complete or assist in completing. If there are quality criteria to meet, the team members should know these quality requirements. More importantly, the data management team members should clearly understand the dependencies and relationships that exist between themselves and the other departments such as regulatory or information technology. The DMSP will include a reporting or organizational chart detailing responsibilities and lines of communication. By delineating the structure and composition of the team, energies may then be directed toward problem solving, task effectiveness, and maximizing resources to achieve the team's purpose. Sound team building recognizes that it is not possible to fully separate one's performance from those of others.

As the data management team structure is comprised of various professionals from diverse disciplines, what then are the primary activities of the data management process? In general, the primary aspects of data management consist of:

1. Handling, processing, and analyzing information for the purpose of supporting clinical research activities
2. Developing protocol-specific case report forms (CRFs)
3. Reviewing and approving programmed validation rules and edit checks in the database systems
4. Collecting data from the medical record or source documents and transcribing it into protocol-specific CRFs
5. Reviewing and evaluating data for inconsistencies, omissions, and errors.

The process of clinical data management knits together various key players who provide specialized skills and knowledge.

Depending on the work environment and setting, there may be different types of personnel performing one or more data management activities. For example, a quality specialist may perform audits in one setting and in another setting may serve as an independent review for final CRF design approval. Various disciplines and personnel comprise the data management team. As such, individual setting or environments may have their own unique set of responsibilities and the position descriptions may vary. Regardless, the key element is that we understand that various disciplines must all be interrelated to manager and process data. Some examples of personnel who are involved in the clinical data management activities are:

- *Data Manager* Core or central person who coordinates the various activities for a specific project assuring project goals and objectives are met. May develop protocol-specific CRFs, define programmed edit checks for electronic CRFs, identify critical safety data fields, and coordinate final approvals of CRFs. The data managers also may develop data entry guidelines and have other responsibilities depending upon the environment.

- *Medical Writers* Develop and prepare study protocol, safety reports, and for interim analyses.
- *Statistician* Review clinical and safety databases (or data structures and CRFs) to assure study endpoints can be measured and protocol-specific data are collected, determine safety and efficacy endpoints, conduct data safety monitoring board interim analyses. Develop or work with database programmers to create specialized statistical programs to support data integration, data reporting, and safety analysis.
- *Regulatory Specialists* Confirm human subject protection and subject safety are preserved, develop and maintain standard operating procedures or work instructions, conduct training and education classes, and maintain regulatory files; reviews initial and continuing protocol submission to the Institutional Review Board (IRB).
- *Programmers* Build study-specific data collection instrument (tool), revise multistudy clinical research database system or create study-specific database, modify eCRF or data system based on protocol amendments or sponsor safety reports. Work with statistician, study coordinator, or data manager to create data queries for reports and analyses.
- *Investigator* Responsible for study conduct at the research site, approves data collected and entered into the CRF, assures compliance with all regulatory, scientific principles, ethical, legal, and GCP standards, assures approved reports are prepared and safety information is reported to IRB, sponsor, and other investigators.
- *Research Nurse or Study Coordinator* May perform data collection and entry into CRFs, supervise data entry personnel, audit data entered by data entry personnel, coordinate site recruitment, subject screening, and enrollment, and monitoring activities, maintain regulatory binder and essential document file, coordinate training of site personnel, records maintenance, provide required documents/reports to regulatory, sponsor, institutional entities, notifies IRB/sponsor immediately of a serious adverse event (SAE).
- *Monitor* Principal link between sponsor and investigator, assure rights and well-being of human subjects protected, oversees progress of the trial and ensure study is conducted and data are handled in accordance to protocol, GCP, ethical and regulatory requirements, implements SOPs, provide site staff education, verify data is accurate, complete, and verifiable, assures CRFs are completed accurately.

Along with delineating the personnel and their roles and responsibilities in the DMSP, the plan should also include standard operating procedures (SOPs) that detail processes for data handling to assure accuracy, reliability, safety, security, and privacy. These processes will be further outlined in specific guidelines such as data entry guidelines, data-coding guidelines, and data-handling and transfer guidelines and case report form design guidelines. The data management team structure serves to clearly define role delineation from a functional framework as well as a division of responsibilities. Therefore, development of the case report forms must rely on the team members to integrate and apply study-specific guidelines, sponsor, protocol and regulatory requirements, clinical guidance, subject safety, and endpoint

measurements. It is from the case report forms that the data collection processes may occur.

## 5 CASE REPORT FORM DESIGN AND GUIDELINES

### 5.1 Overview and Definitions

Case report forms are questionnaires or instruments that are used to collect required data about cases or subjects enrolled in a study in a structured and standardized manner to facilitate reliable, consistent, and clean data for analysis. The CRFs, or data collection instruments, should be designed by investigators in order to measure and define the specific data needed to support or disprove the hypothesis or goal of the study. General purposes of the CRFs are to: (1) meet the objectives of research; (2) obtain most complete and accurate information possible; and (3) do this within the limits of available time and resources.

Ideally, CRF templates are designed to help standardize the information and data collected and facilitate collaboration and reuse for future similar studies. CRFs may be created and used in different modalities or formats including paper or electronic documents. The appearance and functionality of the CRFs will reflect the modality and how they will be used and the capabilities of the environment where they will be implemented.

### 5.2 CRF Design and Development

***CRF Design Best Practices*** When designing CRFs, the content and structure of the data items (or questions) to be included in the CRF should be considered first rather than the modality or appearance of the CRF. It is best to design the CRFs at the same time that the protocol is being created in order to assure that data specified in the protocol is collected and consistent with the hypothesis of the study and feasible to collect. This will help to keep questions, prompts, and instructions clear and concise and help to assure that the CRF design will fit with the data flow from the perspective of the person completing it. The flow of study procedures and typical organization of data in a medical record (or data source) should be taken into account as well.

Other design issues include planning for reusability, collaboration, and standardization when possible. It is helpful to create and maintain a library of standard forms. If there are previously performed studies or pilot work that has been used for measuring similar outcomes or measurements, CRFs and previous items may be used and adapted as necessary using experiential information. Finally, design the CRF with the primary safety and efficacy endpoints in mind as the main goal of data collection and pretest and review them prior to finalization and approval.

Broadly conceptualizing the CRFs in the study. As discussed in previous chapters of the book, clinical studies generally are designed in a highly structured fashion, and many are divided into chronological periods of varying duration, with different evaluations and tests performed or scheduled for subjects at different time periods. Investigators need to determine how and when the data will be collected and

entered and what the reporting requirements will be. This information can assist in determining the design of the CRF that will fit with the data flow of the study procedures and hopefully decrease the tendency to measure data inappropriately and redundantly. Commonly used time points in a clinical study include pretreatment, screening, baseline, treatment, follow-up, and evaluation or summary points. It is important to remember that the time point a CRF is collected is important data as well as the individual data items as CRFs may be used at numerous times or only once during a study. An overall study data schedule will need to be examined along with the individual CRFs that will be collected, prior to study initiation in order to make sure that the proper data is getting collected at the right times and at the correct frequency (repeats) appropriate for analysis to test the study hypothesis(es).

***CRF Level Design*** Once the overall study schedule of procedures or visits has been determined CRF designers should then look at form level issues that will facilitate or hinder data collection or entry and data analysis. First consider the type of data the CRF is used to collect and who will be using and completing it. The goal is to make it easy to use. For example, if a care provider such as a nurse or physician will be using the CRF for data collection, questions may be arranged and grouped differently than if the subjects will enter data into the form. The education level, age, language, and culture all must be considered when designing CRFs for subject entry. Likewise the skill level of the data collector in the research setting is also important to know. Questions may need to be worded for less clinically educated data collectors and more or less help made available to provide explanation of complex items. The following information should be described in the DMSP:

1. Designing for the type of data and the flow of subjects in the study.
2. Designing for who will be completing the CRF in order to assure that it makes sense to that person.
3. Ensure that a data item is collected once and only in one place to avoid conflicts in data. This will help to avoid referential and redundant data points within the CRF.
4. Grouping the same type of data together on the same forms.

For considerations of layout, format, and usability, we will focus primarily on issues related to eCRFs. For overall CRF layout and design, it is important to consider the data flow and the dependency of the items on each other when putting the items in an order or group on the CRF. The order of the items should be determined and interrelationships and dependencies between the questions clarified. For example, if the answer to question 1 determines whether the next few items will be relevant or not and skipped should be made clear on paper CRFs and programmed into electronic CRFs. These "skip patterns" can greatly facilitate either confusion or correct use and entry of data depending on formatting and programming for eCRFs.

***CRF Modality*** Final considerations to overall CRF design will need to be made knowing the modality for use of the form. For example, will it be implemented electronically on a personal digital assistant (PDA) or on a paper form to be

scanned, faxed, or mailed? Once the modality(s) have been determined, different tools can be used to make the CRFs most visually conducive to clear and accurate data entry. The font size should be easily readable, copy and fax well, and the use of different styles for visual emphasis and clarity of reading should be used to help the form filler identify specific sections and complete the CRF.

### 5.3 Measurement Basics

It is not practical to discuss in detail the large corpus of material on measurement theory that describes how to create questionnaires and how to design scales, as there are numerous textbooks and a large body of research on this subject. A few key practical steps that can help investigators and study staff to design data collection instruments or CRFs are contained in Sudman and Bradburn [7].

When designing items to measure standard concepts, it is best to identify standardized and validated previously used items to include in the CRFs. The use of items from standardized instruments will help assure that the items are appropriately defined. When standard validated questions do not exist, the investigator must create new questions or revise existing questions, taking into consideration question formats, phrasing, responses, and categories. Instructions for completion/prompts should be simple, clear, and sufficient for whoever is completing the form. Investigators should consult personnel trained in the development of measurement instruments, as this is complex and error prone.

When creating electronic CRFs, it is best to enter and store raw and original data rather than calculated or summary data. For example, the collection of birth date, rather than age, is more useful. Free text input are less optimal to forced selection of text choices, but many CRFs will require free text entry. Other data types may be based upon standard vocabularies such as those used to code common procedures or diagnoses, and a table look-up will help accurate data entry and collection. For more complex data types such as images, electrocardiograms (EKGs), and the like, different electronic solutions are available such as picture archiving and communication (PAC) systems and radiology information systems.

### 5.4 Standards for Data

Researchers performing epidemiological and clinical research frequently participate in collaborations for reasons such as: joint-funding initiatives, to increase sample size in the case of rare diseases or outcomes, and to pool expertise and resources. When designing CRFs for collaborative research, it is important to have the entire collaborative team review and test pilot versions of CRFs. The use of standard CRFs or CRF templates with standard questions and data items will facilitate the ability of data to be merged for collaboration postcollection if the design of the CRFs cannot be done pre-data collection. Identification of useful, validated CRFs, however, is difficult for various reasons. For example, sample CRFs may not be available; there may be copyright and use restrictions; there may be multiple, different, versions of the same CRF, and the assumptions underlying the use of a particular CRF may not have been clearly documented and items may not be validated. There are efforts underway to standardize different CRFs for use in clinical research from both researchers and regulatory agencies and organizations. Unfor-

tunately, the majority of questionnaires and CRFs in existence are custom designed by individual investigators and do not achieve wide enough use that they would be expected to become candidates for standardization by different organizations.

### 5.5 Testing and Reviewing CRFs

Once the set of CRFs have been designed, they should be reviewed and approved by the entire study team. It is important to have the clinical study team review and test the CRFs before submitting for approval by the various regulatory committees and boards. Data management team members should review pilot and test data for accuracy (of asking, coding, set up) and analyzability. Each member of the team can evaluate the CRFs from a different perspective and purpose. This will then enable comprehensive feedback and review of the CRFs. Specific things to focus on when reviewing CRFs include:

1. The language on CRF is consistent or understandable by the persons who will be using and completing the CRF.
2. The units of the measures collected are consistent with the units used at different sites.
3. For the different types of CRFs (paper, electronic), the coding must be consistent with the database system that will be used to store the data. For example, it is important to check that the correct choice responses and data types are being collected (e.g., one or two decimal points).
4. The statistical team members can confirm that the appropriate analytical data points are complete and in a format needed by entering pilot test data.

As multiteam involvement is helpful during development of CRFs, testing and review after development can facilitate good CRFs that will work for everyone to increase the usability and data quality and reduce the time needed for cleaning data.

### 5.6 Summary: Well-Designed CRFs

It is important to develop and design CRFs when developing the protocol and make sure that only data items are measured and collected that can contribute to measurement of the outcomes and success or failure of the study. The balance between asking too many questions or too many data items or CRFs collected must be made to avoid unnecessary costs and burdens to the subjects and data managers. Too much and/or too little can both contribute to problems in conduct and performance and analysis of the study. By focusing on the overview of data and CRFs in the study, the CRF layout and ergonomics, the usability and CRF modality, along with appropriately designed measures, and standards for data items and databases, and finally by testing and reviewing CRFs with the entire study team, the result will be well-designed CRFs that can be used to collect the right data efficiently to produce data ready to be analyzed. By addressing these CRF design concepts in the DMSP, this will facilitate collection of clean data and will result in a reduction of queries from data managers to clarify items.

## 6 DATA ACQUISITION AND HANDLING GUIDELINES

### 6.1 Definitions and Overview

Data acquisition is the process by which the data is collected [8]. Data handling is a broad concept that we are defining as the multiple processes and activities that are utilized to transport data throughout the clinical study process [8]. The data-handling process covers the entire life cycle of the data from acquisition to archiving. The DMSP should include study-specific procedures for the data acquisition and collection and data-handling processes all necessary in assuring optimal performance, consistency, compliance, and integrity.

### 6.2 Data Acquisition

Data acquisition is the actual collection of protocol-specific data. This collection process can frequently be one of the most difficult and error-prone aspects of a study. Collection of data can occur by one or more methods, all ultimately being deposited or stored in the CRFs. Some methods by which data are acquired and collected include:

- Electronic or computer based—This can include eCRFs on personal computers (PC), laptops, PDAs, or other electronic devices used for data entry. Electronic data capture (EDC) can also be included in this modality type. This can be defined as automatic data collection by a device. An example is the glucose sensor monitors that automatically create electronic glucose data from the subject.
- Through in-person or telephone interviews.
- Paper-based CRFs—These can be scanned, faxed, data entered into a computer, and converted into electronic documents.
- Study subject self-administered diaries and questionnaires.
- Medical chart review and abstraction.
- Electronic imports or transfers of data from other systems (such as laboratory data).
- Sound based, telephone interactive voice response systems (IVRS) that connect telephone users using speech recognition technologies to computer systems [9].

The acquisition of various data utilizing one or more data collection methods presents a challenge. Incremental and systematic checks and monitoring should occur to determine if the data to be collected actually was collected. In addition, was data collected in the format that was expected. And, finally, was data acquired in the time frame expected.

Traditionally, paper has been used to collect and even to store clinical data for research studies. This is primarily due to the convenience of paper and the familiarity of staff and subjects with paper. The move toward more technical approaches is continuing, and for this chapter we will focus on electronic methods of data acquisition or collection for clinical studies. The different modalities that are used for data

collection into a CRF will have varying features that need to be described and included in a DMSP.

Many vendors and data management organizations are moving toward data entry into electronic case report forms often referred to as eCRFs. There are several methods that eCRFs can use to help ensure accurate data entry. The methods used should be described in the DMSP and can include:

- Use predetermined ranges of allowable and expected values.
- Provide simple data type checks, for example, ensuring valid numbers or dates in numeric or date fields. For questions that require text responses, string length checking and validation.
- Preset list(s) of choice values should be defined when possible for a restricted list of responses allowed for data entry selection.
- Skip inapplicable questions and set default values based on the response to specific question(s).
- Compute values for a question calculated based upon other values entered in the CRF.
- Intraform validation checks—In this case one value on a CRF is compared to previously entered data on the same CRF to see if it meets predefined criteria.
- Cross-form validation—In this situation, checks are made that compare the response to a question on one CRF to response(s) on a previously entered CRF, or other previously entered data.
- Spell checking—The software should perform automatic spell checking for certain text fields against online. Domain-specific dictionaries can be used for free text entry to check for spelling errors.
- Support entry and query of missing and approximate values.

Data acquisition guidelines should be included in the DMSP as these will be specific to the study and sponsor requirements. These guidelines will complement the data entry, validation, transfer guidelines, also included in the DMSP, which are based upon the project's scope of work, the study phase, system capacity, and other requirements. Following the acquisition of data, it is the handling processes that address the movement, the storage, archiving, and ultimate security of the data. Appropriate planning and safeguards must be established to address both paper and electronic data. For example, data handled electronically requires the added assurance that the integrity of recorded data are not altered, erased, lost, or accessed by unauthorized users.

### 6.3 Data Handling

Individual data-handling processes may be applicable for a given environment and study, and the DMSP should incorporate appropriate sections for the processes and procedures used in the specific study. Many of the sections to be included in the DMSP for data handling will require the description of validation and testing of the processes and activities involved along with other appropriate ongoing quality

checks. For example, for studies that use a clinical data management system for managing study data, there should be sections in the DMSP on quality checks to be performed including validation and testing of the system and of software programming created for the study. Other examples include quality checks of the data processing and cleaning activities (discrepancy identification and resolution). This might involve describing the queries and reports that will be used for monitoring data entry and data-cleaning activities such as data entry completion and error rate monitoring, missing data reports, discrepancy, and discrepancy resolution reports. Based upon the different capabilities of the software systems used for the study, all or some of these activities can be performed and described in the DMSP.

A systematic approach to how the study will handle data changes or changes to data collection instruments or CRFs should be included in the DMSP. This approach should be an agreed upon process in affect prior to CRF development that details how changes in data collection instruments or CRF versions will be documented and handled. This section should also describe how all changes to data would be documented and copies of previous data stored in electronic audit trails. There should be included clear documentation in the audit trail of who, what, when, and why the change was made.

More data is now available in an electronic format and may be available for import and integration into the clinical study data management system. Examples include laboratory data and subject diary data. It is preferable for this data to be electronically transferred and imported/integrated into existing data management systems rather than be rekeyed into the system. In this case, it will be necessary to include a section in the DMSP describing the approach that will be used to perform the import and check the results in a standard way. Different studies also may require data to be transmitted to outside agencies. Data transfer (both in and out) may be specified to be compliant with standard formats or special formats required by the receiver. The format or use of data format standards such as HL7 [10] and CDISC [11] should be described in the DMSP along with methods to test and validate accuracy of the transfer(s).

***Security and Privacy of Records*** The DMSP must describe [or reference the institutional standard operating procedure(s)], how the security and privacy of clinical study data will be assured throughout the data management process. This must include both electronic data and other forms of data (paper) if relevant. All the computer systems to be used must have methods to prevent unauthorized access, preserve subject confidentiality, and prevent retrospective tampering/falsification of data. Under the FDA's Title 21 Code of Federal Regulations (21 CFR Part 11) [12], access must be restricted to authorized personnel, the system must prevent malicious changes to research data through selective data locking, and an audit trail must exist. Additionally, compliance with the Health Insurance Portability and Accountability Act (HIPAA) must be assured. The HIPAA Privacy Rule describes how protected health information (PHI) may be used and disclosed. Under HIPAA, research is defined as "a systematic investigation, including research development, testing, and evaluation, designed to develop or contribute to generalizable knowledge" [13]. The conduct of research and clinical data management, in specific, must incorporate appropriate safeguards to assure compliance with this law.

Data systems also need to be adequately backed-up and recoverable in the event of catastrophic system failure. The physical and electronic security must be assured to keep the data both safe and secure, yet it must also be available to authorized users with password and authentication. Users should have assignment of role-based privileges. Database management systems should have the ability to store and/or generate deidentified data for purposes of analysis and data sharing. All systems must have robust electronic audit trails and allow for archiving and data locking.

## 7 SUMMARY

The process of data management is complex, error prone, and vital to reliable accurate data. A data management study plan is a structure by which these complex processes and activities can be organized to assist in a consistent performance of the actions. Beginning the DMSP with a study-specific quality plan and applying the quality processes throughout the entire data management process should result in an increase in the reliability of clinical data, decrease variation and errors, and improve on the timeliness of good clinical quality processes and outcomes that can be defined and measured.

## REFERENCES

1. Bohaychuk, W., and Ball, G. (2001), *Conducting GCP-compliant Clinical Research*, Wiley, New York, p. 127.
2. International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH) (1996), Guidance for Industry E6 Good Clinical Practice: Consolidated Guideline; accessed Feb. 6, 2006; www.fda.gov/cder/guidance/959fnl.pdf; glossary section Ref 1.24, p. 8.
3. International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH) (1996), Guidance for Industry E6 Good Clinical Practice: Consolidated Guideline; accessed Feb. 6, 2006; www.fda.gov/cder/guidance/959fnl.pdf; glossary section Ref 1.47, p. 11.
4. International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH) (1996), Guidance for Industry E6 Good Clinical Practice: Consolidated Guideline; accessed Feb. 6, 2006; www.fda.gov/cder/guidance/959fnl.pdf; glossary section Ref 1.46, p. 11.
5. Donabedian, A. (1988), The quality of care. How can it be assessed? *Journal of the American Medical Association*, 260(12), 1743–1748.
6. Walley, P., and Gowland, B. (2004), Completing the circle: From PD to PDSA, *Int. J. Health Care Quality Assurance*, 17(6), 349–358.
7. Sudman, S., and Bradburn, N. M. (1987), *Asking Questions. A Practical Guide to Questionnaire Design*, Jossey-Bass, San Francisco, pp. 281–282.
8. Society for Clinical Data Management (2005), Good Clinical Data Management Practices, Version 4, October.
9. Interactive voice response systems; accessed Feb. 6, 2006; http://www.iec.org/online/tutorials/speech_enabled/.
10. Health Level Seven, HL7 Standards; accessed Feb. 6, 2006; http://www.hl7.org/.

11. Clinical Data Interchange Standards Consortium (CDISC); accessed Feb. 6, 2006, http://www.cdisc.org.
12. U.S Food and Drug Administration (2006), Title 21 Code of Federal Regulations (21 CFR Part 11); accessed Feb. 6, 2006; http://www.fda.gov/ora/compliance_ref/part11/.
13. Department of Health and Human Services (2003, April 3), Health Information Privacy. Research; accessed February 12, 2009; http://www.hhs.gov/ocr/privacy/hipaa/understanding/special/research/.

# *Clinical Trials Data Management*

**Eugenio Santoro[1] and Angelo Tinazzi[2]**

[1]*Laboratory of Medical Informatics, Department of Epidemiology, "Mario Negri" Institute for Pharmacological Research, Milan, Italy*
[2]*Merck Serono, Global Biostatistics, Geneva, Switzerland*

**Contents**

1 Clinical Data Management Aspects
- 1.1 Regulatory Framework for Clinical Data Management
- 1.2 From Clinical Protocol to Data Acquisition Tools
- 1.3 Database Design
- 1.4 Data Processing
- 1.5 Electronic Data Capture Principles
- 1.6 Data Standards
- 1.7 Infrastructure Requirements
- 1.8 Implementation of Clinical Study Data Management System
- 1.9 Computer System Validation
- 1.10 Future: EHR/EDC Integration

2 Web-Based Clinical Trials
- 2.1 Web-Based Clinical Trials
- 2.2 Tools for Participating in Web-Based Clinical Trial
- 2.3 Why a Clinical Trial Website?
- 2.4 Examples of Clinical Trial Websites and Web-Based Clinical Trials
- 2.5 Advantages and Limitations of Web-Based Clinical Trials

References

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

# 1 CLINICAL DATA MANAGEMENT ASPECTS

## 1.1 Regulatory Framework for Clinical Data Management

In recent years the medical community has been involved in a great debate concerning ethical issues that include confidentiality and privacy, especially for data included in medical records or collected in clinical studies for research purposes. Furthermore, the improvements in information technology (IT), which made full paper replacement possible and new solutions for the electronic management of data available, have led the regulatory agencies to better clarify the requirements for processing personal data. For these reasons, new rules have been identified for the clinical data management process including compliance with the privacy and data protection laws when personal data are processed, compliance with good clinical practice (GCP) [1] when medical products or medical interventions are tested, and compliance with other regulatory directives when medical products are submitted to specific regulatory agencies (Table 1).

**TABLE 1 Regulatory and Guideline References**

| |
|---|
| *International* |
| E6: Guideline for Good Clinical Practice—Consolidated Guideline, International Conference on Harmonisation, EU Implementation CPMP/ICH/135/95/Step5 |
| *European Community* |
| The rules governing Medical Products in the EU, Volume IV, 1998, Annex 11: Computerized Systems |
| Directive 95/46/EC—Data Protection |
| Directive 2002/58—Privacy and Electronic Communications |
| Directive 1999/93/EC—Community Framework for Electronic Signatures |
| *Food and Drug Administration* |
| 21 CFR Part 11—Electronic Records; Electronic Signature; 2003 |
| 21 CRF Part 11—Protection of Privacy |
| General Principles of Software Validation; Final Guidance for Industry and FDA Staff; January 2002 |
| Guidance for Industry: Computerized Systems Used in Clinical Trials; September 2004 |
| *Country (Italy)* |
| DRP nr. 318—Use of Personal Data and Privacy Protection |
| DL 30/06/2003 nr. 196—Personal Data Protection |
| *Other Guidelines* |
| GAMP Forum, Good Automated Manufacturing Practice—Supplier Guide for Validation of Automated Systems in Pharmaceutical Manufacture; December 2001 |
| ACDM/PSI—Computer System Validation in Clinical Research: A Practical Guide; ACDM/PSI; 1998 |
| PICS/S Good Practice for Computerised Systems in Regulated GxP Environment, Pharmaceutical Inspection Co-operation Schema Final Guidance, rev. 2; July 2004 |

***Data Privacy*** People involved in the data management process should be familiar with basic data privacy issues and should follow the principles established by their organization to ensure the privacy of research subjects and the compliance with the GCP and other international and/or local regulations [2].

Data privacy is related to the standards surrounding the protection of personal data, defined as "any information about a person who can be identified directly or indirectly" (e.g., patient names, initials, addresses, and genetic code). The privacy of any subject who participates in a clinical trial must be protected from the ethical and legal points of view. The data should be protected from accidental loss, alteration, and unauthorized access. For this reason, the implementation of appropriate security measures are required.

To guarantee data privacy, personal data must be handled separately from the clinical data and made anonymous. In addition, a written, signed, and dated informed consent should be asked and obtained from the owners of the data. The concept of data privacy is enforced in the GCP: "The confidentiality of records that could identify subjects should be protected, respecting the privacy and confidentiality rules in accordance with applicable regulatory requirement(s)" [1].

Sometimes the identification of an individual cannot be fully masked. In fact, the data management staff usually use several sources of information such as primary medical and hospital records, genetic data, economic data, and adverse drug event reports. In these cases, to ensure proper assignment of data in a clinical database, data collection instruments should be designed with the need for the minimum research subject identifiers (in general, a subject identification number and gender can be used to solve any discrepancies that might arise from transcription errors).

In addition, the application of local laws, such as the Italian one, integrates the statements of the European Union (EU) directive with a technical attachment addressed to the electronic data management [3]. These include detailed requirements related to the use of the user name and password for a secure and diversified data access that is based on the user's profile, and the use of backup and restoration procedures in order to ensure data integrity.

***Good Clinical Practice and Other Directives*** The International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH) Good Clinical Practice Guidelines [1] add an additional layer on the requirements for data management of clinical data (Table 2). One of these is related to the quality control process that should be applied to each stage of data handling in order to ensure that all data have been correctly processed. Another important requirement states that any change or correction to a case report form (CRF) should be dated, signed, and explained (if necessary), and should not obscure the original entry. This must be done by maintaining an "audit trail" and should be applied to both written and electronic changes or corrections. Similar requirements have been proposed by other regulatory guidelines. For example, the Food and Drug Administration (FDA) 21 CFR Part 11 [4] establishes the requirements for electronic records and electronic signatures to be trustworthy, reliable, and equivalent to paper records and handwritten signatures. These requirements, for example, imply data encryption, digital signature standards, and other safety requirements (i.e., device checks).

**TABLE 2 Main FDA 21 CRF Part 11 and GCP Requirements**

| Requirement | Part 11 | GCP |
|---|---|---|
| Validation of Computer System | 11.10 (a) | 5.5.3a |
| Data Protection (to enable accurate record retrieval) | 11.10 (c) | 2.10; 4.9.1; 5.5.3f |
| Limiting Access (limiting system access to authorized people) | 11.10 (d) | 2.11; 5.5.3d |
| Audit Trail (secure, computer generated, time-stamped audit trails) | 11.10 (e) | 4.9.3; 5.5.3c |
| Authority Checks (to ensure that only authorized individuals can use the system) | 11.10 (g) | 2.11; 4.1.5; 4.9.3; 5.5.3e |
| System Documentation | 11.10 (k) | 5.5.3b |
| Record Retention (to generate accurate and complete copies of records) | 11.10 (b) | 4.9.7 |

## 1.2 From Clinical Protocol to Data Acquisition Tools

The clinical data management activities should be planned early in the clinical study process. Ideally, they should occur concurrently with the development and writing of the protocol, when the decisions about the study data to be collected and the type of data flow are usually taken. It is a crucial phase for the development of proper data management tools, and it is therefore suggested that statisticians, clinicians, data managers, study monitors, study coordinators, and database programmers be involved in it. The design of the CRF is the first step in translating the protocol requirements into data and is one of the first results of this multidisciplinary work.

A CRF can be defined as a collection of forms including a series of data items, each aimed at obtaining a single response, and composed by a text (item label) and a response field (item value); while the first represents the question to be answered, the second is the space given to the investigator to record its response [5].

## 1.3 Database Design

Clinicians often have difficulty in deciding what patient information is relevant to a clinical trial. In their current practice, they are able to retrieve a summary of the relevant information from individual patients' charts, but are probably unable to give a whole list of the pertinent data items. When these are identified by the multidisciplinary team, they must be organized in a structured way so that they may be easily collected and analyzed. The CRF structure is then used to design the structure of the database where the patients' data will be stored once entered through a data entry application.

Computer science and various software technologies have revolutionized the way medical information is stored, accessed, and retrieved. Today, the relational database model is often considered the founder of many different types of clinical information systems, providing the closest fit to functional requirements of the protocol [6]. In brief, this model comprises data arranged in "tabular" data structure, consisting of several columns (items/variables) and rows (observations). The key aspects of a relational database are the relationships between tables (entities) through the definition of primary and secondary keys (e.g., the patient's ID to link the patient's

demography data to his adverse events), because they ensure database integrity. This means, for example, that a row cannot be added in the "adverse events" table representing an adverse event for a specific patient, unless a row for the same patient exist in the "demography table." In addition, a table will have the required fields for representing the entities (type of data collected), and each field will contain data of a specific type, such as characters, numbers, dates, and times. The interaction between the case report form design process and the database design process is fundamental and unavoidable: The schema identified to collect the data may influence their meaning and the structure of the database in which they are stored.

***Vertical versus Horizontal File Structure*** In designing the database tables containing data about a specific "entity" (such as demography, laboratory data, etc.), the critical question that must be addressed is the following: For multiple occurrence of data items at different time points, is it easier to handle multiple records or one record with repeated data items? There are two options to consider: vertical and horizontal file structures (Fig. 1). The vertical file structure defines the concept of multiple records, while the horizontal file structure is associated to the concept of repeated data items.

To utilize a vertical file structure, a method to distinguish between the multiple occurrences of the data must be identified. This can be done by using the visit date and the visit time. Therefore, in a vertical file structure, the single record contains information on the patient's measurements at a defined date and time. The advantages of this structure include the utilization of cross-tabulation features for reporting and analysis and the minimization of the amount of missing data and space needed for storage in the database. Several statistical packages such as SAS System require this structure for data processing. In addition, programming efforts are reduced due to the need for processing multiple records instead of multiple fields.

When a horizontal file structure is used, the information collected must be fixed and prespecified. The names of the database fields should have suffixes that help

**Vertical**

| Site | Patient | Visit # | Visit Date | Temp | SBP | DBP |
|---|---|---|---|---|---|---|
| 5 | 1 | 1 | 10/11/2005 | 37 | 120 | 70 |
| 5 | 1 | 2 | 10/12/2005 | 37.1 | 125 | 71 |

**Horizontal**

| Site | Patient | Temp1 | Temp2 | SBP1 | SBP2 | DBP1 | DBP2 |
|---|---|---|---|---|---|---|---|
| 5 | 1 | 37 | 37.1 | 120 | 125 | 70 | 71 |

***FIGURE 1*** *Vertical vs. horizontal data structure.*

the identification of the visit time (e.g., in Fig. 1 SBP1 and SBP2 identify the patients' systolic blood pressure taken during visit 1 and visit 2, respectively). This could be a problem if data for additional visits must be collected. In addition, the limit on the number of fields or in the record length for a single table may introduce other problems that cannot easily be solved. Although this structure allows easy comparison of similar fields over time (e.g., to find the worst value), it is not a recommended approach. Furthermore, it requires more programming efforts without any benefit in program flexibility.

## 1.4 Data Processing

Clinical data management deals with complex medical data. Managing this data requires familiarity with medical terminology, anatomy, physiology, and pharmacology, as well as practical knowledge of how data are collected in the health care setting and documented in medical records.

The data processing process includes the following steps: data entry, data validation, data modification, medical coding, and database locking [7].

***Data Entry Methods*** When a paper-based CRF is used to collect patient data, the investigator is asked to access the source patient's documentation (e.g., the medical record), fill in the study forms, and deliver them to the coordinating center (or clinical data center), usually after a data-monitoring process is carried out. Once received, the CRF undergoes an initial inspection and then is entered into the clinical database. Two options are usually available for data entry: single data entry and double data entry. With double data entry two people enter the same data independently. Any discrepancies between the first and second entry are usually solved by a third person (e.g., the data manager in charge of the study) or by the person performing the second entry (interactive or online verification). The main regulations, such as those of the FDA and ICH, do not formally require double entry or any other specific data entry process, although some have questioned the need for duplicate data entry [8]. In general, the single-entry process with proper manual review, or with double data entry of key data (e.g., the main study endpoints) will work better than a sloppy double-entry process.

***Data Cleaning and Checking*** Data are usually collected and entered in the database following detailed rules and conventions that govern how the data are to be recorded (i.e., CRF completion instructions, data entry guidelines, data entry conventions, etc.). However, many errors can be made (voluntary or involuntary) during the data collection and data entry processes including misspelled names, duplication of digits, partial data collection, and inconsistencies between forms.

For this reason, a data cleaning (or data validation) process should be defined. The data-cleaning process is based on a list of activities that are planned to assure the validity and accuracy of the data, including a manual review of the data before their entry, and aggregate descriptive statistics that reveal unusual patterns in the data. However, the most powerful tool provided by the data-cleaning process is the automatic data-checking process to identify errors, protocol violations, and data completeness, inconsistencies, and duplication.

```
CONTROL ON PATIENT DEMOGRAPHY
1. Age <> Missing AND between 18-75
2. Sex <> Missing AND In(1,2)
        ▪  If SEX=2 THEN Inclusion Criteria nr. 3 (Pregnancy) <> Missing
DRUG ADMINISTRATION
1. (Actual Dose/Intended Dose) between 0.95 and 1.05 (5% dose modification
      allowed)
```

***FIGURE 2*** *Example of a data validation plan specifications.*

Example of data-checking include range checks on the variables that may reveal values that fall outside the accepted limits (e.g., age, for a specific eligible criteria of the protocol, must range between 18 and 75), and cross checks between variables, which may reveal inconsistent answers (e.g., if a patient is dead the date of death is required). Other additional types of checks such as checking whether mandatory answers are not missing, checking for invalid dates and invalid date sequences, checking that complex multifile (or cross panel) rules have been followed (e.g., if a specific adverse event occurs, other data such as concomitant medications or procedures might be expected), and the comparison of the values entered to lists of valid patterns of value, either text/character (e.g., the smoking habits could be coded as "smoker," "nonsmoker," and "ex-smoker") or numerical (e.g., the same habits could be coded as 1, 2, and 3) are usually programmed and performed.

A complete validation plan for the data cleaning should be defined, for example, using a pseudocomputer language which will be easily understood and translated into a computer language (Fig. 2).

The data-checking process may be performed as a batch or as an interactive process. In the first case, data checking takes place after data entry. In the second case, it is performed during the data entry, and the data entry person receives a warning message immediately if an error is made.

***Data Modification*** Data may be changed as a result of a data-cleaning procedure. In this case, both the site and the data center must retain a record of all changes to the data. Data changes (with the new and the old values) are usually recorded and documented through a data (or query) clarification form that is signed by the site investigator and kept with its original study record. Under some circumstances, data-cleaning conventions (often called self-evident changes) may specify data that can be modified without a site's acknowledgment. Examples include appropriately qualified personnel correcting obvious spelling errors and converting values when the units are provided. For this reason it is important for the site to receive and maintain a copy of each version of such data conventions.

However, an audit trail should always reflect all deletions, updates, or additions to data that are carried out after the initial entry. In addition, it should include the date and time of the changes, the reasons for the changes, and the name of the user who was in charge of making them. An electronic audit trail can provide a clear documentation of this information (Fig. 3).

| Pt# | Vis# | User Id | Timestamp | Item | Old value | New Value | Reason for Change |
|---|---|---|---|---|---|---|---|
| 1001 | 1 | DM_01 | 12/05/2004 11.46.52 | AEOUT | 1 | 2 | Query following Medical Review |
| 1001 | 1 | DM_01 | 12/04/2003 10.32.46 | AEOUT | | 1 | |

***FIGURE 3*** *Example of an electronic audit trail. An adverse event occurred at visit 1 in patient 1001; the corresponding data was initially entered on 12/04/2003 by user DM_01. Following the data validation process, a query was issued; the investigator reply implied the change of the adverse event outcome (Item AEOUT) from 1 to 2; the modification was performed by user DM_01 on 12/05/2004.*

***Medical Coding*** Wherever possible, data should be collected in a coded format, while the text format should be used to collect summaries and data that do not need to be codified. This makes data entry and the search and data analyses procedures easier. Free text can be coded using an ad hoc dictionary (with a simple list of codes) or a standard dictionary (with thousands of entries).

Data coding can also avoid possible misinterpretations. For example, consider patients with "abnormal transaminases." The investigators from different sites could indicate this condition in the CRF with the phrases "has ALT," "has SGPT," and so forth. When data are analyzed, if we need to identify all cases of "abnormal transaminases," we should search for all possible combinations of words related to the abnormal transaminases concept.

This adds further implications in data analysis. Consider, for example, the statistical tables illustrated in Figure 4 analyzing the safety profile of a particular medical intervention, where worst (maximum intensity) adverse event toxicity by patient is reported. If we consider the first table we may conclude that seven patients had abnormal transaminases of grade 1 or 2. Instead, if we consider the second table based on the same set of patients, we may conclude that five patients had abnormal transaminases. The discrepancy is due to the investigator's use of different terms to identify the same type of adverse event occurring in the same patient (this occurs in two patients as explained in Fig. 4).

The coding process can be automatically performed during data entry by matching the text collected on the CRF to the terms included in the standard dictionary.

Several medical dictionaries are currently available. Among them, the *Medical Dictionary for Regulatory Activities* (MedDRA) [9] has become the standard dictionary for coding a patient's medical history, medical and surgical procedures, and adverse events. In addition, it includes the standard terminology used by regulatory agencies and biopharmaceutical industries within the ICH regions, through all phases of clinical development (including postmarketing). Other well-known medical dictionaries are the *World Health Organization Drug Dictionary* (WHODRUG) [10] and the *Systematized Nomenclature of Medicine—Clinical Terms* (SNOMED-CT) [11], which are used to code medications, the International Classification of Diseases (ICD) oncology for tumor classification, the ICD-9 for classifying morbidity and mortality, and the National Library of Medicine's Unified Medical Language System (UMLS) for classifying general medical terms.

| Investigator term | Worst CTC Grade by Patient | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | 1 | | 2 | | 3 | | Any | |
| | N | % | N | % | N | % | N | % |
| ALT | 1 | 5.2 | 2 | 10.5 | | | 3 | 15.7 |
| SGPT | 2 | 10.5 | 2 | 10.5 | | | 4 | 20.3 |
| .... | … | … | … | … | … | … | … | … |

| MedDRA Primary Preferred Term | Worst CTC Grade by Patient | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | 1 | | 2 | | 3 | | Any | |
| | N | % | N | % | N | % | N | % |
| Alanine Aminotransferase | 1 | 5.2 | 4 | 21.1 | | | 5 | 25.2 |
| .... | … | … | … | … | … | … | … | … |

| Pt# | Vis# | Event Description | MedDRA Primary Preferred Term | Max CTC Grade | Max CTC Grade per patient according to MedDRA term |
|---|---|---|---|---|---|
| 1 | 1 | ALT | Alanine Aminostransferase | 2 | 2 |
| 1 | 2 | SGPT | Alanine Aminostransferase | 1 | |
| 2 | 2 | SGPT | Alanine Aminostransferase | 2 | 2 |
| 3 | 1 | ALT | Alanine Aminostransferase | 1 | 2 |
| 3 | 2 | SGPT | Alanine Aminostransferase | 2 | |
| 4 | 1 | ALT | Alanine Aminostransferase | 2 | 2 |
| 5 | 1 | SGPT | Alanine Aminostransferase | 1 | 1 |

***FIGURE 4*** *Implication of medical coding on statistical analysis.*

***Database Locking*** The study databases must be locked to ensure their integrity for the generation of results, analysis, and submission to the regulatory agencies. Locking a study database is fundamental to prevent inadvertent or unauthorized changes once the final analysis and reporting of the data have begun. Furthermore, database closure is a critical issue in preserving the integrity of blinded randomized clinical trials when the blindness needs to be broken. Therefore, any clinical trial must have a well-defined process for closing its database and clear change-control procedures for reopening the database, if necessary.

Before locking a database, a database quality control should be considered in order to verify that study procedures for collecting and managing the data were correctly applied. This is usually performed by checking the contents of the database against the paper CRF in a sample of patients (e.g., 10% of CRFs).

## 1.5 Electronic Data Capture Principles

Electronic data capture (EDC) is the process of the collection of data into a persistent electronic form: This includes data entry (e.g., keyboard EDC, pen-based systems, voice recognition) and automated (or direct) data acquisition (e.g., bar code scanners, blood pressure cuff devices) [12].

The EDC processes emerged in the 1970s but languished for 20 years without having significant impact on the clinical trials arena. However, in the 1990s the development of tools for clinical trials research became more focused. With these applications data were captured and entered directly into the PC client at the investigator sites where the database application was installed; the data stored at the site were then periodically transmitted to the central server located at the study data

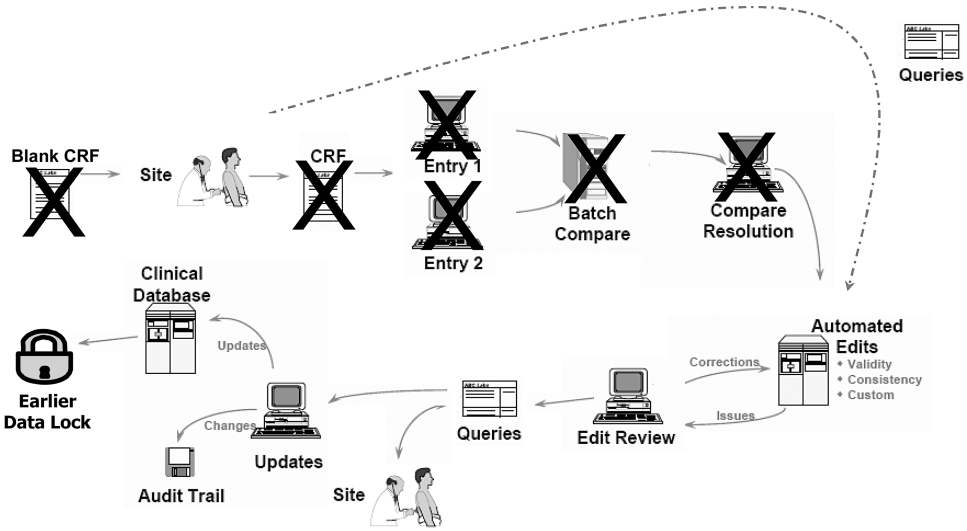


***FIGURE 5*** *Impact of EDC adoption in clinical data management. EDC adoption made the data management process faster by removing time-consuming steps in clinical data flow from source data to clinical trial data.*

center (Fig. 5). Other, more sophisticated EDC systems have been developed during recent years, including those using the Internet and web technology (see following sections), and today they are commonly used by many research organizations to conduct clinical trials. The regulatory agencies themselves are today ready to accept the submissions for drug registration in which EDC tools are used.

The adoption of an EDC system has a great impact in the clinical data management process and introduces new features:

- No formal data entry exists because source data are collected and entered directly into the clinical database without first being captured onto paper; this means that transcription errors are eliminated and a double data entry is therefore not required.
- The data cleaning and editing can be performed during data entry (through the online checks), and the investigator can immediately clarify any discrepancy.
- The source data verification process cannot be eliminated; however, the number of queries issued for the clinical monitors, as well as the time they usually spend at the investigator's site, are reduced.
- The type of training and skills required of the data entry staff are different from those required when a paper-based system is used; in this case, in fact, the data entry personnel should receive training on the specific system being used in the study as well as on study-specific issues (i.e., eCRF completion guideline).

For these reasons, designing a good EDC system needs special attention to make moving from paper to EDC systems and computer programs efficient while maintaining data integrity.

### 1.6 Data Standards

One of the main drawbacks to the adoption of the EDC systems has been the proliferation of software to capture trial data; moreover, without the ability to share data across systems, the value of the EDC systems has been somewhat restricted to clinical researchers.

To solve this issue, the representatives from pharmaceutical industries, academic research institutions, health authorities and other research entities [i.e., clinical research organizations (CROs)], met in 1997 to develop a system of shared standards, known as CDISC (Clinical Data Interchange Standards Consortium, http://www.cdisc.org). Among the CDISC specifications, the Operational Data Model (ODM) describes a standard model to combine both the study data definition and the actual subject data. This standard specification allows two different systems to communicate and share the data if the source system can produce CDISC ODM format and the receiving system can read CDISC ODM format. The portability and sharing of the data is made possible by the use of the eXtended Markup Language (XML), which has been adopted by CDISC to describe the data hierarchy (see example in Fig. 6).

Other CDISC models have been developed. One of the most interesting is the Study Data Tabulation Model (STDM) through which most common clinical data domains, variables, and their attributes are defined (i.e., name, type, length, standard definitions, code lists, etc.). The adoption of such a standard for data structures and conventions permits the sharing and application of the same, unique data structure and data-cleaning procedure to several different studies; moreover, pharmaceutical companies dealing with regulatory submissions or institutions running meta-analysis projects would be able to better integrate data from multiple studies.

In July of 2004 the U.S. Food and Drug Administration (FDA) announced the adoption of the CDISC STDM as the standard format for submitting study data for

```
<ClinicalData StudyOID="S054T231" MetadataVersionOID="101.01">
 <SubjectData SubjectKey="1000" TransactionType="Insert">
  <StudyEventData StudyEventOID="101.Screen">
   <FormData FormOID="101.DEMOG">
    <ItemGroupData ItemGroupOID="101.DM">
      <ItemDataString ItemOID="101.USUBJID">002</ItemDataString>
      <ItemDataString ItemOID="101.SEX">F</ItemDataString>
    </ItemGroupData>
   </FormData>
  </StudyEventData>
 </SubjectData>
</ClinicalData>
```

***FIGURE 6*** *Example of an ODM XML file. An XML file portion describing demography data (DM) of a "Female" patient enrolled in the study S054T231 with the code 002.*

registrative purposes; with such a decision the FDA intends to reduce the time for data evaluation and avoid the reorganization of large amounts of data submitted in varying formats.

### 1.7 Infrastructure Requirements

Setting up an organization for clinical trials data management (i.e., data center, central coordination center) requires the evaluation of a number of issues including the selection of the optimal staff for supporting the various steps of clinical data management flow (i.e., data manager, database programmer, medical coding expert, etc.). Concerning the IT aspects, an appropriate hardware and software selection should be taken into account. The key aspect to consider is the physical security of original data sources (e.g., case report forms, electronic data files, and other original data documents), including the system used to store them. Original paper and electronic documents (servers) should be warehoused in secure rooms, or file cabinets, with controlled access. Direct access to database servers should be restricted to individuals who are responsible for monitoring and backing up the system; all other access to database servers should be controlled by logical security and should occur across a secure network protected by password access and appropriate system controls. Mechanisms should be implemented to capture and prevent unauthorized attempts to access the system (i.e., firewall); if such an attempt takes place, the administrator should be notified immediately.

The investigator sites should be considered part of the infrastructure, especially if data are stored locally at the study site before being sent to a central server (as in the case of use of "offline" systems).

### 1.8 Implementation of Clinical Study Data Management System

An organization, either a company or academic research institute, may decide either to develop an ad hoc system or acquire a commercial clinical study data management system (CSDMS). Both options have some drawbacks: While developing an ad hoc system requires more effort in the system validation process and additional dedicated computer staff, buying an existing commercial system may require a lot of time to choose and evaluate it among many candidates available in the market, and, at least at the beginning, a lot of money.

The choice of the most appropriate CSDMS should take into consideration several aspects, including the availability of tools to compare double data entry sources, design user-friendly data entry forms, easily program checks for data entry validation, support the generation of ad hoc queries, and handle missing values. In addition, multiuser access regulated by user name and password, and the availability of tools to transfer data to other software packages (such as SAS System for performing statistical analysis), to simplify the coding process, to implement data dictionaries, and to perform the data backup and recovering, can be considered good add-ons.

In recent years the market has seen the launch of several solutions. These software programs help users in designing a complete data management system, without having to know any concept of data models. Oracle-based applications, such as Oracle Clinical RDC and PhaseForward Inform/Clintrial, represent the bigger portion of the market and, due to their high costs, are suitable for an environment

in which many trials are conducted. However, a number of alternative solutions, including free open-source CSDMS such as TrialDB (http://ycmi.med.yale.edu/trialdb) and Open Clinica (http://www.openclinica.org) are now available [13, 14].

### 1.9 Computer System Validation

The GCP requirements emphasize the importance of validation of systems, process, and data. In particular, they ask to "ensure and document that the electronic data processing system(s) conforms to the sponsor's established requirements for completeness, accuracy, reliability, and consistent intended performance (i.e., validation)" [CFR GCP 5.5.3 a]. Similar requirements have been proposed by other regulatory guidelines. For example, the FDA 21 CFR Part 11 [4] emphasizes the need for a validation approach for any computerized system that is used to store electronic records or signatures so that they can be trustworthy, reliable, and essentially equivalent to paper records and to a handwritten signature (Table 2). In this context, the term "computer system" is used to describe the combination of hardware (servers, local area network, client PCs), software (operating system and software applications), procedures, documentation [standard operating procedures (SOPs), guidelines, manuals], and people involved in the process.

Data management software purchased off the shelf should have been validated by the vendor who originally developed it. These validations are usually referred to as "design-level validation" and do not need to be repeated by the end user.

However, the documentation of the design-level validation specifications and testing should, ideally, be available; it should ensure the completion and documentation of functional level testing, and include the documentation of the effect of any known limitation, problem, and defect on functions used for the study [15].

Additional validation activities need to be performed by the customer before the system can be used, including testing the data entry screens to ensure that the data are correctly mapped into the database structure, testing the data verification functions, the validation of any generic integrity constraints or data-checking routines running during data entry, and testing the audit trail and import/export (from/to other data formats) procedures.

### 1.10 Future: HER/EDC Integration

The electronic health records (EHR) systems usually include most of the data that are to be collected in a clinical study. Some examples are the demographic data, the medical history, the clinical events, and the concomitant treatments. For this reason many researchers have proposed to consider them as a primary source of information from which to automatically extract, when needed, the data requested in a study protocol [16–19].

However, this solution presents some limits. For example, some specific information related to the study treatment may not be included in the patient's EHR or may not be properly collected. Similarly, some specific laboratory tests may not be usually performed. In addition, patient confidentiality must be addressed and correctly ensured.

Although some organizations have experimented EHR/EDC combinations, this technology is still young, and, as of today, solutions for health care providers and

hospitals are not available. Some projects to integrate the EHR and the EDC systems are still ongoing. The most important is the BRIDG (Biomedical Research Integrated Domain Group, http://www.bridgproject.org) project, which is a collaboration between CDISC, representing research data standards, and HL7 (Health Level Seven, http://www.hl7.org), representing health care data standards.

## 2 WEB-BASED CLINICAL TRIALS

### 2.1 Web-Based Clinical Trials

Telecommunication technology for the management of clinical trials has been used since the early 1990s when PC/modem-based randomization and data-monitoring systems were developed. One of the first applications is the system developed by the Gruppo Italiano per la Sopravvivenza nell'Infarto Miocardico (GISSI) for the management of the GISSI-3 trial, an Italian multicenter large-scale clinical trial in patients with myocardial infarction [20]. A computer/modem-based system allowed investigators from 200 coronary care units to randomize 20,000 patients in the trial using an automated randomization procedure running on a 24-h basis, and provided study reports and reminders [21].

Since from the mid-1990s the Internet and the World Wide Web have been proposed by several pharmaceutical industries, CROs, and international research groups as tools to support clinical research. Speed in communication, strong interaction among people involved in trial conduction, cost reduction, data quality improvement, and the use of simple and standard tools (such as the browser) to interface with study databases are the key elements for the Web's success [22–24].

### 2.2 Tools for Participating in Web-Based Clinical Trial

A potential investigator center who wishes to participate in a Web-based clinical trial (or e-trial) needs a personal computer, a Web browser, an electronic mail system, and access to the Internet via an Internet service provider (IP). Sometimes a printer is needed for printing documents (such as the reminder of the randomized treatment allocation) that need to be archived in the patient's file. Standard software such as Acrobat Reader may also be needed to read documents (such as operating manuals or CRFs) distributed in pdf format by the coordinating center.

A specific release of a browser (Internet Explorer 6.0 or 7.0 are more frequently suggested for this use) is often requested, as well as setting it up to accept cookies for the user's identification by the coordinating center's server.

For security and data safety reasons an Internet connection with a static IP address is suggested because this is one of the best methods to identify the computer of an investigator participating in a trial.

### 2.3 Why a Clinical Trial Website?

There are many reasons for a clinical trial to have a website. First, a clinical trial website provides trial personnel (coordinators, investigators, monitors, sponsors, committee members) with a powerful communication, organization, and monitoring

tool as well as with tools for the decentralization of trial activities (e.g., remote randomization and data entry). Second, the use of such a website can reduce research costs and the time required for trial completion. In addition, the remote data entry allows the building of a more accurate study database because data are entered at the same place where they are collected.

A clinical trial website is also an ideal vehicle for the dissemination of findings and information related to the trial or similar trials conducted in the past by the same group. For these reasons, a clinical trial website is developed mainly for multicenter and international clinical trials, and for those carried out by pharmaceutical companies.

***Communication among Trial Personnel*** A study website can be used to provide the investigators with secretarial support through automatic trial report generation and delivery (Table 3). Examples of online reports are the list of randomized patients, the visit calendar, the list of patients lost to follow-up, and the list of CRFs and queries that are still outstanding [25, 26]. Automated tools may help investigators to organize their work. Web-based systems developed by some study groups include tools to handle the study drugs (how and when new drugs should be ordered, how drugs dispensed to patients should be registered in the database, how to handle the drug inventory, and when expired drugs should be destroyed) [27, 28] and tools to download study materials (protocol, study CRFs, and informed consent forms)

**TABLE 3 Examples of Online Reports and Information Available on a Study Website**

| Information and Reports | Examples |
|---|---|
| *For Site Investigator* | |
| Reports on the study's progress | List of randomized patients |
| | List of planned visits |
| | List of patients lost to follow-up |
| | List of outstanding CRFs/queries |
| | List of uncompleted CRFs |
| Reports about the management of the study drug | Drug inventory monitoring |
| | List of drugs dispensed to the patients |
| | List of expired drugs |
| Study materials | Informed consent form |
| | Study protocol |
| | Case report forms |
| | Study brochure |
| *For Study Coordinating Center* | |
| Geographical distribution of enrolled patients | Frequency distribution by center |
| | Frequency distribution by state |
| | Frequency distribution by country |
| Geographical distribution of recruited centers | Frequency distribution by state |
| | Frequency distribution by country |
| Epidemiological data of enrolled patients | Frequency distribution by gender |
| | Frequency distribution by age |
| | Frequency distribution of the study drug within specific subgroups of patients |
| Statistics on participating centers | List of the centers' quality indicators |

[26]. Automatic and daily procedures can aggregate data that can be hosted on the study website for the study coordinating center staff's use and to plan new strategies to improve data quality, patient recruitment, and center performance. These include epidemiological patient data, geographical distribution of patients enrolled in the trial, and information about the quality of data provided by the participating centers.

A directory on the website of the investigators, committees, sponsors, and monitors with their email addresses, as well as a directory of participating centers and regional coordinators, can also help improve communications, mainly in multicenter large-scale clinical trials [22].

The clinical trial website may be used as a virtual community in which investigators are continuously updated with the latest news on their trial. For this reason a news section of the website can provide information on trial status, trial newsletters, notification of national and international investigator meetings, reports from study committee meetings, results of similar, concomitant trials, and answers to common questions regarding the application of the trial protocol.

***Dissemination of Study Information to Public*** A study website can provide detailed information about the clinical trial presented in a way that the general public can understand. They may include a geographical distribution of the study participating center and a summary of the background, aim, and design of the ongoing trial. The entire protocol (apart from any confidential aspects) can also be made available. Even if some protocols are already included in the public registries of clinical trials such as the National Library of Medicine's ClinicalTrials.gov (http://www.clinicaltrials.gov), in a clinical trial website more detailed information is provided. This information is useful particularly for those clinical trials allowing online recruitment by patients themselves. In this case the website provides the patients with tools to screen their eligibility, collect their data for enrollment into the study, and download the informed consent form.

***Online Randomization*** Randomization is the method of randomly dividing subjects into two or more groups. In the case of two groups, one is allocated to the trial (tested) treatment and the other one to the standard (current) treatment. It is necessary to ensure that any baseline differences between groups are due to chance alone and to prevent selection bias. Several methods for randomizing have been used over the years, including minimization, biased coin, stratification, permuted block, and random number generators on computers. Web-based randomization systems allow investigators to directly include patients in a trial 24 hours a day. They are particularly useful in multicenter clinical trials where a central coordinating center serving 24 hours a day as a randomization center may not be feasible. Web-based randomization systems utilize the client/server technology and may be used directly by the investigators through the browser [29, 30]: The client software running on the investigator's PC provides a friendly interface for data collection; the server software (installed on the server located at the randomization center) processes and archives data and selects treatment allocation. Data checking is performed by the server software or at the client site. A typical online randomization system follows the steps illustrated in Table 4 to randomize a patient [21]. In particular, the checking of data validity and consistency, and the checking of the eligibility criteria may

**TABLE 4 Steps for Typical Online/Web-based Randomization**

| |
|---|
| Automatic identification of the investigator and/or the participating center |
| Entry of the patient's recruitment data (e.g., patient's initials, gender, age, systolic blood pressure, and those data useful to check if a patient can be enrolled in the trial) |
| Automatic data checking and validation |
| Automatic check that the trial's eligibility criteria are met |
| Running of the randomization algorithm |
| Data incorporated in the central database |
| Notification to the investigator of treatment allocation and patient identification code |

reduce the risk of protocol errors and of noneligible patient enrollment. Web-based randomization systems are usually integrated in the study website [25] and developed in-house by the coordinating center staff. As an alternative to this service there are several online randomization programs (some free of charge and some commercial) that can generate random allocations [31].

***Online Data Collection and Validation*** A clinical trial website often includes a Web-based data entry system that allows investigators to enter clinical data online as soon as they are available. Remote data entry to a central database is particularly useful in multicenter clinical trials where participating centers can be geographically separated by great distances across cities and countries [25].

With respect to traditional paper-based clinical trials, data comes directly from the investigator who is responsible for entering them in the data entry system. In this way steps in data collection are reduced, paper source documents are eliminated, and the quality of collected data is improved thanks to a real-time data validation.

The investigator is requested to fill out several electronic forms (often in HTML format) that have a layout and contents that are not too different from those of the traditional paper-based case report. Client/server technology is often used to develop Web-based data entry systems. The architecture of typical Web-based data entry systems is illustrated in Figure 7. The client application runs on the investigator's computer with the application itself running on a central Web server. The browser interacts with the Web server to submit data collected through a Web page; the Web server sends data to the database server where they are saved. Similarly, if the investigator wants to access data saved in the database server, a request is submitted to it through the Web server, which formats the database query results into HTML and sends it back to the browser as a Web page.

One of the main advantages of remote data entry is the possibility to perform real-time data validation (also called "online edit checks") before they are saved into the database server. When an invalid (or incompatible) entry occurs, the investigator is alerted and invited to make the appropriate corrections. The real-time data validation includes the edit checks already illustrated in discussion on data cleaning and checking in Section 1.4: checks on missing data, checks on answers to closed questions (e.g., those with "yes" or "no" as possible answers), and checks on values that must be comprised within a specific range, within a list of (pre)codified answers, or within a medical standard dictionary. Real-time data validation could also be applied to several data fields in order to perform a cross check of data entry. For example, the investigator recording the date of a serious adverse event (e.g., a myo-

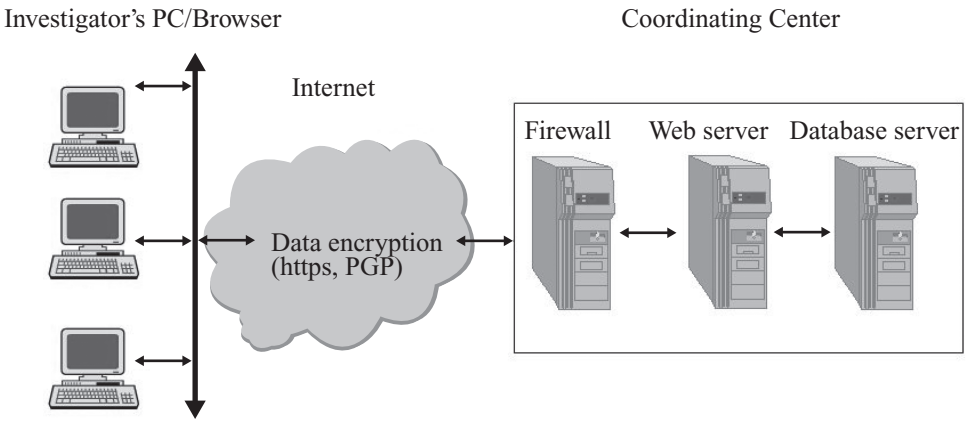


***FIGURE 7*** *Architecture of typical Web-based data entry system.*

cardial infarction) and entering a date that comes before the start of the trial could be prompted immediately, allowing for immediate correction.

Real-time data validation is usually performed on the Web server through Java or Microsoft. NET, which allows for a strong interaction with each data field. In other cases, the data validation process is performed at the end of the data entry process ("offline edit checks") with an automatically generated summary of all the invalid entries that is emailed to the investigator or posted on the trial website for future reminder.

***Data Protection and Security Issues*** Security is a key issue when developing a Web-based clinical trial [22, 32, 33]. This issue becomes crucial and requires more attention when a Web-based data entry system is used to allow remote entry in a centralized database of the patients' sensitive information. Patient's (and investigator's) data that are collected through an electronic form or sent by electronic mail must be protected against any kind of interception by unauthorized users. Similarly, patients' data (and the entire database server) need protection from unauthorized access (by Internet or intranet users) once they are stored on the central database in order to avoid a fraudulent use of them or an invalidation of the trial data and results.

To address the issue of secure Internet transmissions, encryption algorithms are usually used for the data exchanged by electronic mail or data that transit on the Web during a connection between a PC/client and a Web server. The data are coded so that only the user that has the decryption key can read and interpret them. The Pretty Good Privacy (usually available among the facilities provided by many electronic mail software) is one of the most used methods for electronic mail data encryption. On the other hand, a secure Web connection is usually provided by systems such as secure socket layer (SSL) and secure hypertext transfer protocol (S-HTTP). These tools are the same usually used to ensure secure transmissions for the applications of banking and trading online. The last versions of Web browsers support these protocols and so they can be used for a connection to a secure Web server (URLs of secure Web servers start with "https://" instead of "http://").

To protect the database server from unauthorized access, firewalls (a mix of hardware and software that is located between the Web server and the Internet and that protect the local area network where the server is hosted) and tools to check the IP address of the user's computer are used. A login and a password are also assigned to each user to enter the website for data entry and editing. Based on these data, a user's profile can be created and used to enable the user (investigator, monitor, staff of the coordinating center) to enter only the website's sections for which he or she is authorized. Sometimes, a more sophisticated user authentication is reached by using digital signatures.

In addition, the database server must be placed in a secure location and antivirus and updated antivirus software must be installed on each computer on the network. When data are particularly important, experts suggest implementing a backup system (with hardware and software components) that could replace the original one in case of failure.

### 2.4 Examples of Clinical Trial Websites and Web-Based Clinical Trials

Several clinical trial web sites have been developed in the last few years. Some, such as the GISSI website (Gruppo Italiano per lo Studio della Sopravvivenza nell'Infarto Micocarico Acuto, http://www.gissi.org) [20], ALLHAT website (Antihypertensive and Lipid-Lowering Treatment to Prevent Heart Attack Trial, http://allhat.sph.uth.tmc.edu) [34], and STAR website (Study of Tamoxifen and Raloxifene, http://www.nsabp.pitt.edu/STAR/index.html) [35] concern large-scale multicenter (and international) clinical trials and are used to disseminate information to the public related to ongoing or concluded clinical trials. Others, such as the Southwest Oncology Group's SELECT website (and Vitamine E Cancer Prevention Trial, http://www.crab.org/select) [26], provide online patient enrollment, randomization, and reminder notices to the participating centers.

Some have been designed to carry out all the aspects of a clinical trial in ophthalmology [36], orthopedics [30], obstetrics [37, 38], urology [39], and in other medical fields [40]. At present, the most important experience is the INVEST (International Verapamil/Trandolopril Study, http://invest.biostat.ufl.edu) study, a phase IV, international, randomized clinical trial conducted by the University of Florida on the efficacy of the calcium antagonist strategy versus the β blocker strategy for the treatment of hypertension in patients with coronary disease [25, 28, 41]. The trial is the first trial fully conducted online. Investigators from 869 sites located in 14 countries have used the clinical trial website to randomize 23,000 patients between 1997 and 2000, to remotely collect and enter patients' clinical data and adverse reactions, to handle the drug inventory at each site, and to disseminate reports and reminders to all the investigators, monitors, and member committees of the trial. Most of the tools described in this chapter have been used to protect the database server and to be compliant with the GCP, including a firewall, user IDs and passwords, data encryption, electronic signature, and the duplication of critical services to be used in case of system failure. The use of the remote data entry, the automatic data validation, and the automatically generated and delivered reports have reduced the time for the completion of trial and the total trial costs.

Most of these Web-based clinical trial systems are developed "in-house" using the current technology available on the market (database management systems,

HTLM editors, Java, ASP, PHP, etc.). In order to reduce the amount of memory needed to store the data, some of them [33, 42] used the Entity-Attribute-Value (EAV) model [43, 44] to design the database. Others also allow an offline use (e.g., without an Internet connection); in this case a Web-based stand-alone application is installed on the investigator's personal computer, and a data synchronization engine permits the centralized and the local databases to be kept up-to-date during a daily Internet connection [42, 45].

Further, a number of clinical trials are usually carried out by pharmaceutical industries and CROs through the use of Web-based clinical trial systems. In these cases the Web version of commercial clinical trial information systems (such as Oracle Clinical and Clintrial) are usually used.

### 2.5 Advantages and Limitations of Web-Based Clinical Trials

Clinical trial websites may provide a powerful means to conduct clinical trials [46]. Investigators can download study materials, interact more frequently with researchers, and have their problems solved in real time. They can use automatic tools to schedule their work or to enter the clinical data directly. When remote data entry and real-time data validation are implemented, a clean database can be available in a short time for statistical analyses, with reduction of transcription errors and missing data, elimination of data entry costs and reduction of printing costs. For example, researchers of the INVEST study found an 80% reduction in monitoring costs and a 50% reduction in total trial cost [25]. Other cost reductions may be obtained with the implementation of automatic randomization systems and Web-based procedures for organizing drug distribution and accountability, which would allow investigators to order new drugs only when necessary and to destroy them when expired. However, the main advantages in using clinical trial websites are the availability of a friendly and homogeneous interface (the World Wide Web) and of standard software (TCP/IP, browser, electronic mail, Acrobat Reader), which do not require training time, and the ability to centralize study information and coordinate multiple trial processes in real time. Other advantages include security and backup of a whole trial at a single location, simplified data monitoring, and dissemination of study information and results to the public and the scientific community.

Some limitations must be outlined. First, clinical trial websites need the implementation of further security tools (with respect to paper-based or electronic, but not Web-based, clinical trials) in order to prevent unauthorized users from accessing the system and the patient data. This issue is perceived as a threat on data confidentiality by investigators, study centers, and patients and may inhibit them from participating in a Web-based clinical trial. In addition, setting up and maintaining a Web-based clinical trial system requires experienced computer professionals, and further hardware and software to duplicate the key functions in case of system failure. This is particularly important in case of the use of randomization and data entry systems and for the database backup and restoring procedures.

Another limit is the availability of Internet connections. At present, these are still a problem, especially in developing countries where communication facilities are rare. The problem exists also in developed countries, however, where direct connections from hospitals or medical centers are not always available where data are collected.

Other limits include the reluctance on the part of investigators to spend time entering information (in particular when alternative paper-based methods to provide the same data are available), lack of live support personnel (that may lead to the loss of some patients), and high setup cost. Furthermore, every step of the data entry process may be protracted in certain hours of the day due to peaks in Internet traffic. For this reason, many trial groups suggest that Internet connections should be provided by reliable Internet service providers that can assure easy and fast access to the Internet.

Finally, a potential selection bias in patient recruitment should be considered if Internet access is one of the criteria to recruit centers in a trial.

## REFERENCES


1. Guideline for Good Clinical Practice (2003), Consolidated Guideline, International Conference on Harmonisation, EU Implementation CPMP/ICH/135/95/Step5; available at http://www.ich.org/LOB/media/MEDIA482.pdf, accessed Jan. 9, 2008.
2. European Community (1995), Directive 95/46/EC of the European Parliament and the Council of 24 October 1995 on the Protection of Individuals with Regard to the Processing of Personal Data and on the Free Movement of Such Data, Brussels, Belgium, European Community Commission; available at http://ec.europa.eu/justice_home/fsj/privacy/index_en.htm; accessed Jan. 9, 2008.
3. Italian Law DL196/2003 (2003), personal data protection.
4. Food and Drug Administration (2003), 21 CFR Part 11, Electronic Records; Electronic Signatures; Final Rule, Rockville, MD, *Fed. Reg.*; available at http://www.fda.gov/cder/guidance/5667fnl.htm; accessed Jan. 9, 2008.
5. Spilker, B., and Schoenfelder, J. (1991), *Data Collection Form in Clinical Trials*, Raven Press, New York.
6. Kroenke, D. M. (2006), *Database Processing: Fundamentals, Design, and Implementation*, 10th ed., Pearson Education, Upper Saddle River, NJ.
7. Society for Clinical Data Management (2007), Good Clinical Data Management Practices; available at http://www.scdm.org; accessed Jan. 9, 2008.
8. Gibson, D., Harvey, A. J., Everett, V., and Parmar, H. K. B. (1994), Is double data entry necessary? The CHART trials, *Control. Clin. Trials*, 15, 482–488.
9. International Federation of Pharmaceutical Manufacturers and Associations, MedDRA—The Medical Dictionary for Regulatory Activities; available at http://www.meddramsso.com; accessed Jan. 9, 2008.
10. World Health Organization, WHO Drug Dictionary—WHODRUG; available at http://www.umc-products.com; accessed Jan. 9, 2008.
11. International Health Terminology Standards Development Organisation, Systematized Nomenclature of Medicine-Clinical Terms (SNOMED-CT); available at http://www.ihtsdo.org; accessed Jan. 9, 2008.
12. The eClinical Forum and PhRMA EDC/eSource TaskForce (2006), The future vision of electronic health records as e-source for clinical research, Version 1.0; available at http://www.cdisc.org/pdf/Future%20EHR-CR%20Environment%20Version%201.pdf; accessed Jan. 9, 2008.
13. Brandt, C. A., DeshpAnde, A. M., Lu, C., Ananth, G., Sun, K., Gadagkar, R., Morse, R., Rodriguez, C., Miller, P. L., and Nadkarni, P. M. (2003), TrialDB: A web-based Clinical Study Data Management System, *AMIA Annu. Symp. Proc.*, 794.

14. Hanover, J., Golden, J., and Swenson, M. (2004), Clinical Trial Automation and Data Management Solutions. Document Nr. IDC30946, Volume 1; available at http://www.healthindustry-insights.com/HII/getdoc.jsp?containerId=30946; accessed Jan. 9, 2008.
15. Food and Drug Administration (2002), General principles of software validation; final guideline for industry and FDA staff; available at http://www.fda.gov/cdrh/comp/guidance/938.pdf; accessed Jan. 9, 2008.
16. Weiner, M. G., Lyman, J. A., Murphy, S., and Weiner, M. (2007), Electronic health records: High-quality electronic data for higher-quality clinical research, *Inform Prim. Care*, 15(2), 121–127.
17. Murphy, E. C., Ferris, F. L.3rd , and O'Donnell, W. R. (2007), An electronic medical records system for clinical research and the EMR EDC interface, *Invest. Ophthalmol. Vis. Sci.*, 48(10), 4383–4389.
18. Gerdsen, F., Müeller, S., Jablonski, S., and Prokosch, H. U. (2005), Standardized exchange of medical data between a research database, an electronic patient record and an electronic health record using CDA/SCIPHOX, *AMIA Annu. Symp. Proc.*, 963.
19. Powell, J., and Buchan, I. (2005), Electronic health records should support clinical research, *J. Med. Internet Res.*, 7(1); available at http://www.jmir.org/2005/1/e4/.
20. Gruppo Italiano per lo Studio della Sopravvivenza nell'Infarto Miocardio (2004), GISSI-3: Effects of lisinopril and transdermal glyceryl trinitrate singly and together on 6-week mortality and ventricular function after acute myocardial infarction, *Lancet*, 343, 1115–1122.
21. Santoro, E., Nicolis, E., and Franzosi, M. G. (1999), Telecommunication technology for the management of large scale clinical trials: The GISSI experience, *Comput. Methods Programs Biomed.*, 60, 215–223.
22. Santoro, E., Nicolis, E., Franzosi, M. G., and Tognoni, G. (1999), Internet for clinical trials: Past, present, and future, *Controlled Clin. Trials*, 20, 194–201.
23. Santoro, E. (2002), Internet and cardiovascular research: The present and its future potentials and limits, *Minimally Invasive Ther. Allied Tech.*, 11, 73–75.
24. Paul, J., Seib, R., and Prescott, T. (2005), The Internet and clinical trials: Background, online resources, examples and issues, *J. Med. Internet Res.*, 7(1); available at: http://www.jmir.org/2005/1/e5/.
25. Marks, R., Bristol, H., Conlon, M., and Pepine, C. J. (2001), Enhancing clinical trials on the Internet: Lessons from INVEST, *Clin. Cardiol.*, 24(11 Suppl), 17–23.
26. Shaw, P. A., Goodman, P. J., and Brace, J. (2001), The web based management of the selenium and vitamin e cancer prevention trial, *Controlled Clin. Trials*, 22, 57S.
27. Long, D. T., Workman, J., Beck, R., and Moke, P. (2001), A web based procedure for drug distribution and accountability, *Controlled Clin. Trials*, 22, 80S.
28. Cooper-DeHoff, R., Handberg, E., Heissenberg, C., and Johnson, K. (2001), Electronic prescribing via the Internet for a coronary artery disease and hypertension megatrial, *Clin. Cardiol.*, 24(11 Suppl), 14–16.
29. Kiuchi, T., Ohashi, Y., Konishi, M., Bandai, Y., Kosuge, T., and Kakizoe, T. (1996), A world wide web-based user interface for a data management system for use in multi-institutional clinical trials—Development and experimental operation of an automated patient registration and random allocation system, *Controlled Clin. Trials*, 17, 476–493.
30. Dorman, K., Saade, G. R., Smith, H., and Moise, K. J. (2000), Use of the World Wide Web in research: Randomization in a multicenter clinical trial of treatment for twin-twin transfusion syndrome, *Obstet. Gynecol.*, 96(4), 636–639.
31. Bland, M. Directory of randomisation software and services; available at http://www-users.york.ac.uk/~mb55/guide/randsery.htm; accessed Jan. 9, 2008.

32. Marshall, W. W., and Haley, R. W. (2000), Use of a secure Internet web site for collaborative medical research, *JAMA*, 284, 1843–1849.
33. Oliveira, A. G., and Salgado, N. C. (2006), Design aspects of a distributed clinical trials information system, *Clin. Trials*, 3(4), 385–396.
34. ALLHAT Officers and Coordinators for the ALLHAT Collaborative Research Group (2002), Major outcomes in high-risk hypertensive patients randomized to angiotensin-converting enzyme inhibitor or calcium channel blocker vs diuretic: The Antihypertensive and Lipid-Lowering Treatment to Prevent Heart Attack Trial (ALLHAT), *JAMA*, 288, 2981–2997.
35. Vogel, V. G., Costantino, J. P., Wickerham, D. L., Cronin, W. M., Cecchini, R. S., Atkins, J. N., Bevers, T. B., Fehrenbacher, L., Pajon, E. R., Jr., Wade, J. L., 3rd, Robidoux, A., Margolese, R. G., James, J., Lippman, S. M., Runowicz, C. D., Ganz, P. A., Reis, S. E., McCaskill-Stevens, W., Ford, L. G., Jordan, V. C., Wolmark, N., and the National Surgical Adjuvant Breast and Bowel Project (NSABP) (2006), Effects of tamoxifen vs raloxifene on the risk of developing invasive breast cancer and other disease outcomes: The NSABP Study of Tamoxifen and Raloxifene (STAR) P-2 trial, *JAMA*, 295(23), 2727–2741.
36. Kuchenbecker, J., Dick, H. B., Schmitz, K., and Behrens-baumann, W. (2001), Use of Internet technologies for data acquisition in large clinical trials, *Telemed. J.E. Health*, 7(1), 73–76.
37. Kelly, M. A., and Oldham, J. (1997), The Internet and randomised controlled trials, *Int. J. Med. Inform.*, 47(1–2), 91–99.
38. The GRIT Study Group (1996), When do obstetricians recommend delivery for a high-risk preterm growth-retarded fetus? Growth Restriction Intervention Trial, *Eur. J. Obstet. Gynecol. Reprod. Biol.*, 67(2), 121–126.
39. Lallas, C. D., Preminger, G. M., Pearle, M. S., Leveillee, R. J., Lingeman, J. E., Schwope, J. P., Pietrow, P. K., and Auge, B. K. (2004), Internet based multi-institutional clinical research: A convenient and secure option, *J. Urol.*, 171(5), 1880–1885.
40. Tighe, F. P., and Cohen, J. (2001), Web Data Management. Experience with 20,000 case report forms in 14 ongoing studies, *Controlled Clin. Trials*, 22, 51S.
41. Pepine, C. J., Handberg, E. M., Cooper-DeHoff, R. M., Marks, R. G., Kowey, P., Messerli, F. H., Mancia, G., Cangiano, J. L., Garcia-Barreto, D., Keltai, M., Erdine, S., Bristol, H. A., Kolb, H. R., Bakris, G. L., Cohen, J. D., Parmley, W. W., and INVEST Investigators (2003), A calcium antagonist vs a non-calcium antagonist hypertension treatment strategy for patients with coronary artery disease. The International Verapamil-Trandolapril Study (INVEST): A randomized controlled trial, *JAMA*, 290(21), 2805–2816.
42. Clivio, L., Tinazzi, A., Mangano, S., and Santoro, E. (2006), The contribution of information technology: Towards a better clinical data management, *Drug Dev. Res.*, 67, 245–250.
43. Brandt, C. A., Morse, R., Matthews, K., Sun, K., DeshpAnde, A. M., Gadagkar, R., Cohen, D. B., Miller, P. L., and Nadkarni, P. M. (2002), Metadata-driven creation of data marts from an EAV-modeled clinical research database, *Int. J. Med. Inform.*, 65(3), 225–241.
44. Chen, R. S., Nadkarni, P., Marenco, L., Levin, F., Erdos, J., and Miller, P. L. (2000), Exploring performance issues for a clinical database organized using an entity-attribute-value representation, *J. Am. Med. Inform. Assoc.*, 7(5), 475–487.
45. Santoro, E., Clivio, L., and Mangano, S. (2007), GCPBASE: A web-based tool for remote data capture in a clinical trial, *Tech. Health Care*, 15(5), 355.
46. McAlindon, T., Formica, M., Kabbara, K., Lavalley, M., and Lehmer, M. (2003), Conducting clinical trials over the internet: feasibility study, *BMJ*, 327(7413), 484–487.

# *Clinical Trials and the Food and Drug Administration*

**Tarek M. Mahfouz and Janelle S. Crossgrove**
*Raabe College of Pharmacy, Ohio Northern University, Ada, Ohio*

**Contents**

1 Food and Drug Administration: History and Structure
  1.1 Brief History of FDA and U.S. Drug Regulation
  1.2 Structure and Organization of FDA
2 New Drug Development and FDA
  2.1 Preclinical Studies and IND
  2.2 Clinical Trials and New Drug Application (NDA)
3 NDA and Biological License Application (BLA)
4 Timeline
5 Cost and Probability of Success
6 FDA Meetings and Drug Sponsors
7 Regulation of Drugs and Biological Products by FDA
8 Expanded Access and Accelerated Approval
9 Orphan Drugs
10 Pediatric Drugs
11 OTC Drug Products
12 Behind-the-Counter Drugs
13 Drugs for Counterterrorism
14 Globalization and Harmonization: FDA and ICH
References

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

## 1 FOOD AND DRUG ADMINISTRATION: HISTORY AND STRUCTURE

The Food and Drug Administration (FDA) is the agency within the Department of Health and Human Services (HSS) that is responsible for protecting the health of Americans by enforcing the Federal Food, Drug, and Cosmetic Act and other related public health laws. FDA inspectors visit manufacturing facilities to ensure that products are made correctly and labeled truthfully. The FDA also protects consumers in several other ways. It investigates the biological effects of widely used chemicals, it tests medical devices, radiation-emitting products, and radioactive drugs, and it tests food for substances such as pesticide residues. It also monitors biologics such as blood and its products, insulin, and vaccines. In addition, components such as dyes and other additives used in cosmetics, drugs, and foods are all subject to FDA scrutiny.

The most important and critical role of the FDA, however, is in the development of new drugs and medical devices, such as pacemakers. New drugs and medical devices must receive FDA approval before they can be marketed. As the FDA states:

> In deciding whether to approve new drugs, FDA does not itself do research, but rather examines the results of studies done by the sponsor (that is the entity manufacturing or importing the new drug). For a new drug to be approved for marketing by the FDA, the agency must determine that the new drug produces the benefits it's supposed to without causing side effects that would outweigh those benefits [1].

In this chapter, we explain the different steps a new drug goes through before it is marketed and the involvement of the FDA in each step. Special attention will be given to FDA regulation of clinical trials. In order to understand how the FDA has evolved into the regulatory body that exists today, we begin with a brief history of the FDA and an overview of the structure and organization of the agency.

### 1.1 Brief History of FDA and U.S. Drug Regulation

In 1820, and for the three decades that followed, the *United States Pharmacopoeia* (USP) served as a reference for physicians and pharmacists who worked in the extraction or compounding of drugs and drug components that were available at the time. No other laws against food and drug adulteration were in place at that time. In 1848 Dr. M. J. Bailey testified in Congress that more than one half of the imported medicinal products were so much adulterated that they were not only ineffective but could be dangerous [2]. This led to the Drug Importation Act, which required inspection and destruction of drugs that did not meet acceptable standards [3]. The law seemed to be successful at first as very few bad products came into the United States in the year after its passage [2]. However, the law did not achieve its goal, mainly because political influences and bribery affected inspectors who were also not working with fixed standards [2].

In 1862, chemist Charles M. Wetherill was appointed to head the chemical division in the U. S. Department of Agriculture, a precursor to today's FDA, charged with maintaining a safe food supply. In 1901, the chemical division became the Bureau of Chemistry and was headed by Dr. Harvey W. Wiley who was a knowl-

edgeable physician on the issue of food and drug adulteration and was renowned for his "poison squad" experiments [2]. Dr. Wiley's experiments showed that some of the most commonly used preservatives at the time, such as borax, were not safe for human consumption and might lead to permanent stomach and bowel impairments. It was Dr. Wiley's and others efforts that led to the passage of the Food and Drugs Act by the Congress in 1906, and the Bureau of Chemistry was charged to enforce that law. Later, the Sherley amendment in 1911 prohibited the labeling of medications with false therapeutic claims. In 1927, Congress authorized the formation of the Food, Drug, and Insecticide Administration from the regulatory wing of the Bureau of Chemistry, and 3 years later the name was shortened to the Food and Drug Administration. In response to the deaths of over 100 people (mostly children) from the use of the untested compound diethylene glycol to solubilize sulfanilamide, the Food, Drug, and Cosmetic Act was passed in 1938. This new act required that new drugs be tested by their manufacturers for safety and the results be submitted to the government for marketing approval via the new drug application (NDA) [3]. Importantly, the new law authorized the FDA to conduct unannounced inspections of drug manufacturing facilities. In 1940 [4], the FDA left the Department of Agriculture for the Federal Security Agency, and in 1953 the FDA became part of the Department of Health, Education, and Welfare (HEW). In 1968 the FDA became part of the Public Health Service within HEW. When the education function was removed from HEW to create a separate department, HEW became the Department of Health and Human Services in 1980.

### 1.2 Structure and Organization of FDA

The FDA consists of eight centers:

1. *Office of the Commissioner (OC)* The office of the commissioner is formed of several components, such as the history office and the office of combination products, and is responsible for the implementation of the FDA's mission.

2. *Office of Regulatory Affairs (ORA)* The office of regulatory affairs is the office responsible for all field activities of the FDA. It ensures the compliance with FDA's standards by [5]:

   1. Monitoring the clinical trials that are conducted before a product is submitted for approval by the FDA and conducting domestic and foreign inspections of drug manufacturing facilities by the consumer safety officers.
   2. Analyzing product samples to determine their adherence to FDA standards.
   3. Reaching out to consumer groups and health authorities to explain the FDA's policies and encourage compliance with the agency standards by public affairs specialists. The public affairs specialists also respond to public health emergencies caused by natural disasters and product problems.

3. *Center for Biologics Evaluation and Research (CBER)* The CBER regulates biological products [6] such as blood and blood products, vaccines, and protein-based drugs such as monoclonal antibodies and cytokines. CBER also helps advancing and licensing of products to diagnose, treat, or prevent outbreaks from exposure to bioterrorist pathogens by helping the products to rapidly meet the regulatory

requirements. CBER also works on developing procedures and protocols to advance and make available promising experimental products when there is no approved medication for the treatment of victims of terrorism [6].

4. *Center for Devices and Radiological Health (CDRH)* The CDRH regulates [7] medical devices ranging from contact lenses and blood sugar monitors to implanted heart valves. It makes sure that new medical devices are safe and effective before they are marketed, and it monitors these devices throughout their life cycle using a nationwide postmarket surveillance system. CDRH also ensures that radiation-emitting products, such as microwave ovens, TV sets, cell phones, and laser products meet radiation safety standards.

5. *Center for Drug Evaluation and Research (CDER)* The CDER regulates [8] prescription and over-the-counter (OTC) drugs. Its mission is to ensure that all prescription and OTC drugs are safe and effective. To do this, CDER evaluates all new drugs before marketing; and after a drug is marketed CDER serves as a consumer watchdog to be sure these drugs continue to meet the standards. CDER also monitors TV, radio, and print drug ads to ensure they are truthful and balanced. As a measure to counter terrorism, CDER facilitates development of new drugs and new uses for already approved drugs that could be used as medical countermeasures.

6. *Center for Food Safety and Applied Nutrition (CFSAN)* The CFSAN regulates about 80% [9] of all food consumed in the United States. The remaining 20%, which includes meat, poultry, and some egg products, are regulated by the U.S. Department of Agriculture.

7. *Center for Veterinary Medicine (CVM)* The CVM [10] regulates the manufacture and distribution of food additives and veterinarian drugs or devices that will be used on animals including both pets and animals from which human foods are derived. CVM assures that animal drugs and medicated feeds are safe and effective. In approving veterinarian drugs (whether prescription or OTC) for food animals, CVM determines that no unsafe residues or metabolites will result when the drug is used in the approved manner, and that all safety factors are considered when setting the approved levels of use.

8. *National Center for Toxicological Research (NCTR)* The NCTR [11] conducts peer-reviewed scientific research focused toward understanding critical biological events that lead to toxicity and toward developing methods and incorporating new technologies to improve the assessment of human exposure, susceptibility, and risk.

## 2 NEW DRUG DEVELOPMENT AND FDA

New drug development is a multistage process that starts by identifying a drug target, usually an enzyme or a biological process whose function is crucial to treating a disease or ameliorate a medical condition. Many isolated or newly synthesized compounds are screened for activity at the target site using high throughput in vitro assays. If the interaction is strong enough to be of any medical significance, these compounds are then tested in vivo in model animals to evaluate their pharmacological activities and acute toxicity potentials. These initial testing steps are called the preclinical studies. The FDA does not monitor these preclinical studies directly,

assuming that good laboratory practices (GLP) [12] have been followed. The sponsor's primary goal in these studies is to determine if the new compound is reasonably safe for initial use in humans and if it exhibits pharmacological activity that justifies commercial development. If the compound proves to be a promising candidate for further development, the sponsor must collect sufficient data to establish that the new compound will not expose humans to unreasonable risks when used in early-stage clinical studies. The results of these preclinical studies, therefore, are important and they must be included in the investigational new drug (IND) application submitted to the FDA for review before clinical trials can begin.

### 2.1 Preclinical Studies and IND

Direct FDA involvement in the new drug development process begins if data from the preclinical studies indicates that the new drug candidate is effective and reasonably safe to test in humans. In this case, the sponsor can submit an IND to formally notify the FDA of his or her intent to start clinical trials on human subjects. The IND falls in two categories and three types [13]. In this section we will cover the investigative IND, and in later sections we will cover the emergency use IND and the treatment IND. For more information on INDs, the reader is referred to the CDER web page titled Drug Applications (www.fda.gov/Cder/regulatory/applications/ind_page_1.htm#Introduction).

An investigator IND may be submitted by a physician who both initiates and conducts an investigation and under whose immediate direction the investigational drug is administered or dispensed. A physician might submit a research IND to propose studying an unapproved drug, or an approved product for a new indication or in a new patient population. The IND application itself has several components but the information therein falls in three main areas:

1. Results of the preclinical animal pharmacological and toxicological studies together with any previous experience with the drug in humans. This information allows the FDA to decide if the compound is reasonably safe for initial testing in humans.
2. Information on the stability, composition, and manufacturing of the drug and the drug product. This information is provided to assure that the sponsor can supply consistent batches of the drug.
3. The proposed clinical studies protocols and the qualifications of the clinical investigators who will oversee the administration of the experimental compound to assess if people participating in the trial will be exposed to unreasonable risk and if the clinical investigators are qualified to fulfill their clinical trial duties.

Submitted with the IND also are commitments to obtain informed consents [14] from the research subjects and to obtain review of the study by an institutional review board (IRB) [15]. The informed consent document is given to the participants in a clinical trial and states clearly to the participants the purpose of the trial, how long they are expected to participate, what will happen in the study, and what the possible risks or discomforts are, possible benefits, other procedures or treat-

ments that might be available and advantageous to them, and that participation is voluntary and the participants can quit at any time should they so desire. An IRB is a committee composed of at least five medical and ethical experts designated by the institution where the clinical studies are to take place to review and approve clinical trials taking place within their jurisdiction to ensure that medical and scientific standards are maintained and to protect the rights of the human test subjects. The steps of the IND application process are summarized in Figure 1. After the submission of an IND, the FDA has 30 days to respond. During these 30 days, FDA experts review the IND and decide if the new drug is safe to be tested in human

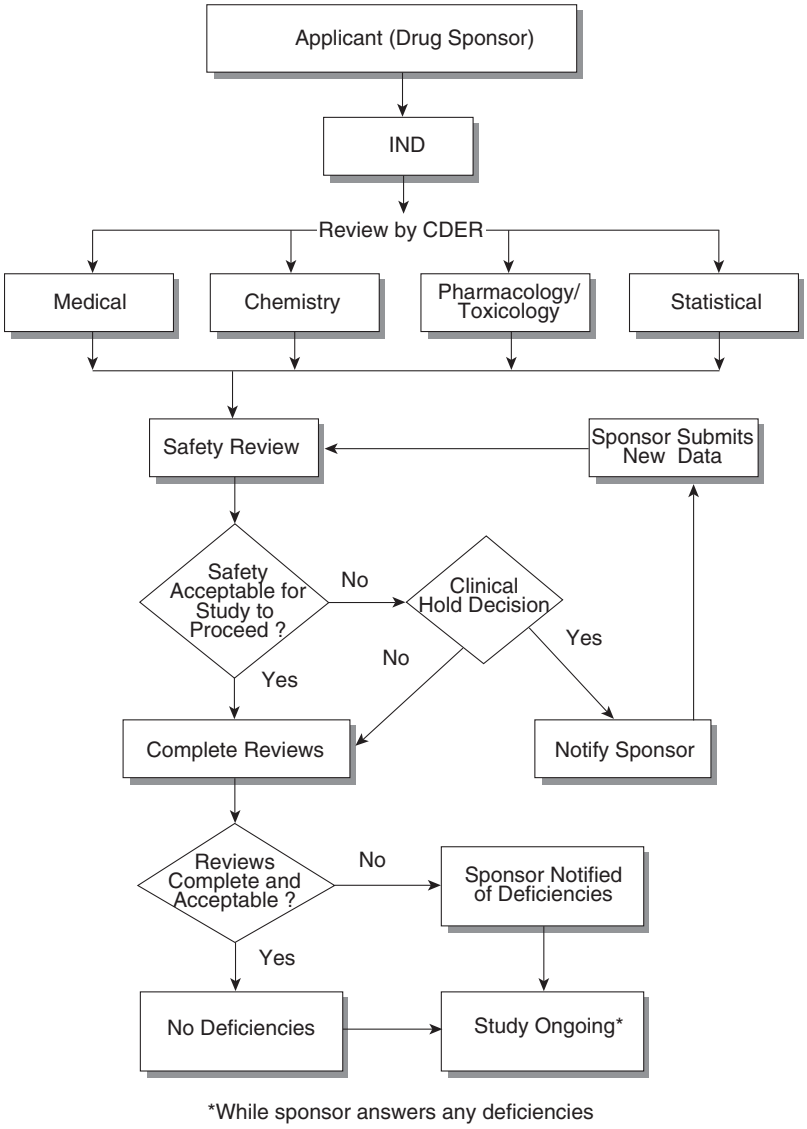


***FIGURE 1*** *Steps an IND takes for review and approval. (Source: FDA.)*

subjects. If the FDA approves the study and the IRB approves the proposed protocols, the clinical trials can begin.

### 2.2 Clinical Trials and New Drug Application (NDA)

In clinical trials, the effectiveness of the new drug in treating a disease or controlling a condition is compared to standard treatments or to no treatment at all (i.e., placebo). Side effects, toxicity, and metabolism are also studied. Because clinical trial participants represent only a small fraction of the target patient population and because drugs can work differently in different populations, it is important that the participants be representative of the wider general population by including people of various age groups, races, ethnic groups, and genders in the trials. Clinical trials usually are conducted in three main phases with a postmarketing phase. The three main phases of clinical trials differ in three main aspects: number and type of subjects required, purpose, and duration [16].

*Phase I* Phase I is the smallest of the three phases lasting up to 3 years and requiring 20–100 healthy volunteers. The main purpose of this phase is to determine dosing, how the drug is metabolized and excreted, and to identify the acute side effects.

*Phase II* Phase II trials are slightly larger than phase I, requiring 100–500 patients with the disease or medical condition that the new drug targets and may last from 2 to 4 years. The purpose of this phase is to collect information on the safety of the new drug and its efficacy. At the end of this phase, the drug sponsor meets with the FDA to discuss the results and how to proceed next. If the results indicate that the drug may be effective and the side effects are considered acceptable, the drug moves on to phase III.

*Phase III* Phase III is the largest of all phases requiring 1000–5000 patients and may take several years to finish. In this phase, the drug safety and effectiveness are further studied; and, if a standard treatment is available, the effectiveness of the new drug compared to that standard is also examined. When more and more participants are tested over longer periods of time, the less common side effects are more likely to be revealed. Phase III also establishes other aspects of the drug development process such as marketing claims, packaging, and storage conditions.

The role of the FDA in clinical trials is (1) to help protect the rights and welfare of the patients participating in the trial and (2) to verify the quality and the integrity of the data. To achieve these goals, the FDA's division of scientific investigation (DSI) conducts inspections of clinical investigators' study sites and reviews the records of the IRBs to make sure they are fulfilling their roles in patient protection. Also, the DSI seeks to determine whether the study was conducted according to the investigational plan, whether all adverse events were recorded, and whether the subjects met the inclusion/exclusion criteria outlined in the study protocol. At the conclusion of each inspection, FDA investigators prepare a report summarizing any deficiencies. In cases where they observe numerous or serious deviations, DSI classifies the inspection as "official action indicated" and sends a warning letter or

Notice of Initiation of Disqualification Proceedings and Opportunity to Explain (NIDPOE) to the clinical investigator specifying the found deviations [17]. The FDA usually has authority over clinical trials, and it can stop an ongoing trial or halt it if serious complications develop that puts the participants at high risk.

## 3 NDA AND BIOLOGICAL LICENSE APPLICATION (BLA)

After the conclusion of phase III of the clinical trials, the sponsor submits an NDA [18] to the FDA asking the agency to consider approving the new drug for marketing in the United States. The NDA contains all animal and human data and analyses of these data together with information on the chemistry, stability, and the proposed manufacturing of the new drug. After receiving an NDA, the FDA has 60 days to decide whether to file the NDA for review by the agency experts, or, if the NDA is incomplete, not to file it. The time the FDA spends in reviewing the NDA will be discussed in the timeline section of this chapter but outcome of the review process can be approved, approvable, or nonapprovable. "Approved" means the new drug product has met all the requirements and the sponsor can begin marketing the new drug. "Approvable" means the new drug product has some minor deficiencies that need to be addressed before approval. "Nonapprovable" means the FDA will not approve the new drug product the way it is submitted due to major deficiencies.

Biologics such as blood and its products, vaccines, antibodies, and the like are dealt with the same way. New biologics are subject to the same regulations and follow the same clinical testing as new drugs. The only difference is that to apply for marketing new biologics, sponsors need to submit a BLA not an NDA. BLAs are submitted to and reviewed by the CBER.

For many drugs approved for marketing, the FDA requires the sponsor to continue submission of clinical data to further validate their safety or effectiveness. In some cases, more studies are needed to find out more about a drug's long-term risks, benefits, and optimal use or to test the drug in different populations of people such as children. These are the reasons behind postmarket surveillance.

## 4 TIMELINE

There is no doubt that drug development is a complicated, costly, and time-consuming process. The various steps a new drug endures from its synthesis to the time it makes it to the market as a drug vary in length from an average of 2 years in the research and development phase to an average of 7 years in clinical trials, as seen in Figure 2. It is hard to estimate how long before a newly synthesized chemical compound will make its way to the market as a drug product because each new drug has a different story and a unique path. Most potential drug candidates do not even make it to the clinical trials and are abandoned in the preclinical stage. In the research and development stage, a large number of chemical compounds are screened by researchers to identify only a few promising candidates called leads. Lead identification is the most time-consuming step in the drug discovery process because once leads are identified, structural modification can optimize

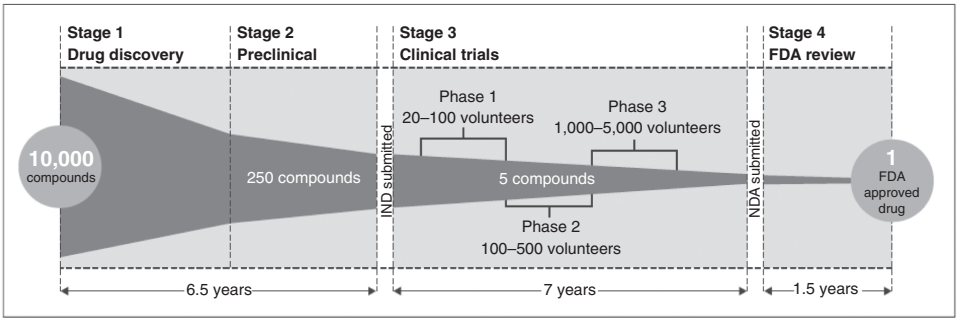


***FIGURE 2*** *Drug discovery, development and review process. (From Ref. 19.)*

their activities or properties as drug candidates. Advances in high-throughput screening technology and computational drug design methods help accelerate this stage a little but, depending on the circumstances, this stage can take up to 5 years.

Once leads have been identified and optimized, toxicity and animal studies can begin. These are less time consuming than the previous step and may take up to 3 years. After the preclinical studies are completed, the new drug sponsor needs to submit an IND and get the FDA approval to begin the clinical studies. As stated above, the FDA has a period of 30 days to decide whether to hold the clinical trials or to give the green light. Clinical trials are the most time-consuming step in the process of new drug development and are the most critical step as well, especially phase II and phase III. Because of the diverse areas investigated and the comprehensive nature of these two phases, they take a long time to finish, sometimes up to 10 years. As the clinical trials evolve and more data become available, changes in the clinical protocols may become necessary. Since each change has to be approved by the IRB, the period of time the drug remains in clinical trials is further extended. All preclinical and clinical data must be included in the NDA submitted to the FDA for approval upon completion of the clinical trials. If the NDA is filed and in accordance with the Prescription Drug User Fee Act (PDUFA; www.fda.gov/cder/pdufa/default.htm) the CDER should complete its initial review and act on at least 90% of all NDAs for standard drugs (those for which there are no perceived significant therapeutic benefits beyond those for available drugs) no later than 10 months after the applications are received and no later than 6 months for priority drugs (those that the FDA expects to provide significant therapeutic benefits beyond drugs already marketed) [19]. A 2002 general accounting office (GAO) report on the effect of user fees on FDA review times [20] has indicated that the median approval time for new drugs has dropped since the implementation of PDUFA in 1992. As shown in Figure 3, from 1993 to 2001, the median approval time for standard new drug applications has dropped from about 27 months to about 14 months and from about 21 months to about 6 months for priority new drugs. The reason for this drop in FDA approval times was attributed to the new resources that allowed the agency to recruit more reviewers and to upgrade its resources in information technology shortening the approval times significantly. Figure 2 summarizes the timeline and the different stages of the drug development process.

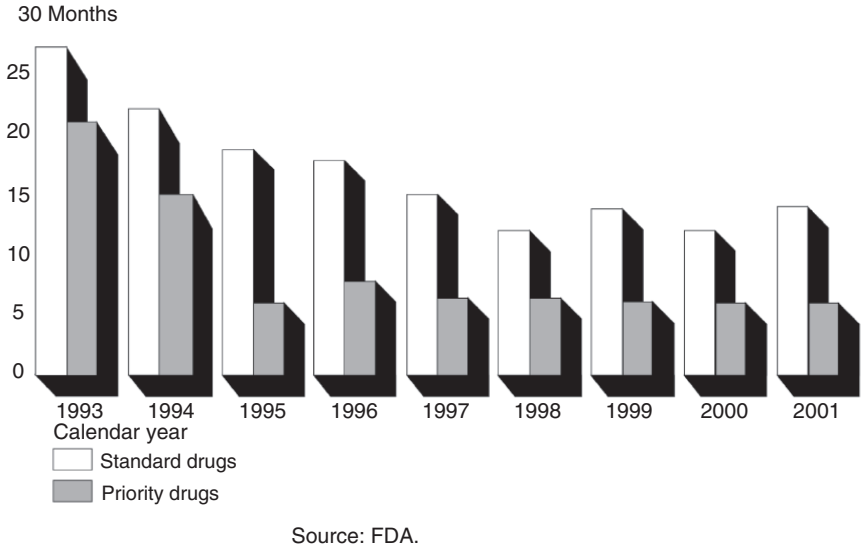


***FIGURE 3*** *Median approval times for standard and priority drug applications based on calendar year of approval, 1993–2001. (From Ref. 19.)*

## 5 COST AND PROBABILITY OF SUCCESS

Much work has gone into determining how much the drug development process may cost from start to finish. Industry estimates on total research and development costs suggest an inflation-adjusted increase from almost $16 billion annually in 1993 to almost $40 billion in 2004. Out-of-pocket expenses for development of a single drug are estimated at $403 million, with capitalized expenses nearing $800 million (2000 dollars) [21]. Time and capital requirements grow at each clinical phase. Phase I studies have been estimated to cost $30 million and take two to three years; Phase II studies may cost $40 million and 2–4 years; phase III studies may cost $86 million [21, 22]. Although size, duration, and design of the study may greatly vary, one estimate suggests that phase II clinical trials can cost anywhere from $2000 to $10,000 per subject [22]. Consumer advocates have questioned whether self-disclosure of drug development costs by industry may present an accurate picture of true costs, as a conflict of interest may be present [23]. Nonetheless, cost analysis of taking a drug to market is a ubiquitous part of the sponsor's decision-making process. Careful study design is essential in clinical studies to maximize data output while minimizing financial and health risks to the participants.

In addition to costs of drug development, a sponsor will also incur user fees when applying for FDA approval. The Prescription Drug User Fee Act (PDUFA) was instituted to provide resources for the FDA to review applications more quickly. Other legislation establishing user fees for medical devices and animal drugs has followed. The user fee encompasses an application fee for each drug, an establishment fee to cover the site(s) of drug production, and a product fee for each drug product in the application. Table 1 provides recent figures for user fees established by the PDUFA.

**TABLE 1 User Fees for Prescription Drug Applications in Selected Fiscal Years**

| | 2006 | 2007 | 2008 |
|---|---|---|---|
| Applications | | | |
| Requiring clinical data | $767,411 | $896,200 | $1,178,000 |
| Not requiring clinical data or supplements requiring clinical data | 383,700 | 448,100 | 589,000 |
| Establishments | 264,000 | 313,100 | 392,700 |
| Products | 42,130 | 49,750 | 65,030 |

*Note*: Values have not been adjusted for inflation or capitalized in any way.

Determining whether or not a drug is likely to be approved is a challenging task. As noted in Figure 2, it is estimated that only one in five drugs that begins clinical trials is approved for market. At the NDA stage, the odds are much better. Of the NDAs submitted from 1993 to 2004, most (76%) were approved by the FDA, lesser amounts ongoing (17%) or withdrawn (7%) [19]. At each phase of clinical trials, good communication between the sponsor and the FDA team is vital. The FDA can suggest specific study designs to optimize the collection of useful and required data for its drug review. Ultimately, though, the drug or product must demonstrate an excellent safety and efficacy profile in order to ensure approval. The chance that a particular drug is approved for market depends on its performance in clinical trials.

## 6 FDA MEETINGS WITH DRUG SPONSORS

The FDA has a guidance document on the subject, Formal Meetings with Sponsors and Applicants for PDUFA Products, available at its website (www.fda.gov/cder/guidance/2125fnl.htm). Briefly, it describes the types of meetings, how and when to request a meeting, when to submit information and what information to submit prior to a meeting, procedures to conduct the meeting, and documentation of the meeting's focus and outcomes. Meetings can and should be scheduled throughout the process of investigating a drug or product, but the most critical meetings are the pre-IND meetings, end of phase II meetings, pre-NDA/BLA meetings, and the labeling meetings. At each of these checkpoints, the sponsor presents its clear scientific evidence of continuing (or discontinuing) the process, while the FDA provides feedback on the sponsor's plan and suggests improvements for a successful application.

The primary goal of the pre-IND meeting is "to introduce the drug to the FDA" [24]. If the sponsor is a small company or is not well known to the FDA, a second goal of this meeting is to present evidence that the company is qualified to perform the studies. The sponsor should present all scientific information about the drug, including any potential side effects identified in the nonclinical studies. The FDA team involved at this meeting will work with the sponsor to assess and reduce risk for the phase I study population. The best outcome from this meeting is that the FDA agrees with the sponsor that an IND can be submitted (which does not guarantee its approval).

A meeting at the end of phase II is almost universally advised. This meeting should be scheduled as soon as (1) phase II clinical data have established an effective dose and revealed pharmacokinetic and pharmacodynamic profiles that support

advancement to phase III and (2) the study design for phase III trials is complete. The phase III study design will be scrutinized thoroughly by the FDA prior to and at the meeting.

The pre-NDA/BLA meeting is recommended as a time to identify any potential pitfalls that might hinder the review of the submission. Depending on the type of drug involved, the FDA may use this meeting to discern the number and type of outside reviewers that may need to join an advisory committee. At this meeting, the FDA routinely confirms that the sponsor understands the NDA submission process and its timeline.

The labeling meeting is viewed as the last step in drug/product development prior to its approval. Although this step may seem small, the prescribing information included in product labeling can make the difference between a drug with minimal impact or one with a maximal impact on the market. A company with so much time, money, and effort in the drug development would certainly be heavily invested in securing a label that will allow the greatest impact on the market that the clinical data allow. Several rounds of negotiation on the final wording can occur. Once that hurdle is reached, the new product can be released onto the market.

The FDA can classify any meetings as type A, B, or C to prioritize the scheduling of the meeting. While sponsors can request a specific designation, the FDA is the authority on the matter. Type A meetings are the most urgent, requiring scheduling within 30 days of receiving the designation. Priority among type A meetings is given to those involving issues with a submitted NDA/BLA. Type B meetings must occur within 60 days of classification and type C within 75 days. The deadline for submission of supporting evidence is specific to the meeting type. Support documents are expected to be received by the FDA 2 weeks in advance of scheduled type A and C meetings. For type B meetings, documentation must be received 1 month in advance. It is vital for sponsors to be prepared to meet these deadlines when requesting a meeting with the FDA. For more information and advice on preparing for meetings with the FDA, see Grignolo's chapter titled "Meetings with the FDA" in *FDA Regulatory Affairs* [24].

## 7 REGULATION OF DRUGS AND BIOLOGICAL PRODUCTS BY FDA

The FDA is involved in regulation of drugs and biological products as they come to market and as they remain on the market. In addition to the regular INDs and NDAs mentioned earlier, the FDA regulates processes for approving generic drugs, over-the-counter (OTC) drugs, and follow-on drugs (modified chemical entities of approved first-in-class drugs). The FDA also has special rules and/or exemptions for expanded access and accelerated approval, orphan and pediatric drugs, and drugs for antiterrorism.

Generic drugs have the same active ingredient as an already approved, brand name drug (sometimes called the innovator drug). Applications for new generic drugs are examined in the FDA by CDER's Office of Generic Drugs. In order to be approved, generic drugs must be shown to be scientifically bioequivalent to the innovator drug. That is, the generic drug must have the same physical, pharmacokinetic, and toxicokinetic profiles as its brand-name counterpart, and it must produce the same pharmacologic effect for the same intended use. The differences between

a generic and a prescription drug are produced solely by changes in inert ingredients and/or changes in pill shape. Generic drugs are approved through an abbreviated new drug application (ANDA). The NDA is abbreviated because some or most of the clinical trials do not need to be repeated. If the drug or drug product is deemed to be bioequivalent to the innovator, that is, having the same bioavailability and pharmacokinetic profile, the same target effect, and the like, then it can be reasonably assumed to be as safe as the innovator drug.

There are a number of resources available to guide companies considering entry into the generic drug market. The FDA Office of Generic Drugs has a comprehensive website (www.fda.gov/cder/ogd/) dedicated to supporting new and worthy applications to approval for market. The FDA, together with industry feedback, has produced several guidance documents for preparation of NDAs and ANDAs. These guidance documents are neither law nor FDA rule, but rather advice from both sides of the application process on what to expect and how to present a well-prepared application. These guidance documents are fluid; they have been reviewed and updated as changes in the process or industry have warranted. A searchable database of guidance documents is available on the FDA website (www.fda.gov/opacom/morechoices/industry/guidedc.htm).

The guidance document on bioavailability and bioequivalence studies for oral drugs provides an excellent overview of the studies needed for ANDAs (www.fda.gov/cder/guidance/3615fnl.htm). It includes a general pharmacokinetic study design and data analysis. Potential generic drug producers can use the pharmacokinetic profile determined in clinical trials of the innovator drug as a benchmark for subsequent bioequivalence studies. The guidance provides several details on the types of studies to be completed, including pharmacokinetic, pharmacodynamic, and in vitro dissolution studies. Distinctions are made between immediate and modified release dosage forms.

The FDA also has a mechanism by which confidential, supporting information can be considered alongside the IND, NDA, ANDA, or export application. Detailed information involving many aspects from production facilities and processes to packaging materials can supplement an existing application under consideration using a confidential document known as the Drug Master File (DMF). The DMF is not required for submission of an IND, NDA, or ANDA, nor does the DMF replace these applications. The DMF may be particularly helpful in cases of patent-pending manufacturing processes, including some that may fall under Section 505(b)(2).

Section 505(b)(2) applications are used when some of the information within the application was not generated by or for the sponsor if the sponsor has no right of reference to the information. This application differs from an NDA in that the Section 505(b)(2) can be delayed to accommodate those with patent rights and/or exclusivity protections of a particular product. For example, an application to combine two previously approved drugs into a combination drug may necessitate the use of the 505(b)(2) application if the idea of that particular combination occurred outside the company. These types of applications have also been used when changes to existing drugs occur, such as a change from a prescription indication to an OTC indication or a change in the drug formulation. Further information on the subject is available in a guidance document titled Applications Covered by Section 505(b)(2) (www.fda.gov/cder/guidance/2853dft.htm).

## 8 EXPANDED ACCESS AND ACCELERATED APPROVAL

The FDA has been making great strides in expanding the access to experimental therapies intended to treat life-threatening diseases and in accelerating the approval process of these promising treatments. The expanded access mechanisms allow severely ill patients with no response to approved therapies to have access to promising treatment drugs or devices. Expanded access can be granted through treatment IND protocols, parallel track protocols, or ordinary open-label studies that are a part of some normal NDAs. Generally, expanded access is only granted after the proposed drug or product has completed much of the clinical trial phase, when safety is established and efficacy is probable. To be approved for expanded access, four conditions must be met: (1) the drug or product must be intended to treat a serious or life-threatening disease, (2) no satisfactory alternative is available to treat that disease stage in that patient population, (3) the drug must either have an established IND protocol or all clinical trials must be completed, and (4) the sponsor must be pursuing marketing approval of the drug or product. Conversely, the parallel track mechanism allows patients with HIV/AIDS who have exhausted all other treatment options access to drugs in IND trials when the patients are not eligible to complete the trials. This mechanism is specific to HIV/AIDS-related products, and it has the potential to provide access earlier in the process than the treatment IND protocol.

In addition to expanded access mechanisms, the FDA supports new drug development in serious and life-threatening diseases through accelerated approval. A sponsor can apply for designation under the fast track drug development program. The drug or device must treat a serious aspect of a serious or life-threatening illness and address an unmet medical need.

Access to unapproved drugs may also be granted by filing a special exemption (also called compassionate exemption) or an emergency IND. Patients who are ineligible for a clinical trial may still be allowed access to the drug by filing as a special exemption. This requires investigator and sponsor approval, FDA consent, and an approved modification to the local IRB. With the permission of the drug supplier, a health care provider can alternatively file an emergency IND directly with the FDA to gain access to the drug or product. The health care provider must notify the IRB of his or her intent to use the drug; in a life-threatening situation in which specific criteria are met, according to 21 CFR 56.102(d), the drug may be administered prior to IRB approval.

## 9 ORPHAN DRUGS

The Office of Orphan Products Development is the subunit of the FDA that oversees, and sometimes funds, the clinical trials of orphan drugs. Enacted in 1983 and since revised, the Orphan Drug Act has provided tax and marketing incentives to entrepreneurs who want to study drug treatments or devices that are projected to be relevant only to small markets. These drugs and devices are considered "therapeutic orphans" if the intended target population is fewer than 200,000 individuals in the United States or in the case of preventative or diagnostic drugs, fewer than 200,000 individuals in the United States per year. When a drug or biologic receives

recognition as an orphan drug, this special status allows not only tax relief for costs related to development but also a waiver of the prescription drug user fee and a 7-year exclusivity of the drug on the market, allowing pharmaceutical companies time to recoup their expenses. Some consumer advocates propose that the exclusive license provides too great a benefit to the drug manufacturer and prevents drug access by prohibitively high end-user cost [25].

A sponsor can ask for designation as an orphan drug during or after the clinical trials. The sponsor submits documentation supporting the designation that the target disease is rare and that the drug will treat the disease. If orphan drug status is established, the investigators become eligible to compete for a research grant to defray the costs of a clinical trial. At the time of writing, the grant only provides funds for clinical trials of orphan drugs. This will facilitate approval of an unapproved drug or an unapproved use of a drug already on the market to treat a rare disease.

## 10 PEDIATRIC DRUGS

Pediatric patient populations experience growth and development that can interfere with drug absorption, distribution, metabolism, and excretion and can require special care in determining safe and effective dosing. The sponsor of a new drug or product intended to treat a disease or condition relevant to the pediatric population may request orphan drug designation, provided the drug is expected to treat 200,000 or less within this population within a given year. In addition to this designation, studies on pediatric populations for more common indications are regulated and required by the Pediatric Research Equity Act (PREA). PREA requires that all new NDAs and new BLAs for new chemical entities or new indication, dosage form, dosage strength, or route of administration contain an assessment of pediatric effectiveness unless a waiver or deferral is obtained. The FDA has posted a draft guidance document (www.fda.gov/cder/guidance/6215dft.pdf).

## 11 OTC DRUG PRODUCTS

Over-the-counter drug products comprise a special class of chemicals intended to treat a health condition that consumers may self-diagnose and self-medicate. In order to approve a drug for OTC use, the FDA must find that it is both safe and effective for the marketed use. OTC drugs are characterized by low health risk and high health benefit, low abuse and misuse potential, and adequate labeling for proper use.

OTC drugs can be marketed under two mechanisms: the NDA and the OTC drug monograph. The NDA process has been described above in detail. A change in existing OTC drug dosage form, dosage strength, or route of administration or the first time an OTC chemical entity comes to market requires new approval through the NDA. A drug previously available only by prescription must undergo NDA approval prior to introduction as an OTC drug, although sometimes this can fall under the guidelines of Section 505(b)(2). The sponsor of a prescription-to-OTC application must show that the drug is safe and effective for customer use without

the aid of a health care professional. Alternatively, an OTC drug can be marketed under an existing OTC drug monograph. The monograph regulates the active ingredients within an OTC product. Changes in inactive ingredients may not require further approval prior to marketing, provided that the active ingredient in the new product meets the standards of the monograph.

## 12 BEHIND-THE-COUNTER DRUGS

At the time of writing, the FDA is considering a third class of drugs to join prescription and nonprescription drugs: the behind-the-counter drugs. This class would be available without a prescription but would require discussion with a pharmacist. It is unknown at this time how a new class of drugs might alter the drug approval process.

## 13 DRUGS FOR COUNTERTERRORISM

In 2002, a measure was put into place to allow fast time-to-market for drugs or products intended to counteract the damaging effects of biological, chemical, radiological, and nuclear agents. The rule is officially titled Approval of Biological Products/New Drugs When Human Efficacy Studies Are Not Ethical or Feasible, but it is commonly called the animal rule. Because testing an antidote to these toxic substances in humans is unethical and unfeasible, the FDA must rely on well-characterized efficacy studies in animals and safety studies in humans and animals to determine whether to approve new drugs or products. Such determinations can only occur when the offending agent's mechanism of toxicity is understood, when the animal endpoints relate directly to human benefits, when the drug product's effect is determined in a species comparable to humans, and when the selection of an effective human dose is possible from the data. To date, two products have been approved by this animal rule: pyridostigmine bromide to combat nerve gas and hydroxycobalamin to treat cyanide poisoning.

## 14 GLOBALIZATION AND HARMONIZATION: FDA AND ICH

The International Conference on Harmonisation (ICH) brings together drug regulatory bodies throughout the world to formulate international standards and to streamline national and international policies for establishing safety and efficacy of new and existing drugs through nonclinical studies and clinical trials. The FDA has a large number of guidance documents relating to harmonization; so many in fact that it has further subdivided each ICH topic into categories involving safety, efficacy, joint safety/efficacy, and quality. The common technical document (CTD) that the FDA uses as a format for all its electronic NDA, ANDA, BLA, and IND submissions was designed by the ICH. The FDA encourages sponsors to use the electronic CTD in hopes that it will increase the efficiency of global marketing approval (www.fda.gov/cder/guidance/7087rev.pdf). In addition to globalization of the drug approval process, the ICH with the FDA and other agencies is examining implications of drug

interactions, including the use of drugs developed outside one's country together with cultural remedies [26].

## REFERENCES


1. FDA (1999), *Food and Drug Administration: An Overview*, Publication No. BG99-2, U.S. Government Printing Office, Washington, DC.
2. Hilts, P. J. (2003), *Protecting America's Health: The FDA, Business, and One Hundred Years of Regulation*, Alfred A. Knopf, New York.
3. Pisano, D. J. (2004), Overview of drug development and the FDA, in Pisano, D. J. and Mantus, D., Eds; *FDA Regulatory Affairs: A Guide for Prescription Drugs, Medical Devices, and Biologics*, CRC Press, Boca Raton, FL.
4. FDA (2002), A Guide to Resources on the History of the Food and Drug Administration; available at: http://www.fda.gov/oc/history/resourceguide/default.htm.
5. FDA (2003), *FDA's Sentinel of Public Health: Field Staff Safeguards High Standards*, Publication No. FS 01-7, U.S. Government Printing Office, Washington, DC.
6. FDA (2002), *FDA's Center on the Front Line of the Biomedical Frontier*, Publication No. FS 01-4, U.S. Government Printing Office, Washington, DC.
7. FDA (2002), *Better Health Care with Quality Medical Devices: FDA on the Cutting Edge of Device Technology*, Publication No. FS 01-5, U.S. Government Printing Office, Washington, DC.
8. FDA (2003), *Improving Public Health: Promoting Safe and Effective Drug Use*, Publication No. FS 01-3, U.S. Government Printing Office, Washington, DC.
9. FDA (2002), *Keeping the Nation's Food Supply Safe: FDA's Big Job Done Well*, Publication No. FS 01-2, U.S. Government Printing Office, Washington, DC.
10. Anon. (2007), CVM Introduction; available at: http://www.fda.gov/cvm/aboutint.htm; accessed July 14, 2007.
11. FDA (2007), NCTR's Mission; available at: http://www.fda.gov/nctr/overview/mission.htm; accessed July 14, 2007.
12. Code of Federal Regulations, 21 CFR Part 58.
13. FDA (2007), Investigational New Drug (IND) Application Process; available at: http://www.fda.gov/Cder/regulatory/applications/ind_page_1.htm#Introduction; accessed July 15, 2007.
14. Code of Federal Regulation, 21 CFR Part 50.
15. Code of Federal Regulation, 21 CFR Part 56.
16. Anon. (2006), Inside clinical trials: Testing medical products in people, FDA Consumer Mag., Publication No. FDA 06-1524G.
17. Anon. (2006), The FDA's drug review process: Ensuring drugs are safe and effective, FDA Consumer Mag., Publication No. FDA 06-1524G.
18. Code of Federal Regulations, 21 CFR Part 314.
19. GAO (2006), New Drug Development: Science, Business, Regulatory, and Intellectual Property Issues Cited as Hampering Drug Development Efforts, Report No. GAO-07-49.
20. GAO (2002), Effect of User Fees on Drug Approval Times, Withdrawals, and Other Agency Activities, Report No. GAO-02-958.
21. DiMasi, J. A., Hansen, R. W., and Grabowski, H. G. (2003), The price of innovation: New estimates of drug development costs, *J. Health Econ.*, 22(2), 151–185.

22. Schacter, B. (2006), *The New Medicines: How Drugs Are Created, Approved, Marketed, and Sold*, Praeger, Westport, CT.
23. Angell, M. (2004), *The Truth about the Drug Companies: How They Deceive Us and What to Do About It*, Random House, New York.
24. Grignolo, A. (2004), Meeting with the FDA, in Pisano, D. J., Mantus, D., Eds., *FDA Regulatory Affairs: A Guide for Prescription Drugs, Medical Devices, and Biologics*, CRC Press, Boca Raton, FL.
25. Thamer, M., Brennan, N., and Semansky, R. (1998), A cross-national comparison of orphan drug policies: Implications for the U.S. Orphan Drug Act, *J. Health Polit. Policy Law*, 23(2), 265–290.
26. Huang, S. M., Temple, R., Throckmorton, D. C., and Lesko, L. J. (2007), Drug interaction studies: Study design, data analysis, and implications for dosing and labeling, *Clin. Pharmacol. Ther.*, 81(2), 298–304.

# *Phase I Clinical Trials*

**Elizabeth Norfleet and Shayne Cox Gad**
*Gad Consulting Services, Cary, North Carolina*

**Contents**

1 Overview
2 Purpose and Objectives
3 Phase I Trial Design
4 Design Types
5 Phase I Trial Characteristics
6 Critical Parameters to Measure
7 PK Parameters to Derive
8 Regulatory Requirements and Issues
References
Bibliography

## 1 OVERVIEW

Phase I trials [also referred to as FIH (first in human) or FIM (first in man)] are the earliest stage clinical trials of a new drug or device typically performed with just a few persons to determine the safety and pharmacokinetics of a new drug or biocompatibility of a new invasive medical device; for drugs, dosage or toxicity limits should be obtained. These rigorously controlled tests of a new drug or a new invasive medical device involve human subjects for such evaluation. In the United States, such trials are conducted with the concurrence of the Food and Drug Administration (FDA) or equivalent regulatory authority before proceeding to further clinical investigation. While it is generally the ideal to perform no more than two (a single-dose escalating and a multiple-dose escalating) or three (for oral drugs, a feed/fasted

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

is also usual) phase I studies before proceeding to phase II, it is not uncommon to perform as many as eight (formulation changes, etc.).

Phase I trials are conducted following prescribed preclinical work that establishes the new molecular entity (NME) to be safe and tolerable in animal and in vitro models. The animal models chosen for such evaluation should reflect the expected response in humans as closely as possible. After preclinical work has been successfully conducted, a complete and thorough copy of all preclinical data must be submitted to the FDA in the format of an investigational new drug (IND). Subsequent to the FDA's evaluation of the data, an initial phase I trial may then commence if the sponsor has not received a response from the FDA within 30 days of the IND submission. Similar pretrial review procedures exist for other countries and for medical devices. In the United States and the European Union (EU), review and approval by an institutional review board (IRB) or ethics committee (EC) prior to trial initiation is also required.

The primary objectives phase I trials seek to accomplish are to assess safety, tolerability, pharmacokinetics, and to determine the MTD (maximum tolerated dose). They typically are small trials ranging from 20 to 80 subjects and are relatively short, only lasting 6 months or less from initiation to trial completion. In general, phase I trials are performed in normal healthy volunteers (though the guinea pig zero effects may be operative [1]). However, in special cases such as drugs for life-threatening diseases such as cancer, AIDS (acquired immodefiency syndrome), amyotropic lateral sclerosis (ALS), or where the drug is a combination of two already approved drugs, patients may constitute the initial trial population.

The importance of FIM trials is crucial to the progression of the new drug molecule. Identifying a safe dose and dosing regimen is vital, and in doing so the utmost precaution and care must be taken during this stage through strict monitoring.

## 2 PURPOSE AND OBJECTIVES

As stated above, the main objectives of phase I trials are to assess tolerability, monitoring the side effects in relation to increase of the doses, and confirm the maximum tolerated dose. Additionally, establishing a safety profile is paramount. These are three of the commanding goals, providing a firm basis for investigators to determine further testing and study design strategies. It is almost certainly the most critical "go/no-go" decisive point.

Other important aims are to obtain adequate information of the drug's basic pharmacokinetics (PK) such as absorption, distribution, metabolism, and excretion (ADME). Pharmacokinetics refers to the behavior of the drug in the body. Absorption is the process of a substance entering the body. Distribution is the dispersion or dissemination of substances throughout the fluids and tissues of the body. Metabolism is the transformation of the substances and its daughter metabolites. Excretion is the elimination of the substances from the body. This data gives rise to analyzing the pharmacological effects of the drug in the human body and developing a sufficient PK model to simulate exposure and response for intended dosing regimens, incorporating variability [2]. A good dose–response curve derived from PK analysis is greatly desired. Additionally, information is sought out to determine the drug's

bioavailability. Bioavailability is defined as the ability to deliver the drug in a usable form to the disease target. Phase I studies are also intended to determine whether the drug is best delivered orally, or by injection, or through the skin and by what regimen.

Furthermore, FIM trials study the drug's structure–activity relationship (SAR) with previously evaluated compounds and mechanism of action (MOA), which provide the basis to investigate biological phenomena or disease processes. However, data regarding efficacy [pharmacodynamics (PD)], though a highly desired aspect of the drug, is rarely gathered during phase I. Some studies will often shed enough early light on possible efficacy to support the decision on further continuance of the study.

## 3 PHASE I TRIAL DESIGN

More recently, it has become the common practice to divide phase I trials into two phases or separate stages, phase Ia and phase Ib. Phase Ia studies include the first dose in humans and are typically short-term single-dose studies to confirm safety before beginning a larger, more extensive trial. The studies on average are usually comprised of six cohorts, with about seven subjects per cohort. The main design usually entails escalating the dose with each new cohort while stringently monitoring for safety, with a final objective of establishing the MTD and a solid PK profile. Oftentimes a placebo is included to provide a more accredited safety evaluation.

Phase Ib studies are typically more comprehensive repeat-dose studies with the same goals being safety, tolerability, and PK on a repeat-dose level in order to assess the drug's therapeutic effect. Other frequent questions addressed in phase Ib studies include food effects and gender differences.

Phase I trials are also different from subsequent trials in that not only are they not conducted at a sponsor's facility (no clinical trials are), but they are almost always conducted by and at contract research organizations (CROs) that have this as their business.

## 4 DESIGN TYPES

When planning a clinical trial, the design and determination of the dose range depends on the MTD established in preclinical studies. The most common trial design is dose escalation/ascending. This study involves the gradual increase of drug dosages to determine the amount that delivers the best balance of high efficacy and acceptable side effects. It is basically viewed as a building block design. The primary objective is to determine MTD. The MTD is defined as the highest dose that is tolerated with no adverse side effects observed. If the dose is not tolerated, then the dose just below is the MTD. Typically, one may dose the subject, look for adverse events, and if none are present, increase dose and repeat. One may have a fixed dose increment or an even more risky approach is to bounce around until the MTD is established. One preferred way to determine AEs (adverse events) is by comparison to a placebo group. Conducting thorough dose-ranging and dose–response studies early in product development reduces the possibility of later failed phase II or III

studies. Of drugs tested in phase I, 50–70% are abandoned because of problems with safety or efficacy [3]. At the end of each phase of dosing, typically a data safety monitoring board (DSMB) meeting is held to decide if the drug is safe enough to continue, and if yes, typically one will want to recruit more patients. When analyzing a dose-escalating study the following should be included:

Comparison between treatment groups (sometimes several doses polled).
Comparison of active groups versus placebo.
Comparison of before and after dosing (temporal factors).
Evaluate AE incidence and complete a summarization.

In some instances, mainly for bioequivalence studies, a crossover design is used. The ICH E9 [4] guideline defines the crossover design as a study in which each subject is randomized to a sequence of two or more treatments and hence acts as his own control for treatment comparisons. This simple maneuver is attractive primarily because it reduces the number of subjects and usually the number of assessments needed to achieve a specific power, sometimes to a marked extent. In the simplest $2 \times 2$ crossover design, each subject receives each of two treatments in randomized order in two successive treatment periods; often separated by a washout period. It is the type of experiment where experimental units are given several treatments in succession. The order of treatments ought to be set by some random procedure and each subject receives both treatments. In this type of trial design comparisons within subjects are performed.

Advantages of a crossover study are that it is useful for reversible effects (e.g., bioequivalence), fewer subjects are needed, and more precise comparisons are obtained.

If applicable, one can have more than two treatments in which an extended crossover design may be applied (which the FDA has been leaning more toward since 2000). An extended crossover design includes two treatments with four periods.

Crossover designs are not flawless, however, and contain a number of problems that can invalidate the concluding results. The chief difficulty concerns carryover, that is, the residual influence of treatments in subsequent treatment periods; therefore a washout period of at least 6 half-lives is desired. The main cause of carryover is that the treatment effects lingering in PK studies, because the washout is too short thus leading to overlap of dose effects.

When analyzing a crossover study the following should be taken into consideration:

$Y$ = period + sequence + treatment + subject (sequence) + carryover + error.
Subject is usually considered a random effect.
In a $2 \times 2$ crossover, sequence is confounded with carryover—if there is a significant carryover, use the first period only.
Comparisons are intrasubject (treatment comparisons).

When analyzing a bioequivalence (BE) crossover study the following should be taken into consideration:

BE is typically testing utilizing the crossover design.

Response is area under the curve (AUC) and the peak concentration of the drug ($C_{max}$).

Key endpoint is the ratio of treatment A AUC over treatment B AUC.

A and B are equivalent if 90% confidence interval for ratio lies entirely between (0.8, 1.25) 80% and 125%.

Perform analysis on log AUC and log $C_{max}$.

## 5 PHASE I TRIAL CHARACTERISTICS

The initial (true FIM) administration of a new drug must be undertaken with some circumspection. Nonclinical (animal) safety studies give good predictions of clinical safety and expected toxicity almost all of the time, with application of current scaling factors for species difference providing confidence that initial doses will present minimal risk. Still, in normal volunteers there are exceptions—particularly highly humanized proteins for which animals lack the mechanisms for response, and therefore provide no assurance of safety (as with the TGN1412 drug trial).

A similar situation occurs when phase I trials are performed in patients (or, indeed, with the first trial in patients if earlier trials were done in normal volunteers). In the latter (patient) treatment case, one should carefully consider which organ, metabolic, or protective systems may already be compromised—and therefore for which one must be very attentive to even weak indication of adverse effect [as with the FIAU (fialuridine) clinical trial].

In either of these cases, or indeed in almost any true FIM dosing, it is generally wise to offset the dosing of the initial cohort—that is, separate dosing of individual cohort members by an appropriate period of time.

Both types of these designs share many similar characteristics indicative of phase I trials. Both serve to provide the information to accurately evaluate safety, tolerability, MTD, and PK. Additionally, they are relatively short and small trials, comprised of about 20–80 subjects usually healthy volunteers, with the exceptions for trials dealing with life-threatening diseases such as HIV/AIDS, cancer, ALS, and the like. The FDA explicitly draws the distinction between a healthy volunteer and a subject or patient. A healthy volunteer is defined as a healthy person who agrees to participate in a clinical trial for reasons other than medical and receives no direct health benefit from participating. Whereas a human subject is defined as an individual who is or becomes a participant in research, either as a recipient of the test article or as a control. A subject may be either a healthy human or a patient (21 CFR 50.3).

Whether it be a healthy volunteer or a patient, they must satisfy the subject/patient criteria outlined in the sponsor's protocol and must provide informed consent. Since phase I trials incorporate dosing a person with an investigational new drug, extremely close monitoring for safety and determination of a tolerated dose is crucial.

A placebo is almost always included as a standard component in phase I trials as well. The main advantage of using a placebo is to serve as a control and decrease bias in an attempt to make some sort of quantitative assessment of the drug's efficacy. A disadvantage to using a placebo is mainly seen in trials for life-threatening illnesses, and in some such cases a placebo is not required. A common approach is

to include a single placebo subject in each cohort. By the end of an ascending dose trial, enough placebo subjects have been accumulated to allow for an estimation of placebo (and nocebo) effects.

The issue of blinding in clinical trials is imperative for the prevention of bias and integrity of the data. The ICH E9 [4] definition of blinding is as follows:

> Blinding or masking is intended to limit the occurrence of conscious and unconscious bias in the conduct and interpretation of a clinical trial arising from the influence that the knowledge of treatment may have on the recruitment and allocation of subjects, their subsequent care, the attitudes of subjects to the treatments, the assessment of endpoints, the handling of withdrawals, the exclusion of data from analysis, and so on. The essential aim is to prevent identification of the treatments until all such opportunities for bias have passed.

Phase I trials are usually double blind. ICH E9 [4] definition of double blind:

> A double-blind trial is one in which neither the subject nor any of the investigator or sponsor staff involved in the treatment or clinical evaluation of the subjects are aware of the treatment received. This includes anyone determining subject eligibility, evaluating endpoints, or assessing compliance with the protocol. This level of blinding is maintained throughout the conduct of the trial, and only when the data are cleaned to an acceptable level of quality will appropriate personnel be unblended.

## 6 CRITICAL PARAMETERS TO MEASURE

### Safety

A. Clinical Examinations
   1. Physical
   2. Vital signs (usually considered as part of the physical examination)
   3. Height and weight (state of dress is usually specified, e.g., socks)
   4. Neurological or other specialized clinical examinations

B. Clinical Laboratory Examinations
   1. Hematology
   2. Clinical chemistry
   3. Urinalysis
   4. Virology (viral cultures or viral serology)
   5. Immunology or immunochemistry (e.g., immunoglobins, complement)
   6. Serology
   7. Microbiology (including bacteriology and mycology)
   8. Parasitology (e.g., stool for ova and protozoa)
   9. Pulmonary function tests (e.g., arterial blood gas)
   10. Other biological tests (e.g., endocrine, toxicology screen)
   11. Stool for occult blood (specify hemoccult or Guaiac method)
   12. Skin tests for immunologic competence
   13. Medicine screen (usually in urine) for detection of illegal or nonprotocol-approved medicines
   14. Bone marrow examination

15. Gonadal function (e.g., sperm count, sperm motility)
16. Genetics studies (e.g., evaluate chromosomal integrity)
17. Stool analysis using in vivo dialysis

C. Probe for Adverse Reactions

D. Psychological and Psychiatric Tests and Examinations
   1. Psychometric and performance examinations
   2. Behavioral rating scales
   3. Dependence liability

E. Examinations Requiring Specialized Equipment (selected examples)
   1. Audiometry
   2. Electrocardiogram (EKG)
   3. Electroencephalogram (EEG)
   4. Electromyography (EMG)
   5. Stress test
   6. Endoscopy
   7. Computed tomography (CT) scans
   8. Ophthalmological examination
   9. Ultrasound
   10. X rays
   11. Others

## 7 PK PARAMETERS TO DERIVE

The purpose of human pharmacokinetic studies is to examine the rate of absorption, distribution, metabolism, and excretion of a drug. Findings from these studies describe how the drug travels through the body and where and how it is eliminated. PK data allows for the detection of drug levels in human blood/urine samples. After PK information is obtained, a dose–response curve should be plotted that describes the change in effect on the subject caused by differing levels of dose exposure to a stressor. Studying dose–response, and developing dose–response models, is central to determining the safe and hazardous levels and dosages for the drug. A good dose–response curve is obviously highly desired.

The dose–response curve (Fig. 1) defines the relationship between dose and response based on the following assumptions: (1) response increases as dose increases; (2) there is a threshold dose—a dose below which there is no effect. The quality of results depend on the placement of time points; therefore you want to see at least a 24-hour profile, with two or three time points for each component. However, sometimes limitations arise such as the ability to draw blood, the patience of the study population, and sometimes the study design or medical needs.

Another consideration taken when evaluating PK data is the food effect, which assesses how food affects the absorption of the drug; but this is only tested for with orally administered compounds. Other vital paramenters to be drawn and evaluated are listed and defined as follows:

- $AUC_{0-\infty}$ represents the total amount of drug absorbed by the body, irrespective of the rate of absorption. This is useful when trying to determine whether two

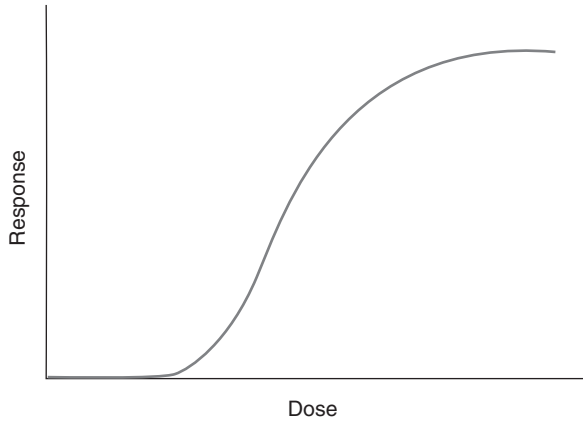


***FIGURE 1*** *Dose–response curve.*

formulations of the same dose (e.g., a capsule and a tablet) release the same dose of drug to the body.

- $AUC_{0-}T$ represents the average concentration over a time interval, AUC/$T$.
- $C_{max}$ represents the peak concentration of the drug in plasma.
- $T_{max}$ represents the duration of time to reach the peak concentration of drug in plasma, beginning from administration of the drug.
- $T_{1/2}$—half life, or the time it takes for half of the administered or absorbed dose to be cleared or metabolized.
- $V_0$—central and peripheral. The theoretical volume in which the drug is homogeneously distributed and is basically dependent upon the lipid or water solubility of the drug and its particular affinity for given tissues or structures.
- Clearance—the volume of plasma that is completely cleared of drug per unit time.
- MRT (mean residence/residual time)—The average total time molecules of a given dose spend in the body. Thus, this can only be measured after instantaneous administration.

## 8 REGULATORY REQUIREMENTS AND ISSUES

In order to proceed into FIM or phase I trials, permission must be obtained from both the national drug regulatory authority in the country where the trial will take place [in the United States it is the Food and Drug Administration (FDA); in the European Union it is the European Medicines Agency (EMEA)], and the national ethics committees, the institutional review board (IRB) in the United States. The IRB is an independent group officially authorized to approve, monitor, and review biomedical and behavioral research involving humans with the objective to protect the rights and welfare of the subjects. In the United States, the FDA and Health and Human Services (HHS) regulations have empowered IRBs to approve, require modifications to gain approval, or disapprove research. An IRB performs critical monitoring tasks for research conducted on human subjects that are scientific,

ethical, and regulatory. IRBs are required to have at least five members with varying backgrounds to promote complete and adequate review of research activities commonly conducted by the institution. The purpose of an IRB review is to assure, both in advance and by periodic review, that appropriate steps are taken to protect the rights and welfare of humans participating as subjects in a research study. The review will cover materials such as the protocol, informed consent documents, and advertisements, with special attention paid to trials that may include vulnerable subjects, such as pregnant women, children, prisoners, the elderly, or persons with diminished comprehension.

The U.S. FDA requires an IND application that includes all the nonclinical and technical data to review. Typically, sponsors will request a pre-IND meeting with the FDA to discuss safety issues related to the proper identification, strength, quality, purity, or potency of the investigational drug, as well as to identify potential clinical hold issues. The pre-IND meeting should focus on the specific questions related to the planned clinical trials. The meeting should also include a discussion of various scientific and regulatory aspects of the drug as they relate to safety and/or potential clinical hold issues. The IND application contents fall primarily into three categories: animal pharmacology and toxicology studies, chemistry and manufacturing information, and clinical protocols and investigator information. Animal pharmacology and toxicology studies are comprised of preclinical data to allow an evaluation as to whether the NME is reasonably safe for initial testing in humans. Also included are any previous experience with the drug in humans (often foreign use). Chemistry and manufacturing information includes information pertaining to the chemical composition, manufacturing methods, stability, and controls used for manufacturing the drug substance and the drug product. The chemical stability and activity of the product must also have been tested. This information is reviewed to ensure that the company can adequately produce and supply consistent and active batches of the drug. Clinical protocols and investigator information consist of detailed protocols for proposed clinical studies to determine whether the initial-phase trials will expose subjects to unnecessary risks. Also, information on the qualifications of the clinical investigators who are to oversee the administration of the experimental compound is included in order to assess whether they are qualified to fulfill their clinical trial duties. An investigator's brochure (IB), which is a document intended to educate the trial investigators of the pertinent facts concerning the trial drug they need to know to conduct their study with the least hazard to the subjects is also submitted. Furthermore, commitments to obtain informed consent from the research subjects, to obtain review of the study by an institutional review board (IRB), and to adhere to the investigational new drug regulations are included. If the sponsor has not received a response from the FDA within 30 days, the trial may begin.

## REFERENCES

1. Helms, R., Ed. (2002), *Guinea Pig Zero: An Anthology of the Journal for Human Research Subjects*, Garret County Press, New Orleans.
2. Chien, J. Y., Friedrich, S., Heathman, M. A., et al. (2005), Pharmacokinetics/pharmacodynamics and the stages of drug development: Role of modeling and simulation, *AAPS J.*, 7(3), E544–549.

3. Lee, C.-J., Lee, L. H., Wu, C. L., et al. (2006), *Clinical Trials of Drugs and Biopharmaceuticals*, CRC Press, Boca Raton, FL.
4. ICH E9 (2005), *Statistical Principles for Clinical Trials*. International Conference on Harmonization.

## BIBLIOGRAPHY

Gallin, J. I., and Ognibene, F. P., Eds. (2007), *Principles and Practices of Clinical Research*, 2nd ed., Academic Press, Burlington, MA.

Green, S., Benedetti, J., and Crowley, J. (2002), *Clinical Trials in Oncology*, 2nd ed., Chapman & Hall/CRC Press, Boca Raton, FL.

Machin, D., Day, S., and Green, S. (2004), *Textbook of Clinical Trials*, Wiley, Hoboken, NJ.

O'Grady, J., and Joubert, P. (1997), *Handbook of Phase I/II Clinical Drug Trials*, CRC Press, Boca Raton, FL.

O'Grady, J., and Linet, O. (1990), *Early Phase Drug Evaluation in Man*, CRC Press, Boca Raton, FL.

Rang, H. P. (2005), *Drug Discovery and Development: Technology in Transition*, Churchill Livingstone Elsevier, Oxford, UK.

Stone, J. (2006), *Conducting Clinical Research*, Mountainside MD Press, Cumberland, MD.

U. S. Food and Drug Administration, Center for Biologics Evaluation and Research. (2001), Guidance for Industry IND Meetings for Human Drugs and Biologics Chemistry, Manufacturing, and Controls Information.

# *Phase II Clinical Trials*

**Say-Beng Tan[1–4] and David Machin[3–5]**

[1]*Singapore Clinical Research Institute, Singapore*
[2]*Duke–NUS Graduate Medical School*
[3]*Division of Clinical Trials and Epidemiological Sciences, National Cancer Centre, Singapore*
[4]*Clinical Trials and Epidemiology Research Unit, Singapore*
[5]*Children's Cancer and Leukaemia Group, University of Leicester, Leicester, United Kingdom*

**Contents**

1 Overview
  1.1 Phase II Trials
2 Planning a Phase II Trial
  2.1 Choice of Endpoint
  2.2 Eligibility
  2.3 Choice of Design
3 Single-Stage Designs
  3.1 Fleming–A'Hern
4 Two-Stage Designs
  4.1 Gehan
  4.2 Simon Optimal and Minimax
5 Phase II Trials with Survival Endpoints
  5.1 Case and Morgan
6 Efficacy and Toxicity in Phase II Trials
  6.1 Bryant and Day
7 Bayesian Approaches
  7.1 Motivation
  7.2 Overview of Bayesian Approaches in the Context of Phase II Trials
  7.3 Bayesian Single and Dual Threshold
8 Randomized Phase II Trials
  8.1 Simon, Wittes, and Ellenberg (SWE)

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

9 Trial Conduct and Reporting
9.1 Trial Conduct
9.2 What to Report
10 Concluding Remarks
References

## 1 OVERVIEW

New compounds are continually being developed with the intended objective that some of these will prove useful in the treatment of human diseases. Any new compound has to undergo a rigorous development process before it can be introduced for standard clinical use. New compounds are first tested for safety in laboratory and animal studies whose objectives are the characterization of the drug's pharmacology, toxicology, metabolism, and other properties.

The new compound is then tested in human subjects in phase I studies that, if successful, are the first of an eventual series in humans. These studies are designed to determine the metabolic and pharmacological actions of the drug, the side effects associated with increasing doses (to establish a safe dose range), and, if possible, to gain early evidence of activity. The focus is typically on determining the toxicity profile of the drug and on finding a potentially therapeutic effective dose.

After a suitable dose has been established in phase I trials, phase II trials are conducted with the main objective being to evaluate the drug's activity for patients with a particular disease or condition, as well as to determine the level of (short-term) side effects and possible risks associated with the use of the drug.

If sufficient preliminary evidence of effectiveness is found in these phase II trials, suggesting worthwhile potential therapeutic efficacy of the drug, the drug may be further evaluated in randomized phase III trials, where it will be compared with the current standard treatment in a larger number of patients.

### 1.1 Phase II Trials

Although there is some variation in terminology and objectives depending on the disease or condition in question, phase II trials are usually single-armed studies with the objective being to investigate the antidisease activity of the new therapeutic regimen. Most of these trials evaluate at least one new treatment relative to a standard. So they are inherently comparative, even though the "standard" treatment information is usually historical and is not obtained prospectively as part of the phase II study itself.

In certain circumstances, there may be several candidate agents for phase II testing; the problem here is to select that which has the most potential to be effective and hence be tested in a phase III trial. This leads to randomized phase II trial designs, which seek to reduce the apparent variability in response rates observed in different studies of the same compound. Some factors that contribute to the variability include patient type, definition of response, interobserver variability in response evaluation, drug dosage and schedule, reporting procedures, and sample

size. Patients in randomized phase II studies are randomized to one of several experimental treatments, but the limited sample size of these trials does not provide sufficient statistical power to make reliable treatment comparisons.

In areas such as oncology, phase II trials often focus on both safety and efficacy of the new therapeutic regimen. Safety is usually assessed in terms of toxicity rates but is not always part of the formal design process. Efficacy is typically measured using tumor response, often the percentage decrease in tumor size compared with that before treatment commences, and so patients enrolled in phase II trials need to have measurable disease.

***Example: Phase II Trial—Carcinosarcoma of Female Genital Tract*** Van Rijswijk et al. [1] conducted a phase II trial in 48 women with carcinosarcoma of the genital tract. Although the activity of the combination of cisplatin, doxorubicin, and ifosfamisde was established (overall response rate 56%), they did not recommend this treatment combination but suggested that combinations "with more favorable toxicity profiles should be explored."

## 2 PLANNING A PHASE II TRIAL

### 2.1 Choice of Endpoint

As already indicated, phase II trials seek to assess whether a new regimen is active enough to warrant a comparison of its efficacy with the standard treatment regimen in a phase III trial. Thus, appropriate endpoints need to be chosen to allow for such an assessment to be made. For example, in human immunodeficiency virus (HIV) research, suitable endpoints might include measures of viral load or immune function. In cardiovascular disease, we might look at blood pressure or lipid levels.

In certain situations, there is not necessarily an obvious measure to take. For example, in oncology, although one may regard tumor shrinkage as a desirable property of a cytotoxic drug when given to a patient, it is not immediately apparent how this is to be measured. Were every tumor of regular spherical shape, the direction in which it is measured is irrelevant. Furthermore the diameter, a single dimension, leads us immediately to the volume of the tumor. However, no real tumor will comply with this ideal geometrical configuration, and this has led to measures such as the product of the two largest (perpendicular) diameters to describe the tumor and then a reduction in this product to indicate response. Precisely what is the best measure to assess tumor shrinkage has been discussed by an international panel and reported in detail by Therasse et al. [2].

More generally, they offer guidelines to encourage more uniform reporting of outcomes particularly for clinical trials. Investigators of future trials may argue about the fine details, and no doubt in time these guidelines will need revision, but they would be foolish to ignore these recommendations when conducting and subsequently reporting their study.

If there are "justifiable" reasons why other criteria should be used, or the recommendations cannot be followed for whatever reason, then before the study commences, these should be reviewed by the investigating team. There is little point in

conducting a study using measures not acceptable to other groups, including referees for the clinical journals, as little note will then be taken of the results. The best option is to follow the guidelines for the primary endpoint, use the "local" measures for secondary reporting, and contrast the two in any discussion.

Sometimes it may be that the true endpoint of interest is difficult to assess for whatever reason. In this case, a surrogate may be sought. For example, when investigating the possibilities of a novel marker for prognosis, it may be tempting to use disease-free survival (DFS) as a "surrogate endpoint" for the overall survival (OS) time of patients with the cancer concerned. The reason being that for many cancers, relapse occurs well before death, and so the evaluation of the marker can occur earlier in time than would the case if OS was to be observed.

More generally, a surrogate endpoint is a biomarker (or other indicator) that is intended to substitute for a (often) clinical endpoint and predict its behavior. If a surrogate is to be used, then there is a real need to ensure that it is an appropriate surrogate for the (true) endpoint of concern.

## 2.2 Eligibility

Common to all phases of clinical trials is the necessity to define precisely who are eligible subjects. This definition may be relatively brief or complex depending on the substance under test.

At the very early stages of the process, it is particularly understandable that great care is taken in subject and particularly patient choice. In these situations, when relatively little is known about the compound, all the possible adverse eventualities have to be considered. This usually results in quite a restricted definition as to those that can be recruited.

Once the possibility of some activity (and hence potential efficacy) becomes indicated, then there is at least a prospect of therapeutic gain for the patient. In this case, the investigators may expand the horizon of eligible patients but simultaneously confine them to those in which a measurable response to the disease can be ascertained.

***Example: Eligibility for Phase II Trial—Gemcitabine in Nasopharyngeal Carcinoma*** Foo et al. [3] specify that patients were to have histologically confirmed undifferentiated carcinoma arising from the nasopharynx, *bidimensionally measurable* disease not within any prior radiotherapy fields, be between 18 and 75 years, with an Eastern Cooperative Oncology Group (ECOG) performance status (PS) $< 2$. In addition there were seven clinical chemistry limits that had to be satisfied before inclusion was possible.

## 2.3 Choice of Design

There are a relatively large number of alternative designs for phase II trials. These include single-stage designs in which a predetermined number of patients are recruited, and two-stage designs in which patients are recruited in two stages and the move to stage 2 being consequential on the results observed in stage 1. Multistage designs have also been proposed, but the practicalities of having several

decision points have limited their use because of the inherent further delays involved with each extra stage. Most phase II trials are of a single-arm, noncomparative design. However, randomized phase II selection designs in which the objective is to select only one, the "best," of several agents tested simultaneously, are strongly recommended in some situations.

Although the usual (single) endpoint for phase II studies typically involves some binary measure of activity, designs are available when activity is assessed by survival time, and when the dual endpoints of activity (response) and acceptable levels of toxicity are stipulated by the design.

With such a plethora of different options for phase II designs, it is clearly important that the investigators choose the design that is best for their purpose. In some cases the choice will be reasonably clear. For example, if one has several compounds to test at the same time, then the randomized selection design will be preferred to (say) a series of parallel single-arm studies. In other circumstances, the patient pool may be very limited and a key consideration will be the maximum numbers of patients that might have to be recruited. Features to guide investigators in their choice are summarized in Table 1.

In this chapter, we will briefly discuss a number of different phase II designs. However, a detailed discussion of the designs and sample size tables is beyond the scope of this chapter. Interested readers are referred to the *Sample Size Tables for Clinical Studies* [4]. The accompanying software also allows for the implementation of all the designs we discuss.

**TABLE 1 Comparative Properties of Alternative Phase II Designs**

| Single Stage | No Stopping Rules | |
|---|---|---|
| Fleming–A'Hern | Sample size fixed. | Size determined at the design stage. |
| Randomized | Sample size fixed. | Size determined at the design stage and depends on the number of compounds under test. |
| **Two Stage** | **Allow Early Termination** | |
| Gehan | Maximum sample size unknown. | Final sample size depends on the number of responses in stage 1. |
| Simon—Optimal | Maximum sample size fixed. | Stage 1 sample size chosen to ensure inactive compound does not go to stage 2. |
| Simon—Minimax | Maximum sample size fixed. | Designed for maximum sample size to be a minimum. |
| Tan–Machin | Maximum sample size fixed. | Stage 1 sample size chosen to ensure inactive compound does not go to stage 2. |
| **Two Stage** | **Allow Early Termination** | **Dual Endpoint** |
| Bryant–Day—Optimal | Maximum sample size fixed. | Stage 1 sample size chosen to ensure inactive or too toxic compound does not go to stage 2. |
| **Two Stage** | **Allow Early Termination** | **Survival Endpoint** |
| Case–Morgan | Maximum sample size fixed. | Sample size chosen to minimize either expected duration of accrual or expected total study length for the trial. |

## 3 SINGLE-STAGE DESIGNS

In planning a phase II trial, there are several options to consider, including the number of stages within each design, alternative designs for essentially the same situation, as well as randomized designs. In this section, we look at the Fleming–A'Hern single-stage design. For this design, the endpoint is some measure of anti-disease activity and this translates into a measure of response.

A key advantage of such single-stage designs is that once the sample size has been determined, this translates directly into the number of patients that need to be recruited. This contrasts with, for example, two-stage designs in which the total numbers eventually recruited depends on the response rate in those recruited to stage 1.

In considering the design of a phase II trial of a new drug, the investigators will usually have some knowledge of the activity of other drugs for the same disease. The anticipated response to the new drug is therefore compared, at the planning stage, with the observed responses to other therapies. This may lead to the investigators prespecifying a response probability that, if the new drug does not achieve, results in no further investigation. They might also have some idea of a response probability that, if achieved or exceeded, would certainly imply that the new drug has activity worthy of further investigation, perhaps in a phase III randomized trial to determine efficacy. If a phase II trial either fails to identify efficacy or overestimates the potential efficacy, there will be adverse consequences for the next stage of the development process.

### 3.1 Fleming–A'Hern

The Fleming [5] single-stage design for phase II trials recruits a predetermined number of patients to the study, and a decision about activity is obtained from the number of responses among these patients.

To use the design, the investigators first set the largest response proportion as $\pi_0$, which, if true, would clearly imply that the treatment does not warrant further investigation. They then judge what is the smallest response proportion, $\pi_{New}$, which would imply that the treatment certainly warrants further investigation. This means that the one-sided hypotheses to be tested in a phase II study are: $H_0$: $\pi \leq \pi_0$ versus $H_A$: $\pi \geq \pi_{New}$, where $\pi$ is the actual probability of response that is to be estimated at the close of the trial.

In addition to specifying $\pi_0$ and $\pi_{New}$, it is necessary to specify $\alpha$, the probability of rejecting the hypothesis $H_0$: $\pi \leq \pi_0$ when it is in fact true, together with $\beta$, the probability of rejecting the hypothesis $H_A$: $\pi \geq \pi_{New}$ when that is true. Note that $\alpha$ is often termed the *test size* or *significance level* and $1 - \beta$ is referred to as the *power*.

With these inputs, Fleming [5] details the appropriate sample size to be recruited, along with the minimum number of responses that would need to be observed in order for the null hypothesis to be rejected.

However, A'Hern [6] repeated the calculations for sample size using exact binomial probabilities and, in general, these are greater than those of Fleming, although the two calculations are often in close agreement. As a consequence, his calculations should supersede those of Fleming.

To implement the design, the appropriate number of patients is recruited, and once all their responses are observed, the response rate (and corresponding confidence interval) is calculated. A decision with respect to efficacy is then made.

***Example: Fleming–A'Hern Design—Sequential Hormonal Therapy in Advanced and Metastatic Breast Cancer*** Iaffaioli et al. [7] used A'Hern's design for two phase II studies of sequential hormonal therapy with first-line anastrozole (study 1) and second-line exemestane (study 2) in advanced and metastatic breast cancer. For study 1 they set $\alpha = 0.05$, $\beta = 0.1$, $\pi_0 = 0.5$, and $\pi_{New} = 0.65$. With these inputs, the design specifies a sample size of 93, with 55 being the minimum number of responses required for a conclusion of "efficacy." The study in the end recruited 100 patients, with 8 complete responses and 19 partial responses observed. These give an estimated response rate of 27% with 95% confidence interval (CI) 19.3–36.4% calculated using the method described by Newcombe and Altman [8]. This is much lower than the desired minimum of $\pi_{New} = 65\%$. For study 2, the investigators set $\alpha = 0.05$, $\beta = 0.1$, $\pi_0 = 0.2$, and $\pi_{New} = 0.4$, giving rise to a sample size of 47 with a minimum of 15 responses required. The trial eventually recruited 50 patients, with 1 complete response and 3 partial responses observed. These give an estimated response rate of 8% (95% CI 3.2–18.8%). Again this is much lower than the desired minimum of 40%. As a consequence, neither drug should be recommended for testing in a phase III trial.

## 4 TWO-STAGE DESIGNS

In many situations, investigators may be reluctant to embark on a single-stage phase II trial requiring a (relatively) large number of patients exposed to a new and uncertain therapy. In such circumstances, a more cautious approach may be to conduct such a study, but in a series of stages and review progress at the end of each stage. In two-stage designs, patients are recruited in two stages, and the move to stage 2 is consequential on the results observed in stage 1. The main advantage of such a design is that the trial may stop, after relatively few patients have been recruited, should the response rate appear to be (unacceptably) low. The disadvantage is that the final number of patients required is not known until after stage 1 is complete.

### 4.1 Gehan

In the approach suggested by Gehan [9], a minimum requirement of efficacy, $\pi_{New}$, is set and patients are recruited in two stages. If no responses are observed in stage 1, patients are not recruited for stage 2. On the other hand, if one or more responses are observed, then the size of the recruitment to the second stage depends on their number.

To implement the design, the appropriate number of patients is recruited in stage 1, and, once all their responses are observed, a decision whether or not to proceed to stage 2 is taken. If stage 2 is implemented, then once recruitment is complete and all assessments made, the response rate (and corresponding CI) is calculated. A decision with respect to efficacy is then made. If stage 2 is not activated, the response

rate (and CI) can still be calculated for the stage 1 patients despite failure to demonstrate efficacy. This procedure applies to all the two-stage designs we will discuss.

***Example: Gehan Design—Dexverapamil and Epirubicin in Nonresponsive Breast Cancer*** Lehnert et al. [10] used the Gehan design for a phase II trial of the combination dexverapamil and epirubicin in patients with breast cancer. For stage 1 they set $\pi_0 = 0.2$ and $\beta = 0.05$ obtaining a stage 1 sample size of 14. Of these 14 patients, 3 responses were observed, resulting in a further 9 patients needing to be recruited for stage 2. Finally a total of 4 (17.4%) responses was observed from the total of $14 + 9 = 23$ patients with 95% CI for $\pi$ from 7 to 37%.

### 4.2 Simon Optimal and Minimax

In the approach suggested by Simon [11], patients are recruited in two stages and there are two alternative designs.

One is *optimal* in that the expected sample size is minimized if the regimen has low activity. This implies that an important focus is to ensure that as few patients as possible receive what turns out to be an ineffective drug by not continuing to stage 2 in these circumstances. In this context, "expected" means, the average sample size that would turn out to have been used, had a whole series of studies been conducted with the same design parameters in situations where the true activity is the same.

The other, the *minimax* design, minimizes the maximum sample size for both stages combined, that is, the sum of patients required for stage 1 and stage 2, is chosen to minimize the maximum trial size within the parameter constraints as set by the design.

For either design, the designs imply that the one-sided hypotheses to be tested in a phase II study are $H_0: \pi \leq \pi_0$ versus $H_A: \pi \geq \pi_{New}$, where $\pi$ is the actual probability of response, and $\pi_0$, $\pi_{New}$ are as defined before. It is also necessary to specify $\alpha$ and $\beta$ as for the Fleming–A'Hern design.

The trial then proceeds by recruiting $n_{S1}$ patients in stage 1 from which $r_{S1}$ responses are observed. Then a decision is made to recruit $n_{S2}$ patients to stage 2 if $r_{S1} > R_{S1}$, where $R_{S1}$ is the minimum number of responders required as indicated by the design. Otherwise the trial is closed at the end of stage 1. At the end of the second stage, the drug is rejected for further use if a predetermined total number of responses are not observed.

***Optimal versus Minimax*** In determining which design to use, the minimax design may be more attractive than the optimal design when the difference in anticipated total sample size is small and the patient accrual rate is low. The optimal designs have smaller stage 1 than the minimax designs, and so this smaller stage 1 reduces the number of patients exposed to an inactive treatment if this turns out to be the case. In cases where the patient population is very heterogeneous, however, a very small stage 1 may not be desirable because the first patients entered into the study may not be entirely representative of the wider eligible population. In this case, a larger stage 1 may be preferred and the minimax design chosen.

***Example: Simon Minimax Design—Gemicitabine in Metastatic Nasopharyngeal Carcinoma*** In a phase II trial of gemicitabine in previously untreated patients with metastatic nasopharyngeal carcinoma (NPC), Foo et al. [3] utilized the Simon minimax design. With $\alpha = 0.05$, $\beta = 0.2$, $\pi_0 = 0.1$, and $\pi_{New} = 0.3$, the design gives:

*Stage 1* Sample size of 15 patients: If responses less than 2, stop the trial and claim gemicitabine lacks efficacy.

*Stage 2* Overall sample size of 25 patients for both stages. Hence 10 more patients were to be recruited. If the total responses for the two stages combined is less than 6, stop the trial as soon as this is evident and claim gemicitabine lacks efficacy.

Once the phase II trial was conducted, the investigators observed 3 and 4 responses in stages 1 and 2, respectively, giving an estimated response rate of 7/25 or 28% (95% CI 14–48%).

***Example: Simon Optimal Design—Gemicitabine in Metastatic Nasopharyngeal Carcinoma*** Suppose in the phase II trial of gemicitabine in metastatic NPC designed by Foo et al. [3] that the Simon optimal, rather than the minimax design, had been planned. With the same design values, we want to investigate what difference this makes to the patient numbers and responses required.

Again we have $\alpha = 0.05$, $\beta = 0.2$, $\pi_0 = 0.1$, and $\pi_{New} = 0.3$, but the use of the optimal design now gives the following results:

*Stage 1* Sample size of 10 patients: If responses less than 2, stop the trial and claim gemicitabine lacks efficacy.

*Stage 2* Overall sample size of 29 patients for both stages. Hence 19 more patients are to be recruited. If total responses for two stages combined is less than 6, stop the trial as soon as this is evident and claim gemicitabine lacks efficacy.

In this case, for the same design parameters, the optimal design has five fewer patients in stage 1 of the design, but four more patients if the trial goes on to complete stage 2, than the corresponding minimax design. The number of responses to be observed are the same, in each stage, for both designs, however.

***Example: Simon Minimax Design—Paclitaxel for Unresectable Hepatocellular Carcinoma*** Chao et al. [12] state in their methods that a Simon [11] design was used in which if the response rate was $\leq 3$ of 19 in the first stage, then the trial would be terminated. The authors set $\alpha = 0.1$, $\beta = 0.1$ but did not specify $\pi_0$ or $\pi_{New}$. With a back calculation, it is possible to deduce that the minimax design was chosen with $\pi_0 = 0.2$, $\pi_{New} = 0.4$, and a stage 1 sample size of 17. In this trial, 0 responses were observed in stage 1 and so stage 2 was not implemented. This implies that the response rate $\pi$ is estimated by 0/17 or 0% with 95% CI of approximately 0–19%. Thus, even with an optimistic view of the true response rate as possibly close to 19%, this is far below the expectations of the investigators who set $\pi_{New}$ as 40%.

## 5 PHASE II TRIALS WITH SURVIVAL ENDPOINTS

Although many phase II trials have disease response as a (binary) outcome, survival times or at least survival proportions at a fixed time are sometimes more relevant.

As already mentioned, a disadvantage of two-stage designs is that the final number of patients required is not known until after recruitment to stage 1 is complete and response in all these patients has been assessed. This poses a particular difficulty if the endpoint of concern is the time from initiation of treatment to some event (perhaps the death of the patient), which is expressed through the corresponding survival time. In this case there will be a variable, and possibly extended, period of observation necessary to make the requisite observations. However, estimates of survival at a prechosen fixed point in interval time (say 1-year poststart of treatment) can be estimated using the Kaplan–Meier (KM) technique, which takes into account censored observations. Censored survival time observations arise when a patient, although entered on the study and followed for a period of time, has not as yet experienced the "event" defined as the outcome for the trial. For survival itself "death" will be the event of concern, whereas if event-free survival was of concern, the event may be recurrence of the disease. Appropriate methods for survival time analysis are described by Machin et al. [13].

In this context, when considering the design of a phase II trial of a new drug, the investigators will usually have some knowledge of the activity of other drugs for the same disease. The anticipated survival rate of the new drug is therefore compared, at the planning stage, with that observed with other therapies. This may lead to the investigators prespecifying a survival probability that, if the new drug does not achieve, results in no further investigation. They might also have some idea of a survival probability that, if achieved or exceeded, would certainly imply that the new drug has activity worthy of further investigation, perhaps in a phase III randomized trial to determine efficacy.

### 5.1 Case and Morgan

In the Case and Morgan [14] two-stage phase II trial designs, "survival" times are utilized in place of binary response variables. The "survival" times usually correspond to the interval between the registration of the patient into the study or the commencement of the phase II treatment, and the time at which the event of primary concern occurs, for example, recurrence of the disease, death, or either of these.

When considering the Case–Morgan designs, it is important to distinguish between *chronological* time—that is, the date on which the trial recruits its first patient, the date of the planned interim analysis, the date the trial closes recruitment, or the date all patient follow-up ends—from the time *interval* between start of therapy and the occurrence of the event. Trial conduct is concerned with chronological time while trial analysis is concerned with interval time. We denote the former by $D$ and the latter by $t$.

The KM estimate at any follow-up time $t$, is denoted $S(t)$. Thus, for example, when $t = 1$ year, the KM estimate at that time point is denoted $S(1)$. In general, a convenient time point, which we denote by $T_{\text{Summary}}$, is chosen by the investigators and the

corresponding $S(T_{\text{Summary}})$ estimates the proportion of patients experiencing the event at that time point.

Typically, observing the event takes longer and is more variable in its time of occurrence than, for example, tumor response rate. This implies that any two-stage phase II design using such an endpoint may require a period between stage 1 and (the potential) stage 2. This time window is to allow sufficient events to accumulate for the stage 1 analysis so that a decision can be taken whether or not to continue to stage 2. The time window is added to the duration of stage 1, and its necessity may require suspending patient recruitment during this interval. Clearly, this will extend the total duration of the study. The Case–Morgan designs eliminate the need for this time window.

To implement the design, the investigators set for a particular interval time, $t = T_{\text{Summary}}$, the largest survival proportion as $S_0(T_{\text{Summary}})$, which, if true, would clearly imply that the treatment does not warrant further investigation. The investigators then judge what is the smallest survival proportion, $S_{\text{New}}(T_{\text{Summary}})$, that would imply the treatment warrants further investigation. This implies that the one-sided hypotheses to be tested in the study are: $H_0$: $S(T_{\text{Summary}}) \leq S_0(T_{\text{Summary}})$ versus $H_{\text{New}}$: $S(T_{\text{Summary}}) \geq S_{\text{New}}(T_{\text{Summary}})$, where $S(T_{\text{Summary}})$ is the actual probability of survival, which is to be estimated at the close of the trial. In addition to specifying $S_0(T_{\text{Summary}})$ and $S_{\text{New}}(T_{\text{Summary}})$, it is necessary to specify $\alpha$ and $\beta$.

With these inputs, there are then two variants of the Case–Morgan design, depending on whether we wish to minimize the expected duration of accrual (EDA) or the expected total study length (ETSL) for the trial. These are defined as follows:

$$\text{EDA} = D_{\text{Stage1}} + \left(1 - P_{\text{Early}}\right) D_{\text{Stage2}}$$
$$\text{ETSL} = D_{\text{Stage1}} + \left(1 - P_{\text{Early}}\right)\left(D_{\text{Stage2}} + T_{\text{Summary}}\right)$$

where $D_{\text{Stage1}}$ and $D_{\text{Stage2}}$ are the durations of stage 1 and stage 2 of the trial, respectively, and $P_{\text{Early}}$ is the probability of stopping at the end of stage 1.

***Example: Case and Morgan—Gemcitabine and External Beam Radiotherapy for Resectable Pancreatic Cancer*** Case and Morgan [14] consider the design of a phase II trial of the effectiveness of adjuvant gemcitabine and radiotherapy in the treatment of patients with resectable pancreatic cancer. The outcome measure used was 1-year survival, and they planned to test the null hypothesis that 1-year survival is 35% against an alternative of 50%. Thus, $T_{\text{Summary}} = 1$, $S_0(1) = 0.35$, $S_{\text{New}}(1) = 0.50$. Further $\beta = 0.1$ and $\alpha = 0.1$.

With these inputs, the ETSL design suggests that stage 1 recruits 54 patients and stage 2 83 patients. With the EDA design the corresponding sample sizes are 46 in stage 1 and 79 in stage 2, resulting in a total sample size of 125.

## 6 EFFICACY AND TOXICITY IN PHASE II TRIALS

In situations where the toxicity of an agent undergoing phase II testing is poorly understood, it may be desirable to incorporate toxicity considerations into the trial

design. We now discuss phase II trial designs for the situation in which both a minimum level of activity and a maximum level of (undesirable) toxicity are stipulated in the design. Such designs expand on the Simon two-stage designs discussed earlier.

### 6.1 Bryant and Day

Bryant and Day [15] point out that a common situation when considering phase I and phase II trials is that although the former primarily focuses on toxicity and the latter on efficacy, each in fact considers both. This provides the rationale for their phase II design that incorporates toxicity and activity considerations. Essentially, they combine a design for activity with a similar design for toxicity in which one is looking for both *acceptable* toxicity and *high* activity.

The design implies that two, *one-sided* hypotheses are to be tested. These are that the true response rate $\pi_R$ is either $\leq\pi_{R0}$, the maximum response rate of no interest, or $\geq\pi_{R\text{New}}$, the minimum response rate of interest. Further the probability of incorrectly rejecting the hypothesis $\pi_R \leq \pi_{R0}$ is set as $\alpha_R$. Similarly, $\alpha_T$ is set for the hypothesis $\pi_T \leq \pi_{T0}$ where $\pi_T$ is the maximum nontoxicity rate of no interest. In addition, the hypothesis $\pi_T \geq \pi_{T\text{New}}$ has to be set together with $\beta$, the probability of failing to recommend a treatment that is acceptable with respect to both activity and (non-)toxicity. (The terminology is a little clumsy here as it is more natural to talk in terms of "acceptable toxicity" rates rather than "acceptable nontoxicity" rates. Thus $1 - \pi_{T0}$ is the highest rate of toxicity above which the drug is unacceptable. In contrast, $1 - \pi_{T\text{New}}$ is the lower toxicity level below which the drug would be regarded as acceptable on this basis.)

In the Bryant and Day design, toxicity monitoring is incorporated into the Simon [11] design by requiring that the trial is terminated after stage 1 if there is an inadequate number of observed responses or an excessive number of observed toxicities. The treatment under investigation is recommended at the end of stage 2 only if there are both a sufficient number of responses and an acceptably small number of toxicities in total.

To implement the designs, the appropriate number of patients is recruited to stage 1, and once all their responses and toxicity experiences are observed, a decision whether or not to proceed to stage 2 is taken. If stage 2 is implemented, then once recruitment is complete and all assessments made, the response and toxicity rates, along with their corresponding CIs, are calculated. A decision with respect to efficacy and toxicity is then made. If stage 2 is not activated, the response rate and toxicity rates can still be calculated for the stage 1 patients despite either failure to demonstrate activity, too much toxicity, or both.

***Example: Bryant and Day Design—Ifosfamide and Vinorelbine in Ovarian Cancer*** González-Martín et al. [16] used the Bryant and Day two-stage design with a cutoff point for the response rate of 10% and for severe toxicity of 25%. Severe toxicity was defined as grade 3 and 4 nonhematological toxicity, neutropenic fever, or grade 4 thrombocytopenia. They do not provide full details of how the sample size was determined, but their choice of design specified a stage 1 of 14 patients and stage 2 a further 20 patients. In the event, in these advanced platinum-resistant ovarian cancer patients, the combination of ifosfamide and vinorelbine was

evidently very toxic. Hence the trial was closed after 12 patients with an observed toxicity level above the 25% contemplated. In fact, this corresponds to a design with $\alpha_R = \alpha_T = 0.1$, $\beta = 0.2$; $\pi_R^0 = 0.1$, $\pi_{R\text{New}} = 0.3$; $\pi_{T0} = 0.25$; and $\pi_{T\text{New}} = 0.45$. On this basis, the completed stage 1 trial of 14 patients proceeds to stage 2 if there are at least 2 responses *and* there are also no more than 2 patients with high toxicity. The stage 2 trial size is a further 20 patients, to a total of 34 for the whole trial, and sufficient efficacy with acceptable toxicity would be concluded if there were 6 or more responses observed *and* 10 or fewer with high toxicity.

## 7 BAYESIAN APPROACHES

### 7.1 Motivation

For most of the designs discussed thus far, the final response rate is estimated by $R/N$, where $R$ is the total number of responses observed from the total number of patients recruited $N$ (whether obtained from a single- or two-stage design). This response rate, together with the corresponding 95% CI, typically provide the basic information for the investigators to decide if a subsequent phase III trial is warranted. However, even after the trial is completed, there often remains considerable uncertainty about the true value of $\pi$. For example, in the trial reported by Lehnert et al. [10] using the Gehan design, a 17% response rate was observed from 23 patients with the corresponding 95% CI for $\pi$ from 7 to 37%. In the trial of previously treated patients with metastatic nasophyargngeal cancer, conducted by Foo et al. [3], a high response rate of 48% (95% CI of 33 to 63%) was reported. This result is consistent with both a true response rate as small as 33% and one as high as 63%, an almost twofold difference. The inevitable uncertainty arising from phase II trials with small sample sizes suggests that Bayesian approaches may be useful for phase II trials.

### 7.2 Overview of Bayesian Approaches in the Context of Phase II Trials

The foundation of the Bayesian approach is Bayes' theorem, which can be expressed as

$$\text{post}(\pi|x) \propto \text{lik}(x|\pi)\,\text{prior}(\pi)$$

which involves combining the likelihood $\text{lik}(x|\pi)$ with the prior distribution, $\text{prior}(\pi)$, to give the posterior distribution, $\text{post}(\pi|x)$. The $\text{prior}(\pi)$ summarizes what we know about $\pi$ before the trial commences, while $\text{lik}(x|\pi)$ describes the data to be collected from the trial itself. Finally $\text{post}(\pi|x)$ summarizes all we know about $\theta$ once the trial is completed.

Many phase II trials involve endpoints that are binary (e.g., whether a response occurs or not). For such endpoints, the prior distribution may be assumed to be of the form

$$\text{prior}(\pi) \propto \pi^{a-1}(1-\pi)^{b-1}$$

This is a Beta distribution with parameters $a$ and $b$, which can take any positive value. When $a$ and $b$ are integers, such a distribution corresponds to a prior belief equivalent to having observed $a$ responses out of a hypothetical $T = (a + b)$ patients. This is then similar to the situation modeled by the binomial likelihood distribution in which we have $x$ as the number of responses from $N$ patients.

Combining the above prior with a binomial likelihood results in a posterior distribution of the form

$$\text{post}(\pi|x) \propto \pi^{a+x-1}(1-\pi)^{b+N-x-1}$$

It can be seen that this too is a Beta distribution, but of the form Beta $(a + x, b + N - x)$.

As mentioned previously, the posterior distribution represents our overall belief at the close of the trial about the distribution of the population parameter $\pi$. Once we have obtained the posterior distribution, we can calculate the exact probabilities of $\pi$ being in any region of interest or obtain summary statistics such as its mean value.

The prior distribution summarizes the information on $\pi$ before the trial commences. The general way in which each of these are derived is as follows. The shape of a Beta distribution is dependent on the values of the parameters $a$ and $b$, and each of the priors will have particular values associated with them. However, eliciting values for $a$ and $b$ is typically not an easy process. Instead, it is often much easier to obtain values for the mean ($M$) and variance ($V$) of the corresponding prior distribution. Once obtained, these values can then be used to obtain $a$ and $b$ by solving the simultaneous equations:

$$M = \frac{a}{a+b} \qquad V = \frac{ab}{(a+b)^2(a+b+1)}$$

which give

$$a = \frac{M[M(1-M)-V]}{V} \qquad b = \frac{(1-M)[M(1-M)-V]}{V}$$

More generally, prior distributions could be elicited either from relevant external data (see e.g., Tan et al. [17]) or from subjective clinical opinion [18] or a combination of the two.

For more detailed overviews of Bayesian approaches, the reader is referred to Berry and Stangl [19], Spiegelhalter et al. [20], and Tan [21]. In the particular situation of two phase II trials conducted "in parallel" using a two-stage design, Bayesian approaches allow for the information from both trials to be taken into account when making decisions regarding whether to proceed to stage 2 of each trial or not [22].

### 7.3 Bayesian Single and Dual Threshold

In the Tan–Machin (TM) two-stage single-threshold design (STD) [23, 24], the focus is to estimate, for example, the posterior probability that $\pi > \pi_{\text{New}}$, so that if this is

high, at the end of the phase II trial, the investigators can be reasonably confident in recommending the compound for testing in a phase III trial.

The investigator first sets the minimum interest response rate $\pi_{New}$ and $\pi_{Prior}$ the anticipated response rate of the drug being tested. However, in place of $\alpha$ and $\beta$, $\lambda_1$ (the required threshold probability following stage 1 that $\pi > \pi_{New}$) and $\lambda_2$ ($>\lambda_1$) (the required threshold probability after completion of stage 2 that $\pi > \pi_{New}$) are specified. Further, once the first stage of the trial is completed, the estimated value of $\lambda_1$, that is, $u_1$, is computed and a decision made whether or not to proceed to stage 2. Should the trial continue to stage 2 then, on trial completion, $u_2$ is computed. Note that the trial only goes into stage 2 if the estimate of $\lambda_1$, at the end of stage 1, exceeds the design value. Efficacy is claimed at the end of stage 2 only if the estimate of $\lambda_2$, obtained from all the data, exceeds the design value.

The design determines the sample sizes for the trial based on the following principle. Suppose that the trial was to be conducted and that $X_1$ and $X_2$ represent the resulting data obtained from stage 1 and stage 2, respectively. Now, suppose also that the (hypothetical) response proportion underlying $X_1$ and $X_2$ is just larger than the prespecified $\pi_{New}$, say $\pi_{New} + \varepsilon$, for some small $\varepsilon > 0$. We then want the smallest overall sample size, $N_{TM}$, that will enable the posterior probability at the end of the trial, denoted $\Pr(\pi > \pi_{New} \mid X_1, X_2)$ or more briefly $\Pr(\pi > \pi_{New})$, to be at least $\lambda_2$. At the same time, we also want the smallest possible stage 1 sample size $n_{TM1}$, which is just large enough so that the posterior probability at the end of stage 1, $\Pr(\pi > \pi_{New} \mid X_1)$ or more briefly $\Pr(\pi > \pi_{New})$, is at least $\lambda_1$. Tan and Machin [23] suggest planning values for $(\lambda_1, \lambda_2)$ as (0.6, 0.7), (0.6, 0.8), or (0.7, 0.8) and also set a value of $\varepsilon = 0.05$.

Tan and Machin [23] also propose an alternative two-stage dual-threshold design (DTD). This design is identical to the STD except that the stage 1 sample size is determined not on the basis of the probability of exceeding $\pi_{New}$ but on the probability that $\pi$ will be less than the "no further interest" proportion, $\pi_0$. This represents the response rate below which the investigator would have no further interest in the new drug. Thus $\pi_0$ functions as a lower threshold on the response rate, as opposed to the upper threshold represented by $\pi_{New}$. The rationale behind this aspect of the DTD is that we want our stage 1 sample size to be large enough so that, if the trial data really does suggest a response rate that is below $\pi_0$, we want the posterior probability of $\pi$ being below $\pi_0$, to be at least $\lambda_1$. The design determines the smallest stage 1 sample size that satisfies this criterion. The trial only goes into stage 2 if the estimate of $\lambda_1$ exceeds the design value and efficacy is claimed at the end of stage 2 only if the estimate of $\lambda_2$ exceeds the design value.

The DTD requires the investigators to set $\pi_{Prior}$ as the anticipated value of $\pi$ for the drug being tested. A convenient choice may be $(\pi_0 + \pi_{New})/2$, but this is not a requirement. Further $\lambda_1$ is set as the required threshold probability following stage 1, that $\pi < \pi_0$, while $\lambda_2$ is the required threshold probability that, after completion of stage 2, $\pi > \pi_{New}$. (Note that unlike in the case of STD, it is no longer a requirement that $\lambda_1 < \lambda_2$.) Once stage 1 of the trial is completed, the estimated value of $\lambda_1$, that is $l_1$, is computed, and, should the trial continue to stage 2 then on its completion, $u_2$ is computed. The latter is then used to help make the decision as to whether or not a phase III trial is suggested. As with the STD, Tan and Machin [23] suggest planning values for $(\lambda_1, \lambda_2)$ as (0.6, 0.7), (0.6, 0.8), or (0.7, 0.8) and also set a value of $\varepsilon = 0.05$.

The original Tan–Machin [23] designs work on the basis of having a "vague" prior distribution. According to Mayo and Gajewski [25], this corresponded to having a prior sample size of 3. Furthermore, Tan and Machin imposed some practical constraints on the designs to encourage their adoption in practice. In particular, they constrained the total study size, $N_{TM}$, to be a minimum of 10 and a maximum of 90, with stage 1 size, $n_{TM1}$, having a minimum size of 5 and a maximum of $N_{TM}$ –5. For these and other reasons, Mayo and Gajewski [25] as well as Wang et al. [26] have suggested modifications to the original Tan–Machin design.

***Example: Tan–Machin STD Design—Gemcitabine in Metastatic Nasopharyngeal Cancer*** Tan and Machin [23] reanalyzed the phase II trial of Foo et al. [3] for previously treated patients as if they had been designed using STD. First, they back-calculated from the two-stage Simon minimax design utilized, that this choice implied for their STD values of $\lambda_1 = 0.728$ and $\lambda_2 = 0.774$, respectively. Using the actual trial data they then computed with the data at the close of stage 1, $\Pr(\pi > \pi_{New}) = u_1 = 0.997$ (which is clearly greater than $\lambda_1 = 0.728$). Further for the data at the close of stage 2, this probability was reestimated to be $\Pr(\pi > \pi_{New}) = u_2 = 0.999$ (which is clearly greater than $\lambda_2 = 0.774$). So *had* the STD been used, this reanalysis suggests that, at the end of stage 1, continuation to stage 2 would have been appropriate. Further information at the end of stage 2 very strongly recommended that gemcitabine can be considered to have sufficient activity for phase III evaluation.

***Example: Tan–Machin DTD Design—Gemcitabine in Metastatic Nasopharyngeal Cancer*** Similarly, reanalyzing the chemonaïve study of Foo et al. [3], but now on the basis that a DTD was conducted, all the actual trial data gives an estimate of $\Pr(\pi < \pi_0) = l_1 = 0.003$ or equivalently expressed by $\Pr(\pi > \pi_0) = 1 - l_1 = 0.997$. The estimate for $\Pr(\pi > \pi_{New}) = u_2 = 0.445$. Thus, we are very confident that $\pi > \pi_0$ but not so sure that $\pi > \pi_{New}$. These together suggest that the response rate lies within the region of uncertainty, $\pi_0 \leq \pi \leq \pi_{New}$, as this has a reasonably high probability of $0.997 - 0.445 = 0.552$ or 55%.

***Example: Tan–Machin DTD—Combination Therapy for Nasopharyngeal Cancer*** A phase II trial using a triplet combination of paclitaxel, carboplatin, and gemcitabine in metastatic nasopharyngeal carcinoma was conducted at the National Cancer Centre, Singapore, by Leong et al. [27]. The trial was expected to yield a minimum interest response rate of 80% and a no further interest response of 60%. The anticipated response rate was assumed to be equal to the minimum interest response rate and the overall threshold probability at the start and end of the trial is set to be 0.65 and 0.7, respectively. The sample size of the trial was calculated using the DTD. With the no further interest response rate $\pi_0 = 0.6$, the minimum interest response rate $\pi_{New} = 0.8$, the anticipated response rate $\pi_{Prior} = 0.8$, the minimum desired threshold probability at the start of the trial $\lambda_1 = 0.65$, and the minimum desired threshold probability at the end of the trial $\lambda_2 = 0.7$, we obtain the following design:

*Stage 1* Sample size of 19 patients; if responses less than 15, stop the trial as soon as this becomes apparent and declare lack of efficacy. Otherwise complete stage 1 and commence stage 2.

*Stage 2* Overall sample size of 32 patients for both stage, hence 13 stage 2 patients to be recruited; if total responses for the two stages combined is less than 28, stop trial as soon as this becomes apparent and declare lack of efficacy.

## 8 RANDOMIZED PHASE II TRIALS

Most phase II trials are of a single-arm, noncomparative design. However, in certain circumstances there may be several compounds available for potential phase III testing in the same type of patients, but practicalities imply that only one of these can go forward for this subsequent assessment. Since there are several options, good practice dictates that the eligible patients should be randomized to the alternatives. This can be achieved by using a randomized phase II selection design in which the objective is to select only one, the "best," of several agents tested simultaneously.

The randomized designs overcome the difficulties pointed out by Estey and Thall [28] when discussing single-arm trials where the actual differences between response rates associated with the treatments (*treatment effects*) are confounded, with differences between the trials (*trial effects*), as there is no randomization to treatment. Consequently, an apparent treatment effect may in reality only be a trial effect.

### 8.1 Simon, Wittes, and Ellenberg (SWE)

The Simon, Wittes, and Ellenberg [29] design is a randomized (single-stage) phase II design that selects from several candidate drugs that with the highest level of activity. This approach chooses the observed best treatment for the phase III trial, however small the advantage over the others. The trial size is determined in such a way that if a treatment exists for which the underlying efficacy is superior to the others by a specified amount, then it will be selected with a high probability. Although details of the random allocation process are not outlined below, this is a vital part of the design implementation. Details are provided by, for example, Machin and Campbell [30].

When the difference in true response rates of the best and next best treatment is $\delta$, Simon et al. [29] allow for the computation of sample sizes depending on the desired probability of correct selection, $P_{CS}$, and the number of treatments being tested, $g$. The response rate of the worst treatment is denoted $\pi_{Worst}$.

To implement the design, the appropriate number of patients are recruited and randomized to the $g$ groups. Once all their responses are observed, the response rates are calculated for each drug under test and that with the highest recommended for phase III testing.

***Example: Gemcitabine, Vinorelbine, or Docetaxel for Advanced Non-Small-Cell Lung Cancer*** A randomized phase II trial of single-agent gemcitabine, vinorelbine, or docetaxel in the treatment of elderly and/or poor performance status patients with advanced non-small-cell lung cancer was conducted at the National Cancer Centre of Singapore [31]. The design was implemented with the probability of correctly selecting the best treatment assumed to be 90%. It was anticipated that the single-agent activity of each drug has a baseline response rate of approximately

20%. In order to detect a 15% superiority of the best treatment over the others, we wished to determine how many patients should be recruited per treatment for the trial. For the difference in response rate $\delta = 0.15$, smallest response rate $\pi_{Worst} = 0.2$, probability of correct selection $P_{CS} = 0.90$, and treatment groups, $g = 3$, the design gives a sample size of $m = 44$ per treatment group. Thus the total number of patients to be recruited is given as $N = 3 \times 44 = 132$.

***Example: Non-Hodgkin's Lymphoma*** Itoh et al. [32] describe a randomized two-group phase II trial comparing dose-escalated (DE) with biweekly dose-intensified (DI) CHOP in newly diagnosed patients with advanced-stage aggressive non-Hodgkin's lymphoma. Their design anticipated at least a 65% complete response (CR) rate in both groups. To achieve a 90% probability of selecting the better arm when the CR rate is 15% higher in one arm than the other, at least 30 patients would be required in each arm. In that event, they recruited 35 patients to each arm and observed response rates with DE and DI of 51 and 60%, respectively. Their follow-on study, a randomized phase III trial, compares DI CHOP with the standard CHOP regimen.

## 9 TRIAL CONDUCT AND REPORTING

In phase III trials, much emphasis has been placed on developing standards for the good conduct and reporting of clinical trials. Among the aspects looked at are issues relating to informed consent, registration of subjects, monitoring the trial, and common standards for the reporting of trials. It is well known that many trials go unreported, leading to publication bias being an important concern. This motivates proposals that all phase III trials need to have their protocols formally registered before a trial can even begin [33].

Unfortunately, the conduct and reporting of phase II trials often does not meet the high standards demanded of the phase III randomized controlled trial. All phases of the clinical trials process crucially affect the final conclusions made regarding the usefulness of new treatments. As such, phase II trials also need to be conducted and reported to the highest standards, and we advocate that the standards applied to phase III trials should also be extended to phase II wherever appropriate. Furthermore, there are some particular considerations that need to be taken into account for phase II trials, which we now discuss.

### 9.1 Trial Conduct

Since most phase II trials are not comparative, in the taking of informed consent, only details of the procedures that are to be involved and any potential side effects and risks need to be explained. It would be important to explain to the patient that any therapeutic benefit hoped for, such as tumor shrinkage, may or may not transfer into benefit for the patient with respect to (say) increased survival or improved quality of life. In the case of randomized phase II trials of Simon et al. [29], then the usual considerations applicable to those for a phase III trial would apply.

Once consent has been taken and patients recruited into the trial, there is then a need to properly register and monitor them as for phase III trials. Moreover, unlike for phase III trials, continuous monitoring of patient responses may occur. This has particular implications should a two-stage design have been used. Typically, such designs would result in a delay in the recruitment process between stage 1 and stage 2, resulting in a longer trial. However with continuous monitoring of responses, stage 2 may be triggered before the formal recruitment to stage 1 is complete if there are already sufficient responses. Nevertheless, in such a situation, we recommend that a formal review of the stage 1 results should still be carried out.

## 9.2 What to Report

Considerable effort is required in order to conduct a clinical trial of whatever type and size, and this effort justifies reporting of the subsequent trial results with careful detail. However, there is a wide variation in the quality of the standard of reporting of clinical trials. In phase III trials, major strides in improving the quality have been made and pivotal to this has been the Consolidation of the Standards of Reporting Trials (CONSORT) statement described by Begg et al. [34] and amplified by Moher et al. [35]. CONSORT describes the essential items that should be reported in a trial publication in order to give assurance that the trial has been conducted to a high standard. This is an internationally agreed recommendation, adopted by many of the leading medical journals, although there are still some who do not appear to insist that their authors comply with the requirements.

Although the CONSORT statement primarily applies to phase III and not phase II trials, many of their principles can and should also be applied to these. In particular, Table 2 highlights some of the relevant items from CONSORT that should be applied. A diagram showing the flow of patients through the trial (Fig. 1) should also be given.

**TABLE 2 Selected Key Items Included in Phase II Clinical Trial Report**

| | |
|---|---|
| Participants | Eligibility criteria for participants and the setting and locations where the data were collected. |
| Intervention | Precise details of the intervention intended, and how and when they were actually administered. |
| Objectives | Specific objectives and hypotheses. |
| Outcomes | Clearly defined primary and secondary outcome measures |
| Sample size | How sample size was determined |
| Randomization and blinding (randomized phase II only) | Details of method used to generate the random allocation sequence—including details of strata and block size. |
| | Method used to implement the random allocation—numbered containers, central telephone, or Web-based. |
| | Description of the extent of the blinding in the trial—investigator, participant. |
| Statistical methods | Statistical methods used for the primary outcome(s). |
| Participant flow | Flow of participants through each stage of the trial (see Figure 1). |
| Recruitment | Dates defining the periods of recruitment and follow-up. |
| Follow-up | As many patients as possible to be followed up. Dropouts should be reported by treatment group. |

*Source*: Adapted from Moher and co-workers [35].

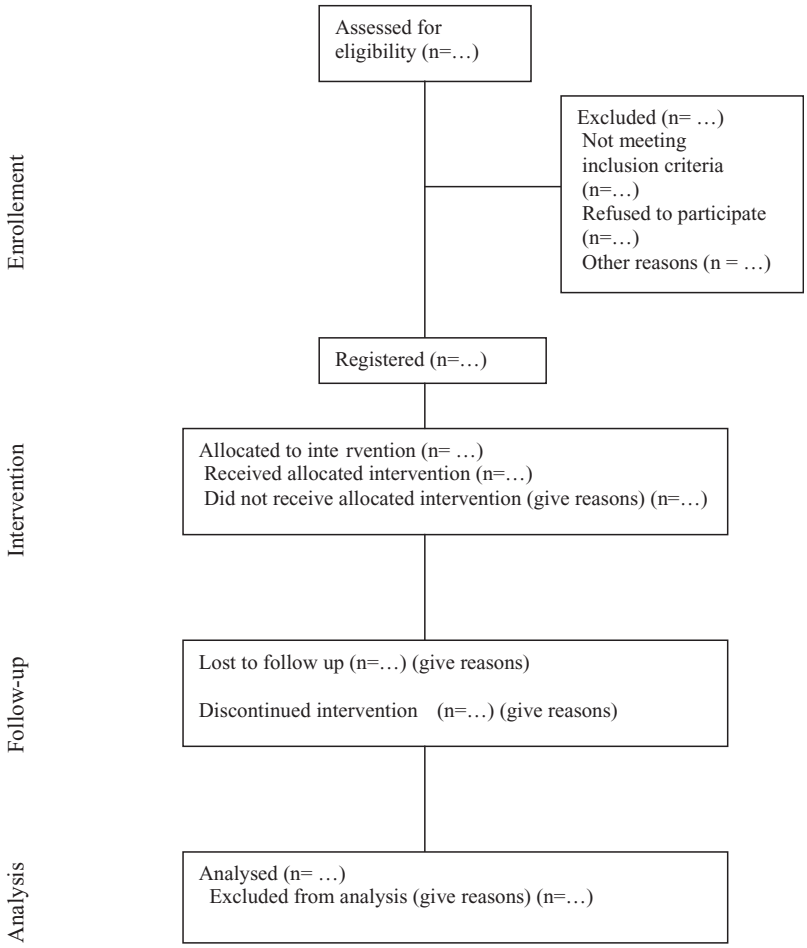


***FIGURE 1*** *Template of diagram showing flow of participants through a phase II trial.*

## 10 CONCLUDING REMARKS

Although phase II studies are typically of modest size relative to phase III trials, the temptation to conduct these studies without due attention to detail should be resisted. In fact, these studies (imprecise though they may be) provide key information for the drug development process. It is therefore essential that they are carefully designed, painstakingly conducted, and meticulously reported in full. Table 3 summarizes the key design and conduct issues.

It is also important to again emphasize that all patients should be registered for the trial (and hence are in the trial database) and that the final report includes information on all these patients. This is particularly important if a review process of, for example, each objective response in a phase II trial may reveal that certain patients admitted to the trial were either not truly eligible, had not received the full treatment as specified by the protocol, or could not be evaluated for the endpoint. Perhaps it is unclear whether or not they had sufficient tumor shrinkage for a satis-

**TABLE 3 Design and Conduct Issues for Phase II Trials**

| |
|---|
| Clearly define patient eligibility. |
| Clearly define the measures of response (and toxicity). |
| Choose a single- or two-stage design. |
| Consider the importance of not proceeding to stage 2 if activity low. |
| Consider whether a CI or threshold probability approach is to be used for interpretation. |
| Consider the possibility of a randomized selection design. |
| Ensure that all patients are registered. |
| Ensure all evaluations are made. |
| Ensure the final report details information on all patients. |

factory response. It must be clear in the study protocol itself, and in the subsequent report of the study results, whether these "ineligible," "noncompliant," and "none-valuable" patients are or are not included in the reported response rates. This equally applies for any assessment of toxicity, whether or not toxicity is a formal endpoint for the design as it is in the Bryant–Day design of phase II.

It should also be emphasized that phase II trials should never be seen as an alternative to well-designed (large) randomized phase III trials. This is because the small sample sizes in phase II trials give rise to estimates with very wide confidence intervals (i.e., a high level of uncertainty). Hence any conclusions drawn from such trials cannot be confirmatory.

Finally, it should be noted that the results of separate single-arm phase II trials should also generally not be used for comparative purposes because of the potential confounding of treatment effects with trial effects. Consequently, an apparent treatment effect may in reality only be a trial effect. As for randomized phase II trials, their objective is to select the "best" of several agents tested for further testing in a phase III trial and not as alternatives to phase III trials.

## REFERENCES


1. van Rijswijk, R. E., Vermorken, J. B., Reed, N., Favalli, G., Mendiola, C., Zanaboni, F., Mangili, G., Vergote, I., Guastalla, J. P., ten Bokkel Huinink, W. W., Lacave, A. J., Bonnefoi, H., Tumulo, S., Rietbroek, R., Teodorovic, I., Coens, C., and Pecorelli, S. (2003), Cisplatin, doxorubicin and ifosfamide in carcinosarcoma of the female genital tract. A phase II study of the European Organization for Research and Treatment of Cancer Gynaecological Cancer Group (EORTC 55923), *Eur. J. Cancer*, 39, 481–487.
2. Therasse, P., Arbuck, S. G., Eisenhauer, E. A., Wanders, J., Kaplan, R. S., Rubinstein, L., Verweij, J., van Glabbeke, M., Van Oosterom, T., Christian, M. C., and Gwyther, S. G. (2000), New guidelines to evaluate the response to treatment in solid tumors, *J. Nat. Cancer Inst.*, 92, 205–216.
3. Foo, K.-F., Tan, E.-H., Leong, S.-S., Wee, J. T. S., Tan, T., Fong, K.-W., Koh, L., Tai, B.-C., Lian, L.-G., and Machin, D. (2002), Gemcitabine in metastatic nasopharyngeal carcinoma of the undifferentiated type, *Ann. Oncol.*, 13, 150–156.
4. Machin, D., Campbell, M. J., Tan, S. B., and Tan, S. H. (2009), *Sample Size Tables for Clinical Studies*, 3rd ed. Wiley-Blackwell, Chichester.
5. Fleming, T. R. (1982), One-sample multiple testing procedure for Phase II clinical trial, *Biometrics*, 38, 143–151.

6. A'Hern, R. P. (2001), Sample size tables for exact single stage Phase II designs, *Statist. Med.*, 20, 859–866.
7. Iaffaioli, R. V., Formato, R., Tortoriello, A., Del Prete, S., Caraglia, M., Pappagallo, G., Pisano, A., Fanelli, F., Ianniello, G., Cigolari, S., Pizza, C., Marano, O., Pezzella, G., Pedicini, T., Febbraro, A., Incoronato, P., Manzione, L., Ferrari, E., Marzano, N., Quattrin, S., Pisconti, S., Nasti, G., Giotta, G., Colucci, G., and Southern Italy Oncology Group (2005), Phase II study of sequential hormonal therapy with anastrozole/exemestane in advanced and metastatic breast cancer, *Br. J. Cancer*, 92, 1621–1625.
8. Newcombe, R. G., and Altman, D. G. (2000), Proportions and their differences, in Altman, D. G., Machin, D., Bryant, T. N., and Gardner, M. J., Eds., *Statistics with Confidence*, 2nd ed., British Medical Journal Books, London, pp. 45–56.
9. Gehan, E. A. (1961), The determination of the number of patients required in a preliminary and follow-up trial of a new chemotherapeutic agent, *J. Chronic Dis.*, 13, 346–353.
10. Lehnert, M., Mross, K., Schueller, J., Thuerlimann, B., Kroeger, N., and Kupper, H. (1998), Phase II trial of dexverapamil and epirubicin in patients with non-responsive metastatic breast cancer, *Br. J. Cancer*, 77, 1155–1163.
11. Simon, R. (1989), Optimal two-stage designs for phase II clinical trials, *Controlled Clin. Trials*, 10, 1–10.
12. Chao, Y., Chan, W.-K., Birkhofer, M. J., Hu, O. Y.-P., Wang, S.-S., Huang, Y.-S., Liu, M., Whang-Peng, J., Chi, K.-H., Lui, W.-Y., and Lee, S.-D. (1998), Phase, II., and pharmacokinetic study of paclitaxel therapy for unresectable hepatocellular carcinoma patients, *Br. J. Cancer*, 78, 34–39.
13. Machin, D., Cheung, Y.-B., and Parmar, M. K. B. (2006), *Survival Analysis: A Practical Approach*, 2nd ed., Wiley, Chichester.
14. Case, L. D., and Morgan, T. M. (2003), Design of Phase II cancer trials evaluating survival probabilities, *BMC Med. Res. Method.*, 3, 6.
15. Bryant, J., and Day, R. (1995), Incorporating toxicity considerations into the design of two-stage Phase II clinical trials, *Biometrics*, 51, 1372–1383.
16. González-Martín, A., Crespo, C., García-López, J. L., Pedraza, M., Garrido, P., Lastra, E., and Moyano, A. (2002), Ifosfamide and vinorelbine in advanced platinum-resistant ovarian cancer: Excessive toxicity with a potentially active regimen, *Gynecol. Oncol.*, 84, 368–373.
17. Tan, S. B., Dear, K. B. G., Bruzzi, P., and Machin, D. (2003), Strategy for randomized clinical trials in rare cancers, *BMJ*, 327, 47–49.
18. Tan, S. B., Chung, Y. F. A., Tai, B. C., Cheung, Y. B., and Machin, D. (2003), Elicitation of prior distributions for a phase III randomized controlled trial of adjuvant therapy with surgery for hepatocellular carcinoma, *Controlled Clin. Trials*, 24, 110–121.
19. Berry, D. A., and Stangl, D. K. (1996), *Bayesian Biostatistics*, Marcel Dekker, New York.
20. Spiegelhalter, D. J., Myles, J. P., Jones, D., and Abrams, K. R. (1999), An introduction to bayesian methods in health technology assessment, *BMJ*, 319, 508–512.
21. Tan, S. B. (2001), Introduction to Bayesian methods for medical research, (invited article), *Ann. Acad. Med. Singapore* 30, 444–446.
22. Tan, S. B., Machin, D., Tai, B. C., Foo, K. F., and Tan, E. H. (2002), A Bayesian re-assessment of two phase II trials of Gemcitabine in metastatic nasopharyngeal cancer, *Br. J. Cancer*, 86, 843–850.
23. Tan, S. B., and Machin, D. (2002), Bayesian two-stage designs for phase II clinical trials, *Stat. Med.*, 21, 1991–2012.

24. Tan, S. B., Wong, E. H., and Machin, D. (2004), Bayesian two-stage design for phase II clinical trials, in Chow, S. C. Ed., *Encyclopedia of Biopharmaceutical Statistics*, 2nd ed. (online), http://www.dekker.com/servlet/product/DOI/101081EEBS120023507, Marcel Dekker, New York.
25. Mayo, M. S., and Gajewski, B. J. (2004), Bayesian sample size calculations in phase II clinical trials using informative conjugate priors, *Controlled Clin. Trials*, 25, 157–167.
26. Wang, Y. G., Leung, D. H. Y., Li, M., and Tan, S. B. (2005), Bayesian designs with frequentist and Bayesian error rate considerations, *Statist. Methods Med. Res.*, 14, 445–456.
27. Leong, S. S., Tay, M. H., Toh, C. K., Tan, S. B., Thng, C. H., Foo, K. F., Wee J. T.S., Lim, D., See, H. T., Tan, T., Fong, K. W., and Tan, E. H. (2005), Paclitaxel, carboplatin and gemcitabine in metastatic nasopharyngeal carcinoma: A phase II trial using a triplet combination, *Cancer*, 103, 569–575.
28. Estey, E. H., and Thall, P. (2003), New designs for phase 2 clinical trials, *Blood*, 102, 442–448.
29. Simon, R., Wittes, R. E., and Ellenberg, S. S. (1985), Randomized phase II clinical trials. *Cancer Treat. Rep.*, 69, 1375–1381.
30. Machin, D., and Campbell, M. J. (2005), *Design of Studies for Medical Research*, Wiley, Chichester.
31. Leong, S. S. (2005), A randomized phase II trial of single agent gemcitabine, vinorelbine or docetaxel in patients with advanced non-small cell lung cancer who have poor performance and/or are elderly, *J. Thoracic Oncol.*, 2, 230–236. *Protocol SQLU01*.
32. Itoh, K., Ohtsu, T., Fukuda, H., Sasaki, Y., Ogura, M., Morishima, Y., Chou, T., Aikawa, K., Uike, N., Mizorogi, F., Ohno, T., Ikeda, S., Sai, T., Taniwaki, M., Kawano, F., Niimi, M., Hotta, T., Shimoyama, M., and Tobinai, K. (2002), Randomized phase II study of biweekly CHOP and dose-escalated CHOP with prophylactic use of lenograstim (glycosylated G-CSF) in aggressive non-Hodgkin's lymphoma: Japan Clinical Oncology Group Study 9505, *Ann. Oncol.*, 13, 1347–1355.
33. Dickerson, K., and Rennie, D. (2003), Registering clinical trials, *JAMA*, 290, 516–523.
34. Begg, C., Cho, M., Eastwood, S., Horton, R., Moher, D., Olkin, I., Pitkin, R., Rennie, D., Schultz, K. F., Simel, D., and Stroup, D. F. (1996), Improving the quality of reporting randomized controlled trials: The CONSORT statement, *JAMA*, 276, 637–639.
35. Moher, D., Schultz, K. F., Altman, D. G., and the CONSORT Group (2001), The CONSORT statement: Revised recommendations for improving the quality of reports of parallel-group randomized trials, *Lancet*, 357, 1191–1194.

# *Designing and Conducting Phase III Studies*

**Nabil Saba, John Kauh, and Dong M. Shin**

*Emory University School of Medicine, Winship Cancer Institute, Department of Hematology and Oncology, Atlanta, Georgia*

**Contents**

1 Overview and Background
2 Drug Background and Information
3 Objectives
4 Primary and Secondary Endpoints
5 Selection Criteria for Patient Population: Inclusion and Exclusion Criteria
6 Trial Design
7 Starting Dose and Dose Modification
8 Evaluation of Drug Efficacy: World Health Organization (WHO) Criteria
9 Toxicity Evaluation
10 Pathology
11 Statistical Considerations
  11.1 Hypothesis
  11.2 Sample Size
  11.3 Power
  11.4 *P* Values
  11.5 Randomization
  11.6 Stratification
  11.7 Blinding
  11.8 Interim Analyses
12 Data-Monitoring Committee
13 Correlative Studies

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

14 Adverse Event Reporting
15 Appendices (Usual Items)
16 Informed Consent Forms and HIPAA
17 Closing
References

## 1 OVERVIEW AND BACKGROUND

Phase III studies are defined as randomized controlled either single- or multicenter studies enrolling large numbers of patients with the purpose of rendering a definitive assessment of how effective the studied intervention is, compared with the current standard of care. The randomized clinical trial is characterized by two or more therapeutic treatment groups. These groups are sometimes referred to as *arms*, particularly in cancer trials. One treatment may be a placebo control in which a biologically inert substance (in a drug trial) is used. Note that while most phase III clinical trials compare two medications or devices, some trials compare three or four medications, doses of medications, or devices against each other. The protocol is the written operating manual of a trial, which ensures that researchers in different locations perform the study uniformly, allowing the data to be pooled and analyzed.

The U.S. National Institutes of Health, (NIH) organize clinical trials into five different types [1]:

*Treatment Trials* Test experimental treatments, new combinations of drugs, or new approaches to surgery or radiation therapy.

*Prevention Trials* Look for better ways to prevent disease in people who have never had the disease or to prevent a disease from returning. These approaches may include medicines, vitamins, vaccines, minerals, or lifestyle changes.

*Diagnostic Trials* Conducted to find better tests or procedures for diagnosing a particular disease or condition.

*Screening Trials* Test the best way to detect certain diseases or health conditions.

*Quality of Life Trials (or Supportive Care Trials)* Explore ways to improve comfort and quality of life for individuals with a chronic illness.

A randomized phase III controlled trial is the study design that can provide the most compelling evidence that the study treatment causes the expected effect on human health. Currently, some phase II and most phase III drug trials are designed as randomized, double blind, and placebo controlled. After preliminary evidence from phase I and II studies suggesting effectiveness of the drug has been obtained, phase III trials are designed to gather additional information to assess the overall benefit to risk relationship of the drug for possible use as a new standard of care.

Many NIH programs encourage or require the use of protocol templates, such as the ones available here at http://ctep.cancer.gov/guidelines/templates.html.

While following template guidelines can help guide authors, care should be exercised when using a template since not every part of a template will necessarily apply in a given study.

As is the case for all research protocols, writing a clinical protocol is a task that requires the collaboration and input of many individuals with various expertise to help reach the goal of a scientifically plausible and sound study. The writing of the protocol is of utmost importance and is a necessary condition on which the quality of the study itself depends.

## 2 DRUG BACKGROUND AND INFORMATION

Phase III protocols, like phase I and II protocols, require a section detailing the scientific rationale for a protocol and the justification medically and scientifically of the hypothesis in question. This section should be organized in a logical and sequential manner.

First, the background section should focus on justifying the need for the study by elaborating first on where we are in the field. Here, a review of disease incidence and its impact on morbidity and mortality may help. Reviewing the results from similar prior studies or phase I and phase II studies testing the drug in question is also warranted. This will be a prerequisite to argue the need to perform the phase III trial and discuss what the trial will potentially add that would be important for patient care.

It is particularly important in phase III trials to indicate what the standard of care is in this patient population, so that if there is no standard of care an observation or placebo arm can be justified. In case there is a known standard, the control arm will usually be that standard, and the investigational drug will be used in the experimental arm(s).

If the drug in question is already established as an acceptable standard of care in the studied patient population, and the investigator wishes to perform a pharmacokinetic (PK) study with different dosing and administration schedules to be tested in several arms to determine the best dosing and schedule method, this design is that of a phase IV study.

It is important to learn about the drug and/or the intervention, and for the principal investigator (PI) to elaborate in the background section on the drug characteristics, the evidence derived from phase I and phase II trials and preclinical studies. Even though issues such as dosing, administration, and route of delivery will be expected to be detailed in the treatment plan section, it is helpful to introduce this information in the background section. Critical information will include the following: dose/dosing schedule, possible need for a vector, route of delivery, and method of preparation.

It is also advisable to discuss the possible benefits of performing the phase III trial in the background section, as this argument is closely linked to the findings of prior phase I or II studies. The benefits may not be restricted here to clinical outcome of disease but may also include ease of administration, convenience to patients, a better side effect profile, or a potential improvement in quality of life. Here, the authors may also allude to the patient population in question and elaborate briefly on the rationale behind the main inclusion and exclusion criteria as the opportunity is there to discuss these. This should be a prelude to the inclusion and exclusion criteria section in which a more detailed and complete description of these criteria is expected.

## 3 OBJECTIVES

In the initial development of a phase III protocol, it is essential for the investigator to try to answer several questions that will help in writing the protocol and define the objectives. Important questions to try to answer include: What is the expected outcome? What is the intervention? For how long will the intervention last? What patients or subjects is the intervention targeting? How many participants are needed? How can the potential benefit be optimized while minimizing potential harm?

The objectives should be stated clearly as they constitute the hypothesis in question. As in any trial, each objective needs to be discussed in the statistical section. The objective section should be concisely and simply written, often as a numbered or bulleted list. Here the authors should not include any arguments or justifications but rather be straightforward in describing the objectives. For a phase III trial the objectives usually are to compare two different approaches. For example, in a therapeutic trial the comparison usually involves two treatment plans, and the objective may seek to compare the survival of patients on each arm or the objective responses to the respective treatments. Other possible parameters may include the comparison of symptoms or quality of life parameters between arms or an assessment of the safety and tolerability of the treatment.

## 4 PRIMARY AND SECONDARY ENDPOINTS

It is important to remember that in a phase III study, as is the case in other designs, the primary endpoint will dictate the statistical method of analysis, the sample size, and stopping rules. Therefore, discussing these questions with a statistician while designing the study objectives cannot be overemphasized. For example, certain considerations such as the number of patients needed to reach statistical significance and the accrual potential in certain diseases may make the objectives difficult to achieve.

In a therapeutic trial, the primary endpoints must usually measure clinical outcome and will have the major impact on the rest of the study design as they will also influence the inclusion and exclusion criteria and the follow-up plan during or after the intervention. Secondary endpoints are usually easier to achieve and often revolve around the primary endpoints. They may include other parameters such as quality of life, assessing the ease of administration of agents in therapeutic trials, but in general they will not have as great an impact on the statistical section as the primary endpoints.

To better elucidate the choice of these endpoints, it is helpful for the author to include a section entitled Rationale for Selection of Endpoints. This will help the investigators review their decisions thoroughly and avoid major pitfalls that may result from a poor selection of endpoints. For example, choosing progression-free survival (PFS) as an endpoint for the trial may be more justifiable for certain types of malignancies such as sarcomas, which are generally incurable if being treated in the metastatic setting. Furthermore, prior interventions have not been shown to affect overall survival (OS) in this disease population—hence the rationale behind

choosing PFS as a primary clinical endpoint. The authors in this case may still be interested in looking at OS since this is a more questionable endpoint that is less likely to be reached in this disease, making it more suitable as a secondary endpoint. The authors may wish to add other endpoints as well that will be labeled exploratory endpoints and added to the secondary endpoints. These may include, for example, assessing changes in cancer-related symptoms.

## 5 SELECTION CRITERIA FOR PATIENT POPULATION: INCLUSION AND EXCLUSION CRITERIA

In any clinical trial, including phase III trials, selection criteria for patient enrollment must be well defined. The purpose of this process is to clearly define the subset of the general population to be investigated in the trial. When determining these criteria, it is important to consider the possible effects that the intervention may have on the subjects and any possible side effects already known or possibly anticipated by the investigator. Other important factors include the ability of the subjects to understand the nature of the intervention and be able to give a valid informed consent. In addition, for studies in which effectiveness is an endpoint, subjects will need to undergo exams that would determine the effectiveness (or lack thereof) of the intervention.

Exclusion criteria are designed to protect subjects with an expected high risk of side effects from the intervention or examinations required by the study, and to prevent the inclusion of biases in the study by including patients with serious comorbidities that could impact the outcome of a phase III trial.

The inclusion or exclusion criteria are, therefore, the medical or social criteria based on which a person may or may not be allowed to enter a clinical trial. These criteria are based on factors such as age, gender, type and stage of a disease, and other medical conditions. Careful attention should be paid while writing the inclusion and exclusion criteria since poorly written criteria have resulted in a number of ineligible and inevaluable patients being enrolled to a study, as well as the unnecessary exclusion of patients who could have been successfully enrolled.

Before proceeding with writing the eligibility criteria for a phase III trial, the investigator should be aware of certain facts: Criteria that are poorly written may dictate a poor rate of accrual to the study and may undermine its scientific validity and the ability to generalize its findings. Eligibility criteria present one of the most important obstacles to accrual to phase III clinical trials. It is imperative when writing these criteria to avoid confusion and to be simple and clear. Problems with being too selective in the inclusion and exclusion criteria include being unable to generalize the findings, limiting patient accrual, and possibly increasing the cost of the study. The investigator should aim to keep the number of criteria listed to a minimum, keeping only those necessary for the validity of the study and for the preservation of patient safety.

The study should enroll sufficient participants to be able to determine whether the endpoints of the study are met but should not enroll greater numbers than are needed to achieve statistical significance.

Factors such as the characteristics of the disease, the availability of alternative interventions, the availability of participants, the study endpoints, and the desired precision of the outcome may all influence the desired number of patients to enroll.

## 6 TRIAL DESIGN

The study design section of a phase III protocol should include sufficient information for the participating site to develop a comprehensive clinical algorithm for enrolled patients on all arms of the study. It usually describes in a stepwise fashion all procedures required by the study.

This section may include the following: a description of the initial evaluation of patients, the treatment plan for both arms of the study, the need to use certain procedures, information on the agent(s) to be used including the investigational agent(s) or other standard agents used in the treatment plan, dose scheduling, and dose modification of all agents.

In phase III clinical trials where it is imperative to compare a new modality with the considered standard of care (or with placebo when there is no established standard of care), some fundamental principles must be followed:

1. The groups must be alike in all important aspects and differ only in the intervention each group receives.
2. The concept of randomization implies that each participant has the same chance of receiving any of the interventions specified in the study.
3. The randomization (allocation process) is carried out using a chance mechanism so that neither the participant nor the investigator will know *in advance* which will be assigned.
4. An effort should be made to avoid conscious or subconscious influences. In designing a clinical trial, a sponsor must decide on the target number of patients who will participate. The sponsor's goal is usually to obtain a statistically significant result showing a significant difference in outcome.

The number of patients required to give a statistically significant result depends on the chosen endpoints of the trial. The larger the sample size or number of participants in the trial, the greater the statistical power. However, in designing a clinical trial, this consideration must be balanced with the fact that more patients will lead to a more expensive trial. Because of their large size and relatively long duration, phase III trials are the most expensive, the most time-consuming and difficult trials to design and run, especially in therapeutic trials addressing the treatment of chronic medical conditions. In determining the sample size, it would help to have information on the nature of the condition being treated, the desired precision of the outcome, some knowledge of the effectiveness of the intervention, the usual outcome of the studied patient population with current standards of care, and the availability of alternative treatments. It is important to make sure that the objectives and study design portion of the statistical section are identical to those described in the objectives section.

It is also imperative to pay attention to the definitions of toxicities in the statistical section. These should match those in the safety and adverse events section.

The advantages of a phase III design over other study designs is that randomization tends to result in comparable groups, which validates the statistical analysis of the data and renders the study more meaningful in addressing the effectiveness of the intervention in terms of patient outcome compared with phase I or phase II designs.

Limitations of phase III studies may possibly include the inability to generalize results to all patients with the same condition or disease because of preselection criteria imposed, as participants may not represent the general study population. Investigators may be faced with challenges in recruiting patients as randomization may be hard to accept by some.

Certain parameters may render a phase I or II study to be more appropriate for a certain investigational agent than a phase III trial. Factors such as lack of effectiveness of the standard of care, therapies that have not been well investigated, or an expected dramatic response may be better investigated in a phase II study.

In summary, when writing the study design section of a phase III protocol, attention to the following items is required:

1. Make sure that a statement of the primary and secondary endpoints to be measured during the trial is included.
2. Describe the design of the phase III trial, for example, double blind, placebo controlled, parallel design. A schematic diagram of trial design, procedures, and stages would also help.
3. Describe the measures taken to minimize/avoid bias including, for example, randomization or blinding.
4. Include a description of the intervention for all arms of the trial including the control arm. This needs to encompass dosage and regimen of the investigational as well as noninvestigational drugs. A description of the dosage form, packaging, and labeling of the investigational and noninvestigational product(s) must be included.
5. The expected duration of subject participation needs to be specified, and a description of the sequence and duration of all trial periods, including follow-up, needs to be well stated.
6. Make sure to include a description of the stopping rules in the statistical section or discontinuation criteria for individual participants.
7. Take measures to assure accountability for the investigational product and the placebo if this applies.

## 7 STARTING DOSE AND DOSE MODIFICATION

In a phase III trial, it is important to reiterate that the drug will be administered only to randomized patients. Once the study drug has been administered to an enrolled patient, this will continue until the discontinuation criteria have been met.

These criteria should have been specified in the design section and may include intolerable side effects, a documentation of progression of disease, significant deviation from the protocol (a protocol violation), noncompliance, or a patient's decision to withdraw (these items do not need to be respecified in this section).

Under this section, it is important to specify the route and schedule of drug administration on each of the phase III study arms. It is also important to specify the time of administration if deemed important by the investigator(s). For oral medications, it should be specified whether the drug should be taken with water and with or without food. For oral medications, a mechanism for patients to report missed doses should be in place and should be specified and described in this section (e.g., a diary card).

All anticipated drug-related events should be described in the protocol as well as the investigator's brochure (IB). The authors should be clear in describing the severity of the potentially reported events, as well as the clinical guidelines they will be using to decide on the appropriate management of any of these events. A table detailing the management of study drug modifications for all the anticipated and possibly nonanticipated drug reactions should be included. This should include side effects expected of the study medication but also should offer information on the drugs used in control arms of the study (i.e., the conventional standard of care agents).

Additional sections should detail the management of specific side effects. For example, how to clinically manage mouth sores in addition to dose modifications. Information should include the method to be used for diagnosis of certain side effects. For example, "pneumonitis is to be diagnosed by bronchoscopy if suspected."

Guidelines should specify when to resume the administration of the drug, what conditions need to be met for the resumption of the drug administration, and what dose to use for each of these conditions.

Dose modification guidelines and interruption guidelines should also be included for unanticipated events such as emergencies unrelated to the drug administration. Here, it is important to specify when is it acceptable to continue the drug and keep patients on the trial, if deemed appropriate (e.g., "within two weeks from the time of interruption").

A section on formulation of the drug(s), packaging, and labelling should also be included. Here all chemical ingredients of the drug in question should be specified. Information on storage and stability should also be given (temperature and expected shelf life of the product). This is not needed for drugs used in the control arm or drugs that are considered to be used as part of the standard of care.

Dispensing of the drug and dosing compliance is usually the pharmacy's responsibility and this should be specified. The site must use the appropriate dispensing and log accountability provided by the sponsor. These logs are to be maintained by the study pharmacist. The principal investigator or a delegate is responsible for discarding all unused study drugs, and this should be clearly stated in the protocol. During the trial and after termination, patients are responsible for returning all unused supplies and any missing items should be investigated. This also applies to study drugs on phase III trials and not conventional approved drugs for treating the disease in question.

## 8 EVALUATION OF DRUG EFFICACY: WORLD HEALTH ORGANIZATION (WHO) CRITERIA

For phase III clinical trials, the investigator must determine the frequency of evaluation of response. This depends on the disease being investigated and the new agent being introduced. For uniformity and to reduce biases, responses to an intervention may need to be reviewed by a single source. For example, a central review where imaging studies are submitted from different participating centers may be designated to assess tumor response to a novel anticancer drug.

For measurements of response to an anticancer drug, measurable lesions are defined as those that can be accurately measured in at least two dimensions. Attention should be made to certain factors. For example, tumor lesions that are situated in a previously irradiated area should not, in general, be considered measurable.

Other potentially nonmeasurable lesions may include small lesions [longest diameter <20 mm with conventional techniques or <10 mm using spiral computed tomography (CT) scan], bone lesions, leptomeningeal disease, ascites, pleural/pericardial effusions, lymphangitis, inflammatory breast disease, abdominal masses [not followed by CT or magnetic resonance imaging (MRI)], and cystic lesions.

In general, for assessing response, index lesions should be selected. This is done on the basis of their measurability in two dimensions and their suitability for accurate repeated measurements (by imaging techniques such as CT or MRI). A sum of the longest diameter (LD) and greatest perpendicular diameter of all index lesions will be calculated and reported as the baseline sum. The baseline sum will be used as a reference by which to characterize the objective tumor response. All other lesions (or sites of disease) including any measurable lesions over and above the index lesions should be identified as nonindex lesions and should also be recorded at baseline. Measurement of these lesions is not required, but the presence or absence of each should be noted throughout follow-up.

Evaluation of index lesions for phase I, II, or III clinical trials involve definitions of response, as set by the WHO or according to the RECIST criteria [2]. The WHO criteria follow.

### *For Index lesions*

*Complete Response (CR)* Disappearance of all known disease, determined by two observations not less than 4 weeks apart.

*Partial Response (PR)* A > 50% decrease in the total tumor load of the lesions that have been measured to determine the effect of therapy not less than 4 weeks apart. The observations must be consecutive.

For single lesion, >50% decrease in tumor area (multiplication of longest diameter by the greatest perpendicular diameter).

For multiple lesions, a 50% decrease in the sum of the products of the perpendicular diameters of the multiple lesions. In addition there can be no appearance of new lesions or progression of any lesion.

*Stable Disease (SD)* A 50% decrease in total tumor area cannot be established nor has a 25% increase in the size of one or more measurable lesions been demonstrated.

*Progressive Disease (PD)* A >25% increase in the area of one or more measurable lesions or the appearance of new lesions.

***For Nonindex Lesions***

*Complete Response (CR)* Complete disappearance of all known disease for at least 4 weeks.

*Partial Response (PR)* Estimated decrease in tumor area of >50% for at least 4 weeks.

*Stable Disease (SD)* No significant change for at least 4 weeks. This includes stable disease, estimated decrease of <50%, and lesions with estimated increase of <25%.

*Progressive Disease (PD)* Appearance of any new lesions not previously identified or an estimated increase of >25% in existent lesions.

***Duration of Overall Response*** The duration of overall response is measured from the time measurement criteria are met for CR or PR (whichever is first recorded) until the first date that recurrent or progressive disease is objectively documented (taking as reference for progressive disease the smallest measurements recorded since the treatment started, best response scan).

The duration of overall CR is measured from the time measurement criteria are first met for CR until the first date that recurrent disease is objectively documented.

Duration of stable disease: Stable disease is measured from the start of the treatment until the criteria for progression are met, taking as reference the smallest measurements recorded since the treatment started.

***Placebo Effect*** In a phase III clinical trial, a placebo may be used in one of the arms. A placebo is an inactive substance given to one group of participants, while the drug being tested is given to another group. A physical or emotional change, occurring after a substance is taken or administered, that is not the result of any special property of the substance is called a placebo effect, which may be beneficial, reflecting the expectations of the participant and, often, the expectations of the person giving the substance. The results obtained in the two groups are then compared to see if the investigational treatment is more effective in treating the condition than the placebo substance [1].

## 9 TOXICITY EVALUATION

For phase I, II, or III clinical trials an adverse event (AE) is defined as any untoward medical occurrence in a patient or clinical investigation subject administered a pharmaceutical product that does not necessarily have a causal relationship with the treatment. An AE can therefore be any unfavorable and unintended sign (including an abnormal laboratory finding), symptom, or disease temporally associated with the use of a medicinal (investigational) product, whether or not considered related to the medicinal (investigational) product. Pre-existing conditions that worsen during a study are to be reported as AEs.

The descriptions found in the revised National Cancer Institute (NCI) Common Terminology Criteria for Adverse Events (CTCAE) version 3.0 will be utilized for AE reporting. All appropriate treatment areas should have access to a copy of the CTCAE version 3.0. A copy of the CTCAE version 3.0 can be downloaded from the Cancer Therapy Evaluation Program (CTEP) web site (http://ctep.cancer.gov).

In addition to the definition of AEs, the safety (or adverse events) section should encompass a list of adverse events expected with the use of each study drug.

If the phase III trial involves blinding, a description of the unblinding process should be included. In addition, the human subject protection section should be included as is the case in all clinical trials and should encompass a discussion of minority representation, the inclusion or exclusion of subjects who are vulnerable because of mental capacity, language barriers or other factors, and a description of the methods for patient recruitment.

An informed consent for a phase III trial should include the following according to the Office of Human Subjects Research (OHSR) [3]:

1. Comprehension of the information provided.
2. Disclosure of relevant information to prospective research subjects.
3. Voluntary agreement of the subject, free from coercion.
4. Statement that the study involves research.
5. Purpose of the research and the length of the study.
6. Description of risks and benefits.
7. Discussion of alternative therapies.
8. Confidentiality policy.
9. Compensation for injury.
10. Contact for further questions/information.
11. Statement of voluntary participation.
12. The use of simple language and the avoidance of technical words, while being thorough and complete. The avoidance of phrases that may be coercive. The insight of a patient's advocate may be very helpful in writing the consent.
13. The principal investigator should personally review the consent form thoroughly.

## 10 PATHOLOGY

As in other clinical trial designs, in a phase III trial, it is imperative especially with cancer-related therapies to be specific as to which pathologic subtype of cancer is to be included in the study. The inclusion and exclusion criteria should specify what pathologic subtypes are admissible to the studies and which are to be excluded. The lack of this information could result in jeopardizing the study by including patients who have heterogeneous diseases with different clinical behaviors.

If the study is multiinstitutional, the investigators may choose to have the pathology slides reviewed by a central pathology team for independent confirmation of the tissue diagnosis. This, as is the case with radiology central reviews, will minimize the introduction of subjective biases. Any discordance between the central pathol-

ogy assessment and the investigator site should be a subject for discussion between the sponsors and the principal investigator.

If the study calls for correlative examination of blood samples or archival tissues, the protocol should specify the mechanism by which the tissue is collected from the different institutions and consenting patients. It also helps to guide the participating institutions about the preferred tissue processing methods, for example, whether it is preferable that archival specimens be formalin fixed, or, if a block is not available, what is required, for example, a specified number of unstained slides to help determine a certain tissue marker. This entails a close coordination between the investigators, the central pathology labs, and the different pathology departments of participating centers.

It is strongly advisable for large phase III or multiinstitutional trials that a laboratory manual detailing the procedure of collection of material including archival material or whole blood specimens from participating patients, and shipping of the material be included as a separate document in the protocol.

## 11 STATISTICAL CONSIDERATIONS

Given the complex nature of large clinical trials, a well-versed biostatistician is an integral part of the study design team and should be consulted in the initial planning stages. Although the study statistician will complete the statistical section of the trial protocol, investigators must have a basic understanding of the concepts involved.

### 11.1 Hypothesis

As with any well-designed scientific experiment, phase III clinical trial investigators must begin with a hypothesis, defined as a supposition that appears to explain a group of phenomena and is advanced as a basis for further investigation [4]. The investigators must then design a trial to test the hypothesis.

In order to test a hypothesis, one must first state the null hypothesis ($H_0$), which is defined as a statement that there is no difference between a standard and experimental treatment. The alternative hypothesis ($H_1$) is a statement that there is a difference between a standard and experimental treatment [5]. If data from a clinical trial reject the null hypothesis, then the alternative hypothesis is proven true. There are three general rules when stating $H_0$ and $H_1$:

1. The null hypothesis is the hypothesis of "no difference."
2. Whatever is to be detected or supported should be the alternative hypothesis.
3. Statistical hypothesis are always set up in the hopes of being able to reject $H_0$ and thereby accept $H_1$.

An example would be that a drug company has developed a new chemotherapy for pancreatic cancer, which it feels has a greater than 50% chance of shrinking tumors. If RR denotes response rates then:

$H_0$: RR $\leq 50\%$
$H_1$: RR $> 50\%$

Once a clinical trial is completed, the investigators must analyze the available data to make a determination on whether the data support $H_0$ or $H_1$. Errors in reaching a conclusion from clinical trials can and do occur. When they occur, two types of errors that can be made are:

1. Type I error—rejecting the null hypothesis when it is actually true; $\alpha$ is the probability of making a type I error.
2. Type II error—not rejecting the null hypothesis when it is false; $\beta$ is the probability of making a type II error.

## 11.2 Sample size

The number of patients to be enrolled to reach a statistically meaningful answer at the conclusion of a clinical trial is based upon the assumptions made during the planning stages of the trial.

## 11.3 Power

Power is defined as the probability of rejecting the null hypothesis when it is indeed false or, conversely, accepting the alternative hypothesis is true when it really is true. In other words, the probability of obtaining a statistically significant result is known as the "power" of a trial. A general rule regarding power calculations of clinical trials is that as the number of planned subjects enrolled in a clinical trial increases, the power of the study increases as well. Conversely, the smaller the planned number of subjects accrued to a clinical trial, the bigger the difference between the standard and the experimental arms must be to demonstrate a statistically significant difference. Generally, the sample size of most phase III trials is adjusted to a power of either 0.80 or 0.90, assuming a difference between the standard and experimental arms is the smallest considered clinically meaningful.

## 11.4 P Values

The $P$ value is the probability of obtaining by chance a result at least as extreme as that observed, even when the null hypothesis is true and no real difference exists [4].

Alternatively, the $P$ value can be thought of as the probability that the observed result is due to chance alone. Generally, $P$ values $<0.05$ are considered significant. The $P$ value is calculated after the statistical test has been performed; if the $P$ value is less than $\alpha$, the null hypothesis is rejected.

## 11.5 Randomization

In clinical trials there are many sources of biases, ranging from physician judgments to patient self-selection, to referral patterns, all of which may skew data. Randomly

selecting trial subjects for either control versus experimental treatment is an effective way to minimize biases [6–10]. It is important to note that randomization does not completely remove all biases, but rather minimizes their effect on overall study results. Some investigators have argued that randomization is not absolutely necessary in all cases as matched historical or current controls can be identified. However, matching can only be performed with respect to known factors, and most investigators would agree there are many more unknown variables that cannot be accounted for in the study design. Thus without a randomization, a "statistically significant result" may be the result of a nonrandom difference in the distribution of unknown prognostic factors. Generally, for phase III clinical trials, which are large and costly to perform, randomization is a necessity to minimize the possibility of obtaining a nonmeaningful result.

There are many ways to randomize subjects, each with their respective advantages and disadvantages.

## 11.6 Stratification

In the case of known prognostic factors, stratifying subjects at the time of randomization may improve overall results as the study is more likely to be comparing two similar population vurses two populations that are skewed with an overabundance of a particular group. This can be accomplished by creating separate randomization lists for each stratum. For example, a clinical trial investigating a new agent in unresectable or metastatic pancreatic cancer with overall survival as the primary endpoint should prospectively stratify patients based upon locally advanced disease or metastatic disease as these two patient populations have known different survival. Otherwise, the study results may be uninterpretable due to a mixture of two patient populations with different expected survival estimates.

## 11.7 Blinding

Trials set so that participants do not know which intervention they are receiving are known as single-blinded trials. Those in which neither researchers nor participants know who is in the investigational or control group are called double-blinded trials. Double-blinded trials ensure that people assessing the outcome will not be influenced by knowing which intervention a participant is receiving and also that ancillary follow-up treatment will be the same.

## 11.8 Interim Analyses

Interim analyses are commonly performed in phase III clinical trials to evaluate current data and determine if a difference between the standard and experimental arms are sufficient enough to warrant premature disclosure of study results. Much controversy surrounds utilizing interim analysis data for stopping trials prematurely for efficacy. Supporters argue if a significant benefit of an experimental treatment is seen, why withhold treatment that may benefit patients; while detractors would argue that given more time the differences in the two treatments may vanish and stopping a trial would prevent fully evaluating the alternative hypothesis. In the setting of an interim analysis demonstrating significantly inferior results from an

experimental treatment, supporters would want to stop that trial from exposing trial subjects to an inferior experimental arm, while detractors would argue in favor of continuing the trial as the differences between the standard and experimental arms could potentially lessen or even reverse if the trial is permitted to run its full course.

In addition to ethical concerns, interim analyses also have potential biostatistical drawbacks. Fleming and colleagues reported that if interim statistical significance tests are performed every 3 months on a study planned for 3 years, the probability of reaching a statistically significant result prematurely can be as high as 26% [11]. This error is known as the type I error of the analysis plan. Therefore, in order for an interim analysis to be valid and useful, the analysis should be preplanned and not occur too frequently.

To minimize the potential negative impact of premature results influencing patients and trial investigators, phase III trials should have a data-monitoring committee (DMC) consisting of experienced research personnel [12, 13]. Due to potential conflict of interest issues with investigators and continuation of the trial, interim results are usually not shared with investigators unless the trial DMC authorizes release of the data. There are several different statistical designs for planning interim analysis [10, 14–16]. Haybittle and colleagues' method involves discounting interim differences unless the difference is statistically significant at the two-sided $P < 0.0025$ level. If the interim differences are not significant at that level, the trial continues until its originally intended size. Other methods have a predetermined number of planned interim analyses and the $P$ value that is used to determine statistical significance depends on the number of analyses performed during the trial.

## 12 DATA-MONITORING COMMITTEE

Data-monitoring committees (also known as data safety monitoring boards) consist of a group of individuals with various areas of expertise charged with overseeing the conduct of a clinical trial in terms of safety and efficacy. Not all phase III trials are required to have a DMC, as in the case of low-risk trials. However, the majority of therapeutic trials will have a DMC. Depending on the sponsor and the trial, DMC membership and responsibilities will differ.

The U.S. Food and Drug Administration (FDA) defines a clinical trial DMC as:

> A group of individuals with pertinent expertise that reviews on a regular basis accumulating data from one or more ongoing clinical trials. The DMC advises the sponsor regarding the continuing safety of trial subjects and those yet to be recruited to the trial, as well as the continuing validity and scientific merit of the trial. When a single DMC is responsible for monitoring multiple trials, the considerations for establishment and operation of the DMC are generally similar to those for a DMC monitoring a single trial, but the logistics may be more complex. [17]

Date-monitoring committees came into use in the 1960s. They were first utilized in large randomized multicenter trials sponsored by the federal government. The first DMC was formed as a result of an NIH external advisory group. This DMC was charged with monitoring the safety, efficacy, and trial conduct of several National

Heart Institute trials [18]. Current DMCs operate based upon the decades of experience gained since the first DMC [19]. Recently, the use of DMCs has grown in popularity. Due to issues related to conflict of interest and the use of mortality endpoints, integration of DMCs into industry-sponsored trials has been increasing.

Some governmental agencies that sponsor clinical trials may require DMCs for certain trials; however, the U.S. FDA does not currently require the use of DMCs except for research studies in emergency settings in which informed consent requirement is exempted. There are no absolute criteria for requiring a DMC, but generally accepted situations are a controlled trial of any size that will compare rates of mortality or major morbidity.

The composition of a DMC is critical; a poorly constituted DMC may put study subjects' lives at risk and also risk the scientific validity of a trial. Criteria commonly used in selecting members are:

1. Relevant experience in area of study
2. Experience on other DMCs
3. Absence of conflict of interest

Data-monitoring committees usually consist of clinicians with expertise in the area of interest and at least one biostatistician with experience with the statistical methods utilized in the trial being monitored. Depending on the trial, other potential members may include medical ethicists, toxicologists, epidemiologists, pharmacologists, or other individuals with potential interest or experience in the area of study. Ideally, members of a DMC should have no relationship to the study sponsors or investigator. In reality this is nearly impossible since members of a DMC are usually selected or nominated by the study sponsors or investigators.

In addition, DMCs should have an established charter describing the duties of the DMC prior to initiation of a study. Important components of the charter should include:

1. Meeting schedule and format (telephone versus face to face).
2. Meeting structure—ideally DMC meetings should be closed to all study sponsors and investigators, but the FDA does allow interaction between the DMC and study investigator/sponsors to clarify issues that may arise during a review.
3. Outline format of interim data to be presented to the DMC.
4. Study plan for interim analysis and stopping rules for futility for superior efficacy.

Although there are potential benefits to be derived from participation in clinical research, the investigational review boards (IRBs) and the NIH must ensure, to the greatest extent possible, the safety of study participants, that they do not incur undue risk and that the risks versus benefits are continually reassessed throughout the study period. Every clinical trial should have provision for data and safety monitoring. A variety of types of monitoring may be anticipated depending on the nature, size, and complexity of the clinical trial. In many cases, the principal investigator would be expected to perform the monitoring function [20].

## 13 CORRELATIVE STUDIES

Secondary endpoints can lead to meaningful trial results, in particular, laboratory or pathology correlatives are useful. Specimens can be obtained during the period of study enrollment and then banked for future use. Certainly, preplanned laboratory or pathological correlative studies can be fruitful, but saved specimens can prove to be invaluable for future study should new information become available or new scientific techniques are developed. The general process for evaluating the usefulness of correlative studies includes:

1. Identifying a clinical hypothesis
2. Identifying the laboratory parameters that may be influenced by the clinical treatment
3. Identifying the correlative hypothesis

An example of the utility of correlative studies is the International Adjuvant Lung Cancer Trial (IALT). This trial was designed to evaluate the effect of cisplatin-based adjuvant chemotherapy on survival after complete resection of non-small-cell lung cancer (NSCLC) [21]. Two years after the publication of the IALT results, Olaussen and colleagues hypothesized that expression of excision repair cross-complementation group 1 (ERCC1) protein could predict survival from cisplatin-based adjuvant chemotherapy in NSCLC as there is a significant amount of in vitro data supporting the rationale of platinum resistance to the expression of the ERCC1 messenger RNA (mRNA) in cell lines. In order to carry out their analysis, the investigators obtained 761 tumor specimens from IALT subjects and evaluated the expression of the ERCC1 protein. In the end, the investigators concluded, "patients with completely resected non-small-cell lung cancer and ERCC1-negative tumors appear to benefit from adjuvant cisplatin-based chemotherapy, whereas patients with ERCC1-positive tumors do not" [22].

An example of a prospectively designed study to bank tumor specimens is the Eastern Cooperative Oncology Group study E3200. The purpose of E3200 was to evaluate response, time to progression, and overall survival of patients with advanced colorectal cancer who have failed therapy with irinotecan and 5-fluorouracil when treated with oxaliplatin and fluorouracil with or without bevacizumab or bevacizumab alone. There were no preplanned correlative studies to be performed on the tumor specimens, but the investigators utilized the study as an opportunity to collect biologic samples prospectively for future use. One can image that these specimens could be analyzed at a later date to search for a biomarker of response as a wealth of clinical data for each patient was also prospectively collected during the clinical trial.

## 14 ADVERSE EVENT REPORTING

The timely reporting of adverse events that occur during the conduct of a trial is highly critical in order to maximize subject safety. Potential adverse events due to an experimental treatment under investigation must be quickly disseminated to all

study investigators so as to alert all study-related personnel to possible events and minimize future adverse events.

The International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH) was convened to bring together experts from the pharmaceutical industry and the regulatory authorities of Europe, Japan, and the United States to discuss scientific and technical aspects of product registration. The overall purpose of this conference was to make recommendations to allow uniform interpretation and application of technical guidelines for product development across the world in order to minimize duplicate testing of new products. ICH defined the following terms to clarify terminology used for drug event reporting:

Adverse event (AE) is defined as "any untoward medical occurrence in a patient or clinical investigation subject administered a pharmaceutical product and which does not necessarily have to have a causal relationship with this treatment" [23]. An AE can therefore be any unfavorable and unintended sign (including an abnormal laboratory result), symptom, or disease temporally associated with the use of a medicinal product, whether or not considered related to the medicinal product.

Adverse drug reaction (ADR) is defined as "a response to a drug which is noxious and unintended and which occurs at doses normally used in man for prophylaxis, diagnosis, or therapy of disease or for modification of physiological function" [23].

Serious adverse event (SAE) is defined as "a serious adverse event (experience) or reaction is any untoward medical occurrence that at any dose:

- Results in death
- Is life-threatening" [23]

Several mechanisms have been put in place by the FDA as well as other governing bodies such as the NCI. In the case of industry-sponsored studies, the sponsor is expected to collect all adverse event data and disseminate to all study sites to alert of potential adverse events.

MedWatch, is a FDA safety information and adverse event reporting program that serves both health care professionals and the medical product-using public. The FDA receives adverse drug reaction reports from manufacturers as required by regulation. Health care professionals and consumers can voluntarily send reports through the MedWatch program as well. These reports become part of a database. The ICH has issued guidance to facilitate the standardization of the data elements for the transmission of individual case safety reports regarding safety reporting (ICH E2B), which describes the content and format for the electronic submission of reports from manufacturers.

The FDA codes all reported adverse events using a standardized international terminology, MedDRA (the Medical Dictionary for Regulatory Activities), then uses this data to monitor for potential unforeseen toxicities. The data collected through the MEDWatch program is then incorporated into the Adverse Event Reporting System (AERS), which is a computerized information database designed to support the FDA's postmarketing safety surveillance program for all approved drug and therapeutic biologic products. The ultimate goal of AERS is to improve

the public health by providing the best available tools for storing and analyzing safety reports.

The CTEP branch of the National Cancer Institute utilizes a Web-based system for the electronic submission of expedited reports on protocols utilizing a CTEP-sponsored investigational new drug (IND) known as the Adverse Event Expedited Reporting System (AdEERS) to rapidly collect and disseminate clinical-trial-related adverse events [24].

## 15 APPENDICES (USUAL ITEMS)

An appendix contains information that is nonessential to the understanding of the clinical trial protocol but may further clarify points without burdening the body of the protocol. Each appendix should be identified by a Roman numeral in sequence, for example, Appendix I, Appendix II, and so on. Each appendix should contain different material.

Appendices should include but are not limited to:

1. Informed consent form
2. Guidelines for study-related procedures not written in the body of the protocol such as:
   a. Pathology submission guidelines and instructions
   b. Correlative laboratory processing guidelines
3. Any additional instructions that are to be given to the study subjects

## 16 INFORMED CONSENT FORMS AND HIPAA

For trials subject to the requirement of the U.S. FDA, which would be virtually all phase III clinical trials, the informed consent form (ICF) must meet the requirements of Title 21 of the Code of Federal Regulations, chapter 50 [25]. The Code of Federal Regulations (CFR) is the codification of the general and permanent rules published in the *Federal Register* by the executive departments and agencies of the federal government. It can be access at http://www.gpoaccess.gov/cfr/about.html. Chapter 50 covers the rules governing the protection of human subjects. Each ICF must also contain the six elements listed in 21 CFR 50.25(b) that are determined to be appropriate for a study. However, the IRB governing the study has the final authority for ensuring the adequacy of the ICF. In addition, the IRB may also impose additional requirements not listed in 21 CFR.

To improve the approval process of a multiinstitutional trial sample or draft consent, documents should be developed that each study site may alter to satisfy local IRB requirements. During the course of a study, many changes and amendments are likely to be made to the protocol as well as the ICF. In order to minimize confusion of multiple versions of ICFs, version numbers and IRB approval dates should be labeled on each version of the ICF approved.

***21 CFR 50.20: General Requirements for Informed Consent*** Language in the ICF must be written in simple and easy to understand terms, as the vast majority of

study participants are not likely to possess medical knowledge sufficient to comprehend the risk and benefits involved in a modern-day clinical trial. 21 CFR 50.20 states the following:

> Except as provided in ß50.23, no investigator may involve a human being as a subject in research covered by these regulations unless the investigator has obtained the legally effective informed consent of the subject or the subject's legally authorized representative. An investigator shall seek such consent only under circumstances that provide the prospective subject or the representative sufficient opportunity to consider whether or not to participate and that minimize the possibility of coercion or undue influence. The information that is given to the subject or the representative shall be in language understandable to the subject or the representative. No informed consent, whether oral or written, may include any exculpatory language through which the subject or the representative is made to waive or appear to waive any of the subject's rights, or releases or appears to release the investigator, the sponsor, the institution, or its agents from liability for negligence.

***Non-English-Speaking Subjects*** In the case of subjects who do not possess a sufficient grasp of the English language to critically evaluate an ICF, a translated ICF document in the native language of the potential study subject should be presented. While a translator may be used intermittently, translators should not be used routinely due to potential variations between individual translators. In addition, local IRBs may have additional rules regarding foreign language ICFs; as such, investigators should consult with the local IRB before proceeding to translate an ICF since this is an expensive procedure.

In the event of an illiterate subject, provided the subject is able to (1) retain the ability to understand the concepts of the study and evaluate the risk and benefit of being in the study when it is explained verbally and (2) is able to indicate approval or disapproval to study entry, they may be entered into the study. To further validate the consenting process of an illiterate individual, it is also suggested to have an impartial third-party witness as well as videotaping the entire consent process.

***Assent of Children*** Guidelines for ICFs regarding children are not specifically addressed in the 21 CFR; instead the FDA defers judgment to the IRB overseeing the particular study. For children old enough to read and comprehend, many IRBs require signatures from both the parents and the child before study entry.

### *21 CFR 50.25: Elements of Informed Consent*

(1) A statement that the study involves research, an explanation of the purposes of the research and the expected duration of the subject's participation, a description of the procedures to be followed, and identification of any procedures which are experimental.

The ICF must explicitly state that the purpose of a clinical trial is research as the relationship between a patient–physician is different than that between a trial participant–investigator.

(2) A description of any reasonably foreseeable risks or discomforts to the subject.

The risks of the trial must be explicitly stated in the ICF. Many IRBs will also require quantification of risks, that is, the expected percentage of patients who are likely to experience a side effect such as nausea or diarrhea.

(3) A description of any benefits to the subject or to others which may reasonably be expected from the research.

The potential benefits of a trial must be stated as a possible benefit and not be overstated as a likely event. Common phrases used in ICFs include "this trial may or not benefit you directly, but may help other people with the same medical problem as you."

(4) A disclosure of appropriate alternative procedures or courses of treatment, if any, that might be advantageous to the subject.

The ICF must offer alternative choices to the patient if available.

(5) A statement describing the extent, if any, to which confidentiality of records identifying the subject will be maintained and that notes the possibility that the Food and Drug Administration may inspect the records.

The ICF must contain a statement explaining the confidentiality of the subject's medical records as they relate to the study and also state who may have access to the subject's medical records such as the FDA, IRB, or study sponsors.

(6) For research involving more than minimal risk, an explanation as to whether any compensation and an explanation as to whether any medical treatments are available if injury occurs and, if so, what they consist of, or where further information may be obtained.

The ICF must contain a statement regarding compensation if any will be provided. The ICF must also explain whether there is compensation available in case of injury but must not waive or appear to waive the rights of the subject. Acceptable language includes "no funds have been set aside for," "[the cost] will be billed to you or your insurance."

(7) An explanation of whom to contact for answers to pertinent questions about the research and research subjects' rights, and whom to contact in the event of a research-related injury to the subject.

The ICF must clearly state who should be contacted in the event of questions that arise regarding study participation. Generally, the contact person listed should be the principle investigator at a study site.

(8) A statement that participation is voluntary, that refusal to participate will involve no penalty or loss of benefits to which the subject is otherwise entitled, and that the subject may discontinue participation at any time without penalty or loss of benefits to which the subject is otherwise entitled.

The ICF must clearly state that participation in a study is voluntary. Common phrases used include "you may participate in this trial if you wish. Your decision to participate in this trial will not affect your care. The options to the trial are _XXXX_. You and your physician can discuss these options and you must decide what is best for you."

The following are additional required elements of an ICF when applicable:

(1) A statement that the particular treatment or procedure may involve risks to the subject (or to the embryo or fetus, if the subject is or may become pregnant) which are currently unforeseeable.
(2) Anticipated circumstances under which the subject's participation may be terminated by the investigator without regard to the subject's consent.
(3) Any additional costs to the subject that may result from participation in the research.
(4) The consequences of a subject's decision to withdraw from the research and procedures for orderly termination of participation by the subject.
(5) A statement that significant new findings developed during the course of the research which may relate to the subject's willingness to continue participation will be provided to the subject.
(6) The approximate number of subjects involved in the study.

***Health Insurance Portability and Accountability Act (HIPAA)*** In 1996, the Department of Health and Human Services (HHS) established national standards for electronic health care transactions and national identifiers for providers, health plans, and employers as a result of the administrative simplification provisions of the Health Insurance Portability and Accountability Act (HIPAA) [26]. These standards seek to prevent the misuse of confidential medical information. Penalties for breaking HIPAA regulations are severe (see below), as such investigators and IRBs must make sure to be HIPAA compliant. Penalties for HIPAA infractions are:

(a) Be fined not more than $50,000, imprisoned not more than 1 year, or both.
(b) If the offense is committed under false pretenses, be fined not more than $100,000, imprisoned not more than 5 years, or both.
(c) If the offense is committed with intent to sell, transfer, or use individually identifiable health information for commercial advantage, personal gain, or malicious harm, be fined not more than $250,000, imprisoned not more than 10 years, or both.

## 17 CLOSING

Designing and implementing phase III clinical trials is a highly complex undertaking that requires good collaboration among all members of a well-trained multidisciplinary team. Involvement of the multidisciplinary team should be encouraged as early as possible in the design and conceptualization stage of protocol development. As with any scientific experiment, obstacles to the activation and implementation

of phase III trials will undoubtedly occur; however, well-written and designed trials will help to complete the study in a timely fashion.

## REFERENCES


1. Clinical trials.gov. (2007), Understanding Clinical Trials; available at http://clinicaltrialsgov/ct2/info/understand.
2. Therasse, P., Arbuck, S. G., Eisenhauer, E. A., et al. (2000), New guidelines to evaluate the response to treatment in solid tumors. European Organization for Research and Treatment of Cancer, National Cancer Institute of the United States, National Cancer Institute of Canada, *J. Nat. Cancer Inst.*, 92(3), 205–216.
3. Office of Human Subjects Research (OHSR). (2006), Information sheet 6: Guidelines for writing informed consents; available at ohsr.od.nih.gov/info/sheet6.html.
4. Anderson, D. M., ed. (2003), *Dorland's Illustrated Medical Dictionary*, Saunders, Philadelphia.
5. Milton, J. S. (1999), *Statistical Methods in the Biological and Health Sciences*, 3rd ed., McGraw-Hill, New York.
6. Byar, D. P., Simon, R. M., Friedewald, W. T., et al. (1976), Randomized clinical trials. Perspectives on some recent ideas, *N. Engl. J. Med.*, 295(2), 74–80.
7. Chalmers, T. C., Block, J. B., and Lee, S. (1972), Controlled studies in clinical cancer research, *N. Engl. J. Med.*, 287(2), 75–78.
8. Hill, A. B. (1951), The clinical trial, *Br. Med. Bull.*, 7(4), 278–282.
9. Peto, R., Pike, M. C., Armitage, P., et al. (1976), Design and analysis of randomized clinical trials requiring prolonged observation of each patient. I. Introduction and design, *Br. J. Cancer*, 34(6), 585–612.
10. Peto, R., Pike, M. C., Armitage, P., et al. (1977), Design and analysis of randomized clinical trials requiring prolonged observation of each patient. II. Analysis and examples, *Br. J. Cancer*, 35(1), 1–39.
11. Fleming, T. R., Green, S. J., and Harrington, D. P. (1984), Considerations for monitoring and evaluating treatment effects in clinical trials, *Control. Clin. Trials*, 5(1), 55–66.
12. Fleming, T. R., Ellenberg, S., and DeMets, D. L. (2002), Monitoring clinical trials: issues and controversies regarding confidentiality, *Stat. Med.*, 21(19), 2843–2851.
13. Smith, M. A., Ungerleider, R. S., Korn, E. L., et al. (1997), Role of independent data-monitoring committees in randomized clinical trials sponsored by the National Cancer Institute, *J. Clin. Oncol.*, 15(7), 2736–2743.
14. Fleming, T. R., Harrington, D. P., and O'Brien, P. C. (1984), Designs for group sequential tests, *Control. Clin. Trials*, 5(4), 348–361.
15. O'Brien, P. C., and Fleming, T. R. (1979), A multiple testing procedure for clinical trials, *Biometrics*, 35(3), 549–556.
16. Pocock, S. J. (1982), Interim analyses for randomized clinical trials: The group sequential approach, *Biometrics*, 38(1), 153–162.
17. US Department of Health and Human Services. (2006), Guidance for Clinical Trial Sponsors—Establishment and Operation of Clinical Trial Data Monitoring committees; available at http://www.fda.gov/cder/guidance/index.htm.
18. Heart Special Project Committee (1988), Organization, Review and Administration of Cooperative Studies (Greenberg Report): A Report from the Heart Special Project Committee to the National Advisory Heart Council, May 1967, *Control. Clin. Trials*, 137–148.

19. Ellenberg, S., Geller, N., Simon, R., et al. (1993), Proceedings of Practical Issues in Data Monitoring of Clinical Trials, Bethesda, Maryland, January 27–28 1992, *Stat. Med.*, 12, 415–616.
20. NIH Clinical Trials Committee. (1979), Clinical Trial Activity, *NIH Guide*, 8(8), 29.
21. Arriagada, R., Bergman, B., Dunant, A., et al. (2004), Cisplatin-based adjuvant chemotherapy in patients with completely resected non-small-cell lung cancer, *N. Engl. J. Med.*, 350(4), 351–360.
22. Olaussen, K. A., Dunant, A., Fouret, P., et al. (2006), DNA repair by ERCC1 in non-small-cell lung cancer and cisplatin-based adjuvant chemotherapy, *N. Engl. J. Med.*, 355(10), 983–991.
23. International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH) Expert Working Group. (1994), ICH Harmonised Tripartite Guideline: Clinical Safety Data Management: Definitions and Standards for Expedited Reporting E2A; available at http://www.ich.org/cache/compo/475-272-1.html#E2A.
24. NCI Cancer Therapy Evaluation Program (CTEP). (2005), CTEP, NCI Guidelines: Adverse Event Reporting Requirements; available at http://ctep.cancer.gov/protocolDevelopment/default.htm#adverse_events_adeers.
25. US Food and Drug Administration (FDA). (2007), Code of Federal Regulations (CFR) Title 21, Chapter 50, Protection of Human Subjects; available at https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfCFR/CFRSearch.cfm?CFRPart=50.
26. US Department of Health and Human Services. (1996), Health Insurance Portability and Accountability Act (HIPAA); available at http://www.cms.hhs.gov/HIPAAGenInfo.

# *Phase IV: Postmarketing Trials*

**Karl Wegscheider**

*Department of Medical Biometry and Epidemiology, University Hospital Eppendorf, Hamburg, Germany*

**Contents**

1 Introduction
  1.1 Life of a Drug
  1.2 Chapter Overview
  1.3 Commitment of Phase IV
2 Definition
3 Objectives
  3.1 Starting Point of Phase IV
  3.2 Open Issues
  3.3 Study Objectives in Phase IV
4 Approaches
  4.1 Controlled Randomized Trials
  4.2 Cluster-Randomized Trials
  4.3 Cross-Sectional Observational Studies
  4.4 Longitudinal Studies/Cohort Studies
  4.5 Case–Control Studies
  4.6 Aggregate-Level Studies/Ecological Studies
  4.7 Medical Registers
  4.8 Surveillance
  4.9 Systematic Review, Meta-Analysis, and Metaregression
  4.10 Health Economics/Health Technology Assessment (HTA)

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

5 Options
5.1 Data Sources
5.2 Experimental versus Observational Approach
5.3 Quality Assessment, Monitoring, and Auditing
5.4 Registration of Studies, Access to Data, and Results
6 Impact of Phase IV Trials
References

## 1 INTRODUCTION

### 1.1 Life of a Drug

The development of a drug has many aspects in common with the raising of a child. Both processes take place in typical phases that follow an unalterable order. The next stage is generally only attainable after the previous stage has been completed. Preclinical development may be seen as the embryonic stage, the first intake of the drug as the moment of birth. This moment bears special risks that cannot be foreseen even by ideal preclinical work. "Birth deformities" of the drug may become apparent only after a few human beings have been exposed. Fortunately, such observations are not frequent since the dosage will be kept low at the beginning and then increased only in small steps, analogous to the steady and rapid growth of a neonate. Phase I can be seen as the early childhood, where the drug is tested in healthy subjects in a protected environment under ideal conditions and the constant monitoring of the producers (parents) and specially trained experts charged with tending to the "education" and "socialization" of the product. Phase II is the first systematic contact with patients. For the first time, performance is demanded, and thus this phase parallels the time in school. From now on, failures have consequences. However, some kind of adaptation is allowed as far as the potential of the new drug is concerned. The ultimate indication and dosage are not yet defined and may be adapted to new experiences as the talents of the child are detected and subsequently supported and promoted. However, at some time (at the passage from phase II to phase III) the intended use of the drug has to be determined, and in phase III the drug has to prove its ability to attain the therapeutic goals for which it was presumed qualified. This point in time resembles the moment when a decision is made on the future career or field of specialization. The evaluation of the phase III trial(s) resembles the final examinations, and the licensing of a drug is the equivalent of the degree that entitles the holder to practice a particular profession.

Many people think that with the awarding of some kind of certificate for vocational or professional qualification and the first "real" employment, that children are grown-up and no longer require their parents' care. But any parent who has had the pleasure of bringing a child this far along the path to maturity knows that is not really the case because the personal life as well as professional careers are still full of challenges, competition, and risks. Support is further required, but not only requested and given by the parents who usually have a personal interest in the well-being of their now grown-up child, but also by friends and family members.

Likewise, many people think that with the licensing of a pharmaceutical the research investment has or should come to an end, and that the marketed drug

should just be used as it was licensed in order to earn money for its producers. But as is the case for any grown-up child, a drug faces competition, has to adapt to new environments and changing demands, may fail in certain situations, and has to be defended against suspicions and accusations. What is more, information is needed and has to be provided on many aspects and has to be propagated to the public. Thus, continuing critical support is needed and will be given not only by the producers but by many other parties such as government, funding institutions, researchers, heath insurances, patient organizations, and the media. Thus, there is a need and a demand for postmarketing trials of a drug, justifying the introduction of another phase, namely phase IV, for this kind of study, and this demand is met more and more frequently as we all become more and more aware of the intricate challenges with which a "grown-up", that is, a licensed drug has to contend.

However, unlike the premarketing phases, in phase IV there is no standard approach, no definite order of study sequence, no unique methodology, nor any agreement on optimality. Instead, there is a still growing diversity of study types, organizational forms, funding models, regulatory guidelines and recommendations, and a rich and complex scientific as well as public debate on licensed drugs.

## 1.2 Chapter Overview

In this chapter, a broad overview is provided on the wealth of methodological approaches to phase IV trials. After definition and an overview of the objectives, each approach is presented separately, including the motivation behind it, the standard design, the challenges one has to meet during conduct, the standard evaluation procedure, the interpretations and common misinterpretations, and the limitations of the approach, shedding light on the place of the approach in the concert of phase IV studies. Phase IV studies on statins, on the use of antibiotics for coughs, and on the safety profile of oral contraceptives will serve as examples. The chapter ends with a summary of the options for phase IV trials and the decisions that have to be met in the planning stage of a phase IV trial by sponsors and investigators.

## 1.3 Commitment of Phase IV

The life of an individual human being inevitably ends with death. At first glance this might not seem to be so inevitable for a drug, but, although its life time is not biologically limited, it is useful to think of a drug's life as finite. A drug may be well known, have a large market share, and may even be identified with its substance class; nevertheless, it is always at risk of being topped by a better (i.e., more effective, more easily administered, more tolerable, or safer) drug, or by a cheaper drug, or by a drug that is simply better promoted or more fashionable. It may be frivolously accused of serious adverse reactions or it may be superfluous because the disease is conquered, for instance, by a new vaccine. Like human beings, a drug may die from an accident, that is, is withdrawn from the market because of safety concerns, but usually it gradually loses its dominance; it remains on the market but exists only in a niche (a kind of "retirement") and disappears some day without many people noticing it, possibly because it is no longer economically viable. It may survive in another form, as a generic or as part of a combination, but the patent and brand name, its identifying characteristics cease to exist.

In summary, with a few exceptions, a drug can, and certainly will, die at some indefinite future date. Undoubtedly, the death will be a severe loss not only for the producers but also for the community since the deep knowledge of the drug gained by experience over an abundance of years of application collected in phase IV studies will at least in part die with the drug. After all, an established drug therapy provides stability and confidence, like an old friendship where you know the strengths and weaknesses of the friend and have learned to live with them. Does this mean that phase IV research is futile or pointless? Should we refrain from any post-marketing research, enjoy the licensing success, and avoid the unavoidable risks of any medical research? Quite the contrary is the case. If we do not study drugs in the market, we are in danger of being surprised by speculations that can be neither proven nor rebutted, and nevertheless can end the life of a drug by ruining its image. Lipobay and Vioxx [1] and Waxman [2] are prominent examples of this. What is more, we may miss opportunities for the use of a drug that may never come to light, and this may shorten its lifetime. A series of trials with different approaches is thus not only recommendable for medical and economic reasons but an ethical commitment for manufacturers, scientists, and society alike and thus not only be taken seriously but encouraged.

## 2 DEFINITION

A phase IV trial is a clinical study where the investigational therapy includes the use of a licensed drug in the approved indication. (Studies of a licensed drug in a new indication are usually referred to as phase IIIb trials.) The term *trial* is usually reserved for studies where the drug under investigation is administered for study purposes. In a strict sense, observational studies on routine use or medical registers are not covered by this definition. Nevertheless, the term *clinical trial* is sometimes used somewhat more loosely for clinical studies in general. In this chapter, we will reserve the term *trial* for interventional studies. It should, however, be noted that this chapter will cover other approaches as well in order to give a more complete account of the diversity of options covered by the concept *phase IV study*.

## 3 OBJECTIVES

### 3.1 Starting Point of Phase IV

At the end of phase III/beginning of phase IV, there is already a considerable body of knowledge about a drug. The proof of principle was given in phases I and II. Phase II/III added some knowledge on dosage and application. From phase III we know that the drug works in a variety of clinical settings and is able to improve clinical relevant endpoints.

On the other hand, the knowledge of the drug at the time of market introduction will always be limited in some sense due to the legal restrictions, economic considerations, and the primary targeting of approval. Up to phase III, the application of the drug has taken place under the conditions of controlled trials, that is, under

steady observation and quality control, in environments that were fully aware of the research motivation behind the application, and with strictly controlled inclusion and exclusion criteria defined in order to receive a homogenous study population to minimize the risk of a negative result. In many phase III trials, the criteria are so cautiously defined that only a small percentage of the future users of the drug can be included, resulting in study populations that are not representatively selected. At the time of licensing, we only have experience on the application of a drug in a very specific population, for a limited range of doses, under absolutely ideal circumstances, and in a number of patients too small to address less frequent complications and safety concerns that nevertheless may become very relevant if the exposure to the drug is extended for routine care of patients.

### 3.2 Open Issues

Correspondingly, optimization of use and safety concerns are the most prominent reasons for conducting postmarketing trials. Just recently, questions of the financial balance have joined the club and sometimes even dominate, starting as add-ons to clinical study projects and then becoming an independent research area. The more traditional form of studies aiming at economic aspects consists of marketing trials that address the propagation and use of a drug. These studies are frequently suspected of being undertaken predominantly for the purpose of propagation by merely performing the study and not for research purposes. This suspicion indicates that there may be a conflict of interest between different study objectives in phase IV. Conflicts of interest are a minor concern in phases I–III but are rather typical for phase IV studies where the scope of research questions is broader and more stakeholders with potentially different interests are participating in the research process.

### 3.3 Study Objectives in Phase IV

Table 1 gives an overview of common study objectives in phase IV, organized according to the four main research areas of phase IV. Optimization of use is required since many decisions made in phases I–III were taken for the purpose of gaining a license for the drug. Thus, only the most promising formulations and dosages are developed and tested in a targeted study population. What is more, the decisions in phases I–III were based on experiences from small samples that were not necessarily conclusive in nature. A further process of fine-tuning is therefore desired and should be based on a solid empirical base, big enough to allow even sophisticated considerations. The first six objectives belong to this category. Furthermore, the prelicensing trials were powered for overall evaluations. But patients' reactions to a drug may vary considerably, and thus an adaptation of the therapy to different subgroups of patients and several influences on drug efficacy are required and necessitate further studies, as the other objectives of this section demonstrate.

Safety issues are critical for the life of a drug. At the time of licensing, the finalized trials have had optimal power for the proof of efficacy. However, the power of the complete database is usually far too small to draw reliable conclusions about

**TABLE 1 Objectives of Phase IV Studies**

| | |
|---|---|
| | *Optimization of Use* |
| IMPROVING TREATMENT STRATEGY | |
| Opt-01 | Optimization of application form |
| Opt-02 | Optimal dosage for different groups of patients, exploration of dose–response curve |
| Opt-03 | Optimal duration of therapy |
| Opt-04 | Optimization of the circadian pattern |
| Opt-05 | Evaluation of further clinical endpoints |
| Opt-06 | Implementation of the drug in complex therapeutic strategies |
| ADAPTATION TO SPECIAL PATIENT GROUPS | |
| Opt-07 | Specification of indication, improvement of diagnostics to establish the indication |
| Opt-08 | Adaptation to age, gender, race, in particular, pediatric or geriatric use of the drug |
| Opt-09 | Use in patients with comorbidity |
| Opt-10 | Assessment of individual variability of drug effects, metabolism |
| EXPLORATION OF EFFECT MODIFIERS | |
| Opt-11 | Identification of patient characteristics that increase or diminish drug effects, in particular, education, habits, lifestyle |
| Opt-12 | Efficacy in various clinical settings |
| Opt-13 | Effect modification due to environmental factors, in particular, food intake, sociodemographic factors |
| Opt-14 | Compliance studies |
| | *Safety Issue* |
| Saf-01 | Evaluation of the frequency of routine side effects/expected adverse effects |
| Saf-02 | Detection and enumeration of unexpected or rare side effects and delayed effects |
| Saf-03 | Cumulative effects of long-term use, accumulation effects in patients |
| Saf-04 | Interactions with concomitant therapies, in particular, drug interactions |
| Saf-05 | Permanent safety surveillance to track new developments that raise safety concerns |
| Saf-06 | Addiction and abuse |
| Saf-07 | Inadequate prescriptions |
| Saf-08 | Risk/benefit ratio in different subgroups |
| | *Health Economy* |
| Eco-01 | Cost/benefit relation, cost effectiveness |
| Eco-02 | Head-to-head comparison to alternative therapeutic approaches, in particular, competing drugs |
| Eco-03 | Economic evaluation of complex therapeutic therapies that incorporate the use of the drug |
| | *Marketing* |
| Mar-01 | Accompanying and assisting the transition from clinical development to marketplace |
| Mar-02 | Demonstration and implementation of the proper use of the drug |
| Mar-03 | Investigation and improvement of the complicance of physicians as well as of patients |
| Mar-04 | Acceptance and performance of the drug in different markets, e.g., changes due to a switch from prescription to over-the counter marketing |
| Mar-05 | Exploration of convenience of application, patient satisfaction, effects on quality of life |

safety issues. This is not only due to the fact that serious adverse events are in most cases rare and may appear with some delay but also because the random fluctuation in the normal population, the treated disease itself and other medication and treatment will result in adverse events that may or may not be caused or aggravated by the drug under study. Extreme sample sizes are required to detect and validate adverse drug reactions caused by a drug in the presence of a natural background of events. Thus, other approaches in addition to randomized trials are required to address these issues, for purely economic reasons. Nevertheless, this kind of evaluation is mandatory since the public is extremely reluctant to accept even small safety risks for the sake of an average benefit in the patient population. The main reason for this reluctance presumably is the fact that a serious adverse event is always personal, while the benefit of a drug is statistical and anonymous. Thus, risk–benefit calculations do not dispense manufacturers from the commitment to meticulously search for each case when a drug might have caused harm. The objectives under the corresponding heading in Table 1 demonstrate how many different aspects of safety have to be addressed in one way or the other as soon as a drug is on the market.

Health economic studies traditionally investigate cost–benefit balances and the cost effectiveness of a drug. The limited availability of funds drives the increase in the number of such studies in phase IV. Economic studies usually require bigger sample sizes and other evaluation methods than clinical trials. Hence, economic evaluations accompanying phase III trials are usually not sufficient, and planned phase IV studies are required. Increasing competition in health markets let people ask for more direct head-to-head comparisons between therapies that help to define the place and potential function of a drug in the health care environment. While the two upper objectives in that section aim at a drug-related evaluation, more and more people take the perspective that the utility of drugs has to be judged according to their contribution to complex interventions in the health system. A statin, for instance, may be studied for its cost effectiveness (Eco-01) (see Table 1); it may be studied in direct comparison to another statin or a different lipid-lowering therapy (Eco-02), or as part of a multiway lipid-lowering program for a certain target population (Eco-03).

Marketing begins with the transition from clinical development to the marketplace. It is useful to accompany this process with a scientific evaluation. However, studies for market introduction are frequently understood as opportunities to promote the drug to physicians and patients as yet unfamiliar with the substance, making the study something akin to advertisement. The scientific aspect of the study may thus be compromised, the more so if the study is explicitly performed for the purpose of the demonstration and establishment of study drugs in the market. While these two study objectives are frequent in phase IV, they are not generally accepted by the scientific community as part of research. The scientific portion plays a greater role in compliance studies, in studies on the performance of the drug under different distribution regimes, and in the evaluation of patients' experiences and opinions when applying the drug. These three objectives are sometimes undervalued. In fact, availability, compliance, and patients' satisfaction may contribute considerably to the success of a theoretically useful drug in a certain health care setting and thus may be very important for all health system participants in order to actually reap the potential benefit of a new drug.

## 4 APPROACHES

The diversity of objectives in phase IV is reflected by a diversity of approaches. This section describes 10 approaches, defined by a characteristic methodology. Each of these approaches is useful for a certain subset of the objectives listed in Table 1. First, the motivation behind the approach is given. The design is described, followed by a summary of the typical challenges in performing such a trial. Then an accompanying example is provided. This is followed by an account of the standard evaluation. In the final sections, recommendations for interpretation of the results are offered and the limitations of the approach discussed.

### 4.1 Controlled Randomized Trials

***Motivation*** Scientific conclusions are always based on comparisons. However, in everyday life as in studies, comparisons are always in danger of bias, that is, it may be that a difference is caused by other factors than the condition (in this case the drug) under study. Bias can never be definitely excluded, but it can be controlled by probability. For this purpose, a random allocation of the drug to the patient is required.

Correspondingly, the randomized controlled trial (RCT) is the gold standard for phase III trials. In phase IV, RCTs are the preferred approach for all objections related to the optimization of use, and for objectives Eco-02 and Mar-03 (see Table 1), whenever it is possible to perform such a trial. RCTs may be performed as superiority trials to prove some advantage of a drug over the comparator or as non-inferiority trials proving at least comparable efficacy of a drug to a competing active treatment.

***Design*** The RCT is prospective. A study protocol has to be prepared that precisely defines the study population and describes the inclusion procedure. After inclusion, patients have to be randomized, preferably by means of an external randomization procedure. Further treatment, observation, and follow-up should be identical for all groups under comparison. A sample size calculation has to be performed in advance for the study to be adequately powered for the primary endpoint.

***Challenges*** There are potential biases that may take place after randomization. To avoid differential reporting, blinding of patients, treating physicians, and readers is recommended. Differential drop-out or missing values may compromise the study and should be avoided as much as possible. Major challenges in conducting such trials are recruitment problems and limited compliance.

***Examples***

1. A phase IV head-to-head comparison between two statins was performed to test the assumption that 40 mg pravastatin daily is at least equivalent to more aggressive lipid reduction by 80 mg atorvastation daily after acute coronary syndrome [3]. The study demonstrated superiority of the more aggressive therapy.

2. To assess the efficacy of pravastatin in the elderly, a double-blind randomized trial versus placebo (PROSPER) was performed in patients aged 70–82 with a history of or risk factors for vascular disease [4]. In the conclusion, the extension of the treatment strategy from middle-aged to elderly people is recommended.

***Evaluation*** In the simplest case, the evaluation of a superiority trial consists of two-group comparisons by Fisher's exact tests or chi-square tests for binary outcomes and by two-sample $t$ tests or nonparametric tests for continuous outcomes. An adjustment for baseline differences between patients may increase the power of the trial but is not mandatory. The evaluation of noninferiority trials is more complex. In any case, the presentation of the results using point estimates and confidence limits is recommended.

***Interpretation*** If the study conduct was regular, cause–effect conclusions can be drawn for the primary endpoint and for the significant secondary endpoints if the study is positive. Results that are not significant are harder to interpret due to the fact that failure to prove an assumed effect may follow from lack of power or low quality of the study.

The intergroup comparisons of a randomized trial may be considered unbiased with a high level of confidence. This is even true if baseline differences between groups are observed. Randomization is expected to evenly distribute chances (success probabilities) between groups, not each of the baseline characteristics. Contrary to a widespread notion, the correctness of randomization cannot be checked empirically but only by checking of the randomization procedure.

Subgroups should be prespecified and never be analyzed separately but using interaction tests. Usually, the trial is not adequately powered for subgroup considerations.

***Limitations*** The ability to draw generalized conclusions about the results crucially depends on the representativeness of the study population and may be limited due to a restrictive patient selection. RCTs are optimized for intergroup comparisons but not for extrapolation to routine health care in general. Analyses that where not prespecified and any kind of intragroup comparisons are not protected by randomization and may be biased to a relevant extent, although they are derived within the framework of an RCT.

### 4.2 Cluster-Randomized Trials

***Motivation*** Drug interventions usually address the individual patient. However, several patients may be cared for by one and the same health care institution or treated by the same physician. The drug intervention itself may even be provided within the framework of a program that includes several physicians and institutions and thus groups of patients. If so, patients within one institution or program tend to be more similar in their results than patients from different programs. This kind of effect is called a *cluster effect*. If cluster effects are present, it may be recommended not to randomize patients to drugs but to randomize institutions to drug interventions on the cluster level.

***Design*** In a cluster-randomized trial, clusters (e.g., general practitioners, hospitals) are randomized to the study drugs, and all patients within a cluster receive the same drug according to the same schedule. Patients should be representative of their cluster, for example, by sequential inclusion or random selection. Patients within a cluster are allowed to communicate and exchange experiences.

In phase IV, cluster-randomized trials are typically applied when the introduction and propagation of a new drug is under study or when drugs are incorporated in prevention programs or general health programs.

***Challenges***

1. A sufficient number of clusters (30 or above) have to participate. The required total number of patients frequently is a multitude (3–5 times) of the required sample size in trials with randomization of individual patients, depending on the extent of the cluster effect.
2. Institutions in the intervention group are confronted with a significant amount of effort and extra work. Institutions in the control group have only few advantages from participating in the study. Thus, recruitment problems are frequent. Special incentives for control group institutions may be helpful.
3. Clusters should not cooperate.
4. Blinding is frequently difficult or impossible.

***Examples***

1. To study the clinical efficacy and the cost effectiveness of a rosuvastatin-based compliance program, general practitioners were randomized to take part with their patients in the program or not [5].
2. In a cluster-randomized trial, Coenen et al. [6] study the effect of a tailored peer intervention to implement a practice guideline for acute cough for optimizing antibiotic prescribing.

***Evaluation*** In the presence of cluster effects, classical statistical evaluation methods are no longer valid, resulting in too many significances. Correct, but not very efficient, is an evaluation where cluster averages are calculated and compared between groups as if they came from individual patients. The best way of analyzing clustered data is the use of a multilevel model that allows the distinction between cluster-level effects and patient-level effects. Since relevant baseline differences are more frequent than in patient-based RCTs, due to the usually small number of clusters, adjustment for baseline is recommended.

***Interpretation*** Cluster-randomized trials are less balanced and thus more vulnerable to biases than individually randomized trials. On the other hand, cluster-randomized trials are better capable to cope with real-world situations and thus may be more externally valid.

***Limitations*** Recruitment problems, baseline imbalances, and differential dropouts are frequent and may compromise the between-group comparisons or the

generalizability of the trial. Despite successful randomization, it may be difficult to maintain a comparable study conduct in all random groups due to external influences on institutions.

### 4.3 Cross-Sectional Observational Studies

***Motivation*** Hypotheses concerning effect modification or suspicions of adverse drug reactions are frequently gained from samples where drug exposure data, patient characteristics, and environmental factors are collected from patients at one point in time. These studies are quick and easy to perform and evaluate since only correlations have to be analyzed, but they do not provide evidence on drug efficacy and are just hypotheses generating. In phase IV, these studies provide first impressions on potential factors for optimization of use or on potential safety risks.

***Design*** Planned observational study. Selection of a representative sample from a target population, collection of data from available sources, and one-time application of questionnaires.

***Challenges*** Completeness of sample and records.

***Examples***

1. By means of a questionnaire, Coenen et al. [7] study the reasons for prescribing antibiotics in general practices. They explore the answers by factor analysis, identifying three latent dimensions, one of them being "nonmedical reasons."
2. In an observational study nested in an RCT, Little et al. [8] studied the influence of patient pressure, perceived patient pressure, and perceived medical need on the investigations, referral, and prescribing decisions of general practitioners by logistic regression.

***Evaluation*** Correlation and regression analyses, data-mining techniques. Frequently, a reduction of dimensionality is helpful, for example, by factor analysis or structural equation modeling.

***Interpretation*** The study results in the demonstration of associations, that is, of coincidences that may or may not be substantial.

***Limitation*** The numerical extent of associations depends heavily on the other variables in the model, and the interpretation is rarely unique. Associations thus depend on other associations and cannot be generalized. Even if the associations are stable over different models and times (if cross-sectional analyses are repeated), the associations do not provide evidence on cause–effect relationships. Likewise, missing associations do not prove the absence of cause–effect relations. Moreover, cross-sectional studies are not able to demonstrate individual courses or any changes with time. Thus, these kinds of studies can only be hypotheses generating and do not provide direct evidence and require additional studies with higher levels of evidence.

### 4.4 Longitudinal Studies/Cohort Studies

***Motivation*** The duration of phase III trials is usually limited to the minimum required to demonstrate clinical efficacy. Long-term efficacy and safety remain to be demonstrated in phase IV. While it is preferable to study long-term effects in RCTs, it is frequently not possible to maintain a randomization regime over many years, for economic, practical, or ethical reasons, the more so as the investigational drug has proven to be advantageous in phase III. A prospective study of drug effects can then be performed by building a well-defined cohort of patients who are routinely treated by the drug under investigation and by following the cohort for the period of drug intake. This cohort can then be compared to other cohorts, treated differently, from the same study or cohorts from other studies ("historical cohorts").

In phase IV, cohort studies are commonly applied to evaluate long-term tolerance and safety but also to study the associations between changes in intermediate or surrogate parameters and clinical endpoints and to learn about long-term morbidity and prognosis and their determinants.

***Design*** Cohort studies are planned epidemiological, purely observational studies. Groups of patients are recruited as for RCTs, but there is no random allocation of study drugs to patients. Instead, patients qualify for the study by the treatment ("exposure") they receive in clinical routine. Data collection and follow-up are predefined and are performed with identical procedures for all patients, that is, additional follow-up examinations may be performed for study purposes.

***Challenges*** Completeness of follow-up is a daunting task, in particular if the drug intervention is lifelong.

***Examples***

1. Simes et al. [9] use the cohorts of the LIPID trial on pravastatin to study the relationship between lipid levels and clinical outcomes on long-term intervention.
2. In order to study the safety profile of rosuvastatin beyond the analyses within phase II and phase III trials, Shepherd et al. [10] formed new cohorts from 27 phase II and phase III trials with different comparators, according to drug and dose ranges. Frequencies of adverse events per year of drug intake were calculated for each of these cohorts.

***Evaluation*** While in RCTs adjustment for baseline is only optional to increase the power of the study, in cohort studies it is mandatory to minimize the selection bias resulting from lack of randomization. Repeated measurements require special statistical model building, called longitudinal data analysis.

***Interpretation*** Treatment effects are usually quantified by regression coefficients. Effects may be time dependent and effect modifiers may be present. The interpretation of results of longitudinal data analysis may thus be involved and requires professional support.

***Limitations*** Even with extended baseline adjustment or inclusion of time-dependent covariates, there is still a danger of hidden bias that cannot be controlled by probabilistic calculations. The evidence level of cohort studies is thus generally considered lower than that of RCTs. On the other hand, actual patient populations are usually better represented by cohorts than by RCT populations.

Cohort studies usually take a long time and thus are not appropriate to raise or to give quick answers to safety concerns. However, in the long run they may contribute substantially to the precise specification of risk and to the exploration of mechanisms.

### 4.5 Case–Control Studies

***Motivation*** Safety concerns, however substantiated, cannot wait for resolution by prospective long-term, follow-up studies such as RCTs or cohort studies. Retrospective studies are an alternative that can strengthen or mitigate the suspicion of a side effect in a shorter time interval. Case–control studies (directly comparing cases with and appropriate controls without the symptom suspected being drug related) and similar study types (such as case cohort studies where exposure and nonexposure time intervals are compared within the same patients) are widespread retrospective study arrangements.

In phase IV, case–control studies are widely used for the analysis of specified safety concerns (Saf-02, Saf-03, Saf-04, Saf-06, Saf-07) or for the study of the influence of environmental factors and lifestyle on drug effects (Opt-11, Opt-13); see Table 1.

***Design*** Cases are systematically detected based on patients' records. Controls matched for other potential factors (e.g., age, sex, primary disease) are selected from a comparable population without symptoms. Drug exposure is compared between cases and controls.

***Challenges*** The validity of the study decisively depends on the use of a comparable control group. Since there is usually no "natural" control group, the choice has to be justified, or several control groups should be used. Further, a bias-free selection of cases and controls has to be guaranteed. Since the exposure (and possibly some baseline characteristic that have to be determined for adjustment) has to be investigated for past periods of time, lack of documentation may compromise the study.

***Example*** Suissa et al. [11] report the results of a case–control study comparing the risks of venous thromboembolism between oral contraceptives of the first and third generations. The data stimulated a fierce debate for years on the existence of an elevated risk in third–generation brands, several new studies, regulatory activities, and even litigation due to the lack of data from trials with a higher level of evidence.

***Evaluation*** The standard statistical analysis requires the fit of a conditional (i.e., stratified) logistic regression model to the data.

***Interpretation*** Drug effects are determined by adjusted odds ratios calculated from the final model that are directly comparable to odds ratios from prospective studies.

***Limitations*** Case–control studies are vulnerable to any type of conscious or subconscious selection. They are discussed predominantly with respect to the appropriateness of the choice of controls. Results heavily depend on the choice of model. Thus, results are rarely unanimously accepted. Evidence is considered to be moderate and only supportive to other findings.

### 4.6 Aggregate-Level Studies/Ecological Studies

***Motivation*** Relevant data may not be available on an individual level but only for communities, states, or countries (e.g., prescription rates, frequencies of side effects, environmental exposure, costs for certain medical interventions). In these cases, analyses can only be performed on an aggregate level. In phase IV, this situation frequently occurs in the exploration of environmental effect modifiers, health economic studies, and studies on safety.

***Design*** Collection of data from different aggregate data collections or linkage with individual data bases.

***Challenges*** Linkage of different data sources requires the existence of comparable identifiers. Different databases may use different definitions of items that do not match. Selection criteria may vary. Data quality may vary.

***Example*** Juurlink et al. [12] use aggregate-level prescription claims data and combine them with hospital records in order to analyze changes in the rates of hyperkalemia after publication of the Randomized Aldactone Evaluation Study (RALES). For this purpose they use time-series analysis methods.

***Evaluation*** Statistical evaluation has to take the clusters and the resulting correlation structure into account. This can be done by using time-series analysis, Generalized Estimating Equations (GEE) models, or multilevel models, by meta-analyses or metaregressions, among others.

***Interpretation*** Aggregate-level studies are purely observational. The interpretation follows the rules for patient-based cross-sectional studies; see above. Additionally, the different data levels have to be taken into account when interpretations are given.

***Limitations*** In addition to the limitations of other cross-sectional studies, aggregate-level studies are vulnerable to the so-called ecological fallacy. It may occur that an association observed between clusters of individuals cannot be confirmed on an individual level or may even reveal an inverse association. Likewise, associations present on an individual level may not show up in the aggregate level analysis. In particular, trend parallelism may be spurious since any simultaneous trends will be correlated even if there is no causal relation, just by random

coincidence. Thus, while aggregate-level analysis are popular with politicians, media, and the public due to its handiness, they may be and frequently are extremely misleading, the more so as the control of other potential influences is usually poor.

### 4.7 Medical Registers

***Motivation*** Due to the rapid development of electronic documentation in medicine, routine data are available for most patients and clinical conditions. The quality of these data is improving since it is being used more frequently for the management and financing of health institutions. However, documentation standards still differ. Nevertheless, it has in the meantime become possible to routinely collect such data for certain groups of patients from comprehensive databases by defining interfaces and by regular data exchange, with a justifiable amount of effort. Such databases, called medical registers or pharmaco-economic databases, allow quick answers to questions arising on the use and the safety of several drugs at the same time. Registers are almost exclusively phase IV, useful for almost all of the objectives of Table 1.

***Design*** Medical registers and pharmacoepidemiological databases try to give an image of a defined target population that is as complete as possible. Cross-sectional as well as longitudinal data are collected on an individual level. There will usually be no interventions, neither for experimental nor for observational purposes.

Registers require a protocol where the procedures are defined. However, other than in planned epidemiological studies, no research hypotheses are formulated in advance but will be defined according to the current state of debate, while further data are collected. If complete and high-quality data are available from routine databases, register population can even be defined retrospectively if organized by inception.

***Challenges*** The validity of a register depends on its completeness and data quality.

***Example*** Graham et al. [13] retrospectively define drug-specific inception cohorts of statin and fibrate users to study the incidence of hospitalized rhabdomyolysis.

***Evaluation*** Registers and pharmacoepidemiological databases include cross-sectional and longitudinal elements and are thus similar to epidemiological studies, but broader and more exploratory than those approaches. The evaluation requires complex statistical modeling. Simple univariate or bivariate statistics usually are misleading.

***Interpretation*** Registers are predestinated for the determination of rates and frequencies under changing conditions. They allow the study of associations but face the same problems as epidemiological studies.

***Limitations*** Registers are purely observational. While their findings are most representative, they do not provide evidence for cause–effect relationships.

### 4.8 Surveillance

***Motivation*** Once a substantiated safety concern has been established, regular surveillance becomes mandatory. Surveillance is a typical phase IV safety activity.

***Design*** Spontaneous reporting systems are the earliest type of surveillance. Other ways of organization are regular survey, sentinels (a selection of institutions that report the rates of interest on a regular base), and specific pharmacoepidemiological databases/registers on drug utilization. The use of Internet platforms is still experimental.

***Challenges*** Traditionally, underreporting by physicians is the greatest challenge of surveillance. Considerable effort is required to improve participation. Independence of the responsible body, transparency, a careful practice of publication, and open access can serve to convince physicians of the usefulness of their contribution.

***Examples*** Several reporting systems and routine databases are run by manufacturers, agencies, and pharmacological academic institutions.

***Evaluation*** Prespecified safety concerns can be addressed with advanced statistical standard methodology. Extraction of unanticipated signals within huge databases is a continuing challenge. Data-mining techniques and in particular graphical tools are used, but no general methodological standards exist.

***Interpretation*** The appraisal of surveillance data should take into account both aspects of statistical significance and those of medical relevance.

***Limitations*** The perception of the data is usually not free from conflicts of interest. At present, no generally accepted rules exist as to when or which actions should be taken.

### 4.9 Systematic Review, Meta-Analysis, and Metaregression

***Motivation*** With the extended use of a drug after approval, the number of studies and the body of evidence increases rapidly. Information has to be summarized not only to facilitate an overview but also to realize the advantages in statistical power and differentiation that can be gained by combining data from different studies. Reviews and meta-analyses that summarize the results are recommended for all objectives in phase IV.

***Design*** Reviews can be performed on a nonformal base as systematic reviews with a qualitative summary, or on a formal base with an established statistical meta-analysis technique with quantitative summaries. The formal analysis can be performed on the patient level if it is possible to combine the individual data of several studies into one database, or on the study level if only summary measures are available by using techniques of formal meta-analysis or metaregression. Every review and meta-analysis requires the prespecification of a protocol.

***Challenges*** Completeness of studies/research results is difficult to attain due to limited access to studies not published in English and in established journals. Literature searches may also be incomplete due to the reluctance of journals to publish negative results (resulting in publication bias). Reporting standards may differ between studies to an extent that prevents pooling of results. Direct access to source data may be refused.

***Examples***

1. The Control Treatment Trialists' Collaborators combined the data of 90,056 participants from 14 randomized statin trials to further evaluate the efficacy and safety of cholesterol-lowering treatment [14].
2. The Cochrane collaboration publishes systematic reviews and aggregate-level meta-analyses to several medical fields, provided by working groups of experts.

***Evaluation*** Systematic reviews should follow the rules of the Cochrane collaboration for literature search and appraisal of the quality of the collected studies. For aggregate-level meta-analysis, an established statistical methodology exists. Random effects models are generally preferred to demonstrate potential heterogeneity between studies. Individual data-based meta-analyses should be analyzed by statistical modeling similar to the model building in epidemiological studies, but including the factor "study" to demonstrate heterogeneity.

There is an open debate as to what extent reviews and meta-analyses should rely on high-quality studies with a high level of evidence (RCTs) only or should include information of nonrandomized studies.

***Interpretation*** Interpretation is usually based on overall summary measures, but heterogeneity should be considered as well.

***Limitations*** Even if only high-quality RCTs are used, this does not automatically translate into a high-quality meta-analysis. Selection biases and an insufficient amount of data may limit interpretation. Aggregate-level meta-analyses and metaregressions may be compromised by the ecologic fallacy, as with individual aggregate-level studies, despite randomization of the contributing studies.

### 4.10 Health Economics/Health Technology Assessment (HTA)

***Motivation*** Costs of a useful drug intervention may be too high to be financed in certain health care settings, or advantages of drug interventions may be judged negligible because the prices for this advantage may be found to be too high. Thus, health economic evaluations or the study of changes to be expected after the introduction of a new health technology are requested to a steadily growing extent.

***Design*** Costs of interventions can be evaluated directly as part of a clinical study or determined separately and then combined with clinical data. HTA uses several data sources and combines them in probabilistic models based on Bayesian decision trees.

***Challenges*** Costs originate at several levels, directly in connection with health care (direct costs) and indirectly due to the consequences of being ill or disabled. A complete evaluation of costs up to the regarded level requires the use of many sources of information. For cost-effectiveness calculations or HTA, a one-dimensional measure of efficacy is required. If several aspects of efficacy are to be taken into regard, a summary measure combining information on different scales has to be used. Quality-adjusted life-years [15] is such a concept that incorporates quality of life into the efficacy measure, but this kind of measure is not unchallenged.

***Example*** Hirsch et al. [16] study the cost effectiveness of rosuvastatin as compared to other statins by retrospective drug cost evaluation for the patients in a head-to-head RCT comparing rosuvastatin to atorvastatin.

***Evaluation*** Costs can be compared between study groups by classical statistical methods, but care is required since cost distributions are usually extremely skewed and variable. Baseline adjustment is thus recommended. Cost effectiveness is mostly measured by incremental cost-effectiveness ratios (ICER) that reveal additional variability since denominator and numerator are subject to random variation. Since costs may vary in different settings and time periods, it is good practice to add sensitivity analyses that demonstrate the sensitivity of the results to price changes or clinical improvements.

Health technology assessment predominantly uses Markov chain Monte Carlo (MCMC) methods for the calculation of probabilities along complex decision trees [15].

***Interpretation*** Health economic analyses result in imprecise estimations of costs for improvements. HTA also gives an impression of the future distribution of outcomes in patient populations.

***Limitations*** The results do not allow automatic decisions on the implementation of new technologies since the definition of boundaries is a question of values and political priorities. The selection of summary measures inevitably requires the definition of subjective weights that may not be accepted by everyone.

## 5 OPTIONS

Before performing a phase IV study, authorities, sponsors, and investigators have to make some decisions crucial to the impact of the results once the study is finished. The following represent some remarks on the decision criteria.

### 5.1 Data Sources

Studies can be performed using data from literature and completed studies or collecting own data. Actually, these two options are less an alternative than usually thought. A study should always start with a review of the present knowledge, including the knowledge to be expected from ongoing studies. If information is judged to be insufficient, existing databases that have not yet been evaluated should be

checked for accessibility, data quality, and potential impact of an analysis. New studies should be planned in a way that they optimally contribute to future reviews and meta-analyses. The development of strategic plans incorporating series of parallel studies, studies with different methodology, and performance of studies conditional on the results of previous studies are as recommended in phase IV as they were in phase III. However, it has to be remembered that in phase IV, other than before approval, the manufacturer is not the only potential sponsor of studies. Since the drug is licensed, several players can perform studies as long as they have a sufficient budget at their disposal. There may be conflicts of interest between different parties. Joint sponsorships resulting in linked databases of different origin may help to regulate such conflicts of interest in advance and certainly are interesting options for future phase IV activities.

### 5.2 Experimental versus Observational Approach

Among methodologists there is a continuing debate whether phase IV studies on safety should be predominantly RCTs or observational studies [17]. As an example, the Cochrane Institute and many supporters of evidence-based medicine prefer RCTs and meta-analyses, attributing them the highest level of evidence since they feel that bias and spurious cause–effect relationships are to be feared most. Pharmacoepidemiologists refer to the lack of representiveness of many RCTs and argue that RCTs clarify safety concerns with an unacceptable time lag insufficient to guarantee patients' safety. The debate was intensified by recent withdrawals or black-box warnings of approved drugs due to new data on previously discussed potential side effects [1]. However, the debate should not be allowed to obscure the fact that the two approaches are not exclusive alternatives. In fact, experimental as well as observational studies contribute unique information on drugs. The question should thus rather be whether the amount of unbiased information and representative information is sufficient when a funding decision has to be met. What is more, both approaches are able to validate each other, the more so if the different study types are linked directly; an epidemiologic study may well be nested within an RCT; an RCT can more easily be performed if patients are recruited within a registry framework. Phase IV gives even more freedom than the premarketing phases to combine different activities into common projects.

### 5.3 Quality Assessment, Monitoring, and Auditing

At present, the standards of phase III are not universally applied in phase IV, as far as data quality and control are concerned. The absence of regulatory requirements and the lower level of evidence on drug efficacy of some types of phase IV studies may have contributed to this situation. However, from a methodological as well as from an economical or political point of view, there is no reason why in a pharmacoepidemiological database on drug safety, for instance, a poor data quality should be more acceptable than in an RCT for optimization of use. The collection and evaluation of spurious data is per se not reasonable and not profitable and should be avoided, the more so as spurious findings, for example, on safety, can have a similar impact on public opinion to that of correct data can have.

### 5.4 Registration of Studies, Access to Data, and Results

In phase III, trials are performed to gain knowledge about a drug and to convince agencies of the approvability of a new drug. Phase IV trials always have to address a broader audience. Not only agencies, but also physicians, patients, the scientific community, and the public have to be convinced as far as the virtue of a drug is concerned. For this purpose, it is essential that the audience trusts the results and has confidence in the proper performance of a study. Important milestones on the way to trustworthiness are the registration of all study activities to avoid the impression that unwanted results are covered up and open access to the data and the results once a study is finished.

## 6 IMPACT OF PHASE IV TRIALS

Up to now phase IV trials and studies are frequently understood as voluntary exercises that may or may not be performed and do not need coordination or comprehensive development plans, special quality investments, transparency, or scientific evaluation. However, this attitude contrasts with the enormous impact that phase IV trials or the lack of these trials can have on the reception and use of a drug. If extended RCTs for the optimization of use are lacking, no safety studies or registries exist, economic consequences are not clear, and the relationship to competing drugs and other therapies is open. When all available data are locked in secret databases and reserved for the management of the manufacturing company, the position of a drug on the market is extremely vulnerable, and even random observations and rumors cannot be challenged due to a lack of data for scientific authorities who are able to clarify concerns. The performance of phase IV thus is in everybody's best interest.

The limited realization of phase IV studies may in part be due to the fact that the postapproval period is only sparsely regulated. However, there are more and more voices asking for a better regulation. Pharmacovigilance in particular is thought to be organized insufficiently and ineffectively. Rules are not clear, the predictability of legal decisions is low, and the twists of public attention and opinion are not predictable either. The only way to overcome the resulting risks is the production of solid data and transparent analyses that establish generally accepted facts upon which future work can rely.

## REFERENCES

1. Psaty, B. M., and Furberg, C. D. (2005), COX-2 inhibitors—lessons in drug safety, *N. Engl. J. Med.*, 352, 1133–1135.
2. Waxmann, H. A. (2005), The lessons of Vioxx—drug safety and sales, *N. Engl. J. Med.*, 352, 2576–2578.
3. Cannon, C. P., Braunwald, E., McCabe, C. H., Rader, D. J., Rouleau, J. L., Belder, R., Joyal, S. V., Hill, K. A., Pfeffer, M. A., and Skene, A. M. (2004), Intensive versus moderate lipid lowering with statins after acute coronary syndromes, *N. Engl. J. Med.*, 350, 1495–1504.
4. Shepherd, J., Blauw, G. J., Murphy, M. B., Bollen, E. L. E. M., Buckley, B. M., Cobbe, S. M., Ford, I., Gaw, A., Hyland, M., Jukema, J. W., Kamper, A. M., Macfarlane, P. W.,

Meinders, A. E., Norrie, J., Packard, C. J., Perry, I. J., Stott, D. J., Sweeney, B. J., Twomey, C., Westendorp, R. G. J., and the PROSPER Study Group (PROspective Study of Pravastatin in the Elderly at Risk). (2002), Pravastatin in elderly individuals at risk of vascular disease (PROSPER): A randomised controlled trial, *Lancet*, 360, 1623–1630.

5. Willich, S. N., Müller-Nordhorn, J., Sonntag, F., Völler, H., Meyer-Sabellek, W., Wegscheider, K., Windler, E., and Katus, H. (2004), Economic evaluation of a compliance-enhancing intervention in patients with hypercholesterolemia: Design and baseline results of the Open Label Primary Care Study: Rosuvastatin Based Compliance Initiatives to Achievements of LDL Goals (ORBITAL) study, *Am. Heart J.*, 148(6), 1060–1067.
6. Coenen, S., van Royen, P., Michiels, B., and Denekens, J. (2004), Optimizing antibiotic prescribing for acute cough in general practice: A cluster-randomized controlled trial, *J. Antimicrobial. Chemother.*, 54, 661–672.
7. Coenen, S., Michiels, B., van Royen, P., Van der Auwera, J.-C., and Denekens, J. (2002), Antibiotics for coughing in general practice: A questionnaire study to quantify and condense the reasons for prescribing, *BMC Family Practice*, 3, 16–25.
8. Little, P., Dorward, M., Warner, G., Stephens, K., Senior, J., and Moore, M. (2004), Importance of patient pressure and perceived pressure and perceived medical need for investigations, referral, and prescribing in primary care: nested observational study, *BMJ*, 328, 444–447.
9. Simes R. J., Marschner, I. C., Hunt, D., Colquhoun, D., Sullivan, D., Stewart, R. A. H., Hague, W., Keech, A., Thompson, P., White, H., Shaw, J., and Tonkin, A. (2002), Relationship between Lipid Levels and Clinical Outcomes in the Long-Term Intervention with Pravastatin in Ischemic Disease (LIPID) Trial, *Circulation*, 105, 1162–1169.
10. Shepherd, J., Hunninghake, D. B., Stein, E. A., Kastelein, J. J. P., Harris, S., Pears, J., and Hutchinson, H. G. (2004), Safety of rosuvastatin, *Am. J. Cardiol.*, 94, 882–888.
11. Suissa, S., Blais, L., Spitzer, W. O., Cusson, J., Lewis, M., and Heinemann, L. (1997), First-time use of newer oral contraceptives and the risk of venous thromboembolism, *Contraception*, 56, 141–146.
12. Juurlink, D. N., Mamdani, M. M., Lee, D. S., Kopp, A., Austin, P. C., Laupacis, A., and Redelmeier, D. A. (2004), Rates of hyperkalemia after publication of the randomized aldactone evaluation study, *N. Engl. J. Med.*, 351, 543–551.
13. Graham, D. J., Staffa, J. A., Shatin, D., Andrade, S. E., Schech, S. D., Grenade, L. L., Gurwitz, J. H., Chan, K. A., Goodman, M. J., and Platt, R. (2004), Incidence of hospitalized rhabdomyolysis in patients treated with lipid-lowering drugs, *JAMA*, 292, 2585–2590.
14. Cholesterol Treatment Trialists (CTT) Collaborators (2005), Efficacy and safety of cholesterol-lowering treatment: Prospective meta-analysis of data from 90,056 participants in 14 randomised trials of statins, *Lancet*, 366, 1267–1278.
15. Friedland, D. J., Go, A. S., Davoren, J. B., Shlipak, M. G., Bent, S. W., Subak, L. L., and Mendelson, T. (1998), *Evidence-Based Medicine—A Framework for Clinical Practice*, Prentice Hall, London.
16. Hirsch, M., O'Donnell, J. C., and Jones, P. (2005), Rosuvastatin is cost-effective in treating patients to low-density lipoprotein-cholesterol goals compared with atorvastatin, pravastatin and simvastatin: analysis of the STELLAR trial, *Eur. J. Cardiovasc. Prevention Rehab.*, 12(1), 18–28.
17. Okie S. (2005), Safety in numbers—Monitoring risk in approved drugs, *N. Engl. J. Med.*, 352, 1173–1176.

# *Phase IV and Postmarketing Clinical Trials*

**Ali Miraj Khan**

**Contents**

1 Introduction
2 Phase IV Clinical Studies
3 Postmarketing Surveillance
4 Adverse Drug Reactions
  4.1 Methods for Monitoring Adverse Drug Reactions
5 Pharmacovigilance and Postmarketing Surveillance
6 PMS: Linkage System and Data Bank
7 Postmarketing Surveillance: FDA Perspective
8 Drug Regulation and the FDA
  8.1 FDA: Globalization and Recent Activities in Drug Safety
9 Postmarketing Surveillance: Australia Perspective
10 Confirmation and Quantification in Denominator-Based Systems
11 United Kingdom Prescription Event-Monitoring (PEM) Program
12 Adverse Drug Reactions and Resource-Poor Countries
  12.1 Pharmacoeconomic Impact of Drug Toxicity in Developing Countries
  12.2 Basic Requirements for Effective Postmarketing Surveillance Program in Resource-Poor Country
13 WHO and Other Authorities for PMS
  13.1 Safety Surveillance of Antiretroviral Drug and RaPID Program
  13.2 Baseline Program in Hospitals of Resource-Poor Countries

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

13.3 Possible Approach to Develop Drug Safety Monitoring Program for Hospital in Resource-Poor Country
13.4 Inference for Future Buildup in Poor Countries
13.5 Data Mining in Postmarketing Surveillance
13.6 Spontaneous Reporting Databases
13.7 Prescription Event-Monitoring Databases
13.8 Linked Administrative Databases
13.9 Electronic Medical Records
13.10 Other Databases
13.11 Data Preprocessing
13.12 Application of Data-Mining Techniques in PMS
14 Discussion
15 Future Perspective
16 Conclusion
Bibliography

## 1 INTRODUCTION

Pharmaceutical products are developed and marketed to promote human health and protect people from various ailments. The identification or development of drugs involves a lot of time and resources, and to bring drugs to the marketing stage involves assessment of different aspects of the product in a very scientific and credible manner. Yet, the assessment of the products does not end just at marketing; assessment of drug effectiveness and, particularly, their safety continues, constituting phase IV clinical trials.

Typically, development of drugs involves several stages of experimentation in animals followed by trials in humans before large-scale phase IV clinical assessments. Animal experimentation precedes trials in humans and are intended to provide information on the acute toxicity, organ damage, dose dependence, absorption and bioavailability, the pharmacokinetics and dynamics, the metabolism, mutagenicity, teratogenicity, species specificity, and lethal dose. After animal experimentations providing encouragement for further development of the drug, clinical trails are planned that typically have three stages: phase I, phase II, and phase III conducted sequentially.

The phase I clinical studies involve a small number of healthy volunteers that allows investigators to understand the effects an investigational drug has in humans, with the goal to understand what happens to the investigational drug in the body from the time it is administered until its elimination from the body, on how human bodies react to the new drug from a safety and tolerability point of view. Study participants are monitored for the occurrence and severity of any side effects they may experience.

Phase II clinical studies are designed to evaluate the safety and efficacy of an investigational drug in patients with a specific disease or condition. Phase II clinical trials are typically conducted in a group of patients who are usually selected based upon their being at the same stage of a disease. The patients are given various doses of the drug and closely monitored to compare the effects and to determine the safest dosing regimen. In many instances, multiple phase II clinical studies are conducted to test the drug in a variety of patient populations with different indications.

Phase III clinical studies are designed to confirm the safety and efficacy of an investigational drug, and the dosage regimen chosen, in large numbers of patients with a specific disease or condition. These studies, as in the earlier phases, may involve one or more "treatment arms," which allow the investigational drug to be compared to other available treatments or to be tested for effectiveness in combination with other therapies. The safety and efficacy of the new drug is compared to that of the currently accepted standard treatment. Information obtained from phase III studies is used to determine how the drug is best prescribed to patients in the future. The complete information available on the new drug is submitted to health authorities. But, for safety reasons, these clinical trials of drugs are designed to exclude women of childbearing age, pregnant women, and/or women who become pregnant during the study. In some cases the male partners of these women are also excluded. Also children and very aged persons are frequently excluded from these clinical studies.

The purpose of the next phase, that is, phase IV clinical study, is to increase knowledge surrounding use of a new drug by increasing the number of humans that provides information on its effectiveness and more importantly adverse events when rolled out onto the market. "There are no therapeutic roses without thorns." Thus, the general use of a particular drug depends on availability of alternatives, the determination of the frequency and severity of adverse effects associated with the new drug, and finally on risk–benefit and cost–benefit assessments. Approval for marketing of a drug is usually granted based on adequate efficacy and safety, as demonstrated through well-designed phase I to phase III clinical trials.

It is, however, important to remember the limitations of these phases, which are conducted on a smaller number of humans and, thus, do not represent real-life situations. They generally represent only a limited population with specific characteristics and thus underrepresent the greater patient populations and provide short-term outcomes. A number of issues are monitored after marketing of drugs to determine their performance as they are prescribed in increasing numbers: Will the drug provide equal or greater benefit to people compared to other available drugs on the market? Will this be more cost effective? Will its safety profile be equal to or better than the alternatives? Will it provide better compliance?

## 2 PHASE IV CLINICAL STUDIES

Phase IV clinical trials involve the surveillance of the safety and adverse reactions of drugs in the market. With progression of medical services, the need for phase IV clinicals trials is definitely greater now. Phase IV clinical trials encompass the detection, assessment, understanding, and prevention of drug-related problems. Both regulatory authorities and the sponsoring company may require phase IV studies for competitive purpose to find a new market for the drug. The safety surveillance is designed to detect any rare or long-term adverse effects of drugs over a larger patient population and for a longer time period than was possible during the phase I–III clinical trials. Withdrawal of a drug from the market or restricted to certain uses may be recommended if harmful effects are discovered by phase IV clinical trials.

Pharmacoepidemiology is the primary scientific discipline now engaged in postmarketing drug safety surveillance. The most common epidemiologic study designs used in phase IV studies are randomized trials, cohort studies, and case–control studies. Regardless of their design, all studies involve human subjects and these studies need to be conducted by qualified investigators according to a written protocol duly reviewed and approved by an institutional review and ethics board. Promotional activities conducted under the guise of postmarketing studies are never acceptable. Changing regulatory environment, increasing concerns about the safety of new drugs, and varied uses for large-scale, real-world data on marketed drugs' safety and efficacy are primary drivers of the growth seen in the phase IV study environment today.

Postmarketing study is an important element of commercialization that enables companies to expand existing markets, enter new markets, develop and deliver messages that directly compare their products with the competition, and secure a better position in competitive markets.

## 3 POSTMARKETING SURVEILLANCE

Management of patients using drugs on a long-term as well as a short-term basis raise several complex issues such as toxicity, interaction with other drugs, as well as efficacy and effectiveness. In addition, age, sex, nutritional status, disease duration, comorbidity, ability to tolerate adverse events, and socioeconomic status are important considerations that influence the outcome of individual patients taking drugs. Randomized clinical trials alone are not sufficient to address all of these issues. Multiple therapies with attendant adverse effects are also common. Thus, postmarketing surveillance is important to accumulate information of adverse effects, influence of personal attributes mentioned above, identification of risk factors, quantitative measurement of safety and long-term safety/toxicity, study of potential risk groups (children/elderly, pregnant women, etc.), and identification of new indications and unexpected beneficial effects. It, additionally, provides useful information on the patterns of use and characteristics of drug users, inappropriate use of drugs, for example, addiction, impact on compliance, the possibility of medication errors and intoxication, quality of life during drug use, and utility and cost assessments.

## 4 ADVERSE DRUG REACTIONS

Drugs are intended to improve and make life comfortable for humans. But, they definitely are not without risk and may cause lesser or greater harm to many people. The unintended effects of drugs are harmful to the recipients and generally are referred to as adverse drug reactions (ADRs). Classical examples of a few ADRs are provided in Table 1.

Adverse drug reactions might account for over 10% of all hospital admissions in some countries. For example, a prospective study conducted in a European country reviewed the cause of hospital admissions of 18,820 patients over a 6-month period and noted 6.5% of admissions to be related to an ADR. The projected annual cost of such admissions to the National Health Service of that country was very high.

**TABLE 1 Adverse Drug Reactions**

| Drug | Adverse Drug Reaction |
|---|---|
| Phenyl butazone | Aplastic anemia |
| Chloramphenicol | Pancytopenia |
| Ethambutol | Optic atropy |
| Erythromycin estolate | Hepatitis |
| Aspirin | GI hemorrhage |
| Steroid | Hyperglycemia |
| Oral contraceptives | Thromboembolism |
| Clindamycin | Pseudomembranous colitis |
| Reserpine | Depression |
| Co-trimoxazole | Agranulocytosis |

According to another survey published in the *Journal of American Pharmacists' Association* in 2001, the cost of identified drug-related morbidities and deaths in a North American country was much higher than expected. Costs associated with drug-related problems have been estimated to be greater since 1995. From developing countries, such information is not readily available, but it is believed that the incidence of adverse events is greater with considerable cost.

### 4.1 Methods for Monitoring Adverse Drug Reactions

There are several methods that can be employed for evaluation and monitoring of adverse drug reactions. Selection of an appropriate method depends on factors such as program objectives and setup and the availability of human resources. A few common methods to evaluate ADRs include:

1. *Case Reports* The publication of single case reports, or a series of cases, on adverse drug events in the medical journals is an important way of compiling unreported and serious reactions.
2. *Anecdotal Reporting* Individual medical practitioners through anecdotal reports provide the majority of first reports of adverse drug reactions. They describe the occurrence of an event in association with use of a particular drug. Because of the anecdotal nature, such events need to be verified; however, confirmation may not be possible.
3. *Spontaneous Reporting* It is the principal continuing method for monitoring safety of marketed drugs. Along with the use of other methods, ADR programs are still based on spontaneous reporting system in many countries, both developing and developed. Clinicians are encouraged, and in some countries physicians are specially instructed, to report any or all reactions associated with medication. Usually, attention is focused on newly marketed drugs. Reporting of ADRs of a serious nature has also been made mandatory in many developed countries. This system is simple to operate and usually does not interfere with clinical practice to any great extent. Absence of an adequate control group is the main drawback of spontaneous reporting and, consequently, the difficulty of knowing the relative risk of treated patients. Despite limitations, voluntary spontaneous reporting systems remain the most

cost-effective approach for postmarket identification of new ADRs. Here the purpose is to register new problems or the occurrence of an increased incidence of previously reported problems. However, underreporting is one of the major limitations in the spontaneous reporting system. Despite regulations, health care professionals report only a variable proportion of adverse drug events with temporal association with a drug.

4. *Intensive Event Recording* In certain medical setups a group of people are given the responsibility to scrutinize a defined population specifically to detect adverse reactions associated with specific drugs. However, because of shorter periods of observation of individual patients and a relatively small number of people targeted in such studies, this method has not been very effective in the detection of drug toxicity.
5. *Cohort Studies (Prospective)* This is a very useful epidemiological method in which a cohort with a particular attribute (e.g., recipients of a drug) is followed prospectively and compared for some outcome (e.g., toxicity of the drug) with another cohort not possessing the attribute. Good control of data, accurate checking of outcome, good estimation of risk, study of many different outcomes, etc., are advantages of this method. Disadvantages are high cost and long wait.
6. *Case–Control Studies (Retrospective)* This is a good epidemiological study design. In this type of study the investigator selects the case group and the control group on the basis of outcome—i.e., having drug-related toxicity of interest versus not having drug-related toxicity of interest—and compares the groups in terms of their frequency of past exposure to possible risk factors. Case-control studies are relatively inexpensive, can be done quickly, and are a frequently used type of epidemiological study that can be carried out by a small team. Its retrospective nature limits the conclusions.
7. *Case–Cohort Studies* The case-cohort study is a recently developed useful modification of the case-control study. It is a hybrid epidemiological design. This design allows direct estimation of the risk ratio from a fixed cohort.
8. *Meta-Analysis* This is a quantitative analysis of two or more independent studies (already carried out) to determine an overall effect and to describe reasons for variation in study results. Meta-analysis aims to establish associations between drugs and adverse events, to estimate the frequency of ADRs, and to identify subgroups at increased risk for ADRs.
9. *Use of Population Statistics* Birth defect registers and cancer registers can be used if a drug-induced event is very frequent. If such assessment raises suspicion, then further observational studies are initiated.

## 5 PHARMACOVIGILANCE AND POSTMARKETING SURVEILLANCE

Pharmacovigilance is a relatively new and emerging science that is expected in the future to serve the purpose of postmarketing surveillance (PMS), but now it is not a well-established academic activity. Although a few universities offer specific courses in pharmacovigilance, the current curricula for training programs of profes-

sions in clinical medicine, clinical pharmacy, clinical pharmacology, and medical biology are not geared to provide all of the skills needed in this discipline. One of the reasons could be that pharmacovigilance involves a wide range of subjects, such as pharmacology, epidemiology, clinical medicine, data management, and drug legislation and communication, and they do not easily fit within the area of competence of any of the existing academic departments. It is anticipated that pharmacovigilance will be taught as an academic subject at the university level in the near future. Currently, the International Society of Pharmacovigilance (ISoP) provides specialized ad hoc training courses in pharmacovigilance. Each country might establish such an organization to provide a stronger base in effective and credible training services for relevant individuals/group of individuals.

## 6 PMS: LINKAGE SYSTEM AND DATA BANK

Many countries and related organizations involved in health care activities either already have or have the capacity to establish comprehensive databases using modern technologies at reasonable expenditure. Existence of such databases would immensely help establish a comprehensive and cost-efficient system for pharmacoepidemiology. With the multicenter nature of such databases, postmarketing surveillance programs have new possibilities. A wide spectrum of data is covered by automated databases of general practice and, therefore, at the population level it provides opportunities for a closer look at the impact of medications. However, special methods need to be established for determining the abuse potential of nonprescription drugs, such as many analgesics, in order to determine the extent of abuse and the temporal changes over periods since marketing a new drug. For nearly last two decades, automated medical record linkage systems have been utilized to evaluate postmarketing drug safety. Recently, for adequate assessment, networks of databases have been employed to identify problems that require large or diverse populations. Despite the potentials, establishment of such an effective system might not be easy. Because it involves merging of data from different systems, their reliability, variations in coding conventions for drug formularies, and the need to protect the confidentiality of medical records are likely to make the task difficult.

The European Active Surveillance Study (EURAS) is an ideal example of record linkage for postmarketing surveillance. This is a prospective, controlled, multinational, postmarketing, nonintervention cohort study of users of oral contraceptives. The study aims to detect the incidence of rare serious adverse side effects associated with the use of new and old oral contraceptives, particularly to determine incidence of thromboembolic events. This surveillance is under routine medical practice in European countries. The study is based on postal questionnaires and semiannual, active follow-up with validation of reported events by the physicians treating those women under surveillance. To ensure low loss to follow-up rates, a multifaceted, four-level follow-up procedure has been established.

Before the introduction of automated databases, studying large cohorts of drug users was extremely expensive and required major administrative and financial efforts to assemble and ensure complete and accurate collection of the required follow-up information. The computerized information system, initiated in the late

1970s, facilitated all these efforts and allowed highly efficient research. In the United Kingdom, the general practitioner (GP) traditionally acts as a gatekeeper to services within the National Health Service (NHS). A comprehensive record of written prescriptions, outpatient diagnoses, and referral letters to hospitals resides in the individual patient's record that GPs maintained in their offices. Since the early 1990s, the database has been known as the General Practice Research Database (GPRD), belonging to the UK Department of Health, and currently it is maintained by the Medicine Control Agency (MCA). Truly population-based data is collected by GRPD, which includes both outpatient and inpatient clinical information with a size that makes follow-up possible on large cohorts of users of specific drugs.

## 7 POSTMARKETING SURVEILLANCE: FDA PERSPECTIVE

Drug review process of the U.S. Food and Drug Administration (FDA) is recognized worldwide as a gold standard for the approval of drugs. To improve the standard, there have been significant additions during the last several decades in response to advances in medical science. Recently, drugs are approved by the FDA after comprehensive assessment, including safety evaluations. Major changes have taken place in the evaluation of new drugs, including complete understanding of their metabolism, interactions with other drugs, and potential differences of effectiveness or safety in people of different age groups, sex, and races. Focusing on the potential problems with the greatest care and importance, internal guidelines are in place to describe systematic approaches for assessment of safety through comprehensive review. Approval of drugs are granted by the FDA after a sponsor demonstrates risk–benefit assessment favoring use in specific population(s) and that the drug meets the standard for accepted safety and efficacy. It is conceivable that premarketing studies will never elucidate all the information about the effectiveness or all the risks from use of a new drug. The FDA recognizes that it is not possible to assess or anticipate all possible effects of a drug from the clinical trials that precede approval. So, the FDA has created a postmarket drug safety program designed to collect and assess adverse events following the drug's approval and marketing with an aim to detect unexpected serious adverse events and to take required action(s). The FDA collects information from adverse event reports filed by drug manufacturers, postmarketing clinical trials, and spontaneous reporting of adverse events by practicing physicians, pharmacists, and consumers.

## 8 DRUG REGULATION AND THE FDA

In the United States, both the prescription and over-the-counter drugs are essential, and their numbers are increasing. Newly marketed drugs are saving lives, reducing suffering, and improving the quality of life of millions, as do older drugs. Despite rising costs of health care, innovative newer medicines are necessary to improve life expectancy. Also, drugs are used to prevent or slow the progress of many diseases, thus avoiding costly invasive treatments, hospitalizations, and nursing home stays. However, there are undesirable consequences by widespread use of drugs. Inherent

side effects, misuse, overuse, underuse, medication mix-ups, and intentional abuse are increasing. So, appropriate monitoring and addressing of all these problem are required. Thus, the responsibilities for oversight of the entire life cycle of drugs has been undertaken by the FDA. It includes premarket testing and development through drug approval, postmarket surveillance, and risk management. The objective of the FDA is to ensure that safe and effective new drugs are available as quickly as possible and that drugs already on the market remain safe and maintain the highest quality. The FDA maintains its efforts to translate new scientific advances into benefits for patients and to take advantage of new developments in monitoring the performance of marketed drugs. The FDA and its Center for Drug Evaluation and Research (CDER) dedicate at least half of their premarket review activities to evaluating the safety of investigational drugs and safety of clinical trial participants. The FDA's Office of Regulatory Affairs (ORA) support the premarket drug review process by monitoring clinical trials and their results.

### 8.1 FDA: Globalization and Recent Activities in Drug Safety

At every stage of the drug life cycle, the FDA is working to improve the mechanism that supports drug safety and risk management. With the evolution of science over drug development, related activities have been expanded, and the focus is now on biomarkers and pharmacogenomics, which are likely to help maximize the safe and effective use of drugs in individual patients. More scientific methods and new tools are now in the works to target a specific drug for specific patients to maximize benefits and minimize associated risks. One of the current important objectives is to prevent adverse events by predicting drug safety problems before they can cause an insult. The aim is to combine new understandings of disease and its origins at the molecular level, including adverse events resulting from drugs, with the emerging knowledge of genetic and biologic characteristics of individual patients that determine how they will react to drug treatment. If successful, these innovative and risk-assessment-based drug development processes will improve FDA's efficiency for early disclosure of safety-related problems. Moreover, the FDA is now exploring, testing, and developing new methods of data mining, signal detection, and analysis of electronic data related to global health care.

The FDA is now very aware of the globalization trends. To respond to this global challenge in a proactive manner, the FDA has developed a variety of solutions. The FDA has entered into a number of cooperative relationships with foreign regulators to formalize arrangements for advancement of drug quality. The FDA is continuing work to encourage greater transparency and to ensure that patients and health care providers have access to emerging new safety information.

## 9 POSTMARKETING SURVEILLANCE: AUSTRALIA PERSPECTIVE

In Australia, guidelines have been applied in industry-sponsored postmarketing surveillance studies on drug safety, and most are observational studies that used cohort design. The guidelines, however, apply for other types of studies, including case–control studies and in intensified monitoring and various forms of recorded release. The term *postmarketing surveillance (PMS) study* need not mean a mere

passive follow-up, but actually it includes scientifically rigorous studies of products that are approved for registration and marketing, and its purpose is to collect reliable information on drug safety. However, they are not clinical trials of registered products or studies designed primarily for marketing purposes despite their scientific validity. PMS studies are generally initiated by the producers but might also be requests from other parties. Their aim is very specific—drug safety question or testing a hypothesis that often results from voluntary reporting. The guidelines for postmarketing surveillance studies have been developed to design and conduct scientifically rigorous and ethical studies to maximize validity of the results and increase potentials for acquiring useful newer information on the product. The ultimate purpose of a PMS study is to find an answer to a specific medical question that is new, significant, and unlikely to be answered by existing voluntary reporting systems.

In company-sponsored PMS studies that are typically observational rather than experimental, the decision to enroll a patient needs to be made after making an independent clinical decision to prescribe a specific drug concerned a particular patient. Prior to commencement, study proposals are sent to the Australian Drug Regulatory and Advisory Council with clear statements with regards to its aims and objectives, the question of clinical significance under investigation, proposed methodology including data collection and analyses procedures, and identification of company officially designated as responsible for the study. The notification of adverse reactions occurring during the PMS study should be done according to the pharmacovigilance guidelines.

The sponsor company is responsible to provide a timely report on the study outcomes. The product to be studied must have been approved for registration in Australia, and only patients having an approved indication for use of the product can be enrolled in PMS studies. Company-sponsored PMS studies should not be a disguised marketing or promotional exercise. Any payment offered to the medical professional must be commensurate with the work involved. The guidelines describe scientific approaches that the pharmaceutical companies are encouraged to follow in designing and conducting the postmarketing surveillance studies.

## 10 CONFIRMATION AND QUANTIFICATION IN DENOMINATOR-BASED SYSTEMS

A number of epidemiological approaches have been developed for postmarketing surveillance and quantification. Some of them have control groups to make relative risk estimates. Population-based claims databases are large databases that can link prescription-dispensing information (e.g., drug, dose, duration, and date) to claims data for medical interventions via unique patient identifiers. These systems can generate event rates (numerators) for the denominator of exposed persons and person-years of exposure and allow comparison of rates observed in unexposed patients. Here, the limitations of the systems are: ability to record the events of interest and the reliability. The availability of medical records in some of these systems allows validation of event reports and more in-depth analysis of adverse events, making formal nested case–control studies possible. However, in the event of very rare events these databases, despite being large, might still be relatively small for acquiring adequate numbers of exposures to the drugs of interest.

## 11 UNITED KINGDOM PRESCRIPTION EVENT-MONITORING (PEM) PROGRAM

In the UK, all prescriptions for newly marketed drugs are entered into an automated database after retrieval from the British Prescription Pricing Authority. Prescription data provided to the Drug Safety Research Unit (DSRU) include the name of the patient and the prescribing general practitioner (GP). Subsequently, each prescribing GP received the questionnaires ("green cards") 6 months after the first prescription for each of the patients. Questionnaires request adverse events during and after treatment, reason(s) for discontinuation, if any, patient age, indication for treatment, starting and stopping dates of treatment, and dosage. They are tailored for individual drugs and additional questions might also be incorporated.

Exposure-specific registry is another cohort technique employed that attempts to identify and follow patients who have been exposed to a drug but have not yet had an adverse outcome (e.g., hypertension treatment registry created before development of side effects). These registries may also be constructed on diseased populations (e.g., patients with gluten-induced enteropathy). Availability of denominator is a major advantage in these systems. Registries may provide also background rates for unexposed or differently exposed patients. Registries also might facilitate more complete follow-up of incident cases. These resources have also limitations common to other databases. Rates are provided for reported events without any control group. Limitations also include the imprecise nature of data linkage of drugs to events, and the temporal relationships between drug and events sometimes are not well established. In most claims data, unless an event leads to hospitalization, the event is not captured. Sometimes, very indirect methods are adopted, e.g., some may capture office visit events and often use of a drug is taken as an indicator of an event. In addition, claims interpreted as representing adverse drug reaction might not actually reflect the actual clinical event or an outcome of interest (e.g., not all claims of "convulsions" in such databases are actually episodes of epilepsy, and similarly, syncope, vasovagal attacks, etc., are often found events). These are limitations that prevent right quantification of the incidence of drug-related adverse reactions.

## 12 ADVERSE DRUG REACTIONS AND RESOURCE-POOR COUNTRIES

The type A reactions include normal and augmented response to drugs and are dose related; reactions are usually predictable and preventable on the basis of known pharmacology of drugs. The incidence of type A reactions is greater and usually associated with higher morbidity. It may be managed to overcome or prevent such reactions by reducing the dose, e.g., bradycardia associated with β-adrenoreceptor blockers and bleeding in association with anticoagulants.

The type B reactions are usually unpredictable and are often caused by immunological and pharmacogenetic mechanisms. The type B reactions are not dose-dependent, are relatively rare, often not predictable and preventable, and sometimes cause serious morbidities and death. Examples of this kind of reactions include malignant hyperthermia caused by anesthetics and many immunological reactions. It is difficult to classify other adverse drug reactions.

In resource-poor countries, frequent use of complementary medicines might interact with conventional allopathic drugs. Thus it may predispose to adverse drug

reactions, factors that need to be considered in their interpretation. A majority of drugs used in resource-poor countries are usually imported from other countries. The ingredients used in drugs may vary and the genetic makeup of the population might also vary, thus may predispose to adverse drug reactions. These factors are not usually verified among populations of resource-poor countries before approval of these drugs. Therefore, the risk of occurrence of drug-related toxicities in such populations could be very high and, in fact, remains unknown.

Due to geographic variations, poor socioeconomic condition, high cost of modern medicines, and nonavailability of physicians in most of the rural areas, the people of developing countries face greater difficulties in accessing standard health care. Drug retail shops frequently serve as the first point of contact of population for seeking health care, thus predisposing the poor people to be more vulnerable to self-medication, a potential contributor to adverse effects of drugs. Moreover, polypharmacy is a very common practice in resource-poor countries, which increases the incidence of adverse drug reactions. It has been demonstrated that drug toxicity sharply rises with the number of drugs consumed.

Age is another very important factor for development of adverse drug reaction—the very old and the very young are more susceptible to adverse effects of drugs due to compromised physiological reserve. The aging population is more likely to have cardiovascular, pulmonary, musculo-skeletal, and metabolic disorders that necessitate several medications; and all these predispose them to increased incidence and increased severity of drug-related toxicities. Patients with impaired kidney and liver functions and those with intercurrent illness are at greater risk for adverse drug reactions. In poor countries these are not meticulously considered by the doctors while prescribing drugs because of limitations of laboratory tests. Race genetic polymorphism and hereditary factors are known to affect the bioavailability of drugs and thus are common important contributing factors for adverse drug reactions in poor regions. Also, in poor countries, the effect of drugs may vary from place to place, depending on climatic diversity or different cultural and social backgrounds.

## 12.1 Pharmacoeconomic Impact of Drug Toxicity in Developing Countries

The economic impact of adverse drug events is less explored and less reported in developing countries. Adverse drug events prolong hospital stay, result in new admissions, and increase treatment cost leading to increased burden on the health care systems in resource-poor countries. It has been observed that the total cost of drug-related morbidity and mortality is exceedingly high, which exceeds the cost for treatment of the primary illness. Therefore, ADR-related admissions in developing countries with limited health care budget stretched that even further.

## 12.2 Basic Requirements for Effective Postmarketing Surveillance Program in Resource-Poor Country

The minimal requirement for the establishment of a working program for postmarketing surveillance of drugs requires infrastructure, funding, expertise, and a trained staff. For resource-poor countries, there may be shortcomings in all of these areas.

However, some useful activities can be established with limited financial resources if sincere efforts are undertaken. It is possible to initiate a program for concurrent surveillance during drug therapy, based on reporting of suspected adverse drug reactions by pharmacists, physicians, nurses, and patients at an affordable cost. Alternatively, a prospective surveillance before drug therapy for patients with a high risk for adverse drug reactions may be initiated.

Doctors or drug prescribers should be adequately notified regarding the drug to be monitored for adverse reactions. Information regarding suspected adverse drug reactions should be reported to the pharmacy for complete collection of data and their analyses, including patients' name, their medical and medication histories, a description of suspected ADR, temporal relationship of the event with drug administration, any remedial treatment required, and sequel and outcome. The cause(s) of each suspected adverse drug reaction need to be evaluated based on the patient's medical and medication histories, the circumstances of the adverse event, the result of de-challenge and re-challenge (if any), consideration of alternative etiologies, and a literature review. Description of each of the suspected adverse drug reactions and their outcomes needs documentation in patients' medical records.

Serious or unexpected adverse drug reactions should be reported to the drug regulatory/health authority and the drug's manufacturer. All adverse drug reaction reports should be reviewed and evaluated by the authorized pharmacy and therapeutics committee. Adverse drug reaction report information should be disseminated to health care professionals for educational purposes. In all of these processes patient confidentiality must be preserved. An organization's ongoing quality assurance activities should incorporate findings from adverse drug reaction monitoring and reporting programs.

## 13 WHO AND OTHER AUTHORITIES FOR PMS

The WHO established its program as a pilot project in 1968 initially in 10 and later in many other countries. The WHO promotes pharmacovigilance at the country level working through the WHO Collaborating Centre for International Drug Monitoring (Uppsala Monitoring Centre). This pharmacovigilance program has following objectives: greater patient care and drug safety, special emphasis on prevention of undesired drug-related toxicities, better public health safety in relation to the use of drugs, provision of reliable information for more rational use of drugs, improved assessment of risk–benefit profile to encourage safer and more effective use of medicines, the resolution of apparent conflicting interests of public health, and ensuring the welfare of individual patients.

Today worldwide many other regulatory authorities are involved in monitoring the safety of drugs. Other authorities include the United Kingdom's Medicine Control Agency (MCA) and the Committee on Safety of Medicine (CSM), the Vaccines Adverse Event Reporting System (VAERS), co-administered by the Department of Health and Human Services (DHHS), the FDA, and the Centers for Disease Control and Prevention (CDC) of the U.S., the Adverse Drug Reactions Advisory Committee (ADRAC) of Australia, and MedWatch Program of the USFDA-Canadian Adverse Drug Reaction Monitoring Program.

### 13.1 Safety Surveillance of Antiretroviral Drug and RaPID Program

In some resource-poor countries, many children have started their lives with antiretroviral therapy, but these countries have no effective and functioning pharmacovigilance system to monitor appropriate use of these drugs and collection of drug safety data. Long-time, adequate financial efforts and expertise are required to establish new pharmacovigilance programs or to strengthen the existing ones. WHO and the Uppsala Monitoring Centre (UMC) have been conducting training seminars for more than 15 years in those countries. Despite these efforts, UMC estimates that excepting a few, these countries have limited capacity; they collect little adverse drug reaction information and it is not sufficient for analysis of formulation of policies to increase patient safety. Drugs exhibiting toxicity of concern need to be removed from use or withdrawn from the market. And those are to be replaced by other suitable drugs.

The most significant benefit of the RaPID program is that it can establish a country- and disease-specific drug safety program very quickly within a reasonable period and concurrently help build institutional capacity at the country level. The RaPID program can help national drug authorities in capturing ADR data and their inclusion into UMC/WHO's Vigiflow software, and analyze them by a team of RaPID program global experts with technical expertise in HIV/AIDS, TB, and malaria and drug-related toxicities. It is important to mention that all activities of the RaPID program are coordinated and jointly developed with the national drug authority. The RaPID program is being implemented in joint technical collaboration with leading pharmacovigilance institutions from the developed and the developing countries including the Uppsala Monitoring Centre in Sweden, the national regulatory authority of Switzerland (Swissmedic), and other organizations, which provide activity-specific expertise, for example, WHO (various departments), Ecumenical Pharmaceutical Network, and the Indian Institute of Health Management Research (a WHO collaborating center). The RaPID program team has significant expertise in designing and implementing pharmacovigilance programs. With WHO (Geneva) support, the RaPID program is currently conducting an evaluation in several countries of Africa to assess their capacity for collection of pharmacovigilance data and their analyses to formulate appropriate policy related to drug use. Currently, the RaPID program team is in discussion with several countries (Nigeria, Sri Lanka, Uganda, India) and nongovernmental organizations (NGOs) (Medicines Sans Frontier, PSI, etc.) to explore collaborations and develop strategies for conducting pharmacovigilance.

### 13.2 Baseline Program in Hospitals of Resource-Poor Countries

Recently, a few hospitals in developing countries have taken initiatives and initiated spontaneous reporting programs at the facility level. ADR reporting forms are placed in different wards and out-patient departments. The doctors, nurses, and pharmacists report the adverse drug reaction, if any, to the pharmacovigillance cell, newly developed units of the hospitals. Based on the report, the cell provides relevant information to the drug regulatory authority of the country. Since its inception, the program looks very promising. Plans are under way to upscale existing spontaneous reporting programs to fully developed pharmacovigilance programs of these hospitals in near the future—its ultimate goal.

## 13.3 Possible Approach to Develop Drug Safety Monitoring Program for Hospital in Resource-Poor Country

Through awareness programs all health care providers need to be encouraged to report all suspected adverse drug reactions using a reporting form, which should be readily available all the time in all patient care areas of the health facilities including wards, out-patient departments (OPDs), and pharmacy departments. After filling out the forms, the health care provider should send them to the drug safety monitoring unit of the hospital. Data collected from these forms can be utilized to evaluate the reactions in detail. Several adverse drug reaction reporting forms are available, such as "yellow card" and "blue card" systems, and one developed by MedWatch. Generic forms can be modified to meet the local needs. The adverse drug reactions and related information of the particular drug(s) should be provided to the concerned authority. The determination of the likelihood of a causal relation between a suspected drug and an adverse drug reaction is an essential step in identifying drug toxicity. Based on the causality assessment, an adverse drug reaction can be classified as one of the following categories: certain, probable, possible, unlikely, conditional/unclassified, and not assessable/not classifiable. There are several causality assessment scales such as the WHO causality assessment scale, Naranjos scale, Karch and Lasagnas scale, Schumock and Thornto scale, European ABO system, and the Bayesian neutral scale. Any one these scales can be used to conduct causality assessments. The preventability and predictability of the adverse drug reactions can be assessed by using the specified scales. Thus, incidence of particular adverse drug reaction can be detected against the total number of drugs used in the hospital during the particular period. All data captured by the program should be entered into its database, which will be data for that particular population group(s). The relevant drug reactions data and their analyses should be periodically disseminated through adverse drug reaction bulletins, journals, and newsletters.

## 13.4 Inference for Future Buildup in Poor Countries

In resource-poor countries, it should be the responsibility of all health care workers to develop strategies for reporting, monitoring, and preventing adverse drug reactions. The drug regulatory authorities in poor countries should take the necessary steps to promote safe use of drugs in their country. Initially, the program might begin at a single hospital for consolidation and training and later might be extended to other hospitals as the program matures. Generally, people also need to be educated on the correct use of drugs and on reporting of undesirable experiences in association with drug use. The medical curriculum should include and emphasize rational use of drugs and aspects of adverse drug reactions. A formal training program for the primary health care professionals on adverse drug reactions should also be established. Postmarketing surveillance studies must be made mandatory after marketing any new drug in the country. A well-developed adverse drug reactions monitoring system can contribute a lot to the health care system if the formed teams/programs are supported by highly dedicated and well-experienced staff and good data management systems. Appropriate addressing of these issues would help establish an effective program of pharmacovigilance in resource-poor situations.

## 13.5 Data Mining in Postmarketing Surveillance

An increasing interest has developed in data mining or knowledge discovery initiatives using robust databases with development of large electronic health data storage systems in recent years. Web mining and information science can simultaneously make their application in postmarketing surveillance of drugs. Data mining processes are often performed on existing databases, for the purposes of postmarketing surveillance considering high cost for collection of data. It is difficult to determine the requirement of the necessary size of the dataset. It depends on quality of data, the background frequency of the event(s), and the strength of the association of the event with the drug; however, for even moderately rare events large databases are required.

## 13.6 Spontaneous Reporting Databases

The spontaneous reporting of adverse drug reactions by health care providers to health authorities or drug companies is an important procedure in postmarketing drug surveillance. However, spontaneous reporting systems are vulnerable to inconsistent reporting. The accuracy of the data may remain uncertain. Yet, spontaneous reporting databases contain large amounts of data. These databases could, therefore, be mined to obtain details of ADRs.

## 13.7 Prescription Event-Monitoring Databases

Prescription event monitoring (PEM) system detects adverse drug reactions by mobilizing data of high quality from family physicians on a select group of patients exposed to a specific (new) drug, for a limited period of time. Lack of an adequate control group is an important limitation of PEM database mining.

## 13.8 Linked Administrative Databases

Large linked health administrative databases, such as Medicaid in the United States, and the Ontario provincial databases contain data on millions of subjects. These might also be used as a source for data mining. The Saskatchewan-linked administrative health care utilization database and the Tayside Medicines Monitoring (MEMO) are examples of linked medical health databases. Both of them have been utilized to identify risks of benzodiazepine therapy.

## 13.9 Electronic Medical Records

Electronic medical records (EMRs) contain large numbers of data: e.g., use of tobacco products, use of nonprescription drugs, symptoms and signs of diseases, laboratory data, and social circumstances on a smaller numbers of patients. All these might be used for data mining. The availability of large numbers of variables and their combined use makes possible the generation of hypotheses and the exploration of new diagnoses or adverse events of drugs. EMR has been used by some to investigate detection of known adverse drug reactions.

## 13.10 Other Databases

Specialist and clinical trials databases and overdose or toxicology databases might provide valuable information. Signals for liver-related ADRs have been reported from a biochemistry laboratory database at a higher rate than that reported by physicians. Poison information centers also record details of ADRs and thus might be a good contributor of pharmacovigilance activity.

## 13.11 Data Preprocessing

Data preprocessing involves sampling and verification of data quality to ensure that data are clean. Medical data might contain erroneous data, for example, putting an age of 120 years instead of 20 years and documentation of hysterectomy in males. Data can also be used in creating new variables of interest for analysis, for example, it might be possible to estimate socioeconomic status from the postal code.

## 13.12 Application of Data-Mining Techniques in PMS

Data mining involves a number of statistical techniques, including cluster analysis, link analysis, deviation detection, and disproportionality assessment, which can be used to determine the presence and assessment of the strength of adverse drug reaction signals. It is also a systematic process whereby large databases are searched. Combinations of variables that occur at a higher frequency than expected are mined. Research has been conducted specifically on the application of data mining techniques in postmarketed drug-related toxicity surveillance. Development of this tool is promising to optimize limited resources and improve the level of consistency in safety reviews of approved marketed drugs. An improvement in the current understanding of adverse drug reaction patterns is expected through the use of data-mining techniques. A successful data-mining tool will allow for earlier detection of safety signals and drug–drug interactions with drugs already in the market. Recently, there has been potential use of data mining and knowledge discovery in databases for detection of adverse drug events in pharmacovigilance. Knowledge discovery in databases (KDD) is a technical process that may be employed to explore potential adverse drug reactions more effectively. KDD involves selection of data variables and databases, data preprocessing, data mining, and data interpretation and utilization. Recently, the spontaneous reporting database of WHO Uppsala Monitoring and the UK yellow card scheme use data-mining methods. These data-mining methods used in pharmacovigilance are those of disproportionality, such as the proportional reporting ratio and information component, which has been used to analyze the drug-related adverse events. Examples are: association of captopril and other angiotensin-converting enzymes with cough, the association of pericarditis with practolol but not with other β-blockers, and the association of terfenadine with heart rate and rhythm disorders; these all could be identified by mining the database. In view of the importance of adverse drug reaction and the development of massive data storage systems and availability of powerful computers, the use of data-mining techniques in knowledge discovery in medical databases is now of increasing importance in the process of pharmacovigilance. They are likely to be able to detect signals earlier.

## 14 DISCUSSION

Postmarketing surveillance programs as well as clinical phase IV studies both are definitely essential to ensure the safety of drugs. These studies disclose beneficial and detrimental effects of drugs, including the existence of causal effects, risk factors, economic consequences, and characterization of drug use in clinical practice. Drug-related diseases impart a considerable burden to health care systems globally, which has been demonstrated well in the Western countries. The situations in the low-income countries are not very clear. There, published reports on the incidence and cost of drug-related problems are few. It is a reasonable assumption that less rigorous control of drug quality, easy availability, and indiscriminate usage create problems associated with drug therapy. A majority of drug-related problems are preventable as demonstrated by studies conducted in developed countries.

Appropriate instructions should be available on the effective use of drugs, and people should be instructed to adhere to them. Then prevention or reduction of drug-related toxicity would be possible for thousands of patients in developing countries. There, suffering could be alleviated as well as many lives could be saved, while millions of dollars would be saved annually by the society. The absence of awareness on the magnitude of drug-related problems in many societies is common. Adverse effects of drugs often manifest gradually and the symptomatology often mimic those of common illness, and usually they manifest at home and thus remain largely undetected and unrecorded. The awareness of drug-related problems is inadequate among politicians, the general public, patients, and sometimes even in professionals to trigger and stimulate a wide discussion on its prevention. So, qualified people are required in the health care system to support industry with genuine interest in disclosing these problems to everyone. An elaborative system of rules and regulations is necessary to govern the drug surveillance process. In postmarketing drug surveillance, pharmaceutical manufacturers, drug regulatory agencies, and the end users (physicians and patients) all need to play essential roles to minimize the risks.

## 15 FUTURE PERSPECTIVE

The future of postmarketing surveillance will need the proper development of research methods that would be capable to cope with the specific problems posed by the newly marketed drugs and would be able to accurately suggest adjustments for reducing severity of drug-related adverse events. Phase IV clinical studies and postmarketing surveillance, including pharmacovigilance, is now a very growing science that would offer great scope for minimizing harm to patients taking drugs. It will also reduce costs of health care systems. From small beginnings, with the appropriate technological knowledge and skills, postmarketing surveillance and phase IV studies can make important contributions to the health of people in any country. A successful postmarketing drug surveillance program needs capacity building, for example, training of professionals, optimization of computerized databases for research purposes, their joint use with traditional epidemiological methods, and so forth. The impact of new vaccines and new drugs on well-defined populations

needs assessment, which will be a critical task and from public health perspectives its impact is important. The pharmaceutical industry and its business environment should change immensely in the future. Growing competition from generic drugs, competitive product market, and limitation of resource and development costs are future problems for pharmaceutical executives, and they should plan a balanced program in this regard. The objective should be to divert resources for promotion of only those drugs that will be unlikely to be rejected by the regulatory agencies or shall not be removed from the market due to safety problems. A pharmaceutical company will need to use appropriate technology to ensure production of a safe and effective drug to ensure a favorable risk–benefit ratio. Thus, to implement a proactive postmarketing surveillance and risk management strategy, drug companies should always adapt to changing regulations to ensure compliance for eventual approval of their drug products. The pharmaceutical industries must adhere to international regulations in addition to local rules and should be regulated by a global enterprise.

Therefore, drug companies should report and communicate drug-related safety issues regularly to every country where they market their products. The adoption of advanced pharmacovigilance tools will greatly help drug companies to generate and submit reports to specific countries or regions based on their regulations. Here, specifically, signal detection and data management tools, management technology solutions, and adverse-event reporting systems will be specially helpful. Effective pharmacovigilance is needed for future successful drug safety surveillance for which radical changes are required in signal detection, use of data mining and management process, and use of these technologies in phase IV clinical trials. Thus, completion of analysis of clinical data would be quick as compared to the past. Time saving shall reduce in development costs of drug and would allow early arrival of new drugs on the market. Simultaneously, termination of future clinical trials would be possible on early detection of safety issues and, therefore, patients would be protected. Thus company resources would be saved that might be spent on modernization of their industry-based technology and development of efficient manpower.

The European Union's postmarketing surveillance system has undergone major changes over the last few years through a stronger drug regulatory and pharmacovigilance system. The EU PMS systems have the power to mandate phase IV studies and impose financial penalties on companies that fail to comply with regulations and their obligations. In recent years, the FDA has adopted many changes and many are currently under review process to have a more proactive approach to pharmacovigilance in the near future. There has been a noticeable shift forwarded by regulatory authorities, in terms of formulation of policies on risk management, and also by pharmaceutical companies in terms of incorporation of pharmacogenomics in their development programs. The purpose of these initiatives are to better predict safety issues and minimize risks as quickly as possible.

In the future, companies need to ensure a close link between premarketing and postmarketing pharmacovigilance activities to increase transparency of their activities and improve communication of information about drugs to begin risk management planning at the inception. Many companies have now agreed to voluntarily register clinical trials and will post all clinical trial data of marketed drugs, whether favorable or unfavorable, on publicly accessible websites.

## 16 CONCLUSION

Despite its four-decade history, postmarketing surveillance remains a dynamic clinical and scientific discipline. It continues to play a crucial role in meeting the challenges posed by the ever increasing range of drugs, all of which carry an unpredictable potential for harm. It is essential that previously unknown adverse effects are reported, analyzed, and their significance communicated effectively to the people, and it should be ensured that they have the knowledge to interpret that information. For all drugs, there is a trade-off between the benefits and the potential for harm. The harm can be minimized by ensuring that drugs are of good quality, safety, and efficacy and are used rationally. At the same time, expectations and concerns of the patient should be taken into consideration when therapeutic decisions are taken. The practitioners should ensure that risks in drug use are anticipated and will be managed if they occur. This will create a sense of trust among patients taking drugs. So, improvement of communication between the health professionals and the public is essential. Also, updating health professionals to understand the efficacy/risk of medicines that they prescribe is necessary for this purpose. As adverse drug reactions are a significant public health problem in every health care system, postmarketing drug surveillance systems need to be improved continuously. Simultaneously, it is a grand challenge with a fundamental problem of great social and economic importance, and the solution needs significant increases in scientific knowledge and technical capabilities.

## BIBLIOGRAPHY

Amery, W. K. (1999), Signal generation from spontaneous adverse event reports, *Pharmacoepidemiol. Drug Saf.*, 8, 147–150.

Anon. (2002), MedDRA compared with WHO-ART, *Drug Saf.*, 25, 445–452.

Anon. (2001), Details of methods: Explanation of data mining methods, *BMJ*, 322, 1207–1209.

Anon. (1975), *Requirements for Adverse Reaction Reporting*, World Health Organization, Geneva, Switzerland.

ASHP (1989), ASHP guidelines on adverse drug reaction monitoring and reporting, *AMJ Hosp. Pharm.*, 46, 336–337.

Bagheri, H., Michel, F., Lapeyre-Mestre, M., Lagier, E., Cambus, J. P., Valdiguié, P., and Montastruc, J. L. (2000), Detection and incidence of drug-induced liver injuries in hospital: A prospective analysis from laboratory signals, *Br. J. Clin. Pharmacol.*, 50, 479–484.

Barbone, F., McMahon, A. D., Davey, P. G., Morris, A. D., Reid, I. C., McDevitt, D. G., and MacDonald, T. M. (1998), Association of road-traffic accidents with benzodiazepine use, *Lancet*, 352, 1331–1336.

Bate, A., Lindquist, M., Edwards, I. R., and Orre, R. (2002), A data mining approach for signal detection and analysis, *Drug Saf.*, 25, 393–397.

Bate, A., Lindquist, M., Orre, R., Edwards, I. R., and Meyboom, R. H. (2002), Data-mining analyses of pharmacovigilance signals in relation to relevant comparison drugs, *Eur. J. Clin. Pharmacol.*, 58, 483–490.

Bates, D. W., Cullen, D. J., Laird, N., Petersen, L. A., Small, S. D., Servi, D., Laffel, G., Sweitzer, B. J., Shea, B. F., and Hallisey, R. (1995), Incidence of adverse drug events and potential adverse drug events. Implications for prevention, *JAMA*, 274, 29–34.

Bates, D. W., Leape, L. L., Cullen, D. J., Laird, N., Petersen, L. A., Teich, J. M., Burdick, E., Hickey, M., Kleefield, S., Shea, B., Vander Vliet, M., and Seger, D. L. (1998), Effect of computerized physician order entry and a team intervention on prevention of serious medication errors, *JAMA*, 280, 1311–1316.

Beard, K. (1992), Adverse reactions as a cause of hospital admission in the aged, *Drugs Aging*, 2, 356–367.

Biriell, C., and Edwards, R. (1997), Reasons for reporting adverse drug reactions—some thoughts based on an international review, *Pharmacoepidemiol. Drug Saf.*, 6, 21–26.

Bombardier, C., Laine, L., Reicin, A., Shapiro, D., Burgos-Vargas, R., Davis, B., Day, R., Ferraz, M. B., Hawkey, C. J., Hochberg, M. C., Kvien, T. K., and Schnitzer, T. J. (2000), Comparison of upper gastrointestinal toxicity of rofecoxib and naproxen in patients with rheumatoid arthritis. VIGOR Study Group, *N. Engl. J. Med.*, 343, 1520–1528.

Brewer, T., and Colditz, G. A. (1999), Postmarketing surveillance and adverse drug reactions: current perspectives and future needs, *JAMA*, 281, 824–829.

Brown, E. G., Wood, L., and Wood, S. (1999), The medical dictionary for regulatory activities (MedDRA), *Drug Saf.*, 20, 109–117.

Bytzer, P., and Hallas, J. (2000), Drug-induced symptoms of functional dyspepsia and nausea. A symmetry analysis of one million prescriptions, *Aliment Pharmacol. Ther.*, 14, 1479–1484.

Cerrito, P. (2001), Application of data mining for examining polypharmacy and adverse effects in cardiology patients, *Cardiovasc. Toxicol.*, 1, 177–179.

Chakrabarti, S. (2002), *Mining the Web: Discovering Knowledge from Hypertext Data*, Morgan Kaufmann, San Francisco.

Chowdhury, G. G. (1999), Template mining for information extraction from digital documents, *Library Trends*, 48, 181–207.

Ciorciaro, C., Hartmann, K., and Kuhn, M. (1998), Differences in the relative incidence of adverse drug reactions in relation to age. An evaluation of the spontaneous reporting system of SANZ, *J. Suisse Med.*, 128, 254–258.

Classen, D. C., Pestotnik, S. L., Evans, R. S., Lloyd, J. F., and Burke, J. P. (1997), Adverse drug events in hospitalized patients. Excess length of stay, extra costs, and attributable mortality, *JAMA*, 277, 301–306.

Colin-Jones, D. G., Langman, M. J., Lawson, D. H., and Vessey, M. P. (1985), Postmarketing surveillance of the safety of cimetidine: Mortality during second, third, and fourth years of follow up, *BMJ (Clin. Res. Ed.)*, 291(6502), 1084–1088.

Crombie, I. K., Brown, S. V., and Hamley, J. G. (1984), Postmarketing drug surveillance by record linkage in Tayside. *J. Epidemiol. Community. Health*, 38(3), 226–231.

Dunn, N., Freemantle, S., and Mann, R. (1999), Nicorandil and diabetes: A nested case–control study to examine a signal generated by prescription-event monitoring, *Eur. J. Clin. Pharmacol.*, 55, 159–162.

Edwards, I. R. (2000), The accelerating need for pharmacovigilance, *J. Roy. Coll. Physicians Lond.*, 34, 48–51.

Edwards, I. R., and Biriell, C. (1994), Harmonisation in pharmacovigilance, *Drug Saf.*, 10, 93–102.

Egberts, A. C., Meyboom, R. H., and van Puijenbroek, E. P. (2002), Use of measures of disproportionality in pharmacovigilance: Three Dutch examples, *Drug Saf.*, 25, 453–458.

Eisenhauer, L. A. (2002), Adverse drug reactions: A concern for clinicians and patients, *Clin. Excellence Nurse Practitioners*, 6, 3–7.

Evans, S. J., Waller, P. C., and Davis, S. (2001), Use of proportional reporting ratios (PRRs) for signal generation from spontaneous adverse drug reaction reports, *Pharmacoepidemiol. Drug Saf.*, 10, 483–486.

Faich, G. A. (1986), Adverse-drug-reaction monitoring, *N. Engl. J. Med.*, 314, 1589–1592.

Fattinger, K., Roos, M., Vergees, P., Holenctein, C., Kind, B., Masche, U., Stocker, D. N., Braunschweig, S., Kullak-Ublick, G. A., Galeazzi, R. L., Follath, F., Gasser, T., and Meier, P. J. (2000), Epidemology of drug exposure and adverse drug reactions in two Swiss departments of internal medicine, *Br. J. Clin. Pharmacol.*, 49, 158–167.

Fisher, S., Bryant, S. G., Solovitz, B. L., and Kluge, R. M. (1987), Patient-initiated postmarketing surveillance: A validation study, *J. Clin. Pharmacol.*, 27, 843–854.

Friedman, M. A., Woodcock, J., Lumpkin, M. M., Shuren, J. E., Hass, A. E., and Thompson, L. J. (1999), The safety of newly approved medicines: Do recent market removals mean there is a problem? *JAMA*, 281, 1728–1734.

Heeley, E., Wilton, L. V., and Shakir, S. A. (2002), Automated signal generation in prescription-event monitoring, *Drug Saf.*, 25, 423–432.

Honig, P. K., Woosley, R. L., Zamani, K., Conner, D. P., and Cantilena, L. R., Jr. (1992), Changes in the pharmacokinetics and electrocardiographic pharmacodynamics of terfenadine with concomitant administration of erythromycin, *Clin. Pharmacol. Ther.*, 52, 231–238.

Honigman, B., Lee, J., Rothschild, J., Light, P., Pulling, R. M., Yu, T., and Bates, D. W. (2001), Using computerized data to identify adverse drug events in outpatients, *J. Am. Med. Inform. Assoc.*, 8, 254–266.

Inman, W. H. (1981), Post marketing surveillance of adverse drug reactions in general practice. II: Prescription-event monitoring at the University of Southampton, *BMJ*, 282, 1216–1217.

Ioannidis, J. P., and Lau, J. (2001), Completeness of safety reporting in randomized trials: An evaluation of 7 medical areas, *JAMA*, 285, 437–443.

Jick, H., Madsen, S., Nudelman, P. M., Perera, D. R., and Stergachis, A. (1984), Postmarketing follow-up at Group Health Cooperative of Puget Sound, *Pharmacother.*, 4, 99–100.

Johnson, J. A., and Bootman, J. L. (1996), Drug-related morbidity and mortality, *Arch. Intern. Med.*, 155, 1949–1956.

Juntti-Patinen, L., and Neuvonen, P. J. (2002), Drug-related deaths in a university central hospital, *Eur. J. Clin. Pharmacol.*, 58, 479–482.

Karch, F. E., and Lasagna, L. (1975), Adverse drug reaction—a critical review, *JAMA*, 234, 1236–1241.

Kaufman, D. W., Rosenberg, L., and Mitchell, A. A. (2001), Signal generation and clarification: Use of case–control data, *Pharmacoepidemiol. Drug Saf.*, 10, 197–203.

Kwansnik, B. H. (1999), The role of classification in knowledge representation and discovery, *Library Trends*, 48, 22–47.

Lazarou, J., Pomeranz, B. H., and Corey, P. N. (1998), Incidence of adverse drug reactions in hospitalized patients—a meta-analysis of prospective studies, *JAMA*, 279, 1200.

Leape, L. L., Cullen, D. J., Clapp, M. D., Burdick, E., Demonaco, H. J., Erickson, J. I., and Bates, D. W. (1999), Pharmacist participation on physician rounds and adverse drug events in the intensive care unit, *JAMA*, 282, 267–270.

Lindquist, M., Stahl, M., Bate, A., Edwards, I. R., and Meyboom, R. H. (2000), A retrospective evaluation of a data mining approach to aid finding new adverse drug reaction signals in the WHO international database, *Drug Saf.*, 23, 533–542.

Mann, R. D. (1998), Prescription-event monitoring—recent progress and future horizons, *Br. J. Clin. Pharmacol.*, 46, 195–201.

Mey, C., Hentschel, H., Hippius, M., and Balogh, A. (2002), Documentation and evaluation of adverse drug reactions (ADR)—contribution from a poison information center, *Int. J. Clin. Pharmacol. Ther.*, 40, 102–107.

Meyboom, R. H., Lindquist, M., Egberts, A. C., and Edwards, I. R. (2002), Signal selection and follow-up in pharmacovigilance, *Drug Saf.*, 25, 459–465.

Mukherjee, D., Nissen, S. E., and Topol, E. J. (2001), Risk of cardiovascular events associated with selective COX-2 inhibitors, *JAMA*, 286, 954–959.

Murphy, B. M., and Frigo, L. C. (1993), Development implementation and results of a successful multidisciplinary adverse drug reaction reporting program in a university teaching hospital, *Hosp. Pharm.*, 28, 1199–1240.

Niu, M. T., Erwin, D. E., and Braun, M. M. (2001), Data mining in the US Vaccine Adverse Event Reporting System (VAERS): Early detection of intussusception and other events after rotavirus vaccination, *Vaccine*, 19, 4627–4634.

Noan, B. A., and Brushwood, D. B. (2000), Adverse drug reactions in elderly patients: alternative approaches to post-surveillance, *J. Health Law*, 33, 383–454.

Olsson, S. (2001), The need for pharmacovigilance, in Gupta, S. K., Ed., *Pharmacology and Therapeutics in the New Millennium*, Narosa, New Delhi, pp. 502–508.

Patel, P., and Zed, P. J. (2002), Drug-related visits to the emergency department: How big is the problem? *Pharmacotherapy*, 22, 915–923.

Perner, P., and Petrou, M. (1999), *Machine Learning and Data Mining in Pattern Recognition*, Berlin: Springer, Berlin.

Pirmohamed, M., Breckenridge, A. M., Kitteringham, N. R., and Park, B. K. (1998), Adverse drug reactions, *BMJ*, 316, 1295–1298.

Ramesh, M., Pandit, J., and Parthasarathi, G. (2003), Adverse drug reactions in a south Indian hospital—their severity and cost involved, *Pharmacoepidemiol. Drug Saf.*, 12, 687–692.

Rawlins, M. D., and Thompson, J. W. (1977), Pathogenesis of adverse drug reactions, in Davies, D. M., Ed., *Textbook of Adverse Drug Reactions*, Oxford University Press, Oxford.

Rawson, N. S., and Rawson, M. J. (1999), Acute adverse event signalling scheme using the Saskatchewan Administrative health care utilization datafiles: Results for two benzodiazepines, *Can. J. Clin. Pharmacol.*, 6, 159–166.

Shakir, S. A., and Layton, D. (2002), Causal association in pharmacovigilance and pharmacoepidemiology: Thoughts on the application of the Austin Bradford-Hill criteria, *Drug Saf.*, 25, 467–471.

Smith, D. L. (1993), The effect of patient non-compliance on healthcare costs, *Med. Interface*, 6, 74–76, 78, 84.

Strom, B. L. (2000), How should one perform pharmacoepidemiology studies? Choosing among the available alternatives, in Strom, B. L., Ed., *Pharmacoepidemiology*, Vol. 3, Chichester: Wiley, Chichester, England, pp. 401–413.

Strom, B. L., Carson, J. L., Morse, M. L., and LeRoy, A. A. (1985), The computerized on-line Medicaid pharmaceutical analysis and surveillance system: A new resource for postmarketing drug surveillance, *Clin. Pharmacol. Ther.*, 38, 359–364.

Szarfman, A., Machado, S. G., and O'Neill, R. T. (2002), Use of screening algorithms and computer systems to efficiently signal higher-than-expected combinations of drugs and events in the US FDA's spontaneous reports database, *Drug Saf.*, 25, 381–392.

van Puijenbroek, E. P., Bate, A., Leufkens, H. G., Lindquist, M., Orre, R., and Egberts, A. C. (2002), A comparison of measures of disproportionality for signal detection in spontaneous reporting systems for adverse drug reactions, *Pharmacoepidemiol. Drug Saf.*, 11, 3–10.

van Puijenbroek, E. P., Egberts, A. C., Meyboom, R. H., and Leufkens, H. G. (1999), Signalling possible drug–drug interactions in a spontaneous reporting system: Delay of withdrawal bleeding during concomitant use of oral contraceptives and itraconazole, *Br. J. Clin. Pharmacol.*, 47, 689–693.

van Puijenbroek, E. P., van Grootheest, K., Diemont, W. L., Leufkens, H. G., and Egberts, A. C. (2001), Determinants of signal selection in a spontaneous reporting system for adverse drug reactions, *Br. J. Clin. Pharmacol.*, 52, 579–586.

# *Regulatory Approval*

**Fred Henry[1] and Weichung J. Shih[2]**

*[1]Drug Development and Regulatory Affairs, Taisho Pharmaceuticals R&D Inc., Morristown, New Jersey*

*[2]Department of Biostatistics, School of Public Health, University of Medicine and Dentistry of New Jersey, Piscataway, New Jersey*

**Contents**

1 FDA Approval
- 1.1 Introduction
- 1.2 Regulations and Guidance Documents
- 1.3 Consultation with the FDA
- 1.4 Investigational New Drug Applications
- 1.5 Clinical Hold: Issues and Resolution
- 1.6 IND Sponsor Reporting Requirements
- 1.7 Marketing Approval Requirements
- 1.8 Summary

2 Fast-Track Approval
- 2.1 Introduction
- 2.2 Fast Track Is More Than "Priority Review"
- 2.3 Request for Designation of Drug as Fast-Track Product
- 2.4 Limitations
- 2.5 Accelerated Approval Issues (I): Clinical and Surrogate Endpoints
- 2.6 Accelerated Approval Issues (II): Postapproval Study
- 2.7 Accelerated Approval Issues (III): Control of Type I Error Rate
- 2.8 Examples of FTDDP

References

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

# 1 FDA APPROVAL

## 1.1 Introduction

The Food and Drug Administration (FDA) is part of the Department of Health and Human Services (DHHS) in the U.S. federal government and is the regulatory agency responsible for marketing approval of new pharmaceutical products. Within the FDA are the Center for Drug Evaluation and Research (CDER), whose activities are the focus of this chapter, along with the Center for Biologics Evaluation and Research (which regulates vaccines and blood products), Center for Devices and Radiological Health, Center for Food Safety and Applied Nutrition, and the Center for Veterinary Medicine. The FDA has authority to write regulations relevant to interstate commerce associated with pharmaceuticals (as well as food, cosmetics, and veterinary medicines), which may then be enacted into law. Regulations may be enforceable by the FDA's compliance office. Guidelines created by the FDA, on the other hand, are meant to provide general direction to drug sponsors on many issues, without legal obligation. However, deviation from guidelines should be discussed with the agency in order to maintain a cooperative and common understanding of a drug's development.

It is important to keep in mind the FDA's concerns during drug development in order to anticipate and address them proactively during all phases of clinical development. The FDA's primary concern is always the protection of the health and rights of subjects involved in human drug research. In addition, during phases II and III the agency becomes concerned about the adequacy and quality of the scientific evaluation of the drug's safety and efficacy as part of a new drug application (NDA). This chapter will summarize the activities necessary to achieve and maintain regulatory clearance for the conduct of clinical trials and also describes some general strategies for the cooperative design of clinical programs to support marketing approval.

## 1.2 Regulations and Guidance Documents

Clinical trial conduct is governed by national laws that are supplemented by international, regional, and national guidance documents. Regulations regarding the conduct of clinical trials and the protection of human subjects involved in clinical trials in the United States, as well as many other regulations regarding food and drugs, are described in Title 21 of the Code of Federal Regulations (21 CFR) [1]. Clinical trials in the United States must conform to the regulations for investigational new drug (IND) applications, which are covered in 21 CFR Part 312 ("Investigational New Drug Applications"). Regulations regarding the content and format of NDAs are described in 21 CFR Part 314 ("Applications for FDA Approval to Market a New Drug"). Application of these regulations to the design and conduct of clinical trials will be addressed in greater detail throughout this chapter. Other important 21 CFR regulations for the clinical development of pharmaceuticals include Part 50, which covers the topic of "Protection of Human Subjects," and Part 56, "Institutional Review Boards."

In addition, the FDA has developed many guidelines for drug sponsors that provide direction regarding the application of the 21 CFR regulations, as well as

more specific information on the expectations that the FDA has for appropriate clinical study design and conduct. A key concept in the regulatory requirements for marketing approval is "substantial evidence" of efficacy from clinical trials, based on the Kefauver–Harris Amendments to the 1962 Food, Drug and Cosmetic Act, which was interpreted to mean that approval for the marketing of a new drug in the United States would be based on evidence from adequate and well-controlled investigations. This requirement would in turn imply that clinical trials had to be of sufficient size to have adequate statistical power to test important clinical hypotheses and had to meet the standard of being adequate and well controlled. To be adequate and well controlled a study has to have:

- A clear statement of objectives
- A design that permits a valid comparison with a control and provides a quantitative assessment of the drug effect
- A method of selection of subjects that ensures that they have the disease under study
- A method of patient assignment that minimizes bias
- A well-defined and reliable method of assessment of responses (outcome measures)

Several regulatory guidance documents from the FDA consider the reporting, evaluation, and interpretation of clinical trial evidence. These guidance documents were issued over a period of nearly 20 years. The first document is the Guideline for the Format and Content of the Clinical and Statistical Section of New Drug Applications [2]. This document addresses how the results of clinical studies are to be reported to the FDA. The second is the Guidance for Industry Providing Clinical Evidence of Effectiveness for Human Drug and Biological Products [3]. The agency participates in the International Conference on Harmonisation (ICH), which is a group of representatives from government and pharmaceutical industry groups in Japan, Europe, and the United States, formed to facilitate the development of mutually agreed upon guidelines for the chemical, nonclinical, and clinical development of new drugs. The guidance documents developed in ICH are then considered for adoption by national authorities in the three represented areas. The ICH guidances are powerful in that they support the design of clinical studies that are accepted internationally, increasing the efficiency of global drug development. For instance, the format issues of the NDA content and format guidelines mentioned above have been largely superseded by the. ICH M4: Common Technical Document for the Registration of Pharmaceuticals for Human Use [4], which provides a structure for marketing applications that can be submitted in most all countries.

Of the ICH guidances relevant to the conduct of trials and adopted by the FDA, those most applicable to the conduct of clinical trials and support for drug registration are:

E1B The Extent of Population Exposure to Assess Clinical Safety: For drugs intended for long–term treatment of non-life-threatening conditions

| | |
|---|---|
| E2A | Clinical Safety Data Management: Definitions and standards for expedited reporting |
| E3 | Structure and Content of Clinical Study Reports |
| E4 | Dose Response Information to Support Drug Registration |
| E6 | Good Clinical Practice: Consolidated Guideline |
| E7 | Studies in Support of Special Populations Geriatrics |
| E8 | General Considerations for Clinical Trials |
| E9 | Guideline on Statistical Principles for Clinical Trials |
| E10 | Choice of Control Group and Related Issues in Clinical Trials |
| M3 | Nonclinical Safety Studies for the Conduct of Human Clinical Trials for Pharmaceuticals |

Throughout this chapter, references will be made to several specific FDA and ICH guidances; however, they are all conveniently included on the FDA's website at http://www.fda.gov/cder/guidance/. Special attention should be paid to the ICH E9, as it also pertains to the aspect of the "substantial evidence" requirement. The FDA's current thinking on the matter of substantial evidence, meshing the views from these guidelines, can be found in a recent paper [5].

Once an organization has determined that it is interested in conducting a clinical trial in the United States, it is appropriate to begin plans for design of the protocol and its submission with supporting documentation to the FDA. The first clinical trial of a new investigational agent will require compilation and submission of an IND [6]. This submission includes information on the chemistry and manufacturing, nonclinical pharmacology, pharmacokinetics, and toxicology, as well as a summary of any previous human experience with the product. Thus, it is prudent to organize an internal team of individuals and external consultants as necessary who will contribute to the summarization of available information on the product and can address issues that may arise with regard to the IND submission. The formation of such a team is essential to establish a communication framework for discussing development issues, including regulatory submissions, meetings with authorities, responses to FDA questions, and so forth. Drug sponsors often do not adequately address cross-functional development issues, such as the implications of findings in animal studies to the clinical setting. A team appropriate to the size and structure of each organization can help to avoid neglect of such issues before they become regulatory roadblocks.

During the preparation of an IND application, questions may arise as to the contents of the submission, protocol design, or other development issues that would benefit from the FDA's input. In order to assist sponsors to best develop new products, the FDA has implemented a guidance on arranging meetings, Formal Meetings with Sponsors and Applicants of PDUFA Products [7], which is described in the next section. The first such opportunity in a product's development for a meeting with the FDA is referred to as the pre-IND meeting.

### 1.3 Consultation with the FDA

Formal meetings between drug sponsors and the FDA regarding the development of new drug products are possible in order to promote mutual understanding of issues relating to matters that arise during a research program (21 CFR 312.47).

Depending on the stage of development, these issues may involve chemistry/manufacturing, nonclinical or clinical research pertaining to the design of a development program, interpretation of study results and/or implications of research on other members of a therapeutic class, or special issues regarding a novel compound. The FDA has identified three basic types of meetings according to the timing and intention of the meeting.

- *Type A Meetings* Immediately necessary to discuss clinical holds, protocol assessments, or other development issues that are considered to be critical to the continued development of a product. Such meetings are normally scheduled within 30 days of a meeting request. An information package that includes the questions and background information for the agency to review, prior to the meeting, and that provide information necessary for their review must be submitted at least 2 weeks prior to the meeting (contents of the information package provided to the agency prior to formal meetings is described in the meetings guidance).
- *Type B Meetings* These meetings include discussions prior to submission of an initial IND (pre-IND meeting), certain meetings after completion of phase I, end of phase II, and meetings to discuss the content and format of new drug applications (pre-NDA meeting). Type B meetings are scheduled within 60 days of the agency's receipt of the meeting request and require submission of a briefing document at least 4 weeks prior to the meeting date.
- *Type C Meetings* Other meetings regarding the development and review of drug products. Type C meetings are scheduled within 75 days of meeting request receipt and briefing documentation is required at least 2 weeks prior to the meeting.

In addition to these sponsor meetings that the FDA allows, special protocol review by the FDA is also possible. A written request is necessary for each meeting desired by a drug sponsor. This request must include a brief summary of the product, type of meeting requested, an agenda and list of objectives for the meeting, along with administrative information regarding the scheduling of the meeting as listed in the meetings guidance. The reviewing division normally responds to meeting requests within 14 days of receipt of the request, at which time a mutually agreeable time and date are set for the meeting.

***Conducting and Documenting Formal FDA Meetings*** After an introduction of the meeting participants and a brief discussion of the background and objectives of the meeting, the agenda of the meeting is typically driven by the list of questions and/or issues that the sponsor has submitted as part of the meeting request and briefing documentation. Agency representatives provide responses to each question or issues raised by the sponsor, and discussion takes place if necessary to clarify the agency's position. Many divisions at the FDA are now providing to sponsors prior to meetings a copy of the agency's internal meeting minutes where they discuss their responses before meeting face to face with the sponsor. Though consideration of the planned duration of the meeting is prudent, sponsors should take advantage of the opportunity of these meetings to ensure understanding of the agency's position and resolve any misunderstandings during the meeting discussion. A summary of

the proceedings normally forms the conclusion of the meeting, and the FDA prepares a written summary of the meeting as the official minutes, which is shared with the sponsor (usually within 30 days of the meeting). Any disputes regarding the minutes of these meetings may be addressed as per 21 CFR 10.75 (Administrative Practices and Procedures), 312.48, and 314.103.

## 1.4 Investigational New Drug Applications

Key to the successful clinical development of a new product is completion and submission of an IND. The FDA's Guidance for Industry on the Content and Format of Investigational New Drug Applications (IND) for Phase I Studies of Drugs, Including Well-Characterized, Therapeutic, Biotechnology-derived Products [6], along with 21 CFR 312.22 and 312.23, provides a clear framework for the content and format of such submissions. While the required components of the IND are standardized by the Guidance and 21 CFR, the FDA has allowed some flexibility in the presentation of data in the application. Also, the agency has begun to accept electronic submissions of INDs, reducing the burdens of archiving and navigating paper-based submissions.

Whatever the therapeutic class, status of development, chemical qualities, or other product characteristics, sponsors of an IND will increase the likelihood of a positive outcome of the FDA's review if they view the application from the perspective of the agency staff who will be considering the documentation. Mindfulness of the agency's concerns for the safety and rights of subjects, and in later stages of development the quality of research design regarding establishment of a product's safety and efficacy, should be maintained throughout the preparation of the application. A logical format will also aid in the review and avoids confusion among the team that will be assembled to consider the application at the agency. Following the FDA's guidance on the content and format is a good start in this direction, but cross-functional critiquing the application by the sponsor throughout the compilation process will help to ensure that the technical information is presented effectively.

***Content and Format*** As per the regulations in 21 CFR 312.23, an IND must include the following components to be considered complete for a new drug product being proposed for clinical study.

*Cover Sheet (Form FDA-1571)* Following a cover letter in standard business style addressed to the Center for Drug Evaluation and Research, a completed Form FDA-1571 (available on the FDA website) is required. This form provides a brief identification of the contents of the submission, along with administrative information regarding the product sponsor's responsible personnel and their contact information.

*Table of Contents* A detailed description of the components of the application and their location within the submission.

*Introductory Statement and General Investigational Plan* This brief section of the IND provides an opportunity to introduce the agency staff to the product, including

intentions for its development, potential indications for use, any regulatory history (within and outside the United States), and any exceptional aspects of the IND that might be important for the review team.

*Investigator's Brochure* The investigator's brochure should follow the ICH guidance on the standards for the content and format of this document.

*Protocol(s)* Phase I study protocols need not be of the same level of detail as those for phase II or III trials; however, they should minimally include an outline of the investigation including number of subjects, description of safety exclusions, doses and duration of planned treatments, and safety monitoring. Since the objectives of phase I trials are normally to assess the safety, tolerability, pharmacokinetics, and perhaps some pharmacodynamic effects of the new drug, plans for assessing each of these aspects of the product's activities should be described in a rational manner.

Phases II and III protocols are required to provide a detailed description of all aspects of the trial. In addition to the assessment of safety, efficacy measurements need to be described thoroughly, remembering that the concerns of the FDA expand to include assessment of the trial's scientific merit toward eventual submission of study results in support of a marketing application.

*Chemistry, Manufacturing, and Controls (CMC) Information* Detailed information regarding the active ingredient (the drug substance) and the drug product (the formulation, such as tablet, solution, etc.) are required as described in 21 CFR 312.23(a)(7) and the Guidance on IND Content and Format for each and all of the products being investigated in the study (including placebo and active comparators). For the drug substance, this information must include:

- A description of the drug substance, including its physical, chemical, and/or biological characteristics
- The name and address of its manufacturer
- The general method of preparation of the drug substance
- The acceptable limits and analytical methods used to assure the identity, strength, quality, and purity of the drug substance
- Information to support the stability of the drug substance during the toxicologic studies and the proposed clinical study(ies)

With regards to the drug product information, similarly detailed information is required, including:

- A list of all components, which may include reasonable alternatives for inactive compounds, used in the manufacture of the investigational drug product, including both those components intended to appear in the drug product and those that may not appear but that are used in the manufacturing process
- Where applicable, the quantitative composition of the investigational new drug product, including any reasonable variations that may be expected during the investigational stage

- The name and address of the drug product manufacturer
- A brief, general description of the method of manufacturing and packaging procedures as appropriate for the product
- The acceptable limits and analytical methods used to assure the identity, strength, quality, and purity of the drug product
- Information to support the stability of the drug substance during the toxicologic studies and the proposed clinical study(ies)
- A brief general description of the composition, manufacture, and control of any placebo to be used in the proposed clinical trial(s)
- A copy of all labels and labeling to be provided to each investigator
- A claim for categorical exclusion from or submission of an environmental assessment

The information presented in the CMC section of the IND should be prepared by or at least with the close involvement of professionals within the organization who have expertise in this area of development and experience in preparing CMC documentation for regulatory submission. An FDA guidance specific to requests for information from the FDA relevant to CMC issues at the pre-IND stage is available to assist with questions that may arise during the preparation of this section of the IND.

*Pharmacology and Toxicology Information [21 CFR 312.23(a)(8)]* A summary of the pharmacology (effect of the drug on animals) and information known about the absorption, distribution, metabolism, and excretion (pharmacokinetics) of the drug should be provided. Individual reports may be provided; however, a summary is often sufficient. The sponsor should consider the stage of development (and the FDA concerns at that stage of development) and what if any implications the results of these studies may have on the assessment of safety.

Because of the safety concerns at the FDA in reviewing IND documentation and regulating clinical trial conduct, much more detailed information is required from the results of toxicology studies compared to other nonclinical studies. Short- and long-term trials in animals need to be presented in a logical manner, paying particular attention to the implications that the results and methods used in the study have for the design of the clinical trial and the characteristics of the subjects involved in the study. ICH has issued several guidances regarding the design and reporting of toxicology studies to support clinical trials and marketing applications that the sponsor should consider. In addition, the ICH M3 guidance Nonclinical Safety Studies for the Conduct of Human Clinical Trials for Pharmaceuticals [8] must be considered to assure that the duration of treatment in the proposed clinical study is supported by the duration and types of toxicology studies reported in the IND. Detailed listings of data from each toxicology study used to support the conduct of the clinical study must be presented in the IND.

*Previous Human Experience* A summary of any and all previous investigational and if available marketed use of the product should be provided. Individual study reports are not normally required.

***IND Review Process*** Once an IND is submitted to the FDA, it is routed to the appropriate reviewing division for assembly of a review team of representatives from each of the functional areas (CMC, nonclinical pharmacology and toxicology, clinical) and initiation of the review process. The review clock starts at the date of receipt of the application at the FDA, and if no comments are received by the sponsor within 30 days, the application is considered to be cleared. No formal approval letter is typically provided by the FDA; however, if issues are identified that delay or halt the clinical trial, the agency will issue a clinical hold to the sponsor (see following section).

***Information Amendments and New Protocols (21 CFR 312.30)*** New protocols and amendments to previously submitted protocols must be submitted to the FDA prior to enactment at the clinical site, in addition to obtaining necessary institutional review board approval. There is no need to await the FDA review; however, the agency may provide comments to a sponsor, which may impact study conduct.

The IND sponsors are also responsible for submitting new technical information (toxicology study reports, chemistry or manufacturing changes, etc.) to the FDA. These "information amendments" should be presented in a format that facilitates the agency's review of the submitted information, whether or not comment from the agency is requested by the sponsor.

## 1.5 Clinical Hold: Issues and Resolution

At any time during the review of an initial IND submission or during the conduct of a clinical program, the FDA may order a drug sponsor to delay or suspend clinical study conduct (21 CFR 312.42). Such "clinical holds" may be limited in nature to a specific protocol or stage of development ("partial clinical hold") or may impact all clinical investigation of a product ("complete clinical hold"). There are many reasons that the agency may take such an action, including:

- Risk to the health of subjects
- Unqualified investigators
- Investigator brochure deficiencies
- Insufficient information in IND to assess subject risks
- Phase II or III study plan or individual study deficient in design

Unless there is immediate and serious risk to patients, the FDA will attempt to discuss and resolve any clinical hold issues prior to imposing the clinical hold. If necessary, the clinical hold is communicated to the sponsor and immediate action is required as described in the clinical hold letter.

Sponsors must submit an official Response to Clinical Hold in order to address the concerns of the agency. Within 30 days, the FDA will provide a response that either removes the hold or describes any outstanding deficiencies. Resumption of the clinical study on hold may not occur until official notification by FDA of its removal.

### 1.6 IND Sponsor Reporting Requirements

***Safety*** Any serious and unexpected drug-related adverse event that occurs during clinical research of a product must be submitted by the drug sponsor (or designee) to the FDA, as well as all investigators participating in trials of the product. Submission of such events must occur within 15 calendar days of the sponsor's awareness of the event. If the event is fatal or life-threatening, the report to the FDA must be submitted within 7 days. Serious adverse events are any untoward medical occurrence that at any dose:

1. Results in death.
2. Is life-threatening.
3. Requires inpatient hospitalization or prolongation of hospitalization.
4. Results in persistent or significant disability/incapacity.
5. Is a congenital anomaly/birth defect.
6. Any other significant medical event that requires intervention to prevent the occurrence of one of the above.

A complete explanation of sponsor responsibilities for submission of serious adverse event reports is found in 21 CRF 312.32.

***Annual Reports*** Each year, within 60 days of the anniversary date of the IND becoming effective, an annual report summarizing accumulated safety and other updates to the IND file must be submitted to the FDA. This IND annual report includes:

- A brief summary of the status of each study in progress or completed during the previous year
- A summary of frequent and serious adverse events
- A summary of all safety reports submitted in the previous year
- A list of subjects who died or dropped out of clinical studies due to adverse events
- A summary of new understandings that have been achieved regarding the drug's activity
- A list of nonclinical studies completed or in progress during the previous year and summary of significant findings
- A summary of chemistry and manufacturing changes
- An investigation plan for the upcoming year
- An updated investigator brochure (if changed in the previous year)
- Any unreported phase I protocol amendments
- Any significant foreign marketing developments (such as approvals or withdrawals)

Due to the need to integrate information from several functional groups (chemistry/manufacturing, nonclinical, clinical) into IND annual reports, a central point

within the sponsor's organization is typically identified to coordinate the compilation of this document. This is typically an activity of the regulatory affairs department of a drug sponsor; however, it may take place within a document management center or be prepared by an external vendor, depending on the size and organization of the drug sponsor. IND annual reports are submitted administratively to the IND file at the FDA. Once a product is marketed, NDA annual reports are also required; however, the IND report must continue to be submitted each year after product commercialization.

## 1.7 Marketing Approval Requirements

***Pivotal Trials*** As discussed earlier in this chapter, discussion and cooperation between a drug sponsor and the FDA is strongly encouraged by the agency to maximize the adequacy of the NDA submission to provide substantial evidence in support of the marketing application, including "adequate and well-controlled" studies that provide the information necessary for the agency to assess the safety and efficacy of the product. These "pivotal" trials will have the statistical power to reach a conclusion regarding the predetermined efficacy endpoint(s) and also contribute significantly to the safety database to allow a well-informed benefit–risk assessment by the agency. Additional studies that examine the drug's effects in special populations (elderly, hepatically impaired, etc.), are also required to elucidate the product's activity and to assist in the evaluation of a drug's safety in a less controlled environment during marketed use.

***NDA Format and Content*** The overall format and content of a new drug application as described in the ICH M4 guidance includes the following administrative, chemistry/manufacturing, nonclinical, and clinical information.

**Module 1: Administrative Information and Prescribing Information**
- 1.1 Table of Contents of the Submission Including Module 1
- 1.2 Documents Specific to Each Region (e.g., application forms, prescribing information)

**Module 2: Common Technical Document Summaries**
- 2.1 CTD Table of Contents
- 2.2 CTD Introduction
- 2.3 Quality Overall Summary
- 2.4 Nonclinical Overview
- 2.5 Clinical Overview
- 2.6 Nonclinical Written and Tabulated Summary
  - Pharmacology
  - Pharmacokinetics
  - Toxicology
- 2.7 Clinical Summary
  - Biopharmaceutics and Associated Analytical Methods
  - Clinical Pharmacology Studies
  - Clinical Efficacy
  - Clinical Safety
  - Synopses of Individual Studies

**Module 3: Quality**
- 3.1 Module 3 Table of Contents
- 3.2 Body of Data
- 3.3 Literature References

**Module 4: Nonclinical Study Reports**
- 4.1 Module 4 Table of Contents
- 4.2 Study Reports
- 4.3 Literature References

**Module 5: Clinical Study Reports**
- 5.1 Module 5 Table of Contents
- 5.2 Tabular Listing of All Clinical Studies
- 5.3 Clinical Study Reports
- 5.4 Literature References

The important clinical sections of the application are modules 2.5 (Clinical Overview) and 2.7 (Clinical Summary). An outline of these sections follows:

2.5 Clinical Overview
- 2.5.1 Product Development Rationale
- 2.5.2 Overview of Biopharmaceutics
- 2.5.3 Overview of Clinical Pharmacology
- 2.5.4 Overview of Efficacy
- 2.5.5 Overview of Safety
- 2.5.6 Benefits and Risks Conclusions
- 2.5.7 References

2.7 Clinical Summary
- 2.7.1 Summary of Biopharmaceutic Studies and Associated Analytical Methods
- 2.7.1.1 Background and Overview
- 2.7.1.2 Summary of Results of Individual Studies
- 2.7.1.3 Comparison and Analyses of Results across Studies
- 2.7.2 Summary of Clinical Pharmacology Studies
- 2.7.2.1 Background and Overview
- 2.7.2.2 Summary of Results of Individual Studies
- 2.7.2.3 Comparison and Analyses of Results across Studies
- 2.7.2.4 Special Studies
- 2.7.3 Summary of Clinical Efficacy
- 2.7.3.1 Background and Overview
- 2.7.3.2 Summary of Results of Individual Studies
- 2.7.3.3 Comparison and Analyses of Results across Studies
- 2.7.3.4 Analysis of Clinical Information Relevant to Dosing Recommendations
- 2.7.3.5 Persistence of Efficacy and/or Tolerance Effects
- 2.7.4 Summary of Clinical Safety
- 2.7.4.1 Exposure to the Drug
- 2.7.4.2 Adverse Events

2.7.4.3 Clinical Laboratory Evaluations
2.7.4.4 Vital Signs, Physical Findings, and Other Observations Related to Safety
2.7.4.5 Safety in Special Groups and Situations
2.7.4.6 Postmarketing Data
2.7.5 References
2.7.6 Synopses of Individual Studies

The contents of each of these sections is described in detail within the M4 guidance; however, sponsors should be mindful again that this is a guidance that should be adapted to each drug product in cooperation with the agency during pre-NDA discussions as per the FDA/sponsor meeting guidance.

***FDA Review of Marketing Applications*** Once the NDA is submitted to the FDA, an initial review is conducted to consider if it contains sufficient information to complete their review. The agency then informs the sponsor of the application's acceptance and the review begins. Questions are usually raised by the agency and discussed with the sponsor during the review process, and labeling negotiations take place after the technical review is complete. Prior to completion of the FDA review, public discussion may take place in which one of the FDA advisory committees is assembled to consider specific questions for which the agency requests external advice. Advisory committees have no regulatory power but rather serve as consultants to the agency, which provides the final, formal approval action. Current standard timelines for completion of marketing application review will lead to an FDA action—approval, nonapproval, or approvable (upon completion of specific additional requirements)—within approximately 10 months.

## 1.8 Summary

In this first section of our chapter on regulatory approval, we have identified the basic requirements that need to be addressed in order to initiate clinical trials in the United States and the obligations of drug trial sponsors during the ongoing conduct of a clinical research program. The objective of investigational clinical studies is to achieve marketing approval of a new product and support the commercial positioning of the drug. Toward this end, we have provided a roadmap for discussions with the FDA and the preparation of essential drug applications. Though these concepts and tactics are common to most drug development projects, a customized strategy must be prepared for each new drug, which reflects its intended use and intrinsic characteristics. The FDA has developed guidance and regulations to assist sponsors in addressing most common and several special circumstances, including orphan drugs (which address diseases with small patient populations), "accelerated" or "priority" review, and accelerated development of products for life-threatening diseases not addressed with available therapies. Each of these categories of new drug products requires thorough discussion with the FDA during development in order to agree upon registration requirements. Accelerated development is considered in greater detail in the following section.

## 2 FAST-TRACK APPROVAL

### 2.1 Introduction

The Food and Drug Administration Modernization Act of 1997 has an intriguing section (No. 112) entitled Expediting Study and Approval of Fast Track Drugs (the act) [9]. In 1998, the FDA issued a Guidance for Industry: The Fast Track Drug Development Programs (the FTDD programs) [10] to meet the requirement of the act. The purpose of FTDD programs is to "facilitate the development and expedite the review of new drugs that are intended to treat serious or life-threatening conditions and that demonstrate the potential to address unmet medical needs" [9]. The act states that an application for approval of a fast track product may be approved if it is determined that "the product has an effect on a clinical endpoint or on a surrogate endpoint that is reasonably likely to predict clinical benefit" [9].

### 2.2 Fast Track Is More Than "Priority Review"

The FDA's Center for Drug Evaluation and Research (CDER) and the Center for Biologics Evaluation and Research (CBER) have long-standing policies that describe criteria for review priority classification of marketing applications. Products regulated by CBER are eligible for priority review if they provide a significant improvement in the safety or effectiveness of the treatment, diagnosis, or prevention of a serious or life-threatening disease. Products regulated by CDER are eligible for priority review if they provide a significant improvement compared to marketed products in the treatment, diagnosis, or prevention of a disease; eligibility is not limited to drugs for a serious or life-threatening disease. A fast track product would ordinarily meet either the center's criteria for priority review. Note, however, that for an NDA or BLA (biologic license application) manufacturers need not seek fast-track designation to be eligible for priority review.

### 2.3 Request for Designation of Drug as Fast-Track Product

Since *fast-track* programs are designed to facilitate the development and expedite the review of new drugs or biologics that are intended to treat serious or life-threatening conditions and that demonstrate the potential to address unmet medical needs, they emphasize the critical nature of close, early communication between the FDA and sponsors. Procedures such as pre-IND and end-of-phase I meetings are convened as ways to improve the efficiency of preclinical and clinical development. These meetings can help FDA and manufacturers reach early agreement on the design of the major clinical efficacy studies that will be needed to support approval. But the manufacturer of a new drug may request the FDA to designate the drug as a fast-track product concurrently with, or at any time after, submission of an application for the investigation of the drug. Within 60 calendar days after the receipt of a request, the FDA shall determine whether the drug that is the subject of the request meets the criteria. If the FDA finds that the drug meets the criteria (i.e., "serious or life-threatening and unmet medical condition" and "valid surrogate endpoint"), it shall designate the drug as a fast-track product and shall take appropriate actions to expedite the development and review of the application for approval

of such product. Under certain circumstances, the FDA may consider reviewing portions of a marketing application in advance of the complete NDA or BLA. Fast-track products may also be eligible to participate in the FDA's Continuous Marketing Applications Pilot 1 or Pilot 2 programs [11]. To facilitate FDA review, a submission for fast-track designation should contain all discussion and supporting documentation necessary to permit a reviewer to assess whether the criteria for fast-track designation are met without having to refer to information located elsewhere, yet should also not be voluminous. The amount of discussion and supporting documentation that shows that the criteria are met will vary. For example, little explanation or supporting documentation may be needed to establish that studying the drug in the treatment of a fatal condition with no approved treatment would qualify if the endpoint were mortality. More extensive explanation and supporting documentation would likely be submitted to show that for a nonfatal condition, serious or life-threatening aspects of the condition will be studied. Where acceptable therapy for the condition already exists, still more extensive discussion and supporting documentation would probably be submitted to establish that the new therapy has the potential to fill a medical need not met by existing therapy.

### 2.4 Limitations

Approval of a fast track product may be subject to the following requirements. The manufacturer needs to: (a) conduct appropriate postapproval studies to validate the surrogate endpoint or otherwise confirm the effect on the clinical endpoint and (b) submit copies of all promotional materials related to the fast track product during the preapproval review period and, following approval and for such period thereafter as the FDA determines to be appropriate, at least 30 days prior to dissemination of the materials. However, the FDA also may withdraw approval of a fast track product using expedited procedures if (a) the manufacturer fails to conduct any required postapproval study of the fast track drug with due diligence; (b) a postapproval study of the fast track product fails to verify clinical benefit of the product; (c) other evidence demonstrates that the fast track product is not safe or effective under the conditions of use; or (d) the manufacturer disseminates false or misleading promotional materials with respect to the product. Therefore, the fast track approval is only a conditional approval for the drug. Nevertheless, the prospect of early availability on the market of the drug has great implications for patients as well as for the drug manufacturer.

### 2.5 Accelerated Approval Issues (I): Clinical and Surrogate Endpoints

Since 1998 many health products have reached patients who suffered from HIV (human immunodeficiency virus), cancer, hepatitis C, and some other diseases sooner by using the fast-track act and the FTDD programs. In the mean time several scientific issues have also surfaced when following the FTDD procedures. The surrogate or intermediate endpoints and their potential in predicting clinical benefit is the major issue [12–15]. Manufacturers whose products are in fast track drug development programs may seek traditional approval based on data demonstrating an effect on clinically meaningful endpoints or well-established surrogate endpoints. Alternatively, they may seek approval under the accelerated approval regulations.

If a manufacturer seeks approval of a product in a fast track drug development program based on evidence of an effect on a less than well-established surrogate endpoint, the FDA may grant accelerated approval based on a determination that the effect on the surrogate endpoint is reasonably likely to predict clinical benefit (21 CFR 314.510 and 601.41). Drug approval under the accelerated approval regulations may also be based on demonstrated clinical effects that are not the desired ultimate benefit but are reasonably likely to predict such benefit (e.g., improved exercise tolerance in refractory heart failure might be considered reasonably likely to predict ultimate benefit) (21 CFR 314.510 and 601.41).

Section 506(b) essentially codifies in statute the FDA's accelerated approval regulations. A surrogate endpoint was defined in the preamble to the accelerated approval rule (57 FR 13234–35, April 15, 1992) as "a laboratory or physical sign that is used in therapeutic trials as a substitute for a clinically meaningful endpoint that is a direct measure of how a patient feels, functions, or survives and that is expected to predict the effect of the therapy." Although some surrogate endpoints are recognized as well-established and have long been a basis for approval (e.g., change in blood pressure or cholesterol), the accelerated approval rule allows reliance in specific circumstances on a "surrogate endpoint that, while 'reasonably likely' to predict clinical benefit, is not so well-established as the surrogates ordinarily used as bases of approval in the past" (57 FR 58942–44, December 11, 1992). To meet the statutory standard for approval, which requires the submission of "substantial evidence" to demonstrate effectiveness, "there must be evidence from adequate and well-controlled studies showing that the drug will have [its claimed] effect … That effect will, in this case, be an effect on a surrogate endpoint." (57 FR 58943-44).

With respect to approval based on clinical endpoints other than survival or irreversible morbidity, the preamble to the final accelerated approval rule pointed out that such approval would usually be considered (like other approvals based on a clinical finding) under traditional procedures (i.e., not under accelerated approval). Approval based on clinical endpoints other than survival or irreversible morbidity would "be considered under the accelerated approval regulations only when it is essential to determine effects on survival or irreversible morbidity in order to confirm the favorable risk/benefit judgment that led to approval" (57 FR 58946). The following examples illustrate types of clinical endpoints that could be a basis for approval with a requirement for further studies under the provisions of the modernization act and the accelerated approval rule:

- Clinical endpoints measuring short-term benefit in a chronic condition where short-term benefit per se does not outweigh risk and where durability of benefit is uncertain but expected.
- Clinical endpoints measuring lesser symptoms or signs of a serious disease (e.g., weight loss, appearance) when the resulting benefits do not per se outweigh risks but are expected to lead to a favorable effect on ultimate outcome, which would outweigh risks.
- Clinical endpoints measuring substantial benefits otherwise suitable for ordinary approval but where there exists a significant but limited concern that the

treatment may adversely affect ultimate outcome. Where such concerns are minimal, ordinary approval would be used. Where the concerns are substantial, data regarding ultimate outcome would be required preapproval. Between these extremes, accelerated approval may be considered.

### 2.6 Accelerated Approval Issues (II): Postapproval Study

The required postapproval study of the fast track drug has a broadened definition. That is, the postapproval study does not have to be a separate study but can be a proper continuation/extension of the same study upon which the early submission for accelerated approval is based. This use of the same study for both accelerated and final approval has been a common practice in HIV. An obvious advantage of such a proper continuation of the same study, instead of initiating a new study, is that much time and resource can be saved. It is also conceivable that in certain situations of severe diseases (such as pancreatic cancer) continuation/extension of the same trial, often very large in sample size, is the only feasible way to study the treatment effect under investigation. This broadened definition may also fit in what has been discussed in the literature [16] as "seamless phases II and III," except that the "phases" may not be labeled as such in this case. However, there are design and conduct considerations to continue a study properly so as to ensure that the final submission of the product satisfies the usual requirements of the traditional approval. For example, the protocol has to design ways to avoid excessive "crossovers" in the continuation to maintain the integrity of the trial for the later definitive endpoint.

### 2.7 Accelerated Approval Issues (III): Control of Type I Error Rate

Controlling the probability of type I error is an important issue from the regulatory and consumer standpoints. It should be recognized that the tentative, conditional approval of a fast track drug requires control of a different kind of type I error rate and that the level of control also needs to be considered. This stems from the fact that often an intermediate (surrogate) endpoint is used at the fast track review and the final approval is based on the confirmation of the treatment effect on the clinical (primary) endpoint.

Since the objective and endpoint often change between two possible approvals according to the design, this raises some interesting questions: Do we view the fast track approval submission as an interim analysis of the whole study or view the final analysis as a "poststudy extension" (i.e., the fast track submission is the main study)? If the fast track approval submission is an interim analysis, how does this type of interim analysis differ from the usual ones? What is the implication of the type I error rate when the first accelerated approval is only a conditional one? How should the type I error rate(s) be calculated and controlled? These issues have been discussed in Shih et al. [17].

The practice of continuation or extension of a study is not uncommon in clinical trials. But the current situation has a unique characteristic in that the continuation is to investigate a more important clinical endpoint (alias the primary endpoint), which requires a longer time to manifest than the surrogate or intermediate endpoint. When the study is to continue for a different mission, it is natural to view the

latter part as an extension. If the latter part is an extension, the first part for the expediting approval must be the main part. Therefore, the type I error rate for the surrogate or intermediate endpoint at the accelerated approval submission (the main study) and that for the primary endpoint at the final submission (the "internal" extension study) may be controlled separately, each at the conventional level.

Another popular view of this situation is that this is a two-stage design with one interim analysis. After all, the first part is an "early" submission. However, it should be noticed that there are differences in the current situation from the usual interim analyses in the group sequential design setting, at least from the outset.

In the usual group sequential design setting, study monitoring is based on a single endpoint. Here, in this situation two endpoints are involved in the sequential monitoring process. Furthermore, the two endpoints are "ordered" differently by their relative emphasis and expected outcomes at the two stages. The usual sense of overall (or experiment-wise) type I error rate means the probability of falsely rejecting any true hypothesis. However, this concept is not useful for the situation where one hypothesis ultimately dominates the other at a later time point with more data.

Shih et al. [17] showed that the final approval type I error rate only involves the primary clinical endpoint. The next question is whether the tentative, conditional approval for an early release of the drug on the market should be protected by controlling another false-positive rate prior to the confirmation. If affirmative, how much? There is no definitive answer to the above question since the answer may well be dependent on how much risk the consumers (patients) need to be protected from using a yet-to-be confirmed drug. That in turn depends on how restrictive the early use of the drug would be. It is reasonable to suggest that, if the time gap between the conditional early approval and the confirmatory final approval is short and that the use of the drug is limited by the label during this gap time, the allowance for the tentative, conditional approval type I error rate could be greater (i.e., more relaxed) than the situation of a wider use of the drug during a longer gap time. In any case, this is clearly another kind of type I error rate, which should not be confused with the one for the final approval, and Shih et al. [17] showed that these two error rates should not be added directly. Specifically, denote

$\alpha_F$ = final approval type I error rate
$\alpha_E$ = early submission type I error rate

Since the early submission can also be the final submission, the above two error rates overlap. Subtracting this overlap,

$\alpha_A$ = tentative, conditional accelerated approval type I error rate during the gap time

Shih et al. [17] suggested considering the following scenarios:

1. Control $\alpha_F$ at 0.05 level.
2. Control $\alpha_F$ at 0.05 level and, in addition, control $\alpha_E$ also at 0.05 level.

3. Control $\alpha_F$ at 0.05 level and, in addition, control $\alpha_A$ at 0.01 level.
4. Control $\alpha_F + \alpha_A$ at 0.05 level, for example, $\alpha_F = 0.04$ and $\alpha_A = 0.01$.

Scenario 1 views that the ultimate approval is the final conclusion of the drug's efficacy, hence only $\alpha_F$ needs to be controlled at the conventional 5% level. Scenario 2 views that the continuation of the trial is an extension study. Each section of the submission process is controlled at the 5% level. Scenario 3 controls the final submission at the conventional 5% and the tentative conditional type I error rate at 1% (implying to impose a stricter risk protection for the accelerated approval). Scenario 4 is the most conservative way of controlling the accelerated and final approvals together at the 5% rate.

In the approaches described above, the clinical and the surrogate endpoints may be controlled at different significance levels at the first stage. This is an advantage in terms of design flexibility, but not necessarily in terms of statistical power. Another approach is to test the endpoints in an ordered fashion at the interim stage so that we may use the primary and secondary nature of the endpoints and gain some power over scenario 4. This is described briefly as follows. At stage 1, we will test the surrogate endpoint first, then the clinical endpoint, since the former is considered more hopeful to show a beneficial effect in terms of statistical power. If both endpoints are "significant" at the first stage, we will stop the trial. Otherwise, we will continue. At the second stage, we reverse the order, test the clinical endpoint, then the surrogate endpoint, since the clinical endpoint is more important in the clinical sense. Compared to the previous approaches, the surrogate and the clinical endpoints are tested at about the same significance level at the first stage (e.g., 0.01) and the second stage (e.g., 0.04). This is a result of the ordering of the test and is set according to the relative importance of the endpoints at the two stages. The strategy may be viewed as an application of Holms generalized weighted procedure [18].

Other scenarios will require different statistical strategy for the type I error rates protection. For example, it might be reasonable for some cases to require early termination of the trial if the secondary endpoint is negatively significant while the primary endpoint is not at the interim stage. Conditional power of the clinical endpoint based on the result of the surrogate endpoint may also be evaluated.

### 2.8 Examples of FTDDP

Table 1 lists the fast-track products approved by CDER since 1998 through March 31, 2005.

As discussed previously, one of the crucial questions in the fast-track approval program is how consumers should be protected from a yet-to-be confirmed but promising drug. It would be interesting to research the statistics of the time for drug manufacturers to fulfill the final confirmation submission after receiving their fast track approvals and the number of failures and successes of the confirmation submission for the list in Table 1. This would give us some idea as to how much protection there should be in comparison with the usual 5% type I error rate required for a regular new drug application.

**TABLE 1 Fast-Track Products Approved by CDER (1998–March 31, 2005)**

| NDA or BLA # | Proprietary Name (Established or Proper Name) | Applicant | Approval Date | Use |
|---|---|---|---|---|
| BL103964 5039.0 | Pegasys (peginterferon alfa-2a) | Hoffmann–La Roche | 2/25/2005 | Alone or in combination with ribavirin (Copegus) for treatment of chronic hepatitis C in patients co-infected with HIV, who have clinically stable HIV disease |
| N21841 | Lamivudine 150 mg; zidovudine 300-mg tablets co-packaged with nevirapine 200-mg tablets | Aspen Pharmacare | 1/24/2005 | Treatment of HIV |
| N21660 | Abraxane (paclitaxel protein-bound particles for injectable suspension) | American BioScience | 1/7/2005 | Treatment of breast cancer after failure of combination chemotherapy for metastatic disease or relapse within 6 months of adjuvant chemotherapy |
| N21673 | Clolar (clofarabine) | Genzyme | 12/28/2004 | Treatment of pediatric patients 1–21 years old with relapsed or refractory acute lymphoblastic leukemia after at least two prior regimens |
| N21756 | Macugen (pegaptanib sodium) | Eyetech Pharms | 12/17/2004 | Treatment of exudative (wet) age-related macular degeneration |
| BL125103 | Kepivance (palifermin) | Amgen | 12/15/2004 | Decrease the incidence and duration of severe oral mucositis in patients with hematologic malignancies receiving myelotoxic therapy requiring hematopoietic stem cell support |
| N20635 SE1-035 | Levaquin (levofloxacin) | Ortho McNeil | 11/24/2004 | Treatment of inhalational anthrax (postexposure) |
| N20634 SE1-035 | Levaquin (levofloxacin) | Ortho McNeil | 11/24/2004 | Treatment of inhalational anthrax (postexposure) |
| N21743 | Tarceva (erlotinib) | OSI Pharms | 11/18/2004 | Treatment of locally advanced or metastatic non-small-cell lung cancer after failure of at least one prior chemotherapy regimen |
| N21677 | Alimta (pemetrexed) | Lilly | 8/19/2004 | Treatment of non-small-cell lung cancer |
| N50794 | Vidaza (azacitidine) | Pharmion | 5/19/2004 | Treatment of patients with certain subtypes of myelodysplastic syndrome |
| N21264 | Apokyn (apomorphine) | Bertek | 4/20/2004 | Treatment of acute, intermittent hypomobility, “off” episodes (“end-of-dose wearing off” and unpredictable “on/off” episodes) associated with advanced Parkinson’s disease |
| BL125085 | Avastin (bevacizumab) | Genentech | 2/26/2004 | For the first-line treatment of patients with metastatic carcinoma of the colon and rectum (in combination with intravenous 5-fluorouracil-based chemotherapy) |

| | | | | |
|---|---|---|---|---|
| BL125084 | Erbitux (cetuximab) | ImClone Systems | 2/12/2004 | Treatment of EGFR-expressing, metastatic colorectal carcinoma in patients who are refractory to irinotecan-based chemotherapy (in combination with irinotecan); treatment of EGFR-expressing, metastatic colorectal carcinoma in patients who are intolerant to irinotecan-based chemotherapy (administered as a single agent) |
| N21462 | Alimta (pemetrexed disodium) | Lilly | 2/4/2004 | Treatment of malignant pleural mesothelioma |
| N21548 | Lexiva (fosamprenavir calcium) | GlaxoSmithKline | 10/20/2003 | Treatment of HIV |
| N21348 | Zavesca (miglustat) | Actelion Pharms | 7/31/2003 | Treatment of type I Gaucher disease |
| N21500 | Emtriva (emtricitabine) | Gilead | 7/2/2003 | Treatment of HIV |
| BL125011 | Bexxar (tositumomab and iodine I 131 tositumomab) | Corixa Corporation | 6/27/2003 | Treatment of patients with CD20 positive, follicular, non-Hodgkin's lymphoma, with and without transformation, whose disease is refractory to Rituximab and has relapsed following chemotherapy |
| N21602 | Velcade (bortezomib) | Millennium Pharms | 5/13/2003 | Treatment of relapsed/refractory multiple myeloma |
| N21399 | Iressa (gefitinib) | AstraZeneca | 5/5/2003 | Treatment of non-small-cell lung cancer |
| BL125058 | Aldurazyme (laronidase) | Biomarin Pharm | 4/30/2003 | Treatment of patients with Hurler and Hurler–Scheie forms of mucopolysacchar- idosis I (MPS I) and for patients with the Scheie form who have moderate to severe symptoms |
| BL103979 | Fabrazyme (agalsidase beta) | Genzyme Corporation | 4/24/2003 | Use in patients with Fabry disease to reduce globotriaosylceramide (GL-3) deposition in capillary endothelium of the kidney and certain other cell types |
| N21106 | Somavert (pegvisomant) | Pharmacia and Upjohn | 3/25/2003 | Treatment of acromegaly |
| N21481 | Fuzeon (enfuvirtide) | Hoffmann–La Roche | 3/13/2003 | Treatment of HIV |
| BL103964 5000 | Pegasys (peginterferon alfa-2a) | Hoffmann–La Roche | 12/3/2002 | PEGASYS, peginterferon alfa-2a, alone or in combination with COPEGUS, for the treatment of adults with chronic hepatitis C virus infection who have compensated liver disease and have not been previously treated with interferon alpha |
| N20541 SE1-010 | Arimidex (anastrozole) | AstraZeneca | 9/5/2002 | Adjuvant treatment of early breast cancer in postmenopausal women |
| N21492 | Eloxatin (oxaliplatin) | Sanofi- Synthelabo | 8/9/2002 | Treatment of colorectal cancer |
| BL125019 | Zevalin (ibritumomab tiuxetan) | IDEC Pharms | 2/19/2002 | Treatment of patients with relapsed or refractory low-grade, follicular, or transformed B-cell non-Hodgkin's lymphoma (not acc. app.): including patients with Rituximab (Rituxan) refractory follicular non-Hodgkin's lymphoma |
| N21232 | Orfadin (nitisinone) | Swedish Orphan | 1/18/2002 | Treatment of hereditary tyrosinemia type I |

**TABLE 1** ***Continued***

| NDA or BLA # | Proprietary Name (Established or Proper Name) | Applicant | Approval Date | Use |
|---|---|---|---|---|
| N21356 | Viread (tenofovir disoproxil fumarate) | Gilead | 10/26/2001 | Treatment of HIV |
| N21335 | Gleevec (imatinib mesylate) | Novartis Pharms | 5/10/2001 | Treatment of chronic myeloid leukemia |
| BL103948 | Campath (alemtuzumab) | ILEX Pharms | 5/7/2001 | Treatment of B-cell chronic lymphocytic leukemia (B-CLL) in patients who have been treated with alkylating agents and who have failed fludarabine therapy |
| N21227 | Candidas (caspofungin acetate) | Merck | 1/26/2001 | Treatment of aspergillus infections for patients who are refractory to or intolerant of other therapies |
| BL103772 1007.0 | Remicade (infliximab) | Centocor | 12/29/2000 | Expand the indication to include the inhibition of progression of structural damage in patients with rheumatoid arthritis who have had an inadequate response to methotrexate |
| N21248 | Trisenox (arsenic trioxide) | Cell Therapeutics | 9/25/2000 | Treatment of acute promyelocytic leukemia |
| N21226 | Kaletra (lopinavir/ritonavir) Capsules | Abbott Labs | 9/15/2000 | Treatment of HIV |
| N21251 | Kaletra (lopinavir/ritonavir) Oral Solution | Abbott Labs | 9/15/2000 | Treatment of HIV |
| BL103836 1001.0 | Actimmune (interferon gamma-1b) | InterMune | 2/10/2000 | Delaying time to disease progression in patients with severe, malignant osteopetrosis |
| N20449 SE1-011 | Taxotere (docetaxel) | Aventis Pharms | 12/23/1999 | For the treatment of locally advanced or metastatic non-small-cell lung cancer |
| N21007 | Agenerase (amprenavir) Capsules | Glaxo Wellcome | 4/15/1999 | Treatment of HIV |
| N21039 | Agenerase (amprenavir) Oral Solution | Glaxo Wellcome | 4/15/1999 | Treatment of HIV |
| N20978 | Ziagen (abacavir sulfate) Oral Solution | Glaxo Wellcome | 12/17/1998 | Treatment of HIV |
| N20977 | Ziagen (abacavir sulfate) Tablets | Glaxo Wellcome | 12/17/1998 | Treatment of HIV |
| BL103795 | Enbrel (etanercept) | Immunex | 10/2/1998 | Treatment of rhuematoid arthritis |
| BL103792 | Herceptin (trastuzumab) | Genentech | 9/25/1998 | Treatment of metastatic breast cancer |
| N020972 | Sustiva (efavirenz) | DuPont Pharms | 9/17/1998 | Treatment of HIV |

## REFERENCES


1. Title 21 Code of Federal Regulations—Food and Drugs; available at: http://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfcfr/CFRSearch.cfm.
2. U.S. Department of Health and Human Services (1988), Guidelines for the Format and Content of the Clinical and Statistical Sections of New Drug Applications. U.S. Department of Health and Human Services, Public Health Services, Food and Drug Administration, Rockville, MD.
3. U.S. Department of Health and Human Services (1998), Guidance for Industry Providing Clinical Evidence of Effectiveness for Human Drug and Biological Products. U.S. Department of Health and Human Services, Public Health Services, Food and Drug Administration.
4. International Conference on Harmonisation Guideline M4; Common Technical Document for the Registration of Pharmaceuticals for Human Use.
5. Anello, C., O'Neill, R. T., and Dubey, S. (2005), Multicentre trials: A US regulatory perspective, *Statist. Methods Med. Res.*, 14, 303–318.
6. U.S. Department of Health and Human Services (1995), Guidance for Industry on the Content and Format of Investigational New Drug Applications (IND) for Phase I Studies of Drugs, Including Well-Characterized, Therapeutic, Biotechnology-derived Products, U.S. Department of Health and Human Services, Public Health Services, Food and Drug Administration, Rockville, MD.
7. U.S. Department of Health and Human Services (2000), Formal Meetings with Sponsors and Applicants of PDUFA Products, U.S. Department of Health and Human Services, Public Health Services, Food and Drug Administration, Rockville, MD.
8. International Conference on Harmonization Guideline M3; Nonclinical Safety Studies for the Conduct of Human Clinical Trials for Pharmaceuticals.
9. U.S. Department of Health and Human Services (1997), Section 112 of the Food and Drug Administration Modernization Act of 1997: Expediting Study and Approval of Fast Track Drugs; available at: http://www.fda.gov/cder/guidance/index.htm.
10. U.S. Department of Health and Human Services (1998), Guidance for Industry: Fast Track Drug Development programs—Designation, Development, and Application Review; available at: http://www.fda.gov/cder/guidance/index.htm.
11. U.S. Department of Health and Human Services (2004), Guidance for Industry Fast Track Drug Development Programs—Designation, Development, and Application Review Procedural Revision 1. U.S. Department of Health and Human Services, Public Health Services, Food and Drug Administration, Center for Drug Evaluation and Research (CDER) and Center for Biologics Evaluation and Research (CBER).
12. Ellenberg, S., and Hamilton, J. M. (1989), Surrogate endpoints in clinical trials: Cancer, *Statist. Med.*, 8, 405–414.
13. Fleming, T. R., Prentice, R. L., Pepe, M. S., and Glidden, D. (1994), Surrogate and auxiliary endpoints in clinical trials, with potential applications in cancer and AIDS research, *Statist. Med.*, 13, 955–968.
14. Fleming, T. R. (1994), Surrogate markers in AIDS and cancer trials (with discussions), *Statist. Med.*, 13, 1423–1435.
15. Neaton, J. D., Wentworth, D. N., Phame, F., Hoagan, C., Abrams, D. I., and Deyton, L. (1994), Considerations in choice of a clinical endpoint for AIDS clinical Trials, *Statist. Med.*, 13, 2107–2125.
16. Chi, G., Hung, H. M. J., and O'Neill, R. O. (2001), Some comments on "Adaptive trials and Bayesian Statistics in Drug Development" by Donald A. Berry, *Biopharmaceut. Rep.*, 9, 7–9.

17. Shih, W. J., Ouyang, P., Quan, H., Lin, Y., Michiels B., and Bijnens, L. (2003), Controlling type I error rate for fast track drug development programmes, *Statist. Med.*, 22, 665–675.
18. Holm, S. (1976), A simple sequentially rejective multiple test procedure, *Scan. J. Statisto.*, 65–70.

# *New Paradigm for Analyzing Adverse Drug Events**

**Ana Szarfman, Jonathan G. Levine, and Joseph M. Tonning**
*Food and Drug Administration, CDER, Silver Spring, Maryland*

**Contents**

1 Introduction
2 Current Paradigms of Analysis
3 Why We Need a Paradigm Change
4 A New Paradigm: Informatics
  4.1 Data Standards and Interoperable Systems
  4.2 High-Quality Data
  4.3 Restructuring Capabilities
  4.4 Data Analysis
  4.5 Reproducibility
  4.6 Maintenance
5 Examples of Practical Computer-Intensive Tools for Systematically Assessing Drug Safety Data
  5.1 Background
  5.2 Analysis of Premarketing Data with WebSDM
  5.3 Analysis of Postmarketing Data with MGPS and HBLR
  5.4 Other Data Resources
  5.5 Validation of New Methods
6 Conclusions
  References

*The opinions expressed in this chapter are solely those of the authors and do not necessarily reflect those of the United States Food and Drug Administration. This chapter originally appeared in *Computer Applications in Pharmaceutical Research and Development*, edited by Sean Ekins, Copyright © 2006 John Wiley & Sons, ISBN: 978-0-471-73779-7.

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

## 1 INTRODUCTION

Only a generation ago, 107 people in 15 states died within a few weeks after a new drug was placed on the market. Many of the victims were children. One victim was the best friend of the doctor who had prescribed the drug for him. The S.E. Massengill Company, which marketed the drug, had been looking for a solvent to dissolve sulfanilamide, a new antibiotic. A company chemist chose diethylene glycol, a chemical normally used as antifreeze. Diethylene glycol was effective in dissolving sulfanilamide but caused renal failure in the unsuspecting patients. The company owner shirked responsibility, stating, "We have been supplying a legitimate professional demand and not once could have foreseen the unlooked-for results." And how could they have foreseen the results? One doesn't find what one doesn't look for. The drug was completely legal. At that time, in 1937, the U.S. Food and Drug Administration (FDA) did not require drug products to be tested for safety [1].

The passage of the Food, Drug, and Cosmetic Act in 1938 greatly helped to improve drug safety. However, challenges in accessing and understanding drug safety data in real time persisted. Computer technology was decades away in 1938. The FDA did not develop a computerized repository of postmarketing voluntary adverse event reports until 1968. FDA scientists analyzed adverse events with paper, pen, typewriter, and perhaps a mechanical calculator. Even in the 1980s, FDA scientists had little more at their disposal to review adverse event data than a typewriter or dedicated word processor. In the 1990s, personal computers and software programs made it possible to eliminate many of the paper processes involved in adverse event analysis. However, there were still no uniform data standards and interoperable systems in place, hindering efforts to truly analyze adverse events in a systematic, computerized way. To rectify this situation, the FDA and the pharmaceutical industry began constructing a computerized repository of premarketing and postmarketing clinical trial data that would enable more efficient data analysis and decision making.

## 2 CURRENT PARADIGMS OF ANALYSIS

Challenges in systematically analyzing drug safety data continue to the present day. The typical product of a traditional analytical method is a static, paper report. Such reports usually consist of a vast number of discrete, personal, ad hoc processes that cannot typically be used to perform subsequent comprehensive *reproducible* analyses as soon as additional analyses are needed. Having an analytical method that can be subsequently audited and *reproduced* is absolutely critical because essential drug safety decisions are made from these analyses. Unfortunately, many drug safety organizations still verify the accuracy of the data they analyze by manual, ad hoc methods when comparing the data stored in a database with the primary medical records. Auditing is done by a second-party review, again by manual ad hoc methods.

Current computerized analyses of adverse events still typically consist of a vast number of discrete, often personal, ad hoc processes that mimic paper and pencil methods. Some commercial-off-the-shelf (COTS) software tools (e.g., Adobe Acrobat, Microsoft Word, Excel) do have the capability to search for specific terms

in electronic documents/case reports and do have navigational tools with hyperlinks and full-text indexing that enable researchers to create their own hyperlinks. Some other COTS software tools (e.g., SAS, Excel, Access, JMP) allow importation of electronic case report tabulations (ECRT) for more detailed analysis.

However, many of these tools, while enabling markedly faster and more detailed analysis than paper-based methods, still mimic static, one-by-one "paperlike" reports with no real-time auditing capability. Moreover, these COTS do not have integrated data analysis and automated data screening capabilities and are not optimized for systematic analyses. Furthermore, the ad hoc analyses that these COTS produce lack interactive, automatic auditing *reproducible* functions. Thus these tools are often used to produce the same dense, unwieldy paper tables of counts and percentages that were created manually before personal computers became ubiquitous.

Humans should use computers to do functional work for them in the most efficient manner possible. However, we must not delude ourselves into thinking that the mere use of a computer to analyze adverse events will magically analyze these events in a systematic, efficient way. Computers do not *automatically* produce coherent, auditable results that can be subsequently reproduced with ease. Computers must be *actively* programmed through an iterative process involving tight communication between analysts and software developers until these processes are totally functional.

## 3 WHY WE NEED A PARADIGM CHANGE

A new paradigm for computer-assisted analysis of adverse drug events is sorely needed. Critics of the current paradigm emphasize the need for more transparency in the data review process, claiming that there is the potential for suppressing negative results or for hiding safety issues in both interventional and observational studies. It is often necessary to reanalyze the data in light of new information about a particular adverse event or class of events for a drug or class of drugs months—or even years—after the initial analysis. In many such situations, it is critical to have ready access to all preclinical, clinical trial, and postmarketing data. The findings from these updated analyses are potentially so influential that they can impact drug therapy recommendations for decades. However, ready access to all of the actual data and results is still lacking in many situations.

To properly assess drug safety, we must be able to systematically tap the information captured in the massive amounts of medical data collected in both premarketing and postmarketing settings. In addition, as stated above, we must also be able to *reproduce* findings in different repositories of medical data in an auditable way. However, two major issues in studying drug safety confront us.

The first issue lies in the whole realm of the human disease process itself. Many adverse drug events mimic diseases and vice versa. Is an "adverse event" really an adverse event, or is it merely a natural occurrence of a disease process that is entirely independent of drug exposure? The science of drug safety is often complicated by the lack of objective markers of drug toxicity that can systematically separate a disease process from an adverse drug event process [2]. Clinical trials, often viewed as the gold standard to assess efficacy, are simply too limited in scope to answer safety questions in a systematic way.

The second issue involves the whole process of data collection, transformation, and presentation. At the study level, important information needed to assess the safety of a new drug is often presented in idiosyncratic ways. For example, concomitant medications are often not translated into standard drug names, and there are often subtle errors in coding of events (which we discuss further). This lack of standards hinders the creation of an integrated safety database. Later, the data that may come from numerous preclinical, clinical, and postmarketing studies are often not collected with a common data standard and are not systematically integrated into a single cumulative database before analysis. Various personnel working in different organizations or in different sections of the same organization may perform analytical tasks with nonstandardized and nonintegrated data. Even when the premarketing adverse event data of a new drug *are* incorporated into an integrated summary of clinical trial safety data, the *totality* of the safety data from *all* pre- and postmarketing research are usually *not* integrated into a coherent, analyzable database that can be used for a comprehensive, real-time analysis.

It is therefore easy to see why this current drug safety paradigm, with its lack of standards in data collection and analysis, hinders the analysis of adverse events. Without data standards in place, it is difficult to build practical, reusable tools for systematic safety analysis. With no standard tools, truly standardized analyses cannot occur. Reviewers may forget their initial analytical processes if they are not using standardized data and tools. Comprehensive reproducibility and auditability, therefore, become nearly impossible. In practice, the same data sets and analytical processes cannot be easily reused, even by the same reviewers who produced the original data sets and analyses. Not using standardized tools slows the real-time systematic analysis and reanalysis of the data because of the large number of restructuring steps needed to perform these analyses and reanalysis. These obstacles restrict the ability of analysts and senior decision makers to gain a full understanding of the entire data and the results in a timely manner. The end result is that analysts cannot routinely access the *computerized* preclinical and clinical data that supported the marketing approval of a drug, a new indication, or a new dosage schedule.

## 4 A NEW PARADIGM: INFORMATICS

We need to transition from quasi-computerized methods, in which the different elements of the analytical process are treated as discrete, "paper report" tasks, to a comprehensive *informatics approach*, in which the entire data collection and analysis is considered as a single reusable, extensible, auditable, and reproducible system. Informatics can be defined as the science of "storing, manipulating, analyzing, and visualizing information using computer systems"[3].

To analyze adverse drug events from an informatics framework, six major infrastructure elements must be in place:

- *Data Standards and Interoperable Systems* When interoperability is in place, standard, automated software tools for systematically analyzing the data can be constructed.

- *High-Quality Data* Miscoding, duplicate records, missing data items, and other data problems must be kept to a minimum.
- *Restructuring Capabilities* There must be the capability of restructuring the data for various types of analyses in a transparent way.
- *Systematic Analysis* The data must be systematically analyzed to gain insight regarding the associations between various treatments and medical conditions. This knowledge can assist in causality assessments.
- *Reproducible Capabilities* The data and analyses must be electronically accessible in real time, easily reanalyzable, and easily reproduced, even years after the adverse event data were collected.
- *Maintenance* The database and software that comprise the data systems must be maintained.

### 4.1 Data Standards and Interoperable Systems

The foundation of any efficient computer-assisted data analysis system is the creation and use of data standards. Data standards consist of standard data file names for each predefined file, standard data elements in each data file, standardized names for each data element, and standard definitions for each data element.

Data standards are exceedingly important because they allow for the use of "interoperable systems." The concept of interoperable systems can be illustrated by the situation that prevailed in the American railroad industry during the nineteenth century. At that time, there were many small, local railroad companies throughout the country. Each company, however, utilized different standards of rail gauges, that is, the distance between the two parallel rails on a track. The Baltimore and Ohio Railroad used a 4-foot, 8.5-inch gauge; railroads in the South used a 5-foot, 0-inch gauge; the Erie and Lackawanna Railroad used a 6-foot, 0-inch gauge, and so forth. The onset of the Civil War exposed the absurdity of the lack of rail gauge standards when such inconsistencies made it nearly impossible to accomplish very fundamental tasks, such as materiel transport [4]. At the 2005 FDA Science Forum, Secretary of Health and Human Services Michael O. Leavitt recounted the story of rail gauge inconsistencies to emphasize the need for interoperability to advance health information technology [5]. It becomes obvious, therefore, why standardized, unique codes for all data values across all stages of drug development and during the entire postmarketing period are so essential. Comprehensive data standards allow for the creation and systematic implementation of reusable software for analyzing medical records and adverse event data. These reusable tools enable the creation of standard analysis tools that can be shared among safety analysts for enhanced communication and learning and refined as existing analytical techniques are evaluated and extended.

All data fields (patient identification, date, sex, drug names, narratives, etc.) require rules regarding the data, because some characters, such as tabs and hard return characters, can interfere with attempts to store, read, and/or reorganize the data. For example, if the data are being migrated from Oracle into other data analysis software and data fields in Oracle contain tabs and hard return characters, the data from these fields may be split into several columns and rows upon migration into the other software. This fact is extremely important because information that

is misplaced in a data file may not be readily apparent to the user without the use of systematic approaches for analysis. This missing information may have profound and unpredictable influences at all levels of drug safety analysis, including studies using preapproval and postapproval pharmacoepidemiological data.

Ideally, the data from the whole drug development program, including accumulating postmarketing data, would be integrated into a single database. This integration would simplify the process of answering important, but previously *unforeseeable* questions (remember the sulfanilamide example). For instance, is a new dose increase as safe as the previously approved dosages? Is it enough to compare the two highest dosages with each other? A better answer could be obtained if the entire integrated clinical trial data across all dosages are studied. However, as mentioned above, drug safety data from multiple sources are not usually integrated into a single database. Compounding this problem is that many people tasked with managing data are focused on preparing a small portion of the data for a specific purpose and are not trying to create a single database that integrates all of the data. Integration of the entire safety data for a drug would enable better and faster communication among decision makers in various organizations (industry, regulatory agencies, etc.)

## 4.2 High-Quality Data

It is essential to have high-quality data in place for interoperable systems to function efficiently. Standard data structures can only be used to full advantage if they are combined with standard terminology for values populating a data element. Yet there are many potential pitfalls in data collection and configuration for analysis. Some of the more common pitfalls are discussed here, but this list is by no means comprehensive.

***Errors and Inconsistencies in Patient Identification*** Whatever system is used to identify patients in a database, it is essential that a single, unique identifier be used for each patient. This identifier must be consistent throughout the database. Errors and inconsistencies in patient identification can significantly interfere with adverse event analysis. Examples include situations in which hyphens, commas, additional zeros, or other characters are not used consistently for identifying patients. In some cases, the *same* patient in a study may have *different* patient identification numbers listed in various tables [5] (Table 1). In such situations, correctly merging patient data from multiple sources/tables becomes essentially impossible because some of the data for the same patient either will be treated as missing or will appear as data for the wrong patient. In one new drug application (NDA) submitted to the FDA, a sponsor built a "unique patient identifier" for some tables by concatenating the study identification number plus the study site identification number plus the

**TABLE 1 Example of Different Types of Unique Patient Identifiers for the Same Patients in a Clinical Trial**

| Unique Patient Identifier in Data Tables | Unique Patient Identifier in Narrative Table |
|---|---|
| 8023007 | Patient 08-023-007 is a 48-year-old Caucasian male, etc. |
| 8031011 | Patient 08-031-011 is a 61-year-old Hispanic female, etc. |

patient within-site identification number, whereas in other tables the numbers were concatenated in a different sequence. In another NDA, a sponsor separated the concatenated identifiers by hyphens in some tables but not in others. In yet another NDA, a sponsor concatenated numeric with character identifiers (the latter with leading zeroes) in some tables, but not in others. In our experience, such problems are easier to detect when systematic approaches for analysis are in place.

***Errors and Inconsistencies in Categorical Variables*** For categorical variables such as sex, race, diagnosis, and the like it is essential that the values used to designate the different categories be precisely defined and consistently used. Precise variable value definitions and their consistent use greatly simplify the analysis at hand and also future analyses. For example, a human being can readily tell that "M," "m," "Male," "male," "hombre," "homme," and so forth all refer to the same sex. A computer, on the other hand, cannot readily make this distinction (unless specifically programmed to do so) and will therefore treat these items as different values for sex. In one FDA submission, a sponsor coded "male" by using the code "1" in some studies and the code "2" in other studies, resulting in the finding of "pregnancies in men." In another submission, a similar problem resulted in males developing "female breast carcinomas."

***Errors and Inconsistencies in Formatting Dates*** It is also necessary to use standardized formats to record dates. For example, Oracle stores a date as the number of seconds since January 1, 4712 BC, and then uses various functions to display the dates in more human-friendly formats. There are many different and personal ways for recording dates; one of the authors (AS) has noted at least 25 different ways in a single clinical trial submitted for FDA review! Should February 1, 2007 be recorded as 1 February 2007, 1 Feb 2007, 1 Feb 07, or 02/01/07? This problem still exists when data for numerical dates are extracted into software programs such as Excel that do not force the user to select a unique format for dates.

***Errors and Inconsistencies in Adverse Event Coding*** Adverse events are also subject to errors and inconsistencies by coders and data entry personnel. Many of these inconsistencies become very important when adverse events are analyzed by automated software.

Adverse events need to be coded consistently with respect to letter case. Problems can occur when there is discordant coding using all capital letters, all lowercase letters, or combinations thereof, as computer software will interpret these capitalization variations as different events. Letter case sensitivity can be important when two or more words are used to describe an adverse event. For example, some databases utilizing the Medical Dictionary for Regulatory Activities (MedDRA) coding dictionary employ a coding system in which only the first letter of the first word of an adverse event is capitalized (e.g., "Atrioventricular block complete"). Failing to adhere to uniform letter case conventions across the data can result in severe errors in data analysis.

*Proper interpretation and coding* of events are also extremely important so that drug safety data can be appropriately analyzed. However, investigators and coders vary widely with respect to their health care training to properly interpret and code these events. Individual investigators may choose different terms to code the same

adverse event. Subjectivity in coding may be due to personal preferences of coders in the selection of terms versus the granularity of coding dictionaries that offer the coder a considerable array of terms to classify the reported adverse event. For example, kidney stones may be coded with a number of terms such as "kidney stones," "nephrolithiasis," "renal calculi," and "renal calculus not otherwise specified." There can also be variations in spelling of the exact same event code, such as the British spelling of "gynaecomastia" versus the American "gynecomastia."

Coding dictionaries also change over time (sometimes even before a study is completed) because of revisions in coding terminology. In one NDA submitted to the FDA, the data were coded with different versions of the same coding dictionary without properly integrating the terms into a single coding version. Such a scenario may lead to partitioning events into too many terms and therefore mischaracterization of adverse events. Again, this misclassification, although not readily apparent, may have a profound impact on the results of the analysis performed.

It is especially difficult to code events when they occur as a constellation of signs, symptoms, and laboratory findings because a coder may inadvertently minimize or exaggerate the severity of an adverse event, depending on the selection of terms to code that event. The use of only one or two seemingly benign terms such as "muscle cramps" and "pain in limb" to describe rhabdomyolysis would not provide a comprehensive picture of the event. On the other hand, categorizing an isolated case of elevated transaminases (with no other associated laboratory findings) as acute liver failure would be inappropriate. Potential signals of adverse events may be obscured or distorted depending on how the events are grouped. Splitting events into multiple terms or grouping unrelated terms may erroneously underestimate the magnitude of a signal.

***Errors and Inconsistencies in Drug Names*** Drug naming conventions are also exceedingly important. Although it may be expected that multiple names of the same medications would occur in postmarketing safety databases, this problem is also often seen in *premarketing* data. In one NDA submitted to the FDA, the data contained 900 different names for 150 unique concomitant drugs. Another NDA recorded 34,000 drug names for 2000 concomitant drugs mentioned in the study because contractors in different countries had submitted different names for the same medications.

***Errors and Inconsistencies in Numerical Data*** It is also important to understand that *character* and *numeric* data are not interchangeable. Errors can occur when there is not a clear understanding between coding with characters versus coding with numbers. A variable may be inadvertently entered with character data in one study and numerical data in another study. For example, studies measuring the effects of a 30-mg dose of a drug may utilize a "30-mg" code in some studies and a "30" code in other studies. The "30-mg" code is treated as character data, but the "30" code is treated as numerical data. Efforts to successfully combine and analyze these studies may be hindered. The "30 mg" and "30" will be treated as two different doses if the numerical variable is treated as a character. If the combined variable is treated as a number, the data for the character variable will be treated as missing in some types of further analyses (e.g., regression analyses).

***Missing Information*** Both premarketing and postmarketing collections of data are perpetually plagued by missing information. In premarketing and postmarketing clinical trials; patients can be lost to follow-up because of:

- Undocumented beginning and/or end of a medication or an adverse event (see Fig. 1)
- Undocumented death
- Undocumented serious adverse events
- Undocumented nonserious adverse events

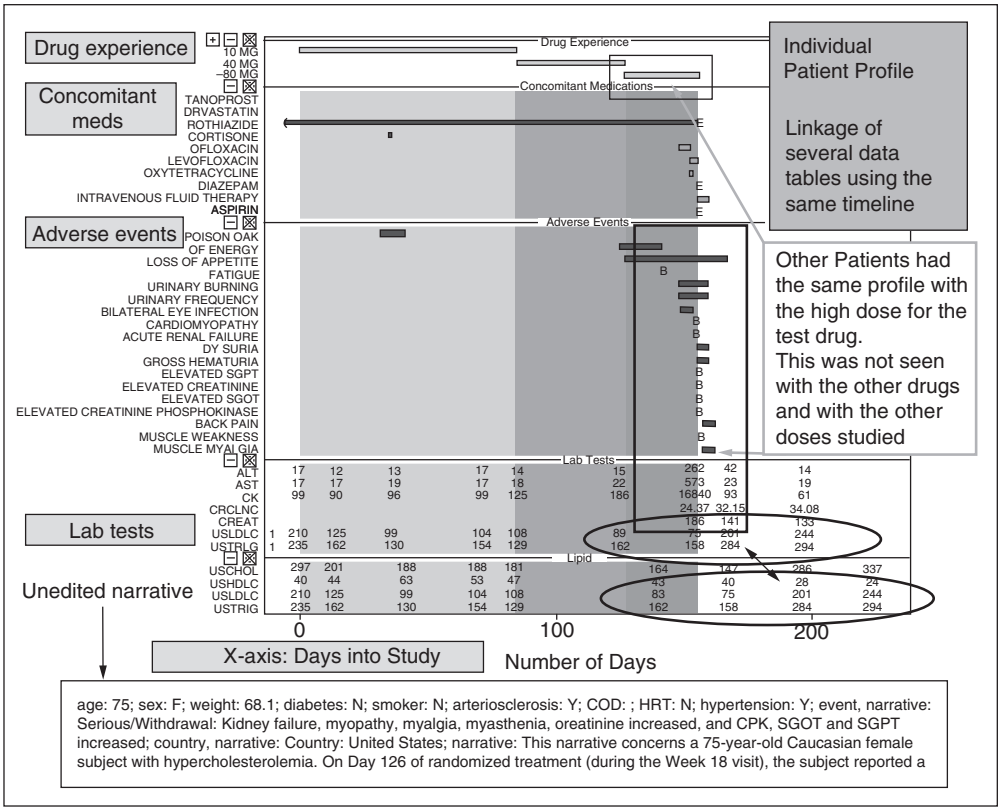


***FIGURE 1*** *An example of new drug application (NDA) data graphically displayed for an individual patient in a dose escalation clinical trial. This graph displays and links drug exposure information, clinical adverse events, concomitant medications, clinical laboratory values, demographic information, and narratives. The graph is divided into four major sections: The* x *axis for all four sections depicts the same time line, and the* y *axis the label of each row for each section. The top section displays drug exposure data for the test drug used in various doses (coded in a gray scale). The second section displays exposure to concomitant medications over time. The third section displays adverse events over time. The bottom sections display when laboratory tests were conducted and the results. Note in the areas highlighted by framed squares that clinical and laboratory adverse events were associated with the high dose of the test drug. Other patients had the same profile with the high dose for the test drug. This was not seen with the other drugs and with the other doses studied. Note that the beginning of an adverse event is displayed (B) but not the end for many adverse events. Note that the end of a concomitant medication is displayed (E) but not the beginning for some medications. Observe in the areas highlighted by framed ovals the discrepancies in the timing of the same laboratory results in different tables, making it difficult to assess whether these values occurred before or after an adverse event or a concomitant drug.*

- Failure to assess adverse events that occur more than 30 days after the last dose of a drug
- Failure to assess adverse events occurring outside a scheduled time window
- Loss of interest of a patient or unwillingness to continue in a study
- Geographical moves by patients
- Loss of insurance coverage (in a postmarketing case–control study)
- Missing records due to technical errors during data migration (see below)

Even when patients remain in a study, information regarding adverse events may not be complete. The event itself may not be coded, even when the narrative includes relevant laboratory or other data (e.g., renal failure may not be coded, even though the narrative mentions abnormal blood urea nitrogen and creatinine levels and the need for dialysis). Moreover, adverse event data (such as laboratory tests, radiological reports, biopsy reports, and hospital records) collected outside a scheduled time window, outside of the study center, or after the study is completed *might not* be included in the final database for the study. This issue is important because unexpected adverse events do not occur at known prespecified time windows. "Tight windows" of adverse event data collection can also be a problem with drugs that have a very long half-life or the potential for causing a delayed, but serious condition.

The problem of missing data is even more pervasive in postmarketing drug safety databases. Most of these systems rely on voluntary reporting for which there are no well-defined protocols. Additionally, there are significant challenges in interpreting such data because of the wide variability of reporting sources (physicians, pharmaceutical companies, patients, attorneys, etc., as well as submissions from different countries). With electronic medical records, the presence of important relevant, but unreachable, data elements may not become apparent without the use of systematic approaches for analysis. This problem may be compounded in some epidemiological studies that ignore the presence of missing information in their analyses.

***Duplicate Information*** Both premarketing and postmarketing databases may contain duplicate reports of adverse events. In postmarketing databases of voluntary reports, duplicate information on the same adverse event case may be submitted by several sources. For example, submission of information on the same event may be duplicated by the treating physician, the dispensing pharmacist, the nurse, the patient's attorney, and/or the patient himself. Submission of information by drug companies may compound the problem, even though they are attempting to comply with regulatory requirements mandating submission of adverse events. Multiple drug companies may submit information on the same case, using their own unique patient identifiers, especially when the adverse event is associated with drugs manufactured by several different companies. There may also be a series of follow-up reports for the same case as additional information becomes available. However, updated patient identifiers for the same patient may not be linked to the original patient identifier. Duplicated information on the same event may also come from several teams working for the same drug company. Despite these problems, removal of duplicate reports is absolutely essential (though challenging) because superfluous information may result in false-positive signals of adverse events and wasted analysis time.

In postmarketing electronic longitudinal medical records, redundant and potentially contradictory information may come from several sources (e.g., reports from medical residents, their supervisors, the attending physicians). In these cases time stamping of each event may help to delineate the important sequences in understanding the adverse events.

***Other Inconsistencies*** Inconsistencies in drug safety data due to difficulties in standardization also include a subject's primary diagnosis, differential diagnoses, relevant medical history, relevant physical exam findings, pertinent information from hospital records, and follow-up information (all of which may be subjective). There may also be a lack of consistency in the narrative summaries for individual patients and the data supporting the narratives. Indeed, it is difficult to clearly describe in the narratives the sequence of adverse events, medication, and laboratory results by using case report forms or line listings as source information. In complex patient records, case report forms and line listings may generate uneven temporal sequences of adverse events, concomitant medications, dosages, and so forth that cannot readily be comprehended. Such data need to be visualized by tools capable of displaying the complex, interrelated information on a common time line (Fig. 1).

Additional examples of variability in data collection (which, in turn, affects data interpretation) include questionnaires and physical exam forms. Questionnaires often utilize open-ended questions that allow great variability in the type and extent of adverse event information gathered. Physical exam forms—even when designed in a checklist format—may elicit variable collection of adverse event data; what is a serious event to one clinician may not be serious to another.

### 4.3 Restructuring Capabilities

***Reconfiguration of Data*** Drug safety data from different sources are often pooled or combined in databases. Reasons for combining data vary. In the case of premarketing studies, data from different sites are routinely combined because one site may not be able to recruit enough patients for a study. Data from different studies are often combined to increase sample size and therefore statistical power for detecting an uncommon adverse event.

In postmarketing safety surveillance databases, data from different countries or from different sources (physicians, patients, drug companies) may be combined in the same database in an attempt to obtain as much information about approved drugs as possible. Pooling or combining data can allow explorations of drug toxicity among various subgroups. Having a large database allows studying possible drug–drug, drug–disease, and drug–demographic associations.

When reconfiguring the data, several issues must be borne in mind. Combining data from different data sources can obscure potentially meaningful signals of adverse drug events [6]. For example, combining data for the term "colitis" with "ischemic colitis" may obscure the presence of ischemic colitis. Also, the criteria for reporting and coding an adverse event may differ among various data sources (e.g., countries with disparate regulatory requirements and different coding dictionaries). Reference ranges for normal values may vary, depending on the reporting source. When reference ranges vary, changes from baseline grouped

by treatment assignment may provide useful information. Patient populations may also vary in different study arms or in different countries where studies are conducted. Different populations may tolerate drugs differently or have varying levels of drug sensitivity. Study design may vary among sites, especially in terms of how outliers are handled and follow-up information is obtained. The selection of analytes and biomarkers of toxicity to measure may vary depending on the reporting source and when the data were collected (criteria for toxicity and case definition may be site specific and can change over time). Duration of drug exposure (as well as drug dose) may vary among studies. Because so many factors can influence the results, the process of transforming and combining data from different sources should be documented in a way that is easy for subsequent investigators to understand.

***Reconfiguration of Databases*** Not only must data from different sources be preprocessed ("cleaned"), reconfigured, and validated before analysis, but entire databases must also sometimes be reconfigured and validated. This is especially the case if the database has evolved and has been maintained over a long period of time.

A good example of an evolving database is the FDA's Adverse Event Reporting System (AERS) database containing reports of spontaneously submitted adverse events. AERS has undergone several configurations since its inception in 1968. This database was known as the Spontaneous Reporting System (SRS) from 1968 to 1997. Adverse events were coded into SRS with the COSTART (Coding Symbols for Thesaurus of Adverse Reaction Terms) dictionary. Only 1200 event codes were present in the COSTART dictionary. COSTART was replaced by the much more granular MedDRA (Medical Dictionary for Regulatory Activities) system of coding in 1997. MedDRA contains over 15,000 preferred term event codes, of which 10,000 are currently in use in the database. When SRS was modified to build AERS, adverse events coded with COSTART terms were mapped to MedDRA terms. Moreover, drug names in AERS have been and are still collected in free text form. There is substantial variation among reporting sources regarding the manner in which drug names are ultimately listed in adverse event reports. Drugs may be listed by their generic or trade names, with numerous and creative variations in spelling, abbreviations, spacing, and punctuation (see Section 4.2 on high-quality data). Thus what is now termed the AERS database is really a data set containing data that have undergone several organizational changes during more than three decades of data collection. This mapping, recoding, organizational reconfiguration, and validation of the database has been necessary to provide a uniform format for data analysis, yet this entire process has, understandably, been labor intensive and challenging.

With electronic medical records, multiple clinical records for the same patient may be treated as belonging to different patients during anonymization and migration of electronic medical records, tainting analytical conclusions. This problem may be difficult to untangle once the anonymized data migration takes place.

Sound analytical assessments require that analysts understand the manner in which the data were collected, reconfigured, migrated, and combined. These processes should be documented in a transparent way so that future investigators can readily understand the anonymization and migration in real time.

### 4.4 Data Analysis

Current practices require that all the data collected be "cleaned," reconfigured, and standardized in order to perform analytical and integrative tasks. These processes are complex, time consuming, and error prone—especially when there are many different personal standards in place. This process of "data cleaning" [7], reconfiguration, standardization, and integration must be done because the data are typically collected by several different contract research organizations, each with its own independent personal data collection standards. Because of personal standards in data cleaning and reconfiguration, many investigators end up analyzing only a small portion of the safety data, resulting in missed rare but serious adverse events or risk factors. If systematic data cleaning and reconfiguration are not done initially, then even seasoned investigators will still waste time constructing a new integrated database prior to analysis.

***Size of the Database*** Database size is important in assessing drug safety in both premarketing and postmarketing settings. During clinical trials in the premarketing period, the number of subjects in a drug safety database often depends on the intended use of a product. For products intended for long-term treatment of non-life-threatening conditions, subjects studied may number in the thousands. For products intended for *short-term* treatment of *rare or life-threatening* conditions for which there are few effective treatment options available, a "smaller" number of subjects are studied, but there can be great subjectivity in defining the word "small" in such cases. This subjectivity is in part a reflection of the wide spectrum of disease severity for which such products might be indicated. For products intended for *chronic* treatment of *life-threatening* conditions, the number of subjects would need to be greater, but again, there is great potential for subjectivity.

Larger databases can help with risk–benefit decisions, but how can we achieve consensus on the exact size of the number of subjects needed for the database? A clinical trial database that contains limited information on a small number of subjects will likely lack the statistical power needed to detect differences in adverse events between control and treatment groups. A researcher may specify criteria for the minimum differences of adverse event rates between the two groups in an attempt to identify important safety signals, but ultimately these criteria are arbitrary. Even when a study enrolls a large number of subjects and records a large volume of data in a database, it is difficult to adequately identify all potential risks associated with a product. Some risks will only become apparent once a product is approved, that is, when hundreds of thousands or even millions in the general population are exposed. Adverse events in the postmarketing period are often collected in a voluntary manner through the use of spontaneous reporting systems such as FDA's previously described AERS database or drug registries (e.g., clozapine). Yet extracting safety information from these databases—even if they are large—can still be challenging because background rates for various events can be difficult to obtain systematically. What is the background rate for headache, rash, decreased appetite, appendicitis, and fatigue—events that frequently occur in the population independent of drug therapy? Moreover, how does one assess the background rate of an event in a prespecified, but non-drug-exposed population compared to the rate of the same adverse event in a similar, but drug-exposed population? For example,

how does one assess the risk of confusion in elderly diabetic patients due to the effects of a drug prescribed for diabetes from baseline rates of confusion in elderly diabetic patients who are not receiving the same diabetes drug and concomitant drugs?

***High-Dimensional Aspect of Data*** The high-dimensional aspect of data collected in a study can make the analysis of these data challenging. Even a simple clinical trial may have recorded dozens of measurements for each patient. More sophisticated studies may have hundreds of measurements per patient. Complex tests measuring the physiology of a specific system, such as pulmonary function tests or echocardiograpy, may be impossible to standardize over many different treatment groups or over a period of time because of technological changes and interpretation of findings. Pathological specimens (e.g., biopsies) may also be difficult to classify in a systematic and objective way. Other high-dimensional data include pharmacokinetic and pharmacodynamic information in phase III clinical trials, both of which are crucial in anticipating potential safety problems, especially in patients with impaired hepatic and renal function or in patients taking many concomitant medications. The challenge then lies in how to analyze such high-dimensional data. Unfortunately, we often do not have systematic methods for reducing the dimensionality of the data to find the subset of variables for which the treatments differ and the key statistics that describe the differences without losing important interdependent information.

***Variations Among Subjects*** In the premarketing phase of assessing drug safety, it is important to have a study population that is not only representative of the target population but also sufficiently diverse in terms of demographics. This diversity will bolster the generalizability of the safety analysis. Diversity in the study population can be enhanced by including both males and females and also patients in different age, racial, body weight, and risk factor groups. Yet this same variability that is so necessary in increasing generalizability also presents challenges in analyzing data. For example, there may be a great deal of variability in renal and hepatic function among elderly patients. Plasma levels of a drug can also vary greatly if patients are given the same dose regardless of body weight, body surface area, or renal and hepatic function. There may also be great differences in sensitivity to the effect of a particular drug among individual patients. These issues may also taint the analysis of postmarketing safety data, including electronic medical records. Computerized safety analysis systems (discussed below) can aid in studying the effect of these variations by automatically generating stratified analyses to adjust for the impact of these patient attributes on adverse events.

***Temporal Relationships of Adverse Events*** The temporal relationship between duration of product exposure and development of an adverse event is important in assessing causality. But how can data on temporal relationships be systematically summarized in a database containing thousands or even hundreds of thousands of subjects? Temporal relationships cannot be clearly elicited if only frequencies of adverse events between treatment and control groups are compared. There can be many disparities in the subjects' time of exposure or time at risk. Toxic manifestations of drugs may not occur until several months or even years after the initial exposure to the drug. How do we systematically assess delayed toxicity of a previ-

ously prescribed drug from the effect of a newly prescribed drug? Such a scenario occurred with reported cases of pancreatitis associated with valproic acid therapy, in which some cases appeared several years after therapy [2].

Sometimes an adverse event that occurs with a different frequency in the treatment group than in the control group is also qualitatively different in the two groups. For example, suppose a rare but serious vascular event occurs more frequently and earlier in the treatment group than the control group and is more likely to lead to discontinuation in the treatment group than in the control group. We need to describe that the event occurs more frequently and earlier in the treatment group and that it is more likely to cause discontinuation of treatment. This information is difficult to convey with tabular data, but often becomes clear when the data are presented in graphical form [8, 9, Figure 2].

***Effect of Concomitant Medication*** Assessing drug–drug interactions is absolutely critical in evaluating the safety profile of a drug. Interactions can occur when one drug affects the absorption, distribution, metabolism, or excretion of another drug or drugs, producing additive or antagonistic effects on the other drugs. Various foods and dietary or herbal supplements (e.g., St. John's wort) can also interact with drugs. Yet how do we systematically assess adverse events from one drug as opposed to adverse events from concomitant drugs or supplements?

***Effect of Preexisting Disease*** The adverse event profile of a drug can be confounded because of the effects of underlying disease for which the drug may or may

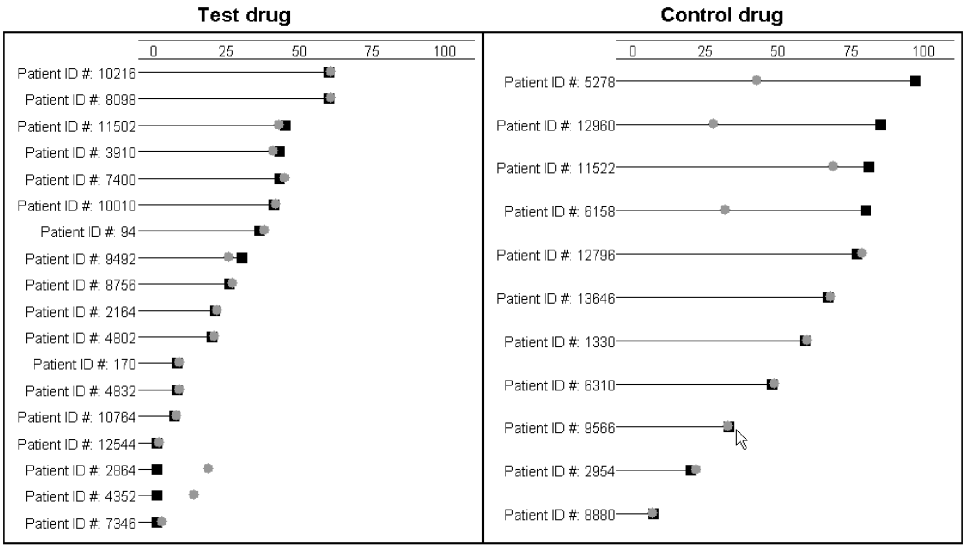


***FIGURE 2*** *A display that summarizes the duration of treatment (black squares) and the timing of serious vascular events (gray circles) for the subset of patients who withdrew from treatment because of an adverse event. Each line represents a single patient's experience over time in days for the test (left panel) and the control drug (right panel). Patients are sorted by decreased duration of treatment. In this 1:1 randomized clinical trial there were 18 withdrawals due to a severe vascular adverse event with the test drug. This is in contrast with the control drug, with 11 withdrawals. Withdrawals with the test drug occurred sooner than with the control drug.*

not be prescribed. Comorbidity can affect a drug's potential for inducing an adverse event. However, it is often difficult to separate the influence of preexisting disease when assessing the potential toxicity of a drug. How do we systematically separate the effects of a drug from a disease, the progression of that disease, or multiple diseases syndromes—each with its own varying rate of progression?

Preexisting disease, such as renal or hepatic disease, can especially influence the metabolism and excretion of certain drugs. In clinical trials, and in electronic longitudinal medical records, it is important to have sufficient variability in disease states and concomitant diseases among subjects in both the study and control groups. Investigators need to consider whether the adverse events that occur are due to abnormalities in the distribution, metabolism, and excretion of drugs as a result of underlying disease. These analyses could be systematically facilitated by having standardized ways of measuring blood (and in some cases, tissue) levels of drugs and their metabolites.

***Lack of Objective Markers of Drug Toxicity*** Some products have well-established, valid biomarkers that can be measured to track certain safety concerns. For example, a dose–response association with proteinuria, creatine phosphokinase, and urine myoglobin levels can be monitored to assess the safety of HMG-CoA reductase inhibitors (Fig. 1). However, there are often no specific markers (or pathognomonic clinical findings) of toxicity for many drugs. For most drugs under investigation—and most marketed drugs—practical tests to measure toxic drug or metabolite levels are not widely available [2]. Additionally, relying on product labeling of a drug in a similar class as a clue to investigate the toxicity of a drug is faulty, as labeling can be influenced by a number of factors including litigation and publicity.

***The Overwhelming Volume of Data to Be Analyzed*** Any given drug safety researcher—whether a statistician, clinician, epidemiologist, or safety evaluator—can only analyze so much nonstandardized data in a given time period. Premarketing databases contain very detailed data on thousands or perhaps tens of thousands of subjects. In some cases, premarketing data may have been collected over several decades for a wide array of indications. Postmarketing drug safety databases may contain millions of adverse event reports. The Composite Health Care System II database maintained by the U.S. Department of Defense contains over 9 million medical records. The FDA's AERS database contains over 2.5 million adverse event reports collected since 1968. As mentioned above, AERS utilizes the MedDRA classification system for its adverse event coding system. For these reports, approximately 10,000 MedDRA-preferred terms have been coded for 4000 generic drugs in the database. Thus over 40 million drug–event combinations are theoretically possible. This situation, compled with the fact that the FDA currently receives over 1000 new reports of adverse events each day, exemplifies the daunting task that safety evaluators face when analyzing postmarketing adverse event data. With such a large volume of reports to review each day, exploring signals based on clinical judgment in combination with threshold reporting frequencies may not always be optimal or even practical. Such an approach makes it difficult to contextualize information regarding adverse events. How do we systematically access such huge databases to select appropriate variables for analyses? Without adequate drug expo-

sure data and baseline rates of disease processes (which may be erroneously attributed to "adverse drug events") in specific populations at risk, how do we determine whether 20 cases of a specific drug–event combination is disproportionately frequent to merit further investigation?

***Subjective Analysis Strategies*** There are also too many *subjective* analysis strategies in place. What is convincing to one analyst is not convincing to another. For example, what should be done when there are outliers in the data, that is, measured values of findings (such as laboratory values) or events that deviate substantially from the reference range? If outliers are ignored, important safety findings may not be identified; on other hand, outliers may represent errors in data collection. This was the case for one patient in an NDA submission. The patient had a serum creatinine value of 13 mg/mL (over 10 times the normal value, by any reference!) but a normal serum blood urea nitrogen (BUN) value.

***Limited Knowledge of Exposure and Reporting Rates in Postmarketing Data*** Unlike clinical trials and electronic medical records in clinical practice, postmarketing voluntarily reported data contain limited information about the total number of patients exposed and the duration of exposure. This problem is compounded by the fact that adverse events are often underreported [2, 9].

### 4.5 Reproducibility

Traditional analytical methods make extensive use of computers, but typically these methods still require constant restructuring of the data and multiple analytical tools. This endless restructuring wastes time and productivity and also makes the analytical processes difficult to document, audit, and reproduce in real time. This situation also makes it difficult to reconstruct and update analyses in real time when *new* adverse event data become available or when *new* questions need to be asked. The application of comprehensive data standards allows the use of integrated, reusable software for analyzing adverse event data. This integration facilitates the reproducibility of the results.

### 4.6 Maintenance

Any computer database system will require maintenance. This maintenance includes such things as actively identifying and correcting data errors, ensuring that data can still be used with upgraded software and that this software can be used with upgraded hardware and data. Maintenance also includes actively testing and identifying computer bugs and adding new features and enhanced functions to the software.

## 5 EXAMPLES OF PRACTICAL COMPUTER-INTENSIVE TOOLS FOR SYSTEMATICALLY ASSESSING DRUG SAFETY DATA

### 5.1 Background

Although 500,000 individuals were enrolled in clinical trials that were submitted to the FDA during 1990–1995 [10], the lack of a repository of clinical trial data, standardized data, and interoperable systems precludes us from efficiently tapping and

reanalyzing these data. This missed opportunity underscores the need for standardization and interoperable systems, as discussed above (see Section 4.1 on data standards and interoperable systems).

Drug safety reviewers spend a great amount of time learning the peculiarities of the data structure format and the variable names used with each NDA (and NDA supplement) submitted for marketing approval. As described above, for some NDAs the data from several studies must be incorporated into an integrated summary of safety data set and validated before performing a safety analysis; if every study uses a different data structure, this is an arduous task. To rectify this situation, the FDA is using the Clinical Data Interchange Standards Consortium (CDISC) Submission Data Tabulation Model (SDTM) format adopted as a standard by the FDA in July 2004. The implementation of such data standards allows for development of standard and comprehensive analytical tools that can automatically generate standardized and comprehensive analyses that can be used across numerous NDAs.

### 5.2 Analysis of Premarketing Data with WebSDM

An example of an analytical tool that utilizes CDISC data standards is WebSDM (Web Submission Data Manager) by Lincoln Technologies. The FDA has recently implemented WebSDM to analyze two NDAs in real time [11]. Although the original NDA data were submitted with nonstandard formats and the data had to be transformed into the Study Data Tabulation Model format before being loaded into WebSDM, the review process for these 2 NDAs was more efficient than with standard methods. In this case the data was transformed to demonstrate the concept that the use of standardized data simplifies the analytical process. WebSDM saves time because it first ensures that the data submitted to the FDA complies with CDISC format and then uses standard methods that enable automation of the analytical processes. Typically, FDA reviewers receive different data formats for each NDA; however, WebSDM eliminates the need for reviewers and supervisors to learn where the variables for analyses for each NDA are located (Fig. 3) and the different data formats for each NDA. For example, assessing potential liver injury by analyzing increases in serum alanine aminotransferase (ALT) and total serum bilirubin (TBILI) is done in one step instead of multiple cumbersome steps. It was also easier for the reviewers to fulfill the requirements of the NDA Review Template in a recent FDA guidance document for reviewers [12]. WebSDM also eliminates the need to reconfigure the data and the analytical tools for each new NDA analysis. WebSDM allows reviewers to use tailor-made, reusable tables and graphs of patient data in any NDA or supplement. The Sector Map graphical tool (with interactive drill-down capabilities) visualizes clinical trial data by highlighting higher-than-expected associations of adverse events compared to control groups. These features greatly simplify interpretation of the data. (See example with postmarketing data in Fig. 4.) These advanced graphical and analytical features are designed to simplify the interactive analysis of the clinical trial data.

### 5.3 Analysis of Postmarketing Data with MGPS and HBLR

For drugs already on the market, the FDA is utilizing the multi-item gamma poisson shrinker (MGPS) statistical algorithm [6, 9, 13] to systematically and simultaneously

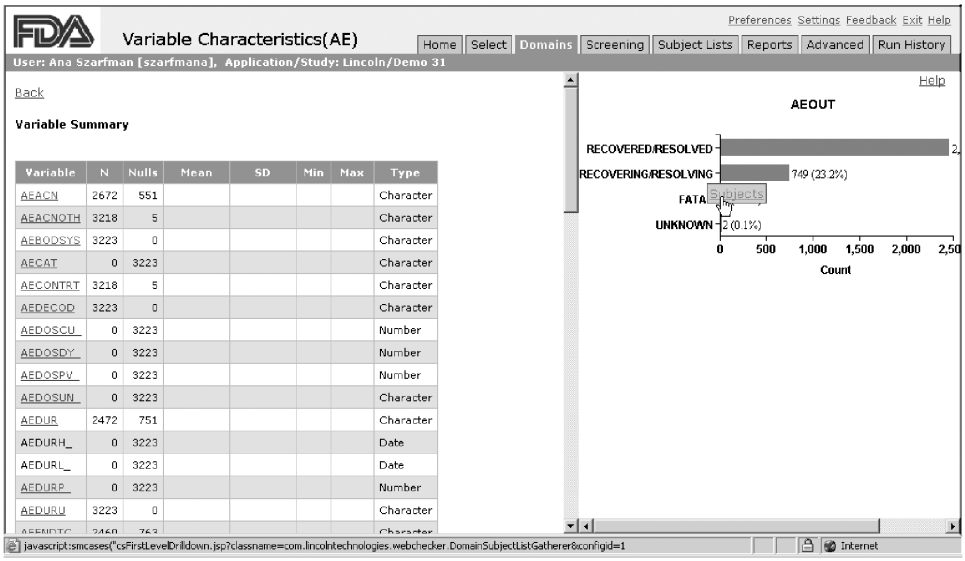


***FIGURE 3*** *Once the data are transformed into CDISC standards and integrated with a drug safety analysis system, the data can be easily analyzed. In this figure we show a sample screen from WebSDM a drug safety analysis system being evaluated by the FDA. This screen allows the user to view different attributes of the variables in a user-specified data set. When a variable is selected, a graphical display of the data is produced on the right-hand side of the window. The user can then select to visualize a graphical display of the individual patient profiles under the variable in a different window.*

detect signals of higher-than-expected drug–adverse event associations in its post-marketing drug safety databases. To identify these signals, MGPS employs a disproportionality analysis of drugs and events, combined with Bayesian shrinkage. MGPS uses the independence model as the basis for computing the drug–event expected counts. MGPS includes a Maentel–Haenszel style approach for adjusting the expected counts for potential strata heterogeneity. When applied to the FDA's AERS database, the MGPS program systematically stratifies the data by over 1000 categories (9 categories for age, 3 for sex, and 38 for year of report) to help adjust for background differences in relative reporting rates by these variables. The FDA thus far has focused its analytical efforts on the AERS database, but MGPS can be applied to any large drug safety database. The British Medicines and Healthcare Products Regulatory Agency (MHRA) has recently begun using MGPS as part of its pharmacovigilance program. MGPS also incorporates advanced graphical tools for analysis, including the Sector Map described previously (Fig. 4).

The FDA is also exploring another statistical algorithm, the hierarchical bayesian logistic regression (HBLR), to study signals that may be influenced by polypharmacy. HBLR corrects for confounding induced by concomitant medications (each with its own potentially strong adverse event associations) throughout the database [14, 15]. HBLR uses a prior distribution (estimated from the data) to improve the modeling of the joint associations for up to hundreds of drugs with a logistic regression response variable. The HBLR adjusts for both "signal absorption" and "signal masking." Signal absorption is a phenomenon whereby an "innocent bystander" drug is falsely signaled as being associated with a particular adverse event simply

**Cerivastatin**

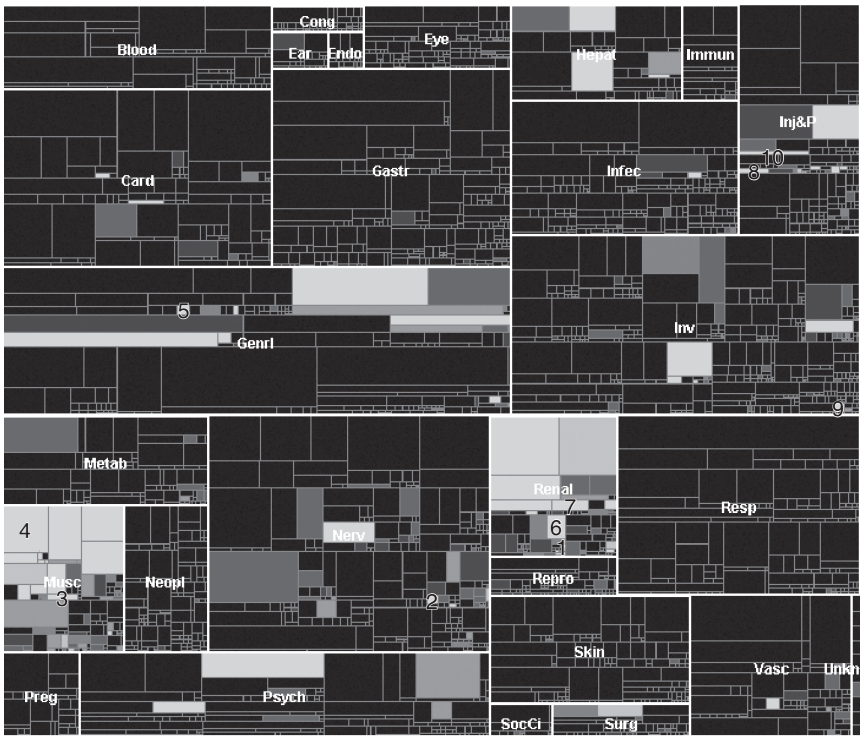


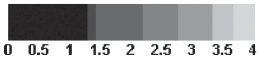


| Rank | SOC | Term (PT) | EBGM | AERS cases |
|---|---|---|---|---|
| 1 | Renal | Myoglobinuria | 15.416 | 104 |
| 2 | Nerv | Myasthenic syndrome | 12.449 | 1021 |
| 3 | Musc | Myositis | 12.198 | 2670 |
| 4 | Musc | Rhabdomyolysis | 11.501 | 12024 |
| 5 | Genrl | Organ failure | 10.121 | 685 |
| 6 | Renal | Chromaturia | 9.437 | 3034 |
| 7 | Renal | Renal tubular necrosis | 9.249 | 2547 |
| 8 | Inj&P | Muscle injury | 9.025 | 998 |
| 9 | Inv | Myoglobin blood increased | 8.807 | 527 |
| 10 | Inj&P | Polytraumatism | 8.497 | 1105 |

**NOTES:** Additional restrictions for graph:

Color controlled by: EBGM.
Size controlled by: relative importance.
Maximum intensity at signal score of 4.0.
Omit rare terms used fewer than 100.0 times
List 10 top scores
Show score indexes.
Group by HLT
Group by HLGT
Group by SOC
Lowest level displayed: PT

Print

Close

because of its frequent coprescription with another drug that is associated with that same event. Signal masking occurs in a database when there is failure to detect a weak signal for a particular drug because of the presence of other strong signals for the same adverse event, usually because of the homogeneity of drugs in the database [16]. Our initial experience indicates that HBLR may be a useful adjunct to MGPS in postmarketing safety assessments, especially in polytherapy regimens [15]. Dr. William DuMouchel has presented an overview of future empirical Bayes

◄

***FIGURE 4*** *Sector map display of the MGPS data-mining profile for a drug, using a dictionary of medical terms. This display shows the safety profile of cerivastatin, a drug withdrawn from the U.S. market in August 2001 because of reports of fatal rhabdomyolysis, renal failure, and other organ failure [24]. The sector map shows strong signals for several serious muscle events including rhabdomyolysis within the Musc System Organ Class (SOC) and for renal failure within the Renal SOC. The strong renal failure signals with this drug were unexpected. In addition, there were huge differences between cerivastatin and other statins regarding the magnitude of the renal failure signals.*

*A* sector map *for data mining results is a visual presentation of data for a particular drug across all System Organ Classes (SOCs).*

*Each SOC is represented by a large tile in the sector map. Smaller tiles within each SOC tile represent Preferred Terms (PTs). A sector map graph is available for two-dimensional results of MGPS data mining runs.*

*PTs are ranked in descending order of EBGM values, a value that indicates the reporting association between a drug and an event (9). A color or a gray scale key (in this article) indicates relative ranking. The user can select to list the ranked PTs below the sector map.*

Note: *The primary path of a PT in the MedDRA hierarchy is used to determine where it appears in the sector map. If a PT is not in the event hierarchy associated with the configuration, it appears in a SOC tile named "Unknown".*

*The list of ranked PTs below the sector map includes the following information:*

*Column Description*
*Rank: Ranking of the term (combined with the drug) according to values of the EBGM values*
*SOC containing the term.*
*Term: (PT) Specific PT.*
*EBGM or other statistics used to configure the sector map*

*If the Notes checkbox is checked, a Notes section provides information about the selection criteria and display options used for the graph.*

- Color or shades of gray (shades of gray in this chapter), size, position in space, grouping, and ranking of tiles *provide a "big picture" overview of the adverse event profile of a drug.*
- Color or shades of gray: *Light gray in this figure corresponds to stronger signals (high EBGM values).*
- Size: *A large tile (with a white border) defines each SOC (System Organ Class) in the MedDRA dictionary.*
- Box size for each PT *(preferred adverse event term)*: *The box size is based on the number of serious cases of the term across all drugs in the AERS database. Thus, the box size of each PT is stable over different displays of different drugs.*
- Position in space of each SOC and PT: *SOCs and PTs are always represented in the same area of the sector map. The position of each SOC and PT is stable over displays of different drugs.*
- Grouping of PTs: *PTs are grouped by high level term (HLT), high-level group (HLGT), and SOC.*
- Ranking of PTs: *PTs are ranked in descending order of EBGM values for each drug. EBGM: Signal Score. AERS cases: number of cases for the term in the AERS database. The PT "renal failure acute" is ranked 43rd with cerivastatin (see the top right rectangle within the Renal SOC, not displayed in the ranking for the top 10 drugs in the list below the graph) and "renal tubular necrosis" is ranked 7th with this drug. The PT "renal failure acute" has a larger box size than "renal tubular necrosis" because it has a much larger number of serious cases across all drugs in the AERS database.*

methods for estimation of adverse event rates in clinical trials and active surveillance [17].

### 5.4 Other Data Resources

The number and size of databases containing drug safety information is growing rapidly, with some databases already containing millions of records [18, 19]. Analyzing several databases can help strengthen or refute a putative safety problem based on the results of the primary database analysis. Databases that can be analyzed include those maintained by various countries (e.g., the British General Practice Research Database or GPRD); various organizations (e.g., the World Health Organization and health maintenance organizations); various agencies (e.g., the U.S. Department of Defense, Department of Veterans Affairs, and Centers for Medicare and Medicaid Services Affairs), and others. Adapting new standard computer-intensive analytical tools to analyze data converted into standardized format will allow different experts to review each other's selection criteria and results so that conclusions can be more objectively studied and understood.

### 5.5 Validation of New Methods

Validation of new methods for analyzing drug safety data is challenging. There is no gold standard tool that can provide complete information about the whole spectrum of toxicity for a given drug and the magnitude and extent of this toxicity in specific subpopulations [9]. These facts, coupled with the discordant manner in which medical data are collected, make it very difficult to systematically analyze drug safety data in real time and to cross-reference multiple collections of medical data and results in a systematic way. The application of advanced computer methods offers a tremendous opportunity to analyze large databases in a timely and consistent manner and to learn about drug safety in a systematic way. These efforts will assist in creating gold-standard positive and negative signal definitions and methods given the data analyzed [20]. With these gold standards in place, we will be better able to further advance the art as well as the science of systematic drug safety assessment across databases.

## 6 CONCLUSIONS

Even though great progress has been made since the passage of the Food, Drug, and Cosmetic Act in 1938, pharmaceutically related adverse events are still, unfortunately, responsible for a tremendous burden of pain and suffering in the United States [21], and old problems still persist around the world [22, 23]. Adverse events are also responsible for tremendous financial costs to taxpayers, insurance policyholders, insurance companies, and pharmaceutical companies. With such public health and economic costs in mind, we should carefully consider the strengths of new pharmacovigilance approaches while clearly acknowledging their limitations as well.

New tools that exploit the power of modern computer technology provide innovative approaches to help identify and investigate potential drug safety problems

in a systematic way. These new computer methods can assist in pharmacovigilance efforts because the results of an analysis from a particular drug safety database can be compared to the results from other, independent databases (e.g., clinical trial, health maintenance organization, or military medical databases). With interoperable standards in place, statisticians, epidemiologists and other analysts will be in a great position to improve drug safety and communication between all parties involved (consumers, clinicians, regulators, industry representatives, and legislators). By using the same software and data we can validate each other's selection criteria, results, and interpretation. As a result, researchers and policy makers will be better equipped to understand the limitations and biases of the data, leading to more objective decisions regarding drug safety.

## REFERENCES


1. http://www.fda.gov/oc/history/elixir.html; accessed April 2009.
2. Szarfman, A., Tonning, J. M., and Doraiswamy, P. M. (2004), Pharmacovigilance in the 21st century: new systematic tools for an old problem, *Pharmacotherapy*, 24, 1099–1104.
3. http://www.janssen-cilag.com/gldisplay.jhtml?itemname=glossary&product=none#gl_informatics; accessed April 2009.
4. http://www.trains.com/Content/Dynamic/Articles/000/000/003/011gsqfq.asp; accessed on March 1, 2006.
5. 2005 FDA Science Forum: Advancing Public Health through Innovative Science. Washington, DC, April 27, 2005.
6. Datamining with applications in Genomics, Clinical Trials, and Post-marketing Drug Risk. Schering—Plough Workshop of May 31–June 1, 2001. Harvard School of Public Health. Section III: Data examples; available at: http://www.hsph.harvard.edu/biostats/events/schering-plough/old/agenda2000-01.html Video Clips. Statistical issues in drug safety monitoring. Schering—Plough Workshop of June 2–3, 2005. Harvard School of Public Health. Section III: Pharmacovigilence and Datamining for Drug Safety Monitoring; available at: http://www.biostat.harvard.edu/events/schering-plough/old/agenda2004-05.html; accessed on April 2009. Szarfman, A., Levine, J. G., Tonning, J. M., Use of Advanced Computer Methods to Simplify the Analysis of Complex Clinical Drug Safety Data; http://webapps.sph.harvard.edu/content/Sch-Plo6205III_Unspecified_2005-06-02_04-03-PM.htm (3rd presentation in the Video Clip); Accessed April 2009.
7. Barnett, S. T., and James, J. A. (1995), Measuring the clinical development process, *Appl. Clin. Trials.*, 4, 44–52.
8. Szarfman, A., Talarico, L., and Levine, J. G. (1997), Analysis and risk assessment of hematological data from clinical trials: toxicology of the hematopoietic system, in Sipes, I. G., McQueen, C. A., Gandolfi, A. J., Eds., *Comprehensive Toxicology*, Vol. 4, Elsevier Science, New York; pp. 363–379.
9. Szarfman, A., Machado, S. G., and O'Neill, R. T. (2002), Use of screening algorithms and computer systems to efficiently signal higher-than-expected combinations of drugs and events in the U.S. FDA's spontaneous reports database, *Drug Saf.*, 25, 381–392.
10. Evelyn, B., Toigo, T., Banks, D., et al. (2001), *Women's Participation in Clinical Trials and Gender-Related Labeling: A Review of New Molecular Entities Approved 1995–1999*, June, Office of Special Health Issues, Office of International and Constituent Relations, Office of the Commissioner, U.S. Food and Drug Administration; available at: http://www.fda.gov/cder/reports/womens_health/women_clin_trials.htm; accessed April 2009.

11. Cooper, C. K., Levine, J. G., Tonning, J. M., et al. (2005), Use of standards-based data and tools to improve the efficiency of the NDA Safety Review. 2005 FDA Science Forum. Abstract and Poster H-03; available at: http://www.accessdata.fda.gov/scripts/oc/scienceforum/sf2005/Search/preview.cfm?abstract_id=339&backto=author; accessed April 2009.
12. Reviewer Guidance Conducting a Clinical Safety Review of a New Product Application and Preparing a Report on the Review. U.S. Department of Health 13a., and Human Services. Food and Drug Administration, Center for Drug Evaluation and Research (CDER) February 2005; available at: http://www.fda.gov/cder/guidance/3580fnl.htm.
13. DuMouchel, W., and Pregibon, D. (2001), Empirical Bayes screening for multi-item associations, in *Seventh ACM SigKDD International Conference on Knowledge Discovery and Data Mining*, ACM Press, San Francisco.
14. DuMouchel, W. Manuscript in preparation.
15. Szarfman, A., DuMouchel, W., Fram, D., et al. (2005), Lactic acidosis: Unraveling the individual toxicities of drugs used in HIV and diabetes polytherapy by hierarchical Bayesian logistic regression data mining (abstract), 11th Annual FDA Science Forum, April 27–28, 2005; available at: http://www.accessdata.fda.gov/scripts/oc/scienceforum/sf2005/Search/preview.cfm?abstract_id=483&backto=author; accessed April 2009.
16. Gould, L. (2003), *Pharmacoepidemiol Drug Saf.*, 12, 559–574.
17. Workshop and Video Clip. Statistical issues in drug safety monitoring. Schering—Plough Workshop of June 2–3, 2005. Harvard School of Public Health. Section III: Pharmacovigilence and Datamining for Drug Safety Monitoring; available at: http://www.biostat.harvard.edu/events/schering-plough/old/agenda2004-05.html; accessed on April 2009. DuMouchel W. "Empirical Bayes Methods for Estimation of Adverse Event Rates in Clinical Trials and Active Surveillance"; available at: http://webapps.sph.harvard.edu/content/Sch-Plo6205III_Unspecified_2005-06-02_04-03-PM.htm (2nd presentation in the Video Clip); Accessed April 2009.
18. http://www.gprd.com/home/; accessed April 2009.
19. http://www.nttc.edu/resources/funding/dod/sbir2003/osd032.htm; accessed March 2006.
20. Levine, J. G., Tonning, J. M., and Szarfman, A. (2006), Reply: The evaluation of data mining methods for the simultaneous and systematic detection of safety signals in large databases: lessons to be learned, *Br. J. Clin. Pharmacol.*, 61, 105–113.
21. Lazarou, J., Pomeranz, B. H., and Corey, P. N. (1998), Incidence of adverse drug reactions in hospitalized patients: A meta-analysis of prospective studies, *JAMA*, 279(15), 1200–1205.
22. Centers for Disease Control and Prevention (CDC). (1996), Fatalities associated with ingestion of diethylene glycol-contaminated glycerin used to manufacture acetaminophen syrup—Haiti, November 1995–June 1996, *MMWR Morb. Mortal. Wkly. Rep.*, 45, 649–650.
23. Ferrari, L. A., and Giannuzzi, L. (2005), Clinical parameters, postmortem analysis and estimation of lethal dose in victims of a massive intoxication with diethylene glycol. *Forensic Sci. Int.*, 153, 45–51.
24. FDA Talk Paper (2001), Bayer Voluntarily Withdraws Baycol, FDA talk paper no T01-34, Aug 8.

# *Clinical Trials in Interventional Cardiology: Focus on XIENCE V Drug-Eluting Stent*

**J. Doostzadeh, S. Bezenek, W.-F. Cheong, P. Sood, L. Schwartz, and K. Sudhir**
*Clinical Science Department, Abbott Vascular Inc., Santa Clara, California*

**Contents**

1 Interventional Cardiology: History and Current Practice
2 Coronary Stent Design
  2.1 Bare-Metal Stents
  2.2 Drug-Eluting Stents
3 Appropriate Control Groups for PCI Clinical Trials
  3.1 Coronary Artery Bypass Graft Surgery
  3.2 Drug Therapy
  3.3 Plain Old Balloon Angioplasty
  3.4 Stenting
4 PCI Study Design Endpoints
  4.1 Clinical Endpoints
  4.2 Angiographic Endpoints
5 Safety
6 Postmarket Registries and Off-Label Stent Use
  6.1 Real-World Issues
  6.2 Characterizing Off-Label Stent Use
7 Peripheral Stents

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

8 Future Technology
8.1 Future Stents
8.2 Cardiac Tissue
8.3 Genetic Advances
9 Conclusions
References

## 1 INTERVENTIONAL CARDIOLOGY: HISTORY AND CURRENT PRACTICE

Coronary artery disease (CAD) affects approximately 13 million individuals in the United States and is considered the number-one killer among both men and women. It is estimated that more than 3 million individuals worldwide suffer myocardial infarctions (MI) annually because of CAD.

Remarkable changes have taken place in interventional cardiology since 1977, when Gruentzig and co-workers [1] performed a novel technique using an expandable balloon to compress obstructive lesions and reduce myocardial ischemia without a surgical procedure. Though the practice of using balloon catheters proved quite successful in the short term, long-term prognosis was less promising because of restenosis, which occurred in ≥30% of patients [2, 3]. This prompted the development of new techniques to remove plaque and the development of mechanical adjuncts, or stents, in order to maintain lumen patency following balloon angioplasty. Limitations in maintaining arterial patency have driven further development of interventional devices such as atherectomy catheters, stents, and laser therapy [4]. Coronary stenting after percutaneous transluminal coronary angioplasty (PTCA) was found in clinical trials to be a promising approach to reducing restenosis compared with angioplasty alone [5, 6]. Bare-metal stents (BMS) were the first type of stents used in percutaneous coronary intervention (PCI). Bare-metal stents reduced angiographic and clinical restenosis rates in de novo lesions compared to PTCA alone and decreased emergency coronary artery bypass graft (CABG) surgery. Stenting became quite prevalent in Europe in the 1980s and was increasingly used as the first line of PCI in the treatment of CAD. Methods developed in the 1990s to reduce or eliminate side effects of stenting included the administration of anticoagulants, intravascular ultrasound, and brachytherapy. Systemic anticoagulation proved very effective and greatly reduced thrombus-associated complications. Brachytherapy was also effective at preventing in-stent restenosis, but its unfavorable side effects such as inhibition of healing around the stent and increased risk of cancer made it unsuitable for widespread use. The shortcomings of brachytherapy prompted further development of technology to maintain long-term arterial patency. This led to the emergence of stents designed to deliver drugs locally in order to inhibit neointimal hyperplasia without the serious effects of radiation or systemic drug administration. These coated, or drug-eluting, stents (DES) used various drugs encapsulated in different polymeric and nonpolymeric formulations. Drug-eluting stents were found to greatly reduce restenosis and maintain arterial patency. These advantages, combined with a lower incidence of side effects and a lower cost compared to surgical interventions such as bypass grafts, makes DES a great option for treatment of CAD [7].

Since the invention of stent devices, many clinical trials with various endpoints, sample size, and complexity have been designed. Device trials are designed to establish the safety and efficacy of the device by achieving a statistically significant difference in defined endpoints. The processes for receiving U.S. FDA approval for BMS and DES have required the submission of a pre-market approval (PMA) application; FDA requirements are evolving as stent technology has advanced. As shown in the Table 1, in clinical trials for BMS approved by the FDA, the sample size, total study duration, and clinical follow-up were lower compared to contemporary trials with DES. The primary objectives required by the FDA in the design of first-generation DES were surrogate endpoints, namely, the angiographic and intravascular ultrasound measures of the lumen diameter and the volume of plaque within the lumen. As a continuous and dichotomous variable that can demonstrate a statistically significant difference in small sample sizes, late luminal loss is a favored endpoint; in addition, its value in predicting target lesion revascularization makes it an excellent surrogate. However, the FDA has currently emphasized the need to demonstrate long-term safety of coronary stents and is shifting the focus of trials from surrogate endpoints [late loss (LL), percent diameter stenosis (%DS), and neointimal hyperplasia (NIH)] to clinical endpoints [cardiac death, MI, stent thrombosis (ST), major adverse cardiac events (MACE), and target vessel failure (TVF)]. Because of this shift, the clinical trial sample sizes for evaluating current DES are becoming larger than previously designed clinical trials. Current clinical trial designs for DES are mostly randomized, active-controlled, single-blind, parallel, two-arm, multicenter trials with clinical endpoints at 1 year and usually compare (newer) DES to (older) DES. The clinical follow-up should be designed for long-term safety (5 years) with angiographic, intravascular ultrasound (IVUS) and clinical endpoints. In addition to the complexity of procedural testing and follow-up duration, prespecified statistical hypothesis testing should be proposed for primary and secondary endpoints.

Another additional FDA requirement for the DES is evidence of continued safety in postmarket trials for the rare events such as ST. The rationale for this decision originated following the FDA warning notification of low-frequency event rates of ST involving the CYPHER stent in 2003 [8]. Current postmarket surveillance plans evaluate a variety of clinical endpoints, including Academic Research Consortium (ARC)-defined ST, as endpoints.

In conclusion, comparison of trial designs between BMS and DES reflect the increasing expectations for PMA approval by the FDA. DES clinical trials have increased in complexity in terms of trial design, study endpoints, hypothesis testing, statistical analysis, randomization, and sample size over the past 4 years. As coronary interventional therapy continues to evolve, efforts are made between the sponsor and the FDA to work collaboratively to generate evidence of efficacy and reasonable assurance of safety.

## 2 CORONARY STENT DESIGN

Stent design affects many properties of the stent, such as elastic recoil and rigidity—two undesirable stent characteristics [9, 10]. Previous studies have shown that differences in stent design and geometry can have a profound influence on late lumen loss [11] and neointimal proliferation [12], which in turn affect restenosis

**TABLE 1 Clinical Trial Summary for Bare-Metal and Drug-Eluting Stents**

| Trial Name | Design[a] | Trial Type | Subjects | Duration | Primary Endpoints |
|---|---|---|---|---|---|
| | | *Selected Bare-Metal Stent Trials* | | | |
| DUET | SA, OL, MC | Registry—PMA | 320[b] | 6 months | MACE at 1 month |
| PIXEL | SA, OL, MC | Registry—PMA | 150 | 6 months | TVF at 30 days |
| ASCENT | R, DA, MC | Pivotal—PMA | 1,390 | 1 year | TVF at 9 months |
| ASCENT Registry | SA, OL, MC | Postmarket | 546 | 5 years | TVF annual rate |
| VISION | SA, OL, MC | Registry—PMA | 265 | 1 year | TVF at 180 days |
| | | *Selected Drug-Eluting Stent Trials* | | | |
| SPIRIT I (everolimus) | R, MC | Feasibility | 60 | 5 years | In-stent late loss at 180 days |
| SPIRIT II (everolimus) | R, MC | Preapproval | 300 | 5 years | In-stent late loss at 180 days |
| SPIRIT III (everolimus) | R, MC | Pivotal | 1,002 | 5 years | In-stent late loss at 8 months |
| SPIRIT USA (everolimus) (pending) | OL, SA, MC | Postmarket | 5,000 | 5 years | Stent thrombosis rates annually through 5 years as defined (ARC) and composite rate of cardiac death and any myocardial infarction (MI) at 1 year |
| RAVEL (sirolimus) | R, MC | Feasibility | 238 | 1 year | In-stent late loss at 6 months |
| SIRIUS (sirolimus) | R, MC | Pivotal | 1,058 | 5 years | TVF at 9 months |
| CYPHERTM (sirolimus) | OL, SA, MC | Postmarket | 2,070 | 1 year | MACE at 30 days, 6 months, and 1 year |
| e-SELECT Registry (sirolimus) | OL, SA, MC | Postmarket | 15,000 | 3 years | Acute, subacute and late stent thrombosis and MACE at 1, 6, 12, 24, and 36 months |
| TAXUS IV (paclitaxel) | R, MC | Pivotal | 1,314 | 5 years | TVR at 9 months |
| TAXUS V (paclitaxel) | R, MC | Pivotal | 1,156 | 5 years | TVR at 9 months |
| TAXUS VI (paclitaxel) | R, MC | Pivotal | 446 | 5 years | TVR at 9 months |
| ARRIVE I (paclitaxel) | OL, SA, MC | Postmarket | 2,585 | 2 years | TAXUS stent related cardiac events at 1 year |
| ARRIVE II (paclitaxel) | OL, SA, MC | Postmarket | 5,016 | 2 years | TAXUS stent related cardiac events at 1 year |
| ENDEAVOR II (zotarolimus) | R, MC | Pivotal | 1,197 | 4 years | TVF at 9 months |
| ENDEAVOR IV (zotarolimus) | R, MC | Pivotal | 1,548 | 5 years | TVF at 9 months |
| PROTECT (zotarolimus) | R, MC | Preapproval | 8,800 | 5 years | ARC definite or probable stent thrombosis at 3 years |
| ENDEAVOR (zotarolimus) | OL, SA, MC | Postmarket | 5,300 | 5 years | ARC definite and probable stent thrombosis annually for 5 years and cardiac death/MI annually for 5 years |

[a]Abbreviations: MC = multicenter; N/R = not reported; OL = open label; R = randomized; SA = single arm; SC = single center.
[b]50 lead in subjects and 270 additional subjects enrolled.

rates and requirements for postprocedural intervention [13–16]. Vascular geometry influences distributions of wall shear stress (WSS), as evidenced by branching and curvature that produce regions of low WSS. These alterations adversely affect the preferential flow environment of intravascular cells and correlate with sites of neointimal hyperplasia [17–19]. Studies have shown that the use of a stent with thinner struts is also associated with a significant reduction of angiographic and clinical restenosis after coronary artery stenting [13, 20–22]. Although most currently approved stents are made of 316 L stainless steel [23], other materials are also used, such as cobalt chromium in the Multi-Link VISION stent (ML VISION). ML VISION is more radiopaque and elicits a less vigorous host response because of thinner struts.

Drug-eluting stents are designed to deliver a drug locally that inhibits intimal thickening by interfering with pathways involved in inflammation, migration, proliferation, and/or secretion of the extracellular matrix. Both the drug and the delivery vehicle must fulfill pharmacological, pharmacokinetic, and mechanical requirements. Besides their biological effects, the drugs' chemical properties influence pharmacokinetic parameters and the possibilities for loading on a stent [24]. Tissue levels depend on lipophilic or lipophobic characteristics, molecular weight, and the degree of protein binding of the used drug. In the FDA-approved DES, drugs used in the DES are loaded into a polymer coating, which forms a reservoir for the drug. Drugs on DES inhibit smooth muscle cell proliferation by targeting cell cycle regulators and arresting the cycle. For example, everolimus, zotarolimus, and sirolimus (the "limus" drugs) differ from paclitaxel in their mechanism of action because of the point in the cell cycle at which the dividing cells are arrested. The limus drugs block an intracellular proliferation signal, which is triggered by IL-2 and other T-cell-specific growth factors. As a result, the cells are arrested in the G1 phase of the cell cycle. Paclitaxel, however, acts later in the cell cycle than the limus drugs: It stabilizes microtubules and has potent activity against proliferation, migration, and signal transduction [25]. Preclinical and clinical data have documented that low doses of everolimus, paclitaxel, sirolimus, or zotarolimus (used, respectively, on the XIENCE V, TAXUS, CYPHER, and ENDEAVOR stents) can achieve therapeutic arterial tissue concentration with insignificant concentrations in the systemic circulation, and that the drugs are safe. The ideal drug to prevent restenosis must have antiproliferative and antimigratory effects on smooth muscle cells, but on the other hand the stent design and the drug used must also minimize the reendothelialization duration [26]. As discussed in the section on next-generation DES, the greatest and quickest reendothelization process was observed with XIENCE V when compared to TAXUS, CYPHER, and ENDEAVOR in preclinical models.

The drugs on the DES are delivered using a vehicle that must fulfill many requirements. Drug release must be controlled and predictable. The delivery vehicle (most often a polymeric carrier) must be biologically inert (e.g., nonthrombogenic, noninflammatory), sterilizable, must follow the changes in stent configuration during expansion and deployment and must be mechanically resistant to abrasion [27–29]. The development of a coating that meets all of these criteria has been challenging. Polymer coatings are needed as carriers for most drugs since these drugs do not adhere to the metallic stent surface. Numerous substances have been proposed as potential biomaterials for stent coatings [30]. Most of the polymers currently used in commercially available DES are proprietary. The safety and biocompatibility of

acrylic polymer and fluorinated copolymer coatings have been demonstrated based upon the XIENCE V everolimus-eluting coronary stent system (EECSS) biocompatibility testing per ISO 10993-1 and the long-term use of these polymers in medical implants.

Ideally, stent design should offer both effectiveness and safety measures and optimize flexibility, tractability, visibility, and biocompatibility [12]. While stent flexibility and tractability are highly dependent on mechanical stent design, radiological visibility and host biocompatibility rely on stent material.

### 2.1 Bare-Metal Stents

The concept of the stent grew directly out of interventional cardiologists' experience with angioplasty balloons in the first decade of use (1977–1987). The rates of restenosis with angioplasty balloons were of concern. Although the artery would be opened successfully using a balloon, in a small percentage of cases, the artery would collapse after the balloon was deflated. BMS stents, inserted after balloon angioplasty, were designed to address the issues met with balloon angioplasty. The stent itself was mounted on a balloon and could be opened once inside the coronary artery. In 1986, in Toulouse, France, Jacques Puel and Ulrich Sigwart inserted the first stent into a human coronary artery. In 1994 the first Palmaz–Schatz stent was approved for use in the United States. Over the next decade, several generations of BMS were developed, with each successive stent being more flexible and easier to deliver. Despite the fact that BMS stents virtually eliminated many of the complications of abrupt artery closure, restenosis typically after 6 months occurred in about 20% of cases, necessitating repeat procedures.

Current BMSs have emerged with advanced design characteristics imparting improved performance compared to prior generations. Stent characteristics of potential importance to procedural performance include metallic alloy content, stent crossing profile, strut thickness and configuration, flexibility, conformability, polish and texture, cell structure, radial support, scaffolding, side branch accessibility, fluoroscopic visibility, delivery balloon technology, and vessel surface coverage. Some of these features also aid in reducing long-term in-stent restenosis (ISR). The importance of strut thickness in reduction of ISR was studied prospectively in the ISAR-STEREO [31] and ISAR-STEREO-2 [21] trials, which compared angiographic and clinical restenosis rates between thick- and thin-strut stents, showed lower ISR rates by approximately 40% with thin-strut stents, with either similar or different strut designs. Similar findings were noted in a retrospective review by Briguori and co-workers [20]. Fluoroscopic visibility and metallic alloy composition are also related to strut thickness. Stainless steel stents become difficult to visualize as strut thickness approaches 0.05 mm. Cobalt alloy stents have superior visibility, as well as improved deliverability, radial strength, and flexibility. Several studies have evaluated BMS versus DES, but proving differences in clinical outcomes between DES and BMS has been problematic because of strut thickness. Speculation has been raised about the thick-strut bare-metal controls used in DES trials. For example, the BASKET trial enrolled 826 patients with 1281 treated lesions. In the BASKET trial patients received either CYPHER or TAXUS in the DES arm or VISION in BMS arm. The study found no difference at 18 months regarding cardiac death/MI between treatments. Interestingly, the BMS stent used in this trial

was the thin-strut VISION stent, which is also used in the everolimus drug-eluting stent, XIENCE V. As mentioned above, studies indicated that strut thickness is important in reducing the ISR [21, 31].

Several recent clinical studies [20–22] have been conducted comparing the performance of stents with varying strut thickness. These studies indicated that when two stents with different designs are compared, the stent with thinner struts elicits less angiographic and clinical restenosis than the thicker strut stent. The VISION stent system is a next-generation BMS that can achieve good acute performance (deliverability, visibility) with thinner strut technology. Because of excellent results obtained with the VISION stent, Abbott Vascular has used the VISION stent as the platform in the XIENCE V. The ML VISION family of stents is fabricated from a L605 cobalt chromium alloy that has a thinner strut stent (0.08 mm or 0.0032 inch) and has a crossing profile of 1.07 mm. The ML VISION stent has been on the market in the European Union since 2002 and the MINI VISION since 2003. The ML VISION and MINI VISION stents have also been on the market in the United States since 2004. Medtronic's Driver cobalt nickel stent, with a smooth, modular design, a strut thickness of 0.09 mm, and crossing profile of 1.09 mm is another future generation BMS. Registry data for the Driver [32] and VISION [33] stents in mostly type B1 and B2 lesions in 3.0–4.0 mm diameter (mean 3.04–3.07 mm) vessels document excellent (100%) device success rates, and remarkably low ischemia-driven target lesion revascularization (TLR) rates at 6 months (4.3% for the VISION and 3.4% for the Driver). Angiographic in-stent restenosis (ISR) rates were 15.7% for both stents at 6 months.

In summary, the latest available BMSs have advanced design characteristics related to strut thickness and cell configuration, metal alloy composition, flexibility, crossing profile, and resultant deliverability. The next generation of DES (e.g., XIENCE V) benefits from the use of these thinner strut BMS platforms. The future generation of BMS may incorporate biodegradable materials, such as magnesium alloys, which may reduce the TLR rates.

### 2.2 Drug-Eluting Stents

The newest minimally invasive intervention alternative to open-heart bypass surgery is PCI with a DES. The first DES, a sirolimus-eluting stent, was approved by the FDA in 2003. The introduction of DES to interventional cardiology practice has resulted in significant improvement in the long-term efficacy of PCI. DES successfully combine mechanical benefits of BMS and stabilizing the lumen with direct delivery and the controlled elution of a drug to the injured vessel wall to reduce further the neointimal proliferation. The dramatic reduction in restenosis has resulted in the implantation of DES in the clinical practice and has rapidly expanded the spectrum of new DES generations [34].

***First-Generation DES*** The first-generation DESs, which combine the mechanical scaffolding properties of metallic stents with the site-specific delivery of an antiproliferative agent, have proved to inhibit vascular responses to arterial injury and reduce restenosis. The polymer-regulated site-specific delivery of an antiproliferative agent has also been shown to inhibit tissue growth after coronary stent implantation and to improve long-term event-free survival compared with BMS [35, 36].

The first-generation DES, CYPHER, was evaluated in the early sirolimus trials RAVEL ($n = 238$), E-SIRIUS ($n = 352$), C-SIRIUS ($n = 100$), and SIRIUS ($n = 1058$) before FDA approval. These trials reported a lower incidence of restenosis in patients who had received sirolimus DES, up to 8 months after PCI [37–39]. The attenuation of neointimal hyperplasia in these trials correlated with fewer TLR, an important clinical marker of restenosis. Rates of death and MI were low and comparable between study arms. These observations were confirmed in the U.S. pivotal trial, SIRIUS. Rates of death and MI, at 9 months, were, respectively, 0.9 and 2.8% for the CYPHER arm compared to 0.6 and 3.2% in the TAXUS arm.

The first-generation DES, TAXUS, was evaluated in TAXUS I ($n = 61$), TAXUS II ($n = 536$), TAXUS IV ($n = 1314$), TAXUS V ($n = 1156$), and TAXUS VI ($n = 446$) trials. These trials showed reductions for the incidence of restenosis and TLR in patients receiving a slow-release polymeric paclitaxel DES compared to BMS [40]. These trends were later confirmed for slow-release and moderate-release formulations by the TAXUS II trial [41]. The pivotal trial, TAXUS IV, showed that the slow-release polymeric paclitaxel DES was associated with significantly reduced rates of in-stent binary restenosis (5.5% compared to 24.4%) and TLR (3.8% compared to 14.6%) at 9 months [42]. Furthermore, no significant adverse events were experienced in DES-treated patients in these early trials. The objective of TAXUS V was to evaluate the SR polymeric paclitaxel DES in complex lesions [43]. Finally, TAXUS VI evaluated a moderate-release polymeric paclitaxel DES, also in complex lesions [44]. Consistent with earlier TAXUS trials, the polymeric paclitaxel DES demonstrated significant reductions in the incidence of restenosis and TLR (72 and 64%, respectively) compared to metallic stents.

The FDA approved a new DES, ENDEAVOR. The ENDEAVOR IV [45] trial compared the performance of ENDEAVOR to TAXUS. Results from the ENDEAVOR IV trial demonstrated that in-stent late-loss, in-segment late-loss, and in-segment angiographic binary restenosis (ABR), respectively, were 60, 57, and 47% higher in the ENDEAVOR arm compared to the TAXUS arm. TLR was 41% higher in the ENDEAVOR compared to TAXUS with comparable MACE and TVF rates between the two arms (Table 3).

Abbott Vascular evaluated first-generation DES in the ACTION, DELIVER, FUTURE, and ZoMaxx clinical trials prior to developing the next generation of DES, XIENCE V. The ACTION trial was a randomized, parallel, three-arm, single-blind, multicenter trial designed to evaluate the safety and performance of the MULTI-LINK TETRA-D, actinomycin-eluting coronary stent system at two different drug concentrations compared to a bare-metal control. Early results of the ACTION trial indicated higher target lesion revascularization rates than expected in DES arms, driven by restenotic events. The ACTION trial was terminated in March 2002.

The DELIVER trial, was a prospective, randomized, single-blinded, parallel-group, multicenter clinical investigation evaluating the RX ACHIEVE paclitaxel drug-coated coronary stent system (CCSS) compared to the bare-metal MULTI-LINK RX PENTA stent in the treatment of patients with de novo native coronary artery lesions. The primary endpoint was target vessel failure (TVF) at 270 days. The primary endpoint was not met as powered; superiority of the RX ACHIEVE paclitaxel-eluting CCSS was not met.

The ZoMaxx DES has a phosphorylcholine coating and releases the drug ABT-578 (zotarolimus), a sirolimus derivative. ZoMaxx I and ZoMaxx II assessed the safety and efficacy of the ZoMaxx stent system compared with the TAXUS Express [2] paclitaxel-eluting stent system (Boston Scientific, Natick, Massachusetts). Although these trials demonstrated that ZoMaxx is safe to be implanted, Abbott Vascular decided not to pursue further the development of the ZoMaxx stent since a next-generation DES by Abbott Vascular, XIENCE V, had demonstrated superiority in terms of late loss to both bare-metal stent and an FDA-approved DES, TAXUS at 8 months, in SPIRIT III (see next section).

In conclusion, clinical trials for first-generation DES have proved the effectiveness of the DES in reducing the rate of ISR by directly inhibiting the progression of the cell cycle and migration of vascular smooth muscle cells, preventing neointimal hyperplasia and injury-induced thickening of the arterial wall. These trials evaluating the first-generation DES were randomized trials with more complex design, and a larger sample size when compared to pivotal studies for approval of BMS by FDA. Surrogate primary endpoints such as angiographic late loss were permitted. However, the search for improved deliverability and greater efficacy has prompted development of the next generation of DES.

***Next-Generation DES*** Next-generation DESs are being designed with the goal of enhanced safety and/or efficacy compared to first-generation devices. The Abbott Vascular DES XIENCE V EECS was approved in the U.S. on July 2, 2008. This is a next-generation DES, which has demonstrated superior clinical outcomes compared to TAXUS. XIENCE V, an everolimus-eluting stent, has been designed so that the drug is released from a thin (7.8 μm), nonadhesive, durable, biocompatible fluorinated copolymer coated onto a low profile, flexible cobalt chromium stent with a 0.0032-inch strut thickness. Several clinical studies have demonstrated that stents with thinner struts elicit less angiographic and clinical restenosis than the thicker strut stent [20–22]. The strut thickness of the ML VISION (81 μm) compares favorably to those of the metallic stent platforms used on the CYPHER (140 μm) and the TAXUS (132 μm) stents [22] (Fig. 1). The thin-strut design is one element of the XIENCE V EECS (based on the ML VISION stent design) that decreases clinical risk.

| CYPHER | TAXUS EXPRESS | ENDEAVOR | XIENCE V |
|---|---|---|---|
| Strut Thickness: **140 μm** | Strut Thickness: **132 μm** | Strut Thickness: **91 μm** | Strut Thickness: **81 μm** |
| Coating Thickness: **12.6 μm** | Coating Thickness: **19.6 μm** | Coating Thickness: **4.8 μm** | Coating Thickness: **7.8 μm** |

***FIGURE 1*** *Progression towards thinner struts in XIENCE. Albluminal coating thickness is represented in all images in figure. Data on file at Abbott Vascular.*

The XIENCE V stent uses the same stent design and L-605 alloy cobalt chromium material as the ML VISION and ML MINI VISION stents. Compared to the 316 L stainless steel used in TAXUS and CYPHER [23], the cobalt chromium in the XIENCE V stent is more radiopaque, thinner, and elicits a less vigorous host response than stainless steel. L-605 cobalt chromium provides both excellent strength and fatigue resistance characteristics, which enables the XIENCE V stent to have significantly thinner struts than contemporary stainless steel stents while maintaining comparable radial strength. Thinner struts, cobalt chromium material, and the MULTI-LINK stent pattern were intended for increased flexibility and conformability, and therefore better deliverability.

More rapid endothelialization with XIENCE V in porcine model, which may be due to the thin struts, has also been reported in preclinical models (Fig. 2). Evaluations of scanning electron micrographs of the luminal surface of various DES demonstrated differences in endothelial cell coverage among DES. As shown in Figure 2, XIENCE V stents had the greatest coverage compared to CYPHER, TAXUS, and Endeavor. The faster reendothelialization associated with the XIENCE V is likely due, at least in part, to the thin stent struts of the XIENCE V. A thin stent strut is incorporated in neointima more rapidly and will require less neointima to completely cover the struts. Similarly, endothelial cells may migrate up and over to incorporate a thin strut more rapidly [46].

Following favorable results with XIENCE V in the SPIRIT FIRST (n = 60) and SPIRIT II ($n$ = 300) clinical trials in Europe [47, 48], the large-scale SPIRIT III trial ($n$ = 1002) was performed to evaluate the XIENCE V in comparison to a widely used TAXUS in patients with CAD [49]. The XIENCE V clinical trials (SPIRIT FIRST, SPIRIT II, and SPIRIT III) were conducted in single- and dual-vessel disease and met their primary and major secondary endpoints with no adverse safety signals to date. The results of these clinical trials are provided in Table 2. As shown in Table 2, the SPIRIT family trials showed significant benefits to using XIENCE V over both commercially available BMS (VISION) and DES (TAXUS).

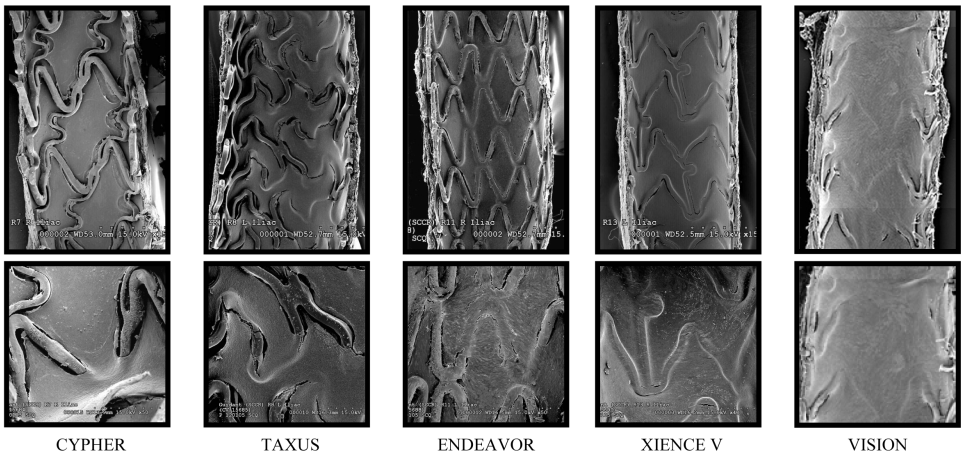


***FIGURE 2*** *Qualitative assessment of endothelial cell coverage (14-day rabbit iliac). Corresponding images of each stent in the top row are shown directly below in greater magnification of the above stent image.*

**TABLE 2 In-Stent Late Loss, In-Segment Late Loss, TVF, MACE, and Cardiac + MI across Spirit Trials**

| | | | Spirit II | | Spirit III | | Spirit III | Combined Spirit II & III RCT | |
|---|---|---|---|---|---|---|---|---|---|
| | Time Point[a] | Spirit First | XIENCE V | TAXUS | XIENCE V | TAXUS | 4.0 mm | XIENCE V | TAXUS |
| In-stent late loss (analysis lesion) | 6M | 0.10 ± 0.23 (23) | 0.10 ± 0.27 (201) | 0.36 ± 0.39 (73) | Not available | Not available | Not available | Not available | Not available |
| MACE | 6M | 7.7% (2/26) | 2.7% (6/222) | 6.5% (5/77) | 2.6% (17/663) | 4.6% (15/326) | 5.9% (4/68) | Not available | Not available |
| Cardiac death + MI | 6M | 3.8% (1/26) | 0.9% (2/222) | 3.9% (3/77) | 1.5% (10/663) | 3.1% (10/326) | 5.9% (4/68) | Not available | Not available |
| TVF | 6M | 7.7% (2/26) | 3.6% (8/222) | 6.5% (5/77) | 3.8% (25/663) | 4.9% (16/326) | 5.9% (4/68) | Not available | Not available |
| In-stent late loss (analysis lesion) | 6M or 8M | 0.10 ± 0.23[b] | 0.11 ± 0.27[b] | 0.36 ± 0.39[b] | 0.16 ± 0.41[c] (301) | 0.31 ± 0.55[c] (134) | 0.12 ± 0.34[c] (49) | Not available | Not available |
| In-segment late loss (analysis lesion) | 6M or 8M | 0.09 ± 0.20 | Not available | Not available | 0.14 ± 0.41[c] (301) | 0.28 ± 0.48[c] (134) | 0.17 ± 0.38[c] (49) | Not available | Not available |
| MACE | 9M | 7.7% (2/26) | 2.7% (6/220) | 6.6% (5/76) | 5.0% (33/657) | 8.8% (28/320) | 5.9% (4/68) | 4.4% (39/877) | 8.6% (34/397) |
| Cardiac death + MI | 9M | 3.8% (1/26) | 0.9% (2/220) | 3.9% (3/76) | 2.9% (19/657) | 3.8% (12/320) | 5.9% (4/68) | 2.4% (21/877) | 3.8% (15/397) |
| TVF | 9M | 7.7% (2/26) | 4.5% (10/220) | 6.6% (5/76) | 7.6% (50/657) | 9.7% (31/320) | 5.9% (4/68) | 6.8% (60/877) | 9.3% (37/397) |
| MACE | 12M | 15.4% (4/26) | 2.7% (6/220) | 9.2% (7/76) | 6.0% (39/653) | 10.3% (33/320) | 5.9% (4/68) | 5.3% (46/873) | 10.1% (40/397) |
| Cardiac death + MI | 12M | 7.7% (2/26) | 0.9% (2/220) | 3.9% (3/76) | 3.4% (22/653) | 4.7% (15/320) | 5.9% (4/68) | 2.7% (24/873) | 4.5% (18/397) |
| TVF | 12M | 15.4% (4/26) | 4.5% (10/220) | 9.2% (7/76) | 8.6% (56/653) | 11.3% (36/320) | 5.9% (4/68) | 7.7% (67/873) | 10.8% (43/397) |

[a]All Spirit First subjects were treated with single, de novo, native coronary artery lesion. 6M = 194 days for Spirit First, Spirit II, Spirit III and Combined Spirit II & III. 8M = 240 days for Spirit II, Spirit III and Combined Spirit II & III. 9M = 270 days for Spirit II, 284 days for Spirit First, Spirit III and Combined Spirit II & III. 12M = 365 days for Spirit II, 393 days for Spirit First, Spirit III and Combined Spirit II & III. Both TAXUS Express[2] (73% of lesions) and TAXUS Liberte (27% of lesions) were used as controls in Spirit II. TAXUS Express[2] was used as the control in Spirit III.
*Source*: Data on file at Abbott Vascular.
[b]Results are from 6 months.
[c]Results are from 8 months.

**TABLE 3 Perspectives from the New DES Compared to TAXUS**

| Study | Stent | In-stent LL | In-seg. LL | In-seg. ABR | TLR | TVR | MACE | TVF |
|---|---|---|---|---|---|---|---|---|
| ***SPIRIT* III**[a] | ***XIENCE V vs.*TAXUS** | ↓ 48% | ↓ 50% | ↓ 47% | ↓ 46% | ↓ 29% | ↓ 43% | ↓ 22% |
| ***ENDEAVOR* IV**[b] | ***ENDEAVOR vs.*TAXUS** | ↑ 60% | ↑ 57% | ↑ 47% | ↑ 41% | ↑ 6% | ↓ 2% | ↓ 8% |

[a]Clinical event rates for XIENCE V calculated from 284 day data.
[b]ENDEAVOR IV results presented by M. Leon at TCT, 2007.

XIENCE V, as compared to TAXUS in the SPIRIT III trial, demonstrated (a) significant reductions in angiographic in-stent late loss ($p = 0.006$), in-segment late loss ($p = 0.004$), and in-segment %DS ($p = 0.009$) at 8 months, (b) significant reduction in IVUS percent volume obstruction ($p = 0.01$) without positive remodeling or late acquired incomplete apposition at 8 months, (c) nonsignificant trends toward less events for composite endpoints cardiac death and MI, and TVF at one year, (d) reductions in TLR and MACE rates at one year, and (e) comparable rates of thrombosis [49]. In the SPIRIT II and SPIRIT III RCT clinical studies, the XIENCE V has demonstrated not only noninferiority but superiority to TAXUS, in terms of the primary endpoints, in-stent and in-segment late loss, respectively. The SPIRIT IV clinical trial is a continued access study for continued access and was initiated in 2006 to further evaluate the safety and effectiveness of the XIENCE V EECSS and to enroll SPIRIT-III-like subjects in order to support the SPIRIT III major secondary endpoint (270-day TVF). The SPIRIT IV clinical trial is a single-blinded, multicenter clinical trial and will enroll approximately 3690 subjects at up to 70 sites in the United States. The primary endpoint of the trial is ischemia-driven MACE at 270 days. Patients will be followed out to 5 years. This study has enrolled approximately 3000 patients as of March 2008. SPIRIT IV will allow for analyzing subgroups of complex patients in a large randomized clinical trial. Furthermore, subjects enrolled in SPIRIT IV are SPIRIT-III-like subjects, so a true pooled analysis can be performed. In addition, postmarketing studies are currently being conducted internationally, which include SPIRIT V and SPIRIT WOMEN. A postapproval study, XIENCE V USA is currently also being planned in the United States. The results of a meta-analysis of SPIRIT II and SPIRIT III clinical studies obtained data at one year to support previous findings from SPIRIT II and III. The analysis indicated reductions in relative risk for both MACE and TVF, suggesting that XIENCE V has lower cardiac event rates in comparison with TAXUS, which was driven by lower TLR (XIENCE V FDA Panel, data are available at Abbott Vascular website; www.abbottvascular.com/av_dotcom/url/home/en_us). In the meta-analysis of the SPIRIT II and SPIRIT III RCT clinical studies, key clinical endpoints observed in XIENCE V arm were lower than those observed in the TAXUS arm. Results from several randomized clinical trials that compared commercially available DES and evaluated various clinical and angiographic endpoints have been published [45, 50–52]. The TAXi study ($n = 202$) demonstrated safety of both TAXUS and CYPHER stents in various clinical and anatomical subsets [50]. Similarly, the REALITY trial confirmed the clinical safety of TAXUS in a larger subject population. In the REALITY trial ($n = 1386$), the primary endpoint of in-lesion ABR rate at 8 months (primary endpoint) was comparable between TAXUS and CYPHER.

Additionally, there was no difference between the two DES in terms of MACE rates at one year [51]. In the SIRTAX trial, although sirolimus-eluting stents showed a 42.6% decrease in MACE at 9 months compared to paclitaxel-eluting stents, the rates of clinical and angiographic restenosis were low for both DES [52]. Although the first-generation DES used BMS as control stents, next-generation DES clinical trials were designed to evaluate (older) DES versus (newer) DES, which resulted in increased sample sizes and new statistical hypothesis testing. Surrogate angiographic endpoints were permitted as primary endpoints (but for longer time points) with the focus on rare safety events such as death, MI, and ST. Composite endpoints such as MACE and TVF were used in these clinical trials to evaluate both safety and effectiveness of next-generation DES.

***Future-Generation DES*** Drug-eluting stents continue to evolve, with enhanced drugs, polymers, and delivery systems designed to improve outcomes for patients with CAD. Improved delivery systems allow access to lesions in vessels with excessive tortuosity. Abbott Vascular plans to evaluate a future everolimus-eluting coronary stent (EECS), which is a balloon-expandable stent fabricated from a single piece of medical-grade L-605 cobalt chromium alloy tubing, which is the same material used in the ML VISION and XIENCE V. This EECS will provide a more flexible, better delivery system profile while maintaining good radiopacity and strength. The sirolimus and paclitaxel stents, CYPHER SELECT Plus Stent (by Cordis), and TAXUS Element Stent (by Boston Scientific) are also designed with improved delivery system features and are currently on the market in the European Union (EU). In addition, clinical trials have recently begun for a sirolimus-eluting stent (by Conor and Cordis) using Conor's unique fully bioresorbable reservoir technology.

The bioabsorbable stent is a future-generation stent type. Bioabsorbable stents are an exciting class of future-generation stents because of their ability to be absorbed over time. Ideally, this allows the lifetime of the stent to be tailored to the clinical need of the disease or condition. In order to open the blocked vessel and offer mechanical support, the bioabsorbable polymeric stent must have high radial strength, low recoil, and a certain degree of flexibility while its degradation product must be nontoxic in body fluids. The bioabsorbable stent holds promise as a device system that will be able to perform its mechanical function (i.e., high radial strength and low recoil), facilitate stent placement, and deliver drugs for the prevention of restenosis. Furthermore, the stent should support the artery while the vessel heals and should gradually transfer the mechanical load to the tissue as the stent degrades over time. It is also worth noting that radiopacity is another consideration in stent design. The BVS (Bioabsorbable Vascular Solutions) stent was developed by Abbott Vascular. In March 2006, BVS started the world's first clinical trial of drug-eluting polymeric bioabsorbable stent called the ABSORB trial.

The BVS stent drug coating layer contains everolimus (antiproliferative agent) and poly-D, L-lactic acid (PDLLA) polymer at a 1 : 1 ratio for controlled drug release. Because of its unique stent pattern design, the BVS stent looks like a regular DES with great mechanical properties. The ABSORB trial enrolled 30 patients [53]. The procedure success rate was 100% and the device success rate was 93.5%. The safety of the device was confirmed in all 30 patients after 12 months follow-up. Only one non-Q-wave MI was reported that was related to the procedure of a nonischemic-

driven target lesion revascularization. There was no ischemia-driven target lesion revascularization, and no thrombosis occurred up to 12 months. Through 12 months, there were no protocol-defined or ARC-defined stent thrombosis [54]. Efficacy measurements of in-stent late loss was 0.44 ± 0.35 mm with angiographic follow-up at 6 months and was mainly due to a mild reduction of the stent area (–11.8%) as measured by intravascular ultrasound (IVUS) [54]. At 24 months, an increase in lumen area on IVUS and optical coherence tomography (OCT), and restored vasomotion in response to vasoactive agents, has been demonstrated.

Apart from Abbott Vascular, Reva Medical Inc. (supported by Boston Scientific, Inc.) is developing its own bioabsorbable DES made from tyrosine-derived polycarbonate material and has started its clinical trial as of late 2007. Reva Medical uses a unique "slide-and-lock" geometry for its stent designs. When expanded, the stent elements slide from the compact state and lock into an expanded state, similar to safety lockouts on extension ladders. Thus the expansion is independent of material deformation and will provide steel-like scaffolding. The Reva stent is delivered by standard balloon deployment and is made radiopaque by a propriety method [55].

Although more long-term beneficial outcomes in controlled clinical studies have yet to be demonstrated, it can be anticipated that biodegradable stents offer the benefits of DES without leaving a metallic stent behind after the treatment of coronary vessel lesion.

As new approaches of stenting continue to evolve, clinical trial design requires new technology and perhaps novel endpoints to demonstrate the safety and effectiveness of these devices. For example, optical coherence tomography (OCT) and serial IVUS are used in the evaluation of biodegradable stents. More long-term safety data are required by the FDA for low-event rate endpoints resulting in larger sample size for these trials.

## 3 APPROPRIATE CONTROL GROUPS FOR PCI CLINICAL TRIALS

Coronary atherosclerotic disease has been dominated in the past by four main therapy options: CABG, optimal medical therapy (OMT), balloon angioplasty (POBA), and percutaneous coronary intervention (PCI) with BMS or DES. From 1960 to 1994, CABG surgical procedures dominated as the treatment for CAD. The CABG was eventually challenged with the introduction of angioplasty in 1977 [56]. The FDA approved the first BMS in 1996, which was followed by the first DES approval in 2003. According to the Centers for Disease Control and Prevention, the most common interventional cardiology procedure performed in 2007 was PCI with DES [57]. Several studies have compared the various treatment options of CAD, POBA, PCI, CABG, and optimal medical therapy. Despite the fact that comparative studies have been conducted, there is still the need for studies directly comparing PCI with DES to these treatments.

In the next paragraphs the key outcome of clinical trials comparing PCI to other CAD treatment options will be discussed.

### 3.1 Coronary Artery Bypass Graft Surgery

Coronary artery bypass graft procedures are a common choice for treatment of advanced CAD. In patients with single-vessel disease though, PCI is the preferred

treatment because of the lower clinical risks associated with PCI [58]. Several studies compared PCI with BMS to CABG including the ERACI II, ARTS, SOS, and the AWESOME trials [59–63]. The results of these trials demonstrated that repeat revascularizations were higher in the PCI with BMS group compared to the CABG group in all trials. However, there was no significant difference in survival benefit between CABG and PCI with BMS. Overall results showed that PCI with BMS was a relatively safe alternative for patients who were not candidates for surgery.

When DESs were initially being introduced, primary endpoints included late lumen loss on angiographic follow-up, death, myocardial infarction, or reinterventions (RAVEL, SIRIUS). Results of these trials showed significant reduction in target lesion revascularization rates for DES. Since the first FDA approval of DES in 2003, numerous studies have shown the benefit of DES for reducing the rate of revascularization [35–41, 43, 44, 64]. As it is well established that DESs out-perform BMS at reducing revascularizations, studies comparing CABG to DES now can compare the rate of repeat revascularization as primary endpoints, such as in the SYNTAX trial.

Furthermore, trials currently underway comparing CABG to DES are designed to compare the rate of cerebrovascular accidents, as CABG have shown higher rates of CVA [64a]. The SYNTAX trial has been designed to compare PCI with DES (TAXUS) to CABG for the treatment of triple-vessel and left main disease [65]. The SYNTAX trial will compare the rate of major adverse cardiac and cerebral events (MACCE) between DES (TAXUS) and surgery with CABG. The study includes two nested registries for subjects ineligible for CABG who are treated with PCI and for subjects ineligible for PCI who are treated with CABG. Complete data from the primary objective of the SYNTAX trial is not available yet, but differences in CABG and PCI practice patterns have been observed between North America and Europe. Thus far, the rate of PCI intervention for triple-vessel disease in Europe is approximately double the rate from North America [65]. Initial reports from the SYNTAX trial in late 2008 showed similar composite rates for death, stroke, and MI in both arms, but reduced repeat revascularization rates in the CABG arm and reduced stroke rate in the PCI arm.

The FREEDOM trial is another current study to investigate CABG versus DES (CYPHER) in the treatment for subjects with CAD and diabetes mellitus [66]. The primary objective will be to assess the effectiveness of PCI compared to CABG (on or off pump) in the prevention of all-cause mortality, nonfatal MI, or stroke. Outcomes of mortality, MI, revascularization, health-related quality of life, and cost effectiveness will be addressed for a mean of 4 years.

The SYNTAX and FREEDOM trials will have completed follow-up in 2012 and 2011, respectively. The SYNTAX primary endpoint results are expected to be communicated in the fall of 2008. SYNTAX and FREEDOM will provide valuable information on CABG treatment versus DES treatment, but data for other type of DES would bring more information for choosing the best treatment options for CAD patients.

### 3.2 Drug Therapy

Pharmaceutical advances in lipid-lowering drugs, antidiabetic agents, antiplatelet therapy, and antihypertensive medications in the 1990s have provided patients and

physicians with drug therapy treatment options. The MASS II trial compared PCI, CABG, and OMT modalities in subjects with stable angina and multivessel CAD [67]. The primary endpoints were total mortality, Q-wave myocardial infarction (QWMI), or refractory angina requiring revascularization. In the MASS II trial, all three therapeutic strategies yielded comparable results with relatively low rates of mortality. The rates of additional revascularization and long-term events were found to be similar for PCI and OMT, but were the lowest for CABG. Devices used for catheter-based PCI treatments included stents, lasers, directional atherectomy, and balloon angioplasty. Since enrollment was completed before DESs were approved by the FDA, no DES was included in the PCI arm of the MASS II trial. Overall, BMS compared to DES represents up to a 60% higher rate of revascularization [68]. Despite this fact for patients with stable multivessel CAD, the MASS II trial showed that routine PCI is as good as CABG or OMT. The MASS II trial results suggest that patients with mild to moderate angina are good candidates for OMT, whereas a PCI or CABG would be appropriate for patients with symptoms that could not be controlled or were more severe.

The COURAGE trial compared outcomes of subjects initially treated by OMT with PCI versus OMT alone for significant CAD. The primary endpoint in the COURAGE trial was all-cause mortality and nonfatal MI, which may well have been a limiting factor. By selecting the primary endpoint of all-cause mortality, the rate of mortality would include the rate of death from any cause and not just a death due to cardiac events. By choosing this type of endpoint, the comparison of mortality rates would include additional noise and would not provide the most accurate comparison. Because the COURAGE trial was comparing cardiac interventions, a more appropriate endpoint might have been cardiac events. In fact, a more appropriate safety endpoint would have included all MI events rather than only the nonfatal MI. Predictably, by selecting this primary endpoint, there was no significant difference in the rates of death distinguishable between the OMT and PCI versus the OMT alone [69]. The trial enrolled ≥97% of PCI subjects treated with BMS compared to 3% PCI subjects treated with DES. The very low rate of DES subjects (<3%) in the COURAGE trial does not allow any comparison of PCI with DES to OMT [56]. This observed bias is because the DES was approved by the FDA at a very late stage of patient enrollment in the COURAGE trial. An additional weakness of the COURAGE trial was that it was powered to show superiority of angioplasty versus OMT with a reduction in death and MI by 22% [56]; this has never been the clinical driver of stenting in stable patients, in whom the major benefit of interventional therapy is reduction of angina, and improved quality of life. Moreover, approximately 33.0% of the OMT patients crossed over to the PCI group at a mean time of 10.2 months for angioplasty, which suggests that OMT was not an adequate treatment. Another weakness of the COURAGE trial was that the investigators modified the definition of MI part way through the study because of low event rates [56]. Currently, no data are available from major trials for comparing PCI with DES to OMT. However, it was suggested that OMT may be preferred for those more stable subjects with less severe progression of CAD.

### 3.3 Plain Old Balloon Angioplasty

Earlier trials comparing PTCA to CABG used primary endpoints that focused on safety; the BARI trial, for example, examined all-cause mortality as a primary

endpoint [70]. The EAST trial also compared PTCA to CABG with a composite endpoint of death, QWMI, and evidence of ischemia at 3 years [71]. Overall findings from these trials did not differ significantly with primary endpoint results. However, revascularization rates were higher in PTCA subjects. Trials were then designed to compare POBA to medical therapy. RITA-II used a primary endpoint of death and definite MI rates at 5 years and showed comparable mortality rates. However, lower rates of restenosis were noted when angioplasty was used in conjunction with coronary stents, compared to using angioplasty alone [72]. The benefit of decreased restenosis led to less repeat procedures that reduced long-term cost of repeat revascularization [73]. As PCI with stenting was introduced, comparisons of POBA with PCI indicated that POBA was not as effective as PCI at reducing restenosis [5, 6]. Study designs then evolved to include composite endpoints including both survival and revascularization rates. Endpoints such as TVF and MACE were implemented. As stent technology expanded through the 1990s, its designs, materials, techniques, and deployment improved, such that stenting has dominated over balloon angioplasty over the last 15 years.

### 3.4 Stenting

The FDA first approved PCI with BMS in 1996. It was later reported that 20% of subjects treated with BMS had experienced recurrent symptoms attributed to restenosis from intimal hyperplasia [74]. Despite several studies showing angiographic and clinical improvements with these catheter-based interventions, restenosis within the BMS stent continued to be a major problem that caused repeat revascularizations [5, 6, 59, 75]. Comparison of PCI with BMS to PCI with DES showed decreased rates of revascularization with DES. Currently, more than 15 analyses have been performed to evaluate the performance of DES versus BMS that have generally shown that DESs are safe and effective [76–91]. One of the largest meta-analyses comparing 19 different trials found that restenosis and MACE were greatly reduced with DES as compared with BMS [92]. A more recent study, published in 2008, showed no significant difference between DES and BMS groups in the rate of death or MI at 1-year follow-up [93].

However, the rate of revascularizations was significantly lower in the DES group, and current studies continue to support the benefit of DES compared to BMS. Trial designs comparing DES to BMS sufficiently demonstrated that safety was comparable between DES and BMS. In addition, by using composite endpoints including revascularization rates, DES showed to be a great improvement to BMS at reducing the need for repeat revascularizations. Studies continue to show comparable safety results and improved efficacy for PCI with DES compared to PCI with BMS. Future trials with next-generation DES versus BMS should continue to investigate the rates of ST and cardiac death, as trials have already shown revascularization rates were dramatically improved with use of the first-generation DES compared to BMS.

Many clinical trials have compared DES to each other. The comparison of two first-generation DES, CYPHER versus TAXUS in the REALITY [51], SIRTAX [52], and ISAR-DIABETES [79] clinical trials did not show superiority of TAXUS over CYPHER. There was no statistically significant difference in the rate of death, MI, and revascularization. ENDEAVOR IV compared the newly approved DES, ENDEAVOR, to TAXUS, and it was concluded that ENDEAVOR was noninferior to TAXUS for the primary endpoint of TVF. The SPIRIT III trial compared the next

generation of DES, XIENCE V, to TAXUS. The trial included the use of an FDA-suggested angiographic primary endpoint of in-segment late loss, in addition to a co-primary endpoint of TVF at 9 months. The follow-up results demonstrated superiority of XIENCE V to TAXUS for the primary endpoints, with lower in-segment LL. In addition, XIENCE V had fewer MIs and less repeat target lesion revascularization procedures than TAXUS, all of which suggest that XIENCE V is safe and effective for treating patients with CAD. Additional DES trials designed after the SPIRIT III trial will probably no longer use surrogate or angiographic endpoints, as the emphasis of the FDA has shifted to clinical and safety endpoints. The SPIRIT IV trial will include the rate of MACE at one year as the primary endpoint.

## 4 PCI STUDY DESIGN ENDPOINTS

Clinical trials are designed to evaluate the safety and effectiveness of DES. Selected endpoints play pivotal roles in device approval and in their adoption for clinical use. These endpoints must serve several purposes. They must have both short- and long-term pathophysiological relevance to device performance, represent clinically meaningful events, be sufficiently defined, preferably through blinded processes, and be subjected to statistical analysis. Because of the intrinsic limitations in the ability to obtain histology, serial examinations, or other mechanistic details from human subjects, clinical endpoints for DES studies are bound to include certain arbitrary assumptions and will frequently vary across clinical trials as the result of different approaches to such assumptions [94].

### 4.1 Clinical Endpoints

Endpoints for clinical trials evaluating devices should measure safety, effectiveness, and patient well-being.

Safety endpoints represent any adverse outcome during the course of the clinical evaluation and whether or not it is specifically related to the use of the device. Safety endpoints are measured by events such as death, MI, stent thrombosis, or composite endpoints such as MACE, TVR, death, and MI. Effectiveness endpoints refer specifically to maintenance of coronary artery luminal patency, which is assessed by measuring the relief of such flow-limiting obstructions, initially through structural mechanisms and later with preservation of the luminal dimension through inhibition of neointimal hyperplasia or restenosis. Effectiveness-driven clinical endpoints are designed to assess clinically significant restenosis, assessed objectively as a requirement for ischemia-driven repeat revascularization, either of the stented segment itself, TLR, or of the stented vessel or its side branches, TVR [94].

Effectiveness endpoints refer specifically to maintenance of coronary artery luminal patency. DESs are implanted for the treatment of obstructive CAD. Their effectiveness is assessed by measuring the relief of such flow-limiting obstructions, initially through structural mechanisms and later with preservation of the luminal dimension through inhibition of neointimal hyperplasia or restenosis. Effectiveness clinical endpoints are designed to assess clinically significant restenosis, assessed objectively as a requirement for ischemia-driven repeat revascularization, either of the stented segment itself, TLR, or of the stented vessel or its side branches, TVR

[95]. Target vessel failure, defined as any TVR, death, or MI attributed to the target vessel, is an even broader metric of failed effectiveness and adjusts for the potential bias introduced when patients who die or sustain MI before the end of the TLR endpoint time are considered to be free from TLR [94].

General considerations for DES evaluation should also include cardiovascular outcomes from the patient's perspective. The outcomes of endpoints should reflect the complex interplay between device performance, revascularization strategy, secondary prevention, and key patient descriptors [96]. Both the time course and the composite selected should characterize patient well-being related to the pathophysiology of the implanted DES device and its impact on underlying CAD outcome.

The most commonly measured clinical indicator of stent efficacy is TLR, which is defined as recurrent ischemia due to angiographic restenosis within the stent or its margins necessitating repeat revascularization with either PCI or coronary artery bypass graft surgery. TLR rates are typically 30–60% lower than the corresponding binary restenosis rates, suggesting discordance between the ischemic thresholds of angiographic and clinical measures of restenosis [95, 97]. Given the low rates of TLR after BMS implantation in the relatively noncomplex lesions typically studied in stent trials, more than 2000 enrolled patients are required in randomized trials to demonstrate a clinically relevant 30% reduction in TLR with DES. Moreover, given the low-frequency clinical event rates with DES, very large potential differences between stents (the "delta") often are allowed in noninferiority trials to make comparative DES studies practical, degrading confidence that the clinical performances of two devices are indeed similar.

### 4.2 Angiographic Endpoints

To reduce the sample size required in superiority and noninferiority trials, continuous angiographic indexes of long-term stent patency such as late LL and follow-up %DS have been proposed as surrogates of TLR for use in randomized studies [35, 96, 98]. However, previous DES studies have suggested that the distribution of individual LL measures is asymmetric with a rightward skew and that the relation between LL and TLR is nonlinear [36, 99]. Recently, Pocock et al. evaluated angiographic measures as valid surrogates for TLR and cutoff values that correspond with clinical efficacy [100]. Using data from 11 randomized trials, the investigators evaluated 4 angiographic measures used as surrogates for the clinical endpoint of TLR. They concluded that all four angiographic measures, instent and in-segment LL, as well as both assessments of %DS, provided a reliable estimation of TLR rates in all the trials and met the accepted criteria for surrogate endpoints. In the same report it was also concluded that %DS may have an advantage over LL as a surrogate endpoint. Therefore, Pocock et al. [100] suggested that angiographic surrogates can be used as the primary outcome measure in smaller patient populations compared to when TLR is the primary outcome, and larger trials should provide information on less frequent safety events.

## 5 SAFETY

Clinically, safety after stenting can be measured by the rate of death, MI, or stent thrombosis or by the rate of composite endpoints such as TVF (composite of death,

MI, TVR) and MACE (composite of death, MI, TLR). Composite endpoints generated by the combination of individual endpoints provide additional statistical power to detect potentially meaningful differences between treatments. The individual components should each represent clinically meaningful events and should be linked by common elements of pathophysiology. The ARC consensus suggests two composite endpoints for DES trials, one that is deviceoriented and one for overall patient-oriented clinical outcome. The device-oriented composite includes cardiac death, MI attributed to the target vessel, and TLR. The broader patient-oriented outcome composite includes all-cause mortality, any MI, and any revascularization (includes TLR), target vessel revascularization, or revascularization of nontarget vessels.

Many trials have evaluated the rates of ST after the PCI with DES versus BMS. However, there are some challenges to comparing ST rates between trials because of the differences in patient and lesion characteristics, the prevalence of diabetes mellitus (which ranged from 14 to 31%), the use of glycoprotein IIb/IIIa inhibitors (which ranged from 0 to 64%), and the use of various definitions for ST in trials performed up until 2006. Recently, the FDA, academia, and industry agreed to use ARC definitions in order to align and standardize the definitions used in clinical trials [94]. However, more than 15 studies evaluated the performance of DES versus BMS. There was no significant and consistent pattern of difference reported between the ST rates of DES and BMS. Meta-analyses of randomized trials performed with DES demonstrated that incidence of ST was not increased in patients receiving DES (0.58% for DES versus 0.54% for BMS) [86]. The overall rate of ST did not differ significantly between patients receiving sirolimus- or paclitaxel-eluting stents (0.57 versus 0.58%). In contrast, a significant relationship between the rate of ST and the stented length was found. In patients with DES, mean stented length was longer in those with ST. Per ARC definitions, the low incidence of ST was reported in the SPIRIT III trial and ST rates were low and comparable through one year for XIENCE V and TAXUS. Premature discontinuation of antiplatelet therapy has been identified as one of the predictors for ST [80]. Recent studies have reported a low incidence of DES thrombosis under prolonged therapy with aspirin plus thienopyridines, comparable to that of BMS, even in unstable clinical settings [101]. The most striking factor associated with DES thrombosis is the absence of treatment with ticlopidine/clopidogrel. In the ASPECT trial, the rate of DES thrombosis in patients receiving cilostazol instead of thienopyridines was 14.8% (4 of 27) [102]. In another study, 30% of patients discontinuing ticlopidine early after DES implantation suffered ST [103]. Furthermore, Park et al. [104] performed a trial in order to investigate the incidence, risk factors, and association of antiplatelet therapy interruption for the development of ST after DES implantation during long-term follow-up (median follow-up 19 months). In this study a total of 1911 consecutive patients with DES were enrolled. The results of this study demonstrated that the incidence of late ST was 0.6%, similar to that of metallic stents. The predictors of ST were premature antiplatelet therapy interruption, primary stenting in acute MI, and total stent length. For a rare event such as ST, using standardized definitions across postmarketing studies and registries may help to further identify predictor of ST. However, as the ST incidence is low, a large sample size is needed to accurately identify these predictors.

## 6 POSTMARKET REGISTRIES AND OFF-LABEL STENT USE

Drug-eluting stent commercialization and use is monitored by the Center for Device and Radiologic Health (CDRH), a branch of the FDA. Recently, the off-label market use of DES has been under scrutiny by the FDA. The issues surrounding the real-world use versus the approved FDA use will be discussed in the next two sections.

### 6.1 Real-World Issues

The FDA has mandated a postmarket safety surveillance plan that must include at least 2000 patients (21 CFR 820). The purpose of the surveillance is to assure that public safety is not compromised and that the product is being used in accordance with its approved indication. Postmarket registries also collect more data on low rate events, such as death, MI, and ST. More long-term follow-up on a larger subset of complex patients can also be accomplished in postmarket studies that could not be done in smaller randomized controlled trials. Complex patients with diabetes, renal dysfunction, acute MI, triple-vessel disease, bifurcations, left main coronary disease, heavily calcified lesions, and longer lesion segments are included in postmarket registries; these patients would otherwise be excluded from most preapproval randomized controlled trials [105]. Additionally, meta-analysis can be completed for indirect comparisons of treatments to determine study inconsistencies [106]. Moreover, the effect of antiplatelet therapy on ST and the compliance with dual antiplatelet therapy can be monitored in postmarket trials. In particular, the duration of clopidogrel use in combination with aspirin has not been established yet for DES. Three to six months of use was initially recommended for the first-generation DES. The American Heart Association/American College of Cardiology/Society for Cardiovascular Angiography and Interventions currently recommends up to 12 months of antiplatelet therapy with DES use for those patients not at high risk for bleeding [107]. Although the duration of clopidogrel appeared to be adequate for the selected patients in the original clinical trials, optimal duration of clopidogrel in more complex patients has not been defined. Many years of additional data are needed to characterize the effects of antiplatelet therapy on DES ST rates before conclusions can be drawn on the optimal antiplatelet therapy duration for DES patients.

Between 2002 and 2004, the CYPHER postmarketing surveillance program enrolled over 15,000 patients who received a CYPHER DES [108]. The results demonstrated safety with low rates of ST and TLR. ARRIVE-1 was a postmarket registry for TAXUS DES and enrolled over 2000 patients. The findings at 2 years showed that discontinuation of clopidogrel before 6 months was a significant predictor of ST at 1 and 2 years. In addition, rates of cardiac events at 6 months were similar for diabetic and nondiabetic patients [109]. The ARRIVE-2 registry was a postapproval study that enrolled over 4000 patients in the real-world setting. The ARRIVE-2 registry showed outcomes and usage patterns similar to those seen in ARRIVE-1 [105]. Boston Scientific pooled the data from these registries and was able to show that TAXUS DES was being implanted in more complex settings. These trials led to several ATLAS trials that aimed to expand the indications for use in these more complex procedures and patients [105]. Medtronic recently

received conditional approval for the ENDEAVOR DES and will initiate a post-approval study with a minimum of 5000 patients. In addition, Abbott Vascular is seeking approval for the XIENCE V DES in the United States and has a postmarket surveillance plan including real-world use trials in populations worldwide. The XIENCE V DES program will include up to 16,000 patients treated with XIENCE V in the United States, India, Japan, China, and Europe. The XIENCE V postmarket surveillance plan is designed to detect low rate events of cardiac death, MI, and ST. Furthermore, the XIENCE V program will monitor patients' long-term compliance with dual antiplatelet therapy.

Collectively, the XIENCE V, TAXUS, ENDEAVOR, and postmarket surveillance programs will provide a comprehensive data set for use in analyzing the low rate of events such as death, MI, and ST.

### 6.2 Characterizing Off-Label Stent Use

Physicians may use a medical device off-label in good faith as the best available treatment option for the patient. The outcome of this off-label use can be a great benefit to the patient or may introduce considerable unnecessary risk or harm to the patient. Currently, the widespread off-label use of FDA-approved DES has prompted increased attention from regulatory agencies. Although the FDA does not regulate physician practice, it does establish guidelines for physicians to use products legally as indicated on the label. Monitoring safety- and efficacy-related information in these off-label use situations has become a considerable challenge for the manufacturer, physicians, and the FDA, making this dilemma a public health policy concern.

According to the FDA, it was estimated that 60% of DES use was off-label in 2006 [74]. Manufacturers and regulatory agencies can gather information regarding the breadth of off-label use through registries, postmarket surveillance/studies, device complaints, publications, payers, investigator-sponsored studies, professional societies, and scientific conferences. The manufacturer can gain considerable insight into the possibility of label expansion from this data because it may indicate that there are unmet needs in the market. In the case of DES, several off-label uses have prompted further studies to support approval of DES for more complex cases [105]. Some current off-label DES use involves bifurcation lesions, saphenous vein grafts, in-stent restenosis, total occlusions, acute MIs, and left main triple-vessel disease. Optimal medical therapy, BMS, and CABG therapies can also be compared against off-label DES use through explorative meta-analysis registries. A recent study compared BMS and DES off-label use versus BMS and DES on-label use. The results demonstrated that the risks of MI, death, or ST were lower for patients treated on-label versus off-label in both BMS and DES. There was also no difference observed between DES types [110]. In another recent trial comparing off-label use in DES to BMS, it was demonstrated that the use of DES was not associated with increased risk of death or MI at 1 year and that DES use continued to show lower revascularization rates in off-label use settings compared to BMS [93]. Without further data, the safety and efficacy of off-label use in all DES remains unknown.

The major disadvantage of off-label DES use is the potential for serious adverse outcomes to patients. The manufacturer may have initially sought FDA approval for an indication that was denied because of safety concerns. Physicians in everyday

practice may not be aware of the reasons why an indication was rejected or of this additional risk to their patients. Currently, there is a lack of process that can accurately analyze the outcomes of all off-label use. Patients not entered into a trial or registry may never receive long-term follow-up and their outcomes may remain unknown. The FDA has an obligation to obtain this information, yet at the same time it cannot condone the off-label practice that would generate the data needed for analysis. The manufacturer is restricted from promoting off-label use in practice. Physicians may use a DES for an unapproved indication instead of waiting for FDA approval, which can be a lengthy and complex process. The label expansion process can take as many or more resources as the initial device approval did [111]. Current FDA limitations for label expansion make this a costly and lengthy process, which ultimately may restrict the best treatment options for the patient. In the future, the industry and regulatory approving bodies must work toward a system that allows for rapid approval process while maintaining safety standards.

The practice of using DES for unapproved product indications is an increasing concern. Additional data are needed to assure the safety and efficacy of DES for the complex cases, lesions, and patients that are being treated off-label. Data should be gathered from randomized control trials, postapproval registries, or through national databases from physicians and manufacturers.

## 7 PERIPHERAL STENTS

Peripheral atherosclerotic occlusive disease or *peripheral vascular disease* (PVD) is exceedingly common in aging western societies; the prevalence of plaque in the femoral arteries of patients over the age of 70 may be as high as 74% [112]. Symptomatic PVD affects approximately 10% of the population over the age of 50, and as much as 20% of the population over the age of 70. Failure to maintain or restore adequate perfusion to the lower extremity results in nearly 250,000 lower extremity amputations annually.

Symptomatic PVD has traditionally been treated by surgical bypass grafting. First conceived in 1949 [113], and popularized in the 1980s through refinements in anesthesia and technique, bypass grafting was the accepted standard of care for nearly 50 years. Unfortunately, patient morbidity is considerable, as those requiring this invasive procedure are older and infirm, and the operation requires relatively long incisions and dissections through ischemic tissue. Hoping to circumvent the excessive morbidity of surgical bypass grafting, techniques of endovascular infrainguinal revascularization were developed in the 1980s. Percutaneous peripheral transluminal balloon angioplasty was first described in 1982, followed closely by reports of adjunctive stent implantation in 1985 [114, 115]. These techniques were adapted and enhanced over the ensuing two decades, including significant improvements in access, imaging, negotiation of chronic occlusions and, most notably, the wider application of self-expanding nitinol stent technology. Despite these refinements, maintenance of long-term patency following endovascular interventions is limited, as restenosis eventually complicates up to 50% of procedures during the first year (Fig. 3) [112, 116].

In the coronary circulation, the problem of restenosis was eventually circumvented through the development of coronary DESs [117]. Given their success, many

***FIGURE 3*** *Peripheral in-stent neointimal hyperplasia and restenosis. [From Scheinert, D., Scheinert, S., Sax, J., et al. (2005), Prevalence and clinical impact of stent fractures after femoropopliteal stenting, J. Am. Coll. Cardiol., 45, 312–315.]*

have theorized that this technology might also be useful in the peripheral arteries. The Sirocco stent (Cordis, a Johnson & Johnson Company, Miami Lakes, Florida) remains the first, and only, peripheral DES to be tested in clinical trials and reported in the scientific literature [118–120]. The Sirocco stent utilized the nitinol self-expanding S.M.A.R.T. stent as its platform, was loaded with 90 μg sirolimus/cm$^2$ stent area using a 5- to 10-μm copolymer matrix (total drug content ~1 mg per 80 mm stent), and delivered its drug load over a period of about 7 days [119]. A total of 93 patients were enrolled in the combined SIROCCO I and SIROCCO II clinical trials. Unfortunately, neither trial achieved its primary endpoint of a reduction in restenosis and, even after 4 years, there was no difference in any metric comparing patients treated with the bare S.M.A.R.T. stent versus the sirolimus-eluting Sirocco stent (Fig. 4) [120]. The development of the Sirocco stent was terminated; as of this writing, no peripheral DES is yet available for clinical use anywhere in the world.

Despite the disappointing results of the Sirocco studies, two other peripheral DESs are currently being tested in clinical trials. The Zilver PTX stent (Cook Inc.,

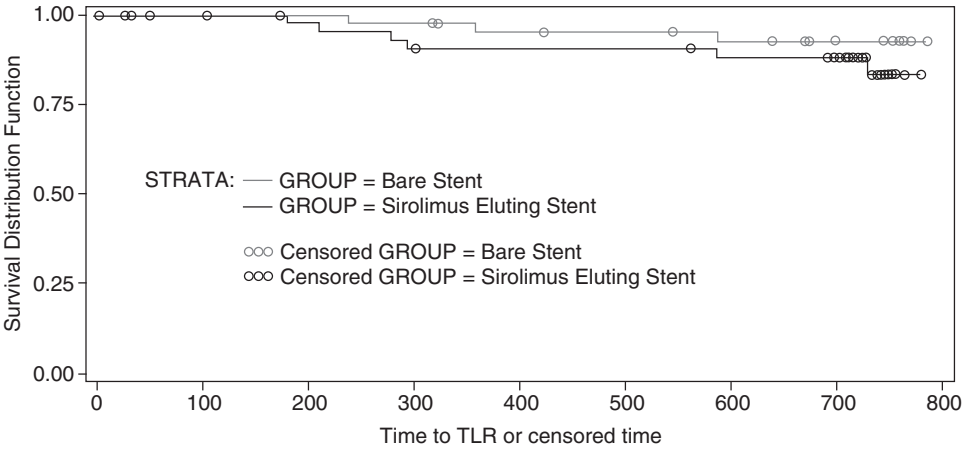


**Kaplan-Meter Analysis of Target Lesion Revascularization (n = 93 pts)**

| Interval ending day | 0 | 30 | 90 | 150 | 210 | 270 | 330 | 360 | 450 | 540 | 600 | 630 | 660 | 720 | 750 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Bare stent (n = 46)** | | | | | | | | | | | | | | | |
| # Censored | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 1 | 0 | 0 | 10 | 17 |
| # At risk | 46 | 45.5 | 44.5 | 44 | 42 | 42 | 41.5 | 39 | 39 | 38.5 | 37.5 | 36 | 36 | 31 | 17 |
| # Events | 0 | 0 | 0 | 0 | 2 | 0 | 2 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 |
| Cumulative survival, % | 100.0 | 100.0 | 100.0 | 100.0 | 95.5 | 95.5 | 90.9 | 90.9 | 90.9 | 90.9 | 88.5 | 88.5 | 88.5 | 88.5 | 83.8 |
| **Sirolimus stent (n = 47)** | | | | | | | | | | | | | | | |
| # Censored | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 7 | 21 |
| # At risk | 47 | 46.5 | 45.5 | 44.5 | 44 | 44 | 42 | 41 | 39.5 | 39 | 38.5 | 37 | 36.5 | 32 | 17.5 |
| # Events | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |
| Cumulative survival, % | 100.0 | 100.0 | 100.0 | 100.0 | 97.7 | 97.7 | 95.3 | 95.3 | 95.3 | 95.3 | 92.8 | 92.8 | 92.8 | 92.8 | 92.8 |

***FIGURE 4*** *Freedom from target lesion revascularization (TLR) in the SIROCCO II trial. Kaplan–Meier event analyses of target lesion revascularization: 92.8% for sirolimus versus 83.8% for control (p = 0.30). Reprinted from Duda, S. H., Bosiers, M., Lammer, J., et al. (2006), Drug-eluting and bare nitinol stents for the treatment of atherosclerotic lesions in the superficial femoral artery: Long-term results from the SIROCCO trial, J. Endovasc. Ther., 13, 701–710.*

Bloomington, Indiana) contains approximately 300 μg paclitaxel (PTX)/cm$^2$ stent area [121]. Similar to the ACHIEVE stent tested in the DELIVER clinical trial [122], paclitaxel is applied directly to the Zilver PTX stent; there is no polymer system serving to control its release. Interim clinical results from a small trial have recently been presented [123]. Sixty patients with symptomatic PVD were randomly assigned to undergo endovascular intervention with either the Zilver PTX stent ($n = 28$) or percutaneous transluminal balloon angioplasty alone (PTA, $n = 32$). Patients in whom PTA yielded suboptimal intraoperative results (16/32) were secondarily randomized to receive either a Zilver PTX stent or a bare nitinol Zilver stent. Although no angiographic results have been reported to date, the safety of this approach is suggested by the equivalent incidence of total events in the two groups after 6 months (incidence of death, revascularization and/or worsened Rutherford score in the PTA and Zilver PTX groups was 91 and 89%, respectively). Further testing of the safety and efficacy of the Zilver stent is ongoing in the 480-patient DESTINY trial using a similar design. Nearly half the patients have been recruited as of this writing.

Lastly, an everolimus-eluting nitinol stent has also recently entered the clinical phase of development. The Dynalink-E device (Abbott Vascular, Santa Clara, California) contains approximately 225 μg everolimus/cm$^2$ stent area, which is released slowly through incorporation in an ethylene vinyl alcohol (EVAL) copolymer. Enrollment was recently completed in the STRIDES First-in-Human clinical trial in Europe and initial results are expected in 2009.

In summary, endovascular intervention has become the treatment of choice for many patients with symptomatic PVD. Similar to percutaneous coronary intervention, it is hoped that the persistent and frequent problem of restenosis can be ameliorated though development of effective DES.

## 8 FUTURE TECHNOLOGY

### 8.1 Future Stents

Although the development and testing of the first-generation DES focused to a considerable degree on efficacy parameters, including restenosis, recent concerns over late clinical events have prompted refinement of the design criteria for succeeding generations of coronary artery devices.

Most new investigational balloon-expandable DES systems have lowered crossing profiles by thinning stent struts with the use of a cobalt chromium alloy, while investigational self-expanding DESs often use nitinol as the platform material. To further lower crossing profile, guidewire-based stents are currently under development [124]. Fully biodegradable stents are also being developed, with deliverability and performance to be determined in future clinical trials. New bifurcation-dedicated stents will secure branch accessibility to offer better deliverability in complex lesion morphologies. As noted, experimentation in stent design is already realizing multiple-lesion stenting and the in situ customization of stent length [124].

First-generation DESs have significantly lowered the rate of TLR compared with BMS. Rather than simply targeting further reductions in restenosis rates, efforts to improve efficacy are shifting toward a lesion-specific approach, including the design of stents dedicated to bifurcation lesions. Another future direction is a disease-specific approach, using DES as local drug delivery devices. So conceived, DESs present the possibility of expanding options for the local pharmacologic modulation of lesions. Potential therapeutic uses still in the investigational stage include myocardial protection at the time of infarct and the stabilization of vulnerable plaque.

Identifying long-term safety issues with the first-generation DES has reignited clinical interest in the development of stents that are more biologically based, including fully biodegradable stents and stents using biomimetic and biodegradable polymers. Important performance criteria for future DES agents include more cell-type specificity, broader safety margins, and greater facility at promoting endothelialization and healing. In addition, future vascular biologic findings on the mechanisms of restenosis will shed light on the potential relationship between gene therapy and local agent delivery. Safety is the perpetual challenging issue in the field of medicine. For interventional cardiologists, the recent recognition of a possible increase in late

clinical events in patients with DES has forced a reexamination and redefinition of the requirements of the ideal DES. Although previous DES clinical trials placed great emphasis on efficacy, especially midterm angiographic and clinical results, these findings pale in comparison to ensuring long-term safety for our patients.

New therapeutic agents such as proteins, nucleic acids [small interfering ribonucleic acids (RNAs) and large deoxyribonucleic acid (DNA) plasmids], viral delivery vectors, and even engineered cell therapies require specific delivery designs distinct from traditional smaller-molecule approaches on DES. While small molecules are currently the clinical standard for coronary stenting, extension of the DES to other lesion types, peripheral vasculature, and nonvasculature therapies will seek to deliver an increasingly sophisticated armada of drug types. In parallel with PCI procedure evolvement, new technologies such as OCT will be introduced for measuring safety and effectiveness of new devices. The application of these new technologies to PCI procedures may require validation of new safety and effectiveness measures.

### 8.2 Cardiac Tissue

Another alternative method for increasing coronary vascularization is the transplantation of stem or progenitor cells [125, 126]. These cells not only produce a variety of growth factors and cytokines but participate structurally in the formation of new vascular tissue and myocytes.

Promising results from experimental studies promoted the initiation of clinical trials of stem cell technologies [127, 128]. Stem and progenitor cells are being tested in patients with both acute MI and chronic ischemic heart failure [125, 129–136]. Improved wall motion or increased perfusion were demonstrated in most study patients. BOOST is a randomized controlled clinical trial [135] in which 60 patients with ST-segment elevation were randomly assigned to either a control group that received optimum postinfarction medical treatment or a bone-marrow-cell group that received optimum medical treatment and intracoronary transfer of autologous bone marrow cells 4.8 days after percutaneous coronary intervention (PCI). After 6 months, mean global left ventricular ejection fraction (LVEF) had significantly increased in the bone-marrow-cell group, compared to the control group. However, most studies at present are limited by the small patient enrollment, and recently some papers reported that bone-marrow-derived hematopoietic stem cells do not transdifferentiate into cardiac myocytes in ischemic myocardium [137, 138].

### 8.3 Genetic Advances

A number of vehicles are used to transfer DNA into heart tissue, including purified DNA, DNA/lipid complexes, adeno-associated viruses, or adenovirus [125, 139]. An advantage of this approach appears to be production of angiogenic factors that is localized and sustained but not indefinite [140]. Apparent disadvantages are the possibilities of vector toxicity or an immune response to the gene therapy vector.

The efficacy of gene transfer approaches to therapeutic angiogenesis is now being tested in clinical trials. Early uncontrolled open-label clinical trials have generally given positive results, although the possibility of a placebo effect has not been excluded [126, 139, 141–147]. Controlled phase II trials are providing positive but

not definitive results. This is promising since the patient population being studied has failed all other therapies and is likely to be refractory to intervention. However, most of the efficacy measures, studied to date, have been surrogate endpoints such as exercise tolerance time, angina, or perfusion. While these measures are useful in suggesting clinical efficacy, hard clinical endpoints such as mortality, MI, and the need for revascularization should be studied. Long-term follow-up data are also needed. An important observation is that the safety results of these trials indicate no major problems. Potential side effects such as worsening of atherosclerosis, retinopathy, or cancer have not been observed in clinical trials.

Two large phase III clinical trials (AGENT 3 and 4) were designed to evaluate further the safety and efficacy of Ad5FGF4. Both trials were designed as randomized, double-blind, and placebo-controlled trials [141]. In each trial, the recruitment goal was 450 patients, and these patients would be randomized to 3 groups (placebo group, adenovirus (Ad) human fibroblast growth factor (FGF) gene (Ad5FGF-4) at a dose of $10^9$ vp group, and Ad5FGF-4 at a dose of $10^{10}$ vp group). In January 2004, enrollment was stopped (416 patients in AGENT 3 and 115 patients in AGENT 4) because interim data analysis of AGENT 3 indicates that the studies will provide insufficient evidence of efficacy. However, enrolled patients' follow-up continues and final data will be presented in the near future.

## 9 CONCLUSIONS

The development of DES represents an exciting area of breakthrough technology, which has generated an enormous literature in parallel with widespread use in a short period of time. Interaction of innovative stent platforms, polymers, and molecular entities, as well as pharmaceutical adjuncts such as dual antiplatelet therapy, present a unique degree of complexity for systematic ongoing evaluation of these devices, their optimal use, and their real safety and performance results. Toward this end, clinical trials and DES industry programs have developed a broad variety of endpoint definitions, which differ across a heterogeneous array of cutoff values, timing of endpoint assessment, and outcome composites. The number of stents currently under investigation is substantial. They are all loaded with drugs that interfere with pathways in the process of inflammation and neointimal proliferation. However, the process of restenosis is a sequence of complex events that has been partly elucidated over the last 2 decades. Locally acting DESs provide the opportunity to interfere with the various mechanisms responsible for each step in the restenotic cascade [148], and a wide variety of different agents are currently available. Johnson and Johnson's sirolimus-eluting stents (SES), Boston Scientific's paclitaxel-eluting stents (PES), Medtronic's zotarolimus-eluting stents, and Abbott Vascular's XIENCE V everolimus-eluting stents have received FDA approval. Recently, in view of late adverse events after stent implantation, the FDA has called for long-term monitoring of safety outcomes in the postapproval context. It is therefore likely that postmarket programs for new DES will have larger sample size and longer follow-up in order to identify the true incidence of late stent thrombosis, a low frequency event. Such programs resemble traditional pharmacovigilance in their scope and size, which are likely to be global, and once implemented and successful, can result in restoration of confidence to the DES arena.

## REFERENCES


1. Gruentzig, A. R., King, S. B., 3rd, Schlumpf, M., and Siegenthaler, W. (1987), Long-term follow-up after percutaneous transluminal coronary angioplasty. The early Zurich experience, *N. Engl. J. Med.*, 316, 1127–1132.
2. Holmes, D. R.Jr, ., Vlietstra, R. E., Smith, H. C., et al. (1984), Restenosis after percutaneous transluminal coronary angioplasty (PTCA): A report from the PTCA Registry of the National Heart, Lung, and Blood Institute, *Am. J. Cardiol.*, 53, 77C–81C.
3. Baim, D. S. (2004), New devices for percutaneous coronary intervention are rapidly making bypass surgery obsolete, *Curr. Opin. Cardiol.*, 19, 593–597.
4. Baim, D. S., Kent, K. M., King, S. B., 3rd, et al. (1994), Evaluating new devices. Acute (in-hospital) results from the New Approaches to Coronary Intervention Registry, *Circulation*, 89, 471–481.
5. Fischman, D. L., Leon, M. B., Baim, D. S., et al. (1994), A randomized comparison of coronary-stent placement and balloon angioplasty in the treatment of coronary artery disease. Stent Restenosis Study Investigators, *N. Engl. J. Med.*, 331, 496–501.
6. Serruys, P. W., de Jaegere, P., Kiemeneij, F., et al. (1994), A comparison of balloon-expandable-stent implantation with balloon angioplasty in patients with coronary artery disease. Benestent Study Group, *N. Engl. J. Med.*, 331, 489–495.
7. Fricke, F. U., Silber, S. (2005), Can PCI with drug-eluting stents replace coronary artery bypass surgery? A comparative economic analysis regarding both therapeutic options based on clinical 12-month data reflecting the German social health care insurance system, *Herz*, 30, 332–338.
8. Baim, D. S., Mehran, R., Kereiakes, D. J., et al. (2000), Postmarket surveillance for drug-eluting coronary stents: A comprehensive approach, *Circulation*, 113, 891–897.
9. Barragan, P., Rieu, R., Garitey, V., et al. (2000), Elastic recoil of coronary stents: A comparative analysis, *Catheter Cardiovasc. Interv.*, 50, 112–119.
10. Ormiston, J. A., Dixon, S. R., Webster, M. W., et al. (2000), Stent longitudinal flexibility: A comparison of 13 stent designs before and after balloon expansion, *Catheter Cardiovasc. Interv.*, 50, 120–124.
11. Kastrati, A., Dirschinger, J., Boekstegers, P., et al. (2000), Influence of stent design on 1-year outcome after coronary stent placement: A randomized comparison of five stent types in 1,147 unselected patients, *Catheter Cardiovasc. Interv.*, 50, 290–297.
12. Hoffmann, R., Jansen, C., Konig, A., et al. (2001), Stent design related neointimal tissue proliferation in human coronary arteries; an intravascular ultrasound study, *Eur. Heart. J.*, 22, 2007–2014.
13. Kastrati, A., Mehilli, J., Dirschinger, J., et al. (2001), Restenosis after coronary placement of various stent types, *Am. J. Cardiol.*, 87, 34–39.
14. McLean, D. R., Eiger, N. L. (2002), Stent design: Implications for restenosis, *Rev. Cardiovasc. Med.*, 3(Suppl 5), S16–22.
15. Rogers, C., Edelman, E. R. (1995), Endovascular stent design dictates experimental restenosis and thrombosis, *Circulation*, 91, 2995–3001.
16. Yoshitomi, Y., Kojima, S., Yano, M., et al. (2001), Does stent design affect probability of restenosis? A randomized trial comparing Multilink stents with GFX stents, *Am. Heart. J.*, 142, 445–451.
17. Ku, D. N., Giddens, D. P., Zarins, C. K., and Glagov, S. (1985), Pulsatile flow and atherosclerosis in the human carotid bifurcation. Positive correlation between plaque location and low oscillating shear stress, *Arteriosclerosis*, 5, 293–302.

18. Malek, A. M., Alper, S. L., Izumo, S. (1999), Hemodynamic shear stress and its role in atherosclerosis, *JAMA*, 282, 2035–2042.
19. Moore, J. E.Jr, ., Xu, C., Glagov, S., Zarins, C. K., and Ku, D. N. (1994), Fluid wall shear stress measurements in a model of the human abdominal aorta: oscillatory behavior and relationship to atherosclerosis, *Atherosclerosis*, 110, 225–240.
20. Briguori, C., Sarais, C., Pagnotta, P., et al. (2002), In-stent restenosis in small coronary arteries: Impact of strut thickness, *J. Am. Coll. Cardiol.*, 40, 403–409.
21. Pache, J., Kastrati, A., Mehilli, J., et al. (2003), Intracoronary stenting and angiographic results: strut thickness effect on restenosis outcome (ISAR-STEREO-2) trial, *J. Am. Coll. Cardiol.*, 41, 1283–1288.
22. Rittersma, S. Z., de Winter, R. J., Koch, K. T., et al. (2004), Impact of strut thickness on late luminal loss after coronary artery stent placement, *Am. J. Cardiol.*, 93, 477–480.
23. Serruys, P. W. (1998), *Handbook of Coronary Stents*, Martin Dunitz Ltd., London.
24. van der Hoeven, B. L., Pires, N. M., Warda, H. M., et al. (2005), Drug-eluting stents: Results, promises and problems, *Int. J. Cardiol.*, 99, 9–17.
25. Axel, D. I., Kunert, W., Goggelmann, C., et al. (1997), Paclitaxel inhibits arterial smooth muscle cell proliferation and migration in vitro and in vivo using local drug delivery, *Circulation*, 96, 636–645.
26. Wong, A., Chan, C. (2004), Drug-eluting stents: The end of restenosis? *Ann. Acad. Med. Singapore*, 33, 423–431.
27. Ong, A. T., Serruys, P. W. (2005), Technology Insight: an overview of research in drug-eluting stents, *Nat. Clin. Pract. Cardiovasc. Med.*, 2, 647–658.
28. Regar, E., Sianos, G., Serruys, P. W. (2001), Stent development and local drug delivery, *Br. Med. Bull.*, 59, 227–248.
29. Winslow, R. D., Sharma, S. K., Kim, M. C. (2005), Restenosis and drug-eluting stents, *Mt. Sinai. J. Med.*, 72, 81–89.
30. Tanabe, K., Regar, E., Lee, C. H., Hoye, A., van der Giessen, W. J., Serruys, P. W. (2004), Local drug delivery using coated stents: New developments and future perspectives, *Curr. Pharm. Des.*, 10, 357–367.
31. Kastrati, A., Mehilli, J., Dirschinger, J., et al. (2001), Intracoronary stenting and angiographic results: strut thinkness effect on restenosis outcoume (ISAR-STEREO) trial, *Circulation*, 103, 2816–2821.
32. Sketch, M. H.Jr, ., Ball, M., Rutherford, B., Popma, J. J., Russell, C., and Kereiakes, D. J. (2005), Evaluation of the Medtronic (Driver) cobalt-chromium alloy coronary stent system, *Am. J. Cardiol.*, 95, 8–12.
33. Kereiakes, D. J., Cox, D. A., Hermiller, J. B., et al. (2003), Usefulness of a cobalt chromium coronary stent alloy, *Am. J. Cardiol.*, 92, 463–466.
34. Kipshidze, N., Dangas, G., Tsapenko, M., et al. (2004), Role of the endothelium in modulating neointimal formation: Vasculoprotective approaches to attenuate restenosis after percutaneous coronary interventions, *J. Am. Coll. Cardiol.*, 44, 733–739.
35. Moses, J. W., Leon, M. B., Popma, J. J., et al. (2003), Sirolimus-eluting stents versus standard stents in patients with stenosis in a native coronary artery, *N. Engl. J. Med.*, 349, 1315–1323.
36. Stone, G. W., Ellis, S. G., Cox, D. A., et al. (2004), A polymer-based, paclitaxel-eluting stent in patients with coronary artery disease, *N. Engl. J. Med.*, 350, 221–231.
37. Morice, M. C., Serruys, P. W., Sousa, J. E., et al. (2002), A randomized comparison of a sirolimus-eluting stent with a standard stent for coronary revascularization, *N. Engl. J. Med.*, 346, 1773–1780.

38. Schampaert, E., Cohen, E. A., Schluter, M., et al. (2004), The Canadian study of the sirolimus-eluting stent in the treatment of patients with long de novo lesions in small native coronary arteries (C-SIRIUS), *J. Am. Coll. Cardiol.*, 43, 1110–1115.
39. Schofer, J., Schluter, M., Gershlick, A. H., et al. (2003), Sirolimus-eluting stents for treatment of patients with long atherosclerotic lesions in small coronary arteries: double-blind, randomised controlled trial (E-SIRIUS), *Lancet*, 362, 1093–1099.
40. Grube, E., Silber, S., Hauptmann, K. E., et al. (2003), TAXUS I: Six- and twelve-month results from a randomized, double-blind trial on a slow-release paclitaxel-eluting stent for de novo coronary lesions, *Circulation*, 107, 38–42.
41. Colombo, A., Drzewiecki, J., Banning, A., et al. (2003), Randomized study to assess the effectiveness of slow- and moderate-release polymer-based paclitaxel-eluting stents for coronary artery lesions, *Circulation*, 108, 788–794.
42. Ellis, S. G., Popma, J. J., Lasala, J. M., et al. (2005), Relationship between angiographic late loss and target lesion revascularization after coronary stent implantation: analysis from the TAXUS-IV trial, *J. Am. Coll. Cardiol.*, 45, 1193–1200.
43. Stone, G. W., Ellis, S. G., Cannon, L., et al. (2005), Comparison of a polymer-based paclitaxel-eluting stent with a bare metal stent in patients with complex coronary artery disease: A randomized controlled trial, *JAMA*, 294, 1215–1223.
44. Dawkins, K. D., Grube, E., Guagliumi, G., et al. (2005), Clinical efficacy of polymer-based paclitaxel-eluting stents in the treatment of complex, long coronary artery lesions from a multicenter, randomized trial: Support for the use of drug-eluting stents in contemporary clinical practice, *Circulation*, 112, 3306–3313.
45. Leon, M. (2007), The ENDEAVOR IV. TCT 2007, Washington DC.
46. Joner, M., Quee, S. C., Coleman, L., et al. (2006), Abstract 2460, Competitive comparison of reendothelialization in drug eluting stents, *Circulation*, 114.
47. Serruys, P., Ruygrok, P., Neuzner, J., et al. (2006), A randomized comparison of an everolimus-eluting coronary stent with a paclitaxel-eluting coronary stent: The SPIRIT II trial, *EuroIntervention*, 2, 286–294.
48. Serruys, P. W., Ong, A. T., Piek, J. J., et al. (2005), A randomized comparison of a durable polymer everolimus-eluting stent with a bare metal coronary stent: The SPIRIT first trial, *EuroIntervention*, 1, 58–65.
49. Stone, G. W. (2008), *SPIRIT III. EuroPCR*. Barcelona, Spain.
50. Goy, J. J., Stauffer, J. C., Siegenthaler, M., Benoit, A., Seydoux, C. (2005) A prospective randomized comparison between paclitaxel and sirolimus stents in the real world of interventional cardiology: The TAXi trial, *J. Am. Coll. Cardiol.*, 45, 308–311.
51. Morice, M.C., Colombo, A., Meier, B., et al. (2006), Sirolimus- vs paclitaxel-eluting stents in de novo coronary artery lesions: The REALITY trial: A randomized controlled trial, *JAMA*, 295, 895–904.
52. Windecker, S., Remondino, A., Eberli, F. R., et al. (2005), Sirolimus-eluting and paclitaxel-eluting stents for coronary revascularization, *N. Engl. J. Med.*, 353, 653–662.
53. Ormiston, J. A., Webster, M. W., Armstrong, G. (2007), First-in-human implantation of a fully bioabsorbable drug-eluting stent: the BVS poly-L-lactic acid everolimus-eluting coronary stent, *Catheter Cardiovasc. Interv.*, 69, 128–131.
54. Ormiston, J. A., Serruys, P. W., Regar, E., et al. (2008), A bioabsorbable everolimus-eluting coronary stent system for patients with single de-novo coronary artery lesions (ABSORB): A prospective open-label trial, *Lancet*, 371, 899–907.
55. http://www.teamreva.com/heart_stents_tech.html RMIW.
56. Kereiakes, D. J., Teirstein, P. S., Sarembock, I. J., et al. (2007), The truth and consequences of the COURAGE trial, *J. Am. Coll. Cardiol.*, 50, 1598–1603.

57. Anon. (2007), Prevalence of heart disease—United States, 2005, *MMWR Weekly*, 56, 113–118.

58. Bravata, D. M., Gienger, A. L., McDonald, K. M., et al. (2007), Systematic review: The comparative effectiveness of percutaneous coronary interventions and coronary artery bypass graft surgery, *Ann. Intern. Med.*, 147, 703–716.

59. Anon. (2002), Coronary artery bypass surgery versus percutaneous coronary intervention with stent implantation in patients with multivessel coronary artery disease (the Stent or Surgery trial): A randomised controlled trial, *Lancet*, 360, 965–970.

60. Morrison, D. A., Sethi, G., Sacks, J., et al. (2001), Percutaneous coronary intervention versus coronary artery bypass graft surgery for patients with medically refractory myocardial ischemia and risk factors for adverse outcomes with bypass: A multicenter, randomized trial. Investigators of the Department of Veterans Affairs Cooperative Study #385, the Angina with Extremely Serious Operative Mortality Evaluation (AWESOME), *J. Am. Coll. Cardiol.*, 38, 143–149.

61. Rodriguez, A., Bernardi, V., Navia, J., et al. (2001), Argentine Randomized Study: Coronary angioplasty with stenting versus coronary bypass surgery in patients with multiple-vessel disease (ERACI II): 30-day and one-year follow-up results. ERACI II Investigators, *J. Am. Coll. Cardiol.*, 37, 51–58.

62. Rodriguez, A. E., Baldi, J., Fernandez Pereira, C., et al. (2005), Five-year follow-up of the Argentine randomized trial of coronary angioplasty with stenting versus coronary bypass surgery in patients with multiple vessel disease (ERACI II), *J. Am. Coll. Cardiol.*, 46, 582–588.

63. Serruys, P. W., Ong, A. T., van Herwerden, L. A., et al. (2005), Five-year outcomes after coronary stenting versus bypass surgery for the treatment of multivessel disease: The final analysis of the Arterial Revascularization Therapies Study (ARTS) randomized trial, *J. Am. Coll. Cardiol.*, 46, 575–581.

64. Leon, M. B., Holmes, D. R., JrWeisz, G., ., et al. (2006), Two-year outcomes after sirolimus-eluting stent implantation: Results from the Sirolimus-Eluting Stent in de Novo Native Coronary Lesions (SIRIUS) trial, *J. Am. Coll. Cardiol.*, 47, 1350–1355.

64a. Valgimigli, M. (2007), Impact of stable versus unstable coronary artery disease on 1-year outcome in elective patients undergoing multivessel revascularization with sirolimus-eluting stents, *JACC*, 49.

65. Kappetein, A. P., Dawkins, K. D., Mohr, F. W., et al. (2006), Current percutaneous coronary intervention and coronary artery bypass grafting practices for three-vessel and left main coronary artery disease. Insights from the SYNTAX run-in phase, *Eur. J. Cardiothorac. Surg.*, 29, 486–491.

66. Farkouh, M. E., Dangas, G., Leon, M. B., et al. (2008), Design of the Future Revascularization Evaluation in patients with Diabetes mellitus: Optimal management of Multivessel disease (FREEDOM) Trial, *Am. Heart J.*, 155, 215–223.

67. Hueb, W., Lopes, N. H., Gersh, B. J., et al. (2007), Five-year follow-up of the Medicine, Angioplasty, or Surgery Study (MASS II): A randomized controlled clinical trial of 3 therapeutic strategies for multivessel coronary artery disease, *Circulation*, 115, 1082–1089.

68. Pasceri, V., Patti, G., Speciale, G., Pristipino, C., Richichi, G., Di Sciascio, G. (2007), Meta-analysis of clinical trials on use of drug-eluting stents for treatment of acute myocardial infarction, *Am. Heart J.*, 153, 749–754.

69. Boden, W. E., O'Rourke, R. A., Teo, K. K., et al. (2007), Optimal medical therapy with or without PCI for stable coronary disease, *N. Engl. J. Med.*, 356, 1503–1516.

70. Anon. (1996), Comparison of coronary bypass surgery with angioplasty in patients with multivessel disease. The Bypass Angioplasty Revascularization Investigation (BARI) Investigators, *N. Engl. J. Med.*, 335, 217–225.

71. King, S. B., 3rd, Lembo, N. J., Weintraub, W. S., et al. (1994), A randomized trial comparing coronary angioplasty with coronary bypass surgery. Emory Angioplasty versus Surgery Trial (EAST), *N. Engl. J. Med.*, 331, 1044–1050.

72. Henderson, R. A., Pocock, S. J., Clayton, T. C., et al. (2003), Seven-year outcome in the RITA-2 trial: Coronary angioplasty versus medical therapy, *J. Am. Coll. Cardiol.*, 42, 1161–1170.

73. Ikeda, S., Bosch, J., Banz, K., Schneller, P. (2000), Economic outcomes analysis of stenting versus percutaneous transluminal coronary angioplasty for patients with coronary artery disease in Japan, *J. Invasive. Cardiol.*, 12, 194–199.

74. Uchida, T. (2006), Summary from the Circulatory Systems Devices Advisory Panel Meeting, Gaithersburg, Maryland, December 7–8.

75. Serruys, P. W., Unger, F., Sousa, J. E., et al. (2001), Comparison of coronary-artery bypass surgery and stenting for the treatment of multivessel disease, *N. Engl. J. Med.*, 344, 1117–1124.

76. Babapulle, M. N., Joseph, L., Belisle, P., Brophy, J. M., Eisenberg, M. J. (2004), A hierarchical Bayesian meta-analysis of randomised clinical trials of drug-eluting stents, *Lancet*, 364, 583–591.

77. Biondi-Zoccai, G. G., Agostoni, P., Abbate, A., et al. (2005), Adjusted indirect comparison of intracoronary drug-eluting stents: evidence from a metaanalysis of randomized bare-metal-stent-controlled trials, *Int. J. Cardiol.*, 100, 119–123.

78. Brunner-La Rocca, H. P., Kaiser, C., Pfisterer, M. (2007), Targeted stent use in clinical practice based on evidence from the Basel Stent Cost Effectiveness Trial (BASKET), *Eur. Heart J.*, 28, 719–725.

79. Dibra, A., Kastrati, A., Alfonso, F., et al. (2007), Effectiveness of drug-eluting stents in patients with bare-metal in-stent restenosis: meta-analysis of randomized trials, *J. Am. Coll. Cardiol.*, 49, 616–623.

80. Ellis, S. G. C. A, Grube, E., Popma, J., Koglin, J., Dawkins, K. D., and Stone, G. W. (2007), Incidence, timing, and correlates of stent thrombosis with the polymeric paclitaxel drug-eluting stent: A TAXUS II, IV, V, and VI meta-analysis of, 3445 patients followed for up to 3 years, *J. Am. Coll. Cardiol.*, 49, 1043–1051.

81. Holmes, D. R.Jr, ., Moses, J. W., Schofer, J., Morice, M. C., Schampaert, E., Leon, M. B. (2006), Cause of death with bare metal and sirolimus-eluting stents, *Eur. Heart. J.*, 27, 2815–2822.

82. Kastrati, A., Dibra, A., Spaulding, C., et al. (2007), Meta-analysis of randomized trials on drug-eluting stents vs. bare-metal stents in patients with acute myocardial infarction, *Eur. Heart J.*, 28, 2706–2713.

83. Lagerqvist, B., James, S. K., Stenestrand, U., Lindback, J., Nilsson, T., Wallentin, L. (2007), Long-term outcomes with drug-eluting stents versus bare-metal stents in Sweden, *N. Engl. J. Med.*, 356, 1009–1019.

84. Mauri, L., Hsieh, W. H., Massaro, J. M., Ho, K. K., D'Agostino, R., Cutlip, D. E. (2007), Stent thrombosis in randomized clinical trials of drug-eluting stents, *N. Engl. J. Med.*, 356, 1020–1029.

85. Moreno, R., Fernandez, C., Calvo, L., et al. (2007), Meta-analysis comparing the effect of drug-eluting versus bare metal stents on risk of acute myocardial infarction during follow-up, *Am. J. Cardiol.*, 99, 621–625.

86. Moreno, R., Fernandez, C., Hernandez, R., et al. (2005), Drug-eluting stent thrombosis: results from a pooled analysis including 10 randomized studies, *J. Am. Coll. Cardiol.*, 45, 954–959.

87. Moses, J. W., Stone, G. W., Nikolsky, E., et al. (2006), Drug-eluting stents in the treatment of intermediate lesions: pooled analysis from four randomized trials, *J. Am. Coll. Cardiol.*, 47, 2164–2171.

88. Nordmann, A. J., Briel, M., Bucher, H. C. (2006), Mortality in randomized controlled trials comparing drug-eluting vs. bare metal stents in coronary artery disease: A meta-analysis, *Eur. Heart J.*, 27, 2784–2814.

89. Spaulding, C., Daemen, J., Boersma, E., Cutlip, D. E., and Serruys, P. W. (2007), A pooled analysis of data comparing sirolimus-eluting stents with bare-metal stents, *N. Engl. J. Med.*, 356, 989–997.

90. Stettler, C., Wandel, S., Allemann, S., et al. (2007) Outcomes associated with drug-eluting and bare-metal stents: A collaborative network meta-analysis, *Lancet*, 370, 937–948.

91. Stone, G. W., Moses, J. W., Ellis, S. G., et al. (2007), Safety and efficacy of sirolimus- and paclitaxel-eluting coronary stents, *N. Engl. J. Med.*, 356, 998–1008.

92. Roiron, C., Sanchez, P., Bouzamondo, A., Lechat, P., Montalescot, G. (2006), Drug eluting stents: An updated meta-analysis of randomised controlled trials, *Heart*, 92, 641–649.

93. Marroquin, O. C., Selzer, F., Mulukutla, S. R, et al. (2008), A comparison of bare-metal and drug-eluting stents for off-label indications, *N. Engl. J. Med.*, 358, 342–352.

94. Cutlip, D. E., Windecker, S., Mehran, R., et al. (2007), Clinical end points in coronary stent trials: A case for standardized definitions, *Circulation*, 115, 2344–2351.

95. Cutlip, D. E., Chauhan, M. S., Baim, D. S., et al. (2002), Clinical restenosis after coronary stenting: Perspectives from multicenter clinical trials, *J. Am. Coll. Cardiol.*, 40, 2082–2089.

96. Cutlip, D. E., Chhabra, A. G., Baim, D. S., et al. (2004), Beyond restenosis: Five-year clinical outcomes from second-generation coronary stent trials, *Circulation*, 110, 1226–1230.

97. Kereiakes, D. J., Kuntz, R. E., Mauri, L., Krucoff, M. W. (2005), Surrogates, substudies, and real clinical end points in trials of drug-eluting stents, *J. Am. Coll. Cardiol.*, 45, 1206–1212.

98. Spertus, J. A., Winder, J. A., Dewhurst, T. A., et al. (1995), Development and evaluation of the Seattle Angina Questionnaire: A new functional status measure for coronary artery disease, *J. Am. Coll. Cardiol.*, 25, 333–341.

99. Anon. (2000), Myocardial infarction redefined—a consensus document of The Joint European Society of Cardiology/American College of Cardiology Committee for the redefinition of myocardial infarction, *Eur. Heart J.*, 21, 1502–1513.

100. Pocock, S. J., Lansky, A. J., Mehran, R., et al. (2008), Angiographic surrogate end points in drug-eluting stent trials: A systematic evaluation based on individual patient data from 11 randomized, controlled trials, *J. Am. Coll. Cardiol.*, 51, 23–32.

101. Lemos, P. A., Lee, C. H., Degertekin, M., et al. (2003), Early outcome after sirolimus-eluting stent implantation in patients with acute coronary syndromes: Insights from the Rapamycin-Eluting Stent Evaluated at Rotterdam Cardiology Hospital (RESEARCH) registry, *J. Am. Coll. Cardiol.*, 41, 2093–2099.

102. Hong, M. K., Mintz, G. S., Lee, C. W., et al. (2003), Paclitaxel coating reduces in-stent intimal hyperplasia in human coronary arteries: A serial volumetric intravascular ultrasound analysis from the Asian Paclitaxel-Eluting Stent Clinical Trial (ASPECT), *Circulation*, 107, 517–520.

103. Pasceri, V., Granatelli, A., Pristipino, C., Pelliccia, F., Pironi, B., and Richichi, G. (2003), High-risk of thrombosis of Cypher stent in patients not taking ticlopidine or clopidogrel, *Am. J. Cardiol.*, 92.

104. Park, S. J., Shim, W. H., Ho, A. E., et al. (2003), A paclitaxel-eluting stent for the prevention of coronary restenosis, *N. Engl. J. Med.*, 348, 1537–1545.

105. Lasala, J. M., Stone, G. W., Dawkins, K. D., et al. (2006), An overview of the TAXUS Express, paclitaxel-eluting stent clinical trial program, *J. Interv. Cardiol.*, 19, 422–431.

106. Lumley, T. (2002), Network meta-analysis for indirect treatment comparisons, *Stat. Med.*, 21, 2313–2324.

107. King, S. B., 3rd, Aversano, T., Ballard, W. L., et al. (2007), ACCF/AHA/SCAI, 2007 update of the Clinical Competence Statement on Cardiac Interventional Procedures: A report of the American College of Cardiology Foundation/American Heart Association/American College of Physicians Task Force on Clinical Competence and Training (Writing Committee to Update the 1998 Clinical Competence Statement on Recommendations for the Assessment and Maintenance of Proficiency in Coronary Interventional Procedures), *Circulation*, 116, 98–124.

108. Urban, P. (2005), CYPHER. PCR, Paris.

109. Dobies, D. (2007), ARRIVE-1. PCR, Barrelona.

110. Applegate, R. J., Sacrinty, M. T., Kutcher, M. A., et al. (2008), "Off-label" stent therapy, 2-year comparison of drug-eluting versus bare-metal stents, *J. Am. Coll. Cardiol.*, 51, 607–614.

111. Russell, M. E., Friedman, M. I., Mascioli, S. R., Stolz, L. E. (2006), Off-label use: An industry perspective on expanding use beyond approved indications, *J. Interv. Cardiol.*, 19, 432–438.

112. Hirsch, A. T., Haskal, Z. J., Hertzer, N. R., et al. (2006), ACC/AHA 2005 guidelines for the management of patients with peripheral arterial disease (lower extremity, renal, mesenteric, and abdominal aortic): executive summary. A collaborative report from the American Association for Vascular Surgery/Society for Vascular Surgery, Society for Cardiovascular Angiography and Interventions, Society for Vascular Medicine and Biology, Society of Interventional Radiology, and the ACC/AHA Task Force on Practice Guidelines (writing committee to develop guidelines for the management of patients with peripheral arterial disease), *J. Am. Coll. Cardiol.*, 1239–1312.

113. Kunlin, J. (1949), Le traitement de l'artèrite oblitèrante par la greffe veineuse, *Arch. Mal. Coeur*, 42, 371–372.

114. Ring, E. J., McLean, G. K., Freiman, D. B. (1982), Selected techniques in percutaneous transluminal angioplasty, *Am. J. Roentgenol.*, 139, 767–773.

115. Palmaz, J. C., Sibbitt, R. R., Reuter, S. R., Tio, F. O., Rice, W. J. (1985), Expandable intraluminal graft: A preliminary study. Work in progress, *Radiology*, 156, 73–77.

116. Norgren, L., Hiatt, W. R., Dormandy, J. A., Nehler, M. R., Harris, K. A., Fowkes, F. G. R. (2007), Inter-society consensus for the management of peripheral arterial disease (TASC II), *J. Vasc. Surg.*, 45(S), S5A–S67A.

117. Serruys, P. W., Kutryk, M. J. B., Ong, A. T. L. (2006), Coronary artery stents, *N. Engl. J. Med.*, 354, 483–495.

118. Duda, S. H., Pusich, B., Richter, G., et al. (2002), Sirolimus-eluting stents for the treatment of obstructive superficial femoral artery disease: Six-months results, *Circulation*, 106, 1505–1509.

119. Duda, S. H., Bosiers, M., Lammer, J., et al. (2005), Sirolimus-eluting versus bare nitinol stent for obstructive superficial femoral artery disease: The SIROCCO II Trial, *J. Vasc. Interv. Radiol.*, 16, 331–338.

120. Duda, S. H., Bosiers, M., Lammer, J., et al. (2006), Drug-eluting and bare nitinol stents for the treatment of atherosclerotic lesions in the superficial femoral artery: Long-term results from the SIROCCO trial, *J. Endovasc. Ther.*, 13, 701–710.
121. Heldman, A. W., Ragheb, A. O. (2004), Paclitaxel persists in vessel wall after rapid delivery from nonpolymeric paclitaxel coating on self-expanding Nitinol stents (abstract), *Am. J. Cardiol.*, 94(suppl, 6A), 27E.
122. Lansky, A. J., Costa, R. A., Mintz, G. S., et al. (2004), Non-polymer-based paclitaxel-coated coronary stents for the treatment of patients with de novo coronary lesions: angiographic follow-up of the DELIVER clinical trial, *Circulation*, 109, 1948–1954.
123. Dake, M. D. (2007), *Interim Report on the Zilver PTX Clinical Trial*. International Society of Endovascular Therapy (ISET), Hollywood, FL.
124. Aoki, J., Rodriguez-Granillo, G. A., Serruys, P. W. (2005), Emergent strategies in interventional cardiology. *Rev. Esp. Cardiol.*, 58, 962–973.
125. Losordo, D. W., Dimmeler, S. (2004), Therapeutic angiogenesis and vasculogenesis for ischemic disease: part II: Cell-based therapies, *Circulation*, 109, 2692–2697.
126. Rafii, S., Lyden, D. (2003), Therapeutic stem and progenitor cell transplantation for organ vascularization and regeneration, *Nat. Med.*, 9, 702–712.
127. Kawamoto, A., Gwon, H. C., Iwaguro, H., et al. (2001), Therapeutic potential of ex vivo expanded endothelial progenitor cells for myocardial ischemia, *Circulation*, 103, 634–637.
128. Kawamoto, A., Tkebuchava, T., Yamaguchi, J., et al. (2003), Intramyocardial transplantation of autologous endothelial progenitor cells for therapeutic neovascularization of myocardial ischemia, *Circulation*, 107, 461–468.
129. Assmus, B., Schachinger, V., Teupe, C., et al. (2002), Transplantation of progenitor cells and regeneration enhancement in acute myocardial infarction (TOPCARE-AMI), *Circulation*, 106, 3009–3017.
130. Britten, M. B., Abolmaali, N. D., Assmus, B., et al. (2003), Infarct remodeling after intracoronary progenitor cell treatment in patients with acute myocardial infarction (TOPCARE-AMI): mechanistic insights from serial contrast-enhanced magnetic resonance imaging, *Circulation*, 108, 2212–2218.
131. Perin, E. C., Dohmann, H. F., Borojevic, R., et al. (2003), Transendocardial, autologous bone marrow cell transplantation for severe, chronic ischemic heart failure, *Circulation*, 107, 2294–2302.
132. Stamm, C., Westphal, B., Kleine, H. D., et al. (2003), Autologous bone-marrow stem-cell transplantation for myocardial regeneration, *Lancet*, 361, 45–46.
133. Strauer, B. E., Brehm, M., Zeus, T., et al. (2002), Repair of infarcted myocardium by autologous intracoronary mononuclear bone marrow cell transplantation in humans, *Circulation*, 106, 1913–1918.
134. Tse, H. F., Kwong, Y. L., Chan, J. K., Lo, G., Ho, C. L., and Lau, C. P. (2003), Angiogenesis in ischaemic myocardium by intramyocardial autologous bone marrow mononuclear cell implantation, *Lancet*, 361, 47–49.
135. Wollert, K. C., Meyer, G. P., Lotz, J., et al. (2004), Intracoronary autologous bone-marrow cell transfer after myocardial infarction: The BOOST randomised controlled clinical trial, *Lancet*, 364, 141–148.
136. Fuchs, S., Satler, L. F., Kornowski, R., et al. (2003), Catheter-based autologous bone marrow myocardial injection in no-option patients with advanced coronary artery disease: A feasibility study, *J. Am. Coll. Cardiol.*, 41, 1721–1724.
137. Balsam, L. B., Wagers, A. J., Christensen, J. L., Kofidis, T., Weissman, I. L., Robbins, R. C. (2004), Haematopoietic stem cells adopt mature haematopoietic fates in ischaemic myocardium, *Nature*, 428, 668–673.

138. Murry, C. E., Soonpaa, M. H., Reinecke, H., et al. (2004), Haematopoietic stem cells do not transdifferentiate into cardiac myocytes in myocardial infarcts, *Nature*, 428, 664–668.
139. Yla-Herttuala, S., Alitalo, K. (2003), Gene transfer as a tool to induce therapeutic vascular growth, *Nat. Med.*, 9, 694–701.
140. Simons, M., Bonow, R. O., Chronos, N. A., et al. (2000), Clinical trials in coronary angiogenesis: issues, problems, consensus: An expert panel summary, *Circulation*, 102, E73–86.
141. Grines, C. L. (2004), The AGENT clinical trials programme, *Eur. Heart J. Suppl.*, 6, E18–E23.
142. Hedman, M., Hartikainen, J., Syvanne, M., et al. (2003), Safety and feasibility of catheter-based local intracoronary vascular endothelial growth factor gene transfer in the prevention of postangioplasty and in-stent restenosis and in the treatment of chronic myocardial ischemia: Phase II results of the Kuopio Angiogenesis Trial (KAT), *Circulation*, 107, 2677–2683.
143. Henry, T. D., Annex, B. H., McKendall, G. R., et al. (2003), The VIVA trial: Vascular endothelial growth factor in Ischemia for Vascular Angiogenesis, *Circulation*, 107, 1359–1365.
144. Kastrup, J., Jorgensen, E., and Ruck, A. (2003), Euroinject Trial. Late breaking clinical trials session. American College of Cardiology, 2003, Chicago, *J. Am. Coll. Cardiol.*, 41, 1603.
145. Seiler, C., Pohl, T., Wustmann, K., et al. (2001), Promotion of collateral growth by granulocyte-macrophage colony-stimulating factor in patients with coronary artery disease: A randomized, double-blind, placebo-controlled study, *Circulation*, 104, 2012–2017.
146. Simons, M., Annex, B. H., Laham, R. J., et al. (2002), Pharmacological treatment of coronary artery disease with recombinant fibroblast growth factor-2: double-blind, randomized, controlled clinical trial, *Circulation*, 105, 788–793.
147. Grines, C. L., Watkins, M. W., Helmer, G., et al. (2002), Angiogenic Gene Therapy (AGENT) trial in patients with stable angina pectoris, *Circulation*, 105, 1291–1297.
148. Virmani, R., Farb, A. (1999), Pathology of in-stent restenosis, *Curr. Opin. Lipidol.*, 10, 499–506.

# *Clinical Trials Involving Oral Diseases*

**Bruce L. Pihlstrom,[1] Bryan Michalowicz,[1] Jane Atkinson,[2] and Albert Kingman[2]**

[1]*School of Dentistry, University of Minnesota, Minneapolis, Minnesota*
[2]*National Institute of Dental and Craniofacial Research, National Institutes of Health, Bethesda, Maryland*

**Contents**

1 Introduction
2 Prominent Design Issues in Oral Health Clinical Trials
  2.1 Split-Mouth and Cross-Over Designs
  2.2 Single-versus Multicenter Trials
  2.3 Use of Active versus Negative Controls
  2.4 Identification of Clinically Meaningful Endpoints
  2.5 Independent Sampling (Sites versus Subjects) and Multilevel Modeling
  2.6 Randomization and Potential Bias Due to Confounding Factors
  2.7 Examiner Training and Calibration
  2.8 Trial Duration
3 Prominent Statistical Issues in Oral Health Clinical Trials
  3.1 Clinical Significance
  3.2 Confidence Intervals versus *p* Values
4 Regulatory Issues
5 Management of Oral Health Clinical Trials
  5.1 Roles and Responsibilities of Trial Personnel
  5.2 Data Coordination Center
  5.3 Data Collection

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

5.4 Recruitment
5.5 Provision of Standard of Care
5.6 Use of Nontreatment and Delayed Treatment Controls
6 Caries Clinical Trials
7 Periodontal Clinical Trials
8 Dental Implant Clinical Trials
9 Other Oral Conditions
9.1 Xerostomia
9.2 Mucosal Candidiasis
9.3 Mucosal Diseases
References

## 1 INTRODUCTION

Dentistry is rapidly entering an era of evidence-based practice, and individual patients and society are demanding that clinical practice and public health policy decisions be based on the best available science. Well-designed randomized oral health clinical trials (RCTs) are needed because they provide the highest level of scientific evidence [1]. It is very difficult for oral health practitioners, researchers, and sponsors who have not been directly involved in large oral health RCTs to truly understand their complexity, challenges, and expense. Moreover, few oral health investigators have appropriate training or clinical trial experience. These factors have contributed to the relative lack of large phase III multicenter oral health clinical trials.

Clinical trials in oral health have mainly focused on the diseases of dental caries and periodontal disease, but a limited number have involved other oral diseases and conditions. These include chronic facial pain, particularly associated with the temporomandibular joint, acute pain after extraction of third molars (wisdom teeth), xerostomia (dry mouth), and various oral mucosal diseases. With the advent of evidence-based dentistry, more clinical trials are needed to establish the efficacy *and* effectiveness of prevention and treatment for a wide variety of oral diseases and conditions.

Oral health trials share many things in common with clinical trials in general, but they have several unique challenges. This chapter focuses on phase III oral health clinical trials in terms of design and analyses, regulatory and management issues, and challenges of conducting trials involving specific oral diseases.

## 2 PROMINENT DESIGN ISSUES IN ORAL HEALTH CLINICAL TRIALS

Design issues of special interest in oral health clinical trials include: advantages and disadvantages of so-called split-mouth and cross-over designs, single versus multicenter trials, use of active versus negative controls, identification of clinically meaningful endpoints, independent sampling (sites versus subjects) and multilevel modeling, randomization and potential bias due to confounding factors, examiner training and calibration, and trial duration.

### 2.1 Split-Mouth and Cross-Over Designs

The parallel design in which subjects are randomly assigned to receive either an active or a control (placebo) intervention is the standard design used in clinical trials. However, oral health clinical trials often use a "split-mouth" design in which various segments of the dentition are randomly assigned to treatment and control arms. Furthermore, many toothpaste and oral rinse studies use a cross-over design that in its simplest form randomizes subjects to one of two arms of a trial, followed by a "washout" period, after which they are assigned to the other arm. An investigator testing the longevity or integrity of dental sealants, composite restorative materials, or dental implants may wonder which study design would be the most appropriate for the trial. One could either use a split-mouth, cross-over, or a parallel design.

The main weakness of the split-mouth design is that treatments may have effects on segments of the dentition other than those to which they were assigned; the outcome of a split-mouth design is the treatment effect plus the sum of these effects [2]. The split-mouth design removes the between-group variance in estimating treatment effect, but it may include effects from treatments administered to other parts of the dentition. As such, it has the potential of biasing the trial against finding differences between treatments. When considering use of this design, it is important to estimate the within-patient correlation of treatment effect; if it is small, little gain in efficiency is obtained by using the split-mouth design; if it is moderate to high, the gain in efficiency should be weighed against the potential disadvantages of the split-mouth design including potential bias, potential difficulty in subject recruitment (there must be similar disease in two quadrants of the mouth), and complexity of the statistical analysis [3]. A split-mouth design may be appropriate if the "carry-across" effect from one tooth site or treatment segment to another is minimal. Unfortunately, there is no easy method to test for a carry-across effect in a split-mouth study, and the investigator would have to be convinced that potential carry-cross effects in the trial are negligible. The parallel design avoids these potential sources of bias and facilitates straightforward statistical analysis. The main weakness of the parallel design compared to the split-mouth and cross-over is it will require many more subjects to achieve comparable statistical significance. In spite of this constraint, the parallel design is generally preferable for diseases such as periodontal disease and caries.

Cross-over designs could be appropriate for short-term trials investigating outcomes that are reversible, such as gingivitis, plaque, and even for some types of chronic pain. However, cross-over trials have limitations due to possible treatment carry-over effect from one period to the next period and the possibility of a patient learning effect that can bias the results. In an attempt to minimize these concerns, a washout period is introduced between active treatment periods. Overall, for plaque and gingivitis trials, the double-blind parallel randomized trial is preferred to the double-blind cross-over design, except possibly for a professionally applied treatment that does not leach into the saliva or affect the oral microflora [4].

### 2.2 Single versus Multicenter Trials

The need for generalizability of results and the number of patients required for a clinical trial are important considerations when designing clinical trials. Multicenter

trials are conducted to generate results that can be generalized to the population at large. For many dental diseases, community water fluoridation, cultural norms, tobacco use, diet, and many other factors may contribute to trial outcomes. Therefore, regardless of size, the outcome of a single-center trial may only be applicable to a particular sample of subjects. Another important reason for using multiple center sites is that there may be insufficient subjects at a single enrollment site that meet trial enrollment criteria. Adequate numbers of subjects are frequently not available for large dental implant and periodontal disease trials; multicenter studies are often needed to enroll enough subjects in these trials. Earlier caries clinical trials were conducted at several schools in a single community (single-site trial). This was possible because most of these trials focused on preventing, rather than treating dental caries. Moreover, caries was very common and it was cost effective to recruit sufficient numbers of qualifying subjects in schools. As caries rates have declined in developed countries, it is much more difficult to find subjects who are at risk for developing new caries. This has necessitated the use of much larger, multicenter caries prevention trials, often involving 800–1500 patients per group.

### 2.3 Use of Active versus Negative Controls

For ethical reasons the majority of recent caries clinical trials have used active rather than negative controls. This is because fluoride has proven to be effective in preventing dental caries and many feel that it is not ethical to include a nonfluoride (negative) control group. If the purpose of a trial is to demonstrate that a new agent or product is superior to the current standard (positive control) agent or product, the minimum difference ($\Delta$) judged to be clinically meaningful between the new agent and the standard would be smaller than if a negative control group were used. Use of a positive control would also require increased sample size. Alternatively, because of improved taste, less cost, ease of application, or reduction in side effects, the purpose of a trial may be to demonstrate, for example, that a new fluoride agent is as good as the standard fluoride agent (positive control). In these trials, a noninferiority trial design generally aims to demonstrate that the new agent or product is not inferior to the standard. A value for clinical comparability or equivalence for the new agent must be defined ($\Delta$) a priori in order to define how close the primary clinical outcome for the new agent ($\mu_N$) must approximate the outcome using the standard agent ($\mu_S$) for it to be considered at least as good as the standard ($\Delta = \mu_N - \mu_S$). To establish efficacy of nonfluoride agents for caries prevention, use of negative controls may be possible. For these trials, the magnitude of the clinically important difference in the development of new caries over a specific time period must be defined. The rate of new caries over time is often called the caries increment.

### 2.4 Identification of Clinically Meaningful Endpoints

Outcomes in randomized clinical trials must be clearly specified, measurable, valid, and reproducible. It is also important that pivotal phase III clinical trials have clinically meaningful endpoints [5]. As defined by a 1999 joint National Institute of Health–Food and Drug Administration (NIH–FDA) workshop on biomarkers and clinical endpoints [6] a *clinical endpoint* is a characteristic or variable that reflects

how a patient feels, functions, or survives; a *surrogate endpoint* is a biomarker intended to substitute for a clinical endpoint and is expected to predict clinical benefit (or harm, or lack of benefit or harm) based on epidemiologic, therapeutic, pathophysiologic, or other scientific evidence; and a *biomarker* is a characteristic that is objectively measured and evaluated as an indicator of normal biologic processes, pathogenic processes, or (pharmacologic) responses to a therapeutic intervention.

True validation of surrogate endpoints that are used in clinical trials in medicine is rare [5] and oral health clinical trials are no exception. Hujoel pointed out that there has been widespread use of periodontal variables such as probing depth, clinical attachment level, and bleeding on probing that have not been validated for use in clinical trials [7]. He suggested that true endpoints such as tooth loss, pain, bleeding after tooth brushing, and other subjective quality of life measurements should be used in pivotal trials of periodontal disease [7]. In a caries trial, reducing pain from caries would be more meaningful than reducing the level of *mutans streptococci*, and in a clinical trial of xerostomia change in subject-perceived oral dryness would be more meaningful than a statistically significant change in inflammatory salivary proteins. Surrogate endpoints can greatly increase efficiency and decrease costs of clinical trials, and more research is needed to validate their use in oral health clinical trials [8].

### 2.5 Independent Sampling (Sites versus Subjects) and Multilevel Modeling

Because clinical measures of oral disease such as caries and periodontal disease can be obtained at many sites in the dentition, there has been confusion as to whether individual sites within subjects or subjects themselves should be the independent unit of analysis in clinical trials. A unique feature of many dental clinical trials is their complex data structure that frequently involves clustered observations of multiple measurements repeated over time within each subject. This introduces multiple levels of structural data dependency; it is clear that sites are correlated or clustered within subjects and that sites cannot be used as independent sampling units in statistical analysis [9].

Collection of a large number of data items from each subject is often necessary because dental disease is tooth- and even tooth-site-specific. That is, diseases such as caries and periodontal disease can progress at very few tooth sites in an otherwise healthy dentition. Unless all sites are evaluated, it is possible that localized progressive disease will remain undetected. The criteria for progressive disease and the protocol for delivering rescue or additional treatment should be specified beforehand in the trial's operations manual. With so many measures, it can be difficult to screen repeated measures to determine if disease is stabile or progressing, and redundant checks are needed to compare measures collected over time. Web-based or electronic chair-side data capture can facilitate this process.

There are 32 teeth in the adult dentition. It is common to measure various periodontal parameters at 6 sites per tooth (mesiobuccal, midbuccal, distobuccal, distolingual, midlingual, and mesiolingual) and caries on 5 surfaces per tooth (occlusal, mesial, distal, buccal, and lingual). Therefore, it is possible to include measurements of up to 168 sites (excluding third molars = 28 teeth × 6 sites per tooth) in periodon-

tal trials and 128 sites in caries trials (excluding third molars there are 16 posterior teeth × 5 surfaces per tooth plus 12 anterior teeth × 4 surfaces per tooth). Since it is clear that sites cannot be treated as independent units of sampling, it is common to aggregate site-based data at the patient level. Depending on the specific trial, typical patient-level periodontal outcome measures have included (1) mean clinical attachment loss, (2) mean probing pocket depth, (3) number of sites with bleeding on probing, (4) extent of clinical attachment loss, and (5) extent of probing pocket depth (using a specific threshold, e.g., 3, 4, or 6 mm). Use of such aggregated data at the patient level may be appropriately used in conventional statistical analyses with the subject as the independent unit of observation. Although statistically correct, this may result in loss of detailed information at the site level; it is clear that the variance in periodontal parameters may be attributed to factors at both the site and the subject level [10]. In this regard, use of multilevel modeling (MLM) can be a valuable statistical method to assess disease progression over time at all levels of the natural hierarchy of tooth sites within subjects [11, 12]. Multilevel modeling of periodontal data has demonstrated that there is considerable variation at all levels of analysis and that MLM can be a more powerful research tool than single-level techniques for the analysis of hierarchical dental data [13].

### 2.6 Randomization and Potential Bias Due to Confounding Factors

Common confounding factors that could lead to selection or confounding bias in oral health clinical trials include age, sex, socioeconomic status, educational level, smoking, diet, oral hygiene, exposure to fluoride and baseline caries level (for caries), use of medications that influence salivary flow, and alcohol use (for oral cancer). Smoking should be considered a strong confounder or effect modifier in periodontal disease, and in most cases it is helpful to separately compare treatment arms among smokers and nonsmokers [14].

Proper attention to randomization procedures assures comparability between treatment groups in clinical trials by minimizing selection and confounding bias [15]. However, small sample clinical trials are prone to imbalances in known and unknown factors that may affect treatment outcome [16]. For relatively small sample trials, use of stratified randomization to achieve balance between test and control arms for factors that may affect treatment outcome should be considered. As noted by Schulz and Grimes, proper randomization takes little time and effort but results in major rewards in terms of scientific accuracy and credibility [15].

### 2.7 Examiner Training and Calibration

As stated by Fleiss in the very first sentence of his classical text *The Design and Analysis of Clinical Experiments* [17]: "The most elegant design of a clinical study will not overcome the damage caused by unreliable or imprecise measurement." Because of the complexity of oral disease measurements, examiner training and calibration are essential components of any oral health clinical trial. For example, it is common for oral health clinical trial examiners to record measurements on 28 teeth in *each subject's mouth* for probing depth, clinical attachment level, gingivitis, dental plaque, and dental calculus ($n = 168$ sites for each parameter) plus tooth mobility ($n = 28$) for a total of 868 separate measurements per patient at each evalu-

ation visit. This is an enormous task and to minimize examiner fatigue and subject discomfort, only essential data that are needed to measure trial outcome should be collected.

Proper training, calibration, and certification of examiners to prespecified standards of discrimination and standardization are essential. The use of multiple examiners makes this task even more complex, time-consuming, and costly because inter- *and* intraexaminer variability must be minimized by training and repeated calibration. Examiners should be trained and calibrated using patients whose severity and extent of disease is similar to the trial population because measures of examiner reliability may be artificially inflated if patients used in calibration exercises are healthier than study participants. For periodontitis trials, calibration statistics such as percent agreement within a specified threshold will be artificially high if subjects used for calibration have only early or localized disease. Examiners should be recalibrated during the trial using a subsample of subjects (e.g., 5%) to ensure consistent evaluation of trial outcomes. Interim assessments are important because experienced examiners trained at the beginning of a trial may slowly revert to their original techniques or methods over time [18]. It is important that measurement error be expressed in terms that are meaningful to the clinician while retaining statistical validity [19]. There is within-subject clustering of agreement for various oral disease measures and failure to account for dependence among site-level measurements can result in a false sense of precision in examiner reliability estimates [20]. Because examiner calibration and reliability is such a major issue in oral health clinical trials, details of examiner training and reproducibility data should be reported in publications.

### 2.8 Trial Duration

As with all clinical trials, the duration of oral health clinical trials depends on many factors including the specific disease being investigated, trial design, numbers and enrollment criteria of subjects to be enrolled, the intervention, and the primary outcome variable. It has been recommended that gingivitis trials be at least 6 months in duration [21] and that periodontitis trials be 9 months [19]. Cross-over designs would require a trial at least twice as long as a two-arm split-mouth or parallel design. Caries trials should be 3 years in duration and include at least one interim examination in addition to an assessment at baseline and at the end of the trial [22]. Ideally, clinical outcomes should be monitored for at least 3 years in trials of direct dental restorations (amalgams, resin-based composites) and for 5 years in trials of indirect restorations (crowns, inlays). In reality, outcomes are rarely assessed after 3 years, even though the materials are designed to last much longer in clinical practice.

## 3 PROMINENT STATISTICAL ISSUES IN ORAL HEALTH CLINICAL TRIALS

Obtaining multiple measurements for each subject in many dental clinical trials provides the opportunity to analyze data at the subject, tooth, and site level. Analysis will depend on the primary purpose of the trial and trial design. Most dental clinical

trials use a parallel design with randomization of treatment to subjects and a statistical analysis focused on a subject-level primary outcome measure. Outcomes could also be analyzed at the individual site or tooth level, but within-mouth correlations must be considered in these analyses. In this regard, generalized estimating equations (GEE) [23, 24] and other methods such as an incidence density method [23, 25], Poisson regression models [26], survival methods [24, 27, 28], and specific nonparametric procedures for ordinal-scaled outcome measures have been proposed [29, 30]. Using GEE is one example of a marginal approach, that is, a population average method that provides estimates at the population level, rather than for individual subjects. Various conditional approaches can be used if the focus is on individual subjects; in dental clinical trials, a conditional model is probably more appropriate. A mixed-model approach can be used to account for repeated correlated observations within subjects.

### 3.1 Clinical Significance

Phase III RCTs must have sufficient sample size to yield results that are statistically *and* clinically significant. Clinical significance is a matter of judgment and may be defined as the minimal magnitude of difference between test and control interventions that would cause clinicians to change their standard of clinical practice or that would inform public health policy. In other words, it answers the question "Is the difference between the groups large enough to be worthwhile?"

It is often difficult to reach consensus on what constitutes clinical significance. It becomes a trade-off between the size of the intervention effect that is considered to be clinically superior and the size that is considered to be equivalent. It also may involve a compromise between what is desirable and what is practical for testing in a clinical trial. In any event, to properly estimate sample size for a trial, clinical significance for the outcome variable must be defined a priori. For example, if a 3-year phase III RCT demonstrated that subjects using an inexpensive over-the-counter (OTC) caries prevention product had a clinically meaningful (i.e., 20% reduction) average reduction in 0.6 new carious surfaces per subject compared to appropriate control subjects, it would have major public health significance and would substantially impact the OTC market for caries prevention. Using the formula from Pocock [31], if we assume a standard deviation in 3-year caries incidence of $\sigma = 3$ surfaces, a difference in caries increment ($\Delta$) = 0.6 surfaces, $\alpha = 0.05$, and $1 - \beta = 0.80$ (power), the sample size for each arm may be estimated at 395. That is, $2 \times 395 = 790$, or almost 800 subjects would be needed for the trial. The actual number of subjects required for screening would be much larger because of the need to identify subjects at high risk for developing dental caries.

### 3.2 Confidence Intervals versus *p* Values

Two basic statistical methods are used to assess the role of chance in estimation and hypothesis testing. The confidence interval (CI) is used in estimation and the *p* value is used in hypothesis testing. A 95% CI is the default value for most clinical trials. It is a frequency-based interval estimate of the unknown true difference between treatment groups based on what would happen under multiple replications of the same study. A desirable property of the 95% CI is that its range or width represents a measure of the precision of the estimate in addition to its significance. For example,

suppose a 3-year caries trial compared two fluoride toothpastes (500 ppm vs. 1000 ppm fluoride) and reported mean (SE) caries DMFS increments of 1.75 (0.1) and 1.55 (0.08), respectively, and that the difference of 0.20 (SE = 0.13; $p = 0.12$) had an associated 95% CI of –0.06 to +0.46. There was no statistically significant difference in the effectiveness of the two toothpastes because $p = 0.12$. It is important to note that the probability value ($p = 0.12$) does not provide an estimate of precision. The clinical significance of the trial outcome may be better appreciated by an inspection of the 95% CI for the difference in DMFS increments. The statistical significance of the difference between treatment outcomes can be seen in the 95% CI because the value 0 (representing no difference between fluoride toothpastes) lies within the confidence interval estimate. Moreover, in this study, a fairly high level of precision of the estimated difference was achieved because the confidence interval was fairly narrow (–0.06 to +0.46). Sample size would need to be increased if a higher level of precision (narrower CI) is desired.

## 4 REGULATORY ISSUES

Most materials used to restore tooth structure are considered medical devices by regulatory agencies such as the U.S. FDA. One specific center within the FDA, the Center for Devices and Radiological Health, regulates products such as dental restorative materials, electric toothbrushes, and dental implants [32]. Each device is grouped by medical specialty and assigned a class (class I, II, or III) based on the level of control needed to ensure safety and effectiveness of the device. Standards to guide the manufacture and performance of various dental materials have been developed by two large organizations, the dental committee of the International Organization for Standardization (ISO), and the American Dental Association, which develops performance standards in conjunction with the American National Standards Institute. Each publishes its own set of standards [33, 34], which are used by dental manufacturers worldwide.

Most individuals seeking FDA clearance for new dental materials must submit a 510(k) application. Many of these products will be classified as class I or class II devices. Class I devices may be exempt from premarket review provided that the device is appropriately registered with the agency. Most class II devices require demonstration of substantial equivalence to another legally U.S.-marketed device. Substantial equivalence means that the new device is at least as safe and effective as the predicate [35]. In some instances, clinical data is needed for FDA clearance of the new device, especially if the manufacturer asserts differences in claims or intended use. Because many new dental materials are similar to older materials, they are classified as class II devices. As such, phase II trials are sufficient and they are not tested in phase III clinical trials.

## 5 MANAGEMENT OF ORAL HEALTH CLINICAL TRIALS

### 5.1 Roles and Responsibilities of Trial Personnel

Each staff member involved in a trial must understand, accept, and be capable of performing his or her role. Specific roles and responsibilities should be assigned

before the trial begins and monitored regularly throughout the trial. While clinical data may be collected by qualified personnel (e.g., dental hygienists), licensure as a dentist is generally required to diagnosis oral disease. For example, in a clinical trial of periodontal disease, it is common for a dental hygienist to perform clinical assessments and collect dental plaque samples or oral biospecimens.

Typically, however, a licensed dentist must also examine each subject or review the clinical and radiographic findings to establish that the patient has the condition or disease specified by the protocol. For caries trials, a nondentist may determine the presence of cavitated lesions (areas of pitted or soft tooth) or photograph and assess fluorescent images of teeth, but a licensed dentist is usually required to establish a diagnosis and provide (or supervise) the trial intervention. Institutional review boards may look unfavorably on a study in which the patient is not examined, diagnosed, and monitored by a dentist.

Therapists must be competent and experienced in treating or managing the disease or condition under study. Patients enrolled in clinical trials should receive the same quality of care that is delivered in nonresearch clinical settings. Dental hygienists or general dentists typically deliver nonsurgical periodontal therapy and in caries trials, treatment is usually provided by licensed and experienced dentists. For severe periodontal disease, treatment involving surgery, or for other interventions outside of the expertise of general dental practitioners, treatment should be provided by specialists or dentists with advanced training or experience.

Study masking or blinding is an important feature of clinical trials that serves to minimize bias during data collection and assessment [16]. Masking can be defined in terms of the participant, the therapist, and the examiner. Treatment of oral disease tends to be procedurally oriented, and therapists in most oral health trials are not masked. Masking participants also can be difficult in some trials. Consider a trial that compares a nontraditional, minimally invasive treatment for dental caries to conventional restorative care (i.e., dental fillings). Because conventional care involves removal of tooth decay with a high speed drill and requires local anesthesia, subjects receiving the minimally invasive therapy would also need to receive local anesthesia (regardless of the need) to maintain masking. For periodontitis trials, it can be difficult to mask examiners because some treatments may produce more gingival recession (resulting in more exposed tooth root) than others. An experienced clinical examiner could unintentionally note such changes following some treatments. To minimize potential bias, patients should be instructed to refrain from discussing treatment or treatment-related signs or symptoms with trial examiners. Patient complaints, such as tooth sensitivity, can be discussed with the study coordinator or lead investigator, provided these individuals do not collect, record, or discuss data with examiners.

### 5.2 Data Coordination Center

Trial coordination and data management are essential to the success of any clinical trial. Large multicenter trials need a data coordinating center (DCC). The DCC helps design the trial, collects data from all centers, manages data, provides day-to-day guidance on protocol implementation, provides trial monitoring, oversees examiner training and calibration, ensures data quality, and analyzes data. Often, the trial statistician is part of the DCC. In most contemporary clinical trials, the

DCC facilitates data input at individual clinical centers by providing computer software and information technology (IT) support.

### 5.3 Data Collection

A primary focus of the investigative team should be to collect quality data. Rigorous monitoring is needed to ensure that data are collected according to trial protocol. Study personnel should carefully review study forms before the end of each subject visit so that errors or omissions can be identified and immediately rectified. Data collected on paper forms should be submitted to the DCC in a timely manner. The DCC should review data collection forms for completeness and send queries to enrollment sites when errors are detected or suspected. Missing data or data entry errors should be corrected as soon as possible because a long duration between error identification and correction leads to greater likelihood for error.

Oral health clinical trials frequently collect information about a participant's medical health. If medical information is collected as part of a trial, it should be collected by staff who are well versed in medical terminology and who are able to verify its accuracy and completeness. If data is abstracted from medical records, this task should be performed by someone who is familiar with the medical record format. In our experience, trained nurses working in the medical care facility are best suited to perform these tasks. Using "in-house" nurses also helps protect patient privacy because medical records do not leave the facility and because nurses are usually well trained in privacy issues.

### 5.4 Recruitment

Even the best designed and most innovative clinical trials will fail if recruitment efforts fail. Investigators should not underestimate the amount of effort needed for successful subject recruitment. Like all trials, oral health trials should have detailed recruitment plans in place before it begins. Investigators often overestimate the prevalence of a condition or disease in the population they plan to study. The "take all comers" environment of clinical practice often causes clinicians to overestimate the number of available patients for trials. Recruitment and participation in oral health clinical trials have been discussed elsewhere [36–41]. It must be emphasized that recruitment will be more successful when trials have relatively few exclusion criteria. Limiting exclusion criteria also increases the likelihood that trial results will be generalizable and embraced by community practitioners.

Institutional review board-approved fliers and advertisements can be posted in busy public places or played on radio or television. While this may yield large numbers of initial inquiries, enrollment success with these methods is often relatively low because people typically have only a cursory understanding of their oral health. For example, enrollment in periodontal disease trials is usually limited to people with disease severity that can only be determined by a periodontal examination. Mass media advertisements for periodontal trials often attract people with a wide variety of periodontal conditions such as gingival recession due to toothbrush abrasion, gingivitis, and even nonbacterial inflammatory conditions such as desquamative gingivitis. If mass media advertising is used, appropriate attention and

resources must be committed to screen potential subjects for trial participation. This can include initial telephone screening for inclusion and exclusion criteria, followed by a clinical examination. Recruitment from dental clinics is often productive and trials can be advertised by posting fliers in waiting areas and patient examination rooms. However, it is critical that investigators establish an ongoing relationship with clinic staff. We have found that simply posting fliers or placing pamphlets in dental clinics is unproductive. Successful recruitment requires that study personnel have regular and substantive contact with clinic personnel to remind them of the trial and its importance. Importantly, recruitment of patients from clinics or private offices should be a "seamless" process. Clinic staff should make patients aware of the study and of the potential benefits of participating. However, once a patient expresses interest in the trial, study personnel, not a clinical staff member, should guide the patient through the consent, screening, and enrollment processes. Because clinic staff are very busy providing patient care, they cannot be expected to perform burdensome tasks such as scheduling trial appointments, distributing enrollment forms, or collecting data. Milgrom et al. [40] reported that recruitment for oral health trials can be enhanced by using concurrent recruitment methods, accurately estimating the costs and time required for recruitment, and regular monitoring recruitment efforts. They also found that newspaper public service advertisements and campus posters were the most effective means for recruiting subjects with periodontitis in a university setting [41].

Active clinical research centers may regularly screen patients recruited from print and broadcast advertisements to identify subjects for possible participation in clinical research. It must be stressed that screening volunteers for research generally requires institutional review board approval. Those who are screened should be encouraged to seek needed dental care if relevant trials are not planned for the near future because some patients for financial or other reasons might defer treatment in the hope they will be enrolled in a trial.

Unfortunately, most dental diseases are quite prevalent in minority and underserved populations [42]. It is critical that research be conducted in partnership with these populations so that disparities in oral health can be addressed. Research investigators should partner with community leaders from underserved populations to make sure that people from underserved populations are made aware of trials that are available for their participation. Moreover, barriers for participation by underserved populations need to be addressed. For example, investigators should consider establishing satellite clinics in underserved communities or provide taxi vouchers if transportation to the main study site is an impediment to trial participation. Provision of on-sight child care and extended or nighttime clinic hours may also make participation in clinical trials more feasible.

While clinical trials must satisfy the Belmont principles of respect, beneficence, and justice, patients and communities may demand more from investigators for their participation. Educational programs on disease prevention may help investigators obtain community trust and reduce the burden of oral diseases. For trials that recruit in underserved communities, study personnel should provide information about community or low-cost clinics where participants can receive care for non-study-related oral diseases. Despite elaborate advertisement and recruitment efforts, enrollment in clinical trials will not be successful if people perceive they are not being treated with respect and compassion. Everyone involved is more likely to be

engaged and enthusiastic if they are treated with respect, appreciate the importance of the trial, and are not unduly burdened as a result of their participation.

### 5.5 Provision of Standard of Care

Participants randomized to the control group in RCTs often receive "standard" therapy; these are termed positive-controlled trials. Alternatively, control subjects may be monitored rather than treated. These are termed negative-controlled trials. Occasionally, trials will include positive and negative controls. In planning a positive-controlled trial, it is necessary for investigators to decide what constitutes the "standard of care." In planning a positive controlled trial, it is necessary for investigators to decide what constitutes the "standard of care" [43].

For dental caries, common and scientifically proven methods to remove caries include lasers, high-pressure water jets, and traditional high-speed rotary drills [44]. The definition of standard of care can impact how a new treatment compares to the standard in terms of clinical outcomes, but also how patients perceive the new treatment. Consider a trial designed to explore patient preferences for a new restorative technique. The new treatment may compare differently if the standard involves the use of a noisy high-speed drill and placement of a dental amalgam filling versus a method in which tooth decay is removed with a laser and replaced with a tooth-colored filling. Because both methods are used in clinical practice, each could satisfy the legal definition of standard of care.

Professional societies typically do not attempt to define standard of care but instead may publish practice guidelines or parameters of care [45]. Nonetheless, the standard of care in dentistry includes an examination to identify disease so it may be appropriately treated. Patients with oral disease should be referred for treatment if it is not part of the trial protocol, and those with oral signs and symptoms of medical disorders should be referred to a health care provider for assessment and care. Examples would include patients with oral disease suggestive of uncontrolled diabetes, or those with mucosal lesions that are suggestive of conditions such as pemphigoid, pemphigus, lichen planus, or a variety of blood dyscrasias.

The American Academy of Periodontology's (AAP) parameter of care for gingivitis [46] recommends that patients with gingivitis receive education, personalized oral hygiene instruction, and a tooth scaling. It is also recommended that the use of antimicrobial and antiplaque agents or devices be considered for patients who are unable to clean their teeth adequately and that overhanging tooth restorations, ill-fitting prostheses, or other conditions that retain dental plaque be corrected. In gingivitis trials, the test intervention should be compared to treatment that is consistent with these guidelines.

The AAP's parameters of care for chronic periodontitis [47, 48] clearly outlines the importance of a personalized oral hygiene instruction and the provision of nonsurgical and maintenance care. In positive-controlled periodontitis trials, control subjects should receive treatment that is consistent with these guidelines. For trials involving periodontal surgery, patients should be managed according to these guidelines either before trial enrollment or before they are randomized to receive surgical treatment. This is important because periodontal surgery is ineffective in the presence of inadequate plaque control [49].

### 5.6 Use of Nontreatment and Delayed Treatment Controls

When considering the use of placebo or untreated controls in oral health clinical trials, the consequences of not providing care or delaying it must be carefully considered. The ethics of using untreated controls may vary according to the oral disease being studied. For example, the ethics of using untreated controls differ in trials of gingivitis versus trials of premalignant oral lesions. Gingivitis is an early form of periodontal disease that does not affect the supporting structures of the teeth; it reversible with simple oral hygiene procedures such as tooth brushing and flossing and does not cause destruction of periodontal support. People with gingivitis could be included in a delayed treatment group, provided there is monitoring to ensure that there is no progression to periodontitis and loss of tooth support. Conversely, delaying treatment of premalignant oral lesions could put patients at risk for developing cancer and would raise serious ethical concerns.

Chronic periodontal disease is a slowly progressing disease that can lead to tooth loss. Tooth loss generally occurs over many years and the risk for progressive disease is relatively low, even in untreated patients over a 3-year period [50]. In a recent study of pregnant women with periodontitis, it was found that only 0.28% of sites in untreated control subjects had progressive disease over 20 weeks [51]. Despite the low risk for disease progression, it is important to monitor subjects in clinical trials and provide early rescue treatment as soon as evidence of progressive disease if found. For nonsurgical periodontitis trials, untreated control subjects with progressive disease should receive scaling, root planning, and oral hygiene instruction. Depending on the patient's periodontal status and oral hygiene, previously treated subjects may be re-treated with mechanical therapy alone, receive local or systemic antimicrobials, or be treated surgically.

Oral health trials of long duration sometimes use community-based controls. In these trials, initial treatment and follow-up care is administered to test subjects while control subjects are referred to community dentists for treatment and follow-up care. These trials are uncommon in oral diseases because the trials tend to be of rather short duration. However, oral disease prevention trials may require long-term follow-up and use of community-based controls should be considered. A concern with this approach is the lack of available dental care in some communities. If a trial recruits from low socioeconomic or underserved populations who have limited access to oral health care, requiring control subjects to seek care in community settings may not be ethical.

In all oral health trials, untreated or delayed treated control subjects should receive follow-up evaluations at the same time intervals as test subjects. Because patients with periodontal disease are typically seen for follow-up care every 3–4 months, participants in periodontitis trials also should be examined and receive follow-up care at this interval. Examinations should include full-mouth periodontal probing with radiographs as needed. For dental caries trials, participants should be examined at least semiannually; more frequently for high-risk subjects. Depending on the specific trial, these examinations may include visual assessment, clinical probing, tooth transillumination, and use of other imaging systems. Intraoral radiographs should obtained as indicated by caries risk.

## 4 CARIES CLINICAL TRIALS

Dental caries is a chronic, slowly progressing infectious disease caused by acid produced from microorganisms and fermentable carbohydrates. Dental caries can affect enamel, dentin, and cementum; clinically, it occurs on a continuum from initial mineral loss to complete tooth destruction [52]. Caries can occur on the coronal or root surfaces of the teeth and on smooth surfaces or occlusal pits and fissures. Historically, the primary purpose of caries prevention and/or treatment has been to restore function and prevent tooth loss.

Traditional methods for assessing caries focused on cavitated lesions using the Radike or the World Health Organization (WHO) DMFS indices [53]. The Radike DMFS index diagnoses active caries (D) at the cavitated level, whereas the WHO index allows for four different levels of active caries, scored as $D_1$, $D_2$, $D_3$, and $D_4$. $D_1$ and $D_2$ scores indicate noncavitated lesions, whereas $D_3$ and $D_4$ are gradations of cavitated lesions. For comparability, the Radike DMFS index is labeled $D_3$MFS to indicate assessment of cavitated lesions. Past evidence of caries is evaluated by counting restorations (F), and missing teeth (M). In caries clinical trials, incidence (new disease or caries increment) is calculated as the difference between the final and initial $D_3$MFS scores. Historically, caries trials were conducted over 3 years and included a baseline, final, and at least one interim $D_3$MFS scoring. The 3-year caries mean increments were compared for the test groups using linear models, after adjusting for important covariates. A common summary measure was the percentage reduction in caries incidence for a test agent compared with a reference (control) agent. $D_3$MFS indices tend to overestimate caries experience because teeth, especially in adults, can be missing or filled for reasons other than caries. Despite its limitations, the $D_3$MFS index has been widely used in caries trials. It is sufficiently sensitive to detect differences in relative efficacy of a wide variety of fluoride delivery systems, including water fluoridation, fluoride mouth rinses, dental office fluoride application, and fluoride dentifrices.

Dental caries has declined in many developed countries over the past few decades [54], and the rate of caries progression through dentin has slowed [55, 56]. As a result, caries activity has extended into and beyond late adolescence. While fluoride can enhance enamel remineralization, a caries-resistant outer shell of enamel can sometimes mask dentinal caries, making caries difficult to diagnose without using radiographs. As a result of changes in caries rates, prevention and treatment trials are more expensive, which may increase the possibility that newer, potentially superior anticaries agents may never be tested in phase III trials.

Since the 1980s, additional methods have been developed for diagnosing caries at the precavitated state. These include modifications of the visual-tactile method, such as the $D_1$MFS [57], the International Caries Diagnostic Assessment System (ICDAS) [58] and sophisticated technologies such as fiber-optic transillumination (FOTI) and digital fiber-optic transillumination (DIFOTI) [59], quantitative light-induced fluorescence (QLF) [60], DIAGNOdent [61], and electrical conductivity measurement (ECM) [62]. Although these methods need to be more fully validated before they are accepted for use in pivotal clinical trials, they have the potential to improve diagnosis of precavitated carious lesions. Use of newer and more sensitive diagnostic methods also could change the focus of therapy from "drilling and filling"

decayed teeth to remineralizing or reversing early caries lesions. It could also reduce the length, size, and cost of caries trials.

A major issue that has plagued caries trials is high rates of subject dropout. Earlier caries trials commonly reported findings based on a per-protocol analysis rather than using some method to impute data for dropouts. Use of a per-protocol analysis that ignores subject dropouts risks introduction of unknown bias. For example, if a test mouth rinse, which has an unpleasant taste and stains the teeth, is compared to a pleasant tasting, nonstaining control rinse, test subjects who use the rinse regularly may be more likely to drop out of the study than irregular users because of the unpleasant side effects. This would introduce a bias against finding a favorable effect. By using data imputation methods, such as an intention to treat analysis, one may minimize the effects of unknown bias in a trial. There are many other methods available to impute missing data from dropouts that are covered elsewhere in this text.

Future caries trials, regardless of whether they score dental caries at the precavitated or cavitated level, will require refined analytical methods such as proposed at the Glasgow Conference on Dental Caries [22]. The new primary caries measures are ordinal scaled variables that are refinements of the traditional DMF scoring system. One new diagnostic system being investigated is the ICDAS method [58], which accounts for possible remineralization or "healing." Previous analytical approaches did not consider that carious lesions could be reversed, and necessarily considered observed longitudinal reversals as examiner error. Today, investigators must distinguish "real" reversals from those due to examiner error. In the future, more emphasis on examiner training and calibration will be required to estimate the frequency of these reversals during caries trials.

Clinical performance of dental materials used to replace carious tooth structure is usually judged with U.S. Public Health Service (USPHS) Guidelines or Ryge criteria developed by Cvar and Ryge in 1971 [63]. These criteria have been modified and expanded over the last 40 years for evaluation of new dental materials. Criteria in the current system include color match, quality of margins, cavosurface marginal discoloration, anatomic form, recurrence of caries around the margin of the restoration, surface texture, postoperative sensitivity, proximal contact, and occlusal contacts [64]. Calibration training materials for investigators evaluating clinical outcomes of dental materials are available online [65].

## 7 PERIODONTAL CLINICAL TRIALS

Periodontal disease is an inflammatory disease that affects the surrounding and supporting tissues of the teeth (periodontium). It is usually divided into the two broad categories of gingivitis and periodontitis. These common infectious diseases are caused by pathogenic microflora in the dental biofilm (dental plaque) that forms daily on the teeth [66]. Gingivitis, the mildest form of periodontal disease affecting 50–90% of adults worldwide [67], is readily reversible by simple, effective oral hygiene. Estimates from a large survey based on partial mouth examinations indicated that about 22% of U.S. adults had mild periodontitis and 13% had moderate or severe disease [68].

A variety of interventions have been tested in clinical trials involving gingivitis and periodontitis ranging from nonsurgical mechanical therapy and attempts to improve oral hygiene to surgical and pharmacologic approaches. Outcomes generally include indices of dental plaque and gingival inflammation for gingivitis [21] and measures of periodontal pocket depth, clinical attachment level, and radiographic measures of alveolar bone support for periodontitis. The American Dental Association (ADA) published guidelines for clinical trials of products designed to control plaque and gingivitis [21]. For ADA acceptance, a product must demonstrate efficacy in two trials, each of which must include a placebo arm and follow subjects for at least 6 months. The guidelines further specify outcome measures for gingivitis and plaque as measures of safety including effects on the oral soft tissues, teeth, toxicology, and microbiology.

The ADA also has published guidelines for periodontitis trials [69] and others have published suggestions for the conduct of periodontitis trials [70]. These recommend that new treatment methods should be compared to basic periodontal therapy consisting of thorough scaling and root planing, oral hygiene instruction, and regular maintenance care. Moreover, it has been suggested that the primary response variable should be clinical attachment level, but that it is important to document changes in probing depth since this is a meaningful measure to many clinicians [19]. Gingival inflammation and bleeding should be used as secondary response variables. Radiographic measures of disease may be useful as primary response variables if safe, reproducible, and valid methods to assess change are utilized. Microbiological monitoring should be a secondary response variable because of numerous questions concerning sampling methodology, quantitative expression of data, and meaningful interpretation in terms of relevance to disease activity. It has also been recommended that the length of periodontitis trials be a minimum of 9 months if claims of superiority or equivalency are made compared to basic periodontal therapy [19]. As with most other diseases and conditions, there is a lack of validated surrogate outcomes for periodontal clinical trials and much more research is needed in this area [8].

## 8 DENTAL IMPLANT CLINICAL TRIALS

Dental implants have been used to replace missing teeth for over 20 years [71]. Usually, implants are made of commercially pure titanium or titanium alloys. Compared to smooth or machined surfaces, roughened surfaces increase surface area and promote earlier and more complete osseointegration [72]. Most implants are placed in the alveolar bone and are termed endosseous implants. Although dental implants can be used to retain maxillofacial prostheses and provide anchorage during orthodontic treatment, this discussion is limited to situations in which implants are used to replace natural teeth.

Implant failures are classified according to time after initial placement in bone. Early failures, occurring weeks or months after the implant is placed, result from excessive surgical trauma, postoperative infection, or early micromovement of the fixture. Late failures, occurring months or years after artificial tooth or teeth are attached to the implant(s), can be caused by peri-implantitis or occlusal trauma. Peri-implantitis is a destructive inflammatory disease caused by many of the same

microorganisms associated with periodontitis of natural teeth. Occlusal trauma occurs when the biting force on dental implants is excessive, causing loss of the intimate contact between the alveolar bone and the implant. This loss of contact may manifest as decreased radiographic evidence of bone support, as a peri-implant radiolucency, or clinical mobility of the implant.

Dental implant studies have several unique features. Much of the information regarding implant success comes from case series and retrospective studies rather than randomized controlled trials. Relatively few trials have been reported that compare two or more implant treatments or designs. The success of implants depends in part on the amount and density of the surrounding alveolar bone. The highest success rates tend to be in the anterior region of the mandible [73] where there is often abundant dense cortical bone for implant anchorage. Therefore, randomization in dental implant trials should take into account jaw location of implants and other possible confounding factors such as number of implants placed, occlusal force distribution, and type of prosthesis supported by the implant(s). This is especially true for trials that have relatively few subjects (even hundreds of subjects) because randomization without appropriate stratification may result in an unequal distribution of important selection and confounding factors among the treatment groups.

Implant trials generally follow participants for 1–5 years, which is longer than most oral health clinical trials. The longer duration is dictated in part by the U.S. FDA but also by the need to evaluate longer-term stability of the bone-to-implant interface. Implant metal or surface treatments that are biocompatible over the short term may have relatively high long-term failure rates if, for example, the oxide or metal surface corrodes or fractures over time. Longer follow-up is also needed to evaluate outcomes after the prosthesis (artificial teeth) are fixed to the implant(s). A variety of outcomes are used in dental implant trials to judge efficacy, including implant and prosthesis survival, annual rate of radiographically determined bone loss, and prosthetic complications (e.g., screw loosening, implant fracture, soft tissue inflammation). While most trials track implant or prosthesis survival as a primary outcome, use of patient-centered outcomes should be considered because the ultimate goal is to improve patient function, comfort, and esthetics. It may also be important to evaluate if specific implant designs require more frequent follow-up care related to, for example, loosening or breakage of screws or other devices that attach prosthesis to implants. Because implants can remain immobile and asymptomatic until bone loss approaches their apex, trials occasionally compare crestal bone loss rates rather than implant survival. The annual rate of bone loss around a "healthy" functioning implant is less than two-tenths of a millimeter a year [74] and "failing" implants typically exhibit a higher rate of loss. By longitudinally comparing sequential, standardized radiographs over time, it may be possible to compare bone loss rates between implant designs. However, at the present time, the rate of alveolar bone loss is an unproven surrogate marker for implant survival. Although less powerful than a randomized clinical trial in terms of design and interpretation, a nonrandomized observational prospective cohort study could be used to compare bone loss to historical "normal" rates of bone remodeling. Occasionally, investigators use composite outcomes to compare implant treatments, where "events" are defined as combinations of implant loss, excessive alveolar bone loss adjacent to the implant, and persistent inflammation or pain. As with periodontitis trials, investigators have studied associations between various biochemical or microbiological measures and

implant survival [75], but there is currently insufficient data to support use of these surrogates in clinical trials. Clinical trials can also be designed to compare various treatments for peri-implantitis. Because peri-implantitis affects the soft tissue cuff surrounding the metal fixture, outcomes for these trials are similar to those used in periodontitis trials, and include probing depth, gingivitis indices, and radiographic measurements of alveolar bone.

With the increasing awareness that dental implants are a predictable treatment option for replacing missing teeth, there is a growing interest in exploring the use of dental implants as an option for addressing extensive dental treatment needs. These trials would randomly assign patients with extensive dental treatment needs to have either: (1) conventional dental treatment involving periodontal, endodontic, and restorative care or (2) tooth extraction and replacement with dental implants. For these trials, it will be necessary to evaluate patient-centered outcomes and cost–benefit ratios of various treatments using methods such as those reported by Zitzmann et al. [76].

## 9 OTHER ORAL CONDITIONS

### 9.1 Xerostomia

Saliva is an important component of the system that maintains oral health. It modulates oral microbial ecosystems, plays a critical role in preventing dental caries, lubricates oral tissues, and facilitates swallowing. Patients with decreased saliva production have an increased risk for dental caries, tooth loss, mucosal candidiasis, and decreased quality of life. While many bodily diseases are associated with decreased salivary output (or salivary flow rate), the most pronounced decreases are found in patients with Sjögren's syndrome and individuals who have received radiation to the head and neck for treatment of cancer.

Sjögren's syndrome is an autoimmune exocrinopathy typified by an intense lymphoplasmocytic infiltration of the salivary and lacrimal glands. Affected individuals have partial to complete loss of their ability to make saliva and tears, but the precise processes responsible for this loss of function are yet to be defined. Medications for Sjögren's syndrome can be classified into three groups: (1) drugs that stimulate remaining functional salivary and lacrimal tissue (cholinergic agonists) [77, 78], (2) anti-inflammatory agents that attempt to decrease autoimmune activity, thereby restoring glandular function [79], and (3) lubricating agents that improve patient comfort [79, 80]. The two agonists approved for use in Sjögren's syndrome in the United States are pilocarpine HCl (Salagen) and cevimeline (Evoxac). In most clinical trials, treatment is evaluated using subjective patient reports of changes in oral and/or ocular dryness and objective measures that quantify saliva and/or tear production. Occasionally, trials involving anti-inflammatory agents measure the amount of lymphoplasmocytic infiltration in the minor salivary glands using a gradation scale [81]. Other outcomes might include levels of serum autoantibodies and other markers of autoimmune disease that are commonly found in these patients.

Cholinergic agonists [82] and mucosal tissue lubricating agents are often used to manage patients with radiation-induced xerostomia. Only small studies have tested the efficacies of different lubricating agents for relief of oral dryness [83] and trials

to *prevent* radiation-induced xerostomia have involved amifostine [84] and other strategies to exclude the salivary glands from the radiotherapy field. These prevention trials need to evaluate patients at least one year postradiotherapy to determine long-term efficacy.

### 9.2 Mucosal Candidiasis

The oral mucosa may become infected with opportunistic fungal pathogens, including *Candida* species. These infections are most common in symptomatic HIV-infected patients, patients who have certain types of immunodeficiences (primarily T cell and phagocyte deficiencies) and individuals undergoing cancer therapy. Both systemic and topical antifungal agents have been tested and approved as therapies for mucosal candidiasis and most phase III trials have been conducted in people with HIV infection [85]. However, candidiasis trials have not used consistent methods to diagnose infection and judge cure. Moreover, erythematous mucosal lesions caused by fungal infections may be ignored because of diagnostic difficulties associated with these infections. The most widely used diagnostic criteria in Phase III trials is a positive culture for *Candida* and presence of a white patch on the oral mucosa that rubs off, leaving bleeding or redness. A "clinical cure" may be defined as the absence of clinical lesions, with or without negative culture. If the trial employs cultures as an outcome measure, a negative culture for *Candida* is termed "mycologic cure."

### 9.3 Mucosal Diseases

The two most common mucosal diseases treated by dentists are aphthous ulcerations and oral lichen planus. These diseases are believed to be mediated by infiltrating T lymphocytes; both can be idiopathic or develop after exposure to certain medications; and both may be chronic and have a major impact on quality of life. Clinical trials of treatment and prevention of these diseases usually enroll subjects with the most severe forms of disease. Since both diseases are mediated by lymphocytes, most drugs used to treat these diseases have been topical or systemic anti-inflammatory agents.

Clinical trials of oral lichen planus (OLP) usually require that patients have histopathologic evidence of OLP as an entry criterion. Various scales using objective methods to assess therapeutic response are available to semiquantify the clinical appearance and mucosal surface area affected by the lesions [86]. The primary outcome of many studies involves symptomatic scoring such as the amount of oral pain or burning sensation. The therapeutic portion of trials may be 4 weeks, with additional follow-up for one month [87, 88]. Unfortunately, most patients with symptomatic OLP of moderate severity have persistent disease; therefore, long-term outcomes (i.e., disease severity at 6 months) should be assessed.

Oral aphthae typically persist for 7–14 days after eruption. Patients with the most severe forms of the disease continually develop new lesions, which can cause significant oral pain and limit food intake. The most common outcome measures in trials for recurrent aphthous ulcerations include amount of oral pain, duration of an individual ulceration, and number of new ulcerations in a defined time period [88]. Therefore, trials should be of sufficient duration (4–6 months) to determine the impact of the therapy on development of new lesions.

## REFERENCES


1. ADA (2007), American Dental Association Policy Statement on Evidence-Based Dentistry; available February 28, 2008, http://www.ada.org/prof/resources/positions/statements/evidencebased.asp; accessed January 2009.
2. Hujoel, P. P., and DeRouen, T. A. (1992), Validity issues in split-mouth trials, *J. Clin. Periodontol.*, 19, 625–627.
3. Hujoel, P. P. (1998), Design and analysis issues in split mouth clinical trials, *Community Dent. Oral Epidemiol.*, 26, 85–86.
4. Chilton, N. W., and Fleiss, J. L. (1986), Design and analysis of plaque and gingivitis clinical trials, *J. Clin. Periodontol.*, 13, 400–410.
5. Fleming, T. R., and DeMets, D. L. (1996), Surrogate end points in clinical trials: Are we being misled? *Ann. Intern. Med.*, 125, 605–613.
6. Downing, D. G., Ed. (2000), *Biomarkers and Surrogate Endpoints: Clinical Research and Applications*, Proceedings of an NIH–FDA Conference, April 15–16, 1999, Bethesda, MD, Elsevier, New York.
7. Hujoel, P. P. (2004), Endpoints in periodontal trials: The need for an evidence-based research approach, *Periodontol. 2000*, 36, 196–204.
8. Barnett, M. L., and Pihlstrom, B. L. (2004), Methods for enhancing the efficiency of dental/oral health clinical trials: Current status, future possibilities, *J. Dent. Res.*, 83, 744–750.
9. Laster, L. L. (1985), The effect of subsampling sites within patients, *J. Periodontal. Res.*, 20, 91–96.
10. Goodson, J. M. (1986), Clinical measurements of periodontitis, *J. Clin. Periodontol.*, 13, 446–460.
11. Tu, Y. K., Gilthorpe, M. S., Griffiths, G. S., et al. (2004), The application of multilevel modeling in the analysis of longitudinal periodontal data—part I: Absolute levels of disease, *J Periodontol.*, 75, 127–136.
12. Tu, Y. K., Gilthorpe, M. S., Griffiths, G. S., et al. (2004), The application of multilevel modeling in the analysis of longitudinal periodontal data—part II: Changes in disease levels over time, *J. Periodontol.*, 75, 137–145.
13. Gilthorpe, M. S., Griffiths, G. S., Maddick, I. H., and Zamzuri, A. T. (2000), The application of multilevel modelling to periodontal research data, *Community Dent. Health*, 17, 227–235.
14. Hyman, J. (2006), The importance of assessing confounding and effect modification in research involving periodontal disease and systemic diseases, *J. Clin. Periodontol.*, 33, 102–113.
15. Schulz, K. F., and Grimes, D. A. (2002), Generation of allocation sequences in randomised trials: Chance, not choice, *Lancet*, 359, 515–519.
16. Friedman, L. M., Furberg, C. D., and DeMets, D. L. (1996), *Fundamentals of Clinical Trials*, Mosby, St. Louis.
17. Fleiss, J. L. (1986), *The Design and Analysis of Clinical Experiments*, Wiley, New York.
18. Kingman, A. (1986), A procedure for evaluating the reliability of a gingivitis index, *J. Clin. Periodontol.*, 13, 385–391.
19. Pihlstrom, B. (1992), Issues in the evaluation of clinical trials of periodontitis: A clinical perspective, *J. Periodontal. Res.*, 27, 433–441.
20. Hill, E. G., Slate, E. H., Wiegand, R. E., Grossi, S. G., and Salinas, C. F. (2006), Study design for calibration of clinical examiners measuring periodontal parameters, *J. Periodontol.*, 77, 1129–1141.

21. ADA (2007), American Dental Association Council on Scientific Affairs. Acceptance Program Guidelines Chemotherapeutic Products for Control of Gingivitis, 1997: available from http://www.ada.org/ada/seal/standards/guide_chemo_ging.pdf; accessed January 2009.
22. Pitts, N. B., and Stamm, J. W. (2004), International Consensus Workshop on Caries Clinical Trials (ICW-CCT)—Final consensus statements: Agreeing where the evidence leads, *J. Dent. Res.*, 83 (Spec No C), C125–128.
23. Beck, J. D., Lawrence, H. P., and Koch, G. G. (1997), Analytic approaches to longitudinal caries data in adults, *Community Dent. Oral Epidemiol.*, 25, 42–51.
24. Mancl, L. A., Hujoel, P. P., and DeRouen, T. A. (2004), Efficiency issues among statistical methods for demonstrating efficacy of caries prevention, *J. Dent. Res.*, 83 (Spec No C) C95–98.
25. Caplan, D. J., Slade, G. D., Biesbrock, A. R., Bartisek, R. D., McClanahan, S. F., and Beck, J. D. (1999), A comparison of increment and incidence density analyses in evaluating the anticaries effects of two dentifrices, *Caries. Res.*, 33, 16–22.
26. Hujoel, P. P., Isokangas, P. J., Tiekso, J., Davis, S., Lamont, R. J., DeRouen, T. A., and Makinen, K. K. (1994), A re-analysis of caries rates in a preventive trial using Poisson regression models, *J. Dent. Res.*, 73, 573–579.
27. Hannigan, A. (2004), Using survival methodologies in demonstrating caries efficacy, *J. Dent. Res.*, 83 (Spec No C), C99–102.
28. Hannigan, A., O'Mullane, D. M., Barry, D., Schafer, F., and Roberts, A. J. (2001), A re-analysis of a caries clinical trial by survival analysis, *J. Dent. Res.*, 80, 427–431.
29. Imrey, P. B., and Kingman, A. (2004), Analysis of clinical trials involving non-cavitated caries lesions, *J. Dent. Res.*, 83 (Spec No C), C103–108.
30. Katz, B. P., and Huntington, E. (2004), Statistical issues for combining multiple caries diagnostics for demonstrating caries efficacy, *J. Dent. Res.*, 83 (Spec No C), C109–112.
31. Pockock, S. J. (2007), *Clinical Trials: A Practical Approach*. Chapter 9. The Size of a Clinical Trial, Wiley, New York.
32. Title 21, Food and Drugs. Subchapter H, Medical Devices, Part 872 Dental Devices. http://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfcfr/CFRSearch.cfm?CFRPart=872, accessed January 2009.
33. International Organization for Standardization. ISO Standards: TC 106 Dentistry; available at: http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_tc_browse.htm?commid=51218; accessed January 2009.
34. American Dental Association Dental Products Specifications; available at: http://www.ada.org/prof/resources/standards/products_specifications.asp; accessed January 2009.
35. USFDA Center for Devices and Radiological Health Standards. Premarket Notification 510(K); available at: http://www.fda.gov/cdrh/devadvice/314.html; accessed January 2009.
36. Witham, M. D., and McMurdo, M. E. (2007), How to get older people included in clinical studies, *Drugs Aging*, 24, 187–196.
37. Walton, J. N., and MacEntee, M. I. (2008), Screening and enrolling subjects in a randomized clinical trial involving implant dentures, *Int. J. Prosthodont.*, 21:210–214.
38. Weinstein, P., Milgrom, P., and Sanghvi, H. (1995), Recruitment issues: Errors of omission in dental research, *J. Dent. Res.*, 74, 1028–1029.
39. Shaya, F. T., Gbarayor, C. M., Yang, H.-K., Agyeman–Duah, M., and Saunders, E. (2007), A perspective on African American participation in clinical trials, *Contemp. Clin. Trials.*, 28, 213–217.

40. Milgrom, P. M., Hujoel, P. P., Weinstein, P., and Holborow, D. W. (1997), Subject recruitment, retention, and compliance in clinical trials in periodontics, *Ann. Periodontol.*, 2, 64–74.

41. Weinstein, P., Milgrom, P., and Sanghvi, H. (1995), Recruitment issues: Errors of omission in dental research, *J. Dent. Res.*, 74, 1028–1029.

42. Edelstein, B. L. (2002), Disparities in oral health and access to care: Findings of national surveys, *Ambul. Pediatr.*, 2, 141–147.

43. Graskemper, J. P. (2004), The standard of care in dentistry: Where did it come from? How has it evolved? *J. Am. Dent. Assoc.*, 135, 1449–1455.

44. Yip, H. K., and Samaranayake, L. P. (1998), Caries removal techniques and instrumentation: A review, *Clin. Oral. Investig.*, 2, 148–154.

45. American Dental Association (2007), Dental Practice Parameters; available at: http://www.ada.org/prof/prac/tools/parameters/index.asp; accessed January 2009.

46. American Academy of Periodontology (2000), Parameter on plaque-induced gingivitis, *J. Periodontol.*, 71, 851–852.

47. American Academy of Periodontology (2000), Parameter on chronic periodontitis with advanced loss of periodontal support, *J. Periodontol.*, 71, 856–858.

48. American Academy of Periodontology (2000), Parameter on chronic periodontitis with slight to moderate loss of periodontal support, *J. Periodontol.*, 71, 853–855.

49. Nyman, S., Lindhe, J., and Rosling, B. (1977), Periodontal surgery in plaque-infected dentitions, *J. Clin. Periodontol.*, 4, 240–249.

50. Lindhe, J., Haffajee, A. D., and Socransky, S. S. (1983), Progression of periodontal disease in adult subjects in the absence of periodontal therapy, *J. Clin. Periodontol.*, 10, 433–442.

51. Michalowicz, B. S., Hodges, J. S., DiAngelis, A. J., Lupo, V. R., Novak, M. J., Fevguson, J. E., Buchanan, W., Bofill, J., Papapanou, P. N., Mitchell, D. A., Matseoane, S., and Tschida, P. A. (2006), Treatment of periodontal disease and the risk of preterm birth, *N. Engl. J. Med.*, 355, 1885–1894.

52. Thylstrup, A., and Fejerskov, O. (1994), Clinical and pathological features of dental caries, in Thylstrup, A., Fejerskov, O., Eds. *Textbook of Clinical Cariology*, Munksgaard, Copenhagen.

53. World Health Organization (1979), *A Guide to Oral Health Epidemiological Investigations*, WHO, Geneva.

54. Whelton, H. (2004), Overview of the impact of changing global patterns of dental caries experience on caries clinical trials, *J. Dent. Res.*, 83 (Spec No C), C29–34.

55. Wenzel, A., Pitts, N., Verdonschot, E. H., Kalsbeek, H., and Wenzel, A. (1993), Developments in radiographic caries diagnosis, *J. Dent.*, 21, 131–140.

56. Kidd, E. A., Ricketts, D. N., and Pitts, N. B. (1993), Occlusal caries diagnosis: A changing challenge for clinicians and epidemiologists, *J. Dent.*, 21, 323–331.

57. Kingman, A., and Selwitz, R. H. (1997), Proposed methods for improving the efficiency of the DMFS index in assessing initiation and progression of dental caries, *Community Dent. Oral Epidemiol.*, 25, 60–68.

58. Ismail, A. I., Banting, D., Eggertsson, H., Ekstrand, K. R., Ferreira-Zandona, A., Longbottom, C., Pitts, N. B., Reich, E., Ricketts, D., Selwitz, R., Sohn, W., Topping, G. V., and Zero, D. (2005), Rationale and evidence for the International Caries Detection and Assessment System (ICDAS II). *In:* G. K. Stookey, ed., *Clinical Models Workshop: Remin-Demin, Precavitation, Caries*. Indiana University School of Dentistry, Indianapolis, IN.

59. Yang, J., and Dutra, V. (2005), Utility of radiology, laser fluorescence, and transillumination, *Dent. Clin. North Am.*, 49, 739–752, vi.
60. Stookey, G. K. (2004), Optical methods—quantitative light fluorescence, *J. Dent. Res.*, 83 (Spec No C), C84–88.
61. Lussi, A., Hibst, R., and Paulus, R. (2004), DIAGNOdent: An optical method for caries detection, *J. Dent. Res.*, 83 (Spec No C), C80–83.
62. Longbottom, C., and Huysmans, M. C. (2004), Electrical measurements for use in caries clinical trials, *J. Dent. Res.*, 83 (Spec No C), C76–79.
63. Cvar, J. F., and Ryge, G. (2005), Reprint of criteria for the clinical evaluation of dental restorative materials. 1971, *Clin. Oral Investig.*, 9, 215–232.
64. Hickel, R., Roulet, J. F., Bayne, S., Heintze, S. D., Mjor, I. A., Peters, M., Rousson, V., Randall, R., Schmalz, G., Tyas, M., and Vanherle, G. (2007), Recommendations for conducting controlled clinical studies of dental restorative materials, *Int. Dent. J.*, 57, 300–302.
65. Clinical Evaluation of Dental Restorations (CER) for Teaching and Research, University of Michigan; available at: http://www.dent.umich.edu/cer/; accessed December 2007.
66. Pihlstrom, B. L., Michalowicz, B. S., and Johnson, N. W. (2005), Periodontal diseases, *Lancet*, 366, 1809–1820.
67. Albandar, J. M., and Rams, T. E. (2002), Global epidemiology of periodontal diseases: an overview, *Periodontol. 2000*, 29, 7–10.
68. Albandar, J. M., Brunelle, J. A., and Kingman, A. (1999), Destructive periodontal disease in adults 30 years of age and older in the United States, 1988–1994, *J. Periodontol.*, 70, 13–29.
69. American Dental Association Council on Scientific Affairs. Acceptance Program Guidelines: Determination Efficay in Product Evaluation 1999; available at: http://www.ada.org/ada/seal/standards/guide_efficacy.pdf; accessed December 2007.
70. Imrey, P. B., Chilton, N. W., Pihlstrom, B. L., Proskin, H. M., Kingman, A., Listgarten, M. A., Zimmerman, S. O., Ciancio, S. G., Cohen, M. E., and O'Agostino, R. (1994), Proposed guidelines for American Dental Association acceptance of products for professional, non-surgical treatment of adult periodontitis. Task Force on Design and Analysis in Dental and Oral Research, *J. Periodontal. Res.*, 29, 348–360.
71. Adell, R., Eriksson, B., Lekholm, U., Branemark, P. I., and Jemt, T. (1990), Long-term follow-up study of osseointegrated implants in the treatment of totally edentulous jaws, *Int. J. Oral Maxillofac. Implants*, 5, 347–359.
72. Puleo, D. A., and Thomas, M. V. (2006), Implant surfaces, *Dent. Clin. North Am.*, 50, 323–338, v.
73. Ferrigno, N., Laureti, M., Fanali, S., and Grippaudo, G. (2002), A long-term follow-up study of non-submerged ITI implants in the treatment of totally edentulous jaws. Part I: Ten-year life table analysis of a prospective multicenter study with 1286 implants, *Clin. Oral. Implants. Res.*, 13, 260–273.
74. Eliasson, A., Eriksson, T., Johansson, A., and Wennerberg, A. (2006), Fixed partial prostheses supported by 2 or 3 implants: a retrospective study up to 18 years, *Int. J. Oral Maxillofac. Implants*, 21, 567–574.
75. Paolantonio, M., Di Placido, G., Tumini, V., Di Stilio, M., Contento, A., and Spoto, G. (2000), Aspartate aminotransferase activity in crevicular fluid from dental implants, *J. Periodontol.*, 71, 1151–1157.
76. Zitzmann, N. U., Marinello, C. P., and Sendi, P. (2006), A cost-effectiveness analysis of implant overdentures, *J. Dent. Res.*, 85, 717–721.

77. Vivino, F. B., Al-Hashimi, I., Khan, Z., Leveque, F. G., Salisbury, P. L. III, Tran-Johnson, T. K., Muscoplat, C. C., Trivedi, M., Goldlust, B., and Gallagher, S. C. (1999), Pilocarpine tablets for the treatment of dry mouth and dry eye symptoms in patients with Sjogren syndrome: A randomized, placebo-controlled, fixed-dose, multicenter trial. P92-01 Study Group, *Arch. Intern. Med.*, 159, 174–181.
78. Petrone, D., Condemi, J. J., Fife, R., Gluck, O., Cohen, S., and Dalgin, P. (2002), A double-blind, randomized, placebo-controlled study of cevimeline in Sjogren's syndrome patients with xerostomia and keratoconjunctivitis sicca, *Arthritis. Rheum.*, 46, 748–754.
79. Brennan, M. T., Shariff, G., Lockhart, P. B., and Fox, P. C. (2002), Treatment of xerostomia: A systematic review of therapeutic trials, *Dent. Clin. North. Am.*, 46, 847–856.
80. von Bultzingslowen, I., Sollecito, T. P., Fox, P. C., Daniels, T., Jonsson, R., Lockhart, P. B., Wray, D., Brennan, M. T., Carrozzo, M., Gandera, B., Fujibayashi, T., Navazesh, M., Rhodus, N. L., and Schiodt, M. (2007), Salivary dysfunction associated with systemic diseases: Systematic review and clinical management recommendations, *Oral Surg. Oral Med. Oral Pathol. Oral Radiol. Endod.*, 103 (Suppl), S57.e1–15.
81. Greenspan, J. S., Daniels, T. E., Talal, N., and Sylvester, R. A. (1974), The histopathology of Sjogren's syndrome in labial salivary gland biopsies, *Oral Surg. Oral Med. Oral Pathol.*, 37, 217–229.
82. Johnson, J. T., Ferretti, G. A., Nethery, W. J., Valdez, I. H., Fox, P. C., Ng, D., Muscoplat, C. C., and Gallagher, S. C. (1993), Oral pilocarpine for post-irradiation xerostomia in patients with head and neck cancer, *N. Engl. J. Med.*, 329, 390–395.
83. Chambers, M. S., Posner, M., Jones, C. U., Biel, M. A., Hodge, K. M., Vitti, R., Armstrong, I., Yen, C., and Weber, R. S. (2007), Cevimeline for the treatment of postirradiation xerostomia in patients with head and neck cancer, *Int. J. Radiat. Oncol. Biol. Phys.*, 68, 1102–1109.
84. Shiboski, C. H., Hodgson, T. A., Ship, J. A., and Schiodt, M. (2007), Management of salivary hypofunction during and after radiotherapy, *Oral Surg. Oral Med. Oral Pathol. Oral Radiol. Endod.*, 103 (Suppl), S66.e1–19.
85. Baccaglini, L., Atkinson, J. C., Patton, L. L., Glick, M., Ficarra, G., and Peterson, D. E. (2007), Management of oral lesions in HIV-positive patients, *Oral Surg. Oral Med. Oral Pathol. Oral Radiol. Endod.*, 103 (Suppl), S50.e1–23.
86. Al-Hashimi, I., Schifter, M., Lockhart, P. B., Wray, D., Brennan, M., Migliorati, C. A., Axell, T., Bruce, A. J., Carpenter, W., Eisenberg, E., Epstein, J. B., Holmstrup, P., Jontell, M., Lozada-Nur, F., Nair, R., Silverman, B., Thongprasom, K., Thornhill, M., Warnakulasuriya, S., and van der Waal, I. (2007), Oral lichen planus and oral lichenoid lesions: Diagnostic and therapeutic considerations, *Oral Surg. Oral Med. Oral Pathol. Oral Radiol. Endod.*, 103 (Suppl), S25.e1–12.
87. Chainani-Wu, N., Silverman, S. Jr., Reingold, A., Bostrom, A., McCulloch, C., Lozada-Nur, F., and Weintraub, J. (2007), A randomized, placebo-controlled, double-blind clinical trial of curcuminoids in oral lichen planus, *Phytomedicine*, 14, 437–446.
88. Thornhill, M. H., Baccaglini, L., Theaker, E., and Pemberton, M. N. (2007), A randomized, double-blind, placebo-controlled trial of pentoxifylline for the treatment of recurrent aphthous stomatitis, *Arch Dermatol.*, 143, 463–470.

# *Dermatology Clinical Trials*

**Maryanne Kazanis, Alicia Van Cott, and Alexa Boer Kimball**
*Department of Dermatology, Massachusetts General Hospital, Boston, Massachusetts*

**Contents**

1 Introduction
2 Overall Study Design Considerations in Dermatology
  2.1 Drug Therapy Trials
  2.2 Device Studies: Phototherapy and Laser Trials
  2.3 Determining Study Size
  2.4 Patient Selection and Recruitment Methods
  2.5 Patient Compliance
  2.6 Disease and Quality of Life Measurement Instruments
  2.7 Source Documentation in Dermatology
  2.8 Clinical Trial Registration
3 Summary
  References

## 1 INTRODUCTION

Dermatology is a relatively young medical specialty [1, 2]. For many years diseases of the skin were not considered processes involving an individual organ but rather the manifestation of greater alterations within the whole human body. As a result of this school of thought, skin conditions fell under the practice of general medicine until the early eighteenth and nineteenth centuries. It was then that European physicians became increasingly curious about skin diseases and began to consider the skin as an organ that could be independently studied, diagnosed, and treated.

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

During this time, an analytical approach to dermatology emerged, and physicians increasingly relied on clinical observations to uncover cause-and-effect relationships between various exposures and skin diseases.

In America, it was not until the late nineteenth century that dermatology was established as a clinical discipline [2]. It was then that American dermatologists began to study in Europe, the academic epicenter for the specialty at the time. These physicians would learn about progress made in the field and would model their American practices after those of their European counterparts. During this period, although progress was made in establishing dermatology as a specialty in America, there was little emphasis placed on dermatological research. The scarcity of research in the field prompted the formation in 1937 of two organizations: the American Academy of Dermatology and the Society for Investigative Dermatology [2–4]. One year later, the Society for Investigative Dermatology published its first *Journal of Investigative Dermatology*. Both the creation of these societies and the publication of this journal were the first of their kind, as no previous journals or organizations had focused primarily on dermatology research.

Today, there are over 200 dermatology specialty journals and the importance of clinical research in the field is unquestionable [5]. In the last 15 years, the approach of medical practice in general has become more evidence based, and there has been a corresponding shift in clinical dermatology [6]. The evidence garnered from various research trials has significantly shaped the way physicians care for their patients, and the results of clinical research studies often guide dermatologists in treatment decision making. In this setting, the accurate design and execution of dermatology clinical trials is paramount.

Given the current diversity within the specialty, there is a large range of dermatology topics under study (Table 1). Although the majority of trials are therapeutic in nature, there are also quality of life studies and research projects investigating the response of skin disease to environmental exposures and other stressors. The structure of therapeutic trials in dermatology varies based upon what disease and treatment method is being studied. Dermatological therapeutic treatments include topical, oral, and subcutaneous drugs, phototherapy, lasers, and other devices. Each of these therapies presents unique considerations when designing and implementing clinical trials, and later these will be discussed in further detail.

Dermatology is a wonderful field for conducting clinical trials, as there are many advantages to studying the skin. First, it is an accessible organ that can be easily

**TABLE 1 Examples of Common versus Rare Debilitating Diseases in Dermatology**

| Most Common Diseases | Rare Diseases |
|---|---|
| Acne vulgaris | Epidemolysis bullosa |
| Actinic keratosis | Blistering diseases: Pemphigus and pemphigoid |
| Nonmelanoma skin cancer | Toxic epidermal necrolysis |
| Contact and atopic dermatitis | Cutaneous T-cell lymphoma |
| Cutaneous fungal infections | Icthyosis |
| Herpes simplex and zoster | Scleroderma |
| Human papillomavirus | Erythrodermic psoriasis |
| Melasma | Lichen planus |
| Psoriasis | Pityriasis rubra pilaris |
| Acne rosacea | Pyoderma gangrenosum |

seen, touched, measured, and sampled. This is a benefit not achievable with other organ systems and allows for unique treatment options, readily observed patient responses, and less invasive study procedures. Further, given the symmetry of dermatological lesions, subjects may serve as their own controls, which is advantageous since it requires the recruitment of fewer subjects to a given trial. Lastly, the duration of clinical trials is generally shorter than in other areas of medicine. Endpoints may be easily recognizable as early as 12 weeks, which makes the skin an ideal organ system to study, especially in pilot programs. Overall, these characteristics permit rapid trials with straightforward treatment schedules and endpoints. In addition, dermatology is one of the only areas of medicine in which study subjects may evaluate their specific responses to treatment, allowing investigators to collect data on patient perceptions of improvement. These data are important when considering quality of life issues and other psychosocial factors that are associated with disease states. When a clinical trial has successful findings, these unique attributes of studying the skin have the potential to not only influence overall patient care but also to greatly impact patients' self-esteem and sense of well-being [7, 8]. An additional benefit to conducting dermatology clinical trials is that diseases of the skin manifest in both adults and children, so there is a demographic of patients from which to draw a study subject population.

Of course, there are multiple challenges associated with conducting dermatology clinical trials. Unlike other medical specialties that use biological markers or other objective tests as indicators of health or disease, many dermatology measurements are inherently subjective. Two physicians may differ in their assessments of the same patient simply because of varying degrees of interpretation. This interobserver variability can potentially lead to different results reporting and reductions in standardization across study centers and clinical facilities.

A second issue involved in conducting dermatology clinical trials is the influence of market constraints on the selection of diseases being studied [6]. Given the high costs associated with conducting studies, the pharmaceutical industry often sponsors research. The focus of industry sponsors is of course affected by fiscal concerns, which sometimes means that rare, serious skin diseases are studied less frequently than more common and sometimes less debilitating disorders.

The progress made in dermatology because of clinical trials research is significant [9]. Acne vulgaris is just one example of a condition that has seen drastic treatment changes because of clinical research [10]. Over the past century, many conclusions regarding the causes and remedies for acne have been made. Previously, sulfur and mercurials were two of the topical treatments used for acne, but today new acne treatments have been derived based on the understanding of acne's pathogenesis. Systemic therapy has evolved to include the use of oral antibiotics, and a number of topical drugs including various antimicrobials and retinoids are now the mainstay of treatment.

Similar advances have been made in psoriasis treatment. The earliest therapy for this skin disease was coal tar. Although today this is rarely used as primary therapy, many preparations for psoriasis do contain coal tar. Clinical trials have provided advancements allowing topical corticosteroids and phototherapy to emerge as methods for improving psoriatic lesions. Current dermatology psoriasis therapies include highly successful intravenous and injectable biologics, many of which are antibodies.

Advances in acne and psoriasis are just two examples of the progress made in dermatology because of the results of clinical trials. The future advancement of dermatology depends greatly on the ability to demonstrate improvements in patient care through clinical research [6]. The remainder of this chapter will focus on the guidelines, considerations, and other factors involved in the effective design and operation of these studies.

## 2 OVERALL STUDY DESIGN CONSIDERATIONS IN DERMATOLOGY

The aim of this section is to give a broad overview of some of the distinctive aspects involved in designing and conducting dermatology clinical trials.

### 2.1 Drug Therapy Trials

The study design used in a dermatology clinical trial largely depends on the particular skin condition under investigation. The gold standard of therapeutic dermatology trials for common skin conditions is the randomized controlled trial [6]. Some drug studies compare placebo to a new medication of interest in order to evaluate its efficacy and safety in treating a particular disease. In dermatology clinical trials, placebos may be in the form of a vehicle (i.e., cream, ointment, lotion, foam, etc.) or in the form of a pill, subcutaneous injection, or infusion. Placebo, however, is really a misnomer when designing a study of an active topical versus its vehicle control. The effect of moisturizing the skin with any topical has beneficial effects on many skin conditions ranging from wrinkles to eczema to psoriasis and can easily result in improvements in the 20–30% range on many scoring scales. Care must be taken when designing these studies to make sure that the sample size is large enough to compensate for this effect.

In a blinded study, it is imperative that the investigator remains unaware of the treatment assignment. In trials that use placebo, avoiding unblinding is not an issue since the vehicle or other placebo can generally be made to look exactly like the actual drug under study.

However, there are a great number of dermatology trials that compare a drug already established in treating a condition to another, newer drug. In these instances, when trials are double blinded or investigator blinded, the potential exists for the subject or investigator to easily distinguish between the two drugs under study, especially in the case of topical treatments. For example, in a study comparing an ointment to a lotion, the study subject is aware of the treatment assignment based on the type of topical she or he is given. As a result, additional considerations are required in the study design to protect against investigator unblinding. In these trials, the protocol should state that study coordinators or other study staff members are responsible for drug administration and collection, drug application instruction, and any other aspects of the trial that involve transactions with the medication(s) under study. Furthermore, protocols should include guidelines for study staff to educate subjects to not discuss with the investigator the type of study medication they are using. For example, enrolled subjects should use the term “study medication” or “study drug” when discussing their participation with the investigator, rather than descriptions such as ointment, lotion, gel, and the like. In these trials,

subject and study staff education is of utmost importance in order to avoid potential investigator unblinding and bias. Study protocols should clearly outline medication dispensing, application, drug collection, and education procedures in accordance with these considerations.

Dermatologists also conduct open-label clinical trials, where both the investigator and patient are aware of the treatment assignment. By and large, open-label trials in dermatology are used to study rare diseases, as the subjects enrolled to these trials in the dermatology setting often require immediate therapy, rendering placebo use unsuitable. The guidelines mentioned above to avoid investigator unblinding do not apply here, given the nature of the study design. However, it is of utmost importance that the investigator remain as impartial as possible when measuring outcomes in this patient population, so as to avoid biasing study results. Whenever possible, the same investigator should be involved in the evaluation of disease throughout an open-label trial in order to minimize additional differences resulting from interobserver variability. Scoring scales, photography use with subsequent evaluation by a blinded observer, and patient perception of improvement assessments may play particularly important roles in these studies. Another valuable element to ensure objectivity in open-label trials is to have multiple assessors conducting subject evaluations or to design a trial where the treating physician is different from the investigator appraising subjects' response to treatment. These features should be incorporated into the study design where applicable to maintain impartiality.

### 2.2 Device Studies: Phototherapy and Laser Trials

The earliest trials involving lasers and phototherapy compared actual therapy to simulated treatment for placebo comparison. However, today very few, if any, trials are done using sham light and laser sources since the efficacy of phototherapy and laser treatment to ameliorate certain disease symptoms has been well established. Most current studies done in these areas investigate whether a particular light source or laser type is superior to another for a given indication, whether a light source or laser is better than no treatment, or whether a light source or laser type works better with or without adjuvant drug therapy. The latter typically involve placebo use for comparison. The double-blind or investigator-blinded study design frequently involves randomizing subjects to treatment regimens and assessing efficacy and monitoring adverse effects of treatment. For example, one group of subjects might be randomly assigned to treatment with conventional ultraviolet B (UVB) phototherapy and compared to another group of subjects that undergoes narrow-band UVB phototherapy to assess which offers better disease response and patient safety. In phototherapy trials of this type, the risk of investigator unblinding is rarely an issue since technicians typically administer light treatment to subjects. This is useful since the investigator may remain unaware of the subjects' treatment assignment and conduct assessments in an unbiased manner. Laser clinical trials differ in this respect since, especially in new studies, the investigator is often involved in the treatment of subjects. Therefore, one way to avoid potential bias in the design of these trials is to state that the treating physician should not be the investigator who is conducting subject skin assessments. Additionally, photography may be incorporated into clinical trials of this type to provide objective evidence of skin modifications.

Some laser and phototherapy studies use symmetrical target lesions on the same subject for the purpose of left–right comparisons. Left–right comparisons involve treating two similar target areas on opposite sides of the body and comparing outcomes. For example, if an investigator is interested in seeing if a topical drug works better than placebo in combination with a light source, a double-blind study might be done where one side of the body is treated with a placebo vehicle, and the opposite side of the body is treated with the topical drug. Both areas would be exposed to the light source and outcomes would be measured to see if the drug did increase the efficacy of the phototherapy treatment. Although this method requires the recruitment of fewer individuals and is advantageous for gathering data on how two particular treatments affect the same subject, this approach does allow for greater chance of error. Subsequently, the study staff must take great care to meticulously record accurate drug/placebo application, outcome measurements, and drug safety assessment information. Protocols using this study design should contain guidelines to avoid error, and source documents should be thoughtfully designed and well labeled in order to capture correct information.

Additionally, there are open-label trials done with phototherapy and lasers in dermatology. For example, some of the light studies involving biologics are open label. Similar to therapeutic trials that use open-label design, these studies typically involve subjects with more advanced disease and require investigators to be as impartial as possible in their subject assessments. Photography and other more objective methods of recording outcomes may be useful in these trials to correlate with physician assessments, and these should be considered when designing such a study.

Finally, for all dermatology studies, regardless of therapy type, timelines for follow-up visits should be outlined in the study protocol and occur at the given time intervals or within the "visit window" identified in the study protocol. The visit window is the amount of time around a designated study visit that a subject has to complete the visit without it being considered a protocol deviation. For example, if the protocol designates week 2 as occurring on day 15 and there is a 3-day visit window, then the day 15 visit can technically occur anytime between days 14 and 16 and still be within study protocol. The same holds true for the laboratory assessments for a particular visit. If, for example, blood specimens were not able to be obtained at the week 2 visit, and it occurred on day 14, then the subject has until day 16 to return to the clinic for a blood draw before being considered a protocol deviation.

Although not obtaining all study data at a given visit is not advisable, on occasion it may be unavoidable. At follow-up visits, all medical exams, investigator assessments, subject questionnaires, laboratory collection, and redispensing of study drug should occur as outlined in the study protocol. If a study drug could potentially be associated with reproductive risks, the protocol should state that women of reproductive age have regular urine pregnancy tests done while on the study and that subjects of both sexes who are sexually active report their method of contraception. In addition, adverse events and concomitant medication taken should be reviewed with the study subject. A copy of the subject's diary should be taken to include as a part of the source documentation if it is to be redispensed to the subject and proper study drug administration instructions should be reviewed with the subject. It is advisable to ask subjects how they take their study drug in order to eliminate

study staff coaxing and ascertain exactly how subjects are taking their study drug. It is also advisable to include a study checklist in the source documentation to verify that all study-related information is collected at follow-up visits.

### 2.3 Determining Study Size

One of the first steps in developing a clinical trial is deciding what number of subjects to enroll. Preliminary studies typically offer a good starting point in this process by providing information on the anticipated size needed to obtain a desired outcome [11, 12]. However, when deciding upon a final sample size, thought must be given to both statistical and clinical significance.

For clinical trials that address common skin conditions and diseases, patient recruitment is typically not a difficult issue, and obtaining large sample sizes for trials is uncomplicated. However, dermatologists conducting these trials should consider that too large a sample size may lead to a situation of a statistically significant finding that holds no real clinical relevance [11]. This is because as sample sizes grow exceedingly larger, even the smallest differences that are not necessarily useful to patient care may be detected. Further, in dermatology it is especially important to prospectively consider the hypothesized central outcome of a trial, given that patient perceptions of improvement play a significant role in evaluating clinical significance. For example, a study that has the statistical power to show that drug A for acne is slightly better than drug B will only really be clinically useful if the subjects on trial felt that their acne indeed improved on drug A [12, 13].

Conversely, for rare, more serious dermatological diseases, the opposite challenge exists. Recruiting adequate numbers of study participants within a reasonable time frame proves to be logistically and economically difficult [11]. Often, these smaller trials may fail to statistically conclude that a relationship exists, even if one does, simply because the study sample size is not large enough. As a result, it is often the most devastating diseases that are the hardest to research. Investigators conducting trials with these patient populations should be aware of the challenges and thoughtfully plan subject recruitment strategies to ensure adequate subject enrollment.

Similar problems exist for detecting differences in rare outcomes or uncommon side effects between groups. For example, even if a sample size is large enough to determine efficacy, it may still be too small to ascertain useful data on safety and dosing schedules [11]. Therefore, it is necessary in these instances to appropriately design a trial that will have a large enough number of subjects to ascertain reliable safety and outcome results.

### 2.4 Patient Selection and Recruitment Methods

At the outset of a clinical trial, it is important for the investigator to define a study's target population. In dermatology, a study population is defined in terms of any combination of the following characteristics: age, gender, race, complexion, skin phototype (type I pale white skin to type VI dark brown or black skin), previous photodamage, duration of disease, and severity of disease. The disease under study dictates which characteristics are of significance. Of note, several major dermato-

logic diseases occur predominantly or prominently in the pediatric population ranging from eczema to acne. Designing studies for this age group, of course, requires additional considerations, ranging from assenting patients, minimizing discomfort, and designing ethically appropriate studies for this age group.

All required baseline characteristics should be clearly outlined in the protocol. This is imperative in understanding which treatments work in a given patient population and is especially useful when translating study results to the patient care setting since a more standardized subject population yields a greater ability to detect true differences in study outcomes. Scoring instruments, which will be discussed in later sections, are used to grade subject disease severity and are helpful tools that may help determine which subjects should be enrolled to a particular trial. Protocols should state exactly which scoring scale or grading method should be used to assess disease at baseline and throughout the study.

Inclusion/exclusion subject criteria should contain directives regarding the use of other therapies to treat the skin disorder under study. Typically, studies require no use of such medications for a period of 14–30 days before the start of study drug. This time frame is often referred to as the "wash-out" period. The wash-out period ensures that all subjects enrolled to a trial have comparable baseline off-treatment characteristics. Enrollment criteria should also include information regarding past therapies used to treat the skin disorder under study. In some cases, it may be necessary to exclude a subject from a study if it is deemed that a prior therapy used could affect study participation and outcomes. Additionally, a protocol should specifically state if moisturizers, cleansers, or other skin products may be used by enrolled subjects. This is because various emollients and cleansers contain ingredients that may improve certain dermatological conditions and potentially affect clinical outcomes.

Dermatology is unique in that people who have skin conditions are able to identify their own diseases. As a result, subject recruitment outside of the clinic setting is a common and successful approach. Unlike other areas of medicine that recruit subjects from clinic patient populations, dermatology clinical trials are often advertised to the general public in the newspaper or in flyers. This method provides a great advantage to subject enrollment strategies since it reaches a much greater sample population.

Given that the sample population is from the general public, potential subjects who respond to trial advertisements must be screened thoroughly. Their medical histories and medications must be completely reviewed prior to enrollment to assure participant safety and subject eligibility. In addition, it is the responsibility of the investigator and study staff to assess subject compliance. For phototherapy trials, especially, where subjects are required to undergo treatments as often as three times per week, it is imperative that participants are willing to cooperate with study protocol procedures. Monetary or other incentives may be offered to increase patient recruitment and compliance. However, these incentives as well as any advertisements used in the study must be approved by the ethics committee of the participating institution to ensure that they coincide with federal regulations.

Regardless of how the subject is recruited, she or he must sign an informed consent form (ICF) before any study procedures are performed, as mandated by 21 Code of Federal Regulations, CFR 50.20. This may or may not occur at the screening visit. Subjects must be given ample time to consider if they want to participate in

the research study and should be allowed time to discuss the study with their physician and families. Furthermore, the ICF should clearly outline any potential risks associated with treatment, including reproductive risks for men and women.

### 2.5 Patient Compliance

Monitoring patient compliance in clinical trials is essential to the acquisition of real outcome data and the conduction of a sound study [14–16]. It is imperative that subjects adhere to protocol visit schedules and that they comply with a stated treatment program. Prior to enrollment, subjects are asked to read the informed consent form, which describes what is entailed in study participation and gives information on potential benefits and risks. Subjects have the opportunity to ask questions and are only enrolled if they are amenable to study procedures. In most dermatology drug trials, studies are designed to be brief with few visits. This typically translates to increased patient compliance because it is less intrusive to the lives of subjects. However, for phototherapy trials, assessing and maintaining patient compliance becomes more of an issue since subject lives may be affected more by numerous, frequent study visits.

There are several ways to maintain an acceptable level of compliance among subjects. First, the simple act of calling patients to confirm appointments helps to remind them about their study visits. Second, short wait times and efficient execution of study visits help to make participation less of a chore for subjects. It is important to consider that subjects have issues that come up that may preclude them from maintaining a strict study schedule. Although these should be kept to a minimum, exceptions can be made for patient scheduling conflicts, and these should be reported to the investigational review board (IRB) of record and/or the study sponsor.

In addition to ensuring that subjects are seen at visits according to protocol, it is also important to make certain that they are properly using their study medication. This involves constant education of study subjects at visits. In dermatology clinical trials, for example, application techniques for topical drugs should be reviewed at each visit. Protocols must clearly and specifically outline how subjects should use medication and emphasize drug application reeducation by study staff. Whenever possible, especially in the case of topical and injected drugs, the first treatment should take place in the clinic so that proper dosing and drug administration practices may be established. In addition, for oral, subcutaneous, and topical treatment administrations that subjects perform at home, drug compliance logs should be completed by the subjects and collected at each visit in order to monitor patient adherence. Study protocols should specify that subjects return all medication at the end of study or at stated intervals throughout the trial to assess consumption. For oral and subcutaneous drugs, this is useful in ensuring that patients did indeed take the medication. An added benefit in collecting study medication for trials involving topical drugs is that bottles and containers may be weighed and the actual amount of drug used by each patient may be collected as data. Typically, drug containers are measured at dispensing, then again when returned by the subject at a time interval determined in the protocol. The difference in the two weights yields the amount used by the subject. When comparing patient outcomes, this information is valuable in evaluating subject responses in relation to the amount of drug applied to the skin.

Additionally, this measurement is helpful for assessing drug safety and efficacy since too much of a study drug could cause adverse reactions, while too little may be associated with inefficacy.

Subjects must also be instructed on the amount of time they have to take a dose of study drug, if one is missed. This is referred to as the "dosing window." It is important that subjects are instructed to call the study staff if they do miss a dose of study drug and to record the missed dose in a study diary. In addition, the protocol should indicate that subjects should be instructed to record in their study diary any adverse events experienced and concomitant medications taken. The diary should be brought with each subject to all study visits. All of the above instructions should be reviewed at all study visits with the subject to make sure that the study protocol is being followed.

Common adverse events experienced in topical clinical trials include erythema, pruritus, irritation, and scale at application sites. These events are expected with most topicals in a small to moderate percentage of patients depending on the ingredients being studied. Some studies will therefore assess for these changes at each visit proactively and separately from the other adverse events in order to ensure uniformity of data collection.

It is important to note that study subjects may differ from "real" patients in that they are more motivated to adhere to treatment schedules and visits [6]. Monetary incentives combined with a general interest in seeking out investigational treatment make subjects in research studies a typically more compliant patient population. For dermatologists involved in clinical trials, it is useful to keep this in mind since a treatment or visit schedule that is inconvenient for study subjects has a high probability of translating into decreased acceptance by dermatology patient populations.

## 2.6 Disease and Quality of Life Measurement Instruments

Capturing appropriate data for dermatology clinical trials almost always involves the use of scoring instruments (Tables 2 and 3). These have been developed, tested, and established as reliable ways to obtain data points [17]. Scoring instruments serve

**TABLE 2 Common Scoring Systems Used in Dermatology Clinical Trials**

| Disease | Scoring System |
|---|---|
| Acne vulgaris | Acne lesion counts (papules, pustules, comedones) |
| | Leeds acne grading technique |
| | Pillsbury scale |
| | Cook's acne grading scale method |
| Atopic dermatitis | Scoring for atopic dermatitis (SCORAD) |
| | Eczema area scoring index (EASI) |
| Psoriasis | Psoriasis area and severity index (PASI) |
| | Psoriasis severity index (PSI) |
| | Lattice system—Physician's global assessment (LS-PGA) |
| | Overall lesion assessment (OLA) |
| Acne rosacea | National Rosacea Society Rosacea Clinical Scorecard |
| Nonspecific skin disease scoring instruments | Physician's global assessment (PGA) |
| | Dermatology index of disease severity (DIDS) |

**TABLE 3 Acne Grading Sytems**

| Name of Grading System | Scale | Method |
|---|---|---|
| Leeds acne grading technique | Number of inflammatory and noninflammatory lesions present | Based on counting lesions and grouping them into noninflammatory and inflammatory categories |
| Pillsbury scale | 1 (mildest) to 4 (severe) | Based on number of inflammatory and noninflammatory lesions |
| Cook's acne grading scale method | 0 (least severe) to 8 (most severe) | Uses photography to document acne severity |
| Global acne grading system | Grade: 0 (no lesions) to 4 (≥ 1 nodule)<br>Global score: 0 (no acne) to ≥39 (very severe acne) | Takes into account location of lesions, and each location is given a factor rate. Factor rate *X* grade = local score. The Global score = sum of local scores. |
| Grading scale for overall disease severity | 0 (clear, no inflammatory lesions) to 6 (severe, numerous comedos, papules, and pustules, with larger inflamed lesions extending over much of the face, erythema may be pronounced) | Based on descriptive definitions of disease severity. Investigator selects the number that most closely identifies the extent of the subject's acne. |

as standardized assessment methods to assure the reliability of clinical trials. The use of grading scales offers better consistency in reporting subject responses and allows for direct comparisons among clinical sites and studies. Global assessments, severity scores, and/or other more specific assessments are used in dermatology clinical trials. In general, global assessments are conducted by the study investigator at regular intervals to evaluate the response of total body disease. Severity scales are used to monitor changes in disease states such as pigmentation, pruritus, and induration that correlate with certain levels of disease involvement. Dermatology scales and scoring systems frequently incorporate body surface area (BSA) calculations in order to provide more accurate descriptions of the extent of disease.

The disease under study dictates the use of a particular scoring system. For example, there is the Psoriasis Area and Severity Index (PASI) for psoriasis and the Scoring for Atopic Dermatitis (SCORAD) for atopic dermatitis. These disease-specific scoring instruments will be considered in detail in later sections. Because many of the scoring systems are subjective, the most reliable way to maintain consistency in disease measurement is to adequately educate investigators and study staff in reporting guidelines and to ensure that within each research site, whenever possible, the same investigator conducts subject assessments throughout the trial. Keeping the number of evaluators to a minimum increases the integrity of reporting due to less interassessor variability.

In addition to disease-scoring systems, dermatology clinical trials often monitor patient perceptions of improvement, patient global disease assessments, and quality of life data [18]. These instruments should be administered at baseline and at the end of study or at another study time point. The administration schedule of these

instruments and the determination of which measurements are used should be clearly indicated in the protocol.

Instruments that assess subject perception of disease vary in the types of questions asked since the data collected depends upon which disease is under study. Generally, questions ask subjects to grade the severity of their disease and rate the level of disease-associated symptoms they have experienced. However, to measure how the life activities of subjects are affected by skin disease across disease states, the majority of therapeutic trials in dermatology use the Dermatology Life Quality Index (DLQI) [19]. The DLQI is not disease specific, and subjects answer by choosing "very much," "a lot," "a little," "not at all," or "not relevant" in response to 10 questions. Topics on the questionnaire range from how skin disease affects work and extracurricular activities to interpersonal relationships and sexual activity. For pediatric trials that enroll subjects between the ages of 5 and 16 years, there is the Children's Dermatology Life Quality Index (CDLQI), which includes questions regarding how skin disease affects friendships, school work, sports, and other pediatric life issues.

Although not as widely used, the Skindex-29 is another measurement of skin disease's impact on the quality of life. Skindex-29 is a shorter version of the original 61-question Skindex quality of life questionnaire. The newer version has been shown to be easier and faster to complete and a more effective measure of patients' quality of life than its earlier form. It is composed of 30 questions that address three subscales: emotion, function, and symptoms. Patients answer questions on a scale of "0 = Never" to "4 = All the time," and the higher the score, the greater impact the disease has on quality of life.

Typically, these quality of life measurements are administered at baseline and then again after an interval of treatment in order to assess whether a given therapy has improved subjects' quality of life (Table 4). Given that dermatological diseases are visible, subjects often have stress related to how people respond to them and interact with them. This can have a major affect on how patients feels about themselves and their disease. Consequently, the quality of life measurements are extremely useful methods for predicting how successful a particular treatment is at ameliorating the psychosocial aspects of having skin disease [19].

## 2.7 Source Documentation in Dermatology

The importance of meticulous data collection in clinical trials cannot be overemphasized. The design of source documents that capture all necessary information for a given trial is vital to the successful execution of a clinical study. This section explains several types of source documents that are useful in conducting dermatology clinical trials. Given the visual characteristic of the specialty and the limited availability of objective testing methods for monitoring disease, dermatology clinical trials typically rely on diagrams, photography, and the use of target area measurements in documenting subject response to treatment.

Target area measurements are typically used to monitor the response of one area of the skin. At baseline, the investigator determines a target area and measures its dimensions using a ruler. The dimensions of the target area are recorded at baseline and at subsequent study visits as a way of monitoring the skin's response to therapy. It is important to designate whether the greatest dimensions should be measured

**TABLE 4 Quality of Life Metrics Used in Dermatology Clinical Trials**

| | |
|---|---|
| Facit fatigue scale | Measures how skin disease affects a subject's energy level and ability to complete daily activities. |
| Beck depression inventory | Measures how skin disease affects a subject's appetite, appearance, sexual desire, health, and weight. |
| EuroQol-5D (EQ-5D) | Measures how skin disease affects a subject's mobility, self-care, usual activities, level of pain/discomfort, and anxiety/depression. |
| EuroQol-Visual analog scale | Feeling "thermometer" where patients rate subjective health status. |
| Dermatology life quality index (DLQI) | Measures how skin disease affects interpersonal relationships, sexual activity, work, and extracurricular activities. |
| Children's dermatology life quality index (CDLQI) | Measures how skin disease affects friendships, school work, participation in sports, and other pediatric life issues. |
| Psoriasis disability index<br>Psoriasis index of quality of life (PSORIQoL)<br>Psoriasis life stress inventory | Measure how psoriasis affects the quality of life, level of discomfort/disability and associated stress of subjects with this skin condition. |
| Skindex-29 | Measures the effect of skin disease on subjects' emotions, symptoms, and day-to-day functioning. |
| Quality of life index for atopic dermatitis (QoLIAD) | Measures the effect of atopic dermatitis on subjects' work and school activities, social functioning, and self-esteem. |
| Melasma quality of life scale (MELASQOL) | Measures the effect of melasma on subjects' daily functioning and level of disease-associated impairment. |
| Dermatology-specific quality of life instrument for contact dermatitis (DSQL-CD) | Measures how contact dermatitis affects subjects' level of discomfort, interpersonal functioning, self-care activities, work or school performance, and self-esteem. |
| Dermatology-specific quality of life instrument for acne DSQL-acne | Measures how acne vulgaris affects subjects' level of discomfort, interpersonal functioning, self-care activities, work or school performance, and self-esteem. |

as lesions may not be uniformly shaped. Monitoring target areas offers the added benefit of assessing other changes in the skin, namely adverse reactions to therapy. Any changes seen in the skin in the target area may be documented by the investigator in writing and followed over the course of a trial.

A major limitation of using target lesions when other areas of the body are also being treated is that they may not capture what is happening clinically elsewhere for a patient. Some skin diseases respond differentially to treatment based on body region. For example, psoriasis on the legs is often substantially harder to treat than other body regions and may lag in terms of improvement.

Body diagrams are a useful method for capturing the location of disease involvement. Investigators illustrate affected skin by using a writing instrument to shade in areas of disease involvement on a blank body diagram. By completing new front and back body diagrams at baseline and at regular intervals throughout a study, investigators may monitor a subject's overall response to disease and have a better clinical picture as to how different body areas respond to a particular treatment.

Photography plays a similar role in monitoring treatment response but warrants additional considerations regarding consistency and subject confidentiality. Study

protocols that incorporate photography should include specific guidelines explaining exactly how photographs should be taken for the trial. An outline of photography procedures should contain parameters to be used for all subjects at all study sites. For example, to maintain consistency the same backdrop should be used for all photos, a color chart and patient identification card should be in the photograph's field of view, the same film, camera, and camera focal length should be used for all photos, the same number of photographs should be taken at the same study intervals, and all subjects should be dressed and positioned for photographs similarly. In addition, appropriate measures must be taken to protect subject confidentiality. Photography of subjects' faces should be avoided unless this area is being monitored for study treatment response. Only designated study staff and the study sponsor (when applicable) should have access to the photographs. In cases where a subject's face is photographed, no subject photos should be used in publications prior to the removal of identifying characteristics, for example, the blacking out of a subject's eyes. Any use of photography in a clinical trial must be included in the study's informed consent form, as it is never appropriate to take photos of study subjects without their understanding and written permission.

### 2.8 Clinical Trial Registration

In January 2005, the International Committee of Medical Journal Editors (ICMJE) issued a statement requiring all phase II, III, and IV clinical trials to be registered in order to be considered for publication in any of their 11 member journals [20, 21]. This directly affects the dermatology clinical trials process since the major dermatology journals are associated with the ICMJE.

The ICMJE provided requirements for the registry database, including that it must offer free public access, must be supervised by a not-for-profit organization, have the ability to be searched electronically, and have a mechanism to confirm registration data legitimacy. The purpose for establishing such a registry is to prevent the discerning coverage of clinical trial results in medical journals [20]. In the past 50 years, nearly half of all clinical studies have not been published, leading to incomplete knowledge about therapeutic effectiveness [21]. Some of this stems from the preference of medical journals to publish positive results, which demonstrate the effectiveness of a new treatment, or those that show noninferiority, with two treatments showing equivalent efficacy. This bias consequently has lead to the underpublication of negative or inconclusive findings and to the nondissemination of these study conclusions. The concern is that this underreporting leads to incomplete understanding of therapeutic efficacy, potential duplication of research efforts, and reduced integrity in study design and methodology. The latter issue is of considerable importance since it is paramount that investigators maintain the original study question and aims throughout a clinical trial. Registration would ensure increased veracity by precluding the potential for outcomes to be altered as data are interpreted.

## 3 SUMMARY

Collectively, the prevalence of skin problems surpasses other medical conditions such as hypertension, obesity, and cancer [22, 23] and has profound effects on health

and quality of life. What makes the skin such a unique and interesting organ to study in clinical trials is that the endpoints are often easily and quickly discernable and that the therapeutic modalities are so varied, ranging from topicals to immunotherapy to devices. The overall relative health of many skin disease patients also reduces the chances of confounding illnesses affecting the safety or interpretation of events in studies. In general, the field has also made tremendous progress in terms of methodology and scoring systems, and advances in imaging techniques will likely improve our ability to detect and measure change in the future. Lastly, while many studies, of course, do not succeed, we have seen tremendous successes in the past decade, with subjects experiencing dramatic improvements in both their health and quality of life, an experience that is greatly gratifying for investigators and subjects alike.

## REFERENCES


1. Stuttgen, G. (1996), Historical observations. Dermatology, *Clin. Dermatol.*, 14(2), 135–142.
2. Baer, R. L. (1994), Historical overview of the evolution of investigative dermatology (1775–1993), *J. Invest. Dermatol.*, 103(1), 3–6.
3. Holubar, K., and Wolff, K. (1989), The genesis of American investigative dermatology from its roots in Europe, *J. Invest. Dermatol.*, 92(4 Suppl), 14S–21S.
4. Potter, B. S. (2003), Bibliographic landmarks in the history of dermatology, *J. Am. Acad. Dermatol.*, 48(6), 919–932.
5. Delamere, F. M., and Williams, H. C. (2001), How can hand searching the dermatological literature benefit people with skin problems? *Arch. Dermatol.*, 137(3), 332–335.
6. Williams, H. (2001), Dowling oration 2001. Evidence-based dermatology—a bridge too far? *Clin. Exp. Dermatol.*, 26(8), 714–724.
7. Ben-Gashir, M. A., Seed, P. T., and Hay, R. J. (2004), Quality of life and disease severity are correlated in children with atopic dermatitis, *Br. J. Dermatol.*, 150(2), 284–290.
8. Weiss, J., Shavin, J., and Davis, M. W. (2003), Improving patient satisfaction and acne severity in patients with mild to moderate acne: The BEST study, *Cutis.*, 71(2 Suppl), 3–4.
9. Katz, S. I. (2001), Dermatological research in the 21st century: Our fantastic future, *J. Dermatol.*, 28(11), 599–601.
10. Witkowski, J. A., and Parish, L. C. (2001), Reflections on dermatology and projections for the 21st century, *Clin Dermatol.*, 19(1), 31–34.
11. Bhardwaj, S. S., Camacho, F., Derrow, A., Fleischer, A. B., Jr., and Feldman, S. R. (2004), Statistical significance and clinical relevance: The importance of power in clinical trials in dermatology, *Arch. Dermatol.*, 140(12), 1520–1523.
12. Lachin, J. M. (1981), Introduction to sample size determination and power analysis for clinical trials, *Control. Clin. Trials*, 2(2), 93–113.
13. Sheps, S. (1993), Sample size and power, *J. Invest. Surg.*, 6(6), 469–475.
14. Carroll, C. L., Feldman, S. R., Camacho, F. T., and Balkrishnan, R. (2004), Better medication adherence results in greater improvement in severity of psoriasis, *Br. J. Dermatol.*, 151(4), 895–897.
15. Zaghloul, S. S., Cunliffe, W. J., and Goodfield, M. J. (2005), Objective assessment of compliance with treatments in acne, *Br. J. Dermatol.*, 152(5), 1015–1021.

16. Zaghloul, S. S., and Goodfield, M. J. (2004), Objective assessment of compliance with psoriasis treatment, *Arch. Dermatol.*, 140(4), 408–414.
17. Gelmetti, C., and Colonna, C. (2004), The value of SCORAD and beyond. Towards a standardized evaluation of severity? *Allergy*, 59(Suppl 78), 61–65.
18. Feldman, S. R., Fleischer, A. B., Jr., Reboussin, D. M., Rapp, S. R., Exum, M. L., Clark, A. R., and Nurre, L. (1996), The self-administered psoriasis area and severity index is valid and reliable, *J. Invest. Dermatol.*, 106(1), 183–186.
19. Lewis, V., and Finlay, A. Y. (2004), 10 years experience of the dermatology life quality index (DLQI), *J. Investig. Dermatol. Symp. Proc.*, 9(2), 169–180.
20. Deangelis, C. D., Drazen, J. M., Frizelle, F. A., DeAngelis, C. D., Drazen, J. M., Frizelle, F. A., Haug, C., Hoey, J., Horton, R., Kotzin, S., Laine, C., Marusic, A., Overbeke, A. J., Schroeder, T. V., Sox, H. C., and Van Der Weyden, M. B. (2005), Clinical trial registration: A statement from the international committee of medical journal editors, *Arch. Dermatol.*, 141(1), 76–77; discussion 75.
21. Kimball, A. B., and Weinstock, M. A. (2005), Mandatory registration of clinical trials: A major step forward for evidence-based medicine, *J. Am. Acad. Dermatol.*, 52(5), 890–892.
22. Johnson, M. L. (2004), Defining the burden of skin disease in the united states—a historical perspective, *J. Investig. Dermatol. Symp. Proc.*, 9(2), 108–110.
23. The Lewin Group (2005), *The Burden of Skin Diseases*, Lewin Group, Falls Church, VA.

# *Emergency Clinical Trials*

**Joaquín Borrás-Blasco,[1] Andrés Navarro-Ruiz,[2] and Consuelo Borrás[3]**

[1]*Pharmacy Service, Hospital de Sagunto, Sagunto, Spain*
[2]*Pharmacy Service, Hospital General Universitario de Elche, Elche, Spain*
[3]*Department of Physiology, University of Valencia, Valencia, Spain*

**Contents**

1 General Information on Clinical Trials
2 Clinical Trials Involving Emergency Patients
  2.1 Subject Recruitment
  2.2 Emergency Clinical Research in Pediatric Population
3 Ethical Issues in Emergency Clinical Trials
  3.1 Informed Consent
4 Types of Clinical Trials in Emergency Departments
  4.1 Cardiovascular Alterations
  4.2 Central Nervous System
  4.3 Sepsis
5 Conclusions
  References

## 1 GENERAL INFORMATION ON EMERGENCY CLINICAL TRIALS

Emergency medicine is a global discipline that functions as a cornerstone for secondary disease prevention; therefore it is one of the many tools for implementing primary disease prevention programs [1].

In general, clinical trials involve the administration of the study drug to healthy human volunteers or to patients under the supervision of a qualified investigator, usually a physician, pursuant to a reviewed protocol. Human clinical trials are typically conducted under protocols that detail the objectives of the study, as well as

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

the parameters used to monitor the safety and efficacy criteria that are going to be evaluated. Moreover, it is very important to recognize that each clinical trial must be conducted under the auspices of an institutional review board (IRB) that considers, among other factors, ethical factors, the safety of human subjects, the possible liability of the institution, and the informed consent disclosure that must be obtained from participants of the clinical trial. Although the goal of this is to obtain safety and efficacy data, the overriding consideration in the studies is the safety of the study participants.

Research in the specialty of emergency medicine continues to blossom because this department is often considered as a desirable place to conduct clinical research due to the broad and undifferentiated spectrum of acute conditions encountered [2]. For this reason, there are some specific aspects when studying emergency medicine that have to be taken into account. One of them is the emphasis on ultra-rapid diagnosis and treatment to save the critically ill emergency patient [3]. Moreover, it is necessary to evaluate the use of a particular treatment; clinicians have to establish the parameters to be evaluated at the onset. These parameters include objective response rate, survival, disease-free survival, or duration of response. Toxicity is also commonly identified. The parameters of interest for a specific trial design are defined before the initiation of the study and analyzed at the completion.

Another characteristic of emergency clinical trials is that usually physicians working in the emergency department do not stay for a long period of time, and therefore it is especially critical for them to standardize the protocol to be used, as well as the ease of using it, for the physicians who are incorporated in the supervision of the trial [4].

Human clinical trials typically are conducted in three sequential phases, although sometimes they are overlapped. Phase I clinical trials represent the initial administration of the investigational drug to a small group of healthy human subjects or, more rarely, to a group of selected patients with the targeted disease or disorder [5]. Normally, this type of clinical trial is not developed in the emergency setting. Phase II clinical trials involve a small sample of the actual intended patient population and seek to assess the efficacy of the drug for specific targeted indications, to determine dose–response and the optimal dose range, and to gather additional information relating to safety and potential adverse effects. In the emergency room, these types of clinical trials involve mainly cardiovascular diseases such as acute ischemic stroke, acute stroke, cardiopulmonary resuscitation, acute myocardial infarction, and refractory angina patients with left ventricular dysfunction [6–9].

Once an investigational drug is found to have some efficacy and an acceptable safety profile in the targeted patient population, phase III clinical trials are initiated to establish further clinical safety and efficacy of the investigational drug in a broader sample of the general patient population. Phase III trials are randomized comparisons of two or more treatment options, often comparing a standard treatment to a new or more complex therapy. Emergency research phase III trials are related to several clinical conditions such as myocardial infarction, acute coronary syndromes, surgery, chronic obstructive pulmonary disease, hemorrhages, surgical pain, sepsis, and traumatic brain injuries [10, 11].

Phase IV trials are generally considered as "postregistration" trials, that is, trials of products that have already a marketing authorization.

It is also very important for the development of clinical research in emergency medicine to design and draw anticipated conclusions in order to note whether the trial is prospectively randomized with concurrent controls or is a clinical trial with historical controls. Proponents of randomized trials believe that one is more certain of equality between the two groups by a concurrent randomization process. This will reduce the bias of selecting control subjects from a historical pool and will also reduce the influence of improvements in management or different physicians who treat the patients in different periods of time. Furthermore, due to the specific characteristics of these clinical trials in emergency medicine, all of them require a special review and approval of trial proposals by an ethics committee.

In conclusion, there are some specific aspects that have to be taken into account when conducting clinical trials in emergency medicine due to the special conditions mentioned above.

## 2 CLINICAL TRIALS INVOLVING EMERGENCY PATIENTS

Clinical trials involving emergency patients should be prospective trials, designed to evaluate emergency care treatments, answer specific questions, and more likely to answer the question correctly. In emergency medicine research it is mandatory that the prospective IRB reviews and approves the protocol designed for the study before it starts.

There are numerous potential barriers to conduct randomized controlled trials with emergency patients because of the specific characteristics mentioned below [12, 13]. Low incidence rates of emergency patient events require pooling of centers to conduct research, that is, in many occasions the trials have to be multicentric depending on the pathology investigated. A large number of emergency patients are required to attain diverse and representative study samples. In order for the precise development of a clinical trial, an infrastructure is needed to test the efficacy of treatments, as well as the transport and care that precede the arrival of emergency patients to the hospital emergency departments. A specific mechanism is needed to study the process of transferring research results to treatment settings.

A related challenge has been the difficulty for even large multicenter networks to gather sufficiently large sample sizes necessary to study treatments in the emergency medicine. Therefore, clinical trials in the emergency department need an infrastructure capable of overcoming inherent barriers to emergency medicine research. This infrastructure provides the capacity to conduct multicenter research studies to support research collaboration among medical investigators [14]. It should provide the test ability to collect, transfer, manage data from all sites, and determine availability, completeness, and agreement of core data from electronic and chart review. Despite these well-designed infrastructure multiinstitutional prospective trials, emergency medicine is complex due to:

- Training for investigators and staff is more difficult than those from single-center studies. This is especially critical for emergency departments where the physicians do not usually work for long periods.

- In the setting of emergency situations, randomization is not a simple process. Since researchers do not know when, and may not know where, the emergency is going to occur; it may be technically difficult to effect a randomization [15].
- Protocol development much more explicit because of the high rate of mobility of the physicians in this department.
- Quality control for maintaining high standards of quality and data management.
- Uniform standards for clinical research. Strong emphasis on training and standardization and include "good clinical practice" training.
- Site monitoring requirements of all centers regardless of whether observational or interventional study.
- Data transmission and security. It is necessary to involve epidemiologists and statisticians at the very beginning of the study. It is also essential to rely on independent data center/data manager. Paper data forms should be simple electronic, Web-based data entry of de-identified data, with logic and range checks. Original data forms/source documents at local site. Double data entry should be considered as a quality check—controversial queries for missing or inconsistent data.
- Multicenter studies are becoming more common, and the variability in local IRB assessments can be problematic if IRBs have different standards.
- Costs are high, and budgets tend to underfund training, site monitoring.
- Keep study simple, focused, and well budgeted.
- Accuracy measurement of predictor and outcome variable.

### 2.1 Subject Recruitment

There is a paucity of clinical trials in emergency patients. This is partly due to difficulties in recruiting participants. Insufficient enrollment of patients is the most common reason for discontinuing emergency medicine studies. Failure to recruit members of specific populations to trials that require adequate representation reduces the external validity of the study results [16]. Frequently, the origin of these problems lies in the ethical issues surrounding clinical trials in this population [17]. In fact, recruitment is one of the most significant challenges to completing a research study on time and on budget. In most trials in emergency research, recruitment takes longer, is more expensive, and produces fewer subjects than planned. Subject recruitment can be difficult in the emergency department environment mainly because when patients present with acute symptoms, the emergency physicians are focused on life-saving interventions, not on study enrollment [18]. Though causes of recruitment difficulty have been documented in medical trials, little is known concerning recruitment in emergency medicine studies. Many reasons for this poor recruitment rate for emergency studies include:

- The subject recruitment should be made following established guidelines for protection of participant's rights. The IRB now has an obligation to consider a strict inclusion and exclusion criteria document about subject recruitment.

- In life-threanning situations, relatives present many inconveniences for their relative to participate in a clinical study.
- Limited size of the emergency population, specially patients in critical situations.
- Plans for sampling and recruitment.
- The increased bureaucracy, reporting, and costs of clinical trials after the implementation of the European Clinical Trials Directive and U.S. Food Drug Administration (FDA) guidance for emergency research has led to a sharp decline in the initiation of clinical trials and in participant recruitment throughout Europe and North America [19].
- Regulations for clinical trials: Researchers have always been concerned about the effects that the regulations might have on clinical trials in emergency situations in patients with impaired consciousness. Examples include research on cardiac arrest, stroke, head trauma, spinal cord injury, gunshot wounds, major trauma, cardiac arrest, and poisoning. These regulations impose the need for prior consent from a personal or professional legal representative before a patient could be recruited into a clinical trial [20]. In most cases the regulations are only applicable to trials of drugs. Nonmedicinal trials and trials of clinical care continue to be governed by common law [21]. Federal regulations allow for waiver of consent in situations where there is more than minimal risk to participants, provided there is a prospect of direct benefit to participants and a number of other conditions are met. The conditions include the following:
  - The study could not be carried out without the waiver.
  - Consultation with community representatives occurs before the study starts.
  - Public disclosure is made before and after the study.
  - A therapeutic window is defined, and the researcher commits to try to locate a surrogate/legally authorized representative who can give consent within that window before proceeding to waive consent [22].

Recently, an amendment of the United Kingdom's Medicines for Human Use (Clinical Trials) Regulations allows unconscious patients in emergency situations to be enrolled in clinical trials without prior consent provided it has been approved by the appropriate ethics committee [23].

These difficulties could involve inadequate recruitment, which could reduce the ability to detect treatment differences.

In conclusion, strategies for the recruitment of the patients susceptible to be included in an emergency clinical trial are fundamental. Successful recruitment depends on multiple factors such as population, potential subjects, pathology, available alternatives, transportation availability, population demographics, and general or specific interest in medical research or science within the community.

### 2.2 Emergency Clinical Research in Pediatric Population

The well-known difficulties of conducting studies in young patients, along with the limited economic returns to pharmaceutical companies for pediatric drugs, have led

to a scarcity of pediatric studies and therefore a scarcity of knowledge about drug safety and efficacy in children [24]. The lack of scientifically evaluated medicines for children has been recognized as an area that requires correction [25]. Moreover, clinical trials in children have special ethical constraints [26].

Experimentation on children is considered by many to be unethical, resulting in difficulties to obtain critical safety data. Clinical trials are subject to detailed scrutiny by the various regulatory bodies. They have recently recognized the need for pharmaceutical companies to invest in pediatric medicines. The majority of marketed drugs are either not labeled, or inadequately labeled, for use in pediatric patients [27]. This is because, as we have stated before, there is a paucity of clinical trial works in children, which leads to the frequent use of off-label and unlicensed medications in this very vulnerable group [28].

The pediatric population is constituted by a wide range of individuals of substantially varied physical size, weight, and stage of physiological development. The differences between adults and children mandate strategies to improve this situation rather than continually relying on extrapolation from adult studies. In the protocols for clinical trials for pediatric emergency room it is mandatory that:

- Each trial should answer one primary study question.
- Sample size estimation based on error rates and clinically important difference must be calculated.
- Allocation concealment, randomization, and blinding are critical to prevent selection and ascertainment bias [29]. Clinical trial design must be practical, and a trial is more likely to succeed if a simple design is utilized, with minimal interference with school and work.
- Standardization of the intervention is important for the results validity, and study centers should be selected on the basis of the trial objective. The effects of dropouts, cross-overs, and missing data should be handled with an intention-to-treat analysis.
- Sample selection, informed consent, and valid outcome measures are specific challenges in the pediatric population [29].

Efforts should be made to increase the number of well-designed, randomized controlled trials in pediatric emergency patients. Therefore, new strategies have to be established between the pharmaceutical industry, the European Union (EU), and the U.S. regulatory bodies and pediatric centers to facilitate effective trials work. In this regard, The EU decided to support the development of a European register of clinical trials in children as part of the Fifth Framework Thematic Programme "Quality of Life" in 2002. DEC-NET (the European register of clinical trials on medicines for children—drug evaluation in children), supported by the EU, was created to provide the scientific community with a flexible tool for promoting communication and collaboration among researchers, disseminating clinical trial results, and facilitating patient access and recruitment to trials [30]. The project DEC-NET currently involves members of four countries: France, Italy, Spain, and the United Kingdom. DEC-NET fits International Committee of Medical Journal Editors criteria, is free of charge, and is designed to be used by the general public and health professionals [31]. It is unique in that it is the first population-oriented

clinical trial register. Such a register represents a useful source for planning new studies, promoting communications and collaborations between researchers, facilitating patient access and recruitment into trials, preventing trial duplication and inappropriate funding, and identifying the therapeutic needs of children that remain neglected. It will also allow active monitoring of new or evolved knowledge of drug therapies [32].

## 3 ETHICAL ISSUES IN EMERGENCY CLINICAL TRIALS

The ethical aspects of an emergency clinical trial cannot be separated from the scientific objectives. Segregation of ethical issues from the full range of study design components demonstrates a flaw in understanding the fundamental nature of research involving human subjects. Recognition that good science and good ethics are inextricably bound together in clinical research is today's reality. Thus, the ethical conduct of a clinical trial does not end with the formulation of study design or the obtainment of a signature on the informed consent form [33].

To improve the study design and conciliate science and ethics, some authors recommend that every protocol include a specific section in which the ethical considerations are exposed: the design and conduct of the study, selection and recruitment of participants, and the conduct of the consent process be presented, discussed, and justified, not in terms of ethics-based regulatory requirements but in true fulfillment of the investigator's responsibilities (as a scientist and moral duty) to ensure that the proposed study is sound, ethically and scientifically [34].

Patients attending an emergency department are entitled to assume that they will be treated in accordance with the best available evidence. The emergency physician has a responsibility to expand the fund of knowledge that underpins best evidence by means of clinical studies of new technologies and interventions. However, there are inherent difficulties in undertaking clinical research in an emergency department setting. Patients are usually frightened, often in pain, and sometimes have impaired cognition. Practitioners use case-based reasoning to apply bioethics to these clinical situations, usually giving most weight to patients' autonomy and values but also incorporating other relevant bioethical principles, including those encompassed in professional oaths and codes. Emergency clinicians must be able to recognize bioethical dilemmas, have action plans based on their readings and discussions, and have a method through which to apply ethical principles in clinical settings [35].

Other aspects that can help for the ethical consideration of the emergency clinical trials could be the randomization to treatment, which must occasionally be completed in the context of a clinical emergency. Furthermore, patients take on trust the fact that the clinician will act in a professional manner and not misuse the authority of his or her position. The researchers must be committed to the principle that self-promotion must not take precedence over the best interests of the patient.

The moral obligation of emergency researchers to do well, beneficence, must be tempered by nonmaleficence. That is, the expected bad effects of an intervention must be weighed against the anticipated benefits. The type, amount, and probability of harm should be effectively communicated to subjects, and the physician should

take account of the fact that what constitutes a harmful effect, or a beneficial effect, may be particular to the individual subject. Justice and equality of access are inalienable rights that must be respected in participants and nonparticipants alike [35]. The researcher must be a good communicator, demonstrating openness and effective listening. Benefits may flow from a research study in favor of the patient or to society. When discussing the benefits of a proposed study, one must distinguish between therapeutic and nontherapeutic research. Researchers are obliged to consider the potential benefit to the patient when balancing potential risks associated with any protocol. Researchers are entitled to expect that their work is regarded as honest until shown to be otherwise.

Although it is uncommon, misconduct occurs in most countries. Several causes of misconduct have been described, including pressure to publish, financial inducements, and simple vanity. Misconduct in researchers must not be tolerated and must be reported promptly. There should be a named person within the organization who is responsible for the receipt of complaints of the misconducted research [36].

In order to ensure the good practices for emergency clinical trials, an IRB was established. The IRB was created to maintain ethical standards of practice in research, to protect subjects of research from harm, to preserve the subjects' rights and to provide reassurance to the public that this is being done. It is not intended to impede good research but facilitate it, approving studies of good quality, not badly planned or poorly designed protocols. The IRB has numerous protection responsibilities that include initial and continuing review of the study protocol and related documents, review of the documentation of informed consent, and review of reports of unanticipated problems and of adverse events [37].

In emergency clinical research the IRB will make an assessment of the risks and benefits represented by the proposal, taking into account the likelihood and severity of any risk to subjects and patients and weighing this against the potential for benefit. IRBs should ensure that a monitoring plan exists for each individual study site and that there is a data-monitoring program for multisite studies. Second, the IRB should certify investigators' understanding of compliance with regulations governing the subjects' safety during the trial, including adverse reaction reporting and the collection of relevant study documents. Finally, the IRB should review the safety monitoring board reports and query investigators as needed to determine whether additional safeguards are necessary [38]. These recommendations seem reasonable because the structure and function of IRBs would enable them to perform these roles better than the other entities involved in the monitoring of clinical trials. In addition, the suggested roles would add an efficient framework to the role that IRBs can play in the safety monitoring of human subjects [39].

### 3.1 Informed Consent

"Informed consent" means the knowledgeable consent of an individual, or his or her legally authorized representative, who is able to exercise free power of choice without undue inducement or any form of force, fraud, deceit, duress, or other form of constraint or coercion. This document will be used with subjects or their legally

authorized representatives in feasible situations. The information in the consent document will be also used when providing an opportunity for a family member to object to a subject's participation in the clinical investigation.

One of the most important ethical dilemmas in emergency department research is the difficulty of obtaining genuine informed consent from research subjects [40]. Patients with life and death emergencies such as those with head injuries and impaired consciousness are unable to give informed consent to participate in clinical trials. Nevertheless, controlled clinical trials are essential in such situations to identify effective ways to prevent death and long-term disability.

On the other hand, emergency department research on new treatments for cardiopulmonary arrest, neurologic emergencies, and major trauma is challenging, controversial, and necessary [41]. The critical nature of the illness requires immediate treatment; emergency medicine research often takes place when patients are in distress or even unconscious, making a thoughtful discussion of risks and benefits impossible. This would particularly affect vulnerable people exposed to acute, life-threatening conditions where available treatments are unproved or unsatisfactory, and the collection of valid scientific evidence, which may include evidence obtained through randomized, placebo-controlled investigations, is necessary to determine the safety and efficacy of particular interventions [42].

Randomized, controlled trials should be designated as minimal risk, allowing IRBs to approve their conduct with a waiver of informed consent if obtaining it is not feasible. These clinical situations are: the patients will not be able to give informed consent as a result of their medical condition; the intervention involved in the research must be administered before consent from the subjects' legally authorized representatives is feasible; or there is no reasonable way to identify prospectively the individuals likely to become eligible for participation in the research [43].

In these clinical trials the potential sources of risk that must be considered are physical risk from study treatments, the loss of individualized care, risk from non-therapeutic components of the research protocol, and the psychological impact of participation. The risks of research participation should be considered in comparison with the risk of nonparticipation; for example, the risks specific to research participation should be considered separately from the risks inherent to the treatment of the potential research participant's underling condition. Therefore the risk–potential benefit balance for every arm of the study is at least as favorable as the risk–potential benefit balance for any therapy available outside of the research protocol. Moreover, participation in the study does not present added medical risk above the risk inherent to the research participant's underlying condition. The incremental risk of such a clinical trial would not exceed the threshold of minimal risk [44].

Participation in an randomized, controlled trial carried out without the waiver from informed consent requirements may pose no more than minimal risk when:

- Genuine clinical equipoise exists [45].
- All of the treatment options included in the research study fall within the current standard of care.
- There is no currently available treatment with a more favorable risk–benefit profile than the treatments included in the study.

- The nontherapeutic components of the research are safely under the minimal risk threshold.
- Research protocol provides sufficient latitude for treating physicians to individualize care when appropriate.

Moreover, it is very important to consider the potential negative psychological impact for research participants or their families in risk assessment [46]. However, a patient involved in a research protocol that has been reviewed for scientific and ethical quality is more protected than the one who receives an unproved treatment at the discretion of a provider who believes that it may help. Abboud et al. [47] reported that some patients in the emergency department would prefer to receive an unproved, potentially helpful drug if they were in cardiac arrest, neurologic emergencies, or major trauma, but a much smaller proportion would participate in a randomized trial comparing a new drug with a placebo. Patients were also less willing to participate in studies that evaluated highly invasive interventions [47].

In 1996, the FDA regulations were revised to include the report 21 CFR 50.24, which allows for an exception from informed consent requirements for a very limited class of research in emergency settings [19].

In this regard, European Union 2001/20/EC directive is a cornerstone of a Europe-wide harmonization of the provisions governing clinical trials and can be expected to foster and facilitate multinational clinical research. Article 5 outlines the conditions for research in incapacitated patients unable to give informed consent. The article, however, is framed to address the needs of individuals who are incapacitated for long periods, many even permanently [48].

A clinical trial can only be done if informed consent of the legal representative has been obtained. The directive may not only affect unconscious people. Patients with acute myocardial infarction have been enrolled in clinical trials so far. Many of these have severe pain on admission and receive treatment with opiates: Can they give informed consent, particularly those with cardiogenic shock. Research in the acute care setting is already difficult and this directive will make it even more difficult [49].

Both, the U.S. federal regulations of 1996 [19] and the 2001 European Union directive combined have resulted in an extremely limited number of emergency research studies that have been completed or are ongoing under an exception from informed consent [50].

On the other hand, poor readability of informed consent forms has been a persistent problem in clinical research. The low educational attainments of many patients with mental illness might suggest a still greater problem in emergency room settings. This problem is even worse in emergency department setting due to the stress of the patient, the severity of illness, or even the conscious level of the patient. This fact not only affects the patient but also his or her family. In most cases, a considerable level of education, consciousness, and calmness is necessary to understand the informed consent documents used to enroll subjects in emergency medicine research. Mader and Playe [51] reported that the length and complexity of the consent forms increased as risk to the subject increased. In conclusion, informed consent documents used in emergency medicine research may be too complex for the average patient to understand [51].

Hence, it is desirable to design a consent form to meet all of the regulatory requirements while maintaining a level of reading comprehension suitable for the general population. This is a difficult task for investigators, methods of reducing the complexity of forms, as part of improving the overall consent process, are much needed. IRBs are charged with safeguarding potential research subjects with limited literacy but may have an inadvertent role in promulgating unreadable consent forms [52].

## 4 TYPES OF CLINICAL TRIALS IN EMERGENCY DEPARTMENTS

In this section, we describe the most common randomized clinical trials carried out in emergency medicine. Diseases with an acute process and high morbidity and mortality are prime candidates to develop research in the emergency department. Observational studies provide useful but potentially biased information about the effect of interventions. Because of their ability to deal with bias, randomized clinical trials are the optimum vehicle to obtain a true estimate of the effect of interventions, including emergency research.

We have selected those randomized clinical trials that include patients suffering from cardiovascular alterations or central nervous system affections conducted in the emergency department.

### 4.1 Cardiovascular Alterations

***Atrial Fibrillation*** Atrial fibrillation is the most prevalent arrhythmia in hospital emergency departments and is a serious disease associated with a high mortality rate. However, the management of atrial fibrillation in this scenario is variable and frequently inadequate. Some randomized controlled trial involved patients diagnosed as having atrial fibrillation in the emergency department, and their main objective was to compare drugs strategies.

Dermican et al. [53] compared the effectiveness of intravenous diltiazem and metoprolol in the management of rapid ventricular rate in atrial fibrillation. Forty patients met the inclusion criteria. All the enrolled patients received information about the study and signed written informed consent. The authors concluded that diltiazem and metoprolol were safe and effective for the management of rapid ventricular rate in atrial fibrillation. However, they concluded that diltiazem allowed patients to recover the rate of control effect earlier and that the percentage of ventricular rate decrease was higher than with metoprolol [53].

Another example was published by Davey and Teubner [54] who conducted a prospective, randomized, double-blind, placebo-controlled trial in an adult emergency department, enrolling patients suffering from rapid atrial fibrillation, to examine the safety and efficacy of magnesium sulfate infusion, in addition to usual care, for acute rate reduction in patients with atrial fibrillation, and a rapid ventricular response rate. In all 199 patients were randomized, 102 received magnesium sulfate and 97 received placebo. The trial results showed that the addition of magnesium sulfate to standard rate reduction therapies enhances rate reduction and conversion to sinus rhythm in patients with rapid atrial fibrillation [54].

Thomas et al. [55] carried out a study to asses the efficacy and safety of rapid high-dose intravenous infusions of amiodarone and sotalol for heart rate control and rapid reversion to sinus rhythm in patients who came to the emergency department with recent-onset symptomatic atrial fibrillation. One hundred and forty patients gave informed consent and were randomized. They received 1.5 mg/kg of sotalol infused in 10 minutes, 10 mg/kg of amiodarone in 30 minutes, or 500 μg of digoxin in 20 minutes. Electrical cardioversion was attempted for patients who did not convert to sinus rhythm within 12 hours. This clinical trial provided evidence that the rapid infusion of sotalol or amiodarone in patients with symptomatic recent-onset atrial fibrillation results in rapid control of ventricular rate. Even with high-dose rapid infusions, all 3 agents are associated with a poor overall reversion rate within 12 hours. Almost all patients were returned to sinus rhythm with the combination of pharmacological therapy and electrical cardioversion [55].

Kim et al. [56] performed a prospective, randomized pilot study of a traditional approach of hospital admission versus an accelerated emergency-department-based strategy with low-molecular-weight heparin and early cardioversion to sinus rhythm, in a cohort of patients with uncomplicated atrial fibrillation of a single university hospital. The primary endpoints were length of stay and total actual direct costs. Eighteen patients were randomized over a 15-month period. The new strategy resulted in a shorter length of stay at potentially lower cost [56].

This kind of assays has to be conducted in a very short period of time, therefore, it is extremely important to design them very accurately and control very sharply their follow-up in order to satisfy the proposed objectives. Especially, the protocol assay and patient recruitment and informed consent have to be strictly defined before starting the clinical trial.

***Myocardial Infarction*** Advances in the treatment of myocardial infarction such as coronary care units, patient education, improved emergency response systems, and reperfusion therapy has reduced mortality. Thus, trials using mortality alone as an endpoint may no longer be feasible even in survivors of a heart attack, unless a very high risk group is studied. This has led to increased use of combination endpoints, such as cardiovascular mortality plus nonfatal myocardial infarction. Development of randomized clinical trials for management of ST-elevation myocardial infarction developed by a community-based emergency medical service, emergency department, or cardiovascular service is recommended.

Moreover, due to the involvement of numerous medical services in clinical trials, there are many types of clinical trials concerning different therapeutic approaches for the treatment of acute myocardial infarction [57]. To identify the characteristics of these assays, we have selected a study carried out to explore the effect of treating with glycoprotein IIb/IIIa receptor antagonist patients suffering from thrombolysis ST-segment elevation myocardial infarction, as an example of acute myocardial infarction clinical trials in the emergency environment design.

*Thrombolysis ST-Segment Elevation Myocardial Infarction* The importance of the dissolution and prevention of thrombosis in treating patients with ST-segment elevation myocardial infarction has motivated the development of novel therapies targeting platelet aggregation and thrombus formation [58].

Several randomized clinical trials have been developed in the emergency setting using glycoprotein IIb/IIIa receptor antagonists.

Abciximab is an antibody fragment that dose-dependently inhibits platelet aggregation and leukocyte adhesion by binding to the glycoprotein (GP) IIb/IIIa, vitronectin, and Mac-1 receptors [59]. The beneficial effects of using abciximab with stent placement in patients with acute MI compared to stent alone were demonstrated in the double-blind ADMIRAL study [60]. Three hundred patients with acute myocardial infarction were enrolled in a double-blind fashion either to abciximab plus stenting (149 patients) or placebo plus stenting (151 patients) before they underwent coronary angiography. The ethics review board of the Pitié–Salpêtrière Hospital approved the protocol, and the study was conducted in accordance with the Declaration of Helsinki. Signed informed consent was obtained from the patients. Clinical outcomes were evaluated 30 days and 6 months after the procedure. The angiographic patency of the infarct-related vessel and the left ventricular ejection fraction were evaluated at 24 hours and 6 months. The results of the trial reported that compared with placebo, early administration of abciximab in patients with acute myocardial infarction improves coronary patency before stenting, as well as the success rate of the stenting procedure, the rate of coronary patency at 6 months, left ventricular function, and clinical outcomes [60].

The large CADILLAC study randomly assigned 2082 patients with acute myocardial infarction to undergo PTCA alone (518 patients), PTCA plus abciximab therapy (528), stenting alone with the MultiLink stent (512), or stenting plus abciximab therapy (524), using a 2-by-2 factorial design [61]. The study was approved by the institutional review board or ethics committee at each participating center, and consecutive, eligible patients provided signed informed consent. In this study, both primary objectives were met; stenting alone was superior to percutaneous transluminal coronary angioplasty, and stenting alone was not inferior to percutaneous transluminal coronary angioplasty plus abciximab [61].

In this regard, a recent meta-analysis of eight randomized trials conducted on abciximab as adjunctive therapy to mechanical revascularization for ST-elevation myocardial infarction showed that abciximab, as adjunctive therapy to both primary angioplasty and fibrinolytictherapy for STEMI, is associated with a significant reduction in short-term reinfarction rate, whereas the benefits in reducing mortality are observed only in association with primary angioplasty [62]. Abciximab is not associated with an increased risk of intracranial hemorrhage (except in combination with fibrinolysis in elderly patients), whereas a higher risk of major bleeding complications is observed only in association with fibrinolysis [62].

Eptifibatide is an intravenously administered glycoprotein IIb/IIIa receptor antagonist that acts at the final step of the platelet aggregation pathway. Two clinical trials showed that concomitant administration of eptifibatide in patients undergoing elective PCI reduced thrombotic complications [63]. One of them was the IMPACT-II (Integrilin to Minimize Platelet Aggregation and Prevent Coronary Thrombosis II), a multicenter, double-blind, placebo-controlled trial, enrolling 4010 patients undergoing elective, urgent, or emergency coronary intervention. Patients were assigned one of three treatments: placebo ($n = 1328$), a bolus of 135 μg/kg eptifibatide followed by an infusion of 0.5 μg/kg/min for 20–24 hours ($n = 1349$), or 135 μg/kg eptifibatide bolus with a 0.75-μg/kg/min infusion ($n = 1333$). All patients provided informed consent. Their results showed that the treatment strategy with eptifibatide

135/0.5 reduced rates of early abrupt closure and ischemic events at 30 days [63]. The other study was the ESPRIT (Enhanced Suppression of the Platelet IIb/IIIa Receptor with Integrilin Therapy), which is a randomized, placebo-controlled trial to assess whether a novel, double-bolus dose of eptifibatide could improve outcomes of patients undergoing coronary stenting. A total of 2064 patients undergoing stent implantation in a native coronary artery were enrolled. Immediately before percutaneous coronary intervention, patients were randomly allocated to receive eptifibatide, given as two 180 μg/kg boluses 10 minutes apart and a continuous infusion of 2.0 μg/kg/min for 18–24 hours, or placebo, in addition to aspirin, heparin, and a thienopyridine. The protocol was approved by each of the respective institutional review boards, and informed consent was obtained from all patients. The authors concluded that routine glycoprotein IIb/IIIa inhibitor pretreatment with eptifibatide substantially reduces ischemic complications in coronary stent intervention and is better than a strategy of reserving treatment to the bailout situation [64].

Moreover, the PURSUIT (Platelet Glycoprotein IIb/IIIa in Unstable Angina: Receptor Suppression Using Integrilin Therapy) trial, which included 10,948 patients with non-ST-elevation acute coronary syndromes, tested the hypothesis that inhibition of platelet aggregation with eptifibatide would have an incremental benefit, beyond that of heparin and aspirin, in reducing the frequency of adverse outcomes. Patients with ischemic chest pain within the previous 24 hours and who had either electrocardiographic changes indicative of ischemia (but not persistent ST-segment elevation) or high serum concentrations of creatine kinase MB isoenzymes were enrolled in the study. They were randomly assigned, in a double-blind manner, to receive a bolus and infusion of either eptifibatide or placebo, in addition to standard therapy, for up to 72 hours (or up to 96 hours if coronary intervention was performed near the end of the 72-hour period). Because this was the first large-scale study using higher doses of eptifibatide than those previously used, it was specified in the protocol that the study would be stopped for the lower-dose group after the independent data safety and monitoring committee had conducted an interim review of safety data, provided the higher dose had an acceptable safety profile. After 3218 patients had been randomly assigned to treatment groups, the committee recommended dropping the lower dose. Eptifibatide significantly reduced the primary endpoint of death and nonfatal myocardial infarction at 30 days compared with placebo [65]. Studies are now evaluating eptifibatide in high-risk patients with non-ST-elevation acute coronary syndromes (NSTE-ACS) and a planned early invasive strategy in the EARLY-ACS (Eptifibatide Administration prior to Diagnostic Catherization and Revascularization to Limit Myocardial Necrosis in Acute Coronary Syndrome) trial and in patients with primary PCI for STEMI in comparison to abciximab in the Eptifibatide versus Abciximab in Primary PCI for Acute Myocardial Infarction trial. After the completion of these trials, the value of eptifibatide in patients undergoing PCI in different indications could be determined based on the evidence thanks to the trials carried out [66].

Tirofiban is a small, synthetic nonpeptide, competitive GP IIb/IIIa antagonist with high specificity and high affinity for the GP IIb/IIIa receptor.

Multicenter phase III open-label, multinational investigator-driven clinical trial of single high-bolus dose tirofiban versus abciximab and sirolimus-eluting versus bare metal stent in acute myocardial infarction was performed. The objective of the study

was to evaluate, with a 2-by-2 factorial design, the safety/efficacy profile of 4 interventional strategies of reperfusion: tirofiban given at high-bolus dose (bolus of 25 μg/kg over 3 minutes), followed by an infusion of 0.15 μg/kg/min for 18–24 hours versus abciximab and sirolimus-eluting stent, as compared to bare-metal stent implantation in primary percutaneous coronary intervention. In a head-to-head comparison, tirofiban 10-μg/kg bolus followed by a 0.15-μg/kg/min infusion was found to be inferior to the standard dose of abciximab in patients undergoing percutaneous coronary intervention. Insufficient platelet inhibition with low-dose tirofiban may likely explain these results. Subsequently, a high-bolus dose of tirofiban followed by standard infusion was tested and evidence suggests that in this dosing tirofiban may be as effective as abciximab and have a comparable safety profile [67].

In conclusion, glycoprotein IIb/IIIa receptor antagonists inhibit the binding of ligands to activated platelet GP IIb/IIIa receptors and, therefore, prevent the formation of platelet thrombi. Glycoprotein IIb/IIIa inhibition to PCI for treatment of STEMI has substantially lowered the incidence of recurrent ischemic events and has improved early survival.

### 4.2 Central Nervous System

Clinical research involving patients with neurologic emergencies has increased in the emergency department. We have highlighted randomized clinical trials studying therapeutic treatments for stroke.

***Stroke*** The National Institute of Neurological Disorders and Stroke (NINDS) rt-PA Stroke Study was a randomized, double-blind trial of intravenous recombinant tissue plasminogen activator (t-PA) for ischemic stroke, when treatment was begun within 3 hours of the onset of stroke. The trial was divided in two parts: Part 1 (in which 291 patients were enrolled) tested whether t-PA had clinical activity, as indicated by an improvement of 4 points over baseline values in the score of the National Institutes of Health stroke scale (NIHSS) or the resolution of the neurological deficit within 24 hours of the onset of stroke. Part 2 (in which 333 patients were enrolled) used a global statistic test to assess clinical outcome at 3 months, according to scores on the Barthel index, modified Rankin scale, Glasgow outcome scale, and NIHSS. Informed consent was obtained for all patients. This study showed that treatment with intravenous t-PA within 3 hours of the onset of ischemic stroke improved clinical outcome at 3 months, despite an increased incidence of symptomatic intracerebral hemorrhage [68].

Randomized double-blind placebo-controlled trial of thrombolytic therapy with intravenous alteplase in acute ischemic stroke (ECASS II) has assessed the safety and efficacy of intravenous thrombolysis with alteplase (0.9 mg/kg body weight) within 6 hours of stroke onset. The trial protocol was reviewed and approved by local independent ethics committees or institutional review boards according to the regulatory requirements of the participating country, and carried out in accordance with the ethical principles of the Declaration of Helsinki. Informed consent was obtained from each patient (or from his or her legally authorized representative) before enrollment in the study. The results do not confirmed a statistical benefit for alteplase. The trend toward efficacy should be interpreted in the light of evidence from previous trials. Despite the increased risk of intracranial hemorrhage,

thrombolysis with alteplase at a dose of 0.9 mg/kg in selected patients may lead to a clinically relevant improvement in outcome [69].

The ATLANTIS study was a randomized controlled trial whose objective was to test the efficacy and safety of rt-PA in patients with acute ischemic stroke when administered between 3 and 5 hours after symptom onset. Its current approved use is limited to within 3 hours of symptom onset, which restricts a lot the number of patients who can be treated, since most stroke patients present more than 3 hours after symptom onset. This study found no significant rt-PA benefit on the 90-day efficacy endpoints in patients treated between 3 and 5 hours. All patients or their legal representatives signed an informed consent, which was previously approved by the correspondent institutional review board [70]. The risk of symptomatic ICH increased with rt-PA treatment. These results do not support the use of intravenous rt-PA for stroke treatment beyond 3 hours [70].

Hacke et al. [71] performed a meta-analysis of the recent large stroke trials, including NINDS [68], ECASS [69], and ATLANTIS [70] in order to draw conclusions about the best use of tissue plasminogen activator (t-PA) for ischemic stroke [71]. This study demonstrates that onset to treatment time is critically associated with improved functional outcome [71].

As opposed to the post hoc analysis of the NINDS trial, where an effect of time was not found in the prespecified analysis of the 90- and 180-minute groups, this larger meta-analysis sample nicely demonstrates a greater likelihood of good outcome, OR 2.8, for treatment in the 0- to 90-minute time frame versus OR of 1.6 and 1.4 for the 90- to 180- and 180- to 270-minute time frame. Interestingly, time from symptom onset did not influence the likelihood for hemorrhage conversion. The meta-analysis solidifies the rationale for the investment of significant public health effort in the development of emergency medicine service and hospital-based teams to treat stroke rapidly [72].

A study to investigate the feasibility and safety of a combined intravenous and intra-arterial approach to recanalization for ischemic stroke has been recently performed: 81 subjects with a baseline NIHSS > or = 10 were enrolled. The median time to initiation of intravenous rt-PA was 142 minutes as compared with 108 minutes for placebo and 90 minutes for rt-PA-treated subjects in the NINDS rt-PA Stroke Trial ($p < 0.0001$) [73]. The 3-month mortality in IMS II subjects was 16% as compared with the mortality of placebo (24%) and rt-PA-treated subjects (21%) in the NINDS rt-PA stroke trial. The rate of symptomatic intracerebral hemorrhage in IMS II subjects (9.9%) was not significantly different than that for rt-PA-treated subjects in the NINDS t-PA stroke trial (6.6%). IMS II subjects had significantly better outcomes at 3 months than NINDS placebo-treated subjects for all endpoints (OR > or = 2.7) and better outcomes than NINDS rt-PA-treated subjects as measured by the Barthel index and global test statistic [73].

### 4.3 Sepsis

Several studies have demonstrated the importance of early and adequate antimicrobial therapy in reducing the mortality and morbidity of patients with severe sepsis. About 6 to approximately 17% of empirical antibiotic selection were judged to be inappropriate according to subsequent microbiology and higher antimicrobial susceptibility as a result. It reflects the diversity in the presentations of infectious

diseases and the limited microbiological reports available from the first-line emergency physicians. Timely diagnosis and selection of appropriate antibiotics/treatment in the treatment of those patients is a challenge for an emergency physician more than ever before.

Rivers et al. [74] performed a prospective, randomized study to evaluate if the efficiency of early goal-directed therapy before admission to the intensive care unit effectively reduces the incidence of multiorgan dysfunction, mortality, and the use of health care resources among patients with severe sepsis or septic shock. This study was approved by the institutional review board for human research and was conducted under the auspices of an independent safety, efficacy, and data-monitoring committee: 263 patients were enrolled, 130 were randomly assigned to early goal-directed therapy and 133 to standard therapy. The result of the trial showed that hospital mortality was 30.5% in the group assigned to early goal-directed therapy, as compared with 46.5% mortality in the group assigned to standard therapy. Furthermore, mean APACHE II scores were significantly lower, indicating less severe organ dysfunction, in the patients assigned to early goal-directed therapy than in those assigned to standard therapy. The authors concluded that early goal-directed therapy provides significant benefits with respect to outcome in patients with severe sepsis and septic shock [74].

Another study tested the hypothesis that in the setting of undifferentiated symptomatic hypotension, the presence of hyperdynamic left ventricular function (LVF) on focused ED echocardiography would be a specific finding for sepsis as the etiology of shock [75]. This clinical trial was preplanned as a secondary analysis of 184 patients enrolled in a randomized clinical trial to investigate the role of an ultrasound protocol in evaluating the etiology of undifferentiated hypotension in the emergency department. Written informed consent was obtained from all patients. A final diagnosis of septic shock was made in 38% (39/103) of patients. Of 103 patients 17 had hyperdynamic LVF with an interobserver agreement of $\kappa = 0.8$. Hyperdynamic LVF had a positive likelihood ratio of 5.3 for the diagnosis of sepsis and was a strong independent predictor of sepsis as the final diagnosis with an odds ratio of 5.5 [95% confidence index (CI) 1.1–45] [75]. They conclude that among emergency department patients with nontraumatic undifferentiated symptomatic hypotension, the presence of hyperdynamic LVF on focused echo is highly specific for sepsis as the etiology of shock [75].

Drotrecogin alfa (activated), or recombinant human activated protein C, produced dose-dependent reductions in the levels of markers of coagulation and inflammation in patients with severe sepsis [76]. PROWESS, a randomized, double-blind, placebo-controlled, multicenter trial, was conducted to evaluate whether treatment with drotrecogin alfa activated reduced the rate of death from any cause among patients with severe sepsis. The institutional review board at each center approved the protocol, and written informed consent was obtained from all participants or their authorized representatives. A total of 1690 randomized patients were treated (840 in the placebo group and 850 in the drotrecogin alfa activated group). The mortality rate was 30.8% in the placebo group and 24.7% in the drotrecogin alfa activated group. However, the incidence of serious bleeding was higher in the drotrecogin alfa activated group than in the placebo group [77]. More studies are required to evaluate the possible beneficial effect of this drug. In this regard, Vincent et al. [78] performed the ENHANCE trial, a multiple-country, single-arm, open-label, trial in

order to provide further evidence for the efficacy and safety of drotrecogin alfa (activated) treatment in severe sepsis. Patients with known or suspected infection, three or four systemic inflammatory response syndrome criteria, and one or more sepsis-induced organ dysfunctions were recruited: 2434 adults entered, 2378 received drotrecogin alfa (activated), and of these, 2375 completed the protocol. Appropriate informed consent was obtained from all patients or their legal representative. The 28-day all-cause mortality was approximately the same as in the PROWESS trial (25.3% vs. 24.7%). However, patients in ENHANCE had increased serious bleeding rates compared with patients in the drotrecogin alfa (activated) arm of PROWESS. Increased postinfusion bleeding suggested a higher background bleeding rate. Intracranial hemorrhage was more common in ENHANCE than PROWESS. The authors concluded that ENHANCE provides supportive evidence for the favorable benefit–risk ratio observed in PROWESS and suggests that more effective use of drotrecogin alfa (activated) might be obtained by initiating therapy earlier [78].

## 5 CONCLUSIONS

Lately some clinical trials have been conducted in emergency departments, thus showing that the emergency-department-associated difficulties can be overcome. In spite of this there is a very short time frame between the admission of the patient and his or her inclusion in the clinical trial. With a well-designed protocol and a strict follow-up of it, as well as with very delimited functions of each physician, it is feasible to conduct a clinical trial in the emergency department, of course, without forgetting to obtain previously the signed informed consent.

To sum up, few clinical trials have been developed up to now in the emergency departments due to the associated difficulties. However, when possible, they have been demonstrated to be very useful for the physicians in order to know which is the best strategy when a patient arrives at the emergency department, especially in diseases related to the cardiovascular or central nervous system. For clinical trials in the emergency department to be feasible, it is crucial to have previously the designed protocol (which has to be very precise and, of course, approved by the corresponding committees), having the informed consent (clear and strictly redacted), the physicians must be trained for the protocol to be conducted, and the infrastructures have to be adequate for conducting the specific protocol.

## REFERENCES

1. Anderson, P., Petrino, R., Halpern, P., et al. (2006), The globalization of emergency medicine and its importance for public health, *Bull. World Health Organ*, 84, 835–839.
2. Walker, D. M., Tolentino, V. R., and Teach, S. J. (2007), Trends and challenges in international pediatric emergency medicine, *Curr. Opin. Pediatr.*, 19, 247–252.
3. Arnold, J. L., and Corte, D. F. (2003), International emergency medicine: Recent trends and future challenges, *Eur. J. Emerg. Med.*, 10, 180–188.
4. Vaslef, S. N., Cairns, C. B., and Falletta, J. M. (2006), Ethical and regulatory challenges associated with the exception from informed consent requirements for emergency

research: From experimental design to institutional review board approval, *Arch. Surg.*, 141, 1019–1023.

5. Patterson, S. D., and Jones, B. (2007), A brief review of Phase 1 and clinical pharmacology statistics in clinical drug development, *Pharm. Stat.*, 6, 79–87.
6. Boisjolie, C. R., Sharkey, S. W., Cannon, C. P., et al. (1995), Impact of a thrombolysis research trial on time to treatment for acute myocardial infarction in the emergency department, *Am. J. Cardiol.*, 76, 396–398.
7. Soran, O., Kennard, E. D., Bart, B. A., et al.; IEPR investigators (2007), Impact of external counterpulsation treatment on emergency department visits and hospitalizations in refractory angina patients with left ventricular dysfunction, *Congest. Heart Fail.*, 13, 36–40.
8. Mehta, S. R., Steg, P. G., Granger, C. B., et al.; ASPIRE Investigators (2005), Randomized, blinded trial comparing fondaparinux with unfractionated heparin in patients undergoing contemporary percutaneous coronary intervention: Arixtra Study in Percutaneous Coronary Intervention: a Randomized Evaluation (ASPIRE) Pilot Trial, *Circulation*, 111, 1390–1397.
9. Saver, J. L., Kidwell, C., Eckstein, M., et al.; FAST-MAG Pilot Trial Investigators (2004), Prehospital neuroprotective therapy for acute stroke: Results of the Field Administration of Stroke Therapy-Magnesium (FAST-MAG) pilot trial, *Stroke*, 35, e106–108.
10. Corneli, H. M., Zorc, J. J., Majahan, P., et al.; Bronchiolitis Study Group of the Pediatric Emergency Care Applied Research Network (PECARN) (2007), A multicenter, randomized, controlled trial of dexamethasone for bronchiolitis, *N. Engl. J. Med.*, 357, 331–339.
11. Gibson, C. M., Kirtane, A. J., Murphy, S. A., et al.; TIMI Study Group (2006), Early initiation of eptifibatide in the emergency department before primary percutaneous coronary intervention for ST-segment elevation myocardia infarction: Results of the Time to Integrilin Therapy in Acute Myocardial Infarction (TITAN)-TIMI 34 trial, *Am. Heart J.*, 152, 668–675.
12. Doney, M. K., and Macias, D. J. (2005), Regional highlights in global emergency medicine development. *Emerg. Med. Clin. North Am.*, 23(1), 31–44.
13. Alagappan, K., Schafermeyer, R., Holliman, C. J., et al. (2007), International emergency medicine and the role for academic emergency medicine, *Ann. Emerg. Med.*, 14, 451–456.
14. Barsan, W. G., Pancioli, A. M., and Conwit, R. A. (2004), Executive summary of the National Institute of Neurological Disorders and Stroke Conference on Emergency Neurologic Clinical Trials Network, *Ann. Emerg. Med.*, 44, 407–412.
15. Hallstrom, A. P., and Paradis, N. A. (2005), Pre-randomization and de-randomization in emergency medical research: New names and rigorous criteria for old methods, *Resuscitation*, 65, 65–69.
16. Roberts, I., Shakur, H., Edwards, P., et al. (2005), Trauma care research and the war on uncertainty, *BMJ*, 331, 1094–1096.
17. Lemaire, F. (2006), The inability to consent in critical care research: Emergency or impairment of cognitive function, *Intensive Care Med.*, 32, 1930–1932.
18. Anon. (2006), ER: The gateway for neurologic emergency trials? *Ann. Neurol.*, 60, A12–14.
19. Food and Drug Administration (1996), Protection of human subjects; informed consent; final rule, *Fed. Reg.*, 61, 51498–51531.
20. Shakur, H., Roberts, I., Barnetson, L., et al. (2007), Clinical trials in emergency situations, *BMJ*, 334, 165–166.

21. Department of Health (2001), *Reference Guide to Consent for Examination or Treatment*, DOH, London.
22. U.S. DHHS (2006), Guidance for Institutional Review Boards, Clinical Investigators and Sponsors Exception from Informed Consent Requirements for Emergency Research Draft Guidance, U.S. Department of Health and Human Services, Food and Drug Administration, Good Clinical Practice Program, Center for Biologics Evaluation and Research, Center for Drug Evaluation and Research Center for Devices and Radiological Health, Office of Regulatory Affairs, July.
23. Anon. (2006), Medicines for Human Use (Clinical Trials) Amendment (No. 2) Regulations 2006, Statutory Instrument 2006 No 2984.
24. Choonara, I. (2000), Clinical trials of medicines in children [editorial], *BMJ*, 321, 1093–1094.
25. Conroy, S., Choonara, I., Impicciatore, P., et al. (2000), Survey of unlicensed and off label drug use in paediatric wards in European countries. European Network for Drug Investigation in Children, *BMJ*, 320, 79–82.
26. Bush, A. (2006), Clinical trials research in pediatrics: Strategies for effective collaboration between investigator sites and the pharmaceutical industry, *Paediatr. Drugs*, 8, 271–277.
27. Ernest, T. B., Elder, D. P., Martini, L. G., et al. (2007), Developing paediatric medicines: Identifying the needs and recognizing the challenges, *J. Pharm. Pharmacol.*, 59, 1043–1055.
28. Cuzzolin, L., Atzei, A., and Fanos, V. (2006), Off-label and unlicensed prescribing for newborns and children in different settings: A review of the literature and a consideration about drug safety, *Expert Opin. Drug Saf.*, 5, 703–718.
29. Kan, P., and Kestle, J. R. (2007), Designing randomized clinical trials in pediatric neurosurgery, *Childs Nerv. Syst.*, 23, 385–390.
30. Bonati, M., and Pandolfini, C. (2005), DEC-net Collaborative Group. Pediatric clinical trials registry, *CMAJ*, 172, 1159–1160.
31. Bonati, M., and Pandolfini, C. (2005), DEC-NET Collaborative Group. More on compulsory registration of clinical trials: Complete clinical trial register is already reality for paediatrics, *BMJ*, 330, 480.
32. Jacqz-Aigrain, E., Zarrabian, S., Pandolfini, C., et al. (2006), A complete clinical trial register is already a reality in the paediatric field, *Therapie*, 61, 121–124.
33. Steinbrook, R. (2002), Improving protection for research subjects. *N. Engl. J. Med.*, 346, 1425–1430.
34. Quinn, S. C. (2004), Ethics in public health research: Protecting human subjects: The role of community advisory boards, *Am. J. Public Health*, 94, 918–922.
35. Nee, P. A., and Griffiths, R. D. (2002), Ethical considerations in accident and emergency research, *Emerg. Med. J.*, 19, 423–427.
36. Royal College of Physicians (1991), Fraud and misconduct in medical research: Causes, Investigations and Prevention. Report of a Working Party, Royal College of Physicians, London.
37. Silverman, H. (2007), Ethical issues during the conduct of clinical trials, *Proc. Am. Thorac. Soc.*, 4, 180–184.
38. Morse, M. A., Califf, R. M., and Sugarman, J. (2001), Monitoring and ensuring safety during clinical research, *JAMA*, 285, 1201–1205.
39. Silverman, H. (2007), Ethical issues during the conduct of clinical trials, *Proc. Am. Thorac. Soc.*, 4, 180–184.
40. Salzman, J. G., Frascone, R. J., Godding, B. K., et al. (2007), Implementing emergency research requiring exception from informed consent, community consultation, and public disclosure, *Ann. Emerg. Med.*, 50, 448–455.

41. Watters, D., Sayre, M. R., and Silbergleit, R. (2005), Research conditions that qualify for emergency exception from informed consent, *Acad. Emerg. Med.*, 12, 1040–1044.
42. Morris, M. C., Nadkarni, V. M., Ward, F. R., et al. (2004), Exception from informed consent for pediatric resuscitation research: Community consultation for a trial of brain cooling after in-hospital cardiac arrest, *Pediatrics*, 114, 776–781.
43. Richardson, L. D. (2005), The ethics of research without consent in emergency situations, *Mt. Sinai J. Med.*, 72, 242–249.
44. Morris, M. C., and Nelson, R. M. (2007), Randomized, controlled trials as minimal risk: An ethical analysis, *Crit. Care Med.*, 35, 940–944.
45. Freedman, B. (1987), Equipoise and the ethics of clinical research, *N. Engl. J. Med.*, 317, 141–145.
46. Sugarman, J., Kass, N. E., Goodman, S. N., et al. (1998), What patients say about medical research, *IRB*, 20, 1–7.
47. Abboud, P. A., Heard, K., Al-Marshad, A. A., et al. (2006), What determines whether patients are willing to participate in resuscitation studies requiring exception from informed consent? *J. Med. Ethics*, 32, 468–472.
48. EU (2001), Directive 2001/20/EC of the European parliament and of the council of 4 April 2001 on the approximation of the laws, regulations and administrative provisions of the member states relating to the implementation of good clinical practice in the conduct of clinical trials on medicinal products for human use, *Official J. Eur. Comm.*, L121, 34–44; http://www.eortc.be/Services/Doc/clinical-EU-directive-04-April-01.pdf.
49. Singer, E. A., and Mullner, M. (2002), Implications of the EU directive on clinical trials for emergency medicine, *BMJ*, 324, 1169–1170.
50. Cone, D. C., and O'Connor, R. E. (2005), Are US informed consent requirements driving resuscitation research overseas? *Resuscitation*, 66, 141–148.
51. Mader, T. J., and Playe, S. J. (1997), Emergency medicine research consent form readability assessment, *Ann. Emerg. Med.*, 29, 534–539.
52. Paasche-Orlow, M. K., Taylor, H. A., and Brancati, F. L. (2003), Readability standards for informed-consent forms as compared with actual readability, *N. Engl. J. Med.*, 348, 721–726.
53. Demircan, C., Cikriklar, H. I., Engindeniz, Z., et al. (2005), Comparison of the effectiveness of intravenous diltiazem and metoprolol in the management of rapid ventricular rate in atrial fibrillation, *Emerg. Med. J.*, 22, 411–414.
54. Davey, M. J., and Teubner, D. (2005), A randomized controlled trial of magnesium sulfate, in addition to usual care, for rate control in atrial fibrillation, *Ann. Emerg. Med.*, 45, 347–353.
55. Thomas, S. P., Guy, D., Wallace, E., et al. (2004), Rapid loading of sotalol or amiodarone for management of recent onset symptomatic atrial fibrillation: A randomized, digoxin-controlled trial, *Am. Heart J.*, 147, E3.
56. Kim, M. H., Morady, F., Conlon, B., et al. (2002), A prospective, randomized, controlled trial of an emergency department-based atrial fibrillation treatment strategy with low-molecular-weight heparin, *Ann. Emerg. Med.*, 40, 187–192.
57. Topol, E. J. (2003), Current status and future prospects for acute myocardial infarction therapy, *Circulation*, 108 (16 Suppl 1), III6–13.
58. Kandzari, D. E. (2006), Evolving antithrombotic treatment strategies for acute ST-elevation myocardial infarction, *Rev. Cardiovasc. Med.*, 7 (Suppl 4), S29–37.
59. Ibbotson, T., McGavin, J. K., and Goa, K. L. (2003), Abciximab: An updated review of its therapeutic use in patients with ischemic heart disease undergoing percutaneous coronary revascularisation, *Drugs*, 63, 1121–1163.

60. Montalescot, G., Barragan, P., Wittenberg, O., et al.; ADMIRAL Investigators (2001), Abciximab before direct angioplasty and stenting in myocardial infarction regarding acute and long-term follow-up. Platelet glycoprotein IIb/IIIa inhibition with coronary stenting for acute myocardial infarction, *N. Engl. J. Med.*, 344, 1895–1903.
61. Stone, G. W., Grines, C. L., Cox, D. A., et al.; Controlled Abciximab and Device Investigation to Lower Late Angioplasty Complications (CADILLAC) Investigators (2002), Comparison of angioplasty with stenting, with or without abciximab, in acute myocardial infarction, *N. Engl. J. Med.*, 346, 957–966.
62. De Luca, G., Suryapranata, H., Stone, G. W., et al. (2005), Abciximab as adjunctive therapy to reperfusion in acute ST-segment elevation myocardial infarction. Meta-analysis of randomized trials, *JAMA*, 293, 1759–1765.
63. Anon. (1997), Randomised placebo-controlled trial of effect of eptifibatide on complications of percutaneous coronary intervention: IMPACT-II. Integrilin to Minimise Platelet Aggregation and Coronary Thrombosis-II, *Lancet*, 349, 1422–1428.
64. ESPRIT Investigators (2000), Enhanced suppression of the platelet IIb/IIIa receptor with integrilin therapy. Novel dosing regimen of eptifibatide in planned coronary stent implantation (ESPRIT): A randomised, placebo-controlled trial, *Lancet*, 356, 2037–2044.
65. PURSUIT Investigators (1998), Inhibition of platelet glycoprotein IIb/IIIa with eptifibatide in patients with acute coronary syndromes. Platelet glycoprotein IIb/IIIa in unstable angina: Receptor suppression using integrilin therapy, *N. Engl. J. Med.*, 339, 436–443.
66. Zeymer, U. (2007), The role of eptifibatide in patients undergoing percutaneous coronary intervention, *Expert Opin. Pharmacother.*, 8, 1147–1154.
67. Valgimigli, M., Bolognese, L., Anselmi, M., et al. (2007), Two-by-two factorial comparison of high-bolus-dose tirofiban followed by standard infusion versus abciximab and sirolimus-eluting versus bare-metal stent implantation in patients with acute myocardial infarction: Design and rationale for the MULTI-STRATEGY trial, *Am. Heart J.*, 154, 39–45.
68. Anon. (1995), Tissue plasminogen activator for acute ischemic stroke. The National Institute of Neurological Disorders and Stroke rt-PA Stroke Study Group, *N. Engl. J. Med.*, 333, 1581–1587.
69. Hacke, W., Kaste, M., Fieschi, C., et al. (1998), Randomised double-blind placebo-controlled trial of thrombolytic therapy with intravenous alteplase in acute ischemic stroke (ECASS II). Second European–Australasian Acute Stroke Study Investigators, *Lancet*, 352, 1245–1251.
70. Clark, W. M., Wissman, S., Albers, G. W., et al. (1999), Recombinant tissue-type plasminogen activator (Alteplase) for ischemic stroke 3 to 5 hours after symptom onset. The ATLANTIS Study: A randomized controlled trial. Alteplase Thrombolysis for Acute Noninterventional Therapy in Ischemic Stroke, *JAMA*, 282, 2019–2026.
71. Hacke, W., Donnan, G., Fieschi, C., et al.; ATLANTIS Trials Investigators, ECASS Trials Investigators, NINDS rt-PA Study Group Investigators (2004), Association of outcome with early stroke treatment: Pooled analysis of ATLANTIS, ECASS, and NINDS rt-PA stroke trials, *Lancet*, 363, 768–774.
72. Hanley, D., and Hacke, W. (2005), Critical care and emergency medicine neurology in stroke, *Stroke*, 36, 205–207.
73. IMS II Trial Investigators (2007), The Interventional Management of Stroke (IMS) II Study, *Stroke*, 38, 2127–2135.
74. Rivers, E., Nguyen, B., Havstad, S., et al. (2001), Early goal-directed therapy in the treatment of severe sepsis and septic shock, *N. Engl. J. Med.*, 345, 1368–1377.

75. Jones, A. E., Craddock, P. A., Tayal, V. S., et al. (2005), Diagnostic accuracy of left ventricular function for identifying sepsis among emergency department patients with non-traumatic symptomatic undifferentiated hypotension, *Shock*, 24, 513–517.
76. Hartman, D. L., Bernard, G. R., Helterbrand, J. D., et al. (1998), Recombinant human activated protein C (rhAPC) improves coagulation abnormalities associated with severe sepsis, *Intensive Care Med.*, 24 (Suppl 1), S77–S77.
77. Bernard, G. R., Vincent, J. L., Laterre, P. F., et al.; Recombinant human protein C Worldwide Evaluation in Severe Sepsis (PROWESS) study group (2001), Efficacy and safety of recombinant human activated protein C for severe sepsis, *N. Engl. J. Med.*, 344, 699–709.
78. Vincent, J. L., Bernard, G. R., Beale, R., et al. (2005), Drotrecogin alfa (activated) treatment in severe sepsis from the global open-label trial ENHANCE: Further evidence for survival and safety and implications for early treatment, *Crit. Care Med.*, 33, 2266–2277.

# *Gastroenterology*

**Lise Lotte Gluud[1] and Jørgen Rask-Madsen[2]**

*[1]Copenhagen Trial Unit, Cochrane Hepato-Biliary Group, Copenhagen, Denmark*
*[2]Department of Medical Gastroenterology, Herlev Hospital, University of Copenhagen, Herlev, Denmark*

**Contents**

1 Preface and Introduction to Evidence-Based Gastroenterology
2 Definitions and Classification of Clinical Trials
3 Observational Studies
4 Randomized Controlled Trials
5 Blinding in Randomized Controlled Trials
6 Sample Size Calculations and Statistical Power
7 Follow-up and Attrition Bias
8 Systematic Reviews
9 Publication Bias and Related Biases
10 Concluding Remarks
References

## 1 PREFACE AND INTRODUCTION TO EVIDENCE-BASED GASTROENTEROLOGY

Traditionally, clinical decisions were based on experience, but sometimes experience may be misleading. We are often prone to remember the best and the worst cases of a disease. The natural course of diseases is fluctuating and humans have the capacity to recover spontaneously. The unsystematic collection of data and limitations in human processing limits the credibility of recommendations based only on experience. Accordingly, evidence-based medicine combining clinical experience with research evidence is gradually replaced by traditional experience-based

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

practice. Evidence-based gastroenterology involves literature searches and quality assessments to identify the most valid research. The internal validity of a trial refers to the credibility of the results and depending on its quality and risk of bias. Low-quality trials have a considerable risk of bias and low internal validity. High-quality trials with adequate bias control have a high internal validity. With a low internal validity of a trial, the assessment of its external validity becomes irrelevant because the results lack credibility. If the internal validity is adequate, assessment of the external validity is still necessary before the trial results are used in clinical practice. The external validity of clinical trials depends on the extent to which the results may be extrapolated. The external validity depends on the characteristics of patients included, treatment regimens, and the trial setting. In clinical practice, treatments are often used for larger patient groups with less strict selection criteria. The results of a trial with highly selective inclusion criteria performed at a specialized unit may, therefore, be difficult to extrapolate and use in general clinical practice.

## 2 DEFINITIONS AND CLASSIFICATION OF CLINICAL TRIALS

Research in gastroenterology deals with diseases of the digestive system, including the esophagus, stomach, intestine, liver, gallbladder, and pancreas. In clinical gastroenterology most trials assess the effects of various drugs, but interventional procedures may also be subject to clinical trials, such as efficacy and safety of endoscopic therapy and primary as well as secondary prevention. Large, high-quality randomized clinical trials and systematic reviews of several randomized clinical trials are considered the gold standard. However, the clinician often has to judge a number of clinical trials with inconsistent results and numerous methodological deficiencies. If no randomized clinical trials are available, observational studies may be considered the best available source of evidence.

## 3 OBSERVATIONAL STUDIES

The classical observational study designs are case–control studies, cohort studies, case series, and case reports. Cohort studies follow a group or patients (cohort) pro- or retrospectively. Case–control studies are based on patients with certain diseases or genetic characteristics (cases) and controls without the specific property but otherwise similar prognostic profile. Cohort studies start out with the identification of the intervention, for example, a group of patients undergoing stricturoplasty for Crohn's disease [1], whereas case–control studies initially identify a group of patients with a certain disease, for example, cases with colorectal cancer and matched controls without the disease, before attempts are made to detect the influence of intervention, for example, use of antidepressents [2]. Case series describe the outcome of a certain group of patients, for example, patients who were treated for presumed exacerbation of Crohn's disease but were subsequently found to have underlying small-bowel carcinoma [3]. Case reports describe unusual cases of individual patients with rare diseases, unusual adverse events, or treatment effects, for example, development of subfulminant hepatitis B after treatment with infliximab for Crohn's disease [4].

Prospective cohort studies are generally considered the most valid observational design. If a disease is rare, or if the course of a disease is protracted, performing a prospective cohort study may, however, not be feasible. For example, the development of cirrhosis and liver failure after infection with hepatitis B or C takes several years [5]. In that case, retrospective cohort studies or case–control studies may be considered.

Although randomized controlled trials are the gold standard for evaluation of intervention effects, this does not mean that observational studies have little or no value. However, if interventions have dramatic effects in line with penicillin for pneumonia as compared with no treatment, randomized controlled trials are not required. Observational studies may also provide important information when we are unable to perform randomized controlled trials, for example, if assessing behavior or if requiring extremely large samples of patients for the assessment of rare adverse events. In other cases, randomized trials may be unethical, for example, if the causative agent is considered potentially harmful. Furthermore, the evidence generated in observational studies may also be useful if complementary or the basis of randomized controlled trials. Although important information may sometimes be contained in observational studies, the information should only be used with due care considering the risk of biases.

One of the most important roles of observational studies lies in the postmarketing surveillance of adverse events. Rare adverse events may be difficult to detect in randomized controlled trials because very large samples are necessary. One example is terlipressin, which has been assessed for treatment of hepatorenal syndrome [6]. Today, only few randomized controlled trials have addressed this question. None of the trials detected serious adverse events. However, one case report suggested that terlipressin may be associated with worsening of cerebral hyperemia [7]. Another case report suggested that terlipressin may be associated with acute ST-segment elevation myocardial infarction [8]. Therefore, performing observational studies that are sufficiently large to detect the risk of serious adverse events seems highly important. One example is the introduction and widespread use of laparoscopic cholecystectomy in the 1990s, which was associated with a dramatic increase in the incidence of bile duct injuries [9]. Several studies have reviewed the outcomes of surgical management of bile duct injuries, but data on the current incidence of injuries are scarce. The completion of a large prospective cohort on the outcomes of laparoscopic cholecystectomy today may provide important information to patients as well as health care workers. The information may also be used in the planning of future randomized controlled trials that are considered necessary to determine whether patients fare better with or without the intervention.

Another important role for observational studies lies in the evaluation of intervention effects on long-term outcomes. The practical difficulties in maintaining long-term prospective randomized controlled trials can make the task impossible. One example of such a situation is the treatment of chronic hepatitis C with antiviral substances, for example, interferon and ribavirin [10]. Chronic hepatitis C has a protracted course. To determine whether antiviral therapy affects morbidity or mortality, patients need to be followed for decades. This may logistically be nearly impossible. Furthermore, the risk of attrition bias is considerable and increases with the duration of follow-up. Finally, the efficacy of today's recommended antiviral treatment for chronic hepatitis C (interferon or peg interferon combined with riba-

virin) may for ethical reasons be prohibitive for preserving a nontreatment control group. Several randomized controlled trials have been performed, but all have focused on biological response, that is, clearance of the hepatitis C virus RNA (ribonucleic acid) from the blood. None of these trials have established whether the demonstrated efficacy of treatment translates into a meaningful effect on clinical outcomes [10]. Therefore, we have to rely on observational studies to determine the outcome of treatment.

As previously described, a retrospective study design, which is used in some observational studies, increases the risk of bias due to factors that change with time, recall bias, and differential measurement errors [11, 12]. However, one of the most important limitations in prospective as well as retrospective observational studies lies in the risk of selection bias. In observational studies, prognostic factors determine whether patients are allocated to intervention or control groups. This is known as confounding by indication, which may lead to systematic differences between comparison groups [13]. When systematic differences exist, estimates of intervention benefits may be incorrect. It is impossible to determine the size of the effect of the differences on the estimated intervention effects, why separate evaluations of individual trials or studies are necessary. In some cases, differences may be derived from analyses comparing groups of several observational studies and randomized controlled trials. One example may be found in a systematic review on artificial and nonartificial support systems for acute and acute-on-chronic liver failure [14]. The review included observational studies as well as randomized controlled trials. Formally, the inclusion criteria used in the different studies and trials were comparable. All patients who were allocated to the control groups received comparable standard medical regimens. When the results of the individual randomized controlled trials were combined in a meta-analysis, the support systems did not appear to reduce mortality significantly compared with standard medical therapy. Although inclusion criteria and the characteristics of included patients at baseline were similar, the control group mortality rates were significantly higher in the observational studies than in the randomized controlled trials. No significant difference was found when comparing the mortality rates of patients allocated to the intervention groups. Accordingly, the observational studies found a statistically significant positive effect of the intervention as compared with the randomized controlled trials. This difference suggests that the observational studies had a skewed allocation of patients with the worst prognosis to control groups. Thus, the observational studies may have overestimated the intervention benefit. The strength of the association warrants separate evaluations in different situations, although observational studies are generally more susceptible to bias than randomized controlled trials. In a similar review, considerable differences were found between estimated intervention effects in comparisons of observational studies and randomized controlled trials [13].

Analyses of all comparisons showed that the effect estimates in the observational studies ranged from an underestimation of effect by 76% to an overestimation of effect by 160% [13]. In another methodological study, odds ratios generated by 168 observational studies and 240 randomized controlled trials within 45 different topics were compared [15]. All trials and studies were included in a meta-analysis with binary outcomes. Overall, the observational studies tended to generate larger summary odds ratios suggesting a more beneficial effect of the intervention. Bias associated with nonrandom allocation has also been analyzed in a review with

studies that compared randomized controlled trials with observational studies [16]. The review reported that nonrandom allocation was related to overestimation as well as underestimation of treatment effects. The variation in the results of observational studies was increased due to haphazard differences in case mix between groups.

Four strategies for case-mix adjustment were subsequently evaluated by generating nonrandomized studies from two large randomized controlled trials [16]. Participants were resampled according to allocated treatment, center, and period. None of the strategies adjusted adequately for bias in historical or concurrent controlled studies. Logistic regression was found to increase bias due to misclassifications and measurement errors in confounding variables as well as differences between conditional and unconditional odds ratio estimates of treatment effects.

## 4 RANDOMIZED CONTROLLED TRIALS

Random allocation means that all patients have a known chance of being allocated to one of the intervention groups and that the allocation of the next patient is unpredictable [17]. To keep the allocation of patients unpredictable, both adequate generation of an allocation sequence and adequate concealment of allocation are required. The allocation sequence may consist of random numbers generated by computers or tables. Allocation concealment may consist of randomization through independent centers or serially numbered identical sealed packages. If the next assignment is known, enrollment of certain patients may be prevented or delayed to ensure that they receive the treatment that is believed to be superior. Theoretically, serially numbered sealed envelopes may provide adequate allocation concealment, although there is some evidence suggesting that this may not be true. In some cases, envelopes were opened before or after patients were excluded [18]. In other cases envelopes have been transilluminated [19]. The adequacy of using serially numbered sealed envelopes therefore seems debatable.

In theory, adequate randomization is necessary to obtain adequate control of bias. On the other hand, empirical evidence is necessary to determine whether the effect of randomization is only hypothetical or whether randomization affects the results and subsequently the conclusions drawn from randomized controlled trials. To address this question, two methodological studies of cohorts of clinical trials estimated the association between the methods of randomization and the effects of intervention [20, 21]. All trials were included in meta-analyses. One study included trials from the field of obstetrics and gynecology [20] and one included trials from a variety of disease areas [21]. Neither of the studies revealed a significant association between allocation sequence generation and intervention effects, but both showed that allocation concealment methods were significantly associated with estimated intervention effects. The analyses made in both methodological studies illustrated that inadequate allocation concealment was associated with significantly more positive estimates of intervention effects. These results suggest that inadequate allocation concealment may lead to exaggerated intervention effects. On the other hand, the results may also show bias due to selective publication of small or low-quality trials with positive results [22, 23]. Because there is no defined gold standard, it may be difficult to determine whether the trials with or the trials without

adequate allocation concealment were more correct. To address these concerns, a subsequent study used very large randomized controlled trials as a reference group [24]. The reference group included randomized controlled trials with more than 1000 participants. Each of the included trials, that is, the trials in the reference group and the smaller trials, were included in the meta-analyses. Each meta-analysis contained at least one large trial. Subsequently, analyses were performed to compare the results of trials in the reference groups with the results of the smaller trials. The analyses showed that, on average, the small trials without adequate allocation sequence generation or allocation concealment overestimated intervention benefits. The results of the small trials with adequate generation of the allocation sequence or adequate allocation concealment were not significantly different from the results of the trials in the reference group.

These results support the importance of adequate randomization to bias control in randomized controlled trials. Similar subsequent methodological studies of randomized controlled trials have also examined the association between randomization and intervention effects [25]. The evidence generated in these studies was combined in a random effects meta-analysis including the summary results from the individual studies. The results suggested that odds ratios were about 12% more positive in trials without adequate allocation sequence generation as compared to odds ratios in trials with adequate allocation sequence generation. Therefore, trials reporting inadequate methods may overestimate intervention benefits due to inadequate bias control. Similarly, odds ratios generated in trials without adequate allocation concealment were about 21% more positive than odds ratios generated in trials with adequate allocation concealment. Both components of randomization process therefore seem to be important. However, the meta-analyses on the effect of allocation sequence generation and allocation concealment also found that there was a considerable heterogeneity between studies. The heterogeneity may be related to the disease area, the type of intervention, trial inclusion criteria, and the classification of adequate randomization in the individual methodological studies. The variation suggests that caution is required when making inferences and recommendations for assessment of bias control. Using the described components as exclusion criteria, for example, disregarding all trials without adequate allocation concealment is not justified. Simply reducing the estimated intervention effect in randomized controlled trials that do not describe adequate allocation concealment seems also problematic. Thus, the quality of individual trials and meta-analyses has to be evaluated separately.

Many trials are described as randomized without reporting randomization methods. A number of cohort studies of randomized controlled trials suggest that the proportion of trials with adequate allocation sequence generation ranges from 1 to 52% (median 37%) [25]. The proportion with adequate allocation concealment ranges from 2 to 39% (median 25%). Some of the variation may depend on the disease areas evaluated or different classifications of randomization methods in the cohort studies, or other factors [26, 27].

Whether the lack of reported randomization methods reflect the actual conduct of a published randomized controlled trial is difficult to establish. In theory, inadequate reporting may hide important flaws in the design of a trial. On the other hand, a high-quality randomized controlled trial may be overlooked because of inaccurate reporting. In a methodological study, the reported methods used for allocation

concealment were extracted from the published reports of 105 randomized controlled trials [28]. The reported allocation concealment methods in full-text publications were subsequently compared with the information on allocation concealment obtained through direct communication with the author(s). The results of the study showed that several trials had adequate allocation concealment methods that were not described in the published report [29]. Another methodological study reached a different conclusion [30]. This study compared the reported descriptions of the methods used for allocation concealment in publications and protocols of 102 randomized controlled trials. The analyses showed that most of the trials with unclear allocation concealment in the published trial reports also had unclear allocation concealment in the protocol. The evidence concerning the discrepancies between the conduct and report of randomized controlled trials is equivocal. Additional evidence is needed to clarify this question.

In 1999, an observational study including all 235 randomized controlled trials published in the journal *Hepatology* from the initiation in 1981 through August 1998 was published [31]. Only 52% of the included trials reported adequate generation of the allocation sequence. The proportion of trials reporting adequate allocation concealment was 34%. In a similar observational study of all 383 randomized clinical trials published in the journal *Gastroenterology* as original articles from 1964 to 2000 were reviewed [26]. The individual authors in all of the included publications had described the trials as being randomized. However, only 42% of the trials reported adequate generation of the allocation sequence. The proportion of trials reporting adequate allocation concealment was 39%. Unlike the study on randomized controlled trials published in *Hepatology*, the reporting quality improved significantly in the mid-1990s.

Nevertheless, both studies found that there was still room for improvement. Whether these findings were specific to these two journals was evaluated in a subsequent study including 616 hepato-biliary randomized controlled trials published from 1985 to 1996 in 12 different MEDLINE-indexed journals [27]. All trials were described as randomized by the individual authors. However, the reported generation of the allocation sequence was only described in 48% of the included trials. Fifty-two percent of the trials did not include a description of the methods used for generation of the allocation sequence. In 38% of the included trials, the allocation concealment was described adequately. In the remaining 62% of the trials, the allocation concealment was not described. A number of analyses were performed to evaluate potential predictors for adequate reporting of the randomization methods, that is, allocation sequence generation and allocation concealment. These analyses focused on the importance of funding and disease area. Based on the published reports, 47% of the trials received no external funding, 30% were funded by nonprofit organizations, and 23% were funded by for-profit organizations. The proportion of trials with adequate allocation sequence generation and allocation concealment was not significantly different among trials funded by profit or nonprofit organizations. When these trials were combined and compared with the trials not reporting external funding, the analyses showed that trials with external funding were significantly more likely to report adequate generation of the allocation sequence. The proportion of trials with adequate allocation concealment was insignificantly different in the two groups. Further analyses revealed that also the proportion of trials with funding was insignificantly different within different disease areas. However,

the proportions of trials with adequate generation of the allocation sequence or adequate allocation concealment were significantly associated with the disease area. Trials dealing with interventions for some disease areas reported adequate randomization methods significantly more often than trials in other areas. Several other aspects, including the sample size and number of clinical sites, have been suggested as potential predictors of the quality of the reported randomization. However, additional studies in this area are still warranted to establish the different patterns.

## 5 BLINDING IN RANDOMIZED CONTROLLED TRIALS

In randomized controlled trials, the term *blinding* refers to keeping participants, health care providers, data collectors, outcome assessors, or data analysts unaware of the assigned intervention. Double blinding may refer to blinding of participants and health care providers, investigators, data collectors, judicial assessors, or data analysts. The specific methods used to maintain blinding in trial reports are often missing. Furthermore, many researchers disagree on the correct definition of double blinding. It is therefore recommended that, in individual trials, the term double blinding is provided with clear information about who were blinded and how the blinding was performed. Sometimes the nature of the intervention precludes double blinding, but blinded outcome assessment and data analyses are usually possible. To ensure adequate double blinding, the interventions compared must be similar. If an intervention is compared to no intervention, an identical placebo must be used. Any difference in taste, smell, or appearance may destroy blinding. One example may be found in a randomized controlled trial on the effect of nicotine gum on smoking cessation [32]. The trial compared the effect of nicotine gum versus a placebo gum. The authors tried to disguise the placebo gum by preparing wrappings, which suggested that the contents included nicotine. However, Wrigley's chewing gum was used as placebo. It seems likely, therefore, that participants correctly guessed whether they belonged to the intervention group or the control group.

Another example of a break in blinding is found in a randomized controlled trial on ascorbic acid for the common cold [33]. The trial was described as using a double-blind randomized design. Participants were employees of the National Institutes of Health. As no established effective treatment was known, a placebo containing lactulose was used to maintain blinding. However, the results of the trial turned out to be questionable because many participants tasted their capsules and guessed in which group they were.

Although there is a clear theoretical association between blinding and the control of bias, empirical evidence is required to determine the actual size and direction of the association. Six methodological studies of randomized controlled trials have analyzed the association between double blinding and intervention effects [25]. Each of the randomized controlled trials was included in meta-analyses assessing binary outcomes. Analyses were subsequently performed to compare odds ratios in randomized controlled trials with or without blinding. Two of these studies revealed that randomized controlled trials without double blinding overestimated intervention effects compared to randomized double-blind trials. The remaining four methodological studies found no significant association between blinding and estimates

of intervention benefits. To combine the empirical evidence generated in the methodological studies, a random effects meta-analysis was performed. This meta-analysis showed no significant differences between odds ratios of intervention effects in groups of double-blind trials compared to trials without double blinding. The meta-analysis did, however, reveal a considerable difference between study variation that may be related to the nature of the disease or the intervention. The meta-analyses included trials from various disease areas including cardiology, gynecology, obstetrics, psychiatry, and smoking cessation. Accordingly, the interventions assessed included diagnostic measures and drugs as well as surgical procedures. The variation may reflect that some interventions are difficult to blind. If we perform double-blind trials on, for example, drugs associated with adverse events blinding may be ineffective. The type of outcome may be equally important. Hard outcomes may be less prone to assessment bias than subjective outcomes. Therefore, trials evaluating the effect of drugs on, for example, mortality may be less susceptible to bias than trials evaluating the effect of drugs on pain. The effect of blinding is highly unpredictable, and separate analyses of the effect of blinding in individual trials and meta-analyses are warranted.

In a cohort study including 616 hepato-biliary randomized controlled trials published during 1985–1996, only 34% were double blind [27]. The proportion of double-blind trials varied significantly in different disease areas. Trials on interventions for gallstones were significantly less often double blind as compared to, for example, trials on portal hypertension. To some extent, the variation reflects that some interventions are more difficult to blind than others. One example is endoscopic procedures. Some trials have attempted to perform "sham" endoscopy, but maintaining the blinding in such cases is obviously difficult. Blinding the effect of drugs may also be difficult if there are specific characteristic effects (e.g., lowering of blood pressure and heart frequency when using β blockers for prevention of bleeding esophageal varices). Characteristic adverse effects associated with treatment may also cause a break in the blinding. If the maintenance of adequate blinding is questionable, it may be relevant to test the possibility of a break in blinding before performing the trial [34]. The results of such pretrial assessments may be used in the development of the final trial protocol. In some cases, we do not need empirical evidence to establish that blinding is impossible for patients or investigators, for example, when assessing endoscopic procedures or interventions such as liver support systems [14]. Trying to maintain blinding may only result in making the trial more complicated. However, it is always possible to maintain some form of blinded outcome assessment or blinded data analyses. Unfortunately, very few studies on randomized controlled trials in gastroenterology report on these aspects.

## 6 SAMPLE SIZE CALCULATIONS AND STATISTICAL POWER

Random error may occur in any direction and subsequently lead to false-positive (type I error) or false-negative results (type II error). In randomized controlled trials, the risk of random error depends on the sample size and the size of the intervention effect. The larger the sample size and the intervention effect, the smaller the risk of random error. Accordingly, small trials on interventions with moderate

effects have a substantial risk of being subject to random error and consequently to produce false-positive or false-negative conclusions. Large randomized controlled trials on the same interventions with moderate effects have a lower risk of random error than the small trials. Small trials on interventions with substantial effects have a smaller risk of generating results that lead to false-negative or false-positive conclusions than small trials on interventions with moderate intervention effects. One way to determine the risk of random error is to calculate confidence intervals or levels of statistical significance. Confidence intervals and the level of statistical significance provide an estimate of the error that may occur and thus reflect the precision of the statistical estimates. However, these estimates do not tell whether they are clinically significant or meaningful.

Confidence intervals are related to the concept of the statistical power. The larger the confidence interval, the less power a study has to detect differences between outcomes in groups of patients allocated to the intervention or the control group. Sample size calculations are required in randomized trials because inadequate statistical power can lead to false-negative results. The calculations should account for the minimum relevant treatment effect, acceptable probabilities of type I and II errors and losses to follow-up [35]. The first parameter, that is, the minimum relevant treatment effect, is adjustable and sensitive. If you reduce the relevant difference by half, four times as many patients are needed. The risk of a type I error ($\alpha$) is usually set to 5% and the risk of a type II error ($\beta$) is usually set to 10 or 20%. Subsequently, the statistical power ($1 - \beta$) is 90 or 80%. The power of a trial reflects the risk of overlooking intervention effects. Suppose you want to perform a trial on a specific treatment, for example, a drug that reduces mortality from 40 to 20% and you set the risk of a type I error ($\alpha$) to 5% and include 90 patients both in the treatment arm and the control arm, your trial will have 80% power to detect the true treatment effect. If you repeat the trial a 100 times, 20 of the trials will overlook the true treatment effect. If you evaluate the same drug, but include 45 patients in each treatment arm instead, the sample size corresponds to a power of 55%. If you repeat the trial a 100 times, 45 of your trials will overlook the true treatment effect. If you search for evidence and identify 100 randomized controlled trials, the entire sample of trials must be evaluated. Including a subgroup of trials, for example, the trials overlooking the intervention effect, means that your conclusions become incorrect. In practice, looking at larger samples of randomized controlled trials may be done through a systematic review, making a meta-analysis of the trials.

In a cohort study including all 235 randomized controlled trials published in the journal *Hepatology*, from the initiation in 1981 through August 1998, were assessed [31]. All trials dealt with hepato-biliary diseases. Only 26% of the trials reported sample size calculations. In similar cohort studies of randomized controlled trials from various disease areas, sample size calculations were only reported in 8–38% of the trials included [25]. This is unfortunate because the sample size calculations provide crucial information about the reliability of the results. If the preset sample size remains unreported, it becomes difficult to evaluate whether the planned sample size was reached or whether the trial was extended beyond the planned size or was terminated at an arbitrary time point.

Two studies evaluated statistical power in trials with statistically insignificant outcomes [36, 37]. Both studies found that most trials had insufficient power to detect clinically relevant treatment effects. The relatively small sample size of randomized

trials suggests that only a few have the recommended statistical power [25]. For example, one observational study included 383 randomized controlled trials published in the journal *Gastroenterology* [26]. On average, the randomized controlled trials included 43 patients per intervention arm (standard error of the mean was 4 patients). If you perform a trial with 45 patients per intervention arm, the trial will have a 90% power to detect a reduction in rates of mortality from 60 to 25% (if the risk of a type I error is set to 0.05). However, very few interventions have such dramatic effects. If we repeat the trial, evaluating a drug that reduces mortality rates from 60 to 40%, the statistical power of the trial will decrease to 39%. Therefore, it seems likely that a number of trials overlooked clinically important significant intervention effects. In a randomized controlled trial including patients admitted with bleeding esophageal varices, the effect of emergency sclerotherapy with sodium tetradecyl sulfate was compared with octreotide infusion [38]. The trial included only 100 patients, although the sample size calculations reported by the authors showed that a sample size of 1800 patients was required. This means that the trial had a 5% chance of demonstrating a statistically significant result, if the treatment effect estimated—also by the authors—was correct. The authors concluded that emergency sclerotherapy and octreotide infusions were equally efficacious in controlling variceal hemorrhage. The conclusion is debatable considering the low statistical power of the trial. Here it is again important to remember that absence of evidence is no evidence of absence. The fact that no significant intervention effect was identified does not mean that it does not exist. It is possible that we simply used the wrong method to look for such an effect. A number of studies have evaluated the sample size of published randomized controlled trials on hepato-biliary and other gastroenterological diseases. Similarly, the average sample size of randomized trials in clinical gastroenterology suggests that effective interventions may have been disregarded based on insufficient grounds. Much larger trials are necessary to conclude that an intervention is ineffective or that two interventions are equally effective.

Sample size calculations are required before performing randomized controlled trials because inadequate statistical power may lead to false-negative results [39]. The calculations should account for the minimum relevant treatment difference, acceptable probabilities of type I and II errors, and losses to follow-up [35, 40]. The first parameter is adjustable and sensitive. If you reduce the relevant difference by half, four times as many patients are needed. The risks of a type I error ($\alpha$) is usually set to 5%. The risk of a type II error ($\beta$) is usually set to 10 or 20%. The corresponding power ($1 - \beta$), which indicates the risk of overlooking an effect of the intervention, is usually set to 90 or 80%. In cohort studies of randomized controlled trials, sample size calculations were reported only in 8–38% of the included trials [25]. Without having preset the sample size beforehand, the reader is unable to assess whether the planned sample size was actually reached or whether the trial was extended beyond the planned size or perhaps terminated at an arbitrary time point.

## 7 FOLLOW-UP AND ATTRITION BIAS

Adequate follow-up is essential to avoid attrition bias. Nearly all clinical trials have some missing data due to losses to follow-up, which may affect the results, because they may be related to prognostic factors [41]. Development of methods to obtain

data on patients who are lost to follow-up and to account for losses to follow-up is important. Also, to achieve an adequate or fair interpretation of the results obtained, a clear description of follow-up is essential [42]. In a study on 235 hepato-biliary randomized controlled trials, the numbers or reasons for dropouts or withdrawals were not reported [31].

Several analytical strategies for dealing with missing data have been proposed [43]. One of the most popular methods is to perform the analysis by using the intention-to-treat principle, including all originally randomized patients. The alternative strategy, which is to perform a per-protocol analysis excluding data from patients who were lost to follow-up or had other protocol deviations [43, 44]. However, there are several problems associated with the design of per-protocol analysis. If, for example, an intervention has adverse effects causing dropouts lost to follow-up, the per-protocol analysis will overestimate the benefit of intervention. In a systematic review about randomized controlled trials on interferon and ribavirin for chronic hepatitis C, several patients were lost to follow-up in the individual trials [10]. Many of the protocol deviations and losses to follow-up reflect the occurrence of adverse events related to the use of interferon and ribavirin. An analysis, accounting for all patients randomized, was therefore required to obtain a valid result. If per-protocol analyses were performed, the intervention benefit may have been overestimated, which is why intention-to-treat analysis is generally the most reliable strategy for performing the analyses in systematic reviews and randomized controlled trials.

## 8 SYSTEMATIC REVIEWS

It is debatable whether large randomized controlled trials or systematic reviews provide the best evidence when comparing different interventions. Large randomized controlled trials are often considered the most reliable sources of evidence for assessment of intervention effects. However, a number of cohort studies of hepato-biliary and gastroenterological trials suggest that many trials are too small and have inadequate bias control [26, 31, 45]. The size of the individual trials suggests that several have inadequate statistical power and are too small for identification of significant intervention benefits. One way to overcome these problems is to perform a systematic review using meta-analysis of the identified randomized controlled trials. In meta-analyses, the results of individual trials are combined in a common statistical analysis to increase the statistical power of inferences. Traditional reviews often count the number of supportive trials and choose the view receiving most votes. This may lead to false-negative conclusions if trials are underpowered. In a systematic review, a meta-analysis of 12 randomized controlled trials was performed to compare the effect of interferon with ribavirin versus interferon alone for patients with chronic active hepatitis C, who had not previously responded to antiviral therapy [10]. The primary outcome was sustained clearance of the hepatitis C virus. The sample sizes of the included trials suggested that no single trial had sufficient statistical power to detect clinically relevant differences in treatment effects, while three trials in the systematic review showed a statistically significant difference, suggesting that interferon with ribavirin was the most effective treatment for viral clearance. A narrative review, counting the number of positive and negative trials, may, therefore, reach the conclusion that no significant differences in effect can be

identified. However, when the results of the individual trials were combined in a meta-analysis, the results clearly showed that adding ribavirin to interferon significantly increased the chance of achieving a virological response.

Performing a meta-analysis has other potential advantages than increasing statistical power. The combination of several trials in a meta-analysis increases the extent to which results can be generalized. Furthermore, systematic reviews and meta-analyses make it possible to identify publication bias, or other biases, and to evaluate the risk of overestimated intervention effects due to inadequate bias control. The main disadvantage of systematic reviews is related to their observational design. Subgroup analyses in systematic reviews generally require prospective evaluation [14]. Methods for identification and selection of trials are necessary to avoid bias. If bias in the individual randomized controlled trials remains undetected [46–48], the results of the meta-analysis may be false positive. Therefore, some systematic reviews may remain inconclusive (in spite of statistically significant results) if trials with inadequate bias control are the only available. In a systematic review of randomized controlled trials, the effect of antibiotics and nonabsorbable disaccharides on the development of hepatic encephalopathy was assessed [49]. The results of the meta-analyses of trials identified and included showed that antibiotics had a significantly more positive effect on hepatic encephalopathy as compared to nonabsorbable disaccharides. However, in the included trials the quality of bias control was inadequate and it was impossible to make recommendations useful for clinical practice, although the findings were both clinically and statistically significant.

## 9 PUBLICATION BIAS AND RELATED BIASES

Observational studies have demonstrated that clinical trials with a positive outcome (i.e., demonstrating a statistically significant superiority of the intervention) are significantly more likely to be published than negative trials showing no statistically significant differences between the experimental intervention versus the placebo or other comparative intervention [22]. Such selective publication of trials is known as publication bias. The risk of publication bias is related to the sample size. Small trials with negative results tend to remain unpublished, while, for example, hepatobiliary randomized controlled trials are cited more often if the reported results show a statistically significant difference. Publication bias and related biases affect the possibility of identifying the trial. Thus, trials with positive results are more likely to be identified and included in, for example, systematic reviews and meta-analyses. Such biases may influence our inferences about intervention effects so that the benefits are overestimated. Different proposals have been made to allow identification of publication bias in meta-analyses so that the results and conclusions may be adjusted for the risk of bias.

## 10 CONCLUDING REMARKS

Evidence-based medicine—although extremely time consuming—is gradually replacing the traditional experience-based clinical practice. The number of randomized controlled trials and observational studies performed within the fields of

gastroenterology and hepatology is steadily increasing. The first step for the evidence-based practitioner is to identify the relevant clinical trial. The second step is to assess the internal validity of the research identified, and the final step is to evaluate its external validity so that the evidence may be used in clinical decision making. A number of methodological studies including clinical trials from several disease areas have been made to develop general guidelines for the validity assessment. These guidelines also apply to clinical trials in gastroenterology. Sample size calculations, randomization methods, blinding, and follow-up are important components in the maintenance of adequate bias control, which is important both in the planning stage and during subsequent interpretation of the results. The general methodology is therefore relevant, not only for researchers but also for clinicians using evidence-based medicine in daily practice. Nevertheless, a number of surveys suggest that there may be considerable gaps between evidence and practice. One such survey included specialists in gastroenterology, hepatology, and internal medicine [50]. The survey showed that several treatments and diagnostic procedures were still used, although there was no significant evidence to support their use. On the other hand, some of the treatments based on statistically and clinically relevant beneficial effects in systematic reviews of randomized controlled trials were not used. Additional measures aiming at bridging the gap between evidence and clinical practice are therefore warranted. Also a number of initiatives have been performed to improve the quality of clinical trials, for example, the Baveno conferences on trials in portal hypertension [51]. Hopefully, similar initiatives will be made in other areas of gastroenterology and hepatology to improve the quality of health care.

## REFERENCES

1. Fearnhead, N. S., Chowdhury, R., Box, B., et al. (2006), Long-term follow-up of strictureplasty for Crohn's disease, *Br. J. Surg.*, 93, 475–482.
2. Xu, W., Tamim, H., Shapiro, S., et al. (2006), Use of antidepressants and risk of colorectal cancer: A nested case-control study, *Lancet Oncol.*, 7, 301–308.
3. Shehendere, R. L., Thompson, N., Mansfield, J. C., et al. (2005), Adenocarcinoma as a complication of small bowel Crohn's disease, *Eur. J. Gastroenterol. Hepatol.*, 17, 1255–1257.
4. Millonig, G., Kern, M., Ludwiczek, O., et al. (2006), Subfulminant hepatitis B after infliximab in Crohn's disease: Need for HBV-screening? *World J. Gastroenterol.*, 12, 974–976.
5. Grønbæk, K., Krarup, H. B., Møller, H., et al. (1999), Natural history and etiology of liver disease in patients with previous community-acquired acute non-A, non-B hepatitis. A follow-up study of 178 Danish patients consecutively enrolled in The Copenhagen Hepatitis Acuta Programme in the period 1969–1987, *J. Hepatol.*, 31, 800–807.
6. Uriz, J., Gines, P., Ortega, R., et al. (2000), Terlipressin plus albumin infusion: An effective and safe therapy of hepatorenal syndrome. *J. Hepatol.*, 33, 43–48.
7. Shawcross, D. L., Davies, N. A., Mookerjee, R. P., et al. (2004), Worsening of cerebral hyperemia by the administration of terlipressin in acute liver failure with severe encephalopathy, *Hepatology*, 39, 471–475.
8. Lee, M. Y., Chu, C. S., Lee, K. T., et al. (2006), Terlipressin-related acute myocardial infarction: A case report and literature review, *Kaohsiung J. Med Sci.*, 20, 604–608.

9. Lillemoe, K. D., Melton, G. B., Cameron, J. L., et al. (2000), Postoperative bile duct strictures: Management and outcome in the 1990s, *Ann. Surg.*, 232, 430–441.
10. Kjaergard, L. L., Krogsgaard, K., and Gluud, C. (2001), Interferon alfa with or without ribavirin for chronic hepatitis C: Systematic review of randomized trials, *BMJ*, 323, 1151–1155.
11. Sacks, H., Chalmers, T. C., and Smith, H., Jr. (1982), Randomized versus historical controls for clinical trials, *Am. J. Med.*, 72, 233–240.
12. White, E., Hunt, J. R., and Casso, O. (1998), Exposure measurement in cohort studies: The challenges of prospective data collection, *Epidemiol. Rev.*, 20, 43–56.
13. Randomization to protect against selection bias in healthcare trials (Cochrane Methodology Review). (2002), *In*: *The Cochrane Library*, Wiley, Chichester, UK.
14. Kjaergard, L. L., Liu, J. P., Als-Nielsen, B., et al. (2003), Artificial and bioartificial support systems for acute and acute-on-chronic liver failure: A systematic review, *JAMA*, 289, 217–222.
15. Ioannidis, J. P., Haidich, A. B., Pappa, M., et al. (2001), Comparison of evidence of treatment effects in randomized and nonrandomized studies, *JAMA*, 286, 821–830.
16. Deeks, J. J., Dinnes, J., D'Amico, R., et al. (2003), Evaluating non-randomized intervention studies, *Health Technol. Assess.*, 7, 1–173.
17. Altman, D. G. (1991), Randomization, *BMJ*, 302, 1481–1482.
18. Swingler, G. H., and Zwarenstein, M. (2000), An effectiveness trial of a diagnostic test in a busy outpatients department in a developing country: Issues around allocation concealment and envelope randomization, *J. Clin. Epidemiol.*, 53, 702–706.
19. Schulz, K. F. (1995), Subverting randomization in controlled trials, *JAMA*, 274, 1456–1458.
20. Schulz, K. F., Chalmers, I., Hayes, R. J., et al. (1995), Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials, *JAMA*, 273, 408–412.
21. Moher, D., Pham, B., Jones, A., et al. (1998), Does quality of reports of randomized trials affect estimates of intervention efficacy reported in meta-analyses? *Lancet*, 352, 609–613.
22. Easterbrook, P. J., Berlin, J. A., Gopalan, R., et al. (1991), Publication bias in clinical research, *Lancet*, 337, 867–872.
23. Cochrane Reviewers' Handbook 4.2.1 (updated December 2003) (2004), in *The Cochrane Library*, Wiley, Chichester, UK.
24. Kjaergard, L. L., Villumsen, J., and Gluud, C. (2001), Reported methodologic quality and discrepancies between large and small randomized trials in meta-analyses, *Ann. Intern. Med.*, 135, 982–989.
25. Gluud, L. L. (2006), Bias in clinical intervention research, *Am. J. Epidemiol.*, 163, 493–501.
26. Kjaergard, L. L., Frederiksen, S. L., and Gluud, C. (2002), Validity of randomized clinical trials in gastroenterology from 1964 to 2000, *Gastroenterology*, 122, 1157–1160.
27. Kjaergard, L. L., and Gluud, C. (2002), Funding, disease area, and internal validity of hepatobiliary randomized clinical trials, *Am. J. Gastroenterol.*, 97, 2708–2713.
28. Devereaux, P. J., Choi, P. T., El-Dika, S., et al. (2005), An observational study found that authors of randomized controlled trials frequently use concealment of randomization and blinding, despite the failure to report these methods, *J. Clin. Epidemiol.*, 57, 1232–1236.
29. Schulz, K. F., and Grimes, D. A. (2002), Blinding in randomized trials: Hiding who got what, *Lancet*, 359, 696–700.

30. Pildal, J., Chan, A. W., Hróbjartsson, A., et al. (2005), Comparison of descriptions of allocation concealment in trial protocols and the published reports: Cohort study, *BMJ*, 330, 1049–1052.
31. Kjaergard, L. L., Nikolova, D., and Gluud, C. (1999), Randomized clinical trials in hepatology: Predictors of quality, *Hepatology*, 30, 1134–1138.
32. Campbell, I. A., Lyons, E., and Prescott, R. J. (1987), Stopping smoking. Do nicotine chewing-gum and postal encouragement add to doctors' advice, *Practitioner*, 231, 114–117.
33. Karlowski, T. R., Chalmers, T. C., Frenkel, L. D., et al. (1975), Ascorbic acid for the common cold. A prophylactic and therapeutic trial, *JAMA*, 231, 1038–1042.
34. Walter, S. D., Awasthi, S., and Jeyseelan, L. (2005), Pre-trial evaluation of the potential for unblinding in drug trials: A prototype example, *Contemp. Clin. Trials*, 26, 459–468.
35. Pocock, S. J. (1996), *Clinical Trials—A Practical Approach*, Wiley, Chichester, UK.
36. Freiman, J. A., Chalmers, T. C., Smith, H., et al. (1978), The importance of beta, the type II error and sample size in the design and interpretation of the randomized control trial. Survey of 71 "negative" trials, *N. Engl. J. Med.*, 299, 690–694.
37. Moher, D., Dulberg, C. S., and Wells, G. A. (1994), Statistical power, sample size, and their reporting in randomized controlled trials, *JAMA*, 272, 122–124.
38. Sung, J. J., Chung, S. C., Lai, C. W., et al. (1993), Octreotide infusion or emergency sclerotherapy for variceal haemorrhage, *Lancet*, 342, 637–641.
39. Altman, D. G., and Bland, J. M. (1995), Absence of evidence is not evidence of absence, *BMJ*, 311, 485.
40. International Conference on Harmonisation Expert Working Group (1997), International conference on harmonisation of technical requirements for registration of pharmaceuticals for human use. ICH harmonised tripartite guideline. Guideline for good clinical practice. 1997 CFR & ICH Guidelines, Barnett International/PAREXEL, Philadelphia.
41. Corrigan, J. D., Harrison-Felix, C., Bogner, J., et al. (2003), Systematic bias in traumatic brain injury outcome studies because of loss to follow-up, *Arch. Phys. Med. Rehabil*, 84, 153–160.
42. Egger, M., Jüni, P., Bartlett, C., et al. (2001), Value of flow diagrams in reports of randomized controlled trials, *JAMA*, 285, 1996–1999.
43. Montori, V. M., and Guyatt, G. H. (2001), Intention-to-treat principle, *CMAJ*, 165, 1339–1341.
44. Millis, S. R. (2003), Emerging standards in statistical practice: Implications for clinical trials in rehabilitation medicine, *Am. J. Phys. Med. Rehabil.*, 82, S32–S37.
45. Gluud, C., and Kjaergard, L. L. (2001), Quality of randomized clinical trials in portal hypertension and other fields of hepatology, in Franchis, R., Ed., *Portal Hypertension III. Proceedings of the Third Baveno International Consensus Workshop on Definitions, Methodology, and Therapeutic Strategies*, Blackwell Science, Oxford.
46. Egger, M., Smith, G. D., and Phillips, A. N. (1997), Meta-analysis: Principles, and procedures, *BMJ*, 315, 1533–1537.
47. Smith, D., and Egger, M. (2000), Meta-analysis. Unresolved issues and future developments, *BMJ*, 316, 221–225.
48. Egger, M., and Smith, G. D. (1997), Meta-analysis. Potentials and promise, *BMJ*, 315, 1371–1374.
49. Als-Nielsen, B., Gluud, L. L., and Gluud, C. (2004), Non-absorbable disaccharides for hepatic encephalopathy: Systematic review of randomized trials, *BMJ*, 328, 1046–1050.

50. Kurstein, P., Gluud, L. L., Willemann, M., et al. (2006), Agreement between reported use of interventions for liver diseases and research evidence in Cochrane systematic reviews, *J. Hepatol.*, 43, 984–989.
51. De Franchis, R. (1996*), Portal Hypertension II Proceesings of the Second Baveno International Consensus Workshop on Definitions, Methodology and Therapeutic Strategies*, Blackwell Science, Oxford.

# *Gynecology Randomized Control Trials*

**Khalid S. Khan,[1] Tara Selman,[1] and Jane Daniels[2]**

[1]*Birmingham Women's Hospital, Birmingham, United Kingdom*
[2]*Clinical Trials Unit and Academic Department of Obstetrics and Gynaecology, University of Birmingham, Birmingham, United Kingdom*

**Contents**

1 Introduction
2 Drug Development Process in Gynecology
  2.1 Phase I: Safety and Pharmacokinetics in Healthy Population
  2.2 Phases II and III: Efficacy and Effectiveness in Target Population
  2.3 Phase IV: Long-Term Safety and Effectiveness
3 Expected Effect Sixe and Sample Sizes of Clinical Trials
4 Avoidance of Systematic Biases
5 Choice of Appropriate Outcome Measures
6 Choice of Appropriate Analysis
7 Multicentered Trials
8 Meta-analysis
9 Conclusion
  References

## 1 INTRODUCTION

Developments in reproductive health are often low profile, but the sheer number of women with gynecological problems means that, if effective interventions exist, a massive overall health and financial benefit can be expected in the population. This chapter focuses on benign gynecology rather than gynecological oncology. Benign gynecological conditions such as chronic pelvic pain, heavy

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

menstrual bleeding, and subfertility may be treated with suboptimal therapies if there is not a commitment to the development of new drugs and a thorough evaluation of the relative merits of existing ones. There remains a burden of disease in women's health that could be alleviated if clinical care was built around evidence-based interventions.

Randomized control trials (RCTs) are widely accepted as a gold standard for scientific evaluation of all treatments. This is as true of gynecology as other specialities, yet there are much fewer trials in this speciality by comparison [1]. In addition to political and financial barriers to drug evaluation, there are methodological problems to be overcome. This chapter will highlight the issues regarding the design and analysis of RCTs that are pertinent to benign gynecology.

## 2 DRUG DEVELOPMENT PROCESS IN GYNECOLOGY

The process of evaluation of new medicinal products will follow the same stepwise escalation of evaluative techniques as within other areas, but there are specific issues to consider at each phase within gynecology.

### 2.1 Phase I: Safety and Pharmacokinetics in Healthy Population

The main difficulty for these trials is to find a "normal" population in gynecology. A large number of premenopausal women are either taking some form of hormonal contraception, so not undergoing normal menstruation, or actively trying to become pregnant, and therefore they would not want to expose themselves or their fetus to the risk of an intervention of unproven safety. Trials limited to postmenopausal women can only comment on the safety and bioavailability of that drug in that limited population.

### 2.2 Phases II and III: Efficacy and Effectiveness in Target Population

Generally, phase II trials recruit a narrow population with the condition of interest, with no comorbidities, and compare the new intervention against placebo. There are ethical considerations regarding placebo where active, albeit not very effective, treatments exist. The acceptability of the route of administration may be of equal importance to its effect, especially if it affects compliance with a drug regime. So whereas oral contraceptives may be as efficacious as injectable alternatives, women may consider the latter preferable and more reliable.

### 2.3 Phase IV: Long-Term Safety and Effectiveness

Interventions for some gynecological conditions, such as endometriosis, may require long-term use without a reduction in effectiveness or tolerability. Long-term follow-up of all trial patients is required to assess overall effectiveness without bias from participant withdrawal. Assessment of safety and teratogenicity may require use of systematic reviews to collate low-frequency events.

## 3 EXPECTED EFFECT SIZE AND SAMPLE SIZES OF CLINICAL TRIALS

In gynecology, it is realistic to expect small to moderate effects, even when compared to a placebo, to be clinically worthwhile. However, because the conditions are often extremely common, for example, menorrhagia or chronic pelvic pain, when aggregated over the population, even a small effect can result in a huge impact on women's quality of life and productivity. Hence trials have to be large to be able to show small differences in effect with sufficient power.

If the process of randomization is strictly performed, the two groups of an RCT should be equivalent, and then it follows that any difference in outcome should be due to chance or genuine treatment effect. This still leaves the difficulty of distinguishing what is due to chance. It is usually very easy to spot a highly successful drug in a common condition with serious outcomes such as myocardial infarction. However, this is not generally the case in gynecology, where the effect size is often moderate or small, especially when looking at pain outcomes and when the condition is chronic or self-limiting. In these cases trials need to be large. Small trials will sometime give nonsignificant results. This may be taken to mean that the treatment does not work, when in fact there were too few participants in the trial to demonstrate a small effect [2]. This is why estimates of a treatment's effectiveness are preferable to simply quoting $p$ values, to highlight the degree of uncertainty around any estimate of effect.

For a given postulated treatment effect, the power of a trial is the probability that a significant result will be obtained in the trial, if the treatment effect is as predicted. Given the expense and immense effort, it becomes unacceptable in gynecology to run a trial that has only a 50, 60, or even 70% chance of spotting a genuine treatment effect. A "power" of 80% is, by convention, the minimum required to be confident of avoiding a falsely negative result and 90% power is preferred. Increasing the power of the trial to detect a difference means increasing the sample size. For example, if the number of women in a particular group with dysmenorrhea who eventually opt for a hysterectomy is 50%, in order to have a 90% chance of detecting whether a new treatment reduces this proportion by half, to 25%, a trial would need to recruit 150 women, whereas at 80% power, it would require 117 women. Unfortunately, intervention in gynecology is seldom that effective.

When anticipating the treatment effect, and hence estimating the sample size, one should consider what the minimum clinically important difference between the two treatments might be, or in other words, the degree of improvement that would lead to a change in the clinician behavior or acceptance by women. For example, if a 25% proportional increase in fertilization rate is considered beneficial and there is a 70% failure rate (or 30% fertilization rate), then to demonstrate this effect with a new therapy, such as low-dose aspirin, would require 256 women to achieve a 53% failure rate, as shown in Table 1. Note that this is not the same as reducing the failure rate by 25 percentage points from 70 to 45%, a proportionally much bigger drop requiring about 132 women. In a high prevalence condition such as dysmenorrhea [3], a reduction from 50% opting for hysterectomy to 45% would still be worthwhile, but would require over 4000 women and hence suddenly a very large trial, requiring an immense effort to complete.

**TABLE 1 Sample Sizes for a Range of Differences between Control and Experimental Arm Rates**

| | Proportional Reduction of Rate[a] | | | | | |
|---|---|---|---|---|---|---|
| | 50% | | 33% | | 25% | |
| Control Arm Rate | Experimental Arm Rate | Sample Size | Experimental Arm Rate | Sample Size | Experimental Arm Rate | Sample Size |
| 0.80 | 0.40 | 46 | 0.53 | 94 | 0.60 | 164 |
| 0.70 | 0.35 | 62 | 0.46 | 132 | 0.53 | 256 |
| 0.60 | 0.30 | 84 | 0.40 | 194 | 0.45 | 346 |
| 0.50 | 0.25 | 116 | 0.33 | 262 | 0.375 | 494 |
| 0.40 | 0.20 | 164 | 0.26 | 390 | 0.30 | 712 |
| 0.30 | 0.15 | 242 | 0.20 | 588 | 0.23 | 1246 |

[a]Assuming 80% power.

**TABLE 2 Sample Sizes for Different Standardized Effect Sizes, Assuming Two Equal Groups**

| | Standardized Effect Size | | |
|---|---|---|---|
| Power | 0.5 | 0.33 | 0.20 |
| 80% | 128 | 292 | 788 |
| 90% | 172 | 388 | 1054 |

When considering continuous outcome measures, the sample size calculation is not based on proportions but on anticipated differences between the mean and the outcome for each group, so a slightly different approach is taken. It is generally accepted that a standardized difference of 0.2 standard deviations can be considered to be a small effect size, 0.5 a medium effect and 0.8 a large effect. As differences between treatments are unlikely to be large, and small to moderate effects can be clinically significant, it is possible to calculate sample sizes for clinical trials based on standardized effect sizes. Table 2 shows the sample sizes for various effect sizes at 80 and 90% power.

If the standard deviation of the outcome measure is known, the difference in means can be predicted too. For example, if the mean length of stay in a hospital following hysterectomy is 5.3 days and standard deviation is 1.3, and the sample size was sufficient to detect a small to moderate effect size of 0.33 standard deviations, a trial of 292 women comparing hysterectomy with uterine artery embolization would be powered to detect increase or reduction in length of stay of at least $1.3 \times 0.33 = 0.43$ days.

One of the alternatives used to avoid very large sample sizes, especially when comparing a new with a standard intervention, is to conduct a "noninferiority" trial. Here the objective is not to demonstrate superiority, which can require large numbers, but to establish that the effect of the experimental intervention in comparison to the standard in not more than some pre-stated small difference, called a noninferiority margin. Hence here clinicians determine the amount of noninferiority they are willing to accept. With the above dysmenorrhea example, if clinicians

were prepared to accept the new intervention does not increase the rate of hysterectomy by more than 5%, then the sample size would be 2460. As the direction of effect being assessed is one sided (the experimental intervention is not inferior to the control intervention), a one-sided hypothesis test is performed. There are limitations with this method: If the experimental group performs better than the control, it cannot be concluded to be superior as the trial was not designed to test this hypothesis and sample sizes are very dependent on the margin at which one accepts noninferiority. So it is important that the noninferiority margin selected is small enough to not exceed that which has clinical relevance, and that the standard intervention is already established as superior to placebo [4].

## 4 AVOIDANCE OF SYSTEMATIC BIASES

In order to detect moderate, but meaningful, differences between treatments, it is important that causes of bias are minimized as much as possible. Biases almost universally tend to exaggerate the true effects of a treatment [5–7]. Figure 1 demonstrates the key biases possible in a badly designed trial. The importance of a sound randomization process to avoid selection bias cannot be overemphasised. This will ensure the comparability of the treatment groups at the start of the trial. An independent telephone service or coded drug containers are appropriate randomization

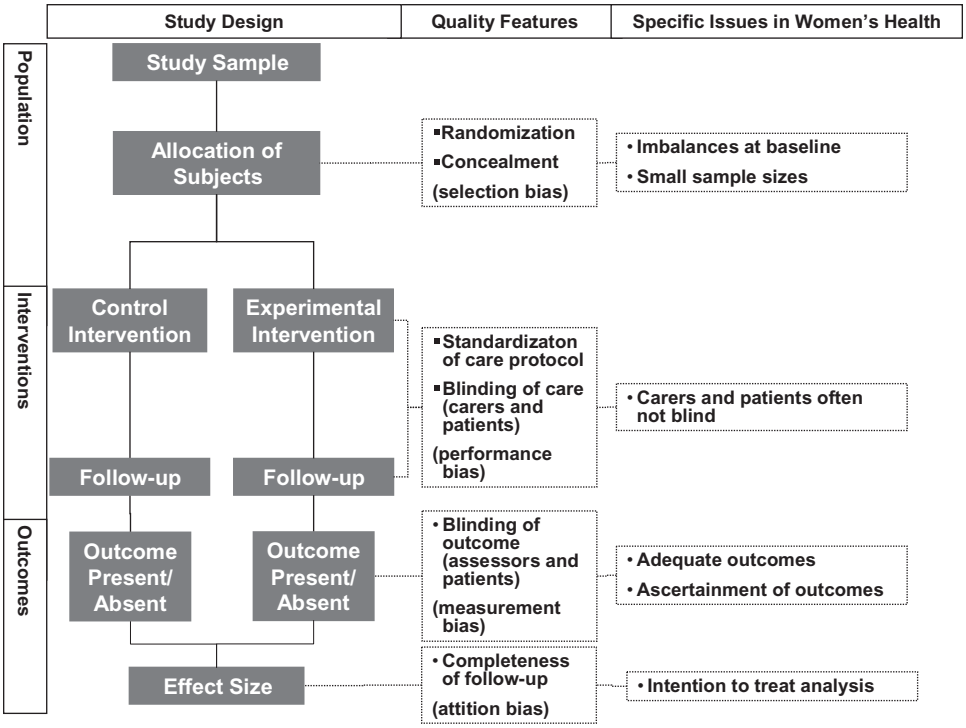


***FIGURE 1*** *Outline of trial, its quality features that minimize the risk of bias and issues specific to trials in benign gynecology.*

methods; tossing a coin, odd and even hospital identification numbers, or allocating participants in order of arrival would not be. The point is not the degree of "randomness" but the extent to which the next allocation cannot be predicted, or in other words, the ability to conceal the allocation from the clinician recruiting the patient. For concealment to be maintained, the randomization process must be indefatigable. For example, if envelopes containing the randomized allocation are used, it is possible for the researchers in this case to manipulate the order of opening of the envelopes, or even resealing opened envelopes that do not contain the preferred allocation, thereby introducing selection bias. For the trial to be concealed, the randomized allocation should be provided by a third party once all eligibility criteria have been confirmed and the participant committed to the trial.

Even the strictest randomization process can be undermined if patients do not receive the treatment to which they are allocated. Postrandomization withdrawals can be minimized by careful screening of potential patients against the eligibility criteria. However, compliance with allocations may diminish if treatment is not instigated soon after randomization. This can be a particular problem in fertility studies, where it is not unusual to randomize women at the start of a cycle, but due to many reasons the embryo transfer procedure may not be carried out in that cycle, and alternative methods are chosen for subsequent cycles [4].

Blinding should not be confused with concealment. Blinding is a method in which trials can attempt to eliminate both performance and measurement biases, by keeping participants and clinicians unaware of the treatment allocation after randomization. Performance bias can occur if there are differences in groups related to co-interventions or supportive care, making it harder to disaggregate the effect of each intervention. Should this potentially be a problem, a standardized care protocol should be agreed upon for all patients and, of course, details of other interventions recorded to detect significant deviations between the groups. Detection or measurement bias can arise if there are differences in the way outcomes are measured or interpreted. Objective measurements are less prone to measurement bias than subjective measurements but may not capture the outcome of interest.

There are different levels of blinding:

- Single blind—Usually when the patient does not know to which treatment arm she has been allocated. Gynecological surgical procedures can be blinded from the woman by use of drapes to obscure her view or sham incisions.
- Double blind—When both clinician and patient do not know which treatment is being given, as in the typical placebo-controlled trial.
- Triple blind—Neither the clinician, patient, or the person performing the outcome assessment are aware of the treatment. This is the most difficult to achieve, but potentially some gynecological assessments, for example, urodynamics, could be blinded to allocation.

Patients should always be analyzed in the group to which they were randomized in an "intention to treat" analysis, as this helps retain the benefits achieved by randomization. If, for any reason, participants move from one arm of the trial into the other, all participants must be analyzed in the group to which they were originally allocated and not in their new group. While intention to treat analysis may underestimate the true treatment effect if it is "diluted" by cross-overs, it will more closely

reflect the clinical reality outside of the trial and provide an anticipated effect of pursing a particular treatment policy.

Attrition bias can arise if there are systematic differences in the degree of follow-up of trial participants. This can arise if a more intensive follow-up for one treatment group is built into the protocol or where there are different rates of participant dropout between the groups. This can be a particular problem in some gynecological conditions such as chronic pelvic pain, where women may decide that the treatment is ineffective and opt out of the trial to seek alternative therapies. Trials where a drug has side effects as troubling as the primary complaint will experience a high degree of noncompliance. This has been observed in a trial of the levonogesterol-releasing intrauterine system when used for heavy menstrual bleeding. The device often has an unpredictable effect on the menstrual cycle in the first 6 months of use that women decide it is preferable to have it removed than wait to see if it has a beneficial effect on their periods.

In these circumstances, every effort should be made to continue to collect follow-up data on all participants, regardless of compliance, again so that an intention to treat analysis can be performed. However, it is inevitable that some women will withdraw their consent to provide data to the trial. One solution is to perform a per-protocol analyses excluding patients with missing data from the analyses. Alternatively, imputation of missing data may be considered. Possible imputation strategies include carrying forward the last observation to the missing time point or estimating the most likely outcome, based on the outcome of other participants in the trial. If patients with missing data are mainly outliers, the precision of the effect size by per-protocol analyses may be increased. But if losses to follow-up are related to prognostic factors, side effects, or lack of response to treatment, per-protocol analyses may overestimate the treatment effects [8]. Reasons for loss to follow-up should be recorded, if possible, to determine the degree of differential loss to follow-up and to establish whether such losses are random or will bias the results.

Unfortunately, following patients through to the prespecified end of the trial is a particular problem with trials of benign gynecology, as participants tend to be otherwise healthy, not have a life-threatening condition and are relatively young and mobile. To avoid, or at least make every attempt to reduce, the number of patients lost to follow-up, multiple identifiers and contact details for participants should be taken at the start of the trial. Follow-up by postal questionnaires direct to the participants is notoriously difficult to sustain and yet is the most appropriate method of collecting quality of life information. A meta-analysis of methods of improving the response rate to postal questionnaires identified a number of useful strategies, such as contact before the questionnaire is sent, provision of a prepaid envelope, and using a short, interesting questionnaire, in addition to obvious monetary incentives—all of these should be considered at the outset of the trial [9].

## 5 CHOICE OF APPROPRIATE OUTCOME MEASURES

Unlike cancer trials, where we are interested primarily in mortality, or obstetric trials where perinatal mortality is a suitable outcome measure, in gynecology the effect of a drug is often on disease-specific symptoms. One has to decide on the most

important outcome measure to use in determining the effectiveness of a drug, that is, whether the treatment is likely to have an effect on that outcome and how the outcome is to be measured.

Ideally, the primary outcome measure should be as important to the women as to the clinician or policy maker. Examples include measuring pain on a visual analog scale for dysmenorrhea [10], assessing sexual function by means of a specific questionnaire [11] in vulvodynia, or counting days of absence from work or usual activities for premenstrual dysphoria. Use of a generic quality of life questionnaire will allow the impact of a treatment to be directly compared to other conditions but tends to focus on physical dimensions of quality of life and may not be sufficiently sensitive to respond to the changes in different aspects of life quality experienced by treating benign gynecological conditions [12]. Disease or symptom-specific quality of life questionnaires capture the more subtle features of a benign condition, particularly those that are important to the patient. The number of available instruments is increasing, but not all exhibit sound psychometric and measurement characteristics; so they should be reviewed for quality before use [13].

Although the main aim of any trial is to determine the effect of a treatment on the primary outcome measure, there are usually other pertinent criteria that are also of interest, such as sexual function or health service resource usage. While the trial sample size is calculated in order to detect a clinically meaningful difference in the primary outcome, the trial may not have sufficient power to detect significant differences in secondary measures. If one looks at too many outcomes simultaneously, it is likely that one will turn out to be statistically significant, due to the play of chance, even in a trial of an ineffective intervention. Thus primary and secondary outcomes should be defined in advance and not chosen once the analysis has been done, on the basis of statistical significance.

Another issue in gynecological research is that the outcome of interest can occur a long time in the future, for example, the ultimate need for hysterectomy. It is often impractical to wait many years to answer such questions, so an alternative is to use a surrogate outcome that predicts the future outcome of interest, for example, using a laboratory marker [14]. Intermediate outcomes are also sometimes proposed as surrogates are for low-frequency events that would otherwise require prohibitively large trials to demonstrate an effect. However, this is not a wholly satisfactory method. If surrogate outcomes are used, it is necessary to ensure that they correlate with clinically relevant measures and they capture the whole of the clinically relevant effect. Hence, ideally, any surrogate should be previously validated. Polycystic ovary syndrome (PCOS) is a common cause of infertility; so the aim of any intervention would be to increase the chance of delivering a healthy baby. Intermediate outcomes are pregnancy and ovulation rates, and there are a host of biochemical and biometric surrogates. Yet the best correlation between surrogate outcomes and the desired clinical endpoint is the correlation between ovulation and pregnancy in women with PCOS taking clomiphene citrate [15] or metformin [16]. However, factors unrelated to PCOS, including male sperm quality and maternal age, will impact on live birth rate, as demonstrated by the meta-analysis of metformin in which there was no evidence of increased clinical pregnancy rate [16].

## 6 CHOICE OF APPROPRIATE ANALYSIS

Many of the common benign gynecological conditions are chronic and lack a definitive "endpoint." Long-term observations of continuous outcomes such as severity of pain or menstrual blood loss are required. Choosing a particular time point for analysis of outcomes may be fairly arbitrary in the course of the condition. Measures that are collected at multiple time points from each trial participant are likely to be more closely related to each other than the variation between different participants. Multilevel modeling takes into account this hierarchy of data and gives the advantage of being able to estimate overall effect over time, utilizing all available information. An overall treatment effect, with confidence intervals can be estimated from the model.

## 7 MULTICENTERED TRIALS

The design, execution, and analysis of a trial might have done all that is possible to eliminate imprecision and biases. The results in this case will be reliable and "true" for the participants studied. But will such results also be generalizable to the wider population? One way to improve generalizability, or external validity, is to recruit a wide and heterogeneous population with the condition of interest, with few exclusion criteria. Another is to recruit from many clinical centers. This approach limits the effect of the peculiarities of single-center studies that make it difficult to replicate results in other settings. Increasing the number of recruitment centers also improves the ability to accrue large numbers rapidly within trials.

## 8 META-ANALYSIS

With the multitude of problems that can arise in gynecological trials, it is imperative that new data should be considered in relation to previous trials and the impact of the results discussed [17]. The best way to achieve this is through the use of meta-analysis to add the results of the trial to those already available. The larger the amount of evidence, the less likely there is to be overemphasis on any particular trial, which can give misleading findings. Investigation of subgroup effects is often not possible from published data meta-analysis due the manner in which studies are reported. Collection of individual patient data from the primary studies to perform a meta-analyses is the most reliable method of assessing the totality of the evidence overall and within subgroups, but this requires considerable effort to organize and goodwill on behalf of the original trial authors [18].

## 9 CONCLUSION

Trials in benign gynecology typically deal with chronic conditions where there is no definite endpoint but an expectation of gradual change on a continuous outcome measure. One exception is infertility studies where the outcome is a successful

pregnancy, but even here intermediate outcomes are often chosen. Most trials require the recruitment of large numbers to evaluate small to moderate effect sizes reliably. Outcome measures need to employ clinically important measures that impact life facets important to patients. In this regard, disease-specific quality of life tools require validation. Data-analytic approaches need to employ intention to treat principles using survival analyses when outcome is time dependent (infertility) or multilevel modeling of repeated measures in chronic conditions such as chronic pelvic pain. In summary, large, multicenter trials with simple entry criteria, robust randomization, consistent execution, and appropriate analysis are required.

## REFERENCES


1. Edwards, A., and Lilford, R. J. (2005), *National Clinical Trials Capacity Review*, NCCRCD, UK.
2. Altman, D. G., and Bland, J. M. (1995), Statistics notes: Absence of evidence is not evidence of absence, *BMJ*, 311(7003), 485.
3. Latthe, P., Latthe, M., Say, L., et al. (2006), WHO systematic review of prevalence of chronic pelvic pain: A neglected reproductive health morbidity, *BMC Public Health*, 6(1), 177.
4. Daya, S. (2006), Methodological issues in infertility research, *Best Practice Res. Clin. Obstetrics Gynecol.*, 20(6), 779–797.
5. Schulz, K. F., Chalmers, I., Hayes, R. J., et al. (1995), Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials, *JAMA*, 273(5), 408–412.
6. Schulz, K. F., and Grimes, D. A. (2002), Allocation concealment in randomized trials: Defending against deciphering, *Lancet*, 359(9306), 614–618.
7. Schulz, K. F. (1995), Subverting randomisation in controlled trials, *JAMA*, 274(18), 1456–1458.
8. Schulz, K. F., Chalmers, I., Hayes, R. J., et al. (1995), Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials, *JAMA*, 273(5), 408–412.
9. Edwards, P., Roberts, I., Clarke, M., et al. (2002), Increasing response rates to postal questionnaires: Systematic review, *BMJ*, 324(7347), 1183.
10. Carlsson, A. M. (1983), Assessment of chronic pain. I. Aspects of the reliability and validity of the visual analogue scale, *Pain*, 16(1), 87–101.
11. Thirlaway, K., Fallowfield, L., and Cuzick, J. (1996), The Sexual Activity Questionnaire: a measure of women's sexual functioning, *Qual. Life Res.*, 5(1), 81–90.
12. Lamping, D. L., Rowe, P., Clarke, A., et al. (1998), Development and validation of the Menorrhagia Outcomes Questionnaire, *Br. J. Obstet. Gynaecol.*, 105(7), 766–779.
13. Clark, T. J., Khan, K. S., Foon, R., et al. (2002), Quality of life instruments in studies of menorrhagia: A systematic review, *Eur. J. Obstet. Gynecol. Reprod. Biol.*, 104, 96–104.
14. Prentice, R. L. (1989), Surrogate endpoints in clinical trials: Definition and operational criteria, *Stat. Med.*, 8(4), 431–440.
15. Imani, B., Eijkemans, M. J., te Velde, E. R., et al. (2002), A nomogram to predict the probability of live birth after clomiphene citrate induction of ovulation in normogonadotropic oligoamenorrheic infertility, *Fertil. Steril.*, 77(1), 91–97.

16. Lord, J. M., Flight, I. H. K., and Norman, R. J. (2003), Insulin-sensitising drugs (metformin, troglitazone, rosiglitazone, pioglitazone, D-chiro-inositol) for polycystic ovary syndrome, *Cochrane Database Syst. Rev.*, 2.
17. Young, C., and Horton, R. (2005), Putting clinical trials in context, *Lancet*, 366(9480) 107–108.
18. Clarke, M. J., and Stewart, L. A. (1994), Systematic reviews: Obtaining data from randomised controlled trials: How much do we need for reliable and informative meta-analyses? *BMJ*, 309(6960), 1007–1010.

# *Special Population Studies (Healthy Patient Studies)*

**Doris K. Weilert**
*Clinical Pharmacology, Quintiles, Inc., Kansas City, Missouri*

**Contents**

1 Introductory Remarks
2 General Considerations for All Special Population Studies
3 Geriatric Population
4 Renal Impairment
  4.1 Design Considerations in RI
  4.2 Special Considerations for Dialysis Patients
5 Hepatic Impairment
6 Women in Clinical Trials
7 Ethnic Considerations
8 Obesity
9 Conclusions
  References

## 1 INTRODUCTORY REMARKS

After a new chemical entity (NCE) has successfully passed the hurdle of the first-in-man study(ies), the pharmacokinetic (PK) and pharmacodynamic (PD) characteristics of the NCE must be fully elucidated to support drug efficacy and drug safety claims for a regulatory submission. Clinical pharmacology/biopharmaceutics components include an understanding of the drug's (1) absorption, distribution, metabolism, and excretion (ADME) profile, (2) behavior at different dosage regimens

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

and/or dosage forms, (3) behavior under diverse or impaired physiological conditions (special populations), (4) potential interactions with coadministered medications or herbal supplements (DDI), and (5) specific PD characteristics that may affect the efficacy and/or safety of the drug. Clinical pharmacology/biopharmaceutics studies are generally conducted in healthy subjects unless the drug is a biological response modifier and cannot be studied in healthy subjects. Exceptions are special population studies which include subjects with physiological characteristics different from the standard healthy volunteer population or typical target patient population. PD studies may be conducted in either healthy subjects or a patient population depending on the PD endpoint of interest.

This chapter focuses on the diversity of physiological effects on NCE PK in people who distinguish themselves from the average adult patient population for which the NCE is being developed. The first section of this chapter provides a definition of criteria which define a special population and general study design considerations common to all special population studies. The subsequent sections address specific objectives, design considerations, statistical analysis, data interpretation, and regulatory requirements for studies conducted in geriatrics, renal impairment (RI), hepatic impairment (HI), women, ethnic, and obese populations.

## 2 GENERAL CONSIDERATIONS FOR ALL SPECIAL POPULATION STUDIES

Special populations encompass a wide variety of people who are physiologically different from the average patient population for which the NCE is being developed and who are frequently not enrolled in sufficient numbers in routine phases I–III clinical trials with a PK sample collection component to allow appropriate characterization of NCE PK and potential risk factors. NCE exposure information in special populations (HI, RI, geriatrics, women, and race) is required as part of the drug registration package. The PK and/or PD data will be used to justify and negotiate the dosage regimen for populations at risk and are a critical component of the drug label. Depending on the PK/PD characteristics of the NCE, dose adjustments may be required in some populations while the use of the NCE may be restricted or contraindicated in others. Potential NCE PK/PD differences have been observed as a result of:

- Physiological progression in life (e.g., pediatric, adolescent, adult, or geriatric populations)
- Lifestyle (e.g., obesity, dietary habits)
- Disease state (e.g., renal or hepatic impairment)
- Gender (males versus females)
- Race (e.g., Japanese populations)
- Genetics/genomics (e.g., poor/extensive metabolizers, differences in receptor composition/affinity resulting in differences in efficacy or adverse-event profiles)

Most formal special population studies are conducted during the later part of phase II development or in parallel with phase III when the therapeutic dose range is

known and the likelihood is high that the NCE will be submitted for registration. Trials such as RI and HI studies are conducted in a multicenter setting, have slow recruitment rates (especially in the severely impaired populations), take a long time to complete, and tend to be expensive. For this reason, they are only conducted to complete the submission package, thus satisfying labeling requirements. Some studies are conducted early(ier) in the development program to either provide data for the subsequent phase II/III program or NCE development in another regions (e.g., Japan).

Dedicated/formal PK studies are routinely conducted in geriatrics, HI, or RI, while data on other special populations may come from integrated population pharmacokinetic (PopPK) analyses of clinical data. For example, formal gender comparison studies have become less common since the U.S. Food and Drug Administration (FDA) 1993 guideline to include women as part of the phase I clinical development and the advent of PopPK approaches [1]. When the nature of the NCE limits drug administration to the intended target population, the regulatory agencies recommend to include subpopulations in the integrated phase II/III development program or to conduct a smaller dedicated "phase I–like" trial in the patient subpopulation.

Study design considerations have many similarities across the various special populations (Table 1). Drug disposition in the special population (sometimes divided further into subpopulations) is compared to that of a control population which has undergone similar study procedures. Standard inclusion and exclusion criteria apply to all special population studies with some modification for elderly, renal, and hepatic impairment populations discussed below. Safety assessments are the same as in other later stage clinical PK study and consist of standard assessments such as physical examination, clinical laboratories (hematology, coagulation, serum chemistry, and urinalysis), vital signs, 12-lead electrocardiograms (ECGs), and adverse events (AEs) plus any safety parameters specific to the NCE. The frequency of safety assessments during the study conduct tends to be held to a minimum unless significantly higher exposure is expected or the drug is known to affect vital signs or cardiac conduction. In special cases, NCE-specific PD markers [e.g., central nervous system (CNS) tests, specific laboratories such as glucose or coagulation assessments] may be evaluated more extensively in order to determine whether PD differences are observed in the population of interest and whether the PD changes are a result of changes in drug disposition.

Most formal special population studies include a control group as one arm of the study design. While special population results can be compared to historical NCE PK data obtained in separate study(ies) of similar design, there is an inherent risk that the regulatory authorities do not agree with the selection of the control study and request other (less favorable) comparisons which then become part of the drug label. The regulatory authorities request control groups to be closely matched to either the special population or the intended patient population. Environmental or behavioral aspects such as diet and smoking habits may also be considered in the selection of the control population since these factors are known to affect NCE PK. It would be difficult to design one "fit-all" control study for the various special population comparisons. Based on the author's experience, inclusion of a control group in each special population study is the preferred approach.

Special population studies are most commonly conducted as single-dose studies assuming the NCE PK is predictive of steady-state PK. Linear PK should be extended

**TABLE 1 Typical Design Features for Special Population Studies**

| Consideration For | Geriatrics >65 years | RI | HI | Gender | Ethnic |
|---|---|---|---|---|---|
| Typical timing of study | Phase II/III | Phase III | Phase III | Phase II/III | As needed in development program |
| General health status | Healthy for age group | Underlying diseases | Underlying diseases | Healthy | Healthy |
| Controls | <65 years | Matched | Matched | Males | Matched |
| Inclusion/exclusion criteria | Reasonably tight | Wide | Wide | Tight | Tight |
| Comedications | Limited | Extensive | Extensive | Only "C/H" | Only "C/H" |
| Confinement | Yes for SD | Yes for SD | Yes for SD | Yes for SD | Yes for SD |
| Meals | Fasted | Fasted (+) | Fasted (+) | Fasted | Fasted/other |
| PK blood sampling | Extended | Extended | Extended | Standard | Standard (+) |
| PK urine sampling | Sometimes | Yes | Sometimes | If key to NCE PK | If key to NCE PK |
| Safety assessments | Minimal (+) | Minimal (+) | Minimal (+) | Minimal | Minimal (+) |
| Dose | Maximum therapeutic dose unless otherwise indicated | | | | |
| Dosing frequency | Single dose (SD) unless NCE or active metabolites have nonlinear PK | | | | |
| PD assessment | If relevant to support NCE safety or to assess population-specific PD changes | | | | |

*Note*:
*Matched*: Demographically matched to special population or indicated patient population.
*C/H*: Oral or depot contraceptives or hormonal replacement therapy; sometimes vitamins, herbal, and limited over counter mediations permitted.
*Fasted*: For −10 hours to 4–5 hours postdose; (+) small snack may be required in more severely ill 1–2 hours prior to dose administration; (other) may receive other than standard meals.
*Extended*: For at least 24–48 hours beyond standard sampling times; (+) may need extension.
*Minimal*: (+) Two to three additional postdose vitals/ECG or more extensive if significant PK differences are expected.

to all analytes measured (parent drug and active/toxic metabolites). The single dose chosen can be the anticipated upper end of the therapeutic dose range, since age or disease state generally affects not $C_{max}$ but instead overall exposure [area under the curve (AUC)]. However, if the NCE has a narrow therapeutic window and/or a significant difference in exposure is expected, a dose in the lower to midrange of the therapeutic window may be more appropriate. For example, due to the lower body weight (WT) in the Japanese population, NCE $C_{max}$ may be different in addition to AUC, which may potentially require a lower dose in this population comparison.

A multiple-dose study may be required if the NCE (and/or metabolite) has either dose or time-dependent PK. Dose selection and dosing frequency should be carefully selected to avoid accumulation of the NCE and/or metabolites to unsafe levels. A dose escalation scheme may be employed to titrate to safe drug levels, or NCE PK parameters may be predicted by modeling the expected effect which the special population is thought to have on drug disposition.

Most of the special population studies, especially those following a single dose of NCE, are conducted under confinement. Subjects are screened and then return to the clinic either the day prior (preferred) or on the morning of the dose administration. Subjects are generally housed during the majority of PK sample

collections and may not be released until the study is completed (last PK sample collected). Healthy (male) volunteers are generally housed in a dormlike setting, while women and elderly tend to have a semiprivate setting. In these healthy populations it is acceptable that the study is conducted under restricted conditions and it is general practice to lock rest rooms (in the case of urine collection), limit access to water fountains, and have controlled access within the facility (e.g., to entertainment rooms). In RI and HI, subjects are managed in a hospital-like setting instead. Considering that many of the subjects are rather ill, every attempt is made to provide as much privacy and freedom (within study restrictions) to keep them comfortable.

General dietary considerations and restrictions are similar to those in any typical phase I study and should reflect the NCE requirements. In most cases, NCE PK is assessed under controlled fasted conditions (e.g., fasted for 10 hours prior to 4–5 hours following dose administration) to avoid introduction of additional variables into the study design. The RI and HI patients may require modifications to the composition of the meal (e.g., salt content) and/or permitted snacks. In a study comparing race (e.g., Japanese versus U.S. population), the content of meals may be purposefully different for the ethnic group of interest to maintain typical dietary habits as part of the factors influencing NCE PK.

The length and frequency of blood (and urine) collections depend on the expected outcome of the study. In geriatric, RI, and HI studies it is worthwhile to extend blood sampling for at least 24–48 hours beyond the sampling scheme employed in routine healthy volunteer studies, even if a minimal effect NCE PK is expected. If the special population is conducted as a multiple-dose study, the PK should be assessed both after single-dose administration and at steady state.

Special population studies are targeted to address potential "extreme" exposure conditions for the NCE which may alter levels of active/toxic metabolites relative to parent concentrations. Metabolites known to be involved in the efficacy and/or safety profile and presenting >10% of circulating parent levels should be considered for analysis. A 10% cutoff is a general guide only, since the decision to evaluate metabolites also depends on relative potency and/or the extent of protein binding of the metabolite compared to parent drug. To reduce costs associated with serial metabolite sampling, collections can be limited to three to four samples at times when the highest metabolite concentrations are expected.

Altered free drug concentrations can change NCE distribution and/or elimination, which may affect safety or efficacy. Blood samples should also be collected to assess protein binding if the NCE (and/or metabolites) are highly bound, the NCE is known to exhibit nonlinear binding, or its binding might be affected by metabolites. In most situations, protein binding is assessed in select samples encompassing the entire concentration range (e.g., around the time of maximum concentration followed by two to three more samples at representative times across the sampling period).

Pooled urine collection is of relevance when significant quantities of parent drug or metabolites are exceeded into urine. If the renal elimination of NCE and metabolites is negligible, urine collections are not necessary with the exception of the RI study. Pooled urine intervals are generally collected for 0–12 and 12–24 hours postdose and in subsequent 24-hour intervals over the entire study period. More

frequent sampling during the initial elimination phase is only recommended for NCEs which are likely not quantifiable over a 12-hour collection interval.

Data are analyzed to describe the PK of the NCE and/or metabolites using either noncompartmental or compartmental methods [2]. The noncompartmental approach is simpler and potentially less subjective. Compartmental modeling approaches are useful if further simulations of the NCE PK are required. Typical PK parameters are time ($t_{max}$) to maximum concentration ($C_{max}$), area under the concentration-versus-time profile (AUC) collected over appropriate intervals, minimum concentration ($C_{min}$) over the dosing interval for multiple dose studies, NCE clearance (CL), volume parameters ($V_d$), apparent terminal elimination rate constant ($k_e$), and half-life ($t_{1/2}$). Following a single dose, the estimation of AUC from time zero to infinity [AUC(0–∞)] should not exceed more than 20% for NCE (and 30% for metabolites) and samples should be collected sufficiently long to adequately characterize $t_{1/2}$, and $k_e$ in the population of interest as well as the control group. For drugs that are highly bound, PK parameters are preferably expressed in terms of unbound concentrations. If urine data are collected in the study, the amount ($A_e$) or fraction of dose ($f_e$) excreted and renal clearance ($CL_R$) are also assessed. Changes in $CL_R$ can be evaluated with regard to its relative impact on total clearance for the population of interest.

The PK parameters are presented descriptively for each treatment. For most studies measures of exposure (AUC and $C_{max}$) are compared as primary PK parameters across categorical population groups using an analysis of variance (ANOVA). The statistical analysis is performed on the log-transformed PK parameters and the model may include other demographic variables if applicable. The 90% confidence interval for the ratio of the least-squares (LS) treatment means is calculated for the various populations. Lack of population difference is established when the calculated confidence interval for the ratio of the LS means falls within the predefined confidence interval limit of the "no-effect" boundary for the study. Statistical methods may also be employed to test for significant differences in other relevant parameters ($t_{max}$, $t_{1/2}$); however, these parameters are generally only summarized descriptively. Where differences in PK parameters can be described as a function of a continuous variable such as creatinine clearance ($CL_{CR}$) in stage II RI studies, a statistical regression model may be more appropriate for data analysis.

Irrespective of the statistical approach, a no-effect boundary should be defined prior to study conduct which is used to assess the significance and clinical implications of any identified PK differences. The criteria should be based on clinical relevant changes in drug PK rather than the narrow statistical window of the 80–125% confidence interval traditionally applied in bioavailability/bioequivalence studies [3, 4]. A wider boundary is acceptable as long as the boundary can be justified with available clinical data. No-effect boundaries can also be defined for the slope of a regression analysis, for a population analysis, or for PK–PD modeling. If "clinically relevant" boundaries are not known prior to study start, the analysis plan should clearly differentiate between predefined narrow bioequivalence boundaries and clinically relevant changes in drug PK. If the 90% confidence interval for the LS mean ratio of the PK measurement falls within the predefined boundary, any changes in drug PK for the special population group can be considered not clinically relevant and do not require any dose adjustments. In the case of RI and HI studies, the small clinical sample size and/or high intersubject variability may preclude meeting tight

statistical criteria and result in statistical interpretation of confidence intervals which are inconclusive.

The data obtained in the special population studies must be evaluated in the context of all available PK, safety, and efficacy information to optimize drug therapy. PopPK analyses assess the contribution of factors to the intersubject variability in modeled NCE volume of distribution and clearance. These factors include demographic covariates [such as age, gender, race, body mass index (BMI) and/or WT, and height], physiological covariates (creatinine clearance, concurrent diseases, etc.), and lifestyle covariates (such as concomitant medications, smoking/drinking habits). The analysis model minimizes PK parameter variability in order of importance (i.e., statistical significance) of the various covariates. Results of PopPK analyses may further support the findings in special population studies, provide mechanistic explanations for apparent NCE PK differences, or alternatively show that statistically significant differences between populations may not have clinical relevance within the overall variability in the population estimates for the PK parameter.

Regulatory expectations have been formulated in various U.S., European, and International Conference on Harmonisation (ICH) guidance documents (see below) and provide detailed information on how the findings of the special population studies should be summarized in the drug label. The special population subsection within the clinical pharmacology section should briefly summarize any findings and describe any dosing adjustments or precautions required in the special population. If clinically relevant changes in drug disposition were observed, a statement should be included in the precautions/warnings section with further reference to the dosage and administration section. The latter should contain detailed instructions for the physician to adjust the NCE dose in the special population or to indicate that the NCE should not be used in the affected population. The reader is referred to the various guidance documents references in the sections below.

## 3 GERIATRIC POPULATION

The geriatric population has been arbitrarily defined as subjects/patients who are 65 years or older. While use of medications generally increases with age, many clinical trials have an upper age cutoff to exclude elderly subjects due to the concern that these subjects (a) are more frail and at potentially higher risk to have treatment-related AEs, (b) have underlying diseases that might affect the NCE PK and thus study objectives, (c) are receiving one or more comedications which are part of the studies' exclusion criteria, (d) have veins that are more difficult to access for blood collection, (e) require different housing conditions than young volunteers, and (f) may not be able to appropriately reason and express a choice when signing an informed consent [5].

As people age, the body undergoes physiological and composition changes [6]. The elderly tend to have lower lean body mass, lower total body water, and higher total body fat, which may affect the volume of distribution and half-life of highly lipophilic drugs. Gastrointestinal function changes with age, gastric pH increases, and adsorptive surface and motility decreases. Unless the NCE has a high first-pass extraction ratio, the absorption of drugs is unlikely to exhibit age-related changes in absorption. Glomerular filtration rate (GFR) and tubular secretion decrease with

age, which may increase half-life for drugs which are predominantly renally excreted [7]. Changes in hepatic metabolism are more complex. Liver mass as well as hepatic blood flow tend to decrease with age, which may affect drugs with low intrinsic clearance as well as those with high extraction ratio; however, there are conflicting reports in the literature whether an actual change in the intrinsic ability of the liver to clear drugs (true intrinsic clearance) exists in the geriatric population or whether the observed changes in metabolism are secondary to an array of related physiological changes [8].

Data in the geriatric population are a regulatory requirement unless the NCE is not indicated in this population. The FDA and ICH guidance documents outline the expectations of the regulatory community [9, 10]. The FDA has further issued a guidance document describing the content and format for geriatric labeling in U.S. submissions [11]. While the FDA focuses on the changes of age-associated conditions with regard to drug disposition characteristics (and that of active/toxic metabolites), the ICH guidances suggest that the elderly should be studied not only with a new NCE but also during the development of new formulations and new combinations of marketed drugs or for new indications which include geriatric patients in the indication. A separate study in the elderly may not be required for drugs with low systemic availability (e.g., some topical drugs) where age differences in PK are unlikely to be of significance.

In a formal elderly PK study, an appropriate number of male and female subjects 65 years of age and older is enrolled along with an equal number of young controls. Generally the number of subjects for each group does not exceed 20; however, the total number of subjects should be large enough to allow statistical comparisons. Since geriatric subjects are generally easy to enroll, a sample size which results in 80% power to detect a clinically significant difference between groups within the specified no-effect boundary is desired. The geriatric population may be further stratified into two to three age subgroups with an equal number enrolled in each age range. Some study designs are also stratified by gender with an equal number of young and elderly male/female subjects enrolled in each age group.

Subjects should be generally healthy within the criteria expected of a geriatric population. Wider inclusion criteria will be required on preadmission body mass index (BMI) values, vital signs, ECGs, and clinical laboratory evaluations to accommodate age-related physiological changes. Elderly are frequently taking multiple medications for various conditions (including hormone replacements, vitamins, and herbal supplements) and the inclusion/exclusion criteria must be specific enough to allow inclusion of a representative geriatric population while ensuring that the PK objectives of the study are not compromised. Geriatric PK studies are generally conducted under standard fasted conditions and meals are withheld for approximately 14 hours. The author has observed that the long fast can lead to a higher incidence of AEs (e.g., dizziness or nausea) in the elderly, who are less tolerant of extended fasting than healthy young subjects.

Pharmacokinetic sampling, safety collections, and data analyses follow the criteria outlined in the previous section. The AUC and $C_{max}$ are compared as primary PK parameters across categorical age groups using an ANOVA. Since age is a continuous variable, a statistical regression model may be employed if a graded age difference in NCE PK is apparent. If the NCE is excreted to a significant extent, it is advisable to collect urine data ($A_e$ or $f_e$ and $CL_R$) and determine $CL_{CR}$ in this

population. In the case of differences in drug disposition, the renal excretion data from the geriatric population can be correlated with the results obtained in the RI population for mechanistic interpretation of the results.

The elderly are more sensitive to CNS agents and may respond differently to cardiovascular agents than younger adults [12]. Phase I studies for CNS or cardiovascular drugs usually include a PD component. PD results can be evaluated using similar statistical methods as for categorical comparison of PK parameters. The objective for the PD component is the determination of whether PD changes occur and, if the case, whether the PD changes are a result of changes in drug disposition or whether the PD changes also occur in the absence of PK changes.

The PK results of definitive geriatric PK studies are frequently supplemented with the modeling results of phase II/III studies. These analyses may substantiate the findings of the definitive elderly PK study or alternatively may identify other factors that may have contributed to the apparent age-related changes in PK (e.g., $CL_{CR}$). At the current time, PopPK results are only considered supplementary in the registration package and have not replaced the formal PK study in the geriatric population.

## 4 RENAL IMPAIRMENT

There are various kinds of kidney diseases. Chronic renal failure, in which the kidney can no longer cope with the load of endogenous and exogenous substances that must be excreted, is the most relevant with regard to affecting the PK of a NCE and/or its circulating metabolites [13]. Patients with RI can be categorized into two distinct subpopulations: those with various degrees of diminished renal function and those who require artificial means to remove the buildup of electrolytes, body waste, and drug product by dialysis methods such as hemodialysis, hemofiltration, hemoperfusion, and peritoneal dialysis. Patients with various degrees of diminished renal function are further categorized into distinct groups based on the severity of their disease. The FDA and European Medicines Agency (EMEA) recommend the categorization of the RI population into five groups, as outlined in Table 2 [14, 15]. The use of categorical renal function groups provides a means of balancing patient enrollment across the RI spectrum and establishes uniformity in data presentation across various drug submissions.

PK data in the RI population are a regulatory requirement unless the NCE is not indicated in this population or a strong case can be made based on disposition

**TABLE 2 Categorization of RI Patients**

| Group | Description | Estimated Creatinine Clearance (mL/min)[a] |
|---|---|---|
| 1 | Normal renal function | >80 mL/min |
| 2 | Mild renal impairment | 50–80 mL/min |
| 3 | Moderate renal impairment | 30–50 mL/min |
| 4 | Severe renal impairment | <30 mL/min |
| 5 | ESRD | Requiring dialysis |

[a]EMEA (2004) further normalizes the renal function groups to body surface area with creatinine clearance expressed in mL/min/1.73 $m^2$ [15].

*Source*: From [14].

and physiochemical properties of the drug that renal disease will not alter drug PK. FDA and EMEA guidelines specify when RI studies are indicated and how the RI data should be utilized in the registration package [14, 15]. If the target patient population is likely to include RI patients and the PK (or PD) of the NCE and/or its active/toxic metabolite is likely to be affected by changes in renal function, studies should be conducted to assess whether dose adjustments are required. RI studies are recommended for NCEs with (a) a narrow therapeutic index and (b) primary elimination of the NCE and/or active/toxic metabolites via the renal route (excretion and metabolism) or (c) a combination of high hepatic clearance and significant protein binding. An RI study is considered unnecessary if an NCE has a wide therapeutic window and (a) is primarily metabolized by the liver, excreted into the bile, or eliminated via the lungs, (b) is a topical drug that is not absorbed, or (c) is intended only for single-dose administration. Confirmation by the agency that a study in RI is not required for drug approval is advised to avoid a potential hold or rejection of a submission until appropriate renal impairment data are available.

The FDA and EMEA distinguish between a traditional approach assessing the NCE PK in patients with varying degrees of renal function (full-study design) and an adaptive two-stage approach (reduced/stage design) [14, 15]. The latter is recommended if the NCE and/or metabolite PK are not likely to require dose adjustments in RI. Stage I consists of a study comparing only the extremes, that is, healthy controls versus patients with severe RI (also see Table 2). If the PK difference between the two extremes does not warrant a dose adjustment, no further studies are needed. Alternatively, if PK in severe RI patients is altered to the extent requiring dose adjustments, patients with a lesser degree of RI should also be studied in a follow-up trial (stage II).

The regulatory agencies recommend RI studies in NCEs with narrow therapeutic margins even if an impact of RI on the NCE PK is unlikely, since RI may affect the disposition of drugs that are exclusively eliminated by nonrenal routes (highly metabolized or excreted into bile). Changes in drug elimination due to changes in protein binding are the most obvious causes of altered PK [16, 17]; however, diseases secondary to RI may also contribute to PK changes. The combination of physiological changes in RI can increase or decrease absorption by altering GI motility, transport mechanisms, and metabolism, blood flow to the kidney and other organs, and hepatic metabolism. Physiological explanations for these changes have been summarized by various authors [18–21]. Other underlying physiological changes secondary to RI include various forms of cardiovascular disorders such as hypertension, coronary artery disease, congestive heart failure, arrhythmias, lipid abnormalities, diabetes mellitus, peripheral neuropathy, and diabetic retinopathy [22–24]. Patients with severe RI (GFR < 30 mL/min) are those most affected by these physiological changes. Low urine production results in a higher accumulation of toxins in the circulation, which puts an additional strain on alternative routes of NCE elimination. Contrary to common belief, patients with end-stage renal disease (ESRD) who complete hemodialysis immediately prior to drug administration can have lower drug exposure for non–renally eliminated drugs compared to the other renal function groups [25]. While the reason for this finding is not well understood, one possible explanation is that the process of dialysis removes a large amount of endogenous material which otherwise would "clog up" the NCE's elimination routes.

Studies of ESRD patients are conducted to determine whether dialysis may result in lower drug concentrations (i.e., patients might require replacement doses) and/or whether dialysis can be used in the treatment of overdose situations. A significant portion of the NCE is likely to be removed during dialysis if an NCE and/or active metabolite is water soluble, has a low nonrenal clearance, a low molecular weight (<500 Da), and low protein binding, and its volume of distribution is small (<1 L/kg) [20, 26]. On the other hand, if the NCE has a large unbound volume of distribution or high nonrenal clearance, dialysis is unlikely to contribute to removal of the NCE to any great extent.

### 4.1 Design Considerations in RI

In the formal renal PK study, an appropriate number of male and female subjects with RI are included in the study along with an equal number of non-RI subjects who match the demographics of either the RI patients or the target population. Ideally, subjects should be known to the investigator long enough to establish the presence of reasonably stable renal function. For example, serum creatinine concentrations should not vary by more than approximately 0.4 mg/dL over a 3-month period prior to the initial screening. Other means of documenting stable serum creatinine concentration or stable creatinine clearance up to 6 months prior to the initial screening can be used as long as they are well specified. Since RI patients are harder to recruit due to restrictions in inclusion/exclusion criteria, the author recommends widening the age range to at least 70 years of age rather than the typical 65-year cutoff. The number of RI patients should be sufficiently large to allow appropriate statistical analyses, but the time it will take to enroll patients should also be considered.

In stage I, an equal number of patients with severe RI are enrolled along with an equal number of healthy controls. In the extended design (stage II), the different RI function groups generally contain a minimum of 8–10 subjects per group and one control group of similar size and demographic characteristics which brackets all RI groups. Patient enrollment should be closely monitored so that patients within each renal function group are evenly distributed (e.g., not all patients with mild renal impairment have a creatinine clearance $CL_{CR}$ of >65 mL/min). This becomes important at the time of data analysis if PK parameters are regressed on $CL_{CR}$.

Creatinine clearance can be directly measured using serum and urine creatinine concentration or calculated by one of several formulas using serum creatinine concentrations only. If $CL_{CR}$ is directly measured using both serum creatinine and urine creatinine, it is advised to collect urine for 24 hours and to draw the serum creatinine sample at the midpoint of the urine collection interval. In ESRD patients who produce urine, serum creatinine concentrations should be obtained at the beginning and end of the urine collection. The average of the two serum creatinine values provides a more accurate estimate since serum creatinine is likely to increase over time between dialyses. If $CL_{CR}$ is measured using serum creatinine only, the FDA recommends using the Cockcroft–Gault formula in adult subjects and the formulas by Schwartz in children and infants from birth to one year of age (see Table 3) [27–29].

Inclusion and exclusion criteria should clearly define which underlying diseases are acceptable. Patients with any major illness (other than renal disease

**TABLE 3 Formulas to Calculate Creatinine Clearance from Serum Creatinine Concentrations**

| Group | Formula | Reference |
|---|---|---|
| Adults (>12 years) | $CL_{CR} = \frac{(140 - \text{age}) \times \text{weight}}{72 \times \text{Serum}_{CR}}$ (0.85 for females) | 27 |
| Infants less than 1 year | $CL_{CR} = \frac{0.45 \times \text{length}}{\text{Serum}_{CR}}$ | 29 |
| Children 1–12 years[a] | $CL_{CR} = \frac{0.55 \times \text{length}}{\text{Serum}_{CR}}$ | 28 |

*Note*: $\text{Serum}_{CR}$: serum creatinine in mg/dL; age in years; weight in kg; length in cm; $CL_{CR}$ in mL/min.
[a]Schwartz initially presented this equation with plasma creatinine for children up to 20 years of age [28].

and associated stable illnesses such as hypertension and diabetes mellitus) such as hepatic, GI, cardiovascular, adrenal, hematological, or obstructive airway disease should be excluded. Patients with mild hepatic and GI diseases should be evaluated carefully to determine if these might affect the pharmacokinetics of the NCE. All underlying diseases should not interfere with the study conduct or pose an unjustifiable risk to the study participant. Wider inclusion and exclusion criteria are required on preadmission vital signs, ECGs, and clinical laboratory evaluations to reflect the nature of the patients to be enrolled. Restrictive inclusion/exclusion criteria will affect the enrollment rate.

Subjects with RI will require concomitant medications. Patients whose medication is likely to interfere with the NCE PK should be excluded. Subjects receiving medications that are unlikely to interfere with the study medication (or vise versa) should receive their regular scheduled medication at a time different from the scheduled NCE dose whenever possible (e.g., withhold all nonessential medications ≥12 hours prior to and ≥4 hours following NCE administration). Patients receiving ranitidine and other $H_2$-receptor antagonists may be asked to withhold their medications for ≥36 hours prior to dosing, if medically acceptable, and to use alternate measures (e.g., antacids) to control their symptoms. Antacids should be held for at least 1 hour prior to dosing and 2 hours postdose. For medication that must be administered near the time of the NCE administration, dosing should be scheduled either 1–2 hours prior to or 1–2 hours after administration of the study medication whenever possible. Some over-the-counter herbal supplements are associated with drug interactions; others may affect physiological parameters such as blood pressure, blood glucose, kidney function, and electrolyte balance in renal disease patients [30]. Patients requiring herbal supplements for medicinal purposes should be carefully evaluated prior to study enrollment.

The timing of meals in renal studies is similar to those in any typical phase I study and should reflect the NCE requirements. With progression of the disease, renal patients may require modifications to the composition of the meal (e.g., salt content) and/or permitted snacks. Patients with severe renal disease and with ESRD may have restrictions with regard to the volume of fluid intake, which needs to be taken into account in the design of the study.

Pharmacokinetic sample and safety collections follow the criteria outlined in Section 2 with the exception that urine samples are almost always collected in RI

studies. The FDA guidance recommends that unbound drug levels are determined in each sample unless the drug exhibits relatively low extent of binding (defined by the FDA as the extent of binding being less than 80%) and that PK parameters are calculated based on unbound concentrations [14]. In the stage I design, the primary exposure PK parameters ($C_{max}$ and AUC) are compared between RI patients and the control group using an ANOVA model with the appropriate confidence interval boundaries. Secondary parameters, including protein binding results, are summarized descriptively. In the stage II design, the primary PK parameters are most commonly regressed against the patients' $CL_{CR}$.

Since screening $CL_{CR}$ values may have been obtained three to four weeks prior to dose administration, it is advised that $CL_{CR}$ is obtained again on the day of pharmacokinetic sampling to avoid potential misclassification of the patient. In RI studies that the author has conducted, the "during study conduct" $CL_{CR}$ estimates generally correlate well with those obtained during screening, but there may be changes in some patients in either direction of renal function. In some cases, patients "categorized" into a renal function group during screening fell into the adjacent renal function group during study conduct (e.g., a patient with $CL_{CR}$ of 27 mL/min at screening had $CL_{CR}$ of 33 mL/min during the actual study). If regression analysis is used to model the PK parameters to renal function, these small shifts do not affect the interpretation of the data, while categorical ANOVA analysis may be affected by shifting one subject from one category to another. The other PK parameters are generally presented descriptively.

Serum creatinine values are generally collected as part of the clinical laboratory package in phase II/III. The $CL_{CR}$ can be calculated for the majority of patients and is available as a potential covariate for PopPK analysis. However, most phase II/III studies do not include a large enough population with decreased RI function and the $CL_{CR}$ range tends to be too narrow to allow determination of meaningful differences. For this reason, the data generated in the definitive phase I study are not always supplemented by PopPK analyses. One of the exceptions would be the situation where the NCE cannot be studied outside the therapeutic target population. The regulatory agencies have requested that an attempt is made to include subpopulations in the integrated phase II/III development program and to utilize PopPK analysis approaches to elucidate the relationship between relevant PK parameters and renal function. However, the number of patients must be large (and diverse) enough to allow detection of PK changes requiring dose adjustments.

### 4.2 Special Considerations for Dialysis Patients

The most common forms of renal replacement therapy in patients with ESRD are hemodialysis and continuous ambulatory peritoneal dialysis (CAPD). Both processes involve the principle of diffusion; however, hemodialysis is considered the most efficient artificial means of toxin and drug removal. During hemodialysis, blood is dialyzed against a dialysate medium across a permeable membrane. The extent of toxin and drug removal during hemodialysis is governed by the rate of blood flow through the dialyzer, the rate of dialysis flow, the surface area and type of dialysis membrane, and the dialysis medium [13]. In CAPD, the peritoneal membrane serves as the dialysis membrane to the dialysate in the peritoneal cavity. Toxin removal in

CAPD occurs through osmosis and diffusion from the small blood vessels in the walls of the abdominal cavity, the viscera, and the interstitium through the peritoneal membrane. For more information on CAPD, the reader is referred to reviews of advantages and disadvantages of CAPD and its influence on drug pharmacokinetics [31, 32]. Studies in CAPD patients are only recommended for the NCEs likely to be administered in this patient population.

Renal dialysis studies are conducted both "between" and "during" the scheduled dialyses regimen. The decision of what time in the dialysis cycle the PK information is collected depends on the objectives of the study. Determination of potential accumulation of the NCE and/or metabolites in the ESRD population will include patients between their scheduled dialysis while the use of hemodialysis as a means of drug removal would require drug administration during the patients' hemodialysis regimen. The length of PK sampling is restricted between dialysis since ESRD patients undergo hemodialysis every 48–72 hours. PK sampling between dialysis regimens is rarely longer than 48 hours, which may introduce error into results for half-life and AUC assessments. Potential shorter PK sample collection should be taken into account when NCE PK data are compared across studies. It is recommended that drug administration be scheduled at least 1 hour after completion of the dialysis before dosing to allow for balance of the patient's fluid level. PK sample collections, data analysis, and statistical comparisons for between-hemodialysis PK studies are performed as previously described and include categorical comparison of the primary exposure PK parameters to a control.

When assessing the NCE PK during hemodialysis, the timing of dose administration and potential fluid loss during analysis should be considered. Hemodialysis sessions generally last around 3–4 hours. NCE administration most often occurs prior to the start of the dialysis. Optimal results, especially if the dialysate is measured, will be obtained if the dialysis is started around the time of maximum exposure, since potential drug removal is proportionate to blood concentrations. Hemodialysis will remove accumulated water along with waste material. Determination of the difference in patient weight immediately before and following hemodialysis accounts for this fluid loss/weight loss and can be used as covariate or correction factor during data analysis and interpretation.

Collection of data regarding relevant equipment and conditions during hemodialysis is important since the extraction efficiency of the hemodialysis is affected by the type of dialyzer, dialysis membrane, and membrane surface area against which the blood is dialyzed. The ionic strength and type of the dialysate (dialysate potassium, dialysate calcium) and blood and dialysate flow rate will also affect the extraction ratio. Dialysis conditions are optimized to a given ESRD patient and are likely different between study subjects and study sites. The subject's case report form should allow for documentation of all pertinent dialysis components.

The PK sample collections, data analysis, and statistical comparisons for during-hemodialysis PK studies include categorical comparison of the primary exposure PK parameters to a control using an ANOVA model with the appropriate confidence interval boundaries. However, AUC comparisons frequently focus on the partial AUC interval obtained during the hemodialysis since the dialysis window is so narrow. In addition, these studies typically include the estimate of dialysis clearance. Lee and Marbury have described a variety of methods to determine dialysis clearance [33]. Calculations may be based solely on blood/plasma concentrations or

can include dialysate concentration. Blood samples are collected during three stages of the study: (a) prior to start of dialysis, (b) during dialysis, and (c) during postdialysis treatment. During dialysis, blood samples can be removed directly from the lines entering (prefilter or arterial sample) and leaving (postfilter or venous sample) the dialysis chamber, while serial venous blood samples are collected prior to and following dialysis. The following information should be recorded in the subject's case record form to allow determination of dialysis clearance: (a) blood flow into the dialyzer, (b) hematocrit in the pre- and postfilter blood samples during dialysis to allow for conversion of blood dialysis clearance to plasma dialysis clearance, and (c) optional endogenous molecules such as blood urea nitrogen (BUN) and/or creatinine to assess the extraction efficiency of the hemodialyser (based on standard markers) and to permit comparison to the NCE and/or metabolites.

If the objective of the study is to determine percentage of dose removed during dialysis, collection of dialysate is required. Collection of dialysate may allow for a more accurate estimate of dialysis clearance than can be obtained from blood/plasma data [33]. It should be noted that dialysate can be a difficult matrix to assay for drug concentrations. Analytical method validation can be difficult to develop due to the high salt content of the dialysis medium. Dialysate flow rate should be collected as part of the study information. Although dialysis flow rate is specified as part of the dialysis procedures, measurement of the total dialysate volume is also recommended since it provides confirmation of the reported dialysis flow setting. Incorrectly recorded flow rates or variation in outputs of the dialyser can be easily detected if the preset flow did not deliver the total intended volume and PK calculations may be performed with an "adjusted" or "corrected" flow rate based on measured dialysate volume.

For the detailed derivation of dialysis clearance the reader is referred to the article by Lee and Marbury [33]. The authors discuss different direct methods such as arteriovenus (AV) difference, recovery rate, or cumulative dialysate methods as well as those derived from plasma pharmacokinetic parameters obtained between and during dialysis. If both plasma and dialysate data are available, data calculated by more than one approach can be compared.

In simplified form, the AV difference approach is based on a flow rate multiplied by the extraction efficiency of the hemodialyser (E). Blood dialysis clearance ($CL_{HD\text{-}b}$) is thus calculated as

$$CL_{HD\text{-}b} = \frac{Q_{bi}[C_{bi} - C_{bo}]}{C_{bi}} \tag{1}$$

$$CL_{HD\text{-}b} = \frac{Q_{do}C_{do}}{C_{bi}} \tag{2}$$

where $Q_{bi}$ is are the blood flow entering and $C_{bi}$ and $C_{bo}$ the blood concentrations entering (i) and leaving (o) the dialysis chamber. The quotient term $(C_{bi} - C_{bo}]/C_{bi}$ is the extraction ratio based on blood/plasma concentration alone. Here $Q_{do}$ is the dialysis flow leaving and $C_{do}$ is the dialysate concentration leaving (o) the dialysis chamber. The assumption is made that ultrafiltration differential and backdiffusion are negligible compared to the blood and dialysis flows.

Since blood flow rather than plasma flow is measured as part of the study and the drug concentrations are derived from plasma samples, $CL_{HD\text{-}b}$ needs to be converted to plasma dialysis clearance ($CL_{HD\text{-}p}$) using

$$CL_{HD\text{-}p} = Q_b \{[K_p * \text{Hct} + (1 - \text{Hct})]\} E \tag{3}$$

where $E = (C_{pi} - C_{po})/C_{pi}$ and $C_{pi}$ and $C_{po}$ are the plasma concentrations of the blood entering (i) and leaving (o) the dialysis chamber; $K_p$ is the partition coefficient between RBCs and plasma and Hct is the hematocrit. If both the hematocrit and the partition coefficient of the NCE between RBCs and plasma are known, the assessment of $CL_{HD\text{-}p}$ using the AV model is most accurate. If the $K_p$ approaches zero (i.e., the analyte does not distribute into RBCs), the equation is reduced to reflect the physiological plasma flow:

$$CL_{HD\text{-}p} = Q_b (1 - \text{Hct}) E \tag{4}$$

Alternatively, assuming that $K_p = 1$, Equation (4) reduces to

$$CL_{HD\text{-}p} = Q_b E \tag{5}$$

Equation (5) is the most common equation used to assess $CL_{HD\text{-}p}$ from plasma data in the absence of any other information. As an approximation, $Q_b$ in Equation (5) can be corrected by the blood-to-plasma NCE concentration ratio ($C_b/C_p$) to estimate plasma blood flow and $CL_{HD\text{-}p}$.

If dialysate data are available, $CL_{HD\text{-}p}$ can be calculated by two basic methods: the recovery rate method and the cumulative dialysate method. In both methods, the assumption is made that the dialysis flow stays constant over the interval of observation. The percent recovery is calculated as the fraction of drug recovered in the dialysate relative to the administered dose. These methods are identical to those employed for the calculations of renal clearance and percent of dose excreted into urine. For this reason, the equations are not presented in this section. Since ESRD patients undergo hemodialysis only every two to three days and the NCE is likely administered more frequently, it might be of interest to assess the percent dialyzed value both for the "actual dose" studied during hemodialysis and for the likely "total dose" (sum of all the doses a patient would have received during one hemodialysis cycle) to determine if dose adjustment in ESRD patients is needed.

## 5 HEPATIC IMPAIRMENT

The liver is the main metabolizing organ in the human body. Hepatic function exhibits a slow decline with increasing age (decrease in liver weight and hepatic blood flow) [8]; however, these decreases are generally not considered of clinical relevance with respect to drug disposition. One the other hand, a wide etiology of pathophysiological mechanisms is known to cause a decrease in liver function and may alter the PK characteristics of drugs. HI tends to be a heterogeneous disorder due to both structural and functional abnormalities. In the western world including the United States, the most common cause of HI is chronic alcohol abuse leading

to liver cirrhosis. On a worldwide basis, viral infections (hepatitis B and C) are the main contributors to HI. Liver disease may also develop secondary to other diseases such as congestive heart failure or cancer and may not always be immediately recognizable.

In contrast to RI, there are no good endogenous markers which allow characterization of hepatic function (such as bilirubin, albumin, or prothrombin time), nor do model drugs which are exclusively cleared from the liver lend themselves as exogenous markers [34, 35]. While various clinical variables such as ascites, encephalopathy, nutritional status, peripheral edema, and histological liver assessments have been studied, none of the various clinical approaches has shown similar predictability of liver function as the use for serum creatinine (or $CL_{CR}$) for RI. For pragmatic reasons, the Child–Pugh classification for alcoholic cirrhosis and portal hypertension has been adapted by both U.S. and European regulatory agencies as a means to allocate HI patients to hepatic function groups [36–37]. The Child–Pugh scale is based on the encephalopathy grade, degree of ascites, serum bilirubin and albumin concentrations, and prolongation in prothrombin time, as illustrated in Table 4. Based on the total number of points for each of the grading criteria, subjects are classified into mild (Child–Pugh grade A: 5–6 points), moderate (Child–Pugh grade B: 7–9 points), or severe (Child–Pugh grade B: 10–15 points) HI groups. A normal control group would have a grading of less than 5.

It should be understood that the Child–Pugh scale was not originally developed to categorize HI patients for the purpose of predicting drug PK. The FDA guidance provides alternative classification scales such as the Maddrey–Carithers discriminant function and Mayo risk scores for primary biliary cirrhosis as alterative clinical measures of liver function, but with the implementation of the FDA and EMEA guidelines, the Child–Pugh scale has become the classification system that is most commonly used [38, 39].

The PK data in the hepatic impairment population are a regulatory requirement unless the NCE is not indicated in this population or a strong case can be made based on disposition of the NCE that HI will not alter drug PK. The FDA and EMEA guidelines specify when HI studies are indicated and how the HI data should be utilized in the registration package [38, 39]. HI studies are indicated for almost all drugs likely to be used in subjects with HI. Drugs with >20% of hepatic

**TABLE 4 Categorization of HI Patients**

| Points | 1 | 2 | 3 |
|---|---|---|---|
| Encephalopathy grade | None (0) | 1–2 | 3–4 |
| Ascites | Absent | Slight | Moderate |
| Bilirubin, mg/dL | <2 | 2–3 | >3 |
| Albumin, g/dL | >3.5 | 2.8–3.5 | <2.8 |
| Prolongation in prothrombin time, sec | <4 | 4–6 | >6 |

*Note*: Encephalopathy grade:
Grade 0: normal consciousness, personality, neurological examination, and electroencephalogram (EEG)
Grade 1: restless, disturbed sleep, irritable or agitated, tremors, impaired handwriting, 5 cps waves on EEG
Grade 2: lethargic, time disoriented, inappropriate, asterixis, ataxia, slow triphasic waves on EEG
Grade 3: somnolent, stuporous, place disoriented, hyperactive reflexes, rigidity, slower waves on EEG
Grade 4: unrousable coma, no personality/behavior, decerebrate, slow 2–3 cps delta waves on EEG

*Source*: From [38, 39].

metabolism or elimination (of parent drug and active/toxic metabolites) are considered to be extensively metabolized and a change in exposure may be clinically significant and thus require dose adjustment or contraindication in this population. If the extent of metabolism or biliary excretion is unknown, the agency assumes the drug to be highly metabolized, and hence a HI study is required. If the drug has a narrow therapeutic window, the HI study is recommended even if <20% of the drug and active/toxic metabolites are eliminated by the liver.

The guidelines consider only a few situations where a HI study may not be required (unless clinical concerns suggest otherwise): the NCE (a) is not to be used in patients with HI or (b) is intended for single-dose use only or (c) has a wide therapeutic window and <20% of parent and/or active/toxic metabolites are eliminated by the liver or (d) is gaseous or volatile and the drug and its active/toxic metabolites are primarily eliminated via the lung.

As in the case of RI, HI is associated with secondary physiological changes or disease states that may affect the NCE PK in this population. For this reason, it is prudent to obtain confirmation by the agency that a study in hepatic impairment is not required for drug approval to avoid a hold or rejection of a submission until appropriate hepatic impairment data are available.

The FDA and EMEA distinguish between two approaches in the design of HI studies [38, 39]. The extended approach assesses the NCE PK in patients with varying degrees of hepatic function compared to a control group. The reduced design only includes subjects with a Child–Pugh category of moderate which are compared to a matched control group. The latter is recommended if the NCE and/or metabolite PK are not likely to require dose adjustments in HI and subjects with a Child–Pugh category of mild are unlikely to provide a signal (e.g., clinically significant change in exposure). Under the reduced scenario, the findings in the moderate Child–Pugh category would be applied to those of the mild category, and dosing in the severe category would likely be contraindicated in the label. This approach has the disadvantage that significant findings in the moderate HI population would be extrapolated to the mild category and could result in contraindication of the drug in any subjects with HI. Under those circumstances, a second study in patients with mild HI may be required if this population is part of the target patient population, which could potentially delay the registration package.

The alternative is the full approach described by the agencies in which at least six subjects with HI are enrolled into each of the three Child–Pugh categories outlined in Table 4 along with a control population of similar or greater size. Six evaluable subjects per treatment arm appear to be a low number; however, subjects with HI are notoriously hard to enroll, especially if they fall into the moderate and severe HI categories. The number of subjects is therefore a compromise between what is feasible and what is minimally needed to obtain data likely to produce meaningful results. Since these studies are not intended to prove statistical equivalence between HI and healthy controls, the sample size in these trials are large enough to provide a signal for a clinically relevant change in NCE exposure. Control subjects should demographically match those in the HI population. In the reduced design, control subjects can be matched one to one, while in the full design their age and gender are matched across the demographic range of the three HI groups. For details with regard to the selection of HI subjects, inclusion/exclusion criteria, and study design the reader is referred to the review by Bonate and Russell [40]. Subjects should be

known to the investigator long enough to establish the presence of reasonably stable disease. They should not have clinically significant changes in their disease status over at least a 3-month period prior to the initial screening. Examples of clinical changes would be worsening clinical signs of hepatic impairment or worsening of total bilirubin or prothrombin time by more than 50%. If a previous diagnosis classified the hepatic dysfunction a result of cirrhosis, copies of documentation of any clinical information used to make the previous diagnosis should be available. Documentation may include clinical evidence of portal hypertension, a positive liver biopsy, hepatic ultrasound, computed tomography scan, and/or magnetic resonance imaging.

Inclusion and exclusion criteria should clearly define which underlying diseases are acceptable. Patients with any major illness (other than liver disease and associated stable illnesses) such as renal (e.g., $CL_{CR} < 50$ mL/min), GI, cardiovascular, adrenal, hematological, or obstructive airway disease should be excluded. Patients with mild renal and GI diseases should be evaluated carefully to determine if these might affect the pharmacokinetics of the NCE. All underlying diseases should not interfere with the study conduct or pose an unjustifiable risk to the study participant. As in the case of RI, subjects with HI tend to require concomitant medications and similar study considerations apply.

The HI study is generally conducted as a single-dose study and the highest proposed therapeutic dose is selected unless safety concerns warrant a lower dose. In the full-study design, treatment administration can be structured to study the NCE PK and tolerability sequentially. Subjects with mild HI are dosed first. Once tolerability has been established, dose administration proceeds to the moderate HI group and then to the severe group. The study could be stopped after completion of any HI category. This design, while safer, will increase the duration of the study since moderate and severe HI populations are increasingly harder to enroll and enrollment of these populations cannot start until the previous category has completed the study.

The PK sample and safety collections as well as data analysis follow the criteria outlined in Section 2. The FDA guidance recommends that protein binding is performed for drugs that are highly extracted by the liver (extraction ratio > 0.7) and are extensively bound to plasma proteins (fraction unbound <10%). Unbound fraction should be determined at least at trough and maximum plasma concentration and clearance and volume parameters are appropriately expressed in terms of both unbound and total concentrations [38]. In both the extended and reduced study, the $C_{max}$ and AUC are compared across categorical population groups using an ANOVA with the appropriate confidence interval boundaries. Secondary parameters including protein binding results are summarized descriptively.

The results of HI impairment studies are frequently difficult to interpret. Due to the small sample size, the study has low power resulting in wide 90% confidence intervals. As a result, the 90% confidence interval for the LS mean ratio of the PK measurement rarely falls within the no-effect boundary or the even more restrictive 80–125% if a no-effect boundary is not available. If the criteria are not met, it has to be concluded that HI has an effect on NCE PK. It is not always clear if the effect is clinically relevant. The FDA guidance states that dose adjustments should be reflected in the label in the case of an obvious effect, that is, in the case that the clearance of the NCE is significantly impaired. The definitions of obvious and

clinical significance need to be put in context with PK and safety data. Exposure–response relationship analysis with emphasis or the upper safety limits of exposure obtained from both phase II/III trials (representative of typical patient exposure) and phase I trials (high-exposure comparison) may be explored to defend the case for the labeling recommendations made to the regulatory agencies. NCI PK metabolism (enzymatic pathways, percent of drug eliminated by the liver, and disposition of metabolites by the liver) and the overall sensitivity of drug disposition to other physiological and external changes will also play a role in the decision if and how dose adjustments in the HI population are proposed. In spite of the difficulties in conducting and interpreting data of HI studies, these investigations are critical to the registration package to support dosing decisions in this population.

## 6 WOMEN IN CLINICAL TRIALS

Inclusion of women in clinical trials has drastically changed over the last 15 years. Up to the early 1990s, women were routinely excluded from phase I clinical trials using the argument that the benefit–risk ratio was too high to allow inclusion of women into all types of phase I trials, especially prior to completion of reproductive toxicity studies which, in the United States, are not required until women of childbearing potential are enrolled into phase III studies [41]. The definitive gender comparison was generally conducted in the later stages of drug development, unless the NCE characteristics required the data to support phase II or III clinical trials.

The FDA Guideline for the Study and Evaluation of Gender Differences in Clinical Evaluation of Drugs reversed the previous FDA policy which excluded women with childbearing potential from early drug studies unless it targets a serious or life-threatening disease [1]. The 1993 guideline requires women to be included in all phases of the clinical drug development program for agents that are indicated for women. The guideline states that the FDA can place an investigational new drug (IND) or specific studies on hold if a sponsor proposes to exclude women or men with reproductive potential because of a risk, or potential risk, or reproductive toxicity from the use of an investigational drug product. In 1998, the FDAMA Women and Minority Working Group Report was issued which reviewed the 1993 guideline and recommended not to generate further gender-related guidelines but to institute a permanent tracking system to gather, search, and evaluate demographic data and potential deficiencies with regard to the inclusion of women in clinical trials [42]. The ICH has a website dedicated to ICH and women which shares the results of a recent review of the existing ICH guidelines and regulatory experiences in the three ICH regions: the United States, European Union (EU), and Japan [43]. In the posting, the ICH concluded that the current ICH and regional guidelines should be consulted for guidance on demographic considerations and that there was no need for a separate ICH guideline on women as a special population in clinical trials. The U.S., EU, and Japanese surveys showed that, in general, women were adequately presented in pivotal trial populations and that some form of gender effect was generally conducted and expected, be it a subpopulation analysis and/or formal PK/PD studies. The EMEA published the ICH findings as part of the ICH Guidelines Step 5 [44].

In today's environment, women (both those of childbearing and non-childbearing potential) are routinely included in phase I clinical trials unless specific NCE

characteristics preclude women from participating. Inclusion/exclusion criteria are worded such that women must adhere to stringent criteria in order to avoid pregnancy over the duration of the study. Women who are pregnant or lactating are always excluded unless the study is specifically conducted in one of these populations. A negative (serum or urine) pregnancy test result prior to enrollment and on check-in prior to drug administration is required for women of childbearing potential or who are perimenopausal. For women of childbearing potential (including perimenopausal women who have had a menstrual period within one year) appropriate birth control [defined as a method which results in a low failure rate, i.e., less than 1% per year, when used consistently and correctly, such as implants, injectables, some intrauterine contraceptive devices (IUDs)], sexual abstinence, or a vasectomized partner during the entire duration of the study are stipulated. It should be noted that many clinical protocols no longer restrict birth control methods to females and also require a barrier contraceptive method for males.

The majority of phase I trials include women who are taking oral contraceptive medications provided that they do not change their regimens throughout the course of the study. Similarly, postmenopausal women are permitted to continue their hormone replacement therapy throughout the study duration. The important point is that the medication is not changed during the study conduct to avoid potential changes in NCE PK across treatments. Exceptions to permitting systemic birth control medication or hormone replacement therapy are the cases where (a) the NCE is known to affect efficacy/safety of the medication (e.g., NCE induces liver enzymes), (b) the NCE is evaluated in a drug–drug interaction study in which the other administered drugs may affect efficacy/safety of the medication, or (c) the contraception/hormonal replacement medication may interfere with the objectives of the study.

Since PK data in women are generated throughout the entire drug development program, PopPK analyses of phases I–III PK data have become accepted not only in the United States but also the EU and Japan and have essentially replaced formal PK studies as the venue to assess potential effects of gender on drug disposition. If an NCE requires a formal PK study in women, the number of male and female subjects enrolled is based on the intrasubject variability of the NCE and is sufficiently large to perform the statistical comparison of the primary PK endpoints. Inclusion/exclusion criteria are described above. PK sampling, analysis, and statistical approaches reflect those described in Section 2.

## 7 ETHNIC CONSIDERATIONS

When an NCE is introduced into a new region, there is always the concern that ethnic differences may affect the drug's efficacy or safety and that dosage adjustment may be required. Ethnic differences in drug disposition may derive from three different sources: differences in (a) subject size and weight, (b) metabolism and transport due to the diversity of the cytochrome P-450 (CYP) family and transport proteins, and (c) food consumption (dietary) habits. With regard to physiological differences, the Asian population tends to have a smaller body size and lower WT and BMI than a typical U.S. or European population, which can result in higher exposure ($C_{max}$ and AUC) in the Asian population. A recent example is the 25.6% higher AUC(inf)

of fosfluconazole in Japanese subjects compared to Caucasians, which could be explained by the difference in BW between the two populations [45].

The diversity of the human CYP family has been well established. At least 57 human CYP isoforms have been identified to date; however, only 10 of these isozymes contribute to drug metabolism. The isoforms CYP3A4, CYP2D6, and CYP2C9 are the main contributors. Most of the CYP isozymes exhibit some form of genetic polymorphism and have structurally altered alleles with reduced or deficient catalytic activity. Ethnic differences in the make-up and polymorphism of CYP isozymes have been extensively reviewed [46–49]. The frequency of the deficient allele expression varies across ethnic groups, and the differences are thought to be likely a result of regional evolutional diversification and environmental factors. There are also instances of variant alleles that may only be expressed in some ethnic groups but not in others [50]. Genetic polymorphism is not restricted to drug-metabolizing enzymes but has also been reported for drug transporters with affinity to chemotherapeutic agents such the multidrug-resistant *p*-glycoprotein (MDR1), multidrug resistance protein 2 (MRP2), and a breast cancer resistance protein [51, 52]. Genetic differences in metabolism and transport can be considered intrinsic ethnic variations independent of any environmental factors.

Food consumption habits are considered external factors and vary across ethnic groups within and across regions. Dietary constituents have been reported to alter the activity of various isozymes of the CYP family, phase II enzymes, and transporter systems and thus contribute to both inter- and intrasubject variability in drug PK. Harris et al. provide an in-depth review of various food constituents and their effect on metabolism enzymes and transport systems [53]. Dietary effects have been described for CYP3A4 (e.g., grapefruit juice, Seville orange juice, red wine, St. John's wort, garlic), CYPA12 (e.g., caffeine, grapefruit juice, jufeng grape juice, cruciferous vegetables, charbroiled meat), CYP2E1 (ethanol, watercress), and phase II enzymes (vegetables) and *p*-glycoprotein (fruit juices, St. John's wort). For this reason, foods that contain complex mixtures of phytochemicals such as herbs, spices, teas, fruits, and vegetables may affect the activity of enzymes or active transport mechanisms and contribute to ethnic variations in drug disposition.

In 1998, the ICH issued the Guideline E5: Ethnic Factions in the Acceptability of Foreign Clinical Data to provide direction to the pharmaceutical industry when drug submissions are considered in more than one region with likely ethnic diversity [54]. The ICH and FDA subsequently issued a follow-up questions-and-answer document to formally respond to common questions regarding the type and extent of bridging studies required to support submissions in new regions [55, 56]. While this guidance is directed towards regulatory strategies for a bridging data package including PD, efficacy, and/or safety bridging studies, it recognizes the importance of appropriate characterization of pharmacokinetics as part of the data package. Pharmacokinetic processes which are biologically or biochemically mediated have the potential to exhibit differences between ethnic groups [57]. If the NCE characteristics are known, it should be possible to identify drugs that are likely to exhibit race differences in PK. Drugs which have a wide therapeutic range, flat concentration-versus-effect profile, linear PK, minimal metabolism, high/nonvariable bioavailability, low protein binding, and low potential for drug–drug interactions fall into the classification of less likely to be sensitive to ethnic factors. Those with

opposite characteristics, metabolized by a single pathway or polymorphic enzyme, and likely coadministered with multiple medications are more likely to be sensitive to ethnic factors. The ICH considerations for NCE classification as either less likely or more likely to be sensitive to ethnic factors based on their PK characteristics are provided in Table 5.

With the emerging globalization of drug development, more drugs from the Western market (United States/Europe) enter the Asian market (such as Japan). For this reason, the comparison of these two populations has become of particular interest. Many of the newer drug applications are employing PopPK approaches to assess ethnic differences in drug disposition. Data may be compiled across one or more phase II/III studies or may also include appropriate PK data from phase I control populations. The data may come from clinical trials conducted either in the population relevant to the new region or in the new region. As long as the PopPK model is robust enough and a sufficiently large number of subjects of the desired ethnic group is available for the analysis, the analysis results should be able to discern race-related differences in PK. Recent examples of PopPK analysis including Japanese subjects are the modeling of PK parameters for (a) olmesartan, which has shown the absence of ethnic differences in drug clearance between Japanese and Westerners, (b) telmisartan, which contributed differences in food intake conditions across populations to the differences in telmisartan PK, and (c) darifenacin, which exhibited a lower bioavailability in Japanese males [58–60]. The above three examples illustrate the diversity of results across different NCEs.

Since formal studies in a different ethnic population are often part of the drug's transition into a new region, the design of the trial is often adapted to a specific

**TABLE 5 NCE Sensitivity to Ethnic Factors**

| Less Likely to be Sensitive to Ethnic Factors | More Likely to be Sensitive to Ethnic Factors |
|---|---|
| Linear PK | Nonlinear PK |
| Flat PD (effect v. concentration) curve for both efficacy and safety in range of recommended dosage and dose regimen | Steep PD curve for both efficacy and safety (small change in dose results in large change in effect) in range of recommended dosage and dose regimen |
| Wide therapeutic dose range | Narrow therapeutic dose range |
| Minimal metabolism or metabolism distributed among multiple pathways | Highly metabolized, especially through single pathway, thereby increasing potential for drug–drug interaction |
| High bioavailability, thus less susceptibility to dietary absorption effects | Metabolism by enzymes know to show genetic polymorphism |
| Low potential for protein binding | Administration of prodrug, with potential for ethnically variable enzymatic conversion |
| Little potential for drug–drug, drug–diet, and drug–disease interaction | High intersubject variation in bioavailability |
| Nonsystemic mode of action | Low bioavailability, thus more susceptibility to dietary absorption effects |
| Little potential for inappropriate use | High likelihood of use in setting of multiple comedications |
| | High likelihood for inappropriate use, e.g., analgesics and tranquilizers |

From [54].

need in the NCE's development program and may take on many forms. Regulatory guidelines do not pose any limits on the study design of formal PK comparisons across ethnic groups but rather focus on the type of data that are required for ethnic comparisons. Traditional ethnic PK studies may follow the design outlined in Section 2 in which a single high-end therapeutic dose of the NCE is administered in an equal number of the healthy ethnic group (e.g., Japanese) and healthy Western or Caucasian controls in a parallel-group design. Alternative approaches may include the administration of more than one dose strength to each subject in either a randomized or ascending-dose cross-over design [61–63]. Administration of multiple dose levels across the anticipated therapeutic dose range allows determination of dose proportionality of the NCE as well as comparative PK. The decision of whether to randomize or escalate doses may depend on the likelihood of ethnic differences in the NCE and the timing of the study within the overall development program. From a statistical perspective, a randomized cross-over design is preferred; however, safety considerations may outweigh the need for a balanced design.

Pharmacodynamic markers such as blood glucose levels to determine the relative changes (RC) in glycemia values may be employed to relate potential differences in NCE exposure to a clinically relevant pharmacological response. Thomsen et al. was able to demonstrate that the higher exposure to repaglinide, an insulin secretogogue, in Japanese subjects was associated with a greater decrease in the RC values in this population [61]. If the drug cannot be safely administered in healthy subjects, a formal PK study may be conducted in the target population. A recent report summarizes the results for cepacitabine, which was administered twice daily for 14 days to an approximately equal number of Caucasians and Japanese subjects with breast cancer [64]. Blood and urine data were collected on the first and last day of drug administration and data analysis followed standard approaches [ANOVA with 90% confidence interval (CI) for the ratio of the LS means]. When drug development is scheduled to occur simultaneously in different regions (e.g., United States and Japan), one consideration might be to include both a Western and Japanese population in a "first-in-man" dose tolerance study conducted in one of the two regions. PK, PD, safety, and tolerance data allow early determination of potential ethnic differences in NCE behavior and appropriate adjustment in a global development program.

The same inclusion/exclusion considerations apply for ethnic comparisons as for other special population comparisons. Where possible, subjects should be healthy and free of concomitant medications, and the study should be powered with a sufficient number of subjects to allow statistical comparisons. Control groups should be age and gender matched whenever possible. The standard age range may require minor adjustment if regional requirements stipulate a higher minimum age for a typical phase I population. The population should be representative of the new ethnic region. If the formal PK study is not conducted in the new ethnic region (e.g., Japanese subjects are studied in the United States), inclusion criteria should clearly define the appropriate population. The study may be restricted to only include first-generation (Japanese) subjects or may be extended to second (or later) generation(s) of the ethnic population if no generation differences are expected. Since ethnic differences in NCE PK are partially attributed to differences in diet and life-style, a decision needs to be made whether the two populations are to receive the same meals during their study conduct period or whether the subjects

should receive meals that are consistent with their typical diet. In the case of a comparison to Japanese subjects, inclusion of smokers may be required since a higher number of Japanese smoke and exclusion of smokers might limit enrollment or is not representative of a typical population in the new ethnic region.

In summary, PK data relevant to all ethnic populations targeted for treatment need to be part of a submission package. Both formal PK studies or PopPK analysis of clinical trials can support the registration package. How the data are generated will depend on the NCE characteristics and the NCE development program.

## 8 OBESITY

Obesity can be considered a nutritional disorder and is defined as a state of excess body fat and body weight in proportion to an individual's height. The weight-to-height ratio, BMI, is used to determine whether an individual is obese. An adult with a BMI in the range of 20–24.9 is considered normal, an adult with a BMI of 25–29.9 is considered overweight, and individuals with a BMI of 30 or greater are considered obese [65, 66]. The number of obese adults and children has tremendously increased over the last 10–20 years, especially in the Western nations. Obesity is now recognized as a global epidemic and a significant health risk. While no linear relationship exists, obesity is associated with hypertension and more severe cardiovascular diseases, dyslipidemia (e.g., high total cholesterol or triglyceride levels), type 2 diabetes, gallbladder disease, osteoarthritis, some cancers (endometrial, breast, and colon), pulmonary disorders, and sleep apnea/sleep-disordered breathing [66–68]. Obesity in children and adolescents between 6 and 19 years of age is increasing more rapidly than obesity in adults. This population is especially at risk for many comorbid conditions such as hypotension and diabetes at an early age, which will affect their long-term health and life expectancy [69].

Obese subjects are becoming a major target population for drugs in the future. Use of medications in this population is twofold: (1) drug products are marketed and in development to treat obesity and (2) this population requires medications to treat the secondary illnesses described above. In the first scenario, the medications are specifically developed for the obese population. Regulatory expectations for the clinical evaluation of obesity drugs are outlined in both U.S. and European guidance documents [70, 71]. In contrast to drug development in most other therapeutic indications, first-in-man and other early clinical pharmacology studies are usually performed in obese subjects who are otherwise healthy. Furthermore, the majority of subsequent clinical pharmacology and biopharmaceutic studies with the exception of specialized phase I studies such as RI and HI studies and formal QTc studies will be conducted in the obese population. For obesity drugs, this population is considered the target population and does not represent a special population.

For the second scenario, obese subjects can be considered a special population rather than a target patient population. Effect of obesity on drug disposition has been studied as early as the late 1980s when obesity was first recognized as an emerging nutritional disorder, and it was found that modifications in body weight can affect drug distribution into tissues [72, 73]. While slightly liposoluble molecules are less likely affected by increases in body fat, lipophilic drugs may have a higher volume of distribution in the obese population and therefore a longer half-life. For

this reason, dose adjustments may be different in obese subjects based on the physicochemical characteristic of the drug.

The effect of obesity on the PK in various drug classes, including anti-infectives, anticancer, CNS, obesity, anesthetics, and beta blockers, has been reviewed [74]. Drug disposition in the obese can be altered not only due an increase and change in body mass (lean WT plus fat WT) but also due to intrinsic physiological changes and concurrent disease. While drug absorption is unlikely to be affected by obesity, both drug distribution and elimination may be altered. Blood flow into fatty tissue is poor, hemodynamic conditions in obesity may be altered, and livers in obese subjects may suffer from fatty infiltration which may lead to HI. Obesity may be associated with RI, and exposure of drugs which are primarily eliminated by the kidney may increase. However, obese subjects have larger kidneys than nonobese subjects [75], and the clearance of some renally eliminated drugs is not substantially affected [74, 76]. The effect of obesity on CYP was reviewed by Kotlyar and Carson, who concluded that changes in CYP activity differed across isozymes [77]. CYP3A4 activity was reported to decrease while that of CYP2E1 appeared to increase. The authors considered the findings regarding other CYP isozymes (1A2, 2C9, 2C19, and 2D6) inconclusive.

Since the potential effect of underlying diseases and intrinsic changes other than body mass are difficult to assess, all dose recommendations to date have been related to weight (i.e., milligram-per-kilogram recommendations). Different weight algorithms have been employed. The most common approaches are total, lean, or ideal body weight [73]. Recently use of predicted normal body weight (PNWT) has been proposed) [78]. The PNWT corrects for excess fat weight (in relationship to lean body weight) that is above the weight of fat normally expected and may therefore be a better standard weight descriptor to describe body size for dose adjustments in obese patients. However, it remains difficult to predict the impact of obesity on PK, and each drug may behave differently. For drugs with a narrow therapeutic index, a study in obese subjects may be indicated. Anticancer agents are one category of drugs where the understanding of the influence of obesity is patchy and the outcome of treatment in this population may be detrimental due to the narrow therapeutic window of these agents. Toxicity for obese patients is impacted by appropriate dose adjustments and outcome data have suggested that obese patients may have a lesser disease-free survival and overall survival [79]. Based on his review, Navarro [79] recommends prospective studies to address optimal chemotherapy dosing in the obese population.

While formal obesity PK studies are currently not a regulatory requirement, it might be advisable to obtain PK information in the obese population if (a) the drug is likely prescribed in this population, (b) the physiochemical properties of the NCE suggest potential differences in drug disposition, and (c) the drug has a narrow therapeutic window. The data can be derived from either PopPK approaches or a formal study. The design of a formal PK study follows that outlined for HI with categorical comparison of the obese population to matched controls. Two subpopulations have been defined for obesity: moderately obese subjects with BMI values in the range of 30–39 $kg/m^2$ and morbidly obese subjects with BMI values equal to or greater than 40 $mg/m^2$. PK data are compared in equal numbers of a group of obese subjects and demographically matched nonobese subjects. Designs may include only moderately obese subjects or may be further stratified between moderately and morbidly obese subjects [76, 80]. The decision to include morbidly obese subjects

depends on the likelihood that these subjects are a substantial part of the target patient population. Inclusion/exclusion criteria for secondary diseases and comedications follow the same considerations described for RI and HI studies and will vary between the obesity categories. Since the onset of secondary illness is not linearly related to obesity, there is a sufficiently large pool of reasonably healthy subjects with moderate obesity; however, morbidly obese subjects will likely require lenient inclusion/exclusion criteria. The number of subjects enrolled in each of the treatment arms will depend on statistical and logistical (in the case of morbidly obese) considerations, while PK sampling and data analysis strategies are the same as described for the other special populations.

In the current environment, formal studies in obese subjects are rarely performed for registration purposes and the majority of data derive from PopPK analyses. PopPK approaches are appropriate if the database contains a sufficiently large number of obese subjects with a wide enough BMI range to allow comparison to nonobese subjects. Since morbidly obese subjects are unlikely to be included in standard phase II/III safety and efficacy studies, PopPK analysis will rarely provide PK information for these subjects and a formal PK study may be needed if data in this population are required for a particular registration.

## 9 CONCLUSIONS

Pharmacokinetic data in special populations are registration requirements if the target patient population is likely to include the special population(s). U.S. and European regulatory guidelines require formal PK studies in geriatric, RI, and HI impairment subjects while PK data in other special populations such as women, ethnic groups, and obese subjects may be obtained by using population analysis approaches. Exceptions to these rules can be found in the specific regulatory guidelines. Study designs for formal PK studies carry common features across all special populations with special considerations to age- or disease-related inclusion/exclusion criteria and comedication requirements that are less restrictive that those in healthy normal subjects.

Over the last 5–10 years, population analysis approaches have obtained greater acceptance in the international community and PopPK results have either supplemented PK data of formal studies or replaced those of formal studies. Population analysis approaches are powerful tools to discern intersubject variability in PK parameters and are only limited by the inclusion/exclusion criteria of clinical studies which may preclude enrollment of (sufficient numbers) of a special population of interest. While the use of population analysis approaches to understand the effect of demographic, physiological, genetic, or life-style covariates on drug disposition, efficacy, or safety response to drugs is likely to increase over the next years, formal PK studies for some populations such as RI and HI will continue to be required to the limited enrollment of these subjects in the phase II and III program.

## REFERENCES

1. Food and Drug Administration (1993), Guidance for Industry: Guidelines for the study and evaluation of gender differences in clinical evaluation of drugs. U.S. Department of

Health and Human Services, Center for Drug Evaluation and Research, Center for Biologics Evaluation and Research, July.

2. Gibaldi, M., and Perrier, D. (1982), *Pharmacokinetics*, 2nd ed., Marcel Dekker, New York.
3. Food and Drug Administration (2001), Guidance for Industry: Statistical approaches to establishing bioequivalence. U.S. Department of Health and Human Services, Center for Drug Evaluation and Research; available at: http://www.fda.gov/cder/guidance/3616fnl.pdf.
4. Food and Drug Administration (2003), Guidance for Industry: Bioavailability and bioequivalence studies for orally administered drug products—General considerations. U.S. Department of Health and Human Services, Center for Drug Evaluation and Research, Center for Biologics Evaluation and Research; available at: http://www.fda.gov/cder/guidance/5356fnl.pdf.
5. Abernethy, D. R., and Azarnoff, D. L. (1990), Pharmacokinetic investigations in elderly patients. Clinical and ethical considerations, *Clin. Pharmacokinet*, 19(5), 89–93.
6. Vestal, R. E., and Gurwitz, J. H. (2000), Geriatric pharmacology, in Curruthers, S. G., Hoffman, B. B., Melmon, K. L., and Nierenberg, D. W., Eds., *Melmon and Morelli's Clinical Pharmacology*, 4th eds, McGraw-Hill, New York, pp. 1151–1177.
7. Rowe, J. W., Andres, R., Tobin, J. D., et al. (1976), The effect of age on creatinine clearance in men: A cross-sectional and longitudinal study, *J. Gerontol.*, 31, 155–163.
8. Woodhouse, K. W., and Wynne, H. A. (1988), Age-related changes in liver size and hepatic blood flow. The influence on drug metabolism in the elderly, *Clin. Pharmacokinet.*, 15, 287–294.
9. Food and Drug Administration (1989), Guidance for Industry: Guidance for the study of drugs likely to be used in the elderly. U.S. Department of Health and Human Services, Center for Drug Evaluation and Research, Center for Biologics Evaluation and Research.
10. International Conference on Harmonisation (ICH) (1993), Studies in support of special populations: Geriatrics (E7); available at: http://www.ich.org/LOB/media/MEDIA483.pdf.
11. Food and Drug Administration (2001), Guidance for Industry: Content and format for geriatric labeling. U.S. Department of Health and Human Services, Center for Drug Evaluation and Research, Center for Biologics Evaluation and Research; available at: http://www.fda.gov/cber/gdlns/gerlab.pdf.
12. Abernethy, D. R. (2001), Drug therapy in the elderly, in Atkinson, A. J., Daniels, C. E., Dedrick, R. L., Grudzinskas, C. V., and Markey, S. P., Eds., *Principles of Clinical Pharmacology*, Academic, New York, pp. 307–317.
13. Guyton, A. C. (1987), The body fluids and the kidneys—renal disease, in Guyton, A. C. Ed., *Human Physiology and Mechanisms of Disease*, 4th ed., W.B. Saunders, Philadelphia, pp. 286–290.
14. Food and Drug Administration (1998), Guidance for Industry: Pharmacokinetics in patients with impaired renal function—study design, data analysis, and impact on dosing and labeling. U.S. Department of Health and Human Services, Center for Drug Evaluation and Research, Center for Biologics Evaluation and Research; available at: http://www.fda.gov/cder/guidance/1449fnl.pdf.
15. European Medicines Agency (2004), Note for Guidance on the evaluation of the pharmacokinetics of medicinal products in patients with impaired renal function. CPMP/EWP/225/02; available at: http://www.emea.eu.int/pdfs/human/ewp/022502en.pdf.

16. Zini, R., Riant, P., Barre, J., et al. (1990), Disease-induced variations in plasma protein levels: Implications for drug dosage regimens (Part I), *Clin. Pharmacokinet.*, 19, 147–159.
17. Zini, R., Riant, P., Barre, J., et al. (1990), Disease-induced variations in plasma protein levels: Implications for drug dosage regimens (Part II), *Clin. Pharmacokinet.*, 19, 218–229.
18. Touchette, M. A., and Slaughter, R. L. (1991), The effect of renal failure on hepatic drug clearance. *DICP: Ann. Pharmacother.*, 25, 1214–1224.
19. Turnheim, K. (1991), Pitfalls of pharmacokinetic dosage guidelines in renal insufficiency, *Eur. J. Clin. Pharmacol.*, 40, 87–93.
20. Lam, Y. W. F., Banerji, S., Hatfield, C., et al. (1997), Principles of drug administration in renal insufficiency. *Clin. Pharmacokinet.*, 32(1), 30–57.
21. Launay-Vacher, V., Storme, T., Izzedine, H., et al. (2001), Pharmacokinetic changes in renal failure, *Presse Med.*, 30, 597–604.
22. Landray, M. J., Thambyrajah, J., McGlynn, F. J., et al. (2001), Epidemiological evaluation of known and suspected cardiovascular risk factors in chronic renal impairment, *Am. J. Kidney Dis.*, 38(3), 537–546.
23. Salem, M. M. (2002), Pathophysiology of hypertension in renal failure, *Semin. Nephrol.*, 22, 17–26.
24. Wheeler, D. C. (2001), Lipid abnormalities in the nephrotic syndrome: The therapeutic role of statins, *J. Nephrol.*, 14(Suppl), S70–75.
25. Robbins-Weilert, D. (2004), Design, conduct, and analysis of studies in patients with renal impairment, in Bonate, P. L., and Howard, D. R., Eds., *Pharmacokinetics in Drug Development: Clinical Study Design and Analysis, Vol. 1*, American Association of Pharmaceutical Sciences, pp. 177–208.
26. Bonate, P. L., Reith, K., and Weir, S. (1998), Drug interactions at the renal level. Implications for drug development, *Clin. Pharmacokinet.*, 34(5), 375–404.
27. Cockcroft, D. W., and Gault, M. H. (1976), Prediction of creatinine clearance from serum creatinine, *Nephron*, 16, 31–41.
28. Schwartz, G. J. (1976), A simple estimate of glomerular filtration rate in children derived from body length and plasma creatinine, *Pediatrics*, 58, 259–264.
29. Schwartz, G. J. (1984), A simple estimate of glomerular filtration rate in full term infants during the first year of life, *J. Pediatr.*, 104, 849–854.
30. Dahl, N. V. (2001), Herbs and supplements in dialysis patients: Panacea or poison, *Semin. Dialysis*, 14(3), 186–192.
31. Bailie, G. R., and Eisele, G. (1992), Continuous ambulatory peritoneal dialysis: A review of its mechanism, advantages, complications, and areas of controversy. *Ann. Pharmacother.*, 26, 1409–1420.
32. Taylor, C. A., 3rd, Abdel-Rahman, E., Zimmerman, S. W., et al. (1996), Clinical pharmacokinetics during continuous ambulatory peritoneal dialysis, *Clin. Pharmacokinet.*, 31(4), 293–308.
33. Lee, C. C., and Marbury, T. C. (1984), Drug therapy in patients undergoing hemodialysis, *Clin. Pharmacokinet.*, 9, 42–66.
34. Oellerich, M., Burdelski, M., Lautz, H.-U., et al. (1990), Lidocaine metabolite formation as a measure of liver function in patients with cirrhosis, *Ther. Drug Monit.*, 12:219–226.
35. Figg, W. D., Dukes, G. E., Lesesne, H. R., et al. (1995), Comparison of quantitative methods to assess hepatic function: Pugh's clarification, indocyanine green, antipyrine, and dextromethorphan, *Pharmacotherapy*, 15, 693–700.

36. Pugh, R. N. H., Murray-Lyon, I. M., Dawson, J. L., et al. (1973), Transection of the oesophagus for bleeding oesophageal varices, *Br. J. Surg.*, 60, 646–649.
37. Zakim, D., and Boyer, T. D. (1996), Hepatology, in *A Textbook of Liver Disease*, W.B. Saunders Company, Philadelphia.
38. Food and Drug Administration (2003), Guidance for Industry: Pharmacokinetics in patients with impaired hepatic function—study design, data analysis, and impact on dosing and labeling. U.S. Department of Health and Human Services, Center for Drug Evaluation and Research, Center for Biologics Evaluation and Research; available at: http://www.fda.gov/cder/guidance/3625fnl.pdf.
39. European Medicines Agency (2005), Guideline on the evaluation of the pharmacokinetics of medicinal products in patients with impaired hepatic function. CPMP/EWP/2339/02; available at: http://www.emea.eu.int/pdfs/human/ewp/233902en.pdf.
40. Bonate, P. L., and Russell, T. (2004), Design, conduct, and analysis of studies in patients with hepatic impairment, in Bonate, P. L., and Howard, D. R., Eds., *Pharmacokinetics in Drug Development: Clinical Study Design and Analysis, Vol. 1*, American Association of Pharmaceutical Sciences, pp. 149–175.
41. Food and Drug Administration (1997), Guidance for Industry: M3 Nonclinical safety studies for the conduct of human clinical trials for pharmaceuticals. U.S. Department of Health and Human Services, Center for Drug Evaluation and Research, Center for Biologics Evaluation and Research, and ICH; available at: http://www.fda.gov/cder/guidance/1855fnl.pdf.
42. FDAMA Women and Minorities Working Group Report (1998), posted by Woodcock, J. at: http://www.fda.gov/cder/guidance/women.pdf.
43. ICH and Women: Gender consideration in the conduct of clinical trials; available at: http://www.ich.org/cache/compo/276-254-1.html.
44. European Medicines Agency (2005), ICH Step 5: Gender consideration in the conduct of clinical trials. EMEA/CHMP/3916/2005; available at: http://www.emea.eu.int/pdfs/human/ich/391605en.pdf.
45. Sobue, S., Tan, K., Shaw, L., et al. (2004), Comparison of the pharmacokinetics of fosfluconazole and fluconazole after single intravenous administration of fosfluconazole in healthy Japanese and Caucasian volunteers, *Eur. J. Clin. Pharmacol.*, 60(4), 247–253.
46. Cascorbi, I. (2003), Pharmacogenetics of cytochrome P4502D6: Genetic background and clinical implications, *Eur. J. Clin. Investigation*, 33, 17–22.
47. Schwarz, U. I. (2003), Clinical relevance of genetic polymorphisms in the human CYP2C9 gene, *Eur. J. Clin. Investigation*, 33, 23–30.
48. Daly, A. K. (2004), Pharmacogenetics of the cytochromes P450, *Curr. Top. Med. Chem.*, 4(16), 1733–1744.
49. Ozawa, S., Soyama, A., Saeki, M., et al. (2004), Ethnic differences in genetic polymorphisms of CYP2D6, CYPC19, CYP3As, and MDR1/ABCB1, *Drug Metabol. Pharmacokinet.*, 19(2), 83–95.
50. Roy, J. N., Lajoie, J., Zijenah, L. S., et al. (2005), CYP3A5 genetic polymorphisms in different ethnic populations, *Drug Metabol. Disp.*, 33(7), 884–887.
51. Sakaeda, T., Nakamura, T., and Okumura, K. (2004), Pharmacogenetics of drug transporters and its impact on the pharmacotherapy, *Curr. Top. Med. Chem.*, 4(13), 1385–1398.
52. Bosch, T. M., Meijerman, I., Beijnen, J. H., et al. (2006), Genetic polymorphisms of drug-metabolizing enzymes and drug transporters in the chemotherapeutic treatment of cancer, *Clin. Pharmacokinet.*, 45(3), 253–285.
53. Harris, R. Z., Jang, G. R., and Tsunoda, S. (2003), Dietary effects on drug metabolism and transport, *Clin. Pharmacokinet.*, 42(13), 1071–1088.

54. International Conference on Harmonisation (ICH) (1998), Ethnic factors in the acceptability of foreign clinical data E5(R1); available at: http://www.ich.org/LOB/media/MEDIA481.pdf.
55. Food and Drug Administration (2004), Guidance for Industry: Ethnic factors in the acceptability of foreign clinical data. Questions and answers. U.S. Department of Health and Human Services, Center for Drug Evaluation and Research, Center for Biologics Evaluation and Research, ICH; available at: http://www.fda.gov/cder/guidance/6200fnl.pdf.
56. International Conference on Harmonisation (ICH) (2003), E5 Ethnic factors: Questions and answers; available at: http://www.ich.org/LOB/media/MEDIA1194.pdf.
57. Johnson, J. A. (1997), Influence of race or ethnicity on pharmacokinetics of drugs, *J. Pharm. Sci.*, 86(12), 1238–1233.
58. Yoshihara, K., Gao, Y., Shiga, H., et al. (2005), Population pharmacokinetics of olmesartan following oral administration of its prodrug, olmesartan medoxomil: in healthy volunteers and hypertensive patients, *Clin. Pharmacokinet.*, 44(12), 1329–1342.
59. Tatami, S., Yamamura, N., Sarashina, A., et al. (2004), Pharmacokinetic comparison of an angiotensin II receptor antagonist, telmisartan, in Japanese and western hypertensive patients using population pharmacokinetic method, *Drug Metab. Pharmacokinet.*, 19(1), 15–23.
60. Kerbusch, T., Wahlby, U., Milligan, P. A., et al. (2003), Population pharmacokinetic modeling of darifenacin and its hydroxylated metabolite using pooled data, incorporating saturable first-pass metabolism, CYP2D6 genotype and formulation-dependent bioavailability, *Br. J. Clin. Pharmacol.*, 56(6), 639–653.
61. Thomsen, M. S., Chassard, D., Evene, E., et al. (2003), Pharmacokinetics of repaglinide in healthy Caucasian and Japanese subjects, *J. Clin. Pharmacol.*, 43(1), 23–28.
62. Jhee, S. S., Lyness, W. H., Rojas, P. B., et al. (2004), Similarity of insulin detemir pharmacokinetics, safety, and tolerability profiles in healthy Caucasian and Japanese American subjects, *J. Clin. Pharmacol.*, 44(3), 258–264.
63. Van Giersbergen, P. L., and Dingemanse, J. (2005), Comparative investigation of the pharmacokinetics of bosentan in Caucasian and Japanese healthy subjects, *J. Clin. Pharmacol.*, 45(1), 42–47.
64. Reigner, B., Watanabe, T., Schuller, J., et al. (2003), Pharmacokinetics of capecitabine (Xeloda) in Japanese and Caucasian patients with breast cancer, *Cancer Chemother. Pharmacol.*, 52(3), 193–201.
65. World Health Organization (1998), Report of a WHO consultation on obesity: Preventing and managing the global epidemic, WHO, Geneva, June 3–5.
66. Center for Disease Control and Prevention (CDC). Overweight and obesity; available at: http://www.cdc.gov/nccdphp/dnpa/obesity/.
67. Li, Z., Bowerman, S., and Heber, D. (2005), Health ramifications of the obesity epidemic, *Surg. Clin. N. Am.*, 85(4), 681–701.
68. Poirier, P., Giles, T. D., Bray, G. A., et al. (2006), Obesity and cardiovascular disease: Pathophysiology, evaluation, and effect of weight loss: An update of the 1997 American Heart Association Scientific Statement on Obesity and Heart Disease from the Obesity Committee of the Council on Nutrition, Physical Activity, and Metabolism, *Circulation*, 113(6), 898–918.
69. Center for Disease Control and Prevention (CDC) (2005), Children and teens told by doctors that they were overweight—United States, 1999–2002, *MMWR, Rep.*, 54(34), 848–849.

70. Food and Drug Administration (1996), Guidance for the clinical evaluation of weight-control drugs. Division of Metabolic and Endocrine Drug Products; available at: http://www.fda.gov/cder/guidance/obesity.pdf.
71. European Medicinal Agency (1997), Note for guidance on clinical investigation of drugs used in weight control. CPMP/EWP/281/96. December 1997; available at: http://www.emea.eu.int/pdfs/human/ewp/028196en.pdf.
72. Cheymol, G. (1988), Drug pharmacokinetics in the obese, *Fundament. Clin. Pharmacol.*, 2(3), 239–256.
73. Cheymol, G. (1993), Clinical pharmacokinetics of drugs in obesity: An Update, *Clin. Pharmacokinet.*, 25(2), 103–114.
74. Cheymol, G. (2000), Effects of obesity on pharmacokinetics. Implications for drug therapy, *Clin. Pharmacokinet.*, 39(3), 215–231.
75. Naeye, K. L., and Rowe, P. (1970), The size and number of cells in several visceral organs in human obesity, *Am. J. Clin. Path.*, 54, 251–253.
76. Dvorchik, B. H., and Damphousse, D. (2005), The pharmacokinetics of daptomycin in moderately obese, morbidly obese, and matched nonobese subjects, *J. Clin. Pharmacol.*, 45, 48–56.
77. Kotlyar, M., and Carson, S. W. (1999), Effects of obesity on the cytochrome P450 enzyme system, *Int. J. Clin. Pharmacol. Therap.*, 37(1), 8–19.
78. Duffull, S. B., Dooley, M. J., Green, B., et al. (2004), A standard weight descriptor for dose adjustment in the obese patient, *Clin. Pharmacokinet.*, 43(15), 1167–1178.
79. Navarro, W. H. (2003), Impact of obesity in the setting of high-dose chemotherapy, *Bone Marrow Transplant.*, 31(11), 961–966.
80. Sarich, T. C., Teng, R., Peters, G. R., et al. (2003), No influence of obesity on pharmacokinetics and pharmacodynamics of melagatran, the active form of the oral direct thrombin inhibitor ximelagatran, *Clin. Pharmacokinet.*, 42(5), 485–492.

# *Musculoskeletal Disorders*

**Masami Akai**
*Director, Rehabilitation Hospital,*
*National Rehabilitation Center Japan, Saitama, Japan*

**Contents**

1 Introduction
2 Definition of Musculoskeletal Disorders
  2.1 Musculoskeletal Disorders
  2.2 Aims of Treatment
  2.3 Concept of "Construct"
  2.4 Assessment of Function and Quality of Life (QOL)
3 Background of Clinical Trials
  3.1 Meaning of Clinical Trial
  3.2 Estimate and Testing
  3.3 Various Biases
  3.4 Methods of Clinical Trials
  3.5 Types of Clinical Studies
  3.6 Random Allocation
4 Methodological Assessment for Clinical Trials
  4.1 Significance of Clinical Trials
  4.2 Critical Evaluation of RCTs
5 Preparing Necessary Tools for Clinical Trials
  5.1 Developing Protocol
  5.2 Patients
  5.3 Interventions
  5.4 Outcomes
  5.5 How to Choose Right Outcome Measures
  5.6 Questionnaires

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

6 Data Analysis and Interpretation
6.1 Sample Size
6.2 Intention to Treat
6.3 Multiple Comparisons
6.4 Post Hoc Analysis and Subgroup Analysis
6.5 Data Synthesis and Combined Result
6.6 Side Effects
6.7 Reporting Clinical Trials
7 RCT and Future Directions
7.1 Impact from RCT to Daily Practice
7.2 Internal Validity and External Validity
7.3 Quantitative Study and Qualitative Study
7.4 Further Suggestions
References

## 1 INTRODUCTION

When we look at past medical history, we can only find a few definite treatments that focus directly on their etiology, such as vaccination for some infectious diseases, vitamin C for scurvy, and, recently, gene therapy. The evidence for presently accepted treatments in our practice is insufficient; it ranges from a scattered presence to total absence. Unfortunately, apart from a few dichotomous results, we cannot expect qualitatively definite results of treatment for all clinical cases. The effect of the interventions is usually assessed quantitatively, comparing before and after treatment, according to certain outcome measures. The key to solving many of the problems in medical practice is to establish valid and reliable methods of evaluating the differences in the outcomes before and after treatment.

I started my postgraduate training as an orthopedic surgeon and then shifted my career to rehabilitation medicine. I am not, by any standard, an expert in all of the subjects covered here: clinical epidemiology, biostatistics, and related fields. However, I have had chances to join teams to organize clinical trials [1, 2]. I would like to describe the necessary steps for conducting a clinical trial from the standpoint of a clinician.

The term *evidence-based medicine* (EBM), which is the integration of best research evidence with clinical expertise and patient values, has already established wide popularity, even in the musculoskeletal field. When we review the clinical trial trends in the musculoskeletal field, we notice several characteristic points, which are obviously different from other areas.

In the musculoskeletal field, there are the typical clinical trials that use placebo-controlled, double-blind methods, such as drug trials for osteoporosis or rheumatoid arthritis. However, if we wish to compare one interventional procedure with another, the double-blind design is often impossible to do methodologically. Only in an open-label trial, where patients know about their treatment, is it possible to conduct the trial under random allocation. With such limitations, we have to suppress various biases as much as possible. Another point that needs attention is the selection of outcome measures suitable for musculoskeletal disorders. Even though there are primary malignant tumors, such as osteosarcoma or a cancer lesion as secondary

metastasis, which are sometimes lethal, the majority of musculoskeletal disorders that develop clinical manifestation are not life-threatening but do impair function. We have to use outcome measures for observing the functional ability of the patients, not for the numerical laboratory data.

In this chapter I would like to discuss the following:

1. Definition of musculoskeletal disorders
2. Background of clinical trials
3. Methodological assessment for clinical trials
4. Preparation of necessary tools for clinical trials
5. Data analysis and interpretation
6. RCT and our future direction

## 2 DEFINITION OF MUSCULOSKELETAL DISORDERS

### 2.1 Musculoskeletal Disorders

The term *musculoskeletal disorders* refers to conditions that involve the supporting structures of the body, such as the trunk and the bones and joints of the extremities. In other words, musculoskeletal disorders involve locomotor functions that have suffered from trauma, diseases, anomalies, and inevitable aging and are mainly attended to by orthopedic surgeons.

Musculoskeletal disorders are the most common cause of severe, long-lasting pain and physical disability affecting hundreds of millions of people in the world. The extent of such problems and the growing burden is described as "age quake" instead of earthquake. The impact on society from these disorders, combined with the recognition that our health care resources need to be used more efficiently, has led to the organized movement of the "Bone and Joint Decade 2000–2010" [3]. This umbrella organization is attempting to raise social awareness of the suffering and pain of musculoskeletal disorders, as well as the growing burden and cost to society that will come with this age quake. We need to advance clinical research in order to reduce the burden of musculoskeletal diseases.

### 2.2 Aims of Treatment

The aims of treatment for musculoskeletal disorders are not to reduce mortality or morbidity but to increase mobility and function, to relieve pain, or to improve quality of life. We have to assess the results of our interventions with a view to the improvement in disability or functional limitation after treatment. In this field, the clinical indicators are neither a 5-year survival rate nor an infantile survival rate. Our aims are to improve the health-related quality of life of the millions of patients suffering from musculoskeletal disorders such as joint diseases, spinal disorders, severe trauma to the body and extremities, and deformity and crippling diseases in children.

When we talk about outcome measures in disability assessment, the key issue to be discussed is what "function" and "disability" mean and how to measure them.

### 2.3 Concept of "Construct"

Our function or ability is measured in accordance with conceptual "construct" [4]. We cannot measure this function or ability directly.

In the past, orthopedic surgeons mainly evaluated the amount of disease involvement in view of physiological or biochemical parameters, pathological findings, and impairments such as range of motion or muscle strength. It was thought that these indicators were objective and independent of the unreliable responses or feelings of patients. However, are these indicators truly reflective of the needs of patients and their families to know how much they can expect as a result of treatment? Their expectations and questions would be more like, "Can I walk again?" or "Will she be able to live by herself as before?"

When a certain standardized scale is not available to measure the function or ability of interests, a new construct, according to a concept for measurement, has to be made.

Conventional clinical indicators, such as a 5-year survival rate or an infantile mortality rate, have been recognized as very important health measures in assessing health conditions in society, but lately, the situation has been changing. The recent close-up of patient-based outcome measure is deeply related to this trend [5]. We have to recognize the demand and organize those types of outcome measures for the individual [6].

### 2.4 Assessment of Function and Quality of Life (QOL)

Functional assessment, from the physician's point of view, has been mainly used to evaluate problems at the impairment level. Traditionally, orthopedic surgeons have used clinician-based outcomes such as pain score, measurement of range of motion, muscle manual test, or X-ray findings. However, due to various biases and poor judgments of accuracy levels, clinician-based outcomes often affect the clinical ratings on the functional status of patients [7].

For the past two decades, new outcome movements have placed importance on patient-based functional assessment and measurement of health-related quality of life [5, 8].

It is because the statistically significant changes of the visual analog scale (VAS) for pains, from 68 to 42, do not mean any improvement for the patient with chronic pain. The improvement of knee flexion range of 35° does not answer the questions from patients and their families.

The end result of health services should take into account the experiences, preferences, and values of the patients [9]. The discomfort described by patients should be managed properly. We have to evaluate and treat the patient not the disease.

The so-called health-related QOL became the main target of the outcome measures related to patients' values. As well, of course, it is clear that health-related QOL assessments alone are not enough. In 2001, the World Health Organization (WHO) proposed a new "health" concept called the International Classification of Functioning, Disability and Health (ICF) [10], which describes the pluralistic conception of functional disabilities [11]. Based on the ICF concept, for example, we would assume the problems associated with patients with knee osteoarthritis were as follows:

| | |
|---|---|
| Health condition | Osteoarthritis |
| Impairment | Knee pain, limited joint movement, muscle weakness |
| Activity limitation | Difficulty in mobility-related daily activities |
| Participation restriction | Difficulty in participating in social life |

Many items and domains in the outcome measures should be constructed to identify all the characteristics of subjects, with a view to disability and impairment. Irrgang and Anderson [12] described the process for developing such a new measure for health-related quality of life when they designed clinical research related to the knee. In many cases, several outcome measures must be combined to cover every aspect from functional impairment to quality of life.

## 3 BACKGROUND OF CLINICAL TRIALS

### 3.1 Meaning of Clinical Trial

Two related statistical factors have to be considered in order to evaluate the quality of clinical studies. This is essential for the understanding of the quality of studies based on the level of evidence. First, in clinical trials we deal with human beings as research objects. Each human being is different; response is always uncertain, backgrounds vary, and mental influence is inevitable. We should compare groups that are similarly distributed, except for the specific factor that is the target of assessment. Thus, observed differences at the assessments can be directly attributed to the study targets. Second, we mainly evaluate an extracted specimen of limited numbers and then extrapolate the conclusion toward the source population. We attribute the presumption from the sample population to the general population, using statistical inference techniques. Using the statistics observed in the investigated samples, the parameters of the source population are estimated. Depending upon these two premises, we use sampling, randomization, comparison, and other statistical methods. This is a statistical outlook of the world around us.

### 3.2 Estimate and Testing

When we treat a patient, we regard him or her not just as a single case but as a representative case among many patients with the same disorder. This concept of patient treatment had already been described by Hippocrates in ancient Greece. There are two important statistical concepts we should consider, probability and confidence interval.

The first topic is the probability, that is, $P$ values that the result would arise by chance. Depending on the choice of cut-off level ($P < 0.05$ or $P < 0.01$), we apply a different level of "statistical significance" to the trial. If we find a result in the statistically significant range, we could reject the null hypothesis that there is no real difference between two groups. But a $P$ value in the nonsignificant range just shows that either there is no difference between the groups or there are too few samples to reveal a difference, if it actually exists.

Also, if we repeated the same trial hundreds of times, we would not get the exact same result each time. We cannot omit an incidence in each trial because we usually conduct only one trial, but we could find a particular level of difference on average

with a 95% confidence interval. If we conducted a trial with a much larger sample size, we would be more likely to show whether a significant result exists or not. The answer is in the comparison of the 95% confidence limits and the clinical significant levels. Very narrow 95% confidence limits suggest definitive results and exclude any clinical obscurity in comparisons. The confidence intervals reveal whether the trials support the result and to what extent and whether any further studies need to be done to reinforce the result.

### 3.3 Various Biases

The results of a trial consist of three parts: (1) true value, (2) random error (imprecision) by chance, and (3) systematic error (bias) by other factors. We cannot eradicate random error, but we can eliminate the other biases by using well-designed study protocols. The goal is to suppress the biases as much as possible and improve the quality of the trial; different study designs require different steps to reduce systematic errors. In clinical epidemiology textbooks, there are various biases pointed out such as sampling bias, selection bias, and information bias.

Confounding is the distortion of an association between two factors brought about by the association of another extraneous factor; for example, the association between lung cancer and alcohol consumption is confounded by smoking. The methods of controlling confounding in the design of a study are: restrict the participants to the study, match individuals with other comparison groups, or use random allocation. In the analysis of a study, restrict once again the participants in the data to stratify the individuals into subgroups according to categories of confounding factor and use multivariate methods.

### 3.4 Methods of Clinical Trials

To assess the effectiveness of a certain intervention, it is necessary to compare the baseline (before state) with the result (after state) of treatment. But generally speaking, the more abnormal value often shows more improvement.

Detected improvement does not always imply that the intervention was effective. Several other factors may have contributed to the same kind of improvement; such possible factors are natural recovery, "regression to the mean" phenomenon, or the psychological effect called placebo effect. We have to compare with control to get the true change rates by test intervention.

In most clinical trials called controlled trials, there is a group receiving a certain treatment and a comparative group. An essential premise of this kind of trial is that there is genuine uncertainty as to which treatments will be best for the patient. It is this uncertainty that justifies random allocation of patients after the consent forms to enter the trial have been given. Therefore, patients should have the targeted disease and satisfy all the required conditions, that is, inclusion criteria.

### 3.5 Types of Clinical Studies

If there were a clinical question about a patient's management, we should perform a clinical study, no matter what the study design is. The basic structures of a study design consist of a combination of some dichotomous divisions.

1. *Experimental (Interventional) Study or Observational Study* In experimental study, a population is selected for a planned trial of a regimen whose effects are measured by comparing the outcome of the regimen in the experimental group with the outcome of another regimen in the control group. Observational study is a nonexperimental study that does not involve any intervention.
2. *Comparative or Noncomparative Design* In a comparative design, the study design has a control group to compare with the active group. In a noncomparative design, there is no control group to compare with the active group.
3. *Intergroup or Intragroup Design* This is comparison of interventional effects among the groups, or comparison in the same single group, before and after the intervention.
4. *Cross-Sectional or Longitudinal Design* This is the directionality of the study. Data are collected at a single time point or collected at two or more time points and followed up later. But some cross-sectional studies refer retrospectively to experiences in the past.
5. *Retrospective or Prospective Design* This is a part of longitudinal studies. Data are collected at baseline points and other points in the past or future. The timing of data collection is also called concurrent, historical, or mixed.

Considering these dichotomous divisions, a widely used basic study design is a parallel design, which compares the groups, or a cross-over design, which compares the before and after. In the parallel design, subjects are divided into two or more groups, each group receives each treatment concurrently, and the results of treatment among the groups are compared. In the cross-over design, subjects receive two or more treatments at stated intervals and the effect of treatment, within the same subject, is evaluated. The cross-over design has the advantage of a smaller sample size than the parallel design but requires caution to minimize the fluctuating effect of disease severity ("order effect") and maintain the stability of patients' symptoms using a "wash-out period."

The randomized controlled trial (RCT) provides the most powerful proof of treatment efficacy:

*Randomized Controlled Trials* Subjects are randomly allocated into two or more groups. The groups are followed up for a specified time period and assessed in terms of outcomes defined at the start of trials. As the groups, on average, are identical except for the intervention as study target, any differences in outcomes are attributed, in theory, to the interventions.

*Open label* Subjects know which treatment they are receiving.

*More Sophisticated Randomized Controlled Trials*

*Single blind* Subjects do not know which treatment they are receiving.

*Double blind* Neither subjects nor investigators know who is receiving which treatment.

### 3.6 Random Allocation

Blinding patients to which treatments they are receiving is thought to be essentially important because the effect of the psychological expectations of the patients are

suppressed. In drug therapy, blinding is achieved with a placebo as a control. However, in certain situations, such as a comparative trial between surgery and nonsurgery, blinding is not easily possible. Double-blind techniques for surgical treatments are tentatively proposed; the anesthesiologists and surgeons in charge will maintain silence about treatment choices and patients and the persons assessing the outcomes will not be aware of the treatment choices. Because of the technical difficulty of being unaware of allocation, we sometimes have to accept that patients would know the treatment they received.

Random allocation is done immediately after completing the original registration for a trial with informed consent. However, patients often hesitate to enter a trial because, at the time of registration, they do not know which treatment is indicated for them. Zelen proposed a new method called "randomized consent design" in which random allocation is already performed at the candidate stage of being involved in a trial. Explanation of informed consent is done only about the allocated treatment [13, 14].

## 4 METHODOLOGICAL ASSESSMENT FOR CLINICAL TRIALS

Because of the explosive increase in health care information, all clinicians must be careful of scientific evidence regarding efficacy of treatment methods based on available medical literature. How are we to distinguish the good from the bad among the flood of information? We need a systematic way to select medical information including RCTs.

### 4.1 Significance of Clinical Trials

The practice of EBM is a process of self-directed, life-long, never-ending learning for our patients. Clinicians need to have the most up-to-date clinical information. Our goal is to continuously strive to improve our medical knowledge by scientific methodology. The introduction of newer information technology and medical knowledge has improved the practice of medicine. If we can properly manage to get all of the clinical information from contemporary medical practice, we will be able to master all of the medical experiences of that same period. The collected results could be the most verified medical knowledge that we have ever had.

In the musculoskeletal field, in-patients are, in most cases, operated on, but the majority of out-patients and some in-patients are treated with a combination of conservative therapeutic methods. We need to develop a systematic strategy for the indication and selection of various treatments. The accumulated results of RCTs could be the essential medical information for the most up-to-date treatments.

### 4.2 Critical Evaluation of RCTs

Randomized controlled trials are now regarded as having superior clinical significance, that is, the gold standard for proper treatment, and are recognized as the primary way to provide rigorous proof of efficacy. However, considering the mixture of facts and errors found in the results of studies, we have to check each step of the trial process, even in the case of RCTs. All trials can be scored on methodological

quality according to a few criteria. Basically, these criteria consist of four main categories [15–17]:

1. Design and study population
2. Description of intervention
3. Measurement of outcomes and follow-up
4. Data presentation and analysis

In the recent analyses, these 4 categories were usually further subdivided into a set of nearly 20 items, each with a given weight [18]. Various scales for quality assessment are now available and after some modifications, those for musculoskeletal disorders have been employed in several systematic reviews and meta-analyses [19, 20]. Introduced here is one of the sophisticated examples of a criteria list for assessing quality of trials in a published article (Table 1).

## 5 PREPARING NECESSARY TOOLS FOR CLINICAL TRIALS

On the basis of my personal experience, the next few sections describe the minimum essential steps in conducting a clinical trial (Table 2); the reader should refer to appropriate textbooks for further reading [21–26]. Recently, this kind of information has also been available on the Internet; the Resource Center for Randomized Trials [27] is an example. The Resource Center has various activities, and one of the useful services is a Web-based library for accessing information about trials; there are checklists for trials, consent forms for participants, and patient information leaflets. Other public information on clinical trials is also provided by the U.S. government [28].

### 5.1 Developing Protocol

The first step in conducting a trial is developing the protocol. When the details of how the trial is to be conducted are determined, it is essential to reduce various relating biases. Farrell and Spark [29] described a protocol checklist as a detailed method to use when conducting a trial (Table 3). Developing a protocol takes a long time, sometimes up to a few years to get funding, secure ethical approval, and organize a trial collaboration team. I think the best thing for anyone planning a trial for the first time to do is to refer to some successful protocols. Many protocols are now available on the Internet. A protocol should clearly show how outcomes will be measured, data collected, and the analysis conducted.

### 5.2 Patients

Use of inclusion/exclusion criteria is important to make it possible for the trial results to be more reliable. Inclusion criteria are used to identify appropriate subjects, keep them safe, and able to answer the research questions. Exclusion criteria are also very important in participant recruitment to avoid unnecessary involvement of inappropriate subjects in the study.

**TABLE 1 Criteria List for Assessing Methodological Quality of Trials**

| Heading | Subheading | Descriptor |
|---|---|---|
| Title | | Identify the study as a randomized trial. |
| Abstract | | Use a structured format. |
| Introduction | | State prospectively defined hypothesis, clinical objectives, and planned subgroup or covariate analyses. |
| Methods | Protocol | Describe |
| | | Planned study population with inclusion or exclusion criteria. |
| | | Planned interventions: their nature, content, and timing. |
| | | Primary and secondary outcome measure(s) and the minimum important difference(s), and indicate how the target sample size was estimated. |
| | | Reasons for statistical analyses chosen, and whether these were completed on an intention-to-treat basis. |
| | | Mechanisms for maintaining intervention quality, adherence to protocol, and assessment of fidelity. |
| | | Prospectively defined stopping rules (if warranted). |
| | Assignment | Describe |
| | | Randomization (e.g., individual, cluster, geographic). |
| | | Allocation schedule method. |
| | | Method of allocation concealment. |
| | Masking (blinding) | Describe |
| | | Mechanism for maintaining blind and allocation schedule control. |
| | | Evidence for successful blinding. |
| Results | Participant flow and follow-up | Provide a trial profile summarizing participant flow, numbers and timing of randomization assignment, interventions, and measurements for each randomized group. |
| | Analysis | State estimated effect of intervention on primary and secondary outcome measures, including a point estimate and measure of precision (confidence interval). |
| | | State results in absolute numbers when feasible (e.g., 10/20, not 50%). |
| | | Present summary data and appropriate descriptive and interferential statistics in sufficient detail to permit alternative analyses and replication. |
| | | Describe prognostic variables by treatment group and any attempt to adjust. |
| | | Describe protocol deviations. |
| | Discussion | State specific interpretations of study findings, including sources of bias and imprecision (internal validity) and discussion of external validity, including appropriate quantitative measures when possible. |
| | | State general interpretation of the data in light of the available evidence. |

*Source*: Modified from Machin et al. [21], with permission.

In spite of application of well-defined inclusion/exclusion criteria, it is commonly thought that about half of all clinical trials do not achieve their planned sample size [30]. As a recruitment strategy, necessary sample size is calculated to cover the expected number of dropout cases. It seems to me that the calculated number of the sample size is just for reference as an appeal to the cautious attitude of the researcher when conducting the trial.

**TABLE 2 Necessary Steps to Conduct Trials in Musculoskeletal Disorders**

| | |
|---|---|
| Developing protocol | Aim and goals with using strategies |
| | Simple pragmatic design |
| | Minimal amount of data correction |
| | Random allocation |
| | Ethical considerations |
| Patients | Clear inclusion/exclusion criteria |
| | Minimal burden for patients |
| | Recruitment strategies |
| Interventions | Drug, therapeutic exercise, physical agents, orthosis, Education or care package |
| Outcomes | Assessment of functions and QOL, good questionnaire |
| | Optimal combining outcome measures |
| | Data management procedures |
| Time constrained | Limited core activity |
| Staff and their training | Minimal demand |
| | Sufficient preparation and supporting |

**TABLE 3 Developing a Protocol for Clinical Trial**

| |
|---|
| Title |
| Summary |
| Background and rationale for the trials |
| Hypothesis to be tested |
| Primary outcome(s) |
| Secondary outcomes |
| Inclusion and exclusion criteria |
| Interventions to be tested |
| Estimated sample size |
| Information for patients and consent |
| Analyses plan, concluding dummy table |
| How patients will be entered into the study, concealment of allocation |
| Duration and methods for follow-up |
| Data collection, including questionnaires |
| Trial management |
| Trial supervision |
| Publication policy |
| Reference |

*Source*: Modified from Duley and Farrell [23], with permission.

## 5.3 Interventions

Treatments in musculoskeletal disorders, apart from surgical interventions, are classified into "drug therapy," "therapeutic exercise," "physical agents," "orthotics and devices," "education or care packages," and "others." When these intervention methods are indicated in the experimental group, it is important how comparative methods are selected in the control group. It is impossible to use the double-blind technique with placebo in musculoskeletal disorders because the comparative control is obviously different from the experimental content for the trial participants. The patients already know which treatments have been allocated and only

the assessors are unaware of the allocation results. This "open-label method" is often indicated in this field.

### 5.4 Outcomes

The content of outcomes used in clinical trials should be important to patients, usually the five D's: death, disease in clinical course, discomfort in symptoms, disability in activities of daily living (ADL), and dissatisfaction in QOL. There are many de facto standards for outcome measures in musculoskeletal disorders (Table 4).

Another important point in the use of outcome measures is the timing of application. Follow-up periods for outcome measures are categorized as short term (less than 6 weeks), intermediate term (6 weeks to 1 year), or long term (more than 1 year follow-up).

In an RCT, the results basically underestimate the difference between the comparative groups because the analysis is conducted with a more conservative standpoint for assessment (an intention-to-treat principle), and that will tend to dilute the estimate of the true difference. The content of interventions often fluctuates depending upon its nature, application timing, and patients' adherence. When the patients receive the strict treatment, they can expect more obvious efficacy than that reported.

### 5.5 How to Choose Right Outcome Measures

As an example, there is a quick-reference book that summarizes and evaluates more than 150 outcome measures for each joint of the extremities [39]. As Suk et al.'s handbook does not include the outcomes for spinal disorders, we have to have proper outcome measures for those, including neck and back problems [40, 41] (Table 4).

The necessary information for selecting outcome measures is as follows:

1. Goal of measurement
2. Nature of measurement: questionnaire, performance rating, physical properties
3. Specific population for which the instrument was developed
4. Format of measurement: number of items, response options, minimum and maximum score
5. Issues related to feasibility: time needed to perform the measurement, required equipment, and training

For clinical researchers, outcome measurements are essential for the advancement of their studies. To assess the overall quality of an outcome measure, three major elements should be considered: content of construct, psychometric evaluation, and clinical utility.

The conventional methods of studying the dimensional structure of measures are principal component analysis and factor analysis; Cronbach's $\alpha$ is calculated to determine the internal consistency of the dimensions. Outcome measures usually consist of one or more domains that reflect the concept of supposed construct.

**TABLE 4 Widely Used or Highly Qualified Outcome Measures for Musculoskeletal Disorders**

| | |
|---|---|
| • Low back pain | |
| Oswestry low back pain disability questionnaire [31] | |
| Roland–Morris disability questionnaire (RDQ) [32, 33] | |
| • Rheumatoid arthritis | |
| Health assessment questionnaire (HAQ) [34] | |
| Arthritis impact measurement scale (AIMS) [35–37] | |
| • Osteoarthritis (hip and knee) | |
| Western Ontario McMaster University (WOMAC) [38] | |
| (In this section of joint functions, the outcome measures that scored higher than 8 are shown) | |
| • Joint functions [39] | |
| Shoulder | Disabilities of the arm, shoulder and hand (DASH) |
| | Flexilevel scale of shoulder function (FLEX-SF) |
| | Oxford shoulder score |
| | Shoulder instability questionnaire |
| | Shoulder pain and disability index (SPADI) |
| | Simple shoulder test (SST) |
| | Upper extremity function scale |
| Elbow | Elbow functional assessment scale (EFA) |
| | Liverpool elbow score |
| | Upper extremity function scale |
| Wrist/hand | Boston questionnaire (also known as Brigham and Women's carpal tunnel questionnaire) |
| | Cochin rheumatoid hand disability scale |
| | Patients rated wrist evaluation (PRWE) |
| | Sequential occupational dexterity assessment (SODA) |
| | Upper extremity function scale |
| Pelvis | — |
| Hip | AAOS hip and knee score |
| | Functional recovery score |
| | Harris hip score |
| | Oxford hip score |
| | Western Ontario and McMaster Universities OA index (WOMAC) |
| Knee | AAOS hip and knee score |
| | Activity rating scale |
| | Fulkerson–Shea patellofemoral joint evaluation score |
| | Knee outcome survey activities of daily living scale |
| | Knee injury and osteoarthritis outcome score (KOOS) |
| | Kujala patellofemoral score (also know as the AKPS—anterior knee pain scale) |
| | Oxford 12-item knee questionnaire |
| | Western Ontario and McMaster Universities OA index (WOMAC) |
| Ankle | Foot health status questionnaire |
| Calcaneus | — |

It is very important to review, integrate, and consolidate the information with measuring instruments for cross-cultural use [6].

## 5.6 Questionnaires

The clinical utility of an outcome measure is divided into two parts: patient friendliness (acceptability) and clinician friendliness (feasibility) [42].

If we want a questionnaire to be user friendly, for both patient and clinician, and to improve clinical utility, the questionnaire has to be refined repeatedly. The content of the questionnaire must be limited to the minimum amount of information needed and be designed to be filled out easily. Even in the case of multiple choice or fill-in-the-numeral forms, there are some problems involved in interpreting the data.

Recently, the content of questionnaires that has only been determined by so-called expert staff is no longer regarded as sufficient for assessing content validity. We have to include the opinions of patients during the development process of the questionnaires in order to check content validity and face validity (these are often used interchangeably).

## 6 DATA ANALYSIS AND INTERPRETATION

### 6.1 Sample Size

The sample size of a trial is calculated based on the estimate of primary outcome [43]. I would like to show an example of how to calculate a sample size from the difference between the two means.

We have to recognize a way of denying null hypothesis, which is always associated with false-positive and false-negative rates. The former (false positive) is known as the type I error, that is, significance level $\alpha$. The latter (false negative) is the type II error $\beta$, and $1 - \beta$ is the power. The sample size is given here in the condition of two groups' comparison, and randomization in equal group numbers.

$$\text{Sample size } N = \frac{4(Z\alpha + Z\beta)^2}{\Delta^2}$$

where $Z\alpha$, $Z\beta$ are standardized normal deviates for given $\alpha$, $\beta$, and $\Delta$ is standardized effect size.

$Z\alpha = Z_{1-\alpha/2}$; $\alpha = 0.05$ in two-sided, then $Z_{1-\alpha/2} = Z_{0.975} = 1.96$

$Z\beta = Z_{1-\beta}$; a power of $1 - \beta = 0.9$, then $Z_{1-\beta} = Z_{0.9} = 1.645$

$\Delta = \mu A - \mu B/\sigma$; $(\mu A - \mu B)$ is difference between the two means, and $\sigma$ is standard deviation of the endpoint

If you set $\Delta$ less than 0.1, approximately 0.5, or more than 1, it means a small, moderate, or large standardized effect, respectively.

Sample size is noted in the majority of articles on clinical trials. Sample size depends on whether we want much more power to test the difference among the groups or we want to estimate the precision of the confidence interval. If we have an available working hypothesis, we can calculate the necessary sample size according to other similar studies. The result of the calculations only indicates the number required for the data analysis. Usually, we have to add a few more cases, expecting that there will be some dropouts.

### 6.2 Intention to Treat

Intention to treat (ITT) is a conceptual principle and is not defined as a specific technical procedure. One widely used procedure is to review the data in detail when the treatment and follow-up have been completed and all the patient information has been collected, that is, at the time that all the data has been frozen. Once allocation has been done, it is the ITT principle to analyze as if the original allocation was continued, even after the intervention itself was changed. The "last observation carry-forward" method is able to cover the deficit data when the trial has been in progress for some time (more than two check points).

However, even such principles cannot be maintained at analysis in the following cases:

1. Criteria unfit for inclusion, even after allocation
2. Complete loss to follow-up

### 6.3 Multiple Comparisons

We must cut down the working hypotheses to one main one. In a usual trial comparing two groups, a statistical test provides a figure of $P$ value as the boundary line for the verdict of hypothesis. If the calculated $P$ value for the primary endpoint is less than the predefined line (usually $P < 0.05$), we will deny the null hypothesis of no difference between the groups. However, if we use more than one endpoint, there are more than two comparisons arising (sometimes this situation becomes much more complex with multiple comparisons). In such a situation, a false-positive rate is no longer $\alpha \times 100\%$ and increases according to the number of multiple comparisons.

### 6.4 Post Hoc Analysis and Subgroup Analysis

It is the nature of things that the therapy-responder group has a better prognosis and the nonresponder group has a worse result. That is a typical post hoc analysis. Grouping allocation should be provided at the beginning and after disclosing the result, retrograde grouping is not permitted.

At times, clinicians like to know whether certain types of patients show more obvious benefits from interventions than others, knowing that a sample group of interest is a mixture of various types of patients. This is the subgroup analysis, for example, mild or severe cases, young or old patients. Some subgroups could easily reveal positive effects by chance, even if the overall trial results are negative. A common mistake is to mix the $P$ value of more than 0.05 in all data-combined analysis and less than 0.05 in one or more subgroup(s). This may lead to the false conclusion that a certain subgroup has a favored result by chance, in spite of the fact that the true result is negative. When you plan a subgroup analysis at the design stage, advance adjustment in sample size should be considered.

### 6.5 Data Synthesis and Combined Result

If there are several numbers of articles available for review, systematic review, is regarded as an "infrastructure" of the information system supporting EBM practice.

In regards to treatment, the pathway of systematic review is very logical; so the results provide the most accurate and authoritative guidelines for therapy. To avoid potential problems in interventional studies (clinical trials), only RCTs should be included, with complete follow-up information, blinded outcome assessment, and analysis based on intention to treat. This is applicable to an epidemiological research in observational studies, but it is still unable to solve all the problems of biases.

Meta-analysis is defined as a statistical synthesis of the numerical results of some trials (quantitative systematic review) that all examined the same question. It is a type of research that attempts to reanalyze and combine the results already reported, mainly as RCTs.

The assessment of methodological quality for meta-analysis has been recognized to be very important because even the conduct of an RCT is no guarantee of unbiased outcomes from such a study. Even meta-analysis and systematic reviews should be scrutinized carefully and analyses based on small studies should especially be treated with caution.

## 6.6 Side Effects

We have to pay sufficient attention to avoid adverse effects. In the case of surgery, examples of adverse side effects are nerve injury, infection, bleeding, secondary osteoarthritis, and even death. Some side effects occur even in conservative therapy.

We cannot say that treatment side effects are out of the question, even if there is no obvious difference in the frequency and degree of adverse effects between the treatment group and the control group in a certain RCT. An RCT is not an appropriate design to assess side effects because an RCT basically suppresses the number of patients to a minimum and the incidence of side effects is usually low. The role of the trial steering committee and the data-monitoring committee should be established and the range of their responsibilities defined. These committees have to take necessary actions such as early stopping, considering the results of the interim analysis, or the frequency and content of side effects. I realize that relatively small size trials do not often have the side effects and other problems that would stop a trial, but large multicentered trials have the latent possibility of such troubles.

## 6.7 Reporting Clinical Trials

After completing a clinical trial, an important step is to report the results, whether positive, negative, or equivocal. Selective reporting, in which only positive studies are published, often distorts a true situation; this practice is called "publication bias." When meta-analysis is performed as a clinical trial overview, it is definitely important to include all relevant unpublished trials to gain overall results.

The International Committee of Medical Journal Editors, which consists of several leading medical core journals, decided that clinical trials should be registered in advance [44, 45]. If clinical trials are not registered in advance, journals might refuse to publish a report of a trial. This is an important step in avoiding publication bias and other inappropriate analyses.

Article writing is regarded as a form of communication or "dialog," even in the case of a scientific journal; that is the object of our communication. The writer of an article must be aware of the target audience. Consolidated Standards of

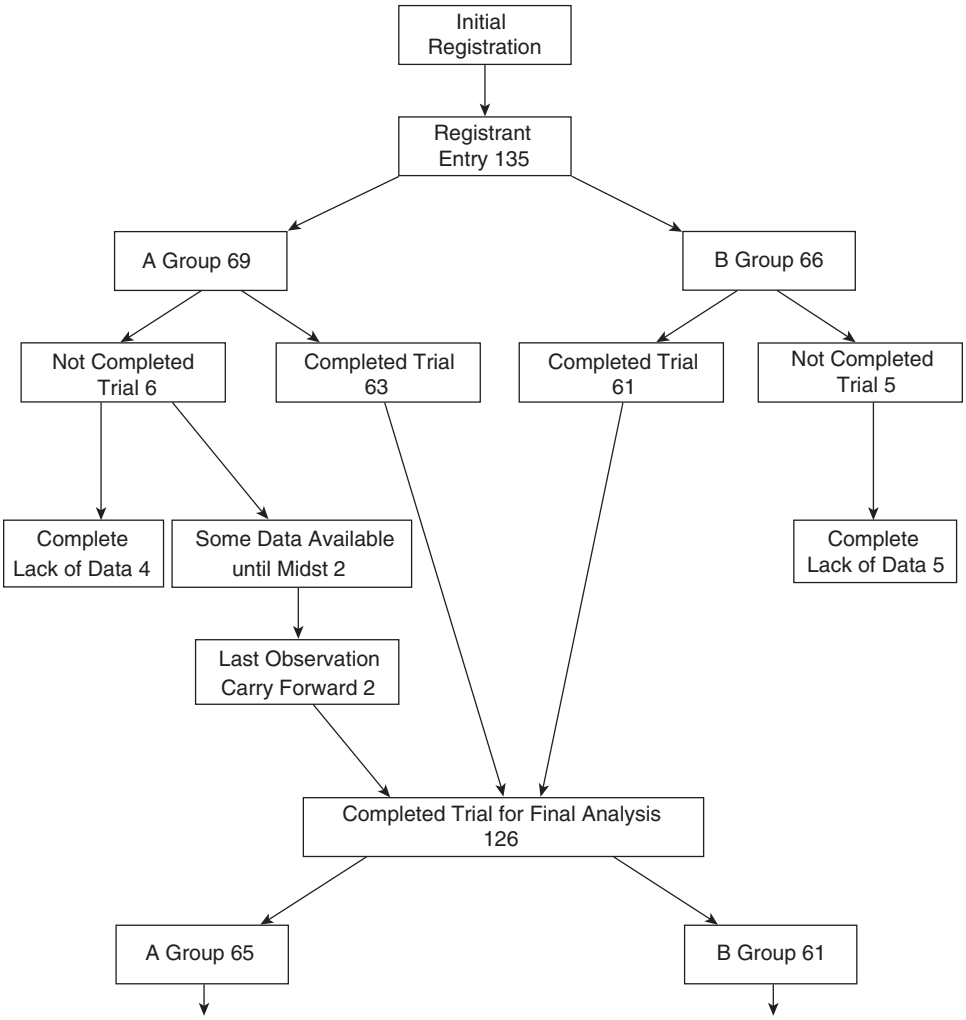


***FIGURE 1*** *Example flowchart of trial according to CONSORT statement.*

Reporting Trials (CONSORT) is one of the examples of such a checklist or flowchart to be used, as well as the instructions for improving the quality of reporting randomized controlled trials when a paper is to be submitted [46, 47] (Fig. 1). Also suggested through discussions at a conference was a guideline for reporting systematic reviews or meta-analysis referred to as the Quality of Reporting of Meta-analysis (QUOROM). The discussions resulted in the creation of the QUOROM Statement, which consists of a checklist and flowchart [48].

## 7 RCT AND FUTURE DIRECTIONS

At the close of this section, I would like to talk about how to apply the results of an RCT to daily practice, as well as mention the potential problems when reaching the limit of a statistical outlook.

### 7.1 Impact from RCT to Daily Practice

Medical practice is a life-long, continuous process of self-learning, and, as clinicians, we are required to keep up-to-date on various medical developments. Evidence-based medicine is our way of integrating individual clinical expertise with the best available evidence to assist us in making decisions about each patient's care. Reported results of RCTs make it possible to cover the majority of our activities systematically, from daily practice for patient care to writing and reading scientific papers.

Evidence-based medicine is regarded as a new paradigm in medical practice, equal to the Human Genome Project. However, there is still much confusion and misunderstanding about the concept and content of EBM; it is often limited to searching the literature and reading articles, and to serving cost cutters and suppressing clinical freedom. The use of clinical guidelines or the managed care system is seen as intimidating the discretion of doctors in clinical practice. EBM is also seen as the fashionable trend of a group of medical academics armed with epidemiological and statistical jargon, but there are more serious basic problems behind such a statistical paradigm.

### 7.2 Internal Validity and External Validity

In clinical practice we are always surrounded by a condition called the "gray zone" [49]. Of course, there are many factors involved in this obscurity. I would like to focus closer attention on a statistical outlook of the world around us, that is, internal validity and external validity.

These are concepts relating to the extrapolation of the relationship between sample and source populations. In an experimental trial like an RCT, internal validity is high because the difference between the experimental group and the control group is thoroughly controlled only for the target issue, and by chance. However, external validity is low when the results from the sample with several specific premises are brought into a general finding in the mother population. On the contrary, in an observational study, such as a cohort study, there are many confounding factors involved and internal validity is low in surveyed samples. External validity in an observational study is high because a real situation in a resource population reflects just what it is (Fig. 2).

### 7.3 Quantitative Study and Qualitative Study

Clinical trials cover a wide range of medical fields and use very diverse methods. The data used is basically divided into quantitative, such as laboratory data, and qualitative, based on verbal information collected (mainly used in psychology or nursing science at present). Behind the methods of investigation and analysis in these quantitative and qualitative approaches, there is an obvious difference in concepts [50].

The general concept for study design consists of the following characteristics (Table 5):

1. Aims of study
2. Methods of outcome measurements

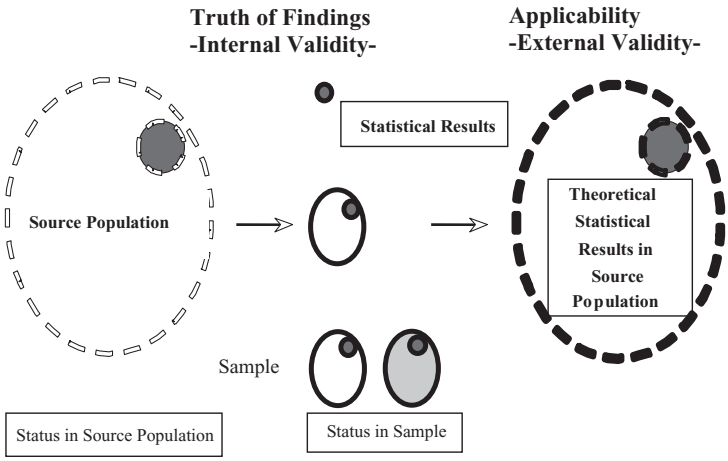


**Observational Study**

| | |
|---|---|
| Cross-sectional Study or Cohort Study | Internal Validity: low<br>External Validity: high |

**Interventional Study**

| | |
|---|---|
| Randomized Controlled Trial Blind & Placebo-controlled | Internal Validity: high<br>External Validity: low |

***FIGURE 2*** *Relationship between sample and mother population.*

**TABLE 5 Comparison Between Quantitative Research and Qualitative Research**

| | Quantitative Research | Qualitative Research |
|---|---|---|
| Study design | Intervention | Observation |
| Methods | Survey, experiment | Interview |
| Question setting | Enumeration | Classification |
| Reasoning | Deductive | Inductive |
| Sampling method | Statistical model | Theoretical model |
| Closeness to the truth | Internal validity | Credibility |
| Applicability/generalizability | External validity | Concept transferability |
| Consistency | Reliability | Dependability |
| Neutrality | Objectivity | "Confirmability" |

3. Selection of subjects
4. Analytic methods and provided results

In a conventional quantitative study, the main strategy is to make up a working hypothesis and test it. An RCT is a typical example of this type of strategy and is inseparable from biostatistics for data analysis. To generalize the results from samples, such as the rejection of null hypothesis in the majority of confirmatory studies, requires estimating and testing as a basic procedure. The advance of computer technology makes possible the use of multivariate analysis and the ability to search and verify potential factorial structure, even without a working hypothesis.

This technique makes it easy to analyze latent factorial structure, which occurs in some observational studies, from a calculated relationship of how the weight distribution among explanatory variables accounts for the tendency of objective variables. Therefore, this exploratory study makes it possible to investigate and verify risk factors or to produce a model of a certain phenomenon. The distinction between the term "quantitative" study and "qualitative" study will be more obscure once the quantitative studies move toward an exploratory nature from a confirmatory one for testing hypothesis.

On the contrary, a qualitative study does not always follow the deductive logic of applying the result from a sample to the mother population. Researchers try to categorize the verbal information received from subjects to establish a coding system for the data and generalize a new concept in order to build a new theory or model. It is good enough to select the subjects that fit the aim of a study as one likes. It would be possible to produce an epochal qualitative study that could produce a unique, clinically significant theory that would never have been yielded by a confirmatory study using hypotheses testing. A qualitative study is able to make progress in building up a new and significant theoretical system [51]. However, good or bad, theory is totally dependent upon the capability of the researcher.

### 7.4 Further Suggestions

If a certain intervention shows obviously successful results for an otherwise lethal condition, we do not require RCTs and do not wait for more studies to be conducted. We can get the evidence from the basic sciences, properly designed follow-up studies, or proper cross-sectional trials.

We need to be careful not to become so skeptical as to assume that if a study is not an RCT it would have no value or use. We also need to recognize the limitations of a statistical approach through RCTs. At each step, we continue our efforts to track down the best evidence to answer our clinical questions and promote cost-effective prevention and treatment for musculoskeletal disorders.

## REFERENCES

1. Akai, M. (2002), Evidence-based medicine for orthopedic practice, *J. Orthop. Sci.*, 7, 731–742.
2. Akai, M., Doi, T., Fujino, K., et al. (2005), An outcome measure for Japanese people with knee osteoarthritis, *J. Rheumatol.*, 32, 1524–1532.
3. BJD on line (Bone and Joint Decade's musculoskeletal Portal); http://www.boneandjointdecade.org/.
4. McDowell, I., and Newell, C. (1996), *The Theoretical and Technical Foundations of Health Measurement. Measuring Health; A Guide to Rating Scales and Questionnaires*, 2nd ed., Oxford University Press, New York, pp. 10–46.
5. Guyatt, G. H., Kirshner, B., and Jaeschke, R. (1992), Measuring health status; What are the necessary measurement properties? *J. Clin. Epidemiol.*, 45, 1341–1345.
6. Dekker, J., Dallmeijer, A. J., and Lankhorst, G. J. (2005), Clinimetrics in rehabilitation medicine; current issues in developing and applying measurement instruments, *J. Rehabil. Med.*, 37, 193–201.

7. Wolfson, A. M., Doctor, J. N., and Burns, S. P. (2000), Clinician judgments of functional outcomes; how bias and perceived accuracy affect rating, *Arch. Phys. Med. Rehabil.*, 81, 1567–1574.
8. Guyatt, G. H., Feeny, D. H., and Patrick, D.L. (1993), Measuring health-related quality of life, *Ann. Int. Med.*, 118, 622–629.
9. Clancy, C. M., and Eisenberg, J. M. (1998), Outcomes research; measuring the end results of health care, *Science*, 282, 245–246.
10. World Health Organization (2001), *International Classification of Functioning, Disability and Health*, WHO, Geneva.
11. Stucki, G., and Grimby, G. (2004), Applying the ICF in medicine, *J. Rehabil. Med.*, 36 (suppl 44), 5–7.
12. Irrgang, J. J., and Anderson, A. F. (2002), Development and validation of health-related quality of life measures for the knee, *Clin. Orthop.*, 402, 95–109.
13. Zelen, M. (1979), A new design for randomized clinical trials, *N. Engl. J. Med.*, 300, 1242–1245.
14. Zelen, M. (1990), A new design for randomized consent designs for clinical trials; an update, *Stat. Med.*, 9, 645–656.
15. ter Riet, G., Kleijnen, J., and Knipschild, P. (1990), Acupuncture and chronic pain; A criteria-based meta-analysis, *J. Clin. Epidemiol.*, 43, 1191–1199.
16. Koes, B. W., Bouter, L. M., Beckerman, H., et al. (1991), Physiotherapy exercises and back pain; A blinded review, *BMJ*, 302, 1572–1576.
17. Koes, B. W., Assendelft, W. J. J., van der Heijden, G. J. M. G., et al. (1991), Spinal manipulation and mobilisation for back and neck pain; A blinded review, *BMJ*, 303, 1298–1303.
18. Koes, B. W., Bouter, L. M., and van der Heijden, G. J. M. G. (1995), Methodological quality of randomized clinical trials on treatment efficacy in low back pain, *Spine*, 20, 228–235.
19. Akai, M., and Hayashi, K. (2002), Effect of electrical stimulation on musculoskeletal systems; A meta-analysis of controlled clinical trials, *Bioelectromagnetics*, 23, 132–143.
20. Akai, M., Kawashima, N., Kimura, T., et al. (2002), Electrical stimulation as an adjunct to spinal fusions; A meta-analysis of clinical controlled trials, *Bioelectromagnetics*, 23, 496–504.
21. Machin, D., Day, S., and Green, S. Eds. (2004), *Textbook of Clinical Trials*, Wiley, West Sussex, UK.
22. Bowling, A. (2001), *Measuring Disease; A Review of Disease-Specific Quality of Life Measurement Scales*, 2nd ed., Open University Press, Buckingham, UK.
23. Duley, L., and Farrell, B. Eds. (2002), *Clinical Trials*, BMJ Books, London.
24. McDowell, I., and Newell, C. (1996), *The Theoretical and Technical Foundations of Health Measurement. Measuring Health; A Guide to Rating Scales and Questionnaires*, 2nd ed., Oxford University Press, New York.
25. Riegelman, R. K. (2005), *Studying a Study and Testing a Test; How to Read the Medical Evidence*, 5th ed., Lippincott Williams and Wilkins, Philadelphia.
26. Suk, M., Hanson, B. P., Norvell, D. C., et al. (2005), *AO Handbook; Musculoskeletal Outcomes Measures and Instruments; 155 Instruments Evaluated and Assessed*, AO Publishers, Thieme.
27. Resource Center for Randomized Trials; http://www.rcrt.ox.ac.uk/.
28. An introduction to clinical trials; http://www.clinicaltrials.gov/.
29. Farrell, B., and Spark, P. (2002), Building resources for randomized trials, in Duley, L., and Farrell, B., Eds., *Clinical Trials*, BMJ Books, London, p. 84.

30. Fayers, P. M., and Machin, D. (1995), Sample size; how many patients are necessary? (editorial review), *Br. J. Cancer*, 72, 1–9.
31. Fairbank, J. C. T., Couper, J., Davies, J. B., et al. (1980), The Oswestry low back pain disability questionnaire, *Physiotherapy*, 66, 271–273.
32. Roland, M., and Morris, A. (1983), A study of the natural history of back pain. Part II. Development of a reliable and sensitive measure of disability in low-back pain, *Spine*, 8, 141–144.
33. Roland, M., and Fairbank, J. (2000), The Roland-Morris disability questionnaire and the Oswestry disability questionnaire, *Spine*, 25, 3115–3124.
34. Fries, J. F., Spitz, P., Kraines, R. G., et al. (1980), Measurement of patient outcome in arthritis, *Arthritis Rheum.*, 23, 137–145.
35. Meenan, R. F., Gertman, P. M., and Mason, J. H. (1980), Measuring health status in arthritis; The arthritis impact measurement scales, *Arthritis Rheum.*, 23, 146–152.
36. Meenan, R. F., Gertman, P. M., Mason, J. H., et al. (1982), The arthritis impact measurement scales; Further investigations of a health status measure, *Arthritis Rheum.*, 25, 1048–1053.
37. Meenan, R. F., Mason, J. H., Anderson, J. J., et al. (1992), AIMS2; The content and properties of a revised and expanded arthritis impact measurement scales health status questionnaire, *Arthritis Rheum.*, 35, 1–10.
38. Bellamy, N., Buchanan, W. W., Goldsmith, C. H., et al. (1988), Validation study of WOMAC; A health status instrument for measuring clinically important patient relevant outcomes to antirheumatic drug therapy in patients with osteoarthritis of the hip or knee, *J. Rheumatol.* 15, 1833–1840.
39. Suk, M., Hanson, B. P., Norvell, D. C., et al. (2005), *AO Handbook; Musculoskeletal Outcomes Measures and Instruments; 155 Instruments Evaluated and Assessed*, AO Publishers, Thieme, pp. 37–40, 47–405.
40. Bombardier, C. (2000), Outcome assessments in the evaluation of treatment of spinal disorders; summary and general recommendations, *Spine*, 25, 3100–3103.
41. Stratford, P. W., Binkley, J., Solomon, P., et al. (1994), Assessing change over time in patients with low back pain, *Phys. Ther.*, 74, 528–533.
42. Suk, M., Hanson, B. P., Norvell, D. C., et al. (2005), *AO Handbook; Musculoskeletal Outcomes Measures and Instruments; 155 Instruments Evaluated and Assessed*, AO Publishers, Thieme, pp. 20–24.
43. Machin, D., Campbell, M. J., Fayers, P. M., et al. (1997), *Sample Size Tables for Clinical Studies*, 2nd ed., Blackwell Science, Oxford, pp. 40–78.
44. DeAngelis, C. D., Drazen, J. M., Frizelle, F. A., et al. (2004), Clinical trial registration; a statement from the international committee of medical journal editors. *JAMA*, 292, 1363–1364.
45. DeAngelis, C. D., Drazen, J. M., Frizelle, F. A., et al. (2005), Is this clinical trial fully registered? A statement from the international committee of medical journal editors, *JAMA*, 293, 2927–2929.
46. Begg, C., Cho, M., Eastwood, S., et al. (1996), Improving the quality of reporting of randomized controlled trials; The CONSORT statement, *JAMA*, 276, 637–639.
47. Moher, D., Schulz, K. F., Altman, D. G., for the CONSORT group (2001), The CONSORT statement; revised recommendations for improving the quality of reports of parallel-group randomized trials, *Lancet*, 357, 191–194.
48. Moher, D., Cook, D. J., Eastwood, S., et al. (1999), Improving the quality of reporting of meta-analyses of randomised controlled trials; The QUOROM statement, *Lancet*, 354, 1896–1900.

49. Naylor, C. D. (1995), Grey zones of clinical practice; Some limits to evidence-based medicine, *Lancet*, 345, 840–842.
50. Greenhalgh, T., and Taylor, R. (1997), Papers that go beyond numbers (qualitative research), *BMJ*, 315, 740–743.
51. Öhman, A. (2005), Qualitative methodology for rehabilitation research, *J. Rehabil. Med.*, 37, 273–280.

# *Oncology*

**Matjaz Zwitter**

*Institute of Oncology, Ljubljana, Slovenia, and Department of Medical Ethics, Medical School, University of Maribor, Slovenia*

**Contents**

1 Introduction
2 General Comments on Ethical Issues
  2.1 Ethical Demand for Research
  2.2 Ethical Obligation of Patients
  2.3 ALARA Principle and Positive Ethical Balance
  2.4 Patient with Cancer with Comments on Informed Consent and Real Patient Autonomy
3 Different Medical Problems in Clinical Trials
  3.1 Prevention of Cancer
  3.2 Induction (Neoadjuvant) Therapy
  3.3 Adjuvant Therapy
  3.4 Trials of Concomittant Radiochemotherapy
  3.5 Systemic Treatment of Advanced Cancer
  3.6 Palliative Treatment
4 Types of Clinical Trials
  4.1 Phase I Trials
  4.2 Phase II Trials
  4.3 Phase III Trials
5 Endpoints
  5.1 Response to Treatment
  5.2 Time to Progression
  5.3 Survival
  5.4 Quality of Life
6 Conclusion
References

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

## 1 INTRODUCTION

Patients with cancer have their diagnostics and treatment in different departments of virtually every hospital. Still, when it comes to research, most of it would be conducted either in specialized cancer centers or in oncology departments of larger hospitals. We can say that oncology as a distinct branch of medicine indeed developed due to clinical research. Without clinical research, oncology as a discipline would not exist. Without clinical research, our only guidelines would be statements like "I think ..." or "I remember two similar patients ...".

Many oncologists are sceptical in their attitude toward new developments. This is not unexpected. The nature of the disease itself with its unpredictable course and many false hopes do not support overt optimism at every new idea. Nevertheless, a healthy degree of scepticism should not be understood as pessimism. Looking from the perspective of one or two years, no dramatic progress may be apparent. Still, in the long run, the battle against cancer is a story of hundreds of small victories. Fifty years ago, only a tiny minority of cancer patients with truly limited disease could hope for a successful cure by surgery. For the rest, cancer stood as an invincible stone wall. Systematic clinical research led to erosion of this wall, piece by piece. Hematological malignancies and pediatric tumors, testicular cancer, breast, prostate, gastrointestinal, lung cancer—the prognosis of each of these tumors in terms of cure, prolonged survival, or quality of life has improved significantly. Thousands of clinical trials have contributed to this gradual progress.

Our understanding of how the treatment works and why it is effective in one patient but fails in another apparently similar case lags behind clinical experience. In the past, we used to apply treatments and hope for some 30% chance of remission, at the same time often ignoring the fact that the majority of patients did not benefit from the treatment and experienced only its side effects. In the past, we used to say that cancer was not one disease but about 200 different diseases. We then gave up hope for a single miracle drug against cancer and accepted the fact that each of these diseases demands a different treatment. In the future, we may need to go further in our individualized approach. Relatively common diseases such as breast cancer or non-small-cell lung cancer are now subdivided into separate entities, depending on phenotypic and genetic characteristics of tumor cells. Clinical research in oncology is turning into a truly team venture: pathology, molecular biology, genetics, and immunology provide essential information for individual classification of a tumor. So deep is the change in methodology that we can speak of a new era of clinical research in oncology [1]. Better understanding of tumor biology and of pharmacogenomics will lead to much more individualized assignment of treatment and, consequently, to more predictable outcome.

Final introductory remarks deal with professionalization of clinical research. What used to be an amateur activity several decades ago is now almost entirely in the hands of professionals supported by pharmaceutical companies or (rarely) public funds. Among clinicians, only a tiny minority of senior "opinion leaders" will be asked to participate in the design of clinical trials; for the rest, clinical research has become synonymous with recruiting patients and filling clinical record forms (CRFs). While increased funding from private funds and from for-profit companies led to important new discoveries, professionalization of clinical research has also

its dark side. It typically produces large multicenter trials focusing on common cancers where substantial profits are expected and often ignores rare tumors; yet, taken together, these rare tumors comprise more than half of all patients with cancer. Commercial interests often prevail over honest scientific evaluation of research proposals, and many promising initiatives and ideas remain unexplored. In the phase of designing a new trial, sponsors do not listen to creative young physicians. While we need large multicenter trials, we should also support simple, innovative, small trials not necessarily sponsored by industry. In a world where every new idea is tested and retested, size of a trial is of minor importance. What matters are innovative potential, reliability of research methodology, quality of the data, and fair reporting [2].

## 2 GENERAL COMMENTS ON ETHICAL ISSUES

In any branch of medicine, ethics is at the beginning and at the end of every discussion on clinical research. Ethical issues are crucial when planning a clinical trial, throughout its actual performance, and when formulating experience and implications for routine treatment and for further research. Oncology is not an exception to this general rule.

We here offer four general statements on ethics of clinical cancer research. Some specific comments and recommendations will be made also in other parts of the text.

### 2.1 Ethical Demand for Research

Considering the enormous burden imposed by cancer on every patient and on society as a whole, it would be unethical not to search for new effective and safer approaches to treatment. During the past five decades, clinical research led to an improved outlook for virtually every cancer. Systematic clinical research is not only ethically acceptable but indeed an obligation for everybody in the field of oncology.

It is estimated that only between 3 and 10% of cancer patients participate in clinical trials [3], with even lower representation of elderly patients and minorities [4]. Since it is clear that our progress in the field of oncology critically depends upon the proportion of patients enrolled in clinical research, low recruitment of patients into clinical trials is not only a problem for the researchers but indeed for society as a whole. While public understanding and support for medical research generally increases, refusal of an eligible patient to participate in a clinical trial is not uncommon [5]. Time-consuming procedures for actual performance of a clinical trial with physician's time as a diminishing resource, and the impact of greater demands in a climate of decreasing health care resources have a negative effect on accrual [6, 7]. To increase trial participation, there is a critical need for infrastructure to support trials, especially additional support staff and research nurses [8]. Other obstacles to clinical research include legal restrictions, as seen in the European Union Clinical Trials Directive [9] or in the Declaration of Helsinki [10]. These documents should be critically discussed and, hopefully, amended. Please see the excellent discussion on this matter by Harris [11].

### 2.2 Ethical Obligation of Patients

From the legal standpoint, a patient with cancer is absolutely free to accept or to refuse participation in clinical research. From the ethical standpoint, however, a patient has a moral duty to participate in a clinical trial that has passed an independent scientific and ethical review and that does not present an unproportionate burden or risk. This moral duty is based on the fact that patients of today expect to be treated according to the best current knowledge. This knowledge increases continuously and is derived from the past clinical research. Patients of today, therefore, benefit from the fact that patients of yesterday participated in clinical trials—and have a moral duty to contribute to the medical knowledge that will also benefit them, future generations of patients, and society as a whole [11].

### 2.3 ALARA Principle and Positive Ethical Balance

In the practice of clinical research, ethical puritanism is rarely achievable. In spite of meticulous plan and execution of a trial, some "ethical costs" are inevitable. Instead of speaking of zero ethical costs, we follow the philosophy of ALARA—*as low as reasonably achievable* [12]. We also speak about a positive balance between the benefits of a trial (as applicable to patients in the trial and to society as a whole) and ethical costs.

Practical possibilities for lowering ethical costs and for a positive balance between benefits and ethical costs in clinical research will be presented in Sections 4 and 5.

### 2.4 Patient with Cancer with Comments on Informed Consent and Real Patient Autonomy

Among patients with cancer, we can rarely speak about full autonomy. Patients are not only medically uneducated but also under serious physical and emotional burden of the disease.

In spite of this well-recognized burden of malignant disease, information offered to cancer patients during recruitment for clinical trials is often very extensive, written in a complicated medical and legal jargon, and with detailed description of all possible and rare complications of the treatment. Such extensive and frightening information does not add to patient's autonomy. Rather, many patients sign consent forms without an attempt read such an extensive text, leading to a diminished autonomy. According to a survey among physicians recruiting patients for randomized clinical trials, 42% of patients were not fully informed about the trial [13]. In another survey, the majority of patients who recently signed consent for participation in a trial did not recognize the nonstandard treatment nature of the trial (74%), the potential for incremental risk from participation (63%), or the unproven nature of the treatment (70%) [14]. An additional argument against extensive text of informed consent comes from Eastern Cooperative Oncology Group, which randomized patients between the standard consent procedure to an easy-to-read consent statement. A simplified text did not affect patient comprehension and resulted in significantly lower consent anxiety and higher satisfaction [15].

Even from a purly legal standpoint, it should not be too difficult to prove that many patients with cancer sign consent without being fully competent and/or do not have a free choice.

If we accept limited patient's autonomy as a reality, then we have to accept also the fact that many patients do not rely on rational considerations but simply trust their doctor. Thus, while we need patient's consent, the responsibility stays with the physician. Physician's paternalism, whether we agree with it or not, is often a reality. A physician who is also a clinical researcher should be aware of the delicacy of his or her role. He should be aware that his relation to the patient does not derive from a contract, but rather from a patient's trust. As he cannot get rid of a certain degree of paternalism, he has to strive constantly to push his role of a physician ahead of his role of a researcher.

In concluding this section, we recommend that the information for patient is adjusted to his or her real condition: brief, understandable, with realistic presentation of the problem and also offering some hope. While patients have the right to full information, they should not be excluded from research in case they opt for a simplified procedure. Rather than continuing with what is increasingly obvious to be a fiction—that fully informed consent has been obtained—we should perhaps move to a consent protocol that is basically simple consent in the presence of full information. A summary paragraph to each information sheet would allow the patients to focus on that alone if they choose to do so [16].

## 3 DIFFERENT MEDICAL PROBLEMS IN CLINICAL TRIALS

Understanding the medical problem is essential for choosing appropriate methodology for a clinical trial. We will briefly review some specific medical problems in oncology and offer comments on the research methodology.

### 3.1 Prevention of Cancer

Various food supplements, hormones, hormone antagonists, antiviral vaccines, other substances, and even surgical procedures have been studied with the aim of cancer prevention [17–21]. While a critical coverage of this area would be clearly out of the scope of our discussion, we here offer some specific recommendations.

- Before initiating a trial, its scientific background should be thoroughly discussed. A clear distinction should be made between the two possibilities: to delay or to prevent occurrence of a particular cancer. Cancerogenesis is a long process. From the experience on secondary cancers induced by radiation or anticancer drugs, we know that there is an interval of at least 5 years between the carcinogenic mutation and clinical detection of cancer. On short terms of a few years, we may expect only a delay of occurrence of a cancer that was already in its subclinical form. If we speak about real prevention, we cannot expect a detectable effect on clinical incidence of a particular cancer in less than 8–10 years.
- The population of persons at an increased risk for a particular cancer should be clearly defined. Even so, we are dealing with healthy persons and not with

patients. The intervention studied should impose a negligible risk and minimal discomfort to participants.

- The control groups are either "pure" controls or are given placebo. Compliance is often poor: Subjects in the intervention group may not follow the procedure, while subjects in the control group may get the substance (such as a food supplement) from other sources. Besides, it is unrealistic to expect good compliance in a trial that continues over many years. Suboptimal compliance leads to dilution of the results and demands a greater number of subjects involved.
- To prove that an intervention lowers the incidence of a relatively rare event, many thousands of individuals have to be followed over a long period.
- Trials of cancer prevention are commonly linked to a program for early detection of a particular cancer. Impact of the intervention upon the incidence of a particular cancer is the obvious objective of such a trial. Still, the real value of an intervention among healthy individuals would be reduction in mortality from any cause, coupled to uncompromised quality of life. As an example, antiestrogens given to healthy women at an increased risk for breast cancer led to reduction in the incidence of breast cancer; yet, most of the prevented tumors were highly curable hormone-dependent cancers. The incidence of the more malignant hormone-resistant cancers remained almost the same, and there was no reduction in mortality. Furthermore, an increase in endometrial cancer and more frequent thrombo-embolic events were observed; the impact of antiestrogens upon subtle differences in the quality of life of healthy women remains unknown. Therefore, incidence of a particular cancer should not be the principal objective of a trial and should be regarded as a temporary surrogate endpoint.

### 3.2 Induction (Neoadjuvant) Therapy

Induction systemic therapy prior to definite local treatment with surgery and/or radiotherapy is a very promising approach that is increasingly used in apparently localized tumors [22–28]. Compared to adjuvant (postoperative) systemic treatment, induction therapy has several advantages: Treatment of micrometastatic disease begins immediately; the tumor may shrink and become operable in a higher proportion of patients; clinical, radiologic, and pathologic evaluation of response to systemic therapy will be important for further treatment strategy. The potential dangers of induction therapy in operable disease are resistance to chemotherapy, in which case a tumor may actually grow during induction treatment and become inoperable, and the risk of an increase in the incidence of perioperative complications. Prior to randomized trials, solid data from phase II trials should provide a reliable estimate of these risks.

It is essential that induction therapy and local treatment with surgery and/or radiotherapy are part of the same treatment protocol with precisely defined timing and as short a gap as possible between systemic and local treament. Tumors that shrink during induction therapy will often regrow rapidly. If the interval between induction therapy and local treatment is too long, accelerated repopulation of the tumor may quickly annihilate all the benefits of combined treatment [29].

### 3.3 Adjuvant Therapy

Patients at an increased risk for relapse after local treatment are often given adjuvant chemotherapy, aimed at reducing this risk. Such a strategy has been widely accepted in breast cancer [30], lung cancer [23], and several gastrointestinal cancers [22]. Many new trials are in progress to improve the ratio between efficacy and side effects for the existing indications and to study this approach for new indications.

A substantial proportion of patients offered adjuvant treatment might be cured of cancer with local treatment alone. As an example, a recent trial among 1146 patients with early breast cancer has shown a 65% 10-year disease-free survival after surgery and adjuvant chemotherapy, compared to 60% after surgery alone ($p < 0.01$) [31]. This experience has been presented as evidence for the indication of adjuvant therapy. Yet, these figures can be interpreted also from a different angle. We can say that 60% of patients were cured by surgery alone; that 35% had a recurrence in spite of adjuvant therapy; and that only 5% actually benefited from adjuvant treatment. It is clear, then, that the benefit of a prolonged disease-free interval and of improved survival should be weighted against side effects of the adjuvant therapy, and it is therefore essential that studies of adjuvant therapy include assessment of quality of life.

Planning a trial of adjuvant therapy has several elements in common with trials of cancer prevention. As there is no possibility for individual assessment of response, we have to enroll a large population of healthy individuals or patients after successful local treatment, and we ususally measure the success with time to event (diagnosis of cancer in the case of prevention trials, time to progression in adjuvant treatment). In both instances, we should understand that time to event is only a surrogate endpoint: It is only an improvement in overall survival that validates the results. While some cancers or some recurrences of cancer that we wish to prevent are still curable, it is also important that the prognosis of what we wish to prevent is generally dismal. If this were not the case, it would make no sense to engage in prevention of a highly curable situation.

### 3.4 Trials of Concomitant Radiochemotherapy

In this section, we are dealing with two distinct categories of patients. In the first are patients with trully localized cancer such as early laryngeal or cervical cancer. A high proportion of these patients are cured either by surgery, by surgery and postoperative radiotherapy, by radiotherapy alone, or by combined irradiation and chemotherapy. When the prospects with each particular treatment are good, the decision regarding the treatment depends on patient's preferences, on side effects of each of these treatments, and on the expertise and influence of particular oncologists. The second category are patients with inoperable locally advanced disease without overt distant metastases for whom surgery is not an option.

During the past decade, there was an explosion of trials of concomittant radiochemotherapy. When nonsurgical treatment with curative intent is applied, head and neck cancers, lung, esophageal, gastric, colorectal, pancreatic, cervical and bladder carcinomas, many brain tumors, sarcomas, and most pediatric tumors are now all treated with concomitant radiotherapy and chemotherapy [32–41].

While a certain degree of uncertainty regarding the true advantage of combined-modality treatment persists, we can be confident that the trend of combined-

modality treatment will continue. Irradiation, probably to higher total doses and possibly with different schedules of fractionation will be combined with new drugs, all with the aim of achieving a better local and systemic control of the disease at the expense of acceptable toxicity. It is more than clear that both the effect against the tumor and the pattern of acute and late toxicity of combined-modality treatment are unpredictable. "Wishful thinking" should be avoided, and well-designed phase I and II clinical trials are needed [42]. In this section, we will present a selection of recent accomplishments and of basic knowledge important for the design of combined-modality trials.

Radiotherapy of today is virtually incomparable to what this treatment used to be just two decades ago. Modern technology has allowed very precise targeting of the tumor and of critical healthy tissues. Since the coverage of the tumor by radiation is much more precise and the margin of healthy tissues is getting smaller, we see fewer side effects along with higher efficacy of radiotherapy in eradication of a tumor. Instances of "individual hypersensitivity to radiation" (which in fact often represented cases with grossly inprecise dosimetry) are now rare and treatment with radiotherapy can proceed to higher total doses with predictable and acceptable toxicity.

The technological advances of radiotherapy are important not only for trials of radiation alone but also for the design of combined-modality trials. It is extremely important to understand that radiotherapy with the total dose and the fractionation that used to be standard radical radiotherapy one or two decades ago is now a suboptimal treatment. As an example, 60 Gy in 30 fractions over 6 weeks used to be a standard for radiotherapy of lung cancer in the 1980s and early 1990s; today, the optimal dose is from 64 to 74 Gy, and in some trials even higher [43].

A reader familiar with radiobiology may skip this paragraph. For the rest who are just interested in combining radiation with drugs, it is essential to understand the importance of fractionation upon tumor control and upon the pattern of acute and late side effects of irradiation. Regarding *tumor control*, the classical and still valid advice is: "Give as many fractions as possible in as short a total time as possible." Hyperfractionated irradiation and hyperfractionated accelerated irradiation with more than one small fraction per day were tested and proven to be superior to standard radiotherapy for head and neck and lung cancers [44, 45]. Hyperfractionation produced more *acute side effects*, which were usually transient and rarely extended into chronic toxicity. On the other side of the spectrum of fractionation, we have hypofractionation with fewer large fractions of radiotherapy. While this approach is useful for palliation where the total dose is low, hypofractionation cannot be recommended for treatment with curative intent. Severe *late and permanent toxicity* was observed in trials of hypofractionated radiotherapy.

It is clear that even at very high doses, radiotherapy alone cannot cure many tumors. The concept of synergistic effect of radiotherapy and chemotherapy has been tested in vitro and confirmed in dozens of clinical studies. It was shown that most cytotoxic drugs enhance the effects of radiotherapy [46]. The broad term "radiosensitization" encompasses several phenomena: additive effect (most cytotoxic drugs) [47], true radiosensitization (gemcitabine, cisplatin when applied less than 30 minutes prior to irradiation) [48, 49], and selective sensitization of hypoxic tumor cells, which are generally resistant to irradiation (mytomicin C) [50].

Drugs that enhance the effect of irradiation on tumor also contribute toward greater side effects upon normal tissues. In radiobiological terms, adding drugs to irradiation means that each fraction of irradiation is "worth more". At the same total dose, not only can we expect more control of the tumor but also more pronounced acute and late side effects. The simplest, but possibly not the optimal solution, is lowering of the total dose of irradiation. From the logistical perspective, adhering to the standard fractionation is the easiest way; proper attention to late side effects is essential. Hyperfractionation is another approach [51]. In addition to possible better tumor control, its clear advantage is that the maximal tolerated dose would be defined by acute and transient, rather than late and permanent, toxicity.

A detailed discussion of each particular clinical situation would be clearly beyond the scope of this chapter. In concluding this section, we wish to emphasize that the design of combined-modality treatment is a true team venture. Only optimal collaboration of radiation oncologists and medical oncologists will result in a trial that will improve the prospects of this category of patients.

### 3.5 Systemic Treatment of Advanced Cancer

Systemic treatment of advanced cancer is a rapidly expanding and exciting area of medical research. Experience on systemic treatment on advanced cancer is indeed the rostrum of cancer research. Virtually all anticancer drugs and their combinations are first tested in patients with advanced cancer.

Clinical trials of advanced cancer differ from the other previously mentioned topics in three fundamental aspects. As already mentioned, the first difference is their innovative potential. The second one is willingness of patients to participate and to agree to virtually every proposal that might prolong, improve, or save life. When suffering from advanced cancer, many patients do not wish to exert their full autonomy. A researcher in charge of the trial and a physician in charge of recruitment and care for patients in a trial should be aware of limited patient's autonomy and, consequently, of physician's great responsibility.

The third difference is in the choice of endpoints. For all other previously mentioned areas, survival is (or at least should be) the main endpoint. In clinical trials of advanced cancer, we have a choice of several endpoints. While survival is important, it is not always the main endpoint: With incurable disease, other factors such as quality of life come into play. A detailed discussion on endpoints follows in the next section.

Since thousands of clinical trials on treatment of advanced cancer are reported each year, an attempt to present a list of references would be clearly inappropriate. What needs to be mentioned, however, is a recent trend for biological or targeted treatments tailored to specific phenotypic or genotypic characteristics of a tumor. In common and in rare cancers, molecules that exert very specific inhibition of proliferation of tumor cells on a variety of levels from the cell membrane to the nucleus have found rapid application in clinical trials and in regular medical practice [52–57]. Clinical research with these drugs often requires additional information from pathologists; additional consent from the patient to study the tissue blocks is also required. While collection of tissue samples during the trial is a significant burden for all participants, it is virtually impossible once the trial has been

completed. Hence, it is important that these elements of research are provided during the initial version of the protocol.

Eligibility of a particular patient for the new treatment often depends upon results of detailed characterization of the tumor. Many patients will therefore sign informed consent, only to find out that they are not eligible for the expected treatment. It is highly desirable that these patients are offered an alternative treatment within the same treatment protocol. Such a provision in the treatment protocol is not only desirable from the ethical standpoint but will also enable the researchers to follow all subjects in the group, regardless of the characteristics of the tumor. In this way, the results may be compared to trials without such characterization.

### 3.6 Palliative Treatment

This section would be incomplete without a word on the need for clinical research in palliative treatment. For obvious reasons, informed consent should be simplified. Challenges include identifying a target population, avoiding selection bias in the face of clinician and patient denial of serious illness, developing eligibility criteria for a seriously ill population, minimizing high patient refusals due to illness, and accurate reporting of all screened and eligible participants [58].

## 4 TYPES OF CLINICAL TRIALS

The main mission of clinical research is testing ideas for improved diagnostics and treatment under controlled setting and then offering the experience either for further research or for clinical practice. If this is to be accomplished, the experience from clinical research should be relevant for its further application. At the same time, we should not forget that patients participating in clinical research have their individual priorities, which may change with time. Design and conduct of a clinical trial, therefore, involves a delicate balance between patient's interests and objectives of research.

### 4.1 Phase I Trials

For obvious reasons of acute toxicity and long-term mutagenic effects of anticancer drugs, phase I trials in cancer research are not performed on healthy volunteers. The research subjects in phase I trials in oncology are patients wit cancer who have progressed after all conventional therapies. In addition, patients participating in phase I clinical trials are selected for absence of any significant comorbidity. This assures that dose-limiting toxicity and maximal tolerated dose are established with a proper degree of certainty. Since the response to treatment is not among the main objectives, patients in phase I research often suffer from a spectrum of different cancers.

Phase I clinical trials test new drugs on patients resistant to one or several standard combinations of drugs. While the possibility of life-threatening toxicity is not negligible, the chances for an objective response to therapy with a new drug are below 5% [59, 60]. These chances are even smaller for the initial subgroups of patients who are offered only very small doses of the new drug. A reasonable compromise should seek a balance between patient's safety, which demands only very

gradual increases of the dose of the drug, and patient's expectations to derive benefit from the new drug, a possibility that can only be expected with higher doses of the drug [61].

In addition to the commonly discussed situation of testing a new drug as a single agent, phase I trials are done (or at least should be done) also for new combinations of drugs that have already been approved for use and for new combinations of drugs and concomitant irradiation. In testing such a treatment, one drug, an existing combination of drugs, or irradiation are applied in the standard dosing and intervals, while a new drug is added upon this treatment in progressive doses or at progressively shorter intervals. Patients in this type of trial usually suffer from a well-defined diagnosis and stage of disease. In addition to assessment of toxicity, objectives of a trial include response rate, time to progression, and survival. For this reason, the term phase I–II trial is often used for this type of research. While terminology is of minor importance, it is very important that this step is not omitted. Testing a new idea straight in a phase II or even phase III trial often leads to premature closure of the trial due to unexpected toxicity.

### 4.2 Phase II Trials

Phase II clinical trials stand between the very early experience of phase I research and eventual testing of this experience in a phase III trial. Early promising experience in phase I, information on pharmacokinetics and metabolism of drugs, and (to be sincere) commercial interests lead to definition of patients in whom a new drug, or a new combination of drugs, will be tested. Eligibility criteria for phase II trials include a precise diagnosis and stage of disease. Precise evaluation of toxicity remains among the main endpoints. For this reason, patients with severe comorbidity are usually excluded from early phase II trials.

When compared to phase I or phase III trials, phase II trials are easier to organize and cheaper. Number of patients is much smaller than in randomized trials. Many phase II trials are designed by individual researchers in academic institutions and may be conducted even without substantial support from industry. Flexibility in design and short time between idea and actual activation of a trial are the main advantages of such trials.

Experience from phase II trials is a most valuable basis for phase III trials and for preparing meta-analyses. However, two remarks are appropriate. The first one goes to honesty in performing small or single-institution trials. The key factor is unbiased registration of patients into the trial. In final reports, many single-institution trials include only patients who have completed a certain number of treatment cycles and exclude those with early withdrawal due to complications, disease progression, or patient's refusal. Biased registration of patients in single-institution trials is probably the most common form of fraud in clinical research and is the easiest way of improving the results. Unless we are strict in implication of a rigid formal procedure for registration of patients, this form of scientific dishonesty is virtually impossible to discover: In an institution treating thousands of patients, an outside observer cannot identify additional patients who also started the same treatment and were later excluded from the report. To prevent this, patient's entry into a clinical trial should be formal and in written form registered with a person who is not under the physician's authority. From the moment of registration, the patient is included in the trial, regardless of the actual course of disease or treatment.

The second remark is on application of experience from phase II trials for clinical practice. Experience on selected groups of a few dozens or a few hundred patients is far too scarce to be taken as "pure gold." Let us repeat that promising experience from phase II trials gives support for further phase II research (possibly with wider eligibility criteria) and for phase III trials but should not be uncritically applied in clinical practice.

### 4.3 Phase III Trials

Phase III clinical trials present a bridge between clinical research and the application of its experience in clinical practice. It is, therefore, essential that the experience from a phase III trial is applicable to a broad population of patients. Elderly patients and those with common comorbidity (hypertension, diabetes) should not be excluded from research if they are later to be treated with the same drugs. Rather, they may be offered an adjusted treatment schedule [62]. If this is not feasible, it is essential that the authors make a clear statement on limited applicability of the experience.

The ethics of randomized clinical trials have their basis in the uncertainty principle [63]. To maintain the uncertainty throughout the trial, it is essential to keep the recruitment period as short as possible. We have to avoid a possibility that an interim analysis would show a clear, yet statistically nonsignificant, superiority of one treatment over the other. As a rule of the thumb, we propose that the recruitment period should not be longer than double the expected median time to event (e.g., with 12 months as expected median survival, recruitment period should not extend beyond 24 months).

Many randomized clinical trials face the problem of slow recruitment. The most common reaction to slow recruitment is to seek participation of additional cancer centers. Sometimes this works, sometimes it does not; and it is not rare that a trial has to close due to poor recruitment. A general rule that "prevention is better than treatment" may be applied also to the problem of slow recruitment. Quite often, the problem of slow recruitment is not due to a small number of centers but due to a small proportion of eligible patients in each center who actually participate in a trial. Even in lung cancer, the most common cancer worldwide, it is quite common that a single institution recruites fewer than 10 patients within, say, 3 years, out of 100 or more eligible patients. The burden of complicated and time-consuming CRFs has an important negative impact upon recruitment of patients into clinical trials. A busy physician treating several hundred new patients per year does not dare to register more than a dozen patients for multi-institutional clinical trial. The narrow point is not the number of eligible patients, nor registration, nor treatment: It is the fear of CRFs with thousands of unnecessary data and hundreds of replies to queries. Developing a simple CRF is therefore essential for quick recruitment and, indeed, for success (or failure) of a trial. Whenever feasible, a randomized trial should follow the principles of a "large simple trial" and collect only the essential information [64].

## 5 ENDPOINTS

Endpoints in clinical trials in oncology are an interesting topic, yet they are rarely discussed. Few physicians really understand all biases behind simple words such as

response, time to progression, survival, and quality of life [65]. Let us briefly discuss these biases.

### 5.1 Response to Treatment

Response to treatment is now understood as objective evidence of regression of the tumor. Other indications of response such as symptomatic improvement are assessed under quality of life.

In objective evaluation of response, most clinical trials nowadays follow the RECIST guidelines (Response Evaluation Criteria In Solid Tumors) [66]. The procedure involves three steps:

1. Definition of target lesions. At least one and not more than five lesions in any organ affected by cancer should be defined as measurable. Other lesions are declared as nonmeasurable.
2. Measurement of target lesions during or after treatment. The same method (e.g., CT scan) should be used prior to treatment and at evaluation of response. Nonmeasurable lesions are also evaluated for their complete disappearance, persistence, or obvious progression.
3. Confirmation of response after a minimum of 1 month since establishing a response.

By definition, *complete response* means complete resolution of all measurable and nonmeasurable disease for more than 1 month. *Partial response* is defined as >30% reduction in the sum of all measurable lesions for more than 1 month, along with no evidence of progression of nonmeasurable disease. *Progression* is either >20% increase (in comparison with the smallest measurement) in the sum of all measurable lesions, or obvious progression of nonmeasurable disease, or any appearance of a few lesions. *Stable disease* is what remains between partial response and stable disease.

While these three steps are straightforward and should be easy to follow, evaluation of response in the practice of clinical trials is far from devoid of any bias. *Definition of target lesions* is often not done during registration of a patient for a clinical trial. Researchers are allowed to enter patients on the basis of meeting the eligibility criteria, while the target lesions are defined later in the course of treatment. This would be acceptable if all lesions of a tumor would follow the same curve of regression or progression. Still, this is not the case. The tumor population on different sites of disease is heterogeneous. In addition, there is a variable proportion of the accompanying inflammatory reaction that contributes to the bulk of a tumor, as measured radiographically. Chemotherapy itself also does not reach the same concentration in all tissues. For these reasons, we often see considerable heterogeneiety in response among different organs affected by the disease, and even within the same organ. A posteriori definition of target lesions opens a possibility that those responding better will be measured, leading to an increase in response rate.

*Measurement of target lesions* is also not free of a bias. Most tumors are not round and do not shrink or progress evenly in all directions. The longest diameter is often not relevant: A lung tumor may extend into an interlobar fissure, and its longest diameter may remain unchanged even if the tumor shrinks considerably. Measuring

the tumor in other directions allows a bias similar to the one described in the previous paragraph: A posteriori definition of the direction in which response to treatment is most clearly seen leads to an increase in response rate.

*Confirmation of response* is a demand that is not always strictly followed. Even if they declare adherence to the RECIST criteria, not all researchers confirm partial remission after another month. From the clinical point of view, this seems logical: After an objective response has been documented, there is no urgent need for early repeat examinations; reevaluation every 2 or 3 months is sufficient. Finally, confirmation of response is not feasible in certain situations such as induction chemotherapy prior to surgery or irradiation where local treatment immediately follows induction chemotherapy.

#### *Recommendations and Personal View on Evaluation for Response to Treatment*

1. Target lesions for measurable disease should be defined at the time of registration of the patient for the trial.
2. Unidimensional measurement of measurable disease is not free of a bias. Precise volumetric analysis (which is quite feasible with modern computerized radiology) of predefined lesions might offer more information than unidimensional measurements and would better reflect true regression or progression of the disease.
3. Early confirmation of response does not contribute to objective evaluation and adds the burden of unnecessary diagnostics.

### 5.2 Time to Progression

Time to progression is defined as the interval from start of treatment until progression (see previous section for definition of progression). On the first glance, time to progression is clearly defined. Still, at least two comments may be added—one from statistical point of view and the other from the clinical standpoint.

Quite often, we read about very small differences in time to progression of 1 or 2 weeks; yet, the average interval at which the tumor is evaluated may be once every 2 or 3 months. Looking at such data with the eyes of a statistitian, the situation is similar to measuring centimeters with a 1-meter scale. Unless the number of patients is really very large, differences in time to progression smaller than half of the interval between measurements should be taken with great caution.

From the clinical standpoint, very frequent exams for eventual progression may not be justified. Also, a clinician may feel that a certain treatment is still considered beneficial, in spite of minor radiological progression. A useful addition to time of radiologic progression might be *time to clinically meaningful progression*, defined as a moment when the treatment has to be changed or a new treatment modality introduced.

### 5.3 Survival

Survival appears as the most clear endpoint. Yet, even survival is not free of a bias.

Survival is a function of prognostic factors (such as stage, histology, age, gender, and performance status), of a specific treatment under consideration in a particular

trial, and of the treatment after progression (sometimes called "salvage treatment," a term that is often too ambitious or misleading). Although second-line treatment rarely cures patients, it may considerably prolong survival. A trial protocol should therefore define the general guidelines for treatment at the moment of progression. Type of second-line treatment and proportion of patients who actually received such treatment should be included in the report. This will ensure that eventual difference in survival is not due to unbalanced second-line treatment.

When speaking about survival as "the ultimate endpoint," we most often focus on statistically significant difference and do not consider the fact that a difference in survival also has to be clinically meaningful. Provided the trial is large enough, even a 2% difference in survival can become statistically significant. It is important to note that although the difference between two groups may be statistically significant to a very small probability value, the difference may be of no clinical significance [67].

In general, advantage of one treatment over another will shrink when the treatment is taken from the research setting into general use: Broader selection of patients and physicians with varying degree of expertise in the treatment and in management of complications contribute to inferior results. When improvement in survival is small, statistical significance is not the only criterion for accepting a new treatment. In such instances, the concept of clinically meaningful difference should be applied. A judgment on the applicability and relevance of the research data for the general population, on the burden of the new treatment for patients, and on the costs should be thoroughly discussed before the new treatment is accepted as a routine.

### 5.4 Quality of Life

When it comes to incurable disease—and advanced cancer still most often falls under this category—quality of life is of at least equal importance as the other endpoints we just discussed. While we all recognize the importance of quality of life, we are far from agreement on how to approach this issue in clinical trials.

Assessment of quality of life is now regularly included in most treatment protocols [68–70]. Still, the majority of published reports do not present data on quality of life. Most often, we are left with the data on toxicity, which partially reflect quality of life but cannot offer a comprehensive picture. In a survey of randomized clinical trials for advanced breast cancer, assessment of quality of life added relatively little value to other endpoints in helping select the best treatment option, apparently largely because of suboptimal methodological standards [71].

Two fundamental reasons are behind this lack on information on quality of life. The first one is that in any particular trial, data on quality of life are virtually always incomplete. They are most often missing for those patients who do not do well—and for whom evaluation of quality of life would be most important. Patients with progression and/or severe toxicity fail to return for follow-up examinations and do not respond to the questionnaire. In such instances, one cannot avoid a bias in an analysis of an incomplete series of questionnaires [72].

The second reason for rare inclusion of quality of life issues in reports of clinical trials is the difficulty in analyzing the data. Instruments for assessment of quality of life include from 10 to more than 30 questions; a protocol often includes two

instruments (such as observer's and patient's scale). If presenting only one or two items from the lengthy quality of life questionnaire, the author could be blamed for a bias. If all the data are presented and analyzed, the volume of information is such that a separate paper might be needed.

Unlike other endpoints, quality of life data are rarely—if ever—presented in meta-analyses. The reason is inconsistency in instruments and in reporting.

The current trend in assessing quality of life of cancer patients is to use increasingly complex instruments. I am not in favor of this approach. That the mental, physical, and social domains, each containing many dimensions and items, all contribute to quality of life is uncontroversial. What is controversial is the weight of the different dimensions in overall quality of life. It has been shown to be very different between different patient populations. For individuals, assuredly complex systems, the many dimensions and items of quality of life interact, probably sometimes in chaotic ways. In these conditions, the weights of isolated items in individuals become for all practical purposes meaningless. The classical endpoints of discrete health-related functions and duration of survival are increasingly perceived as unacceptably reductionistic [73].

In our single-institutional clinical trials, we use our own simplified scale for assessment of quality of life: How do you feel, in comparison with your feeling prior to treatment?

1. Much worse
2. Worse
3. About the same
4. Better
5. Much better

Such a simple scale follows the idea that it is the patient who can best describe his or her quality of life; It is of lesser importance what precisely this means for an individual patient. The approach may be unscientific, but it is reliable, easy to use, and easy to analyze.

## 6 CONCLUSION

Most methodological issues of clinical research are similar in every field of medicine. An attempt to present a comprehensive overview of the methods of clinical research in oncology would inevitably lead to overlapping with other chapters. General questions of design of a clinical trial, its organization, statistics, and regulatory issues are to be found elsewhere in this volume. In this chapter, we focused on specific questions of design and conduct of clinical research in oncology. The choice of issues was admittedly personal, as were the views and proposals. Some dilemmas of a clinical oncologist who is also involved in research were discussed, all with the aim of facilitating research.

Our progress against cancer critically depends upon the quality and quantity of clinical research. A positive attitude of patients toward participation in clinical research, removing the obstacles that restrain physicians to enter patients into

clinical trials, and the relevance of research for regular medical practice are the three crucial points. We hope that this chapter will contribute toward a better performance on all these three points.

## REFERENCES


1. Sargent, D. J., Conley, B. A., Allegra, C., et al. (2005), Clinical trial designs for predictive marker validation in cancer treatment trials, *J. Clin. Oncol.*, 23, 2020–2027.
2. Meyerson, L. J., Wiens, B. L., LaVange, L. M., et al. (2000), Quality control of oncology clinical trials, *Hematol. Oncol. Clin. North. Am.*, 14, 953–971.
3. Go, R. S., Frisby, K. A., Lee, J. A., et al. (2006), Clinical trial accrual among new cancer patients at a community-based cancer center, *Cancer*, 106, 426–433.
4. Murthy, V. H., Krumholz, H. M., and Gross, C. P. (2004), Participation in cancer clinical trials: Race-, sex-, and age-based disparities, *JAMA*, 291, 2720–2726.
5. Lara, P. N. Jr, Higdon, R., Lim, N., et al. (2001), Prospective evaluation of cancer clinical trial accrual patterns: Identifying potential barriers to enrollment, *J. Clin. Oncol.*, 19, 1728–1733.
6. Sateren, W. B., Trimble, E. L., Abrams, J., et al. (2002), How sociodemographics, presence of oncology specialists, and hospital cancer programs affect accrual to cancer treatment trials, *J. Clin. Oncol.*, 20, 2109–2117.
7. Grunfeld, E., Zitzelsberger, L., Coristine, M., et al. (2002), Barriers and facilitators to enrollment in cancer clinical trials: Qualitative study of the perspectives of clinical research associates, *Cancer*, 95, 1577–1583.
8. Somkin, C. P., Altschuler, A., Ackerson, L., et al. (2005), Organizational barriers to physician participation in cancer clinical trials, *Am. J. Manag. Care*, 11, 413–421.
9. European Union Clinical Trials Directive; available at http://www.wctn.org.uk/downloads/EU_Directive/Directive.pdf.
10. Declaration of Helsinki; available at http://www.wma.net/e/policy/b3.htm.
11. Harris, J. (2005), Scientific research is a moral duty, *J. Med. Ethics*, 31, 242–248.
12. Zwitter, M. (1999), Ethics of randomized clinical trials and the "ALARA" approach, *Acta Oncol.*, 38, 99–105.
13. Williams, C. J., and Zwitter, M. (1994), Informed consent in European multicentre randomised clinical trials. Are patients really informed? *Eur. J. Cancer*, 30A, 907–910.
14. Joffe, S., Cook, E. F., Cleary, P. D., et al. (2001), Quality of informed consent in cancer clinical trials: A cross-sectional survey, *Lancet*, 358, 1772–1777.
15. Coyne, C. A., Xu, R., Raich, P., et al. (2003), Randomized, controlled trial of an easy-to-read informed consent statement for clinical trial participation: A study of the Eastern Cooperative Oncology Group, *J. Clin. Oncol.*, 21, 836–842.
16. Jayson, G., and Harris, J. (2006), How participants in cancer trials are chosen: Ethics and conflicting interests, *Nat. Rev. Cancer*, 6(4), 330–336.
17. Lippman, S. M., and Lee J. J. (2006), Reducing, the "risk" of chemoprevention: Defining and targeting high risk—2005 AACR Cancer Research and Prevention Foundation Award Lecture, *Cancer Res.*, 66, 2893–2903.
18. Klein, E. A. (2006), Chemoprevention of prostate cancer, *Annu. Rev. Med.*, 57, 49–63.
19. Demierre, M. F., Higgins, P. D., Gruber, S. B., et al. (2005), Statins and cancer prevention, *Nat. Rev. Cancer*, 5, 930–942.

20. Kahn, J. A. (2005), Vaccination as a prevention strategy for human papillomavirus-related diseases, *J. Adolesc. Hlth.*, 37, S10–16.
21. Villa, L. L., Costa, R. L., Petta, C. A., et al. (2005), Prophylactic quadrivalent human papillomavirus (types 6, 11, 16, and 18) L1 virus-like particle vaccine in young women: A randomised double-blind placebo-controlled multicentre phase II efficacy trial, *Lancet Oncol.*, 6(5), 271–278.
22. Arnold, D., and Schmoll, H. J. (2005), (Neo-)adjuvant treatments in colorectal cancer, *Ann. Oncol.*, 16(Suppl 2), 133–140.
23. Betticher, D. C. (2005), Adjuvant and neoadjuvant chemotherapy in NSCLC: A paradigm shift, *Lung Cancer*, 50(Suppl 2), S9–16.
24. Glynne-Jones, R., Grainger, J., Harrison, M., et al. (2006), Neoadjuvant chemotherapy prior to preoperative chemoradiation or radiation in rectal cancer: Should we be more cautious? *Br. J. Cancer*, 94, 363–371.
25. Amiel, G. E., and Lerner, S. P. (2006), Combining surgery and chemotherapy for invasive bladder cancer: Current and future directions, *Expert Rev. Anticancer Ther.*, 6, 281–291.
26. Smith, I., and Chua, S. (2006), Medical treatment of early breast cancer. IV: Neoadjuvant treatment, *BMJ*, 332, 223–224.
27. Evans, D. B. (2005), Preoperative chemoradiation for pancreatic cancer, *Semin. Oncol.*, 32(6 Suppl 9), S25–29.
28. Gallo, A., and Frigerio, L. (2003), Neoadjuvant chemotherapy and surgical considerations in ovarian cancer, *Curr. Opin. Obstet. Gynecol.*, 15, 25–31.
29. El Sharouni, S. Y., Kal, H. B., and Battermann, J. J. (2003), Accelerated regrowth of non-small-cell lung tumours after induction chemotherapy, *Br. J. Cancer*, 89, 2184–2189.
30. Carlson, R. W., Brown, E., Burstein, H. J., et al. (2006), National Comprehensive Cancer Network. NCCN Task Force Report: Adjuvant therapy for breast cancer, *J. Natl. Compr. Cancer Net.*, 4(Suppl 1), S1–26.
31. Arriagada, R., Spielmann, M., Koscielny, S., et al. (2005), Results of two randomized trials evaluating adjuvant anthracycline-based chemotherapy in 1146 patients with early breast cancer, *Acta Oncol.*, 44(5), 458–466.
32. Merlano, M., and Mattiot, V. P. (2006), Future chemotherapy and radiotherapy options in head and neck cancer, *Expert Rev. Anticancer Ther.*, 6, 395–403.
33. Leonard, G. D., McCaffrey, J. A., and Maher, M. (2003), Optimal therapy for oesophageal cancer, *Cancer Treat. Rev.*, 29, 275–282.
34. Psyrri, A., and Fountzilas, G. (2006), Advances in the treatment of locally advanced non-nasopharyngeal squamous cell carcinoma of the head and neck region, *Med. Oncol.*, 23, 1–15.
35. Rigas, J. R., and Lara, P. N. Jr. (2005), Current perspectives on treatment strategies for locally advanced, unresectable stage III non-small cell lung cancer, *Lung Cancer*, 50(Suppl 2), S17–24.
36. Henson, J. W. (2006), Treatment of glioblastoma multiforme: A new standard, *Arch. Neurol.*, 63, 337–341.
37. Oehler, C., and Ciernik, I. F. (2006), Radiation therapy and combined modality treatment of gastrointestinal carcinomas, *Cancer Treat. Rev.*, 32, 119–138.
38. Sastre, J., Garcia-Saenz, J. A., and Diaz-Rubio, E. (2006), Chemotherapy for gastric cancer, *World J. Gastroenterol.*, 12, 204–213.
39. Gillespie, M. B., Marshall, D. T., Day, T. A., et al. (2006), Pediatric rhabdomyosarcoma of the head and neck, *Curr. Treat. Options Oncol.*, 7, 13–22.

40. Bosset, J. F., Lorchel, F., Mantion, G., et al. (2005), Radiation and chemoradiation therapy for esophageal adenocarcinoma, *J. Surg. Oncol.*, 92, 239–245.
41. Roukos, D. H., and Kappas, A. M. (2005), Perspectives in the treatment of gastric cancer, *Natl. Clin. Pract. Oncol.*, 2, 98–107.
42. Deutsch, E., Soria, J. C., and Armand, J. P. (2005), New concepts for phase I trials: Evaluating new drugs combined with radiation therapy, *Natl. Clin. Pract. Oncol.*, 2, 456–465.
43. Bradley, J. (2005), A review of radiation dose escalation trials for non-small cell lung cancer within the Radiation Therapy Oncology Group, *Semin. Oncol.*, 32(2 Suppl 3), S111–113.
44. Bernier, J., and Bentzen, S. M. (2003), Altered fractionation and combined radiochemotherapy approaches: Pioneering new opportunities in head and neck oncology, *Eur. J. Cancer*, 39, 560–571.
45. Baumann, M., Appold, S., Petersen, C., et al. (2001), Dose and fractionation concepts in the primary radiotherapy of non-small cell lung cancer, *Lung Cancer*, 33(Suppl 1), S35–45.
46. Wilson, G. D., Bentzen, S. M., and Harari, P. M. (2006), Biologic basis for combining drugs with radiation, *Semin. Radiat. Oncol.*, 16, 2–9.
47. Lawrence, T. S., Blackstock, A. W., and Mcginn, C. (2003), The mechanism of action of radiosensitization of conventional chemotherapeutic agents, *Semin. Radiat. Oncol.*, 13, 13–21.
48. Pauwels, B., Korst, A. E., Pattyn, G. G., et al. (2003), Cell cycle effect of gemcitabine and its role in the radiosensitizing mechanism in vitro, *Int. J. Radiat. Oncol. Biol. Phys.*, 57, 1075–1083.
49. Zwitter, M., Kovac, V., Smrdel, U., et al. (2006), Gemcitabine, cisplatin and hyperfractionated accelerated radiotherapy for locally advanced non-small cell lung cancer, *J. Thorac. Oncol.*, 1, 662–666.
50. Budihna, M., Soba, E., Smid, L., et al. (2005), Inoperable oropharyngeal carcinoma treated with concomitant irradiation, mitomycin C and bleomycin—long term results, *Neoplasma*, 52(2), 165–174.
51. Bernier, J. (2005), Alteration of radiotherapy fractionation and concurrent chemotherapy: A new frontier in head and neck oncology? *Nat. Clin. Pract. Oncol.*, 2, 305–314.
52. Lynch, T. Jr, and Kim, E. (2005), Optimizing chemotherapy and targeted agent combinations in NSCLC, *Lung Cancer*, 50(Suppl 2), S25–32.
53. Nahta, R., and Esteva, F. J. (2006), Herceptin: Mechanisms of action and resistance, *Cancer Lett.*, 232, 123–138.
54. Tarn, C., and Godwin, A. K. (2005), Molecular research directions in the management of gastrointestinal stromal tumors, *Curr. Treat. Options Oncol.*, 6, 473–486.
55. Sanborn, R. E., and Blanke, C. D. (2005), Gastrointestinal stromal tumors and the evolution of targeted therapy, *Clin. Adv. Hematol. Oncol.*, 3, 647–657.
56. Cortes, J., and Kantarjian, H. (2005), New targeted approaches in chronic myeloid leukemia, *J. Clin. Oncol.*, 23, 6316–6324.
57. Boehrer, S., Nowak, D., Hoelzer, D., et al. (2006), Novel agents aiming at specific molecular targets increase chemosensitivity and overcome chemoresistance in hematopoietic malignancies, *Curr. Pharm. Des.*, 12, 111–128.
58. Bakitas, M. A., Lyons, K. D., Dixon, J., et al. (2006), Palliative care program effectiveness research: Developing rigor in sampling design, conduct, and reporting, *J. Pain Symptom Manage.*, 31, 270–284.

59. Rosa, D. D., Harris, J., and Jayson, G. C. (2006), The best guess approach to phase I trial design, *J. Clin. Oncol.*, 24, 206–208.
60. Horstmann, E., McCabe, M. S., Grochow, L., et al. (2005), Risks and benefits of phase 1 oncology trials, 1991 through 2002, *N. Engl. J. Med.*, 352, 895–904.
61. Rogatko, A., Babb, J. S., Tighiouart, M., et al. (2005), New paradigm in dose-finding trials: Patient-specific dosing and beyond phase I, *Clin. Cancer Res.*, 11, 5342–5346.
62. Townsley, C. A., Selby, R., and Siu, L. L. (2005), Systematic review of barriers to the recruitment of older patients with cancer onto clinical trials, *J. Clin. Oncol.*, 23, 3112–3124.
63. Hansson, S. O. (2006), Uncertainty and the ethics of clinical trials, *Theor. Med. Bioeth.*, 27, 149–167.
64. Thom, E. A., and Klebanoff, M. A. (2005), Issues in clinical trial design: Stopping a trial early and the large and simple trial, *Am. J. Obstet. Gynecol.*, 193, 619–625.
65. Sakamoto, J., and Teramukai, S. (2002), Data handling in cancer clinical trials-how we can minimize potential biases, *Jpn. J. Clin. Oncol.*, 32, 1–2.
66. Therasse, P., Arbuck, S. G., Eisenhauer, E. A., et al. (2000), New guidelines to evaluate the response to treatment in solid tumors. European Organization for Research and Treatment of Cancer, National Cancer Institute of the United States, National Cancer Institute of Canada, *J. Natl. Cancer Inst.*, 92, 205–216.
67. Lader, E. W., Cannon, C. P., Ohman, E. M., et al.; the American Heart Association (2004), The clinician as investigator: Participating in clinical trials in the practice setting: Appendix 2: statistical concepts in study design and analysis, *Circulation* 109, e305–307.
68. Kirkova, J., Davis, M. P., Walsh, D., et al. (2006), Cancer symptom assessment instruments: A systematic review, *J. Clin. Oncol.*, 24, 1459–1473.
69. Gunnars, B., Nygren, P., Glimelius, B., et al. (2001), Swedish Council of Technology Assessment in Health Care. Assessment of quality of life during chemotherapy, *Acta Oncol.*, 40, 175–184.
70. Kiebert, G. M., Curran, D., and Aaronson, N. K. (1998), Quality of life as an endpoint in EORTC clinical trials. European Organization for Research and Treatment for Cancer, *Stat. Med.*, 17, 561–569.
71. Fossati, R., Confalonieri, C., Mosconi, P., et al. (2004), Quality of life in randomized trials of cytotoxic or hormonal treatment of advanced breast cancer. Is there added value? *Breast Cancer Res. Treat.*, 87, 233–243.
72. Fayers, P. M., Hopwood, P., Harvey, A., et al. (1997), Quality of life assessment in clinical trials—guidelines and a checklist for protocol writers: The U.K. Medical Research Council experience. MRC Cancer Trials Office, *Eur. J. Cancer*, 33, 20–28.
73. Bernheim, J. L. (1999), How to get serious answers to the serious question: "How have you been?" Subjective quality of life (QOL) as an individual experiential emergent construct, *Bioethics*, 13, 272–287.

# *Pharmacological Treatment Options for Nonexudative and Exudative Age-Related Macular Degeneration*

**Alejandro Oliver, Thomas A. Ciulla, and Alon Harris**
*Department of Ophthalmology, Indiana University, Indianapolis, Indiana*

**Contents**

1 Introduction
2 Diagnosis
3 Nonexudative Age-Related Macular Degeneration
  3.1 Antioxidants
  3.2 Drusen Ablation
  3.3 Rheopheresis
4 Exudative Age-Related Macular Degeneration
  4.1 Thermal Laser Photocoagulation
  4.2 Transpupillary Thermotherapy
  4.3 Photodynamic Therapy
  4.4 Radiation Therapy
  4.5 Surgical Therapy
  4.6 Antiangiogenic Therapy
  References

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

## 1 INTRODUCTION

Age-related macular degeneration (AMD) is the leading cause of vision loss in the developed world, and it is estimated that approximately 1.2 million Americans currently suffer from severe vision loss caused by this disease [1–4]. This number is likely to increase considerably as the population in developed countries ages.

Two main types of macular degeneration have been traditionally recognized: the nonexudative (dry), composed of atrophic changes of the retinal pigment epithelium as well as deposits beneath it, and the exudative (wet) in which abnormal blood vessels develop under the retina causing blood and fluid leakage. Approximately 10–20% of nonexudative AMD will eventually progress to the exudative form, which is responsible for the majority of cases of severe visual loss from AMD [4, 5].

The traditional AMD classification criteria were revised in 1995 by the Age-Related Maculopathy Epidemiological Study Group, and the criteria for diagnosis of AMD became stricter. Patients with minimal or moderate nonexudative age-related changes in the macula were reclassified as having age-related maculopathy (ARM). By definition, advanced retinal pigment epithelium (RPE) atrophy and clumping is now required for nonexudative AMD, and the presence of choroidal neovascularization (CNV) is a requisite for the diagnosis of exudative age-related macular degeneration [6]. Currently, an estimated 85–90% of patients with age-related macular changes are ARM patients who exhibit drusen and only mild to moderate RPE changes and are typically minimally symptomatic with mild blurred central vision, difficulty reading, color and contrast disturbances, and metamorphopsia. The remaining 10–15% of patients with macular changes, who fall under the modern definition of AMD, tend to describe painless, progressive, moderate to severe blurring of central vision and moderate to severe metamorphopsia, which can be acute or insidious in onset [6].

Although some subtypes of exudative AMD are potentially treatable, the currently available treatment modalities offer limited efficacy; therefore, great interest exists in delaying or ceasing the progression of ARM or more effectively treating the factors leading to vision loss once it becomes AMD. At present, the only widely accepted method of intervention for ARM is the use of high-dose antioxidants; however, this only slows progression in some patients and does not reverse any damage already present. Once AMD becomes exudative, the treatment scheme offered to patients varies largely among physicians, as no standard therapy has been established and approved. Options currently available include laser photocoagulation, photodynamic therapy (PDT) with verteporfin, and intravitreal pegaptanib sodium. Only a minority of patients with exudative AMD shows well-demarcated "classic" CNV amenable to laser treatment, and at least half of the patients undergoing thermal laser photocoagulation suffer persistent or recurrent CNV formation within 2 years. In addition, since the treatment itself causes a blinding central scotoma when the CNV is located subfoveally, many clinicians do not treat subfoveal CNV with thermal laser. In 2000, PDT was approved by the U.S. Food and Drug Administration (FDA) as treatment for subfoveal CNV; however, it only limits vision loss and often requires multiple retreatments. Pegaptanib sodium, a vascular endothelial growth factor (VEGF) inhibitor, was approved by the FDA on December 17, 2004, and was available to physicians in January of 2005; however, its administration requires intravitreal injections every 6 weeks. Because of these treat-

ment limitations, alternative therapies for exudative AMD are being developed and include new types of photodynamic therapy, transpupillary thermotherapy, growth factor modulators, radiation, and surgical therapy. A major limitation for the design of effective treatment is still our lack of true understanding of the underlying etiology of the disease.

## 2 DIAGNOSIS

Any patient who exhibits signs and symptoms consistent with exudative AMD undergoes a thorough dilated fundus exam, stereo color fundus photography, and rapid sequence fluorescein angiography (RSFA). RSFA, which is usually the initial angiographic study, will reveal leakage—the hallmark of CNV. According to the distance from the foveal avascualar zone, the leakage is classified as subfoveal, juxtafoveal (1–199 μm) or extrafoveal (200–2500 μm). In addition, indocyanine green (ICG) angiography is performed as an adjunctive study in patients with poorly delineated CNV. ICG can better delineate choroidal circulation because the near-infrared light (795–810 nm) absorbed by ICG tends to penetrate the retinal pigment epithelium better than can the shorter wavelength of light absorbed by fluorescein. Also, unlike fluorescein, ICG is strongly bound to plasma proteins, which prevents diffusion of the compound through the fenestrated choroidal capillaries, and permits better delineation of choroidal details. When a CNV is suspected, angiography is customarily performed within 72 hours of any planned treatment since CNV morphology and resulting treatment parameters can evolve rapidly.

The Macular Photocoagulation Study (MPS) defined two RSFA leakage patterns for CNV [7]. "Classic" CNV presents as discrete early hyperfluorescence with late leakage of dye into an overlying neurosensory retinal detachment. "Occult" CNV is categorized into two basic forms: late leakage of undetermined source, or fibrovascular pigment epithelial detachments. (PEDs). Late leakage of undetermined source manifests as regions of stippled leakage into an overlying neurosensory retinal detachment, without a distinct source identified on the early frames of the angiogram. Fibrovascular PEDs present as irregular elevations of the RPE associated with stippled leakage into an overlying neurosensory retinal detachment in the early and late frames of the angiogram.

## 3 NONEXUDATIVE AGE-RELATED MACULAR DEGENERATION

### 3.1 Antioxidants

The treatment options for ARM and nonexudative AMD are limited; therefore, prevention of disease progression is viewed as critical. Currently, the only evidence-based intervention comes from the Age-Related Eye Disease Study (AREDS) [8]. This multicenter, U.S. National Institutes of Health (NIH)–supported investigation was a double-masked, randomized, prospective, clinical trial that enrolled 4357 subjects into 1 of 4 treatment groups: placebo, antioxidants (vitamins C and E plus β-carotene), zinc/copper, and antioxidants plus zinc/copper. The study was based on the theory that oxidative damage to the retina may contribute to the development of AMD [9–12] and a smaller, randomized, placebo-controlled clinical trial

that suggested zinc might provide protection from vision loss due to AMD [13]. AREDS concluded that "patients with extensive intermediate sized drusen, at least 1 large drusen or noncentral geographic atrophy in one or both eyes, or advanced AMD or vision loss due to AMD in one eye should consider taking a supplement of antioxidants plus zinc" [8]. The formulation (Ocuvite PreserVision, Bausch & Lomb, Rochester, New York, and ICaps, Alcon, Inc., Fort Worth, Texas) containing high doses of vitamin C, vitamin E, β-carotene, copper, and zinc lowered the risk of developing advanced AMD by 25% in the study [8]. Importantly, another study demonstrated that β-carotene supplementation increased the risk of developing lung cancer in smokers; therefore, the antioxidant formulation should exclude β-carotene when taken by tobacco users [14].

### 3.2 Drusen Ablation

It has been proposed that the aging RPE accumulates remnants of incomplete degradation of phagocytosed rod and cone membranes, and such accumulation results in the presence of metabolic debris and, over time, drusen formation [15, 16]. Laser treatment at or near a large drusen will often lead to resolution of the drusen, and sometimes this will be accompanied by an improvement of visual acuity. Based on the knowledge that drusen constitute a risk factor for exudative AMD, it has been proposed that laser treatment to promote clearing of drusen may result in a reduction of the risk of CNV formation. A randomized multicenter clinical trial, known as CNVPT (Choroidal Neovascularization Prevention Trial) evaluated 432 eyes treated with argon green laser. The study showed significant drusen reduction in the treated group; however, it also suggested that argon green laser treatment might increase the risk of CNV development [17]. Another study also showed that treatment with laser photocoagulation results in significant drusen reduction compared with observation at 2 years, but no differences were observed in CNV occurrence between groups [18]. In addition, two prospective randomized clinical trials are currently in progress to evaluate the effects of laser-induced drusen reduction on AMD. The NIH has supported the CAPT (Complications of AMD Prevention Trial) study, which is evaluating low-intensity argon green laser photocoagulation in patients with bilateral drusen. The study completed enrollment of 1052 patients in March 2001, and the 5-year results should be available this year [19]. The PTAMD (Prophylactic Treatment of AMD) study, a multicenter randomized prospective placebo-controlled clinical trial sponsored by Iridex Corp (Mountain View, CA) has completed enrollment and is currently in progress [20]. It is designed to evaluate the effects of extrafoveal subthreshold infrared diode treatment on stopping the progression of exudative AMD.

### 3.3 Rheopheresis

Some researchers believe that lipid deposition in sclera and Bruch's membrane leads to scleral stiffening and impaired choroidal perfusion, which in turn could adversely affect the metabolic transport function of RPE. The affected RPE would in consequence be incapable of efficiently metabolizing and removing material shed from the photoreceptors, leading to its accumulation and drusen formation [21, 22]. Based on this theory, it has been speculated that apheresis could be of benefit to

AMD patients by improving ocular blood flow and by limiting the concentration of circulating macromolecules involved in drusen formation, Bruch's membrane degradation, and retinal pigment epithelial cell dysfunction. OccuLogix LP (a joint venture between TLC Vision Co, Mississaugua, Canada, and Vascular Sciences CO, Tampa, Florida) developed the Rheofilter membrane differential filter (MDF) system based on the concept of apheresis, to filter high-molecular-weight proteins and lipoproteins from the blood. A randomized prospective, double-masked pilot study involving 30 patients showed vision improvement in a significant number of patients undergoing rheophoresis. Currently, in the United States the MIRA-1 (Multicenter Investigation of Rheophoresis for AMD) is a multicenter, randomized, placebo-controlled trial of patients with large soft drusen without advanced AMD in at least one eye, and elevated serum cholesterol, IgA or fibrinogen at screening. This phase III trial was interrupted after about one-half of the planned 180 patients had been recruited due to loss of capitalization by the sponsor. An interim analysis performed on 43 patients who completed the one-year visit revealed an improvement in mean acuity of greater than one line in treated patients, compared to a mean loss of almost two lines in controls. The study has now been resumed and 185 patients have been enrolled. On November 17, 2005, OccuLogix announced that all final study visits had been completed for the MIRA-1 trial, and results would be analyzed by the end of December 2005. They were planning to file for FDA approval in the first quarter of 2006 [23].

## 4 EXUDATIVE AGE-RELATED MACULAR DEGENERATION

### 4.1 Thermal Laser Photocoagulation

Traditionally, ophthalmologists have used thermal laser destruction of CNV as the primary treatment of exudative AMD based on the results of the Macular Photocoagulation Study (MPS), a large, randomized, multicenter, prospective set of clinical trials comparing laser photocoagulation to observation. These studies, which were initiated in the 1980s and supported by the NIH, demonstrated that laser photocoagulation of certain types of CNV lowered the risk of large reductions in visual acuity compared to observation alone. In these studies, patients were deemed eligible for laser photocoagulation if they manifested classic CNV as determined by RSFA.

Unfortunately, only 13–26% of patients with exudative AMD presented with classic CNV eligible for laser treatment, and it became unclear whether laser photocoagulation was beneficial in a majority of patients, as they were not eligible for laser therapy in the MPS [7, 24]. Moreover, at least half of the enrolled subjects suffered from persistent or recurrent CNV formation within 2 years of treatment [24–26]. Although the arm of the MPS exploring treatment of CNV under the fovea suggested that laser photocoagulation is better than observation, treating subfoveal CNV with thermal photocoagulation is not a common practice because of the immediate central scotoma from the collateral retinal destruction. Given the large number of limitations posed by thermal laser photocoagulation, researches have searched for alternative means of subfoveal CNV treatment using a variety of laser derivatives [27, 28].

### 4.2 Transpupillary Thermotherapy

Transpupillary thermotherapy (TTT) occludes the CNV by slowly heating the subfoveal choroidal neovascular complex with infrared (810 nm) diode laser light. The infrared wavelength is thought to traverse the retina and RPE to maximally affect the CNV, while minimizing thermal injury to the neurosensory retina. Laser application covers the entire CNV complex with a single large spot. Although the precise mechanism of CNV destruction is unclear, one study, using color Doppler imaging, suggested TTT leads to alterations in choroidal blood flow [29]. An uncontrolled phase I/II safety and efficacy study involving 113 patients showed that patients with occult CNV receiving TTT compared similarly to the verteporfin-treated patients in the verteporfin in photodynamic therapy (VIP) trial at 6 and 12 months [30]. Similarly, another uncontrolled trial with 69 patients found that TTT use compared favorably to the natural history of occult CNV [31]. The Transpupillary Thermotherapy of Occult Subfoveal Choroidal Neovascular Membranes in Patients with Age-Related Macular Degeneration Trial (TTT4CNV), the first randomized, prospective, double-blind, placebo-controlled study evaluating the effectiveness of TTT for occult CNV (or up to 10% classic CNV), enrolled 303 patients between March 2000 and March 2003 and is now following subjects [32].

### 4.3 Photodynamic Therapy

Photodynamic therapy (PDT) utilizes laser light and intravascular dyes (i.e., photosensitizers). After intravenous injection, once sufficient time passes to concentrate the photosensitizer in neovascular tissue, the CNV is stimulated with a specific wavelength of light to react with water and create oxygen and hydroxyl free radicals [33]. These free radicals, in turn, react with cell membranes of the pathologic endothelium to induce occlusion by massive platelet activation and thrombosis, while still preserving the normal choroidal vasculature and nonvascular tissue [34, 35]. Ideally, the intensity of the exciting wavelength is low enough to spare the non-neovascular irradiated tissues from thermal damage. Important variables in this reaction include the intravascular concentration of dye, the photochemical behavior of the dye, the interval between the injection and the onset of irradiation, the intensity and specificity of the exciting irradiation, and the duration of irradiation [36–38].

***Verteporfin*** The U.S. FDA approved verteporfin (Visudyne; QLT Therapeutics, Inc., Vancouver, Canada, and Novartis Ophthalmics, Bulach, Switzerland) in April 2000 for patients with "predominantly classic" subfoveal CNV caused by AMD, which demonstrates a characteristic early and well-defined RSFA stain to over 50% of the CNV complex. Similarly, marketing approval was granted in Europe in July 2000, and it is currently commercially available in over 70 countries for predominantly classic CNV.

In 1999 and 2001, the 1- and 2-year results of the Treatment of AMD with PDT (TAP) study were published. TAP consisted of two randomized, prospective, double-blind, placebo-controlled phase III trials with 609 subjects. First-year data reported the proportion of eyes with less than 15 letters of visual acuity loss on a standardized eye chart was 67% in the treated group versus 39% in control group

($p < 0.001$) when the CNV was predominantly classic; however, no significant differences in visual acuity were demonstrated when the area of classic CNV was less than 50% of the entire complex. In addition, it was noted that 90% of the subjects required retreatment at 3 months, and an average of more than three retreatments over the first year [39]. Second-year follow-up data reported that 59% of treated eyes had a favorable visual outcome versus 31% in the control group when the lesion was predominantly classic [40]. The TAP trial was unmasked at 2 years of follow-up. An open-label extension to 36 months of 124 of the 159 original TAP participants with predominantly classic CNV revealed that visual acuity remained nearly constant and required fewer retreatments [41]. Because of the success of the TAP trial, the verteporfin in photodynamic therapy (VIP) trial, another randomized, prospective, double-blind, placebo-controlled clinical trial, was developed to examine many of the patients who fell outside of the inclusion guidelines set by TAP.

The VIP trial was designed to evaluate the treatment efficacy of PDT in 339 subjects with total occult subfoveal CNV, classic CNV with a visual acuity better than 20/40, or CNV secondary to pathological myopia. One-year results of the occult AMD arm showed no significant difference between visual acuity outcomes in exudative AMD patients treated with verteporfin and placebo (51% PDT and 54% placebo treated had unfavorable visual outcomes, respectively). However, 2-year follow-up data revealed that 55% of the treated subjects with occult CNV had an unfavorable outcome versus 68% in the placebo group ($p = 0.023$). On average, the verteporfin-treated patients received five treatments over 24 months of follow-up. Based on this data, the study group recommended verteporfin for purely occult subfoveal CNV that demonstrated recent disease progression in all patients except those with large lesions with good visual acuity [42]. Because the FDA desired additional data before approving verteporfin for occult CNV, the Visudyne in Occult (VIO) trial was developed as a 24-month study to analyze patients with only occult CNV. Enrollment of 364 subjects was completed and is currently in the second year of follow-up as per the recommendation of the Data and Safety Monitoring Committee [43].

Several other trials have evaluated the efficacy of verteporfin in a variety of clinical situations previously lacking sufficient data. Retrospective TAP and VIP data suggested some treatment benefit for smaller minimally classic lesions. The Visudyne in Minimally (VIM) classic trial was thus initiated as a randomized, prospective, double-blind, placebo-controlled clinical trial designed to study the use of verteporfin in patients with minimally classic CNV. Phase II data on 117 patients suggests that small, recently progressive, minimally classic CNV might benefit from verteporfin therapy [44]. Two-year follow-up data revealed that fewer verteporfin-treated eyes lost three or more lines of vision on a standard visual acuity chart or converted to a predominantly classic lesion as compared to placebo [45]. Consequently, a phase III study [the visudyne minimally classic (VMC) trial] was started in late 2003 to further evaluate verteporfin in minimally classic CNV. On April 1, 2004, the U.S. Centers for Medicare and Medicaid Services (CMS) agreed to reimburse physicians for PDT of occult and minimally classic subfoveal CNV (less than 50% of CNV complex with early well-defined hyperfluorescence on RSFA) from AMD, provided that the lesion is four disk areas or less in size at least three months prior to initial treatment and evidence of progression (i.e., loss of five or more letters

on standard visual acuity charts, increase of at least one disk diameter, or appearance of blood) within 3 months of treatment.

Because 80% of vision loss in verteporfin-treated patients occurs within 6 months of developing CNV, the Verteporfin Early Retreatment (VER) trial was designed as a phase III study of 323 patients to compare the benefit of retreatment in 6-week intervals versus the standard 3 months. Twelve-month interim results of the 2-year trial did not show improved outcomes when compared to the standard treatment [46]. Additionally, the Verteporfin with Altered (Delayed) Light in Occult (VALIO) study was developed to evaluate whether delaying the light application to 30 minutes after the initiation of verteporfin infusion (versus the standard 15 minutes) would improve outcomes in occult CNV. Phase II data at 6 months of follow-up show the group treated at 30 minutes postinfusion lost 1.3 lines of vision while the standard 15 minute postinfusion treatment group lost 2–3 lines, which was not statistically significant. One-year data substantiated the 6-month findings [47, 48].

At the moment verteporfin is the only approved PDT agent, but additional photosensitizing products are under study and development.

***Rostaporfin*** Rostaporfin (Photrex; Miravant Medical Technologies, Santa Barbara, California) is a purpurin with a structure similar to chlorophyll that absorbs maximally at 664 nm. Like verteporfin, the preconstituted solution is intravenously infused over 10–20 minutes. In December 2001, enrollment for a phase III placebo-controlled, double-masked clinical trial involving 920 patients was completed. Two-year follow-up data found that 58% of patients receiving a 0.5 mg/kg dose of SnET2 lost less than 15 letters compared to 42% of placebo patients ($p = 0.0045$). Rostaporfin was well tolerated and demonstrated an acceptable safety profile [49]. On September 30, 2004, the FDA requested an additional confirmatory clinical trial before final marketing approval.

### 4.4 Radiation Therapy

Since choroidal neovascular membranes are composed of rapidly proliferating pathologic endothelial cells, they may be sensitive to agents that inhibit cell division. Consequently, radiation therapy has been suggested as a treatment for subfoveal CNV. Given an apparent dose–response effect, some groups have delivered ionizing radiation to the macula using modalities that may limit the exposure of ionizing radiation to normal radiosensitive structures of the eye, such as the optic nerve or lens. These methods have included stereotactic external photon beam irradiation of the posterior pole, brachytherapy, in which radioactive plaques are sutured to the posterior pole of the eye and explanted several days later, and proton beam irradiation, which deposits almost all of its energy at the desired depth in the eye at a point called the Bragg peak and undergoes little scattering [50, 51].

Although some of the early pilot studies suggested a possible benefit, conflicting reports regarding the efficacy of radiation therapy for exudative AMD have since been published. Two prospective controlled studies, using a relatively large number of low ionizing radiation fractions, failed to show a treatment benefit for external beam radiation [52, 53]. However, two smaller prospective controlled studies using a smaller number of higher radiation fractions demonstrated a statistically significant vision benefit over controls [54, 55]. Because of the positive outcomes, a

prospective, controlled, pilot study to evaluate external beam radiation on CNV in a small number of high-energy fractions was sponsored by the National Eye Institute. An interim analysis of this study, known as the AMD Radiotherapy Trial (AMDRT), found that at 12 months follow-up 43% of radiated eyes and 50% of nonradiated eyes demonstrated a moderate visual loss ($p = 0.60$) [56].

### 4.5 Surgical Therapy

Some vitreoretinal surgeons have attempted to remove CNV with direct surgical excision, which can yield impressive results in CNV secondary to histoplasmosis and multifocal choroiditis. However, the results were disappointing for exudative age-related macular degeneration. Researchers speculate the CNV of AMD has a different morphology and grows both anterior and posterior to the RPE. The damaged RPE that remains after CNV removal causes atrophy of the underlying choriocapillaris leading to retinal disorganization [57–60]. In 1998, the National Eye Institute of the National Institutes of Health awarded funding to the submacular surgery trial (SST). This study was designed as a randomized, multicenter, prospective clinical trial comparing surgery with observation to specifically evaluate patients with large or poorly demarcated new subfoveal CNV, submacular hemorrhage from CNV associated with exudative AMD, or subfoveal CNV due to presumed ocular histoplasmosis (POHS) or idiopathic causes.

Patients were followed for 2 years and assessed for stabilization or deterioration of visual acuity (VA), change in contrast sensitivity, cataract development, surgical complications, and quality of life. Of 454 patients with subfoveal choroidal neovascularization enrolled, 228 study eyes were assigned to observation and 226 to surgery. Median VA losses from baseline to the 24-month examination were 2.1 lines (10.5 letters) in the observation arm and 2.0 lines (10 letters) in the surgery arm. Median VA declined from 20/100 at baseline to 20/400 at 24 months in both arms. Moreover, rhegmatogenous retinal detachment occurred in 12 surgery eyes (5%) and 1 observation eye. In conclusion, it was determined that submacular surgery does not improve or preserve VA for 24 months better than observation, and it is therefore not recommended for patients with subfoveal choroidal neovascularization caused by AMD or with submacular hemorrhage from CNV [61, 62].

### 4.6 Antiangiogenic Therapy

***Vascular Endothelial Growth Factor Inhibitors*** Animal and clinical studies have identified vascular endothelial growth factor (VEGF) as a key mediator of ocular angiogenesis [63]. Upregulation of VEGF expression has been reported in experimentally induced CNV in rats, and it has also been shown that VEGF is capable of inducing intraretinal and subretinal neovascularization [64]. In human clinical trials, particular attention has focused on the development of pharmaceutical agents to block VEGF expression or neutralize it once expressed. Investigators have inhibited preretinal neovascularization in experimental models with antibodies against VEGF. Others have shown similar effects using VEGF-neutralizing chimeric proteins, which were constructed by joining the extracellular domain of high-affinity VEGF receptors with IgG [65].

*Pegaptanib Sodium* The anti-VEGF pegylated aptamer, pegaptanib sodium (Macugen; Eyetech Pharmaceuticals, Inc., New York, and Pfizer, Inc., New York), demonstrated both safety and efficacy in clinical trials. This intravitreally administered polyethylene glycol (PEG)-conjugated oligonucleotide was specifically designed to bind and neutralize $VEGF_{165}$, hypothesized to be the predominant VEGF isomer associated with CNV in humans. A phase I trial, involving 15 subjects receiving a single injection of pegaptanib sodium, demonstrated 80% with stable or improved vision at 3 months. More impressively, 27% of the patients had significantly improved vision: A finding missing from many of the other standard AMD treatment modalities [66]. Although small, a phase II trial involving 21 patients supported the phase I data, and, when pegaptanib sodium injections were combined with PDT, 6 of 10 (60%) patients had significantly improved vision versus 2.2% treated with PDT alone [67].

The VEGF Inhibition Study in Ocular Neovascularization (VISION), two phase II/III, multicenter, randomized, placebo-controlled studies, completed enrollment of 1186 subjects in July 2002. The 2-year follow-up data revealed less vision loss for subjects maintained on pegaptanib sodium than those who only received the medication during one year ($p < 0.05$). In the group given pegaptanib at 0.3 mg, 70% of patients lost fewer than 15 letters of visual acuity, as compared with 55% among the controls ($p < 0.001$). The risk of severe loss of visual acuity (loss of 30 letters or more) was reduced from 22% in the sham-injection group to 10% in the group receiving 0.3 mg of pegaptanib ($p < 0.001$). More patients receiving pegaptanib (0.3 mg), as compared with sham injection, maintained their visual acuity or gained acuity (33% vs. 23%; $p = 0.003$). As early as 6 weeks after beginning therapy with the study drug, and at all subsequent points, the mean visual acuity among patients receiving 0.3 mg of pegaptanib was better than in those receiving sham injections ($p < 0.002$). Among the adverse events that occurred, endophthalmitis (in 1.3% of patients), traumatic injury to the lens (in 0.7%), and retinal detachment (in 0.6%) were the most serious and required vigilance. These events were associated with a severe loss of visual acuity in 0.1% of patients [68]. Based on these results, the FDA approved the use of pegaptanib sodium to slow vision loss in people with neovascular AMD on December 20, 2004 [69].

*Ranibizumab* Ranibizumab (Lucentis, Genentech Inc., San Francisco, and Novartis Ophthalmics, Basel, Switzerland), an intravitreally injected, recombinant, humanized, monoclonal antibody Fab fragment designed to actively bind and inhibit all isoforms of VEGF, has shown promise in early human trials. A phase Ib/II randomized, single-agent study found that 94% of the 50 patients receiving ranibizumab had stable vision and 44% had significantly improved vision at 6 months [70]. Additional trials have since then been initiated to provide more definitive evaluation of the clinical benefit of ranibizumab in patients with predominantly classic or minimally classic/occult CNV.

The MARINA (Minimally Classic/Occult Trial of the Anti-VEGF Antibody Ranibizumab in the Tretament of Neovascular AMD) is a phase III randomized, prospective, double-blind, placebo-controlled trial initiated in 2003 with the objective of comparing ranibizumab against verteporfin for minimally classic or occult CNV. A total of 716 patients were enrolled in this study and randomized 1:1:1 to sham injection or to ranibizumab (0.3 or 0.5 mg) injected intravitreally monthly for

24 months. Preliminary analysis of one-year MARINA data revealed that approximately 95% of patients treated with ranibizumab lost fewer than 15 letters at one year, compared with approximately 62% in the control group ($p < 0.0001$). On average, the patients treated with ranibizumab had a significant improvement in visual acuity relative to their visual acuity at study entry, whereas the control group experienced a substantial decrease from baseline in mean visual acuity [71].

The ANCHOR (Anti-VEGF Antibody for the Treatment of Predominantly Classic Choroidal Neovascularization in AMD) is a multicenter, prospective, randomized, multicenter, double-masked phase III trial designed to compare a combination ranibizumab/PDT theraoy to verteporfin PDT alone in 423 subjects with predominantly classic exudative AMD. This trial is still ongoing in centers in United States, Europe, and Australia [71].

The FOCUS (RhuFab V2 Ocular Treatment Combining the Use of Visudyne to Evaluate Safety) study is a randomized, single-masked phase I/II trial investigating the safety, tolerability, and efficacy of ranibizumab in combination with verteporfin PDT versus verteporfin PDT alone in patents with subfoveal predominantly classic CNV due to AMD. Enrollment of 162 patients has been completed and a preliminary analysis of one-year data indicate that approximately 90% of patients treated with the combination of ranibizumab and PDT had stable or improved visual acuity, compared with approximately 68% of patients in the control arm of PDT alone ($p = 0.0003$) [71].

The Phase IIIb, multicenter, randomized, double-masked, sham Injection-controlled study of the Efficacy and safety of Ranibizumab (PIER) study started enrolling approximately 180 patients in September 2004 with the objective of comparing 3-month intravitreal dosing intervals to the standard 1-month intervals [72].

***Pigment Epithelium-Derived Factor Inducer*** Researchers have attempted to stimulate intravitreal production of native pigment epithelium-derived factor (PEDF), a naturally occurring potent antiangiogenic protein deficient in eyes with CNV, using gene therapy [73]. PEDF inhibits angiogenesis by inducing apoptotic death of endothelial cells stimulated to form new vessels [74]. In a laser-induced CNV murine model, choroidal neovascularization was reduced after intravitreal PEDF was produced from an adenoviral vector [75]. One study demonstrated that increased intravitreal PEDF results in 85% inhibition of neovascularization in laser-induced CNV, transgenic VEGF, and retinopathy of prematurity models. GenVec, Inc. (Gaithersburg, Maryland) has developed a PEDF producing adenovirus vector, AdPEDF (pigment epithelium-derived factor on an adenovirus vector), and completed recruitment of 51 patients in five states for phase I human trials in August 2004. Interim 12-month results of 24 patients revealed no dose-limiting toxicities or related severe adverse events [76].

***Cyclo-Oxygenase-2 Inhibitor*** It is likely that a whole cascade of soluble factors play a role in ocular angiogenesis and CNV development. Cyclooxygenase-2 (COX-2) is expressed in neovascular structures, especially human cancers. Thus, it has been proposed that a cyclooxygenase-2 inhibitor might control neovascularization associated with AMD. Orally administered celecoxib (Celebrex; Pfizer, Inc., New York), a COX-2 inhibitor, significantly reduced angiogenesis and prostaglandin production in basic fibroblast growth factor (bFGF)-induced neovascularization of

rat corneas [77]. The National Eye Institute is conducting a phase I/II safety and efficacy trial comparing the use of PDT and celecoxib to PDT alone. The double-masked, randomized, placebo-controlled prospective study completed enrollment of 60 participants.

***Squalamine*** Squalamine (Genaera Co., Plymouth Meeting, PA), an antiangiogenic aminosterol originally found in the body tissues of the cancer-resistant dogfish shark, acts as an inhibitor of growth factor signaling, including VEGF, integrin expression, and cytoskeletal formation. Systemic intravenous administration has inhibited iris neovascularization in primate models, oxygen-induced retinopathy in murine models, and laser-induced CNV in a rat model [78–80].

Three phase II clinical trials are currently underway to evaluate its role in AMD treatment. The largest, MSI-1256F-209, is a 100-patient prospective, randomized, controlled trial evaluating the effects of 20 or 40 mg given intravenously every week for 4 weeks followed by maintenance every 4 weeks for 48 weeks followed by 12 months of observation for exudative AMD. The second trial, MSI-1256F-208, is a 45-patient prospective controlled trial evaluating the effects of 10, 20, or 40 mg of intravenous squalamine given initially in combination with verteporfin PDT and then alone for an additional 6 months followed by 12 months of observation for exudative AMD. Preliminary results from this trial showed that subjects treated with 40 mg of squalamine lactate and concomitant PDT gained an average of 0.4 letters in visual acuity compared to study entry, while those treated with PDT alone lost an average of 4.8 letters [81]. The last trial, MSI-1256F-207, is an 18-patient open-label, parallel group trial comparing three doses of intravenous squalamine given weekly for 4 weeks followed by 4 months of follow-up on exudative AMD. On October 4, 2004, the U.S. FDA granted fast-track designation to squalamine.

***Steroid Compounds*** Corticosteroid compounds have long been known to possess angiostatic properties as they alter extracellular matrix degradation and inhibit inflammatory cells, which invariably participate in neovascular responses [82]. Intravitreal administration of corticosteroids has become very popular as the blood–ocular barrier is bypassed, more constant therapeutic steroid levels are achieved, and systemic side effects are minimized. These injections have demonstrated efficacy in subretinal and preretinal neovascularization in animal models [83, 84].

*Triamcinolone Acetonide* Uncontrolled pilot studies evaluating CNV in AMD have employed the off-label use of intravitreally administered triamcinolone acetonide (Kenalog; Bristol-Myers Squibb, New York) because of its long half-life and corticosteroid properties. One study of 30 eyes receiving a single triamcinolone acetonide injection reported that 11 eyes experienced improved or stabilized vision within 1–3 months of treatment with regression of the CNV to inactive fibrosis, 15 experienced a similar outcome, except for slow extension and exudation from recurrent CNV, while 4 experienced no obvious treatment benefit [85]. In later publications, studies reported a favorable effect on the course of the disease over 6-, 12-, and 18-month follow-up; however, the lack of controls complicated the ability to compare treatment efficacy to the natural course of the disease [86–88]. The authors proposed that intravitreal triamcinolone had a beneficial effect on AMD-related CNV through inhibition of leukocytes, including macrophages, which normally

release angiogenic factors [85–87]. A randomized, double-masked, placebo-controlled clinical trial of 151 eyes receiving a single 4-mg injection of intravitreal triamcinolone found significant antiangiogenic effects at 3 months after treatment; however, no beneficial visual acuity effect was seen at 1 year [89]. The authors speculated that triamcinolone might be efficacious at a higher or more sustained dose or in concert with other modalities.

Visagen (Regenera Limited, Nedlands, Australia) is developing a triamcinolone acetonide formulation that is being developed strictly for intraocular applications. Regenera Ltd. anticipates sponsoring clinical trials in an attempt to formally gain approval for several ophthalmic indications in the near future. A different group is developing a preservative-free formulation that theoretically decreases the 0.8% sterile endophthalmitis rate observed with traditional intravitreal triamcinolone acetonide injections [90].

*Anecortave Acetate* In 1985, a class of steroid with minimal glucocorticoid or mineralocorticoid activity was developed and is now undergoing evaluation in human trials as anecortave acetate (Retaane; Alcon Laboratories, Inc., Fort Worth, Texas) [91]. The lack of corticosteroid activity minimizes commonly encountered intraocular pressure elevation and accelerated cataract formation [92]. In addition, anecortave acetate was formulated for injection into the subtenon space with a specially designed cannula. A phase II/III randomized, prospective, placebo-controlled trial involving 128 patients designed to evaluate the clinical safety and efficacy of juxtascleral injection of anecortave acetate versus placebo for the treatment of subfoveal CNV, found that baseline vision ($p = 0.01$), stabilization of vision ($p = 0.03$), and prevention of severe vision loss ($p = 0.02$) were statistically superior to baseline at 12 months; however, the dropout rate was nearly 50% [93].

Another phase III study designed to compare anecortave acetate 15-mg suspension to visudyne PDT in patients with exudative AMD was carried out over a 12-month period. The study enrolled 530 patients with the primary objective of demonstrating that anecortave acetate 15 mg is noninferior to PDT in patients with predominantly classic subfoveal CNV. It was shown that no difference existed between the two treatment groups ($p = 0.4305$) [94].

A new study (C-02-60) evaluating anecortave acetate suspension versus a sham administration procedure for prevention of progression from dry AMD to exudative AMD is currently enrolling patients. The objective of this study is to determine the safety and efficacy of anecortave acetate suspension for treatment of patients with nonexudative AMD who are at risk of progression to exudative AMD, in an attempt to arrest the development of choroidal neovacularization. A total of 2500 patients are to be enrolled in this 4-year study [95].

*Implantable Corticosteroids* Because intraocular corticosteroids have shown antiangiogenic effects with repeated intravitreal administration, the efforts to develop sustained-release intraocular implants have received special attention, in an attempt to achieve near constant intraocular steroid concentrations without repeated injections. One study demonstrated CNV inhibition using triamcinolone acetate microimplants in a laser-induced CNV rat model [96]. Furthermore, researchers at Bausch & Lomb (Rochester, New York) and Control Delivery Systems (Watertown, MA) have developed Retisert (also known as Envision TD), a nonbiodegradable intra-

vitreal implant that releases fluocinolone acetonide for up to 3 years. Phase III studies involving diabetic macular edema were promising for resolution of retinal leakage compared to placebo. However, the study also reported that after one year 58.5% of subjects receiving the 0.5-mg implant developed serious side effects such as increased intraocular pressure, vitreal hemorrhage, and cataracts, complications that occurred in only 10.7% of the standard care group. The patients will be followed for an additional 3 years to monitor the safely of the implant. Enrollment is complete for a phase II study to evaluate the effects of Retisert on exudative AMD. However, development in this indication has been discontinued [97]. Similarly, a biodegradable dexamethasone implant (Posurdex; Allergan, Irvine, California) has shown safety and benefit in recent phase II trials for macular edema from diabetes mellitus, branch or central retinal vein occlusion, uveitis, or surgery. No trials are currently evaluating Posurdex on AMD.

Given the complexity of AMD and our limited understanding of its pathophysiology, the current design of therapeutic modalities relies largely on the evaluation of a large number of candidate treatments in order to define those who offer clinical benefit. Until our knowledge of physiological processes responsible for macular degeneration increases, treatment will likely depend on a combination of approaches to limit the underlying CNV. This combination approach is already largely reflected in the design of currently ongoing clinical trials.

## REFERENCES


1. Kahn, H. A., Leibowitz, H. M., Ganley, J. P., et al. (1977), The Framingham Eye Study. I. Outline and major prevalence findings, *Am. J. Epidemiol.*, 106, 17–32.
2. Attebo. K., Mitchell, P., and Smith, W. (1996), Visual acuity and the causes of visual loss in Australia. The Blue Mountains Eye Study, *Ophthalmology*, 103, 357–364.
3. Klaver, C. C., Wolfs, R. C., Vingerling, J. R., et al. (1998), Age-specific prevalence and causes of blindness and visual impairment in an older population: the Rotterdam Study, *Arch. Ophthalmol.*, 116, 653–658.
4. Seddon, J. M. (2001), Epidemiology of age-related macular degeneration, in Ryan, S. J., Ed, *Retina*, Mosby, St Louis, pp. 1039–1050.
5. Friedman, D. S., O'Colmain, B. J., Munoz. B., et al. (2004), Prevalence of age-related macular degeneration in the United States, *Arch. Ophthalmol.*, 122, 564–572.
6. Bird, A. C., Bressler, N. M., Bressler, S. B., et al. (1995), An international classification and grading system for age-related maculopathy and age-related macular degeneration. The International ARM Epidemiological Study Group, *Surv. Ophthalmol.*, 39, 367–374.
7. Macular Photocoagulation Study Group (1982), Argon laser photocoagulation for senile macular degeneration. Results of a randomized clinical trial, *Arch. Ophthalmol.*, 100, 912–918.
8. Age-Related Eye Disease Study Research Group (2001), A randomized, placebo-controlled, clinical trial of high-dose supplementation with vitamins C and. E., beta carotene, and zinc for age-related macular degeneration and vision loss: AREDS report no. 8, *Arch. Ophthalmol.*, 119, 1417–1436.
9. Fliesler, S. J., and Anderson, R. E. (1983), Chemistry and metabolism of lipids in the vertebrate retina, *Prog. Lipid. Res.*, 22, 79–131.

10. Young, R. W. (1988), Solar radiation and age-related macular degeneration, *Surv. Ophthalmol.*, 32, 252–269.
11. Gerster, H. (1991), Review: Antioxidant protection of the ageing macula, *Age Ageing*, 20, 60–69.
12. Beatty, S., Koh, H., Phil, M., et al. (2000), The role of oxidative stress in the pathogenesis of age-related macular degeneration, *Surv. Ophthalmol.*, 45, 115–134.
13. Newsome, D. A., Swartz, M., Leone, N. C., et al. (1988), Oral zinc in macular degeneration, *Arch. Ophthalmol.*, 106, 192–198.
14. The Alpha-Tocopherol Beta Carotene Cancer Prevention Study Group (1994), The effect of vitamin E and beta carotene on the incidence of lung cancer and other cancers in male smokers, *N. Engl. J. Med.*, 330, 1029–1035.
15. Eagle, R. C. J. (1984), Mechanisms of maculopathy, *Ophthalmology*, 91, 613–625.
16. Young, R. W. (1987), Pathophysiology of age-related macular degeneration, *Surv. Ophthalmol.*, 31, 291–306.
17. Ho, A. C., Maguire, M. G., Yoken, J., et al. (1999), Laser-induced drusen reduction improves visual function at 1 year. Choroidal Neovascularization Prevention Trial Research Group, *Ophthalmology*, 106, 1367–1373.
18. Olk, R. J., Friberg, T. R., Stickney, K. L., et al. (1999), Therapeutic benefits of infrared (810-nm) diode laser macular grid photocoagulation in prophylactic treatment of nonexudative age-related macular degeneration: two-year results of a randomized pilot study, *Ophthalmology*, 106, 2082–2090.
19. National Eye Institute (2006), Complications of Age-Related Macular Degeneration Prevention Trial (CAPT); available from http://www.nei.nih.gov/neitrials/viewStudyWeb.aspx?id=70; accesed Jan. 14, 2006.
20. PTAMD Clinical Trial (2006); available from http://www.iridex.com/ophthalmology/ptamd_clinical_trial.html; accessed Jan. 15, 2006.
21. Friedman, E., Krupsky, S., Lane, A. M., et al. (1995), Ocular blood flow velocity in age-related macular degeneration, *Ophthalmology*, 102, 640–646.
22. Friedman, E. (1997), A hemodynamic model of the pathogenesis of age-related macular degeneration, *Am. J. Epidemiol.*, 124, 677–682.
23. Boyer, D., and Gallemore, R. (2006), Clinical trials assess rheophoresis; available from http://www.mdsupport.org/library/rheotri2.html; accessed Jan. 16, 2006.
24. Macular Photocoagulation Study Group (1986), Argon laser photocoagulation for neovascular maculopathy. Three-year results from randomized clinical trials. Macular Photocoagulation Study Group, *Arch. Ophthalmol.*, 104, 503–512.
25. Macular Photocoagulation Study Group (1986), Recurrent choroidal neovascularization after argon laser photocoagulation for neovascular maculopathy. Macular Photocoagulation Study Group, *Arch. Ophthalmol.*, 104, 503–512.
26. Macular Photocoagulation Study Group (1991), Argon laser photocoagulation for neovascular maculopathy. Five-year results from randomized clinical trials. Macular Photocoagulation Study Group, *Arch. Ophthalmol.*, 109, 110–114.
27. Macular Photocoagulation Study Group (1991), Subfoveal neovascular lesions in age-related macular degeneration. Guidelines for evaluation and treatment in the macular photocoagulation study. Macular Photocoagulation Study Group, *Arch. Ophthalmol.*, 109, 1242–1257.
28. Macular Photocoagulation Study Group (1993), Laser photocoagulation of subfoveal neovascular lesions of age-related macular degeneration. Updated findings from two clinical trials. Macular Photocoagulation Study Group, *Arch. Ophthalmol.*, 111, 1200–1209.

29. Ciulla, T. A., Harris, A., Kagemann, L., et al. (2001), Transpupillary thermotherapy for subfoveal occult choroidal neovascularization: effect on ocular perfusion, *Invest. Ophthalmol. Vis. Sci.*, 42, 3337–3340.
30. Algvere, P. V., Libert, C., Lindgarde, G., et al. (2003), Transpupillary thermotherapy of predominantly occult choroidal neovascularization in age-related macular degeneration with 12 months follow-up, *Acta. Ophthalmol. Scand.*, 81, 110–117.
31. Thach, A. B., Sipperley, J. O., Dugel, P. U., et al. (2003), Large-spot size transpupillary thermotherapy for the treatment of occult choroidal neovascularization associated with age-related macular degeneration, *Arch. Ophthalmol.*, 121, 817–820.
32. TTT4CNV Clinical Trial (2003), Iridex.
33. Aveline, B., Hasan, T., and Redmond, R. W. (1994), Photophysical and photosensitizing properties of benzoporphyrin derivative monoacid ring A (BPD-MA), *Photochem. Photobiol.*, 59, 328–335.
34. Allison, B. A., Waterfield, E., Richter, A. M., et al. (1991), The effects of plasma lipoproteins on in vitro tumor cell killing and in vivo tumor photosensitization with benzoporphyrin derivative, *Photochem. Photobiol.*, 54, 709–715.
35. Hunt, D. W., Jiang, H., Granville, D. J., et al. (1999), Consequences of the photodynamic treatment of resting and activated peripheral T lymphocytes, *Immunopharmacology*, 41, 31–44.
36. Reichel, E., Puliafito, C. A., Duker, J. S., et al. (1994), Indocyanine green dye-enhanced diode laser photocoagulation of poorly defined subfoveal choroidal neovascularization, *Ophthalmic. Surg.*, 25, 195–201.
37. Hope-Ross, M. W., Gibson, J. M., Chell, P. B., et al. (1994), Dye enhanced laser photocoagulation in the treatment of a peripapillary subretinal neovascular membrane, *Acta. Ophthalmol.*, (Copenh) 72, 134–137.
38. Moriarty, A. P. (1994), Indocyanine green enhanced diode laser photocoagulation of subretinal neovascular membranes, *Br. J. Ophthalmol.*, 78, 238–239.
39. Treatment of age-related macular degeneration with photodynamic therapy (TAP) Study Group (1999), Photodynamic therapy of subfoveal choroidal neovascularization in age-related macular degeneration with verteporfin: one-year results of 2 randomized clinical trials–TAP report, *Arch. Ophthalmol.*, 117101329–1345.
40. Bressler, N. M. (2001), Treatment of age-related macular degeneration with photodynamic therapy (TAP) Study Group. Photodynamic therapy of subfoveal choroidal neovascularization in age-related macular degeneration with verteporfin: two-year results of 2 randomized clinical trials-tap report 2, *Arch. Ophthalmol.*, 119, 198–207.
41. Blumenkranz, M. S., Bressler, N. M., Bressler, S. B., et al. (2002), Verteporfin therapy for subfoveal choroidal neovascularization in age-related macular degeneration. three-year results of an open-label extension of 2 randomized clinical trials–TAP Report no. 5, *Arch. Ophthalmol.*, 120, 1307–1314.
42. Verteporfin In Photodynamic Therapy Study Group (2001), Verteporfin therapy of subfoveal choroidal neovascularization in age-related macular degeneration: two-year results of a randomized clinical trial including lesions with occult with no classic choroidal neovascularization–verteporfin in photodynamic therapy report 2, *Am. J. Ophthalmol.*, 131, 541–560.
43. QLT Inc. Occult AMD (2006); available at http://www.qltinc.com/Qltinc/main/mainpages.cfm?InternetPageID=143; accesed Jan. 16, 2006.
44. Bressler, N., Rosenfeld, P., and Lim, J. (2003), VIM Study Group: A phase II placebo-controlled, double-masked, randomized trial—verteporfin in minimally classic CNV due to AMD (VIM), *Invest. Ophthalmol. Vis. Sci.*, 44, E-abstract 1100.

45. Azab, M., Boyer, D. S., Bressler, N. M., et al. (2005), Verteporfin therapy of subfoveal minimally classic choroidal neovascularization in age-related macular degeneration: 2-year results of a randomized clinical trial, *Arch. Ophthalmol.*, 123, 448–457.
46. Stur, M. (2004), VER STudy Group: Verteporfin Early Retreatment (VER)—12-month results of a phase IIIB controlled clinical trial, *Invest. Ophthalmol. Vis. Sci.*, 45, E-abstract 2275.
47. Slakter, J., and Rosenfeld, P. (2003), VALIO Study Group: Verteporfin with altered (delayed) light in occult CNV (VALIO)—Results of a phase II controlled clinical trial, *Invest. Ophthalmol. Vis. Sci.*, 44, E-abstract 1101.
48. Singerman, L., and Rosenfeld, P. (2004), VALIO Study Group: Verteporfin with altered (delayed) light in occult (VALIO)—12-month results of a phase II controlled clinical trial, *Invest. Ophthalmol. Vis. Sci.*, 45, E-abstract 2274.
49. Thomas, E. (2004), SnET2 Study Group: SnET2 photodynamic therapy for age-related macular degeneration. visual acuity efficacy outcomes from two parallel phase III trials, *Invest. Ophthalmol. Vis. Sci.*, 45, E-abstract 2214.
50. Yonemoto, L. T., Slater, J. D., Friedrichsen, E. J., et al. (1996), Phase I/II study of proton beam irradiation for the treatment of subfoveal choroidal neovascularization in age-related macular degeneration: treatment techniques and preliminary results, *Int. J. Radiat. Oncol. Biol. Phys.*, 36, 867–871.
51. Finger, P. T., Berson, A., Ng, T., et al. (1999), Ophthalmic plaque radiotherapy for age-related macular degeneration associated with subretinal neovascularization, *Am. J. Ophthalmol.*, 127, 170–177.
52. Radiation Therapy for Age-related Macular Degeneration (RAD) Study Group (1999), A prospective, randomized, double-masked trial on radiation therapy for neovascular age-related macular degeneration (RAD Study), *Ophthalmology*, 106, 2239–2247.
53. Marcus, D. M., Sheils, W. C., Young, J. O., et al. (2004), Radiotherapy for recurrent choroidal neovascularisation complicating age related macular degeneration, *Br. J. Ophthalmol.*, 88, 114–119.
54. Bergink, G. J., Hoyng, C. B., van der Maazen, R. W., et al. (1998), A randomized controlled clinical trial on the efficacy of radiation therapy in the control of subfoveal choroidal neovascularization in age-related macular degeneration: radiation versus observation, *Graefes Arch. Clin. Exp. Ophthalmol.*, 236, 321–325.
55. Char, D. H., Irvine, A. I., Posner, M. D., et al. (1999), Randomized trial of radiation for age-related macular degeneration, *Am. J. Epidemiol.*, 127, 574–578.
56. Marcus, D., Peskin, E., Alexander, J., et al. (2003), The age-related macular degeneration radiotherapy trial (AMDRT). 1 year results, *Invest. Ophthalmol. Vis. Sci.*, 44, E-abstract 3158.
57. Lambert, H. M., Capone, A. J., Aaberg, T. M., et al. (1992), Surgical excision of subfoveal neovascular membranes in age-related macular degeneration, *Am. J. Ophthalmol.*, 113, 257–262.
58. Gass, J. (1994), Biomicroscopic and histopathologic considerations regarding the feasibility of surgical excision of subfoveal neovascular membranes, *Am. J. Ophthalmol.*, 118, 285–298.
59. Ormerod, L. D., Puklin, J. E., and Frank, R. N. (1994), Long-term outcomes after the surgical removal of advanced subfoveal neovascular membranes in age-related macular degeneration, *Ophthalmology*, 101, 1201–1210.
60. Thomas, M. A., Dickinson, J. D., Melberg, N. S., et al. (1994), Visual results after surgical removal of subfoveal choroidal neovascular membranes, *Ophthalmology*, 101, 1384–1396.

61. Bressler, N. M., Bressler, S. B., Childs, A. L., et al. (2004), Surgery for hemorrhagic choroidal neovascular lesions of age-related macular degeneration: ophthalmic findings. SST report no. 13, *Ophthalmology*, 111, 1993–2006.
62. Hawkins, B. S., Bressler, N. M., Miskala, P. H., et al. (2004), Surgery for subfoveal choroidal neovascularization in age-related macular degeneration: ophthalmic findings. SST report no. 11, *Ophthalmology*, 111, 1967–1980.
63. Yi, X., Ogata, N., Komada, M., et al. (1997), Vascular endothelial growth factor expression in choroidal neovascularization in rats, *Graefes Arch. Clin. Exp. Ophthalmol.*, 235, 313–319.
64. Okamoto, N., Tobe, T., Hackett, S. F., et al. (1997), Transgenic mice with increased expression of vascular endothelial growth factor in the retina: a new model of intraretinal and subretinal neovascularization, *Am. J. Pathol.*, 151, 281–291.
65. Aiello, L. P., Pierce, E. A., Foley, E. D., et al. (1995), Suppression of retinal neovascularization in vivo by inhibition of vascular endothelial growth factor (VEGF) using soluble VEGF-receptor chimeric proteins, *Proc. Natl. Acad. Sci. USA*, 92, 10457–10461.
66. Eyetech Study Group (2002), Preclinical and phase 1A clinical evaluation of an anti-VEGF pegylated aptamer (EYE001) for the treatment of exudative age-related macular degeneration, *Retina*, 22, 143–152.
67. Eyetech Study Group (2003), Anti-vascular endothelial growth factor therapy for subfoveal choroidal neovascularization secondary to age-related macular degeneration: phase II study results, *Ophthalmology*, 110, 979–986.
68. Gragoudas, E. S., Adamis, A. P., Cunningham, E. T. J., et al. (2004), VEGF Inhibition Study in Ocular Neovascularization Clinical Trial Group: Pegaptanib for neovascular age-related macular degeneration, *N. Engl. J. Med.*, 351, 2805–2816.
69. FDA (2006), FDA Approves New Drug Treatment for Age-Related Macular Degeneration, *FDA news*; available at http://www.fda.gov/bbs/topics/news/2004/new01146.html; accesed Jan. 16, 2006.
70. Heier, J., Sy, J., and McCluskey, E. (2003), RhuFab V2 Study Group: RhuFab V2 in wet AMD—6 month continued improvement following multiple intravitreal injections, *Invest. Ophthalmol. Vis. Sci.*, 44, E-abstract 972.
71. Heier, J. S. (2005), Lucentis update in *American Academy of Ophthalmology Annual Meeting 2004, Retina Specialty Day Supplement*, Chicago.
72. Heier, J. S. (2004), Anti-VEGF: Genetech Ranibizumab in *American Academy of Ophthalmology Annual Meeting 2004. Retina Specialty Day*, New Orleans.
73. Takita, H., Yoneya, S., Gehlbach, P. L., et al. (2003), Retinal neuroprotection against ischemic injury mediated by intraocular gene transfer of pigment epithelium-derived factor, *Invest. Ophthalmol. Vis. Sci.*, 44, 4497–4504.
74. Stellmach, V., Crawford, S. E., Zhou, W., et al. (2001), Prevention of ischemia-induced retinopathy by the natural ocular antiangiogenic agent pigment epithelium-derived factor, *Proc. Natl. Acad. Sci. USA*, 98, 2593–2597.
75. Mori, K., Duh, E., Gehlbach, P., et al. (2001), Pigment epithelium-derived factor inhibits retinal and choroidal neovascularization, *J. Cell. Physiol.* 188, 253–263.
76. Campochiaro, P., Klein, M., Holtz, E., et al. (2004), AdPEDF therapy for subfoveal choroidal neovascularization (CNV): preliminary phase I results, *Invest. Ophthalmol. Vis. Sci.*, 45, E-abstract 2361.
77. Leahy, K. M., Ornberg, R. L., Wang, Y., et al. (2002), Cyclooxygenase-2 inhibition by celecoxib reduces proliferation and induces apoptosis in angiogenic endothelial cells in vivo, *Cancer Res.*, 62, 625–631.

78. Genaidy, M., Kazi, A. A., Peyman, G. A., et al. (2002), Effect of squalamine on iris neovascularization in monkeys, *Retina*, 22, 772–778.
79. Higgins, R. D., Sanders, R. J., Yan, Y., et al. (2000), Squalamine improves retinal neovascularization, *Invest. Ophthalmol. Vis. Sci.*, 41, 1507–1512.
80. Ciulla, T. A., Criswell, M. H., Danis, R. P., et al. (2003), Squalamine lactate reduces choroidal neovascularization in a laser-injury model in the rat, *Retina*, 23, 808–814.
81. AAO (2006), Genaera Presents Positive Preliminary Clinical Results for EVIZON for Treatment of Age-Related Macular Degeneration at the Annual AAO Meeting; available at http://www.genaera.com/pressreleases/October%2019,%202005.pdf; accesed 2006 Jan. 16, 2006.
82. Folkman, J., and Ingber, D. E. (1987), Angiostatic steroids. Method of discovery and mechanism of action, *Ann. Surg.*, 206, 374–383.
83. Ishibashi, T., Miki, K., Sorgente, N., et al. (1985), Effects of intravitreal administration of steroids on experimental subretinal neovascularization in the subhuman primate, *Arch. Ophthalmol.*, 103, 708–711.
84. Ciulla, T. A., Criswell, M. H., Danis, R. P., et al. (2001), Intravitreal triamcinolone acetonide inhibits choroidal neovascularization in a laser-treated rat model, *Arch. Ophthalmol.*, 119, 399–404.
85. Penfold, P. L., Gyory, J. F., Hunyor, A. B., et al. (1995), Exudative macular degeneration and intravitreal triamcinolone. A pilot study, *Aust. N. Z. J. Ophthalmol.*, 23, 293–298.
86. Challa, J. K., Gillies, M. C., Penfold, P. L., et al. (1998), Exudative macular degeneration and intravitreal triamcinolone: 18 month follow up, *Aust. N. Z. J. Ophthalmol.*, 26, 277–281.
87. Danis, R. P., Ciulla, T. A., Pratt, L. M., et al. (2000), Intravitreal triamcinolone acetonide in exudative age-related macular degeneration, *Retina*, 20, 244–250.
88. Ranson, N. T., Danis, R. P., Ciulla, T. A., et al. (2002), Intravitreal triamcinolone in subfoveal recurrence of choroidal neovascularisation after laser treatment in macular degeneration, *Br. J. Ophthalmol.*, 86, 527–529.
89. Gillies, M. C., Simpson, J. M., Luo, W., et al. (2003), A randomized clinical trial of a single dose of intravitreal triamcinolone acetonide for neovascular age-related macular degeneration: one-year results, *Arch. Ophthalmol.*, 121, 667–673.
90. Heriot, W. (2004), Corticosteroids for AMD in *American Academy of Ophthalmology Annual Meeting 2004. Retina Subspecialty Day*, New Orleans.
91. Crum, R., Szabo, S., and Folkman, J. (1985), A new class of steroids inhibits angiogenesis in the presence of heparin or a heparin fragment, *Science*, 230, 1375–1378.
92. Clark, A. F. (1997), AL-3789: a novel ophthalmic angiostatic steroid, *Expert Opin. Investig. Drugs*, 6, 1867–1877.
93. D'Amico, D. J., Goldberg, M. F., Hudson, H., et al. (2003), Anecortave acetate as monotherapy for treatment of subfoveal neovascularization in age-related macular degeneration: twelve-month clinical outcomes, *Ophthalmology*, 110, 2372–2383.
94. Slakter, J. S., Bochow, T. W., D'Amico, D. J., et al. (2006), Anecortave acetate (15 milligrams) versus photodynamic therapy for treatment of subfoveal neovascularization in age-related macular degeneration, *Ophthalmology*, 113, 3–13.
95. Slakter, J. S. (2005), Retaane Update in *American Academy of Ophthalmology Annual Meeting 2005. Retina Specialty Day*, Chicago.
96. Ciulla, T. A., Criswell, M. H., Danis, R. P., et al. (2003), Choroidal neovascular membrane inhibition in a laser treated rat model with intraocular sustained release triamcinolone acetonide microimplants, *Br. J. Ophthalmol.*, 87, 1032–1037.
97. Anon. (2005), Fluocinolone acetonide ophthalmic–Bausch & Lomb: fluocinolone acetonide Envision TD implant, *Drugs R. D.*, 6, 116–119.

# *Paediatrics*

**Anne Cusick,[1] Natasha Lannin,[2] and Iona Novak[3]**

[1]*School of Biomedical and Health Sciences, University of Western Sydney, Sydney, Australia*
[2]*Rehabilitation Research Studies Unit, Faculty of Medicine, University of Sydney, Sydney, Australia*
[3]*Cerebral Palsy Institute, Sydney, Australia*

**Contents**

1 Introduction
2 Definitions
  2.1 Paediatric Population
  2.2 Off-Label Use
  2.3 Therapeutic Orphan
  2.4 Paediatric Investigation Plan
  2.5 Minimal Risk
3 Overview of Unique Aspects of Paediatric Trials
4 Conventions
5 Benchmark Regulations
6 Recommendations and Guidelines
7 Institutional Review Boards
8 Trial Questions
9 Participant Characteristics
10 Paediatric Investigation Plans
11 Assent and Consent
12 Safety and Monitoring
13 Conclusion
  Appendix: Facts Summary
  References

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

## 1 INTRODUCTION

This chapter introduces then explores in detail, issues and practical aspects of the context and conduct of paediatric clinical trials. Investigators need to be alert to the unique obligations that come into play when study participants are infants, children or young people. These obligations are ethical, procedural, legal and social. Investigators also need to be aware of the multiplicity of interests that operate in paediatric research so that projects can be successfully managed to completion. Ideally, realistic and reasonable paediatric trials will cause minimal harm and will meet expectations of parents, caregivers, investigators, sponsors and the public for meaningful and clinically useful outcomes.

This chapter presents the scope of challenges and opportunities in the paediatric speciality, to help support clinical trials that are child-centred, worthwhile and rigorous. Paediatric research is an area characterised by local procedural variation and rapid change in regulation, case-law, scientific evidence, scholarly opinion, public concern and professional guidelines. Consequently, this chapter provides common signposts, rather than prescriptive maps, that can be set in place by investigators as they lead the way in particular trial journeys. By the end of this chapter, investigators should have a breadth of view on issues involved in paediatric trials to enable them to apply technical information to the context of paediatrics. Investigators should also be able to reflect on their own standpoint and responsibility as trial leaders, sponsors or team members who are knowingly putting infants, children or young people at some level of risk in order to answer a question. Investigators should be able to seek out specialized paediatric trial sources after reading this chapter, having first gained a broad understanding of issues that must be proactively managed for successful trial completion.

## 2 DEFINITIONS

### 2.1 Paediatric Population

The United Nations Convention on the rights of the child [1] defines a minor as anyone under the age of 18. While there are local variations to the legal age of adulthood, and while the clinical responsibility of the paediatric speciality may extend to 21 [2], the UN Convention should normally be observed for the purposes of trial research. For this chapter, the previable or viable foetus is not included however these are recognised as responsibilities of paediatrics [2] and specialist research sources should be consulted commencing with policy and guideline statements of relevant bodies for example, the American Academy of Pediatrics [3].

### 2.2 Off-Label Use

Many drugs used in paediatrics are off-label and one reason underpinning the need for clinical trials is to reduce this practice:

> "new drugs and biologicals [need to] include adequate pediatric labeling for the claimed indications at the time of, or soon after, approval. However, because such labelling may

not immediately be available, off-label use (or use that is not included in the approved label) of therapeutic agents is likely to remain common in the practice of pediatrics ... The purpose of off-label use is to benefit the individual patient. Practitioners may use their professional judgment to determine these uses ... The off-label use of a drug should be based on sound scientific evidence, expert medical judgement, or published literature" [4, p. 181].

### 2.3 Therapeutic Orphan

Drugs which are not approved by the Food and Drug Administration [USA] as safe and effective in children are prescribed daily. This is due in part to the fact that many drugs released since 1962 carry an "orphaning clause" in the package insert such as, "not to be used in children, since clinical studies have been insufficient to establish recommendations for its use ... Is the physician breaking the [USA] law when he prescribes drugs ... which carry the 'orphaning clause'?. No, he is not. The physician may exercise his professional judgement in the use of any drug. However, if he deviates from the instructions in the package insert and adverse reactions occur, he must be prepared to defend himself in court if there is a malpractice suit" [5, p. 811].

### 2.4 Paediatric Investigation Plan

This is the term used in the European Union by the European Medicines Agency (EMEA).

[It is] a development plan aimed at ensuring that the necessary data are obtained through studies in children, when it is safe to do so, to support the authorisation of the medicine for children. ... The paediatric investigation plan includes a description of the studies and of the measures to adapt the way the medicine is presented (formulation) to make its use more acceptable in children ... The plan should cover the needs of all age groups of children, from birth to adolescence. The plan also defines the timing of studies in children compared to adults. In some cases, studies will be deferred until after the studies in adults have been conducted, to ensure that research with children is done only when it is safe and ethical to do so [6].

### 2.5 Minimal Risk

Minimal (the least possible) risk describes procedures such as questioning, observing and measuring children, provided that procedures are carried out in a sensitive way, respecting the child's autonomy and that consent has been given ... It is expected that research of minimal risk would not result in more than a very slight and temporary negative impact on the health of the person concerned [7, p. 15].

## 3 OVERVIEW OF UNIQUE ASPECTS OF PEDIATRIC TRIALS

This section overviews key issues that make paediatric clinical trials unique. Paediatric clinical trials seek answers about intervention effectiveness and safety—there is nothing unique about this—but to get answers, these trials are unique in requiring infants, children and youth as participants. The sample characteristics and needs thus dominate trial decisions. Paediatric populations present design challenges for trial

investigators. As participants they have inherent and continuous change in body structures, function, activities and participation, there is constant change in their social relations, exposure to physical environments, family and community influences—this continuous change is evident even when they are in good health. The scale and variability of change increases if there is illness, disease, disability or injury. Continuous underlying change in participants must therefore be anticipated in trial question, design and protocol decisions.

Paediatric populations also bring investigator responsibilities and accountabilities that extend well beyond those normally encountered in trials with adults. Young participants are inherently *vulnerable* and those who have illness or disease appear particularly exposed to the possibility of suffering. Adults make trial decisions on behalf of youngsters and act towards them in ways that may help or harm both their daily life and their life chances. Their vulnerability presents dilemmas for all involved in clinical trials. For adults to purposefully involve youngsters in studies with potential or known risk seems incompatible with obligations to protect and nurture them. For institutions, like hospitals, established to care and help, knowingly exposing youngsters to risk seems a betrayal of civic duty. These moral contradictions are a necessary and inevitable part of paediatric trials. They underpin the heightened emotion and public interest that accompanies paediatric trials. But without sound paediatric clinical trials greater harms may be perpetrated as parents, carers, medical and health personnel use clinical interventions on children at large without adequate or sometimes any scientific evidence to inform their decisions.

Clinical trials must be conducted, but their planning and accountabilities must also anticipate and accommodate paediatric participant vulnerability. Investigators who are aware of this *moral context* will ensure that questions are not only worthwhile, but also that protocols are explicitly child-centred, of rigorous design, and well conducted. It is the heightened moral context of paediatric clinical trials that is unique as infants, children and young people, not investigators, parents or sponsors, must live with the trial experience and the short and long term consequences of participation. Trial investigators who are aware of the moral context will also anticipate potential public interest and consider in advance how the layperson might construe issues such as recruitment, design, funding, sponsorship, personnel and study procedure. Paediatric trial project planning thus needs to include strategies for public communication, public liaison and accountability trails.

The inherent vulnerability also means that a paediatric trial is one where child participants are the focus of many interested parties. *Stakeholders* include parents and guardians, medical, legal, health, and policy personnel who act as gatekeepers to and watchdogs of paediatric population participation. The media, pharmaceutical companies, organised advocacy and lobby groups play vigilant and vigorous roles in the initiation, public presentation and interpretation of trials. Stakeholder authority and influence can make the planning and conduct of paediatric trials a complex and delicate campaign of personal, professional, industry and community politics—in addition to the usual demands of a complicated scientific project. Proactive communication with stakeholders is needed regarding the question, study rationale, study conduct, regulation adherence, avoidance of conflict of interest, transparency and accountability of processes and records, and importantly the need for trials to

prevent current and future suffering caused by the use of interventions that have no scientific support.

Paediatric trials necessitate a careful balance of common sense, technical precision, and adherence to highly prescriptive regulation. Given the sensitivity of research with children, successfully engaging in paediatric trials is as much about identifying a *moral purpose*, communicating trial values and visions, ensuring justice for participants, and scanning the strategic context for threats and opportunities as it is about the study question, procedural diligence, protocol development and project management.

These "soft" project factors can publicly or professionally derail an otherwise well intentioned and well constructed trial if not proactively managed. They can destroy an investigator's reputation even when no malicious intent or negligence was involved. The "front page of the paper" by-line may be all a public needs to discredit a researcher, damage institutional confidence and undermine a much-needed program of paediatric health research. Careful consideration of not only the scientific merit of the question, design and regulatory obligations, but also of the moral *politics* of paediatric trials is thus required. For some investigators, the interests and authority of stakeholders and gatekeepers seem like obstacles to research, however any experienced trial researcher knows these factors are just part of the "package" that is a paediatric trial.

Paediatric practice is also inherently *multidisciplinary* and the design of clinical trials must be robust enough to anticipate an array of potentially confounding factors in a child's life emanating from health services, schools, community and family, including parental use of off-the-shelf medications, complementary medicines and alternative therapies. A clinical trial that is child-focused rather than variable focused will take into account the multiplicity of influences that may affect body structures, ability to participate in protocol requirements, recruitment, retention, confounding factors, and the short and long term consequences of trial participation.

Finally, participation, protocol adherence, study retention and the safety of young participants ultimately relies on *parent and caregiver* expertise and their understanding of and commitment to trials. Engaging parents and caregivers as protocol partners is critical to study success. Providing appropriate learning opportunities for parents to understand not only the study purpose and participants demands but also children's rights to assent or decline participation are needed. This understanding goes to the heart of informed consent. Consideration of parent and caregiver perspectives in protocol design is essential, particularly in relation to the logistic and time demands on parents for intervention adherence and presentation of the child for outcome data measure collection. Paediatric clinical trials work better if the protocol can be reasonably integrated into the daily life of families as part of a sustainable routine.

This section of the chapter has provided paediatric investigators with an orientation to unique issues in paediatric trials. The following sections explore issues in more depth, however the following *caveats* apply. This chapter is introductory. Specialized paediatric sources, such as the Helms and Stonier (Eds) *Pediatric Clinical Research Manual* [8] which is regularly updated and has sub-speciality supplements, or the many speciality research related web-sites of regulatory bodies or professional societies should also be consulted. Research, scholarship policy and

commentary associated with paediatric research changes quickly. Investigators must therefore ensure they inform themselves about local requirements, contemporary conventions, up-to-date regulation, current public concerns, scholarly opinion and relevant scientific evidence available at the time of study commencement. A matter of months can dramatically change the paediatric research context as case law, public interest, media "frenzies", local administration or new evidence can change what could be reasonably expected in an investigation plan. Particular attention should also be given to the most recent scientific evidence emerging from the paediatric sub-specialty under study, be that oncology, cerebral palsy, infectious disease or whatever, including evidence relating to effective sub-speciality trial methodologies. Different sub-specialities bring unique challenges relating to measurement, recruitment, consent and retention. Successful sub-speciality precedent studies can also help inform trial plan decisions. But caution! Precedent studies may have been conducted in different regulatory and social contexts making their methodologies unsuitable for contemporary replication even though the scientific findings may be rigorous and relevant.

The remainder of the chapter begins with the policy context. Policies have been developed to protect vulnerable infants, children and young people in research and to promote ethical practice. Conventions, regulations, recommendations, guidelines, and institutional review boards are introduced. The chapter continues with an exploration of issues that are particularly challenging in paediatrics: trial questions, participant characteristics, investigation plans, consent, assent, safety and monitoring.

## 4 CONVENTIONS

Regulatory requirements compel investigators, sponsors and research partners to scope and conduct research in certain ways. Investigators and trial partners must comply with regulations or face consequences that may include prosecution and penalties ranging from public "naming and shaming" to fines or in cases of criminal conduct imprisonment. The existence of regulations means that clinical research can not only be weak or rigorous, it can also be lawful or unlawful. Investigators and members of institutional ethical review boards therefore have a duty to keep up-to-date with regulatory requirements as ignorance is normally no excuse. Understanding legal obligations, particularly those for paediatric populations, is as important as understanding scientific methods and the needs of youngsters. Investigators cannot work on the basis of what has been done in past practice or research, nor can they apply the same level of autonomous discretion they may use in their clinical life, as research-related regulations change over time, the standards for research decision-making are more prescribed and tests of due diligence, fair hearing and procedural fairness in research may be tighter.

A good place to start in understanding the regulatory context of paediatric research is with landmark conventions. These state agreed international positions on matters of importance that relate to the human condition. The United Nations (UN) General Assembly "Convention on the Rights of the Child" 1989 is probably one of the most important foundations for the paediatric speciality [1]. Although

not all countries have ratified this convention, the influence of the convention on national standards is enormous. In summary, the convention identifies that: human rights apply to children without exception; the child's best interests are the primary consideration and highest priority; children have a right to the highest attainable level of health; and they have a right to information and respect of their opinion [1]. While clinical trial plans would almost never cite the Conventions on the Rights of the Child as a methodological source and there is no mechanism to directly register trials as convention compliant with the UN, it is principles in this convention that local regulations around the world usually aim to embed and enforce.

Practically, investigators can use this convention to reflect on whether or not their study question or plan is "just". Investigators have a duty to act justly towards participants [9]and tests of justice may go beyond local regulatory requirements. The convention can provide some insight into what might reasonably be considered just. If one accepts the principle that children everywhere in any society should have human rights, then just study protocols should seek to preserve and protect those rights. Just protocols will thus include strategies to inform, listen to the opinion of and seek assent of child participants, even when parental permission has already been granted. The principle of "best interests of children", if accepted, also means that the best interests of child participants and children in general are high priorities in study decisions. In some way children must benefit, whether that is directly through trial participation or indirectly through study outcomes that enhance the well-being of children in general. Finally, the just trial will support the principle that children have a right to the highest attainable health, particularly in relation to weighing up trial benefits and risks to individual children and children in general.

Other international agreements relating to research may also apply to paediatric investigation. The most notable is the Declaration of Helsinki [10]. Clause 25 specifically relates to child involvement in medical research and it focuses on consent and assent. The United Nations Convention on the Rights of Persons with Disabilities [11], and the Declaration on the Rights of Indigenous Peoples [12] may also be relevant for studies that have targeted or incidental recruitment of youngsters from these groups. Both these conventions are relatively new, and not all countries are signatories, but again they provide a benchmark for investigators to consider whether or not the study question and plan is "just".

Conventions are thus an important and a useful background to investigator development of an ethical standpoint towards young participants. They help identify what might be considered just treatment. But most investigators do not use them. Instead they follow local regulations and procedures that codify ethical requirements. Local regulations may or may not be adequate for the moral context of paediatric clinical trials. Here is where the utility of "benchmark" regulations comes into play for investigators around the world. Although they may not apply locally, benchmark regulations provide guidelines and procedures that usually embed principles from conventions or declarations in their construction. Benchmark regulations can thus act as a guide, along with local requirements, for investigators to consider what a reasonable person would expect in a just paediatric study. The benchmark regulations of most influence are now explored. Both relate to medicines for children however the principles and processes are useful to inform research-

ers who work with other clinical interventions as they highlight the need for clear standards, accountabilities and procedural precision.

## 5 BENCHMARK REGULATIONS

One of the most significant developments in Twenty-first century paediatric clinical research has been the release of regulations in the European Union (EU) and the United States of America (U.S.). While other countries have local standards and regulations that must be consulted by investigators and adhered to in plan development and reporting, the sheer scale of the EU and U.S. regulations impact on numbers of trials and participants makes them global benchmarks that can inform and guide clinical trial decisions anywhere. This section first overviews where regulations and related sources are located, as any paediatric investigator will need to continually update and check rulings and applications of regulations. Then the EU and USA regulations themselves will be introduced, and examples of regulations from other jurisdictions that may be helpful will be provided.

The regulations and support material of the EU are easily located. A web-search using any popular search engine and the general term "paediatric clinical trial" will reveal links to the European Agency for the Evaluation of Medicinal Products (also known as European Medicines Agency, EMEA) (http://www.emea.eu) in addition to independent sites that hold related articles, opinion pieces, conferences and training announcements on the general topic of paediatric clinical trials and often the EMEA initiative specifically. These related items can be EMEA sponsored, or independent and they often maintain a lively watch on the EMEA initiative from an investigator and industry perspective. Sites such as these, together with subspeciality resources in particular fields available through scholarly journals, professional societies and consumer groups, help investigators understand the motivations, agendas, obligations and impacts of the EU regulation from a broader perspective. The latter can help inform the strategic decisions that need to be made by trial leaders in project planning and management. From the home page of the EMEA Medicines for Children a wealth of official resources is also available to investigators including the regulation itself, guidance for applicants, access to scientific advice, decisions and opinions on applications and importantly, paediatric related information. The latter includes information on paediatric needs, clinical trials, priority list off-patent medicines and presentations. Importantly, the decisions and opinions on particular trial applications, including class waivers and product-specific decisions are included on this site.

The EU *Regulation (EC) No 1901/2006 amended*, the "Paediatric Regulation", came into force in January 2007, and investigators, industry and the public are still exploring the material effect it may have on paediatric research activity. The paediatric regulation aims to increase the number and availability of medicines that can be used in the paediatric population, by providing rulings, guidance and incentives to investigators, sponsors and institutions to develop paediatric specific products and to develop paediatric prescribing information for other medicines. Mechanisms to license and provide clinician and parent information are also included. One of the features of the Paediatric Regulation is the establishment of a new Paediatric Committee of the European Medicines Agency [13]. This scientific committee has

the authority to make decisions and provide opinions on applications to do research and on outcomes of that research in relation to product use. The Committee is multi-disciplinary, brings together renowned experts in fields of general practice, paediatric medicine, pharmacy, pharmacology, research, pharmacovigilance, ethics and public health. Health care professionals and patient associations are also part of the collective expertise.

The Paediatric Committee has an onerous task to ensure paediatric medicine approval is based on rigorous quality, safety and efficacy data. Strategies used by the Committee include: requiring paediatric investigation plans (PIPs) and data to be submitted to regulatory authorities; assessing PIPs and providing decisions and opinions; monitoring of the PIP compliance; supporting a paediatric research network; implementating key public communication strategies such as the use of a common symbol for medicines that have an approved paediatric use; and training investigators. In addition the regulations provide an incentive to investigators and sponsors by providing an additional six months on the supplementary protection certificate if completed PIP information is included in the Summary of Product Characteristics. For off-patent products, there is the incentive of a paediatric use marketing authorisation which has a ten year data and market protection period. The regulation also provides for establishment of a European data base of paediatric clinical trials, part of which will be publicly available.

Access to U.S. regulations is less straightforward. There are multiple web-routes to access relevant information, and the best route will depend upon the study question. One is to go direct to the U.S. Department of Health & Human Services (HHS) (http://www.hhs.gov), thence to the Office for Human Research Protections (OHRP) (http://www.hhs.gov/ohrp/). Another is to start with the U.S. Food and Drug Administration (FDA) (http://www.fda.gov) and consult the various pages to do with clinical trial practice that may include adults as well as children (http://www.fda.gov/cder/pediatric). One page, for example, provides direct links to all FDA regulations relating to good clinical practice and clinical trials (http://www.fda.gov/oc/gcp/regulations.html). There is also the Office of Pediatric Therapeutics (http://www.fda.gov.oc.opt/).

The history and recent state of USA regulations that inform paediatric research has been reviewed by Diekema [14]. He provides a concise guide to critical incidents and key sources, most notably the *Code of Federal Regulations, 45 CFR 46, Subpart D Additional Protections for Children Involved as subjects in research* 1983. Others are the *Best Pharmaceuticals for Children Act* 2002 and 2007 [15], and the *Pediatric Research Equity Act* 2003 [16] that re-established the FDA's authority to mandate paediatric drug development [17]. Informative guides and regular updates on paediatric issues have been developed and are available through HHS sites of the OHRP and FDA. An example is the *Guidance for Clinical Investigators, Institutional Review Boards and Sponsors: Process for Handling Referrals to FDA under 21 CFR 50.54; Additional Safeguards for Children in Clinical Investigations* [18]. Apart from linked guidelines, these sites provide "current thinking" of the agencies in relation to key issues variously presented as "frequently asked questions" or "guidelines". Further they provide aide-memoirs for investigators such as the "pediatric points to consider" that include a summary of unique review concerns for paediatrics, including study justification, study design, ethical issues, and pediatric protocol checklists. Later stage trial results should be reported to the clinical trial registry

and data base in accord with Regulation S3807. There are also incentives in regulations to encourage paediatric pharmaceutical research, such as six month exclusivity on manufacturer marketing licensing if the company "fairly responds" to FDA requests.

While the OHRP site provides guidance and recommendations for paediatric study conduct, it also highlights areas where compliance is required under the U.S. HHS Regulation. Compliance is specifically required for HHS supported research, but the aspects identified may provide useful benchmarks to alert investigators outside the HHS or USA for potentially high stake paediatric trial issues. As thinking may change and as there is acknowledgement that alternative approaches may be considered by the authorities, investigators are advised to check these sites for changes, opinions and rulings as part of trial planning.

Internationally, many countries have regulations that apply to research with human subjects and specifically to paediatric populations. Most feature principles and practice requirements or recommendations that are consistent with the broad approach of the EU or U.S. and related international conventions. Australia for example, has the National Health and Medical Research Council statements including the *Australian Code for the Responsible Conduct of Research* [19] and the *National Statement on Ethical Conduct in Human Research* [20]. Canada has the Medical Research Council Tri-Council policy statement on *Ethical Conduct for Research Involving Humans* [21]. Some other countries have statements on the conduct of research in humans, but have little that specifically relates to paediatric populations. For example, India has the *Ethical Guidelines for Biomedical Research on Human Participants* [22]; and South Africa has a *Code of Research Ethics* [23]. Both have only limited clauses relating to child research.

In addition to benchmark and local regulations for paediatric research in general, investigators need to be alert to any special provisions for particular groups. In Australia, for example, there are particular guidelines and requirements for research involving indigenous people [24]. Other countries may have similar provisions that should be consulted and integrated into PIPs. There are also special provisions in Australia for research that is conducted outside the country [25] and similar approaches may be taken in other nations. The particular vulnerabilities of unaccompanied children, foster children, wards of the state, and emancipated children also need attention if they are to be target or incidental participants. A lack of attention to any of these regulations can have a significant impact on the conduct of a trial or the reputation of investigators, sponsors or institutions involved.

## 6 RECOMMENDATIONS AND GUIDELINES

In addition to regulatory requirements and recommendations from statutory or government sponsored institutions, trial investigators, coordinators and employees may also be need to be registered or accredited professionals, subject to the Codes of Conduct, recommendations, guidelines or prescriptions of their professional societies or registration Acts. Depending on local conditions these may be enforceable. Each paediatric speciality and sub-speciality may also have requirements, guides

and resources. Investigators are urged to consult these at an early stage. Investigators should also satisfy themselves that they and their team members are meeting relevant professional obligations in addition to prescribed regulations, as the conduct of some assessment, intervention or outcome measure procedures may require a registered or accredited practitioner. Examples of professional societies that have guidelines for research are: the Royal College of Paediatrics and Child Health [26]; the European Academy of Paediatrics (formerly the Confederation of European Specialists in Paediatrics) [27] that provides guidance ranging from official statements to commentaries and summary presentations, for example that provided by Kurz [28]; and the American Academy of Pediatrics [3]. Profession and subspeciality specific guides must be sought out by investigators to inform trial decisions, and if none exist, this should be noted somewhere in trial planning records so that the due diligence of trial leaders in this regard can be noted.

There are also guidelines from esteemed practice and research institutes that could be considered and used as methodological sources in trial plans. One is the International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH). This organisation has produced many resources including the oft-cited, *Guideline for Good Clinical Practice E6 (R1)* [29] which applies to the conduct of clinical trials, including essential documentation and archive guidelines; and the guideline for *Clinical Investigation of Medicinal Products in the Pediatric Population E11* [30], that contains a series of guidelines for drug development and registration processes. These supplement more general ethical guidelines for biomedical research [31]. The Medical Research Council [7] provides an ethics guide for *Medical Research Involving Children* that clearly summarizes key points unique to paediatric trials. Guidelines for international research [25, 31], and research involving human participants in developing societies may also be relevant to investigators working across international boundaries.

Guidelines from esteemed bodies can relate to particular paediatric populations and sub-specialities. For example, the Society for Adolescent Medicine has issued *Guidelines for Adolescent Health Research* [32]. Others such as the National Institute for Clinical Excellence (NICE) (http://www.nice.org.uk/) provide specialist resources on "best practice" approaches that can be incorporated into research protocols and general care of young people. While NICE guidelines cover public health, health technologies (including medicines, treatments and procedures) and clinical practice (for specific diseases and conditions) applicable within the National Health Service of the United Kingdom, they are useful practice benchmarks. Examples of approved guidelines include: *Improving Outcomes with Children and Young People with Cancer* [33] and *Feverish Illness in Children—Assessment and Initial Management in Children Younger than 5 years* [34]. Other guideline projects are underway for children. For example, *Prevention of Unintentional Injury in Children under 15* (due April 2010) and *Guidance on Looked After Children* (due September 2010). While there are no obligations on researchers to observe these guidelines in trial plans, they are useful practice benchmarks which may be required when arguing the case for equipoise in intervention or control groups.

In addition to professional society and advisory bodies, there are industry developed guidelines and issues papers. One example is the Association of the British Pharmaceutical Industry publication *Current Issues in Paediatric Clinical Trials* [35]

that reported on a conference covering matters such as regulation, ethics, parent perspectives, national frameworks and the industry perspective.

## 7 INSTITUTIONAL REVIEW BOARDS

Institutional review boards are often referred to in regulation as the local means to assess, approve and monitor studies. New investigators generally focus their approval efforts on these local bodies, and may not initially be aware of the recommendations, guidelines, regulations and conventions referred to earlier. Paediatric trial investigators should, however, set their views beyond local board requirements. A local compliance approach that breaches standards set in conventions or benchmark regulations may not be considered "reasonable" or "just" in the mind of the public or the law if a trial goes awry.

Notwithstanding the need for a broad view by investigators, many local boards are subject to regulations and guidelines at a national level and this may make local compliance adequate. For example in Australia, Human Research Ethics Committees are established through a formal process involving registration with the Australian National Health and Medical Council. Their conduct and reporting lines are mandated. In another example, local Institutional Review Boards in the U.S. have limits to their approval authority indicated by the level of risk and benefit incurred by research participants [14, 36]. So in some areas, there may be natural links from local boards to national regulations and international conventions, but this should not be assumed.

Institutional review boards that consider and approve paediatric trial research need to ensure that they have access to appropriate expertise to make informed decisions about research with youngsters. Local boards that approve paediatric research may later be on the defensive if problems arise and they did not have adequate paediatric expertise in place to make their decisions robust. Barret [37] for example, suggests that research ethics committees should have members with practical experience in working with sick children so that they can assess whether or not risks to young participants are acceptable, protocols are workable, opportunities are provided for children to withdraw, and whether their autonomy is respected. Ethics committees also need to consider whether the research team has the paediatric capacity to do the study. Drawing on ICH [30], Kurz and Gill [38] recommend that paediatric trials should be done by "medical and scientific personnel who are familiar with GCP [good clinical practice] guidelines and are capable of a trusting relationship and communication with the child and parents ... in ... a child-friendly atmosphere with a paediatric infrastructure and personnel" (p. 43).

While ethics committees can interrogate researcher profiles as part of the approval process, local boards may make assumptions based on researcher reputation that is not backed up by documentation in the applications. This can easily happen if the researcher is a local who is well known for their expertise, or if a leading international researcher is involved and local boards do not feel comfortable interrogating his or her expertise. This can create later problems for boards if the appropriateness of the decision to approve research is challenged and the grounds for the decision about researcher capacity were scant. Paediatric investigators should therefore take some care in preparing the expertise statement that most institutional review board

applications require. It may be worthwhile to detail paediatric clinical, professional and research expertise relevant to the study topic, and to demonstrate that, between all team members, there is capacity for all study demands. In Australia, many institutions also require risk assessment prior to or following institutional review board approval for trial insurance purposes. It is not uncommon, for example, for subject and/or study-specific insurance to be required by some institutions for invasive clinical trials. The risk assessment outcome may depend upon the demonstrated paediatric research capacity of the investigation team.

## 8 TRIAL QUESTIONS

Trial questions need to be worthwhile and ethical. They need to be "honest and valid" [39, p. 836]. The technical skills involved in the development of trial research questions have been covered elsewhere in this handbook, as have the general ethical principles of beneficence, malfeasance and clinical equipoise, so this discussion will explore what principles underpin a worthwhile ethical paediatric trial question.

Paediatric clinical trials ask questions about the safety and efficacy of interventions for infants, children or youth. Kurz and Gill [38] proposed that paediatric should therefore "focus on the knowledge, cure, relief, or prevention of diseases of children. Biomedical studies must be devoted to reducing suffering and improving the prognosis of diseases" (pp. 42–43). To do this ethically, questions need to meet tests of relevance, benefit, originality, achievability, timing, and minimal harm. These are now explored.

The first issue facing investigators is whether or not the question is *relevant* to children and thus whether the involvement of young participants is essential rather than desirable. A high threshold needs to be met in this regard. The Royal College of Paediatrics and Child Health [26] identifies that research should only be conducted in children if the question cannot be answered by studying adults. Further, it identifies that involvement of children in clinical research will be required when an illness or condition only occurs in children, or has features that are more pronounced or have greater impact in young people. In such instances, research will be needed if there is no treatment information available or when what is known is inadequate. Paediatric research participation will also be required if the condition is in the general population but there is no paediatric treatment or there is only adult intervention evidence available [26]. Generally, paediatric inferences should not be drawn from adult studies, although occasionally a treatment will have a long history of use in children that initially relied on adult data but has subsequently been complemented by consensus expert paediatric opinion [26]. In such instances, the weight of consensus expert opinion may mean that an exception can be made and a paediatric inference can be drawn. The Medical Research Council [7] provides five questions for researchers to help determine whether or not the involvement of children is essential or whether findings from adult studies would be adequate. These questions cover [7]: age specificity, developmental understanding, implications for pharmacokinetics, applicability of adult-style therapy and issues of later life disease prevention.

The second issue facing investigators is whether or not answers to the trial question will *benefit* children. Benefit to children is essential [30]. The Royal College of

Paediatrics and Child Health [26] identifies that paediatric research must not only be well designed and well conducted but it must have a real prospect of benefiting children. Benefits to children may be direct though trial participation in treatment or control groups or indirectly to children in general—even though there may not be benefits to individual trial participants. Benefit needs to be self-evident in the trial question and study description [40]. The Royal College of Paediatricians proposed that the following issues help make potential benefits to children clear [26]: *magnitude* of the condition including severity, how *common* it is and how findings will be used; how *probable* it is that research will achieve aims; who specifically will *benefit* from the research (whether it is child participants or children in general); whether benefit to children will be *limited* because treatment is expensive or hard to deliver; the *type* of intervention and whether a less invasive one could be used; the *timing* of benefits in terms of duration or later impact; and finally the whether the *range* of child participants is adequate in terms of potential benefits [26].

One simple way for investigators to consider benefit, is to imagine that the study is finished, the trial question has been answered, then adopt the perspective of a public health official, a general practitioner, a parent or a child and ask "so-what?" What material benefit to children would accrue from the result? Does it merely confirm something already known? Is it interesting but not essential for child health and care? Is there any reasonable likelihood of direct benefit to participants or benefit to children in general? What would the "reasonable person" think about the study if they also knew the costs, risks and demands made on health care staff, participants and families to get the answer? Would the reasonable person agree that some children need to be exposed to risk to benefit children at large? Would the lay-person feel as Smyth [41] does, that "we are all aware of the dramatic impact which results of clinical trials have had on the care and survival of children" (p. 835), and "there is an energy and dynamism which is both exciting and invigorating for paediatric clinical research" (p. 837).

While common sense, careful scholarship and the "so-what?" test suffice for many trials, some studies and teams may benefit from involving ethicists in early stages of question development to ensure that a careful and considered approach to justice and benefit is taken. This is particularly important when trial interventions or measures may involve discomfort, distress or pain; or where the illness or condition brings inherent suffering or risk of death that may be exacerbated or alleviated by trial processes. A clear and scholarly beginning position on the issue of benefit to children not only helps investigators ensure their trial question is child-centred, but it also helps sponsors and participating institutions monitor progress of the trial and develop public and stakeholder communication plans.

The third challenge for investigators is to ask a question that can actually be answered. Is it *achievable*? Is it a question that the investigation team has the capacity to answer from the point of view of expertise, population access, resources and infrastructure? A question may not be realistic if there are insufficient potential participants for the study time frame; if the testing regimen is not physically or emotionally tolerable, or fails to accommodate family routines or responses; if sufficient funding, expertise, infrastructure or consumables are not guaranteed; if a-priori protocols for intervention, child care and data analysis are not transparent or cannot be adhered to; or if the findings are not going to be available or disseminated in ways that will grow the knowledge base to benefit children in general.

One of the most important issues to consider in weighing up whether a trial question is achievable is the capacity of the team. Clinical proficiency and research interest is not enough—the team as a whole needs to have the expertise to conduct all aspects of the trial. This may mean investigators must expand team membership to fill skill, knowledge and labour gaps that range from statistical analysis, ethical question design, budget management, regulation compliance, use of clinical procedures, to interpretation of findings and writing. Clinical professionals new to research sometimes learn the hard way that a good research idea is not necessarily question that can be investigated. Alternatively these researchers don't learn, they don't use structural and capacity building strategies to get the study done, and instead they can blame anyone involved in or connected to the project! The experience can leave clinical staff, managers, families, patients and researchers themselves disillusioned about the research experience and sometimes about each other. This is a particular risk for professionals who are attempting trial research in environments where they are already overloaded with clinical responsibilities and have limited prior experience of the technical and time demands of trials.

Turning a good idea into an achievable question requires a combination of scholarly, managerial and political skill that almost always involves long term collaboration of multidisciplinary experts. It almost never involves the "deity-like-researcher" leading followers or directing an operational team from a geographical or organizational distance. Such studies have inherent structural risks that make them prone to mistakes. Successful paediatric trials are always a team effort. They always involve the building of relationships over time with trust, respect and recognition. They always take effort on the part of the research leader to build and maintain a climate of open scholarly enquiry. Without such effort, mistakes may not be reported, good staff may leave, people fear betrayal or theft of their ideas or reputations, team politics rather than child participation can absorb emotional energy and the high test of ethical practice required in the moral context of paediatric practice is undermined. For paediatric researcher leaders there is, quite simply, no way to "delegate" a paediatric trial. Senior investigators must be involved in every aspect of trial planning, conduct, interpretation and writing as directors, collaborators or hands-on players. To do anything else is to risk allegations of being "front-men" or "poster girls" for trial sponsors, research institutes, or servants of their own "brilliant" careers—even when this is not true.

In addition to issues of team capacity, achievable questions rely on practical details such as estimating whether or not required sample sizes are attainable given the incidence of the condition, the recruitment target population, recruitment methods and time available. Many paediatric clinical populations are very small, hard to access, and may have high decline rates in recruitment. Trial planners may need more than epidemiological data to estimate whether or not they have a reasonable likelihood of recruiting the needed sample—they may need local "on-the-ground' informants who can estimate the impact of recruitment methods on limited potential participants. The target sample size must not only account for the clinical effect size of the intervention in question but also appropriately deal with the confounding variable of developmental maturation, and practical issues such as likely decline rates and drop outs. Trial questions that require good luck to achieve sample sizes should be put to one side. This is hard for researchers to do, particularly if the question is their passion, however it must be done. Underpowered trial findings can

be worse than no findings at all—they give the illusion of reliable evidence. They are all too common in paediatric research. A review of trials published in the *Archives of Disease in Childhood* from 1982 to 1996, found that half the trials had 40 participants or less, which in the case of the trials reviewed meant that they were often under-powered [42]. Researchers should ask themselves "what is the point?" if an adequate sample size cannot be assured. Their attention should turn to more realistic questions or less rigorous research designs.

The fourth challenge for investigators is to ensure that the trial question is *new*. Questions should not be asked when answers are already known. The case of originality should be made clear in "study rationale" sections of institutional review board and investigation plan applications. The case for originality must be strong, scholarly and set in an international context. The strength of evidence should be compelling: systematic reviews, when rigorously conducted, provide a strong evidence base to demonstrate gaps or failings in current knowledge that can be used to justify originality.

Fifth, investigators need to be confident about the *timing* of trials that involve children—they should only be done when the knowledge base is "ready". The nature of the trial should have a good fit with knowledge already available. ICH guidelines suggest [29, 30]: Phase 1 or 2 paediatric trials are acceptable only when diseases being targeted are entirely or predominantly found in children; phase 2 or 3 trials are acceptable in children for serious diseases where no adequate treatment exists but only after safety and tolerability information has been gained from adult studies; and, Phase 2 or 3 paediatric trials are acceptable for conditions in the general population after there has been considerable research work in adults [29, 30].

Finally, trial questions need to balance anticipated harm with expected benefit and harm should be of "*minimal risk*". Minimal risk can denote the type of procedure for data collection or intervention where only "very slight or temporary negative impact" might occur [7, p. 15], or where the risk is about the same as that in daily life, or with comparable treatments. Under regulation, [43] potential risks to research participants must be identified and minimized and the prospect of direct benefit to research participants must be maximised. Risks in this instance refer to: "any harm including physical injury, pain, distress, psychological harm, social economic or legal harms that might occur of physical injury occurs, or the potential harms that may be caused if research related information is shared with others" [43]. This places heavy obligations on investigators to identify and describe what potential risks might be. The Royal College Paediatrics and Child Health [26] provides a guide to likely harm description by identifying five aspects to consider: the *magnitude* of severity; the *probability* of harms occurring; whether the *type* of intervention is invasive or non-invasive including psychosocial procedures; the *timing* of potential harm in terms of immediate duration or later effects; and finally issues of *equity* relating to the overuse of children who have many medical problems and are used in research because they are more easily accessible. In assessing harm control, intervention and placebo conditions need to be considered—even placebos may cause harm—antihypertensive drug studies is a good case in point as this raises ethical issues regarding consent and trial design for a condition known to cause harm if left untreated [44].

While there is a requirement to assess likely risk, there is also an obligation to minimize whatever harm must be done to answer the trial question. There needs to

be an appropriate balance between the harm done and the benefit achieved: "expected benefit must exceed recognizable risks [and] serious predictable risks should be avoided" (38, p. 43). These are subjective judgements that must be made by investigators, institutional review boards and families of participants. Some judgements are easy—effective treatment should not be withheld from child participants. Other decisions about design and procedures are harder. The National Academy of the Sciences [45] provides a guide that may be helpful: researchers should consider the potential for age-related risks of harm; whether or not children are really needed; screening for known vulnerabilities; how demanding the protocol adherence is and risks arising from this; the use of only necessary procedures to answer the question; use of rigorous research designs; use of existing knowledge to estimate likely type and magnitude of risks; inclusion of adverse event information in data collection and reports of findings; ensuring investigator research and paediatric capacity; assuring appropriateness of the research setting for children; inclusion of safety monitoring, emergency arrangements, stopping rules for discontinuation; having clear guidelines for data use; and a plan for secure archiving [45].

To estimate and minimize likely harm, investigators must describe the level of risk of an intervention or outcome measure procedure. Defining level of risk is, however, a fraught task. It is one that involves value judgements because currently, risk assessment lacks an empirical standard even though risks are supposed to be identified, quantified and compared [39]. Some paediatric leaders have identified the need for guidelines [46, 47], although the use of guidelines is not universally supported [48]. Inadequate though they are, investigators should thus consult whatever regulation guidelines underpin their institutional review board requirements and ensure they benchmark their risk rating to the relevant regulation using scientific evidence for support. If there is limited scientific evidence available, the National Academy of Sciences [45] guidelines provide a framework for researchers to describe, as best they can, their strategies to minimize harm.

Researchers should also be aware that risk perception varies from the lay-person to the expert. As Afshar et al. [39] suggest: "experts usually assess harm in terms of mortality or morbidity, while a lay person may perceive harm in terms of severity, reversibility, the effect on future generations or influence on personal life. The main consideration should be the acceptability to non-experts, which in pediatric research are parents and older children. The most direct way of determining risk acceptability is to inform participants of the probability and magnitude of harm, and ask them about their preference" (p. 837). Whether or not the risk, discomfort and suffering caused is ultimately reasonable will depend on the question, interests of the child, preferences of the parents, likely benefit and comparison to likely harm in usual clinical practice.

## 9 PARTICIPANT CHARACTERISTICS

Infants, children and youth have unique body system and social attributes that need to be considered in trial research. Probably one of the most quoted phrases in paediatric research that emphasizes their unique position comes from the Royal College of Paediatrics and Child Health [26]: children are "not small adults". While every trial will need to investigate participant characteristics relevant to the sub-speciality,

there are some common issues relating to developmental change that should be considered in all fields. These include ages, stages, incidence and heterogeneity of paediatric disease and vulnerability. These issues can be managed as potential confounders as they are known and some aspects to consider are reviewed below. Paediatric trial investigators must also be alert to the possibility of confounders that may be unknown but somehow inherent in the fact their participants are youngsters.

Investigators need to consider how the variable of *age* will be managed in the trial. Age is a proxi-indicator of developmental change and age, as a covariate, needs to be measured and controlled for. The ICH [30] recommends age should be defined in completed days, months or years, using the following stage categories: preterm newborn infants; term new born infants (0 to 27 days); infants and toddlers (28 days to 23 months); children (2 to 11 years); and adolescents (12 to 16 or 18 years depending upon the region). These categories may be used as participant inclusion/exclusion criteria in a study to restrict age variability; alternatively study specific age limits can be set that reflect the sub-speciality study question. In some studies, age may be less of a concern, however it must always be prospectively accounted for, usually by being treated as a continuous covariant in analysis particularly if participants "move" from one age category to another in the course of the study. Different end-points can be set for different age ranges, however every age strata and end-point increases the numbers required in samples.

ICH [30] age related *stages* also provide investigators with specific guidance on factors to consider in relation to drug trials. These factors may also be useful in other types of clinical studies. Investigators are, for example, alerted to [30]: the need for assessment of causes of low-birth weight in preterm newborn infants to determine whether they are immature or growth retarded; or the increased reliability of oral absorption in infants and toddlers; or the increased drug clearance (hepatic and renal) for most pathways in children; or the possibility of hormonal change affecting results of clinical studies in adolescents [30]. Kurz and Gill [38] point out that there are many differences in "physiology, pathology, pharmacokinetics and pharmacodynamics between children and adults" (p. 42); and further, growth and development can influence side effects, dose relative to body weight or surface area, severity of disease, pathological agents and natural history [38]. Investigators should consult the ICH [30], speciality trial resources such as Helms and Stonier [8], and make specific enquiries regarding age related factors in their sub-speciality. This is particularly important when [37]: considering invasive procedures such as the use of repeat blood samples in infants and children as blood volume may limit what can be done; testing off-label or unlicensed applications; or when using surrogate markers, measures of quality of life, pain or other outcomes not validated for use in children [37].

The *incidence and heterogeneity* of disease in children has also been identified as a practical challenge in paediatric research [37]. Some diseases in children are rare which creates challenges for recruiting adequate trial sample sizes. For others, individual responses can vary with age.

The *vulnerability* of young participants is a critical aspect to consider in trial research. As the *European Union Clinical Trials Directive 2001/20/EC, Article 3* states, "children represent a vulnerable population with developmental, physiologi-

cal and psychological differences from adults". In the distant and not-so-distant past vulnerable children were exploited for medical research purposes particularly if they were institutionalised, disadvantaged or had disabilities [39]. Today, conventions, regulations and guidelines identify these actions were unacceptable, however a child's inherent vulnerability means the risk of exploitation is always present. It was their vulnerability and perceived inability to give informed consent that lead to the post-World War Two tradition of excluding children from medical research following the establishment of the Nuremburg Code Directives for Human Experimentation 1949. Since post-war years, there is growing recognition of the need for high quality paediatric research "so that tomorrow's children receive new and better treatments and clinicians have real evidence on which to base their decisions" [41, p. 837]. So children are now permitted to be involved in research, but with a particularly cautious and sensitive approach to acknowledge their vulnerability—this caution and the complexity of regulatory and administrative arrangements to protect them are perceived by some as "barriers" particularly in relation to drug development, but it applies equally to other clinical interventions [49]. "Barriers" are, however, better than the alternative if they are efficiently designed and managed. Barrett [37] says children are the most vulnerable patient group in research—they have fewer rights, problems expressing themselves and the potential for lasting benefit or harm from a research experience. Notwithstanding their vulnerability, she argues that paediatric trial research is essential: "because of the failure to recognise their needs and to perform appropriate research, children are denied access to safe and effective treatment that adults would demand as a fundamental right. Children have become therapeutic orphans" [37]. These "orphans" have been identified since the mid 1960's when use of adult drugs for children through off-label or off-licence prescription was recognised as a common practice [50] and the term has since been used to identify these medicines [5, 51, 52]. In most cases physicians have little "wriggle room" as adult drugs may be potentially or demonstrably effective for children and the alternative is to provide nothing. Professional societies have tried to grapple with this problem as off-label use can mean that third-party reimbursement or government subsidy of these drugs is not permitted, and physicians may be exposed to law suits in the event of adverse reactions. The ultimate solution to the expanding family of "orphans" is the production of research that involves child participants in rigorous trials under the "parental" scrutiny of regulation, rules, reporting and peer review.

## 10 PAEDIATRIC INVESTIGATION PLANS

Other parts of his handbook have explored the preparation of trial protocols and plans in detail. These principles and practices apply to paediatric trials too and will not be reviewed again. This section therefore highlights the special features of paediatric investigation plans that researchers should be aware of at study commencement. These features include the need for plans to be child-centred and family focussed, publicly accountable, regulation compliant, and multi-disciplinary.

Investigation plans need to have the *needs and experience of the child* as the centre-piece of development. Only the minimum number of children required for

adequately powered study design should be used. Measures should aim to demonstrate the quality, safety and efficacy of interventions for infants, children and young people. Consequently, measures selected, data collection procedures, the timing of collection, personnel involved and the study environment should be planned in such a way that developmentally appropriate and supportive approaches are used. The ICH [29, 30] recommends the following ways to minimize discomfort and distress in paediatric participants: only use personnel knowledgeable and skilled in dealing with infants, children and youth and their age-appropriate needs including skill in performing paediatric procedures; use physical settings with furniture, play equipment, activities, and food appropriate for participant's ages; aim to have the study conducted in familiar environments that may include usual places of care; minimize pain and discomfort in procedures wherever possible for example by appropriate anaesthesia, and collect research data through routine clinical tests rather than additional procedures.

Selection of methods should be based on *research evidence* from development and sub-speciality literature, together with advice of paediatric research specialists [8]. Trial plans from adult studies should not be "adapted", but rather a fresh approach taken with the needs and attributes of children at the core of plan decisions. All personnel involved in the trial should be trained not only for the technical requirements of their role, but also for interaction with children. The trial manual or handbook should focus on the behaviour required of study personnel towards children and their families and this should be monitored for consistency throughout the study. Special training for the identification and management of paediatric adverse events should be required in study plans and a climate of openness encouraged by trial leaders so personnel feel compelled to report mistakes or gaps and are respected and valued for doing so—even if they made the mistake. Outcome measures should consider the child as a person and the demands and influences that will be made on him or her throughout the course of the study. How will the study requirements affect their daily routine, their opportunity for play or rest, their contact with friends and family, their use of favourite toys or electrical devices, the protection of their personal privacy? Things that are of marginal or no interest to researchers may be critical for the wellbeing a child or adolescent.

Apart from the notion of "minimal harm" and "benefit" there is the practical matter of what a child can *reasonably tolerate*. There is little point, for example, in setting a battery of performance tests to measure outcomes if the child can only concentrate long enough for one of them; or setting a number of blood collections if total blood volume in the child is too small; or administering medications or conducting procedures that assume adolescents are not sexually active or that they will openly tell guardians if they are. Obvious errors can be made by even experienced investigators if they focus on outcome measures and forget that they come from a child.

Trial investigation plans should also be *family focussed*. Parents and caregivers know their infants, children and youth better than anyone else. They must feel committed to and interested in the trial to first give consent, and then to maintain involvement. For any recruitment, retention, intervention or outcome measurement strategy to be successful in a paediatric clinical trial, the commitment, expertise and adherence of parents and caregivers is critical. For children in care, consent by authorities as well as carers may also be required. Investigation plans should use

recruitment and retention strategies that have been found to be effective in other trials with participants who are minors including: contact and scheduling methods (such as the use of multiple contacts for recruitment, maintaining a log of cohort follow-ups, making phone calls to contacts to locate participants), reminders (for example letters to remind of appointments), family friendly visit arrangements (such as scheduling data collection visits at the same time as routine health care visits), reimbursements (for example of transport costs), financial incentives (reimbursing for their time), non-financial incentives (such as thankyou letters), and tracking methods to monitor participation (such as regular team meetings to follow up cohort participation) [53–55]. A concerted effort to retain participants is essential. Even though "intention to treat" principles can be used in analysis, drop-outs do threaten internal and external validity of studies and limit the level of inference that can be drawn from findings.

Practically, families may need to adapt their schedule, expend resources, provide transport or other support to enable a child to participate in a trial. The question, the demands on them, the likely benefit and harm to their child will influence their decision to be involved and stay involved. If investigation plans fit into family priorities and schedules their continued involvement is more likely as it is part of a *sustainable* routine. This is particularly the case if trials require special visits for data collection. Even if a child is continually available to researchers, for example if they are hospitalised, the family must still be the focus of plan activities. Many families, for example, would expect that any trial related activity would occur in their presence and that may mean adjusting data collection or intervention schedules for study personnel. Further, families are important sources of trial information—they are usually more alert to changes in the child, are aware of more potentially confounding factors, and if their involvement as "trial partners" [17] is acknowledged by study personnel, they are more likely to adhere to protocols and provide helpful information regarding the study conduct or the child's response.

In multi-cultural or multi-lingual societies it is important to provide information that can be understood by parents and care-givers. A common practice is translation—this should be done by professionals with expertise in health using usual process of "back-translation" to verify accuracy. But there may be more to consider—the content, presentation, method of recruitment and parent communication may require accommodations to ensure understanding and cultural acceptability. Local advisers and community members should be involved in trial planning stage rather than making adjustments as the variability of the population presents itself.

PIPs need to be publicly accountable and they need to adhere to regulatory requirements. The EMEA, for example, requires all PIPs to be submitted for consideration and to be approved by the Paediatric Committee before study commencement. Once approved, the plan is binding on investigators and sponsors. Results need to be submitted at study series conclusion in accord with the plan specifications. Investigators conducting studies outside these EMEA requirements, are well advised to mimic the transparency and and provide a publicly accountable research plan as it is probably one of the best ways to manage potential criticisms or allegations of child "exploitation". This is particularly important if clinical research organisations are employed to conduct studies: as contractors they can benefit from a clear public plan as it holds them to account at the same time as protecting the investigators and the children.

Perceptions or actual cases of exploitation in medical research do happen. Niles [56], for example, quotes Dembner, an author from the "Boston Globe" February 18, 2001 who, under the gut-wrenching title of "Dangerous dosage to make pediatric medicine safer: Thousands of children are being used to test drugs originally designed for adults" states: the "potent combination of vulnerable children, ambitious researchers, potential profits, and weak oversight can hold great peril for these children". While the particular facts and opinions surrounding this article are not pertinent here, it is apparent that the title and commentary are, in general terms, both true and false when it comes to pediatric medicines. Yes there are thousands if not millions of children receiving adult drugs—and yes some are involved in testing these drugs—but what is the alternative? To let children go untreated and suffer? To wait until specific paediatric drugs have been developed? Not to test? To ensure only those researchers with no ambition, or those companies with no obligation to shareholders, or those universities with no interest in private funding, engage in pediatric research? To consider that the elaborate regulatory and reporting arrangements in many places are weak and can be bundled together with those that are genuinely sloppy? No wonder we are all so cautious and careful in our PIPs and no wonder there are so many "barriers" to pediatric trial research when the newspaper by-lines such as this can come hurtling into view. The best we can do as pediatric investigators is ensure that our research does meet high standards and that there is opportunity for transparency and accountability in what we do and how. PIPs that comply with reporting requirements and are clearly child-centered can protect investigators and children, and encourage the public to engage in a more informed dialogue. If investigators do not need to register with the EMEA other strategies such as submission to "trial banks" or publication of the PIP/protocol with the trial findings, or publication in open-access or independent journals can be a way to achieve transparency. Such strategies may address some of the concerns regarding potential and actual conflict of interest that have been expressed by researchers themselves [57]. There is an urgent need for informed, realistic and open public, researcher and sponsor dialogue around complex issues of trial funding, independence and 'cross subsidizing' of research, researchers, clinical or administrative staff through research funding and PIP transparency could help.

Investigation plans must also be tailored to *multidisciplinary* study teams. In particular, communication strategies for multidisciplinary personnel within and outside the trial are needed. All trial personnel should have a trial manual or handbook that specifies what they need to do, that uses "quick reference" and visual cues to help guide and reinforce required behaviour. Photographs, flow-charts, check-lists and so on may be useful. Small reminder posters, cue notes on files, regular follow up, thankyou or reminder calls, occasional presentations by researchers can help keep a multi-disciplinary team consistent and equivalent in their trial behaviour. These strategies can also help people feel valued and retain their commitment and enthusiasm. For clinical staff not connected with any trial activity, there may also be a need to provide information. Many parents seek the opinion of their doctor about whether or not their child should get involved in a trial [56]. Consequently, researchers should prepare information and communication strategies. Researchers could for example, contact local doctors, teachers, case-workers, therapists etc as a matter of routine. They could provide parents with information sheets that can

be given to those people parents believe are important in helping them make decisions.

## 11 ASSENT AND CONSENT

This section of the chapter explores issues of consent by parents and adults and of assent by participating children. Not enough paediatric research is conducted as there are so many apparent barriers [49] involved—trying to avoid harm, fear of litigation by sponsors and investigators, and concerns about ethics on the part of potential investigators. While great care must be taken by investigators to protect children from unnecessary risk, children ultimately have the most to gain from well intentioned and well constructed paediatric trials. Requirements relating to consent and assent can be arduous and can vary from place to place [e.g., 6, 7, 15]. Investigators must be aware of regulations that apply in jurisdictions in which their research is being conducted, as consent has legal meanings and implications, particularly in relation to the "competence" of a person, including a child, to give consent.

One way of conducting responsible research is to ensure that those who act on the behalf of children are well informed [40, 58]. If adults who make decisions for children and act towards them are well informed about research and ethical issues, paediatric needs and the autonomy and potential apprehension of children about medical procedures, then agreeing to participate in paediatric trials is done properly. The Royal College of Paediatrics and Child health [26] provides guidelines on consent that cover strategies to ensure consent is freely given, informed and that explanation, information and where possible study findings are available. But even with these safeguards not all parents read consent documents and not all have these documents explained properly to them [59]. Parents and caregivers are vulnerable to coercion, influence and intimidation even when they are well intentioned if they are not properly informed [31]. Consequently, researchers need to take the time and consider how best to explain to parents and children what will happen and why, why it is important and why children in general will benefit [60, 61]. When parents have the study specific and context relevant information they can reflect on their child's involvement with all relevant issues laid bare. They need to be able to understand their right to withdraw and the difference between trial interventions and usual therapy [59]. If an open approach is taken, "it is now widely accepted from an ethical perspective research can be carried out on children, when there is no expected benefit for them individually, provided there is minimum risk, strict safeguards and no objection from either the child or parents" [62, p. 202]. The Medical Research Council [7] provides a helpful flow chart of processes involved in seeking consent that may be a useful procedural guide.

Explicit informed consent to participate in a clinical trial is a standard ethical requirement. In the case of paediatric trials, parental consent or consent from a person with parental responsibility on behalf of the child is necessary [63]. As noted in the last paragraph, for parental consent to be valid it must be freely given and informed. The parent can permit, approve, or agree to anything that is clearly not against the interests of the child [63]. Parental consent to something that might harm the child is not valid. The Belmont Report, cornerstone of all international ethics guidelines [64], requires protection of children regardless of parental permission.

"Minimal risk" is acceptable for parent consent and is generally accepted to be equivalent to risk encountered in the normal course of a child's everyday life. Parents have the responsibility and authority to choose activities that define their child's risks and benefits [65]. thus, parents can consent on their child's behalf when the risks are equivalent to the normal risks of childhood [65]. They can also consent when the risks are comparable to other available options—referred to as "clinical equipoise" [66]. Demonstrating clinical equipoise is particularly difficult in paediatric studies—Afshar et al. [39] recommend that "genuine uncertainty about the superiority of 1 treatment over another is essential to motivate clinicians and patients to participate" (p. 838). Parents cannot and should not consent to risks that are not comparable to other available options or that go beyond minimal risk.

The process of obtaining informed parental consent is complex. Consent to treatment is usually in the child's best interest but consent to research may not be [37]. Research benefits may be for the paediatric population at large and may not necessarily be for the individual. Or there may be little benefit at all. Why, then, do parents consent to research? The decision to participate in a trial is known to be influenced by parental, child, trial and investigator factors [67, 68, 69]. The two main reasons parents say they consent to trial participation are—contributing to clinical research and to benefit their child [70]. Caldwell, Butow and Craig [67] found parents may perceive benefits to include the offer of hope, better care, access to new treatments, access to help and information, parent-to-parent support, and the altruistic motivation of helping others.

Even though parents may support the general notion of paediatric research trials they may not want their own child involved [39]—parents fear causing harm or hurt. Parents may object to the notion of their child being used as a "guinea pig", especially where the trial methodology involves random assignment and placebo controls [67]. Preparation of participant information sheets can help counter fears and facilitate successful recruitment. Investigator factors such as doctor recommendations, doctor invitations and communication of trial information also affect decision-making from the parent's perspective [67]. Parents also seek out the views of their own doctors to help inform trial participation decisions [56, 68]. Parent knowledge, beliefs and emotional responses coupled with an understanding of their child's preferences will affect decision-making [67, 68]. Other parent factors that can influence consent rates are the parent's socio-economic background with people from lower strata agreeing more often [71, 72], parent understanding of right to withdraw, trial versus usual therapy, and the voluntary nature of participation [59]. The context of study invitations and recruitment conditions also affect consent rates and the validity of consent given. Parents in emergency or acute medical situations may not be in a state any reasonable person would consider was acceptable for making consent decisions on matters relating to harm and short or long term potential effects. Time for reflection may affect agreement rates—the longer parents are given to reflect on risks, the less likely consent will eventuate [37]. Parents of chronically ill children or children with disabilities may also be "research savvy" and able to weigh up issues in the light of previous research experience, while others may not be able to differentiate research activity from usual treatment even when this is pointed out. There is a debate in medical practice emerging about how consent for interventions should best be given [73] and this will have implications for research consent procedures.

In paediatric research, consent of parents is not enough, Even if it is given in an informed and fair manner, the interests of the child remain paramount [1]. Since children may have only a limited understanding of what the research involves, it is hard to think of their participation as anything other than "involuntary" [61]. So how do investigators observe the same ethical principles of respect, justice and beneficence as they do in adult research? Assent provides the answer. Where positive agreement can be obtained from children capable of giving it, studies should be discussed with them in age appropriate ways such as stories or photographs, and assent sought [31]. Determining whether or not a child can decline or assent in an informed manner involves a judgement about their capacity to reason [74]. Researchers must be cognisant of the child's developmental level both when explaining research procedures and the likely outcomes, and when judging the child's capacity to give consent [19, 20, 25]. It is understood that "children become capable of assent when they are capable of understanding the research in question and making a prospective decision whether to participate" [77, p. 233]. If a child refuses to participate and gives a reason for this that makes sense given their age, then this may be enough to indicate "competence" and hence their ability to decline in an informed manner. Welthorn and Campbell [75] suggest the age of nine years might be a useful guide and this is the intellectual age recommended by the American Academy of Pediatrics [76]; Wendler recommends the age of 14 years as suitable because at this age they can usually understand research questions [77]; and Koren et al. [78] suggest more general indicators of "maturity"—such as when they are old enough to be a "baby-sitter"—are adequate. Regardless of competence to assent, children must be given the opportunity to object. Sustained dissent should be respected in all cases even if the child is too young or unable to give assent [79].

The process for acquiring assent involves the researcher thinking about the issue from the child's point of view. Assent statements should be written in language that is at the comprehension level of a 6-year old child, in large font using the active form [39]. It is similar to the process of converting medical jargon into plain English for adult participants on information sheets. Study participation benefits and risks are different when viewed from the child's perspective. A child wants to know [80]: Will it be fun? Will it hurt? What will happen to me? Do I have to? Will there be parental consequences if I say no? Will assent lead to other desirable incentives, such as time off school, undivided attention of a parent, new toys and treats as rewards for good behaviour, or perhaps even the possibility of making a worried parent happy?

Children should be asked to assent only to research procedures that they are capable of understanding [79]. A child cannot become fully informed via a standard written participant information sheet and so will rely on environmental cues to confirm their impressions as part of the assent process—Is my parent looking comfortable or anxious around the investigator? Does the setting give any clues that it will be painful despite what I am being told? Does the investigator seem trustworthy to interact with? Do they provide toys and play with me? Stop when I say no? Make me fail repeatedly? An investigator who attends to the child's emotional and environmental context in addition to providing factual information is more likely to gain the child's agreement. The investigator should keep a written record of the assent procedure which demonstrates that they "provided the child with all the necessary information in an age-appropriate fashion, that the child understood the informa-

tion, and that the child voluntarily agreed to participate in the research project" [80, p. S32]. The investigator records are probably more important than obtaining a child's signature, because it is not until children are older that they understand the symbolic meaning of a signature [80].

The process of gaining assent from children and parents together gives rise to many grey areas for ethics. Children and parents do not always agree. Balancing the wishes of a child who refuses to assent to research procedures against the parent's decision and right to enrol their child in a potentially beneficial trial is a complex ethical dilemma. Ethical issues regarding consent and confidentiality become even more complex when adolescents are participants, as they may not want their parents' involvement, but in studies where harm is greater they may need adult assistance [32].

Other issues involved in consent relate to processes and delegations. Children may have parents or care-givers, but they may not be the people who hold legal responsibility for the child. Families may be separated and careful attention to agreement of relevant adults is needed. Agencies, government or other adults may hold the authority to consent. Once in a study, parents and the children always have a right to withdraw and this should be told at the beginning and throughout the study.

Adequate recruitment and retention of homogenous participants, to paediatric research trials, is extremely challenging. Many diseases common in adults are rare in children, so it is hard to recruit sufficient numbers for statistical significance. There are temptations to "water down" high standards and consent procedure requirements in the interests of study completion. Pressure on investigators to attain sample sizes, increases the risk of parents being inappropriately pressured to enrol children to participate in trials, because the available numbers are so small. Coercion at any level is unethical even though good numbers are needed for rigorous research. Investigators must carefully plan and design trials that are clinically feasible in terms of likely sample size. It is unprincipled to enrol child in studies that cannot ever be completed due to sampling frame limitations. Recruitment and referral strategies can influence consent rates: referral by health care professionals leads to high consent rates [55], sample size attainment is more likely if large numbers of potential candidates are approached [55]. Recruitment incentives also positively influence consent rates, but their use should be informed by appropriate ethical standards [3]. Payment and reimbursement of costs have been used to provide compensation for study participation and retention to study completion [39, 81].

## 12 SAFETY AND MONITORING

Safety in clinical trials is both a process and an outcome. Safety information is an aim of trial studies. At the same time, the duty of care for trial participants means that trial processes must plan for prompt action in response to suspected and identified adverse events. Trial plans and reports therefore need to clearly outline what strategies, decision-points and decision-makers will be involved in identifying, referring, confirming and responding to suspected adverse events.

Safety of trial participants needs to be considered before, during and after the trial. Before the trial, a careful exploration of the existing evidence base should be

made to assess potential adverse events and these should be monitored and measured. If the information base relates only to adults, then a careful watch on paediatric participants may be needed as developing systems and structures may respond differently to adults [29, 30]. Before the trial, processes to monitor protocol adherence, inspect participant files and receive reports from trial and other clinical staff about suspected adverse events need to be in place. Training should be provided to everyone involved in the trial so that they know how and when to make reports of suspected adverse events. A culture of openness and respect is needed so that mistakes or adverse events are reported promptly and the honesty and vigilance involved in making that report is valued by the team. After a trial, there may be a need to continue safety and monitoring processes through long-term follow up studies. This is particularly important for paediatric trials or trials where the condition is chronic and treatment long term, as participants are growing and adverse consequences of an intervention may not be apparent initially but may emerge later in life [30]. Sammons and co-investigators [82], identified through a literature review of therapeutic clinical trials, that only 2% of studies reported use of safety monitoring committees, and only 11% of studies reported adverse events including deaths. They came to the conclusion that every paediatric trial should have an independent safety monitoring committee to assess the likelihood of risk, monitor the progress of participants and quickly respond to observed differences [82]. Independent risk assessment and safety monitoring is helpful not only for study integrity, but also for protecting children, increasing parent confidence in the well-being of their children, and demonstrating due diligence by investigators. Safety monitoring should be done by personnel with paediatric expertise, access to expert advice regarding the subspeciality, ready access to trial material, and the authority to compel action if the safety of trial participants is in doubt.

## 13 CONCLUSION

This chapter provided an introduction to common issues that are unique or pronounced in paediatric clinical trials and an outline of those issues are critical to their initiation, conduct and success. The tone adopted in the chapter was that of the "guide on the side". These trials, perhaps more than others, require sponsors, investigators, trial coordinators and project managers to adopt a personal moral standpoint towards participants and appreciate the social context of clinical research. This chapter provided common signposts that should be helpful in whatever unique trial journey investigators, sponsors and workers may lead, recognising that the technical and procedural issues covered in other Handbook chapters will apply, and the unique evidence of sub-specialities will need to be sought.

While morality, ethics and participant characteristics feature in every trial, in paediatrics they are critical to design integrity and the public success of a trial. A perfectly crafted paediatric trial can bring down an investigator's lifetime reputation, a company's global brand, or reduce the esteem of institutions if the study was not child-centred and astute to the multiplicity of factors associated with participant vulnerability and their capacity to reasonably and meaningfully participate in trial requirements. The complexity and demands of paediatric clinical trials can drive some investigators, sponsors and institutions to wonder whether or not they are

worth doing. There are many perceived and actual barriers [49]. Paediatric trials are not for the faint hearted. They demand the highest intellectual capacity and ethical stance for planning, implementing, monitoring and reporting. These trials require enormous good will and generosity from sponsors, investigators, trial employees, parents/carers, and the children themselves. They are inevitably time intensive in ways that are hard to anticipate, they can be very costly with small, hard to locate and recruit samples. The financial and reputation rewards may be few and the risks are great for all concerned. There is, however, an urgent need for the brightest and best in health care to commit to paediatric clinical trials and do more of them.

Routine and widespread use of interventions with little or often no paediatric research evidence continues apace with potentially catastrophic results. Investigators, sponsors and institutions can sometimes feel caught between a rock and a hard place, as they are castigated for off-label, unscientific practice whilst suffering sometimes justified but often unwarranted public torment regarding involvement of children in trials. The irony of public concern for paediatric trial participants is that there appears little or no concern for the millions who daily receive interventions without any scientific evidence to inform parent or professional decisions. Notwithstanding the demands, risks and rigours of paediatric trial research, there is an urgent and compelling moral and practical need to involve outstanding investigators, sponsors and institutions in rigorous, community supported and well understood paediatric trials so that we are able to benefit children.

## APPENDIX: FACTS SUMMARY

Paediatric clinical trials are urgently needed to support use of clinical interventions with infants, children and youth who regularly receive off-label, off-licence or scientifically unsupported treatment.

Paediatric clinical trials require investigators to understand the moral context of their research as participants are vulnerable.

Conventions, regulations and policies from esteemed professional societies provide guidance on principles and practices that deal with implications of the moral context.

Paediatric trial questions must be worthwhile, relevant, new, achievable, timely, pose minimal risk and must benefit children.

Paediatric investigation plans need to be rigorous, evidence based, equal to the capacity of the team and resources available and benchmarked to appropriate standards and regulation particularly in relation to consent.

Safety of participants and external monitoring of adverse events are critical study processes that underpin the quality of study findings

## REFERENCES

1. United Nations (November, 1989), Convention on the rights of the child, General Assembly resolution 44/25 of 20 November 1989 Geneva, Switzerland; available at: http://www.unicef.org/crc/

2. American Academy of Pediatrics Council on Child and Adolescent Health (1988), Age limits of Pediatrics, *Pediatrics*, 81(5), 736; reaffirmed 2006 in *Pediatrics*, 117, 1846–1847.
3. American Academy of Pediatrics (2008), *Policy Statements* (various). Available at: http://aappolicy.aappublications.org/policy_statement/index.dtl
4. American Academy of Pediatrics Committee on Drugs (2002), Use of drugs not described in the package insert (off-label uses), *Pediatrics*, 110, 181–183.
5. American Academy of Pediatrics Committee on Drugs (1970), Policy Statement: Therapeutic orphans and the package insert, *Pediatrics*, 46, 811–813.
6. European Medicines Agency (2008), *Paediatric Investigation Plans*; available at: http://www.emea.europa.eu/htms/human/paediatrics/pips.htm
7. Medical Research Council (2004), *MRC Ethics Guide: Medical research involving children*, MRC Publications, London; available at: www.mrc.ac.uk
8. Helms, P., and Stonier, P., eds (2005), *Paediatric Clinical Research Manual*, Euromed Communications, Surrey, UK.
9. Kahn, J., Mastroianni, A. C., and Sugarman, J., eds (1998), *Beyond consent: Seeking justice in research*, Oxford University Press, New York.
10. World Medical Association (2004), Declaration of Helsinki, *Ethical principles for medical research involving human subjects, Document 17.C* amended. Geneva, Switzerland; available at: http://www.wma.net/e/policy/b3.htm
11. United Nations (2008), *Convention on Rights of Persons with Disabilities, General Assembly Resolution of 3 April, 2008*, Geneva, Switzerland; available at: http://www.un.org/disabilities/default.asp?id=259
12. United Nations (2007), *Declaration on the Rights of Indigenous Peoples, General Assembly Resolution 13 of September 2007*, Geneva, Switzerland; available at: http://www2.ohchr.org/englis/law/
13. European Medicines Agency EMEA (2009), *Paediatric Committee*; available at: http://www.emea.europa.eu/htms/human/paediatrics/pdco.htm
14. Diekema, D. S. (2006), Conducting ethical research in pediatrics: a brief historical overview and review of pediatrics regulations. *J. Pediatr.*, 149, S3–S11.
15. USA Food and Drug Administration (2002), *Best Pharmaceuticals for Children Act*, Jan 4, 2002 and 2007 (Updated provisions 2008 and 2009), Washington, DC; available at: http://www.fda.gov/cder/pediatric/#bpca2007
16. USA Food and Drug Administration (2003), *Pediatric Research Equity Act of 2003*. July 23, 2003. Washington, USA: available at: http://fda.gov/cder/pediatric/#prea
17. Rose, K. (2005), Pediatric drug development. *Appl. Clin. Trials, January* Article 140819, available at: http://appliedclinicaltrialsonline.findpharma.com/appliedclinicaltrials/author/authorDetail.jsp?id=19055
18. USA Food and Drug Administration (2006), *Guidance for Clinical Investigators, Institutional Review Boards and Sponsors, process for handling referrals to FDA Under 21 CFR 50.54 Additional Safeguards for Children in Clinical Investigations*, FDA, Rockville, MD.
19. National Health and Medical Research Council (2007), *Council Statements Including the Australian Code for the Responsible Conduct of Research*, NHMRC Publications, Canberra.
20. National Health and Medical Research Council (2007), *National Statement on Ethical Conduct in Human Research*, NHMRC Publications, Canberra.
21. Medical Research Council of Canada, Natural Sciences and Engineering Research Council of Canada and Social Sciences and Humanities Research Council of Canada

(2005), *Tri-Council Policy Statement. Ethical Conduct for Research Involving Humans*, Ontario Public Works and Government, Ottowa; available at http://pre.ethics.gc.ca/english/policystatement/policystatement.cfm

22. Indian Council of Medical Research (2006), *Ethical Guidelines for Biomedical Research on Human Participants*, ICMR, New Delhi; available at: http://www.icmr.nic.in/ethical_guidelines.pdf
23. Human Sciences Research Council (2009), *Code of Research Ethics*. HSRC, Pretoria, South Africa. Available at: http://www.hsrc.ac.za/Corporate_Information-6.phtml
24. National Health and Medical Research Council (2003), *Guidelines for ethical conduct in Aboriginal and Torres Strait Islander Health Research*, NHMRC Publications, Canberra; available at: http://www.nhmrc.gov.au/users/indig.htm
25. National Health and Medical Research Council (2007), *National Statement on Ethical Conduct in Research Involving Humans: People in Other Countries*, NHMRC Publications, Canberra; available at: http://www.nhmrc.gov.au/publications/2007_humans/section4.8.htm
26. Royal College of Paediatrics and Child Health: Ethics Advisory Committee (2002), Guidelines for the ethical conduct of medical research involving children, Reprinted in *Arch. Dis. Ch.*, 82(2), 177–182.
27. European Academy of Paediatrics (2009), *What is the E.A.P.?* available at: http://www.eapaediatrics.eu/v3/lay_eap.cfm
28. Kurz, R. (2004), *Paediatric research demands child-specific guidelines for ethics and good clinical practice*. European Academy of Paediatrics: Powerpoint document. Available at: http://www.cesp-eap.org/_public/lay_docs.cfm
29. International Conference on Harmonisation (1996), *Guideline for Good Clinical Practice E6(R1)*; available at: http://www.emea.europa.eu/htms/human/ich/ichefficacy
30. International Conference on Harmonisation (2000), *Clinical Investigation of Medicinal Products in the Paediatric Population E11*; available at: http://emea.europa.eu/htms/human/ich/ichefficacy.htm
31. Council for International Organizations of Medical Sciences (2002), *International Ethical Guidelines for Biomedical Research Involving Human Subjects*; available at: http://www.cioms.ch/frame_guidelines_nov_2002.htm
32. Santelli, J. S., Rosenfeld, W. D., DuRant, H. R., et al. (1995), Guidelines for adolescent health research: a position paper of the Society for Adolescent Medicine, *J. Adol. Health*, 17, 270–322.
33. National Institute for Clinical Excellence (2005), *Improving outcomes with children and young people with cancer*. August 2005; Available at: http://www.nice.org.uk/guidance/index.jsp?action=byID&o=10899
34. National Institute for Clinical Excellence (2007), *Feverish illness in children—assessment and initial management in children younger than 5 years*, May 2007; Available at: http://www.nice.org.uk/guidance/index.jsp?action=byID&o=11010
35. Association of the British Pharmaceutical Industry (2005), *Current Issues in Paediatric Clinical Trials*, ABPI Publications, London, UK.
36. Shah, S., Whittle, A., Wilfond, B., et al. (2004), How do institutional review boards apply the federal risk and benefit standards for pediatric research? *JAMA*, 291, 476–482.
37. Barrett, J. (2002), Why aren't more pediatric trials performed? *Applied Clinical Trials*, July, Article 83729. Available at: http://appliedclinicaltrialsonline.findpharma.com/appliedclinicaltrials/author/authorDetail.jsp?id=5016
38. Kurz, R., and Gill, D. (2003), Practical and ethical issues in pediatric clinical trials. *Applied Clinical Trials, September*, Article 79923; available at: http://appliedclinicaltrialsonline.findpharma.com/appliedclinicaltrials/author/authorDetail.jsp?id=5124

39. Afshar, K., Lodha, A., Costei, A., et al. (2005), Recruitment in pediatric clinical trials: an ethical perspective, *J. Urol.*, 174(3), 835–840.
40. US Food and Drug Administration (2009), *Should your child be in a clinical trial?* available at: http://www.fda.gov/consumer/updates/pediatrictrial101507.html
41. Smyth, R. L. (2007), Making a difference: the clinical research programme for children, *Arch. Dis. Child.*, 92, 835–837.
42. Campbell, H., Surry, S., and Royle, E. (1998), A review of randomised controlled trials published in Archives of Disease in Childhood from 1982–1996, *Arch. Dis. Child*, 79, 192–197.
43. US Department of Health and Human Services (2001), *Protection of human subjects: additional protections for children involved as subjects in research*. Code of Federal regulations Title 45, Part 46 Subpart D as revised October 1, 2001; available at: http://www.hhs.gov/ohrp/children
44. Flynn, J. T. (2003), Ethics of placebo use in pediatric clinical trials: the case of antihypertensive drug studies, *Hypertension*, 42, 865–869.
45. National Academy of Sciences (2004), *Ethical Conduct of Clinical Research Involving Children*, NAS, Washington DC.
46. Nelson, R. M. (2007), Minimal risk, yet again. *J. Pediatr.*, 150, 570–572.
47. Nelson, R. M., and Ross, L. F. (2005), In defense of a single standard of research risk for all children, *J. Pediatr.*, 147, 565–566.
48. Wendler, D., and Glantz, L. (2007), A standard for assessing the risks of paediatric research: pro and con, *J. Pediatr.*, 150, 579–582.
49. Vanchieri, C., Butler, A. S., Khutsen, A. (Rapporteurs) (2008), *Addressing the barriers to pediatric drug development workshop summary*, The National Academies Press, Washington, D.C.
50. Shirkey, H. (1968), Therapeutic orphans, *J. Pediatr.*, 72, 119–120.
51. Shirkey, H. (1999), Editorial comment: therapeutic orphans, *J. Pediatr.*, 104, 583–584.
52. Wilson, J. T. (1999), An update on the therapeutic orphan, *J. Pediatr.*, 104, 585–590.
53. Meyers, K., Webb, A., Frantz, J., et al. (2003), What does it take to retain substance-abusing adolescents in research protocols? Delineation of effort required, strategies undertaken, costs incurred, and 6-month post-treatment differences by retention difficulty, *Drug Alch. Depend.*, 69, 73–85.
54. Robinson, J. L., Fuerch, J. H., Winiewicz, D. D., et al. (2007a), Cost effectiveness of recruitment methods in an obesity prevention trial for young children. *Prev. Med.*, 44, 499–503.
55. Robinson, K. A., Dennison, C. R., Wayman, D. M., et al. (2007b), Systematic review identifies number of strategies important for retaining study participants. *J. Clin. Epidem.*, 60, 757–765.
56. Niles, J. P. (2003), Pediatric subjects and their parents. *Appl. Clin. Tr.*, Sept, 46–48. Article 79923; available at: http://appliedclinicaltrialsonline.findpharma.com/appliedclinicaltrials/author/authorDetail.jsp?id=5489
57. Smith, R. (2005), Medical journals are an extension of the marketing arm of pharmaceutical companies. *PLoS Medicine*, 2(5), e138. doi: 10.1371/journal.pmed.0020138.
58. Diekema, D. S., and Stapleton, F. B. (2006), Current controversies in pediatric research ethics: Proceedings introduction. *J. Pediatr.*, 149, S1–S2.
59. Hazen, R. A., Drotar, D., and Kodish, E. (2007), The role of the consent document in informed consent for pediatric leukaemia trials, *Contemp. Clin. Trials*, 28, 401–408.

60. Brown, K. E., Barton, R. P., Short, M. A., et al. (2006). Positive approach to pediatric informed consent. *Appl. Clin. Trials*, June, Article: 334576. available at: http://appliedclinicaltrialsonline.findpharma.com/appliedclinicaltrials/author/authorDetail.jsp?id=33197
61. Royal Australasian College of Physicians (2008), *The Royal Australasian College of Physicians' Paediatric Study policy on Ethics of Research in Children*, RACP, Sydney.
62. British Medical Association (2001), *Consent rights and choices in health care for children and young people*. BMJ Books: London, UK.
63. Royal College of Physicians (2007), *Guidelines on the practice of ethics committees in medical research with human participants*, Fourth Edition RCP, London, UK.
64. National Commission for the protection of human subjects of biomedical and behavioural research (1979), *The Belmont Report: Ethical principles and guidelines for the protection of human subjects of research*, Department of Health Education and Welfare, U.U. Government, Washington; available at: http://www.biomethics.gov/reports/part_commissions/index.html
65. National Commission for the Protection of Human Subjects of Biomedical and Behavioural Research (1977), *Report and recommendations—research involving children*. Department of Health Education and Welfare, US Government, Washington; available at: http://www.bioethics.gov/reports/past_commissions/index.html
66. Freedman, B. (1987), Equipoise and the ethics of clinical research. *New Eng. J. Med.*, 317, 141–145.
67. Caldwell, P. H., Butow, P. N., and Craig, J. C. (2003), Parents' attitudes to children's participation in randomized controlled trials. *J. Pediatr.*, 142, 554–559.
68. Caldwell, P. H., Murphy, S. B., Butow, P. N., et al. (2004), Clinical trials in children. *Lancet*, 364(9436), 803–811.
69. Tait, A. R., Voepel-Lewis, T., Robinson, A., et al. (2001), Priorities for disclosure of the elements of informed consent for research: a comparison between parents and investigators. *Paediatr. Anaesth.*, 12, 332–336.
70. van Stuijvenberg, M., Suur, M. H., de Vos, S., et al. (1998). Informed consent, parental awareness, and reasons for participating in a randomised controlled study, *Arch. Dis. Child.*, 79, 120–125.
71. Harth, S. C., and Thong, Y. H. (1990), Sociodemographic and motivational characteristics of parents who volunteer their children for clinical research: a controlled study. *BMJ*, 300(6736), 1372–1375.
72. Harth, S. C., Johnstone, R. R., and Thong, Y. H. (1992), The psychological profile of parents who volunteer their children for clinical research: a controlled study. *J. Med. Ethics*, 18, 86–93.
73. Elwyn, G. (2008), Patient consent-decision or assumption? *BMJ*, 336, 1259–1260.
74. Broome, M. E. (2001). *Children's assent to clinical trial participation: a unique kind of informed consent*. Available at: http://cancertrials.nci.hih.gov/understanding/indepth/protections/assent/index.html
75. Welthorn, L. A., and Campbell, S. B. (1982), The competency of children and adolescents to make informed treatment decisions. *Child Dev.*, 53, 1589–1598.
76. American Academy of Pediatrics Committee on Bioethics (1995), Informed consent, parental permission, and assent in pediatric practice. *J. Pediatr.*, 95, 314.
77. Wendler, D. S. (2006). Assent in paediatric research: theoretical and practical considerations. *J. Med. Ethics*, 32, 229–234.
78. Koren, G., Carmeli, D. B., Carmeli, Y. S., et al. (1993), Maturity of children to consent to medical research: the babysitter test. *J. Med. Ethics*, 19, 142–147.

79. Wendler, D., and Jenkins, T. (2008), Children's and their parents' views on facing research risks for the benefit of others. *Arch. Pediatr. Adolesc. Med.*, 162, 9–14.
80. Ungar, D., Joffe, D., and Kodish, E. (2006), Children are not small adults: documentation of assent for research involving children. *J. Pediatr.*, 149, S31–S33.
81. Wendler, D., Rackoff, J. E., Emanuel, E. J., et al. (2002). The ethics of paying for children's participation in research. *J. Pediatr.*, 141, 166–171.
82. Sammons, HM, Gray, C, Hudson, H., et al. (2008). Safety in paediatric clinical trials—a 7 year review. *Acta. Paediatr.*, 97(4), 474–477.

# *Clinical Trials in Dementia*

**Encarnita Raya-Ampil[1] and Jeffrey L. Cummings[2]**

[1]*Department of Neurology and Psychiatry, University of Santo Tomas, Manila, Philippines*
[2]*Departments of Neurology and Psychiatry and Biobehavioral Sciences, David Geffen School of Medicine at UCLA, Los Angeles, California*

**Contents**

1 Introduction
2 Defining AD, MCI, and VaD
3 Severity of Dementia
4 Ethical Conduct of Dementia Trials and Informed Consent
5 Generalizability of Clinical Trial Results
6 Outcome Assessments in Dementia Trials
  6.1 Primary Outcome Measures
  6.2 Secondary Outcome Measures
7 Clinical Trial Designs
  7.1 Special Clinical Trial Design Features
  7.2 Randomization
  7.3 Length of Clinical Trials
8 Statistical Analyses
9 Drug–Placebo Difference
10 Placebo Responses
11 Attrition and Adverse Effects
12 Presenting Clinical Trial Results
13 Disease-Modifying Trials
  Acknowledgment
  References

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

## 1 INTRODUCTION

Dementia is a health problem affecting millions of people worldwide. Its prevalence increases with age and it is almost always a disease of the elderly. The impact of this disorder lies not only in the loss of patient autonomy and caregiver burden that ensue as it progresses but also in its marked economic effects. With improving health care management in both affluent and developing countries, a rise in the aging population and subsequent rise in dementia cases are anticipated.

Alzheimer's disease (AD) is the leading cause of dementia. In the United States alone, AD was estimated to affect 4 million Americans in 1990 [1]. This number is expected to rise to 8.5 million by the year (2030) [2] and 14 million by the year (2050) [1]. The overall prevalence rate of AD is 2–3% at age 65 [3]. This is estimated to double every 5 years so that almost 50% of individuals 85 years and older may be affected by this disorder [3]. The United States is spending $100 billion per year to care for individuals with AD [4, 5].

The large population affected by dementia and the unmet need for more efficacious treatment have led to randomized controlled trials. The largest numbers of trials have been done with cholinesterase inhibitors (ChE-Is), the first class of agents approved by the U.S. Food and Drug Administration (FDA) for the treatment of AD. Recently, trials led to the approval of memantine, an *N*-methyl-D-aspartate (NMDA) antagonist as a treatment for patients with moderate to severe AD. These trials of approved agents are very influential in determining how future trials of antidementia agents will be conducted for AD, vascular dementia (VaD), and other entities such as mild cognitive impairment (MCI). Lessons learned from these trials will guide trial conduct not only for symptomatic agents with AD but also for compounds that may have disease-modifying effects. This chapter reviews published trials to derive guidelines on how future trials may be conducted. We concentrate on trials of AD and MCI only.

## 2 DEFINING AD, MCI, AND VaD

Precision in clinical diagnosis is essential to ensure valid outcomes in clinical trials. Randomized controlled trials in AD and MCI have utilized various diagnostic criteria to guarantee subject homogeneity.

Mild cognitive impairment had been a diagnostic dilemma from the time that the term was coined. Most regard it as a transition state between normal aging and early AD. More recently, new concepts of MCI emerged, making assessment and management more complex. Some regard MCI as incipient AD [6] (see Table 1).

Currently, MCI can be classified on the basis of the affected cognitive domain/s [7]—single memory domain, single non–memory domain, and multiple cognitive domain with or without involvement of memory. The amnestic type or the single memory domain deficit is the one that is closely correlated with AD. The rest of the MCI types may lead to either AD or other dementia syndromes, broadening the possibility of patient outcome. At the moment, therapeutic trials have limited recruitment to the amnestic type of MCI [8, 9] since this is the most certain prelude to AD. A majority of studies have adopted delay to progression to AD using a survival type of outcome as the research design approach.

**TABLE 1 Criteria for Amnestic Mild Cognitive Impairment**

1. Memory complaint, preferably corroborated by informant
2. Impaired memory function for age and education
3. Preserved general cognitive function
4. Intact activities of daily living
5. Not demented

**TABLE 2 DSM-IV Criteria for Diagnosis of Alzheimer's Disease**

A. Alzheimer's disease is characterized by progressive decline and ultimately loss of multiple cognitive functions, including both:
   1. Memory impairment—impaired ability to learn new information or to recall previously learned information.
   2. At least one of the following:
      a. Loss of word comprehension ability (aphasia)
      b. Loss of ability to perform complex tasks involving muscle coordination (apraxia)
      c. Loss of ability to recognize and use familiar objects (agnosia)
      d. Loss of ability to plan, organize, and execute normal activities
B. The problems in A represent a substantial decline from previous abilities and cause significant problems in everyday functioning.
C. The problems in A begin slowly and gradually become more severe.
D. The problems in A are not due to:
   - Other conditions that cause progressive cognitive decline, among them stroke, Parkinson's disease, Huntington's chorea, brain tumor, etc.
   - Other conditions that cause dementia, among them hypothyroidism, HIV infection, syphilis, and deficiencies in niacin, vitamin $B_{12}$, and folic acid
E. The problems in A are not caused by episodes of delirium.
F. The problems in A are not caused by another mental illness: depression, schizophrenia, etc.

*Source*: From [10].

Alzheimer's disease is a progressive degenerative disorder which leads to cognitive decline that is severe enough to cause functional deterioration. Several criteria are available to clearly define this entity: *Diagnostic and Statistical Manual of Mental Disorders*, fourth edition (DSM-IV) [10] (Table 2); *International Classification of Diseases*, tenth revision (ICD-10) [11]; and the National Institute of Neurological and Communicative Disorders and Stroke and the Alzheimer's Disease and Related Disorders Association (NINCDS-ADRDA) [12] criteria (Table 3). The level of diagnostic certainty is indicated in the NINCDS-ADRDA criteria. Neuroimaging has been included for the purpose of excluding other dementias in the differential diagnoses. Its postconsensus sensitivity and specificity of the diagnosis of AD range from 0.83 to 0.95 and 0.79 to 0.84 [13–15] while clinicopathological sensitivity and specificity are 0.64–0.86 to 0.89–0.91 [14, 16]. Interrater reliability is moderate [14]. The NINCDS-ADRDA criteria for AD has been regularly used in trials due to its established validity.

## 3 SEVERITY OF DEMENTIA

It is essential for dementia severity to be specified in clinical trials to ensure patient homogeneity and gauge treatment response. The spectrum of dementia depends on

**TABLE 3 NINCDS/ADRDA Criteria for Diagnosis of Probable Alzheimer's Disease**

I. Dementia established by clinical examination and documented by a standard test of cognitive function (e.g., Mini–Mental State Examination, Blessed Dementia Scale) and confirmed by neuropsychological tests
II. Significant deficiencies in two or more areas of cognition, for example, word comprehension and task completion ability
III. Progressive deterioration of memory and other cognitive functions
IV. No loss of consciousness
V. Onset from age 40 to 90, typically after 65
VI. No other diseases or disorders that could account for the loss of memory and cognition
VII. Diagnosis of probable AD is supported by:
    1. Progressive deterioration of specific cognitive functions: language (aphasia), motor skills (apraxia), and perception (agnosia)
    2. Impaired activities of daily living and altered patterns of behavior
    3. A family history of similar problems, particularly if confirmed by neurological testing
    4. The following laboratory results: normal cerebrospinal fluid (lumbar puncture test), normal electroencephalogram (EEG) test of brain activity, and evidence of cerebral atrophy in a series of computerized tomography (CT) scans
VIII. Other Features consistent with AD:
    1. Plateaus in the course of illness progression
    2. CT findings normal for the person's age
    3. Associated symptoms, including depression, insomnia, incontinence, delusions, hallucinations, weight loss, sex problems, and significant verbal, emotional, and physical outbursts
    4. Other neurological abnormalities, especially in advanced disease: including increased muscle tone and a shuffling gait
IX. Features that decrease the likelihood of AD:
    1. Sudden onset
    2. Such early symptoms as seizures, gait problems, and loss of vision and coordination

*Source*: From [12].

its longitudinal course, which is defined by three domains: cognition, behavior, and function. Generally, dementia severity can be divided into three stages. In the early and mild stage the initial manifestations of minor memory impairment emerge with concomitant decline in complex activities. In the moderate stage, behavioral changes become more prominent, with involvement of cognitive domains other than memory, typically language and visuospatial skills. Behavioral changes become more apparent at this point and patients have more difficulty in coping with daily activities such as housework and hobbies. In the severe stage, patients are unable to live without assistance and they manifest with more disturbing behavioral symptoms. Institutionalization is common in this stage of the illness.

The Global Deterioration Scale (GDS) [17], Clinical Dementia Rating (CDR) [18], and Mini–Mental State Examination (MMSE) [19] are instruments that are frequently used for assessing dementia severity. Both the GDS and CDR are global assessment instruments which examine cognition, function, and behavior. The MMSE, however, is limited to the measurement of cognition in terms of orientation, attention, memory, language, and figure copying. It has been regularly used as an instrument for the staging of dementia since it examines the "core" manifestations of the disorder which are the main target of therapeutic trials.

The MMSE was formulated by Folstein as a practical method of grading the cognitive state [19] of psychiatric inpatients. Advantages of this instrument are its

**TABLE 4 MMSE Range of Subjects in AD and MCI Randomized Controlled Trials**

| Diagnosis | Agent | RCTs | Duration of RCT (weeks) | MMSE/sMMSE Range |
|---|---|---|---|---|
| Alzheimer's disease | | | | |
| Mild to moderate | Metrifonate | [21] | 30 (12[a]), 36 (26[a]) | 10–26 |
| | Tacrine | [22–25] | 6, 12, 30 | 10–26 |
| | Donepezil | [26–28] | 12, 24, 52 | 10–26 |
| | Galantamine | [29] | 20 | 10–22 |
| | | [30, 31] | 12, 24 | 11–24 |
| | Rivastigmine | [32, 33] | 26, 26 | 10–26 |
| | Diclofenac/Misoprostol | [34] | 25 | 11–25 |
| | Prednisone | [35] | 52 | 13–26 |
| | Estrogen | [36] | 52 | 14–28 |
| | Rofecoxib or naproxen | [37] | 52 | 13–26 |
| | Rofecoxib | [38] | 52 | 14–26 |
| | Acetyl-L-carnitine | [39] | 52 | 13–26 |
| Moderate to severe | Donepezil | [40] | 24 | 5–17[b] |
| | Memantine | [41] | 28 | 3–14 |
| Mile cognitive impairment | Donepezil | [8, 9] | 24, Survival design (time to reach endpoint) | ≥24 |

*Note*: SMMSE: standardized Mini–Mental State Examination; RCT, randomized controlled trial.
[a]Duration of active treatment.
[b]In this trial sMMSE was used instead of MMSE.

brevity and ease of administration. It has an adequate sensitivity of 86% and a specificity of 92% when a cutoff score of 23–24/30 is used [20]. Its ceiling and floor effect, though, lessens its sensitivity in detecting mild and severe cognitive impairment, respectively. MMSE scores are influenced by age and education. The presence of within-subject and between-subject variability challenges the applicability of MMSE scores; they have wide standard errors of measurement. The natural variability of the disease also contributes to score fluctuations. Despite these caveats, the use of MMSE is widespread in the staging of dementia and is commonly used to define a restricted range of dementia severity for patients included in clinical trials (Table 4).

## 4 ETHICAL CONDUCT OF DEMENTIA TRIALS AND INFORMED CONSENT

Research trials involve experimentation and, to protect human subjects, various guidelines have been implemented. As in all research studies involving human subjects, the principles of minimization of harm, beneficence, and veracity are maintained in dementia trials. Conduct of dementia trials is complicated by inclusion of subjects with varying severity of cognitive impairment, a condition that makes them vulnerable to exploitation.

A multitude of "ethically approved" randomized controlled trials in dementia therapy have been conducted in the past years. From these studies, acetylcholinesterase inhibitors have been approved as a class of drugs that is effective and safe in AD patients. In 2001, the American Academy of Neurology issued guidelines in the management of AD and it described acetylcholinesterase inhibitors as the standard drug for this disorder [42]. Thus, assignment of subjects to placebo in future clinical

trials will be unethical since a standard drug is already available. Administration of this standard drug to all the groups in a dementia clinical trial becomes necessary.

Determination of the capacity to consent is the most discussed aspect of dementia trials. Differing levels of disease severity produce differing levels of incapacity and challenge the ability of the patient to participate in consent discussions. The presentation of the purpose, methodology and procedures, risks, benefits, and alternatives must be very clear and simple for cognitively impaired subjects to comprehend. The extent of the subject's grasp of the details of the study and decision-making capacity should be evaluated. This involves examination of (1) understanding of the relevant facts of the trial disclosed to the patient, (2) appreciation of the research risks and potential benefits, and (3) reasoning in terms of comparing options and drawing consequences from these options [43].

If a subject is deemed incompetent to give informed consent, proxy consent is obtained from an individual who has the capacity and legal authority to give it. Typically, the caregiver is given this task since he or she is the one who addresses the daily needs of the subject and knows the latter's preferences. Competence and benevolence of the caregiver must be ensured in these circumstances.

In the mild to moderate stages of the disease process, the subject will still be able to contribute to research trial decisions. The investigator must ensure that the subject is still part of the choices made by the proxy. Later, the subject loses the ability to participate meaningfully and the caregiver becomes the sole decision maker. It is at this point that assent (or dissent) should be obtained from the subject. This is judged behaviorally based on the cooperativeness of the subject with the study procedures [43]. Consistent resistance of the subject with study procedures may be taken as dissent, a probable basis for discontinuation from the study.

## 5 GENERALIZABILITY OF CLINICAL TRIAL RESULTS

A drug's value is ultimately tested once it is marketed in the targeted population. The terms "efficacy" and "effectiveness" distinguish between how well the intervention/drug *can* work under ideal circumstances such as that in clinical trials and how the intervention/drug *does* work under "field conditions" such as that in the community [44]. The drug's applicability depends on the inherent characteristics (demographic and medical) of the cohort of subjects enrolled in the clinical trial and how representative they are of the population who will eventually use it. Subject selection by using inclusion and exclusion criteria potentially limit the generalizability of the results of the clinical trial. The National Institute of Mental Health Clinical Antipsychotic Trials of Intervention Effectiveness (NIMH-CATIE) is an example of an effectiveness study proposal of psychotropic use in AD with a methodological design that assures applicability of the results in the community setting [45].

In randomized controlled trials of acetylcholinesterase inhibitors, recruitment bias was observed toward subjects who are healthier, better educated, younger, and of higher socioeconomic status [46, 47]. This has the effect of excluding subjects who have complicated medical histories and are taking specific medications. Caregivers may be more aggressive in seeking medical help when their relatives are in the early stage of dementia, exposing the patients to a higher probability of being recruited.

Another confounding factor on the generalizability of trials is the drug's eventual applicability in different ethnic populations. Randomized controlled trials require major funding and, for this reason, these are conducted in affluent countries. This fundamentally explains why Caucasians compose the greater part of the subjects in trials globally. However, there is a mixture of ethnic populations in the United States and Caucasians should not dominate disproportionately. Examination of the population recruited in acetylcholinesterase inhibitor trials which were conducted in the United States show that the samples were 91–99% white [46]. Problem may arise due to differences in the pharmacokinetics of the drug in various ethnic populations resulting in varied first-pass metabolism and systemic bioavailability. Activity of the cytochrome P450 enzymes differ among various ethnic groups, possibly leading to disparate side-effect profiles, dosaging, and efficacy.

Highly specialized centers conduct clinical trials. Involved physicians have expertise in handling dementia cases and are highly motivated to recruit patients. Complete laboratory and neuroimaging work-up are also on hand. Subjects in these centers are diagnosed with precision and handled differently as compared to those in the community setting. There is stronger and more accessible patient support system in trial centers. This may have great impact in the care of dementia patients since it would result in better compliance and fewer problem behaviors. This setting again distinguishes clinical trials from routine care.

## 6 OUTCOME ASSESSMENTS IN DEMENTIA TRIALS

Alzheimer's disease affects primarily the memory with subsequent involvement of other cognitive domains. This change causes concomitant deterioration in function and behavior. The efficacy of antidementia drugs therefore should be evaluated based on these domains: cognition, behavior, and function.

The efficacy of an antidementia drug is measured through various outcome assessment scales and psychometric tests. As required by the FDA, the cognitive improvement that results from drug administration should be supported by (1) positive change in a performance-based cognitive instrument and (2) clinically meaningful effect seen globally [48]. These two factors comprise the primary outcome measures in dementia clinical trials. Secondary outcome measures are included to further document the effect of a drug on the other aspects of the subject's life. These secondary outcome measures do not need to be positive for the drug to be approved and marketed. Assessment scales examining the subject's quality of life, noncognitive behavioral symptoms, and economic impact of the illness usually compose the secondary outcomes.

Outcome measures must possess certain properties before these can be employed in clinical trials. The instruments must have been proven valid, reliable, and sensitive. Validity confirms whether the tool measures what it is intended to measure. Reliability (test/retest, interrater and/or intrarater) refers to the replicability of results so that the same value will be obtained under multiple circumstances. Sensitivity is the capacity to detect change over time, especially with treatment. It is best if the instruments are effective in detecting changes even in the extremes of the spectrum of the illness, eliminating floor and ceiling effects. Other important properties are ease of use and short administration time and availability of multiple equivalent forms to avoid practice effects with repeated use. Ideally, these should

be independent of demographic and socioeconomic factors such as gender, education, and cultural background.

### 6.1 Primary Outcome Measures

***Performance-Based Cognitive Assessment*** Cognition is composed of multiple domains, all of which are inevitably affected at the late stage of dementia. Consensus on cognitive domains that may be assessed include the following: memory, attention, processing speed, visuospatial function, praxis, language, executive function, and abstraction [49]. A combination of different psychometric tests can fully evaluate these domains. However, it is more appropriate for disease-specific instruments to be used in a chronically progressive illness so that it can appropriately reflect the outcome at study endpoint.

The Alzheimer's Disease Assessment Scale (ADAS) was designed to evaluate the severity of both cognitive and noncognitive manifestations of AD patients [50]. Cognitive domains include memory, orientation, language, and praxis while the noncognitive domains include mood state and behavioral changes. The cognitive portion has a maximum score of 70 with a higher score indicating more severe impairment. Its advantages include short administration time, ability to detect changes from the mild to severe stages, and being appropriate for patients in different environments [50]. There is excellent interrater and test/retest reliability of the ADAS-cog at 0.99 and 0.92, respectively. It is the most widely used objective cognitive assessment scale in dementia randomized controlled trials (RCTs). The ADAS-cog is the prevailing primary outcome measure of efficacy that evaluates changes in the core manifestations of dementia. The ADAS-cog lacks tests examining executive function, a domain that is affected frequently. The Alzheimer's Disease Cooperative Study (ADCS) has extended the ADAS-cog by adding two executive tests, a cancellation and a maze task.

***Global Measures*** The overall clinical impact of an antidementia drug is evaluated using global assessment scales. These assess the multidimensional manifestations of the illness in terms of cognition, behavior, and function. There are two categories of global measures: (1) global severity scales, which ascertain the absolute severity of the patient's condition, and (2) global change scales, which determine the overall improvement or deterioration of the subject. Global scales have less structure, thus partially avoiding the influence of subject characteristics and rating variance. The available global scales were specifically designed for use in AD or primary degenerative dementia so that results are specific to the illness. These are sensitive measures for long-term assessment of efficacy since changes can be quantified when other assessments are affected by floor effects.

Global assessment scales of change were developed to measure a "clinically meaningful" treatment effect that translates into practical usefulness. All of the measures are seven-point scales which are rated as follows: 1 = very much improved, 2 = much improved, 3 = minimally improved, 4 = no change, 5 = minimally worse, 6 = much worse, 7 = very much worse. These assess change from a specified baseline. Unlike the symptom scales, these are relatively unstructured, relying on an experienced clinician to conduct a thorough and accurate interview on which the

**TABLE 5 Global Outcome Measures (Severity and Change Scales) Utilized in AD and MCI Trials as Either Primary or Secondary Measure of Efficacy**

| Global Outcome Measures | RCT That Used Outcome as Primary Measure of Efficacy | RCT That Used Outcome as Secondary Measure of Efficacy |
|---|---|---|
| Severity | | |
| CDR (global and/or sum of boxes) | [39] | [8, 26, 27, 35–37, 54, 58] |
| GDS | [34] | [8, 22, 25, 28, 32, 41] |
| Change | | |
| CGI [51] | [23, 24, 34, 55] | [39] |
| CIBI [52] | [25] | |
| CIBIC Plus [52] | [21, 22, 26, 27, 29–33, 38, 40, 41] | |
| ADCS-CGIC [53] | [36], MCI version [9] | |

*Note*: CIBIC-plus: Clinical Interview Based Impression of Change plus caregiver information; CGIC: Clinical Global Impression of Change.

rating will be based. Even though reliability is compromised with this format, the sensitivity to measure meaningful change is retained so that it remains as one of the primary efficacy measures.

The first global assessment scale was the Clinician Global Impression (CGI), an unstructured instrument that was widely used in neuropsychopharmacological trials [52]. The CGI was first utilized in a dementia RCT involving tacrine by Davis et al. with ADAS-cog as the other primary measure of efficacy [23]. However, significant improvement was only noted in the ADAS-cog but not in the CGI, which indicated the latter's lack of sensitivity to the treatment effect of tacrine. A guideline-based global assessment scale, the Clinical Interview Based Impression (CIBI), was then developed which was used in the 30-week tacrine trial by Knapp et al. [25]. Both ADAS-cog and CIBI yielded significantly positive results which led to the FDA approval of tacrine. Several global assessment scales emerged to further improve the intrument's reliability (Table 5). To date, CIBIC plus is the most frequently used global assessment instrument in dementia RCTs.

Global severity or staging scales such as the CDR scale [18] and the GDS [17] are frequently used as entry criteria or as secondary outcome measures. The CDR is a worksheet-based semistructured interview that evaluates six domains: three cognitive (memory, orientation, judgment, and problem solving) and three functional (community affairs, home and hobbies, and personal care). Rating is based on a five-point scale in which 0 = none, 0.5 = questionable dementia, 1 = mild impairment, 2 = moderate impairment, and 3 = severe impairment. It can be scored in two ways: (1) as a sum of boxes (SB) by obtaining the sum of the ratings of each of the six CDR domains/boxes and (2) as a global rating based on a scoring system wherein the memory box/domain is the main consideration. Preference for this instrument is due to its clinically based assessment and high interrater reliability resulting in a level of agreement of 80% [56]. The GDS is also a useful instrument in staging primary degenerative dementia. It is capable of accurately delineating stages of dementia throughout the course of AD [17]. It rates cognitive decline based on a seven-point scale with the following scoring system: 1 = none, 2 = very mild, 3 = mild, 4 = moderate, 5 = moderately severe, 6 = severe, 7 = very severe. Interrater and test/retest reliability are high, both at 0.92 [57].

### 6.2 Secondary Outcome Measures

***Mini–Mental State Examination*** The MMSE may be included as an entry criterion or as a supplementary measure of cognition [8, 21, 23, 25–28, 32, 34, 36, 38, 39, 41, 54, 55, 58]. Compared to the ADAS-cog, its result is better understood by non-AD specialists and can be easily translated into more practical terms. However, its limited sensitivity makes it a poor primary outcome measure.

***Activities of Daily Living*** Functional impairment is an essential component of the clinical syndrome of dementia. It is required for the clinical diagnosis of dementia and included in the NINCDS-ADRDA criteria [12] and the DSM-IV criteria [10]. The resulting dependence affects not only the patient but also the quality of life of the caregiver and becomes an important factor that leads to institutionalization. Changes in activities of daily living (ADLs) are frequently included as a secondary outcome measure. Functional deterioration is only moderately correlated with the cognitive status of patients with AD [59] and seems to be an expression of other integrative abilities of the individual. This makes functional assessment all the more important in clinical trials since cognitive tests cannot fully gauge improvement in other aspects of the patient's life.

Drug effect in terms of reversibility, stabilization, or slower deterioration of ADLs can be monitored through the use of functional assessment scales [60]. Different instruments for this purpose have been developed for AD (Table 6), incorporating either basic or instrumental/complex ADLs or both. It is necessary for complex ADLs to be incorporated in the examination since functional deterioration occurs in hierarchical order, initially affecting difficult tasks before simpler ones. Ideally, caregiver information should be obtained for data reliability since loss of insight as the disease progresses makes self-report impossible.

These functional assessment scales were used as either primary or secondary outcome measures of efficacy in different dementia RCTs (Table 7)

***Neuropsychiatric Symptoms*** Neuropsychiatric manifestations are common in AD. They reflect the underlying neuropathological and neurotransmitter changes in the brain. Incidence increases with disease severity but symptoms are variable and differ among afflicted individuals. Assessment of these behavioral manifestations is essential since it can influence the individual's state, further aggravating the existing cognitive and functional impairment. Presence of disruptive behavior increases caregiver burden and is one of the determinants of eventual nursing home placement. These symptoms can also predate the onset of dementia and may present in MCI.

Several instruments that measure behavioral symptoms are available and employed in dementia clinical trials. These are used as primary outcome measures when the target symptom is behavior (e.g., with psychotropic agents) and as secondary outcome measures when the target symptom is cognition (e.g., with cholinergic agents) and behavior is evaluated as an auxiliary effect. Characteristically, these can be categorized into broad-spectrum scales when they sample a comprehensive range of symptoms and focused scales when more items are dedicated into the subtleties of a particular behavioral domain [64]. Some characterize the symptoms by their frequency and severity, which is very helpful in gauging the disruptiveness of the behavior and becomes an indirect measure of caregiver burden. Frequently used

**TABLE 6 Comparison of General Characteristics and Reliability of Different Functional Assessment Scales Utilized in Dementia RCT**

| Description of Instrument | PSMS and IADL [61] | PDS [62] | DAD [59] | ADCS-ADL [63] | ADFACS [54] |
|---|---|---|---|---|---|
| General characteristic | Caregiver questionnaire | Caregiver questionnaire developed as measure of QOL changes; bipolar analog scale | Proxy respondent scale (completed either by caregiver as questionnaire or as structured interview) | Informant-based questionnaire | Informant-based questionnaire |
| Population/environment | AD/community and residential care | AD/not indicated | AD/community dwelling | AD/community dwelling | AD/not indicated |
| Administration time | Not indicated | 10–15 min | <15 min | 30–45 min | Not indicated |
| Gender influence | IADL different for males and females (5 pts. male; 8 pts. female) | Not indicated | No influence (including age and education) | Gender-biased items removed | No influence |
| ADL inclusion | Both BADL and IADL | Both BADL and IADL | Both BADL and IADL | Both BADL and IADL | Both BADL and IADL |
| Number of items (score range) | PSMS 6 items, IADL 8 items | 29 QOL factors (0–100) | 40 items (0–100) | 24 items (0–78) | 16 items (0–54) |
| Correlation with dementia severity | Not indicated | Yes (GDS) | Yes (MMSE and GDS) | Yes (MMSE) | Not indicated |
| Reliability | Test/retest: reproducibility coefficient 0.96; interrater: 0.87–0.91 (Pearson *r*) | Test/retest: 0.898 (Pearson product-moment correlation) | Test/retest: 0.96 (intraclass correlation coefficient); interrater: 0.95 (intraclass correlation coefficient) | Test/retest: 0.4–0.75 (*k* statistics) | Not indicated |
| Special qualities | | | Evaluates aspects of activities that are impaired (initiation, planning, organization, effective performance); nonapplicable domain does not influence scoring since total score is converted into a percentage. | | |

*Note*: PSMS: Physical Self-Maintenance Scale; IADL: instrumental activities of daily living; PDS: Progressive Deterioration Scale; DAD: Disability Assessment in Dementia; ADCS-ADL: Alzheimer's Disease Cooperative Study—Activities of Daily Living; ADFACS: Alzheimer's Disease Functional Assessment and Change Scale; BADL: Basic Activities of Daily Living; QOL: quality of life.

**TABLE 7 Functional Assessment Scales Utilized in Different Dementia RCTs as Primary or Secondary Outcome Measures of Efficacy**

| Functional Assessment Scale | RCT That Used Scale as Primary Outcome Measure | RCT That Used Scale as Secondary Outcome Measure |
|---|---|---|
| PSMS and IADL | | [21, 23, 25, 34, 39, 40] |
| PDS | [32, 33] | [23, 25, 28] |
| DAD | | [22, 30, 31, 40] |
| ADCS-ADL | Adapted to severe state [41] | MCI version [8]; [29, 37, 38] |
| ADFACS | [54] | |

assessment scales in dementia clinical trials are the following: Neuropsychiatric Inventory (NPI) [65], Behavioral Pathology in Alzheimer's Disease Rating Scale (BEHAVE-AD) [66], CERAD Behavior Rating Scale for Dementia (CBRSD) [67], Cohen Mansfield Agitation Inventory (CMAI) [68], Cornell Scale for Depression in Dementia (CSSD) [69], Brief Psychiatric Rating Scale [70, 71], and Positive and Negative Syndrome Scale (PANSS) [72].

The NPI, BEHAVE-AD, and CBRSD are examples of broad-spectrum scales developed primarily for the evaluation of dementia patients (Table 8). These are valid and reliable instruments that can be used in clinical and research settings. The NPI is widely used in dementia clinical trials due to its comprehensive list of symptoms that commonly occur in individuals with dementia. It also has accompanying questions for every symptom so that appropriate probing can be done and includes a frequency and severity component for every symptom category. Major dementia RCTs [22, 28–30, 37, 40, 41] have used the NPI as a standard gauge of behavior outcome. CBRSD was used in the trial by Sano et al. [58] as a secondary outcome measure. The BEHAVE-AD was more frequently employed in trials involving psychotropic agents [74–76].

The CMAI and CSDD are focused scales. These are useful when particular symptoms are targeted by the clinical trial. For example, the CMAI was used as a primary outcome measure to demonstrate the efficacy of risperidone in agitation and aggression in subjects with AD [76]. The CSDD was likewise the primary outcome measure in the sertraline RCT [77] that determined the drug's effect on depression in subjects with AD.

The BPRS and PANSS are broad-spectrum scales that were developed for use in psychiatry populations. These have been employed in several dementia trials involving psychotropic agents since it covers a variety of symptoms aside from psychosis. The BPRS was the primary measure of efficacy for the control of agitation and aggression in dementia in carbamazepine [78] and divalproex sodium RCTs [79]. The PANSS has five symptomatic dimensions (negative, positive, excitation, cognitive, and depression) and the quetiapine trial [80] used the excitation component (PANSS-EC) to evaluate the drug's efficacy in agitation that occurs in dementia.

***Other Measures of Efficacy*** Aside from treatment efficacy, pharmacoeconomics, caregiver burden, and quality of life in dementia have been addressed in dementia RCTs. Although these are not required as outcome measures for drug approval,

their usefulness is based on a more practical aspect of efficacy, one that can be seen and experienced in daily life. Caregiver burden scales determine the impact of dementia on the people who are considered direct health care providers. They experience stress and burden. One way of determining burden is by recording how much time is spent attending to the patient. This can be assessed using the Caregiver Activities Time Survey, the Caregiver Activity Survey [81], or the Caregiver Time Questionnaire [82].

The economic impact of dementia is overwhelming since the burden extends to the national level. It is directly related to the number of people affected by the disorder and the duration of the disease [83]. Incurred costs include (1) medical resources used to treat the illness (medication, cost of institutionalization, etc.), (2) nonmedical resources, and (3) lost productivity caused by disability (both patient and caregiver) [83]. Since maintenance of patients with dementia eventually translates into cost of care, some RCTs include pharmacoecomomic measures. The AD2000 trial [84] determined the health care cost of dementia from the societal perspective in terms of delaying institutionalization or requiring other health services and the cost-effectiveness of the drug (donepezil). Reisberg et al. [41] similarly evaluated this effect by utilizing the Resource Utilization in Dementia (RUD) [85], which assesses the burden on the caregiver and provides AD-related health economics data, in a memantine trial.

Quality-of-life scales encompass the overall effect of the treatment on the patient in terms of social, psychological, cognitive, and functional well-being. Either the patient or the caregiver's quality of life can be assessed. However, information will be derived mainly from the caregiver once the patient's communication problems emerge.

Transcultural application of these instruments may alter their validity and reliability. Interpretation of results should therefore be done with caution when these are used outside the cultures in which they were developed.

## 7 CLINICAL TRIAL DESIGNS

The FDA requires that agents tested should be superior to placebo in terms of its cognitive and global effects. The standard format of antidementia drug RCTs is the parallel group design wherein there is direct comparison between the experimental and placebo group. In this head-to-head comparison, the active drug and placebo are allocated to different subjects randomly. The 12-week tacrine study by Farlow et al. [24] utilized this design, although randomization of subjects was stratified. Alzheimer's disease drug trials (acetylcholinesterase inhibitors, NMDA antagonists, psychotopics, SSRIs, etc.) follow this standard format due to its straightforward approach. This design was also used in the MCI trial of Salloway et al. [9] comparing donepezil with placebo. Between-patient variability can deleteriously affect the outcome of parallel group studies and cannot be eliminated with this format. The crossover design eradicates this confounding variable by applying the treatment sequentially to the same subject. Aside from reducing error variance, sample size is diminished since subjects can be used several times in the analysis while increasing the statistical power of the clinical trial. However, this strategy is rarely used primarily due to difficulty in reobtaining baseline values during the

**TABLE 8 Comparison of Characteristics of NPI, BEHAVE-AD, and CERAD, CBRSD**

| Instrument Characteristics | BEHAVE-AD | CBRSD | NPI |
|---|---|---|---|
| Overview | Developed through retrospective review of 57 AD outpatient charts using information from nursing staff, physicians, and family members | Items from literature review and expert panel; some items drawn from other scales; items are homogeneously scaled, anchored; emphasize frequency rather than severity | Assesses a wide range of behaviors with potential to differentiate between dementia syndromes; items suggested by expert panel by "Delphi" method; format used minimizes administration time and optimizes capture of information |
| Population | AD outpatients with mild to severe cognitive loss | AD outpatients | Elderly dementia patients |
| Informant | Caregiver | Caregiver | Caregiver |
| Administration | Clinician in person or by phone or self-administered | Structured interview by technician | Clinician interview |
| Administration time | 45 min | 30 min | 20 min |
| Time interval | 2 weeks | 1 month (behaviors occurring prior to 1 month, since dementia onset, are also noted) | 1 month |
| Items | 25 plus 2 global ratings | 48 | 10 items in 5 domains: mood changes (dysphoria, euphoria), agitation (aggression, aberrant motor behavior), personality alterations (apathy/indifference, irritability/lability, disinhibition), psychosis (delusions, hallucinations), and anxiety; sleep and appetite changes included later on |

| | | | |
|---|---|---|---|
| Frequency | No | 0 = not since illness began; 1 = 1–2 days/past month, 2 = 3–8 days, 3 = 9–15 days, 4 => 16 days, 8 = since illness began but not in past month | 1 =< 1 time/wk, 2 = 1 time/wk, 3 = several times/week but <1 time/day, 4 =≥ 1/day |
| Severity | Rating differs for each specific item; includes severity estimates that indirectly consider caregiver impact: absent = 0, mild = 1, moderate = 2, severe = 3 | No | 1 = mild, 2 = moderate, 3 = severe |
| Caregiver impact | Directly assessed by items 21 and 23 and part II | No | No |
| Reliability | Interrater: $r = 0.90$ (severity ratings for scale as a whole); reliability of symptom presence or absence = 0.62–1.0 | Interrater: $k = 0.77–1.0$ ($n = 104$) | Interrater: $r = 0.96–1.0$ (frequency), 0.98–1.0 (severity); Test/retest: $r = 0.79$ (frequency), $r = 0.86$ (severity); Overall (Cronbach *A*) $r = 0.88$ |
| Validity | Has content and construct validity (derived from patient records) | Has contruct and content validity; convergent validity with CMAI | Has construct and content validity; convergent validity with HAM-D and BEHAVE-AD |
| Scoring | Total score or in seven domains (paranoid/delusional ideation, hallucinations, activity disturbances, aggressive behavior, sleep disturbance, affective symptoms, anxiety/phobic symptoms) | Initial study suggested eight factors: depressive symptoms, psychotic symptoms, poor self-regulation, irritability/agitation, vegetative features, apathy, aggression, affective lability | For each of the 10 items, total score = frequency × severity |
| Suggested use | Demonstration of behavioral disturbance and quantification of behavioral effects of antipsychotics or other interventions | Detailed elicitation/quantification of broad range of psychopathology in mild to moderate AD patients | Assessment and quantification of symptoms |

*Source*: From [73].

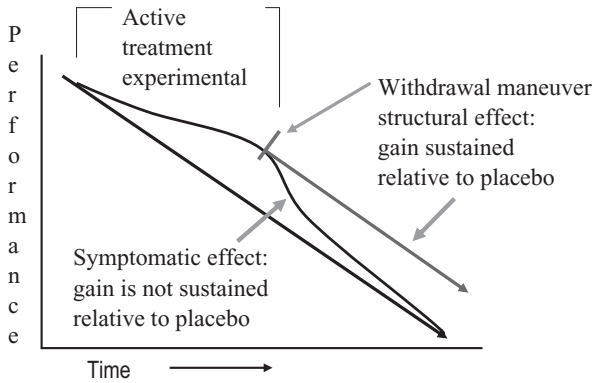


***FIGURE 1*** *Randomized withdrawal design. (Adapted from [86])*

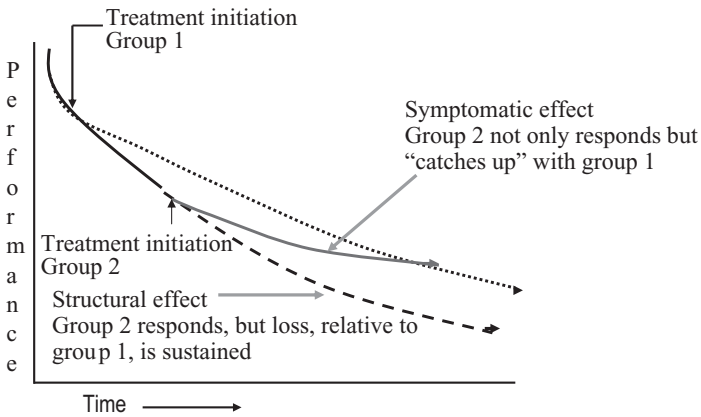


***FIGURE 2*** *Randomized start design. (Adapted from [86])*

treatment shift. There is also uncertainty if the changes noted with the second treatment are due to the remaining or carry-over effect of the first one. The required washout periods of the drugs are unknown and may be lengthy, causing prolongation of the clinical trial.

The current trend in antidementia drug development is toward disease modification. The above-cited study designs are adequate for measuring symptomatic change (has no effect in the survival and function of degenerating neurons) and are less applicable to assessing disease modification. A variety of trial designs have been suggested to assist in establishing disease-modifying effects. Paul Leber [86] suggested the randomized withdrawal and randomized start designs to determine the disease-modifying effect of the antidementia drug (Figures 1 and 2). The randomized withdrawal design (Figure 1) is based on the hypothesis that withdrawal of a disease-modifying drug will not produce loss of the incremental changes noted compared to the withdrawal of a symptomatic drug. The major drawback of this design is the uncertainty as to how long the subjects need to be observed to determine the outcome. Consequently, it will require a considerable number of subjects since attrition is expected to be high secondary to the length of the trial and with-

drawal of a treatment that may be beneficial to the subject. Blinding is also compromised since everyone becomes aware that all patients are taking placebo at the end of the trial. These potential limitations led to the formulation of the staggered start design (Figure 2) which has two phases: The first is similar to a placebo-controlled, randomized, parallel study while the second entails all subjects to be given the active treatment. A significant drug placebo difference should be existent after the first phase before the second phase can be started. If the active drug has a disease-modifying effect, the placebo group's performance cannot catch up with that of the treatment group in the second phase of the study.

Sano et al. [87], utilized a time-oriented, valid outcome measure to assess disease progression that includes the main domains affected in AD—cognition and function. The survival design examines the longitudinal course of dementia to measure the antidementia drug effect on disease progression. Endpoints are not primarily based on neuropsychological tests, unlike in other trial designs. Instead, readily observable, clinically meaningful endpoints that represent disease progression are utilized. The primary efficacy outcome measure is the time to reach specified endpoints. Similarly, Mohs et al. [54] used a survival design to evaluate the length of time the function is preserved in AD patients with the use of donepezil.

Amnestic MCI lends itself to use of survival designs since longitudinal examination of patients can reveal either progression or nonprogression to AD. Most MCI clinical trials used this study design with progression to AD as the endpoint [8, 88]. A limitation of this scheme is the difficulty in precisely defining progression to AD.

An enrichment design was utilized in the tacrine study of Davis et al. [23]. The first or enrichment phase consisted of finding the "best dose" response of each patient to tacrine as defined by the ADAS without producing intolerable side effects. Subjects who continued in the subsequent parallel group randomized controlled trial were the "enriched population" or those who had a potential to respond to the treatment as seen in the first phase of the study. RCTs which include subjects who have a higher risk of acquiring AD such as those with positive family history, presence of ApoE4 alleles, presence of cerebrospinal fluid (CSF) AD biomarkers, hippocampal atrophy on magnetic resonance imaging/computerized tomography (MRI/CT) scan, or bilateral temporoparietal hypometabolism on positron emission tomography (PET) scan are recruiting an "enriched population." A drawback of this study design is the potential to skew the outcome favorably toward the active drug and in the end limit the generalizability of the findings. Enrichment design for an antidementia drug is useful in proof-of-concept studies or as a means of entering the market by showing evidence of its efficacy. However, it should be followed by a trial that employs less restrictive designs so as not to limit the generalizability of findings.

### 7.1 Special Clinical Trial Design Features

Long-term drug trials may incorporate special design features to improve accuracy of findings and to confirm additional hypotheses. Single-blind run-ins added before initiation of the double-blind phase involve administration of placebo to both treatment groups for a specified period of time without the knowledge of subjects and caregivers. Its primary aim is to minimize the "placebo effect" to which the initial changes from baseline may be contributed. By doing this, both groups equally start

at the same level. The galantamine efficacy and safety trials utilized this feature [29–31].

Single-blind washout periods are sometimes done at the end of the double-blind treatment. This entails administration of placebo to both treatment groups in an effort to determine the withdrawal effect of the study drug. In the donepezil trials [26, 27], the washout periods showed the return of the performance of the donepezil group to baseline values—an indication that the observed improvement after drug administration was a symptomatic and not a disease-modifying effect. This is also useful for the assessment of untoward reactions secondary to drug removal and to gain information on how long the clinical response is sustained after drug discontinuation.

Open-label extensions conducted after the double-blind study is a method of determining safety of prolonged treatment. Limited efficacy data are collected in open-label extensions.

### 7.2 Randomization

Randomization allocates subjects to treatment arms wherein each arm is exposed to a different condition. Usually only two arms are employed in phase III trials: experimental and control. In phase II or when several treatment strategies are available, additional treatment arms become necessary. For example, in the MCI trial by Petersen et al. [8], three arms were utilized and subjects received (1) 2000 IU of vitamin E daily, (2) 10 mg of donepezil daily, or (3) placebo. Sano et al. [58] likewise expanded the treatment allocation by means of a $2 \times 2$ factorial design to further examine treatment combinations. Four treatment arms were used: (1) selegiline and α-tocopherol, (2) placebo and α-tocopherol, (3) selegiline and placebo, and (4) placebo only. Multiple treatment arms are used to compare response to different dosages of the drug [26–32, 36].

### 7.3 Length of Clinical Trials

Duration of the trial and number of recruited patients are very important factors that contribute to the validity of the outcomes of the RCTs. Duration of phases II and III should be long enough to ensure that the efficacy and safety of the drug are adequately explored. Efficacy of antidementia drugs cannot be seen before 3 months of drug exposure and may be evident only after 6 months. Although longer duration (>6 months) is ideal, the number of subjects completing the study and their compliance are concerns in longer studies.

Adequate clinical trial length is essential in demonstrating treatment effect in a progressively deteriorating disorder. Drug effect (symptomatic vs. disease modifying), target symptoms (cognitive vs. noncognitive/behavioral), and outcome measures (biological vs. assessment scales) are important factors that should be considered in determining the trial length. Trials with insufficient duration can yield inaccurate and potentially misleading results in view of the fact that small but important changes from baseline may remain undetected. Longer trials (≥1 year) are ideal considering the chronicity of the disorder being examined, but ethical, attrition, and compliance problems may be encountered which could com-

promise study interpretation. Winblad et al. [28] conducted the first published long-term (1-year) efficacy and safety study of donepezil on AD while the AD (2000) trial [84] has extended the duration to 2 years. Most trials involving agents that may have a neuroprotective or disease-modifying effect are conducted for 1 year (Table 4) since time must be allotted for their structural effect to become apparent.

Neuropsychiatric symptoms are encountered throughout the spectrum of dementia and may even herald its onset. Numerous studies have been conducted involving psychotropics (typical and atypical), anticonvulsants (carbamazepine and valproate), and antidepressants [selective serotonin reuptake inhibitors (SSRIs)] to determine which agent alleviates these symptoms. Patterns of analysis may be either (1) reduction in emergence of behavioral symptoms wherein asymptomatic patients are followed up longitudinally to determine which regimen has fewer symptoms at endpoint or (2) comparing which regimen produced more change/reduction in the behavioral symptoms from baseline to endpoint. The length of the trials is notably shorter than those addressing cognitive symptoms, ranging from 6 to 12 weeks, since acute symptomatic improvement is the goal.

## 8 STATISTICAL ANALYSES

Subject noncompliance and dropouts always complicate clinical trials. These subjects cannot be excluded from analysis since they may have demographic or disease characteristics that are different from those who completed the trial and adhered to the treatment randomly assigned to them. It is for this reason that most clinical trials use the intention-to-treat principle, which provides unbiased and reliable interpretation of treatment effect. Intention-to-treat analyses typically include all subjects who were randomized to treatment, received at least one study drug dose, and provided a baseline assessment and at least one postbaseline assessment. It is a conservative method that makes it possible for the potential treatment benefit on patients to be evaluated regardless of whether or not they completed the study. With this, the random assignment of subjects to treatment groups is preserved during data analysis and potential bias is reduced [44]. This differs from the fully evaluable (per-protocol) population analysis wherein those who completed the entire phase of the study and remained compliant to the treatment regimen based on compliance rules set prior to study initiation are the only ones included.

Analysis of longitudinal data can be problematic since missing data are evident. Different methods of statistical analysis have been adopted to treat these data sets. The most widely used technique that addresses missing data is the last observation carried forward (LOCF), a method in which the subject's last available assessment is imputed for all remaining unobserved response measurements. It has the advantage of preserving the sample size, but it presumes that the subjects' responses have been constant from the last observed value to the trial endpoint. This unwarranted assumption about the missing data may result in either underestimating or overestimating the treatment effects. Type I errors can be generated from this so that a treatment difference can be falsely endorsed when in fact there is none. Despite these caveats, the method is frequently used due to its simplicity and ease of implementation and relatively conservative method of treating data. On the other hand, an observed cases (OC) analysis utilizes only the data of subjects remaining in the

trial at a specified point in time. In this method, a direct relationship of the data used and the obtained results is observed. However, loss of power and subsequent validity of results may occur since data of the noncompleters are unexploited. Results that are statistically significant for both types of analysis clearly support their accuracy. Conversely, cautious interpretation of results should be done when results are significant only in the OC analysis.

All statistical analysis should be planned in detail prior to the initiation of the clinical trial since changes afterward may introduce bias in the system. However, changes that are made prior to breaking of the blind still have limited implications for study interpretation. Analyses that are made afterward are less compelling. This also applies for specification of subgroup analysis, which is done on the basis of an expectation of a larger treatment effect in some subgroups than in others [44].

Clinical trials with survival data require different statistical treatment since subjects have varying endpoints producing asymmetries in data distribution. If the data from the subjects who did not make it to the endpoint are excluded, bias may be introduced into the results. Studies using the survival design have three general objectives: to estimate the time to event, to compare the time to event between/among the groups, and to determine the relationship of the covariables to the time to event. The hazard ratio and survival time are two important functions that are examined to generate the answers to these questions [89]. The Cox proportional-hazards regression and Kaplan–Meier analysis were used to evaluate these, respectively, in the trials of Mohs et al. [54], Sano et al. [58], and Petersen et al. [8]. The Cox proportional-hazards model controls for any bias in the predetermined covariates among the treatment arms since these change over time. It is used to estimate the hazard ratio, which is the risk of progression to an event over time in the treatment group versus the placebo group. The Kaplan–Meier method provides survival time estimates to clinically evident decline or to a chosen event (e.g., death, institutionalization, or loss of the ability to perform basic ADLs).

An adequate sample size ensures that the treatment effect can be reliably derived from the clinical trial at a specified endpoint. For ethical reasons, the sample size should be well justified; samples too small or too large are not warranted. Factors that affect sample size determination are the power of the study to detect a drug–placebo difference and the chosen level of significance of the statistical tests. Confounders and attrition rate should be considered in sample size determination, for which it should be adjusted appropriately.

The study should have enough "power" to accurately detect the smallest possible difference in the primary outcome measure that has clinical significance produced by the treatment. Power is usually set at 80% so that there is a 20% probability of missing the difference between the treatment and placebo group. Some clinical trials use a power of 90% to further reduce the chance of a false-negative result. The level of significance, or $p$ value, is the probability of incorrectly identifying a treatment difference between the treatment and placebo arm when actually there is none (false-positive result). By convention, a value of $\leq 0.05$ is frequently used. Sample size is inversely proportional to the chosen level of significance while it is directly proportional to the power of the study.

Sample size calculation is based on the primary outcome measure and how much change is required to produce a clinically meaningful effect. Previous phases II and

III clinical trials and longitudinal studies establish these changes. For example, in the ADAS-cog a four-point change is utilized for 6-month trials and seven-point change for 1-year trials for clinically significant change to be detected [90]. For the CIBIC-plus a 0.3–0.4 change is usually targeted [26, 40]. In the 1-year study of Mohs et al. [54] where preservation of function was determined as an effect of donepezil, power was calculated based on functional performance, that is, the 1-year value for significant functional decline in the placebo and donepezil group based on a previous study. Most base the power of the study on one of the primary outcomes (either ADAS-cog or the global assessment) while some base it on dual outcomes (both ADAS-cog and global assessment) [21, 22].

## 9 DRUG–PLACEBO DIFFERENCE

The drug–placebo difference is the discrepancy between the deterioration of the placebo group and the improvement, stabilization, or reduced deterioration in the treatment group [46]. It is derived by determining the difference between the mean change from baseline scores of the actively treated and placebo groups at a specified endpoint. The FDA requires proof of efficacy in terms of statistically significant improvement on specified outcome measures in the treatment group.

The effect size is the definitive basis of a drug's efficacy. It is determined by dividing the drug–placebo difference by the standard deviation using specific outcome measures.

A summary of the drug–placebo differences of the RCTs in MCI and AD based on outcome measures can be seen in Tables 9, 10, and 11. The drug–placebo difference varies among the studies and among the class of therapeutic agents. The range of treatment effect produced by acetylcholinesterase inhibitors in the ADAS-cog is 2–3.9 and 0.2–0.47 in the global assessment scales among subjects with mild to moderate AD. Among the other class of therapeutic agents, only Ginkgo biloba [55] produced significant drug–placebo difference in the ADAS-cog, but this was not supported by the global evaluation.

## 10 PLACEBO RESPONSES

Use of a placebo arm in a RCT is a standard procedure as long as it is ethical and feasible. Comparison of drug–placebo outcomes determines the investigational drug's efficacy. In AD, it is expected that an efficacious drug will produce improvement or stabilization in primary assessments while the placebo group continues its course of deterioration.

Irregularities in the placebo response are apparent in some clinical trials (Figure 3). Factors that contribute to this effect are fluctuation of the symptoms, methodological inconsistencies, and the beneficial effects of improved medical care provided during the study. It is hypothesized that placebo effects or trial effects can result from the attention that subjects receive from health care providers involved in the study. In some studies, the occurrence of adverse events in the placebo group can approach the level of occurrence observed in the treatment group. Placebo and trial effects may account for the initial improvement that is seen in the placebo

**TABLE 9 Drug–placebo Differences on ADAS-cog and Global Assessment Scales among Various Therapeutic RCTs in MCI and Mild to Moderate AD (Intent-to-Treat Analyses[a])**

| | ADAS-cog | | Global Assessment Scale (CGI, CGIC, CIBIC-Plus) | |
|---|---|---|---|---|
| Agent/Study | D/P Difference | Significance (*p* Value) | D/P Difference | Significance (*p* Value) |
| Tacrine | | | | |
| Farlow et al.[b] [24] | | | | |
| 20 mg | 0.9 | NS | 0 | NS |
| 40 mg | 1.4 | NS | 0.1 | NS |
| 80 mg | 3.8 | 0.015 | 0.5 | 0.015 |
| Davis et al. [23] | | | | |
| 40 or 80 mg | 2.4 | <0.001 | NR | NS |
| Knapp et al. [25] | | | | |
| 80 mg | 1.4 | NS | 0.1 | NS |
| 120 mg | 2 | 0.008 | 0.2 | 0.04 |
| 160 mg | 2.2 | 0.002 | 0.2 | 0.04 |
| Donepezil | | | | |
| Rogers et al. [27] | | | | |
| 5 mg | 2.49 | <0.0001 | 0.36 | 0.0047 |
| 10 mg | 2.88 | <0.0001 | 0.44 | 0.0001 |
| Rogers et al. [26] | | | | |
| 5 mg | 2.5 | <0.001 | 0.3 | 0.003 |
| 10 mg | 3.1 | <0.001 | 0.4 | 0.008 |
| Galantamine | | | | |
| Tariot et al. [29] | | | | |
| 16 mg | 3.1 | <0.001 | 0.41 | <0.001 |
| 24 mg | 3.1 | <0.001 | 0.44 | <0.001 |
| Raskind et al. [31] | | | | |
| 24 mg | 3.9 | <0.001 | 0.28 | <0.01 |
| 32 mg | 3.4 | <0.001 | 0.29 | <0.05 |
| Rockwood et al. [30] | | | | |
| 24 or 32 mg | 1.7 | <0.01 | NR | 0.01 |
| Rivastigmine | | | | |
| Corey-Bloom et al. [32] | | | | |
| 1–4 mg | 1.73 | NR | 0.26 | NR |
| 6–12 mg | 3.78 | <0.001 | 0.29 | <0.01 |
| Rosler et al. [33] | | | | |
| 1–4 mg | 0.03 | NS | 0.14 | NS |
| 6–12 mg | 1.6 | NS | 0.47 | <0.001 |
| Metrifonate | | | | |
| Cummings et al. [21] | | | | |
| 10–20 mg | 1.5 | 0.02 | 0.04 | NS |
| 15–25 mg | 1.3 | NS | 0.29 | 0.005 |
| 30–60 mg | 2.94 | 0.0001 | 0.35 | 0.0007 |
| Morris et al. [22] | | | | |
| 30–60 mg | 2.86 | 0.0001 | 0.28 | 0.0071 |
| Mulnard et al. [36] | | | | |
| Estrogen 0.625 or 1.25 mg/d | 2 | NS | 0.1 | NS |
| Le Bars et al. [55] | | | | |
| Ginkgo biloba 120 mg/d | | | | |
| a. AD and MID | 1.4 | 0.04 | 0 | NS |
| b. AD only | 1.7 | 0.02 | 0 | NS |
| Reines et al. [38] | | | | |
| Rofecoxib 25 mg/d | 0.6 | NS | 0.03 | NS |
| Scharf et al. [34] | | | | |
| Diclofenac/Misoprostol | 1.14 | NS | 0.24 | NS |

*Note*: Only studies with complete data are included. ADAS-cog: Alzheimer's Disease Assessment Scale; D/P: drug–placebo.

[a]Except Farlow et al. [24].

[b]Evaluable population analysis.

**TABLE 10 Drug–Placebo Differences on MMSE among Various Therapeutic Randomized Controlled Trials in MCI and Mild to Moderate AD (Intent-to-Treat Analyses[a])**

| | MMSE | |
|---|---|---|
| Agent/Study | D/P Difference | Significance (*p* Value) |
| Tacrine | | |
| Farlow et al.[b] [24] | | |
| 20 mg | 0.1 | NS |
| 40 mg | 0.4 | NS |
| 80 mg | 0.7 | NS |
| Davis et al. [23] | | |
| 40 or 80 mg | 0.7 | NS |
| Knapp et al. [25] | | |
| 80 mg | 0.6 | NS |
| 120 mg | 0.4 | NS |
| 160 mg | 0.9 | 0.02 |
| Donepezil | | |
| Rogers et al. [27] | | |
| 5 mg | 1.21 | 0.0007 |
| 10 mg | 1.36 | 0.0002 |
| Rogers et al. [26] | | |
| 5 mg | 0.96 | <0.004 |
| 10 mg | 1.26 | <0.001 |
| Mohs et al. [54] | | |
| 10 mg | NR | NS |
| Winblad et al. [28] | | |
| 10 mg | NR | Significant |
| Rivastigmine | | |
| Rosler et al. [33] | | |
| 1–4 mg | 0.15 | NR |
| 6–12 mg | 0.68 | <0.05 |
| Metrifonate | | |
| Cummings et al. [21] | | |
| 10–20 mg | 1.11 | 0.003 |
| 15–25 mg | 0.63 | NS |
| 30–60 mg | 1.37 | 0.003 |
| Morris et al. [22] | | |
| 30–60 mg | 0.43 | NS |
| Sano et al. [58] | | |
| Vitamin E, selegiline, or both | NR | NS |
| Mulnard et al. [36] | | |
| Estrogen 0.625 or 1.25 mg/d | 0.4 | NS |
| Thal et al. [39] | | |
| Acetyl-L-carnitine 3 g/d | 0.4 | NS |
| Reines et al. [38] | | |
| Rofecoxig 25 mg/d | 0.44 | NS |
| Scharf et al. [34] | | |
| Diclofenac/misoprostol | 1.37 | NS |
| Feldman et al.[c] [40] | | |
| Donepezil 10 mg | 1.79 | <0.0001 |
| Reisberg et al. [41] | | |
| Memantine 20 mg | 0.7 | NS |
| Petersen et al. [8] | | |
| Donepezil 10 mg | 0.44 | NS |

*Note*: Only studies with complete data are included. NS, not significant; NR, not reported.
[a]Except Farlow et al. [24].
[b]Evaluable population analysis.
[c]sMMSE—standardized Mini–Mental State Examination.

**TABLE 11 Comparative Effects of Therapeutic Agents in AD Clinical Trials on ADAS-cog and MMSE at 6 months (Intent-to-Treat Analyses[a])**

| Agent/ RCT | Year | Treatment Duration (weeks) | ADAS-cog Drug–Placebo Difference | MMSE Drug–Placebo Difference |
|---|---|---|---|---|
| Metrifonate [22], 30–60 mg | 1998 | 26 | 2.86 | |
| Tacrine [25] | 1994 | 30 | | |
| 80 mg | | | 2.2 | 0.8 |
| 120 mg | | | 3.1 | 0.9 |
| 160 mg | | | 4.2 | 2.5 |
| Donepezil [28] | 2001 | | — | |
| 10 mg | | 52 | | 1.4 |
| Rivastigmine | | | | |
| [33] | 1999 | 26 | | |
| 1–4 mg | | | 0.03 | 0.15 |
| 6–12 mg | | | 1.6 | 0.68 |
| [32] | 1998 | 26 | | |
| Galantamine [31] | 2000 | 52 | | |
| 24 mg | | | 3.9 | |
| 32 mg | | | 3.4 | |
| Ginkgo biloba [55] | 1997 | 52 | 1.33 | |
| Estrogen [36] | 2000 | 52 | 1.3 | 0.7 |

*Note*: Included trials have duration of ≥6 months and either have data at 6 months or have graphs from which data can be extrapolated.

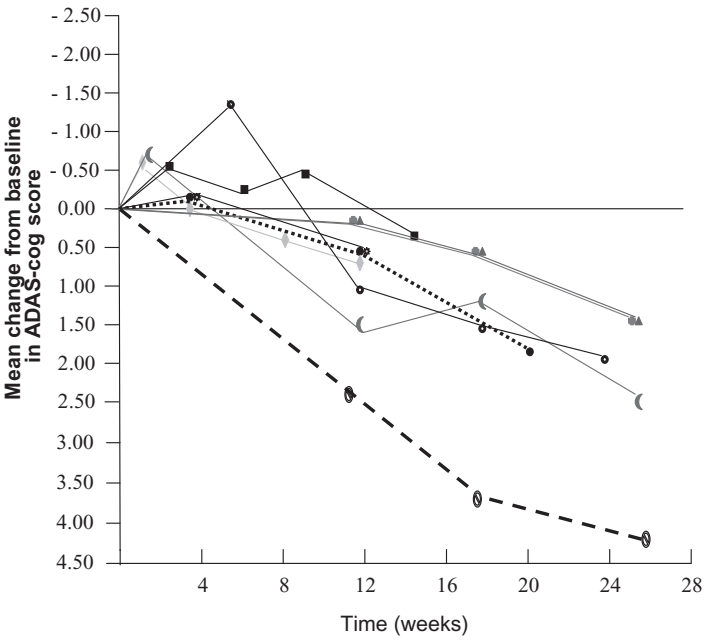


***FIGURE 3*** *Mean change from baseline (intent-to-treat analyses) in ADAS-cog scores of placebo groups of different acetylcholinesterase inhibitor clinical trials with duration of 3–6 months in mild to moderate AD (mean change from baseline derived from available data while other values were extrapolated from graphs): ▲ [31], ■ [33], ☾ [22], ● [29], ○ [27], ♦ [21], ✱ [30], ■ [26], ◊ [32].*

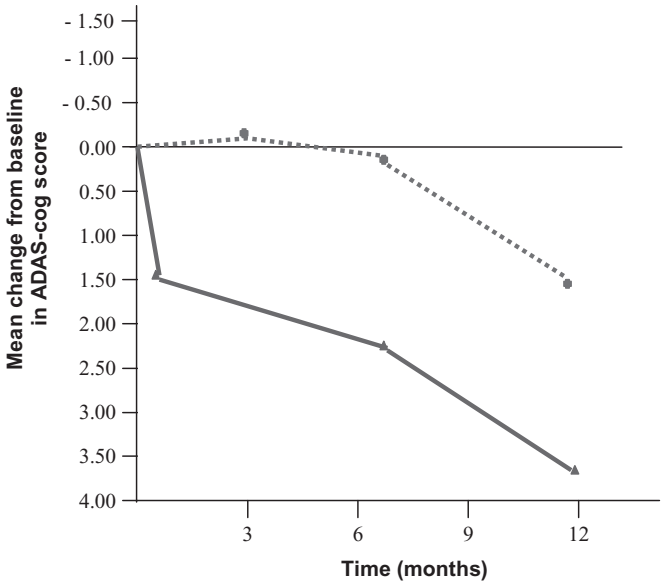


***FIGURE 4*** *Mean change from baseline (intent-to-treat analyses) in ADAS-cog scores of placebo groups of two 1-year clinical trials involving estrogen replacement and Ginkgo biloba (mean change from baseline derived from available data while other values were extrapolated from graphs): ▲ [36], ● [55]. Patients in the estrogen clinical trial had an MMSE range of 14–28 while those in the Ginkgo biloba trial had an MMSE range of 9–26. Other trials not included due to unavailable mean ADAS-cog scores of change from baseline along the course of the trial.*

group in the first few weeks of some clinical trials [21, 22, 26, 27, 29, 30]. Deterioration is inevitable in trials of symptomatic agents despite these initial improvements and fluctuations.

The expected rate of change in the ADAS-cog scores annually in patients with AD is seven to eight points [91]. This varies depending on the severity of the cognitive impairment; less severe patients have a slower rate of decline [90]. This is evident in the Ginkgo biloba trial [55], for example, which included subjects with a lower range of MMSE scores and higher magnitude of decline, compared to those in the estrogen trial, which included subjects with higher MMSE scores (Figure 4).

## 11 ATTRITION AND ADVERSE EFFECTS

Clinical trials suffer from subject dropout and the effects of attrition must be anticipated in trial design. Dropouts are more commonly due to refusal to follow up, lack of efficacy, violation of study protocol, and presence of adverse events.

Diminution of the subject pool is expected as the length of the trial is increased due to dropouts. Attrition is a potential problem because each dropout is equivalent to lost data, which may eventually lead to inadequate sample or, worse, skewing of the outcomes. Selective attrition can produce changes in the subject pool so that it inadvertently becomes unrepresentative of the subjects that originally entered into the study. Results become unreliable in this setting.

**TABLE 12 Gastrointestinal and Selected Other Adverse Events in Placebo Group (Percent of Occurrence) of Acetylcholinesterase Inhibitors in Mild to Moderate AD Clinical Trials**

| | Donepezil | | | | Rivastigmine | Galantamine | | |
|---|---|---|---|---|---|---|---|---|
| Adverse Events | [26] | [27] | [28] | [54] | [33] | [30] | [29] | [31] |
| Anorexia | — | 2 | — | 6 | 2 | 2.4 | 3.1 | 5.6 |
| Nausea | 8 | 4 | | 9 | 10 | 11.2 | 4.5 | 13.1 |
| Vomiting | — | 2 | — | — | 6 | 4 | 1.4 | 7.5 |
| Abdominal pain | — | — | 5.6 | — | 3 | 1.6 | — | 4.2 |
| Diarrhea | 3 | 7 | 6.9 | 17 | 9 | — | 5.9 | 9.9 |
| Constipation | — | — | 6.3 | | | | | |
| Dyspepsia | — | — | — | 6 | | | | |
| Weight loss | — | — | — | 6 | — | — | — | 4.7 |
| Agitation | — | — | — | 13 | — | 0.8 | 9.4 | |
| Confusion | — | — | 6.3 | | | | | |
| Insomnia | — | — | 6.9 | 8 | | | | |
| Somnolence | — | — | — | — | — | 0.8 | | |
| Dizziness | — | 4 | 4.2 | — | 7 | 4 | — | 11.3 |
| Headache | 8 | — | 6.3 | 9 | 8 | | | |
| Fatigue | — | 2 | — | | 3 | | | |
| Syncope | — | — | 2.8 | | | | | |
| Vertigo | — | — | 2.1 | | | | | |
| Anxiety | — | — | 5.6 | | | | | |
| Depression | — | — | 7.6 | | | | | |
| Asthenia | — | — | 3.5 | 7 | | | | |
| Urinary tract infection | 13 | — | 6.9 | 13 | | | | |

*Note*: Included are adverse events with rate of occurrence that differed by at least 5% between the treatment group and the placebo.

Adverse events (AEs) is another cause of subject dropout. Adverse events are defined as newly emergent events/symptoms occurring after the first dose of the study drug or clinically significant worsening of a preexisting condition after the first dose of the study drug. Treatment emergent signs and symptoms (TESSs) is another term that is frequently used to indicate AEs. A serious adverse event (SAE) pertains to any event that is considered life threatening, resulting in hospitalization, permanent disability, or death. Occurrence of a serious AE that is deemed related to drug exposure may be grounds for withdrawal of a drug that is already released in the market. Clinical trials include tolerability and safety studies critical to determining any AEs that may occur with a drug.

Both the active treatment and placebo group experience AEs and all are reported whether these are deemed related or not related to the drug exposure. AEs range from incidental symptoms, part of the inherent deterioration of the disorder, secondary to existent medical disorder or related to drug exposure. Subject variability may account for the assortment of symptoms encountered as AEs and they can be observed among placebo groups as well as active intervention groups (Table 12).

Most AEs in acetylcholinesterase inhibitor trials are causally related to the drug's mechanism of action. Consistent findings are cholinergic manifestations that are usually transient and mild, although some may be moderate in severity. These are dose related and studies which utilized a forced titration rate increased cholinergic

AEs in subjects [22, 26]. In comparison, those that employed a longer dosage titration schedule had lower incidence of these AEs and better drug tolerability [27].

Deaths occur during clinical trials. The probability of its occurrence is increased in a dementia trial due to an expected lower survival rate in this population and the age of the subjects included. Deaths have been rare in most trials and were judged to be unrelated to the drug intake.

## 12 PRESENTING CLINICAL TRIAL RESULTS

The CONSORT statement was developed in (1996) [92] with the purpose of improving the accuracy of conducting and reporting RCTs. This obligates investigators to provide transparency as to the details of the trial from recruitment of subjects to its results. By following the CONSORT criteria, investigators are more likely to conduct ethical trials and produce valid, unbiased results. CONSORT criteria make it easier for readers to evaluate and understand the RCTs and there is an assurance that the results can be relied upon. Inadequate reporting of RCTs has led to biased interpretation of results with overestimation or underestimation of treatment effects. Currently, a 22-item CONSORT checklist and a flow diagram are available for use and these are constantly updated and modified [93].

Table 13 shows how CONSORT criteria ca be applied to trials of AD and MCI.

## 13 DISEASE-MODIFYING TRIALS

Disease-modifying drugs have been the focus of dementia management for some time. The methodological strategies of clinical trials involving this class of therapeutic agent are apparently different from those involving symptomatic agents since the former target the neurodegenerative process of AD. Patients in the early stage of AD should ideally be included since they will receive the most benefit from disease intervention. Evidence of disease course alteration can be established in longer trials, that is, ≥1 year, since the natural progression of AD needs to be observed. This will also ensure that the treatment effect is not due to disease fluctuation that is present in AD. A larger sample population is necessary since attrition rate is expectedly higher in clinical trials with longer duration. Outcome measures other than neuropsychological tests such as biomarkers need to be employed to support the disease-modifying effect of the therapeutic agent. Use of biomarkers as surrogate endpoints can also reduce the required sample size and the duration of clinical trial, saving on time and resources. Pharmacoeconomic measures should also be integrated to justify long term societal benefits from the drug.

## ACKNOWLEDGMENT

The authors wish to acknowledge Lynn Fairbanks, Ph.D., a professor in the Department of Psychiatry and Biobehavioral Sciences, UCLA, for her special contribution in the statistical section of this chapter.

**TABLE 13 Features of Current MCI and ADRCT as These Are Adapted to CONSORT Checklist**

| Paper Section and Topic | Item | Examples of Features of MCI and Dementia Trials |
|---|---|---|
| Title and abstract | 1 | • Intervention stated (duration of trial sometimes included) in randomized, double-blind, placebo-controlled study |
| Background | 2 | Introduction<br>• Basis for trial of intervention in human subjects: e.g., cholinergic hypothesis for AchEI's/glutaminergic excitotoxicity via NMDA receptor for memantine/oxidative damage and free-radical generation for antioxidants/ inflammatory response for prednisone and NSAIDS/estrogen replacement |
| Participants | 3 | Methods<br>• Subjects of both sexes, any race, ≥50 years old (or other age range), no significant medical disorder, adequate hearing and vision for testing, recruited from community/nursing home/other institutions, ambulant/or assisted<br>• Diagnosis of MCI, AD, VaD based on criteria ([7] for MCI; NINDS-AIREN [94] for VaD; NINCDS-ADRDA [12] for AD)<br>• Severity of dementia specified via MMSE range or through other measures (e.g., CDR)<br>• Acceptable concomitant medications<br>• Exclusion of those with contraindication to administration of intervention<br>• Thorough discussion of approval by ethical board, if ethical standards are met, and how consent is taken<br>• Multicenter involving tertiary centers with adequate facilities and personnel trained to diagnose, treat, and handle MCI and dementia cases |
| Interventions | 4 | • Placebo and interventional drug compared via parallel-group trial; treatment arms specified<br>• Fixed dose/dose titration/titration sequence if dose-finding study<br>• Pill distribution (both placebo and interventional drug) including time |
| Objectives | 5 | • Efficacy and safety of intervention<br>• Time to reach an endpoint (cognitive and/or functional deterioration) |
| Outcomes | 6 | • Primary and secondary outcome measures assessing cognitive, global, functional states, and/or quality of life<br>• Intervals of assessment stated, including endpoint |
| Sample size | 7 | • Derived through power analysis (usually power of 80% with specified alpha value)<br>• Based on review of clinical studies of drugs of same class and results of earlier phase II studies<br>• Determined using primary efficacy variable to achieve certain power; estimate based on clinically relevant change in the score of said variable from baseline |
| Randomization (sequence allocation/allocation concealment/ implementation) | 8–10 | • Sequence allocation by simple/blocked or stratified randomization<br>• Via computer-generated randomization list by pharmaceutical company |
| Blinding (masking) | 11 | • Double blind with occasional inclusion of single-blind run-in or single-blind wash-out |

**TABLE 13** ***Continued***

| Paper Section and Topic | Item | Examples of Features of MCI and Dementia Trials |
|---|---|---|
| Statistical methods | 12 | Efficacy:<br>• Drug–placebo difference between baseline and endpoint in specified outcome measures using least-squares (LS) mean<br>• Statistical analysis used specified for both categorical and continuous variables<br>• ITT or PP (LOCF and OC): population defined<br>• Fully evaluable population (EP): population defined<br>• Survival analysis<br>Safety:<br>• Included data defined (e.g., all patients who received at least one dose of study medications and who provided any postbaseline follow-up data)<br>• Incidence of AEs or TESS compared among treatment arms |
| Results | 13 | Participant flow<br>• Diagram of number of subjects in every stage of trial (enrollment, intervention, allocation, follow-up, and analysis) |
| Recruitment | 14 | • Recruitment period specified |
| Baseline data | 15 | • Demographic and clinical characteristics of population allocated to placebo and intervention shown<br>• Comparison made between groups to determine any significant difference in characteristics that may contribute to biased outcome |
| Numbers analyzed | 16 | • Number of subjects who were included in each analysis (dependent on type of analysis, e.g., ITT, LOCF, and OC) |
| Outcomes and estimation | 17 | • Effect size based on clinically meaningful change produced by the intervention<br>• Statistically significant drug–placebo difference based on each of primary and secondary efficacy variables (level of significance specified)<br>• For survival studies: statistically significant difference between placebo and intervention in delaying time to reach endpoint |
| Ancilllary analyses | 18 | • Subgroup analysis based on certain features of subjects where efficacy of drug may be different (e.g., early onset vs. late onset) |
| Adverse events | 19 | • Including serious AEs and any deaths<br>• Determination if occurrence is beyond acceptable limit and if there is association with exposure to intervention |
| Interpretation | 20 | Discussion<br>• Interpretation and explanation of positive and/or negative results |
| Generalizability | 21 | • Limitations of study that will limit its application to targeted population in community setting (e.g., differential recruitment in terms of race, highly specialized centers) |
| Overall evidence | 22 | • Evidence-based conclusion regarding efficacy and safety of intervention in subset of patients studied<br>• Recommendations in future studies |

ITT: Intent-to-treat analysis; PP: per protocol.

## REFERENCES


1. Evans, D. A. (1990), Estimated prevalence of Alzheimer's disease in the United States, *Milbank Q*, 68, 267–289.
2. National Institute of Health, National Institute on Aging (1999), *Progress Report on Alzheimer's Disease, 1999* (NIH Pub. No. 99-4664), U.S. Department of Health and Human Services, Bethesda, MD.
3. Evans, D. A., Funkenstein, H. H., Albert, M. S., et al. (1989), Prevalence of Alzheimer's disease in a community population of older persons. Higher than previously reported, *JAMA*, 26, 2551–2556.
4. Institute on Aging (1996), *Progress Report on Alzheimer's Disease, 1996* (NIH Pub. No. 96-4137), National Institute on Aging, Bethesda, MD.
5. Ernst, R. L., and Hay, J. W. (1994), The US economic and social costs of AD revisited, *Am. J. Public Health*, 84, 1261–1264.
6. Morris, J. C., Storandt, M., Miller, P., et al. (2001), Mild cognitive impairment represents early-stage Alzheimer's disease, *Arch. Neurol.*, 58, 397–405.
7. Petersen, R. C., Doody, R., Kurz, A., et al. (2001), Current concepts in mild cognitive impairment, *Arch. Neurol.*, 58, 1985–1992.
8. Petersen, R. C., Thomas, R. G., Grundman, M., et al. (2005), Vitamin E and donepezil for the treatment of mild cognitive impairment, *N. Engl. J. Med.*, 352, 2379–2388.
9. Salloway, S., Ferris, S., Kluger, A., et al. (2004), Efficacy of donepezil in mild cognitive impairment: A randomized placebo-controlled trial, *Neurology*, 63, 651–657.
10. American Psychiatric Association (1994), *Diagnostic and Statistical Manual of Mental Disorders*, 4th ed. (DSM-IV), American Psychiatric Association, Washington, DC, pp. 143–147.
11. World Health Organization. (1993), *The ICD-10 Classification of Mental and Behavioral Disorders. Diagnostic Criteria for Research*, World Health Organization, Geneva, Switzerland.
12. McKhann, G., Drachman, D., Folstein, M., et al. (1984), Clinical diagnosis of Alzheimer's disease: Report of the NINCDS-ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease, *Neurology*, 34, 939–944.
13. Holmes, C., Clairns, N., Lantos, P., et al. (1999), Validity of current clinical criteria for Alzheimer's disease, vascular dementia and dementia with Lewy bodies, *Bri. J. Psych.*, 174, 45–50.
14. Blacker, D., Albert, M., Bassett, S., et al. (1994), Reliability and validity of NINCDS-ADRDA criteria for Alzheimer's disease. The National Institute of Mental Health Genetics Initiative, *Arch. Neurol.*, 51, 1198–1204.
15. Lopez, O. L., Litvan, I., Catt, K. E., et al. (1999), Accuracy of four clinical diagnostic criteria for the diagnosis of neurodegenerative dementias, *Neurology*, 53, 1292–1299.
16. Tierney, M., Fisher, R., Anthony, J., et al. (1988), The NINCDS-ADRDA Work Group criteria for the clinical diagnosis of probable Alzheimer's disease: A clinicopathologic study of 57 cases, *Neurology*, 38, 359–364.
17. Reisberg, B., Ferris, S., de Leon, M., et al. (1982), The Global Deterioration Scale for assessment of primary degenerative dementia, *Am. J. Psych.*, 138, 1136–1139.
18. Hughes, C., Berg, L., Denziger, W., et al. (1982), A new clinical scale for the staging of dementia, *Bri. J. Psych.*, 140, 566–572.

19. Folstein, M., Folstein, S., and McHugh, P. (1975), Mini-Mental State: A practical method for grading the cognitive state of patients for the clinician, *J. Psychiatric Res.*, 12, 189–198.
20. O'Connor, D. W., Pollitt, P., Hyde, J. B., et al. (1989), The reliability and validity of the Mini-Mental state in a British community survey, *J. Psychiatric Res.*, 23, 87–96.
21. Cummings, J., Cyrus, P., Bieber, F., et al. (1998), Metrifonate treatment of the cognitive deficits of Alzheimer's disease, *Neurology*, 50, 1214–1221.
22. Morris, J., Cyrus, P., Orazem, J., et al. (1998), Metrifonate benefits cognitive, behavioral, and global function in patients with Alzheimer's disease, *Neurology*, 50, 1222–1230.
23. Davis, K., Thal, L., Gamzu, E., et al. (1992), A double-blind, placebo-controlled multicenter study of tacrine for Alzheimer's disease, *N. Engl. J. Med.*, 327, 1253–1259.
24. Farlow, M., Gracon, S., Hershey, L., et al. (1992), A controlled trial of tacrine in Alzheimer's disease, *JAMA*, 268, 2523–2529.
25. Knapp, M., Knopman, D., Solomon, P., et al. (1994), A 30-week randomized controlled trial of high-dose tacrine in patients with Alzheimer's disease, *JAMA*, 271, 985–991.
26. Rogers, S., Doody, R., Mohs, R., et al. (1998), Donepezil improves cognition and global function in Alzheimer's disease, *Arch. Intern. Med.*, 158, 1021–1031.
27. Rogers, S., Farlow, M., Doody, R., et al. (1998), A 24-week, double-blind, placebo-controlled trial of donepezil in patients with Alzheimer's disease, *Neurology*, 50, 136–145.
28. Winblad, B., Engedal, K., Soininen, H., et al. (2001), A 1-year, randomized, placebo-controlled study of donepezil in patients with mild to moderate AD, *Neurology*, 57, 489–495.
29. Tariot, P., Solomon, P., Morris, J., et al. (2000), A 5-month, randomized, placebo-controlled trial of galantamine in AD, *Neurology*, 54, 2269–2276.
30. Rockwood, K., Mintzer, J., Truyen, L., et al. (2001), Effects of a flexible galantamine dose in Alzheimer's disease: A randomized, controlled trial, *J. Neurol. Neurosurg. Psych.*, 71, 589–595.
31. Raskind, M., Peskind, E., Wessel, T., et al. (2000), Galantamine in AD: A 6-month randomized, placebo-controlled trial with a 6-month extension, *Neurology*, 54, 2261–2268.
32. Corey-Bloom, J., Anand, R., Veach, J., et al. (1998), A randomized trial evaluating the efficacy and safety of ENA 73 (rivastigmine tartrate), a new acetylcholinesterase inhibitor, in patients with mild to moderately severe Alzheimer's disease, *Int. J. Ger. Psychopharm.*, 1, 55–65.
33. Rosler, M., Anand, R., Cicin-Sain, A., et al. (1999), Efficacy and safety of rivastigmine in patients with Alzheimer's disease: International randomized controlled trial, *BMJ*, 318, 633–638.
34. Scharf, S., Mander, A., Ugoni, A., et al. (1999), A double-blind, placebo-controlled trial of diclofenac/misoprostol in Alzheimer's disease, *Neurology*, 53, 197–201.
35. Aisen, P., Davis, K., Berg, J., et al. (2000), A randomized controlled trial of prednisone in Alzheimer's disease, *Neurology*, 54, 588–593.
36. Mulnard, R., Cotman, C., Kawas, C., et al. (2000), Estrogen replacement therapy for treatment of mild to moderate alzheimer disease, *JAMA*, 283, 1007–1015.
37. Aisen, P., Schafer, K., Grundman, M., et al. (2003), Effects of rofecoxib or naproxen vs. placebo on Alzheimer disease progression, *JAMA*, 289, 2819–2826.
38. Reines, S., Block, G., Morris, J., et al. (2004), Rofecoxib: No effect on Alzheimer's disease in a 1-year, randomized, blinded, controlled study, *Neurology*, 62, 66–71.

39. Thal, L., Carta, A., Clarke, W., et al. (1996), A 1-year multicenter placebo-controlled study of acetyl-L-carnitine in patients with Alzheimer's disease, *Neurology*, 47, 705–711.
40. Feldman, H., Gauthier, S., Hecker, J., et al. (2001), A 24-week, randomized, double-blind study of donepezil in moderate to severe Alzheimer's disease, *Neurology*, 57, 613–620.
41. Reisberg, B., Doody, R., Stoffler, A., et al. (2003), Memantine in moderate-to-severe Alzheimer's disease, *N. Engl. J. Med.*, 348, 1333–1341.
42. Doody, R., Stevens, J., Beck, C., et al. (2001), Practice parameter: Management of dementia (an evidence-based review): Report of the Quality Standards Subcommittee of the American Academy of Neurology, *Neurology*, 56, 1154–1166.
43. Alzheimer Association (2004), Research consent for cognitively impaired adults: Recommendations for institutional review boards and investigators, *Alzheimer Dis. Assoc. Disord.*, 18, 171–175.
44. Koepsell, T., and Weiss, N. (2003), *Epidemiologic Methods*, Oxford University Press, New York, pp. 64–67, 261, 327–329.
45. Schneider, L., Tariot, P., Lyketsos, C., et al. (2001), National Institute of Mental Health Clinical Antipsychotic Trial of Inervention Effectiveness (CATIE): Alzheimer Disease Trial Methodology, *Am. J. Ger. Psych.*, 9, 346–390.
46. Cummings, J. (2003), Use of acetylcholinesterase inhibitors in clinical practice: Evidence-based recommendations, *Am. J. Ger. Psych.*, 11, 131–145.
47. Schoenmaker, N., and Van Gool, W. (2004), The age gap between patients in clinical studies and in the general population: A pitfall for dementia research, *Lancet Neurol.*, 3, 627–630.
48. Leber, P. (1990), *Guidelines for the Clinical Evaluation of Antidementia Drugs*.
49. Ferris, S., Lucca, U., Mohs, R., et al. (1997), Objective psychometric tests in clinical trials of dementia drugs, *Alzheimer Dis. Assoc. Disord.*, 11, 34–38.
50. Rosen, W., Mohs, R., Davis, K., et al. (1984), A new rating scale for Alzheimer's disease, *Am. J. Psych.*, 141, 1356–1364.
51. Guy, W., ed. (1976), Clinical Global Impression (CGI). *ECDEU Assessment Manual for Psychopharmacology*, U.S. Department of Health and Human Services, Public Health Service, Alcohol Drug Abuse and Mental Health Administration. NIMH Psychopharmacology Research Branch, Rockville, MD, pp. 218–222.
52. Knopman, D., Knapp, M., Bracon, S., et al. (1994), The Clinician Interview-Based Impression (CIBI): A clinician's global change rating scale in Alzheimer's disease, *Neurology*, 44, 2315–2321.
53. Schneider, L., Olin, J., Doody, R., et al. (1997), Validity and Reliability of the Alzheimer's Disease Cooperative Study—Clinical Global Impression of Change, *Alzheimer Dis. Assoc. Disord.*, 11 (Suppl 2), S22–S32.
54. Mohs, R., Doody, R., Morris, J., et al. (2001), A 1-year, placebo-controlled preservation of function survival study of donepezil in AD patients, *Neurology*, 57, 481–488.
55. Le Bars, P., Katz, M., Berman, N., et al. (1997), A placebo-controlled, double-blind, randomized trial of an extract of ginkgo biloba for dementia, *JAMA*, 278, 1327–1332.
56. Burke, W., Miller, O., Rubin, E., et al. (1988), Reliability of the Washington University Clinical Dementia Rating, *Arch. Neurol.*, 45, 31–32.
57. Reisberg, B., Scaln, S., Franssen, E., et al. (1994), Dementia Staging in chronic care populations, *Alzheimer Dis. Assoc. Disord.*, 8 (Suppl. 1), S188–S205.
58. Sano, M., Ernesto, C., Thomas, R., et al. (1997), A Controlled trial of selegiline, alpha-tocopherol, or both as treatment for Alzheimer's disease, *N. Engl. J. Med.*, 336, 1216–1222.

59. Gelinas, I., Gauthier, L., McIntyre, M., et al. (1999), Development of a functional measure for persons with Alzheimer's disease: The disability assessment for dementia, *Am. J. Occup. Ther.*, 53, 471–481.
60. Gauthier, S., Bodick, N., Erzigkeit, E., et al. (1997), Activities of daily living as an outcome measure in clinical trials of dementia drugs, *Alzheimer Dis. Assoc. Disord.*, 11 (Suppl 3), 6–7.
61. Lawton, M., and Brody, E. (1969), Assessment of older people: Self-maintaining and instrumental activities of daily living, *Gerontologist*, 9, 179–186.
62. DeJong, R., Osterlund, O., and Roy, G. (1989), Measurement of quality-of-life changes in patients with Alzheimer's disease, *Clin. Ther.*, 11, 545–554.
63. Galasko, D., Bennett, D., Sano, M., et al. (1997), An inventory to assess activities of daily living for clinical trials in Alzheimer's disease, *Alzheimer Dis. Assoc. Disord.*, 11 (Suppl 2), S33–39.
64. Patterson, M., and Bolger, J. (1994), Assessment of behavioral symptoms in Alzheimer's disease, *Alzheimer Dis. Assoc. Disord.*, 8 (Suppl 3), 4–20.
65. Cummings, J., Mega, M., Gray, K., et al. (1994), The Neuropsyciatric Inventory: An efficient tool for comprehensively assessing psychopathology in dementia, *Neurology*, 44, 2308–2314.
66. Reisberg, B., Auer, S. R., Monteiro, I. M., et al. (1996), Behavioral pathology in Alzheimer's disease (BEHAVE-AD) rating scale, *Int. Psychogeriatr.*, 8 (Suppl 3), 301–308.
67. Tariot, P., Mack, J., Patterson, M., et al.; the CERAD Behavioral Pathology Committee (1995), The CERAD Behavior Rating Scale for Dementia (BRSD), *Am. J. Psych.*, 152, 1349–1357.
68. Cohen-Mansfield, J. (1985), Agitated behaviors in the elderly ill. Preliminary results in the cognitively deteriorated, *J. Am. Ger. Soc.*, 34, 722–727.
69. Alexopoulos, G., Abrams, R., Young, R., et al. (1988), Cornell Scale for Depression in Dementia, *Biol. Psych.*, 23, 271–284.
70. Overall, J., and Gorham, D. (1962), The Brief Psychiatric Rating Scale, *Psychol. Rep.*, 10, 799–812.
71. Overall, J., and Gorhal, D. (1988), Introduction: The Brief Psychiatric Rating Scale (BPRS): Recent developments in ascertainment and scaling, *Psychopharml. Bull.*, 24, 97–99.
72. Kay, S., Opler, L., Lindenmayer, J., et al. (1989), The Positive and Negative Syndrome Scale: A rationale and standardization, *Br. J. Psych.*, 155 (Suppl 7), 59–65.
73. Weiner, M., Koss, E., Wild, K., et al. (1996), Measures of psychiatric symptoms in Alzheimer patients: A review, *Alzheimer Dis. Assoc. Disord.*, 10, 20–30.
74. De Deyn, P., Rabheru, K., Rasmussen, A., et al. (1999), A randomized trial of risperidone, placebo, and haloperidol for behavioral symptoms of dementia, *Neurology*, 53, 946–955.
75. Katz, I., Jeste, D., Mintzer, J., et al. (1999), Comparison of risperidone and placebo for psychosis and behavioral disturbances associated with dementia: A randomized, double-blind trial, *J. Clin. Psych.*, 60, 107–115.
76. Brodaty, H., Ames, D., Snowdon, J., et al. (2003), A randomized placebo-controlled trial of risperidone for the treatment of aggression, agitation and psychosis of dementia, *J. Clin. Psych.*, 64, 134–143.
77. Lyketsos, C., Sheppard, J., Steele, C., et al. (2000), Randomized, placebo-controlled, double-blind clinical trial of sertraline in the treatment of depression complicating Alzheimer's disease: Initial results from the depression in Alzheimer's disease study, *Am. J. Psych.*, 157, 1686–1689.

78. Tariot, P., Erb, R., Podgorski, C., et al. (1988), Efficacy and tolerability of carbamazepine for agitation and aggression in dementia, *Am. J. Psych.*, 155, 54–61.
79. Porteinsson, A., Tariot, P., Erb, R., et al. (2001), Placebo-controlled study of divalproex sodium for agitation in dementia, *Am. J. Ger. Psych.*, 9, 58–66.
80. Zhong, K., Tariot, P., Minkwitz, M. C., et al. (2004), Quetiapine for the treatment of agitation in the elderly institutionalized patients with dementia: A randomized, double blind trial. International Conference on Alzheimer's Disease and Related Disorders meeting, poster # P2–442.
81. Davis, K., Marin, D., Kane, R., et al. (1997), The Caregiver Activity Survey (CAS): Development and validation of a new measure for caregivers of persons with Alzheimer's disease, *Int. J. Ger. Psych.*, 12, 978–988.
82. Wimo, A., Wetterholm, A., Mastey, Y., et al. (1998), Evaluation of the healthcare resource utilization and caregiver time in anti-dementia drug trials, in Wimo, A., Jonsson, B., Karlson, G., and Wilblad, B., Eds., *Health Economics of Dementia*, Wiley, Chichester, England, pp. 465–499.
83. Whitehouse, P. (1997), Pharmacoeconomics of dementia, *Alzheimer Dis. Assoc. Disord.*, 11 (Suppl 5), S22–33.
84. AD2000 Collaborative Group (2004), Long term donepezil treatment in 565 patients with Alzheimer's disease (AD2000): Randomized double-blind trial, *Lancet*, 353, 2105–2115.
85. Winblad, B., Wimo, A., and Almkvist, O. (2000), Outcome measures in Alzheimer's disease: Do they go far enough? *Dem. Ger. Cogn. Dis.*, 11 (Suppl 1), 3–10.
86. Leber, P. (1997), Observations and suggestions on antidementia drug development, *Alzheimer Dis. Assoc. Disord.*, 10 (S1), 31–35.
87. Sano, M., Ernesto, C., Klauber, M., et al. (1996), Rationale and design of a multicenter study of selegiline and α-tocopherol in the treatment of Alzheimer disease using novel clinical outcomes, *Alzheimer Dis. Assoc. Disord.*, 10, 132–140.
88. A Randomized Double-Blind, Placebo-Controlled Trial to Evaluate the Efficacy and Safety of GalantamIne in Subjects with Mild Cognitive Impairment (MCI) Clinically at Risk for Development of Clinically Probably Alzheimer's Disease. Gal–INT-18 and Gal-INT-11. http://www.alz.org/news/05q1/012405.asp; accessed November 17, (2005).
89. Everitt, B., and Pickler, A. (2004), *Statistical Aspects of the Design and Analysis of Clinical Trials*, Imperial College Press, London, pp. 45–50, 203–204.
90. Doraiswamy, P., Kaiser, L., Bieber, F., et al. (2001), The Alzheimer's Disease Assessment Scale: Evaluation of psychometric properties and patterns of cogntivie decline in multicenter clinical trials of mild-to-moderate Alzheimer's disease, *Alzheimer Dis. Assoc. Disord.*, 15, 174–183.
91. Morris, J., Edland, S., Clark, C., et al. (1993), The Consortium to Establish a Registry for Alzheimer's Disease (CERAD). Part, IV. Rates of cognitive change in the longitudinal assessment of probable Alzheimer's disease, *Neurology*, 43, 2457–2465.
92. Begg, C., Cho, M., Eastwood, S., et al. (1996), Improving the quality of reporting of randomized controlled trials: The CONSORT statement, *JAMA*, 276, 637–639.
93. Moher, D., Schulz, K., Altman, D., et al. (2001), The CONSORT Statement: Revised recommendations for improving the quality of reports of parallel-group randomized trials, *JAMA*, 285, 1987–1991.
94. Roman, G. C., Tatemichi, T. K., Erkinjuntti, T., et al. (1993), Vascular dementia: diagnostic criteria for research studies. Report of the NINDS-AIREN International Work Group, *Neurology*, 43, 250–260.

# *Clinical Trials in Urology*

**Geoffrey R. Wignall, Carol Wernecke, Linda Nott, and Hassan Razvi**
*Schulich School of Medicine and Dentistry, University of Western Ontario, London, Ontario, Canada*

**Contents**

1 Introduction
2 Basics
  2.1 Clinical Trial Objectives
  2.2 Study Population
  2.3 Ethical Considerations
  2.4 Review Boards
3 Specific Issues with Surgical Device Trials
  3.1 Regulatory Issues
  3.2 "First-in-Man" Studies
4 Summary
  References

## 1 INTRODUCTION

The wide scope of current urological practice has led to a proliferation of research and clinical trials in this specialty. As a surgical subspecialty devoted to the medical and surgical care of patients with genitourinary disorders, both pharmaceutical and device trials have equally important roles in the advancement of the future care of the urologic patient. Pharmaceutical trials are ongoing for conditions such as erectile dysfunction, urinary incontinence, and oncology to name a few. Significant advancements in surgical instrumentation such as robotics and prosthetic devices for the treatment of genitourinary diseases are also being witnessed and are the

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

subject of clinical investigation. The objective of this chapter is to provide an overview of the objectives, design, and implementation of clinical trials specifically relevant to those conducting urological research. While many of the domains covered in the chapter are subjects commonly faced by researchers in other specialties, we will attempt to highlight issues of particular importance to urological investigators.

## 2 BASICS

In this section we provide an overview of the issues related to the development of clinical trial objectives and the study populations of importance in urological research.

### 2.1 Clinical Trial Objectives

The first and most important step in designing any clinical trial is defining the study objective [1]. Careful attention to the formulation of the primary and secondary study objectives will allow investigators to design a quality clinical trial that is adequately powered to answer the study question and avoid many potential pitfalls.

This crucial first step attempts to formulate the primary question or that question that is most significant to be answered by the trial. This may seem self-evident, as this objective is likely the reason the trial was initiated in the first place. However, more often than not the initial concept is too broad or poorly defined to be useful as a primary question. Investigators must ask themselves whether this primary objective can be realistically addressed with a clinical trial. The sample size calculation is based on the primary question. In some instances, that calculation may reveal that more participants would be needed than would be expected to be available in a reasonable time frame. Reevaluation of the primary outcome measurement may provide a more realistic number of participants that would be needed for the trial.

When designing a study to evaluate a new treatment or drug to treat a disease such as cancer, the urological researcher should ensure the most clinically relevant outcomes will be assessed. Time to disease progression, disease-specific survival, and overall survival are just a few of the commonly determined measurements. The effect of the study treatment may or may not be evident for several years, depending on the disease process, and the effect of the study treatment. Therefore, the investigators will need to define how these study endpoints will be assessed over time.

In formulating the primary question, investigators may either seek to prove a hypothesis [2] or to show a lack of a difference in outcomes between participants and controls. An example of the former would be the situation where investigators hypothesize that patients receiving a study drug for the treatment of prostate cancer may have a decrease in cancer-specific mortality. At other times it is beneficial to show that there is no difference between subjects and controls such as when an intervention is deemed to be unnecessary. An example of this would be the argument against the routine placement of a ureteral stent following uncomplicated ureteroscopy where investigators demonstrated that there was no advantage in the placement of a stent [3].

While formation of a clear primary objective is of the utmost importance, secondary objectives must also be defined. These may be evident at the start of the study or may become clear through collection of data designed to address the primary objective. Using the above example of a drug for the treatment of cancer, if the primary objective is to demonstrate an improvement in cancer-specific mortality, a secondary objective may be to see which subgroups in particular benefit from this effect. Whenever possible, these secondary objectives should be outlined at the commencement of the trial. Additionally, secondary objectives may be distinctly separate from the primary question such as determining whether there are any adverse effects from the drug or a difference between subjects and controls with respect to quality of life issues. The primary objective is used for power and sample size calculation; therefore, it is not uncommon for secondary objectives to be inconclusive and lack power due to small sample size.

Urologists have several tools to evaluate outcomes in clinical trials. While questionnaires are able to evaluate subjective outcomes, the study design should also include objective measurements. Objective measurements must be relative to the disease under study and be considered to be part of the standard of care or evaluation of that same patient population.

The International Prostate Symptom Score includes symptom, bother, and quality of life scores that are useful in accessing primary and secondary outcomes in several types of urology trials [4]. The International Index of Erectile Function (IIEF) can be used to evaluate the primary outcome in erectile dysfunction trials, and secondary outcomes of other trials affecting the male population where the evaluation of sexual function is relevant [5]. Quality of life measurements are commonly utilized in urologic trials as secondary objectives. The SF-36 Health is considered the most comprehensive instrument to measure general health-related quality of life (HRQOL) [6]. Various validated instruments are also available to assess a number of specific urologic conditions. The UCLA Prostate Cancer Index was devised to assess health-related quality of life in patients with prostate cancer [7]. The Ureteral Stent Questionnaire (USSQ) uses both the visual analog scale (VAS) and quality of life questions to evaluate the patient's stent-related symptoms [8]. Specific urological questions, either as primary or secondary outcomes, can be evaluated in the trial by using the VAS to assess the effect of the study treatment on participants' perceived results, ranging from pain to treatment effectiveness or reduction of urinary symptoms [9].

## 2.2 Study Population

Careful selection of the study population will improve the chances of successfully answering the study question in a meaningful and convincing fashion. Well-defined inclusion and exclusion criteria must be clearly defined in the study protocol. In general, a simple approach to defining the study population begins with including the entire population at large and methodically narrowing the definition of the eligible subject. To illustrate this process we consider the example of a large prospective urological study, The Prostate Cancer Prevention Trial (PCPT) [10, 11]. The objective of PCPT was to assess the effect of the drug finasteride for the prevention of prostate cancer. The first step toward selecting the study population in this trial was to identify those individuals with or at risk for prostate cancer. In the case of

prostate cancer, young men are much less likely to manifest the disease; therefore, the designers of the PCPT trial opted to include only those men 55 years of age or older. As this study looked at the development of prostate cancer in men without the disease, it was important to include only men with a normal digital rectal exam (DRE). One must also be aware of potential confounding conditions that may detract from the ability to assess for the condition being studied. Benign prostatic hyperplasia (BPH) may result in a benign rise in prostate specific antigen that may be perceived as potential prostate cancer. To minimize the impact of BPH, men with an American Urological Association symptom score of 20 or greater were excluded from the study.

Randomized clinical trials usually have at least one control group as well as one or more treatment groups. While this tends to increase the costs of the trial due to the greater number of subjects required, it is essential to statistically answer the primary study question.

Control subjects may be assigned as either negative or positive controls. In the setting of a pharmaceutical trial, a patient receiving a placebo (i.e., sugar pill) would be a negative or passive control, meaning that they are not getting active treatment. Still, patients in placebo arms of studies often report improvement or deterioration in their condition, demonstrating the so-called placebo effect. This is a well-described phenomenon in urology, especially noteworthy in benign prostatic hyperplasia clinical trials. Improvements in symptom severity in particular may range from 5 to 15% for placebo or sham treated cohorts for both medical and minimally invasive device treatment trials of BPH patients [12].

A negative control may be used either alone (if no gold standard treatment exists) or in conjunction with a positive control arm. A positive or active control may be used when there is a current treatment that is considered the standard of care for a particular disease or condition. The control arm consists of patients receiving the standard treatment while subjects in the experiment arm receive the treatment under investigation. By actively treating both treatment and control groups, investigators hope to demonstrate superiority of the experimental treatment over the established standard. It has been suggested that active controls may protect patients and should be routinely utilized [13, 14].

Subject-to-subject variability is controlled by choosing a cross-over design. In a cross-over medical device trial, subjects can be used as their own control. Cross-over trials can be considered when the study includes two or more interventions. The interventions may be two separate treatment devices or one may be a placebo treatment. Each subject in the clinical study is assigned a randomized treatment group, allowing a period between interventions for a "wash out" of any carry-over from the previous intervention. Patients treated as part of the cross-over design are then followed from the baseline of the study's schedule as outlined in the protocol. In studies with a placebo arm, patients who were randomized to the placebo arm are offered open-label study treatment at a predetermined time. Often this is defined as when the study treatment is unblinded and revealed to all of the participants [15].

### 2.3 Ethical Considerations

Research has been ongoing throughout human history; however, it has not always been carried out in a fashion that would be deemed ethical by today's standards.

Modern investigators must be acutely aware of the ethical aspects of clinical trials and should strive to exceed the expectations of their research review boards. Given the nature of the urological patient population, significant ethical issues are not infrequently encountered in the course of urological clinical trials.

The formulation of a relevant and important research question helps to lay the foundation for ethically sound research [16]. Researchers must thoroughly consider the relevance of their research proposal, ensuring that it is based on solid scientific principles and that the results of the study are likely to improve patient care or enhance knowledge in the field [17]. Patients enrolled in a clinical trial should be confident that their participation should ultimately result in a valuable scientific contribution [16].

In order to best protect the rights of individual patients, investigators should adhere to the principles of autonomy, beneficence, and justice [18]. Autonomy implies that the individual should make important decisions intentionally and free of external influence. Beneficence centers on the responsibility of the investigator to maximize the positive outcome of the trial for each patient. This is a fundamental principle of the Hippocratic oath along with the responsibility to "do no harm" and thereby minimize the potential negative outcomes of the trial [19]. Justice dictates that the design and implementation of a clinical trial be fair to participants. The selection of subjects should be free of bias, and drawn from a relevant patient population.

The Council for International Organizations of Medical Sciences (CIOMS) and the World Medical Association (WMA) have both published guidelines for the ethical management of biomedical research. The WMA Declaration of Helsinki (Ethical Principles for Medical Research Involving Human Subjects) was first adopted in 1964 and has undergone several revisions with the most recent in 2000 [20]. The CIOMS International Ethical Guidelines for Biomedical Research Involving Human Subjects consists of 21 guidelines relating to ethical justification scientific validity of research, informed consent, and equity regarding burdens and benefits among other issues [21]. Each set of guidelines provides an excellent framework for the ethical conduct of biomedical research and are of benefit when preparing a research proposal for institutional review.

No clinical trial may be deemed ethical without a properly formulated and freely given informed consent. Not only must this consent be legal according to the laws of the nation where the study is to be performed but should also conform to the ethical guidelines laid out by the WMA and/or CIOMS. Additionally, in the event that pertinent information becomes available during the course of a trial, it is the investigators' responsibility to advise the participants of any information that pertains to the informed consent [1]. First, the potential study subjects should be provided with an overview of the proposed research to decide whether they would be interested in inclusion into the trial. This includes informing subjects of their responsibilities during the trial (i.e., taking of medications, follow-up appointments), as well as any procedures that will be performed. There should be an honest and open dialog between the individual obtaining the consent and the prospective subject with no attempt to mislead or deceive. The consent must be given voluntarily with participants entering the trial of their own free will with no fear of reprisal should they decline (i.e., withholding of care). Subjects must be told why they are appropriate candidates for the trial (inclusion criteria) and what would make them ineligible (exclusion criteria).

Participants must be informed of all potential risks and benefits of participation in the trial at the time of consent. Risks may include but are not limited to potential physical or mental injury as a result of participation in the study. Additionally, the investigators should provide the patient with an explanation of what will be done to minimize these risks. The benefits of inclusion in the trial may include an improvement to the patient's symptoms or condition, although the patient should be made aware that this is not guaranteed. It is essential to explain any alternate treatments that are available to the subject, including the standard drugs or procedures that would be used in his or her situation and make the patient aware if any of these will be withheld due to inclusion in the study. An example of this would be a trial studying a new form of ureteroscopic lithotripsy. The patient should be made aware of the potential risks of the procedure (i.e., ureteral injury, failure to fragment the stone) as well as possible benefits (i.e., improved stone fragmentation). Additionally, participants must be made aware of alternative procedures such as laser lithotripsy that would be withheld in order to study the new device. The described benefits should also include the potential benefits to society as a whole such as improved stone-free rates using the new technology.

Confidentiality is an important ethical consideration when carrying out a clinical trial and should be maintained from consent through to the study conclusion. Personalized information including subject name, date of birth, and ethnicity along with any other personal identifiers must remain confidential. Study participants should be notified at the time of consent if any personal identifiers will be disclosed during the course of the trial and what measures will be taken to maintain confidentiality. In the event that the trial requires the use of video or photographic data, patients should be aware that they have the right to access such files should they wish to do so.

Participants should be provided with written material summarizing the above information in lay language that is clear and understandable. This information should include appropriate contact information for the reference during the course of the study. The informed consent document must be signed and dated by both the participant and the individual obtaining consent, although verbal consent may be allowed at the discretion of the review board in some instances.

### 2.4 Review Boards

While the actual requirements to conduct research vary based on the country in which the trial is to be conducted, most nations have adopted a model that incorporates an institutional review board (IRB) or a research ethics board (REB). The general mandate of the IRB/REB is to objectively review research proposals involving the use of human subjects and to ensure that trials are ethically sound according to the previously described principles. The board protects the rights of individual participants in research trials through external review to assess the relevance and ethics of proposals and to make certain that confidentiality and safety are maintained by investigators. In general, ethics boards are comprised of multiple individuals of both genders, at least one of whom should be from a nonscientific discipline. Additionally, one or more members may be appointed from outside of the institution at which the research is to be conducted [16].

Researchers embarking on a trial should be aware of the principles of ethical research and identify any potential pitfalls that may delay approval of the ethics application. The primary research question must be clear and relevant, and all secondary objectives should be stated. Relevant statistics such as the sample size calculation must be performed ahead of time to justify the required number of participants. Any potential risks to participants should be addressed in the application as well as measures through which the investigators will minimize these risks and deal with adverse events should they arise. While guidelines are, for the most part, consistent between institutions, researchers should make themselves familiar with the requirement at their institution. This preparation will help to avoid any unnecessary delays.

## 3 SPECIFIC ISSUES WITH SURGICAL DEVICE TRIALS

### 3.1 Regulatory Issues

Urologists employ a wide array of surgical devices ranging in complexity from simple guidewires to robotic operating systems. Each of these devices must pass through regulatory safeguards prior to introduction into clinical use. Issues relating to drug trials are covered elsewhere in this text; however, we feel it necessary to briefly touch on matters relating to device trials and how this may differ from pharmaceutical trials.

The U.S. Food and Drug Administration (FDA) Center for Devices and Radiological Health oversees the development and marketing of new devices in the United States. If the investigational device study presents a significant risk to the subjects, an investigational device exemption (IDE) application must be completed [22]. The IDE must be approved by the FDA, and the individual site(s) institutional review board prior to study initiation. The process of IDE approval is often lengthy, and once approved the clinical study must be conducted according to the IDE regulations (CFR 812). If an investigator or a sponsor proposes that the device under study does not pose a significant risk (NSR) to the participants, and the ethics board agrees and approves the study, no FDA/IDE application is needed.

Medical devices are classified by risk, with class I devices being lowest risk, and class III devices being the highest risk as defined by FDA standards [22]. FDA class I devices (i.e., handheld surgical instruments) present minimal potential harm to the user and require the least amount of regulatory control. Class II devices include instruments such as uroflowmeters, monitors, pump systems, and X-ray devices. These devices require additional special controls including special labeling requirements, mandatory performance standards, and postmarket surveillance. Class III devices (class III/IV in Canada) have the strictest regulatory controls and include devices that support or sustain human life or lack sufficient safety information to assure safety and efficacy (e.g., ureteral stents, penile prostheses, microwave thermotherapy units).

The process in Canada is similar. Medical devices are regulated by Health Canada's Therapeutic Products Directorate under the Food and Drugs Act [23]. The goal of this regulation is to ensure that medical devices available in Canada are safe and effective and of sufficient quality for general distribution. Once Health Canada

approval has been secured, the local institutional review boards are then able to proceed with their review of the research protocol. Health Canada is entitled to ask the sponsor for any information it deems necessary to make a sufficiently informed decision on the safety of the product (additional information request). The cycle of Health Canada issuing the request, the time for the sponsor to compose its response, and a new review time of 30 days from Health Canada's receipt of the response may further prolong the review cycle. Health Canada differs slightly by stratifying devices from class I to class IV but follows the same general principles as the FDA.

If an investigator or a sponsor proposes a study using a device previously approved for the same application, a Health Canada Clinical Trial Application (CTA) is not required. However, if the study proposes to use an approved device for another indication or in association with a drug not previously approved for the same indication, a CTA must be submitted to Health Canada [24].

For specific details on the regulatory process in countries outside North America, interested researchers should contact the relevant national health department's regulatory branch.

### 3.2 "First-in-Man" Studies

First-in-man trials present a particular concern for drug studies, especially since March of 2006 when six healthy volunteers were administered the study drug TGN1412 and subsequently developed multiorgan failure requiring intensive care [25]. Fortunately, there were no deaths involved with this trial. However, this brought to light particular issues regarding first-in-man trials. In 2007, the Royal Statistical Society published a study focusing on statistical consideration for first-in-man studies [26]. They recognized that first-in-man studies are vital in the development of new drugs and devices and such trials must have a high level of safety. Recommendations were made for improvement in the areas of preparatory clinical work, experimental design (particularly with respect to drug dosing), risk assessment, and communication.

With respect to device trials, first-in-man studies may be slightly less cumbersome to undertake. Regulatory bodies require evidence that the safety and effectiveness of the device has been established as far as possible prior to human use, that the study will accomplish its objectives, and that trial subjects are protected to every possible extent. For first-in-man studies, it is generally easier to justify small subject numbers. However, the "pilot" study can often be omitted through the appropriate use of preclinical and bench-top studies where the risks to patients can be shown to be minimal.

## 4 SUMMARY

The specialty of urology is a fertile ground for the astute scientist with many opportunities for cutting-edge drug and device research. Only through the conduct of properly constructed clinical trials and the proper interpretation of study findings, however, will our patients truly reap the benefits of this activity. It is hoped the

information provided in this chapter will aid those considering urological investigation and inspire a commitment to the highest quality research endeavor.

## REFERENCES


1. Friedman, L. M., Furberg, C., and DeMets, D. L. (1998), *Fundamentals of Clinical Trials*, 3rd ed., Springer, New York, pp. 41–78.
2. Cutler, S. J., Greenhouse, S. W., Cornfield, J., et al. (1996), The role of hypothesis testing in clinical trials. Biometrics seminar, *J. Chronic Dis.*, 19, 857–882.
3. Denstedt, J. D., Wollin, T. A., Sofer, M., et al. (2001), A prospective randomized controlled trial comparing nonstented versus stented ureteroscopic lithotripsy. *J. Urol.*, 165, 1419–1422.
4. Anon. (1992), American Urological Association symptom index for benign prostatic hyperplasia, *J. Urol.*, 148, 1549–1557.
5. Rosen, R., Riley, A., Wagner, G., et al. (1997), The international index of erectile dysfunction (IIEF): A multidimensional scale for assessment of erectile dysfunction, *Urology*, 49, 822–830.
6. Health Assessment Lab, Medical Outcomes Trust and QualityMetric Incorporated, 1992–2002; available at http://www.outcomes-trust.org/instruments.htm#SF-36.
7. Litwin, M. S., Hays, R. D., Fink, A., et al. (1998), The UCLA Prostate Cancer Index: Development, reliability and validity of a health-related quality of life measure, *Med. Care*, 36, 1002–1112; available at http://roadrunner.cancer.med.umich.edu/epic/epicmain.html.
8. Joshi, H. B., Newns, N., Stainthorpe, A., et al. (2003), Ureteral stent symptom questionnaire: Development and validation of a multidimensional quality of life measure, *J. Urol.*, 169, 1060–1064.
9. Beiko, D. T., Watterson, J. D., Knudsen, et al. (2004), Double-blind randomized collected trial assessing the safety and efficacy of intravesical agents for ureteral stent symptoms after extracorporeal shockwave lithotripsy, *J. Endourol.*, 18, 723–730.
10. Higgins, B., and Thompson, I. M. (2004), The prostate cancer prevention trial: Current status. *J. Urol.*, 171, S15.
11. Thompson, I. M., Goodman, P. J., Tangen, C. M., et al. (2003), The influence of finasteride on the development of prostate cancer. *N. Engl. J. Med.*, 349, 215–224.
12. Roehrborn, C. G. (1996), The placebo effect in the treatment of benign prostatic hyperplasia, in Kirby, R. S., McConnell, J. D., Fitzpatrick, J. M., Roehrborn, C. G., and Boyle, P., Eds., *Textbook of Benign Prostatic Hyperplasia*, Oxford, UK, Isis Medical Media, pp. 239–253.
13. Good, P. I. (2006), *A Manager's Guide to the Design and Conduct of Clinical Trials*, 2nd ed., Wiley-Liss, Hoboken, NJ, pp. 23–36.
14. Barbui, C., Violante, A., and Garattini, S. (2000), Does placebo help establish equivalence in trials of new antidepressants? *Eur. Psychiatry*, 15, 268–273.
15. Tan, A. H., Nott, L., Hardie, R., et al. (2005), Long-term results of microwave thermotherapy for symptomatic benign prostatic hyperplasia, *J. Endourol.*, 19, 1191–1195.
16. Portney, L. G., and Watkins, M. P. (2000), *Foundations of Clinical Research: Applications to Practice*, 2nd ed., Prentice Hall Health, Upper Saddle River, NJ, pp. 33–133.
17. Weijer, C., Dickens, B., and Meslin, E. M. (1997), Bioethics for clinicians: 10. research ethics, *Can. Med. Assoc. J.*, 156, 1153–1157.

18. Levine, R. J. (1986), *Ethics and Regulation of Clinical Research*, 2nd ed., Urban & Schwarzenberg, Baltimore, pp. 87–98.
19. The Belmont Report: Ethical principles and guidelines for the protection of human subjects of research; available at http://www.impactcg.com/docs/BelmontReport.pdf.
20. Anon. (2000), World Medical Association Declaration of Helsinki: Ethical principals for medical research involving human subjects, *JAMA*, 284, 3043–2045.
21. CIOMS (2002), Council for International Organizations of Medical Sciences (CIOMS). International ethical guidelines for biomedical research involving human subjects, *Bull. Med. Ethics*, 182, 17–23.
22. U.S. Food and Drug AdministrAtion. Center for Devices and Radiographic Health; available At: http://www.fda.gov/CDRH/.
23. Health Canada; available at: http://www.hc-sc.gc.ca.
24. Beiko, D. T., Watterson, J. D., Knudsen, B. E., et al. (2004), Double-blinded prospective randomized controlled trial assessing the safety and efficacy of intravesical agents for ureteral stent symptoms following extracorporeal shock wave lithotripsy, *J. Endourol.*, 18, 723–730.
25. Hemelaar, J. (2007), Minimising risk in first-in-man trials, *Lancet*, 369, 1496–1497.
26. Royal Statistical Society's working party on statistical issues in first-in-man studies; available at: http://www.rss.org.uk/.

# *Clinical Trials on Cognitive Drugs*

**Elisabetta Farina and Francesca Baglio**

*Neurorehabilitation Unit, Don Carlo Gnocchi Foundation, Scientific Institute and University, IRCCS, Milan, Italy*

**Contents**

1 Kinds of Diseases Representing Target in This Field
  1.1 Dementia
  1.2 Trauma
  1.3 Stroke
2 Selecting Population for Study
  2.1 Common Criteria
  2.2 Particular Criteria
3 Endpoint selection and Assessment Tools for Efficacy and Safety Evaluation
  3.1 Efficacy Evaluation
  3.2 Cognitive Tools
  3.3 Outcome Measures of Function
  3.4 Global and Quality-of-Life Measures
  3.5 Measures of Neuropsychiatric and Behavioural Changes
  3.6 Other Measures: Outcomes Specific To Caregivers
  3.7 New and Future Measures
  3.8 Safety Evaluation
4 Duration of study and Possible bias in Trials on Cognitive Impairment
  4.1 Duration
  4.2 Possible Bias
5 Conclusion
  References

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

## 1 KINDS OF DISEASES REPRESENTING TARGET IN THIS FIELD

Cognitive disabilities, especially memory disturbances and cognitive deficits, posttraumatic brain injury (TBI), and poststroke are important targets for clinical pharmacological trials given their important epidemiological and social impact (see Table 1).

Until a few years ago, patients with dementia, stroke, or TBI faced a bleak future with a diminished quality of life, reduced productivity in terms of their capacity to work and contribute to family life, and dependence on caregivers. Without a cure or any means of alleviating the patient's symptoms, for example, by prescribing drugs that slowed the progression of a degenerative disease or neurorehabilitation to restore some of the brain function lost through stroke or traumatic brain injury, the doctor's role was reduced to advising the patient on coping strategies. However, over recent years and thanks to new scientific data, the treatment of brain diseases has undergone a revolution.

### 1.1 Dementia

Brain changes leading to cognitive decline and dementia, and their relation to disability, are key topics requiring investigation at the population level. Global aging is a recent phenomenon. Its potential impact on social and economic aspects of more developed countries highlighted aging as an important issue toward which resources must be directed. Dementia is an increasingly common diagnosis in our aging population, and the numbers are expected to rise exponentially in coming years. Alzheimer's disease (AD) alone actually affects 4.5 million people in the United States [2], while millions more are currently affected by vascular dementia (17.6% of all incident dementia [3]), Lewy body disease, and frontotemporal dementia. Each of these is a distinct entity, though overlapping symptoms and comorbidities occur frequently. Within the past two decades research has progressed rapidly on multiple fronts, including epidemiology, etiology, pathology, diagnosis, and treatment. It is important for clinicians to recognize early signs and symptoms of dementia and to note critical differences among them. Dementia research has moved beyond description of symptoms and clinicopathological correlation to the elucidation of risk factors, the pathobiology of the disease process, and, most important, to the first generation of dementia treatments. Furthermore, the question of how to divide aging phenomena into categories of healthy, normal, or disease-related remains a difficult but important goal in medicine: We are entering an era of dementia care that will be based upon the identification of potentially modifiable risk factors and early disease markers and the application of new disease-specific diagnostic tools and treatment modalities. Literature data showed, for example, that older persons

**TABLE 1 Epidemiological and Social Impact of Dementia, Stroke, and TBI**

| Disease | Prevalence | Annual Cost (euro)/Year |
|---|---|---|
| Dementia | 5 million people | 55 billion |
| Traumatic brain injury | 700,000 | 3 billion |
| Stroke | 1 million people | 22 billion |

*Source*: Data from Olsen et al. [1].

can develop demonstrable cognitive impairment ("mild cognitive impairment," MCI; [4]), especially memory deficits without crossing the threshold for dementia, but that these patients have an increasing risk of developing dementia, especially Alzheimer disease [5]. We expect that pharmacological approaches will become increasingly integrated with behavioral, genetic, and neuroimaging methods to investigate not only disease processes but also the normal and preclinical individual differences that underlie successful—or unsuccessful—aging. These future directions share an important feature with the current focus of the field: a shift from the dismal characterization of aging as an inevitable process of brain damage and decline to the concept (emerging from cognitive neuroscience) that aging can be successful associated with gains and not only with the losses. It is not necessarily a unidirectional process but rather a complex phenomenon characterized by reorganization, optimization, and enduring functional plasticity that can enable the maintenance of a productive and happy life.

### 1.2 Trauma

Traumatic brain injury is another leading cause of disability, and survivors often suffer cognitive, mood, and behavioral disorders. TBI is an insult to the brain that leads to temporary or permanent impairments of cognitive abilities and physical functioning. TBI results principally from vehicular incidents, falls, acts of violence, and sports injuries and is more than twice as likely to occur in men as in women. The estimated incidence in U.S. rate is 100 per 100,000 persons, with 52,000 annual deaths. The highest incidence is among persons aged 15–24 years and 75 years or older, with a less striking peak in incidence in children aged 5 years or younger [6]. Since TBI may result in lifelong impairment of physical, cognitive, and psychosocial functioning and prevalence is estimated at 2.5–6.5 million individuals, it is a disorder of major public health significance. Mild TBI is significantly underdiagnosed, and the likely social burden is therefore even greater. Given the large toll of TBI and absence of a cure, prevention is of paramount importance. More recently, a hospitalized incidence rate of 229 per 100,000 was calculated for England for TBI in 2001–2002 [7]. Again, significant local variation was noted (91–419 per 100,000 across health authorities). In the United Kingdom, "serious" TBI is estimated to have an incidence rate of 52 per 100,000, while an incidence of 12 per 100,000 was reported in Australia for "severe" brain injury [8].

Research in traumatic brain injury has shown that the primary traumatic insult sets in training a cascade of events in the brain that leads to secondary injury, and that this secondary injury can be exacerbated by systemic insults such as hypoxia and shock [9]. Strategies to minimize both secondary insults and the evolution of secondary damage in the acute phase, while maximizing plasticity in the following course, are essential goals for those patients, but pharmacological studies in this field are lacking. The same can be said for long-term cognitive impairment deriving from a previous TBI, a condition affecting a significant percentage of TBI survivors.

### 1.3 Stroke

Stroke is a vascular disorder characterized by the sudden death of brain cells due to a reduced or blocked supply of blood. The symptoms depend on the site of the

stroke, but they can include motor, speech deficits (aphasia), visuospatial deficits (including neglect), cognitive decline, and dementia. The incidence of stroke is reaching pandemic proportions: Stroke is the third leading cause of death and one of the leading causes of adult disability in North America, Europe, and Asia [10–12]. Each year across the world, 5.5 million people die as result of a stroke. A further 15 million people survive but are disabled and face the prospect of a later, even more serious stroke [13]. Costs of care, which have to be met by either the individual or the state, can also prove crippling [14].

A number of well-designed randomized stroke trials and case series have now been reported in the literature to evaluate the safety and efficacy of pharmacologic therapy for the treatment of acute ischemic stroke, but there are few data on long-term cognitive outcomes and possible treatments after first-ever stroke. Literature data showed that recurrent stroke significantly contributes to global cognitive decline after a first-ever stroke. Furthermore, studies explained that stroke prevention would be important in reducing dementia related to stroke (e.g., hypertension treatments; [15]). There is a growing interest in the latter because patients with only mild vascular cognitive impairment and no dementia are at increased risk of cognitive deterioration and should have more opportunities for treatment and prevention [16]. Further investigation of these issues may help to understand mechanisms underlying stroke-related cognitive decline and assist in planning interventions to prevent dementia related to stroke, and to ameliorate existing cognitive deficits.

## 2 SELECTING POPULATION FOR STUDY

### 2.1 Common Criteria

***Diagnostic Accuracy*** Before starting a clinical pharmacological trial, clear classification of the diseases and well-recognized diagnostic criteria are necessary to realize the selection of the study population. This is relevant not only for scientific reasons (standardization and comparability between data collected by several scientific groups) but also for practical aspects (e.g., without a clear stadiation of disease, it is unthinkable to lay down different therapeutic options). A detailed anamnesis as well as physical examination at presentation are necessary for the overall evaluation of the single subject; the employment of consensus criteria for the diagnosis is essential for the group analysis comprehension. Actually, consensus criteria are well defined in dementia (Table 2 cites the most recent clinical criteria). Unlike dementia, cognitive and behavioral sequelae in stroke and TBI recognize only clinical classifications of the specific symptoms. Nevertheless, attributing single individuals to the classic syndrome categories is not easy, due to the complexity of the underlying brain damage and the existence of clinical variants. For example, aphasia, the loss or impairment of language caused by damage to several discrete cognitive systems, is one of the most devastating cognitive sequelae of stroke. Well-defined classic aphasic profiles (anomia, Broca's, conduction, Wernicke's, and transcortical) are frequent when patients with only single lesions are considered [24], whereas global and nonclassified aphasias account for 50% of cases admitted to acute stroke units, especially among patients with previous strokes [25]. A difficult to ascertain proportion of aphasic individuals cannot easily be assigned to the classic

**TABLE 2 Recent Clinical Criteria in Dementia**

| Type of Dementia | Criteria |
|---|---|
| Alzheimer's disease | Criteria of the Diagnostic and Statistical Manual of Mental Disorders [17]<br>The National Institute of Neurologic, Communicative Disorders and Stroke—Alzheimer Disease and Related Disorders Association (NINCDS-ADRDA) criteria [18];<br>NINCDS-ADRDA research criteria [5] |
| Dementia with Lewy body | The Consortium for DLB diagnostic criteria [19, 20] |
| Frontotemporal dementia | Frontotemporal lobar degeneration: a consensus on clinical diagnostic criteria [21]<br>Report of the Work Group on Frontotemporal Dementia and Pick's Disease [22] |
| Vascular dementia | The National Institute of Neurologic Disorders and Stroke and the Association Internationale pour la Recherche et l'Enseignement en Neuroscience (NINDS-AIREN) diagnostic criteria [23] |

*Note*: The clinical diagnosis should rely on criteria that have been proposed to increase the reliability and accuracy of the diagnosis. The accuracy of these diagnostic criteria varies as a function of the dementia.

syndrome categories, whereas others display atypical clinicoradiological correlations (e.g., Wernicke's aphasia associated with frontal lobe lesions), in part, a result of the idiosyncratic brain organization of language networks [26]. Indeed, other right-handed individuals show no aphasia despite having large lesions in the left hemisphere because in such cases language is innately lateralized to the right hemisphere [27]. However, functional imaging [functional magnetic resonance imaging (MRI)], mapping language brain areas, may currently theoretically provide means to improve the diagnostic accuracy by clinicians [28, 29].

In any case, it is now recommended submitting patient to baseline imaging evaluation before beginning trial [computed tomography (CT) or magnetic resonance imaging (MRI), within 6–12 months]. This is important to confirm the original diagnosis and order to exclude other concomitant pathologies (e.g., vascular impairment for degenerative dementia, brain tumor).

***Severity Criteria*** Another relevant aspect to consider when planning a trial is the stadiation of disease that includes both temporal dimension and severity. In pharmacological trials stadiation is important because different phases of illness require different pharmacological approaches. For example, this is true for the treatment of AD: Literature data showed that cholinesterase inhibitors for AD treatment are useful in the mild-to-moderate phase and ineffective in the severe stage. On the contrary, memantine, an *N*-methyl-D-aspartate (NMDA) antagonist, is approved for the treatment of moderate to severe AD. Predefined severity criteria should be made explicit in future trials to ensure transparency with regard to comparability and general applicability of results.

***Cultural and Educational Background*** Cultural and educational background of the population under study is another relevant characteristic to hold in consideration. Really, cognitive and behavioral measures are particularly vulnerable to the effect of culture, language, and education. The relevance of some domains may

vary between cultures; functional disability, for example, may be less important in cultural contexts where independence and autonomy in activities of daily living are not part of the older person's role [30]. Similarly, memory impairment is not regarded as important in all cultures [31]. Moreover, the recruited subjects must have the same range of years in age and in school attendance. The age and school attendance setting is an essential requisite for making a comparison between different group performances on assessment tools.

***Baseline Safety Evaluation*** Regardless of the drug that we want to try, it is also essential to assess adverse effects of treatment in order to determine the risk–benefit ratio. For this reason, it is well advised to have a safety assessment that at least includes laboratory examinations (hematology and urinalysis) and an electrocardiogram (ECG). Further clinical or instrumental evaluations before starting and during trial are established in accordance with the preliminary safety data of the drug to try.

## 2.2 Particular Criteria

***Population with Dementia*** In patients with dementia some aspects are typical, and they must be kept in mind before starting the clinical trial. The power of clinical trials in dementia is significantly reduced by diagnostic inaccuracy; therefore, there is a tension between requirements for diagnostic accuracy and the necessity for clinical trial results to be generally applicable to the clinical population that might subsequently receive the intervention.

First, it is well recognized that diagnostic accuracy and the detection of treatment effects may be easier to attain in moderate than in mild dementia, but the potential benefits of disease modification are likely to be greatest in the early or prodromal stages. Earlier intervention with disease-modifying therapies is likely to be more effective when there is a lower burden of amyloid and hyperphosphorylated tau and may truncate the devastating effects of secondary events due to inflammatory, oxidation, excitotoxicity, and apoptosis. Also for this reason revised research criteria for AD [National Institute of Neurologic, Communicative Disorders and Stroke-Alzheimer Disease and Related Disorders Association (NINCDS-ADRDA) research criteria; [5]] are now available. These new criteria would allow diagnosis when symptoms first appear, before full-blown dementia, thus supporting earlier intervention at the prodromal stage. They are centered on a clinical core of early and significant episodic memory impairment, and they stipulate that there must also be at least one or more abnormal biomarkers [among structural neuroimaging with MRI, molecular neuroimaging with positron emission tomography (PET), and cerebrospinal fluid analysis of amyloid β or tau proteins]. However, these criteria require significant expertise, technical skills, and financial resources to allow the comprehensive assessment of MRI, PET, and cerebrospinal fluid, and this diagnostic framework may not yet be feasible in all memory clinics and certainly not in most epidemiological studies and all clinical trials: The research criteria will need to be adapted for use in standard clinical settings.

The appearance of these criteria, however, could overcome problems encountered in clinical trials targeting MCI population, where the clinical heterogeneity of patients with this diagnosis makes the demonstration of clinical effects of a drug very difficult.

A second aspect to consider is the coexistence of several pathologies in these patients given their old age. Indeed, dementia onset is most often situated in old age (subjects who are 65 years old or more), and it is common knowledge that age is an independent risk factor for other pathologies (hypertension, diabetes, depression, etc.). Therefore, it would be necessary to exclude those conditions that may adversely affect the cognitive domain and/or giving safety problems. However, it is important to avoid a hyperselection of the study population in order to obtain a group of subjects similar to the general patient population, that is, patients seen in everyday clinical practice. This is also important to avoid insurmountable difficulties in recruitment.

Another point to consider is the common presence in dementia of behavioral and psychological symptoms [Behavioral and Psychological Symptoms of Dementia (BPSD) or neuropsychiatric symptoms; [32]]. These disturbances are distressing, present major difficulties for caregivers, and accelerate nursing home placement [33]. For these reasons antipsychotic drugs are commonly prescribed to many people with dementia (up to 45%) in residential or nursing homes but also at home, often for prolonged periods. However, studies reported that antipsychotics have a negative impact on cognition in patients with dementia [34, 35]. Even though it would be appropriate to exclude patients treated with these drugs from clinical trials, also in this case it is important to avoid a population hyperselection in order to reproduce as much as possible the real clinical situation in everyday practice. Therefore, you can evaluate whether to include patients in some specific cases: when the patient assumes a low dose of benzodiazepines with hypnotic purpose or when he has been in steady therapy with antidepressant for at least 6 months. Furthermore, it is possible to consider in some specific trials the sporadic use of low doses of neuroleptics (e.g., haloperidol).

Approaching dementia, another critical aspect must be well known: the mixed dementia. Mixed dementia, a combination of definite AD and vascular encephalopathy, is a controversial but important issue to consider given the frequency of this condition in the oldest patients. In particular, it is unclear how many vascular lacunar lesions are compatible with a diagnosis of AD (generally it is considered only one), taking into account that isolated leukoaraiosis is not considered an exclusion criteria for the diagnosis of neurodegenerative diseases. In the future, it could be useful to include into pharmacological trials also patients with mixed dementia.

### *Population with Cognitive and Behavioral Deficits after TBI and Stroke*

Different than in dementia, several cognitive and behavioral clinical situations can occur following TBI or stroke, depending on the degree and type of brain injury. The variability of pretraumatic and prestroke brain function further adds elements to the clinical complexity. For example, although some brain structures are more vulnerable in TBI than others, the vast heterogeneity of the TBI patients is likely

to cause high variability in the therapeutic response to any agent. Moreover, a certain cognitive symptom may have a variable biochemical background. For example, apathy and poor initiation may arise from deficits in the dopaminergic, serotonergic, noradrenergic, or cholinergic systems or from an imbalance between these systems. At the clinical level, however, the manifestations may be very similar.

Therefore, in order to evaluate the efficacy of a pharmacological treatment in these patients, it is important to chose the population using stringent and accurate multicriteria. These criteria can be both clinical (e.g., clinical presentation at the neurological examination: neglect, aphasia, etc.) and founded on neuroimaging findings (injury level established by CT or MRI examination). Functional imaging, in particular, may theoretically provide means to evaluate the basis for treatment interventions [36], but for practical and economical reasons it seems unrealistic that, at an individual level and at larger scale, we could effectively use such tools for treatment decisions in the foreseeable future.

Another important point is that, since the damage can occur at any age, age is an important prognostic factor, and the recovery can be different between several patients, it is necessary that patients recruited for the trial are similar not only in the demographic characteristics (age, sex, and education) but also as far as premorbid conditions are concerned (e.g., it is necessary to exclude preexisting psychiatric or substance abuse problems, poor general health). The same can be said for comorbidities (such as chronic pain, depression, substance abuse, life stress, unemployment, and protracted litigation). The use of poorly matched experimental and control groups increases the likelihood that neuropsychological group differences will reflect premorbid group differences rather than the effect of damage postinjury. Moreover, it is of primary importance the good knowledge of TBI clinical details (e.g., duration of posttraumatic amnesia, coma duration, patient level at the Glasgow Coma Scale). Clearly, according to previous cited common criteria, another relevant aspect to consider is the stadiation of disease: temporal dimension postinjury and severity of the pathology at the time of the enrollment. The temporal dimension postinjury is relevant: The overall effect size for the influence of mild TBI on neuropsychological functioning, for example, in the *postacute* phase (more than 3 months postinjury) is very different than in the *acute* phase. Indeed, in a meta-analysis on neuropsychological outcome following TBI, Schretlen and Shapiro [37] concluded that mild TBI results in clear decrements in cognitive functioning that resolve in 1–3 months. In contrast, they illustrated that patients with moderate or severe brain injuries show clear evidence of early recovery but they are at risk for frank deficits more than 2 years after injury. Moreover, studies of spontaneous recovery in poststroke aphasia have shown that the greatest improvement occurs in the first 2 or 3 months, with a discernible but less evident improvement in the following months, and most patients reaching a plateau after 1 year [38]. If the stability of cognitive sequelae is ascertained, TBI and stroke survivors can be candidates to cross-over studies and not only to classical prospective clinical trial. This is not possible in dementia due to the progressive nature of the disease.

To evaluate the severity of TBI pathology at the time of the enrollment, specific scales are actually available to establish a global cognitive evaluation of the patients. They are also useful to provide follow-up and comparison of the patients [e.g., Rancho Los Amigos (RLA) Levels of Cognitive Functional Scale for TBI].

## 3 ENDPOINT SELECTION AND ASSESSMENT TOOLS FOR EFFICACY AND SAFETY EVALUATION

### 3.1 Efficacy Evaluation

Dementia, stroke, and TBI are multifactor disorders that negatively impact cognitive, behavioral, and functional abilities. Complete and accurate evaluation of the efficacy of treatments is crucial, but the heterogeneity of symptoms makes uniform assessment difficult. Due to the complex nature of these diseases, it is crucial to design assessment procedures that accurately track disease progression across a wide variety of symptom domains. Although a number of outcome measures have been developed for this purpose, it is suitable to use the diagnostic tools that have sensibility, specificity, and relevancy enough to test modifications of estimated characteristics (primary and secondary endpoints). Generally, primary outcome measures assess key pathology domains (such as cognition, function, global impression, and quality of life in dementia), while secondary endpoint measures assess other meaningful and measurable domains, which may constitute appropriate targets for interventions (e.g., caregiver outcomes, behavioral measures, activity of daily living).

These assessment tools must be based on reliable and validated scales and scores and must be usable in many settings. Ideally, the administration of the instrument should be brief; however, many of the measurement tools used in clinical research take too much time to administer to be employed in clinical settings.

***Dementia*** The goals of treatments in this field can be to prevent dementia emerging in patients with MCI, or, in those with established dementia, to evaluate improvement and stability versus deterioration. The outcome measures used to evaluate efficacy of dementia treatments should be sensitive to both improvement and decline and have minimal floor and ceiling effect [39]. The primary outcomes include key domains of dementia: cognition, function, global impression, and/or quality of life. Secondary measures assess behavioral ratings and other outcomes such as nursing home placement, caregiver burden, and others.

Table 3 shows the principal primary and secondary outcome measures used to evaluate efficacy in dementia treatments. For MCI patients, one of the primary outcome measures should be the conversion to dementia, that is, how much a drug is able to reduce the absolute and relative risk of conversion. However, this is a critical primary endpoint to evaluate, due to difficulties in establishing when the individual patient exactly passes the threshold of dementia (it basically depends from loss of independence in everyday life), and to the necessity of planning a long-term study to catch enough cases of conversion to dementia (at least 2 years).

### 3.2 Cognitive Tools

The cognitive tools should provide an assessment of multiple cognitive functions that are affected by dementia (memory, orientation, language, frontal abilities, praxis, and others), plus a measure of global cognitive decline. The Alzheimer's Disease Assessment Scale–Cognitive subscale (ADAS-Cog) is the de facto standard primary outcome neuropsychological measure for AD trials [40]. It measures several

**TABLE 3 Primary and Secondary Outcomes Used to Evaluate Efficacy in Dementia Treatments (see text for further details)**

| Outcome | Scale |
|---|---|
| Cognition | Alzheimer's Disease Assessment Scale—cognitive subscale (ADAS-Cog) |
| | Mini-Mental State Examination (MMSE) |
| | Severe Impairment Battery (SIB) |
| Function | Activities of daily living (ADL): basic ADL (BADL) and instrumental ADL (IADL) |
| | The Alzheimer's Disease Activities of Daily Living International Scale (ADL-IS) |
| | Performance measures: |
| | The Physical Performance Test (PPT) |
| | The Direct Assessment of Functional Status (DAFS) |
| | The Functional Living Skills Assessment (FLSA) |
| Global impression | Clinical Interview-Based Impression of Change (CIBIC, CIBIC-plus) |
| Quality of life | Quality of Life–Alzheimer Disease Scale (QOL-AD) |
| | The Late-Stage Dementia Scale (QUALID) |
| | The Alzheimer Disease-Related Quality of Life Scale (ADRQL) |
| | The Dementia Quality of Life Scale (D-QoL) |
| Behavioral and psychological symptoms | Neuropsychiatric Inventory (NPI) |
| | ADAS non-Cog |
| | The Behavioral Pathology in Alzheimer's Disease Rating Scale (BEHAVE-AD) |
| | Cornell Geriatric Scale |
| Other | Global measures of burden |

cognitive domains, including memory, language, and praxis. Total scores range from 0 to 70, with scores greater than 40 indicating severe impairment and a score of 10–25 indicating mild impairment. Many regulatory authorities recognize a four-point change on the ADAS-Cog at 6 months as indicating a clinically important difference [41–44]. Interestingly, the ADAS-Cog scale has the advantage of having parallel forms that avoid learning as a confounding factor. Unfortunately, it has limitations that are worth noting. The first of these is important when comparing results in different kinds of dementia. Since the ADAS-Cog is an assessment designed for AD, other dementias require the use of further neuropsychological instruments to assess the cognitive domain particularly affected in that specific type of disease. In patients with frontotemporal dementia, for example, frontal abilities that are specifically involved should be investigated (e.g., with the Frontal Assessment Battery [45]); in patients with Lewy body dementia, visuospatial deficits and cognitive fluctuations should be taken into consideration (for visuospatial deficits, several tests exist to evaluate visuoperceptive or visuospatial abilities, for example, some subtests of the Birmingham Object Recognition Test and the Benton Line Orientation Test, respectively: For the latter, short forms are also available [46]; cognitive fluctuations should be evaluated with specific questionnaires, such as the DLB fluctuation scale by Ferman et al. [47]).

The ADAS-Cog takes 30–45 minutes to administer; therefore it is widely used only by researchers. On the contrary, Mini-Mental State Examination (MMSE; [48]) is a brief cognitive screening tool that only takes 5–15 minute to administer. Therefore, it is used both in research and clinical settings.

As previously noted, the measures should be sensitive to disease progression over time and useful for a broad range of severity. Another limitation of ADAS-Cog is that there is a floor effect in severe AD patients [49]. For this reason, this scale is

not generally used to evaluate the cognitive changes in advanced patients, and the Severe Impairment Battery (SIB; [50–52]) is currently the most accredited instrument for tracking progression in severely demented patients [53]. The SIB includes six subscales (attention, orientation, language, memory, visual perception, and construction) and a brief assessment of social skills, praxis, and response to name. Total score ranges from 1 to 100, with a score of 63 or less being considered "very severely impaired"; this battery provides accurate assessment in patients with MMSE scores of 13 or less [54]. Like the ADAS-Cog, the original SIB is a long tool to be administered to the patient (it takes approximately 30 minutes). For this reason, a short form of the Severe Impairment Battery (SIB-S) has been developed that only takes 10–15 minutes to administer, making it more appropriate for the use in clinical pharmacological trials in patients with very severe dementia, while it maintains the attributes of the original SIB [55].

### 3.3 Outcome Measures of Function

Outcome measures of function are other important data to collect in order to demonstrate the efficacy of a drug in clinical pharmacological trials. Loss of functioning in complex tasks of everyday life is a hallmark feature of dementing illness: In fact, current clinical diagnostic criteria for dementia [17, 18] require documentation of cognitive decline as well as loss of competence in either social or occupational domains. Furthermore, measures of function allow clinicians to assess the need for personal and institutional care, and to design individually tailored interventions, so that a demented individual receives sufficient help while avoiding unnecessary assistance that may lessen patient self-esteem, increase caregiver burden, and accelerate functional deterioration. The ability to perform activities of daily living (ADLs) has a significant impact on the patient, caregiver, and society. Among the functional disability scales based on patient or informant response, we can distinguish generic tools.

Generic scales are designed to measure function regardless of the person's specific medical history: With these scales, the disability of a patient with dementia can be compared with the disability of a patient with arthritis, for example. This type of tool evaluates basic ADLs (BADLs) and/or instrumental ADLs (IADLs). IADLs include activities that concern the ability to adapt to the environment: Some of these activities (such as telephone usage, financial management, and shopping) are particularly dependent on cognitive functioning, and often serve as earlier predictors of poor outcome; this subset of IADLs is known as advanced cognitive IADL [56]. The most popular generic BADL scale is the Katz ADL scale [57]: This measure of basic functioning requires an observer to rate the patient's ability in bathing, dressing, going to the toilet, transferring, maintaining continence, and feeding. The most commonly used IADL scale is the Lawton IADL scale [58].

This last one includes eight items: telephoning, shopping, meal preparation, housekeeping, laundry, transportation, medications, and handling finances. A problem with the Katz IADL is that the original version of this scale uses all items for women and only five tasks for men (three items are removed for men to avoid tasks that men did not usually perform because of societal norms associated with the division of labor in household management: But these norms are changing in the last decades). A low performance in IADL scales is associated with mild

dementia, while ADL deficits characterize more severe dementia [59]. Although, measuring ability to perform ADLs and IADLs is important when assessing the therapeutic efficacy of a drug, these scales are very poorly sensitive to change, and these generic scales have been criticized for a relative insensitivity to functional losses due to cognitive impairment [60]. Therefore, alternative tools should be identified, such as more detailed functional scales or other performance instruments.

The most popular specific informant-based scales are the Blessed scale [61], the Nurses' Observation Scale for Geriatric Patients [62] (it includes a subscale specifically designed to assess IADLs, but this last one contains only five items), and the Alzheimer's disease activities of daily living international scale (ADL-IS) [63]. The ADL-IS was launched to develop an ADL scale aimed to measure pharmacological response in AD patients [63]; it was carefully developed in a 10-year period and covers a large sample of ADLs. It shows high correlations with measures of cognitive functions and stage of dementia, thus appearing a promising tool also for clinical setting. However, clinical experience with this scale is still limited. It must be added that the use of informant reports is appealing because it can allow comprehensive and longitudinal evaluations but has potential disadvantages: The source of information is filtered through an observer, and it may contain distortions based on the personal needs and opinions of the informant [64]. For this reason, in the last years, some performance measures of everyday functioning have been proposed to assess demented patients.

Performance measures are tools in which an individual is asked to perform an activity and is evaluated in a formal manner on that performance with standardized criteria (counting, timing, observation of the need for assistance, and appropriateness or completeness of the task). These measures have been used in medical rehabilitation for diagnosis and to demonstrate therapeutic progress, and have shown to predict survival, hospitalization, use of assistance, long-term-care, and nursing home placement [65–69]. The main advantage to this approach is that it directly measures ability, thus eliminating faulty perceptions. Moreover, with performance measures it is possible to take into account a large variability of the ability to perform a task, thus increasing the sensitivity of the tool [70]. Other claimed advantages are good patient acceptability and good interrater reliability [68, 71–73]. Examples are the ADL Situational Test (AST; [74]), the Physical Performance test (PPT; [72]), the Structured Assessment of Independent Living Skills (SAILS; [73]), the Direct Assessment of Functional Abilities (DAFA; [75]), the Everyday Problems Test for the Cognitively Challenged Elderly (EPCCE; [76]), and the Direct Assessment of Functional Status (DAFS; [71]). In the AST [74], evaluation is limited to one BADL (dressing) and three IADL (meal preparation, telephoning, and purchasing), each ADL item broken down into subtasks. For each item there are two scores: the time required to complete each ADL task and a performance score derived from the level of assistance needed by the subject to complete the task.

As far as the PPT [72] is concerned, this tool has been proposed to evaluate a general geriatric population rather than patients affected by dementia. The test includes some motor activities, three BADLs (eating, dressing, and walking), and writing a sentence. Therefore, IADLs are not evaluated in the PPT.

The DAFS [71] features mainly IADL skills (communication abilities, transportation, financial skills, shopping skills), along with two BADL skills (eating and

dressing/grooming skills), and time orientation. Performance is rated as correct or incorrect. The DAFS, which represents an interesting tool, but it explores only four IADL skills and does not define the degree of dependence of the patient.

The SAILS [73] includes a subset of BADL (dressing, eating) and IADL (meal preparation, telephoning, and money-related skills), as well as fine and gross motor tasks, and high-order abilities (communication, social interaction, and orientation). Each item includes five subtasks. Each subtask is scored from "unable" to "normal," with an intermediate level of disability, with both time (on timed tasks) and accuracy (in all tasks) contributing to the score.

The DAFA [75] is a direct measure of IADLs conceived to be compared with the Pfeffer Functional Activities Questionnaire (PFAQ; [77]). Ten items are adapted to explore seven functional domains: money management, shopping, hobbies, meal preparation, awareness, reading, and transportation. Items on the DAFA and PFAQ are both evaluated by using an integer score from 0 (independent functioning) to 3 (dependent functioning).

Finally, the EPCCE [76] has been specifically designed to provide an objective measure of problem solving with respect to cognitively demanding tasks encountered in daily livings. The patient is shown 16 stimuli and asked to solve two problems related to each stimulus. Each item is scored right or wrong. The authors propose the EPCCE as a complement to other performance measures (e.g., the DAFS). These performance instruments, for example, the DAFS, represent interesting tools, but it must be recognized that these instruments could show some limitation in terms of applicability because of different sociocultural backgrounds and temporal bias (the amount of time required to administer the scales).

Our group has developed a new direct performance measure for very mild to moderate patients with dementia: the Functional Living Skills Assessment (FLSA; [78]). Eight areas of interest are evaluated (resources, consumer skills, public transportation, time management, leisure, telephone skills, self-care, and health). Performance is scored according to completeness and level of assistance. We propose its use when a high sensibility to different levels of functional impairment is needed, as evaluation of treatment efficacy, diagnosis of dementia in the very initial phase, and identification of relatively intact functional areas to plan cognitive rehabilitation.

### 3.4 Global and Quality-of-Life Measures

Global measures are one of the most important data used in clinical pharmacological trials. Global assessment scales assess whether treatment effects are clinically meaningful and whether a given agent leads to a change, and they appear to be one of the principal modality to detect changes in mild cognitive impairment patients. The global scales encompass two distinct categories: first, the clinician's interview-based global severity scales, and, second, the clinician's interview-based global change scales [79]. The Clinical Dementia Rating (CDR; [80]) and the related CDR-sum of boxes (CDR-SB), the Global Deterioration Scale (GDS; [81]), and the Functional Assessment Staging (FASTM; [82]) procedure are the global severity scales most often used in clinical trials. These measures are clearly useful in subject categorization in treatment trials (they are relatively free of many of the sociocultural biases inherent in psychometric descriptors and psychobehavioral measures).

They can also be used to demonstrate therapeutic efficacy in terms of the general progression of the dementia process and have also proven to be useful in sensitively assessing pharmacotherapeutic effects in AD treatment trials.

On the other hand, the most popular clinician's interview-based global change scales are the Clinician's Interview-Based Impression of Change (CIBIC; [83]) and the Clinical Global Impressions of Change (CGIC; [84]). With the CIBIC the clinician assesses four areas of patient functioning (general, cognitive, behavioral, and ADLs) by interviewing the patient. Another version of this scale is also available: the Clinician's Interview-Based Impression of Change Plus Caregiver Input (CIBIC-Plus; [85]), which is much more elaborate and includes information derived from caregiver input. In both versions, disease severity is assessed at baseline, and the rate of change from baseline is assessed at each follow-up examination. Score ranges from 1 to 7, where a score of 1 indicates marked improvement, a score of 4 indicates no change, and a score of 7 indicates marked worsening. The CIBIC-Plus procedures require an independent clinician assessment and can provide independent, comprehensive evidence of therapeutic efficacy. The CIBIC-Plus procedures may also be useful in sensitively assessing efficacy in prevention trials (e.g., MCI therapeutic trials, and perhaps in future trials with people referring subjective cognitive complaints).

The CGIC assesses cognitive, behavioral, social, and daily functioning domains, allows for caregiver input, and provides a minimal number of sample probes to guide clinicians in their assessments. Although CIBIC and CGIC are useful global measures, the lack of structure of global assessment reduces the interrater reliability, and it can be difficult to compare these measures across drug studies [84]. Assessment terms, such as "minimal improvement," are not well defined, and the rating of change by the clinician is a subjective measure that can be prone to biases (e.g., the clinician's perception of what constitutes change; sources of information about the patient; how information is recorded; the tests used to assess cognitive functioning; how the change is rated) [83, 86].

Dementia is one of the main causes of decreased quality of life (the degree of need satisfaction within the physical, psychological, social, activity, material, and structural area) [87] among older adults. Quality of life is recognized as an abstract and broad concept encompassing physical well-being, perceptions of well-being, satisfaction, and sense of self-worth. Actually, a substantial number of researchers agree that measuring quality of life is just as important as measuring cognition, disease severity, symptom response, behavioral disturbance, functional abilities, caregiver burden, and resource utilization [88]. Currently, the consensus group for measuring treatment benefits in dementia [89] have indicated that an improvement in quality of life could be an acceptable alternative to global improvement and have identified a number of other meaningful and measurable domains that may constitute appropriate targets for interventions (caregiver outcomes, including psychological consequences and time spent in delivering care, BPSD, quality of life, activities of daily living, global assessment, health utility, and health care costs). Furthermore, quality of life is mentioned as one of the primary outcomes of interest in dementia drug trials in a recent Cochrane update on cholinesterase inhibitors for AD [90].

However, quality-of-life scales are used as outcome measure in only 4.4% of all dementia/MCI-related randomized controlled trials [91]. Quality-of-life measures

cover a range of domains: physical status, functional ability including role functioning, social and community interactions, economic status, psychological status, well-being, somatic sensation, and life satisfaction [92, 93]. The most frequent quality-of-life measures applied in pharmacological trials in dementia or MCI are the Quality-of-Life–Alzheimer Disease Scale (QOL-AD), the patient-rated scale according to Blau (PRB), the Late-Stage Dementia scale (QUALID), and the Goal Attainment Scaling (GAS). Other currently used scales are the Dementia Care Mapping (DCM), the Alzheimer Disease Related Quality-of-Life scale (ADRQL), the Dementia Quality-of-Life scale (D-QoL), the psychosocial domain of the Functional Limitation Profile (FLP), and the European Quality-of-Life scale (Euro-QoL) (for the review of these tools, see [94]). Rating of quality of life in dementia trials, three methods are generally used: self-rating (patient-reported outcomes), proxy rating, and proxy observation scales. Patient-reported outcomes (including measures of subjective symptom report, health-related quality of life, and treatment satisfaction) are unique indicators of disease activity, complementary to established outcome measures [95]. Persons with MCI and mild to moderately severe dementia can be considered good informants of their own subjective states. At more severe stages proxy measures or direct observation may be preferred. The disadvantage of proxy ratings is that they represent a subjective measure, passed throughout the filter of the caregiver's own expectations, mood, burden of care, and the specific relationship with the person being rated [96–98]. In case a patient-reported outcome is impossible, observational evaluation by an uninvolved professional caregiver is the best alternative [96].

### 3.5 Measures of Neuropsychiatric and Behavioral Changes

Clinicians are now aware that neuropsychiatric symptoms (also known as BPSD) are very common and diverse, and that scores on specific symptom clusters, such as measures of affect, apathy, and psychosis, may be more meaningful than an overall score [99]. BPSDs are distressing, present major difficulties for caregivers, and accelerate nursing home placement. These concepts need to be further developed to facilitate valid measurement of treatment effects on specific symptoms or domains. The measure of neuropsychiatric symptoms and behavioral changes most commonly used in dementia trials is the Neuropsychiatric Inventory (NPI; [100]), which assesses 10 common BPSD areas (e.g., delusions, irritability, disinhibition, agitation, sleep disorders) through an interview with the patient's caregiver. The frequency and severity of each behavior is measured, as well as the resulting caregiver distress. Each behavioral domain score ranges from 0 to 12, with a total NPI range of 0–120 (a score of less than 20 indicates mild disturbance; a score ranging from 20 to 50 indicates moderate disturbance; and a score superior to 50 indicates severe disturbance). The NPI takes approximately 10–15 minutes to administer; therefore, it is used both in research and clinical settings. A subscale of the Alzheimer's Disease Assessment Scale (ADAS-non-Cog; [40]) and the behavioral pathology in Alzheimer's disease rating scale (BEHAVE-AD; [101]) are other global neuropsychiatric measures used in AD trials. The ADAS-non-Cog assesses 10 domains (e.g., hallucinations, delusions, depression, power of concentration) and the score range is from 0 to 50 (maximum of deficits). The BEHAVE-AD is a good measure in clinical pharmacological trials to follow up the efficacy of behavioral

drug treatments, taking into account behavioral changes independent from cognitive deficit.

Depression is a relevant psychiatric symptom in dementia, and it may deserve a specific evaluation: the Cornell Geriatric Scale, for example, is a useful scale specifically conceived to detect depression in patients affected by dementia [102].

Because specific neuropsychiatric symptoms vary in prevalence between the different forms of dementia, in clinical trials recruiting non-AD dementia patients, other instruments conceived to evaluate particular neuropsychiatric and behavioral areas according to the dementia type can be selected between the different options offered by the dementia literature.

### 3.6 Other Measures: Outcomes Specific to Caregivers

Measures of caregiver-related outcomes are various and are useful measures as endpoints for AD and related disorders studies. In clinical trials burden and time used to assist the patient represent the most extensively examined outcomes specific to caregivers, followed by other caregiver-specific variables (such as psychological well-being, health care costs, and satisfaction with treatment) and other caregiver-dependent outcomes (institutionalization, proxy survey of interview).

Burden outcome is generally reported using the Neuropsychiatric Inventory Distress rating (NPI-D; [103]) or more global measures of burden included in the Caregiver Burden Inventory (CBI; [104]), the Screen for Caregiver Burden (SCB; [105]), and the Caregivers Stress Scale (CSS; [106]). The NPI-D measures distress in response to 10 behaviors (e.g., delusions, agitation, apathy, disinhibition, sleep disorders) exhibited by the patient.

The CBI, SCB, and CSS are measures based on a conceptual model of stress of caregivers of patients with AD. The CBI is a 24-item, multidimensional 5-subscale (e.g., it measures time, dependence burden, and developmental burden) instrument to evaluate the impact of providing care to AD patients on family members' life. The SCB is a 25-item assessment of subjective and objective burdens that has been validated against a theoretical model of caregiver stress. The CSS is a 15-domain assessment that targets caregiver's primary stressors (e.g., cognitive and behavioral symptoms), secondary role strains (e.g., family conflict), secondary intrapsychic strains (e.g., role captivity, loss of self), and mediators (e.g., management of situation, expressive support). Other burden instruments include the Relatives' Stress Scale [107], which measures caregivers' personal distress, life upset, and negative feelings associated with the caregiving role, and the Cognitive Subscale of the Poulshock and Deimling tool [108], which assesses distress responses to care recipient cognitive symptoms.

Another outcome specific to caregivers is the "time use": caregivers' time spent assisting patients to perform basic or instrumental activities of daily living. Generally, the active time used by informal caregivers is examined, and the estimations of passive care time is reported (time providing supervision but not engaging in a task per se), and of paid care time. Measures of time use includes the Allocation of Caregiver Time Survey [109] and the Caregiver Activity Time Survey [110]. Both measures assess the amount of time spent assisting the patient in activities of daily living. Similarly, the Instrumental Activities of Daily Living Scale-Plus [57], the Physical Self-Maintenance Scale Plus [111], and the Resource Utilization in Demen-

tia questionnaire [112] generate caregiver estimates of the time spent providing such instrumental activities of daily living and self-care assistance.

Although caregiver outcomes generally represent secondary endpoints in AD and related disorders, drug development, with little time and effort information derived from such instruments, could be considerably improved. In order to implement and enrich the quality and interpretability of dementia clinical research with caregiver-specific outcomes, some recommendations are useful to remember. According to Lingler et al. [113], we suggest that investigators should specify entry criteria for caregiver study participants and should collect basic caregiver sociodemographic data (caregiver sociodemographic characteristics—sex, relationship to patient, and age, years caring for the patient, primary or secondary caregiver status, caregiver marital status if a child, co-residence with the patient and other relatives, caregiver ethnicity, income, physical, and mental health status, level of social support). This will aid in the interpretation of data about drug effects on caregiver measures and data about proxy-dependent patient effects. It will also facilitate the identification of variables influencing such effects, and it will allow researchers to investigate not only whether the benefits of antidementia drugs extend to caregivers, but also to what type of caregiver (e.g., spouses or children) and under what circumstances.

### 3.7 New and Future Measures

In the past two decades much progress has been made in refining our understanding of the neurobiology and clinical phenomenology of dementia, particularly in Alzheimer's dementia (see new research criteria for Alzheimer's disease [5]. Actually, several "candidate" biomarkers for diagnosis, severity, progression, or prediction of response are recognized (e.g., structural brain changes visible on MRI with early and extensive involvement of the medial temporal lobe in AD, molecular neuroimaging changes seen with PET with hypometabolism or hypoperfusion in specific areas, and changes in cerebrospinal fluid biomarkers). Volumetric neuroimaging, which seems particularly promising as a measure of progression, and markers of oxidative stress may also become useful. Biomarkers may be informative indicators of the mechanism of action of novel treatments or may help identify potential responders (possible strategies to facilitate work in this area in the future include stratification of study populations to identify predictors of response such as known or putative biomarkers, e.g., Apolipoprotein E—APOE—isoform ApoE-$\varepsilon$4), but it would be considered as supplementary data to collect and not a proxy for clinically significant measures of treatment response.

***Patients with Cognitive and Behavioral Deficits after TBI and Stroke*** Currently, poor level of scientific evidence in treating the cognitive and behavioral deficits due to TBI is available. Future studies are necessary, and, in order to gather scientific evidence, selecting appropriate and reliable outcome measures is a critical issue. An important obstacle in scientific studies is the problem of how to measure cognition and behavior. It is well known that victims of TBI may show fairly good test performance, yet their everyday life is severely restricted or vice versa. In addition, pronounced fluctuation in daily functioning is a common sequel of TBI, reflecting the low cognitive reserve of the injured brain. The Functional Independence

Measurement (FIM) [114], a widely used index of rehabilitation outcome, measures the level of assistance that an individual requires to perform basic life activities. It is an 18-item, 7-level scale that rates the ability of a person to perform independently in self-care, sphincter control, transfers, locomotion, communication, and social activity. The total score is obtained by summing the partial scores, and it ranges from 18 (maximally dependent) to 126 (maximally independent). Two motor and cognitive subscales can be obtained by summing the 13 motor items (range 13–91) and the 5 cognitive items (range 5–35). The Rancho Los Amigos (RLA) Levels of Cognitive Functional Scale is another valid tool to measure global outcome in cognitive sequelae of TBI [115]. This scale was developed to use in the planning of treatment, tracking of recovery, and classifying outcome levels in TBI. There are 8 classification levels, ranging from no response (level I) to confused and agitated (level IV) to purposeful and appropriate (level VIII). It is an easy and simple to perform tool for the global cognitive evaluation of the TBI patients and an appropriate and valid scale as it can both provide follow-up and comparison of the patients.

Several cognitive and behavioral clinical situations can occur following stroke, depending on the degree and type of brain injury. We will consider in the text only poststroke aphasia as an example of cognitive sequela. While speech-language therapy actually remains the mainstay treatment of aphasia, there have been some attempts to introduce the use of pharmacological agents (piracetam, drugs acting on catecholamine systems—bromocriptine, dexamfetamine—and on the cholinergic system—donepezil—as therapeutic strategies in aphasia and other stroke sequelae (e.g., spatial neglect).

Currently, there is unanimous consensus that evaluation of aphasic deficits should be comprehensive to allow the planning of rational therapies [116]. However, there is less agreement regarding the language assessment methodology with debated problems about the different levels of analysis that should be implemented to formally assess language impairment. Some researchers [116, 117] recommend evaluation of groups of aphasic patients using standardized batteries. These batteries examine oral language, reading, and writing and combine the information gathered from different language subtests that assess spontaneous speech, comprehension, repetition, and naming to obtain a graphic profile of performance (e.g., Boston Diagnostic Aphasia Examination; [118]), overall scores of aphasia severity (e.g., Western Aphasia Battery, WAB; [119]), or measures of communicative ability (e.g., Porch Index of Communicative Abilities, PICA; [120]). The estimation of the initial aphasia severity and its clinical profile using standardized assessment batteries usually couple with mapping of lesion size. The overall score of these batteries are increasingly being used as primary outcome measures to estimate changes in the global level of performance after pharmacological treatments [121, 122].

The alternative position to evaluate aphasia is the neuropsychological approach. This approach examines the nature of both normal and abnormal language functioning in terms of current information-processing models. One of the assessment tools based on neuropsychological point of view is the Psycholinguistic Assessment of Language Processing in Aphasia (PALPA; [123]). The PALPA is useful for identifying the different clusters of linguistic impairment as well as the residual areas of strength in the aphasic patient. However, administration of the PALPA or

similar batteries is not always feasible because they are time consuming and not applicable to patients with severe aphasia. In recent years, there has been a growing convergence on the usefulness of both these two assessment approaches: The surface symptoms of aphasia and its global severity are better recognized using standardized aphasia batteries such as the WAB or PICA, whereas evaluations using the cognitive neuropsychological approach and tools such as the PALPA more appropriately assess the nature of language deficits [124].

Like in dementia, given that aphasia and cognitive-behavioral sequelae of TBI negatively affect quality of life, there has been increasing interest in developing reliable and valid assessment instruments in this field (e.g., Stroke and Aphasia Quality of Life Scale-39; [125]). Future studies also need to include psychiatric comorbidity as a predictive factor of outcome in these diseases because depression, anxiety, and social withdrawal can have a negative impact on rehabilitation, potential benefits of drugs treatment, and psychosocial functioning.

### 3.8 Safety Evaluation

Safety evaluation does not depart from the normal pharmacological trial procedures: Assessment of safety included reports of all possible adverse events, laboratory examinations (chemistry, hematology, and urinalysis), and vital sign measurements (electrocardiogram). Especially in the population affected by dementia, given the old age, it is important to monitor laboratory findings (renal and hepatic function, etc.) and possible cardiac adverse events by repeating electrocardiogram.

## 4 DURATION OF STUDY AND POSSIBLE BIAS IN TRIALS ON COGNITIVE IMPAIRMENT

### 4.1 Duration

When performing pharmacological trials for cognitive disturbances, it is necessary to study the patient for a long time in order to detect changes. Recently, Cortes et al. [125], trying to determine the best duration of follow-up necessary to demonstrate the impact of new drugs in dementia, showed that changes undergone by AD patients under cholinesterase inhibitors after 6 months are not adequate to demonstrate the effect of a new treatment. The authors propose that an 18-month trial appear to have the potential to demonstrate clearly the effect of a new drug.

Even if previous randomized controlled trials were of limited duration (often 24 weeks) in order to contain costs and to not loose too many patients, and changes were statistically significant in some cases at 6 months, it is natural to think such a time span is far shorter than the duration of the disease. Therefore, the duration of a randomized controlled study should be at least of 9 or 12 months, with planned evaluations, in our opinion, every 3 months. In addition, it is possible to consider from the beginning an open extension of the randomized trial to obtain further data on patient evolution. Sure, a careful planning of the power of the study (i.e., of the number of patients necessary to demonstrate a treatment effect, calculated by taking into account previous data about the amplitude of the effect of existing drugs for cognitive deficits) is necessary, also taking into consideration the possibility to loose patients at the last follow-up evaluations.

### 4.2 Possible Bias

A bias is a systematic error that alters the results of a study. We will have a bias when the population of treated patients is different from the control population, according to a characteristic that can change results in primary or secondary efficacy measures. To correctly evaluate drug efficacy, the two populations (treated patients and controls) must be identical in all their characteristics, with the exception of the experimental treatment. Possible confounders in studies on cognitive impairment may be, for example, age, education, severity of the disease, initial severity of trauma for TBI, lesion location for cognitive sequelae of stroke, comorbidity, or treatment with other drugs. The best way to avoid systematic bias is to perform an accurate randomization of subjects included in the study; randomization must follow the general rules, particularly it must be performed by a center that is external to centers involved in the study. Good randomization avoids selection bias. Another important system to avoid bias is to maintain the double-blind condition; this is not always possible, however, particularly when the drug has frequent and easily detected side effects (this was the case for interferon in multiple sclerosis trials). Sometimes it can be appropriate to perform stratification of the studied population according to main possible confounders before randomization: This is important also to identify subpopulations of patients who better respond to therapy. At the statistical level, it is indicated to perform analysis by covariating for different factors (e.g., treatment with other drugs to avoid execution bias, comorbidity, age, sex, education, center, and severity).

As in all drug studies, the intention to treat analysis allows maintaining advantages assured by randomization, and it avoids the introduction of systematic errors due to reduced compliance (which can be particularly problematic in cognitively impaired patients), that is, exclusion bias.

To avoid bias, it is also necessary to diversify the sources to recruit patients and controls. Actually, all pharmacological trials are multricentric; generally, centers of different countries are involved. This also allows to recruit a sufficient number of subjects for the study: Calculation of the study power is very important to avoid underscoring drug efficacy due to recruitment of an insufficient patient number. On the other hand, when considering standard statistical significance ($p = 0.05$), we will find a casual association (and not an association due to treatment) once every 20 times. Therefore, it is necessary to limit the number of comparisons, which is the number of primary and secondary endpoints.

Other bias can derive from some deficiency of instruments used to evaluate efficacy. For example, measures of caregiver burden do not take into account some important variables, such as caregiver education, social status, relationship with the patient. (Is he/she a family member or a professional caregiver? Does he/she live with the patient?) Functional scales based on informant reports may be influenced by the same variables. Some cognitive instruments may not have parallel forms to avoid learning bias. Another bias can be represented by the interpretation bias: This arises when different stakeholders assign their individual values to the interpretation of the final results of randomized controlled trials. This can be a frequent bias in the particular domain of trials on diseases causing cognitive impairment, where really effective and satisfying treatments are lacking.

## 5 CONCLUSION

The field of drug trials in medical situations leading to cognitive impairment is a fascinating one. The population of patients who can potentially benefit from therapy is enormous. Actually, treatments available for AD are of limited efficacy, and they do not succeed in stopping cognitive deterioration and loss of function, or to significantly reduce caregiver and social burden. In other forms of dementia (with the exception of dementia with Lewy bodies, for which limited data indicate some efficacy of cholinesterase inhibitors) there is no treatment. Nearly the same can be said for long-term cognitive sequelae of TBI and stroke.

We need the pharmaceutical industry to invest its energy and financial resources in trials aimed at improving this sad situation. However, in order not to waste resources, it is necessary to plan very careful drug trials.

## REFERENCES


1. Olsen, J., Baker, M., Freund, T., et al. (2006), Consensus document on European brain research, *J. Neurol. Neurosurg. Psychiatry*, 77(Suppl I), i1–I49.
2. Hebert, L. E., Scherr, P. A., Bienias, J. L., et al. (2003), Alzheimer disease in the U.S. population: Prevalence estimates using the 2000 Census, *Arch. Neurol.*, 60, 1119–1122.
3. Fratiglioni, L., Launer, L. J., Andersen, K., et al. (2000), Incidence of dementia and major subtypes in Europe: A collaborative study of population-based cohorts. Neurologic Diseases in the Elderly Research Group, *Neurology*, 54, S10–S15.
4. Petersen, R. C., Smith, G. E., Waring, S. C., et al. (1999), Mild cognitive impairment: Clinical characterization and outcome, *Arch. Neurol.*, 56(3), 303–308.
5. Dubois, B., Feldman, H. H., Jacova, C., et al. (2007), Research criteria for the diagnosis of Alzheimer's disease: Revising the NINCDS-ADRDA criteria, *Lancet Neurol.*, 6(8), 734–746.
6. Consensus Conference (1999), Rehabilitation of persons with traumatic brain injury. NIH Consensus Development Panel on Rehabilitation of Persons with Traumatic Brain Injury, *JAMA*, 282(10), 974–983.
7. Tennant, A. (2005), Admission to hospital following head injury in England: Incidence and socio-economic associations, *BMC Public Health*, 5, 21.
8. Wenden, F. J., Crawford, S., and Wade, D. T. (1998), Assault, post-traumatic amnesia and other variables related to outcome following head injury, *Clin. Rehabil.*, 12, 53–63.
9. Bullock, R., Chesnut, R. M., Clifton, G., et al. (2002), Guidelines for the management of severe head injury. BrainTrauma Foundation, American Association of Neurological Surgeons, Joint Section on Neurotrauma and Critical Care, *J. Neurotrauma*, 17, 449–627.
10. Adelman, S. M. (1981), The national survey of stroke: Economic impact, *Stroke*, 12, 169–187.
11. Taylor, T. N., Davis, P. H., Torner, J. C., et al. (1996), Lifetime cost of stroke in the United States, *Stroke*, 27, 1459–1466.
12. Bergen, D. C., and Silberberg, D. (2002), Nervous system disorders: A global epidemic, *Arch. Neurol.*, 59, 1194–1196.
13. MacKay, J., and Mensah, G. A., eds. (2004), *The Atlas of Heart Disease and Stroke*, World Health Organisation, Geneva.

14. Palmer, A. J., Valentine, W. J., Roze, S., et al. (2005), Overview of costs of stroke from published, incidence-based studies spanning 16 industrialized countries, *Curr. Med. Res. Opin.*, 21, 19–26.
15. Srikanth, V. K., Quinn, S. J., Donnan, G. A., et al. (2006), Long-term cognitive transitions, rates of cognitive change, and predictors of incident dementia in a population-based first-ever stroke cohort, *Stroke*, 37, 2479–2483.
16. Serrano, S., Domingo, J., Rodriguez-Garcia, E., et al. (2007), Frequency of cognitive impairment without dementia in patients with stroke: A two-year follow-up study, *Stroke*, 38, 105–110.
17. American Psychiatric Association (APA) (2000), *Diagnostic and Statistical Manual of Mental Disorders*, DSM-IV, 4th ed., APA, Washington, DC.
18. McKhann, G., Drachman, D., Folstein, M., et al. (1984), Clinial diagnosis of Alzheimer's disease: Report of the NINCDS-ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease, *Neurology*, 34, 939–944.
19. McKeith, I. G., Galasko, D., Kosaka, K., et al. (1996), Consensus guidelines for the clinical and pathological diagnosis of dementia with Lewy bodies (DLB): Report of the consortium on DLB international workshop, *Neurology*, 47, 1113–1124.
20. McKeith, I. G., Dickson, D. W., Lowe, J., et al. (2005), Diagnosis and management of dementia with Lewy bodies: Third report of the DLB Consortium, *Neurology*, 65(12), 1863–1872.
21. Neary, D., Snowden, J. S., Gustafson, L., et al. (1998), Frontotemporal lobar degeneration: A consensus on clinical diagnostic criteria, *Neurology*, 51, 1546–1554.
22. McKhann, G. M., Albert, M. S., Grossman, M., et al. (2001), Clinical and pathological diagnosis of frontotemporal dementia: Report of the Work Group on Frontotemporal Dementia and Pick's Disease, *Arch. Neurol.*, 58, 1803–1809.
23. Roman, G. C., Tatemichi, T. K., Erkinjuntti, T., et al. (1993), Vascular dementia: Diagnostic criteria for research studies. Report of the NINDS-AIREN International Workshop, *Neurology*, 43, 250–260.
24. Pedersen, P. M., Vinter, K., and Olsen, T. S. (2004), Aphasia after stroke: Type, severity and prognosis. The Copenhagen aphasia study, *Cerebrovasc. Dis.* 17(1): 35–43.
25. Godefroy, O., Dubois, C., Debachy, B., et al. (2002), Vascular aphasias: Main characteristics of patients hospitalized in acute stroke units, *Stroke*, 33(3), 702–705.
26. Basso, A., Lecours, A. R., Moraschini, S., et al. (1985), Anatomoclinical correlations of the aphasias as defined through computerized tomography: Exceptions, *Brain Lang.*, 26, 201–229.
27. Ferro, J. M., and Madureira, S. (1997), Aphasia type, age and cerebral infarct localisation, *J. Neurol.*, 244, 505–509.
28. Binder, J. (1997), Functional magnetic resonance imaging. Language mapping, *Neurosurg. Clin. N. Am.*, 8(3), 383–392.
29. Drobyshevsky, A., Baumann, S. B., and Schneide, W. (2006), A rapid fMRI task battery for mapping of visual, motor, cognitive, and emotional function, *Neuroimage*, 31(2), 732–744.
30. Gureje, O., Ogunniyi, A., Kola, L., et al. (2006), Functional disability in elderly Nigerians: Results from the Ibadan Study of Aging, *J. Am. Geriatr. Soc.*, 54(11), 1784–1789.
31. Chiu, H. F., and Zhang, M. (2000), Dementia research in China, *Int. J. Geriatr. Psychiatry*, 15(10), 947–953.

32. Ballard, C., Holmes, C., McKeith, I., et al. (1999), Psychiatric morbidity in dementia with Lewy bodies: A prospective clinical and neuropathological comparative study with Alzheimer's disease, *Am. J. Psychiatry*, 156, 1039–1045.
33. Ryu, S. H., Katona, C., Rive, B., et al. (2005), Persistence of and changes in neuropsychiatric symptoms in Alzheimer disease over 6 months: The LASER-AD study, *Am. J. Geriatr. Psychiatry*, 13, 976–983.
34. McShane, R., Keene, J., Gedling, K., et al. (1997), Do neuroleptic drugs hasten cognitive decline in dementia? Prospective study with necropsy follow up, *BMJ*, 314, 266–270.
35. Ballard, C., et al. (2005), Quetiapine and rivastigmine and cognitive decline in Alzheimer's disease: Randomised double blind placebo controlled trial, *BMJ*, 330(7496), 874.
36. Strangman, G., O'Neil-Pirozzi, T. M., Burke, D., et al. (2005), Functional neuroimaging and cognitive rehabilitation for people with traumatic brain injury, *Am. J. Phys. Med. Rehabil.*, 84, 62–75.
37. Schretlen, D. J., and Shapiro, A. M. (2003), A quantitative review of the effects of traumatic brain injury on cognitive functioning, *Int. Rev. Psychiatry*, 15, 341–349.
38. Lendrem, W., and Lincoln, N. B. (1985), Spontaneous recovery of language in patients with aphasia between 4 and 34 weeks after stroke, *J. Neurol. Neurosurg. Psychiatry*, 48, 743–748.
39. Jarvik, L. F., Berg, L., Bartus, R., et al. (1990), Clinical drug trials in Alzheimer disease. What are some of the issues? *Alzheimer Dis. Assoc. Disrod.*, 4, 193–202.
40. Rosen, W. G., Mohs, R. C., and Davis, K. L. (1984), A new rating scale for Alzheimer's disease, *Am. J. Psychiatr.*, 141, 1356–1364.
41. Matthews, H. P., Korbey, J., Wilkinson, D. G., et al. (2000), Donepezil in Alzheimer's disease: Eighteen month results from Southampton Memory Clinic, *Int. J. Geriatr. Psychiatry*, 5, 713–720.
42. Le Bars, P. L., Kieser, M., and Itil, K. Z. (2000), A 26-week analysis of a doubleblind, placebo-controlled trial of the ginkgo biloba extract EGb 761 in dementia, *Dement. Geriatr. Cogn. Disord.*, 11, 230–237.
43. Farlow, M., Potkin, S., Koumaras, B., et al. (2003), Analysis of outcome in retrieved dropout patients in a rivastigmine vs placebo, 26-week, Alzheimer disease trial, *Arch Neurol.*, 60, 843–848.
44. Aisen, P. S., Schafer, K. A., Grundman, M., et al. (2003), Effects of rofecoxib or naproxen vs placebo on Alzheimer disease progression: A randomized controlled trial, *JAMA*, 289, 2819–2826.
45. Dubois, B., Slachevsky, A., Litvan, I., et al. (2000), The FAB: A Frontal Assessment Battery at bedside, *Neurology*, 55(11), 1621–1626.
46. Qualls, C. E., Bliwise, N. G., and Stringer, A. Y. (2000), Short forms of the Benton Judgment of Line Orientation Test: Development and psychometric properties, *Arch. Clin. Neuropsychol.*, 15(2), 159–163.
47. Ferman, T. J., Smith, G. E., Boeve, B. F., et al. (2004), DLB fluctuations: Specific features that reliably differentiate DLB from AD and normal aging, *Neurology*, 62, 181–187.
48. Folstein, M. F., Folstein, S. E., and McHugh, P. R. (1975), 'Mini-mental state'. A practical method for grading the cognitive state of patients for the clinician, *J. Psychiatr. Res.*, 12, 189–198.
49. Caban-Holt, A., Bottiggi, K., and Schmitt, F. A. (2005), Measuring treatment response in Alzheimer's disease clinical trials, *Geriatrics*, Jun (Suppl):3–8.

50. Saxton, J., McGonigle Gibson, K. L., et al. (1990), Assestment of the severely impaired patient: Decription and validation of a new neuropsychological test battery, *Psychol. Assess.*, 2, 298–303.
51. Panisset, M., Roudier, M., Saxton, J., et al. (1994), Severe impairment battery: A neuropsychological test for severely demented patients, *Arch. Neurol.*, 51, 41–45.
52. Schmitt, F. A., Ashford, W., Ernesto, C., et al. (1997), The severe impairment battery: Concurrent validity and the assessment of longitudinal change in Alzheimer's disease, *Alzheimer Dis. Assoc. Disord.*, 11(Suppl 2), S51–S56.
53. Reisberg, B., Doody, R., Stöffler, A., et al. (2003), Memantine Study Group. Memantine in moderate-to-severe Alzheimer's disease, *N. Engl. J. Med.*, 348(14), 1333–1341.
54. Schmitt, F. A., Cragar, D., Ashford, J. W., et al. (2002), Measuring cognition in advanced Alzheimer's disease for clinical trials, *J. Neural. Transm. Suppl.*, 62, 135–148.
55. Saxton, J., Kastango, K. B., Hugonot-Diener, L., et al. (2005), Development of a short form of the Severe Impairment Battery, *Am. J. Geriatr. Psychiatry*, 13(11), 999–1005.
56. Wolinsky, F. D., Callahan, C. M., Fitzgerald, J. F., et al. (1992), The risk of nursing home placement and subsequent death among older adults, *J. Gerontol: Social Sci.*, 47, S173–S182.
57. Katz, S., Ford, A. B., Moskowitz, R. W., et al. (1963), The index of ADL: A standardized measure of biological and psychosocial function, *JAMA*, 185, 914–919.
58. Lawton, M. P., and Brody, E. M. (1969), Assessment of older people: Self-maintaining and instrumental activities of daily living, *Gerontologist*, 9, 179–186.
59. Stern, Y., Hesdorrfer, D., Sano, M., et al. (1990), Measurement and prediction of functional capacity in Alzheimer's disease, *Neurology*, 40, 8–14.
60. Spector, W. D. (1997), Measuring functioning in daily activities for persons with dementia, *Alzheimer Dis. Assoc. Disord.*, 11(Suppl 6), 81–90.
61. Blessed, G., Tomlinson, B. E., and Roth, M. (1968), The association between quantitative measures of dementia and of senile change in the cerebral gray matter of elderly subjects, *Br. J. Psychiatry*, 114, 797–811.
62. Spiegel, R., Brunner, C., Ermini-Fünfschilling, D., et al. (1991), A new behavioral assessment scale for geriatric out and in-patients the NOSGER (Nurses' Observation Scale for Geriatric Patients), *J. Am. Geriatr. Soc.*, 39, 339–347.
63. Reisberg, B., Finkel, S., Overall, J., et al. (2001), The Alzheimer's Disease Activities of Daily Living International Scale (ADL-IS), *Int. Psychogeriatr.*, 13, 163–181.
64. Rubenstein, L. Z., Schairer, C., Wieland, G. D., et al. (1984), Systematic biases in functional status assessment of elderly adults: Effects of different data sources, *J. Gerontol.*, 39, 686–691.
65. Kuriansky, J., and Gurland, B. (1976), The performance test of activities of daily living, *Int. J. Aging Hum. Dev.*, 7, 343–352.
66. Jette, A. M., and Branch, L. G. (1985), Impairment and disability in the aged, *J. Chronic Dis.*, 38, 59–65.
67. Williams, M. E. (1987), Identifying the older person likely to require long-term care services, *J. Am. Geriatr. Soc.*, 35, 761–766.
68. Guralnik, J. M., Branch, L. G., Cummings, S. R., et al. (1989), Physical performance measures in aging research, *J. Gerontol.*, 44, 141–146.
69. Reuben, D. B., Siu, A. L., and Kimpau, S. (1992), The predictive validity of self-report and performance-based measures of function and health, *J. Gerontol.*, 47, M106–110.
70. Zimmerman, S. I., and Magaziner, J. (1994), Methodological issues in measuring functional status of cognitively impaired nursing home residents: The use of proxies

and performance based measures, *Alzheimer Dis. Assoc. Disord.*, 8(Suppl 1), S281–S290.

71. Lowenstein, D. A., Amigo, E., Duara, R., et al. (1989), A new scale for the assessment of functional status in Alzheimer's disease and related disorders, *J. Gerontol.*, 44, 114–121.
72. Reuben, D. B., and Siu, A. L. (1990), An objective measure of physical function of elderly outpatients. The Physical Performance Test, *J. Am. Geriatr. Soc.*, 38, 1105–1112.
73. Mahurin, R. K., De Bettignies, B. H., and Pirozzolo, F. J. (1991), Structured assessment of independent living skills: preliminary report of a performance measure of functional abilities in dementia, *J. Gerontol.*, 46, 58–66.
74. Skurla, E., Rogers, J. C., and Sunderland, T. (1988), Direct assessment of activities of daily living in Alzheimer's disease, *J. Am. Geriatr. Soc.*, 36, 97–103.
75. Karagiozis, H., Gray, S., Sacco, J., et al. (1998), The Direct Assessment of Functional Abilities (DAFA): A comparison to an indirect measure of instrumental activities of daily living, *Gerontologist*, 38, 113–121.
76. Willis, S. L., Allen-Burge, R., Dolan, M. M., et al. (1998), Everyday problem solving among individuals with Alzheimer's disease, *Gerontologist*, 38, 569–577.
77. Pfeffer, R. I., Kurosaki, T. T., Harrah, C. H., et al. (1982), Measurement of functional activities in older adults in the community, *J. Gerontol.* 37, 323–329.
78. Farina, E., Fioravanti, R., Chiavari, L., et al. (1999), Functional Living Skills Assessment: A standardized instrument built to monitor activities of daily living in patients with dementia: Preliminary data, *Journal of Neurology*, 246(suppl. 1), I/101.
79. Reisberg, B. (2007), Global measures: utility in defining and measuring treatment response in dementia, *Int. Psychogeriatr.*, 19, 421–456.
80. Hughes, C. P., and Berg, L. (1982), A new clinical scale for the staging of dementia, *Br. J. Psychiatry*, 140, 566–572.
81. Reisberg, B., Ferris, S. H., deLeon, M. J., et al. (1982), The Global Deterioration Scale for assessment of primary degenerative dementia, *Am. J. Psychiatry,* 139, 1136–1139.
82. Sclan, S. G., and Reisberg, B. (1992), Functional assessments staging (FAST) in Alzheimer's disease: Reliability, validility, and ordinality, *Int. Psichogeriatr.*, 4(Suppl. 1), 55–69.
83. Knopman, D. S., Knapp, M. J., Gracon, S. I., et al. (1994), The Clinician Interview-Based Impression (CIBI): A clinician's global change rating scale in Alzheimer's disease, *Neurology*, 44, 2315–2321.
84. Schneider, L. S., Olin, J. T., Doody, R. S., et al. (1997), Validity and reliability of the Alzheimer's Disease Cooperative Study—Clinical Global Impression of Change. The Alzheimer's Disease Cooperative Study, *Alzheimer Dis. Assoc. Disord.*, 11(Suppl 2), S22–S32.
85. Reisberg, B., Schneider, L., Doody, R., et al. (1997), Clinical global measures of dementia. Position paper of the International Wotking Group on Harmonization of Dementia Drug Guidelines, *Alzheimer Dis. Assoc. Disord.*, 11(Suppl 3), 8–18.
86. Quinn, J., Moore, M., Benson, D. F., et al. (2002), A videotaped CIBIC for dementia patients: Validity and reliability in a simulated clinical trial, *Neurology*, 58, 433–437.
87. Joyce, C. R. B., O'Boyle, C. A., and McGee, H. (1999), *Individual Quality of Life. Approaches to Conceptualisation and Assessment*, Harwood Academic, Singapore.
88. Rockwood, K., and Gauthier, S., eds. (2006), *Trial Designs and Outcomes in Dementia Therapeutic Research*, Taylor & Francis, London and New York.

89. Katona, C., Livingston, G., Cooper, C., et al. (2007), International Psychogeriatric Association consensus statement on defining and measuring treatment benefits in dementia, *Int. Psychogeriatr.*, 19(3), 345–354.
90. Birks, J. (2006), Cholinesterase inhibitors for Alzheimer's disease, *Cochrane Database Syst. Rev.*, CD005593.
91. Scholzel-Dorenbos, C., Van der Steen, M., Engels, L., et al. (2007), Assessment of quality of life as outcome in dementia and MCI intervention trials, *Alzheimer Dis. Assoc. Disord.*, 21, 172–178.
92. Bowling, A. (1993), in Keynes, M., Ed., *Measuring Health: A Review of Quality of Life Measurement Scales*, Open University Press, Philadelphia.
93. Spilker, B. (1990), *Quality of Life Assessment in Clinical Trials*, Raven Press, New York.
94. Naglie, G. (2007), Quality of life in dementia, *Can. J. Neurol. Sci.*, 34(Suppl 1), S57–61.
95. Wiklund, I. (2004), Assessment of patient-reported outcomes in clinical trials: The example of health-related quality of life, *Fundam. Clin. Pharmacol.*, 18, 351–363.
96. Brod, M., Stewart, A. L., Sands, L., et al. (1999), Conceptualization and measurement of quality of life in dementia: The dementia quality of life instrument, *Gerontologist*, 39, 25–35.
97. Logsdon, R. G., Gibbons, L. E., McCurry, S. M., et al. (2002), Assessing quality of life in older adults with cognitive impairment, *Psychosom. Med.*, 64, 510–519.
98. Scocco, P., Fantoni, G., and Caon, F. (2006), Role of depressive and cognitive status in self-reported evaluation of quality of life in older people: Comparing proxy and physician perspectives, *Age Ageing*, 35, 166–171.
99. Lyketsos, C. G., Breitner, S., and Rabins, P. V. (2001), An evidence-based proposal for the classification of neuropsychiatric disturbance in Alzheimer's disease, *Int. J. Geriatr. Psychiatry*, 16, 1037–1042.
100. Cummings, J. L., Mega, M., Gray, K., et al. (1994), The Neuropsychiatric Inventory: Comprehensive assesment of psychopathology in dementia, *Neurology*, 44, 2308–2314.
101. Reisberg, B., et al. (1988), BEHAVE-AD. A clinical rating scale for the assessment of pharmacologically remediable behavioral symptomatology in Alzheimer's disease, in Altman, H., (Ed.), *Alzheimer's Disease and Dementia: Problems, Prospects, and Perspectives*, Plenum, New York.
102. Conn, T. (2007), Assessment of behavioral and psychological symptoms associated with dementia, *Can. J. Neurol. Sci.*, 34(Suppl 1), S67–71. Review.
103. Kaufer, D. I., Cummings, J. L., Christine, D., et al. (1998), Assessing the impact of neuropsychiatric symptoms in Alzheimer's disease: The Neuropsychiatric Inventory Caregiver Distress Scale, *J. Am. Geriatr. Soc.*, 46, 210–215.
104. Novak, M., and Guest, C. (1980), Application of a multidimensional caregiver burden inventory, *Gerontologist*, 29, 798–780.
105. Vitaliano, P. P., Russo, J., and Young, H. M. (1991), The screen for caregiver burden, *Gerontologist*, 31, 76–83.
106. Pearlin, L. I., Mullan, J. T., Semple, S. J., et al. (1990), Caregiving and the stress process: An overview of concepts and their measures, *Gerontologist*, 30, 583–594.
107. Greene, J. G., Smith, R., and Timbury, G. C. (1982), Measuring behavioral disturbances of elderly demented patients in the community and its effects on relatives: A factor analytic study, *Age Aging*, 11, 121–126.
108. Poulshock, S. W., and Deimling, G. T. (1984), Families caring for elders in residence. Issues in the measurement of burden, *J. Gerontol.* 39, 230–239.

109. Blesa, R. (2000), Galantamine: Therapeutic effects beyond cognition, *Dement. Geriatr. Cogn. Disord.*, 11(Suppl 1), 28–34.
110. Clipp, E. C., and Moore, M. J. (1995), Caregiver time use. An outcome measure in clinical trial research on Alzheimer's disease, *Clin. Pharmacol. Ther.*, 58, 228–236.
111. Mohs, R. C., Doody, R. S., Morris, J. L., et al. (2001), A 1-year placebo-controlled preservation of function survival study of donepezil in AD patients, *Neurology*, 57, 481–488.
112. Wimo, A., Wetterholm, A. L., Mastey, V., et al. (1988), Evaluation of the resource utilization and caregiver time in anti-dementia drug trials: A quantitative battery, in Wimo, A., Jonsson, B., Karlsson, G., Eds., *The Health Economics of Dementia*, Wiley, London. pp. 465–499.
113. Lingler, J. H., Martire, L. M., and Schulz, R. (2005), Caregiver-specific outcomes in antidementia clinical drug trials: A systematic review and meta-analysis, *J. Am. Geriatr. Soc.*, 53(6), 983–990.
114. Wright, J. (2000), The Functional Assessment Measure. The Center for Outcome Measurement in Brain Injury.
115. Hagen, C., Malkmus, D., and Durham, P. (1972), Rancho Los Amigos cognitive scale. Rancho Los Amigos Hospital.
116. Blumstein, S. E. (1997), A perspective on the neurobiology of language, *Brain Lang.*, 60, 335–346.
117. Grodzinsky, Y., Pinango, M. M., Zurif, E., et al. (1999), The critical role of group studies in neuropsychology: Comprehension regularities in Broca's aphasia, *Brain Lang.*, 67, 134–147.
118. Goodglass, H., and Kaplan, E. (1972), *Assessment of Aphasia and Related Disorders*, Lea and Febiger, Philadelphia.
119. Kertesz, A. (1982), *The Western Aphasia Battery*, Grune and Stratton, New York.
120. Porch, B. (1982), *The Porch Index of Communicative Abilities*, Consulting Psychologists Press, Palo Alto, CA.
121. Walker-Batson, D., Curtis, S., Natarajan, R., et al. (2001), A double-blind, placebo-controlled study of the use of amphetamine in the treatment of aphasia, *Stroke*, 32, 2093–2098.
122. Berthier, M. L., Hinojosa, J., Martin, M. C., et al. (2003), Open-label study of donepezil in chronic poststroke aphasia, *Neurology*, 1218–1219.
123. Kay, J., Lesser, R., and Colthearth, M. (1992), *Psycholinguistic Assessments of Language Processing in Aphasia (PALPA)*, Lawrence Erlbaum, Hove.
124. Nickels, L. (2002), Theoretical and methodological issues in the cognitive neuropsychology of spoken word production, *Aphasiology*, 16, 3–19.
125. Hilari, K., Byng, S., Lamping, D. L., et al. (2003), Stroke and aphasia quality of life scale-39 (SAQL-39): evaluation of acceptability, reliability and validity, *Stroke* 34, 1944–1950.
126. Cortes, F., Portet, F., Touchon, J., et al. (2007), Six and 18-month changes in mild to moderate Alzheimer's patients treated with acetylcholinesterase inhibitors: What can we learn for clinical outcomes of therapeutic trials? *J. Nutr. Health Aging*, 11(4), 330–337.

# *Bridging Studies in Pharmaceutical Safety Assessment*

**Jon Ruckle**

*Covance Clinical Research Unit, Honolulu, Hawaii*

**Contents**

1 Introduction
  1.1 Are Ethnic Factors Real?
  1.2 What Are "Bridging Studies"?
2 Why Do Bridging Studies?
  2.1 Traditional Practice of Medicine in Japan and the United States
  2.2 Pharmaceutical Development in Japan
  2.3 Safety versus Efficacy
  2.4 Market Factors
  2.5 Social and Political Factors
3 Regulatory Guidance: ICH and E5
  3.1 ICH History
  3.2 E5 Guidelines
  3.3 Implications for Non-ICH Jurisdictions
4 Ethnic Pharmaceutical Safety Issues: Pharmacokinetics
  4.1 ADME: Absorption
  4.2 ADME: Distribution
  4.3 ADME: Metabolism
  4.4 ADME: Elimination
  4.5 ADME: Dietary Factors
  4.6 ADME: Multiple Mechanisms

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

5 Ethnic Pharmaceutical Safety and Efficacy Issues: Pharmacodynamics
6 Bridging Study Design
6.1 Candidate Compound
6.2 Basic Study Design
6.3 Subject Population
6.4 Diet and Food Effects
6.5 Restrictions
7 Bridging Study Conduct
7.1 Recruitment
7.2 Informed Consent
7.3 GCP Practices
7.4 Adverse Events
7.5 Documentation
7.6 Monitoring and Audits
8 Bridging Study Experience
8.1 Examples of Success
8.2 Challenges
9 Summary and Future Directions
References

## 1 INTRODUCTION

Globalization increasingly characterizes the pharmaceutical industry. We see mergers and acquisitions within industry, rapid international information sharing, and emerging centers for drug discovery. With this, we see more urgency for international regulatory approval and marketing of pharmaceutical compounds. However, regulatory approval in a new region is generally not a simple matter of reciprocity. Practitioners and regulatory agencies have long been concerned with ethnic differences that may affect the optimal use of a medication in a new region, even if it is well understood in the original region. Agencies have understandable concerns about defining the indication, dose selection, product labeling, safety, tolerability, and monitoring of the medication in the new region. "Bridging studies" are one important mechanism for addressing these questions.

### 1.1 Are Ethnic Factors Real?

As an American working in Honolulu, Hawaii, arguably one of the world's major cross roads for Eastern and Western cultural exchange, I have had the opportunity to participate in discussions regarding clinical pharmacology development programs with people worldwide. In these discussions, I hear two distinct schools of thought regarding "ethnic differences."

One school maintains that ethnic differences are real and important considerations in evaluating new medications. For example, some see individuals composing Japanese society as unique and different from those in other societies, thus foreign data must be viewed with great caution. Support for this view comes from review of the historic and cultural differences between Japanese and Western societies, including differences in diet, average weight, expected life span, social homogeneity, and epidemiological patterns of disease. In addition, Japanese practitioners typically

utilized considerably lower doses of medications than their Western colleagues, even when using the same medication for the same indication. Data from controlled studies illustrating substantially different pharmacokinetic profiles in Japanese versus Caucasians for certain compounds, and differences in adverse event profiles in Japanese versus Caucasians for certain medications are cited to add scientific support to this perspective.

The other school asserts that there is no such thing as "ethnic differences." This school maintains that "under the skin" all the world's peoples are essentially the same, and that apparent ethnic differences can be explained on the basis of identifiable variables or individual differences. For example, Japanese have traditionally been smaller and lighter than Westerners and many apparent pharmacokinetic ethnic differences disappear once the profiles are adjusted for weight or percent body fat. Other differences can be largely explained by identifiable factors such as heterogeneity in hepatic cytochromes or other metabolizing enzymes, or known food effects and nutritional/dietary differences. This view is supported by data from controlled trials showing that individual intraethnic differences in pharmacokinetic profiles for many compounds are often greater than interethnic differences.

Which school is correct? In this chapter, we will discuss these issues and then revisit this question.

## 1.2 What Are "Bridging Studies"?

When considering two regions with ethnic differences, bridging studies are performed to facilitate approval of a medication in the new region after it has already been studied or approved in another region. A bridging study is defined as: "A supplemental study performed in the new region to provide pharmacodynamic or clinical data on efficacy, safety, dosage and dose regimen in the new region that will allow extrapolation of the foreign clinical data to the new region. Such studies could include additional pharmacokinetic information" [1].

The concept of a "bridge" is a useful metaphor. Physical bridges have long enabled people to travel and share goods and services across rivers and other natural barriers. Similarly, a bridging study provides data to members of the new jurisdiction, enabling them to interpret and apply data from the original jurisdiction. Especially if the results allow the new region to believe the drug will behave similarly in both regions, or at least reasonably similarly with a dose adjustment, the study creates a bridge allowing data from the first region to be used to support approval and labeling in the new region.

In their simplest form, bridging studies are clinical pharmacology studies done in healthy individuals or patients to compare the pharmacokinetic (PK) profile, safety and tolerability of single or multiple doses in subjects representing the new region, compared with subjects representing the original region. Such studies may be performed in either or both regions and are usually conducted in a relatively small number of participants, who may be "healthy" and "normal." These studies are generally modeled after phase I studies performed in the United States and primarily determine if a dose adjustment is appropriate in the new region.

In the larger context, additional studies performed in the new region confirming pharmacokinetics and/or the efficacy, safety, and tolerability are also considered bridging studies if their design is based on data from foreign studies and their

purpose is to confirm or clarify the drug's behavior in the new population compared with the reference population. These studies are generally performed within the new region and enroll the target patient population. Their design is similar to phase II or phase III studies performed in the United States.

## 2 WHY DO BRIDGING STUDIES?

From a business standpoint, it is about saving time and money. The bridge allows a large body of existing safety and efficacy data to be used that would otherwise have to be re-created, allowing pivotal efficacy and safety studies in the new region to be smaller and shorter than they would be otherwise.

The key variable is the dose. Should the dose in the new region be the same, higher, or lower? Coupled closely with identifying the appropriate dose, product labeling and safety monitoring may be different in a new region, reflecting both the pharmacodynamic (PD) effects, definition of disease and identification of the target population, and the priorities and medical practices of the new region.

In this section, we will be mainly discussing the United States, the European Union (EU), and Japan, with primary attention to medications studied or approved in the United States and the European Union being considered for approval in Japan. To better understand the background for bridging studies, it may be helpful to review the cultural context for medical practice in Japan and the traditional Japanese approach to pharmacological research.

### 2.1 Traditional Practice of Medicine in Japan and the United States

The practice of medicine in Japan has traditionally been quite different from that in the United States. Medical practice in the United States has long incorporated a fee for service system with multiple third-party payers, a competitive marketplace for drug pricing, and a physician–patient relationship based on considerable patient autonomy and expectation of information, second opinions, and informed consent.

Not so in Japan. Rihito Kimura, a Japanese bioethicist, summarizes as follows:

> In the long tradition of Japanese medical practice, the Confucian notion of *jin* (benevolence) has been one of the most important ethical elements; medicine itself is known as *jinjyutsu* (the art of *jin*). Physicians, as conduits of *jin*, were required to act with benevolonce toward their patients, and were responsible for the welfare of patients in a fiduciary (trust) relationship (Kimura, 1991a). It was obligatory to use medicine, a gift of benevolence, for the good of others even without payment. Physicians fulfilled their responsibility toward their patients and the patients' family members by acting in a paternalistic and authoritative way; the Japanese, nurtured in the Confucian ethos to respect law, order, authority, and social status, acquiesced without murmur to the superior knowledge of the physician. [2]

Kimura further notes:

> It is important to note that owing to the character of Japanese society and its distinctive historical understanding of medicine and the role and responsibilities of the physician, it was not until the 1960s that the bioethical and sociolegal concerns about the practice

> of medicine began to be deliberately reflected in Japanese society, and only during the 1980s that the notions of autonomy and rights in medicine, and of bioethics in general, became gradually influential. [2]

In addition to these cultural factors and relationship patterns, Japanese medical practice also often incorporated acupuncture and herbal medications. A universal health insurance covered most medical services and medication prices were standardized centrally.

### 2.2 Pharmaceutical Development in Japan

The traditional approach to pharmaceutical development in Japan reflected their cultural and dynamics of health care delivery in general. Trials were generally performed under the direction of a chief investigator who wrote the protocol and reviewed the data. It was not until 1980 that the Pharmaceutical Affairs Law (PAL) included an explicit clause regulating clinical trials, including the sponsor's obligation to submit a clinical trial plan, and the minister's authority to instruct sponsors to modify the trials to avoid possible health hazards.

Even still, investigators were allowed to enroll subjects with oral rather than written consent, the discussion about possible adverse events was generally abbreviated (by Western standards), protocols were not routinely reviewed by independent ethics review boards, sponsors and auditors were not allowed access to the original source documents to establish data validity, accountability for investigational products at the trial site was informal, and reporting of adverse events was essentially up to the investigator. Studies were usually done without much if any support from clinical research coordinators, research nurses, and data management staff.

These practices changed in 1998 when the Japanese regulations were revised to meet the standards of the International Conference on Harmonisation (ICH) on good clinical practices (GCP) regulations, reflecting pharmaceutical development practices in Europe and the United States.

For cultural reasons, incorporation of the new regulations was particularly difficult regarding obtaining written informed consent, and enrollment of subjects decreased dramatically for several years. Japanese people were not accustomed to making a written contract for their daily activities, especially regarding health care. Physicians were accustomed to making patient care decisions, and many patients were more comfortable letting physicians do whatever they thought was right rather than being informed fully and being asked to make a decision themselves. In particular, Japanese people are sensitive about possible side effects of drugs. A full description of the major undesirable possibilities is not usual in daily medical practice in Japan, thus the tendency is to refrain from study participation after seeing these in written form, even though the risk is small.

Obtaining informed consent presented a dilemma for Japanese researchers. Although they recognized the importance, obtaining consent is not common in daily medical practice and contradicts the traditional doctor–patient relationships deeply rooted in the culture.

Furthermore, study participation often lacked incentives. With universal health insurance coverage, study participation did not necessarily provide services otherwise unavailable and instead required more frequent visits, procedures, and

observations than normal treatment. Stipend payments for study participation have only been recently introduced. In general, there were negative perceptions regarding study participation, rather than seeing study participation as an opportunity to contribute to society and thus an honorable duty, such as donating blood.

### 2.3 Safety Versus Efficacy

Both safety and efficacy are important everywhere. But traditionally, there is a different cultural emphasis on safety versus efficacy in Japan compared with the United States. While Americans want safe medications, the cultural bias also insists "give me something that works," with high interest in efficacy. If there are potential adverse effects, most individuals are willing to weigh these risks (once properly advised) against the expected benefit.

In Japan, it is almost the opposite. While they want effective medications, both providers and patients highly desire to avoid anything that could make them worse. As a cultural practice, they are generally willing to use low doses or less efficacious medications if such will also offer a safer approach.

### 2.4 Market Factors

Of necessity, economic considerations are major factors in drug development, particularly regarding bridging studies. The largest market for pharmaceutical products is the United States, followed by the European Union (when considered collectively), then Japan, with Japan as second largest when considering individual nations. To illustrate, in 2002, the global market for sales of pharmaceutical products was approximately $401 billion. Sales are markedly disproportional to population. U.S. sales accounted for over 50% of the world's pharmaceutical sales, of about $204 billion, and continues to grow at 12% per year. The European Union followed the United States with combined sales of about $91 billion, with Japan at $47 billion. Thus these three jurisdictions account for less than 15% of the world's population, yet about 85% of the world's pharmaceutical sales.

Active drug discovery is ongoing in all three jurisdictions, and all three have and continue to produce important compounds. Historically, the European Union invested the most in pharmaceutical research and development, a leadership role that continued through 1990. Growth in the U.S. pharmaceutical industry was such that U.S. investment exceeded Europe in the early 1990s and has continued to grow rapidly. In 2004 research input by U.S. pharmaceuticals nearly matched the investments of Europe and Japan combined.

Each jurisdiction has its own approach to patent protection, marketing, distribution, and pricing. For example, U.S. pharmaceutical companies generally base pricing on market considerations. In Japan, prices have been established under the oversight of the Ministry of Health, Labor, and Welfare (MHLW), with downward adjustments applied, on the average, every 2 years. Even considering these differences, if a pharmaceutical company has a product that is commercially successful in one jurisdiction, there is usually sufficient potential for financial gain to interest industry in offering the compound in the other jurisdictions.

Market factors interface with political structures and regulatory practices and affect product development. During the 1980s and 1990s, it took a U.S.

pharmaceutical company 10 years, on average, to develop and launch a new product, whereas it required 15–17 years for the Japanese counterparts. Partially for this reason, of the 178 new chemical entities launched worldwide between 1999 and 2003, just 47 were launched in Japan. On the other hand, drugs discovered in Japan tended to be rapidly acquired and successfully launched by Western companies. Examples include Pravachol, Crestor, and Aricept, discovered, respectively, by Sankyo, Shionogi, and Eisai laboratories yet marketed outside of Japan by Bristol Meyers Squibb, AstraZeneca, and Pfizer.

Market factors are mentioned here because they are such powerful considerations, although an adequate discussion of this topic is beyond the scope of this chapter.

### 2.5 Social and Political Factors

Collective social opinion is a significant factor in the pharmaceutical industry, influencing regulation, pricing strategies, and standards for determining what is "generally safe and well tolerated." Social influences are particularly important in bridging studies and reflect a continuation of the usual perspectives of one's own culture and the natural comfort of seeing and doing things in ways that are familiar. It takes substantial effort to understand perspectives and traditions that are different from your own. As will be discussed in more detail below, cultural factors are significant considerations in the practice of medicine in general and the role of pharmaceuticals in particular. Thus it is understandable that the regulatory agencies in each jurisdiction will operate with primary concern for their own population, in ways which generally represent local medical and business practices. These concerns and practices vary from one country to another. Successful global pharmaceutical development strategies require appreciation for these political factors.

The agencies in each jurisdiction and the approval process for new compounds operate within this milieu and change over time. The milestones for the U.S. Food and Drug Administration (FDA) include a long list of laws and regulations. Historically, most were "political" responses to problems within the industry [3]. Recently, the emphasis has been on increased efficiency.

Historical development in Japan was somewhat similar, with a recent emphasis on greater efficiency. Japan introduced a revised Pharmaceutical Affairs Law (PAL) in 2002, with compliance effective by 2005, with organizational and regulatory changes in product registration, industry standards, bridging studies, clinical trials labeling, advertising, product classification, and intellectual property. Restructuring under the revised PAL includes the creation of the Pharmaceutical and Medical Devices Agency (PMDA), combining the Organization for Pharmaceutical Safety and Research (OPSR), the Pharmaceutical and Medical Device Center (PMDEC), and the Japan Association for the Advancement of Medical Equipment (JAAME). The purpose of the PMDA is to expedite what had historically been a slow review process. The PMDA revised systems and regulations involving the import, marketing, and contract manufacturing requirements of non-Japanese pharmaceutical companies and signaled an intent to gradually increase the acceptance of new drug application (NDA) submissions featuring bridging studies.

Political factors and the structure of each jurisdiction's regulatory agencies are also topics beyond the scope of this chapter. They are introduced here out of

recognition for their significance and as a transition to the discussion of regulatory guidance.

## 3 REGULATORY GUIDANCE: ICH AND E5

Bridging studies would not be possible without consensus regulatory guidance to allow data from studies done in one jurisdiction to be accepted in another. Consensus at an appropriate level of detail on issues this important is not easily achieved. Despite recognition of the substantial benefits to all concerned, consensus guidelines required diligent effort for nearly a decade before they were agreed, adopted, and utilized.

The regulatory guidance currently utilized is the effort of the ICH. The history is briefly summarized below, with attention to the topic (E5) describing use of foreign data.

### 3.1 ICH History

The importance of independent evaluation of medicinal products before they are allowed on the market was realized at different times in different regions. In the United States, the role of the FDA in reviewing medications was markedly increased following a tragic mistake in the formulation of a sulfanilamide syrup in 1937. Japan began requiring all medicinal products to be registered for sale starting in the 1950s. Response to the thalidomide tragedy of the 1960s expanded regulatory oversight in Europe.

The different regulatory systems were based on the same fundamental obligations to evaluate the quality, safety, and efficacy of the pharmaceutical products, yet the detailed technical requirements diverged over time to such an extent that industry found it necessary to duplicate many time-consuming and expensive test procedures in order to market new products internationally.

By the late 1980s, the urgent need to rationalize and harmonize regulation in an increasingly global marketplace was impelled by concerns over the rising costs of health care, including pharmaceutical R&D, in juxtaposition to the need to meet the public expectation that there should be a minimum of delay in making safe and efficacious new treatments available to patients in need.

Harmonization of regulatory requirements was pioneered by the European Community, in the 1980s, as the European Community (now the European Union) moved toward the development of a single market for pharmaceuticals. The success achieved in Europe demonstrated that harmonization was feasible. At the same time there were bilateral discussions between Europe, Japan, and the United States on possibilities for harmonization.

The birth of the ICH took place at a meeting in April 1990. Representatives of the regulatory agencies and industry associations of Europe, Japan, and the United States met, primarily, to plan an international conference. The ICH Steering Committee, which was established at that meeting, has since met at least twice a year, with the location rotating between the three regions. ICH work groups have met frequently, and six ICH meetings have been held, most recently in November 2003 in Osaka, Japan. Although the scope of the ICH topics has expanded over time, the

fundamental goals are still generally consistent with the early goals, as expressed in a statement by the ICH Steering Committee from October, 1990:

> The Parties cosponsoring this Conference, represented at the 2nd Steering Committee Meeting in Tokyo, 23–24 October 1990 re-affirmed their commitment to increased international harmonization, aimed at ensuring that good quality, safe and effective medicines are developed and registered in the most efficient and cost-effective manner. These activities are pursued in the interest of the consumer and public health, to prevent unnecessary duplication of clinical trials in humans and to minimize the use of animal testing without compromising the regulatory obligations of safety and effectiveness.

ICH Guidelines cover a variety of topics, generally arranged as:

- Quality Topics (e.g., Q7, on Good Manufacturing Practices and Q8 on Pharmaceutical Development)
- Safety Topics (e.g., S2 discussing genotoxicity studies, and S5 addressing reproductive toxicity)
- Efficacy Topics relating to studies in human subjects (e.g., E6 on good clinical practices)
- Multidisciplinary Topics, which do not fit uniquely into the above categories (e.g., M1: Medical Terminology).

### 3.2 E5 Guidelines

Of special relevance for bridging studies, ICH Efficacy Topic E5 discusses Ethnic Factors in the Acceptability of Foreign Clinical Data. The current document was recommended for adoption by the ICH Steering Committee on February 5, 1998, approved by the EU in March, 1998, published in the U.S. *Federal Register* in June 1998, and adopted by the MHLW in Japan in August 1998 [1].

Early experience in all ICH regions indicated the need for some clarification, and a Questions and Answers document was issued in November 2003. This was adopted by the European Union in November 2003, by Japan in February 2004, and by the United States in June 2004. The bridging studies discussed in this chapter were conducted within the regulatory environmental matrix described in these two documents. Although classified by ICH as an "efficacy" topic, the E5 guidelines are equally concerned with issues of patient safety.

The perspective of this author is that E5 comprises a particularly key component of ICH, providing the essential mechanism for the success of the ICH effort in achieving its goals. All topics are important. However, if data cannot be shared from one jurisdiction to the other, the pharmaceutical sponsors are essentially back where they started: duplicating the investigational program in a new region.

***E5 Objectives*** The objectives of E5 are:

- To describe the characteristics of foreign clinical data that will facilitate their extrapolation to different populations and support their acceptance as a basis for registration of a medicine in a new region.

- To describe regulatory strategies that minimize duplication of clinical data and facilitate acceptance of foreign clinical data in the new region.
- To describe the use of bridging studies, when necessary, to allow extrapolation of foreign clinical data to a new region.
- To describe development strategies capable of characterizing ethnic factor influences on safety, efficacy, dosage, and dose regimen.

The guidelines further state "All regions acknowledge the desirability of utilizing foreign clinical data. ... However, concern that ethnic differences may affect the medication's safety, efficacy, dosage and dose regimen in the new region has limited the willingness to rely on foreign clinical data," and thus explains the reason why there was so much duplication of the research program. The guidelines are based on the assumption that many medicines will have comparable characteristics and effects across regions, although there is acknowledgment that ethnic differences may affect the safety, efficacy, dosage, or dose regimen.

E5 "is based on the premise that it is not necessary to repeat the entire clinical drug development program in the new region." It is also "not intended to alter the data requirements for registration in the new region; it seeks to recommend when these data requirements may be satisfied with foreign clinical data," noting that "additional studies conducted in any region may be required by the new region to complete the clinical data package."

***Regional Requirements*** The three ICH regions have developed a basic consensus on the fundamentals of appropriate study conduct, summarized as good clinical practices (E6). Still, the regions vary in what constitutes an approvable clinical data package. At a minimum, studies need to provide data that contain adequate characterization of pharmacokinetics, pharmacodynamics, dose response, efficacy, and safety. Once these are established, further clinical trials should be performed according to GCP with appropriate choice of controls and endpoints using medical and diagnostic definitions acceptable to the new region. In particular, the pharmacodynamic endpoints for safety, tolerability, and efficacy must be performed either in the new foreign region or in a population representative of the new region.

Several ICH guidelines address these aspects of program design and provide guidance for a complete clinical data package. These include GCPs (E6), evaluation of dose–response (E4), adequacy of safety data (E1 and E2), conduct of studies in the elderly (E7), conduct of studies in a pediatric population (E11), reporting of study results (E3), general considerations for clinical trials (E8), statistical considerations (E9), and choice of control groups (E10). Guidelines are also emerging regarding specific therapeutic areas (e.g., hypertension, E12A) and clinical safety issues (e.g., QT/QTc interval prolongation for non-antiarrhythmic drugs, E14).

***Ethnic Factors*** Within E5, "ethnic factors are factors relating to races or large populations grouped according to common traits and customs." The guidance specifically notes that ethnic factors go beyond racial factors per se. The focus of attention for ethnic factors includes diet and lifestyle considerations with implications for the expected safety and efficacy of the medication in the region. For example, the Japanese authorities are more concerned with the effects of a medication within

Japan than the effects of the medication when taken by Japanese individuals within the United States.

Ethnic factors are broadly classified as Extrinsic and Intrinsic.

*Extrinsic Factors* Extrinsic factors are associated with the environment and culture in which a person resides and are more behaviorally than genetically derived. Examples include diet, use of tobacco and alcohol, exposure to pollution and sunshine, socioeconomic status, medical practice, compliance with prescribed medications, and practices in clinical trials design and conduct.

*Intrinsic Factors* Intrinsic factors apply more to the genetic features of the individual and the collective subpopulation of individuals within a region, including the physical aspects of genetics plus the lifestyle of the region. Examples include genetic polymorphism, age, gender, height, weight, lean body mass, body composition, and organ dysfunction. Table 1 illustrates intrinsic and extrinsic ethnic factors as defined in E5.

*Sensitivity to Ethnic Factors* Ethnic sensitivity is compound specific, depending on its pharmacologic class, indication, and the usual age and gender of the target patient population. No single factor is predictive of the relative sensitivity, but the following properties as a whole are generally indicative.

Examples of properties making a compound less sensitive to ethnic factors include:

- Linear pharmacokinetics
- A flat pharmacodynamic curve for both efficacy and safety

**TABLE 1 Classification of Intrinsic and Extrinsic Ethnic Factors**

| INTRINSIC | | EXTRINSIC |
|---|---|---|
| Genetic | Physiological and pathological conditions | Environmental |
| Gender | Age (children-elderly) | Climate<br>Sunlight<br>Pollution |
| Height<br>Bodyweight | | |
| | Liver<br>Kidney<br>Cardiovascular functions | **Culture**<br>Socioeconomic factors<br>Educational status<br>Language |
| ADME<br>Receptor sensitivity | | |
| Race | | Medical practice<br>Disease definition/Diagnostic<br>Therapeutic approach<br>Drug compliance |
| Genetic polymorphism of the drug metabolism | | |
| | Smoking<br>Alcohol | |
| | Food habits<br>Stress | |
| Genetic diseases | Diseases | |
| | | Regulatory practice/GCP<br>Methodology/Endpoints |

- A wide therapeutic dose range
- Minimal metabolism or metabolism distributed among multiple pathways
- High bioavailability, that is, less susceptibility to dietary absorption effects
- Low potential for protein binding
- Little potential for drug–drug, drug–diet, and drug–disease interactions
- Nonsystemic mode of action
- Little potential for inappropriate use and abuse

The opposite properties make a compound more sensitive to ethnic factors, especially:

- Metabolism by enzymes known to show genetic polymorphism
- Administration as a prodrug, with potential for ethnically variable enzymatic conversion
- Low bioavailability, with more susceptibility to dietary absorption effects

***Bridging Data Package*** E5 specifies that the complete clinical data package intended for registration contain clinical data that fulfill the regulatory requirements of the new region and contain pharmacokinetic data relevant to the population of the new region. Use of a bridging study allows inclusion of selected information from the original region, including pharmacokinetic, dose–response, and pharmacodynamic data, with appropriate extrapolation to the population of the new region based on the results of the bridging study.

***Global Development Strategies*** Sponsors are increasingly interested in concurrent international development, and use of ICH guidelines allow data to be used across the three major jurisdictions. Thus we are seeing Japanese–Caucasian bridging studies with a clinical pharmacology focus performed concurrently with phase II studies in the United States and European Union, with the expectation that phases II and III studies will be performed in Japan concurrently with the phase III program in the United States.

### 3.3 Implications for Non-ICH Jurisdictions

This may be all well and fine for Japan, the European Union, and the United States. How about the rest of the world? The ICH regions account for approximately 85% of global pharmaceutical sales, but less than 15% of the global population. The non-ICH regions contain some of the most dynamic and changing pharmaceutical markets. These regions are also seeking innovative and better quality medicines, are also interested in benefiting from the medicines developed within the ICH regions, and are also attempting to develop world class drug/biological regulatory systems.

Long recognizing these global implications, the ICH has made efforts to involve stakeholders with similar needs and related interests from other regions. For example, the ICH Steering Committee statement of July 1997 included language that "the widespread adoption and use of ICH guidelines ... is essential if the long-term

benefit of international harmonization in terms of quicker access to effective new medicines, is to be available to patients throughout the world."

In 1999, the ICH steering committee created a Global Cooperation Group (GCG) subcommittee to share information and harmonization practices with other nations. Participants include Canada and the World Health Organization (WHO), representing 180 countries.

Harmonization initiatives are also being coordinated on a regional basis by the following organizations:

- Asia-Pacific Economic Cooperation (APEC)
- Association of Southeast Asian Nations (ASEAN)
- Gulf Cooperation Countries (GCC)
- Pan American Network on Drug Regulatory Harmonization (PANDRH)
- South African Development Community (SADC)

These agencies are participants or interested observers in the ICH process and adopt or incorporate ICH guidelines in their regions where applicable. Thus ICH guidelines, or guidelines similar to ICH, may facilitate global availability of new compounds, global development of new compounds, and globalization of clinical research.

## 4 ETHNIC PHARMACEUTICAL SAFETY ISSUES: PHARMACOKINETICS

Pharmacology is a science. The ethnic, historical, and regulatory issues primarily provide a context for understanding protocol design and the regulatory approval process. The heart and soul of the clinical trials is to provide scientific data according to recognized practices that answer the most important questions regarding the behavior and effects of a specific compound following human exposure.

Sections 4 and 5 provide an overview of pharmacokinetics (PK), that is, the absorption, distribution, metabolism, and elimination (ADME) of the subject compound and the pharmacodynamic (PD) considerations and experience of bridging studies. Only selected examples are cited here, in the interest of illustrating ethnic issues relevant to planning bridging studies.

### 4.1 ADME: Absorption

This discussion of absorption will focus primarily on the gastrointestinal tract. For our purposes, intravenous, intramuscular, and subcutaneous routes largely bypass the absorption process. There may be ethnic differences in absorption of transdermal medications, as these may vary with skin type, including racial differences in skin thickness, barrier function, and transepidermal water loss. In terms of ethnic differences, intraocular, intravaginal, intrarectal, and other routes of administration have not received extensive investigation.

Absorption from the gut is a passive process for many medications. In this case, ethnic differences in absorption would not be expected, and none have been demonstrated. However, a number of medications are actively metabolized within

the intestinal membranes or absorbed via transport proteins. At last three major enzymes systems are involved with drug absorption across the intestinal membrane, CYP (especially CYP 3A4), the multidrug resistance proteins (MDR1-5) including P-glycoprotein (MDR-1), and the organic anion transporting polypeptide (OATP) family, which transport a large array of structurally divergent drugs [4]. While their relevance for ethnic differences at the population level have yet to be defined, individual heterogeneity and single nucleotide polymorphism (SNP) variation have been identified for a number of these transport proteins. Undoubtedly, other transport proteins have yet to be defined.

The bridging studies done with Japanese to date have shown several compounds, such as risedronate and rosuvastatin, with a higher time to maximum concentration ($T_{max}$) and area under the curve (AUC) yet a similar half-life ($T_{1/2}$). These observations remain after adjustment for weight, gender, percent body fat, and related individual factors, suggesting the ADME difference is primarily one of increased absorption rather than differences in distribution, metabolism, or elimination. Bridging studies have documented these observations, even though the precise mechanism remains to be elucidated.

### 4.2 ADME: Distribution

Insofar as a medication's distribution varies throughout body compartments, generally lipophilic medications would be expected to show a difference in the volume of distribution, with ethnic effects reflecting the population's typical percent body fat versus lean body mass. In Japan, obesity is defined as a body mass index (BMI) of greater than 25, whereas in the United States an individual is considered "overweight" with BMI 27–30 and "obese" when the BMI is over 30. This may partially account for the greater effect of diazepam in Asians than in Caucasians.

Some ethnic differences in protein binding have been described. The two major drug-binding proteins in blood plasma are albumin and $\alpha_1$-acid glycoprotein (AGP). Albumin has a relatively low drug affinity, but a high binding capacity with three major binding sites and the ability to bind acids, bases, and neutral compounds. Conversely, AGP has a high drug affinity but low capacity. AGP binds many basic psychotropic compounds, including haloperidol, chlorpromazine, and fluvoxamine. The structures of albumin and AGP are genetically determined, with polymorphic variations across racial groups in several studies. In general, evidence indicates that Caucasians (especially those of European origin) have higher levels of AGP than Asians and blacks, although albumin levels are similar.

Finally, ethnic differences in transport proteins such as P-glycoprotein and the OATP enzyme systems may account for differences in intracellular concentration within the liver, kidey, brain, and other tissues.

### 4.3 ADME: Metabolism

A substantial body of literature exists regarding heterogeneity of metabolizing enzymes, particularly the hepatic cytochrome P450 (CYP) enzyme families. Considering the wide range of intraethnic mutant alleles, some authorities may argue that these approach "individual" factors. Nonetheless, the incidence of certain genotypes shows ethnic patterns and must be considered when discussing ethnic differences.

Wider application of genotyping has clarified our understanding while adding considerable complexity, as we have identified multiple mutant genotypes with similar phenotypical function. As the CYP system is also involved with detoxification or metabolism of a wide range of dietary components and environmental toxins, implications for risk of certain cancers in various ethnic populations may be as important clinically as the implications for drug metabolism and remains an area of active research.

Overall, enzymes in the CYP 2 superfamily show far greater heterogeneity than CYP 1 and CYP 3. In addition to the allele heterogeneity, some studies also show considerable individual and ethnic differences in enzyme content and activity [5–7]. Selected CYP enzyme systems with ethnic relevance for drug metabolism are considered next.

***CYP 1A1, 1A2, and 1B1*** CYP 1A2 substrates include caffeine, theophylline, and tacrine. Four types of CYP 1A1 and three types of CYP 1B1 polymorphisms have been described, with distinct incidence patterns in Japanese compared with Caucasians [8], and may play a role in cancer risk. CYP 1A2 polymorphisms vary in the Chinese population, with some more similar to Japanese and some more similar to Caucasians [9]. Additional CYP 1A2 polymorphisms have been identified and shown to have variable incidence in Taiwanese, Caucasian, and African American populations. However, the functional differences of these variants is fairly small, and relevance for dose selection has not been established [10, 11].

***CYP 2C9*** CYP 2C9 is a major enzyme, accounting for about 20% of the total P450 enzyme pool in both Japanese and Caucasians. Substrates include many nonsteroid anti-inflammatory medications, glipizide, losartan, and S-warfarin. There is large interindividual variability in 2C9 content and considerable intraethnic variation in both population. Some mutant alleles are shared in both populations and some are unique to each population [12].

CYP 2C9 is polymorphic, with six major allele variants. The different alleles vary in ethnic distribution between Asian, Caucasian, and African populations, and different alleles vary in their metabolic activity for common substrates.

***CYP 2C19*** CYP 2C19 substrates include omeprazole and phenytoin. Along with CYP 2D6, CYP 2C19 contributes to the metabolism of selective serotonin reuptake inhibitors, tricyclic antidepressants, and a variety of other psychotropic medications.

***CYP 2D6*** CYP 2D6 substrates include codeine and other opiate medications, β blockers, and antidysrhythmic medications such as flecainide and propafenone. CYP 2D6 also contributes significantly to the metabolism of tricyclic antidepressants, selective serotonin reuptake inhibitors, and other psychotropic medications such as haloperidol and thioridazine.

***CYP 2E1*** The classic pharmacological substrate for CYP 2E1 is chlorzoxazone, although 2E1 is also a metabolic pathway for *p*-nitrophenol, ethanol, and dimethylnitrosamine. Unlike most of the other CYP enzymes, 2E1 metabolism has a tendency to activate substrates into more toxic metabolites, such as acetaminophen,

ethanol, carbon tetrachloride, thioacetamide, and the N-nitroso compounds found in red meat and processed meats. Thus a higher level of 2E1 activity may increase the risk of alcoholic liver injury and various malignancies [13–15].

A comparison of 39 Japanese and 45 Caucasians showed three types of polymorphisms, with seven genotypes found in the Japanese and four in the Caucasians [7a]. Although little difference in metabolic activity was found between the ethnic groups with one probe substrate despite the genotype heterogeneity, there is a typical difference between Japanese and Caucasians in the metabolism of chlorzoxazone, with lower clearance for the Japanese even after adjusting for differences in body weight, suggesting a lower level of catalytic activity in Japanese [4].

***CYP 3A4 and 3A5*** CYP 3A4 is present in the intestinal brush border as well as hepatocytes and is the major "workhorse" for drug metabolism, that is, the major metabolic pathway for as many as 60% of our currently known pharmaceutical compounds. This lack of substrate specificity also lends itself to a large number of drug and diet interactions [16]. Individual 34A content and activity varies widely, with up to 40-fold individual content variation and up to 11-fold individual variability in midazolam oral clearance [5, 17]. More than 30 SNP variations have been identified, which vary in ethnic distribution between Japanese, Caucasian, and African American subjects. Despite this variety, allele frequency is low (<5%) and appears as heterozygous with the wild type. None of the alleles have been associated with clinically significant differences in drug metabolism, especially regarding interethnic comparisons [18].

CYP 3A5 is a coenzyme for many 3A4 substrates. CYP 3A5 is also polymorphically expressed with varying ethnic distribution, but the functional significance has not been established [19].

As we learn more about individual intraethnic differences in content and activity, allele function, and diet interactions, ethnic differences previously reported in CYP 3A4 activity may be partly due to individual differences in studies with small sample sizes and/or dietary factors [6, 20].

***Non-CYP Metabolic Enzymes*** A number of enzymes other than the CYP system are also important in drug metabolism and demonstrate ethnic heterogeneity. The intra- and interethnic genotype variation is not always linked with phenotype expression, and this remains an area of active inquiry. Selected examples include:

- Aldehyde dehydrogenase (ALDH) is primarily significant for metabolism of ethanol, especially ALDH2. An atypical allele, ALDH2(2) results in accumulation of acetaldehyde with histamine release and flushing. Approximately 30% of Asians are ALDH deficient [21].
- *N*-acetyltransferases (NAT) demonstrate polymorphism and ethnic variation and are involved with metabolism of environmental carcinogens and a number of arylamine and hydrazine drugs [22]. As expected, SNP and allele analysis show substantial heterogeneity, both within Japanese and within Caucasians [23].
- Thiopurine *S*-methyltransferase (TPMT) is responsible for the metabolism of thiopurine medications such as azothioprine. Patients with deficient TPMT

activity are at high risk for toxicity and can be successfully treated with 10- to 15-fold lower dose. Mutant alleles vary in their ethnic distribution [24].

- Glucuronsyltransferase (UGT) enzymes catalyze the transfer of glucuronic acid to alcohols, carboxylic acids, amines, and free sulfhydryl groups. Twenty-four distinct UGT genes have been identified, with significant ethnic variation in Caucasian, Asian, and African populations [25].

### 4.4 ADME: Elimination

The major routes of elimination are hepatic and renal. Ethnic differences in hepatic function are primarily metabolic. To my knowledge, no ethnic differences in post-hepatocyte biliary elimination have been described.

Renal clearance is the net effect of glomerular filtration, tubular secretion, and tubular reabsorption. Glomerular filtration and tubular reabsorption are largely passive processes, where ethnic differences have not been described. However, tubular secretion is an energy-requiring active process, and ethnic differences could possibly result from genetic heterogeneity in this process although they are particularly compound specific. For example, renal clearance for morphine and metabolites morphine-3-glucoronide and morphine-6-glucoronide are higher in Chinese than Caucasians, primarily due to tubular secretion. However, other medications that undergo tubular secretion have not shown ethnic differences, such as furosemide and procainamide.

### 4.5 ADME: Dietary Factors

Most of the ADME factors described above are primarily genetically determined. But diet can have a significant influence on how these enzymes function, which may exaggerate or introduce new ethnic differences.

As a context, the "traditional diet" in Japan is quite different from most regions within the United States. Despite globalization, certain generalities still apply:

- Source of complex carbohydrates: Japanese consume more rice, and Americans more wheat, corn, and potatoes.
- Source of protein: Japanese consume more fish and soy products, Americans more beef and dairy products.
- Sources of fat: Japanese consume less fat overall, and the fat consumed is more likely to be from fish and vegetable sources. Americans consume more fat overall, and a higher percent is saturated fat from red meat and other animal sources.
- Typical vegetables and fruits vary between Japan and the United States, in addition to the seasonal variations in both regions.
- Common food preparation techniques and preferred sauces, flavorings, dressings, and spices are typically quite different.

Macronutrient differences may have some affect on drug absorption, as will physiological changes in the fed versus the fasted state such as gastric pH, gastric and intestinal motility, pancreatic and biliary secretion, drug adherence to dietary

fiber, passive drug absorption along with meal uptake, and the like. However, we should draw particular attention to the vegetables, fruits, herbs, and spices. Consumption of these items often varies by culture or region, and many fruits, vegetables, and herbs have been shown to have isoflavones or other compounds that can affect drug absorption, metabolism, or both.

One well-known example is grapefruit and grapefruit juice, which contain isoflavones that significantly inhibit CYP 3A4 in both the intestinal wall and the liver. For medications metabolized by 3A4, grapefruit intake reduces metabolism in the gut, thus increasing absorption and the $C_{max}$. Grapefruit also prolongs the $T_{1/2}$ by inhibiting hepatic metabolism. Both mechanisms contribute to a higher AUC. These effects persists at least 24 hours after dietary exposure. This grapefruit effect has presumably been present throughout human biological history, even though it was only defined in the 1990s [26]. This effect apparently applies across all ethnic groups, yet some would classify this as an "ethnic factor" as grapefruit consumption varies by region and subpopulation.

Herbal products can interact with prescription medications. St. John's wort (hyperforin and quercetin) is commonly used as an antidepressant and has been shown to be a potent inducer of CYP 3A4 and P-glycoprotein, although it may inhibit other CYP enzymes [27]. St. John's wort has been shown to lower the plasma concentration (and/or the pharmacological effect) of a number of drugs including alprazolam, amitriptyline, cyclosporine, digoxin, fexofenadine, indinavir, irinotecan, methadone, nevirapine, simvastatin, tacrolimus, theophylline, warfarin, phenprocoumon, and oral contraceptives, with interactions that may cause a variety of adverse clinical effects [28].

Other herbs, spices, and "natural medications" have significant effects on CYP enzymes. Ginkgo biloba (Gingkofar) can exert potent inhibition on 1A2, 2C9, and 2C19, with less inhibition of 2D6 [29]. A variety of spices, herbal and black teas, and soybean products were tested for their in vitro inhibition of CYP enzymes. All were found to inhibit 3A4 mechanisms to a variable extent, and 26 showed significant inhibition of 2C9, 2C19, or 2D6. Examples of products tested include black tea, herbal tea mixtures, sage, thyme, cloves, goldenseal, and seven soybean varieties [30]. Similar inhibition of CYP 1A2, 2C9, 2C19, 2D6, and 3A4 was noted by herbal extracts of devil's claw root (*Harpagophytum procumbens*), feverfew herb (*Tanacetum parthenium*), fo-ti root (*Polygonum multiflorum*), kava-kava root (*Piper methysticum*), peppermint oil (*Mentha piperita*), eucalyptus oil (*Eucalyptus globules*), red clover blossom (*Trifolium pretense*), garlic (allicin), and evening primrose oil (cis-linoleic acid) [29, 31]. Gallic acid, a plant phenol found in wines and teas, can inhibit CYP 3A4 under certain conditions [32].

In general, a wide variety of clinically significant adverse events have been documented from herb–drug interactions, through a combination of pharmacokinetic and pharmacodynamic mechanisms [33]. To the extent that herbal use characterizes an ethnic group in a particular region, herbal use can be considered an ethnic factor.

Diet components may also affect transport proteins. Grapefruit, orange, and apple juice have been shown to inhibit human OATP activity, with reduced oral drug bioavailability [34]. Another study of 20 naturally occurring flavonoids and some of their corresponding glycosides demonstrated significant inhibition of OATP drug uptake, indicating the potential for drug interactions [35].

P-glycoprotein is also inhibited by a wide variety of dietary and herbal compounds, including capsaicin (in chili peppers), curcumin (in turmeric), [6]-gingerol (in ginger), and resveratrol (in grapes) [36]. Eight dietary flavonoids commonly present in fruits, vegetables, and plant-derived beverages have been shown to inhibit P-glycoprotein [37, 38]. Citrus (rutaceous) herbs are often used in traditional Japanese medicine and cuisine and contain compounds known as monoterpenoids. A number of these have been shown to be potent P-gp inhibitors [39].

Even food preparation preferences can affect drug pharmacokinetics. Charcoal-broiled and smoked foods contain polycyclic aromatic hydrocarbons formed during their preparation, which can induce intestinal CYP 1A [40].

Much of the research in this area is fairly recent and more remains. We are likely to define new transport proteins, new receptors, and new dietary influences on the pharmacokinetic or pharmacodynamic effects of drugs.

### 4.6 ADME: Multiple Mechanisms

Certain compounds will be influenced by combinations of the above factors. Before we complete our discussion about ethnic factors in pharmacokinetics, let us use warfarin, one of the world's most widely prescribed anticoagulant medications, as an example. Although effective at reducing the incidence of thromboembolic disease in common clinical settings, there is a greater than 10-fold interpatient variability in the dose required, necessitating careful monitoring and dose adjustment for each patient.

Warfarin exists as optical S and R isomers. S-warfarin has three to five times more anticoagulant potency than its optical congener. Warfarin is well absorbed orally, although delayed in the presence of food. Absorption through the bowel as well as target intracellular concentrations may be affected by P-glycoprotein. Once absorbed, warfarin is highly protein bound, mainly to albumin. Warfarin is primarily metabolized by CYP 2C9, with minor activity by CYP 1A2 and CYP 3A, and elimination of inactive metabolites in the urine and the stool.

As discussed previously, CYP 2C9 enzymes show marked phenotypical and genotypical variability in various ethnic groups worldwide, with multiple alleles demonstrating considerable differences in rates of S-warfarin metabolism. Multiple variants have also been described in the promoter, exonic, intronic, and 3′-untranslated regions of CYP 2C9.

A wide variety and a large number of medications have reported interactions with warfarin, including azole, macrolide and quinolone antibiotics, nonsteroidal anti-inflammatory drugs, selective serotonin reuptake inhibitors, omeprazole, lipid-lowering agents, amiodarone, and fluorouracil.

Interactions with a wide variety of food and herbal medications have also been described, including vitamin K intake, geen tea, danshen, fish oil, soy milk, ginseng, and quilinggao.

While there are wide interindividual differences, Asians in general tend to achieve therapeutic benefit at lower warfarin doses than Caucasians. When assessing "what are the ethnic effects," studies have compared Caucasian with Asian subjects matched by CYP 2C9 allele variants and still found unexplained differences. Likewise, interethnic differences were present with or without consideration of CYP 2C9

coding region variants, suggesting other genetic, dietary, or environmental influences [41, 42].

In this case a simple explanation for ethnic differences remains elusive. Even accounting for the genetic heterogeneity in the primary metabolic pathway, multiple mechanisms appear to be involved. Undefined genetic factors, general dietary patterns, and other environmental influences are all ethnic factors, both intrinsic and extrinsic.

## 5 ETHNIC PHARMACEUTICAL SAFETY AND EFFICACY ISSUES: PHARMACODYNAMICS

As we discussed earlier, safety and tolerability are as important, if not more important than efficacy. Once the PK profile is understood and necessary dose adjustments taken into consideration, we are prepared to assess efficacy, safety, and tolerability of the selected doses.

Pharmacodynamic effects are often about the same in Japan as elsewhere but need to be evaluated in the context of the dose and related PK characteristics. Consider risedronate sodium as an example, approved in Japan in 2002 for the treatment of osteoporosis with use of a bridging strategy. The dose was set at 2.5 mg/day in Japan, compared with 5 mg/day in the United States and the European Union because of a PK finding that a 2.5-mg dose in Japanese produced a similar AUC to a 5-mg dose in Caucasians. In addition, the efficacy of 2.5 mg/day in Japan regarding the effect on bone density was comparable to the benefit of 5 mg/day in the United States and the European Union. The main difference in the PK was increased absorption. Although these findings primarily involve the PK profile, note that confirmation of the same PD effect at the lower yet equivalent dose was necessary for dose labeling.

The same was observed with rosuvastatin calcium, for the treatment of hypercholesterolemia. Interestingly, this compound was discovered by Shionogi in Japan but developed by AstraZeneca in the United States and the European Union, with subsequent bridging studies leading to approval in Japan. Rosuvastatin received U.S. approval in August 2003 compared with approval in Japan in December 2004. Japanese and other Asian subjects showed increased absorption, with $C_{max}$ about twice as high as Caucasians at the same doses. Efficacy at the lower doses paralleled the PK profile. Thus labeling in the United States is from 5 to 40 mg/day with a recommended starting dose of 10 mg/day, compared with labeling in Japan of 2.5 to 20 mg/day with a recommended starting dose of 5 mg/day.

In other cases, for reasons not always well understood, the PD effects and adverse event profile may differ in Japan from the United States and the European Union. Eletriptan hydrobromide, for the treatment of migraine headaches, is another medication approved based on a bridging strategy. The maximal doses were set at 40 mg/day in Japan, compared with 80 mg/day in the European Union, after the bridging study demonstrated higher rates of adverse events such as nausea and somnolence in the Japanese at 80 mg/day, although the $C_{max}$ and AUC of single doses were lower in the Japanese than the foreign population. Likewise, sildenafil citrate is recommended at 50 mg, ranging from 25 to 100 mg in the United States and the European Union, but limited to 25–50 mg in Japan as the 100-mg dose caused increased

adverse events (e.g., abnormal vision) in Japanese subjects but added no further effectiveness.

After allowance for ADME profiles, one possible mechanism for different PD responses is ethnic differences in the target receptor. Ethnic heterogeneity has been described for some important receptors. Selected examples include:

- Three functional β-adrenergic receptor polymorphisms have been described with clinical relevance to several disease, including asthma, hypertension, congestive heart failure, cystic fibrosis, and obesity, as well as response to β-agonist therapy such as salbutamol or β-blocker therapy such as propranolol. Distribution of these alleles varies in Caucasian, black, and Asian populations, providing at least a partial explanation for their different clinical responses to therapy [43–46].
- The effects of serotonin-related gene polymorphisms on central nervous system (CNS) serotonergic function vary as a function of both ethnicity and gender [47] and may affect response to therapy.
- The mu opioid receptor gene (OPRM) has several known variations, with different distribution patterns in Caucasian, Asian, and black populations [48]. Ethnic variations in receptor function contribute to differences in the response to morphine [49], have been associated with methamphetamine psychosis [50], and may affect patterns of opiate addiction [51].

Receptor heterogeneity is an area of active investigation, and ethnic patterns are likely to be defined with relevance for numerous clinical domains [52].

## 6 BRIDGING STUDY DESIGN

Genetically determined metabolizing enzymes, transport proteins, and receptors are significant intrinsic ethnic factors. Diet is one of the major extrinsic ethnic factors. The data cited above show some reasons to believe drugs may behave differently in different populations or new regions. With the regulatory and physiological context as above, let us turn our attention to bridging study design.

For simplicity of discussion, this section will use a scenario where the bridging goes from the European Union or the United States to Japan, that is, where a compound is approved in the European Union or the United States, or a significant body of data exist in one of these jurisdictions and the compound is well on its way to regulatory approval, and Japan is the new jurisdiction of interest. In the first 5 years of E5 guideline implementation, this has been the most common scenario. Presumably, the same general principles would work when bridging from Japan to the United States or the European Union or bridging from an ICH jurisdiction to a non-ICH jurisdiction.

The comments that follow in this section and the next are based on experience between 1998 and 2005. Requirements for data acceptability are inherently a "moving target," as we accrue experience with bridging programs in the context of emerging scientific data, medical practice changes, and political realities. In Japan, the Organization for Pharmaceutical Safety and Research (OPSR) offers consultative guidance to sponsors planning a bridging program, and sponsors are generally well advised to seek guidance prospectively.

### 6.1 Candidate Compound

As expected, the ADME characteristics as well as the intended clinical application of the candidate compound will determine the bridging strategy. The sponsor must assess the potential for intrinisic and ethnic effects based on the criteria discussed above.

### 6.2 Basic Study Design

The successful bridging study programs to date have often utilized a phase I style clinical pharmacology study to establish the dose selection, most commonly in normal healthy individuals, followed by patient-based longer term Phase II or Phase III style studies demonstrating safety and efficacy within the new jurisdiction. In the case of applying for registration in Japan, the safety and efficacy studies for successful NDAs so far have been performed in Japan. Utilizing ethnic Japanese populations residing out of Japan to demonstrate safety and efficacy has been problematical. As per the examples discussed above, we do not yet know all the variables and mechanisms affecting a compound's PK profile or PD effects. These considerations make it understandable to perform certain safety and efficacy studies within the new target region.

Study design needs to comply with good clinical practices and related ICH guidelines, with ethical principles consistent with the most current internationally recognized standard, currently the Declaration of Helsinki.

Single-dose PK/PD studies ordinarily include a range of at least two or three doses; the usual dose administered in the jurisdiction of reference and one or more lower or higher doses to clarify dose selection in the new jurisdiction. For compounds where single-dose kinetics do not reliably predict steady-state kinetics, a multiple-dose study (again, with a range of doses administered) will likely be advised in the new ethnic group. Pharmacokinetics may need to be confirmed in a patient population if the pivotal PK study was performed in healthy individuals.

The phase II/III efficacy studies may or may not employee a placebo, versus a range of two or three active doses or an already approved active comparator. Dose selection for these studies is based on the bridging ADME data. If the doses selected are lower for the Japanese than commonly used in the United States or the European Union, pharmacodynamic efficacy endpoints need to be confirmed as well as the safety and tolerability at the lower doses selected.

***Variables Affecting ADME Study Data*** As major decisions regarding the phase II/III program will depend on the pharmacokinetic profile, the reference PK study needs to be carefully designed.

A host of variables may affect ADME study data, some of which can be forseen and anticipated, others are more subtle. Such variables are especially relevant when considering the decision to use historical data for comparison versus concurrent controls, and when considering whether to perform a study at one center or two. At some level, these variables may affect study conduct, data capture, or the regulatory agency's perception of data quality.

Examples include:

- Staff experience, training, and expertise
- Characteristics of the local population from which subjects are recruited
- Procedures for obtaining and documenting informed consent
- Overall GCP compliance
- Temperature, humidity, mold, lighting, ventilation, and other internal environmental variables within the study unit
- Staffing and work flow regarding study procedure performance
- Work habits and involvement of the principal investigator
- Interpretation and application of the protocol
- Source documentation data capture
- Methods and style for inquiry and documentation of adverse events
- Specimen processing and handling
- Calibration and maintenance of scales, centrifuges, and related devices
- Diet and beverages provided while confined, including food preparation techniques
- Interpretation and application of study restrictions
- Medical monitor approval of protocol deviations
- Quality and completeness of the regulatory file
- Monitoring and audit team findings, reports, and interaction with study staff
- Local lab methodology differences and reported ranges of normal
- Unexpected external events (e.g., local electrical power surges or failure, and storms, earthquakes, or other natural disasters)

Bridging PK studies have been successfully performed utilizing a one-protocol, two-center approach [53]. Some studies require placement at multiple centers for timely completion. Nevertheless, for the reasons listed above and many others, combining or comparing data across centers may involve unexpected findings or subtle differences. If unexpected findings occur—are they the result of ethnic differences or study conduct variables? Unexplained findings may require repeating the study under more rigorously controlled conditions.

***Caucasian ADME Data: Historical or Direct Comparison?*** The fundamental assumptions of bridging strategies involve use of historical data gathered in a foreign region, with assessment for applicability to the new region.

Partially for this reason, the use of historical Caucasian data seems like an efficient and cost-saving approach when planning the ADME study. However, if the new data show unexpected differences, it may be difficult to tell if this is due to "real" ethnic differences or the results of differences in study design, conduct, or other nonethnic variables. Considering the pivotal nature of the PK profile for the rest of the bridging program, it is generally preferable to plan the PK study with concurrent enrollment of subjects from both ethnic groups.

### 6.3 Subject Population

***Definition of "Japanese"*** Intuitively, defining "Japanese" seems like a simple issue, but it is not. Every ethnic group contains a certain amount of heterogeneity.

Japan is no exception, that is, sumo wrestlers and geisha may both have unquestioned Japanese genealogy. Ancient migration patterns produced different subpopulations in regions spreading from the Kyushu District (Okinawa and other islands in the south) to Hokkaido (northern islands), and immigration within the past century or so has introduced many Korean, Chinese, and others to Japan. How many generations does it take for descendents of such individuals to become "Japanese?" Conversely, if an individual is born and raised in Japan, with a pedigree of pure Japanese parents, grandparents, and great-grandparents, does he or she cease to be "Japanese" as soon as he leaves the confines of Japan's airspace?

Even more debatable is how many generations can pass after leaving Japan before an individual is no longer considered "Japanese." Generational terms for Japanese are as follows, each assuming that the pedigree is pure Japanese ethnicity:

- Issei (first generation, subject was born in Japan)
- Nissei (second generation, subject born elsewhere but parents born in Japan)
- Sansei (third generation, subject and one or both parents born elsewhere but grandparents born in Japan)

My Japanese colleagues who perform phase I studies in Japan report that if a subject in Japan chooses to participate in a clinical pharmacology study conducted in one of their units in Japan, there is relatively little requirement to document Japanese ethnicity. Yet for many studies, if that same subject wishes to participate in a clinical study in Honolulu, we must document Japanese birth and citizenship by passport or visa, and fill a three-generation genealogy tree documenting that both parents and all four grandparents are Japanese and were born in Japan.

The stricter standard for "foreign data" is understandable, in that "Japanese" citizens born, raised, and living in Japan represent the population of interest to the MHLW—regardless of exact ethnic pedigree or regional origin. Subjects out of Japan may or may not represent their population, and it is important to know that the subjects enrolled as "Japanese" meet acceptable criteria. Consequently, virtually all bridging studies performed outside of Japan involving Japanese subjects require at least a "four grandparents" rule, that all four grandparents are known to be of Japanese ethnicity. This allows Issei, Nissei, and Sansei participants, but not those of mixed Japanese/non-Japanese descent.

Depending on the goals of the study, some protocols limit Japanese enrollment to Issei (with documentation of a three-generation geneology chart), sometimes with a requirement that the subject has been outside of Japan for a limited time, for example, no longer than 5 or 10 years. Additionally, many protocols also inquire regarding dietary patterns and lifestyle, discussed in further detail below.

***Definition of "Caucasian"*** Likewise, defining "Caucasian" may seem to be simple, but it is not. When we have asked sponsors for their definitions, responses vary considerably, including one to the effect "anyone with blonde hair, light skin, and blue or green eyes!" Fortunately, most sponsors include more geographically based definitions. Of course, many non-Caucasians are born or raised in Europe. However, without quibbling about ancient migration patterns or attempts to nar-

rowly define genetic origins, we deal with a reasonable sense of what it means to be of "European" origin genetically, that is, the people characteristically and traditionally known as English, German, French, Italian, Russian, and so forth. Specific protocols may or may not include those of the Middle East or North African descent, partially depending on the population in which the majority of clinical pharmacology data were originally generated: If performed in Europe the definition of "Caucasian" tends to reflect the participants of the original studies; if performed in the United States the definition of "Caucasian" is generally more broadly defined.

The Guidance for Industry Regarding Collection of Race and Ethnicity Data in Clinical Trials published by the U.S. FDA, January 2003, includes guidance for clinical trials conducted outside of the United States and recognizes racial designations of:

- American Indian or Alaska Native
- Asian
- Black, of African heritage
- Native Hawaiian or Other Pacific Islander
- White

The guidance also includes ethnic categories of "Hispanic or Latino" or "Not Hispanic or Latino." Per this guidance document, the racial category "white" can reflect origins in Europe, the Middle East, or North Africa, and the category "Asian" can reflect origins from areas ranging from India to Japan.

***Healthy Normal Versus Patient Population*** Most bridging studies follow the model established in the original jurisdiction and are most commonly performed in normal healthy volunteers. The exception usually involves medications with significant potential toxicities, with ethical constraints limiting exposure in healthy individuals who will not be receiving any direct medical benefit. Precedents exist for performing PK studies in patients as well as healthy volunteers, for measuring the PK profile in patients as part of a study to demonstrate efficacy, and for confirming the PK in patient studies as well as performing a PK study in healthy volunteers.

***Male and Female Participants*** Unless the subject compound was intended exclusively for a female population (e.g., oral contraceptives), pharmacokinetic studies were traditionally performed in healthy male volunteers in all jurisdictions. Current regulatory guidance favors inclusion of female subjects, as long as there are appropriate measures to minimize the risk of in utero fetal exposure to an investigational drug. Bridging studies follow this same trend.

"Adequate" contraceptive measures vary with the protocol. Oral or injected hormonal contraception is generally allowed except when concerns exist regarding direct or indirect drug interactions. Requiring surgical sterilization or postmenopausal status often introduces an age difference between the male and female participants. Intrauterine devices are not commonly used. Abstinence or various forms of "double-barrier" contraception are increasingly allowed. Regardless of the method of conception, enrollment of women of child-bearing potential requires negative tests for pregnancy at the initial screening visit and again shortly prior to dose administration.

***Age*** What is the difference between an 18-year old and a 20-year old? Not much, from a clinical pharmacology perspective. But this is an important distinction from a regulatory standpoint. In most of the United States, an 18-year old is legally an adult and has the autonomy and legal authority to sign his or her own informed consent. In Japan, this occurs at age 20. Consequently, many bridging studies define the lower age limit as 20 rather than 18.

The upper age limit is increasingly flexible. Through the 1990s many studies used age 45 as the upper limit for normal healthy studies. As the average population age increases—a trend occurring in Japan as well as the United States—there is growing interest in the inclusion of more mature participants, particularly for compounds whose clinical applications will largely apply to an elderly population.

Bridging studies will typically follow the trends established in the original jurisdiction, and more commonly include individuals up to age 50 or 55 in their "normal healthy" population, and sometimes to age 65. The trend established in the original jurisdiction is also followed for compounds that will be almost exclusively used by an elderly population, and bridging studies have been performed with a lower age limit of 50.

At the other end of the spectrum, pediatric trials have their own unique issues, including the requirement for the parent/guardian to sign the informed consent and for the participant to sign an age appropriate document indicating his or her assent. Bridging studies may also include a pediatric population. As of 2003 two compounds for pediatrics use were approved in Japan utilizing bridging studies: Palivizumab (for the treatment of respiratory syncytial virus in pediatrics) and oseltamivir phosphate for pediatric influenza infection. The pediatrics approval of oseltamivir followed the approval for use in adults at higher doses.

***Weight*** Historically, the Japanese population has been smaller and lighter than the American population, and weight differences account for some of the differences in typical doses administered. Ideally, bridging studies will reflect the populations in both jurisdictions, and yet still provide comparable data. In a study enrolling concurrent Japanese and Caucasian participants, we ordinarily see a reasonable range for body mass index (BMI) applicable to both populations. Perceptions of appropriate body weight vary: In Japan a BMI over 25 is considered obese, whereas in the United States obesity is a BMI over 30. To better represent the Japanese population, the lower limit is often around 18, especially if the study enrolls female participants. To better represent the Caucasian population, the upper limit is often 27 or 29. Within these ranges, weight-adjusted pharmacokinetics can be determined.

### 6.4 Diet and Food Effects

***Japanese Versus Western Diet*** Paradoxically, as globalization progresses, diets worldwide are both increasingly diverse and increasingly similar. On a short tour of a major city in Japan, one observes Japanese consuming coffee, hamburgers and fries from American-based "fast food" outlets, and Americans have a growing taste for sushi, miso soup, and saimin. Both populations enjoy pizza and pasta.

Despite this globalization, we have discussed the significant differences between the traditional Japanese diet and the traditional American diet, and many generali-

ties still apply when we look at the overall population. To illustrate, I believe most international travelers would agree that menu choices in a typical local restaurant in Japan look quite different from menu choices in a typical local restaurant in America!

Do these differences affect a compound's pharmacokinetics? Data discussed previously regarding diet effects on metabolizing enzymes and transport proteins indicates it is prudent to assume "yes" until proven otherwise. Such effects may be related to specific diet contents and effects on pharmacologically related enzymes, beyond the usual fed versus fasted effects.

For this reason, establishing an appropriate diet for bridging study requires considerable attention. If differences exist, are they dietary or genetic? We have seen some studies structured where the Caucasians consume a typical "Western" diet, to match the conditions of the original studies, and the Japanese participants consume a more typical "Japanese" diet to match the conditions expected in Japan. Other studies require that all subjects consume the same diet, which may be Western, Japanese, or selected from common denominator food items shared by both cultures. Priorities vary with the compound, any known food effects, and the mechanics of absorption.

If a specific diet is not defined in the protocol, it may be prudent to provide a sample menu along with other study data when submitted for approval.

***Fed/Fasted Studies*** Taking a medication with a meal may have no effect or may increase or decrease absorption. For compliance and optimal labeling, this is an important question for orally administered medications.

At this time, the OPSR has not defined a Japanese equivalent of the FDA's standard "high-fat" breakfast to evaluate food effects. Although other specific diet items may be of scientific interest, when it comes to regulatory submission, a high-fat meal conforming to the FDA's guidance is the place to start when assessing the effects of taking a medication fed versus fasted.

## 6.5 Restrictions

The approach to study restrictions varies with the compound, the intended clinical application, and the conditions under which data were generated in the original jurisdiction. Often, sponsors will retain certain study restrictions from the original PK studies when they are planning bridging studies so the data will be more easily compared, even though subsequent experience has shown that the original design was overly conservative.

Common restrictions include the following examples.

***Concomitant Medications*** Unless a drug interaction is the goal of the study, concomitant medications are usually restricted for anywhere from 1 to 2 weeks, or up to 5 half lives, whichever is longer. Medications known to inhibit or induce the metabolic enzyme pathways of the study compound are generally prohibited for a longer period of time. Both over-the-counter and prescription medications are commonly cited, and usually herbal products and nutritional supplements are limited or prohibited. Certain exceptions often apply, such as acetaminophen or continued use of oral contraceptives.

***Smoking and Use of Tobacco Products*** Nicotine may not have a direct effect on the metabolism of the study drug but is well known to affect blood pressure and heart rate, and the occurrence of nicotine withdrawals during a clinical trial skews the adverse events profile. However, smoking up to 10 cigarettes a day is still quite common in a young Japanese population, and the exclusion of smokers may cause recruitment delays.

A compromise measure is to limit cigarette consumption to a total of 10 cigarettes a day and to prohibit smoking for 2 hours before vital signs or electrocardiograms are obtained.

***Caffeine/Xanthine Products*** Intake of caffeinated food and beverages is nearly ubiquitous in both Japanese and Caucasian participants. Again, there may or may not be direct effects on the ADME profile, but there may be effects on urine output and vital signs, and caffeine withdrawal headaches and related symptoms affect the adverse events. Additionally, different forms and quantities consumed by different subjects adds another variable that could complicate data analysis.

For these reasons, use of xanthine-containing products is often excluded during a confinement period, and typically for 48–72 hours prior to the first dose and sometimes through the entire course of the study, even if there are outpatient visits following the confinement. However, use is so common and xanthines are contained in so many different products that it is difficult to get through a large study, especially with cross-over periods, multiple confinements, or multiple follow-up exams, without one or more subjects violating these conditions. So as to not have an unacceptable number of protocol deviations, or lose subjects prior to study completion, it is prudent to limit these restrictions or include language in the protocol such that a medical monitor may allow participation if the quantity consumed is deemed too small to have a clinically or pharmacologically significant effect.

***Exercise*** At least in Hawaii, most young healthy study participants are quite active physically. While most medical authorities support an individual program of appropriate regular aerobic exercise, vigorous exercise while enrolled in a clinical pharmacology study introduces variables in cardiac output, renal and hepatic blood flow, hydration, and other circulation changes that can affect ADME data [54]. These are especially problematic if there are marked subject-to-subject differences in the intensity or duration of the exercise, especially if the subjects baseline level of fitness are dissimilar. Vigorous exercise can also cause elevation of creatinine phosphokinase (CPK), aspartate aminotransferase (AST), and other enzymes, which may raise unnecessary questions about the compounds potential to cause myositis, rhabdomyolysis, or cardiac or hepatic toxicity.

For these and related reasons, most bridging studies collecting ADME data limit or prohibit vigorous exercise during critical portions of the study conduct, and often for 48–72 hours before the first dose. When dealing with a young healthy population, it is also advisable to obtain a chemistry panel including CPK and AST at admission or prior to dose in addition to screening, to clarify exercise versus pharmacological effects if elevations are observed postdose exposure.

## 7 BRIDGING STUDY CONDUCT

### 7.1 Recruitment

If the study unit is located in an area with a large Caucasian population, recruitment of Caucasian subjects is about the same as recruiting for any other phase I study. Or sometimes easier, if the study compound is already approved by the FDA or there is a large body of safety data in Caucasians with dose exposure similar to the study of interest and the risk of significant adverse events is small.

Recruiting Caucasian subjects in Japan is limited by the available local Caucasian population, and performing bridging studies in Japan reflect the timelines and cost structure for other clinical pharmacology studies.

Recruiting Japanese participants outside of Japan is primarily limited by the local population of eligible Japanese subjects. Yet a sizeable population alone does not necessarily assure that they are willing to be study participants. Study participation is still viewed with considerable suspicion by many Japanese. When recruiting Issei participants, cultural biases exist against exposure to medications with unknown risks, as well as known adverse events even if the risks are small. Overcoming this bias requires establishing trust, credibility, and considerable discussion, a process usually facilitated with the assistance of Issei Japanese staff members and prior study participants.

As with any "special population" study, the narrower the population is defined the more difficult it is to recruit. Limiting the Japanese participants to Issei (versus four grandparents, allowing Sansei participants), restricting BMI to less than 25, prohibiting smokers, and restricting the upper age to 45 narrows the pool of eligible candidates, especially if the Issei subjects have a time restriction on how long they have been out of Japan. Inclusion/exclusion criteria allowing both male and female participants, but requiring the females to be Issei, out of Japan less than 5 years, under age 45, and of non-child-bearing potential means that for practical purposes, the study will only enroll male subjects.

"Matching" Japanese and Caucasian subjects by one or more criteria also complicates the recruitment process. Matching subjects by sex, weight, or age may increase the confidence in pharmacokinetic analysis but must be weighed against the greater time and expense required for recruitment.

### 7.2 Informed Consent

ICH and GCP standards require that the consent be truly "informed," with the subject's signature only after adequate discussion whereby the study staff is confident the subject truly understands the nature of the study, including its attendant risks, goals, benefits, procedures, and restrictions. The major concern with bridging study is language: If recruiting Sansei subjects who grew up with English as their primary language (for studies performed at American centers), obtaining informed consent is much the same as for English-speaking Caucasian subjects.

However, if the protocol requires Issei subjects, especially with a fairly short time restriction out of Japan, one cannot necessarily assume that a discussion in English and use of an English language consent is adequate. While Issei candidates may be conversationally fluent in English, the medical and legal terms in many consent

documents are not necessarily familiar to them, yet they may not want to admit they are uncertain about the meaning. At times, I have found this to be the case even when dealing with university students who attend classes conducted in English and read textbooks written in English. As mentioned earlier, informed consents are not a familiar item in Japanese culture, and signing a document listing possible risks in detail is uncomfortable for many Japanese.

For these reasons, it is usually advantageous to provide an informed consent document in Japanese, and utilize staff fluent in Japanese to facilitate the discussion. This means obtaining approval and maintaining two versions of the consent: one in English and one in Japanese. The institutional review board (IRB) must approve the Japanese translation and often require that the translation be performed by their own staff. Japanese sponsors often request that this translation be submitted to them for their review and approval before it is utilized. Maintaining dual-language informed consents adds time and cost, and needs to be considered in advance.

### 7.3 GCP Practices

Good clinical practices are well specified elsewhere, especially ICH Section E6, and will not be discussed here other than to note that bridging studies must be compliant with these guidelines. Were it not for these consensus guidelines, bridging studies as we know them could not even be performed. Study planning and conduct compliance with these guidelines is what allows data from bridging studies to be acceptable when submitted to the new jurisdiction. Their importance cannot be overstated.

### 7.4 Adverse Events

Definition and documentation requirements for adverse event collection and reporting are specified in the good clinical practices. Bridging studies follow these guidelines the same as other studies. There are two particular challenges: language and culture.

As explained when discussing the informed consent process, Japanese participants may not be familiar with English medical terminology even if fluent in ordinary conversation. And, as with all subjects, they may only understand and explain their symptoms using informal, colloquial terms with which they are familiar. If Issei subjects are enrolled, it is advisable to utilize staff fluent in Japanese to inquire regarding adverse events and translate their results.

Related to this, participant's use of concomitant medications from Japan often requires a search for drug names or components that can be coded for Western regulatory agencies.

Japanese culture generally downplays spontaneous adverse event reporting, as many feel that there is some level of shame associated with experiencing an adverse event. This makes active inquiry even more important. At a minimum, the "expected" adverse events would have been listed in the informed consent, and subjects should be made to feel comfortable describing these should they occur. A private environment with supportive staff fluent in Japanese should inquire at regular intervals regarding adverse events in general, make sure they are adequately characterized and "translated" into codeable terms, and continue follow-up until resolved.

### 7.5 Documentation

Documentation again follows GCP guidelines, generally utilizing source documents and separate case report forms (CRFs). The standard approach to "document everything" is especially important for bridging studies. Free-form progress notes and explanatory notes to file are important items even with excellent source documents, particularly for any protocol deviations or events out of the norm.

American study centers will generally utilize English language source documents. It is advisable to capture data using metric units for weights, measures, and temperature as much as possible. CRFs may be written or electronic. They will be in English if the study center uses English-speaking staff for data entry, but good CRFs will have already defined how the data will be coded in Japanese for their own NDA submission.

### 7.6 Monitoring and Audits

As the pharmaceutical industry becomes increasingly globalized, monitoring and audits are often shared between staff in Japan, the United States, and the European Union. These processes also follow GCPs. Adequate monitoring and detailed audits, with observation of study conduct and data verification, is a feature distinguishing ICH-compliant studies from clinical trials performed prior to ICH implementation.

Because bridging studies are pivotal by nature, most are audited, and may well be audited by more than one team. Bridging studies are also subject to audits from the regulatory agencies in both jurisdictions.

## 8 BRIDGING STUDY EXPERIENCE

Since the E5 guidelines were adopted in 1998, industry has responded by implementing bridging strategies and submitting NDAs utilizing bridging data. Starting with 3.2% in 1999, use of bridging studies increased steadily. In 2003, 25% of NDAs approved in Japan were based on a bridging strategy. Dr. Kato reported at the ICH Sixth International Conference on Harmonization (Osaka, Japan, November 2003) that at least 67 submissions had involved the application of the E5 guideline. Of course, this number does not include the bridging studies performed from 1998 to 2003 for development programs for which the NDA had not been submitted.

The experience of these early bridging studies remains instructive. We will cite examples of success and illustrate challenges encountered.

### 8.1 Examples of Success

The first bridging study for which I was the principal investigator was performed in July 1998, shortly after the E5 guidelines were adopted. We studied the single- and multiple-dose pharmacokinetic profiles of oseltamivir phosphate (Tamiflu), an antiviral product to treat acute influenza infection, in 24 healthy young males—12 Japanese and 12 Caucasians. The Japanese were all Issei. The data showed that

pharmacokinetics were similar in both populations and no safety issues were observed. These data allowed the sponsor to implement a clinical trial program in Japan to demonstrate safety and efficacy during the 1998–1999 and 1999–2000 influenza seasons, and this product was approved in Japan in 2000 with the bridging study allowing use of safety and efficacy data generated in the United States and the European Union coupled with data from the studies in Japan. Early approval and introduction into clinical use has potentially saved many lives in Japan, which would have been otherwise lost to influenza. Early approval would not have been possible without the use of bridging data.

A recent review selected 26 total NDAs approved in Japan from 1999 to 2003 based on a bridging strategy. This list only included new drugs. Target clinical domains vary considerably and include cancer, viral infections, diabetes, osteoporosis, rheumatoid arthritis, migraine headache, age-related macular degeneration, glaucoma, Parkinson's disease, renal transplant rejection, allergic rhinitis, erectile dysfunction, and *Helicobacter pylori* eradication. Six of these compounds start at lower doses or limit the highest dose. Others have differences in labeling. Examples have been cited previously in this chapter.

The bridging strategy for these compounds typically included an ADME study coupled with one or more phase II/III safety and efficacy studies performed in Japan, with their design based on foreign data and the results of the ADME study. The ADME study enrolled healthy subjects in 12 of these 26 cases, healthy subjects plus patients in 4 cases, and 5 cases used an independent ADME study plus measuring the PK in their efficacy trials.

The bridging strategy "worked" in terms of the goal to save time and money. Measuring the clinical development period from the time when a phase II/III efficacy and safety study began in Japan to the time of NDA submission, the median time for NDAs using a bridging strategy was 32 months versus 56 months for compounds without a bridging strategy, a savings of 24 months [55].

## 8.2 Challenges

Discussion at the ICH Sixth International Conference on Harmonisation also included some of the challenges and difficulties encountered with some bridging programs. All concerned recognized there was some misunderstanding and confusion regarding the E5 guidelines. Attempts to clarify included a Questions and Answers document approved in November 2003, based on actual experience.

In short, issues included differences of opinion between industry and the MHLW on what constituted intrinsic and extrinsic differences, assessment of a compound's "ethnic sensitivity," how to interpret GCP compliance, definitions of disease and medical practice in defining patient populations, criteria for evaluating risk–benefit ratio if disease incidence varied between regions, labeling for multiple indications once bridging established efficacy for the primary indication, criteria for determining if foreign data met regulatory requirements for acceptability, and mutally acceptable surrogate endpoints when assessing efficacy. Still, industry felt that the E5 guidance overall saved them time and money in their development program.

## 9 SUMMARY AND FUTURE DIRECTIONS

Again we ask the question: "Do bridging studies work?" The examples provided illustrate that the answer to this is "yes," that bridging studies have allowed use of existing data with an abbreviated and streamlined drug development program in the new jurisdiction, saving both time and money in NDA approval.

Do bridging studies always work? No, not all have produced the expected outcome for the resources invested.

Let us come back to the question posed at the beginning of this chapter: Are "ethnic differences" real?

My answer to this is "yes," at least for certain compounds. Some compounds have similar properties across all ethnic groups studied, and individual intraethnic variation is greater than interethnic variation. Yet for other compounds, we see characteristic pharmacokinetic or pharmacodynamic differences in one ethnic group that are quite distinct from our observations in other ethnic groups. In some cases, we have defined one or more mechanisms that account for these findings. In other cases, there are multiple mechanisms, or we do not have a clear explanation for all of our findings. We may not be able to distinguish the particular contribution of genetics versus diet, environment, lifestyle, medical practice, or the interaction of these factors within a region, all cumulatively considered "ethnic differences." Yet we know that certain medications need to be used differently in one region than another.

As we continue our investigations, it is prudent to respect the unique culture and traditions of each jurisdiction. While we are answering scientific questions, it is prudent to design our studies with awareness of the historical, social, political, and medical context affecting the respective regulatory agencies.

## REFERENCES


1. ICH Efficacy Topic ES (1998), U.S. *Federal Register*, June.
2. Kimura, R. (1995), History of medical ethics: contemporary Japan, *Encyclopedia of Bioethics*, rev. ed., edited by W. T. Reich, pp. 1496–1505.
3. FDA Backgrounder (2002), Milestones in U.S. Food and Drug Law History, updated Aug 5; www.FDA.gov/opacom/backgrounders/miles.html.
4. Kim, R. B., Yamazaki, H., Chiba, K., et al. (1996), In vivo and in vitro characterization of CYP2E1 activity in Japanese and Caucasians, *J. Pharmacol. Exp. Ther.*, 279(1), 4–11.
5. Shu, Y., Cheng, Z. N., Liu, Z. Q., et al. (2001), Interindividual variations in levels and activities on cytochrome P-450 in liver microsomes of Chinese subjects, *Acta Pharmacol. Sin.*, 22(3), 283–288.
6. Mizutani, T. (2003), PM frequencies of major CYPs in Asians and Caucasians, *Drug Metab. Rev.*, 35(2–3), 99–106.
7. Solus, J. F., Arietta, B. J., Harris, J. R., et al. (2004), Genetic variation in eleven phase I drug metabolism genes in an ethnically diverse population, *Pharmacogenomics*, 5(7), 895–931.
8. Inoue, K., Yamazaki, H., and Shimada, T. (2000), Characterization of liver microsomal 7-ethoxycoumarin *O*-deethylation and chlorzaxozone 6-hydroxylation activities in

Japanese and Caucasian subjects genotyped for CYP2E1 gene, *Arch. Toxicol.*, 74(7), 372–378.

8a. Inoue, K., Asao, T., and Shimada, T. (2000), Ethnic-related differences in the frequency distribution of genetic polymorphisms in the CYP1A1 and CYP1B1 genes in Japanese and Caucasian populations, *Xenobiotica*, 30(3), 285–295.

9. Ham, X. M., Chen, S. P., Wu, Q. N., et al. (2000), G-2964A and C734A genetic polymorphisms of CYP1A2 in Chinese population, *Acta Pharmacol. Sin.*, 21(11), 1031–1034.

10. Aitchison, K. J., Gonzalez, F. J., Quattrochi, L. C., et al. (2000), Identification of novel polymorphisms in the 5′ flanking region of CYP1A2, characterization of interethnic variability, and investigation of their functional significance, *Pharmacogenetics*, 10(8), 695–704.

11. Bartoli, A., Xiaodong, S., Gatti, G., et al. (1996), The influence of ethnic factors and gender on CYP1A2-mediated drug disposition: A comparative study in Caucasian and Chinese subjects using phenacetin as a marker substrate, *Ther. Drug Monit.*, 18(5), 586–591.

12. Inoue, K., Yamazaki, H., Imiya, K., et al. (1997), Relatioship between CYP2C9 and 2C19 genotypes and tolbutamide metyl hydroxylation and S-mephenytoin 4′-hydroxylation activities in livers of Japanese and Caucasian populations, *Pharmacogenetics*, 7(2), 103–113.

13. Caro, A. A., and Cederbaum, A. I. (2005), Inhibition of CYP2E1 catalytic activity in vitro by S-adenosyl-L-methionine, *Biochem. Pharmacol.*, 69(7), 1081–1093.

14. Kessova, I., and Cederbaum, A. I. (2003), CYP2E1: Biochemistry, toxicology, regulation and function in ethanol-induced liver injury, *Curr. Mol. Med.*, 3(6), 509–518.

15. Le Marchand, L., Donlon, T., Seifried, A., et al. (2002), Red meat intake, CYP2E1 genetic polymorphisms, and colorectal cancer risk, *Cancer Epidemiol. Biomarkers Prev.*, 11(10Pt1), 1019–1024.

16. Zhou, S., Yung Chan, S., Gher Goh, B., et al. (2005), Mechanism-based inhibition of cytochrome P450 3A4 by therapeutic drugs, *Clin. Pharmacokinet.*, 44(3), 279–304.

17. He, P., Court, M. H., Greenblatt, D. J., et al. (2005), Genotype-phenotype associations of cytochrome P450 3A4 and 3A5 polymorphism with midazolam clearance in vivo, *Clin. Pharmacol. Ther.*, 77(5), 373–387.

18. Eap, C. B., Buclin, T., Hustert, E., et al. (2004), Pharmacokinetics of midazolam in CYP3A4- and CYP3A5-genotyped subjects, *Eur. J. Clin. Pharmacol.*, 60(4), 231–236.

19. Lamba, J. K., Lin, Y. S., Schuetz, E.G., et al. (2002), Genetic contribution to variable human CYP3A-mediated metabolism, *Adv. Drug Deliv. Rev.*, 54(10), 1271–1294.

20. Harris, R. Z., Jang, G. R., and Tsunoda, S. (2003), Dietary effects on drug metabolism and transport, *Clin. Pharmacokinetic*, 42(13), 1071–1088.

21. Yoshida, A. (1992), Molecular genetics of human aldehyde dehydrogenase, *Pharmacogenetics*, 2(4), 139–147.

22. Pande, J. N., Pande, A., and Singh, S. P. (2003), Acetylator status, drug metabolism and disease, *Natl. Med. J. India*, 16(1), 24–26.

23. Sekine, A., Saito, S., Iida, A., et al. (2001), Identification of single-nucleotide polymorphisms (SNPs) of human *N*-acetyltransferase genes NAT1, NAT2, AANAT, ARD1 and L1CAM in the Japanese population, *J. Human Genet.*, 46(6), 314–319.

24. Krynetski, E. Y., and Evans, W. E. (2000), Genetic polymorphism of thiopurine *S*-methyltransferase: Molecular mechanisms and clinical importance, *Pharmacology*, 61(3), 136–146.

25. Mackenzie, P. I., Miners, J. O., and McKinnon, R. A. (2000), Polymorphisms in UDP glycyronosyltransferase genes: Functional consequences and clinical relevance, *Clin. Chem. Lab. Med.*, 38(9), 889–892.
26. Bailey, D. G., Malcom, J., Arnold, O., et al. (1998), Grapefruit juice-drug interactions, *Br. J. Clin. Pharmacol.*, 46(2), 101–110.
27. Zhou, S., Lim, L. Y., and Chowbay, B. (2004), Herbal modulation of P-glycoprotein, *Drug Metab. Rev.*, 36(1), 57–104.
28. Izzo, A. A. (2004), Drug interactions with St. John's Wort (*Hypericum perforatum*): A review of the clinical evidence, *Int. J. Clin. Pharmacol. Ther.*, 42(3), 139–148.
29. Zou, L., Harkey, M. R., and Henderson, G. L. (2002), Effects of herbal components on cDNA-expressed cytochrome P450 enzyme catalytic activity, *Life Sci.*, 71(13), 1579–1589.
30. Foster, B. C., Vandenhoek, S., Hanna, J., et al. (2003), In vitro inhibition of human cytochrome P450-mediated metabolism of marker substrates by natural products, *Phytomedicine*, 10(4), 334–342.
31. Unger, M., and Frank, A. (2004), Simultaneous determination of the inhibitory potency of herbal extracs on the activity of six major cytochrome P450 enzymes using liquid chromatography/mass spectrometry and automated online extraction, *Rapid Commun. Mass Spectrom.*, 18(19), 2273–2281.
32. Stupans, L., Tan, H. W., Kirlich, A., et al. (2002), Inhibition of CYP3A-mediated oxidation in human hepatic microsomes by the dietary derived complex phenol, gallic acid, *J. Pharm. Pharmacol.*, 54(2), 269–275.
33. Hu, Z., Yang, X., Ho, P. C., et al. (2005), Herb-drug interactions: A literature review, *Drugs*, 65(9), 1239–1282.
34. Dresser, G. K., Bailey, D. G., Leake, B. F., et al. (2002), Fruit juices inhibit organic anion transporting polypeptide-mediated drug uptake to decrease the oral availability of fexofenadine, *Clin. Pharmacol. Ther.*, 71(1), 11–20.
35. Wang, X., Wolkoff, A. W., and Morris, M. E. (2005), Flavonoids as a novel class of human organic anion transporting polypeptide OATP1B1 (OATP-C) modulators, *Drug Metab. Dispos.*, 33(11), 1666–1672.
36. Nabekura, T., Kamiyama, S., and Kitagawa, S. (2005), Effects of dietary chemopreventive phytochemicals on P-glycoprotein function, *Biochem. Biophys. Res. Commun.*, 327(3), 866–870.
37. Nguyen, H., Zhang, S., and Morris, M. E. (2003), Effect of flavonoids on MRP1-mediated transport in Panc-1 cells, *J. Pharm. Sci.*, 92(2), 250–257.
38. Zhang, S., and Morris, M. E. (2003), Effects of the flavonoids biochanin A, morin, phloretin, and silymarin on P-glycoprotein-mediated transport, *J. Pharmacol. Exp. Ther.*, 304(3), 1258–1267.
39. Yoshida, N., Takagi, A., Katazawa, H., et al. (2005), Inhibition of P-glycoprotein-mediated transport by extracts and monoterpenoids contained in *Zanthoxyli fructus*, *Toxicol. Appl. Pharmacol.*, 209(2), 167–173.
40. Wilkinson, G. R. (1997), The effects of diet, aging and disease-stats on presystemic elimination and oral drug bioavailability in humans, *Adv. Drug Deliv. Rev.*, 27(2–3), 129–159.
41. Zhao, F., Loke, C., Rankin, S. C., et al. (2004), Novel CYP 2C9 genetic variants in Asian subjects and their influence on maintenance warfarin dose, *Clin. Pharmacol. Ther.*, 76(3), 210–219.
42. Takahashi, H., Wilkinson, G. R., Caraco, Y., et al. (2003), Population differences in S-warfarin metabolism between CYP2C9 genotype-matched Caucasian and Japanese patients, *Clin. Pharmacol. Ther.*, 73(3), 253–263.

43. Maxwell, T. J., Ameyaw, M. M., Pritchard, S., et al. (2005), Beta-2 adrenergic receptor genotypes and halotypes in different ethnic groups, *Int. J. Mol. Med.*, 16(4), 573–580.
44. Xie, H. G., Kim, R. B., Wood, A. J., et al. (2001), Molecular basis of ethnic differences in drug disposition and response, *Annu. Rev. Pharmacol. Toxicol.*, 41, 815–850.
45. Xie, H. G., Stein, C. M., Kim, R. B., et al. (1999), Frequency of functionally important beta-2 adrenoceptor polymorphisms varies markedly among Africa-American, Caucasian and Chinese individuals, *Pharmacogenetics*, 9(4), 511–516.
46. Zhou, H. H., Koshakji, R. P., Silberstein, D. J., et al. (1989), Altered sensitivity to and clearance of propranolol in men of Chinese descent as compared with American whites, *N. Engl. J. Med.*, 320(9), 565–570.
47. Williams, R. B., Marchuk, D. A., Gadde, K. M., et al. (2003), Serotonin-related gene polymorphisms and central nervous system serotonin function, *Neuropsychopharmacology*, 28(3), 533–541.
48. Crowley, J. J., Oslin, D. W., Patkar, A. A., et al. (2003), A genetic associaton study of the mu receptor and severe opioid dependence, *Psychiatr. Genet.*, 13(3), 169–173.
49. Klepstad, P., Dale, O., Skorpen, F., et al. (2005), Genetic variability and clinical efficacy of morphine, *Acta Anaesthesiol. Scand.*, 49(7), 902–908.
50. Ide, S., Kobayashi, H., Tanaka, K., et al. (2004), Gene polymorphisms of the mu opioid receptor in methamphetamine abusers, *Ann. N. Y. Acad. Sci.*, 1025, 316–324.
51. Tan, E. C., Tan, C. H., Karupathivan, U., et al. (2003), Mu opioid receptor gene polymorphisms and heroid dependence in Asian populations, *Neuroreport*, 14(4), 569–572.
52. Evans, D. A., McLeod, H. L., Pritchard, S., et al. (2001), Interethnic variability in human drug responses, *Drug Metab. Dispos.*, 29(4 Pt 2), 606–610.
53. van Gerven, J. M., Uchida, E., Uchida, N., et al. (1998), Pharmacodynamics and pharmacokinetics of a single oral dose of nitrazepam in healthy volunteers: An interethnic comparative study between Japanese and European volunteers, *J. Clin. Pharmacol.*, 38(12), 1129–1136.
54. Lenz, T. L., Lenz, N. J., and Faulkner, M. A. (2004), Potential interactions between exercise and drug therapy, *Sports Med.*, 34(5), 293–306.
55. Uyama, Y., Shibata, T., Nagai, N., et al. (2005), Successful bridging straegy based on ICH E5 guideline for drugs approved in Japan, *Clin. Pharmacol. Ther.*, 78(2), 102–113.

# *Brief History of Clinical Trials on Viral Vaccines*

**Megan J. Brooks,[1] Joseph J. Sasadeusz,[1] and Gregory A. Tannock[2]**

[1]*Victorian Infectious Diseases Service, Centre for Clinical Research Excellence in Infectious Diseases, The Royal Melbourne Hospital, Parkville, Victoria, Australia*
[2]*Department of Biotechnology and Environmental Biology, RMIT University, Bundoora, Victoria, Australia*

**Contents**

1 Introduction
2 Recent Trends in Human Vaccine Manufacture
3 Events over 60 Years that Have Contributed to Changes to Regulatory Environment for Development of Human Viral Vaccines
4 Expanded Programs for Use of Vaccines in Developing Countries
5 Clinical Studies Necessary for Development of New Vaccines
6 The Future
References

## 1 INTRODUCTION

Although our recognition of viruses as the etiologic agents of infectious disease only dates from the early years of the twentieth century, it is interesting to note that vaccines against smallpox and rabies were introduced much earlier. In the case of smallpox the procedure referred to as *variolation*, in which skin lesions from infected individuals were administered to those without lesions, dates back to the late Middle Ages [1]. Despite this, the observation by Jenner in 1798 that milkmaids continuously exposed to cows infected with cowpox were largely resistant to endemic

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

smallpox is rightly regarded as the origin of modern vaccination practice. Rabies vaccines, consisting of formolized suspensions of infected animal brain material and introduced in the late nineteenth century, were highly reactogenic but were used as recently as the 1960s when they were replaced by vaccines prepared from viruses grown in cell culture [2].

Most viral vaccines in use today were developed as a consequence of our ability to cultivate viruses in embryonated eggs (from the 1930s) or cell cultures (the 1950s and 1960s) and so fulfill the second of Koch's postulates for establishment of the etiology of an infectious disease. Until their discontinuation for general use in the early 1980s, vaccines against smallpox were prepared from vaccinia virus grown on the skin of calves, a procedure that would now be completely unacceptable. In vitro cultivation of bacteria was first achieved in the late nineteenth century, and many highly effective bacterial vaccines used today (notably those for the prevention of tetanus and diphtheria) were available by the 1920s. For other significant human viruses, such as hepatitis B, C, and the human papilloma viruses, the second of Koch's postulates has never been fulfilled; until the late 1980s the antigens that comprise the active ingredients of hepatitis B vaccines were actually prepared from chronic hepatitis B carriers. Fortunately, over the past 15 years hepatitis B vaccines have been prepared by the application of molecular techniques and are far more acceptable from the standpoint of safety and at the same time much less expensive. Molecular technology has allowed the realization of preventive vaccines against human papilloma viruses [3]. However, despite much effort and an urgent need, vaccines against hepatitis C (HCV) or the human immunodeficiency virus (HIV) seem much further away. For other reasons, urgently needed vaccines against herpes simplex and cytomegaloviruses, both of which can be grown in cell culture, are still unavailable, and the same could have been said until recently of rotaviruses.

## 2 RECENT TRENDS IN HUMAN VACCINE MANUFACTURE

Greater concerns about safety over the past 30 years have increased the costs of viral vaccine development to the point where all but a handful of very large multinational companies have the resources to fully comply with all the steps necessary for registration in the two major regulatory jurisdictions, the Food and Drug Administration (FDA) and the European Union. These costs have been largely driven by the willingness of U.S. courts to grant extraordinary damages against manufacturers for claims, some inspired by negligence but others due to historical gaps in our knowledge at the time of the original product registration. Examples of the latter will be given below. Until the last 20 years human viral and bacterial vaccines were prepared in Western countries by about 30–40 medium to large sized companies and a comparable number of state-run institutions or statutory corporations. The *state serum institutes* of several European countries were founded early in the twentieth century and are good examples of the latter. Their existence followed advances in microbial cultivation and the identification of blood groups around the turn of the twentieth century and the consequent development of vaccines and serum therapies, largely for use within the same country. Similar institutions had a

prominent role in the countries of the former Soviet Bloc, with their strong emphasis on immunization and public health planning. In several countries of the developing world similar institutes undertook vaccine manufacture under the aegis of colonial powers, as exemplified by the various Pasteur Institutes of some Francophone countries of Africa and Asia. The importance of these institutes in the provision of biologics was greatly enhanced by two world wars. In Australia, for example, the Commonwealth Serum Laboratories (now CSL Ltd) was instituted by the government of the day in 1917 for the express purpose of supplying smallpox vaccine for use by the Australian army in World War I in response to a likely shortfall in supply from manufacturers in the United Kingdom. CSL Ltd has greatly expanded its range of vaccines and other biologics since 1917 but is no longer a government-owned institution. As with comparable organizations, nearly all of its output, until comparatively recently, was reserved for use within Australia. The *raison d'etre* for their survival over almost a century has been based primarily on perceptions of national interest!

Over the past 20–30 years most small manufacturers of human vaccines have either disappeared and/or have been subsumed by or developed strategic alliances with four giant multinational companies (*Bigpharma*). Only Bigpharma has the resources to undertake the complete development of human vaccines from early research to final clinical application, although, for many vaccines, early research relating to proof of concept is frequently undertaken by smaller start-up companies—often in association with research institutes. Because of the international nature of their operations, vaccine manufacture and accreditation of existing vaccines by Bigpharma may be carried out in more than one country. The cornerstone of modern manufacture has been the emergence of rigidly enforced international codes of good manufacturing and laboratory practice (GMP and GLP). Most countries with only limited capacity to manufacture human viral vaccines may, nevertheless, impose regulatory conditions for efficacy, safety, and standardization that are relevant to local needs. In practice, because of the sheer cost and complexity of manufacture, requirements for registration by the federal drug authority in the United States and the European Union [often with guidelines set by the World Health Organization (WHO)] remain points of reference for most national regulatory authorities.

## 3 EVENTS OVER 60 YEARS THAT HAVE CONTRIBUTED TO CHANGES TO REGULATORY ENVIRONMENT FOR DEVELOPMENT OF HUMAN VIRAL VACCINES

1. *Contamination of Yellow Fever Vaccine by Hepatitis B by Serum Preservatives (1940s)* The development of live vaccines against yellow fever and, in particular, the introduction of the egg-grown attenuated 17D strain by Theiler's group at the Rockerfeller Institute just prior to World War II was a major achievement in human medicine. By April 1942 some 7 million doses had been administered to the U.S. Armed Forces alone, using vaccines prepared by the Rockerfeller Institute. Yellow fever viruses are highly labile, and thermostability was conferred by the addition of human serum to the vaccine, at a time when nothing was known of the etiology of

the hepatitis B virus. By December 1942 over 50,000 cases of hepatitis and 84 deaths were recognized from infection arising from contaminated lots, representing some 2.5 million administered doses. The involvement of human serum was quickly recognized, despite the absence of definitive tests (which only became available in the 1970s), and the problem was overcome by elimination of the serum from the vaccine [4].

2. *Cutter Incident: Release of Inactivated Poliomyelitis Vaccines (IPV) Containing Residual Amounts of Live Poliovirus (1955)* Soon after the commencement of programs for the mass immunization of children in the United States, a batch of IPV, prepared by Cutter Laboratories in California, was released and subsequently shown to be responsible for poliomyelitis in 79 vaccinated children, 105 family members, and 20 community members [5]. Despite the overwhelming success of the immunization program, the *Cutter incident* and the investigations that followed provided the catalyst for the introduction of powerful national agencies for the registration of vaccines.

3. *Contamination of Early Batches of IPV by Simian Virus 40 ($SV_{40}$)* Between 1955 and 1963 successive batches of IPV prepared in primary monkey kidney cultures were shown, retrospectively, to have been contaminated with $SV_{40}$, a commensal of certain species of monkeys and referred to as the vacuolating agent from the cytopathic effect produced after inoculation to certain cell lines. In the United States alone, it was estimated that 10–30 million individuals were inoculated with contaminated vaccines. Because $SV_{40}$ virus is a member of the family Papovaviridae, members of which are oncogenic for several animal species, concerns were expressed about long-term risks to vaccine recipients. Some 30 years later and, after comprehensive epidemiological studies in the United States and Europe, the balance of evidence indicates no causal relationship between the receipt of contaminated vaccine and cancer [6].

4. *Enhancement of Pathogenicity in Infants Receiving Experimental Formalin-Inactivated Respiratory Syncytial Virus (RSV) Vaccine* Respiratory syncytial virus is the most significant respiratory pathogen of the neonatal child for which a vaccine is urgently required. In a traditional approach to control, a formalin-inactivated vaccine was administered parenterally to infants in 1968 at the Children's Hospital in Washington, D.C. Results of the study indicated that not only was the vaccine unprotective but recipients who were subsequently subjected to natural challenge experienced enhanced immunopathology in the form of Arthus-type hypersensitivity reactions [7]. These results have had profound consequences for the development of effective vaccines against RSV, and most efforts since have been directed toward the use of intranasally administered live attenuated vaccines. Experimental vaccines for use in neonatal children present unusual difficulties both in immunologic terms and from an ethical standpoint. Despite considerable progress, vaccines against RSV are still unavailable almost 40 years after the Washington study.

5. *Molecular Instability of Oral Poliomyelitis Vaccines (OPVs): Recognition of Vaccine-Associated Paralysis (1970s)* Oral poliomyelitis Vaccines have been crucial for the elimination of poliomyelitis from all but a few regions of a very small number of countries, under programs sponsored by the WHO in recent years [8].

These vaccines were developed in the early 1960s at a time virtually nothing was known as to the genetic stability of their live virus components. Molecular techniques developed over the past 20 years now indicate instability for some of these viruses following human passage, which is associated with reversion to neurovirulence [9]. If molecular evidence for instability had been available in the 1960s, OPVs would almost certainly not have been licensed and there would have been little prospect of eliminating poliomyelitis from the world some 40 years later. IPVs can only be afforded in developed countries where they have largely replaced OPVs in order to overcome concerns associated with instability.

6. *Association between Guillain–Barré Syndrome (GBS) and Receipt of Swine Influenza Vaccine (1976–1977)* In 1976 following the death of a soldier at Fort Dix, New Jersey, after infection with an H1N1 (swine) virus with apparent antigenic similarities to the agent responsible for the 1918–1919 pandemic, a decision was made to offer the entire U.S. population specially prepared monovalent vaccines. The program was curtailed after an association was recognized that resulted in the attribution of 8–10 excess cases of GBS per million during the first 6 weeks after administration of the A/swine vaccine [10]. For reasons unknown, the incidence of GBS in subsequent vaccines has been much lower, and egg-grown inactivated vaccines remain the principal public health measure for the prevention of influenza.

7. *Possible Association between Intussusception and Administration of Oral Rotavirus Vaccines (1999)* A quadrivalent reassortant simian rotavirus vaccine was licensed in 1998 for manufacture and distribution by Wyeth Laboratories and approved by the FDA after extensive testing in several countries. The vaccine was prepared for administration to infants in 3 doses at 2, 4, and 6 months. Within 12 months, subsequent evaluation of U.S. clinical data suggested a small increase in the frequency of intussception in infants who received the vaccine, especially after the first dose [11]. Despite an urgent public health need for effective preventive measures against rotaviruses and developmental research programs extending over many years that were largely publicly funded, the vaccine was withdrawn. Fortunately, a reassortant vaccine of bovine origin and prepared by Merck was approved by the FDA in 2006, which, it is hoped, will have an impact on neonatal rotavirus infections, especially in the developing world [12].

8. *Association between Intranasal Administration of Inactivated Influenza Vaccine and Bell's Palsy (2000)* An influenza vaccine was prepared by a Swiss company (Berna Biotech) in which vaccine surface antigens were incorporated into carrier liposomes. The heat-labile *Escherichia coli* enterotoxin, a powerful mucosal adjuvant, was incorporated into the formulation. After its introduction in Switzerland, the risk of Bell's palsy was increased by a factor of 19 among those who received the vaccine intranasally. An increase did not occur in recipients who received the vaccine by parenteral injection [13]. These adverse events were not detected in prelicensure trials and the vaccine has since been withdrawn.

The list is incomplete but these events have profoundly affected the regulatory environment and have been accompanied over the same time by a vast increase in our understanding of viral replication and pathogenesis, with flow-on consequences

for product liability. Consequent advances over the period have also occurred in vaccine safety testing and in the development of models for vaccine evaluation in test populations that allow the power of a study to be estimated well in advance of its commencement. We now have a vastly increased array of standardized tests, both biological and molecular for estimating (1) the active ingredient(s) of a vaccine, (2) the presence of known adventitious contaminants, and (3) the likely toxicological consequences of its use in humans. However, the need still exists for extensive and costly clinical testing in a climate where ethical constraints on the use of volunteers for clinical trials have been greatly increased in the years immediately before and after the Declaration of Helsinki in 1964 [14]. Many trials that were conducted on earlier vaccines would be simply disallowed these days by institutional review boards, which place strong emphasis on need to provide comprehensive information on the nature of a trial to each volunteer at the point of entry. They also require to be adequately briefed concerning inducements, financial and otherwise, that are provided to volunteers. Overall, the manufacture of viral vaccines and other biologics has become commercially unattractive for all but a handful of very large manufacturers. As a consequence, the role of governments in sponsoring and meeting the costs of developing new vaccines has increased greatly.

Many highly successful vaccines for the prevention of measles, poliomyelitis, rubella, mumps, and yellow fever were developed in a much less stringent regulatory environment, and it is arguable whether many would have been registered today. The situation with OPVs has already been discussed but, to take another example, measles vaccines were first developed in the early 1960s at a time when very little was known about subacute sclerosing panencephalomyelitis (SSPE), a rare and unpredictable immunologic complication of measles associated with chronic neural infection. Effective vaccines against measles consist of live virus preparations and, if SSPE had been widely recognized at the time, the onus would have been placed on vaccine developers to prove that mass administration was not associated with SSPE. Given that there are still no satisfactory in vitro markers for SSPE, this would have been an impossible task and measles vaccines would probably never have become available.

Over the same period and in a very different regulatory environment, the countries of the former Soviet Bloc have made enormous and often unrecognized contributions to vaccinology. Nowhere was this more evident than in the adoption of OPV for the prevention of poliomyelitis in Western countries in the early 1960s, following the conduct of very large field trials under the direction of Smorodintsev and his colleagues in the former Leningrad [15]. The recent reemergence of childhood infectious diseases, such as diphtheria, in former Soviet countries is ample testimony to the success of former state-sponsored programs, many of which have not been maintained over the past 10–15 years. [16].

## 4 EXPANDED PROGRAMS FOR USE OF VACCINES IN DEVELOPING COUNTRIES

The global program sponsored by the WHO for the elimination of smallpox between the mid-1960s and the early 1980s is widely regarded as perhaps the greatest medical achievement of the twentieth century. Its success has removed the need for smallpox

vaccination—often associated with unacceptable adverse reactions—and delivered significant economic benefits, mainly to the developed world, by eliminating the need to maintain vaccination programs.

Because of the success of that program, several initiatives have commenced over the past 20 years for the elimination of pediatric viral diseases, using vaccines widely used in Western countries. These programs have been supported by philanthropic organizations, such as the Gates and Rockerfeller Foundations, usually in close collaboration with the WHO. Mention has already been made of programs for the elimination of poliomyelitis by the strategic use of OPV. Another program is concerned with immunization against measles, a cause of high mortality in developing countries especially in association with malnutrition. Logistical problems in this and similar programs arise from difficulties in vaccine delivery due to poor infrastructure and problems in maintenance of the *cold chain.* Protective responses to the vaccine may be diminished by the presence of maternal antibody that is associated with high levels of measles endemicity. Here, the window of opportunity to vaccinate a child after the decline of maternal antibody but before natural exposure to the virus is often relatively narrow. It should be remembered that the use of combination measles–mumps–rubella vaccines in infants was only possible in developed countries by the elimination of endemic measles by earlier vaccination programs! Reservoirs of measles, poliomyelitis, and other preventable pediatric viruses are much more widespread, and these viruses are responsible for much higher rates of subclinical disease than smallpox viruses. Vaccination programs against these viruses will, therefore, need to be maintained until well after the elimination of reported clinical disease.

## 5 CLINICAL STUDIES NECESSARY FOR DEVELOPMENT OF NEW VACCINES

Figure 1 provides an indication of the cost and time scale of any program for the development and registration of a hypothetical new vaccine for the prevention of a specific viral disease. The entire process is broadly divided into four phases of

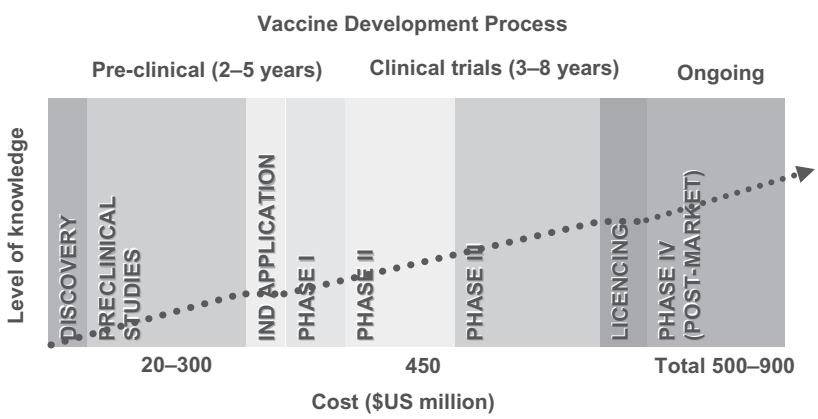


***FIGURE 1*** *Vaccine development process.*

variable length and complexity, which relate to the virus itself, its pathogenicity, and technical impediments that may stand in the way of evaluating its effectiveness as a preventive measure. For the registration of many vaccines, data from phases II and III or II and IV trials are considered as single categories by regulators. These studies involve the participation of increasing numbers of volunteers in clinical trials and are the most expensive part of the registration process. They are generally beyond the resources of most small-to medium-sized manufacturers and research institutes and can only be undertaken by BigPharma to achieve near-simultaneous registration in many countries.

Before clinical studies can commence, a dossier of preclinical study results, and the proposed protocols for human studies are submitted to the regulatory agencies. For the FDA, an investigational new drug (IND) application is made. Clinical trials are classified as phase I, II, III, or IV, depending on the stage of clinical development of a vaccine. Phase I studies are *first-in-human* studies in which the drug is administered to small numbers of healthy volunteers to determine safety and tolerability, as well as pharmacological activity. Results can also provide information as to appropriate dose ranges for future studies. If there is a high possibility of toxic side effects, then the drug may be only administered to the patients for whom it is primarily intended. Therapeutic vaccines against HIV and hepatitis C could be examples of such an approach. Phase I studies are usually conducted in purpose-built facilities that allow close monitoring of patients.

The principal purpose of a phase II study is to determine vaccine efficacy in small groups of the targeted patients, although safety data is also collected. A range of doses is generally used to determine the maximum tolerated dose and the therapeutic range. Phase III studies are conducted in large groups of patients if phase II studies indicate that the drug is safe and has potential clinical benefit. Efficacy data continues to be collected in phase III studies, together with data on the incidence and nature of adverse effects. Phase IIIb studies may be undertaken while the application for licensing is being assessed to provide further safety and/or efficacy data (periapproval studies).

Once approval for marketing is granted, phase IV studies are conducted to determine efficacy in the typical clinical setting (using inclusion/exclusion criteria less stringent than those in the protocols for the phase III studies). Postmarketing surveillance is conducted to track adverse effects and studies comparing the new drug to existing therapies, or its role in combination therapy, may be conducted as part of phase IV studies.

## 6 THE FUTURE

The advent of recombinant deoxyribonucleic acid (DNA), together with the emergence of HIV and, more recently, HCV viruses as major human pathogens have given new impetus over the past 20 years to technologies associated with viral vaccine development. Ironically, the very success of commonly used vaccines for the prevention of poliomyelitis, rubella, mumps, and yellow fever was a major reason for a diminution of interest in vaccine technology from the mid-1960s to the early 1980s. Both HIV and HCV have been targets for an unprecedented increase in research funding on vaccines in recent years. However, for several technical reasons

not discussed here, vaccines are still unavailable against either agent, either for preventive or therapeutic use.

In recent years great changes have occurred in most countries following the acceptance by governments and other public bodies of the need to maintain vaccination programs as a basic human right. These changes have been accompanied by commensurate improvements to vaccine production, which should have resulted in improved efficiencies and lower costs and a reduction in the time and cost required for the registration of new vaccines. Unfortunately, this has not been the case. Regulatory agencies are increasingly able to fast-track approvals for new products for use in situations of immediate need, such as for the development of avian influenza vaccines, that make use of newer molecular and older technologies. However, the majority of new vaccines are only approved many years after an etiology has been established for a previously unrecognized disease or a novel approach has been put forward as an alternative to a new vaccine. Contemporary examples of these delays include (1) live attenuated vaccines against influenza, which were first proposed in the 1970s and achieved registration for restricted use by the FDA in 2002 and (2) rotavirus vaccines, referred to above, in 2006—some 33 years after an etiology was established! Costs associated with bringing a new vaccine to the community are reported to be $US500–900 million, although BigPharma often avoid disclosing specific details, citing commercial sensitivities. National regulatory agencies are highly risk averse and are responsible for ensuring the quality, efficacy, safety, and timely availability for vaccines for the community, through a uniform system of regulatory controls. They are involved in all aspects of vaccine development from initial discovery to postapproval surveillance for adverse effects in close consultation with vaccine developers. The reasons for their increased involvement are historic and have been outlined above. However, it is fair to state that many developers of highly successful pediatric viral vaccines would have experienced great difficulties in achieving registration today. It is also highly problematic whether the manufacture of such vaccines in the current regulatory environment would be profitable, even allowing for the economies of scale enjoyed by BigPharma.

While the prevention of viral diseases by vaccination has been a singular triumph for medical science, many challenges still exist. Quite apart from HIV and HCV, vaccines are still unavailable for the prevention of most respiratory virus infections, and improvements to existing influenza vaccines are universally regarded as a matter of urgency. Vaccines are also unavailable for the prevention of all human (but not veterinary) herpesvirus infections, again despite an urgent need. Much work needs to be done to expand the number of adjuvants to improve the performance of existing and future vaccines (at present only alum is acceptable), despite the formidable problems that have been encountered over the past 20 years in this area. It would be a great pity if the constraints referred to above gave rise to a diminution in the forces of innovation that have been responsible for so many improvements to public health over the past century.

## REFERENCES

1. Blake, J. B. (1959), *Public Health in the Town of Boston, 1630–1832*, Harvard University Press, Cambridge, MA.

2. Dreesen, D. W. (1997), A global review of rabies vaccines for human use, *Vaccine*, 15(Suppl), S2–6.
3. U.S. Food and Drug Administration (2006), FDA licenses new vaccine for prevention of cervical cancer and other diseases in females caused by human papillomaviruses, FDA 2006-06-08.
4. Seeff, L. B., Beebe, G. W., Hoofnagle, J. H., et al. (1987), A serologic follow-up of the 1942 epidemic of post-vaccination hepatitis in the United States Army, *N. Engl. J. Med.*, 316(16), 965–970.
5. Nathanson, N., and Longmuir, A. D. (1963), The Cutter incident. Poliomyelitis following formaldehyde-inactivated poliovirus vaccination in the United States during the spring of 1955. II. Relationship of poliomyelitis to Cutter vaccine, *Am. J. Hyg.*, 78, 29–60.
6. Strickler, H. D., Rosenberg, P. S., Devesa, S. S., et al. (1998), Contamination of poliovirus vaccines with simian virus 40 (1955–1963) and subsequent cancer rates, *JAMA*, 279(4), 292–295.
7. Kapikian, A. Z., Mitchell, R. H., Chanock, R. M., et al. (1969), An epidemiologic study of altered clinical reactivity to respiratory syncytial (RS) virus infection in children previously vaccinated with in inactivated RS virus vaccine, *Am. J. Epidemiol.*, 89, 405–421.
8. World Health Organization (2003), *Global Polio Eradication Initiative: Strategic Plan* 2004–2008, Geneva.
9. Evans, D. M., Dunn, G., Minor, P. D., et al. (1985), Increased neurovirulence associated with a single nucleotide change in a noncoding region of the Sabin type 3 poliovaccine genome, *Nature*, 314, 548–550.
10. Houff, S. A., Miller, G. L., Lovell, C. R., et al. (1977), The Guillain-Barre syndrome and swine influenza vaccination, *Trans. Am. Neurol. Assoc.*, 102, 120–123.
11. Centers for Disease Control and Prevention (CDC) (1999), Intussusception among recipients of rotavirus vaccine–United States, 1998–1999, *MMWR Morb. Mortal Wkly. Rep.*, 48(27), 577–581.
12. Glass, R. I., and Parashar, U. D. (2006), The promise of new rotavirus vaccines, *N. Engl. J. Med.*, 354, 75–77.
13. Mutsch, M., Zhou, W., Rhodes, P., et al. (2004), Use of the inactivated intranasal influenza vaccine and the risk of Bell's palsy in Switzerland, *N. Engl. J. Med.*, 350(9), 896–903.
14. World Medical Association (2000), *Declaration of Helsinki. Ethical Principles for Medical Research Involving Human Subjects*, Geneva.
15. Smorodintsev, A. A., Davidenkova, E. F., Drobyshevska, Y. A., et al. (1959), Results of a study of the reactogenic and immunogenic properties of live anti-poliomyelitis vaccine, *Bull. World Health Organ.*, 20, 1053–1074.
16. Vitek, C. R., and Wharton, M. (1998), Diphtheria in the former Soviet Union: Reemergence of a pandemic disease, *Emerg. Infectious Dis.*, 4(4), 539–550.

# *Methods of Randomization*

**Gladys McPherson and Marion Campbell**

*Health Services Research Unit, University of Aberdeen, Aberdeen, Scotland*

**Contents**

1 Patients Registration
- 1.1 Patient Identification and Recruitment
- 1.2 Confirmation of Eligibility
- 1.3 Agreement to Randomize
- 1.4 Informed Patient Consent
- 1.5 Formal Entry to Trial
- 1.6 Treatment Assignment
- 1.7 Completion of Relevant Documentation
- 1.8 Commencement of Treatment

2 Generation the Random Sequence
- 2.1 Simple Randomization
- 2.2 Restricted Randomization

3 Covariate-Adaptive Randomization
- 3.1 Stratification
- 3.2 Random Permuted Blocks within Strata
- 3.3 Minimization
- 3.4 Minimization Compared with Simple Randomization
- 3.5 Other Constrained Methods

4 Response-Adaptive Methods

5 Special-Case Scenarios
- 5.1 Unequal Randomization
- 5.2 Cluster Randomization

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

6 Summary
Appendix I
Appendix II
Appendix III
Appendix VI
References

The randomized controlled trial (RCT) is one of the simplest, most powerful, and revolutionary tools of research. In essence, the RCT is a comparison of two or more interventions in which people are allocated at random to a treatment group. Randomization is a method of eliminating bias in the way that treatments are allocated to patients, but it does not guarantee that the characteristics of the groups will be completely balanced; in that if imbalance occurs it will have occurred by chance rather than by the introduction of some systematic bias. Some factors are known in advance to be important risk factors associated with the outcome of the patient. Imbalance in these prognostic variables could have a marked effect on the results of a trial and on their credibility. However, if the randomization is performed fairly, then any difference between groups is due to chance, but the groups might differ in a way that might affect their response to treatment (this is a question of clinical importance rather than statistical significance). While it is possible to modify the statistical analysis to take account of any differences between groups at baseline, it is better to control the problem at the design stage.

There are four classes of randomization procedure: complete (simple) randomization, restricted randomization, covariate-adaptive randomization, and response-adaptive randomization. The simplest randomization procedure is complete randomization with equal allocation and does not depend on prognostic factors or previous treatment assignments. Restricted randomization is used when it is desired to have equal numbers of patients assigned to each treatment group. Covariate-adaptive randomization is used to achieve similar numbers of patients in each treatment group and also to ensure that patient groups are similar with respect to prognostic factors such as age or gender. In response-adaptive randomization the treatment assignments depend upon previous patient responses to treatment. These different approaches to generating the random sequence will be discussed in detail in Sections 2–4.

## 1 PATIENT REGISTRATION

In a randomized controlled trial patients are identified for entry to the trial and, if they are eligible and consent to trial entry, they are then randomized to one of the study treatments. For a patient to be included into a clinical trial the following steps are necessary:

- Patient identification and recruitment
- Confirmation of eligibility
- Agreement to randomize
- Informed patient consent

- Formal entry to trial
- Treatment assignment
- Completion of relevant documentation
- Commencement of treatment

### 1.1 Patient Identification and Recruitment

How potentially eligible patients are to be identified and how and where they are to be approached should be outlined in the study protocol. The study design team should have made these decisions so as to ensure that the study population will be as representative as possible of the particular disease area under investigation to ensure the generalizability of the results. All appropriate patients should be logged even if they are ineligible or missed so that a full record of potential recruits is obtained. This can then be used as a check that systematic selection bias of patients into the trial had not occurred. The recruitment process should be outlined fully and could consist of a script for the recruitment officer to follow.

### 1.2 Confirmation of Eligibility

Eligibility criteria should be agreed upon before recruitment begins and should be recorded in the study protocol. The recruitment officer should check that a potential participant satisfies the eligibility criteria before asking them to consent to the trial. If the patient is recruited and then subsequently found to be ineligible, this should be logged and the decision whether to include them in an intention-to-treat analysis, or exclude them from analysis according to some consistent algorithm, made at a later date. Postrandomization exclusions unrelated to noncompliance, withdrawal, or losses to follow up occur when patients are inappropriately randomized into a clinical trial [1]. These patients can be removed from both study arms without risking bias. The number of patients randomized but not included in the primary analysis should be clearly stated along with the reasons for exclusion.

### 1.3 Agreement to Randomize

It is vitally important that the patient understand the concept of randomization and agree to accept the outcome of the treatment assignment. The recruitment officer will have to explain this very carefully and be satisfied that the patient has a good understanding of the process. Patients may have a preference for one treatment over another, and it is sometimes desirable to record this as it can help to better understand subsequent noncompliance. However, the recruitment officer must be assured that this preference is not so strong that the patient will then not comply with the allocated treatment. In this situation patients are uncertain about the treatment choices and therefore should not agree to be randomized. This concept also applies to any clinician who has a strong treatment preference or who is unwilling to randomize the patients. There are study designs that allow treatment allocation according to either patient or collaborator preference, which may be adopted in order to increase recruitment especially in surgical trials.

### 1.4 Informed Patient Consent

There is a strong ethical and legal requirement to obtain informed consent to a randomized control trial before the patient is randomized. Therefore, it is important that the recruitment officer ensures that the patient fully understand the purpose of the study as well as agreeing to the randomization process. A well-designed patient information leaflet will describe these concepts. It should also outline what patient involvement is expected for the duration of the study (medication to take or surgery to be undertaken, clinics to attend, questionnaires to be filled in, etc.) and what steps should be taken if, at any time, the patient wishes to withdraw from the study. Patients are usually ineligible if they cannot give informed consent. For example, they cannot understand English (if there are no translations available) or they have mental health problems. Such exclusions would be stipulated in the protocol.

### 1.5 Formal Entry to Trial

If all the previous steps have been taken and a patient is identified as being eligible for entry to the trial, then it is necessary at this stage to record some basic demographic details about that patient. The patient's name and address can now be logged for entry into the study database. If any important prognostic factors have been identified and are to be used in the randomization process, the values should be recorded on a designated pro-forma prior to randomization. The patient might be allocated a study number at this stage (or this might be allocated at the time of randomization) if randomization is done by an external randomization service.

### 1.6 Treatment Assignment

Adequate concealment of treatment allocation is widely recognized to be of paramount importance to the scientific validity of any clinical trial [2]. As such, randomization should not be performed by the recruitment staff. Randomization may be performed by staff in the central study office or by an independent trials center offering this service. It is common to use either a randomization list or computer program to do this, and the process involves either sending paper forms to the randomization center for manual data entry and treatment allocation, by telephoning an automated system, or using a Web-based randomization system. Treatment allocation by sealed opaque envelopes is not recommended as this method has been shown not to be tamper-proof (e.g., envelopes have been observed to have been opened out of sequence; and there has been evidence that envelopes have been opened in advance to seek prior knowledge of treatment assignment).

Some common study scenarios are as follows:

(a) *Multicenter Trial with Central Study Office* If the trial design requires immediate treatment assignment (or randomization outside normal working hours), then it is preferable to use an automated telephone randomization service often referred to as an IVR (interactive voice response) system. A benefit of this system is that a full audit of the randomization process is automatically created. The date and time

of randomization is logged, along with at least one patient identifier so that treatment allocations can be proven to be fair and unbiased. IVR systems also help to maintain blinding since there is no staff involvement and no need for code break envelopes. It is necessary, however, to ensure that backup procedures are documented in case of very occasional system failure. It is now routine that these systems can be configured and maintained relatively cheaply, with minimal denial of service achieved for affordable cost.

This system may be backed up by IWR (interactive web response) systems, which may help overcome language difficulties (e.g., understanding of written English can often be easier than listening to a voice over the telephone). The person performing the randomization also has more time to read and digest instructions, and helpful tips can also be included on the page. It is also possible to construct multilingual Web or telephone systems. The use of an IWR system offers further benefits as it is possible to extend the system to include real-time reporting (giving up-to-date information about recruitment, randomization, and patient status). Some recent trials have also used this technology for the collection of electronic patient-reported outcome data [3].

If randomization is required only within normal working hours, then treatment assignments can be read off the randomization list and given to the recruitment officer over the telephone. Ideally, a small, customized computer program would be used to store the list and output the next treatment assignment as required to minimize breaking allocation concealment. If sufficient funds are available, then an automated system is preferred as this does not rely on someone being available to answer the telephone.

If patients are screened first and then only considered for randomization later, for example, after responding appropriately to a screening questionnaire, then an automated system may not be necessary and a computerized randomization system may be written into the trial management program. If the code is kept secure, then this would offer all the benefits of an automated telephone response system without the expense.

The randomization scheme itself could consist of an allocation list constructed from permuted blocks or might incorporate a program using minimization if important prognostic variables had been identified.

(b) *Double-blind Trial of Drug Therapy* For this type of trial design the pharmacist preparing the drugs needs to have access to the treatment allocation. While the randomization lists continue to be created by the study office or the trials unit performing the randomization, the treatment allocation should now also include a corresponding drug treatment identifier that will be used to label the drug. These lists are then sent to the pharmacy where the drugs are prepared and packaged according to the list. Treatment unblinding can be performed by the pharmacy and/or the central randomization service.

When a patient is randomized into the trial, the recruitment officer telephones the automated service. The patient is allocated a treatment allocation number, which is known to be in stock at that site. If the randomization algorithm uses random permuted blocks, then the allocated treatment will simply be the next available one on the list for that site. If a minimization algorithm is being used, then the allocation will be the next available pack of the appropriate drug type. The randomization center must be aware at all times of the current stock levels at each site and an early

warning system can be written into the IVR system to alert the pharmacy when stocks are low.

(c) *Single-Center Trial* For a single-center trial there are several possible solutions. The randomization can be performed by an independent person who is not involved with the trial; a small local computer program that incorporates a full audit trail can be written to perform the randomization; or a central randomizing service may be used if funds permit.

### 1.7 Completion of Relevant Documentation

At this stage in the study the following documentation should be available:

- A log sheet of all patients identified as appropriate for the study
- Ineligibility forms for those not eligible
- Eligibility forms for those considered eligible for entry to the study (with the study number)
- A consent form appropriately signed and dated (with the study number)
- A baseline data form containing basic demographic data (with the study number), previous medical history, relevant information about their clinical condition, and any baseline questionnaires (such as SF36, EQ-5D)
- (*Optional*) A randomization form containing the values for any prognostic factors, the treatment allocation code, and the study number

Any forms should be piloted before the study commences and committed to final printing only after this has proved successful. Consideration must be given as to where these forms are to be stored, and for what length of time, as any patient identifiable material should be kept in a secure environment.

### 1.8 Commencement of Treatment

Treatment should commence as soon as possible after randomization; so in a multitherapy trial each randomization should take place only when the patient is ready to receive that treatment. In a drug therapy trial each dispatch of study treatment requires rigorous control, and batch numbers should be checked and recorded. It is not sufficient to record only the drug pack allocated at randomization; the actual drug pack received should also be recorded.

## 2 GENERATING THE RANDOM SEQUENCE

In Sections 2.1–5 a description of each randomization method including advantages and disadvantages is given. This outlines the resources required, ease of use of method, allocation concealment, and predictability of the method.

In most clinical trials the complete list of participants is not available at the start of the trial, and patients will be recruited and randomized one at a time as they are identified. For both single-center and multicenter trials it is preferable that the randomization is performed by an independent organization or a central study

office to ensure that the randomization is carried out according to the required specification.

## 2.1 Simple Randomization

With simple randomization each patient has a known chance, usually equal, of being given each treatment, and the treatment to be given cannot be predicted in advance. The simplest method is tossing a coin, but one can also use random tables, or a random-number generator in a calculator or computer. Table 1 gives an example of a table of random digits that were generated by computer (see Appendix I for the code). To randomize a patient, (a) choose an arbitrary starting point in the list and (b) work along the rows assigning treatments by using the numbers on the list.

***Example*** For *two* treatments assume the digits 0–4 correspond to treatment A and digits 5–9 correspond to treatment B. Therefore the top row of Table 1 would produce the following treatment assignments:

| | | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 5 | 5 | 6 | 3 | 1 | 3 | 9 | 1 | 1 | 4 | 5 | 5 | 4 | 8 | 1 | 0 | 7 | 3 | 4 | 8 | 3 | 4 | 7 |
| B | B | B | A | A | A | B | A | A | A | B | B | A | B | A | A | B | A | A | B | A | A | B |

For *three* treatments assume the digits 1–3 correspond to treatment A, digits 4–6 correspond to treatment B, and digits 7–9 correspond to treatment C. Ignore 0 (ignoring a digit has no effect on the inherent "randomness" of the sequence). The top row of Table 1 would then produce the following treatment assignments:

| | | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 5 | 5 | 6 | 3 | 1 | 3 | 9 | 1 | 1 | 4 | 5 | 5 | 4 | 8 | 1 | 0 | 7 | 3 | 4 | 8 | 3 | 4 | 7 |
| B | B | B | A | A | A | C | A | A | B | B | B | B | C | A | — | C | A | B | C | A | B | C |

For *four* treatments assume the digits 1–2 correspond to treatment A, digits 3–4 correspond to treatment B, digits 5–6 correspond to treatment C, and digits 7–8 correspond to treatment D. Ignore 0 and 9. The top row of Table 1 would then produce the following treatment assignments:

| | | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 5 | 5 | 6 | 3 | 1 | 3 | 9 | 1 | 1 | 4 | 5 | 5 | 4 | 8 | 1 | 0 | 7 | 3 | 4 | 8 | 3 | 4 | 7 |
| C | C | C | B | A | B | — | A | A | B | C | C | B | D | A | — | D | B | B | D | B | B | D |

The list should be produced before the start of the trial and should be large enough to complete the trial.

### Advantages of This Method

- Each patient assignment is completely unpredictable; and, if the numbers are large enough (say ≥ 500), then probability theory allows us to be confident that the numbers in each treatment group will be similar.
- The method is very straightforward and easy to implement.

### Disadvantages of This Method

- Table 2 illustrates the difference in treatment numbers that may occur in a two-treatment trial with probability at least 0.05 and at least 0.01. The probability of ≥10% overall imbalance between treatment groups is also shown.

**TABLE 1 Table of Random Numbers**

| | | | | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 5 | 5 | 6 | 3 | 1 | 3 | 9 | 1 | 1 | 4 | 5 | 5 | 4 | 8 | 1 | 0 | 7 | 3 | 4 | 8 | 3 | 4 | 7 | 9 | 3 |
| 5 | 3 | 0 | 6 | 0 | 0 | 5 | 2 | 5 | 2 | 3 | 8 | 3 | 5 | 4 | 8 | 7 | 1 | 6 | 0 | 9 | 8 | 0 | 3 | 6 |
| 0 | 8 | 0 | 4 | 3 | 9 | 9 | 6 | 0 | 1 | 2 | 0 | 6 | 8 | 8 | 8 | 9 | 3 | 5 | 6 | 2 | 4 | 4 | 7 | 8 |
| 8 | 7 | 2 | 5 | 3 | 0 | 7 | 7 | 1 | 6 | 7 | 0 | 9 | 6 | 9 | 4 | 4 | 9 | 0 | 3 | 8 | 2 | 8 | 3 | 6 |
| 1 | 6 | 8 | 7 | 1 | 7 | 2 | 9 | 1 | 2 | 1 | 3 | 7 | 9 | 1 | 1 | 3 | 8 | 7 | 2 | 3 | 9 | 0 | 4 | 5 |
| 3 | 9 | 9 | 9 | 8 | 9 | 6 | 8 | 9 | 2 | 7 | 7 | 0 | 8 | 9 | 7 | 2 | 8 | 4 | 3 | 0 | 8 | 2 | 0 | 8 |
| 0 | 2 | 9 | 9 | 6 | 6 | 1 | 3 | 5 | 1 | 5 | 1 | 5 | 9 | 1 | 9 | 8 | 7 | 1 | 0 | 4 | 1 | 3 | 4 | 4 |
| 6 | 9 | 9 | 4 | 4 | 6 | 0 | 6 | 2 | 5 | 1 | 3 | 6 | 8 | 6 | 2 | 1 | 1 | 1 | 2 | 9 | 8 | 9 | 5 | 0 |
| 7 | 0 | 1 | 1 | 0 | 5 | 0 | 6 | 3 | 2 | 0 | 4 | 2 | 0 | 3 | 5 | 2 | 2 | 9 | 7 | 2 | 6 | 9 | 1 | 8 |
| 1 | 8 | 9 | 4 | 2 | 1 | 0 | 4 | 3 | 8 | 2 | 6 | 8 | 7 | 9 | 0 | 3 | 7 | 7 | 9 | 1 | 0 | 5 | 8 | 2 |
| 6 | 0 | 8 | 6 | 2 | 4 | 4 | 4 | 5 | 2 | 1 | 2 | 8 | 0 | 0 | 3 | 4 | 5 | 3 | 8 | 1 | 9 | 6 | 3 | 3 |
| 7 | 3 | 4 | 6 | 8 | 2 | 5 | 2 | 0 | 5 | 9 | 9 | 4 | 1 | 4 | 2 | 2 | 7 | 2 | 5 | 3 | 7 | 6 | 1 | 5 |
| 0 | 8 | 7 | 0 | 3 | 6 | 0 | 7 | 9 | 4 | 3 | 8 | 9 | 5 | 9 | 0 | 2 | 0 | 6 | 7 | 7 | 7 | 8 | 3 | 4 |
| 5 | 1 | 4 | 3 | 7 | 7 | 8 | 6 | 1 | 4 | 3 | 9 | 6 | 6 | 0 | 6 | 7 | 9 | 9 | 5 | 8 | 3 | 3 | 8 | 0 |
| 2 | 8 | 1 | 7 | 2 | 5 | 0 | 6 | 8 | 7 | 0 | 7 | 1 | 8 | 0 | 1 | 1 | 3 | 6 | 9 | 0 | 1 | 2 | 7 | 6 |
| 2 | 2 | 8 | 9 | 3 | 5 | 3 | 9 | 8 | 8 | 8 | 9 | 6 | 7 | 8 | 2 | 7 | 7 | 0 | 8 | 8 | 7 | 2 | 7 | 3 |
| 3 | 9 | 7 | 1 | 9 | 7 | 0 | 2 | 9 | 9 | 6 | 6 | 1 | 3 | 5 | 1 | 5 | 1 | 5 | 9 | 1 | 9 | 8 | 7 | 1 |
| 0 | 4 | 1 | 3 | 4 | 4 | 6 | 9 | 9 | 4 | 4 | 6 | 0 | 6 | 2 | 5 | 1 | 3 | 6 | 8 | 6 | 2 | 1 | 1 | 1 |
| 2 | 9 | 8 | 9 | 5 | 0 | 7 | 0 | 1 | 1 | 0 | 5 | 0 | 6 | 3 | 2 | 0 | 4 | 2 | 0 | 3 | 5 | 2 | 2 | 9 |
| 7 | 2 | 6 | 9 | 1 | 8 | 1 | 8 | 9 | 4 | 2 | 1 | 0 | 4 | 3 | 8 | 2 | 6 | 8 | 7 | 9 | 0 | 3 | 7 | 7 |
| 9 | 1 | 0 | 5 | 8 | 2 | 6 | 0 | 8 | 6 | 2 | 4 | 4 | 4 | 5 | 2 | 1 | 2 | 8 | 0 | 0 | 3 | 4 | 5 | 3 |
| 8 | 1 | 9 | 6 | 3 | 3 | 7 | 3 | 4 | 6 | 8 | 2 | 5 | 2 | 0 | 5 | 9 | 9 | 4 | 1 | 4 | 2 | 2 | 7 | 2 |
| 5 | 3 | 7 | 6 | 1 | 5 | 0 | 8 | 7 | 0 | 3 | 6 | 0 | 7 | 9 | 4 | 3 | 8 | 9 | 5 | 9 | 0 | 2 | 0 | 6 |
| 7 | 7 | 7 | 8 | 3 | 4 | 5 | 1 | 4 | 3 | 7 | 7 | 8 | 6 | 1 | 4 | 3 | 9 | 6 | 6 | 0 | 6 | 7 | 9 | 9 |
| 5 | 8 | 3 | 3 | 8 | 0 | 2 | 8 | 1 | 7 | 2 | 5 | 0 | 6 | 8 | 7 | 0 | 7 | 1 | 8 | 0 | 1 | 1 | 3 | 6 |
| 9 | 0 | 1 | 2 | 7 | 6 | 2 | 2 | 8 | 9 | 3 | 5 | 3 | 9 | 8 | 8 | 8 | 9 | 6 | 7 | 8 | 2 | 7 | 7 | 0 |
| 8 | 8 | 7 | 2 | 7 | 3 | 3 | 9 | 7 | 1 | 9 | 7 | 0 | 2 | 9 | 9 | 6 | 6 | 1 | 3 | 5 | 1 | 5 | 1 | 5 |
| 9 | 1 | 9 | 8 | 7 | 1 | 0 | 4 | 1 | 3 | 4 | 4 | 6 | 9 | 9 | 4 | 4 | 6 | 0 | 6 | 2 | 5 | 1 | 3 | 6 |
| 8 | 6 | 2 | 1 | 1 | 1 | 2 | 9 | 8 | 9 | 5 | 0 | 7 | 0 | 1 | 1 | 0 | 5 | 0 | 6 | 3 | 2 | 0 | 4 | 2 |
| 0 | 3 | 5 | 2 | 2 | 9 | 7 | 2 | 6 | 9 | 1 | 8 | 1 | 8 | 9 | 4 | 2 | 1 | 0 | 4 | 3 | 8 | 2 | 6 | 8 |
| 7 | 9 | 0 | 3 | 7 | 7 | 9 | 1 | 0 | 5 | 8 | 2 | 6 | 0 | 8 | 6 | 2 | 4 | 4 | 4 | 5 | 2 | 1 | 2 | 8 |
| 0 | 0 | 3 | 4 | 5 | 3 | 8 | 1 | 9 | 6 | 3 | 3 | 7 | 3 | 4 | 6 | 8 | 2 | 5 | 2 | 0 | 5 | 9 | 9 | 4 |
| 1 | 4 | 2 | 2 | 7 | 2 | 5 | 3 | 7 | 6 | 1 | 5 | 0 | 8 | 7 | 0 | 3 | 6 | 0 | 7 | 9 | 4 | 3 | 8 | 9 |
| 5 | 9 | 0 | 2 | 0 | 6 | 7 | 7 | 7 | 8 | 3 | 4 | 5 | 1 | 4 | 3 | 7 | 7 | 8 | 6 | 1 | 4 | 3 | 9 | 6 |
| 6 | 0 | 6 | 7 | 9 | 9 | 5 | 8 | 3 | 3 | 8 | 0 | 2 | 8 | 1 | 7 | 2 | 5 | 0 | 6 | 8 | 7 | 0 | 7 | 1 |
| 8 | 0 | 1 | 1 | 3 | 6 | 9 | 0 | 1 | 2 | 7 | 6 | 2 | 2 | 8 | 9 | 3 | 5 | 3 | 9 | 8 | 8 | 8 | 9 | 6 |
| 7 | 8 | 2 | 7 | 7 | 0 | 8 | 8 | 7 | 2 | 7 | 3 | 3 | 9 | 7 | 1 | 9 | 7 | 0 | 2 | 9 | 9 | 6 | 6 | 1 |
| 3 | 5 | 1 | 5 | 1 | 5 | 9 | 1 | 9 | 8 | 7 | 1 | 0 | 4 | 1 | 3 | 4 | 4 | 6 | 9 | 9 | 4 | 4 | 6 | 0 |
| 6 | 2 | 5 | 1 | 3 | 6 | 8 | 6 | 2 | 1 | 1 | 1 | 2 | 9 | 8 | 9 | 5 | 0 | 7 | 0 | 1 | 1 | 0 | 5 | 0 |
| 6 | 3 | 2 | 0 | 4 | 2 | 0 | 3 | 5 | 2 | 2 | 9 | 7 | 2 | 6 | 9 | 1 | 8 | 1 | 8 | 9 | 4 | 2 | 1 | 0 |
| 4 | 3 | 8 | 2 | 6 | 8 | 7 | 9 | 0 | 3 | 7 | 7 | 9 | 1 | 0 | 5 | 8 | 2 | 6 | 0 | 8 | 6 | 2 | 4 | 4 |
| 4 | 5 | 2 | 1 | 2 | 8 | 0 | 0 | 3 | 4 | 5 | 3 | 8 | 1 | 9 | 6 | 3 | 3 | 7 | 3 | 4 | 6 | 8 | 2 | 5 |
| 2 | 0 | 5 | 9 | 9 | 4 | 1 | 4 | 2 | 2 | 7 | 2 | 5 | 3 | 7 | 6 | 1 | 5 | 0 | 8 | 7 | 0 | 3 | 6 | 0 |
| 7 | 9 | 4 | 3 | 8 | 9 | 5 | 9 | 0 | 2 | 0 | 6 | 7 | 7 | 7 | 8 | 3 | 4 | 5 | 1 | 4 | 3 | 7 | 7 | 8 |
| 6 | 1 | 4 | 3 | 9 | 6 | 6 | 0 | 6 | 7 | 9 | 9 | 5 | 8 | 3 | 3 | 8 | 0 | 2 | 8 | 1 | 7 | 2 | 5 | 0 |
| 6 | 8 | 7 | 0 | 7 | 1 | 8 | 0 | 1 | 1 | 3 | 6 | 9 | 0 | 1 | 2 | 7 | 6 | 2 | 2 | 8 | 9 | 3 | 5 | 3 |
| 9 | 8 | 8 | 8 | 9 | 6 | 7 | 8 | 2 | 7 | 7 | 0 | 8 | 8 | 7 | 2 | 7 | 3 | 3 | 9 | 7 | 1 | 9 | 7 | 0 |
| 2 | 9 | 9 | 6 | 6 | 1 | 3 | 5 | 1 | 5 | 1 | 5 | 9 | 1 | 9 | 8 | 7 | 1 | 0 | 4 | 1 | 3 | 4 | 4 | 6 |
| 9 | 9 | 4 | 4 | 6 | 0 | 6 | 2 | 5 | 1 | 3 | 6 | 8 | 6 | 2 | 1 | 1 | 1 | 2 | 9 | 8 | 9 | 5 | 0 | 7 |
| 0 | 1 | 1 | 0 | 5 | 0 | 6 | 3 | 2 | 0 | 4 | 2 | 0 | 3 | 5 | 2 | 2 | 9 | 7 | 2 | 6 | 9 | 1 | 8 | 1 |
| 8 | 9 | 4 | 2 | 1 | 0 | 4 | 3 | 8 | 2 | 6 | 8 | 7 | 9 | 0 | 3 | 7 | 7 | 9 | 1 | 0 | 5 | 8 | 2 | 6 |

**TABLE 1** ***Continued***

| | | | | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 8 | 6 | 2 | 4 | 4 | 4 | 5 | 2 | 1 | 2 | 8 | 0 | 0 | 3 | 4 | 5 | 3 | 8 | 1 | 9 | 6 | 3 | 3 | 7 |
| 3 | 4 | 6 | 8 | 2 | 5 | 2 | 0 | 5 | 9 | 9 | 4 | 1 | 4 | 2 | 2 | 7 | 2 | 5 | 3 | 7 | 6 | 1 | 5 | 0 |
| 8 | 7 | 0 | 3 | 6 | 0 | 7 | 9 | 4 | 3 | 8 | 9 | 5 | 9 | 0 | 2 | 0 | 6 | 7 | 7 | 7 | 8 | 3 | 4 | 5 |
| 1 | 4 | 3 | 7 | 7 | 8 | 6 | 1 | 4 | 3 | 9 | 6 | 6 | 0 | 6 | 7 | 9 | 9 | 5 | 8 | 3 | 3 | 8 | 0 | 2 |
| 8 | 1 | 7 | 2 | 5 | 0 | 6 | 8 | 7 | 0 | 7 | 1 | 8 | 0 | 1 | 1 | 3 | 6 | 9 | 0 | 1 | 2 | 7 | 6 | 2 |
| 2 | 8 | 9 | 3 | 5 | 3 | 9 | 8 | 8 | 8 | 9 | 6 | 7 | 8 | 2 | 7 | 7 | 0 | 8 | 8 | 7 | 2 | 7 | 3 | 3 |
| 9 | 7 | 1 | 9 | 7 | 0 | 2 | 9 | 9 | 6 | 6 | 1 | 3 | 5 | 1 | 5 | 1 | 5 | 9 | 1 | 9 | 8 | 7 | 1 | 0 |
| 4 | 1 | 3 | 4 | 4 | 6 | 9 | 9 | 4 | 4 | 6 | 0 | 6 | 2 | 5 | 1 | 3 | 6 | 8 | 6 | 2 | 1 | 1 | 1 | 2 |
| 9 | 8 | 9 | 5 | 0 | 7 | 0 | 1 | 1 | 0 | 5 | 0 | 6 | 3 | 2 | 0 | 4 | 2 | 0 | 3 | 5 | 2 | 2 | 9 | 7 |
| 2 | 6 | 9 | 1 | 8 | 1 | 8 | 9 | 4 | 2 | 1 | 0 | 4 | 3 | 8 | 2 | 6 | 8 | 7 | 9 | 0 | 3 | 7 | 7 | 9 |
| 1 | 0 | 5 | 8 | 2 | 6 | 0 | 8 | 6 | 2 | 4 | 4 | 4 | 5 | 2 | 1 | 2 | 8 | 0 | 0 | 3 | 4 | 5 | 3 | 8 |
| 1 | 9 | 6 | 3 | 3 | 7 | 3 | 4 | 6 | 8 | 2 | 5 | 2 | 0 | 5 | 9 | 9 | 4 | 1 | 4 | 2 | 2 | 7 | 2 | 5 |
| 3 | 7 | 6 | 1 | 5 | 0 | 8 | 7 | 0 | 3 | 6 | 0 | 7 | 9 | 4 | 3 | 8 | 9 | 5 | 9 | 0 | 2 | 0 | 6 | 7 |
| 7 | 7 | 8 | 3 | 4 | 5 | 1 | 4 | 3 | 7 | 7 | 8 | 6 | 1 | 4 | 3 | 9 | 6 | 6 | 0 | 6 | 7 | 9 | 9 | 5 |
| 8 | 3 | 3 | 8 | 0 | 2 | 8 | 1 | 7 | 2 | 5 | 0 | 6 | 8 | 7 | 0 | 7 | 1 | 8 | 0 | 1 | 1 | 3 | 6 | 9 |
| 0 | 1 | 2 | 7 | 6 | 2 | 2 | 8 | 9 | 3 | 5 | 3 | 9 | 8 | 8 | 8 | 9 | 6 | 7 | 8 | 2 | 7 | 7 | 0 | 8 |
| 8 | 7 | 2 | 7 | 3 | 3 | 9 | 7 | 1 | 9 | 7 | 0 | 2 | 9 | 9 | 6 | 6 | 1 | 3 | 5 | 1 | 5 | 1 | 5 | 9 |
| 1 | 9 | 8 | 7 | 1 | 0 | 4 | 1 | 3 | 4 | 4 | 6 | 9 | 9 | 4 | 4 | 6 | 0 | 6 | 2 | 5 | 1 | 3 | 6 | 8 |
| 6 | 2 | 1 | 1 | 1 | 2 | 9 | 8 | 9 | 5 | 0 | 7 | 0 | 1 | 1 | 0 | 5 | 0 | 6 | 3 | 2 | 0 | 4 | 2 | 0 |
| 3 | 5 | 2 | 2 | 9 | 7 | 2 | 6 | 9 | 1 | 8 | 1 | 8 | 9 | 4 | 2 | 1 | 0 | 4 | 3 | 8 | 2 | 6 | 8 | 7 |
| 9 | 0 | 3 | 7 | 7 | 9 | 1 | 0 | 5 | 8 | 2 | 6 | 0 | 8 | 6 | 2 | 4 | 4 | 4 | 5 | 2 | 1 | 2 | 8 | 0 |
| 0 | 3 | 4 | 5 | 3 | 8 | 1 | 9 | 6 | 3 | 3 | 7 | 3 | 4 | 6 | 8 | 2 | 5 | 2 | 0 | 5 | 9 | 9 | 4 | 1 |
| 4 | 2 | 2 | 7 | 2 | 5 | 3 | 7 | 6 | 1 | 5 | 0 | 8 | 7 | 0 | 3 | 6 | 0 | 7 | 9 | 4 | 3 | 8 | 9 | 5 |
| 9 | 0 | 2 | 0 | 6 | 7 | 7 | 7 | 8 | 3 | 4 | 5 | 1 | 4 | 3 | 7 | 7 | 8 | 6 | 1 | 4 | 3 | 9 | 6 | 6 |
| 0 | 6 | 7 | 9 | 9 | 5 | 8 | 3 | 3 | 8 | 0 | 2 | 8 | 1 | 7 | 2 | 5 | 0 | 6 | 8 | 7 | 0 | 7 | 1 | 8 |
| 0 | 1 | 1 | 3 | 6 | 9 | 0 | 1 | 2 | 7 | 6 | 2 | 2 | 8 | 9 | 3 | 5 | 3 | 9 | 8 | 8 | 8 | 9 | 6 | 7 |
| 8 | 2 | 7 | 7 | 0 | 8 | 8 | 7 | 2 | 7 | 3 | 3 | 9 | 7 | 1 | 9 | 7 | 0 | 2 | 9 | 9 | 6 | 6 | 1 | 3 |
| 5 | 1 | 5 | 1 | 5 | 9 | 1 | 9 | 8 | 7 | 1 | 0 | 4 | 1 | 3 | 4 | 4 | 6 | 9 | 9 | 4 | 4 | 6 | 0 | 6 |
| 2 | 5 | 1 | 3 | 6 | 8 | 6 | 2 | 1 | 1 | 1 | 2 | 9 | 8 | 9 | 5 | 0 | 7 | 0 | 1 | 1 | 0 | 5 | 0 | 6 |
| 3 | 2 | 0 | 4 | 2 | 0 | 3 | 5 | 2 | 2 | 9 | 7 | 2 | 6 | 9 | 1 | 8 | 1 | 8 | 9 | 4 | 2 | 1 | 0 | 4 |
| 3 | 8 | 2 | 6 | 8 | 7 | 9 | 0 | 3 | 7 | 7 | 9 | 1 | 0 | 5 | 8 | 2 | 6 | 0 | 8 | 6 | 2 | 4 | 4 | 4 |
| 5 | 2 | 1 | 2 | 8 | 0 | 0 | 3 | 4 | 5 | 3 | 8 | 1 | 9 | 6 | 3 | 3 | 7 | 3 | 4 | 6 | 8 | 2 | 5 | 2 |
| 0 | 5 | 9 | 9 | 4 | 1 | 4 | 2 | 2 | 7 | 2 | 5 | 3 | 7 | 6 | 1 | 5 | 0 | 8 | 7 | 0 | 3 | 6 | 0 | 7 |
| 1 | 6 | 5 | 1 | 1 | 7 | 1 | 2 | 4 | 2 | 8 | 9 | 0 | 9 | 0 | 5 | 6 | 7 | 3 | 6 | 6 | 9 | 9 | 0 | 8 |
| 3 | 6 | 5 | 1 | 8 | 8 | 2 | 8 | 9 | 1 | 1 | 7 | 0 | 5 | 5 | 0 | 2 | 4 | 0 | 3 | 9 | 4 | 8 | 3 | 8 |
| 0 | 9 | 2 | 9 | 3 | 0 | 2 | 3 | 3 | 5 | 8 | 1 | 2 | 3 | 4 | 9 | 8 | 4 | 4 | 0 | 2 | 5 | 7 | 5 | 1 |
| 0 | 0 | 0 | 1 | 8 | 9 | 0 | 4 | 9 | 9 | 2 | 0 | 0 | 9 | 4 | 9 | 5 | 5 | 1 | 9 | 3 | 1 | 9 | 2 | 4 |
| 1 | 1 | 8 | 8 | 3 | 5 | 7 | 3 | 7 | 3 | 7 | 1 | 3 | 1 | 0 | 9 | 3 | 2 | 6 | 3 | 5 | 6 | 6 | 8 | 1 |
| 1 | 6 | 6 | 8 | 2 | 9 | 4 | 7 | 3 | 5 | 8 | 0 | 8 | 4 | 3 | 3 | 3 | 4 | 1 | 0 | 1 | 7 | 2 | 9 | 2 |
| 3 | 4 | 2 | 7 | 2 | 8 | 5 | 4 | 2 | 6 | 4 | 2 | 5 | 7 | 4 | 4 | 1 | 9 | 4 | 8 | 1 | 3 | 0 | 3 | 0 |
| 1 | 6 | 4 | 3 | 2 | 6 | 5 | 0 | 4 | 8 | 0 | 9 | 1 | 2 | 5 | 9 | 9 | 1 | 3 | 2 | 7 | 0 | 4 | 8 | 2 |
| 0 | 8 | 5 | 6 | 6 | 6 | 7 | 4 | 3 | 4 | 0 | 2 | 2 | 5 | 6 | 7 | 5 | 0 | 3 | 1 | 8 | 5 | 5 | 9 | 5 |
| 6 | 8 | 0 | 4 | 7 | 4 | 2 | 7 | 1 | 1 | 7 | 3 | 6 | 4 | 4 | 9 | 4 | 7 | 5 | 9 | 8 | 3 | 7 | 2 | 0 |
| 9 | 2 | 5 | 8 | 2 | 9 | 1 | 6 | 5 | 1 | 1 | 7 | 1 | 2 | 4 | 2 | 8 | 9 | 0 | 9 | 0 | 5 | 6 | 7 | 3 |
| 6 | 6 | 9 | 9 | 0 | 8 | 3 | 6 | 5 | 1 | 8 | 8 | 2 | 8 | 9 | 1 | 1 | 7 | 0 | 5 | 5 | 0 | 2 | 4 | 0 |
| 3 | 9 | 4 | 8 | 3 | 8 | 0 | 9 | 2 | 9 | 3 | 0 | 2 | 3 | 3 | 5 | 8 | 1 | 2 | 3 | 4 | 9 | 8 | 4 | 4 |
| 0 | 2 | 5 | 7 | 5 | 1 | 0 | 0 | 0 | 1 | 8 | 9 | 0 | 4 | 9 | 9 | 2 | 0 | 0 | 9 | 4 | 9 | 5 | 5 | 1 |
| 9 | 3 | 1 | 9 | 2 | 4 | 1 | 1 | 8 | 8 | 3 | 5 | 7 | 3 | 7 | 3 | 7 | 1 | 3 | 1 | 0 | 9 | 3 | 2 | 6 |
| 3 | 5 | 6 | 6 | 8 | 1 | 1 | 6 | 6 | 8 | 2 | 9 | 4 | 7 | 3 | 5 | 8 | 0 | 8 | 4 | 3 | 3 | 3 | 4 | 1 |

**TABLE 2 Possible Imbalance between Treatment Groups in Simple Randomization with Two Treatments**

| Total Number of Patients | Difference in Numbers | | Probability of ≥10% Treatment Imbalance |
|---|---|---|---|
| | Probability ≥ 0.05 | Probability ≥ 0.01 | |
| 10 | 2:8 | 1:9 | 0.744 |
| 20 | 6:14 | 5:15 | 0.819 |
| 50 | 18:32 | 16:34 | 0.477 |
| 100 | 40:60 | 37:63 | 0.371 |
| 200 | 86:114 | 82:118 | 0.188 |
| 500 | 228:272 | 221:279 | 0.036 |
| 1000 | 470:530 | 459:541 | 0 |

**TABLE 3 Possible Imbalance between Treatment Groups in Simple Randomization with Three Treatments**

| Total Number of Patients | Lowest Observation in a Group | | Probability of ≥10% Treatment Imbalance |
|---|---|---|---|
| | Probability ≥ 0.05 | Probability ≥ 0.01 | |
| 10 | 1 in any group | 0 in any group | 1 |
| 20 | 2 in any group | 2 in any group | 1 |
| 50 | 9 in any group | 8 in any group | 0.732 |
| 100 | 24 in any group | 21 in any group | 0.474 |
| 200 | 54 in any group | 49 in any group | 0.196 |
| 500 | 144 in any group | 137 in any group | 0.022 |
| 1000 | 300 in any group | 290 in any group | 0.001 |

**TABLE 4 Possible Imbalance between Treatment Groups in Simple Randomization with Four Treatments**

| Total Number of Patients | Lowest Observation in a Group | | Probability of ≥10% Treatment Imbalance |
|---|---|---|---|
| | Probability ≥ 0.05 | Probability ≥ 0.01 | |
| 10 | 0 in any group | 0 in any group | 1 |
| 20 | 1 in any group | 0 in any group | 0.944 |
| 50 | 6 in any group | 4 in any group | 0.806 |
| 100 | 15 in any group | 13 in any group | 0.530 |
| 200 | 37 in any group | 33 in any group | 0.193 |
| 500 | 104 in any group | 98 in any group | 0.004 |
| 1000 | 220 in any group | 213 in any group | 0.001 |

- For three or four treatments, Tables 3 and 4 show the probability of no patients being allocated to one of the treatment groups, and the probability of ≥10% overall imbalance between treatment groups is also shown.
- There is also no guarantee that the groups are similar with respect to important prognostic factors, and for small to medium sized trials they are unlikely to be balanced. Therefore, in practice, some sort of restricted randomization is almost always used.

### 2.2 Restricted Randomization

Restricted randomization is used to balance treatment assignments assigned to each treatment group. Three possible approaches are described here.

***Random Permuted Blocks*** A list of allocations is prepared before the start of the trial to be assigned in sequence order. This method is used to keep numbers in each treatment group closely balanced at all times, so there is a greater chance of balance whenever recruitment ends (which may be early). Blocks are made up of all possible permutations of treatment assignments. Therefore block sizes that are multiples of the number of treatments are easier to manage.

For two treatments, blocks of an even size would be used. It is unusual to use a block size as small as 2 as this makes guessing easier, but for a small trial with few patients it may be justified. The random numbers in Table 1 can be used to produce the randomization list.

*Example: Two Treatments, Block Size 2* Let 0–4 correspond to block AB and 5–9 correspond to block BA. The top row of Table 1 would then produce the following treatment assignments:

| | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 5 | 5 | 6 | 3 | 1 | 3 | 9 | 1 | 1 | 4 | 5 | 5 | 4 | 8 | 1 |
| BA | BA | BA | AB | AB | AB | BA | AB | AB | AB | BA | BA | AB | BA | AB |

*Example: Three Treatments, Block Size 3* Let 1 correspond to block ABC, 2 to block ACB, 3 to block BAC, 4 to block BCA, 5 to block CAB, and 6 to block CBA. Ignore 0 and 7–9. The top row of Table 1 would then produce the following treatment assignments:

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| 5 | 5 | 6 | 3 | 1 | 3 | 9 | 1 | 1 | 4 | 5 |
| CAB | CAB | CBA | BAC | ABC | BAC | — | ABC | ABC | BCA | CAB |

*Example: Two Treatments, Block Size 4* Let 1 correspond to block AABB, 2 to block ABAB, 3 to block ABBA, 4 to block BBAA, 5 to block BABA, and 6 to block BAAB. Ignore 0 and 7–9. The top row of Table 1 would then produce the following treatment assignments:

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 5 | 5 | 6 | 3 | 1 | 3 | 9 | 1 | 1 | 4 |
| BABA | BABA | BAAB | ABBA | AABB | ABBA | — | AABB | AABB | BBAA |

It is also common practice to make the second block the mirror image of the first block, for example, Block 1 (A B A B) becomes Block 2 (B A B A). Producing a list of randomly ordered numbers is easily done within Microsoft Access. To randomly order the numbers 0–19 first create a table, say called TblRandom20, with one field in it (called ID) and add the data values 0–19 into this table. Then create a query with the following SQL (Structured Query Language) syntax:

```
SELECT TblRandom20.ID
FROM TblRandom20
ORDER BY Rnd([ID]);
```

Each time this query is opened in the datasheet view it will produce the numbers 0–19 in a random order. For 20 digits the number of possible permutations (and therefore the number of differently ordered blocks) are:

$$20! = 20 \times 19 \times 18 \times \cdots \times 3 \times 2 \times 1 = 2432902008176640000$$

This can be used by a computer program (see Appendix II) to produce a list of random permutations as illustrated in Table 5, which can then be used to produce the randomization list.

*Example: Two Treatments, Block Size 20* Let 0–9 correspond to treatment A and 10–19 to treatment B.

The top row of Table 5 would then produce block 1:

| | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 9 | 18 | 17 | 5 | 3 | 10 | 2 | 4 | 11 | 13 | 12 | 14 | 6 | 7 | 15 | 0 | 19 | 8 | 1 | 16 |
| A | B | B | A | A | B | A | A | B | B | B | B | A | A | B | A | B | A | A | B |

And the second row would produce block 2:

| | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 4 | 1 | 15 | 10 | 5 | 14 | 3 | 18 | 17 | 9 | 0 | 19 | 2 | 12 | 6 | 11 | 8 | 13 | 7 | 16 |
| A | A | B | B | A | B | A | B | B | A | A | B | A | B | A | B | A | B | A | B |

When using a fixed block size, it is possible that the block size may be guessed (resulting in allocation concealment being subverted). If the block size has been correctly guessed, then the last assignment in each block could be predicted with certainty, and potentially half the treatments in a block might be guessed. Varying the block length can be used to prevent this occurring. Implementing a variable-block length design is most easily done by a computer program. The program given in Appendix III was used to produce a list of treatment assignments for a study design with two treatments and random block sizes of 4 and 6 for an intended study population of 200 patients. The output from this program was stored in a table that, when read sequentially, gave the following treatment allocation schedule (assigning treatment A to 0 and treatment B to 1):

B A B A B A A A B B A A A B B B A B B A B A B A B A B B A A B
A B A B A A A B B A B A B B A A B A A B B A A B B B A B A B A
A B A A A B B B A A B A B B A B B A B B A B A A A B A B A A A
B B B A A A B B B A B B B A A B B A A A B A B B A A B A B A B
B A A B B A B B A A B A B B A A B A B A A A B B B B A A A A B
A B B B A A B A B B A B B A B A A B A A B A B B A B B A A A B
B A B A A B B B A A B A A B A B B A A B A B B B A A A B A B B
A A B A B A B B A B A B A A A B B A A A B B A

**Advantages of This Method**

- The list can be prepared in advance and will ensure roughly equal numbers in each treatment group even if recruitment stops early.

**TABLE 5 Random Permutations of 20 Numbers (0–19)**

| | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 9 | 18 | 17 | 5 | 3 | 10 | 2 | 4 | 11 | 13 | 12 | 14 | 6 | 7 | 15 | 0 | 19 | 8 | 1 | 16 |
| 4 | 1 | 15 | 10 | 5 | 14 | 3 | 18 | 17 | 9 | 0 | 19 | 2 | 12 | 6 | 11 | 8 | 13 | 7 | 16 |
| 19 | 0 | 5 | 11 | 12 | 4 | 7 | 1 | 13 | 16 | 3 | 10 | 17 | 9 | 18 | 14 | 15 | 6 | 8 | 2 |
| 11 | 10 | 7 | 12 | 18 | 1 | 15 | 8 | 17 | 2 | 3 | 0 | 19 | 4 | 6 | 14 | 9 | 13 | 5 | 16 |
| 7 | 5 | 3 | 16 | 2 | 6 | 12 | 4 | 14 | 1 | 17 | 19 | 0 | 11 | 18 | 8 | 9 | 13 | 15 | 10 |
| 3 | 2 | 12 | 8 | 17 | 16 | 19 | 0 | 6 | 5 | 11 | 9 | 15 | 14 | 13 | 1 | 10 | 4 | 18 | 7 |
| 6 | 17 | 19 | 0 | 3 | 15 | 7 | 13 | 8 | 5 | 4 | 18 | 2 | 16 | 9 | 11 | 1 | 14 | 12 | 10 |
| 5 | 9 | 11 | 10 | 4 | 14 | 15 | 2 | 8 | 12 | 18 | 3 | 17 | 0 | 19 | 13 | 6 | 7 | 16 | 1 |
| 12 | 15 | 19 | 0 | 3 | 2 | 9 | 13 | 16 | 1 | 5 | 17 | 18 | 4 | 6 | 11 | 10 | 8 | 7 | 14 |
| 10 | 12 | 4 | 11 | 1 | 3 | 8 | 5 | 16 | 9 | 2 | 14 | 15 | 13 | 17 | 6 | 7 | 18 | 0 | 19 |
| 8 | 13 | 10 | 18 | 19 | 0 | 5 | 12 | 9 | 17 | 3 | 2 | 14 | 11 | 15 | 4 | 1 | 7 | 16 | 6 |
| 8 | 6 | 19 | 0 | 13 | 16 | 7 | 1 | 9 | 3 | 2 | 12 | 18 | 14 | 10 | 11 | 4 | 5 | 15 | 17 |
| 3 | 7 | 12 | 10 | 4 | 6 | 9 | 17 | 15 | 2 | 1 | 5 | 14 | 8 | 13 | 18 | 11 | 16 | 0 | 19 |
| 17 | 12 | 4 | 2 | 16 | 15 | 5 | 10 | 6 | 13 | 8 | 19 | 0 | 9 | 3 | 18 | 7 | 1 | 14 | 11 |
| 18 | 17 | 10 | 7 | 12 | 11 | 4 | 13 | 6 | 16 | 0 | 19 | 1 | 2 | 15 | 3 | 9 | 8 | 5 | 14 |
| 11 | 4 | 3 | 9 | 18 | 8 | 7 | 14 | 10 | 5 | 16 | 2 | 13 | 6 | 12 | 15 | 1 | 0 | 19 | 17 |
| 5 | 13 | 9 | 4 | 18 | 7 | 1 | 12 | 14 | 2 | 16 | 3 | 17 | 10 | 11 | 6 | 15 | 8 | 0 | 19 |
| 4 | 7 | 9 | 6 | 18 | 13 | 11 | 10 | 16 | 3 | 17 | 2 | 5 | 8 | 14 | 1 | 12 | 15 | 0 | 19 |
| 2 | 9 | 19 | 0 | 5 | 17 | 6 | 15 | 14 | 10 | 4 | 1 | 12 | 3 | 16 | 8 | 7 | 13 | 11 | 18 |
| 9 | 3 | 13 | 11 | 14 | 5 | 8 | 10 | 18 | 7 | 16 | 0 | 19 | 12 | 1 | 17 | 15 | 2 | 4 | 6 |
| 14 | 10 | 15 | 2 | 13 | 4 | 11 | 17 | 1 | 16 | 18 | 6 | 0 | 19 | 7 | 3 | 12 | 8 | 9 | 5 |
| 2 | 14 | 18 | 11 | 5 | 13 | 6 | 16 | 17 | 10 | 8 | 7 | 3 | 15 | 9 | 0 | 19 | 1 | 12 | 4 |
| 11 | 16 | 7 | 3 | 9 | 1 | 6 | 10 | 15 | 18 | 12 | 14 | 13 | 0 | 19 | 4 | 17 | 8 | 2 | 5 |
| 8 | 19 | 0 | 16 | 2 | 3 | 6 | 1 | 11 | 4 | 5 | 9 | 7 | 12 | 10 | 18 | 13 | 15 | 17 | 14 |
| 13 | 10 | 5 | 8 | 16 | 9 | 7 | 14 | 4 | 15 | 3 | 0 | 19 | 6 | 17 | 2 | 12 | 18 | 1 | 11 |
| 18 | 7 | 9 | 14 | 8 | 5 | 19 | 0 | 15 | 3 | 13 | 11 | 10 | 2 | 17 | 1 | 16 | 4 | 12 | 6 |
| 16 | 19 | 0 | 14 | 15 | 9 | 6 | 8 | 2 | 11 | 12 | 4 | 1 | 7 | 5 | 17 | 3 | 18 | 13 | 10 |
| 16 | 6 | 15 | 13 | 9 | 14 | 0 | 19 | 2 | 1 | 18 | 4 | 8 | 7 | 17 | 5 | 3 | 11 | 12 | 10 |
| 12 | 4 | 16 | 7 | 10 | 5 | 8 | 1 | 11 | 13 | 3 | 15 | 14 | 9 | 6 | 2 | 18 | 17 | 0 | 19 |
| 12 | 6 | 5 | 14 | 10 | 2 | 18 | 4 | 17 | 1 | 3 | 15 | 0 | 19 | 16 | 13 | 7 | 11 | 8 | 9 |
| 19 | 0 | 9 | 13 | 5 | 12 | 16 | 18 | 8 | 7 | 4 | 10 | 11 | 6 | 15 | 17 | 2 | 1 | 3 | 14 |
| 19 | 0 | 6 | 9 | 4 | 11 | 13 | 1 | 3 | 10 | 7 | 14 | 5 | 16 | 18 | 17 | 2 | 12 | 15 | 8 |
| 5 | 2 | 12 | 4 | 3 | 10 | 1 | 11 | 17 | 15 | 14 | 6 | 16 | 0 | 19 | 8 | 13 | 18 | 9 | 7 |
| 17 | 2 | 3 | 6 | 14 | 12 | 11 | 8 | 19 | 0 | 18 | 16 | 15 | 13 | 4 | 10 | 7 | 9 | 1 | 5 |
| 18 | 2 | 11 | 9 | 3 | 5 | 10 | 12 | 16 | 4 | 19 | 0 | 7 | 8 | 1 | 13 | 15 | 6 | 17 | 14 |
| 5 | 10 | 12 | 19 | 0 | 18 | 15 | 17 | 2 | 14 | 3 | 1 | 9 | 8 | 13 | 7 | 11 | 4 | 16 | 6 |
| 13 | 6 | 18 | 14 | 3 | 19 | 0 | 8 | 9 | 10 | 7 | 16 | 17 | 4 | 2 | 15 | 11 | 5 | 12 | 1 |
| 18 | 9 | 17 | 6 | 2 | 8 | 3 | 16 | 10 | 5 | 11 | 12 | 4 | 7 | 15 | 1 | 0 | 19 | 13 | 14 |
| 16 | 18 | 13 | 4 | 10 | 12 | 14 | 3 | 9 | 0 | 19 | 17 | 2 | 11 | 15 | 5 | 7 | 6 | 8 | 1 |
| 7 | 13 | 9 | 8 | 19 | 0 | 6 | 12 | 2 | 10 | 11 | 1 | 15 | 16 | 5 | 17 | 3 | 18 | 14 | 4 |
| 3 | 5 | 14 | 4 | 11 | 9 | 12 | 15 | 13 | 8 | 18 | 2 | 1 | 16 | 7 | 17 | 6 | 10 | 0 | 19 |
| 14 | 16 | 15 | 18 | 12 | 17 | 5 | 11 | 7 | 10 | 2 | 6 | 1 | 13 | 0 | 19 | 8 | 9 | 3 | 4 |
| 11 | 6 | 10 | 14 | 5 | 13 | 9 | 15 | 17 | 1 | 12 | 2 | 7 | 18 | 3 | 19 | 0 | 16 | 8 | 4 |
| 2 | 3 | 16 | 7 | 12 | 5 | 13 | 4 | 14 | 6 | 19 | 0 | 18 | 15 | 17 | 1 | 9 | 10 | 11 | 8 |
| 8 | 12 | 2 | 9 | 15 | 13 | 14 | 5 | 16 | 18 | 3 | 17 | 4 | 19 | 0 | 6 | 7 | 1 | 10 | 11 |
| 12 | 6 | 5 | 13 | 3 | 19 | 0 | 2 | 7 | 9 | 11 | 18 | 10 | 4 | 15 | 8 | 17 | 1 | 14 | 16 |
| 9 | 12 | 8 | 5 | 14 | 19 | 0 | 7 | 6 | 11 | 4 | 10 | 13 | 18 | 1 | 15 | 2 | 17 | 16 | 3 |
| 8 | 14 | 4 | 16 | 13 | 17 | 7 | 19 | 0 | 12 | 5 | 11 | 10 | 15 | 18 | 3 | 9 | 2 | 6 | 1 |
| 14 | 12 | 13 | 19 | 0 | 18 | 10 | 5 | 15 | 8 | 6 | 4 | 17 | 11 | 16 | 9 | 7 | 2 | 3 | 1 |

**Disadvantages of This Method**

- The smaller the block size, the greater the risk of predicting the next treatment assignment.
- Also, this method does not attempt to balance prognostic factors.

***The Biased Coin Method*** Although blocking has been widely used, it may not be necessary to enforce such strict equality. To prevent major inequalities in treatment numbers, the biased coin design [4] can be used. This fairly simple method limits selection bias and maintains groups of roughly equal size. When a patient is to be randomized the probability of assignment to the treatment group with the least patients so far is assigned a probability $p > \frac{1}{2}$. If the two treatment groups are equal, then the probability $p = \frac{1}{2}$. Efron favored $p = \frac{2}{3}$ and maintained that although $p = \frac{3}{4}$ will maintain stricter balance, there is a greater chance of predicting the next allocation when there is an inequality. For a trial with $N = 100$, an inequality of 45:55 is not excessive but is extremely unlikely to occur with $p = \frac{2}{3}$ [5].

*Example: Two Treatments, $p = \frac{2}{3}$* Let 1–6 correspond to assignment to the treatment with least patients, 7–9 to assignment to the treatment with most patients, and ignore 0.

If treatment numbers are equal, use simple randomization: Let 0–4 correspond to treatment A and 5–9 to treatment B. The top row of Table 1 would then produce the following treatment assignments:

| | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 5 | 5 | 6 | 3 | 1 | 3 | 9 | 1 | 1 | 4 | 5 | 5 | 4 | 8 | 1 | 0 | 7 | 3 | 4 | 8 | 3 |
| B* | A | B* | A | A* | B | B* | A | A* | B | B* | A | A* | A | B | — | A | B | B | B* | A |

where * indicates simple randomization has been used.

To keep the allocation process as "random" as possible and to minimize the risk of prediction, it is possible to use a combination of methods. For example, use simple randomization unless the treatment imbalance exceeds some specified limit and then introduce a biased coin method to address the imbalance.

***Urn Method*** Using Efron's design, the bias of the coin, $p$, is a constant whatever the degree of imbalance. Wei [6] developed an adaptive biased coin method where the probability of assignment depends upon the magnitude of imbalance and number of patients treated.

For the urn design, assume an urn contains an equal number of balls of two types, A and B. When a patient is randomized, a ball is drawn and replaced. If the ball is type A, then the patient receives treatment A, and type B balls are added to the urn. If the ball is type B, then the patient receives treatment B, and type A balls are added to the urn. So the probability of assignment to the treatment with fewer numbers is increased.

The urn design is designated as UD ($\alpha$, $\beta$) with $\alpha$ being the number of balls of each type in the urn at the start and $\beta$ being the number of balls added to the urn after each allocation. The simplest model has one ball of each type in the urn initially ($\alpha = 1$), and one ball of the opposite type is added to the urn following each treatment allocation ($\beta = 1$).

The program listed in Appendix IV was used to produce the following randomization list:

B A A A B B A A B B B A B B B A A B B B B A B A B
A B A B A A B A A A A A B B A A B A A B B A A B B

**Advantages of These Methods**

- The biased coin and urn designs control the likelihood of treatment imbalance without imposing strict balance. Therefore, they are less predictable than blocked designs and less susceptible to selection bias.

**Disadvantages of These Methods**

- These methods are slightly more complex, requiring that totals of previous assignments are kept in order to calculate the probabilities to be used for the current assignment. This is best done with the aid of a computer program.
- Does not attempt to balance prognostic factors.

## 3 COVARIATE-ADAPTIVE RANDOMIZATION

In any clinical trial it is often desirable not only to achieve similar numbers of patients in each treatment group but also to ensure that patient groups are similar with respect to important prognostic factors such as age or gender. When using simple randomization or blocks, there is a chance that the treatment groups will differ with respect to important prognostic variables. Stratifying the randomization by the prognostic factors will reduce this imbalance. This will ensure that roughly equal numbers of patients will be allocated within each stratum to each of the treatment options either by simple randomization or by using permuted blocks. However, stratified randomization becomes unworkable as the number of prognostic factors increases because the number of strata required can quickly exceed the number of patients in the trial (see next section).

### 3.1 Stratification

Each combination of prognostic factors represents one stratum within the allocation list, and lists of random allocations are drawn up for each stratum. Multiple strata can become unmanageable and may lead to empty cells (the number of cells are the product of all the levels of the prognostic factors), therefore this method may not be suitable for small trials with several factors. The maximum desirable number of strata is unknown but according to Therneau [7] the number of strata in a two-treatment trial should be

$$\text{No. of strata} < \frac{N}{\text{block size}} \quad \text{and maybe} \quad \text{No. of strata} < \frac{N}{\text{block size} \times 2}$$

Imbalance between treatment groups can occur if there are strata with incomplete randomized blocks, and there is a possibility of overall treatment imbalance even

though each stratum has more or less equal numbers in each treatment. Retrospective balancing by means of an analysis involving covariates can take care of possible imbalances, but balancing provides a more efficient comparison of treatments and trials in which the prognostic factors balanced are far more convincing than sophisticated analysis alone.

International Conference on Harmonisation (ICH) Guideline E9 [8] recommends stratifying by center and recognizes that balancing by prognostic variables may sometimes be valuable, having greater potential benefits in small trials. If stratifying by center is not possible, the Committee for Proprietary Medicinal Products (CPMP) [9] recommend stratifying by country or region. However, if there are a large number of strata compared to the number of patients, there is no guarantee of balance at site level and no guarantee of treatment balance.

### 3.2 Random Permuted Blocks within Strata

The most commonly used method of stratification is to categorize patients into several types (or strata) and then to allocate patients within strata using permuted blocks. The block size within each stratum will normally need to be somewhat smaller than for unstratified random permuted blocks, since the number of patients within a given stratum may be quite small. The smaller the size of the trial and the larger the number of strata, the smaller the block size should be. However, blocks of size 2 should be avoided if there is any chance that an investigator could predict the next assignment.

***Example*** In a recent study to investigate whether vaginal progesterone gel is effective in preventing preterm delivery in twin pregnancy (before 34 weeks of pregnancy), patients were randomized either to daily vaginal progesterone gel 90 mg or to placebo. The chorionicity of the pregnancy (monochorionic or dichorionic) was considered to be of major prognostic importance and was therefore included as a stratification factor along with the center. There were 6 recruiting centers that produced 12 ($6 \times 2$) strata consisting of all combinations of center and chorionicity.

Before the trial started, a randomization list would be produced for each of the patient strata using the method described in Section 2.2. In this example a block size of 4 could be used, or a mix of block sizes 2 and 4, or 4 and 6, to prevent predictability of assignment. Table 6 shows the randomization lists for 240 patients using the output from the random permuted blocks of sizes 4 and 6 given in Section 2.2.

As patients enter the trial, they are identified as belonging to one stratum. Each patient is then allocated the next available treatment assignment from the corresponding randomization list. For example, if the first patient was from Center 1 and had monochorionicity, she would be allocated to treatment B. If the second patient was from Center 5 and had dichorionicity, she would be allocated to treatment A.

Some studies will have just one major prognostic factor for stratified randomization while others may have several. However, using too many strata increases the number of lists required and so increases the administrative burden. Another problem with having too many strata is the possibility of empty cells or cells with very few patients. Table 7 illustrates the characteristics of the first 136 patients randomized by showing their distribution across the 12 strata.

**TABLE 6 Example of Random Permuted Blocks within Strata for a Trial in Twin Pregnancies**[a]

| Center | 1 | 2 | 3 | 4 | 5 | 6 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Chorionicity | M | M | M | M | M | M | D | D | D | D | D | D |
| | B | B | A | B | B | B | A | B | A | B | A | A |
| | A | A | B | A | B | B | B | A | B | B | B | B |
| | B | B | A | A | A | A | A | B | A | A | B | A |
| | A | A | B | B | B | B | B | A | B | A | A | B |
| | B | B | B | A | A | B | B | A | B | A | A | B |
| | A | A | A | A | A | A | A | A | A | B | B | A |
| | A | B | A | A | B | A | A | B | B | B | A | B |
| | A | B | B | B | B | A | B | B | B | A | B | A |
| | B | A | A | B | A | B | B | B | A | B | B | B |
| | B | A | A | B | B | B | A | B | B | A | B | A |
| | A | B | B | A | A | A | B | A | A | A | A | A |
| | A | A | B | A | A | A | B | A | A | B | A | A |
| | A | B | A | B | A | A | A | A | B | B | A | B |
| | B | A | A | A | B | B | A | A | A | B | B | B |
| | B | B | B | B | B | A | B | B | A | A | A | A |
| | B | A | B | B | B | B | A | A | B | A | B | A |
| | A | A | B | A | A | B | B | B | A | B | B | A |
| | B | A | A | B | A | A | B | B | B | A | A | B |
| | B | B | B | B | A | A | A | B | B | A | A | B |
| | A | B | A | A | B | B | A | A | A | B | B | A |

[a]M = Monochorionic, D = Dichorionic; A = Intervention; B = Control.

**TABLE 7 Distribution of Patients across Strata in a Twin Pregnancy Trial**

| Chorionicity \ Center No. | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| Monochorionic | 4 | 1 | 5 | 5 | 2 | 2 |
| Dichorionic | 33 | 8 | 17 | 13 | 32 | 14 |

Such an uneven distribution of patients across strata is typical and may possibly lead to imbalance. For a small trial with even more strata it is obvious that the chance of empty cells is even greater. In multicenter trials "center" is very often included as a stratifying variable as it is in this example. Indeed, this is recommended by the ICH Guidelines [8] and the Food and Drug Administration (FDA). However, for a trial involving a large number of centers this could lead to a very large number of strata. Therefore, in these cases it is worth considering using center as a minimization factor. This also reduces predictability of the next treatment allocation as centers are unaware of the characteristics of patients recruited by other centers. Overstratification can be particularly evident in small trials, but at the same time the possibility of serious imbalance is greater in small trials without stratification. An alternative approach would be to use minimization.

### Advantages of this method

- Randomization lists can be made up before the trial starts.
- Treatment allocation is unpredictable as long as the block size is kept secret.

- Allows for interaction between variables (treatment groups are balanced across *combinations* of prognostic factors).

**Disadvantages of This Method**

- The number of strata can quickly become large compared to the number of patients especially if center is one of the stratifying variables.
- When stratifying by center, it is desirable to keep the block size to a minimum as there is a risk of severe imbalance resulting from the cumulative effect of imbalance within each stratum. Recruitment officers do not need to know that blocked assignment is being used, and they should never be made aware of the block size.

### 3.3 Minimization

Minimization was first described by Taves in 1974 [10] and Pocock and Simon in 1975 [11] and aims to ensure treatment arms are balanced with respect to predefined patient factors as well as for the number of patients in each group. Minimization can be used when it is really important to achieve close similarity between treatment groups for several variables as it ensures balance between groups for several prognostic factors, even in small samples. Simulations show that minimization provides better balanced treatment groups when compared with restricted or unrestricted randomization, and it can incorporate more prognostic factors than stratified randomization methods such as permuted blocks within strata. Some more computationally complex methods may result in an even better performance. The loss in power resulting from treatment groups of unequal size is not likely to compromise a trial, but a serious imbalance in treatment groups with regard to a factor of prognostic importance can have severe consequences. Adjustment should always be made for minimization factors in the analysis (e.g., analysis of covariance).

***Example*** For a hypothetical trial with two treatments, intervention and control, assume there are three prognostic factors: sex (male or female), age (<40, 40–60 or >60), and previous surgery (yes or no). A minimization program was prepared and 10 hypothetical patients randomized. Table 8 demonstrates how minimization works by considering the treatment assignment of the next patient. If the 11th patient considered is male, aged 45, and has had previous surgery, the decision to allocate to the treatment or control group can be made by comparing the totals in these categories for both treatments and allocating the patient to the treatment group that gives most balance overall with a high degree of probability. Using Taves' method, the patient would always (i.e., with 100% probability) be assigned to the treatment group that gives best balance, and the allocation is said to be deterministic as no random factor has actually been used.

Pocock and Simon [11] defined a more general method where treatment assignment depends on three functions: the amount of variation among assignments for any given factor level, a measure of the total imbalance in treatment numbers and assignment probabilities to the $k$ arms of the trial $\{p_i\}$ (where $p_1$ is the probability of assignment to the arm that would lead to the least overall imbalance). Total imbalance is usually calculated by taking the sum of the individual imbalances, but

**TABLE 8 Example of How Minimization Works Using Hypothetical Trial Data**

| Prognostic Factor | Intervention | Control |
|---|---|---|
| Sex | | |
| Male | 4 | 3 |
| Female | 1 | 2 |
| Age group | | |
| <40 | 2 | 2 |
| 40–60 | 1 | 1 |
| >60 | 2 | 2 |
| Previous surgery | | |
| Yes | 2 | 4 |
| No | 3 | 1 |

a weighted sum can be used if some factors are considered more important than others. If $p_1 = 1$, the assignment is deterministic (as in Taves) and goes automatically to the treatment with least overall imbalance, but the $\{p_i\}$ can also be chosen to decrease the predictability of assignment. Table 8 provides a demonstration of how the range method works in practice.

If the 11th patient has factors: Male, Age 45, Yes, Has had previous surgery:

**Taves' Minimization**

Total in intervention group $4 + 1 + 2 = 7$

Total in control group $3 + 1 + 4 = 8$

The patient is allocated to the group with the lowest marginal totals.

Therefore the 11th patient is allocated to the intervention group because $7 < 8$.

**Pocock and Simon's Range Method** (Using Unweighted Sum and $p_1 = 1$)

If allocated to intervention group, total imbalance is $|(4 + 1) - 3| + |(1 + 1) - 1| + |(2 + 1) - 4| = 4$

If allocated to control group, total imbalance is $|4 - (3 + 1)| + |1 - (1 + 1)| + |2 - (4 + 1)| = 4$

The patient is allocated to the group that would lead to less overall imbalance. However, in this case the groups are equal. Therefore, simple randomization would be used to allocate a treatment to the 11th patient.

Although minimization does not ensure balance between combinations of factors, it does consider frequently occurring combinations of subcategories indirectly since they tend to work together, giving them a greater chance of being equally distributed than if they were independent.

This method cannot be implemented simply by preparing a randomization list since it is necessary to keep a record of all patient factors and allocations to date, as in Table 8. The easiest way to keep track is to have a custom-designed program (there are some generic minimization programs available). Alternatively, many clinical trials units and commercial companies sell minimization services, and the use of these can help to eliminate allocation or investigator bias.

### Advantages of the Method

- Balances treatment assignment across several prognostic factors simultaneously and across treatment groups overall.
- There is increased credibility by presenting data where prognostic variables are closely balanced within each treatment group [12].
- Planning to use minimization is a good discipline for making trialists think about prognostic factors before a study starts and for helping ensure adherence to the protocol as the trial progresses [13].

### Disadvantages of the Method

- For these dynamic allocation methods assignment lists cannot be prepared in advance.
- Although balance in the whole trial will be good, balance in a subgroup may not be. This is because minimization regards all the factors as independent (no interaction).
- If $p = 1$, then the assignment is deterministic, but if no investigator can predict the assignment, this may not be relevant. Decreasing the probability value can lead to treatment imbalance, so there is a need to select that $p$ that strikes a balance between minimizing imbalance and avoiding predictability of assignment. There is a counter argument that states that even with $p = 1$ it would still be possible for any individual to be allocated to any treatment; this is determined by the order in which they are recruited to the study, in effect by chance. The random element in the assignment arises because the factors of the next patient come from a random distribution.
- It may be more difficult to administer than other methods especially when the number of prognostic factors increases.
- A computer program may be necessary to perform the randomization and, this has potential implications for cost, programming error, system failure, and training. If a well-tested generic program is not available, it is important to use simulations to check that the method is working correctly before starting the trial.

ICH E9 recognizes that dynamic allocation procedures such as minimization can achieve balance across a number of factors simultaneously but recommend avoiding fully deterministic assignment. The ICH guidelines state that deterministic allocation procedures should be avoided, and an appropriate element of randomization should be incorporated, and that details of the randomization that facilitate predictability should not be contained in the protocol. Minimization may be used with a biased coin in favor of the winner with a suggested value of 0.8. Further extensions to the method include using combinations of factors, or some factors may be weighted more heavily if they are considered to be more important. However, minimization does consider frequently occurring combinations of subcategories indirectly since they tend to work together giving them a greater chance of being equally distributed than if they were independent [10]. The CPMP states that not only does balancing protect against chance imbalance but it also increases the statistical power for small trials and increases the power and credibility of subgroup analyses.

### 3.4 Minimization Compared with Simple Randomization

For varying trial sizes, the imbalance that might occur following simple randomization was simulated using the characteristics of the first 3000 participants recruited to a UK multicenter trial evaluating two interventions for the secondary prevention of bone fractures as the population. Key prognostic factors for this trial were sex, age group, presenting fracture, and time since fracture. The trial size was varied from small (100 patients) to large (2000 patients). Imbalance was deemed to occur if the percent of participants in any of the prognostic categories varied by more than 10% between trial groups. The rate of imbalance decreased nonlinearly with trial size. Small trials ($N \leq 100$) regularly displayed imbalance (over 80% of occasions), but in trials of 1000 participants or more, imbalance rarely occurred. Minimization ensured that balance was obtained on all occasions. There is, therefore, evidence of a trade-off between trial size and the need for minimization. Minimization is particularly useful for small to medium sized trials, but simple randomization may suffice when the trial is large. However, even large trials are small at the time of interim analysis and when producing reports for committees such as the Data and Safety Monitoring Committee (DSMC), so there is an argument that we should always use minimization.

### 3.5 Other Constrained Methods

Extensions to the minimization method include Atkinson's method, which uses optimum design theory [14]. Here the probability of allocation to the under-represented treatment responds to increasing imbalance rather than just being an arbitrary value. Klotz's method [15] addresses the question of a trade-off between uncertainty and imbalance. His method is similar to that of Pocock and Simon but is more computationally intensive and describes a function for calculating optimal treatment randomization probabilities. Titterington proposes a similar method [16] that involves minimizing a quadratic criterion subject to a balance constraint. The method aims to compromise between unpredictability of assignment and good stratification balance and claims to be simpler than Klotz's method. Signorini and co-workers outlined a dynamic balancing method [17] described as an easily implemented method that balances treatment allocations both within strata and across the trial as a whole. The method keeps a tally of total treatment allocations across all strata. Simulations have been used to demonstrate that the major imbalances that are possible with other schemes do not occur using this method, and the potential for selection bias is much reduced.

However, within the United Kingdom, these methods are very seldom used [18], and the results of a recent survey of trialists conducted by our group was to keep the method of randomization as simple as possible. When selecting a method, one must only balance on scientifically justifiable factors. If a stratified method can cope, then this is the method of choice; if there are too many factors, then use a dynamic method. Predictability may be an important factor as there is a risk of selection bias if the recruitment officer can guess the next allocation. Dynamic techniques help to make the allocation less predictable.

## 4 RESPONSE-ADAPTIVE METHODS

As a trial progresses, it may become evident that one treatment is performing better than the other. It may be possible to make use of this information by varying the allocation ratio to ensure more patients are allocated to the apparently better treatment. In the "play-the-winner" rule of Wei and Durham [19] the probability of assignment varies according to the success of the treatments. Initially, they conceptualize an urn containing equal numbers of balls representing the two treatments, A and B. A ball is drawn and replaced and the next patient is allocated to that treatment. If a treatment produces a successful outcome, then a ball of that type is added to the urn. If a treatment produces a nonfavorable outcome, then a ball of the other type is added to the urn. Therefore the probability of assignment is skewed to favor the treatment performing "better".

This adaptive approach to treatment assignment results in a substantial reduction of patients on an inferior treatment compared to a conventional randomized trial. However, this design is not often used as the analysis of such designs is problematic. It is recommended that assignment procedures be simple, objective, and foolproof, and this is difficult to achieve with an adaptive design.

## 5 SPECIAL-CASE SCENARIOS

### 5.1 Unequal Randomization

In the majority of trials patients are randomized equally between the experimental and control groups. Sometimes it is preferable to favor the experimental treatment where anecdotal evidence suggests there may be a great health benefit, and patients may be reluctant to have only a 50% chance of receiving the new treatment. Such unequal allocation schemes may help to improve recruitment. It should be cautioned, however, that if the experimental treatment is harmful, unequal randomization in this situation may subject more patients to a harmful treatment.

If research costs between the treatments differ, it may be more cost effective to randomize more patients to the cheaper treatment. However, economic reasons are not the only criterion for selecting unequal randomization ratios, and in some cases it may even be preferable to favor the more expensive treatment. For example, when evaluating the learning curve for a new surgical intervention, more patients may be randomized to the experimental (or expensive) treatment [20]. For many trials there is a substantial amount of previous experience relating to the control treatment, and there is more to be learnt about the new treatment [5].

When experimental treatments differ in their costs, then for a given statistical power, unequal randomization will produce the least-cost trial. Unequal randomization is unnecessary if there is no cost difference between the treatments. However, it may be possible to alter the randomization ratio as the trial proceeds if large cost differences become apparent. If unbalanced allocation is used, it is generally justified on the grounds of ethics or cost. Randomization ratios of 2:1 or 3:2 do not greatly alter the statistical properties of a trial, but loss of power appears to be substantial for a ratio of 3:1 or more extreme [5]. The chance of obtaining a statistically significant difference between the two treatments is not significantly

reduced, as long as the chosen ratio is less extreme than 70:30 [21]. Formulas exist to adjust standard sample size estimates to accommodate for unequal randomization [22].

### 5.2 Cluster Randomization

A cluster-randomized trial is one in which clusters of individuals, rather than individuals themselves, are randomized to different treatment groups. Cluster-randomized trials have become particularly widespread in the evaluation of non-therapeutic interventions such as lifestyle modification or educational programs. The units of randomization range from relatively small clusters such as families to entire communities such as neighbourhoods, hospitals, or medical practices. In a trial to investigate if training in a certain consultation procedure is effective, then some physicians would receive training and others would not. The patients being seen by a physician would form a cluster. If the individual patients were randomized, then they may have to be seen by another physician as it would prove difficult for the physician to switch consultation practice between patients and contamination could occur.

The clusters are often decided in advance and randomized before the intervention is applied. However, the subjects themselves are sometimes not identified at the time of randomization, and so the randomization will not balance the prognostic factors as in individual randomization. The purpose of randomization in cluster trials is to try and balance the confounding factors associated with the cluster [23]. Cluster unit characteristics will balance as the number of clusters increases, and so the patient characteristics will balance as a result. The methods described in Sections 2–4 can all be used to allocate clusters.

The design and conduct of cluster-randomized trials requires special considerations [24]. They are often more complex to design, require more participants, and require more complex analyses than individually randomized trials. Cluster randomization is less efficient in a statistical sense than randomizing individuals to treatment groups [25] because the responses of individuals within a cluster tend to be more similar than responses of individuals in different clusters. The theory of experimental design assumes that the experimental unit that is randomized is also the unit of analysis, and this is not necessarily the case in a cluster randomization trial. Analyses that do not take the clustering into account may report significance when none exists. Increasing the number of clusters enhances power more efficiently than increasing the number of subjects within a cluster.

**Advantages of This Method**

- Cluster randomization is often used for practical or ethical reasons when patient randomization is not possible.
- Clustering helps to avoid contamination.

**Disadvantages of This Method**

- Loss of statistical efficiency
- Need to recruit more participants

## 6 SUMMARY

In conclusion, the most important factors when choosing a randomization method are the size of the trial, practicality of the method, maintaining the blind, and reducing predictability. Smaller trials may require some form of restricted allocation to maintain balance, whereas simple randomization may suffice for large trials. Keeping the randomization method as simple as possible will reduce time, cost, and programming errors.

If the number of important prognostic factors and layers within these is sufficiently small, then the preferred method of randomization, recommended by ICH, is permuted blocks of varying random length within strata. This method eliminates the problem of predictability and at the same time balances across combinations of factors, whereas minimization only balances over single factors. If the number of prognostic factors is large, then minimization can be used to provide treatment balance as well as balance over these factors. However, only those factors known to affect outcome should be considered.

Web-based systems may become more widely used in the future, but there are some problems to be overcome, including access to the Internet, speed and reliability of access, and the lack of willingness in trial centers to adopt new technologies.

## APPENDIX I

Visual Basic for Applications (VBA) code to produce a table of 2500 random numbers.

```
Private Function GenerateList()
   Dim Num As Integer
   Dim MyValue
   Dim Numbers(25) As Integer
   Dim intI As Integer

   Open "random.txt" For Output As #1 ' Open file.
   For Num = 1 To 100

   For intI = 0 To 24
              Randomize ' Initialize random-number
              generator.
              MyValue = Int((10 * Rnd))' Generate random
              value between 0 and 9.
              Numbers(intI) = MyValue
       Next

        Print #1, Numbers(0); " "; Numbers(1); " ";
        Numbers(2); " "; Numbers(3); " "; Numbers(4); " ";
        Numbers(5); " "; Numbers(6); " "; Numbers(7); " ";
```

```
        Numbers(8); " "; Numbers(9); " "; Numbers(10);
        " "; Numbers(11); " "; Numbers(12); " ";
        Numbers(13); " "; Numbers(14); " "; Numbers(15);
        " "; Numbers(16); " "; Numbers(17); " ";
        Numbers(18); " "; Numbers(19); " "; Numbers(20);
        " "; Numbers(21); " "; Numbers(22); " ";
        Numbers(23); " "; Numbers(24);
        Print #1,

        Next Num
        Close #1

   End Function
```

## APPENDIX II

Visual Basic for Applications (VBA) code to produce a table of 100 permutations of 20 numbers (0–19).

```
Private Function Permutations()
  Dim Num As Integer
  Dim intI As Integer
  Dim db As Database
  Dim rst As Recordset
  Dim rstAdd As Recordset

  Set db = CurrentDb
  Set rstAdd = db.OpenRecordset("TblPermute20")

  For Num = 1 To 100
    rstAdd.AddNew
    Set rst = db.OpenRecordset("QrySort20")
    For intI = 0 To 19
      rstAdd(intI) = rst!ID
      rst.MoveNext
    Next intI
    rstAdd.Update
  Next Num

  rst.Close
  rsAdd.Close

  MsgBox "Done!", vbOKOnly

End Function
```

## APPENDIX III

Visual Basic for Applications (VBA) code to produce a table of approximately 200 treatment allocations in random blocks of 2 and 4.

```
Private Function Blocks()
  Dim db As Database
  Dim arrBlock() As Integer
  Dim intBlock As Integer
  Dim intBlockSize As Integer
  Dim intCount As Integer
  Dim intResult As Integer
  Dim strInsert As String

  Set db = CurrentDb

  For intBlock = 1 To 45
    Randomize

    If (Rnd > 0.5) Then
       intBlockSize = 6
      Else
        intBlockSize = 4
      End If

      ReDim arrBlock(intBlockSize - 1)

      For intCount = LBound(arrBlock) To UBound(arrBlock)
        arrBlock(intCount) = 0
      Next

      intCount = 1
      While intCount <= intBlockSize / 2
       intResult = Int((intBlockSize - 1 + 1) * Rnd + 1)
       If arrBlock(intResult - 1) <> 1 Then
           arrBlock(intResult - 1) = 1
           intCount = intCount + 1
        End If
      Wend

      For intCount = LBound(arrBlock) To UBound(arrBlock)
        strInsert = "INSERT INTO TblAllocate(Treatment)
        VALUES(" & arrBlock(intCount) & ")"
        db.Execute strInsert
       Next
      Next
    End Function
```

## APPENDIX IV

Visual Basic for Applications (VBA) code to produce a table of 50 treatment allocations using the Urn method.

```
Private Function Urn()
  Dim db As Database
  Dim strInsert, Treat As String
  Dim Num, x As Integer
  Dim randnum, p As Double

  Set db = CurrentDb

  Num = 0
  For x = 1 To 50
    Randomize

    If x = 1 Then
       p = 0.5
    Else
       p = Num / (x - 1)
    End If
    If (Rnd < p) Then
       Treat = "A"
    Else
       Treat = "B"
    End If
    If Treat = "B" Then
       Num = Num + 1
    End If

    strInsert = "INSERT INTO TblUrn(Treatment) VALUES('" &
    Treat & "')"
    db.Execute strInsert
  Next x

End Function
```

## REFERENCES


1. Fergusson, D., Aaron, S. D., Guyatt, G., et al. (2002), Post-randomisation exclusions: The intention to treat principle and excluding patients from analysis, *BMJ*, 325, 652–654.
2. Altman, D. G., and Schulz, K. F. (2001), Statistics notes: Concealing treatment allocation in randomised trials, *BMJ*, 323, 446–447.
3. Byrom, B. (2004), Electronic diary solutions: Enhanced collection of patient reported outcome data, *EBR*, Autumn, 90–94.

4. Efron, B. (1971), Forcing a sequential experiment to be balanced, *Biometrika*, 58, 403–417.
5. Pocock, S. J. (1979), Allocation of patients to treatment in clinical trials, *Biometrics*, 35, 183–197.
6. Wei, L. J. (1978), An application of an urn model to the design of sequential controlled clinical trials, *J. Am. Statist. Assoc.*, 73, 559–563.
7. Therneau, T. M. (1993), How many stratification factors are "too many" to use in a randomization plan? *Controlled Clin. Trials*, 14, 98–108.
8. ICH Steering Committee (1998), ICH Harmonised Tripartite Guideline E9; Statistical Principles for Clinical Trials.
9. Committee for Proprietary Medicinal Products (CPMP) (2003), The European Agency for the Evaluation of Medicinal Products (EMEA). Points to consider on adjustment for baseline covariates. CPMP/EWP/2863/99 at www.emea.europa.eu.
10. Taves, D. R. (1974), Minimization: A new method of assigning patients to treatment and control groups, *Clin. Pharmacol. Ther.* 15, 443–453.
11. Pocock, S. J., and Simon, R. (1975), Sequential treatment assignment with balancing for prognostic factors in the controlled clinical trial, *Biometrics*, 31, 103–115.
12. Brown, B. W. (1980), Statistical controversies in the design of clinical trials, *Controlled Clin. Trials*, 1, 13–27.
13. Day, S. (1999), Commentary: Treatment allocation by the method of minimisation, *BMJ*, 319, 947–948.
14. Atkinson, A. C. (1982), Optimum biased coin designs for sequential clinical trials with prognostic factors, *Biometrics*, 69, 61–67.
15. Klotz, J. H. (1978), Maximum entropy constrained balance randomization for clinical trials, *Biometrics*, 34, 283–287.
16. Titterington, D. M. (1983), On constrained balance randomization for clinical trials, *Biometrics*, 39, 1083–1086.
17. Signorini, D. F., Leung, O., Simes, R. J., et al. (1993), Dynamic balanced randomization for clinical trials, *Stat. Med.*, 12, 2343–2350.
18. Scott, N. W., McPherson, G. C., Ramsay, C. R., et al. (2002), The method of minimization for allocation to clinical trials: A review. *Controlled Clin. Trials*, 23, 662–674.
19. Wei, L. J., and Durham, S. (1978), The randomized play-the-winner rule in medical trials, *J. Am. Statist. Assoc.*, 73, 840–843.
20. Torgerson, D., and Campbell, M. (1997), Unequal randomisation can improve the economic efficiency of clinical trials, *J. Health Services Res. Policy*, 2, 81–85.
21. Peto, R., Pike, M. C., Armitage, P., et al. (1976), Design and analysis of randomized clinical trials requiring prolonged observation of each patient, *Br. J. Cancer*, 34, 585–612.
22. Campbell, M. J., Julious, S. A., and Altman, D. G. (1995), Estimating sample sizes for binary, ordered categorical, and continuous outcomes in two group comparisons, *BMJ*, 311, 1145–1148.
23. Machin, D., and Campbell, M. J. (2005), *Design of Studies for Medical Research*, Wiley, Chichester, p. 274.
24. Campbell, M. K., Grimshaw, J. M., and Elbourne, D. R. (2004), Intracluster correlation coefficients in cluster randomized trials: Empirical insights into how they should be reported, *BMC Med. Res. Method.* 4.
25. Donner, A., and Klar, N. (2000), *Design and Analysis of Cluster Randomization Trials in Health Research*, Arnold, London.

# *Randomized Controlled Trials*

**Giuseppe Garcea[1] and David P. Berry[2]**

[1]*Cancer Studies and Molecular Medicine, The Leicester Royal Infirmary, United Kingdom*

[2]*Department of Hepatobiliary and Pancreatic Surgery, The Leicester General Hospital, United Kingdom*

**Contents**

1 Introduction
2 Uncontrolled Trials
3 Problems with Uncontrolled Trials
4 Historical Controls
5 Problems with Historical Controls
6 Concurrent Nonrandomized Controlled Trials
7 Problems with Concurrent Nonrandomized Controlled Trials
8 Is Randomization Feasible?
9 Conclusion
References

## 1 INTRODUCTION

A randomized controlled trial is a study in which subjects are allocated at random to receive one of two possible interventions. The outcomes measured are events or quantifiable changes that are present or absent following the proposed interventions. These outcomes are known as endpoints. In general, one of the interventions chosen is a standard of comparison: the control. The control intervention can be a drug or procedure with no known clinical effect: a placebo. Alternatively, the

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

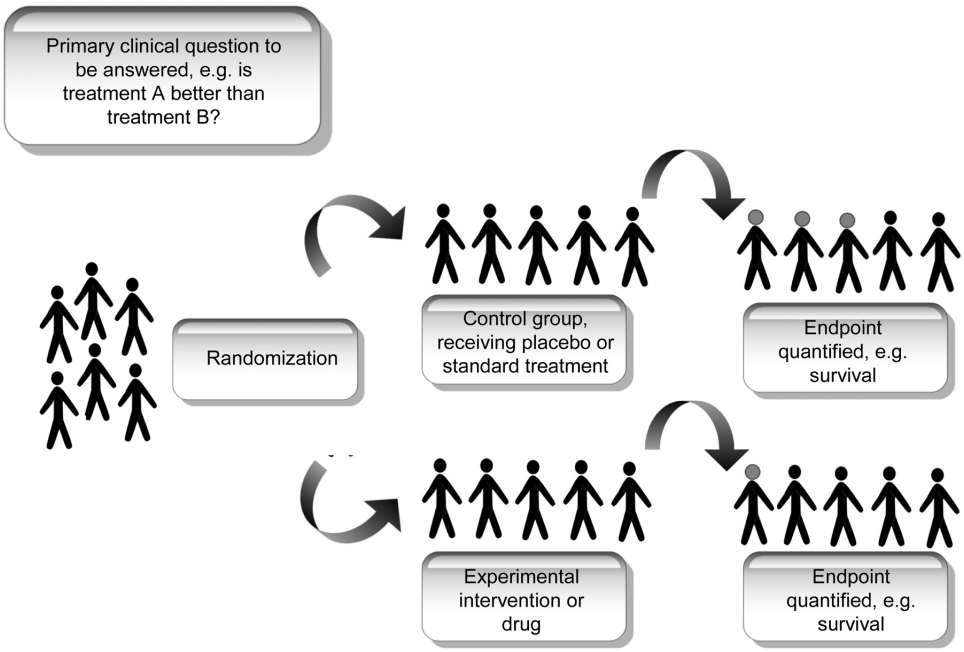


***FIGURE 1*** *Structure of randomized controlled trials.*

control can be a therapeutic measure that is currently the accepted treatment or intervention for a disease process. The other group is the experimental intervention or trial drug (Fig. 1).

Randomization refers to the random allocation of the subjects to one of the intervention groups. Therefore, the allocation of patients is not determined by the clinical investigators or the subjects within the study. The random allocation of subjects to the treatment arms of the study ensures that the demographics of the subjects are likely to remain similar. Randomization also removes investigator bias in the allocation of patients to the various treatment arms. Hence, randomization reduces the risk of imbalance between the subjects receiving the trialled intervention or drug and those receiving the standard intervention or placebo drug. This balancing of prognostic factors between the two groups increases the likelihood that any observed difference is attributable solely to the proposed treatment.

A number of different endpoints can be employed to quantify the response to a trialled intervention. Endpoints can vary from laboratory-measured responses, such as changes in cell-signaling molecules or histological changes, known as biomarkers, through to clinical responses. Ideally, randomized controlled trials should have one primary objective, asking one specific question and one primary outcome or endpoint. Both the primary objective and endpoints should be well defined and easily measured. A quantifiable objective endpoint is superior to a subjective one. Many randomized controlled trials will also have secondary objectives and endpoints. In general, conclusions drawn from these secondary objectives should be regarded as tentative since the study may lack the power to address them.

A randomized controlled trial (RCT) should answer a clinical need (Fig. 2). Any intervention studied should be feasible in clinical practice and remain a viable treat-

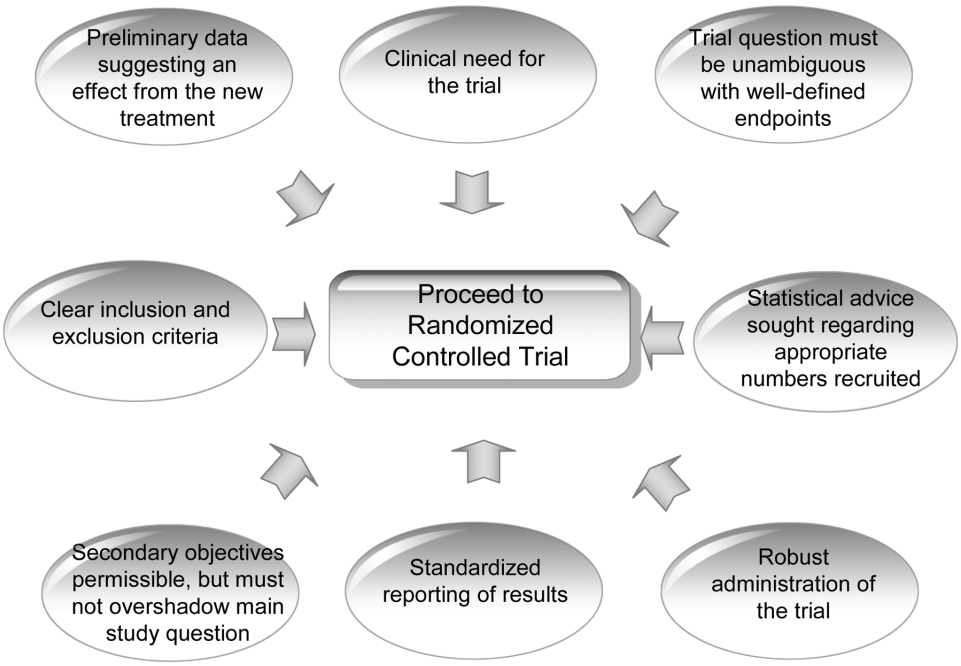


***FIGURE 2*** *Issues in designing a randomized controlled trial.*

ment following the results of the study. The disease process in question should be of a serious nature or impact upon a significant proportion of the population. There must be preliminary evidence (from phase I and phase II trials) that the trial intervention has a promise of efficacy. However, there also needs to be a genuine uncertainty regarding the efficacy of the new proposed treatment over the standard treatment or placebo treatment, known as equipoise. The design of the RCT should be robust and proper administration of the study is essential. Standardized reporting of the results of any RCT, combined with valid statistical analyses allows data from multiple RCTs to be compared for a more precise estimate of the effects of a treatment.

In practice other study methodologies have been used to answer the same clinical questions tackled by RCTs. This chapter will examine these methods and discuss their potential flaws.

## 2 UNCONTROLLED TRIALS

In some circumstances it may not be possible to conduct randomized controlled trials. This may be due to ethical problems (e.g., examining the effect of cessation of smoking and the risk of lung cancer), financial considerations, or, as occurs frequently in surgical practice, problems in recruiting adequate numbers of patients. In these scenarios, uncontrolled trials may give important clinical information. Uncontrolled studies are essentially observational trials where the effect of an intervention is assessed on a population of patients (a cohort) over a period of time. In uncontrolled trials, there is no control group and hence all endpoints are essen-

tially qualitative (in respect to the patient's initial symptom or disease status). In drug therapy, uncontrolled trials may be a useful bridging step from translating in vitro and in vivo evidence of drug efficacy into a clinical setting. Hence, the results from uncontrolled clinical trials may be used to rationalize and plan subsequent RCTs of the same intervention. In particular, uncontrolled clinical trials can provide important safety and tolerability data in the setting of drug trials.

Surgery is one field where the RCT has suffered a notable decline [1–4]. In one report, less that 40% of the published hypotheses involving surgical operations for gastrointestinal surgery could have been answered successfully by an RCT [5]. The main problems encountered included accrual of patients due to a low disease incidence and patient preference. For example, the mastectomy versus lumpectomy RCT for breast cancer suffered from significant poor accrual into the latter treatment group [6]. In such circumstances a noncontrolled cohort study may be the only means of evaluating the efficacy of a surgical intervention.

Uncontrolled trials are also useful for chronic conditions when compared to placebo controlled RCTs. In this group of patients long-term follow-up during RCTs is plagued by poor patient adherence to treatment regimens, especially in the placebo control group. There are also ethical considerations in administering a nontherapeutic drug for long periods to patients with chronic illness, which is severely restricting their quality of life. In such scenarios patients taking their preferred (and therapeutic) treatments are more likely to comply with the trial, and hence longer-term follow-up can be achieved with minimal loss of patients recruited to the trial. Uncontrolled observational studies can also recruit larger numbers of patients with fewer inclusion and exclusion criteria than RCTs. One criticism of RCTs is that they often recruit relatively small numbers of patients (in comparison to the disease burden) who are then treated and monitored under conditions that do not reflect everyday practice [7, 8].

## 3 PROBLEMS WITH UNCONTROLLED TRIALS

While we have discussed the clinical pressures that necessitate the undertaking of noncontrolled trials, there are major disadvantages in this type of study and assessing the reliability of the evidence reported. The lack of a control group makes it impossible to compare any new treatment with no treatment or with the "gold standard" of treatment currently adopted with any particular disorder. An example would include a hypothetical series of laparoscopic hernia repairs reported by surgeon A, which reveals a low recurrence rate, a high rate of same-day discharge, and a high level of patient satisfaction following the procedure. With no control group, we cannot compare this treatment with the current standard method of treating hernias (i.e., an open repair). As a result, although the results may indicate that laparoscopic surgery is a viable technique in dealing with hernias, we cannot assess if it is more efficacious than the currently adopted treatment regimen. In the field of testing new drugs, this may have important implications since many new drugs available on the market would incur significantly greater costs if their use was to be adopted universally, when compared to cheaper and better established medication.

Furthermore, with the lack of a control, any observed effect of drug or intervention is subject to systematic error or bias within the study. This bias may arise from

subject selection, outcome measurement, or other confounding factors. For example, in the case of a new drug treatment for insomnia, the patients selected may be well-motivated individuals who, in addition to taking the new trial drugs, have also instituted other measurements to ease stress and hence improve their ability to sleep at night. The investigator is clearly aware that all individuals assessed following the drug treatment have received the trial drug and this may affect their data collection. In studies where qualitative assessments are to be used (such as the quality of sleep overnight), this observer bias is likely to be of particular importance. Finally, the placebo effect of the new drug has not been accounted for. The placebo effect is a well-recognized response of patients to new treatments or drugs. The fact that these patients with insomnia have been selected to trial a new and potentially efficacious medication may in itself help to ease their symptoms, even if the trial drug had no pharmaceutical action at all.

Uncontrolled trials can be a useful adjunct in determining drug tolerance and acceptability in phase I and phase II clinical trials. However, results from these trials, however encouraging, should be interpreted with caution pending control group studies.

## 4 HISTORICAL CONTROLS

It is possible to compare outcome from new treatments in trials with data from previous studies examining alternative treatments or no treatment. Hence, the control group is named a historical control (Fig. 3). A major advantage in using

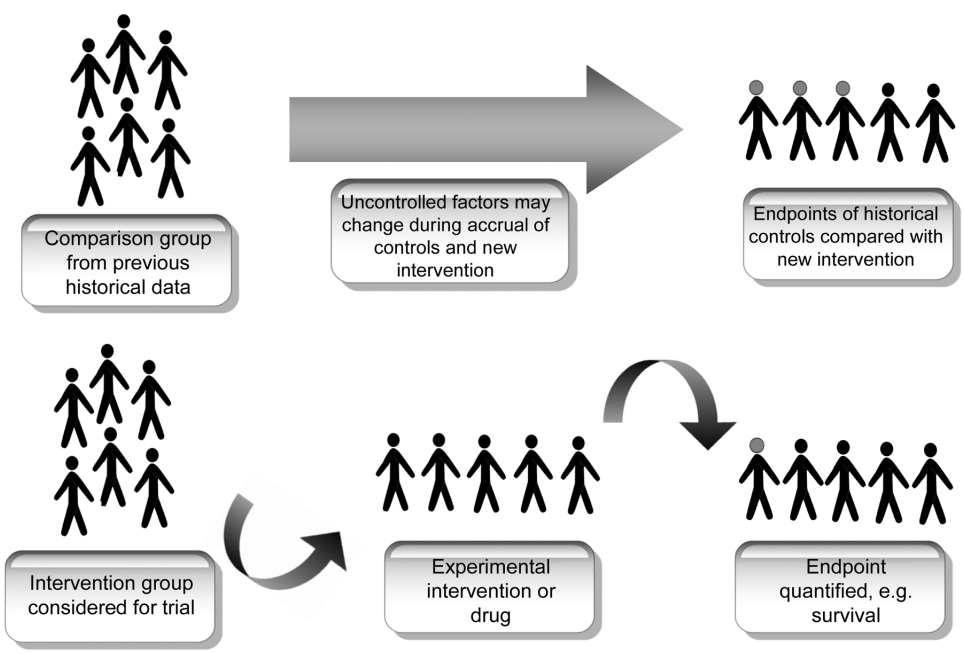


***FIGURE 3*** *Structure of historical control trials.*

historical controls is that the studies are easier to perform, with fewer patients and shorter study period duration than studies that use concurrent controls. The treatment of diabetes mellitus with insulin, chemotherapy for childhood leukemia, and vitamin $B_{12}$ therapy for pernicious anemia all represent major medical advances that have been instituted as a result of data obtained from historical controls [9–12].

Historical controls may also be instituted for conditions that are sufficiently rare to result in a low rate of accrual of eligible patients for concurrent control studies. By comparing data from historical controls, it may be possible to evaluate new treatments in a much shorter time period than would otherwise be possible. In areas where financial considerations may be a limiting factor, historical control groups are an attractive proposition since they avoid the need to recruit control groups and randomization for treatment allocation. In doing so, historical controls reduce the need of support personnel required and cost.

The final argument for historical controls is an ethical one. In general, RCTs compare a newer and presumed better drug with standard therapy. If no effective established therapy exists for a particular condition, it may be argued that ethically all patients should receive the new treatment. For example, if comparing insulin treatment for the treatment of diabetes when compared to no treatment at all, those patients allocated to the control arm of an RCT would be denied a life-saving treatment.

## 5 PROBLEMS WITH HISTORICAL CONTROLS

The major drawback to using historical controls is that many factors may change in the time period from accrual of data from historical controls to the implementation of the new drug or intervention in the intervention subjects. These differences may reflect the continuous improvement of medical care, changes in the demographics of patients referred, changes in referral patterns (such as earlier referral due to increased awareness of a particular condition), and improvements in epidemiological factors (better diet, increased exercise). For the intervention group, the increased monitoring and attention will also have a strong placebo effect, which will not be adequately controlled for among the historical controls.

In the setting of cancer drug trials, for example, the use of historical controls relies on the assumption that the survival rate among the controls would be the same as that among the control group in a concurrent RCT. The factors discussed in the preceding paragraph suggest that this may not be true. These confounding factors can be minimized if the data from the historical controls is from patients who were all treated at the same institution. This would ensure that referral patterns, medical care, nursing care, and other support would be as similar as possible to the intervention group [10]. It is probable, however, that some differences will inevitably exist between the control and intervention group.

Another potential flaw in the use of historical controls is that for most trialled treatments there are restrictions on which patients are selected for the intervention group. For example, in the case of chemotherapy agents the patients with worst prognosis may be excluded from the trial since they may be least likely to benefit

from the therapy. These exclusion criteria cannot be applied to the historical controls. Since exclusion and inclusion criteria cannot be adequately controlled for in the historical group, this invalidates any comparison between them.

Due to the overall improvements in medical and nursing care over the time interval separating intervention groups from their historical controls, there is a tendency for historical control trials to overemphasize the magnitude of any particular intervention. For example, the use of anticoagulant therapy following acute myocardial infarction using data from historical controls revealed a 54% reduction in mortality [13]. Subsequent concurrent RCTs showed a 21% reduction in mortality. This discrepancy occurred in spite of broad database consisting of 18 studies and 9000 patients in all. While it may not be possible to eliminate this "historical error," it is possible to give an approximation as to the degree of error for a historically controlled study. This may be achieved by portioning the period of control entry into different intervals. The results obtained between those groups can be then compared to the overall outcome. If a significant reduction in the benefit from a new intervention is observed across the different groups, then the degree of error due to changes in clinical practice is significant (Fig. 4).

In a study by Diehl and Perry the survival times and disease-free survival times of historical controls from chemotherapy trials were compared with those from control groups derived from matched RCTs [9]. The RCTs were matched for location of disease, stage, and follow-up. This study showed that 42% of studies varied by more than 10 percentage points, 21% varied by more that 20 percentage points, and 5%

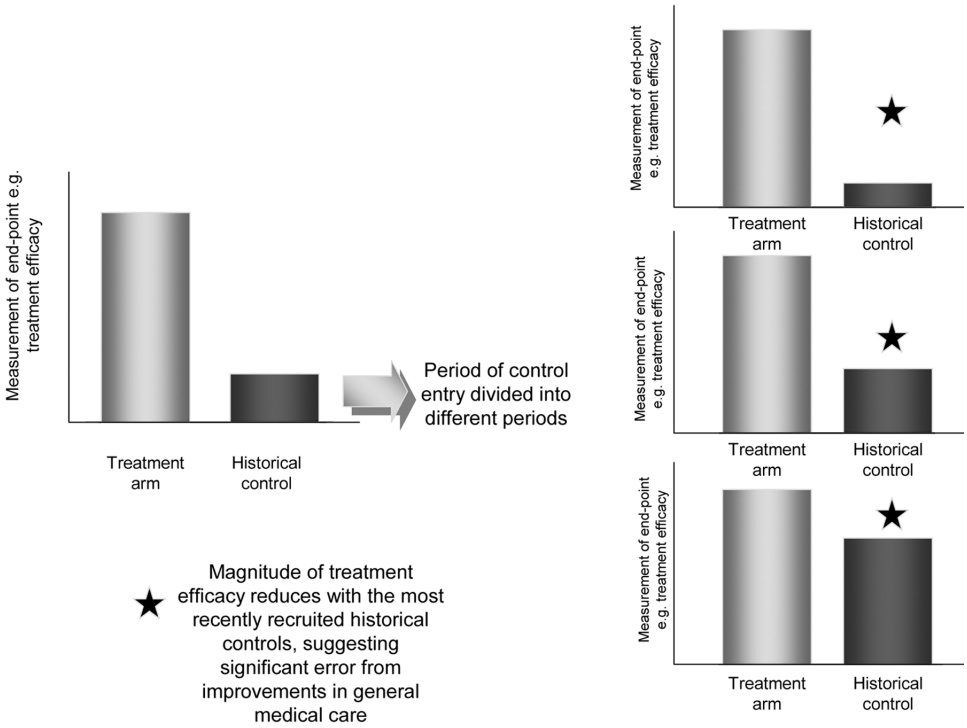


***FIGURE 4*** *Schematic representation from error occurring in historically controlled trials.*

varied by more that 30 percentage points. It also revealed that even for institutions that had both a historical control and randomized control group, there were still significant differences between survival and disease-free survival. In summary, although historical controls offer certain advantages over RCTs in terms of simplicity, speed, and cost, they cannot be used to reliably control treatment studies.

## 6 CONCURRENT NONRANDOMIZED CONTROLLED TRIALS

In certain scenarios a nonrandomized design may be applicable. Nonrandomized trials still include a control and intervention group. However, allocation of patients to control or treatment arms is determined by the investigator or the patient themselves (Fig. 5). There are numerous valid reasons for the use of nonrandomized controlled trials in clinical practice. Surgery is an area where the conduction of RCTs is especially problematic. These relate to the difficulty in patients allowing themselves to be randomized. Patients who may be eligible for randomization may withhold consent to be randomized. This may be due to the fact that the new treatment under trial may be less cumbersome, less traumatic, or less disfiguring that the standard therapeutic procedure. Patients undergoing surgery will need to coordinate the timing of their procedure and recovery time around their employment and hence may be unwilling to be randomized into a treatment arm that would cause more disruption for them.

A similar circumstance can be envisaged in a hypothetical chemotherapy trial comparing an intravenous chemotherapy regime, with one using orally administered medication. The former treatment arm will necessitate the positioning of

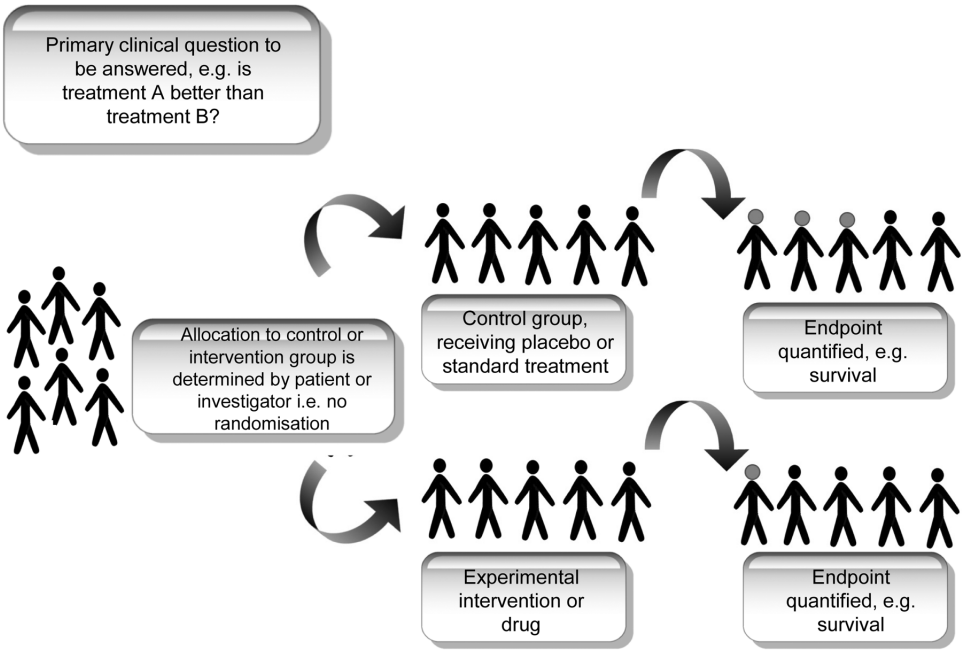


***FIGURE 5*** *Structure of nonrandomized concurrent control trials.*

intravenous cannulae and a visit to a hospital daycare unit, while the oral preparation may be taken at home. Many patients would prefer the less inconvenient option, particularly if the oral preparation is the newer and perceived (if only by the patients) more effective treatment for their cancer.

For certain diseases, the population of patients may be too small to accrue sufficient numbers for an effective RCT. In such situations, every patient is valuable, and hence the investigators may allow individuals to choose their treatment in order to accumulate sufficient numbers of patients. Finally, there may be ethical considerations that make randomization unfeasible. For example, in the case of the hypothetical orally administered chemotherapy agent discussed above, patients with extensive carcinomatosis and a poor prognosis, randomization into the standard treatment arm will have a significant impact on their remaining quality of life. It may be a better option to allow patients to choose their treatment, which, in turn, may result in better compliance among those patients who have chosen to be in the treatment arm examining standard therapy.

The perceived disadvantages in losing randomization has been challenged by Abel and Koch [14]. The argument for randomization is that it leads to comparable and balanced groups and so allows comparison of the new treatment arm of the study with the standard treatment or placebo arm. Having ensured that baseline characteristics have been eliminated by randomization, any differences in effect are attributable to the interventions applied. However, this assumption is inaccurate as it does not take into account the many variables that may affect outcome following allocation to treatment or control groups. Figure 6 summarizes the causes of differ-

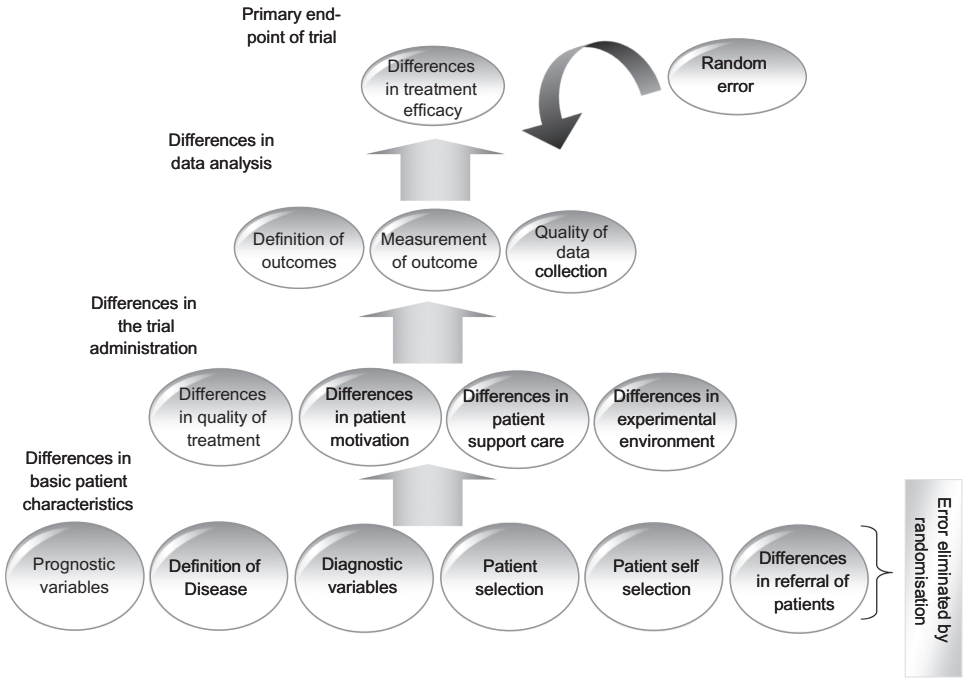


***FIGURE 6*** *Pyramid of possible reasons for differences observed between treatment in clinical trials and the effect of randomization on eliminating errors. (Adapted from Abel and Koch [14].)*

ences between control and treatment groups and which ones are eliminated by randomization. It is clear that there are still many more confounding factors that may alter the findings of a study erroneously.

Blinding is a concept in which patients and physicians are unaware of which treatment arm the patient has been assigned. Blinding prevents patients from overstating the benefits of a new medication in order to build rapport or please their attending doctor. It also prevents physicians from introducing observer bias when recording and interpreting the results of a tested intervention. Another stated disadvantage to nonrandomized controlled trials is that blinding is not possible. While randomization is essential for blinding of patients, it is not a prerequisite for blinding of treatment evaluation.

Nonrandomized controlled studies are a response to a clinical need to evaluate a new therapy, where the applicability of a conventional RCT is in doubt. A comparison of 30 randomized and nonrandomized controlled studies comparing treatment with no treatment found no influence of randomization on the mean observed treatment effect [15]. Other studies, however, have shown that the magnitude of effect observed may vary with randomization [16–19] for both medical and surgical clinical trials. In summary, while nonrandomized controlled trials can probably reliably predict a treatment effect, the magnitude of the effect may vary from that observed in RCTs.

## 7 PROBLEMS WITH CONCURRENT NONRANDOMIZED CONTROLLED TRIALS

While the loss of the traditional advantages of randomization do not necessarily excessively penalize a nonrandomized study, there are other benefits from randomization that are now discussed. Figure 6 demonstrates that there are many other points of error that are not eliminated by randomization; however, randomization of the population study group does at least remove one tier of possible error in that prognostic factors, both known and unknown are balanced between intervention and comparison groups.

In the conduct of clinical trials, nonconsent bias is removed. For example, in our hypothetical orally administered chemotherapy agent, it may be that patients who are not willing to travel to the hospital for the intravenous standard treatment are those too frail or immobile to make such repeated journeys. In such circumstances, the true benefit of the new oral agent may not be fully appreciated due to the confounding factors of age and general fragility affecting long-term survival of the intervention arm. In addition, it would be relatively easy for an investigator to sway a patient's decision making by recommending one intervention over another. In this manner, patients with the smallest disease burden or better prognosis may be chosen for the trial arm, which would of course affect the validity of the study.

Randomized controlled trials require a significant support framework to administer treatment to patients and for appropriate supervision and data collection. In this manner, randomization compels the investigator to plan the conduct of the trial and introduce written protocols. While this is not a direct advantage of randomization, it is an advantage offered by the RCT trial framework. A second indirect benefit from introducing an RCT is that in many cases multicenter trials are required

due to a number of reasons, such as ensuring adequate recruitment of patients, and pooling of financial resources. By using a multicenter framework, many of the differences between the deliveries of patient care between centers is diluted. In this manner, the population group of the RCT is more representative of the population of patients with the disease out in the community.

The implementation of an RCT may have other beneficial effects within the trial center. RCTs may be a learning experience for clinicians in the planning, data collection, and analysis of trials. The RCT teaches a methodological approach to implementing trial therapies. The skills learnt from this can be used in the critical evaluation of other published evidence by clinicians and can directly cross over to the daily ward management of patients. The written protocols for RCTs will stipulate the manner in which the trial intervention is administered and how it is monitored. These trial tests need to run along the normal administration of care. As a result, this requires medical and nursing staff reevaluate the patient pathway through their unit. This may frequently be an opportunity to further improve the delivery of care, even for nontrial patients.

In conclusion, the pitfalls of not randomizing controlled studies are illustrated by a trial of vitamin C in patients with advanced colorectal cancer [20]. One hundred patients with terminal colorectal cancer were given high-dose vitamin C. The survival of these patients was then compared with 1000 control patients who received no vitamin C. The study found that the mean survival time in the vitamin C patients was 4.2 times greater than those in the control group. Despite these results, two further RCTs using the same inclusion and exclusion criteria as the previous study found no survival benefit from vitamin C therapy, with no differences in symptoms, performance status, appetite, or weight loss [21].

## 8 IS RANDOMIZATION FEASIBLE?

The final point to consider in this analysis is how applicable is randomization in providing the best medical care to patients. Earlier in this chapter the technical considerations on why randomization is not always feasible in practice were discussed. These may include poor patient recruitment due to dislike of one the intervention options, for example, the mastectomy versus lumpectomy RCT [6], where patients preferred the less disfiguring operation over a mastectomy. Trials evaluating treatments for chronic illness may also suffer from low recruitment rates and high dropout rates, due to failure to comply with less effective medication or medication with greater side effects. It is possible to minimize these issues using the randomized discontinuation design [22]. The purpose behind these study designs is to select a subset of enrolled patients who are relatively homogenous with respect to their prognostic factors and randomize only this subgroup to treatment or intervention. These patients may be selected by their ability to tolerate medication without excessive side effects or their ability to comply with treatment regimens.

In chemotherapy trials, it is common for patients to deviate from their randomized treatment during follow-up. This is often due to worsening of their disease, which requires rescue medication, either in addition to their trial medication or in place of it. In such cases, the additional medication plays a major confounding factor since the rescue medication will doubtless contribute to outcome differences

between the intervention and control arms. In scientific and statistical terms, the best approach to this problem is to avoid making any changes to medication once the patient has been enrolled and randomized, however, this would be unethical. Excluding patients from the trial outcome analysis would also mask a proportion of patients who may not be responding to the trial therapy and hence may give a false impression of the true efficacy of the drug under evaluation. In addition, any synergistic effect of trial drug plus rescue medication combination would be lost.

Finally, there are ethical issues in the process of randomization of patients in clinical trials. In the Declaration of Helsinki it states that a subject may be enrolled in an RCT if there is genuine uncertainty (equipoise) about which treatment (standard or trial treatment) would benefit the patient [23]. New drug treatments undergo extensive preclinical, phase II and phase III testing before entering RCTs. This laborious evaluation process occurs as a result of huge expenditure on the part of pharmaceutical companies, who would only pursue such clinical testing if there was clear and dramatic evidence of the trial drug offering significant benefits over standard treatment. Doubtless, the principal investigator will be familiar with the evidence. In light of these observations, how much uncertainty is there that the new trial drug is no better than the control drug that patients will receive? The counterargument to this observation concerns whether it is ethical to leave a patient's treatment to chance, or should he or she receive the best treatment available according to the knowledge of the day?

Another important ethical issue when considering RCTs is what to do if the trialled treatment is significantly worse or better that the standard treatment for a particular condition, during the trial period. There is provision for calling RCTs to a halt in either circumstance, but at what point would it be ethical to intervene? The azidothymidine (AZT) study was a double-blind randomized controlled study evaluating the efficacy of AZT in patients with AIDS (acquired immune deficiency syndrome) [24]. The study was terminated early before full recruitment of patients had occurred, when 19 patients receiving placebo had died and only one patient receiving AZT. The study confirmed the efficacy of AZT in the management of AIDS, but 19 patients who had been randomized to the study were unable to benefit from the new treatment.

Ultimately, the physician has a duty to provide the optimum care for his or her patient, but what if this means breaking the study protocol? The first RCT in Britain in the 1940s was assessing streptomycin in the treatment of tuberculosis. During the conduct of the trial, one of the senior research council members contracted tuberculosis and received supplies of streptomycin outside of the trial. This example highlights the fact that clinical judgment can be at conflict with the rigid protocols of RCTs.

The ethical issues touched upon do not, however, preclude the use of RCTs or nullify their usefulness in modern medicine. Sensitivity to these issues is important to the principle investigator in these trials and can be used to optimize study design. With regard to the subject of equipoise, there are numerous examples of RCTs where the trialled drug treatment has been found to be ineffective. In such circumstances, had it not been for the unbiased data collection afforded by RCTs, the ineffectual experimental treatment would be that chosen by many clinicians, having been seduced by the wealth of preclinical data supporting the efficacy of the drug. There is a subtle distinction between theoretical equipoise, as defined by the

Helsinki declaration, and clinical equipoise. Clinical equipoise is best defined as a lack of consensus among the medical community over which treatment is the most effective. Hence, although a new drug treatment is theoretically more efficacious (supported by preclinical and phase I/II data), it is still not actually proven to be so.

Most RCTs compare new treatments with standard treatments and hence are powered to pick a 5–10% improvement in efficacy. For this reason, examples such as the AZT treatment in AIDS, where a new effective treatment was compared with the standard (i.e., no treatment or placebo), are rare. In such circumstances, a well-designed and monitored RCT can be stopped and the efficacious new drug offered to all trial participants and the disease population. The AZT trial is a good example of an RCT that was well monitored by a central oversight group that intervened suitably when it became clear that the AZT treatment should be available to all participants. Although an RCT may not be appropriate or applicable to all clinical scenarios, where possible it can provide unbiased assessment of new drugs or treatments.

The technical details of randomization are discussed elsewhere in this book. In this chapter the phenomenon of "code breaking" (deciphering the allocation sequence) will be briefly touched upon. Investigators rarely document code breaking, but in anonymous questionnaires, there have been investigators who admit to sabotaging randomization codes [25]. Awareness that investigators may actively attempt to unravel the allocation sequence is important in the design of clinical trials, and hence robust steps must be taken to ensure that this not occur. RCTs may frequently alienate the clinicians involved by frustrating their clinical inclinations, and allocation deciphering is an attempt to provide the perceived best care for their patient. It is unlikely that investigators are attempting to deliberately sabotage the results of the trial. As we have seen, randomization, when possible, offers key advantages and credibility to the results of a clinical trial. It is essential then that the word *randomization* is not used in vain, and concealment of allocation codes is essential in the planning of any RCT.

## 9 CONCLUSION

Randomized controlled trials are the best statistical method of evaluating one treatment over another. In general, a well-conducted and monitored RCT provides an unbiased, balanced, and reliable method of determining the efficacy of new drugs or devices. Figure 7 is a schematic illustration of the level of evidence provided by different trial methodology. RCTs also carry with them important logistic and ethical issues, which must be thoroughly addressed before commencing the trial. A review of phase III RCTs evaluating chemotherapy agents observed that the methods used were satisfactory in 67.5% of cases, but 30% of authors were not respecting fundamental ethical principles [26]. In certain circumstances, RCTs may not be possible, in which case recourse to alternative trial designs may be necessary, but this must be with realization that the level of evidence provided will be diminished.

Green and Raley reported that a well-designed clinical trial should incorporate a proper randomization procedure allocating patients to treatment groups in an unbiased manner with a sample size large enough to provide adequate statistical

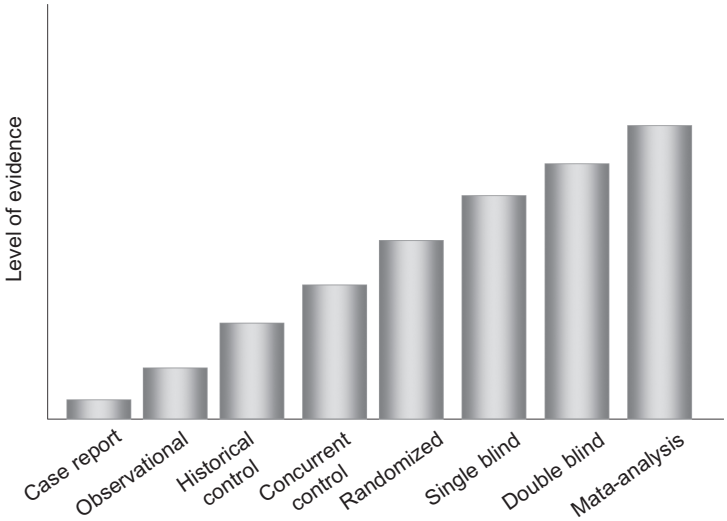


***FIGURE 7*** *Schematic representation of the tiers of evidence provided by different trial methodology. (Reproduced with permission from Lyman and Kudere [28].)*

power [27]. In addition simple endpoints are required and an intention-to-treat analysis that includes all individuals randomized to the treatment groups regardless of whether or not they completed treatment. Results should be presented clearly with a confidence interval that quantifies uncertainty. A final point to consider is the quality of trial results reporting. Well-reported and designed controlled trials can have their results combined to produce a more precise estimate of the effects of treatment.

## REFERENCES

1. Russell, P. S. (1989), *Theoret. Surg.*, 4, 169–170.
2. Haines, S. J. (1979), *J. Neurosurg.*, 4, 163–170.
3. Oettinger, W., and Berger, H. G. (1989), *Theoret. Surg.*, 4, 170.
4. Pollack, A. V. (1989), *Theoret. Surg.*, 4, 163–170.
5. Solomon, M. J., and McLeod, R. S. (1995), *Surgery*, 118, 459–467.
6. Taylor, K. M., Mangolese, R. G., and Soskolne, C. L. (1984), *N. Engl. J. Med.*, 310, 1363–1367.
7. Mant, D. (1999), *Lancet*, 353, 743–746.
8. Charlson, M. E. (1984), *BMJ*, 289, 1281–1284.
9. Diehl, L. F., and Perry, D. J. (1986), *J. Clin. Oncol.*, 4, 114–1120.
10. Gehan, E. A., and Feicrich, E. J. (1974), *N. Engl. J. Med.*, 290, 198–203.
11. Cranberg, L. (1979), *BMJ*, 2, 1265–1266.
12. Gehan, E. A. (1978), *Biomedicine*, 28, 13–19.
13. Chalmers, T. C. (1977), *N. Engl. J. Med.*, 297, 285.
14. Abel, U., and Koch, A. (1999), *J. Clin. Epidemiol.*, 52, 487–497.
15. Ottenbacher, K. (1991), *Controlled Clin. Trial*, 13, 50–61.

16. Pak, C. Y., Adams-Huet, B., Sakhaee, K., et al. (1996), *J. Bone Mineral Res.*, 11, 160–168.
17. Miller, J. N., Colditz, G. A., and Mosteller, F. (1989), *Stat. Med.*, 8, 455–454.
18. Colditz, G. A., Miller, J. N., and Mosteller, F. (1989), *Stat. Med.*, 4, 441–454.
19. Chalmers, T. C., Celano, P., Sacks, H. S., et al. (1993), *N. Engl. J. Med.*, 309, 1358–1361.
20. Cameron, E., and Pauling, L. (1976), *Proc. Natl. Acad. Sci. U.S.A.*, 73, 3685–3689.
21. Creagan, E. T., Moertel, C. G., O'Fallon, J. R., et al. (1979), *N. Engl. J. Med.*, 27, 687–690.
22. Kopec, J. A., Abrahamowicz, M., and Esdaile, J. M. (1993), *J. Clin. Epidemiol.*, 46, 959–971.
23. Edwards, S. J. L., Lilford, R. J., Braunholtz, D. A., et al. (1998), *Health Technol. Assessment*, 2, 1–32.
24. Fischl, M. A., Richman, D. D., and Grieco, M. H. (1987), *N. Engl. J. Med.*, 387, 185–191.
25. Schulz, K. F. (1995), *JAMA*, 274, 1456–1458.
26. Tuech, J. J., Pessaux, P., Moutel, G., et al. (2005), *J. Med. Ethics*, 31, 251–255.
27. Green, S. B., and Raley, P. L. (2000), *Sci. Ed.*, 23, 157.
28. Lyman, G. H., and Kudere, N. M. (1997), *Cancer Control*, 4, 413–418.

# *Cross-Over Designs*

**Raphaël Porcher and Sylvie Chevret**

*Département de Biostatistique et Informatique Médicale, Hôpital Saint-Louis, France*

**Contents**

1 Introduction
2 Main Issues
  2.1 Advantages
  2.2 Disadvantages
  2.3 Should Cross-Over Trials Be Used and When?
3 Analysis of AB/BA Design
  3.1 Basic Analysis
  3.2 Adjustment for Period Effect
  3.3 Model-Based Analysis
  3.4 Two-Stage Analysis
  3.5 Carry-Over
  3.6 Use of Baselines
  3.7 Bayesian Analysis
4 Other Endpoints
  4.1 Nonparametric Analyses
  4.2 Binary or Categorical Data
  4.3 Censored Endpoints
5 Power and Sample Size Considerations
  5.1 Sample Size
  5.2 Efficiency of Cross-Over versus Parallel-Group Design
6 More Complicated Designs
  6.1 Higher Order Designs for Two Treatments
  6.2 Designs for Three or More Treatments
7 Summary
  References

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

## 1 INTRODUCTION

Cross-over designs refer to clinical trials where patients sequentially receive several of the compared treatments, instead of receiving only one of them as in the classical parallel-arm designs. Each subject hence acts as his (her) own control. The main purpose of such a design is thus to perform within-subject comparison of treatment effects instead of between-subject comparisons, which increases efficiency. As a consequence, a cross-over trial usually requires less patients than a parallel-arm design.

The simplest cross-over design is the two-treatment two-period cross-over design, also referred to as the AB/BA design, since patients receive either treatment A followed by treatment B or treatment B followed by treatment A. A schematic representation of the AB/BA design is given Figure 1. As depicted in Figure 1, patients included in an AB/BA cross-over trial are randomized between each *group* (or sequence) A → B or B → A. After a potential *run-in* period where subjects are given neither A nor B, the *first period* (period 1 on the figure) begins with the administration of one of the two treatments, according to the randomization group. During the run-in period, baseline measurements can be performed. Between the two treatment periods, a *wash-out* period is generally used to remove any residual effects of the treatment received in the first period. As for the run-in period, neither A nor B is administered during the wash-out period. Then, patients switch to the treatment they did not receive yet.

Provided the disease condition of subjects is similar at the beginning of each treatment period, the cross-over design is intuitively appealing to compare treatments. Indeed, each patient will serve as his (her) own control. However, the effect of the first treatment received may persist during the second treatment period, leading to the so-called *carry-over* effect of the treatment. Several considerations on both the use of two-stage designs to cope with carry-over and the low power of the test for carry-over have generated a large literature on cross-over designs [1–6] and somewhat contradictory guidelines from regulatory agencies [7, 8]. More recently, three textbooks dedicated to cross-over trials have been published, where details and discussion regarding cross-over trials may be found [9–11].

In this chapter, we will review the main issues concerning the use and analysis of cross-over designs. We will focus on the AB/BA design and more particularly on the analysis of such trials. More complex design will be mentioned later. Another type of cross-over design can be frequently encountered, for example, in malignant diseases. It is basically a parallel-arm design where patients who fail to respond to the

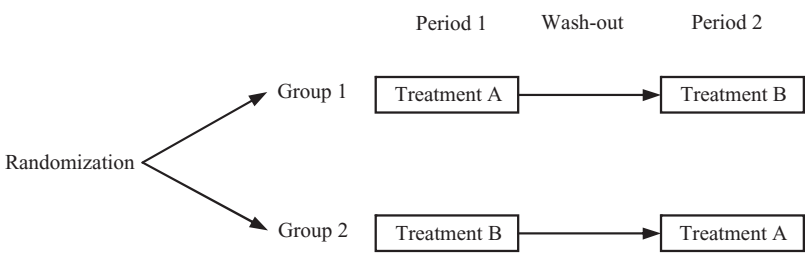


***FIGURE 1*** *Representation of an AB/BA cross-over trial.*

first treatment they were assigned to are switched to the other one. These treatment cross-overs are planned for ethical reasons in order to give each patient a chance to receive the "superior" treatment, and the main analysis relies on comparing the initial treatments only. This type of cross-over design will thus not be considered in this chapter.

## 2 MAIN ISSUES

### 2.1 Advantages

The cross-over design is intuitively well suited to compare treatments as all, or at least some, are administered to each subject. Nevertheless, the main advantage of cross-over trials is that inference on treatment effect is based on within-subject information, each subject being used as his (her) own control. Between-subject variability is eliminated, which leads to increased efficiency. In general, fewer patients, and sometimes much fewer, will be required to obtain the same precision in estimation than classical parallel designs. Cross-over trials thus may save considerable resources.

Another advantage of cross-over trials is their acceptability. As all subject know they will receive the superior treatment at some point, their willingness to participate may increase, as well as their compliance.

### 2.2 Disadvantages

The principal disadvantages of cross-over trials stem from carry-over. Carry-over is defined as the persistence of the effect of the treatment received in a period into the subsequent periods. In the presence of carry-over, the outcome measured after the second treatment will be also affected by the first treatment, which is likely to bias the estimate of the treatment effect. For example, a drug agent may physiologically persist during the next period. The first treatment could also change permanently the condition of the patient, for example, by curing him (her). If the first case can be corrected by using an appropriate wash-out period between treatment periods, the second will preclude using a cross-over design. This is the case of acute infections, for instance. Carry-over has focused most concerns toward cross-over trials, and some approaches have dealt with removing carry-over, or at least accounting for it, as will be seen below. The best approach is, however, to design the study in order to avoid carry-over. Another problem related to carry-over is treatment-by-period interaction, that is, when the treatment effect is not the same in all treatment periods. As will be seen in the following, treatment-by-period interaction cannot be separated from carry-over in the AB/BA design, but both can be estimated separately in more complex designs.

Dropouts can also be more problematic in cross-over trials than in parallel-group trials, as patients who do not participate to the second study period do not provide any information on the treatment effect. This is the case for patients who die during the study.

Cross-over trials can also be used in particular situations of stable diseases, where the patient condition will not be permanently modified by the treatment.

Other less important disadvantages of cross-over trials are that all patients will receive the inferior treatment (at some point). Moreover, the duration of participation to cross-over trials is longer, due to the several period design particularly for high-order cross-over designs. This can cause some inconvenience to patients and decrease their willingness to participate.

At last, it should be underlined that cross-over trials are more complex to analyze than parallel-group designs.

### 2.3 Should Cross-Over Trials Be Used and When?

As already seen, cross-over trials are obviously not adapted to all situations. As each patient receives each treatment sequentially, a cross-over trial cannot be reasonably considered in diseases where patients have a nonnegligible probability of dying or considerably deteriorating during the trial period. Cross-over trials can also not be used if one of the treatments is supposed to have a permanent, for instance, curative, effect. Moreover, carry-over leads to biased estimates of treatment effect. Therefore, the Biometric and Epidemiology Methodology Advisory Committee (BEMAC) of the Food and Drug Administration (FDA) recommended during the 1970s that cross-over trials should be avoided as much as possible. One should prefer completely randomized designs where the estimate of treatment effect is unbiased, without any modeling assumption [7].

Since then, considerable research has been published on the design and analysis of cross-over trials. In contrast to this recommendation, several authors have published in defense of cross-over trials, at least in certain cases [5]. Actually, cross-over trials may be very useful but require more careful planning than parallel-group trials, as it is important to avoid carry-over [8, 11]. They should be used only in the appropriate disease settings and when dropouts are also expected to be small. For instance, it is well suited for chronic diseases such as hypertension or chronic renal failure. When planning a cross-over trial, everything should be done to avoid carry-over. This will not arise from a particular modeling strategy but from design only. It is therefore of primary importance to plan sufficient wash-out periods [11]. Additionally, random and balanced allocation to each group is advisable, with adjustment for period effect in the analysis. At last, baseline measurements may be used to increase precision. A satisfactory use of cross-over trials can also be found in demonstrating the bioequivalence of two formulations of the same drug. This is particularly the case in trials involving healthy volunteers, where carry-over is likely to be properly eliminated using an adequate wash-out period.

## 3 ANALYSIS OF AB/BA DESIGN

Even for the apparently simple AB/BA design, analysis of cross-over designs has generated a wide literature in the last three decades. The main reason has to do with carry-over. This section introduces the most common methods used for analysis of cross-over designs and sketches the implications of the use of these methods. More details may be found in dedicated monographs [9, 11] or in the issue of *Statistical Methods in Medical Research* dedicated to cross-over trials (1994, volume 4, issue 3).

To simplify the notations and the comparison between the different methods reviewed here, we propose to use the general model formulation for the cross-over design with a continuous outcome. Such a model has been very commonly used in the cross-over literature [1, 12–14]. Let us then note $Y_{ijk}$ as the outcome measure of the $j$th subject within the $i$th treatment group (or sequence) at the $k$th period. For the AB/BA design, $i = 1,2$; $j = 1, \ldots, n_i$; $k = 1, 2$. The model for $Y_{ijk}$ we consider is

$$Y_{ijk} = \mu + \pi_k + \tau_l + \eta_k \lambda_i + s_{ij} + \varepsilon_{ijk} \tag{1}$$

where $l = 1$ if $i = k$ and 2 if $i \neq k$, and $\eta_k = 0$ if $k = 1$ and 1 if $k = 2$.

In this model, $\mu$ in the overall mean, $\pi_k$ is the period effect, $\tau_l$ the treatment effect and $\lambda_i$ the first-order carry-over effect. The subject effect is represented by $s_{ij}$ and may be regarded as a fixed or a random effect. We will adopt the latter formulation and consider that the $s_{ij}$ are normally and independently distributed with mean 0 and variance $\sigma_s^2$. The within-patient errors $\varepsilon_{ijk}$ are also assumed to be normally distributed, with mean 0 and variance $\sigma^2$, and both independent from each other and from the $s_{ij}$. We will note $\rho$, the correlation between observations on the same patient, which is $\rho = \sigma_s^2 / (\sigma_s^2 + \sigma^2)$. In this model, the whole variance of observations is $\sigma_T^2 = \sigma_s^2 + \sigma^2$, which can also be rewritten as $\sigma^2/(1 - \rho)$. Table 1 presents the expected cell means for model (1). This model explicitly assumes that the only effects are due to treatments or periods and that they do not vary from a patient to another. It also assumes that the total variability can be split into a between-subject and a within-subject variability, that are both equal for all patients. It is thus clearly not the only possible model for the data. Freeman presents the same model with baseline measurements corresponding to a third period noted $k = 0$ [14].

Let us note the treatment difference, or treatment effect, $\tau = \tau_2 - \tau_1$, the carry-over difference $\lambda = \lambda_2 - \lambda_1$, thereafter referred as carry-over for simplicity, and the period effect between periods 2 and 1, $\pi = \pi_2 - \pi_1$. Let $y_{ijk}$ denote the observation of the random variable $Y_{ijk}$ thereafter.

### 3.1 Basic Analysis

The basic analysis of the AB/BA cross-over design overlooks any period or carry-over effect. As each patient in a cross-over trial acts as its own control, the natural idea is to use within-patient differences. The basic analysis consists in computing for each patient the difference in outcome between both treatments, which is called the cross-over difference. Suppose that we use treatment A as a control for B, the cross-over difference will be $Y_{1j2} - Y_{1j1}$ for a patient of group 1 and $Y_{2j1} - Y_{2j2}$ for a patient

**TABLE 1 Expected Cell Means for the Two-Period Cross-Over Model**

| | | Period | |
|---|---|---|---|
| Randomization Group | | First Treatment (1) | Second Treatment (2) |
| 1 | Treatment | A | B |
| | Expected mean | $\mu + \pi_1 + \tau_1$ | $\mu + \pi_2 + \tau_2 + \lambda_1$ |
| 2 | Treatment | B | A |
| | Expected mean | $\mu + \pi_1 + \tau_2$ | $\mu + \pi_2 + \tau_1 + \lambda_2$ |

of group 2, as treatment B is given in the second period in group 1 and in the first period in group 2. Once these cross-over differences are calculated, the analysis is carried-out ignoring the distinction between both groups. This is equivalent to comparing the outcomes obtained with treatment A to the outcomes obtained with treatment B using an appropriate paired test statistic, irrespective of the randomization group. The parametric approach relies on the matched paired *t* test, while the nonparametric approach is based on Wilcoxon's signed rank test. This is illustrated on the example displayed in Table 2. These data have been used by Grieve [15] and come from a study carried out by CIBA-GEIGY to assess the effectiveness of transdermal nitroglycerine patches in the prophylaxis of angina in general practice [16]. A total of 63 patients were randomized between either 3-week treatment with placebo (PL) followed by transdermal nitroglycerine (TN) (group 1, PL → TN) or vice-versa (group 2, TN → PL). The main endpoint was the weekly anginal attack rate during the third week of treatment.

**TABLE 2 Weekly Anginal Attack Rates in Transdermal Nitroglycerine Trial**

| Group 1 | | | | Group 2 | | | |
|---|---|---|---|---|---|---|---|
| Patient | Period 1 | Period 2 | Cross-Over Difference | Patient | Period 1 | Period 2 | Cross-Over Difference |
| 19 | 3 | 10 | 7 | 20 | 12 | 16 | –4 |
| 22 | 8 | 6 | –2 | 21 | 4 | 11 | –7 |
| 26 | 6 | 4 | –2 | 23 | 6 | 5 | 1 |
| 35 | 1 | 0 | –1 | 36 | 7 | 14 | –7 |
| 38 | 12 | 6 | –6 | 37 | 13 | 25 | –12 |
| 39 | 4 | 2 | –2 | 40 | 9 | 11 | –2 |
| 42 | 6 | 3 | –3 | 41 | 1 | 1 | 0 |
| 59 | 11 | 3 | –8 | 43 | 4 | 0 | 4 |
| 64 | 3 | 4 | 1 | 56 | 4 | 10 | –6 |
| 73 | 11 | 3 | –8 | 57 | 2 | 5 | –3 |
| 76 | 8 | 8 | 0 | 60 | 0 | 8 | –8 |
| 78 | 8 | 9 | 1 | 61 | 17 | 13 | 4 |
| 80 | 18 | 4 | –14 | 65 | 1 | 1 | 0 |
| 81 | 12 | 5 | –7 | 67 | 6 | 8 | –2 |
| 84 | 12 | 2 | –10 | 75 | 8 | 8 | 0 |
| 85 | 3 | 1 | –2 | 77 | 7 | 4 | 3 |
| 115 | 1 | 3 | 2 | 79 | 3 | 19 | –16 |
| 122 | 12 | 4 | –8 | 82 | 4 | 19 | –15 |
| 124 | 8 | 6 | –2 | 83 | 3 | 12 | –9 |
| 126 | 7 | 12 | 5 | 86 | 2 | 4 | –2 |
| 128 | 1 | 1 | 0 | 87 | 2 | 1 | 1 |
| 140 | 2 | 0 | –2 | 121 | 4 | 7 | –3 |
| 142 | 3 | 0 | –3 | 123 | 3 | 1 | 2 |
| 146 | 21 | 10 | –11 | 125 | 3 | 3 | 0 |
| 147 | 17 | 7 | –10 | 127 | 1 | 0 | 1 |
| 150 | 12 | 5 | –7 | 130 | 41 | 36 | 5 |
| 201 | 4 | 5 | 1 | 145 | 10 | 24 | –14 |
| 209 | 0 | 1 | 1 | 148 | 9 | 18 | –9 |
| 211 | 7 | 0 | –7 | 149 | 4 | 13 | –9 |
| 233 | 11 | 0 | –11 | 210 | 8 | 1 | 7 |
| 236 | 18 | 7 | –11 | 234 | 5 | 7 | –2 |
| | | | | 235 | 0 | 9 | –9 |

In the example, the cross-over differences are displayed for each patient. Their mean would produce the basic estimator of the treatment effect. Numerically, we obtain $\hat{\tau}^b = -3.651$, with standard error 0.70. Using a $t$ test, the observed $t$ statistics is –5.188 with 62 degrees of freedom, leading to a $P$ value of $2.5 \times 10^{-6}$ (for a two-sided test). The 95% confidence interval (CI) of the treatment effect is –5.058 to –2.244.

This is a simple way to analyze cross-over trials. It is, however, only valid under particular conditions. Actually, from model (1), the cross-over differences will have expectation $\tau + \lambda_1 + \pi$ in group 1 and $\tau - \lambda_2 - \pi$ in group 2. With $n_1$ patients in group 1 and $n_2$ patients in group 2, the basic estimator has expectation $\tau + \pi(n_1 - n_2)/(n_1 + n_2) + (n_1\lambda_1 - n_2\lambda_2)/(n_1 + n_2)$. It is unbiased when there is no period effect, that is, no global time trend and no carry-over. However, for an unbalanced design, a period effect is sufficient to produce a biased estimator. Moreover, and even more seriously, a period effect would also result in an inflated variance of the estimator, as deterministic variations would have been attributed to random variation. Senn lists three other factors affecting the value of this basic analysis, namely a treatment by period interaction, patient by treatment interaction, and patient by period interaction [11]. Of these factors, carry-over and period by treatment interactions have been regarded as the major problems with cross-over trials.

### 3.2 Adjustment for Period Effect

Hills and Armitage [4] used a stratified analysis to adjust for a period effect. They still use within-patient differences but take the unweighted average of the estimator of the treatment effect obtained in each group. This defines a new estimator of the treatment effect, $\hat{\tau}^c = \{(\bar{Y}_{1.2} - \bar{Y}_{1.1}) + (\bar{Y}_{2.1} - \bar{Y}_{2.2})\}/2$, where the subscript "·" indicates averaging with respect to the corresponding index. This method thus combines some of the cell means of Table 1. Under the assumptions of model (1), this estimator has variance $\sigma^2(1/n_1 + 1/n_2)/2$, where $\sigma^2$ is the within-subject variance. For a balanced trial where $n_1 = n_2 = n$, this variance reduces to $\sigma^2/n$. In such a case, the estimate is equal to the basic estimate, but with a different estimate of the variance. Actually, the within-subject variance $\sigma^2$ is half the variance of a cross-over difference, which may be estimated by pooling the two within-groups sum of squares and dividing the sum by the appropriate $n_1 + n_2 - 2$ degrees of freedom. This allows to calculate a $t$ statistic, which also has $n_1 + n_2 - 2$ degrees of freedom. This can be illustrated on the preceding example, with useful cell means presented Table 3.

For the two randomization groups, we obtain means equal to –3.839 and –3.469, respectively, and estimated variances $S_1 = 26.740$ and $S_2 = 36.451$. The estimate of the treatment effect is thus simply $\hat{\tau}^c = -3.654$, and the pooled variance estimate

**TABLE 3 Cell Means for Transdermal Nitroglycerine Trial Data**

| Randomization Group | | Period 1 | Period 2 | Difference TN – PL |
|---|---|---|---|---|
| 1 | Treatment | PL | TN | |
| | Cell mean | 8.065 | 4.226 | –3.839 |
| 2 | Treatment | TN | PL | |
| | Cell mean | 6.344 | 9.813 | –3.469 |

$$2\hat{\sigma}^2 = \frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2} = 31.675$$

Hence, the within-subject variance is $\hat{\sigma}^2 = 15.837$, the estimated standard error of $\hat{\tau}^c$ is 0.709, and the resulting $t$ statistic is –5.152 with 61 degrees of freedom. The corresponding $P$ value is $2.9 \times 10^{-6}$. These results are very close to those obtained with the basic estimator.

This estimator is the one recommended by Senn [11]. It also corresponds to the standard weighted least-squares estimate of the treatment effect in model (1) [17]. Freeman referred to it as the CROS estimator [14], and the term was subsequently used in several reviews [11, 18]. We will thus use this terminology thereafter.

### 3.3 Model-Based Analysis

The alternative is to fit directly model (1) to the data. Note that, up to now, we did not consider baselines. Whether to allow separate variation for groups and patients within groups or not does not modify the estimate of treatment effect but allows to estimate a carry-over effect. The detailed analysis of variance (ANOVA) table for model (1) with fixed group, period, and treatment effects and random patient effect nested in group is not straightforward and is not detailed here. It can be found in the textbook by Jones and Kenward [9].

Numerical values for the ANOVA table together with the fixed effects estimate are presented in Table 4 for the weekly angina attack data. Using this model, the treatment effect estimator is identical to the CROS estimate presented above. The model also allows to test for carry-over or period effects, although the latter is generally of minor interest in a cross-over trial. We will detail the test of carry-over effect later. Results suggest a strong evidence for a reduction of angina attacks, about three or four per week. More precisely, the 95% confidence interval of the treatment effect is estimated at –5.072 to –2.235. Moreover, there seems to be little evidence of a carry-over (group) effect nor a period effect. For purpose of comparison, note that model (1) uses a different parameterization than that used in Grieve [15], the effects estimated being minus twice the ones he chose to parameterize the model. Also note that the group effect is half the carry-over effect $\lambda$. The results of Table 4 perfectly match those presented by Grieve [15].

Another parameterization leading to exactly the same model would be to replace the group effect by a treatment-by-period interaction. The model used is obviously unable to attribute the difference of treatment effect in the second period to a

**TABLE 4 Simplified ANOVA Table and Parameter Estimates for Transdermal Nitroglycerin Trial**[a]

| Source | df | *F* Value | *P* Value | Parameter Estimate | SE |
|---|---|---|---|---|---|
| Group (2 vs. 1) | 1, 61 | 1.693 | 0.198 | 1.933 | 1.486 |
| Period (2 vs. 1) | 1, 61 | 0.068 | 0.795 | –0.185 | 0.709 |
| Treatment (TN vs. PL) | 1, 61 | 26.545 | $2.9 \times 10^{-6}$ | –3.654 | 0.709 |

[a]df = degrees of freedom; SE = standard ends.

**TABLE 5 Distribution Parameters for Contrasts in Model (1)**

| Contrast | Mean | Variance |
|---|---|---|
| $\hat{\tau}^c$ | $\tau - \frac{\lambda}{2}$ | $\frac{m\sigma^2}{2}$ |
| $\hat{\lambda}$ | $\lambda$ | $2m\sigma^2\left(\frac{1+\rho}{1-\rho}\right)$ |
| $\hat{\tau}^p$ | $\tau$ | $m\sigma^2\left(\frac{1}{1-\rho}\right)$ |

carry-over [as parameterized in model (1) formulation] or to an interaction between treatment and period. Both can thus not be separated using the observed data.

Let us consider two additional contrasts:

$$\hat{\lambda} = \left(\bar{Y}_{2.1} + \bar{Y}_{2.2}\right) - \left(\bar{Y}_{1.1} + \bar{Y}_{1.2}\right) \quad \text{and} \quad \hat{\tau}^p = \bar{Y}_{2.1} - \bar{Y}_{1.1}$$

where $\hat{\lambda}$ is the difference between the two groups, and has expectation $\lambda$, of which it is the weighted least-squares estimate. The contrast $\hat{\tau}^p$ is simply the between-patient contrast for treatment effect using the first period data only. They are linked to the CROS estimator by $\hat{\tau}^p = \hat{\lambda}/2 + \hat{\tau}^c$. These three contrasts have a Gaussian distribution, with respective mean and variance given in Table 5, where $m$ denotes $(n_1 + n_2)/(n_1 - n_2)$.

It appears in Table 5 that the CROS estimator is biased in the presence of carry-over, except if the carry-over is the same after each treatment ($\lambda_1 = \lambda_2$), which does not seem much more plausible than the absence of any carry-over in practice. An unbiased estimator of $\tau$ would be obtained using $\hat{\tau}^p$ instead, but this would imply discarding all data recorded during the second period. This analysis has been referred to as PAR (for parallel, as no cross-over data are used). Consequently, the variance of such an estimator would be at least twice that of the CROS estimator. The contrast $\hat{\lambda}$ provides an estimate and test for the carry-over effect. Given the variance of $\hat{\lambda}$ as compared to that of the test of treatment effect (based either on CROS or PAR), it is clear that the test for carry-over will not be very powerful. This is, however, the basis for the general two-stage strategy proposed by Grizzle [12].

### 3.4 Two-Stage Analysis

Initially proposed by Grizzle [12], a two-stage approach has been recommended for many years and was commonly used for analysis. This approach begins with a formal test for carry-over, usually at a 10% level, given the limited power of this test. If no evidence of a carry-over effect is brought by this test, the treatment effect is estimated from within-patient differences using the CROS estimator. However, if the test for carry-over is considered as significant, a between-patient estimator is used instead. This estimator is the one referred as PAR above and only used the first period of the trial. Since it cannot be known in advance which test will be used,

some authors have advised the trial should always be adequately powered for the PAR analysis [1]. At first sight, this design was considered ingenious and regarded as the method of choice. Freeman conducted a thorough analysis of the two-stage design as a whole and showed that it had highly inflated type I error rates, mainly because the test for carry-over and the PAR test are not independent [14]. While some work was more recently published in defense of this procedure [5], main monographs on cross-over trials do not recommend it anymore [9–11].

For the weekly anginal attack rates of the transdermal nitroglycerin data, no significant evidence of carry-over was found at the 0.10 level. We would thus simply use the CROS analysis and conclude at a significant reduction of the weekly anginal attack rate with nitrogylcerin. The estimated rate reduction would be 3.654 per week (95%CI 2.235–5.072). For comparison, if a more liberal level was used for the test of carry-over, say 0.30 for instance, the PAR analysis would be preferred. The estimated anginal attack reduction would be 1.721 per week (95%CI −5.056 to +1.615), and we would have concluded to a nonsignificant anginal attack rate reduction ($P = 0.306$). Although the CROS analysis yields here a treatment effect estimate very close to the basic estimate, this two-stage procedure leads to a clearly different conclusion.

Freedman showed that if CROS behaved badly for carry-over effects departing from zero, substantial carry-over making it almost certain that spurious treatment differences would be reported [14]. The two-stage procedure had possibly highly inflated type I error rates, even in the case of absence of carry-over. For example, if $\rho = 0.6$, and using a 0.1 level for the test of carry-over effect and a 0.05 level for the second stage test, the actual type I error rates of the two-stage procedure would range from 0.085 in case of no carry-over effect up to 0.5 for increased carry-over effect (though it has been argued that high carry-over difference values may be quite unrealistic in case of no treatment effect). Nevertheless, the two-stage procedure always has increased significance level. On the contrary, in case of no cross-over, the two-stage procedure shows a better power than PAR, but purchased at the expense of increased type I error rates. These properties are explained by the joint distribution of the three contrasts $(\hat{\tau}^c, \hat{\lambda}, \hat{\tau}^p)$. If $\hat{\lambda}$ and $\hat{\tau}^c$ are independent, the correlation between $\hat{\lambda}$ and $\hat{\tau}^p$ is $\sqrt{(1+\rho)/2}$, which can be remarkably high, for example, about 0.866 if the within-patient variability $\sigma^2$ is equal to the between-patient variability $\sigma_s^2$. The tests based on PAR and the test for carry-over are thus not independent. Actually, Senn showed that the conditional error of PAR given a significant test for carry-over at the 0.1 level lies between 0.25 and 0.5, depending on the balance between within-subject and between-subject variability [19]. On the contrary, the type I error rate of the CROS analysis is 0.05 when the test for carry-over is nonsignificant. The global type I error of the whole procedure thus lies between $0.9 \times 0.05 + 0.1 \times 0.25 = 0.07$ and $0.9 \times 0.05 + 0.1 \times 0.5 = 0.095$.

Another particular property of the two-stage design pointed out by Freeman is that its performance does not increase when the sample size increases [14]. This makes worthless the proposal of some authors that the trial should be adequately powered for the carry-over test (see, e.g., Brown [1]). Actually, as results of the first stage test are more likely to yield reliable conclusions, results of the second stage will be all the more sensitive to small carry-over values. The overall performance thus worsens as sample size increases, with inflated type I error rates. For example, with $\rho = 0.6$ and a 0.1 level for the test of carry-over effect and a 0.05 level for the

second stage test, an actual type I error rate of 0.5 occurs at $\lambda = 1.25\sigma$ when $n$ ($= n_1 = n_2$) is about 20, $\lambda = \sigma$ when $n$ is about 30 and $\lambda = 0.55\sigma$ when $n = 100$. At last, it should be noted that if $\hat{\tau}^c$ is a biased estimator of $\tau$, with bias depending on carry-over $\lambda$, $\hat{\tau}^p$ even unbiased, is correlated to $\hat{\lambda}$, the estimate of $\lambda$.

Based on these considerations, both Senn [11, 19] and Jones and Kenward [9] recommend not using the two-stage procedure. They advise the best should be done to remove carry-over, using, for instance, adequate wash-out periods. Then, a sensible analysis can be performed by CROS. Freeman would recommend the use of a Bayesian approach [15, 20, 21].

### 3.5 Carry-Over

As seen in Section 3.3, carry-over and treatment-by-period interaction cannot be estimated separately using the observed data. Thus, carry-over or treatment-by-period interaction is rather a modeling point of view [14]. If carry-over cannot be ruled out by design, then it may be preferable to perform a parallel-group trial [11, 14]. On the contrary, if a treatment-by-period interaction is suspected, any form of trial is questionable, as estimating an absolute difference between treatments is meaningless.

Of note, both may also exist in parallel-group trials. For example, in a trial comparing an experimental treatment to a standard treatment of a chronic disease, patients allocated to the experimental group may have received the standard treatment in the past, and carry-over could occur in the same way as in a cross-over trial if a (run-in) wash-out period is not planned. Conversely, one may also imagine that some patients already treated by the standard therapy and allocated to the control group may have developed some form of resistance to this treatment, which would amplify the difference between treatment arms.

Different approaches concerning carry-over may be found in the literature on cross-over trials. For instance, Senn recommends to design the trial at best to avoid carry-over and then to not consider further this issue [6, 11, 22]. Willan and Pater would also likely use a CROS analysis, but in cases where between-subject variability exceeds within-subject variability and the carry-over effect is small [13]. This would be consistent with the point of view of Poloniecki and Pearce, who argue that, since the test for carry-over has a low power, it will allow to detect only large values of carry-over [23].

### 3.6 Use of Baselines

Carry-over has been the focus of much concern in preceding sections. Baseline measurements may be used to add information and eliminate parameters such as carry-over from the contrasts of interest. Baselines may be intended as genuine baseline measurements of the outcome (i.e., before treatment) or more generally as other covariates. Additionally, baselines may be measured at several time points, as during a run-in period, before the first treatment period, at the end of the wash-out period, or even at the end of a second wash-out period after the last treatment period [9–11]. For example, the formulation for model (1) used in Freeman only considers baseline measurements during a run-in period [14]. Such a situation can be treated in practice in the same way as for genuine covariates. In both cases, if

the response is related to the covariate (or baseline measurement), the between-subject residual variance may be sensibly reduced by adjusting for the covariate in the model. This is the same as for parallel group trials. Considering two or more baselines requires a modification of the way data are to be analyzed [9]. Several approaches have been proposed to incorporate baseline measurements into the analysis. They differ with regard to the assumptions that may be made or not [9, 11, 24]:

- Is the period effect applicable to a baseline measurement equal to the one applying to the following outcome measurement?
- Does carry-over also apply to baseline measurements and, if so, is it the same as carry-over for the subsequent treatment?
- Do the observations on the same subject have a uniform covariance structure?

To summarize, it has been shown that incorporating baselines can reduce the variance of the treatment effect estimate, as expected, and Kenward and Jones showed how to consider these baselines without assuming a particular covariance structure for the mesurements [24]. They also showed that if information from baseline could increase the power of the test for carry-over, it remained unlikely that it would be increased up to the power of the test for direct treatment effect. The methods are not detailed here but can be found in appropriate references [9, 24].

### 3.7 Bayesian Analysis

The reader unfamiliar with Bayesian analysis should perhaps read first the introduction to Bayesian methods in health science by Spiegelhalter et al. [25]. A first Bayesian analysis of the $2 \times 2$ cross-over trial as used for bioequivalence assessment was proposed by Selwyn et al. [26] Grieve and co-workers have proposed a Bayesian approach to the analysis of the AB/BA cross-over design, which has been extended to incorporate baseline measurements and to other cross-over designs [15, 20, 21, 27]. According to Freeman, this is the only satisfactory analysis of the two-period two-treatment cross-over design [14]. We here present a brief summary of this approach.

The assumed model is model (1). Grieve considers a joint uninformative prior distribution for the model parameters [20]. Let us note $p(\tau, \lambda, \sigma^2, \rho)$ the prior distribution of model parameters of interest (for simplicity, the period effect is not explicitated here). The prior distribution has the form [28]

$$p(\tau,\lambda,\sigma^2,\rho) \propto \frac{1}{\sigma^2(1+\rho)}$$

Noting $Y$ the observed data, the following posterior distributions are obtained for the location parameters $\tau$, $\lambda$ given the observations and the second-order parameters $\sigma^2$, $\rho$ are obtained from standard properties of the normal distribution as:

$$p(\tau|\lambda,\sigma^2,\rho,Y) = N\left(\hat{\tau}^c + \frac{\lambda}{2}; \frac{m\sigma^2}{2}\right)$$

$$p(\tau|\sigma^2,\rho,Y) = N\left(\hat{\tau}^p; \frac{m\sigma^2}{1-\rho}\right)$$

$$p(\lambda|\sigma^2,\rho,Y) = N\left(\hat{\lambda}; 2m\sigma^2\frac{1+\rho}{1-\rho}\right)$$

$$p(\tau,\lambda/\sigma^2,\rho,Y) = N\begin{pmatrix}\hat{\tau}^p \\ \hat{\lambda}\end{pmatrix}; \frac{m\sigma^2}{1-\rho}\Bigg)\begin{bmatrix}1 & 1+\rho \\ 1+\rho & 2(1+\rho)\end{bmatrix}$$

Unconditional distributions are computed by reparameterizing the model with $\sigma_1^2 = \sigma^2 + 2\sigma_s^2 = \sigma^2(1-\rho)(1+\rho)$ and $\sigma^2$ instead of $\sigma^2$ and $\rho$. They depend on the sum of squares associated with residual error (SSE) and with subject within group (SSS) of the ANOVA table. The following posterior distributions are obtained:

$$p(\lambda|Y) = t\left(\hat{\lambda}; \frac{2m\text{SSS}}{n_1+n_2-2}; n_1+n_2-2\right) \tag{2}$$

$$p(\tau|\lambda,Y) = t\left(\hat{\tau}^c + \frac{\lambda}{2}; \frac{m\text{SSE}}{2(n_1+n_2-2)}; n_1+n_2-2\right) \tag{3}$$

where $t(\mu, \theta, \upsilon)$ denotes the noncentral and scaled $t$ distribution with $\upsilon$ degrees of freedom, location parameter $\mu$ and scale parameter $\theta^{1/2}$. The integral necessary to obtain the posterior density of $\tau$ unconditional on $\lambda$ is intractable, but the following approximation was derived by Grieve:

$$p(\tau|Y) \approx t\left(\hat{\tau}^p; \frac{m\phi}{2\nu}; \nu\right) \tag{4}$$

where

$$\nu = \frac{(\text{SSE}+\text{SSS})^2(n_1+n_2-6)}{\text{SSE}^2+\text{SSS}^2} + 4$$

$$\phi = \frac{(\phi-2)(\text{SSE}+\text{SSS})}{n_1+n_2-4}$$

Grieve shows how to additionally take the constraint $\rho > 0$ into account in this analysis [15, 20]. However, if negative correlations between different measurements on a same patient were expected, a cross-over design would unlikely be conducted. Additionally, the implied modifications only make little difference, so that these refinements are not further considered here.

The results of a Bayesian analysis of the transermal nitroglycerin trial are displayed in Figure 2. As shown in Grieve, the posterior distribution $P(\tau < 0|\lambda = 0, Y)$ and $P(\tau < 0|Y)$ differ markedly [15]. For example, $P(\tau < 0|Y) = 0.85$, whereas $P(\tau < 0|\lambda = 0, Y) > 0.9999$. Note that despite the modified parameterization, these

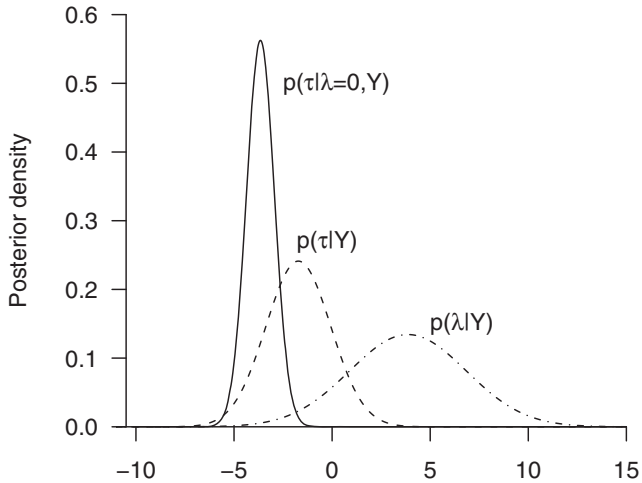


***FIGURE 2*** *Posterior marginal distribution of λ and τ for transdermal nitroglycerin trial data.*

values are identical to Grieve's analyses. Recall that the $P$ value associated to a (frequentist) test of carry-over is 0.198, thus not supporting a model with a carry-over effect in the nitroglycerin trial data. On the contrary, the posterior density of $\lambda$ seems rather in favour of a carry-over, with $P(\lambda > 0|Y) = 0.90$.

Results of the example well illustrate that a model without carry-over, noted $M_0$, and a model with carry-over, $M_1$, can lead to different conclusions. Grieve proposed a Bayesian approach of differentiating between the models $M_0$ and $M_1$, using a prior specification allowing to incorporate information on prior belief on carry-over [20]. The proposed method models the set up as a mixture of $M_0$ and $M_1$, and then uses a Bayes factor. The posterior probability of each of $M_0$ and $M_1$ can be written as

$$P(M_i|Y) = P(Y|M_i)\frac{P(M_i)}{P(Y)} \qquad i = 0,1$$

Noting the Bayes factor

$$B_{01} = \frac{P(M_0|Y)P(M_1)}{P(M_0)P(M_1|Y)} = \frac{P(Y|M_0)}{P(Y|M_1)}$$

and the prior odds on $M_0$, $\kappa = P(M_0)/P(M_1)$, the posterior probability of the two models are

$$P(M_0|Y) = \frac{\kappa B_{01}}{1+\kappa B_{01}} \quad \text{and} \quad P(M_1|Y) = \frac{1}{1+\kappa B_{01}}$$

Inference on the treatment effect can then use the mixture posterior distribution

$$p(\tau|Y) = \frac{\kappa B_{01}}{1+\kappa B_{01}} p(\tau|M_0, Y) + \frac{1}{1+\kappa B_{01}} p(\tau|M_1, Y)$$

Where $p(\tau|M_0, Y)$ and are given $p(\tau|M_1, Y)$ are given by Equations (3) and (4), respectively. The Bayes factor $B_{01}$ can be calculated following Spiegelhalter and Smith as [29]:

$$B_{01} = \sqrt{\frac{3}{2m}} \left(1 + \frac{F_\lambda}{n_1 + n_2 - 2}\right)^{(n_1+n_2)/2}$$

where $F_\lambda$ is the $F$-test statistic for assessing significance of carry-over effect $\lambda$. The value of $\kappa$ one chooses represents his (her) prior belief in carry-over. A value of $\kappa = 1$ indicates indifference to the choice of a model.

A graph of posterior treatment effect as a function of prior belief in carry-over can be plotted. It represents either the posterior mean and the 95% highest posterior density (HPD) interval or either $P(\tau < 0|Y)$ or $P(\tau > 0|Y)$ according to model parameterization against $P(M_1) = (1 + \kappa)^{-1}$. Both are given in Figure 3.

Figure 3 displays some of the possible summaries of the posterior distribution of $\tau$ according to the prior belief in carry-over for the example data. These are the posterior mean with 95% HPD interval and the probability of a benefit of transdermal nitroglycerin in the prevention of anginal attacks, $P(\tau < 0|Y)$. Details on the interpretation of such graphs are well described in the works of Grieve [15, 20, 27]. Briefly, the results show here that the conclusions strongly depend on the prior belief in carry-over, but it has been pointed out that it should be due to the relatively small value of $B_{01}$ (2.052, in the present case) [20].

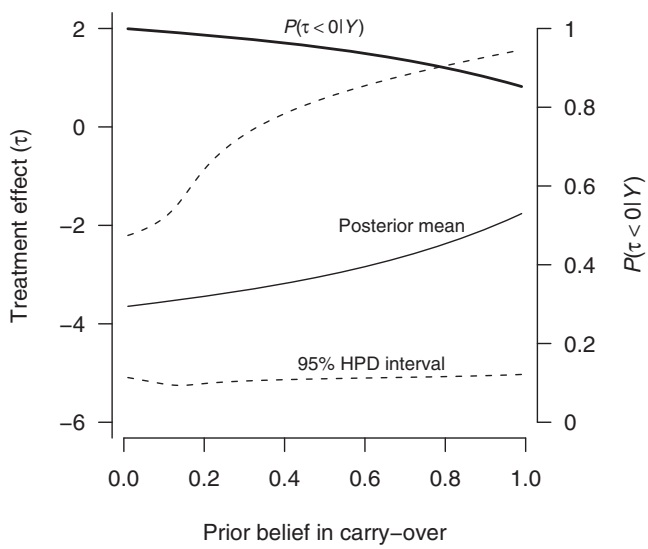


***FIGURE 3*** *Relationship between posterior inference on treatment effect $\tau$ and prior belief in a carry-over effect (expressed as probability of a model with carry-over) for the weekly anginal attack rates.*

The main advantage of Bayesian analysis of a cross-over trial is that it allows to stand in between the options of CROS and PAR analyses of a classical two-stage design. Several extensions of this technique have been considered but not reviewed here. Notably, it has been developed to consider baseline measurements [27] or nonuniform covariance matrices [30]. The use of informative priors has also been outlined [15].

## 4 OTHER ENDPOINTS

Analysis of AB/BA cross-over trials has generated a wide literature for normally distributed outcomes, as summarized in Section 3. We review here some methods for other situations.

### 4.1 Nonparametric Analyses

Model (1) and $t$ tests rely on the assumption that the outcome measurements $Y_{ijk}$ are normally distributed random variables. In some cases, however, such an assumption is clearly unverified. A first approach is to perform some variance stabilizing transformation, for example, taking the logarithm of data, in order to obtain a distribution closer to a normal distribution. Another choice is to prefer a nonparametric analysis that does not rely on a particular distribution for data. Nonparametric analyses are also well suited for ordinal variables with more than 4 or 5 classes. An excellent review of nonparametric analyses of cross-over trials is given by Tudor and Koch [31].

If a nonparametric analysis is found better adapted, the basic analysis described in Section 3.1 can be very simply carried out using a nonparametric test based on ranks such as Wilcoxon signed rank-sum test or a sign test, instead of the paired $t$ test. This corresponds to test whether cross-over differences obtained for each patients are centered around zero.

Several nonparametric methods accounting for a period effect are also available [11]. Koch, in an important study on nonparametric analysis of cross-over trials proposed to use a Wilcoxon rank-sum test (or equivalently a Mann–Whitney U test) to compare period differences across groups [32]. Contrary to the basic analysis, the period differences are constructed for each patient, that is, $Y_{ij2} - Y_{ij1}$, whatever $i$ is. A nonparametric test for equality of the distribution of period differences across both groups is then performed. A significant difference results in the conclusion that the average period difference is not the same in both group, which would imply a treatment effect. An estimator of this treatment effect can be obtained using Hodges–Lehman estimator, and 95% confidence limits may also be computed [33]. Senn has underlined the parallel between such a procedure and the CROS estimator of Hills–Armitage [11].

A more general procedure was also proposed by Koch that resembles the two-stage design discussed in previous sections [9, 31, 32]. In case of a significant test for treatment effect, as described above, a similar nonparametric test for carry-over is performed. Contrary to Grizzle's two-stage procedure, this test is not carried out first but has a supportive role in a second stage, in order to address the concerns

regarding the classical two-stage procedure [11, 14]. A test of period effect given no carry-over can also be performed.

Nonparametric methods have been extended to account for multiple centers or strata, or baseline measurements. For further reading, see Tudor and Koch [31] or Jones and Kenward [9].

### 4.2 Binary or Categorical Data

For binary data, analysis has long relied on more or less appropriate methods for contingency tables, with the exception of the Mainland–Gart test. More recent works share the common feature to use an explicit model [34–37] while former methods often relied on implicit models [38]. We briefly present here some of the possible approaches to the analysis of binary AB/BA cross-over data, as well as their extension to count data.

***Mainland–Gart Test*** The Mainland–Gart test examines the association between the period and group for patient preference [39, 40]. Subjects are classified according to their group into three categories: those who responded to treatment in period 1 and not in period 2, whose response is noted (1,0); those who responded to treatment in period 2 but not in period 1, that is, response (0,1); and those who had identical response after each period, noted (1,1) or (0,0) and who exhibited no preference. An example is given Table 6.

Under the null hypothesis, the preference (i.e., response in period 1 or in period 2) should only reflect the period effect. Ignoring patients who exhibited no preference such as in McNemar's test, that is, ignoring the counts $b$ and $e$ in Table 6, a difference between treatment should lead to an association between group and the category of response (1,0) or (0,1). Such an association can be tested using any test for unpaired proportions, such as Fisher's exact test or a chi-square test with or without continuity correction.

One of the disadvantages of such a procedure is that the number of subjects actually contributing to the test procedure may be small, as observations with the same response to both treatments are simply discarded. On the contrary, the Mainland–Gart test does not depend on the randomization of subjects into groups [9]. A test relying on randomization, but exploiting all information, was subsequently proposed by Prescott [41]. It will not be detailed here.

***Model-Based Methods*** As shown for continuous outcomes, cross-over trials rely on within-subject information for the comparison of treatments. As a consequence, they result in correlated observations. Methods for modeling correlated binary data, although theory exists for many years, have long suffered from unavailability in statistical software packages. Recent advances in this domain have led to a rapid

**TABLE 6 Contingency Table for Mainland–Gart Test**

| | Overall Response | | |
|---|---|---|---|
| Randomization Group | (1,0) | (1,1) or (0,0) | (0,1) |
| 1 | *a* | *b* | *c* |
| 2 | *d* | *e* | *f* |

diffusion of new methods for analysis of clustered binary data in the last years [34].

Contrary to the case of a linear model such as (1) for normally distributed data, the effect of a covariate on binary correlated data may have two different interpretations. The marginal effect corresponds to the average difference between two populations defined by differing covariate values, whereas the subject-specific or conditional effect corresponds to the expected difference for a same subject under different covariate values. These effect interpretations correspond to different possible models, namely the marginal or population-averaged model and the marginal or subject-specific model [34, 36, 42, 43]. Actually, for a linear model as (1), the outcome and the linear combination of covariates are on the same scale, so that parameters in both modeling approaches have the same interpretation, the difference lying in the error structure. For binary data, where a generalized model with a nonlinear link function relating the linear combination of covariates to the probability of response has to be used, both models cannot have the same interpretation anymore [36, 38].

Let us briefly detail these models and summarize their condition of use for AB/BA cross-over data. Paralleling model (1), a marginal logistic model can be expressed as

$$\text{logit } P(Y_{ijk}=1)=\log\left[\frac{P(Y_{ijk}=1)}{P(Y_{ijk}=0)}\right]=\mu+\pi_j+\delta_k\tau_l \tag{5}$$

where $\pi_j$, $\delta_k$, and $\tau_l$ have the same definition as for Equation (1). Other choices for the link function, such as a probit, for instance, are possible. Subject effect does not appear explicitly in model (5). Rather, dependence among observations from the same subject is accounted for via a robust covariance matrix. Model parameters are only related to the whole population, and estimates will be similar to classical logistic regression. The difference will lie in the standard errors of the parameter estimates. Several approaches have been proposed for model inference, one of the most popular being the generalized estimating equations (GEE) approach of Liang and Zeger [44]. These population-averaged models have been favored for practical computational aspects and availability in standard statistical software packages [38]. However, they are best adapted to assess the effects of cluster-level, that is, subject level in the case of cross-over trial data, covariates [36].

A conditional model can be written in a form even closer to model (1), as

$$\text{logit } P(Y_{ijk}=1)=\mu+s_{ij}+\pi_j+\delta_k\tau_l \tag{6}$$

with $s_{ij}$ denoting the effect associated with subject $ij$. As for model (1), observations are assumed independent conditionally of the subject effect, and the subject effect can be considered as either fixed or random. Two approaches to conditional modeling can thus be considered in practice. The conditional likelihood approach eliminates the subject effect [45] and corresponds to the Mainland–Gart test for the AB/BA design [38]. The other approach assumes a particular distribution for the subject effect, usually a normal distribution, which leads to a generalized linear mixed model (GLMM). This is the closest approach to model (1). In a conditional model, the covariate effects are measured conditionally on the subject effect $s_{ij}$. Thus,

regression parameters corresponding to covariates that do not change within subject may suffer from interpretation problems when using a conditional likelihood approach, as such effects are actually not observable [42, 43]. Conversely, GLMMs are based on both within-subject and between-subject comparisons, at the cost of distributional assumptions for the subject effects [36, 42]. For cross-over trials, however, where treatment effect estimation relies on inference about individual differences, conditional models are particularly well suited [36]. It must be noted that GLMMs have been long unavailable for use in routine practice, as major statistical packages did not offer this type of model. This was due in particular to the intractable likelihood function leading to very computer-intensive algorithms involving numerical integration. More recently, there has been a large development of methods, making the use of these models easier, though it still remains a topic of current research.

***Categorical Data*** For categorical data, the same types of model as described above for binary data exist. More detailed can be found in appropriate references [9, 38, 46].

### 4.3 Censored Endpoints

Few works in the literature have concerned the analysis of censored endpoints in cross-over trials [47–49]. At first, cross-over trials and censored data may seem not to fit with each other because one usually thinks of survival. However, a cross-over may be adapted if the treatments are aimed for recurrent events, such as infection or headache. Another domain could be measures of some form of endurance [49].

For a single duration response, France et al. [48] have proposed to use a Cox proportional hazards model stratified by patient [50]. Such a model cannot incorporate a group effect and thus explicitly does not address carry-over. A further analysis of the same example by Bristol proposed to use a group effect without stratification [51]. Both methods led, however, to very different results. Another approach used a parametric log-linear model for a Weibull process [49, 52]. The results showed that a stratified model should be preferred to an unstratified model with period effect.

A markedly different approach was adopted by Feingold and Gillespie [47]. It used an additive model instead of the proportional hazards model of France et al. [48], which was additionally based on the usual methods for complete (uncensored) data. Two tests were considered, namely the Wilcoxon test, which only applies to two-period designs, and a more general but less powerful score transformation procedure. It was shown that these methods were more powerful than that of France et al. [48].

Multiple duration responses were only considered by Lindsey, through the log-linear model evoked above [49].

## 5 POWER AND SAMPLE SIZE CONSIDERATIONS

It is sometimes stated that cross-over trials require half the sample size of a parallel-arm trial to demonstrate a given (direct) treatment effect because each patient

receives both treatments. Recent work based on a review of published cross-over studies has shown that a parallel design is expected to need between 4 and 10 times more subjects than a cross-over design in order to achieve the same power, leading to costs between 2 and 5 times higher [53]. In this section, we give sample size equations for a two-period two-treatment cross-over trial with a continuous outcome. We then provide some discussion on the comparative efficiency of cross-over and parallel trials. We place ourselves in the situation where there is no carry-over and analysis is performed using ANOVA in model (1) or, equivalently, CROS.

### 5.1 Sample Size

Sample size is usually computed to ensure the trial will have sufficient power against a given alternative hypothesis. Under this alternative, both treatments A and B have different direct effects and the treatment effect $\tau = \tau_2 - \tau_1$ is supposed equal to $\Delta$. This value is often referred as the minimal clinically relevant difference or the expected treatment effect. Sample size for cross-over trials are computed using usual methodology for the test of the null hypothesis $\tau = 0$. Let $\alpha$ and $\beta$ denote the type I and II error rates, that is, the probability of rejecting the null hypothesis when it is actually true and the probability of not rejecting the null hypothesis when the alternative is actually true, respectively. These error rates are usually fixed by convention at 0.05 for $\alpha$ and 0.10 or 0.20 for $\beta$ [54], though some different values can be sometimes found, such as lower $\beta$ or $\alpha = 0.025$ for one-sided tests. Moreover, the type II error rate is more frequently expressed using its counterpart $1 - \beta$, called power, which is the probability of rejecting the null hypothesis when there is actually a treatment effect. The null hypothesis can be rejected when $\hat{\tau}^c > 0$ or $\hat{\tau}^c < 0$ by a sufficient amount. Here, we consider the case when both chances of rejecting the null hypothesis are taken into account, which is referred to as a two-tailed or two-sided test.

As noted above, the null hypothesis of no treatment effect can be tested using $\hat{\tau}^c$. Recalling Table 1 for a balanced cross-over trial, that is, when $n = n_1 = n_2$, $\hat{\tau}^c$ has variance $\sigma^2/n$, and expectation 0 under the null hypothesis and $\Delta$ under the alternative. Assuming the within-subject variance $\sigma^2$ is known, $\hat{\tau}^c$ has a normal distribution and usual sample size for Gaussian variables can be used [9, 11, 54]:

$$n = \frac{\left(z_{1-\alpha/2} + z_{1-\beta}\right)^2 \sigma^2}{\Delta^2} \tag{7}$$

where $z_{1-\gamma}$ denotes the $(1 - \gamma)$ quantile of the standard normal distribution.

If $\sigma^2$ is unknown, it has to be estimated from the trial's data, and $\hat{\tau}^c$ has a Student distribution. One could use appropriate modification of Equation (7) using the noncentral $t$ distribution. As an approximation, a correction factor $\frac{1}{4} z_{1-\alpha/2}$ may be added to $n$, which improves the approximation of Equation (7) [54]:

$$n = \frac{\left(z_{1-\alpha/2} + z_{1-\beta}\right)^2 \sigma^2}{\Delta^2} + \frac{z_{1-\alpha/2}}{4} \tag{8}$$

To use Equation (7) or (8) in practice, one has to choose a value for $\Delta$ but also to obtain an a priori estimate for $\sigma^2$. This is naturally also the case when planning

any type of trial, and such values are usually chosen using existing trials or studies in the domain considered. However, for cross-over trials, it might be a little more difficult to obtain a value for $\sigma^2$, as it represents the within-subject variance only. It may be worth noting that $\sigma^2$ is half the variance of cross-over (or within-subject) differences, which may be more easily obtained using prior published information, for instance. The within-subject variance is also $(1-\rho)\sigma_T^2$, where $\sigma_T^2$ is the total variance of an observation.

As an example, consider the transdermal nitroglycerin trial used for illustration throughout this chapter. Suppose it was considered at the planning stage that a mean diminution of two attacks per week would be clinically relevant and that the within-subject variance was not known, but that in prior early phase studies where patients received transdermal nitroglycerin the within-subject before–after differences had variance of about 30. A value of $\sigma^2 = 15$ would then seem reasonable. With $\Delta = -2$ and, for example, $\alpha = 0.05$ and $\beta = 0.20$, formula (8) would lead to

$$n = \frac{(1.96+0.84)^2 \times 15}{4} + \frac{1.96}{4} = 29.89$$

By rounding up to 30, it would then be decided to recruit 30 patients in each group.

Sample size for cross-over trials in other situations, such as an equivalence or a noninferiority trial, or for bioequivalence, have been considered in detail in the review of Julious for normally distributed data [54]. A method for binary endpoints is given in Ezzet and Whitehead [46].

### 5.2 Efficiency of Cross-Over Versus Parallel-Group Design

A complete review of efficiency considerations for cross-over and parallel designs is given in Chapter 9 of Senn [11]. We sketch here some of these issues.

Consider a cross-over trial where $n$ patients are randomly assigned to each of the sequences AB and BA and a parallel-group trial where $p$ patients are randomly assigned to treatment by A and $p$ patients are randomly assigned to treatment by B. As considered higher, the estimate of treatment effect in the cross-over trial is $\hat{\tau}^c$, which has variance $\sigma^2/n$. Conversely, the estimate of treatment effect in the parallel design is analogous of $\hat{\tau}^p$, as all data will be recorded in a single (first) period. From Table 5, replacing $n_1$ and $n_2$ by $p$, the variance of the estimator of treatment effect in this design is $2\sigma^2/[p(1-\rho)] = 2\sigma_T^2/p$. To obtain equal variance estimates, it is then necessary to recruit $p = 2n/(1-\rho)$ patients per arm in the parallel-group trial. In the variance component model (1), $\rho \geq 0$, so that $p$ will always be larger than $n$, even in the case where between-subject variance is negligible as compared to within-patient variance. This is because the parallel-group design is also subject to within-subject variation, which is often overlooked, as remarked by Senn [11]. Once again, if the variance component model does not hold and negative values of $\rho$ are expected, it is anyway most improbable that a cross-over trial would be considered at all.

For choosing, whenever possible, between a parallel and a cross-over design, one may also take the duration of the trial into account, and not only the sample size. This should account for the two treatment periods of the cross-over trial but also for the

possible run-in and wash-out periods. Senn shows that when there is no run-in nor wash-out, the cross-over trial duration is always shorter than that of a parallel trial [11]. With no run-in but a wash-out duration equal to treatment period duration, the cross-over trial is shorter as soon as $1/(1 - \rho) > 3/2$, which is the case when between-subject variation is at least half the within-subject variation, a reasonable situation in practice. For more general cases, the reader may directly refer to Senn [11].

## 6 MORE COMPLICATED DESIGNS

There has been some work done toward finding the optimal design for a particular situation [9]. In particular, designs for two treatments but with more than two periods, have been studied, as a way of estimating carry-over and treatment-by-period interactions. Cross-over trials for more than two treatments, and usually at least three periods, have also been considered.

### 6.1 Higher Order Designs for Two Treatments

The main limitation of the AB/BA design without baseline measurements is that carry-over effects, treatment-by-period interaction, and group effects cannot be separated. Higher order designs, under specific and possibly disputable assumptions, may allow estimation of carry-over or treatment-by-period interactions.

Several assumptions or models for carry-over have been considered to examine design optimality. Some are, for example, the *simple carry-over* model, where the carry-over lasts exactly for one period and only depends on the last treatment, or the *steady-state carry-over* model, where the carry-over also lasts exactly for one period but treatment does not induce carry-over into itself.

Assuming either model, there have been numerous works on "optimality" for a wide range of cross-over designs with more groups and two periods (e.g., AB/AA/BA/BB) or with more periods (e.g., ABB/BAA) or both, such as ABA/BAB/AAB/BBA or AABB/BBAA/ABBA/BAAB [55–58]. The choice between these designs is not easy. Generally, increasing the number of periods increases the precision of the estimation of carry-over effects, but this gain is not always more efficient in terms of costs than including more patients. If the major cost is recruitment, then it has been advised to keep patients as many periods as is practical and ethical [9].

However, a greater number of periods may also lead to consider issues such as second-order carry-over or treatment-by-carry-over interactions. Overall, if a particular design should be recommended, the AABB/BBAA/ABBA/BAAB or the ABB/BAA designs would have the favor of Jones and Kenward [9].

These approaches, and particularly the assumptions concerning carry-over, have been quite sharply criticized [2, 11]. Some authors thus do not recommend multiperiod cross-over trials to be conducted except if prior knowledge on carry-over supports the underlying assumptions.

### 6.2 Designs for Three or More Treatments

Cross-over designs have been extended to compare more than two treatments. In practice, such designs have been preferred in situations where intervals between

treatments were short (e.g., some days). They have been quite popular in early phase I trials, with treatments also intended as several doses of the same drug.

Planning of such trials often relies on choosing sequences that form a Latin square. An example of such a design for four treatments (A, B, C, and D) may be the one displayed in Table 7, where patients would be randomly assigned in equal number to each of the four groups.

This Latin square ensures that every subject receives all four treatments, and each treatment occurs only once in each group. Moreover, the example given Table 7 also ensures that each ordered pair of treatments (e.g., A followed by B) occurs once and only once. The literature refers to these Latin squares as balanced. The efficiency of Latin square designs is the highest possible if the Latin square is balanced. As an example, the Latin square given in Table 8 is not balanced because treatment A is followed by treatment B more than once.

These designs are also most efficient if the simple carry-over model holds. Alternative models for other carry-over situations are considered in Jones and Donev [59].

Additionally, to achieve balance, it may be desirable that each treatment be followed by each other exactly once. This is not always possible in a simple Latin square, and sets of orthogonal Latin squares have been used for this purpose. Often, very few patients are included in each group. An example of a complete set of orthogonal Latin squares is presented in Table 9.

Such designs show the advantage of being balanced over all preceding treatments, so they are balanced for all first-order, second-order, and so on carry-over effects. Conversely, they may require much more subjects. For instance, using the design given in Table 9 implies including multiples of 12 subjects. Moreover, having an increased number of groups, they have a higher probability of a sequence being incorrectly administered to a patient. Williams showed that balance could be achieved by using a particular Latin square for an even number of treatments and

**TABLE 7 Balanced Latin Square Design for Four Treatments**

| | Period | | | |
|---|---|---|---|---|
| Group | 1 | 2 | 3 | 4 |
| 1 | A | D | B | C |
| 2 | B | A | C | D |
| 3 | C | B | D | A |
| 4 | D | C | A | B |

**TABLE 8 Another Latin Square Design for Four Treatments**

| | Period | | | |
|---|---|---|---|---|
| Group | 1 | 2 | 3 | 4 |
| 1 | A | B | C | D |
| 2 | B | C | D | A |
| 3 | C | D | A | B |
| 4 | D | A | B | C |

**TABLE 9 Orthogonal Latin Squares Design for Four Treatments**

| | Period | | | |
|---|---|---|---|---|
| Group | 1 | 2 | 3 | 4 |
| 1 | A | B | C | D |
| 2 | B | A | D | C |
| 3 | C | D | A | B |
| 4 | D | C | B | A |
| 5 | A | D | B | C |
| 6 | B | C | A | B |
| 7 | C | B | D | A |
| 8 | D | A | C | B |
| 9 | A | C | D | B |
| 10 | B | D | C | A |
| 11 | C | A | B | D |
| 12 | D | B | A | C |

two particular squares for an odd number of treatments [60, 61]. Such designs require less subjects than using complete sets of orthogonal squares.

All these designs have the same number of treatments and periods. Sometimes, more treatments are investigated than can be reasonably given to a single patient. Some designs thus may have less periods than treatments. Conversely, to better estimate carry-over effects, more periods than treatments could be necessary, as in the two-treatment case. Details on both types of designs can be found, for example, in Jones and Kenward [9]. Briefly, in the situation where the number of treatments is greater than the number of periods, the simplest method is to delete periods from a full set of orthogonal Latin squares [62, 63]. Another popular method is the balanced incomplete block design, where each treatment is replicated the same number of times, and each pair of treatments occurs in the same block the same number of times [63]. These designs are, however, not balanced for carry-over, as each treatment does not follow every other with the same frequency. More precisions can be found in cross-over textbooks [9, 11].

## 7 SUMMARY

Cross-over trials are trials where subjects subsequently receive all the treatments being compared (or at least some of them). Between the treatment periods, washout periods are used to avoid a residual or carry-over effect of a treatment into the next period. The estimate of treatment effect being based on within-subject rather that between-subject information, cross-over trials are usually more efficient, but they can only be used in situations where treatments will not durably modify patient conditions.

Although the use of complex models for cross-over trials was long limited by availability in common statistical software packages, all analyses presented here can be easily performed using standard software such as SAS (SAS Institute Inc., Cary, NC) or R (The R Foundation for Statistical Computing, Vienna, Austria).

## REFERENCES


1. Brown, B. W. (1980), The crossover experiment for clinical trials, *Biometrics*, 36, 69–79.
2. Fleiss, J. L. (1989), A critique of recent research on the two-treatment crossover design, *Controlled Clin. Trials*, 10, 237–243.
3. Grieve, A. P. (1994), Crossover studies, *Clin. Pharmacol. Ther.*, 56, 112–114.
4. Hills, M., and Armitage, P. (1979), The two-period cross-over clinical trial, *Br. J. Clin. Pharmacol.*, 8, 7–20.
5. Jones, B., and Lewis, J. A. (1995), The case for cross-over trials in phase III, *Statist. Med.*, 14, 1025–1038.
6. Senn, S. J., and Hildebrand, H. (1991), Crossover trials, degrees of freedom, the carryover problem and its dual, *Statist. Med.*, 10, 1361–1374.
7. Food and Drug Administration. (1977), A report on the two-period crossover design and its applicability in trials of clinical effectiveness. Minutes of the Biometric and Epidemiology Methodology Advisory Comittee (BEMAC) meeting.
8. ICH International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use. (1998). *Note for the Guidance on Statistical Principles in Clinical Trials (ICH-E9)*. Geneva, Switzerland. http://www.ich.org/LOB/media/MEDIA485.pdf.
9. Jones, B. J., and Kenward, M. G. (2003), *Design and Analysis of Cross-Over Trials*, 2nd ed., Chapman & Hall, London.
10. Ratkowsky, D. A., Evans, M. A., and Alldredge, J. R. (1993), *Cross-over Experiments, Design, Analysis and Application*, Marcel Dekker, New York.
11. Senn, S. (2002), *Cross-over Trials in Clinical Research*, 2nd ed., Wiley, Chichester.
12. Grizzle, J. E. (1965), The two-period change-over design an its use in clinical trials, *Biometrics*, 21, 467–480.
13. Willan, A. R., and Pater, J. L. (1986), Carryover and the two-period crossover clinical trial, *Biometrics*, 42, 593–599.
14. Freeman, P. R. (1989), The performance of the two-stage analysis of two-treatment, two-period crossover trials, *Statist. Med.*, 8, 1421–1432.
15. Grieve, A. P. (1994), Bayesian analyses of two-treatment crossover studies, *Stat. Methods Med. Res.*, 3, 407–429.
16. Wheatley, D. (1987), Transdermal nitroglycerin in angina pectoris, *Stress Med.*, 3, 199–203.
17. Draper, N. R., and Smith, H. (1981), *Applied Regression Analysis*, 2nd ed., Wiley, New York.
18. Senn, S. (1994), The AB/BA crossover: past, present and future? *Stat. Methods Med. Res.*, 3, 303–324.
19. Senn, S. (1998), Crossover designs, in Armitage, P., and Colton, T., Eds. *The Encyclopedia of Biostatistics*, 2 vols, Wiley, New York, pp. 1033–1049.
20. Grieve, A. P. (1985), A Bayesian analysis of the two-period crossover design for clinical trials, *Biometrics*, 41, 979–990.
21. Racine, A., Grieve, A. P., Flühler, H., et al. (1986), Bayesian methods in practice: Experiences in the pharmaceutical industry (with discussion), *Appl. Statist*, 35, 93–150.
22. Senn, S. J. (1991), Problems with the two stage analysis of crossover trials, *Br. J. Clin. Pharmacol.*, 32, 133.
23. Poloniecki, J. D., and Pearce, A. C. (1983), Interaction in the two-way crossover trial (Letter to the Editor), *Biometrics*, 39, 798.

24. Kenward, M. G., and Jones, B. (1987), The analysis of data from $2 \times 2$ cross-over trials with baseline measurements, *Statist. Med.*, 6, 911–926.
25. Spiegelhalter, D. J., Myles, J. P., Jones, D. R., et al. (1999), Methods in health service research. An introduction to Bayesian methods in health technology assessment, *BMJ*, 319, 508–512.
26. Selwyn, M. R., Dempster, A. R., and Hall, N. R. (1981), A Bayesian approach to bioequivalence for the $2 \times 2$ changeover design, *Biometrics*, 37, 11–21.
27. Grieve, A. P. (1994), Extending a Bayesian analysis of the two-period crossover to allow for baseline measurements, *Statist. Med.*, 13, 905–929.
28. Geisser, S. (1964), Estimation in the uniform covariance case, *J. Roy. Stat. Soc., B*, 26, 477–483.
29. Spiegelhalter, D. J., and Smith, A. F. M. (1982), Bayes factors for linear and log-linear models with vague prior information, *J. Roy. Stat. Soc., B*, 44, 377–387.
30. Grieve, A. P. (1992), Implementation of Bayesian methods in the pharmaceutical industry, PhD Thesis, University of Nottingham.
31. Tudor, G., and Koch, G. G. (1994), Review of nonparametric methods for the analysis of crossover studies, *Stat. Methods Med. Res.*, 3, 345–381.
32. Koch, G. G. (1972), The use of non-parametric methods in the statistical analysis of the two-period change-over design, *Biometrics*, 28, 577–584.
33. Hauschke, D., Steinijans, V. W., and Diletti, E. (1990), A distribution-free procedure for the statitsical analysis of bioequivalence studies, *Int. J. Clin. Pharmacol, Therapy Toxicol.*, 28, 72–78.
34. Carlin, J. B., Wolfe, R., Brown, C. H., et al. (2001), A case study on the choice, interpretation and checking of multilevel models for longitudinal binary outcomes, *Biostatistics*, 2, 397–416.
35. Kenward, M. G., and Jones, B. (1992), Alternative approaches to the analysis of binary and repeated measurements, *J. Biopharma. Stat.*, 2, 137–170.
36. Neuhaus, J. M. (1992), Statistical methods for longitudinal and clustered designs with binary responses, *Stat. Methods Med. Res.*, 1, 249–73.
37. Zeger, S. L., and Liang, K.-Y. (1992), An overview of methods for the analysis of longitudinal data, *Stat. Med.*, 11, 1825–1839.
38. Kenward, M. G., and Jones, B. (1994), The analysis of binary and categorical data from crossover trials, *Statist. Methods Med. Res.*, 3, 325–344.
39. Gart, J. J. (1969), An exact test for comparing matched proportions in crossover designs, *Biometrika*, 56, 75–80.
40. Mainland, D. (1963), *Elementary Medical Statistics*, 2nd ed., Saunders, Philadelphia.
41. Prescott, R. J. (1981), The comparison of success rates in cross-over trials in the presence of order effect, *Appl. Stat.*, 30, 9–15.
42. Hu, F. B., Goldberg, J., Hedeker, D., et al. (1998), Comparison of population-averaged and subject-specific approaches for analyzing repeated binary outcomes, *Am. J. Epidemiol.*, 147, 694–703.
43. Neuhaus, J. M., Kalbfleisch, J. D., and Hauck, W. W. (1991), A comparison of cluster-specific and population-averaged approaches for analyzing correlated binary data, *Int. Stat. Rev.*, 59, 25–35.
44. Liang, K.-Y., and Zeger, S. L. (1986), Longitudinal analysis using generalized linear models, *Biometrika*, 73, 13–22.
45. Breslow, N. E., and Day, N. E. (1980), *Statistical Methods in Cancer Research I: The Analysis of Case Control Studies*, International Agency for Research on Cancer, Lyon.

46. Ezzet, F., and Whitehead, J. (1991), A random effects model for ordinal responses from a crossover trial, *Stat. Med.*, 10, 901–906; discussion 906–907.
47. Feingold, M., and Gillespie, B. W. (1996), Cross-over trials with censored data, *Stat. Med.*, 15, 953–967.
48. France, L. A., Lewis, J. A., and Kay, R. (1991), The analysis of failure time data in cross-over studies, *Stat. Med.*, 10, 1099–1113.
49. Lindsey, J. K., Jones, B., and Lewis, J. A. (1996), Analysis of cross-over trials for duration data, *Stat. Med.*, 15, 527–535.
50. Cox, D. R. (1972), Regression models and life tables (with discussion), *J. Roy. Stat. Soc., B*, 34, 187–220.
51. Bristol, D. R. (1992), The analysis of failure time data in crossover studies (Letter to the Editor), *Stat. Med.*, 11, 975–977.
52. Lindsey, J. K. (1994), Fitting parametric counting process by using log linear models, *Appl. Stat.*, 44, 201–212.
53. Garcia, R., Benet, M., Arnau, C., et al. (2004), Efficiency of cross-over designs: An empirical estimation, *Stat. Med.*, 23, 3773–3780.
54. Julious, S. A. (2004), Sample sizes for clinical trials with normal data, *Stat. Med.*, 23, 1921–1986.
55. Ebbutt, A. F. (1984), Three-period crossover designs for two treatments, *Biometrics*, 40, 219–224.
56. Kershner, R. P., and Federer, W. I. (1981), Two-treatment crossover designs for estimating a variety of effects, *J. Am. Stat. Assoc.*, 76, 612–619.
57. Laska, E. M., and Meisner, M. (1985), A variational approach to optimal two-treatment crossover designs: Applications to carryover effects models, *J. Am. Stat. Assoc.*, 80, 704–710.
58. Laska, E. M., Meisner, M., and Kushner, H. B. (1983), Optimal crossover desings in the presence of carryover effects, *Biometrics*, 39, 1087–1091.
59. Jones, B., and Donev, A. N. (1996), Modelling and design of cross-over trials, *Stat. Med.*, 15, 1435–1446.
60. Williams, E. J. (1949), Experimental designs balanced for the estimation of residual effetcts of treatments, *Austral. J. Scient. Res.*, 2, 149–168.
61. Williams, E. J. (1950), Experimental designs balanced for pairs of residual effects, *Austral. J. Scient. Res.*, 3, 351–363.
62. Patterson, H. D. (1951), Change-over trials, *J. Roy. Stat. Soc., B*, 13, 256–271.
63. Patterson, H. D. (1952), The construction of balanced designs for experiments involving sequences of treatments, *Biometrika*, 39.

# Biomarkers

**Michael R. Bleavins,[1] Claudio Carini,[2] Malle Jurima-Romet,[3] and Ramin Rahbari[4]**

[1]*Michigan Technology and Research Institute, Ann Arbor, Michigan*
[2]*Fresnius Biotech of North America, Waltham, Massachusetts*
[3]*MDS Pharma Services, Montreal, Quebec*
[4]*Innovative Scientific Management, New York, New York*

**Contents**

1 Biomarker Definition and Criteria
2 Biomarkers and Decision Making
3 Life Cycle of Biomarker
4 Drug Safety
5 Efficacy
6 Patient Stratification
References

## 1 BIOMARKER DEFINITION AND CRITERIA

Biomarkers are valuable tools for drug development that provide greater accuracy or more complete information regarding drug performance and/or disease progression. Developments in genomics, genetics, proteomics, and imaging have highlighted the importance of biomarkers as useful clinical indicators. The Biomarkers Definitions Working Group [1], sponsored by the National Institutes of Health, has defined a biomarker as:

> A characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention.

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

In most cases, a biomarker should satisfy two criteria: (1) be associated with the biological mechanisms involved within a disease or treatment of a disease and (2) have the potential to correlate statistically with clinical outcomes. This broad definition means that a vast array of tests falls within this scope, spanning methods as varied as measurements of serum troponins and cardiac computed tomography. It is also important to clarify that most biomarkers are *not* surrogate endpoints. Under regulatory agency definitions, surrogate endpoints are findings or measurements that may be substituted for disease endpoints in clinical trials to evaluate the safety or effectiveness of a medical therapy for treating disease. These well-defined and accepted biomarkers represent a small subcategory of the biomarker world.

While there are different categories of biomarkers, three categories are most relevant to this discussion: target, mechanism, and outcome. Although a biomarker used in any stage of drug development may be from the target, mechanism, or outcome categories, in general, target biomarkers are more often part of discovery applications and outcome biomarkers are more likely to be seen in clinical situations. Qualification and acceptance of biomarkers, in applications paralleling the test's ultimate use and in cost-effective manners that minimize the possibility of patient harm, also are key considerations [2]. The primary goals of incorporating biomarkers into a drug development program, either clinically or preclinically, are to make better informed decisions earlier, more accurately, and to reduce the overall cost of drug development. The basic strategy is to define the critical question to be answered and then employ an appropriate biomarker in a relevant biological system that will answer that question as early as possible. When the purpose of the biomarker is to confirm hitting the putative therapeutic target, it then is possible to assess the relevance of the presumed pathophysiological mechanism and if altering this mechanism will affect clinical status.

Strategies for incorporating new biomarkers into clinical practice form the basis for translational medicine. One of the most tangible problems that scientists have been facing in recent years is finding biomarkers that are predictive, or "translate well," from animal or in silico models to humans. For example, inhibiting an enzyme in animal models may have a dramatic effect in the animal, whereas inhibiting the same enzyme in humans may have very limited or no clinical impact. The reasons for lack of correlation are many and varied: (1) the therapy can impact the biomarker chosen, but the parameter is irrelevant to the disease, (2) the biomarker may be indicative of the disease but not reflect clinically important effects of the treatment, and (3) a drug may work in humans through several different pathways not reflected in animal models. It is important to remember that a drug or other intervention should treat the disease, not a biomarker. Proper biomarker progress requires close dialog between biomarker scientists and practicing physicians, and this ideally can be accomplished under the banner of translational medicine.

The cost and duration of clinical trials are two of the primary drivers of biomarker research. Improving the understanding of a new drug's effects also is fundamental in using biomarkers within the pharmaceutical industry and for current regulatory focus on improving the success rates in drug development. Development of new therapeutic agents is estimated to cost between $800 million [3] and $1.7 billion [4] and requires 7–12 years. Reducing the high attrition rate of drugs by selecting better molecules, and being able to make decisions on compound viability earlier, have tremendous potential for reducing the high costs of bringing new

therapeutic agents to patients. The poor success rate for many compounds makes this an especially attractive goal. Attrition rates in phase II clinical research programs are as high as 55% [5] to 90% [6], with lack of efficacy as the primary cause. Thus, biomarkers that accurately predict human efficacy at the preclinical or phase I clinical trial stages are particularly valuable in reducing costs and increasing confidence in the compound. Similarly, biomarkers of chronic disease progression/resolution would substantially enhance clinical trial design and characterization of research efforts in diseases such as rheumatoid arthritis, type II diabetes, Alzheimer's disease, and Parkinson's disease.

Although many think of *omics* (proteomics, genomics, metabonomics, etc.) when the term *biomarkers* is used, many biomarkers are less exotic or "high-tech" and yet have contributed significantly to our understanding of disease processes. In 1733, Stephen Hales first measured blood pressure [7], and subsequent to this initial measurement, many investigators found blood pressure to be elevated in some disorders and reduced in others. Over many years, blood pressure was proven to have such clear clinical utility and predictivity to warrant acceptance as a surrogate for many forms of cardiovascular disease. Upon first description, however, blood pressure characterization was a new tool with unknown real potential. Although now deeply ingrained in medical practice, blood pressure initially presented a situation akin to the current status of C-reactive protein (CRP). The utility and significance of CRP [8, 9] are yet to be defined for inflammation and other diseases and only will be understood after more clinical experience is gained.

Although CRP is becoming a standard marker for inflammation, in the absence of known sources of inflammation, CRP is associated with the pathogenic inflammatory process in several diseases such as rheumatoid arthritis. Clearly, CRP and other newly discovered biomarkers need to stand the test of time in large clinical trials and long-term population studies. In addition, a biomarker like CRP must be reliably and conveniently measured. Clinical laboratories must be able to reproduce the measure, and national standards must be established so that consistent interpretation of measures from different laboratories can be achieved [10]. As the significance of inflammation has been identified in a variety of diseases, other inflammatory biomarkers are also the subject of current research interest. Lind [11] has described fibrinogen, leukocyte count, interleukin-6, tumor necrosis factor-α (TNFα), immunoglobulins, cell-adhesion molecules, complement, serum amyloid A, phospholipase-A2, and neopterin as other biomarkers of inflammation that might be useful in predicting coronary diseases. Like blood pressure, if these biomarkers prove important in mediating disease processes, they also will be targets for drugs that block their action, and thereby reduce the risk for disease.

## 2 BIOMARKERS AND DECISION MAKING

Biomarkers constitute a rational approach that, at its most optimal, reflects both the biology of the disease and the effect of the drug candidate. Proper incorporation of biomarkers in drug development strategy enables the concept of "fail fast, fail early," allowing early identification of the extremely high proportion of compounds that fail during drug development. In addition to minimizing human exposure to drugs unlikely to be effective or with safety concerns, substantial cost savings can

be achieved by shifting resources to those molecules most likely to become effective new medicines. A properly selected and characterized biomarker also facilitates the choice of the proper critical path toward approval and can differentiate the new product from approved drugs in a competitive marketplace. The challenge, therefore, is to identify relevant biomarkers early enough to implement them for "go, no-go" decisions at critical stages of the development process.

Traditional clinical trial endpoints, such as morbidity and mortality, often require extended time frames and may be difficult to evaluate. Imaging-based biomarkers are providing objective endpoints that may be confidently evaluated in a reasonable time frame. Imaging techniques tend to be expensive but can be cost effective when used in well-defined situations where subjective assessment has been the only approach available. Examples of therapeutic areas where imaging is reshaping clinical trial design include Alzheimer's disease, pain, and osteoarthritis. In affected joints, as well as in tissues expressing specific receptors, magnetic resonance imaging (MRI), computed tomography (CT), and position emission tomography (PET) imaging are delivering new information to clinicians and researchers. These biomarker technologies are rapidly developing, requiring close collaboration between basic researchers and physicians treating patients.

Biomarker development forms one of the cornerstones of a new working paradigm in the pharmaceutical industry by increasing the importance of linking diagnostic technologies with the use of drugs. The position of the Food and Drug Administration (FDA) is rather clear on this. Biomarkers are crucial to generate safe and efficacious drugs and are essential for deciding what patients should receive which treatment. It appears likely that development of pharmaceuticals also will drive the development of biomarkers as diagnostic probes for clinical decisions and stratification of patients. Preclinical biomarkers are essential to identifying compounds with the greatest potential for therapeutic intervention. Work at this stage of the process has tight controls of environmental conditions, diet, and genetics of the laboratory animals and provides access to tissues that cannot be collected during normal clinical trials. These studies provide opportunities for high-content data gathering using functional genomics, proteomics, metabonomics, simulations, computational methods/informatics, analytical technologies, and interventions. Results from the preclinical studies then can be distilled and used in the clinical phases of the most promising compounds.

In early-stage clinical trials, the use of suitable biomarkers should be incorporated to assist in demonstrating proof of concept (POC) and to identify appropriate dosage regimens for safe and efficacious clinical trials. Assessing which subpopulations are most likely to benefit from a new treatment can help in the planning of subsequent efficacy clinical trials. Although this approach may appear to be straightforward, integrating biomarkers can involve several practical problems and pitfalls. Clinical observations in a given patient may serve multiple purposes, with implications for the way in which the observed characteristics are documented, validated, interpreted, and used.

Aside from their use in diagnosis and prognosis, biomarkers currently have applications in drug discovery and development, aiding in formulating dosing and in guiding pharmacokinetic/pharmacodynamic (PK/PD) indices, and assisting in determining efficacy. These PK/PD indices are important factors in determining efficacious doses, particularly for evaluation in early-phase clinical trials. A deter-

mination of efficacy using biomarkers serves an important and cost-effective role in lieu of traditional large-scale clinical trials with standard clinical endpoints. Another important area that is well served by biomarkers is drug safety and the characterization of toxicity. Safety biomarkers can ensure exclusion of high-risk patients before treatment is administered, and monitor the need for early withdrawal of treatment before adverse side effects are manifested. These biomarkers also can be critical for assessing the relevance of preclinical findings and preventing or monitoring for their occurrence in humans.

## 3 LIFE CYCLE OF BIOMARKER

Biomarkers do not come to preclinical or clinical practice fully formed, fully validated, applicable across all patient populations, and with the guidance as to when to use or not use the particular marker. The tests are introduced poorly characterized with limited directions for understanding their limitations, potential, and applicability, much like an infant enters the life cycle. With care and further research, over time the biomarker can become a powerful tool to be utilized in drug development or patient care. In addition to the biological component of the biomarker requiring nurturing, the testing platform utilized to measure this biomarker usually does not exist as a point of care device that can be administered at the bedside. Rather, it is usually and necessarily by definition in the early stages, an untried, unreliable, and briefly tested platform that will be appropriately labeled "exploratory."

To illustrate the life cycle of a biomarker, it can be viewed from late maturity, the stage at which most individuals outside of the research laboratory will first encounter the test, working backwards to inception. At its most mature stage, an accepted biomarker is commonly referred to as "validated," meaning the biology is solid and the testing platform is reliable. But for the purpose of making what sort of decision? The term *validated* also conjures quite varied images in the minds of different people, with the same assay's robustness assumed to be greater as one ascends each step on the management ladder. Currently, the concept of "fit-for-purpose" is gaining acceptance in the pharmaceutical industry to provide more precise terminology and recognizing that the requirements for a biomarker when screening compounds is fundamentally different than when deciding to escalate doses in a clinical trial. Earlier in this chapter the different classifications of biomarkers were described (target, mechanism, outcome), in addition to the decisions that can be made with the information the biomarker provides and the types of testing platforms that can be utilized. Since the nature of questions asked in different situations varies greatly, the real answer for when a given biomarker is appropriate lies in whether it can be expected to reliably provide the information necessary to confidently make a decision. The risks and benefits to be weighed when conducting in vitro testing of a chemical library using high-throughput screening are very different than determining whether to recommend angioplasty or choosing between orally active compounds acting through Factor Xa inhibition versus coumarin therapy.

Determining a biomarker to be fit-for-purpose in the indication for which it is to be used is a more precise designator of the confidence in the combination of the

biology, platform, patient population, and the proposed action. This approach also takes into account that the magnitude of a decision can be substantially different even when the underlying questions are similar. For example, the acceptable false-positive/false-negative rates for a biomarker may be higher when one is prioritizing a large number of early-stage molecules than when determining whether to cancel research on a potential therapeutic target for which only one or two candidate compounds have been identified.

The degree of testing also can be affected by knowledge of similar drugs in the same class or current clinical practice. A biomarker supporting a new statin compound with a safety profile comparable to the marketed products is unlikely to require the same degree of characterization as a biomarker for drugs with a completely new indication (i.e., neovascularization). The potential of a biomarker to provide actionable information is directly proportional to its maturity, fitness for the intended purpose, and the gravity of the decision to be made using this information.

The development of a new biomarker minimally takes 6–9 months and can span many years (Fig. 1). Testing and building confidence in the biology can strengthen or wane, and the testing platform can improve by becoming smaller, more portable, and more widely available. In many ways, a biomarker goes through phases similar to the development of a child. The life cycle of a biomarker can be viewed as progressing from infant to toddler, then to child, preteen, adolescent, young adult, independent individual, and finally a mature person. The progression of blood pressure as a biomarker through these stages is illustrated below.

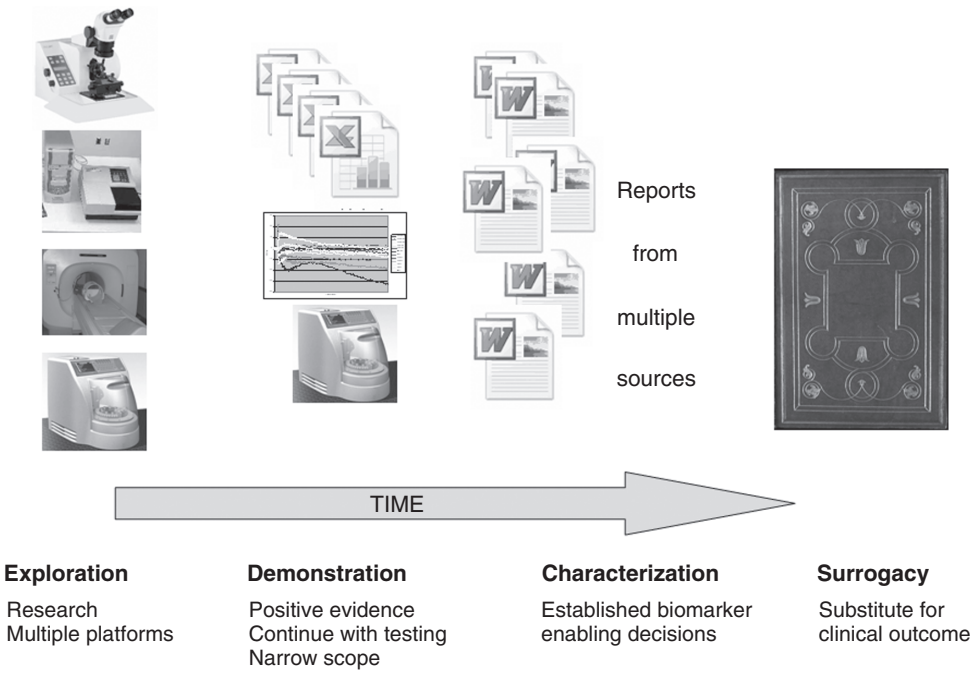


***FIGURE 1*** *Life cycle of a biomarker: exploration, demonstration, characterization, and surrogacy.*

In 1913, Janeway published a well-documented study where blood pressure was measured via sphygmomanometer in a large group of patients [12]. This study was made possible because of the blood pressure cuff developed by Scipione Riva-Rocci [13] and a large relevant patient population (~8000) examined over nearly a 10-year period. Janeway noted that 11% of the patients had systolic blood pressures over 160 mmHg, and this group of patients survived for only 4–5 years after being identified with the condition he named "hypertensive cardiovascular disease" [12]. The underlying physiology of hypertension was poorly understood at this time, and therapy options largely were limited to rest and reduced salt intake. Extensive research efforts in multiple laboratories led researchers in numerous directions, but it was not until 1939 that the renin–angiotensin–aldosterone system (RAAS) became the coherent focus of primary research [14]. Much more bench research and clinical findings, in addition to a consensus opinion of the community, led to the National High Blood Pressure Education Program Task Force I (NHLBI/NIH); *Report to the Hypertension Information and Education Advisory Committee* in 1973, and the report of the Joint National Committee (JNC I) on *Detection, Evaluation, and Treatment of High Blood Pressure* [15, 16]. These reports address reference ranges for the biomarker of blood pressure, compare recommended therapies and pharmaceutical interventions, and in the most recent release discuss patient stratification. This developing application and acceptance of blood pressure as a biomarker beyond its original use demonstrates its increasing maturity to the current status as one of the very few accepted surrogate biomarkers by both clinical practitioners and regulatory agencies. The life cycle of blood pressure is nearly 100 years and continues to grow as knowledge in the field accumulates and scientific and medical confidence in the biomarker increases.

Further illustrating that biomarkers are not a new concept, and their progression through various stages of development/acceptance, can be seen in routine medical practice. The likely tests performed at an annual physical examination include several biomarkers that have changed over the years, as has the clinical confidence in using them to make decisions. Among the tests administered will be body weight, diastolic/systolic blood pressure, heart rate, and serum lipid profile to provide an initial diagnosis of cardiovascular health. If any of these tests are skewed or abnormal, addition biomarker measurements would be initiated. Electrocardiogram (EKG), stress tests, troponins, CRP, and a host of other tests can characterize the presence or absence of heart disease, as well as inform the physician on appropriate therapy. Tests/biomarkers such as glucose, glycated hemoglobin, and alanine aminotransferase (ALT) provide information on other possible clinically important alterations and diseases that may indirectly impact cardiovascular health. Among these biomarkers, some can be considered "adolescent" (CRP) as "young adult" (prostate-specific antigen,), "independent individual" (troponin, stress test, body mass index, glycated hemoglobin), and "mature person" (blood pressure, EKG, lipid profile, glucose, ALT).

The two initial reports on hypertension mentioned above can be considered the forerunners of biomarker initiatives currently underway as consortia between the pharmaceutical industry and the FDA, International Life Sciences Institute/Health and Environmental Sciences Institute (ILSI/HESI), and National Institutes of Health (NIH). The driver for these present-day programs clearly is the economics of drug development. The process of drug development has grown significantly in

cost at the same time as the number of drugs submitted and approved per year has steadily decreased. The decline has resulted from a wide range of causes from the complexity of new disease targets, multifactorial diseases, increased regulatory requirements, and higher expectations of drug safety. As the acute causes of mortality and morbidity have been addressed by the medical community, the population now is experiencing the ill effects of more chronic diseases, and novel therapeutic targets and methods have become necessary. Many of these new therapeutic areas are still in need of predictive animal models for early-stage testing, as well as more comprehensive tools for defining the condition. Due to the greater uncertainty in these areas, the compound attrition rates have been higher. Inefficiencies in the drug development process and marketing expenses also have contributed to the increases in drug development costs. The marketplace also has a substantial impact on whether a new drug, even if approved, will actually be distributed to patients. A compound that cannot be sufficiently differentiated from similar approved molecules in terms of safety, efficacy, or cost, is unlikely to be included in the formularies of large insurers or become an approved course of treatment in countries with publicly funded health care systems. Biomarkers can be used to establish differentiation in class and this use will likely increase.

In the early-attrition paradigm designed to identify nonviable compounds at the first possible stage of development, many new biomarkers are being utilized by the pharmaceutical industry to achieve the cost-containment goals. These biomarkers are only sufficiently characterized to be considered a discovery tool and are generally limited to uses that prioritize similar early-stage molecules. Nevertheless, the tests are valuable for proving viability or proof of nonviability of a therapeutic target or class of compounds. These biomarkers most often are mechanism of action or safety markers. In contrast, disease-related outcome biomarkers can be difficult to validate in the absence of sizable efforts like the one for hypertension. Without additional research involving multiple clinical situations, possibly necessitating consortia to share resources, very few biomarkers currently being developed will mature past the point in their life cycle where they are used for internal decision making during pharmaceutical testing.

Biomarkers are seen as among the most promising tools for reducing the cost of developing drugs and ensuring that the appropriate patient populations are treated. Proper use of these tests can improve the quality and quantity of mechanistic information obtained at each step of the drug development process, allow better understanding of other results and the factors relevant to their interpretation, aid in the allocation of scarce resources, and provide clearer indications of a molecule's activity or potential safety liabilities. Clearly there is a wide spectrum of technologies encompassing biomarkers, with individuals from many specialties involved in the field with varying degrees of success. There also are numerous considerations to keep in mind when selecting a biomarker, and which must be balanced against each other, to ensure the test developed will meet the intended goal and actually be used. Among the critical biomarker parameters to be defined are:

- Nature of the biomarker's use (activity, efficacy, safety, differentiation)
- Throughput, capacity, and turn-around time requirements
- Reproducibility, sources of physiologic variation, and sensitivity
- Sample requirements and practicality of collection from the patients or preclinical species being evaluated

- Stability of parameter and compatibility with the study design
- Safety considerations in collecting and/or processing sample
- Existence of similar assay or approach that can be modified to the situation
- Time frame in which an appropriate method needs to be developed
- Defining how the biomarker adds value

Biomarkers of pharmacological or therapeutic activity and safety can be critical indicators of an early compound's activity and safety. This is especially important as drug development costs rise and the number of new compounds registered decreases. Appropriate selection and application of biomarkers can allow companies to terminate nonviable programs earlier, and reallocate resources to compounds with the highest probability of helping patients. Biomarkers also can be important for small companies in highlighting a new chemical entity's attractiveness and potential when looking for partnerships or licensing opportunities with big pharmaceutical companies.

## 4 DRUG SAFETY

Biomarkers in drug safety can be key determinants in expediting decisions and may foretell the survival of a molecule, or even an entire program. Serious effects seen in preclinical testing, particularly at low multiples of presumed efficacious exposure, necessitate either a biomarker that allows safe progression or discontinuation of the compound. Many drug safety biomarkers are clinical pathology parameters, with long histories and years of clinical use. Alanine aminotrasferase (ALT), troponins, urea nitrogen, bile acids, creatinine, electrolytes, and a host of other tests allow monitoring of tissue injury and have good correlations between humans and preclinical species. Safety departments within the pharmaceutical industry also are actively developing new biomarkers using the older platforms, as well as the technologies of proteomics, genomics, and metabonomics.

Discovering, developing, validating, and utilizing safety biomarkers present a variety of challenges in advancing pharmaceutical compounds. Safety has implications at all stages of the drug development process and carries significant relevance for clinical trial design and regulatory approval. Since the purpose of preclinical toxicology testing is to assist in defining relevant safety concerns for the clinical trials, biomarkers developed and evaluated in these experiments can be adapted for studies conducted in healthy volunteers and patients. Acceptable tests for the adverse effects seen in the preclinical testing that can be used in healthy human volunteers or patients are essential for safely advancing a compound into clinical trials. This connection between drug safety and clinical groups is one of the most important aspects of translational medicine.

Early identification of safety concerns, or indications that they will not be an issue, allows more effective prioritization and resource allocation. It also is often necessary to work backwards from a phenotype (liver injury, vasculitis, renal damage, etc.) caused by an unknown mechanism to a specific parameter, or group of markers, that caused the injury. The "wealth of opportunities and potential mechanisms" can be the most difficult and valuable aspect of the problem. In

contrast to many activity/efficacy biomarkers, biomarkers of specific tissue damage often can be applicable for monitoring toxicity of a range of molecules acting via a variety of mechanisms, but on the same tissue or subset of tissues.

Often even more challenging to manage are effects seen in preclinical species for which the clinical significance is unknown. Vasculitis has been an issue for a variety of therapeutic targets [antivirals, phosphodiesterase inhibitors, endothelin antagonists, norepinephrine reuptake inhibitors]. In these instances effects seen in animals may or may not be predictive of human risk, with the further complication that no good measures of the injury exist for monitoring in clinical trials. A biomarker panel capable of clearly monitoring vascular injury and early changes would allow compounds to be safely tested in human volunteers and patients, as well as having significant value in improving clinical practice related to vasculitis.

Identification, development, and validation of a safety biomarker follow approaches very similar to that required for efficacy assays. These include:

- Based on relevant parameters, with a foundation in a known mechanism.
- Coordinated engagement across departments (discovery, clinical, pharmacokinetics and drug metabolism, safety/toxicology) to optimize acceptance.
- Confidence in the test's performance that it can be used to make decisions.
- Use technology to efficiently demonstrate role in the toxicity, while remembering that the platforms are the tools and not the goal.
- Keep in mind that the question to be answered/characterized may be tissue rather than specific enzyme based—do not limit options.
- Comparative biology may be important in determining if the toxicity is relevant for human volunteers/patients or unique to the preclinical species.
- Good science is essential, as is the ability to show why the assay should be accepted as premonitory or able to detect changes before serious and irreversible injury has occurred.
- Ability to translate from preclinical to clinical application, as well as reverse engineering clinical tests, is critical for acceptance when entering new areas.

In many instances it is possible to utilize an efficacy biomarker to predict undesirable effects, particularly in areas such as anticoagulants. Thrombin inhibitors, low-molecular-weight heparins, Factor Xa inhibitors, and vitamin K antagonists are valuable therapeutic agents. Since most of these molecules at high doses exhibit only minimal toxicity other than those mediated through the same pathways as their pharmacologic methods of action (bruising, internal bleeding, etc.), the same assays that predict their activity can be used to assess toxicity. Tests such as ecarin clotting time, activated partial thromboplastin time, prothrombin time, and Factor Xa activity all represent useful safety and efficacy biomarkers. Since these assays show considerable translatability across species, and are already within the repertoire of most hospital laboratories, their acceptance by clinicians and regulatory agencies is enhanced. An added benefit for the development of drugs for which toxicity is generally limited to exaggerated pharmacology is that frequently early indications of efficacy and maximum tolerated doses can more accurately be estimated in phase I clinical trials than for most other compounds.

Additionally, it is highly desirable to have a biomarker that precedes overt tissue injury. Many of the oldest and most accepted biomarkers of toxicity represent leakage of cellular components and therefore only are seen at significant levels after at least minimal damage to an organ system has occurred.

In many instances, significant corporate overhead is automatically assumed whenever the situation requires a safety biomarker. Concerns are frequently voiced regarding regulatory agency acceptance, interpretation of new tests, and clinical practice. While at Pfizer, two of the authors of this chapter (MRB and RR) were among the scientists who advanced and advocated the "fit-for-purpose" approach (Bleavins and Rahbari, personal communication). Careful consideration of the decisions to be made for the test results, nature of the toxicity, and population being evaluated are all critical. However, open discussion across toxicology and clinical departments, as well as early engagement of the relevant regulatory agencies has been productive. It has been our experience that if the science is solid, and the rationale for use of the biomarker as part of the compound development strategy is well communicated, the physicians involved with the clinical trials and the regulatory agency reviewers will feel comfortable with the monitoring, and compound development is expedited.

## 5 EFFICACY

One of the most important areas in drug development for the application of biomarkers is defining and characterizing efficacy in clinical trials. As previously described, few validated biomarkers of disease efficacy and even fewer surrogate endpoints currently exist. Nevertheless, biomarkers are increasingly being incorporated into clinical trials in various therapeutic areas, often at early stages (e.g., phase I, phase IIa), and sometimes relying on advanced technologies such as genomic, proteomic, or imaging analyses. At this stage of drug development, there is a strong desire to have very quick assurances of hitting the target and a justification for continued development of the compound. In the later phases of development, especially after the initial proof of concept has been established, the more established measures of efficacy ("gold standard") may be employed.

Initially seen as an extension of the mantra of "fail fast, fail early" that was widely adopted by large pharmaceutical companies in the 1990s in preclinical development programs, the value of biomarker use in clinical development is increasingly being recognized for more than just identifying potential drug failures at an earlier stage. Integration of a drug's molecular response profile in animal disease models with clinical response enables better drug development decisions and efficiency. The current drug development paradigm is being reshaped by patient stratification using biomarker profiles.

An example of a clinical efficacy biomarker is an individual's BCR-ABL tyrosine kinase genotype. Characterization of the chromosomal rearrangement creating this abnormal protein, and the multiple subsequent mutations that can occur, are used to identify patients with chronic myeloid leukemia who are likely to respond to the drug, imatinib mesylate (Gleevac), which inhibits this kinase [17]. Additionally, imatinib also is an inhibitor of the C-KIT tyrosine kinase. Based on convincing clinical biomarker data, the U.S. FDA approved the use of imatinib to treat

BCR-ABL-positive chronic myeloid leukemia and gastrointestinal stromal tumors associated with activating mutations in C-KIT.

An additional well-characterized example is the anticancer drug trastuzumab (Herceptin), which achieved global sales in excess of $1.3 billion despite being efficacious in only a subset of patients [18]. Herceptin is an antibody that targets the HER2 receptor protein. Amplification of the human epidermal growth factor receptor *(HER)2* (ErbB2) gene in approximately 25–30% of patients with breast cancer results in elevated expression of HER2. The clinical response rate to treatment with Herceptin is high at 35% in patients with overexpression of HER2-while the drug is ineffective in patients who do not have increased protein levels. Therefore, clinical genotyping of breast cancer patients to identify HER2-positive cancers and determine a patient's suitability for Herceptin therapy is now commonly practiced. This approach maximizes the drug's use in patients who will benefit from the drug, while minimizing the exposure and cost to other patients of potential toxicity when those individuals are unlikely to respond. Although not a perfect biomarker, HER2 expression is a valuable tool for the oncologist, and additional research is being conducted to refine the predictivity of this measure, either by alternative tests or better characterization of key variables.

Efficacy biomarkers are, of course, not limited to genotypes and can be based on gene expression patterns, individual proteins or proteomic patterns, metabonomics, imaging, histology, physicians' clinical observations, or patients' self-reported observations (e.g., pain intensity). Thus, an efficacy biomarker is not defined by technology but by its reliable and predictive correlation to a differential response in patient subpopulations. It should be noted that rarely is the genomic or other molecular biomarker profile used alone to determine the drug therapy and dosage regimen. Many other factors such as age, body weight, concurrent diseases, and concomitant medications need to be considered as well, and the ultimate decision has to be made by physicians who are cognizant of all of these factors as they apply to the individual patient and his/her biomarker profile. The biomarker offers a valuable tool for enhancing and supplementing existing options, and "individualizing" the patient's treatment.

Interindividual differences in drug absorption, distribution, metabolism, and excretion (ADME) are known to be associated with differential efficacy or toxicity responses to many compounds. Genetic polymorphisms in drug metabolizing enzymes (DMEs) are a major cause of this substantial variability among individuals. This is especially important for drugs with narrow therapeutic windows that undergo extensive biotransformation to pharmacologically inactive metabolite(s) by one or more polymorphic drug metabolizing enzymes. Poor metabolizers (PMs) are typically at increased risk of adverse drug reactions, while ultra-rapid metabolizers may not be able to achieve therapeutic exposures when the parent compound is the active form. Prodrugs requiring metabolic activation by a polymorphic enzyme to an active metabolite also may have reduced efficacy or higher incidences of treatment failure in PMs. Three clinically important cytochrome P450s (CYP2D6, CYP2C9, CYP2C19) collectively are involved in the metabolism of approximately 40% of marketed drugs [19], and have a large number of functionally significant genetic polymorphisms. One example of an efficacy failure in PMs of CYP2D6 pertains to codeine, a central analgesic and cough suppressant that is metabolized to its pharmacologically active metabolite, morphine, by CYP2D6. The analgesic

effects of codeine are dramatically reduced in CYP2D6 PMs [20], whereas subjects who have multiple copies of the *CYP2D6* gene, so-called ultra-rapid metabolizers (UMs), have an increased analgesic response to normal doses of codeine, and in some cases, experience severe toxicity [21].

For the past several decades, pharmaceutical sponsors and clinical investigators have been conducting clinical pharmacology studies to evaluate the effects of genetic polymorphisms in DMEs using subpopulations of extensive and PMs (and UMs where appropriate). In some situations, this has led to postapproval labeling changes as new information in specific populations has become available. Several recent examples include the addition of deoxyribonucleic acid (DNA)-based DME biomarkers to the labeling of 6-mercaptopurine (Purinenthol), azathioprine (Imuran), and irinotecan (Camptosar) [22].

More recently, regulatory agencies have recognized that pharmacogenomic testing has evolved to the point that regulatory policies and guidance are required. In January 2005, the FDA approved the first clinical diagnostic test, Roche's AmpliChip for the rapid genotyping of *CYP2D6* and *CYP2C19* variants. To encourage more extensive sharing of genomic data, increase understanding on how and when to use this type of information, and determine opportunities for genomic studies to expedite drug development, the FDA published its *Guidance for Industry: Pharmacogenomic Data Submissions* in March 2005 [23].

Whether intended to improve efficacy or to avoid toxicity, biomarkers for selecting appropriate patients to enroll in clinical trials are becoming increasingly prospective rather than retrospective. Historically, as mentioned above, stratification has occurred after market approval when variation in the drug's effects has become better understood and could be correlated to results of a biomarker assay or diagnostic test. For example, the clinical utility of the kinase inhibitor gefitinib (Iressa) was defined after the drug failed to show efficacy in an unselected population of patients with lung cancer. The drug was, however, associated with marked responses in a small subset of patients with non-small-cell lung carcinoma [24]. There are several other examples where a nonstratified approach to drug therapy failed because of poor efficacy or unacceptable toxicity, and subsequent patient stratification based on biomarkers restored drug efficacy in a subset of patients. Human immune deficiency virus (HIV) treatment regimens are based on matching therapies to the viral strain present in the patient in consideration of resistance-conferring mutations, and antibiotics are matched to specific resistant infections. The anticancer drug irinotecan (Camptosar) has been associated with severe toxicity (diarrhea and leukopenia) in patients with a genetic variant of the UDP-glucuronosyltransferase 1A1 enzyme (UGT1A1) [25]. An FDA-approved diagnostic test based on the UGT1A1 biomarker is now available to identify patients potentially at higher risk for developing toxicity [26].

From a drug development perspective, lessons learned from these retrospective applications of biomarkers are increasingly leading to more prospective incorporation of safety and efficacy biomarkers into clinical trials. By enriching patient populations in clinical trials with likely responders, and excluding patients at greater risk of developing toxicity, drug development risks are reduced at the same time as the sample size necessary to demonstrate efficacy is reduced. Most importantly for diseases with traditional long-term endpoints such as survival, efficacy biomarkers allow for shortened endpoint observation times. In chronic and slowly progressive

diseases (arthritis, Alzheimer's disease, diabetes, lysosomal storage diseases), biomarkers can be evaluated after several weeks/months of drug treatment versus years for traditional clinical endpoints. The escalating cost of drug development and poor survival of drug candidates in phase II are causing pharmaceutical companies to reevaluate programs for chronic diseases for which there are no predictive biomarkers of efficacy. This is an even more critical concern for biotechnology and emerging pharmaceutical companies, reliant on venture capital or other short-term financing, for which it is not financially feasible to conduct multiyear clinical trials with compounds of unknown activity. Demonstrating proof-of-concept in trials of shorter duration by employing efficacy biomarkers can make the difference between continuing to pursue a new molecular target and/or molecular entity and eliminating a research program due to poor clinical promise. Indeed, the use of biomarkers has become common practice in phase IIa first-in-patient and even in some phase I first-in-human trials. This approach has the potential to provide useful information to validate preclinical disease models, assist in dose and biomarker selection to take forward into larger late-stage clinical trials, as well as excite management and investors with early "hints" of efficacy. However, the complexity and cost of early clinical trials involving biomarkers increases, and one should be careful that "more information" does not quickly become "too much information." Particularly for molecular profiling or multivariant biomarker strategies, the amount of data generated even for relatively small patient sample sizes can be enormous, and may require sophisticated and time-consuming data analysis and interpretation. For novel biomarkers, data on the normal variation in baseline (predose) levels may be limited, and a preliminary "survey" study may be necessary to define normal ranges. Unless the biomarker incorporated into the study design is decision making for the compound's future, biomarkers are rarely useful when they result in equivocal findings.

It is apparent that the increasing use of efficacy biomarkers will continue to make important contributions to clinical drug development by lowering late-stage failure rates, improving decision making and the efficiency of clinical trials, improving patient care, and paving the way toward personalized medicine.

## 6 PATIENT STRATIFICATION

In traditional clinical trial designs, patients with a particular condition were enrolled in clinical trials assuming a near homogenous group (Fig. 2), with random selection to prevent bias. The goal of random sampling employed in traditional drug development was to ensure the general population was well represented. In reality, the groups were heterogeneous with respect to stage of disease, drug metabolizing capabilities, environmental conditions (e.g., diet, smoking status, lifestyle), concurrent or previous medication exposures, and often even underlying cause of the symptomology. By using biomarkers to better characterize the biological makeup of the participants in clinical trials, drug developers are streamlining the testing process. The most obvious and often quoted use of patient stratification is to select only patients who would be expected to respond to a particular investigational new drug based on the mechanism of action of the drug. The drug's statistically significant benefits, in this targeted subset of patients, could be shown in a smaller and shorter clinical trial than those needed for randomly selected patients. Under some

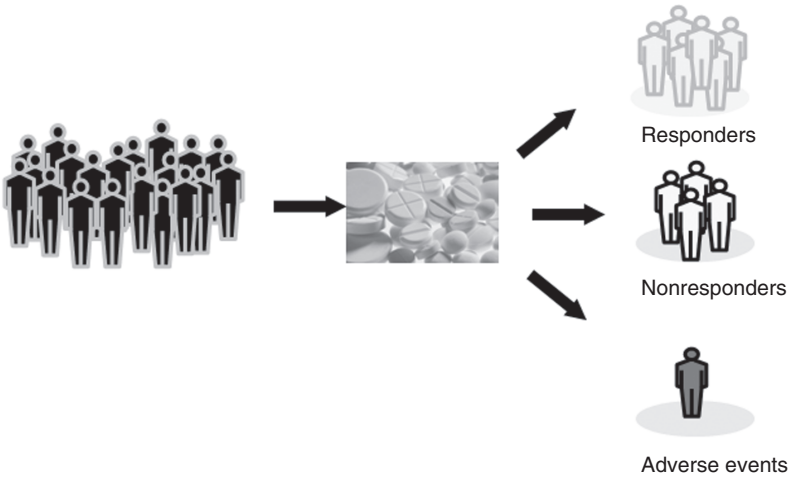


***FIGURE 2*** *Traditional treatment approaches often assume a relatively homogeneous patient population, with limited differentiation of individuals before administration of a new drug.*

circumstances, stratification of responders could also be performed retrospectively on the data collected to demonstrate efficacy of a compound in a subpopulation. While generally not acceptable for registration, retrospective findings can be important for identifying potential new therapeutic applications and redirecting programs (e.g., Neurontin for neuropathic pain, Rogaine for hair growth, Viagra for erectile dysfunction).

Minimally, stratification can speed approval for drug candidates intended for a subset of patients while leaving the door open for further testing and market expansion in the more heterogeneous population of patients. Maximally, it can unmask a useful therapeutic agent that would be lost in the noise generated by the nonresponders in the study, as described above for trastuzumab and gefitinib.

Since the challenge of testing new drugs in chronic diseases is one of the most difficult aspects of tackling these devastating conditions, an example in this area seems appropriate. Alzheimer's disease (AD) is a major health care concern that will only become more pressing as the baby-boomer generation ages. However, clinical studies for AD are difficult due to the generally slow progression of the disease, subjective nature of diagnostic tests, and limited ability of the affected individuals to actively participate in their care. Selecting early-onset AD patients may provide advantages in studying this disease because their symptomology presents with faster progression. Clinical trials could be shorter, thereby providing a quicker determination on a new molecule's efficacy. Unfortunately, this approach also has some rather severe limitations. Initial clinical trials still need to be many months in duration to demonstrate a compound's activity, and the early-onset form of AD may or may not effectively predict effects in the much larger population developing AD later in life. Far more valuable to identifying new therapies for AD would be a biomarker that accurately predicts disease progression or resolution in weeks or months. A biomarker of this type would revolutionize clinical trial design and allow significantly more new chemical entities to be tested for activity. It also would provide desperately needed information on the relevance of animal models,

further increasing the speed and productivity of basic research into this disease. AD is at the forefront of biomarker development for neurodegenerative diseases, and several promising candidate biomarkers have emerged in recent years, including isoprotanes, tau, Aβ, homocysteine, apoliporotein E genotype, and several others [27]. Imaging technologies such as resting glucose PET and volumetric MRI are also being investigated as biomarkers for AD.

The focus of pharmacogenetics is to identify the molecular causes of differential therapeutic response across patient populations. It is generally appreciated that patients and their disease show significant variations in response to a given treatment based on their genetic background in addition to the better recognized environmental causes. Advances in understanding the mechanisms underlying diseases, as well as drug response, are increasingly creating opportunities to match the right patients with the right therapies that are more likely to be effective and safe. At the extreme of patient matching is personalized medicine (Fig. 3). Currently, there are few examples of patient matching to therapies for which some form of biomarkers are available to indicate that a given patient is likely to show a response to a specific therapy. In addition to the BCR-ABL example described above, for the chemokine receptor 5 (CCR5) antagonist anti-HIV drug Maraviroc, prospective stratification based on CCR5 and HIV-1 genotyping results was included in the new drug application (NDA) submission. These data also were key components in the successful registration of the compound in 2007.

Patient stratification strengthens traditional medicine by associating an individual patient with a specific therapy. Patient stratification is already practiced in several contexts, though often after first-line empirical treatment proves unsatisfactory because of poor efficacy or intolerable toxicity. Examples can be seen in antibiotics that are matched to specific resistant infections, or the often protracted process of selecting the appropriate medicine for patients with schizophrenia. Biomarkers enable the progression from empirical therapy to stratification therapy, linking patient subsets with therapies that have an increased likelihood of success. In breast cancer, patients with different levels of receptor tyrosine kinase HER2 expression characterize not only their cancer type but even their candidacy for receiving treatment. Similarly, HIV therapies that are tailored to the viral variant

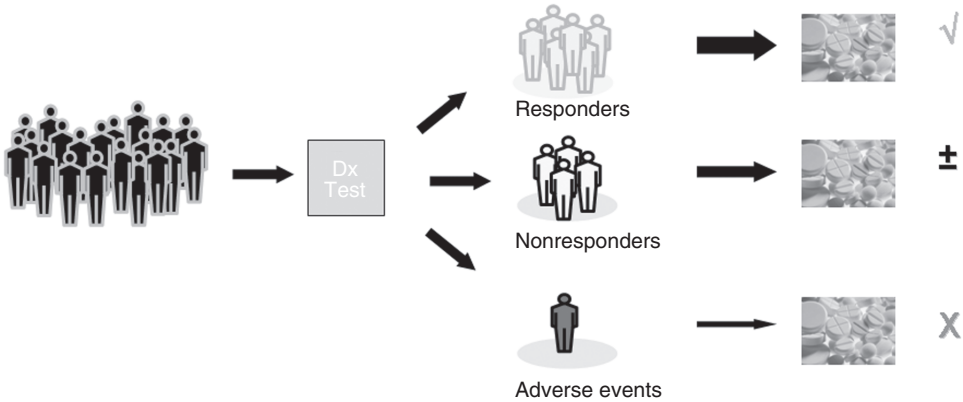


***FIGURE 3*** *Personalized medicine utilizes a diagnostic test/biomarker to optimize therapeutic decision making.*

present and oncology treatment regimes prioritize agents based on histological and molecular tests. In the absence of a biomarker, patient stratification may not be possible. For instance, the lack of accepted biomarkers in depression that are reflective of disease status or progression often results in patients receiving several different therapies to identify one that leads to a satisfactory response. At present the treatment for depression remains limited to empirical, trial-and-error therapy.

Patient's stratification has considerable economic impact on the model of the pharmaceutical industry and society. By identifying the populations most likely to benefit from a new therapy, drug development costs can be reduced at the same time as treatment of people unlikely to respond is minimized. Exposure to drugs that do not address a patient's underlying disease is minimized, thereby also reducing the likelihood of adverse events in those for whom the compound does not provide help. Overall, the risk–benefit assessment is substantively refined when the potential for adverse drug–drug interactions is reduced, and the process enhanced when biomarker data enabling patient stratification are available. Patient compliance also will be improved as treatment becomes more closely linked with disease resolution in a higher proportion of the individuals treated. Ultimately, better targeting of specific disease and responsive patient through biomarkers can reshape clinical practice by providing clinicians with new opportunities for better patient care.

## REFERENCES


1. Biomarkers Definitions Working Group (2001), Biomarkers and surrogate endpoints: Preferred definitions and conceptual framework, *Clin. Pharmacol. Ther.*, 69, 89–95.
2. Williams, S. A., Slavin, D. E., Wagner, J. A., et al. (2006), A cost-effective approach to the qualification and acceptance of biomarkers, *Nature Rev. Drug Dis.*, 5, 897–902.
3. DiMasi, J. A., Hansen, R. W., and Grabowski, H. D. (2003), The price of innovation: New estimates of drug development costs, *J. Health Econ.*, 22, 151–158.
4. Bain & Co. (2003), Cost estimates quoted in *Chem. Eng. News*, 81(50), 8.
5. Hurko, O. (2006), Understanding the strategic importance of biomarkers for discovery and early development phases, *Drug Disc. World*, 16, 63–74.
6. Bleavins, M. R. (2006), Personal communication.
7. Hamilton, W. F., and Richards, D. W. (1982), The output of the heart, in Fishman, A. P., and Richards, D. W., Eds., *Circulation of the Blood: Men and Ideas*, Bethesda, MD: American Physiological Society, pp. 83–85.
8. Ridker, P. M., Cushman, M., Stampfer, M. J., et al. (1997), Inflammation, aspirin, and the risk of cardiovascular disease in apparently healthy men, *N. Engl. J. Med.*, 336, 973–979.
9. Ridker, P. M., Buring, J. E., Cook, N. R., et al. (2003), C-reactive protein, the metabolic syndrome, and risk of incident cardiovascular events: An 8-year follow-up of 14,719 initially healthy American women, *Circulation*, 107, 391–397.
10. Pearson, T. A., Mensah, G. A., Alexander, R. W., et al. (2003), Markers of inflammation and cardiovascular disease: Application to clinical and public health practice. A statement for healthcare professionals from the Centers for Disease Control and Prevention and the American Heart Association, *Circulation*, 107, 499–511.
11. Lind, L. (2003), Circulating markers of inflammation and atherosclerosis, *Atherosclerosis*, 169, 203–214.

12. Janeway, T. C. (1913), A clinical study of hypertensive cardiovascular disease, *Arch. Intern. Med.*, 12, 755–798.
13. Riva-Rocci, S. (1896), Un nuovo sfigmomanometro, *Gazetta medica di torino* 47, 981–996; 1001–1017. [English translation: Faulconer, A. Jr., Keys, T. E., Eds., *Foundations of Anesthesiology*, Vol. 2, Charles C Thomas, Springfield, IL, 1965, pp. 1043–1075.]
14. Vertes, V., Tobias, L., and Galvin, S. (1991), Historical reflections on hypertension, *Prim. Care*, 18(3), 471–482.
15. National High Blood Pressure Education Program, Task Force I (1973), *Report to the Hypertension Information and Education Advisory Committee*, Recommendations for a national high blood pressure data base for effective antihypertensive therapy, U.S. Department of Health, Education and Welfare Publication No. (NIH) 74-593, Government Printing Office, Washington, DC.
16. JNC I (1977), Report of the Joint National Committee on Detection, Evaluation, and Treatment of High Blood Pressure (JNC I), *JAMA*, 237, 255–261.
17. Druker, B. J., Talpaz, M., Resta, D. J., et al. (2001), Efficacy and safety of a specific inhibitor of the BCR-ABL tyrosine kinase in chronic myeloid leukemia, *N. Engl. J. Med.*, 344, 1031–1037.
18. Burstein, H. J. (2005), The distinctive nature of HER2-positive breast cancer, *N. Engl. J. Med.* 353, 1652–1654.
19. Xie, H.-G., and Frueh, F. W. (2006), Pharmacogenomics steps toward personalized medicine, *Personalized Med.*, 2, 325–337.
20. Sindrup, S. H., and Brosen, K. (1995), The pharmacogenetics of codeine hypoalgesia, *Pharmacogenetics*, 6, 335–346.
21. Gasche, Y., Daali, Y., Fathi, M., et al. (2004), Codeine intoxication associated with ultrarapid CYP2D6 metabolism, *N. Engl. J. Med.*, 351, 2827–2831.
22. Huang, S.-M., Goodsaid, F., Rahman, A., et al. (2006), Application of pharmacogenomics in clinical pharmacology, *Toxicol Mechanisms Methods*, 16, 89–99.
23. U.S. FDA (2005), *Guidance for Industry: Pharmacogenomic Data Submissions*; available at: http://www.fda.gov/cder/genomics.
24. Dowell, J. E., and Minna, J. D. (2005), Chasing mutations in the epidermal growth factor in lung cancer, *N. Engl. J. Med.*, 352, 830–832.
25. Andersson, T., Flockhart, D. A., Goldstein, D. B., et al. (2005), Drug-metabolizing enzymes: Evidence for clinical utility of pharmacogenomic test, *Clin. Pharmacol. Ther.*, 78, 559–581.
26. U.S. FDA (2005), FDA clears genetic test that advances personalized medicine: Test helps determine safety of drug therapy; available at: http://www.fda.gov/bbs/topics/news/2005/new01220.html.
27. Shaw, L. M., Korecka, M., Clark, C. M., et al. (2007), Biomarkers of neurodegeneration for diagnosis and monitoring therapeutics, *Nature Rev. Drug Disc.* 6, 295–303.

# *Biomarkers in Clinical Drug Development: Parallel Analysis of Alzheimer Disease and Multiple Sclerosis*

**Christine Betard,[1] Filippo Martinelli Boneschi,[2] and Paulo Caramelli[3]**

[1]*Global Strategic Drug Development Unit, Quintiles, Levallois-Perret Cedex, France*
[2]*Neuro-Rehabilitation Unit, Department of Neurology, San Raffaele Scientific Milano, Italy*
[3]*Cognitive Neurology Unit, Department of Internal Medicine, Faculty of Medicine, Federal University of Minas Gerais, Belo Horizonte, Brazil*

**Contents**

1 Introduction
  1.1 Working Definitions
2 When to Use Biomarkers
  2.1 Diagnosing AD and MS and Biomarkers to Identify Disease: Diagnostic Biomarkers
  2.2 Identifying Risk of Developing MS and AD: Presymptomatic Biomarkers
3 Clinical Development Application: Drug Efficacy and Safety
  3.1 Using Biomarkers to Select Patient Subgroups
  3.2 Predicting Changes in Response to Therapy and Disease Course: Predictive Biomarkers
4 Conclusion
  Acknowledgments
  References

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

## 1 INTRODUCTION

Biomarkers are being used increasingly in clinical drug development by research-based pharmaceutical and biotechnology companies. Examples of their application to the development of new medicines to treat chronic disorders will be discussed in two central nervous system (CNS) diseases: multiple sclerosis (MS) and Alzheimer disease (AD). Combined, these two neurological disorders are creating great suffering to the affected patients and their families and represent a public health burden to societies. MS and AD are common and devastating neurological diseases for which pathophysiological processes leading to inflammation and neurodegeneration are thought to begin long before early clinical symptoms emerge and thus can be detected. For many years, AD and MS have been considered as two completely different diseases with no common features, MS being characterized by diffuse inflammation and demyelination, inflammation resulting in sclerotic scars (MS plaques) during healing [1], and AD clinically characterized by an insidious and progressive loss of memory and cognitive abilities, with *definite* diagnosis requiring pathological confirmation of amyloid plaques and neurofibrillary tangles associated with brain atrophy resulting from neuronal and synaptic loss [2–4].

Despite substantial progress, existing therapeutic interventions are only moderately effective in improving clinical symptoms, and research efforts are focusing on the development of disease-modifying therapies to impede disease progression. There is a widely shared perception that biomarkers are critical to identify the presymptomatic phase of the disease and to allow the development of these new therapies, since limitations in animal models and inadequate understanding of the pathophysiology still persist [U.S. Food and Drug Administration (FDA) Critical Path Initiative, European Medicines Agency (EMEA) Innovative Medicine Initiative].

In recent years, transcriptomics and neuroproteomics analyses as well as genetic association studies have revealed that AD and MS pathophysiologies share common features of inflammation and neurodegeneration [1, 5–8] and also glucose metabolism impairment [9]. One of the main challenges in improving care of MS and AD patients resides in the poor performance of clinical diagnostic criteria that, in their current state, do not allow MS or AD diagnosis before the appearance of characteristic clinical symptoms but instead many years after the neurodegenerative and inflammatory processes have been started (Figure 1). Consequently, the diagnosis is made and the therapy is initiated when the disease process has reached irreversible steps, rendering the impact of potential disease-modifying therapies almost negligible in terms of clinical benefit. As basic molecular events in the initiation and progression of the diseases are still poorly understood, a rather long gap between disease onset and diagnosis remains. In addition, MS and AD show a high degree of individual variability in severity and clinical course, highlighting the existence of disease subtypes that current diagnostic criteria do not allow differentiation and which could necessitate different treatments.

A large panel of clinical biomarkers is beginning to be used in early drug development to determine whether the drug is reaching and affecting the appropriate molecular target in humans, allowing comparisons with preclinical data and providing measurable endpoints that predict desired or undesired clinical effects. The ultimate goal of this approach is to increase the success rate in the confirmatory stage of clinical development. Noninvasive neuroimaging techniques, particularly magnetic reso-

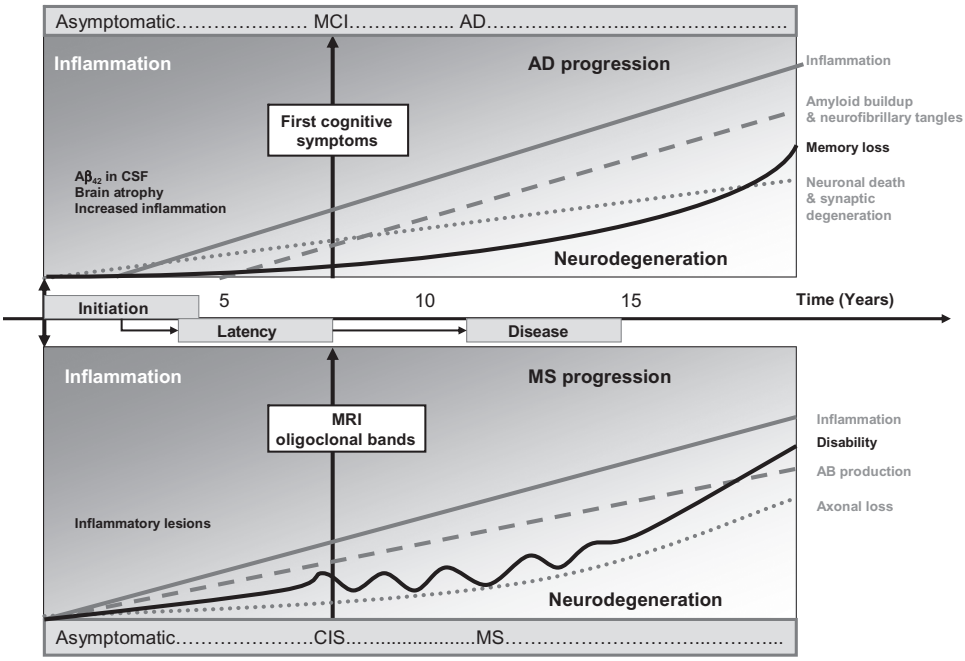


***FIGURE 1*** *Different stages of MS and AD related to time frame and window for therapeutic intervention. MCI: mild cognitive impairment; CIS: clinically isolated syndrome.*

nance and positron emission tomography (PET), are promising methods applied to clinical trials to assess potential therapeutic mechanisms in MS and AD, in addition to molecular biomarkers emerging from genomics and proteomics research.

In this chapter, molecular and neuroimaging biomarkers are reviewed in parallel for MS and AD, focusing on inflammatory and neurodegenerative common features and highlighting unique biomarkers specific to each condition. Possible applications of biomarkers in clinical development are also discussed.

Very intensive research efforts in genetic research in these two multifactorial and genetically complex disorders have led to the identification of a number of genetic risk factors associated with disease, such as apolipoprotein E (ApoE) polymorphism in sporadic AD, amyloid precursor protein (APP), presenilin 1 (PSEN1) and presenilin 2 (PSEN2) mutations in familial forms of AD, and human leukocyte antigen (HLA) locus in MS. None of the biological indicators tentatively associated with the disorders has demonstrated enough sensitivity and specificity to allow their use as symptomatic or presymptomatic diagnostic tools.

This well-known limitation to diagnose MS and AD patients in a clinical course window between disease onset and clinical manifestation is undoubtedly linked to the absence of cure for these neurological disorders. Analysis of the clinical success rate from first-in-man to registration is showing that CNS compounds have one of the lowest success rates compared to other therapeutic areas, the lack of efficacy being the major cause of attrition in the clinic and accounting for approximately 30% of failures, with high failure rates in phases II and III trials [10] (Figure 2).

Different approaches have been proposed to reduce attrition in drug development, starting from the earliest stages of discovery. These include getting very strong

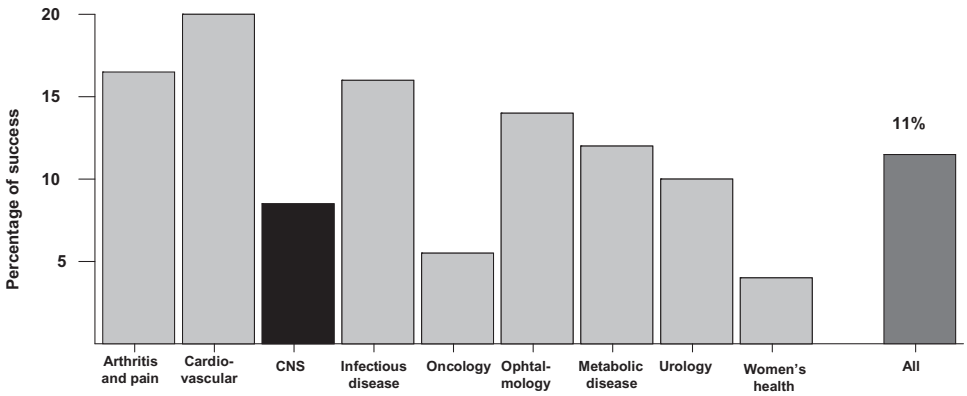


***FIGURE 2*** *Success rates from first-in-man to registration. The overall success rate is 11%. However, if the analysis is carried out by the therapeutic areas, big differences emerge. The data are from the ten biggest drug companies during 1991–2000. (The companies are AstraZeneca, Bristol-Myers Squibb, Eli Lilly, F. Hoffman-LaRoche, GlaxoWellcome Johnson & Johnson, Novartis, Pfizer, Pharmacia, Schering-Plough, and SmithKline Beecham; data were obtained by Datamonitor in the Pharmaceutical Benchmarking Study.) CNS: central nervous system. Reprinted by permission from Macmillan Publishers Ltd: Nature Reviews Drug Discovery, vol. 3, Kola, I., and Landis, J., Can the pharmaceutical industry reduce attrition rate? 711–715, copyright 2004.*

evidence for proof of mechanism with modulation of target involved in relevant disease pathways, eliminating compounds with poor toxicity profiles, and using appropriate animal models for efficacy testing in preclinical studies. It is interesting to note that CNS pharmacological research, one of the areas with the highest attrition rate, is also the area where animal models are very poor predictors of human pathophysiology for AD as well as for MS.

In this context, clinical biomarkers have emerged as a critical tool to help define disease and to determine pharmacological response [11–13]. Biomarker research is intensively focusing on diagnostic biomarkers to track disease at its initiation or when only mild symptoms occur, as disease-modifying therapies could more effectively slow down the disease process if prescribed early. Intensive biomarker research is also conducted on prognostic and predictive biomarkers to evaluate disease severity, progression rate, and patients at risk who would most benefit from novel treatments to anticipate efficacy and/or safety of symptomatic and/or disease-modifying neurotherapies [14, 15].

## 1.1 Working Definitions

The National Institutes of Health Biomarkers Definitions Working Group has agreed on the classification of biomarkers allowing a clear distinction between biomarkers and clinical endpoints [16], with surrogate endpoints being a subset of biomarkers:

*Biological Marker = Biomarker* A characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes, or pharmacological responses to a therapeutic intervention.

*Surrogate Endpoint* A biomarker that is intended to substitute for a clinical endpoint. A surrogate endpoint is expected to predict clinical benefit (or harm

or lack of benefit or harm) based on epidemiological, therapeutic, pathophysiological, or other scientific evidence.

*Clinical Endpoint* A characteristic or variable that reflects how a patient feels, functions, or survives.

The optimal biomarker should identify a universal and fundamental feature of the pathophysiology of the disease and should be validated in patients who have developed the disease, such as in neuropathologically confirmed patients for AD or in morphologically characterized MS patients with demyelination, inflammation, gliosis, and axonal damage. The optimal biomarker should also be independent of the clinical symptoms, have high specificity and sensitivity, and possess specific technical characteristics, such as being noninvasive, allowing repeated measurements, and being easy to test and inexpensive. A combination of biomarkers may be necessary to achieve all/most of these components.

## 2 WHEN TO USE BIOMARKERS

### 2.1 Diagnosing AD and MS and Biomarkers to Identify Disease: Diagnostic Biomarkers

***Diagnosing AD in Clinical Practice*** Over the last two decades, a wealth of evidence has supported the amyloid hypothesis in AD, with a central role attributed to the accumulation and deposition of fibrillar β-amyloid driving neurodegeneration and cognitive decline and leading to dementia [17–19]. Advances in understanding the neurobiology of the normal aging brain as well as the mechanisms of AD have allowed increasing agreement between clinical (antimortem) and autopsy (postmortem) diagnoses in demented aged individuals.

The diagnosis of AD is currently based on clinical criteria [2–4]. According to the NINCDS-ADRDA criteria [4], the diagnosis of AD is classified as probable, possible, or definite. Definite AD depends on the observation of amyloid neuritic plaques and neurofibrillary tangles at neuropathological examination. The diagnoses of probable and possible AD rely on proper clinical information. Probable AD is defined by progressive memory decline associated with impairment of at least one additional cognitive function, such as language, visuospatial, or executive abilities, in the absence of other brain or systemic diseases that may cause cognitive decline. The diagnosis of possible AD admits the occurrence of an atypical clinical course, such as plateaus, or the presence of a concomitant clinical abnormality capable of producing cognitive impairment but not considered by the clinician to be sufficient to produce dementia.

Although being essentially based on exclusion, the diagnostic criteria of probable AD displays good sensitivity (81%), although with lower specificity (70%), in relation to the gold-standard pathological diagnosis, and appropriate clinical follow-up substantially increases this diagnostic accuracy [20]. Disease duration is typically 8–10 years but ranges from 2 to 25 years.

Some ancillary tests have been found to increase the diagnostic accuracy of AD. Hence, atrophy of the hippocampus and of the entorhinal cortex, indicated either by volumetric analysis [21–23] or by visual inspection [24, 25], on magnetic resonance imaging (MRI) and PET scan showing decreased metabolic rates in posterior temporoparietal, posterior cingulate, and prefrontal cortices [26, 27], increased levels

of total and hyperphosphorylated tau and reduced concentrations of amyloid beta (Aβ) in the cerebrospinal fluid (CSF) [28, 29], and high $A\beta_{42}$ plasma levels [30] have all repeatedly shown to improve the clinical diagnosis of AD. Moreover, the presence of one or two copies of the *ε4* allele of the *APOE* gene also increases diagnostic specificity [31, 32].

***Diagnosing MS in Clinical Practice*** MS is the most frequent chronic inflammatory demyelinating disorder of the CNS, and its diagnosis is based on the demonstration of a temporal and spatial dissemination of inflammatory lesions in the CNS. This demonstration is obtained on a clinical basis and also using surrogate outcome measures like brain and spinal cord MRI and CSF examination. A recent revision of diagnostic criteria has been performed by a panel of experts [33], with recent modifications [34]. This is further complicated by the lack of existence of a definite histopathological diagnosis, like in AD, and by the relatively low number of autoptic cases so far analyzed. These criteria are universally accepted, but their validity and concordance among different examiners are yet to be established. On MRI, findings of multifocal lesions of various ages, especially those involving the periventricular white matter, the brainstem, the cerebellum, and the white matter of the spinal cord, support the diagnosis, and the appearance of gadolinium-enhancing lesions on MRI indicates sites of active inflammation and rupture of the blood–brain barrier. CSF analysis supports the diagnosis if it shows an increased intrathecal synthesis of immunoglobulins which are not present in the serum—so-called oligoclonal band (OB) positivity at the isoelectrofocusing. In the absence of gold standard for the diagnosis of MS, it is difficult to give specific accuracy features of MRI and CSF OB methods. There is a considerable clinical heterogeneity of the disease, for which there are essentially two different modes of presentation. The most common is the relapsing–remitting (RR), which occurs in about 85% of the patients, and it is characterized by the occurrence of episodes of neurological manifestations followed by complete or partial recovery. About 80% of RR patients develop a secondary progressive (SP) course, with or without superimposed relapses at a certain time after the onset: nearly 50% within 10 years and more than 80% within 25 years from the onset [35], while the remaining 20% are fully functional after 15 years of disease and are defined as benign cases. Ten to 15% of patients are affected by the primary progressive (PP) form, which is characterized by a progressive accumulation of irreversible disability since the disease onset. In MS, between 55 and 65% of patients experience cognitive decline as revealed by neuropsychological test battery, which is more frequent and severe in patients with the progressive phase [36–39] and in association with the *APOE4* genotype in severe cases of cognitive impairment [40]. The cognitive decline resembles the pattern seen in subcortical dementias, and it is clinically characterized by slow information processing, impaired memory retrieval, deficient problem solving, and personality and mood disturbances as a consequence of the disruption of the white matter connections between the frontal lobes and the subcortical structures.

## 2.2 Identifying Risk of Developing MS and AD: Presymptomatic Biomarkers

Many studies have suggested that pathological processes occurring in MS and AD begin long before the onset of clinical symptoms. For instance, in relation to AD, neuropathological studies have demonstrated that a large proportion of cognitively

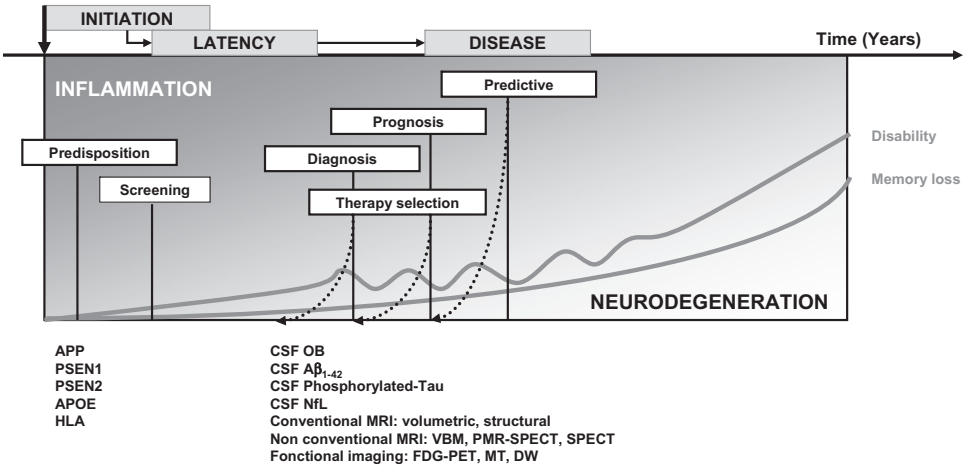


***FIGURE 3*** *Biomarkers in AD and MS. The goal of the next generation of biomarkers is to allow a shift in the detection of disease stages at earlier stages: from symptomatic AD or MS to presymptomatic MCI or CIS to asymptomatic, as indicated by the dotted line arrows for each biomarker subtype making disease-modifying therapies effective in preventing disease or slowing down its progression.*

normal elderly individuals have amyloid plaques when autopsy is performed [41, 42]. These presymptomatic phases have been characterized as clinically isolated syndrome (CIS) for MS, which is defined as the occurrence of a first acute episode of neurological worsening due to a single white matter lesion and usually happens in about 85% of patients [43] and mild cognitive impairment (MCI) for AD [44]. The identification of these early disease stages are critically important for the development of disease-modifying therapies designed to slow the underlying progression of the disease, therapies likely to be the most valuable for patients with mild symptoms or asymptomatic patients. The development and validation of biological markers, neuroimaging tests, and other objective indicators of preclinical AD and preclinical MS are regarded as the most important research areas to aid in identifying individuals at high risk for developing symptomatic AD or MS [45–47] (see Figure 3).

### *Neuroimaging Biomarkers*

*Structural Imaging: Conventional and Nonconventional Biomarkers* Structural neuroimaging techniques, especially MRI, have been intensely investigated as diagnostic tools to identify patients in the very early stages of AD or in individuals with MCI, who are at high risk of conversion to AD. Great attention has been focused to medial temporal lobe structures, especially the hippocampus and the entorhinal cortex, since these areas are known to be the first to be affected in AD [48]. Some studies have shown that direct (volume of the structures) or indirect measures (such as volume of the temporal horn of the lateral ventricle) of medial temporal lobe involvement, particularly when longitudinal information is available, may be useful as indicators of increased risk of progression from MCI to dementia [49–51]. Even the assessment of medial temporal lobe atrophy by visual rating has been shown to be a significant independent predictor of conversion from MCI, especially when memory is impaired, to AD [52].

In a longitudinal study in which patients with MCI and cognitively normal controls were followed over two years, longitudinal hippocampal volume losses were

associated with increasing hyperphosphorylated tau and decreasing $A\beta_{42}$ CSF levels, suggesting that the increasing atrophy over time of the hippocampal formation might be related to underlying pathological abnormalities that occur in AD [50].

In another study, in which amnestic MCI patients and normal controls were submitted to serial MRI, two quantitative MRI measures, namely whole-brain and ventricle atrophy rates, obtained 1–2 years before baseline were associated with an increased likelihood of conversion from MCI to AD. The combination of these two atrophy rates with cross-sectional hippocampal volume at baseline provided additional predictive information regarding conversion from MCI to AD. Nonetheless, some degree of overlap between converters and nonconverters was observed, indicating that these measures cannot be used as definite prognostic indicators for individual patients [49].

An additional MRI-based method evaluated in prospective studies of MCI is voxel-based morphometry (VBM). In two recent studies using this technique, MCI patients who converted to AD on follow-up (from 18 to 24 months) showed significant reductions in gray matter density in brain regions known to be affected early by AD pathology, such as the hippocampus and posterior cingulate, and also in other neocortical areas in the frontal, temporal, and parietal lobes [53, 54]. MRI spectroscopy seems to be another promising tool for early diagnosis of AD (in the MCI stage), although the studies published so far were all cross-sectional [55, 56].

As previously stated, CIS is considered to be the presymptomatic phase of MS. From a clinical point of view, according to a large database of patients affected by CIS, 52% presented with long-tract symptoms, 18% with a optic neuritis, 9% with a brainstem syndrome, and 21% with multifocal abnormalities [57]. By definition, the CIS phase evolves into MS if there is a demonstration of dissemination in time of inflammatory lesions. Differently from the past when the occurrence of a second clinical relapse resulted in a diagnosis of clinically definite MS [58], the recent McDonald criteria [33] with further revisions [34] consider that the presence of a gadolinium-enhanced lesion on an MRI done at least 3 months after the first episode or a new T2 lesion confirmed on two MRI scans done no less than 3 months apart is sufficient to demonstrate a dissemination in time of lesions and hence a diagnosis of MS. As regards to the sensitivity, specificity, and accuracy of the McDonald criteria in predicting the 3-year conversion to MS, Dalton et al. [59] found sensitivity, specificity, and accuracy of 83%, while Tintore et al. [60] reported respectively figures of 74, 86, and 80%. Therefore, conventional MRI is considered to be extremely important in making MS diagnosis and in helping to reduce the latency between the disease onset and the clinical diagnosis. However, conventional MRI techniques, namely the assessment of the number of T2 lesions, T1 lesions, and gadolinium-enhancing lesions, are of limited value for prognosis in the early phases of the disorder, while nonconventional MRI techniques are more useful.

We can distinguish among different techniques which can be used in the assessment and prognosis of CIS patients. It is now known that brains of MS patients show early axonal pathology that correlates with immune cell damage [61] and that even in a classical neuroinflammatory disorder like MS there is an unexpected collateral damage to neurons, which however do not represent the target of the immune attack, which is the myelin sheet. It is speculated that the extent and size of irreversible axonal loss in CIS patients are predictive of the long-term disability, and the measurement of progressive brain atrophy is a marker of irreversible tissue damage [62].

Dalton et al. [63] performed a longitudinal 3-year follow-up study on 55 CIS patients and demonstrated that patients who progressed into MS (53% of the total) developed more important gray matter atrophy than patients without MS diagnosis. Interestingly enough, the white matter volume did not change over time in both converters and nonconverters, and the number of T2 lesions was moderately correlated with atrophy measures, showing that they are less sensitive and less indicated to longitudinal changes. Another study [64] found a decrease in the normalized brain parenchymal volume and percentage brain volume change of CIS patients. Magnetization transfer (MT) and diffusion-weighted (DW) MRI has the potential to detect and quantify the extent of tissue damage present in the brain and cervical cord, which is not visible by using other techniques. A low MT ratio can be detected in normal-appearing white matter and brain tissue of CIS patients, but it is still unclear whether it can be used as a measure of disability prediction, as only one study found that the measure was predictive of subsequent disease progression [65], while others did not [66, 67].

Proton magnetic resonance spectroscopy can add information on the biochemical nature of changes in the CNS. It has been used in a recent study including 96 patients with CIS who have been tested within 6 months of their clinical episode [68], and high myoinositol and creatine levels have been found in the white matter of diseased patients. It is interesting to notice that a lower ratio of *N*-acetylaspartate and creatine and a lower brain parenchymal fraction have been found in patients with CIS without cognitive impairment versus patients affected by CIS and cognitive impairment, possibly suggesting a more severe axonal loss in the latter group.

Another interesting research area is represented by the MRI of the optic nerve as a model of damage and recovery in demyelination. The best model of demyelination is represented by CIS patients who experience a retrobulbar optic neuritis in which there is a visual loss which is determined by a demyelination of the optic nerve.

*Functional Imaging* As functional neuroimaging methods, particularly functional MRI (fMRI) and PET, became more widely available in recent years, they have been increasingly investigated as potential biomarkers for preclinical or very early AD and MS diagnosis.

In MCI, fMRI studies assessing the risk of conversion to AD are still scarce. However, a recent study in which two carriers of a *PSEN1* mutation were scanned (one presenting subtle memory problems on comprehensive neuropsychological testing and another with amnestic MCI), significant differences related to the pattern of brain activation during an episodic memory task were observed in both subjects in comparison to healthy controls [69]. It is noteworthy that the very mild symptomatic carrier was 20 years old, almost 30 years younger than the mean age of clinical manifestation of the disease in the family. This observation supports further investigation of fMRI as a possible biomarker in very early AD. Moreover, fMRI may also be viewed as a promising tool for monitoring the neurochemical and brain function changes induced by pharmacological agents [70].

PET scan, especially using fluoro-2-deoxy-*d*-glucose (FDG), has proven to be an efficient method for early diagnosis of AD, even prior to the MCI stage. Indeed, in a 3-year longitudinal study enrolling 48 cognitively healthy elderly, from whom 11 declined to MCI, De Leon et al. [71] found that reductions in the metabolic rate in the entorhinal cortex at baseline accurately predicted this conversion. Moreover,

among those subjects who declined, baseline metabolism anticipated the emergence of memory deficits as well as of hippocampal and temporal neocortex hypometabolism. Reduced metabolic rates in additional brain regions, such as the temporoparietal cortex, posterior cingulate, and prefrontal areas, have also shown to be good predictors of conversion from MCI to AD [72–74]. FDG PET was found to be superior to neuropsychological testing as an indicator of subsequent global cognitive decline in patients with MCI [53], although some degree of heterogeneity in the patterns of brain glucose metabolism reduction has been reported in these individuals, a feature that probably will have to be taken into account in future studies [75].

Perhaps the most promising neuroimaging method for very early AD diagnosis is the amyloid PET tracer named Pittsburgh Compound-B [(11)C-PIB], which has been shown to detect in vivo amyloid plaques in the brain of patients with mild AD [76]. Subsequent studies found an inverse correlation between in vivo amyloid load and the concentration of $A\beta_{42}$ in the CSF in AD [77], a positive correlation between whole-brain atrophy rates and whole-brain and regional (11)C-PIB uptake [78] and a negative correlation between amyloid load in the parietal cortex and the performance in a verbal memory test [79]. More interesting and promising, though, is the finding of elevated (11)C-PIB uptake in some nondemented individuals, suggesting that it may turn out to be a sensitive method for detection of preclinical AD in MCI or even in totally asymptomatic individuals [80]. On the other hand, the observation of a relatively stable (11)C-PIB uptake in mild AD patients after a 2-year follow-up suggests that amyloid deposition in the brain tissue reaches a plateau by the early stages of the disease and thus the method may not be ideal for tracking amyloid reduction induced by future treatments [79].

As regards to MS, fMRI studies on CIS patients are extremely useful in assessing the cortical reorganization and plasticity secondary to lesions in the CNS. A 1-year follow-up study of patients with a CIS found that those who developed clinically definite MS had a different response on motor tasks at fMRI than those who did not convert [81]; more specifically, the nonconverters had more substantial activations in the contralateral primary somatomotory cortex, supplementary motor area, ipsilateral paracentral lobule, and cerebellar hemisphere, while MS patients who converted had a more widespread recruitment of additional areas not involved in motor response.

***Predisposition Biomarkers: Susceptibility Genes and Modifiers*** Genetic epidemiology studies have demonstrated that genetic factors are clearly involved in the etiology of AD and MS [82–85]. Despite the fact that only about 25% of AD is familial (FAD), tremendous progress has been made in understanding the etiology of AD by dissecting the molecular genetics of the disease. Mutations in the Aβ precursor protein (β*APP*) coding gene, *PSEN1* [86], and *PSEN2* are associated with early-onset familial AD (EOFAD). Mutations in these genes result in altered processing of βAPP and the relative overproduction of all forms of the β-amyloid peptide or specific overproduction of isoforms ending at residue 42. The relative proportion of EOFAD patients carrying mutations are 10–15% for *APP*, 20–70% for *PSEN1*, and <1% for *PSEN2* [82–84, 87]. Other genes are causing EOFAD as kindreds with autosomal dominant FAD with no known mutations in *PSEN1, PSEN2*, or *APP* have been described [88]. About 1–6% of all AD cases are early onset (<60 years) and about 60% of early-onset AD is familial, with 13% appearing

to be inherited in an autosomal dominant manner [82, 89]. Therefore, the EOFAD group comprises less than 2% of all AD cases [90].

Despite the application of the same molecular tools to MS, identification of specific or definitive MS genes has failed. Association of MS with HLA haplotype within the major histocompatibility complex (MHC) on chromosome 6, extending over a long interval containing over 200 genes, has consistently been reported by several independent studies for more than three decades [91, 92]. However, the vast majority of MS patients do not share the same haplotype, the strength of association is declining from Northern to Southern Europe, and calculations suggest that the MHC could explain between 17 and 62% of the genetic etiology of MS [93]. Genetic data reject a single-locus model of MS susceptibility but suggest that MS depends on independent or epistatic effects of several genes, with a small contribution of each of them. As efforts in genetic research are pursued, it is likely that a limited number of genetic profiles determined by genetic variants in different genes can clearly identify MS patients and be used for presymptomatic diagnosis and selection criteria for treatment. In the meantime, apart from the HLA haplotype, no genetic biomarker can contribute to MS diagnosis in the early stage of the disease.

Several studies have found an increased frequency of the *HLA-A2* allele in patients with EOAD and others have demonstrated an association between the *A2* allele and an earlier age of onset of AD [94], with *HLA-A2* homozygotes having an onset of AD 5 years earlier, on average, than either *A2* heterozygotes or those without *A2*, reflecting a gene dosage effect independent of *APOE4* status. It has been suggested that the *HLA-A2* allele may have a role in regulating an immune response in the pathogenesis of AD or that there may be a responsible gene in close linkage to *A2* [94].

The *APOE4* (Cys112Arg) variant of apolipoprotein E is associated with increased risk for late-onset AD (LOAD, after age 60) in a dose-dependent fashion [95, 96]. The *APOE4* allele, by unknown mechanisms, appears to affect age of onset by shifting the onset curve toward an earlier age [97]. Estimates of the total contribution of *APOE* to the variance in onset of LOAD vary widely. As the inheritance of one or more *APOE4* alleles is not deterministic for AD and the mechanism by which inheritance of one or more *APOE4* alleles causes AD is unidentified, the usefulness of *APOE* genotyping in clinical diagnosis and risk assessment remains unclear [98]. *APOE* genotyping may have an adjunct role in the diagnosis of AD because a large proportion of demented individuals with *APOE4* alleles have been found to have neuropathological confirmation of AD at autopsy [31, 32, 99, 100]. Increased risk of LOAD associated with the *APOE4* allele has also been shown in African-Americans [101] and Caribbean Hispanics [102]. In a segregation analysis of LOAD families, Daw et al. [103] found evidence that, in addition to *APOE*, four additional loci make a contribution to the variance in age at onset of LOAD similar to or greater in magnitude than that made by *APOE* alone, with one locus making a contribution several times greater than that of *APOE*. Their results suggested that several genes not yet localized to that time may play a larger role than does *APOE* in LOAD.

No result indicates that homozygosity or heterozygosity for *APOE4* allele causes a greater risk of developing MS, but there are several studies which found an association between the allele *APOE4* and a faster disease progression and also two longitudinal studies which found a relationship between the allele and the progression rate. A recent meta-analysis confirmed such data [104].

Numerous candidate gene studies have been conducted both in MS and AD to identify additional disease modifiers. Some positive findings have been reported, but modestly significant associations were observed, suggesting that the studied genetic variants have only modest effects on risk/protection, age at onset, and disease duration. This observation reinforces the existence of not yet identified additional putative functional variants and demonstrates that none of the investigated single variants taken alone contributed substantially to the development of AD or MS in the respective studied samples. Among the analyzed genes, the following can be cited: *UBQLN1* [105], *APBB2* [106], *TNF alpha* [107], *LRP1, ACE, A2M, BLMH, DLST, TNFRSF6, NOS3, PSEN1, PSEN2, BCHE, APBB1, ESR1, CTSD, MTHFR*, and *IL1A* [108]. Croes et al. [109] argued against using genetic testing for AD as a diagnostic tool, suggesting that the contribution of genetic testing to clinical diagnosis is small and does not counterbalance the problems associated either with interpretation or with secondary effects on family members. In contrast to the utility of *APOE* testing as an adjunct diagnostic test in individuals with dementia, there is general agreement that *APOE* testing should not be used for predictive testing for AD in asymptomatic persons.

***Soluble Biomarkers: Analysis of Body Fluids*** Biological analyses of body fluids such as CSF, blood, and urine samples have the potential to deliver important biomarkers in the clinic due to the accessibility of these biological materials. However, lumbar puncture is a relatively invasive procedure and repeated collection of CSF needs to be restricted, representing a limitation for its use in the clinical setting and also as a potential surrogate marker in clinical trials. Blood–brain barrier damage, both in MS and AD, allows relatively large amounts of CSF to be absorbed in the blood every day [110], but the main challenge resides in the relevance of brain leakage proteins likely to be present in CSF and/or plasma in poor to very poor levels in comparison to abundant proteins such as albumin. This is due to the fact that peripheral blood represents a traffic compartment for both regulatory and effector lymphomonuclear cells, and it is very unlikely that the processes occurring in the CNS would be reflected and paralleled in the peripheral circulation [111]. Moreover, soluble molecules, which could play a relevant role in the ethiopathogenesis of MS and AD, act in an autocrine or paracrine fashion, which typically show a short half-life [112].

*Common Features between MS and AD* CSF levels of protein tau (total tau) have been shown to be significantly increased in patients with AD, with elevations of two- to threefold CSF total tau, but weak correlations present with changes in cognitive scores [113, 114]. In MS, increased total tau levels in patients with active disease compared to clinically nonactive states indicate axonal pathology in active disease [115], and the report was confirmed in a second study [116]. The association of tau concentrations with AD suggests that the protein could be related to cognitive dysfunction in MS. Over the last 20 years, several studies have confirmed the association of MS with the HLA complex, the most polymorphic gene cluster known in humans with >200 genes, but the MS-associated DR2 haplotype, also known as HLA-DRB1*1501-DQA1*0102-DQB1*0602, extends over a long interval containing multiple genes [93]. Most recent studies based on linkage disequilibrium analysis with highly dense single-nucleotide polymorphism (SNP) maps have demonstrated

that it is the HLA class II genotype patterns that determine susceptibility and resistance to MS, this pattern of susceptibility strongly supporting an autoimmune etiology for MS [117, 118]. *HLA-A2* allele association with earlier age of AD onset has been reported with an additive effect of *HLA-A2* and *APOE4* demonstrating a 3-year shift [94, 119]. Although sensitivity is high, poor specificity limits the diagnostic value of these biomarkers.

*Specific Features of AD* Aβ is recognized as a normal product of cellular metabolism throughout life and circulates as a soluble peptide in biological fluids. The most reproduced finding is the decrease of CSF $A\beta_{1-42}$ levels [28, 120, 121] in association with higher levels of total tau in the CSF of AD patients in comparison with the CSF of controls while CSF $A\beta_{1-40}$ levels (or $A\beta_{total}$) are not different from age-matched controls. The high positive likelihood ratio indicates that combined biomarker tests are useful in confirming the diagnosis of AD, but the low negative likelihood ratio indicates that a negative test result cannot rule out the disease [122]. Changes in CSF biomarkers occur early in the course of AD in most patients. No consistent correlation of $A\beta_{1-40}$ or $A\beta_{1-42}$ plasma concentrations with CSF levels has been observed [123]. Based on CSF levels of $A\beta_{42}$, tau, and ubiquitin, subgroups presenting different clinical profiles can be distinguished; these different subgroups may benefit from different therapeutic drugs [124].

Analysis of the different CSF phosphorylated tau isoform (p-tau 181, 199, 231, 235, 396, and 404) levels revealed that CSF tau-199 (phosphorylation at serine 199) is significantly increased in AD patients compared to non-AD subjects, exceeding both sensitivity and specificity over 85% as a sole biomarker for AD [125], but with many of the non-AD tauopathy and degenerative dementias also showing increased tau-199 levels. Cross-sectional studies examining three p-tau species as diagnostic measures (tau-181, tau-199, and tau-231) have revealed that tau-231 may have a greater specificity to AD [126]. The observed correlations between levels of CSF phosphorylated tau and hippocampal atrophy are independent of disease duration and severity, suggesting that CSF phosphorylated tau levels may reflect neuronal damage in AD [127].

Combining CSF tau and CSF $A\beta_{1-42}$ levels improves specificity, allowing differentiation between AD, normal aging, and vascular dementia [128]. Other biological markers such as CSF levels of neurotransmitters, cytokines, or superoxide dismutase were shown to have less diagnostic value [129].

*Specific Features of MS* As regards biomarkers related to the inflammatory phase, it is well known that the detection in the CSF of an intrathecal antibody production, namely the presence of two or more oligoclonal IgG bands (OB) detected by separation of CSF proteins, while not demonstrable in corresponding serum, is supportive of a diagnosis of MS. Several studies suggest that CSF analysis increases diagnostic sensitivity, though perhaps at the cost of specificity and accuracy [130]. However, findings in one large study of PPMS [131] have led to a recommendation to revise the CSF criteria for this population of patients, and it is no longer a requirement for diagnosis [34]. MS patients with and without CSF bands may be clinically indistinguishable but may represent two immunologically distinct populations [132].

Other components of the immune system, and mostly cytokine and chemokine blood levels, have been extensively studied as potential biomarkers with

disappointing results, with only few studies trying to test more than one potential biomarker at the same time in the same subjects [133].

Given the recent evidence that axonal loss occurs in MS since the early stages of the disease, there is a current strong interest in the identification and analysis of biomarkers which reflect the axonal damage, which can be divided into cytoskeleton markers, membrane markers, and other axonal markers [134]. Several studies have shown that concentration of light-chain neurofilament (NfL), one of the major axonal cytoskeleton proteins, is increased in CSF of patients with MS [135], with peak concentrations during relapses. In addition, the correlation of NfL and heavy-chain neurofilament (NfH) autoantibody indices with surrogate markers for atrophy on MRI suggests a relation with disease progression [136]. Combining CSF markers (tau and NfH) can increase sensitivity in predicting the conversion from CIS to MS (40% compared with 34% for MRI), which could be further increased to 60% if CSF and MRI criteria were combined [137]. Similarly, the combination of tau and NfH showed higher specificity (94%) than MRI (82%). The only marker related to neuronal damage that has shown promising results in peripheral blood is 24*S*-hydroxycholesterol [138], with concentrations decreased in serum of patients with MS, especially if affected by the primary progressive form of MS. Additional markers of axonal loss which have been tested are the ApoE levels, the 14–3–3 protein, and the neuron-specific enolase with, however, conflicting results.

## 3 CLINICAL DEVELOPMENT APPLICATION: DRUG EFFICACY AND SAFETY

### 3.1 Using Biomarkers to Select Patient Subgroups

Biomarkers may be applied to drug development in a number of distinct ways [129]. First, additional diagnostic measures in clinically identified populations are provided (high specificity and sensitivity are therefore required). Second, new biomarkers of presymptomatic disease are very important for population enrichment strategies and confirmation of efficacy during the assessment of novel disease-modifying therapies. Third, biomarkers can help to establish a differential therapeutic approach and to treat individuals selectively according to their pathophysiological/genetic subtype and disease status (Table 1). In clinical development, the most important benefit offered by biomarkers is to limit investigational drugs to high-risk groups, who would gain the most of novel therapies and/or in which efficacy is most likely to be observed. Heterogeneity in the patient population can be limited by subgrouping patients according to these additional biomarker criteria. In the absence of subgrouping patients, dilution effect on efficacy may lead to no statistically significant results and arrest of drug development of molecule even if the molecule may be valuable for some patient subgroups. Enrichment trial designs using the same methodology as in cancer drug development can be extremely valuable strategy, especially for unprecedented targets [139]. The improved accuracy in diagnosis with additional diagnostic measurements can lead to better chance of success in proof-of-principle/proof-of-concept (POP/POC) studies, allowing us to concentrate our efforts on successful POP/POC studies to continue to late-phase clinical development and, therefore, to speed up the testing of novel treatments. The reduction in

**TABLE 1 Different Types of Biomarkers**

| | |
|---|---|
| Predisposition biomarkers | To use in families where risk of developing disease is high and/or age dependent, such as in AD families with mutations in APP, PSEN1, PSEN2, the penetrance may not be 100%. |
| Screening biomarkers | To be used for screening in high-risk presymptomatic populations—mutations in APP, PSEN1, and PSEN2 are too rare to justify the effort and expense, and ApoE is not sufficiently strong to be used in presymptomatic populations. |
| Diagnosis biomarkers | Neuroimaging biomarkers and oligoclonal bands for MS. In AD, neuroimaging biomarkers are the most promising, ApoE being used as an adjunct. |
| Therapy selection biomarkers | To use to define targeted populations where disease-modifying therapies have been efficacious (targeted therapies). |
| Prognosis biomarkers | Once diagnosis is performed, markers defining disease severity and/or rate of progression. |
| Predictive biomarkers | To anticipate efficacy and/or safety of symptomatic and/or disease-modifying neurotherapies. |

sample size estimates is also very important: Enrollment of fewer subjects is needed if using neuroimaging and soluble biomarkers rather than when using cognitive measurements (for AD) or clinical criteria (for MS) [39, 140].

From a broader clinical standpoint, redefining disease is the challenge of future medicine. Because of heterogeneity in MS and in AD, subtyping patients by genetic, clinical, neuroradiological, and neuroimmunological parameters will be necessary, and biomarkers will be extremely important contributors to the subtyping algorithm.

### 3.2 Predicting Changes in Response to Therapy and Disease Course: Predictive Biomarkers

As new targets have emerged from genomics, it has been put forward that insufficient target validation has caused the increase in failure rates of drug development programs. As a consequence, experimental medicine groups frequently develop and validate biomarkers to enable them to test if the pharmacology can be safely expressed in humans and if the pharmacological activity provides sufficient efficacy and differentiation in the disease indication [141].

A number of strategic considerations are addressed in the course of drug development and questions frequently emerge: Why and how to use biomarkers in clinical drug development? With rare exceptions, biomarkers cannot substitute primary endpoints in drug registration trials and, by definition, a surrogate can only approximate the gold standard. Biomarkers may be applied to drug development in a number of distinct ways [129]. If deployed as early as possible, biomarkers can contribute to the development of three phases of the clinical plan: (1) to confirm if the drug is hitting the target; (2) to test if hitting the target alters the pathophysiological mechanism of disease (POP/POC); and (3) to test if altering the mechanism affects clinical status [142].

Consequently, classes of biomarkers differ not only for each therapeutic area but also according to their objectives. In risk reduction, specific biomarkers may be developed for selecting drug dose, for selecting patient population, for assessing early efficacy, or for predicting toxicity, making some biomarkers suited for one

role and inappropriate for another. Specificity and sensitivity required will depend on the application. Clearly, they are add-ons to facilitate decision making and not a substitute for clinical endpoints requested by regulatory authorities (FDA or EMEA) [143].

Development and validation of biomarkers can be cost effective in drug development programs, aiming for high-risk unprecedented targets (such as biologicals) coming from genomics discoveries and carrying increased risk in early development, as well as in chronic diseases and in pre–phase II studies to separate more quickly and cheaply than traditional phase II studies. In theses instances, biomarker studies contribute not necessarily to go–no go drug development decisions but rather to gaining confidence in a novel compound and promoting it for further development, keeping in mind that validation and reliability of biomarkers are frequently difficult points and that safety and efficacy biomarker data are to be weighed against "traditional" safety and efficacy data. In parallel, for the measurement of disease severity, observations of changes in biomarker with disease progression are to be observed in longitudinal studies (high specificity not specially required in this case).

## 4 CONCLUSION

The development and validation of diagnostic and predictive biomarkers in diseases with usually extended asymptomatic phases, like MS and AD, are particularly challenging and can take a long time as their validation is by necessity often linked to long-term clinical outcomes. Nonetheless, such biomarkers could have a big impact on health care by facilitating the assessment of new drugs that prevent costly chronic progressive disease [144]. Progress in identifying diagnostic biomarkers has been rate limited due to incomplete understanding of the pathogenesis of these disorders despite the advances in the identification of genes associated with inherited risk. Biomarkers for AD are far from reaching clinical practice while MRI and CSF OB bands are the only validated biomarkers currently used in MS. Yet, if validated diagnostic biomarkers as well as disease-modifying treatments with ability to slow down the disease progression became available—as the first Aβ modulators in AD and interferons in MS—patients would highly benefit from management of the early disease stage.

Biological and neuroimaging biomarkers with strong potential are currently being used in early drug development programs to allow for the exploration of pharmacodynamic effects of targeted therapeutics and to aid early diagnosis, alone or in combination, but require additional validation to be considered for their ability to reveal effect of investigational drugs [11, 12, 145]. CSF Aβ and tau are biological markers of AD pathophysiology, and their measures may have potential clinical utility in the future diagnosis of AD. They might be combined with neuroimaging techniques, and the PET scan with (11)C-PIB tracer is perhaps the most promising candidate [77]. Combinations of biomarkers hold great promise: For example, tau and NfH are valuable biomarkers for axonal damage in CIS patients. Predicting conversion from CIS to MS can be improved if CSF markers are combined with MRI [137]. However, soluble biological biomarkers show considerable variance, with significant overlap between values obtained in disease groups and controls for AD or MS. Meta-analyses of previous studies comparing these markers need to demonstrate similar findings. For most biomarkers, the differences are not sensitive

and specific enough for discrimination between disease and control groups or prediction of disease outcome in individual patients. Few studies on specific markers for neurodegeneration in body fluids have been done. The studies have typically included only small numbers of patients, and, more importantly for monitoring of neurological decline, follow-up needs to be performed on longer time periods. Validation of biomarkers requires longitudinal studies to collect data on disease progression and on neuropathology, especially for AD.

The search for biomarkers in MS and AD has just started and is already yielding promising results. Intensive research efforts are currently being accomplished using mass spectrometry–based proteomics as well as other promising technologies to provide a deep coverage of the body fluid proteomes. In the immediate future, emerging biomarkers for MS and AD are more likely to be used in clinical trials than in clinical practice, but undoubtedly, some of them will bring interesting diagnostic and prognostic tools to molecular medicine [146].

## ACKNOWLEDGMENTS

The authors would like to thank Jose Maria Badenas, MD, Director, Medical & Scientific Services, Quintiles, for his insight and editorial comments.

## REFERENCES

1. Hauser, S. L., and Goodin, D. S. (2005), Multiple sclerosis and other demyelinating diseases, in Kasper, D. L., Braunwald, E., Fauci, A. D., Hauser, S. L., Longo, D. L., and Jameson, J. L., Eds., *Harrison's Principle of Internal Medicine*, 16th ed., Mc-Graw Hill, New York.
2. Khachaturian, Z. S. (1985), Diagnosis of Alzheimer's disease, *Arch. Neurol.*, 42, 1097–1105.
3. Khachaturian, Z. S. (2006), Diagnosis of Alzheimer's disease: Two-decades of progress, *J. Alzheimer's Dis.*, 9, 409–415.
4. McKhann, G., Drachman, D., Folstein, M., et al. (1984), Clinical diagnosis of Alzheimer's disease: Report of the NINCDS-ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease, *Neurology*, 34, 939–944.
5. Lue, L. F., Rydel, R., Brigham, E. F., et al. (2001), Inflammatory repertoire of Alzheimer's disease and nondemented elderly microglia in vitro, *Glia*, 35, 72–79.
6. Akiyama, H., Barger, S., Barnum, S., et al. (2000), Inflammation and Alzheimer's disease, *Neurobiol. Aging*, 21, 383–421.
7. Mrak, R. E., and Griffinbc, W. S. (2001), The role of activated astrocytes and of the neurotrophic cytokine S100B in the pathogenesis of Alzheimer's disease, *Neurobiol. Aging*, 22, 915–922.
8. Zipp, F., and Aktas, O. (2006), The brain as a target of inflammation: common pathways link inflammatory and neurodegenerative diseases, *Trends Neurosci.*, 29, 518–527.
9. Watson, G. S., and Craft, S. (2006), Insulin resistance, inflammation, and cognition in Alzheimer's disease: Lessons for multiple sclerosis, *J. Neurol. Sci.*, 24521–24533.
10. Kola, I., and Landis, J. (2004), Can the pharmaceutical industry reduce attrition rate? *Nat. Rev. Drug Discov.*, 3, 711–715.

11. Matthews, P. M., and Wise, R. G. (2006), Noninvasive brain imaging for experimental medicine in drug development, *expert Opin. Drug Discov.*, 1(2), 111–121.
12. Frank, R. A., Galasko, D., Hampel, H., et al. (2003), National Institute on Aging Biological Markers Working Group. Biological markers for therapeutic trials in Alzheimer's disease. Proceedings of the biological markers working group; NIA initiative on neuroimaging in Alzheimer's disease, *Neurobiol. Aging*, 24, 521–536.
13. Rolan, P., Atkinson, A. J. Jr., and Lesko, L. J. (2003), Scientific Organizing Committee; Conference Report Committee. Use of biomarkers from drug discovery through clinical practice: Report of the Ninth European Federation of Pharmaceutical Sciences Conference on Optimizing Drug Development, *Clin. Pharmacol. Ther.*, 73, 284–291.
14. Tardif, J. C., Heinonen, T., Orloff, D., et al. (2006), Vascular biomarkers and surrogates in cardiovascular disease, *Circulation*, 113, 2936–2942.
15. White, T. J., Clark, A. G., and Broder, S. (2006), Genome-based biomarkers for adverse drug effects, patient enrichment and prediction of drug response, and their incorporation into clinical trial design, *Personalized Med.*, 3(2), 177–185.
16. Atkinson, A. J. Jr., Biomarkers Definitions Working Group (2001), Biomarkers and surrogate endpoints: Preferred definitions and conceptual framework, *Clin. Pharmacol. Ther.*, 69, 89–95.
17. Glenner, G. G., and Wong, C. W. (1984), Alzheimer's disease: Initial report of the purification and characterization of a novel cerebrovascular amyloid protein, *Biochem. Biophys. Res. Commun.*, 120, 885–890.
18. Selkoe, D. J. (1991), The molecular pathology of Alzheimer's disease, *Neuron*, 6(4), 487–498.
19. Caramelli, P., Robitaille, Y., Laroche-Cholette, A., et al. (1998), Structural correlates of cognitive deficits in a selected group of patients with Alzheimer's disease, *Neuropsychiatry Neuropsychol. Behav. Neurol.*, 11, 184–190.
20. Knopman, D. S., DeKosky, S. T., Cummings, J. L., et al. (2001), Practice parameter: Diagnosis of dementia (an evidence-based review). Report of the Quality Standards Subcommittee of the American Academy of Neurology, *Neurology*, 56, 1143–1153.
21. De Leon, M. J., George, A. E., Golomb, J., et al. (1997), Frequency of hippocampal formation atrophy in normal aging and Alzheimer's disease, *Neurobiol. Aging*, 18, 1–11.
22. Bottino, C. M., Castro, C. C., Gomes, R. L., et al. (2002), Volumetric MRI measurements can differentiate Alzheimer's disease, mild cognitive impairment, and normal aging, *Int. Psychogeriatr.*, 14, 59–72.
23. Gosche, K. M., Mortimer, J. A., Smith, C. D., et al. (2002), Hippocampal volume as an index of Alzheimer neuropathology: Findings from the Nun Study, *Neurology*, 58, 1476–1482.
24. Bresciani, L., Rossi, R., Testa, C., et al. (2005), Visual assessment of medial temporal atrophy on MR films in Alzheimer's disease: Comparison with volumetry, *Aging Clin. Exp. Res.*, 17, 8–13.
25. Wahlund, L. O., Julin, P., Johansson, S. E., et al. (2000), Visual rating and volumetry of the medial temporal lobe on magnetic resonance imaging in dementia: A comparative study, *J. Neurol. Neurosurg. Psychiatry*, 69, 630–635.
26. Herholz, K., Salmon, E., Perani, D., et al. (2002), Discrimination between Alzheimer dementia and controls by automated analysis of multicenter FDG PET, *Neuroimage*, 17, 302–316.
27. Mosconi, L. (2005), Brain glucose metabolism in the early and specific diagnosis of Alzheimer's disease. FDG-PET studies in MCI and AD, *Eur. J. Nucl. Med. Mol. Imaging*, 32, 486–510.

28. Sunderland, T., Linker, G., Mirza, N., et al. (2003), Decreased β-amyloid1-42 and increased tau levels in cerebrospinal fluid of patients with Alzheimer disease, *JAMA*, 289, 2094–2103.
29. Blennow, K. (2004), Cerebrospinal fluid protein biomarkers in Alzheimer's disease, *NeuroRx*, 1, 213–225.
30. Mayeux, R., Honig, L. S., Tang, M. X., et al. (2003), Plasma A[beta]40 and A[beta]42 and Alzheimer's disease: Relation to age, mortality, and risk, *Neurology*, 61, 1185–1190.
31. Bétard, C., Robitaille, Y., Gee, M., et al. (1994), Apo E allele frequencies in Alzheimer's disease, Lewy body dementia, Alzheimer's disease with cerebrovascular disease and vascular dementia, *Neuroreport*, 5, 1893–1896.
32. Mayeux, R., Saunders, A. M., Shea, S., et al. (1998), Utility of the apolipoprotein E genotype in the diagnosis of Alzheimer's disease. Alzheimer's Disease Centers Consortium on Apolipoprotein E and Alzheimer's Disease, *N. Engl. J. Med*, 338, 506–511.
33. McDonald, W. I., Compston, A., Edan, G., et al. (2001), Recommended diagnostic criteria for multiple sclerosis: Guidelines from the International Panel on the diagnosis of multiple sclerosis, *Ann. Neurol.*, 50, 121–127.
34. Polman, C. H., Reingold, S. C., and Edan, G. (2005), Diagnostic criteria for multiple sclerosis: 2005 revisions to the "McDonald Criteria," *Ann. Neurol.*, 58, 840–846.
35. Weinshenker, B. G., Bass, B., Rice, G. P., et al. (1989), The natural history of multiple sclerosis: A geographically based study. 2. Predictive value of the early clinical course, *Brain*, 112, 1419–1428.
36. Comi, G., Filippi, M., Martinelli, V., et al. (1995), Brain MRI correlates of cognitive impairment in primary and secondary progressive multiple sclerosis, *J. Neurol. Sci.*, 132, 222–227.
37. Kujala, P., Portin, R., and Ruutiainen, J. (1997), The progress of cognitive decline in multiple sclerosis. A controlled 3-year follow-up, *Brain*, 120, 289–297.
38. Nocentini, U., Pasqualetti, P., Bonavita, S., et al. (2006), Cognitive dysfunction in patients with relapsing-remitting multiple sclerosis, *Mult. Scler.*, 12, 77–87.
39. Brass, S. D., Benedict, R. H., Weinstock-Guttman, B., et al. (2006), Cognitive impairment is associated with subcortical magnetic resonance imaging grey matter T2 hypointensity in multiple sclerosis, *Mult. Scler.*, 12, 437–444.
40. Parmenter, B. A., Denney, D. R., Lynch, S. G., et al. (2007), Cognitive impairment in patients with multiple sclerosis: Association with the APOE gene and promoter polymorphisms, *Mult. Scler.*, 13, 25–32.
41. Knopman, D. S., Parisi, J. E., Salviati, A., et al. (2003), RC2003. Neuropathology of cognitively normal elderly, *J. Neuropathol. Exp. Neurol.*, 62, 1087–1095.
42. Bennett, D. A., Schneider, J. A., Arvanitakis, Z., et al. (2006), Neuropathology of older persons without cognitive impairment from two community-based studies, *Neurology*, 66, 1837–1844.
43. Miller, D. H., Ormerod, I. E., Rudge, P., et al. (1989), The early risk of multiple sclerosis following isolated acute syndromes of the brainstem and spinal cord, *Ann. Neurol.*, 26, 635–639.
44. Petersen, R. C. (2004), Mild cognitive impairment as a diagnostic entity, *J. Intern. Med.*, 256, 183–194.
45. Bielekova, B., and Roland, M. (2004), Development of biomarkers in multiple sclerosis, *Brain*, 127, 1463–1478.
46. Morris, J. C., Kimberly, A., Quaid, K., et al. (2005), Role of biomarkers in studies of presymptomatic Alzheimer's disease, *Alzheimer's Dementia*, 145–151.

47. Chong, M. S., Lim, W. S., and Sahadevan, S. (2006), Biomarkers in preclinical Alzheimer's disease, *Curr. Opin. Investig Drugs*, 7, 600–607.
48. Braak, H., and Braak, E. (1991), Neuropathological staging of Alzheimer-related changes, *Acta Neuropathol.*, 82, 239–259.
49. Jack, C. R. Jr., Shiung, M. M., Weigand, S. D., et al. (2005), Brain atrophy rates predict subsequent clinical conversion in normal elderly and amnestic MCI, *Neurology*, 65, 1227–1231.
50. De Leon, M. J., DeSanti, S., Zinkowski, R., et al. (2001), Longitudinal CSF and MRI biomarkers improve the diagnosis of mild cognitive impairment, *Neurobiol. Aging*, 27, 394–401.
51. Erten-Lyons, D., Howieson, D., Moore, M. M., et al. (2006), Brain volume loss in MCI predicts dementia, *Neurology*, 66, 233–235.
52. Geroldi, C., Rossi, R., Calvagna, C., et al. (2006), Medial temporal atrophy but not memory deficit predicts progression to dementia in patients with mild cognitive impairment, *J. Neurol. Neurosurg. Psychiatry*, 77, 1219–1222.
53. Chetelat, G., Landeau, B., Eustache, F., et al. (2005), Using voxel-based morphometry to map the structural changes associated with rapid conversion in MCI: A longitudinal MRI study. *Neuroimage*, 27, 934–946.
54. Bozzali, M., Filippi, M., Magnani, G., et al. (2006), The contribution of voxel-based morphometry in staging patients with mild cognitive impairment, *Neurology*, 67, 453–460.
55. Kantarci, K., Xu, Y., Shiung, M. M., et al. (2002), Comparative diagnostic utility of different MR modalities in mild cognitive impairment and Alzheimer's disease. *Dement. Geriatr. Cogn. Disord.*, 14, 198–207.
56. Falini, A., Bozzali, M., Magnani, G., et al. (2005), A whole brain MR spectroscopy study from patients with Alzheimer's disease and mild cognitive impairment, *Neuroimage*, 26, 1159–1163.
57. Confavreux, C., Vukusic, S., Moreau, T., et al. (2000), Relapses and progression of disability in multiple sclerosis, *N. Engl. J. Med.*, 343, 1430–1438.
58. Poser, C. M., Paty, D. W., Scheinberg, L., et al. (1983), New diagnostic criteria for multiple sclerosis: Guidelines for research protocols, *Ann. Neurol.*, 13, 227–231.
59. Dalton, C. M., Brex, P. A., Miszkiel, K. A., et al. (2002), Application of the new McDonald criteria to patients with clinically isolated syndromes suggestive of multiple sclerosis, *Ann. Neurol.*, 52, 47–53.
60. Tintore, M., Rovira, A., Rio, J., et al. (2003), New diagnostic criteria for multiple sclerosis: Application in first demyelinating episode, *Neurology*, 60, 27–30.
61. Trapp, B. D., Peterson, J., Ransohoff, R. M., et al. (1998), Axonal transection in the lesions of multiple sclerosis, *N. Engl. J. Med.*, 338, 278–285.
62. Miller, D. H., Barkhof, F., Frank, J. A., et al. (2002), Measurement of atrophy in multiple sclerosis: Pathological basis, methodological aspects and clinical relevance, *Brain*, 125, 1676–1695.
63. Dalton, C. M., Brex, P. A., Jenkins, R., et al. (2002), Progressive ventricular enlargement in patients with clinically isolated syndromes is associated with the early development of multiple sclerosis, *J. Neurol. Neurosurg Psychiatry*, 73, 141–147.
64. Filippi, M., Rovaris, M., Inglese, M., et al. (2004), Interferon beta-1a for brain tissue loss in patients at presentation with syndromes suggestive of multiple sclerosis: A randomised, double-blind, placebo-controlled trial, *Lancet*, 364(9444), 1489–1496.
65. Iannucci, G., Tortorella, C., Rovaris, M., et al. (2000), Prognostic value of MR and magnetization transfer imaging findings in patients with clinically isolated syndromes suggestive of multiple sclerosis at presentation, *Am. J. Neuroradiol.*, 21, 1034–1038.

66. Kaiser, J. S., Grossman, R. I., Polansky, M., et al. (2000), Magnetization transfer histogram analysis of monosymptomatic episodes of neurologic dysfunction: preliminary findings, *Am. J. Neuroradiol.*, 21, 1043–1047.
67. Brex, P. A., Leary, S. M., Plant, G. T., et al. (2001), Magnetization transfer imaging in patients with clinically isolated syndromes suggestive of multiple sclerosis. *Am. J. Neuroradiol.*, 22, 947–951.
68. Fernando, K. T., McLean, M. A., Chard, D. T., et al. (2004), Elevated white matter myo-inositol in clinically isolated syndromes suggestive of multiple sclerosis, *Brain*, 127, 1361–1369.
69. Mondadori, C. R., Buchmann, A., Mustovic, H., et al. (2006), Enhanced brain activity may precede the diagnosis of Alzheimer's disease by 30 years, *Brain*, 129, 2908–2922.
70. Goekoop, R., Scheltens, P., Barkhof, F., et al. (2006), Cholinergic challenge in Alzheimer patients and mild cognitive impairment differentially affects hippocampal activation—a pharmacological fMRI study, *Brain*, 129, 141–157.
71. De Leon, M. J., Convit, A., Wolf, O. T., et al. (2001), Prediction of cognitive decline in normal elderly subjects with 2-[(18)F]fluoro-2-deoxy-D-glucose/poitron-emission tomography (FDG/PET), *Proc. Natl. Acad. Sci.*, 98, 10966–10971.
72. Chetelat, G., Desgranges, B., de la Sayette, V., et al. (2003), Mild cognitive impairment: Can FDG-PET predict who is to rapidly convert to Alzheimer's disease? *Neurology*, 60, 1374–1377.
73. Drzezga, A., Lautenschlager, N., Siebner, H., et al. (2003), Cerebral metabolic changes accompanying conversion of mild cognitive impairment into Alzheimer's disease: A PET follow-up study, *Eur. J. Nucl. Med. Mol. Imaging*, 30, 1104–1113.
74. Mosconi, L., Perani, D., Sorbi, S., et al. (2004), MCI conversion to dementia and the APOE genotype: A prediction study with FDG-PET, *Neurology*, 63, 2332–2340.
75. Anchisi, D., Borroni, B., Franceschi, M., et al. (2005), Heterogeneity of brain glucose metabolism in mild cognitive impairment and clinical progression to Alzheimer disease, *Arch. Neurol.*, 62, 1728–1733.
76. Klunk, W. E., Engler, H., Nordberg, A., et al. (2004), Imaging brain amyloid in Alzheimer's disease with Pittsburgh Compound-B, *Ann. Neurol.*, 55, 306–319.
77. Fagan, A. M., Mintun, M. A., Mach, R. H., et al. (2006), Inverse relation between in vivo amyloid imaging load and cerebrospinal fluid Abeta42 in humans, *Ann. Neurol.*, 59, 512–519.
78. Archer, H. A., Edison, P., Brooks, D. J., et al. (2006), Amyloid load and cerebral atrophy in Alzheimer's disease: An 11C-PIB positron emission tomography study, *Ann. Neurol.*, 60, 145–147.
79. Engler, H., Forsberg, A., Almkvist, O., et al. (2006), Two-year follow-up of amyloid deposition in patients with Alzheimer's disease, *Brain*, 9, 2856–2866.
80. Mintun, M. A., Larossa, G. N., Sheline, Y. I., et al. (2006), [11C]PIB in a nondemented population: Potential antecedent marker of Alzheimer disease, *Neurology*, 67, 446–452.
81. Rocca, M. A., Mezzapesa, D. M., Ghezzi, A., et al. (2005), A widespread pattern of cortical activations in patients at presentation with clinically isolated symptoms is associated with evolution to definite multiple sclerosis, *Am. J. Neuroradiol.*, 26, 1136–1139.
82. Campion, D., Dumanchin, C., Hannequin, D., et al. (1999), Early-onset autosomal dominant Alzheimer disease: prevalence, genetic heterogeneity, and mutation spectrum, *Am. J. Hum. Genet.*, 65, 664–670.
83. Sherrington, R., Froelich, S., Sorbi, S., et al. (1996), Alzheimer's disease associated with mutations in presenilin 2 is rare and variably penetrant, *Hum. Mol. Genet.*, 5, 985–988.
84. Janssen, J. C., Beck, J. A., Campbell, T. A., et al. (2003), Early onset familial Alzheimer's disease: Mutation frequency in 31 families, *Neurology*, 60, 235–239.

85. Dyment, D. A., Ebers, G. C., and Sadovnick, A. D. (2004), Genetics of multiple sclerosis. *Lancet Neurol.*, 3, 104–110.
86. Larner, A. J., and Doran, M. (2006), Clinical phenotypic heterogeneity of Alzheimer's disease associated with mutations of the presenilin-1 gene, *J. Neurol.*, 253, 139–158.
87. Campion, D., Flaman, J. M., Brice, A., et al. (1995), Mutations of the presenilin I gene in families with early-onset Alzheimer's disease, *Hum. Mol. Genet.*, 4, 2373–2377.
88. Cruts, M., van Duijn, C. M., Backhovens, H., et al. (1998), Estimation of the genetic contribution of presenilin-1 and -2 mutations in a population-based study of presenile Alzheimer disease, *Hum. Mol. Genet.*, 7, 43–51.
89. Rocca, W. A., Hofman, A., Brayne, C., et al. (1991), Frequency and distribution of Alzheimer's disease in Europe: A collaborative study of 1980–1990 prevalence findings. The EURODEM-Prevalence Research Group, *Ann. Neurol.*, 30, 381–390.
90. Bird, T. D., Sumi, S. M., Nemens, E. J., et al. (1989), Phenotypic heterogeneity in familial Alzheimer's disease: A study of 24 kindreds, *Ann. Neurol.*, 25, 12–25.
91. Jersild, C., Svejgaard, A., and Fog, T. (1972), HLA antigens and multiple sclerosis, *Lancet*, 1(7762), 1240–1241.
92. Peltonen, L., Saarela, J., and Kuokkanen, S. (2002), *Multiple Sclerosis. The Genetic Basis of Common Diseases*, 2nd ed., Oxford Monographs on Medical Genetics, New York: Oxford University Press, pp. 805–817.
93. Haines, J. L., Terwedow, H. A., Burgess, K., et al. (1998), Linkage of the MHC to familial multiple sclerosis suggests genetic heterogeneity. The Multiple Sclerosis Genetics Group, *Hum. Mol. Genet.*, 7, 1229–1234.
94. Zareparsi, S., James, D. M., Kaye, J. A., et al. (2002), HLA-A2 homozygosity but not heterozygosity is associated with Alzheimer disease. *Neurology*, 58, 973–975.
95. Corder, E. H., Saunders, A. M., Strittmatter, W. J., et al. (1993), Gene dose of apolipoprotein E type 4 allele and the risk of Alzheimer's disease in late onset families, *Science*, 261, 921–923.
96. Strittmatter, W. J., Saunders, A. M., Schmechel, D., et al. (1993), Apolipoprotein E: High-avidity binding to beta-amyloid and increased frequency of type 4 allele in late-onset familial Alzheimer disease, *Proc. Natl. Acad. Sci. U.S.A.*, 90, 1977–1981.
97. Meyer, M. R., Tschanz, J. T., Norton, M. C., et al. (1998), APOE genotype predicts when—not whether—one is predisposed to develop Alzheimer disease, *Nat. Genet.*, 19, 321–322.
98. Roses, A. D. (1995), Apolipoprotein E genotyping in the differential diagnosis, not prediction, of Alzheimer's disease, *Ann. Neurol.*, 38, 6–14.
99. Saunders, A. M., Hulette, O., Welsh-Bohmer, K. A., et al. (1996), Specificity, sensitivity, and predictive value of apolipoprotein-E genotyping for sporadic Alzheimer's disease, *Lancet*, 348, 90–93.
100. Welsh-Bohmer, K. A., Gearing, M., Saunders, A. M., et al. (1997), Apolipoprotein E genotypes in a neuropathological series from the Consortium to Establish a Registry for Alzheimer's Disease (CERAD). *Annals of Neurology*, 42, 319–325.
101. Green, R. C. (2002), Risk assessment for Alzheimer's disease with genetic susceptibility testing: Has the moment arrived? *Alzheimer's Care Quart.*, 3, 208–214.
102. Romas, S. N., Santana, V., Williamson, J., et al. (2002), Familial Alzheimer disease among Caribbean Hispanics: A reexamination of its association with APOE, *Arch. Neurol.*, 59, 87–91.
103. Daw, E. W., Payami, H., Nemens, E. J., et al. (2000), The number of trait loci in late-onset Alzheimer disease, *Am. J. Hum. Genet.*, 66, 196–204.
104. Pinholt, M., Frederiksen, J. L., and Christiansen, M. (2006), The association between apolipoprotein E and multiple sclerosis, *Eur. J. Neurol.*, 13, 573–580.

105. Kamboh, M. I., Minster, R. L., Feingold, E., et al. (2006), Genetic association of ubiquilin with Alzheimer's disease and related quantitative measures, *Molec. Psychiat.*, 11, 273–279.
106. Li, Y., Hollingworth, P., Moore, P., et al. (2005), Genetic association of the APP binding protein 2 gene (APBB2) with late onset Alzheimer disease, *Hum. Mutat.*, 25, 270–277.
107. Ramos, E. M., Lin, M.-T., Larson, E. B., et al. (2006), Tumor necrosis factor-alpha and interleukin 10 promoter region polymorphisms and risk of late-onset Alzheimer disease, *Arch. Neurol.*, 63, 1165–1169.
108. Prince, J. A., Feuk, L., Sawyer, S. L., et al. (2001), Lack of replication of association findings in complex disease: An analysis of 15 polymorphisms in prior candidate genes for sporadic Alzheimer's disease, *Eur. J. Hum. Genet.*, 9, 437–444.
109. Croes, E. A., Dermaut, B., van der Cammen, T. J. M., et al. (2000), Genetic testing should not be advocated as a diagnostic tool in familial forms of dementia, *Am. J. Hum. Genet.*, 67, 1033–1035.
110. Zipser, B. D., Johanson, C. E., Gonzalez, L., et al. (2007), Microvascular injury and blood-brain barrier leakage in Alzheimer's disease, *Neurobiol. Aging*, 28, 977–986.
111. Saunders, N. R., Habgood, M. D., and Dziegielewska, K. M. (1999), Barrier mechanisms in the brain. I. Adult brain. *Clin. Exp. Pharmacol. Physiol.*, 26, 11–19.
112. Martino, G., Adorini, L., Rieckmann, P., et al. (2002), Inflammation in multiple sclerosis: The good, the bad, and the complex, *Lancet Neurol.*, 1, 499–509.
113. Blennow, K., and Hampel, H. (2003), CSF markers for incipient Alzheimer's disease, *Lancet Neurol.*, 2, 605–613.
114. Wallin, A. K., Blennow, K., Andreasen, N., et al. (2006), CSF biomarkers for Alzheimer's Disease: Levels of beta-amyloid, tau, phosphorylated tau relate to clinical symptoms and survival, *Dement. Geriatr. Cogn. Disord.*, 21, 131–138.
115. Kapaki, E., Paraskevas, G. P., Michalopoulou, M., et al. (2000), Increased cerebrospinal fluid tau protein in multiple sclerosis, *Eur. Neurol.*, 43, 228–232.
116. Sussmuth, S. D., Reiber, H., and Tumani, H. (2001), Tau protein in cerebrospinal fluid (CSF): A blood-CSF barrier related evaluation in patients with various neurological diseases, *Neurosci. Lett.*, 300, 95–98.
117. Dyment, D. A., Herrera, B. M., Cader, M. Z., et al. (2005), Complex interactions among MHC haplotypes in multiple sclerosis: Susceptibility and resistance, *Hum. Mol. Genet.*, 14, 2019–2026.
118. Lincoln, M. R., Montpetit, A., Cader, M. Z., et al. (2005), A predominant role for the HLA class II region in the association of the MHC region with multiple sclerosis, *Nat. Genet.*, 37, 1108–1112.
119. Payami, H., Schellenberg, G. D., Zareparsi, S., et al. (1997), Evidence for association of HLA-A2 allele with onset age of Alzheimer's disease, *Neurology*, 49, 512–518.
120. Andreasen, N., Minthon, L., Davidsson, P., et al. (2001), Evaluation of CSF-tau and CSF-Abeta42 as diagnostic markers for Alzheimer disease in clinical practice, *Arch. Neurol.*, 58, 373–379.
121. Sunderland, T., Gur, R. E., and Arnold, S. E. (2005), The use of biomarkers in the elderly: Current and future challenges, *Biol. Psychiatry*, 58, 272–276.
122. Herukka, S.-K., Hallikainen, M., Soininen, H., et al. (2005), CSF amyloid-beta-42 and tau or phosphorylated tau and prediction of progressive mild cognitive impairment, *Neurology*, 64, 1294–1297.
123. Mehta, P. D., Pirttila, T., Patrick, B. A., et al. (2001), Amyloid beta protein 1-40 and 1-42 levels in matched cerebrospinal fluid and plasma from patients with Alzheimer disease, *Neurosci. Lett.*, 304102–304106.
124. Iqbal, K., Flory, M., Khatoon, S., et al. (2005), Subgroups of Alzheimer's disease based on cerebrospinal fluid molecular markers, *Ann. Neurol.*, 58, 748–757.

125. Itoh, N., Arai, H., Urakami, K., et al. (2001), Large-scale, multicenter study of cerebrospinal fluid tau protein phosphorylated at serine 199 for the antemortem diagnosis of Alzheimer's disease, *Ann. Neurol.*, 50, 150–156.

126. Hampel, H., Buerger, K., Zinkowski, R., et al. (2004), Measurement of phosphorylated tau epitopes in the differential diagnosis of Alzheimer disease: A comparative cerebrospinal fluid study, *Arch. Gen. Psychiatry*, 61, 95–102.

127. Hampel, H., Burger, K., Pruessner, J. C., et al. (2005), Correlation of cerebrospinal fluid levels of tau protein phosphorylated at threonine 231 with rates of hippocampal atrophy in Alzheimer disease, *Arch. Neurol.*, 62, 770–773.

128. De Jong, D., Jansen, R. W., Kremer, B. P., et al. (2006), Cerebrospinal fluid amyloid beta42/phosphorylated tau ratio discriminates between Alzheimer's disease and vascular dementia, *J. Gerontol. Biol. Sci. Med. Sci.* 61, 755–758.

129. Thal, L. J., Kantarci, K., Reiman, E. M., et al. (2006), The role of biomarkers in clinical trials for Alzheimer disease, *Alzheimer Dis. Assoc. Disord.* 20, 6–15.

130. Link, H., and Huang, Y. M. (2006), Oligoclonal bands in multiple sclerosis cerebrospinal fluid: An update on methodology and clinical usefulness, *J. Neuroimmunol.*, 180(1–2), 17–28.

131. Wolinsky, J. S.,the PROMiSe Study Group (2003), The diagnosis of primary progressive MS, *J. Neurol. Sci.*, 206, 145–152.

132. Imrell, K., Landtblom, A. M., Hillert, J., et al. (2006), Multiple sclerosis with and without CSF bands: Clinically indistinguishable but immunogenetically distinct, *Neurology*, 67, 1062–1064.

133. Furlan, R., Rovaris, M., Martinelli Boneschi, F., et al. (2005), Immunological patterns identifying disease course and evolution in multiple sclerosis patients, *J. Neuroimmunol.*, 165, 192–200.

134. Teunissen, C. E., Dijkstra, C., and Polman, C. (2005), Biological markers in CSF and blood for axonal degeneration in multiple sclerosis, *Lancet Neurol.*, 4, 32–41.

135. Haghighi, S., Andersen, O., Oden, A., et al. (2004), Cerebrospinal fluid markers in MS patients and their healthy siblings, *Acta Neurol. Scand.*, 109, 97–99.

136. Eikelenboom, M. J., Petzold, A., Lazeron, R. H., et al. (2003), Multiple sclerosis: Neurofilament light chain antibodies are correlated to cerebral atrophy, *Neurology*, 60, 219–223.

137. Brettschneider, J., Petzold, A., Junker, A., et al. (2006), Axonal damage markers in the cerebrospinal fluid of patients with clinically isolated syndrome improve predicting conversion to definite multiple sclerosis, *Mult. Scler.*, 12, 143–148.

138. Leoni, V., Masterman, T., Diczfalusy, U., et al. (2002), Changes in human plasma levels of the brain specific oxysterol 24S hydroxycholesterol during progression of multiple sclerosis, *Neurosci. Lett.*, 331, 163.

139. Potti, A., Dressman, H. K., Bild, A., et al. (2006), Genomic signatures to guide the use of chemotherapeutics, *Nat. Med.*, 12, 1294–1300.

140. Dickerson, B. C., and Sperling, R. A. (2005), Neuroimaging biomarkers for clinical trials of disease-modifying therapies in Alzheimer's disease, *NeuroRx*, 2, 348–360.

141. Littman, B. H., and Williams, S. A. (2005), The ultimate model organism: Progress in experimental medicine, *Nat. Rev. Drug Discov.*, 4, 631–638.

142. Patterson, S. D., and DuBose, R. F. (2006), The role of biomarkers in the future of drug development, *Expert Opin. Drug Discov.*, 1, 199–204.

143. Mohs, R. C., Kawas, C., and Carillo, M. C. (2006), Optimal design of clinical trials for drugs designed to slow the course of Alzheimer's disease, *Alzheimer's Dementia*, 2, 131–139.

144. Patterson, S. D., and DuBose, R. F. (2006), The role of biomarkers in the future of drug development, *Expert Opin. Drug Discov.*, 1(3), 199–204.

145. Severino, M. E., Dubose, R. F., and Patterson, S. D. (2006), A strategic view on the use of pharmacodynamic biomarkers in early clinical drug development, *IDrugs*, 9, 849–853.

146. Hye, A., Lynham, S., Thambisetty, M., et al. (2006), Proteome-based plasma biomarkers for Alzheimer's disease, *Brain*, 129, 3042–3050.

147. Simonsen, A. H., McGuire, J., Podust, V. N., et al. (2007), A novel panel of cerebrospinal fluid biomarkers for the differential diagnosis of Alzheimer's disease versus normal aging and frontotemporal dementia, *Dement. Geriatr. Cogn. Disord.*, 24, 434–440.

148. Shi, M., Caudle, W. M., and Zhang, J. (2008), Biomarker discovery in neurodegenerative diseases: A proteomic approach, *Neurobiol. Dis.*, Doi:10.1016/j.nbd.2008.09.004.

149. Akuffo, E. L., Davis, J. B., Fox, S. M., et al. (2008), The discovery and early validation of novel plasma biomarkers in mild-to-moderate Alzheimer's disease patients responding to treatment with rosiglitazone, *Biomarkers*, 13, 618–636.

150. Vellas, B., Andrieu, S., Sampaio, C., et al.; A European Task Force Group (2008), Endpoints for trials in Alzheimer's disease: A European task force consensus, *Lancet Neurol.*, 7, 436–450.

# *Review Boards*

**Maureen N. Hood,[1] Jason F. Kaar,[2] and Vincent B. Ho[1]**

*[1]Department of Radiology and Radiological Sciences and [2]Office of General Counsel, Uniformed Services University of the Health Sciences, Bethesda, Maryland 20814*

**Contents**

1 Background
  1.1 Development of Human Research Guidelines
  1.2 History and Landmark Doctrines
2 Definitions
  2.1 Institutional Review Board
  2.2 Ethics Committee
  2.3 Data Monitoring Safety Boards
  2.4 Assurances
  2.5 Human Research
  2.6 Informed Consent
  2.7 Institution
  2.8 IRB Registration
3 Global Nature of Research Review: International Committee on Harmonisation and Good Clinical Practice
4 Composition, Roles and Activities of an Institutional Review Board/Institutional Ethics Committee
  4.1 Composition
  4.2 Oversight
  4.3 Informed Consent
  4.4 Other Concerns

The opinions or assertions contained herein are the private views of the authors and are not to be construed as official or reflecting the views of the Uniformed Services University of the Health Sciences or the Department of Defense.

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

5 Types of IRB Reviews
5.1 Initiating a Clinical Trial and Types of IRB/IECs
5.2 Exempted Review
5.3 Expedited Review (45 CFR § 46.110)
5.4 Full Review
5.5 Continuing/Annual Review
6 Adverse Events
7 IRB Recordkeeping and Administrative Actions
7.1 Information That Must Be Retained and Duration
7.2 Training and Credentials
7.3 Privacy and Confidentiality
8 Investigator Responsibilities
8.1 Principal Investigator
8.2 Qualifications and Training
8.3 Study Design and Performance
8.4 Logistical Support and Budgetary Planning
8.5 Study Security and Maintenance
8.6 Communication with Stakeholders
9 Summary
Appendix: Websites
References

> The fundamental lesson to be understood by doctors is that when they stray from the traditional concepts of medicine—saving lives, healing, and ameliorating pain and suffering—they enter into a territory that is fraught with danger, not only for their patients, but also for the morality of medicine as a profession.
>
> —Lounsbury DE, Bellamy RF, Beam TE, and Sparacino LR.
> Editors of *Military Medical Ethics*, Volume 2, 2003,
> in memory of Sheldon H. Harris, Ph.D.

## 1 BACKGROUND

### 1.1 Development of Human Research Guidelines

Adherence to the highest ethical and moral standards is important not only for routine clinical practice but also for all research involving human subjects. In the United States, the Food and Drug Administration (FDA) regulates the evaluation of medications and biomedical devices and has the sole responsibility of approval for their commercial sale for use in humans. The Office of Human Research Protections (OHRP) oversees the regulation of federally funded human research [1, 2]. The primary oversight authority of investigators, however, is the responsibility of the institutional review board (IRB) or institutional ethics committee (IEC). The delegation of oversight responsibility to the IRB was made in recognition of the regional and institutional variations that exist in ethical and treatment standards of care [3]. The geographic proximity of the IRB/IEC to the location of the research facilitates timely direct communication between the regulatory body and the investigator. However, there is a growing move toward regional or community review boards, as well as a move toward independent review boards. Open communication between investigators and the IRB/IEC is critical for the timely performance of

audits, reviews, notifications, and amendments and ultimately provides improved safety and welfare for the research participants.

## 1.2 History and Landmark Doctrines

The history of IRBs/IECs has not been without significant challenges and problems, however. Prior to the early to mid-twentieth century, human research was governed by individual conscience, professional bodies, societal norms, and government customs of the time [4]. The earliest known legal requirement pertaining to informing human subjects of medical experimentation was the Berlin Code of 1900 [5]. The United States made some initial efforts toward forming human research guidelines and ethics in the early 1900s with studies such as the U.S. Army's Yellow Fever Project; but, unfortunately, most physicians and scientists either knew little of these guidelines or refused to adhere to such strict rules [5, 6].

Formalization of human research standards came as a reaction to the atrocities committed by Nazi medical researchers found during the Nuremberg Military Tribunal [7]. In addition, it is estimated that several hundred thousand deaths occurred from the Japanese biomedical, biological, and chemical warfare experimentation on involuntary subjects from the mid-1920s until 1945 [8]. The atrocities of forced human testing performed by Nazi and Japanese scientists on prisoners and civilians during World War II resulted in a global outcry and the establishment of a variety of ethical guidelines for research involving human subjects. In 1953, the Department of Defense instituted the first directive for an official process for conducting human-use research, called the Wilson Memorandum, which unfortunately, was kept mostly secret [6]. The National Institutes of Health attempted to establish policies for approving research studies on humans, but no consensus could be reached as to which human-use research projects should be subject to regulation [9]. A more substantial and formal systematic protection of human subjects in medical research remained dormant for many years.

In 1964, the World Medical Association issued the Declaration of Helsinki, a groundbreaking treatise for international research protections. The Declaration of Helsinki continues to be followed today and has been revised periodically, most recently in 2000 during the 52nd World Medical Association General Assembly in Edinburgh, Scotland [10]. However, the United States did not formulate human research regulations until after Henry Beecher published his ethics paper in 1966 [11] outlining some of the unethical human research projects being conducted in the United States, which sparked the U.S. Congress to pass the National Research Act in 1974 [12]. The National Research Act established the National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research, which is best known for formulating the links between ethical principles and human research practices in the Belmont Report of 1979. This legislation formally required the establishment of IRBs to oversee all federally funded research activities [12, 13]. The Belmont Report (available at http://www.hhs.gov/ohrp/humansubjects/guidance/belmont.htm) articulates the three main ethical principles upon which all IRBs should operate: respect for persons, beneficence, and justice [13]. Various international and federal regulations that govern human subject research require the oversight of research by an IRB or its equivalent (e.g., ethics committees, institutional ethics committees, health research ethics committees, etc.).

## 2 DEFINITIONS

### 2.1 Institutional Review Board

An independent group, committee, or similar such entity consisting of a combination of medical/scientific and nonmedical/nonscientific members that is formally organized to oversee biomedical research involving human subjects at an institution. An IRB's primary objective is to protect the rights, safety, and welfare of human subjects participating in biomedical research [14].

### 2.2 Ethics Committee

Essentially equivalent to an IRB but usually uses the name ethics committee (EC) or human research ethics committees (HREC) in countries outside the United States. Each country sets specific minimum requirements for the composition of ECs as guided by the International Conference on Harmonisation (ICH) E6 Good Clinical Practice (GCP) guidelines [15].

### 2.3 Data Monitoring Safety Boards

An independent, objective group that is charged with monitoring the safety of clinical trials. Clinical trials that are evaluating new drugs, biologics, or devices are required by 21 CFR § 312.50 and 21 CFR § 312.56 to monitor their studies [16]. Data monitoring safety boards (DSMBs) are an additional group that may perform a monitoring function [17]. Although DSMB composition may vary from place to place and country to country, they should be composed of scientists, clinicians, biostatisticians, and bioethicists who have the expertise necessary to be able to evaluate independently an objective safety analysis of the clinical trial [18].

### 2.4 Assurances

Documents for acknowledging compliance by an institution of mandated federal policies for human subject research and for the protection of research subjects. Assurances are a written commitment obtained through the Office of Human Research Protections (OHRP) that must be approved before federal funding can be awarded. The most common type of assurance accepted by OHRP is the Federalwide Assurance (FWA). All organizations participating in federal research that have an IRB must obtain an assurance and a written agreement between IRBs and/or organizations outlining their relationships and commitments to the recognized assurance [19].

### 2.5 Human Research

A systematic investigation conducted by an investigator involving living individuals or identifiable information of the human subjects that is conducted in a manner that is designed to contribute new knowledge. All research activities that fall under the regulation of Title 45 CFR part 46 are required to abide by IRB/IEC oversight [3].

### 2.6 Informed Consent

The process by which human subjects are informed of all aspects of a research study including potential risks and benefits that may influence the subject's decision on whether or not to volunteer to participate in the study. The informed consent document must be reviewed and approved by the IRB/IEC. In the event that a research subject is not legally competent to consent, as in the case with a minor or a person who lacks mental capacity (such as a person in a coma), the law normally allows a third person, typically a family member, to consent on the person's behalf. Check the requirements of the jurisdiction you are in to determine whether it is appropriate and the procedure to be followed. Care should be taken to ensure the IRB/IEC is acting in the best interests of the research subjects. In at least one instance an IRB was cited for erroneously putting the interests of investigators over the interests of subjects [20]. Informed consent may be oral or written depending on the risk level. More than minimal risk studies typically require formal written consent. Documentation of informed consent is done through a document that is signed and dated [15].

### 2.7 Institution

Any medical or dental entity or facility, public or private, where clinical trials can be conducted.

### 2.8 IRB Registration

Registration is required by any IRB/IEC that reviews U.S. federally supported human research or any institution that is subject to the Common Rule [3]. IRBs/IECs must be registered with OHRP before they may be designated on an FWA. IRBs/IECs that are registered are listed on the OHRP website at http://ohrp.cit.nih.gov/search/asearch.asp#ASUR. An OHRP assurance is also required.

## 3 GLOBAL NATURE OF RESEARCH REVIEW: INTERNATIONAL COMMITTEE ON HARMONISATION AND GOOD CLINICAL PRACTICE

All human subject research, regardless of the country in which the research is taking place, should be conducted under the International Conference on Harmonisation (ICH), E6 Good Clinical Practice (GCP) guidance (available at http://www.fda.gov/cder/guidance/959fnl.pdf). The International Conference on Harmonisation of Technical Requirements for the Registration of Pharmaceuticals for Human Use grew out of a meeting hosted by the European Federation of Pharmaceutical Industries and Associations in Brussels in 1990 that brought together the regulatory authorities and pharmaceutical industries in the European Union, United States, and Japan to discuss joint regulation of the pharmaceutical industry. The early intent of the ICH was to establish an international process for the development and regulation of new medical products in a more efficient manner.

The ICH has continued to expand its guidance through harmonization of therapeutic advances, technical developments, and other specific medical products based

on scientific consensus and the protection of public health and individual safety on an international perspective. The World Health Organization (WHO) has endorsed these ICH GCP international standards as well as the Declaration of Helsinki and the Council for International Organizations of Medical Sciences (CIOMS) International Ethical Guidelines for Biomedical Research Involving Human Subjects as the ethical and scientific standards for conducting biomedical human subject research [21]. These guidelines are intended to help facilitate biomedical research and clinical trials worldwide. The Office for Human Research Protections has been working on compiling an international listing of countries that have laws, regulations, or guidelines to cover human research. The 2005 international compilation listing includes 72 countries with links to standards and international organizations (available at http://www.hhs.gov/ohrp/international/HSPCompilation.pdf).

## 4 COMPOSITION, ROLES, AND ACTIVITIES OF INSTITUTIONAL REVIEW BOARD/INSTITUTIONAL ETHICS COMMITTEE

### 4.1 Composition

The composition of an IRB/IEC will vary but should consist of a reasonable number of members with qualifications and expertise in a variety of aspects so as to ensure the rights, safety, and welfare of subjects involved in research are protected. In the United States, 45 CFR § 46.107 strongly encourages diversity, not only as to race, creed, sex, and national origin, but also professional and social experience [3]. If a vulnerable population, such as prisoners are used, there must be someone on the committee representing their interests (45 CFR § 46.304). Minimum requirements: at least five members, at least one member who is from a nonscientific background, and at least one member who is independent of the site or institution. Both genders must be represented, including making quorum, which must be at least 50% of the primary membership (45 CFR § 46.108) [14]. Many institutions are moving to a two-tiered system that adds a scientific review component step before the protocol is presented to the IRB/IEC. The scientific review is intended to aid in promoting research excellence. This two-step system is intended to help streamline the IRB/IEC review process by having a scientific review by experts on the subject matter first address the study design aspects of the protocol, which then allows the IRB/IEC to focus on the human use issues. However, the IRB/IEC still needs to review the report from the scientific review as well as review the protocol as a whole in the human use context.

### 4.2 Oversight

The main goals of an IRB/IEC are to ensure that the rights of the research subject especially in regards to privacy and informed consent are maintained, that investigators conduct their research in a professional and ethical manner, and that the potential benefits of the study outcomes outweigh the potential risks to the patient. IRBs/IECs are also tasked with the responsibilities to ensure that the studies that fall under their purview are methodologically sound in their design, whereby the objectives of the research will provide meaningful data and the risks and benefits

for study involvement are suitable for human participation. If an institution uses a scientific review committee, the findings from the scientific review should be reviewed by the IRB/IEC. These roles for the IRB are not exclusive to the initial protocol approval but are continuous throughout the duration of the study and for at least 3 years after the study has completed, and even longer for pharmaceutical clinical trials according to 45 CFR § 46.115 [3].

### 4.3 Informed Consent

The informed consent process and documentation is a critical element the IRB/IEC must review. The information contained within the informed consent document must be at a language level appropriate for the study population (usually sixth grade level) and may need to be made available in more than one language in order to support all participants. The information must be clear to the potential volunteer so that the decision of whether to participate can be made freely by the subject. When the research involves a medical product such as a device or pharmaceutical, an adequate summary of the product must be communicated to the potential subject that includes the safety information and clinical experience or the product. Any compensation for subject participation should be contained in the informed consent document as well. Compensation may not be at a level that is seen as improper inducement to participate in the research. The process of recruiting the potential subjects as well as a description of the informed consent process must be approved by the IRB/IEC in order to reduce potential coercion.

### 4.4 Other Concerns

There are other important concerns that an IRB/IEC must also consider for each research protocol such as legal issues and intellectual property rights. Members of the IRB must remember that research protocols may contain sensitive or proprietary information. It is imperative that members of the IRBs/IECs respect the confidentiality of the research protocols. "Products of human ingenuity" such as a new medication or device should be considered intellectual property as guided by *Diamond v. Chakrabarty*, 447 U.S. 303, 1980, [22, 23]. The IRB will also be assessing a number of other issues as outlined in Table 1, all in accordance with respect for persons, beneficence, and justice [24].

## 5 TYPES OF IRB REVIEWS

### 5.1 Initiating Clinical Trial and Types of IRBs/IECs

Before starting any human research or clinical trial, an investigator must attain written approval from his or her IRB/IEC prior to initiation of the study. This approval process includes approval of informed consent forms, advertisements, and other written documents as well as specific procedures, data collection, analysis, and storage. When an institution does not have an IRB/IEC, there are independent IRBs/IECs that can be contracted to serve as an IRB/IEC to review research under your institution's OHRP accepted assurance. For multicenter clinical trials, central IRBs/IECs can be used [25]. Clinical trials that fall under the investigative new drug

**TABLE 1 Overview of Basic Items in an IRB Review**

| Purpose of Study | Background, Summary, and Aims |
|---|---|
| Sponsor | The name of the organization that takes the responsibility for and initiates a clinical trial. It may be a private or public organization. |
| Investigator | Qualifications, financial disclosures, conflicts of interest. |
| Study population | Inclusion/exclusion criteria, subject recruitment, appropriateness of subject selection, advertisement. |
| Scientific design | Objectives, design description adequate to answer question, research procedures. |
| Informed consent | Written documentation of process is essential. |
| Potential risks and discomforts to subjects | Risk relative to alternatives, risk–benefit analysis. |
| Confidentiality | Are procedures adequate to protect the privacy and confidentiality of the subject? |
| Data oversight | Stopping rules, data monitoring board. |
| Data analysis | Appropriate data analysis and statistical analysis. |
| Compensation and costs | Is compensation or reimbursement of subject reasonable? |
| Other | Depending on the type of clinical trial, other checks may be considered by the IRB. |

(IND) regulations must be approved by an IRB/IEC. Central IRBs/IECs help facilitate the IRB/IEC process by reducing redundancy in effort, and for small institutions, can provide a more diverse perspective for reviewing the research protocol. Centralized IRBs/IECs are consistent with 21 CFR § 56.114 that provides for institutions involved in multi-institutional studies and joint reviews. Centralized IRBs/IECs can be used by institutions in whole or in part for the review of a multicenter clinical trial [25]. In addition, if there is a registered IRB/IEC in your geographic region, you can negotiate having them conduct reviews for your institution, or an institution may also consider establishing its own IRB/IEC.

In all cases in which research is conducted using funds from the U.S. government each institution must obtain its own assurance that is accepted by OHRP (45 CFR § 46.103). Regardless of the type of review system, there are a number of elements in common that must be addressed in the protocol submission (Table 1). Local IRBs/IECs have the flexibility to specify the format and detailed content required for the review as long as the information meets federal requirements. The type of review a protocol receives is up to the IRB/IEC, not the investigator. It is prudent for an investigator to discuss a potential protocol with the IRB/IEC for guidance before submitting a protocol. However, the IRB/IEC or its properly established subcommittee must review the protocol before deciding which type of review is most appropriate. This is normally determined by local guidelines.

### 5.2 Exempted Review

Certain types of research can be exempted from the full detailed review of the IRB. The research must meet all of the criteria in 45 CFR § 46.101(b) (1-6) and involve no greater than minimal risk to be eligible for exempted review (e.g., retrospective review of last 100 abdominal CT studies) (Table 2). The IRB typically does not conduct continuing reviews of exempted protocols, but the investigator will be

**TABLE 2 Exempted Research Criteria as Found in 45 CFR§ 46.101 (b)**

| Category | Conditions |
|---|---|
| (1) Research conducted in established or commonly accepted educational settings, involving normal educational practices, such as: | (a) Research on regular and special education instructional strategies,<br>OR<br>(b) Research on the effectiveness of or the comparison among instructional techniques, curricula, or classroom management methods. |
| (2) Research involving the use of educational tests (cognitive, diagnostic, aptitude, achievement), survey procedures, interview procedures or observation of public behavior, if:<br>NOTE: all three conditions must apply. | (a) Information obtained is recorded in such a manner that human subjects cannot be identified, directly or through identifiers linked to the subjects;<br>AND<br>(b) Any disclosure of the human subjects' responses outside the research would not reasonably place the subjects at risk of criminal or civil liability or be damaging to the subject's financial standing, employability, or reputation;<br>AND<br>(c) Subjects are not under the age of 18 or members of a vulnerable class, including prisoners, pregnant women, individuals who are mentally disabled or economically or educationally disadvantaged. |
| (3) Research involving the use of educational tests (cognitive, diagnostic, aptitude, achievement), survey procedures, interview procedures, or observation of public behavior that is not exempt under paragraph (2) above if: | (a) The human subjects are elected or appointed public officials or candidates for public office;<br>OR<br>(b) Federal statutes(s) require(s) without exception that the confidentiality of the personally identifiable information will be maintained throughout the research and thereafter. |
| (4) Research, involving the collection or study of existing data, documents, records, pathological specimens, or diagnostic specimens, if: | (a) These sources are publicly available;<br>OR<br>(b) If the information is recorded by the investigator in such a manner that subjects cannot be identified, directly or through identifiers linked to the subjects. |
| (5) Research and demonstration projects which are conducted by or subject to the approval of department or agency heads, and which are designed to study, evaluate, or otherwise examine: | (a) Public benefit or service programs;<br>OR<br>(b) Procedures for obtaining benefits or services under those programs;<br>OR<br>(c) Possible changes in or alternatives to those programs or procedures;<br>OR<br>(d) Possible changes in methods or levels of payment for benefits or services under those programs. |
| (6) Taste and food quality evaluation and consumer acceptance studies, if: | (a) Wholesome foods without additives are consumed;<br>OR<br>(b) If a food is consumed that contains a food ingredient at or below the level **and** for a use found to be safe, **or** agricultural chemical or environmental contaminant at or below the level found to be safe, by the Food and Drug Administration or approved by the Environmental Protection Agency or the Food Safety and Inspection Service of the USDA. |

*Note*: In order to qualify for exemption, all of a project's activities must qualify as one or more categories from the left column. (To qualify for a category, a project must meet the conditions to the right of the category). The IRB/IEC makes the decision regarding exempted status. This is presented as a guideline only.

*Source*: Adapted from Ref. [3].

required to complete an annual report. The investigator is also still bound by all the federal ethical requirements and regulations for good clinical practice (GCP) [15]. Informed consent is typically not required by law for exempted protocols, however, the sponsor or local IRB/IEC may require it. In the United States, the Health Insurance Portability and Accountability Act (HIPAA) authorization or waiver may also be needed. The IRB/IEC will make the determination as to which protocols fit exempted review status and the human use requirements for the investigator.

### 5.3 Expedited Review (45 CFR § 46.110)

Certain kinds of research can be conducted through an expedited review process. Protocols that may be subject to expedited review need to have certain characteristics that may include: (1) must be no more than minimal risk; (2) may be exempted research; (3) must fit within the categories outlined by federal regulations 45 CFR § 46.110; (4) may be a review of minor revisions for a previously approved research study; and (5) may still require informed consent. The local IRB/IEC has the authority to make the decision for expedited reviews, not the investigator.

### 5.4 Full Review

Most prospective studies and some retrospective studies will require a full review. Consult with the local IRB/IEC when in doubt. The review must be conducted at a convened IRB meeting that has a quorum of at least half of its primary members. The IRB must determine that all of the federal requirements specified in 45 CFR § 46.111 and 21 CFR § 56.111 are satisfied [3, 14]. A protocol must receive the approval of the majority of the IRB members present at the meeting to be approved. The IRB must inform the investigators and the institution in writing of its decision to approve, modify, or disapprove the research. Most protocols go through at least one round of revisions prior to final approval.

### 5.5 Continuing/Annual Review

Continuing reviews of each protocol will occur at intervals appropriate for the degree of risk, but must not be less than once a year. If 365 days elapse without continuing review (or a shorter period of time specified by the IRB), the study is suspended [26]. There is NO grace period. The purpose of the continuing review is to review the progress of the entire study, not just the changes to it. The review process not only protects the study subjects, but also exists to make sure enrollment is being done reasonably, and that the study is being conducted as outline by the research protocol. Studies with higher risk levels are reviewed more frequently (e.g., every 3 months or after every 3 subjects). The IRB/IEC will make the determination for the frequency, plus the IRB/IEC has the authority to change the frequency of the reviews based on what it considers appropriate for the safety of the subjects and the integrity of the study.

The IRB/IEC reviews all amendment requests by the investigator/sponsor. The amendments must not deviate significantly from the original proposal and need to be critically evaluated for potential risk to the subjects and to scientific validity of the study design.

## 6 ADVERSE EVENTS

Adverse events are taken very seriously by IRBs/IECs. Serious and unexpected adverse events must be reported expediently to the IRB [24]. A serious event is considered to be any experience that results in death or a life-threatening situation where the subject is admitted to the hospital or the outcome is a persistent or significant disability, incapacity, or congenital anomaly in a progeny. An unexpected event is any adverse experience not consistent with the current description of the IRB-approved protocol and informed consent form. Not all adverse events need to be evaluated individually by the IRB/IEC or DSMB. Events that are not significant or not clinically meaningful to the safety of the subjects should be documented, but normally do not require special review [24]. The IRB/IEC and/or DSMB will review protocols at least annually to assess the protocol for safety and documentation. Sometimes, unexpected frequencies of adverse events may cause the IRB/IEC or DSMB to either require a modification to the risk–benefit section of the protocol and informed consent document, or in rare cases, terminate the study.

## 7 IRB RECORDKEEPING AND ADMINISTRATIVE ACTIONS

### 7.1 Information That Must Be Retained and Duration

Institutional review boards are required to maintain accurate records of every protocol and investigator under their authority [1]. In the United States, the FDA requires IRBs/IECs to maintain accurate records of all research activities of the institution and store them for at least 3 years after completion or termination of a protocol [1]. The international guidelines similarly recommend the same careful documentation and 3-year minimum storage of records [27]. Some sponsored clinical trials and institutions may keep their research records longer.

### 7.2 Training and Credentials

The main functions of the IRB/IEC are to thoroughly review each research protocol, perform continuing reviews, and review adverse events. The breadth of an IRB/IEC's responsibilities is truly great and even includes the responsibility for ensuring that investigators have adequate training to perform the research procedures and have the necessary credentials to adequately care for the participants. In addition, the members of the IRB/IEC need to have periodic continuing education and training for their roles in the IRB/IEC process.

### 7.3 Privacy and Confidentiality

One additional aspect of the IRB/IEC's responsibility that cannot be overlooked concerns the protection of the participants' privacy and confidentiality rights. The IRB/IEC must assure that the investigators maintain the participants' privacy in all aspects of the research protocol. In the United States, there are added security requirements (HIPAA), which went into effect April 14, 2003 [28]. Table 3 outlines the protected health information.

**TABLE 3 Health Information Protected by 45 CFR§ 164.514(b)**

| | |
|---|---|
| Names | Biometric identifiers |
| Dates that directly relate to an individual | Persons over the age of 89 |
| Telephone numbers | Fax numbers |
| Medical record number | Social Security number |
| Health plan beneficiary number | Electronic e-mail address |
| Certificate/license numbers | Account numbers |
| Web universal resource locators (URLs) | Internet protocol (IP) address |
| Full-face photograph | Geocodes |
| Any copyrighted/unique identifying number, characteristic or code | Geographic subdivisions smaller than a state (some exceptions apply) |

*Source*: Adapted from 45 CFR § 164.514(b) [29].

The IRBs/IECs must also make sure the intellectual property rights of each protocol and the investigators are protected as well. IRB/IEC research protocols are considered confidential material for the participants and the investigators and must be kept in a secure or locked file. Each institution must also maintain accurate and adequate files and databases of their research that are kept secure.

## 8 INVESTIGATOR RESPONSIBILITIES

### 8.1 Principal Investigator

Investigators that participate in human research have a variety of responsibilities that include qualifications and expectations, study design and performance, and communication with research stakeholders. The principle investigator, or PI, has the ultimate responsibility to ensure that all investigators (i.e., associate investigators) and staff on the study have the appropriate qualifications to perform their roles within the study design. The PI bears the primary responsibility for study adherence to regulatory and ethical standards and the integrity of both the scientific and clinical aspects of the research.

### 8.2 Qualifications and Training

A primary responsibility of any investigator is that he or she has the adequate education, training, and experience to participate in the proposed research. If an investigator is a novice, it is generally expected that additional experience under the tutelage of an expert is obtained prior to the individual's performance of the task independently. If a formal certification process exists for that procedure or task, the investigator is expected to have obtained that certification. Furthermore, it is expected that investigators adhere to the laws, regulations, and ethics of the country and locality that pertain to the tasks or research that are being performed. Financial disclosure is required when the investigator has a monetary relationship with sponsors. This is to avoid possible conflicts of interest in the clinical trial [30].

In addition to being qualified to perform their roles within a project, all investigators must also receive proper human research protections training. This can be

achieved various ways such as via a traditional classroom instruction, interactive CD ROM program, or Web-based tutorial. In special cases, if an IRB/IEC feels that a researcher has not met full qualifications to participate in research, the IRB/IEC can require additional training or the involvement of a more senior investigator to oversee the work and mentor the less seasoned investigator [26]. The investigators' responsibilities, regardless of the type of study or type of scientific or medical discipline, are based on the ICH and GCP documents. In the United States, in addition to human research protection training, investigators should be given a copy of the Belmont Report (http://www.hhs.gov/ohrp/humansubjects/guidance/belmont.htm) as an ethical guide.

### 8.3 Study Design and Performance

Investigators are responsible for designing studies that have not only scientific merit but also do not unduly jeopardize the well-being of the participants. Accordingly, human research studies must be designed such that the study can be validated or replicated, the study participants are not exposed to unnecessary risks, and the benefits of the research outweigh the risks to the individuals volunteering their time for the research. The investigator and the IRB should note that "a poorly designed human research protocol is unethical" [31, p. 109]. Inherent to this concept is that the study design includes proper surveillance methods for detecting potential risks and/or qualified individuals to identify abnormalities.

### 8.4 Logistical Support and Budgetary Planning

Prior to the performance of the study, the investigators must ensure that there is adequate logistical support (e.g., personnel, equipment, space, etc.) and funding for the performance of the proposed study. Should resources lapse (e.g., key investigator leaves the institution or funding abruptly be terminated) during the conduct of a study, it is the responsibility of the investigators to notify the IRB and potentially suspend or amend the protocol to more modest goals should the resource not be replenished.

### 8.5 Study Security and Maintenance

Investigators are responsible for the integrity of the data, the human subject recruitment, informed consent process, and maintenance not only of the regulatory documentation but also of subject data files. Investigators are also responsible for maintenance of subject privacy and confidentiality, which includes development of the processes for the assurance of these elements. It is common practice for the secure storage of study records and materials. Secure storage of research documents is usually required to be no less than 2 years following FDA market application approval or discontinuation (21 CFR § 312.62) [16], but some sponsors and institutions require much longer storage periods, so investigators are responsible for finding out specific requirements. The records that must be maintained by the investigator are: investigational drug disposition, dates, quantities of drugs given, unused supplies, case histories, all recorded observations, supporting data, signed informed consent documents, progress notes, and any other pertinent medical records or data

collected for the trial (21 CFR § 312.62) [16]. These recordkeeping requirements may also be required by the institution. Investigators have the responsibility to ensure that records for the research study are being properly completed and maintained.

## 8.6 Communication with Stakeholders

Investigators are expected to communicate honestly, openly, and freely with not only the subjects but also the IRB/IEC or other outside regulatory (e.g., FDA) or interested bodies (e.g., sponsor or funding agency)—that is all persons with a "stake" in the research. Investigators, thus, are responsible not only for the preparation of the research proposal document but also for the presentation and defense of the proposed study to the IRB/IEC and any other applicable regulatory committee. A clinical study which entails the use of chest X rays, for example, typically must also receive approval from the institution's radiation safety committee. Investigators also have a responsibility to maintain the communication with the IRB/IEC and other stakeholders even after the study has been initially approved.

***New Findings*** Investigators are expected to notify the IRB/IEC of any new findings such as unanticipated risks or benefits. There is a similar expectation that subjects also be notified. Investigators are expected to incorporate their new information into the study and request amendment to the study design or documentation (e.g., informed consent documentation) in accordance to the new information. Note that this pertains not only to new information that the investigator may have learned from the specific research undertaken but also from the reported literature or available information from studies performed by other investigators. If the study is an investigational new drug trial, the pharmaceutical company (i.e., sponsor) typically provides regular updates on new findings, namely serious adverse events, that arise from all sites such that individual sites can benefit from the experience of other investigators. In response to any new study finding, investigators have the responsibility of determining if the protocol needs amendment and the responsibility of filing the appropriate regulatory paperwork for protocol amendment.

***Amendment Requests and Protocol Deviation Reporting*** Any changes or amendment requests to the protocol require IRB/IEC approval. If, after the fact, a deviation for the protocol is noted, the deviation must be documented and explained to the IRB/IEC and the sponsor of the study. Deviations are considered protocol violations and are taken seriously by an IRB/IEC. Deviations made for patient safety may be done prior to IRB/IEC approval but must be reported and an amendment request made immediately to the IRB/IEC, the sponsor, and any regulatory authorities, or the trial may be suspended [15]. The protection of the human subject is the primary responsibility of the investigator. The investigator must not coerce subjects into volunteering for trials, and the investigator must be careful to fully disclose the potential risks and benefits to the subjects prior to enrollment (45 CFR § 46.116) [3]. Only investigators on the IRB/IEC-approved protocol are allowed to administer investigation drugs and with disposing of or returning unused supplies to the sponsor.

***Complying with Audits*** Investigators also are responsible of complying with regulatory audits, which can be performed as a scheduled audit or impromptu inspection by the IRB or on occasion, the FDA or other regulatory body. IRBs/IECs are required to review human subject research no less than annually. In studies associated with higher risk, IRBs/IECs may choose to review the protocol more frequently. In any case, the investigators are responsible for adhering to these regulatory responsibilities which typically include an interview with an IRB/IEC representative, a reporting of study progress, and preparation of data files for review.

***Safety and Data Safety Monitoring Boards*** Investigators must also report regularly to the sponsor any data or information specified by the agreement between the sponsor and investigator for the conduction of the trial. Types of reports are IRB/IEC approvals and rosters, progress reports, safety reports, and final reports. In addition, the investigators may be monitored by study monitors that are supplied by the sponsor, in addition to any monitoring requirement made by the IRB/IEC. DSMBs are growing in use by IRBs/IECs. However, there is some confusion and overlap of responsibilities between the IRBs/IECs and the DSMBs because the FDA requirements (21 CFR § 312.66) for reporting unanticipated problems are vague [16]. The overall goal should be to adequately protect the research participants through a process that can reliably detect harm to participants and then respond quickly and appropriately to the particular instances [32]. DMSBs are meant to help the research process, but more standardization of roles and adverse event reporting need to be established on an international basis [32]. Nonetheless, investigators must comply with approved protocol procedures and safety regulations. Noncompliance by investigators such as failure to report changes in the protocol, failure to report safety problems, or misuse of informed consent documents may result in the IRB/IEC halting the protocol and investigating the issues [26]. Serious lapses or continued lapses by the investigator may lead to termination of the protocol. IRBs/IECs want to ensure that the research being performed with human subjects is being done in a manner that protects the rights and welfare of the participants. Similarly, sponsors and researchers hope to bring safe and beneficial health care products, techniques, and devices to clinical practice.

## 9 SUMMARY

Institutional review boards or institutional ethics committees are an essential part of the clinical trial system. IRBs/IECs should help the researchers conduct research in an ethical manner that protects the rights and welfare of the human participants. There are many international regulations and guidelines such as the ICH/GCP that help guide IRBs/IECs in clinical trials. These boards are in place because of the atrocities that have occurred in past medical research, especially during World War II. This chapter presented an overview of the IRB/IEC process along with resources for researchers who want to be involved in clinical trial research. The bottom line is to perform clinical research that is safe and ethical to the human participant such that new knowledge can be gained for the advancement of health care.

## APPENDIX: WEBSITES

### General Regulations and Guidance

http://www.fda.gov/oc/gcp/default.htm
http://www.fda.gov/oc/ohrt/irbs/default.htm
http://www.hhs.gov/ohrp/humansubjects/guidance/45cfr46.htm
http://www.fda.gov/cder/guidance/959fnl.pdf
http://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfcfr/CFRSearch.cfm
http://www.hhs.gov/ohrp/international/HSPCompilation.pdf
http://www.cioms.ch/frame_guidelines_nov_2002.htm
http://www.pre.ethics.gc.ca/english/policystatement/policystatement.cfm
http://www.dh.gov.uk/assetRoot/04/01/86/39/04018639.pdf
http://eudract.emea.europa.eu
http://ec.europa.eu/enterprise/pharmaceuticals/eudralex/eudralex_en.htm
http://www.nihs.go.jp/hse/drug/drug-e.html
http://www.tga.gov.au/index.htm
http://www.hhs.gov/ohrp/documents/OHRPRegulations.pdf

### Clinical Trial Guidance

http://www.fda.gov/cber/gdlns/clindatmon.pdf
http://www.emea.europa.eu
http://www.hc-sc.gc.ca/dhp-mps/prodpharma/applic-demande/guide-ld/clini/qual_cta_dec_e.html

### Ethics

http://www.ich.org/cache/compo/276-254-1.html
http://www.wma.net/e/policy/b3.htm
http://www.hhs.gov/ohrp/humansubjects/guidance/belmont.htm

### Investigator Agreement

http://www.hhs.gov/ohrp/humansubjects/assurance/spa-aii.htm

## REFERENCES

1. U.S. Department of Health and Human Services Food and Drug Administration. Information Sheets: Guidance for Institutional Review Boards and Clinical Investigators 1998 Update; available at: http://www.fda.gov/oc/ohrt/irbs/appendixc.html; accessed April 28, 2002.
2. Department of Health and Human Services, Office of the Secretary (2000), Office of Public Health and Science, and National Institutes of Health, Office of the Director, Statement of organization, functions and delegations of authority, *Fed. Reg.*, 65(114), 37136–37137.

3. National Archives and Records Administration (2006), Electronic Code of Federal Regulations (e-CFR), Title 45, Public Welfare, Part 46, Protection of Human Subjects (45 CRF §46), Government Printing Office, Washington, D.C; available at: http://www.hhs.gov/ohrp/documents/OHRPRegulations.pdf.
4. Schneider, W. H. (2002), The establishment of institutional review boards in the U.S. background history; available at: http://www.iupui.edu/~histwhs/G504.dir/irbhist.html; accessed November 2, 2002.
5. Nicholson R. H. (2003), The regulation of medical research: A historical overview, in Eckstein S., Ed., *Manual for Research Ethics Committees*, 6th ed., Cambridge University Press, New York.
6. Advisory Committee on Human Radiation Experiments (1996), *Human Radiation Experiments: Final Report of the Presidents Advisory Committee*, Oxford University Press, New York.
7. *United States v. Karl Brandt*, et al. (1949), The Medical Case: Trials of War Criminals before the Nuremberg Military Tribunals under Control Council Law 10. *Nuremberg. Volumes I–II*, Government Printing Office, Washington, D.C.
8. Harris, S. H. (2003), Japanese biomedical experimentation during the World-War-II era, in Lounsbury D. E., Bellamy, R. F., Beam, T. E., and Sparacino, L. R., Eds., *Military Medical Ethics*, Vol. 2, Office of the Surgeon General, Department of the Army, Washongton, D. C.
9. National Institutes of Health (2002), A short history of the National Institutes of Health—exhibits and galleries; available at: http://history.nih.gov/exhibits/history/main.html; accessed December 31, 2003.
10. World Medical Association General Assembly (2002), World Medical Association Declaration of Helsinki: Ethical Principles for Medical Research Involving Humans; available at: http://www.wma.net/e/policy/pdf/17c.pdf; accessed December 31, 2003.
11. Beecher, H. K. (1966), Ethics and clinical research, *N. Engl. J. Med.*, 274(24), 1354–1360.
12. *National Research Act of 1974*. P.L. 348, 93d Cong., 2d Sess (12 July 1974). A bill to amend the Public Health Service Act to establish a national program of biomedical research fellowships, traineeships, and training to assure the continued excellence of biomedical research in the United States, and for other purposes, 1974.
13. National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research (1979), The Belmont Report Ethical Principles and Guidelines for the Protection of Human Subjects of Research; available at: http://ohsr.od.nih.gov/mpa/belmont.php3; accessed October 27, 2002.
14. National Archives and Records Administration (2002), *Code of Federal Regulations, Title 21, Chapter I—Food and Drug Administration*, Department of Health and Human Services, Part 56, Institutional Review Boards (21 CRF §56), Government Printing Office, Washington, D.C.
15. ICH (2006), International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use. ICH Harmonised Tripartite Guideline: Guideline for good clinical practice, E6 (R1); available at: http://www.ich.org/LOB/media/MEDIA482.pdf; accessed April 6, 2006.
16. National Archives and Records Administration (2002), *Code of Federal Regulations, Title 21, Chapter I—Food And Drug Administration*, Department of Health and Human Services, Part 312, Investigational New Drug Application (21 CRF §312), Government Printing Office, Washington, D.C.
17. FDA (2001), U. S. Department of Health and Human Services, Food and Drug Administration, Guidance for Clinical Trial Sponsors. Establishment and Operation of Clinical

Trial Data Monitoring Committees; available at: http://www.fda.gov/OHRMS/DOCKETS/98fr/01d-0489-gdl0003.pdf; accessed April 6, 2006.

18. Slutsky, A. S., and Lavery, J. V. (2004), Data safety and monitoring boards, *N. Engl. J. Med.*, 350(11), 1143–1147.
19. HHS (2006), U.S. Department of Health & Human Services, Office for Human Research Protections (OHRP), available at: http://www.hhs.gov/ohrp/; accessed April 6, 2006.
20. *Grimes v. Kennedy Krieger Institute* (2001), Case 129, Maryland Court of Appeals.
21. Council for International Organizations of Medical Sciences (2006), International Ethical Guidelines for Biomedical Research Involving Human Subjects; available at: http://www.cioms.ch/frame_guidelines_nov_2002.htm; accessed April 6, 2006.
22. *Diamond v. Chakrabarty* (1980), 447 US 303-310. Diamond, Commissioner of Patents and Trademarks V. Chakrabarty. Certiorari to the United States Court of Customs and Patent Appeals. No. 79–136.
23. Resnick, D. (2001), DNA patents and scientific discovery and innovation: Assessing benefits and risk, *Sci. Eng. Ethics*, 7(1), 29–62.
24. Amdur, R., Kornetsky, and Bankert, L. (2002), Reviewing a research proposal, in Amdur, R. Ed., *The Institutional Review Board Member Handbook*, Jones and Bartlett, Sudbury, MA.
25. HSS (2006), U.S. Department of Health and Human Services, Food and Drug Administration, Guidance for industry. Using a centralized IRB review process in multicenter clinical trials; available at: http://www.fda.gov/cder/guidance/index.htm; accessed April 6, 2006.
26. U.S. Department of Health & Human Services (1993), Protecting Human Research Subjects: Institutional Review Board Guidebook; available at: http://www.hhs.gov/ohrp/irb/irb_guidebook.htm; accessed April 6, 2006.
27. World Health Organization (2000), Operations Guidelines for Ethics Committees That Review Biomedical Research; available at: http://www.who.int/tdr/publications/publications/pdf/ethics.pdf; accessed April 6, 2006.
28. Steinbrook, R. (2002), Heath Policy Report: Improving protection for research subjects, *N. Engl. J. Med.*, 346(18), 1425–1430.
29. National Archives and Records Administration (2006), *Electronic Code of Federal Regulations (e-CFR)*, Title 45, Public Welfare, Part 162.514, Administrative Requirements (45 CRF §162.514), Government Printing Office, Washington, D.C.; available at: http://ecfr.gpoaccess.gov/cgi/t/text/text-idx?c=ecfr&tpl=%2Findex.tpl; accessed April 6, 2006.
30. National Archives and Records Administration. (2002), *Code of Federal Regulations, Title 21, Chapter I—Food And Drug* Administration, Department of Health and Human Services, Part 54.4, Financial Disclosure by Clinical Investigators (21 CRF §54.4), Government Printing Office, Washington, D.C.
31. Federman, D. D., Hanna, K. E., and Rodriguez, L. L. (2003), *Responsible Research: A Systems Approach to Protecting Research Participants*, National Academies Press, Washington, D.C., p. 109.
32. Califf, R. M., Morse, M. A., Wittes, J., et al. (2003), Toward protecting the safety of participants in clinical trials, *Controlled Clin. Trials*, 24, 256–271.

# *Size of Clinical Trials*

**Jitendra Ganju**

*Amgen Inc., South San Francisco, California*

**Contents**

1 Introduction
  1.1 Challenges Faced in Practice
2 Statistical Methods for Determining Trial Size
  2.1 Difference in Means
  2.2 Differences in Means Adjusted for Baseline Data
  2.3 Differences in Proportions
  2.4 Time to Event
3 Realistic Assessment of Trial Size
  3.1 Misspecification of Rates
  3.2 Misrepresentation of Estimate of Treatment Effect and Variance
  3.3 Sensitivity to Projected Number of Events
  3.4 Deviations from Assumptions and from Protocol Procedures
4 Sample Size Reestimation
5 Points to Consider When Planning Size of Trial
  5.1 Noninferiority/Equivalence Trials
  5.2 Unequal Allocation
6 Conclusion
  References

## 1 INTRODUCTION

On the face of it, the number of subjects required for a clinical trial is obtained from a mechanical application of a formula. A number representing the trial size pops out after presumed values of efficacy and variance, and specified levels of false-

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

positive and false-negative rates are plugged into the formula. However, arriving at a well-planned trial size is much more than a routine calculation. Because the derivations of formulas for different analysis methods are amply and adequately covered in the literature, they will be kept to a minimum here. The emphasis instead will be on the many challenges that arise when practical aspects of the size of a trial are considered. For purposes of keeping the focus on certain practical aspects, the chapter will be limited to randomized trials comparing two treatments and unless stated otherwise, a 1:1 allocation ratio.

Scientific, strategic, operational, and ethical deliberations exert a push and pull on the appropriate number of subjects needed for a trial. The tension this creates is described later in the section. Section 2 will give a limited overview of the technical details. In Sections 3 and 4 methods to avoid certain pitfalls will be described. Section 5 will briefly discuss noninferiority/equivalence testing and unequal allocation.

The minimum inputs to trial size (better known as sample size) assessment are the presumed values of efficacy and variability, the type I error (false-positive) rate or $\alpha$ and the type II error (false-negative) rate or $\beta$. Power is $1 - \beta$, or in words, the probability of obtaining a statistically significant result assuming the presumed value of efficacy to be true. When power is held fixed, the number of subjects and $\alpha$ have an inverse relationship. Determining the sample size at fixed power is equivalent to determining the power at fixed sample size. Typically, for confirmatory phase III trials, the one-sided $\alpha$ is set at 0.025 and $\beta$ at either 0.1 or 0.2. The fact that $\alpha$ is chosen to be less than $\beta$ is a reflection of how we, the clinical trials community, including regulatory bodies such as the U.S. Food and Drug Administration (FDA), are more concerned about the risk of declaring an ineffective treatment effective than we are of the risk of failing to declare an effective treatment effective. We have agreed to reach our conclusions by risking on average, for every 40 trials, no more than 1 false positive or no more than 4–8 false negatives.

The presumed value of efficacy used in sample size calculations is stated under the alternative hypothesis for the primary variable (also referred to as the response or outcome variable or endpoint). Its complement, the null hypothesis, states that the presumed value of efficacy is 0—that is, the efficacy of the experimental treatment is no different from control. (But see noninferiority/equivalence testing in Section 5 for different formulations of the null and alternative hypotheses.) For example, for a trial comparing cholesterol-lowering drugs, the null hypothesis may state that the average low-density lipoprotein, LDL or "bad cholesterol," in the two groups is the same, and the alternative may state that the experimental treatment, compared to the control, lowers the average LDL by 15 mg/dL.

We now get into how the scientific, strategic, operational, and ethical matters exert a push and pull on how many subjects are appropriate for addressing the trial's objectives.

### 1.1 Challenges Faced in Practice

***Scientific*** Let $\Delta$ denote the presumed value of efficacy. (For our cholesterol example, we would write $\Delta = 15$ mg/dL.) Since the sample size formula relies heavily on the value specified for efficacy (the sample size is a function of $\Delta^2$), it is important

to investigate where Δ came from. In assessing its reliability, the following sorts of questions need to be investigated.

1. Did Δ come from an earlier trial conducted by the same institution/sponsor? If so, how is the trial being planned different from the earlier one? Was the earlier one a small-sized trial conducted in a few centers? Will the upcoming trial include a more heterogeneous population? Are there differences in the lengths of follow-up between the two trials? Is the control group the same?
2. If Δ was provided to a statistician by a clinician, what was their source of information?
3. Was Δ inferred from a published article that reported results on a recently approved treatment with a similar mechanism of action?
4. Could medical advances in the field influence the magnitude of Δ estimated from earlier trials? For example, organ transplant patients receive immunosuppressants to weaken the body's immune system to prevent it from attacking the transplanted organ. Advances in surgical methods for organ transplant may influence Δ independent of the drug effect.
5. If Δ was obtained from a peer-reviewed journal article, then taking that Δ at face value will likely be an optimistic value. Positive trials tend to get published, whereas negative trials tend not to. If we had results from positive and negative trials, then our presumed value of efficacy would weigh the evidence from all trials. Overlooking the fact that negative results get censored leads to an overly favorable Δ.

All things considered, if Δ seems optimistic, it would be advisable to err on the side of being conservative. Even in the simple case of Δ defined as the difference in means, the statistician, if only as a watchdog, has an important role to play in checking whether the disparate bits of information contribute a well-inferred assessment of the presumed value of efficacy.

If the primary variable is instead time to some event (e.g., time to heart attack in our cholesterol-lowering example), then other challenges are present. Here Δ is the ratio of hazards, a concept generally not readily understood by nonstatisticians when adjustments to it are made depending on the subject accrual rate, the follow-up duration, nonproportional hazards, censoring due to loss to follow-up, and competing risks. For such an endpoint the statistician needs to be even more involved in carefully projecting the number of events and in ensuring that Δ has been carefully selected.

Some of the same questions also arise when choosing a value for the variance. Let $s^2$ denote the estimate of variability to be used in sample size calculations that is associated with the estimate of Δ. Often in a covariate-adjusted analysis $s^2$ is obtained by conditioning on marginal totals (e.g., the number of subjects per center). The estimate is valid only when there is no interaction between treatment and the stratification factor. When interaction exists, the estimated treatment effect is biased and the estimated variance is an underestimate (Section 3).

Additionally, $s^2$ obtained from an earlier trial does not represent, in the statistical repeated sampling sense, an estimate of variability for the upcoming trial. No two trials are ever conducted under identical conditions subject only to random variation. For instance, the inclusion/exclusion criteria or the centers may be different or

the definition of the disease may have been refined over time. It is more appropriate, therefore, to refer to $s^2$ as a "guesstimate" of the true variance. The estimate–guesstimate delineation is a reminder that the upcoming trial will not be identical, except for random variation, to the completed one from which $s^2$ was calculated. This realization prepares us for questioning or challenging our assumptions while the trial is ongoing and possibly even reestimating the sample size from interim trial data.

Values of $\Delta$ and $s^2$ that get plugged into the sample size formula often assume compliance to the protocol. But the assumption of compliance itself is one to guard against. As Efron [1] notes: "There could not be worse experimental animals on earth than human beings; they complain, they go on vacations, they take things they are not supposed to take, they lead incredibly complicated lives, and, sometimes, they do not take their medicine." Too often little attention is paid at the *planning* stage about the consequences of protocol deviations on the power of the trial. Standard sample size formulas do not automatically adjust for imperfect compliance. When subjects prematurely terminate from the trial, their efficacy assessment at the final visit will be missing. If the intention is to analyze per the intent-to-treat principle—a principle that mandates inclusion of all randomized subjects—then the results depend on the imputation method. Careful planning of sample size should acknowledge the pervasiveness of missing data, make realistic projections on the proportion likely to be missing, and calculate the sample size consistent with the proposed imputation method.

Other challenges to sample size calculation arise when subjects midstream change their trial conduct. In blinded trials with long-term follow-up, some subjects may modify their behavior after they become unblinded to the blinded treatment. Unblinding can occur when subjects are able to infer the treatment administered by monitoring the trajectories of their outcomes. Upon becoming unblinded, subjects may then, for example, start taking prohibited study medications, thus confounding the causal assessment of efficacy. Adjusting for such a deviation in the sample size is difficult. The combination of even slightly unrealistic values for $\Delta$ and $s^2$, and deviations from protocol procedures can drastically impact the chance of achieving the trial's objectives. Section 3 discusses this in more detail.

***Strategic*** The size of a trial is generally governed by its primary objective. Sometimes, however, trialists have strategic objectives in mind that go beyond the trial's primary objective. These concern finding expeditious ways to license a drug or terminate the trial early for futility. This situation is best described with some examples. (1) Suppose a sponsor is initiating its phase II trial around the time news on the approval status of its competitor's drug is forthcoming. The sponsor adopts the following strategy. If the competitor receives approval, then the sponsor will amend the protocol to add another cohort of subjects who would be randomized to the competitor's newly approved drug or to the sponsor's experimental treatment. Results from the additional cohort would facilitate planning of the sponsor's phase III trials by giving them an "early read" of the efficacy of their experimental treatment relative to the newly approved one. If the competitor does not receive approval, then the trial size remains unchanged. (2) In a phase II trial with long-term follow-up, if enrollment turns out to be much slower than expected, the sponsor may no longer wish to wait for trial completion. Instead the sponsor may decide to test for superiority (or noninferiority) on a surrogate endpoint that is measured before the primary clinical endpoint, while continuing enrollment. If superiority (or noninfe-

riority) is established, then enrollment and follow-up for evaluation of the clinical endpoint continues, otherwise the trial terminates for futility. Because the trial may be underpowered with interim data, the sponsor may choose to lower the threshold for demonstration of its objective on the surrogate endpoint.

Such strategic objectives can influence the trial's size, but they may not get written into the trial protocol, particularly when the sponsor's thinking is still evolving at the time of initiating the trial.

***Operational*** Trials can be very expensive, especially late-phase trials. The National Heart, Lung, Blood Institute–funded Coronary Drug Project Trial, designed to assess lipid-influencing regimens, randomized 8341 subjects and was reported to cost $42 million [2], amounting to an average cost of approximately $5000 per subject. In 2010 this amount is about $11,000 per subject after adjusting for inflation, a cost by no means out of the ordinary for a typical trial. More expensive trials, particularly those sponsored by nascent pharmaceutical or biotechnology companies, may simply not get off the ground. (The possibility of joint sponsorship with institutions or larger companies is not always viable.) The solution, however, is not necessarily to reduce drastically the trial size because that would also lower the chance of success. Indeed, the failure to demonstrate success has been attributed to an inadequate number of subjects [3]. As aptly put by Lachin [4]: "Clinical trials with inadequate sample size are thus doomed to failure before they begin and serve only to confuse the issue of determining the most effective therapy for a given condition." A trial that has failed to demonstrate efficacy because it was not sufficiently sized, may simply not get repeated, risking termination of not just the trial but the development of a potentially important therapy. Innovative alternatives to reducing the trial cost need to be considered without substantially compromising statistical power.

***Ethical*** If trials could be done on plots of land, many of the challenges noted above would not arise. There would be full adherence to the protocol and ample data, none of it missing. But ethical concerns with human participation bring the size of trials into glare. Trials with mortality or irreversible morbidity as the primary variable lend themselves to extra scrutiny if they are too small or too large. Small-sized trials have a low probability of success and a high probability of equivocal results, whereas excessively large trials expose more subjects than necessary to at least one suboptimal or detrimental therapy. Ethical concerns sharply bring to focus the point that subjects are, after all, volunteers who provide data to help trialists decide fairly conclusively whether or not the experimental treatment (a) can proceed to the next stage for further testing or (b) is safe and effective for licensure.

## 2 STATISTICAL METHODS FOR DETERMINING TRIAL SIZE

In this section an outline of the methods for calculating the sample size for different analysis methods will be described. The methods will be described in some detail for differences in means and proportions, but sketched for the time-to-event endpoint. The level of detail in this section is intended to provide a basic understanding. For an in-depth understanding the references in this section and the next should be consulted. The excellent paper by Lachin [4] belongs to the "must read" category.

### 2.1 Difference in Means

Let $Z_\alpha$ and $Z_\beta$ denote the $\alpha$ and $\beta$ percentiles of the normal distribution, respectively. Power is mathematically expressed as

$$1-\beta = P(Z > Z_{1-a} | \Delta) \tag{1}$$

Un-normalizing the $Z$ test gives

$$1-\beta = P\left(\frac{\hat{\Delta}}{\sqrt{\mathrm{Var}(\hat{\Delta})}} > Z_{1-\alpha} | \Delta\right) \tag{2}$$

where $\hat{\Delta}$ is the estimated treatment effect and where $\mathrm{Var}(\hat{\Delta}) = 4\sigma^2/N$, with $\sigma^2$ denoting the variance of each observation and $N$ denoting the *total* trial size. The above equation simplifies to

$$1-\beta = P\left(Z > \frac{Z_{1-\alpha}\sqrt{\mathrm{Var}(\hat{\Delta})} - \Delta}{\sqrt{\mathrm{Var}(\hat{\Delta})}}\right) \tag{3}$$

If the two treatments are identical (i.e., $\Delta = 0$), then regardless of the size of the trial, power equals the false-positive rate $\alpha$. From (3)

$$Z_\beta = \frac{Z_{1-\alpha}\sqrt{\mathrm{Var}(\hat{\Delta})} - \Delta}{\sqrt{\mathrm{Var}(\hat{\Delta})}}$$

We almost always choose power to exceed 50%, so $Z_\beta$ is negative. Noting that $-Z_\beta = Z_{1-\beta}$ we solve for $N$ to get the total sample size:

$$N = \frac{4\sigma^2 (Z_{1-a} + Z_{1-\beta})^2}{\Delta^2} \tag{4}$$

For $\alpha = 0.025$ it is simpler to remember the total sample size as $N = 42/(\text{signal-to-noise ratio})^2$ for 90% power or $N = 32/(\text{signal-to-noise ratio})^2$ for 80% power where $\Delta/\sigma$ denotes the "signal-to-noise ratio." When testing on the difference scale a trial with 90% power requires about 30% more subjects than a trial with 80% power (and $\alpha = 0.025$).

The sample size calculation relies on the $Z$ test, which assumes that the variance is known. The data are analyzed, however, using the $t$ test instead, which is like the $Z$ test, except that the variance is estimated from the data. There is very little loss in power caused by the discrepancy in using one statistic ($Z$ test) for calculating power and another ($t$ test) for analyzing data. The $Z$ and $t$ tests make assumptions that (a) the variances associated with each treatment are equal, and (b) the data are normally distributed. Do violations of any, or both, of these assumptions make the sample size formula unreliable? There may be some loss in power due to violations of assumptions but (4) is still reliable. The $t$ test is remarkably robust to violations in assumptions. Posten, Yeh and Owen [5] study robustness when the equal variance

assumption is violated. A powerful conclusion they reach in the case of normal data and a sample size of 15 per group is that "no matter how much $\sigma_1^2$ varies from $\sigma_2^2$, the true significance level [of the *t* test] is within 0.01 units of the desired level $\alpha = 0.05$." Heeren and D'Agostino [6] demonstrate the robustness of the *t* test for data on 3, 4, or 5-point ordinal scales. Aside from invoking normality to analytically calculate power, power may also be empirically calculated by simulating data under the alternative for the type of nonnormal data at hand (e.g., ordinal or skewed data).

Some investigators [7] have argued that the sample size calculated by the above method is "really large." They say that this method "permits an illogical interpretation that tempts investigators and plagues readers. The large sample size can lead one to declare an unimportant difference important." What they mean by this is that an unimportant but statistically significant difference in means may be interpreted as clinically meaningful. It would be inappropriate to conclude, as the authors assume readers are wont to do, that a statistically significant result is clinically meaningful. They instead advocate the following sample size that was proposed by Feinstein [8]:

$$N \geq \frac{4\sigma^2 Z_{1-\alpha}^2}{\Delta^2} \tag{5}$$

According to them power is not the issue—instead one needs to pick an "intellectually honest" $\Delta$ before data collection and then draw inferences by reference to $\Delta$ and 0. The following explains their recommendation (wherein positive values of $\Delta$ signify improvement, LL and UL denote the lower and upper confidence limits around $\hat{\Delta}$):

If $\hat{\Delta} > \Delta$ and LL > 0, conclude an important difference.
If $\hat{\Delta} < \Delta$ and LL < 0, concede the null.
If $\hat{\Delta} < \Delta$ but LL > 0 and UL > $\Delta$, conclude an important difference may exist.
If LL > 0 and UL < $\Delta$, conclude result is unimportant even though statistically significant.

Their approach, however, is not feasible. First, as an inequality (5) does not indicate how large *N* should be relative to $4\sigma^2 Z_{1-\alpha}^2/\Delta^2$. Second, power is paramount, although the authors do not believe in it. In the end it is necessary to know what chance of success a sponsor has of meeting the trial's objectives. For example, with $\Delta = \sigma = 1$, an *N* equal to $4\sigma^2 Z_{1-\alpha}^2/\Delta^2$ would require 16 subjects ($\alpha = 0.025$). The power of declaring that $\hat{\Delta}$ is statistically significantly different from 0 is approximately 50%; the power of declaring that an important difference exists (as they have defined it) would be even less. Why would a sponsor undertake a trial that has less than a 50% chance of finding an important difference? Third, presupposing an "intellectually honest" $\Delta$ belies the reality that proposing a carefully thought-through $\Delta$ may involve the weighing of evidence from fragmentary and conflicting data. It would be unrealistic to expect a sponsor, a regulatory agency, or a committee advising the agency to all agree on the same value of $\Delta$. If clinical rather than statistical significance is of interest, then it would be better to power the trial so that the LL

exceeds the clinically important difference. However, the power required for a clinically important difference would result in an even *larger* sample size than (4).

### 2.2 Differences in Means Adjusted for Baseline Data

When baseline data on the same endpoint are available, variability can be reduced by taking the change from baseline or by fitting an analysis of covariance (ANCOVA) model. Under standard assumptions ANCOVA improves variability by nearly $1 - r^2$ [9] and $0.5(1 + r)$ [10] compared to the unadjusted and change from baseline models, respectively, where $r$, which takes values between –1 and 1 [11], denotes the correlation of the baseline and posttreatment data. (For pre/post data, $r$ usually takes values between 0 and 1.) The sample size for the ANCOVA model is obtained as $N_{\text{adj}} = N(1 - r^2)$ where $N$ is as shown in (4). Since ANCOVA results in a smaller variance than change from baseline, there is no reason to continue to use change from baseline as the endpoint.

### 2.3 Difference in Proportions

A simple method for calculating the sample size for comparing the difference in proportions makes use of the fact that the proportions are approximately normally distributed. The derivation proceeds along the same lines as for the difference in means with one important distinction explained below.

The null and alternative hypotheses are: $H_0$: $p_1 - p_2 = 0$ and $H_1$: $p_1 - p_2 = \Delta$, where $p_1$ and $p_2$ denote population proportions. The estimate of the treatment effect is $\hat{\Delta} = \hat{p}_1 - \hat{p}_2$. The important distinction is that there are two formulas for the variance of $\hat{\Delta}$ that feature in the sample size formula: one under the null hypothesis and the other under the alternative. We start by rewriting (2) as

$$1-\beta = P\left(\frac{\hat{\Delta}}{\sqrt{\mathrm{Var}_{H_0}\left(\hat{\Delta}\right)}} > Z_{1-\alpha} \,|\, \Delta\right)$$

where the variance under the null is $\mathrm{Var}_{H_0}\left(\hat{\Delta}\right) = (4/N)\,\bar{p}(1-\bar{p})$ and $\bar{p} = (p_1 + p_2)/2$ is the average of the proportions specified under the alternative. Recognizing that power is calculated under the alternative, this can be reexpressed as

$$1-\beta = P\left(Z > \frac{Z_{1-\alpha}\sqrt{\mathrm{Var}_{H_0}\left(\hat{\Delta}\right)} - \Delta}{\sqrt{\mathrm{Var}_{H_1}\left(\hat{\Delta}\right)}}\right) \tag{6}$$

where the variance under the alternative is

$$\mathrm{Var}_{H_1}\left(\hat{\Delta}\right) = \frac{2}{N}\{p_1(1-p_1) + p_2(1-p_2)\}$$

The total sample size derived from (6) is

$$N = \frac{\left[Z_{1-\alpha}\sqrt{4\bar{p}(1-\bar{p})} + Z_{1-\beta}\sqrt{2p_1(1-p_1) + 2p_2(1-p_2)}\right]^2}{\Delta^2} \tag{7}$$

Noting that $\mathrm{Var}_{H_0}\left(\hat{\Delta}\right) \geq \mathrm{Var}_{H_1}\left(\hat{\Delta}\right)$ a more conservative sample size formula is obtained by replacing $\mathrm{Var}_{H_1}\left(\hat{\Delta}\right)$ in (6) with $\mathrm{Var}_{H_0}\left(\hat{\Delta}\right)$ giving

$$N = \frac{4\bar{p}(1-\bar{p})(Z_{1-\alpha} + Z_{1-\beta})^2}{\Delta^2} \tag{8}$$

Formulas (7) and (8) yield approximate but not exact power. This is because the difference in proportions has a discrete distribution that is only approximately normal. Adjustments have been proposed so that power of the test gets closer to the stated value of $1 - \beta$ [12]. The formula implemented in the sample size software nQuery Advisor [13] is the one proposed by Fleiss, Tytun, and Ury [14]; their formula is

$$N' = N + \frac{4}{|p_1 - p_2|} \tag{9}$$

where $N$ is shown in (7).

When either proportion is expected to be very small (≤0.05) or very large (≥0.95), Fisher's or Barnard's [15] exact test is used for analysis because tests that rely on the normal approximation are less reliable. The sample size calculation should then reflect the power of the exact test and not one based on the normal approximation. It is beyond the level of technical detail intended for this chapter to describe the exact methods, but the interested reader is referred to [16]. Gordon [17] provides an excellent review of the sample size methods for comparing proportions.

### 2.4 Time to Event

Instead of analyzing data on a binary scale, the time to the occurrence of the event (i.e., time to event) is often a more powerful analysis [18]. The sample size calculation is sketched for one of the more commonly used time-to-event test statistics, the log-rank test [19].

The concept of a hazard is central in time to event analysis. Loosely put, the hazard at time $t$ is the instantaneous risk that the event will occur at time $t$. A large portion of the sample size literature rests on the assumption that the hazard of an event at any given time for one treatment group is proportional to the hazard at that time for the other treatment group. The assumption of proportional hazards enables the derivation of the sample size formula for the log-rank statistic, although once the data become available, the log-rank test remains valid even if that assumption is unmet.

The null and alternative hypotheses are: $H_0$: $\lambda_1(t)/\lambda_2(t) = 1$ and $H_1$: $\lambda_1(t)/\lambda_2(t) = \Delta$, where $\lambda(t)$ denotes the hazard of an event at time $t$ for an individual. On a log scale we test $H_0$: $\log_e \lambda_1(t) - \log_e \lambda_2(t) = 0$ against $H_1$: $\log_e \lambda_1(t) - \log_e \lambda_2(t) = \log_e \Delta$. For large samples, the log-rank statistic has an approximate normal distribution with mean $\log_e \Delta$ and, under a 1:1 allocation ratio, variance $4/d$. Following the same steps in this section in going from Equations (1)–(3), we see that the variance of the log-rank statistic can also be written as $(Z_{1-\alpha} + Z_{1-\beta})^2/(\log_e \Delta)^2$. Therefore, the

number of events required to achieve power $1 - \beta$ at a type I error rate of $\alpha$ is $d = [4(Z_{1-\alpha} + Z_{1-\beta})^2]/(\log_e \Delta)^2$.

The required number of subjects is obtained from $N = d/P(\text{event})$ where $P(\text{event})$ is the probability of the event and is a function of efficacy, the accrual rate, and follow-up period. For randomized trials $P(\text{event})$ is the average of $P(\text{event in group 1})$ and $P(\text{event in group 2})$. To obtain this probability for each treatment group, let $a$ denote the accrual period, $f$ denote the follow-up period going from the end of accrual to a fixed period (in the same time unit), and assume that enrollment occurs at a uniform rate. Then it can be shown that $P(\text{event in group } i)$ equals

$$1 - 1/6\{P_i(T > f) + 4P_i(T > 0.5a + f) + P_i(T > a + f)\} \tag{10}$$

for $i = 1,2$ [20]. In the equation above, $T$ is a random variable denoting the time to event, and the notation $P_i(T > x)$ denotes the probability that the time to event in group $i$ exceeds $x$. When enrollment is simultaneous, $a = 0$, and $P(\text{event in group } i)$ simplifies to $1 - P_i(T > f)$. When enrollment is staggered $P_i(T > f)$, $P_i(T > 0.5a + f)$, and $P_i(T > a + f)$ are all decreased as the accrual period increases, and (10) shows how this increases the probability of the occurrence of an event.

Often a terminal event such as death or irreversible morbidity is the "event" in time-to-event trials. (Such trials also get the most media attention.) It is therefore important to understand the concept of hazards, the derivation of the formula, and the assumptions that the formula depends upon beyond the silhouette presented above. A highly accessible account of the sample size derivation is in Collett [20]. Other useful references include George and Desu [21], Schoenfeld [22], and Lachin and Foulkes [23]. A novel method to calculate the sample size that does not make any of the restrictive assumptions is discussed in Section 3.

## 3 REALISTIC ASSESSMENT OF TRIAL SIZE

PhRMA (the Pharmaceutical Research and Manufacturers of America) says that of drugs completing Phase II trials, about 50% fail in Phase III, often for lack of efficacy (quoted in [24]). But Temple [24] points out that phase II trials are meant to demonstrate efficacy, so it is startling that the phase III failure rate is as high as 50%. It may well be that when sizing phase III trials misjudgments in $\Delta$ or $s^2$ contribute to the high failure rate. This section will discuss the sensitivity of the sample size to various assumptions regarding $\Delta$, $s^2$, and protocol compliance.

### 3.1 Misspecification of Rates

A simple hypothetical example is presented to help understand better why the sample size (or, equivalently, power) is so sensitive to the initial assumptions. Table 1 displays the sample sizes for 80 and 90% power for testing $H_0$: $p_1 - p_2 = 0$ against $H_1$: $p_1 - p_2 = 0.07$. As explained in Section 2, to calculate the sample size or power we also need the presumed population proportions. Suppose we believe that $p_1 = 0.12$ and $p_2 = 0.05$ are the true population proportions. Table 1 shows that we need a total of 552 subjects to have 80% power and 720 subjects for 90% power. If in reality the population proportions are instead $p_1 = 0.13$ and $p_2 = 0.06$, a slight change from our

**TABLE 1 Sample Sizes under Various Assumptions ($\alpha = 0.025$)**

| | | | | |
|---|---|---|---|---|
| $p_1$ | 0.12 | 0.13 | 0.14 | 0.13 |
| $p_2$ | 0.05 | 0.06 | 0.07 | 0.07 |
| $p_1 - p_2$ | 0.07 | 0.07 | 0.07 | 0.06 |
| $N$ (power = 80%) | 552 | 606 | 656 | 848 |
| $N$ (power = 90%) | 720 | 790 | 860 | 1114 |

*Note*: Samples sizes obtained from nQuery Advisor Version 5.0 or equivalently from (9).

**TABLE 2 Outcome Is Measure of Neurological Functioning**

| | | Treatment A | | | Treatment 2 | | |
|---|---|---|---|---|---|---|---|
| Stratum | Social Class and Gender | $n_A$ | $\bar{y}_A$ | $\hat{\sigma}_A$ | $n_B$ | $\bar{y}_B$ | $\hat{\sigma}_B$ |
| 1 | Low, female | 41 | 1.38 | 0.22 | 40 | 1.36 | 0.28 |
| 2 | Low, male | 41 | 1.26 | 0.25 | 38 | 1.28 | 0.19 |
| 3 | Medium, female | 33 | 1.51 | 0.31 | 35 | 1.41 | 0.27 |
| 4 | Medium, male | 45 | 1.46 | 0.28 | 46 | 1.39 | 0.33 |
| 5 | High, female | 18 | 1.61 | 0.34 | 20 | 1.51 | 0.41 |
| 6 | High, male | 23 | 1.59 | 0.46 | 23 | 1.44 | 0.30 |

*Source*: From Fleiss [25].

assumptions although the difference of 0.07 is maintained, we need 606 and 790 subjects for 80 and 90% power, respectively. The sample size increases even though $\Delta$ is unchanged because $\mathrm{Var}(\hat{\Delta})$ increases. The next column shows a similar increase in sample sizes (656 and 860 subjects for 80 and 90% power) for a similar increase in the proportions. The last column shows how significant the impact can be if both the proportions and the difference in proportions are slightly off target. For $p_1 = 0.13$ and $p_2 = 0.07$ we need 848 and 1114 subjects, respectively, for 80 and 90%. (Note that in each case the sample size for 90% power is roughly 30% more than the sample size for 80% power.)

### 3.2 Misrepresentation of Estimate of Treatment Effect and Variance

Consider the data in Table 2 on 403 subjects displayed by social class and gender for two treatments [25]. We assume the data were obtained from a trial that employed the method of simple (permuted block) randomization. Further assume that the design of the upcoming trial is similar to the completed one. When data become available for the upcoming trial, the analysis will adjust for the six stratification factors to reduce variability. Although the data in Table 2 suggest that there is no evidence of treatment by stratum (social class or gender) interaction, we will assume for the sake of illustration that such an interaction exists. We use this information to estimate the variability of $\hat{\Delta}$.

It is common to estimate the treatment effect as $\hat{\Delta} \sum_{i=1}^{6} w_i \hat{\Delta}_i \Big/ \left(\sum_{i=1}^{6} w_i\right)$ where $\hat{\Delta}_i$ is the estimated effect for the $i$th stratification level, and where $w_i$ is a weight that is a function of the number of subjects in each cell in the $i$th stratum. For example, for $i = 1$, $\hat{\Delta}_1 = 1.38 - 1.36 \equiv 0.2$ and $w_1 = 41 \times 40/(41 + 40) \equiv 20.25$. When interaction exists, which is another way of saying that $\Delta_i$'s are not all equal, then $\hat{\Delta}$

is biased [26]. The reason $\hat{\Delta}$ is biased is because the weights, $w_i$'s, are treated as fixed constants, whereas according to the trial design the weights are random. (If randomization was stratified within each stratum, then $\hat{\Delta}$ is unbiased [27]). Moreover, the pooled estimate of variance is as an underestimate. The pooled variance estimate, $s^2$, is calculated as a weighted average of the variance in each cell in Table 2:

$$\left\{\begin{matrix}(41-1)\times 0.22^2+(40-1)\times 0.28^2 \cdots \\ +(23-1)\times 0.30^2\end{matrix}\right\}\Big/(40+39+\cdots+22)=0.0883$$

Typically, this is the value of $s^2$ that gets plugged in (4) to calculate the sample size. This variance is known as the conditional variance where the condition is the observed number of subjects in each of the 12 cells. Because $\hat{\Delta}=\sum_{i=1}^{6} w_i\hat{\Delta}_i \Big/ \left(\sum_{i=1}^{6} w_i\right)$ is biased when interaction exists, the unweighted and unbiased estimate of the treatment effect may be preferred. Accordingly, we need to use the appropriate variance formula. It has been shown that [26] the unconditional variance is

$$\mathrm{Var}\left(\hat{\Delta}\right)\approx\frac{4}{N}\left(\sigma^2+\frac{K}{2}\right)$$

where $K=\sum_{i}^{6}\pi_i\left\{(\mu_{Ai}-\mu_A)^2+(\mu_{Bi}-\mu_B)^2\right\}$; $\pi_i$ denotes the proportion of subjects who belong to the $i$th stratification level ($i = 1, 2, \ldots, 6$), $\mu_{Ai}$ and $\mu_{Bi}$ denote the population means of treatments A and B for the $i$th strata with $\mu_A = \Sigma\pi_i\mu_{Ai}$ and $\mu_B = \Sigma\pi_i\mu_{Bi}$ denoting the population means. The unconditional variance adds $K/2$ to the pooled variance $\sigma^2$ to free the conditional variance from requiring inferences limited to fixed cell sizes. Just as the data in Table 2 provided an estimate of $\sigma^2$, so too we use the data to estimate $\pi_i$, $\mu_{Ai}$, $\mu_{Bi}$, $\mu_A$, and $\mu_B$. For example, the estimate of the proportion of subjects who belong to the "medium, female" level, $\pi_3$, is 68/403 = 0.17 and the estimates of $\mu_{A3}$ and $\mu_{B3}$ are 1.51 and 1.41, respectively. Substituting estimates for population parameters, we get $K/2 = 0.0065$, so the pooled estimate of variance goes from 0.0883 to 0.0883 + 0.0065 = 0.0968.

The sample size formula corresponding to the unconditional variance becomes

$$N=\frac{(4\sigma^2+2K)(Z_{1-\alpha}+Z_{1-\beta})^2}{\Delta^2} \tag{11}$$

Using the conventional sample size formula (4) gives a total sample size $N$ of 772 subjects to achieve 80% power, whereas (11) requires 964 subjects to achieve the same power, an increase of 25%. Put another way, the power that will be achieved with 772 subjects is 70% and not, as claimed, 80%. The tradition of using an unconditional variance for sample size is common in survey sampling but not, however, in the clinical trials literature. [Because the population proportions are known in survey sampling, the unconditional variance expression is different from (11)]. For survey sampling the unconditional variance has been recommended by none other than Deming [28] who says that it is the "formula which one will use at the planning stage."

### 3.3 Sensitivity to Projected Number of Events

A large vaccines trial was designed to reduce the burden of illness due to herpes zoster in a population at least 60 years of age with a minimum of 6 months of follow-up [29]. Approximately 38,000 subjects included in the analysis received either an investigational vaccine or placebo, in a 1:1, randomized, double-blind manner. Although the primary analysis variable for the trial was the relative reduction in the herpes zoster burden of illness score, we will consider the relative reduction in the proportion of subjects with herpes zoster.

There were a total of 957 confirmed cases of herpes zoster included in the analysis, 315 among vaccine recipients and 642 among placebo recipients. The relative risk is 0.49 and the 95% confidence interval (CI) is (0.43, 0.56). Because the interval excludes 1, the result is statistically significant. (In fact, since the upper limit is much below 1, the investigational vaccine substantially reduced the rate of herpes zoster.) Table 3 shows the observed result in the first row and hypothetical ones in the next two rows. Comparing the first and second rows, we see that width of the CI is the same even though $N$ in the second row is less by 28,000 subjects. The massive reduction in sample size has no effect on the width of the interval because the variance of ratios of estimates of event rates is driven by the number of events (and only trivially by the total sample size). Comparing the second row with the third, we see that $N$ is the same but the number of cases are halved and that this only slightly increases the width of the interval. Although the reduction in number of cases from 957 to 450 is large, the result is, in effect, unchanged. This is because the relative risk $Z$ values for both are very large, raising questions whether the trial was overpowered. An interim analysis allowing for early termination upon demonstration of a significant reduction in herpes zoster cases may have markedly reduced the trial size. Indeed, if mortality were the outcome variable, an interim analysis would be mandatory so that the trial not enroll more subjects than necessary.

### 3.4 Deviations from Assumptions and from Protocol Procedures

Here we consider sample size assessment when the outcome variable is binary or time to some event and when there is loss to follow-up, noncompliance, "drop-in," nonproportional, or nonconstant hazards. (Drop-in refers to subjects changing their randomized treatment in violation of protocol procedures.) For binary outcomes Lakatos [30] has presented a method for calculating the sample size under such realistic conditions. His method uses a Markov model for adjusting the proportions, $p_1$ and $p_2$, for losses due to noncompliance, drop-in, and so forth. Each treatment group is modeled separately. In Lakatos' Markov model, a transition matrix is

**TABLE 3 Number of Cases of Herpes Zoster,[a] Relative Risk Estimates and 95% CI**

| Sample Size | Vaccine | Placebo | Relative Risk | 95% CI |
|---|---|---|---|---|
| $N \approx 38{,}000$ | 315 | 642 | 0.49 | (0.43, 0.56) |
| $N^* = 10{,}000$ | 315 | 642 | 0.49 | (0.43, 0.56) |
| $N^* = 10{,}000$ | 150 | 300 | 0.50 | (0.41, 0.61) |

*Note*: The number randomized to each group is roughly $N/2$. The first row shows real data, the next two rows (indicated by $N^*$) show hypothetical data.
[a]From [29].

created for each time interval (the length of the time interval is user specified). The rows and columns of the transition matrices are states subjects belong to and will transition to with probabilities that are also specified by the user. Typical examples of states that subjects belong to include: a loss to follow-up state (no further information available for that subject), an event state (subject has had the event), an at-risk state for a subject who is a complier (subject is at risk of experiencing the event assuming compliance), and an at-risk state for a subject who is a noncomplier (subject is at risk of experiencing the event assuming noncompliance). Needless to say, subjects would transition from at-risk states to the same or other states, but not from a loss to follow-up or an event state to an at-risk state. At the beginning of the trial the probability of belonging to the at-risk state is 1 for compliers and is 0 for all other states. At the end of the trial the probability of belonging to any particular state is the Markov process, obtained as the product of the transition matrices. The adjusted proportions, $p_1$ and $p_2$, obtained from the Markov process for the end of the trial get plugged in the sample size formula.

Lakatos compares the sample size for the SHEP (systolic hypertension in the elderly program) trial calculated the traditional unadjusted way and upon application of the Markov model. The outcome variable for the trial was fatal or nonfatal stroke. Subjects were to be followed for a minimum of 4 years. The experimental treatment was assumed to lower the rate of stroke relative to control by 40%. Without adjusting for losses, noncompliance, and drop-ins the control and treatment rates were assumed to be 0.0775 and 0.0471, respectively. Application of (9) results in a sample size of 2784 to achieve 90% power ($\alpha = 0.025$). After making reasonable adjustments for noncompliance, losses to follow-up, and the like, the Markov model gave rates of 0.0677 and 0.0463. With these rates, the required sample size is 5116.

Lakatos later [31] extended his method for the log-rank statistic enabling calculation of the sample size under unrestrictive conditions. Unlike other methods in the literature, Lakatos' method does not rely on the assumption of proportional hazards. Indeed, the advantage of this method is that it can accommodate any arbitrary pattern of data projected for the future. The difference in sample sizes calculated with and without adjustments for loss to follow-up, noncompliance, and drop-in for nonproportional hazards data is striking. For example, in Table 2 of Lakatos [30], assuming constant recruitment the sample size to achieve 90% power is 4397 ($\alpha = 0.025$). For the same constant recruitment assumption but adjusting for a lag effect (i.e., a certain type of nonproportional hazard), modest loss to follow-up, noncompliance, and drop-in rates the required sample size is 8009.

These examples exemplify the point that trialists should pay careful consideration at the planning stage of the impact on power when rates are misspecified or when protocol compliance is less than perfect.

## 4 SAMPLE SIZE REESTIMATION

Methods have been proposed to estimate the variance while the blinded trial is still ongoing to allow for reestimation of the sample size. One option is to break the treatment code after data on some fraction of subjects has become available

and estimate the variance. However, because this step unblinds the trial, it has to be done with much forethought. One has to be concerned about protecting the integrity of trial results at its scheduled termination due to interim unblinding. Here we discuss a method proposed for continuous outcome variables that does not require breaking the blind [32, 33]. To apply the method it suffices to know the enrollment order and randomization block size. The method is simple to apply and works reasonably well for small randomization block sizes when compared to its unblinded counterpart. It works as follows. At an interim stage we have data on the outcome variable for, say, $\tilde{N}$ subjects. Suppose the randomization block size is $n$ and there are $k$ blocks, so that $\tilde{N} = nk$. Denote the data on the outcome variable as

$$Y_{11}, Y_{21}, \ldots, Y_{n1} | Y_{12}, Y_{22}, \ldots, Y_{n2} | \ldots | Y_{1k}, Y_{2k}, \ldots, Y_{nk}$$

The data structure written this way means that of the first set of observations, $Y_{11}, Y_{21}, \ldots, Y_{n1}$, half of those belong to one treatment group and the other half to the other treatment group. Of the second set of $n$ observations, $Y_{12}, Y_{22}, \ldots, Y_{n2}$, half belong to one group and half belong to the other, and so on until the last set $Y_{1k}, Y_{2k}, \ldots, Y_{nk}$.

Take the sum of the $Y$'s within each set. For the $j$th set denote this sum by $T_j$. The blinded variance estimator is (variance of $T_j$)/block size or:

$$\tilde{\sigma}^2 = \frac{1}{n} \frac{\sum_{j=1}^{k} \left(T_j - \bar{T}\right)^2}{k-1} \tag{12}$$

This blinded variance estimator is unbiased and achieves minimum variance—that is, the variation in $\tilde{\sigma}^2$ is smallest—when $k$ equals the number of treatments. The following example demonstrates how the method works.

Normally distributed data of size 20 with seed 1234 were generated in SAS version 8.2 with mean 0 and variance 1 for group 1 and mean 2 and variance 1 for group 2. If the means are equal, then blinded data estimation of variability would amount to unblinded estimation. For this reason the means chosen were sufficiently different. The data are shown in Table 4. Assume the randomization block size is 2. Data that belonged to the blocks were taken to be those generated sequentially by SAS for the chosen seed. The first two observations, one from each group, were assigned to block 1, the next two to block 2, and so on. The unblinded estimates of the variances of the data in groups 1 and 2 are 1.19 and 1.81, respectively. Since they are estimating the same variance, we take their average, 1.24, as the unblinded variance estimate. From (12) the blinded variance estimate is 1.28, a result similar to the unblinded estimate. As the block size gets larger, the performance of the blinded variance estimator is less impressive. For normally distributed data, the ratio of the standard deviation of the blinded variance estimator relative to its unblinded counterpart is $\sqrt{n\left(\tilde{N}-2\right)/\left(\tilde{N}-n\right)}$; for $n = 2$ the ratio equals $\sqrt{2}$ [32]. This method based on block sums has been extended to estimate the variance after adjusting for covariates [33]. Other simple methods are also available for estimating the within-group variance [34, 35]. These methods depend, however, on guessing the true treatment effect and are limited to two-treatment trials.

**TABLE 4 Blinded Variance Reestimation Example**

| Block | Group 1 | Group 2 | Block Sum |
|---|---|---|---|
| 1 | 0.277 | 3.462 | 3.74 |
| 2 | 0.299 | 0.293 | 0.59 |
| 3 | 0.895 | 1.396 | 2.29 |
| 4 | 0.591 | 2.857 | 3.45 |
| 5 | –0.331 | 2.874 | 2.54 |
| 6 | –1.839 | 1.990 | 0.15 |
| 7 | –1.261 | 1.321 | 0.06 |
| 8 | 0.902 | 0.867 | 1.77 |
| 9 | –0.701 | 1.378 | 0.68 |
| 10 | –0.829 | 3.049 | 2.22 |
| 11 | –1.21 | 0.141 | –0.98 |
| 12 | 0.657 | 4.126 | 4.78 |
| 13 | –1.487 | 4.192 | 2.71 |
| 14 | –1.409 | 1.880 | 0.47 |
| 15 | 0.648 | 1.491 | 2.14 |
| 16 | 1.246 | 2.351 | 3.60 |
| 17 | –1.072 | 2.978 | 1.91 |
| 18 | –0.067 | 1.913 | 1.85 |
| 19 | 0.925 | 2.208 | 3.13 |
| 20 | –2.490 | 1.393 | –1.10 |

*Note*: Group 1 is $N(0, 1)$ and group 2 is $N(2, 1)$. Data generated in SAS (version 8.2) using seed 1234. The unblinded pooled variance is 1.24, the blinded variance, for block size 2, is 1.28.

## 5 POINTS TO CONSIDER WHEN PLANNING SIZE OF TRIAL

The topics discussed above will be summarized in this section after briefly discussing noninferiority/equivalence trials and unequal allocation.

### 5.1 Noninferiority/Equivalence Trials

When the control is a marketed product, it is not uncommon to design a noninferiority trial in which the objective is to demonstrate that the experimental treatment is not inferior to control. For example, a sponsor may wish to demonstrate that the once-a-day dosing regimen of its experimental treatment is not inferior to the twice-a-day dosing regimen of an approved drug. The critical issue is the choice of a "margin" that defines noninferiority. For superiority trials, the margin is unambiguously defined to be 0; in other words, to establish superiority, the objective is to demonstrate that $\hat{\Delta}$ is statistically significantly different from 0. For noninferiority trials the objective is to demonstrate that $\hat{\Delta}$ is statistically significantly different from some margin $\delta$. For example, to demonstrate whether the experimental treatment is not inferior to a marketed control, the sponsor may wish to test whether the difference (experimental minus control) in the proportion of subjects surviving at the end of the trial is greater than $\delta = -10\%$. A regulatory agency, such as the FDA, may consider it more appropriate if instead $\delta = -5\%$. The choice of $\delta = -5\%$ is a more difficult objective to meet than $\delta = -10\%$ and requires a larger sample size.

Equivalence trials involve a lower and an upper margin, $(\delta_L, \delta_U)$. An example of an equivalence trial is a vaccine lot consistency trial whose goal is to demonstrate

that the vaccine lots (as assessed, e.g., by the titer values obtained in a clinical trial across the different vaccine lots) are equivalent.

The null hypothesis for noninferiority is $H_0$: $\Delta \leq \delta$ and for equivalence is $H_0$: $\Delta \leq \delta_L$ or $H_0$: $\Delta \geq \delta_U$ with alternative hypotheses, respectively, $H_1$: $\Delta > \delta$ and $H_1$: $\delta_L < \Delta < \delta_U$. The sample size derivation for two-treatment noninferiority/equivalence trials is conceptually similar to that for superiority. However, since the sample size is calculated under the alternative hypothesis, a value of $\Delta$ needs to be specified. In some equivalence trials it has been common to let $\Delta = 0$; this is often a mistake. For vaccine lot consistency (or equivalence) trials, it is demonstrated why the $\Delta = 0$ assumption may lead to a severely underpowered trial [36]. Following the steps in Section 2 for calculating the sample size for noninferiority/equivalence will make evident why the choice of the margin(s) influences the trial size. In particular, the closer the margin(s) is to 0, the larger the sample size. Before planning the size of a noninferiority or equivalence trial, the sponsor needs to be confident in its choice of margin(s). Useful references for determining the sample size in noninferiority/equivalence settings include Farrington and Manning [37] and Nam [38].

### 5.2 Unequal Allocation

Although the variance is minimized when the allocation to the two treatments is equal, in large placebo-controlled trials it is not uncommon to allocate more subjects to the experimental treatment. Such allocation provides more safety data for the experimental treatment in subjects who are representative of the population to be treated. In small- to midsized dose-ranging trials where the objective is often to compare increasing doses of the drug to control, it is statistically preferable to allocate more subjects to control, yet it is not uncommon to allocate an equal number of subjects to each group. Such an allocation provides more data on various doses of the drug at a stage in the drug's development when there is little prior evidence of its dose-ranging activity or efficacy. In the large or the dose-ranging trial pragmatism was elevated over statistical idealism. Before determining the sample size, first an allocation ratio deemed to be pragmatically (but not necessarily statistically) appropriate should be determined.

## 6 CONCLUSION

In conclusion, the gist of the chapter can be organized around a few assertions. Determining the number of subjects needed is much more than a calculation. Unfortunately, too often the sample size assessment is done mechanically. The inputs to the sample size formula require carefully combining pieces of information. This is where collaboration between statisticians and nonstatisticians is most important. In general, the presumed values of efficacy and variance should be on the conservative side. The impact of missing data should be considered upfront. As stated by Lavori [39]: "Do not expect nature to be kind ... Many power calculations are based on expected causal differences (what would happen under complete control), and not on expected practical differences (when uncontrolled extra treatments or undelivered study treatments intervene)." It is very helpful to calculate sample sizes under various values of presumed efficacy and variance after adjusting those values by the

intended imputation method for missing data expected in the trial. When the trial is ongoing, examine the blinded data for noncompliance. The impact of disproportionate noncompliance should be assessed. If possible, estimate the variance without breaking the blind after applying the imputation method that was stated in the trial protocol. If the blinded variance estimate is very different from the assumed variance, then recalculate the power of the trial and evaluate the options. Monitor the number of events for time-to-event or relative risk analyses, recalling that the variance is a function of the number of events not the total number of subjects. It will be informative to compare the total number of events observed during the blinded portion with the number that was projected in the sample size calculation.

The calculation of the size of a trial, like any other learning process, is iterative, and not a one-step solution. A more reliable sample size will be obtained if the factors that influence it are given due attention and subject to debate, the assumptions challenged, the blinded data evaluated than if the calculation is performed mechanically.

## REFERENCES


1. Efron, B. (1998), Foreword: Limburg compliance symposium, *Stat. Med.*, 17, 249–250.
2. Canner, P. L. (1984), How much data should be collected in a clinical trial? Experience of the coronary drug project, *Stat. Med.*, 3, 423–432.
3. Friedman, J., Chalmers, T., Smith, H., and Kuebler, R. (1978), The importance of Beta, the Type II error and sample size in the design and interpretation of the randomized controlled trial, *N. Engl. J. Med.*, 299, 690–694.
4. Lachin, J. (1981), Introduction to sample size determination and power analysis for clinical trials, *Controlled Clin. Trials*, 2, 93–113.
5. Posten, H. O., Yeh, H. C., and Owen, D. B. (1982), Robustness of the two-sample t-test under violations of the homogeneity of variance assumption, *Commun. Stat. Theory Methods*, 11, 109–126.
6. Heeren, T., and D'Agostino, R. B. (1983), Robustness of the two independent samples t-test when applied to ordinal scaled data, *Stat. Med.*, 6, 79–90.
7. Jacobson, R., and Poland, G. A. (2005), Sample sizes and negative studies in clinical vaccine research, *Vaccine*, 23, 2318–2321.
8. Feinstein, A. (1984), *Principles of Medical Statistics*, Chapman and Hall/CRC, Boca Raton, FL.
9. Cox, D. R. (1958), *Planning of Experiments*, Wiley, New York, pp. 53–58.
10. Ganju, J. (2004), Some unexamined aspects of analysis of covariance in pretest–posttest studies, *Biometrics*, 60, 829–833.
11. Crager, M. R. (1987), Analysis of covariance in parallel-group clinical trials with pretreatment baselines, *Biometrics*, 43, 895–901.
12. Casagrande, J. T., Pike, M. C., and Smith, P. G. (1978), An improved approximate formula for calculating sample sizes for comparing two binomial distributions, *Biometrics*, 34, 483–496.
13. Statistical Solutions (2005), nQuery Advisor, version 5.0, MA.
14. Fleiss, J. L., Tytun, A., and Ury, H. K. (1980), A simple approximation for calculating sample sizes for comparing independent proportions, *Biometrics*, 36, 343–346.
15. Barnard, G. A. (1947), A new test for $2 \times 2$ tables, *Nature*, 156, 177.

16. Mehta, C. R., and Hilton, J. F. (1993), Exact power of conditional and unconditional tests: Going beyond the $2 \times 2$ contingency table, *Am. Statist.*, 47, 91–98.
17. Gordon, I. (1994), Sample size for two independent proportions: A review, *Australian J. Stat.*, 36, 199–209.
18. Cuzick, J. (1982), The efficiency of the proportions test and the log-rank test for censored survival data, *Biometrics*, 38, 1033–1039.
19. Mantel, N. (1966), Evaluation of survival data and two new rank order statistics arising in its consideration, *Cancer Chemother. Rept.*, 50, 163–170.
20. Collett, D. (1994), *Modelling Survival Data in Medical Research*, Chapman and Hall, London, pp. 255–264.
21. George, S., and Desu, M. (1974), Planning the size and duration of a clinical trial studying time to some critical event, *J. Chronic Dis.*, 27, 15–24.
22. Schoenfeld, D. A. (1983), Sample size formula for the proportional-hazards regression model, *Biometrics*, 39, 499–503.
23. Lachin, J. M., and Foulkes, M. A. (1986), Evaluation of sample size and power for analyses of survival with allowance for non-uniform patient entry, losses to follow-up, noncompliance, and stratification, *Biometrics*, 42, 507–519.
24. Temple, R. J. (2004). *The Critical Path Opportunities for Efficiency in Development.* FDA Science Board Advisory Committee Meeting Maryland, April 22.
25. Fleiss, J. L. (1986), *The Design and Analysis of Clinical Experiments*, Wiley, New York, p. 152.
26. Ganju, J. (2008), Post stratified analysis of clinical trial data, under preparation.
27. Ganju, J., and Mehrotra, D. V. (2003), Stratified experiments re-examined with emphasis on multicenter trials, *Controlled Clin. Trials*, 24, 167–181. Correction: 24, 830.
28. Deming, W. E. (1960), *Sample Design in Business Research*, McGraw-Hill, New York.
29. Oxman, M. N., Levin, M. J., Johnson, G. R., et al. (2005), A vaccine to prevent herpes zoster and postherpetic neuralgia in older adults, *N. Engl. J. Med.*, 352, 2271–2284.
30. Lakatos, E. (1986), Sample size determination in clinical trials with time-dependent rates of losses and noncompliance, *Controlled Clin. Trials*, 7, 189–199.
31. Lakatos, E. (1988), Samples sizes based on the log-rank statistic in complex clinical trials, *Biometrics*, 44, 229–241.
32. Xing, B., and Ganju, J. (2005), A method to estimate the variance of an endpoint from an on-going blinded trial, *Stat. Med.*, 24, 1808–1814.
33. Ganju, J., and Xing, B. (2009), Re-estimating the sample size of an on-going blinded trial based on the method of randomization block sums, *Stat. Med.*, 28, 24–38.
34. Gould, A. L., and Shih, W. J. (1992), Sample size re-estimation without unblinding for normally distributed outcomes with unknown variance, *Commun. Stat. Theory Methods*, 21, 2833–2853.
35. Zucker, D. M., Wittes, J. T., Schabenberger, O., and Brittain, E. (1999), Internal pilot studies II: Comparison of various procedures, *Stat. Med.*, 18, 3493–3509.
36. Ganju, J., Izu, A., and Anemona, A. (2008), Sample size for equivalence trials: A case study from a vaccine lot consistency trial, *Stat. Med.*, 27, 3743–3754.
37. Farrington, C. P., and Manning, G. (1990), Test statistics and sample size formulae for comparative binomial trials with null hypotheses of non-zero risk difference or non-unity relative risk, *Stat. Med.*, 9, 1447–1454.
38. Nam, J. (1995), Sample size determination in stratified trials to establish the equivalence of two treatments, *Stat. Med.*, 14, 2037–2049.
39. Lavori, P. W. (1992), Clinical trials in psychiatry: Should protocol deviation censor patient data, *Neuropsychopharmacology*, 6, 39–48.

# *Blinding and Placebo*

**Artur Bauhofer**
*Institute of Theoretical Surgery, Philipps-University Marburg, Marburg, Germany*

**Contents**

1 Introduction
2 Need for Placebo
3 Features of Placebo
4 Coding and Randomization
5 Blinding More Than Patients and Physicians
6 Blinding Trials Other Than Oral Drug Comparisons
7 Waiving Blindness
8 Safety Mechanisms: Breaking Code in Case of Emergency
9 Assessment of Blinding and Expectation
10 Indices for Assessment of Blinding
11 Breaking Code at End of Trial
12 Conclusion
Appendix: Critical Questions for Blinded, Placebo-Controlled Randomized Trials
References

## 1 INTRODUCTION

Treatment recommendations in guidelines are mainly based on the results of randomized controlled trials (RCTs) [1]. Results of RCTs should provide the closest possible approximation to the truth. Bias-reducing safeguards (e.g., concealments of randomization and blinding) are important because their omission can exaggerate a treatment effect by 20–45% relative to the true treatment effect [2]. Overestimation in this order may be important since most RCTs seek to detect treatment

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

**TABLE 1 Quality of Different Allocation Methods**[a]

| Allocation by | Constant Probability | Equal Probability | Independent Probability |
|---|---|---|---|
| Randomization | Yes | Yes | Yes |
| Alternation | Yes | Yes | No |
| Birthday | (Yes) | No | Yes |
| Initials | (Yes) | No | Yes |

[a]Different allocation methods with limitations compared to randomization.

*Source*: Adapted from Lorenz et al. [10].

effects of moderate size of about 20–35% [3]. For this reason, moderate degrees of bias can lead to important distortions in treatment effect estimates; and authorities [4] encourage the use of RCT methodology and researchers conducting systematic reviews of RCTs. In systematic reviews heterogeneity in trial results are often explained by the possibility of differences in trial methodology. Sometimes the effect size is weight based on the methodology used in the RCT [5]. Such corrections are problematic since a trial that is not double blinded must per se not be of inferior quality than a trial in which only data collectors and outcome assessors were blinded or even where an open trial methodology was used.

Often only wordings without common sense such as double blind were used without exactly reporting who was blinded and who was not. Despite efforts to improve the reporting of RCTs, such as the Consolidated Standards of Reporting Trials (CONSORT), studies have documented in a low number concealment of randomization and blinding [2, 6].

Most researchers appreciate the meaning of blinding, but beyond general appreciation there is confusion. Terms such as single blind, double blind, and triple blind mean different things to different people (Fig. 1) [7]. Moreover, many medical researchers confuse the term blinding with allocation concealment. Allocation concealment is primarily used to prevent selection bias and to protect an assignment sequence before and until allocation. In contrast, blinding refers to keeping persons (participants, investigators, and data assessors) unaware of the allocated treatment or therapy, so that they are not influenced psychologically or physically by that knowledge. In well-blinded trials one can have more certainty that any differential effect between groups is due to the treatment rather than the subjects' or researchers' bias [8].

About 44–93% of publications of randomized controlled trials lack a clear description of allocation concealment [9]. This is really a poor rate since a clear description without quantitative information needs no additional trial effort and inadequate information exaggerates the treatment effect [6].

There is agreement that the allocation should be performed by chance, but randomization is not always used as it should be. There are still other allocation procedures in use (Table 1), such as allocation by alternation, by birthday, and by initials. These methods do not fulfill all three criteria: constant, equal, and independent probability [10].

## 2 NEED FOR PLACEBO

In contrast to several other study types, randomized controlled trials always have a control group. Having control patients who are completely without treatment does

(A)

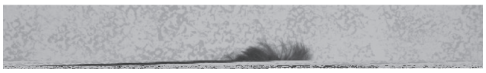

(B)

(C)

***FIGURE 1*** *An exact description of who is blinded when is preferable to stating the trial was "single" (A), "double" (B), or (C) "triple" blinded. In the pictures the author is (A) "single", (B) "double", and (C) "triple" blinded.*

not allow to decipher whether any response improvement in the treatment group is due to therapy or due to the fact of being treated in some way. Even if therapy is irrelevant to the patients condition, the patient's attitude to his or her illness and indeed the illness itself may be improved by a feeling that something is being done to improve his or her condition [11].

Gribbin [12] argues that many patients could be effectively treated by placebo, especially by the use of attractive pills and a convincing statement by the physician as to their value. Hence, in any randomized drug trial versus untreated controls, it is worth considering giving the latter a placebo. The use of a placebo allows to eliminate the "placebo effect" from the therapeutic comparison. For this reason placebos are commonly used now in many disease trials.

The decision whether some standard active drug therapy should be a control and the new drug is given as an add-on is an entirely separate issue. One basic principle is that patients cannot ethically be assigned to only a placebo if there exists an alternative standard therapy of established efficacy [11]. The problem is that in many areas of medicine the standard therapy was never tested by sound evidence-based trials. Without supportive evidence the physician has to relay on clinical experience and opinion in deciding whether it is ethical to withhold what has become accepted as standard therapy.

Sometimes two active components should be compared. In this scenario often a "double-dummy" method is used. For example, if we want to compare two medicines, one presented as a blue tablet and one as a red capsule, we could also supply blue placebo tablets and red placebo capsules. Both patient groups have to take one blue tablet and one red capsule [13].

Placebos are used to make patients' and observers' attitudes to the trial as similar as possible in treatment and control groups. Under many circumstances the use of placebos is a prerequisite to perform a double-blinded trial.

## 3 FEATURES OF PLACEBO

Placebos have to be identical in all respects to the active drug, except that the active drug is absent. This means for oral placebos that they are not distinguishable by color, size, shape, taste, or texture. For oral therapy there are several modes of application available: capsules, tablets, or liquids. Since many active drugs have a distinctive taste, the use of capsules is most often feasible.

When using liquids for oral therapy or for injections, in addition to the color the viscosity and the potential to generate foam by shaking should also be considered. For example, liquids containing protein generate foam but buffer solutions do not. In this case, the ideal placebo will contain also an inactive protein in low concentrations such as human albumin. Some clinicians are very clever in breaking the code; they not only shake or identify the drugs by taste and smell but even photometers were used to determine differences between active drug and placebo. For such reasons it is sometimes very difficult to guarantee blinding.

## 4 CODING AND RANDOMIZATION

Double-blinded randomized trials require most careful organization in the allocation of treatments. The allocation to different treatments should have similar probabilities (see above and Table 1).

Before the generation of a randomization list is performed, principal considerations have to be made. Should the randomization list be generated by simple

random permutation or is there a need for stratification in the trial. In some trials, for example, stratification for different operations and age are useful. To allocate approximately the same number of patients to the placebo and the treatment group in each center, a block randomization is often used. This means, for example, in a trial with 100 patients it could have a block size of 20 patients randomizing 10 placebo patients and 10 treatment patients to one center. If one block is assigned, the next block will be used by the individual study center.

Another topic to be discussed is the use of an unequal randomization to placebo and treatment group. Driven by the enthusiasm for a new treatment, sometimes more patients were randomized to the treatment group, even though it would involve some loss of statistical efficiency [14]. Thus randomization in a 2:1 or 3:1 ratio for the new:standard treatment is a realistic consideration; however, there is an even greater loss in statistical power in the 3:1 approach.

There are several methods for the generation of randomization lists available. A detailed description of which method was used is given in Table 2. In general, the preparation of random lists and drug packages should be performed by persons not otherwise involved in conducting the trial. Most times random lists are generated by the study statistician. In all cases it is important to have a simple coding system linking the drug packages to the randomization list. Each package must have a unique trial code number that is also written on the randomization list and patient record form to safeguard a traceable link between, patients, packs, and the list.

Today, more and more central computers are used for randomization and for distribution and allocation of medication packages to the different study centers over a interface with the telephone (interactive voice response, IVR) or the Internet (interactive Web response, IWR). These methods led to significant savings in the amount of medication needed, but also to an increased risk of breaking the code [15]. A deduction from the dispensing order is in some cases possible. An example of such a situation is given by McEntegart et al. [15]. Four packs—two of each group—were delivered to the site in a trial with two treatments (A and B) stratified by gender. The first two patients randomized into the study were male (numbers

**TABLE 2 Generation of Randomization Lists and Allocation Concealment[a]**

| Method | Adequate Descriptions | Inadequate, Incomplete Descriptions |
|---|---|---|
| Computer | Random lists generated by a computer were used by the physician to sequentially allocate the patients to a treatment. | Just mentioning the use of a computer is not enough information. Further details should be given since computers are rarely involved in the allocation process. |
| Envelopes | For allocation serially numbered, sealed and opaque envelopes were used. | Allocation was done by the use of opaque envelopes |
| Vehicles | Vehicles were indistinguishable, sequentially numbered, and sequentially administered without knowledge of the content. | For allocation indistinguishable vehicles were used. |
| Pharmacy unit | Drug preparations from the pharmacy unit, using indistinguishable vials and content were used. | Allocation was performed by the pharmacist. |

[a]Criteria for adequate description of allocation concealment were taken in part from Schulz [34].

001 and 002). They were assigned to treatments A and B. The next patient, a female, was assigned to treatment A, as defined by the random list and received a matching pack, which was in this example number 004. From this numbering can be concluded that pack 003 must contain different medication from pack 004 because 003 was not given to the patient. Since pack 003 and 004 are different from each other also 001 and 002 must be different from each other.

Another problem is called pack separation, which occurs when the medication is supplied or used in a different ratio to the one used for packaging. More packs from one type will be left, for example, due to a higher withdrawal rate on active treatment. Repeated resupply of packs with equal numbers of packs will allow a division of packs into two distinct groups (one group with more and one with fewer packs), which allows assumptions on their difference. This clearly demonstrates also that sophisticated blinding and drug supply procedures with IVR and IWR technology are susceptible to bias unblinding.

To overcome these problems two strategies are possible. In the first approach the number of delivered packs in each block is increased, but then a part of the drug savings obtained by the individual delivery of packs is lost. Another approach is a double randomization [15]. In this case packs are randomly shuffled around so that there is no longer any association between the order in the file and the pack numbers. This method probably increases the complexity of labeling and distribution, and care has to be taken that no confusion occurs.

## 5 BLINDING MORE THAN PATIENTS AND PHYSICIANS

For the term *blinding* sometimes the synonym *masking* is used but blinding is used more often. Blinding, however, is used to reduce bias, but it is not always easy to obtain. Before an *open (nonblinded) trial* is performed, the question should be asked as to who can be blinded of the persons involved in the trial. A list of the individual groups is defined in Table 3 in accordance with Montori et al. [16].

*Single-blind trials* (where either only the investigator or the patient is blind to the allocation) are most times preferable to *open trials*. In *double-blinded trials* it

**TABLE 3 Individual Groups That Could Be Blinded**

| Group | Definition[a] |
|---|---|
| Participants | Individuals who are randomly assigned to the interventions under evaluation. |
| Health care providers | The physicians, nurses, or other personnel that actually care for the participants during the study period and/or administer the interventions. |
| Data collectors | The individuals who actually collect data for the study outcomes. Data collection could include a quality-of-life questionnaire, talking and/or recording a blood pressure measurement. |
| Outcome assessors | The individuals who ultimately decide if a patient has suffered the outcome of interest. |
| Data analysts | The individual who conducts the data analysis. |
| Manuscript writers | The individuals who write alternative versions of the manuscript before breaking the randomization code. |

[a]Definitions were taken from Montori et al. [16].

is implicated that the assessment of patient outcome is done without knowing the treatment received [13]. Such blind assessment of outcome can often also be achieved in trials that are open. For example, pathological findings can be assessed by someone else who was not involved in running the trial. Blinded assessment of patient outcome is also valuable in epidemiological studies, such as cohort studies. In diagnostic tests and other trials, persons evaluating the performance of those performing the test should be unaware of the true diagnosis. Or in studies for the evaluation of the reproducibility of a measurement technique, the observers should be unaware of previous measurements on the same individuals [13].

One particular advantage of double-blind trials is that they allow the objective evaluation of side effects, both by the patient and by the physician. For instance, side effects (usually minor effects such as headache, fatigue, nausea) were reported by patients on placebo. This enables to correct for the overreporting of side effects on active therapy to get an unbiased estimate of adverse reactions attributed to the treatment itself [11].

Beyond the widely used term *double blinding* sometimes the term *triple blinding* is also used. This term usually means a double-blind trial that also maintains a blinded data analysis [11], but this is not always the case. Some people think it denotes that investigators and assessors as well as the participants are all unaware of the assignment.

Since there is some confusion on the use of the terminology *single blind, double blind*, and *triple blind* in the report of trials, clear statements should be given on who was blinded at what part of the study (Fig. 1). In general, the quality of the reports of concealment and blinding is poor (Table 4), but there seems to be an improvement in the reports from 2002 to 2004, although the studies from Montori et al. [16] and from Devereaux et al. [2] did not use identical criteria for trial selection. Montori et al. [16] analyzed RCT reports from five leading journals, and Devereaux et al. [2] used reports from internal medicine patients. Poor reporting does not mean automatically poor conduction of the trials. Some authors stated that they performed concealing randomization and blinding in adequate ways [2]. From this observation it can be concluded that readers should not assume at least in general that bias-reducing procedures not reported in RCTs did not occur.

However, on average, randomized trials that have not used appropriate levels of blinding show larger treatment effects than blinded studies [17]. In parallel,

**TABLE 4 Reporting Allocation Concealment and Blinding in Randomized Trials**[a]

| | Montori et al. [16] 2002 | Devereaux et al. [2] 2004 |
|---|---|---|
| Number of trials | 191 | 105 |
| Report of concealment | n.a. | 45% |
| Blinding status of | | |
| Participants | 15% | 74% |
| Health care providers | 5% | 36% |
| Data collectors | 12% | 16% |
| Outcome assessors | 23% | 17% |
| Data analysts | 5% | 4% |
| Manuscript writers | 0% | n.a. |

[a]Information given on allocation concealment and blinding in randomized controlled trials. n.a. = not assessed in the study.

diagnostic test performance is often overestimated when the reference test is interpreted with knowledge of the test results [18]. Quantification of the blinding bias by adjusting the results for a single trial or a meta-analysis of several trials for trial quality is problematic. However, without quantification blinding makes it difficult to bias results intentionally or unintentionally and so helps to ensure the credibility of study results [13].

Beyond the blinding of participants, health care providers, data collectors, outcome assessors, and data analysts (Table 3) some trialists postulate even the blinding of writers [19]. In this approach two or more manuscripts were prepared before the code of the trial is broken. In one manuscript the test drug is significantly better than placebo and in the other there is no benefit of the new treatment. In my opinion this approach is too much directed by statistical rigidity, aside of clinical reality. There should be a clear hypothesis published in advance for the primary outcome of the trial, but the manuscripts will differ gradually. For example, the new treatment can be statistically better than the standard, as assumed beforehand, or even better at a lower significance level, or unchanged or even significantly adverse. However, hardly anyone would write five different manuscripts and interpretations of the results in advance.

## 6 BLINDING TRIALS OTHER THAN ORAL DRUG COMPARISONS

Blinding in oral drug trials is most times the easiest part, but it can be also more complex, for example, in the comparison of repetitive applications with a single application. In this case multiple placebo applications are needed to safeguard blinding.

Placebo injections present a greater practical and ethical problem than oral placebos. Additionally, subcutaneous and peripheral injections, for instance, are more easily accepted than more invasive procedures such as insertions of central venous lines.

The problem of blinding is most prominent in surgical trials. Blinding of surgeons is in most cases impossible, and a placebo operation (sham operation) will be in most cases unethical, but it is under some circumstances possible. In a trial by Freed et al. [20] patients received a sham surgery by drilling holes in the skull without penetration of the dura mater. This trial was accepted by the local ethics committee and published in the *New England Journal of Medicine*. A sham treatment such as this one is probably not possible in a lot of other centers. This example demonstrates that ethical considerations may be influenced by local and cultural differences.

Even if blinding of the surgeon is impossible, very elegant designs were used to reduce bias in surgical trials. For instance, in several trials comparing laparoscopic cholecystectomy with conventional surgery, an improved outcome for the new technique was reported [21]. Expectation bias was excluded for the first time in a randomized controlled trial from Majeed et al. [22] comparing laparoscopic and small-incision cholecystectomy. In this trial identical wound dressings in both groups were used so that nurses and trial personnel were blinded to the type of operation. Investigators found no difference between the groups with regard to hospital stay, time back to work, and time to full activity. In this case blinding has destroyed the illusion of a significant improvement by the use of laparoscopy.

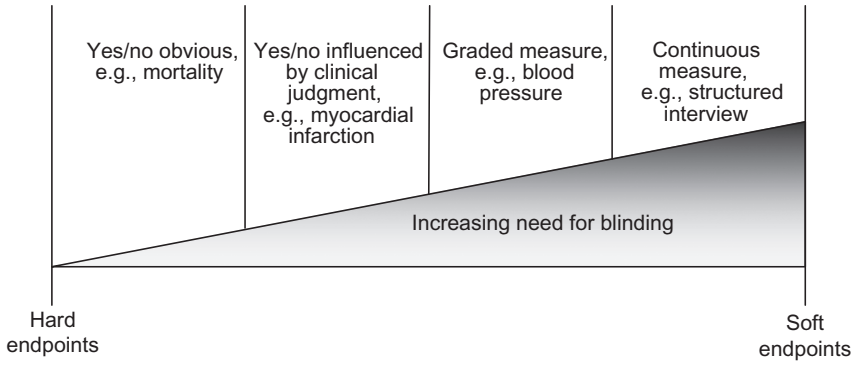


***FIGURE 2*** *Need for blinding: In parallel with the decrease of objectivity from "hard" endpoints to "soft" endpoints, the influence of expectation increases. Blinding is more important for the assessment of outcomes with high expectation influence.*

## 7 WAIVING BLINDNESS

Ethical considerations often rule out a double-blind trial design. As already mentioned, in most cases of surgical trials it would be unethical to subject a control group to incisions under anesthesia, mimicking genuine surgery. But, except for the surgeon, all others involved in the trial can be blinded (Table 3).

For some treatments it is totally impossible to arrange a double-blind design. For instance, the evaluation of cytotoxic drugs in cancer therapy is often not double blinded because of complicated dose schedules, the likelihood of serious side effects, and dose modifications to suit each patient's needs. All these points make it necessary for the treating physician to know the patient's therapy [11].

It should be considered how serious the bias might be without blinding. In general, for more subjective endpoints a double-blind design is more important than for more objective ones such as mortality (Fig. 2). In each trial organizers have to weigh the pros and cons when planning the trial and writing the trial protocol. Both, ethical and practical problems with blinding have to be considered.

## 8 SAFETY MECHANISMS: BREAKING CODE IN CASE OF EMERGENCY

In the case of an emergency a rapid decoding must be possible. For this reason the sponsor and investigators at the different centers must be immediately able to break the code of each individual patient. For this event in most randomized, blinded trials envelopes are prepared for decoding. With each drug package an opaque, sealed, unique coded envelope is delivered to the responsible physician that allows decoding. An additional set of envelopes are left with the sponsor. Another method is the preparation of lists that allow one to scratch off the number code to identify the drug assigned to the participant. In no case should the entire randomization list be transferred to the investigator, which would allow decoding all patients in the trial. Decoding should be allowed only when necessary for the care of the individual patient. Each time the code is broken, the reasons why, the date, and person who broke the code should be recorded. The condition of envelopes will be checked at

each monitoring visit. At the end of the trial the envelopes will be collected, inspected, and mentioned in the final study report [23].

## 9 ASSESSMENT OF BLINDING AND EXPECTATION

Ideally, investigators state in the report if blinding was successful. The success of blinding can be easily determined by asking participants, health care providers, data collectors, and outcome assessors to guess which intervention was provided. Table 5 demonstrates an example from a randomized, placebo-controlled trial. Patients were nicely blinded, but health care providers and data collectors were not. In this trial [24, 25] the study drug was G-CSF (granulocyte colony-stimulating factor), which increases the number of granulocytes. Unblinding was probably performed by the knowledge of the leukocyte count. Furthermore, in our trial there was no change over time in the different groups of persons asked. Others have found a change over time, so a repeated measurement of blinding is favored by them [26].

In general, the number of trials providing evidence on the success of blinding is poor. In a analysis by Fergusson et al. [27] it was shown that only 8% (15 of 191 trials) provided qualitative or quantitative information about blinding. Only in 5 of these 15 trials was blinding successful, and only two presented qualitative data. Due to the poor information given, the reporting of randomized controlled trials in regard to this topic should be improved.

In surgery, independent of the treatment, whether the new test drug or placebo is applied, patients believed that the disease will be cured and symptoms reduced by the operation (Table 6). Surgeons had a much more realistic expectation knowing the current literature and their own results. Positive expectations by the patient have a strong influence on the outcome as demonstrated by Koller et al. [28]. Furthermore, a comparison of patients included in the randomized controlled trial and patients treated under routine conditions in the same institution demonstrated an improved outcome of the patients under trial conditions [29]. This observation is not new but often neglected at the time of interpretation of trial efficacy data and extrapolation in regard to clinical efficiency in routine.

In many trials participants showed a strong tendency to believe they had been assigned to the active intervention (Table 6). Expectation can be analyzed also with questionnaires. In our G-CSF trial [24, 25] 68/75 patients thought immediately before the operation that they got the active substance and not the placebo. This belief may be influenced by the time of asking (beginning or end of the trial) and

**TABLE 5 Guess of Treatment (Blindness) of Patients, Surgeons, Ward Assistants, and Data Collectors at Day 3 after Operation (and at Discharge)[a]**

| Group | Correct Guess (%) | Phi Coefficient | *P* value |
|---|---|---|---|
| Patients | 41 (50) | –0.16 (0.10) | 0.155 (0.37) |
| Surgeons | 73 (73) | 0.46 (0.49) | <0.001 (<0.001) |
| Ward assistants | 67 (76) | 0.33 (0.54) | <0.005 (<0.001) |
| Data collectors | 73 (76) | 0.46 (0.54) | <0.001 (<0.001) |

[a]Blinding was evaluated in a randomized controlled trial with filgrastim prophylaxis versus placebo in patients with increased risk (ASA class 3 and 4) for an adverse outcome after colorectal cancer surgery [24, 25]. Most patients guessed (68/75) that they received the active study drug and not the placebo. *P* values were determined with the Chi-square test.

**TABLE 6 Expectation of Patients and Surgeons before Operation**[a]

| | Patient | | | Surgeon | | |
|---|---|---|---|---|---|---|
| Do You Expect: | G-CSF | Placebo | *P* Value | G-CSF | Placebo | *P* Value |
| Stop tumor growth | 36/36 | 39/40 | 1 | 33/36 | 35/40 | 0.71 |
| Reduction of tumor size | 36/36 | 35/40 | 0.06 | 31/36 | 37/40 | 0.47 |
| Healing | 36/36 | 39/40 | 1 | 32/36 | 30/40 | 0.15 |
| Prevention of metastasis | 35/36 | 39/40 | 1 | 32/36 | 30/40 | 0.15 |
| Prevention of tumor relapse | 36/36 | 40/40 | 1 | 31/36 | 34/40 | 0.58 |
| Prevention of pain increase | 36/36 | 39/40 | 1 | 34/36 | 33/40 | 0.16 |
| Pain relief | 34/36 | 38/40 | 1 | 20/36 | 25/40 | 0.49 |
| Free of pain without medication | 34/36 | 38/40 | 1 | 31/36 | 29/40 | 0.17 |
| Relief of tumor-related symptoms | 34/35 | 37/40 | 0.61 | 31/36 | 35/40 | 1 |
| Psychological stabilization | 35/36 | 39/40 | 1 | 32/36 | 38/40 | 0.41 |

[a]Expectation of patients and surgeons was evaluated in a randomized controlled trial with filgrastim prophylaxis versus placebo in patients with increased risk (ASA class 3 and 4) for an adverse outcome after colorectal cancer surgery. *P* values were determined with the Chi-square test.

by repeated asking itself to guess what treatment they got. Differentiated outcome analysis of blinded and unblinded patients revealed that correct guess of treatment by the study personnel had no influence on patients' outcomes in all major endpoints like: quality of life, the McPeek recovery index, length of stay, mortality, complication rate, and re-operation rate.

Expectation bias can be further reduced by using different persons for study coordination and psychometric rating [30]. Since during the trial patients become increasingly familiar and comfortable with the staff and trial office, this often triggers a therapeutic response to the patients [31].

To reduce expectation effects in randomized trials with psychotropic drugs, often a placebo run-in phase is included, which is a single-blind placebo period that lasts about 7–14 days. The placebo run-in phase occurs before randomization, and all study-eligible subjects are given the placebo treatment and are withdrawn most times from any psychoactive drugs during this interval. Responders to the placebo in this preliminary phase were excluded from the trial to reduce expectation and to increase the size of the treatment effect. A recent meta-analysis showed that trials with a placebo run-in phase and elimination of the placebo responders before randomization tended to result in a larger drug effect size, but without statistical significance [32].

Another approach to reduce expectation is to orientate patient expectations with educational statements. This may include statements about the possibility of the study drug working as intended; they may get a placebo and consequently may not improve, and they may receive admonitions to report the actual symptoms honestly without attempting to "help" in accentuating the positive [30]. For instance, site personnel would read the educational statement to the patient during the consent and review it at each succeeding visit. Information about these complex interactions between patients and staff is also important and will further reduce expectation.

## 10 INDICES FOR ASSESSMENT OF BLINDING

The success of blinding can be assessed not only by questionnaires but also by validated indices that are available [8, 33]. In the simple and easy-to-calculate index of

James et al. [8] the value increases from 0 to 1 with the success of blinding, while 0.0 is the complete lack of blinding, and 1.0 perfect blinding. In contrast to this index the more complex index of Bang et al. [33] includes also the important opinion of uncertainty. The index is scaled to an interval of –1 to 1; 1 being completely lack of blinding, 0 being consistent with perfect blinding, and –1 indicating opposite guessing, which may be related to unblinding.

It is the decision of the persons responsible for the particular trial to use a score or questionnaire. They should be aware that these tools can help to assess the rate of unblinding and the potential influence on the outcome.

## 11 BREAKING CODE AT END OF TRIAL

Before the code is broken, all parameters and endpoints listed in the trial protocol should be determined and saved in a log file. This procedure should guarantee as much as possible blindness of data collectors, outcome assessors, and data analysis. At that time point when all steps of data analysis were performed without group comparison, it is time to break the code. Unblinding is a milestone in randomized controlled trials and is for this reason often performed as a ceremonial act.

When the whole trial is completed and the results are interpreted, then it is desirable to inform investigators of each patient treatment. This allows one to get feedback from the experience of the investigators and understanding of the treatments. This procedure allows further the improvement of future trials. Some patients also want to know to which treatment group they were assigned. At this time the information should be accessible for them.

## 12 CONCLUSION

There is no uniform methodology for all randomized controlled blinded trials. The need for blinding is increasing in parallel with the decrease of objectivity from "hard" endpoints to "soft" endpoints as the influence of expectation increases. Blinding is more important for the assessment of outcomes with high expectation influence. Ethical and practical problems with blinding have to be considered early in trial development. Reducing bias by blinding is also possible in trials where health care providers (e.g., surgeons) cannot be blinded; but data collectors, outcome assessors, and data analysts can effectively be blinded. Some critical questions to be asked are listed in the Appendix. The success of blinding can be determined at various time points by asking participants, health care providers, data collectors, and outcome assessors to guess which intervention was provided. Blinding can be also assessed by the use of a scoring system. The decision of which system of quantification is used is not so important, but the trial report should include a clear statement of how randomization and allocation were performed and who was at which time point blinded. Following the general consideration mentioned here, bias can be reduced and the plausibility of trial results can be increased.

## APPENDIX CRITICAL QUESTIONS FOR BLINDED, PLACEBO-CONTROLLED RANDOMIZED TRIALS

| Item | Questions to Be Answered |
|---|---|
| Use of placebo | What is the current standard for the treatment of the specific disease under investigation? |
| | Is the use of a placebo ethically sound? |
| Preparing the placebo | Which pharmaceutical form is the best to hide the identity of the drug in regard to smell, taste, color, and structure? |
| | Is the placebo identical to the active medication? |
| Coding and randomization | Which method is used for coding and randomization? |
| Blinding | Does the study allow to blind: participants, health care providers, data collectors, outcome assessors, data analysts? |
| | Dose the blinding result in any harm or undo risk to the patient? |
| Breaking the code | Which safety mechanisms are provided to break the code in an emergency? |
| Testing blinding | Should the study include a test of blinding of the different individuals involved in the trial? |
| Reporting | Is the generation of the randomization code, allocation process, and blinding adequately described? |

## REFERENCES


1. Ellis, J., Mulligan, I., Rowe, J., et al. (1995), Inpatient general medicine is evidence based. A-Team, Nuffield Department of Clinical Medicine, *Lancet*, 346, 407–410.
2. Devereaux, P. J., Choi, P. T., El-Dika, S., et al. (2004), An observational study found that authors of randomized controlled trials frequently use concealment of randomization and blinding, despite the failure to report these methods, *J. Clin. Epidemiol.*, 57, 1232–1236.
3. Yusuf, S., Collins, R., and Peto, R. (1984), Why do we need some large, simple randomized trials? *Stat. Med.*, 3, 409–422.
4. The European Agency for the Evaluation of Medical Products (2001), ICH Topic E10—Choice of Control Group in Clinical Trials: Note for Guidance on Choice of Control Group in Clinical Trials. CPMP/ICH364/96, 1–29.
5. Moher, D., Jadad, A. R., Nichol, G., et al. (1995), Assessing the quality of randomized controlled trials: An annotated bibliography of scales and checklists, *Controlled Clin. Trials*, 16, 62–73.
6. Pildal, J., Chan, A. W., Hrobjartsson, A., et al. (2005), Comparison of descriptions of allocation concealment in trial protocols and the published reports: Cohort study, *BMJ*, 330, 1049.
7. Devereaux, P. J., Manns, B. J., Ghali, W. A., et al. (2001), Physician interpretations and textbook definitions of blinding terminology in randomized controlled trials, *JAMA*, 285, 2000–2003.
8. James, K. E., Lee, K. K., Kraemer, H. C., et al. (1990), An index for assessing blindness in a multicenter clinical trial: Disulfiram for alcohol cessation—a VA cooperative study, *Stat. Med.*, 15, 1421–1434.
9. Moher, D., Schulz, K. F., and Altman, D. G. (2001), The CONSORT statement: Revised recommendations for improving the quality of reports of parallel-group randomised trials, *Lancet*, 357, 1191–1194.

10. Lorenz, W., Ohmann, C., Immich, H., et al. (1982), Patientenzuteilung bei kontrollierten klinischen Studien, *Chirurg*, 53, 514–519.
11. Pocock, S. J. (1983), Blinding and placebo, in Pocock, S. J., Ed., *Clinical Trials—A Practical Approach*, Wiley, Chichester, pp. 90–99.
12. Gribbin, M. (1981), Placebos: Cheapest medicine in the world, *New Sci.*, 89, 64–65.
13. Day, S. J., and Altman, D. G. (2000) Statistics notes: Blinding in clinical trials and other studies, *BMJ*, 321, 504.
14. Pocock, S. J. (1979), Allocation of patients to treatment in clinical trials, *Biometrics*, 35, 183–197.
15. McEntegart, D., Lang, M., and Wood, R. (2005), Protecting the blind, *Good Clin. Pract. J.*, 11, 10–13.
16. Montori, V. M., Bhandari, M., Devereaux, P. J., et al. (2002), In the dark: The reporting of blinding status in randomized controlled trials, *J. Clin. Epidemiol.*, 55, 787–790.
17. Schulz, K. F., Chalmers, I., Hayes, R. J., et al. (1995), Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials, *JAMA*, 273, 408–412.
18. Lijmer, J. G., Mol, B. W., Heisterkamp, S., et al. (1999), Empirical evidence of design-related bias in studies of diagnostic tests, *JAMA*, 282, 1061–1066.
19. Gotzsche, P. C. (1996), Blinding during data analysis and writing of manuscripts, *Control Clin. Trials*, 17, 285–290.
20. Freed, C. R., Greene, P. E., Breeze, R. E., et al. (2001), Transplantation of embryonic dopamine neurons for severe Parkinson's disease, *N. Engl. J. Med.*, 344, 710–719.
21. McGinn, F. P., Miles, A. J., Uglow, M., et al. (1995), Randomized trial of laparoscopic cholecystectomy and mini-cholecystectomy, *Br. J. Surg.*, 82, 1374–1377.
22. Majeed, A. W., Troy, G., Nicholl, J. P., et al. (1996), Randomised, prospective, single-blind comparison of laparoscopic versus small-incision cholecystectomy, *Lancet*, 347, 989–994.
23. Spriet, A., and Dupin-Spriet, T. (1994), *Good Practice of Clinical Drug Trials*, Karger, Basel.
24. Bauhofer, A., Lorenz, W., Stinner, B., et al. (2001), Granulocyte-colony stimulating factor in the prevention of postoperative infectious complications and sub-optimum recovery from operation in patients with colorectal cancer and increased preoperative risk (ASA 3 and 4). Protocol of a controlled clinical trial developed by consensus of an international study group. Part two: Design of the study, *Inflamm. Res.*, 50, 187–205.
25. Bauhofer, A., Plaul, U., Torossian, A., et al. (2007), Perioperative prophylaxis with granulocyte colony-stimulating factor (G-CSF) in high-risk colorectal cancer patients for an improved recovery: A randomized controlled trial, *Surgery*, 141, 501–510.
26. Rees, J. R., Wade, T. J., Levy, D. A., et al. (2005), Changes in beliefs identify unblinding in randomized controlled trials: A method to meet CONSORT guidelines, *Contemp. Clin. Trials*, 26, 25–37.
27. Fergusson, D., Glass, K. C., Waring, D., et al. (2004), Turning a blind eye: The success of blinding reported in a random sample of randomised, placebo controlled trials, *BMJ*, 328, 432–437.
28. Koller, M., Lorenz, W., Wagner, K., et al. (2000), Expectations and quality of life of cancer patients undergoing radiotherapy, *J. R. Soc. Med.*, 93, 621–628.
29. Kopp, I., Bauhofer, A., and Koller, M. (2004), Understanding quality of life in patients with colorectal cancer: Comparison of data from a randomised controlled trial, a population based cohort study and the norm reference population, *Inflamm. Res.*, 53 (Suppl. 2), S130–S135.

30. Kirby, L., Borwege, S., Christensen, J., et al. (2005), Reducing placebo response: Triple blinding & setting expectations, *Apl. Clin. Trials*, November, 48–52.
31. Fritze, J., and Moller, H. J. (2001), Design of clinical trials of antidepressants: Should a placebo control arm be included? *CNS. Drugs*, 15, 755–764.
32. Lee, S., Walker, J. R., Jakul, L., et al. (2004), Does elimination of placebo responders in a placebo run-in increase the treatment effect in randomized clinical trials? A meta-analytic evaluation, *Depress. Anxiety*, 19, 10–19.
33. Bang, H., Ni, L., and Davis, C. E. (2004), Assessment of blinding in clinical trials, *Control Clin. Trials*, 25, 143–156.
34. Schulz, K. F., Altman, D. G., Moher, D., et al. (2002), Allocation concealment in clinical trials, *JAMA*, 288, 2406–2408.

# *Pharmacology*

**Thierry Buclin**
*Division of Clinical Pharmacology and Toxicology, University Hospital of Lausanne (CHUV), Lausanne, Switzerland*

**Contents**

1 Introduction
  1.1 Ethical Issues in Human Pharmacology Trials
  1.2 Roles and Importance of Clinical Pharmacology in Drug Development
  1.3 Classical Study Questions and Designs in Clinical Pharmacology
2 Pharmacokinetic Characterization of Drugs
  2.1 Drug Assay Methods
  2.2 Sample-Rich Trials
  2.3 Population PK Studies
3 Pharmacokinetic–Pharmacodynamic Evaluations
  3.1 Pharmacodynamic Variables
  3.2 Concentration–Response Relationships
  3.3 Design and Analysis of PKPD Trials
4 Conclusion
  Appendix: Note on Geometric Averages and Coefficients of Variation
  References

## 1 INTRODUCTION

A vast majority of clinical trials are designed to scientifically assess the efficacy, safety, effectiveness, or usefulness of a given medical procedure, most often a therapeutic intervention (drug treatment, surgery, physical therapy, etc.) and sometimes a diagnostic test or a prognostic marker. Such clinical trials challenge a small

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

segment of actual or upcoming medical practice. Their goal is to improve the welfare of the enrolled patients, which represents a determinant condition for the acceptance of new medical techniques. Proper conduct of clinical research is necessary to ensure the scientific validity of such claims for improvement. However, besides this important body of medically oriented trials, one can identify a minority of clinical experiments not strictly directed toward the improvement of the participants condition. Such trials are typically aimed at studying a given scientific question, which is to be tested during a clinical investigation, however, without therapeutic purpose. Depending on the type of question, one may distinguish between clinical studies in *physiology*, which address normal functions of the organism, *pathophysiology*, dealing with disease mechanisms, and *pharmacology*, oriented toward characterization of drug properties in humans. This chapter is devoted to the latter type of trials. Clinical pharmacology trials play an important role both in drug development and in optimization of patient care, and their specific methodological characteristics will be reviewed.

## 1.1 Ethical Issues in Human Pharmacology Trials

Even if scientific progress has indirect repercussions beneficial for patient care, clinical trials in pharmacology are most often without clear benefit to the subjects included. Those subjects may be either healthy volunteers or patients presenting with a given condition to investigate, such as renal or hepatic insufficiency susceptible to alter the disposition of a drug. Even in the latter case, the drug tested is generally given without medical indication (e.g., a single dose of an antibiotic may be given to hemodialysis patients not currently suffering from infection). This has definite ethical implications [1]:

- Obtaining a fully informed consent is absolutely necessary, and the inclusion of subjects not able to understand extensively the study conditions is unacceptable (while it may be considered feasible to some extent in therapeutically oriented trials).
- Not the slightest foreseeable risk is acceptable for study participants (while an expectedly favorable benefit–risk ratio is required in therapeutic trials, even when potentially hazardous treatments are administered).
- Study subjects will usually receive significant financial compensation for their participation (which is usually not the case in therapeutic trials).
- No pressure is to be exerted on the candidates to stimulate their participation in the study, making it problematic to include economically disadvantaged people, who may depend too strongly on financial compensation, or collaborators of the investigator, bound by professional dependency.
- Insofar as altruistic motivations play some role in the decision of subjects to participate in a trial, the investigators have a moral responsibility of exploiting the study results with the best consideration to public health.
- Eventually, particular attention must be paid to psychological and relational aspects of the study: Most people will not easily accept unnecessary medical investigations and drug treatments, even with compensation. Participating subjects may experience irrational concerns and anxiety or unpleasant feelings of

"hiring one's body." Thus, it is important that investigators in clinical pharmacology be capable of psychological attention, receptive attitude, recognition of valuable motivations, and thankfulness toward study subjects.

A detailed discussion of ethical issues raised by the inclusion of healthy volunteers in clinical trials will be found elsewhere in this textbook (see Chapters Ethics of Clinical Research in Drug Trials and Ethical Issues in Clinical Research). It is worth noting that the ethical standards applied to the conduct of human pharmacology trials have changed over the time and do not apply equally all over the world. The preferential involvement of prisoners as study subjects had been advocated for centuries; it culminated under the Nazi regime in Germany, and it was still practiced in the United States up to the 1970s [2]. It is now considered incompatible with human rights. The systematic inclusion of homeless people is still current practice in some places. Similarly, numerous academic centers would regularly ask their employees to participate in clinical investigations requiring healthy subjects. On the other hand, the importance of financial aspects in pharmacological research becomes of increasing public awareness. For-profit contract research organizations (CROs) play nowadays a leading role in the conduct of phase I trials, a vast majority of them being exclusively supported by industrial sponsors. In addition, the general trust of western populations in pharmaceutical companies is decreasing [3]. In this changing world, a strong commitment is required from investigators to apply rigorous ethical standards and protect human pharmacology research from any suspicion of implication in postmodern forms of slavery.

### 1.2 Roles and Importance of Clinical Pharmacology in Drug Development

Clinical pharmacologists working in pharmaceutical companies are traditionally in charge of phase I trials, mostly devoted to the characterization of the pharmacological profile of new compounds. Issues related to the organization and performance of phase I trials are discussed elsewhere in this textbook (see Chapter Phase I Clinical Trials). However, human pharmacology studies cannot be merely equated with phase I trials. Pharmacological questions are to be investigated at every step of drug development [4]. Pharmacokinetic analyses are increasingly included in phase II and III trials. Many drug interaction studies are conducted during phase IV. Moreover, it is often only after the commercialization of a drug that the suitability of therapeutic drug monitoring (TDM) for dosage individualization is examined, justifying further pharmacological assessments. In other instances, a biomarker useful for patient follow-up is identified and requires proper pharmacodynamic validation before large-scale evaluation and implementation.

The pharmacological characterization of new drugs has been considered by many authors as not entirely satisfactory at the time of product registration. Numerous drugs have been launched at suboptimal dosage, either too high or (less frequently) too low, because of a poor description of their dose–response relationship. A systematic analysis of 499 new prescription drugs approved by the Food and Drug Administration (FDA) over 20 years revealed that dosage changes had occurred in 73 of the drugs over the following years, the vast majority of which were decreases in dosage due to safety reasons [5]. Similar findings were reported in Europe [6].

Recommended dosing schedules happen to prove inadequate, either globally or in specific subgroups of patients. For example, while most intravenous β-lactam antibiotics should probably be proposed as 24-hour infusions [7], they are still prescribed as injections on a t.i.d. (three times a day) or q.i.d. (four times a day) basis. Significant drug interactions frequently remain unknown at launch time, which may cause serious adverse outcomes in patients and even the withdrawal of products from the market [8, 9]. This is also the case for pharmacogenetic influences, which should receive more consideration [10]. Lastly, treatment monitoring and dosage adaptation issues are paid insufficient attention from drug developers, preventing the patients from obaining the best possible benefit from new drugs until rational individualization strategies have been elaborated. Typical examples can be found in HIV (human immunodeficiency virus) treatment with protease and reverse transcriptase inhibitors [11], immunosuppression [12], and probably cancer treatment with new signal transduction inhibitors.

Clinical pharmacology has much to bring to drug development not only during the clinical phases I–IV but may even complement or replace some parts of the preclinical development. The sensitivity of drug assay methods is continuously increasing, with accelerator mass spectrometry techniques currently enabling determination of minute concentrations of radiolabeled compounds in biological fluids [13]. The administration of tiny amounts of new drug candidates to human volunteers, well below efficacy and toxicity dose thresholds, has been advocated for screening purposes, before the initiation of costly preclinical pharmacology and toxicology programs. This so-called phase-zero approach puts emphasis on the human pharmacokinetic profile as a relevant criterion for early drug development decisions [14]. It must be recalled that less than one in five drug candidates entering phase I eventually reach the market; this attrition plays a significant role in increasing the global costs of pharmaceutical research [15]. Clinical pharmacologists employed in the pharmaceutical industry are those who regularly bring the elements leading to a decision to kill a drug candidate during development. An earlier acquisition of such elements might therefore allow significant savings in research expenditures.

Finally, clinical pharmacology is still a science under development, and presently unknown concepts probably remain to be discovered. The whole story of cytochrome P450 enzymes has occupied many scientists since the 1970s, with numerous clinical trials devoted to elucidate their metabolic roles, genetic variations, and modulation by interacting drugs [16]. More recently, drug transporters went into scrutiny and justified a further amount of clinical investigations, which are still currently underway [17]. Other mechanisms of drug disposition are still poorly characterized and deserve definite research efforts (e.g., degradation pathways for peptide products). Advances in biopharmacy and new drug delivery techniques raise further issues that require to be tested at the clinical level. Pharmacodynamic mechanisms, besides their detailed dissection through in vitro experiments, also warrant to be better described through clinical experiments. Much remains to be discovered, too, in the field of adverse effect causation. Pharmacogenetics and ethnopharmacology represent further areas for intensive scientific exploration. Such "knowledge-oriented" research has been traditionally ascribed to academic groups, in opposition to "product-oriented" development. It is, however, important that both types of research conform to similar quality standards. Noticeably, a part of the scientific findings produced by universities lead to patent protection and industrial development. Conversely, phase I trials devoted

to a specific product development can identify new pharmacological concepts and produce generalizable findings. Thus, not only is clinical pharmacology important for drug development, but drug development itself invigorates clinical pharmacology in various scientific aspects.

## 1.3 Classical Study Questions and Designs in Clinical Pharmacology

Clinical pharmacology aims to describe, explain, and predict the fate and the effects of drugs in the human organism and to provide scientific foundations for the rational use of drugs in therapeutics. It relies on established knowledge in physiology and pathophysiology, and it uses specific methodological approaches to derive useful conclusions from clinical measurements, obtained during drug administration trials designed for that purpose. The general concepts governing clinical pharmacology can be found in dedicated textbooks [18, 19], and only their practical investigation through human trials will be covered here.

***Main Types of Study Questions*** Traditionally, the scientific questions addressed during clinical pharmacology studies have been classified into two main groups:

- *Pharmacokinetic* (PK) issues are those dealing with the fate of a drug in the human organism, that is, *absorption* from administration site, *distribution*, *metabolism*, and *excretion* (metabolism and excretion are regarded together as *elimination*, and distribution and elimination are referred to collectively as *disposition*) [19]. The variables to be measured in a pharmacokinetic trial are mainly circulating concentrations or excreted amounts of the drug and/or its metabolites.
- *Pharmacodynamic* (PD) issues are those related to the effects of a drug on the human organism, that is, dose and time dependency of the response, effect modulation over repeated administration, such as *sensitization* or *tolerance*, and interference with other endogenous or exogenous agents, such as *synergism* or *antagonism*. The variables to be assessed during a pharmacodynamic trial are biological effect markers, or *biomarkers*.

During the past decade, the advantages of simultaneously addressing both groups of questions in clinical pharmacology trials have been recognized. Such *pharmacokinetic–pharmacodynamic* (PKPD) studies are increasingly performed [20]. They require an elaborated methodology for the analysis of their results, which take specifically into account the mechanisms of translation of concentration profiles into effects, as they can be observed clinically [21].

***Usual Study Objectives*** The aim of a clinical pharmacology study can be a mere *description* of the average trends that characterize dose–concentration (PK), or dose–response or concentration–response (PD) relationships. However, many investigators will also propose an *interpretation* of study observations, with reference to a theoretical model (e.g., compartmental modeling in PK trials, or $E_{max}$ modeling in PD trials). This approach provides clinical estimates of a set of parameters describing the drug PK, PD, or PKPD profile, according to the model chosen. Those parameters can then be used to explain or predict the influence of *specific factors* on the

pharmacological profile. The type of factors entering into consideration delineate another way to classify clinical pharmacology investigations, based on the comparisons planned in the study protocol:

- *Dose–range trials* investigate the influence of the dose administered on drug PK descriptors (e.g., pharmacokinetic *linearity* issues), or PD or PKPD profile (*dose–response* issues).
- *Demographic characteristics trials* examine the influence of gender, age, race, or body weight on a drug PK, PD, or PKPD parameters. This basic approach is applied, for example, to the display of results in most PK trials.
- *Drug formulation trials* examine how the route of administration and galenical conditioning of an active ingredient influences its PK, PD, or PKPD profile. *Bioavailability* and *bioequivalence* trials enter into this category, the latter being oriented toward the demonstration of an acceptable degree of similitude between two different formulations of a given pharmacological agent, for example, an original and a generic brand (which raises specific statistical issues).
- *Food effect trials* investigate the influence of food intake, mainly on the PK parameters of the study drug.
- *Pregnancy trials* aim to determine how childbearing alters the pharmacological profile of a drug. Some investigations also include an evaluation of the transplacental passage of the drug agent to the fetus (e.g., through measure of the drug in cord blood at delivery). The secretion of a drug agent into breast milk can be determined in women subjects during *lactation*.
- *Drug interaction trials* examine the modifications induced by a specific co-medication in the PK, PD, or PKPD characteristics of a given drug.
- *Organ insufficiency trials* assess the impact of a given degree of failure affecting a given physiological system on the PK, PD, or PKPD profile of a drug. While most such trials investigate alterations of drug disposition in subjects suffering from kidney or liver failure, other types of *drug–disease interaction* are sometimes conducted. Drug *dializability* studies are to be cited here as well.
- *Pharmacogenetic trials* try to demonstrate the influence of specific individual genetic traits on the pharmacological profile of a given drug; among classical examples for such investigations are those addressing genetic polymorphisms of cytochrome P450 enzymes or drug receptors.

The combination of those study objectives with the PK, PD, or PKPD principal aims outlines a general typology of human pharmacology trials. Not infrequently, several objectives are combined in a single trial (it is, however, recommended to hierarchize the study questions, and to predefine one or a few main questions for confirmatory analysis, while keeping an exploratory status for all other comparisons).

***Classes of Study Designs*** A further axis for study classification relies on the *trial design*. While many variations can be imagined with regard to the questions addressed and the context of the study, trials can be roughly classified into one of the following main categories:

- *Sample-rich trials* correspond to the classical pharmacological experiments in human subjects, as they were performed by innumerable investigators over more than one century. Such trials continue to be required for new drug registration and represent the most straightforward way to bring answers to well-defined scientific questions of clinical pharmacology [22]. Traditionally, they include a limited number of subjects (between 5 and 50), while each subject provides a large amount of data (repeated concentration and/or effect measurements).
  (a) *Single-dose trials* apply PK and/or PD measurements to the follow-up of one drug administration. Such trials are often built up in *cross-over*, that is, the study subjects receive several single doses (using different dose levels, formulations, interacting medications, etc.) according to a given sequence, the administrations being separated by wash-out periods. This design enables the subtraction of difficult-to-control individual determinants of drug exposure and response, which cancel out during within-subject comparisons. Ideally, the sequence would be randomized to minimize confounding influences such as time-related trends. However, during early development of new products, a *dose-rising sequence* is often preferable for safety reasons; the inclusion of interspersed placebo periods allows randomization and blinding. Placebo control and blinding are usually considered of little relevance in PK trials but are recommended in principle when PD measurements are performed, to evaluate and subtract the baseline response profile observed after a zero dose of the test drug. Factors such as demographic or genetic characteristics can only be tested according to a *parallel* design.
  (b) *Repeated-dose trials* imply a prolonged period of observation and therefore often follow a parallel design. Such trials with *longitudinal* follow-up and analysis are necessary to observe time-related trends in exposure and response, for example, concentration accumulation, autoinduction of metabolism, or tolerance to drug effects. Sometimes, a *transversal* snapshot over one dosing interval is considered sufficient to assess some aspects of the drug PK or PD profile in patients treated on the long term.
- *Population trials* have been proposed since about the last three decades and rely upon data analysis techniques based on pharmacostatistical modeling. Such techniques allow for the estimation of pharmacologically relevant PK and PD parameters from observations drawn in sparse sample conditions, that is, when a large number of subjects provide few data each. Observations can thus be collected on a large scale directly among the target population of patients receiving the treatment, under less stringent study conditions than in classical trials. Population modeling techniques can accommodate varying dosing levels and schedules, unbalanced designs, and fragmentary data. They enable the estimation of both the mean values and the variability of PK and PD parameters among the population. Moreover, the influence of demographic factors, genetic markers, food effects, organ insufficiency, and co-medication can be assessed, provided that the trial population covers a sufficient range of individual differences. While subject heterogeneity is considered as a nuisance in traditional sample-rich trials, it is rather an advantage in population trials. Thus, population approaches have received much attention from both drug

developers and registration authorities [23], and they contribute currently to extend the scope of pharmacological characterization of drugs throughout their clinical development [24–26].

## 2 PHARMACOKINETIC CHARACTERIZATION OF DRUGS

During clinical PK trials, drug assay techniques are applied to a series of biological samples to determine the concentration profile that develops following drug administration; the results are analyzed to describe the fate of the drug, and possibly its derivatives, in the organism. This analysis, interpreted with reference to the classical laws governing the absorption and disposition of drugs, provides estimates of the *pharmacokinetic parameters* that characterize the drug studied. Such parameters represent specific values appearing in the pharmacokinetic equations used to model the drug PK profile. The parameter values are specific not only for a given drug but also for a given type of individual in a given condition. Many pharmacokinetic investigations will include a search for important factors that influence the parameters to a clinically significant extent. Knowing a drug PK profile is a prerequisite for designing rational therapeutic regimens and individualization strategies.

### 2.1 Drug Assay Methods

***Key Steps of Drug Assay*** A PK trial has to rely on efficient and valid laboratory methods. The measurement of drugs and their metabolites in plasma, urine, or other fluids is most often achieved using chromatographic methods. Without entering into details of analytical chemistry, we just summarize here the main steps of a drug assay:

- *Preanalytical processing* covers the operations done on the sample between its collection and assay. Blood will usually be separated into either plasma or serum by centrifugation, with or without anticoagulant, respectively (although certain assay methods are developed specifically for whole blood). For poorly stable compounds, the sample may have to be added with preservatives (pH or oxidation modifiers, chelators, or enzyme inhibitors). In general, a series of samples will be stored under refrigeration until analysis. The influence of storage conditions on sample stability must be checked. Not infrequently, the samples will have to be sent frozen (e.g., on carbonic ice) to a distant lab. Preanalytical processing can be a source of serious bias in PK results: The tested drug may undergo spontaneous degradation or adsorption on the walls of sample tubes, or additives such as anticoagulants may interfere with the assay. Another important issue is *sample labeling*, which has to be thoroughly designed and organized to avoid any confusion between samples. A proper sample label should indicate the study name, the study site, the patient identification number, and the date and time of its collection; it must survive in storage conditions. All sequential tubes used in preanalytical processing must be labeled, working conditions must be organized to minimize the risk of confusion, and the technicians must be taught to "always read the tube they hold."

- *Extraction* is often necessary to avoid slandering of analytical devices by protein-rich fluids. Protein denaturation, for example, by trichloroacetic acid, followed by centrifugation is often sufficient. *Solid-phase extraction* on microcolumns or *liquid-phase extraction* in an organic solvent may achieve a rough preliminary separation and compensate for variations in the sample matrix (e.g., urine concentration). It is important to check the degree of completeness of extraction through recovery experiments at various concentration levels. *Microdialysis* or *ultracentrifugation* techniques will enable the specific extraction of the unbound moiety of a drug in plasma or serum and the determination of free circulating drug concentrations. Such techniques are delicate to implement and highly sensitive to preanalytical processing conditions.
- *Separation* represents the central step in a chromatographic assay. It may be achieved in liquid phase, with a solvent eluting the components of the sample extract along a solid column under high pressure (*high-performance liquid chromatography*, HPLC) or in gaseous phase along a tiny capillary tube (*gas chromatography*, GC). Numerous choices are available to optimize the separation process, which depends on the relative physicochemical affinity of the analytes for the mobile and fixed phases. It is frequently possible to develop a method enabling the simultaneous separation of a drug and its metabolites, or of several drugs in a single run. On the other hand, different components of the sample happen to come out of the separation column at the same time, which represents an important risk of interference when nondiscriminative detection methods are used.
- *Detection* of the analytes at the exit of the separation column enables their quantification. Traditional HPLC methods use spectrometric, fluorimetric, or amperometric detection, with a modest discriminative power. More recently, mass spectrometry methods have been implemented, enabling detection of both high specificity and sensitivity, and thus decreasing the requirements on separation. *Gas chromatography* and *liquid chromatography with mass spectrometry* (GCMS and LCMS) assays have imposed a new standard on PK analytical methods, while both increasing the sensitivity and specificity of drug determination and decreasing sample processing time. The addition of ion fragmentation and mass spectrum analysis of fragments adds a further degree of specificity; it even enables the qualitative identification of chemical components in the sample, with reference to a database of fragmentation spectra.
- *Quantification* uses the signal issued by the detection apparatus. It usually consists of integrating the electronic signal over the duration of the peak corresponding to the analyte of interest. The value is then transformed into a concentration level according to a calibration curve. Many choices are available for the calibration method and may significantly impact on the performance of the assay (curve shape, regression method, coverage of the dynamic range by calibrators). It is frequently recommended to use an *internal standard*, that is, a chemical compound different from the analyte, which is added in fixed amount into the sample before extraction and separation, and detected and quantified on its own. Normalizing the measurement signal of the analyte by the signal of the internal standard will compensate for variations in the whole assay process and can significantly improve its reproducibility.

- *Quality control* determines the acceptance of a series of results. It is usually done by including a few *internal control* samples of known concentration and checking whether their results lie in a predefined acceptance range (e.g., ±10%). In case of problematic precision, one may decide to perform all measurements in duplicate or triplicate. For drug measurements widely performed, programs of *external quality control* may be available, with one center sending samples of undisclosed concentration to all participating labs, receiving the results and analyzing the respective performance of each lab. If only one lab performs a given drug assay, it can be a good idea to provide *blinded duplicates*, that is, to separate certain study samples into two tubes and send one with a coded label.

Other analytical methods are used as well in clinical PK besides chromatographic assays and can be roughly described according to similar steps. For example, biotech products are often poorly characterized at the molecular level, as they represent a mixture of numerous chemical species (e.g., proteins with different stages of glycosylation). For such types of drugs, immunoassays are often preferable, with antibodies ensuring the step of analyte separation; the detection is ensured by radioactivity counting (*radioimmunoassay*, RIA), enzymatic reactions followed by spectrometry (*enzyme immunoassay*, EIA), or *fluorescence polarization* (FPIA). One frequent problem is the cross-reactivity of drug metabolites, either active or inactive, or other sample constituents with the assay method. Another interesting family of separation techniques uses a preparation of drug receptors instead of antibodies: such *receptor assays* are able to specifically measure the sum of chemical species showing affinity for the pharmacological drug target. Going one further step in the direction of pharmacodynamics, *bioassay* techniques directly measure the pharmacological activity contained in the sample (e.g., inhibition of bacterial growth by antibiotics). Such assay methods are often of high relevance, however, of limited performance.

***Elements of Assay Validation*** Whatever the assay technique considered for a PK trial, it must be set up methodically, and its performance and reliability must be controlled in a preliminary validation step, before it can be applied to study samples. The elements to be collected during this step have been listed in the recommendations published following the Washington Conference on Analytical Methods Validation [27], considered as a basic document for good laboratory practice in the field of PK assay methodology. Similar elements are mentioned in the Q2 guideline of the International Conference on Harmonisation (ICH), which, however formally, address pharmaceutical quality control rather than PK determinations [28]. Briefly, the following pieces of information are required to characterize the performance of an assay method:

- *Accuracy* is the extent to which results generated by the method agree with the "true value." It is usually determined by assaying control samples of known concentration (e.g., serum spiked with precise amounts of the analyte). Accuracy is expressed as a *relative bias*, which should not depart statistically from zero.
- *Precision* is the extent to which the individual results of multiple measurements of a series of standards agree together. It is expressed as a *coefficient of*

*variation*. Intraassay (also termed *repeatability*), interassay and possibly interlab (also termed *reproducibility*) components of variability can be identified using appropriate tests.

- *Sensitivity* is given by the *limit of detection*, that is, the point at which a measured signal is significantly larger than the background noise (signals from blank samples), allowing to conclude that some analyte is definitely present in the sample; and by the *limit of quantification*, that is, the minimum concentration that allows precise measurement. The latter limit, together with the upper level of concentrations still measurable with linearity and precision, define the *dynamic range* of the analytical method.
- *Linearity* is the ability of the assay to elicit results that are (either directly or by means of a given mathematical transformation), proportional to the concentration of analyte in the samples within a given range. It can be determined by serial dilutions of a highly concentrated sample.
- *Specificity* (or *selectivity*) is the ability of the assay to measure unequivocally an analyte in the presence of chemical interferences, such as drug excipients, degradation products, co-medications, impurities, or other components of the sample matrix. Its assessment requires extensive search for cross-reactive substances.
- *Robustness* is the resistance of the assay methods toward changes in technical conditions of operation (e.g., temperature, processing time, batch of reactants, etc.).
- Finally, the period of assay validation is a suitable time to check *sample stability* issues under various collection and storage conditions (type of tubes, matrix effects, and length and temperature of storage).

### 2.2 Sample-Rich Trials

In traditional pharmacokinetic experiments, a panel of individuals are administered one or several standard doses of the drug studied and provide a series of blood samples (and possibly urine collections) for the determination of drug concentrations [29]. Practical aspects of subject selection, drug administration, sample collection, and data analysis are presented in this section.

***Selection of Subjects*** Many PK trials are performed in healthy young volunteers, considered as reference subjects for the assessment of PK parameters of a given drug in "normal conditions." Any departure from normality would raise a risk of biased evaluation of the drug PK profile. Therefore, rather stringent inclusion criteria are usually applied to ensure that the study population will actually correspond to this goal. The list below enumerates inclusion criteria frequently applied in PK study protocols on healthy subjects.

- As a principle, both male and female subjects should be included so that the results reflect possible sex-related differences and avoid selection bias [30]. Exceptions are for treatments indicated only in either gender or expected to seriously interfere with reproduction (impact on menstrual cycle, potential for teratogenic activity). Pregnant women should not be included, as pregnancy

significantly influences the absorption and disposition of numerous drugs [31]. Contraceptive use is usually permitted, however, potential interaction of the study drug with contraceptive agents should be taken into account. A pregnancy test may be included in lab screen.

- "Young volunteers" are usually considered between 18 and 45 years. Beyond age 30, a progressive decline of many physiological functions by roughly 1% per year is reported in the average population [19]. It probably results less from programmed physiologic changes than from a trivial increase in the probability of facing sequels of diseases and traumas. In any case, there must be some predefined cutoff to avoid significant impact of such changes on the PK study results.
- Significant underweight or overweight subjects should be avoided as well. Usual weight limits range between 55 and 95 kg for men and from 45 to 80 kg for women, with a body mass index (BMI, i.e., body weight divided by squared height) between 18 and 28 kg/m$^2$.
- The medical history and physical examination should not reveal "significant findings," that is, any condition susceptible to bias the evaluation of normal PK parameters or to impair the safety of exposure to the study drug. A history of severe drug allergy or atopic condition is frequently considered as an exclusion criterion, as are hypertension, cardiac disease history, or significant electrocardiographic abnormalities (including prolonged QT interval).
- Routine laboratory tests should be ordered (blood count, electrolytes, proteins, kidney, and liver tests), and subjects having clinically significant laboratory abnormalities should not be included. Noticeably, athletic young men can have rather high serum creatinine values without any renal impairment due to abundant production by the muscular mass. Gilbert's syndrome (a rather frequent condition with increased total and unconjugated bilirubin on fasting, due to impaired glucuroconjugation) can be accepted if the study drug is not a substrate for glucuroconjugation reactions.
- Urine drug screens for common street drugs are usually requested, to refuse candidate volunteers with current substance use disorder, in which unwanted interactions might happen. The existence of false-positive results for those tests is to be remembered (e.g., poppy seed bread consumption producing positive opiate screen).
- In general, there should be no limitation of race in PK studies. However, it is to be remembered that ethnic factors play a definite role in drug disposition, mainly because the frequency of genetic polymorphisms affecting drug metabolism or transport varies between the different populations across the world (e.g., higher number of CYP2C19-poor metabolizers in Asians, of CYP2D6 ultrarapid metabolizers in Ethiopians [32]). Genotyping of the study subjects for important pharmacogenetic polymorphisms is increasingly requested in PK trials and is now recommended by registration authorities to explore the influence of given genes on the PK profile of new drugs.
- The use of any medication is to be forbidden 2 weeks prior to study drug administration and during the follow-up, including aspirin or other over-the-counter preparations. The use of long half-life drugs with induction or inhibition

potential may affect drug metabolism even more durably (e.g., phenobarbital, amiodarone). Acetaminophen may be permitted before and during the study as a rescue medication with investigator permission.

- The subjects must generally be asked not to consume quinine or grapefruit beverages, as they inhibit CYP2D6 and CYP3A4/5, respectively. Tobacco smoking or exposure to hydrocarbons induce CYP1A2, and alcohol abuse induces CYP2E1. More generally, the study subjects should be asked to refrain from excessive food consumption, prolonged fasting, strenuous exercise, dehydration, and blood donation before and during the whole study period.
- Finally, ability to understand the study conditions, willingness to participate, provision of written informed consent, and appropriate compliance toward study restrictions represent obvious inclusion criteria.

***Dose Determination*** The pharmacokinetic literature reveals a striking contrast between the marked attention devoted to analytical and mathematical aspects of the exploitation of biological samples and the poor consideration given to the pharmaceutical aspects of drug administration. A recent literature survey reported that the assessment of doses actually administered was mentioned only in a small minority of pharmacokinetic trials published in clinical pharmacology journals, while about one quarter of those trials met criteria indicating a theoretical relevance for such an assessment [33]. No mention of procedures for actual dose assessment is found in the recommendations of registration authorities for the performance of pharmacokinetic studies. However, current pharmacopeia requirements set acceptance criteria to ±10% around the nominal strength for the pharmaceutical quality control of most manufactured drugs; moreover, analytical methods used in quality control are usually not identical to assays used for pharmacokinetic determinations, which can quantify similar drug amounts fairly differently (especially when based on immunological reactivity and/or applied to complex agents or biologicals); salt factors may be neglected; and finally, the preparation of solutions for infusion introduces a supplementary risk of dilution errors, leakage, residues, adsorption on syringes or lines, or inaccuracy in infusion rate and volume. All those factors can be expected to introduce significant bias in the determination of pharmacokinetic or pharmacodynamic parameters if the calculations rely upon the labeled dose instead of the best estimate of the dose actually administered. The following pragmatic steps are thus recommended to implement actual dose assessment when necessary in pharmacological trials involving human subjects:

- While dose determination is certainly not necessary in all kinds of trials, the question of the relevance of such assessment should be raised for each study at the time of protocol elaboration. Various criteria should suggest investigators to plan such procedures in a given trial: use of intravenous injections or infusions, experimental drug manufactured on a small scale, especially peptides, proteins, or other complex agents (e.g., phytopharmaceuticals), immunologically based analytical methods (e.g., RIA or EIA), parallel-group design with qualitatively different treatment assigned to each group, and estimation of a drug absolute clearance, volume, or effective dose parameters (the calculation of which include a dose term).

- For trials using a ready-for-use drug formulation, such as an oral tablet, a global evaluation of the average actual dose present in the trial drug is usually sufficient. Pharmaceutical quality controls should be checked, and their results should be used if they indicate any significant departure from the nominal dose of the formulation. The identity and actual titer of the substance used to prepare the calibrators for drug assay in biological fluids should be checked as well (taking a possible salt factor into account). For complex preparations such as biological agents, one may consider determining actual dose amounts using the assay method developed for plasma samples.
- For trials using a drug in solution, typically for intravenous infusion, the actual dose delivered to each study subject should be assessed individually. First, the volume of solution actually infused will be carefully measured and recorded (use the indications given by an electronic pump after a preliminary validation of its flow rate; alternatively, weight precisely the solution vials or bags connected with lines before and after drug administration, and divide the difference by the specific gravity of the solution measured on a sample). Second, the concentration of active drug in the infusion solution will be determined using the assay applied to plasma samples (which must have been validated for aqueous matrix). Whenever feasible, prepare more infusion solution than required, and retain samples from each treatment at the outlet of the line (after postinfusion weighting).
- The differences between assigned dose levels and the amounts actually administered should be reported, and the cause for significant departures should be elucidated. Attention should be paid to the accuracy and precision of the measurement and assay methods used for dose assessment, to avoid the unnecessary introduction of variability in PK or PD calculations (the reproducibility should be reported as standard deviation or coefficient of variation).
- Whatever the approach, the procedures applied to actual dose assessment should be described in study reports and publications, and the results used in the calculation of PK or PD parameters.

In PK experiments, it is important to properly standardize the conditions of drug administration. Unless specified by the protocol, oral dose intake should occur on an empty stomach, hence the subjects are to remain fasted over several hours before. A predefined amount of water is to be given with the oral dose. Even the posture of the subjects should be standardized, as gastrointestinal absorption, hepatic first pass, and renal excretions are influenced by body position. Similarly, the site of a subcutaneous drug injection is to be standardized across the whole trial. Further precautions are necessary for topical drug administration. Finally, it is important to hold a comprehensive drug accountability throughout the study, and to preserve all drug containers and packages until the end of the study.

***Sample Collection*** Study protocols of PK trials usually define a series of sequential sampling times. The goal is to cover with sufficient thoroughness the concentration curves to be described. Ideally, the duration of PK follow-up should last over at least three to four half-lives of the drug. Sampling should be more frequent immediately after drug administration and around the peak time or immediately

after infusion cessation; thereafter, the samples can be separated by larger time intervals. However, the terminal phase is often the most relevant part of the PK profile to be described, as it determines effect duration and accumulation.

Several pragmatic recommendations deserve to be mentioned with regard to sample collection:

- Prepare a gallery of sampling tubes with appropriate labels, along with a worksheet to be filled in during the experiment. The worksheet comprises a sample record log (Fig. 1). Ensure that the labels correspond to worksheet designations for each tube and that consistent designation rules apply to all subsequent produres.
- Record on the worksheet the precise clock time of drug administration(s) and blood samplings. While the protocol defines relative sampling time after dose administration, it is convenient to write down on the worksheet the absolute times scheduled for sampling (which depend on the actual dosage time). In case of an infusion, the worksheet will mention start and end times, infusion flow rate, weight of the pouch before and after infusion, and collection of a retain sample.
- For repetitive blood samplings, it is suitable that the subjects are equipped with a venous cannula with cap and lateral valve. Large forearm veins are best suited to draw blood (avoid antecubital and dorsal hand veins for subjects convenience). In case of a drug infusion, use different venous accesses for drug administration and blood sampling, as sampling through the same line has every chance to contaminate the samples with the infusion solution (which is about 1000 times more concentrated).
- Before drawing samples, discard 0.5 mL of blood (to flush the cannula dead volume); immediately after sampling, recap properly the cannula and rinse it with 2 mL normal saline through the lateral plug. Normal saline is sufficient to main the permeability of venous access over several hours; anticoagulants are unnecessary and may only cause problems (interference with assays, sensitization). Tell the subjects to ask for supplemental rinsing in case of blood reflux observable in the tip of the cannula.
- When collecting a blood sample takes a significant amount of time, record the midsample time on the worksheet and mention this difficulty in a remark. Whatever the system used for blood collection, drawing blood gently rather than trying to pull it out with a strong suction pressure will always do it better. Blood collection may be facilitated by warming the arm bearing the cannula (as a cold skin associated with increased venous tone and decreased blood flux through superficial veins). If it is impossible to obtain blood, report the problem and keep the empty tube in the series.
- If urine collections are scheduled in the study protocol, record precisely each micturition time, and measure the total collection volume before taking samples (it is advisable to weight precisely the recipient and subtract the tare; as urine density varies between 1000 and 1030 g/L, the measurement error is not expected to exceed 3%, which is often more precise than direct measurement of the volume in a column). Offer the subjects a quiet place and ask them to empty their bladder as completely as possible. Ensure sufficient provision of water

**Fluoxetine Bioequivalence Study – Clinical Investigation Unit**
**Worksheet for Sample Collection**

| | | | |
|---|---|---|---|
| **Subject** | LAR | **AN** | 101 |
| **Period** | 1 | **Date** | Apr 14, 05 |

| Sample label | Relative time (h) | Scheduled time (hh:mm) | Actual time (hh:mm) | Visa | Remarks |
|---|---|---|---|---|---|
| -30 min | -0.5 | 07:30 | 7:43 | TB | |
| (dose intake) | 0 | 08:00 | 8:12 | TB | Dose intake with 200 ml water; OK |
| 1.0 h | 1 | 9:12 | 9:12 | TB | |
| 2.0 h | 2 | 10:12 | 10:12 | TB | |
| 3.0 h | 3 | 11:12 | 11:17 | TB | Delay due to difficult blood collection |
| 4.0 h | 4 | 12:12 | 12:12 | TB | Subject mentions nausea (see AE form) |
| 5.0 h | 5 | 13:12 | 13:12 | TB | |
| 6.0 h | 6 | 14:12 | 14:12 | TB | |
| 7.0 h | 7 | 15:12 | 15:12 | TB | |
| 8.0 h | 8 | 16:12 | 16:12 | MA | Difficult blood collection |
| 9.0 h | 9 | 17:12 | 17:12 | MA | |
| 10.0 h | 10 | 18:12 | 18:18 | MA | Inadvertent delay |
| 12.0 h | 12 | 20:12 | 20:05 | MA | Subject on a hurry to leave |
| 24.0 h | 24 | Apr 15, 8:12 | 8:22 | TB | Inadvertent delay |
| 48.0 h | 48 | Apr 16, 8:12 | 8:12 | TB | |
| 72.0 h | 72 | Apr 17, 8:12 | 8:12 | TB | |
| day 5 | 96 | Apr 18, 8:12 | 8:18 | MA | |
| day 7 | 144 | Apr 20, 8:12 | void | TB | Subject forgot the visit; unreacheable |
| day 14 | 312 | Apr 27, 8:12 | 9:25 | TB | Subject late on visit; warned |
| day 21 | 480 | May 4, 8:12 | 8:12 | MA | |
| day 28 | 648 | May 11, 8:12 | 8:17 | JB | |
| day 35 | 816 | May 18, 8:12 | 8:12 | MA | Subject late on visit |
| day 42 | 984 | May 25, 8:12 | 9:02 | MA | Subject late on visit |

***FIGURE 1*** *Example of worksheet for sample recording over one period in a PK trial, where 30 subjects received twice 20 mg of fluoxetine orally (Prozac versus generic) at 2-month interval.*

over the investigation. Calculate the duration of each collection as the difference between end of collection time and initial micturition time; recall that the initial urine does not belong to the collection and is to be discarded. If a subject needs to urinate before the scheduled collection end, save that urine in the container. If a subject forgets to give his or her urine, ask for an evaluation of the missing volume and record the deviation on the worksheet.

- Process the sample tubes as soon as possible and store them in structured racks and boxes, at the appropriate temperature. It makes sense to dispatch the plasma or serum of each blood sample into two tubes, so as to obtain two parallel collections, including one for backup. Ensure that the storage conditions are respected up to sending the samples to the lab. Temperature "spy" recorders may help in monitoring the storage conditions. Refrigerators should be connected to an alarm system to prevent unrecognized malfunction.
- Have all samples shipped to the laboratory, along with the list of tubes. The study samples will most often need to be shipped frozen on dry ice, and packaging issues are of critical importance. The lab should immediately check the completeness of the sample series. If possible, all samples of one subject should be analyzed on the same day.

***Description of PK Results*** Once the lab has produced measurement data for the series of samples, after careful quality check, the first step is to produce graphics of *individual time–concentration profiles* (Fig. 2). Not infrequently, the graphics will identify remaining errors in data handling or suspect data points requiring reanalysis. The actual sampling times are to be used to draw individual PK curves. In most situations, the use of a log *Y* scale is advisable, as it will linearize the exponential components of the time–concentration relationship and indicate the most suitable model to describe the data. The issue of results below quantification limit (BQL) has been abundantly discussed in the literature, and many authors have issued recommendations on that specific question. When sufficient amounts of data are

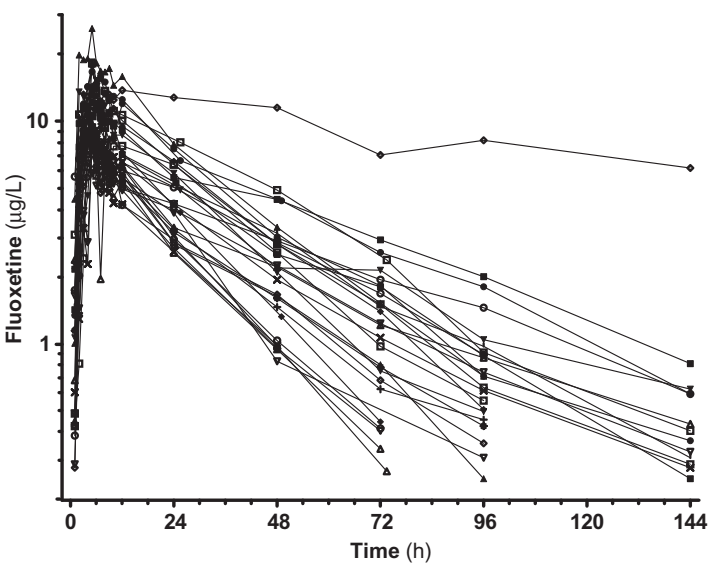


***FIGURE 2*** *Individual graphic of raw data in a PK trial, where 30 subjects received 20 mg of fluoxetine orally (generic formulation). A log Y scale is used. Observe several delayed samples, one suspect point at 6 hours, the fairly log-linear shape of the terminal phase, the decrease in relative measurement precision at low concentration levels, and the presence of one slow eliminator among the subjects.*

available (e.g., in Fig. 2), the best solution is to merely drop such data points from individual curves and calculations. For the calculation of means and during population analyses, a simple approach is to equate the first BQL point to half the value of the quantification limit, drop all subsequent points, and force the residual error to include at least the range between zero and the quantification limit. This will somewhat correct a selection bias of data points, which would otherwise lead to overrepresentation of higher values over BQL, and thus to overestimation of terminal half-life.

In a second step, an *average time–concentration profile* can be computed and plotted (Fig. 3). It is highly recommended to compute geometric means with corresponding coefficients of variation and to plot them along a log $Y$ scale (see Appendix ). Several average curves can be compared on a common graph when the PK experiment involves different doses, formulations, co-medications or subgroups of subjects (according to gender, age, genotype, clinical condition, etc.).

Most sample-rich PK trials will apply a "two-step" approach in data analysis, that is, the characteristic values describing the PK profile are first calculated for each subject individually, and then submitted to statistical tests. The classical descriptors of PK profile curves are as follows:

- The *maximum concentration* ($C_{max}$) corresponds to the highest observed value in the curve. This descriptor is of significance mainly for drugs administered through an extravascular route (oral, subcutaneous, etc.). It is clearly dose dependent, and it represents for some drugs the best predictor of pharmacodynamic efficacy.

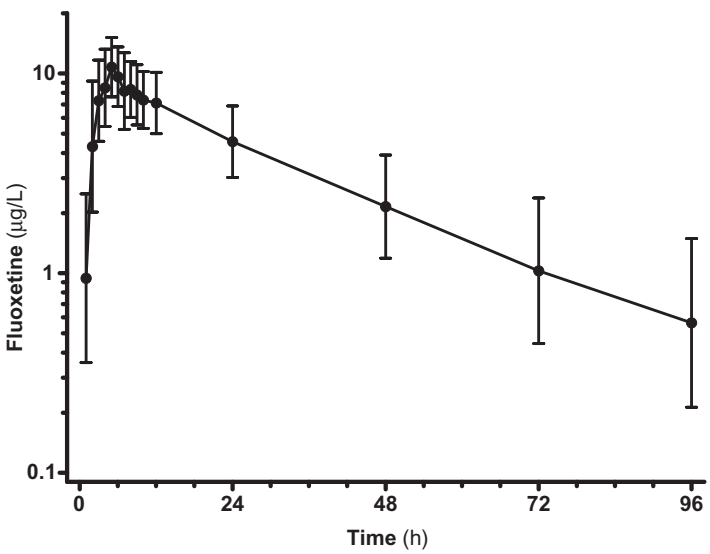


***FIGURE 3*** *Individual graphic of average data in a PK trial, where 30 subjects received 20 mg of fluoxetine orally (generic formulation). A log Y scale with geometric mean and standard deviations is used. Based on the graphic of individual values (Fig. 2), the characterization of average curve appears appropriate up to 96 hours after dose. Observe the relative homogeneity of the dispersion, however, slightly increasing toward low concentration levels.*

- The *time of maximum concentration* ($t_{max}$) is merely the time corresponding to the observation of $C_{max}$. It is affected by both the rate of absorption and the rate of disposition of the drug, according to the pharmaceutical formulation studied. It is usually determined with less precision than other PK descriptors, due to the sparse distribution of sampling times and to the sensitivity of the concentration plateau to slight measurement errors.
- The *rate constant of the terminal phase* ($\lambda_z$) is the absolute value of the slope of log-transformed concentration values versus time (only natural logarithms must be used). It is computed by linear regression over the terminal part of the curve, including all points that appear sufficiently well aligned on the graph (e.g., in Fig. 1, the data points would be included beyond 10 hours after dose intake). Various algorithms have been devised to select formally the points to be included, but visual inspection is often sufficient. When a drug is given at different doses, the quantification limit of the assay often restricts the possibility of estimating $\lambda_z$ at the lowest doses, producing artificially high $\lambda_z$ values; this must be differentiated from the case of true PK nonlinearity, where increase in dosage actually slows down drug elimination.
- A distribution phase is sometimes identifiable on the log concentration versus time curve, translating into a steeper initial segment; this is mainly the case after intravenous drug injection. A *rate constant of the initial phase* ($\lambda_1$) can be computed in such situations.
- After administration of an intravenous bolus dose, a *concentration at time zero* ($C_{(0)}$) can be back-extrapolated as the intercept of the log-linear regression line fitted over the points of the initial phase.
- The *area under the curve* (AUC) is the most important descriptor to calculate after single-dose administration. It is estimated by numerical integration over all segments of the concentration curve. It is dose dependent, and it represents for many drugs the best exposure marker correlated with pharmacodynamic efficacy. The *trapezoidal rule* is often used to calculate AUC segments: if a concentration $C_i$ is measured on time $t_i$, followed by a concentration $C_j$ on time $t_j$, the corresponding segment has the area

$$\mathrm{AUC}_{i-j} = (t_j - t_i)\left[\frac{1}{2}(C_j + C_i)\right]$$

During the decay phase, especially when samples are taken at sparse intervals, the trapezoidal rule tends to overestimate segment areas, as their upper boundary is in fact curvilinear; using the *log-trapezoidal rule* is thus considered as preferable:

$$\mathrm{AUC}_{i-j} = (t_j - t_i)\left[\frac{C_i - C_j}{\log(C_i) - \log(C_j)}\right]$$

The sum of all segments is then calculated from the first to the last observed concentration ($\mathrm{AUC}_{last}$). However, this sum still underestimates the true AUC, and it must be complemented with extrapolations. An extrapolation between dosage time and the first measured concentration will use $C_{(0)}$ for the estimation

of the first segment area in case of an intravenous bolus injection; for other routes of administration, the concentration at dosage time can be considered to be zero, and the first segment of AUC can be equated to a triangle having an area of $AUC_{0\text{-}1} = 0.5C_1/t_1$. At the other end of the time–concentration curve, an extrapolation of the AUC toward infinity can be based on the last measurable point ($C_{last}$) and the rate constant of the terminal phase:

$$AUC_{last\text{-}\infty} = \frac{C_{last}}{\lambda_z}$$

Some authors recommend to replace the observed $C_{last}$ by a fitted value based on all points of the terminal phase. Both extrapolations $AUC_{0\text{-}1}$ and $AUC_{last\text{-}\infty}$ are added to the sum of segments to compute $AUC_{0\text{-}\infty}$, the best available estimate of total AUC. As the extrapolations represent the least robust part of AUC estimation, it is often required that they do not represent more than 10% of the total AUC, to ensure an accurate estimation of the concentration exposure. In repeated dose trials, when concentrations are followed over a dosing interval at steady state (i.e., once accumulation is complete), no extrapolation to infinity must be applied, as the AUC limited to one dosing interval is equivalent to the AUC to infinity, which would have been observed after a single dose.

- The *area under the first moment curve* (AUMC) is another descriptor of PK curves used sometimes. The first moment curve is the curve of concentration values multiplied by corresponding sampling times. Its integration can be interpreted as an exposure marker assigning bonus to concentration levels according to their duration. The trapezoidal rule for AUMC estimation is

$$AUMC_{i-j} = (t_j - t_i)\left(\frac{t_j C_j + t_i C_i}{2}\right)$$

A log-trapezoidal rule has also been proposed but it is far more complicated and rarely used. The estimation of AUMC has also to be extrapolated toward infinity:

$$AUMC_{last\text{-}\infty} = \frac{t_{last} C_{last}}{\lambda_z} + \frac{C_{last}}{\lambda_z^2}$$

As this descriptor weights concentrations by sampling times, the percent extrapolated is often fairly high, which limits the confidence that can be placed on AUMC estimation.

***Noncompartmental PK Parameters*** Once the descriptors mentioned above have been computed for each individual PK curve, they can serve to estimate various dose-independent parameters, taking into account the actual dose administered ($D$), the administration route, and the study design [34]. The following PK parameters are usually calculated:

- The *terminal half-life* is taken as the inverse of the rate constant of the terminal phase multiplied by the natural logarithm of 2 (i.e., 0.69):

$$t_{1/2z} = \frac{\log(2)}{\lambda_z}$$

In case of intravenous injection or rapid extravascular absorption, $t_{1/2z}$ reflects the drug *elimination half-life*; however, for slow-absorption formulation, it may reflect the *absorption half-life*, which becomes the factor limiting drug disposition. Similarly, the terminal half-life of a drug metabolite measured in blood can reflect the kinetics of either its formation or its elimination. When an initial log-linear segment can be identified on the time–concentration curve, its rate constant can be used to estimate the *initial half-life*, which most often represents a *distribution half-life*.

- The *total drug clearance* is calculated as the ratio of the dose over the resulting AUC:

$$\mathrm{CL} = \frac{D}{\mathrm{AUC}}$$

Notice that after extravascular administration, this formula gives the *apparent clearance* (CL/$F$) rather than the absolute clearance, as the concentration exposure results only from the fraction of the dose able to reach the circulation, which is the *bioavailability* ($F$). When urinary determinations are performed, the amount of drug excreted into the urine ($A_e$), which may require extrapolation toward infinity, can be used to determine the *renal clearance* of the drug:

$$\mathrm{CL}_R = \frac{A_e}{\mathrm{AUC}}$$

The *fraction of dose undergoing renal excretion* is calculated as the ration $A_e/D$. When a drug metabolite is measured in plasma or serum, the AUC of this metabolite is usually presented along with its half-life. However, to determine the fraction of the dose eliminated through the corresponding metabolic pathway requires that the absolute clearance of the metabolite is evaluated after intravenous administration of the metabolite itself. Alternatively, the fraction of the dose recovered in urine as metabolite can provide some indication. Otherwise, the metabolite AUC cannot be assigned a precise meaning, as it results from both the drug transformation into metabolite and the disposition kinetics of the metabolite, which are unidentifiable.

Finally, it is to be recalled that any estimation of a drug clearance refers always to a specific medium where drug concentrations are measured: for most drugs, the plasma clearance will significantly differ from the whole blood clearance. By default, total plasma is considered as the reference medium. There is growing interest for the determination of free drug concentrations in the plasma; for drugs highly bound to plasma carriers (albumin, α-1-glycoprotein, lipoproteins, etc.), as free concentrations are very low, *free drug clearance* can

reach impressive values, especially in case of high liver extraction ratio, when the metabolic activity of the liver is able to displace the binding equilibrium in plasma during blood transit through the liver. Free drug concentrations are considered to be more tightly related to pharmacodynamic activity than total plasma levels.

- The determination of the *absolute bioavailability* of an extravascular formulation requires to compare the AUC after its administration with the AUC after intravenous injection of the active ingredient, accounting for possible dose differences:

$$F_{\text{extravascular}} = \frac{\text{AUC}_{\text{extravascular}} D_{\text{intravenous}}}{\text{AUC}_{\text{intravenous}} D_{\text{extravascular}}}$$

When two extravascular formulations are compared, this formula gives the *relative bioavailability* of one formulation with respect to the other (taken as reference). This is the key parameter estimated in bioequivalence trials.

- Finally, several types of *volume of distribution* can be defined. The *volume of the terminal phase* is simply derived from the total clearance and the terminal constant rate:

$$V_z = \frac{\text{CL}}{\lambda_z}$$

Notice that after extravascular administration, this formula gives the *apparent terminal volume* ($V_z/F$) rather than the absolute one. After intravenous bolus injection, an *initial* or *central volume* can be deduced from the concentration extrapolated at time zero:

$$V_1 = \frac{D}{C_{(0)}}$$

When an AUMC value has been estimated after intravenous injection, it can be used to calculate the *mean residence time*: MRT = AUMC/AUC, a weighted estimator of the average duration of sojourn of drug molecules in the organism before their elimination. After intravenous infusion, the ratio AUMC/AUC is equal to the MRT plus one half the infusion duration. After extravascular administration, it is equal to the MRT plus the *mean absorption time* (MAT), which can thus be determined only by comparison with the intravenous administration. Knowing the MRT of a drug enables one to calculate its *volume at steady state* (i.e., when concentrations in the different compartments are at equilibrium):

$$V_{\text{ss}} = (\text{CL})(\text{MRT})$$

Like clearances, volumes of distribution always refer to a given measurement medium, and for agents highly bound to plasma carriers, the distribution volume of the free moiety can exceed many times that of the total plasma drug.

***Pharmacokinetic Modeling*** Beyond the basic calculations outlined above, many more sophisticated calculations can be devoted to the analysis of PK data, depending on the study aims and conditions. A detailed description of all possible approaches is beyond the scope of this chapter [19]. Briefly, the models that are regularly applied belong to the following classes:

- *Exponential models* use nonlinear regression techniques to fit the data with a general equation, usually a sum of exponential terms; for example, a two-exponential concentration curve after intravenous injection will be described according to the equation:

$$y = Ae^{-\alpha t} + Be^{-\beta t}$$

  As PK data are characterized by noticeable heteroscedasticity (i.e., heterogeneity in variability), the issue of appropriate weighting schemes for the regression is of relevance. The coefficients estimated ($A$, $\alpha$, $B$, $\beta$, termed PK *macroconstants*) enable the calculation of classical PK parameters ($\lambda_1$, $\lambda_z$, CL, $V_1$, $V_z$). Notice that the model equation can also be built up using directly clearance and volume parameters. Parameter values estimated this way not infrequently show fair differences with their noncompartmental counterparts (with the exception of CL, which is the most robust regarding estimation methods). Exponential models are ideally suited to draw continuous curves across the data points and calculate various simple extrapolations (e.g., dose repetition).
- *Compartmental models* aim to fit the data with a set of differential equations. The equations characteristically describe the evolution of drug amounts throughout body compartments. For example, the bicompartmental disposition of a drug after intravenous injection will be described according to the set of differential equations:

$$\frac{dY_1}{dt} = k_{21}Y_2 - k_{12}Y_1 - k_{10}Y_1$$

$$\frac{dY_2}{dt} = k_{12}Y_1 - k_{21}Y_2$$

  where $Y_1$ is the amount of drug in the central compartment (related to the concentration through the volume $V_1$ of this compartment), and $Y_2$ the amount of drug in the peripheral compartment. The coefficients ($k_{12}$, $k_{21}$, $k_{10}$, $V_1$, termed PK *microconstants*) enable as well the derivation of macroconstants and classical PK parameters. Compartmental models can accommodate a wider spectrum of conditions, such as time-varying infusion rates, first-pass effect and enterohepatic cycle, multiple metabolites production, or situations where some PK processes are nonlinear, for example, saturable elimination pathways following Michaelis–Menten kinetics. The results are more susceptible to be interpreted mechanistically and can be used in simulations where PK mechanisms are selectively modified (e.g., prediction of the effects of a change in formulation or organ function).

- *Physiological models* try to incorporate physiological variables, such as anatomical volumes of tissues and organs, or circulation flow rates. While those models are mainly used in animal experiments, they can be useful in selected clinical situations, such as the determination of drug kinetics through an organ, either natural or artificial (e.g., renal replacement machines), especially when it is accessible to measurement (e.g., cerebral imaging of drug distribution into the brain).

While the basic PK calculations described above can be achieved using a simple spreadsheet, fitting exponential, compartmental, or physiological models requires specific computer facilities. Numerous pieces of software are available to perform such advanced PK data analyses, most of them being represented on the Internet (see, e.g., the online catalog maintained by Bourne [35]). Among such programs, WinNonLin and Kinetica currently experience a high level of recognition among pharmacokineticists. In addition, several computer packages for general statistics have now implemented specific routines for PK data analysis (e.g., S-Plus, R, and Stata). Nowadays, the limiting factor for a sound exploitation of PK data using computer resources is only the degree of knowledge and experience required from users.

***Statistical Analysis of PK Results*** Once the PK parameters have been deduced from the raw concentration data, they are ready for description, analysis, and comparisons according to the study objectives and design. It is standard practice to compare PK parameters rather than the concentration results themselves. Many specialized PK computer programs incorporate statistical tools to perform tests on the parameter estimates that have been calculated. Alternatively, the parameters may be copied into a standard statistical computer package.

If the main study objective is a mere *description* of a drug PK profile, the importance will be put on the summary table of PK parameters. Several choices are available for statistical description. The calculation of geometric means and associated percent variations should be encouraged, as it is often the most appropriate alternative (see Appendix). Nevertheless, arithmetic means and standard deviations are still often encountered. When parameter values appear to follow an irregular distribution pattern, nonparametric descriptors should be preferred (i.e., median and interquartile range). The descriptive presentation of PK results is often complemented with some statistical comparisons of the parameters according to various factors, such as gender, age, or the genotype regarding given polymorphisms. Such tests must be viewed as *exploratory*, that is, the $P$ values have only an indicative interest, and no Bonferroni-like correction is applied for multiple testing.

When a PK trial is built up toward the assessment of a given hypothesis, the main study objective must be specified in terms of one or a few PK parameters to be tested regarding the influence of relevant factors, and *confirmatory* statistical tests must be predefined in the protocol. A preference should be given to parametric tests after preliminary log transformation of the parameters. For example, in a metabolic interaction study following a cross-over design, the effect of a given treatment on the log of the clearance of the tested drug is to be assessed through an analysis of variance; in a renal failure study including patients with various degrees of renal impairment, the correlation of the log of the clearance of the tested drug with the log of the glomerular filtration rate is to be evaluated. When the distribution of PK

parameters departs significantly from a log-normal, nonparametric tests should be applied instead.

A particular case is represented by *bioequivalence* trials, which are intended to establish the absence of clinically relevant difference between two formulations of the same agent. In this situation, it is not sufficient to show an absence of statistically significant differences. The ratio of the parameters of interest (AUC, $C_{max}$, $T_{max}$) between both formulations is to be calculated in each study subject, and the bioequivalence can be claimed only if the statistical confidence interval at the 90% level around the geometric average of this ratio lies entirely inside a predefined range. Registration authorities set this range between 0.8 and 1.25 for AUC and $C_{max}$ and between 0.7 and 1.4 for $T_{max}$ [36, 37]. This is roughly equivalent to performing two one-sided *t* tests, and to conclude that the probabilities that the "true" ratio lies either below the lower limit or above the upper limit are less than 5% each [38]. A similar approach can be used to establish the absence of influence of food or of a potentially interacting drug on the PK profile of a given agent. Another investigation oriented toward showing an absence of difference is the assessment of PK linearity, usually performed by comparing dose-corrected AUC values. Here again, all the statistical calculations are to be done on log-transformed values (as suggested by the asymmetrical acceptance range for the confidence interval around the ratio of parameters of interest).

### 2.3 Population PK Studies

Population PK studies differ in many aspects from the classical sample-rich trials, while their general aims remain similar, namely to characterize the absorption and disposition profile of drugs in humans. They must rely as well on efficient, reliable, and validated analytical methods. Their importance is increasingly recognized in both academic circles and registration authorities, and they are progressively invading drug development [39]. They are not aimed to replace, but rather to complement, the small-sized trials performed during early drug development in healthy volunteers. They can be easily engrafted, indeed, on other types of studies, such as phase II or phase III therapeutic trials.

***Specificities of Population PK Study Protocols*** While classical PK trials mainly derive from the tradition of physiological experimentation, population PK studies owe something to epidemiological concerns: drug characteristics should be established from observations gathered in field conditions, where the drug is employed for therapeutic purposes. This explains substantial differences in study protocols and data analysis techniques [40].

The *study subjects* will thus preferably be representative patients having to receive the drug for a medical condition, rather than healthy subjects selected according to strict normality criteria. This does not mean that population approaches cannot be used in phase I trials: They can indeed, only with less relevance. Subject heterogeneity, which is considered as a nuisance in experimental trials, is regarded in population studies as a valuable opportunity to assess the variability of PK parameters and to identify determinant factors affecting their value. A condition, however, is to carefully record what the relevant subject characteristics are (gender, age, body weight, medication, morbidity, renal and liver function tests, genotype results for

polymorphic enzymes or transporters, etc.), so that they can be used as explanatory *covariates* during data analysis.

Population *study designs* are also less stringent: While healthy subjects are able to undergo intensive investigations, involving numerous blood samplings over a short time, medical patients should not be exposed to cumbersome study protocols. Specific population data analysis techniques can accommodate sparse-sample designs. Therefore, instead of obtaining a lot of measures in a restricted group of subjects, population studies will tend to include a larger number of patients (several dozen at least), with each patient providing only a limited number of observations (usually between 3 and 12). This makes, for instance, population studies represent the only way to address pharmacokinetic issues in pediatrics. Moreover, neither the dosing regimen nor the blood sampling schedule need to be rigorously standardized, as population techniques will accommodate unevenness in doses, intervals, blood sampling times, and even administration routes. So the study patients may take their drug and undergo venipuncture on convenient times (this may be initially accepted with perplexity by health professionals, who often associate clinical research with the requirement for strict timetables). Of prime importance is, however, the thorough recording of all actual dosing and sampling times. Dose determination issues need also to be considered (see above). In addition, compliance may represent a noteworthy problem, while it is rarely an issue in classical PK experiments. Compliance determination using electronic devices may thus be suitable in some population PK trials.

This flexibility makes population PK studies relatively easy to append to a variety of clinical trials. In particular, phase II and phase III therapeutic trials may clearly benefit from this kind of extension (which can be applied only to a subgroup of patients, e.g., those in the hospital). This will help to assess concentration–efficacy and concentration–toxicity relationships in the target patient population, to uncover causes of treatment failure or toxicity, to design dosing adaptation strategies for specific patient subgroups, and to optimize rationally the medical utilization of new drugs. A further step will be to pool the PK data gathered during all phase I, II, and III trials of a new drug into a large database, and run a systematic population analysis with proper exploration strategies and confirmation cycles, which will extract all clinically relevant knowledge from concentration data. This has been achieved for some newly developed drugs; however, it represents nonnegligible technical requirements and expenses, in terms of data management, computer facilities, and specialist time [24].

In summary, population PK studies are remarkably flexible toward patient inclusion criteria, drug administration, and sampling time schedules, but this implies that each blood sample tube comes associated with a set of carefully recorded clinical data. The data to record vary from one study to another, and the following tentative list is only indicative:

Patient identity and unique number in the study

Study group and investigation center, when relevant

Usual dosing regimen regarding the study drug

Last dose administered, along with precise dosing time and other relevant details (e.g., route, formulation, infusion rate, compliance assessments, etc.)

Accurate sampling time

Demographic covariates, such as gender, age, body weight, height, and ethnicity

Scores of organ insufficiency (heart, kidney, liver), global functional scores (in chronic disease patients)

Biochemical markers of organ function, such as typically creatinine or bilirubin

Concentration of plasma carriers (albumin, $\alpha_1$-acid glycoprotein), for drugs highly bound

Genotype results regarding relevant polymorphisms susceptible to affect drug disposition (whenever tested, which is increasingly the case)

Concomitant medications, to be classified at the end of data collection according to their potential for interaction with the study drug (e.g., inhibitors of inducers of a given cytochrome P540 isoenzyme)

***Analysis of Population PK Data*** The analysis of a set of observations collected during a population PK study represents a nontrivial assignment. It requires both specific computer facilities and specialized skills [40]. An analysis will include in a single run the PK data obtained in all study subjects. The system will fit them with a family of PK models, based on a common equation (or system of equations), with variable parameters allowed to take different values between each subject.

Formally, if we call $f$ the common equation, without specifying it, the model for the $j$th concentration value observed in the $i$th subject will correspond to the general expression:

$$y_{ij} = f(x_{ij}, \varphi_i) + \varepsilon_{ij}$$

where the $x_{ij}$ are the *fixed effects* that apply to sample $y_{ij}$, such as dose and relative sampling time, the $\phi_i$ are the *individual PK parameters* values that characterize subject $i$, and $\varepsilon_{ij}$ is a *residual error* (i.e., the difference between model prediction and actual observation). The errors $\varepsilon_{ij}$ are assumed to be randomly distributed around zero with a variance $\Sigma$, which quantifies the residual or *intrasubject* variability, a mixture of measurement errors, biological oscillations, and model misfit. More complex error structures may be used, for example, involving both addition and multiplication of error terms and necessitating to define $\Sigma$ as a matrix of variance components. This model for observations is complemented with a second layer of models, which express the variability of the subjects PK parameters and their dependency toward individual factors:

$$\phi_i = g(z_i, \theta) + \eta_i$$

where the $z_i$ are the *covariates* characterizing subject $i$ (which may even vary along time: $z_{ij}$), the $\theta$ are the average *population PK parameters* including constants that correlate individual parameter values with covariates, and $\eta_i$ is a *random effect* affecting the parameter value in subject $i$ (i.e., the difference between its "actual" value and the value predicted equation $g$). The random effects $\eta_i$ are distributed around zero with a variance $\Omega$, which quantifies the unexplained *intersubject* variability, while equation $g$ captures the part of variability that is explained by the covariates. Here again, additive and multiplicative random effect models may be

used. Moreover, there may exist some degree of correlation between the random effects affecting various PK parameters, necessitating to define $\Omega$ as a variance–covariance matrix.

Thus, fitting a population model represents a problem of hierachical, or *mixed-effect* regression, which involves the presence of two levels of errors (a third level is sometimes introduced, leading to distinguish between intersubject, interoccasion, and residual variability terms). The individual PK parameters $\varphi_i$ cannot be directly measured and represent latent variables. In general, the mathematical expressions for the functions $f$ and $g$ above are not reducible to linear equations. The population PK parameters to be found are not only the values of the constants $\theta$ but also the components of the variance matrices $\Sigma$ and $\Omega$ (hence the designation of *pharmacostatistical model*). Their estimation requires one to decide on appropriate fitting criteria and to apply efficient numerical algorithms. Many choices must be done regarding the equations $f$ and $g$, the inclusion of influent covariates among the available $z_i$, and the acceptability of various underlying assumptions. That is why a series of sequential models are usually adapted to a given set of population PK data, requiring the adoption of appropriate decision rules, and giving utmost importance to a rational and clinically guided model building strategy.

Several software tools have been developed for this type of calculations [35]. NonMEM, developed by L. B. Sheiner and S. L. Beal at UCSF, remains the most widely accepted program in the field. While depending on a whole set of assumptions often difficult to validate, it provides a highly flexible framework for population data analysis, which can accommodate a unrivaled variety of study conditions. Several computer tools have been devised to facilitate NonMEM operation, lacking notably from user-friendliness (Wings, PDx-POP, Xpose). Other tools are currently available for population PK analysis, which use different algorithms and will suit a more restricted range of study problems (WinNonMix, P-Pharm, Kinetica, WinBUGS, Monolix, PopKinetics, USC*Pack, etc.). Extension modules for population PK analysis are also available in all-purpose statistical packages (NLME for R and S-Plus, NLMIX for SAS, etc.).

Without going into much detail, we just summarize here the main concrete steps of a population data analysis [40]:

- The preparation of the data set may take a while, as stringent requirements of the computer program must be followed. The drug measurement results must be entered in a structured way, along with all corresponding clinical data (subject identification, dosing schedule, sampling times, and covariate values). Automatized extraction from a database is tricky to set up and requires careful checking.
- The simplest possible population model, chosen according to preliminary knowledge of the drug PK profile, should be adapted initially, without including any covariate. The resulting set of population PK parameters will serve as a baseline for further model building.
- A set of decision rules for the acceptance of refined models should be chosen. It is usual to consider indices of goodness-of-fit such as the improvement in log-likelihood or Akaike information criterion, diagnostic graphics such as scatterplots of predictions versus observations or of residuals versus predictions (taking

both the individual and the population predictions), and commonsense arguments such as the clinical relevance of refinements introduced in the model.

- Usually, model building will begin with the search of an appropriate structural PK model, that is, deciding on the number of exponentials or compartments. The measurement of metabolites or urinary excretion may further complicate the problem. A second issue is the choice of the most appropriate error model, both for intraindividual and interindividual variability. Thereafter, the sequential inclusion of covariates will be tested. *Covariate selection* may be valuably assisted by graphical explorations, showing possible trends between individual parameter estimates and known subject characteristics. It is important as well to be aware of the meaning of each sophistication proposed for the model and to judge its clinical pertinence. Covariate centering is usually recommended; for example, when testing a dependency of distribution volume toward body weight (BW), the expression

$$V = \theta_1 + \theta_2 \frac{\mathrm{BW} - 70}{70} + \eta_V$$

will be more satisfactory than

$$V = \theta_1 + \theta_2(\mathrm{BW}) + \eta_V$$

As in the first case, $\theta_1$ represents the volume of a typical 70-kg individual and $\theta_2$ the change predicted a doubling in body weight, while in the second case, no particular meaning can be attributed to $\theta_1$ and $\theta_2$. Moreover, convergence issues are better dealt with using the first type of expression.

- Once the final model is established, it must undergo a formal validation. Internal validation will be based on careful examination of diagnostic graphics, standard errors and cross-correlations of each parameter, sequential removal of each covariate to test its marginal significance level, estimation of the lack of fit, and the like. Whenever possible, external validation on a different set of data is recommended. Overparametrization is often an issue, and the principle of parsimony should be kept in mind during data modeling.
- The equations and population parameters of the final model should be presented, along with the model building history. A graphical representation of the model is often useful to show the typical concentration curve expected in an average subject receiving a usual dosage; a 90% prediction interval should ideally be drawn around the typical curve, using either simulations or error propagation calculations. Sophisticated visual predictive check techniques are increasingly recommended.

## 3 PHARMACOKINETIC–PHARMACODYNAMIC EVALUATIONS

Most drugs induce measurable effects, which mediate their therapeutic efficacy, and at least in part the adverse reactions that they can cause. Ultimately, human pharmacology trials aim to describe, explain, and predict not so much the concentration

profile than the effect profile of pharmaceutical agents, taking into account the administration scheme and the clinical characteristics of recipients. It is suitable that clinical pharmacodynamic trials include drug concentration measurements whenever possible, as the pharmacokinetic phase of drug action in humans represents usually the major source of variability in pharmacological response [19]. PKPD trials thus tend to break down the analysis of dose–response relationships into the sequential assessment of dose–concentration (PK) and concentration–response (PD) relationships. This section presents some considerations on the latter part of the work.

### 3.1 Pharmacodynamic Variables

Whereas a single type of data is encountered in PK modeling, namely drug concentration measures, PD trials happen to deal with a variety of response modalities. The global aims of both PK and PD investigations in humans remain, however, to record the profile of the respective variables along time. Hence many methodological considerations discussed above for PK studies apply similarly to PD and PKPD trials: necessity to develop and validate accurate, sensitive, specific, and reproducible measurement methods, importance of dose determination, accurate recording of dosage and measurement times, choice of an appropriate study design with respect to the scientific questions addressed, rational approaches for data description, interpretation and modeling, and the like.

All kinds of clinical, physiological, or biochemical measures can be employed as PD variables, as soon as they are shown to change in response to a given drug treatment. Their clinical relevance may, however, vary considerably. Certain variables correspond themselves directly to the *clinically relevant outcome* of a treatment, for example, the level of pain after administration of an analgesic. Other variables are used as *surrogate markers* for treatment outcome, based on sufficient preliminary evidence showing that they play a significant role in the pathophysiology of the disease under treatment and that their correction correlates with clinical improvement in the patient, for example, the levels of viremia and CD4-positive lymphocytes in chronic HIV infection. The evidence allowing a marker to be recognized as a valid surrogate for clinical outcome may be difficult to establish. For example, it has taken decades to confirm that blood pressure or plasma cholesterol levels, beyond their predictive potential regarding the occurrence of cardiovascular diseases, were amenable to pharmacological correction with an actual benefit for the patients; a famous counterexample is the case of ventricular arrhythmia, a condition apparently improved by antiarrhythmic agents, however, at the expense of increased patient mortality. Finally, mere *biomarkers* reflecting drug action are frequently measured during PKPD trials, however, without claim for a strong predictive value with regard to clinical outcome. Such markers may, however, help to understand the clinical pharmacology of new agents, to build up rational dosage strategies, and to develop useful treatment monitoring methods. Noticeably enough, some biomarkers are so tightly correlated with circulating drug concentrations that they can even be used for PK assessments: for example, circulating angiotensin-converting-enzyme (ACE) activity closely reflects blood concentration of ACE inhibitors.

The pharmacometric characteristics of PD variables are important to take into account when designing a PKPD trial. Whenever possible, *graded response markers* are preferable to accurately assess concentration–response relationships using

regression techniques and formal PKPD modeling. This kind of measures are therefore most frequently encountered in clinical pharmacology studies and include, for example, biochemical markers (e.g., glucose, lipids, electrolytes, hormones, cytokines, coagulation tests, etc.) and functional variables (e.g., electrocardiographic or electroencephalographic variables, spirometric flows, psychometric tests, etc.). The measurement of subjective symptoms can also be performed using continuous variables: This is the principal merit of *visual analog scales* (VAS), such as those used to assess pain or sleepiness. The subjects are presented with a 100-mm horizontal line printed on a paper, labeled with "none at all" on the left end and "worst imaginable" on the right end. They are instructed to place a mark on the line to report the intensity or quality of the sensation being experienced. The distance between the left end and the mark is recorded precisely and provides a suitable quantification of the subjective response. Symptoms and clinical signs may also be recorded using validated scores, typically on a 1–4 scale corresponding to clear definitions, for example, for the assessment of tolerability and adverse events (1, mild; 2, moderate; 3, severe; 4, life threatening). Such *ordinal variables* deserve a different treatment during data analysis. Formal validation is not less important for scoring methods than for graded response markers, however, slightly different approaches are used, including test–retest and interinvestigator concordance assessments. Lastly, *binary variables* can intervene in PD studies, such as typically success/failure assessment, for example, for infection cure.

Most effect variables measured during PD and PKPD trials are less precise and less specific than PK measurements, thus making effect profiles more irregular and noisy than concentration profiles. Evaluating the *signal–to–noise ratio* of a response marker is one of the objectives of the procedure of validation and may help to chose the most efficient measure among several available ones. Highly variable markers further increase the requirement for appropriate data analysis approaches, able to deal with imprecision and to extract pharmacologically pertinent trends. The choice of an appropriate *scale* for marker measurement is of importance, as it may dramatically influence data analysis and results interpretation (e.g., expressing the response to antacids as a change in either pH or $H^+$ concentration in gastric juice). Frequently, PD response measures start from a certain *baseline level* that is to be thoroughly assessed before drug administration. Moreover, the level of the response variable may display systematic *fluctuations* even in the absence of an active substance, for example, due to cyclic modifications (circadian, circannual), spontaneous disease evolution, psychological influences, and the like. Therefore, PD and PKPD trials will often include a placebo arm, while this is useless in purely PK trials. Effect variables may also be distinguished according to their degree of *reactivity* or inertia. Some markers quickly respond to drug exposure and return to baseline thereafter, while others integrate the response over prolonged times. A striking example can be found in osteoporosis treatment: A marked response to bone resorption inhibitors can be recorded over a few hours in healthy volunteers while following specific bone resorption markers such as plasma or urine collagen degradation products; on the other hand, a radiological follow-up of bone density in patients takes years to reveal significant changes in this marker, which seems, however, better correlated with the fracture risk. In addition, the sensitivity of a response variable may decrease over time, due to a *tolerance* phenomenon, such as a decrease in number and/or excitability of cell receptors (e.g., with opioids), or the induction of physiological

counteraction mechanisms (e.g., with α-blocker antihypertensives). Inversely, some responses display a phenomenon of *sensitization*, meaning an increase in amplitude over prolonged exposure (e.g., drug addiction or allergy). Appropriate PKPD models are thus required to account for such time-related changes in concentration–response relationships. Finally, the description of a drug effect over fairly long time periods may necessitate to take into account the proper kinetics of pathophysiological events that characterize the disease and that modulate its response to treatment. Such *disease evolution modeling* represents an important research area, leading clinical pharmacology to meet the traditional questions of therapeutics.

### 3.2 Concentration–Response Relationships

***Pharmacodynamic Data Description*** Effect measurement results gathered during sample-rich trials can first be described using simple calculations fairly similar to those developed for PK data description (see above). This basic approach will be useful to summarize the results, apply statistical comparisons, and reveal important trends. The following calculations are applicable even when the effect curve shows some retardation compared with the concentration curve, provided the effect has been followed over a sufficiently long time to capture the major part of its development and resolution:

- The *maximum effect observed* ($E_{max}$) corresponds to the highest marker value in the curve.
- The *time of maximum effect* ($t_{Emax}$) is merely the time corresponding to the observation of $E_{max}$. It is rarely determined with good precision, as it is measured on effect plateau.
- The *area under the effect curve* (AUEC) can be estimated by numerical integration over all segments of the effect curve. The *trapezoidal rule* is generally used to calculate AUEC segments. As it is exceptional that a model is proposed to fit the terminal part of the effect curve, no extrapolation to infinity is usually added to the sum of AUEC segments.
- Even if some usefulness could be imagined for an area under the moment of the effect curve, no such calculations are usually performed.

When sufficient amounts of data are available, those PD descriptors can be submitted to simple statistical tests to assess the effects of dose, demographic factors, medical condition, co-medication, or genotype and to explore their correlation with the elementary PK descriptors. Among various questions in consideration, the identification of the best PK predictor for the effect is of relevance: For example, among the antibiotics, one can distinguish agents with an efficacy depending on peak concentration (e.g., aminoglycosides), on the time of concentrations exceeding a threshold (e.g., penicillins), or on the AUC (e.g., macrolides) [41]. When different effects depend on diverse PK descriptors, manipulating the dosing regimen will help to enhance the specificity of the treatment for a given effect. This is a way to optimize the efficacy/toxicity ratio for anticancer agents, for example. Finally, such basic explorations are useful as an orientation for appropriate model selection in further PKPD analyses.

***Pharmacokinetic–Pharmacodynamic Models*** The simplest representation of the dependency between circulating drug concentrations and effect levels postulates merely an instantaneous correlation between both variables. This view characterizes the *immediate PKPD models*, where the development and the resolution of the effect are directly driven by the shape of the PK curve. In the situations satisfying the assumption of immediacy, a simple correlation analysis between simultaneous effect and concentration measures is sufficient to describe the effect profile. Only exceptionally is a *linear* correlation appropriate for an accurate description. In most conditions, the level of response tends rather to increase proportionally to the logarithm of the circulating concentration, making relative concentration changes more accurately predict the effect level than absolute concentration values. *Log-linear* relationships are thus sometimes used in immediate PD models to relate an effect to a circulating drug concentration. This model is, however, not entirely satisfactory, as it predicts negative effects at very small concentration values. Moreover, in many pharmacological situations, the effect is observed to saturate at the highest values of concentration. Thus, a hyperbolic equation derived from the receptor theory is most often used to describe clinical concentration–effect relationships. This is the $E_{max}$ *model* relating an effect $E$ to the drug concentration $C$:

$$E = E_0 + \frac{E_{\max} C}{\mathrm{EC}_{50} + C}$$

where $E_0$ is the baseline response level predicted for a zero concentration, $E_{\max}$ the maximal increase in response predicted at infinite concentration, and $\mathrm{EC}_{50}$ the concentration associated with one-half of the asymptotic maximal effect. This expression corresponds to a sigmoid curve when represented on a log $X$ axis. Inhibition of a response variable may require subtraction of the right-sided term from the baseline $E_0$. The value of $E_{\max}$ is traditionally considered to reflect the *efficacy* of the drug, while $\mathrm{EC}_{50}$ expresses its *potency*. These parameters can serve to compare various drugs among a pharmacological class. For a given drug, different $\mathrm{EC}_{50}$ values characterize various types of responses (e.g., therapeutic and adverse effects), which may help to assess the *therapeutic margin* and define a suitable therapeutic range. Notice that the parameters of the clinical $E_{\max}$ model have little to do with their counterparts found during in vitro pharmacological experiments. A supplemental coefficient of steepness $\gamma$ may be added in the concentration–response curve, giving *Hill's* or *sigmoid* $E_{\max}$ *model* (initially developed to describe the affinity of oxygen for hemoglobin):

$$E = E_0 + \frac{E_{\max} C^{\gamma}}{\mathrm{EC}_{50}^{\gamma} + C^{\gamma}}$$

Further PD models have been developed to account for less usual shapes of concentration–effect curves.

For binary response variables, a convenient description of the PK dependency uses the *logistic regression* model, which associates to each concentration level $C$ a probability $\pi(E)$ of successful achievement of the effect:

$$\pi(E) = \frac{e^{\alpha+\beta C}}{1+e^{\alpha+\beta C}}$$

where the logit of $\alpha$ gives the baseline probability in the absence of drug, and $\beta$ indicates the log of the odds associated with a unit increase in concentration. A less sophisticated model sometimes used for binary responses is the step function curve, characterized by a single threshold concentration value, above which the effect is assumed to be achieved (e.g., the minimum inhibitory concentration for antibiotics).

Not infrequently, the expansion of the PD curve does not parallel the PK curve, and the effect appears with retardation, while still depending on the circulating concentration profile. In such situations, a concentration–effect scatterplot will not reveal a steady correlation but rather separate lines over the periods of concentration ascent and descent, respectively. Such a looping appearance of the concentration–effect curve calls for the application of *hysteresis PKPD models*. Several approaches have been devised to handle this phenomenon in modeling. The *link model* derives from the compartment theory and assumes that the circulating drug distributes in an effect compartment, where it induces the effect through an immediate relation as shown above [42]. Thus, the drug concentration at effect site is constructed as a latent variable. The retardation between plasma concentration and effect curves is explained by the time constant that determines drug accumulation and persistence in this peripheral effect compartment (this is in fact the rate constant of elimination from this compartment, $k_{e0}$). This model is in accordance with physiological concepts, with receptors sometimes actually located in peripheral tissues, or receptors themselves found to progressively accumulate and slowly release ligands. Another class of models deserving mention are the *indirect models* [43]. These models, based on a compartmental approach, consider the effect marker as an endogenous substance, which is produced and eliminated with its proper kinetics. The drug administered is assumed to act either on the production or on the elimination rate of the endogenous marker, either to stimulate or to inhibit the process, through one of the direct concentration–effect relationships introduced above. While the drug concentration is supposed to influence immediately the rate of change of the endogenous marker, the resulting absolute marker level integrates this response over time with a definite inertia, thus accounting for the retardation between concentration and response curves. This model finds physiological justifications for a series of drugs known to act through second messengers, sometimes amenable to measurement (e.g., oral anticoagulants, which inhibit the synthesis of coagulation factors having specific turnover rates). It can also be applied to other situations where the turnover of a latent endogenous marker is nothing but a virtual representation.

A number of sophisticated PKPD models have been developed to describe more complex dose–concentration–effect relationships [44]. The production of a virtual antagonistic metabolite can be invoked to describe a phenomenon of tolerance, which is characterized by an antihysteresis loop. Feedback stimulation or inhibition loops can be added, either on PK or on PD processes. Further intermediary effect compartments can be designed. A complex interplay between the drug and living

pharmacological targets (e.g., viruses, cancer cells) can be modeled. Nonlinearity in PKPD models explain the schedule dependency shown by some drugs, meaning that the effect of a given total amount of drugs strongly depends on the way it is administered over time. Even paradoxical responses can be accounted for, with different administration schemes resulting in opposite effects (e.g., gonadrotrophin-releasing-hormone analogs, which stimulate ovulation and hormone production when given in pulses but achieve a so-called chemical castration when released slowly from a depot formulation [45]).

### 3.3 Design and Analysis of PKPD Trials

As outlined above, PKPD studies are much more versatile than traditional PK trials due to the variety of measurable response markers, each of them having its own clinical relevance, pharmacometric properties, and endogenous profile. Hence PKPD study protocols have to be somewhat creative, but care must be given to have them produce good science anyway. The following recommendations deserve consideration for the design, execution, and analysis of PKPD trials:

- Precise study objectives should always be defined in advance, and one or a few of them should be specified as primary. The study design should be decided in accordance with the study aims, with consideration for preliminary knowledge of the characteristics of the drug tested. PKPD trials may involve single or multiple dosages, parallel or cross-over designs, and sample-rich or sparse-sample measurement schedules.
- The inclusion of a *placebo* arm is often useful in PKPD trials to assess the evolution of response markers in the absence of pharmacological agents. For trials testing an experimental treatment, the inclusion of a *verum* arm, that is, a reference treatment of known impact on the response markers, may be useful as positive control.
- A restricted choice of effect variables should be done, based on a selection of the most appropriate response markers. When a large number of response markers are available, the selection should be oriented to minimize the redundancy and to mainly retain variables that complement each other. For example, a rapidly reactive variable will be chosen along with a slowly integrative one. Marker measurement methods should undergo formal validation. A preliminary pilot study may be useful to check the feasibility of the methods before undertaking a large-scale trial.
- Whenever possible, basic robust descriptors should be calculated separately on both PK and PD results data, before initiating any formal PKPD modeling. In rich-sample trials, the questions corresponding to main study objectives are frequently addressed only with reference to such simple descriptors, while the fitting of a formal PKPD model is left among the secondary objectives. Individual and average response curves should be graphed for all study variables.
- In sparse-sample trials, population approaches are perfectly suited for data analysis and may work even better in this context than in purely PK trials. Most

computer facilities developed for this purpose allow the inclusion of both concentration and effect data, even though it makes the problem technically more demanding. *Population PKPD analysis* represents probably the best available tool for a rational exploitation of the results of human pharmacology trials [46]. Sometimes, only the PK data deserve a population analysis, while the PD data and PKPD associations are merely analyzed by simple statistical approaches applied to basic descriptors outlined above (e.g., AUEC), which can be correlated to the individual PK parameters produced by the population analysis. Another alternative is to first fit the data from the PK part, then fix all PK parameters, and then rerun the population analysis to estimate the PD parameters.

## 4 CONCLUSION

Clinical pharmacologists have the task to gather, produce and summarize relevant knowledge on the pharmacological profile of therapeutic drugs in order to optimize their medical utilization. Rational dosing regimens must be established, and consistent recommendations for dosage individualization must be issued. Many choices are to be discussed during early clinical drug development: selection of the best candidate among a family of compounds, route of administration, pharmaceutical formulation, dose level and dosing frequency, duration of treatment, dosage adaptation strategies (according to body weight, age, organ failure, coexistent diseases, interacting drugs or pharmacogenetic tests), requirement for some form of monitoring based on either biomarkers or drug concentrations, and the like. The answers to such questions should ideally be found before a candidate drug enters into large phase III confirmatory trials, where it should be ready for use as in clinical practice [4]. Considering all those questions, and the number of effect variables susceptible to be examined, one may envision vast drug development projects with hundreds of studies to explore any imaginable combination of choices. However, this is neither feasible nor really suitable. The aim of early clinical drug development is to find the most efficient pathway reaching the optimal combination of choices, across the tortuous hills and valleys of a multidimensional space of possible decisions. This requires to arrange a series of efficient, small-sized clinical experiments, each of them being designed according to the results of the previous ones. Exploration–confirmation cycles are continued until sufficient information has been obtained to devise rational dosing and monitoring recommendations. PKPD modeling and simulation represent highly valuable tools to improve this process [47, 48]. This effort does not end up at authorization to market. It is to be continued not only by drug manufacturers but also by academic researchers taking up the responsibility to raise critical questions and to challenge the therapeutic tools, so that the patients and the public obtain the best service from the drugs they buy. This is how clinical drug development definitely resembles the art of "response surface analysis," to quote Sheiner [4] referring to an elegant theoretical framework developed in engineering [49]. This amazing vision must not be restricted to the sole domain of clinical pharmacology applied to drug development but will probably prove relevant in many conditions where human trials are planned, performed, and analyzed.

## APPENDIX NOTE ON GEOMETRIC AVERAGES AND COEFFICIENTS OF VARIATION

Both drug concentration data and pharmacokinetic parameters usually follow a skewed distribution among individuals, which is best described according to a *log-normal* rather than a Gaussian statistical model. This is because most pharmacokinetic processes are governed by *first-order rates* (i.e., transfer or reaction rates are proportional to the amount of substrate available), giving importance to *relative* rather than absolute changes.

In this situation, it is not appropriate to summarize results using ordinary mean and standard deviation (SD), as the thin and tall upper tail of the distribution will drag the mean well above the bulk of typical values and increase the magnitude of SD up to exceeding the size of the mean. A more suitable summary descriptor is the *geometric mean*, which is defined as the *n*th root of the product of *n* individual values. Practically, it is obtained by transforming the individual values into logarithms, computing the usual (arithmetic) mean of log transforms, and calculating back the antilog of this mean. For example, the three values 80, 100, and 125 correspond to a geometric mean of 100.

Computing the SD of those log-transformed data, and then taking its antilog, provides an estimate of the relative spread of log-normally distributed values: It can be interpreted as the typical factor either multiplying or dividing the geometric mean, which encompasses about two thirds of the data. For our three values 80, 100, and 125, this "geometric SD" represents a factor of 1.25. (Remember that the classical SD around an arithmetic mean is the typical difference either added to or subtracted from the mean, which encompasses about two thirds of a set of data that follow a Gaussian distribution). It may be more convenient to express this relative spread as a *percentage of variability*, by simply subtracting 1 from the geometric SD and multiplying by 100. In our example, this leads to a value of 25%. This value is higher than the "arithmetic" coefficient of variation, defined as the (arithmetic) SD divided by the mean (which would give 22.6/101.7 = 22% in our example). Nevertheless, it is sometimes called a *geometric coefficient of variation* (a more proper denomination would be the "coefficient of variation of the multiplicative model"). A more exact estimate for this coefficient can be calculated as the square root of the antilog of the variance of log-transformed data minus one (giving $\sqrt{0.051} = 22.6\%$ in our example) [50].

For graphic purposes, it is recommended to represent geometric means associated with error bars showing their multiplication and division by the corresponding geometric SD. This is the easiest way to compute error bars that appear symmetric along a log *Y* axis (Fig. 3). Likewise, when parametric statistical tests are to be used for treatment or group comparisons on log-normally distributed data, a preliminary log transformation of the data is clearly indicated. Subsequently, it is possible to translate a difference between average log values back into a ratio of relative change. Study factors are thus assumed to impact multiplicatively rather than additively on the variables of interest, which they do most of the time (e.g., an enzyme inhibitor will be said to increase the concentrations of an interacting drug by a given percentage, rather than by a fixed amount in milligrams/liter). While this log-transform approach is standard practice in clinical pharmacokinetics, it would probably be suitable in many other areas of data analysis for clinical research.

## REFERENCES


1. Miller, F. G. (2003), Ethical issues in research with healthy volunteers: Risk-benefit assessment, *Clin. Pharmacol. Ther.*, 74, 513–515.
2. Hornblum, A. M. (1997), They were cheap and available: Prisoners as research subjects in twentieth century America, *BMJ*, 315, 1437–1441.
3. Lenzer, J. (2004), Scandals have eroded U.S. public's confidence in drug industry, *BMJ*, 329, 247.
4. Sheiner, L. B. (1997), Learning versus confirming in clinical drug development, *Clin. Pharmacol. Ther.*, 61, 275–291.
5. Cross, J., Lee, H., Westelinck, A., et al. (2002), Postmarketing drug dosage changes of 499 FDA-approved new molecular entities, 1980–1999, *Pharmacoepidemiol. Drug Safety*, 11, 439–446.
6. Heerdink, E. R., Urquhart, J., and Leufkens, H. G. (2002), Changes in prescribed drug doses after market introduction, *Pharmacoepidemiol. Drug Safety*, 11, 447–453.
7. Mouton, J. V., and Vinks, A. A. T. M. M. (1996), Is continuous infusion of β-lactam antibiotics worthwhile?—efficacy and pharmacokinetic considerations, *J. Antimicrob. Chemother.*, 38, 5–15.
8. Krayenbuhl, J. C., Vozeh, S., Kondo-Oestreicher, M., et al. (1999), Drug-drug interactions of new active substances: Mibefradil example. Euro. *J. Clin. Pharmacol.*, 55, 559–565.
9. Williams, D., and Feely, J. (2002), Pharmacokinetic-pharmacodynamic drug interactions with HMG-CoA reductase inhibitors, *Clin. Pharmacokinet.*, 41, 343–370.
10. Shah, R. R. (2005), Pharmacogenetics in drug regulation: Promise, potential and pitfalls, *Philo. Trans. Roy. Soc. London*, 360, 1617–1638.
11. Rendon, A., Nunez, M., Jimenez-Nacher, I., et al. (2005), Clinical benefit of interventions driven by therapeutic drug monitoring, *HIV Med.*, 6, 360–365.
12. Van Gelder, T., Meur, Y. L., Shaw, L. M., et al. (2006), Therapeutic drug monitoring of mycophenolate mofetil in transplantation, *Ther. Drug Monitor.*, 28, 145–154.
13. Brown, K., Tompkins, E. M., and White, I. N. (2006), Applications of accelerator mass spectrometry for pharmacological and toxicological research, *Mass Spectrom. Rev.*, 25, 127–145.
14. Garner, R. C., and Lappin, G. (2006), The phase 0 microdosing concept, *Br. J. Clin. Pharmacol.*, 61, 367–370.
15. DiMasi, J. A., Hansen, R. W., and Grabowski, H. G. (2003), The price of innovation: New estimates of drug development costs, *J. Health Economy*, 22, 151–185.
16. Ortiz de Montellano, P. R. (2005), *Cytochrome P450: Structure, Mechanism, and Biochemistry*, Kluwer Academic/Plenum, New York.
17. Ho, R. H., and Kim, R. B. (2005), Transporters and drug therapy: Implications for drug disposition and disease, *Clin. Pharmacol. Ther.*, 78, 260–277.
18. Atkinson, A. J. Jr., Daniels, C. E., Dedrick, R. L., et al. (2001), *Principles of Clinical Pharmacology*, Academic, San Diego.
19. Tozer, T. N., and Rowland, M. (2006), *Introduction to Pharmacokinetics and Pharmacodynamics: The Quantitative Basis of Drug Therapy*, Lippincott Williams & Wilkins, Philadelphia.
20. Colburn, W. A., and Lee, J. W. (2003), Biomarkers, validation and pharmacokinetic-pharmacodynamic modelling, *Clin. Pharmacokinet.*, 42, 997–1022.

21. Gabrielsson, J., and Weiner, D. (2001), *Pharmacokinetic and Pharmacodynamic Data Analysis: Concepts and Applications*, 3rd ed., Taylor & Francis, London.
22. FDA Guidance for Industry (2003), *Exposure-Response Relationships—Study Design, Data Analysis, and Regulatory Applications*; available at: www.fda.gov/cber/guidelines.htm.
23. FDA Guidance for Industry (1999), *Population Pharmacokinetics*; available at: www.fda.gov/cber/guidelines.htm.
24. Williams, P. J., and Ette, E. I. (2000), The role of population pharmacokinetics in drug development in light of the Food and Drug Administration's "Guidance for Industry: Population pharmacokinetics," *Clin. Pharmacokinet.*, 39, 385–395.
25. Minto, C., and Schnider, T. (1998), Expanding clinical applications of population pharmacodynamic modelling, *Br. J. Clin. Pharmacol.*, 46, 321–333.
26. Vozeh, S., Steimer, J. L., Rowland, M., et al. (1997), The use of population pharmacokinetics in drug development, *Clin. Pharmacokinet.*, 30, 81–93.
27. Shah, V. P., Midha, K. K., Dighe, S., et al. (1991), Analytical methods validation: Bioavailability, bioequivalence and pharmacokinetic studies. Conference report. *Euro. J. Drug Metabolism Pharmacokinet.*, 16, 249–255.
28. International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (1994, rev. 1996, 2005), *Validation of Analytical Procedures*; available at: www.ich.org.
29. Directive 75/318/EEC as amended (1987), *Pharmacokinetic Studies in Man*; available at: ec.europa.eu/enterprise/pharmaceuticals/eudralex/vol-3/pdfs-en/3cc3aen.pdf.
30. Schwartz, J. B. (2003), The influence of sex on pharmacokinetics, *Clin. Pharmacokinet.*, 42, 107–121.
31. Loebstein, R., Lalkin, A., and Koren, G. (1997), Pharmacokinetic changes during pregnancy and their clinical relevance, *Clin. Pharmacokinet.*, 33, 328–343.
32. Mizutani, T. (2003), PM frequencies of major CYPs in Asians and Caucasians, *Drug Metabolism Rev.*, 35, 99–106.
33. Buclin, T., Perrottet, N., and Biollaz, J. (2005), The importance of assessing the dose actually administered in pharmacokinetic trials, *Clin. Pharmacol. Ther.*, 77, 235–240.
34. Gillespie, W. R. (1991), Noncompartmental versus compartmental modeling in clinical pharmacokinetics, *Clin. Pharmacokinet.*, 20, 253–262.
35. Boomer, D. (2006), Pharmacokinetic and Pharmacodynamic Resources; accessible at: http://www.boomer.org/pkin/soft.html.
36. FDA Guidance for Industry (2001), *Statistical Approaches to Establishing Bioequivalence*; available at: www.fda.gov/cber/guidelines.htm.
37. EU Directive 65/65/EEC and 75/318/EEC as amended (1991), *Investigation of Bioavailability and Bioequivalence*; available at: ec.europa.eu/enterprise/pharmaceuticals/eudralex/vol-3/pdfs-en/3cc15aen.pdf.
38. Senn, S. (2001), Statistical issues in bioequivalence, *Stat. Med.*, 20, 2785–2799.
39. Williams, P. J., and Ette, E. I. (2000), The role of population pharmacokinetics in drug development in light of the Food and Drug Administration's "Guidance for Industry: Population Pharmacokinetics," *Clin. Pharmacokinet.*, 39, 385–395.
40. Ette, E. I., and Williams, P. J. (2004), Population Pharmacokinetics. I: Background, Concepts, and Models; II: Estimation Methods; III: Design, Analysis, and Application of Population Pharmacokinetic Studies, *Ann. Pharmacother.*, 38, 1702–1706; 1907–1915; 2136–2144.

41. Craig, W. A. (2002), Pharmacodynamics of antimicrobials: General concepts and applications, in Nightingale, C. H., Murakawa, T., and Ambrose, P. G. Eds., *Antimicrobial Pharmacodynamics in Theory and Clinical Practice*, Marcel Dekker, New York, pp. 1–22.
42. Sheiner, L. B., Stanski, D. R., Vozeh, S., et al. (1979), Simultaneous modeling of pharmacokinetics and pharmacodynamics: Application to d-tubocurarine, *Clin. Pharmacol. Ther.*, 25, 358–371.
43. Jusko, W. J., and Ko, H. C. (1994), Physiologic indirect response models characterize diverse types of pharmacodynamic effects, *Clin. Pharmacol. Ther.*, 56, 406–419.
44. Csajka, C., and Verotta, D. (2006), Pharmacokinetic-pharmacodynamic modelling: History and perspectives, *J. Pharmacokinet. Pharmacodyn.*, 33, 227–279.
45. Stanislaus, D., Pinter, J. H., Janovick, J. A., et al. (1998), Mechanisms mediating multiple physiological responses to gonadotropin-releasing hormone, *Mol. Cell. Endocrinol.*, 144, 1–10.
46. Aarons, L., Karlsson, M. O., Mentre, F., et al. (2001), Role of modelling and simulation in Phase I drug development, *Euro. J. Pharma. Sci.*, 13, 115–122.
47. Meibohm, B., and Derendorf, H. (2002), Pharmacokinetic/pharmacodynamic studies in drug product development, *J. Pharma. Sci.*, 91, 18–31.
48. Chien, J. Y., Friedrich, S., Heathman, M. A., et al. (2005), Pharmacokinetics/pharmacodynamics and the stages of drug development: Role of modeling and simulation, *AAPS J.*, 7, 544–559.
49. Myers, R. H., and Montgomery, D. C. (2002), *Response Surface Methodology: Process and Product Optimization Using Designed Experiments*, Wiley, Hoboken, NJ.
50. Limpert, E., Stahel, W. A., and Abbt, M. (2001), Log-normal distributions across the sciences: Keys and clues. *Bioscience*, 51, 341–352.

# *Modeling and Simulation in Clinical Drug Development*

**Jerry Nedelman,[1] Frank Bretz,[2] Roland Fisch,[3] Anna Georgieva,[1] Chyi-Hung Hsu,[2] Joseph Kahn,[1] Ryosei Kawai,[4] Phil Lowe,[3] Jeff Maca,[2] José Pinheiro,[2] Anthony Rossini,[3] Heinz Schmidli,[5] Jean-Louis Steimer,[3] and Jing Yu[4]**

*[1]Modeling and Simulation and [2]Clinical Information Sciences, Novartis Pharmaceuticals Corp., East Hanover, New Jersey*
*[3]Modeling and Simulation and [5]Clinical Information Sciences, Novartis Pharma AG, Basel, Switzerland*
*[4]Modeling and Simulation, Novartis Institutes for BioMedical Research, Inc., Cambridge, Massachusetts*

**Contents**

1 Introduction
- 1.1 What Is a Model?
- 1.2 What Is Simulation?
- 1.3 Model Building
- 1.4 Bayesian Methods
- 1.5 What Lies Ahead

2 Types of Modeling and Simulation
- 2.1 Regression Models
- 2.2 Longitudinal Models
- 2.3 Exploratory Modeling and Data Mining
- 2.4 Dose–Exposure–Response Modeling

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

2.5 Pathways Modeling
2.6 Physiological Modeling
2.7 Disease Progression Modeling
2.8 Clinical Trial Simulation
2.9 Decision Analysis
3 M&S-Based Clinical Trials
3.1 First-in-Man Studies
3.2 Oncology Phase I Studies
3.3 Proof-of-Concept Studies
3.4 MCP-Mod: Unified Strategy for Dose-Finding Studies
3.5 Adaptive Trials
4 Discussion
References

## 1 INTRODUCTION

This chapter describes modeling and simulation (M&S) applied to clinical drug development within a pharmaceutical company. In that context, M&S serves the two goals of drug development: learning and confirming [1].

From the learning perspective, M&S is a tool to learn how the drug works. M&S helps us design and interpret clinical trials, the learning experiences of clinical drug development. M&S is also a framework for representing what we learn in the precise language of mathematics.

For confirming, M&S provides a rigorous context in which to make statistical judgments about relationships, such as whether increases in dose are significantly associated with increases in response.

As you can see from the previous paragraphs, it is conventional to use the phrase "modeling and simulation" as a singular noun. But modeling and simulation are different things. Let us begin with a separate consideration of each. We will then discuss some general notions about how models are built and briefly introduce Bayesian methodology, which plays a large role in M&S. Then we will pause to outline how the remainder of the chapter describes M&S in clinical drug development.

### 1.1 What Is a Model?

A model is a way of predicting outputs from inputs.

For example, consider a model to predict the steady-state concentration $C_{ss}$ obtained by infusing a drug at rate $R$ into a subject who has clearance CL:

$$C_{ss} = \frac{R}{\mathrm{CL}} \tag{1}$$

The output is $C_{ss}$ and the inputs are $R$ and CL. Note that different inputs may play different roles. The infusion rate is at the discretion of the experimenter (physician, pharmacologist, . . .). The clearance is a characteristic of the subject and drug. Both might be called variables, or we might refer to the clearance, which may be unknown, as a parameter.

Now consider the systolic blood pressures at baseline, the start of the trial, among patients enrolled in a clinical trial for an antihypertensive drug. The protocol will specify some severity of hypertension for included patients, but there will still be variability among patients at baseline. We use the language of probability to create a model for such variability. Suppose the average baseline blood pressure among all eligible patients is $\mu$ and the variance is $\sigma^2$. It is often the case that a collection of blood pressures follows a bell-shaped curve that we might characterize as a normal probability distribution. We then say the output—baseline blood pressure $B$—is predicted from the inputs—$\mu$ and $\sigma^2$—according to the Normal probability model:

$$B \sim N(\mu, \sigma^2) \tag{2}$$

In (2), $B$ is called a random variable; $\mu$ and $\sigma^2$ are both considered parameters. Note that to represent a random variable as an output from its distributional model we use ~ instead of =.

Imagine observing steady-state concentrations by taking blood samples and assaying them. Even if model (1) is reasonable, and even if we know CL as well as $R$, our observed concentration will probably not exactly equal $C_{ss}$ because of measurement error and because of physiological fluctuations that render the model incomplete. We account for this discrepancy by introducing a new variable, usually denoted $\varepsilon$, called the residual error. It is modeled as a random variable, analogous to the baseline blood pressures (2), and we often use the same Normal probability model. We usually assume a mean of zero for residual errors, so the only input is the variance, $\sigma^2$. We then expand (1) to

$$C_{ss} = \frac{R}{\mathrm{CL}} + \varepsilon \tag{3}$$

$$\varepsilon \sim N(0, \sigma^2) \tag{4}$$

Note that in (3) $\varepsilon$ plays the role of an input, but in (4) it plays the role of an output. Chaining models together in this way gives modeling great power while maintaining conceptual simplicity.

Suppose we administer a single intravenous bolus dose $D$ of a drug at time 0, and suppose the systemic circulation behaves as if it is a single well-mixed compartment. Then the systemic concentration $C$ of drug may be predicted as a function of the inputs dose ($D$), time post dose ($t$), clearance (CL), and compartmental volume ($V$) as

$$C(t) = D \exp\left(\frac{-\mathrm{CL}\, t}{V}\right) \tag{5}$$

Models with time as one of the inputs play a very important role in clinical trials (and throughout science), so it is common to highlight the input time by using it as an arugment in a representation of the output as a function, as $C(t)$ in (5).

The one-comparment model (5) is somewhat special because we can write the dependence of concentration on time algebraically. Frequently, models with time as input can only be expressed by describing how the rate of change of the output

depends on time and other inputs, that is, as a differential equation. A differential equation corresponding to (5) is

$$\frac{dC}{dt} = -\left(\frac{\mathrm{CL}}{V}\right)C \qquad C(0) = D \tag{6}$$

The algebraic representation (5) is the solution to the differential equation (6). But often no algebraic solution can be found, and we must solve the differential equation numerically, which takes us to simulation.

### 1.2 What Is Simulation?

Simulation is the act of computing values of the output for given values of the input. In (1), if we plug in values for $R$ and CL and compute the resulting value of $C_{ss}$, we have accomplished a very simple simulation.

Usually, however, simulation refers to repeating such calculations many times for different values of the inputs to see how changes in the inputs translate into changes in the outputs. The three most common situations are:

1. One of the inputs is time, and the computations are done to see how the output evolves as time progresses.
2. One or more of the inputs is a random variable, and the computations are done to see how the output varies due to the random process(es) represented by the random variable(s).
3. Both 1 and 2 together.

### 1.3 Model Building

Models seldom spring fully formed from the heads of modelers, at least if "fully formed" includes specification of the values of the model's parameters. Usually, the parameters must be estimated by fitting the model to data. Often the structure of the model (e.g., algebraic functions, differential equations, or probability distribution) and the inputs that best predict the outputs are determined partly if not wholly by fitting different models to the data and choosing the one that seems best, a process called model building.

Fitting the model to data means selecting parameter values of the model to minimize some measure of discrepancy or maximize some measure of similarity between observed values of the output variable (the data) and predictions of the output variable from the model. The sum of squared differences between data and predictions is a measure of discrepancy; minimizing it yields a least-squares fit. A common similarity measure is the likelihood, which, as its name implies, quantifies how probable the observed data are according to the model; maximizing it yields a maximum-likelihood fit.

The more complex the model the better the fit in terms of least squares or likelihood. To decide when to stop building, called model selection, criteria may be used that penalize models for being too complex, both for the aesthetic appeal of parsimony and because theory shows that these criteria effectively home in on the best model. Two such criteria are the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) [2].

## 1.4 Bayesian Methods

In the Bayesian approach to modeling, parameters are considered random variables not unknown constants. A parameter's probablity distribution represents our uncertainty regarding its value. To learn is to reduce that uncertainty, corresponding to a reduction in the variance of the parameter's probability distribution. At the start of a new learning step, the probability distribution is called the prior. When we acquire new data and thereby learn something about the parameter, the updated probability distribution is called the posterior. This process of updating our knowledge, called Bayesian inference, is accomplished via Bayes rule, a mathematical formula. Berry [3] provides an introduction to Bayesian methods and their application in clinical trials.

## 1.5 What Lies Ahead

The remainder of this chapter will describe types of M&S used in clinical drug development and types of clinical trials that strongly depend on M&S. Each section will begin with a statement of purpose for the type of M&S covered or for the use of M&S in the clinical trial.

The sections will frequently cross reference one another. Curly braces { } will denote section number cross references, whereas square brackets [ ] will denote bibliographical citations and parentheses ( ) will denote equation numbers.

Although the chapter's parts must be sequentially ordered in a print medium such as this, there is no perfect sequencing of the material. Figure 1 displays one possible configuration of the dependencies among the sections.

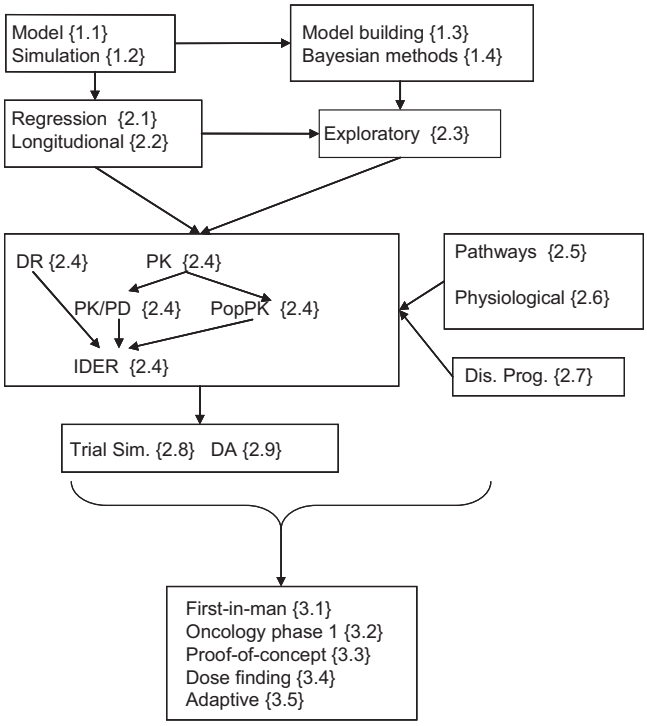


***FIGURE 1*** *Dependencies among chapter sections.*

The heart of the chapter is {2.4}, which covers dose–exposure–response (DER) modeling, the fundamental clinical characterization of how a drug works and how M&S helps determine what dose to give to which patients. {1.1}–{2.3} provide the fundamentals of M&S needed to understand how DER models are built and used. {2.5}–{2.6} introduce exciting new areas of biological modeling that support DER modeling with mechanistic underpinnings. {2.7} describes methods for characterizing higher level disease progression. Within the process of drug development, an important application of models is to support decision making; {2.8}–{2.9} cover that. Finally {3} showcases five types of clinical trials whose design and interpretation benefit from M&S; they are presented roughly in the order they might appear in a drug development program.

## 2 TYPES OF MODELING AND SIMULATION

### 2.1 Regression Models

**Purpose:** *To infer how one variable may be predicted empirically from one or more others using independent observations of those variables in a sample of subjects.*

One variable, such as blood pressure (BP) at the end of a hypertensive trial, is the output, called the response; and one or more others, such as baseline blood pressure, treatment, dose, age, and sex, are inputs, called predictors. The goal of relating the response to the predictors may be learning—for example, to devise a rule to predict future observations of the response—or confirming—for example, to test if one or more of the predictors, such as treatment, is statistically significant. The model is fitted to data comprising one record per subject, each record containing observations of the response and the predictors, with records from different subjects considered independent.

Instead of being derived from any scientific theory, the form of a regression model is selected as a compromise between mathematical convenience and the empirical representation of the data's shape. Often a linear model (7) suffices because relationships tend to be approximately linear over limited ranges of the data.

Simple linear regression has one continuous response, for example, BP, that depends on a continuous predictor, for example, dose, and residual error:

$$\mathrm{BP}_i = \beta_0 + \beta_1 \,\mathrm{dose}_i + \varepsilon_i \tag{7}$$

$$\varepsilon_i \sim N\left(0, \sigma^2\right) \tag{8}$$

where $\beta_0$, $\beta_1$, and $\sigma^2$ are parameters that are estimated by fitting the model to data and where the subscript $i$ indexes the data, that is, here the subjects. Presumably, if the drug were an antihypertensive, $\beta_1 < 0$.

The adjective "linear" refers to the linearity of (7) with respect to the parameters $\beta_0$ and $\beta_1$, not dose. If in (7) dose were replaced by, for example, the logarithm of dose, it would still be a linear model even though a plot of $\beta_0 + \beta_1 \log(\mathrm{dose})$ versus dose would be curved.

If there are several predictors, such a model becomes a multiple regression model.

Nonlinear regression is when some parameters appear nonlinearly, such as

$$\mathrm{BP}_i = \beta_0 + \frac{\beta_1 \,\mathrm{dose}_i}{\beta_2 + \mathrm{dose}_i} + \varepsilon_i \tag{9}$$

Suppose we create a new response variable, $Y$, by considering only whether BP dropped by 3 mmHg or not, for which we assign $Y = 1$ or 0, respectively. Logistic regression is where the logit transformation of the probability that $Y = 1$, say $p$, is modeled linearly as

$$\mathrm{logit}(p) = \log\left[\frac{p}{(1-p)}\right] = \beta_0 + \beta_1 \,\mathrm{dose} \tag{10}$$

Logistic regression is an example of a broader class of models called generalized linear models [4], where some function of the mean of the response is related linearly to the predictors.

A clinical outcome may be the time to some event, such as death. Regression models for such a response time are a major component of survival analysis. The chance of the event occurring in the next instant of time, given that it has not happened yet, is called the hazard. In the Cox proportional hazards model, the logarithm of the hazard is related linearly to the predictors.

Many statistics textbooks cover regression modeling. See [5] for a nontechnical introduction or [6] for a technical one.

### 2.2 Longitudinal Models

**Purpose:** *To infer how one variable may be predicted empirically from one or more others using observations measured at more than one time per subject in a sample of subjects.*

Whereas regression models {2.1} are appropriate when subjects independently contribute one observation each, longitudinal models describe multiple observations collected over time for each subject. The several observations of the same individual tend to be correlated, which must be accounted for by the model. This additional complication buys, for confirming purposes, additional precision and power, and, for learning purposes, additional insight into how the response evolves as conditions change.

Regression models generalize to longitudinal models in two ways: marginal models and mixed-effect models.

The focus of a marginal model is to predict the population's average response, generally for confirming purposes. As in generalized linear models {2.1}, some transformation of the average response is assumed to admit a linear structure in the predictors. The correlation structure is handled somewhat empirically using methods that are known to be robust with sufficiently large samples; see [7].

Marginal models have two main limitations. First, they require that all subjects have the same number of observations under the same design. Second, they generally are not informative about individual responses.

Mixed-effect models (MEMs) overcome these limitations. An MEM is constructed hierarchically, by chaining models together as described in {1.1}, starting from a model for the response of an individual. The parameters of the individuals' separate models are themselves outputs of regression-type models. This accounts for the necessary correlation structure of the data and allows the models to be fitted to data even where subjects have different numbers of observations at different times.

Let us consider an MEM generalizing (7–8) to describe a study where each subject receives successively higher doses at each of which BP is measured. Let $\mathrm{BP}_{ij}$ and $\mathrm{dose}_{ij}$ be the $j$th observed BP and dose for the $i$th subject. The MEM is

$$\mathrm{BP}_{ij} = b_{0i} + b_{1i}\,\mathrm{dose}_{ij} + \varepsilon_{ij} \tag{11}$$

$$b_{0i} = \beta_0 + \eta_{0i} \qquad b_{1i} = \beta_1 + \eta_{1i} \tag{12}$$

$$\eta_{0i} \sim N(0, \omega_0^2) \qquad \eta_{1i} \sim N(0, \omega_1^2) \qquad \mathrm{corr}(\eta_{0i}, \eta_{1i}) = \rho \tag{13}$$

$$\varepsilon_{ij} \sim N(0, \sigma^2) \tag{14}$$

Each subject has his or her own intercept and slope $b_{0i}$ and $b_{1i}$. These inputs to model (11) are outputs of the models (12). The random variables $\eta_{0i}$ and $\eta_{1i}$ are called random effects. They characterize the variation among subjects of the intercepts and slopes of the BP-vs.-dose relationships, and their presence accounts for the correlations among the several $\mathrm{BP}_{ij}$ values for the $i$th subject.

Suppose we believe that some of the variation in the dependence of BP on dose is due to age. We need only make age an input to the model for $b_1$ in (12):

$$b_{1i} = \beta_1 + \beta_2\,\mathrm{age}_i + \eta_{1i} \tag{15}$$

If any of the parameters in an MEM besides the variances appears nonlinearly, then the model is called a nonlinear MEM or NLMEM, which will figure prominently in {2.4}.

Longitudinal models for clinical trials are reviewed in [8]. See also [9].

## 2.3 Exploratory Modeling and Data Mining

**Purpose:** *In large data sets, to find patterns that can be used as hypotheses to guide further experimentation or as retrospective explanations of surprising signals.*

Exploratory modeling in clinical drug development has two main applications. The first is part of the guided knowledge generation that is common to all science, when some general goals of discovery are pursued in the hopes of generating hypotheses that might be confirmed with further experimentation. Examples are mining genomic data to find potential biomarkers, or mining clinical data to find possible new indications for marketed drugs.

An alternative setting occurs when a clinical trial produces some surprising signal, such as a worrisome toxicity. Sometimes called fire-fighting, exploratory modeling then seeks to propose a plausible story for the observed results. Such a story would ultimately require confirmation, and thus this second application is really one of

hypothesis generation like the first; but if strong enough, such modeling results might be proposed as tentatively confirmatory or at least satisfactorily descriptive for, say, a regulatory submission.

The steps of exploratory modeling are (a) defining the problem, (b) exploring the data with graphical, tabular, and numerical summaries, (c) building models, and (d) validating models.

In the first two steps, the scope of potential models is clarified by learning the structure and content of the data and identifying any important substructures such as grouping in study centers, or local laboratories with different reference ranges, or time profiles of observations.

The nomenclature of data mining, derived largely from computer science, often refers to modeling as the application of various algorithms. Commonly used algorighms are listed below. To link with our view of modeling {1.1}, for each algorithm we identify the kind of outputs. In every case, the inputs might be anything else in the database; finding the right inputs and their relationship with the outputs is the model building {1.3}.

*Classification Algorithms* The output is one or more discrete variables, such as responder versus nonresponder. Methods include logistic regression {2.1}, rule induction, decision trees, neural networks, *K*-nearest neighbors, and support vector machines.

*Regression Algorithms* The output is one or more continuous variables, such as blood pressure. Methods include linear regression {2.1}, regression trees, and neural networks.

*Segmentation Algorithms* The output is the assignment of a patient into one of several groups, or clusters, of patients that have similar properties. Methods include clustering techniques, neural networks, and visualization methods.

*Dependency Algorithms* The output is any variable in the data set, and often outputs are inputs for other relations in a chain or network that represents associations among different variables. Methods include correlation analysis, association rules, and Bayesian networks.

Finally, after a model is built, its prediction performance is evaluated. Cross-validation, bootstrapping, and bagging are the techniques usually used. Sensitivity, specificity, positive predictive value, negative predictive value, accuracy, and receiver operating characteristic (ROC) curve are the mostly used evaluation metrics. For a comprehensive introduction to these ideas and data mining generally, see [10].

### 2.4 Dose–Exposure–Response Modeling

A fundamental goal of drug development is to learn what dose of the drug to give to which patients. Most drugs reach their site of action via the systemic circulation, so drug concentrations in the blood or plasma are often more predictive of response than dose. Hence, one way of building a dose–response model is to concatenate a dose–exposure (or pharmacokinetic) model with an exposure–response (or pharmacokinetic/pharmacodynamic) model. But direct modeling of dose–response is also done. This section describes these various models.

### *Dose–Response Modeling*

**Purpose:** *To characterize the relationship between dose and any measurable response.*

Dose–response modeling is regression modeling {2.1} or longitudinal modeling {2.2} with dose as input and response as output. Examples are models (7–14). Responses may be biomarkers, clinical outcomes, adverse events, or even pharmacokinetics, although the latter merits separate consideration (see below). See {3.4} for use of such models in dose-finding trials.

### *Pharmacokinetic (PK) Modeling*

**Purpose:** *To describe measured drug concentrations as a function of dosing regimens and, usually, time.*

Pharmacokinetics is "what the body does to the drug," as in absorption, distribution, metabolism, and elimination (ADME). Systemic drug concententrations are measured in repeated blood (or plasma) samples over time after one or multiple doses administered to healthy volunteers or patients. PK models characterize such concentrations by mathematically quantifying the ADME processes as parameters such as clearance (the flow at which the drug leaves the body), volume of distribution (the apparent volume in which the drug is distributing), half-life (the rate at which the drug is declining, once transient events in the body have taken place, a parameter that combines influences of elimination and distribution). Examples of such models are (1, 3–6).

Models (5–6) are appropriate when the PK behaves as if drug distributes within a single, well-mixed compartment. Quite often, however, the PK data require assuming the drug distributes among multiple compartments; usually two or three are adequate to describe a PK profile. All such models are referred to as compartment models.

A comprehensive reference on PK models is [11].

Ultimately, the number of compartments in the body is (at least) the number of homogeneous tissues. This is recognized by physiologically based PK (PBPK) modeling, which, in the situation of transitioning from animal studies to first-study-in-man, attempts to predict the key human pharmacokinetic parameters in advance before doing the actual experiments {3.1}. A PBPK model is based on literature data in human and animal models about blood flows and sizes of respective organs, and on actual compound-specific PK data in animals (e.g., the rat, mouse, dog, and less often the monkey). For animal-to-human extrapolation, the approach is extremely resource intensive and problematic, despite recent advances. In rare situations, retrospective attempts to develop PBPK models have been largely successful, for instance, for well-known drugs (oncology, morphine) or in other applications, like extrapolations from adults to children [12].

Population PK is another extension of PK modeling; see discussion below for details.

Pharmacokinetics results provide instrumental knowledge along the development of a compound and are essential for the drug submission dossier to help characterize the clinical pharmacology of the new chemical entity [13]. The classical method for data analysis initially avoids models and is therefore sometimes called "noncompartmental analysis." Metrics of exposure to the drug like $C_{max}$ (the

maximum concentration value) and AUC (the area under the measured concentration time points) are determined through simple calculations. This approach, combined with cross-over experiments assessed by analysis of variance, is extremely powerful for descriptive and comparative purposes (e.g., across formulations or dosing regimens).

Quite often, the exposure metrics $C_{max}$ and AUC are considered as outputs in dose–response models (see discussion above). A commonly used model is the so-called power model:

$$\mathrm{AUC} = \beta_0\ \mathrm{dose}^{\beta_1} \tag{16}$$

Of interest is whether or not $\beta_1 = 1$, in which case the PK is said to be dose proportional or linear. Testing whether $\beta_1 = 1$ is a confirming role for (16).

### *Pharmacokinetic/Pharmacodynamic (PK/PD) Modeling*

**Purpose:** *To characterize the mutual variation of drug concentrations and biomarkers or clinical responses to the drug.*

Pharmacodynamics is "what the drug does to the body." PD data comprises observations of a marker of drug effect in a subject, be it a systemic marker (e.g., a hormone or a pathway intermediate), a diagnostic assessment [e.g., a QT interval within the electrocardiogram (ECG)], a surrogate marker [e.g., HbA1c for diabetes or CD4 counts for human immunodeficiency virus (HIV)], or a clinical outcome (e.g., fractures in osteoporosis, death in oncology). Usually, the observations are repeated over time, or, for single events such as death, the time to event is relevant.

A very simple class of PK/PD models, commonly used for clinical outcomes in chronic diseases, uses the same empirical forms as dose–response models (see discussion above) but replaces the input dose with a summary measure of systemic exposure such as AUC.

The objective of PK/PD modeling is often to describe temporal delays of the PD response relative to the PK. A classical type of PK/PD modeling for this purpose is the effect-compartment model, when the drug exhibits its effect in a compartment that is not in rapid equilibrium with the systemic circulation [14]. Such a compartment may represent a real physiological space, such as the brain for anesthetics, or it may be empirically imputed solely to account for the observed delay. The drug concentration in such a compartment is an output of the PK part of the model and an input for the effect of the drug on the PD. In some cases, when the distribution to the site of action is extremely slow, the maximal effect can be achieved late after the administration of the drug, when the systemic concentrations are no longer detectable (have reached levels below the limit of assay quantification).

Another more physiological approach for accounting for delays between the time course of the drug effect relative to drug concentration is the turnover model (initially introduced as the "indirect-response model") [15]. Under normal physiological conditions, the rate of production, $k_{in}$, of the marker of drug effect is equal to the rate of elimination, $k_{out}$; "nothing happens"—the level of the marker remains constant at $k_{in}/k_{out}$. Depending on the compound and the marker, the drug effect is described as a stimulation or inhibition of the rate of production or of the rate of elimination. For instance, a single turnover model is adequate to describe, across a broad range of bisphosphonate doses, the kinetics of selected markers of bone

degradation which are inhibited by 0–80% even in healthy subjects. These markers are fast, reacting over hours/days. Bone mineral density is a slow bone marker, where changes are detectable over months/years.

It is important that the PD assessment is part of the causal chain linking the dose of drug to the ultimate clinical effect (e.g., reduction of fracture in case of a drug for osteoporosis, diminishing the risk of heart failure for a cardiovascular drug). Biomarkers are important tools to make drug development more efficient. Reacting more quickly to therapeutic interventions than clinical responses, sometimes even in healthy subjects, they may be signals to guide rapid and inexpensive decision making. The number and type of links in the causal chain will vary; for example, a biomarker may be a parallel process to the clinical effect (as shown in Fig. 2) or it may be directly in the causal chain. Nonetheless, the biomarker should clearly indicate that the drug has bound to the intended target.

In the effect-compartment model and in the turnover model, concentration serves as the stimulus to effect. In contrast, there are situations where PK and PD are intimately linked and cannot be separated, for example, monoclonal antibody kinetics [16], where the PD, the interaction of the drug with its receptor, plays an important role in the actual disposition and elimination of the drug, the PK. With such target-mediated drug disposition, further complexities arise; for example, the system may become nonlinear, with overproportional increase in AUC with dose. PK/PD models are necessary; noncompartmental approaches for PK analysis are largely insufficient.

Further mechanism-based PK/PD models have been reviewed recently [17]. These authors highlight that the principles are well known and illustrate a large amount of experience with these models in a wide variety of therapeutic areas. Much of this experience has accumulated presently in academic settings. However, experience within the pharmaceutical industry is increasing dramatically as the potential of the approaches reaches the attention of senior management and regulators.

In early-phase drug development, PK and PK/PD models are often used to predict the concentration (and response) after multiple dosing based on the first human trial after single-dose administration. Such models may be used in clinical trial simulation {2.8} to help in planning future studies, such as proof-of-concept studies {3.3}.

Since the uncertainty under such circumstances is great, one cannot expect a model to be always right, even if it is based on the most accurate quantitative data and knowledge available at this particular point. Learning in the form of ongoing revision to the model ought to occur all along the development of the compound and should incorporate insights from the development of similar pharmacological entities or different compounds addressing the same therapeutic indication.

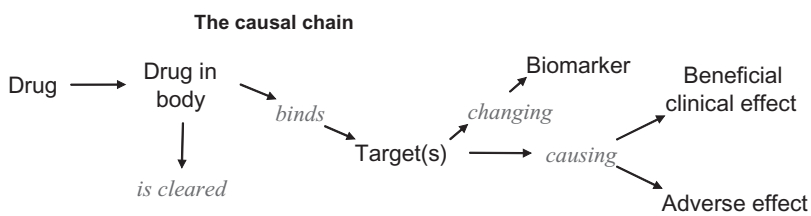


***FIGURE 2*** *Causal chain.*

In terms of regulatory submission, when a model is used for extrapolation, it must be validated beforehand or agreed upon with the regulators that the approach taken is convincing [18].

As an integrative tool later in development, population PK/PD modeling is more and more used for dose–exposure–response assessment [19]; see discussion below.

### *Population PK Modeling*

**Purpose:** *To explain and predict how systemic drug exposure varies among subjects.*

Population pharmacokinetics (PopPK) originated as an attempt to infer the patterns of PK response within a patient population from sparse PK data collected under routine clinical care [20]. Nonlinear mixed-effect modeling {2.2} was employed, according to which the parameters of compartmental PK models are themselves considered outputs of models where patient covariates (and random effects) comprise inputs. For example, clearance might depend on body size, age, and the presence of some concomitant therapy that induces a drug–drug interaction.

PopPK has evolved into a widely used tool in clinical drug development for characterizing the variability of PK in patient populations by means of nonlinear mixed-effect models [21, 22].

PopPK is valuable when "a reasonable a priori expectation exists that intersubject kinetic variation may warrant altered dosing for some subgroups in the target population" [23]. It can be used strategically to help assess a need for altered dosing, as in ethnic bridging [24]. It is also useful whenever PK understanding is needed but only sparse PK data can be collected, for example, pediatric trials [25]. Regulators frequently request a PopPK analysis as part of a submission dossier, and many drug development scientists believe that a PopPK characterization of variability is an essential component of a development program.

The steps of using PopPK for clinical trials are design, data collection, model building, model validation, and model application.

The design goal for PopPK is usually to extract useful PK information from few blood samples per patient. Optimal design methods have been implemented in software for such purposes [26].

Data collection can be problematic because the clinical sites are not always used to it. Care must be taken that dates and times of dosing and blood sampling are accurately recorded because errors in that data can lead to biased and imprecise estimates of the model's important PK parameters.

Building the submodels that describe how the compartmental PK parameters depend on covariates is an intensively data-driven task. Optimal strategies for this step are the subject of much current research [27].

Model validation is demonstrating that the model is credible for its intended purpose. If the only purpose is quantification and explanation of variability, then validation might consist of checking stability of the parameter estimates when refitting the model to random subsets of the data [28]. Ever more frequently, concentration outputs from PopPK models are used as inputs into models for clinical responses as part of integrated dose–exposure–response models (see next section) and for clinical trial simulation {2.8}. To be used for such a purpose, the model should be validated by demonstrating that its predictions are consistent with observed data

(predictive checking [29]), preferably new data that was not used as part of the model building (external validation [30]).

#### *Integrated Dose–Exposure–Response Modeling*

**Purpose:** *To comprehensively describe relationships among dose, exposure, and clinical response (efficacy and safety), often by pooling data from several clinical studies.*

The clinical response (effectiveness, undesirable effects) to a specific dose of a drug varies from patient to patient. With an integrative dose–exposure–response (IDER) model, one tries to understand which factors are responsible for individually differing responses. This information is important for the safe and effective use of a drug in individual patients. For example, it can help to specify an appropriate individualized starting dose, or the best way to adjust doses for individual patients [31].

The IDER analysis is typically done in two steps: First, the relationship between dose and drug concentration in blood (exposure) is described, and factors responsible for variations in exposure are identified (such as body weight or renal function); a PopPK analysis (see above) can provide this information. Second, the PK/PD relationship between drug concentration in blood and clinical response is investigated, and factors responsible for individual differences (such as gender or ethnicity) are identified. The PK and PK/PD models are cumulatively learned throughout the drug development process. Both mechanistic models (e.g., pathway models {2.5} or physiological models {2.6}) and empirical models (e.g., regression models {2.1} or longitudinal models {2.2}) are used for the development of the IDER relationships.

An IDER model can have a powerful impact in getting the drug approved and providing guidance on dosing. An example is the development of Certican (everolimus) in combination with Neoral (cyclosporine) for prophylaxis of acute organ rejection in renal transplant recipients. Studies in about 1200 patients showed good efficacy but unsatisfactory renal function when everolimus was administered as 1.5 or 3.0 mg daily together with full dose cyclosporine [32, 33]. An IDER model [34] revealed that improved efficacy was associated with everolimus exposures above 3 ng/mL, but that efficacy did not vary with observed cyclosporine exposures. On the other hand, better renal function was associated with lower cyclosporine exposures, whereas renal function did not vary with observed everolimus exposures. The IDER model thus suggested that therapeutic drug monitoring (TDM) should be used for everolimus to ensure that everolimus exposure is above 3 ng/mL in each patient, and that cyclosporine exposure should be individually reduced using TDM. This regimen was then prospectively tested in an additional study with about 250 patients that confirmed the predicted good efficacy and renal function [35, 36].

### 2.5 Pathways Modeling

**Purpose:** *To describe and quantify the dynamic behavior of biological signaling pathways.*

As discussed in {2.4}, clinical response to a given drug varies from patient to patient. This variability can be due to routinely collected demographic factors such as gender and race, but it can also be due to different protein and gene expression

levels, as well as mutations in relevant biological entities, such as the drug target or the drug metabolizing enzymes.

Mathematical models of signaling pathways elucidate the dynamical behavior of such relevant biological entities. In terms of target discovery and validation, such models may suggest what are the optimal ways that the system can be manipulated by drugs [37]. In targeted therapies used in oncology, such models can help elucidate possible combinations of drugs that would together achieve efficacy without interfering with the physiological function of the drugs' targets in normal tissue, thus limiting toxicity [38, 39]. In combination with molecular dynamics simulations, such models can address the role that a mutation in the drug target has in the overall clinical efficacy seen with a drug [40]. Schoeberl et al. [41] have used mathematical modeling to show that different protein expression levels from different cell lines and tissues can lead to largely different behavior of the same signaling network and thus helped to reconcile seemingly contradictory literature results.

Signaling pathways are usually modeled by a set of differential equations {1.1}, where the state variables are proteins or protein complexes within the signaling pathway. The kinetic parameters in the system include the Michaelis–Menten constant, turnover numbers, and rate constants of association and dissociation [42].

Development of highly predictive pathway models is difficult because such models require (1) complete characterization of the molecular interactions in the pathway and (2) numerical values for the parameters associated with the model equations. Parameter values are often determined via numerical optimization against experimentally measured quantities (model fitting, {1.3}). This, in turn, creates an additional complexity for models of realistic size since the number of the unknown parameters is too large as compared to the available experimental data [37]. Sensitivity analysis can determine the dominant parameters, which helps both by pointing to data that can best inform and constrain the model, as well as by suggesting possible biomarkers [43]. In addition, specialized techniques such as extended metabolic control analysis [44], determining the components that control the dynamics of signaling pathways, facilitate the extraction of the relevant pathway modules to be included in a larger disease model {2.7} for clinical development.

## 2.6 Physiological Modeling

**Purpose:** *To describe mathematical models of tissues and organs and outline possible utility in clinical drug development.*

Mathematical models of human physiology have been of interest for a long time [45]. The IUPS Physiome project (http://www.physiome.org.nz) represents a worldwide public domain to build a computational framework and quantitative understanding of human physiology and pathophysiology. As new data accumulates, these models have become increasingly more sophisticated and biologically plausible. Organs that have been described by such mathematical models include kidney [46], respiratory tract [47], heart [48], and even the whole human physiology [49]. Because constructing these models is complex, data intensive, and computationally expensive, such modeling efforts have been largely the subject of academic research with somewhat limited applications to human health. However, the situation appears to be changing with the accumulation of data, government funding, and establishment of consortia and collaborative efforts that pool data across different organizations.

One example for a tissue-level model that is being used to guide clinical decision making is that of cardiac electrophysiology. The last decade has been marked by the withdrawal of several medicines whose use in patients has been associated with the development of torsade de pointe (TdP), a potentially fatal polymorphic tachycardia [50]. Regulatory guidelines [51] propose the prolongation of the QT interval on the electrocardiogram as a biomarker for torsadogenic risk of a given compound. However, TdP is triggered by a dynamic combination of multiple factors, and so, in order to reduce the number of false positives and false negatives, pharmaceutical companies are looking into combining the strengths of different methodologies (both experiemental and modeling) to make the best informed decision. A detailed description of the cardiac system in humans allows inclusion of experimental data on the effects of different drugs in a relatively simple manner [52] and prediction of the effect of the molecules both in reference to healthy subjects as well as high-risk individuals. It is worthwhile to note that as the models become more predictive and computation becomes cheaper, such prospective simulations can be used routinely in the drug development process and can serve as data integration tools.

## 2.7 Disease Progression Modeling

**Purpose:** *To improve the assessment and predictability of a drug's impact on a disease by distinguishing drug-specific properties from disease-specific properties.*

Many chronic diseases that are targets of drug development, such as Alzheimer's, diabetes, and osteoporosis, are degenerative. The durations of clinical trials may not be insignificant on the time scales of disease progression. So disease progression is a source of variability, accounting for which can improve the detection of drug effect. On the other hand, clinical trials are almost always short relative to the expected duration of a drug's use by a chronic-disease patient. So a reliable forecast of a drug's long-term impact is important for developers, prescribers, patients, and payors.

Disease progression models first quantify the evolving status of a disease absent a drug and then characterize alterations induced by the drug. This hierarchical approach facilitates clarity, extrapolation to different dose regimens, and comparison of different drugs.

Although models for different diseases might be expected to have little in common, certain simple model forms have been found practical as well as conceptually enlightening. Chan and Holford [53, 54] considered the constant progression model,

$$\frac{dS}{dt} = \alpha \qquad S(t) = S_0 \tag{17}$$

where $S(t)$ is the disease status at time $t$, $S_0$ is the initial status, and $\alpha$ is the rate of disease progression. They cited use of such a model for Alzheimer's, Parkinson's, chronic obstructive pulmonary disease (COPD), and diabetic nephropathy.

A drug may alter $S(t)$ by affecting $S_0$, $\alpha$, or both. A drug effect on $S_0$ is considered a symptomatic effect; if drug treatment stops, $S(t)$ reverts to its original trajectory. A drug effect on $\alpha$ is considered a protective effect; if drug treatment stops, $S(t)$

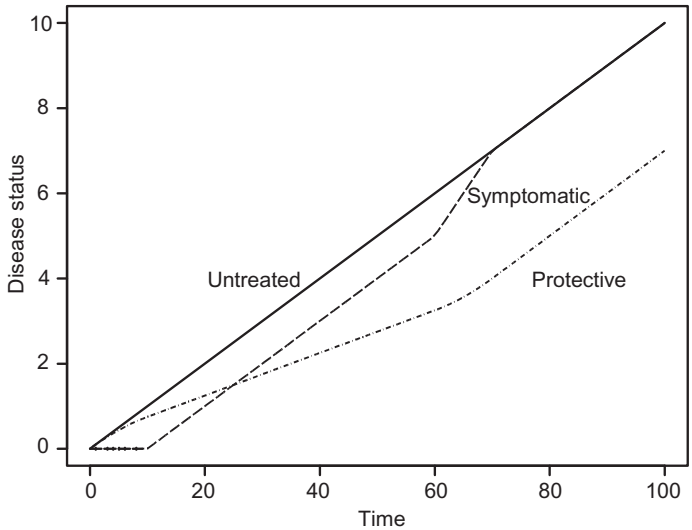


***FIGURE 3*** *Untreated disease progression and symptomatic and protective drug effects.*

continues on a trajectory displaced from the original at the position where treatment stopped. See Figure 3.

Post et al. [55] introduced a class of models based on the disease process being a perturbation of a turnover model {2.4}. The disease alters $k_{in}$ or $k_{out}$ in a progressive way. Symptomatic drug effects add to or multiply the perturbed $k_{in}$ or $k_{out}$ without affecting the process of perturbation. Protective drug effects affect the process of perturbation directly. Models of this sort were applied to type 2 diabetes in [56].

## 2.8 Clinical Trial Simulation

**Purpose:** *To assess the expected performance of clinical trial designs and analysis methods.*

Clinical trials often have one or both of two objectives related to the two goals of drug development {1}: (i) learning—estimate some characteristic of the drug's behavior, such as the maximum tolerated dose; and (ii) confirming—test some hypothesis about the drug, such as whether the drug is superior to placebo. Modelers have long studied how different experimental designs, estimates, and tests affect the accuracy and precision of such inferences [57]. When their algebra and calculus are not powerful enough to quantify these properties, modelers resort to simulating {1.2} their experiments on the computer, for example, [58, 59].

Holford et al. [60] identified three broad classes of questions for which simulation experiments might be useful: (1) To investigate the properties of new model-based methods, such as nonlinear mixed-effect modeling for PopPK {2.4} [61]. (2) To study the properties of novel trial designs, such as concentration-controlled trials [62]. (3) Applications to specific clinical trials, for example, [63, 64]. For the latter, different designs can be input to the model and the best design selected for actual implementation based on comparison of the outputs. Simulating the different designs may

cost tens of dollars; actually conducting the trial may cost millions; the small investment of model-based forethought can have a big impact.

A trial simulation model may be thought of as one large model, the key input to which is the design of a clinical trial, and the key output of which is some performance measure of the trial. However, it is useful to consider component models. Commonly, these include an integrated dose–exposure–response (IDER) model {2.4}, a covariate distribution model, and a trial execution model [60].

The covariate distribution model describes the trial's target patient population in terms of the distribution of relevant covariates, such as the distribution of blood pressures in (2).

The trial execution model is a computer algorithm for simulating the trial. It generates simulated patients according to the covariate distribution model who experience outcomes according to the IDER model. The trial execution model may allow deviations from the trial protocol through noncompliance, missing observations, and the like. The trial's simulated data are statistically analyzed according to the protocol. This is done repeatedly for many simulated replications of the trial. Then the statistical outomes of all of the trials are summarized as a performance measure that addresses the question that motivated the simulation experiment.

An excellent compendium on clinical trial simulation is [65].

## 2.9 Decision Analysis

**Purpose:** *To help clinical teams create, compare, and evaluate strategies for clinical development, and to achieve clarity of action in choosing among strategies in the face of uncertainty.*

Modeling and simulation is a powerful tool for decision making due to its role in knowledge integration and its ability to quantify "what-if" scenarios via trial simulation {2.8}. Decision analysis is a structured process designed to facilitate effective communication between decision makers and a project team (including modelers and experts), driving toward clarity of action [66, 67]. Matheson and Matheson [68] characterize six components of a decision analysis: (1) framing, (2) strategy generation, (3) meaningful, reliable information, (4) clear values and trade-offs, (5) logically correct reasoning, and (6) commitment to action. Decision quality is achieved when further improvement in each component is not worth the additional time and resources that would be required. Each of these six components is now explained:

1. *Framing* consists of understanding the key questions. A decision hierarchy is developed to distinguish between those decisions already made, those decisions that are to be made now, and those that can be settled at a later time.
2. In *strategy generation*, creative doable strategies are assembled from different alternatives associated with the decisions to be made now.
3. *Meaningful, reliable information.* An integral feature of decision-analytic thinking is to accept the existence of uncertainties and to quantify them by means of probability distributions. Probabilities over clinical (and sometimes also market) possibilities may be developed via modeling from data, but also may be carefully assessed from experts in a manner designed to reduce various biases [69].

4. *Clear values and trade-offs* are selected to allow meaningful, quantitative comparisons of strategies. For-profit companies will typically be concerned with expected (i.e., probability averaged) net present (dollar) value. However, some useful proxies may also be evaluated, and then explicitly balanced against resources (i.e., cost and time). In early-phase "proof-of-concept" trials {3.3}, the proxy focus is typically the chance of a correct "go/no-go" after completion of the trial. In later phase (dose finding) trials {3.4}, the focus shifts toward the chance for correctly choosing the best dose regimen(s). For phase III trials, the focus is the chance of correctly confirming a compound's safety and efficacy, leading to successful registration. Beyond the clinical development of a single compound, larger decisions may be analyzed, such as allocation of resources between many development projects in a portfolio [70].
5. *Logically correct reasoning* proceeds comparing the performance of selected strategies against the chosen evaluation criteria. Typically, trial simulation and probabilistic (Bayesian) inference {1.4} must be used. Sensitivity analysis can also be performed and the value of information computed [67], allowing better focus on the most salient information.
6. *Commitment to action* among stakeholders is achieved by involving the right people in the above steps. This includes a planned sequence of opportunities for the decision maker to provide input into the frame, the strategies chosen for consideration, and the evaluation criteria.

Other authors might use the term "decision analysis" to refer to steps 3–5. Examples of such decision-theoretic M&S applied to clinical trial designs can be found in [71, 72]. However, the fuller decision analysis process that includes all of steps 1–6 often generates insights leading to improved strategies; teams who engage in such a systematic process frequently report that the open, guided discussion of assumptions, uncertainties, alternatives, and values along with emergent insights provide as much or more benefit compared with the mathematical model outputs.

## 3 M&S-BASED CLINICAL TRIALS

### 3.1 First-in-Man Studies

**Purpose of M&S:** *To estimate doses for first-in-man studies.*

A hallmark of a good development candidate is to have a high probability of surviving through the clinical proof-of-concept PoC study {3.3}, showing pharmacologically useful activity. To assess this one answers three questions: (1) Is the site of action exposed to sufficiently efficacious drug concentration for a suitable duration? (2) Do these concentrations have a suitable safety profile? (3) Can formulations that achieve these concentrations be developed? Several research and development functions generate data, each utilizing different methodologies. M&S should play a key role to fully integrate these data, anticipate doses likely to elicit pharmacological and adverse effects in humans, and then optimize the early clinical development program.

The dose for a first-in-man (FIM) study is generally determined by following international, public guidances of the International Conference on Harmonisation (ICH) E8 [73] and ICH M3 [74]. These provide the rationale for estimating a safe starting dose by specifying target organs, dose dependency, relationships to exposure, and potential reversibility in FIM trials. Further, ICH S6 [75] (for biologicals) defines the principle of relative safety exposure between test animals and humans. The no-observed-adverse-effect level (NOAEL) and the toxic dose in animals are the bases of anticipating the FIM dose, and uncertainty of estimation is accounted for by setting safety margins. The Food and Drug Administration (FDA) [76] gives guidance on the maximum recommended starting dose (MRSD) for FIM studies by $mg/m^2$-based scaling of Good Laboratory Practice (GLP) toxicology data with a typical 10-fold safety margin.

That said, the guidance [76] does not cover how to determine the pharmacologically active dose, that is, the lowest dose with the intended pharmacological activity, making MRSD essentially a safety secured dose. However, an MRSD far below the pharmacologically active dose requires a long dose escalation scheme, while an MRSD far exceeding the right dose range will not allow us to detect the complete dose–effect relationship [77]. In addtion, the guidance-compliant approach does not necessarily fully secure the safety, as highlighted in 2006 by TGN1412, a monoclonal antibody whose MRSD was calculated by conventional approaches with a 500-fold safety margin, but six healthy volunteers were seriously injured in the FIM study.

Collectively, a rational approach is demanded to efficiently anticipate FIM doses, in addtion to the guidances, following the core scientific principle that drug effects are a function of target exposure. A three-step process is practical to anticipate the human dose–response [78] (Fig. 4): (1) Experimentally, determine exposure–response relationship in preclinical (animal) studies. Blood (plasma) concentration is often considered as a surrogate to the target exposure, assuming drug distribution equilibirum between circulating blood and drug target. Summary measures such as AUC or $C_{max}$ {2.4} may be used if appropriate. (2) Adjust the exposure response curve from preclinical species to human. This will involve correcting for interspecies differences (e.g., known binding affinities to blood components and/or the drug target). (3) Convert the anticipated exposure–response to dose–response in humans using

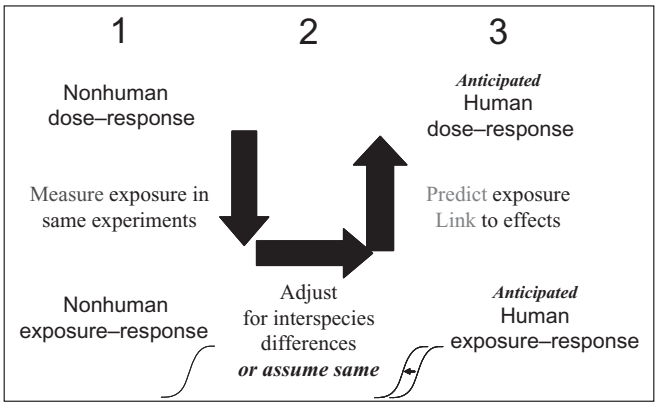


***FIGURE 4*** *Anticipating the first-in-man dose.*

pharmacokinetic scaling, such as allometry [79] or physiologically based pharmacokinetic modeling {2.4} [80].

### 3.2 Oncology Phase I Studies

**Purpose of M&S:** *To optimally design and analyze first-in-man trials for oncology drugs.*

First-in-man trials for oncology trials differ from other therapeutic areas because oncology drugs often cannot be tested in healthy volunteers due to their toxicity at therapeutic doses. The typical design for such trials is therefore a dose escalation scheme in patients, with interim analysis and dose escalation/stopping decisions after each cohort. The primary endpoint is the occurrence of a dose-limiting toxicity (DLT), and the primary objective is the assessment of a maximally tolerated dose (MTD), corresponding to an acceptable target rate of DLTs. In order to optimally design such trials, the M&S challenge is to define decision rules for dose escalation and MTD selection (specifically, model-based rules), and to evaluate their operating characteristics through simulation.

Heuristic rules to decide upon dose escalation and MTD declaration have been proposed [81, 82], the most popular being the so-called "3+3" method, which bases decisions on whether to escalate, keep, or deescalate doses on how many DLTs are observed in the previous cohorts of size 3. Starting with the so-called continual reassessment method (CRM) [83], many model-based methods have been introduced, where decision rules and MTD selection are based on a model fitted to the available data. In its simplest form, where DLT is binary (i.e., occurrence or nonoccurrence), a model for DLT has the rate of DLTs as output and the dose as input along with parameters: rate of DLTs = $f$(dose, parameters). Due to the sparseness of the data (binary data from few patients), complicated models with many parameters cannot be fitted; thus, the proposed models typically have one or two parameters. An example is the two-parameter logistic function: logit(rate of DLTs) = $\alpha + \beta \log(\text{dose})$, where $\text{logit}(p) = p/(1-p)$ {2.1}. After each cohort, the model is fitted to the data, either through maximum likelihood {1.3} or using a Bayesian approach {1.4}. The latter has some distinct advantages since it allows to base decisions on posterior probabilities of the quantities of interest [84]. For example, one may want to specify the optimal target rate of DLTs to be in the range [0.25, 0.33]; anything beyond 0.33 would be deemed unacceptable toxicity. In a Bayesian framework, this is easily transformed to a decision rule, by choosing a dose maximizing the posterior probability of the rate of DLTs being in the interval [0.25, 0.33], while keeping the posterior probability that the rate of DLTs exceeds 0.33 below a threshold, for example, 0.05; this may be complemented with restrictions on how many dose levels to escalate.

The evaluation of operating characteristics is done through clinical trial simulation {2.8}. Generic features can be evaluated through the simulation of typical situations (see, e.g., [84]). However, specific trials with a unique set of prior assumptions and logistic restrictions will very often require specific simulation-based evaluation.

For a discussion of different approaches, with an emphasis on Bayesian methods, see [84] and the references given therein. For a nontechnical introduction see [85]. For references to extensions such as time-to-event and continuous outcomes, covari-

ates, combination trials, and combined safety–efficacy outcome, see references given in [84–87].

### 3.3 Proof-of-Concept Studies

**Purpose of M&S:** *To improve the information value of proof-of-concept trials with regard to efficacy and safety evaluation.*

Early development studies for establishing proof-of-concept (PoC) generally use small patient cohorts (typically between 10 and 20) observed for a relatively short period of time (several weeks) in a dose escalation design to evaluate early efficacy and safety signals. Cohorts are assigned, in sequence, to increasing doses until the maximum tolerated dose, generally determined in a previous study, is reached, or unacceptable safety is observed for a given cohort. A new cohort is only allowed to start once acceptable safety signals are verified for all previous doses. At the end of the study, one hopes either to determine a dose range for further exploration in phase IIb or to conclude that no PoC can be established based on the efficacy/safety trade-off.

Because of the small cohort sizes, only safety problems occurring in a relatively large percentage of patients can be reliably detected. Likewise, only relatively strong efficacy signals can be detected with reasonable statistical power using traditional pairwise hypothesis tests.

Safety and efficacy variables are often measured on a continuous scale and observed several times over the duration of the study. However, typically, the endpoints for the go/no-go decision are based on a single time point (e.g., change from baseline at end of study) and on dichotomized versions of the original variables to characterize responder/nonresponder behavior. An example of the latter is the transformation of continuous liver function test measurements [e.g., alanine aminotransferase (ALT) and aspartate aminotransferase (AST)] into binary indicators of exceeding three times the upper limit of normal. There are, therefore, two types of information loss often present in PoC studies: (1) not using of all available longitudinal measurements collected in the study and (2) the dichotomization of continuous endpoints.

Safety and efficacy signal detection can be made more efficient in a variety ways: by using information external to the trial and longitudinal modeling approaches {2.2} to exploit all available information. Furthermore, the utility of PoC studies within the drug development program can be enhanced by incorporating the information obtained in them directly into later phase trials. Bayesian modeling techniques {1.4} are particularly useful in implementing these different approaches.

The key idea is to use a Bayesian mixed-effect model (typically nonlinear) {2.2} to characterize the longitudinal dose–response behavior of efficacy and safety variables, measured on a change from baseline scale (e.g., log of measurement divided by baseline). Preclinical or other external data can be used in the elicitation of priors for the Bayesian model. The estimated model is then used to simulate longitudinal profiles of change from baseline for different doses, which are combined with a sample of baseline values, obtained from a database of previous trials with a similar population, to produce dose–response longitudinal profiles. Dichotomization and other transformations can be applied to the simulated profiles to evaluate efficacy and safety more efficiently compared to using a single dichotomized time point. An

additional benefit of the approach is that it allows the use of external information, frequently available in pharmaceutical companies.

### 3.4 MCP-Mod: Unified Strategy for Dose-Finding Studies

**Purpose of M&S:** *To combine multiple comparison procedures and modeling in dose finding.*

Understanding and adequately representing the dose response (DR) profile of a compound, for both efficacy and safety, is a fundamental objective of clinical drug development. Proper understanding of this relationship has both a confirming and a learning component: confirming whether there is an overall DR effect (proof-of-concept {3.3}); and if so learning which doses to select for further development (dose finding). Selecting too low a dose decreases the chance of showing efficacy in later studies, while selecting too high a dose may result in tolerability or safety problems.

The analysis of dose-finding studies can be classified into two major strategies: modeling [88,89] and multiple comparison procedures (MCP) [90,91]. Dose–response (DR) modeling {2.4} assumes a functional relationship between dose (taken as a quantitative factor) and the response, according to a prespecified parametric model (defined in the study protocol). While this provides flexibility in investigating the effect of doses not used in the actual study, the validity of its conclusions highly depends on the correct choice of the DR model, which is typically unknown.

A multiple comparison procedure regards dose as a qualitative factor and makes very few, if any, assumptions about the underlying DR model. It can either be used for detecting an overall dose signal by means of trend tests or for the estimation of target doses by stepwise testing strategies, while preserving the overall type I error rate at a prespecified level. Such procedures are relatively robust to the underlying DR shape but are not designed for extrapolation of information beyond the observed dose levels.

MCP-Mod is a hybrid methodology combining features of MCP and modeling to provide a flexible and principled approach for designing and analyzing dose-ranging studies. Its central idea is the use of a set of candidate DR models to cover the possible shapes anticipated for the DR relationship. MCP are applied to a set of test statistics, determined optimally for the models in the candidate set, to decide which shapes give statistically significant signals. If no candidate model is statistically significant, the procedure stops and declares that no DR relationship can be established from the observed data (i.e., no proof of concept). Otherwise, out of the statistically significant models a best model is selected for dose estimation in the last stage of the procedure. The selection of the dose estimation model can be based on the minimum $P$ value (of the candidate models) or some other relevant model selection criteria such as AIC or BIC {1.3}. The selected dose–response model is then employed to estimate target doses using inverse regression techniques [where the output (response) and input (dose) exchange roles] and possibly incorporating information on clinically relevant effects.

Simulation results suggest that MCP-Mod is as powerful as standard trend tests, while allowing more precise estimation of target doses than MCP, due to its modeling component [88, 92]. It is worth pointing out that this method can be seen as a

seamless design that combines proof of concept (phase Ib/IIa) with dose finding (phase IIb) into one single study.

### 3.5 Adaptive Trials

**Purpose of M&S:** *To determine the operating characteristics of an adaptive design, and ensure that the optimal design will be conducted.*

Adaptive designs are a relatively new type of clinical trial in which data from the trial can be used to modify some design aspect of the trial while the trial is still ongoing. This greater flexibility allows for design aspects to be modified in the case where some original trial design assumptions were incorrect, without having to wait for the conclusion of the trial. Such flexibility can save a considerable amount of development time for a novel drug. Adaptive trials can also combine what traditionally might be two separate studies into a single trial, thereby eliminating the time between the two trials. However, in order to design these trials, the decision process must be carefully thought out in advance and specified in the protocol. Often, clinical trial simulation {2.8} is the best way to understand and modify the operating characteristics of an adaptive trial.

Although there are many types of adaptation, one common type, sample size reestimation, is very appealing because often the parameters needed for sample size calculations are completely unknown at the beginning of the trial. Therefore, the idea is to try to estimate these needed parameters during the trial and modify the sample size accordingly. However, it has to be understood that at this interim time, the estimates for these parameters have considerable variability in them as well. Therefore, it might not be optimal to use such estimates directly, but instead use them to guide the calculation for the final sample size. A decision rule can be defined for how the sample size will be changed, and then computer simulation can be used to determine how the sample size would change under various scenarios.

Another type of adaptive design is a seamless adaptive design in which two trials are combined together. For example, a trial could be designed such that multiple doses are started in the beginning of the trial, and then the optimal dose is selected during an interim analysis. Additional patients would then only be randomized to the selected dosage group (and control), and all the data would be used at the conclusion of the trial to determine if the selected dose was indeed efficacious. If proper statistical testing procedures are in place at the end of the trial, then this type of design could cover objectives of both a phase II dose-finding study {3.4} and a confirmatory phase III trial.

For this type of seamless adaptive trial, simulation must also be used to determine the operating characteristics of the trial. For example, once a dose–response model {2.4} is assumed, and the dose selection algorithm is defined, the trial could be simulated to determine how often the selected dose would have a significant statistical test for efficacy at the conclusion of the trial (the statistical power for the trial). Another important characteristic of the trial that could be determined via simulation is the selection probabilities for each of the doses. Simulations should then be performed using different assumed dose responses, selection rules, and sample sizes to ensure that the trial has robust operating characteristics under many different scenarios.

For an introduction to adaptive trials, see [93].

## 4 DISCUSSION

This chapter introduces M&S in clinical drug development. Just a chapter cannot do the topic full justice, and indeed a thick book has recently been published that surveys the field [94]. The references herein provide some historical markers as well as some general introductions to specific methodologies. Readers interested in following trends would do well to monitor the journals *Clinical Pharmacology and Therapeutics, Journal of Pharmacokinetics and Pharmacodynamics, Statistics in Medicine*, and others cited in the references.

## REFERENCES


1. Sheiner, L. B. (1997), Learning versus confirming in clinical drug development, *Clin. Pharmacol. Therap.*, 61, 275–291.
2. Linhart, H., and Zucchini, W. (1986), *Model Selection*, Wiley, New York.
3. Berry, D. A. (2006), Bayesian clinical trials, *Nature Rev. Drug Dis.*, 5, 27–36.
4. McCullagh, P., and Nelder, J. A. (1989), *Generalized Linear Models*, Chapman and Hall, New York.
5. Katz, M. H. (2006), *Multivariable Analysis: A Practical Guide for Clinicians*, 2nd ed., Cambridge University Press, Cambridge, UK.
6. Harrell, F. E., Jr. (2001), *Regression Modeling Strategies*, Springer, New York.
7. Liang, K.-Y., and Zeger, S. L. (1986), Longitudinal data analysis using generalized linear models, *Biometrika*, 73, 13–22.
8. Albert, P. S. (1999), Longitudinal data analysis (repeated measures) in clinical trials, *Stat. Med.*, 18, 1707–1732.
9. Diggle, P. J., Heagerty, P., Liang, K.-Y., et al. (2002), *Analysis of Longitudinal Data*, 2nd ed., Oxford University Press, Oxford, UK.
10. Witten, I. H., and Frank, E. (2005), *Data Mining: Practical Machine Learning Tools and Techniques*, Elsevier, San Francisco.
11. Gibaldi, M., and Perrier, D. (1982), *Pharmacokinetics*, Marcel Dekker, New York.
12. Edgington, A. N., Schmitt, W., Voith, B., et al. (2006), A mechanistic approach for the scaling of clearance in children, *Clin. Pharmacokin.*, 45, 683–704.
13. Welling, P. E., and Balant, L. P., eds. (1994), *Handbook of Experimental Pharmacology, Vol. 110, Pharmacokinetics of Drugs*, Springer, Heidelberg.
14. Sheiner, L. B., Stanski, D. R., Vozeh, S., et al. (1979), Simultaneous modeling of pharmacokinetics and pharmacodynamics: Application to d-tubocurarine, *Clin. Pharmacol. Therapeut.*, 25, 358–371.
15. Dayneka, N. L., Garg, V., and Jusko, W. J. (1993), Comparison of four basic models of indirect pharmacodynamic response, *J. Pharmacokinet. Biopharmaceut.*, 21, 457–478.
16. Mager, D. E., Wyska, E., and Jusko, W. J. (2003), Diversity of mechanism-based pharmacodynamic models, *Drug Metabolism Disposition*, 31, 510–519.
17. Danhof, M., de Jongh, J., De Lange, E. C. M., et al. (2007), Mechanism-based pharmacokinetic-pharmacodynamic modeling: Biophase distribution, receptor theory, and dynamical systems analysis, *Ann. Rev. Pharmacol. Toxicol.*, 47, 357–400.
18. Stanski, D. R., and Orloff, J. J. (2008), Communicating with the FDA: The "third rail" of a new model for drug development, *J. Clin. Pharmacol.*, 48, 144–145.
19. Sheiner, L. B., and Steimer, J. L. (2000), Pharmacokinetic/pharmacodynamic modeling in drug development, *Ann. Rev. Pharmacol. Toxicol.*, 40, 67–95.

20. Sheiner, L. B., Rosenberg, B., and Marathe, V. V. (1977), Estimation of population characteristics of pharmacokinetic parameters from routine clinical data, *J. Pharmacokinet. Pharmacodynam.*, 5, 445–479.
21. Yuh, L., Beal, S., Davidian, M., et al. (1994), Population pharmacokinetic/pharmacodynamic methodology and applications: A bibliography, *Biometrics*, 50, 566–575.
22. Bonate, P. (2005), Recommended reading in population pharmacokinetic pharmacodynamics, *AAPS J.*, 7, E363–E373.
23. Food and Drug Administration (1999), *Guidance for Industry: Population Pharmacokinetics*; available at: http://www.fda.gov/cder/guidance/index.htm.
24. Uyama, Y., Shibata, T., Nagai, N., et al. (2005), Successful bridging strategy based on ICH E5 guideline for drugs approved in Japan, *Clin. Pharmacol. Therapeut.*, 78, 102–113.
25. Meibohm, P., Laer, S., Panetta, J. C., et al. (2005), Population pharmacokinetics studies in pediatrics: Issues in design and analysis, *AAPS J.*, 7, E475–E487.
26. Duffull, S., Waterhouse, T., and Eccleston, J. (2005), Some considerations on the design of population pharmacokinetic studies, *J. Pharmacokinet. Pharmacodynam.*, 32, 441–457.
27. Ribbing, J., and Jonsson, E. N. (2004), Power, selection bias and predictive performance of the population pharmacokinetics covariate mode, *J. Pharmacokinet. Pharmacodynam.*, 31, 109–134.
28. Ette, E. I. (1997), Stability and performance of a population pharmacokinetic model, *J. Clin. Pharmacol.*, 37, 486–495.
29. Yano, Y., Beal, S. L., and Sheiner, L. B. (2001), Evaluating pharmacokinetic/pharmacodynamic models using the posterior predictive check. *J. Pharmacokinet. Pharmacodynam.*, 28, 171–192.
30. Brendel, K., Comets, E., Laffont, C., et al. (2006), Metrics for external model evaluation with an application to the population pharmacokinetics of gliclazide, *Pharma. Res.*, 23, 2036–2049.
31. International Conference on Harmonisation (1994), *Dose Response Information to Support Drug Registration*; available at: http://www.emea.europa.eu/pdfs/human/ich/037895en.pdf.
32. Vítko, S., Margreiter, R., Weimar, W., et al. (2004), Everolimus (Certican) 12-month safety and efficacy versus mycophenolate mofetil in de novo renal transplant recipients, *Transplantation*, 78, 1532–1540.
33. Lorber, M. I., Mulgaonkar, S., Butt, K. M., et al. (2005), Everolimus versus mycophenolate mofetil in the prevention of rejection in de novo renal transplant recipients: A 3-year randomized, multicenter, phase III study, *Transplantation*, 80, 244–252.
34. Lorber, M. I., Ponticelli, C., Whelchel, J., et al. (2005), Therapeutic drug monitoring for everolimus in kidney transplantation using 12-month exposure, efficacy, and safety data, *Clin. Transplant.*, 19, 145–152.
35. Vitko, S., Tedesco, H., Eris, J., et al. (2004), Everolimus with optimized cyclosporine dosing in renal transplant recipients: 6-month safety and efficacy results of two randomized studies, *Am. J. Transplant.*, 4, 626–635.
36. Kovarik, J. M., Tedesco, H., Pascual, J., et al. (2004), Everolimus therapeutic concentration range defined from a prospective trial with reduced-exposure cyclosporine in de novo kidney transplantation, *Therap. Drug Monit.*, 26, 499–505.
37. Aksenov, S., Church, B., Dhiman, A., et al. (2005), An integrated approach for inference and mechanistic modeling for advancing drug development, *FEBS Lett.*, 579, 1878–1883.

38. Christopher, R., Dhiman, A., Fox, J., et al. (2004), Data-driven computer simulation of human cancer cell, *Ann. NY Acad. Sci.*, 1020, 132–153.
39. Fitzgerald, J., Schoeberl, B., Nielsen, U., et al. (2006), Systems biology and combination therapy in the quest of clinical efficacy, *Nature Chem. Biol.*, 2, 458–466.
40. Liu, Y., Purvis, J., Shih, A., et al. (2007), A multiscale computational approach to dissect early events in the Erb family receptor mediated activation, differential signaling and relevance to oncogenic transformations, *Ann. Biomed. Eng.*, 35, 1012–1025.
41. Schoeberl, B., Pace, E., Shavonne, H., et al. (2006), A data-driven computational model of the ErbB receptor signaling network, *Conference Proceedings: Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 1, 53–54.
42. Schoeberl, B., Eichler-Jonhson, C., Gilles, E., et al. (2002), Computational modeling of the dynamics of the MAP kinase cascade activated by surface and internalized receptors, *Nature Biotechnol.*, 20, 370–375.
43. Nielsen, U., and Schoeberl, B. (2005), Using computational modeling to drive the development of targeted therapeutics, *IDrugs*, 8, 822–826.
44. Hornberg, J., Binder, B., Bruggeman, F., et al. (2002), Control of MAPK signalling: From complexity to what really matters, *Oncogene*, 24, 5533–5542.
45. Hoppensteadt, F. C., and Peskin, C. S. (2002), *Modeling and Simulation in Medicine and the Life Sciences*, Springer, New York.
46. Thomas, S. R., Layton, A. T., Layton, H. E., et al. (2006), Kidney modeling: Status and perspectives, *Proc. IEEE*, 94, 740–752.
47. Subramaniam, R., Ashagarian, B., Freijer, J., et al. (2003), Analysis of lobar differences in particle deposition in the human lung, *Inhal. Toxicol.*, 15, 1–21.
48. McCulloch, A., Bassingthwaighte, J. B., Hunter, P. J., et al. (1998), Computational biology of the heart: From structure to function. *Prog. Biophys. Mol. Biol.*, 69, 151–572.
49. Abram, S., Hodnett, B., Summers, R., et al. (2007), Quantitative circulatory physiology: An integrative mathematical model of human physiology for medical education, *Adv. Physiol. Ed.*, 31, 202–210.
50. Dumotier, B., and Georgieva, A. (2007), Preclinical cardio-safety assessment of torsadogenic risk and alternative methods to animal experimentation: The inseperable twins, *Cell Biol. Toxicol.*, 23, 293–302.
51. International Conference on Harmonisation. (2005), *The Non-Clinical Evaluation of the Potential for Delayed Ventricular Repolarization (QT Interval Prolongation) by Human Pharmaceuticals*; available at: http://www.emea.europa.eu/pdfs/human/ich/042302en.pdf.
52. Bottino, D., Penland, C., Stamps, A., et al. (2006), Preclinical cardiac safety assessment of pharmaceutical compounds using an integrated systems-based computer model of the heart, *Prog. Biophys. Mol. Bio.*, 90, 414–443.
53. Chan, P. L. S., and Holford, N. H. G. (2001), Drug treatment effects on disease progression, *Ann. Rev. Pharmacol. Toxicol.*, 41, 625–659.
54. Holford, N. (2007), Modelling disease progression; available at: http://www.page-meeting.org/page/page2007/ModellingDiseaseProgression.pdf.
55. Post, T. M., Freijer, J. I., DeJongh, J., et al. (2005), Disease system analysis: Basic disease progression models in degenerative disease, *Pharma. Res.*, 22, 1038–1049.
56. de Winter, W., DeJongh, J., Post, T., et al. (2006), A mechanism-based disease progression model for comparison of long-term effects of pioglitazone, metformin and gliclazide on disease processes underlying type 2 diabetes mellitus, *J. Pharmacokinet. Pharmacodynam.*, 33, 313–343.

57. Bacchieri, A., and Della Cioppa, G. (2007), *Fundamentals of Clinical Research*, Springer, Milan.
58. Tang, D., Geller, N. L., and Pocock, S. J. (1993), On the design and analysis of randomized clinical trials with multiple endpoints, *Biometrics*, 49, 23–30.
59. O'Quigley, J., and Chevret, S. (1991), Methods for dose finding studies in cancer clinical trials: A review and results of a monte carlo study, *Stat. Med.*, 10, 1647–1664.
60. Holford, N. H. G., Kimko, H. C., Monteleone, J. P. R., et al. (2000), Simulation of clinical trials, *Ann. Rev. Pharmacol. Therapeut.*, 40, 209–234.
61. Sheiner, L. B., and Beal, S. L. (1983), Evaluation of methods for estimating population pharmacokinetic parameters. III. Monoexponential model: routine pharmacokinetic data, *J. Pharmacokinet. Biopharmaceut.*, 11, 303–319.
62. Sanathanan, L. P., and Peck, C. C. (1991), The randomized concentration-controlled trial: An evaluation of its sample size efficiency, *Controlled Clin. Trials*, 12, 781–794.
63. Hale, M. D. (1997), Using population pharmacokinetics for planning a randomized concentration-controlled trial with a binary response, in Aarons, L., Balant, L. P., Danhof, M., Eds., *European Cooperation in the Field of Scientific and Technical Research*, European Commission, Geneva, pp. 227–235.
64. Hale, M. D., Nicholls, A. J., Bullingham, R. E. S., et al. (1998), The pharmacokinetic-pharmacodynamic relationship for mycophenolate mofetil in renal transplantation, *Clin. Pharmacol. Therapeut.*, 64, 672–683.
65. Kimko, H. C., and Duffull, S. B. (2003), *Simulation for Designing Clinical Trials*, Dekker, New York.
66. Howard, R. A. (1989), The evolution of decision analysis, in Howard, R. A., and Matheson, J. E., Eds., *Readings on the Principles and Applications of Decision Analysis*, Strategic Decisions Group, Menlo Park, CA, pp. 7–16.
67. Clemen, R. T. (1996), *Making Hard Decisions, An Introduction to Decision Analysis*, 2nd ed., Brooks/Cole, Pacific Grove, CA.
68. Matheson, D., and Matheson, J. (1998), *The Smart Organization*, Harvard Business School Press, Cambridge, MA.
69. Spetzler, C., and Stael von Holstein, C. (1989), Probability encoding in decision analysis, in Howard and Matheson, Eds. *Readings on the Principles and Applications of Decision Analysis*, Vol. II, Strategic Decisions Group, Menlo Park, CA, pp. 601–626.
70. Sharpe, P., and Keelin, T. (1998), How Smith Kline and Beecham makes better research allocation decisions, *Harvard Business Rev.*, March–April, 5–10.
71. Pallay, A. (2000), A decision analytic approach to determining sample sizes in a Phase III program, *Drug Info. J.*, 34, 365–377.
72. Berry, D. (2004), Bayesian statistics and the efficiency and ethics of clinical trials, *Stat. Sci.*, 19, 175–187.
73. International Conference on Harmonisation (1998), *General Considerations for Clinical Trials*; available at: http://www.emea.europa.eu/pdfs/human/ich/029195en.pdf.
74. International Conference on Harmonisation (2000), *Non-Clinical Safety Studies for the Conduct of Human Clinical Trials for Pharmaceuticals*; available at: http://www.emea.europa.eu/pdfs/human/ich/028695en.pdf.
75. International Conference on Harmonisation (1998), *Preclinical Safety Evaluation of Biotechnology-Derived Pharmaceuticals*; available at: http://www.emea.europa.eu/pdfs/human/ich/030295en.pdf.
76. U. S. Department of Health and Human Services Food and Drug Administration Center for Drug Evaluation and Research (2005), *Guidance for Industry: Estimating the Maximum*

*Safe Starting Dose in Initial Clinical Trials for Therapeutics in Adult Healthy Volunteers*; available at: http://www.fda.gov/cder/guidance/5541fnl.pdf.

77. Reigner, B. G., and Blesch, K. S. (2002), Estimating the starting dose for entry into humans: Principles and practice. *Eur. J. Clin. Pharmacol.*, 57, 835–845.
78. Lowe, P. J., Hijazi, Y., Luttringer, O., et al. (2007), On the anticipation of the human dose in first-in-man trials from preclinical or prior clinical information in early drug development, *Xenobiotica*, 37, 1331–1354.
79. Feng, M. R., Lou, X., Brown, R., et al. (2000), Allometric pharmacokinetic scaling: Towards the prediction of human oral pharmacokinetics. *Pharma. Res.*, 17, 410–418.
80. Kawai, R., Mathew, D., Tanaka, C., et al. (1998), Physiologically based pharmaockinetics of cyclosporine A: Extension to tissue distribution kinetics in rats and scale-up to human, *J. Pharmacol. Exper. Therapeut.*, 287, 457–468.
81. Dixon, W., and Mood, A. (1948), A method for obtaining and analyzing sensitivity data, *J. Am. Statist. Assoc.*, 43, 109–126.
82. Storer, B. (1989), Design and analysis of Phase I clinical trials, *Biometrics*, 45, 925–937.
83. O'Quigley, J., Pepe, M., and Fisher, L. (1990), Continual reassessment method: A practical design for Phase I clinical trials in cancer, *Biometrics*, 46, 33–48.
84. Neuenschwander, B., Branson, M., and Gsponer, T. (2008), Critical aspects of the Bayesian approach to phase I cancer trials, *Stat. Med.*, 27, 2420–2439.
85. Garrett-Mayer, E. (2005), *Understanding the Continual Reassessment Method for Dose Finding Studies: An Overview for Non-Statisticians*, Johns Hopkins University, Dept. of Biostatistics Working Papers, Baltimore, MD, Paper 74.
86. Thall, P. F., and Cook, J. D. (2004), Dose finding based on efficacy toxicity trade offs, *Biometrics*, 60, 684–693.
87. Whitehead, J., Zhou, Y., Stevens, J., et al. (2004), An evaluation of a Bayesian method of dose escalation based on bivariate binary responses, *J. Biopharma. Stat.*, 14, 969–983.
88. Pinheiro, J., Bretz, F., and Branson, M. (2006), Analysis of dose response studies: Modeling approaches, in Ting, N., Ed., *Dose Finding in Drug Development*, Springer, New York, pp. 146–171.
89. Bates, D. M., and Watts, D. G. (1988), *Nonlinear Regression Analysis and Its Applications*, Wiley, New York.
90. Hochberg, Y., and Tamhane, A. C. (1987), *Multiple Comparisons Procedures*, Wiley, New York.
91. Hsu, J. C. (1996), *Multiple Comparisons*, Chapman and Hall, New York.
92. Bretz, F., Pinheiro, J., and Branson, M. (2005), Combining multiple comparisons and modeling techniques in dose response studies, *Biometrics*, 61, 738–748.
93. Gallo, P., Chuang-Stein, C., Dragalin, V., et al. (2006), Adaptive designs in clinical drug development—an executive summary of the PhRMA working group, *J. Biopharma. Stat.*, 16, 275–283.
94. Ette, E. I., and Williams, P. J. (2007), *Pharmacometrics: The Science of Quantitative Pharmacology*, Wiley, Hoboken, NJ.

# *Monitoring*

**Nigel Stallard[1] and Susan Todd[2]**

[1]*Warwick Medical School, University of Warwick, Warwick, United Kingdom*
[2]*Applied Statistics, University of Reading, Reading, United Kingdom*

**Contents**

1 Reasons for Monitoring Clinical Trial Data
2 Monitoring of Clinical Trial Data without Unblinding Treatment Allocations
  2.1 Administrative and Safety Data Monitoring
  2.2 Monitoring for Estimation of Nuisance Parameters and Sample-Size Reviews
3 Dangers of Unplanned Clinical Trial Monitoring
4 Methodology for Sequential Monitoring of Unblinded Efficacy Data
  4.1 Parameterization of Treatment Difference and Test Statistics
  4.2 Repeated Significance Testing
  4.3 Spending Function Approach
  4.4 Boundaries Approach
  4.5 Analysis after a Sequential Trial
  4.6 Monitoring and Analysis of More Than One Endpoint
5 Implementation of Sequential Designs
6 Example: Lithium Gamolenate in Patients with Advanced Pancreatic Adenocarcinoma
7 Response-Driven and Adaptive Designs
References

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

## 1 REASONS FOR MONITORING CLINICAL TRIAL DATA

Many traditional methods for the analysis of data assume that the data are all available contemporaneously. This reflects the setting of agricultural experimentation in which much of the methodology was developed, as data on crop yields from different fields are collected simultaneously after harvesting. In the majority of clinical trials, however, patients are recruited over a period of time, so that data accumulate gradually through the course of the trial, with data from some patients available prior to the recruitment of others. This accumulation of data makes possible the statistical analysis of some of the data collected before the trial is completed. Such analysis of data during the course of the trial is termed *monitoring*.

There are a number reasons why monitoring of clinical trial data might be considered desirable. The first is an ethical reason. In a randomized trial, randomization of subjects to receive one treatment or another is generally considered ethically acceptable provided there is *equipoise*, that is, an uncertainty as to which treatment is superior. If clear evidence was available that one treatment was of superior efficacy to the other, or that one treatment was less safe than the other, continued randomization would be considered unethical. In a single-arm trial, enrollment in the trial would be similarly considered unethical if the treatment being assessed clearly led to an unacceptable level of adverse events. Monitoring of efficacy and safety data allows the results of an ongoing trial to be regularly assessed and enables termination of the trial as soon as the results are sufficiently convincing that continued recruitment cannot be ethically defended.

The second reason for monitoring a clinical trial is administrative. As the trial progresses, it is desirable to ensure protocol compliance, check data collection procedures, and to monitor recruitment rates. The estimation of nuisance parameters (such as the unknown variance for normally distributed responses) used in power calculations can also be conducted to ensure that the trial will have a sufficient sample size before the close of recruitment.

The third reason for monitoring a clinical trial is economic. Clinical trials are expensive, and it is not in the interest of sponsors or patients to continue a trial once the results are clear. By monitoring the accumulating data, a clinical trial can be stopped as soon as sufficient data are available to justify a conclusion that an experimental treatment is superior or inferior to the control, or that the data clearly indicate that any difference is so small that continuation of the trial is futile.

In a blinded randomized trial we may consider data monitoring to be divided into three types: first, administrative monitoring of clinical trial conduct and monitoring of safety data with no monitoring of efficacy data; second, monitoring of efficacy data without unblinding of treatment allocation, for example, for the estimation of nuisance parameters; third, monitoring of efficacy data with treatment allocation unblinded to allow the estimation of the difference in efficacy between the treatments being compared. In the latter two instances, administrative and safety monitoring will most likely be conducted in addition. As will be discussed in Section 3, it is the third of these types of monitoring that presents the most ethical and statistical challenges. It is therefore on this type of monitoring that the majority of the chapter will be focused. Before moving on, however, the first two types of monitoring will be discussed briefly in the next section.

## 2 MONITORING OF CLINICAL TRIAL DATA WITHOUT UNBLINDING TREATMENT ALLOCATIONS

### 2.1 Administrative and Safety Data Monitoring

Monitoring of administrative data such as recruitment rates, accuracy of completion of record forms, treatment compliance, and data collection/transfer procedures is common in large clinical trials. The aim of such monitoring is to ensure that any problems in trial conduct or data collection are identified and rectified as soon as possible. Identification of incomplete record forms from a particular investigator could lead, for example, to collection of the missing data prior to the end of the trial, while observation of a lower than anticipated recruitment rate might justify the introduction of new centers in a multicenter trial. An example of a form to assess ongoing treatment compliance as part of a trial is given in Figure 1. Administrative monitoring of this kind is generally conducted without unblinding treatment allocation since treatment information is generally irrelevant. The risks of increased error rates or release of information about the performance of the study drugs that could lead to bias as discussed in Section 3 therefore do not arise from this sort of monitoring.

Many large trials also involve some monitoring of safety data. As discussed above, in a trial in which serious adverse events may be observed, this type of monitoring serves an important ethical function. In order to assess whether there is an excess number of adverse events in one treatment group relative to the other, it is necessary to unblind treatment allocation. In some cases, the two treatment groups are presented simply as groups A and B, without indicating which group is receiving the experimental treatment and which the control. However, it is generally easier to interpret adverse events in the light of the treatment administered, so that full unblinding is considered to be a better practice [1]. The unblinding of treatments introduces the risk of bias as discussed in Section 3. In order to minimize this risk, particularly in large clinical trials, the safety data may be reviewed by a *data and safety monitoring board* (DSMB) (sometimes alternatively called a *data monitoring committee*), so that only members of this committee need have access to the unblinded data. This committee decides on the basis of the unblinded data reviewed whether or not it is safe for the trial to continue and makes its recommendation to the *trial steering committee* responsible for the overall conduct of the trial. Further details of the role and function of DSMBs are given by Ellenberg et al. [2] and Whitehead [3]. Two example forms for monitoring adverse events are given in Figures 2 and 3. Figure 2 illustrates the categorizations to be used when determining adverse events and their attributability. Figure 3 shows how a case report form for serious adverse events might be structured.

As the purpose of safety monitoring is solely to recommend termination of the clinical trial if an experimental treatment is associated with too many adverse events, this type of monitoring cannot lead to a positive claim of effectiveness for the experimental treatment based on the monitored data. This means that, in contrast to the type of efficacy monitoring discussed in Section 3, there is no risk of a false-positive conclusion arising as a result of this monitoring; that is, the type I error rate cannot be increased. The special statistical methodology that is discussed in the rest of this chapter is therefore not required in this case.

| Treatment Period | PATIENT STUDY NUMBER ☐☐☐☐ |
|---|---|
| | PATIENT INITIALS ☐☐☐ |

**PHARMACY DRUG DISPENSING RECORD**

**ISSUES**

Date study medication dispensed ☐☐☐

Number of capsules dispensed ☐☐☐

or Number of ampoules dispensed ☐☐☐

Signature of dispensing pharmacist _____

**RETURNS**

Date study medication returned ☐☐☐

Number of capsules returned ☐☐☐

or Number of unopened ampoules returned ☐☐☐

Signature of receiving pharmacist _____

**VERIFICATION** *(to be completed by monitor)*

Date of verification ☐☐☐

Number of capsules checked ☐☐☐

or Number of unopened ampoules ☐☐☐

Signature of monitor _____

***FIGURE 1*** *Example form for monitoring treatment compliance.*

## 2.2 Monitoring for Estimation of Nuisance Parameters and Sample-Size Reviews

Sample size calculations at the planning stage of a clinical trial often require the specification of values for unknown parameters other than the treatment effect of primary interest. These are often known as *nuisance parameters*. For example, if data are assumed to be normally distributed, the primary measure of relative treatment

**GUIDELINES FOR RECORDING OF ADVERSE EVENTS**

| | | |
|---|---|---|
| Severity | 1 | Mild |
| | 2 | Moderate |
| | 3 | Severe |
| | 4 | Serious – notify the company immediately by telephone. Complete and return Serious Adverse Event form, which can be found in the Investigator File |
| Relation to Treatment | 1 | Probably |
| | 2 | Possibly |
| | 3 | Probably not |
| | 2 | Temporary interruption |
| | 3 | Permanent discontinuation |
| | 4 | Drug therapy changed or added – note changes on medication page |
| | 5 | Other therapy – give details below |
| | 6 | Hospitalization or prolonged hospitalization – notify the company immediately by telephone. Complete and return Serious Adverse Event form, which can be found in the Investigator File |
| Outcome | 1 | Complete recovery |
| | 2 | Recovery with sequelae |

***FIGURE 2*** *Example categorizations for adverse event monitoring.*

efficacy will often be the difference in the mean response for the two treatments. Specification of the variance of the response, which is commonly assumed to be the same in the two groups, is, however, needed for calculation of the required sample size. A second example is for binary data, when the sample-size calculation depends on the average probability of success for the two groups. In practice, in such cases, the guess of the nuisance parameter required before the start of the trial may be inaccurate. This might lead to an inappropriate choice of sample size and to an underpowered or overpowered trial.

**1. PATIENT AND REACTION INFORMATION**

| Patient Initials | Study Code<br>Patient No. | Country | Date of Birth<br>Day Mth Yr | Age | Sex | Reaction onset<br>Day Mth Yr | Tick all appropriate to adverse reaction |
|---|---|---|---|---|---|---|---|
| DESCRIBE REACTION(S)<br>Give signs or symptoms, diagnoses, course (underlying the main event)<br>(Include relevant lab. data) | | | | | | | ☐ Patient died<br>☐ Involved or prolonged inpatient hospitalization<br>☐ Involved significant disability or incapacity<br>☐ Life threatening<br>☐ None of the above |

**2. SUSPECT DRUG INFORMATION**

| SUSPECT DRUG (include all information available: generic name, form and strength, dosage, batch number). | | DID REACTION ABATE AFTER STOPPING DRUG?<br>Yes ☐ No ☐ NA* ☐ |
|---|---|---|
| DAILY DOSE (with unit) | ROUTE OF ADMINISTRATION | DID REACTION REAPPEAR AFTER REINTRODUCTION? |
| INDICATION FOR USE | | |

***FIGURE 3*** *Example form for monitoring serious adverse events.*

Gould and Shi [4] proposed, as a solution to this problem, the use of an analysis of the accumulating data part way through the trial to estimate the unknown nuisance parameter in what is usually termed a *sample-size review*. A review of the method and further advances in the methodology is given by Gould [5]. The required sample size for the trial can then be obtained based on this estimate of the nuisance parameter, and the sample size for the remainder of the trial adjusted accordingly. For example, suppose a trial with a normally distributed response was initially designed with a sample size of 100 patients. An analysis could be conducted when data were available from 50 patients to calculate the unknown common variance. At this point it might be revealed that the estimated variance based on the data was larger than the value anticipated when the trial was planned, so that the trial is likely to be underpowered. Based on the estimate from the data, the required sample size might be 120 patients rather than the initial 100 planned. The second stage of the trial could then recruit a total of 70 rather than 50 patients to give this total sample size. Simulation studies have shown that such an approach is effective in maintaining the power even if the initial guess of the unknown nuisance parameter is inaccurate.

In the case of normally distributed data, as in the illustration above, the nuisance parameter that is estimated at the sample size review is the common variance. The usual pooled variance estimate is given by

$$s^2 = \frac{1}{n_{E1} + n_{C1} - 2}\left[\sum_{j=1}^{n_{E1}} (x_{Ej} - \bar{x}_{E1})^2 + \sum_{j=1}^{n_{C1}} (x_{Cj} - \bar{x}_{C1})^2\right] \tag{1}$$

where $x_{E1}, \ldots, x_{En_{E1}}$, with mean $\bar{x}_{E1}$, are the $n_{E1}$ observations on the experimental arm and $x_{C1}, \ldots, x_{Cn_{C1}}$, with mean $\bar{x}_{C1}$, are the $n_{C1}$ observations on the placebo arm. The calculation of this estimate requires a knowledge of which patients are in which treatment group, so that the treatment allocation must be unblinded. As explained below, this is undesirable in clinical trial monitoring as it may lead to bias through the differential treatment and interpretation of further patients in the trial.

Expression (1) for $s^2$ can be rearranged to give

$$s^2 = \frac{1}{n_{E1} + n_{C1} - 2}\left[(n_{E1} + n_{C1} - 1)s_0^2 - \frac{n_{E1}n_{C1}}{n_{E1} + n_{C1}}(\bar{x}_{E1} - \bar{x}_{C1})^2\right] \tag{2}$$

where $s_0^2$ is the overall estimate of variance, which can be calculated without breaking the blind, given by

$$s_0^2 = \frac{1}{n_{E1} + n_{C1} - 1}\left[\sum_{j=1}^{n_{E1}} (x_{Ej} - \bar{x})^2 + \sum_{j=1}^{n_{C1}} (x_{Cj} - \bar{x})^2\right] \tag{3}$$

with $\bar{x}$ the overall mean.

This suggests a method for estimation of the standard deviation from blinded data. The first term in (2), depending on $s_0^2$, can be found from the unblinded data. An estimate of the standard deviation can thus be obtained from blinded data together with an estimate of $(\bar{x}_{E1} - \bar{x}_{C1})$, the treatment difference. Gould [5] suggests

using the treatment difference under which the power is calculated to give an estimate of the standard deviation on which a new sample size can be calculated. If we denote this difference by $\delta_R$, replacing $(\bar{x}_{E1} - \bar{x}_{C1})$ in (2) by $\delta_R$ gives an estimate of the variance of

$$\frac{1}{n_{E1} + n_{C1} - 2}\left[(n_{E1} + n_{C1} - 1)s_0^2 - \frac{n_{E1}n_{C1}}{n_{E1} + n_{C1}}\delta_R^2\right]. \tag{4}$$

If an estimate of the unknown variance is based on blinded data, the sample size reestimation procedure does not have any impact on the overall type I error rate, in contrast to the more general sequential monitoring that is discussed in the next section. The type I error rate may be inflated, however, if unblinded data are used to provide the variance estimate used for the calculation of the sample size for the second stage [6]. This gives further reason for the use of methods that do not require knowledge of treatment allocations.

## 3 DANGERS OF UNPLANNED CLINICAL TRIAL MONITORING

Although, for the reasons just described, monitoring of accumulating data in a clinical trial is very appealing, there are a number of dangers, particularly when considering treatment efficacy, to which careful attention should be given whenever data monitoring is considered.

The main danger is a statistical one: The conduct of multiple analyses can lead to difficulty in the interpretation of the results [7]. In particular, if significance tests comparing experimental and control treatments are performed more than once, with testing stopped as soon as a significant result is obtained, the type I error rate is inflated as explained below.

When a single significance test is conducted, it is performed in such a way that the probability of declaring the treatments to be different if in fact they are identical, that is, of making a type I error, is controlled, usually to be 5%. If the test is repeated at a number of analyses, with each test conducted at the 5% level and the procedure stopped as soon as a significant result is obtained, there is risk of a type I error at each of these analyses. The overall probability of such an error, the type I error rate, will thus be more than 5%. The effect of repeated looks conducted at the 5% level in the simplified setting of comparison of normal data with known variance was assessed by Armitage et al. [8]. They showed that in this setting while a single analysis has a type I error rate of 5% as planned, if two analyses are conducted, the overall type I error rate is increased to 8%, with this increasing to 14% if five analyses are conducted, and increasing further if the number of analyses is more than five.

The problem of type I error rate inflation means that monitoring of clinical trials cannot be conducted without the use of specialist statistical methodology to ensure the type I error rate is controlled at the nominal level. The remainder of this chapter gives descriptions and examples of such statistical methods, explaining how valid tests may be obtained. A clinical trial conducted using such methodology is called a *sequential clinical trial*, and the formal analysis of data part way through the trial is known as an *interim analysis*.

A further statistical problem, related to that of type I error inflation, is that of estimation at the end of a monitored clinical trial. Just as the type I error rate is inflated because the trial stops exactly when extreme results are obtained, so a conventional estimate obtained at the end of a sequential clinical trial will be biased because the trial stops exactly when positive results are seen at an interim analysis. Again, special statistical techniques are required to provide a valid analysis, as described below.

In addition to the statistical problems associated with the conduct of interim analyses, there are some practical problems that may make monitoring of efficacy data difficult or undesirable. In some trials the primary endpoint may be some measure, such as long-term survival, that is observed only after considerable follow-up. In this case, all patients may be treated before the primary endpoint data from even the first patients enrolled in the trial are observed. This means there is no possibility of stopping the trial as a result of interim analyses. Although there may be some advantage in monitoring the efficacy data even in this case, for example, to obtain the results of the trial more quickly, there is no potential saving in terms of trial length or sample size.

If the data observed early in a trial are not considered typical of all of the data that are likely to be observed, the results of an interim analysis may be misleading. Monitoring of efficacy data in this case is thus undesirable and should be avoided. An example might be a comparison of treatments for cancer when the primary efficacy endpoint is survival, and one treatment is expected to lead to excess early mortality, but to better long-term prognosis. If only early death data were available at an interim analysis, the long-term advantages of this treatment could be missed, and the trial halted erroneously.

The final danger associated with many data-monitoring strategies is the need to unblind the treatment allocation. If treated patients continue to be followed-up, there is a danger that knowing which treatment they are receiving may lead to a difference in the level of care for or interpretation of results from patients in the experimental treatment and control groups. This may give rise to bias in the treatment comparison. The release of interim results may also lead to reluctance among investigators to recruit patients, or reluctance among the patients themselves to consent to randomization, if there is some evidence of a treatment difference, so that equipoise is no longer considered to exist.

## 4 METHODOLOGY FOR SEQUENTIAL MONITORING OF UNBLINDED EFFICACY DATA

As described above, some clinical trial monitoring, such as monitoring of adverse events for safety, or the estimation of nuisance parameters from blinded data for sample-size review may be conducted with no risk of type I error rate inflation or effect on the final analysis. The monitoring of unblinded efficacy data, with the chance to stop the trial early if a large difference between the treatments is observed, does, however, lead to an increase in the type I error and bias in the analysis unless specific statistical methods are employed. The remainder of this chapter describes such methods and illustrates how valid monitoring of efficacy data can be conducted. Most of the statistical methodology in this area has been developed for clinical trials

in which two treatment groups, often corresponding to an experimental treatment and a control treatment, are compared. It is on this type of trial that the rest of the chapter will be specifically focused.

Whitehead [9] gives a list of the four key elements that must be considered in the design and conduct of a sequential clinical trial to assess treatment difference. These are a parameter giving some measure of the advantage of the experimental treatment over the control treatment; test statistics that give information on the size of this advantage on the basis of the data observed at an interim analysis and the amount of information about the treatment difference that is given by the observed data; a stopping rule that determines, on the basis of the observed test statistic, values at an interim analysis whether the trial should continue or should stop; and a final analysis method, valid for the stopping rule used, that enables a conclusion of whether or not the experimental treatment is superior to the control and provides a $p$ value and point estimate and confidence interval for the treatment difference at the end of the trial.

The four elements are now considered in detail. Specification of the first two, the treatment difference parameter and the test statistics, is required whether fixed sample size or sequential methods are employed, but is discussed here for completeness. The specification of the third, the stopping boundary, represents a solution to the problem of control of the type I error rate. A number of solutions to this problem have been proposed, leading to several different approaches to the way in which the stopping boundary is calculated. These are described below. Although the fourth element, the analysis, is very important in the interpretation of the results from the sequential trial, statistical methodology in this area has generally lagged behind that for the construction of stopping rules. In practice, the use of an appropriate final analysis has often been neglected, in part due to the lack of availability of suitable software. Methods for the valid analysis have been developed, however, and the analysis can now be conducted using commercially available software. The methods are described briefly below and the alternative computer programs are discussed in the next section.

### 4.1 Parameterization of Treatment Difference and Test Statistics

As in a fixed sample-size trial, the first stage in the design of a sequential clinical trial is the specification of some primary measure of treatment efficacy. This measure should be chosen on the basis of clinical relevance, ease and accuracy of measurement, and familiarity to clinicians. Following the definition of the primary endpoint, an associated parameter measuring the difference between the experimental and control treatments must be chosen. This will depend on the type of data collected on the primary endpoint, the interpretability of the parameter, for example, whether a difference or a ratio is more familiar, and the precision of the resulting analysis.

For example, if the primary response is a dichotomous variable such as whether a cancer patient is in remission or not after one year, the chance of being in remission in each group can be measured by the odds of remission, and the difference between the treatment groups can be measured by the ratio of the odds in the two groups, or more commonly the logarithm of this ratio, the *log-odds ratio*. If the primary response is a continuous measure such as weight loss, a difference in true unknown mean weight loss for the two groups would be of interest. If the primary

response is the time to some event, such as death, the difference between treatment groups might be measured by the ratio of the hazards in the two groups or the logarithm of this ratio.

At each interim analysis, some test statistic measuring the difference between the experimental and control treatments based on the observed data is calculated together with a test statistic measuring the amount of information given by the observed data at that interim analysis. Whitehead [10] and Jennison and Turnbull [11] suggest using a measure of difference known as the *efficient score* statistic and a measure of information called the *observed Fisher's information*, though other pairs of statistics may also be used, such as an estimate of the treatment effect and its standard error as suggested by Jennison and Turnbull [12] or a $p$ value for comparison of the treatments and the planned sample size, as is used in *repeated significance testing* described in the next paragraph and in the *adaptive design* approach described in more detail in the last section of this chapter.

### 4.2 Repeated Significance Testing

As discussed above, if a simple significance test comparing the experimental and control treatment groups is conducted repeatedly at the 5% level until a significant result is obtained, the overall type I error rate will be increased above 5%. Armitage et al. [8] showed how the actual error rate could be calculated when the number of repeated tests is specified. Similar calculations to those performed by Armitage et al. can be used to enable repeated significance testing with overall type I error rate controlled at the 5% level. By conducting each individual test at a more stringent level, the overall type I error rate is reduced, and so may be reduced to be 5% as desired. Numerical computation may be used to calculate the required error rate for each of the repeated tests to ensure that the overall level is 5%.

Pocock [13] proposed such a repeated significance testing approach, giving tables of the nominal type I level at which each test should be conducted to give an overall type I error rate of 5 or 1% for different numbers of tests. For example, to give an overall type I error rate of 5%, if two tests are performed each should be carried out at the 2.94% level, whereas if five tests are conducted, the nominal level for each should be 1.58%.

The repeated significance tests described by Pocock [13] are assumed to all be conducted at the same nominal level so that the same critical value is used for a standardized test statistic at each of the interim analyses. In general, there is no requirement that this should be so. If, say, five interim analyses are planned, it is possible to use different nominal levels for the five significance tests and control the overall type I error rate to be 5%. An alternative pattern of significance levels was proposed by O'Brien and Fleming [14]. They constructed the repeated significance tests in such a way that the level of evidence required to reject the null hypothesis that there is no difference between the two treatments at the earlier analyses is greater than that required to reject it at later analyses. In particular, the critical value for the $i$th interim analysis is chosen to be $k_1/\sqrt{i}$ for some constant $k_1$ chosen to maintain the overall type I error rate at the desired level. This choice of nominal levels, while maintaining the overall type I error rate, makes stopping at early interim analyses almost impossible, negating some of the advantages of sequential analysis. The Pocock [13] design, on the contrary, has been criticized, for example,

by Pocock and White [15], for leading to too high a chance of early stopping. Alternative approaches have been proposed [16], although the repeated significance testing approach has been largely superceded by the more flexible spending function approach described in the next section.

### 4.3 Spending Function Approach

A disadvantage of the repeated significance testing approach is that the number of interim analyses and the timing of these in terms of the amount of information available at each analysis need to be specified in advance in order to calculate the critical values required to control the overall error rate. Lan and DeMets [17] introduced a more flexible approach in which neither the timing nor the number of interim analyses need be given at the design stage.

Rather than specifying the critical values at a number of interim analyses, Lan and DeMets [17] proposed that the stopping rule should be given by a *spending function* that describes how the overall type I error rate, say 5%, is spent through the course of the trial. At the design stage, the maximum amount of information that will be attained if the trial does not stop early is calculated. At each interim analysis the level of information currently available is evaluated and the *information time* calculated to be the proportion of the maximum information available from the current data. The spending function gives, as a function of this information time, the amount of type I error that is to be spent, and the hypothesis test is conducted at such a level to ensure that the overall type I error spent up to and including this interim analysis is equal to this amount.

The spending function is thus an increasing function of the information time. It takes value 0 for information time 0, that is, at the start of the trial, and value equal to the overall type I error rate for information time 1, that is, at the end of the trial. It is common to denote the information time by $t$, and the spending function by $\alpha^*(t)$. Suppose that interim analyses are taken at information times $t_1, t_2, \ldots$. The critical values used for the hypothesis tests at each of the interim analyses are chosen so that the probability of stopping and rejecting the null hypothesis at any of the interim analyses up to and including the $i$th is equal to $\alpha^*(t_i)$ if the null hypothesis is really true. Equivalently, if the null hypothesis is true, the probability of stopping and rejecting the null hypothesis at the $i$th interim analysis given that the trial has not stopped at an earlier interim analysis is equal to $\alpha^*(t_i) - \alpha^*(t_{i-1})$. The critical values to give these stopping probabilities can be calculated using a recursive numerical integration technique similar to that used by Armitage et al. [8] to evaluate error rates for repeated significance tests. A similar technique enables calculation of the probability of stopping under a specified alternative hypothesis, and this allows the maximum information level to be chosen to give a test that satisfies some power requirement. Further details of the spending function approach and the computations required to calculate the critical values and sample size are given by Jennison and Turnbull [11].

As in the repeated significance testing approach, there is considerable flexibility regarding how likely the trial is to stop at the different interim analyses within the constraint that the overall type I error rate is controlled at the desired level. In the spending function approach this flexibility is expressed in the choice of the spending function, which may be any increasing function rising from 0 to the overall type I

error rate. A number of forms for spending functions have been proposed. Lan and DeMets [17] give spending functions to give approximations to the repeated significance testing procedures of Pocock [13] and O'Brien and Fleming [14], while Kim and DeMets [18] and Hwang et al. [19] suggest possible families of spending functions.

The flexibility of the sequential stopping boundaries in terms of the form of the spending function was utilized further by Kim and DeMets [18]. They showed how two different spending functions may be used to construct tests that have different properties for positive and negative treatment differences.

The sequential approach can be extended further to give tests that have different power to detect positive and negative treatment differences. A lower power is associated with a smaller sample size. Since the sample size is no longer fixed, but is now a random variable depending on the observed data, it is possible to stop the trial earlier for certain observed treatment differences than others. In particular, the stopping rule may not be symmetric, so that different decisions may be taken for positive and negative observed differences of the same magnitude. This is valuable, for example, in a clinical trial comparing an experimental treatment with a placebo. If the experimental treatment is superior to the placebo, we desire a high probability of rejecting the null hypothesis that there is no difference, and we are willing to randomize the large number of patients required to demonstrate this. If the experimental treatment is actually inferior to the placebo, demonstrating this is of less interest. It is sufficient to show that the experimental treatment is either the same or worse than the placebo and to continue to randomize patients to discover whether the disadvantage associated with the experimental treatment is significant or not is undesirable. Whitehead [10] calls asymmetric tests of this type tests with *power requirement I*, in contrast to the more usual symmetric tests, which he calls tests with *power requirement II*. Sequential procedures with power requirement type I can also be constructed using the spending function approach [20, 21].

### 4.4 Boundaries Approach

The repeated significance testing approach and the spending function approach provide a stopping rule. In each case, for each interim analysis conducted, a pair of critical values is obtained, and the trial is stopped if the test statistic lies outside of the interval defined by these two values. The critical values are said to form a *stopping boundary* and the set of intervals to form a *continuation region*. The observed values of the test statistic measuring the treatment difference at the interim analyses are said to form a *sample path*, which is compared with the stopping boundary to decide whether or not the trial should continue.

An alternative to the methods for construction of the stopping boundary described above is based on the abstract concept of continuous monitoring, in which it is imagined that the value of the test is observed at all times rather than just at the discrete times given by the interim analyses. In this case, a plot of the test statistic measuring the treatment difference against that measuring the amount of information available forms a continuous line. In the *boundaries approach*, this continuous line is compared with continuous stopping boundaries that are expressed as functions of the information level.

The boundaries approach stems from the work of Wald [22] who developed the sequential probability ratio test (SPRT), initially for the testing of armaments during World War II. In the SPRT the likelihood ratio for a simple alternative hypothesis relative to the null hypothesis is calculated at each interim analysis. The test continues so long as this likelihood ratio falls within some fixed range. Wald derived stopping limits so as to give a test with specified type I error rate and power under the assumption of continuous monitoring. The SPRT has the important optimality property that among all tests with the same type I error rate and power, the SPRT minimizes the expected sample size when either the null or alternative hypothesis holds. Although the expected sample size is minimized by the SPRT, however, the SPRT has no maximum sample size, so that it cannot be guaranteed that the sample size will be below any fixed value. This feature makes the SPRT unsuitable for many clinical trials.

Following the work of Wald [22], a number of alternative forms for boundaries that maintain the overall type I error rate have been proposed. The choice of the form of boundary in this approach is analogous to the choice of spending function in the approach described above, and, as above, trials may be designed with asymmetric (power requirement I) or symmetric (power requirement II) boundaries. Whitehead [10] describes a wide range of such tests, including approximations to the SPRT that do place a limit on the maximum sample size required. One family of boundary shapes that is particularly common in sequential clinical trials is based around the triangular test. If the efficient score statistic is plotted against the observed Fisher information, this test has straight boundaries that form a triangular-shaped continuation region. An example of a double triangular test is illustrated in the example in Section 6. The triangular test approximately minimizes the maximum expected sample size among all tests with the same error rates. Unlike the SPRT the triangular test has a maximum sample size. It can be shown that for a triangular test there is a high probability of stopping with a sample size below that of the equivalent fixed sample size test, so that on average the sample size will be reduced by the use of this sequential design.

The stopping boundaries obtained using the boundaries approach maintain the overall type I error rate under the assumption of continuous monitoring of the test statistic measuring the treatment difference. In practice, monitoring is necessarily discrete, even if an interim analysis is conducted after observation of every patient. This means that if the critical values from the boundaries approach are used, the type I error rate will be less than the planned level. Whitehead [10] has proposed a correction, known as the *Christmas tree correction*, to modify the continuous boundaries to allow for the discretely monitored sample path. This correction allows interim analyses to be made at any time provided that the timing is not dependent on the observed treatment difference.

### 4.5 Analysis after a Sequential Trial

At the end of a sequential trial a final analysis of the data observed should be conducted. This analysis should provide a $p$ value for testing the null hypothesis that there is no difference between the experimental treatment and the control, and a point estimate and confidence interval for the difference between the treatments,

usually as measured by the primary parameter of treatment difference discussed above.

The fact that a sequential stopping rule has been used means that many of the standard analysis methods that would be employed after a fixed sample sized trial are no longer valid. Facey and Lewis [23] reported in 1998 that a standard analysis is performed inappropriately in many sequential trials. This is hopefully now changing.

Suppose that a sequential trial stopped at some interim analysis with the upper critical value exceeded by the test statistic measuring observed treatment difference, leading to the conclusion that the experimental treatment was superior to the control treatment. The trial has stopped precisely because of the large observed value of the treatment difference. This means that standard unbiased estimates of treatment difference will, on average, overestimate the true value of the treatment difference. In a similar way, the $p$ value from a standard analysis will, on average, be too small, that is, it will tend to overstate the evidence against the null hypothesis. Special methods of analysis are therefore required to allow for the interim analyses that have been conducted in a sequential trial.

The meaning and interpretation of the $p$ values, point estimates, and confidence intervals obtained after a sequential clinical trial are the same as for a fixed sample size trial, but a method must be employed in their calculation that is valid for the sequential stopping rule used. A number of methods of analysis have been proposed. Most are based on an ordering of all possible data sets to reflect which are considered to provide strongest evidence against the null hypothesis being tested. Further details are given by Whitehead [10] and Jennison and Turnbull [11]. The implementation of these methods in commercially available software, as described in more detail in Section 5, enables the easy conduct of a valid analysis following a sequential trial.

### 4.6 Monitoring and Analysis of More Than One Endpoint

In a clinical trial it may be the case that other endpoints in addition to the primary endpoint are important in the clinical assessment of a treatment. If these endpoints are only to be analyzed at the end of the study after complete data on all patients are available, then they are usually termed *secondary endpoints*. No formal monitoring of these is conducted during the trial. If, however, one or more of the endpoints can really be regarded as additional primary endpoints that are important in determining whether a study should stop or continue, then further statistical methodology is required. There are two general approaches to monitoring these *co-primary endpoints* depending on the nature of the endpoints involved.

If several endpoints are measuring a similar aspect of treatment performance, then it may be acceptable to demonstrate efficacy in a collective way across the endpoints. Combining the various endpoints together into a single global measure and then using methods already described in this chapter would then be a suitable approach. As an example, consider the therapeutic area of stroke. There is no single endpoint that is universally accepted. Outcomes of interest include the Barthel index, the modified Rankin score, the Glasgow outcome scale, and the National Institutes of Health (NIH) stroke scale. Investigators do not believe that a positive result in terms of a single one of those outcomes is sufficient. Tilley et al. [24]

describe using a global test statistic in precisely this indication and discuss the advantages and disadvantages of such an approach.

In some indications, with some endpoints, it is not clinically meaningful to combine the endpoints into a single measure. Examples include the need to demonstrate efficacy for two outcomes to gain licensing, such as in Alzheimer's disease where it is generally necessary to consider both mental and physical aspects of a patient's progress. The simultaneous monitoring of efficacy and safety responses would be another example. In these instances, the endpoints should be kept separate and the correlation between them accounted for in the design and analysis of any sequential trial. Due to the complexity of computations, most authors have concentrated on developing *bivariate* methodology, that is, for the case of two co-primary endpoints, see, for example, Jennison and Turnbull [25], Cook and Farewell [26], and Todd [27].

## 5 IMPLEMENTATION OF SEQUENTIAL DESIGNS

As illustrated in Section 4, there are a number of different approaches to the specification of a sequential design and in particular the form of a stopping rule. When considering the implementation of such methodology, the most important issues to focus upon are the reasons why it is desirable to either stop or continue the trial at an interim analysis. Reasons for stopping were outlined above and may include evidence that the experimental treatment is obviously better than the control, evidence that the experimental treatment is already obviously worse, or indication that there is little chance of showing that the experimental treatment is better than the control so that it is considered futile to continue. Reasons for continuing may include belief that a moderate advantage of the experimental treatment is likely and it is desired to estimate this magnitude carefully, or evidence that the event rate is low and more patients are needed to achieve required power. Criteria such as these will first determine the symmetry requirement that is appropriate for the study under consideration, that is, whether an asymmetric (power requirement I) or a symmetric (power requirement II) design should be considered. Once this has been established, attention turns to the specific properties of different tests within a broad category. A choice between the tests is then made based upon consideration of likely sample sizes under various possible treatment effect scenarios. The wide range of stopping rules now available allow sequential designs to be devised for testing *superiority*, *noninferiority*, *equivalence*, and even formal *safety* aspects of clinical trials. As discussed above, some more complex designs even aim to deal with both efficacy *and* safety aspects in a combined stopping rule.

The availability of appropriate designs is just one element of enabling the implementation of sequential methodology. The key requirement for widespread take-up of the techniques is access to software. There are four major commercial software packages currently available. The package PEST [28] is based on the boundaries approach. The package EaSt [29] implements the spending function approach. An addition to the package S-Plus is the S+SeqTrial module [30], which also implements the spending function approach, in this case for a family of designs based on the methodology described by Kittelson and Emerson [31]. The package AddPlan (see www.addplan.com) is based on the adaptive design approach mentioned in Section 7. If a study is to be a comparison of two treatments in respect of a single primary endpoint, with the objective of discovering whether one treatment is superior, non-

inferior, or equivalent to the other, then it will be possible to implement a suitable sequential method using one of the available packages.

When planning to include interim analyses in any clinical trial, the implications of introducing a stopping rule need to be thought through carefully in advance of the study. All parties involved in the production of the protocol should be consulted on the choice of clinically relevant difference, specification of an appropriate power requirement, and the selection of a suitable stopping rule. The operation of any sequential procedure should be clearly described in the statistical section of the trial protocol. Once the trial is underway, it is extremely important that the interim results of an ongoing trial are not circulated widely. Such knowledge may have an undesirable effect on the future progress of the study. Whether a treatment looks good or bad as the trial progresses will obviously affect an investigator's attitude toward the trial. For this reason, it is often the case that a DSMB may be established to consider the progress of a sequential trial as well as assess safety data as discussed in Section 2.

If a DSMB is appointed, one of its first duties should be to review the protocol and to scrutinize any proposed sequential stopping rule prior to the start of the trial. The procedure for undertaking the interim analyses should also be finalized in advance. The DSMB would then review results of the interim analyses as they are reported. It is usual for the board to be supplied with full, unblinded information. Ideally, the only other individual to have knowledge of the treatment comparison would be the statistician who performs the actual analyses of the accumulating data. After each interim examination of the data, the DSMB should issue a short report detailing the outcome of its deliberations. The most common outcome of a meeting is likely to be that the study should continue without modification. However, the DSMB may recommend continuation with some protocol modifications, stopping the study due to safety concerns, or stopping as a result of the formal stopping rule for efficacy having been met. An example of the form of report issued by a DSMB following a meeting is given in Figure 4.

Decision making as part of a sequential trial (whether by a DSMB or another party involved in the trial) is both important and time sensitive. A decision taken to stop a study affects both the current trial and impacts on future trials in the same therapeutic area. However, continuing a trial too long puts participants at unnecessary risk and delays dissemination of important results. The issue applies not only to monitoring efficacy data but also to reacting in a timely manner to safety problems. Wondering whether data are accurate and up-to-date makes the decision process harder. It is therefore necessary to have both timely and accurate data. Unfortunately, a trade-off exists. It takes time to ensure accuracy. One option is to establish a "fast-track" system, processing efficacy data and key safety data separately from other information. Fewer data means that they can be validated more quickly. This is a particularly important issue that requires careful thought and planning.

## 6 EXAMPLE: LITHIUM GAMOLENATE IN PATIENTS WITH ADVANCED PANCREATIC ADENOCARCINOMA

The prognosis for pancreatic cancer patients is poor, with an average survival time of a matter of months from diagnosis. Chemotherapy for pancreatic cancer sufferers

Dear

[redacted]

analysis presented by the independent statistics committee, and then held discussions with representatives of the company.

We noted the incidence of severe side effects but acknowledge that this particular form of cancer is classically associated with thrombotic complications!

We were reassured to hear that centers were now treating patients more expeditiously and with greater familiarity and expertise.

Our decision as to whether to continue the trial was *not* an easy one, but we unanimously agree that the trial should continue under close monitoring and we await the next analysis (70 deaths) with interest.

**Other Comments**

- Would it be possible to look statistically at the first few versus the later patients treated at main centers

[redacted]

***FIGURE 4*** *Example letter sent from the DSMB to the Steering Committee.*

offers only very small survival benefits and is associated with considerable side effects. It was essentially felt that there was no standard treatment available. It has been found that unsaturated fatty acids (lithium gamolenate) have an antitumour effect in experimental studies, and in phase II trials few side effects of this treatment were observed. This prompted the conduct of a randomized dose-finding phase III study [32]. It was required to reach a reliable conclusion concerning the optimum dose when considering survival time from randomization.

Patients with a diagnosis of inoperable pancreatic adenocarcinoma were treated with one of three doses of the active treatment, either oral dose, or low or high intravenous doses. When the objectives of the trial were considered, it was decided that a sequential approach was appropriate. The methodology followed was the boundaries approach, as discussed in Section 4. The comparison employed was

between the high intravenous (IV) dose and the oral dose. The plan was to recruit patients into the study until a sufficiently large difference in the survival rate was seen between these two doses. Analyses involving the low dose IV treatment were considered secondary to this main analysis. The study was designed to have 90% power to detect a difference in median survival of 4 months between the oral dose and the high dose treatments as significant at the 5% level. The reference median survival for oral treatment was taken to be 120 days.

In evaluating the appropriate dose of the treatment, it was desirable to stop the trial as soon as there was sufficient evidence to indicate that either dose was significantly superior to the other. If there was no difference between the doses, then it was also important to distinguish this quickly. In the latter case, further development of oral dose could be justified as this is a less invasive form of the treatment than the intravenous administration. A sequential procedure that satisfies these requirements and has smaller expected sample sizes than the equivalent fixed sample test for any size of treatment difference is the double triangular test [10].

The data concerning survival times following entry into the trial were modeled as survival data, with the difference between the two treatment groups measured by the log hazard ratio. The statistic signifying the difference between the two doses seen so far was taken to be the efficient score for the log hazard ratio. This is the usual log-rank test for survival data, which will be denoted by $Z$. The statistic measuring the amount of information upon which the comparison was based will be denoted by $V$. This is the variance of $Z$ and is approximately one quarter the number of deaths. An independent data and safety monitoring board was established to monitor the progress of the trial at the interim analyses. Figure 5 shows the outcome of five interim analyses. On the diagram the inner dotted boundaries show the Christmas tree correction for discrete looks. These form the stopping boundary, so that the trial should be stopped when one of these values is reached.

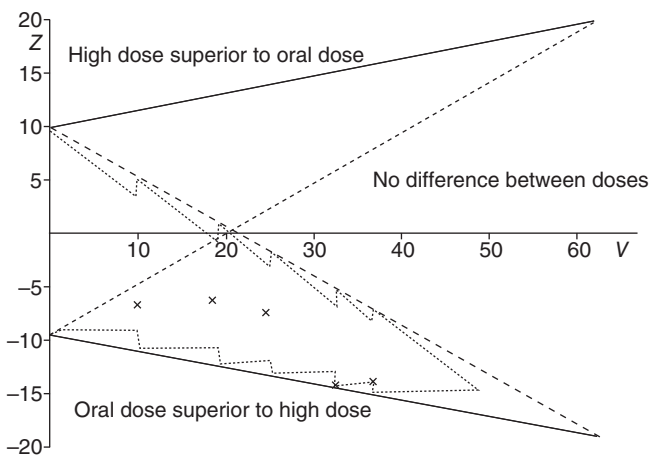


*Z*: Efficient score statistic measuring the treatment difference between high and oral doses

*V*: Statistic measuring the information on which the comparison is based

***FIGURE 5*** *Sequential plot for the trial of lithium gamolenate.*

A total of 278 patients were recruited to the trial. Of these 93 patients received oral dose, 90 patients received low dose, and 95 patients received high dose. The log-rank statistic $Z$ crossed the boundary, which indicated that the oral dose was significantly superior to high dose, which was actually contrary to expectations. This occurred at the fourth interim analysis, and recruitment to the trial was stopped at this time. Continuing follow-up of patients already recruited when the trial stopped led to one further *overrunning* analysis, the results of which are also illustrated in Figure 5. A final analysis of the data was conducted using the computer package PEST [28]. The $p$ value obtained was 0.03, a significant result in favor of the oral dose. Median survival after oral treatment was 129 days, while after high dose treatment it was 94 days.

## 7 RESPONSE-DRIVEN AND ADAPTIVE DESIGNS

In the methodology described in detail in Sections 4 and 5 and illustrated by the example, a single decision is made at each interim analysis: whether the trial should be stopped or should continue to the next interim analysis. Much more general sequential methods can be envisaged in which many aspects of a clinical trial design are reconsidered on the basis of the results from the interim analyses. Decisions could be made after each interim analysis regarding, for example, how many patients should be recruited to the next stage of the trial, whether treatment arms should be dropped from a multiarmed clinical trial, which doses should be used for the treatment of the next group of patients in a dose-finding study, or how the treatment allocation ratio might be changed to favor an apparently superior treatment, in addition to the decision of whether or not to stop the trial. Such general methods are often referred to as *response-driven*, *response-adaptive*, or *data-dependent* designs. They are also sometimes called *sequential designs* or *adaptive designs*, although increasingly these terms are reserved for the methods described in Section 4 and methods based on the work of Bauer and Köhne [33], which are described in the next paragraph, respectively.

A statistical approach for clinical trials that compare two treatments that allows more flexibility than the sequential methods described in Section 4 is the *adaptive design* approach [33]. In this approach, a $p$ value is obtained at each interim analysis to test the null hypothesis that there is no difference between the two treatments on the basis of the new data obtained since the previous analysis. The $p$ values from the different interim analyses are then combined by taking their product, and the value of this product is used to decide whether or not to stop the trial and whether the experimental treatment is superior to the control treatment at the end of the trial. Provided the data from the different stages of the trial are independent, any adaptation of the trial can be conducted without increasing the type I error rate. This method thus provides considerable flexibility in terms of design modification while still enabling valid inference to be drawn. It has been shown, however, that if the flexibility is not utilized, there is a slight reduction in power relative to the sequential methods described in Section 4 [34]. Further details of the adaptive design approach are given, together with a review of recent advances, by Posch et al. [35].

Although there has been much interest in the methodology for more general response-driven designs, there has been relative little use of such approaches in

practice. There are two main areas where such designs have been used. The first area is that of small two-arm studies that attempt to modify treatment allocation toward the treatment observed to be more effective. These methods extend the *play-the-winner* designs in which an interim analysis comparing the two treatments is made after observation of each patient, and the next patient is treated with the treatment currently observed to be most effective. A review of statistical methods for the design and analysis of such studies is given by Rosenberger [36].

The second area where response-driven designs have been used is in phase I dose-finding studies. In these studies the aim is to identify a dose that leads to an acceptably low risk of some adverse event in the trial subjects, who may be patients or healthy volunteers. Subjects are often treated in small groups, or *cohorts*, perhaps of three subjects, and the number of subjects to demonstrate the adverse event is recorded. The responses from each cohort are used to decide the dose level or levels to be used for the next cohort. Reviews of methodology for studies of this type are given by Cavalli and Sessa [37] and Whitehead et al. [38].

## REFERENCES


1. Food and Drug Administration (2001), Draft Guidance for Clinical Trial Sponsors on the Establishment and Operation of Clinical Trial Data Monitoring Committees; available at: http://www.fda.gov/cber/gdlns/clindatmon.pdf.
2. Ellenberg, S. S., Fleming, T. R., and DeMets, D. L. (2003), *Data Monitoring Committees in Clinical Trials: A Practical Perspective*, Wiley, Chichester.
3. Whitehead, J. (1999), On being the statistician on a data and safety monitoring board, *Stat. Med.*, 18, 3425–3434.
4. Gould, A. L., and Shi, W. J. (1992), Sample size re-estimation without unblinding for normally distributed data with unknown variance, *Commun. Stat. Theory Methods*, 21, 2833–2853.
5. Gould, A. L. (1995), Planning and revising the sample size for a trial, *Stat. Med.*, 14, 1039–1052.
6. Wittes, J., Schabenberger, O., Zucker, D., et al. (1999), Internal pilot studies I: Type I error rate of the naive t-test, *Stat. Med.*, 18, 3481–3491.
7. McPherson, K. (1974), Statistics: The problem of examining accumulating data more than once, *N. Engl. J. Med.*, 290, 501–502.
8. Armitage, P., McPherson, C. K., and Rowe, B. C. (1969), Repeated significance tests on accumulating data, *J. Roy. Stat. Soc. A*, 132, 235–244.
9. Whitehead, J. (1999), A unified theory for sequential clinical trials, *Stat. Med.*, 18, 2271–2286.
10. Whitehead, J. (1997), *The Design and Analysis of Sequential Clinical Trials*, Wiley, Chichester.
11. Jennison, C., and Turnbull, B. W. (2000), *Group Sequential Methods with Applications to Clinical Trials*, Chapman and Hall/CRC, Boca Raton, FL.
12. Jennison, C., and Turnbull, B. W. (1997), Group sequential analysis incorporating covariate information, *J. Am. Stat. Assoc.*, 92, 1330–1341.
13. Pocock, S. J. (1977), Group sequential methods in the design and analysis of clinical trials, *Biometrika*, 64, 191–199.

14. O'Brien, P. C., and Fleming, T. R. (1979), A multiple testing procedure for clinical trials, *Biometrics*, 35, 549–556.
15. Pocock, S. J., and White, I. (1999), Trials stopped early: Too good to be true? *Lancet*, 353, 943–944.
16. Pocock, S. J. (1982), Interim analyses for randomised clinical trials: The group sequential approach, *Biometrics*, 38, 153–162.
17. Lan, K. K. G., and DeMets, D. L. (1983), Discrete sequential boundaries for clinical trials, *Bioemtrika*, 70, 659–663.
18. Kim, K., and DeMets, D. L. (1987), Design and analysis of group sequential tests based on the type I error spending rate function, *Biometrika*, 74, 149–154.
19. Hwang, I. K., Shi, W. J., and DeCani, J. S. (1990), Group sequential designs using a family of type I error probability spending functions, *Stat. Med.*, 9, 1439–1445.
20. Stallard, N., and Facey, K. M. (1996), Comparison of the spending function method and the Christmas tree correction for group sequential trials, *J. Biopharma. Stat.*, 6, 361–373.
21. Pampallona, S., Tsiatis, A. A., and Kim, K. (2001), Interim monitoring of group sequential trials using spending functions for the Type I and Type II error probabilities, *Drug Info. J.*, 35, 1113–1121.
22. Wald, A. (1947), *Sequential Analysis*, Wiley, New York.
23. Facey, K. M., and Lewis, J. A. (1998), The management of interim analyses in drug development, *Stat. Med.*, 17, 1801–1809.
24. Tilley, B. C., Marler, J., Geller, N. L., et al., for the National Institute of Neurological Disorders and Stroke (NINDS) rt-PA Stroke Trial Study Group. (1996), Use of a global test for multiple outcomes in stroke trials with application to the National Institute of Neurological Disorders and Stroke t-PA stroke trial, *Stroke*, 27, 2136–2142.
25. Jennison, C., and Turnbull, B. W. (1993), Group sequential tests for bivariate response: Interim analyses of clinical trials with both efficacy and safety endpoints, *Biometrics*, 49, 741–752.
26. Cook, R. J., and Farewell, V. T. (1994), Guidelines for monitoring efficacy and toxicity responses in clinical trials. *Biometrics*, 50, 1146–1152.
27. Todd, S. (1999), Sequential designs for monitoring two endpoints in a clinical trial. *Drug Info. J.*, 33, 417–426.
28. MPS Research Unit (2000), *PEST 4: Operating Manual*, University of Reading, Reading, UK.
29. Cytel Software Corporation (2000), *EaSt 2000: A Software Package for the Design and Interim Monitoring of Group-Sequential Clinical Trials*, Cytel Software Corporation, Cambridge, MA.
30. MathSoft Inc (2000), *S-Plus 2000*, MathSoft Inc., Seattle, WA.
31. Kittelson, J. M., and Emerson, S. S. (1999), A unifying family of group sequential test designs, *Biometrics*, 55, 874–882.
32. Johnson, C. D., Puntis, M., Davidson, N., et al. (2001), Randomized, dose-finding phase III study of lithium gamolenate in patients with advanced pancreatic adenocarcinoma, *Br. J. Surg.*, 88, 662–668.
33. Bauer, P., and Köhne, K. (1994), Evaluation of experiments with adaptive interim analyses, *Biometrics*, 50, 1029–1041.
34. Tsiatis, A. A., and Mehta, C. (2003), On the inefficiency of the adaptive design for monitoring clinical trials. *Biometrika*, 90, 376–378.
35. Posch, M., Bauer, P., and Brannath, W. (2003), Issues in designing flexible trials, *Stat. Med.*, 22, 953–969.

36. Rosenberger, W. F. (1996), New directions in adaptive designs, *Stat. Sci.*, 11, 137–149.
37. Cavalli, F., and Sessa, C. (1999), Current issues in phase I trials: New study designs and informed consent procedures, *Ann. Oncol.*, 10(Suppl. 6), S147–S148.
38. Whitehead, J., Zhou, Y., Stallard, N., et al. (2001), Learning from previous responses in phase I dose-escalation studies, *Br. J. Clin. Pharmacol.*, 52, 1–7.

# *Inference Following Sequential Clinical Trials*

**Aiyi Liu and Kai F. Yu**

*Biostatistics and Bioinformatics Branch, Eunice Kennedy Shriver National Institute of Child Health and Human Development, National Institutes of Health, Rockville, Maryland*

**Contents**

1 Introduction
2 Brownian Motion Paradigm in Sequential Trials
3 Estimating Magnitude of Treatment Difference
4 Calculation of *P* Values as Evidence Against Null Hypothesis
5 Construction of Confidence Intervals
6 Inference Concerning Secondary Endpoint
References

## 1 INTRODUCTION

Mainly in response to ethical and cost concerns, it has become a common practice that clinical trials, especially phase III studies comparing the efficacy of two or more treatment agents, are monitored using the so-called *group sequential testing procedures*. In a group sequential clinical trial a set of stopping boundaries are usually specified for a test statistic at a few time points, either in terms of real calendar time or statistical information. The stopping boundaries are chosen in a way that the type I error (false-positive) and type II error (false-negative) rates are controlled at the desired levels. In the past several decades, extensive research efforts have been devoted to the determination of these stopping boundaries, resulting in a number

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

of popular methods such as those developed by Pocock [1], O'Brien and Fleming [2], Lan and DeMets [3], Whitehead and Stratton [4], and Jennison [5], to name a few. Jennison and Turnbull [6] and Proschan, and co-workers [7] provide excellent overview of these group sequential methods.

Statistically speaking, the primary aim of a clinical trial is to test the null hypothesis that there is no significant difference between treatment agents under investigation, while maintaining a satisfactory statistical power if certain treatment difference indeed exists. A sequential clinical trial should be terminated as soon as the monitoring test statistic, derived based on the so-termed *primary endpoint*, reaches a stopping boundary, although the Data Safety and Monitoring Board (DSMB) may recommend the trial be terminated earlier before reaching a boundary or be continued after reaching a stoping boundary, based on various reasons, such as safety concerns. Once a boundary is reached and the trial stops, a decision can then be made on whether the null hypothesis should be rejected and one treatment agent is better than the other.

When reporting clinical trial findings, it is not sufficient to simply state *rejection* or *no rejection* of the null hypothesis. Point estimates and confidence intervals, which measure the magnitude of the treatment difference, and $P$ values, which give the strength of the trial against the null hypothesis, should also be provided as an integrated part of the trial findings. However, from a frequentist's point of view, the usual maximum-likelihood approach, which provides (asymptotically) most efficient fixed sample-size inference, yields biased estimation due to the randomness of the observed sample size and its dependence with the test statistic and is thus no longer appropriate. In fact such randomness and dependence have been long well recognized as the cause of the type I error inflation when the usual fixed-size critical values are used repeatedly at each interim analysis. Unfortunately, when it comes to inferential analysis following a sequential trial, randomness and dependence are often ignored and the usual maximum-likelihood inference is conducted.

In what follows, we summarize several popular approaches developed in the literature to construct point and interval estimation and $P$ values with respect to the primary endpoint following the termination of a sequential phase III trial. A brief overview will also be given to methods for inference concerning a secondary endpoint. Sequential phase II clinical trials in which interim monitoring is based on a Bernoulli variable (binary outcome) will not be discussed. However, most concepts and principles described here apply analogously to analysis following a sequential phase II trial.

## 2 BROWNIAN MOTION PARADIGM IN SEQUENTIAL TRIALS

A phase III clinical trial is usually monitored based on a Brownian motion $X(t)$ with drift $\delta$, provided that the sample size is large enough so that the test statistics can be well approximated by a Brownian motion. A Brownian motion $X(t)$ with drift $\delta$ satisfies several nice properties: $X(0) = 0$, and $X(t) \sim N(\delta t, t)$ for $t > 0$. Furthermore, for any $0 < t_1 < t_2 < \cdots < t_k < \infty$, $X(t_1)$, $X(t_2) - X(t_1)$, $\ldots$ , $X(t_k) - X(t_{k-1})$ are mutually independent, a proposition so well termed as *independent increments*, each normally distributed as $X(t_j) - X(t_i) \sim N(\delta(t_j - t_i), t_j - t_i)$ and $\mathrm{cov}(X(t_j), X(t_i)) = t_i$, for $i < j$. In practice, $t$ is usually the Fisher information and $X(t)$ is usually a test statistic, and

the drift parameter $\delta$ is a measure of treatment difference with $\delta = 0$ if there is no treatment difference.

To give the readers a flavor of how such Brownian motion may be constructed, let us consider the case of comparing, with equal allocation, two normal means with known common variance. Let $X_{1A}, \ldots, X_{nA}$ be the outcomes from $n$ subjects allocated to treatment $A$, and $X_{1B}, \ldots, X_{nB}$ be that to treatment $B$. Assume that $X_{iA} \sim N(\mu_A, \sigma^2)$ and $X_{iB} \sim N(\mu_B, \sigma^2)$. The null hypothesis is $H_0$: $\mu_A - \mu_B = 0$. The standardized test statistic is $Z_n = \sum_{i=1}^{n}(X_{iA} - X_{iB})/\sqrt{2n\sigma^2}$, which follows standard normal $N(0, 1)$, and a two-sided test with level of significance $\alpha$ rejects the null hypothesis if $|Z_n| > \Phi^{-1}(1 - \alpha/2)$, where $\Phi$ is the standard normal distribution. To put this into a Brownian motion framework, we redefine (after a little bit pleasant algebraic manipulation) the test statistic as

$$X = \sqrt{\frac{n}{2\sigma^2}} Z_n$$

then $X$ resembles a Brownian motion $X(t)$ with drift $\delta = \mu_A - \mu_B$ and in this example the (discrete) time $t = n/(2\sigma^2)$.

This Brownian motion paradigm applies to the two-arm comparison of many clinical endpoints such as proportions, ordinal data, and survival rates. Whitehead [8] presented a nice review of this structure and named it *a unified theory*. See also Lan and Zucker [9] and Proschan and co-workers [7].

With this framework, a group-sequential clinical trial can be formulated as interim monitoring of a Brownian motion $X(t)$ with drift $\delta$ at several discrete time points $t_1, \ldots, t_K$, where $K$ is the prespecified number of interim and looks to be carried out during the whole course of the trial. For each look $k$, let $C_k = (a_k, b_k)$ be the continuation region and $S_k = (-\infty, a_k] \cup [b_k, \infty)$ be the stopping region, where $a_k < 0 < b_k$ define the lower and upper boundary of the sequential test. Then the trial stops at the $k$th look if the $k$th boundary is reached first, that is, if

$$X(t_i) \in C_i, i \le k-1 \qquad X(t_k) \in S_k$$

Upon stopping at the $k$th interim look, we observe two statistics, the stopping time $T = t_k$ and the position of the Brownian motion $X(T) = X(t_k)$. These two statistics together are sufficient for the drift parameter $\delta$ and hence inference on $\delta$ can be based solely on $(T, X(T))$. The joint distribution $f_\delta(k, x)$ of $(T, X(T))$ at $(T, X(T)) = (t_k, x)$ is $f_\delta(k, x) = f_0(k, x)\exp\{x\delta - \frac{1}{2}t_k\delta^2\}$, where the baseline functions $f_0(k, x)$ can be computed numerically using the following recursive formulas:

$$f_0(1, x) = \frac{1}{\sqrt{t_1}}\phi\left(\frac{x}{\sqrt{t_1}}\right)$$

and

$$f_0(k, x) = \frac{1}{\sqrt{t_k - t_{k-1}}}\int_{C_{k-1}} \phi\left(\frac{x-s}{\sqrt{t_k - t_{k-1}}}\right) f_0(k-1, s)\,ds \qquad k = 2, \ldots, K$$

where $\phi$ is the standard normal density function. Thus, for a statistic $h(T, X(T))$, its expectation is given by

$$E(h(T, X(T))) = \sum_{k=1}^{K} \exp\left\{-\frac{1}{2} t_k \delta^2\right\} \int_{S_k} h(k, x) f_0(k, x) \exp\{x\delta\} dx \tag{1}$$

## 3 ESTIMATING MAGNITUDE OF TREATMENT DIFFERENCE

The maximum-likelihood estimate of $\delta$ is $\hat{\delta}_{\mathrm{ML}}(T, X(T)) = X(T)/T$, the same form as that following a fixed-size test ($K = 1$). However, its statistical properties change substantially due to the randomness of $T$ and further complication of the dependence between $T$ and $X(T)$. As a consequence, many optimal properties that the estimate enjoys under fixed-size sampling vanish under sequential sampling. For instance, the maximum-likelihood estimate is no longer unbiased. When $K = 2$, the bias function $b(\delta)$ can be expressed in a closed form as

$$b(\delta) = \sqrt{t_1}\left(\frac{1}{t_1} - \frac{1}{t_2}\right)\left\{\phi\left(\frac{b_1 - \delta t_1}{\sqrt{t_1}}\right) - \phi\left(\frac{a_1 - \delta t_1}{\sqrt{t_1}}\right)\right\}$$

See Emerson [10].

A naive approach to reduce the bias is to subtract from $\hat{\delta}_{\mathrm{ML}}$ the maximum-likelihood estimate of the bias $b(\delta)$, yielding an estimate as $\hat{\delta} = \hat{\delta}_{\mathrm{ML}} - b\left(\hat{\delta}_{\mathrm{ML}}\right)$. Since $\hat{\delta}_{\mathrm{ML}}$ is biased for $\delta$, $b\left(\hat{\delta}_{\mathrm{ML}}\right)$ may not be a good estimate for $b(\delta)$. Therefore, the resulting estimate $\hat{\delta}$ may not be a good estimate for $\delta$. Whitehead [11] instead suggested replacing $\delta$ in the bias function by the so-called *bias-adjusted estimate*, $\hat{\delta}_{\mathrm{W}}$, to be solved from the equation

$$\hat{\delta}_{\mathrm{W}} + b\left(\hat{\delta}_{\mathrm{W}}\right) = \hat{\delta}_{\mathrm{ML}}$$

Whitehead [11] initially proposed his bias-adjusted estimate in the context of *fully sequential tests* with parallel or triangular stopping boundaries. He presented substantial numerical evidence for both boundaries that the bias-adjusted estimate indeed has smaller bias (and mean-squared error as well!) than the maximum-likelihood estimate.

However, the idea is quite general and applicable in reducing the bias of an estimate, so long as the bias function of the estimate can be worked out. The performance of such estimate depends on the shape of the bias functions. If the bias function of the estimate is nondecreasing and can be differentiated at least twice, then the bias-adjusted estimate has uniformly smaller mean-squared error than the initial estimate; see Whitehead [11] and Liu [12].

Emerson and Fleming [13] investigated Whitehead's bias-adjusted estimate in the context of group-sequential tests and compared with an unbiased estimate constructed in their study using conditioning argument:

$$\hat{\delta}_{\mathrm{EF}} = E\left\{\frac{X(t_1)}{t_1}\middle| (T, X(T))\right\}$$

This is indeed an unbiased estimate since the expectation of $X(t_1)/t_1$ is always $\delta$, regardless at which stage the test stops. Although $\hat{\delta}_{EF}$ nullifies the bias, numerical results from Emerson and Fleming [13] showed that in general it has a larger mean-squared error than Whitehead's bias-adjusted estimate, a sacrifice to be made for enforcing unbiasedness. As expected, the two estimates both have smaller bias and mean-squared error than the maximum-likelihood estimate, as considerable numerical results have shown. However, analytic results are limited, perhaps due to the complicated forms of the density functions, on the statistical properties and comparison of the two estimates.

More can be said about how the sequential sampling [resulting in the randomness of $T$ and dependence between $T$ and $X(T)$] changes the statistical property of a parameter estimate. Under fixed-size sampling, the maximum-likelihood estimate $\hat{\delta}_{ML}$ is the unique unbiased estimate based on the sufficient statistic $X(T)$ ($T$ is a constant now) and has uniformly minimum variance among the unbiased estimates, a fact attributed to the well-known Rao–Blackwell theorem. Under group-sequential sampling, however, as shown in Liu and Hall [14], there are infinitely many sufficient-statistic-based unbiased estimates and none, $\hat{\delta}_{EF}$ particularly included, has uniformly minimum variance. Only when restricted to unbiased estimates that are independent of future stopping rules does $\hat{\delta}_{EF}$ have uniformly minimum variance.

## 4 CALCULATION OF *P* VALUES AS EVIDENCE AGAINST NULL HYPOTHESIS

The $P$ values in a clinical trial measure the statistical strength (evidence) against the null hypothesis that there is no difference between treatment agents. It is defined as the probability, computed under the null hypothesis, of obtaining data that yields a statistic of interest being at least as extreme as the observed value of the statistic. Thus, smaller $P$ values correspond to more extreme data from which the statistic is computed and reflect stronger evidence against the null hypothesis.

In a fixed-size clinical trial ($K = 1$ and $t_1$ is a constant), the calculation and interpretation of the $P$ values is more straightforward and much less problematic. So long as the statistic of interest is a monotone function of $X(t_1)$, as is true for most cases, the upper- and lower-tail $P$ values are given by, respectively,

$$P_{\delta=0}(X(t_1) > \tilde{x}) = 1 - \Phi\left(\frac{\tilde{x}}{\sqrt{t_1}}\right) \qquad P_{\delta=0}(X(t_1) < \tilde{x}) = \Phi\left(\frac{\tilde{x}}{\sqrt{t_1}}\right)$$

where $\tilde{x}$ is the realization (observed value) of $X(t_1)$. Doubling the smaller of the upper- and lower-tail $P$ values yields the two-sided $P$ value, which, because of symmetry, is given by

$$P \text{ value} = P_{\delta=0}(|X(t_1)| > |\tilde{x}|) = 2\left\{1 - \Phi\left(\frac{\tilde{x}}{\sqrt{t_1}}\right)\right\}$$

Unfortunately, the definition of a $P$ value in a group-sequential trial is not as clear as it would be in a fixed-size trial and sometime brings considerable ambiguity and confusion to many practitioners. What does it mean by data being *extreme*? What statistic should be used in defining *extreme*? The problem comes from the fact that, instead of summarizing the data by a one-dimensional sufficient statistic $X(t_1)$ in a fixed-size trial, we now have to summarize the data by two dependent statistics, $T$ and $X(T)$, both of which contain information on the treatment difference $\delta$. A sample path (or data) with smaller $T$ (earlier stopping) and larger $X(T)$ is without doubt more extreme than a path with larger $T$ and smaller $X(T)$. But what about a path with $T$ and $X(T)$ both smaller and a path with $T$ and $X(T)$ both larger? In a word, how should the points $(T, X(T))$ in the sample space in a group-sequential trial,

$$\Omega = \{X(t_1) \in S_1\} \cup \left[\cup_{k=2}^{K} \{X_{t_i} \in C_i, i < k, X_{t_k} \in S_k\}\right]$$

be ordered so that *being extreme* is well defined? Is there an ordering that is better than the others, statistically and/or practically?

A number of popular orderings have been proposed in the literature, most based on a one-dimensional statistic $s(T, X(T))$ defined on the sample space $\Omega$. A point $(t_i, X(t_i))$ is more extreme than another point $(t_j, X(t_j))$ if $s(t_i, X(t_i)) > s(t_j, X(t_j))$. With this ordering, suppose the trial stops at stage $k$ with $X(t_k) = \tilde{x}$, then the upper-tail $P$ value is given by

$$P_{\delta=0}\{s(T, X(T)) > s(t_k, \tilde{x})\}$$

which can be computed numerically using the recurse density functions. The lower-tail $P$ value is given accordingly, and the two-sided $P$ value is twice the smaller of the two $P$ values.

Roser and Tsiatis [15] defined $s(T, X(T))$ to be the score statistic (the position of the Brownian motion) $X(T)$, while Chang [16] suggested using the standardized test statistic $X(T)/\sqrt{T}$, and Emerson and Fleming [13] investigated ordering based on the maximum-likelihood estimate $X(T)/T$.

The so-called *stagewise ordering* (e.g., Siegmund [17], Fairbanks and Madsen [18], and Tsiatis and co-workers [19] is somewhat different from the above three orderings and is not based on a one-dimensional statistic. A data point $(t_i, X(t_i))$ is more extreme than the other $(t_j, X(t_j))$ if it reaches the upper or lower boundary earlier $[i < j$ and $X(t_i) > b_i$ or $< a_i]$. For data points with the same information time, the one with larger value of $X(t)$ is more extreme.

All four orderings reduce to the same one in the fixed-size setting. However, there are no deterministic results on which ordering is better than the others. Some argue that the stagewise ordering is more preferable since it yields $P$ values in agreement with the decision from the hypothesis testing; it is less than the significance level $\alpha$ if and only if the null hypothesis is rejected. Moreover, the stagewise ordering does not depend on the stages beyond the stopping one, making it applicable in some more flexible designs in which the information time points at which interim analysis is conducted are not specified in advance. See Jennison and Turnbull [6].

## 5 CONSTRUCTION OF CONFIDENCE INTERVALS

A confidence interval of $\delta$ with level $(1-\alpha)$ is an interval with lower bound $\delta_L^* = \delta_L^*(T, X(T))$ and upper bound $\delta_U^* = \delta_U^*(T, X(T))$ such that for any $\delta$

$$P_\delta\{\delta_L^* < \delta < \delta_U^*\} \geq 1-\alpha$$

One approach to construct such an interval is to use an ordering of the sample space described previously for calculating $P$ values. Let the terminal data be $(t_k, \tilde{x})$. Then $\delta_L^*$ is the drift value at which the probability is half $\alpha$ of observing data being at least as extreme as $(t_k, \tilde{x})$, and $\delta_U^*$ is the drift value at which the probability is $1-\alpha/2$ of observing data being at least as extreme as $(t_k, \tilde{x})$. That is, $\lambda(\delta_L^*; t_k, \tilde{x}) = \alpha/2$ and $\lambda(\delta_U^*; t_k, \tilde{x}) = 1-\alpha/2$, where

$$\lambda(\delta; t_k, \tilde{x}) = P_\delta\{(T, X(T)) \text{ is at least as extreme as } (t_k, \tilde{x})\}$$

In order for the two equations to yield an interval, the function $\lambda(\delta; t_k, \tilde{x})$ needs to be increasing in $\delta$ for any given data point $(t_k, \tilde{x})$. With such monotonicity each equation has a unique solution and $\delta_L^* < \delta_U^*$ holds. If the $\lambda$ function is nondecreasing, then $\delta_L^*$ can be set to be $\{\delta: \lambda(\delta; t_k, \tilde{x}) = \alpha/2\}$ and $\delta_U^*$ to be $\{\delta: \lambda(\delta; t_k, \tilde{x}) = 1-\alpha/2\}$. It can be shown analytically that Emerson and Fleming's [13] ordering and the stagewise ordering both yield increasing $\lambda$ functions; see Emerson [10] or Hall and Liu [20] for the former and Jennison and Turnbull [6, Chapter 8] for the latter. There is no analytic results on this matter for the other two orderings though limited numerical results do not seem to support the monotonicity property.

Jennison and Turnbull [6, Chapter 8] discussed several criteria for defining a good confidence interval such as agreement with the hypothesis testing results and inclusion of the maximum-likelihood estimate. They come to a conclusion that the stagewise ordering gives a more preferable confidence interval, which, in addition, is the only feasible method to use when future information time for interim analysis is not specified.

Woodroof [21] proposed a different approach to construct a confidence interval by modifying the fixed-size procedure to take the bias into consideration. Write $Z(\delta) = (X(T) - T\delta)/\sqrt{T}$. In fixed-size sampling, $Z$ is normally distributed with zero mean and unit variance and serves as a pivot to construct the usual $(1-\alpha)$-level confidence interval with limits $X(T)/T \mp \Phi^{-1}(1-\alpha/2)/\sqrt{T}$. In a group-sequential sampling, this pivot no longer has a standard normal distribution, so a certain adjustment is needed. A new pivot is defined as $Z(\delta) = (Z(\delta) - \mu(\delta))/s(\delta)$ where $\mu(\delta)$ and $s(\delta)$ are the mean and standard deviation function of $Z(\delta)$, respectively. If we assume that this new pivot approximately follows a standard normal distribution, then an approximate $(1-\alpha)$-level confidence interval for $\delta$ has confidence limits being

$$\left\{\frac{X(T)}{T} - \frac{\hat{\mu}}{\sqrt{T}} \mp \frac{\hat{s}}{\sqrt{T}}\Phi^{-1}\left(1-\frac{\alpha}{2}\right)\right\}$$

where $\hat{\mu}$ and $\hat{s}$ are proper estimates of $\mu(\delta)$ and $s(\delta)$, respectively, which are often obtained with $\delta$ replaced by its maximum-likelihood estimate.

Unlike the confidence intervals based on a sample space ordering, the above pivotal method does not always guarantee the exact coverage probability of $1 - \alpha$. However, it works pretty well in most cases, and the coverage probability of the resulting confidence intervals are quite close to the nominal level of $1 - \alpha$.

We have seen applications of the orderings of the sample space to constructing confidence intervals and calculating $P$ values. Another usage of a proper ordering of the sample space is to compute a *median-unbiased* estimate $\hat{\delta}_M = \hat{\delta}_M(T, X(T))$ of $\delta$ such that

$$P_\delta\left\{\hat{\delta}_M(T, X(T)) < \delta\right\} = P_\delta\left\{\hat{\delta}_M(T, X(T)) > \delta\right\} = \frac{1}{2}$$

That is, a median-unbiased estimate has equal chances to over- or underestimate the parameter. (In contrast, the usual unbiased estimates such as $\hat{\delta}_{EF}$ are referred to as *mean-unbiased* estimates, if such distinction needs to be made.) For an ordering of sample space with monotonicity property, the corresponding median-unbiased estimate at $(t_k, \tilde{x})$ can be found by solving for $\delta$ the equation:

$$\frac{1}{2} = \lambda(\delta; t_k, \tilde{x}) = P_\delta\{(T, X(T)) \text{ is at least as extreme as } (t_k, \tilde{x})\}$$

Emerson and Fleming [13] investigated the bias and mean-squared error of the median-unbiased estimates based on certain orderings. Compared with the bias-adjusted estimate $\hat{\delta}_W$ and the unbiased estimate $\hat{\delta}_{EF}$, the median-unbiased estimates in general tend to have larger bias, often coupled with larger mean-squared error as well.

## 6 INFERENCE CONCERNING SECONDARY ENDPOINT

Termination of a sequential clinical trial is based on the *primary endpoint*, such as patients' survival in a cancer trial. In contrast, a *secondary endpoint*, such as patients' disease-free survival or treatment × strata interaction, refers to an endpoint whose values are also observed from each patient along with the primary endpoint but do not contribute in any way to the termination of the trial. The secondary endpoints are usually analyzed only after the trial stops, though they may be evaluated at certain interim analysis, and such evaluation plays no role in sequential monitoring of the trial.

A secondary endpoint is usually correlated with the primary endpoint since they both are observed from the same patients. Due to the correlation, the usual likelihood analysis should also be adjusted for the random stopping of the trial so that a valid inference on secondary endpoints can be made. For most trials the primary and a secondary endpoint can be put into a random-sampling framework and be (asymptotically) modeled as two correlated Brownian motions with constant (often assumed to be known) correlation and proportional information time. The first Brownian motion, as given in the previous sections, governs the stopping of the trial, and the second one summarizes the observations of the secondary endpoint with

its drift parameter usually measuring the difference in the endpoint between treatment arms. For this drift parameter, Whitehead [22] constructed a bias-adjusted estimate to reduce the bias of the maximum-likelihood estimate, and Liu and Hall [23] investigated its unbiased estimates. Certain confidence intervals are developed by Whithead and co-workers [24] using Woodroofe's pivotal method; these confidence intervals can also be used for testing hypotheses concerning the secondary endpoints.

Not all secondary endpoints can be modeled (asymptotically) with the primary endpoint as two Brownian motion with constant correlation coefficients. Yakir and Hall [25] and Hall and Yakir [26] gave examples that do not satisfy this model. One example is that a sequential trial with staggered entry is carried out to compare the survival rate of two treatment arms based on the log-rank statistic. Upon termination of the trial, one wants to know whether treatment × gender interaction exists. In these examples, the primary and secondary endpoints are modeled as two Gaussian processes with correlation being a function of the information time. Hall and Yakir [26] derived distributional theory and certain optimal properties to construct point estimates and confidence intervals of the parameters associated with the secondary endpoint.

The inference procedures discussed so far are parametric in nature. Chuang and Lai [27, 28] developed nonparametric methods based on a resampling technique for confidence intervals concerning both the primary and secondary endpoints.

## REFERENCES


1. Pocock, S. J. (1977), Group sequential methods in the design and analysis of clinical trials, *Biometrika*, 64, 191–199.
2. O'Brien, P. C., and Fleming, T. R. (1979), A multiple testing procedure for clinical trials, *Biometrics*, 35, 549–556.
3. Lan, K. K. G., and DeMets, D. L. (1983), Discrete sequential boundaries for clinical trials, *Biometrika*, 70, 659–663.
4. Whitehead, J., and Stratton, I. (1983), Group sequential clinical trials with triangular continuation regions, *Biometrics*, 39, 227–236.
5. Jennison, C. (1987), Efficient group sequential tests with unpredictable group sizes, *Biometrika*, 74, 155–165.
6. Jennison, C., and Turnbull, B. W. (2000), *Group Sequential Methods with Applications to Clinical Trials*, Chapman and Hall/CRC, New York.
7. Proschan, M. A., Lan, K. K. G., and Wittes, J. T. (2006), *Statistical Monitoring of Clinical Trials: A Unified Approach*, Springer, New York.
8. Whitehead, J. (1999), A unified theory for sequential clinical trials, *Stat. Med.*, 18, 2271–2286.
9. Lan, K. K. G., and Zuker, D. (1993), Sequential monitoring of clinical trials: The role of information and Brownian motion, *Stat. Med.*, 12, 753–765.
10. Emerson, S. S. (1988), Parameter estimation following group sequential hypothesis testing [dissertation], University of Washington, Seattle.
11. Whitehead, J. (1986), On the bias of maximum likelihood estimation following a sequential test, *Biometrika*, 73, 573–558.

12. Liu, A. (2003), A simple low-bias estimate following a sequential test with linear boundaries, in Kolassa, J., and Oakes, D., Ed., *Crossing Boundaries: Statistical Essays in Honor of Jack Hall*, Institute of Mathematical Statistics Lecture Notes Monograph Series, Beachwood, OH, Vol. 43, pp. 47–58.
13. Emerson, S. S., and Fleming, T. R. (1990), Parameter estimation following sequential hypothesis testing, *Biometrika*, 77, 875–892.
14. Liu, A., and Hall, W. J. (1999), Unbiased estimation following a group sequential test. *Biometrika*, 86, 71–78.
15. Rosner, G. L., and Tsiatis, A. A. (1988), Exact confidence intervals following a group sequential trial: A comparison of methods, *Biometrika*, 75, 723–729.
16. Chang, M. N. (1989), Confidence intervals for a normal mean following a group sequential test, *Biometrics*, 45, 247–254.
17. Siegmund, D. (1978), Estimation following sequential tests. *Biometrika*, 65, 295–297.
18. Fairbanks, K., and Madsen, R. (1982), *P* values for tests using a repeated significance design, *Biometrika*, 69, 69–74.
19. Tsiatis, A. A., Rosner, G. L., and Metha, C. R. (1984), Exact confidence intervals following a group sequential test, *Biometrics*, 40, 797–803.
20. Hall, W. J., and Liu, A. (2002), Sequential tests and estimates after overrunning based on maximum-likelihood ordering, *Biometrika*, 89, 699–707.
21. Woodroofe, M. (1992), Estimation after sequential testing: A simple approach for truncated sequential probability ratio test, *Biometrika*, 79, 347–353.
22. Whitehead, J. (1986), Supplementary analysis at the conclusion of a sequential clinical trial, *Biometrics*, 42, 461–471.
23. Liu, A., and Hall, W. J. (2001), Unbiased estimation of secondary parameters following a sequential test. *Biometika*, 88, 895–900.
24. Whitehead, J., Todd, S., and Hall, W. J. (2000), Confidence interval for secondary parameters following a sequential test, *J. Roy. Statist. Soc., B*, 62, 731–745.
25. Yakir, B., and Hall, W. J. (2003), Testing for a treatment-by-stratum interaction in a sequential clinical trial, in Kolassa, J., and Oakes, D., Ed., *Crossing Boundaries: Statistical Essays in Honor of Jack Hall*, Institute of Mathematical Statistics Lecture Notes Monograph Series, Beachwood, OH, Vol. 43, pp. 1–12.
26. Hall, W. J., and Yakir, B. (2003), Inference about a secondary process after a sequential trial, *Biometrika*, 90, 597–611.
27. Chuang, C. S., and Lai, T. L. (2000), Hybrid resampling methods for confidence intervals. With discussion and rejoinder by the authors, *Statistica Sinica*, 10, 1–50.
28. Chuang, C. S., and Lai, T. L. (1998), Resampling methods for confidence intervals in group sequential trials, *Biometrika*, 85, 317–332.

# *Statistical Methods for Analysis of Clinical Trials*

**Duolao Wang,[1] Ameet Bakhai,[2] and Nicola Maffulli[3]**

[1]*Medical Statistics Unit, London School of Hygiene and Tropical Medicine, London, United Kingdom*
[2]*Barnet General & Royal Free Hospitals, London, United Kingdom*
[3]*Department of Trauma and Orthopaedic Surgery, Keele University School of Medicine, Keele, Staffordshire, United Kingdom*

**Contents**

1 Introduction
  1.1 Bias and Systematic Errors
  1.2 Confounding
  1.3 Random Error
2 Types of Data, Summary, and Data Presentation
  2.1 Types of Data
  2.2 Data Description and Presentation
  2.3 Summarizing Quantitative Variables
3 Normal Distribution: Symmetric Frequency Distribution
  3.1 What Is a Normal Distribution?
  3.2 Properties of Normal Distribution
4 Principles of Statistical Inference
  4.1 Hypothesis Testing
  4.2 Alpha (Type I) and Beta (Type II) Errors
  4.3 Confidence Intervals
  4.4 Relationship between Significant Testing and Confidence Intervals
  4.5 Examples
5 Comparison of Two Means

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

6 Comparison of Two Proportions
6.1 Assessing Size of Treatment Effects in Two-Arm Trial
7 Survival Analysis
7.1 Example: Pancreatic Cancer Trial
7.2 Basic Concepts in Survival Analysis
7.3 Assessing Size of Treatment Effect in Two-Arm Trial for Survival Data
8 Commonly Used Terms in Clinical Trials
9 Concluding Remarks
References

## 1 INTRODUCTION

Evidence-based medicine is a cornerstone of current medical practice. Randomized controlled trials are regarded as the most robust form of evidence-based medicine. An appreciation of statistical methods is fundamental to understanding randomized trial methods and results.

A randomized controlled trial aims to provide unbiased treatment effect regarding the efficacy and safety of a medicinal product or a therapeutic procedure. The observed treatment effect, however, may represent the "true" difference between the new drug and the comparative treatment or it may not. Therefore, if the trial were repeated with all the available patients in the world, then the outcome would either be the same as the trial (a true result) or different (making the trial result a chance event, or an erroneous false result). Understanding the possible sources of erroneous results is critical in the appreciation of clinical trials. The reasons for erroneous results fall into three main categories.

- First, the trial may have been biased in some predictable fashion.
- Second, it could have been contaminated (confounded) by an unpredictable factor.
- Third, the result may simply have occurred by chance.

### 1.1 Bias and Systematic Errors

Bias can influence a trial by the occurrence of systematic errors that are associated with the design, conduct, analysis, and reporting of the results of a clinical trial. Bias can also make the trial-derived estimate of a treatment effect deviate from its true value [1, 2]. The most common types of bias in clinical trials are those related to subject selection and outcome measurement. For example, if the investigators are aware of which treatment a patient is receiving, it could affect the way they collect information on the outcome during the trial, or they might recruit patients in a way that could favor the new treatment, resulting in a selection bias.

In addition, exclusion of subjects from statistical analysis because of noncompliance or missing data could bias an estimate of the true benefit of a treatment, particularly if more patients were removed from analysis in one group than the other [3, 4]. Much of the advanced design strategies seek to reduce these systematic errors.

### 1.2 Confounding

Confounding represents the distortion of the true relationship between treatment and outcome by another factor, for example, the severity of disease [5]. Confounding occurs when an extra factor is associated with both the outcome of interest and treatment group assignment. Confounding can both obscure an existing treatment difference and create an apparent difference that does not exist.

If we divided patients into treatment groups based on inherent differences (such as mean age) at the start of a trial, then we would be very likely to find the benefit of the new treatment to be influenced by those preexisting differences. For example, if we assign only smokers to get treatment A, only nonsmokers to get treatment B, and then assess which treatment protects better against cardiovascular disease, we might find that the benefit seen with treatment B is due to the lack of smoking in this group. The effect of treatment B on cardiovascular disease development would therefore be confounded by smoking.

Randomization in conjunction with a large sample size is the most effective way to restrict such confounding, by evenly distributing both known and unknown confounding factors between treatment groups. If, before the study begins, we know which factors may confound the trial, then we can use randomization techniques that force a balance of these factors (stratified randomization). In the analysis stage of a trial, we might be able to restrict confounding using appropriate statistical techniques such as stratified analysis and regression analysis [6].

### 1.3 Random Error

Even if a trial has an ideal design and is conducted to minimize bias and confounding, the observed treatment effect could still be due to random error or chance [1, 2, 7]. The random error can result from sampling, biologic, or measurement variation in outcome variables. Since (given specific selection criteria) the patients in a clinical trial are only a sample of all possible available patients, the sample might yet show a chance false result compared to the overall population. This is known as a sampling error. Sampling errors can be reduced by choosing a very large group of patients. Other causes of random error are described elsewhere [2].

Statistical analyses deal with random error by providing an estimate of how likely the measured treatment effect reflects the true effect [7–9]. Statistical testing or inference involves an assessment of the probability of obtaining an observed treatment difference or more extreme difference for an outcome, assuming that there is no difference between treatments. This probability is often called the $P$ value or false-positive rate. If the $P$ value is less than a specified critical value (e.g., 5%), the observed difference is considered to be statistically significant. The smaller the $P$ value, the stronger the evidence is for a true difference between treatments. On the other hand, if the $P$ value is greater than the specified critical value, then the observed difference is regarded as not statistically significant and is considered to be potentially due to random error or chance. The traditional statistical threshold is a $P$ value of 0.05 (or 5%), which means that we only accept a result when the likelihood of the conclusion being wrong is less than 1 in 20. In other words we conclude that only 1 out of a hypothetical 20 trials will show a treatment difference when in truth there is none.

Statistical estimates summarize the treatment differences for an outcome in the forms of point estimates (e.g., means or proportions) and measures of precision [e.g., confidence intervals (CIs)] [7]. A 95% CI for a treatment difference means that the range presented for the treatment effect contains (when calculated in 95 out of 100 hypothetical trials assessing the same treatment effect) the true value of treatment difference, that is, the value we would obtain if we were to use the entire available patient population is 95% likely to be contained in the 95% CI.

Finally, testing several different hypotheses with the same trial (e.g., comparing treatments with respect to different outcomes or for several smaller subpopulations within the trial population) will increase the chance of observing a statistically significant difference purely due to chance [10]. Even examining the difference between treatments at many time points (interim analyses) throughout the length of the trial could lead to a spurious result due to multiple testing [10, 11]. Therefore, the aim should be to plan a trial in such a way that the occurrence of any such errors is minimal.

## 2 TYPES OF DATA, SUMMARY, AND DATA PRESENTATION

### 2.1 Types of Data

Clinical studies compare effects of medical treatment or interventions on two or more groups of subjects. Data are collected, however, at the individual subject or patient level. Data are composed of baseline characteristics such as age, gender, height, weight, and so forth, or of disease factors such as presence of arthritis, coronary disease, or severity of a fracture, or of treatment response variables such as reduction in pain, improvement of disease, or prolongation of life. These data come from variables of different types—either continuous (or quantitative), such as age or hemoglobin levels, or they may be categorical (or qualitative), such as race or gender.

Data can be classified further into four main groups:

- Binary (categorical): for example, sex (male and female)
- Unordered (categorical): for example, race (white, black, other)
- Ordered (categorical): for example, severity (mild, moderate, severe)
- Numerical (continuous): for example, age (in years)

### 2.2 Data Description and Presentation

In clinical studies, data are summarized for presentation by treatment groups using frequency distributions. These present the distribution of both qualitative and quantitative data, summarizing how often each value of a data point is repeated. With quantitative data, we mostly present a grouped frequency distribution. From a frequency table we can appreciate:

- The frequency (number of cases) occurring for each category or interval—for example, number of 70-year-old patients

- The relative frequency (percentage) of the total sample in each category or interval—for example, 70-year-old patients were 10% of the overall sample
- The highest and lowest or the range of possible values from our patient groups—for example, in that group the oldest patient was 96 and the youngest patient was 25

Although a frequency table provides a detailed summary of the distribution of the data, presentation of the distribution in a graph/chart makes the message from the data more informative. The type of graph depends on the type of data. Generally, if the data are categorical, we use a bar graph or a pie chart. If the data are continuous, a histogram or frequency polygon is more appropriate. As well as for the entire study population, the frequency distributions can then be presented for each treatment group. This is a fast way of ascertaining whether there are broad similarities or differences between treatment groups. Later we can use statistical tests to ascertain whether any differences between the groups are significant.

### 2.3 Summarizing Quantitative Variables

Categorical variables may be expressed as percentages and compared between different groups. There is little more that can be done to describe such variables. However, for a quantitative variable we can do more than this. We have other measures with which to summarize the data (summary measures). From the frequency distribution we can calculate the location (or central tendency) that summarizes where the center of the distribution lies and we can also summarize the spread/range (or variation) of the distribution, which describes how widely the values are spread above and below the central value.

There are two measures commonly used to describe the location for the central tendency, depending on whether the distribution is even (such as age) or more biased in one direction (such as physical fitness—many more people are unfit than are Olympic athletes).

- *Mean:* The mean can be calculated by summing all the quantitative observations and divided the sum by the number of observations.
- *Median:* The median is the value that divides the distribution equally. The median is more appropriate for distributions that are skewed, such as physical fitness. When the distribution is symmetrical, the median equals the mean.

There are three measures commonly used to summarize the spread of a variable:

- *Standard deviation:* This gives the average distance of all observations from the mean. The standard deviation has an important role in statistical analysis.
- *Range:* Any data between the lowest and highest values is known as the full range of values.
- *Percentile:* This is the value below which a given percentage of the data observations occur. The most commonly used percentiles are 5 and 95%. Using these overcomes the problem of extreme data values away from the mean or median.

## 3 NORMAL DISTRIBUTION: SYMMETRIC FREQUENCY DISTRIBUTION

### 3.1 What Is a Normal Distribution?

Quantitative variables can range from very small to very large values, and, in some situations, from negative to positive values. Consider the systolic blood pressure (BP) measurements of 1000 subjects participating in a lifestyle survey. The frequency distribution of these systolic BPs is shown in Figure 1. The height of each vertical bar in this graph shows the proportion (or probability) of subjects whose systolic blood level was between the values at the base of the bar. So, if we summed the heights of all the bars in the histogram, we should find that the total of all the proportions is 1. Proportion is the same as the percentage expressed as a fraction over 100%.

The distribution of these blood pressures conforms approximately to a bell-shaped curve or normal distribution typical for common continuous biological measurements. The curve has particular mathematical properties expressed by the

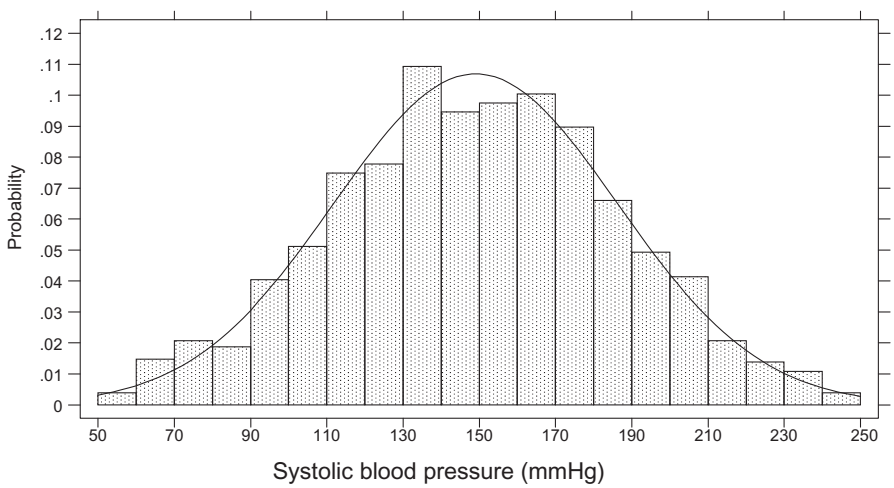


Normal distribution function:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[\frac{-(x-\mu)^2}{2\sigma^2}\right] \qquad \sigma > 0, -\infty < \mu < \infty, -\infty < x < \infty$$

Properties of normal distribution (curve):

The curve has a single peak at the center; this peak occurs at the mean ($\mu$).
The curve is symmetrical about the mean.
The curve never touches the horizontal axis.
The total area under the curve is equal to 1.
The width or shape of the curve is described by the variance ($\sigma^2$), the squared root of which is the standard deviation ($\sigma$).

***FIGURE 1*** *Histogram and fitted normal distribution curve for systolic blood pressures from 1000 subjects.*

equation shown with the curve. This distribution is one of the most important distributions in statistics and is known as the Gaussian distribution [12].

If we have enough values with smaller bars, the heights of these bars would form the bell-shaped curve shown superimposed on the chart. This curve is defined by the mean, or central value ($\mu$), of 149.044 mmHg, and by its spread, or standard deviation ($\sigma$), of 37.317 mmHg.

### 3.2 Properties of Normal Distribution

The properties of a normal distribution are illustrated in the legend of Figure 1. The standard deviation helps to describe the spread of the observations, with about 68% of all observations being captured within one standard deviation at either side of the mean and about 95% of all observations captured within two standard deviations at either side of the mean.

## 4 PRINCIPLES OF STATISTICAL INFERENCE

Let us suppose that it is necessary to measure the average systolic blood pressure (SBP) level of all males aged ≥16 years in the United Kingdom in 2005. For practical and financial reasons, it is not possible to directly measure the SBP of every adult male in the United Kingdom. Instead, we can conduct a survey among a subset (or "sample") of 500 males within this population. Through statistical inference, we can measure the properties of the sample (such as the mean and standard deviation) and use these values to infer the properties of the entire UK adult male population [11–13].

Population properties are usually determined by population parameters (numerical characteristics of a population) that are fixed and usually unknown quantities, such as the mean ($\mu$) and standard deviation ($\sigma$) in a normal distribution $N(\mu,\sigma^2)$ [14, 15]. The statistical properties of the sample, such as the mean ($\bar{X}$) and standard deviation ($S$), can be used to provide estimates of the corresponding population parameters. Conventionally, Greek letters are used to refer to population parameters, while the Roman alphabet is used to refer to sample estimates.

Two strategies that are often used to make statistical inference are [7, 13, 14]:

1. Hypothesis testing
2. Confidence intervals (CIs)

### 4.1 Hypothesis Testing

Statistical inference can be made by performing a hypothesis (or significance) test, which involves a series of statistical calculations [7, 12–14]. In the sample of 500 adult males, the mean SBP ($\bar{X}$) was 130 mmHg, with a standard deviation ($S$) of 10 mmHg. The empirical estimate for the mean SBP of this population from previous medical literature is reported as 129 mmHg (denoted by $\mu_0$). So, we want to know whether there is any evidence that the mean SBP value for all adult males in the United Kingdom in 2005 ($\mu$) is different from 129 mmHg ($\mu_0$).

We start by stating a hypothesis that the population mean SBP for all adult men in 2005 is 129 mmHg, or $\mu = \mu_0$ (i.e., no different to that reported in the medical literature). This is referred to as the null hypothesis and is usually written as $H_0$, representing a theory that has been put forward as a basis for argument [7, 12–14]. The hypothesis test is a means to assess the strength of evidence against this null hypothesis of no difference.

The alternative hypothesis, usually written as $H_a$, is that the mean SBP for the study population is not equal to the specified value, that is, $\mu \neq \mu_0$. Note that under the alternative hypothesis, the 2005 population mean could be higher or lower than the reference mean. The statistical test for the above hypotheses is usually referred to as a two-sided test.

Once the null hypothesis has been chosen, we need to calculate the probability that, if the null hypothesis is true, the observed data (or data that were more extreme) could have been obtained [7, 12–14]. To reach this probability, we need to calculate a test statistic from the sample data (e.g., $\bar{X}$, $S$, and $n$ for quantitative outcomes) using an appropriate statistical method. This test statistic is then compared to the distribution (e.g., the normal distribution) implied by the null hypothesis to obtain the probability of observing our data or more extreme data.

For the SBP data, given the relatively large sample size, we can use the $Z$ test to calculate the value of the test statistic $Z$. The $Z$ test is expressed by the following formula [7, 12–14]:

$$Z = \frac{\bar{X} - \mu_0}{S/\sqrt{n}}$$

This statistic follows a normal standard distribution under the null hypothesis [7, 12–14]. For the SBP data:

- $\bar{X} = 130$ mmHg
- $S = 10$ mmHg
- $n = 500$
- $\mu_0 = 129$ mmHg

Replacing the values in the formula generates $Z = 2.24$. A variety of statistical methods can be used to address different study questions (e.g., comparing treatment difference in means and proportions). The choice of statistical test will depend on the types of data and hypotheses under question [7, 12–14].

Having obtained the appropriate test statistic (in our example, the $Z$ value), the next step is to specify a significance level. This is a fixed probability of wrongly rejecting the null hypothesis, $H_0$, if it is in fact true. This probability is always chosen by the investigators taking into account the consequences of such an error. That is, the significance level is kept low to reduce the chance of inadvertently making a false claim. The significance level, denoted by $\alpha$, is usually chosen to be 0.05 (5%). The corresponding $Z_{\alpha/2}$ is called the critical value of the $Z$ test. The critical value for a hypothesis test is a threshold with which the value of the test statistic calculated from a sample is compared in order to determine the $P$ value to be introduced next. For example, if $\alpha = 0.05$, we have $Z_{0.05/2} = 1.96$; if $\alpha = 0.01$, we have $Z_{0.01/2} = 2.58$.

A $P$ value is the probability of our result ($Z = 2.24$ for the SBP data) or a more extreme result ($Z < -2.24$ or $Z > 2.24$) being observed, assuming that the null hypothesis is true. The exact $P$ value in the $Z$ test is the probability of $Z \leq -Z_{\alpha/2}$ or $Z \geq Z_{\alpha/2}$, which can always be determined by calculating the area under the curve in two-sided symmetric tails from a statistical table specifically of a normal distribution [7, 14]. For the SBP data, the exact $P$ value = 0.025.

In a practical application, we often need to determine whether the $P$ value is smaller than a specified significance level, $\alpha$. This is performed by comparing the value of the test statistic with the critical value. Statistically $P \leq \alpha$ if, and only if, $Z = -Z_{\alpha/2}$ or $Z = Z_{\alpha/2}$. For the SBP data, since $Z = 2.24 > Z_{0.05/2} = 1.96$, we can conclude that $P < 0.05$. A smaller $P$ value indicates that $Z$ is further away from the center (i.e., the null value $\mu - \mu_0 = 0$) and consequently provides stronger evidence to support the alternative hypothesis of a difference.

Although the $P$ value measures the strength of evidence for a difference, which is largely dependent on the sample size, it does not provide the size and direction of that difference. Therefore, in a statistical report, $P$ values should be provided together with CIs (described in detail later) for the main outcomes [7, 14].

We are now in a position to interpret the $P$ value in relation to our data and decide whether there is sufficient evidence to reject the null hypothesis. Essentially, if $P = \alpha$, the prespecified significance level, then there is evidence against the null hypothesis, and we accept the alternative hypothesis stating that there is a statistically significant difference. The smaller the $P$ value, the lower the chance of obtaining a difference as big as the one observed if the null hypothesis were true, and, therefore, the stronger the evidence against the null hypothesis. Otherwise, if $P > \alpha$, there is not sufficient evidence to reject the null hypothesis or there is no statistically significant difference.

For our SBP data, since $P < 0.05$, we can state that there is evidence to reject the null hypothesis of no difference at the 5% significance level, and, therefore, that the mean SBP for the adult male population is statistically significantly different from 129 mmHg. Furthermore, the actual $P$ value equals 0.025, which suggests that the probability of falsely rejecting the null hypothesis is 1 in 40 if the null hypothesis is indeed true. On the other hand, $Z_{0.005} = 2.58 > Z = 2.24$, and $P$ is >0.01. Now we can state that there is no evidence to reject the null hypothesis of no difference if the significance level $\alpha$ is chosen as 0.01.

The implementation of the above procedures for hypothesis testing with the SBP data is summarized in Table 1.

### 4.2 Alpha (Type I) and Beta (Type II) Errors

When testing a hypothesis, two types of errors can occur. To explain these two types of errors, we will use the example of a randomized, double-blind, placebo-controlled clinical trial on a cholesterol-lowering drug A in middle-aged men and women considered to be at high risk for a heart attack. The primary endpoint is the reduction in the total cholesterol level at 6 months from randomization.

The null hypothesis is that there is no difference in mean cholesterol reduction level at 6 months postdose between patients receiving drug A ($\mu_1$) and patients receiving a placebo ($\mu_2$) ($H_0$: $\mu_1 = \mu_2$); the alternative hypothesis is that there is a difference ($H_a$: $\mu_1 \neq \mu_2$). If the null hypothesis is rejected when it is in fact true, then

**TABLE 1 Practical Procedures for Hypothesis Testing**

| Step | Procedure | Illustration with SBP[a] Data |
|---|---|---|
| 1 | Set up a null hypothesis and alternative hypothesis that is of particular interest to study. | $H_0$: $\mu = \mu_0$ (= 129), i.e., population mean SBP is equal to 129 mmHg<br>$H_a$: $\mu \neq \mu_0$, i.e., population mean SBP is different from 129 mmHg |
| 2 | Choose a statistical method according to data type and distribution and calculate its test statistic from the data collected. | $Z = \dfrac{\bar{X} - \mu_0}{S/\sqrt{n}} = 2.24$<br>$\bar{X} = 130$ mmHg<br>$S = 10$ mmHg<br>$n = 500$ |
| 3 | Define a significance level $\alpha$ and its corresponding critical value. | $\alpha = 0.05$ and $Z_{\alpha/2} = 1.96$<br>$\alpha = 0.01$ and $Z_{\alpha/2} = 2.58$ |
| 4 | Determine the $P$ value by comparing the test statistic and the critical value, or calculate the exact $P$ value. | Since $Z = 2.24 > 1.96$, $P < 0.05$<br>Since $Z = 2.24 < 2.58$, $P > 0.01$<br>Exact $P$ value = 0.025 |
| 5 | Make your conclusion according to the $P$ value. | As $0.01 < P < 0.05$, there is evidence to reject the null hypothesis of no difference at the 5% level of significance, but there is no evidence to reject the null hypothesis at the 1% level. The $P$ value of 0.025 means that the probability of falsely rejecting the null hypothesis is 1 in 40 if the null hypothesis is true |

[a]SBP, systolic blood pressure.

a type I error (or false-positive result) occurs. For example, a type I error is made if the trial result suggests that drug A reduced cholesterol levels when in fact there is no difference between drug A and placebo. The chosen probability of committing a type I error is known as the significance level [7, 12–14]. As discussed above, the level of significance is denoted by $\alpha$. In practice, $\alpha$ represents the consumer's risk [2], which is often chosen to be 5% (1 in 20).

On the other hand, if the null hypothesis is not rejected when it is actually false, then a type II error (or false-negative result) occurs [7, 12–14]. For example, a type II error is made if the trial result suggests that there is no difference between drug A and placebo in lowering the cholesterol level when in fact drug A does reduce the total cholesterol. The probability of committing a type II error, denoted by $\beta$, is sometimes referred to as the manufacturer's risk [2]. The power of the test is given by $1 - \beta$, representing the probability of correctly rejecting the null hypothesis when it is in fact false. It relates to detecting a prespecified difference.

### 4.3 Confidence Intervals

The second strategy for making statistical inference is through the use of CIs. In making inference about a population, we might want to know the likely value of the unknown population parameter [e.g., mean ($\mu$), proportion]. This is estimated from the sample statistics. For example, mean ($\bar{X}$), and we call $\bar{X}$ a point estimate of $\mu$.

In addition, we might want to provide some measure of our uncertainty as to how close the sample mean is to the true mean. This is done by calculating a CI (or interval estimate)—a range of values that has a specified probability of containing the true population parameter being estimated. For example, a 95% CI for the mean is usually interpreted as a range of values containing the true population mean with a probability of 0.95 [2]. The formula for the $(1-\alpha)\%$ CI around the sample mean ($\bar{X}$) corresponding to the $Z$ test, is given by

$$\bar{X} \pm Z_{\alpha/2}\mathrm{SE}(\bar{X})$$

where $\mathrm{SE}(\bar{X})$ is the standard error of $\bar{X}$, calculated by $S/\sqrt{n}$. This is a measure of the uncertainty of a single sample mean ($\bar{X}$) as an estimate of the population mean [7]. This uncertainty decreases as the sample size increases. The larger the sample size, the smaller the standard error. Therefore, the narrower the interval, the more precise the point estimate.

For our SBP example, the 95% CI for the population mean ($\mu$) can be calculated with the following formula:

$$\bar{X} \pm 1.96S/\sqrt{n} = 129.1\text{–}130.9\,\mathrm{mmHg}$$

This means that the interval between 129.1 and 130.9 mmHg has a 0.95 probability of containing the population mean $\mu$. In other words, we are 95% confident that the true population mean is between 129.1 and 130.9 mmHg, with the best estimate being 130 mmHg.

Confidence intervals can be calculated not just for a mean but also for any estimated parameter depending on the data types and statistical methods used [12, 14]. For example, one could estimate the proportion of people who smoke in a population, or the difference between the mean SBP in subjects taking an antihypertensive drug and those taking a placebo.

### 4.4 Relationship between Significant Testing and Confidence Intervals

When comparing, for example, two treatments, the purpose of significance testing is to assess the evidence for a difference in some outcome between the two groups, while the CI provides a range of values around the estimated treatment effect within which the unknown population parameter is expected to be with a given level of confidence.

There is a close relationship between the results of significance testing and CIs. This can be illustrated using the previously described $Z$ test for the SBP data analysis. If $H_0$: $\mu = \mu_0$ is rejected at the $\alpha\%$ significance level, the corresponding $(1-\alpha)\%$ CI will not include $\mu_0$. On the other hand, if $H_0$: $\mu = \mu_0$ is not rejected at the $\alpha\%$ significance level, then $(1-\alpha)\%$ CI will include $\mu_0$.

For the SBP data of adult males, the significance test shows that $\mu$ is significantly different from $\mu_0$ (= 129 mmHg) at the 5% level, and the 95% CI (= 129.1–130.9 mmHg) did not include 129 mmHg. On the other hand, the difference between $\mu$ and $\mu_0$ is not significant at the 1% level; the 99% CI $[129 \pm (2.58 \times 10)/\sqrt{500} = 128.8\text{–}131.2\,\mathrm{mmHg}]$ for $\mu$ does indeed contain $\mu_0$. Further information about the proper use of the above two statistical methods can be found in [7, 16].

### 4.5 Examples

Let us assume that four randomized, double-blind, placebo-controlled trials are conducted to establish the efficacy of two weight loss drugs (A and B) against placebo, with all subjects, whether on a drug or placebo, receiving similar instructions as to diet, exercise, behavior modification, and other lifestyle changes. The primary endpoint is the weight change (in kilograms) at 2 months from baseline. The difference in the mean weight change between an active drug and placebo groups can be considered as weight reduction for the active drug against placebo. Table 2 presents the results of hypothesis tests and CIs for the four hypothetical trials. The null hypothesis for each trial is that there is no difference between the active drug treatment and placebo in mean weight change. In trial 1 of drug A, the reduction of drug A over placebo was 6 kg with only 40 subjects in each group. The *P* value of 0.074 suggests that there is no evidence against the null hypothesis of no effect of drug A at the 5% significance level. The 95% CI shows that the results of the trial are consistent with a difference ranging from a large reduction of 12.6 kg in favor of drug A to a reduction of 0.6 kg in favor of placebo.

The results for trial 2 among 400 patients, again for drug A, suggest that mean weight was again reduced by 6 kg. This trial was much larger, and the *P* value ($P < 0.001$) shows strong evidence against the null hypothesis of no drug effect. The 95% CI suggests that the effect of drug A is a greater reduction in mean weight over placebo of between 3.9 and 8.1 kg. Because this trial was large, the 95% CI was narrow and the treatment effect was therefore measured more precisely. In trial 3, for drug B, the reduction in weight was 4 kg. Since the *P* value was 0.233, there was no evidence against the null hypothesis that drug B has no statistically significant benefit effect over placebo. Again this was a small trial with a wide 95% CI, ranging from a reduction of 10.6 kg to an increase of 2.6 kg for the drug B against the placebo.

The fourth trial on drug B was a large trial in which a relatively small, 2-kg reduction in mean weight was observed in the active treatment group compared with the placebo group. The *P* value (0.008) suggests that there is strong evidence against the null hypothesis of no drug effect. However, the 95% CI shows that the reduction is as little as 0.5 kg and as high as 3.5 kg. Even though this is convincing statistically, any recommendation for its use should consider the small reduction achieved alongside other benefits, disadvantages, and cost of this treatment.

Key points from the four trials are summarized in Table 3.

**TABLE 2 Point Estimate and 95% Confidence Interval (CI) for Difference in Mean Weight Change from Baseline between the Active Drug and Placebo Groups in Four Hypothetical Trials of Two Weight Reduction Drugs**

| Trial | Drug | No. of Patients per Group | Difference in Mean Weight Change from Baseline (kg) between the Active Drug and Placebo Groups | Standard Deviation of Difference | Standard Error of Difference | 95% CI for Difference | *P* value |
|---|---|---|---|---|---|---|---|
| 1 | A | 40 | –6 | 15 | 3.4 | –12.6, 0.6 | 0.074 |
| 2 | A | 400 | –6 | 15 | 1.1 | –8.1, –3.9 | <0.001 |
| 3 | B | 40 | –4 | 15 | 3.4 | –10.6, 2.6 | 0.233 |
| 4 | B | 800 | –2 | 15 | 0.8 | –3.5, –0.5 | 0.008 |

**TABLE 3 Summary of Key Points from Results Described in Table 2**

| Key Points from Significance Test and CI | Examples |
|---|---|
| In a small study, a large *P* value does not mean that the null hypothesis is true—absence of evidence is not evidence of absence | Trials 1 and 3 |
| A large study has a better chance of detecting a given treatment effect than a small study and is therefore more powerful | Trials 2 and 4 |
| A small study usually produces a CI for the treatment effect that is too wide to allow any useful conclusion | Trials 1 and 3 |
| A large study usually produces a narrow CI, and therefore a precise estimate of treatment effect | Trials 2 and 4 |
| The smaller the *P* value, the lower the chance of falsely rejecting the null hypothesis, and the stronger the evidence for rejecting the null hypothesis | Trials 2 and 4 |
| Even if the *P* value shows a statistically significant result, it does not mean that the treatment effect is clinically significant. The clinical importance of the estimated effects should always be assessed | Trial 4 |

## 5 COMPARISON OF TWO MEANS

The statistical methods used to compare two means of different treatment groups depend on how the two means were obtained. Data may be obtained from paired or unpaired samples. Paired samples occur when the source of the observations in the first sample are from the same source as the observations of a second or repeat sample. For quantitative data this means that repeated observations are from the same person using a before and after treatment method. Unpaired samples occur when individual observations in one group are independent of individual observations in the other. That is to say that the groups of subjects are different.

The most commonly used statistical test for the comparison of two means is the *t test*. The *t* test compares the means of two sets of continuous variables and expresses the probability that any differences are due to the play of chance (i.e., the null hypothesis is supported), or that the differences may be "real" (i.e., the alternative hypothesis is supported). This probability is often determined by a *t* value. The *t* tests can be used to compare two sample means for both paired and unpaired data, but formulas for calculating *t* values are different [12, 14]. An example of a *t* test for unpaired data is shown in Table 4.

From Table 4, we see that the *t* value of 2.055 corresponds to an observed *P* value of 0.040. The *P* value is the probability of observing the difference (3.440) or even larger (e.g., absolute difference >3.440) from the samples if the null hypothesis were true. The null hypothesis is that the difference in means of the two populations is zero. If there were no difference in systolic blood pressure between two treatment groups, there would be a small chance ($P = 0.040$) that we would observe the difference we did. We can turn this around and say that it is more likely that the two treatment groups differ. We say: "The difference in means between two treatment groups is statistically significant" since the observed *P* value is lower than our significance threshold value of $P = 0.05$.

The *t* test can be performed equivalently by calculating a confidence interval for the difference in means. A confidence interval is a range of values within which the "true" population parameter (such as difference in two means) is likely to lie. Usually 95% confidence limits are quoted, which implies that there is 95%

**TABLE 4 Example of Comparison of Two Means from Independent Samples—*t* Test**

An example calculation in a study to evaluate the effects of a treatment to reduce blood pressure are given. A randomized placebo-controlled trial comparing the effect of an antihypertensive therapy against placebo was conducted. The primary endpoint is systolic blood pressure at 6 months after randomization. The results of the two groups of patients were as follows:

Results of Blood Pressures

- In the control group of 548 patients the mean systolic blood pressure was 150.22 with a standard deviation of 27.12 mmHg
- In the treatment group of 550 patients the mean systolic blood pressure was 146.78 with a standard deviation of 28.32 mmHg

Using these data the *t* test can be calculated in the following manner:

Suppose two populations of sizes $n_1$ and $n_2$ have blood pressure results forming normal distributions with means $\bar{X}_1$ and $\bar{X}_2$ and standard deviations $S_1$ and $S_2$, respectively. Then the efficacy of the treatment, as compared to placebo, can be examined by testing the following hypotheses:

$$H_0 : \mu_1 = \mu_2$$

$$H_1 : \mu_1 \neq \mu_2$$

The statistic ($t$) for the above test is given by

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{[(n_1-1)S_1^2 + (n_2-1)S_2^2]}{(n_1+n_2-2)}}\sqrt{1/n_1 + 1/n_2}}$$

Therefore the null hypothesis can be rejected ($H_0$: $\mu_1 = \mu_2$) at the $\alpha$ level of significance if $|t| \geq t(\alpha/2, n_1 + n_2 - 2)$, where $t(\alpha/2, n_1 + n_2 - 2)$ is the upper ($\alpha/2$)th percentile of the $t$ distribution with $n_1 + n_2 - 2$ degrees of freedom.

In this case, we have $n_1 = 548$, $n_2 = 550$, $\bar{X}_1 = 150.22$, $\bar{X}_2 = 146.78$, $S_1 = 27.12$, $S_2 = 28.32$. Substituting the above sample statistics into the $t$-test formula, we obtain $t = 2.055$, which equates to a probability of 0.040 that the two populations are the same. The estimated difference in means together with its 95% confidence interval is 3.440 [0.156, 6.724].

As $P = 0.040 < 0.05$ (or 95% confidence interval does not contain 0), we can say that there is a statistically significant difference in the systolic blood pressures between the two treatment groups. In other words, the drug appears to reduce systolic blood pressure compared to placebo, and the result has a low likelihood of arising by chance.

confidence in the statement that the "true" population parameter will lie somewhere between the upper and lower limits. For the Table 4 data, the estimated 95% confidence interval for the difference in means is [0.156, 6.724]. If the 95% confidence interval does not contain zero, we can say that there is a statistically significant difference in means between two populations. For Table 4 data, as the lower limit of the 95% confidence interval is greater than zero, we can say that two means are statistically significantly different.

Four key assumptions are required in the two-sample $t$ test. Firstly, we assume that the two treatment group populations from which the samples are drawn are distributed normally [8, 12, 14]. Second, we assume that the variances (or standard deviations) of the two populations are equal [8, 12, 14], that is, $\sigma_1^2 = \sigma_2^2 = \sigma^2$. The equality of variances assumption can be formally verified with an $F$ test [8, 12, 14]. We can also perform an informal check by examining the relative magnitude of the two-sample variances $S_1^2$ and $S_2^2$. For example, if $S_1^2$ / $S_2^2$ is considerably different

from 1, then the assumption that $\sigma_1^2 = \sigma_2^2 = \sigma^2$ will be in doubt. In cases where $\sigma_1^2 \neq \sigma_2^2$, we need to use a modified *t* test or nonparametric method [8, 12, 14, 17]. Third, we assume that the observations in the two treatment groups are independent of each other, that is, no observation in one group is influenced by another observation in the second group [8, 12, 14, 17]. In other words, the value in one treatment group is not affected by that in another group. Finally, we assume that the two populations are homogeneous in terms of the observed and unobserved characteristics of patients (i.e., free from confounding) [7]. These characteristics might be demographics (such as age), prognosis (such as clinical history or disease severity), or baseline measurements of outcome variables.

Although we might never know the unobservable heterogeneity (differences) between two populations, we can assess whether the two populations are comparable by examining the observed summary statistics, such as means or proportions at baseline by treatment. This is why a table that summarizes the baseline information in a clinical trial by treatment group is always provided in a clinical report. If the two treatment groups are not balanced with regard to some of the predictors of outcome, covariate adjustment by means of stratification or regression modeling can be employed [1, 14, 17].

## 6 COMPARISON OF TWO PROPORTIONS

Quite often, we have a binary variable result or proportion to analyze from both groups. For example, in a clinical trial comparing a new treatment to reduce mortality after myocardial infarction, the primary endpoint is a binary outcome of death or survival. The numbers of subjects who die or survive in each of the two treatment groups form a $2 \times 2$ contingency table, from which we can calculate the rate of death or proportion dead for each treatment group [9, 12, 14].

The most common approaches for comparing two proportions are the chi-squared ($\chi^2$) test and the Fisher exact test. The $\chi^2$ test involves determining the expected number of deaths in the new and standard treatment arms and then comparing these to those observed numbers. The test statistic used to assess these differences in deaths can be expressed as $\chi^2 = \Sigma(O - E)^2/E$, where $O$ represents the observed frequencies and $E$ the expected frequencies in each cell of the $2 \times 2$ table. Under the null hypothesis, $\chi^2$ should follow a chi-squared distribution with one degree of freedom. The Fisher exact test is a little more complex and consists of evaluating the sum of probabilities associated with the observed frequency table and all possible two-by-two tables that have the same row and column totals as the observed data [9, 12, 14]. The $\chi^2$ test can also be used to compare more than two proportions [9, 12, 14].

When the total study size is large (say over 200), a test for the difference between two proportions also uses a normally distributed test statistic $Z$ (known as $Z$ test), which can be easily calculated by hand (Table 5).

The null hypothesis in Table 5 is that there is no difference between the two treatment groups in the death rate among patients after myocardial infarction. The $P$ value (<0.001) gives us the probability of observing the difference seen in the death rates (–0.028) if the null hypothesis were true. As the $P$ value is far less than 0.001, we can reject the null hypothesis and say that the difference in proportions is statistically significant.

**TABLE 5 Example of Comparison of Two Proportions—*Z* test**

A randomized placebo-controlled clinical trial was conducted to evaluate a cardiovascular drug on reducing mortality on patients with myocardial infarction (MI). The primary endpoint was the death after MI at 30 days after randomization. Following data were obtained from the trial:

- Group 1 (treatment) 110 deaths out of a total of 2045 patients
- Group 2 (placebo) 165 deaths out of a total of 2022 patients

Statistical inference of binary data such as death from parallel two-group trial involves testing the hypotheses regarding the difference in proportions with a yes response between the treatment and the placebo group. This can be expressed as

$$H_0 : \pi_1 = \pi_2$$

$$H_a : \pi_1 \neq \pi_2$$

The statistic for the above test is given by

$$Z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}_c(1-\hat{p}_c)(1/n_1 + 1/n_2)}}$$

where $n_1$ and $n_2$ are the sample sizes (2045 and 2022), $X_1$ and $X_2$ the number of events in the two groups (110 and 165), respectively, and $\hat{p}_1 = X_1/n_1$ (= 110/2045 = 0.054), $\hat{p}_2 = X_2/n_2$ (= 165/2022 = 0.082), and $\hat{p}_c = (X_1 + X_2)/(n_1 + n_2)$ (= [(110 + 165)/(2045 + 2022)] = 0.068).

Therefore, the null hypothesis can be rejected ($H_0$: $\pi_1 = \pi_2$) at the $\alpha$ level of significance if $|Z| \geq Z_{\alpha/2}$, where $Z_{\alpha/2}$ is the upper ($\alpha$/2)th percentile of the standard normal distribution.

In this case the $Z = -3.548$, which equates to a probability of $P < 0.001$. The estimated difference is –0.028 (95% confidence interval: [–0.043, –0.012]).

The above results suggest that patients in the treatment group have a statistically significant lower death rate than patients in the placebo group.

Alternatively, we can calculate the confidence interval for the difference between the two rates. For the Table 5 data, the observed difference is –0.028 (–2.8%) with a 95% confidence interval [–4.3%, –1.2%]. From these results, we can say that there is a 95% probability that the difference that would be seen, if the entire population of all myocardial infarction patients were tested with either the new or standard treatment, would be between –4.3 and –1.2%. As the 95% confidence interval does not include zero, we can say that rate of death in the treatment group would be lower than that for standard treatment.

There is one common important assumption for the $\chi^2$ *t* test, the Fisher exact test, and the *Z* test: The two treatment groups are homogeneous in terms of the patients' characteristics (e.g., demographic information, disease-related risk factors, medical histories, and concurrent medical treatments). We can check whether the two groups are comparable by examining the observed baseline summary statistics. If the two treatment groups are not balanced with regard to some predictor(s) of outcome, logistic regression modeling could be used to adjust for these potential confounding factors [7, 9].

## 6.1 Assessing Size of Treatment Effect in Two-Arm Trial

In the previous sections, we have introduced different methods for assessing whether there is any evidence against the null hypothesis of no difference. In this section, we describe three commonly used measurements for assessing the size of any treatment effect for a binary outcome.

*Risk Difference* The difference in the proportion of the outcomes between two groups, $p_1 - p_2$, is called the risk difference. In the case of the myocardial infarction (MI) trial data, the risk difference is the risk of death between the active drug and placebo, that is, 5.4% – 8.2% = –2.8%. This means that the estimated absolute risk of death is 2.8% lower (about 3 in 100 MI patients) in the active drug group compared to the placebo group. Statistical inferences about the risk difference, such as point estimate and CIs, can be made by means of a $Z$ test, as described in the last section.

*Risk Ratio* The *risk ratio* is the ratio of the risks in the active drug treatment group compared to the placebo group. The risk ratio is often abbreviated to RR, and is also sometimes called the relative risk:

$$\mathrm{RR} = \frac{p_1}{p_2} = \frac{X_1/n_1}{X_2/n_2}$$

For the MI trial, the RR = 0.66 (5.4%/8.2%), meaning that the risk of death for the patients in the active drug treatment group is 66% of the risk in the placebo group. Equivalently, we could say that the drug treatment is associated with a 34% (100% – 66%) reduction in mortality at 30 days.

*Odds Ratio* A third measure of treatment effect is the *odds ratio* (OR). The odds of an outcome event are calculated as the number of events divided by the number of nonevents. For example, in the active treatment arm in the MI trial, the number of deaths is 110 and the number of survivals is 1935, so the odds of death are 110/1935 = 0.057. If the odds of an event are >1, the event is more likely to happen than not. In particular, the odds of an event that is certain to happen are infinite, and the odds of an impossible event are zero. The OR is calculated by dividing the odds in the active treatment group by the odds in the placebo group.
For the MI trial data, OR is calculated as $(110 \times 1857)/(1936 \times 165) = 0.64$, meaning that the odds of deaths after MI in the drug group is 64% of the odds in the placebo group. Clinical trials typically study treatments that reduce the proportion of patients with an event or equivalently have an OR < 1. In these cases, a percentage reduction in OR is often quoted instead of the OR itself. For the preceding OR, we can say that there is a 36% (100% – 64%) reduction in the odds of deaths in the active treatment group.

## 7 SURVIVAL ANALYSIS

In many clinical trials, the primary outcome is not just whether an event occurs but also the time it takes for the event to occur. For example, in a cancer study comparing the relative merits of surgery and chemotherapy treatments, the outcome measured could be the time from the start of therapy to the death of the subject. In this case the event of interest is death, but in other situations it might be the end of a period spent in remission from cancer spread, relief of symptoms, or a further admission to hospital. These types of data are generally referred to as time-to-event data or survival data even when the endpoint or the event being studied is something

other than the death of a subject. The term "survival analysis" encompasses the methods and models for analyzing such data representing time free of events of interest.

### 7.1 Example: Pancreatic Cancer Trial

The death rate from pancreatic cancer is among the highest of all cancers. A randomized controlled clinical trial was conducted on 36 patients diagnosed with pancreatic cancer. The aim of this trial was to assess whether the use of a new treatment (A) could increase the survival of patients compared to the standard treatment (B). Patients were followed-up for 48 months and the primary endpoint was the time, in months, from randomization to death. Table 6 displays the survival data for the 36 patients. We will use this example to illustrate some fundamental survival analysis methods and their applications.

### 7.2 Basic Concepts in Survival Analysis

*Censoring* In survival analysis, not all subjects are involved in the study for the same length of time due to *censoring*. This term denotes when information on the outcome status of a subject stops being available. This can be because the patient is lost to follow-up (e.g., they have moved away) or stops participating in the study, or because the end of study observation period is reached without the subject having an event. Censoring is a nearly universal feature of survival data. Table 7 summarizes

**TABLE 6 Survival Data for 36 Patients with Pancreatic Cancer in Trial of New Treatment versus Standard Treatment**

| New Treatment | | Standard Treatment | |
|---|---|---|---|
| Survival Time (months) | Survival Status (0 = survival, 1 = dead) | Survival Time (months) | Survival Status (0 = survival, 1 = dead) |
| 2 | 0 | 3 | 0 |
| 5 | 0 | 5 | 1 |
| 10 | 1 | 6 | 0 |
| 12 | 1 | 7 | 1 |
| 15 | 0 | 8 | 1 |
| 27 | 1 | 10 | 1 |
| 36 | 0 | 11 | 0 |
| 36 | 0 | 12 | 1 |
| 37 | 1 | 13 | 0 |
| 38 | 0 | 15 | 1 |
| 39 | 0 | 16 | 0 |
| 41 | 0 | 23 | 1 |
| 42 | 0 | 30 | 1 |
| 44 | 0 | 39 | 1 |
| 45 | 1 | 40 | 1 |
| 46 | 0 | 45 | 1 |
| 48 | 0 | 48 | 0 |
| 48 | 0 | 48 | 0 |

**TABLE 7 Reasons for Censoring Observations in Clinical Trials**

| Reason | Example |
|---|---|
| Lost to follow-up | Patient moved away or did not wish to continue participation. |
| Patient withdrawn | Patient withdraws from the study due to side effects. |
| Patient has an outcome that prevents the possibility of the primary endpoint (competing risk) | Death from cancer where death from cardiac causes is the primary endpoint. |
| Study termination | All patients who have not died are considered censored at the end of the study. |

the main reasons for censoring that can occur in a clinical trial. Survival analysis takes into account censored data, and, therefore, uses the information available from a clinical trial more fully.

*Survival Function and Hazard Function* In survival analysis, two functions are of central interest, namely *survival function* and *hazard function* [18, 19]. The survival function, $S(t)$, is the probability that the survival time of an individual is greater than or equal to time $t$. Since $S(t)$ is the probability of surviving (or remaining event-free) to time $t$, $1 - S(t)$ is the probability of experiencing an event by time $t$. Plotting a graph of probability against time produces a *survival curve*, which is a useful component in the analysis of such data.

The hazard function, $h(t)$, represents the instantaneous event rate at time $t$ for an individual surviving to time $t$ and, in the case of the pancreatic cancer trial, it represents the instantaneous death rate. With regard to numerical magnitude, the hazard is a quantity that has the form of "number of events per time unit" (or "per person-time unit" in an epidemiological study). For this reason, the hazard is sometimes interpreted as an incidence rate. To interpret the value of the hazard, we must know the unit in which time is measured.

For the pancreatic cancer trial, suppose that the hazard of death for a patient is 0.02, with time measured in months. This means that if the hazard remains constant over one month, then the death rate will be 0.02 deaths per month (or per person-months). In reality, the 36 patients contributed a total of 950 person-months and 16 deaths. Assuming that the hazard is constant over the 48-month period and across all patients, an estimate of the overall hazard is $16/950 = 0.017$ deaths per person-months.

*Kaplan–Meier Method* The Kaplan–Meier (KM) approach estimated the proportion of individuals surviving (i.e., who have not died or had an event) by any given time in the study [18, 19]. When there is no censoring in the survival data, the KM estimator is simple and intuitive. $S(t)$ is the probability that an event time is greater than $t$. Therefore, when no censoring occurs, the KM estimator, $\hat{S}(t)$, is the proportion of observations in the sample with event times greater than $t$. For example, if 50% of observations have times >10, we have $\hat{S}(10) = 0.50$.

The KM estimates of the survival curves by the two treatment groups for the pancreatic cancer trial data are displayed in Figure 2. The survival curve is shown in a step function: The curve is horizontal at all times at which there is no event, with a vertical drop corresponding to the change in the survival function at each time, $t_j$, when an event occurs.

In reports, KM curves are usually displayed in one of two ways. The curves can decrease with time from 1 (or 100%), denoting how many people survive (or remain event free). However, in general it is recommended that the increase in event rates is shown starting from 0 (or 0%) subjects, with an increasing curve $(1 - \hat{S}[t])$ unless the event rate is high [20]. Placing the curves for different treatment groups on the same graph allows us to graphically review any treatment differences.

*Log-Rank Test* For the two KM curves by treatment group shown in Figure 2, the obvious question to ask is: Did the new treatment make a difference in the survival experience of the two groups? A natural approach to answering this question is to test the null hypothesis that the survival function is the same in the two groups: that is, $H_0$: $S_1(t) = S_2(t)$ for all $t$, where 1 and 2 represent the new treatment and the standard treatment, respectively.

The above hypothesis can be assessed by performing a log-rank test equivalent to a $\chi^2$ test [18, 19, 21]. The main purpose of this test is to calculate the number of events expected in each treatment group, and to compare this expected number of events with the observed number of events in each treatment group if the null hypothesis is true.

For the pancreatic cancer trial, the resulting $\chi^2$ value is 5.424 [22], converted to a $P$ value of 0.020. As $P < 0.05$, the log-rank test has shown a significant survival difference between the new treatment A and standard treatment B. This test readily generalizes to three or more groups, with the null hypothesis that all groups have the same survival function. If the null hypothesis is true, the test statistic has a chi-

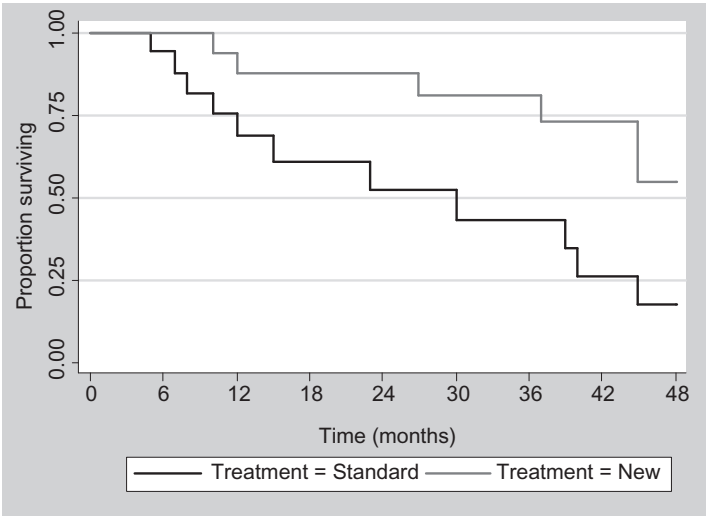


***FIGURE 2*** *Kaplan–Meier survival functions by treatment group for the pancreatic cancer trial data.*

squared distribution with the degrees of freedom equal to the number of groups minus 1.

*Cox Proportional-Hazards Model* The proportional-hazards model relates the hazard function to a number of covariates (such as patient's characteristics at randomization and the treatment received in a clinical trial) as follows [18, 19, 22]:

$$h_i(t) = h_0(t)\exp(b_1x_{1i} + b_2x_{2i} + K + b_px_{pi}) \tag{1}$$

where $x_{ki}$ is the value of the covariate $x_k$ $(k = 1, 2, \ldots, p)$ for an individual $i(i = 1, 2, \ldots, n)$.

The equation says that the hazard for individual $i$ at time $t$ is the product of two factors:

- A baseline hazard function $h_0(t)$ that is left unspecified, except that it cannot be negative.
- A linear function of a set of $p$ fixed covariates, which is then exponentiated.

The baseline hazard function can be regarded as the hazard function for an individual whose covariates all have values of 0 and changes according to time $t$.

This is called a proportional-hazards model because, while the baseline hazard can constantly change over time, the hazard for any individual is assumed to be proportional to the hazard for any other individual and will depend on the individual values of covariates. To see this, let us assume that the model only has one covariate (treatment, $x_{1i}$, $x_{1i} = 0$ for standard treatment and 1 for new treatment). We first calculate the hazards for two individuals 1 and 2 according the Equation (1) and then take the ratio of the two hazards:

$$\begin{aligned} h_1(t) &= h_0(t)\exp(b_1x_{11}) \\ h_2(t) &= h_0(t)\exp(b_1x_{12}) \\ \frac{h_1(t)}{h_2(t)} &= \exp[b_1(x_{11} - x_{12})] \end{aligned} \tag{2}$$

What is important about the equation is that $h_0(t)$ is canceled out of the numerator and denominator. As a result, the ratio of hazards, $\exp[b_1(x_{11} - x_{12})]$, is a constant over time or proportional.

### 7.3 Assessing Size of Treatment Effect in Two-Arm Trial for Survival Data

*Incidence Rate Difference and Ratio* The incidence rate is defined as the number of events divided by the number of units of time [14, 22]. By comparing the incidence rates between treatment groups, we can derive the incidence rate difference and ratios following the procedures described by Kirkwood and Sterne [14, 22]. For the pancreatic cancer trial data, the incidence rates are calculated as 0.9 and 2.9 deaths

per 100 person-months for the new treatment group and the standard treatment group, respectively. The estimates of incidence rate difference and rate ratio together with their 95% CI and $P$ value are as follows:

- Incidence rate difference: –2.0, 95% CI (–3.9, –0.1), $P = 0.034$
- Incidence rate ratio: 0.30, 95% CI (0.08, 0.94), $P = 0.026$

The above results suggest that the new treatment reduces deaths by 2 per 100 person-months, with a 95% CI of 0.1–3.9 per 100 person-months, and that the incidence rate for patients in the new treatment group is only about 30% of the incidence rate for those in the standard treatment group.

Although the incidence rate uses the information on censored observations, it is based on the assumption that the hazard of an event is constant during the study period or has an exponential distribution. In the case of the pancreatic cancer trial, it means that the hazard of death is constant over the 48-month period. However, the risk of an event can change with time. To overcome this problem, the Cox model, which does not require such assumptions, can be used to derive a better measurement for the treatment effect.

*Hazard Ratio* The treatment can be simply measured as a binary covariate (1 for new treatment A and 0 for standard treatment B in the pancreatic cancer trial) and introduced into a Cox proportional-hazards model. In the pancreatic cancer trial, the estimated hazard ratio of death for patients who received treatment A to those who received standard treatment B is 0.31, with 95% CI (0.11, 0.89), $P = 0.030$. This means that the new treatment is estimated to reduce the hazard of death by 69%, with 95% CI (11%, 89%) and the reduction in hazard is statistically significant at the 5% significance level. As there is only one covariate (treatment) in the Cox model, the estimated hazard ratio is called a crude or unadjusted treatment effect. The adjusted hazard ratio for the treatment will be generated if other baseline patient characteristics are introduced in the model.

## 8 COMMONLY USED TERMS IN CLINICAL TRIALS

***ANOVA (Analysis of Variance)*** A statistical method for comparing several means by comparing variances. It concerns a normally distributed outcome (response) variable and a single categorical (predictor) variable representing treatments or groups. ANOVA is a special case of a linear regression model by which group means can be easily compared.

***Bias*** Systematic errors associated with the inadequacies in the design, conduct, or analysis of a trial on the part of any of the participants of that trial (patients, medical personnel, trial coordinators, or researchers), or in publication of its the results, that make the estimate of a treatment effect deviate from its true value. Systematic errors are difficult to detect and cannot be analyzed statistically but can be reduced by using randomization, treatment concealment, blinding, and standardized study procedures.

***Confidence Intervals*** A range of values within which the "true" population parameter (e.g., mean, proportion, treatment effect) is likely to lie. Usually, 95% confidence limits are quoted, implying that there is 95% confidence in the statement that the "true" population parameter will lie somewhere between the lower and upper limits.

***Confounding*** A situation in which a variable (or factor) is related to both the study variable and the outcome so that the effect of the study variable on the outcome is distorted. For example, if a study found that coffee consumption (study variable) is associated with the risk of lung cancer (outcome), the confounding factor here would be cigarette smoking since coffee is often drunk while smoking a cigarette, which is the true risk factor for lung cancer. Thus, we can say that the apparent association of coffee drinking with lung cancer is due to confounding by cigarette smoking (confounding factor). In clinical trials, confounding occurs when a baseline characteristic (or variable) of patients is associated with the outcome, but unevenly distributed between treatment groups. As a result, the observed treatment difference from the unadjusted (univariate) analysis can be explained by the imbalanced distribution of this variable.

***Correlation Coefficient (r)*** A measure of the linear association between two continuous variables. The correlation coefficient varies between –1.0 and +1.0. The closer it is to 0, the weaker the association. When both variables go in the same direction (e.g., height and weight), $r$ has a positive value between 0 and 1.0 depending on the strength of the relationship. When the variables go in opposite directions (e.g., left ventricular function and life span), $r$ has a negative value between 0 and –1.0 depending on the strength of this inverse relationship.

***Covariates*** This term is generally used as an alternative to explanatory variables in regression analysis. However, more specifically refer to variables that are not of primary interest in an investigation. Covariates are often measured at baseline in clinical trials because it is believed that they are likely to affect the outcome variable and, consequently, need to be included to estimate the adjusted treatment effect.

***Descriptive/Inferential Statistics*** Descriptive statistics are used to summarize and describe data collected in a study. To summarize a quantitative (continuous) variable, measures of central location (i.e., mean, median, mode) and spread (e.g., range and standard deviation) are often used, whereas frequency distributions and percentages (proportions) are usually used to summarize a qualitative variable. Inferential statistics are used to make inferences or judgments about a larger population based on the data collected from a small sample drawn from the population. A key component of inferential statistics is hypothesis testing. Examples of inferential statistical methods are $t$ test and regression analysis.

***Endpoint*** Clearly defined outcome associated with an individual subject in a clinical research. Outcomes may be based on safety, efficacy, or other study objectives (e.g., pharmacokinetic parameters). An endpoint can be quantitative (e.g., systolic blood pressure, cell count), qualitative (e.g., death, severity of disease), or time to event (e.g., time to first hospitalization from randomization).

***Hazard Ratio*** In survival analysis, hazard (rate) represents instantaneous event rate (incidence rate) at a certain time for an individual who has not experienced an event at that time. Hazard ratio compares two hazards of having an event between two groups. If the hazard ratio is 2.0, then the hazard of having an event in one group is twice the hazard in the other group. The computation of the hazard ratio assumes that the ratio is consistent over time (proportional hazards assumption).

***Hypothesis Testing or Significance Testing*** Statistical procedure for assessing whether an observed treatment difference was due to random error (chance) by calculating a *P* value using the observed sample statistics such as mean, standard deviation, and so on. The *P* value is the probability that the observed data or more extreme data would have occurred if the null hypothesis (i.e., no true difference) were true. If the calculated *P* value is a small value (like <0.05), the null hypothesis is then rejected, and we state that there is a statistically significant difference.

***Intention-to-Treat Analysis*** A method of data analysis on the basis of the intention to treat a subject (i.e., the treatment regimen a patient was assigned at randomization) rather than the actual treatment regimen he or she received. It has the consequence that subjects allocated to a treatment group should be followed up, assessed, and analyzed as members of that group regardless of their compliance to that therapy or the protocol, irrespective of whether they later crossed over to the other treatment group or not or whether they discontinued treatment.

***Kaplan–Meier Estimate and Survival Curve*** A survival curve shows an estimate of the fraction of patients who survive over the follow-up period of the study without an event of interest (e.g., death). The Kaplan–Meier estimate is a simple way of computing the survival curve taking into account patients who were lost to follow-up or any other reasons for incomplete results (known as censored observations). It usually provides a staircase graph of the fraction of patients remaining free of event over time.

***Meta-Analysis*** The systematic review and evaluation of the evidence from two or more independent studies asking the same clinical question to yield an overall answer to the question.

***Normal Distribution*** A bell-shaped symmetric distribution for a continuous variable having highest frequency at a mean value and less frequency further away from this mean value. A normal distribution can be completely described by two parameters: mean ($\mu$) and variance ($\sigma^2$). In the special case of $\mu = 0$ and $\sigma^2 = 1$, it is called the standard normal distribution.

***Number Needed to Treat (NNT)*** This term is often used to describe how many patients would need to be given a treatment to prevent one event. It is determined from the absolute difference between one treatment and another. In a randomized study the group receiving treatment A had a death rate of 12.5%, and the group receiving treatment B had a death rate of 15.0%. Both groups are matched for size and length of follow-up. Comparing the two treatments there was an absolute risk reduction of 15% − 12.5% = 2.5% for treatment A. From this we can derive that the

NNT (= 1/0.025) is 40. This means 40 patients need to be given treatment A rather than B to prevent 1 additional death.

***Odds Ratio (OR) and Risk Ratio (RR)*** These terms compare the probability of having an event between two groups exposed to a risk factor or treatment. The RR is the ratio of the probability of occurrence of an event between two groups. The OR is the ratio of the ratio of patients with and without an event in each group. If the number of deaths in the treatment and control arms (both of sample size 100) of a randomized study are 50 and 25, respectively, the RR = (50/100) / (25/100) = 2. The treatment group has a twofold relative risk of dying compared with the control group. The OR = (50/50) / (25/75) = 3 indicates that the odds of death in the treatment arm is threefold of the control arm.

***Per-Protocol Analysis*** A method of analysis in which only the subset of subjects who complied sufficiently with the protocol are included. Protocol compliance includes exposure to treatment, availability of measurements, correct eligibility, and absence of any other major protocol violations. This approach contrasts with the more conservative and widely accepted "intention-to-treat" analysis.

***Power*** The probability of rejecting the null hypothesis (e.g., no treatment difference) when it is false. It is the basis of procedures for calculating the sample size required to detect an expected treatment effect of a particular magnitude.

***Random Error*** An unpredictable deviation of an observed value from a true value resulting from sampling variability. It is a reflection of the fact that the sample is smaller than the population; for larger samples, the random error is smaller, as opposed to systematic errors (bias) that keep adding up because they all go in the same direction.

***Regression Analyses*** Methods of explaining or predicting outcome variables using information from explanatory variables. Regression analyses are often used in clinical trials to estimate the adjusted treatment effect taking into account differences in baseline characteristics, and in epidemiological studies to identify prognostic factors while controlling for potential confounders. Commonly used regression models include linear, logistic, and Cox regression methods.

***Risk Factor*** A risk factor can be defined as anything in the environment, personal characteristics, or events that make it more or less likely one might develop a given disease, reach an adverse event or experience a change in health status. For example, raised cholesterol is a risk factor for heart attacks.

***Standard Error*** A measure of the random variability of a statistic (e.g., mean, proportion, treatment effect) indicating how far the statistic is likely to be from its true value. For example, standard error of the mean (SEM) indicates uncertainty of a single sample mean ($\bar{X}$) as an estimate of the population mean ($\mu$). A smaller SEM implies a more reliable estimate of the population mean. Standard error can be used to calculate confidence interval of an estimated population parameter. The

smaller the standard error, therefore, the narrower the confidence interval, and more precise the point estimate of the population parameter.

***Treatment Effect*** An effect attributed to a treatment in a clinical trial, often measured as the difference in a summary measure of an outcome variable between treatment groups. Commonly expressed as difference in means for a continuous outcome, a risk difference, risk ratio, or odds ratio for a binary outcome, and hazard ratio for a time-to-event outcome.

***Univariate/Multivariate Analysis*** The term variate refers to the term variable. A univariate analysis examines the association between a single variable and an outcome variable (correctly called a bivariate analysis), for example, age and occurrence of stroke. In a multivariate analysis, associations between many variables are examined simultaneously. In particular, multivariate regression analysis can be used to assess the relative importance and contribution of each predictor variable to the outcome variable. For example, a multivariate logistic regression can be undertaken to identify the most important prognostic factors among several risk factors (e.g., age, sex, systolic blood pressure, and cholesterol level) that predict the occurrence of stroke.

## 9 CONCLUDING REMARKS

We have introduced some basic statistical approaches to dealing with different types of data and samples. To apply these methods for data analysis, there are a few basic principles we should adhere to. We first need to make the purpose of the analysis clear. For example, is the analysis to compare groups or to estimate associations? Different purposes will require different methods. Second, we need to identify the nature of the data. Are they data from paired samples or two independent samples? Third, we need to examine the characteristics of the data. Are the variables qualitative or quantitative? Symmetrically distributed or not? These properties of the data will guide to use the correct assumptions for analysis. Important as the appropriate use of statistical methods is, we must also ensure that the data are clean, accurate, meaningful, and complete and were gathered reducing any biases occurring in the design, conduct, and analysis of the clinical trial. The advice of a statistician with experience in medical research should always be sought from the design through to completion and analysis phases of a clinical trial.

## REFERENCES

1. Pocock, S. J. (1983), *Clinical Trials: A Practical Approach*, Wiley, Chichester.
2. Chow, S. C., and Liu, J. P. (1998), *Design and Analysis of Clinical Trials: Concept and Methodologies*, Wiley, Chichester.
3. Everitt, B. S., and Pickles, A. (1999), *Statistical Aspects of the Design and Analysis of Clinical Trials*, Imperial College Press, London.
4. Wang, D., and Bakhai, A. (2005), Intention-to-treat Analysis, in Wang, D., and Bakhai, A., Eds., *Clinical Trials: A Practical Guide to Design, Analysis and Reporting*, Remedica, London, pp. 255–264.

5. Wang, D., Clayton, T., and Bakhai, A. (2005), Confounding, in Wang, D., and Bakhai, A., Eds., *Clinical Trials: A Practical Guide to Design, Analysis and Reporting*, Remedica, London, pp. 295–304.
6. Steele, F., and Wang, D. (2005), Regression analyses, in Wang, D., and Bakhai, A., Eds., *Clinical Trials: A Practical Guide to Design, Analysis and Reporting*, Remedica, London, pp. 273–286.
7. Wang, D., Clayton, T., and Yan, H. (2005), Significance tests and confidence intervals, in Wang, D., and Bakhai, A., Eds., *Clinical Trials: A Practical Guide to Design, Analysis and Reporting*, Remedica, London, pp. 185–196.
8. Wang, D., Clemens, F., and Clayton, T. (2005), Comparison of means, in Wang, D., and Bakhai, A., Eds., *Clinical Trials: A Practical Guide to Design, Analysis and Reporting*, Remedica, London, pp. 197–216.
9. Wang, D., Clayton, T., and Clemens, F. (2005), Comparison of proportions, in Wang, D., and Bakhai, A., Eds., *Clinical Trials: A Practical Guide to Design, Analysis and Reporting*, Remedica, London, pp. 217–234.
10. Nitsch, D., Wang, D., Clayton, T., et al. (2005), Multiplicity, in Wang, D., and Bakhai, A., Eds., *Clinical Trials: A Practical Guide to Design, Analysis and Reporting*, Remedica, London, pp. 329–338.
11. Jennison, C., and Turnbull, B. W. (2000), *Group Sequential Methods with Applications to Clinical Trials*, Chapman & Hall/CRC, New York.
12. Altman, D. G. (1999), *Practical Statistics for Medical Research*, Chapman and Hall, London.
13. Zelen, M. (1998), Inference, in Armitage, P., and Colton, T., Eds., *Encyclopedia of Biostatistics*, Wiley, New York, pp. 2035–2046.
14. Kirkwood, B., and Sterne, J. (2003), *Essential Medical Statistics*, 2nd ed., Blackwell, Oxford.
15. Wang, D., Bakhai, A., and Gupta, A. (2005), Types of data and normal distribution, in Wang, D., and Bakhai, A., Eds., *Clinical Trials: A Practical Guide to Design, Analysis and Reporting*, Remedica, London, pp. 167–184.
16. Sterne, J. A., and Davey Smith, G. (2001), Sifting the evidence—what's wrong with significance tests? *BMJ*, 322, 226–231.
17. Lee, A. F. S. (1998), Student's *t* distribution and student's *t* statistics, in Armitage, P., and Colton, T., Eds., *Encyclopedia of Biostatistics*, Wiley, New York, pp. 4396–4397.
18. Collett, D. (2003), *Modelling Survival Data in Medical Research*, 2nd ed., Chapman and Hall, London.
19. Cox, D. R., and Oakes, D. (1984), *Analysis of Survival Data*, Chapman and Hall, London.
20. Pocock, S. J., Clayton, T. C., and Altman, D. G. (2002), Survival plots of time-to-event outcomes in clinical trials: Good practice and pitfalls, *Lancet*, 359, 1686–1689.
21. Bland, J. M., and Altman, D. G. (1998), Survival probabilities (the Kaplan–Meier method), *BMJ*, 317, 1572.
22. Wang, D., Clayton, T., and Bakhai, A. (2005), Analysis of survival data, in Wang, D., and Bakhai, A., Eds., *Clinical Trials: A Practical Guide to Design, Analysis and Reporting*, Remedica, London, pp. 235–254.

# *Explanatory and Pragmatic Clinical Trials*

**Rob Herbert**

*The George Institute for International Health, Sydney, Australia*

**Contents**

1 Pragmatic and Explanatory Perspectives
2 Participants
  2.1 Inclusion and Exclusion Criteria
  2.2 Run-In Periods
3 Sample Size
  3.1 Criteria Used to Calculate Sample Size
  3.2 An Even More Pragmatic Approach: Value of Information
  3.3 Adjusting Sample Size Calculations for Noncompliance
4 Interventions
  4.1 Experimental Intervention
  4.2 Comparison Intervention
5 Choice of Outcomes
6 Analysis
  6.1 Who Should Be Included in Analysis?
  6.2 Adjustment for Noncompliance
7 Reporting
8 Recommendations
  Acknowledgments
  References

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

## 1 PRAGMATIC AND EXPLANATORY PERSPECTIVES

In 1967 Schwartz and Lellouch published an article in the *Journal of Chronic Diseases* (the precursor to the *Journal of Clinical Epidemiology*) titled "Explanatory and pragmatic attitudes in therapeutical trials" [1]. The article was significant because it clearly distinguished, for the first time, different philosophical attitudes of clinical trialists. Since that time the article has become a classic in the field of clinical trial methodology (the 2007 edition of *Web of Science* records 421 citations), and the ensuing discussion has stimulated development of several methodological advances.

Schwartz and Lellouch argued that clinical trialists can have explanatory or pragmatic attitudes (or perspectives, or orientations) to clinical trials. These attitudes influence many aspects of trial design. In the contemporary literature, explanatory trials may be called "efficacy trials" [2, 3]. Pragmatic trials may be called "effectiveness trials" [2, 3] or "practical clinical trials" [4] or, nearly synonymously, "large simple trials" [5].

Explanatory trialists conceive of clinical trials as being clinical analogs of classic laboratory experiments. The aim is to learn of the potential effects of the specific intervention. Explanatory trials are designed to establish "proof of concept" of a new intervention [3] (Table 1).

In contrast, pragmatic trialists think of clinical trials as a mechanism for informing clinical decisions. Pragmatic trials tell us about the importance of interventions in the real world of clinical practice. They are designed to inform clinical decisions about whether to implement a particular intervention.

We need both explanatory and pragmatic trials. It would appear sensible to first conduct explanatory trials on a particular intervention and then subsequently subject the same intervention to a pragmatic trial. An exception might be where health interventions become established clinical practice before they are subject to clinical trials. Arguably, the first trials of such interventions should adopt a pragmatic approach.

Trialists rarely make explicit statements, in trial reports, about whether their intentions are explanatory or pragmatic. Many trials incorporate some features of explanatory trials and some features of pragmatic trials. This suggests that at least some trialists do not clearly think through whether their intentions are explanatory or pragmatic.

**TABLE 1 Explanatory and Pragmatic Attitudes to Clinical Trials**

| Explanatory Attitudes | Pragmatic Attitudes |
|---|---|
| Aim is to learn about effects of an intervention. | Aim is to make a decision about whether to administer an intervention. |
| Interest is in any effect. | Interest is in clinically important effects. |
| Ascertains if an intervention *can* produce effects (proof of concept). | Ascertains if an intervention *does* confer effects. |
| Used to make inferences about specific effects of narrowly defined interventions given under optimal conditions. | Used to make inferences about the real-world effects of complex interventions in routine clinical practice. |

Those trialists who explicitly adopt a particular perspective may find themselves misunderstood by reviewers of grant applications or journals. Sometimes pragmatic trials are criticized for failing to satisfy requirements that are particular to explanatory trials and vice versa. Thus it appears that the issues raised by Schwartz and Lellouch over 40 years ago are still not fully appreciated by clinical trialists.

This chapter discusses some of the consequences of the trialist's perspective (explanatory or pragmatic) for design and analysis of randomized trials. The chapter concludes by making some recommendations for the design, analysis, and reporting of randomized trials.

## 2 PARTICIPANTS

### 2.1 Inclusion and Exclusion Criteria

One of the first issues that confront trialists, when developing a trial protocol, is which patients to include in the trial.

If the intention is to learn whether the intervention *could* be effective (the explanatory perspective), then a sensible course of action would be to recruit those patients in whom effects of intervention are most likely to be apparent. Usually, this involves sampling from a carefully selected and homogenous population. Careful formulation of inclusion and exclusion criteria can ensure that only participants with characteristics likely to be associated with large treatment effects are sampled, and homogeneous sampling can ensure that the variance of outcomes is minimized. Selective, homogenous sampling maximizes the possibility of detecting effects of intervention, but it may also exclude patients who might, in the course of normal clinical practice, be offered the intervention. People with comorbidities, pregnant women, and people who do not speak the local language are typically excluded from explanatory trials.

Pragmatic trialists seek to make inferences about the effects of an intervention on the population for whom the intervention is indicated. Thus pragmatic trials typically sample broadly. Typically, they have broad inclusion criteria and few exclusion criteria. (In some trials the primary inclusion criterion is simply that the relevant clinician might choose to offer the intervention and the participant might choose to accept it. An example of such a trial is the MRC Spinal Stabilisation Trial in which the primary inclusion criterion was that "patients who were candidates for surgical stabilisation of the spine were eligible if the clinician and patient were uncertain which of the study treatment strategies was best" [6].) Pragmatic trials often sample patients under the care of many clinicians and from multiple sites [4].

Ideally, pragmatic trials would do more than sample broadly. Theoretically, at least, pragmatic trials should sample in a representative way from the population about which inferences are to be made. Random sampling is usually difficult or impossible, so there are very few examples of trials that randomly sample participants. However, a small number of studies have made explicit attempts to sample as representatively as possible. Fransen and colleagues describe how they attempted to obtain a representative sample of general medical practitioners and patients in the DIAMOND study of two protocols for management of dyspepsia in primary care [7].

### 2.2 Run-In Periods

In some trials, patients are not randomized at the time they are recruited into the trial. Instead, they are asked to participate in a run-in period, during which their compliance with the intervention (or, more usually, with the placebo intervention) is monitored. Only those patients who demonstrate a sufficiently high level of compliance are subsequently randomized into the trial. Thus the run-in period serves to winnow out potential participants who are unlikely to comply with the intervention specified in the trial protocol.

The use of run-in periods offers several potential benefits to explanatory trialists. First, run-in periods can reduce loss to follow-up. Second, the use of a run-in period potentially increases compliance, enabling better explanatory estimates of effects of intervention (see Section 6). Third, simulation studies show that when there is both a high degree of noncompliance and the run-in period gives an accurate indication of who will and will not comply with the intervention protocol, the use of a run-in period can produce modest increases in statistical efficiency [8]. (That is, slightly fewer patients would need to be recruited to a trial with a run-in period to yield the same level of statistical power as a trial without a run-in period.) However, these benefits are not guaranteed: When compliance is high or the ability to discriminate between compliant and noncompliant patients is low, the statistical efficiency of trials with run-in periods may actually be lower than in trials without a run-in period [8]. Trials with run-in periods require a longer period of participation in the trial, which could paradoxically reduce participation rates and increase rates of noncompliance and loss to follow-up. Thus, while the use of run-in periods may seem attractive to explanatory trialists, they can be counterproductive.

The use of run-in periods is much less attractive to pragmatic trialists because pragmatists are much more tolerant of noncompliance. From the pragmatic perspective, noncompliance is as much a characteristic of the intervention as it is of the patient. If patients do not comply with an intervention, then the intervention is unlikely to be of any real practical effect, regardless of its explanatory effects. Insofar as a pragmatic estimate of the effect of an intervention depends on the degree of compliance with the intervention, a pragmatic trial of the intervention should involve patients with similar levels of compliance to the population about whom inferences are to be made. This means that, theoretically at least, pragmatic trialists should seek to recruit participants with representative levels of compliance. As discussed above, pragmatic trialists have rarely made explicit efforts to sample representatively, so it may be often true that pragmatic trials do not sample patients with representative levels of compliance.

## 3 SAMPLE SIZE

### 3.1 Criteria Used to Calculate Sample Size

The classic hypothesis testing approach to sample size calculation is to calculate the sample size required to provide adequate statistical power to detect an effect of intervention while keeping the probability of falsely claiming an effect to acceptably low levels.

A fairly general form of the equation for estimating sample size in two-armed clinical trials with equally sized groups is

$$n \geq 2\left(z_{1-\alpha/2} + z_{1-\beta}\right)^2 \left(\frac{\mathrm{SD}^2}{\delta^2}\right)$$

where $n$ is sample size per group, $z_{1-\alpha/2}$ is the $z$ score corresponding to the probability, $\alpha$, of falsely rejecting the null hypothesis, $z_{1-\beta}$ is the $z$ score corresponding to the probability, $\beta$, of falsely retaining the null hypothesis, SD is the within-group standard deviation (usually assumed to be the same for both groups), and $\delta$ is the smallest effect worth detecting [9].

This approach to sample size calculation is widely used but, because it is built on a hypothesis testing approach to statistical inference, and because hypothesis testing is becoming increasingly regarded with suspicion by methodologists, it has been strongly criticized (see, e.g., [10]). Here discussion of issues of sample size calculation is restricted to comparison of explanatory and pragmatic approaches to sample size calculation.

The explanatory trialist might wish to detect *any* effect of intervention, no matter how small. This is because the explanatory trialist's interest is in proof of concept rather than in whether the effect of intervention is large enough to make the intervention worth implementing in clinical practice. But it is not possible to design trials to reliably detect *any* effect because the sample sizes required to detect very small effects become impossibly large. The hypothesis testing approach requires the trialist to nominate $\delta$, the smallest effect that is worth knowing about, and calculate the sample size required to detect effects of at least that size. This is problematic for explanatory trialists because in explanatory trials the choice of $\delta$ seems arbitrary. Consequently, sample sizes calculated for explanatory trials seem arbitrary too.

Pragmatic trialists find the task of nominating $\delta$ more straightforward. Pragmatic trials are intended to inform decisions about whether or not to implement an intervention, so they should be designed to reliably detect effects of intervention that are just large enough to warrant the risks, costs, discomforts, and inconveniences of intervention [11, 12]. Recently, Barrett and colleagues have developed the benefit–harm trade-off method to establish patient-focused estimates of the smallest important effect of intervention [13–15]. This method can be used to obtain empirical estimates of $\delta$.

Paradoxically, while pragmatic trialists may find it easier to nominate $\delta$ than explanatory trialists, it has been argued that it is not necessary for pragmatic trialists to nominate $\delta$ [1, 16]. The argument is as follows. The pragmatic trialist does not care about falsely rejecting the null hypothesis because, from the pragmatic trialist's perspective, if the interventions given in the two arms of the trial are equally effective, it does not matter which of the two interventions is administered in clinical practice. Insofar as pragmatic trialists do not care about falsely rejecting the null hypothesis, they should always consider that the intervention with the larger observed effect is the more effective intervention, regardless of any considerations of sampling error. This strategy results in one of three possible scenarios:

1. The two interventions have identical effects, but the trialist claims one is more effective than the other.

2. One intervention is more effective than the other, and the trialist correctly identifies the more effective intervention.
3. One intervention is more effective than the other, and the trialist incorrectly concludes the less effective intervention is the more effective one.

According to Schwartz and Lellouch [1], only the last type of error is of any concern to the pragmatic trialist. They refer to the probability of this kind of error as $\gamma$. As the pragmatic trialist wishes only to minimize $\gamma$, such studies requires only a sample size of

$$n \geq 2\left(z_{1-\gamma}\right)^2\left(\frac{\mathrm{SD}^2}{\delta^2}\right)$$

Even if $\gamma$ is set at just 1%, the sample size calculated in this way will be nearly one-third less than with conventional calculations that assume $\alpha = 0.05$ and $\beta = 0.2$. This observation has been used to defend the conduct of small pragmatic trials [17].

As MacRae has pointed out, Schwartz and Lellouch's approach to sample size calculation for pragmatic trials should only be used when the trialist is prepared to apply an intervention that is no more effective than alternative interventions [16]. In practice, that is often not the case because competing interventions often have quite different risks, inconveniences, and costs. For example, most surgeons would be disinclined to offer spinal stabilization surgery to patients with chronic low back pain if alternatives such as an intensive rehabilitation program have similar effects because the surgery is associated with significantly greater risks, discomfort, inconvenience, and costs than rehabilitation [6].

From a pragmatic perspective, the hypothesis testing approach to sample size calculation is quite unsatisfactory. Pragmatic trials should be designed to assist clinicians and health policymakers make decisions about whether the effects of particular interventions are big enough to be worthwhile. The most useful trials are those that provide precise estimates of the effects of intervention: The more precise the estimate, the more likely it will be that the decision maker will be able to conclude that the intervention definitely is or is not worth implementing [11, 12]. In practice, the hypothesis testing approach to sample size calculation often yields estimates of treatment effects that are too imprecise to be useful for clinical decision making. (That is, confidence intervals or credible intervals are usually so broad that they include both clinically worthless and clinically worthwhile effects.) This suggests an alternative approach to sample size calculation is needed. A better approach may be to explicitly specify the precision required of estimates of effect, and then estimate the sample size required to achieve that precision [18]. Admittedly, the choice of the acceptable precision will usually be quite arbitrary.

### 3.2 An Even More Pragmatic Approach: Value of Information

One response to the issues outlined above is to use decision-theoretic approaches to sample size calculation. Decision-theoretic approaches are explicitly pragmatic because they seek to maximize the difference between the cost of a trial and the expected value of the results.

Willan and Pinto have described a decision theoretic approach to sample size calculation in clinical trials [19]. They take a societal perspective, although the broad approach could be modified to take other perspectives such as the perspective of a corporate entity. They estimate the expected value of a trial (the expected value of sample information), the cost of conducting the trial, and the opportunity cost of treating some patients in the trial with the inferior intervention. All three quantities (the value of sample information and the cost and opportunity costs of the trial) are functions of sample size. The optimal sample size is that which maximizes the value of the trial minus its costs. This approach will often result in radically different estimates of sample size to those obtained with traditional methods. Indeed Willan and Pinto show that in some circumstances optimal trial size is zero, and the intervention should be implemented without first conducting a randomized trial.

Willan and Pinto's [19] approach to sample size calculation is attractive because it is rational and because it does not require the trialist to go through the arbitrary exercise of nominating $\alpha$, $\beta$, and $\delta$. The disadvantage is that the calculations are complex and require that the trialist nominate other difficult-to-estimate quantities such as the number of patients who would, in the future, be eligible to receive the experimental or control interventions.

### 3.3 Adjusting Sample Size Calculations for Noncompliance

As noted above, explanatory trialists try to minimize noncompliance with intervention. From an explanatory trialist's perspective, participants who do not receive the experimental intervention cannot provide information about the effects of that intervention. In most trials, however, some degree of noncompliance is inevitable and explanatory trialists need to decide how to deal with it.

One strategy is to omit noncompliant cases in a "per protocol analysis." If the primary analysis is to be conducted per protocol (though it is not clear that it ever should—see Section 6), the sample size calculations should anticipate the loss of participants due to noncompliance and apply an appropriate correction. In the simplest scenario in which similar proportions of participants are randomly excluded from two equally sized groups, the sample size should be increased by a factor of 1/(1 – anticipated proportion excluded).

A better alternative is to retain all participants in the analysis and apply a correction for attenuation of the explanatory effect by noncompliance (see Section 6). If the primary analysis is to involve obtaining estimates of explanatory effects of intervention calculated in this way, then the sample size should be adjusted accordingly. Newcombe [20] describes a method for adjusting conventional sample size calculations when an explanatory effect of intervention is to be estimated from the difference between two means, and Sato [21] provides sample size formulas for trials in which the explanatory effect of intervention is to be estimated as a risk difference. Sommer and Zeger tabulate the relative efficiency of pragmatic and explanatory estimates of risk ratios [22].

As discussed in Section 2, pragmatic trialists consider that noncompliance is a nonissue. Pragmatic trialists are uninterested in the hypothetical effect that an intervention could have in the presence of complete compliance. To them the parameter of most interest is the average effect of the intervention in clinical populations. It is understood that clinical populations contain some patients who do not

comply with intervention. For this reason pragmatic trialists generally do not adjust sample size estimates for noncompliance.

## 4 INTERVENTIONS

### 4.1 Experimental Intervention

Some health interventions are constrained by tight protocols that ensure uniformity of clinical practice. More often clinical practice is highly variable. Many clinicians have quite a high degree of autonomy, and they tend to develop their own formal or informal protocols for administering particular interventions. Individual clinicians may develop idiosyncratic decision rules that are not shared by other clinicians. In some settings a particular intervention may be administered in a technically adept manner using optimal protocols by clinicians who appear competent and demonstrate empathy. In other settings the same intervention may be administered tardily, incompetently, or indifferently.

The explanatory trialist wants to know about the effects of an intervention when it is administered optimally. Consequently, explanatory trialists expend considerable effort ensuring that the intervention is administered as well as is possible. Usually, the intervention is administered by a clinician with recognized expertise, and the intervention is administered according to strict protocols. This confers an incidental advantage: It ensures that it is clear precisely what intervention has been provided in the trial, so there need be little ambiguity about precisely what sorts of intervention the trial can be used to make inferences about.

The pragmatic trialist's interest is, instead, in the effects of the intervention as it is administered in real practice. (Or, when the trial is of a *new* intervention, the pragmatic trialist's interest is in what the effects of the intervention *would be* if it was to be administered in real practice.) Seen from this perspective, there is no necessity to enforce optimal intervention in clinical trials other than with the usual mechanisms that regulate the quality of clinical care. The interventions should be applied by clinicians who are representative of the spectrum of clinicians who apply (or would apply) the intervention in clinical practice, not exclusively by expert clinicians.

Useful pragmatic estimates of the effects of an intervention may be difficult to obtain when the intervention is new and administration of the intervention is difficult. An important example is new surgical techniques. The problem arises because new surgical techniques are pioneered by surgeons who have specific expertise in that type of surgery, but with time those techniques come to be more widely used by surgeons with less specific expertise. Moreover, the techniques themselves usually evolve rapidly, so that over the months and years that follow the introduction of the technique the standard approach becomes refined. Pragmatic interest should be in the effects of the intervention some time after the introduction of the new technique, at a time when the intervention has become, or could have become, part of routine surgical practice. By that time the technique may have been greatly refined, many more surgeons (including those with less expertise in the field) may be using the technique, and good training programs may have been introduced. A trial conducted soon after the introduction of a new surgical technique might neces-

sarily involve sampling unusually gifted surgeons who are relatively unpracticed in the technique, and might involve the administration of relatively unrefined surgical methods, compared to the same trial conducted several years later. Thus, while it may be possible to conduct explanatory trials seeking proof of concept soon after the introduction of a new surgical technique, it may be impossible to obtain realistic and generalizable pragmatic estimates of effect until the surgical technique has become widely established. Only then may it become possible to observe the effects of surgeons with widely representative competence performing refined surgical techniques.

### 4.2 Comparison Intervention

The choice of comparison intervention should reflect the orientation of a trial. Explanatory trialists usually aim to determine the specific effects of an intervention. To them, nonspecific effects associated with the rituals of administration of the intervention and the interactions between the clinician and patient are usually of no interest and should not be allowed to contaminate estimates of the specific effects of the intervention. It is possible to estimate the specific effects of intervention by ensuring that both the experimental and control groups experience exactly the same nonspecific effects. That is, explanatory trialists can ensure that estimates of the effects of intervention are uncontaminated by nonspecific effects by ensuring that experimental and control interventions are "equalized" [1].

Experimental and control interventions can be equalized by providing a placebo (or sham) intervention to the control group. The ideal placebo intervention is one that involves identical nonspecific rituals of administration but lacks the specific effects of the experimental intervention. Thus most phase III drug trials compare the outcomes of a group of patients given pills (or injections, or ...) containing the experimental compound with the outcomes of a group given pills that are identical in shape, size, and color but that do not contain the experimental compound. Then both the experimental and control groups experience, on average, similar nonspecific effects associated with administration of the intervention. As a consequence, estimates of the effects of the intervention (the difference in the outcomes of the two groups) are believed to be uncontaminated by these nonspecific effects.

Similar reasoning is used to justify blinding of the clinician who administers the intervention. Blinding of the clinician ensures that, on average, the clinician provides the same degree of enthusiasm and positive regard when administering both interventions [2].

There is almost universal agreement that phase III drug trials should blind both the participants (patients) and the clinician administering the intervention. In contrast, there is much more ambivalence among researchers about blinding patients and clinicians in trials of other sorts of interventions. Trials of surgical interventions, medical devices, and programs of care such as educational interventions, exercise programs, and psychological or behavioural interventions frequently do not blind patients or clinicians. No doubt this is due, at least in part, to issues of feasibility: It can be difficult to design a sham intervention that is both comparable to the experimental intervention and therapeutically inert, ethical considerations may preclude the provision of a convincing sham, and it may be difficult to blind clinicians to whether they have administered the experimental or sham intervention. But there

are also theoretical objections to the use of sham interventions: Placebo controls are important in explanatory trials but may be unnecessary, or less necessary, in pragmatic trials.

To the pragmatic trialist, nonspecific effects are a potentially important component of the total effect of an intervention. Therefore, from the pragmatic trialist's perspective, there is no need to separately estimate specific and nonspecific effects. The pragmatic trialist needs only to know the total effect (the sum of nonspecific and specific effects) of administering the intervention. It is this effect that can be anticipated when the intervention is administered to patients in clinical practice. From this perspective there is no need to equalize exposure to nonspecific factors in experimental and control groups. Instead it is more important that the control group is offered a clinically relevant alternative intervention. In Schwartz and Lellouch's terminology, the goal of the pragmatic approach is to optimize rather than equalize the control intervention [1].

More generally, pragmatic trialists need to know whether their patients will experience better outcomes if prescribed the experimental intervention than if prescribed the best available alternative intervention. In general, then, pragmatic trials should compare outcomes of participants receiving the experimental intervention with the outcomes of participants receiving the best available alternative intervention. When the best alternative intervention is minimal intervention (such as advice and support), the most useful trial is one that compares the outcome of participants given the experimental intervention with the outcomes of participants that receives minimal intervention. When the best alternative intervention is a particular program of treatment, the most useful trial is one that compares the outcome of participants given the experimental intervention with the outcome of participants that receive that program of treatment. The principle is that pragmatic trialists should conduct trials with clinically reasonable (relevant) comparison groups [4]. For this reason pragmatic trialists should seriously consider *not* conducting placebo-controlled trials [23].

This advice needs to be qualified. First, the nonspecific effects of intervention may be context-specific. For example, it may be that a particular intervention has placebo effects that manifest in the context of a particular trial (where, e.g., there may be a particularly high degree of regard paid to the patient) but not in the context of broader clinical practice. Context-specific nonspecific effects are of little interest and should be controlled for. Also, the imperative to obtain clinically relevant comparisons may be outweighed by the need to prevent bias. In clinical trials, blinding of participants and clinicians serves additional purposes, over and above controlling for nonspecific effects associated with the rituals of administration of the intervention. At least theoretically, blinding could reduce differential loss to follow-up. More importantly, whenever outcomes are measured subjectively (as they often must be in pragmatic trials) blinding of participants prevents any possibility that the effects of treatment are biased by participants anticipating the responses that they believe the researchers want to hear (the "polite-patient effect" [12]). One of the greatest advantages of blinding of assessors in trials is that it means subjective outcomes can be evaluated without fear of assessor bias ("the clinical impression can be included and given full weight in the analysis" [24]). These important considerations may cause even the most pragmatic trialist to implement placebo controls in their trials.

## 5 CHOICE OF OUTCOMES

Unlike pragmatic trialists, explanatory trialists are not preoccupied with the issue of whether the effects of intervention justify adoption of the intervention in clinical practice; they simply want proof of the concept that the intervention produces certain specific effects. Thus explanatory trialists often use measures of disease activity or disease severity as the primary trial outcomes or endpoints. Examples include measures of tumor size, blood pressure, biochemical markers, muscle strength, bone density, and so on.

In the clinical trials literature these sorts of outcomes are often called surrogate outcomes [25, 26]. Trials that monitor surrogate outcomes are generally considered to be less useful than trials that monitor clinical outcomes such as mortality or quality of life because, as history has shown, surrogate outcomes can provide a misleading indication of the true clinical effects of intervention on mortality or quality of life [26].

The contemporary suspicion of surrogate outcomes is underpinned by two very pragmatic ideas. First, pragmatic decisions about whether or not to implement an intervention involve weighing up all of the potential good and all of the potential harm caused by the intervention. This is more easily achieved by examining global measures of benefit and harm than by assessing the effects of the intervention on specific measures of disease activity. Second, an intervention is ultimately only of clinical benefit if it increases length of life or quality of life. For this reason, the primary outcomes of pragmatic trials should be measures of mortality (survival) or quality of life.

The assessment of mortality is relatively straightforward. A minor controversy concerns which deaths should be counted. Explanatory trialists may prefer to measure disease-specific mortality because analyses of disease-specific mortality may have more statistical power, and therefore be more able to detect effects of intervention, than analyses of all-cause mortality [27]. Pragmatic trialists are more inclined to count all deaths, regardless of their cause, rather than disease-specific mortality. All-cause mortality provides a more global measure of the effects of an intervention because it allows for the possibility that reductions in disease-specific mortality are offset by increases in other causes of mortality [28].

Measurement of quality of life is a more complex task. Quality of life can be measured with generic scales or with any of numerous disease-specific quality-of-life measures [29]. Some trialists are reluctant to measure quality of life because they believe measures of quality of life are relatively insensitive to change and may obscure real effects of intervention. But such reluctance may not always be justified. The *measure* of quality of life must be sensitive to changes in the underlying construct (the latent variable of true quality of life) but, at least from a pragmatic perspective, there need not be any requirement that the underlying construct of quality of life should be sensitive to change.

Outcome measures can be considered to lie on a spectrum. At the explanatory end are measures of disease activity, like biochemistry results, and at the pragmatic end are mortality and quality of life. Somewhere between the ends of this spectrum are outcome measures that are not formal measures of quality of life but could reasonably be expected to reflect quality of life. Examples include measures of depression, or disability, or incidence of injury, or severity of incontinence. From an

extremely pragmatic perspective these measures are really only of interest insofar as they can be used to predict quality of life or costs. Nonetheless, because it is reasonable to expect that these sorts of outcomes are likely to be important determinants of quality of life, they may provide suitable endpoints for pragmatic trials.

Policy-level decisions about whether to implement a particular intervention should be informed by cost-effectiveness (or cost-utility) analyses. For this reason pragmatic trials often measure costs of intervention. Costs to the consumer and health service providers can usually be quantified with relatively little difficulty, although costs to society may be considerably more difficult to estimate. Cost-effectiveness and cost-utility analyses are naturally pragmatic in their orientation so they should always be based on pragmatic outcomes such as quality of life.

## 6 ANALYSIS

### 6.1 Who Should Be Included in Analysis?

Even the best conducted trials experience protocol violations. For example, participants in the experimental group may fail to be given, or comply with, the experimental intervention. Participants in the control group may fail to comply with the control intervention. Participants in either group may be given co-interventions other than those specified in the protocol. People who do not satisfy the inclusion criteria may be incorrectly admitted to the trial. Outcome measures may be obtained at times other than (usually later than) specified in the protocol.

Some trialists may be tempted to identify participants who experience serious protocol violations and exclude those participants' data from the analysis (a per protocol analysis). This strategy might be particularly attractive to explanatory trialists because it might appear that excluding noncompliant participants' data from the analysis would provide a better estimate of the effect of intervention in compliant participants. However, per protocol analyses are potentially seriously biased [30]. Insofar as protocol violations are related to allocation, exclusion of randomized participants can produce imbalance in observed or unobserved baseline covariates. The potential for bias is great when the proportion of excluded participants is moderate or large. In addition, because per protocol analyses are conducted on a subsample of data, they may be underpowered. Another problem is that inference in per protocol analyses is to a relatively poorly and retrospectively defined population.

For these reasons the primary analysis of most clinical trials should be conducted on all available data from all randomized patients. More generally, the analysis should be conducted on an "intention-to-treat" basis. Different meanings are ascribed to the term intention to treat [31], but here the term is used to mean that (a) wherever possible outcome data are obtained from all trial participants (i.e., from every person who was randomized, not just those for whom the trial protocol was followed), (b) all available data (not just data from participants for whom the protocol was followed) are included in analysis, and (c) each participant's data are analyzed in the groups to which the participant was assigned (not according to treatment actually received). In short, the intention-to-treat policy is one of "ever randomised, always analysed" [32].

There may be some circumstances under which departures from an intention-to-treat policy may be reasonable. Fergusson and colleagues [33] have argued that it is reasonable to exclude from the analysis those participants who were clearly ineligible to participate in the trial at the time they were randomized (i.e., people whose inclusion in the trial was a result of the researchers' failure to follow the protocol). On the other hand, participants should not be excluded from the analysis if the information required to determine they did not meet eligibility criteria only became apparent sometime after randomization. For example, participants who were thought to have a particular diagnosis, but who were found some time after randomization to have an alternative diagnosis that would have excluded them from the trial had the alternative diagnosis been known at the time of allocation, should not be excluded from the primary analysis. These recommendations have a distinctly pragmatic flavor.

### 6.2 Adjustment for Noncompliance

In the preceding section it was argued that per protocol analyses should not be used to obtain explanatory estimates of the effects of intervention. The primary analysis should be by intention to treat.

The intention-to-treat approach presents little difficulty for pragmatic trialists. Pragmatic trialists understand estimates of the effects of interventions as estimates of the effect of the *intention* to provide the intervention, or of a *policy* to provide the intervention, rather than in terms of the effect of the intervention itself. More subtly, pragmatic trialists appreciate that the estimate is of the effect of the intervention when it is given in good faith to the people for whom it is thought to be appropriate, even though the intervention may sometimes, for one reason or another, actually be given to patients for whom it may be inappropriate. Similarly, effects are considered to be those measured at as close as possible to a target follow-up time, understanding that in practice the timing of this effect may be variable in practice. An intention-to-treat analysis provides estimates of effects of intervention that can be understood in these pragmatic terms.

Explanatory trialists may, however, be unsatisfied by estimates of effects of intervention based on simple intention-to-treat analyses. Explanatory trialists wish to know the effects of actually receiving an intervention, as distinct from the effects of intending to provide the intervention. This interest may stem from a desire to understand how the effects of intervention are mediated or, alternatively, from a concern about the generalizability of pragmatic estimates of effects of interventions [34, 35]. In recent years there has been a rapid development of methods that can be used to adjust intention-to-treat-based estimates of the effects of interventions to take account of protocol deviations. The focus has been on obtaining intention-to-treat-based estimates of the effects of intervention in compliant patients.

The simplest method for obtaining compliance-adjusted estimates of effects of intervention from all randomized participants applies to the situation in which participants either complied fully with the intervention or received, from the moment of allocation, the intervention they were not allocated. Newcombe has shown that, in this special case, the effect of intervention is attenuated by an amount that depends only on the frequency with which participants receive the intervention they were not allocated to [20]. The explanatory estimate of the treatment effect is the

conventional pragmatic estimate divided by an attenuation factor, which is simply the proportion of subjects in one group receiving the allocated intervention minus the proportion of subjects in the other group receiving the same intervention. The attenuation factor is always less than or equal to one, so the explanatory effect is always greater than or equal to the pragmatic effect. The validity of this estimate is contingent on the assumption that participants complied perfectly either with the allocated intervention or the alternative intervention from the time of randomization (i.e., it assumes participants did not "cross-over" some time after they had already experienced some of the effect of the allocated intervention or too late to experience some of the effect of the intervention they crossed over to). Insofar as participants cross-over after they have already experienced some of the effect of the allocated intervention, or too late to experience some of the effect of the intervention they crossed over to, this explanatory estimate will be biased upward.

Pocock and Abdallah describe a number of simple analyses they conducted to obtain explanatory estimates of the effects of intervention from a placebo-controlled trial of dexfenfluramine for obesity [36]. As they had data on the degree of compliance, they were able to regress outcome against compliance for participants in the experimental group. The regression equation was used to estimate the outcome of perfect compliers in the experimental group. An explanatory measure of the effect of the intervention could then be obtained by subtracting from this value the mean outcome for the control group. (A better approach, though one that was not possible with the data available in that trial, would be to subtract the estimated outcome of perfect compliers in the control group from the estimated outcome of perfect compliers in the experimental group.) A difficulty with this approach is that compliance may be both a determinant and an effect of outcome [37]. Comparison of outcomes of treated and control groups at the same level of compliance are potentially biased when the treatment and control conditions have different influences on compliance.

Sommer and Zeger circumvented this problem in their analysis of a trial of vitamin A supplementation on child mortality [22]. Compliance was treated as a dichotomous variable, so participants were considered to be either compliant or noncompliant. The approach was to infer virtual compliance levels in the control group by assuming that compliance levels in the control group were the same in the control and experimental groups. Unbiased estimates of the effects of intervention in compliant participants could be obtained by comparing mortality risk in the subgroup of compliant participants in the experimental group with the inferred mortality risk in the virtual subgroup of compliant participants in the control group.

Little and Rubin show that both Newcombe's approach and Sommer and Zeger's approach can be derived with causal models [38]. Causal modeling provides a rigorous framework for development of explanatory estimates of effects of intervention. The fundamental idea is that the explanatory effect of an intervention on any individual is the difference between the outcome that individual would have experienced with and without intervention. The obvious statistical problem is that, in a clinical trial, each individual either receives the intervention or does not. Nonetheless, given some often plausible assumptions, it is possible to estimate the average causal effect in compliant people. Causal models have been subject of active development in recent years. Bellamy and colleagues provide an introduction to the use

of causal models for obtaining explanatory estimates of effects of intervention in randomized trials [39].

There are two distinctly different ways of conceiving of explanatory effects of intervention. One possibility is to think of the explanatory effect of intervention in terms of the effect of intervention on the subpopulation of patients who comply (the conditional compliance-related treatment gain [35]). Alternatively, the explanatory effect of intervention could be thought of as the effect intervention would have had if all patients had complied with the intervention (the unconditional compliance-related treatment gain [35]). Estimation of these two quantities involves invoking different assumptions and involves different types of statistical models [39].

It goes without saying that any attempts to estimate the effects of intervention in compliant people can only be conducted if good data are available on compliance. Compliance is a complex construct that includes dimensions of consistency, intensity, and sustainability. Noncompliance may manifest as missed treatments, treatments in which the intervention was administered at an inappropriately high or low dose, or failure to continue with treatment. Unfortunately, it is often difficult to obtain good data on all these dimensions. Measurement of compliance is most often based on pill counts, patient diaries, and participants' self-reports [40], all of which may be biased. Usually, compliance data are reduced to a simple dichotomy, so that participants are classified as compliant or noncompliant.

## 7 REPORTING

The primary theme of this chapter has been that the design of a clinical trial should be strongly influenced by whether the intention is explanatory or pragmatic. Explanatory trials and pragmatic trials differ in the way they sample participants, administer the intervention and control interventions, measure outcomes, and analyze data. Authors of clinical trials should state explicitly, both in trial protocols and trial reports, whether their intention was explanatory or pragmatic. This will go some way to reducing misunderstandings that arise from the different intentions of trialists and readers of trial reports.

Pragmatic trialists are more interested than explanatory trialists in generalizing their findings to clinical contexts. (Perhaps a more accurate way to say this is that pragmatic trialists are more concerned that their findings can be used to make inferences about the care of individual patients.) The extent to which this is possible depends on the extent to which the characteristics of the trial participants and trial interventions resemble the participants and interventions about which inferences are to be made. Consequently, it is particularly important that reports of pragmatic trials provide detailed information about the characteristics of trial participants (patients and clinicians) and how the intervention was administered. Specific details should be provided about the strategies used to facilitate compliance with the intervention.

The conclusions of trial reports should be consistent with the design of the trial. Importantly, authors of explanatory trials should restrict their conclusions to claims about proof of concept. In the event that an explanatory trial provides evidence in support of the specific effects that were hypothesized, the conclusions should state just that. Explanatory trials should not make global claims about the effectiveness

of an intervention. Such claims should await demonstration of clinically important effects with pragmatic trials.

## 8 RECOMMENDATIONS

1. Clinical trials of interventions that are already in widespread practice should usually be pragmatic in orientation.
2. Run-in periods should not be used in pragmatic trials.
3. Pragmatic trials should sample participants (patients and clinicians) in as representative a way as is possible.
4. If sample size calculations for pragmatic trials are to be based on a hypothesis testing approach, they should incorporate empirical estimates of the smallest worthwhile effect of intervention, such as can be obtained with the benefit–harm trade-off method. Alternatively, sample size calculations for pragmatic trials could be based on considerations of the precision of estimates of effect, or of the value of sample information.
5. If the primary analysis of an explanatory trial is to involve adjustment of estimates for noncompliance, sample size should be adjusted accordingly.
6. Explanatory trials should blind participants and clinicians wherever possible. Pragmatic trialists should consider not blinding participants or clinicians.
7. In pragmatic trials the control group should receive the best available treatment alternative.
8. Pragmatic trials should have as endpoints either survival or measures that reflect quality of life or both.
9. The primary analysis of all trials, whether explanatory or pragmatic, should be conducted on an intention to treat basis.
10. Reports of clinical trials should explicitly indicate if the intention is to answer an explanatory or pragmatic question.
11. Pragmatic trials should provide detailed information about the characteristics of trial participants (patients and clinicians), how the intervention was administered, and the degree of compliance with the intervention.
12. Explanatory trials should not make global claims about the effectiveness of an intervention. Such claims should await demonstration of clinically important effects with pragmatic trials.

## ACKNOWLEDGMENTS

The author acknowledges helpful comments on the manuscript made by Luciana Machado and Lisa Harvey. He is supported by a fellowship from the Australian NHMRC.

## REFERENCES

1. Schwartz, D., and Lellouch, J. (1967), Explanatory and pragmatic attitudes in therapeutical trials, *J. Chronic Dis.*, 20, 637–648.

2. Nathan, P. E., Stuart, S. P., and Dolan, S. L. (2000), Research on psychotherapy efficacy and effectiveness: Between Scylla and Charybdis? *Psychol. Bull.*, 126, 964–981.
3. Depp, C., and Lebowitz, B. D. (2007), Clinical trials: Bridging the gap between efficacy and effectiveness, *Int. Rev. Psychiatry*, 19, 531–539.
4. Tunis, S. R., Stryer, D. B., and Clancy, C. M. (2003), Practical clinical trials: Increasing the value of clinical research for decision making in clinical and health policy, *JAMA*, 290, 1624–1632.
5. Yusuf, S., Collins, R., and Peto, R. (1984), Why do we need some large, simple randomized trials? *Stat. Med.*, 3, 409–422.
6. Fairbank, J., Frost, H., Wilson-MacDonald, J., et al. (2005), Randomised controlled trial to compare surgical stabilisation of the lumbar spine with an intensive rehabilitation programme for patients with chronic low back pain: The MRC spine stabilisation trial, *BMJ*, 330, 1233.
7. Fransen, G. A., van Marrewijk, C. J., Mujakovic, S., et al. (2007), Pragmatic trials in primary care. Methodological challenges and solutions demonstrated by the DIAMOND study, *BMC. Med. Res. Methodol.*, 7, 16.
8. Brittain, E., and Wittes, J. (1990), The run-in period in clinical trials. The effect of misclassification on efficiency, *Control. Clin. Trials*, 11, 327–338.
9. Julious, S. A. (2004), Sample sizes for clinical trials with normal data, *Stat. Med.*, 23, 1921–1986.
10. Rothman, K. J., and Greenland, S. (1998), *Modern Epidemiology*, 2nd ed., Lippincott-Raven, Philadelphia.
11. Herbert, R. D. (2000), How to estimate treatment effects from reports of clinical trials. I: Continuous outcomes, *Australian J. Physiother.*, 46, 229–235.
12. Herbert, R. D., Jamtvedt, G., Mead, J., et al. (2005), *Practical Evidence-Based Physiotherapy*, Elsevier, Oxford.
13. Barrett, B., Brown, R., Mundt, M., et al. (2005), Using benefit harm tradeoffs to estimate sufficiently important difference: The case of the common cold, *Med. Decis. Making*, 25, 47–55.
14. Barrett, B., Brown, D., Mundt, M., et al. (2005), Sufficiently important difference: Expanding the framework of clinical significance, *Med. Decis. Making*, 25, 250–261.
15. Barrett, B., Harahan, B., Brown, D., et al. (2007), Sufficiently important difference for common cold: Severity reduction, *Ann. Family Med.*, 5, 216–223.
16. MacRae, K. D. (1989), Pragmatic versus explanatory trials, *Int. J. Technol. Assess. Health Care*, 5, 333–339.
17. Powell-Tuck, J., MacRae, K. D., Healy, M. J., et al. (1986), A defence of the small clinical trial: Evaluation of three gastroenterological studies, *BMJ*, 292, 599–602.
18. Beal, S. L. (1989), Sample size determination for confidence intervals on the population mean and on the difference between two population means, *Biometrics*, 45, 969–977.
19. Willan, A. R., and Pinto, E. M. (2005), The value of information and optimal clinical trial design, *Stat. Med.*, 24, 1791–1806.
20. Newcombe, R. G. (1988), Explanatory and pragmatic estimates of the treatment effect when deviations from allocated treatment occur, *Stat. Med.*, 7, 1179–1186.
21. Sato, T. (2000), Sample size calculations with compliance information, *Stat. Med.*, 19, 2689–2697.
22. Sommer, A., and Zeger, S. L. (1991), On estimating efficacy from clinical trials, *Stat. Med.*, 10, 45–52.
23. Vickers, A. J., and de Craen, A. J. (2000), Why use placebos in clinical trials? A narrative review of the methodological literature, *J. Clin. Epidemiol.*, 53, 157–161.

24. Bradford Hill, A. (1961), *Principles of Medical Statistics*, 7th ed. London: The Lancet.
25. International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (1998), *ICH harmonised tripartite guideline*. Statistical principles for clinical trials.
26. Biomarkers Definitions Working Group (2001), Biomarkers and surrogate endpoints: Preferred definitions and conceptual framework, *Clin. Pharmacol. Ther.*, 69, 89–95.
27. Yusuf, S., and Negassa, A. (2002), Choice of clinical outcomes in randomized trials of heart failure therapies: Disease-specific or overall outcomes? *Am. Heart J.*, 143, 22–28.
28. Black, W. C., Haggstrom, D. A., and Welch, H. G. (2002), All-cause mortality in randomized trials of cancer screening, *J. Natl. Cancer Inst.*, 94, 167–173.
29. Guyatt, G. H., Feeny, D. H., and Patrick, D. L. (1993), Measuring health-related quality of life, *Ann. Intern. Med.*, 118, 622–629.
30. Lee, Y. J., Ellenberg, J. H., Hirtz, D. G., et al. (1991), Analysis of clinical trials by treatment actually received: Is it really an option? *Stat. Med.*, 10, 1595–1605.
31. Hollis, S., and Campbell, F. (1999), What is meant by intention to treat analysis? Survey of published randomised controlled trials, *BMJ*, 319, 670–674.
32. Cox, D. (1998), Discussion, *Stat. Med.*, 17, 387–389.
33. Fergusson, D., Aaron, S. D., Guyatt, G., et al. (2002), Post-randomisation exclusions: The intention to treat principle and excluding patients from analysis, *BMJ*, 325, 652–654.
34. Mark, S. D., and Robins, J. M. (1993), A method for the analysis of randomized trials with compliance information: An application to the Multiple Risk Factor Intervention Trial, *Control. Clin. Trials*, 14, 79–97.
35. Goetghebeur, E. J., and Shapiro, S. H. (1996), Analysing non-compliance in clinical trials: Ethical imperative or mission impossible? *Stat. Med.*, 15, 2813–2826.
36. Pocock, S. J., and Abdalla, M. (1998), The hope and the hazards of using compliance data in randomized controlled trials, *Stat. Med.*, 17, 303–317.
37. Rochon, J. (1995), Supplementing the intent-to-treat analysis: Accounting for covariates observed postrandomization in clinical trials, *J. Am. Stat. Assoc.*, 90, 292–300.
38. Little, R. J., and Rubin, D. B. (2000), Causal effects in clinical and epidemiological studies via potential outcomes: Concepts and analytical approaches, *Ann. Rev. Pub. Health*, 21, 121–145.
39. Bellamy, S. L., Lin, J. Y., and Ten Have, T. R. (2007), An introduction to causal modeling in clinical trials, *Clin. Trials*, 4, 58–73.
40. Gossec, L., Tubach, F., Dougados, M., et al. (2007), Reporting of adherence to medication in recent randomized controlled trials of 6 chronic diseases: A systematic literature review, *Am. J. Med. Sci.*, 334, 248–254.

# *Ethics of Clinical Research in Drug Trials*

**Roy G. Beran**
*Strategic Health Evaluators, Chatswood NSW 2067 Australia*

**Contents**

1 Foundations for Trial Design
2 Investigator Selection
3 External Study Approval
4 Consent Process
5 Conduct of Trial
6 After Study Completion
7 Conclusion
References

## 1 FOUNDATIONS FOR TRIAL DESIGN

The ethics of clinical drug trialing start with the concept of the trial that is to be undertaken. The trial must aim to answer a question that has not yet been answered, thereby necessitating the need for such a trial [1]. It should not be a substitute for a marketing strategy in which investigators place patients onto expensive agents for which either the patient or the formulary (namely the list of drugs available within a specific facility) cover the real costs of the trial, while the yield from the study contributes very little to scientific/clinical knowledge and more to treatment commitment [2].

The latter group of studies have a distinct advantage for the pharmaceutical company that is sponsoring them because they represent a pool of patients who would, otherwise,

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

possibly not receive the medication, were it not for the conduct of the clinical trial [3]. These studies generally lack scientific validity, and investigators need to think very carefully before committing to such questionable trials. Ethics committees also need to review the content of studies to make sure that they are not merely marketing tools masquerading as clinical trials.

The design of a study should include calculations to determine the sample size necessary to allow sufficient mathematical power to achieve a reliable response with adequate scientific rigor [4].

There needs to be consideration of the size of the response expected, the right dosage needed to attract such a response, the correct comparator (be it placebo, which theoretically equates to no comparator, or a proven therapeutic agent already known to be effective in the management of the condition under review), and the potential subject "dropout rate" that needs to be accommodated to achieve the desired final sample size upon which to base the study's analysis with adequate power [5].

It is often impossible to determine these parameters accurately as they may rely on earlier animal studies for extrapolation, and the animal data may not properly translate into human experience [6]. Alternatively, the concept may require an initial pilot study, in which far fewer subjects are involved but in which either the idea or the methodology under investigation can be tested. Such pilot studies may not truly reflect definitive data, but they will provide invaluable evidence to either support or refute the initial hypothesis [7, 8].

Pilot trials may be the only source of information upon which to base a foundation with which to ascertain indicative information to be used as the basis for calculations to determine a subsequent study that allows adequate power to truly answer the scientific question being posed. Often such pilot studies are difficult to publish as they lack the power to answer the question reliably, but their value should not be underestimated nor should the ethics of publication of such information be dismissed as they may represent the only proof of concept information [9].

The dosage of medication to be used in the trial may prove critical. Studies that have used too large a dosage cause unnecessary adverse events and hence delay or impede a drug's development or acceptance, as was the case with topiramate, which used a dosage of 200 mg in its trials [10, 11]. Conversely, the dosage that was used within the trial may be too low to achieve adequate response rate, thereby suggesting the agent to be less effective than it truly is, as was found with gabapentin [12]. Such errors have direct negative implications for the target patient population and may deny patient access to what amounts to an effective therapy because the study design was formulated on false expectations. It can be seen that innocent, but incorrect, assumptions may carry with them ethical consequences that may ultimately affect patient well-being because they affect the availability of appropriate therapeutic modalities [13].

The cost of conducting a trial has direct implications upon the adopted methodology. A crossover design is far less expensive than is a parallel arm design, and the purpose of the study must be decided. If the purpose of the study is to satisfy a regulatory requirement, then a parallel arms study design may be mandated by those regulatory authorities that are to determine the fate of the product under review [14].

The reason for reduced costs, in a crossover design, is that the subject acts as his or her own control, thereby allowing a significantly reduced sample size, as one patient serves in both arms of the trial to answer the study's question, which may effectively half the necessary sample size. This obviously reduces both the time delay and the effort it takes to recruit double the sample size and hence also reduces the need for additional investigators with inherent monitoring and infrastructure costs. Thus the purpose of the study may determine the study design, beyond any strict scientific or ethical requirements.

Where there is a need to have large numbers in a trial, then it often follows that the trial sponsors will resort to cost-cutting measures that may impact on the scientific rigor adopted for the protocol. An example of this is the use of large open-label studies to answer questions of comparative efficacy where the cost of a double-blinded methodology would be prohibitive due to the need for a "double-dummy approach" (in which both treatment A and treatment B need to have mirrored therapy with matched placebos such that the patient receives either treatment A, plus placebo, or placebo plus treatment B, thereby offering a potentially truly double-blind methodology) [15, 16]. The costs of such an approach might mean that the study would never have been feasible while the open-label design may still provide some very useful information.

The need to have large numbers recruited to a study may also translate into the inclusion of less skilled investigators, with far greater propensity for protocol violation and hence possible pollution of data. It is hoped that such potential problems do not translate into ethical expedience with insufficient external scrutiny of procedure.

It can be seen that the ethics of clinical research, and drug trialing, commence long before the first investigators are approached, the first study site is selected, or the first patient is recruited into the trial. The consideration of ethics starts with the selection of the scientific question to be asked, the capacity to answer that question, and the study design to be adopted for that purpose [17].

## 2 INVESTIGATOR SELECTION

For an investigator to accept to be part of a clinical trial requires that the investigator be confident that he or she can fulfill the requirements of the trial. This dictates that the investigator conducts a review of the purpose of the study, the question(s) to be asked in the trial, the methodology to be adopted, the types and numbers of patients sought for inclusion (reviewing inclusion and exclusion criteria), and the feasibility of conducting the study within the constraints of his or her center, while recognizing that all conduct within the trial must adhere to good clinical practice [18].

Many international trial sponsors will conduct a feasibility study of proposed investigators, but, as these are based on personal estimations, their validity is quite suspect [19]. Despite this shortcoming, the concept that each proposed investigator should at least assess his or her capacity to adequately conduct such a trial is very appropriate.

It is inappropriate, and ethically unjustified, to nominate oneself for inclusion as a prospective investigator within a trial if it is highly doubtful that one could accom-

modate the proposed trial's needs [20]. If selected as an investigator, the study sponsor automatically accumulates direct costs: such as investigator meetings, with its transportation and accommodation costs, costs for ethics committee submissions, and infrastructure costs [21].

There are basic considerations applicable to each nominated investigator. These include having adequate knowledge of the subject matter under consideration, having the resources to conduct the trial (be it space, staff, storage facilities, or standard operating and accounting procedures), having access to an acceptable and properly constituted ethics committee, and expecting to be able to recruit adequate numbers of appropriate patients to justify inclusion [22].

Without these ingredients, the proposed investigator should exclude him or herself.

The purpose of this discussion is to demonstrate that there are unequivocal ethical obligations on the investigator even before agreeing to be part of any investigative team. These obligations are often overlooked but to do so may cause unnecessary burden to a prospective sponsor. This may cause a sponsor to think twice about underwriting future studies, if investigators failed to meet their agreed commitments.

## 3 EXTERNAL STUDY APPROVAL

Once the aims of the study have been determined, the protocol established, and investigator sites selected, the study must then be subjected to external scrutiny.

All those who are involved with the conduct of the study (namely the company initiating the trial, the sponsor, the clinical research organization that will supervise the study, and the investigators) must review all the requirements [23], which may include contractual negotiations, privacy agreements, financial incentives, and infrastructure requirements, all of which should be available to the external review process as required by the relevant committees.

Submission to review boards, be they for scientific validity or for human research ethics committee (HREC) evaluation, are usually prepared in a predetermined fashion. The purpose of the HRECs is to protect the interests of trial subjects. Some, but not all, HRECs ask for a relevant expert scientific evaluation of a proposed protocol, before deliberating on the trial, so as to ensure that it is seeking valid scientific answers within the study design. This process should ensure that the study is not simply a marketing ploy. Such appraisal also seeks to confirm the feasibility of a study, its merits, and any potential risk that may, or may not, have been identified at the time of the submission. Some HRECs rely on their own members, which might include the relevant scientific expertise to provide the necessary scientific input. In either context, the role of the HREC is less to do with the scientific value of the proposed research (for which it can rely on others), and its main function is to focus its principal deliberations on the ethics of patient involvement [24].

The HREC reviews the protocols, the reasons for the study, patient safety, access to necessary information, format, content, and language of both patient information and consent documentation, access to investigators in case of any problems, and clear evidence of lack of coercion when recruiting patients [25].

This is of particular importance when the sample population is recruited from the clinical practice of an investigator, be it in the private or public practice of medicine. Patient protection is the fundamental obligation of the HREC [25], while at the same time it should not obstruct worthwhile research that may benefit the wider community.

Clinical trials can expose research candidates to new, and often unknown, risks when consenting to be subjects for a trial. To protect patients against such risks, the sponsors are expected to take out insurance to cover those risks should they eventuate [26]. More recently, members of HRECs have also sought insurance to indemnify their membership against potential litigation consequent to any adverse events that may occur within a trial. The HREC will often seek confirmation of appropriate insurance being in place before approving any study that involves human participation.

Studies that include live subjects (human or animal) cannot proceed without ethics committee approval, and, in the case of humans, a written consent or signed surrogate consent is needed from either the patient or someone responsible. Surrogate consent is "ideally a substituted judgment made by a person responsible for healthcare decision-making for a particular patient under the relevant legislation" [27]. The situations in which such surrogate consent may be required include evaluation of therapy for certain conditions, such as treatment in stroke patients for whom the capacity to decide is no longer valid. It may also be necessary to seek surrogate consent in underage subjects for whom guardian approval is mandatory. A similar situation exists for patients who lack the intellectual capacity to comprehend what is being asked of them within the trial protocol, as is the case with intellectually handicapped patients. For these patients it is also necessary to seek guardianship board approval, so as to relegate the authority to consent for such a patient to be included within any given trial, to a "responsible other" person who will have the patient's best interests as his or her primary concern [28].

## 4 CONSENT PROCESS

"Consent is widely believed to affirm the autonomous decision-making rights of prospective research participants, and their capacity to protect their own interests" [27]. Once the study has been approved by the external review process, each investigator assumes responsibility for the conduct of the trial within his or her institution, be it in private or public practice. This starts with a site initiation visit in which the sponsor, or surrogate, usually a clinic, also termed contract research organization (CRO), which is specifically contracted for the task, confirms that all those at the site are fully conversant with their obligations and the procedures to be adopted within the trial and that the center has the wherewithal to conduct the study [29].

With the emergence of CROs there has also emerged an understanding that they too have unequivocal ethical obligations. The level of sophistication impacting upon CROs have developed to the point of these being umbrella organizations designed to oversee compliance of its voluntary membership [30].

The investigator(s) then approaches suitable candidates regarding their willingness to participate within the trial. Each candidate must provide informed consent, based on provision of HREC approved study information, which clearly outlines

the purpose of study, the obligations imposed on both subjects and investigators, and the potential known risks and hazards that accrue consequent to participation in the trial. Each subject must have exercised free will, devoid of the imposition of undue influence, and must have the capacity to ask any questions that may have arisen, following the reading of the study information, before signing the informed consent document [22].

This process must be conducted in an open and unbiased fashion and the written consent should be witnessed and signed by a competent adult, recognizing that the written consent may be provided by a surrogate, representing the patient's best interests. Some may allow for the lack of equipoise in the doctor–patient relationship and have the trial coordinator initiate the consent process, following which the doctor will be available to answer any outstanding questions [26, 31].

It is imperative that the investigator must still be available to answer more difficult questions (which were beyond the capacity of the trial coordinator) and that investigator should countersign the consent form and confirm its authenticity [26, 31].

This goes some way to overcome any suggestion that patients, selected from the clinical practice of an investigator, have been coerced into participating in the trial. It is far easier to refuse participation within a trial if the request comes from a research coordinator rather than the doctor, especially if the HREC-approved study information includes a passage that guarantees ongoing clinical care by the doctor, irrespective of whether the prospective subject is willing, or not, to participate in the trial [26, 31].

The consent documents must acknowledge that trials of new therapeutic agents may not have encountered all possible risks and hazards that attach to the medication under experimentation. Potential subjects must be warned of the known risks and must also be advised that no one can guarantee that the trial will provide a direct benefit for them. Where a trial compares an investigative agent to a placebo, trial subjects must understand that they may not receive the medication under review, within the trial, and may receive the equivalent of a "sugar pill." They also need to be appraised of the risk ratio of this happening [29]. Similar advice needs to be given where differing doses of the study medication are to be given so that consent is truly informed. In short, the study protocol must be adequately described within the patient information and consent documents to clearly demonstrate that there has been no attempt to hide vital information or to give a false risk–benefit ratio [29]. The HREC has the responsibility to ensure that this has occurred and that the language used is of a sufficiently simple nature to avoid excessive use of jargon that may confuse a prospective subject. It must also be satisfied that the concepts, including the consent of information documents, are provided in a manner that is not too complicated for the average man/woman to understand [27].

Many trials of new medications adopt a blinded methodology, be it single or double blinded, and the potential subject must appreciate that he or she will not know what medication he or she will receive in a blinded study [15]. It is even more important for the subject to understand that, in the double-blinded methodology, neither the patient nor the doctor knows which medication the patient is taking during the trial. As mentioned earlier, the trial may use a placebo as the comparator. In such trials subjects will not know if they are receiving an active compound at all, in addition to the standard medications they were receiving before the trial. This

encapsulates the very essence of a randomized, placebo-controlled, double-blinded protocol. Most placebo-controlled studies do make provision for the patient to receive the experimental compound in an open-labeled, long-term safety and efficacy trial that follows the blinded study and that continues until the study medication is freely available. This must also be explained to potential trial subjects [23, 32].

Only subsequent to the free giving of informed consent, based on full disclosure, can the patient be included within the trial procedures [33]. Part of the consent process also includes provision for the withdrawal of consent, at any time throughout the study, without fear of compromising future ongoing clinical care by the doctor [33].

## 5 CONDUCT OF TRIAL

There is usually very little latitude to alter the conduct of a trial. All requirements and expectations of a trial are set out in the HREC-approved trial protocol, which stipulates the timing and conduct of the trial visit [22]. Unfortunately, all trials that include human subjects will also include errors made by either the investigating team or the subjects themselves. Some of these errors represent only minor infringements for which a waiver may be sought from the sponsor [29]. Other errors may be so significant as to dictate immediate exclusion and withdrawal from any further trial involvement. Should a waiver be approved, then the patient can continue in the trial and allowance is made for the deviation from trial protocol requirements.

If the patient experiences a serious adverse event (SAE), such as the occurrence of something that causes hospitalization or is life-threatening, then both the HREC and the sponsor need to be advised, as soon as is possible, following the investigator having been made aware of the SAE [34]. The necessary action that is to be taken, regarding the experimental agent, is at the discretion of the investigator, although everyone needs to be advised of the action taken. Depending on the nature of the SAE, and its cause, it may be necessary to advise all patients within a trial of the SAE, and it may also be necessary to amend the patient information and informed consent documents. The HREC, itself, may insist on this action and it may go so far as to terminate the trial if its members believe that continuation of the trial is no longer ethical or in the patients' best interests. Investigators must make the same considerations, and advise the HREC accordingly, thereby demonstrating that the investigator also holds the well-being of the trial subjects as his or her primary concern for continuation of any given trial [34].

Others who are trialing the same medication—either for a similar indication or an alternatively indication—and who encounter an SAE—must also notify their appropriate authorities, such as their HREC, sponsor, or CRO. They, meaning the sponsor or CRO, must notify all other investigators who are/were trialing the same agent. Once notified of these SAEs encountered by others, it is imperative that the investigator also bring these to the attention of their HREC in the same fashion as they would have done had they personally encountered the SAE [29]. This further ensures maintenance of informed consent, should any of the SAEs have a bearing on the decision to continue a trial, be it from the perspective of the investigating team or subject.

Based on the SAE reporting, the HREC may seek clarification to determine if: the trial should continue with its currently recruited subjects; if continuation of recruitment should be allowed, if the entire process should be aborted, or if there needs to be an amendment to the current informed consent and information documents; and if the informed consent process needs to be repeated with each subject who is already within the trial to expedite their withdrawal should they choose not to face further previously unforeseen risk [29].

Depending on the nature of an adverse event, there may be need to unblind the blinded randomized data to determine whether the patient was exposed to the experimental compound and, if so exposed, the dosage of the exposure. Again this is at the discretion of the investigator, but it remains imperative that the patient's well-being must assume primal consideration [34]. This takes priority over any secondary ethical obligation that the investigator may have to the sponsor of the study to adhere to trial protocol as it relates to blinding of treatment arms.

Upon completion of the trial, the subject may either elect to continue on the study agent, if an open-label study is to follow, or choose to stop whatever it was that they were taking during the trial [23]. It must be remembered that trial subjects may not be advised of what they were taking during a double-blinded study, even after they complete that part of the trial, if not all subjects have completed the study. This is considered important to maintain mathematical integrity when analyzing the study data, but it does raise important ethical considerations as the decision to continue within the open-label arm is often based on inadequate information, in the absence of knowing what was taken during the trial. Some patients may think that they had a real benefit from the study medication when in fact they were on placebo, while others who were on the study medication may have thought they were on placebo [35]. Thus far there has not been an adequate resolution to this mathematical and ethical dilemma, which still needs to be resolved on a consensus basis.

The question of conducting an open-label study, following completion of a blinded protocol, is also of ethical importance [32]. Access to the study medication, following involvement in a blinded study, is often considered as an inherent ethical consideration, even at the time of the initial HREC approval for the blinded study to be undertaken. It is considered important to ensure that patients, who consider that they have benefited from the study medication, be allowed to continue on it until the medication is available on the market at a reasonable price, especially where a formulary exists, as is the case with the government-sponsored Pharmaceutical Benefits Scheme (PBS) in Australia [23, 36, 37].

## 6 AFTER STUDY COMPLETION

Responsibility for the trial consequences does not automatically stop once the last subject has completed his or her trial commitments. Insurance against the delayed expression of adverse effects is expected to be maintained long after the study has been completed and at least for one year. Ongoing support of patients, access to effective treatment, and notification of subsequent SAEs remains obligatory to ensure that patient safety is maintained [29, 34].

What is often less clear-cut is the ethical requirement to publish the findings of the study once the results become available. Current opinion supports an ethical

obligation to make the findings of any study publicly accessible, to ensure the widest dissemination of all results, be they positive or negative [38]. To satisfy this obligation, it is mandatory for journals to provide equal space for both negative and positive studies, something that is often claimed but not necessarily satisfied or realized with the current perception being that positive studies are far easier to publish than our negative studies [39, 40].

## 7 CONCLUSION

The ethics of clinical drug trials start at the time of establishing what question(s) should be asked, and answered, as a consequence of the trial. The response to these issues influences every step along the way, up to and including posttrial follow-up.

Issues that have not been canvassed within this treatise include those ethical obligations that represent the responsibility of trial subjects to adhere to their commitments to follow trial dictates. It behoves all those concerned with the conduct of the trial to recognize that trials represent a very expensive and demanding exercise that must adhere to the requirements as stipulated within the relevant protocol. Other issues, not addressed in the above analysis, include the ethical obligations that should be addressed by the company that sponsors the trial. These obligations should include the maintenance of rigorous postmarketing surveillance concerning the product being trialed, to protect against unexpected SAEs, as occurred with such medications as vigabatrin, which caused specific visual field defects that were only recognized years after the drug had come to market [41]. It may also be relevant to conduct postmarketing studies that better reflect the true use of products to see if clinical practice truly reflects what was demonstrated as a consequence of regulatory determined trials [42–44].

## REFERENCES

1. Ablett, S. (1997), Clinical trials: Towards good practice, *Arch. Dis. Child*, 55, 283–286.
2. Altman, D. I. (2002), Poor-quality medical research: What can journals do? *JAMA*, 287, 2765–2767.
3. Editorial (2001), Sponsorship, authorship and accountability. Revision of statement on publication ethics by the International Committee of Medical Journal Editors, *MJA*, 175, 294–296.
4. Chan, Y. H. (2003), Randomised controlled trials (RCTs)—Sample size: The magic number? *SMJ*, 44(4), 172–174.
5. Kirby, A., Gebski, V., and Keech, A. C. (2002), Determining the sample size in a clinical trial, *MJA*, 177(5), 256–257.
6. Roberts, I., Kwan, I., Evans, P., et al. (2002), Does animal experimentation inform human healthcare? Observations from a systematic review of international animal experiments on fluid resuscitation, *BMJ*, 324, 474–476.
7. Lancaster, G. A., Dodd, S., and Williamson, P. R. (2004), Design and analysis of pilot studies: Recommendations for good practice, *J. Eval. Clin. Pract.*, 10(2), 307–312.
8. Wittes, J., and Brittain, E. (1990), The role of internal pilot studies in increasing the efficiency of clinical trials, *Stat. Med.*, 9(1–2), 65–71.

9. Workman, P. (2004), Inhibiting the phosphoinositide 3-kinase pathway for cancer treatment, *Biochem. Soc. Trans.*, 32, 393–396.
10. Stephen, L. J., Sills, G. J., and Brodie, M. J. (2000), Topiramate in refractory epilepsy: A prospective observational study, *Epilepsia*, 41, 977–980.
11. Salinsky, M. C., Storzback, D., Spencer, D. C., Oken, B. S., Landry, T., and Dodrill, C. B. (2005), Effects of topiramate and gabapentin on cognitive abilities in healthy volunteers, *Neurology*, 64, 792–798.
12. Chadwick, D., Leiderman, D. B., Sauermann, W., et al. (1996), Gabapentin in generalised seizures, *Epilepsy Res.*, 25(3), 191–197.
13. Kane, J. M. (2002), Issues in clinical trials designs, in Davis, K. L., Charney, D., Coyle, J. T., Nemeroff, C., Eds. *Neuropsychopharmacology: The Fifth Generation of Progress*, Lippincott Williams & Wilkins, Philadelphia, pp. 537–546.
14. FDA (1999), Guidance for industry: Clinical development programs for drugs, devices and biological products for the treatment of rheumatoid drugs, devices and biological products for the treatment of rheumatoid arthritis (RA), U.S. Food and Drug Administration, Rockville, MD.
15. Grimes, D. A., and Schulz, K. F. (2002), Clinical research in obstetrics and gynecology: A Baedeker for busy clinicians, *Obstet. Gynecol. Surv.*, 57(9), S35–S53.
16. Talley, N. J., Moore, M. G., Sprogis, A., et al. (2002), Randomised coltrolled trial of pantoprazole versus ranitidine for the treatment of uninvestigated heartburn in primary care, *MJA*, 177(8), 423–427.
17. The Therapeutic Goods Administration (1991), *Guidelines for Good Clinical Research Practice (GCRP) in Australia*, Cth Dept of Health Housing and Community Services (DEB 3), Aust. Gov. Printers, Canberra, Commonwealth Dept. of Health, Housing & Community Services.
18. World Health Organisation (2002), *Handbook for Good Clinical Research Practice (GCP): Guidance for Implementation*, WHO, Geneva.
19. Committee for Proprietary Medical Products, *Guidelines for Good Clinical Pracrice (GCP): ICH Harmonised Tripartite Guideline* (originally approved 17 July 1996).
20. Lader, E. W., Cannon, C. P., Ohman, E. M., et al. (2004), The clinician as investigator: Participating in clinical trials in the practice setting, *Circulation*, 109, 2672–2679.
21. Kovacevic, M., Odeleye, O. E., Sietsema, W. K., et al. (2001), Financial concepts to conducting and managing clinical trials within budget, *Drug. Info. J.*, 35, 1031–1038.
22. The Therapeutic Goods Administration (2006), *The Australian Clinical Trial Handbook: A Simple Practical Guide to the Conduct of Clinical Trials to International Standards of Good Clinical Practice (GCP) in the Australian Context*, Dept. of Health and Ageing, Aust. Gov., Canberra.
23. World Medical Association Declaration of Health, Helsinki (1989), *Recommendations Guiding Physicians in Biomedical Research Involving Human Subjects*, Adopted by the 18th World Medical Assembly, Helskinki, Finland, June 1964, and amended by 29th World Medical Assembly, Tokyo, Japan, Oct. 1975; 35th World Medical Assembly, Venice, Italy, Oct 1983; 41st World Medical Assembly, Hong Kong, Sept 1989; 48th General Assembly Somerset West, Republic of South Africa, Oct 1996; 52nd World Medical Assembly Edinburgh, Scotland, Oct 2000; AMA Code of Ethics, Canberra 1996, 2003, 2004; Fluss S S, *International Guidelines on Bioethics*—Informal listing of selected International Codes, Declarations, Guidelines etc. on medical ethics/bioethics, health care ethics, human rights

aspects of health, *EFCGP News*, Sept. 1998; rev. ed., Dec 1999; 2nd rev. ed., Autumn 2000.

24. National Health & Medical Research Council—Human Research Ethics Committee (1999), *National Statement on Ethical Conduct in Research Involving Humans*, Commonwealth of Australia, Canberra.
25. Commission of the European Communities (1997), Proposal for a European Parliament and Council Directive on the Approximation of Provisions Laid down by Law, Regulation or Administrative Action Relating to the Implementation of Good Clinical Practice in the Conduct of Clinical Trials on Medicinal Products for Human Use [COM (97) 369 final], Brussels.
26. Beran, R. G. (2005), Ethical considerations within clinical research with special focus upon clinical drug trials, *Med. Law*, 25, 411–436.
27. National Health and Medical Research Council (2001), *Human Research Ethics Handbook*, NHMRC, Canberra.
28. Devereux, J. A. (2002), Guardianship and Consent, in Beran, R. G., Eds., *Epilepsy: A Question of Ethics*, Yozmot, Tel Aviv.
29. Anon (2000), *Note for Guidance on Good Clinical Practice* (CPMP/ICH/135/95): Annotated with TGA comments. Canberra: Therapeutic Goods Administration, Department of Health and Aged Care. Available at: http://www.tga.gov.au/docs/html/ich13595.htm; accessed April 4, 2006.
30. Beran, R. G. (2003), Informed consent, a legal requirement in the management of patients with epilepsy, *Med. Law*, 22(1), 155–184.
31. Chase, D., Gierend, M., Rettig, S., et al. (2005), Quality assurance in clinical trials: results of the follow-up systems of member companies in the BVMA, *Int. J. Pharm. Med.*, 19(5–6), 275–276.
32. Wainwright, P. (2002), Consent to open label extension studies: Some ethical issues, *J. Med. Ethics*, 28, 373–376.
33. Donnellan, P., and Smyth, J. (2001), Informed consent and randomised controlled trials, *J. R. Coll. Surg. Edinb.*, 46(2), 100–102.
34. Keech, A. C., Wonders, S. M., Cook, D. I., et al. (2004), Balancing the outcomes: Reporting adverse events, *MJA*, 181(4), 215–218.
35. Oh, V. M. S. (1994), The placebo effect: Can we use it better? *BMJ*, 309, 69–70.
36. Graham, D. (1995), The Australian pharmaceutical benefits scheme, *Aust. Prescr.*, 18, 42–44.
37. Eadie, M. J. (2004), The Australian pharmaceutical benefits scheme, in Beran, R. G., Eds., *Epilepsy and the Law of Therapeutics*, Yozmot, Tel Aviv, Israel.
38. Wager, E., Field, E. A., and Grossman, L. (2003), Good publication practice for pharmaceutical companies, *Curr. Med. Res. Op.*, 19, 149–154.
39. Hensley, S., and Abboud, L. (2004), Medical research has "black hole": Negative results often fail to get published in journals; some blame drug industry, *Wall. St. J. (East Ed)*, June 4, B3.
40. Berlin, J. A., and Wacholtz, M. C. (2005), Selective reporting, publication bias and clinical trial registry: An industry perspective, *Int. J. Pharm. Med.*, 19(5–6), 277–284.
41. Beran, R. G. (2001), The ethics of post-marketing surveillance of therapeutic agents, *Med. Law*, 20(4), 587–594.
42. Beran, R. G., Berkovic, S. F., Black, A. B., et al. (2005), Efficiency and safety of levetiracetam 1,000 to 3,000 mg/day in patients with refractory partial seizures: A multicentre, open-label single arm study, *Ep. Res.*, 63, 1–9.

43. Beran, R. G., Berkovic, S., Black, A., et al. (2001), Australian Study of Titration to Effect Profile of Safety (AUS-STEPS): High-dose gabapentin (Neurontin®) in partial seizures, *Epilepsia*, 42(10), 1335–1339.
44. Berry, D. J., Beran, R. G., Plunkett, M., et al. (2003), The absorption of gabapentin following high dose escalation, *Seizure*, 12, 28–36.

# *Ethical Issues in Clinical Research*

**Kelton Tremellen[1] and David Belford[2]**

[1]*Repromed, Dulwich, South Australia*
[2]*GroPep Limited, Adelaide, South Australia*

**Contents**

1 Introduction
2 Ethical Dilemmas Central to Conduct of Clinical Research
3 Philosophical Theory and Clinical Trial Ethics
  3.1 Deontological Theory (Kantianism)
  3.2 Utilitarianism
  3.3 Principlism
4 Historical Triggers for Development of Codes of Ethical Research Practice
5 Development of Guidelines for Ethical Conduct of Medical Research
  5.1 Hippocratic Oath (Fourth Century BC)
  5.2 Nuremberg (1949)
  5.3 Declaration of Helsinki (1964)
  5.4 National Research Act (1974) and Belmont Report (1979)
  5.5 CIOMS (1982)
  5.6 National Guidelines on Ethical Conduct of Human Research
6 Ethical Issues Design, Conduct, and Reporting of Medical Research
  6.1 Gold Standard: Randomized Clinical Trial
  6.2 Ethical Frameworks for Conduct of Clinical Trials
  6.3 Ethical Issues in Clinical Trial Design: Study Power and Control Groups
  6.4 Informed Consent
  6.5 Transparent Clinical Trial Culture: Financial Disclosure, Trial Registration, and Reporting of Results

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

7 Ethically Challenging Scenarios in Clinical Research
7.1 Children
7.2 Mentally Impaired
7.3 Women and Pregnancy
7.4 Critically Ill Patient
7.5 Research in Undeveloped World
8 Four Golden Rules of Ethical Conduct in Clinical Research
8.1 A: Respect for Patient Autonomy
8.2 I: Maximization of Research Impact on Medical Treatment
8.3 M: Minimization of Risk to Research Participants
8.4 S: Scientific Integrity
9 Conclusion
Appendix A: Conclusion
Appendix B: World Medical Association Declaration of Helsinki: Ethical Principles for Medical Research Involving Human Subjects
References

## 1 INTRODUCTION

> "It is not enough that you should understand about applied science in order that your work may increase man's blessings. Concern for man himself and his fate must always form the chief interest of all technical endeavors."
>
> —*Albert Einstein, Address to the California Institute of Technology, 1931*

> "There is the more reliable safeguard provided by the presence of an intelligent, informed, conscientious, compassionate, responsible investigator."
>
> —*Harry Beecher, New England Journal of Medicine, 1966*

The quest for new treatments to alleviate suffering and prevent disease is indeed a very noble cause. Medical research has the potential to create huge benefits to society and enhance the professional reputation of the researchers who are responsible for these discoveries. However, a researcher must never forget that concern for the participants of human research is paramount, irrespective of how important the research outcomes are to society in general. A researcher must never knowingly seriously compromise the welfare of a research participant—"the ends must never justify the means." According to Harry Beecher, careful training of researchers in both biomedical and ethical concerns is the key to preventing or minimizing serious mishaps. Therefore, the aim of this chapter is to educate clinical researchers in the common pitfalls of human experimentation and ways to avoid these ethical problems in pharmaceutical trials.

## 2 ETHICAL DILEMMAS CENTRAL TO CONDUCT OF CLINICAL RESEARCH

Ethics is a branch of philosophy that is concerned with answering the age-old questions about duty, honor, integrity, and justice. Research ethics is a branch of applied ethics that studies the ethical problems, dilemmas, and issues that arise in the conduct of research. Within the field of clinical trials three key ethical dilemmas often arise [1]:

1. *Good of Individual Versus Good of Society* Central to all moral theories is the need to respect and promote the rights of an individual. Natural or universal rights are those rights which are inherent in the nature of people and are not contingent on human actions or beliefs nor do they require legal enforcement by government or society. Philosophers believe that the rights to life and liberty (ability to act according to his or her own will) are the two guiding principles of natural rights. On the other hand, the promotion of social welfare of the community in general is also a central tenet in moral philosophy. Unfortunately these moral principles of respect for an individual's natural rights and the needs of society are often in conflict in the setting of clinical research. It is vitally important to use human subjects in research in order to gain valuable scientific knowledge that can benefit all of society, but this research has the potential to place individuals in harm's way and compromise their natural rights. Therefore the central ethical question in all clinical research is how best to protect the universal natural rights of an individual without compromising the scientific validity and social value of the research.

2. *Net Result of Risks Versus Benefits* In any clinical research setting with the potential to harm the participant physically or mentally, the researcher has an ethical obligation to ask the question whether the benefits of the research to the participant and/or the community in general outweigh the risks to the individual. It is normal to assess risks/benefits in terms of probability and magnitude and in the context of the underlying health status of the trial participants. For example, it would be considered unethical to conduct a research study for a medication with even a low risk of death (1%) in the setting of a non-life-threatening medical condition such as dermatitis or alopecia. However, it may be ethically acceptable to enroll terminally ill patients with grade IV astrocytoma in a trial of a new chemotherapy medication with a 1% chance of medication-induced death if this trial treatment gives them any real chance of cure. Conversely, it may be acceptable to place healthy phase I trial volunteers at significant risk of mild reversible adverse events such as headaches or nausea, but it is totally unacceptable to place them at any material risk of death or irreversible disability.

3. *Just Distribution of Benefits and Harms* All human experimentation raises the important ethical dilemma of equitable distribution of benefits and harms. The potential benefits to an individual of being involved in medical research include access to new potentially more efficacious treatments, supervision of clinical care by highly qualified medical staff with intensive monitoring, identification of secondary pathology during the trial entry examination, financial support, and improved self-esteem. Many people gain considerable benefit from knowing that their participation in research is advancing medical science, thereby helping other people in the future with the same medical condition that they are experiencing. Such altruistic motivation is a common positive psychological benefit of trial participation. Conversely, involvement in a clinical trial may create harm. The trial medication may produce physical or psychological adverse effects ranging from the relatively benign through to death. For example, if the participant is allocated to the placebo arm of a study, their involvement in research may delay or block their access to already established effective treatments. Finally, many trial designs require extensive monitoring which can be intrusive and tiresome to an individual whose quality of life may already be diminished by their underlying disease. If the benefits

of participation in a trial are likely to exceed the potential risks, natural justice would dictate that any individual, regardless of their age, race, gender, religion, or financial resources, should have equal access to the trial.

## 3 PHILOSOPHICAL THEORY AND CLINICAL TRIAL ETHICS

Philosophers have developed several different ethical theories/principles which may assist in making the morally correct decision in the field of clinical research [2]. The following discussion will summarize the main components of the three most useful theories/principles in the field of research ethics. Each of these theories/principles has its own strengths and weaknesses, making it important to not analyze a clinical research dilemma from a single philosophical point of view.

### 3.1 Deontological Theory (Kantianism)

Deontological theory is an approach to ethics that focuses on the rightness or wrongness of actions themselves, as opposed to the rightness or wrongness of the consequences of those actions [2]. It can be summarized succinctly as an ethical theory based on "moral duty or obligation." The German philosopher Immanuel Kant (1724–1804) is generally considered as the father of deontological theory. Kant claimed that an action is morally wrong, irrespective of the final outcome, if it is inconsistent with the status of a person as a free and rational being [3]. The duty to promote an individual's freedom and rationality is the guiding ethical principle of Kantianism, referred to as the categorical imperative. Stated differently, Kantianism requires an individual to treat both themselves and others as ends in themselves, never only as means to an end. All human beings have intrinsic moral dignity and we should not abuse, manipulate, harm, exploit, or deceive people in order to achieve specific goals, no matter how desirable that goal may be. Put quite simply, deontological theory can be condensed into "the ends never justify the means" and "do unto others as you would have done to yourself."

### 3.2 Utilitarianism

The basic guiding principle of utilitarianism is that the morally right thing to do is to produce the best overall outcome for the most number of people. The English philosophers Jeremy Bentham (1748–1832) and John Mill (1806–1873) developed the theory of utilitarianism [4]. Mills stated that the morally correct action was the one which created the greatest balance of happiness for the greatest number of people—somewhat akin to a moral "cost–benefit" analysis. Where deontological theory prohibits placing clinical trial participants at significant personal risk for the benefit of science or the community in general, utilitarianism states that this is ethically sound provided the expected outcomes for society are favorable. Utilitarianism is a commonly used theory in medical research as investigators and institutional review boards (IRBs) are always trying to find the appropriate balance between personal risk and benefit for the trial participant versus the benefits to society in general.

### 3.3 Principlism

The term *principlism* is used to describe four standard principles outlined in the Declaration of Helsinki [5] and the Nuremberg Code [6] that govern ethically responsible research practice. These principles include respect for autonomy, nonmaleficence, beneficence, and justice [7]. Principlism is not derived from a single philosophical theory, but rather combines several different philosophical perspectives [1]. For example, the right of individual freedom or autonomy is central to deontological theory while beneficence or the generation of "greater good" is a key feature of utilitarianism. As principlism borrows the best features of the available philosophical theories, it is generally considered the most applicable to resolving ethical dilemmas in medical research:

(a) *Autonomy* Autonomy is the basic human right to respect an individual's ability to make their own decisions regarding their own health and future. If a researcher is to respect an individual's right for self-determination, he or she must create a situation where the research subject is able to make an autonomous decision regarding participation in the research. This can only be enabled by providing adequate information about the participant's involvement in the trial (informed consent) and avoiding undue influence or coercion. To achieve informed consent, subjects must be given adequate information regarding the research procedures that they will undergo, the risks and benefits of being involved, and alternative medical therapies of proven benefit outside the clinical trial. This information must be provided in both oral and written form in simple lay language that can be easily understood by a person with only basic language skills. Furthermore, all study participants must enter voluntarily, free from coercision or undue influence, and understand that they may withdraw at any time. The provision of excessive financial reward or the threat of withdrawal of standard medical care would be considered an unethical violation of the ethical principle of subject autonomy.

(b) *Beneficence* Beneficence refers to actions likely to promote individual well-being and/or benefit others.

(c) *Nonmaleficence* The concept of nonmaleficence is best embodied by the Latin phrase *primum non nocere* ("first do no harm"). To avoid doing harm has been a guiding principle of medical treatment since the advent of the Hippocratic oath. Medical researchers have both a moral and legal obligation to avoid "unjustified" harm to trial participants. Some harm may be justifiable provided that the benefits outweigh the negatives. All medical research must be assessed on the balanced probability of harm versus benefit according to the ethical principles of beneficence and nonmaleficence.

(d) *Justice* Justice requires that there be fair procedures and outcomes in the selection of research subjects. The burdens and benefits of research must be equally distributed to include individuals of both genders, different ages, and various racial and ethnic backgrounds, not just the easily available or vulnerable to facilitate trial recruitment. In the past researchers have targeted middle-aged white men for trials, avoiding recruitment of children and women of reproductive age. This produces an injustice as it creates an inequality in medical understanding of how a particular treatment may work (or not work) in children and women while robbing these

disadvantaged groups of the opportunity to benefit from new clinical trial treatments.

## 4 HISTORICAL TRIGGERS FOR DEVELOPMENT OF CODES OF ETHICAL RESEARCH PRACTICE

Clinical research using the established scientific method of hypothesis generation, observation, and experimentation has been conducted throughout the world for many centuries. However, only relatively recently have such medical experiments been placed under the scrutiny of ethical analysis and the nature of human experimentation limited by published ethical guidelines. Unfortunately many researchers have either been ignorant or intentionally subjugated these established ethical principles, often leading to disastrous consequences. Some of the best known and studied examples of human experimentation gone awry are outlined in the paragraphs below as examples of why clinical research requires enforceable ethical guidelines to protect research participants and society alike.

The Tuskegee experiment is an example of medical research that breaks almost every ethical principle [1]. In 1929 the U.S. Public Health Service commenced a study to examine the prevalence of syphilis among poor uneducated blacks and its possible mechanisms of treatment. The study confirmed that mass treatment of syphilis was feasible and desirable, but by 1932 funding for treatment was suspended due to lack of money created by the Great Depression. However, in the town with the highest prevalence of syphilis (Tuskegee, Alabama) it was decided that researchers would continue to monitor participants for the long-term complications of untreated syphilis, thereby ascertaining the natural history of the disease in African Americans [8, 9]. Participants were administered "therapy" which was known to be ineffective (mercurial ointment and spinal taps) and coerced into being involved in the research by being offered financial incentives such as provision of free medical "care" and payment of funeral expenses! The Tuskegee research participants were not told that they had syphilis but rather "bad blood" and they were not offered alternative medical treatment, despite penicillin becoming an established medical cure for syphilis as early as 1945. It was not until 1972, following considerable public outcry, that the U.S. Department of Heath, Education and Welfare (DHHS) closed the study. This unfortunate experiment breaks several established ethical principles, including the right of participants to make an informed unbiased choice regarding their involvement in research (autonomy), the need to maximize individuals' well-being while minimizing harm (beneficence, nonmaleficence), and justice [9].

Medical experiments conducted by Nazi doctors of the Dachau, Auschwitz, and Buchenwald concentration camps are one of the worst examples in history of a researchers' subjugation of their ethical duties to research participants. These doctors were proponents of racial hygiene theory, which stated that people of non-Aryan background such as Jews were subhuman and therefore should not be afforded the same basic human rights as the Aryan people. As such, the Nazis believed that they were free to conduct experiments on concentration camp prisoners that can only be described as barbaric. These experiments included subjecting prisoners to extremes of temperature and pressure, ingestion of seawater, limb transplants

without medical need, injection of infectious agents to determine the effectiveness of new antibacterial drugs or vaccines, deliberate inflicting of skin wounds to provide an experimental model to assess new forms of wound therapy and exposure of subjects to high doses of radiation [1, 10]. In all of the experiments the aim was to gain knowledge that was helpful to the German military, not the treatment of underlying prisoner ailments. Furthermore, prisoners were not told what would happen to them, were not given the opportunity to refuse consent, and were not offered established medical treatment when complications arose. These gross violations of basic human rights and ethical principles by the Nazis during World War II were the impetus for the development of the first formal international guideline on the ethical conduct of medical research, the Nuremberg Code (1949).

There are many examples of researchers using vulnerable subjects for research to help facilitate recruitment [1]. In the 1950s New York University conducted experiments in which mentally retarded institutionalized children from Willowbrook State School were deliberately infected with hepatitis virus to enable study of the natural progression of the disease. These children's parents had unfair inducement applied by researchers to give their consent by offering to expedite admission of their children to the school if they agreed to be involved. Between 1960 and 1972 the Defense Atomic Support Agency funded a research program in which low-income, poorly educated principally black patients with cancer were administered massive doses of whole-body irradiation as a "treatment." Despite there being no scientific justification for treating a localized cancer with whole-body irradiation, this study was conducted so that the military could assess human tolerance to radiation exposure and effective methods of treatment. Finally there are multiple examples of prisoners being subjected to dangerous exposures of infectious agents and radiation in the interests of medical research. While many of these prisoners were informed of the nature of the experiments, they were coerced into involvement by incentives such as early release from prison or gifts of money and cigarettes. In the prison system, where an individual's liberties are severely limited, such coercion does not allow unbiased informed consent.

## 5 DEVELOPMENT OF GUIDELINES FOR ETHICAL CONDUCT OF MEDICAL RESEARCH

### 5.1 Hippocratic Oath (Fourth Century BC)

The Hippocratic oath is an oath taken by physicians governing the ethical practice of medicine. It is believed that the oath was written by Hippocrates, the father of medicine, in the fourth century BC. While the contents of the oath do not pertain directly to medical research, they do set down some valid ethical practices to which physicians conducting clinical trials should aspire. Two components of the oath have special relevance to the ethical practice of clinical research. First, Hippocrates outlines the need for nonmaleficence and beneficence in the phrase "I will prescribe regimes for the good of my patients according to my ability and my judgment and never do harm to anyone" and the decree to keep "myself far from all intentional ill-doing." Second, he speaks of the importance of educating the next generation of physicians. In the broadest sense, one can draw parallels between this requirement

for education and the ethical requirement for researchers to share knowledge gained during research with the general community.

### 5.2 Nuremberg (1949)

Following the end of World War II trials of war criminals were conducted before the Nuremberg Military Tribunal in order to investigate and punish those individuals who committed crimes against humanity. Specifically, they were in response to the inhumane Nazi human experimentation carried out during the war by individuals such as Dr. Josef Mengele [2, 10]. During the trial several of the accused argued that their experiments differed little from prewar experiments and that there was no law that set the boundaries differentiating legal from illegal human experimentation. Dr. Leo Alexander then submitted to the tribunal a list of six points that defined what should be considered legitimate human medical research. These six points were adopted by the trial verdict and an additional four points were added—together making the 10-point Nuremberg Code governing ethical human experimentation [6]. The contents of the Nuremberg Code are outlined in Appendix A. In summary, the code outlines the need for ensuring participant autonomy (Articles 1 and 9), beneficence (Article 2), and nonmaleficence (Articles 3–8 and 10) during human experimentation.

While an excellent starting point in the development of guidelines to govern ethical human experimentation, the Nuremberg Code was lacking in several key areas. First, the code contained no guidelines concerning the ethical conduct of research on individuals who are unable to given informed consent (e.g., children, mentally impaired, and critically ill). Second, it totally ignored the need for maintaining participant confidentiality and the need for justice regarding individual's opportunity to engage in medical research. These deficiencies were later addressed in the Declaration of Helsinki [5].

### 5.3 Declaration of Helsinki (1964)

The Declaration of Helsinki [5] was developed by the World Medical Association (WMA) as a set of ethical principles for physicians and other participants in medical research. It was originally adopted in 1964 in Helsinki, Finland, and has since undergone eight revisions, the latest significant revision occurring in 2000 (Appendix B). The declaration covers all of the major ethical protections outlined in the Nuremberg Code but also contains several important additions:

1. *Article 1* The declaration states that human material and identifiable human data require the same ethical protections given to individuals undergoing direct human experimentation. This means that researchers are not able to conduct experiments using human tissue/blood samples collected for a nonrelated purpose without the consent of the original research participant. Similarly, researchers are not able to use databases containing confidential medical details of individuals for the purposes of research without first gaining these individuals' informed consent.
2. *Articles 8 and 23–25* Here the declaration discusses restraints required for the ethical conduct of research on "vulnerable persons" such as children, the

mentally incompetent, and subjects in a dependent therapeutic doctor–patient relationship with the researcher.

3. *Article 9* The declaration states that the ethical protections set forward by the declaration cannot be reduced or eliminated by local legal or regulatory requirements but these protections should be considered universal worldwide.
4. *Article 12* Here the declaration states that due care for the environment and welfare of animals must be taken into consideration during the conduct of medical research.
5. *Article 13* There is an absolute requirement that all human research protocols should first be reviewed by an independent committee of relevant experts (ethical review committee) to ensure that the study design is scientifically robust and provides adequate protection of research participants.
6. *Article 21* This stresses the need for maintaining participant privacy and confidentiality.
7. *Article 27* Both authors and publishers of scientific journals have an obligation to honestly disclose the results of both positive and negative outcomes of human experimentation. Furthermore, any potential source of conflict of interest must be declared in all publications.
8. *Articles 29 and 30 and Associated Clarifications* Following public outcry concerning the use of a placebo in a U.S. Center for Disease Control study of the antiviral zidovudene to prevent perinatal transmission of HIV in the developing world, the declaration's provisions on acceptable "alternative care" were revised (2000). This revision held that developed-world standards of care should apply to the conduct of clinical trials in the developing world, irrespective of whether this standard of care is actually available to the public through the existing heath system. While the current declaration does not expressly prohibit placebo-controlled studies, it does set out some significant limitations. Article 29 states that "the benefits, risks, burdens and effectiveness of a new method should be tested against those of the best current prophylactic, diagnostic, and therapeutic methods. This does not exclude the use of placebo, or no treatment, in studies where no proven prophylactic, diagnostic, or therapeutic method exists." The declaration then goes on to state that it is acceptable to use a placebo in a research study where there is a proven alternative therapy provided that the condition being studied is a relatively minor one and the research participant will not suffer any irreversible harm and all research participants allocated to the placebo arm will later be offered the best available proven therapy at the conclusion of the study.

### 5.4 National Research Act (1974) and Belmont Report (1979)

Although the Nuremberg Code and Declaration of Helsinki provided excellent guidelines on what constituted the ethical practice of medical research, they did not have the power of legal mandated enforcement within individual countries. Starting in the United States, the Surgeon General (1966) and later the U.S. Food and Drug Administration (FDA) (1971) mandated prior peer review of all

human research protocols funded by the National Institutes of Heath or regulated by the FDA. This in turn led to the development of the National Research Act (1974), which mandated that any government-funded medical research would first require review and a pass judgment on the acceptability of the proposed research by an independent IRB. Furthermore, the regulations set down rules which would govern the composition of these IRBs, such as a mixture of men and women, scientific and nonscientific members, as well as general community members. The composition of these boards was to ensure that they had the capacity to assess the scientific validity of the research proposal as well as that ethical principles such as distributive justice, participant safety, and autonomy were respected while blocking research with the potential for inciting moral outrage in the general community.

When the National Research Act was signed into law on the July 12, 1974, The National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research was formed. This commission was charged with developing guidelines that should be followed during all human medical experimentation. The results of nearly 4 years of deliberations by the commission were published in 1979 as the landmark Belmont Report [11]. As would be expected, the report focuses on the need to respect participant autonomy, the need to maximize good (beneficence), and the application of justice.

### 5.5 CIOMS (1982)

The Council for International Organizations and Medical Sciences (CIOMS) was formed in 1949 by the World Health Organization (WHO) and the United Nations Scientific and Cultural Organization (UNESCO) with the goal to facilitate and promote international activities in the field of the biomedical sciences. As part of this mandate, CIOMS undertook important work in the field of bioethics during the 1970s, culminating in the publication of the "International Ethical Guidelines for Biomedical Research Involving Human Subjects" in 1982. This work has since been updated in 1993 and 2002 and is available in print [12] or online via the CIOMS website (http://www.cioms.ch/; CIOMS 2002 guidelines). The CIOMS document consists of 21 guidelines governing the ethical conduct of human biomedical research. These guidelines are heavily based on those contained within the Declaration of Helsinki but also contain several important practical discussions on how best to implement ethical principles in the design and conduct of a research. For example, guideline 4 stresses that informed consent is a continuing process that should be conducted throughout the life of a study, not just an event at the commencement of a participant's enrollment in a research study. The CIOMS document provides a very useful 26-point checklist that helps ensure that all the requirements of proper informed consent are met by a study [12]. Guideline 7 outlines in very practical terms what level of financial reimbursement is ethically acceptable for involvement in a trial and what would be considered unacceptable financial inducement. Appendix 1 of the CIOMS guidelines outlines the 48 individual points of discussion that need to be addressed by all research submission to a Human Research Ethics Committee/IRB to enable proper determination of the ethical merits of the proposed research.

### 5.6 National Guidelines on Ethical Conduct of Human Research

Most countries have their own guidelines governing what is ethically acceptable practice in the field of human experimentation. The Office of Human Research Protections, DHHS, publishes a summary of the individual regulations and legislation covering human research in 79 individual countries and two confederations (European Union and the Commonwealth of Independent States). This summary is updated regularly and can be located on the Office for Human Research Protections website [13]. While the legislation enacted in each country may differ, it should be stressed that no national ethical, legal, or regulatory requirement should be allowed to reduce or eliminate any of the protections for human subjects set forth in the Declaration of Helsinki (Article 9, 2000).

## 6 ETHICAL ISSUES IN DESIGN, CONDUCT, AND REPORTING OF MEDICAL RESEARCH

The aim of this section is to make the reader aware of aspects of the ethical debate surrounding the design, conduct, and reporting of a clinical trial. Several issues which are discussed in previous chapters are not discussed in detail here. These include stopping a clinical trial (Chapter Monitoring), the role of Data and Safety Monitoring Committees (DSMCs) and ethics committees (Chapter Clinical Trials Data Management), and patient records and privacy (Chapter Review Boards). It is further recognized that as the ethical debate is evolving, guidance documents and ethical codes are revised at regular intervals, and new legislative changes are under consideration. For example, at the time of this writing, the Declaration of Helsinki is being revised [5], an update of the CONSORT statement is being prepared [14], and federal legislation requiring disclosure of clinical trial results is being considered in the United States [15]. Therefore, where possible, we have provided current website addresses that direct the reader to the latest versions of the relevant documents (Table 1).

It is further noted that the ethical conduct of clinical trials extends beyond the study design, conduct, and reporting. Appropriate attention to the quality of manufacture of investigational drugs, validation of methods, and demonstration of stability underpins ethical evaluation of new drug candidates. Such considerations are now mandatory in most regulatory territories. In addition, when conducting nonclinical safety studies to support clinical trial programs, researchers must be aware of the applicable quality systems—Organisation for Economic Co-operation and Development (OECD) principles for good labor practices (GLP) [16]; GLP regulations, Code of Federal Regulations (CFR), Title 21, Part 58] [17]—and the Declaration of Helsinki, Article 12, which states that "the welfare of animals used for research must be respected" [5].

### 6.1 Gold Standard: Randomized Clinical Trial

A necessary, although not sufficient, condition for any experiment on human subjects is that the study is scientifically sound [18]. The principle of beneficence includes the requirement that the hypothesis being tested has scientific merit, that the design of the study adequately addresses the hypothesis, and the outcomes

**TABLE 1 Summary of Ethical Codes for Clinical Research—Web-Based Resources**

| Guidance/Ethical Code | URL or Literature Reference |
|---|---|
| Ethical codes and acts | |
| Nuremberg Code (1949) | http://www.hhs.gov/ohrp/references/nurcode.htm |
| Declaration of Helsinki (2004; first adopted by the 18th WMA General Assembly, Helsinki, Finland, 1964) | http://www.wma.net/e/ethicsunit/helsinki.htm |
| Belmont Report: Ethical Principles and Guidelines for the Protection of Human Subjects of Research (1979) | http://www.hhs.gov/ohrp/humansubjects/guidance/belmont.htm |
| CIOMS International Ethical Guidelines for Biomedical Research Involving Human Subjects (2002; first issued 1982). | http://www.cioms.ch |
| U.S. FDA 21 CFR 50, Protection of Human Subjects | www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfcfr/CFRsearch.cfm?CFRPart=50 |
| U.S. FDA 21 CFR 56, Institutional Review Boards | www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfcfr/CFRsearch.cfm?CFRPart=56 |
| ICH Harmonised Tripartite Guideline: Guideline for Good Clinical Practice E6 (R1) | www.ich.org |
| International Compilation of Human Research Protections, 2008 Edition, compiled by Office for Human Research Protections, U.S. Department of Health and Human Services. | http://www.hhs.gov/ohrp/international/HSPCompilation.pdf |
| U.S. FDA Best Pharmaceuticals for Children Act | http://www.fda.gov/opacom/laws/pharmkids/contents.html |
| Clinical trial registration | |
| WHO International Clinical Trials Registry Platform (ICTPR) | http://www.who.int/ictrp/en/ |
| Ottawa Statements | http://ottawagroup.ohri.ca/ |
| Financial conflict of interest: U.S. Department of Health and Human Services, Guidance Document: Financial Relationships and Interests in Research Involving Human Subjects: Guidance for Human Subject Protection | http://www.hhs.gov/ohrp/humansubjects/finreltn/fguid.pdf |
| Publication of clinical trials | |
| International Committee of Medical Journal Editors (ICMJE) Uniform Requirements for Manuscripts Submitted to Biomedical Journals: Writing and Editing for Biomedical Publication | http://www.icmje.org/index.html |
| The CONSORT Statement | http://www.consort-statement.org/, *JAMA* 2001; 285(15),1987–1991. |
| Good Publication Practice (GPP) for Pharmaceutical Companies | http://www.gpp-guidelines.org/, *Curr. Med. Res. Opin.* 2003; 19(3), 149–154. |
| Design of clinical trials: ICH Harmonised Tripartite Guideline: Choice of Control Group and Related Issues in Clinical Trials E10 | www.ich.org |

*Note*: All URLs last accessed December 2007.

of the study are valid. Beneficence further requires that investigators are competent to conduct the study and they seek to minimize the risks and maximize the benefits of the research [12]. The "gold standard" for determining the efficacy and safety of a new therapeutic intervention is the randomized controlled trial (RCT).

*Randomization* aims to eliminate systematic differences between treatment groups [19]. It removes any selection bias that may be introduced by those conducting the study and, ideally, balances confounding variables (some of which may be unknown) among the treatment groups, thus preventing "covariate imbalance" [20]. By diminishing the chance that baseline confounding variables differ among the treatment groups, the validity of conclusions relating the study outcomes to the intervention under investigation is enhanced. A more detailed discussion on randomization procedures is provided in Chapters Methods of Randomization and Randomized Controlled Trials.

*Blinding* involves the effective masking of both the investigational drug and control treatments (e.g., a placebo or comparator drug) in an attempt to ensure that all those involved in the study (i.e., trial participants and trial staff) do not alter their behavior in response to the knowledge of which treatment is being administered. In a "double-blind" trial, treatment allocation is unknown to the participant, investigator, sponsor, and trial staff, including those assessing outcome measures. Blinding thus aims to minimize potential bias in the management, treatment, and assessment of trial subjects that may otherwise occur if the treatments were known [19]. The term "single blind" refers to studies where one party (usually the subject) is blinded to treatment allocation.

The design of the RCT raises many complex ethical issues. For example, how can investigators be asked to suspend the principles of individualized patient care inherent in the Hippocratic oath and, blinded to prospective treatments, allow their patients to be randomized receive an experimental or control therapy? Further, can the use of placebo as a control be justified if currently available therapies are known to effectively treat the condition under investigation? How can a clinician condone trial-based medical procedures (e.g., the taking of biopsies) that have no therapeutic benefit for the patient? Indeed, should the roles of treating physician and clinical investigator be separated and filled by different individuals?

From the trial participant's perspective, does an RCT—in requiring patients to delay knowledge of their treatment until the randomization code is broken—contravene the principle of autonomy that would allow individual patients to determine which treatment they receive?

Ethical frameworks for the conduct of a clinical trial, with reference to ethical codes, are briefly introduced in the following sections.

## 6.2 Ethical Frameworks for Conduct of Clinical Trials

The concept of uncertainty in outcome, or "equipoise," has been invoked by many authors to resolve ethical considerations inherent in the conduct of an RCT. According to this paradigm, equipoise exists if the outcome of a trial is sufficiently uncertain such that an individual participating in the trial will not be disadvantaged; under these circumstances, enrollment in the trial is therefore considered "an equal bet in prospect" [21] and participants should not be harmed by random

allocation to a given treatment arm [22]. A trial in equipoise should, therefore, be acceptable to both Kantian and utilitarian philosophies [21], both of which seek to balance the duties and rights of the individual and society (see a brief description of each above). Thus, in the presence of true equipoise, individuals can, in theory, contribute to the wider medical knowledge at no cost to themselves [23]. The application of the concept of equipoise does, however, raise several questions. For example:

1. *From whose point of view should equipoise exist?* Options include the individual investigators, the clinical research community ("clinical equipoise," after Freedman [24]), or the prospective patient. Many commentators would consider equipoise as essentially a characteristic of the informed, competent patient [21]. Importantly, a trial that may be considered to be in equipoise for one trial participant may not be for another [23].
2. *What degree of equipoise should exist?* Any given outcome in a clinical study is rarely an "even bet." As a therapeutic candidate moves through the development process, a significant body of knowledge accrues, and it is unlikely sponsors would continue development unless previous nonclinical and clinical studies were favorable. Moreover, most published industry-funded trials report a positive outcome in favor of the drug candidate under investigation [22, 25]. This finding is not explained by publication bias alone and suggests that, in reality, very few industry-sponsored trials are in true equipoise. The question then becomes: At what point of "equipoise imbalance" does a trial become unethical? One study found that 50% of subjects perceived a trial to be unethical when 70% of experts favored one arm of the study; if this figure rose to 80% of experts, less than 3% of subjects thought a trial is justifiable [26].
3. *At what point is a decision regarding equipoise made?* The key decision faced by most trial participants is whether or not to consent to randomization. Equipoise in this context would mean the prospective participant perceives that agreeing to randomization gives them an equal chance of a good outcome as not participating in the trial [27].

The concept of equipoise has many critics. Some consider that the term lacks consistency, reality, and utility [28], while others maintain that it disregards the autonomy of the subject [22] and is fundamentally incoherent [29, 30]. Alternative ethical frameworks for the conduct of clinical trials have been proposed, including the "positive expected outcomes" [22]. Under this model, the trial participant would base a decision to participate in a given study on the pooled average of expected outcomes in all arms of the trial before randomization. In this context, some authors have noted that many trial participants achieve above average clinical outcomes even though the trial may have reported a negative result or the patient received a control treatment [31, 32]. Several reasons for this observation have been cited, including the cyclical nature of the disease being studied [20], the contention that clinicians who recruit to trials tend to be better clinicians [31], and the possibility that the strict protocol-driven procedures (often overseen by expert committees) provide a superior standard of care. It is emphasized, however, that improved medical care should not be offered as an inducement to participate in a clinical trial.

### 6.3 Ethical Issues in Clinical Trial Design: Study Power and Control Groups

***Study Power*** The social value and scientific validity of a study are critical to the ethical conduct of clinical research. Exposure of trial participants to even minor risks cannot be justified if the study is poorly conducted or is of little scientific merit. It if often argued in this context that underpowered trials that yield outcomes that are not clear-cut have the potential to mislead and are therefore unethical. In contrast, others would maintain that, if equipoise truly exits, then there is no ethical objection per se to participating in an underpowered study [21]; the patient does not "lose out in prospect." While this contention would seem to contravene a central tenet of ethical clinical research that demands validity and scientific rigor, there are examples where an underpowered trial may be better than no trial at all or where an underpowered trial represents a key step in the drug development process. These include studies on rare conditions, where alternative sources of data may be used to support the trial results, and early-phase studies that aim to determine estimates of efficacy and safety [33]. The authors point out that, in both cases, prospective participants should be informed of how the trial data may be used (e.g., inclusion in a subsequent meta-analysis) and that the trial may only indirectly contribute to health care benefits.

***Placebo-Controlled Trials*** Significant ethical concern and debate surround the use of placebo treatments as a control in a clinical study. Some authors have invoked the concept of equipoise to argue that, if an effective treatment is available, then placebo-controlled trials are unethical because equipoise cannot exist [21]. In contrast, others have argued that methodologically sound placebo-controlled trials, where participants are not exposed to excessive risks of harm, are ethical and that equipoise in this context does not provide an adequate ethical framework [30, 34].

Article 29 of the Declaration of Helsinki (issued in October 2000) states: "The benefits, risks, burdens and effectiveness of a new method should be tested against those of the best current prophylactic, diagnostic, and therapeutic methods. This does not exclude the use of placebo, or no treatment, in studies where no proven prophylactic, diagnostic or therapeutic method exists" [5]. This article was initially interpreted by some, in its strictest sense, to mean that the use of placebo controls was ruled out whenever an effective therapy was available. In 2002, the WMA added a note of clarification to Article 29, reaffirming "its position that extreme care must be taken in making use of a placebo-controlled trial and that in general this methodology should only be used in the absence of existing proven therapy. However, a placebo-controlled trial may be ethically acceptable, even if proven therapy is available, under the following circumstances:

- Where for compelling and scientifically sound methodological reasons its use is necessary to determine the efficacy or safety of a prophylactic, diagnostic or therapeutic method; or
- Where a prophylactic, diagnostic or therapeutic method is being investigated for a minor condition and the patients who receive placebo will not be subject to any additional risk of serious or irreversible harm" [5].

The CIOMs guideline 11 [12] elaborates on these points, further stating that a placebo control can be used "when withholding an established effective intervention would expose subjects to, at most, temporary discomfort or delay in relief of symptoms." Examples cited include a slightly raised blood pressure or serum cholesterol, common headache, or treatments for baldness or common cold.

In relation to the term "compelling and scientifically sound methodological reasons," it has been argued that the use of "active controls" does pose methodological concerns [35, 36]. Such views would hold that a noninferiority or equivalence study showing no difference between a study drug and active comparator could be interpreted to mean that both were effective, both were ineffective, or the trial (for whatever reason) was simply not able to detect a difference between them. In this regard it has been noted that, in several areas of medicine, not all "known" active treatments will reliably prove to be superior to a placebo in every study. This particularly applies to analgesics, antidepressants, antianxiety drugs, antihypertensives, antiangina drugs, anti–heart failure drugs, antihistamines, and drugs for asthma prophylaxis [35]. Moreover, trials designed to demonstrate that a new drug and an active control possess similar efficacy do not provide sponsors and investigators with the same "built-in incentives for trial excellence" [19] as studies designed to show a treatment difference. It is relatively easy for poorly designed and conducted studies to fail to show a difference between treatment arms. Thus, for some therapeutic areas, it is widely considered that reliable evidence of efficacy can only be obtained by superiority trials or the internal validation of active-comparator equivalence or noninferiority trials using a placebo. In the former case, the requirement to demonstrate superior efficacy over an authorized treatment may be considered too restrictive. Moreover, an active comparator trial may expose patients to risk and fail to deliver a clear-cut answer.

Thus, if invoking this clause to justify inclusion of a placebo control, the possibility of "additional risk of serious and irreversible harm," including the period of time a placebo treatment may be administered, needs to be carefully considered. All other ethical safeguards should apply to placebo-controlled trials including ethical and scientific review, full disclosure and informed consent, and the right to withdraw at any time. Additional strategies to minimize the risk of participants include the incorporation of predefined "stop" rules in the study protocol and the monitoring of data collected during the study by a Data Safety and Monitoring Board (Chapter Monitoring). Where appropriate, interim analyses should be undertaken to ensure trial participants are not exposed to the placebo for any longer than necessary or to assess the likelihood that the trial will deliver a clear-cut answer. If, in the latter case, this probability is low, the trial should be terminated. Study designs that seek to address the ethical and practical issues associated with the conduct of placebo-controlled trials are described in the International Conference on Harmonisation (ICH) guideline "Choice of Control Group and Related Issues in Clinical Trials" [19], including "early escape" and "randomized withdrawal" designs.

### 6.4 Informed Consent

Informed consent is a central tenet of clinical research. It is fundamental to all codes of ethical clinical research. The first principle of the Nuremberg Code states that "the voluntary consent of the human subject is absolutely essential" [6]. The

Declaration of Helsinki (Articles 20–26) clearly addresses the principles of informed consent, including reference to subjects who may have a dependent relationship with their physician, subjects who are legally incompetent or physically and mentally incapable of giving consent, and subjects with a condition that renders them unable to give informed consent. In obtaining consent, clinical researchers should adhere to the requirements outlined in the ICH "Guideline for Good Clinical Practice" (Section 4.8) [37], the CIOMS guidelines 5 and 6 [12], and any additional codes published by regulatory authorities (e.g., 21 CFR 50) [38].

Consent must be voluntary and based on sufficient information about the proposed study and the implications of participation to allow potential trial participants to make an autonomous decision whether or not to enroll. The Belmont report states that "no undisclosed risks may be more than minimal" [11]. Importantly, subjects must have realistic expectations as to the study outcomes, including any potential benefit to both the individual participant and society. The language used on the informed consent documents should be "as non-technical as practical and should be understandable to the subject or the subject's legally acceptable representative and the impartial witness, where applicable" [37]. It is important that the informed consent process continue throughout the course of the study [12].

The source of funding, financial relationships, and financial interests of the parties involved should be disclosed in line with the recommendations in the FDA guidance "Financial Relationships and Interests in Research Involving Human Subjects: Guidance for Human Subject Protection" [39] or local institutional policies. This FDA guidance further recommends that, if a potential or actual financial conflict of interest exists that could "influence the tone presentation or type of information presented during the consent process," then a third party who has no conflict of interest should be involved in the consent process.

***Coercion and Inducement*** Coercion and undue inducement are ethically unacceptable and can invalidate consent. While the issue of patient dependence is addressed in Article 23 of the Declaration of Helsinki, researchers should be alert to the fact that more subtle forms of coercion (or intimidation) can result from the position of authority held by an investigating clinician. Indeed, in one study of patients consenting to a major surgical procedure, consent was based more on the trust rather than an assessment of information provided to them [40]. All subjects should be assured that a decision whether or not to participate in the proposed clinical trial will not affect their medical care [12].

It is generally accepted that, for studies that aim to recruit patients into later phases (phases II and III trials), financial reimbursement should only compensate for direct costs of participation (such as travel to and from the clinic). Exceptions to this position may be considered in situations where administration or outcome-testing schedules cause significant inconvenience or when little or no benefit is expected to result from trial participation.

In the case of phase I trials, bioequivalence or pharmacokinetic studies, where participants receive no medical treatment that may be of benefit, a payment above out-of-pocket costs as compensation for inconvenience can be offered. However, such a payment should not be so great as to encourage participants to expose themselves to excessive risk. Moreover, payment contingent on completion of the trial

may prevent participants from leaving a study against their better judgment and could be deemed to constitute undue inducement.

Reasons for leaving a trial should be documented. In cases of withdrawal due to adverse reactions to the study drug, trial subjects should be encouraged to fully articulate all reasons for leaving the trial, and full payment should be offered. Subjects who withdraw due to personal reasons should be considered for a prorated payment. Willful noncompliance on the part of a participant, resulting in expulsion from the study, may be grounds for withholding all payment.

Any incentives offered to trial participants must be reviewed by an ethics committee before recruitment activities commence (see the Declaration of Helsinki, Article 13).

One interpretation of inducement deserves further comment. Although the benefits of participation in clinical trials have been discussed by several authors [20, 31, 32], the prospect of improved medical care cannot be offered as inducement to participate in a trial. However, application of this ethical principle in the context of (for example) teaching hospitals that service populations who may not have access to or cannot afford ongoing quality medical care presents several ethical challenges [18]. Does the prospect of access to medical care (including, e.g., transport to and from the clinic) constitute undue inducement? Does the recruitment of such individuals raise issues of distributive justice that would require the burdens and benefits of clinical research to be evenly distributed (see CIOMs guideline 12 [12])? Does participation in a trial by those seeking access to treatments not otherwise available to them constitute exploitation? In considering these issues, clinical researchers should be cognizant of the principles of justice implicit in relevant articles in the Declaration of Helsinki. In particular, Article 19 states: "Medical research is only justified if there is a reasonable likelihood that the populations in which the research is carried out stand to benefit from the results of the research." It is noted here that the definition of "populations" is open to interpretation, and there are many clinical research situations where the intended benefit is directed at the wider community. Further, Article 22 states: "The subject should be informed of the right to abstain from participation in the study or to withdraw consent to participate at any time without reprisal." Lastly, in relation to posttrial access of beneficial medicines, in its clarification to Article 30 of the Declaration of Helsinki, the WMA "hereby reaffirms its position that is necessary during the study planning process to identify post-trial access by study participants to prophylactic, diagnostic and therapeutic procedures identified as beneficial in the study or access to other appropriate care." This article raises several ethical and practical issues; importantly, posttrial access arrangements should be agreed to prior to commencing the study and such arrangements require prior review by an ethics committee. For a discussion of the wider ethical issues of research in populations or communities with limited resources, the reader is referred to CIOMS guidelines 10 and 11 [12].

***Competence*** Consent should be given by competent individuals with the capacity to understand the information provided to them. It is recognized that conditions for informed consent may vary with the nature of the study and the cultural and religious sensitivities of the participants and their communities. The Belmont report states: "it is necessary to adapt the presentation of the information to the subject's capacities" [11]. It is important that the ethical review process (see Chapter Methods

of Randomization) provides a further level of protection to the informed consent process. The ICH "Guideline for Good Clinical Practice" [37] requires that, before initiating a trial, the investigator must have obtained "written approval of the informed consent form and any other written information to be provided to subjects." Further, consent by a subject alone is not reason enough to justify participation in the trial. The Declaration of Helsinki (Article 15) states: "The responsibility for the human subject must always rest with a medically qualified person and never rest on the subject of the research, even though the subject has given consent."

Despite the best attempts of sponsors and investigators to clearly present information about a clinical trial to potential participants, several studies have shown that some degree of communication difficulty is inevitable [21]. For example, the concept of randomization and its justification may not be fully grasped by all individuals [29, 41], and patients who have recently received a diagnosis of a life-threatening disease may not be in a psychological state that enables them to absorb all the information provided to them [42]. In this context, provision of information should not be conducted in a manner that may cause distress; "respect for persons" does not only equate with autonomy [11, 43]. The degree of protection provided by investigators and IRB/Institutional Ethics Committees (IECs) should take into account both a patient's capacity for self-determination and the risk of harm and degree of benefit that participation in a clinical trial may entail.

Finally, all investigators and sponsors should carefully review their patient information sheets and informed consent forms before initiating their study. Are all risks adequately disclosed? Is the language in any way coercive? Are the chances of receiving a placebo treatment or active therapy fully spelled out? Are financial conflicts of interest disclosed? On a more operational note, Good Clinical Practice (GCP) audits commonly reveal shortcomings in the informed consent process, including use of outdated forms or forms that have not been approved by the relevant ethics committee, timing of signing of the informed consent form in relation to commencement of trial procedures, and forms not appropriately signed or dated by the subject or witness.

### 6.5 Transparent Clinical Trial Culture: Financial Disclosure, Trial Registration, and Reporting of Results

Several related initiatives that seek to increase the transparency of the clinical trials process have arisen in response to (1) concerns that financial interests in clinical research may be placing trial participants at risk [44]; (2) the delayed disclosure or nondisclosure of trial results and publication bias [45], including selective publication [46], selective reporting [47], and inadequate reporting [48]; and (3) inappropriate authorship and ghost writing of publications [49]. There is little doubt these concerns have significant ethical implications and momentum is building on several fronts to address them.

***Financial Disclosure*** Clinical researchers should be aware of and adhere to the ethical codes and guidance documents relevant to both their regulatory territories and institutional affiliations. The Declaration of Helsinki (Article 13) states: "The researcher should also submit to the (ethical review) committee, for review, information regarding the funding, sponsors, institutional affiliations, other potential conflicts of interest and incentives for subjects."

The document "Financial Relationships and Interests in Research Involving Human Subjects: Guidance for Human Subject Protection" was released by the DHHS in 2004 to help research institutions, IRBs, and investigators identify potential and actual conflicts of interest and suggest means to eliminate or manage such conflicts with the interests with the welfare of the trial subjects in mind. This guidance document, which is based on the ethical principles described in the Belmont report (respect for persons, beneficence, and justice), poses several questions that IRB members, investigators, and researchers should consider when determining whether or not a conflict of interest exists; it then suggests several "actions to consider" in each case. Many research institutions have adopted these principles.

It should be noted that some consider that the requirement for investigators to disclose any actual or potential conflicts of interest on the informed consent documents does not go far enough to adequately protect research subjects. Indeed, there is a lack of data on how potential clinical trial participants use disclosures in their decision-making process [44]. Moreover, disclosure alone, without further safeguards, may even confer a "moral license to exaggerate" on the part of investigators that could potentially increase any bias in the study that exists as a result of a conflict of interest [43]. Some therefore propose that additional mechanisms must be found to protect subjects; in many cases, the only answer would seem to be to encourage investigators conducting a trial to divest any possible conflicts of interest.

***Registration of Clinical Trials*** The participation of any subject in a clinical trial that does not contribute to the medical knowledge of the wider research community contravenes the obligations of clinical researchers. Nondisclosure or delayed disclosure of trial results may mean that the welfare and safety or trial participants could be adversely affected by subsequent clinical trial designs and ethical review processes that do not build on prior research knowledge. On a wider scale, failure to consider all available data may misdirect those responsible for health policy decisions and resource allocation. The Declaration of Helsinki includes several statements relevant to these considerations:

*Article 16* "Every medical research project involving human subjects should be preceded by careful assessment of predictable risks and burdens in comparison with foreseeable benefits to the subject or to others.

The design of all studies should be publicly available."

*Article 11* "Medical research involving human subjects must ... be based on a thorough knowledge of the scientific literature, other relevant sources of information. ..."

*Article 19* "Medical research is only justified if there is a reasonable likelihood that the populations in which the research is carried out stand to benefit from the results of the research."

Registration of clinical trials on publicly available databases represents a key mechanism to promote transparency, knowledge sharing, and the timely and accurate disclosure of clinical information. A global movement for registration of clinical trials is gaining momentum, driven by bodies such as the Ottawa Group [50, 51], WHO [52], and the International Committee of Medical Journal Editors (ICJME)

[53, 54] (see below) and the trial registries. Many trial registries currently exist, including the International Standard Randomized Controlled Trial Number (ISRCTN) register and ClinicalTrials.gov, which is administered by the National Library of Medicine. The WHO International Clinical Trials Registry Platform (ICTRP) project aims to set international standards for trial registration, ensure a minimum set of results are reported, and provide global trial identification and search capability. The ICJME policy, requiring trial registration as a prerequisite to publication, has resulted in a marked increase in the number of trials registered [15, 54].

However, registration of trial information does not currently require that trial outcomes will be disclosed and there are no international standards for reporting of clinical trial results [55]. In response, many now advocate systems to facilitate full public disclosure of published and unpublished trial results. Significantly, in the United States, both federal and state legislation are addressing such concerns. In Maine, a law that came into effect in October 2005 requires companies wishing to advertise prescription drugs to disclose trial results, including adverse effects. In addition, federal legislation currently under consideration will require reporting of results in government databases [15]. For example, the Fair Access to Clinical Trial Act of 2005 would see the ClinicalTrials.gov database expanding to include a clinical trials registry and clinical trials results.

Significant challenges to the registration of clinical trials and disclosure of results lie ahead, including agreement between the pharmaceutical industry and other stakeholders as to the minimal data set for trial registration [51], validation of the registry data, prevention of duplicate registrations, and defining and naming of interventions and the format of reporting and validation of results [15]. Nevertheless, the climate is changing. As noted by a recent ICMJE editorial, "three years ago trial registration was the exception; now it is the rule" [54].

***Publication of Clinical Trial Results*** "Both authors and publishers have ethical obligations. In publication of the results of research, the investigators are obliged to preserve the accuracy of the results. Negative as well as positive results should be published or otherwise publicly available" (Declaration of Helsinki, Article 29).

In addition to the registration of trial information and initiatives under way to promote the public disclosure of trial results, several important policy and guidance documents have been developed by professional bodies to promote the quality of reporting of clinical trials in biomedical journals.

*International Committee of Medical Journal Editors* Three policy initiatives of the ICMJE, a committee representing 12 member biomedical journals, are discussed below. These include the requirement to register trials as a prerequisite to publication and requirements for authorship and disclosure of any conflicts of interest on published manuscripts. Additional information can be found in the Uniform Requirements for Manuscripts (URM) Submitted to Biomedical Journals, a set of ethical principles and recommendations developed by the ICJME to promote the accurate reporting of biomedical studies [56]. These guidelines have been adopted by many journals.

1. *Requirement to Register Trial: Prerequisite to Publication* In 2005 the ICMJE established a policy that requires information about trial design to be prospectively registered in recognized databases as a prerequisite to publication in member journals as follows [53]:

> The ICMJE member journals will require, as a condition of consideration for publication, registration in a public trials registry. Trials must register at or before the onset of patient enrollment. This policy applies to any clinical trial starting enrollment after July 1, 2005. For trials that began enrollment prior to this date, the ICMJE member journals will require registration by September 13, 2005, before considering the trial for publication.

A marked increase in the number of registrations lodged with the five recognized registries was seen around the time of implementation of the policy. The largest existing registry, ClinicalTrials.gov, which registered around 30 trials per week before September 2005, received over 200 registrations per week after September 2005 [15]. The number of registrations continued to grow as many other biomedical journals embraced this policy.

Although early-phase studies (including phase 1 trials and pharmacokinetic studies) were initially excluded from this requirement [53], the ICMJE has recently adopted the WHO definition of clinical trials as "any research study that prospectively assigns human participants or groups of humans to one or more health-related interventions to evaluate the effects on health outcomes" [54]. Thus, as of July 1, 2008, all early-phase studies will require, as prerequisite to their publication, registration on a recognized database. Importantly, in addition to the five previously recognized registries, the ICJME will now accept registration on primary registries participating in the WHO ICTRP project [54].

2. *Requirements for Authorship* The URM lists three requirements for authorship as follows [56]:

- Substantial contributions to conception and design, or acquisition of data, or analysis and interpretation of data
- Drafting the article or revising it critically for important intellectual content
- Final approval of the version to be published

Authors should meet all these conditions and no authors who fulfill these criteria should be omitted. The URM further states that "acquisition of funding, collection of data, or general supervision of the research group, alone, does not justify authorship." While these guidelines have been widely adopted, the requirements of individual journals should be observed.

3. *Conflict of Interest* In-line with the Declaration of Helsinki (Article 27), which states, "Sources of funding, institutional affiliations and any possible conflicts of interest should be declared in the publication," the URM requires: "When authors submit a manuscript, whether an article or a letter, they are responsible for disclosing all financial and personal relationships that might bias their work. To prevent ambiguity, authors must state explicitly whether potential conflicts do or do not exist." Again, individual journals often have their own conflict of interest policies.

*Good Publication Practice (GPP) for Pharmaceutical Companies* The GPP guidelines set forth principles that aim to increase the transparency of the processes

involved in the publication of industry-sponsored trials, reduce publication bias, and clarify the roles of the sponsor company and academic investigators [57, 58]. The guideline includes topics such as publication standards, premature publication, duplicate publication, authorship, and the role of professional medical writers.

A further comment on the role of professional medical writers is made here. Ghost writing occurs when a professional writer controls the content of the manuscript and discloses neither their involvement nor their funding source. This is ethically unacceptable. However, working with a professional medical writer should pose no ethical concerns provided that (1) the medical writing assistance is acknowledged and (2) their source of funding is disclosed [59]. Throughout the writing of the manuscript, authors should maintain responsibility and control for the content of the publication, including key messages and data to be presented.

*CONSORT (Consolidated Standards of Reporting Trials) Statement* The CONSORT group (which comprises trialists, methodologists, and medical journal editors [60]) has developed several initiatives to improve the reporting of randomized clinical trials, including the CONSORT statement. The CONSORT statement is an evolving guideline that was originally published in 1996 [61] and revised in 2001 [48]. Following a meeting of the CONSORT group in January 2007, a further update of the CONSORT statement is due for release. The latest versions of all CONSORT documents can be found on the CONSORT group website [60].

In its current form, the CONSORT statement includes a 22-item checklist and flow diagram that is intended to improve the quality of the writing and assist in the review and evaluation of RCT reports. In particular, the checklist items encourage transparency in the reporting of the trial methods and results, which together account for 17 of the 22 items. The flow diagram depicts the number of participants progressing through the enrollment, intervention allocation, follow-up, and data analysis stages of a clinical trial, allowing the reader to evaluate the basis of any analysis undertaken. Although the main CONSORT statement is directed at two-group parallel RCTs, extensions to the statement address alternative trial designs, including cluster trials [62] and noninferiority and equivalence randomized trials [63]. A checklist of items that should be included in conference or journal abstracts is due for publication in 2007 [60]. It is highly recommended that the requirements of the CONSORT statement be considered early in the trial design process.

## 7 ETHICALLY CHALLENGING SCENARIOS IN CLINICAL RESEARCH

### 7.1 Children

Children, by virtue of their developmental and cognitive abilities, are a vulnerable population in relation to research participation. The Willowbrook Hepatitis Study [2] and many other similar questionable pediatric research studies have shown us that unscrupulous researchers have resorted to the use of children in research trials because they were seen as easy to recruit and manipulate. With the advent of the Nuremberg Code, which required informed consent, research in children declined because it was seen as almost impossible to get fully informed consent from a child of limited cognitive ability. For this reason and the fact that pharmaceutical manu-

facturers saw pediatric markets as too small to warrant expensive pediatric clinical trials, the majority of medications in clinical use today have not been properly tested in children but are being used "off label." This in itself is a major ethical dilemma as several important differences in disease pathophysiology, pharmacokinetics, pharmacodymanics, and adverse drug reactions occur between children and adults. For example, the effective dose of a single application of an asthma nebulizer will be higher in an adult with effective hand/breathing coordination than a relatively uncoordinated small child. The ability of diazepam to cause sedation in adults but agitation in some children is an example of the wildly varying physiological responses to a drug. The absence of adequate studies in the pediatric population hinders the effective provision of the medical care of children. As such, there has been a recent push by U.S. and European regulators to require drug manufacturers to assess the safety and effectiveness of medications in pediatric patients when pediatric use is anticipated [64, 65]. The U.S. Best Pharmaceuticals for Children Act—2002 creates financial incentives for pharmaceutical companies to engage in pediatric research by allowing a 6-month extension to their exclusive patent protection and funding for research of drugs that no longer have patent protection.

While distributive justice mandates that research should be conducted on children to enhance their health care, there are some instances when it is absolutely inappropriate to conduct research on children [66]. It is ethically appropriate to exclude children from research when:

(a) The research topic is irrelevant to children (e.g., ischemic heart disease, alzheimer's disease).
(b) Existing data in adults are sufficient to judge risks in children.
(c) Local laws or regulations bar inclusion of children in this type of research.
(d) A nontherapeutic study which places a child at greater than small increase in minimal risk. Minimal risk is defined as a risk that is "greater than those already encountered in daily life or during the performance of routine physical or psychological examinations or tests." The FDA, DHHS, and National Institutes of Health (NIH) have produced guidelines on categories of risk in pediatric research and the associated approval requirements [66]. While the degree of risk associated with "routine life" is open to varying interpretations by individual IRBs, it would generally be considered unethical to expose healthy children to any material risk of a serious adverse outcome, such as a phase I study of a new class of drug.

Pediatric research raises two potential dilemmas above and beyond the ethical dilemmas routinely seen in adult research [65–67]:

(a) *Informed Consent* Informed consent, a process where research participants are given sufficient information about a study to allow them to determine if it the risks and benefits warrant their own involvement, is not possible for a child with limited cognitive powers below the legal age of "consent." Pediatric research therefore requires a two-part consent process—parental permission and the child's assent. Central to the requirement of parental permission is the assumption that a loving, protective relationship exists between the

parent/guardian and the child. This assumption has two weaknesses. First, the theologian Paul Ramsey [68] states that a parent does not have the moral authority to permit a child to be put at risk of harm in an experiment from which the child could not benefit medically as a child is not a piece of property but an autonomous individual. Second, in dysfunctional relationships the parents may not always have the child's best welfare at hand. Therefore, if the researcher has any suspicion of a dysfunctional relationship between child and parent, they should defer any decisions concerning that child's participation to an independent IRB. After a parent has given permission for their child to be involved in research, the child's assent must be sought. The child should be given a simple explanation about why the research is being conducted and what procedures and discomfort may occur if they are to be involved. Researchers also need to define risks with the child's perspective in mind, not that of an adult. For example, separation from parents and unfamiliar surroundings may be considered a risk from the perspective of a child but not an adult. Assent should be sought apart from the parents' direct influence and the child must be informed that it is their right to refuse involvement. While childhood assent is always preferable, there are several instances where it is not mandated. For example, assent is not required if the child's capacity to understand the research is limited (intellectual age of less than 7 years, American Academy of Pediatrics) or the potential direct benefit to the heath of the child is high and only available through direct participation in the research (e.g., new chemotherapy for a childhood cancer). Apart from these limitations, failure to obtain a child's assent overrides a parent's permission and prohibits that child's involvement in research.

(b) *Payment for Research Participation* Payment to pediatric research participants becomes particularly problematic because of the child's vulnerability and dependence on their parents as the primary decision makers. Poor families are particularly vulnerable to monetary rewards, and therefore offering large financial incentives may create unnecessary coercion by the parents for their children to participate in research. Only small token gestures of appreciation or reimbursement of direct costs of involvement should be made to ensure payment is not the reason parents are enrolling their children in the research. Furthermore, incentives given to children should not be excessive and, according to American Academy of Paediatrics guidelines [69], should only be revealed after enrollment so that the child's participation is voluntary. This however is contrary to usual informed consent guidelines where incentives should be discussed before enrollment. A reasonable compromise may be to mention in the consent process that a gift will be given to participants without outlining its exact nature.

### 7.2 Mentally Impaired

Over the last 50 years very effective pharmacological treatments for mental disorders such as depression, psychosis, bipolar disorder, and obsessive-compulsive disorder have seen the transition of mental health care from institutions to the general community, resulting in huge benefits to patients and society in general. Without research, none of these life-changing benefits would have materialized.

Guideline 15 of the International Ethical Guidelines for Biomedical Research Involving Human Subjects [12] states that research on mentally impaired individuals is permissible provided that the purpose of the research is to obtain knowledge relevant to the particular heath needs of the mentally impaired and that the research cannot be equally well carried out on persons with full mental capacity.

According to Michels [70] the conduct of clinical research in mentally ill patients raises three key ethical questions. Are mentally ill subjects especially vulnerable to exploitation? Are they competent to give informed consent? Are psychiatric research methods particularly dangerous? Following the suicide of a patient with schizophrenia after being asked to cease their antipsychotic medication in the "wash out" period of a clinical trial, regulators have been struggling with the ethical dilemma of whether mentally ill patients should be allowed to participate in research. The U.S. National Bioethics Advisory Commission (NBAC) and DHHS have both published guidelines regarding the ethical conduct of clinical research in mentally ill patients [71, 72]. A summary and discussion of these recommendations are presented below:

(a) *Impaired Capacity* According to the NBAC, mental capacity to consent to research requires that the subject understand the difference between treatment and research, the nature of the research being conducted, its risks and benefits, available alternatives, and the fact that he or she is making a decision and that this decision can be changed. The subject must not be swayed by a pathological affective state, a false belief, or a dependent relationship that might interfere with the decision or his or her autonomy and must be capable of making a stable, reasoned choice and communicating it. It should be noted that merely having a mental illness does not mean that an individual's mental capacity to give informed consent is impaired. Different mental illnesses have varying effects on a person's decision-making capacity. Furthermore, illnesses such as schizophrenia can produce fluctuating levels of capacity as an individual passes through stages of good and bad mental heath. It is therefore advisable that researchers carefully consider the timing of recruitment of participants into research studies to avoid periods of heightened vulnerability due to acute mental illness. Furthermore, while not yet mandated, it may be preferable to get an independent expert in mental health to assess a potential research subject's capacity to give informed consent, especially if this research places a participant at more than minimal risk. Finally, as informed consent is a continual process, a research participant's capacity to consent should be continually monitored throughout the entire life of the study, preferably by an independent monitor.

(b) *IRB Provisions* Under U.S. regulations, IRBs that regularly review research involving vulnerable persons should have one or more individuals on that IRB who are expert within the field. In the area of mental health research it has been suggested that an expert in the area of mental health research and an individual from a relevant mental health support group would provide the best assessment of research protocols and therefore protection of research participants.

(c) *Surrogate Permission* Where permitted by law, individuals with impaired capacity may have a family member or other legally authorized representa-

tive serve as a surrogate for research decisions. Surrogates should be informed of the risks, benefits, and alternatives to the research when they are providing permission. The surrogate should be educated that it is their duty to first act in a manner that reflected the views of the mentally incapacitated individual while they were decisionally capable (substituted judgment). If the values of the individual are unknown, a "best interest" standard should be used. The autonomy of the patient should be respected and therefore their assent to participation in research should be obtained whenever possible. Their right to withdraw from research should be respected unless the research treatment may be in the best interests of the mentally incapacitated patient and not available outside of a research setting.

### 7.3 Women and Pregnancy

Prior to the 1990s women were systematically excluded from participating in research studies primarily because of the fear that unrecognized pregnancy might place the early embryo at risk, increasing liability risk for researchers. Such a policy of exclusion is not just since it leads to an unequal knowledge base regarding the correct use of medications in the male and female population, potentially reducing the quality of clinical care provided to women. Differences in body composition (fat to muscle ratio), hormone fluctuations during the menstrual cycle, and changes in physiology specific to pregnancy make it highly likely that significant differences in pharmacokinetics and pharmacodynamics exist between the two sexes, therefore mandating more balance research protocols [73].

Guideline 16 of CIOMS International Ethical Guidelines [12] states that the potential for becoming pregnant during a study should not in itself be used as a reason for precluding or limiting participation. A paternalistic approach where an investigator, sponsor, or IRB limits a woman's capacity to be involved in research because of a potential risk to the fetus infringes a woman's autonomy. McCullough [74] argues that a previable fetus does not have the moral or legal status of an autonomous individual and as such a woman should have the absolute right to make an autonomous decision concerning her own involvement in potentially harmful research without regard for the well-being of a nonviable fetus. However, in order to sensibly minimize risk to a potential unborn child, ethical guidelines mandate two precautions when conducting research in women of reproductive age [75]. First, research participants should be screened for pregnancy with sensitive and reliable pregnancy tests before enrolment if their participation in the research poses more than a minimal risk to the fetus or if being pregnant may materially increase their own risk during the conduct of the study. Second, researchers should guarantee access to effective contraceptive methods before enrollment in research. Research protocols should not mandate certain types of contraceptive use (e.g., IUCD, contraceptive pill) and respect the women's autonomy to choose a method, including abstinence, according to her own needs and values.

Clinical research on pregnant women carrying a potentially viable fetus (post 23 weeks gestation) is a larger ethical dilemma as here the fetus is an individual by virtue of its ability to survive ex utero. A general prohibition of termination of pregnancy beyond the period of potential viability provides legal support for this moral concept that a fetus's own heath must be considered when conducting research

on women in the later stages of pregnancy. McCullough [74] outlines four criteria for the ethical conduct of phases I–II studies in pregnant women:

(a) The medication should be reliably predicted to alter the course of pregnant women's condition based on prior animal and human studies.
(b) Previous animal and/or human studies have documented no deaths or serious or irreversible injury to pregnant women that was caused by the medication.
(c) Previous animal and/or human studies have documented no deaths or serious or irreversible injury to the fetus that was caused by the medication.
(d) Previous animal and/or human studies have reported no or only a very low risk of less serious injury to the fetus that was caused by the medication.

Pregnant women should be provided with adequate information regarding the known benefits and risks of research to their own health and that of the fetus to ensure adequate informed consent. Appropriate follow-up of heath outcomes of children exposed to an investigational agent during pregnancy should also be arranged to confirm that the investigational agent is not harmful to the fetus.

Consent of a pregnant woman alone is usually sufficient for most research which may have an impact on the fetus. In the United States 45 CFR 46 states that "when research holds out the prospect of direct benefit solely to the fetus then the consent of the pregnant woman and the father should be obtained." However, international guidelines such as CIOMS [12] and the American College of Obstetricians [73] do not support this need for paternal consent as paternal rights may weaken or infringe maternal autonomy—the right of a woman to take action in decisions that affect her own body and heath. At present, the need for maternal, not paternal, informed consent is standard ethical practice in medical research.

### 7.4 Critically Ill Patient

Clinical research involving critically ill patients is necessary to reduce the extreme morbidity and mortality encountered in the intensive care unit (ICU). However, this type of research is ethically challenging for two principal reasons [76]. First, critically ill patients are captive and thus vulnerable as they are dependent on the ICU team for all aspects of care or even life itself. This vulnerability raises concerns about the patient or their surrogate's ability to give informed, autonomous consent to participate in clinical research, even if they can judge the pros and cons of participation. Second, most critically ill patients lack decisional capacity due to their illness (head injury, delirium, and drug-induced coma) or the drugs (sedatives, analgesics) that they are receiving. As a result, their family members are usually asked to provide permission for research participation. While a recent study concerning research in the ICU suggests that surrogates do predict the choices of patients with a reasonably high degree of accuracy [77], there is the potential for a surrogate to make a decision that ultimately contravenes the participant's real wishes.

At the present moment there is no legislation in place governing research in the setting of the critically ill patient. The American Thoracic Society [78] and the American Heart Association [79] have published some very useful guidelines concerning involvement of critically ill patients in clinical research. These guidelines

mirror issues discussed in previous paragraphs related to research in vulnerable persons such as the mentally ill. However, we will outline some potential solutions specific to the ICU setting:

(a) *Separation of Clinical and Experimental Team* Due to the extreme vulnerability of the critically ill patient and their dependence on the ICU team, it is probably best that researchers separate from the clinical team ask patients or their surrogates for permission to be involved in any study. This ensures that patients/surrogates do not confuse research with therapeutic intent and can feel free to refuse involvement without any perceived pressure that such refusal may harm the standard of routine care being provided to the patient.

(b) *Consent* As most patients are not in a position to give their own consent, it is permissible for family members to provide surrogate consent provided that the research is of minimal risk or may have some direct benefit to the patient. It is probably unethical to allow a surrogate to make a decision regarding a third person (the patient) that places that individual at more than minimal risk. Alternatives to the use of surrogate consent include retrospective consent, prospective consent, and waiver of consent by the IRB. An example of retrospective consent would be when a small sample of blood is taken as part of routine clinical care and then stored for potential later use in research once the patient has recovered and given consent. In prospective consent a noncritically ill patient is consented for research in the event that he or she may become critically ill later. For example, preoperatively a patient undergoing major surgery may be consented to research on a new therapy to prevent lung infections related to prolonged artificial ventilation in the ICU. Finally, in limited studies on true emergencies where the intervention is minimal, an IRB may waive the need for patient or surrogate consent. For example, a randomized trial between a manual CPR technique and a new but FDA-approved mechanical CPR machine may be undertaken without consent as both are reasonable care and there is no time for consent in a cardiac arrest [29].

### 7.5 Research in Undeveloped World

The conduct of clinical research in undeveloped countries raises three principal ethical concerns [80, 81]. First, many individuals from the undeveloped world have only very basic education, making it difficult for them to understand abstract concepts such as placebos, making informed decisions on the benefits and disadvantages of research involvement difficult. Furthermore, health care expenditure in the undeveloped world is paltry compared to western society, with even basic medical care often lacking. If an individual has the chance of receiving some medical treatment in the form of a trial medication, compared to no chance of medical care outside of a research protocol, does this really allow for an autonomous unpressured decision? For these reasons it is potentially more difficult to obtain truly informed consent in the undeveloped world.

The second major ethical issue surrounding research in the developing world is exploitation of communities that are unlikely to benefit themselves from the medical

advances developed by the research. Guideline 10 of the CIOMS ethical guidelines [12] states that "research must be responsive to the needs and the priorities of the population or community in which it is carried out." Furthermore, the guideline goes on to state that "any intervention or product developed, or knowledge generated, must be made reasonably available for the benefit of that population or community." It is only just that a community that may potentially be harmed by research should also share in its final benefits (new proven therapy). Unfortunately this has not always been the case. For example, the AIDS Clinical Trials Group protocol 076 was a study in which pregnant women from the undeveloped world were randomized to receive either placebo/routine care or oral antiviral medication throughout pregnancy, intravenous antiviral during labor, and postpartum oral antiviral for the baby in order to determine if such intensive treatment could reduce the vertical transmission of HIV to the child. While this trial did show a very significant reduction in vertical transmission, it was heavily criticized as unethical because the proposed treatment is unlikely to ever be reasonably available to mothers in the third world because of significant cost. The trial protocol cost $800 per woman and child, 600 times the annual per-capita allocation for heath care in Malawi [82]. As such, while proven to be an effective therapy for women in the western world, the results of the trial are unlikely to ever be practically applicable to the community in which the information was gained, making the study exploitative and unethical. Clinical trials can only be ethically performed in the developing world if the pathology in question is clinically relevant to the developing community and the trial intervention, if proven to be effective, is likely to be reasonably available to the community after completion of the research.

The final ethical concern for research in the developed world is an appropriate choice for the "alternative" control treatment in a trial. Some commentators suggest that it is ethical to use a placebo in a clinical trial in the undeveloped world, as no treatment is the default "standard of care" present in these poor countries [83]. These individuals suggest that since half the participants will be exposed to a potentially useful study medication, the "greater good" is still being done, making the study ethical. However, the majority view is to the contrary. The CIOMs guideline [12] states that research participants in the control arm of a trial should receive an established effective intervention (guideline 11). The use of a placebo is only ethically justifiable if there is no established effective intervention or delaying the use of this established intervention would not add any serious risk or cause any irreversible harm. As the CIOMs Guideline 11 states [12], "An economic reason for unavailability of an established effective intervention cannot justify a placebo-controlled study in a country of limited resources when it would be unethical to conduct a study with the same design in a population with general access to the effective treatment outside the study." The reader is referred to the CIOMs Guideline for a wider discussion of this topic.

## 8 FOUR GOLDEN RULES OF ETHICAL CONDUCT IN CLINICAL RESEARCH

After consideration of the various philosophical theories related to medical ethics and review of the relevant published guidelines it is possible to summarize ethical

research practice into four guiding principles. These principals, or golden rules, are summarized by the acronym AIMS to help facilitate recall, where A refers to respecting patient autonomy, I stands for maximizing the potential impact of the research on society in general, M refers to the mitigation of potential harm to research participants, and S stands for maintance of scientific integrity in all aspects of clinical research. The following sections outline the key features of each of these golden rules.

### 8.1 A: Respect for Patient Autonomy

Researchers should respect an individual's right to make decisions regarding their future by providing them with information which honestly outlines the risks and benefits of participating in the research in terms that are easily understandable to the average lay person of limited education (informed consent). Researchers should avoid offering excessive financial or other incentives that may impede unbiased decision making on the part of the potential research participant. When a subject lacks the ability to make an informed decision themselves (child, mental incapacity), the researcher must gain informed consent from a legally authorized representative. Research participants must feel free to exit a study at any stage without fear of compromising their standard medical care. Research participants should expect the same degree of respect of their privacy and dignity as would be afforded patients undergoing standard medical care. Maintenance of individual participant's privacy during the conduct of the trial and in the publication of trial results is paramount.

### 8.2 I: Maximization of Research Impact on Medical Treatment

The impact of a piece of research on contempary medical understanding/treatment can be maximized by considering four issues. First, research protocols should be scientifically well designed so that they have the capacity to answer the question of interest. Poorly designed trials with clinically irrelevant endpoints, inadequate sample sizes, or inherent biases that make firm conclusions impossible are unethical as they consume scarce resources while placing participants at risk of adverse events for no real gain. Proper peer review and IRB assessment should prevent such studies from even commencing. Second, it is important that trials are not designed with very limited inclusion criteria (e.g., white, male, normal weight, 30–40 years age) as the results of these types of trials may not be applicable to the population in general. Justice requires that, if possible, a broad cross section of the community should be approached to be involved in research so that the potential benefits are equitably distributed throughout society. Third, it is a moral imperative that the results of research must be published within the scientific literature, irrespective of the research outcome. Failure to publish negative outcomes could create a situation where an identical study is replicated by an independent research group unaware of the original study outcome, thereby exposing the second group of research participants to unnecessary risk while wasting valuable research resources. Finally, researchers should only conduct experiments that have scientific and social worth and not use human beings in frivolous research.

### 8.3 M: Minimization of Risk to Research Participants

According to the theory of utilitarianism, it is ethically acceptable to expose an individual to justifiable risk provided the greater good is served. However, there is a moral obligation on the part of the researcher to mitigate risk by designing protocols that minimize risk or provide backup that reduces the consequences of an adverse event if it were to occur. For example, before conducting a phase I study investigating a new drug in healthy volunteers it is imperative that extensive testing in several different animal models has shown this investigational agent to be relatively safe. Furthermore, clinical trials should be supervised by appropriately trained and qualified medical personnel that are able to identify and treat any adverse reaction to a trial medication. The research should be conducted in a facility that has the appropriate equipment available to treat any medical complication of the research. Data-monitoring committees, independent of the research sponsor, should monitor trial data during the conduct of the trial to ensure the welfare of trial participants. These committees significantly enhance the safety of clinical research as they enable early termination of unsafe trials. Finally, it is a legal and moral imperative that all research participants are covered by insurance that will provide free medical care and compensation if they are inadvertently injured as a result of their participation in the research.

### 8.4 S: Scientific Integrity

Researchers have an ethical obligation to objectively analyze and report data. Improper data analysis or biased reporting of results in the scientific literature is unethical as it could lead to the adoption of medical treatments as standard proven therapy when in fact these treatments may be at best useless or worse even harmful. Careful analysis and recording of data, sharing of data and resources, and respect of research colleagues are all important components of ethical research conduct.

## 9 CONCLUSION

Our understanding of what constitutes ethical conduct in clinical research has come a long way since the advent of the Nuremberg Code in 1949. Today, the clinical researcher has several excellent guidelines [5, 11, 12] to help with the design and conduct of clinical studies that maximize participant safety and autonomy while still allowing appropriate scientific experimentation. This chapter has outlined the ethical principles used in the development of these clinical research guidelines. It is hoped that with an in-depth knowledge of these principles, researchers will be in a better position to design and conduct scientifically and ethically appropriate clinical trials. We hope that this chapter will facilitate a better understanding of ethical principles and the importance of ethics to clinical research practice.

## APPENDIX A NUREMBERG CODE*

1. The voluntary consent of the human subject is absolutely essential. This means that the person involved should have legal capacity to give consent; should be so situated as to be able to exercise free power of choice, without the intervention of any element of force, fraud, deceit, duress, over-reaching, or other ulterior form of constraint or coercion; and should have sufficient knowledge and comprehension of the elements of the subject matter involved as to enable him to make an understanding and enlightened decision. This latter element requires that before the acceptance of an affirmative decision by the experimental subject there should be made known to him the nature, duration, and purpose of the experiment; the method and means by which it is to be conducted; all inconveniences and hazards reasonable to be expected; and the effects upon his health or person which may possibly come from his participation in the experiment.

   The duty and responsibility for ascertaining the quality of the consent rests upon each individual who initiates, directs or engages in the experiment. It is a personal duty and responsibility which may not be delegated to another with impunity.
2. The experiment should be such as to yield fruitful results for the good of society, unprocurable by other methods or means of study, and not random and unnecessary in nature.
3. The experiment should be so designed and based on the results of animal experimentation and a knowledge of the natural history of the disease or other problem under study that the anticipated results will justify the performance of the experiment.
4. The experiment should be so conducted as to avoid all unnecessary physical and mental suffering and injury.
5. No experiment should be conducted where there is an a priori reason to believe that death or disabling injury will occur; except, perhaps, in those experiments where the experimental physicians also serve as subjects.
6. The degree of risk to be taken should never exceed that determined by the humanitarian importance of the problem to be solved by the experiment.
7. Proper preparations should be made and adequate facilities provided to protect the experimental subject against even remote possibilities of injury, disability, or death.
8. The experiment should be conducted only by scientifically qualified persons. The highest degree of skill and care should be required through all stages of the experiment of those who conduct or engage in the experiment.
9. During the course of the experiment the human subject should be at liberty to bring the experiment to an end if he has reached the physical or mental state where continuation of the experiment seems to him to be impossible.

*Reprinted from *Trials of War Criminals before the Nuremberg Military Tribunals under Control Council Law No. 10, Vol. 2, pp. 181–182*. Washington, DC: U.S. Government Printing Office, 1949.

10. During the course of the experiment the scientist in charge must be prepared to terminate the experiment at any stage, if he has probable cause to believe, in the exercise of the good faith, superior skill and careful judgment required of him that a continuation of the experiment is likely to result in injury, disability, or death to the experimental subject.

## APPENDIX B WORLD MEDICAL ASSOCIATION DECLARATION OF HELSINKI: ETHICAL PRINCIPLES FOR MEDICAL RESEARCH INVOLVING HUMAN SUBJECTS*

Adopted by the 18th WMA General Assembly, Helsinki, Finland, June 1964, and amended by the

29th WMA General Assembly, Tokyo, Japan, October 1975
35th WMA General Assembly, Venice, Italy, October 1983
41st WMA General Assembly, Hong Kong, September 1989
48th WMA General Assembly, Somerset West, Republic of South Africa, October 1996 and the 52nd WMAN General Assembly, Edinburgh, Scotland, October 2000

Note of Clarification on Paragraph 29 added by the WMA General Assembly, Washington 2002.

Note of Clarification on Paragraph 30 added by the WMA General Assembly, Tokyo 2004.

### A. Introduction

1. The World Medical Association has developed the Declaration of Helsinki as a statement of ethical principles to provide guidance to physicians and other participants in medical research involving human subjects. Medical research involving human subjects includes research on identifiable human material or identifiable data.

**Clarification of paragraph 29*: The WMA hereby reaffirms its position that extreme care must be taken in making use of a placebo-controlled trial and that in general this methodology should only be used in the absence of existing proven therapy. However, a placebo-controlled trial may be ethically acceptable, even if proven therapy is available, under the following circumstances:

- where for compelling and scientifically sound methodological reasons its use is necessary to determine the efficacy or safety of a prophylactic, diagnostic, or therapeutic method or
- where a prophylactic, diagnostic, or therapeutic method is being investigated for a minor condition and the patients who receive placebo will not be subject to any additional risk of serious or irreversible harm.

All other provisions of the Declaration of Helsinki must be adhered to, especially the need for appropriate ethical and scientific review.

*Clarification of paragraph 30*: The WMA hereby reaffirms its position that it is necessary during the study-planning process to identify posttrial access by study participants to prophylactic, diagnostic, and therapeutic procedures identified as beneficial in the study or access to other appropriate care. Posttrial access arrangements or other care must be described in the study protocol so the ethical review committee may consider such arrangements during its review.

2. It is the duty of the physician to promote and safeguard the health of the people. The physician's knowledge and conscience are dedicated to the fulfillment of this duty.
3. The Declaration of Geneva of the World Medical Association binds the physician with the words, "The health of my patient will be my first consideration," and the International Code of Medical Ethics declares that, "A physician shall act only in the patient's interest when providing medical care which might have the effect of weakening the physical and mental condition of the patient."
4. Medical progress is based on research which ultimately must rest in part on experimentation involving human subjects.
5. In medical research on human subjects, considerations related to the well-being of the human subject should take precedence over the interests of science and society.
6. The primary purpose of medical research involving human subjects is to improve prophylactic, diagnostic and therapeutic procedures and the understanding of the aetiology and pathogenesis of disease. Even the best proven prophylactic, diagnostic, and therapeutic methods must continuously be challenged through research for their effectiveness, efficiency, accessibility and quality.
7. In current medical practice and in medical research, most prophylactic, diagnostic and therapeutic procedures involve risks and burdens.
8. Medical research is subject to ethical standards that promote respect for all human beings and protect their health and rights. Some research populations are vulnerable and need special protection. The particular needs of the economically and medically disadvantaged must be recognized. Special attention is also required for those who cannot give or refuse consent for themselves, for those who may be subject to giving consent under duress, for those who will not benefit personally from the research and for those for whom the research is combined with care.
9. Research Investigators should be aware of the ethical, legal and regulatory requirements for research on human subjects in their own countries as well as applicable international requirements. No national ethical, legal or regulatory requirement should be allowed to reduce or eliminate any of the protections for human subjects set forth in this Declaration.

### B. Basic Principles for All Medical Research

10. It is the duty of the physician in medical research to protect the life, health, privacy, and dignity of the human subject.
11. Medical research involving human subjects must conform to generally accepted scientific principles, be based on a thorough knowledge of the scientific literature, other relevant sources of information, and on adequate laboratory and, where appropriate, animal experimentation.
12. Appropriate caution must be exercised in the conduct of research which may affect the environment, and the welfare of animals used for research must be respected.

13. The design and performance of each experimental procedure involving human subjects should be clearly formulated in an experimental protocol. This protocol should be submitted for consideration, comment, guidance, and where appropriate, approval to a specially appointed ethical review committee, which must be independent of the investigator, the sponsor or any other kind of undue influence. This independent committee should be in conformity with the laws and regulations of the country in which the research experiment is performed. The committee has the right to monitor ongoing trials. The researcher has the obligation to provide monitoring information to the committee, especially any serious adverse events. The researcher should also submit to the committee, for review, information regarding funding, sponsors, institutional affiliations, other potential conflicts of interest and incentives for subjects.
14. The research protocol should always contain a statement of the ethical considerations involved and should indicate that there is compliance with the principles enunciated in this Declaration.
15. Medical research involving human subjects should be conducted only by scientifically qualified persons and under the supervision of a clinically competent medical person. The responsibility for the human subject must always rest with a medically qualified person and never rest on the subject of the research, even though the subject has given consent.
16. Every medical research project involving human subjects should be preceded by careful assessment of predictable risk and burdens in comparison with foreseeable benefits to the subject or to others. This does not preclude the participation of healthy volunteers in medical research. The design of all studies should be publicly available.
17. Physicians should abstain from engaging in research projects involving human subjects unless they are confident that the risks involved have been adequately assessed and can be satisfactorily managed. Physicians should cease any investigation if the risks are found to outweigh the potential benefits or if there is conclusive proof of positive and beneficial results.
18. Medical research involving human subjects should only be conducted if the importance of the objective outweighs the inherent risks and burdens to the subject. This is especially important when the human subjects are healthy volunteers.
19. Medical research is only justified if there is a reasonable likelihood that the populations in which the research is carried out stand to benefit from the results of the research.
20. The subjects must be volunteers and informed participants in the research project.
21. The right of research subjects to safeguard their integrity must always be respected. Every precaution should be taken to respect the privacy of the subject, the confidentiality of the patient's information and to minimize the impact of the study on the subject's physical and mental integrity and on the personality of the subject.
22. In any research on human beings, each potential subject must be adequately informed of the aims, methods, sources of funding, any possible conflicts of interest, institutional affiliations of the researcher, the anticipated benefits and potential risks of the study and the discomfort it may entail. The subject should

be informed of the right to abstain from participation in the study or to withdraw consent to participate at any time without reprisal. After ensuring that the subject has understood the information, the physician should then obtain the subject's freely-given informed consent, preferably in writing. If the consent cannot be obtained in writing, the non-written consent must be formally documented and witnessed.

23. When obtaining informed consent for the research project the physician should be particularly cautious if the subject is in a dependent relationship with the physician or may consent under duress. In that case the informed consent should be obtained by a well-informed physician who is not engaged in the investigation and who is completely independent of this relationship.
24. For a research subject who is legally incompetent, physically or mentally incapable of giving consent or is a legally incompetent minor, the investigator must obtain informed consent from the legally authorized representative in accordance with applicable law. These groups should not be included in research unless the research is necessary to promote the health of the population represented and this research cannot instead be performed on legally competent persons.
25. When a subject deemed legally incompetent, such as a minor child, is able to give assent to decisions about participation in research, the investigator must obtain that assent in addition to the consent of the legally authorized representative.
26. Research on individuals from whom it is not possible to obtain consent, including proxy or advance consent, should be done only if the physical/mental condition that prevents obtaining informed consent is a necessary characteristic of the research population. The specific reasons for involving research subjects with a condition that renders them unable to give informed consent should be stated in the experimental protocol for consideration and approval of the review committee. The protocol should state that consent to remain in the research should be obtained as soon as possible from the individuals or a legally authorized surrogate.
27. Both authors and publishers have ethical obligations. In publication of the results of research, the investigators are obliged to preserve the accuracy of the results. Negative as well as positive results should be published or otherwise publicly available. Sources of funding, institutional affiliations and any possible conflicts of interest should be declared in the publication. Reports of experimentation not in accordance with the principles laid down in this Declaration should not be accepted for publication.

### C. Additional Principles for Medical Research Combined with Medical Care

28. The physician may combine medical research with medical care, only to the extent that the research is justified by its potential prophylactic, diagnostic or therapeutic value. When medical research is combined with medical care, additional standards apply to protect the patients who are research subjects.
29. The benefits, risks, burdens and effectiveness of a new method should be tested against those of the best current prophylactic, diagnostic, and therapeutic

methods. This does not exclude the use of placebo, or no treatment, in studies where no proven prophylactic, diagnostic or therapeutic method exists.

30. At the conclusion of the study, every patient entered into the study should be assured of access to the best proven prophylactic, diagnostic and therapeutic methods identified by the study.
31. The physician should fully inform the patient which aspects of the care are related to the research. The refusal of a patient to participate in a study must never interfere with the patient-physician relationship.
32. In the treatment of a patient, where proven prophylactic, diagnostic and therapeutic methods do not exist or have been ineffective, the physician, with informed consent from the patient, must be free to use unproven or new prophylactic, diagnostic and therapeutic measures, if in the physician's judgement it offers hope of saving life, re-establishing health or alleviating suffering. Where possible, these measures should be made the object of research, designed to evaluate their safety and efficacy. In all cases, new information should be recorded and, where appropriate, published. The other relevant guidelines of this Declaration should be followed.

## REFERENCES


1. Shamoo, A., and Resnik, D. (2003), *Responsible Conduct of Research*, Oxford University Press, Oxford.
2. Loue, S. (2000), *Textbook of Research Ethics. Theory and Practice*, Kluwer Academic/Plenum, New York.
3. Kant, I. (1998), *Groundwork of the Metaphysics of Morals*, Cambridge University Press, Cambridge.
4. Shaw, W. (1998), *Contemporary Ethics. Taking Account of Utilitarianism*, Blackwell, London.
5. World Medical Association Declaration of Helsinki (2000), Ethical Principals for Medical Research Involving Human Subjects; available at: http://www.wma.net/e/policy/b3.htm.
6. The Nuremberg Code (1949), *International Principals for Human Experimentation, from the Trials of War Criminals before the Nuremberg Military Tribunals under the Control Council Law No. 10*, Vol. 2, Government Printing Office, Washington D.C., pp. 181–182.
7. Beauchamp, T., and Childress, J. (1994), *Principals of Biomedical Ethics*, Oxford University Press, New York.
8. Vonderlehr, R., Clark, T., Wenger, O., et al. (1936), Untreated syphilis in the male Negro: A comparative study of treated and untreated cases, *JAMA*, 107, 856–860.
9. Jones, J. H. (1993), *Bad Blood: The Tuskegee Syphilis Experiment*, Free Press, New York.
10. Roelcke, V. (2004), Nazi medicine and research on human beings, *Lancet*, 364(Suppl.), 6–7.
11. The Belmont Report (1979), *Ethical Principals and Guidelines for the Protection of Human Subjects of Research*, Government Printing Office, Washington, D.C.
12. Council for International Organizations of Medical Sciences (CIOMS) (2002), *International Ethical Guidelines for Biomedical Research Involving Human Subjects*, World Health Organization, Geneva.

13. Office of Human Research Protections, U.S. Department of Heath and Human Services; available at: http://www.hhs.gov/ohrp/international/HSPCompilation.pdf, accessed December 2007.
14. CONSORT Statement; available at: http://www.consort-statement.org/?o=1399, accessed December 2007.
15. Zarin, D. A., Ide, N. C., Tse, T., et al. (2007), Issues in the registration of clinical trials, *JAMA*, 297(19), 2112–2120.
16. OECD Principles of Good Laboratory Practice (GLP), available at: http://www.oecd.org, accessed December 2007.
17. Good Laboratory Practice (GLP) regulations; available at: http://www.cfsan.fda.gov/~dms/opa-pt58.html; and http://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfcfr/CFRSearch.cfm?CFRPart=58, accessed December 2007.
18. Ashcroft, R. E., Chadwick, D. W., Clark, S. R., et al. (1997), Implications of socio-cultural contexts for the ethics of clinical trials, *Health Technol. Assess*, 1(9), i–iv, 1–65.
19. ICH (2000), *ICH Harmonised Tripartite Guideline: Choice of Control Group and Related Issues in Clinical Trials E10*, Federal Register, vol. 66, no. 93, 24390–24391. May 14, 2001
20. Simon, S. D. (2001), Is the randomized clinical trial the gold standard of research? *J. Androl.*, 22(6), 938–943.
21. Edwards, S. J., Lilford, R. J., Braunholtz, D. A., et al. (1998), Ethical issues in the design and conduct of randomised controlled trials. *Health Technol. Assess*, 2(15), i–vi, 1–132.
22. Fries, J. F., and Krishnan, E. (2004), Equipoise, design bias, and randomized controlled trials: The elusive ethics of new drug development, *Arthritis Res. Ther.*, 6(3), R250–255.
23. Lilford, R. J. (2003), Ethics of clinical trials from a bayesian and decision analytic perspective: Whose equipoise is it anyway? *BMJ*, 326(7396), 980–981.
24. Freedman, B. (1987), Equipoise and the ethics of clinical research, *N. Engl. J. Med.*, 317(3), 141–145.
25. Djulbegovic, B., Lacevic, M., Cantor, A., et al. (2000), The uncertainty principle and industry-sponsored research, *Lancet*, 356(9230), 635–638.
26. Johnson, N., Lilford, R. J., and Brazier, W. (1991), At what level of collective equipoise does a clinical trial become ethical? *J. Med. Ethics*, 17(1), 30–34.
27. Olson, L. G. (2002), Patient-centred equipoise and the ethics of randomised controlled trials, *Monash Bioeth. Rev.*, 21(2), S55–67.
28. Sackett, D. L. (2000), Equipoise, a term whose time (if it ever came) has surely gone, *CMAJ*, 163(7), 835–836.
29. Ashcroft, R. (1999), Equipoise, knowledge and ethics in clinical research and practice, *Bioethics*, 13(3–4), 314–326.
30. Miller, F. G., and Brody, H. (2007), Clinical equipoise and the incoherence of research ethics, *J. Med. Philos.*, 32(2), 151–165.
31. Braunholtz, D. A., Edwards, S. J., and Lilford, R. J. (2001), Are randomized clinical trials good for us (in the short term)? Evidence for a "trial effect," *J. Clin. Epidemiol.*, 54(3), 217–224.
32. Walach, H., Sadaghiani, C., Dehm, C., et al. (2005), The therapeutic effect of clinical trials: Understanding placebo response rates in clinical trials—a secondary analysis, *BMC Med. Res. Methodol.*, 5, 26.
33. Halpern, S. D., Karlawish, J. H., and Berlin, J. A. (2002), The continuing unethical conduct of underpowered clinical trials, *JAMA*, 288(3), 358–362.
34. Miller, F. G., and Brody, H. (2002), What makes placebo-controlled trials unethical? *Am. J. Bioeth.*, 2(2), 3–9.

35. FDA (1998), FDA Guidance for Institutional Review Boards and Clinical Investigators; available at: http://www.fda.gov/oc/ohrt/irbs/drugsbiologics.html#study.
36. EMEA/CPMP (2001), *Position Statement on the Use of Placebo in Clinical Trials with Regard to the Revised Declaration of Helsinki*, EMEA Workshop on Ethical Considerations in Clinical Trials, London.
37. ICH (1996), *ICH Harmonised Tripartite Guideline: Guideline for Good Clinical Practice E6*, Federal Register, Vol. 62, no. 90, 25691–25709. May 9, 1997.
38. 21 CFR part 50, Protection of Human Subjects; available at: www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfcfr/CFRsearch.cfm?CFRPart=50.
39. Guidance Document: Financial Relationships and Interests in Research Involving Human Subjects: Guidance for Human Subject Protection; available at: http://www.hhs.gov/ohrp/humansubjects/finreltn/fguid.pdf, accessed December 2007.
40. McKneally, M. F., and Martin, D. K. (2000), An entrustment model of consent for surgical treatment of life-threatening illness: Perspective of patients requiring esophagectomy, *J. Thorac. Cardiovasc. Surg.*, 120(2), 264–269.
41. Robinson, E. J., Kerr, C. E., Stevens, A. J., et al. (2005), Lay public's understanding of equipoise and randomisation in randomised controlled trials, *Health Technol. Assess.*, 9(8), 1–192, iii–iv.
42. Bernstein, M. (2005), Fully informed consent is impossible in surgical clinical trials, *Can. J. Surg.*, 48(4), 271–272.
43. De Vries, R., and Elliott, C. (2006), Why disclosure? *J. Gen. Intern. Med.*, 21(9), 1003–1004.
44. Weinfurt, K. P., Friedman, J. Y., Dinan, M. A., et al. (2006), Disclosing conflicts of interest in clinical research: Views of institutional review boards, conflict of interest committees, and investigators, *J. Law Med. Ethics*, 34(3), 581–591, 481.
45. Dickersin, K. (1990), The existence of publication bias and risk factors for its occurrence, *JAMA*, 263(10), 1385–1389.
46. Song, F., Eastwood, A. J., Gilbody, S., et al. (2000), Publication and related biases, *Health Technol. Assess.*, 4(10), 1–115.
47. Chan, A. W., Hrobjartsson, A., Haahr, M. T., et al. (2004), Empirical evidence for selective reporting of outcomes in randomized trials: Comparison of protocols to published articles, *JAMA*, 291(20), 2457–2465.
48. Moher, D., Schulz, K. F., and Altman, D. G. (2001), The CONSORT statement: Revised recommendations for improving the quality of reports of parallel-group randomized trials, *Ann. Intern. Med.*, 134(8), 657–662.
49. Sismondo, S. (2007), Ghost management: How much of the medical literature is shaped behind the scenes by the pharmaceutical industry? *PLoS Med.*, 4(9), e286.
50. http://ottawagroup.ohri.ca/.
51. Krleza-Jeric, K. (2005), Clinical trial registration: The differing views of industry, the WHO, and the Ottawa Group, *PLoS Med.*, 2(11), e378.
52. The WHO International Clinical Trials Registry Platform; available at: http://www.who.int/ictrp/en/.
53. Clinical trial registration: A statement from the International Committee of Medical Journal Editors; available at: http://www.icmje.org/clin_trial.pdf.
54. Laine, C., Horton, R., DeAngelis, C. D., et al. (2007), Clinical trial registration: Looking back and moving ahead, *JAMA*, 298(1), 93–94.
55. http://www.who.int/ictrp/results/en/.

56. International Committee of Medical Journal Editors: Uniform Requirements for Manuscripts Submitted to Biomedical Journals: Writing and Editing for Biomedical Publication; available at: http://www.icmje.org/.
57. Good Publication Practice Guidelines for Pharmaceutical Companies; available at: http://www.gpp-guidelines.org/.
58. Wager, E., Field, E. A., and Grossman, L. (2003), Good publication practice for pharmaceutical companies, *Curr. Med. Res. Opin.*, 19(3), 149–154.
59. Woolley, K. L. (2006), Goodbye ghostwriters!: How to work ethically and efficiently with professional medical writers, *Chest*, 130(3), 921–923.
60. Consort: Transparent Reporting of Clinical Trials; available at: http://www.consort-statement.org/.
61. Begg, C., Cho, M., Eastwood, S., et al. (1996), Improving the quality of reporting of randomized controlled trials. The CONSORT statement, *JAMA*, 276(8), 637–639.
62. Campbell, M. K., Elbourne, D. R., and Altman, D. G. (2004), CONSORT statement: Extension to cluster randomised trials, *BMJ*, 328(7441), 702–708.
63. Piaggio, G., Elbourne, D. R., Altman, D. G., et al. (2006), Reporting of noninferiority and equivalence randomized trials: An extension of the CONSORT statement, *JAMA*, 295(10), 1152–1160.
64. McKinney, R. E. (2003), Congress, the FDA, and the fair development of new medications for children, *Pediatrics*, 112, 669–670.
65. Raymond, A. S., and Brasseur, D. (2005), Development of medicines for children in Europe: Ethical implications, *Paediatric Respir. Rev.*, 6, 45–51.
66. Knox, C. A., and Burkhart, P. V. (2007), Issues related to children participating in clinical research, *J. Pediatric Nurs.*, 22(4), 310–318.
67. Burns, J. P. (2003), Research in children, *Crit. Care Med.*, 31(3), S131–136.
68. Ramsey, P. (1976), The enforcement of morals: Nontherapeutic research on children, *Hastings Center Rep.*, 6, 21–30.
69. American Academy of Pediatrics, Committee on Drugs (1995), Guidelines for the ethical conduct of studies to evaluate drugs in pediatric populations, *Pediatrics*, 95, 286–294.
70. Michels, R. (1999), Are research ethics bad for our mental health? *N. Engl. J. Med.*, 340, 1427–1430.
71. NBAC (1998), *Research Involving Persons with Mental Disorders That May Affect Decision Making Capacity*, National Bioethics Advisory Commission, Rockville, MD.
72. Research involving individuals with questionable capacity to consent: Points to consider, U.S. Departments of Health and Human Services; available at: http://grants1.nih.gov/grants/policy/questionalblecapacity.htm.
73. American College of Obstetricians and Gynecologists (2007), Research involving women, *Obstet. Gynecol.*, 110, 731–736.
74. McCullough, L. B., Coverdale, J. H., and Chervenak, F. A. (2005), A comprehensive ethical framework for responsibly designing and conducting pharmacologic research that involves pregnant women, *Am. J. Obstet. Gynecol.*, 193, 901–907.
75. McCullough, L. B., Coverdale, J. H., and Chervenak, F. A. (2006), Preventative ethics for including women of childbearing potential in clinical trials, *Am. J. Obstet. Gynecol.*, 194, 1221–1227.
76. Bigatello, L. M., George, E., and Hurford, W. E. (2003), Ethical considerations for research in critically ill patients, *Crit. Care Med.*, 31(Suppl 3), s178–181.
77. Coppolino, M., and Ackerson, L. (2001), So surrogate decision makers provide accurate consent for intensive care research? *Chest*, 119, 603–612.

78. American Thoracic Society (2004), The ethical conduct of clinical research involving critically ill patients in the United States and Canada, *Am. J. Respir. Crit. Care Med.*, 170, 1375–1384.
79. American Heart Association (2006), Recommendations for implementation of community consultation and public disclosure under the FDA "exception from informed consent requirements from informed consent"; available at: http://www.americanheart.org/emergencyexception.
80. Lackey, D. P. (2001), Clinical trials in developing countries: A review of the moral issues, *M. Sinai J. Med.*, 68, 4–12.
81. Wendler, D., Emanuel, E., and Lie, R. K. (2004), The standard of care debate: Can research in developing countries be both ethical and responsive to those countries needs? *Am. J. Public Hlth.*, 94, 923–928.
82. Varmus, H., and Satcher, D. (1997), Ethical complexities of conducting research in developing countries, *New Engl. J. Med.*, 337, 1003–1005.
83. Phanuphak, P. (1997), Ethical issues in studies in Thailand of the vertical transmission of HIV, *New Engl. J. Med.*, 338, 834–835.

# Regulations

**Ramzi Dagher,[1] Rajeshwari Sridhara,[2] Nallaperumal Chidambaram,[3] and Brian P. Booth[4]**

[1]*Pfizer, Inc., New London, Connecticut*
[2]*Office of Translational Science, Office of Biostatistics, Division of Biometrics, Food and Drug Administration, Rockville, Maryland*
[3]*Office of New Drug Quality Assessment, Division of Post-Marketing Evaluation, Food and Drug Administration, Rockville, Maryland*
[4]*Office of Translational Science, Office of Clinical Pharmacology, Division of Clinical Pharmacology, Food and Drug Administration, Rockville, Maryland*

**Contents**

1 Introduction
2 Regulatory History
3 Clinical Investigations
4 NDA Review Process
5 Pediatric Initiatives
6 Role of Statistics in Regulatory Setting
7 Role of Chemistry, Manufacturing, and Controls (CMC) Information in Regulatory Setting
8 Role of Clinical Pharmacology in Regulatory Setting
  8.1 Biopharmaceutics Guidances
  8.2 Clinical Pharmacology Guidances
  8.3 Timins of Studies
9 Drug Regulation in a Global Environment
Appendix: Clinical Pharmacology Drug Development Questions
References

The views expressed are those of the authors and do not reflect official policy of the FDA. No official endorsement by the FDA is intended or should be inferred.

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

## 1 INTRODUCTION

The U.S. Food and Drug Administration (USFDA) is a scientific, regulatory, and public health agency that oversees what is estimated to be 25% of the consumer economy. The agency has jurisdiction over most food products, human and animal drugs, therapeutic biologic agents, medical devices, radiation-emitting products, cosmetics, and animal feed [1]. The FDA has a staff of approximately 10,000 employees including chemists, pharmacologists, toxicologists, physicians, microbiologists, veterinarians, pharmacists, lawyers, statisticians, project managers, and others. In this chapter, the role of the FDA and different scientific disciplines in the marketing approval of drug products in the United States are presented.

## 2 REGULATORY HISTORY

The FDA's mandate has evolved based on a number of legislative acts that have been introduced over the last 100 years. These are discussed below and outlined in Table 1. Regulatory functions were incorporated into the agency's mission in 1906 with passage of the Federal Food and Drugs Act. The 1906 law forbade interstate and foreign commerce in adulterated and misbranded foods and drugs. Drugs had to abide by standards of purity and quality set forth in the *U.S. Pharmacopeia* and the *National Formulary*, works prepared by committees of physicians and pharmacists, or meet individual standards chosen by their manufacturers and stated on their labels. The law prohibited the adulteration of foods by the removal of valuable constituents, the substitution of ingredients so as to reduce quality, the addition of deleterious ingredients, or the use of spoiled animal and vegetable products.

Another important milestone in the agency's history took place in 1938 with passage of the Federal Food, Drug, and Cosmetic (FDC) Act, which, for the first time, required new drugs to be shown safe before marketing. The FDC Act also extended the FDA's jurisdiction to cosmetics and therapeutic devices, authorized factory inspections, and added the remedy of court injunctions to the previous penalties of seizures and prosecutions.

In the late 1950s and early 1960s, a drug used in Europe for its sedative effects (thalidomide) was found to have caused birth defects in thousands of infants. The drug had been kept off the market in the United States. This event, as well as other factors, aroused public support for stronger drug regulation. In 1962, the

**TABLE 1 FDA Regulatory Landmarks**

| Year | Legislation | Features |
|---|---|---|
| 1906 | Food and Drugs Act | Defined FDA regulatory functions |
| 1938 | Food, Drug and Cosmetics Act | Requirement for safety data |
| 1962 | Kefauver–Harris Drug Amendments | Required demonstration of safety and effectiveness prior to marketing |
| 1992 | Subpart H | Accelerated approval of drugs for serious or life-threatening diseases |
| 1992 | Prescription Drug User Fee Act | Fees for product applications |
| 1997 | Food and Drug Modernization Act | Several reforms |

Kefauver–Harris Drug Amendments passed. In addition to more rigorous requirements for the demonstration of drug safety, this legislation required manufacturers to demonstrate to the FDA the effectiveness of their products prior to marketing.

In the 1980s and early 1990s, there was a growing recognition for the need to encourage the timely development and approval of drugs for serious or life-threatening illnesses such as HIV/AIDS (the human immunodeficiency virus/acquired immunodeficiency syndrome) and cancer. In 1992, subpart H was added to the new drug approval (NDA) regulations, allowing for accelerated approval (AA) of drugs for serious or life-threatening diseases where a drug demonstrates an advantage over available therapy or demonstrates its effects in a setting where no available therapy exists. Approval is based on a surrogate endpoint "reasonably likely to predict clinical benefit." The sponsor must study the drug further to demonstrate clinical benefit in subsequent or ongoing clinical trials [2, 3].

Also in 1992, the Prescription Drug User Fee Act (PDUFA) was enacted, requiring drug and biologic drug manufacturers to pay fees for product applications and supplements. The act also required the FDA to use these funds to increase the number and breadth of reviewer expertise for product applications. In 1997, the Food and Drug Modernization Act reauthorized PDUFA and mandated several reforms in agency practices [4]. Measures were introduced to accelerate review of devices, to regulate advertising of unapproved uses of approved products, and to regulate health claims for foods.

## 3 CLINICAL INVESTIGATIONS

In the United States, investigational drugs must be administered under an investigational new drug (IND) application submitted to the FDA. The regulations describe two parties involved in IND submission, the *sponsor*, who is responsible for reporting to the FDA, and the *investigator*, who performs the trial [5]. The sponsor may be a pharmaceutical company, an academic institution, or an individual. Sponsors are to select only investigators "qualified by training and experience as appropriate experts to investigate the drug."

When an IND is submitted to the FDA, a team of scientific reviewers evaluate the safety data from animals or other sources, evaluate the proposed study, and determine whether patients would be exposed to an unreasonable and significant risk. These issues are discussed individually in the sections below. All studies of nonapproved drugs must be done under an IND. For approved drugs, however, some studies require an IND and others are exempt from the IND requirement.

To determine that an IND is not needed, the following conditions should be met: The study (1) is not intended to support approval of a new indication or a significant change in the product labeling, (2) is not intended to support a significant change in advertising, (3) does not involve a route of administration or dosage level or use in a patient population or other factor that significantly increases the risks associated with the use of the drug product, (4) is conducted in compliance with institutional review board (IRB) and informed consent regulations, and (5) will not be used to promote unapproved indications [5].

The IND process spans the entire period of drug investigation (Fig. 1). It includes the initial IND application and later amendments to provide safety reports or

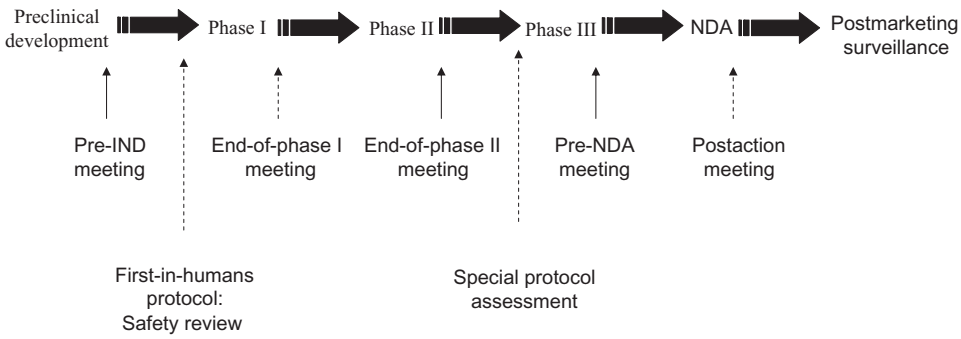


***FIGURE 1*** *Drug development timeline. The solid arrows indicate standard time points for meeting with drug developers. The dashed arrows indicate additional time points when the FDA sometimes meets with sponsors.*

additional protocols. The initial IND application usually consists of a phase I clinical protocol and data to support the safety of the proposal. The latter would include in vitro, animal, and/or human evidence describing drug toxicity and allowing prediction of a safe starting dose. Manufacturing data describing the composition, manufacture, and control of the drug substance and drug product is also included. After the FDA receives the initial IND, sponsors are required to wait 30 days prior to initiating the proposed study unless they request and receive a waiver of the 30-day review period from the FDA. A multidisciplinary team of FDA reviewers, including physicians, pharmacologist/toxicologists, chemists, statisticians, and clinical pharmacologists determine whether the study is safe to proceed. The FDA may put an IND "on hold" if the agency believes subjects would be exposed to unreasonable risk of injury or if there is insufficient information to assess the risks.

The FDA frequently meets with sponsors and investigators in pre-IND meetings to review proposed IND plans and to clarify IND requirements. Animal studies should use the same or a similar schedule of drug administration proposed for the phase I clinical study. The starting dose for investigational drugs used in human studies is usually based on a preclinical animal model, which may differ based on the type of product and the therapeutic area of development. FDA advice on the specific requirements for a given product should be sought from representatives of the responsible FDA review division.

A critical FDA role in drug development is to meet with sponsors to provide advice on the design and conduct of phase III (and sometimes phase II) clinical trials that will support marketing applications. The multidisciplinary FDA team attending these meetings includes physicians, statisticians, pharmacologists/toxicologists, clinical pharmacologists, and often external expert consultants and patient advocates. FDA chemists also meet with sponsors, often separately, to discuss manufacturing and quality control issues. Sponsors can submit protocols subsequent to these meetings and request a Special Protocol Assessment, which provides for a binding agreement on protocol design [6]. After the clinical trials have been conducted and trial results are available, sponsors again meet with the FDA in pre-NDA meetings to discuss whether a new drug application (NDA) may be warranted and, if so, to discuss details of an NDA submission.

After clinical trials have been completed and an NDA has been submitted, the FDA verifies data quality and judges whether trial results demonstrate that the drug is safe and effective for the proposed use. After approval, the FDA continues to evaluate drug safety and regulate drug marketing.

## 4 NDA REVIEW PROCESS

The package insert describes clinical trial results from data that have been reviewed and validated by FDA review teams. Regulations require that NDAs contain all relevant information about manufacturing, preclinical pharmacology and toxicology, human pharmacokinetics and bioavailability, clinical data, and statistical analyses. NDA applicants must submit financial disclosure information about the investigators. The FDA review of the NDA involves a team of chemists, pharmacologists/toxicologists, clinical pharmacologists, physicians, statisticians, microbiologists, site inspectors, and a project manager. The FDA reviewers evaluate the primary data, verify analyses, and where appropriate perform additional analyses. The FDA field inspectors verify that information on case report forms is supported by source data such as hospital charts and that biological drug concentrations are documented. Inspectors also inspect manufacturing facilities for suitability.

Based on the PDUFA, the FDA performs NDA reviews with either a 6-month or a 10-month goal. Applications representing a significant improvement compared to products that are currently available are assigned priority status and a 6-month review period, whereas standard applications have a 10-month review goal.

The FDA routinely seeks external advice on the design, analysis, and interpretation of clinical trials. These consultants are screened to exclude any potential conflicts of interest. Individual consultants advise the FDA during the design of clinical trials and during the NDA review. After an initial NDA review, the FDA presents selected NDAs to the appropriate advisory committee. This group is composed of medical and/or surgical specialists, statisticians, patient advocates, consumer representatives, and a nonvoting industry representative. At the public meetings of an advisory committee, the NDA applicant summarizes the study results, and the FDA presents its review of the findings. Members of the public may comment on the issues being discussed during the public speaker portion of the meeting. The committee finally votes on the specific questions submitted by the FDA. Although the discussion and the votes by the committee are taken into consideration, the FDA is not obligated to adhere to the advice provided.

## 5 PEDIATRIC INITIATIVES

There has been general recognition of the lack of adequate labeling of drug products for children. In order to encourage evaluation of new therapies in pediatric populations and to optimize submission of pediatric data to support labeling of commercially available drugs, the FDA has undertaken two initiatives, which are outlined in Table 2 and summarized as follows.

A voluntary program, commonly referred to as "pediatric exclusivity," was outlined in the Food and Drug Modernization Act of 1997. Under this program,

**TABLE 2 FDA Pediatric Initiatives**

| PREA | BPCA |
|---|---|
| Applies to drugs and biologics | Applies only to drugs |
| Studies mandatory | Studies voluntary |
| Required only for drug/indication under review | Studies on entire active moiety |
| Orphan indications exempted | Includes orphan drugs |

a commercial sponsor may receive a 6-month extension of existing marketing exclusivity by submitting data from pediatric clinical trials in support of an NDA. The general design of studies to be conducted and data to be submitted are outlined in a pediatric written request letter issued by the FDA to the sponsor. This program applies to drug products, including those with orphan designation. However, biological therapies such as monoclonal antibodies and vaccines are excluded. The Best Pharmaceuticals for Children Act (BPCA) subsequently provided some refinements to the program, including a framework for making summaries of medical and pharmacology reviews available to the public [7].

A mandatory program applying to indications where the disease in pediatric patients is similar to the adult disease, previously known as the "pediatric rule," was subsequently formally legislated in the Pediatric Research Equity Act (PREA) [8]. This mandatory requirement for sponsors to submit data from pediatric studies when an adult indication is granted applies to a specific drug and indication under review. In order to avoid the potential delay in drug development and in providing access to therapies for life-threatening diseases that can arise when the adult data support approval and pediatric studies are still ongoing, a deferral for submission of the pediatric data may be granted. Both drug and biologic products are subject to the PREA. However, products with orphan designation are excluded from this requirement.

## 6 ROLE OF STATISTICS IN REGULATORY SETTING

In 1962 the Kefauver–Harris amendment to the Food, Drug and Cosmetic Act (FD&C Act) led to the provision that the manufacturing companies need to provide substantial evidence of effectiveness and that the evidence show adequate and well-controlled investigations, including clinical investigations, by qualified scientific experts, that proves the drug will have the effect claimed by its labeling" [Section 505(d) FD&C Act of 1962 as amended]. Statistical methodology provides tools for objectively planning, conducting, and evaluating clinical trials. When evaluating the efficacy and safety of a drug from a regulatory perspective, two objectives have to be met: First, the claimed effect is true and, second, the effect is clinically meaningful with acceptable toxicity. The role of a statistical expert at the FDA is, using the data submitted by the applicant, to evaluate whether the claimed efficacy is in fact true. The statistical team works in conjunction with the medical team to evaluate whether the claimed effect is clinically meaningful and has an acceptable toxicity profile.

Before an application [an NDA or biologics license application (BLA)] for marketing approval is submitted to the FDA, product development would typically have

gone through several phases: preclinical phase and clinical phases I–III. Although statisticians are consulted at the design stage of each of these phases, the greatest impact of statistical advice is in planning and designing comparative, randomized, well-controlled studies (phase II or III registration studies). A randomized study mitigates potential selection and reporting biases and balances any known and unknown prognostic factors that may influence the outcome of the study. As detailed in Section 3, new product manufacturers can consult FDA staff by submitting the planned study protocol as an IND application. These protocols are reviewed by the FDA staff and comments are sent to the applicant. In reviewing these protocols including special protocol assessments (SPAs), the statisticians critique both the design and planned analysis of the study.

The protocol should clearly identify the study objective, the primary hypothesis (the benefit that is to be demonstrated), study population, and the primary outcome (or endpoint) that is being measured based on which efficacy claim is planned. The primary endpoint could be a binary outcome (e.g., proportion of patients cured of an infection), a continuous variable (e.g., lowering of cholesterol), or a time-to-event endpoint (e.g., overall survival, time to progression of disease). An endpoint may also be defined as a composite endpoint (e.g., myocardial infarction or death). In certain disease settings more than one endpoint may be of primary interest (e.g., the rate of acute rejection and survival for an organ transplant drug). The design and analysis of a trial depends on the endpoint that is being measured. In conducting and evaluating clinical studies, it is important to control false-positive claims or the so-called type I error rates ($\alpha$). The study also should have enough power or low type II error rates ($\beta$), so that a true effect, if it exists, can be detected with the number of patients enrolled in the study. Subjective endpoints, such as symptom measurements, are difficult to evaluate in open-label studies, and potential biases can be minimized in double-blinded studies [9].

In confirmatory, phase III comparative studies [10, 11], there are basically three types of hypotheses that are tested: (1) the new product is superior [new product compared to placebo; new product compared to an active control (standard treatment); the new product has a dose–response relationship]; (2) the new product is noninferior (the new product is not worse than the existing treatment—it may be better or it may have a similar effect); or (3) two products are bioequivalent [absence of any difference between the two products (e.g., generic drugs)]. Marketing applications with claims of superiority or noninferiority of new drugs are reviewed at the Center for Drug Evaluation and Research. For consideration of a noninferiority claim, the assumption that the patient population and the conditions of the current study with the new treatment are similar to the historical placebo-controlled trials with the existing treatment should be met, the effect of size of the existing treatment must be well established, and that a certain prespecified percentage of the historical control effect size is maintained. Because of these assumptions, in general, noninferiority claims are difficult to establish and require large studies. In hypothesis 3, bioequivalence studies of generic products are generally reviewed by the Office of Generic Drugs at the FDA.

The sample size required to conduct a clinical trial that will ensure that the study is informative to test the primary hypothesis depends on the primary endpoint of interest, the smallest difference in treatment efficacy, which is important enough to detect, the probability of type I error (significance level), the probability of type II

error (1, power of the test), the number of interim analyses that are planned, the number of endpoints, and the comparisons that are being evaluated. If interim analyses or multiple comparisons are planned, then the type I error rate allocated for testing the final analysis has to be adjusted so that the overall familywise type I error rate is maintained at the specified level [10]. Sometimes, meta-analysis, a technique combining results from different studies, is used to provide benchmarks for new agents and new indications and in the design stage of clinical trials.

The statistical analysis plan should be prespecified. The statistical analysis plan should include definitions of the primary and secondary endpoints, and the hypotheses and the analysis plan for each of the endpoints. Comparative analyses based on secondary endpoints are considered exploratory and hypothesis generating if the familywise type I error rate is not maintained at the prespecified level. A post hoc analysis and claim based on a subgroup of patients, when the efficacy could not be demonstrated in the overall study population, is considered exploratory and hypothesis generating, and are generally not considered suitable as a basis for drug approval.

During an ongoing clinical trial, any changes in the definition of endpoint, changes in the type of hypothesis being tested, change in study population, additional interim analyses, increasing sample size, and other study modifications are problematic if the modifications or adaptations are not based on established statistical methods. Such modifications could compromise on the integrity of the study and introduce bias in the interpretation of the results, particularly in an open-label study. Too many modifications and adaptations of the study design and conduct are considered as exploratory. In general drug product approval is based on statistically persuasive results that are clinically meaningful with adequate efficacy and safety.

## 7 ROLE OF CHEMISTRY, MANUFACTURING, AND CONTROLS (CMC) INFORMATION IN REGULATORY SETTING

As per regulations [12], the sponsor must submit CMC information that will describe the general method of preparation of the drug substance, physical, chemical, and biological characteristics, drug product formulation, manufacturing process, and controls to provide assurance of the identity, strength, quality, purity, and potency of the investigational drug. The CMC information that is expected in an initial IND (phase I) is more to assure patient safety and less on product and process controls. However, for phase II and phase III studies, as more manufacturing experience is attained with the manufacturing process and the final formulation is chosen, more exhaustive details regarding process and product controls should be provided. The amount of CMC information required is also dependent on the type of molecule to be tested (e.g., synthetic, semisynthetic, naturally derived, biotechnological) and whether there is any previous clinical experience with the study drug.

The amount of CMC information that will be required in an IND will also be dependent on the marketing status of the drug product. For a drug that is already marketed in the United States, the investigator should provide the current approved labeling and/or package insert of the drug. If a drug is not marketed in the United States, but studied under a cross-referenced IND, the sponsor should provide a letter of authorization (LOA) from the sponsor of the cross-referenced IND to allow the

agency to access to the referenced IND. The LOA should clearly specify the name, strength, and dosage form of the drug. If a drug is not marketed in the United States or it is not referenced to an IND, the sponsor should provide CMC information on the study drug appropriate to the phase of the study. If the drug is intended to be administered for intravenous administration after dilution, compatibility data with different diluents and in-use stability data for the duration of administration should be submitted. Please refer to the following guidance documents regarding the CMC information that needs to be submitted to support the proposed clinical study in the IND [13, 14].

If a drug is not marketed in the United States, but approved and marketed in a foreign market, or it is a legally marketed dietary supplement, the sponsor should provide full CMC information [13, 14]. For a limited phase I study, this requirement may be fulfilled by providing the following information: components and composition of the drug product, name of manufacturer/supplier of the drug product, and the labeling and/or package insert.

The sponsor should also provide a certificate of analysis for the lot that is intended to be used in the proposed clinical trial. Additional product-specific information may be required for the drugs approved and marketed outside the United States or for a legally marketed dietary supplement in the United States. When a sponsor is faced with a situation that is unique and not covered by the above guidance(s) [13, 14], the FDA recommends that the sponsor obtain clarification from the review division with regard to the CMC information necessary to support the proposed clinical study. For a botanical drug, the CMC information that needs to be submitted may be different from that of conventional small-molecule drugs [15].

For all drugs, the sponsor should provide a statement in the CMC section (as well as under dosage and administration section of the clinical protocol), as to the labeled dosage form and the strength of the drug product and how the drug will be administered. If the sponsor intends to change from the labeled dosage form, strength, or route of administration, the sponsor should provide relevant information such as release and stability data to support the change as well as to assure that the change will not adversely affect the identity, strength, quality, purity, and potency of the investigational drug. For all investigational drugs, the following labeling and environmental assessment information should be provided; the label for the immediate packaging of the drug product should be provided that includes the following cautionary statement: "Caution: New Drug—Limited by Federal (or United States) law to investigational use" [16].

A statement requesting a categorical exclusion from environmental assessment under provisions provided for in 21 CFR § 31(e); Part 25, Environmental Impact Considerations; Subpart C, Categorical Exclusions; Section 31, Human Drugs and Biologics, Paragraph (e)—Action on an IND [17].

If a device is used in conjunction with the study drug (e.g., a nebulizer for an inhalation drug or a pump for continuous infusion for home use), information on the manufacturer and model of the device to be employed and a general description of relevant conditions of use (e.g., carrier gas, flow rate, temperature) should be provided. A statement as to whether the device is FDA approved or not approved should also be included. When appropriate, submit adequate information regarding safety instructions for use of equipment and disposal of unused portion of the drug.

## 8 ROLE OF CLINICAL PHARMACOLOGY IN REGULATORY SETTING

The purpose of clinical pharmacology and biopharmaceutics data is to support the determination that the drug to be marketed is safe and effective for use in the indicated patient population. The clinical pharmacology data in conjunction with the safety and effectiveness data from the registration clinical trial(s) can be used to understand the relationships between drug dose or concentration and effects, namely patient outcome and toxicity (exposure–response relationships, ER). These relationships allow for better choices regarding optimizing the dose for the patient population, adjusting doses in special clinical situations, and providing information in the product labeling. Some of these studies provide direct support for the assessment of ER, such as pharmacokinetic information derived from clinical studies (see Fig. 1). On the other hand, some clinical pharmacology and biopharmaceutic data provide indirectly to understanding the behavior of the drug in the patient. For instance, in vitro cytochrome P-450 (CYP 450) screening often indicates whether the drug is metabolized by or modulates any of these enzymes. These in vitro data may suggest whether or not in vivo drug–drug interactions might occur between the drug being developed and other medications that may be commonly administered to a particular patient population that may lead to toxicity or reduced effectiveness.

The summation of these studies should provide appropriate information on drug behavior and its safe and effective use in the product labeling. With that goal in mind, the clinical pharmacology and biopharmaceutics studies can be used collectively to address some basic questions about the drug and how to use it appropriately. With respect to labeling the drug, the main questions are: Is this dose the optimal dose for this patient population? Are any dose or dose regimen adjustments needed for patients with specific clinical characteristics? The answer to these questions are quite complex. Many different chemical, pharmacological, and clinical aspects of the drug contribute to the answers. In the Appendix, many questions about the clinical pharmacology of the drug are posed, and these specific answers, once determined, contribute to the overall issues regarding the optimal dose, dose adjustments, and the safe labeling of the drug.

To address many of these specific issues, the FDA has publicly posted or is planning to post Guidances for Industry that describe the current thinking of the FDA on a specific topic [18]. Each of these is intended to be a general set of guidelines to deal with a specific issue; they may not address every situation, but they should be the starting point for dealing with the issue of interest. Essentially, these are located on the FDA website under the Biopharmaceutics, Chemistry, Clinical, and Clinical Pharmacology headings [18]. The Biopharmaceutics and Chemistry Guidances address issues that affect the drug formulation and may have an impact on how the drug product behaves after administration to the patient (i.e., how much drug may be absorbed). For instance, changing the ingredients of a tablet may alter how quickly the tablet dissolves after administration to the patient, which may have an impact on the absorption of the drug and therefore its effect(s). The Clinical Pharmacology Guidances describe what happens to the drug, and hence its impact on the patient once it is administered to the patient. For instance, renal impairment may cause the exposure of drug to increase because drug elimination is reduced. This in turn may increase the rate of drug toxicity and may necessitate a dose adjustment in these patients.

### 8.1 Biopharmaceutics Guidances

The Bioanalytical Method Validation Guidance [19] is designed to verify that the analytic method used to measure drug concentration is accurate and reliable for the purpose it is intended. The Food Effect Guidance [20] describes the general issues to be considered when assessing the effect of food on a per oral formulation. It addresses the biopharmaceutics issue of how food taken with a medication can alter the dissolution and absorption of a given product. The outcome of this study is often used to determine how the drug should be administered in future trials and, eventually, the intended patient population. During drug development, it is common for a drug developer to alter the drug formulation. These changes may impact the pharmacokinetics and pharmacodynamics of the drug in the patient, and these formulation changes raise questions about how to "bridge" (compare) data from different studies developed with different formulations. The Dissolution Guidance [21] and the Bioavailability and Bioequivalence Guidance [22] are closely related in purpose. Both methodologies are intended to assess the effect of changes in the product formulation on the pharmacokinetics and, by extension, the pharmacodynamics of a drug. The dissolution guidance describes how an in vitro method can be developed using well-defined apparati and criteria. The purpose is to develop a discriminatory in vitro method to assess alterations in the formulation of a per oral drug product (similar approaches can be used to develop an in vitro drug release methodology for a liposome drug product). For instance, changing an ingredient may have an impact on the rate that the tablet dissolves in the gastrointestinal milieu. One could expect that this could alter the extent to which a drug is absorbed and the timing of this event. This could result in undesired effects in the patient (reduced effectiveness or increased toxicity). However, significant changes in dissolution do not necessarily indicate clinically significant changes in the behavior of the drug, so monitoring changes in dissolution is the first step in assessing the impact of a formulation change. Such an in vitro observation is likely to be further assessed by observing the effect of the new formulation in vivo. The Bioavailability/Bioequivalence Guidance [22] describes in general terms the elements of a study for comparing the pharmacokinetics and pharmacodynamics of a preexisting drug formulation with a newly developed one. Studies of dissolution and pharmacokinetics conducted early in drug development can lead to the development of in vitro–in vivo correlations [23] in which a clinically meaningful dissolution method can be used to predict in vivo performance of new formulations. These IVIVC relationships are often used in the development of modified-release drug products.

With regard to clinical pharmacology studies, the characterization of the drug often begins with in vitro studies. Protein binding, although not the subject of a specific guidance, is usually assessed with in vitro incubations for each new drug. (There are examples of ex vivo protein binding determinations as well.) Current scientific thinking is that only free or unbound drug can mediate drug action. Therefore, highly protein-bound drugs (≥90%) have their pharmacological effects mediated by only a small fraction of drug. Furthermore, small changes in protein binding for these drugs might be expected to mediate a substantial effect. Therefore, the extent of protein binding to human serum proteins (e.g., albumin, α-acidic glycoprotein) is commonly performed.

The Drug Metabolism Guidance describes both in vitro and in vivo studies [24, 25]. The in vitro studies describe cytochrome P-450 (CYP 450) screening. The CYP 450 superfamily of enzymes accounts for the majority of drug metabolizing capability in humans. In vitro screening often identifies the main isozyme(s) responsible for the metabolism of a particular drug and the extent of its contribution. New drugs can also be screened for their ability to either inhibit CYP 450 isozyme(s) activity or induce their activities and expression. The end result may be reduced drug clearance or increased drug clearance, respectively, which may affect effectiveness and/or safety. These in vitro findings can then be used as the basis for deciding on the need to conduct in vivo drug–drug interaction studies. It is noteworthy that these in vitro studies do not account for the activity of all metabolizing enzymes, and the metabolism and potential drug–drug interactions on non-CYP 450 metabolized drugs are often difficult to assess. One non-CYP 450 enzyme system that has been studied further is the P-glycoprotein (PGP) transporter system. A newly published draft Guidance on Drug Metabolism describes both in vitro and in vivo studies to assess the impact of new drugs on PGP [26].

Related to the in vitro metabolism studies in purpose are the in vivo "mass balance" studies. These studies are not the subject of a specific guidance but are referred to in some agency publications. The purposes of these types of studies are generally twofold. The first is to determine whether a novel metabolite is generated in humans that has not been previously characterized (at least in part) in preclinical studies. The second purpose is to assess the route of excretion/elimination from the patient. This latter purpose is of importance because it can indicate that patients with organ dysfunction may have difficulty eliminating the drug, which in turn can lead to excessive drug exposure and toxicity. The results of these studies are used to determine whether pharmacokinetic studies in patients with either renal and/or hepatic impairment should be conducted. One common misinterpretation regarding mass balance studies is that the administration of radiolabeled drug is required. This, in fact, is untrue. Any appropriate methodology may be used to assess mass balance. However, use of radiolabeled drug is often employed because of the considerable sensitivity imparted by the methodology.

## 8.2 Clinical Pharmacology Guidances

The basic elements of the pharmacokinetic characterization (clearance, CL, volume of distribution, $V_d$, and elimination half-life, $t_{1/2}$) are characterized in the Format and Content of Human Pharmacokinetics and Bioavailability Section of an Application Guidance [27]. Essentially, the kinetics should be determined to better understand the behavior of the drug. Typically, single- and multiple-dose pharmacokinetics are assessed, usually in early studies in volunteers or patients. Dose proportionality is often assessed with the data generated from these studies. The exposure–response guidance describes the different measures that can be used to evaluate the relationships between drug amount and its effectiveness and toxicity(ies) [28]. These relationships are of great importance because they are the basis for understanding the behavior of the drug and for making rational decisions about choosing the optimum dose or altering the dose when necessary. The Population Pharmacokinetic Guidance [29] describes alternative approaches to assessing the pharmacokinetics of a drug. In some drug development circumstances, the execution of a dedicated phar-

macokinetic study with dense sampling may not be feasible. In these cases, an alternative approach can be employed to collect pharmacokinetic data in a more limited fashion. These data are often combined with previously collected data, and the effects of different factors (covariates) such as disease, patient demographics, and patient characteristics can be assessed, often by using sparse sampling.

The guidances on the effect of renal [30] or hepatic impairment [31] on the pharmacokinetics of a drug also describe important studies. Most drugs are renally excreted or metabolized (predominantly by the liver) prior to elimination or sometimes both. As drug exposure is a function of both the administered dose, and the rate of elimination, renal or hepatic impairment can lead to significant reductions in drug elimination, which in turn produce excessive drug exposures (e.g., higher than anticipated plasma concentrations). Sometimes these elevated concentrations lead to toxicities. Ideally, by understanding the pharmacokinetics (PK) and the pharmacodynamics (PD) of the drug in both the general patient population and in those patients with organ dysfunction, rational dose modifications can be made.

The agency has also published a guidance on the evaluation of QT prolongation [32], which describes how to assess the potential of a drug to adversely affect cardiac conduction.

The agency has also been developing new guidances that deal with issues of drug PK in pregnant women [33] or lactating women [34], as well as in pediatric patients [35].

Recently, the FDA also implemented the EOP1/2a Guidance [36]. The purpose of this guidance is to encourage drug developers to employ all of the available preclinical biomarker, toxicity, and pharmacokinetic and chemistry information along with existing clinical data on effectiveness, biomarkers, and pharmacokinetics in a comprehensive model and to subsequently simulate the outcomes of different drug development trials with different regimens, combinations, and dosages to help choose the most effective trial design for the pivotal trial. In a similar vein, the agency has also issued the Guidance for Industry on Pharmacogenomic Data Submission [37] in a effort to stimulate the development of testing platforms and analysis and utilization of pharmacogenomic data in clinical trials in order to develop better, targeted therapies.

### 8.3 Timing of Studies

With respect to when these studies should be conducted, there is no regulatory requirement. It is up to the drug development sponsor to decide on when and how to allocate resources to these studies. However, there is some sense to the order and timing of these studies. For instance, there may not be too much value in conducting a mass balance study in parallel with the phase III safety and effectiveness trial, if the active metabolites have been identified and organ dysfunction studies have been conducted (or planned). Figure 2 depicts one possible scheme of clinical pharmacology and biopharmaceutics study conduct. According to this scheme, development begins with preclinical studies. The extent of protein binding is determined, and the compound is screened in vitro to evaluate it as a substrate for, or inhibitor and/or inducer of, CYP 450 enzymes. The compound is also screened as a substrate or inhibitor of P-glycoprotein. Additionally, biomarkers may be selected as potential indicators of activity based on animal and in vitro models. Other preclinical studies

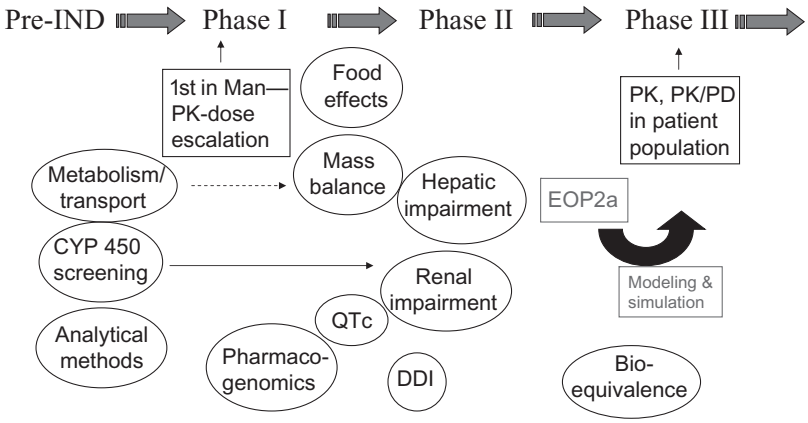


***FIGURE 2*** *One possible clinical pharmacology drug development scenario is depicted. The FDA does not regulate the timing of studies in drug development.*

generate information regarding the toxicity of the compound, including the potential to alter cardiac function (e.g., hERG assay). The analytical method chosen to measure in vivo drug concentrations is often adapted from preclinical to clinical use and validated.

With the first-in-man studies, and ideally throughout the completion of the phase III trial(s), the pharmacokinetics are assessed. Single-dose and multiple-dose pharmacokinetics are usually determined in the dose escalation phase I study(ies), and dose proportionality and time-dependent changes in the pharmacokinetics can be assessed with this data. In some cases, population PK/PD modeling is started with the development of the base model using this data. Collection of samples for pharmacokinetic analyses from phase II may be correlated with markers of effectiveness and toxicity. Drug development according to this scheme would lend itself to an EOP2a analysis, and subsequent simulation of possible phase III trial designs, which may be very helpful in choosing the most efficient regimens, doses, combinations, and/or patient populations. In the late phase I early phase II period, it may be advisable to assess the effect of food on the pharmacokinetics of orally administered drugs in order to decide how to administer the drug in the pivotal trial and the intended patient population. It may also be helpful to consider the effect of two drugs that may be co-administered together. For instance, in oncology, drugs developed for glioblastoma multiforme are often administered with antiepileptic medications, such as phenytoin. The latter drug induces CYP 450 3A4 expression and often reduces the plasma concentrations of the new drug to levels that are too low to be effective. In these cases, the new drug is often developed with concomitant use in all later studies and is labeled for use specifically in combination with each other.

Often in this time frame, thorough QTc studies are performed to assess the potential for the drug to prolong the QT interval, and possibly generate torsades de pointes. Mass balance studies are often conducted some time between the first-in-man studies and phase II. Drug–drug interaction studies and organ dysfunction studies are often conducted in parallel to the phase III, when the decision has been made to pursue the drug marketing. Bioequivalence studies are conducted at various

points during development, from phase I to postdrug approval, depending on the development of the drug formulation and oral preparations with new strengths.

## 9 DRUG REGULATION IN GLOBAL ENVIRONMENT

Drug development has become a global effort with clinical trials in several disease areas often involving a multitude of centers across the globe. Even when some individual clinical trials are conducted in one country or region, a package submitted for approval purposes often includes trials conducted in a number of regions around the world. Even when individual trials are conducted in one region, commercial sponsors usually have a global strategy aimed at marketing approval in a number of countries and regions. While international activities take place in virtually every component of the FDA, the Office of International Programs (OIP) is the focal point for the agency's international activities. It provides leadership for FDA's international activities by guiding and supporting agency programs and by developing strategic relationships with other U.S. and foreign governmental agencies and international organizations. While the FDA takes a proactive role in these activities, it is also worthwhile to discuss not only the role of the FDA but to also describe the roles of other regulatory agencies around the world. The following is a brief introduction to drug regulation in Canada, Europe, Australia, and Japan.

The Therapeutic Products Directorate (TPD) is the Canadian federal authority that regulates pharmaceutical drugs and medical devices for human use and is housed within Health Canada (Ministry of Health) [38]. Prior to being given market authorization, a manufacturer must present substantive scientific evidence of a product's safety, efficacy, and quality as required by the Food and Drugs Act. The Biologics and Genetic Therapies Directorate (BGTD) has regulatory oversight for biologics, radiopharmaceuticals, genetic therapies, and the transplantation of tissues, cells, and organs.

With headquarters in London, the European Medicines Agency (EMEA) is a decentralized body of the European Union (EU) that coordinates a network of 42 national agencies and has undergone significant changes concurrent with the enlargement of the European Union [39, 40]. Its main responsibility is the protection and promotion of public and animal health through the evaluation and supervision of medicines for human and veterinary use. The EMEA began its activities in 1995, when the European system for authorizing medicinal products was introduced, providing for a centralized and a mutual recognition procedure. The EMEA has a role in both, but is primarily involved in, the centralized procedure. Where the centralized procedure is used, companies submit one single marketing authorization application to the EMEA. A single evaluation is carried out through the Committee for Medicinal Products for Human Use (CHMP) or Committee for Medicinal Products for Veterinary Use (CVMP). If the relevant committee concludes that safety and efficacy is proven, it adopts a positive opinion. This is sent to the Commission to be transformed into a single marketing authorization for the whole European Union. Since 2005, applications for marketing authorization within the European Union for diabetes, AIDS, cancer, neurodegenerative disorders, and for products with orphan designation have required submission exclusively through the centralized procedure [40].

In Australia, the Therapeutics Goods Administration (TGA) has authority in the regulation of prescription and nonprescription medicines, medical devices, and gene technology products. The TGA guidelines are aligned with those of the European Union and the International Conference on Harmonisation (ICH) of drug regulatory activities. Further information on drug regulation in Australia can be found on the official webpage [41].

In Japan, regulatory activities have been consolidated into the Pharmaceuticals and Medical Devices Agency (PMDA) since 2004. The PMDA is authorized by the Ministry of Health, Labor and Welfare (MHLW) and includes a staff of approximately 250 personnel including reviewers. The PMDA is comprised of four offices: Office of Relief Funds, Office of Review, Office of Safety, and Office of Research and Development Promotion. Further information on drug regulation in Japan can be found on the official PMDA webpage [42].

In China, regulation of foods, drugs, cosmetics, and traditional Chinese medicinal preparations is the purview of the State Food and Drug Administration (SFDA). Drugs are regulated by the Center for Drug Evaluation (CDE). Further information on drug regulation in China can be found on the official SFDA webpage [43].

## APPENDIX CLINICAL PHARMACOLOGY DRUG DEVELOPMENT QUESTIONS

### Exposure–Response Questions

1. What are the concentration/dose–response relationships? (ER, population PK)
   What are the effectiveness endpoints? In vivo
   - Biomarkers?
   - Clinical outcomes?

   What are the safety endpoints? In vivo
2. What are the factors that affect the concentration–response relationships? (ER)
   Do pharmacogenomics affect the ER? In vitro, In vivo (VGDS)
   What is the extent of protein binding? Does it impact ER? In vitro
   Are there active metabolites that contribute to ER? In vitro (drug metabolism/drug interaction)
   What are the characteristics of metabolism and elimination that affect the ER?
   Is the drug eliminated by metabolism, renal excretion, or both (some other route?)? in vitro, in vivo (drug metabolism/drug interaction)
   Is the drug a CYP 450 isozyme(s) substrate, inhibitor, induce? In vitro
   Are in vivo drug–drug interaction studies needed to assess impact? In vivo
   Are labeling instructions necessary?
   Is the drug a P-glycoprotein substrate or inhibitor? In vitro (drug metabolism/drug interaction)
   Are in vivo drug–drug interaction studies needed to assess impact? In vivo
   Are labeling instructions necessary?
   Are there metabolic polymorphisms that affect metabolism/elimination? In vitro, In vivo (drug metabolism/drug interaction)

Are in vivo drug–drug interaction studies needed to assess impact?
Are labeling instructions necessary?
Are in vivo studies in patients with hepatic impairment needed to assess the need for dose adjustments? In vivo (hepatic impairment)
Are in vivo studies in patients with renal impairment needed to assess the need for dose adjustments? In vivo (renal impairment)
Does age have any effect? In vivo (21 CFR 314)
Does race have any effect? In vivo (21 CFR 314)
Does gender have any effect? In vivo (21 CFR 314)
Are the ER the same in pediatric populations? In vivo (Pediatric PK)
Are specific dosing instructions/regimens required?

3. What biopharmaceutics issues affect ER of the drug?
   Can the active moieties of the drug be accurately measured? In vitro, In vivo (Bioanalytical method validation)
   What is the extent of absorption? In vitro, in vivo (21 CFR 320, BA/BE, BCS Classification)
   Are there different formulations of the drug that affect the ER? In vitro, In vivo (BA/BE, dissolution, IVIVC, SUPAC, IR, MR)
   Are there different oral formulations that absorbed differently?
   Is there a modified intravenous formulation that affect ER? In vitro, in vivo (Liposome drug products) e.g., liposomal drug, antibody-conjugated drug, etc?
   How does food affect absorption of an oral formulation? In vivo (Food effect)

## REFERENCES


1. Hilts, P. J. (2003), *Protecting America's Health: The FDA, Business, and One Hundred Years of Regulation*, Random House, New York.
2. 21 Code of Federal Regulations Part 314.530.
3. Dagher, R., Johnson, J., Williams, G., Keegan, P. and Pazdur, R. (2004), Accelerated approval of oncology drug products: A decade of experience, *J. Natl. Cancer Inst.*, 96, 1500–1509.
4. The Food and Drug Modernization Act, 21 USC 355a, Public Law # 105–115, 1997.
5. 21 Code of Federal Regulations Part 3122(b)(1).
6. Guidance for Industry: Special Protocol Assessment, May 2002; available at: http://www.fda.gov/cder/guidance/3764.htm.
7. Best Pharmaceuticals for Children Act (BPCA) Public Law #107–109, January 4, 2002.
8. Pediatric Research Equity Act (PREA) Public Law #108–155, December 3, 2003.
9. Guidance for Industry Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims; available at: http://www.fda.gov/cder/guidance/index.htm.
10. FDA (1998), Guidance for Industry: International Conference on Harmonization E9 Statistical principles for Clinical Trials Center for Drug Evaluation and Research.
11. Guidance for Industry: International Conference on Harmonization E10 Choice of Control Group and Related Issues in Clinical Trials, 2001.
12. The Code of Federal Regulations Title 21 part 312.23(a) (7), 2006.

13. Guidance for Industry: Content and Format of Investigational New Drug Applications (INDs) for Phase 1 Studies of Drugs, Including Well Characterized, Therapeutic, Biotechnology-derived Products, 1995.
14. Guidance for Industry: INDs for Phase 2 and Phase 3 Studies Chemistry, Manufacturing, and Controls Information, 2003.
15. Guidance for Industry: Botanical Drug Products, 2004.
16. The Code of Federal Regulations Title 21 CFR § 312.6 (a), 2006.
17. The Code of Federal Regulations 21 CFR § 25.31(e); Part 25, 2006.
18. Available at: http://www.fda.gov/cder/guidance/index.htm.
19. Guidance for Industry: Bioanalytical Method Validation. Center for Drug Evaluation and Research, United States Food and Drug Administration, 2001.
20. Guidance for Industry: Food-Effect Bioavailability and Fed Bioequivalence Studies, 2002.
21. Guidance for Industry: Dissolution Testing for Immediate Release Solid Oral Dosage Forms. Center for Drug Evaluation and Research, United States Food and Drug Administration, 1997.
22. Guidance for Industry: Bioavailability and Bioequivalence Studies for Orally Administered Drug Products. Center for Drug Evaluation and Research, United States Food and Drug Administration, 2003.
23. Guidance for Industry: Extended Release Oral Dosage Forms: Development, Evaluation, and Application of In Vitro/In Vivo Correlations, 1997.
24. Guidance for Industry: Drug Metabolism/Drug Interaction Studies in the Drug Development Process: Studies In Vitro, 1997.
25. Guidance for Industry: In Vivo Drug Metabolism/Drug Interaction Studies—Study Design, Data Analysis, and Recommendations for Dosing and Labeling, 1999.
26. Guidance for Industry: Drug Interaction Studies—Study Design, Data Analysis, and implications for Dosing and Labeling, 2006.
27. Guidance for Industry: Format and Content of the Human Pharmacokinetics and Bioavailability Section of an Application, 1987.
28. Guidance for Industry: Exposure-Response Relationships—Study Design, Data Analysis, and Regulatory Applications. Center for Drug Evaluation and Research, United States Food and Drug Administration, 2003.
29. Guidance for Industry: Population Pharmacokinetics. Center for Drug Evaluation and Research, United States Food and Drug Administration, 1999.
30. Guidance for Industry: Pharmacokinetics in Patients with Impaired Renal Function. Center for Drug Evaluation and Research, United States Food and Drug Administration, 1998.
31. Guidance for Industry: Pharmacokinetics in Patients with impaired Hepatic Function: Study Design, Data Analysis and Impact on Dosing and Labeling. Center for Drug Evaluation and Research, United States Food and Drug Administration, 2003.
32. Guidance for Industry: International Conference on Harmonization E14 Clinical Evaluation of QT/QTc Interval Prolongation and Proarrhythmic Potential for Non-Antiarrhythmic Drugs, 2005.
33. Guidance for Industry (draft): Pharmacokinetics in Pregnancy—Study Design, Data Analysis, and Impact on Dosing and Labeling, 2004.
34. Guidance for Industry (draft): Clinical Lactation Studies–Study Design, Data Analysis, and Recommendations for Labeling, 2005.

35. Guidance for Industry (draft): General Considerations for Pediatric Pharmacokinetic Studies for Drugs and Biological Products, 1998.
36. Guidance for Industry: End of Phase 1/2a Meetings, 2004.
37. Guidance for Industry: Pharmacogenomic Data Submission, 2005.
38. Farrell, A. T., Papadouli, I., Hori, A., Harczy, M., Harrison, B., Asakura, W., Marty, M., Dagher, R., and Pazdur, R. (2006), The advisory process for anti-cancer drug regulation: A global perspective, *Ann. Oncol.* 17, 889–896.
39. Pignatti, F., Boone, H., and Moulon, I. (2004), Overview of the European regulatory approval system, *J. Ambul. Care Manage.* 27(2), 89–97.
40. Available at: http://www.EMEA.eu.int.
41. Available at: http://www.tga.gov.au.
42. Available at: http://www.pmda.go.jp.
43. Available at: http://www.sfda.gov.cn.

# *Future Challenges in Design and Ethics of Clinical Trials*

**Carl-Fredrik Burman[1] and Axel Carlberg[2]**

[1]*Technical & Scientific Development, AstraZeneca, Mölndal, Sweden*
[2]*Department of Cardiothoracic Surgery, Lund University Hospital, Lund, Sweden*

**Contents**

1 Introduction
  1.1 Introducing Public Trust
  1.2 Introducing Productivity
2 Challenge of Public Trust
  2.1 Requirement of Genuine Consent
  2.2 Requirement of Patient Benefit Optimization (PBO)
  2.3 Requirement of Nonexploitation
3 Challenge of Clinical Trial Efficiency
  3.1 Background
  3.2 Decision Analysis Approach to Study Design
  3.3 Optimizing Information per Patient
  3.4 Optimal Amount of Information to Collect in Trial
  3.5 Adapting to New Information
4 Discussion and Conclusions
  References

Disclaimer: The views expressed in this chapter do not necessarily reflect those of our respective affiliating institutions.

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

## 1 INTRODUCTION

Clinical development faces two major challenges: public trust and productivity. Costs in drug development have increased substantially over the past decades [1]. Notwithstanding this development, the number of new drugs that are implemented in clinical practice has not increased but rather decreased in recent years [2]. At the same time, recent opinion polls [3] show that the public trust in pharmaceutical companies has fallen sharply and is now almost at a level comparable to that of the oil and tobacco industries. While some may argue that these challenges of public trust and productivity are difficult to align, we consider both to be essential and indeed mutually dependent. If public trust in clinical trials decreases, recruitment of patients for clinical trials will be more difficult, which in turn will result in even poorer productivity.

The ambition of improving public trust, especially in view of the tragic incident at Northwick Park Hospital in March 2006, has much to do with heightening ethical standards. The dual objective of this chapter is therefore to clarify the nature of the ethical challenges that lie ahead and to specify the design features that can increase productivity in individual clinical trials under ethically acceptable standards. Our question could be formulated as follows: Given that clinical trials should meet stringent requirements, how can productivity be maximized? In order to formulate an answer to this question, we will first have to specify the ethical requirements. Once these have been formulated, we will provide a method of analyzing trial efficiency and will give examples of how productivity can be increased. In doing so, we will consider specifically how to optimize the information per patient, approach sample size determination from a business perspective, and choose between adaptive design options.

This reflection is inspired by decision analysis (DA), which is a method that aims at optimizing different kinds of decisions, often by analyzing trade-offs. The essential assumption of DA is that different objectives can be valued on a common scale. For example, if one treatment offers better health benefits while the other provides better safety, it is possible to compare the two and choose the treatment that maximizes the net utility for the patient. Furthermore, even if the respective health benefits and safety features of these treatments are uncertain, DA offers the possibility of comparing the expected utility of the two treatment options. In this study, the first DA will be patient-centered, meaning that the net utility of the *patient* will be focused. After and secondarily to that, the interests of science and the sponsor of the trial will be considered, and a commercial optimization of clinical trials will be outlined. This priority is, in our minds, the only one that is ethically acceptable.

### 1.1 Introducing Public Trust

The study will therefore begin by a broad review of the ethical issues facing clinical development and more specifically randomized clinical trials (RCTs). In doing so, we formulate three minimal ethical requirements that should be met before patients can be enrolled in RCTs: genuine consent, patient benefit optimisation (PBO), and nonexploitation. These requirements are addressed principally to physicians who ultimately are responsible for their patients' health and for the enrollment of their

patients in any RCT. Our requirements do not cover properly scientific or regulatory concerns, and it is taken for granted that any clinical trial must also be examined and approved by an ethical committee or an investigational review board (IRB). In addition, we take for granted that regulatory approval is needed to administer new pharmaceuticals outside clinical trials. A more detailed list of ethical requirements covering such scientific and regulatory issues has been formulated by Emanuel et al. [4].

The core issue being treated here is therefore the physician/patient relationship and how it is affected by the collaboration of the former in clinical development. We argue that this collaboration does call into question the physician's fiduciary commitment to the well-being of his patient, as he or she must concurrently serve the interests of science. Such collaboration, which constitutes a threat to patient trust, can only be ethically justified through the consent process in which the patient acquiesces to participate in an RCT. The patient's informed consent in this matter should therefore not only be a formal but genuine and be based on relevant and appropriate information allowing the patient to formulate a well-reasoned judgment about such participation.

While this genuine consent is a necessary condition, it is not a sufficient prerequisite for trial participation. There are instances when the patient might accept to take part in a trial but when such participation would nonetheless be unethical. Such situations might arise if the patient is not sufficiently informed or free to consent or decline trial participation or when the trial design itself is morally skewed. That is why the physician, who is ultimately responsible for the patient's medical well-being, must be certain that he or she can, in good conscience, propose an RCT to his or her patient. Traditionally, this ethical peace of mind has been sought in the state of equipoise, understood as a state of "genuine uncertainty" on the part of the physician as to the merits of the respective arms of the trials. We argue, in a Bayesian perspective, that such genuine uncertainty in reality never exists. The physician has always a hunch or can second guess, on the basis of medical competence and experience, which arm in the RCT will receive the best treatment. An appeal to equipoise alone will therefore not do. We believe that the only way to ethically justify physician collaboration and patient participation in a RCT is the requirement of patient benefit optimization, that is, the judgment that the patient will benefit more from inclusion in the trial than from exclusion [5].

Besides genuine consent and patient benefit optimization, there is a third ethical requirement that we believe is essential: the principle of nonexploitation. It stipulates that no coercion should be used in the recruitment of patients. In other words, a patient or volunteer must be provided with sufficient freedom to willingly accept or decline an invitation to enroll in a RCT. The research subject might not, for instance, feel free to decline an offer to participate in a trial if this would result in, for example, inferior treatment. Such a situation can arise when research is carried out in developing countries or if patients are required to participate in a clinical study in order to qualify for medical coverage.

### 1.2 Introducing Productivity

Within the framework of these ethical requirements, the sponsor must design the most efficient trial. What does efficiency mean, however? Just as different

therapeutic treatments have medical pros and cons, trial designs have also relative benefits and drawbacks. Thus, DA is useful, not only in evaluating the ethical issues involved but even more so in the analysis of efficiency. In Section 3.2 we propose a detailed 10-step DA approach to clinical development. In this introduction, we limit ourselves to outlining the four most important steps, which are (1) modeling background information, (2) specifying trial objectives and utilities, (3) identifying design options, and (4) optimizing the decision of the three previous steps.

Experimentation is driven by the need to know more about the medical effects and safety features of a certain drug. It is therefore important to evaluate the information that is already available and to quantify the level of uncertainty. Existing information about the drug's characteristics can usually be compiled from a number of sources. Previous clinical trial data are often limited, especially in early clinical development. To fill in the information gap, investigators can refer to preclinical data, literature on the disease and related drugs, and expert input. In a DA approach, a quantitative model must first be constructed based on already existing information and a set of assumptions. These assumptions are necessarily subjective and should therefore be challenged during the analysis process. Then the objective of the trial must be formulated from a wide enough context in order to avoid suboptimization. The overall objective is then converted into quantitative preferences, which can be called utilities. The utility is thus a function of the outcome of the trial and reflects both the value of information as well as the cost of experimentation. In Section 3, we will take the commercial net present value (NPV) as an approximation of the utility of the trial sponsor. Third, DA can help to identify the designs and to tailor them in such a way that they are compatible with the objectives of the experiment. Still today, many trials are given standard designs, without taking into account its specific objectives. We therefore argue that standard procedure, at the outset of any trial, should be to generate, through project team brainstorming, a number of designs by considering common options such as cross-over, adaptive, and enrichment designs. Finally, the fourth step consists in predicting the trial results for the different designs. These predictions are made from the model and are typically carried out by computer simulations. When this is done, the predicted trial outcomes can be combined with the utility function. An expected utility can then be calculated and the optimal trial be designed in view of maximizing this value. By questioning the underlying assumptions and making tentative modifications of them, the robustness of this result can be verified.

The DA approach will thus be the dominating framework when discussing different aspects of study design. We will therefore discuss how much information should be obtained from each patient in view of optimizing the design (Section 3.3). Multiple measurements and multiple treatments (cross-over study) for each individual will yield more information but might also cost more and cause methodological problems. It is therefore important to analyze the optimal amount of information to purchase, that is, the optimal sample size when information value is compared with the cost of the experiment (Section 3.4). Finally, we will consider some sequential and adaptive designs where the trial design decisions depend on accumulating information (Section 3.5).

A number of recent trends in clinical study methodology and technology are supporting the DA approach to trial optimization. Model-based drug development,

and especially pharmacokinetic/pharmacodynamic (PK/PD) modeling, are being used more and more by the industry and are being promoted by regulators. The Food and Drug Administration (FDA) has stated that a drug may, in some instances, be approved on the basis of only one single confirmatory trial when supported by modeling results. The increase in computer speed has opened the way for nonlinear mixed-effect models as well as for Bayesian methods, where a probability model is updated with accumulating data. New design options are identified, and there has been an increasing interest in adaptive designs during the last decade. Web-based data capture has greatly facilitated the practical implementation of sequential and adaptive designs. The interest in decision-analytic approaches is on the upswing, and methodological work is being done in related areas, ranging from the psychology of decision making to the integration of commercial values in sample size determination.

## 2 CHALLENGE OF PUBLIC TRUST

From the point of view of medical ethics, clinical trials are problematic because they challenge the generally held view that the principal responsibility of physicians is to further the well-being of the individual patient. This responsibility is the cornerstone of the physician–patient relationship. As Ruth Purtilo argues [6], the professional's priority to further the welfare of the patient:

> is interpreted to have the stringency of a special moral obligation on the part of the professional to seek a patient's well-being, guided by that person's health-related concerns and needs. Any other worthy goal, such as furthering knowledge about disease and its cure, or earning a just wage, or maintaining the efficiency or financial solvency of the institution, is not an appropriate beacon to guide the professional in this relationship.

This hallow responsibility is enshrined in most codes of ethics regulating medical conduct and has enjoyed almost universal recognition. In the Oath of Hippocrates (Greece, 4th century B.C.E.), the physician swears to apply his art "for the benefit of the sick according to my ability and judgment" adding that he will keep the sick "from harm and injustice." In the Indian tradition, the Oath of Initiation (Caraka Samhita) from the first century C.E. puts it another way: "Day and night, however you may be engaged, you shall endeavour for the relief of patients with all your heart and soul." Similar enjoinders can be found in the Advice to a Physician (Persian tradition 10th century C.E.) and in the 17 Rules of Enjuin (Japanese tradition 16th century C.E.). In modern times, the International Code of Medical Ethics adopted by the World Medical Association in 1949 phrases this obligation as such: "A physician shall act only in the patient's interest when providing medical care which might have the effect of weakening the physical and mental condition of the patient" [7].

Insofar as the objective of clinical trials is to advance science through the randomization of patient treatment, the collaboration of physicians and the involvement of their patients in such trials would seem at first glance to be ethically problematic. The question that looms at the outset of this study is whether clinical

trials may be conducted under ethically acceptable standards. To answer that question we must begin by examining the nature of clinical practice and especially the character of the physician–patient relationship. This reflection will help us to determine whether the collaboration of physicians and the participation of patients in such trials can ever be justified, and if so, under what conditions.

## 2.1 Requirement of Genuine Consent

The startling breakthroughs of biomedical science during the last decades, and the promise of more revolutionary treatments and drugs, obscure the simple fact that medicine is primarily a relationship between two individuals. As Thomasma [8] has pointed out, "medicine is not a theory about the body or how the body works, so much as a theory about practice, about the ways in which physicians and patients, other health professionals, and institutions interact to bring about healing" (p. 246). This interaction begins when a person in a vulnerable state due to physical or psychological illness or distress requests assistance and is admitted to medical care. In doing so, the patient assumes that the physician, as the person in charge of his or her care, will act in his or her best interests and will prescribe the most effective medication or treatment available. What characterizes this relationship is therefore trust and responsibility. The patient trusts that the physician will act in his or her best interest while the physician accepts to shoulder the responsibility of taking care of the patient to the best of his or her abilities.

In the Hippocratic tradition, which still shapes our understanding of the ethos of medicine, the physician–patient relationship is understood in terms of a certain kind of friendship [9, p. 20]. The friendship of the physician (*philantropía*) sees to the well-being of the patient and is expressed through the love for his art (*philotekhnía*), which is healing. While there were many practitioners of the art of healing in ancient Greece, such as magicians and exorcists, the Hippocratic physician loved his art as a *techné*, that is, as an empirically established technical body of knowledge useful to his craft. The Hippocratic Corpus [10, p. 314] states unequivocally the commitment of the physician to empirical science:

> Do not rely on conclusions that result from mere reasoning, but (rely) on evidence. Arguments in the form of plain rhetoric are false and easily defeated. Therefore, you should stick to the facts and scrutinize them, if you are going to acquire faultless capability, which we call Medicine.

Thus, the Hippocratic physician loved his art through his concern for the well-being of his patient and his patient through the practice of his scientifically established craft [9, p. 45]. In other words, the physician loves his patient as a friend because he can come to his aid by applying his craft and knowing that this craft is well founded and likely to produce good results. In this view, science is an essential medium but still plays an ancillary role to the practical object of medical practice, which is the healing of the patient.

In the context of the discussions of the morality of randomized clinical trials, this view of the physician–patient relationship has undergone a radical change. Medical science is no longer seen as primarily an instrument that permits the physician to

exercise his art of healing but as a competing object of his professional responsibility. In other words, it is no longer only a medium but equally an end of the physician's work and thus regulated by an independent set of values. In a very influential article, Lellouch and Schwartz [11] characterize the conflict of interest facing the physician as one between an individual ethics and a collective ethics.

Clayton [12] spells out what these competing sets of values involve. Individual ethics refers, on the one hand, to the duty of the physician "to apply existing knowledge for the best possible treatment of each individual patient." Collective ethics, on the other hand, refers to two duties of the physician: (1) "to acquire new knowledge so that, by that advance, future patients might benefit" and (2) "having acquired that knowledge, to accurately communicate it to other physicians." Gifford [13] notes the existence of the physician's "therapeutic responsibility," which he believes, prima facie, to be incompatible with his collaboration in randomized clinical trials. However, mindful of the benefits that clinical trials bring them, patients might allow the physician to collaborate and would willingly submit themselves to randomized treatments. In this contractual scheme inspired by the philosophy of John Rawls and Robert Nozick, randomized clinical trials could be considered ethical to the extent to which the patients acquiesce to participate and permit the physician to override his or her therapeutic responsibility in order to further a more collective good.

Gifford's contractual approach defers the physician's fiduciary responsibility and places the decision of RCT participation on the shoulders of patients themselves. This approach corresponds well with the insistence upon the informed consent process and the principle of autonomy, which permeates today's ethical climate. This principle of autonomy or permission is widely considered today as the lowest common ethical denominator and the foundation of modern ethics. The reasoning behind one of the major proponent of this view, H. T. Engelhardt [14], is the following. There is a plurality of moral goods, which human beings can pursue and enjoy such as freedom, work, health, pleasure, religious faith, and many others. How these values are ranked in the pursuit of a good life depends very much upon one's political, religious, and philosophical convictions. Furthermore, these different values cannot be arbitrated by appealing to an objective standard, to which everybody can agree. In light of this plurality of moral opinions, it is up to the individual to determine what is good for him or her. A man suffering from prostate cancer might, for instance, want to choose between radical prostatectomy and more conservative methods depending on what outcome he values. If he is younger and still wants to father a child, he might value the preservation of fertility very highly. However, if he has no such thoughts, he might opt for the treatment that offers the best possible outcome in terms of life preservation. It all depends on what values he espouses and what plans he has for the future. Free and informed consent is, in this light, the gold standard of medical ethics.

The principle of autonomy or permission, with its stipulation of free and informed consent, is, in the view of most ethicists, a necessary condition for patient participation in any RCTs, with the possible exception of trials carried out in emergency situations where patients or their next of kin cannot give immediate consent. There is also general agreement that participants should be informed specifically about the relevant details of the randomization process. It is not the quantity of information

given to the research subject that determines whether the consent is free and informed but his or her ability to understand relevant facts in order to make an informed decision. That is why some ethicists talk about "genuine" rather than informed consent [15]. This wording emphasizes the fact that true consent requires health care providers to do their best to communicate accurately and in a way that is understood by patients, volunteers, or relatives. It presupposes that these can understand procedures and risks and react to the limits of their understanding. Moreover, it requires care in detecting and eliminating lack of consent.

While there is general consensus about the necessity of informing patients on randomization procedures as part of the consent process, and prior to trial participation opinions differ on the necessity of informing research subjects about interim data. Those who argue against such disclosure [16] fear that interim data can be unreliable and misunderstood by research participants and the media, which in turn could compromise the completion of RCTs. In a contractual understanding of RCT participation, it is possible to conceive that patients would waive the right to such data, and would accept an "incomplete disclosure" [17] as part of the consent process. Lilford et al. [18] disagree and argue that the praxis of withholding interim data has no public endorsement, is contrary to the spirit of regulations, and is the reflection of a "culture of secrecy and scientific imperialism" that is self-fulfilling. Making interim data available, they argue, would not necessarily reduce recruitment to clinical trials but would most likely increase it because it would enhance public trust in the study. As a result, it would also widen the base of the participants, thus improving the trial's generalizability. The authors provide the example of the second international study of infarct survival (ISIS 2) in which disclosure of improved survival rates of heart attack patients taking clot busting drugs increased recruitment as sceptical physicians became more interested in the interim results. This example shows that more stringent ethical standards do not necessarily decrease productivity. In some instances they could even increase it.

### 2.2 Requirement of Patient Benefit Optimization (PBO)

The Hippocratic model of the physician–patient relationship, where the physician bears the main burden of responsibility and decision making, is considered by many as too paternalistic. Today, patients want to be full-fledged partners in the clinical process, together with the physician. This is all the more important as the physician's loyalty to the patient competes with his or her scientific ambitions. The patient must therefore also fend for himself. Informed consent procedures are one step on the way to establishing more equality between patient and physician. Furthermore, a more democratic approach might give added credence to clinical trials and does not necessarily result in less efficiency. On the contrary, more public trust through greater transparency could result in better recruitment and more generalizability. But there are instances where the requirements of informed consent might not be sufficient to make a clinical trial ethically acceptable.

In addition, the principle of autonomy with its requirement of informed or genuine consent builds upon a model of the patient as a rational agent. This model does not describe accurately how patients make clinical decisions concerning their own welfare. Individuals often behave and make decisions irrationally. To compare treatments is often a complex task, involving a multitude of effect, safety, and con-

venience dimensions. When asked to choose between two treatments, we may therefore—irrationally—prefer treatment A to treatment B, B to C, but C to A, thus violating the transitivity axiom in decision theory. Furthermore, many people have difficulties in understanding small risks. Lastly, the patient is not always able to give informed or genuine consent as, for example, in an emergency or situation of distress. Such a situation may arise when the patient is informed that he suffers from a chronic or fatal disease. He may feel that the physician's time is limited and he may interpret a question about trial participation as an advice or wish from his physician. If the physician clearly communicates that he does not know which treatment is best, and has no advice about study participation, the patient may give his formal consent by signing a piece of paper without really understanding the content and relevance of the information received and thus without acting in a truly free and informed manner.

Because patients are not always capable or even willing to act rationally, and bearing in mind that genuine consent is sometimes difficult if not impossible to retrieve, the attending physician ultimately bears the responsibility of furthering the well-being of his patient and assuring that the trial itself is ethically sound. How can this be assured?

Over the past 30 years, the notion of equipoise has been used to ethically justify physician collaboration in RCTs. It was first used in this context by Charles Fried [19] in 1974 when speaking about the "balance of opinion" among physicians and investigators, when no treatment option in a multiarm RCT is perceived to be inferior to the other or others. From this balance of opinion the term has taken a more psychological turn to connote the "state of genuine uncertainty on the part of the clinical investigator regarding the comparative therapeutic merits of each arm in a trial" [20]. According to this standard, RCTs would be morally justified to the extent to which the clinical investigator can honestly claim a "state of genuine uncertainty" between the treatment options incorporated in the study design.

Most ethicists and clinicians agree that true equipoise rarely exists. Clinical trials are financed and conducted because there are reasons to believe that a new drug or treatment option will improve care, which is also why clinicians choose to implement them in their practice. Notwithstanding this a priori, there might be, however, scholarly discussions in the scientific community about the relative efficacy of a novel therapy. So while absolute or "theoretical equipoise" might not exist, there might still be a nagging uncertainty among investigators and health care providers about the benefits of the different treatment arms. This practical uncertainty or "clinical equipoise" would be the requirement that scientifically justifies such trials. Another term that is used in this context is the ethical theory of "null hypothesis" pertaining to genuine medical uncertainty. The hypothesis entails that clinical trials are justifiable when no consensus exists about which is the better treatment [12, p. 472].

The null hypothesis theory is an easy and questionable justification of RCTs. Consensus among clinicians and researchers about the benefits of a certain treatment versus an alternative is rare, even regarding established treatments. Given that consensus rarely exists, most if not all RCTs could be considered, according to the equipoise theory, as morally justifiable. Moreover, as Lilford [21] has pointed out, even the justification of "uncertainty" regarding treatment outcomes is indeed a

questionable criterion. "Unknown" or "uncertain" is not equivalent to saying that all possible effects are equally likely. Clinicians and investigators always possess some evidence about the merits of the novel treatment or drug before the trial begins. Such information, which can be the basis for drug modeling, is obtained through animal or in vitro experiments, the same treatment in other diseases or, in some cases, even randomized trials. Saying therefore that the effects of the novel treatment are uncertain is certainly ambiguous. The patient might interpret such a statement as meaning that the recruiter has no idea at all about the benefits of the new treatment, which is not the case. Certainty and uncertainty are not absolute terms, and the physician is therefore most likely to have some idea about relative advantages and disadvantages. As Lilford states, "the blanket term of unknown sidesteps any indication of the magnitude of possible effects, reducing the chance that potential participants will be able to appreciate what is really at stake" [21]. In addition, the whole theory of equipoise and ethical null hypothesis is paradoxical because the more an investigator knows about a compound or a treatment the more he would draw away from equipoise. Oddly, it would encourage ignorance.

The notion of equipoise, as used in the debate over RCTs, is therefore problematic because it places too much emphasis on the physician's state of mind. Pleading uncertainty, as Lilford points out, says very little about what is known and what should be told to a patient concerning the merits or drawbacks of a certain treatment option. What is really ethically at stake in RCTs is not whether the physician considers himself or herself in a state of equipoise but whether the patient will benefit more from inclusion in than from exclusion from the trial. This is precisely what the requirement of PBO entails. If the patient stands to benefit personally by inclusion in the trial, the RCT would meet the requirement. If he or she does not, the requirement would not be met.

Studies in which patients in the control arm receive the standard treatment while patients in the experimental arm are given a novel drug expected to be superior would fulfill the requirement of patient benefit optimization. With all patients receiving standard treatment as background medication, it is often acceptable to compare the addition of a new drug with the addition of placebo [5]. But what about placebo-controlled trials where patients in the control arm would be given no pharmaceutical treatment at all? Most ethicists agree that there are conditions when placebo-controlled trials are acceptable and others when they are clearly not. If the ailment under investigation is not serious, and when there is minimal risk and minimal burden for research participants, the use of placebo is generally considered as ethical. A trial studying allergic rhinitis is a good example since this ailment typically does not impair health or causes severe discomfort. Since the possibility of harm is minimal, the benefit derived for the patient in the control arm could be the satisfaction of helping others. Such a trial, despite the use of placebo, could therefore meet the requirement of patient benefit optimization. There are, however, other situations where the use of placebo is unethical. If effective, life-saving, or life-prolonging treatment is available, and if research subjects who receive placebo are more likely to suffer serious harm, a placebo-controlled trial is unethical. The disagreements among ethicists, investigators, and clinicians center on whether placebo controls can ever be used in trials where a treatment known to be effective is available and when there is some potential harm to participants being randomized [22].

The Declaration of Helsinki [23], the most authoritative text on matters of research ethics, is not totally clear on this matter. Paragraph 32 states: "The benefits, risks, burdens and effectiveness of a new intervention must be tested against those of the best current proven intervention, except in the following circumstances: the use of placebo, or no treatment, is acceptable in studies where no current proven intervention exists; or where for compelling and scientifically sound methodological reasons the use of placebo is necessary to determine the efficacy or safety of an intervention and the patients who receive placebo or no treatment will not be subject to any risk of serious or irreversible harm. Extreme care must be taken to avoid abuse of this option." Commentators argue that this wording in the latest revision of the declaration signals a move to a more permissive attitude toward placebo use. This interpretation is motivated by the fact that the 1996 version included the provision that "in any medical study, every patient—including those of a control group, if any—must be assured the best proven diagnostic and therapeutic method" (Paragraph II.3). This sentence, once again, has been eliminated. Despite this omission, the overall tenet of these both versions seems to rule out the use of placebo wherever proven treatment exists. Paragraph 32 in the latest version adopted in 2008 in Seoul, lists two situations where placebo is acceptable: where there are "compelling and scientifically sound methodological reasons," *or* where the condition under study will not lead "to any additional risk of serious and irreversible harm" [23]. What has led commentators to interpret more leniency regarding placebos is the option of the connector "or" instead of "and" between the two provisions. It would seem therefore that the Declaration of Helsinki, in its present version, could be interpreted as saying that the use of placebos for "compelling and scientifically sound methodological reasons" is permissible even when there is an increased risk of serious harm.

The spirit of the text does not go in that direction and the wording of Paragraph 6 ultimately invalidates such an interpretation. This paragraph states that "in medical research on human subjects, considerations related to the well-being of the human subject should take preference over all other interests." However, much of the confusion surrounding Paragraph 32 has to do with the terms "best current proven" or "proven" intervention. Does the text mean "proven" in existence, in a local context, according to the judgment and knowledge of the attending physician, or according to the best interests of the patient? The Nuffield Council on Bioethics [25] suggests that "best proven" should mean "the minimum standard of care that should be offered [in the control arm] is the best intervention available as part of the national public health system."

The Declaration of Helsinki encourages ethically optimized designs that exclude the use of placebos. Consider, for example, a situation where it is scientifically interesting to compare two different drugs, X and Z, which both are available on the market. By giving each patient the drug that is expected to be best for him or her, and assuming that this choice will be different for different patients, it is possible to learn more about the relative merits of the two drugs. Say that X is perceived to be better in patients with mild disease, expressed as a baseline value of a certain covariate below a cut-off $c$, while Y is thought to be better for patients with covariate values above $c$. If the responses are collected, the results from both groups can be modeled as functions of the covariate and extrapolated so that a comparison of the expected responses of the two treatments can be done. As more and more observations are collected, the estimated cut-off will change and treatment practice will

improve over time. This type of trial would be in conformity with the Declaration of Helsinki ethically optimal but, unfortunately, it is not very efficient or viable compared to standard RCTs.

More often than not, it is desirable for methodological and regulatory reasons that the experimental treatment Y is tested against placebo without using X. Can any such trial withstand the PBO requirement? Imagine a trial where participation is beneficial on average but unfavorable for the patients randomized to one of the treatment arms. Say that mortality is estimated to be 10% for placebo, 9% for the "best proven" therapeutic method C, and 8% for a new nonapproved drug candidate X. In order to get regulatory approval for X, the trial would need to prove its superiority versus placebo, which is much easier and cheaper than to show superiority, or even noninferiority, versus C [26]. Therefore, a blinded trial is proposed where ¾ of the patients are randomized to X and ¼ to placebo. The expected death risk for a patient taking part in this trial is ¾(8%) + ¼(10%) = 8.5%, which is lower than the 9% risk experienced by the patients that do not take part in the trial and receive C. This trial would meet the requirement of PBO because, statistically, the patient would benefit more from trial inclusion than from exclusion. It is still, however, very debatable whether the trial would be ethically acceptable because it inevitably exposes certain patients to less than "best proven" treatment. This example shows that while the PBO is a necessary ethical requirement, it is not alone a sufficient requirement for ethically validating a trial.

The example above also indicates that the PBO should not be interpreted as an average utility for all participating patients in the trial but should also be relevant and applicable to every individual participant. Retsas [27] has formulated a thought experiment, which illustrates this point through what we will call the VIP principle:

> Imagine that you are an acclaimed expert on a particular disease and you are invited to offer your services to your head of state. Will the head of state be enticed to participate in a randomised trial leaving his or her treatment to chance, or will treatment be provided on the best available knowledge of the day? It can be observed safely that there are hardly any prime ministers, secretaries of health, doctors, professors, or similar persons who enlist in the ranks of randomised cohorts that serve the progress of medicine!

Retsas' thought experiment is very general and suggests that all RCTs are unethical. While some ethicists have argued that such is the case [28], most believe that it is not. Retsas is also wrong in stating that "doctors, professors or similar persons" do not participate in RCTs. They often do, and the history of medicine is full of examples of clinical researchers risking their lives with self experiments in the zealous pursuit of scientific knowledge. But the point is well taken that physicians seldom recruit VIPs, be they heads of state or their own children, to RCTs. When human life is truly at stake, leaving treatment option to chance, even when there are reasons to believe that the experimental group stands to benefit, is indeed counterintuitive. The fact that a high percentage of drugs tested in phase III clinical trials never reach the market would lead us to exercise extreme caution. The poignant question that Retsas addresses is this: If VIPs are seldom, if ever, randomized by physicians, why should anybody else be? While the VIP principle might be intuitively sound, reason can still ethically motivate RCTs if every individual patient

can be assured that he or she will receive the best possible treatment or its equivalent.

### 2.3 Requirement of Nonexploitation

It has been argued that genuine consent should be a necessary condition for enrollment in a clinical trial. Given that RCTs distort the fiduciary relationship between patient and physician through the latter's participation in what is essentially research and not primarily therapy, the patient's genuine consent is essential in order to reestablish the contractual balance between the two parties. Such a requirement, it has also been argued, is not a sufficient condition for patient enrollment in RCTs. Besides genuine consent, the physician must also ensure that the design of the trial, especially its expected utility, can be aligned with the patient's best overall interests. From an expected utility perspective, the patient should stand to benefit more from inclusion in the trial than from exclusion. There are, furthermore, other cases involving research subjects where the ethical issues involved in the physician–patient relationship are not covered by the two aforementioned requirements, hence the necessity of formulating the requirement of nonexploitation.

A good example of where this requirement is at stake is first-time-in-man (FTIM) trials planned in healthy volunteers. Usually, trial participants are offered economic compensation for the time and discomfort associated with the experiment. Is it unethical to pay these participants for taking medical risks, and would such payment constitute a form of coercion? Assume that there are no discomforts in such a trial, that trial participation is short, that there are no health gains for the volunteers but a slight risk of experiencing an unexpected serious adverse effect. With the possible exception of altruism, there would then be no other reason to participate in the trial other than economic compensation. Assume, furthermore, that a firefighter in the United States is paid to take a small risk in a clinical trial in much the same way that he is compensated for work-related risks. Although it is hard to assess trial risks objectively, it is certainly possible to give some ball-park estimates. Severe adverse events are extremely rare but, as the incident in Northwick Park Hospital in March 2006 shows, they do occur and could potentially lead to death. As a very approximate risk calculation, assume an expected 0.01–0.1 deaths per year globally in FTIM trials, 100 new approved drugs per year, 10 times more drugs tested in humans, and 10 individuals per trial. Then the average death risk for participants in FTIM trials would be about 1 in 100,000 or 1 in 1,000,000. A more careful analysis, using historical data and expert input, may give much better estimates. The risks will also depend on the different kinds of drugs. Hence, there are nonnegligible risks, but these are small and comparable to those taken by some groups as part of their professions. As a comparison, the fatality rate per year in the United States is almost 1 in 1000 for logging workers and fishers and 1 in 4000 for truck and taxi drivers [29]. Healthy volunteers in FTIM studies are paid for their time and for the discomforts they might suffer as a result of the study. From a DA perspective, however, risks and economic gains can be factored in the decision about whether or not to take part in the trial.

The requirement of nonexploitation stipulates that research subjects must be sufficiently free to be able to consent genuinely or, by the same token, to decline trial participation. Generous financial benefits would cast a doubt over certain

persons' capacity to decline an offer of participation, especially those who are economically vulnerable. In this respect, the additional protocol to the Bioethics Convention on biomedical research, adopted by the Council of Europe in January 2005, states very clearly that "no undue influence, including that of a financial nature, will be exerted on persons to participate in research. In this respect, particular attention must be given to vulnerable or dependent persons" [30, Article 12]. The explanatory report of the protocol does not strictly define the term "undue influence" but suggests that it is related to coercion and "may be exerted in particular on a person in a weak or feeble condition, so that very little pressure will overbear the person's will, and make the individual feel that he or she must agree" Furthermore the text specifies that "compensation should not be provided at a level that might encourage participants to take risks that they would not otherwise find acceptable" (Paragraph 64). From an expected utility perspective those who stand to gain the most from trial participation as volunteers, especially the economically weak, would paradoxically be those who are the most discriminated by the requirement of nonexploitation. This short-term discrimination is necessary, however, to protect their long-term interests.

Biomedical research in developing countries is yet another context where the requirement of nonexploitation is especially relevant. Given the accrued risk of exploitation of vulnerable individuals and the potentially huge economic and cultural impact of RCTs in the communities where they take place, the requirement of individual informed consent is not enough to protect the interests of research subjects in this particular context. The European Group on Ethics [31, p. 14] has thus pointed out that "it may be appropriate to seek agreement on the implementation of a research project from persons representative of or invested with a certain authority within the community, or the family." In developing countries, where the capacity of individuals able to consent has been weakened by earlier exploitation or where the principle of autonomy is not widely acknowledged, the consent process must therefore be strengthened by the implication of the larger community. The UK Nuffield Council on Bioethics has also written on the subject and flagged the requirement of nonexploitation as one of three others that should guide biomedical research in developing countries: (1) the duty to alleviate suffering, (2) respect for persons, and (3) sensitivity to cultural differences [25, p. 52].

In most cases, patients in developing countries would receive better medical care while taking part in clinical trials and would therefore generally stand to benefit more from inclusion than exclusion. Seen from an overall benefit perspective, the requirement of nonexploitation could at firsthand discriminate them by placing more obstacles in the way of trial sponsors wishing to conduct such RCTs in developing countries. Such short-term losses would, however, be compensated by creating a better long-term perspectives for individuals as well as developing communities.

The situations detailed above deal with cases where persons able to give consent, and who might wish to participate in a trial, need nonetheless additional protection against possible exploitation at the hands of trial sponsors and collaborating physicians. But what about cases where research might be a condition for treatment? In the United States, the CMS (Centers for Medicare and Medicaid Services) have recently stipulated, for example, that coverage for new users of approved anticancer drugs depends on the willingness of patients to enroll in a clinical trial sponsored by the National Cancer Institute. The CMS policy affects research on tests or treatments

whose efficacy has not been established for the patients being studied. Orentlicher [32, p. 22] argues that such a condition should also extend to research aimed at determining which treatment is most effective when a patient can be offered one of multiple established treatments. Orentlicher's position is motivated by his concern that important studies are being delayed—and medical progress impeded—by difficulties in securing the participation of individuals. In one example he cites, more than half of the patients invited to participate in the study declined with the total recruitment taking twice as long as expected (16 months instead of 8 months).

The Declaration of Helsinki stipulates that participation in any trial is voluntary, that refusal to participate will involve no loss of benefits, and that such participation may be discontinued at any time. Orentlicher's hypothetical protocol would also require voluntary participation and would not deprive the patient of any benefits to which he or she would otherwise be entitled. What he argues for is the individual physician's right to deny care to any patient unwilling to participate in the study. Such a patient would be referred to another physician. The physician–patient relationship is a contractual one and does not oblige the specialized physician to admit any patient who requests medical attention if such medical attention can be received elsewhere and does not represent a danger to the patient in question. Orentlicher argues that trial participation in view of the common good is a valid and ethically justifiable condition for patient care in case the patient can receive equivalent care elsewhere.

Orentlicher's unorthodox position is certainly worth thinking about because it emphasizes that there are societal considerations in any RCT that a strict interpretation of the physician–spatient relationship does not cover. Moreover, in the hypothetical protocol that he proposes, the medical interests of the patients are secured given that the methods being compared are well established and that the patients would be treated by another physician if they declined the offer of participation. However, as we stressed in the beginning, the nature of the physician–patient relationship should be inspired both by the sentiments of friendship and care of the doctor toward his patient and the love for his healing art. It is this *philanthropía* that transcends the terms of different protocols or the provisions of trial designs and that ultimately ethically motivates the physician's *philotekhnía* and hence the possibility of carrying out clinical trials.

## 3 CHALLENGE OF CLINICAL TRIAL EFFICIENCY

Most experts in the field consider that efficiency in clinical development is far too poor. One reason that explains this is the rather simple design imperfections that can still be found in some trial publications. In addition to weeding out these imperfections, trials can be further enhanced by implementing new design methodology and combining them with novel supporting technologies, such as Web-based data capture. The FDA's Critical Path has identified a large number of possibilities that could increase the productivity in different stages of drug development [33, 34]. We will focus here principally on some of the possibilities that relate specifically to the improvement of clinical trial design. After an overview of the current situation (Section 3.1), we will present a DA approach (Section 3.2) to a number of design questions. This paradigm is then applied to the different key

aspects of a trial design. Primarily, we will discuss how each patient can contribute more information (Section 3.3), analyze the optimal size of a trial (Section 3.4), and outline the possibilities that exist to be more adaptive to incoming information (Section 3.5).

### 3.1 Background

The great advances in medicine have contributed significantly to the prolongation of life expectancy during the last century. As part of this, the role of clinical trials during the second half of the century can hardly be overestimated. New pharmaceutical treatments have almost eliminated many diseases and reduced the risks for and consequences of many others.

Trends in productivity are, however, discouraging for drug development. Despite rapidly increasing development costs, the number of approved new drugs has been declining recently. One important reason for the productivity fall is the increased regulatory demands. A further reason is that drug developers try to solve increasingly difficult medical problems. In areas where there are already at least partly effective treatments, a new drug can often not be compared directly to placebo for ethical reasons. To show an effect versus placebo must thus be done indirectly, by testing the new drug and an active comparator in clinical trials, which requires considerably larger sample sizes [26]. Moreover, diseases for which no effective drugs exist are usually much more complicated and difficult to treat.

As many as half of the drugs entering confirmatory studies (phase III) fail [2]. Some drugs, such as the best-selling Vioxx, have also been withdrawn from the market due to safety findings postapproval. Such failures undermine the public trust in medicines and put increased pressure on regulatory agencies. Regulators may react by increasing the demands for safety documentation before drug approval, which will lead to accelerating development costs. A trend can be seen toward a more frequent need for large outcome trials in order to generate enough data to convince regulators of a positive benefit–risk relation. A net result is that, despite huge investments, too few new pharmaceuticals reach the patients in need.

There are a number of hopeful signs, however. First of all, the FDA has clearly recognized the productivity problems and has launched its Critical Path Initiative [33] as a response. This initiative aims at finding innovative ways to improve drug development, in partnership with academia and industry. The Critical Path Opportunity List [34] from 2006 formulates a large number of suggestions, including the use of different biomarkers, innovative/adaptive study designs, and modeling and simulation. It also flags the need to develop methods for indirect comparisons with placebo, in cases where placebo-controlled trials are unethical [26]. Below, we will focus on issues dealing with clinical study design and will not highlight changes in drug discovery technologies and IT per se. It should be noted, however, that the development of information technologies, such as Web-based data capture, is important for the possibility of conducting sequential and adaptive trials in practice.

### 3.2 Decision Analysis Approach to Study Design

As we have noted, the great need for increased efficiency in clinical development is accompanied by new opportunities to improve clinical trial design. A crucial issue

is how to best utilize these possibilities. We will address this challenge by proposing the following systematic, integrated, and quantitative approach:

1. Gather relevant information.
2. Integrate the information into a quantitative model.
3. Define the objectives of the trial.
4. Translate the objectives into a utility function, depending on the trial outcome.
5. Generate a variety of design alternatives.
6. Apply the ethical sift.
7. Simulate the trial outcomes for the interesting designs.
8. Calculate expected utility.
9. Choose the design with maximal expected utility.
10. Check the robustness of the solution to changes in the model assumptions.

The items on this list do not necessarily come in a strict chronological order but are often strongly interrelated. Given that the overall goal is to find a good enough design within the ethical framework, the modeling and DA process will depend on what is needed to accurately answer the design question. Consequently, the proposed approach does not necessarily have to be complicated or time consuming. In many cases, relatively limited work is enough to improve the study design.

This chapter will not describe all the items on the list in depth, as many of them are accurately covered in the existing literature. Much of the literature on modeling is topic specific. A classic text on model building, based in statistics, is Box and coworkers [35]. Pharmacokinetic/pharmacodynamic modeling is described in [36, 37]. Hoppensteadt and Peskin [38] present mathematical modeling for medicine, emphasizing physical processes in physiologic systems but ignoring pharmaceutical interventions. Burman et al. [39] focus on clinical development and provide a shorter, but very similar, version of the list above. Parmigiani [40] treats modeling in medical decision making, taking the perspective of the patient and his treating doctor. This book applies DA and Bayesian methods and describes computer simulations in a relatively nontechnical way. Hunink et al. [41] is an easy read on DA, taking a similar perspective as Parmigiani. It includes a chapter on information gathering (Chapter Clinical Trials Data Management) and notes that subjective components are often needed in the analysis (p. 13 of Chapter "Clinical Trials Data Management"). DA can also be viewed from a management perspective [42] and applied to drug development [43]. A common component of decision analysis and decision theory is Bayesian models. Both Spiegelhalter et al. [44] and Berry [45] provide excellent accounts of Bayesian methods in connection to clinical trials and include sections on ethics. More general texts on DA are Raiffa [46] and the broad overview in Howard and Matheson [47].

Simulations, expected utility calculation, and optimization are quite technical matters but specialized programmers and mathematicians do the job. Literature on random simulations includes [48–50]. Mathematical and numerical analysis can sometimes substitute, and often complement, simulations. Many textbooks cover such areas, including optimization. Although items 7–9 on the list above can be carried out by a sufficiently skilled programmer, the robustness checks should be a

concern to the whole scientific team, in interaction with the programmer. The team, and preferably also experts who have not taken part in the modeling, should change the key assumptions and look at how these changes affect the conclusions. If the solution is very sensitive to the value of a parameter, there is a strong indication that more knowledge should be sought regarding its true value. Renewed information gathering or focused experimentation may be indicated.

Naturally, the utilities may be quite different depending on what perspective you take. From a patient perspective, and in order to comply with the spirit and the letter of the Declaration of Helsinki, placebo should preferably not be used. FDA and European regulations stipulate, however, that a positive effect has to be demonstrated versus placebo, by direct or indirect comparison. Finally, a trial sponsor has to consider the return on investments. Although there are complementing company perspectives, we will take the net present value (NPV) as a proxy for the company utility. Profit maximization from the company's side must, however, be made under the constraints outlined above. Among the possible ethically compatible designs, the company will try to maximize the expected NPV for the project or, more generally, for the company. The value of the project will reflect also the regulatory hurdles and the possibilities of convincing payers about the cost-effectiveness of the drug.

Let us now study the key elements of clinical trial design. The following three subsections will address different aspects of this question.

### 3.3 Optimizing Information per Patient

The general aim of all experiments is to reduce uncertainty. The obvious way to decrease the uncertainty of trial results is to increase its sample size. In Section 3.4, we will apply DA to the problem of finding the optimal sample size. However, the most cost-effective way of retrieving trial information is often to reduce the variance for each patient. There are a number of ways to do this such as devising more than one treatment and several measurements per patient, choosing a more informative response variable, and choosing a patient population with a larger response.

***Covariates and Multiple Measurements*** For many chronic diseases, there is a response variable [e.g., cholesterol level, forced expiratory volume ($FEV_1$), subjective symptoms, or quality of life (QoL)] that fluctuates over time. Technically, we will assume that the variable follows a linear normal model, implying a linear dependence on covariates as well as normally distributed residual terms. The variability in the response variable during treatment has several different causes. It is standard procedure to include the baseline value of the response variable as a covariate in the linear model, together with other potentially predictive covariates. This will reduce the variability due to explainable differences between patients and may lead to considerably higher efficiency of the trial. These covariate adjustments are part of good statistical practice and concern more the analysis than the study design. Some of the other ways of reducing variability require larger modifications of the study design.

A straightforward example of how DA can be used concerns the decision of how many measurements should be taken per patient. One cause of variability that is not handled by covariate adjustments is the day-to-day fluctuations within each

patient. This variability can be reduced simply by taking measurements on different days and using the average of the observations as the response variable. Many studies collect the necessary information without using it in the analysis. Say, for example, that the objective (item 3 on the list in Section 3.2) is to estimate the primary variable in the most cost-effective way as possible. We then proceed by setting up the utility function (item 4). In this case, the cost per information is a function of the number of measurements $M$ per patient. Whether more than one measurement should be taken is essentially a question of how much of the total (unexplained) variability is attributed to day-to-day variability and how much it costs to reduce it. In a straightforward model (item 2), the variance of the average for a patient is proportional to $1 + V/M$, where $V/(1 + V)$ is the fraction of the observation variance that is attributed to day-to-day variability. Assume (item 2 again) that the cost per patient is proportional to $1 + CM$. As information is the inverse of the variance, the cost per unit of information is proportional to $(1 + CM)(1 + V/M)$. The number of measurements, $M$, can easily be chosen (item 9) to minimize this cost/information ratio, given that the value of the parameters can be estimated (modeled) from existing information (item 1).

Say, for example, that the cost per patient is \$11,000 if only one measurement is taken and that the additional cost per extra measurement is \$1000. Assume as well that 50% of the total response variance of 200 units is due to day-to-day variability. The cost and variance per patient are then $10{,}000(1 + 0.1M)$ and $100(1 + 1/M)$, respectively. The cost/information ratio, $1{,}000{,}000(1 + 0.1M)/(1 + 1/M)$, is minimized if $M = 3$ measurements are made. The cost per patient increases by 18% if $M = 3$ compared to $M = 1$, but the variance per patient decreases by as much as one third. Consequently, the sample size can be decreased by one third, from 200 to 133 per arm if a standard error of 1.0 is to be achieved. The cost per unit of information is reduced by 22%; in a two-arm study the total cost decreases from \$4.4 million to \$3.52 million. A robustness check (item 10) indicates that $M = 2$ or $M = 4$ give most of the benefit that is achieved with $M = 3$, and that $M = 2$ is a viable solution for a range of values for the parameters, for example, if the cost per measurement is twice as large as originally assumed.

This example does not indicate that the number of measurements should generally be increased. In other situations the analysis would indicate that costs should be cut by decreasing the number of visits. The point is that the efficiency of a design is often a trade-off between estimation precision and costs. By clearly defining the trial objective, integrating information about variability and economical costs, and quantifying a utility function, the decision problem is clarified. This kind of reasoning is considerably more far-reaching than just deciding on a hunch the number of measurements. In the example above, there could be complications given that the true underlying effect is changing over time, and that there is a nontrivial correlation structure between observations taken on different days. Typically, the effect is increasing toward a steady-state value and the correlation between two observations decrease with the time between the observations. The solution to both these problems is essentially the same: Model the time–effect process and the variability structure and then apply DA to the problem of finding the optimal sample scheme and test variable.

A common way to waste information is to dichotomize data, that is, to replace a continuous response with a zero–one variable in order, for example, to classify

patients as "responders" and "nonresponders" and ignore the rest of the information regarding the response. Occasionally, there are good reasons to classify patients as responders or nonresponders, based on whether a continuous response variable is above or below a specified cut-off value, and base the primary analysis of the constructed zero–one responder variable. However, in most situations, a direct analysis of the original continuous variable gives much more information [51, Section 8.2.4]. If the treatment effect is small, for example, the efficiency gained by using original data is about 57% when the patient population is divided in two approximately equally sized groups based on a normally distributed variable. The efficiency is more than doubled and almost threefold when the dichotomous population split is 80%–20% and 90%–10%, respectively. It is usually possible to model the distribution of the response (item 2), and then the effects of dichotomization are easily investigated by simulating (item 7) the power for the alternative response variables.

***Cross-Over Designs*** The between-patient variability is completely eliminated if the patient can serve as his or her own control. This is the idea behind cross-over trials, where two or more treatments are given to each patient during different periods. Such designs are well studied [52, 53] and have been utilized systematically in some areas for many years, especially in phase I trials of pharmacokinetics and also for patient trials in some specific medical indications. However, there are possibilities of expanding the scope of these designs considerably in the future. When a cross-over design can be applied, the efficiency of the trial can normally be reduced drastically. The number of patients in the trial can often be reduced by considerably more than 50%.

There are some clear limitations to the use of cross-over designs, however. Cross-over trials are only possible for chronic diseases and obviously not suitable to investigate, for example, mortality. Their potential is largest in proof of principle and dose finding and for discriminating against similar competitors. They are best suited when the treatment response is relatively fast (asthma, lipids, some lab values, etc.). Carry-over effects have been seen as an important problem. However, there are in many cases possibilities to handle such potential effects by wash-out periods, statistical corrections, or pharmacological modeling.

When a cross-over design is considered as an alternative to a parallel group design, the usual tools of modeling, simulation, and DA can be of help. Using available information (item 1) in terms of earlier trials with the same and similar drugs, between-patient and within-patient variances can often be estimated or at least guessed. The time–effect curve and potential carry-over effects can be modeled in a similar way (item 2). Based on some models, it is possible to predict (item 7), often via simulations, the likely effects in the planned trial. There will often be trade-offs (items 4 and 9), for example, between the treatment duration and factors such as inclusion time, potential bias, and variability. A cross-over design will probably have a shorter inclusion period, as the sample size is lower, but a longer total treatment time. It is sometimes reasonable to shorten the treatment time per period, although this may lead to the steady-state effect not being fully reached.

Cross-over trials make it possible to assess whether there is a between-patient variability in the relative treatment effect. It may even be that treatment A is better in some patients while other patients are better off on treatment B. Parallel trials cannot assess such individual differences in effect between patients with similar

prognostic variables. One extreme version of the cross-over design is worth mentioning in this context. A trial may be conducted in one single patient, exposed to treatments A and B in a series of treatment periods. This design appears to be quite natural. It resembles what a patient suffering from migraine attacks would do to find out which "treatments" are causing the attacks. She occasionally eats chocolate, drinks wine, goes out in heavy sunshine, and notices when the migraine problems occur. Gradually, she will understand her personal risk factors and avoid them. The design, called *N*-of-1, can be much more rigorous, involving blinding and randomization to two or more treatment groups, comparing two or more pharmaceuticals [52, Section 7.4].

*N*-of-1 trials may be attractive from an ethical perspective when the aim of the experiment is to find the best treatment for the individual patient, and DA is used to find the optimal criterion for when the testing should stop and the patient receive the better treatment henceforth. However, the signal-to-noise ratio is often too poor to allow firm conclusions on the best treatment for an individual. A combination of data from *N*-of-1 trials in several patients can be performed in order to estimate treatment differences more precisely and also estimate how a large percentage of patients would benefit from each treatment.

### 3.4 Optimal Amount of Information to Collect in Trial

Having discussed how to optimize the information gathering for each patient, we will now turn to the problem of choosing the optimal amount of information to purchase, that is, to determine the sample size. Clinical trials are commonly dimensioned by specifying a least clinically relevant effect, taking this as the alternative hypothesis, and then calculating the sample size that gives, say, 90% power of detecting any effect. This approach can be severely misleading, we think, as it ignores all relevant information except the least clinically relevant effect. We will sketch how the knowledge and uncertainty regarding effect and safety, regulatory demands, price, market size, and study costs could all help to find the best sample size. However, let us first analyze the notion of "the least clinically relevant effect." If you have to choose between two drugs with known different effects but identical safety profile, price, dosing, tablet size, and the like, which would you choose? If your 1-year risk of dying is 1% with drug A and 1.1% with drug B, you would clearly choose A. Will the choice be different if the risk is instead 1.001% with drug B? Our conclusion is that all differences are "clinically relevant." The question about which drug to prefer arises only if both drugs have relative advantages in different aspects. For the patient, there can be a trade-off between effect and safety, for the payor between effect (and safety) and cost.

For the drug developer that tries to maximize the expected net present value (ENPV), study sizing is a trade-off between generating enough information regarding effect and safety on one hand and the cost of the trial on the other. A related and even more fundamental problem is whether any investment at all should be made in developing the drug or if the project should be terminated immediately. It is very complex to optimize a complete clinical development program, as the optimal design of a later phase depends on the data from the earlier phase, and the optimal design of the earlier phase depends on how the results from this phase will be used for designing the later phase. The principle solution to the problem is backward

induction [54], where one starts by analyzing the last phase, given different possible outcomes of earlier phases. After finding how phase III should be designed as a function of the information from phase IIB, phase IIB can be optimized. This process can in principle be carried back to even earlier phases, although computational difficulties and specification of the complex problem may be obstacles.

Consider the problem of dimensioning a phase III trial, assuming that supportive evidence exist and that one successful phase III trial will be enough to get the drug approved. Generally, the utility, as a function of the sample size $N$ of the trial, is the gain by running the trial minus the cost for doing so. The cost can often be put rather directly in monetary terms. However, the personnel resources may be limited and hard to expand rapidly. Consequently, one may have to consider that if the phase III trial is run, other projects may be delayed or stopped and the net consequences of this should be factored into the cost analysis. As the sample size affects the possible time to market introduction, the time delay may be viewed as a cost. Time effects are, however, related to the trial outcome and are easier to analyze as part of the gain function.

The gain can be complex to analyze, although some market value calculations are routinely done in current practice. The simplest model is to say that the drug is approved if and only if a statistically significant effect versus placebo is shown in phase III, and if it is approved the commercial value is a certain known amount. This model can easily be improved by modeling how the market value will decrease with a delay in study ending time as a result of an increased sample size. There are, however, other factors, such as the market value's dependence on the trial results, that are harder to model. It is often too simplistic, although common, to assume a fixed market value if the drug is approved without considering how good the trial data are. One way of modeling the sales of a drug is to start by analyzing the decision analyses that the patient is doing together with his or her physicians. What is required to choose the new drug in terms of estimated effect, standard error for the effect estimate, safety, price, and the like? Pharmaceutical companies often perform limited market research about such questions by asking physicians to compare different hypothetic drugs. Individual treatment decisions could then be aggregated to form an effect–sales model. It is likely that historical data on, for example, how fast a market is penetrated have to be added in order to find a suitable model.

The commercial and cost models together give a utility function. The design alternatives are all possible sample sizes, including the possibility of stopping the project and not run the trial. Regulators may have requirements for a minimal sample size in order to get enough safety information. With a model for the effect of the drug and for how large the uncertainty is, the trial results can be predicted, by numerical methods or simulations. It is often important to consider the uncertainty in the effect estimate and run Bayesian clinical trial simulations. Combining predicted trial results with the utility function, the expected utility can be calculated for each sample size, and the optimal sample size can be found. The expected utility is relatively flat around the optimum, indicating that the exact sample size is not crucial. The ENPV can be very sensitive to such changes in assumptions concerning costs and gains. However, given that it is optimal to run the trial, the sample size is typically relatively robust against such changes, meaning that the same sample size often gives an ENPV close to the optimal ENPV.

The time to market is essential. If there is an indication that the sample size should be increased, one might therefore consider including a larger number of centers. In analyzing this option, both direct costs and the risk for reduced quality, leading, for example, to a larger unexplained variability, should be considered. The optimal size of an experiment is discussed by Lindley [55] and a review on applications to clinical trials is provided by Pezeshk [56].

***Commercial Limitations*** Medical research will likely look for smaller and smaller improvements in many therapeutic areas. Smaller effect differences than what most historic studies have set out to detect may still be of great importance. The vast majority of mortality trials have been dimensioned to detect risk reductions of 20% or more. However, a treatment with 10% additional reduction in, for example, cardiovascular mortality, would of course be of great value. The problem is that smaller effects call for larger trials. The sample size is, at least approximately, proportional to the square of the relative treatment effect. Thus, looking for a risk reduction of 10% instead of 20% will mean that four times more patients are needed in the trial. While the cost for a trial is approximately quadratic in the effect size, the benefit, in the number of spared lives, say, is linear. Thus, small enough relative effect sizes will not be found by clinical trials, although they may translate into a large number of deaths.

Assuming rational payors, the commercial gain of introducing a new pharmaceutical will always be lower than the societal gain, as the price will always be lower, at least for some payors, than the willingness to pay. There are, consequently, research projects that would give a good expected return-on-investment for society, although they would never be run by commercial interest. Noncommercial research therefore has a clear value.

### 3.5 Adapting to New Information

For both economic and ethical reasons, it is clearly desirable to adapt the design of a clinical trial in response to accumulating data. Interim data may show, for example, that there is a clear difference between the treatments. The trial can then be stopped, permitting the patients in the control arms of the study to stop receiving the inferior treatment. Alternatively, interim data may indicate that a proposed treatment has no clear benefit over the placebo. The trial can in such a case also be stopped for reasons of futility.

***Group-Sequential Designs*** The most well-known example of adaptive designs is the classic group-sequential design [57, 58], which permits a number of interim analyses to be conducted. It allows for the trial to be stopped if interim results are convincing enough according to preset criteria. As a statistically significant result can be obtained both at interim analyses and at the preplanned full sample size, the critical limits for when to claim significance must be adjusted. The result is that a group-sequential trial has to have a somewhat larger maximal sample size than a fixed-size trial if the same power should be obtained. This is the price to pay for the possibility that the trial sometimes can be stopped much earlier than in a trial without interim analyses. With a reasonable group-sequential design, the expected sample size is decreased. Consequently, the cost related to the number of patients

is on average decreased. Among the drawbacks of a sequential design are the cost for the interim analysis per se. Given these costs, it is straightforward to simulate trials and calculate the expected total cost to find which design is best. One can investigate whether it is optimal to include interim analyses, when they should come and what the critical limits for significance should be. There are some additional drawbacks, such as problems with overrunning, that is, that additional trial data may be obtained after an interim analysis that suggested a trial termination. Such data may potentially change the conclusion in the trial. Still, it is possible to simulate this problem and factor it into the DA.

***Flexible Designs*** From a DA perspective, the optimal design of a trial may change during a trial if external information is received that changes the utility function or the effect/safety model. Examples of such information is the approval of a competitor drug or the finding of unexpected safety problems in substances with similar chemical properties as the one being developed. It is therefore desirable to be able to change the design during the conduct of a trial. An article by Bauer and Köhne [59] in 1994 gave the impetus to research in flexible designs that allow several types of design modifications in response to interim data or external information. According to Müller and Schäfer [60], "design changes may include increase or reduction of the sample size and of the number and time points of interim analyses, and even changes of the type of statistical test, of the outcome variable, or of the null-hypothesis." Their stand is that such changes do not have to be preplanned. However, if such flexible designs are getting acceptance and penetrate clinical trial practice, it will constitute a fundamental change compared to the hitherto dominating methodology. Preplanning has always been one of the cornerstones of controlled clinical trials and a basis for the classical statistical analysis of data. Other authors, including Bauer and Köhne [59] in their original work, have been more moderate than Müller and Schäfer [60] regarding when and how flexible designs could be used. The EU reflection study [61] on adaptive designs is quite restrictive to flexible design changes, especially in confirmatory studies.

Sample size reestimation (SSRE) has attracted the greatest interest among the flexible designs. The sample size may, for example, be increased in order to achieve more statistical power if the interim estimate of the treatment effect is lower than anticipated. The fundamentally new technical idea behind flexible designs is to tentatively consider different stages of a trial as different trials altogether. One $p$ value is calculated for each stage in isolation. These $p$ values are then combined into a $p$ value for the complete trial, by weighting together the stagewise $p$ values. The type I error can be guaranteed, for essentially all kinds of design modifications, as long as the weight for each stage is determined independently of the data from that stage.

The weighted inference is controversial. Jennison and Turnbull [58] have suggested that an adequately developed group sequential design is preferable to an adaptive SSRE that only depends on interim data. The flexible design has, however, the advantage of being able to be modified according to information external to the trial. Another problem with the methodology proposes by Bauer and Köhne [59] is that equally informative observations may receive unequal weight in the analysis, depending on which stage of the trial they are collected. This feature has been criticized for violating the sufficiency principle, leading not only to inefficiency [58] but also in extreme cases to erroneous conclusions [62].

Our tentative conclusions are that careful preplanning of trials is crucial, that there can occasionally still be a genuine need for flexibility, that the weighted test is questionable but that a modification of this test could be of value. The works on flexible designs mentioned above have been based on classical, frequentist statistics. Bayesian statistics can more naturally incorporate sample size modifications and other design changes. Such modifications can be useful in earlier phases but may not be accepted in the confirmatory phase.

## 4 DISCUSSION AND CONCLUSIONS

Since World War II, the number of clinical trials has increased rapidly. The importance of these trials can hardly be underestimated. The medical needs continue to be manifold. Despite the clinical needs and the steady increase in the number of clinical trials carried out, there are factors that could threaten their future. In this chapter, we have highlighted the problems concerning ethics and efficiency. Although RCTs represent the gold standard in clinical development, there are alternatives. Observational studies and historic controls can give valuable information, even if their interpretation is not as straightforward as blinded clinical trials. Preclinical experiments and modeling can also substitute some trials.

Clinical trials should be performed in an ethical manner. Failure to do so is not only unethical, per definition, but it also jeopardizes public trust in clinical development. A patient is unlikely to give informed consent unless he or she is convinced that the physician will put the patient's interests first. The concept of informed consent is a fundamental requirement that we have highlighted. In the rare situations where informed consent cannot be given by the patients themselves (studies on unconscious patients or children), the informed consent should instead be given by a guardian *on behalf of and in the interest of* the patient as soon as possible. Likewise, the physician should see principally to the best of his or her patient first and only secondarily to the interests of science. Thus, when the physician also acts as an investigator, he or she is still personally responsible for ensuring that the patient's best interests are safeguarded during trial participation. In addition to genuine consent and patient benefit optimization, an ethical requirement of nonexploitation is needed. Additional requirements may be indicated to cover issues about, for example, payment and trials that are beneficial on average but include nonoptimal treatment arms. While patient enrollment is the responsibility of the physician, the ethical standard of clinical trial is the common responsibility of the investigator, the ethical committee, the sponsor, and the personnel responsible for the planning and the execution of the trial. Personal responsibilities should be stressed to avoid the risk that all actors put the responsibility on someone else, leading to unethical and risky trials.

Ethical requirements constrain trial design, in particular the treatments that can be tested. Depending on the situation and background information, these constraints will sometimes complicate or even obstruct clinical development. Within these limits there are, however, considerable possibilities of optimizing designs that can also be efficient. Several features of a trial design should be considered in the process of trial optimization, such as sample size, patient population, and treatment duration. There are also great possibilities to apply alternatives to the standard

parallel-group nonadaptive design and there is a continuing methodological development. How to choose between all the available options is perhaps the major challenge. Consequently, we anticipate that the most significant changes in trial design will take place in decision process rather than design methodology per se. Model-based drug development is already receiving considerable attention. Still, modeling and DA are clearly underestimated.

## REFERENCES


1. DiMasi, J. A., Hansen, R. W., and Grabowski, H. G. (2003), The price of innovation: New estimates of drug development costs, *J. Health Econom.*, 22, 151–185.
2. Kola, I., and Landis, J. (2004), Can the pharmaceutical industry reduce attrition rates? *Nature Rev. Drug Dis.*, 3, 711–716.
3. Santoro, M. A., and Gorrie, T. M., eds. (2005), *Ethics and the Pharmaceutical Industry*, Cambridge University Press. Cambridge, UK, pp. 3–4.
4. Emanuel, E. J., Wendler, D., and Grady, C. (2000), What makes clinical research ethical? *JAHA*, 283, 2701–2711.
5. Senn, S. (2002), Ethical considerations concerning treatment allocation in drug development trials, *Stat. Methods Med. Res.*, 11, 403–411.
6. Purtilo, R. (1995), Professional-patient relationship: Ethical issues, in *Encyclopedia of Bioethics*, vol. 4, Macmillan, New York, pp. 2094–2101.
7. Mason Spicer, C. (1995), Codes, oaths, and directives related to bioethics. Section II. Ethical directives for the practice of medicine, in *Encyclopedia of Bioethic*, vol. 4, Macmillan, New York, pp. 2630–2649.
8. Thomasma, D. (1990), Establishing the moral basis of medicine: Edmund D. Pellegrino's philosophy of medicine, *J. Phil. Med.*, 15, 245–267.
9. Laín Entralgo, P. (1969), *El Médico y el Enfermo*, Ediciones Guadarrama, Madrid.
10. Hippocrates (1984), *Parangeliai (Recommendations, Precepts) 2 Hippocrates 1*, Loeb Classical Library, Harvard University Press, Cambride, MA.
11. Lellouch, J., and Schwartz, D. (1971), L'essai thérapeutique: Éthique individuelle ou éthique collective? *Rev. Inst. Int. Stat.*, 39, 127–136.
12. Clayton, D. G. (1982), Ethically optimised designs, *Br. J. Clin. Pharma.*, 13, 469–480.
13. Gifford, F. (1986), The conflict between randomized clinical trials and the therapeutic obligation, *J. Med. Phil.*, 11, 347–366.
14. Engelhardt, H. T. (1996), *The Foundations of Bioethics*, 2nd ed., Oxford University Press, New York.
15. Nuffield Council on Bioethics (1995), *Human Tissue: Ethical and Legal Issues*, Nuffield Council, London.
16. Schaffner, K. F. (1986), Ethical problems in clinical trials, *J. Med. Phil.*, 11, 297–315.
17. Kopelman, L. (1986), Consent and randomized clinical trials. Are there moral or design problems? *J. Med. Phil.*, 11, 317–346.
18. Lilford, R. J., Braunholtz, D., Edwards, S., et al. (2001), Monitoring clinical trials—interim data should be publicly available, *BMJ*, 323, 441–442.
19. Fried, C. (1974), *Medical Experimentation: Personal Integrity and Social Policy*, American Elsevier, New York.
20. Freedman, B. (1987), Equipoise and the ethics of clinical research, *N. Engl. J. Med.*, 317, 141–145.

21. Lilford, R. J. (2003), Ethics of clinical trials from a bayesian and decision analytic perspective: Whose equipoise is it anyway? *BMJ*, 326, 980–981.
22. Emanuel, E. J., and Miller, F. G. (2002), The ethics of placebo-controlled trials—a middle ground, *Am. J. Ophthalmol.*, 133, 174–175.
23. World Medical Association (2008), Declaration of Helsinki: Ethical principles for medical research involving human subjects, available at: http://www.wma.net/e/policy/b3.htm; accessed February 2009.
24. Carlson, R. V., Boyd, K. M., and Webb, D. J. (2004), The revision of the Declaration of Helsinki: Past, present and future, *Br. J. Clin. Pharmac.*, 57, 695–713.
25. Nuffield Council on Bioethics (2002), *The Ethics of Research Related to Healthcare in Developing Countries*, Nuffield Council, London.
26. D'Agostino, R. B., Massaro, J. M., and Sullivan, L. M. (2003), Non-inferiority trials: Design concepts and issues of the encounters of academic—consultants in statistics, *Stat. Med.*, 22, 169–186.
27. Retsas, R. (2004), Treatment at random: The ultimate science or the betrayal of Hippocrates? *J. Clin. Oncol.*, 22, 5005–5008.
28. Marquis, D. (1986), An argument that all prerandomised clinical trials are unethical, *J. Med. Phil.*, 11, 367–383.
29. Bureau of Labor Statistics (2005), News: National census of fatal occupational injuries in 2004. Bureau of Labor Statistics, U.S. Department of Labor, Washington D.C., August 25; available at: http://www.bls.gov/news.release/pdf/cfoi.pdf; accessed May 2006.
30. Council of Europe (2005), Additional protocol to the convention on human rights and biomedicine, concerning biomedical research. Council of Europe Treaty Series, No 195; available at: http://conventions.coe.int/; accessed May 2006.
31. European Group on Ethics (2001), Ethical Aspects of Clinical Research in Developing Countries; available at: http://ec.europa.eu/european_group_ethics/docs/avis17_en.pdf; accessed May 2006.
32. Orentlicher, D. (2005), Making research a requirement of treatment. Why we should sometimes let doctors pressure patients to participate in research, *Hastings Center Rep.*, 35/5, 20–28.
33. FDA (2004), Challenge and opportunity on the critical path to new medical products; available at: http://www.fda.gov/oc/initiatives/criticalpath/whitepaper.html; accessed May 2006.
34. FDA (2006), Critical Path Opportunities List; available at: http://www.fda.gov/oc/initiatives/criticalpath/reports/opp_list.pdf; accessed May 2006.
35. Box, G. E. P., Hunter, W. G., and Hunter, J. S. (1978), *Statistics for Experimenters. An Introduction to Design, Data Analysis and Model Building*, Wiley, New York.
36. Gabrielsson, J., and Weiner, D. (1997), *Pharmacokinetic and Pharmacodynamic Data Analysis: Concepts and Applications*, 3rd ed., Swedish Pharmaceutical Press, Stockholm.
37. Sheiner, L. B., and Steimer, J.-L. (2000), Pharmacokinetic/pharmacodynamic modeling in drug development, *Ann. Rev. Pharma. Toxicol.*, 40, 67–95.
38. Hoppensteadt, F. C., and Peskin, C. S. (2000), *Modeling and Simulation in Medicine and Health Sciences*, 2nd ed., Springer, New York.
39. Burman, C.-F., Hamrén, B., and Olsson, P. (2005), Modelling and simulation to improve decision-making in clinical development, *Pharma. Stat.*, 4, 47–58.
40. Parmigiani, G. (2002), *Modeling in Medical Decision Making: A Bayesian Approach*, Wiley, Chichester.
41. Hunink, M. G. M., Glasziou, P. P., Siegel, J. E., et al. (2001), *Decision Making in Health and Medicine: Integrating Evidence and Values*, Cambridge University Press, Cambridge, UK.

42. Goodwin, P., and Wright, G. (1998), *Decision Analysis for Management Judgment*, 2nd ed., Wiley, Chichester.
43. Burman, C.-F., Grieve, A., and Senn, S. (2006), Decision analysis in drug development, in Dmitrienko, A., Chuang-Stein, C., D'Agostino, R., Eds., *Pharmaceutical Statistics*, SAS Institute, Cary, NC.
44. Spiegelhalter, D. J., Abrams, K. R., and Myles, J. P. (2004), *Bayesian Approaches to Clinical Trials and Health-Care Evaluations*, Wiley, Chichester.
45. Berry, D. A. (1993), A case for Bayesianism in clinical trials (with discussion), *Stat. Med.*, 12, 1377–1404.
46. Raiffa, H. (1968), *Decision Analysis: Introductory Lectures on Choices under Uncertainty*, McGraw-Hill, New York.
47. Howard, R. A., and Matheson, J. E., eds. (1983), *Readings on the Principles and Applications of Decision Analysis. Volume I–II*, Strategic Decision Group, Menlo Park, California.
48. Ross, S. M. (1997), *Simulation*, 2nd ed., Academic. San Diego.
49. Robert, C. P., and Casella, G. (1999), *Monte Carlo Statistical Methods*, Springer, New York.
50. Thompson, J. R. (2000), *Simulation: A Modeler's Approach*, Wiley, New York.
51. Senn, S. (1997), *Statistical Issues in Drug Development*, Wiley, Chichester.
52. Senn, S. (1993), *Cross-Over Trials in Clinical Research*, Wiley, Chichester.
53. Jones, B., and Kenward, M. G. (2003), *Design and Analysis of Cross-Over Trials*, 2nd ed., Chapman & Hall, Boca Raton, FL.
54. Bather, J. (2000), *Decision Theory: An Introduction to Dynamic Programming and Sequential Decisions*, Wiley, Chichester.
55. Lindley, D. V. (1997), The choice of sample size, *Statistician*, 46, 129–138.
56. Pezeshk, H. (2003), Bayesian techniques for sample size determination in clinical trials: A short review, *Stat. Methods Med. Res.*, 12, 489–504.
57. Whitehead, J. (1997), *The Design and Analysis of Sequential Clinical Trials*, 2nd ed., Wiley, Chichester.
58. Jennison, C., and Turnbull, B. W. (2000) *Group Sequential Methods with Applications to Clinical Trials*, Chapman and Hall, London.
59. Bauer, P., and Köhne, K. (1994), Evaluation of experiments with adaptive interim analyses, *Biometrics*, 50, 1029–1041.
60. Müller, H.-H., and Schäfer, H. (2004), A general statistical principle for changing a design any time during the course of a trial, *Stat. Med.*, 23, 2497–2508.
61. EMEA/CHMP (2006), Reflection paper on methodological issues in confirmatory clinical trials with flexible design and analysis plan; available at: http://www.emea.eu.int/pdfs/human/ewp/245902en.pdf; accessed May 2006.
62. Burman, C.-F., and Sonesson, C. (2006), Are flexible designs sound? *Biometrics*, 62, 662–669.

# *Proof-of-Principle/ Proof-of-Concept Trials in Drug Development*

**Ayman Al-Shurbaji**
*Experimental Medicine, International PharmaScience Center, Ferring Pharmaceuticals A/S, Copenhagen S, Denmark*

**Contents**

1 Background
  1.1 Why Is Proof-of-Principle/Proof-of-Concept (PoP/PoC) Crucial in Drug Development?
  1.2 Definitions
  1.3 When Should PoP/PoC Be Established?
2 Constantly Improving Toolbox for PoP/PoC
3 Biomarkers in PoP/PoC Trials
4 Pharmacogenomics, Proteomics, and Metabonomics
5 Imaging
6 Pharmacological and Nonpharmacological Models for PoP/PoC Trials
  6.1 Pharmacological Models in Healthy Volunteers
  6.2 Nonpharmacological Models
  6.3 Provocation Models in Patients
7 Different Types of PoP/PoC Studies
  7.1 Exploratory Investigational New Drug PoC Trials
  7.2 Clinical Trials Using Biomarkers or Models in Healthy Subjects
  7.3 Clinical Trials Using Biomarkers in Patients
  7.4 Trials Using Clinical Endpoints for PoC

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

8 Knowledge of Exposure–Response Relationship of Drug Is Essential for Successful PoP/PoC
9 PoP/PoC Outcomes Can Be "False Dawns"
9.1 Summary of Selected Human Models Quoted in This Chapter That Can Be Used in PoP/PoC Trials
References

## 1 BACKGROUND

### 1.1 Why Is Proof-of-Principle/Proof-of-Concept (PoP/PoC) Crucial in Drug Development?

Drug development is complex, costly, and time consuming. After decades of fruitful endeavors leading to novel therapies that radically changed the way we treat human disease (e.g., β-blockers, proton pump inhibitors, SSRIs, HMG-CoA reductase inhibitors, and many others), the pharmaceutical industry began to face increasing and significant challenges. Everyone who is involved in drug development is well aware of issues such as the "innovation gap," "productivity decline," and "late-stage attrition." Indeed, these issues have been on the agenda for some years now and have triggered waves of new paradigms in drug development. Essentially, the pharmaceutical industry has been investing more and more resources in developing new drugs, but this has not been translated into more successful medicines, that is, medicines that are approved by the regulators and well-received by the market. In other words, the output from the R&D machinery has not kept pace with spending. This productivity decline is illustrated in Figure 1 showing escalating drug development costs with no parallel increase in the number of regulatory approvals. The average cost of bringing a new compound to the market is in the neighborhood of $1.3 billion [1] and the overall clinical success rate is 11% [2]. A nearly 50% attrition rate in phase III is still plaguing pharmaceutical companies [2]; and there is no clear evidence that it is on the decline. Late-stage failures are severe blows that hurt large pharmaceutical companies and might ruin smaller ones, and it is in this context that the importance of early PoP/PoC becomes apparent. Drug developers need to ascertain that only compounds with a high chance of success are selected and progressed through the development process. Although most pharmaceutical companies are trying to adopt "kill early" approaches, this is often hampered by the lack of reliable "killer experiments" and the fear of erroneously discarding valuable assets. For smaller companies such as many in the biotechnology sector, which often lack the resources to take compounds through full clinical development, a positive PoP/PoC significantly improves the chance of finding a partner or successfully licensing out the compound. Regardless of company size, a robust early PoP/PoC has become a key component in contemporary drug development.

### 1.2 Definitions

A human PoP/PoC trial is simplistically a trial that provides indications that the compound might work in the intended therapeutic area. Some companies distinguish between a PoP trial in which the pharmacological principle has been

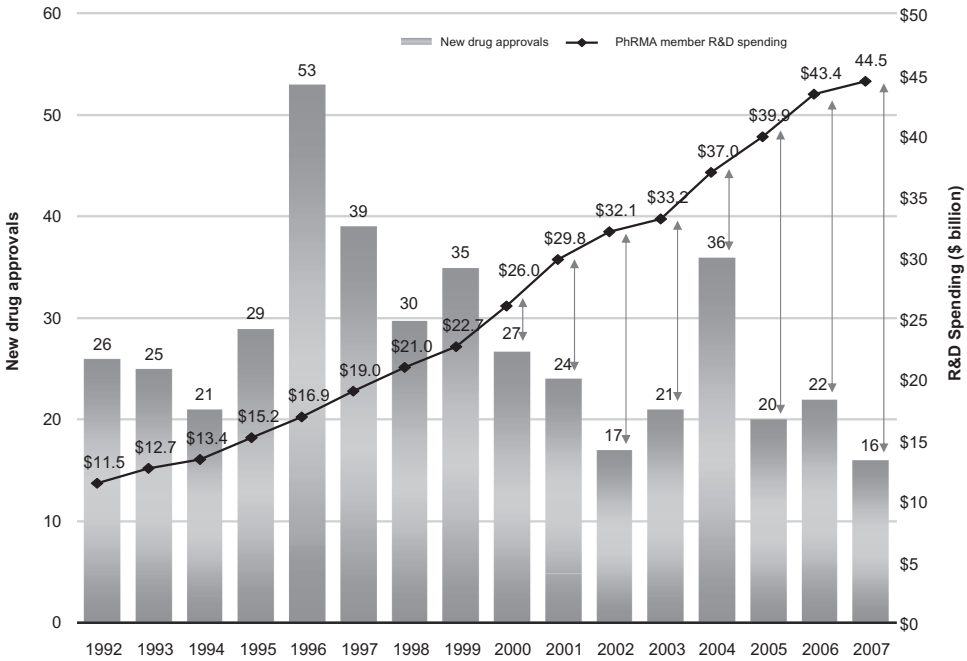


***FIGURE 1*** *Escalating R&D spending with no increase in output (FDA drug approvals): 2004–2007 approvals include new biologics. (Source: FDA and Pharmaceutical Research and Manufacturers of America, Pharmaceutical Industry Profile 2008, PhRMA, Washington, D.C. March 2008).*

demonstrated in humans (healthy subjects or patients), usually with the use of a biomarker and a clinical PoC trial where early signs of efficacy are detected, based on solid clinical endpoints or validated surrogates. However, PoP and PoC are often used interchangeably. A broad definition has been proposed by fifth EUFEPS Conference: "A human trial that provides scientifically sound evidence supporting the postulated effects of a new therapeutic drug product, where effects may be relevant pharmacological action or a change in disease biomarkers, established surrogate endpoints, or clinical outcomes, and be beneficial and/or toxic in nature" [3].

## 1.3 When Should PoP/PoC Be Established?

As indicated above, spiralling development costs and high attrition rates are forcing drug developers to search carefully for signs of efficacy and potential safety concerns as early in the development process as possible. It is common practice today to conduct a PoP/PoC trial as soon as phase I trials have provided evidence that the compound appears be to reasonably safe in healthy subjects and possesses pharmacokinetic characteristics that are in line with projected clinical use. As shown in Figure 2, a PoP/PoC trial is conducted as a part of phase II.

The PoP/PoC is located at a most crucial "go–no go" decision point in most pharmaceutical R&D organizations. Once PoP/PoC has been established, drug developers (and investors) feel more confident to invest in the subsequent, more costly clinical trials.

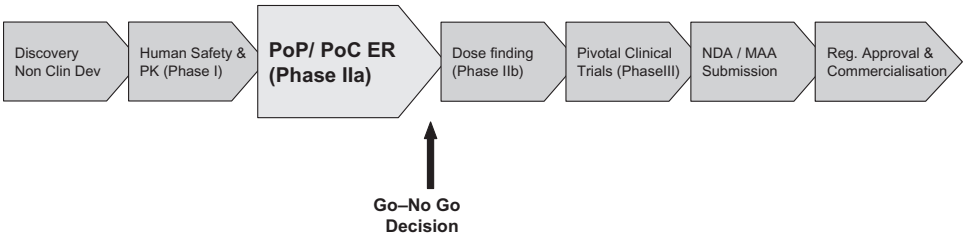


***FIGURE 2*** *Location of PoP/PoC in the drug development process. ER, Exposure–response; NDA, new drug application; MAA, marketing authorization application.*

Although PoP/PoC is often linked to a single trial that provides a dichotomous "yes or no" answer, it is more prudent to think of PoP/PoC as a package of key data that have emerged during the early exploratory phase of development. Thus, a positron emission tomography (PET) study in healthy volunteers may provide evidence that a certain compound crosses the brain–blood barrier and binds to a specific receptor in the brain. A human model (Table 3) may show that the compound is effective in a disease-like setting. Pharmacokinetic/pharmacodynamic (PK/PD) modeling of preclinical and human data is used to characterize the exposure–response (ER) relationship and ensure that safe doses are potentially efficacious. Finally, the integrated data package from all of the above provides a reasonable level of confidence to pursue development, that is, a PoP/PoC-based "go decision."

## 2 CONSTANTLY IMPROVING TOOLBOX FOR POP/POC

Despite the serious issues facing the pharmaceutical industry, the scientific advances in biosciences (chemical and imaging biomarkers, pharmacogenetics and pharmacogenomics, proteomics, toxicogenomics, metabonomics) are providing drug developers with useful tools to explore the potential clinical utility of new targets and novel therapeutic agents. These areas of research have become independent disciplines and there is a wealth of literature on each of them. Since biomarkers constitute a key component in PoP/PoC paradigms, they are discussed in some detail next.

## 3 BIOMARKERS IN POP/POC TRIALS

The National Institutes of Health (NIH) Biomarkers Definitions Working Group has defined a biomarker as "a characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention" [4]. Although biomarkers have been used in drug development for decades, the new potential of using biomarkers for regulatory purposes (label claims) [5] has intensified the efforts to develop biomarkers that are accepted for ordinary and accelerated approvals. Regulatory initiatives, most notably the Food and Drug Administration's (FDA's) Critical Path Initiative have opened new avenues for collaboration between academia,

**TABLE 1 Examples of Biomarkers Accepted by FDA as Surrogate Endpoints**

| Condition | Surrogate |
|---|---|
| Ischemic heart disease | Plasma cholesterol |
| Stroke | Blood pressure |
| Osteoporosis | Bone density |
| Diabetes mellitus | HbA1c |
| Prostate cancer | Serum testosterone |
| HIV/AIDS | HIV-RNA load and $CD4^+$ cell count |

pharmaceutical companies, and regulators aimed at facilitating innovative approaches in the development of novel medicines [6]. A key area in the Critical Path Initiative is biomarkers with a special consortium focusing on the qualification of biomarkers for regulatory decision making. Qualification of biomarkers refers to the process of linking a biomarker with biological processes and clinical endpoints [7, 8]. The highest level of qualification of a biomarker is surrogacy, that is, when the biomarker can substitute for a clinical endpoint [7, 9]. Table 1 shows examples of biomarkers that have been accepted by the FDA as surrogate endpoints.

It should be underscored that while surrogate endpoints and other "valid" biomarkers can and should be used in PoP/PoC trials when available, early clinical development is an exploratory phase, and less well-qualified biomarkers can provide valuable information for internal company decision making. For example, "distal biomarkers"—relative to the clinical endpoint—can be very useful in confirming the pharmacological activity of a given compound in humans [10]. This is especially valuable in two situations: the selection of follow-up or backup compound when the "pathfinder" has demonstrated clinical utility; and the validation of a novel mechanism of action with an uncertain role in the pathogenesis of the disease. Lack of clinical efficacy despite adequate pharmacological effect would disqualify the mechanism of action and stop the development of similar compounds. Also, biomarkers can be very useful in establishing the ER relationship preclinically and translating that relationship into the clinical phases of development (the concept of translational science), facilitating dose selection in PoC studies and beyond [11]. Good examples of the use of biomarkers in early clinical development have been described by Kuhlmann [12]. These included the use of ex vivo inhibition of cholesteryl ester transfer protein (CETP) in healthy subjects to establish the pharmacological activity of a CETP inhibitor, the use of leukotriene D4-induced bronchoconstriction to establish PoP and dose–response in healthy subjects for a leukotriene receptor antagonist for asthma, and the use of pupillography as a marker for effects of 5-hydroxytryptamine 1A (5-HT1A) compounds.

## 4 PHARMACOGENOMICS, PROTEOMICS, AND METABONOMICS

The concept of identifying genomic [single nucleotide polymorphisms (SNP), haplotypes, ribonucleic acid (RNA) expression signatures], proteomic [protein profiles in plasma, cerebrospinal fluid (CSF), synovial fluid, etc.], and metabonomic (metabolite patterns in body fluids) markers that identify patients who are likely to respond to treatment and/or to develop serious adverse drug reactions is very exciting albeit

**TABLE 2 Examples of Valid Genetic/Genomic Markers Predicting Clinical Response or Specific ADRs**

| Marker | Drug | Clinical Relevance |
|---|---|---|
| Her2/neu overexpression | Trastuzumab | Only patients with tumors overexpressing Her2 are eligible for treatment. |
| Philadelphia chromosome (Ph1) | Busulfan | Poor response in patients with AML who lack Ph1 chromosome. |
| PML/PAR alpha gene expression | Tretinoin | Clinical response to tretinoin in patients with APL have not been observed in the absence of the genetic marker. |
| Thiopurine methyltransferase (TPMT) variants | Azathioprine | Increased risk of myelotoxicity in patients with low or absent enzyme activity. |
| HLA-B*1502 allele | Carbamazepine | Risk for serious dermatologic reactions in patients positive for the allele. |
| G6PD deficiency | Rasburicase | Drug is contraindicated in patients with G6PD deficiency due to the risk of severe hemolysis. |

AML, acute myelogenous leukemia; APL, acute promyelocytic leukemia.

extremely challenging. There are few examples of valid genomic markers that predict clinical outcomes [responder vs. nonresponders; adverse drug reaction (ADR) prone genotypes] and that have been accepted by the FDA as part of the label (Table 2). A comprehensive list of valid genomic biomarkers in the context of FDA-approved drugs is available [13].

However, in the context of early clinical development of novel compounds acting on nonvalidated targets, the identification of such markers remains a major hurdle to overcome. Roses and colleagues described some examples of the use of early pipeline genetics to predict efficacy and side effects of new compounds [14]. These included the identification of gene variant candidates related to efficacy in a phase II obesity drug trial, and the early discovery of the genetic basis (CYP 2C9 polymorphisms) of some of the side effects of the kinase inhibitor lapatinib (Tykerb). The authors rightly underscored that such early findings can only be used to generate hypotheses and that the validation of these markers should await larger clinical trials [14]. The limitations of pharmacogenomic predictor discovery in early clinical trials are discussed in Pusztai [15].

Needless to say, this field is still in its infancy, and the ultimate impact of the "omics revolution" on expediting drug development and reducing attrition rates remains to be seen.

In many pharmaceutical companies, default banking of samples from clinical trials for future genetic/genomic analyses is becoming common practice, although this is sometimes challenged ethically and legally. Since genetically determined variability in drug metabolism, most notably involving CYP 450 enzyme system, impacts significantly on the pharmacokinetics and consequently pharmacodynamics of many drugs, it is prudent to obtain relevant genotypes in healthy volunteers and patients in early clinical trials of compounds known to be mainly metabolized by CYP 450 isoenzymes, primarily CYP 2D6, CYP 2C9, and CYP 2C19. This would be useful if unexpected PK/PD profiles are observed and potentially reduce the risk of selecting inappropriate doses in PoP/PoC trials resulting in erroneous go–no go decisions.

## 5 IMAGING

Advances in medical imaging present a great opportunity in drug development. A number of different imaging technologies are available. These include computed tomography (CT), magnetic resonance imaging (MRI), magnetic resonance spectroscopy (MRS), positron emission tomography (PET), and single photon emission computed tomography (SPECT). If adequately qualified, imaging biomarkers can be very helpful in the early phases of clinical development. Early evidence of biological/pharmacological activity of a new compound as well as useful pharmacokinetic profiles for future dose selection can be obtained. For example, using fluorodeoxyglucose PET (FDG-PET), it could be shown that glucose metabolism was reduced in gastrointestinal stromal tumors in patients treated with imatinib mesylate (Gleevec), indicating an early response that preceded tumor size reduction (on CT) by weeks [16]. The development of a new antidepressant was terminated based on the finding that the compound lacked specific receptor binding, as shown by cerebral PK imaging [17]. Other good examples of the use of imaging in early clinical trials in different therapeutics are described by Pien et al. [18]. The potential advantages of imaging techniques in the different phases of drug development have recently been reviewed by Strack [19].

## 6 PHARMACOLOGICAL AND NONPHARMACOLOGICAL MODELS FOR POP/POC TRIALS

### 6.1 Pharmacological Models in Healthy Volunteers

In these models, a pharmacological agent, for example, scopolamine is given to healthy subjects to induce symptoms typical of the targeted disease, in this case dementia (cognitive impairment with deterioration of episodic memory and attention) [20, 21]. Compounds with cognition-enhancing characteristics would reduce these symptoms [22], providing early evidence of potential clinical efficacy. Pharmacological models are often used in psychopharmacological research, and, if adequately adapted, some of them can potentially be useful in early clinical development of psychotropic drugs (see [23] for a review). Asthma is another therapeutic area where pharmacological provocation tests in healthy volunteers; using, for example, methacholine [24, 25] or leukotriene D4 [26] are utilized to demonstrate a pharmacological activity and capture early signs of potential clinical efficacy. It should be pointed out that unless the model used involves a mechanism that is relevant for the pathophysiology of the disease, a positive effect of a new compound in the model merely demonstrates a pharmacological antagonistic activity and does not provide early evidence of efficacy in the targeted disease. For leukotriene receptor antagonists, a class of drugs that target a mechanism known to be involved in the pathophysiology of asthma, a reduction of bronchoconstriction could be demonstrated in the leukotriene challenge as well as in other provocation models using, for example, allergens, methacoline, and ultrasound-nebulized distilled water [27, 28].

If models are to be reliable, a number of criteria should be fulfilled. These include face validity, test–retest consistency, responsiveness to reference drugs [29], and a reasonably high response rate in healthy subjects [23]. Since the clinical

**TABLE 3 Pharmacological and Nonpharmacological Models Used in PoP/PoC Trials**

| Model | Therapeutic Area |
|---|---|
| Carbon dioxide inhalation | General anxiety disorder, panic disorder |
| Conditioned aversive anxiety | Anxiety |
| Scopolamine test | Dementia/cognitive disorders |
| Pain models (mechanical, chemical, thermal or electrical stimulation) | Pain |
| Bronchial challenges (allergens, histamine, methacholine, leukotriene, cold air, UNDW) | Asthma |
| Pentagastrin-induced gastric acid secretion | Peptic ulcer, GERD |
| Human endotoxemia model | Sepsis |
| Phenylephrine challenge | Benign prostatic hyperplasia |

UNDW, Ultrasound-nebulized distilled water, GERD, gastroesophageal reflux disease.

manifestations of many diseases are multifaceted and not possible to reproduce in a model involving the administration of a single pharmacological agent, it has been suggested that the model can be useful if it provokes the cardinal symptoms of the disease [23]. Table 3 shows examples of pharmacological models that can be used in PoP/PoC trials.

### 6.2 Nonpharmacological Models

These use the same general principles as pharmacological models. However, they require rigorous standardization for reliability [23]. Examples of nonpharmacological models used in PoP/PoC paradigms include models of anxiety and panic such as simulated public speaking [30] and conditioned aversive anxiety [30], and nonpharmacological provocation tests of asthma and human experimental pain models (Table 3). The latter have evolved to provide a useful translational tool linking animal findings with data obtained in healthy subjects and patients. Moreover, by adopting a multimodal (mechanical, chemical, thermal, electrical stimuli), multitissue (skin, muscles, viscera) approach, it is now possible to explore different pain pathways and mechanisms and assess responses induced by novel putative analgesics in PoP/PoC trials. This approach is described in a recent review by Arendt-Nielsen and colleagues [31].

### 6.3 Provocation Models in Patients

PoP/PoC can also be explored using provocation models in patients. Both pharmacological and nonpharmacological models are potentially useful to progress a new compound through early clinical development. Asthma and allergic diseases are therapeutic areas where provocation models in patients are most frequently used, and the practical setup of the trials does not deviate from that in healthy volunteers. However, since the trials are conducted in the target patient population [27, 28, 32, 33], a positive outcome is more indicative of a successful PoC compared with a positive result from a provocation experiment in healthy subjects. Provocation tests for PoP/PoC in patients have also been used in other therapeutic areas, such as epilepsy. Among the models tested, the technically challenging intermittent photic stimulation in patients with photosensitive epilepsy has been found reliable in

predicting the antiseizure activity of a number of efficacious drugs that were eventually approved [34]. PoP was recently obtained for two novel antiepileptic drugs, brivaracetam, an SV2A ligand, and carisbamate, a new carbamate using this model [35, 36]. Both drugs suppressed generalized photoparoxysmal electroencephalogram (EEG) responses in treated patients. PoP studies in epilepsy using this and other models are discussed in Schmidt [37].

## 7 DIFFERENT TYPES OF POP/POC STUDIES

Considering the broad definition of PoP/PoC (see above), PoP/PoC trials are quite variable in terms of objectives, overall design, trial population, endpoints, as well as regulatory requirements.

### 7.1 Exploratory Investigational New Drug PoC Trials

These are early clinical studies conducted under the auspices of FDA's Exploratory Investigational New Drug Guidance [38], which covers microdose studies of pharmacokinetics or imaging as well as phase 0 clinical trials to study pharmacologically relevant doses. A microdose is defined as 1/100th of the dose calculated to yield a pharmacologic effect with a maximum dose of 100 μg (or 30 nmol for proteins). Imaging studies using microdoses can be very useful in establishing PoP, for example, by demonstrating significant binding of the investigational compound to the target receptor. Since these studies imply single exposures to minute amounts of the test compound, they can be conducted without the standard safety pharmacology and genotoxicity studies.

Phase 0 trials that are designed to study the pharmacologic effects of the test compounds provide a new avenue to early PoC. The FDA accepts the use of doses at which a PK/PD response or target modulation is observed in such trials. However, prerequisites are the lack of an adverse clinical response and that the dose is less than one fourth of the 2-week rodent no adverse effect level (NOAEL), or that the total exposure to drug expressed as the area under the concentration curve (AUC) does not exceed one half of the AUC at NOAEL in the most sensitive species. Due to the lower doses and consequently lower toxicity risks implied, the preclinical pharmacology and toxicology package is less extensive than that required for traditional phase I investigational new drugs (INDs), although safety pharmacology and genotoxicity studies are required. Moreover, if a "go decision" is made, based on the outcome of the study, that is, target modulation is proven, a traditional IND will have to be submitted in order to progress the compound further in clinical development.

One key therapeutic area in which phase 0 exploratory studies are performed is oncology, where there is a great need for new safe and efficacious agents and where early evidence of target modulation (molecular PoC) would potentially expedite the clinical evaluation and facilitate the development of the most promising agents. In phase 0 cancer trials, low doses are administered to a limited number of patients (10–15 patients) [39].

The most critical factor for the success of a phase 0 PoC study in oncology—and other therapeutic areas—is the availability of a sensitive, robust, and qualified/validated pharmacodynamic assay. Kinders and colleagues proposed the following

set of performance criteria for PD assay validation: accuracy, dynamic range, precision, reproducibility and sensitivity [39]. The same group reported the results of an elegant phase 0 trial [40] in which 14 patients with advanced malignancies were treated with low doses of ABT-888, a poly(ADP-ribose) polymerase inhibitor. Using a PD assay that measured the amount of the enzyme polymer in tumor tissue and blood cells as a surrogate for enzyme inhibition, the authors were able to establish the PK/PD relationship and select a dose that had a potential for efficacy with minimized risk of toxicity.

Although the exploratory IND initiative allows for a more flexible mode of developing drugs in oncology and other therapeutic areas, it remains to be seen whether phase 0 PoC studies will result in a more speedy development of safe and efficacious novel drugs.

Apart from the technical challenges that phase 0 PoC trials pose (e.g., PD assay validation), the use of patients in a trial where there is no therapeutic intent (due to subtherapeutic dosing) introduces new ethical standards in clinical research where "benefit to others" is accepted as the sole benefit. This is discussed in a thoughtful editorial [41], where phase 0 trials are described as "ethically challenged and ethically challenging."

### 7.2 Clinical Trials Using Biomarkers or Models in Healthy Subjects

These trials exploit validated or reasonably qualified "fit-for-purpose" biomarkers or models to elicit a clinically relevant pharmacological response in healthy subjects. For example, an early PoP was obtained for degarelix, a GnRH receptor antagonist currently in clinical development for prostate cancer by Ferring Pharmaceuticals, by demonstrating a rapid, profound and sustained reduction in serum levels of testosterone and LH in healthy volunteers (Figs. 3 and 4). The findings were later confirmed in large trials in the target patient population (unpublished data). The availability of a well-validated biomarker/surrogate (testosterone deprivation in prostate cancer) was crucial for the successful PoP/PoC achieved.

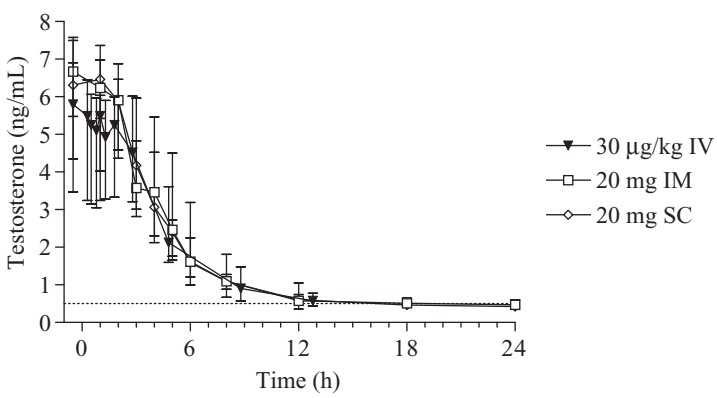


***FIGURE 3*** *Serum testosterone levels in healthy volunteers after the administration of degarelix intravenously, intramuscularly, and subcutaneously. The horizontal dotted line indicates the level of castration (serum testosterone of 0.5 ng/mL) required for efficacy in prostate cancer trials. Median values and quartiles are shown.*

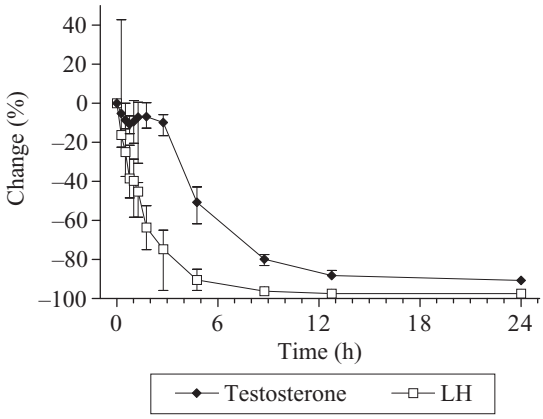


***FIGURE 4*** *Changes (median and quartiles) in serum levels of testosterone and luteinising hormone (LH) after intravenous administration of degarelix (30μg/kg) to healthy subjects. The response confirms the mechanism of action of the compound (GnRH antagonism leading to LH decline with resultant chemical castration).*

### 7.3 Clinical Trials Using Biomarkers in Patients

In these trials, biomarkers related to the pathophysiology of the disease are analyzed and drug-induced modulation of these markers is evaluated. For truly innovative, first-in-class therapeutic agents, a biomarker-based PoP/PoC poses significant challenges. The pathophysiology of the disease is often not completely understood. Consequently, the link between the different biomarkers and the disease biology might not be evident. Therefore, a positive change in a biomarker in a patient trial may constitute a positive PoP but does not guarantee a positive clinical outcome. This is unavoidable in all cases where the biomarkers have not yet been clinically validated. However, as discussed above, the use of biomarkers in PoP/PoC trials is a way of increasing the confidence in a particular therapeutic agent or class of agents to make better internal go–no go decisions. The alternative is to rely primarily on huge clinical trials that might fail late and at a much higher cost.

One therapeutic area in which the use of biomarkers during early drug development is of particular interest is Alzheimer's disease (AD). Despite extensive R&D efforts by many large and small pharmaceutical companies, disease-modifying therapies have been elusive. The identification of presymptomatic neurodegenerative changes that are mirrored by changes in specific markers is crucial and remains to be in focus for AD research. A wide array of biochemical markers such as β-amyloid and tau proteins in CSF [42] and imaging-based biomarkers [43, 44] are available, and evidence seems to suggest that a multiple biomarker approach will be necessary [45]. The logistics of large clinical trials in AD and the fact that available symptomatic treatment of AD (cholinesterase inhibitors, memantine) influences the primary clinical endpoint (cognition), thus blurring the potential effects of a truly disease-modifying agent, make biomarker-based PoP/PoC the most attractive approach to develop such important therapies in a huge area of unmet medical need. Other proposed strategies to enhance drug development in this area include enriched population

PoC studies (patients with mild cognitive impairment who are at higher risk of developing AD), implementing alternative, more sensitive clinical outcome measurements and adaptive dose-finding trials (seamless phase IIb–III) [46].

### 7.4 Trials Using Clinical Endpoints for PoC

In the absence of qualified biomarkers or acceptable surrogates, PoC can only be established by a successful clinical trial conducted in the target patient population using solid clinical endpoints for efficacy. As discussed earlier, biomarkers that are not fully validated can still be useful in demonstrating a pharmacological effect and selecting doses for the "ultimate" PoC trial. PoC trials using clinical outcomes are usually large and can be of long duration, depending on the natural course of the disease, the drug's mode of action, and the sensitivity of the clinical measurements to therapeutic intervention.

## 8 KNOWLEDGE OF EXPOSURE–RESPONSE RELATIONSHIP OF DRUG IS ESSENTIAL FOR SUCCESSFUL POP/POC

Adequate characterization of the ER relationship of the drug is vital for robust PoP/PoC conclusions and consequently solid go–no go decisions. This is easily understood if one realizes that a primary objective of PoP/PoC trials is to demonstrate some signs of efficacy, be it direct or via a biomarker or other surrogates at a dose that does not elicit unacceptable adverse effects. Failure to explore the ER relationship properly early on carries a risk of selecting inappropriate dose(s) in the PoP/PoC trial and, consequently, making the wrong go–no go decision. Some drug developers still advocate the use of a high, tolerated dose close to the maximal tolerated dose (MTD) in a PoP/PoC trial; "if this doesn't work, why bother?" While this approach might work when developing compounds with wide therapeutic windows, it ignores the fundamental principles of clinical pharmacology and carries a significant risk of adverse drug reactions arising later in larger trials or even registering too high a dose of the drug. Finally, with increasing safety awareness, regulatory requirements often include the demonstration of the minimum dose that is maximally effective [11].

It should be underscored that a thoughtful integrated learning approach to PoP/PoC is a prerequisite for well-founded decisions. Notwithstanding the cost constraints that many pharmaceuticals companies experience, PoP/PoC strategies should not be "quick and dirty" or "lean and mean" development approaches. A rational scientifically solid and well-engineered PoP/PoC strategy indeed saves time and money by ensuring the progress of good drug candidates and the termination of compounds with poor chances of making it to the marketplace.

## 9 PROOF-OF-PRINCIPLE/PROOF-OF-CONCEPT OUTCOMES CAN BE "FALSE DAWNS"

Despite the importance of PoP/PoC trials in drug development, as underscored in the literature and reiterated in this chapter, positive trial outcomes are by no means

a guarantee of clinical success. As discussed above, PoP/PoC trials usually rely on biomarkers/surrogates/models that might not adequately predict the clinical outcome and therefore fail the ultimate test of efficacy and safety, that is, the confirmatory phase III pivotal trials. Fleming and DeMets reviewed experiences with biomarkers/surrogates that failed to predict clinical outcomes in different therapeutic areas and discussed reasons for failure [47]. These are: (a) The biomarker/surrogate is not in the causal pathway of the disease. (b) The disease involves several pathways and the drug affects only the one mediated by the biomarker. (c) The biomarker is not in the pathway affected by the drug. (d) The drug affects mechanisms that are independent of the disease process [47]. One of the most classic examples of biomarkers/surrogates failing badly in predicting serious clinical outcomes is the use of reduction of ventricular extra systoles as a surrogate for decreased cardiovascular mortality after myocardial infarction. Flecainide, encainide, and moricizine were found to increase mortality [48, 49] despite the fact that they effectively suppressed ventricular arrhythmias. All three drugs were later withdrawn from the market. This and many other sobering examples reflect the fact that conclusive answers on the efficacy and safety of drugs can only be answered by large clinical trials using solid clinical endpoints. Moreover, long-term safety evaluation often requires extensive clinical use after drug approval. However, this does not decrease the value of biomarker-based PoP/PoC as a means of demonstrating relevant pharmacological activity of new compounds and capturing early signs of potential clinical utility. After all, PoP/PoC is primarily an indication that a certain compound is reasonably good enough to develop further.

### 9.1 Summary of Selected Human Models Quoted in This Chapter that Can be Used in PoP/PoC Trials

- *Human Endotoxemia Model* This is an experimental model of sepsis using intravenously administered, purified lipopolysaccharide (LPS) from *Escherichia coli* to healthy volunteers. The endotoxin induces changes that mimic early inflammatory and cardiovascular responses in septic shock [50–54]. The model has been widely used to study the mechanisms underlying the pathophysiology of sepsis and can potentially be useful to establish early PoP/PoC for drugs targeting these mechanisms. One limitation of the model is that it may not capture the increased vascular permeability, an important feature of severe sepsis/septic shock [55, 56].
- *Carbon Dioxide Inhalation Model* Inhalation of 7.5% carbon dioxide induces anxiety manifestations in healthy subjects with a sense of fear and tension together with increases in blood pressure and heart rate [57]. This model has been shown to be sensitive to the effects of lorazepam and to some extent paroxetine [58], and might therefore be useful to establish early PoP for new compounds intended for the treatment of generalized anxiety disorder (GAD) and panic disorder.
- *Scopolamine-Induced Cognitive Impairment Model* Cognitive impairment is induced in healthy subjects using subcutaneously or intravenously administered scopolamine; and the effect of cognition enhancers can be monitored, for example, using computerized neuropsychological test battery (CNTB) and quantitative electroencephalography (qEEG) [20–22].

- *Conditioned Aversive Anxiety Model* In this model, skin conductance responses to auditory stimuli are measured. Ten neutral tones are initially presented (habituation phase), and tone 11 is immediately followed by an aversive white noise (unconditioned stimulus). The response to a second presentation of the tones (extinction phase) is recorded. Amplitude as well as spontaneous fluctuations (variability) of the skin conductance response is registered. A positive control such as diazepam is used, and compounds with putative anxiolytic effects are compared with the control [30, 59].
- *Pentagastrin-Induced Gastric Acid Secretion* In this model, maximal/submaximal gastric acid secretion is induced in healthy subjects by intravenous infusion of pentagastrin. The effect of antiulcer/GERD drugs such as proton pump inhibitors on gastric acid output and intragastric pH is measured [60, 61].
- *Phenylephrine Challenge* This model is used to investigate the effect of alpha blockers on urethral tone and blood pressure in order to estimate the selectivity for the effect on urethral smooth muscles relative to the hypotensive response [62]. Phenylephrine, which raises urethral and blood pressure, is administered by dose escalation infusions at the estimated $T_{max}$ for the compound, urethral pressure is monitored using a sensor-equipped self-retaining urethral catheter, and blood pressure is recorded simultaneously. New, more selective compounds can be tested in this model establishing PoP/PoC [62].

## REFERENCES


1. Pharmaceutical Research and Manufacturers of America (2008), PhRMA, Pharmaceutical Industry Profile 2008, PhRMA, Washington, D.C.; available at: http://www.phrma.org/files/2008%20Profile.pdf.
2. Kola, I., and Landis, J. (2004), Can the pharmaceutical industry reduce attrition rates? *Nat. Rev. Drug. Discov.*, 3(8), 711–715.
3. Lesko, L. J., Rowland, M., Peck, C. C., et al. (2000), Optimizing the science of drug development: Opportunities for better candidate selection and accelerated evaluation in humans, *Pharmaceutical Res.*, 17(11), 1335–1344.
4. Atkinson, A. J., Colburn, W. A., DeGruttola, V. G., et al. (2001), Biomarkers Definitions Working Group: Biomarkers and surrogate endpoints: Preferred definitions and conceptional framework, *Clin. Pharmacol. Ther.*, 69, 89–95.
5. Lesko, L. J., and Atkinson, A. J. (2001), Use of biomarkers and surrogate endpoints in drug development and regulatory decision making: Criteria, validation, strategies, *Annu. Rev. Pharmacol. Toxicol.*, 41, 347–366.
6. Food and Drug Administration. The "Critical Path Initiative"—Innovation–Stagnation: Challenge and opportunity on the critical path to new medical products; available at: http://www.fda.gov/oc/initiatives/criticalpath/whitepaper.pdf.
7. Wagner, J. A., Williams, S. A., and Webster, C. J. (2007), Biomarkers and surrogate end points for fit-for-purpose development and regulatory evaluation of new drugs. *Clin. Pharmacol. Ther.*, 81(1), 104–107.
8. Williams, S. A., Slavin, D. E., Wagner, J. A., et al. (2006), A cost-effectiveness approach to the qualification and acceptance of biomarkers, *Nat. Rev. Drug. Discov.*, 5(11), 897–902.

9. Berns, B., Démolis, P., and Scheulen, M. E. (2007), How can biomarkers become surrogate endpoints? *EJC* Suppl., 5(9), 37–40.
10. Rolan, P. (1997), The contribution of clinical pharmacology surrogates and models to drug development—a critical appraisal, *Br. J. Clin. Pharmacol.*, 44(3), 219–225.
11. Food and Drug Administration (2003), Guidance for industry: Exposure-response relationships—study design, data analysis, and regulatory applications.
12. Kuhlmann, J. (2007), The applications of biomarkers in early clinical drug development to improve decision-making processes, *Ernst Schering Res. Found. Workshop, Rev.*, (59), 29–45.
13. US Food and Drug Administration. Table of valid genomic biomarkers in the context of approved drug labels; available at: http://www.fda.gov/cder/genomics/genomic_biomarkers_table.htm.
14. Roses, A. D., Saunders, A. M., Huang, Y., et al. (2007), Complex disease-associated pharmacogenetics: Drug efficacy, drug safety, and confirmation of a pathogenetic hypothesis (Alzheimer's disease), *Pharmacogenomics J. Rev.*, 7(1), 10–28.
15. Pusztai, L. (2007), Limitations of pharmacogenomic predictor discovery in Phase II clinical trials, *Pharmacogenomics*, 8(10), 1443–1448.
16. Stroobants, S., Goeminne, J., Seegers, M., et al. (2003), 18FDG-Positron emission tomography for the early prediction of response in advanced soft tissue sarcoma treated with imatinib mesylate (Glivec), *Eur. J. Cancer.*, 39(14), 2012–2020.
17. Christian, B. T., Livni, E., Babich, J. W., et al. (1996), Evaluation of cerebral pharmacokinetics of the novel antidepressant drug, BMS-181101, by positron emission tomography, *J. Pharmacol. Exp. Ther.*, 279(1), 325–331.
18. Pien, H. H., Fishman, A. J., Thrall, J. H., (2005), Sorensen, A. G. Using imaging biomarkers to accelerate drug development and clinical trials, *Drug Disc. Today*, 10(4), 259–266.
19. Strack, T. (2007), Imaging as a tool in drug development, *Drugs Today*, 43(10), 725–736.
20. Christensen, H., Maltby, N., Jorm, A. F., et al. (1992), Cholinergic "blockade" as a model of the cognitive deficits in Alzheimer's disease, *Brain*, 115(Pt 6), 1681–1699.
21. Broks, P., Preston, G. C., Traub, M., et al. (1988), Modelling dementia: Effects of scopolamine on memory and attention, *Neuropsychologia*, 26(5), 685–700.
22. Brass, E. P., Polinsky, R., Sramek, J. J., et al. (1995), Effects of the cholinomimetic SDZ ENS-163 on scopolamine-induced cognitive impairment in humans, *J. Clin. Psychopharmacol.*, 15(1), 58–62.
23. Gilles, C., and Luthringer, R. (2007), Pharmacological models in healthy volunteers: their use in the clinical development of psychotropic drugs, *J. Psychopharmacol. Rev.*, May; 21(3), 272–282.
24. Adamus, W.S., Jansen, U., Schilling, C., et al. (1988), Bronchospasm induced by methacholine inhalation as a model for testing of bronchospasmolytics in healthy volunteers, *Methods Find. Exp. Clin. Pharmacol.*, 10(2), 135–139.
25. Foster, R. W., Jubber, A. S., Hassan, N. A., et al. (1993), Trials of the bronchodilator activity of the xanthine analogue SDZ MKS 492 in healthy volunteers during a methacholine challenge test, *Eur. J. Clin. Pharmacol.*, 45(3), 227–234.
26. O'Shaughnessy, T.C., Georgiou, P., Howland, K., et al. (1997), Effect of pranlukast, an oral leukotriene receptor antagonist, on leukotriene D4 (LTD4) challenge in normal volunteers, *Thorax*, 52(6), 519–522.
27. Carratù, P., Morelli, N., Freire, A. X., et al. (2003), Effect of zafirlukast on methacholine and ultrasonically nebulized distilled water challenge in patients with mild asthma, *Respiration*, 70(3), 249–253.

28. Diamant, Z., Grootendorst, D. C., Veselic-Charvat, M., et al. (1999), The effect of montelukast (MK-0476), a cysteinyl leukotriene receptor antagonist, on allergen-induced airway responses and sputum cell counts in asthma, *Clin. Exp. Allergy*, 29(1), 42–51.
29. Guttmacher, L. B., Murphy, D. L., and Insel, T. R. (1983), Pharmacologic models of anxiety, *Compr. Psychiatry*, 24, 312–326.
30. Deakin, J. F. W., Guimaraes, F. S., and Graeff, F. G. (1994), Testing 5HT theories of anxiety in normal volunteers, in Palomo, T., and Archer, T., Eds., *Strategies for Studying Brain Disorders, Vol. 1, Depressive, Anxiety and Drug Abuse Disorders*, Farrand, London, pp. 211–238.
31. Arendt-Nielsen, L., Curatolo, M., and Drewes, A. (2007), Human experimental pain models in drug development: Translational pain research, *Curr. Opin. Investig. Drugs. Rev.*, 8(1), 41–53.
32. Allocco, F. T., Votypka, V., deTineo, M., et al. (2002), Effects of fexofenadine on the early response to nasal allergen challenge, *Ann. Allergy Asthma Immunol.*, 89(6), 578–584.
33. Greiff, L., Persson, C. G., Svensson, C., et al. (1995), Loratadine reduces allergen-induced mucosal output of alpha 2-macroglobulin and tryptase in allergic rhinitis, *J. Allergy Clin. Immunol.*, 96(1), 97–103.
34. Kasteleijn-Nolst Trenité, D. G. A., Marescaux, C., Stodieck, S., et al. (1996), Photosensitive epilepsy: A model to study the effects of antiepileptic drugs. Evaluation of the piracetam analogue, levetiracetam, *Epilepsy Res.*, 25(3), 225–230.
35. Kasteleijn-Nolst Trenité, D. G. A., Genton, P., Parain, D., et al. (2007), Evaluation of brivaracetam, a novel SV2A ligand, in the photosensitivity model, *Neurology*, 69(10), 1027–1034.
36. Trenité, D. G., French, J. A., Hirsch, E., et al. (2007), Evaluation of carisbamate, a novel antiepileptic drug, in photosensitive patients: An exploratory, placebo-controlled study, *Epilepsy Res.*, 74(2–3), 193–200.
37. Schmidt, B. (2006), Proof of principle studies, *Epilepsy Res.*, 68, 49–52.
38. Food and Drug Administration (2006), Guidance for Industry, Investigators and Reviewers, Exploratory IND Studies.
39. Kinders, R., Parchment, R. E., Ji, J., et al. (2007), Phase 0 clinical trials in cancer drug development: From FDA guidance to clinical practice, *Mol. Interv.*, 7(6), 325–334.
40. Kummar, S., Kinders, R., Gutierrez, M., et al. (2007), Phase 0 Working Group: Inhibition of poly (ADP-ribose) polymerase (PARP) by ABT-888 in patients with advanced malignancies: Results of a phase 0 trial, *J. Clin. Oncol.*, ASCO Annual Meeting Proceedings Part I, 25(18S), 3518.
41. Hill, T. P. (2007), Phase 0 trials: Are they ethically challenged? *Clin. Cancer Res.*, 13(3), 783–784.
42. Sunderland, T., Linker, G., Mirza, N., et al. (2003), Decreased beta-amyloid1-42 and increased tau levels in cerebrospinal fluid of patients with Alzheimer disease, *JAMA*, 289(16), 2094–2103.
43. Frank, R. A., Galasko, D., Hampel, H., et al. (2003), National Institute on Aging Biological Markers Working Group. Biological markers for therapeutic trials in Alzheimer's disease. Proceedings of the biological markers working group; NIA initiative on neuroimaging in Alzheimer's disease, *Neurobiol. Aging Rev.*, 24(4), 521–536.
44. Dickerson, B. C., and Sperling, R. A. (2005), Neuroimaging biomarkers for clinical trials of disease-modifying therapies in Alzheimer's disease, *NeuroRx. Rev.* 2(2), 348–360.
45. Frank, R., and Hargreaves, R. (2003), Clinical biomarkers in drug discovery and development, *Nat. Rev. Drug Discov.*, 2(7), 566–580.

46. Cummings, J. L. (2008), Optimizing phase II of drug development for disease-modifying compounds, *Alzheimer dementia*, 4, S15–S20.
47. Fleming, T. R., and DeMets, D. L. (1996), Surrogate endpoints in clinical trials: Are we being misled? *Ann. Intern. Med.*, 125(7), 605–613.
48. Echt, D. S., Liebson, P. R., Mitchell, L. B., et al. (1991), Mortality and morbidity in patients receiving encainide, flecainide, or placebo. The Cardiac Arrhythmia Suppression Trial, *N. Engl. J. Med.*, 324(12), 781–788.
49. Epstein, A. E., Hallstrom, A. P., Rogers, W. J., et al. (1993), Mortality following ventricular arrhythmia suppression by encainide, flecainide, and moricizine after myocardial infarction. The original design concept of the Cardiac Arrhythmia Suppression Trial (CAST), *JAMA*, 270(20), 2451–2455.
50. Martich, G. D., Boujoukos, A. J., and Suffredini, A. F. (1993), Response of man to endotoxin, *Immunobiology*, 87, 403–416.
51. Suffredini, A. F., Fromm, R. E., Parker, M. M., et al. (1989), The cardiovascular response of normal humans to the administration of endotoxin, *N. Engl. J. Med.*, 321, 280–287.
52. Suffredini, A. F., Shelhamer, J. H., Neumann, R. D., et al. (1992), Pulmonary and oxygen transport effects of intravenously administered endotoxin in normal humans, *Am. Rev. Resp. Dis.*, 45, 1398–1403.
53. Suffredini, A. F., Harpel, P. C., and Parrillo, J. E. (1989), Promotion and subsequent inhibition of plasminogen activation after administration of intravenous endotoxin to normal subjects, *N. Engl. J. Med.*, 320, 1165–1172.
54. Kumar, A., Bunnell, E., Lynn, M., et al. (2004), Experimental human endotoxemia is associated with depression of load-independent contractility indices: Prevention by the lipid a analogue E5531, *Chest*, 126, 860–867.
55. van Eijk, L. T. G. J., Pickkers, P., Smits, P., et al. (2005), Microvascular permeability during experimental human endotoxemia: An open intervention study, *Crit. Care*, 9, R157–R164.
56. Anel, R., and Kumar, A. (2005), Human endotoxemia and human sepsis: Limits to the model, *Crit. Care*, 9, 151–152.
57. Poma, S. Z., Milleri, S., Squassante, L., et al. (2005), Characterization of a 7% carbon dioxide ($CO_2$) inhalation paradigm to evoke anxiety symptoms in healthy subjects, *J. Psychopharmacol.*, 19, 494–503.
58. Bailey, J. E., Kendrick, A., Diaper, A., et al. (2007), A validation of the 7.5% $CO_2$ model of GAD using paroxetine and lorazepam in healthy volunteers, *J. Psychopharmacol.*, 21(1), 42–49.
59. Hellewell, J. S. E., Guimaraes, F. S., Wang, M., et al. (1999), Comparison of buspirone with diazepam and fluvoxamine on aversive classical conditioning in humans, *J. Psychopharmacol.*, 13, 122–127.
60. Lind, T., Cederberg, C., Ekenved, G., et al. (1983), Effect of omeprazole—a gastric proton pump inhibitor on pentagastrin stimulated acid secretion in man, *Gut*, 24, 270–276.
61. Pratha, V. S., Hogan, D. L., Lane, J. R., et al. (2006), Inhibition of pentagastrin-stimulated gastric acid secretion by pantoprazole and omeprazole in healthy adults, *Digestive Diseases Sci.*, 51(1), 123–131.
62. Sultana, S. R., Marshall, S., Davis, J., et al. (2007), Experiences with dose finding in patients in early drug development: The use of biomarkers in early decision making, *Ernst Schering Res. Found. Workshop*, 59, 65–79.

# BIOMARKERS ARE NOT NEW

Ian Dews, MRCP, FFPM
*Envestia Ltd., Thame, Oxfordshire, UK*

## INTRODUCTION

The word *biomarker* in its medical context is a little over 30 years old, having first been used by Karpetsky, Humphrey, and Levy in the April 1977 edition of the *Journal of the National Cancer Institute*, where they reported that the "serum RNase level … was not a biomarker either for the presence or extent of the plasma cell tumor." Few new words can have proved so popular—a recent PubMed search lists more than 370,000 publications that use it! Part of this success can no doubt be attributed to the fact that the word gave a long-overdue name to a phenomenon that has been around at least since the seventh century B.C., when Sushustra, the "father of Ayurvedic surgery," recorded that the urine of patients with diabetes attracted ants because of its sweetness. However, although the origins of biomarkers are indeed ancient, it is fair to point out that the pace of progress over the first 2500 years was somewhat less than frenetic.

## UROSCOPY

Because of its easy availability for inspection, urine was for many centuries the focus of attention. The foundation of the "science" of uroscopy is generally attributed to Hippocrates (460–355 B.C.), who hypothesized that urine was a

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

filtrate of the "humors," taken from the blood and filtered through the kidneys, a reasonably accurate description. One of his more astute observations was that bubbles on the surface of the urine (now known to be due to proteinuria) were a sign of long-term kidney disease. Galen (A.D. 129–200), the most influential of the ancient Greco-Roman physicians, sought to make uroscopy more specific but in reality added little to the subject beyond the weight of his reputation, which served to hinder further progress in this as in many other areas of medicine.

Five hundred years later, Theophilus Protospatharius, another Greek writer, moved things one step nearer to the modern world when he investigated the effects of heating urine and hence established the world's first medical laboratory test. He discovered that heating urine from patients with symptoms of kidney disease caused cloudiness (in fact, the precipitation of proteins). In the sixteenth century, Paracelsus (1493–1541) in Switzerland used vinegar to bring out the same cloudiness (acid, like heat, will precipitate proteins).

Events continued to move both farther north and closer to modernity when in 1695 Frederick Deckers of Leiden in the Netherlands identified this cloudiness as resulting from the presence of albumin. The loop was finally closed when Richard Bright (1789–1858), a physician at Guy's Hospital in London, made the connection between proteinuria and autopsy findings of abnormal kidneys.

The progress from Hippocrates' bubbles to Bright disease represents the successful side of uroscopy, but other aspects of the subject now strike us as a mixture of common sense and bizarre superstition. The technique of collecting urine was thought to be of paramount importance for accurate interpretation. In the eleventh century, Ismail of Jurjani insisted on a full 24-hour collection in a vessel that was large and clean (very sensible) and shaped like a bladder, so that the urine would not lose its "form" (not at all sensible). His advice to keep the sample out of the sun and away from heat continues, however, to be wise counsel.

Gilles de Corbeil (1165–1213), physician to King Philip Augustus of France, recorded differences in sediment and color of urine which he related to 20 different bodily conditions. He also invented the matula, or jorden, a glass vessel through which the color, consistency, and clarity of the sample could be assessed. Shaped like a bladder rounded at the bottom and made of thin clear glass, the matula was to be held up in the right (not the left) hand for careful inspection against the light. De Corbeil taught that different areas of the body were represented by the urine in different parts of the matula. These connections, which became ever more complex, were recorded on uroscopy charts that were published only in Latin, thus ensuring that the knowledge, and its well-rewarded use in treating wealthy patients, was confined to appropriately educated men. To further this education, de Corbeil, in his role as a professor at the Medical School of Salerno, set out his own ideas and those of the ancient Greek and Persian writers in a work called *Poem on the Judgment*

*of Urines*, which was set to music in order that medical students could memorize it more easily. It remained popular for several centuries.

## BLOOD PRESSURE

One of the first excursions away from urine in the search for markers of function and disease came in 1555 with the publication of a book called *Sphygmicae artis iam mille ducentos annos perditae & desideratae Libri V* by a physician from Poznán in Poland named Józef Struś (better known by his Latinized name, Iosephus Struthius). In this 366-page work, Struthius described placing increasing weights on the skin over an artery until the pulse was no longer able to lift the load. The weight needed to achieve this gave a crude measure of what he called "the strength of the pulse" or, as we would call it today, blood pressure.

Early attempts at quantitative measurement of blood pressure had to be conducted in animals rather than human subjects because of the invasiveness of the technique. The first recorded success with these techniques dates from 1733, when the Reverend Stephen Hales, a British veterinary surgeon, inserted a brass pipe into a horse's artery and connected the pipe to a glass tube. Hales observed the blood rising in the tube and concluded not only that the rise was due to the pressure of the blood in the artery but also that the height of the rise was a measure of that pressure.

By 1847, experimental technique had progressed to the point where it was feasible to measure blood pressure in humans, albeit still invasively. Carl Ludwig inserted brass cannulas directly into an artery and connected them via further brass pipework to a U-shaped manometer. An ivory float on the water in the manometer was arranged to move a quill against a rotating drum, and the instrument was known as a *kymograph* ("wave-writer" in Greek).

Meanwhile, in 1834, Jules Hérisson had described his *sphygmomètre*, which consisted of a steel cup containing mercury, covered by a thin membrane, with a calibrated glass tube projecting from it. The membrane was placed over the skin covering an artery and the pressure in the artery could be gauged from the movements of the mercury into the glass tube.

Although minor improvements were suggested by a number of authors over the next few years, credit for the invention of the true *sphygmomanometer* goes to Samuel Siegfried Karl Ritter von Basch, whose original 1881 model used water in both the cuff and the manometer tube. Five years later, Scipione Riva-Rocci introduced an improved version in which an inflatable bag in the cuff was connected to a mercury manometer, but neither of these early machines attracted widespread interest. Only in 1901, when the famous American surgeon Harvey Cushing brought back one of Riva-Rocci's machines on his return from a trip to Italy, did noninvasive blood pressure measurement really take off.

Sphygmomanometers of the late nineteenth century relied on palpation of the pulse and so could only be used to determine systolic blood pressure. Measurement of diastolic pressure only became possible when Nikolai Korotkoff observed in 1905 that characteristic sounds were made by the constriction of the artery at certain points in the inflation and deflation of the cuff. The greater accuracy allowed by auscultation of these *Korotkoff sounds* opened the way for the massive expansion in blood pressure research that characterized the twentieth century.

## IMAGING

To physicians keen to understand the hidden secrets of the human body, few ideas can have been more appealing than the dream of looking through the skin to examine the tissues beneath. The means for achieving this did not appear until a little over a century ago, and then very much by accident. On the evening of November 8, 1895, Wilhem Roentgen, a German physicist working at the University of Würzburg, noticed that light was coming from fluorescent material in his laboratory and worked out that this was the result of radiation escaping from a shielded gas discharge tube with which he was working. He was fascinated by the ability of this radiation to pass through apparently opaque materials and promptly set about investigating its properties in more detail. While conducting experiments with different thicknesses of tinfoil, he noticed that that if the rays passed though his hand, they cast a shadow of the bones.

Quick to see the potential medical uses for his new discovery, Roentgen immediately wrote a paper entitled "On a new kind of ray: a preliminary communication" for the Würzburg Physical Medical Society, reprints of which he sent to a number of eminent scientists with whom he was friendly. One of these, Franz Exner of Vienna, was the son of the editor of the *Vienna Presse*, and hence the news was published quickly, first in that paper and then across Europe. Whereas we are inclined to believe that rapid publication is a feature of the Internet age, the Victorians were no slouches in this matter, and by January 24, 1896 a reprint of the Würzburg paper had appeared in the London *Electrician*, a major journal able to bring details of the new invention to a much wider technical audience.

The speed of the response was remarkable. Many physics laboratories already had gas discharge tubes, and within a month physicists in a dozen countries were reproducing Roentgen's findings. Edwin Frost produced an x-ray image of a patient's fractured wrist for his physician brother, Gilmon Frost, at Dartmouth College in the United States, while at McGill University in Montreal, John Cox used the new rays to locate a bullet in a gunshot victim's leg. Similar results were obtained in cities as far apart as Copenhagen, Prague, and Rijeka in Croatia. Inevitably, not everyone was initially quite so impressed; *The Lancet* of February 1, 1896 expressed considerable surprise that the

Belgians had decided to bring x-rays into practical use in hospitals throughout the country! Nevertheless, it was soon clear that a major new diagnostic tool had been presented to the medical world, and there was little surprise when Roentgen received a Nobel Prize in Physics in 1901.

Meanwhile, in March 1896, Henri Becquerel, professor of physics at the Muséum National d'Histoire Naturelle in Paris, while investigating Roentgen's work, wrapped a fluorescent mineral, potassium uranyl sulfate, in photographic plates and black material in preparation for an experiment requiring bright sunlight. However, a period of dull weather intervened, and prior to actually performing the experiment, Becquerel found that the photographic plates were fully exposed. This led him to write: "One must conclude from these experiments that the phosphorescent substance in question emits rays which pass through the opaque paper and reduce silver salts." Becquerel received a Nobel prize, which he shared with Marie and Pierre Curie, in 1903, but it was to be many years before the use of spontaneous radioactivity reached maturity in medical investigation in such applications as isotope scanning and radioimmunoassay.

The use of a fluoroscopic screen on which to view x-ray pictures was implicit in Roentgen's original discovery and soon became part of the routine equipment not only of hospitals but even of shoe shops, where large numbers of children's shoe fittings were carried out in the days before the true dangers of radiation were appreciated. However, the greatest value of the real-time viewing approach only emerged following the introduction of electronic image intensifiers by the Philips company in 1955.

Within months of the introduction of planar x-rays, physicians were asking for a technique that would demonstrate the body in three dimensions. This challenge was taken up by a number of scientists in different countries, but because of the deeply ingrained habit of reviewing only the national, not the international, literature, these workers remained ignorant of each other's progress for many years.

Carl Mayer, a Polish physician, first suggested the idea of tomography in 1914. André-Edmund-Marie Bocage in France, Gustav Grossmann in Germany, and Allesandro Vallebona in Italy all developed the idea further and built their own equipment. George Ziedses des Plantes in the Netherlands pulled all these strands together in the 1930s and is generally considered the founder of conventional tomography.

Further progress had to wait for the development of powerful computers, and it was not until 1972 that Godfrey Hounsfield, an engineer at EMI, designed the first computer-assisted tomographic device, the EMI scanner, installed at Atkinson Morley Hospital, London, an achievement for which he received both a Nobel prize and a knighthood.

Parallel with these advances in x-ray imaging were ongoing attempts to make similar use of the spontaneous radioactivity discovered by Becquerel. In 1925, Herrman Blumgart and Otto Yens made the first use of radioactivity as a biomarker when they used bismuth-214 to determine the arm-to-arm

circulation time in patients. Sodium-24, the first artificially created biomarker radioisotope, was used by Joseph Hamilton to investigate electrolyte metabolism in 1937.

Unlike x-rays, however, radiation from isotopes weak enough to be safe was not powerful enough to create an image merely by letting it fall on a photographic plate. This problem was solved when Hal Anger of the University of California, building on the efficient gamma-ray capture system using large flat crystals of sodium iodide doped with thallium developed by Robert Hofstadter in 1948, constructed the first gamma camera in 1957.

The desire for three-dimensional images that led to tomography with x-rays also influenced radioisotope imaging and drove the development of single-photon-emission computed tomography (SPECT) by David Kuhl and Roy Edwards in 1968. Positron-emission tomography (PET) also builds images by detecting energy given off by decaying radioactive isotopes in the form of positrons that collide with electrons and produce gamma rays that shoot off in nearly opposite directions. The collisions can be located in space by interpreting the paths of the gamma rays, and this information is then converted into a three-dimensional image slice. The first PET camera for human studies was built by Edward Hoffman, Michael Ter-Pogossian, and Michael Phelps in 1973 at Washington University. The first whole-body PET scanner appeared in 1977.

Radiation, whether from x-ray tubes or from radioisotopes, came to be recognized as having dangers both for the patient and for personnel operating the equipment, and efforts were made to discover media that would produce images without these dangers. In the late 1940s, George Ludwig, a junior lieutenant at the Naval Medical Research Institute in Bethseda, Maryland, undertook experiments using industrial ultrasonic flaw-detection equipment in an attempt to determine the acoustic impedance of various tissues, including human gallstones surgically implanted into the gallbladders of dogs. His observations were detailed in a 30-page project report to the Naval Medical Research Institute dated June 16, 1949, now considered the first report of its kind on the diagnostic use of ultrasound. However, a substantial portion of Ludwig's work was considered classified information by the Navy and was not published in medical journals.

Civilian research into what became the two biggest areas of early ultrasonic diagnosis—cardiology and obstetrics—began in Sweden and Scotland, respectively, both making use of gadgetry initially designed for shipbuilding. In 1953, Inge Edler, a cardiologist at Lund University collaborated with Carl Hellmuth Hertz, a graduate student in the department of nuclear physics who was familiar with using ultrasonic reflectoscopes for nondestructive materials testing, and together they developed the idea of using this method in medicine. They made the first successful measurement of heart activity on October 29, 1953 using a device borrowed from Kockums, a Malmö shipyard. On December 16 of the same year, the method was used to generate an echo encephalogram. Edler and Hertz published their findings in 1954.

At around the same time, Ian Donald of the Glasgow Royal Maternity Hospital struck up a relationship with boilermakers Babcock & Wilcox in Renfrew, where he used their industrial ultrasound equipment to conduct experiments assessing the ultrasonic characteristics of various in vitro preparations. With fellow obstetrician John MacVicar and medical physicist Tom Brown, Donald refined the equipment to the point where it could be used successfully on live volunteer patients. These findings were reported in *The Lancet* on June 7, 1958 as "Investigation of abdominal masses by pulsed ultrasound."

Nuclear magnetic resonance (NMR) in molecules was first described by Isidor Rabi in 1938. His work was followed up eight years later by Felix Bloch and Edward Mills Purcell, who, working independently, noticed that magnetic nuclei such as hydrogen and phosphorus, when placed in a magnetic field of a specific strength, absorb radio-frequency energy, a situation described as being "in resonance."

For the next 20 years NMR found purely physical applications in chemistry and physics, and it was not until 1971 that Raymond Damadian showed that the nuclear magnetic relaxation times of different tissues, especially tumors, differed, thus raising the possibility of using the technique to detect disease. Magnetic resonance imaging (MRI) was first demonstrated on small test tube samples in 1973 by Paul Lauterbur, and in 1975 Richard Ernst proposed using phase and frequency encoding and the Fourier transform, the technique that still forms the basis of MRI.

The first commercial nuclear magnetic imaging scanner allowing imaging of the body appeared in 1980 using Ernst's technique, which allowed a single image to be acquired in approximately 5 minutes. By 1986, the imaging time was reduced to about 5 seconds without sacrificing too much image quality. In the same year, the NMR microscope was developed, which allowed approximately 10-mm resolution on approximately 1-cm samples. In 1993, functional MRI (fMRI) was developed, thus permitting the mapping of function in various regions of the brain.

## ELECTROCARDIOGRAPHY

Roentgen's discovery of x-rays grew out of the detailed investigation of electricity that was a core scientific concern of the nineteenth century, and it is little surprise that investigators also took a keen interests in the electricity generated by the human body itself. Foremost among these was Willem Einthoven. Before his day, although it was known that the body produced electrical currents, the technology was inadequate to measure or record them with any sort of accuracy. Starting in 1901, Einthoven, a professor at the University of Leiden, conducted a series of experiments using a string galvanometer. In his device, electric currents picked up from electrodes on the patient's skin passed through a thin filament running between very

strong electromagnets. The interaction of the electric and magnetic fields caused the filament or "string" to move, and this was detected by using a light to cast a shadow of the moving string onto a moving roll of photographic paper.

It was not, at first, an easy technique. The apparatus weighed 600 lb, including the water circulation system essential for cooling the electromagnets, and was operated by a team of five technicians. Over the next two decades Einthoven gradually refined his machine and used it to establish the electrocardiographic (ECG) features of many different heart conditions, work that was eventually recognized with a Nobel prize in 1924.

As the ECG became a routine part of medical investigations it was realized that a system that gave only a "snapshot" of a few seconds of the heart's activity could be unhelpful or even misleading in the investigation of intermittent conditions such as arrhythmias. This problem was addressed by Norman Holter, an American biophysicist, who created his first suitcase-sized "ambulatory" monitor as early as 1949, but whose technique is dated in many sources to the major paper that he published on the subject in 1957, and other authors cite an even later, 1961 publication.

## HEMATOLOGY

The scientific examination of blood in order to learn more about the health of the patient from whom it was taken can be dated to 1642, when Anthony van Leeuwenhoek first observed blood cells through his newly invented microscope. Progress was at first slow, and it was not until 1770 that leucocytes were discovered by William Hewson, an English surgeon, who also observed that red cells were flat rather than spherical, as had earlier been supposed.

Association of blood cell counts with clinical illness depended on the development of a technical method by which blood cells could be counted. In 1852, Karl Vierordt at the University of Tübingen developed such a technique, which, although too tedious for routine use, was used by one of his students, H. Welcher, to count red blood cells in a patient with "chlorosis" (an old word for what is probably our modern iron-deficiency anemia). He found, in 1854, that an anemic patient had significantly fewer red blood cells than did a normal person. Platelets, the third major cellular constituent of blood, were identified in 1862 by a German anatomist, Max Schultze.

Remarkably, all these discoveries were made without the benefit of cell staining, an aid to microscopic visualization that was not introduced until 1877 in Paul Ehrlich's doctoral dissertation at the University of Leipzig. The movement of blood cell studies from the research laboratory to routine support of patient care needed a fast automatic technique for separating and counting cells, which was eventually provided by the Coulter brothers, Wallace and Joseph. In 1953 they patented a machine that detected the change in electrical conductance of a small aperture as fluid containing cells was drawn through.

Cells, being nonconducting particles, alter the effective cross section of the conductive channel and so signal both their presence and their size.

An alternative technique, flow cytometry, was also developed in stages between the late 1940s and the early 1970s. Frank Gucker at Northwestern University developed a machine for counting bacteria in a laminar stream of air during World War II and used it to test gas masks, the work subsequently being declassified and published in 1947. Louis Kamentsky at IBM Laboratories and Mack Fulwyler at the Los Alamos National Laboratory experimented with fluidic switching and electrostatic cell detectors, respectively, and both described cell sorters in 1965. The modern approach of detecting cells stained with fluorescent antibodies was developed in 1972 by Leonard Herzenberg and his team at Stanford University, who coined the term *fluorescence-activated cell sorter* (FACS).

## BLOOD AND URINE CHEMISTRY

As with hematology, real progress in measuring the chemical constituents of plasma depended largely on the development of the necessary technology. Until such techniques became available, however, ingenious use was made of bioassays, developed in living organisms or preparations made from them, to detect and in some cases quantify complex molecules. A good example of this is the detection of human chorionic gonadotrophin (hCG) in urine as a test for pregnancy. Selmar Aschheim and Bernhard Zondek in Berlin, who first isolated this hormone in 1928, went on to devise the Aschheim–Zondek pregnancy test, which involved five days of injecting urine from the patient repeatedly into an infantile female mouse which was subsequently killed and dissected. The finding of ovulation in the mouse indicated that the injected urine contained hCG and meant that the patient was pregnant.

In the early 1940s, the mouse test gave way to the frog test, introduced by Lancelot Hogben in England. This was a considerable improvement, in that injection of urine or serum from a pregnant woman into the dorsal lymph sac of the female African clawed frog (*Xenopus laevis*) resulted in ovulation within 4 to 12 hours. Although this test was known to give a relatively high proportion of false negatives, it was regarded as an outstanding step forward in diagnosis. One story from the 1950s recounts that with regard to the possible pregnancy of a particular patient, "opinions were sought from an experienced general practitioner, an eminent gynecologist, and a frog; only the frog proved to be correct."

Pregnancy testing, and many other "biomarker" activities, subsequently moved from out-and-out bioassays to the "halfway house" of immunological tests based on antibodies to the test compound generated in a convenient species but then used in an ex vivo laboratory setting, and in 1960 a hemagglutination inhibition test for pregnancy was developed by Leif Wide and Carl Gemzell in Uppsala.

Not all immune reactions can be made to modulate hemagglutination, and a problem with the development of immunoassays was finding a simple way to detect whether the relevant antibody or antigen was present. One answer lay in the use of radiolabeled reagents. Radioimmunoassay was first described in a paper by Rosalyn Sussman Yalow and Solomon Berson published in 1960.

Radioactivity is difficult to work with because of its safety concerns, so an alternative was sought. This came with the recognition that certain enzymes (such as ABTS or 3,3′,5,5′-tetramethylbenzidine) which react with appropriate substrates to give a color change could be linked to an appropriate antibody. This linking process was developed independently by Stratis Avrameas and G. B. Pierce. Since it is necessary to remove any unbound antibody or antigen by washing, the antibody or antigen must be fixed to the surface of the container, a technique first published by Wide and Porath in 1966.

In 1971, Peter Perlmann and Eva Engvall at Stockholm University, as well as Anton Schuurs and Bauke van Weemen in the Netherlands, independently published papers that synthesized this knowledge into methods to perform enzyme-linked immunosorbent assay (ELISA).

A further step toward physical methods was the development of chromatography. The word was coined in 1903 by the Russian botanist Mikhail Tswett to describe his use of a liquid–solid form of a technique to isolate various plant pigments. His work was not widely accepted at first, partly because it was published in Russian and partly because Arthur Stoll and Richard Willstätter, a much better known Swiss–German research team, were unable to repeat the findings.

However, in the late 1930s and early 1940s, Archer Martin and Richard Synge at the Wool Industries Research Association in Leeds devised a form of liquid–liquid chromatography by supporting the stationary phase, in this case water, on silica gel in the form of a packed bed and used it to separate some acetyl amino acids derived from wool. Their 1941 paper included a recommendation that the liquid mobile phase be replaced with a suitable gas that would accelerate the transfer between the two phases and provide more efficient separation: the first mention of the concept of gas chromatography. In fact, their insight went even further, in that they also suggested the use of small particles and high pressures to improve the separation, the starting point for high-performance liquid chromatography (HPLC).

Gas chromatography was the first of these concepts to be taken forward. Erika Cremer working with Fritz Prior in Germany developed gas–solid chromatography, while in the UK, Martin himself cooperated with Anthony James in the early work on gas–liquid chromatography published in 1952. Real progress in HPLC began in 1966 with the work of Csaba Horváth at Yale. The popularity of the technique grew rapidly through the 1970s, so that by 1980, this had become the standard laboratory approach to a wide range of analytes. The continuing problem with liquid or gas chromatography was the identification of the molecule eluting from the system, a facet of the techniques that was to be revolutionized by mass spectrometry.

The foundations of mass spectrometry were laid in the Cavendish Laboratories of Cambridge University in the early years of the twentieth century. Francis Aston built the first fully functional mass spectrometer in 1919 using electrostatic and magnetic fields to separate isotope ions by their masses and focus them onto a photographic plate. By the end of the 1930s, mass spectrometry had become an established technique for the separation of atomic ions by mass.

The early 1950s saw attempts to apply the technique to small organic molecules, but the mass spectrometers of that era were extremely limited by mass and resolution. Positive theoretical steps were taken, however, with the description of time-of-flight (TOF) analysis by W. C. Wiley and I. H. Maclaren. and quadruple analysis by Wolfgang Pauli.

The next major development was the coupling of gas chromatography to mass spectrometry in 1959 by Roland Gohlke and Fred McLafferty at the Dow Chemical Research Laboratory in Midland, Michigan. This allowed, for the first time, an analysis of mixtures of analytes without laborious separation by hand. This, in turn, was the trigger for the development of modern mass spectrometry of biological molecules.

The introduction of liquid chromatography–mass spectrometry (LC-MS) in the early 1970s, together with new ionization techniques developed over the last 25 years (i.e., fast particle desorption, electrospray ionization, and matrix-assisted laser desorption/ionzation), have made it possible to analyze almost every class of biological compound class right up into the megadalton range.

## FASHIONABLE "OMICS"

In Benet Street, Cambridge, stands a rather ordinary pub which on Saturday, February 28, 1953, enjoyed 15 minutes of fame far beyond Andy Warhol's wildest dreams. Two young men arrived for lunch and, as James Watson watched, Francis Crick announced to the regulars in the bar that "we have found the secret of life." The more formal announcement of the structure of DNA appeared in *Nature* on April 2 in a commendably brief paper of two pages with six references. Watson and Crick shared a Nobel prize with Maurice Wilkins, whose work with Rosalind Franklin at King's College, London had laid the groundwork. Sadly, Franklin's early death robbed her of a share of the prize, which is never awarded posthumously.

Over the next two decades a large number of researchers teased out the details of the genetic control of cells, and by 1972 a team at the Laboratory of Molecular Biology of the University of Ghent, led by Walter Fiers, were the first to determine the sequence of a gene (a coat protein from a bacteriophage). The same team followed up in 1976 by publishing the complete RNA nucleotide sequence of the bacteriophage. The first DNA-based genome to be sequenced in its entirety was the 5368-base-pair sequence of bacteriophage

Φ-X174 elucidated by Frederick Sanger in 1977. The science of genomics had been born.

Although the rush to sequence the genomes of ever more complex species (including humans in 2001) initially held out considerable hope of yielding new biomarkers, focus gradually shifted to the protein products of the genes. This process is dated by many to the introduction in 1977 by Patrick O'Farrell at the University of Colorado in Boulder of two-dimensional polyacrylamide gel electrophoresis (2-D PAGE). The subject really took off in the 1990s, however, with technical improvements in mass spectrometers combined with computing hardware and software to support the extremely complex analyses involved.

The next "omics" to become fashionable was metabolomics, based on the realization that the quantitative and qualitative pattern of metabolites in body fluids reflects the functional status of an organism. The concept is by no means new, the first paper addressing the idea (but not using the word) having been "Quantitative Analysis of Urine Vapor and Breath by Gas–Liquid Partition Chromatography" by Robinson and Pauling in 1971. The word *metabolomics*, however, was not coined until the 1990s.

## THE FUTURE

Two generalizations may perhaps be drawn from the accelerating history of biomarkers over the last 2700 years. The first is that each new step depends on an interaction between increasing understanding of the biology and technical improvement of the tools leading to a continuous spiral of innovation. The second is the need for an open but cautious mind. Sushustra's recognition of the implications of sweet urine has stood the test of time; de Corbeil's *Poem on the Judgment of Urines* has not. The ultimate fate of more recent biomarkers will only be revealed by time.

# BIOMARKERS: FACING THE CHALLENGES AT THE CROSSROADS OF RESEARCH AND HEALTH CARE

Gregory J. Downing, D.O., Ph.D.
*U.S. Department of Health and Human Services, Washington, DC*

## INTRODUCTION

Across many segments of the biomedical research enterprise and the health care delivery sectors, the impact of biomarkers has been transforming in many ways: from business and economics to policy and planning of disease management. The pace of basic discovery research progress has been profound worldwide, with the intertwining of innovative technologies and knowledge providing extensive and comprehensive lists of biological factors now known to play integral roles in disease pathways. These discoveries have had a vast impact on pharmaceutical and biotechnology industries, with tremendous growth in investment in biomarker research reaching into the laboratory technology and services sector. These investments have spawned new biomedical industry sectors, boosted the roles of contract research organizations, supported vast new biomarker discovery programs in large corporate organizations, and prompted the emergence of information management in research. Similarly, growth in the academic research programs supporting biomarker research has greatly expanded training capacity, bench and clinical research capacity, and infrastructure, while fueling the growth of intellectual property.

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

By many reports, private-sector applications of biomarkers in toxicity and early efficacy trials have been fruitful in developing decision-making priorities that are introducing greater efficiency in early to midstage medical product development. Despite the heavy emphasis in private and publicly funded research, the reach of the impact of biomarkers into clinical practice interventions at present is challenging to quantify. The costs of development remain high for many drugs, and the numbers of new chemical entities reaching the marketplace have continued to remain relatively low compared to prior years and expectations following robust research expansion of the 1980s and 1990s. Industry concerns about the sustainability of research and development programs have grown in the backdrop of the clinical challenges that are attendant on biomarker applications in clinical trials. Understanding of the clinical implications of disease markers as a consequence of relatively slow evidence development has taken much longer to discern than many had predicted. There have been many challenges in establishing a translational research infrastructure that serves to verify and validate the clinical value of biomarkers as disease endpoints and their value as independent measures of health conditions. The lack of the equivalent of the clinical trial infrastructure for biomarker validation and diagnostics has slowed progress compared to therapeutic and device development. Evidence development processes and evaluations have only now begun to emerge for biomarkers, and wide adoption of them in clinical practice measures has not yet matured. For some, the enthusiasm and economic balance sheets have not been squared, as the clinical measure indices that had been hoped for have been viewed as moderately successful by some and by others as bottlenecks in the pipelines of therapeutic and diagnostic development.

## BRIEF HISTORY OF BIOMARKER RESEARCH, 1998–2008: THE FIRST DECADE

During the last decade of the twentieth century, biomedical research underwent one of the most dramatic periods of change in history. Influenced by a multitude of factors—some scientific, others economic, and still others of policy—new frontiers of science emerged as technology and knowledge converged, and diverged—bringing new discoveries and hope to the forefront of medicine and health. These capabilities came about as a generation's worth of science that brought to the mainstream of biomedical research the foundation for a molecular basis of disease: recombinant DNA technology. Innovative applications of lasers, novel medical imaging platforms, and other advanced technologies began to yield a remarkable body of knowledge that provided unheralded opportunities for discovery of new approaches to the management of human health and disease.

Here we briefly revisit a part of the medical research history that led to the shaping of new directions that is, for now, captured simply by the term

*biomarker*, a biological indicator of health or disease. In looking backward to the 1980s and 1990s and the larger scheme of health care, many new challenges were being faced. The international challenges and global economic threats posed by human immunodeficiency virus (HIV) and AIDS provided the impetus for one of the first steps in target-designed therapies and the use of viral and immune indicators of disease. For the first time, strategically directed efforts in discovery and clinical research paradigms were coordinated at the international level using clinical measures of disease at the molecular level. The first impact of biomarkers on discovery and translational research, both privately and publicly funded, as related to biological measures of viral load, CD4+ T-lymphocyte counts, and other parameters of immune function and viral resistance came to be a mainstay in research and development. Regulatory authority was put in place to allow "accelerated approval" of medical products using surrogate endpoints for health conditions with grave mortality and morbidity. Simultaneously, clinical cancer therapeutics programs had some initial advances with the use of clinical laboratory tests that aided in the distinction between responders and nonresponders to targeted therapies. The relation of *Her2/neu* tyrosine kinase receptor in aggressive breast cancer and response to (Herceptin) [1], and similarly, the association of imatinib (Gleevac) responsiveness with the association of the presence of Philadelphia chromosome translocation involving *BCR/Abl* genes in chronic myelogenous leukemia [2], represented some of the cases where targeted molecular therapies were based on a biomarker test as a surrogate endpoint for patient clinical response. These represented the entry point of pharmaceutical science moving toward co-development, using diagnostic tests to guide selection of therapy around a biomarker.

Diverse changes were occurring throughout the health care innovation pipeline in the 1990s. The rise of the biotechnology industry became an economic success story underpinned by successful products in recombinant DNA technology, monoclonal antibody production, and vaccines. The device manufacturing and commercial laboratory industries became major forces. In the United States, the health care delivery system underwent changes with the widespread adoption of managed care programs, and an effort at health care reform failed. For U.S.-based academic research institutions, it was a time of particular tumult for clinical research programs, often supported through clinical care finances, downsized in response to financial shortfalls. At a time when scientific opportunity in biomedicine was, arguably, reaching its zenith, there were cracks in the enterprise that was responsible for advancing basic biomedical discovery research to the clinic and marketplace.

In late 1997, the director of the National Institutes of Health, Harold Varmus, met with biomedical research leaders from academic, industrial, governmental, and clinical research organizations, technology developers, and public advocacy groups to discuss mutual challenges, opportunities, and responsibilities in clinical research. In this setting, some of the first strategic considerations regarding "clinical markers" began to emerge among stake-

holders in clinical research. From a science policy perspective, steps were taken to explore and organize information that brought to light the need for new paradigms in clinical development. Some of these efforts led to the framing of definitions of terms to be used in clinical development, such as *biomarkers* (a characteristic that is measured and evaluated objectively as an indicator of normal biological processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention) and *surrogate endpoints* (a biomarker that is intended to substitute for a clinical endpoint and is expected to predict clinical benefit or harm, or lack of benefit or harm, based on epidemiologic, therapeutic, pathophysiologic, or other scientific evidence) and descriptions of the information needs and strategic and tactical approaches needed to apply them in clinical development [3]. A workshop was held to address statistical analysis, methodology, and research design issues in bridging empirical and mechanism-based knowledge in evaluating potential surrogate endpoints [4]. In-depth analyses were held to examine information needs, clinical training skills, database issues, regulatory policies, technology applications, and candidate disease conditions and clinical trials that were suitable for exploring biomarker research programs. As a confluence of these organizational activities, in April 1999 an international conference was hosted by the National Institutes of Health (NIH) and U.S. Food and Drug Administration (FDA) [5]. The leadership focused on innovations in technology applications, such as multiplexed gene analysis using polymerase chain reaction technologies, large-scale gel analysis of proteins, and positron-emission tomography (PET) and magnetic resonance imaging (MRI). A summary analysis was crafted for all candidate markers for a wide variety of disease states, and a framework was formed for multiple disease-based public–private partnerships in biomarker development. A series of research initiatives supported by industry, NIH, and FDA were planned and executed in ensuring months. New infrastructure for discovery and validation of cancer biomarkers was put in place. Public–private partnerships for biomarker discovery and characterization were initiated in osteoarthritis, Alzheimer disease, and multiple sclerosis. Research activities in toxicology markers for cardiovascular disease and metabolism by renal and hepatic transformation systems were initiated by FDA. These events did not yield a cross-sector strategic action plan, but did serve as a framework for further engagement across governmental, academic, industrial, and nongovernmental organizations. Among the breakthroughs was the recognition that new statistical analysis methods and clinical research designs would be needed to address multiple variables measured simultaneously and to conduct metaanalyses from various clinical studies to comprehend the effects of a biomarker over time and its role as a reliable surrogate endpoint. Further, it was recognized that there would be needs for data management, informatics, clinical registries, and repositories of biological specimens, imaging files, and common reagents.

Over the next several years, swift movement across the research and development enterprise was under way. It is obvious that future biomarker research

**TABLE 1 Major Scientific Contributions and Research Infrastructure Supporting Biomarker Discovery**

| |
|---|
| Human Genome Project |
| Mouse models of disease (recombinant DNA technology) |
| Information management (informatics tools, open-source databases, open-source publishing, biomarker reference services) |
| Population-based studies and gene–environment interaction studies |
| Computational biology and biophysics |
| Medical imaging: structural and functional |
| High-throughput technologies: in vitro cell-based screening, nanotechnology platforms, molecular separation techniques, robotics, automated microassays, high-resolution optics |
| Proteomics, metabolomics, epigenomics |
| Pharmacogenomics |
| Molecular toxicology |
| Genome-wide association studies |
| Molecular pathways, systems biology, and systems engineering |

was driven in the 1990s and early years of the twenty-first century by the rapid pace of genome mapping and the fall in cost of large-scale genomic sequencing technology, driven by the Human Genome Project. A decade later, it is now apparent that biomarker research in the realm of clinical application has acquired a momentum of its own and is self-sustaining.

The major schemes for applications of biomarkers can be described in a generalized fashion in four areas: (1) molecular target discovery, (2) early-phase drug development, (3) clinical trials and late-stage therapeutic development, and (4) clinical applications for health status and disease monitoring. The building blocks for biomarker discovery and early-stage validation over the last decade are reflected in Table 1. Notable to completion of the international Human Genome Project was the vast investment in technology, database development, training, and infrastructure that have been applied throughout industry toward clinical research applications.

## SCIENCE AND TECHNOLOGY ADVANCES IN BIOMARKER RESEARCH

In the past decade of biomarker research, far and away the most influential driving force was completion of the Human Genome Project in 2003. The impact of this project on biomarker research has many facets beyond establishment of the reference data for human DNA sequences. This mammoth undertaking initiated in 1990 led to the sequence for the nearly 25,000 human genes and to making them accessible for further biological study. Beyond this and the other species genomes that have been characterized,

human initiatives to define individual differences in the genome provided some of the earliest large-scale biomarker discovery efforts. The human haplotype map (HapMap) project defined differences in single-nucleotide polymorphisms (SNPs) in various populations around the world to provide insights into the genetic basis of disease and into genes that have relevance for individual differences in health outcomes. A collaboration among 10 pharmaceutical industry companies and the Wellcome Trust Foundation, known as the SNP consortium, was formed in 1999 to produce a public resource of SNPs in the human genome [6]. The SNP consortium used DNA resources from a pool of samples obtained from 24 people representing several racial groups. The initial goal was to discover 300,000 SNPs in two years, but the final results exceeded this, as 1.8 million SNPs had been released into the public domain at the end of 2002 when the discovery phase was completed. The SNP consortium was notable, as it would serve as a foundation for further cross-industry public–private partnerships that would be spawned as a wide variety of community-based efforts to hasten the discovery of genomic biomarkers (see below).

The next phase of establishing the basic infrastructure to support biomarker discovery, particularly for common chronic diseases, came in 2002 through the International HapMap Project, a collaboration among scientists and funding agencies from Japan, the United Kingdom, Canada, China, Nigeria, and the United States [7]. A haplotype is a set of SNPs on a single chromatid that are associated statistically. This rich resource not only mapped over 3.1 million SNPs, but established additional capacity for identifying specific gene markers in chronic diseases and represented a critical reference set for enabling population-based genomic studies to be done that could establish a gene–environmental basis for many diseases [8].

Within a short time of completing the description of the human genome, a substantial information base was in place to enable disease–gene discoveries on a larger scale. This approach to referencing populations to the well-described SNP maps is now the major undertaking for defining gene-based biomarkers. In recent years, research groups around the world have rapidly been establishing genome-wide association studies to identify specific gene sets associated with diseases for a wide range of chronic diseases.

This new era in population-based genetics began with a small-scale study that led to the finding that age-related macular degeneration is associated with a variation in the gene for complement factor H, which produces a protein that regulates inflammation [9]. The first major implication in a common disease was revealed in 2007 through a study of type II diabetes variants [10]. To demonstrate the rapid pace of discovery of disease gene variants: At the time of this writing, within 18 months following the study, there are now 18 disease gene variants associated with defects in insulin secretion [11].

The rapid growth in genome-wide association studies (GWASs) is identifying a large number of multigene variants that are leading to subclassification

of diseases with common phenotype presentations. Among the databases being established for enabling researchers public access to these association studies is dbGaP, the database of genotype and phenotype. The database, which was developed and is operated by the National Library of Medicine's National Center for Biotechnology Information, archives and distributes data from studies that have investigated the relationship between phenotype and genotype, such as GWASs. At present, dbGAP contains 36 population-based studies that include genotype and phenotype information. Worldwide, dozens if not hundreds of GWASs are under way for a plethora of health and disease conditions associated with genetic features. Many of these projects are collaborative, involve many countries, and are supported through public–private partnerships. An example is the Genomics Association Information Network (GAIN), which is making genotype–phenotype information publicly available for a variety of studies in mental health disorders, psoriasis, and diabetic nephropathy [12]. For the foreseeable future, substantial large-scale efforts will continue to characterize disease states and catalog genes associated with clinically manifested diseases.

As technology and information structures advance, other parameters of genetic modifications represent new biomarker discovery opportunities. The use of metabolomics, proteomics, and epigenomics in clinical and translational research is now being actively engaged. A new large-scale project to sequence human cancers, the Cancer Genome Atlas, is focused on applying large-scale biology in the hunt for new tumor genes, drug targets, and regulatory pathways. This project is focused not only on polymorphisms but also on DNA methylation patterns and copy numbers as biomarker parameters [13]. Again, technological advances are providing scientists with novel approaches to inferring sites of DNA methylation at nucleotide-level resolution using a technique known as high-throughput bisulfite sequencing (HTBS). Large-scale initiatives are also under way to bring a structured approach to relating protein biomarkers into focus for disease conditions. Advances in mass spectrometry, protein structure resolution, bioinformatics for archiving protein-based information, and worldwide teams devoted to disease proteomes have solidified in recent years. Although at a more nascent stage of progress in disease characterization, each of these emerging new fields is playing a key complementary role in biomarker discovery in genetics and genomics.

Supporting this growth in biomarker discovery is massive investment over the last 10 years worldwide by public and private financers that has spawned hundreds of new commercial entities worldwide. Private-sector financing for biomarker discovery and financing has become a major component of biomedical research and development (R&D) costs in pharmaceutical development. Although detailed budget summaries have not been established for U.S. federal funding of biomarker research, in a recent survey by McKinsey and Co., biomarker R&D expenditures in 2009 were estimated at $5.3 billion, up from $2.2 billion in 2003 [14].

## POLICIES AND PARTNERSHIPS

Although progress in biomarker R&D has accelerated, the clinical translation of disease biomarkers as endpoints in disease management and as the foundation for diagnostic products has had more extensive challenges. A broad array of international policy matters over the past decade have moved to facilitate biomarker discovery and validation (Table 2). In the United States, the FDA has taken a series of actions to facilitate applications of biomarkers in drug development and use in clinical practices as diagnostic and therapeutic monitoring. A voluntary submission process of genomic data from therapeutic development was initiated by the pharmaceutical industry and the FDA in 2002 [15]. This program has yielded many insights into the role of drug-metabolizing enzymes in the clinical pharmacodynamic parameters of biomarkers in drug development. In July 2007, guidelines for the use of multiplexed genetic tests in clinical practice to monitor drug therapy were issued by the FDA [16] More recently, the FDA has begun providing label requirements indicating those therapeutic agents for which biomarker assessment can be recommended to avoid toxicity and enhance the achievement of therapeutic responses [17]. In 2007, Congress authorized the establishment of a private–public resource to support collaborative research with the FDA.

One of the major obstacles to clinical genomic research expressed over the years has been a concern that research participants may be discriminated against in employment and provision of health insurance benefits as a result of the association of genetic disease markers. After many years of deliberation, the U.S. Congress passed legislation known as the Genetic Information Non-discrimination Act of 2008, preventing the use of genetic information to deny employment and health insurance.

The past decade has seen many new cross-organizational collaborations and organizations developed to support biomarker development. For example, the

**TABLE 2 Major International Policy Issues Related to Biomarker Research**

| |
|---|
| Partnerships and collaborations: industry, team science |
| Expanded clinical research capacity through increases in public and private financing |
| Open-source publishing, data-release policies |
| Standards development and harmonization |
| FDA Critical Path Initiative |
| Regulatory guidances for medical product development |
| Biomarkers Consortium |
| Evidence-based medicine and quality measures of disease |
| International regulatory harmonization efforts |
| Public advocacy in medical research |
| Genetic Information Non-discrimination Act of 2008 (U.S.) |

American Society for Clinical Oncology, the American Association for Cancer Research, and the FDA established collaborations in workshops and research discussions regarding the use of biomarkers for ovarian cancer as surrogate endpoints in clinical trials [18]. The FDA Critical Path Initiative was developed in 2006, with many opportunities described for advancing biomarkers and surrogate endpoints in a broad range of areas for therapeutic development [19,20]. Progress in these areas have augmented industry knowledge of application of biomarkers in clinical development programs and fostered harmony with international regulatory organizations in the ever-expanding global research environment. This program has been making progress on expanding the toolbox for clinical development—many of the components foster development and application of biomarkers. As an example of international coordination among regulatory bodies, recently the FDA and the European Medicines Agency (EMEA) for the first time worked together to develop a framework allowing submission, in a single application to the two agencies, of the results of seven new biomarker tests that evaluate kidney damage during animal testing of new drugs. The new biomarkers are KIM-1, albumin, total protein, $\beta_2$-microglobulin, cystatin C, clusterin, and trefoil factor-3, replacing blood urea nitrogen (BUN) and creatinine in assessing acute toxicity [21]. The development of this framework is discussed in more detail by Goodsaid in Chapter Characterization and Validation Biomarkers in Drug Development: Regulatory Perspective.

In 2007, Congress established legislation that formed the Reagan–Udall Foundation, a not-for-profit corporation to advance the FDA's mission to modernize medical, veterinary, food, food ingredient, and cosmetic product development, accelerate innovation, and enhance product safety. Another important policy in biomarker development occurred with the above-mentioned establishment of the Critical Path Institute in 2006 to facilitate precompetitive collaborative research among pharmaceutical developers. In working closely with the FDA, these collaborations have focused on toxicology and therapeutic biomarker validation [22]. In 2007, a new collaboration building on public–private partnerships with industry was formed to develop clinical biomarkers. In 2006, the Biomarkers Consortium was established as a public–private initiative with industry and government to spur biomarker development and validation projects in cancer, central nervous system, and metabolic disorders in its initial phase [23]. These programs all support information exchange and optimize the potential to apply well-characterized biomarkers to facilitate pharmaceutical and diagnostic development programs.

Other policies that are broadening the dissemination of research findings relate to the growing directions toward open-source publishing. In 2003, the Public Library of Science began an open-source publication process that provides instant access to publications [24]. Many scientific journals have moved to make their archives available six to 12 months after publication. In 2008, the National Institutes of Health implemented policy that requires publications of scientific research with U.S. federal funding to be placed in the public domain within 12 months of publication [25]. All of these policy actions are

favoring biomarker research by accelerating the transfer of knowledge from discovery to development. New commercial management tools have been developed to provide extensive descriptions of biomarkers and their state of development. Such resources can help enhance industry application of well-characterized descriptive information and increase efficiency of research by avoiding duplication and establishing centralized credentialing of biomarker information [26].

New business models are emerging among industry and patient advocacy organizations to increase the diversity of financing options for early-stage clinical development [27]. Private philanthropy conducted with key roles of patient groups are supporting research in proof-of-concept research and target validation with the expectation that these targeted approaches will lead to commercial interests in therapeutic development. Patient advocacy foundations are supporting translational science in muscular dystrophy, amyotrophic lateral sclerosis, juvenile diabetes, multiple myeloma, and Pompe disease, often with partnerships from private companies [28,29].

## CHALLENGES AND SETBACKS

While progress in biomarker R&D has accelerated, the clinical translation of disease biomarkers as endpoints in disease management and as the foundation for diagnostic products has had more extensive challenges [30]. For example, we have not observed a large number of surrogate endpoints emerging as clinical trial decision points. Notable exceptions to this include imaging endpoints that have grown in substantial numbers. In most cases for drug development, biomarkers are being applied in therapeutic development to stratify patients into subgroups of responders, to aid in pharmacodynamic assessment, and to identify early toxicity indicators to avoid late-stage failures in therapeutic development.

There are difficulties in aligning the biomarker science to clinical outcome parameters to establish the clinical value in medical practice decision making. As applied in clinical practice, biomarkers have their most anticipated applications in pharmacotherapeutic decisions in treatment selection and dosing, risk assessment, and stratification of populations for disease preemption and prevention. In the United States, challenges to the marketplace are presented by the lack of extensive experience with pathways for medical product review and reimbursement systems that establish financial incentives for biomarker development as diagnostic assays. Clinical practice guidelines for biomarker application in many diseases are lacking, leaving clinicians uncertain about what roles biomarker assays play in disease management. In addition, few studies have been done to evaluate the cost-effectiveness of inclusion of biomarkers and molecular diagnostics in disease management [31]. The lack of these key pieces in a system of modern health care can cripple plans for integration of valuable technologies into clinical practice.

Scientific setbacks have also occurred across the frontier of discovery and development. Among notable instances was the use of pattern recognition of tandem mass spectrometric measurements of blood specimens in ovarian cancer patients. After enthusiastic support of the application in clinical settings, early successes were erased when technical errors and study design issues led to faulty assumptions about the findings. Across clinical development areas, deficiencies in clinical study design have left initial study findings unconfirmed, often due to overfitting of sample size to populations, and improper control for selection and design bias [32].

Commercial development of large-scale biology companies has also struggled in some ways to identify workable commercial models. Initial enthusiasm early in the decade about private marketing of genomic studies in disease models faltered as public data resources emerged. Corporate models for developing large proteomic databases faltered based on a lack of distinguished market value, little documented clinical benefit, and wide variability in quality of clinical biospecimens.

Evidence development to support clinical utility of many biomarkers to be used as clinical diagnostics is difficult to establish, as clinical trial infrastructure has not yet been established to validate candidate biomarkers for clinical practice. An obstacle to this has been access to well-characterized biospecimens coupled with clinical phenotype information. This has led to calls for centralized approaches to biospecimen collection and archiving to support molecular analysis and biomarker research [33]. Furthermore, a wide variety of tissue collection methods and DNA and protein preparation for molecular analysis has been at the root of many problems of irreproducibility. Standards development and best practices have been represented as cornerstones in facilitating biomarker validation [34,35]. Similarly, reacting to the lack of reproducibility of findings in some studies, proposals have been made for standards in study design for biomarker validation for risk classification and prediction [36].

## LOOKING FORWARD

The next decade of biomarker research is promising, with a push toward more clinical applications to be anticipated. Key factors on the horizon that will be integral to clinical adoption are summarized in Table 3. The confluence of basic and translational research has set the stage for *personalized medicine*, a term of art used widely now, which indicates that health care practices can be customized to meet specific biological and patient differences. The term was not part of the lexicon in 1997 but speaks to the consumer-directed aspects of biomedical research. Genomic services offered to consumers have emerged with the use of GWASs, although clinical value and impact are not known. It is clear that the emergence of biomarkers in an ever-changing health care delivery system will in some fashion incorporate the consumer marketplace.

**TABLE 3 Looking Ahead: Implementing Biomarkers in Clinical Care**

| |
|---|
| Intellectual property policy |
| Phenotypic disease characterization |
| Clinical translation: biomarker validation and verification |
| Clinical trials stratification based on biological diversity |
| Surrogate endpoints: managing uncertainty and defining boundaries in medical practice |
| Data-sharing models |
| Dynamic forces in industry financing |
| Co-development of diagnostics and therapeutics |
| Clinical infrastructure for evidence development and clinical utility of diagnostics |
| Health information exchange and network services |
| Consumer genomic information services |

Prospects for biomarkers to continue to play a major role in the transformation of pharmaceutical industry research remain high as new technology platforms, bioinformatics infrastructure, and credentialed biomarkers evolve. The emergence of a clearer role for federal regulators and increased attention to appraisal of value from genomic-based diagnostics will help provide guideposts for investment and a landscape for clinical application. One can anticipate that the impact of genomics in particular will probably provide a clinical benefit in chronic diseases and disorders, where multiple biomarker analyses reflect models and pathways of diseases. An emerging clinical marketplace is evolving for the development and application of biomarker assays as clinical diagnostics. The pathway for laboratory-developed tests will probably evolve to include FDA oversight of certain tests with added complexity of multiple variables integrated into index scoring approaches to assist in therapeutic selection. Clinical practice guidelines are beginning to emerge for the inclusion of biomarkers to guide stratification of patients and therapeutic decision making. The early impact of these approaches is now evident in oncology, cardiovascular, and infectious disease and immune disorders.

Advancing clinical biomarker to improve safety and quality of health care as a mainstay remains many years away. The clinical evaluation processes for diagnostics and targeted molecular therapy as a systematic approach have not yet been firmly established. The use of electronic health information, and integration of information from health plans and longitudinal data collection and randomized clinical trials will need integration and coordination for effective implementation in medical decision making.

In 2008, the first impact was felt for over-the-counter or electronically available consumer services based on genetic tests. Utilizing public and private genome-wide association databases, private-sector resources using powerful search engines coupled with family history information and SNP analysis developed a consumer service that identifies possible health risks. Although

the medical benefit of such services remains undocumented, the successful entry of several services and the growth of online commercial genomic services indicates interest among health-conscious citizens for understanding inherited disease risk. Another noteworthy factor that will probably play an important role in the next decade of clinical biomarker adoption is the development of standards and interoperability specifications for the health care delivery system and consumers. Interoperable health record environments will probably provide more flexibility and mobility of laboratory information and provide advantages to consumer empowerment in prevention and disease management.

One of the most important and sweeping challenges with biomarkers is the development of intellectual property policies that will bring opportunity and entrepreneurship in balance with meeting unmet market needs and clinical value. Because of the likelihood that single gene mutations or simple protein assays are not by themselves a discovery equivalent to diagnostic tests or new clinical markers, the arraying of the convergence of circles of technologies and knowledge will need new approaches to management for the combined aspects of technology to be brokered as real value in health care. Indeed, the challenges ahead in noting the importance overall for this notion is underscored more globally by Alan Greenspan in noting that "arguably, the single most important economic decision our lawmakers and courts will face in the next twenty-five years is to clarify the rules of intellectual property" [37].

Overall, a decade's worth of work has charted a robust and vibrant course for biomarkers across the biomedical research and development landscape. Clinical applications of biomarkers in medical practice are coming more into focus through diagnostics and molecularly targeted therapies, but a long period of time may pass before biomarker-based medicine becomes a standard in all areas of health care practice.

## REFERENCES

1. Ross JS, Fletcher JA, Linette GP, et al. (2003). The *Her-2/neu* gene and protein in breast cancer 2003: biomarker and target of therapy. *Oncologist*, 8:307–325.
2. Deininger M, Druker BJ (2003). Specific targeted therapy of chronic myelogenous leukemia with Imatinib. *Pharmacol Rev*, 55:401–423.
3. Biomarkers Definitions Working Group (2001). Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. *Clin Pharmacol Ther*, 69:89–95.
4. De Grottola VG, et al. (2001). Considerations in the evaluation of surrogate endpoints in clinical trials: summary of a National Institutes of Health Workshop. *Control Clin Trials*, 22:485–502.
5. Downing GJ (ed.) (2000). *Biomarkers and Surrogate Endpoints: Clinical Research and Applications*. Elsevier Science, Amsterdam.

6. Sachidanandam R, Weissman D, Schmidt S, et al. (The International SNP Map Working Group) (2001). A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature*, 409:928–933.
7. The International HapMap Consortium (2003). The International HapMap Project. *Nature*, 426:789–796.
8. The International HapMap Consortium (2007). A second generation human haplotype map of over 3.1 million SNPs. *Nature*, 449:851–861.
9. Klein RJ, Zeiss C, Chew EY, et al. (2005). Complement factor H polymorphism in age-related macular degeneration. *Science*, 308:385–389.
10. Sladek R, Rocheleau G, Rung J, et al. (2007). A genome-wide association study identifies novel risk loci for type 2 diabetes. *Nature*, 445:881–885.
11. Perry JR, Frayling TN (2008). New gene variants alter type 2 diabetes risk predominantly through reduced beta-cell function. *Curr Opin Clin Nutr Metab Care*, 11:371–378.
12. GAIN Collaborative Research Group (2007). New models of collaboration in genome-wide association studies: the Genetic Association Information Network. *Nat Genet*, 39(9):1045–1051.
13. Collis FS, Barker AD (2007). Mapping the cancer genome: pinpointing the genes involved in cancer will help chart a new course across the complex landscape of human malignancies. *Sci Am*, 296;50–57.
14. Conway M, McKinsey and Co. (2007). Personalized medicine: deep impact on the health care landscape. http://sequencing.hpcgg.org/PM/presentations/Tue_08_Conway_061120%20Michael%20Conway%20Harvard%20Pers%20Med%20Presentation.pdf (accessed Sept. 5, 2008).
15. Orr MS, Goodsaid F, Amur S, Rudman A, Frueh FW (2007). The experience with voluntary genomic data submissions at the FDA and a vision for the future of the voluntary data submission program. *Clin Pharmacol*, 81:294–297.
16. FDA (2007). Guidance for Industry and FDA Staff: Pharmacogenetic tests and genetic tests for heritable markers. http://www.fda.gov/cdrh/oivd/guidance/1549.html.
17. Frueh FW, et al. (2008).Pharmacogenomic biomarker information in drug labels approved by the United States Food and Drug Administration: prevalence of related drug use. *Pharmacotherapy*, 28:992–998.
18. Bast RC, Thigpen JT, Arbuck SG, et al. (2007). Clinical trial endpoints in ovarian cancer: report of an FDA/ASCO/AACR Public Workshop. *Gynecol Oncol*, 107(2):173–176.
19. The critical path to new medical products. http://www.fda.gov/oc/initiatives/criticalpath/report2007.html.
20. FDA (2004). Challenge and opportunity on the critical path to new medical products. http://www.fda.gov/oc/initiatives/criticalpath/whitepaper.html (accessed Aug. 23, 2008).
21. FDA (2008). European Medicines Agency to consider additional test results when assessing new drug safety. http://www.fda.gov/bbs/topics/NEWS/2008/NEW01850.html.
22. Woolsey RL, Cossman J (2007). Drug development and the FDA's Critical Path Initiative. *Clin Pharmacol Ther*, 81:129–133.

23. The Biomarkers Consortium (2008). On the critical path of drug discovery. *Clin Pharmacol Ther*, 83:361–364.
24. Public Library of Science. http://www.plos.org.
25. National Institutes of Health (2008). Revised policy on enhancing public access to archived publications resulting from NIH-funded research. NOT 08–033. http://grants.nih.gov/grants/guide/notice-files/not-od-08-033.html.
26. Thomson Reuters. BIOMARKERcenter. http://scientific.thomsonreuters.com/products/biomarkercenter/ (accessed Sept. 23, 2008).
27. Kessel M, Frank F (2007). A better prescription for drug-development financing. *Nat Biotechnol*, 25:859–866.
28. PriceWaterhouseCoopers (2007). Personalized medicine: the emerging pharmacogenomics revolution. Blogal Technology Centre, Health Research Institute, San Jose, CA.
29. Trusheim MR, Berndt ER, Douglas FL (2007). Stratified medicine: strategic and economic implications of combiing drugs and clinical biomarkers. *Nat Rev Drug Discov*, 6(4):287–293.
30. Phillips KA, Van Bebber S, Issa A (2006). Priming the pipeline: a review of the clinical research and policy agenda for diagnostics and biomarker development. *Nat Revi Drug Discov*, 5(6):463–469.
31. Phillips KA, Van Bebber SL (2004). A systematic review of cost-effectiveness analyses of pharmacogenomic interventions. *Pharmacogenomics*, 5(8):1139–1149.
32. Ransohoff DW (2005). Lessons from controversy: ovarian cancer screening and serum proteomics. *J Natl Cancer Inst*, 97:315–319.
33. Ginsburg GS, Burke TW, Febbo T (2008). Centralized biospecimen repositories for genetic and genomic research. *JAMA*, 299:1359–1361.
34. National Cancer Institute (2007). National Cancer Institute best practices for biospecimen resources. http://biospecimens.cancer.gov/practices/.
35. Thomson Reuters (2008). Establishing the standards for biomarkers research. http://scientific.thomsonreuters.com/pm/biomarkers_white_paper_0308.pdf (accessed Sept. 4, 2008).
36. Pepe MS, Feng Z, Janes H, Bossoyt PM, Potter JD (2008). Pivotal evaluation of the accuracy of a biomarker used for classification or prediction: standards for study design. *J Natl Cancer Inst*, eprint, Oct. 8.
37. Greenspan A (ed.) (2007). *The Age of Turbulence: Adventures in a New World*. Penguin Press, New York.

# ENABLING GO/NO GO DECISIONS

J. Fred Pritchard, Ph.D.
*MDS Pharma Services, Raleigh, North Carolina*

Mallé Jurima-Romet, Ph.D.
*MDS Pharma Services, Montreal, Quebec, Canada*

## UNDERSTANDING RISK

There is no question that most people who know something about our industry consider developing a drug product as a risky business. Usually, they mean that the investment of time and money is high while the chance of a successful outcome is low compared to other industries that create new products. Yet the rewards can be great, not only in terms of monetary return on investment (ROI) but also in the social value of contributing an important product to the treatment of human disease.

*Risk* is defined as "the possibility of loss or injury" [1]. Therefore, inherent in the concept is the mathematical sense of probability of occurrence of something unwanted. Everyday decisions and actions that people take are guided by conscious and unconscious assessments of risk, and we are comfortable with compartmentalized schemes where we sense that a situation is very high, high, medium, low, or very low risk. We often deal with the concept of a relative risk (e.g., in comparison to other options). Some risks can be defined in more absolute terms, such as some type of population measure based on trend analysis of prior incidence statistics (e.g., current risk of postmenopausal Caucasian women in the United States being diagnosed with breast cancer).

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

These types of population-based risk data, while often much debated in the scientific and popular press, do affect decision making at the individual level.

As opposed to an individual's skill at assessing risks, decisions and actions taken during drug development require objective and systematic risk assessment by groups of people. There are many stakeholders involved in the development of a drug product. These include the specialists that do the science required in drug development. Also included are the investors and managers who make decisions about how limited resources will be used. Clinical investigators who administer the drug and the patients who agree to participate in clinical trials are stakeholders, as are the regulators and institutional review boards (IRBs) or ethics committees that approve use of an experimental drug in humans. Each stakeholder has his or her unique perspective of risk. The prime focus of some is the business risk involved, including how much work and money is invested to progress the drug at each phase of development. On the other hand, IRBs, regulators, investigators, and patients are concerned primarily with the safety risk to the patient.

Figure 1 depicts the major factors that contribute to three different perspectives of risk, broadly classified as risk to patient, business risk, and risk of therapeutic failure. Investigators and patients are concerned primarily with safety and efficacy factors associated with a particular therapy. Although regulators are concerned primarily with patient safety, regulations do affect how much development will be required for a particular drug candidate. The business risk can be greatly affected by the investment and development partners involved and the expectations set with owners and managers. Potential competitors in the marketplace affect risk, as does the available pool of money for development. The common factor for all three risk perspectives is the novelty

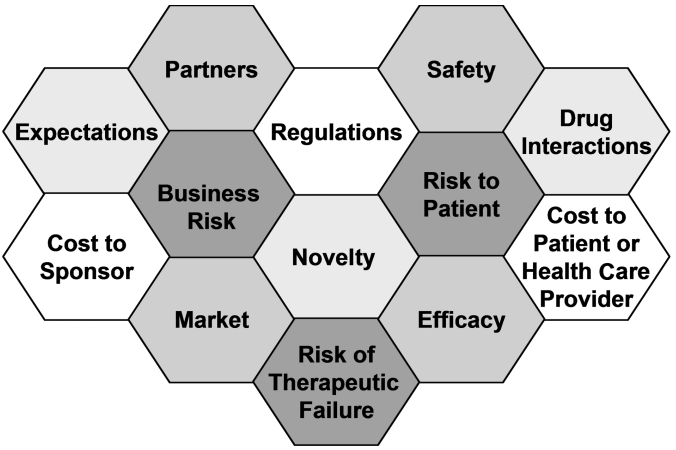


**Figure 1** Major factors affecting three main perspectives of risk in drug development: business risk, risk to patient, and risk of therapeutic failure.

of the product. Drug candidates that represent new targets for therapy while representing a hope for improved efficacy do require more interactions with regulators and investigators, adding to development expense. In addition, because the target is unproven, there is a greater relative risk for therapeutic failure compared to proven pharmacological targets of disease intervention. Therefore, when attempting to express the risks involved in developing a drug, it is important to understand the varying perspectives of each major group of stakeholders.

Before risk assessment can be fully appreciated in the context of decision making in drug development, it must be balanced with the perceived benefits of the product. For example, the patient and physician would tolerate a high level of risk to the patient if the potential benefits offered by a novel therapeutic are for a life-threatening disease for which there is no effective treatment. In a similar way, a high degree of business risk may be tolerated by investors in novel drug candidates if the return on investment, if successful, were very high. Like risk, benefits are also in the eye of the beholder.

Ideally, on many occasions during drug development, each stakeholder is asked to assess his or her view of risk versus benefit based on current data. This assessment will become part of the decision-making process that drives drug development in a logical and, hopefully, collaborative way. Effective decision making requires integrating these varying risk–benefit assessments in a balanced way. The decision gate approach is a useful way to integrate these needs into the drug development process.

## DECISION GATES

Drug development is a process that proceeds through several high-level *decision gates* from identification of a potential therapeutic agent through to marketing a new drug product [2]. A decision gate is a good analogy. For a drug candidate to progress further in drug development it must meet a set of criteria that have been agreed to by the decision makers before they will open the gate. It is "go/no go" because the future product life of the drug hangs in the balance. Once a new drug candidate has progressed through a decision gate, the organization should be committed to expend even greater resources (money and scientists' time) to do the studies to address criteria for the next decision gate along the development path.

Disciplined planning and decision making is required to leverage the value of the decision gate approach. An example of a decision grid applicable to the early phase of drug development is depicted in Table 1. Initially, a clear set of questions at each decision gate needs to be agreed upon and understood by the decision makers. In later stages of drug development these questions and criteria are often presented in a *target product profile* (TPP), which is, in essence, a summary of a drug development plan described in terms of labeling concepts. The FDA has formalized the agency's expectations for a TPP in a

**TABLE 1 Decision Gate Grid Defining Criteria Required to Move Through the Decision Gate "Safe to Give to Humans?"**

| | Efficacy | Animal Safety | PK/ADME | CMC | Phase I Study Design |
|---|---|---|---|---|---|
| Best achievable (enhance investment) | Rodent model: $ED_{90}$<0.3 mg/kg; human receptor: $IC_{50}$<1 μM | Rodent: NOAEL >50 mg/kg; dog: NOAEL >50 mg/kg; no genotoxicity; no CV effects | Dog half-life >8 h; dog BA >90%; no active metabolites; no *CYP2D6* inhibition | API stable for at least 6 mo Cost of goods for API <$10 K/kg GMP tablet | Cohorts of healthy normal subjects: staggered (SD to MD) dose escalation; normal safety monitoring |
| Base case (invest in next phase) | Rodent model: $ED_{90}$<1 mg/kg; human receptor: $IC_{50}$<10 μM | Rodent: NOAEL >10 mg/kg; dog: NOAEL >10 mg/kg; no genotoxicity no CV; effects <100 mg/kg | Dog half-life >4 h; dog BA >50%; <10% active metabolites; *CYP2D6* inhibition: $K_i$>10 μM | API stable for at least 3 mo Cost of goods for API <$20 K/kg GMP tablet | Cohorts of healthy normal subjects: sequential dose escalation; intensive CV monitoring |
| Minimum required (proceed with caution) | Rodent model: $ED_{90}$<2 mg/kg; human receptor: $IC_{50}$<50 μM | Rodent: NOAEL >5 mg/kg; dog: NOAEL >5 mg/kg; equivocal genotoxicity at highest exposure; no CV effects <30 mg/kg | Dog half-life >2 h; dog BA >30%; <30% active metabolites; *CYP2D6* inhibition: $K_i$>1 μM | API stable for at least 3 mo Cost of goods for API <$30 K/kg GMP tablet | Patients: multisite dose escalation; intensive CV monitoring; FDA review after each dose escalation |

guidance document [3]. Many companies start evolving a TPP even at the earliest go/no go decision gates of drug development. In this way, a common way of thinking is preserved throughout the life of the product. The development program plan is assembled by determining what studies need to be done and their design that would provide information critical to answering the key questions at each go/no go decision gate.

The value of a solid drug development plan based on a decision gate approach can be leveraged only if there is discipline in the decision-making process. What method will be used to make a decision: a majority vote of a committee or a single decider who is advised by others? Each stakeholder needs to understand clearly his or her role in making the decision at each gate. Is he or she a decider, a consultant, or just someone who needs to know what the decision is in order to do the job effectively? Go/no go decisions need to be made when all information required to answer the key questions is available. Go/no go decisions should not be revisited. If this occurs, there is either a lack of decision discipline or key information was missing when the original decision was made, representing a failure in planning.

Some go/no go decision gates require agreement to proceed by different groups of stakeholders. This is particularly true when the question "Is the drug candidate safe to give to humans?" is addressed. The sponsor must decide whether to file an investigational new drug (IND) application based on the data and information collected so far from animal tests and in vitro assays. The same data are also evaluated by the regulators (e.g., the FDA), who must have time (30 days) to object if they feel the safety of the subjects in the first few clinical trials will be compromised unduly. Finally, the data are reviewed again by the IRB, who look specifically at how safety will be evaluated and managed in the clinical trials and who represent the interests of the volunteers. It is as if the gate has three locks, each with a key owned by a separate entity which will decide independently whether they open their lock. One cannot pass through the gate until all three locks are open.

## ROLE OF BIOMARKERS IN DISCOVERY AND PRECLINICAL DEVELOPMENT

Less than 10% of drug candidates that enter clinical testing result in a marketed product [4]. This means that at some point during clinical development over 90% of drug candidates fail. If failure occurs late in clinical development, up to hundreds of millions of dollars and a great deal of time and effort will have been invested, with little or no return. Proper incorporation of biomarkers in drug development strategy enables the concept "fail fast, fail early." Early failure actually lowers overall risk because it enables one to move resources and utilize available patients on other promising therapies.

Biomarker data can be critical to a go/no go decision. By definition, biomarkers may reflect the biology or the progression of disease and/or the effect

of drug treatment. Therefore, information provided by properly selected biomarkers can greatly influence the decision as to whether or not to progress through a go/no go decision gate. The challenge is to identify relevant biomarkers early enough to implement them for go/no go decisions at the critical early stages of the discovery–development process.

Biomarkers in discovery are valuable tools for understanding the pathobiology of a disease and the pharmacology of a target and/or compounds (hits or leads) under investigation. Moreover, even at this very early stage, a biomarker may already be identified as a potential clinical safety or efficacy biomarker. Decision gates in discovery are frequently directed toward selection of a lead or a small number of leads to take forward to preclinical development, and prioritization of the leads. For this purpose, a number of commonly employed screening assays can generate useful biomarker data. For example, membrane permeability coefficients and metabolic stability data from *Caco-2* and hepatocyte screening assays, respectively, are predictive biomarkers of in vivo intestinal absorption and liver metabolism, and can be used to weed out compounds unlikely to have acceptable pharmacokinetic (PK) properties for the clinical use intended.

During preclinical development, the focus of studies is on revealing potential toxicity of the drug candidate. Accordingly, there is interest in identifying and monitoring safety biomarkers that can be used for go/no go decisions. Particularly for classes of compounds whose development has been plagued by unusual findings in animals that have not translated into human risk (e.g., fibrates and peroxisome proliferation in rodents), safety biomarkers that translate well across species are highly desirable but not always easy to find.

In both discovery and preclinical development, the abundance of new technologies, high content screening, imaging, genomics, proteomics, metabonomics, and systems biology, to name a few, has opened the door to the identification of theoretically innumerable novel biomarkers along with an expectation that this will result in enhanced efficiency and better decision making. However, the technologies, and the biomarkers that they produce, have to be evaluated as to how they provide insight into understanding the biology and pharmacology being investigated, and how they can be used effectively by scientists and managers in decision making. As discussed in Chapter Imaging as a Localized Biomarker: Opportunities and Challenges, increased volume and complexity of data require sophisticated data analysis and informatics infrastructure to be in place before the new knowledge generated can be applied meaningfully to decision making.

## ROLE OF BIOMARKERS IN EARLY CLINICAL RESEARCH

Biomarkers may be the only way to address a decision gate that asks the question: Is there evidence that the drug is working in humans using the same

mechanism of action defined by animal studies? This question is answered by a clinical proof-of-concept study: often, a phase IIa study in a small number of patients. However, occasionally, clinical proof of concept can be addressed during phase I development if the mechanism of action can be demonstrated in healthy volunteers (e.g., effects on blood pressure, body weight, lipids). The changes in the data collected during this study do not achieve statistical significance; rather, there needs to be enough indication of potential efficacy to convince decision makers to open the gate and spend the resources to progress the drug further. Often, the minimal result for progression can be defined ahead of time, making the actual decision-making process much easier, more objective, and transparent.

Early in clinical development, multiple biomarker assays or technologies may be utilized in order to understand more fully the safety and actions of the drug in humans. One may use a collection of biomarker methods, most of which are not likely to be fully validated surrogate markers of effect. For early decision making, it is not necessary to have results that meet the same standards of accuracy and quality as those used for drug approval. Indeed, it is during these early trials that promising tests for use in later clinical development or marketing are first identified from among the pool of available experimental biomarkers. Resources can then be put to further developing the selected biomarker methodology to meet the quality standards expected for continued future use.

Knowledge of the biology of the disease may be evolving at the same time as the drug is being developed. Therefore, new biomarkers may need to be added during the development process as the science evolves. It should not be assumed that the "tried and true" biomarkers are the only ones worthy of consideration or that a biomarker that was used to obtain marketing approval in the past is the one that should be used today. Good science supported by convincing data can convince stakeholders that a novel biomarker is better than one already established.

## ROLE OF BIOMARKERS IN PHASE III AND POSTMARKETING DECISIONS

Traditional clinical trial endpoints, such as morbidity and mortality, often require extended time frames and may be difficult to evaluate. Fully validated surrogate markers can be used as primary endpoints for pivotal phase IIb and III safety and efficacy trials. Lowering blood pressure and lowering LDL cholesterol are both examples of surrogate biomarker endpoints upon which several drugs have been approved, because it has been demonstrated that effects on these markers directly affect morbidity and mortality several years later.

New technologies are now rapidly emerging that show similar promise in providing efficacy signals earlier in the treatment of certain slowly progressing diseases. Imaging-based biomarkers are providing objective endpoints that may be confidently evaluated in a reasonable time frame. Imaging techniques tend to be expensive, but can be cost-effective when used in well-defined situations where subjective assessment has been the only approach available. As discussed in greater detail elsewhere in this book, examples of therapeutic areas where imaging is reshaping clinical trial design include Alzheimer disease, pain, and osteoarthritis. In affected joints, as well as in tissues expressing specific receptors, magnetic resonance imaging, x-ray computed tomography, position-emission tomography, and single-photon-emission computed tomography imaging are delivering new information to clinicians and researchers.

Patient selection is another area where decisions are made using biomarker information. For example, in developing Herceptin, patients who tested negative for the *Her2Neu* receptor responded poorly to drug treatment, whereas good effects were observed in patients who expressed the receptor. This is consistent with the mechanism of action of the drug and resulted in regulatory approval for patients positive for *Her2Neu*. The outcome was that a diagnostic test was required before this drug could be marketed effectively. So this is now a biomarker that drives the treatment decisions of breast cancer oncologists around the world.

Identifying potential responders or nonresponders prior to phase III clinical trials can have a profound effect on the size of a clinical trial. Eliminating nonresponders can dramatically reduce the intersubject variability in response, thereby reducing the number of subjects in each group of a pivotal clinical trial required to demonstrate effect. The result is a saving in time and large savings in costs. However, if the biomarker is novel, a diagnostic will need to be developed along with the drug before one can market the product effectively.

## SUMMARY

Biomarkers have the potential to provide valuable information that can be translated into knowledge to aid decision making at all stages of the drug discovery–development process. Different stakeholders have different perspectives of risk associated with development and clinical use of a drug. Although biomarkers cannot eliminate the inherent challenges to integrating risk assessment in this complex environment, the associated data can be evaluated objectively. As informatic tools evolve to analyze more efficiently and effectively the large data sets associated with the new biomarker technologies, decision making, in turn, will become better. Regardless of the technologies employed and the complexity of data, a disciplined planning and decision-making process is essential to successful drug development.

## REFERENCES

1. Merriam-Webster (1988). *Webster's Ninth New Collegiate Dictionary*. Merriam-Webster, Inc., Springfield, MA.
2. Pritchard JF, Jurima-Romet M, Reimer MLJ, Mortimer E, Rolfe B, Cayen MN (2003). Making better drugs: decision gates in non-clinical drug development. *Nat Rev Drug Discov*, 2:542–553.
3. FDA (2007). Guidance for Industry and Review Staff: Target product profile—a strategic development process tool. http://www.fda.gov/cder/guidance/6910dft.htm.
4. Kola I, Landis J (2004). Can the pharmaceutical industry reduce attrition rates? *Nat Rev Drug Discov*, 3:711–715.

# IMAGING AS A LOCALIZED BIOMARKER: OPPORTUNITIES AND CHALLENGES

Jonathan B. Moody, Ph.D.
*INVIA Medical Imaging Solutions, Ann Arbor, Michigan*

Philip S. Murphy, Ph.D.
*GlaxoSmithKline Research and Development, Uxbridge, UK*

Edward P. Ficaro, Ph.D.
*INVIA Medical Imaging Solutions, Ann Arbor, Michigan*

## INTRODUCTION

Medical imaging is well known for its ability to produce pictures of patient anatomy for the diagnosis of disease, but perhaps more significant for drug development is its ability to produce images of physiological and physicochemical processes in a living subject. It is the latter that gives rise to the concept of imaging biomarkers, which we may define as the subset of all biomarkers that utilize medical imaging technologies to acquire data. Medical imaging technologies include a range of modalities used in a routine clinical setting: x-ray computed tomography (CT), positron-emission tomography (PET), single-photon-emission computed tomography (SPECT), magnetic resonance imaging (MRI), ultrasound, and optical imaging.

In its 2006 Critical Path Opportunities Report, the U.S. Food and Drug Administration (FDA) defined *biomarkers* as "measurable characteristics that

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

reflect physiological, pharmacological, or disease processes in animals or humans. Changes in biomarkers following treatments reflect the clinical response to the product. Biomarkers can reduce uncertainty by providing quantitative predictions about performance" [1]. In this chapter we will discuss imaging as a localized biomarker. Although imaging biomarkers hold great potential for accelerating drug discovery and development, in practice they are often difficult and costly to implement for specific applications, may be used inappropriately in a development setting, or may fail to provide a robust answer to the development questions being asked. We focus our discussion on understanding the opportunities for using imaging biomarkers as tools in drug development while recognizing the current challenges and how efforts to overcome these challenges are evolving.

We begin by presenting a brief overview of biological imaging. Then we discuss the general characteristics and context of an imaging biomarker along with key advantages and limitations of imaging biomarkers. We place the definition of a biomarker given above into the context of biological imaging in general and diagnostic medical imaging in particular. We follow this with an overview of the scope of imaging in drug development, highlight specific examples of established and emerging imaging biomarkers from oncology that have found utility in drug development and illustrate considerations for imaging biomarker selection. A key issue in the implementation and use of imaging biomarkers is the challenge of standardization which we then examine in some detail. We conclude by discussing current efforts to address the unique needs of imaging biomarker development.

## OVERVIEW OF IMAGING IN BIOLOGICAL RESEARCH

From their inception, the various noninvasive imaging modalities have been used widely in basic biological research. The broad utility of imaging in biological research reflects the breadth of fundamental physical processes that underlie all imaging technologies. The same types of probes and methods of detection that are used in biochemistry and in molecular and cellular biology [e.g., optical, nuclear (gamma emission), nuclear magnetic resonance (NMR) spectroscopy, x-ray] also form the basis of imaging technologies. When the electron microscope was introduced in the 1930s, the morphology of submicroscopic cellular structures that had only been inferred by indirect methods such as x-ray diffraction or structural chemistry could be visualized directly for the first time in electron micrographs [2]. By analogy, the development of medical imaging technologies (e.g., SPECT [3], PET [4,5], CT [6], and MRI [7,8]) has provided the ability to visualize in vivo biological processes previously accessible only by indirect or invasive means. As the science of biological imaging has grown, the development of new imaging applications has been driven by the unique ability of noninvasive imaging to reveal information

about anatomical structure and physiological function as well as various disease states in living subjects.

It will be helpful to review briefly the use of imaging in biological research in order to understand the context and source of imaging biomarkers. We describe some common clinical imaging technologies and then highlight some of the image-based methods for measuring biological function that have been proposed in the literature, providing a background for detailed examples. Several examples of biological imaging applications in drug discovery are also provided in a recent monograph [9].

## Clinical Imaging Technologies

Although we limit our discussion here to nuclear PET and SPECT, CT, and MRI modalities, the following applies equally well to other imaging technologies, including ultrasound, which is widely used clinically, and various forms of optical imaging which are becoming more prevalent [10]. An accessible general introduction to imaging technologies is available [11], as is a comprehensive treatment of the theoretical basis of imaging [12].

CT, the most commonly used clinical imaging modality, provides three-dimensional images of anatomical structure. The main source of contrast in CT images arises from the differential attenuation of x-rays by tissues in the body. In CT images, mineralized tissues appear brighter due to higher attenuation of the x-rays, and soft tissues have intermediate intensity due to lower attenuation. Because of limited contrast differences between soft tissues, an injectable iodine-based contrast agent is often necessary to better highlight anatomical lesions. Despite this limitation, CT provides the highest spatial resolution (less than 0.5 mm), and acquisition of CT data is straightforward and rapid.

MRI may be characterized as a large family of methods to detect molecular- and cellular-dependent image contrasts through the manipulation of nuclear spin magnetic resonance during data acquisition. The proton nuclei of hydrogen in water and lipids are the most commonly imaged, although in principle the NMR basis of detection permits imaging or spatially localized spectroscopy of other nuclei, such as $^{13}C$, $^{19}F$, $^{23}Na$, and $^{31}P$. The sensitivity of MRI is relatively poor, generally requiring tens of micromolar to millimolar volumes of nuclei for detection. However, excellent soft tissue contrast with high spatial resolution (ca. 1 mm) may be obtained. Acquisition of MRI data is controlled by a pulse sequence, which is a computer program that orchestrates the transmission of radio-frequency pulses and magnetic field gradients, as well as the reception of radio-frequency signals from the subject [13]. Vendor-supplied pulse sequences provide a broad range of available acquisition methods; however, additional pulse sequences and improved methods are constantly being introduced in MRI research.

While MRI applications tend to be focused on the seemingly limitless variations in MR acquisition methods, data acquisition in nuclear imaging (PET

and SPECT) is relatively straightforward, being based on either detection of single gamma photons (SPECT) or coincident collinear pairs of gamma photons (PET). Image contrast in nuclear imaging relies on the use of targeted biomolecular probes or tracers labeled with radionuclides to detect in vivo molecular processes and biological function. Thus, the scope of radionuclide imaging is limited primarily by the ability to efficiently radiolabel biologically relevant molecules, and applications are centered primarily on the skill and creativity of the radiochemist. In general, the sensitivity of nuclear imaging methods is the highest of all clinical modalities, with the picomolar sensitivity of PET being about two to three orders of magnitude higher than that of SPECT [14]. The low intrinsic resolution of PET (ca. 5 mm) and SPECT (ca. 15 mm) has been mitigated somewhat through the development of multimodality image acquisition combining PET/CT or SPECT/CT [15], and more recently PET/MRI [16], which provide high-resolution anatomical images registered with functional nuclear images.

In all imaging modalities, the resulting images must be reconstructed from the data acquired using computational methods of varying complexity [17,18], and additionally, in some cases, a series of images may be used to calculate parametric images representing image-based measures corresponding to particular biological functions.

## Image-Based Measures of Biological Function

Many imaged-based measures of biological and physiological function have arisen in imaging and medical research. For example, the Center for Biomarkers in Imaging (http://www.biomarkers.org) makes available an online catalog of approximately 350 image-based biological measures that have been investigated primarily at Massachusetts General Hospital in Boston. Apart from differences in modality, such image-based measures may be classified into three general types according to the way in which image contrast is generated: by endogenous biophysical mechanisms, exogenous contrast agents, or molecular probes [19]. For example, endogenous contrast can be produced by differences in MRI relaxation times such as $T_1$, $T_2$, $T_2^*$, macromolecular cross-relaxation, or transport mechanisms such as blood flow, perfusion, and diffusion. These contrast mechanisms are often enhanced by the use of exogenous contrast agents in CT and MRI as well as ultrasound. Table 1 lists some examples of endogenous and contrast agent–enhanced biological measures using MRI. Numerous other MRI approaches providing image-based measures of disease processes are available [20], as well as MRI-based measures specific to diseases of the brain [21].

Imaging with the use of molecular probes, known as *molecular imaging* [22], also involves the injection of an exogenous agent to provide image contrast. In this case the imaging agent is designed for cellular uptake or binding with a high affinity for a specific in vivo molecular target or process [23]. Molecular imaging is able to provide new information on disease processes not available

**TABLE 1 Examples of Image-Based Measures of Biological Function in MRI Using Endogenous Mechanisms or Exogenous Contrast Agents**

| MRI Imaging Method[a] | Image-Based Measure or Parameter | Biological Function or Parameter | Reference |
|---|---|---|---|
| DSC bolus tracking | $T_2$- and $T_2^*$-weighted intensity changes | Relative cerebral blood volume, relative cerebral blood flow, mean transit time | [149] |
| BOLD | $T_2$- and $T_2^*$-weighted intensity correlation with stimulus | Task-induced functional activation in the brain | [195] |
| ASL | Difference between spin-inverted and control images | Blood flow | [149] |
| Chemical shift encoding | Phase-shifted images from water–fat chemical shift | In vivo water distribution, in vivo fat distribution | [196] |
| $T_2$-weighted MRI | Integral increase in $T_2$ values | Peripheral edema | [70] |
| | Weighted sum of biexponential $T_2$ values | Liver iron concentration | [197] |
| MRTI | Temperature-dependent proton resonance frequency shift | In vivo temperature distribution | [198] |
| CEST/ PARACEST | Amide proton transfer | Spatial distribution of pH | [199,200] |
| CEST | Amide proton transfer | Glycogen concentration | [201] |

[a]DCS, dynamic susceptibility contrast; BOLD, blood-oxygenation-level-dependent; ASL, arterial spin labeling; MRTI, magnetic resonance temperature imaging; CEST, chemical exchange saturation transfer.

from other methods. For example, molecular imaging was used recently to establish the relationship between inflammation and microcalcification in early-stage atherosclerotic plaques [24]; it has been used to visualize the in vivo spatial distribution and time course of vascular endothelial growth factor receptor expression in a model of peripheral arterial disease [25]; and it has been combined with conditional disease models using transgenic mice and genetic profiling to yield a powerful system for in vivo preclinical study of disease processes and potential therapeutic targets [26–29]. The most common molecular imaging applications utilize radiolabeled tracers for detection by the nuclear modalities. However, many nonnuclear approaches have also been developed with optical [10], MRI [20,22], and ultrasound modalities [30,31].

The majority of such image-based measures are distinguished from direct imaging of anatomy by being connected indirectly to the biology. A biological,

biophysical, or molecular model is introduced to provide a bridge from the specific image contrast to the physiology, and a correlation is typically established using invasive (nonimaging) or histological measures. In the case of molecular imaging, the model is represented by the underlying signal generation and targeting mechanisms of the probe, which may be verified independent of the imaging experiment. In other cases, the model is implicit and an empirical correlation provides only a phenomenological link between the image-based measurement and a physiological parameter or process. The invasive or histological correlate often provides an initial rationale for the imaging approach, and may guide development of a model and subsequent interpretation of the measurements.

It is helpful to think of such methods as image-based noninvasive assays, in the sense that the measurements extract or infer some biological, physiological, or molecular characteristic from the subject as a whole. Although not always clinically diagnostic for disease, such image-based assays can often provide the ability to monitor disease progression [19]. The associated biological or biophysical models provide widely varying degrees of linkage between measurement and biology. Molecular imaging provides perhaps the most direct linkage since the characteristics of target, signal generation, and the probe's affinity for the target can be determined empirically, leading to relative confidence in the functional or molecular basis of the image-based signal. The use of [$^{18}$F]fluorodeoxyglucose, a glucose analog, as a radiotracer in PET imaging relies on the empirical facts of glucose uptake and phosphorylation by metabolically active cells. In other cases, a single image-derived parameter corresponds to a complex of multiple functional parameters. For example, kinetic modeling of dynamic contrast-enhanced (DCE) MRI data produces a model parameter, $K^{\text{trans}}$, that can be influenced by several factors, such as blood flow, vascular permeability, and vascular surface area, potentially complicating interpretation of this image-based measure.

Even when an explicit biophysical model is available, as is the case for magnetization transfer MRI [32], it may be impractical to acquire the image-based measurements necessary to apply that model. In this particular case, a related phenomenological model has proven useful to relate a simpler measurement (the magnetization transfer ratio) to the diffuse macromolecular changes and demyelination associated with multiple sclerosis [33]. The lack of specificity of some image-based measures warrants caution in their use prior to adequate assay validation. Early ex vivo NMR measurements of bulk $T_1$ and $T_2$ relaxation of tumors [34] were initially believed to distinguish malignancy from nonmalignancy, and even inspired development of the first whole-body MRI system [35]. However, these results have not subsequently be validated in vivo, although several other MRI methods for characterizing tumors and evaluating therapy have since been developed [36]. Anatomical and structural imaging can also provide functional measures if appropriate image quantification is applied, such as measuring tumor size or volume, which leads to an image-based measure of tumor growth.

Despite the exploratory nature of many imaging research applications, the diversity of image-based functional measures often furnishes much of the interest and enthusiasm for potential imaging biomarkers. However, numerous additional factors must be addressed to establish the reliability, validity, and suitability of a particular image-based measurement as a tool for evaluating novel therapeutics.

## ANATOMY OF AN IMAGING BIOMARKER

Like the drug development process itself, developing, implementing, and using imaging biomarkers is a highly collaborative multidisciplinary effort. As depicted in Figure 1, an imaging biomarker may be defined conceptually by two aspects: image acquisition and image analysis. *Image acquisition* refers to all activities and technologies related and leading to the acquisition of image data, such as imaging hardware operation, maintenance and quality control, subject positioning, contrast agent or radiotracer preparation, injection protocols, and physiological monitoring. *Image analysis* includes all postacquisition activities, such as image reconstruction, processing, quantification, and interpretation. These two aspects of an imaging biomarker require significant

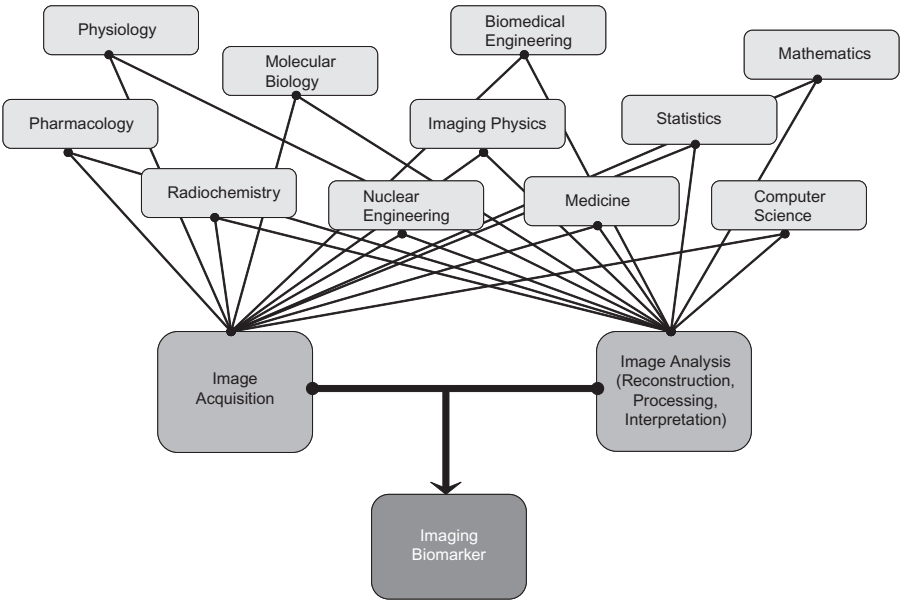


**Figure 1** An imaging biomarker may be defined by its methods for image acquisition and analysis. Both aspects require input from numerous scientific and engineering disciplines.

input from a diverse array of scientific and engineering disciplines, including biology, chemistry, physics, biomedical and nuclear engineering, computer science, and mathematics as well as medicine. Each of these specialties has several critical roles in the implementation and use of imaging biomarkers:

- *Biologist/physiologist/molecular biologist/pathologist:* help establish the biological relationship between the image-based measurement and the physiological function or disease process; help establish the rationale for studying particular image-based measures; and help interpret the quantitative results.
- *Pharmacologist/physician/pathologist:* help establish the clinical relationship between the imaged-based measurement and the relevant therapeutic outcomes during treatment; and help understand all possible actions of the treatment (beneficial and deleterious) [37] to aid in the interpretation of imaging results.
- *Biochemist/radiochemist:* provide understanding of physiological processes and molecular targets at the biochemical level; and design, synthesize, characterize, and test optical, magnetic, and molecular probes and radiotracers.
- *Mathematician/statistician/computer scientist:* provide the fundamental mathematical, statistical, and quantitative basis for image acquisition, reconstruction, and analysis; and provide software tools for performing many tasks associated with image acquisition, reconstruction, quantification, analysis, visualization, and interpretation.
- *Imaging physicist/biomedical and nuclear engineer/technologist:* provide the physical basis for imaging technologies, signal generation and detection, hardware design, and optimization; provide the technological expertise for isotope production to be used for generating radiotracers; and operate the imaging equipment, performing image acquisition, processing, and analysis tasks.

## Essential Characteristics of Imaging Biomarkers

Earlier we highlighted the use of imaging in biological research. In such studies, when a correlation is demonstrated between an image-based measurement and a particular physiological characteristic or disease process, it is tempting to identify that measurement as an imaging biomarker. However, this is not always useful. Here we identify several additional characteristics of imaging biomarkers that will generally become essential as an image-based functional measure progresses through the process of validation and clinical qualification (Figure 2).

***Scientific Characteristics*** For an image-based measurement to be an effective tool in drug development, the "model" linking the measurement to a

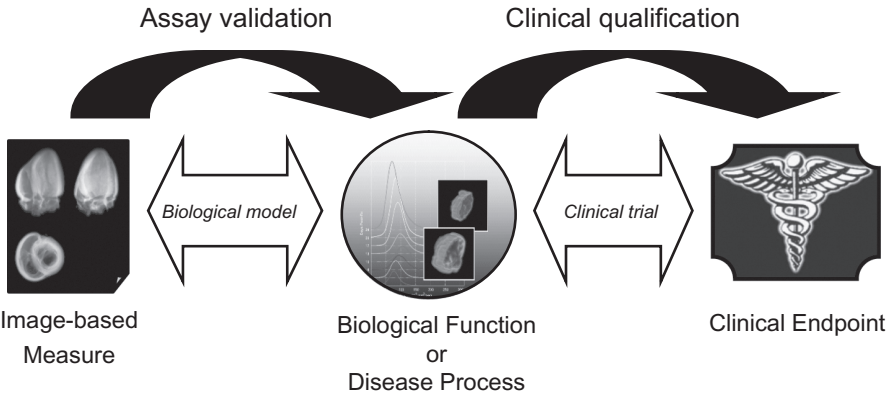


**Figure 2** Two-step process for imaging biomarker validation/qualification. *Assay validation* establishes the link between the image-based measure and a specific biological function or disease process by way of a biological, biophysical, or molecular "model." This includes characterizing the measurement's intra- and interobserver and test–retest variability (measurement reliability), as well as correlating the measurement to an established (nonimaging or invasive) standard. The goal of the second step, *clinical qualification*, is to establish empirically, by clinical trials, the relationship between the image-based functional measure and a relevant clinical endpoint or outcome. (*See insert for color reproduction of the figure.*)

biological or disease process must be characterized, which is termed *assay validation*. This process includes:

- Characterizing the validity of the model by correlation with an accepted standard
- Characterizing the reliability of the measurement:
  - Intraobserver variability
  - Interobserver variability
  - Test–retest variability
- Defining the conditions under which the image-based measurement will result in the expected model validity and measurement reliability:
  - Biological conditions
  - State and stage of disease
  - Class of therapy
- Optimizing the image acquisition and analysis systems for the task of parameter *estimation* as opposed to *classification*

Basic research studies of image-based biological measures may provide only partial validation by attempting to establish the validity of the model and occasionally reporting some measure of reliability in a limited sample size,

often in preclinical studies. Additional development may be necessary to fully characterize and optimize the measure in both animal models and in humans. Like the underlying image-based methodology, the imaging biomarker is necessarily quantitative, which requires parameter estimation from the images, a process that should aim to be as efficient and objective as possible.

***Clinical Characteristics*** In addition to the scientific concerns of assay validation, several clinical issues must be addressed. The relationship of an image-based measure of biological function to a specified clinical endpoint must be established empirically through clinical trials, which may be termed *clinical qualification* (Figure 2). Clinical qualification seeks to:

- Define a relevant clinical endpoint or outcome
- Establish empirically the relationship between the image-based measure and the clinical endpoint
- Characterize the propagation of measurement uncertainty to uncertainty in clinical outcome
- Establish the safety of the imaging procedure

These aims are often hindered by the need for large amounts of data to support clinical qualification. The imaged-based measure may be more sensitive to biological changes than the corresponding clinical endpoint, reflecting fluctuations that are not clinically relevant; or the clinical endpoint might be somewhat subjective, subject to large variability, or require lengthy trials.

***Additional Criteria*** Apart from the scientific and clinical issues of biomarker qualification, further logistical and economic concerns must be addressed as an imaging biomarker progresses through clinical qualification.

- The necessary acquisition hardware, methods, and image analysis tools must be readily available.
- The image acquisition and analysis methods must be standardized across hardware platforms and imaging centers.
- The imaging biomarker must be cost-effective and efficient compared to other available imaging and nonimaging biomarkers with similar purpose.

***Development Stages*** It is clear that most imaging biomarkers today fall short of achieving all of these characteristics. To describe the development stage of imaging biomarkers, the FDA has identified four distinct classes of imaging biomarkers which are focused on drug development applications: *pre-biomarkers, biomarkers, surrogate biomarkers*, and *clinical diagnostic surrogate biomarkers* [38]. According to this classification, *pre-biomarkers* are those at an incomplete stage of assay validation; *biomarkers* are those that have completed assay validation along with an appropriate safety profile; *surrogate*

*biomarkers* are those for which sufficient clinical evidence exists to qualify the biomarker for use in drug development; and *clinical diagnostic surrogate biomarkers* are those that have been validated and qualified to such an extent as to be "integrated into an approved/licensed therapeutic regimen" and are able to "assure/improve safety and/or efficacy" [38].

Imaging biomarkers are useful tools for a variety of tasks in addition to drug development. Therefore, another perspective, which perhaps includes a wider spectrum of methods, is to stage imaging biomarkers according to their intended use [39]. There are at least four distinct *categories of use* for imaging biomarkers in the wider sense:

1. Elucidate biological function and disease processes in vivo (*basic biological research*).
2. Evaluate novel therapies in vivo for industry decision making (*drug development—industry decisions*).
3. Evaluate novel therapeutics in vivo for government licensure (*drug development—government registration*).
4. Enhance routine diagnosis of disease with better disease management, prognosis, and therapeutic decision making (*clinical practice*).

The four stages correspond roughly to the FDA imaging biomarker classes described above in terms of the level of validation and qualification within each stage (Table 2). They also parallel the conceptual biomarker hierarchy

**TABLE 2 Classification of Imaging Biomarkers According to Their Intended Use[a] and the Corresponding Stage of Validation and Qualification**

| Stage | Imaging Biomarker Intended Use | Assay Validation | Clinical Qualification | Additional Criteria[b] | Potential Impact[c] |
|---|---|---|---|---|---|
| 1 | Basic research | Partial | No | No | * |
| 2 | Drug development: industry decisions | Yes | Partial | No | ** |
| 3 | Drug development: government registration | Yes | Yes | Partial | *** |
| 4 | Clinical practice | Yes | Yes | Yes | **** |

[a]Intended uses are "backward compatible," in that an imaging biomarker that achieves a given stage of validation and qualification will fulfill the requirements for all preceding uses. Note that regulatory scrutiny would be limited to imaging biomarkers at stages 3 and 4.
[b]Nonscientific and nonclinical criteria, including logistical and economic issues such as availability of technology, standardization of methodology, and cost-effectiveness.
[c]At each development stage, the potential impact of an imaging biomarker on drug discovery and development and on human health.

suggested by Frank et al. [40] using the body of evidence supporting each class: *plausibility, correlation/association, predictive power,* and *cause*. The categories of "probable valid biomarker" and "known valid biomarker" suggested by the FDA [41] would be contained in stage 2, while stage 3 corresponds to "surrogate endpoints" [39].

These four categories of use emphasize that there may be an appropriate context and use of imaging biomarkers at each stage of validation or qualification. Basic biological research provides fundamental knowledge of disease processes as well as a ready source for new biomarker discovery; the use of imaging biomarkers in clinical practice clearly overlaps with the routine use of imaging in clinical diagnosis. The significance of these potential imaging biomarker uses is that they can be leveraged to facilitate further development and adoption of imaging biomarkers as drug development tools, particularly as surrogate endpoints for regulatory decision making.

## Imaging Biomarkers and Diagnostic Imaging

The foremost application of medical imaging is the routine clinical diagnosis of disease. Historically, the development of imaging systems capable of scanning humans followed shortly after the initial technological development of the various imaging modalities [3,5–7,35,42–44]. Moreover, medical imaging applications emphasizing medical diagnosis have driven much of the technological improvements to the present day.

Although the primary diagnostic purpose of medical imaging overlaps considerably with imaging biomarker applications, there are important distinctions between the context and goals of their use [45]. Diagnostic imaging is generally focused on detection of lesions or abnormality, and image assessments are usually made in a subjective, qualitative manner relying on the training and experience of a medical professional. In this case, the primary goal is the medical management of individual patients. By contrast, imaging biomarkers are focused on the quantitative characterization of disease which is generally already known to be present, with the aim of objectively measuring effects related to a medical treatment or intervention. Typically, serial measurements from one or more follow-up imaging sessions are compared to a baseline measurement, and measurements are collected from patient groups drawn from well-defined patient populations.

Why is it important to draw a distinction between imaging biomarkers and diagnostic imaging? It is widely recognized that the most reliable and objective approach to assess image quality, and thus optimize the imaging process, is, first, to specify the purpose or task for which the image is produced, and second, to determine quantitatively how well that task is performed [46]. Two general categories of tasks that relate to medical imaging are *classification* tasks, which are the typical diagnostic tasks of deciding whether a patient has or does not have disease, and *estimation* tasks, which involve estimating quantities of interest from medical images [46]. An example of a classification task

would be an oncologist using images to detect or locate abnormal lesions, and attempting to answer the questions: Are the lesions malignant or nonmalignant, and if malignant, what is the grade or stage of the cancer? Examples of estimation tasks would be a cardiologist estimating the extent and severity of a myocardial perfusion defect or an oncologist estimating the metabolic uptake of [$^{18}$F] fluorodeoxyglucose in a tumor.

Diagnostic imaging platforms and analysis tools are by definition concerned with classification (detection) tasks and have been uniformly optimized for those purposes. It should be clear that imaging biomarkers, being inherently quantitative, are concerned with either estimation tasks or combined classification/estimation tasks [47,48]. Thus, the requirements and standards for diagnostic imaging systems are somewhat different from that of imaging biomarkers. For example, the demands of image quantification and the extraction of objective image features for imaging biomarkers place strict limits on image quality that may differ from that required for diagnostic purposes, where the experience and skill of the clinician may compensate for typical nonidealities in the images. Evidently, qualitative visual image evaluation suffices for a majority of diagnostic imaging purposes. On the other hand, objective quantitative image assessment is essential for imaging biomarkers, and image acquisition and analysis systems must be optimized for this purpose. Such optimization may involve differences in hardware design, evaluation of image quality control, image reconstruction, processing and quantification methods, and software.

***Nuclear Cardiology as an Imaging Biomarker*** SPECT myocardial perfusion imaging (MPI) is a nuclear cardiology procedure developed for the clinical diagnosis of coronary artery disease [49]. The imaging procedure is based on cellular uptake of a gamma-emitting radiopharmaceutical (such as thallium-201 or various technetium-99m-labeled agents) in proportion to the regional blood flow in the myocardium [50]. The imaging procedure is typically performed twice: first, with the subject at rest, and second, after inducing hyperemia through exercise or pharmacological stress. The resulting images provide a relative measure of myocardial perfusion and allow the identification of perfusion defects. In general, perfusion defects may represent regions of possible ischemia (reversible defects) or necrotic tissue resulting from infarction (scar or fixed defects). Quantitative analysis of rest and stress images provides estimates of the perfusion defect extent and severity as well as discrimination between ischemic and scar regions [51]. In addition, numerous other quantitative measures of cardiac function can be derived when the images are acquired with electrocardiogram (ECG) gating. Such image-based measures play a critical role in guiding subsequent medical intervention and clinical management of coronary artery disease.

As a mature imaging technology, SPECT MPI has been developed and improved continuously for nearly 25 years, and apparently has many characteristics of a qualified stage 4 imaging biomarker:

- Extensive validation of image-based quantitative measures of myocardial perfusion (perfusion defect type, extent, and severity)
- Extensive clinical qualification for the diagnosis of coronary artery disease
- Standardized protocols for image acquisition and reporting
- Standardized automated image processing and quantification methods [52]
- Validated imaging hardware and software tools readily available [53]
- Cost-effective imaging procedures used widely in clinical practice

In addition to the availability of standardized image quantification, an important aspect of SPECT MPI that sets it apart from many other diagnostic imaging applications is the routine use of normal subject databases. Image-based assessment of abnormal perfusion is done by computer-automated comparison of perfusion parameters for a given patient with normal limits defined by databases of subjects with a low likelihood of coronary artery disease [51]. Moreover, the accuracy of automated assessment is comparable to human visual interpretation of the images, but with potentially lower variability [54].

Despite these characteristics and its broad acceptance in diagnostic imaging, SPECT MPI appears to have only limited utility currently as a tool for drug development. As a diagnostic tool, SPECT MPI technology and standards have been optimized for the clinical diagnostic setting. In particular, recent use of SPECT MPI for serial imaging of patients has highlighted limitations and the need for further research to optimize SPECT MPI for serial image assessment [55,56]. Furthermore, despite the availability of more objective quantitative measures of myocardial perfusion deficit, recent clinical trials that have used SPECT MPI evaluation as a primary endpoint have been compelled to use visual categorical interpretations and semiquantitative measures for drug registration [57,58]. We may speculate that these limitations are due in part to imaging systems that have been optimized for the classification task rather than the estimation task required of imaging biomarkers in drug development. For example, the quantitative thresholds that have been established for the use of normal subject databases in SPECT MPI software have generally been optimized by reference to expert visual reads in the single study setting, which may not be optimal for quantitative evaluation of serial studies. Thus, SPECT MPI technology illustrates the subtle, yet important differences between conventional diagnostic imaging and imaging biomarkers.

## Key Advantages of Imaging Biomarkers

As a tool in drug development, imaging biomarkers share several general advantages among all imaging modalities. Imaging biomarkers provide the unique ability to visualize anatomy as well as to spatially localize physiological and molecular function in vivo. Additionally, most imaging modalities have the capability to acquire repeated or temporally resolved images which allow a study of various dynamic biological processes. In vivo image-based evalua-

tion of disease processes or drug targets occurs in their respective biological microenvironments, accounting for both systemic and local modulators. As in vivo assays, imaging biomarkers also share with in vitro biomarkers the ability to determine drug–target interactions, to help optimize drug dose, and to evaluate longitudinally the response to therapy in vivo [59].

A broad range of sensitivity and spatial resolution is also available, providing great flexibility in the choice of applications. High-resolution CT and MRI anatomical imaging biomarkers can evaluate gross therapeutic effects at the level of organ systems; nuclear, CT, and MRI perfusion imaging biomarkers can probe functional effects at a regional tissue and vascular level; nuclear, MRI, and optical molecular imaging biomarkers can provide readouts down to the level of cellular and molecular effects. This range can sometimes be an advantage when considering the process of clinical qualification. For example, although molecular imaging applications may attain the level of detail necessary to test for proof of mechanism, clinical qualification of imaging biomarkers will probably be facilitated by coarser biological "resolution" at the macroscopic level of organ systems which may more readily be related to a clinical endpoint.

In addition to the clinical relevance of the various imaging modalities, there is also the possibility for translation of preclinical imaging applications to the clinic, which provides a potential for continuity of readout parameters between preclinical safety and efficacy studies in animal models and eventual clinical evaluations in humans. Furthermore, a majority of hospitals and medical centers (often, the sites participating in later-phase clinical drug trials) typically have direct access to clinical imaging centers or imaging equipment.

Longitudinal assessment of individual subjects provides a potential for reduced statistical variability, and consequently a reduction in the number of study subjects. Clearly, this can benefit drug development through faster studies and reduced costs; however, in practice, many additional factors can influence the net cost savings and efficiency achieved in using an imaging biomarker. These factors are related to the implementation of imaging biomarkers and must be evaluated carefully for each imaging biomarker application.

## Limitations of Imaging Biomarkers

There are generally three categories of imaging biomarker limitations. First, there are limitations that are likely to be addressed in the near future by the accelerating rate of imaging biomarker development. The complex and multidisciplinary nature of image acquisition hardware and analysis methods has hindered the translation of promising image-based functional measures into validated and qualified imaging biomarkers. Similarly, a lack of standardization has also impeded imaging biomarker development, since variations in methodologies, as well as hardware platforms with differing capabilities, have made it difficult to compare and evaluate potential imaging biomarkers across sites. Compared to in vitro biomarkers, it may be costly and time consuming

to acquire the clinical evidence necessary to establish a biological relationship between an imaging biomarker and a particular disease state. All of these factors have limited the introduction of imaging biomarkers into clinical practice for patient management, diminishing the incentive to commercialize such technologies. Additionally, despite the impressive preclinical accomplishments of molecular imaging to date, expansion of such applications to the clinic has been limited by the fact that relatively few molecular imaging agents have been approved for clinical use in humans [22,60].

However, these limitations are starting to be addressed by novel approaches to imaging biomarker development, broad collaborations to standardize technology, and federal initiatives to foster the development of imaging agents and increase the availability of component tools and methods. In addition, radioisotope production for some important radionuclides, such as $^{18}F$, $^{11}C$, $^{13}N$, and $^{15}O$, although complex and expensive, is beginning to be addressed by recent innovations in cyclotron technology [61].

Second, there are limitations in the underlying imaging technologies and methodologies that we might expect to be addressed by ongoing basic research and technology development already occurring independent of imaging biomarker development. The widespread diagnostic use of imaging motivates continual research and vendor improvements in hardware, software, and acquisition methods. Such improvements have recently included faster imaging and hardware improvements for MRI, innovations that reduce radiation dose in CT, new hardware designs that improve sensitivity and resolution of nuclear imaging modalities, and new capabilities afforded by combining modalities, as in SPECT/CT, PET/CT, and most recently, PET/MRI. Furthermore, many image analysis approaches and tools have been borrowed from ongoing computer vision research. All such advances have the potential to benefit imaging biomarkers' performance; however, careful attention must be paid to the differing criteria for system optimization between diagnostic and biomarker applications.

Finally, there are inherent limitations in imaging biomarkers that are not likely to be fully mitigated by future development. Within the spectrum of potential imaging biomarkers, those that rely on endogenous contrast mechanisms have the advantage of avoiding the added complexity and safety concerns introduced by the use of exogenous imaging agents. However, such image-based functional measures are often inherently nonspecific, relying on biological models in which the connection between measurement and function may not be interpreted easily or directly. In addition, there is a growing perception that the increased use of nuclear and CT modalities, which utilize ionizing radiation, presents a public health concern [62,63]. Diagnostic imaging has been guided by the principle that the risks of ionizing radiation should always be weighed against the benefits of the imaging procedure to the patient [64]; however, in some cases, radiation doses from medical tests may exceed limits known to increase the risk of cancer, particularly in the case of serial imaging [64]. Although progress has recently been made in reducing radiation

doses in CT procedures [65], the image quality of CT depends inherently on a nonnegligible radiation exposure. Additional advantages and limitations specific to different imaging modalities and methods are discussed further in a later section.

## Importance of Context: Disease, Therapeutic Strategy, Development Stage, and Alternative Approaches

The development or implementation of any particular imaging biomarker must be preceded by consideration of the current research environment, the local availability of technology and resources, and an attempt to evaluate the relative utility, efficiency, and cost-effectiveness of alternative approaches. Local implementation of an imaging biomarker is a time-consuming process, even for methods validated and published previously. Commercially available imaging systems are generally tailored toward diagnostic rather than quantitative applications, and as a consequence, the implementation of an imaging biomarker often involves applying customized acquisition, radiotracer synthesis, or analysis methods that must be locally optimized and compared to published results. In a preclinical setting, the implementation process places great demands on the local availability of resources such as MRI pulse sequences and acquisition methods, radiotracer production facilities, appropriate animal models, and image analysis tools, among many others. In a clinical research setting, an imaging biomarker may have to be implemented across multiple centers. An overriding concern in the usability of any particular imaging biomarker once implemented is the overall throughput that can be achieved [66].

Alongside these practical concerns are various scientific questions about which technique among several will be optimal. Within any one imaging biomarker application, there may be multiple methods for acquiring similar data, and multiple approaches for image reconstruction, processing, and quantification. Similarly, there may be multiple imaging or in vitro biomarkers that address the same biological function or clinical endpoint. Such choices typically have to account for incomplete data comparing the methods and limited available resources.

When considering the use of imaging biomarkers, the dominant focus is often on the technology, implementation issues, or necessary tools. Although implementation choices must be weighed carefully, it is important to emphasize that imaging biomarkers, like any other biomarker, should be specified in terms of the disease context and therapeutic strategy, as well as the development stage and questions to be answered. Defining the clinical context of a given imaging biomarker is part of the process of biomarker validation, which answers the question: What are the biological conditions, disease states, and therapies under which the use of the imaging biomarker will result in the empirically determined reliability and validity? For example, the location of a tumor may limit our ability to obtain a robust quantitative assessment of treatment effects due to excessive respiratory or cardiac motion. An imaging

biomarker that is validated and qualified for a given tumor type and therapeutic class cannot be assumed to be validated and qualified for tumors of other types or anatomical locations or other therapies without additional validation data.

Conversely, in drug development, time efficiency and the value of investing in imaging technology resources can be maximized by selecting and implementing cross-mechanism imaging biomarkers [67]. For example, multiple imaging biomarkers may share a common MRI pulse sequence or acquisition method, such as $T_2$*-weighted dynamic imaging for dynamic susceptibility contrast (DSC) bolus tracking and blood-oxygenation-level-dependent (BOLD) functional MRI, or $T_2$ mapping for measuring peripheral edema or liver iron concentration. Similarly, a custom PET radiotracer may be developed for a specific development program but may have utility in other therapeutic areas [45]. Such coordination requires significant planning and broad alignment across discovery and development groups and therapeutic areas, which may be difficult to achieve in large organizations.

## SCOPE OF IMAGING BIOMARKERS IN DRUG DEVELOPMENT

Parallel developments in our knowledge and understanding of biology and disease, and advances in biomedical technologies, have dramatically increased our ability to discover and develop new therapies as well as to evaluate the biological effects of disease interventions. Imaging technologies have played an important role across the spectrum of drug discovery and development [9,20]. Apart from basic biological research, the roles of imaging biomarkers in drug development generally fall into two categories: supporting internal decision making, and to support the demonstration of efficacy in pivotal clinical trials for registration.

Within the category of imaging biomarkers used for decision making in drug development, there are three fundamental applications that imaging biomarkers have in common with in vitro biomarker technologies: confirming that a drug candidate hits the intended biological target, testing whether hitting the target, alters the disease process (proof of mechanism), and testing whether altering the disease process affects the clinical status of the patient (proof-of-therapeutic concept) [40,45]. Imaging biomarkers have provided early critical information at each of these development stage gates through in vivo molecular imaging studies of competitive binding or receptor occupancy [45], imaging of target distribution and radiolabeled drug distribution [68,69], lead optimization, and preclinical and early clinical efficacy testing [19,70]. Other tasks in early clinical development that imaging biomarkers have addressed include appropriate dose and regimen selection and in vivo pharmacokinetic/pharmacodynamic studies [71,72], patient stratification, and study population enrichment [40].

Importantly, imaging biomarkers for internal industry decision making have demonstrated value for drug development without comprehensive vali-

dation or clinical qualification [67]. Indeed, many such biomarkers may be proprietary and specific to a particular development program, therapeutic target, or mechanism [45]. However, achieving a stage of assay validation appropriate to a given task and the associated risk is a crucial but often time-consuming process. The use of novel imaging biomarkers is most effective if development and characterization activities, such as PET radiotracer development or MRI pulse sequence and method development, are begun as early as possible and ideally in tandem with the corresponding drug development program [45].

Although imaging biomarkers have proven useful in many aspects of drug development without being fully clinically qualified, there is renewed interest from multiple stakeholders to develop qualified surrogate endpoints and expedite the lengthy and costly drug approval process [39]. In addition to the well-documented difficulties of clinical qualification [45], the potential impact of such tools for regulatory decision making will be limited unless they achieve some measure of standardization and cost-effectiveness, and become widely accepted and widely available. In the next section we examine in some detail a few of the imaging biomarkers that are currently, or are poised to become, such high-impact tools for drug development in oncology.

## PROFILES OF IMAGING BIOMARKERS FROM ONCOLOGY

### Morphological Evaluation of Tumor Response

There is a broad range of imaging biomarkers applicable to oncology drug development [59]. This spans exploratory imaging methodology that may contribute to early development decision making in small studies (e.g., DCE-MRI assessment of anti-angiogenics [73]) to robust and broadly applicable assessments of disease burden (e.g., RECIST (response evaluation criteria in solid tumors) [74]). RECIST provides a framework to use imaging (predominantly CT) to document baseline disease burden and define change over time.

It could be argued that RECIST should not be considered a biomarker since the aim of RECIST is to provide a broad categorization of response. However, the quantitative components of RECIST, the lesion size measurements, are measurable characteristics reflecting disease and response to treatment, and it therefore seems reasonable to consider it alongside other imaging biomarkers. Furthermore, understanding the attributes of RECIST will help inform the development of other imaging-derived biomarkers: (1) how it was developed; (2) how it became an international standard of assessment; (3) how it addresses the challenges of multisite application, including multiple modalities, acquisition parameters, interpreting radiologists; and (4) how it is implemented into large-scale clinical trials.

***Background*** In the early 1980s the World Health Organization (WHO) published recommendations aimed at standardizing tumor response assess-

ment [75] based on criteria formulated in 1979. The recommendations described minimum requirements for reporting response assessments using clinical, radiological, biochemical, and pathological factors. In terms of radiological assessment, this was to be performed on single or multiple lesions using a bidimensional approach to estimate tumor size. Further, clinical or radiological assessment could be used to review new lesions which would lead to disease progression irrespective of other changes. Four groups were described: complete response (CR), partial response (PR), stable disease (SD), and progressive disease (PD). This thresholding into discrete response categories was subsequently applied widely and continues to this day. Despite the longevity of this simple classification scheme it was noted early that there are significant advantages of reporting continuous tumor size estimates, "especially in the phase II setting" [76]. Although adoption of the broad principles was seen, subsequent modification based on different expert group preferences and tumor-specific requirements resulted in a heterogeneous use of the WHO criteria as described in the commentary by Michaelis and Ratain [77]. It was apparent that a new standard was required for measuring response of solid tumors in clinical trials. This was with a backdrop of dramatic change in standard imaging methodology and application (including improvements in CT, ultrasound, MRI, and PET) that offered potentially many new or improved ways in which response could be evaluated. With such a diversity of potential assessments the new criteria had to specifically address how lesion sizes needed to be estimated to avoid continual diversification of applied methodology.

A collaboration between the European Organization for Research and Treatment of Cancer, the National Cancer Institute, and the National Cancer Institute of Canada culminated in a publication of the new guidelines in 2000 [78]. The new guidelines maintained the classification scheme of CR, PR, SD, and PD but made a number of important changes: adoption of a simplified tumor size estimation (i.e., longest diameter), a definition of how many lesions (target lesions) should contribute to the size estimation, a description of how nontarget lesions should be assessed, and a description of which imaging methodologies should be considered for response assessment. RECIST has been adopted widely and continues to contribute to many endpoints in cancer clinical trials, including progression-free survival.

### *Limitations*

*Tumor-Specific Limitations* As with all such generalized criteria, there are instances when the approach becomes problematic. Since the publication of the original RECIST guidelines, many subsequent publications have detailed the limitations of applying RECIST for certain tumor types and certain therapies, such as cytostatic agents and general comparisons of RECIST with WHO [79]. The review by Therasse et al., in 2006 [80] documents the tumor-specific issues pertinent to RECIST.

It is clear that with challenges associated with different tumor types, modifications to the RECIST framework may provide benefits. Examples include the evaluation of gastrointestinal stromal tumors (GISTs) treated with imatinib. It was found that response by RECIST did not sufficiently translate into survival. However, when response was defined as a 10% reduction in tumor size or a 15% reduction in contrast-enhanced CT density, better linkage with outcome was observed [81,82]. Another example is that of mesothelioma, where Byrne and Nowak describe modifications to the RECIST criteria [83] that improve correlation with both survival and lung function.

*Limitations of Pure Radiological Assessment* RECIST offers a generalized framework for defining response in clinical trials which has been integral to many successful drug development programs. However, there are many instances where radiological evaluation alone is limited and does not typify decision making in clinical practice. Often, the clinical decision making will be based on multiple factors (such as *CA-125* for ovarian cancer), and clinical trial assessments may need to incorporate nonradiological parameters alongside radiological assessment.

*Centralized Radiological Review* One of the recommendations in the RECIST guidelines is that response assessments should be made based on independent assessment, and this is often considered a requirement for regulatory submission. Central and independent image review aims to reduce the variability of assessment and remove or reduce the potential for investigator-introduced bias. Central review of high volumes of imaging data from hundreds of sites across many countries does come with high complexity and high cost to the study sponsor. Furthermore, how do we deal with disagreements between the central reviewer and the site-based assessment? The site-based radiologist is likely to be performing RECIST differently from the central reviewer: selecting a different number of target lesions, defining the longest diameter of the lesion differently, a different judgment on nontarget lesions, and equivocal new lesions may be interpreted differently. When there is subjectivity associated with the application of criteria such as RECIST it is important to identify sources of variability associated with image interpretation such that different results from different readers can be understood.

*Unidimensional Evaluations* A frequently raised issue with RECIST and cause of discordance with WHO is the reliance on a single-dimensional size estimation. Schwartz et al. [84,85] discussed this issue in relation to esophageal cancer. With multidetector CT technology and MRI a three-dimensional lesion assessment can be performed and offers a theoretical advantage in terms of likely sensitivity to change and avoidance of lesion eccentricity problems. However, one of the most significant advantages of RECIST is the simplicity of the imaging required. If three-dimensional imaging was required

in a 100-site clinical trial, there would be many challenges to ensuring that this is done to a level that would allow robust volumetric assessments with acceptable variability. Further, the additional radiology review time (central and site based) required to outline multiple lesions on multiple slices would be substantial. It is likely that three-dimensional assessments will have a clear benefit in a number of applications, especially when deployment of a standardized three-dimensional acquisition can be achieved readily. A significant amount of data is required, however, to demonstrate that the theoretical advantages translate broadly into a measure more predictive of response and that the practical issues can be overcome for multisite clinical trials.

*Imaging Technology* The ability to define meaningful response is strongly influenced by the methodology available and the mass of data that links a method with clinical outcome. RECIST focuses on morphological assessment with CT since the technology is available across all clinical trial centers, provides a relatively consistent acquisition, and is used most widely across the majority of cancers. However, it is recognized that technologies such as FDG-PET as indicators of metabolic response could provide improved response assessments in the future. Owing to the tumor-specific utility of metabolic imaging, it is likely that approaches such as FDG-PET will become very important for some indications, but not broadly applicable. In lymphoma, for example, FDG-PET is an integral component of the current recommendations for response evaluation [86].

***Continual Development of Response Criteria*** Response criteria will need to evolve continually with imaging technologies, whether new molecular imaging approaches become available or there are broad improvements in standard morphological assessment with CT or MRI. After extensive meta-analyses and in response to changes in imaging practice since RECIST was first published, modifications were proposed in 2009 in criteria termed *RECIST 1.1* [202]. One challenge in applying the new criteria will be to ensure consistency within any a given clinical trial. Since oncology trials may often take years to complete, there will be an overlap between use of the original RECIST and the new guidelines.

***Learning from RECIST Development*** There are many important lessons to learn from the way in which RECIST has evolved to become an integral component of many clinical trials involving solid tumors:

- Developing a consensus around response criteria takes a considerable effort requiring multidisciplinary input and broad representation from stakeholders.
- Criteria should be sufficiently broad that revision from year to year is not required if, for example, imaging technologies change.

- To demonstrate a robust relationship to clinical outcome, very significant data sets are required. The new RECIST modifications use data from over 6500 subjects.
- Response criteria applicable to large multicenter trials need to be pragmatic to ensure utility across many different imaging platforms and alignment with country-specific practices and radiologist preferences.

Although RECIST may seem far removed from the concept of an imaging-biomarker it represents the most widely used imaging-based clinical trial methodology in oncology. For this reason, an understanding of how it has evolved and is deployed practically is important to consider when other techniques are being developed.

## [$^{18}$F]FDG-PET

The increased glucose metabolism associated with many cancers has been established for some time and the origin and consequences remain areas of continued discussion [87]. [$^{18}$F]fluorodeoxyglucose (FDG) PET ([$^{18}$F]FDG-PET) exploits this general phenomenon, facilitating the widespread application of a molecular imaging technique for oncology. FDG is taken in cells via glucose transporters (Glut-1), phosphorylated by hexokinase, and becomes trapped. The rate of uptake is related to glucose utilization and therefore an [$^{18}$F]FDG-PET examination interrogates the glucose uptake and subsequent metabolism. The [$^{18}$F]FDG-PET study therefore broadly defines tumor metabolic rate and is also considered to reflect viable cell fraction, although the relationship is not necessarily straightforward [88].

The extensive review by Gambhir et al. outlines the broad use of [$^{18}$F]FDG-PET for diagnosis, staging, assessing recurrence, and monitoring response [89]. It is interesting to note the varied sensitivity to detect primary cancers relative to the generally high avidity to metastases. A more recent review by Kelloff et al. [90] provides an extensive review of [$^{18}$F]FDG-PET applied to clinical practice and oncology drug development. Although the role in all tumor types is yet to be established, such reviews emphasize the high availability of [$^{18}$F]FDG-PET and broad applicability, both of which are important defining points in the role of [$^{18}$F]FDG-PET in drug development.

***Application of FDG-PET to Drug Development*** As explained by Kelloff et al. [90], current data suggest that [$^{18}$F]FDG-PET response is likely to be sufficiently predictive of clinical benefit. Such surrogacy status would broaden the impact of [$^{18}$F]FDG-PET in its ability to be used as a provider of key trial endpoints, including increased support of proof of concept through phase II studies to even the potential to provide a true surrogate endpoint for phase III studies. Although full validation studies are yet to be completed, data from some indications (NSCLC [91] and lymphoma, for example) suggest that [$^{18}$F] FDG-PET can be used successfully to define response to approved therapies

and predict endpoints such as overall survival and progression-free survival [92–94]. Sarcoma is also an indication where [$^{18}$F]FDG-PET response provides a better prediction than tumor size change of histopathological response to neoadjuvant therapy [95]. Published and ongoing studies of response with approved therapies are likely to provide a significant body of data to understand the scope of [$^{18}$F]FDG-PET-based endpoints in clinical trials of many tumor types.

With an increased understanding of [$^{18}$F]FDG-PET responses to standard chemotherapies will come an accelerated application to support the clinical development of novel therapeutics. The demonstration by Stroobants et al. [96] that PET provides an early indicator of metabolic response to imatinib mesylate for subjects with soft tissues sarcoma was a convincing example of how metabolic response can significantly precede any morphological change. Even 48 hours after beginning treatment, it was demonstrated that response could be identified in the subgroup contributing this early time point. Additional data supported the sensitivity of [$^{18}$F]FDG-PET to define treatment response in gastrointestinal stromal tumors (GISTs) [97,98]. A review of the rapidly evolving role of [$^{18}$F]FDG-PET for GISTs is given by Van den Abbeele [99], who emphasizes that the experiences with GISTs are being seen to provide more personalized assessment of response of targeted therapy.

***Method Standardization*** Consistent acquisition and analysis of [$^{18}$F]FDG-PET data within clinical trials is paramount to successful interpretation of the clinical trial data. [$^{18}$F]FDG-PET analysis can simply be undertaken by visual inspection of the PET images. Whilst commonly done in clinical practice it is often limited for clinical trials. However, there are some instances where visual inspection is considered sufficient for clinical trials such as for response criteria for lymphoma [86]. Generally, a more quantitative assessment is considered necessary for applying [$^{18}$F]FDG-PET to clinical trials. The most comprehensive quantification of FDG uptake is achieved by acquiring a dynamic scan with simultaneous assessment of tracer concentration in arterial blood. Subsequent Patlak analysis (and other methods) can yield metabolic rate constants. The dynamic scanning, arterial line, and complex analysis are not suitable for all clinical trials. The most widely used approach for [$^{18}$F]FDG-PET analysis in clinical trials is a semiquantitative approach providing the standardized uptake value (SUV). This approach simply normalizes tumor uptake to injected dose and body weight, and the acquisition procedures are favorable to a busy PET center.

In 1999 the EORTC (European Organization for Research and Treatment of Cancer) PET study group provided recommendations on how PET should be performed for studying clinical response [100]. These recommendations emphasized the need for consistent measurement, analysis, and interpretation of response. In 2006, Shankar et al. [101] provided recommendations on how [$^{18}$F]FDG-PET should be applied to study therapeutic response in National Cancer Institute trials. Other groups from both industry and academia have

defined protocol standards for clinical trials [102,103]. Other publications provide important insight into the practical issues associated with [$^{18}$F]FDG-PET in clinical trials. For example, Weber [104] describes many pitfalls of seemingly straightforward SUV analyses, including errors associated with incorrect scanner calibration and varied period between injection and imaging. Furthermore, Westerterp et al. [105] describe the impact of image processing and ROI definition procedures on the SUV analysis. It demonstrates that despite the conceptual simplicity of SUV, every attention needs to be paid to all components in the imaging chain to ensure successful measurement in a multicenter setting. Also important may be tumor-type specific nuances of the PET approach as described by Lind et al. [106], where specific technical considerations are important to optimize sensitivity to detect disease during follow-up of breast cancer.

The published recommendations and further practical insights can be used to form an [$^{18}$F]FDG-PET protocol to define unambiguously how the patient is to be prepared, how the images are to be acquired, and how the images are to be processed and reported for a given clinical trial. Although there are still multiple standards that could be adopted for [$^{18}$F]FDG-PET, the attention focused on standardization encourages wide adoption of quantitative [$^{18}$F]FDG-PET. Many investigating sites now have experience of clinical trial imaging using [$^{18}$F]FDG-PET and consistently demonstrate a willingness to follow a standard protocol. Given the varied approaches by which analysis could be performed, centralized review of PET data for SUV analysis is preferable. This is achieved either using a contract research organization with expertise in quantitative imaging or coordinated via an academic network of centers where one center takes the lead role in analysis.

In addition to SUV analyses in the study of treatment-induced change, it is important to note the use of [$^{18}$F]FDG-PET with visual interpretation for screening bone lesions, where it may have sensitivity advantages over bone scintigraphy [107]. Such bone lesion screening remains important for assessments of disease burden in many clinical studies. In some instances it may replace the need for scintigraphy where whole-body [$^{18}$F]FDG-PET is undertaken as a matter of routine. However, given the high cost of PET relative to bone scintigraphy, it remains uncertain what the role of [$^{18}$F]FDG-PET will ultimately be in this setting.

***New Tracers*** Whereas [$^{18}$F]FDG-PET remains a developing tool in terms of its role in oncology drug development, PET has the potential to do much more [108]. Specifically, FDG has significant limitations in a number of settings. Kelloff et al. [109] review the extensive array of molecular imaging probes able to define molecular targets and understand processes to support oncology drug development. The tracers in development (including probes for SPECT) facilitate the study of proliferation, apoptosis, hypoxia, and angiogenesis and can study androgen and estrogen receptors. These molecular imaging platforms will provide an array of tools applicable for different stages of drug development, including:

1. *Biodistribution and in vivo concentration.* Studies of radiolabeled drugs to interrogate in vivo concentration and distribution [110], in particular the concept of microdosing where subtherapeutic doses of labeled drugs can be given, hold great promise [111].
2. *Receptor analyses.* Understanding receptors pertinent to cancer growth continues to be of interest, particularly receptors highly relevant to therapeutic intervention (e.g., estrogen; see Katzenellenbogen et al. [112], and for *HER2* expression, [113]). In a review by Mankoff et al. [114] the potential of tumor receptor imaging is reviewed with an emphasis on the complementary role that imaging can play vis-à-vis biopsy by studying receptor expression heterogeneity.
3. *Proliferation.* Labeled thymidine analogs offer a route to noninvasive assessment of tumor proliferation [115]. Of the potential agents, [$^{18}$F]FLT (3′-deoxy-3′-fluorothymidine) has been established as a suitable agent for wide use [116]. Reports have demonstrated the key advantages of [$^{18}$F]FLT-PET over [$^{18}$F]FDG-PET in a number of tumor types [117–119]. The role of [$^{18}$F]FLT-PET continues to grow, and increased availability of the tracer is likely to further establish the technique within clinical drug development.
4. *Apoptosis.* A noninvasive probe of apoptosis would undoubtedly provide general insight into early therapeutic success both for clinical assessments and for drug development decision making. Annexin-V has received considerable attention as an apoptosis probe labeled with either $^{18}$F [120] or $^{124}$I [121]. Different approaches to measure apoptosis are also in development, as described in the review of apoptosis tracer labeling strategies by Lahorte et al. [122]. It is a clear that a robust apoptosis imaging agent will play an important role in clinical drug development.
5. *Angiogenesis.* The specificity and sensitivity of currently available surrogates of angiogenic activity (e.g., dynamic contrast-enhanced MRI) are limited. PET tracers targeting receptors up-regulated with angiogenesis, such as the integrin $\alpha_v\beta_3$, make attractive approaches to a study of angiogenesis [123] with close linkage to the tumor biology. Such agents are currently in clinical development [124,125] for cancer imaging and can be used for quantitative assessment of integrin levels in tumors [126]. Given the preponderance of antiangiogenic targets in different stages of clinical development, PET tracers able to study change in this process remain highly attractive.

The diversity of labeling strategies means that many drugs and additional tracers will become available to complement the already developing array of PET tracers. Some of these agents are clearly broadly applicable (such as proliferation and apoptosis) and could follow the route that [$^{18}$F]FDG-PET has followed: to wide clinical and drug development use. The cost–benefit

relationship of such agents is clear. Tracer development for specific drug programs (e.g., labeling a drug for biodistribution studies) may be costly in terms of finance, time, and resource, but if the biodistribution analysis identifies an important safety signal, for example, the investment is likely to be rewarding. Molecular imaging development with required multidisciplinary teams and substantial infrastructure remains a costly endeavor. However, relative to potentially late-failing drug development programs the investment can be justified in many instances.

## Dynamic Contrast-Enhanced MRI

Dynamic contrast-enhanced MRI (DCE-MRI) is a functional imaging technique performed by rapidly acquiring multiple $T_1$-weighted MR images from a given tissue during injection of a paramagnetic contrast agent. The consequential tissue signal changes are quantified, and parameters with biological linkage are extracted. A review of the experiment with a focus on the analysis required is given by Parker and Padhani [127]. Further, standardized terminology and definitions of DCE-MRI-derived parameters have been defined by Tofts et al. [128]. Parameters typically extracted and used in the study of pharmacological response include:

- $K^{\text{trans}}$ ($\text{min}^{-1}$): the volume transfer constant between plasma and the extravascular extracellular space. The parameter is related to both flow and vascular permeability and, as such, change cannot be interpreted precisely. However, successful pharmacological intervention is expected to reduce this parameter by either flow and/or permeability change.
- $v_e$ (unitless): the volume fraction of the extravascular extracellular space. This is assumed to relate to interstitial space, although the relationship remains complex and posttreatment changes are not predictable.
- $k_{\text{ep}}$ ($\text{min}^{-1}$): the rate constant between the extravascular extracellular space and the plasma.
- IAUC: the initial area under the curve, typically over the first 60s. This is a semiquantitative expression of the initial contrast enhancement, which requires no data modeling. The relationship to physiology is complex [129], although both flow and permeability are likely to affect the parameter, and successful antiangiogenic treatment is expected to reduce the value.

In the recommendations by Leach et al. [130] both $K^{\text{trans}}$ and IAUC should be used as primary DCE-MRI endpoints, with $v_e$ and $k_{\text{ep}}$ as secondary endpoints.

As emphasized in the review by O'Connor et al. [131], DCE-MRI has been used widely for the study of angiogenesis inhibitors. Since this class of agents has received considerable attention across industry, interest in DCE-MRI has

become widespread. The review by Jayson and Waterton [132] outlines the progress in use of the technique for drug development, focusing on fundamental aspects such as design of the protocol, measurement validity, and reproducibility. Work focusing on DCE-MRI as a pharmacodynamic endpoint has been complemented by developing our understanding of the technique for more routine clinical use: for prognostication, for example.

The use of DCE-MRI offers significant advantages over other methods able to interrogate aspects of tissue vascularity. With careful implementation, multisite standardization can be achieved and sufficient reproducibility relative to typical pharmacodynamic effects can be realized. However, it could be argued that angiogenesis inhibitors either currently in development or marketed may have proceeded without DCE-MRI within the development program. So what does DCE-MRI add to the development paradigm, and why, if such a localized indicator of antiangiogenic activity exists does it not form a fundamental step in the development of such agents? First, there are technical challenges to successful implementation that form impediments to broader use. Second, there is still insufficient understanding of the relationship between given DCE-MRI parameters and both histopathology and clinical outcome. These uncertainties lead to difficulties in interpreting DCE-MRI studies. For example, how much of a change in a DCE-MRI parameter is enough to predict benefit?

### *Method Standardization and Deployment Challenges*

*Site Limitations* DCE-MRI remains a specialized technique with limited routine application, and therefore many radiology departments will have limited experience with its use. It is important to consider the practicalities of consistent DCE-MRI implementation across multiple sites required for a given clinical trial. The first step requiring consideration is the selection of sites capable of performing DCE-MRI by adhering to a common acquisition for the study. One factor to consider is what experience site personnel have in performing DCE-MRI. However, this is not to say that sites with few experienced personnel are not capable of undertaking DCE-MRI, but the users must be committed to adhering to a prescribed protocol.

Statistically, it may be better to have fewer subjects scanned at a small number of sites where the imaging can be tightly controlled versus many sites with greater diversity of equipment and experience. A pharmaceutical company considering multiple studies could consider developing a network of specialized imaging sites able to perform consistent imaging and where good communication between sites can help optimize imaging.

*Implementing DCE-MRI Within the Clinical Protocol* A great variety of options exist regarding incorporating DCE-MRI within the clinical protocol. Some of the most important factors to consider are:

- Is a repeat baseline examination necessary to establish methodological reproducibility?
- When should posttherapeutic administration imaging be performed?
- What are the primary DCE-MRI endpoints?

These questions should be answered by considering what is known about the expected pharmacology of the drug and what has been done previously with similar agents. Important practical parameters include how many examinations will be acceptable to the subject (with a careful review of other required procedures) and whether close time points present scheduling challenges to the radiology department.

It has been recommended that reproducibility examinations be built into the study where possible in the form of a repeat baseline [130]. This would typically be done by performing two examinations within a week prior to beginning of therapy. An example of analysis of a repeat baseline study includes that by Galbraith et al. [133], where repeat examinations performed one week apart provided confidence intervals for $K^{trans}$ and $IAUC_{90}$ (64% and 61%, respectively). Other examples include that by Roberts et al. [134], where investigators assessed repeat examinations from data across two centers in multiple tumor types, comparing compartmental modeling versus model-free analysis. The 95% confidence intervals for $K^{trans}$ and $IAUC_{60}$ were calculated to be 35% and 55%, respectively. Clear differences in the reproducibility are apparent from the published examples and are expected to arise from differences in acquisition, modeling techniques, and assumptions in the model. This gives credence to the idea that a reproducibility assessment is often important, and that if performed it must be representative of subsequent study conditions. Although changes following administration of drug are often large relative to the methodological variability, in many instances it may be important to minimize the variability as much as possible. In a publication by Parker et al. [135] it is demonstrated that significant improvement in reproducibility of $K^{trans}$ and other parameters can be attained by using a population-based arterial input function. Such improvements in reproducibility are likely to be important. In a study where dose is varied, using DCE-MRI to understanding the lower limit of biological effectiveness could be important, and the measurement reproducibility will have a direct impact on the ability to define subtle change at low doses.

Published DCE-MRI studies show a range of selected posttherapy measurement time points [131]. Between different drugs there will be significant dosing differences, pharmacokinetic properties, and differences in mechanism likely to lead to varied time scales of DCE-MRI response. Therefore, there is no consistent formula to select posttreatment time points. However, where studies do include an early time point (such as one or two days after therapy has begun), a large response is generally seen and is similar in magnitude to subsequent time points. The data by Morgan et al. [136], Mross et al. [137], and Thomas et al. [138] emphasize the early response (day 2) that can be

measured consistently between studies, which then persists (out to day 28). It would thus seem that in most instances an early time point provides significant data to evaluate vascular response.

*Choosing a Standardized Protocol* A number of acquisition and analysis options exist to perform DCE-MRI. For any one study, consistent application for patients being investigated is an obvious requirement to keep variability minimized. However, for a given company there are clear benefits in maintaining a consistent acquisition and analysis protocol across multiple studies: reduced operational burden to develop acquisition guidelines and the ability to compare one study to the next potentially to compare the magnitude of response between different drugs, doses, or dosing schedules. There are also benefits to developing consistent practices for both academic- and industry-sponsored DCE-MRI-based studies:

- Increased awareness of a standard protocol to minimize the complexity encountered by radiology departments in dealing with many different DCE-MRI protocols for different studies.
- Potential to perform metaanalyses on data from multiple studies to query relationships between DCE-MRI response and clinical outcome.

Leach et al., on behalf of Cancer Research UK, proposed recommendations for performing DCE-MRI for studying antiangiogenic and antivascular therapies [130]. Their technique provides an academic and industry perspective on the key attributes of a standardized DCE-MRI acquisition. Owing to the differences in the way the technique is implemented by different scanners, different acquisition protocols will often be required from one center to the next. However, provided that cross-site implementation is undertaken carefully, a DCE-MRI measurement at one center should be comparable with that taken at another.

MRI technology changes at a significant pace, evidenced by multichannel technology and availability of scanners operating at static magnetic field strengths of 3T and above. Standardizing measurements from one study to the next is a challenge with a background of technological evolution. When a new technology (e.g., 3T MRI) is introduced, a careful comparison with existing technology is required to ensure equivalence of measurement.

*Responsibilities for Deployment and Centralized Analysis* Ideally, multiple sites with quantitative imaging expertise will be selectable to ensure that significant recruitment goals can be achieved in a reasonable period of time. Often, this may not be the case, and sites with limited imaging experience are required. For this reason, deployment of the DCE-MRI acquisition is complex, costly, and time consuming, requiring the following activities: review of the adequacy of scanner specifications, implementation of the MRI sequences required, scanner performance evaluation, training of site technologists,

instructions to upload the data in a timely manner on scan completion, robust quality control steps, and centralized analysis. Often, a third-party company, typically a contract research organization with imaging expertise, can be given responsibility for these activities. Alternatively, one clinical site in a network can be responsible for defining the acquisition protocol and communicating imaging details with partnering sites. The latter is particularly applicable where all sites have expertise in quantitative imaging. Whatever the approach selected, it is important that data acquired be transported readily (ideally via electronic transfer) to a central analysis lab in order that any important deviations from the acquisition protocol can be identified to ensure future consistency.

Owing to the diversity of software platforms able to analyze DCE-MRI and the options available for analysis, it is imperative that data sets be reviewed centrally using a consistent process. This will ensure that variability in the analysis is kept to an absolute minimum. Typically, each analyzable lesion has an associated DCE-MRI parameter created ($K^{trans}$, IAUC, and others) following pixel-by-pixel analysis. A mean value associated with each lesion is typically assumed to provide the simplest output with more advanced analysis (such as heterogeneity analyses) potentially able to provide further insight into the data.

*Toxicity of Gadolinium-Based Contrast Agents* Data show that administration of gadolinium-based contrast media is associated with a risk of developing nephrogenic systemic fibrosis (NSF) [139–141]. Although understanding is still limited, it appears that risk factors for NSF include reduced kidney function; multiple exposures to gadolinium-based chelates; the use of linear chelates, indicating a potential relationship to thermodynamic stability [142]; and the presence of major tissue injury, termed *pro-inflammatory conditions* [143].

It is clear that much is still to be understood about NSF and risk factors involved. However, it is prudent to minimize the risk of NSF for all subjects being exposed to gadolinium as part of a DCE-MRI examination or other MRI study based on existing data. DCE-MRI studies used for drug development differ from clinical practice and should approach the issue with particular caution, owing to the following factors:

- The need for multiple injections of gadolinium-based contrast agents within a short period of time, particularly if a repeat baseline and early posttreatment follow-up is required.
- For clinical diagnostic purposes the justification of gadolinium in high-risk subjects can be clearly defined, but for a patient in a clinical drug study the benefits of receiving multiple examinations are less clear.

The safety of the patient is paramount at the same time as acquisition of data to assist the development of novel therapies remains important. A conservative approach is recommended to minimize risk. This could include:

- Minimizing the number of DCE-MRI examinations while still being able to provide an understanding of the pharmacology
- Use of a single-dose gadolinium agent
- Careful monitoring of renal function and setting boundaries below which subjects should not be studied
- Use of contrast agents that, based on current data, appear to have the lowest association with NSF
- Continual review of emerging data with necessary regular revision of guidelines

***DCE-MRI for Drug Development Decision Making*** As reviewed by O'Connor et al. [131] DCE-MRI has been used extensively to study multiple antiangiogenic and antivascular therapeutics. Many of the DCE-MRI experiments performed to date have been undertaken on drugs still in development. DCE-MRI certainly appears to be a methodology applied consistently within antiangiogenic development programs. Examples include the evaluation of AG-013736 by Liu et al. [144], where a relationship was observed between drug exposure and change in both $K^{trans}$ and IAUC from baseline to day 2. DCE-MRI was also incorporated into phase I investigations of the kinase inhibitor (VEGFR, PDGFR, FGFR) BIBF 1120 [145,146]. Both studies reported DCE-MRI findings after baseline, early (2 days [145] and 3 days [146]), and later time points (28 days [145] and 30 days [146]). Although DCE-MRI was obtained at multiple drug doses, it is not clear whether there was any dose–response relationship to support a dose decision for future studies.

A methodology able to define dose versus vascular response for antiangiogenic therapies should in theory play a significant role in many drug development programs. There do remain, however, a number of significant challenges. First, most published studies where DCE-MRI has been applied to novel therapies have been undertaken on subjects with a range of tumor types, typical of the phase I setting. This creates methodological challenges likely to add to measurement variability (and measurement failure in a percentage of patients), and biological response will be less uniform than a study of a given tumor type. Another challenge is interpretation of data. How much of a change in $K^{trans}$ is sufficient to predict subsequent treatment success? Without data linking DCE-MRI parameter change with metrics of clinical outcome in large studies, with different tumor types, and different therapies, it is not possible to establish a broad threshold of, say, the $K^{trans}$ change required to predict therapeutic success. If this could be realized, a firm understanding of radiological effectiveness from DCE-MRI could be used to inform dose decisions for a drug development program. Until such data are available, DCE-MRI is likely to be used as a supportive function contributing at least to dose decisions, evaluating early signals of efficacy, and understanding scheduling. The value that DCE-MRI brings, however, would increase significantly, and

implementation cost would clearly be justified if there were greater understanding of different levels of DCE-MRI response.

Collins [147] has critiqued how imaging (both PET and MRI) contributed to decision making in the development of combretastatin A4 phosphate. Such analyses are important in order to understand how imaging has actually contributed to decision making, the many instances when it has not, and what can be done to enhance the role of imaging in clinical drug development.

***Future Improvements in the Use of DCE-MRI*** As described previously, published recommendations will help to encourage consistent DCE-MRI measurements. These should enable pharmaceutical companies and academic groups to focus on the challenges of drug development over methodological optimization. However, the current recommendations that exist remain appropriately broad since they require applicability to scanners from different manufacturers. Greater harmonization of methods between scanner manufacturers would considerably benefit the DCE-MRI experiment, lowering the barrier to implementation and enabling broader use in drug development.

Increased standards of measurement would deliver more studies that could be subjected to metaanalysis: for example, to study DCE-MRI response versus outcome. With such relationships understood, the role of DCE-MRI in early drug development can potentially move from an interesting secondary objective to delivering primary study goals.

Clinical trials are also likely to benefit increasingly from the multiparameter assessments possible within a single MR scan session. In a study of the pan-VEGF receptor tyrosine kinase inhibitor AZD2171 in glioblastoma patients, Batchelor et al. [148] demonstrated that an extensive MRI evaluation provides significant benefit over DCE-MRI alone. The study incorporated lesion volume and apparent diffusion coefficient measurements, together with DCE-MRI-derived indices, including $K^{trans}$ and relative vessel size estimations. Such approaches are likely to be necessary to fully interpret the complex changes occurring following antiangiogenic therapy and to promote understanding of drug scheduling and how optimally to combine different therapies. Improved standardization will both enhance the ability to incorporate DCE-MRI into drug development and promote understanding of how different DCE-MRI response levels are interpreted.

## Apparent Diffusion Coefficient MRI

A number of other image-based functional measures are being investigated as potential imaging biomarkers in ongoing exploratory research. Here we briefly highlight one additional cancer application. Diffusion MRI is an image-based method used to measure the molecular diffusion of water molecules in biological tissues, and derives from the well-established NMR method of

molecular diffusion measurement [149]. The parameter measured, the apparent diffusion coefficient (ADC), reflects the diffusivity or mobility of water molecules and is termed *apparent* because of the undetermined effects of perfusion, multiple tissue water compartments, restricted diffusion due to cellular and extracellular structures, and tissue anisotropy [149].

Measurement of tissue water diffusion has proven to be remarkably versatile for characterizing tissue structure, as well as tissue changes due to pathology or the effects of therapy. Diffusion MRI has been well studied in the clinical evaluation of ischemic stroke [150]. In oncology, the application of diffusion MRI has focused on the image-based measurement of tumor ADC to detect early changes associated with treatment response [151]. ADC-MRI has been applied with some preliminary success in preclinical and clinical research as an imaging biomarker of treatment response in cancers of the brain [152,153], head and neck [154], cervix [155], breast [156], and prostate [157]. A related MRI technique, diffusion tensor imaging (DTI), has been used for noninvasive characterization of tissue anisotropy, structural integrity, and connectivity in myocardium [158] and white matter tracts in the brain [159]. In particular, the latter application has been useful in preoperative planning for brain tumor patients [160].

The concept of using ADC-MRI to evaluate the effects of therapy in solid tumors is based on the general observation of an increase in tumor ADC after cytotoxic treatments in several tumor types [161–163]. Lyng et al. [163] demonstrated in four different melanoma xenograft models that tumor ADC is inversely proportional to viable tissue cell density. It is thought that an increase in tumor ADC upon treatment reflects the development of necrosis through increased cell membrane permeability and extracellular fraction, and decreased cellularity which ultimately results in greater water mobility [164]. In some tumor types such as intracranial gliomas, an early transient decrease in ADC has also been observed prior to subsequent ADC elevation, which may be due to treatment-induced cellular swelling [152,165]. These changes generally precede gross tumor regression, so that changes in tumor ADC appear to predict such regression.

ADC-MRI offers some key advantages as a potential imaging biomarker of tumor treatment response. ADC-MRI does not require the injection of exogenous contrast agents, since it relies on the endogenous contrast mechanism of water diffusion. In addition, as an MRI-based method it does not involve exposure to ionizing radiation. Moreover, the ADC is an absolute biophysical quantity which, in principle, should be comparable across MRI hardware platforms. However, varying acquisition methodologies currently lead to some variability across hardware platforms and between different imaging centers, which could be addressed through better standardization of acquisition and processing protocols.

A primary limitation currently with ADC-MRI is the enhanced sensitivity of diffusion MRI to patient motion, such that involuntary or unrelated physiological motion can produce quantitative artifacts. A variety of acquisition

and image-processing techniques have been devised to correct or minimize such image artifacts (e.g., [165–167]), which must be addressed and optimized in the process of standardizing the imaging protocol. An additional concern is the fact that tumor ADC response to therapy can also reflect formation of edema secondary to treatment [151]. Recently developed voxel-based statistical classification approaches have improved the robustness of ADC-MRI against such potentially confounding factors [152,153]. Although ADC-MRI appears to be a promising imaging biomarker for oncology treatment monitoring, further clinical studies are needed to fully characterize the validity and reliability of ADC-MRI for different tumor types and therapeutic classes, as well as to compare the relative value of ADC-MRI with similar measures of tumor viability, such as FDG-PET.

## Selecting the Right Technique

Very often, similar measurements of biological and disease processes are provided by multiple imaging modalities or even distinct approaches within a single modality. In selecting the right technique for a given problem, it is important to consider the range of possible biomarkers (both imaging and nonimaging), as well as the varying technical requirements and available resources, any significant differences between hardware platforms, and the limitations of specific modalities. Here we illustrate the considerations for selecting a particular imaging biomarker for the problem of assessing tumor vascularity.

***Example: Studying Aspects of Tissue Vascularity*** Let us assume that we wish to evaluate the tumor response to a new antiangiogenic agent for the purposes of defining the lower limit of biological effectiveness across a dose range. What parameters need consideration when deciding on the appropriate methodology?

The advantages and disadvantages of techniques are seldom reviewed comprehensively, and more specifically, there are few studies actually comparing methodologies for their potential to study therapy. De Langen et al. provide an insightful comparison of $[^{15}O]H_2O$-PET versus DCE-MRI for tumor blood flow measurements [168]. The review focuses on both the practical aspects of the two techniques and the ability of the different techniques to probe flow. The conclusion from the comparison is that both approaches are viable methods for the study of flow, yet important differences need to be considered for each study. Goh and Padhani compare the merits of DCE-CT versus DCE-MRI in the study of tumor angiogenesis [169]. Again, it is concluded that both techniques should be considered, but a number of factors should determine selection, including drug mechanism of action, tumor location, patient characteristics, and available infrastructure and expertise [169].

The potential techniques able to define tumor vascular characteristics include the following:

*Dynamic Contrast-Enhanced CT*

- *Endpoints.* DCE-CT can measure blood flow, blood volume, permeability, and mean transit time.
- *Tumor localization.* Multidetector CT technology enables a significant coverage of the tumor during the dynamic assessment, overcoming previous limitations of the approach. Furthermore, CT offers some key advantages when imaging certain anatomical regions, as summarized by Goh and Padhani [169].
- *Clinical trial practicalities.* The technique is widely available, although it is preferable that sites have previous expertise in dynamic CT. However, training imaging centers with a standard protocol is achievable. Also, since CT is generally required for RECIST assessment of disease burden, the use of CT for a pharmacodynamic investigation may result in fewer scanning sessions.
- *Risks.* Dynamic CT may add a significant radiation dose burden that may become problematic in many patient populations. This is particularly true if these subjects are already receiving regular CT scans for standard radiological assessment, perhaps also with nuclear medicine investigations accumulating a significant dose. This is likely to limit the number of time points that are measurable within a short clinical trial, particularly compared to DCE-MRI and microbubble ultrasound.
- *Quantification.* The straightforward relationship between CT contrast agent concentration and tissue enhancement is often cited as a significant advantage that DCE-CT has over DCE-MRI, making absolute quantification of perfusion achievable. For analysis, the availability of commercial software on the scanning platform holds a significant advantage relative to other methods, although centralized analysis is often required anyway, so this advantage may be limited.

In summary, DCE-CT is a widely available, robust methodology for providing perfusion quantification [170]. However, the associated radiation dose is likely to limit the use of CT in certain patient populations and certainly keep the number of time points low.

*Dynamic Contrast-Enhanced MRI*

- *Endpoints.* DCE-MRI typically provides indices relating to permeability and flow ($K^{\text{trans}}$ and IAUC) in addition to indices relating to the tissue microenvironment ($v_e$).
- *Tumor localization.* The localization can be similar to CT in that a single level is typically scanned using multiple slices. However, coronal scanning, for example, may facilitate greater coverage if multiple lesions

require measurement. Compared to DCE-CT, DCE-MRI is not optimal for all lesions and may often suffer from artifacts rendering measurement impossible (e.g., in the mediastinum).

- *Clinical trial practicalities.* The equipment required to perform DCE-MRI is widely available, yet the expertise in quantitative MRI measurement is limited. This can often be overcome by working closely with the imaging center and providing rapid feedback on data generated. In many instances, willingness of a site to adhere to a consistent imaging protocol is preferable to a site with extensive experience, insisting that the measurement be performed a particular way. The number of time points available from DCE-MRI is one advantage, with many publications demonstrating five measurements or more are feasible. However, recent concerns regarding NSF are likely to limit the number of time points, especially in patient populations with compromised renal function.
- *Risks.* The only significant risk to the subject is from the exposure to gadolinium-containing contrast media. These risks can be minimized, but in patient populations with compromised renal function an alternative imaging technique may be deemed appropriate.
- *Quantification.* With the lack of consistent and widely available image analysis platforms for DCE-MRI, central analysis of multicenter data is paramount. A significant limitation is the inability to define flow separately from permeability. Although $K^{trans}$ and IAUC may provide useful general indicators of vascular pharmacology, interpretation can be challenging.

Without the radiation dose burden and the ability to perform multiple post-therapy time points, DCE-MRI remains attractive for many studies. However, the challenge to implement across multiple centers with the associated cost and time should not be overlooked. Furthermore, ambiguity of the imaging parameters needs to be considered carefully to ensure that DCE-MRI will contribute sufficient insight into the pharmacology.

### *Microbubble Ultrasound*

- *Endpoints.* The reported endpoints are varied and include parameters such as area under the enhancement curve, arrival time, time to peak enhancement, and counting of identified vessels.
- *Tumor localization.* Bone and air attenuation sets significant limits on accessibility of this technique to some lesions. This is a fundamental limitation that should be considered when establishing the appropriate patient population to study.
- *Clinical trial practicalities.* Ultrasound remains widely available, although centers experienced in quantitative microbubble techniques remain limited. Furthermore, differences in techniques at those centers need

harmonizing before a clinical trial can begin. Several repeat measurements are possible [171]. This could allow studies in multiple lesions or repeat measurement in the event of failure or repeated measurements averaged to minimize variability. The technique is likely to be favorable to the patient, requiring relatively quick examination times, and comfort during the examination is likely to be high. The low cost may also be an important factor, and if implemented appropriately would constitute only a small fraction of a total study budget. A frequently quoted criticism of quantitative ultrasound is the operator dependence of the measurement. However, it has been shown that with careful implementation, such dependence can be controlled [172]. How this would translate into large multicenter studies, however, is yet to be established.

- *Risks.* No significant risks are expected with repeated microbubble ultrasound investigations. Microbubble contrast media are considered generally safe, with serious adverse events rarely observed [173].
- *Quantification.* Analysis is currently limited to empirical analyses of the microbubble dynamics. As such, there is likely to be significant variability of measurement from one scanner to the next where different performance characteristics and software implementations will affect measurement. A cross-site calibration exercise combined with central analysis could overcome many of the limitations. Understanding the relationship of the ultrasound parameters to the biology may be challenging.

Microbubble ultrasound remains a quick, low-cost, highly accessible methodology for gauging vascular change [174]. Further understanding is needed of the multisite performance together with more insight into the biological linkage. For single or small multisite studies it could provide a useful method to enable many repeat measurements over a short period of time to understand relationships between pharmacokinetics and pharmacodynamic response.

### *[$^{15}$O]$H_2O$-PET*

- *Endpoints.* Flow and distribution volume are typically derived. Since the tracer is rapidly diffusible, permeability is not extracted.
- *Tumor localization.* The poor spatial resolution of PET will limit the study of small lesions. Even if such lesions are identifiable, partial volume effects are likely to make robust quantification challenging [175]. Careful screening of subjects is required to ensure that there exist, for example, at least 2-cm lesions prior to entering into the study.
- *Clinical trial practicalities.* Owing to the short half-life of $^{15}$O (2 minutes), an onsite cyclotron is required to generate the tracer and facilitate rapid administration. Even sites with a cyclotron may not produce [$^{15}$O]$H_2O$

routinely. If this technique is required at multiple sites, the infrastructure and expertise of potential sites must be assessed carefully. One benefit of the short half-life is that repeat examinations can be performed within 10 minutes of each other. This allows either repeat perfusion assessments early after therapy or use of correlative imaging such as [$^{18}$F]FDG-PET without interference. Since an arterial input function needs to be defined for quantification, rapid arterial blood sampling is required during the study. This adds significant complexity onto an imaging examination, and the practicalities need consideration when selecting this technique.

- *Risks.* Although the effective radiation dose is lower than an extensive CT scan or an [$^{18}$F]FDG-PET scan, the additional radiation dose this brings to a subject who has been scanned extensively should be considered carefully. This is particularly true where there may be no direct benefit to the subject of such a pharmacodynamic endpoint versus a diagnostic CT or PET scan.
- *Quantification.* A significant body of work exists on quantification of dynamic PET data, particularly from the brain and heart. Adaptation of such analyses can facilitate robust measurement of tumor pathophysiology [176]. Owing to different implementation options, central analysis is required if data are obtained from multiple centers.

PET provides a robust technique to quantify absolute tumor perfusion, and the quantification has many advantages over other methodologies. However, the limited availability across sites, the relatively complex patient setup, and limited performance in small lesions need careful consideration.

As outlined above, there exist many practical and basic differences between measurement techniques able to study different aspects of the same phenomenon. All decisions will be dependent on the study question being asked. For example, if analysis of lung tumor perfusion across multiple centers is required, DCE-CT is likely to be an appropriate technology. If multiple time points are required to assess relative vascular change of liver lesions, microbubble ultrasound may be an appropriate methodology.

The following parameters will define the appropriateness of the various techniques: the ability to perform centralized analysis; the patient characteristics (radiation dose considerations and sensitivity to both MR and CT contrast media), the number of subjects and study centers required for delivery of the primary study endpoint, the number of time points required to fully assess the pharmacodynamics, the drug mechanism, and the probable importance of perfusion analysis versus a measure also sensitive to permeability.

There is no single technique ideally suited to address all questions in all types of clinical trials. DCE-MRI has tended to receive considerable attention for studies of anti-angiogenics. This is probably because although it has

limitations in terms of the quantification, it comes with high-resolution robust measurement in a variety of tumor types, and multiple repeat measurements are possible in most patient populations. Despite these advantages, the limitations of DCE-MRI should be evaluated and all measurement options considered.

## THE ONGOING CHALLENGE OF IMAGING STANDARDIZATION

Without standardization of imaging methods, the variety of imaging practices, tracers, equipment, and interpretation would lead to variability far greater than any drug-induced change. Since the extent of the drug-induced change is generally unknown before the experiment, study setup should aim to minimize measurement variability as far as practically possible.

Standardization of imaging is required at a number of levels:

1. *Clinical protocol.* The clinical protocol defines how the imaging endpoint is incorporated into the clinical protocol. For example, when are imaging time points relative to drug dosing? What size lesions are expected to be measurable by the technique chosen? Consistency of how imaging is incorporated is certainly required within any drug development program to ensure that cross-comparison of studies can be made.
2. *Technique selection.* What technique is chosen to deliver a given endpoint? Deciding what techniques would generally be used for a given drug development program or throughout a development organization would aid consistency of decision making and, through familiarization, lower implementation costs and resource requirements.
3. *Patient preparation.* Standardization of patient preparation for an imaging examination includes management within the scanner: fasting requirements, patient position, and cannulation. Patient preparation is particularly important for metabolic imaging (e.g., [$^{18}$F]FDG-PET) and functional imaging (e.g., DCE-CT), where multiple factors can significantly impact the imaging result (e.g., via modulation of blood glucose level or liver blood flow).
4. *Scanner setup.* How the imaging technique is implemented on a given scanner and how this is done between different scanners at different institutions defines the most significant complexity of multisite imaging. The variety of equipment is growing as manufacturers offer more hardware and software options. This diversification increases the demands to set up imaging between two different scanners with equivalence.
5. *Image analysis and reporting.* Generally, the software platforms used by centers (on the scanner, third-party, or site proprietary) are so diverse that data must be analyzed centrally. When this is done, standardization within a study and between studies can readily be controlled.

### Unique Challenges to Standardized Imaging

The diversity of imaging hardware creates one of the most significant sources of measurement variability. Seemingly equivalent imaging methods deployed on two different scanners may provide equivalent diagnostic information and images of similar appearance. However, for quantitative imaging where imaging-derived biomarkers are required, very significant differences in derived values can occur. Also, there may exist proprietary algorithms for image reconstruction or subsequent analysis, the details of which are not known and therefore cannot be compensated for.

When implementing complex imaging across multiple centers, expert resources are required to (1) compile a site-specific imaging protocol (e.g., detail acquisition requirements for a given manufacturer), (2) provide training to the relevant site staff likely to undertake the imaging (technologist and radiologist), (3) send standard phantoms to sites and review data to ensure expected performance and equivalence between sites, (4) ensure that data are received rapidly, preferably via electronic image transfer, so that the data can be reviewed rapidly and feedback given on issues, and (5) analyze the data and provide the imaging endpoints to the sponsor for statistical analysis.

### Achieving Protocol Consensus by Relevant Stakeholders

An effective method to provide consistent imaging biomarker application is to develop an agreed protocol between experts internationally. Examples of this include DCE-MRI and [$^{18}$F]FDG-PET [101,130]. Development of such recommendations is challenging to achieve given the complexity and likely disagreement on certain components during rapid evolution of imaging technology. Such efforts tend to result in general principles that are sufficiently flexible to work across many centers and ensure that similar endpoints are reported out from different studies.

### How Much Standardization Is Enough?

The degree of standardization needs to be considered for each technique requiring deployment. Standardization also comes at a cost in terms of resource and time in particular. There should be realistic expectations that imaging cannot, for example, be comprehensively standardized at 50 centers to the extent that it can be controlled at three centers. Some imaging techniques (e.g., [$^{18}$F]FDG-PET) can be standardized by a well-considered imaging protocol likely to be adhered to by the site imaging experts. However, complex MRI methodologies may require phantoms to be shipped to all participating sites and phantom data submitted for central review prior to any patients being scanned. If budget is limited, it may be preferable to perform consistent imaging on fewer subjects at a small number of centers rather than acquiring data from more subjects without investing in standardization.

It is important to remember that standardization is never likely to be achieved completely. With the rapid developments in imaging hardware, radiology best practice and contrast media/tracers standardization need to adapt. For example, with the increasing availability of 3T MRI scanners relative to the standard 1.5T scanner, when should 3T be recommended? A reasonable approach is that where data support the use of new hardware for a given application, equivalence should be defined prior to wider adoption.

## Mechanisms for Implementing Standardized Imaging

A number of bodies that represent imaging experts or users of imaging endpoints are optimally placed to develop consensus on imaging techniques and to disseminate recommendations. This includes government bodies such as the National Cancer Institute [101,177] and Cancer Research UK [130], accreditation bodies such as the American College of Radiology (ACR; http://www.acr.org) and the Intersocietal Accreditation Commission (IAC; http://www.intersocietal.org/intersocietal.htm), as well as the many professional societies related to medical imaging. ACR and IAC have traditionally provided voluntary accreditation for diagnostic imaging providers, although a recent U.S. health insurance quality improvement initiative and changes in Medicare law have mandated accreditation as a condition for insurance reimbursement. Standards have been established by IAC and ACR that specify the qualifications of imaging personnel, hardware quality control, conduct of imaging studies, image interpretation, and reporting. The scope of these standards is diagnostic imaging in clinical practice, and they are not directly applicable to imaging biomarkers. However, with the development of appropriate standards for quantitative imaging biomarkers, ACR and IAC may be ideally suited to facilitate the accreditation and education of imaging centers for imaging biomarkers.

In response to the need for quantitative imaging standards, the Radiological Society of North America (RSNA) has recently initiated the Quantitative Imaging Biomarkers Alliance (QIBA; http://www.rsna.org/Research/qiba_intro.cfm), represented by pharmaceutical companies, imaging hardware and software vendors, government agencies, imaging societies, RSNA leadership, and clinical trialists [178]. Modeled on the collaboration known as Integrating the Healthcare Enterprise (IHE; http://www.ihe.net/), QIBA aims "to advance quantitative imaging by focusing on standardizing the use of imaging biomarkers in clinical trials" with the long-term goal of "enhancing the use of quantitative imaging methods in clinical practice." Initially, planned efforts will address standardizing quantitative anatomical CT, FDG-PET, and DCE-MRI. Although QIBA has been formed under the auspices of RSNA, it seems probable that a dedicated standard setting organization may eventually be needed for ongoing development, maintenance, and dissemination of new imaging biomarker standards.

## MECHANISMS TO FACILITATE IMAGING BIOMARKER DEVELOPMENT

Despite the potential of imaging biomarkers as unique tools in drug development, the multidisciplinary nature and complexity of these technologies have somewhat hindered biomarker development. In addition, implementing and using imaging biomarkers has tended to be isolated to research institutions and centers with the necessary range of resources. The difficulties of imaging standardization and the expense of imaging biomarker qualification has also limited the incentive for commercialization of these technologies.

Many stakeholders in government and industry have recognized the inadequacy of traditional research and development approaches as applied to biomarker development. Other biomarkers in use today, such as blood pressure or plasma cholesterol, have taken decades to achieve their current utility [40]. In its Critical Path Initiative the FDA acknowledges the urgent need for new approaches and broad collaboration to address the unique challenges of biomarker development [179]. Two new approaches that are beginning to have a positive impact are public–private partnerships and the open-source development paradigm.

### Public–Private Partnerships

The Foundation for the National Institutes of Health (FNIH), in collaboration with the National Institutes of Health (NIH), the FDA, and the Pharmaceutical Research and Manufacturers of America (PhRMA), established in 2006 the Biomarkers Consortium [180], a public–private partnership focused on the discovery, development, and qualification of biomarkers, including imaging biomarkers, through collaborative biomedical research [181]. The Biomarkers Consortium is unprecedented in its size and coordination of government, academic, and industry aims for biomedical research focused on new biomarker technologies.

Government partners, including the NIH, FDA, and Centers for Medicare and Medicaid Services, contribute scientific and technical expertise as well as financial resources and project management to the consortium. Membership in the consortium is open to nonprofit and for-profit organizations and patient advocacy groups with an interest in helping the consortium reach its goals. Consortium members may nominate subject-area specialists as representatives on steering committees devoted to various therapeutic areas that are charged with evaluating biomarker project proposals. According to its Web site, "the Biomarkers Consortium wants all members to be actively engaged in this collaborative multi-sector approach to speed the discovery and validation of disease biomarkers and surrogates. Given the expense and logistical challenges of moving biomarker science forward, we've adopted this collaborative multi-sector approach that encompasses all relevant stakeholders."

Following biomarker project approval by the appropriate steering committee, private-sector fundraising for biomarker projects is facilitated by the FNIH.

Several large-scale imaging biomarker development and qualification studies have been initiated with assistance from the Biomarker Consortium. Two studies, to be conducted by the National Cancer Institute, are evaluating the utility of FDG-PET to monitor cytotoxic treatment of non-Hodgkin lymphoma and non-small cell lung carcinoma, providing data to support clinical qualification of FDG-PET as a possible surrogate endpoint [182]. A multicenter MRI study of carotid atherosclerotic plaque is determining the test–retest and intraobserver variability of MRI-based measures of plaque size and composition to facilitate imaging assay validation [183]. The Osteoarthritis Initiative (OAI) is a four-year, multicenter, longitudinal study of the clinical onset and progression of knee osteoarthritis which seeks to "characterize imaging, biochemical and genetic biomarkers that predict and track the course of [osteoarthritic] disease" [184]. Clinical data and images from almost 5000 patients will be made publicly available for further biomarker research worldwide [185]. The Alzheimer's Disease Neuroimaging Initiative (ADNI) is the largest project to date, with $25 million of its $60 million funding being raised through the FNIH. ADNI is a five-year study tracking 823 normal, mildly cognitively impaired, and Alzheimer disease patients using MRI and PET as well as laboratory and cognitive tests [186]. As with the OAI study, ADNI clinical data and images have been made available to the general scientific community for continued biomarker research and development [187].

These and other imaging-related projects supported by the Biomarker Consortium are addressing the need for clinical trials to establish the validity and clinical qualification of imaging biomarkers. Such trials would be too costly and would make little sense for any single organization to pursue but are made possible through collaborative sharing of the costs and associated risks among consortium members.

## Open-Source Collaboration

The success of the open-source paradigm for software development [188] has led to the open-source approach being applied to the development of biomedical informatics tools [e.g., Open Bioinformatics Foundation (http://www.open-bio.org) and Bioinformatics Organization (http://www.bioinformatics.org/)] as well as a gradual adoption of "open-source" collaborative models of biomedical research [189]. Key features of the open-source model are: (1) participation is open to anyone with the interest and skills to contribute; (2) development and communication is centralized through an Internet-based organization; and (3) the results are freely shared. Among drug development public–private partnerships, a hybrid approach is emerging in which the knowledge-based aspects of the research follow an open-source model, while "rule-based" aspects (such as toxicology, chemistry manufacturing and controls, clinical trials) are outsourced [190,191].

Large publicly funded research initiatives are similarly adopting an open-source approach. The cancer Biomedical Informatics Grid (caBIG; https://cabig.nci.nih.gov/) is an ambitious collaboration launched by the National Cancer Institute in 2003 to develop and share computing tools, infrastructure, and data across the spectrum of cancer research [192]. Data sharing follows a federated model [193] and integrates diverse cancer data from basic and clinical sources, as well as cancer imaging and tissue pathology data. A broad set of tools have been developed based on widely available open-source components and standards, including bioinformatics software, image analysis and database tools [194], and a complete clinical trials management system. Sharing of tools and data is facilitated by an open-source grid-based computing infrastructure. Although the focus of the caBIG research enterprise is cancer, the infrastructure and many of the tools are generic and may be useful in other therapeutic areas. One recent example has been the launch of the CardioVascular Research Grid (cvrGRID; http://cvrgrid.org) by the National Heart Lung and Blood Institute.

Given the complexity and difficulties of imaging biomarker development, it may be advantageous to apply the open-source paradigm to an integrated development effort for imaging biomarkers. As the foundational image-related tools from caBIG reach maturity, they could form the basis of an open-source development model that integrates image acquisition, analysis, protocol specification, cross-platform compatibility, data standards, validation, and qualification. Such an approach could address development and resource hurdles through the open participation of any interested party, standardization of protocols through centralized development and dissemination, and wide availability of common image analysis tools. Although some development issues would still need to be resolved, such as intellectual property protection and regulatory concerns, an open-source model may ultimately be the most efficient approach to developing, maintaining, and delivering the imaging biomarker tools that can affect drug development.

## CONCLUSIONS: CHALLENGES AND FUTURE OPPORTUNITIES

One of the most salient features of imaging science, which is relevant to both diagnostic imaging and imaging biomarkers, is that the state of the technology does not stand still. Continuous research and development for more than 30 years has produced a steady stream of new imaging technologies as well as incremental improvements that address the limitations of existing technologies. Development and implementation strategies for imaging biomarkers must account for this continuous change and attempt to balance the needs of standardization with the adoption of new innovative technologies. Like software, imaging biomarkers will probably need to be "versioned" to account for changes to the underlying technologies. However, this will also require a continuous process of development and maintenance, for which an open-source

model may be the most appropriate. Moreover, with the continued introduction of new imaging technologies, in the future we might expect the translation of additional image-based biological measures to promising new imaging biomarkers. This will become increasingly feasible as innovative systems make imaging biomarker development more routine.

It seems likely that no one biomarker will be able to adequately capture all of the important aspects of a given disease process or biological response to therapy. A combination of imaging and other biomarkers may be the most effective surrogate endpoint tool for drug development. In addition to the requirements of biomarker validation and qualification, this will require that the component biomarker technologies are cost-effective as a combination.

Current collaborative efforts being led by public–private partnerships such as the FNIH and by the NIH and QIBA are starting to address many of the hurdles of imaging biomarker development and standardization. New approaches that allow contributions from all stakeholders will accelerate the work of imaging biomarker validation and qualification, and we might expect that these initiatives will soon lead to broader availability of new imaging biomarker tools for drug development.

## REFERENCES

1. FDA (2006). Critical path opportunities report. http://www.fda.gov/oc/initiatives/criticalpath/reports/opp_report.pdf.
2. Frey-Wyssling A (1953). *Submicroscopic Morphology of Protoplasm*. Elsevier, Amsterdam.
3. Kuhl DE, Edwards RQ (1963). Image separation radioisotope scanning. *Radiology*, 80:653–662.
4. Wrenn FR, Good ML, Handler P (1951). The use of positron-emitting radioisotopes for the localization of brain tumors. *Science*, 113:525–527.
5. Brownell GL, Sweet WH (1953). Localization of brain tumors with positron emitters. *Nucleonics*, 11:40–45.
6. Hounsfield GN (1973). Computerized transverse axial scanning (tomography): 1. Description of system. *Br J Radiol*, 46:1016–1022.
7. Lauterbur PC (1973). Image formation by induced local interactions: examples employing nuclear magnetic resonance. *Nature*, 242:190–191.
8. Mansfield P, Grannell PK (1973). NMR diffraction in solids. *J Phys C*, 6:L422–L426.
9. Rudin M (ed.) (2005). *Imaging in Drug Discovery and Early Clinical Trials*. Birkhauser, Basel, Switzerland.
10. Weissleder R, Pittet MJ (2008). Imaging in the era of molecular oncology. *Nature*, 452:580–589.
11. Schaeffter T (2005). Imaging modalities: principles and information content. *Imaging in Drug Discovery and Early Clinical Trials*. In Rudin M (ed.), Birkhauser, Basel, Switzerland, pp. 15–81.

12. Barrett HH, Myers K (2003). *Foundations of Image Science*. Wiley–Interscience, Hoboken, NJ.
13. Bernstein MA, King KF, Zhou XJ (2004). *Handbook of MRI Pulse Sequences*. Academic Press, San Diego, CA.
14. Rahmim A, Zaidi H (2008). PET versus SPECT: strengths, limitations and challenges. *Nucl Med Commun*, 29:193–207.
15. Zaidi H (ed.) (2005). *Quantitative Analysis in Nuclear Medicine Imaging*. Springer Verlag, New York.
16. Judenhofer MS, Wehrl HF, Newport DF, et al. (2008). Simultaneous PET-MRI: a new approach for functional and morphological imaging. *Nat Med*, 14:459–465.
17. Defrise M (2001). A short reader's guide to 3D tomographic reconstruction. *Comput Med Imaging Graphics*, 25:113–116.
18. Haacke EM, Brown RW, Thompson MR, Venkatesan R (1999). *Magnetic Resonance Imaging: Physical Principles and Sequence Design*, Wiley–Liss, New York.
19. Rudin M, Beckmann N, Rausch M (2005). Evaluation of drug candidates: efficacy readouts during lead optimization. In Rudin M (ed.), *Imaging in Drug Discovery and Early Clinical Trials*. Birkhauser, Basel, Switzerland, pp. 185–255.
20. Beckmann N (ed.) (2006). *In Vivo MR Techniques in Drug Discovery and Development*. Informa HealthCare, London.
21. Tofts P (ed.) (2003). *Quantitative MRI of the Brain: Measuring Changes Caused by Disease*. Wiley, Hoboken, NJ.
22. Jaffer FA, Weissleder R (2005). Molecular imaging in the clinical arena. *JAMA*, 293:855–862.
23. Massoud TF, Gambhir SS (2003). Molecular imaging in living subjects: seeing fundamental biological processes in a new light. *Genes Dev*, 17:545–580.
24. Aikawa E, Nahrendorf M, Figueiredo J, et al. (2007). Osteogenesis associates with inflammation in early-stage atherosclerosis evaluated by molecular imaging in vivo. *Circulation*, 116:2841–2850.
25. Willmann JK, Chen K, Wang H, et al. (2008). Monitoring of the biological response to murine hindlimb ischemia with $^{64}$Cu-labeled vascular endothelial growth factor-121 positron emission tomography. *Circulation*, 117:915–922.
26. Grimm J, Kirsch DG, Windsor SD, et al. (2005). Use of gene expression profiling to direct in vivo molecular imaging of lung cancer. *Proc Natl Acad Sci USA*, 102:14404–14409.
27. Koo V, Hamilton PW, Williamson K (2006). Non-invasive in vivo imaging in small animal research. *Cell Oncol*, 28:127–139.
28. Pomper MG, Lee JS (2005). Small animal imaging in drug development. *Curr Pharm Des*, 11:3247–3272.
29. Singh M, Johnson L (2006). Using genetically engineered mouse models of cancer to aid drug development: an industry perspective. *Clin Cancer Res*, 12: 5312–5328.
30. Lanza GM, Wickline SA (2003). Targeted ultrasonic contrast agents for molecular imaging and therapy. *Curr Probl Cardiol*, 28:625–653.

31. Kaufmann BA, Lindner JR (2007). Molecular imaging with targeted contrast ultrasound. *Curr Opin Biotechnol*, 18:11–16.
32. Henkelman RM, Stanisz GJ, Graham SJ (2001). Magnetization transfer in MRI: a review. *NMR Biomed*, 14:57–64.
33. Khaleeli Z, Sastre-Garriga J, Ciccarelli O, Miller DH, Thompson AJ (2007). Magnetisation transfer ratio in the normal appearing white matter predicts progression of disability over one year in early primary progressive multiple sclerosis. *J Neurol Neurosurg Psychiatry*, 78:1076–1082.
34. Damadian R (1971). Tumor detection by nuclear magnetic resonance. *Science*, 171:1151–1153.
35. Damadian R, Goldsmith M, Minkoff L (1977). NMR in cancer: XVI. FONAR image of the live human body. *Physiol Chem Phys*, 9:97–100, 108.
36. Sorensen AG (2006). Magnetic resonance as a cancer imaging biomarker. *J Clin Oncol*, 24:3274–3281.
37. Katz R (2004). Biomarkers and surrogate markers: an FDA perspective. *NeuroRX*, 1:189–195.
38. Mills G (2005). Biomarker imaging in drug development and licensed products. http://www.fda.gov/CDER/REGULATORY/medImaging/ImagingWorkshop.ppt.
39. Wagner JA, Williams SA, Webster CJ (2007). Biomarkers and surrogate end points for fit-for-purpose development and regulatory evaluation of new drugs. *Clin Pharmacol Ther*, 81:104–107.
40. Frank R, Hargreaves R (2003). Clinical biomarkers in drug discovery and development. *Nat Rev Drug Discov*, 2:566–580.
41. FDA (2005). Guidance for Industry: Pharmacogenomic data submissions. http://www.fda.gov/CDER/GUIDANCE/6400fnl.pdf.
42. Kuhl D, Edwards R (1970). The Mark III scanner: a compact device for multiple-view and section scanning of the brain. *Radiology*, 96:563–570.
43. Phelps ME, Hoffman EJ, Mullani NA, Higgins CS, Pogossian MMT (1976). Design considerations for a positron emission transaxial tomograph (PETT III). *IEEE Trans Nucl Sci*, 23:516–522.
44. Mansfield P, Pykett IL, Morris PG (1978). Human whole body line-scan imaging by NMR. *Br J Radiol*, 51:921–922.
45. Hargreaves R (2008). The role of molecular imaging in drug discovery and development. *Clin Pharmacol Ther*, 83:349–353.
46. Barrett HH (1990). Objective assessment of image quality: effects of quantum noise and object variability. *J Opt Soc Am A*, 7:1266–1278.
47. Kupinski MA, Clarkson E, Gross K, Hoppin JW (2003). Optimizing imaging hardware for estimation tasks. In *Medical Imaging 2003: Image Perception, Observer Performance, and Technology Assessment*, San Diego, CA. SPIE, Bellingham, WA, pp. 309–313.
48. Clarkson E (2007). Estimation receiver operating characteristic curve and ideal observers for combined detection/estimation tasks. *J Opt Soc Am A*, 24: B91–B98.
49. DePuey EG, Garcia EV, Berman DS (eds.) (2001). *Cardiac SPECT Imaging*. Lippincott Williams & Wilkins, Philadelphia.

50. Taillefer EG (2001). Radiopharmaceuticals. In DePuey EG, Garcia EV, Berman DS (eds.), *Cardiac SPECT Imaging*. Lippincott Williams & Wilkins, Philadelphia, pp. 117–152.
51. Van Train CF, Garcia EV, Cooke CD, Areeda JS (2001). Quantitative analysis of SPECT myocardial perfusion. In DePeuy EG, Garcia EV, Berman DS (eds.), *Cardiac SPECT Imaging*. Lippincott Williams & Wilkins, Philadelphia, pp. 41–64.
52. Lin GS, Hines HH, Grant G, Taylor K, Ryals C (2006). Automated quantification of myocardial ischemia and wall motion defects by use of cardiac spect polar mapping and 4-dimensional surface rendering. *J Nucl Med Technol*, 34: 3–17.
53. Ficaro EP, Lee BC, Kritzman JN, Corbett JR (2007). Corridor4DM: the Michigan method for quantitative nuclear cardiology. *J Nucl Cardiol*, 14:455–465.
54. Leslie WD, Tully SA, Yogendran MS, Ward LM, Nour KA, Metge CJ (2004). Automated quantification of $^{99m}$Tc sestamibi myocardial perfusion compared with visual analysis. *Nucl Med Commun*, 25:833–838.
55. Iskandrian AE, Garcia EV, Faber T (2008). Analysis of serial images: a challenge and an opportunity. *J Nuclear Cardiol*, 15:23–26.
56. Slomka PJ, Berman DS, Germano G (2004). Quantification of serial changes in myocardial perfusion. *J Nucl Med*, 45:1978–1980.
57. Iskandrian AE, Bateman TM, Belardinelli L, et al. (2007). Adenosine versus regadenoson comparative evaluation in myocardial perfusion imaging: results of the ADVANCE phase 3 multicenter international trial. *J Nucl Cardiol*, 14: 645–658.
58. Udelson JE (2008). Lessons from the development of new adenosine A2A receptor agonists. *J Am Coll Cardiol Imaging*, 1:317–320.
59. Shields AF, Price P (2007). *In Vivo Imaging of Cancer Therapy*. Humana Press, Totowa, NJ.
60. Weissleder R (2006). Molecular imaging in cancer. *Science*, 312:1168–1171.
61. Nutt R, Vento LJ, Ridinger MHT (2007). In vivo molecular imaging biomarkers: clinical pharmacology's new PET? *Clin Pharmacol Ther*, 81:792–795.
62. Thompson RC, Cullom SJ (2006). Issues regarding radiation dosage of cardiac nuclear and radiography procedures. *J Nucl Cardiol*, 13:19–23.
63. Einstein AJ, Henzlova MJ, Rajagopalan S (2007). Estimating risk of cancer associated with radiation exposure from 64-slice computed tomography coronary angiography. *JAMA*, 298:317–323.
64. Amis ES, Butler PF, Applegate KE, et al. (2007). *American College of Radiology White Paper on Radiation Dose in Medicine*, 4:272–284.
65. McCollough CH, Bruesewitz MR, Kofler JM (2006). CT dose reduction and dose management tools: overview of available options. *Radiographics*, 26:503–512.
66. McConville PJ, Moody JB, Moffat BA (2005). High-throughput magnetic resonance imaging in mice for phenotyping and therapeutic evaluation. *Curr Opin Chem Biol*, 9:413–420.
67. Hargreaves R, Wagner JA (2006). Imaging as biomarker for decision-making in drug development. In Beckmann N (ed.), *In Vivo MR Techniques in Drug Discovery and Development*. Informa HealthCare, London, pp. 31–44.

68. Bergström M, Långström B (2005). Pharmacokinetic studies with PET. In Rudin M (ed.), *Imaging in Drug Discovery and Early Clinical Trials*. Birkhauser, Basel Switzerland, pp. 279–317.
69. Willmann JK, van Bruggen N, Dinkelborg LM, Gambhir SS (2008). Molecular imaging in drug development. *Nat Rev Drug Discov*, 7:591–607.
70. Major TC, Dhamija S, Black N, et al. (2008). The T- and L-type calcium channel blocker (CCB) mibefradil attenuates leg edema induced by the L-type CCB nifedipine in the spontaneously hypertensive rat: a novel differentiating assay. *J Pharmacol Exp Ther*, 325:723–731.
71. Seddon BM, Workman P (2003). The role of functional and molecular imaging in cancer drug discovery and development. *Br J Radiol*, 76:S128–S138.
72. Uppoor RS, Mummaneni P, Cooper E, et al. (2008). The use of imaging in the early development of neuropharmacological drugs: a survey of approved NDAs. *Clin Pharmacol Ther*, 84:69–74.
73. Evelhoch JL (2007). Magnetic resonance measurement of tumor perfusion and vascularity. In Shields AF, Price P (eds.), *In Vivo Imaging of Cancer Therapy*. Humana Press, Totowa, NJ, pp. 73–84.
74. Gwyther SJ (2007). Anatomical measure of tumor growth with computed tomography and magnetic resonance imaging. In Shields AF, Price P (eds.), *In Vivo Imaging of Cancer Therapy*. Humana Press, Totowa, NJ, pp. 33–46.
75. Miller AB, Hoogstraten B, Staquet M, Winkler A (1981). Reporting results of cancer treatment. *Cancer*, 47:207–214.
76. Lavin PT (1981). An alternative model for the evaluation of antitumor activity. *Cancer Clin Trials*, 4:451–457.
77. Michaelis LC, Ratain MJ (2006). Measuring response in a post-RECIST world: from black and white to shades of grey. *Nat Rev Cancer*, 6:409–414.
78. Therasse P, Arbuck SG, Eisenhauer EA, et al. (2000). New guidelines to evaluate the response to treatment in solid tumors. *J Natl Cancer Inst*, 92:205–216.
79. Tuma RS (2006). Sometimes size doesn't matter: reevaluating RECIST and tumor response rate endpoints. *J Natl Cancer Inst*, 98:1272–1274.
80. Therasse P, Eisenhauer E, Verweij J (2006). RECIST revisited: a review of validation studies on tumour assessment. *Eur J Cancer*, 42:1031–1039.
81. Benjamin RS, Choi H, Macapinlac HA, et al. (2007). We should desist using RECIST, at least in GIST. *J Clin Oncol*, 25:1760–1764.
82. Choi H (2008). Response evaluation of gastrointestinal stromal tumors. *Oncologist*, 13:4–7.
83. Byrne MJ, Nowak AK (2004). Modified RECIST criteria for assessment of response in malignant pleural mesothelioma. *Ann Oncol*, 15:257–260.
84. Schwartz LH, Curran S, Trocola R, et al. (2007). Volumetric 3D CT analysis: an early predictor of response to therapy. *J Clin Oncol* (*Meet Abstr*), 25:4576.
85. Schwartz LH, Colville JAC, Ginsberg MS, et al. (2006). Measuring tumor response and shape change on CT: esophageal cancer as a paradigm. *Ann Oncol*, 17:1018–1023.
86. Cheson BD, Pfistner B, Juweid ME, et al. (2007). Revised response criteria for malignant lymphoma. *J Clin Oncol*, 25:579–586.

87. Gillies RJ, Robey I, Gatenby RA (2008). Causes and consequences of increased glucose metabolism of cancers. *J Nucl Med*, 49:24S–42S.
88. Spaepen K, Stroobants S, Dupont P, et al. (2003). [$^{18}$F]FDG PET monitoring of tumour response to chemotherapy: Does [$^{18}$F]FDG uptake correlate with the viable tumour cell fraction? *Eur J Nucl Med Mol Imaging*, 30:682–688.
89. Gambhir SS, Czernin J, Schwimmer J, Silverman DH, Coleman RE, Phelps ME (2001). A tabulated summary of the FDG PET literature. *J Nucl Med*, 42:1S–93S.
90. Kelloff GJ, Hoffman JM, Johnson B, et al. (2005). Progress and promise of FDG-PET imaging for cancer patient management and oncologic drug development. *Clin Cancer Res*, 11:2785–2808.
91. Weber WA, Petersen V, Schmidt B, et al. (2003). Positron emission tomography in non-small-cell lung cancer: prediction of response to chemotherapy by quantitative assessment of glucose use. *J Clin Oncol*, 21:2651–2657.
92. Mikhaeel NG, Hutchings M, Fields PA, O'Doherty MJ, Timothy AR (2005). FDG-PET after two to three cycles of chemotherapy predicts progression-free and overall survival in high-grade non-Hodgkin lymphoma. *Ann Oncol*, 16:1514–1523.
93. Lin C, Itti E, Haioun C, et al. (2007). Early $^{18}$F-FDG PET for prediction of prognosis in patients with diffuse large B-cell lymphoma: SUV-based assessment versus visual analysis. *J Nucl Med*, 48:1626–1632.
94. Hutchings M, Loft A, Hansen M, et al. (2006). FDG-PET after two cycles of chemotherapy predicts treatment failure and progression-free survival in Hodgkin lymphoma. *Blood*, 107:52–59.
95. Evilevitch V, Weber WA, Tap WD, et al. (2008). Reduction of glucose metabolic activity is more accurate than change in size at predicting histopathologic response to neoadjuvant therapy in high-grade soft-tissue sarcomas. *Clin Cancer Res*, 14:715–720.
96. Stroobants S, Goeminne J, Seegers M, et al. (2003). $^{18}$FDG-positron emission tomography for the early prediction of response in advanced soft tissue sarcoma treated with imatinib mesylate (Glivec). *Eur J Cancer*, 39:2012–2020.
97. Jager PL, Gietema JA, van der Graaf WTA (2004). Imatinib mesylate for the treatment of gastrointestinal stromal tumours: best monitored with FDG PET. *Nucl Med Commun*, 25:433–438.
98. Antoch G, Kanja J, Bauer S, et al. (2004). Comparison of PET, CT, and dual-modality PET/CT imaging for monitoring of Imatinib (STI571) therapy in patients with gastrointestinal stromal tumors. *J Nucl Med*, 45:357–365.
99. Van den Abbeele AD (2008). The lessons of GIST–PET and PET/CT: a new paradigm for imaging. *Oncologist*, 13:8–13.
100. Young H, Baum R, Cremerius U, et al. (1999). Measurement of clinical and subclinical tumour response using [$^{18}$F]fluorodeoxyglucose and positron emission tomography: review and 1999 EORTC recommendations. *Eur J Cancer*, 35:1773–1782.
101. Shankar LK, Hoffman JM, Bacharach S, et al. (2006). Consensus recommendations for the use of $^{18}$F-FDG PET as an indicator of therapeutic response in patients in National Cancer Institute trials. *J Nucl Med*, 47:1059–1066.

102. Boellaard R, Oyen W, Hoekstra C, et al. (2008). The Netherlands protocol for standardisation and quantification of FDG whole body PET studies in multi-centre trials. *Eur J Nucl Med Mol Imaging*, http://dx.doi.org/10.1007/s00259-008-0874-2.
103. Hallett WA, Maguire RP, McCarthy TJ, Schmidt ME, Young H (2007). Considerations for generic oncology FDG-PET/CT protocol preparation in drug development. *IDrugs*, 10:791–796.
104. Weber WA (2005). Use of PET for monitoring cancer therapy and for predicting outcome. *J Nucl Med*, 46:983–995.
105. Westerterp M, Pruim J, Oyen W, et al. (2007). Quantification of FDG PET studies using standardised uptake values in multi-centre trials: effects of image reconstruction, resolution and ROI definition parameters. *Eur J Nucl Med Mol Imaging*, 34:392–404.
106. Lind P, Igerc I, Beyer T, Reinprecht P, Hausegger K (2004). Advantages and limitations of FDG PET in the follow-up of breast cancer. *Eur J Nucl Med Mol Imaging*, 31:S125–S134.
107. Hamaoka T, Madewell JE, Podoloff DA, Hortobagyi GN, Ueno NT (2004). Bone imaging in metastatic breast cancer. *J Clin Oncol*, 22:2942–2953.
108. Mankoff DA, Eary JF, Link JM, et al. (2007). Tumor-specific positron emission tomography imaging in patients: [$^{18}$F]fluorodeoxyglucose and beyond. *Clin Cancer Res*, 13:3460–3469.
109. Kelloff GJ, Krohn KA, Larson SM, et al. (2005). The progress and promise of molecular imaging probes in oncologic drug development. *Clin Cancer Res*, 11:7967–7985.
110. Gupta N, Price PM, Aboagye EO (2002). PET for in vivo pharmacokinetic and pharmacodynamic measurements. *Eur J Cancer*, 38:2094–2107.
111. Bergström M, Grahnén A, Långström B (2003). Positron emission tomography microdosing: a new concept with application in tracer and early clinical drug development. *Eur J Clin Pharmacol*, 59:357–366.
112. Katzenellenbogen JA, Welch MJ, Dehdashti F (1997). The development of estrogen and progestin radiopharmaceuticals for imaging breast cancer. *Anticancer Res*, 17:1573–1576.
113. Smith-Jones PM, Solit D, Afroze F, Rosen N, Larson SM (2006). Early tumor response to Hsp90 therapy using *HER2* PET: comparison with $^{18}$F-FDG PET. *J Nucl Med*, 47:793–796.
114. Mankoff DA, Link JM, Linden HM, Sundararajan L, Krohn KA (2008). Tumor receptor imaging. *J Nucl Med*, 49:149S–163S.
115. Shields AF, Grierson JR, Kozawa SM, Zheng M (1996). Development of labeled thymidine analogs for imaging tumor proliferation. *Nucl Med Biol*, 23:17–22.
116. Shields AF, Grierson JR, Dohmen BM, et al. (1998). Imaging proliferation in vivo with [F-18]FLT and positron emission tomography. *Nat Med*, 4:1334–1336.
117. Buck AK, Halter G, Schirrmeister H, et al. (2003). Imaging proliferation in lung tumors with PET: $^{18}$F-FLT versus $^{18}$F-FDG. *J Nucl Med*, 44:1426–1431.
118. Chen W, Cloughesy T, Kamdar N, et al. (2005). Imaging proliferation in brain tumors with $^{18}$F-FLT PET: comparison with $^{18}$F-FDG. *J Nucl Med*, 46:945–952.

119. Smyczek-Gargya B, Fersis N, Dittmann H, et al. (2004). PET with [$^{18}$F]fluorothymidine for imaging of primary breast cancer: a pilot study. *Eur J Nucl Med Mol Imaging*, 31:720–724.
120. Grierson J, Yagle K, Eary J, et al. (2004). Production of [F-18]fluoroannexin for imaging apoptosis with PET. *Bioconjug Chem*, 15:373–379.
121. Glaser M, Collingridge DR, Aboagye EO, et al. (2003). Iodine-124 labelled Annexin-V as a potential radiotracer to study apoptosis using positron emission tomography. *Appl Radiat Isotopes*, 58:55–62.
122. Lahorte C, Vanderheyden J, Steinmetz N, Wiele C, Dierckx R, Slegers G (2004). Apoptosis-detecting radioligands: current state of the art and future perspectives. *Eur J Nucl Med Mol Imaging*, 31:887–919.
123. Haubner R, Wester H, Burkhart F, et al. (2001). Glycosylated RGD-containing peptides: tracer for tumor targeting and angiogenesis imaging with improved biokinetics. *J Nucl Med*, 42:326–336.
124. Kenny LM, Coombes RC, Oulie I, et al. (2008). Phase I trial of the positron-emitting Arg-Gly-Asp (RGD) peptide radioligand $^{18}$F-*AH111585* in breast cancer patients. *J Nucl Med*, 49:879–886.
125. Beer AJ, Haubner R, Goebel M, et al. (2005). Biodistribution and pharmacokinetics of the $\alpha v\beta 3$-selective tracer $^{18}$F-Galacto-RGD in cancer patients. *J Nucl Med*, 46:1333–1341.
126. Zhang X, Xiong Z, Wu Y, et al. (2006). Quantitative PET imaging of tumor integrin $\alpha v\beta 3$ expression with $^{18}$F-FRGD2. *J Nucl Med*, 47:113–121.
127. Parker GJM, Padhani AR (2003). T1-W DCE-MRI: T1-weighted dynamic contrast-enhanced MRI. In Tofts P (ed.), *Quantitative MRI of the Brain*. Wiley, Hoboken, NJ, pp. 341–364. http://dx.doi.org/10.1002/0470869526.ch10.
128. Tofts PS, Brix G, Buckley DL, et al. (1999). Estimating kinetic parameters from dynamic contrast-enhanced *T*1-weighted MRI of a diffusable tracer: standardized quantities and symbols. *J Magn Reson Imaging*, 10:223–232.
129. He ZQ, Evelhoch JL (1998). Analysis of dynamic contrast-enhanced MRI in tumors: relationship of derived parameters with physiologic factors. In *Proceedings of the International Society for Magnetic Resonance in Medicine*, Sydney, Australia. Wiley, New York, p. 1652. http://cds.ismrm.org/ismrm-1998/PDF6/P1652.PDF.
130. Leach M, Brindle K, Evelhoch J, et al. (2005). The assessment of antiangiogenic and antivascular therapies in early-stage clinical trials using magnetic resonance imaging: issues and recommendations. *Br J Cancer*, 92:1599–1610.
131. O'Connor JPB, Jackson A, Parker GJM, Jayson GC (2007). DCE-MRI biomarkers in the clinical evaluation of antiangiogenic and vascular disrupting agents. *Br J Cancer*, 96:189–195.
132. Jayson G, Waterton J (2005). Applications of dynamic contrast-enhanced MRI in oncology drug development. In Baert A, Jackson A, Buckley D, Parker G (eds.), *Dynamic Contrast-Enhanced Magnetic Resonance Imaging in Oncology*. Springer, New York, 281–298.
133. Galbraith SM, Lodge MA, Taylor NJ, et al. (2002). Reproducibility of dynamic contrast-enhanced MRI in human muscle and tumours: comparison of quantitative and semi-quantitative analysis. *NMR Biomed*, 15:132–142.

134. Roberts C, Issa B, Stone A, Jackson A, Waterton JC, Parker GJ (2006). Comparative study into the robustness of compartmental modeling and model-free analysis in DCE-MRI studies. *J Magn Reson Imaging*, 23:554–563.
135. Parker GJM, Roberts C, Macdonald A, et al. (2006). Experimentally-derived functional form for a population-averaged high-temporal-resolution arterial input function for dynamic contrast-enhanced MRI. *Magn Reson Med*, 56: 993–1000.
136. Morgan B, Thomas AL, Drevs J, et al. (2003). Dynamic contrast-enhanced magnetic resonance imaging as a biomarker for the pharmacological response of PTK787/ZK222584, an inhibitor of the vascular endothelial growth factor receptor tyrosine kinases, in patients with advanced colorectal cancer and liver metastases: results from two phase I studies. *J Clin Oncol*, 21:3955–3964.
137. Mross K, Drevs J, Müller M, et al. (2005). Phase I clinical and pharmacokinetic study of PTK/ZK, a multiple VEGF receptor inhibitor, in patients with liver metastases from solid tumours. *Eur J Cancer*, 41:1291–1299.
138. Thomas AL, Morgan B, Horsfield MA, et al. (2005). Phase I study of the safety, tolerability, pharmacokinetics, and pharmacodynamics of PTK787/ZK222584 administered twice daily in patients with advanced cancer. *J Clin Oncol*, 23: 4162–4171.
139. Prince MR, Zhang H, Morris M, et al. (2008). Incidence of nephrogenic systemic fibrosis at two large medical centers. *Radiology*, 248:807–816.
140. Deo A, Fogel M, Cowper SE (2007). Nephrogenic systemic fibrosis: a population study examining the relationship of disease development to gadolinium exposure. *Clin J Am Soc Nephrol*, 2:264–267.
141. Wertman R, Altun E, Martin DR, et al. (2008). Risk of nephrogenic systemic fibrosis: evaluation of Gadolinium chelate contrast agents at four American universities. *Radiology*, 248:799–806.
142. Penfield JG, Reilly RF (2008). Nephrogenic systemic fibrosis risk: Is there a difference between Gadolinium-based contrast agents? *Semin Dial*, 21:129–134.
143. Sadowski EA, Bennett LK, Chan MR, et al. (2007). Nephrogenic systemic fibrosis: risk factors and incidence estimation. *Radiology*, 243:148–157.
144. Liu G, Rugo HS, Wilding G, et al. (2005). Dynamic contrast-enhanced magnetic resonance imaging as a pharmacodynamic measure of response after acute dosing of AG-013736, an oral angiogenesis inhibitor, in patients with advanced solid tumors: results from a phase I study. *J Clin Oncol*, 23:5464–5473.
145. Lee CP, Taylor NJ, Attard G, et al. (2006). A phase I study of BIBF 1120, an orally active triple angiokinase inhibitor (VEGFR, PDGFR, FGFR) given continuously to patients with advanced solid tumours, incorporating dynamic contrast enhanced magnetic resonance imaging (DCE-MRI). *J Clin Oncol* (*Meet Abstr*), p. 3015. http://meeting.ascopubs.org/cgi/content/abstract/24/18_suppl/3015.
146. Mross KB, Gmehling D, Frost A, et al. (2005). A clinical phase I, pharmacokinetic (PK), and pharmacodynamic study of twice daily BIBF 1120 in advanced cancer patients. *J Clin Oncol* (*Meet Abstr*), p. 3031. http://meeting.ascopubs.org/cgi/content/abstract/23/16_suppl/3031.
147. Collins JM (2003). Functional imaging in phase I studies: decorations or decision making? *J Clin Oncol*, 21:2807–2809.

148. Batchelor TT, Sorensen AG, di Tomaso E, et al. (2007). AZD2171, a pan-VEGF receptor tyrosine kinase inhibitor, normalizes tumor vasculature and alleviates edema in glioblastoma patients. *Cancer Cell*, 11:83–95.
149. Thomas DL, Lythgoe MF, Pell GS, Calamante F, Ordidge RJ (2000). The measurement of diffusion and perfusion in biological systems using magnetic resonance imaging. *Phys Med Biol*, 45:R97–R138.
150. Warach S, Gaa J, Siewert B, Wielopolski P, Edelman RR (1995). Acute human stroke studied by whole brain echo planar diffusion-weighted magnetic resonance imaging. *Ann Neurol*, 37:231–241.
151. Chenevert T, McKeever P, Ross B (1997). Monitoring early response of experimental brain tumors to therapy using diffusion magnetic resonance imaging. *Clin Cancer Res*, 3:1457–1466.
152. Moffat BA, Chenevert TL, Lawrence TS, et al. (2005). Functional diffusion map: a noninvasive MRI biomarker for early stratification of clinical brain tumor response. *Proc Natl Acad Sci USA*, 102:5524–5529.
153. Hamstra DA, Galbán CJ, Meyer CR, et al. (2008). Functional diffusion map as an early imaging biomarker for high-grade glioma: correlation with conventional radiologic response and overall survival. *J Clin Oncol*, 26: 3387–3394.
154. Razek AAKA, Megahed AS, Denewer A, Motamed A, Tawfik A, Nada N (2008). Role of diffusion-weighted magnetic resonance imaging in differentiation between the viable and necrotic parts of head and neck tumors. *Acta Radiol*, 49:364–370.
155. McVeigh P, Syed A, Milosevic M, Fyles A, Haider M (2008). Diffusion-weighted MRI in cervical cancer. *Eur Radiol*, 18:1058–1064.
156. Lee KC, Moffat BA, Schott AF, et al. (2007). Prospective early response imaging biomarker for neoadjuvant breast cancer chemotherapy. *Clin Cancer Res*, 13:443–450.
157. Lee KC, Sud S, Meyer CR, et al. (2007). An imaging biomarker of early treatment response in prostate cancer that has metastasized to the bone. *Cancer Res*, 67: 3524–3528.
158. Helm P, Beg MF, Miller MI, Winslow RL (2005). Measuring and mapping cardiac fiber and laminar architecture using diffusion tensor MR imaging. *Ann NY Acad Sci*, 1047:296–307.
159. Jellison BJ, Field AS, Medow J, Lazar M, Salamat MS, Alexander AL (2004). Diffusion tensor imaging of cerebral white matter: a pictorial review of physics, fiber tract anatomy, and tumor imaging patterns. *Am J Neuroradiol*, 25: 356–369.
160. Mori S, Frederiksen K, van Zijl PCM, et al. (2002). Brain white matter anatomy of tumor patients evaluated with diffusion tensor imaging. *Ann Neurol*, 51: 377–380.
161. Ross BD, Chenevert TL, Kim B, Ben-Yoseph O (1994). Magnetic resonance imaging and spectroscopy: application to experimental neuro-oncology. *Q Magn Reson Biol Med*, 1:89–106.
162. Zhao M, Pipe JG, Bonnett J, Evelhoch JL (1996). Early detection of treatment response by diffusion-weighted $^{1}$H-NMR spectroscopy in a murine tumour in vivo. *Br J Cancer*, 73:61–64.

163. Lyng H, Haraldseth O, Rofstad EK (2000). Measurement of cell density and necrotic fraction in human melanoma xenografts by diffusion weighted magnetic resonance imaging. *Magn Reson Med*, 43:828–836.
164. Ross BD, Moffat BA, Lawrence TS, et al. (2003). Evaluation of cancer therapy using diffusion magnetic resonance imaging. *Mol Cancer Ther*, 2:581–587.
165. Koh D, Collins DJ (2007). Diffusion-weighted MRI in the body: applications and challenges in oncology. *Am J Roentgenol*, 188:1622–1635.
166. Rohde G, Barnett A, Basser P, Marenco S, Pierpaoli C (2004). Comprehensive approach for correction of motion and distortion in diffusion-weighted MRI. *Magn Reson Med*, 51:103–114.
167. Pipe JG, Zwart N (2006). Turboprop: improved PROPELLER imaging. *Magn Reson Med*, 55:380–385.
168. de Langen AJ, van den Boogaart VEM, Marcus JT, Lubberink M (2008). Use of H215O-PET and DCE-MRI to measure tumor blood flow. *Oncologist*, 13:631–644.
169. Goh V, Padhani A (2006). Imaging tumor angiogenesis: functional assessment using MDCT or MRI? *Abdominal Imaging*, 31:194–199.
170. Miles KA (2002). Functional computed tomography in oncology. *Eur J Cancer*, 38:2079–2084.
171. Lamuraglia M, Escudier B, Chami L, et al. (2006). To predict progression-free survival and overall survival in metastatic renal cancer treated with sorafenib: pilot study using dynamic contrast-enhanced Doppler ultrasound. *Eur J Cancer*, 42:2472–2479.
172. Rouffiac V, Bouquet C, Lassau N, et al. (2004). Validation of a new method for quantifying in vivo murine tumor necrosis by sonography. *Invest Radiol*, 39:350–356.
173. Jakobsen JÅ, Oyen R, Thomsen HS, Morcos SK (2005). Safety of ultrasound contrast agents. *Eur Radiol*, 15:941–945.
174. Lassau N, Chami L, Benatsou B, Peronneau P, Roche A (2007). Dynamic contrast-enhanced ultrasonography (DCE-US) with quantification of tumor perfusion: a new diagnostic tool to evaluate the early effects of antiangiogenic treatment. *Eur Radiol Suppl*, 17:89–98.
175. Soret M, Bacharach SL, Buvat I (2007). Partial-volume effect in PET tumor imaging. *J Nucl Med*, 48:932–945.
176. Bacharach SL, Libutti SK, Carrasquillo JA (2000). Measuring tumor blood flow with H215O: practical considerations. *Nucl Med Biol*, 27:671–676.
177. Evelhoch J, Garwood M, Vigneron D, et al. (2005). Expanding the use of magnetic resonance in the assessment of tumor response to therapy: workshop report. *Cancer Res*, 65:7041–7044.
178. Frank R (2008). Quantitative Imaging Biomarkers Alliance FDG-PET/CT Working Group Report. *Mol Imaging Biol*, 10:305.
179. Woodcock J (2005). The Critical Path Initiative: one year later. http://www.fda.gov/CDER/REGULATORY/medImaging/woodcock.ppt.
180. The Biomarkers Consortium (2008). http://www.biomarkersconsortium.org.
181. Altar C (2008). The Biomarkers Consortium: on the critical path of drug discovery. *Clin Pharmacol Ther*, 83:361–364.

182. Foundation for the National Institutes of Health (2008). Biomarkers: FDG-PET lung and lymphoma. http://www.fnih.org/index.php?option=com_content&task=view&id=503&Itemid=638.
183. Foundation for the National Institutes of Health (2008). Biomarkers: carotid MRI reproducibility study. http://www.fnih.org/index.php?option=com_content&task=view&id=489&Itemid=600.
184. Nevitt MC (2007). Osteoarthritis Initiative (OAI): Design, subject characteristics, data and images. http://www.oai.ucsf.edu/datarelease/docs/presentations/oarsi12062007/Nevitt_OARSI2007.pdf.
185. OAI:Home (2008). http://www.oai.ucsf.edu/datarelease/.
186. Foundation for the National Institutes of Health–Alzheimer's Disease Neuroimaging Initiative (2008). http://fnih.org/index.php?option=com_content&task=view&id=103&Itemid=227.
187. Alzheimer's Disease Neuroimaging Initiative LONI–ADNI (2008). http://www.loni.ucla.edu/ADNI/.
188. Maurer SM, Scotchmer S (2006). *Open Source Software: The New Intellectual Property Paradigm*. National Bureau of Economic Research Working Paper Series, no. 12148. http://www.nber.org/papers/w12148.
189. Rai AK (2004). Open and collaborative research: a new model for biomedicine. http://ssrn.com/paper=574863.
190. Maurer SM (2007). Open source drug discovery: finding a niche (or maybe several). http://ssrn.com/paper=1114371.
191. Munos B (2006). Can open-source R&D reinvigorate drug research? *Nat Rev Drug Discov*, 5:723–729.
192. Eschenbach ACV, Buetow K (2006). Cancer Informatics Vision: caBIG. *Cancer Inf*, 2:22–24.
193. Piwowar HA, Becich MJ, Bilofsky H, Crowley RS (2008). Towards a data sharing culture: recommendations for leadership from academic health centers. *PLoS Med*, 5:e183 EP.
194. Prior FW, Erickson BJ, Tarbox L (2007). Open Source Software Projects of the caBIG In Vivo Imaging Workspace Software Special Interest Group. *J Digital Imaging*, 20:94–100.
195. Jezzard P, Ramsey NF (2003). Functional MRI. In Tofts P (ed.), *Quantitative MRI of the Brain: Measuring Changes Caused by Disease*. Wiley, Hoboken, NJ, pp. 415–453.
196. Börnert P, Keupp J, Eggers H, Aldefeld B (2007). Whole-body 3D water/fat resolved continuously moving table imaging. *J Magn Reson Imaging*, 25:660–665.
197. St Pierre TG, Clark PR, Chua-anusorn W, et al. (2005). Noninvasive measurement and imaging of liver iron concentrations using proton magnetic resonance. *Blood*, 105:855–861.
198. Stafford RJ, Hazle JD (2006). Magnetic resonance temperature imaging for focused ultrasound surgery: a review. *Top Magn Reson Imaging*, 17:153–163.
199. Zhou J, Payen J, Wilson DA, Traystman RJ, van Zijl PCM (2003). Using the amide proton signals of intracellular proteins and peptides to detect pH effects in MRI. *Nat Med*, 9:1085–1090.

200. Woods M, Woessner DE, Sherry AD (2006). Paramagnetic lanthanide complexes as PARACEST agents for medical imaging. *Chem Soc Rev*, 35:500–511.
201. van Zijl PCM, Jones CK, Ren J, Malloy CR, Sherry AD (2007). MRI detection of glycogen in vivo by using chemical exchange saturation transfer imaging (glycoCEST). *Proc Natl Acad Sci USA*, 104:4359–4364.
202. Eisenhauer EA, Therasse P, Bogaerts J et al. (2009) New response evaluation criteria in solid tumors: revised RECIST guideline (version 1.1). *Eur J Cancer*, 45:228–247.

# PROTEIN BIOMARKER DISCOVERY USING MASS SPECTROMETRY–BASED PROTEOMICS

Joanna M. Hunter, Ph.D., and Daniel Chelsky, Ph.D.
*Caprion Proteomics, Inc., Montreal, Quebec, Canada*

## INTRODUCTION

Recent advances in biological sample preparation, automated sample handling, and sensitive mass spectrometry with a variety of sample ionization sources have catalyzed the emergence of these methods as tools for clinical proteomic analyses [1,2]. Concurrently, proteomic data analysis has been aided by the expansion and improvement of publicly available protein and gene databases. Among the methods reported for quantitative comparison of proteins of biological origin, several techniques based on liquid-phase separation of proteins and peptides rather than gel electrophoresis have been optimized and implemented within the last few years. These include stable isotope labeling as well as label-free methods [3–9]. Direct measurement and comparison of unlabeled peptide ion peak intensities has become the standard, particularly for high-throughput applications. Quantitative, label-free mass spectrometry–based methods for protein expression profiling include those based on pattern recognition [10,11], peptide counting [12–14], and peptide ion intensity or area [4,5,15–20]. Of these, the latter seems to be the most robust and applicable to clinical studies. The correlation between observed peptide intensity differences and protein abundance differences is the fundamental basis of label-free protein expression profiling. In particular, the rela-

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

tionship between liquid chromatography–mass spectrometry (LC-MS) signals and peptide abundance has been shown to vary linearly with concentration, even in complex samples [5,21–23].

Mass spectrometry–based protein expression profiling, without the use of labels or internal standards, is applicable to a wide range of sample types, including tissues, cells, organelles, and organisms, as well as body fluids such as plasma, cerebrospinal fluid, and urine. It has the additional advantage that comparisons can be made post hoc on a per patient basis, thereby dramatically increasing the granularity of information that can be obtained from the data. Multiple approaches can be taken for the label-free identification of differentially expressed proteins in large sample sets. At Caprion we have developed a specific approach, called Cell*Carta*, which has proven to be very effective and reliable. Features of this platform will be used as an example in describing issues and solutions for biomarker discovery.

The platform has low overall variability [<15% coefficient of variation (CV) in the absence of biological variability] and provides an accurate and quantitative measure of the biological modulation of protein abundance. Thus, statistically significant differences in plasma protein abundance can be detected, making this technology highly applicable to biomarker discovery in clinical samples. Plasma markers of disease, drug reversion of disease, predictive markers of drug response, pharmacodynamic markers of drug action, and early markers of both efficacy and toxicity can be detected using this technology. Studies may be conducted in animal models (preclinical) as well as with human clinical samples [24–26].

One requirement for a useful biomarker is that it can be measured in an easily accessible body fluid. In this regard, blood (serum or plasma) is preferred, and it is used routinely for many clinical assays [27]. However, the discovery of biomarkers in blood brings some technical challenges. Plasma has been estimated to contain anywhere from several hundred to several thousand proteins [28,29], whose concentrations are known to vary by up to 10 orders of magnitude. Many of the proteins with desirable biomarker characteristics are expected to be at the low end of the plasma protein dynamic range of concentration, typically at or below 100 ng/mL. Two approaches to this issue are described here. One approach depends on the immunoaffinity-based depletion of some of the most abundant plasma proteins, thus making it possible to see more of the lower-abundance proteins. The second approach is based on identifying and comparing plasma proteins just prior to release from their cells or tissue of origin. While residing in the secretory apparatus, proteins destined for the blood are highly concentrated, thus allowing for a comprehensive picture of tissue-specific markers.

## PROTEIN PROFILING PLATFORM FOR BIOMARKER DISCOVERY

One of the major challenges of biomarker discovery is the execution of unbiased large-scale studies to avoid the confusion of process artifacts with actual

biomarkers. Unbiased experimental design and study implementation are arguably among the most critical components of a differential protein expression platform [30]. Because subtle errors in a work plan can result in large biases, careful study planning and implementation are necessary in order to discover biomarkers that are robust to population variation, sample collection variability, processing artifacts, and other confounding factors. The first step in setting up a study is to clearly define the questions that must be addressed. From these goals, the minimum number of patients in each condition to be compared can be fixed. Although the group sizes can be small from a purely analytical perspective, the number of samples required is dictated by the possibility of confounding factors within the target group and/or by the need to sample a diverse target population. Multiple comparisons across large sample sets are achievable and fully exploit the analytical platform. In addition, the order in which samples are processed and analyzed must be considered. Depending on the comparisons to be made, samples must be interleaved and/or randomized between processing steps. Alternatively, the samples to be compared are block randomized, with comparisons made within each block, thus ensuring the greatest processing consistency within a comparison. After samples have been processed, data analysis techniques are employed to ensure that biases have not occurred. If biases appear, appropriate sample pairing and statistical tests are utilized to ensure that the biases do not influence the conclusions drawn from the study.

To generate highly reproducible results, stringent control of sample preparation, analysis, and data normalization procedures are required. Low peptide signal intensity variation across study samples is required to detect statistically significant and subtle protein changes against background variation. Protocols followed rigorously, including quality control checks, are necessary throughout the study, from sample collection to data analysis. For example, strictly following a standard operating procedure (SOP) for blood collection and generation of plasma enables blood samples to be collected from multiple sites for the same study. Furthermore, a well-documented sample "chain of custody" is obligatory when handling clinical samples. As part of this documentation process, samples are barcoded, logged into a laboratory information management system (LIMS), and stored in locked freezers. All subsequent steps in sample handling and analysis are monitored and logged, and all data are secured. Thus, SOPs and quality control procedures ensure sample integrity throughout the process. Specific features of the Cell*Carta* work flow are detailed below.

## Focusing on the Relevant Proteome

Without enrichment of a targeted proteome, samples contain far too many proteins to be visualized in a reasonable number of analytical runs, and only the most abundant proteins are detected. Enrichment of subproteomes is necessary to reduce the sample complexity and to decrease the limit of detection. More comprehensive protein identification as well as information on

subcellular localization is obtained by isolating organelles such as plasma membranes, endosomes, phagosomes, mitochondria, endoplasmic reticulum, nuclei, or Golgi apparatus from cells or tissue homogenates. Similarly, blood (plasma or serum) is enriched for lower abundance proteins by the removal of high-abundance proteins. For example, depletion of the 14 most abundant proteins using the Multiple Affinity Removal System (MARS, Agilent Technologies) removes over 96% of total protein and consequently enhances the detection of the remaining lower-abundance components [31,32]. Various depletion approaches are being used to get deeper into the plasma proteome [33]. The key criteria for implementing enrichment or depletion strategies are that they provide minimal contamination of the nontargeted proteins and are highly reproducible.

## Sample Fractionation

Regardless of the origin of the purified protein samples, the complexity is typically too great—both in terms of the number of proteins and the dynamic range of concentration—for direct analysis by mass spectrometry. Reversed-phase liquid chromatography coupled to mass spectrometry (LC-MS) can reproducibly distinguish and quantify over 5000 peptides in an hour-long analytical run. A single LC-MS injection is therefore suitable for samples comprised of some 500 to 1000 proteins. Subproteome enrichment typically reduces the total number of proteins to fewer than 10,000. To characterize these complex samples more comprehensively, a fractionation step must be applied before the LC-MS. At the protein level, separations such as strong anion exchange provide effective fractionation prior to trypsin digestion and LC-MS. Alternatively, the proteins are first proteolyzed to tryptic peptides, with subsequent separation of the peptides by strong cation exchange (SCX) liquid chromatography [34]. The Cell*Carta* work flow employs this method.

## Differential Protein Expression by LC-MS

For optimal matching of peptide signals across many samples, the LC-MS platform must be as reproducible as possible (Figure 1). Technically, the requirements of stable chromatography, high resolution, and mass accuracy may be addressed by a variety of instruments. Peptide profiling takes place on mass spectrometers that are capable of providing sensitive, high-mass-accuracy measurements, such as time-of-flight or high-resolution ion-trapping instruments. Peptide sequencing is performed on quadrupole-time-of-flight instruments such as the QSTAR (Applied Biosystems) or the QTOF (Waters), or on an ion trap such as the LTQ-Orbitrap (ThermoFisher), which are sensitive and can provide high-quality MS/MS spectra. Regardless of the mass spectrometer type, capillary reversed-phase LC columns are coupled to these instruments via an electrospray interface. The Cell*Carta* mass spectrometry platform utilizes capillary liquid chromatography (CapLC) and QTOF LC-MS

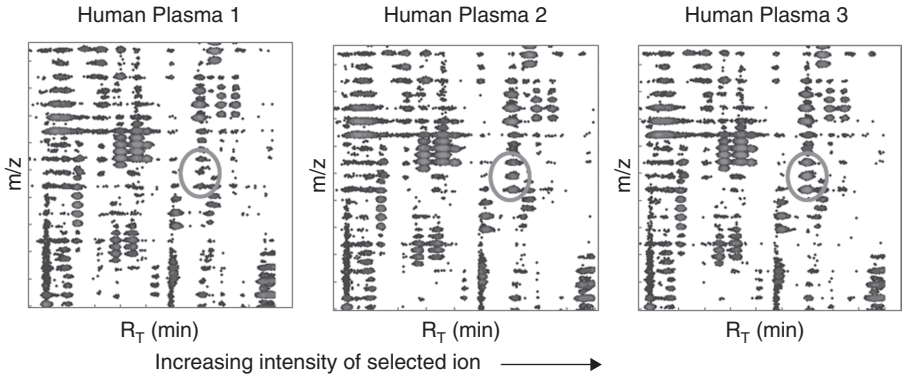


**Figure 1** Consistent peptide intensity across all samples allows the detection of differentially expressed peptide ions. Shown is a partial view of peptide ion maps (as measured by LC-MS) from the plasma of three individuals. The horizontal axis is chromatographic retention time, the vertical axis is mass-to-charge ratio ($m/z$), and the peptide ion intensity is denoted by the size and color of the spots. The peptide ion circled shows differential expression across patients and increases in abundance from sample 1 to sample 3. (*See insert for color reproduction of the figure.*)

systems (Waters). The LC-MS measurements are reliable and reproducible, providing a median CV of peak intensity of 8 to 9% for peptides matched across six replicate injections of a standard sample.

To quantify the differentially expressed peptides, a suite of proprietary bioinformatics tools was developed at Caprion and implemented into Cell*Carta*. Matching peptides across large sample sets (peak alignment) is the first step in the differential expression analysis (Figure 2). Algorithms to perform mass-to-charge ratio ($m/z$) and chromatographic retention-time alignment result in the confident detection of significant peptide intensity differences between samples. Peptide ions are detected and matched across all the samples in the study. Each study peptide is characterized by $m/z$, charge, retention time, and intensity. Bioinformatics software tools map peptides across all samples, comparing ion intensity for all reproducible peptide ions across all fractions. Those peptides that show a statistically significant differential abundance are targeted for protein identification.

## Protein Identification

Identification of the candidate protein biomarkers is accomplished using two different and complementary tools. First, database searching of LC-MS/MS spectra for peptide identification is accomplished using Mascot (MatrixScience), which searches for the best fit between the recorded spectra and theoretical spectra calculated from tryptic peptides from proteins in the reference

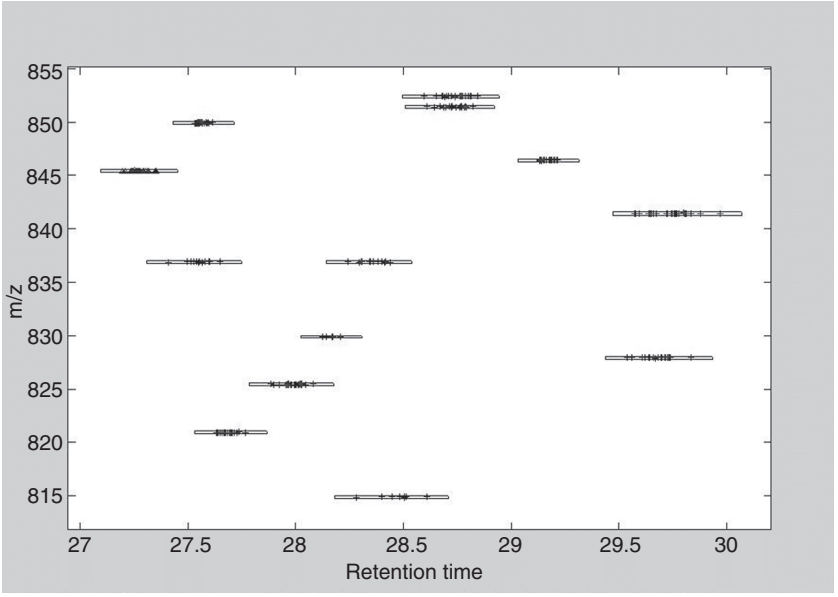


**Figure 2** Related peptide ions are clustered across the entire study. High confidence matching is achieved, independent of small variations in mass and retention time detected. Each symbol represents a detected ion in one sample, plotted at the observed mass-to-charge ratio and chromatographic retention time.

database. For human plasma studies, the human International Protein Index (IPI; European Bioinformatics Institute) is searched. For samples not of human origin, other public databases may be used, depending on the species. The search results are parsed based on peptide and protein score thresholds that are determined by setting an acceptable false-positive rate of identification calculated by searching a randomized database [35].

In addition, we have developed novel software that adds a differential abundance correlation filter to mass and retention-time fingerprinting [virtual mass spectrometry (VMS)]. Protein identification using mass and retention-time fingerprinting alone suffers from the high level of mass redundancy for peptides from complex species. VMS incorporates the protein identification and peptide expression resolver (PIPER) filter, which requires all peptides pointing to the same protein to share the same relative intensity across all samples, a logical requirement for association. Application of the PIPER filter to the identified peptides results in false-positive protein identification rates below 5% [36]. This method provides for the identification of proteins where the peptide ions can be detected but are not sufficiently intense to provide an interpretable MS/MS fragmentation pattern [37]. In addition to adding new protein identifications, VMS can also be used to supplement LC-MS/MS by

increasing confidence in low-scoring sequence matches or single-peptide identifications. Thus, the combined use of LC-MS/MS sequencing with VMS greatly expands the ability to identify and quantify novel proteins by expanding the dynamic range (number and concentration) of identifiable proteins as well as the confidence (number of identified peptides per protein) of protein identification.

## Candidate Biomarker Results Visualization

Proteomics analyses can often generate large data sets, including all of the associated peptide sequence and expression data. To enable rapid evaluation and interpretation of results, it is very helpful to organize the output into a searchable database with a specialized but simple user interface. We have created such a tool that is Java-based and accessible on a secure Web portal. The Data Report provides query, filtering, and data analysis capabilities as well as visualization of peptide differential expression and protein identification across all of the samples in the study. Included in the Data Report are listings of all the peptides, protein accession numbers, and gene names identified, along with differential expression information for each. Graphical displays of peptide intensities, differential expression, and MS/MS spectra are also available (Figure 3). In addition, selected data can be exported into standard formats.

## Candidate Biomarker Panels

Historically, circulating biomarkers have been defined by a single analyte that can be used to distinguish given groups of individuals with high specificity and sensitivity (positive and negative predictive values). In the last two decades, the development of high-throughput multivariate analytical platforms, such as protein and gene chips, has enabled the identification and validation of multivariate biomarkers. Due to population variation and the multitude of effects that treatment can have within a person and across the population, it is acknowledged that panels of proteins are likely to produce a biomarker that is significantly more sensitive and specific than any single protein alone [38].

In a typical biomarker discovery study performed on the Cell*Carta* platform, hundreds of proteins are found to be significantly differentially modulated. These proteins are then prioritized into panels of 3 to 10 proteins. This process involves filtering based on several criteria, including diagnostic strength [i.e., the area under the receiver operator characteristic (ROC) curve or AUC] and linear regression to clinical factors. In particular, the panel of proteins must have a high composite AUC in addition to strong individual protein AUC scores. Biomarker panels with high discriminating power (positive and negative predictive values) can thus be composed from the individual biomarker candidates. The number of protein candidates included in the panel

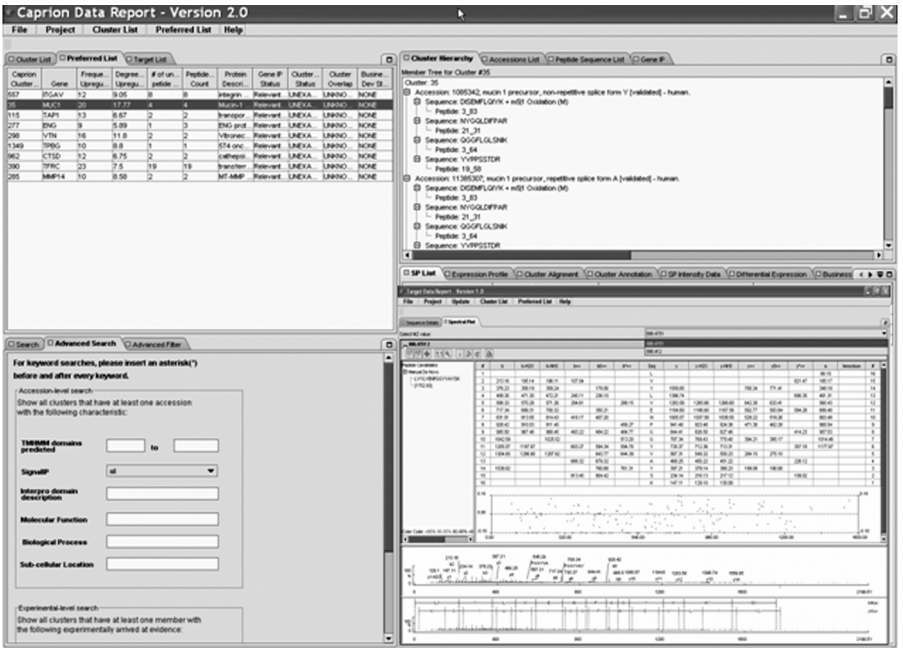


**Figure 3** Caprion Data Report user interface. An example Data Report screenshot is shown. The upper left panel contains a summary of the proteomic data for the proteins selected by the user. The upper right panel contains several tabs that permit additional details to be displayed, including a hierarchical clustering of the peptides. The lower left panel contains search options, while the lower right panel contains several tabs that provide additional proteomic data, such as peptide intensities, differential expression, and MS/MS spectra.

represents a trade-off between cost and risk. Although larger protein panels are more costly to verify, they are more resilient to population variability. Prioritized candidate biomarkers are then qualified and verified further using immunoassays such as enzyme-linked immunosorbent assay (ELISA) or mass spectrometry–based assays such as multiple reaction monitoring (MRM) [39,40].

## PROTEIN BIOMARKER DISCOVERY IN BLOOD PLASMA

The Cell*Carta* platform enables the measurement of biological changes of interest in human plasma profiling studies for protein biomarker discovery. The following examples illustrate that protein biomarkers can be found in plasma using a highly controlled proteomics technology platform and employing an appropriate study design.

## Global Proteomics Applied to Alzheimer Disease

A collaboration with Hyman Schipper, a neuropsychologist at the McGill University Memory Clinic, part of the Jewish General Hospital (JGH; Montreal, Canada), was undertaken to investigate the effect of drug treatment on Alzheimer disease (AD) patients. Plasma was collected from patients after obtaining written informed consent and with approval of the research and ethics committee of the JGH. Included in the study were 33 age-matched healthy control, 19 untreated AD, and 25 donepezil-treated AD patients. Each patient was administered the Folstein Mini-Mental State Examination (MMSE), a qualitative test of disease severity [41]. In addition, clinical histories were obtained.

In the course of the study, a "global proteomics" approach to data visualization and analysis was developed and implemented. Following unsupervised clustering by multidimensional scaling (MDS), the untreated disease group was found to cluster apart from the healthy group (Figure 4). Subsequently, a centroid (median value in three dimensions) was calculated for each group. The line intersecting the normal and disease group centroids is defined as the

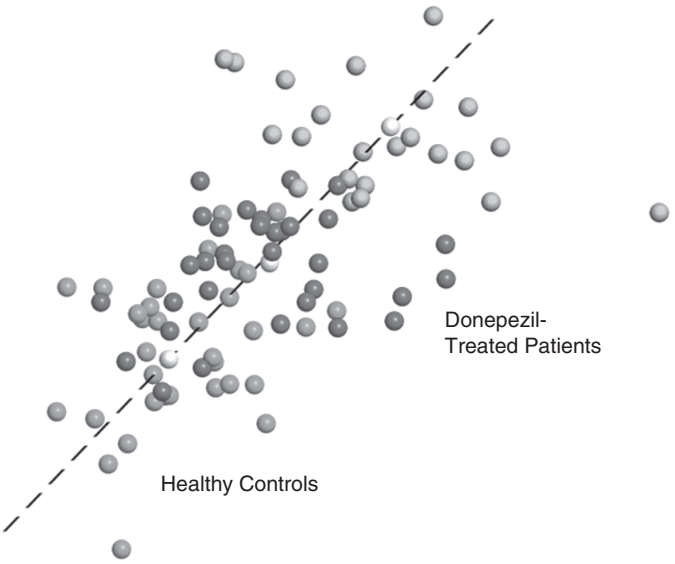


**Figure 4** Global proteomics analysis on Alzheimer disease and normal patients. Multidimensional scaling of proteomics data demonstrates the separation of healthy individuals (green spheres) from Alzheimer patients (red spheres). Caprion has defined a disease axis that is used to quantify relative disease state. The axis is a line that passes through the disease and healthy centroids (yellow spheres). Each patient is then positioned on the axis according to its orthogonal intercept. Donepezil-treated Alzheimer patients (purple spheres), as a group, are shifted on the disease axis from the disease group toward the healthy group. (*See insert for color reproduction of the figure.*)

*disease axis* (Figure 4). The disease severity for each patient is indicated by the distance between its orthogonal intercept with the disease axis and the normal centroid. The disease severity of each patient, as determined by this approach, correlated well with the patient MMSE score, with a Pearson correlation coefficient of 0.75. When the relative position on the disease axis was matched to relative peptide intensity, 282 peptides were found to correlate with the disease severity profile. The proteins corresponding to these peptides were subsequently identified. Although there is currently no blood-based diagnostic test for AD, these results suggest that a test monitoring blood protein levels could measure AD severity.

The global proteomics approach was applied to assess the effect on AD patients of treatment with donepezil. A total of 75 peptides were found to be highly correlated with the disease severity profile of treated versus untreated AD patients. These peptides are a subset of the 282 disease-related peptides. The Cell*Carta* platform can thus be employed as an assay that enables the analysis of drug response for patient segregation. As an extension, pharmacodynamic analyses such as dose optimization and drug efficacy could be performed using this method. This study demonstrates that global proteomics can be an effective approach to pharmacodynamic biomarker identification.

## Oncology Biomarkers: Blinded Case Study

The goal of this second study was to identify circulating markers of ovarian and breast cancer. Two sets of patient samples were examined. The first set was composed of eight samples each from patients with breast cancer or ovarian cancer as well as eight matched healthy subjects. The second set of samples had the same composition, but the identities were blinded to Caprion. All samples were processed concomitantly though the Cell*Carta* platform. Of the 53,628 peptide ions detected across the study, 4089 were differentially modulated reproducibly and significantly ($p < 0.005$) in one of the three cohorts. These peptides were further analyzed by the bioinformatics tool, multidimensional scaling, to determine the relationship between each of the patients at the global proteomics level. Results, shown in Figure 5, indicate that each of the three groups of samples cluster separately. The samples in each group are therefore more similar to other members of the same group than to members of other groups. This similarity within a group translates to peptides (and proteins) that distinguish and separate each group.

The behavior of the peptides that were found to distinguish each of the three groups was analyzed for each of the 24 blinded samples. Each sample was therefore categorized according to which of the three groups it was most related. When the samples were unblinded, it was determined that 22 of the 24 assignments were correct. Only one pair, consisting of a normal and a breast cancer plasma sample, was assigned incorrectly.

In this example, all differentially expressed peptides were used to discriminate the three groups in order to maximize the ability of the small learning

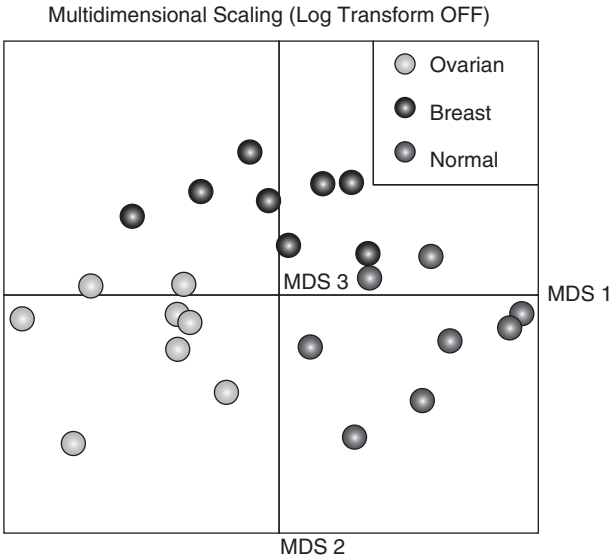


**Figure 5** Sample groups distinguished by differentially expressed peptides. Multidimensional scaling analysis was performed using the intensity values for 4089 differentially expressed peptide ions from 24 samples. Separation along three axes of variance (MDS1 to MDS3) is shown, where each sphere represents a patient sample. The groups are identified by the colors indicated. (*See insert for color reproduction of the figure.*)

set (first 24 samples) to determine the identity of the blinded sample test set. None of the individual peptide ions could discriminate the three groups very well. However, sets of five or 10 peptides were much more effective, as shown in the receiver operating characteristic (ROC) plots in Figure 6. In the case of multiple peptide analysis, peptides were chosen randomly from the population of differentially expressed peptides in order to provide multiple examples of the range of discrimination possible. The AUC required to fully discriminate groups is a value of 1. This was most frequently achieved as the panel size increased from one to 10 proteins (Figure 6, right panels).

To better understand the biology underlying the differentially expressed peptides, their parent proteins were identified. This was accomplished by LC-MS/MS and VMS to identify groups of peptides representing a single protein. This result was further filtered by requiring that the relative intensity of a minimum of three peptide ions be well correlated across all samples. Approximately 200 proteins met these and other strict confidence criteria. Half of these proteins distinguished either breast cancer or ovarian cancer from the other two conditions. The remaining half of the proteins distinguished cancer (breast or ovarian) from normal healthy subjects. These results suggest that circulating blood proteins provide a rich source of high-quality

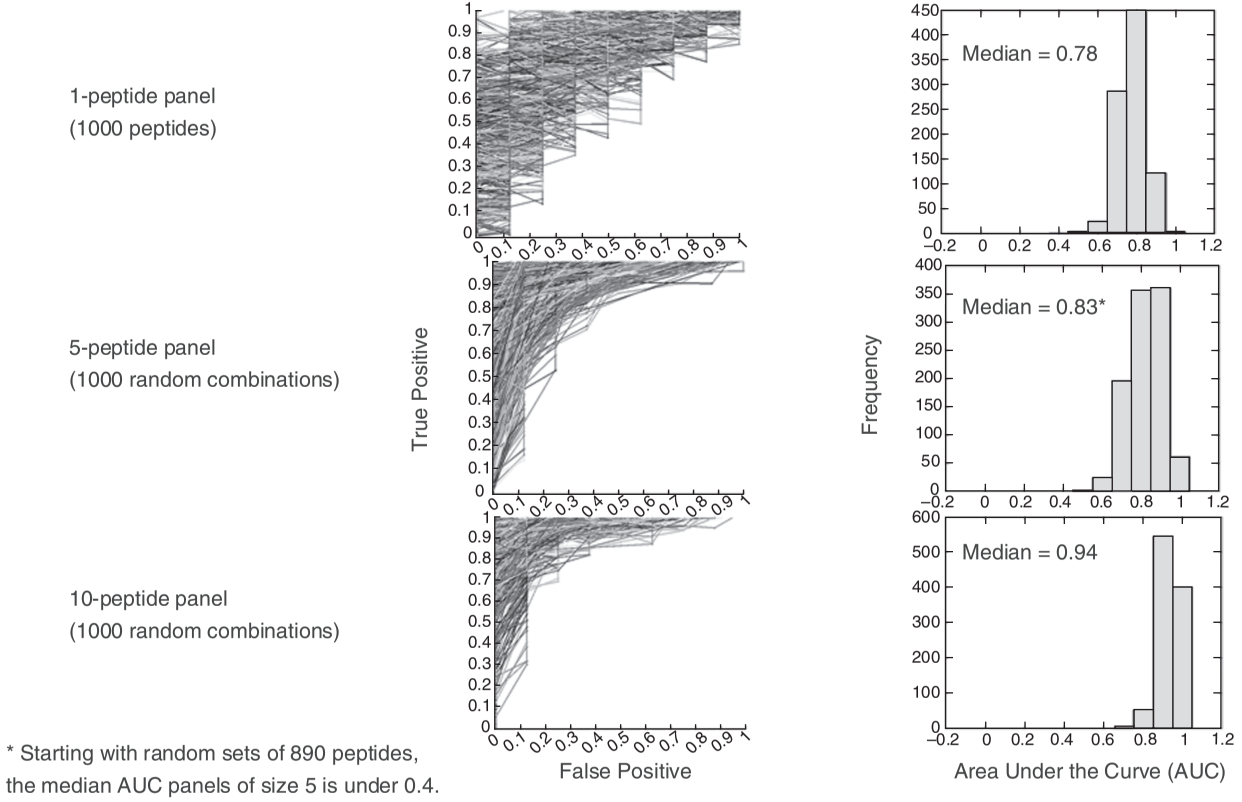


**Figure 6** Multiple peptide panels are effective at discriminating groups. ROC plots show improved performance when going from one to five to 10 peptides. Displayed are curves (left) representing true-positive ($x$-axis) and false-positive ($y$-axis) ratios and the area under the curve (AUC; right column) for single-peptide (top), five-peptide (middle) and 10-peptide panels (bottom). In the case of multiple-peptide analysis, peptides were chosen randomly from the population of differentially expressed peptides. The optimal AUC for discriminating groups is a value of 1. This was best achieved as the panel size increased. (*See insert for color reproduction of the figure.*)

disease biomarkers that can be discovered using mass spectrometry–based protein expression profiling.

## PROTEIN BIOMARKER DISCOVERY IN THE SECRETORY APPARATUS

### Secreted Proteins as Biomarkers

In the quest for circulating biomarkers of disease or drug response, much effort has been made to avoid the overwhelming effect of high-abundance proteins in the blood that obscure lower abundance and potentially more interesting proteins. Removal of the high-abundance proteins with multiple affinity antibody columns has been very effective in this regard. An alternative approach that has been successful as well is to isolate plasma proteins as they are being released from the source tissues or cells. One common such method involves harvesting proteins released by tissue culture cells into conditioned media [42,43]. Although this approach has been productive, proteins may be lost due to dilution or adherence to the culture container, and it is generally limited to cultured cells rather than whole tissue. Analysis of tissues can provide a more physiological context, as they include primary cells as well as the natural composition of multiple cell types that may modulate secretion. Tissues can also be taken from the host, typically animal models of disease, following drug treatment, in order to assess the impact of treatment on the secretome. This can lead to a better understanding of the mechanism of action of the drug treatment as well as the identification of possible markers of therapeutic response. Proteins released from tissue into culture medium have also been studied [44], but contamination from lysed cells and culture medium can obscure the actual proteins secreted. Reliance on those proteins with a signal sequence as well as isotopic labeling of de novo synthesized proteins from the tissue are usually necessary.

An alternative approach to identifying and quantifying secreted proteins relies on the isolation of such proteins directly from the secretory pathway of the cells or tissue. The contents of the Golgi and related secretory vesicles provide a highly concentrated collection of proteins, just prior to their release from the cell [45,46]. This method avoids dilution into the plasma, which would obscure all but the more abundant secreted proteins. Dilution as well as contamination by cell lysis proteins in culture medium is also eliminated as a concern. Tissues can be as easily investigated as cultured cells, with the added advantage that the samples do not need to be viable at the time of analysis. Thus, frozen surgical samples from patients or animal models are amenable to study and comparison, even when samples are collected at different time points. To isolate secretory proteins directly from cell culture or tissue, the homogenized samples are separated on a density gradient to isolate secretory vesicles.

## Secreted Proteins Differentially Expressed in Prostate Cancer

The prostate secretes a range of proteins related to its function, including proteolytic enzymes, acid phosphatase (ACPP), and prostate-specific antigen (PSA). In prostate tumors, some of these proteins are known to increase or decrease in the amount secreted. To test the application of secretory vesicle isolation to frozen human surgical specimens, samples were obtained from patients with prostate cancer, with institutional review body (IRB) approval and patient consent. Tumor and normal tissue were separated as described [47] and the content of the secretory apparatus was isolated. Proteins were digested with trypsin and analyzed by LC-MS in order to compare expression between each normal and tumor pair from all six patients. Significantly differentially expressed peptides were identified. Four of the best known prostate cancer–associated secreted proteins were found to be up-regulated, as expected, in the tumors. PSA, ACPP, kallekrin 2 (KLK2), and macrophage migration inhibitory factor (MIF) are all low-abundance proteins in the blood, at levels ranging from 0.2 to 3.5 ng/mL in normal individuals. Detection of these proteins by direct LC-MS analysis of plasma would not be possible, yet was readily accomplished by examination of the secretory pathway. Two of these proteins, PSA and MIF, were also evaluated by commercial ELISA in the plasma of the same patients. The levels observed in the secretory vesicles by unlabeled mass spectrometry were directly comparable to levels in the plasma from the same patients, with correlation coefficients of 0.69 and 0.74, respectively [46]. Thus, analysis of differentially expressed proteins in the secretory apparatus is predictive of relative expression in the plasma. This correlation means that discovery can be done in the highly concentrated milieu of the vesicles, while verification and validation studies can be conducted with more sensitive antibody-based assays directly in blood.

In addition to the well-known prostate cancer markers, other known cancer-related proteins were found, for a total of 40 proteins with known cancer association. A further 20 proteins were identified as differentially expressed in the tumors that were not previously known to have a cancer association. The known function of the proteins typically fit well with their observed expression patterns. For example, of 11 proteins involved with sugar metabolism, all were found to go up in the tumors. On the other hand, of nine proteins involved with contraction and adhesion, six were found to be expressed at lower levels in the tumor samples than in the matching normal tissue.

## Proteins Secreted by Visceral Adipose Tissue

Visceral adipose tissue is much more than a storage medium for excess energy. This realization came with the discovery in 1994 of leptin and its wide-ranging activities, which include reduction of appetite, angiogenesis, hematopoiesis, and bone formation [48–50]. Visceral adipose tissue is now recognized as an endocrine organ that releases a large number of biologically active molecules

called adipokines, which have pleiotropic effects on a variety of metabolic pathways [48,51]. Proteins released from adipose tissue include regulators of inflammation, inducers of angiogenesis, and modulators of hypertension, among others. Visceral adipose tissue, in particular, has been implicated in the release of proteins responsible for metabolic syndrome and type 2 diabetes [52]. For this reason, evaluation of visceral adipose tissue, either alone or in comparison to subcutaneous adipose tissue, is of great interest [52,53].

As a pilot study of the eventual comparison of proteins secreted from visceral and subcutaneous adipose tissue, visceral adipose tissue was obtained from three obese patients undergoing gastric bypass surgery. The contents of the secretory apparatus were isolated from each subject and analyzed by LC-MS/MS to determine feasibility and yield, as well as to catalog the proteins and verify that known adipocyte-secreted proteins could be found. In a limited analysis, 155 proteins were identified, including many of the well-known adipokines. Among the proteins identified were cytokines (TGFb1; IL 25), acute-phase proteins (SPARC; lipocalin 2/NGAL), proteins that regulate blood pressure (angiotensinogen; ARTS-1), regulators of lipid metabolism (retinol-binding protein 4; lipoprotein lipase), and signaling molecules (galectin-1; prohibitin). Only 65 of the proteins were also found in one of three secreted-protein databases [54–56]. Although this could suggest that some of the proteins are contaminants, it could also point to the possibility that many secreted proteins are missed by conventional discovery methods that typically screen for proteins with an N-terminal signal sequence [44,57]. Some well-known

**TABLE 1 Five Proteins Found in the Secretory Apparatus of Human Visceral Adipose Tissue That Are Targets for Therapies in Clinical Trials**

| Target | Condition | Drug | MoA | Phase |
|---|---|---|---|---|
| Leucocyte elastase | Acute respiratory distress syndrome | Depelestat | Inhibitor of elastase activity | II |
| Transforming growth factor-b-1 | Pulmonary fibrosis | GC1008 | Neutralizing antibody | I |
| Clusterin | Lung and prostate cancers | OGX-011 | Inhibition of gene expression | II |
| Transthyretin (TTR) | Familial amyloidosis | Diflunisal | Stabilization of TTR to prevent formation of amyloid fibrils | III |
| Pigment epithelium-derived protein | Macular degeneration | AdGVPEDF.11D | Gene therapy | I |

secreted proteins, including FGF-1, FGF-2, IL-1, and galectins, do not have a classical N-terminal signal sequence [58]. Direct detection of these nonclassical secreted proteins will be an important approach to expanding on the secretome as well as validating the predictive tools currently under development for these proteins, such as SecretomeP [59].

An important aspect of studying the contents of the secretory apparatus is the ability to identify low-abundance secreted proteins that would be difficult to detect without specific antibodies, such as IL-25, TGFb1, and lipocalin 2. Also important is that the proteins identified have important biological functions that could eventually aid in the discovery of novel approaches to therapeutics. Of interest, therefore, is the observation that five of the proteins identified in the pilot adipose tissue study are current targets of therapies in clinical trials (Table 1). These therapies are not necessarily involved in adipose-related diseases, but still serve to demonstrate that proteins identified by this approach include those that are important to drug discovery.

## SUMMARY

The global proteomics approach for candidate biomarker discovery using label-free mass spectrometry-based methods is able to detect significantly differentially expressed proteins in a variety of biological systems. An advantage of this technology is that it enables varied study designs involving comparisons between large cohorts or multiple experimental conditions. For example, pharmacodynamic markers of drug treatment, markers of disease severity, or predictive markers of efficacy may be discovered, depending on the study design. Candidate biomarkers are combined into panels having sufficient discriminating power (positive and negative predictive values) to provide meaningful diagnostic or prognostic information and perform better than single-protein biomarkers against broad populations. Consequently, these high-confidence panels provide biomarkers that appear suitable for development into high-throughput screening assays.

Multiple approaches to identifying lower-abundance proteins in plasma have demonstrated value, including antibody depletion of high-abundance proteins as well as the analysis of the contents of the secretory apparatus. These two techniques are seen as complementary in terms of the plasma protein concentration range interrogated. Typically, the limit of detection of direct analysis of plasma or serum proteins is approximately 10 to 100 ng/mL, which has proven to be a very productive range for candidate biomarker identification. The blood concentrations of proteins secreted from Golgi lumen are approximately two orders of magnitude below this, allowing for an even more comprehensive view of the low-abundance tissue-specific blood proteome. Further improvements are expected in the future, but clearly the tools now exist to identify novel and useful biomarkers of disease and drug response.

## REFERENCES


1. Marko-Varga G, Lindberg H, Löfdahl CG, et al. (2005). Discovery of biomarker candidates within disease by protein profiling: principles and concepts. *J Proteome Res*, 4:1200–1212.
2. Engwegen JY, Gast MC, Schellens JH, Beijnen JH (2006). Clinical proteomics: searching for better tumour markers with SELDI-TOF mass spectrometry. *Trends Pharmacol Sci*, 27:251–259.
3. Julka S, Regnier FJ (2005). Recent advancements in differential proteomics based on stable isotope coding. *Brief Funct Genom Proteom*, 4:158–177.
4. Silva JC, Denny R, Dorschel CA, et al. (2005). Quantitative proteomic analysis by accurate mass retention time pairs. *Anal Chem*, 77:2187–2200.
5. Wang W, Zhou H, Lin H, et al. (2003). Quantification of proteins and metabolites by mass spectrometry without isotopic labeling or spiked standards. *Anal Chem*, 75:4818–4826.
6. Wiener MC, Sachs JR, Deyanova EG, Yates NA (2004). Differential mass spectrometry: a label-free LC-MS method for finding significant differences in complex peptide and protein mixtures. *Anal Chem*, 76:6085–6096.
7. Wang G, Wu WW, Zeng W, Chou CL, Shen RF (2006). Label-free protein quantification using LC-coupled ion trap or FT mass spectrometry: reproducibility, linearity, and application with complex proteomes. *J Proteome Res*, 5:1214–1223.
8. DeSouza LV, Grigull J, Ghanny S, et al. (2007). Endometrial carcinoma biomarker discovery and verification using differentially tagged clinical samples with multidimensional liquid chromatography and tandem mass spectrometry. *Mol Cell Proteom*, 6:1170–1182.
9. Wu WW, Wang G, Baek SJ, Shen RF (2006). Comparative study of three proteomic quantitative methods, DIGE, cICAT, and iTRAQ, using 2D gel or LC-MALDI TOF/TOF. *J Proteome Res*, 5:651–658.
10. Kislinger K, Gramolini AO, MacLennan DH, Emili A (2005). Multidimensional protein identification technology (MudPIT): technical overview of a profiling method optimized for the comprehensive proteomic investigation of normal and diseased heart tissue. *J Am Soc Mass Spectrom*, 16:1207–1220.
11. Calvo KR, Liotta LA, Petricoin EF (2005). Clinical proteomics: from biomarker discovery and cell signaling profiles to individualized personal therapy. *Biosci Rep*, 25:107–125.
12. Gao J, Friedrichs MS, Dongre AR, Opiteck GJ (2005). Guidelines for the routine application of the peptide hits technique. *J Am Soc Mass Spectrom*, 16:1231–1238.
13. Old WM, Meyer-Arendt K, Aveline-Wolf L, et al. (2005). Comparison of label-free methods for quantifying human proteins by shotgun proteomics. *Mol Cell Proteom*, 4:1487–1502.
14. Lu P, Vogel C, Wang R, Yao X, Marcotte EM (2007). Absolute protein expression profiling estimates the relative contributions of transcriptional and translational regulation. *Nat Biotechnol*, 25:117–124.
15. Patil ST, Higgs RE, Brandt JE, et al. (2007). Identifying pharmacodynamic protein markers of centrally active drugs in humans: a pilot study in a novel clinical model. *J Proteome Res*, 6:955–966.

16. Roy SM, Becker C (2007). Quantification of proteins and metabolites by mass spectrometry without isotopic labeling. *Methods Mol Biol*, 359:87–105.
17. Finney GL, Blackler AR, Hoopmann MR, Canterbury AD, Wu CC, MacCoss MJ (2008). Label-free comparative analysis of proteomics mixtures using chromatographic alignment of high-resolution LC-MS data. *Anal Chem*, 80(4):961–971.
18. Prakash A, Piening B, Whiteaker J, et al. (2007). Assessing bias in experiment design for large scale mass spectrometry-based quantitative proteomics. *Mol Cell Proteom*, 6:1741–1748.
19. Fang R, Elias DA, Monroe ME, et al. (2006). *Mol Cell Proteom*, 5:714–725.
20. Gaspari M, Verhoeckx KCM, Verheij ER, van der Greef J (2006). Integration of two-dimensional LC-MS with multivariate statistics for comparative analysis of proteomic samples. *Anal Chem*, 78:2286–2296.
21. Follettie MT, Pinard M, Keith JC Jr, et al. (2006). Organ messenger ribonucleic acid and plasma proteome changes in the adjuvant-induced arthritis model: responses to disease induction and therapy with the estrogen receptor-b selective agonist ERB-041. *Endocrinology*, 147:714–723.
22. Chelius D, Bondarenko P (2002). Quantitative profiling of proteins in complex mixtures using liquid chromatography and mass spectrometry. *J Proteome Res*, 1:317–323.
23. Lamontagne J, Butler H, Chaves-Olarte E, et al. (2007). Extensive cell envelope modulation is associated with virulence in *Brucella abortus*. *J Proteome Res*, 6:1519–1529.
24. Sonnen JA, Keene CD, Montine KS, et al. (2007). Biomarkers for Alzheimer's disease. *Expert Rev Neurother*, 7:1021.
25. Maurya P, Meleady P, Dowling P, Clynes M (2007). Proteomic approaches for serum biomarker discovery in cancer. *Anticancer Res*, 27:1247.
26. Fu Q, Van Eyk JE (2006). Proteomics and heart disease: identifying biomarkers of clinical utility. *Expert Rev Proteom*, 3:237.
27. Burtis CA, Aswood EA, Burns DE (2005). *Tietz Textbook of Clinical Chemistry*. Elsevier Saunders, Philadelphia.
28. Anderson NL, Anderson NG (2002). The human plasma proteome: history, character, and diagnostic prospects. *Mol Cell Proteom*, 1:845.
29. Thadikkaran L, Siegenthaler MA, Crettaz D, Queloz PA, Schneider P, Tissot JD (2005). Recent advances in blood-related proteomics. *Proteomics*, 5:3019–3034.
30. Hendriks MM, Smit S, Akkermans WL, et al. (2007). How to distinguish healthy from diseased? Classification strategy for mass spectrometry-based clinical proteomics. *Proteomics*, 7:3672–3680.
31. Sitnikov D, Chan D, Thibaudeau E, Pinard M, Hunter JM (2006). Protein depletion from blood plasma using a volatile buffer. *J Chromatogn B*, 832:41–46.
32. Bjorhall K, Miliotis T, Davidsson P (2005). Comparison of different depletion strategies for improved resolution in proteomic analysis of human serum samples. *Proteomics*, 5:307–317.
33. Whiteaker J, Zhang H, Eng JK, et al. (2007). Head-to-head comparison of serum fractionation techniques. *J Proteome Res*, 6:828–836.
34. Sitnikov D, Hunter JM, Hayward C, et al. (2007). Peptide shifter: enhancing separation reproducibility using correlated expression profiles. *J Am Soc Mass Spectrom*, 18:1638–1645.

35. Peng J, Elias JE, Thoreen CC, Licklider LJ, Gygi SP (2003). Evaluation of multidimensional chromatography coupled with tandem mass spectrometry (LC/LCMS/MS) for large-scale protein analysis: The yeast proteome. *J Proteome Res*, 2:43–50.
36. Kearney P, Butler H, Eng K, Hugo P (2008). Harmonizing protein identification with protein expression data. *J Proteome Res*, 7:234–244.
37. Lekpor K, Benoit MJ, Butler H, et al. (2007). An evaluation of multidimensional fingerprinting in the context of clinical proteomics. *Proteom Clin Appl*, 1:457–466.
38. Rifai N, Gillette MA, Carr SA (2006). Protein biomarker discovery and validation: the long and uncertain path to clinical utility. *Nat Biotechnol*, 24:971.
39. Barnidge DR, Goodmanson MK, Klee GC, Muddiman DC (2004). Absolute quantification of the model biomarker prostate-specific antigen in serum by LC-MS/MS using protein cleavage and isotope dilution mass spectrometry. *J Proteome Res*, 3:644–652.
40. Bondar OP, Barnidge DR, Klee EW, Davis BJ, Klee GG (2007). LC-MS/MS quantification of Zn-α2 glycoprotein: a potential serum biomarker for prostate cancer. *Clin Chem*, 53:673–678.
41. Folstein MF, Folstein SE, McHugh PR (1975). Mini-Mental State: A practical method for grading the state of patients for the clinician. *J Psych Res*, 12:189–198.
42. Kratchmarova I, Kalume DE, Blagoev B, et al. (2002). A proteomic approach for identification of secreted proteins during the differentiation of 3T3-L1 preadipocytes to adipocytes. *Mol Cell Proteom*, 1:213–222.
43. Chen X, Cushman SW, Pannell LK, Hess S (2004). Quantitative proteomic analysis of the secretory proteins from rat adipose cells using 2D liquid chromatography-MS/MS approach. *J Proteome Res*, 4:570–577.
44. Alvarez-Llamas G, Szalowska E, de Vries MP, et al. (2007). Characterization of the human visceral adipose tissue secretome. *Mol Cell Proteom*, 6:589–600.
45. Lee MC, Miller EA, Goldberg J, Orci L, Schekman R (2004). Bi-directional protein transport between the ER and Golgi. *Annu Rev Cell Dev Biol*, 20: 87–123.
46. Lanoix J, Paramithiotis E (2008). Secretory vesicle analysis for discovery of low abundance plasma biomarkers. *Expert Opin Med Diagn*, 2(5):475–485.
47. Sircar K, Gaboury L, Ouadi L, et al. (2006). Isolation of human prostatic epithelial plasma membranes for proteomics using mirror image tissue banking of radical prostatectomy specimens. *Clin Cancer Res*, 12:4178–4184.
48. Yang R, Barouch LA (2007). Leptin signaling and obesity: cardiovascular consequences. *Circ Res*, 101:545–559.
49. Malendowicz LK, Rucinski M, Belloni AS, Ziolkowska A, Nussdorfer GG (2007). Leptin and the regulation of the hypothalamic–pituitary–adrenal axis. *Int Rev Cytol*, 263:63–102.
50. Louis GW, Myers MG (2007). The role of leptin in the regulation of neuroendocrine function and CNS development. *Rev Endocr Metab Disord*, 8:85–94.
51. Lago F, Dieguez C, Gomez-Reino J, Gualillo O (2007). Adipokines as emerging mediators of immune response and inflammation. *Nat Clin Pract Rheumatol*, 3:716–724.

52. Iannucci CV, Capoccia D, Calabria M, Leonetti F (2007). Metabolic syndrome and adipose tissue: new clinical aspects and therapeutic targets. *Curr Pharm Des*, 13:2148.
53. Vohl MC, Sladek R, Robitaille J, et al. (2004). A survey of genes differentially expressed in subcutaneous and visceral adipose tissue in men. *Obes Res*, 12:1217–1222.
54. Secreted Protein Database. http://spd.cbi.pku.edu.cn/.
55. Polanski M, Anderson NL (2006). A list of candidate cancer biomarkers for targeted proteomics. *Biomarker Insights*, 2:1–48.
56. States DJ, Omenn GS, Blackwell TW, et al. (2006). Challenges in deriving high-confidence protein identifications from data gathered by a HUPO plasma proteome collaborative study. *Nat Biotechnol*, 24:333–338.
57. Arnoys EJ, Wang JL (2007). Dual localization: proteins in extracellular and intracellular compartments. *Acta Histochem*, 109:89–110.
58. Walter N (2005). Unconventional secretory routes: direct protein export across the plasma membrane of mammalian cells. *Traffic*, 6:607–614.
59. Bendtsen JD, Jensen LJ, Blom N, von Heijne G, Brunak S (2004). Feature-based prediction of non-classical and leaderless protein secretion. *Protein Eng Des Sel*, 17(4):349–356.

# QUANTITATIVE MULTIPLEXED PATTERNING OF IMMUNE-RELATED BIOMARKERS

Dominic Eisinger, Ph.D., Ralph McDade, Ph.D., and Thomas Joos, Ph.D.
*Rules Based Medicine, Inc., Austin, Texas*

## INTRODUCTION

The human immune system is a complex cellular and molecular network unlike any other biological system. The complex interplay of multiple immune cells that rid the body of foreign invaders and diseased tissues is governed by a vast number of cell-to-cell communications. In this chapter we focus on the expanding utility of measuring these secreted protein messengers as biomarkers for drug development and diagnostic efforts that are relevant to a variety of diseases. Herein we use a broad acronym, IB (immune-related biomarker), to represent protein biomarkers for a variety of immune-related disorders. The IBs most prominently monitored in drug development efforts are the cytokines and chemokines, but also include acute-phase reactants, tissue remodeling factors, vascular markers, and growth factors. This complexity has led to the rapid adoption of newer multiplexed measurement tools. Key features of these include the quantitation, sensitivity, and perhaps most important, precision of each measurement. Ecosanoids, prostaglandins, prostacyclins, the thromboxanes, and the leukotrienes are also key elements of immune processes but do not fit under the paradigm of protein biomarkers and are not addressed in this chapter.

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

This ability to measure multiple IBs simultaneously from a small biological sample has ushered in an era of exploration known as *biomarker patterning*. It would be convenient if the single biomarker paradigm of the past were to continue unabated. However, the rate of introduction of new U.S. Food and Drug Administration (FDA)–approved diagnostic protein assays has fallen dramatically to an average of one or less per year [1]. This unfortunate trend does not suggest relying on a single biomarker but, rather, a group or set of biomarkers for drug development and diagnostic efforts. The theory of biomarker stacking is that by combining the predictive power of each marker, the optimal sensitivity and specificity of the multiplexed test will be achieved [2]. Rather than a single biomarker to fit with a specific disease or drug effect, a pattern of multiple biomarkers provides stronger predictive value, due to their "stacking" effects. In this chapter we focus on the multiplexed immunoassay measurement technologies as well as some recent and relevant examples of successful biomarker patterning. These approaches are rapidly bringing multiplexed biomarker assays into everyday use in drug development and diagnostics for personalized medicine.

## MEASUREMENT PROTEOMICS VS. TRADITIONAL PROTEOMICS

Multiplexed immunoassays used to identify novel biomarker patterns is a form of proteomics. However, it is often confused with traditional proteomics, which typically use complex separation technologies such as two-dimensional gel electrophoresis and/or mass spectrometry to identify patterns of peaks that are peptides or proteins of unknown origin. This approach provides new targets for future research but lacks the quantitation and certainly the precision necessary for clinical use. In contrast, the approach discussed here measures only those analytes where a specific assay has been developed. It requires specific ligand-binding reagents, generally antibodies, and a physical address to localize that binding reaction. This technique has been described as *measurement proteomics*, where a restricted set of analytes is measured and data reported in absolute terms (e.g., pg/mL) [3].

One caveat of measurement proteomics that must be addressed is the inherent variability among individuals for expression of different IBs. This variation results from a blend of genetic and environmental factors. Before one can accurately predict the biological response to a disease or its therapy, the natural variance within a population as well as within a person must be determined for the protein analytes in question. This is typically performed by assessing the biomarker patterning data for a control group of people while simultaneously measuring the identical markers in the diseased or drug-treated group. Statistical tools and means are widely available to use either cutoff values for individual analytes or multivariate analysis for multiple biomarkers in order to achieve a valuable test. Experience has shown that the level of a given protein biomarker or biomarker pattern often correlates with a particu-

lar disease state. In addition, these biomarker levels in serum or plasma have been shown to be stable in a person over time [1,2].

## MEASUREMENT PROTEOMICS PLATFORM TECHNOLOGIES

### ELISA-Based Immunometric (Sandwich) Immunoassays

Many of the key IBs (cytokines and chemokines in particular) exist at relatively low concentrations in serum or plasma, requiring the femtomolar sensitivity provided by the sandwich immunoassay. Sandwich immunoassays get their name from the fact that the antigen is "sandwiched" in between two antibodies. This is typically a more sensitive technique than single-antibody competitive assays. Traditional single-plex methods such as the 96-well microtiter plate–based enzyme-linked immunoassay (ELISA) use a capture antibody anchored either hydrophobically or covalently to the polystyrene wall of each well. After binding of antigen from the liquid sample, the assay is completed by adding a second antibody solution specific for the same target, thus completing the sandwich. A reporter signal attached to the second antibody provides the quantitative signal. ELISAs have been the standard in the research lab for decades, while the same sandwich immunoassay technique has been applied in a more industrial fashion in fully automated random access immunoanalyzers such as the Immulite from Siemens, built for FDA-approved assays in the clinical lab.

### Multiplexed Immunometric (Sandwich) Immunoassays

The random access immunoanalyzers found in the clinical lab do provide a form of multiplexing by physically splitting the loaded serum or plasma sample into multiple single-plex assays. Although this is an efficient approach for the work flow in a lab, it still requires a minimum of sample volume for each assay and does not reduce the cost per assay significantly. True multiplexing, as delivered by today's newer platforms, does provide significant savings in both the sample volume requirements and the cost per assay. Although a number of interesting multiplexed immunoassay platforms have been launched in the last decade, two primary types of platforms have survived as viable commercial entities: planar arrays and microsphere-based arrays.

### Planar Arrays

Planar microarrays are highly miniaturized and parallelized solid-phase assay systems that use a large number of different capture molecules immobilized in physically addressable locations called *microspots*. These spots, which typically have diameters of less than 250 μm and are arrayed in rows and columns,

are where the capture antibody for each assay is deposited. This deposition by physical touching with pins or spraying with piezoelectric tips is a technologically demanding step that was first pioneered in the gene expression industry for detection of nucleic acids by sequence-specific hybridization [4]. This approach has been used to generate antibody arrays consisting of several hundred capture antibodies by depositing them on nitrocellulose-coated microscope slides. Typically, these large arrays have then been probed either by using a direct labeling approach, where the mixture of antigens is fluorescently labeled; or conversely, a cocktail of second antibodies labeled with a fluorescent tag have been used to complete the sandwich assays. In either case, quantitation was achieved by imaging the spots with a change-coupled device (CCD) camera. The two limiting features of this approach have been (1) the irreproducibility of the spotting approach, where each new array is considered a new lot of material from a quality control point of view; and (2) the high probability of antibody–antibody interactions when greater than 25 antigen–antibody reactions are performed in a shared physical space. The second is a limiting factor for all multiplexed immunoassays and has been considered in the three most popular platforms.

### Microtiter-Based Planar Arrays

Glass slide–based immunoassay microarrays have been adapted to the microtiter plate format to increase throughput and to use existing automated solutions. Although this limits the number of microspots available per well, it helps to avoid the problem of antibody–antibody interactions. Such an approach was adapted by Thermo Scientific with their SearchLight Protein Array Technology. On the bottom of each microtiter well, an array consisting of up to 16 microspot features are generated, which allows the measurement of 16 analytes from 50 μL of sample. After incubation with sample, the wells are washed as with an ELISA and the captured antigen visualized using a secondary antibody labeled with a chemiluminescent probe. Quantitation of captured analytes is provided by a CCD imaging system using a calibration curve set up in parallel wells (www.piercenet.com).

Meso Scale Discovery (http://www.meso-scale.com) also uses a microtiter well microarray strategy called Multi-Array. Each well is arranged with 1 to 25 carbon microspots that are in essence miniature electrodes integrated into the bottom of the plate. Onto each carbon electrode the capture antibodies are immobilized, which after multiplexed immunoassay development results in a bound electrochemiluminescent (ECL) label. Quantitation of each assay is performed by passing electrical current through each electrode, producing a luminescent signal which is detected by a CCD camera imaging system. The microtiter plate format allows the use of standard liquid-handling automation with a multiplexed control and calibration strategy. As each Multi-Array plate can be imaged across all 96 wells within a few seconds, the sample throughput of this approach is large and quite scalable. The ECL detection system is also

thought to provide significantly better sensitivity for immunoassays than can be provided by chemiluminescence or fluorescence.

The two planar array systems described above are typically operated in a manual mode, although both are capable of interfacing with automated liquid-handling systems. One fully automated planar array system in the marketplace is the Evidence, an automated biochip system developed by Randox Laboratories (http://www.randox.com). It provides multiplexed analysis of miniaturized and parallelized immunoassays in a macroarray format. This macroarray contains 25 features and uses a chemiluminescence-based readout. Several immunoassays panels are available, including fertility, cardiac disease, tumors, cytokines and growth factors, and drug residue panels.

## Microsphere Arrays

Robust and flexible microsphere-based array systems have been developed over the last decade [5]. For the purposes of this review, there is only one microsphere-based array developed and commercialized in a manner parallel to the planar array discussed previously that has succeeded in garnering the most significant portion of the multiplexed immunoassay market: xMAP from Luminex Corporation (http://www.luminexcorp.com). Luminex technology can perform up to 100 multiplexed, microsphere-based assays in a single reaction vessel by combining optical classification schemes, biochemical assays, flow cytometry, and advanced digital signal-processing hardware and software.

Multiplexing is accomplished by assigning each analyte-specific assay a microsphere set labeled with a unique fluorescence signature. To attain 100 distinct microsphere signatures, two fluorescent dyes, red and infrared, are mixed in various combinations using 10 intensity levels of each dye (i.e., $10 \times 10$). Each batch or set of microspheres is encoded with a fluorescent signature by impregnating the microspheres with one of these dye combinations. After the encoding process, assay-specific capture antibodies are conjugated covalently to each unique set of microspheres. Coupling is performed on large numbers of individual microspheres ($10^7$ to $10^9$ microspheres) simultaneously within each unique set, resulting in low microsphere-to-microsphere variability.

After optimizing the parameters of each assay separately, multianalyte profiles (MAP) are performed by mixing up to 100 different sets of the microspheres in a single well of a 96-well microtiter plate. A few microliters of sample is added to the well and allowed to react with the microspheres. The assay-specific capture reagent on each microsphere binds the analyte of interest. A cocktail of assay-specific, biotinylated antibodies is reacted with the microsphere mixture, followed by a streptavidin-labeled fluorescent "reporter" molecule (typically, phycoerythrin). Finally, the multiplex is washed to remove unbound detecting reagents.

After washing, the mixture of microspheres is analyzed using a Luminex instrument which uses hydrodynamic focusing to pass the microspheres in

single file through two lasers. As each individual microsphere passes through the excitation beams, it is analyzed for size, encoded fluorescence signature, and the amount of fluorescence generated in proportion to the analyte. Microsphere size, determined by measuring the 90° light scatter as the microspheres pass through a red diode laser (633 nm), is used to eliminate microsphere aggregates from the analysis. While in the red excitation beam, the encoded red and far-red dyes are excited and the resulting fluorescence signature (ratio 660 nm/720 nm) is filtered, measured using avalanche photodiodes, and classified to a microsphere set. Since each microsphere is encoded with a unique signature, the classification identifies the analyte being measured on that individual microsphere. As the microsphere passes through a green diode-pumped solid-state laser (532 nm), the fluorescence "reporter" signal (580 nm) is generated in proportion to the analyte concentration, filtered, and measured using a photomultiplier tube. Data acquisition, analysis, and reporting are performed in real time on all microsphere sets included in the MAP. A minimum of 50 individual microspheres from each unique set are analyzed and the median value of the analyte-specific, or reporter, fluorescence is logged. Using calibrators and controls of known analyte quantity, sensitive and quantitative results are achieved with precision enhanced by the redundant oversampling at each data point.

xMAP provided several key advantages over the two-dimensional planar arrays. First, consistent coating of the polystyrene microsphere surface with antibodies has been performed in the diagnostic industry for decades, so it is a reproducible technique. Microsphere coating does not suffer from the irreproducibility that plagues planar arrays. Second, the lot size for manufacturing of a microsphere array is generally in the thousands to millions of tests per lot. Lower lot-to-lot variability coupled with extremely large manufacturing lots means that the problem of lot variation is greatly diminished. The primary disadvantage of microsphere arrays is throughput, as each sample must be read in turn in the flow system. So far, no group has solved this problem be developing multiple-channel flow systems, but this is the next logical step of the technology.

## APPLICATIONS AND EXAMPLES IN DRUG DEVELOPMENT AND DIAGNOSTICS

Whether one uses planar arrays or microsphere arrays, the resulting quantitative, multiplexed data are then mined for distinctive patterns of biomarkers that indicate disease, drug toxicity or efficacy, or some other useful biochemical phenotype. We will not discuss the data mining techniques available but simply describe several examples where data mining of these complex data sets has revealed meaningful biomarker patterns. Typically, data mining is first done using traditional statistics to identify those analytes whose values are significantly different between two experimental groups. This is often done by

determining mean analyte values and applying Student's *t*-tests to determine *p*-values of significance. The next step is often to use more complex multivariate analysis tools, such as principal component analysis, for pattern recognition.

The reports of IB monitoring cited most often are related to inflammatory disease. Asthma, rheumatoid arthritis (RA), multiple sclerosis (MS), systemic lupus erythematosus (SLE), chronic obstructive pulmonary disease (COPD), and psoriasis (PS) are thought of as classic inflammatory diseases. In recent years, a host of other diseases have been found to have a significant inflammatory component. These include cardiovascular disease (CVD), central nervous system diseases such as Alzheimer disease, most infectious diseases, and various cancers [6]. IBs are thus very useful for quantifying, characterizing, and monitoring the inflammatory and anti-inflammatory components of a disease in addition to therapeutic benefits during drug trials.

Measurement of inflammatory markers is a popular application of multiplexed profiling in drug development and diagnostic efforts. The duration and pattern of cytokine and chemokine expression in serum or plasma have been the most common uses of multiplexed immunoassays. Common inflammatory mediators are often cited (e.g., TNFα, IL-1, IL-6, IL-8, MIP-1) due to the large set of publications describing the association of these classical inflammatory markers with numerous diseases, conditions, and drug treatments. Although important, other IBs that are related directly or indirectly to the disease and/or therapy in question will often add significant prognostic value to biomarker patterning efforts.

## Inflammatory Bowel Disease in a Mouse Model

Crohn disease and ulcerative colitis are chronic inflammatory diseases of the gut that cause tremendous human suffering and loss of productivity. There is a clear need for better diagnostic tools that are less invasive or costly than endoscopy and histology. Torrence et al. published a study [7] of an inflammatory bowel disease model in mouse where they identified a serum-based IB pattern using data generated from a microsphere array. These data were generated using serum samples of mice infected with *Helicobacter bilis*, which causes an inflammation in the gut. These IB patterns were then compared with those of control animals. Samples were sent to Rules-Based Medicine (RBM), Inc., Austin, Texas for biomarker analysis using their RodentMAP (http://www.rbmmaps.com). RBM has an automated testing laboratory that utilizes xMAP technology under regulatory compliance to both good laboratory practices (GLPs) and the Clinical Laboratory Improvement Act (CLIA). RBM performed a multianalyte profile consisting of 58 different quantitative immunoassays and reported the data for each analyte in mass per milliliter of sample.

Results indicated that serum levels of IL-11, IL-17, interferon gamma-inducible protein (IP-10), lymphotactin, monocyte chemoattractant protein

(MCP-1), and vascular cell adhesion molecule (VCAM-1) were elevated in early disease. In later, more severe disease, IL-11, IP-10, haptoglobin, matrix metalloproteinase-9 (MMP-9), macrophage inhibitory protein 1 alpha (MIP 1α), fibrinogen, immunoglobulin A, apolipoprotein A1, and IL-18 were elevated. All of these biomarkers are considered IBs with the exception of apolipoprotein A1, which is considered a metabolic biomarker. Interestingly, all of the IBs correlated with histopathological scores. Antibiotic treatment of the infected mice both improved the histopathology scores and decreased the mean serum values for most of the IBs. These data suggest that the serum IB patterns could be useful for both diagnosis and prognosis of similar human diseases.

## Scleroderma

Scleroderma is an autoimmune disease that manifests itself in the skin, resulting in fibrosis and damage to the vasculature. It is thought that IBs play a major role in the course of the disease and therefore should be good targets for both therapeutic intervention and diagnostic/prognostic development. Recently, Duan et al. [8] have compared the serum IB patterns of scleroderma patients with controls using the RBM testing approach, measuring 188 different biomarkers. In addition, this group compared the serum protein expression patterns with the messenger RNA expression from the monocytes and lymphocytes isolated from these patients' peripheral blood.

As in the previous mouse study, the majority of biomarkers that were differentially expressed were IBs. Acute-phase reactants (fibrinogen, haptoglobin, and von Willebrand factor), cytokines (IL-1α, IL-16, soluble tumor necrosis factor receptor 2, and MIP-1α), tissue remodeling proteins (intercellular adhesion molecule type 1 and tissue inhibitor of metalloproteinase type 1), and one metabolic biomarker, apolipoprotein CIII, were all elevated in the scleroderma patient sera compared with normal age-matched controls. Comparison of the gene expression patterns from the various immune cell populations correlated poorly with the serum IB profiling data. This was not surprising, as these two profiles often do not correlate well. The reason(s) for the gene expression and IB profiling differences are unknown.

Results indicated that the pattern of IB elevation in scleroderma could be used as a more specific diagnostic to help the rheumatologist differentiate scleroderma from other autoimmune conditions or skin maladies. More important, the intensity of IB elevation could be used to monitor therapeutic intervention in clinical trials as well as after drugs developed for this indication are approved.

## Rheumatoid Arthritis

Rheumatoid arthritis (RA) is an autoimmune disease where the current focus of drug development efforts is devoted to specific blockade of certain cytokine

activities. It is therefore highly likely that IB monitoring will play an important role in RA disease prognosis and drug development programs. A brief review of the cytokine blockade strategies for RA starts with the inflammatory cytokine TNFα, a major therapeutic advancement in the treatment of RA [9]. There are now several TNFα biological drugs on the market (Humira, Abbott; Enbrel, Amgen; Remicade, Centicor) that in addition to treatment of RA, are also being used off-label for psoriasis and for Crohn disease.

Blockade of IL-1 is another option available for RA therapy. Anakinra (Kineret, Amgen), a recombinant interleukin-1 receptor antagonist, is used for treatment of rheumatoid arthritis. The next major advancement in anticytokine therapy for RA appears to be the humanized anti-IL-6 receptor blocking antibody tocilizumab (Actemra, Roche). The acute-phase reactant serum amyloid A was identified in a serum profile that differentiated response from no response to Actemra therapy [10]. Intervention of IL-15 also looks promising for treatment of RA and psoriatic arthritis in addition to pulmonary inflammatory diseases [11]. Targeted cytokine neutralization as a therapeutic option in autoimmune diseases such as RA will probably lead to novel IB patterns with both prognostic and diagnostic value during drug efficacy trials. For example, in a recent study of Sjögren syndrome, a panel of 25 IBs provided diagnosis and disease management in this debilitating condition [12].

## Cardiovascular Disease

Over the past 25 years, the number of risk factors for coronary artery disease has increased dramatically. Systemic inflammation and abnormal lipoprotein metabolism are important contributors to the progression of atherosclerotic disease leading to plaque instability [13]. Even so, the vast majority of patients who experience an acute coronary event have no prior symptoms. Further complicating the diagnosis of acute coronary syndrome (ACS) is the frequent occurrence of patients who present symptoms of chest pain that can be attributable to other completely nonrelated events, such as acute gastroesophageal reflux disease. Given the serious life-threatening nature of ACS, an improvement in the early diagnosis of heart attacks would be a major medical advancement.

Matrix metalloproteinases (MMPs) are promising IBs for early identification of high-risk patients. Pro-inflammatory cytokines stimulate the secretion of MMPs, and an increase in systemic MMPs has been documented in ACS and correlated to unstable plaque metabolism [13]. In a recent study by Gurbel and co-workers, MMP-2, MMP-3, and MMP-9 were sensitive and specific markers differentiating asymptomatic coronary artery disease (A-CAD) from symptomatic CAD (S-CAD). The stable quiescent disease group of A-CAD patients were documented by angiography, a history of coronary artery bypass grafting, or revascularization (stenting or balloon angioplasty). The progressive S-CAD group of patients were enrolled immediately before coronary intervention and presented either myocardial infarction or stable angina.

Patients with S-CAD had markedly elevated levels of specific MMP-2 and MMP-9, whereas patients with A-CAD had significantly greater levels of MMP-3. In addition to the MMPs, other IBs (TIMP-1, C-reactive protein IL-8, IL-10, RANTES, endothelin, plasminogen activator inhibitor type 1, apolipoprotein C-III, and IL-1a) reinforced the MMP results for distinguishing the two groups [13].

### Hepatitis C

Wright and co-workers documented cytokine profiles of patients grouped based on their viral titers before and after standard therapy [14]. They found that they were able to demonstrate the prognostic power of serum cytokine profiling in chronic hepatitis C virus (HCV) infection. The study demonstrated that overall serum cytokine levels were significantly higher in patients than in controls and that the levels dropped significantly after therapy in concert with their viral titer. Distinct cohort subgroups based on changes in viral titers correlated to specific sets of cytokines that decreased in each group. This could allow for new efforts to stratify patients and increase efforts to find improved therapies.

### Ovarian Cancer

Ovarian cancer has been called the silent killer, as it is often not diagnosed until late in the disease progression, when therapeutic intervention is rarely successful. The unmet medical need for a sensitive and specific diagnostic test that will detect early-stage ovarian cancer and differentiate it from benign ovarian disease is strong. Bertenshaw et al. recently reported the identification of a biomarker pattern of eight serum analytes that using a multivariate algorithm could differentiate stage 2 ovarian cancer patients from benign ovarian disease [15]. In addition to *Cancer Antigen (CA)-125*, seven IBs were identified (C-reactive protein, epidermal growth factor, IL-10, IL-8, connective tissue growth factor, haptoglobin, and MMP-1) as being of importance to a potential diagnostic. These IBs, as well as *CA-125*, were identified by comparing the plasma biomarker patterns of samples from ovarian cancers in various stages, common benign gynecological conditions, as well as age-matched normal controls.

## EMERGING PARADIGM: EX VIVO IMMUNE BIOMARKER MONITORING WITH WHOLE-BLOOD CELL CULTURES

A fortuitous aspect of applying measurement proteomics to the immune system is the relative ease with which many properties of the immune system can be quantified outside the human body, ex vivo. (Herein, *ex vivo* refers to

the in vitro culture and ex vivo analysis of the immune system outside the body.) Both innate and adaptive immune system responses in blood leukocytes can be measured with multiplex IB measurements. Indeed, much of our understanding of the immune system has been generated by experimentation on human leukocytes cultures ex vivo. Given the complexity of all the different types of immune cells present in blood, a reductionist approach of isolating subpopulations of leukocytes was invaluable. For example, density gradient centrifugation of blood to obtain peripheral blood mononuclear cells (PBMCs) is a common procedure to obtain cell populations enriched for lymphocytes and monocytes but depleted of granulocytes, platelets, and red blood cells [16].

The implementation of ex vivo analysis of the immune system has only recently been deployed in the clinical drug development setting with the primary applications of safety (immunotoxicity) and pharmacodynamics. An example of an immunotoxicity application would be the Tegenero incident of 2006, in which the CD28 superagonistic antibody *TGN1412* was administered to healthy phase I volunteers. Within a few hours, a single intravenous dose induced severe lymphopenia and a systemic cytokine storm, leading to multiorgan failure [17]. Clearly, it is preferable to identify the potential for such a catastrophic drug effect during an ex vivo study. An example of a pharmacocynamic application would be quantifying IBs in the culture media that correlate with the drug mechanism of action.

The utility of ex vivo analysis of the immune system for drug development programs can be divided into two broad categories. Initially, a candidate drug can be tested ex vivo for its effects on the immune system in a resting inactivated state in addition to an activated state created by the addition of an immune stimulant to mimic inflammation. Later in the drug development process, a candidate drug can be tested in vivo in a trial subject, and the subject's immune system tested ex vivo with and without immune stimulation.

Unfortunately, the historic research methodology of utilizing PBMCs for ex vivo studies is not very adaptable to the clinical setting, nor does it provide a physiologically relevant situation for drug development studies. Compared to isolated PBMC cultures, the whole-blood culture approach has a major advantage, due to the fact that all components of the blood, including cellular (e.g., granulocytes: neutrophils, basophils, eosinophils; lymphocytes, monocytes, NK-cells), pseudocellular (e.g., red blood cells, platelets), or subcellular components (e.g., enzyme inhibitors, complement system, kininogens), are present. Therefore, functional assays performed with a whole-blood culture reflect the situation of a person's whole immune repertoire much better than do PBMC-based cultures [18–20]. In a whole-blood stimulation assay, the leukocytes are within the center of a well-composed concert of regulatory components that relies on an entire series of feedback loops, which modulate in a positive or negative manner the function of the immune system. T-cells, for example, can amplify the activities of monocytes and macrophages, and platelets produce and release precursors of leukotrienes, which are taken up

and processed further to become full-fledged mediators by polymorphonuclear granulocytes [21,22].

Red blood cells, which are not present in PBMC preparations, can produce leukotrienes and modulate the action of chemokines by expressing receptors for chemokines on their surface [23]. Plasma proteins like the complement system and the kininogens form readily available precursors of peptides exhibiting strong immunoregulating activities [24]. Some of these precursors are cleaved to become active only in the presence of enzymes, released for example by activated polymorphonuclear granulocytes [25].

Using whole-blood cultures, the functionality of the human immune system can be analyzed as close as possible to the in vivo situation. The difficulty with implementing whole-blood cultures in the clinical setting is the need for a specialized cell culture laboratory and highly trained technicians, a very expensive prospect. Shipping samples to a centralized cell culture facility dramatically reduces the quality and reproducibility of results. Recently, a standardized, fully closed, whole-blood culture system has been developed for the clinical setting that does not require cell culture facilities or specialized personnel [26,27]. The TruCulture system incorporates a 3-mL blood collection tube with cell culture media to which test substances can be added. Blood is drawn from a subject directly into a collection tube that contains a medium for an immediate and standardized initiation of the whole-blood culture. After incubation in a 37 °C heating block, the cells are physically separated from the culture medium by insertion of a valve separator, and the unit is frozen until IBs in the culture medium can be analyzed. This system will greatly aid the routine implementation of ex vivo immune monitoring in clinical trials.

## CONCLUSIONS

History has shown repeatedly that the application of a new breakthrough technology is the driving force into the next phase of meaningful discovery. The development and commercialization of quantitative multiplexed immunoassay platforms is a prime example. The two primary array platforms, planar and microsphere, provide investigators with more comprehensive biomarker data for multiple analytes with the same sensitivity and precision of single-plex testing. The technology has been invaluable to basic research and drug discovery efforts and is now moving rapidly into the clinical arena of drug development, with the goal that useful biomarker patterns will allow more informed drug development decisions to be made.

Immune-related biomarker patterning has great potential to improve the success rate of drug approval in a highly cost-effective manner. The efficiency of simultaneous profiling of multiple immune-related biomarkers with very small sample volumes has revolutionized our ability to discover and validate biomarker patterns in order to stratify patients and find markers of efficacy and safety. Central to many diseases are the immune-related biomarkers dis-

cussed in this chapter. Their modulation as drug targets themselves has also become a core theme for many of the autoimmune diseases. Finally, innovative products designed for the clinical trial setting, such as the TruCulture whole-blood culture system, will help bring immune-related biomarker patterning into the everyday realm of clinical drug development and companion diagnostics.

## REFERENCES


1. Anderson NL, Anderson NG (2002). The human plasma proteome: history, character, and diagnostic prospects. *Mol Cell Proteom*, 1(11):845–867.
2. Bodovitz S, Patterson S (2003). Protein biomarker strategies. *Drug Discov World*, Fall, pp. 67–78.
3. Heuer J, Cummins D, Edmonds B (2005). Multiplex proteomic approaches to sepsis research: case studies employing new technologies. *Expert Rev Proteom*, 2(5):669–680.
4. Kricka LJ, Master SR, Joos TO, Fortina P (2006). Current perspectives in protein array technology. *Ann Clin Biochem*, 43(6):457–467.
5. Templin M, Stoll D, Bachmann J, Joos T (2004). Protein microarrays and multiplexed sandwich immunoassays: What beats the beads? *Combin Chem High Throughput Screening*, 7(3):223–229.
6. Lucas S, Rothwell NJ, Gibson RM (2006). The role of inflammation in CNS injury and disease. *Br J Pharmacol*, 147:S232–S240.
7. Torrence A, Brabb T, Bielefeldt-Ohmann H, et al. (2008). Serum biomarkers in a mouse model of bacterial-induced inflammatory bowel disease. *Inflamm Bowel Dis*, 14(4):480–490.
8. Duan H, Fleming J, Pritchard D, et al. (2008). Combined analysis of monocyte and lymphocyte messenger RNA expression with serum protein profiles in patients with scleroderma. *Arthritis Rheum*, 58(5):1465–1474.
9. Fogler WE (2008). Treating rheumatoid arthritis with DMARDs and biologics. *Drug Discov World*, Fall, pp. 15–18.
10. Miyamae T, Malehorn D, Lemster B, et al. (2005). Serum protein profile in systemic-onset juvenile idiopathic arthritis differentiates response versus nonresponse to therapy. *Arthritis Res Ther*, 7(4):R746–R755.
11. McInnes IB, Gracie JA (2004). Interleukin-15: a new cytokine target for the treatment of inflammatory diseases. *Curr Opin Pharmacol*, 4(4):392–397.
12. Szodoray P, Alex P, Brun JG, Centola M, Jonsson R (2004). Circulating cytokines in primary Sjogren's syndrome determined by a multiplex cytokine array system. *Scand J Immunol*, 59(6):592–599.
13. Gurbel P, Kreutz R, Bliden K, DiChiara J, Tantry U (2008). Biomarker analysis by fluorokine multianalyte profiling distinguishes patients requiring intervention from patients with long-term quiescent coronary artery disease: a potential approach to identify atherosclerotic disease progression. *Am Heart J*, 155(1): 56–61.

14. Wright H, Alex P, Nguyen T, et al. (2005). Multiplex cytokine profiling of initial therapeutic response in patients with chronic hepatitis C virus infection. *Dig Dis Sci*, 50(10):1793–1803.
15. Bertenshaw GP, Yip P, Seshaiah P, et al. (2008). Multianalyte profiling of serum antigens and autoimmune and infectious disease molecules to identify biomarkers dysregulated in epithelial ovarian cancer. *Cancer Epidemiol Biomarkers Prev*, 17(10):2872–2881.
16. Jackson A (1990). Basic phenotyping of lymphocytes: selection and testing of reagents and interpretation of data. *Clin Immunol Newsl*, 10:43–55.
17. Suntharalingam G, Perry MR, Ward S, et al. (2006). Cytokine storm in a phase 1 trial of the anti-CD28 monoclonal antibody TGN1412. *N Engl J Med*, 355(10):1018–1028.
18. Esteve E, Ricart W, Fernandez-Real JM (2004). Dyslipidemia and inflammation: an evolutionary conserved mechanism. *Clin Nutr*, 24(1):16–31.
19. Jensen LE, Whitehead AS (1998). Regulation of serum amyloid A protein expression during the acute-phase response. *Biochem J*, 334(3):489–503.
20. Kemper C, Atkinson JP (2007). T-cell regulation: with complements from innate immunity. *Nat Rev Immunol*, 7(1):9–18.
21. Danese S, de la Motte C, Reyes BMR, Sans M, Levine AD, Fiocchi C (2004). Cutting Edge: T cells trigger CD40-dependent platelet activation and granular RANTES release: a novel pathway for immune response amplification. *J Immunol*, 172(4):2011–2015.
22. Murdoch C, Finn A (2000). Chemokine receptors and their role in inflammation and infectious diseases. *Blood*, 95(10):3032–3043.
23. Pruenster M, Rot A (2006). Throwing light on DARC. *Biochem Soc Trans*, 34:1005–1008.
24. Bank U, Ansorge S (2001). More than destructive: neutrophil-derived serine proteases in cytokine bioactivity control. *J Leukoc Biol*, 69:97–206.
25. Ellis TN, Beaman BL (2004). Interferon-gamma activation of polymorphonuclear neutrophil function. *Immunology*, 112(1):2–12.
26. Schmolz M, Hurst TL, Bailey DM, et al. (2004). Validation of a new highly standardised, lab-independent whole-blood leukocyte function assay for clinical trials (ILCS). *Exp Gerontol*, 39(4):667–671.
27. http://www.rulesbasedmedicine.com/products-services/TruCulture.asp.

# GENE EXPRESSION PROFILES AS PRECLINICAL AND CLINICAL CANCER BIOMARKERS OF PROGNOSIS, DRUG RESPONSE, AND DRUG TOXICITY

Jason A. Sprowl, Ph.D., and Amadeo M. Parissenti, Ph.D.
*Laurentian University, Sudbury, Ontario, Canada*

## INTRODUCTION

Measurement of the expression level of specific genetic or protein biomarkers in patient serum or biopsies can be extremely valuable in the diagnosis and treatment of a variety of human neoplasms. Often, the careful and time-consuming evaluation of several biomarkers by highly trained pathologists is required to construct a definitive diagnosis and treatment regimen. The expression level of single or several biomarkers can also be useful in the prediction of patient prognosis or response to chemotherapy, but these approaches generally have only limited accuracy [1–4]. Current studies suggest that no single biomarker will suffice for the accurate prediction of patient prognosis or tumor response to chemotherapy [5–7]. It is therefore thought that the tumor or serum level of large groups of proteins or transcripts is necessary to predict patient prognosis or outcome reliably after chemotherapy. In this chapter we review the current progress that has been achieved in the use of gene profiling to predict patient prognosis and response or toxicity to specific

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

chemotherapy regimens. Moreover, the impact of experimental design and the approach to data analysis on the ability of gene profiling experiments to reliably identify large groups of genes that can serve as prognostic or predictive biomarkers is discussed. A critical assessment of the prospect for gene profiling experiments to affect the clinical management of cancer patients and drug development is also provided.

## PLATFORMS AND TOOLS FOR GENOME PROFILING EXPERIMENTS

The task of gene expression profiling is complex, and knowledge of the platforms and tools used to accomplish such a task is essential. The most popular approaches for profiling of gene expression in biological samples are DNA microarray analysis and quantitative reverse-transcription polymerase chain reaction (Q-PCR). In a typical microarray experiment (reviewed by Villeneuve and Parissenti [8]), RNA is isolated from cells, and mRNAs present within the sample are reverse-transcribed with oligo-dT primers in the presence of fluorescent or radioactive-labeled nucleotides to yield a series of labeled cDNA probes. After denaturation, the labeled cDNA probes are hybridized to known denatured PCR products or single-stranded oligonucleotides immobilized in a grid pattern on either glass slides or nylon membranes. After washing to remove unbound probe, the intensity of label associated with the various PCR products or oligonucleotides on the array can be quantified by autoradiography or by using an array scanner. The amount of labeled cDNA hybridizing to a specific gene on the microarray is a measure of its expression (level of transcription). Often, two samples can be compared for differences in gene expression by labeling the cDNA preparations from the two samples with nucleotides conjugated to dyes that fluoresce at different wavelengths (e.g., Cy3 and Cy5 dyes). In addition, cDNAs can be transcribed using three wild-type nucleotides and one aminoallyl-labeled ribonucleotide to generate labeled aminoallyl RNAs as probes [9]. While many investigators use microarrays that are prepared within their laboratories [10–13], a variety of companies now produce microarrays with widely varying numbers of oligonucleotides or PCR products immobilized to them. This creates significant variability across experiments. The oligonucleotides can also be immobilized on microarrays via different methods, including photolithography (Affymetrix) or by printing directly on glass slides. This introduces additional variability in microarray data across laboratories, resulting in significant difficulty in identifying common genetic profiles across experiments. Moreover, while standards for the performance of microarray experiments have been established (minimal information about microarray experiments (MIAME) standards [14]), not all microarray studies follow these standards.

Some genome profiling studies are also conducted using high-throughput Q-PCR. This approach is highly quantitative, providing a more reliable

measure of gene expression, but the number of genes that can be quantified is limited, since two gene-specific primers are required for each gene whose expression is measured. In addition, gene expression is measured relative to specific reference genes whose expression must also be measured. The choice of reference genes also varies significantly between studies. *Multiplexing* of reactions, and approaches in which each PCR reaction product has a distinct wavelength for fluorescence emission (e.g., the Luminex system), have helped increase the number of genes whose expression can be profiled by PCR-based approaches. PCR-based methods of genome profiling are quite distinct from those used in microarray analyses, making comparisons between Q-PCR- and microarray-based findings more difficult. Nevertheless, various PCR-based gene profiling methods have been used to identify sets of genes associated with response to chemotherapy or recurrence of disease post-chemotherapy [15–17].

## DATA ANALYSIS PITFALLS ASSOCIATED WITH GENE PROFILING EXPERIMENTS

Other difficulties in comparing data across genome profiling experiments relate to the large genetic variation that exists in humans, the design of microarray or Q-PCR experiments, and the widely varying approaches used to analyze data from such experiments. This is particularly the case for DNA microarray studies. While microarray and, in particular, PCR-based analyses are capable of measuring the expression of a large number of genes with acceptable accuracy, gene expression varies widely across patients as well as in the same person over time. In addition, the probability of identifying genes correlating with a specific phenomenon such as chemotherapy drug response is very high when tens of thousands of genes are surveyed. Unfortunately, such correlations are often not replicated in additional independent data sets, suggesting that the false discovery rate is quite high using these approaches. The use of multiple replicate arrays per sample in microarray experiments helps considerably to reduce the false discovery rate, as does the use of multiple independent data sets from highly controlled experiments.

In addition to variable gene expression among individuals, microarray procedures often yield noisy data sets with differing efficiencies of labeling and background labeling across arrays. Therefore, a variety of algorithms and statistical approaches have been developed to normalize and analyze microarray data to improve the accuracy of identifying significant changes in gene expression between groups. Statistical tests include the computation of *t*-tests for each gene and/or the use of SAM, a significance analysis for microarrays algorithm that permits one to identify significant changes in gene expression after a statistical computation of a false discovery rate based on the data set. Common additional algorithms used for microarray data analysis include weighted voting (WV) [16,18,19], supporting vector machine (SVM) [16,20,21],

and *k*-nearest-neighbor (K-NN) [16,22]. The resulting list of differentially expressed genes is commonly quite large, and each gene in this training set should then be validated by several methods, including a "leave-one-out" cross-validation. This involves the removal of one patient sample from the training set and determining if the remaining gene set is capable of classifying correctly the left-out sample. Once the training set has demonstrated accurate prediction within the samples from which it was derived, the genetic profile should be further tested within a large patient population, ideally outside of that used to create the training set. Should the genetic profile result continuously in high predictability in the classification of interest, useful sets of biomarkers have been established. However, the use of varying data analysis methods can often result in completely different sets of differentially expressed genes. Consequently, multiple data analysis techniques should be used to obtain a list of genes commonly identified across methods [23–25]. These differences in gene expression can be further validated using other independent approaches, such as Q-PCR and immunoblotting experiments.

In addition to the data analysis tools above, clustering algorithms are being used routinely to identify sets of genes whose expression is significantly and consistently different between various groups of patients (e.g., between responders and nonresponders of chemotherapy). However, it is important to note that recent findings suggest that unsupervised methods of data analysis, such as data clustering, are often not effective in identifying predictive or prognostic biomarkers. The reason for this may be that cluster analysis does not utilize information related to the original sample grouping, and it is therefore a subjective strategy for data analysis. Furthermore, the number of genes that distinguish between classifications with the use of clustering algorithms is often small, and the distances calculated in cluster analysis may not properly portray the importance of certain genes, while excluding others of great importance [26].

## GENE PROFILES AS BIOMARKERS FOR THE DIAGNOSIS AND CLASSIFICATION OF HUMAN CANCERS

Gene profiling has proven quite valuable in identifying genes whose expression can be used to diagnose specific cancers, including many that are difficult to distinguish by other methods. For example, Kohlman et al. in 2003 used gene profiling to differentiate between acute lymphoblastic leukemia (ALL) and acute myeloid leukemia (AML) [18]. Bone marrow samples were isolated from 90 patients, 25 having been diagnosed with ALL, and 65 with AML. RNA was extracted from the samples and labeled cDNAs hybridized to Affymetrix U95Av2 and U133 microarrays. The WV algorithm and the leave-one-out cross-validation approach were then used to identify a subset of genes that were expressed differentially between the two leukemias. The expression of 24 genes was then used to accurately identify patients with AML, while 19

genes were sufficient to diagnose ALL. However, upon comparing the genes identified by the two types of microarrays, it was observed that the 24 genes identified by the U133 arrays were different from those identified by U95Av2 chips. Furthermore, only five genes were observed to be in common within the 19-gene set when the two types of array tools were used. These results show the importance of using the same array platforms when comparing across experiments.

The utility of the U133 arrays was studied further in the diagnosis of leukemia using 937 patients (892 with clinically relevant leukemia subtypes and 45 nonleukemic patients) [27]. Patients were divided into 13 subgroups of leukemia type, and each subgroup was split equally into training and validation sets. Class-specific gene expression was determined using an SVM approach, with an overall classification accuracy of 95.1% when the top 100 genes for leukemia class discrimination were used. The studies cited above clearly demonstrate the utility of gene profiling for the diagnosis of cancers and suggest that this approach can be used further in the classification of cancers into various subgroups. Sørlie et al.'s study in 2001 further illustrates the latter utility by demonstrating that breast cancers could be classified into a basal epithelial-like group, an *ERBB2*-overexpressing group, a normal breastlike group, and a luminal epithelial/estrogen receptor–positive group, which could be further divided into at least two additional subgroups, each with a distinct gene expression profile. Interestingly, survival analyses using a subset of uniformly treated patients with locally advanced breast cancer revealed clear differences in outcomes among the various groups. For example, the basal-like group had a poor prognosis, and the two estrogen receptor–positive groups had clear differences in treatment outcome [28]. These findings have been well corroborated in additional experiments performed across array platforms and using RNA extracted from paraffin-embedded tissues [29–31]. Similar approaches and findings have been observed for endometrial cancer, hepatocellular carcinoma, sarcomas, neuroblastoma, and other cancers [32–35].

## GENE PROFILES AS BIOMARKERS OF PROGNOSIS IN CANCER PATIENTS

Despite the problems associated with gene profiling approaches that were noted above, there have recently been significant advances in using gene profiling to predict prognosis in cancer patients. For example, van't Veer and colleagues isolated at diagnosis 78 core samples from breast cancer patients under the age of 55 [36]. Of these patients, all were lymph node–free at the time of diagnosis. By measuring patient outcome over five years (regardless of treatment regimen), it was observed that 34 patients developed distant metastases, whereas 44 patients did not. Microarrays containing 25,000 human genes were prepared by the investigators, to which labeled cDNA samples from the patients and a pooled reference RNA sample were hybridized.

Following supervised data analysis and correct classification by a leave-one-out cross-validation approach, a classifier of 70 genes was developed in which disease outcome could be predicted. When used within the 78-patient sample training set, the genetic profile predicted patient outcome with an accuracy of 83%. In support of the classifier, the 70-gene set predicted with 89% accuracy the prognosis of an additional 19 patients who were not included within the training set.

The utility of the van't Veer genes as prognostic biomarkers has been tested further using a larger study, which included 295 patients [13]. Sixty-one patient samples (from the study in which the 70-gene classifier was developed) were included as a control. Of the patients used, 151 were lymph node–negative and 144 were lymph node–positive. Kaplan–Meier analysis was performed to determine the probability of patients remaining metastatis-free. The 70-gene classifier showed excellent accuracy at predicting prognosis in all patients ($p < 0.001$), including lymph node–negative patients (which generally have better outcomes). These findings further suggested that the acquisition of metastases in patients is already defined through gene expression in their tumors rather than being acquired late during tumorgenesis. The findings from the two studies clearly indicate that the 70-gene classifier can reliably predict prognosis in breast cancer patients and therefore may be useful in preventing overtreatment in some patients while identifying those that would probably benefit from adjuvant therapy [13]. The 70-gene set (also known as MammaPrint) has recently been approved for use in the United States by the Food and Drug Administration (FDA). The utility of the van't Veer classifier in predicting prognosis for cancers of tissue origins other than breast is currently unknown.

A recent investigation attempted to predict outcome in breast cancer patients using RNA from tumors of 162 patients [11] and from several cell lines to serve as the reference RNA (SW872, WM115, NTERA2, MCF-7, HEPG2, MOLT4, Hs578t, HL60, OVCAR3, COLO205, and RPMI 8226 cells). Both labeled reference and tumor RNAs were hybridized to arrays consisting of 10,368 unique genes. Analysis of the hybridized arrays by SAM, prediction analysis for microarrays (PAM), and the approaches described in the van't Veer et al. study [36] resulted in the identification of 49 genes that correlated with patient outcome. The gene list, when further reduced to 21 genes by including only those genes observed in all three analyses, displayed 69% accuracy by the leave-one-out cross-validation strategy. Using the gene set, the classifier appeared to predict patient outcome with 65% accuracy using the van't Veer data and 62% accuracy using data from another study [37]. Considering chance alone can yield as much as 60% accuracy; the gene set above was not validated. Furthermore, the expression of the van't Veer prognostic genes was unable to classify the patients successfully in terms of disease outcome, though over half of the genes identified in the van't Veer study were not present on the microarrays used. This underscores the difficulties in comparing across microarray experiments when different array platforms are used.

In addition, successful prediction of prognosis using this subset of the 70-gene classifier may have been increased by dividing patients according to lymph node status as was done in the van't Veer follow-up study [13].

While the van't Veer classifier shows significant promise, it has yet to be used widely in a clinical setting. An alternative gene classifier has undergone considerable validation in a clinical setting. This classifier was developed by selecting 250 candidate genes that showed promise as prognostic biomarkers, based on literature searches and previous microarray experiments [13,19,28,38]. The expression of these genes was then measured in tumor biopsies of 447 patients associated with three independent clinical trials. Analysis of the relationship between the tumor expression of each gene and patient outcome resulted in the identification of 21 prognostic genes, including 16 cancer-related genes and five reference genes. This 21-gene prognostic classifier is now known as the Oncotype DX profile [15]. Interestingly, five of the 16 genes in the Oncotype DX profile were identified previously in the study by van't Veer et al. [36]. The Oncotype DX genetic profile measures the probability of disease recurrence in patients diagnosed with estrogen receptor–positive, lymph node–negative breast cancer who were treated with tamoxifen. The likelihood of recurrence is measured using a recurrence score, which is calculated by measuring the expression of the 21 genes and converts these values into a score ranging from 1 to 100. Low risk of recurrence is determined by a score of <18, while ranges from 18 to 30 and ≥31 indicates intermediate and high risk of recurrence, respectively. The Oncotype DX gene set has been validated in three additional clinical trials [15,39,40] and is currently undergoing further validation in a phase III study known as TAILORx. Furthermore, the Oncotype DX gene set has been studied further for its ability to predict factors other than disease recurrence. The recurrence score was recently used to measure recurrence risk in 651 patients who were randomly assigned to treatment with tamoxifen or tamoxifen and chemotherapy [41]. The 10-year distant recurrence rates were determined for the two treatment groups and it was revealed that low- and intermediate-risk patients (as predicted by RS) did not benefit from chemotherapy. In contrast, high-risk patients derived great benefit from chemotherapy. Therefore, the Oncotype DX recurrence score appears capable of identifying in a group of patients who typically have a relatively low risk of disease recurrence those who will probably benefit from chemotherapy. This, in turn, would enable us to spend health care resources more efficiently by providing chemotherapy only to those who will benefit from treatment.

## CLASSIFICATION OF TOXINS AND PREDICTION OF DRUG TOXICITY USING GENE EXPRESSION PROFILING

Profiling of gene expression in cells is not limited to the diagnosis, classification, or prognosis of disease. One can also profile changes in gene expression

elicited by external agents such as chemotherapy drugs or toxicants. This can provide significant insight into their mechanisms of action. It is theorized that certain families of drugs or toxicants induce similar changes in gene expression in various cell lines, such that gene expression profiles or *signatures* can be used to classify known and unknown drugs or toxicants.

Nuwaysir and colleagues reviewed the potential utility of microarrays in the identification, classification, and characterization of toxicants in toxicology studies [42]. The authors suggested a procedure involving the treatment of specific cell lines or animals with or without a known toxicant, isolation of RNA from the cell lines or animal organs, reverse transcription of the RNAs to create differentially labeled cDNAs, and the profiling of gene expression by microarray analysis. The use of toxicants of a specific class with a known mechanism of action in gene profiling experiments would then allow researchers reliably to identify changes in gene expression associated with a known class of toxicant (a *toxicant signature*). After a wide variety of toxicant signatures have been established for known toxicants, the toxicant signature for an unknown toxicant could be established and compared to standard toxicant signatures in order to identify and/or classify the unknown toxicant. Moreover, the toxicant signature could provide significant insight into the biochemical mechanisms responsible for its toxicity. Nuwaysir et al. further discussed how microarray technology could be used in toxicology studies involving animal models [42]. Current methods of toxicology measure toxin levels or the expression of specific hepatic enzymes levels within blood [42,43]. The use of animal models often requires that studies take place over an extensive period of time, particularly for chronic exposure to low levels of toxins. In contrast, gene expression changes typically occur quickly (in the manner of hours) and may require only low levels of toxicants. Therefore, specific posttreatment changes in gene expression may be effective measures of exposure to certain toxicants and may provide insight both into the mechanism of toxicity and into effective measures to combat this toxicity. As raised by Nuwaysir and colleagues [42], specific standards would have to be established, including the animal model to be used, the organs to be assessed for changes in gene expression, the concentration of compounds to be administered, and the time following exposure when gene expression will be measured. Such factors will probably be dependent on the toxicant, its properties, and its target organ. Other influences on the degree of toxicity and on gene expression would also have to be minimized, including the age of subjects or animals, variations in diet, and gender.

Recently, there have been several attempts to use the approach described above to identify changes in gene expression associated with exposure to known hepatotoxic or nephrotoxic compounds in either human or animal models. Fielden et al. [44] used gene profiling approaches in rats to identify changes in gene expression induced by nephrotoxic agents, after confirmation of renal tubular degeneration by histopathology. By identifying gene profiles associated with the onset of nephrotoxicity in rats, these authors theorized

that drug development could be strongly facilitated by focusing on lead compounds that, when profiled, did not exhibit nephrotoxicity-associated changes in gene expression. In the study, rats were subjected to short, repeated exposures to well-classified toxicants or the vehicle in which they were dissolved. Changes in gene expression induced in the rats by 15 nephrotoxic and 49 nonnephrotoxic compounds were then studied by microarray analysis. Blood samples were assessed for changes in cholesterol or albumin levels, which would be consistent with the onset of nephrotoxicity. Rats were only sacrificed upon detection of nephrotoxicity-associated changes in blood levels of cholesterol or albumin. RNA was isolated from kidneys, and toxicant-induced changes in gene expression were determined for these compounds by cDNA microarray analysis using Amersham Codelink Uniset Rat1 Bioarrays and the sparse linear programming (SPLP) algorithm. Thirty-five genes were identified as changing expression only upon administration of nephrotoxic compounds to rats. Interestingly, within the training set, the 35-gene classifier was found to have 83% accuracy for identifying nephrotoxic compounds with renal tubular degeneration and 94% accuracy for identification of compounds that are not nephrotoxic. Studies were then conducted for cross-validation of the classifier with an additional 21 independent rodent samples. In the validation experiments, nephrotoxic agents were identified with 79% accuracy and nonnephrotoxic samples with 75% accuracy. Although complete accuracy at identifying nephrotoxic agents was not obtained, this approach would nevertheless help in drug discovery programs to eliminate lead compounds that have a significantly higher likelihood of being nephrotoxic. However, the identity of the 35 genes did not provide significant insight into likely mechanisms for nephrotoxicity. It is also unknown whether the gene expression changes identified in rats were specific to kidney tissue. In addition, it should be noted that there are significant differences between rats and humans in terms of their metabolism of and sensitivity to drugs and toxicants. Thus, while the study did appear able to show a significant link between the expression of specific genes and nephrotoxicity, it remains unclear whether rats serve as a good model for nephrotoxicity in humans.

An attempt to address the differences in toxicant metabolism between rats and humans was addressed in a study by Martin et al. [45]. This study utilized the potential of two toxicogenomic knowledge bases to estimate the hepatotoxicity of various compounds in humans based on changes in gene expression induced by the toxic compounds within rat in vitro and in vivo models. The in vitro analysis consisted of observing through microarray analysis changes in gene expression induced by 86 compounds, 36 of which were nonhepatotoxins. In contrast, the in vivo study analyzed gene expression changes induced by 111 compounds, 32 of which were nonhepatotoxic. Affymetrix RGU34 microarrays were used containing probes for 8799 genes. Toxicant-induced changes in gene expression were identified through comparison to control cell lines or rats exposed to the solvent in which the toxicant was dissolved. Using a tree-based machine learning technique known as *gradient boosting*, 173

genes were identified whose expression would be expected to be modified by hepatotoxins in humans. In the in vivo experiments, 168 genes would be expected to change expression in humans by hepatotoxins. Unfortunately, while the in vitro and in vivo data identified real gene classifiers associated with hepatotoxicity, only five genes were identified in both the in vitro and in vivo experiments. Furthermore, none of the common genes identified were within the top 30 changes in gene expression, and validation experiments strongly suggested that both the in vitro and in vivo gene sets were found to be poor predictors of toxicity. To increase the reliability of predictive gene sets and decrease the degree of variability within the data, it is recommended that higher numbers of nontoxic compounds also be monitored for changes in gene expression (as controls) and that associated toxicology databases have better coverage of human studies.

Given the findings described above, it appears that significant additional work is required before reliable gene classifiers for the identification of nephrotoxic or hepatotoxic compounds are available. Most toxicology studies are performed in animal model systems due to ethical concerns and it appears unlikely that such studies can be extrapolated easily to humans (particularly if rodents are used). Human cell lines could be used, but such in vitro experiments are unlikely to be replicated in living human subjects. Nevertheless, the development of gene classifiers for toxicants in nonhuman systems could assist the drug discovery process by avoiding lead compounds that induce changes in gene expression in both rodents and human cell lines which resemble strongly those of known nephrotoxic and hepatotoxic compounds. In addition, gene classifiers for agents capable of inducing toxicity to other organs (e.g., cardiotoxicity) can possibly be developed using this approach.

## USE OF GENE PROFILING TO PREDICT DRUG SENSITIVITY OR RESISTANCE

Approaches similar to those described above for toxicants can also be used to classify drugs and their modes of action. Moreover, it may be possible to use genome profiling of patients to identify those that would probably benefit from administration of a particular drug. Considerable emphasis has been placed recently on the identification of genes whose expression in the tumors or host tissues of cancer patients can predict or measure the response to specific chemotherapy agents or regimens. Various families of chemotherapy drugs act by different mechanisms within tumors, such as the inhibition of topoisomerase II or DNA replication [46–49] or the stabilization of microtubules [50]. Destruction of such essential functions would be expected to trigger cell cycle checkpoints and promote cell death mechanisms, resulting in specific changes in gene expression that are characteristic to the administered drug. In addition, since tumors can often be heterogeneous, there may exist within the tumor population variants that before or after drug treatment are better able to

survive in the presence of chemotherapy agents. These or additional variants may also be capable of inactivating specific chemotherapy agents or blocking drug-induced cell death pathways. This would result in drug resistance, which is often observed in cancer patients before or after treatment with chemotherapy agents. By genome profiling of tumor or host tissues from cancer patients, it may be possible to identify which tumors or patients are likely to exhibit resistance to a given chemotherapy agent. Moreover, the genome profile may enable the oncologist to select a chemotherapy agent that has the highest probability of killing a given tumor. The profiling and classification of tumors into specific subtypes (as described above) may aid further in this goal.

Chang et al. were the first researchers to use gene profiling successfully to identify possible genes whose expression in tumors pretreatment could differentiate between responders and nonresponders of docetaxel chemotherapy [10]. Patient tumor size was measured throughout treatment, allowing classification of 24 patients into 11 responders and 13 nonresponders (who had progressive disease or no response to treatment). RNA was isolated from tumors before chemotherapy, amplified, reverse-transcribed to cDNA, labeled with Cy3 or Cy5 fluorescent dyes, and hybridized to HgU95-Av2 Affymetrix microarrays. Student's *t*-tests were used to identify 92 genes whose expression was significantly different between responders and nonresponders. Using these genes in a leave-one-out cross-validation approach, the gene set appeared to correctly classify responders and nonresponders with an accuracy of 88% within its training set. The genes identified were found to be involved in various cellular functions, such as cell signaling, immunologic response, DNA damage detection or repair, cell cycle regulation, and tumor suppressors. To confirm this gene set, the expression level of 15 differentially expressed genes was determined by reverse-transcription polymerase chain reaction. Thirteen positive correlations were obtained, of which six were found to be highly significant between responders and nonresponders. To test this genetic profile further , RNA was isolated from six additional patients who were not involved in the training set. Quantitation of expression of the 92 genes using the same approach correctly identified all six as drug responders. The accuracy of the 92-gene profile to predict response to docetaxel chemotherapy was then further evaluated in a follow-up study by the same investigators [51]. The authors observed gene expression of resistant and sensitive tumors after 3 weeks of treatment in a subset of 13 patients. Surprisingly, very few differences in gene expression were observed. Therefore, it would appear that changes in tumor gene expression shortly after administration of docetaxel cannot be used to differentiate between responding and nonresponding tumors.

A study by Iwao-Koizumi et al. also used high-throughput Q-PCR to identify genes whose expression could predict response to docetaxel in 44 breast tumors [16]. The expression of 2453 genes was profiled in the tumors relative to controlling RNA from 78 primary breast cancers. WV, K-NN, and SVM approaches were used in data analyses to identify gene sets, which were then

subjected to leave-one-out verification studies. It was observed that the most consistent results were obtained with the use of the WV algorithm, which identified 85 genes that were capable of identifying response to docetaxel with 80.7% accuracy in verification studies. To provide further evidence of their utility, expression vectors containing cDNAs for a number of the genes associated with resistance to docetaxel chemotherapy were transfected into MCF-7 cells and were found to increase survival in the presence of docetaxel compared to nontransfected controls. This suggests that observations from in vitro experiments using cell lines may have relevance to drug resistance in cancer patients. Interestingly, the genes determined to be associated with docetaxel response in this study differed greatly from those identified in the study by Chang and colleagues [10]. Only three genes were in common between the two studies. A number of differences in the studies may account for this discordance, including different patient populations, widely varying sets of genes being profiled, and differences in the algorithms used in data analysis. Use of a more quantitative method for gene profiling, the larger set of patients, and the higher prediction accuracy would suggest that the 85-gene set identified by Iwao-Koizumi has a higher probability of being replicated in other independent studies.

Genome profiling was also used by Cleator et al. to identify genes whose expression is predictive of response to doxorubicin/cyclophosphamide chemotherapy in breast cancer patients [52]. In this study, RNA samples were isolated from 12 patients whose tumors responded to chemotherapy and from six patients whose tumors did not. After amplification and labeling, the RNAs were hybridized to Affymetrix HgU133A arrays. After scanning the arrays, the output data were assessed using a "randomized variance model," and 253 genes were identified to be differentially expressed between responding and nonresponding tumors. Leave-one-out cross-validation studies were then performed using the training set, and only 67% of samples were correctly classified as responders or nonresponders. This accuracy is only slightly better than chance. Thus, this approach could not be used to identify genes associated with response to doxorubicin/cyclophosphamide chemotherapy, or, more likely, the number of patients in the training set was of insufficient size to reliably identify genetic biomarkers for prediction of response to doxorubicin/cyclophosphamide chemotherapy. The authors of this study also tested the ability of the Chang 92-gene set for docetaxel responsiveness [10] (described above) to predict response to doxorubicin/cyclophosphamide treatment. The 92-gene profile was unable to distinguish between responders and nonresponders of doxorubicin/cyclophosphamide chemotherapy, suggesting that genes for prediction of response to chemotherapy regimens are probably regimen-specific and not general biomarkers of drug response.

A similar approach to that described above was used to identify 74 genes whose expression correlates positively or negatively with response to T/FAC chemotherapy (paclitaxel, followed by 5-fluorouracil, doxorubicin, and cyclophosphamide). Accuracy at measuring response to this regimen was deter-

mined to be 78% [53]. Independent expression of any one of the 74 genes could not be used to predict response to T/FAC chemotherapy.

As mentioned previously, false discovery rates are often high in genome profiling experiments, particularly when tens of thousands of genes are surveyed. Thus, in the gene profiling studies above, the genes identified may correlate with drug response, but, in reality, may have no relationship to drug action. This would then explain the many failures that have been observed in identifying genes whose expression correlates with drug response across a wide variety of independent clinical studies. To circumvent this problem, it can often be advantageous to restrict the analysis of genes in patient populations to those that have demonstrated a clear relationship with drug response or resistance in in vitro studies. A number of studies have identified clear differences in gene expression between drug-sensitive and drug-resistant tumor cells. For example, Villeneuve et al. used microarray studies to identify a number of genes whose expression was altered in doxorubicin- or paclitaxel-resistant MCF-7 breast tumor cells compared to wild-type cells [12]. The majority of genes (>98%) exhibited no difference in expression between the wild-type and drug-resistant cell lines, confirming the isogenicity of the established cell line. One strength of this study was that 91 to 95% of the differences in gene expression identified by microarray analysis were corroborated through Q-PCR verification experiments. In addition, control cell lines were created to account for differences in gene expression during selection for drug resistance that were simply due to extended propagation of cells in culture. The identities of the genes uncovered in this study were associated with drug transport, drug metabolism, growth promotion, and cell death inhibition, consistent with the types of genes that one might expect to observe in drug-resistant cells. Thus, this approach has the prospect of providing significant insight into the many and varied mechanisms by which tumor cells acquire resistance to chemotherapy drugs.

The limitation of the study by Villeneuve et al. [12] is that gene expression was compared between single sets of isogenic wild-type and drug-resistant breast tumor cell lines. Although the genes identified may play a role in drug resistance in a single breast tumor cell line, some may have little relevance to drug resistance in other breast tumor cell lines or other cell lines of different tissue origin. In contrast, genes having altered expression in cell lines from several tissues upon acquisition of resistance to a specific drug have a higher likelihood of playing a bona fide role in drug resistance, although probably not in a tissue-specific manner. Györffy et al. profiled gene expression across 30 different cancer cell lines using HGU133 Affymetrix arrays and identified genes whose expression correlated with sensitivity or resistance to 11 different chemotherapy drugs [24]. The accuracy of the gene sets to identify a cell line as sensitive or resistant to a given chemotherapy agent was 86%. Using this approach, 1481 genes were identified as playing a role in resistance to a diverse set of anticancer agents. Not surprisingly, the drug-resistance genes identified were different depending on the chemotherapy agent. Only 67 genes were

observed to be associated with resistance to four or more anticancer agents, and the expression level of only two genes correlated with resistance to all 11 drugs. In a similar study [54], HGU133 Affymetrix arrays were also used in the study by Györffy et al. to identify genes associated with resistance in gastric cancer cells to three of the agents used. Despite similar array platforms, the number and identities of genes associated with resistance to 5-fluorouracil, doxorubicin, or cisplatin varied dramatically between the two studies. A number of factors could account for the differences in the gene numbers and identities. Whereas the former study assessed a large number of wild-type and drug-resistant cell lines of varying tissue origin for sensitivity or resistance to 11 chemotherapy agents, the latter study examined only 14 gastric cancer cell lines, of which 10 were resistant to one of three chemotherapy agents. Thus, it is possible that the genes associated with drug resistance in gastric cancer cells are distinct from those that would have relevance across cell lines. Alternatively, variations in experimental design, the number of cell lines, and approaches to microarray data analysis resulted in the identification of different gene sets, which may or may not be relevant to drug resistance.

## STRATEGIES TO IMPROVE THE ACCURACY OF GENE PROFILING IN IDENTIFYING PROGNOSTIC AND PREDICTIVE BIOMARKERS

As stated previously, very few gene sets have proven reliable in predicting response to chemotherapy agents across large groups of patients [10,11,16]. Numerous factors may contribute to the poor reproducibility of findings from gene profiling experiments, including differences in the tissues studied, the array platforms, the number and nature of genes on the microarrays, RNA extraction procedures, the number of replicate arrays used in experiments, and the data analysis and verification methods. Simon and colleagues reviewed approaches to improving the accuracy of gene profiling experiments in the identification of reliable predictive or prognostic biomarkers [26]. The first limitation to the reliable identification of predictive and prognostic biomarkers is the size of the training and validation sets, which in many studies is from 30 to 50 patients. Larger data sets of 500 patients or more have a much higher probability than small data sets of identifying useful biomarkers that can be validated across experiments. The use of more than one method for obtaining gene classifiers is also preferred. Moreover, in addition to larger data sets, there should be equal representation between patients from which the positive and negative classifiers will be developed. In addition, a number of studies only validate the utility of a gene set to positively identify patients with a particular attribute, without showing that the positive classifier has no ability to identify patients without the attribute.

Minna et al. [23] discussed further the reproducibly of genome profiling experiments across laboratories in an editorial written in response to a study

by Hsu et al. [25] which identified a set of genes that could serve as a classifier of cisplatin responsiveness. While acknowledging the promising prospects of the study by Hsu and colleagues, Minna et al. commented on what additional information could have been described or collected in this and other microarray experiments. These include the method used for classification of patients, the analytical procedures used in the genome profiling experiments, and access to *all* microarray data so that it is available for analysis by researchers. The free availability of well-described data sets will enable future researchers to use this information to identify the best analytical methods for microarray data and to compare microarray data analysis tools effectively across experiments. This will enable researchers to determine if future gene sets can accurately predict drug response in groups of patients outside the initial training and validation sets. It will also enable researchers to repeat or apply new analytical procedures to previous data sets in order to make effective comparisons across experiments. Interestingly, it has been shown that with the use of identical gene arrays, extraction methods, and statistical analyses, identical gene signatures can be obtained across microarray experiments, despite differences in operators, sample handling, and time of extraction [55].

## FUTURE DIRECTIONS

There can be little doubt that genome profiling experiments are beginning to yield important new biomarkers of clinical relevance, particularly for the diagnosis of disease states. However, considerable variability exists in the performance of genome profiling experiments across studies, in particular for microarray studies. Differences in array platforms, the number and identities of genes on arrays, sample size, experimental design, methods of patient classification and RNA extraction, and algorithms for data analysis all contribute to the wide variability observed in microarray findings. Only through comparisons across well-characterized microarray experiments can a clear conclusion be made as to the optimal design of a microarray experiment and optimal methods for both data analysis and patient classification. Moreover, by describing array experiments in detail and ensuring full access to microarray data, it should be possible to repeat microarray data analyses using such optimal methods (once defined). In the future, the preferred method of microarray experimental design and microarray data analysis will enable more accurate and reproducible findings to be achieved. This will dramatically increase the capacity of microarray studies to affect the clinical diagnosis and management of cancer patients.

Microarray or PCR-based gene profiling experiments are powerful tools which, used optimally will dramatically increase the discovery of gene classifiers for the accurate diagnosis of cancer patients and for the optimal management of their care. This would probably include gene classifiers for the prediction of patient prognosis pre- and posttherapy, the optimal selection of

chemotherapy agents, and the measurement of patient response and/or toxicity to these agents. Moreover, genome profiling is likely to have a major impact on drug discovery programs by enabling investigators to classify and characterize drugs and toxicants, with the goal of optimizing drug structures to achieve maximal therapeutic benefit with minimal toxicity to patients. Gene profiling of patients and their tumors before and after drug treatment may also enable us to tailor our approaches to the management of cancer patients by accurately identifying those that will derive benefit, with minimal toxicity, from a given chemotherapy regimen. For the pharmaceutical industry, this may also permit the introduction or reintroduction of highly effective drugs into the marketplace that exhibit toxicities in rare subsets of patients.

Recent evidence confirms that gene profiling experiments are beginning to revolutionize the practice of oncology and that physicians are prepared to embrace their use, provided that the utility of genetic biomarkers identified from gene profiling experiments are widely validated in appropriately conducted clinical trials. This appears to be the case for both Oncotype DX [15,41] and Mammaprint [56] gene sets for prediction of prognosis in women with breast cancer. The Oncotype DX gene expression assay will probably be widely used by breast oncologists for a number of reasons:

1. The recurrence score associated with the gene set used in this assay greatly aids the oncologist in assessing the prognosis of a patient and the need for adjuvant chemotherapy, such as MF/CMF [41]. This is particularly useful for low-risk patients with node-negative, ER-positive tumors, where the physician is likely to recommend adjuvant chemotherapy for patients with high recurrence scores but not for patients with low recurrence scores.
2. Respected clinical organizations such as the American Society of Clinical Oncology [57] and the National Comprehensive Cancer Network have established clear clinical guidelines for use of the Oncotype DX assay in patients with tumors of 0.6 to 1 cm with moderate to poorly differentiated or unfavorable features, and all tumors larger than 1 cm.
3. Gene expression is assessed by a highly quantitative procedure (Q-PCR) conducted centrally in a recognized laboratory following good laboratory practices (GLP) guidelines.
4. There is now evidence that an expanded Oncotype DX assay, which includes estrogen and progesterone receptor scores, provides a quantitative measure of the expression of these receptors with a high degree of concordance with standard immunohistochemical approaches [58]. This provides greater information to the physician on the degree of expression of the receptors and the confidence that can be placed on the diagnosis and treatment plans based on receptor status. Such developments further increase the adoption of this assay by breast oncologists. In addition to the benefits of the Oncotype DX assay, recent studies using the new molecular classifications of breast cancer (based on earlier genome

profiling studies) suggest that responses to specific therapies may vary significantly among the various subgroups.

Successful adoption of the Oncotype DX assay into clinical trials and into clinical practice guidelines clearly illustrates that oncologists are willing to accept quantitative gene profiling of patient tumors as an important tool in the clinical management of a cancer patient, but this requires the successful translation of findings in research laboratories into validated robust assays that can be conducted in licensed clinical service laboratories. The support and cooperation of clinical societies is critical to the adoption of gene-based assays into the clinic, including the establishment of clear practice guidelines for the physician.

## REFERENCES


1. Goldhirsch A, Glick JH, Gelber RD, Senn HJ (1998). Meeting highlights: International Consensus Panel on the Treatment of Primary Breast Cancer. *J Natl Cancer Inst*, 90:1601–1608.
2. Eifel P, Axelson JA, Costa J, et al. (2001). National Institutes of Health Consensus Development Conference Statement: adjuvant therapy for breast cancer, Nov. 1–3. *J Natl Cancer Inst*, 93:979–989.
3. McGuire WL (1991). Breast cancer prognostic factors: evaluation guidelines. *J Natl Cancer Inst*, 83:154–155.
4. Bast RC, Ravdin P, Hayes DF, et al. (2001). 2000 update of recommendations for the use of tumor markers in breast and colorectal cancer: clinical practice guidelines of the American Society of Clinical Oncology. *J Clin Oncol*, 19: 1865–1878.
5. Gottesman MM, Fojo T, Bates SE (2002). Multidrug resistance in cancer: role of ATP-dependent transporters. *Natl Rev Cancer*, 2:48–58.
6. Jänicke F, Prechtl A, Thomssen C, et al. (German N0 Study Group) (2001). Randomized adjuvant chemotherapy trial in high-risk, lymph node-negative breast cancer patients identified by urokinase-type plasminogen activator and plasminogen activator inhibitor type 1. *J Natl Cancer Inst*, 93:913–920.
7. van Diest PJ, Michalides RJ, Jannink L, et al. (1997). Cyclin D1 expression in invasive breast cancer: correlations and prognostic value. *Am J Pathol*, 150:705–711.
8. Villeneuve DJ, Parissenti AM (2004). The use of DNA microarrays to investigate the pharmacogenomics of drug response in living systems. *Curr Top Med Chem*, 4:1329–1345.
9. 't Hoen PA, de KF, van Ommen GJ, den Dunnen JT (2003). Fluorescent labelling of cRNA for microarray applications. *Nucl Acids Res*, 31:e20.
10. Chang JC, Wooten EC, Tsimelzon A, et al. (2003). Gene expression profiling for the prediction of therapeutic response to docetaxel in patients with breast cancer. *Lancet*, 362:362–369.

11. Korkola JE, Blaveri E, DeVries S, et al. (2007). Identification of a robust gene signature that predicts breast cancer outcome in independent data sets. *BMC Cancer*, 7:61.
12. Villeneuve DJ, Hembruff SL, Veitch Z, Cecchetto M, Dew WA, Parissenti AM (2006). cDNA microarray analysis of isogenic paclitaxel- and doxorubicin-resistant breast tumor cell lines reveals distinct drug-specific genetic signatures of resistance. *Breast Cancer Res Treat*, 96:17–39.
13. van de Vijver MJ, He YD, van't Veer LJ, et al. (2002). A gene-expression signature as a predictor of survival in breast cancer. *N Engl J Med*, 347:1999–2009.
14. Brazma A, Hingamp P, Quackenbush J, et al. (2001). Minimum information about a microarray experiment (MIAME)-toward standards for microarray data. *Nat Geneti*, 29:365–371.
15. Paik S, Shak S, Tang G, et al. (2004). A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. *N Engl J Med*, 351:2817–2826.
16. Iwao-Koizumi K, Matoba R, Ueno N, et al. (2005). Prediction of docetaxel response in human breast cancer by gene expression profiling. *J Clin Oncol*, 23:422–431.
17. Chang JC, Makris A, Gutierrez MC, et al. (2007). Gene expression patterns in formalin-fixed, paraffin-embedded core biopsies predict docetaxel chemosensitivity in breast cancer patients. *Breast Cancer Res Treat*, 108:233–240.
18. Kohlmann A, Schoch C, Schnittger S, et al. (2003). Molecular characterization of acute leukemias by use of microarray technology. *Genes Chromosomes Cancer*, 37:396–405.
19. Golub TR, Slonim DK, Tamayo P, et al. (1999). Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science*, 286:531–537.
20. Zhao YP, Chen G, Feng B, et al. (2007). Microarray analysis of gene expression profile of multidrug resistance in pancreatic cancer. *Chin Medi J*, 120:1743–1752.
21. Furey TS, Cristianini N, Duffy N, Bednarski DW, Schummer M, Haussler D (2000). Support vector machine classification and validation of cancer tissue samples using microarray expression data. *Bioinformatics*, 16:906–914.
22. Pomeroy SL, Tamayo P, Gaasenbeek M, et al. (2002). Prediction of central nervous system embryonal tumour outcome based on gene expression. *Nature*, 415:436–442.
23. Minna JD, Girard L, Xie Y (2007). Tumor mRNA expression profiles predict responses to chemotherapy. *J Clin Oncol*, 25:4329–4336.
24. Györffy B, Surowiak P, Kiesslich O, et al. (2006). Gene expression profiling of 30 cancer cell lines predicts resistance towards 11 anticancer drugs at clinically achieved concentrations. *Intl J Cancer*, 118:1699–1712.
25. Hsu DS, Balakumaran BS, Acharya CR, et al. (2007). Pharmacogenomic strategies provide a rational approach to the treatment of cisplatin-resistant patients with advanced cancer. *J Clin Oncol*, 25:4350–4357.
26. Simon R, Radmacher MD, Dobbin K, McShane LM (2003). Pitfalls in the use of DNA microarray data for diagnostic and prognostic classification. *J Natl Cancer Inst*, 95:14–18.
27. Haferlach T, Kohlmann A, Schnittger S, et al. (2005). Global approach to the diagnosis of leukemia using gene expression profiling. *Blood*, 106:1189–1198.

28. Sørlie T, Perou CM, Tibshirani R, et al. (2001). Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc Natl Acad Sci USA*, 98:10869–10874.
29. Perreard L, Fan C, Quackenbush JF, et al. (2006). Classification and risk stratification of invasive breast carcinomas using a real-time quantitative RT-PCR assay. *Breast Cancer Res*, 8:R23.
30. Hu Z, Fan C, Oh DS, et al. (2006). The molecular portraits of breast tumors are conserved across microarray platforms. *BMC Genomi*, 7:96.
31. Mullins M, Perreard L, Quackenbush JF, et al. (2007). Agreement in breast cancer classification between microarray and quantitative reverse transcription PCR from fresh-frozen and formalin-fixed, paraffin-embedded tissues. *Clini Chem*, 53: 1273–1279.
32. Cai B, Liu L, Xi W, et al. (2007). Comparison of the molecular classification with FIGO stage and histological grade on endometrial cancer. *Eur J Gynaecol Oncol*, 28:451–460.
33. Seike M, Yanaihara N, Bowman ED, et al. (2007). Use of a cytokine gene expression signature in lung adenocarcinoma and the surrounding tissue as a prognostic classifier. *J Natl Cancer Inst*, 99:1257–1269.
34. Tschoep K, Kohlmann A, Schlemmer M, Haferlach T, Issels RD (2007). Gene expression profiling in sarcomas. *Critl Rev Oncol Hematol*, 63:111–124.
35. Warnat P, Oberthuer A, Fischer M, Westermann F, Eils R, Brors B (2007). Cross-study analysis of gene expression data for intermediate neuroblastoma identifies two biological subtypes. *BMC Cancer*, 7:89.
36. van't Veer LJ, Dai H, van de Vijver MJ, et al. (2002). Gene expression profiling predicts clinical outcome of breast cancer. *Nature*, 415:530–536.
37. Sotiriou C, Neo SY, McShane LM, et al. (2003). Breast cancer classification and prognosis based on gene expression profiles from a population-based study. *Proc Natl Acad Sci USA*, 100:10393–10398.
38. Perou CM, Sorlie T, Eisen MB, et al. (2000). Molecular portraits of human breast tumours. *Nature*, 406:747–752.
39. Cronin M, Sangli C, Liu ML, et al. (2007). Analytical validation of the Oncotype DX genomic diagnostic test for recurrence prognosis and therapeutic response prediction in node-negative, estrogen receptor-positive breast cancer. *Clin Chem*, 53:1084–1091.
40. Habel LA, Shak S, Jacobs MK, et al. (2006). A population-based study of tumor gene expression and risk of breast cancer death among lymph node-negative patients. *Breast Cancer Res*, 8:R25.
41. Paik S, Tang G, Shak S, et al. (2006). Gene expression and benefit of chemotherapy in women with node-negative, estrogen receptor-positive breast cancer. *J Clin Oncol*, 24:3726–3734.
42. Nuwaysir EF, Bittner M, Trent J, Barrett JC, Afshari CA (1999). Microarrays and toxicology: the advent of toxicogenomics. *Mol Carcinog*, 24:153–159.
43. Batt AM, Ferrari L (1995). Manifestations of chemically induced liver damage. *Clin Chem*, 41:1882–1887.
44. Fielden MR, Eynon BP, Natsoulis G, Jarnagin K, Banas D, Kolaja KL (2005). A gene expression signature that predicts the future onset of drug-induced renal tubular toxicity. *Toxicol Pathol*, 33:675–683.

45. Martin R, Rose D, Yu K, Barros S (2006). Toxicogenomics strategies for predicting drug toxicity. *Pharmacogenomics*, 7:1003–1016.
46. Tarr M, van Helden PD (1990). Inhibition of transcription by adriamycin is a consequence of the loss of negative superhelicity in DNA mediated by topoisomerase II. *Mol Cell Biochem*, 93:141–146.
47. Zijlstra JG, de Jong S, de Vries EG, Mulder NH (1990). Topoisomerases, new targets in cancer chemotherapy. *Med Oncol Tumor Pharmacother*, 7:11–18.
48. Binaschi M, Bigioni M, Cipollone A, et al. (2001). Anthracyclines: selected new developments. *Curr Medi Chem Anticancer Agents*, 1:113–130.
49. Galmarini CM, Mackey JR, Dumontet C (2002). Nucleoside analogues and nucleobases in cancer treatment. *Lancet Oncol*, 3:415–424.
50. Chazard M, Pellae-Cosset B, Garet F, et al. (1994). [Taxol (paclitaxel), first molecule of a new class of cytotoxic agents: taxanes.] *Bull Cancer*, 81:173–181.
51. Chang JC, Wooten EC, Tsimelzon A, et al. (2005). Patterns of resistance and incomplete response to docetaxel by gene expression profiling in breast cancer patients. *J Clin Oncol*, 23:1169–1177.
52. Cleator S, Tsimelzon A, Ashworth A, et al. (2006). Gene expression patterns for doxorubicin (Adriamycin) and cyclophosphamide (Cytoxan) (AC) response and resistance. *Breast Cancer Res Treat*, 95:229–233.
53. Ayers M, Symmans WF, Stec J, et al. (2004). Gene expression profiles predict complete pathologic response to neoadjuvant paclitaxel and fluorouracil, doxorubicin, and cyclophosphamide chemotherapy in breast cancer. *J Clin Oncol*, 22:2284–2293.
54. Kang HC, Kim IJ, Park JH, et al. (2004). Identification of genes with differential expression in acquired drug-resistant gastric cancer cells using high-density oligonucleotide microarrays. *Clin Cancer Res*, 10:272–284.
55. Kohlmann A, Schoch C, Dugas M, et al. (2005). Pattern robustness of diagnostic gene expression signatures in leukemia. *Genes Chromosomes Cancer*, 42:299–307.
56. Wittner BS, Sgroi DC, Ryan PD, et al. (2008). Analysis of the MammaPrint breast cancer assay in a predominantly postmenopausal cohort. *Clin Cancer Res*, 14:2988–2993.
57. Harris L, Fritsche H, Mennel R, et al. (2007). American Society of Clinical Oncology 2007 update of recommendations for the use of tumor markers in breast cancer. *J Clin Oncol*, 25:5287–5312.
58. Badve SS, Baehner FL, Gray RP, et al. (2008). Estrogen- and progesterone-receptor status in ECOG 2197: comparison of immunohistochemistry by local and central laboratories and quantitative reverse transcription polymerase chain reaction by central laboratory. *J Clin Oncol*, 26:2473–2481.

# USE OF HIGH-THROUGHPUT PROTEOMIC ARRAYS FOR THE DISCOVERY OF DISEASE-ASSOCIATED MOLECULES

Douglas M. Molina, Ph.D., W. John W. Morrow, Ph.D., and Xiaowu Liang, Ph.D.
*Antigen Discovery, Inc., Irvine, California*

## INTRODUCTION

The discovery of biologically relevant disease-state markers, or biomarkers, is a crucial step in the maturation of accurate disease diagnosis and appropriate treatment. With all the advances in modern research, the need for better diagnostics as well as markers of disease is paramount. The management of infection, cancer, and autoimmune disorders could benefit greatly from new biomarkers, or combinations of biomarkers, that could provide quick and accurate identification of the causative agent of disease, disease state, and even the stage of the disease. The lack of progress in identifying such molecules is due primarily to the way that, to date, researchers have approached the problem. When using traditional approaches to the discovery of serodiagnostic antigens, researchers have been limited to the number of molecules they could study. Typically, most laboratories have studied just a handful of proteins at any given time and therefore have only a limited understanding of the pathobiology of disease. Much like the children's game *Flip 'n' Match*, inves-

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

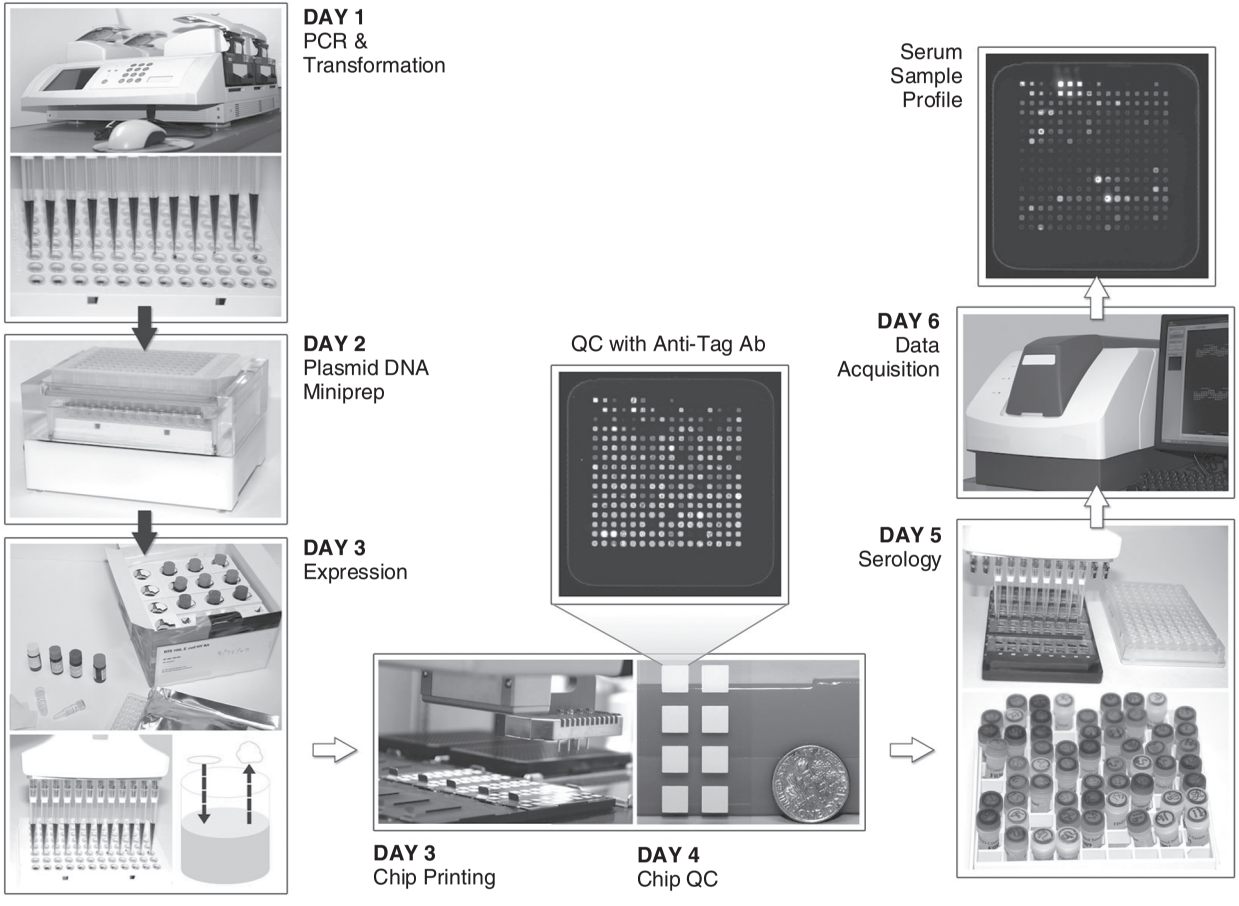


**Figure 1** High-throughput proteomics screening technique used by the authors.

tigators have employed the "best guess" approach to which "pieces" (i.e., antigens or proteins) may be most appropriate as disease markers. Such best-guess hypotheses have usually been based on the body of literature available as well as extrapolations from our understanding of the immune response. Much as in the game, once in a while the correct choice is made. However, a much more productive approach would be to view all the pieces simultaneously and assess the complete picture, without prejudice, in the context of the immune response profile. In this way we would be in a much better position to make determinations as to which proteins, or combinations of proteins, would allow us to diagnose and monitor disease accurately. Very recently, such an approach has been made possible, and in this chapter we discuss how researchers have taken advantage of the sequencing boom sparked by the human genome project. We also examine assay miniaturization technology, which has carried over from the nucleic acid microarray field, and how new advances in cloning and gene expression have allowed the use of those resources to profile the immune response to infection, cancer, and autoimmune disease.

## OVERVIEW OF THE PRODUCTION AND USE OF MICROARRAYS FOR THE DISCOVERY OF ANTIGENIC BIOMARKERS

To take advantage of genomic sequence data for the development of novel therapeutics and clinical diagnostics, high-throughput cloning and screening technologies are required. The authors and own collaborators have developed and described a gene cloning and expression platform that allows whole proteome microarrays to be constructed from genomic sequence information in a relatively short time. The platform can then be screened with large numbers of well-characterized sera to identify antibody targets. This powerful high-throughput approach lends itself to the rapid identification of reactive antigens on an epidemiological scale, and should lead to the development of improved serodiagnostic tests for many diseases (Figure 1).

## PROTEIN MICROARRAYS

In the context of the discovery of disease-associated molecules, protein microarrays can be thought of as miniaturized dot blots or highly multiplexed enzyme-linked immunosorbent assays (ELISAs) for serology, and highly multiplexed sedimentation, co-immunoprecipitation assays, or enzyme function assays for protein biochemistry. They allow data collection from hundreds of parallel interaction studies in an area the size of a single well of a 96-well ELISA plate. To achieve this type of miniaturization, the chemistry of the surface being used for spotting the proteins is important. Reagents can

be coated onto activated plastics or glass to accommodate the deposition and subsequent binding of the protein onto the surface. Some examples of the chemistries used to bind proteins to the substrates are described below.

### Two-Dimensional Substrates

Well-characterized protein-binding substrates, used extensively in other assays, have been transferred or "ported" for the use of microarray fabrication. Most of these surface substrates were used extensively in DNA microarray research and have been used with success with purified protein arrays. There are two principal ways that proteins bind to surfaces: (1) by the formation of covalent bonds with reactive groups on the surface of the slide, and (2) by noncovalent interactions with the slide substrate.

In slide surface chemistry, the use of aldehyde and epoxy coatings allows for covalent bonds between the proteins by virtue of the amine group on amino acids and the activated aldehyde of the epoxy group on the slide [1,2]. The primary amine group reacts with the aldehyde, or epoxy, groups on the slide surface. The electrons on the nitrogen attach to the carbon with a partial positive charge on the reactive group and form a nitrogen–carbon bond. In protein–aldehyde coupling, the attachment is stabilized by a dehydration reaction. Another option for two-dimensional substrates is the coat the slide surface with a molecule that is part of a high-affinity binding interaction. Binding-partner substrates are not as common but are still useful. There are examples of protein-binding surfaces that have been treated with avidin, streptavidin, glutathione, and monoclonal antitag antibodies. Unlike the aldehyde and epoxy examples, these are very specific and require that the appropriate binding partner be coupled to the protein(s) being spotted.

### Three-Dimensional Substrates

Polyacrylamide-coated slides form a three-dimensional substrate for protein microarrays. Much like the polyacrylamide gels used for electrophoresis, they are reported to provide high probe loading capacity and a hydrophilic, solution phase–like environment, which is thought to preserve protein probe function [3–6]. Consequently, this flexible and versatile platform is well suited for many types of proteins and has been used to develop a variety of assays in microarray format. A three-dimensional hydrogel coating preserves the native three-dimensional structure of proteins, thereby maintaining stability and functionality. The reactive chemistry is stable and remains active even during very long spotting (also know as "printing") runs. Agarose is an economical, protein-friendly medium that reduces the background of the slide significantly, which can interfere with signal detection of low-abundance proteins or pep-

tides. There are other proprietary three-dimensional porous polymer surfaces, such as those produced by Full Moon Biosystems, that claim to provide a high level of sensitivity and specificity for fluorescence-based assays. All seem to have similar properties, which include covalent multifunctional bonding sites that provide high binding specificity by covalent attachment to surface functional groups.

Nitrocellulose is the best-characterized protein-binding chemistry that has been adapted successfully to microarray fabrication. Nitrocellulose membranes are used for immobilizing a wide variety of molecules, such as nucleic acids in Southern and Northern blots, as well as proteins in Western blots. Nitrocellulose makes an ideal surface for a wide variety of assays, including antibody-based assays, due to its nonspecific affinity for proteins. Nitrocellulose is usually coated onto a nylon backbone to produce membranes. It has also been coated on activated glass microscope slide, and as such has become a very popular tool for the fabrication of protein microarrays. A variety of nitrocellulose-coated slides are currently available which are ideal for microarrays because of their high binding capacity per unit volume. At one end of the spectrum, the nitrocellulose surface may be as thin as 5 μm (Path Slides, GenTel), and at the other, the membrane can be as thick as 150 μm (SuperNitro, Telechem). In the case of nitrocellulose-coated slides, it is critically important to pick the correct thickness for the application. The thickness of the coating chosen will be influenced greatly by the type of microarray printing that is to be used.

## Contact vs. Noncontact Printing

Like DNA microarray fabrication, protein microarray fabrication is dependent on applying the protein onto the substrate reliably and reproducibly. To achieve this objective, two major classes of spotting, or microarray printing, have evolved. Noncontact printing is the mechanism of spotting onto the slide or plate surface without any contact. The volumes range from picoliters up to hundreds of microliters having coefficients of variation (CVs) of less than 10%, depending on the technology used. One such technology, solenoid-based dispensing systems, relies on the rapid opening and closing of valves, which permits the flow of pressurized liquid. Rapid opening and closing of a valve, or pulsing, results in a very small volume being deposited. Leaving the valve open longer results in the deposition of larger volumes. Solenoid-based systems typically spot between 10 nl and hundreds of microliters. Another noncontact technology, piezoelectric-based dispensing systems, can dispense as little as picoliter quantities of liquid. They rely on electrical pulses that create a pressure wave when a piezoelectric element is deformed and squeezes the tubing, the pressure created results in the displacement of a small volume of liquid. By opening the valve at the precise moment that the piezoelectric element is deformed, a small volume of liquid is deposited on the substrate.

The main drawback of noncontact printing is that it is not the ideal tool for arraying more viscous or concentrated protein mixtures. Cell lysates immediately come to mind when considering heterogeneous materials that may cause problems for noncontact arrayers. There is a high potential for clogging not only the dispensing solenoids, but the tubing as well. Another drawback is that the noncontact dispensing robots available are limited by the number of valves that can be used to array spots onto a slide or plate; thus, throughput becomes a problem. When printing thousands of different samples, the number of solenoids becomes the limiting step.

Contact printers address these shortcomings directly. Although there are numerous variants of contact printing technologies, the two most prevalent are the solid pin and split pin methods. Unlike the solenoid- and valve-based systems, pin-based printing systems are not limited by size. Some printers use the traditional 9-mm pin spacing used in 96-well microtiter plates, but the majority use 4.5-mm pin spacing. More recently, printers that use 2.25-mm pin spacing have become available. Depending on the size of the printhead, as many as 256 pins in an $8 \times 32$ configuration can be fitted when the 2.25-mm spacing is used. This results in 256 samples for every sample uptake, or over 1000 samples with only four sample uptakes. The basics are the same in solid pin and split pin printing. The robot dips the pins into the samples, the pins deposit the sample on a substrate by making contact, the pins re-dip when necessary, and the pins are washed before the next sample. Solid pins must re-dip after every deposit on a substrate; split pins act more like a quill and fill up a reservoir, of which only a fraction is deposited every time the pins make contact.

The number of spots that can be made after sample uptake by split pins varies greatly depending on a number of factors. The viscosity of the sample, the substrate type due to varying amounts of adsorption, the humidity, the pin design, and whether or not it has a reservoir are all factors and must be tested for empirically. When making DNA arrays, the samples are more homogeneous; they have similar viscosities, similar concentrations (moles of DNA), and similar molecule substrate interactions. When dealing with protein microarrays these characteristics are seldom the same. Proteins are usually dialyzed after purification with whichever buffer will keep the protein in solution. This can very widely from protein to protein. This will contribute to potential viscosity differences. Protein concentrations can vary greatly depending on the success of expression and subsequent purification of the protein. In case the concentrations are similar, proteins vary in size and there will be a heterogeneous distribution of the number of molecules per spot. The amino acid sequence of the protein(s) will also influence how it interacts with the substrate on which it is being spotted. In addition, the isoelectric point of the proteins will undoubtedly have an effect on how they bind to the substrate, and furthermore, hydrophilic proteins will behave differently than hydrophobic proteins. There is also the problem of protein aggregation, or too high a concentration, clogging a split pin. There are various-sized split pins with chan-

nels up to 400 μm across which would accommodate most substances. As attempts to make higher-density arrays to increase throughput take place, this option become less feasible.

When the substrate to be arrayed is not compatible with the split pins necessary to meet the array density required, solid pins must be employed. The need for solid pins is more of an issue when arraying cell lysates, cells of bacterial culture fractions. Solid pins are a better choice for these situations, but at the expense of throughput. Even the fastest solid pin machine, the Aushon 2470, requires twice the number of pins to outperform the fastest split pin systems, such as the Omnigrid line of microarray printers manufactured by Genomic Solutions. Throughput may or may not be an issue, but when dealing with proteins, most researchers need the process to take as little time as possible.

## GENOME–PROTEOME CONTENT CURRENTLY USED FOR BIOMARKER RESEARCH

There is a need for an expressible open reading frame (ORF) eome that can be made into an arrayable proteome in the search for disease biomarkers. The number of proteins that are available for screening limits disease biomarker discovery. For example, in the search for human autoantigens, the availability of proteins to use in screening studies has been a bottleneck. The National Institutes of Health (NIH) has a collection of approximately 29,636 human ORFs that contain 17,526 (source: http://mgc.nci.nih.gov/) nonredundant gene sequences. Although Invitrogen has cloned all the human ORFs into their Gateway system, the creation of an assayable human proteome has been limited by the protein purification process. After almost a decade, the number of human proteins contained on their commercial Human ProtoArray chip remains at 8000 (<30% coverage). Recent work has shown that that these ProtoArray human protein chips, although limited in scope, have been used successfully to identify serodiagnostic autoantigens (cancer biomarkers). Hudson et al. profiled serum samples from 30 cancer patients and 30 healthy persons using microarrays containing 5005 human proteins. Ninety-four discriminatory serodiagnostic antigens were identified that exhibited enhanced reactivity from sera in cancer patients relative to control sera [7]. This is a good proof-of-principle experiment that should be expanded to verify that the 30 serum samples used for this experiment are actually representative of the ovarian cancer population as a whole, not just markers for families that are predisposed to this type of cancer but may never develop the disease. An additional control would have included serum samples from the sisters of the patients who did not have ovarian cancer.

To assay even a small population with an appropriate number of controls is financially prohibitive using the commercially available human ProtoArray. The array is available for $3000 each. For this reason the Human ProtoArray

chip offered by Invitrogen has not been utilized for large-scale screening. Small-scale experiments may be fine for a publication, but to identify robust biomarkers there needs to be a massive screening that allows profiling as many healthy, predisposed, and diseased persons as possible so that the data derived from the experiment are biologically relevant. It is only at this scale that we can truly make any conclusions about the immune response. The reason that we still have not found the ideal biomarkers is due to the complexity of immune response and its variability from person to person. The current human protein chip is being distributed as a consumable with a sizable price tag. To attempt a study of the size required to determine disease-state discriminatory antigens, or groups thereof, there needs to be a higher-throughput, more cost-effective method of generating the human proteome chip.

Infectious disease biomarker discovery faces challenges as well. Purified proteins of many infectious agents are scarce. Purified protein microarrays of pathogens usually include only a handful of proteins. The lack of content has been the major reason for the slow rate of emergence of new serodiagnostic markers. For example, only 149 out of the 4198 ORFs (3.5%) constituting the *Yersinia pestis* proteome were used to profile the antibody response to live vaccine [8]. In addition, 156 out of approximately 1000 (15.6%) *Chlamydia trachomatis* proteins were used to profile the human humoral immune response to urogenital tract infections in human subjects [9]. The ability to express and purify recombinant proteins in the lab limits the amount of unique content available for serological screening. This low percentage of proteome coverage is common, and most laboratories in this field are studying a few antigens at a time from thousands of possible candidates.

One of the innovative ways is which researchers have overcome the lack of coverage is to apply the lessons learned from whole-cell Western blots and two-dimensional gels to microarray research. Native protein microarrays are heterogeneous protein pools that have been fractionated using chromatography. Sartain et al. developed a technique for the separation of native *Mycobacterium tuberculosis* cytosol and culture filtrate proteins that resulted in 960 unique protein fractions that were used to generate protein microarrays. These 960 fractions represented all the expressed proteins, having some proteins represented in different fractions. When these microarrays were used to profile the reactivity of different disease states, various previously characterized proteins as well as some novel proteins were identified from these fractions using mass spectrometry [10]. A similar approach was utilized in a different field to profile the sera of prostate cancer patients and controls. Two-dimensional liquid chromatography was used to separate proteins from the prostate cancer cell line LNCaP into 1760 fractions, which were subsequently spotted onto a microarray. The microarrays were probed with serum samples from 25 men with prostate cancer and 25 male controls. Statistical analysis revealed that 38 of the fractions showed significantly more reactivity in the cancer group. Samples were classified with up to 98% accuracy using the same sera [11]. Although a good academic exercise, the protein fraction microarray

needs further studies to uncover the proteins responsible for the reactivity. In the tuberculosis case, mass spectrometry was able to characterize a number of reactive proteins, but the authors conceded that this approach was a stopgap until the complete proteome is cloned, expressed, and arrayed. In the case of prostate cancer the reactive fractions must be analyzed further to determine which of the many proteins represented in the fractions are the important ones. Another challenge that faces users of protein fraction microarrays is that proteins are not all represented equally in these fractions. Antigens that potentially are extremely good biomarkers may be underrepresented, due to an expression level regulated by cell degradation pathways that may target these proteins.

## HIGH-THROUGHPUT GENE CLONING AND EXPRESSION FOR THE FABRICATION OF PROTEOME MICROARRAYS

The genomics era ushered in a new type of research. For the first time in history the decoded "blueprint of life" was available for study. The Human Genome Project (HGP) was the first major undertaking of its kind. The HGP was a 13-year project coordinated by the U.S. Department of Energy and the National Institutes of Health that was completed in 2003. The goal was to identify all of the approximately 20,000 to 30,000 genes in human DNA, determine the sequences of the 3 billion chemical base pairs that make up human DNA, and store this information in databases. This initiative sparked the boom in genome sequencing, and today there is a vast amount of publicly available data stored in numerous databases. The genomes of more than 180 organisms have been sequenced since 1995, and the data from these projects are available to the scientific community. The availability of genome sequence data leads to the first big boom in microarray research. Researchers used this sequence to create libraries of oligo nucleotides representing all the ORFs of the organism being studied. Robotic machines were then used to transfer and arrange nanoliter amounts of thousands of gene sequences representing a cell expression state on a single microscope slide. These high-density, highly organized arrays of DNA are called *microarrays*. DNA microarrays allowed for the miniaturization and high-level multiplexing of Southern blots. Affymetrix and Agilent technologies pioneered the use of DNA microarrays for full genome gene expression analysis. Researchers were now able to track the expression of genes in response to various stimuli and incorporated this new tool in their existing research. They could track the synthesis and degradation at the mRNA level, to determine which genes are turned on and which are turned off in a given cell. We now knew the number and sequence of the expressible genes and the expression pattern of the genes in response to a particular stimulus. An area that still needs to be explored, however, is how the proteins performing their function once expressed. This conundrum has ushered in the age of protein microarrays.

Currently, protein microarrays are being used for a variety of assays. There has been much trial and error when attempting to miniaturize binding assays, enzyme activity assays, and serological assays. The same tools that helped propel the era of the DNA microarray are being applied to protein microarrays. Protein microarrays permit the miniaturization and high-level multiplexing of protein-based assays. Companies like Protometrix cloned, expressed, and purified thousands of proteins to make the Yeast Proteome chip and the Human Partial Proteome chip. The Human chip, now on version 8, is available commercially from Invitrogen, with approximately 8000 human proteins that can be used for screening binding partners, substrates for enzymes, small molecules, and antibody profiling. That is the equivalent of 8000 co-sedimentation assays, 8000 phosphorylation assays, or 8000 ELISA assays on a surface area of 3 inches × 1 inch. When all the approximately 29,000 ORFs that have been cloned are eventually expressed and printed on a microarray, all proteins expressed by human cells will be contained on two or three microscope slides and ready for assay.

Traditional multistep cloning methods have been a major stumbling block because of their innate low throughput. Starting first by amplifying the desired insert by polymerase chain reaction (PCR), then digesting it and the vector, followed by ligation, subsequent transformation, plating, colony picking, and verification of clones, the volume of work is the rate-limiting step. Technical problems may be encountered in any of the aforementioned steps, and most often than not, researchers spend months cloning a few genes of interest. There are, however, ways to circumvent this laborious cloning methodology when attempting to manufacture a high-content microarray.

One way is to take advantage of the vast genomic sequence data and the high-throughput friendly PCR reaction to amplify out all of the ORFs, followed by a second PCR step to make the products transcriptionally active. The secondary PCR step adds a promoter region at the 5′ end and a stop codon at the 3′ end of the PCR product, making a transcriptionally active PCR (TAP) product. The production of TAP fragments for expression can yield a large library of expressible ORFs in a relatively short amount of time. TAP fragments have been reported to express very well in a variety of expression systems [12]. Organisms with large genomes that would otherwise take decades to clone into an expressible form will now only take two rounds of PCR. A recent report by Regis et al. has shown that in the case of *Plasmodium falciparum*, TAP fragments could be used for a large-scale screening of the humoral immune response and eventual selection of antigens for further study [13]. While providing the throughput, TAP fragments are limited by the amount of product produced during the second PCR step.

Recent research has shown that serological screening of cDNA libraries created from cancer cell lines can be used for the discovery of cancer biomarkers. Cancer is the result of a synergistic malfunctioning of multiple signaling pathways. This may or may not be the same set of pathways, proteins, and mutations in every cancer type and tissue, so it becomes even more critical to

cast the widest net to find the most potential biomarkers [14]. Lung cancer cell lines were used to make cDNA libraries that have been used successfully for serological identification of serodominant antigens [16,17]. This is encouraging, but the task at hand with this technology is more daunting. After the creation of the reverse complement of the mRNA, there is still a need to use brute-force cloning techniques to accomplish the task. One must keep in mind that the majority of mRNAs in most tissues are in low abundance, and thus quite a few of the mRNAs will be underrepresented in the cDNA library [15,18]. Extremely large number of clones are required to ensure a good representation of these low-abundance genes. There is also the very real problem of partial cDNA clones, which do not have the complete sequence. If we are fortunate enough to minimize these two issues, there is still the very real bottleneck of screening the clones. High-density gridding is becoming more and more common and makes it possible to screen large amounts of mRNA complementary DNA. This technology is designed largely for gene discovery, but if refined to the point that all mRNA transcripts are represented equally, we may have a tool for more reliable screening of human serum samples.

The next rung on the ladder would be the creation of an expressible genome library that contains all the ORFs in the genome. To make the genome library quickly and efficiently, we need to consolidate all the traditional cloning steps into one efficient step. Recombination in vivo provides the all-encompassing single step. Recombination cloning is widely used in research, but for a long time it was thought that bacteria lacked the mechanism to allow for recombination. However, bacterial in vivo homologous recombination has been an efficient and heavily used tool in the genetic field. Recombination-based cloning allows DNA sequences to be inserted or deleted without regard to location of restriction sites [19,20]. One very widely used methodology for recombination cloning is the Gateway system now offered commercially by Invitrogen. The first step to Gateway cloning is inserting the gene of interest into the Gateway entry vector. There are two ways to clone your gene of interest into a gateway entry clone. The first is the standard cut-and-paste protocol using restriction enzymes and ligase. The second way is to create a PCR product with terminal attB sites, using primers containing a 25-base pair attB sequence, plus four terminal G's. This product will then be inserted in the entry clone using reconstituted recombination machinery. PCR-based inserts can be made using genomic DNA, a cDNA library, or an plasmid clone containing your gene of interest. An entry clone is a vector containing your gene of interest and flanked by Gateway *att* recombination sites. Bacteriophage lambda *att* site recombination is a well-characterized phenomenon [21]. Bacteria have a stretch of DNA sequence called *att* encoded into their genome, and bacteria phage have the same stretch of DNA sequence. When the phage infects a bacterium, the lambda DNA injected recombines with the corresponding bacterial DNA via the *att* sites in the presence of integration-specific enzymes. The enzyme-assisted recombination results in integration of the phage DNA into the bacterial genome. Using the same reconstituted lambda

*att* site recombination system, entry clones containing a gene of interest can then be transferred into any expression vector for subsequent protein expression. Although the Gateway System works and increases cloning throughput dramatically, it is not efficient enough to work as a true high-throughput alternative.

In vivo recombination cloning directly into an optimized expressible vector, however, appears to have solved the throughput dilemma. Recent advances in cloning methodologies have resulted in the ability to clone entire bacterial ORFeomes, comprising thousands of genes, in a relatively short time. Figure 2 illustrates the typical results one would see beginning with the PCR step followed by the in vivo recombination cloning, and checking the clones by PCR using sequence specific primers. Specifically, in vivo recombination cloning [22] in *Erscherichia coli* has broken the barriers posed by traditional brute-force cloning methodologies that rely on multiple-step cloning techniques: from PCR, restriction enzyme digestion of the PCR product and vector, ligation, to single colony screen and selection. Available sequence data are used to design 5′ and 3′ gene-specific primers for all ORFs encoded in the genome. The primers contain 53 nucleotides, which are comprised of a 33-nucleotide recombination adapter sequence and a 20-nucleotidegene specific sequence. The PCR products are then mixed with a T7-based linear expression vector as described previously [22] and transformed into supercompetent DH5a cells (Antigen Discovery Inc.). The cells transformed are grown overnight and are checked for turbidity the following day. DNA can then be purified from the mixed cultures using 96-well plate–based mini-prep protocols such as Qiaprep Turbo 96.

The plasmid DNA purified from overnight culture can then be used to express the protein encoded by the insert. Proteins are expressed in a coupled in vitro transcription-translation (IVT) reaction in an *E. coli*–based cell-free

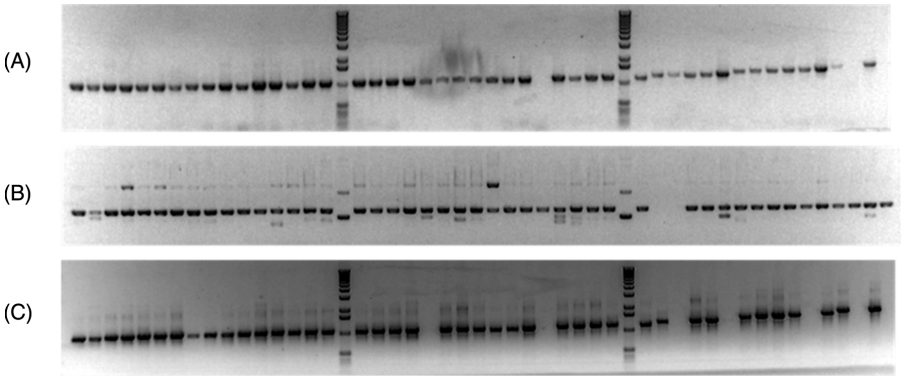


**Figure 2** Synthesis and verification of clones. Representative images of whole Vv ORFeome PCR (A), cloning (B), and QC-PCR (C).

expression system such as RTS 100 from Roche. The unpurified proteins can then be printed directly to nitrocellulose-coated slides along with a set of negative and positive controls using a contact microarray printer. It is crucial that the mixture of unpurified proteins be spotted as quickly as possible. The quality of the arrays is inversely proportional to the amount of time from the end of the reaction to the deposition onto the nitrocellulose. Once the unpurified protein mixture is spotted and dry, it is very stable.

In the high-throughput microarray fabrication first reported by Davies et al. in 2005 [22], the cell-free expressed proteins can be detected using antibodies against the N-terminal polypolyhistidine (polyHis) tag and the C-terminal hemagglutinin (HA) tag directly on the chip. The antibodies were used to monitor expression of the large numbers of parallel reactions. The arrays probed using a mouse monoclonal antibody raised against the polyHis epitope and rat monoclonal antibody raised against the HA epitope are visualized using a fluorescent-based microarray scanner. An example of one such hybridization and scanning output can be seen in Figure 3. The data can then be quantified using a microarray data analysis software package that can quantify the intensity of the spots on the microarray chip. Each array contains positive control spots printed from serial dilutions of whole immunoglobulin G (IgG). Each array also contains no DNA negative control spots, and the reactivity of

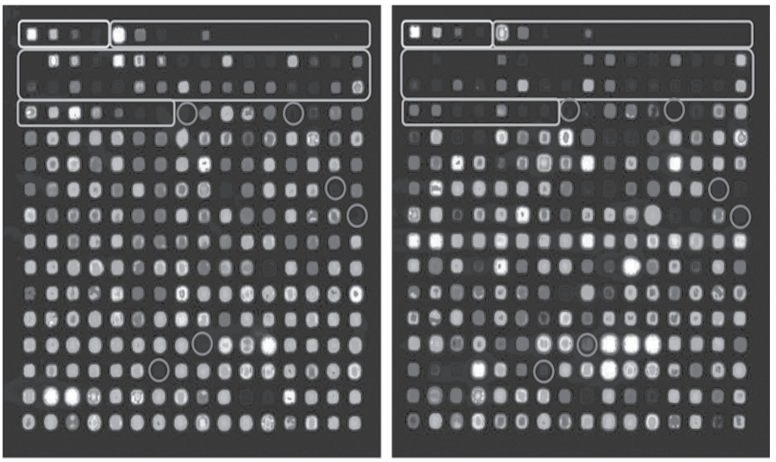

(A) (B)

**Figure 3** Quality assessment for protein expression. Images of whole Vv proteome microarray probed with mAb against polyhistidine (A) and hemoagglutinin tags (B). 99.0% of Vv proteins were reactive to antipolyhistidine antibody and 88.2% to anti-hemoagglutinin antibody. The images show four dilutions of human IgG (yellow box), positive controls (green box), and six negative controls (mock transcription/translation reaction) (red circles). The remaining spots are Vv proteins. (*See insert for color reproduction of the figure.*)

these spots is low for both serum samples. The positive and negative controls are used to normalize the data using a modified VSN package in a statistical environment named R [23,24] from arrays that are probed on different days. There are also serially diluted EBNA1 protein control spots, which have been shown to be reactive to varying degrees in different human subjects [25–27]. Once an array has passed quality assurance, it is ready to be used in serological studies.

## USING WHOLE PROTEOME MICROARRAYS TO SCREEN FOR DISEASE-STATE DISCRIMINATORY ANTIGENS

The ability to find biomarkers relies on the ability to profile the immune response to entire proteomes. Antigens that are highly reactive to infection, or auto-antibodies, can be used in diagnostics and therapeutics. They are also potential candidates for the development of vaccines.

Using the model organism vaccinia virus Western Reserve strain, researchers were able to use in vivo recombination cloning to clone the 185Vv proteins, or greater than 90% of the proteome, from genomic DNA into an expressible vector. This proteome vector library was subsequently expressed and arrayed onto nitrocellulose-coated slides. Each slide contains 16 nitrocellulose pads with a Vv proteome array per pad. Each slide is capable of producing serological data for 16 serum samples interrogating approximately 200 spots (185Vv proteins plus controls), yielding 3200 data points per slide, the equivalent of thirty-three 96-well ELISA plates. Davies et al. [22] showed that there are specific vaccination-state and disease-state profiles that are discernible using whole proteome screening. Comparing the pre- and postvaccination data for each species, naive mice, monkeys, and humans have a very different reactivity pattern than that of vaccinated mice, monkey, and humans. In Figure 4 we see one such example comparing the hyperimmune serum vaccinia immunoglobulin (panel A) from Cangene and a naive subject (panel B). These results verified that this platform was a rapid way to scan humoral immunity comprehensively from vaccinated or infected humans and animals on a whole proteome scale [22]. In 2008, Davies et al. showed the power of the technology for profiling the immune response. Figure 5 is a heat map of an experiment comparing the immune response of two different stains of vaccinia virus. The U.S. government is considering using modified vaccinia virus Ankara (MVA) as a smallpox vaccine to replace live vaccinia virus, otherwise known as Dryvax, because it is safer. The heat map is a global view that allows for comparison of the immune response elicited by MVA and Dryvax and can give us an idea of whether it elicits a similar response to confer protection.

Barbour et al. recently published an article that takes advantage of this technology to clone and express 1292 *Borrelia burgdorferi* proteins. Employing enzymeless recombination cloning and *E. coli* cell-free expression, they produced a protein array representing approximately 80% of the genome. The

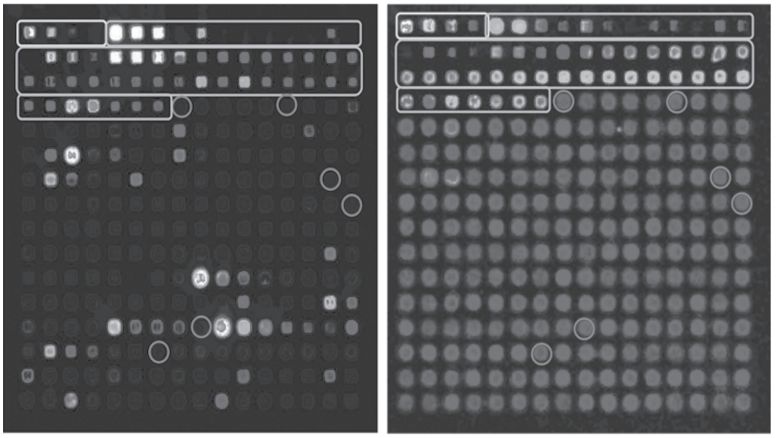

(A) (B)

**Figure 4** Naive subject serum and VIG. Images of whole vaccinia virus proteome microarray probed with vaccinia immunoglobulin (Cangene) (A) and naive individual (B). The IgG control spots (shown in yellow) show signal intensities for both samples; both objects do not react with the no-DNA control spots (shown in red). The images show four dilutions of human IgG (yellow box), positive controls (green box), and six negative controls (mock transcription/translation reaction) (red circles). The remaining spots are Vv proteins. (*See insert for color reproduction of the figure.*)

antibody reactivities of sera from patients' Lyme disease were compared to the antibody reactivities of sera from controls. Overall, approximately 15% of the (ORFs) of *B. burgdorferi* in the array elicited an antibody response in humans with natural infections. Among the immunogens, 103 stood out on the basis of statistical criteria [26].

In 2007 this same platform technology allowed fabrication of the *Francisella tularensis* proteome. The proteome was interrogated with serum samples to find biomarkers that are specific to tularemia, the disease caused by *F. tularensis*. This particular study was looking at the efficacy of subunit vaccines, and the proteome arrays were used to take a look at the humoral response to different vaccination protocols. The stimulation of a protective immune response against intracellular pathogens using nonreplicating vaccines was investigated. Many of today's vaccines still use attenuated organism vaccines or a different less virulent strain as a vaccine. Eyles et al. showed the effect of different adjuvants on the response to the nonreplicating vaccine. The *F. tularensis* proteome chips were used to effectively differentiate which adjuvant was stimulating an IgG1 or IgG2 response. An IgG1 response is related to the TH2 receptor and does not confer protection. An IgG2 response is associated with the TH1 receptors and has been shown to be associated with protective immunity. Furthermore, these data give insight into the protective immune response

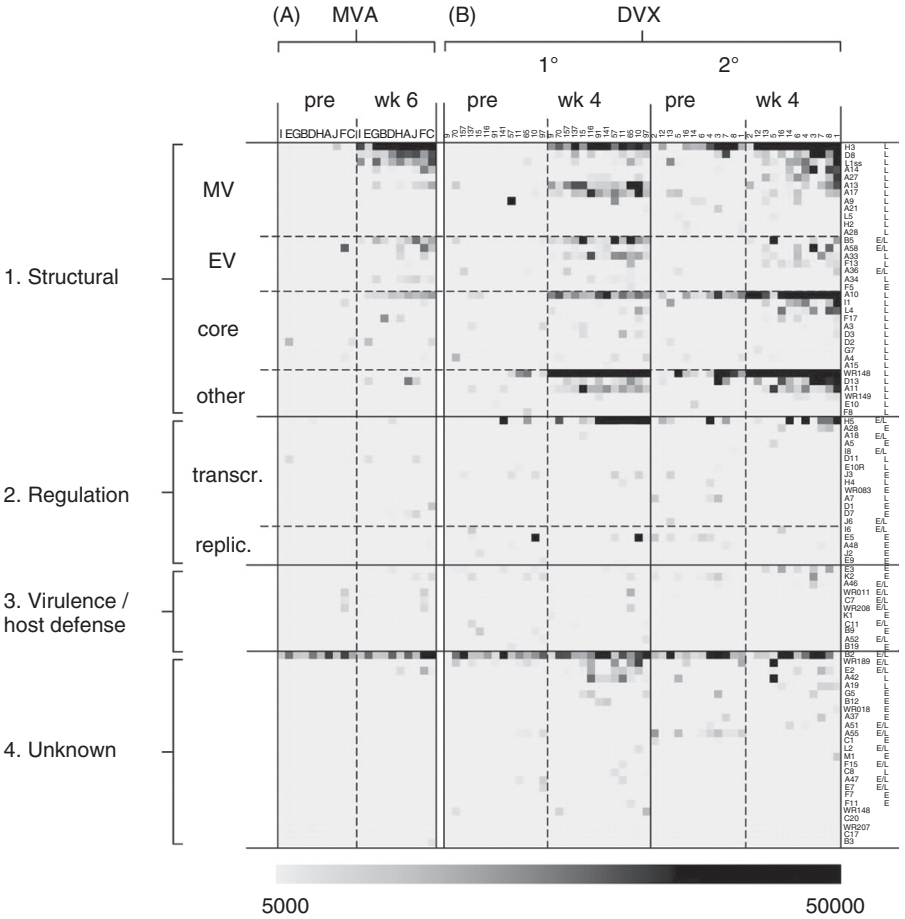


**Figure 5** Antibody profiling of human antibody profiles for pre- and postvaccination with MVA (A) and WR (B). For Dryvax responses, primary ($n$ = 13) and secondary ($n$ = 12) infections are shown. (From ref. 25.) (*See insert for color reproduction of the figure.*)

and have potentially important implications for the rational design of nonliving vaccines for tularemia and other intracellular pathogens.

Recent developments in the fabrication of protein microarrays allow the multiplexing of the entire proteome in profiling the immune response. This approach is powerful enough to discover content to finally replace tests used currently to diagnose infectious disease, some of which were developed over a century ago. The diagnosis of tuberculosis (TB) is an excellent example of such an opportunity. TB is a major cause of illness and death worldwide. Globally, 9.2 million new cases and 1.7 million deaths from TB occurred in 2006 [28]. Multiple tests must be performed to diagnose TB. A person sus-

pected of having TB is first tested using the tuberculin skin test (TST), also known as the purified protein derivative (PPD) test, pioneered by French physician Charles Mantoux in 1907. The TST is used routinely in humans, but the results are subjective, variable, and their interpretation is rarely consistent [29]. Sputum testing identifies *M. tuberculosis* directly, providing specificity, but current methodologies do not give good sensitivity [30,31]. The gold standard for the diagnosis of TB is a bacterial culture, which is a time-consuming process that requires a dedicated microbiology lab and several weeks to obtain the results [32]. More novel and sophisticated methods such as real-time PCR (RT-PCR) and cell-mediated-immunity–based detection assays, such as interferon gamma release assays (IGRAs), are accurate but require sophisticated equipment and highly trained personnel [33,34]. The limitations of the current TB diagnosis techniques have sparked renewed interest in antibody detection–based tests, which are relatively straightforward. The immune response to TB infection is variable enough that an infected population may not recognize a single antigen [35–40]. Methodology using previously characterized *M. tuberculosis* proteins using immunoassays such as ELISAs and microbead suspension assays for the purpose of TB diagnosis have shown the need of a multiplex approach at diagnosing TB [41,42]. There is a dire need for the discovery of an optimal panel of antigens that will allow accurate and reliable diagnosis.

Cancer is another field that is in urgent need of diagnostics with improved accuracy. We have known that cancer cells produce auto-antibodies since the 1950s [43]. Researchers found that these auto-antibodies were against proteins that were associated with these cells. Auto-antibodies against the tumor suppressor p53 have been found in breast cancer, lung cancer, and ovarian cancer [44–47]. Proteins that are overexpressed in a cancer cells also elicit the production of auto-antibodies. For example, the overexpression of c-Myc is linked to the presence of auto-antibodies against c-Myc in lung cancer [48]. There is evidence that the immune system begins to produce auto-antibodies against cancer cell–associated proteins years, perhaps even decades, before the onset of disease [49–51].

There is a dire need for earlier and more accurate cancer detection. The development of "early warning" tests will ensure a better prognosis for the patient. Serological biomarkers can also be a tool for the development of antibody-based therapeutics to treat cancer. To achieve this objective, it is important to identify sets of cancer-specific antigens, or biomarkers, that can differentiate disease states as well as different forms of cancer. Serum auto-antibody profiling is a promising approach for early detection and diagnosis of breast cancer. It is becoming more and more clear that it will take more than one marker to achieve the early warning system that cancer clinicians so desperately seek, and it is likely that well-characterized panels of biomarkers will be required. Zong et al. recently tested a breast cancer T7 phage library with healthy control sera and breast cancer sera and concluded that a panel of auto-antibodies, not a single auto-antibody, is required for optimal accuracy of disease-state discrimination [49]. A similar study tested colon cancer sera

with a colon cancer T7 phage library. Ran et al. were able to identify 24 autoantigens associated with colon cancer. A six-marker (six antigens) achieved 91.7% sensitivity and 91.7% specificity in the training set of sera. In a testing set with this marker panel, they predicted 85% of the samples correctly [52].

There are approximately 30,000 ORFs in the NIH human gene collection, representing 17,500 nonredundant genes and variants there of [53]. Using enzymeless recombination cloning, a human proteome chip could be made with upward of 90% of the ORFs represented. With the ability to assay 27,000 proteins at once, the discovery of new biomarkers for human disease will be limited only by the number of serum samples used to determine disease-state discriminatory sets of antigens, and potentially, antigen panels that differentiate between manifestations of the same disease, such as cancer.

## SUMMARY

The high-throughput proteomics approach described here is the natural progression of the genome sequencing projects of the 1990s, but instead of sequencing the genome, we are fabricating the proteomes. Organisms such as *M. tuberculosis* with 4000 ORFs, or malaria (*P. falciparum* and *P. vivax*) with even larger genomes, can now be assayed on a whole-proteome scale. By providing a comprehensive tool to test all the proteins at once, we open the door to a new nonbiased approach in biomarker discovery. Instead of working with a handful of proteins and attaining very limited data sets, we can now look at these organisms from a whole-proteome perspective, allowing the immune system to tell us what is and is not important. Although this approach has been used primarily for infectious disease research, it also holds promise for the discovery and study of biomarkers for human disease.

## REFERENCES

1. Redkar RJ, Schultz NA, Scheumann V, et al. (2006). Signal and sensitivity enhancement through optical interference coating for DNA and protein microarray applications. *J Biomol Tech*, 17(2):122–130.
2. Angenendt P, Glökler J, Sobek J, Lehrach H, Cahill DJ (2003). Next generation of protein microarray support materials: evaluation for protein and antibody microarray applications. *J Chromatogr A*, 1009(1–2):97–104.
3. Chiari M, Cretich M, Corti A, Damin F, Pirri G, Longhi R (2005). Peptide microarrays for the characterization of antigenic regions of human chromogranin *A*. *Proteomics*, 5(14):3600–3603.
4. Brueggemeier SB, Wu D, Kron SJ, Palecek SP (2005). Protein–acrylamide copolymer hydrogels for array-based detection of tyrosine kinase activity from cell lysates. *Biomacromolecules*, 6(5):2765–2675.
5. Cretich M, Pirri G, Damin F, Solinas I, Chiari M (2004). A new polymeric coating for protein microarrays. *Anal Biochem*, 332(1):67–74.

6. Brueggemeier SB, Kron SJ, Palecek SP (2004). Use of protein–acrylamide copolymer hydrogels for measuring protein concentration and activity. *Anal Biochem*, 329(2):180–189.
7. Hudson ME, Pozdnyakova I, Haines K, Mor G, Snyder M (2007). Identification of differentially expressed proteins in ovarian cancer using high-density protein microarrays. *Proc Natl Acad Sci USA*, 104(44):17494–17499.
8. Li B, Jiang L, Song Q, et al. (2005). Protein microarray for profiling antibody responses to Yersinia pestis live vaccine. *Infect Immun*, 73(6):3734–3739.
9. Sharma J, Zhong Y, Dong F, Piper JM, Wang G, Zhong G (2006). Profiling of human antibody responses to *Chlamydia trachomatis* urogenital tract infection using microplates arrayed with 156 chlamydial fusion proteins. *Infect Immun*, 74(3):1490–1499.
10. Sartain MJ, Slayden RA, Singh KK, Laal S, Belisle JT (2006). Disease state differentiation and identification of tuberculosis biomarkers via native antigen array profiling. *Mol Cell Proteom*, 5(11):2102–2113.
11. Bouwman K, Qiu J, Zhou H, et al. (2003). Microarrays of tumor cell derived proteins uncover a distinct pattern of prostate cancer serum immunoreactivity. *Proteomics*, 3(11):2200–2207.
12. Liang X, Teng A, Braun DM, et al. (2002). Transcriptionally active polymerase chain reaction (TAP): high throughput gene expression using genome sequence data. *J Biol Chem*, 277(5):3593–3598.
13. Regis DP, Dobaño C, Quiñones-Olson P, et al. (2008). Transcriptionally active PCR for antigen identification and vaccine development: in vitro genome-wide screening and in vivo immunogenicity. *Mol Biochem Parasitol*, 158(1):32–45.
14. McCarrey JR, Williams SA (1994). Construction of cDNA libraries from limiting amounts of material. *Curr Opin Biotechnol*, 5(1):34–39.
15. Suzuki Y, Sugano S (2001). Construction of full-length-enriched cDNA libraries: the oligo-capping method. *Methods Mol Biol*, 175:143–153.
16. Ali Eldib AM, Ono T, Shimono M, et al. (2004). Immunoscreening of a cDNA library from a lung cancer cell line using autologous patient serum: identification of XAGE-1b as a dominant antigen and its immunogenicity in lung adenocarcinoma. *Int J Cancer*, 108(4):558–563.
17. Li HH, Ma LJ, Wang QQ (2004). [Construction and characterization of cDNA library for lung cancer cell line YTMLC-90 isolated from Gejiu, Yunnan province]. *Xi Bao Yu Fen Zi Mian Yi Xue Za Zhi*, 20(4):465–468.
18. Fulton LL, Hillier LD, Wilson RK (1995). Large-scale complementary DNA sequencing methods. *Methods Cell Biol*, 48:571–582.
19. Thomason L, Court DL, Bubunenko M, et al. (2007). Recombineering: genetic engineering in bacteria using homologous recombination. *Curr Protoc Mol Biol*, (1):16.
20. Wendland J (2003). PCR-based methods facilitate targeted gene manipulations and cloning procedures. *Curr Genet*, 44(3):115–123.
21. Groth AC, Calos MP (2004). Phage integrases: biology and applications. *J Mol Biol*, 335(3):667–678.
22. Davies DH, Liang X, Hernandez JE, et al. (2005). Profiling the humoral immune response to infection by using proteome microarrays: high-throughput

vaccine and diagnostic antigen discovery. *Proc Natl Acad Sci USA*, 102(3): 547–552.

23. Sundaresh S, Doolan DL, Hirst S, et al. (2006). Identification of humoral immune responses in protein microarrays using DNA microarray data analysis techniques. *Bioinformatics*, 22(14):1760–1766.
24. Sundaresh S, Randall A, Unal B, et al. (2007). From protein microarrays to diagnostic antigen discovery: a study of the pathogen *Francisella tularensis*. *Bioinformatics*, 23(13):i508–i518.
25. Davies DH, Wyatt LS, Newman FK, et al. (2008). Antibody profiling by proteome microarray reveals the immunogenicity of the attenuated smallpox vaccine modified vaccinia virus ankara is comparable to that of Dryvax. *J Virol*, 82(2): 652–663.
26. Barbour AG, Jasinskas A, Kayala MA, et al. (2008). A genome-wide proteome array reveals a limited set of immunogens in natural infections of humans and white-footed mice with *Borrelia burgdorferi*. *Infect Immun*, 76(8):3374–3389.
27. Eyles JE, et al., (2007). Immunodominant *Francisella tularensis* antigens identified using proteome microarray. Crown Copyright 2007 Dstl. *Proteomics*, 7(13): 2172–2183.
28. WHO (2008).*Global Tuberculosis Control: Surveillance, Planning, Financing*. WHO, Geneva, Switzerland.
29. Eyles JE, Unal B, Hartley MG, et al. (2007). A systematic review of rapid diagnostic tests for the detection of tuberculosis infection. *Health Technol Assess*, 11(3):1–196.
30. Gilpin C, Kim SJ, Lumb R, Rieder HL, Van Deun A (2007). Critical appraisal of current recommendations and practices for tuberculosis sputum smear microscopy. *Int J Tuberc Lung Dis*, 11(9):946–952.
31. Steingart KR, Henry M, Ng V, et al. (2006). Fluorescence versus conventional sputum smear microscopy for tuberculosis: a systematic review. *Lancet Infect Dis*, 6(9):570–581.
32. Baylan O (2005). [Culture based diagnostic methods for tuberculosis]. *Mikrobiyol Bul*, 39(1):107–124.
33. Parashar D, Chauhan DS, Sharma VD, Katoch VM (2006). Applications of real-time PCR technology to mycobacterial research. *Indian J Med Res*, 124(4): 385–398.
34. Hoffmann H, Loytved G, Bodmer T (2007). [Interferon-gamma release assays in tuberculosis diagnostics]. *Internist (Berl)*, 48(5):497–498, 500–506.
35. Steingart KR, Ramsay A, Pai M (2007). Commercial serological tests for the diagnosis of tuberculosis: do they work? *Future Microbiol*, 2:355–359.
36. Steingart KR, Henry M, Laal S, et al. (2007). Commercial serological antibody detection tests for the diagnosis of pulmonary tuberculosis: a systematic review. *PLoS Med*, 4(6):e202.
37. Bothamley GH (1995). Serological diagnosis of tuberculosis. *Eur Respir J Suppl*, 20:676s–688s.
38. Raja A, Ranganathan UD, Bethunaickan R (2008). Improved diagnosis of pulmonary tuberculosis by detection of antibodies against multiple *Mycobacterium tuberculosis* antigens. *Diagn Microbiol Infect Dis*, 60(4):361–368.

39. Wang EL, Liu WT, Li T (2006). [The serodiagnostic value of antigens secreted from *Mycobacterium tuberculosis*]. *Zhonghua Jie He He Hu Xi Za Zhi*, 29(7): 466–469.
40. Perkins MD, Conde MB, Martins M, Kritski AL (2003). Serologic diagnosis of tuberculosis using a simple commercial multiantigen assay. *Chest*, 123(1):107–112.
41. Khan IH, Ravindran R, Yee J, et al. (2008). Profiling antibodies to *Mycobacterium tuberculosis* by multiplex microbead suspension arrays for serodiagnosis of tuberculosis. *Clin Vaccine Immunol*, 15(3):433–438.
42. Verma RK, Jain A (2007). Antibodies to mycobacterial antigens for diagnosis of tuberculosis. *FEMS Immunol Med Microbiol*, 51(3):453–461.
43. Giraud G, Latour H, Levy A, Puech P, Roujon J (1955). [Cancer of pancreas with auto-antibodies in hemolytic anemia]. *Montpellier Med*, 48(3):465–467.
44. Bergqvist M, Brattström D, Larsson A, et al. (1998). P53 auto-antibodies in non-small cell lung cancer patients can predict increased life expectancy after radiotherapy. *Anticancer Res*, 18(3B):1999–2002.
45. Angelopoulou K, Diamandis EP (1997). Detection of the TP53 tumour suppressor gene product and p53 auto-antibodies in the ascites of women with ovarian cancer. *Eur J Cancer*, 33(1):115–121.
46. Regidor PA, Regidor M, Callies R, et al. (1996). Detection of p53 auto-antibodies in the sera of breast cancer patients with a new recurrence using an ELISA assay: Does a correlation with the recurrence free interval exist? *Eur J Gynaecol Oncol*, 17(3):192–199.
47. Green JA, Mudenda B, Jenkins J, et al. (1994). Serum p53 auto-antibodies: incidence in familial breast cancer. *Eur J Cancer*, 30A(5):580–584.
48. Yamamoto A, Shimizu E, Sumitomo K, et al. (1997). *L-Myc* overexpression and detection of auto-antibodies against *L-Myc* in both the serum and pleural effusion from a patient with non-small cell lung cancer. *Intern Med*, 36(10):724–727.
49. Zhong L, Ge K, Zu JC, et al. (2008). Autoantibodies as potential biomarkers for breast cancer. *Breast Cancer Res*, 10(3):R40.
50. Chapman CJ, Murray A, McElveen JE, et al. (2008). Autoantibodies in lung cancer: possibilities for early detection and subsequent cure. *Thorax*, 63(3):228–233.
51. Chapman C, Murray A, Chakrabarti J, et al. (2007). Autoantibodies in breast cancer: their use as an aid to early diagnosis. *Ann Oncol*, 18(5):868–873.
52. Ran Y, Hu H, Zhou Z, et al. (2008). Profiling tumor-associated autoantibodies for the detection of colon cancer. *Clin Cancer Res*, 14(9):2696–2700.
53. Zabarovskii ER, Allikmets RL (1989). [Construction of libraries of structural genes (cDNA) and "jumping" gene libraries in lambda vectors]. *Mol Biol (Mosk)*, 23(5):1205–1220.

# CHARACTERIZATION AND VALIDATION BIOMARKERS IN DRUG DEVELOPMENT: REGULATORY PERSPECTIVE

Federico Goodsaid, Ph.D.
*Center for Drug Evaluation and Research, U.S. Food and Drug Administration, Silver Spring, Maryland*

## INTRODUCTION

Biomarkers can be considered tools to be qualified in close association with individual submissions for drug approvals. The U.S. Food and Drug Administration (FDA) Drug-Diagnostic Co-development Concept Paper [1] described a specific example of this association, where the approval of a drug and a test are closely linked with each other in both their product concepts and the timelines of their regulatory approvals. Several examples of biomarkers approved through this co-development process are shown in the FDA's Table of Valid Genomic Biomarkers in the Context of Approved Drug Labels [2]. Biomarkers in this table include both genetic and translational entries. Genetic biomarkers are often integrated in drug development as clinical or nonclinical markers of drug efficacy or safety for the purpose of patient selection in clinical trials, response prediction through stratification or enrichment, or dose optimization. Translational biomarkers have applications similar to those of genetic biomarkers but may also be useful for response monitoring and as early indicators of toxicity or adverse reactions.

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

Novel approaches are being tested continuously for the successful integration of biomarkers in drug development. Their applications range from early compound selection through postmarketing applications. However, integration of these novel biomarkers into routine nonclinical and clinical practice, and regulatory submissions, have often been slow. Hesitation in the application of these tests is often associated with fear not only about how comprehensive data supporting these applications are but also about the regulatory interpretation of the context of use for these applications [3]. Biomarker tests can be integrated into drug development when we have a consensus about the context in which we are measuring the biomarker and the evidence supporting this measurement. These levels of consensus need to be reflected in the regulatory review of biomarker data.

## REGULATORY PATHS IN BIOMARKER EVALUATION AND QUALIFICATION

The path from an exploratory biomarker to a biomarker qualified for a specific application context can be long and unpredictable [4]. Application of these biomarkers requires an objective record for their nonclinical or clinical context and supporting qualification evidence. Information from the development of exploratory biomarkers has been shared between the pharmaceutical industry and the FDA through voluntary exploratory data submissions (VXDSs) [5]. Submissions of exploratory biomarker data have allowed reviewers at the FDA to share with scientists in the pharmaceutical industry study designs, sample isolation and storage protocols, technology platforms, analysis algorithms, biological pathway interpretation, and electronic data submission formats. This experience has been valuable in training our reviewers for the analysis and interpretation of biomarker data.

VXDS have stressed the need for a regulatory path from exploratory biomarkers to biomarkers qualified for a specific context. Such a path is currently being tested at the FDA through a pilot process for biomarker qualification [6]. This process is focused on the specific needs of the regulatory environment to ensure scientifically accurate and clinically (or preclinically) useful decision making. In the first use of this new joint-agency review process put in place by U.S. and European drug regulators, the FDA and the European Medicines Association (EMEA) [7] allow drug companies to submit the results of seven new tests as evidence of nephrotoxicity by new drugs. At this time, the qualification of these biomarkers covers voluntary submission of these data for rat studies. On a case-by-case basis, the FDA will also consider possible application of these biomarkers in phase I human trials. The tests measure levels of seven key proteins or biomarkers that scientists from the FDA and EMEA believe provide important new safety information about the effect of drugs on the kidney. When reviewing investigational new drug (IND) applications, new drug applications (NDAs), or BLAs, both regulatory agencies will now

consider the test results in addition to blood urea nitrogen (BUN) and creatinine.

The development of the new renal toxicity biomarkers was led by the Predictive Safety Testing Consortium (PSTC) [8], whose members include scientists from 16 pharmaceutical companies. The PSTC was organized and led by the Critical Path Institute [9]. Researchers from Merck and Novartis identified the new biomarkers, tested them to prove their sensitivity, specificity, and positive and negative predictive value, and then shared their findings with the other consortium members for further study. The consortium then submitted applications for their qualification to FDA and EMEA.

This is a unique example of how a group of drug companies can work together to propose and generate qualification data for new safety tests and then present them jointly to the FDA and EMEA for qualification. The FDA and EMEA laid the groundwork for such joint-agency reviews in 2004 with the development of the VXDS framework. The VXDS review served as the baseline model around which to design the pilot process for biomarker qualification in 2006. A similar biomarker qualification data submission (BQDS) meeting is held in this pilot process to allow an exchange of questions with the sponsor about scientific and clinical information submitted for qualification.

The pilot process for biomarker qualification allowed the PSTC to submit a single application for biomarker qualification to both regulatory agencies, and then to meet jointly with scientists from both agencies to discuss it in detail and to address additional scientific questions posed by the regulators. Each regulatory agency reviewed the application separately and made independent decisions on whether each would allow the new biomarkers to be used. The new biomarkers qualified by FDA and EMEA are KIM-1, albumin, total protein, $\beta_2$-microglobulin, cystatin C, clusterin, and trefoil factor-3. Testing for these proteins will help scientists assess whether a drug is likely to cause damage to the kidneys, a toxic side effect of some drugs. At this time, both FDA and EMEA require drug companies to submit the results of two other tests, BUN and serum creatinine, to show whether such kidney damage has occurred.

The seven new tests may provide important advantages over these two tests. For example, in the rat model, once kidney damage has begun to occur, it takes a week before the two current tests can detect it [10]. The new tests are more sensitive and can reveal cellular damage within hours [11]. BUN and serum creatinine show that damage has occurred somewhere in the kidneys, but the new tests can also pinpoint which parts of the kidney have been affected [12]. Although additional studies are needed, the new biomarkers may one day allow promising drugs to advance into clinical trials that otherwise would have been abandoned, because currently there are no tests available to detect early-onset renal injury. The seven new tests were developed and will be carried out initially in rats, but they were selected because other

studies have shown that similar biomarkers are produced in human kidney cells [13]. Although initially the FDA and EMEA will consider only data from rat studies, the PSTC will begin work to qualify these biomarkers in human studies. If these studies are successful, the PSTC will present a new application seeking acceptance of the human biomarkers over the next two years.

The need for an accurate, comprehensive, and efficient process for biomarker qualification is closely linked with our ability to quickly integrate new biomarkers in drug development and regulatory review. The biomarker qualification pilot process at the FDA is testing the scientific, clinical, and regulatory components for a biomarker qualification process. Experience gained with this pilot process will be useful in the development of a formal regulatory process for biomarker qualification.

## EVIDENTIARY RECOMMENDATIONS

The most difficult part of this process will be to define incremental contexts of use and the corresponding evidence with which biomarkers may be qualified. We should not misrepresent the industry goal: qualified biomarkers capable of managing new drugs in the clinic with nonclinical or clinical findings of nephrotoxicity somewhere in their previous developmental stages. We also cannot misrepresent the goal as far as public health is concerned, which is to obtain better biomarkers of nephrotoxicity for routine clinical use as quickly as the data will allow. Intermediate qualification contexts and data need to be defined so that investment in biomarker qualification studies will be productive both for clinics and for the pharmaceutical industry. Initial studies proposed by consortia are unlikely to match a clear context for qualification for a full clinical application of biomarkers. What intermediate contexts for qualification can we define, and what study characteristics can we propose for qualification in these intermediate contexts of use? Several authors [14,15] have proposed evidentiary recommendations for biomarker qualification.

Unlike the incremental process for biomarker qualification embodied in the pilot process for biomarker qualification at the FDA [7], papers on evidentiary recommendations often propose all-or-nothing qualification contexts, where if the ultimate goal is a clinical qualification, no intermediate qualification contexts are expected to be defined or qualified. This approach is not only time consuming but is also not likely to encourage the investment needed to generate data for biomarker qualification. At each stage, whether the context of use for a biomarker is to be in vitro, in a nonclinical animal model, or in the clinic, a company or consortium proposing a qualification will probably seek a quick return on the qualification of a biomarker after data are available to qualify a biomarker in a specific context in drug development. An effective process for biomarker qualification should include incremental application

context steps, so that these incremental steps can quickly benefit the drug development process.

## HARMONIZATION

The application of biomarkers to drug discovery and development has the potential of improving the efficacy and speed of bringing more effective and safer new drugs to market. This requires that biomarkers which may be applicable to such uses be qualified for a specific application context. To achieve this, both the process of qualification and the evidentiary criteria and standards for qualification will need to be described and defined. The International Committee on Harmonization's (ICH) E16 [16] is a harmonization effort to define the context, structure, and format of the biomarker qualification submission. It is based on the previous experience by the FDA and EMEA regarding biomarker qualification. This harmonization effort does not address the evidentiary requirements for biomarker qualification.

The structure, format, and content of a submission of biomarker data for qualification depend on the context in which the biomarker is intended to be used. The first step in drafting a submission for qualification of a biomarker is to determine its context of use, preceding specific decisions on applicable structure and format. The context of use for a biomarker is (1) the general area of biomarker application, (2) the specific applications/implementations, and (3) the critical factors that define where a biomarker is to be used and how the information from measurement of this biomarker is to be integrated in drug development and regulatory review. To demonstrate the alignment between proposed context and data, the initial context proposal must be supported by data available at the initial application step or expected to be available throughout the data evaluation process in biomarker qualification. There is a convergent relationship between an initial qualification context and the data supporting it. The initial gap between proposed context and data may need to be filled throughout the qualification process. Initial context proposals, however, should project a significant improvement over currently available biomarkers and/or endpoints.

The context of a biomarker drives data requirements to demonstrate its qualification for the intended application. The structure of a submission document ensures that the context and data can be submitted in a package consistent for consortia submitting qualifications as well as for reviewers in regulatory agencies evaluating a qualification package. The structure of a qualification submission is independent of the context of this submission, but must also be flexible enough to deal with the specific requirements of each context. On the other hand, the format of data required to qualify a biomarker may vary significantly with the context in which it is to be used. It is therefore only possible

to harmonize general regulatory guidelines on data format for biomarker qualification submissions.

## SUMMARY

The qualification of novel biomarkers for drug development and regulatory review is made possible by the development and testing of regulatory mechanisms for biomarker qualification. The urgency for getting more and better biomarkers to improve new drug development is clear to the pharmaceutical industry and regulatory agencies. Harmonized processes for biomarker qualification are being actively developed through ICHs harmonization procedures.

## REFERENCES

1. FDA. Drug-Diagnostic Co-development Concept Paper. http://www.fda.gov/cder/genomics/pharmacoconceptfn.pdf (accessed Oct. 19, 2008).
2. FDA. Table of valid genomic biomarkers in the context of approved drug labels. http://www.fda.gov/cder/genomics/genomic_biomarkers_table.htm (accessed Oct. 19, 2008).
3. Wagner JA (2008). Strategic approach to fit-for-purpose biomarkers in drug development. *Annu Rev Pharmacol Toxicol*, 48:631–651.
4. Rifai N, Gillette MA, Carr SA (2006). Protein biomarker discovery and validation: the long and uncertain path to clinical utility. *Nat Biotechnol*, 24(8):971–983.
5. Goodsaid F, Frueh FW (2007). Implementing the U.S. FDA guidance on pharmacogenomic data submissions. *Environ Mol Mutagen*, 48(5):354–358.
6. Goodsaid F, Frueh F (2006). Process map proposal for the validation of genomic biomarkers. *Pharmacogenomics*, 7(5):773–782.
7. Goodsaid F, Frueh F (2007). Biomarker qualification pilot process at the US Food and Drug Administration. *AAPS J*, 9(1):E105–E108.
8. Goodsaid FM, Frueh FW, Mattes W (2008). Strategic paths for biomarker qualification. *Toxicology*, 245(3):219–223.
9. Anon. (2008). Public consortium efforts in toxicogenomics. *Methods Mol Biol*, 460:221–238.
10. Duarte CG, Preuss HG (1993). Assessment of renal function-glomerular and tubular. *Clin Lab Med*, 13:33–52.
11. Vaidya VS, Ramirez V, Ichimura T, Bobadilla NA, Bonventre JV (2006). Urinary kidney injury molecule: 1. A sensitive quantitative biomarker for early detection of kidney tubular injury. *Am J Physiol Renal Physiol*, 290(2):F517–F529.
12. Zhang J, Brown RP, Shaw M, et al. (2008). Immunolocalization of Kim-1, RPA-1, and RPA-2 in kidney of gentamicin-, mercury-, or chromium-treated rats: relationship to renal distributions of iNOS and nitrotyrosine. *Toxicol Pathol*, 36(3):397–409.

13. Dieterle F, Maurer E, Suzuki E, Grenet O, Cordier A, Vonderscher J (2008). Monitoring kidney safety in drug development: emerging technologies and their implications. *Curr Opin Drug Discov Dev*, 11(1):60–71.
14. Altar CA, Amakye D, Bounos D, et al. (2008). A prototypical process for creating evidentiary standards for biomarkers and diagnostics. *Clin Pharmacol Ther*, 83(2): 368–371.
15. Wagner JA, Williams SA, Webster CJ (2007). Biomarkers and surrogate end points for fit-for-purpose development and regulatory evaluation of new drugs. *Clin Pharmacol Ther*, 81(1):104–107.
16. International Committee on Harmonization (ICH) E16. Genomic biomarkers related to drug response: context, structure and format of qualification submissions. http://www.ich.org/cache/html/4773-616-1.html. (accessed Oct. 19, 2008).

# FIT-FOR-PURPOSE METHOD VALIDATION AND ASSAYS FOR BIOMARKER CHARACTERIZATION TO SUPPORT DRUG DEVELOPMENT

Jean W. Lee, Ph.D., Yuling Wu, Ph.D., and Jin Wang, M.S.
*Amgen, Inc., Thousand Oaks, California*

## INTRODUCTION

Novel biomarkers have been discovered from recent research in genomic and proteomic studies. Differences in gene expression and protein abundance or modification provide the basis for novel biomarkers to be tested further with in vitro and preclinical disease models. Interactions of putative biomarkers at interwoven pathways and the intervening drug actions are relevant to polygenic disorders associated with the complex interplay between genetics, epigenetics, and the environment. Activities of novel biomarker development have increased, especially in diabetes, cancer, rheumatoid arthritis, and cardiovascular disease. In addition to clinical biomarkers proven previously, putative biomarkers have been included in clinical trials of new, mechanistic-based drug candidates for exploratory, demonstrative, or characterization applications in the process of developing mechanism-specific biomarkers toward the ultimate goal of surrogacy (Pepe et al., 2001; Wagner, 2002a; Bjornsson, 2005; Wagner et al., 2007). For example, to the commonly employed measurement of blood glucose, hemoglobin A1c, and circulating insulin, novel

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

biomarkers could be included in diabetes trials, depending on the mechanism of action. For peroxisome proliferator–activated receptor gamma (PPARγ) agonists in type 2 diabetes, free fatty acid was linked to insulin resistance as mechanism-specific, and adiponectin was identified as a proximal biomarker correlated with insulin sensitivity (Berger and Wagner, 2002; Wagner 2002b), while glycosylated hemoglobin ($Hb_{a1c}$) was qualified as a surrogate marker (Krishnamurti and Steffes, 2001). Other mechanisms and the relevant biomarkers (in parentheses), such as lipid and bone metabolism (free fatty acids, leptin, osteocalcin), incretins (glucagon, GLP-1), and inflammation (cytokines, hsCRP, PAI -1, fibrinogen, adhesion molecules) are also involved in the complex disease of diabetes (Lee and Pratley, 2005; Chu et al., 2006; van Doorn et al., 2006; Haffner, 2007; Liu et al., 2007). In addition to circulating prostate-specific antigen (PSA), carcinoembryonic antigen (CEA), and other tumor markers, the mRNA of specific genes, various growth factors, and imaging are also used in clinical trials to monitor drug responses for cancer treatments (Sweep et al., 2003; Rubin, 2004; Quraishi et al., 2007; Yu and Veenstra, 2007; Khatami, 2007).

The use of novel biomarkers has become a prominent component of decision-making processes in drug development. They are used in the in vitro and preclinical models and early clinical phase for quick-hit and early-attrition decisions (Bloom and Dean, 2003; Bjornsson, 2005; Lee et al., 2005; FDA CDER, 2006). Characterization of a novel biomarker in the translational phase requires data collection to show preclinical sensitivity and specificity, and linkage to clinical outcomes in multiple clinical studies in humans (Wagner et al., 2007). The purposes at this phase are different from the exploratory or demonstrative phase. In exploratory and demonstrative studies, pharmacodynamic (PD) correlations are typically unknown, data are used mainly for internal decision making, and the output will generally not be subject to regulatory review. The extent of method validation can thus be limited to a few basic components to expedite the process and preserve resources without unduly affecting commercialization (Lee et al., 2006). In contrast, the purposes of characterization application are to provide pivotal data to establish linkage to clinical outcome and to monitor patient progress upon treatment (Figure 1). The data are often used for critical decisions (such as supporting dose selection and patient stratification, and demonstration of drug safety or efficacy), submitted to review by regulatory agencies, or used for postmarketing patient monitoring (Kummar et al., 2007; Stoch and Wagner, 2007). In addition, the same method used for characterization is likely to be used in the qualification phase toward surrogacy, confirmed over multiple drugs of similar mechanism, and used during surveillance studies. Therefore, biomarker characterization studies would require more intense rigor in assay method validation and in traceable and more detailed documentation than that of previous phases to meet the study objectives in a defined context of use.

A biomarker is useful if there are distinctions in concentrations in the disease state from that of a healthy status (or disease stabilization, in the case

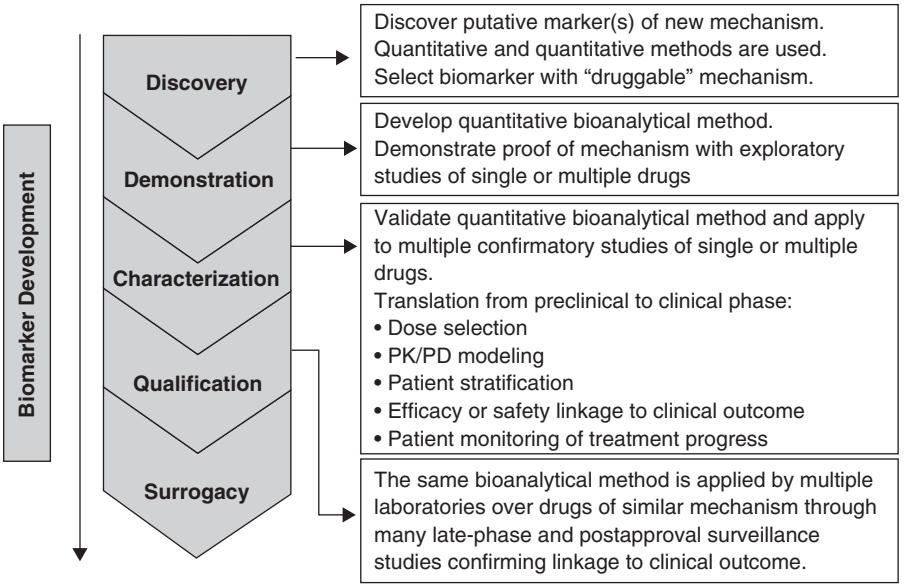


**Figure 1** Path from putative to confirmed mechanism of a novel biomarker through the processes from discovery to surrogacy. The side boxes illustrate the relationship of the processes to drug development applications and the methods required.

of some oncology trials). The assay performance of the biomarker analytical method must be able to meet this goal of disease status distinction and estimation of the drug effect. Therefore, contrary to the general perception of many analytical laboratories, the required assay acceptance criteria do not depend solely on the deliverable method's accuracy and precision. Instead, consideration should be on the assay suitability for the intended applications based on three major factors: (1) intended use of the data during various stages of drug development, (2) nature of the assay methodology and the different types of data that it provides, and (3) biological variability of the biomarker that exists within and between populations. The first factor helps shape the assay tolerance or acceptance criteria for biomarkers.

A fit-for-purpose method validation approach at various phases of biomarker application has been described by Lee and colleagues (2006). In this chapter we focus on how the fit-for-purpose method validation and assay application during drug development clinical trials can contribute to biomarker characterization.

## GENERAL PROCESSES

Quantification of protein expression is important when considering the translation of a candidate protein biomarker. If there are changes in posttransla-

tional modification of the biomarker, it would be preferable to be able to quantify levels of both protein expression and posttranslational modification, utilizing technologies such as ligand-binding assays (LBA), enzymatic assays, liquid chromatography–mass spectrometry (LC-MS) in multiple-reaction mode (MRM-LC-MS), protein assay, and real-time polymerase chain reaction (RT-PCR) genomics. LBAs can be performed in high throughput at a relatively low cost per samples, which is most advantageous for application through phases of characterization and qualification development of a novel biomarker. Therefore, LBAs have been widely used for protein biomarker quantification and will be the typical method for discussion in this chapter.

The developmental processes of a novel biomarker and a new chemical drug entity are intertwined (Lee et al., 2003, 2007). Biomarker development may happen concurrently with the development of a single or multiple drug candidates, a refined new chemical or biological entity, or for extended indications and/or additional mechanisms. The progress of a potential biomarker does not always coincide with or parallel that of a new drug candidate development. To utilize and integrate information from development of drugs in therapeutic programs and the development of relevant putative biomarkers, a detailed work plan can be prepared for the novel biomarker and the study objectives identified in the corresponding study protocols of each drug candidate. Innovative companies often organize biomarker work groups to facilitate timely input and communication among therapeutic areas and supporting teams.

As depicted in Figure 2, novel target (BMKa) and/or proximal biomarkers (BMKb) of a specific mechanism can be used for multiple drug candidates to prove the mechanism of action by drug intervention during the exploratory and demonstration phases. If the circulatory biomarkers are reflecting their actions in the target cells/organs, measurements of blood levels of the target and/or proximal biomarkers can be included during clinical trials for further characterization purposes. In addition, distal biomarkers (BMKi,j) of a specific disease indication that have been used for other drug compounds can be applied for a new drug of a different mechanism toward the same indication. In general, distal biomarkers are closer to the disease symptom magnification for efficacy and off-target indirect effects, while the target and proximal biomarkers are closer to the direct exposure effect. Thus, the establishment of PD relationships to pharmacokinetics (PK) is usually simpler for the target or proximal biomarkers, while linkage to a clinical outcome may be shown more readily by distal biomarkers. PK-PD modeling is used to correlate the dose–exposure effects. PK-PD modeling based on various mechanisms of action, direct and indirect drug effects, and reactions of biological cascades has been developed (Mager et al., 2003). For proof of biology, the biochemical coverage of biomarkers from target hit to distal to clinical outcome would provide a thorough understanding of drug effect and characterization of a novel biomarker, whether it is a target, proximal, or distal. Additionally, off-target negative effects should be monitored carefully by toxicity biomarkers (BMKt). Such information is vital in the selection of lead candidates and appropriate dose ranges for clinical trials (FDA, 2004).

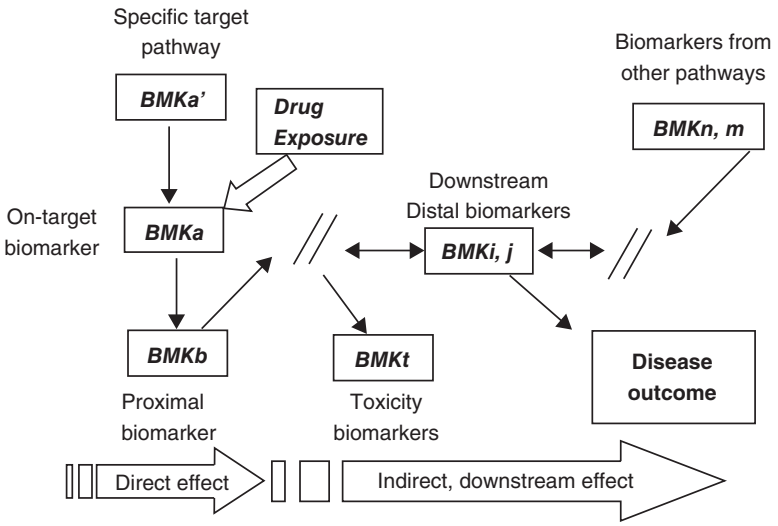


**Figure 2** On- and off-target biomarkers of a specific mechanism of drug intervention. Possible relationship of biomarkers from a specific pathway to disease outcome. Other pathways may have common distal biomarkers that may be used as predictive index of the same disease outcome.

The application of a novel biomarker in clinical trials for the purpose of biomarker characterization is often included as a secondary objective in a clinical protocol, while toxicity and commonly recognized efficacy biomarkers may be included in the primary objectives. Characterization of a specific novel biomarker may be carried out on different new chemical entity development programs over several therapeutic areas. Therefore, it is important for the assay laboratory to have both a "big picture" perspective on the development of a certain novel biomarker and the specific purpose of the application of that biomarker in a clinical trial. The time required for biomarker assay development and method validation, the operational and logistical issues, including preanalytical factors, and the limitations in data interpretation are key elements for careful planning and execution to support biomarker development and its applications.

## METHOD VALIDATION AND ASSAY APPLICATION IN DRUG DEVELOPMENT

### Population Range and Assay Range

In the process of biomarker assay development and validation, samples from the target populations of healthy persons and patients are used to establish expected ranges and biological variations for a given biomarker. Knowledge

of these ranges and assay performance in these matrices are necessary for a number of reasons. The initial concentration ranges of healthy populations are usually available from literature or vendor brochures for previously identified biomarkers. During the characterization phase the range-finding experiments should include sufficient samples from individuals of the target populations (and extended populations if applicable) to enrich the existing information. Biological variation exists in some biomarkers. It is important to note the variable levels over time (i.e., diurnal, seasonal, and food effects) to plan the correct conditions and time of sample collection appropriately. In addition, biological variation can come from the changes in the binding-protein components in the biological matrices, affecting the biomarker measurement in some patient samples. The effect of matrix from patient samples is discussed further later in the chapter.

The results of the range-finding studies and the expected drug effect on the biomarker are used in the design of method validation and assay application in clinical studies. Knowledge of variation and modulation will power the sample size necessary to confidently detect relevant changes in a biomarker after therapeutic intervention. It is desirable to have the biomarker assay range cover the range of the target populations and the expected drug effect. However, this is often not possible. For example, most protein biomarkers are analyzed by LBA, with the assay working range dependent on the antibody reagents used, which is often different from the biological range. Additionally, the biomarker levels of one disease are often different from that of another. For example, inflammatory biomarkers such as IL-6 and TNFα in serum samples for healthy and rheumatoid arthritis patients are in pg/mL range, whereas those of sepsis patients are in the ng/mL range. Use of the inflammatory biomarkers for sepsis will require dilutions into the working range. Thus, application of these biomarkers to sepsis studies would require dilutional linearity during method validation, and the variability introduced by dilution should be monitored during in-study assays.

## Sample Integrity: Sample Collection at the Clinic

Sample integrity means that the measurement of a biomarker after collection, transport, and storage should yield a result as close to that which could be measured in as fresh a sample as possible. Sample characteristics can vary over time, in the fed state, or by age- and gender-specific factors, and by the methods used to prepare the sample. Sample handling can alter a sample irrevocably, so care must be used in specimen collection. Delays in specimen processing, addition of anticoagulants and other additives, type of collection vessel and storage tube, and a myriad of other preanalytical circumstances can affect a biomarker. A careful review of potential confounding variables is essential prior to method validation. Preanalytical variables have been shown to compromise data utility in proteomic biomarker discovery and validation (Ferguson et al., 2007; Banks, 2008).

Serum, plasma, or urine samples are the most common matrices during the characterization phase. For diseases involving the brain or central nervous system, cerebral spinal fluids are commonly used. Accessible cells from sputum, bronchial lavage, and blood cells are used for receptor functional measurements. The preservation of samples must be explicitly defined in the study protocol. Control and standardization of sample collection are especially necessary to minimize variability from multiple clinical sites. For both tissue collection and ex vivo stimulation assays, it is best to validate sample collection procedures using conditions that closely mimic the planned sampling protocol and to examine sample stability as soon as possible. If sample handling is complicated, it would be prudent to establish tolerance boundaries surrounding sample collection and processing for the clinical sites to abide and document. In addition, staff training and oversight procedures are essential for the monitoring and maintenance of sample integrity.

In some biomarker studies, normalization is used to correct concentration variability. For example, urine creatinine concentration is used to normalize hydration variability in urine samples. In that case, the assay of the normalizing factor should also be validated on stability and reliability prior to the clinical study.

During biomarker characterization, multiple clinical studies from multiple drug candidate development programs may be involved. It is important for the biomarker team to coordinate the study protocols to include standardized procedures for sample collection, shipping, and analysis and for sample storage and disposal. Sample collection and storage stability should be tested by the bioanalytical laboratory to cover the conceivable conditions during in-life study sample analysis. This includes process stability, assessing the impact of short- and long-term storage, and multiple freeze–thaw cycles.

## Method Development

***What Will Be Measured*** Many biomarkers of interest are proteins whose function is affected by a change in disease progression and drug intervention. The ultimate goal of a biomarker assay is typically to evaluate the effect on the biomarker activity in vivo, and a method that measures this activity accurately is most desirable. Direct measurement of changes to target cells, tissues, or organs in response to a drug is rarely feasible, and given current drug potencies, activity assays rarely have adequate sensitivity and precision. Moreover, cell-based and activity assays are usually laborious, low in throughput, and often lack definitive reference standards. LBA are often chosen for biomarker analysis over activity measurements because of their sensitivity, versatility in application, relatively high throughput, and low cost.

Many biomarkers are endogenous protein that can be physicochemically heterogeneous, with multiple forms present simultaneously. The concentration and biological activity of each component are often unknown and may vary with health status, over time, and between individuals. However, it is

desirable to know, even in the case of a well-defined regulatory peptide, the related species of the analyte existing in the biological matrix of measurement. For example, measurements of proinsulin, insulin, and C-peptide provide different information for clinical interpretation and assessment. For large molecules, the heterogeneity of the molecules poses challenges to identification of the exact species being measured. Therefore, various approaches in designing the analytical method, including choices of the reference standard and critical reagents, are worth considering.

LBA methods measure the binding reaction of the mixed components in the matrix. The majority of the data generated by LBA is relative quantification (Lee et al., 2006). Limitations of LBA methods are usually a lack of chemical information or correlation of the binding activity to the biological activity. In addition, the reference standard used for data regression may be a recombinant protein or one of several components which may or may not be representative of the entire endogenous mixture. However, as the purpose during biomarker characterization is to monitor population and longitudinal effects of drug treatment and disease progression, this type of relative quantification is adequately "fit for use."

***Reference Standard and Alternative Matrix for Standard Preparation*** The reference material may not be fully representative of the endogenous analyte because the endogenous analyte may exist in various forms in the biological matrix. A well-defined reference standard is used as a scalar for the relative measurement of the endogenous species. For example, the recombinant form of a biomarker can be produced and characterized with defined molecular size and purity to enable concentration and molar equivalence calculations. Documentation of the reference material is necessary for its use.

Most biomarkers are endogenous molecules with variable measurable amounts in the matrix. It is difficult to find "blank" control matrix for preparation of reference standards. An initial standard curve can be constructed by spiking the reference standard into a protein buffer solution. An initial screen on multiple matrix lots against the buffer standard curve may identify some blank matrix lots for standard preparation. When that is not possible, an alternative matrix can be prepared by depleting the endogenous analyte using methods such as charcoal stripping, high-temperature incubation, acid or alkaline hydrolysis, or affinity chromatography. Commonly, a protein-containing buffer or the matrix from another species with a non-cross-reactive homolog is used. The use of alternative matrices require studies of matrix effects and parallelism during method development and validation to understand the impact that differences from patient samples will have on assay results.

Parallelism experiments are performed through serial dilutions of a high-concentration sample with the standard matrix. Multiple individual matrix lots (three or more) should be tested to demonstrate lot-to-lot consistency. There are several ways to assess data of parallelism (Lee et al., 2006). The concentra-

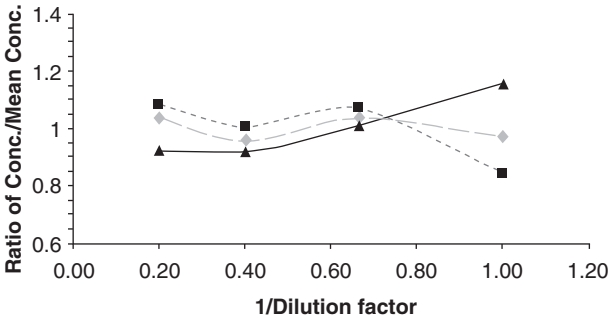


**Figure 3** Parallelism of three individual matrix lots. Each lot is represented by a separate symbol. The sample was diluted with the standard diluent at dilution factors of 1.5, 2.5, and 5. The concentrations of the diluted and undiluted samples observed were all within the standard curve assay range. The mean of the calculated concentrations (observed concentration times the dilution factor) of each lot was determined. The ratios of the individual calculated concentration over the mean of the lot were plotted against 1/dilution factor. The data showed that the ratios were all within the range 0.8 to 1.2, which was the defined acceptance criterion of the assay. Therefore, the parallelism test results were acceptable for the test lots.

tions of diluted and undiluted samples observed should all be within the standard curve range. An example is depicted in Figure 3: The mean of the concentrations calculated (concentration observed times the dilution factor) of each lot was determined. The ratios of the individual concentrations calculated over the mean of the lot were plot against 1/dilution factor. Parallelism was demonstrated for the test lots because the ratio was not affected by the variable amounts of standard matrix introduced by dilution. When parallelism is not possible due to the unavailability of samples with sufficiently high concentrations of analyte, dilutional linearity should be tested in a manner similar to parallelism, except that high-concentration spike samples are used in place of the endogenous samples.

The failure to demonstrate parallelism should be taken into consideration in data assessment and interpretation. For example, it may mean that the values obtained should be treated as quasiquantitative instead of as relative quantitative measurements (Lee et al., 2006). In addition, within-subject comparison on a longitudinal time sequence would be more useful than between subjects or populations, which affects the clinical study design. All of these considerations would be dependent on the a priori goals of the study.

***Selection of Critical Reagents in Ligand-Binding Assays*** For the development of a first-in-class drug compound that targets a novel biomarker, assay reagents will need to be developed for both the drug candidate and the target protein biomarker. In the case of a macromolecular drug, often an analog of the drug candidate may be used as the binding reagent(s). On the other hand,

reagents for most off-target biomarkers may be available from commercial sources. These reagents may be established assay kits [U.S. Food and Drug Administration (FDA) approved for diagnostic use] or kits for research use only.

For the purpose of characterization of the biomarker of interest, fit-for-purpose method validation should be conducted with sufficient rigor to provide reliable data from multiple clinical studies. This should include assay range finding, accuracy and precision, selectivity, specificity, stability, and robustness (reproducibility) during prestudy method validation. Additional data will be collected from in-study assay performance, change control method validation, and long-term storage stability. Except in the case of FDA-approved diagnostic kits, many of the assay performance parameters will have to be established by the bioanalytical laboratory. For FDA-approved kits, because the application will be for drug development instead of diagnosis, some of the assay performance parameters will need to be established. These often include selectivity against the target patient sample matrices and specificity against the drug compound(s). If the drug is expected to decrease the biomarker to a level lower than the lowest standard of the kit, the method may need to be modified to increase assay sensitivity.

Antibody pairs are typically chosen as capture and detection reagents. In general, the more selective antibody would be chosen as the capturing agent, especially if it is more readily available than the other member of the pair. A tertiary detector antibody can be used that is conjugated to a reporter enzyme such as horseradish peroxidase. Alternatively, a biotinylated detector antibody can be used together with a biotin-binding protein (e.g., antibiotin antibody or an avidin-type protein) conjugated to a reporter enzyme. The sensitivity of an assay can be increased by varying the number of reporter enzyme molecules on the detection reagents, or by using multivalent strategies to increase the effective signal from each analyte captured.

Some assays use receptors or their fragments as binding partners, most often in concert with a specific second antibody. This arrangement can improve selectivity for specific ligands (e.g., a cytokine activated from a latent precursor, or a particular subtype of ligand with distinct binding characteristics from its homologs). The binding selectivity of such reagents can offer added biological relevance to quantification, and can suggest results that might otherwise be obtained only via functional assays.

***Critical Reagents Characterization, Consistency, Stability, and Documentation*** Characterization of a novel biomarker often involves many clinical studies that last for several years. If the studies are to be conducted within the same company or from a joint program from different institutions, plans should be made to provide consistent supplies of the same reference material and assay reagents throughout the studies. If the materials are from in-house sources, sufficient production of the reference standard and primary reagents should be assured. In addition, multiple preparations of labeled

detector reagents should be assessed during method validation. If the materials are from commercial sources, negotiation with the vendor should take place to assure a consistent and sufficient supply of the same batch material, if possible. Often, multiple batches of materials (reference standards and critical reagents) should be tested during method validation. Documentation should be obtained for the stability of the respective lots over the time span of a clinical study. Documentation of chain of custody should be similar to that of a good laboratory practices (GLP) study. For late-phase studies using multiple clinical sites, the use of a central sample repository offers numerous advantages in controlling the process of specimen collection and labeling of multiple biomarker assays.

In-study sample control (SC) charts such as those shown in Figure 4 can be used for trend analysis of assay performance and stability. A big pool of endogenous sample at low, middle, and high concentrations of a biomarker were prepared during method validation and monitored throughout in-study.

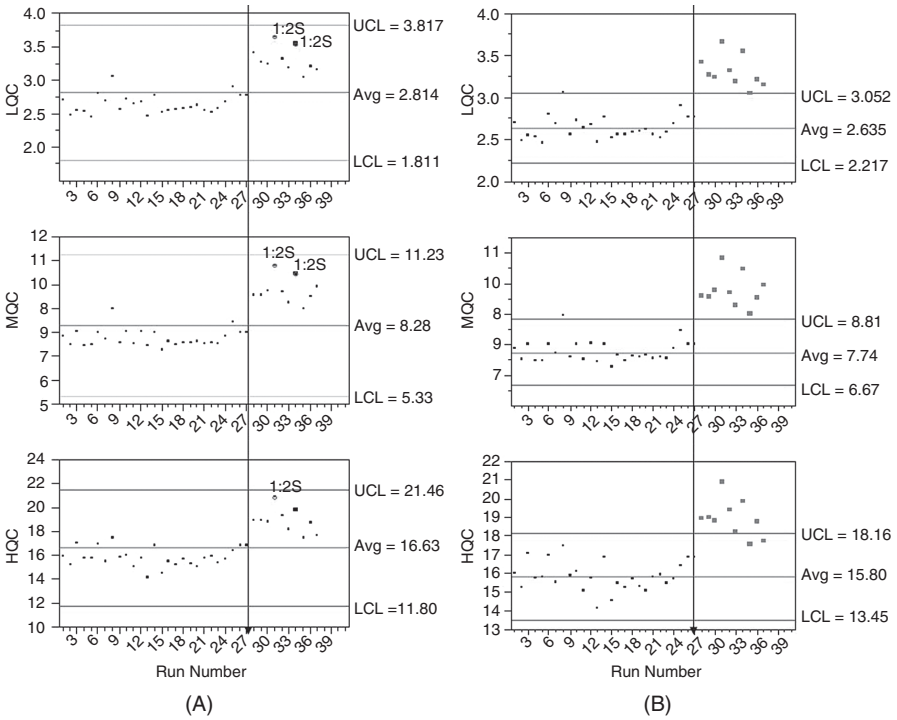


**Figure 4** Use of sample controls for trend analysis on variability. Low, middle, and high sample controls of a biomarker were monitored in Levey Jennings control charts. There was a noticeable shift in trend in all the sample control levels after run 27. The average, upper and lower control limits of analytical runs up to run 27 (B) were compared to the overall parameters (A). (*See insert for color reproduction of the figure.*)

The Levey Jennings control charts showed a noticeable shift in trend at all the SC levels after run 27, indicating a systemic bias. The average, upper, and lower control limits of analytical runs up to 27 (right panels) were different from the overall values (left panels). This shift was traced to a change in commercial kit lot.

## Method Validation on Performance Parameters

The foremost assay performance parameters to be established for quantitative methods are precision and accuracy. In addition to these two parameters, sensitivity, selectivity, specificity, stability, and reproducibility should also be demonstrated. These parameters are established during prestudy method validation to demonstrate reliably that the method is fit for the purpose of biomarker characterization over multiple clinical studies. If more than one bioanalytical laboratory is used, each additional laboratory should also demonstrate its ability to perform by conducting a full or partial method validation.

***Accuracy and Precision*** Method performance is consistently better understood and validated through the appropriate use of control samples. To ensure data quality, assay performance is evaluated during method validation with validation samples (VSs) and monitored during sample analysis with quality control samples (QCs). VSs are used in method qualification or validation to define intra- and interrun accuracy and precision, providing data to demonstrate the robustness and suitability of the assay for its intended application. On the other hand, QCs are essential in determining run acceptability during specimen analysis. However, many laboratories have not made the distinction, using the term QC for both VS (for prestudy method validation) and QC (for in-study run acceptance).

For a well-defined biomarker assay, at least six accuracy-and-precision validation runs should be performed to provide statistical data to calculate these assay parameters (DeSilva et al., 2003; Lee et al., 2008). Each run should consist of standards prepared by spiking the reference standard into blank biological matrix or an alternative blank matrix. QC/VS at low, middle, and high concentrations should be prepared in the biological matrix or alternative matrix by spiking the reference standard. In addition, sample controls (SCs) should be prepared by pooling authentic samples at the low and high levels of the biomarker to reflect the performance of the endogenous biomarker in assay precision, stability, and reproducibility. At least two sets of QC/VS and SC should be run with the standards in each accuracy and precision experiment.

Accuracy and precision can be evaluated from the total error of VS data from the validation runs in a way similar to that of the macromolecular protein drug as analyte (DeSilva et al., 2003). However, given biological variability and other factors in biomarker research, more lenient acceptance criteria may

be used for biomarker PD than that for PK studies. Still, it should be recognized that accuracy and precision data of VS in buffer provide only a relative quantification, which may be quite different from measurements in the authentic matrix.

Concentrations of the endogenous SCs will be determined by multiple validation runs. The true values are then defined after sufficient data collection from pre- and in-study validation. For example, the mean of 30 runs and 2 standard deviations can be used to define the target concentration and acceptable range of the SCs. Because the reference material may not fully represent the endogenous biomarkers, the SCs should be used for stability tests (such as minimum tests of exposure to ambient temperature and freeze–thaw cycle). In addition, the SC can be used for in-study, long-term storage stability analysis and for assessment of lot-to-lot variability of key assay reagents.

Regression models are essential for data calculation from sigmoid curves like those in LBAs. The most commonly used four- or five-parameter logistic regression should be evaluated in conjunction with weighting factors during method development. Final decisions on which curve-fitting model to use should rest on which offers the best fit for all the standards in the precision profile. In some cases, a less than optimal fit will suffice to allow for greater assay sensitivity.

***Sensitivity*** To provide data for PK/PD studies, assay sensitivity is usually defined by the assay lower limit of quantification (LLOQ), which is the lowest concentration that has been demonstrated to be measurable with acceptable levels of bias and precision and total error (the sum of bias and precision). However, low-concentration clinical samples may fall below the LLOQ (i.e., the method lacks the required sensitivity). In some instances, the investigator may want to use values below the LLOQ but above the limit of detection (LOD), to obtain a numerical estimate of the changes while recognizing the high variability below the LLOQ region.

LOD is often used as the analytical "sensitivity" of the assay in a commercial diagnostic kit. A common practice of diagnostic kits to determine the LOD is the use of extrapolated concentrations from a response signal of +3 SD (or −3 SD for a competitive assay) of the mean background signal from 30 or more blank samples. The National Committee for Clinical Laboratory Standards (CLSI) has recommended an approach for determining LOD (National Committee for Clinical Laboratory Standards, 1999; Tholen et al., 2004). This statistically sound approach evaluates the limit of blank (LOB) as type I error and LOD as type II error from sufficient numbers of blank and low concentration samples with normally distributed analyte concentrations. One approach of calculation is described briefly as follows: LOB is estimated from the blank result at the 5th percentile position. LOD is calculated as

$$\mathrm{LOD} = \mathrm{LOB} + c\beta\ \mathrm{SDs}$$

SDs is the estimated standard deviation of the low-concentration sample measurement, and $c\beta$ is a correction factor for the population bias:

$$c\beta = \frac{1.645}{1 - 1/[4(N-K)]}$$

Here $N$ is the number of determinations and $K$ is the number of sample sets.

The LOD determined by this approach would avoid the inclusion of analytical false negative and false positive values. Due to assay conditions being variable, each bioanalytical laboratory should determine the LOD using the CLSI approach instead of using the LOD stated in a kit brochure. If the LOD is used as the lower limit for data inclusion instead of the LLOQ, the user should justify such a choice and be aware of the risk being taken with the higher variability at the LOD to LLOQ range, interpreting the data with caution.

***Selectivity and Matrix Effect*** Selectivity is the ability of the assay to discriminate the analyte unequivocally in the presence of components that may be expected to be present in the sample and alter assay results. Lack of selectivity can result in signal inhibition or enhancement. The results can appear as false or negative positives (an over or underestimation of the analyte concentration). In general, signal suppression from binding proteins occurs more often for LBAs than that of enhancement, resulting in a negative bias.

Interference in an assay is the product of concentration of the interfering substance times the cross-reactivity (or binding inhibition). However, the concentration–response relationship of LBA is nonlinear, and often the magnitude of cross-reactivity (or inhibition) is not evenly dispersed over the entire assay range. Thus, a standard curve prepared in a protein buffer solution has higher optical density (OD) responses than that in biological matrices, while the differences in responses are variable over the standard curve. Moreover, a standard curve prepared in biological matrices from one person would have different responses than that of other people. Therefore, the estimate of interference should not simply extrapolate across the entire assay range or be represented by an $IC_{50}$ value.

Matrix interference should be surveyed in samples from normal and diseased donors, especially in samples from the anticipated patient population. Spike recovery using samples from persons in the target population should be performed. A later selectivity evaluation of LBA involves a parallelism test of high-concentration endogenous samples from many people diluted at ≥4 dilution factors with the standard diluent.

***Specificity*** Specificity reflects the ability of an assay to distinguish between the analyte of interest and other structurally similar components of the sample. For LBAs a specificity test should include test of interference from components that may specifically affect the binding reaction. For example, inter-

ference from the drug compound may alter ligand interactions with the capture reagent. If the ligand reagent of an LBA is the target drug or its analog, the presence of the drug compound at certain concentrations can interfere with the assay. In addition, biomarker homologs or endogenous molecules of the same family may contribute to specificity problems. If reference material of the potential interferent is available, specificity can be tested using various amounts of the test material spiking into various levels of VSs (Lee and Ma, 2007).

Specificity for the biomarker in the presence of the drug under study should also be tested, using samples from endogenous pools (such as the SCs), with and without the addition of the drug compound at various concentrations spanning the expected therapeutic range. Figure 5 illustrates a specificity experiment of a target biomarker (A) and its proximal biomarker (B), each tested with three sample pools at pM levels against various nanomolar drug concentrations. As expected, the test drug inhibited the quantification of the target biomarker at concentrations greater than 2 nM. For the proximal

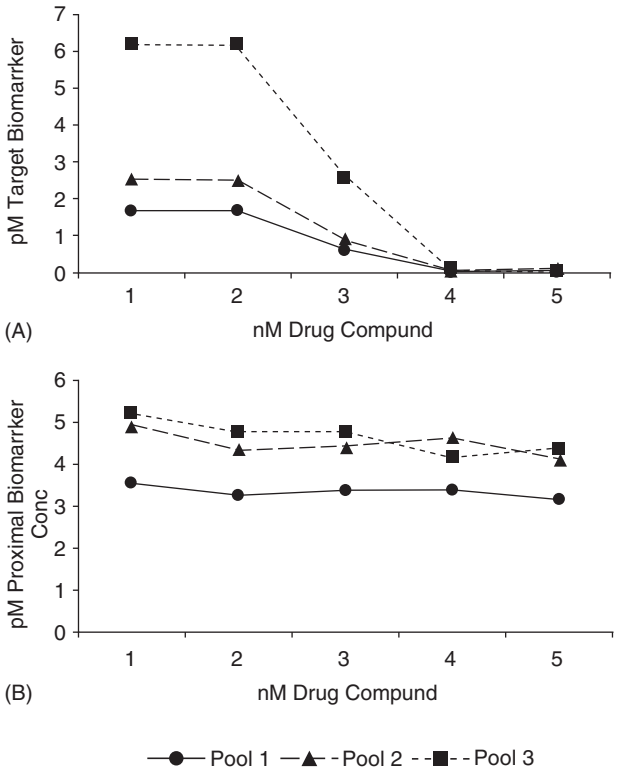


**Figure 5** Method validation specificity tests of a target biomarker (A) and a proximal biomarker (B) against a drug compound.

biomarker, the test drug did not affect the assay at all test concentrations, which exceeds the highest expected concentration of the drug in the serum. The results illustrated that the drug affected only its target protein.

***Stability*** In addition to sample integrity during collection, stability of protein biomarkers and assay reagents is required to show that the assay is not compromised by this preanalytical factor. Stability information is gathered during method validation on short-term storage and process stability, and is extended through in-study validation for long-term storage stability. Reagent storage stability is tested on reference material stock solutions and ligand-binding reagent stock solution if information from the supplier is lacking. Working solutions will be prepared from the stored stock solution and assay performance compared to those from that of a freshly reconstituted solution. Process stability is evaluated at different time durations and various conditions to mimic possible sample handling in the laboratory from the results of SCs subjected to stress from multiple freeze–thaw cycles and temperature exposures (e.g., four cycles and up to 16 to 72 hours to at 2 to 8 °C and ambient temperature).

Long-term storage stability of the SCs should be started during validation and continued to support the drug development program. The acceptance criteria of storage stability are according to those established during method validation, prior to conducting the experiment. For long-term stability evaluation, in addition to comparison against the initial concentration, trend analysis can be more useful than the acceptance criteria strictly applied. This is due to the challenges of a kit lot change, potentially shifting the mean value of the SCs, and a new preparation of the standard curve, introducing systemic bias to add an error component that is not caused by the storage conditions (similar to that illustrated in Figure 4).

## Assay Application

Before sample analysis begins, a written method standard operating procedure (SOP) must be in place. Often, a validation report should also be issued. Documentation of sample assay should follow the SOP in a GLP compliance manner (see the discussion later on regulatory issues), with traceable records. The data should be stored in a secure place for auditing.

***Standard Operating Procedure*** A validated method should have defined procedures, assay performance parameters collected from experiments using the same method during prestudy validation, and acceptance criteria established reflecting the acceptability of the method performance. These elements constitute a method SOP and are discussed in the following section.

*Method Procedures* A written method procedure should clearly describe the following:

- The intended purpose of the method
- The principle of the method
- A validation performance summary
- Materials and reagents, including lot numbers and expiration dates
- Solutions preparation
- Standards and QC, SC preparations
- Sample pretreatment (processing)
- Analytical steps
- Data regression
- Acceptance criteria
- References

During the development of a novel biomarker, a series of method modifications can take place. It is important that specific names (version number) and effective dates be given to the same method as that used in a clinical study. A historical log should be kept on the modifications made for each version, and its application in specific studies identified. Sometimes, for various reasons, if a method change has to happen within a clinical study, a crossover comparison of the old and new methods applied to a subset of clinical samples must be designed with a statistical approach to show method equivalence.

*Regression Model* The response–concentration relationship of the LBA method is nonlinear. Curve fitting would require regression model selection based on multiple runs of standard curves from the method validation accuracy and precision experiments. Typically, a four- or five-parameter logistic model will fit most methods with a choice of weighting factors (such as 1/response or 1/variance). The initial selection is based on the residual of the standard mean. It can be chosen to obtain the best fit over the entire range and with regard to the low end if sensitivity is an issue. The minimal total error of the VS data generated in the validation batches should be used to determine the regression algorithm and weighting. Once the regression model is chosen, it is a basic element of the method that should not be changed during assay implementation. Justification must be given and documented for a regression model change.

*Acceptance Criteria* Without sufficient data from the healthy and patient samples, the acceptance criteria for a novel biomarker can initially be set as determined by the assay performance from prestudy method validation. The biological data obtained from in-study validation from subject samples can then be used to refine the initial acceptance criteria set by the prestudy validation. The process of setting acceptance criteria for a protein biomarker measurement follows an evolving path. During the exploratory phase, the acceptance criteria were initially set according to the initial assay performance

of prestudy method validation. After use in pilot studies, assessment should be made on the suitability of the assay stringency vs. the effect observed. The biological data obtained from the in-study validation from subject samples during in-study phase can then be used to refine the initial acceptance criteria set by the prestudy validation. For example, an assay with 50% total error may still be acceptable for detecting a twofold treatment effect observed in the in-study phase clinical trial. Setting the same acceptance criteria of a given method may be most convenient for an analytical laboratory. However, for a novel biomarker, it may not be the most appropriate for all of its applications. One could take into account the intended purpose of the application and the possible outcomes in the analytical phase. For example, the effect of one population or indication may be different from another (e.g., change from a twofold treatment effect into only 30%), which may require a more stringent method and/or more subjects to increase the predictive power in such an application for the other population.

***Controls in Clinical Studies*** For a well-planned assay application, both analytical quality controls and biological sample controls should be available to assess the assay variability. Data of analytical variability are gathered from standards, QC and SC of each analytical run of sample analysis. A common set of SCs (or pools of incurred samples from a previous study) analyzed by multiple bioanalytical laboratories can provide an assessment of assay performance among the laboratories. Variability of supplies in reagents and reference standard materials can be detected by tracking the assay performance of a common set of SCs used within and between studies. An example is shown by an SC chart for serum C-terminal telopeptides of type I collagen (CTx) assay in Figure 6, using the Westgard rule for monitoring (Westgard, 2003) as well as a priori acceptance criteria of 25%. The low and high SCs were prepared by pooling a large volume of authentic samples, and the mean values were established during method validation. Figure 6 shows the Levey Jennings charts of more than 117 analytical runs from five clinical studies (A to F) over more than 2.5 years. During this time span, two reference stock materials and three kit batches were used. The plots show a slight positive bias in the SC values after the initial sequence of studies B and C. In addition, a slight negative bias was observed prior to study A. However, since the SC values were still acceptable within ±25%, no action was taken. Other statistical analysis can be use to assess SC data on other assay variability sources, such as operator differences and sample storage stability trend.

At the same time, data from predosed samples can furnish information on biological variability among the subjects and between studies. Figure 7 shows the baseline concentration distribution of TRACP 5b, a bone resorption biomarker, in various patient populations from four clinical studies. Statistical tools can be used to construct distribution charts of the predosed samples of each population, and cohorts can be compared. For pilot clinical trials with small number of cohorts, one should note that the data may not be normally

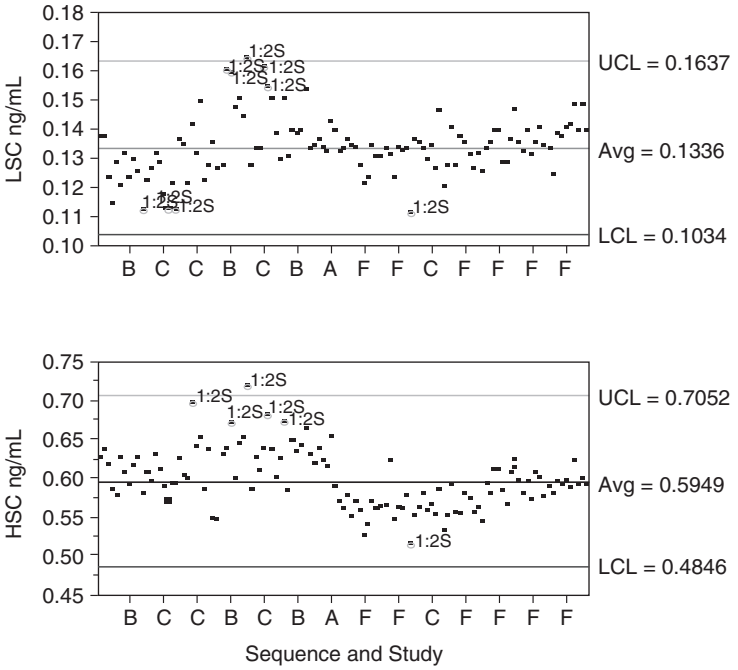


**Figure 6** Levey Jenning plots of low and high sample controls of serum CTx. Concentrations were plotted against run sequences labeled with the specified studies from A to F. UCL, upper confidence limit; LCL, lower confidence limit. Sample results exceeding the Westgard rule of 1:2 S (beyond 2 standard deviations) were marked.

distributed (example in Figure 7A and C) as compared to those with the late-phase clinical trials (example in Figure 7B and D). In addition, multiple collections of predose samples from the same subject can provide sample collection variables of multiple clinical sites.

***Confirming Selectivity in Patient Populations*** Biomarkers can reflect the dynamic changes in disease progression and response to therapy, but homeostatic and compensatory biological responses can alter the expected profile in response to drug treatment and concomitant therapies (Heeschen et al, 2003; Vural et al., 2006). In the latter case, biomarker assay selectivity is a critical component for interpreting biomarker responses and clinical data, particularly when a concomitant therapy is a potential interferent in the assay. For example, the bone turnover biomarker, serum CTx, is being used in clinical studies involving many different populations. During method validation, serum from healthy males and females and from patients with breast cancer, multiple myeloma, osteoporosis, and rheumatoid arthritis were analyzed with and without the addition of CTx at 0.145 ng/mL. The spike recovery was calculated

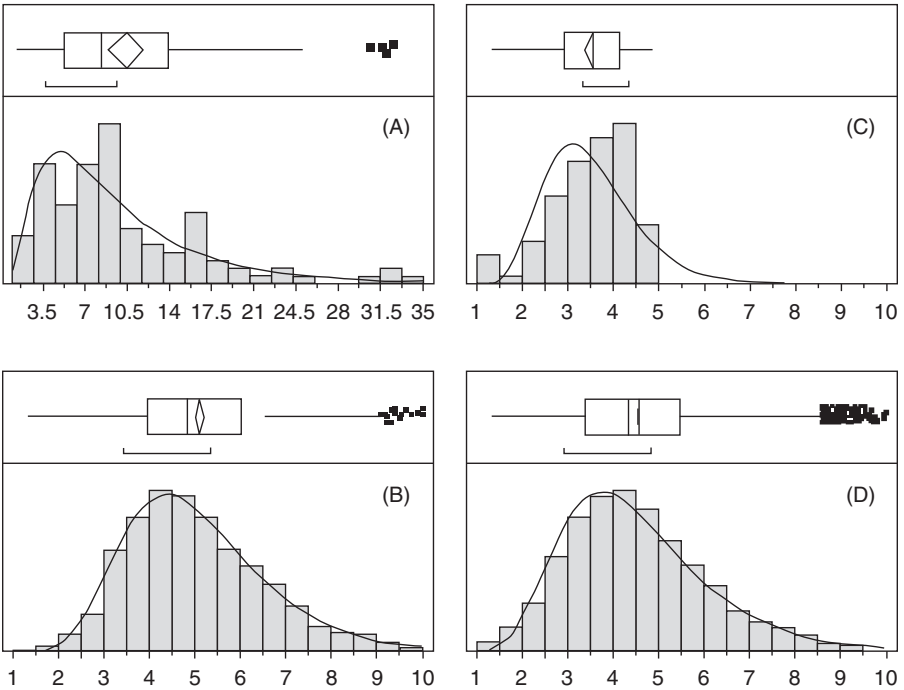


**Figure 7** Baseline concentration distribution of TRACP5b in various patient populations from four clinical studies: (A) patients with advanced cancer ($n$ = 105); (B) patients with prostate cancer ($n$ = 1427); (C) postmenopausal patients ($n$ = 222); (D) postmenopausal patients ($n$ = 7828). The curved line represents a fit by lognormal distribution. Outlier box plots are shown in the panels above the histograms. $X$-axis, concentrations of TRACP5b in ng/mL; $Y$-axis, relative frequency.

by subtracting the endogenous concentration of the unspiked sample from the spiked sample. The recovery data of the percent difference from the nominal value of the various populations were presented in Figure 8. The results show that all normal male and female serum recovered with the acceptance criteria (dashed line of ±25%). Recoveries were more variable with patient samples, although most of them were acceptable. The recovery data set of normal females had the best precision and was used as the control in Dunnett's method to compare with the other groups. The Figure 8 side bar shows that the circles of all groups were overlapping with the control group (bold circle). All groups had $p$-values less than an $\alpha$ of 0.05, while that of the rheumatoid arthritis group was 0.0587, possibly due to a slight interference from the rheumatoid factor in the samples.

*Specificity and Sensitivity of Novel Biomarker on Drug Effect* In-study data should be compiled and assessed to determine if (1) assay performances of

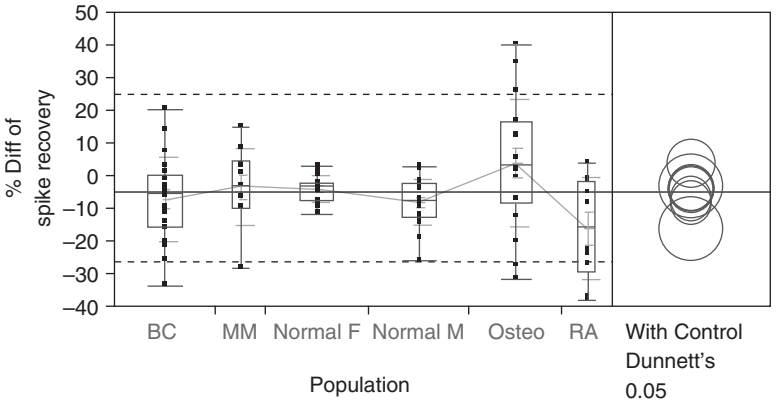


**Figure 8** Selectivity test of CTx in sera from healthy and patient populations. The populations included healthy male (normal M, $n = 20$) and female (normal F, $n = 20$); patients with breast cancer (BC, $n = 20$), multiple myeloma (MM, $n = 10$), osteoporosis (Osteo, $n = 18$) and rheumatoid arthritis (RA, $n = 10$). The normal male age range was 45 to 74, that of female was 38 to 67. Serum samples were analyzed without and with the addition of CTx at 0.145 ng/mL. The spike recovery was calculated by subtracting the unspiked sample concentration from that of spiked sample for each person. Percent (%) difference from the nominal value of 0.145 was calculated and the data presented in box plot for each group of population. Each box presents the mean, standard deviations, and range. Side panel: Comparisons with a control using Dunnett's method.

standards and SCs are within acceptance criteria, (2) patient sample concentrations are in the ranges expected for predosed samples, and (3) postdosed patient sample concentrations are consistent with the values expected for previous pilot study (such as PD profile, dose effect, and recovery to normal after treatment change, if applicable). Data should be evaluated for false positive (specificity) and false negative (sensitivity). For example, the variability of the SCs could reflect the assay noise of incurred samples. This variance should be smaller than that of the difference in disease subject vs. normal (specificity) and that of concentration change due to drug versus placebo effect (sensitivity).

The data mining, interpretation, and linkage of biomarker data to clinical outcome is beyond the scope of this chapter. It should be borne in mind that a useful predictive index of a biomarker is dependent on whether it is used as one of a panel of biomarkers. The statistical power is affected by the sampling number (number of subject and time point) and assay stringency. Therefore, the bioanalytical laboratory should be included in the biomarker team to interpret data use, the model being tested, and the analytical component required for the predictive index desired.

***Regulatory Issues*** Regulatory guidance for bioanalytical validation indicates that the paramount objective of a bioanalytical method validation is to ensure

that the assay is reliable for its intended use (ICH Guidelines, 1994; Shah et al., 1992). This philosophical viewpoint is central to many guidance documents, conference reports, and "white papers" for bioanalytical validation of assays for conventional small-molecule drugs and macromolecular therapeutics (Shah et al., 2000; *Code of Federal Regulations*, 2001a; FDA, 2001; Findlay et al., 2000; Miller et al., 2001; DeSilva et al., 2003). Currently, two basic bioanalytical approaches are being applied to biomarker validation. One is based on the FDA bioanalytical drug assay guidance and is generally referred to as "GLP-like" (*Code of Federal Regulations*, 2001a; FDA, 2001). The second is based on the CLSI and CLIA (National Committee for Clinical Laboratory Standards, 1999; *Code of Federal Regulations*, 2001b). Presently, no regulatory guidance has been given concerning what experiments should be performed and what data are necessary and appropriate for biomarker assay validation (Lee et al., 2007). This has led to inconsistent application of validation procedures and acceptance criteria for biomarkers in clinical trials.

For biomarker assays to be characterized and used for drug development support, they should be GLP-like (Lee et al., 2005, 2006). However, the guidelines must be flexible, to allow assay- and technology-specific applications to meet specific clinical study objectives on a case-by-case basis. Thus, general guidelines for bioanalytical laboratories should be developed to include both good scientific and clinical sense with flexibility to develop specific applications (Lee et al., 2007). At the same time, documentation of change and version control of biomarker methods is important, especially to those who produce elements of the assay or kit.

The FDA has set up an Interdisciplinary Pharmacogenomic Review Group (IPRG) to carry out specific review function for the assessment of biomarker qualification data. The team will evaluate study protocols and review study results of novel biomarkers of drug safety, using appropriate preclinical, clinical, and statistical considerations. The team will then develop recommendations and guidance for the submission of biomarker data, assess the original biomarker context proposal through voluntary data submission (VXDS, where the X indicates a wide range of data sources), and then evaluate the qualification study protocol together with the sponsor to reach a consensus protocol (Goodsaid and Frueh, 2007). Characterization data of novel biomarkers to support efficacy can be categorically in the voluntary data submission, and discussion with the IPRG will clarify the data requirement for drug labeling.

***Documentation*** Documents and data on biomarker characterization to support drug development must be kept in a secure place and made accessible for internal or external audits. All processes and data generation should be traceable. It is critically important to define and document the reference standard. Because novel biomarkers do not usually have established reference standards, characterization of the standard material is dependent on the supplier laboratories from commercial or in-house sources. Documentation such as a certificate or record of analysis and a certificate of stability should be

obtained. If these are not available, at least the source, identity, potency (or concentration), and lot number of the reference material should be documented. A consistent and sufficient supply of the same reference material lot should be reserved and used for the duration of a study and program if inter-study comparisons are intended.

Similar to the reference material, documentation of the source, identity, potency or concentration, and lot number of critical ligand reagents and their stability should be kept. If possible, the same lot of the capturing ligand should be used throughout a study. It is prudent to test for method robustness using multiple lots of reagents during advanced method validation.

***Partial Validation for Change Control*** The time span of biomarker characterization can be a long one. Changes in operators, assay equipment, reagent lots, and a variety of other unforeseen events usually happen and will require additional validation. The significance of the alteration in a method will determine the extent of validation necessary, leading to cross-validation, partial validation, or complete revalidation of an assay to control for changes during or between studies. A validation plan should be prepared for partial validations, describing the intended purpose of change, scope, experiments to be carried out, data comparison, and a prior acceptance criterion for method equivalence. A common set of clinical samples analyzed by both methods can be used to show method equivalence. The sampling size should be powered by statistics based on the desired criterion of acceptance and the method variability. If the method change is expected to be extensive, a new validation would be conducted according to the new analytical procedure, followed by a new written validation report.

## CONCLUSIONS AND PERSPECTIVES

In this chapter we have focused on a discussion of quantitative method validation and assay for biomarker characterization to support drug development using the fit-for-purpose approach. The main processes include: defining the purpose of application, assay development, validation and use meeting the purpose and statistical treatment, and interpretation of analytical and clinical data.

Due to the heterogeneous nature of many protein biomarkers, discrepancies may exist between the endogenous forms and the reference standard. It is important to understand what is being measured and the clinical meaning of the moieties measured as related to disease progression and drug effect. For example, during the course of development of a LBA method, biophysical methods (such as mass spectrometry coupled to ligand binding solid-phase extraction) can be used to select reagents and identify the chemical species being captured in the ligand solid phase.

In addition, there are differences in the standard curve blank control matrix from the intended biological matrix. Parallelism is a characteristic that must

be tested using authentic samples from the target population. Another important consideration is that binding of the biomarker to endogenous binding proteins, receptors, and the xenobiotic drug candidate may cause assay anomaly. Different methods may measure different forms of the biomarker, which often leads to the confounding data in the literature. Specific binding reagents and processing steps can be designed and manipulated to measure "bound" or "free" species of the biomarkers.

In contrast to the exploratory phase, biomarker characterization requires more intense rigor in a quantitative method. For example, method robustness and long-term storage stability must be established; sufficient and consistent supplies of reagents should be assured to sustain the applications through multiple studies; and the same methods should be validated and applied in the participating clinical sites on sample collection and in bioanalytical laboratories on sample assays. Only when a robust method has been established and the assay performance defined, is it possible for the biological variability (clinical variance) to be distinguished from analytical variability and be assigned its clinical significance.

Statistical treatment of biomarker assay performance enables assessing if the method will meet the study purpose before its application. Statistical data from in-study performance continue to add information to the method appropriateness, especially with data from the sample controls, which provide trends in authentic sample stability and reagent lot variability. Data from multiple studies and drug candidates (from the first-in-class to later backups, or to another candidate of related pathways) contribute to the accumulated knowledge of the biomarker.

Whereas the exercise of utilizing biomarkers to support drug development of a single or several candidates can be controlled within a pharmaceutical company, the movement of characterization and qualifications of a novel biomarker will depend on the concerted efforts of multiple institutes/organizations. The collaborative effort of sharing and using the same method and reagents may be a challenge. If different methods are to be used, the data interpretation and comparison will add complexity to this challenge. For example, metaanalysis of clinical biomarker data from publications has been problematic because method details and validation parameters were missing to enable proper comparisons. It will be helpful if a consortium organization will handle data repository and communication (such as standardization of publication) of novel biomarkers. This is the goal of the Biomarker Consortium (launched in October 2006) by the FDA, Foundation for the National Institutes of Health, and Pharmaceutic Research and Manufactures of America through their various working groups.

## Acknowledgments

The authors are grateful to the Biomarker Subcommittee in the Ligand Binding Assay Bioanalytical Focus Group of the American Association of

Pharmaceutical Scientists for ideas and collaborations in the fit-for-purpose method validation concept. We thank Binodh DeSilva, Daniel Burns, Chad Ray, Han Gunn, and Lennie Uy for their contributions to the content and critical review of this manuscript.

## REFERENCES

Banks RE (2008). Preanalytical influences in clinical proteomic studies: Raising awareness of fundamental issue in sample banking. *Clin Chem*, 54;16–17.

Berger J, Wagner JA (2002). Physiological and therapeutic roles of peroxisome proliferator-activated receptors. *Diabetes Technol Ther*, 4:163–174.

Bjornsson TD (2005). Biomarkers: applications in drug development. *Eur Pharm Rev*, 1:17–21.

Bloom JC, Dean RA (eds.) (2003). *Biomarkers in Clinical Drug Development*. Marcel Dekker, New York.

Chu CS, Lee KT, Lee MY, et al. *(*2006). Effects of rosiglitazone alone and in combination with atorvastatin on nontraditional markers of cardiovascular disease in patients with type 2 diabetes mellitus. *Am J Cardiol*, 97:646–650.

*Code of Federal Regulations* (2001a). CFR Title 21, vol. 1, *Good Laboratory Practice for Nonclinical Laboratory Studies*, rev. Apr. 1.

*Code of Federal Regulations* (2001b) CFR Title 42, vol. 3, *Clinical Laboratory Improvement Amendment*, rev. Oct. 1

DeSilva B, Smith W, Weiner R, et al. (2003). Recommendations for the bioanalytical method validation of ligand-binding assays to support pharmacokinetic assessments of macromolecules. *Pharm Res*, 20:1885–1900.

FDA (2001). Guidance for industry on bioanalytical method validation: availability. *Fed Reg*, 66:28526–28527.

FDA (2004). Innovation or stagnation? Challenge and opportunity on the critical path to new medical products. http://www.fda.gov/opacom/hpview.html.

FDA (2006). Table of valid genomic biomarkers in the context of approved drug labels. http://www.fda.gov/cder/genomics/genomic_biomarkers_table.htm.

FDA CDER (2006). Using disease, placebo, and drug prior knowledge to improve decisions in drug development and at FDA. *Case Studies Across Companies Disease Models at FDA: Overview and Case Studies (Diabetes and Obesity)*, p. 65.

Ferguson RE, Hoschstrasser DF, Banks RE (2007). Impact of preanalytical variable on the analysis of biological fluids in proteomic studies. *Proteom Clin Appl*, 1:739–746.

Findlay JWA, Smith WC, Lee JW, et al. (2000). Validation of immunoassays for bioanalysis: a pharmaceutical industry perspective. *J Pharm Biomed Anal*, 21: 1249–1273.

Goodsaid F, Frueh F (2007). Biomarker qualification pilot process at the US Food and Drug Administration. *AAPS J*, 9:E105–E108.

Haffner SM (2007). Abdominal adiposity and cardiometabolic risk: Do we have all the answers? *Am J Med.* 120(9 Suppl 1):S10–S106.

Heeschen C, Dimmeler S, Hamm CW, Boersma E, Zeiher AM, Simoons ML (2003). Prognostic significance of angiogenic growth factor serum levels in patients with acute coronary syndromes. *Circulation*, 107:524–530.

ICH Guidelines. (1994). *Text on Validation of Analytical Procedures, Q2A*. International Conference on Harmonization, Geneva, Switzerland.

Khatami M (2007). Standardizing cancer biomarkers criteria: data elements as a foundation for a database. Inflammatory mediator/M-CSF as model marker. *Cell Biochem Biophys*, 47:187–198.

Krishnamurti U, Steffes MW (2001). Glycohemoglobin: a primary predictor of the development or reversal of complications of diabetes mellitus. *Clin Chem*, 47:1157–1165.

Kummar S, Kinders R, Rubinstein L, et al. (2007). Opinion: Compressing drug development timelines in oncology using phase '0' trials. *Nat Rev Cancer*, 7:131–139.

Lee JW, Ma H (2007). Specificity and selectivity evaluations of ligand binding assay of protein therapeutics against concomitant drugs and related endogenous proteins. *AAPS J*, 9:E164–E170.

Lee JW, Smith WC, Nordblom GD, Bowsher RR (2003). Validation of assays for the bioanalysis of novel biomarkers. In Bloom JC, Dean RA (eds.), *Biomarkers in Clinical Drug Development*. Marcel Dekker, New York, pp. 119–149.

Lee JW, Weiner RS, Sailstad JM, et al. (2005). Method validation and measurement of biomarkers in nonclinical and clinical samples in drug development: a conference report. *Pharm Res*, 22(4):499–511.

Lee JW, Devanarayan V, Barrett YC, et al. (2006). Fit-for-purpose method development and validation for successful biomarker measurement. *Pharm Res*, 23(2):312–328.

Lee JW, Figeys D, Vasilescu J (2007). Biomarker assay translation from discovery to clinical studies in cancer drug development: quantification of emerging protein biomarkers. In Hampton GM, Sikora K (eds.), *Genomics in Cancer Drug Discovery and Development*. Advances in Cancer Research. Elsevier, Amsterdam, pp. 269–298.

Lee JW, O'Brien P, Pan Y, et al. (2008). Development and validation of ligand binding assays for biomarkers. In Khan M, Findlay JWA (eds.), *Ligand-Binding Assays: Development, Validation and Implementation in the Drug Development Arena*. Wiley, Hoboken, NJ.

Lee YH, Pratley RE (2005). The evolving role of inflammation in obesity and the metabolic syndrome. *Curr Diabetes Rep*, 5:70–75.

Liu S, Tinker L, Song Y, et al. (2007). A prospective study of inflammatory cytokines and diabetes mellitus in a multiethnic cohort of postmenopausal women. *Arch Intern Med*, 167:1676–1685.

Mager DE, Wyska E, Jusko WJ (2003). Diversity of mechanism-based pharmacodynamic models. *Drug Metab Dispos*, 31:510–518.

Miller KJ, Bowsher RR, Celniker A, et al. (2001). Workshop on bioanalytical methods validation for macromolecules: summary report. *Pharm Res*, 18:1373–1383.

National Committee for Clinical Laboratory Standards (1999). *Evaluation of Precision Performance of Clinical Chemistry Devices: Approved Guideline*. CLSI Document EP5-A.

Pepe MS, Etzioni R, Feng Z, et al. (2001). Phases of biomarker development for early detection of cancer. *J Natl Cancer Inst*, 93:1054–1061.

Quraishi I, Rishi M, Feldman M, Wargovich MJ, Weber B (2007). Clinical validation of breast cancer biomarkers using tissue microarray technology. *App Immunohistochem Mol Morphol*, 15:45–49.

Rubin MA (2004). Using molecular markers to predict outcome. *J Urol*, 172(5 Pt 2): S18–S21.

Stoch SA, Wagner JA (2007). Biomarker analysis as a decision-making tool in drug discovery and development: implications for peroxisome proliferator–activator receptors. *Int J Pharm Med*, 21:271–277.

Sweep FC, Fritsche HA, Gion M, Klee GG, Schmitt M (2003). Considerations on development, validation, application, and quality control of immuno(metric) biomarker assays in clinical cancer research: an EORTC-NCI working group report. *Int J Oncol*, 23:1715–1726.

Shah VP, Midha KK, Dighe S, et al. (1992). Analytical methods validation: bioavailability, bioequivalence, and pharmacokinetic studies. *Pharm Res*, 9:588–592.

Shah VP, Midha KK, Findlay JWA, et al. (2000). Bioanalytical method validation: a revisit with a decade of progress. *Pharm Res*, 17:1551–1557.

Tholen DW, Linnet K, Kondratovich M, et al. (2004). *Protocols for Determination of Limits of Detection and Limits of Quantitation; Proposed Guideline*. NCCLS EP17-P.

van Doorn M, Kemme M, Ouwens M, et al. (2006). Evaluation of proinflammatory cytokines and inflammation markers as biomarkers for the action of thiazolidinediones in type 2 diabetes mellitus patients and healthy volunteers. *Br J Clin Pharmacol*, 62:391–402.

Vural P, Akgul C, Canbaz M (2006). Effects of hormone replacement therapy on plasma pro-inflammatory and anti-inflammatory cytokines and some bone turnover markers in postmenopausal women. *Pharm Res*, 54:298–302.

Wagner JA (2002a). Overview of biomarkers and surrogate endpoints in drug development. *Dis Markers*, 18:41–46.

Wagner JA (2002b). Early clinical development of pharmaceuticals for type 2 diabetes mellitus: from pre-clinical models to human investigation. *J Clin Endocrinol Metab*, 87:5362–5366.

Wagner JA, Williams SA, Webster CJ (2007). Biomarkers and surrogate end points for fit-for-purpose development and regulatory evaluation of new drugs. *Clin Pharm Ther*, 81:104–107.

Westgard JO (2003). Internal quality control: planning and implementation strategies, *Ann Clin Biochem*, 40:593–611.

Yu LR, Veenstra TD (2007). AACR-FDA-NCI cancer biomarkers collaborative. *Expert Rev Mol Diagn*, 7:507–509.

# MOLECULAR BIOMARKERS FROM A DIAGNOSTIC PERSPECTIVE

Klaus Lindpaintner, M.D., M.P.H.*
*F. Hoffmann–La Roche AG, Basel, Switzerland*

## INTRODUCTION

Molecular biomarkers may be useful in a number of applications, including mechanistic studies, preclinical safety and efficacy testing, early clinical pharmacodynamic assessment, dose-finding studies, and stratification for efficacy or safety and dose determination or dose guidance for a marketed pharmaceutical product. Although applications for a marketed drug product will pose the most demanding requirements in terms of the characterization of the test and its development into a commercially available product, all biomarker tests will need to satisfy certain requirements to be useful for derivation of the information they are intended to deliver.

A broader definition of the term *molecular biomarker* comprises any and all in vitro testing for biologically informative molecules in tissues or body fluids, be they for purposes of diagnosing disease, preventive health screening applications, prognostication of the natural history of a disorder, prediction of likely efficacy or safety parameters of a treatment, monitoring of treatment efficacy, or screening health care product manufacture or processing, including blood products.

*This communication represents the author's personal views and not necessarily those of any of his institutional or corporate affiliations, including F. Hoffmann–La Roche.

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

For purposes of the discussion below, we focus primarily on the diagnostic perspective as relevant to novel biomarker tests that accompany the development of a drug, in particular those that will require the development of a commercially available diagnostic test. Since such tests, which are often referred to as *companion diagnostics*, represent the most demanding scenario, they serve as an example for the spectrum of possible considerations that may be applicable to any diagnostic test—even though, in reality, most tests that add value at various points during the discovery or development of a drug will not need to live up to such scrutiny.

The fundamental question determining whether a biomarker is useful is the added information content that it conveys regarding the question it is intended to answer, and whether or to what extent this information adds value in the treatment of patients. Determination of these parameters will generally require characterization of the test on multiple levels that may be defined, broadly, as analytical and clinical. Analytical performance includes a number of metrological aspects, and clinical performance is largely subdivided into validation and utility.

## ANALYTICAL PERFORMANCE

Describing and characterizing analytical performance is very much oriented toward metrology. The general conventions and standards to consider include precision, trueness, and accuracy, along with a number of other more technical characteristics that describe the performance of a test.

### Analytical Precision and Imprecision

*Precision* is a measure of the random error or variability observed in measurement results and is a product of the sample handling and analytical process. Precision is typically expressed as the standard deviation found in sample results. Precision includes the definitions of repeatability and reproducibility. *Repeatability* is the precision with which, in a series of measurements carried out in sequence, results of a measurement agree with each other. *Reproducibility*, on the other hand, means the precision with which measurements of the same standard specimen carried out not in sequence, but under conditions that differ in some specified way (common examples are day to day, or lab to lab), agree with each other. The difference between actual and perfect agreement of these values is referred to as *imprecision*.

### Analytical Trueness and Bias

*Trueness* describes the closeness of agreement of an average value from a large series of measurements with a "true value" or an accepted reference value. The numerical value that represents the difference between the two is gener-

ally referred to as *bias*. Bias refers to systematic differences between measurement results and the true value of a parameter that is being measured. Bias in measurement results can be introduced in a number of ways, including through sampling, sample handling, sample preparation, matrix interference, cleanup, and determinative processes.

## Analytical Accuracy

*Accuracy* refers to a composite assessment that comprises both random and systematic influences (i.e., both precision and trueness). Its numerical value is the total error of measurement.

## Other Metrics Describing Analytical Performance

Aside from metrics for analytical accuracy, a number of parameters are of importance in defining and/or determining the utility of a particular test:

1. The *limit of detection* (LOD) is the smallest amount of an analyte that can reliably be detected by an assay, with a stated confidence limit. The definition includes a number of different *detection limits* that define different properties of the assay, including the lower limit of detection (LLOD), the instrument detection limit (IDL), the method detection limit (MDL), the (lower) limit of quantitation (LOQ or LLOQ), and the practical quantitation limit (PQL). The detection limits are estimated from the mean of a number of repeated measurements of the blank, the standard deviation of the blank measurements, and a defined confidence factor. The PQL is defined simply as about five times the MDL.

   In practical terms, the *lower limit of detection* is the lowest level of analyte that can be statistically distinguished from the blank (i.e., from background noise). LLOD is a function of the variability of the blank and the sensitivity of the assay. The LLOD is usually considered to be a value that is 3 standard deviations above the mean of the blank. Using this formula, the chance of misclassification is 7%. If 2 standard deviations are used, the chance of misclassification is 16%. Values below the LLOD should be reported as "less than the LLOD value" rather than as a finite value.

   The LLOD can commonly be distinguished from an additional variable that is not assay-specific but instrument-specific. Most analytical instruments produce a signal even when a blank (matrix without analyte) is analyzed. This signal is referred to as the *instrument detection level* or *instrument noise level*. The IDL is the analyte concentration that is required to produce a signal greater than three times the standard deviation of the noise level. Ideally, this would be equivalent to the assay's LLOD (as determined under optimal conditions), but is usually somewhat higher.

2. The *method detection limit* (MDL) is a metric that is similar to the IDL, but is based on samples that have gone through the entire sample preparation scheme prior to analysis, such as extractions, digestions, concentrations or dilutions, or fractionations, as well as interference by other components present in a complex matrix. The *recovery* of an analyte in an assay is the detector response obtained from an amount of the analyte added to and extracted from the biological matrix, compared to the detector response obtained for the true concentration of the pure authentic standard.

   Recovery pertains to the extraction efficiency of an analytical method within the limits of variability. Recovery of the analyte need not be 100%, but the extent of recovery of an analyte and of the internal standard should be consistent, precise, and reproducible. Recovery experiments should be performed by comparing the analytical results for extracted samples at three concentrations (low, medium, and high) with unextracted standards that represent 100% recovery.

3. The *limit of quantitation* (or quantification), also referred to as *lower limit of quantification* (LLOQ), is set at a higher concentration than the LLOD; in the statistical method, it generally is defined as 10 standard deviations above the mean blank value, thus presenting a greater probability that a value at the LLOQ is "real" and not just a random fluctuation of the blank reading. The lowest standard on the calibration curve should be accepted as the limit of quantification if the analyte response at the LLOQ is at least five times the response compared to the blank response, and if the analyte peak (the response) is identifiable, discrete, and reproducible with a precision of 20% and an accuracy of 80 to 120%.

   The LLOQ can differ drastically between laboratories, so another parameter for detection limit is commonly used, the *practical quantitation limit* (PQL). The PQL is commonly defined as 3 to 10 times the MDL.

4. *Selectivity* is the ability of an analytical method to differentiate and quantify the analyte in the presence of other components in the sample. For selectivity, analyses of blank samples of the appropriate biological matrix (plasma, urine, or other matrix) should be obtained from a sufficiently large and representative number of sources. Each blank sample should be tested for interference, and selectivity should be ensured at the lower limit of quantification (LLOQ).

5. The *coefficient of variation* (CV) is a normalized measure of dispersion of a probability distribution. It is defined as the ratio of the standard deviation to the mean. It is often reported as a percentage (%) by multiplying the calculation by 100. The coefficient of variation is useful because the standard deviation of data must always be understood in the context of the mean of the data. The coefficient of variation is a

dimensionless number, so when comparing between data sets with different units or wildly different means, one should use the coefficient of variation for comparison instead of the standard deviation.

## DIAGNOSTIC ACCURACY AND CLINICAL VALIDATION

*Diagnostic accuracy*, determined by a process commonly referred to as *clinical validation*, refers to the degree with which results of a test concur with what would be considered the current gold standard of clinical assessment of the question under study. This may be another biomarker (i.e., an established reference test) or—the real gold standard—a clinical outcome or endpoint, such as survival or death. Accuracy can be expressed through sensitivity and specificity, positive and negative predictive values, or positive and negative diagnostic likelihood ratios. Each measure of accuracy should be used in combination with its complementary measure: sensitivity complements specificity, positive predictive value complements negative predictive value, and positive diagnostic likelihood ratio complements negative diagnostic likelihood ratio [1,2]. All of these parameters are not intrinsic to the test and are determined by the clinical context in which the test is employed. A summary of the characteristics, and the strengths and weaknesses of these metrics, is presented in Table 1.

### Sensitivity

The *sensitivity* ("positivity in disease") of a test is the probability that it will produce a true positive result when used on a diseased population (compared to a reference test or gold standard assessment). For the purpose of biomarkers, this would translate to the probability of, or the frequency with which, a particular qualitative marker, or a certain value of a quantitative marker is observed in the group of people harboring the condition of interest. After inserting the test results into a table set up like Table 2, the sensitivity of a test can be determined by calculating TP/(TP + FN). High sensitivity corresponds to high negative predictive value and is the ideal property of a *rule-out test*.

### Specificity

The *specificity* ("negativity in health") of a test is the probability that a test will produce a true negative result when used on a nondiseased population (as determined by a reference test or gold standard assessment, such as clinical outcome). For the purpose of biomarkers, this would translate to the probability or the frequency with which a particular qualitative marker, or a certain value of a quantitative marker, is not observed in the reference group of nonaffected individuals. After inserting the test results into a table set up like

**TABLE 1 Summary of Characteristics for Various Parameters Used in Describing Clinical Test Performance**

| Metric | Definition[a] | Strengths | Weaknesses |
|---|---|---|---|
| Accuracy | (TP + TN)/N | Intuitive | Depends on prevalence |
| Sensitivity | TP/(TP + FN) | Does not depend on prevalence | Applies only to diseased persons |
| Specificity | TN/(TN + FP) | Does not depend on prevalence | Applies only to nondiseased persons |
| Positive predictive value | TP/(TP + FP) | Clinical relevance | Depends on prevalence |
| Negative predictive value | TN/(TN + FN) | Clinical relevance | Depends on prevalence |
| Positive likelihood ratio | [TP/(TP + FN)]/[FP/(TN + FP)] | Does not depend on prevalence | Applies only to positive tests |
| Negative likelihood ratio | [FN/(TP + FN)]/[TN/(TN + FP)] | Does not depend on prevalence | Applies only to negative tests |
| Odds ratio | (TP × TN)/(FN × FP) | Does not depend on prevalence; combines sensitivity and specificity | Values FP and FN errors equally; not intuitive |
| Area under curve | Area under ROC curve | Does not depend on prevalence; combines sensitivity and specificity | Lack of clinical interpretation |

[a]FN, false negative; FP, false positive; N, sample size; ROC, receiver operating characteristic; TN, true negative; TP, true positive.

Table 2, the specificity of a test can be determined by calculating TN/(TN + FP). High specificity corresponds to high positive predictive value and is the ideal property of a *rule-in test*.

## Positive Predictive Value

The *positive predictive value* of a test is the probability that a person is diseased when a positive test result is observed. In practice, predictive values should only be calculated from cohort studies or studies that legitimately reflect the number of people in that population who are diseased with the disease of interest at that time [i.e., when the prevalence (or incidence, for predictive tests) of the disorder is known]. This is because predictive values are inher-

**TABLE 2 Basic Model for Clinical Test Performance**[a]

| | Gold Standard (Reference Test Results or Clinical Outcome) | |
|---|---|---|
| New Test Results | + | – |
| + | TP | FP |
| – | FN | TN |

[a]TP, number of true-positive specimens; FP, number of false-positive specimens; FN, number of false-negative specimens; TN, number of true-negative specimens.

ently dependent on the prevalence of disease, called the *prior probability*. After inserting results into a table set up like Table 2, the positive predictive value of a test can be determined by calculating TP/(TP + FP).

### Negative Predictive Value

The *negative predictive value* of a test is the probability that a person is not diseased when a negative test result is observed. Again, this measure of accuracy should be used only if prevalence is available from the data. After inserting test results into a table set up like Table 2, the negative predictive value of a test can be determined by calculating TN/(TN + FN).

### Positive Diagnostic Likelihood Ratios

*Diagnostic likelihood ratios* (DLRs) are not yet commonly reported in peer-reviewed literature or in marketing information provided by test manufacturers, but they can be a valuable tool for comparing the accuracy of several tests to the gold standard, and they are not dependent on the prevalence of disease [3]. The *positive DLR* represents the odds ratio that a positive test result will be observed in a diseased population compared to the odds that the same result will be observed among a nondiseased population. After inserting test results into a table set up like Table 2, the positive DLR of a test can be determined by calculating [TP/(TP + FN)]/[FP/(FP + TN)], or it can also be expressed as sensitivity/(1 – specificity).

Useful tests will therefore have larger positive DLRs, and less useful tests will have smaller positive DLRs. An example interpretation of a positive diagnostic likelihood ratio equal to 5.0 is that for every 1% of nondiseased subjects that test as positive, 5% of the diseased subjects will test as positive. DLRs leverage pretest into posttest probabilities of a condition of interest, and there is some evidence that they are more intelligible to users.

## Negative Diagnostic Likelihood Ratios

The *negative DLR* represents the odds ratio that a negative test result will be observed in an diseased population compared to the odds that the same result will be observed among a nondiseased population. After inserting the test results into a table set up like Table 2, the negative DLR for a test can be determined by calculating [FN/(TP + FN)]/[TN/(FP + TN], or as false negative rate/true negative rate.

Useful tests will therefore have negative DLRs close to zero, and less useful tests will have higher negative DLRs. As an example, interpretation of a negative diagnostic likelihood ratio equal to 2.5 is for every one false negative, we observe 2.5 true negatives.

## Receiver Operating Characteristic

The determination of an "ideal" cutoff value of a quantitative test that delivers continuous values is almost always a trade-off between sensitivity (true positives) and specificity (true negatives). As both change with each cutoff value, it becomes difficult to determine which cutoff is ideal. A graph of sensitivity against (1 – specificity) is called a *receiver operating characteristic* (ROC) *curve*. The ROC curve offers a graphical illustration of these trade-offs at each cutoff for any diagnostic test that uses a continuous variable. Ideally, the best cutoff value provides both the highest sensitivity and the highest specificity, easily located on the ROC curve by finding the highest point on the vertical axis and the farthest to the left on the horizontal axis (upper left corner) (Figure 1). ROC curves compare sensitivity versus specificity across a range

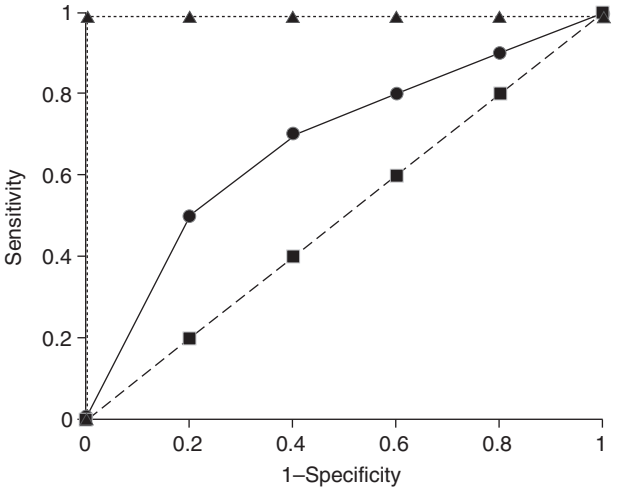


**Figure 1** Receiver operating characteristic curve.

of values of a quantitative biomarker for the ability to predict a dichotomous outcome. When the cutoff value for a continuous diagnostic variable is increased (assuming that larger values indicate an increased chance of a positive outcome), the proportions of both true and false positives decrease. These proportions are the sensitivity and 1 − specificity, respectively.

A perfect test would have sensitivity and specificity both equal to 1. If a cutoff value existed to produce such a test, the sensitivity would be 1 for any nonzero values of 1 − specificity. The ROC curve would start at the origin (0,0), go vertically up the *y*-axis to (0,1), and then go horizontally across to (1,1), as shown in Figure 1 (dotted line). A good test would be somewhere close to this ideal.

If a variable has no diagnostic capability, a test based on that variable would be equally likely to produce a false positive or a true positive: sensitivity = 1 − specificity, or sensitivity + specificity = 1. This equality is represented by a diagonal line from (0,0) to (1,1) on the graph of the ROC curve, as shown in Figure 1 (dashed line). The curve for a hypothetical test shown in Figure 1 (solid line) suggests that this marker does not provide a very good indication of the outcome of interest, but that it is better than a random guess.

The performance of a diagnostic variable can be quantified by calculating the area under the ROC curve (AUROC) [4]. The ideal test would have an AUROC of 1, whereas a random guess would have an AUROC of 0.5. The AUROC can be calculated as a sum of the areas of trapeziums. For example, in Figure 1, the area under the curve between points (0.2, 0.5) and (0.4, 0.7) is given by $(0.4 - 0.2) \times (0.5 + 0.7)/2 = 0.12$, or in other words, the difference between the *x*-values multiplied by half the sum of the *y*-values. In the example shown, the AUROC for the hypothetical marker is 0.68. This is interpreted as the probability that a patient who has the disease has a test value greater than that for a patient who does not. Like all summary measures, however, there are confidence intervals around this value that must also be taken into consideration for proper interpretation. The area under the ROC curve has become a particularly important metric for evaluating diagnostic procedures because it is the average sensitivity over all possible specificities.

## Precision of Diagnostic Accuracy

It is important to emphasize that as in other empirical studies, specific values of diagnostic accuracy are merely estimates. Therefore, when evaluations of diagnostic accuracy are reported, the precision of the sensitivity and specificity estimates or likelihood ratios should be stated. If sensitivity and specificity estimates are reported without a measure of precision, clinicians cannot know the range within which the true values of the indices are likely to lie. Evaluations of diagnostic accuracy should therefore be described with the inclusion of confidence intervals.

In the case of agreement with (an)other biomarker(s), the driving factor of such an evaluation may be the replacement of an existing, clinically validated

test with a new one that may be operationally or economically more attractive, or one that may have greater diagnostic precision.

## Odds Ratio and Relative Risk

The *odds ratio* is a way of comparing whether the relative odds of a certain event (e.g., of harboring a certain biomarker) is the same for two groups (e.g., those who have the disease as compared to those who do not). An odds ratio of 1 implies that the presence of the biomarker is equally likely in both groups (i.e., that the biomarker does not distinguish between the two). An odds ratio greater than 1 implies that the biomarker is more frequent or prevalent in the first group, and may thus be a useful test to distinguish between the two groups. From Table 2, the odds ratio would be calculated as (TP × TN)/(FN × FP). Odds ratios are applicable solely to retrospective case–control studies, whereas in prospective studies, randomized clinical trials, or cohort studies, calculation of relative risk is appropriate.

The *relative risk* (sometimes called the *risk ratio*) compares the probability of disease in the biomarker-positive and biomarker-negative groups, rather than the odds. It is calculated as TP/(TP + TN)/FN/(FN + TN), which is the quotient of the positive and negative predictive values. It is applicable only to prospective studies, randomized clinical trials, or cohort studies (i.e., to situations where the incidence/prevalence of the outcome is not biased by retrospective selection). However, in rare conditions (i.e., conditions with low incidence or prevalence), where both TP and FN are very small, the odds ratio asymptotically approaches the relative risk and is therefore sometimes substituted for the latter. For small probabilities, TP and FN are very small compared to TN and FP.

## Translating Different Metrics of Information Content

Academic biomarker research tends to be concerned primarily with the statistically determined reliability of a finding. Thus, if the association of a biomarker with a particular phenotype is reproducible, it is interpreted to convey real biological findings that may be of great interest to fundamental understanding of biological mechanisms, even if the magnitude of effect that can be seen under a particular set of experimental constraints is small. Similarly, the Geoffrey Rose paradox [5] explains that true, reproducible observations may be of great importance from the perspective of public health even if the overall magnitude of effect is modest. Thus, targeted modulation of variables that affect disease risk by small amounts may have important effects if applied to large populations and are thus commonly used to guide health policy. Academic epidemiological research evaluating associations of predictors for outcomes often uses odds ratios or relative risk as its primary parameters for reporting results, particularly in the field of genetic association studies. In complex polygenic diseases, the magnitude of effects commonly found for associations

for individual genetic variants generally ranges between odds ratios of 1 to 2. Demonstrating statistical significance for such modest-sized effects requires usually fairly large studies, particularly if multiple comparisons are made (as is the case with genome-wide association studies).

On the other hand, when tests are to be applied to clinical decision making in individual patients, information content with regard to the magnitude of association becomes critically important. Only reliable tests with an acceptably low rate of false positive and/or false negative results can responsibly be used in this setting. These performance parameters tend to be gleaned most directly from stating sensitivity and specificity (or positive predictive value and negative predictive value), and these are therefore the parameters commonly referred to in clinical diagnostics. In general, tests are not considered particularly useful if the area under the ROC curve does not exceed 0.8, which translates into balanced sensitivities and specificities of about 0.75. On the other hand, a balanced sensitivity and specificity of 0.75 translates into an odds ratio of about 9; if an effect of this magnitude is indeed present, it will be readily recognizable even in quite modestly sized studies.

## CLINICAL UTILITY

The term *clinical utility*, although used widely, is ill-defined. It is commonly used as a synonym for studies of clinical effectiveness and/or economic evaluations. The most basic definition of clinical utility refers to an estimation of the respective benefits and risks resulting from test use. Risks and benefits are, at this higher-level perspective, to be seen as encompassing both medical and economic connotations and considerations, even though the discussion of benefits and risks is often restricted to the former.

As health care payers are becoming increasingly cost-conscious, and reimbursement decisions are being more commonly influenced by medical–economic considerations, clinical utility is quickly becoming the overriding consideration with regard to the introduction of a companion or other diagnostic. This adds significantly to the burden of research and development expenses for diagnostic companies since reliable estimates of clinical utility will usually require prospective, controlled studies in which clinical endpoints are reached and where interventions are or are not guided by testing for the biomarker of interest.

### Medical Considerations

On the level of medical considerations, in the narrowest sense of the term, clinical utility refers to the ability of a screening or diagnostic test to prevent or ameliorate adverse health outcomes such as mortality, morbidity, or disability through the adoption of efficacious treatments conditioned on test results. A screening or diagnostic test in isolation does not have inherent

utility; because it is the adoption of therapeutic or preventive interventions that influence health outcomes, the clinical utility of a test depends on effective access to appropriate interventions, or to the way it can beneficially affect the choice of an intervention. This use of the term *utility* is consistent with standard practice in evidence-based medicine, which focuses on objective measures of health status to evaluate interventions. Clinical utility can more broadly refer to any use of test results to inform clinical decision making. Finally, in its broadest sense, the medical interpretation of clinical utility can refer to any outcome considered important to individuals and their families, as well as to other societal strata.

## Medical–Economic Considerations

If economic considerations are included in the definition of clinical utility, the question of whether the actual medical benefit derived from the use of the test indeed results in good value for the cost incurred becomes part of the consideration. *Value* in this context can be assessed by either cost–benefit or cost-effectiveness/utility analyses. In the case of a cost–benefit analysis, costs and benefits are expressed in the same units of measurement, usually monetary. Cost-effectiveness and cost-utility analyses compare the respective monetary cost of using or not using the test with defined clinical outcomes, either in physical terms or by using an index that also includes quality-of-life aspects.

As summarized in Table 3, when considering such *cost–benefit analysis* to evaluate the utility of a companion diagnostic that aids in the selection of a particular treatment, one would compare the aggregate costs of testing all eligible patients, of treating test-positive patients with the associated therapy, and of treating those that are test-negative with the conventional *standard-of-care* approach against the cost of using the standard of care, without any testing, on all patients. In a cost–benefit analysis, the net (pecuniary) gain or loss of the two approaches would be calculated based on direct medical and nonmedical, and on indirect (productivity) costs of either alternative. As an example, if the test is intended to increase the efficacy of a particular treatment, the cost–benefit ratio may be advantageous if the savings in productivity losses due to faster recovery from an illness based on the test-guided stratification of treatments offsets the added cost of testing and use of test-directed treatment alternatives (which may have a greater direct medical cost than the standard of care). Similarly, if a test is used to avoid serious adverse effects, the cost–benefit analysis may indicate advantages if the cost savings realized by avoiding the treatment of adverse reactions outweighs the cost of testing.

In a *cost-effectiveness analysis*, typically the overall cost per outcome unit (such as relapses or survival years), adjusted for a global quality of life in the case of a cost-utility analysis, is compared among the treatment alternatives. In the case of a companion diagnostic, the incremental cost of using the test in all patients and then treating only those in whom the test-driven therapy is either effective or safe may result in an acceptable cost-effectiveness ratio (as

**TABLE 3 Cost–Benefit Considerations for Companion Diagnostics**[a]

| | Efficacy Marker | | Adverse Reaction Marker | |
|---|---|---|---|---|
| | Direct Costs | Indirect Costs | Direct Costs | Indirect Costs |
| Cost of testing | ↑ | | ↑ | |
| **plus** | | | | |
| Cost of diagnostics-driven therapy in selected group | ↑ Novel, presumably more expensive drug | ↓ Faster return to productivity | — | ↓ No productivity loss due to ADR |
| **plus** | | | | |
| Cost of alternative therapy in de-selected group | — | — | — | — |
| **versus** | | | | |
| Cost of alternative therapy in all, no testing | ↓ Standard of care presumably cheaper | ↑ Later return to productivity | ↑ Cost of treating ADR | ↑ Productivity loss due to ADR |

[a]↑, increased cost; ↓, decreased cost.

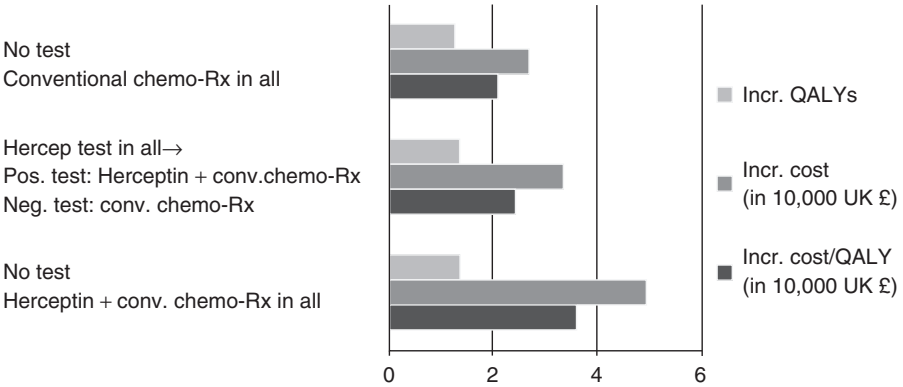


**Figure 2** Incremental cost-effectiveness ratio Herceptin/HerCep test.

expressed cost per quality-adjusted years of life) for an expensive novel treatment where a less discerning use of the novel treatment may be associated with an unacceptable ratio based on the inclusion of nonresponders. An example for such an analysis was recently conducted for the case of Herceptin and its companion diagnostic test (HerCep test) (Figure 2). In this example

[6], it could be clearly shown that the use of the test in all patients, followed by a targeted prescription of the new drug only to those in whom it was indicated based on the test, showed an acceptable incremental cost-effectiveness ratio, and one that was clearly superior to a more indiscriminate use of the drug.

Cost-effectiveness calculations are affected by numerous factors, including the prevalence of a (positive) test result which materially affects the choice of therapy, the test performance as outlined above in terms of analytical performance and clinical validity, the context with regard to indication and intended use (see below), the magnitude of the advantage over conventional treatments that the test-guided alternative therapy provides, by the cost of tests and different treatments, and last but certainly not least, by the reimbursement environment, including pricing policies that will differ from market to market.

Yet another way of looking at the question of clinical utility with regard to medical economics would be to calculate the *number needed to screen* (i.e., the number of patients that need to undergo the diagnostic test to achieve a particular outcome in at least one case). The outcome could either be defined as the prevention of a death or as an adverse event. Such a definition could be reached for both efficacy markers (where the choice of a certain drug based on screening may result in more successful treatment of eligible patients, and thus lives saved as compared with standard of care therapy) and—perhaps more obviously—for safety markers where the result would be the avoidance of an adverse event.

## CONTEXTUAL CONSIDERATIONS FOR COMPANION DIAGNOSTICS' CLINICAL UTILITY

While clinical performance, expressed as sensitivity and specificity, is of obvious importance for the utility of any test, it is important to recognize that with regard to companion diagnostics, a set of specific considerations apply depending on the intended use of the test. In this respect, both the clinical indication and the magnitude of expected benefit from the treatment, and the intended use to either enhance efficacy or avoid adverse reactions, are variables that will greatly influence the requirements for sensitivity and specificity, respectively, of the desired test (Table 4).

Thus, on one end of the spectrum, in life-threatening indications (e.g., oncology, HIV) and in the case of a drug that may be lifesaving, and for which feasible alternatives do not exist, an efficacy marker would be expected to show a high sensitivity (i.e., few false negatives, as they would result in inappropriate withholding of the treatment), but may be acceptable at lower specificity (false positives will result, for an otherwise safe drug, in unnecessary and ineffective treatment). In contrast, a safety marker in such an indication would be required to demonstrate high specificity (i.e., few false positives, which again would result in unnecessary withholding of the treatment), while

**TABLE 4 Contextual Considerations Regarding Companion Diagnostics**[a]

| | Efficacy Marker | | Adverse Event Marker | |
|---|---|---|---|---|
| Indication | Sensitivity | Specificity | Sensitivity | Specificity |
| Serious or life threatening | ⬆ | ⬇ | ⬇ | ⬆ |
| Less serious or trivial | ⬇ | ⬆ | ⬆ | ⬇ |

[a]⬆, High requirements, important; ⬇, low requirements, less important.

sensitivity is less important [use of the marker would still be expected to result in fewer adverse reactions (ARs) than would otherwise be encountered; in indications where ARs are often viewed as necessary trade-offs].

On the other end of the spectrum, in more trivial indications where treatment is less urgent, the requirement for a safety marker would be very high sensitivity, as a false-negative test would result in the occurrence of an AR, which in this setting is much less acceptable than in a life-threatening illness. Specificity is less important, as "wrongful" withholding in a condition of modest severity is clearly less of an issue. Conversely, again, an efficacy marker would be expected to show high specificity: If the added cost of the test is to be acceptable in such a condition, it would presumably need to be justified more by cost–benefit considerations that would be set off by inappropriate inclusion in the treatment group of patients with false-positive test results.

One may view these considerations from the standpoint of that old medical principle: first, do no harm. In serious disease, the threat may come primarily from the illness, so harm would more likely result from inappropriate withholding of treatment based on a false test result. In contrast, in less serious disease, harm may, rather, come from inappropriate administration of a treatment, and one would perhaps be more aligned with the Hippocratic principle by avoiding noncritical treatments.

## CO-DEVELOPMENT OF A NEW MEDICAL ENTITY AND COMPANION DIAGNOSTIC

If the label of a new medicine is to include the obligatory use of a test, the manufacturer needs to take care that drug and test are developed in tandem and receive regulatory approval in such a fashion that they can reach the market simultaneously. Authorities are paying increasing attention to the challenges that this process may present. The U.S. Food and Drug Administration has issued a concept paper on "Drug–Diagnostic Co-development" (http://www.fda.gov/Cder/genomics/pharmacoconceptfn.pdf) that specifically addresses this issue. A number of working groups, including

the International Conference on Harmonization, are also currently addressing aspects related to this issue, from coining unified terminology and language around biomarkers and companion diagnostics, to formal issues relating to regulatory submissions and the evidentiary standards that will be required to accept a companion diagnostic as valid and useful.

## CONCLUSIONS

At the transition of a biomarker discovery and assessment program into the development for the use of a biomarker as a clinically applicable decision tool, a number of considerations that are not applicable to the discovery stage must come into focus. Many of them represent well-established concepts in analytical metronomy but are not necessarily familiar to investigators at the cutting edge of biomedical research. It is critical that these concepts be considered carefully, as this will directly influence the feasibility of any biomarker test that is to be used in the setting of clinical practice. Although many of the considerations are of a technical nature, the ultimate driver of such workability is, of course, rooted in biology and in the information content conveyed by the data.

## REFERENCES

1. Jaeschke A, Guyatt GH, Sackett DL (1994). Users' guides to the medical literature: III. How to use an article about a diagnostic test. A. Are the results of the study valid? *JAMA*, 271:389–391.
2. Jaeschke A, Guyatt GH, Sackett DL (1994). Users' guides to the medical literature: III. How to use an article about a diagnostic test. B. What are the results and will they help me in caring for patients? *JAMA*, 271:703–707.
3. Halkin A, Reichman J, Schwaber M, Paltiel O, Brezis M (1998). Likelihood ratios: getting diagnostic testing into perspective. *Q J Med*, 91:247–258.
4. Faraggi D, Reiser B (2002). Estimation of the area under the ROC curve. *Stat Med*, 21:3093–3106.
5. Rose G (1981). Strategy of prevention: lessons from cardiovascular disease. *Br Med J*, 282:1847–1851.
6. Elkin EB, Weinstein MC, Winer EP, et al. (2004). HER-2 testing and trastuzumab therapy for metastatic breast cancer: a cost-effectiveness analysis. *J Clin Oncol*, 22(5):854–863.

# STRATEGIES FOR THE CO-DEVELOPMENT OF DRUGS AND DIAGNOSTICS: FDA PERSPECTIVE ON DIAGNOSTICS REGULATION

Francis Kalush, Ph.D., and Steven Gutman, M.D., M.B.A.*
*U.S. Food and Drug Administration, Rockville, Maryland*

## INTRODUCTION

The U.S. Food and Drug Administration (FDA) introduced its critical path initiative in 2004 as an effort to stimulate and facilitate the scientific process for the development of new drugs, biological products, and medical devices [1]. The initiative was grounded in the sobering reality that between 1993 and 2003, despite a doubling in research and development spending by both industry and the National Institutes of Health (NIH), the FDA observed a decrease in the number of new drug applications (NDAs) and an erratic but flat curve in the approval of new molecular entities. Of particular concern was the dramatic increase in the developmental cost of new drugs and the high rate of new drug failures occurring relatively late in the developmental life cycle (phase III or phase IV drug failures). The critical path was developed to ensure that FDA fosters a collaborative and robust scientific environment to

*The views presented in this chapter do not necessarily reflect those of the U.S. Food and Drug Administration.

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

help strengthen the assortment of tools that might be used to bring new medical products to the marketplace in an expedited manner. Since its initiation in 2004, a new office, the Critical Path Office, has been created, a comprehensive and ambitious opportunities list has been generated [9], and numerous collaborative projects and initiatives have been born, all with the singular goal of using good science to address the challenge of timely and cost-effective product development.

With the unveiling of the complete human genomic map in 2001–2003 [2,3] and the development of sophisticated new technologies for molecular diagnostics (including use of microarrays and bioinformatics to create integrated multiplex tests with complex molecular signals), considerable attention has been focused on biomarkers (especially those with a pharmacogenomics base) to help improve drug candidacy and to streamline drug studies.

The introduction of a diagnostic biomarker to refine drug discovery has obvious appeal. Although biomarkers in a broad sense can include both imaging technologies and in vitro diagnostic devices (IVDs), for the sake of this chapter the focus is on the latter. When used clinically, IVDs are subject to the same regulations as those applied to all other medical devices. Both the science and regulation required for proper deployment of these companion diagnostics products is quite different than the science and regulation of drugs. As a result, to bring closely linked products to the market in a timely manner, both FDA and sponsors must work collaboratively to address what may often be dual device and drug regulatory requirements.

## FDA REGULATION OF IN VITRO DIAGNOSTIC DEVICES

A *medical device* is defined as an "instrument, apparatus, implement, machine, contrivance, implant, in vitro reagent, or other similar or related article, including a component part, or accessory which is: recognized in the official *National Formulary*, or the *United States Pharmacopoeia*, or any supplement to them, intended for use in the diagnosis of disease or other conditions, or in the cure, mitigation, treatment, or prevention of disease, in man or other animals, or intended to affect the structure or any function of the body of man or other animals, and which does not achieve any of its primary intended purposes through chemical action within or on the body of man or other animals and which is not dependent upon being metabolized for the achievement of any of its primary intended purposes" by the Federal Food, Drug, and Cosmetic Act [21 United States Code [321] (h)]. *In vitro diagnostics* (IVDs) are defined as reagents, instruments, and systems intended for use in the diagnosis of disease or other conditions, including a determination of the state of health, in order to cure, mitigate, treat, or prevent disease or its sequelae. Such products are intended for use in the collection, preparation, and examination of specimens taken from the human body [21 CFR 809.3(a)]. IVDs are devices, as defined in section 201(h) of the act, and may also be

biological products subject to section 351 of the Public Health Service Act. While drugs are regulated in the Center for Drug Evaluation and Research (CDER) under the provisions of the state law, IVDs are regulated in the Center for Devices and Radiological Health (CDRH) under provisions of the Medical Device Amendments of 1976.

Like the regulation of all medical devices, the regulation of IVDs by the FDA is risk-based with devices classified into low-risk (class I, e.g., adjunctive immunohistochemical stains used in conjunction with standard microscopic analysis to subtype tumors), moderate-risk (class II, e.g., prognosis, monitoring in already diagnosed cancer patients), or high-risk (class III, e.g., cancer diagnosis, screening) categories. The FDA regulatory program is comprehensive and includes requirements for registration and listing of products, for high-quality production using good manufacturing practices, and for postmarket reporting of adverse events. For some class I, most class II, and all class III devices, FDA review is required before a new medical device can enter the marketplace.

- *Class I devices:* typically present minimal potential for harm to the user and the person being tested. They are subject to general controls, which include registration and listing, labeling, and adverse event reporting requirements [section 513(a)(1)(A) of the act]. Most class I devices are exempt from premarket notification (see definition below), subject to certain limitations found in section 510(l) of the act and in 21 CFR 862.9, 864.9, and 866.9. An IVD example of class I devices is complement reagent (21 CFR 866.4100).
- *Class II devices:* devices for which general controls alone are insufficient to provide reasonable assurance of their safety and effectiveness and for which establishment of special controls can provide such assurances. Special controls may include special labeling, mandatory performance standards, risk mitigation measures identified in guidance, and postmarket surveillance [section 513(a)(1)(B) of the act]. Most class II devices require premarket notification. IVD examples of class II devices include glucose test systems (21 CFR 1345), antinuclear antibody immunological test systems (21 CFR 866.5100), and coagulation instruments (21 CFR 864.5400).
- *Class III devices:* devices for which insufficient information exists to provide reasonable assurance of safety and effectiveness through general or special controls. Class III devices are usually those that support or sustain human life, are of substantial importance in preventing impairment of human health, or which present a potential, unreasonable risk of illness or injury [section 513(a)(1)(C) of the act]. Most class III devices require premarket approval (PMA), defined below. IVD examples of these include automated PAP smear readers, nucleic acid amplification devices for tuberculosis, and total prostate-specific antigen (PSA) for the detection of cancer (21 CFR 866.6010).

Laboratory-developed tests (LDTs), in-house tests, and "home-brew" tests are tests developed in a laboratory for use only in that laboratory. Historically, FDA has applied enforcement discretion to most LDTs and not regulated them actively. Instead, FDA has generally chosen to regulate the components of LDTs, such as instruments and reagents. Considerable discussion has occurred over the past 10 years on the existence of two parallel but noncongruent regulatory pathways to market (IVD "kits" vs. LDTs); these have been the subject of a number of reports, including the Task Force on Genetic Testing and the Secretary's Advisory Committee on Genetic testing [4]. The Secretary's Advisory Committee on Genomics, Health and Society (SACGHS) is in the process of discussing the issue of genetic tests with regard to potential scientific or regulatory gaps and is currently seeking public comment on a draft report to the Secretary of Health and Human Services (HHS) on the oversight of genetic testing [5].

## Major Elements of a Submission

Major elements of a submission can be found by browsing the Office of In Vitro Diagnostic Devices and Safety (OIVD) 510(k) data templates posted on the OIVD Web page [6]. These summarize information related to specific cleared products, including the intended use/indications for use, analytical and clinical validation information, device description including platform and software information, information on instrument and software validation when applicable, and labeling (package insert information). IVD labeling is unique among medical devices since it is described specifically in regulation [21 CFR 809.10(b)]. For PMAs, information describing manufacturing, design controls, and adherence to quality system regulations are also needed.

The intended use should specify what analyte the test measures, the clinical indication for which the test is to be used, and the target population for which the test is intended. It should also indicate whether the test is qualitative and quantitative.

Analytical validation includes precision (repeatability and reproducibility), accuracy, limit of detection, interferences, cross-reactivity, software, performance around the cutoff, carryover, cross hybridization, sample preparation/conditions, and assay limitations. FDA recognizes dozens of standards by the Clinical Laboratory Standards Institute (CLSI) to assist companies in developing this information. A list of recognized standards [10] may also be found on the OIVD Web page.

Clinical validation should be established in appropriate clinical studies to support the indication for use and claims of the device. These studies generally need to be provided by the sponsor or, if available, sponsors may also cite applicable clinical literature. For clinical literature to be acceptable for product clearance or approval, the quality must be carefully assessed and the link between the published studies and the device being reviewed must be carefully established. In instances where tests are developed based on sequential use of

training sets followed by independent test validation, the FDA encourages that rigorous and meticulous attention be paid to the training sets used for new diagnostic markers. However, FDA review is focused on the independent validation that supports the merits of the device itself based on feasibility studies and development of a clear hypothesis in a defined population under established conditions of use.

When the system includes software, software documentation including software design, hazard analysis, and complete verification and validation should be provided [11]. Labeling needs to be detailed sufficiently to satisfy the requirements of 21 CFR 807.87(e). Final labeling for in vitro diagnostic devices must comply with the requirements of 21 CFR 809.10 before the device is introduced into interstate commerce. Among the particularly important components of required labeling are an intended use, device description, directions for use, quality control, precautions, warnings and limitations, performance characteristics, interpretation of results, and values expected.

## REGULATORY TOOLS AND SOLUTIONS FOR PHARMACOGENOMICS MARKERS

Over the past several years, the Office of In Vitro Diagnosis (OIVD) has made flexible use of its regulatory tools to help bring important new diagnostics to market. These include the use of the preinvestigational device exemption (pre-IDE process), expedited review, real-time work interactions, and *de novo classification* of new devices into classes II and I under section 513(f)(2) of the FDCA. The purpose of a pre-IDE protocol review with the FDA is to assist the sponsor and familiarize FDA with cutting-edge technology in advance of the submission. Furthermore, assistance may be needed in defining possible regulatory pathways, study design and analysis of complex data, and statistical approaches before initiation. This is a work process performed by review staff in the Center for Devices and Radiological Health at no cost to sponsors with a 60-day time line. OIVD now processes almost 300 pre-IDEs per year and finds that use of this process decreases uncertainty in the review process, provides an early opportunity for question and answers, and drives the least burdensome approach for sponsors.

Both 510(k)s and PMAs for cutting-edge IVDs may also be designated for expedited review if the device is expected to provide unique public health benefits. Expedited review ensures that at every stage in the review process, the device under consideration goes to the front of the review queue. Even so, the outcome of an expedited review, like the outcome of all reviews, will depend ultimately on the quality of the submission.

A subcategory of PMA supplements is eligible for what is known as *real-time review*, which calls for rapid interactive review between the FDA and sponsors. Although OIVD gets only a small number of real-time applications

per year, it has found rapid interactive review with sponsors to be so successful that it has been applied more broadly to review pre-IDEs and IDEs and processing of some devices subject to expedited review. In some cases, such interaction has facilitated review of cutting-edge new products in two weeks or less.

OIVD has also made use of the de novo classification process authorized by section 513(f)(2) of the FDCA. A device that measures a new analyte or has an intended use for which no submission has previously been cleared or classified as a class I or II device is automatically assigned a class III designation, which would require a PMA submission. Under the de novo process, a sponsor that receives a determination that its device is not substantially equivalent may petition for classification directly into class II or class I if it can provide sufficient scientific basis for the FDA to conclude that a reasonable assurance of the safety and effectiveness of that device can be assured without a PMA. In this case, the new device may be tested by the manufacturer in the same way as it would be for a PMA to establish that the test is safe and effective, but the review is performed following the administrative procedures commonly applied to the 510(k) process. As a result, many PMA requirements, including premarket review of manufacturing processes, premarket inspection of manufacturing facilities, and requirements for annual reports, are all not required. Mandatory time lines for review are also shorter (90 vs. 180 days).

For de novo classification requests, OIVD has been able to use clinical data and supportive literature information, current U.S. clinical guidelines and practices, as well as product-specific data generated by the sponsor to conclude that classification into class II is appropriate and to establish special controls. This type of classification occurs more quickly than other classification processes and more quickly than review of a PMA, and the newly classified device not only may be legally marketed but becomes the first predicate device for a new classification, creating a means for subsequent similar devices to reach the market through the 510(k) process. Devices to detect cytochrome P450 DNA polymorphisms, cystic fibrosis mutations, West Nile antibody, breast cancer gene expression signatures, and the hemagglutinin 5 (H5) subtype of influenza are among the products cleared using the 510(k) de novo process and the establishment of special control guidances.

## REGULATORY FRAMEWORK TO ACCOMMODATE CO-DEVELOPING DRUG DIAGNOSTICS

The lifecycle to obtain marketing approval for the development of a new drug is outlined in FDA law and regulation and progresses through a well-defined multistep process: basic research, prototype design or discovery, preclinical development, clinical development (phases I, II, and III), FDA filing/approval/launch, and postlaunch (phase IV).

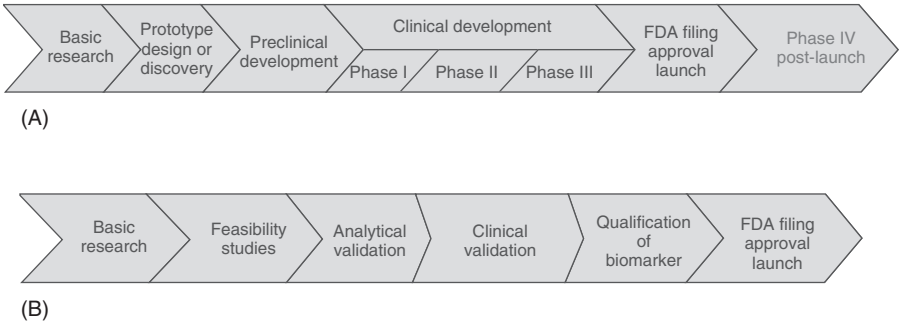


**Figure 1** (A) Drug and (B) device development pathways.

The development of a new diagnostic also goes through a well-defined multistep process: basic research, feasibility studies, analytical validation, clinical validation, and FDA approval/clearance. The developmental pathways for a drug and device are illustrated in Figure 1. When use of a diagnostic test is identified as important in selecting which patients receive or avoid a new drug or are to be given a higher or lower dose of a drug, the efficacy/safety of the drug becomes inextricably linked to the effectiveness of the diagnostic. Drug performance will only be as good as the ability of the diagnostic to properly select patients for that drug treatment.

Drug performance is judged based on an assessment of the drug to meet appropriate clinical endpoints or responses with a reasonable safety profile. Diagnostic studies are based on demonstrating that a test properly identifies an outcome or target of interest. When the diagnostic is used to detect drug response, the outcome of interest becomes the drug response or avoidance of drug toxicity. The test parameters of interest in the context of co-development with a new drug do not differ from those of ubiquitous interest in the study of a new diagnostic.

Standardized techniques to establishing test performance have been promoted in the literature and in FDA guidances. Clinical parameters of importance include sensitivity—in this case the ability of the test to identify patients who will exhibit the desired drug response of interest—and specificity—in this case the ability of the test to identify patients who will not demonstrate the drug response of interest. Alternative useful techniques include a definition of the predictive value of a positive result (the fraction of test positives who respond), the predictive value of a negative result (the fraction of test negatives who do not respond), or the likelihood ratio of drug response after testing. The hazard ratio (HR) in prospective studies/analysis and the odds ratio (OR) in retrospective analysis could be used when interpreting probabilities in a different context of use.

## CHANGES IN STRATEGIC THINKING

Because there are now two separate medical products under investigation which may have different life cycles, companies involved in co-development are facing new challenges and may have to create complex arrangements that allow for regulatory and business issues to be addressed. FDA is committed to approaching products being submitted as part of a co-development paradigm to assuring that good science and appropriate regulatory thresholds be met but also bringing to bear a regulatory framework that is as flexible as possible.

Companies may either wish to rethink their business models, shift strategies to accommodate new types of products, and change administrative structures to allow drugs and devices to be developed either in parallel or at least in a scientifically streamlined manner. Moreover, companies may be inclined to develop business alliances, seek out business partners and financial solutions to developmental costs that will encourage product development, and demonstrate the proper premium on testing so that value-based reimbursement becomes possible. Some authors have called for health system reforms that promote value-based, flexible reimbursement for innovative diagnostics and therapeutic products, and others have highlighted potential value to creating stronger economic incentives for the development of personalize medicine [7]. As health care costs continue to skyrocket, the role of FDA regulation appears to be a less important barrier to development of product than obtaining reimbursement or widespread clinical acceptance. Increasingly, both third-party payers and health care gatekeepers are demanding studies grounded in evidence-based medicine before making coverage decisions or before choosing to order new tests. There is also heightened interest among third-party payers to performing complex cost-effectiveness analysis to help make payment determinations.

## PATHS IN CO-DEVELOPMENT

There are several major paths of co-developing a drug and associated test: (1) drug and test are developed in parallel (true co-development); (2) drug is developed first, followed by the diagnostic test(s) development (drug rescue in case there are safety issues discovered in phase IV); and (3) the already existing test can be used for drug(s) selection (patient enrichment/stratification). Multiple tests may be developed for the selection of a certain drug, either in parallel with the drug or at a later date. Figure 2 illustrates major drug diagnostics collaboration scenarios.

In the end, regardless of the path taken, co-development results in cross-labeling/relabeling of drug and/or the diagnostics. FDA drug package inserts frequently cite the relevant in vitro diagnostic test(s), and FDA cleared/approved devices reference the targeted drug(s). Table 1 presents current

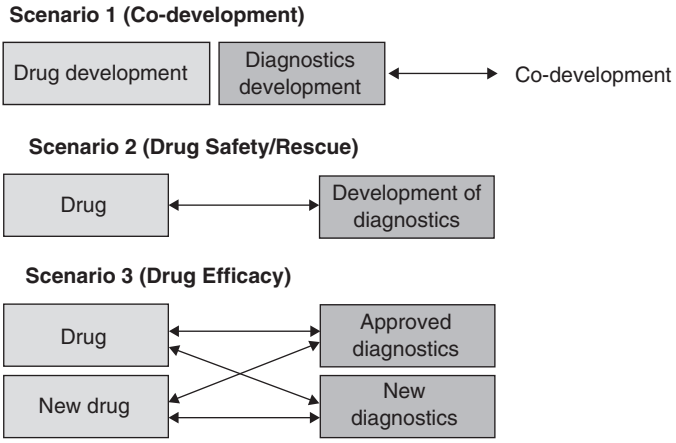


**Figure 2** Drug-diagnostics collaboration scenarios.

**TABLE 1 Current Pharmacogenomics Examples**

| Drug | Efficacy Test | Safety Test |
|---|---|---|
| Trastuzumab (Herceptin)[a] | HER-2/neu IHC and FISH[b] | |
| Imatinib mesylate (Gleevec)[a] | C-kit IHC[b] | |
| Cetuximab, Panitumumab[a] | EGFR IHC[b] | |
| Imatinib mesylate (Gleevec)[a] | bcr/abl or 9:22 tr | |
| Gefitinib | EGFR mutations | |
| Irinotecan[a] | — | UGT1A1 mut[b] |
| 6MP and azathioprine[a] | — | TPMT |
| 5-HT3R antagonist, codeine derivative | — | CYP450(2D6) mut[b] |
| Warfarin[a] | — | CYP2C9&VKORC1[b] |

[a]FDA drug package insert cites the relevant test.
[b]FDA cleared/approved device in reference to the drug.

examples. This document is intended primarily to provide direction for the development of diagnostics that are used to select or avoid drugs (predictive tests). A predictive test/biomarker would be defined as a single trait or a signature of traits that separates different populations with respect to the outcome of interest in response to a particular (targeted) treatment (e.g., response to Herceptin, Erbitux, Panitumumab).

Predictive tests could be categorized into three classes: (1) a predictive test for efficacy, to identify patients most likely to respond beneficially to a

targeted treatment; (2) a predictive test for safety, to identify patients most likely to respond adversely to a targeted treatment; and (3) a predictive test for dosing, to identify a patient's specific dosing to optimize the benefit of or minimize the risk of a targeted treatment.

Following Sargent et al. [8], a predictive biomarker provides information about the effect of treatment and is useful for selecting patients for the drug. In this document, diagnostic tests are assumed to be predictive biomarkers. However, much of the discussion also applies to a prognostic test or biomarkers defined as follows: A prognostic biomarker (e.g., risk for cancer recurrence) is a single trait or a signature of traits that separates different populations with respect to the risk of an outcome of interest in the absence of treatment or despite nontargeted "standard" treatment.

## Co-Development and Phase I and II Drug Studies

Ideally, when a drug and a diagnostic are to be used in combination, the development of the drug and diagnostic should be coordinated as much as possible. For the drug candidate, CDER requires phase I studies to demonstrate drug safety and phase II studies to confirm proof of concept, determine dosing and expand on drug safety. In cases where biomarkers are being studied for possible use with a drug, the diagnostic work is generally exploratory but, ideally, should include initial biomarker validation.

Phase I and II drug studies do offer opportunities to gather preliminary data about associations between biomarkers and drug use. This information may include descriptive data on prevalence of the biomarker, association with adverse events or the safety profile being identified with early drug use, and/or information on possible ability of the test to predict drug response. In addition, the information gathered helps improve benefit–risk analysis at an earlier stage of development. A number of interesting proposals have been made to streamline the path to market by expanding phase II studies to allow better selection and characterization of biomarkers for drug use. Jiang et al.'s [12] proposed design combines a test for overall treatment effect in all randomly assigned patients with the establishment and validation of a cut point for a prespecified biomarker for identifying the sensitive subpopulation. The procedure provides prospective tests of the hypotheses that the new treatment is beneficial for the entire patient population or that it is beneficial for a subset of patients defined by the biomarker. The good news is that this allows an opportunity to introduce a drug without or with a biomarker as circumstances may dictate. However, the bad news is that each separate hypothesis cannot be performed without a $p$-value penalty, so a recommendation is made that testing for the drug affect in the all-comers population be with $p = 0.04$ and, if needed, follow-up testing in the biomarker positive population occurs with a $p = 0.01$. Proposals have also been made to find mechanisms of adaptive design applying Bayesian statistics to allow phase II and III studies to be performed in a more seamless manner.

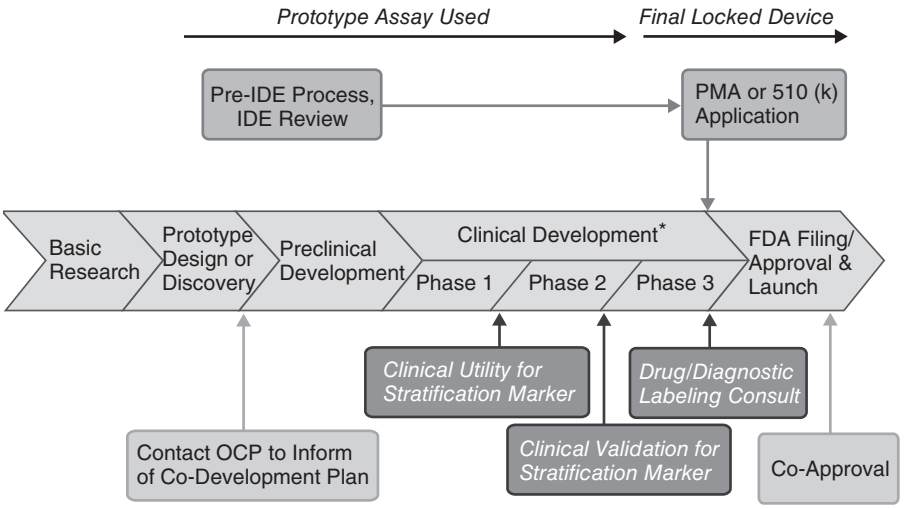


*Adapted from Drug-Diagnostic Co-Development Concept Paper (draft).

**Figure 3** Co-development strategy timing.

## Co-Development and Phase III Drug Studies

If a diagnostic is actually intended for use in patient selection, the safety and efficacy of the drug may in fact depend on the ability of the test to identify patients properly. Clinical trials for drug approval are generally multicenter. Therefore, during co-development the test needs to be standardized and a consensus methodology for test validation among the centers should be established. Ideally, analytical validation should be performed prior to phase III trials, so that clinical validation can be established using an analytical robust test.

When the drug approval is contingent on use of the diagnostic, the diagnostic should optimally be studied either at the time of the original phase III study or in a follow-up phase III–like study before drug approval can be obtained. These studies would most commonly be based on an appropriately designed and powered prospective blinded and randomized clinical trial usually associated with drug approvals and validated in follow-up trials. Co-development strategy timing is illustrated in Figure 3.

## Co-Development After Drug Approval

In those cases in which the drug approval is not contingent upon the use of the diagnostic or when changes in drug labeling for improving safety or effectiveness profiles are being requested, studies may be performed during phase IV drug studies or at any interval following drug approval (e.g., postmarket

commitments). In some cases, prospective studies may be needed; in others, use of banked samples from prospective studies, retrospective studies, and/or cross-sectional studies may be sufficient to support diagnostic claims.

If it is not possible to conduct coordinated studies of the drug and diagnostic, the phase III clinical trial may represent the only opportunity available to obtain and bank prospectively collected samples for future diagnostic test study. Particular attention should be paid to ensuring that samples are collected properly with appropriate informed consent to support future bridging studies (e.g., platform change), diagnostic studies, and evaluation of the analyte of interest. The samples collected should be banked in storage conditions that maintain sample integrity. Concerns about any bias, missing samples, or selection in accrual of banked samples should be addressed appropriately.

## INTERCENTER REVIEW CONSIDERATIONS IN CO-DEVELOPMENT

Overcoming potential regulatory barriers to speed up introduction of innovative technologies and products, while ensuring their safety and effectiveness, is a major challenge. Co-development, to be most efficient, requires coordination and harmonization between FDA reviewing centers. The success of this coordination in part depends on the ability of all stakeholders to foster an increase in the dialogue between the FDA and industry and the ability of the FDA to engage sponsors in collaborative discussion early in the process as science develops.

Co-developed products that would be used together may or may not be combination products as defined in 21 CFR 3.2(e).[1] FDA anticipates that many therapeutic drug and diagnostic test products will be marketed separately. For the purposes of this document, co-development refers to products that raise development issues that affect both the drug therapy and the diagnostic test, regardless of their regulatory status as a combination product or as a noncombination product. For example, when co-developed products are considered together, unique questions may arise that would not exist for either product alone. Scientific or technological issues for one product alone may be minimal, but they may have substantial implications for the other product. Also, postapproval changes in one may affect the safety and effectiveness of the other. The FDA has recently established the Office of Combination Products (OCP)[2] to work with sponsors and FDA work groups to ensure that proper regulatory tools are applied to products used in combination and to assist in addressing problems that joint reviews may encounter.

If the diagnostic becomes integral to approval of the drug, diagnostic approval is needed in parallel with the drug. FDA review centers (CDER and CDRH) are willing to work with single sponsors or collaborating sponsors

(drug–diagnostics companies) to coordinate review processes and have in the past had conjoint teams, tandem panel meetings, and same-day approval of linked products. If the diagnostic has broad use not specifically linked to the drug or if the drug has use not specifically linked to the diagnostic, approvals would follow usual CDER or CDRH procedures and timelines.

In some cases, development of a diagnostic may follow development of a drug and be used to refine or improve the safety or effectiveness profile of the drug. CDER and CDRH will work collaboratively with OCP to determine best practices for addressing these situations in a timely manner.

## CONSULTATION PROGRAMS

The parallel development of a drug and a diagnostic is a relatively new aspect of drug development and calls for careful coordination. FDA has three mechanisms for an early interaction with companies about the use of drug–diagnostic combinations. CDER has developed a new approach for data review referred to as *voluntary genomic data submissions* (VGDSs). These submissions can be used throughout this development process to present and discuss data with the agency that are not used for regulatory decision making but could have an effect on the overall development strategy. Such data submitted voluntarily will not be used for regulatory decision making by the FDA and is not included in the evaluation of an investigational new drug (IND), investigational device exemption (IDE), or market application. Both CDER and CDRH representatives participate in VGDSs and will work with the sponsor to identify when pre-IDE interactions for development of the test in conjunction with the drug are advisable.

CDER has also developed a pre-IND process for early interactions with a sponsor to improve a development program and expedite entry into clinical trials. Pre-IND advice may be requested for issues related to data needed to support the rationale for testing a drug in humans; the design of nonclinical pharmacology, toxicology, and drug activity studies, including design and potential uses of any proposed treatment studies in animal models; data requirements for an investigational new drug (IND) application; initial drug development plans; and regulatory requirements for demonstrating safety and efficacy. Pre-IND interactions should be considered as preliminary communications based on early development information and will generally take the form of written comments that may be supplemented by teleconferences or meetings as needed and appropriate.

The pre-IDE process is the term CDRH uses for a voluntary system available to sponsors for protocol review and has already been described in this chapter. Additional regulatory tools that CDRH uses to support introduction of new molecular diagnostics and co-development products are expedite review submissions (the front of the queue), the de novo process for regulating new unclassified devices (first of a kind) and real-time reviews by reducing

regulatory working time and improving the quality of results (see the Section "Regulatory Tools and Solutions for Pharmacogenomics Markers).

By employing these types of review processes, the intent of CDER and CDRH is to provide useful information to sponsors as they work on co-developed products. The FDA is willing to coordinate work between these reviews as appropriate. The FDA recommends that sponsors of new drug–diagnostic products interact early with the FDA to determine best practice information and current scientific and regulatory thinking.

## SUMMARY

One of multiple important parts of the critical path initiative at the FDA has been a growing interest and attention to the issue of biomarkers, particularly with regard to the role they play as tools to aid in development of drugs and other new therapies. The FDA appreciates that the coordination of reviews across these two product lines may be challenging and has both administrative and regulatory tools in place that can contribute to review successes for these products. Scientific models for optimal evaluation are currently a matter of great interest and discussion at the FDA, and scientists in all three centers regulating human products (CDER, CBER, and CDRH) have had multiple interactions with stakeholders and sponsors of new technologies. It is our belief that the FDA is well positioned to help foster the use of this new technology. We believe that the challenge at hand is taming the science itself and that the FDA can in a nimble manner review new products, assure they are safe and effective, and bring them to market in a manner consistent with the precepts of the critical path program.

## NOTES

1. Under 21 CFR 3.2(e), a combination product is defined to include:
   (1) A product comprised of two or more regulated components, i.e., drug/device, biologic/device, drug/biologic, or drug/device/biologic, that are physically, chemically, or otherwise combined or mixed and produced as a single entity;
   (2) Two or more separate products packaged together in a single package or as a unit and comprised of drug and device products, device and biological products, or biological and drug products;
   (3) A drug, device, or biological product packaged separately that according to its investigational plan or proposed labeling is intended for use only with an approved individually specified drug, device, or biological product where both are required to achieve the intended use, indication, or effect and where upon approval of the proposed product the labeling of the approved product would need to be changed, e.g., to reflect a change in intended use, dosage form, strength, route of administration, or significant change in dose; or

(4) Any investigational drug, device, or biological product packaged separately that according to its proposed labeling is for use only with another individually specified investigational drug, device, or biological product where both are required to achieve the intended use, indication, or effect.

2. Contact information and information on combination products and intercenter review processes may be found on the OCP Web site: http://www.fda.gov/oc/combination.

## REFERENCES


1. FDA (2004). Innovation or stagnation: challenge and opportunity on the critical path to new medical products. http://www.fda.gov/oc/initiatives/criticalpath/white-paper.html.
2. Venter JC, Adams MD, Myers EW, et al. (2001). The sequence of the human genome. *Science*, 291(5507):1304–1351.
3. Lander ES, Linton LM, Birren B, et al. (2001). Initial sequencing and analysis of the human genome. *Nature*, 409(6822):860–921.
4. http://www4.od.nih.gov/oba/sacgt/reports/oversight_report.pdf.
5. http://www4.od.nih.gov/oba/sacghs/public_comments.htm.
6. http://www.fda.gov/cdrh/oivd.
7. Garrison LP, Austin F (2007). The economics of personalized medicine: a model of incentives for value creation and capture. *Drug Inf J*, 41.
8. Sargent DJ, Conley BA, Allegra C, Collette L (2005). Clinical trial designs for predictive marker validation in cancer treatment trials. *J Clin Oncol*, 23(9):2020–2027.
9. http://www.fda.gov/oc/initiatives/criticalpath/reports/opp_report.pdf.
10. http://www.fda.gov/cdrh/oivd/regulatory-standards.html.
11. http://www.fda.gov/cdrh/ode/guidance/337.pdf.
12. Jiang W, Freidlin B, Simon R (2007). Biomarker-adaptive threshhold design: a procedure for evaluating treatment with possible biomarker-defined subset effect. *J Natl Cancer Inst*, 99(13):1036–1043.

# IMPORTANCE OF STATISTICS IN THE QUALIFICATION AND APPLICATION OF BIOMARKERS

Mary Zacour, Ph.D.
*BioZac Consulting, Montreal, Quebec, Canada*

## INTRODUCTION

> Learning how to apply statistical analyses … may take a lifetime of trial (and sometimes error), as it has done in the author's case. There is no evidence that biomedical investigators of the present generation are on a steeper learning curve. [1]

Sitting in the Drill Hall at the University of Toronto with row upon row of shifting, sighing, and coughing co-sufferers, cold fingers clutching one of my two well-sharpened HB pencils while I carefully write out each step of my standard deviation calculation. Statistics 101 Christmas exam. One finger rasping down rows and across columns of tables, $p$-values, degrees of freedom, these are the memories I share with countless scientists who have studied statistics in preparation for their research careers. Although I left that course and others with respectable marks and a seemingly clear understanding of the null hypothesis and the normal curve, I found myself some years later, scratching my head as I considered a sheet whose bland columns of numbers belied the painstaking hours spent injecting, isolating, pipetting, incubating, and centrifuging, from which they had been distilled. Looked at one way, it seemed

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

to suggest one thing, but was that the correct approach? Looked at another way, the picture was different. The conclusion to be drawn would influence the direction that the entire project would take.

How to apply the right statistical test correctly to a particular research question, given the many choices, assumptions, and caveats, is not always obvious; indeed "gross misunderstandings of the purpose and functions of statistical analysis are apparent in applications to research grant-giving bodies and ethics committees, in manuscripts submitted to journals and sometimes in published papers" [1]. In fact, there seems to be a growing field of forensic biostatistics, pointing out data analysis errors that have crept into respected publications [2–4]. This is not surprising, considering the advent of genomics, proteomics, and other-omic-style research, in which the ability to produce huge quantities of complex data are outstripping the common biologist's understanding of the mathematical issues concerning its interpretation.

Scientists focused on biomarker research in the drug development industry face an additional set of challenges, where the pressures of the business world to use less time and fewer resources, to focus only on the immediate goal, and to quickly release positive findings to investors or other onlookers are often at odds with the expert, thorough, and careful consideration that the more challenging research problems may require. Understandably, biomedical scientists who are often highly expert in their scientific discipline are not necessarily so in the mathematical issues underlying analysis of their data, and many are also without the luxury of a biostatistician with whom to consult at times of indecision. Although some lucky others may feel that there is no need to master the field when they can simply rely on in-house expertise, it is important to recognize that a working knowledge of the factors affecting data analysis facilitates better communication between the statistician and researcher, which, in turn, generally leads to better research design. Furthermore, the question of how to interpret data wisely is not limited to applying the correct statistical test in the correct manner, but also includes understanding the limits and pitfalls of the technical procedures used to generate the data. In this regard, biostatisticians, too, face challenges, such as being unfamiliar with or intimidated by the rapidly evolving and complex technical methods used in the lab. As so well voiced by Ransohoff [5], "how well do biologists … appreciate the nature or seriousness of problems that can be introduced by methods of handling data … ? How well do epidemiologists or biostatisticians understand technical details of specimen collection, handling, and analysis in a way they can use to anticipate and manage specific sources of bias?"

In addition to an understanding of both technical and mathematical issues, there is a third element important to teasing out the elusive meaning of the data, which I like to think of as the HP factor (Hercule Poirot, that is). Born of intellect and experience, modeled by a higher-order view, a wider vision of extraneous factors that may play roles in the outcome, infused with a measure of skepticism and questioning of assumptions, this factor may more commonly be known as judgment.

Now, I sincerely hope that I am not marked, Rushdie-like, by the statisticians guild for this comment, since this softer science is undoubtedly in the realm of the subjective, and as such, in danger of being classified in the same bin as bias, data selection, and the bubonic plague. To my mind, however, judgment is critical to the success of most endeavors, including data evaluation and statistical analysis. Even though the rules of application of such analyses may be straightforward (in our day, in fact, completed for us by a simple twitch of an index finger on a mouse), understanding potential errors in our data and/or determining which of many analytical possibilities best fit the situation presented by our own experimental paradigm and data set are not, and this is where judgment often comes in.

There is certainly a large mass of statistical reference material already published and available to answer questions that most biologists might have, however, most texts and articles focused on statistical analysis issues are written by, well … statisticians. This is a good thing, of course, but sometimes brings with it the negative corollary of a specialized dialect and viewpoint. Joe or Josephine Biologist may find these, at best, difficult to adapt to their own needs, or, at worst, frankly unintelligible.

This chapter is addressed to fellow Joes and Josephines. It is written in plain language by a nonstatistician who has often grappled at the interface where experimental results are transduced through an intermediate state as digits, before manifesting themselves, newly metamorphosed, into conclusions that flutter their wings right into our collective scientific knowledge database. Whereas some basic statistics explanations may embellish its text, this is by no means intended to serve as a statistics manual, nor to explore the many intricacies of the field; in fact, statisticians may well be appalled by its gaps and oversimplifications. The publisher, on the other hand, is likely to be pleased at the paucity of equations, since it is common knowledge in the industry that book sales drop with each equation in a text. This chapter addresses data analysis issues pertinent to biomarker research in drug development. More specifically, it is intended to clarify some common misconceptions or confusion concerning which situations are appropriate for alternative choices of statistical analysis and to integrate statistical issues with nonstatistical issues that also affect our data interpretation.

## FUNDAMENTALS

### Is That Significance Significant? (And Conversely, the Significant Insignificance)

Statistics texts all deal with the issue of type I and type II errors, or, respectively, the false finding of a significant difference where none exists and the false finding of no significant difference where one indeed exists. They seem invariably to omit mention, however, of how a statistically significant

difference that truly exists and is truly shown can yet remain truly irrelevant. This particular possibility gets no press time in statistics manuals because it remains in the realm of that softer science mentioned above, judgment.

As trained scientists sensitized to the dangers of making inferences based on anecdotal or potentially chance findings, we seem as a group to have acquired the tendency to overexalt the statistically significant finding when it makes its appearance. A $p$-value of 0.05 or less has become our holy grail, the golden chalice that shines from the paragraphs of our grant and investigational new drug (IND) applications, presaging, we hope, the sweet wine to spill forth from them. This, of course, is not unnatural when one considers how often we are faced with budgetary or time constraints on the number of subjects that can be tested and therefore on the ease of statistically confirming such differences when they truly exist. The converse situation of which we need to be mindful, however, is that with enough samples being evaluated, even tiny differences can be statistically significant, and furthermore, even differences that are not tiny might still not be relevant. What constitutes a "significant" increment of an analyte, in the physiological or clinical sense, is likely to vary with the situation; it is not necessarily on the same scale as a statistically significant increment. An additional point to consider, addressed below in the section on measurement error, is whether the magnitude of a statistically significant difference surpasses the magnitude of inaccuracy of measurement.

A qualitative judgment on the significance of significance, rather than simply adjusting our blinders to keep out the sun while logging $p$-values, is certainly worth a moment of silence on our parts. Furthermore, the converse case is also relevant; for example, what if one's inner Hercule is suspicious that a clinical significance may be lurking behind the harmless façade of $p > 0.05$?. How could we reconcile the lack of statistical significance in that case? One possibility is that too few samples were evaluated to achieve the statistical power to reveal significance (discussed further later in the chapter). Another is that part of the group of samples tested reacts differently than (an) other part(s), resulting in a total effect across the disparate subgroups that does not reach statistical significance, even though it reflects a clinical situation of great significance.

The latter concept is central to the development of personalized medicine, an emerging field that seeks to improve disease treatment by targeting molecularly distinct subpopulations within a given group, such that response rates to treatments are increased and the administration of ineffective and potentially toxic treatments to inappropriate individuals is diminished. As an example, selection of breast cancer patients according to *ErbB2* tyrosine kinase status has had a significant impact on management of the disease; moderate, but significant, response rates to *ErbB2*-targeted antibody therapeutics in some studies might have been overlooked had study groups not first been selected for its overexpression [6]. Some basic concepts concerning statistical methods of relevance to pharmacogenomics are discussed later, in addition to some of the common errors in their application.

## Air the Error

Because statistical descriptions and inferences are based on probability, a certain amount of chance (random) error is always inherent in any statistically based conclusion. This error is routinely quantified by upper and lower confidence limits and mentioned in results sections. Unfortunately, such routine error expression seems to induce a false sense of security in readers that *all* the error inherent in the results has been taken into account.

In fact, there are numerous sources of nonrandom error, generally covered by the blanket term *bias* (i.e., when subjects, specimens, or data in the groups being compared are inherently different or are handled differently in a way that systematically introduces a signal into data for one of the groups compared). Differences such as types of tubes used or length of time until spinning or time samples are maintained frozen prior to analysis would fall into this category. Bias has been deemed such a serious problem in nonexperimental research that some experts consider such studies guilty of bias and erroneous results until proven innocent [7]. The term *nonexperimental research* refers to research where the effects of some perturbation to the system are not being tested, but rather, information is simply gathered (i.e., in epidemiological studies) or quantified (i.e., in laboratory analyses, such as most proteomic analysis). Bias is also a serious and often overlooked problem in experimental research [i.e., where effects of deliberate maniuplation of (a) variable(s) are tested]. A recent perspective article on how potential flaws in preclinical research may play a major role in late-phase drug failures discusses this in detail, particularly as it relates to animal models [8].

Measurement imprecision, included by some under the term *bias*, but by others considered separately, constitutes an additional significant source of error. In the case of continuous data quantifications, although measurement error almost always occurs, it is almost never quantified, or at least such quantification is in no way accounted for in the result reported. In the case of noncontinuous data, such as in diagnostic tests that may generate a yes/no dichotomy, measurement error can have a different but equally profound effect, potentially contributing to false positives or negatives.

***Continuous Data Quantifications*** Every assay result is only the best approximation to the true value that a particular assay is able to produce. This is readily acknowledged by all, and, in fact, it is standard practice in drug development to quantify the inaccuracy of assays using known amounts of a reference standard. How the quantification should best be accomplished is beyond the scope of this chapter, but the total error approach wisely incorporates both the components of random error and bias/measurement error in a single measure and is discussed further elsewhere [9–11].

Despite the depth of consideration given to measurement error, there appears to be an element of lip service to its acknowledgment, since error quantifications are generally used only to determine acceptance or rejection

of an assay's results, rather than to modify or qualify in any way the numerical result produced by the assay. Measurement error is thus consistently brushed under the carpet, with results often reported in ways that dramatically overstate their precision. As an end result, differences from one measure to another in some cases become infused with spurious meaning, when they actually represent nothing more than noise in the signals. One can only speculate on how many drugs that failed during late-phase trials might never have gotten that far had the measurement error in their "promising" preclinical results been accounted for in the data analysis.

Sometimes measurement error is compounded dramatically by introducing more than one erroneous measurement into a final calculation of the result, and again the norm is to ignore this inflated error rather than account for it in the results reported. Let us consider the example of a clinical trial examining the effect of a treatment on the activity of a disease biomarker in white blood cells. In this example, two assays are required: one to quantify the disease biomarker in a lysate prepared from the cells, and one to quantify the amount of protein that was culled in each cell preparation. The activity result from the first assay is then normalized against the protein result from the second assay, to give an activity per mass of protein as the final assay result. Looking more closely at the potential error in this example, we should note that standard acceptance criteria for bioassays generally include specification that a certain level of demonstrated measurement error will be considered acceptable, with data still included "as is" in the analysis. Assuming a typical acceptance criterion of up to ±20% inaccuracy, two measurements of a single identical sample of true value $X$ could therefore quantify as differing from each other by 40% of $X$, and no one would raise an eyebrow. In fact, Example 1 shows how once you account for the compound error of the two assays in

---

**Example 1.** One assay for enzyme activity and one for protein content of sample are performed, with the final activity reported as per mass of protein. Assuming ±20% measurement inaccuracy in each of the two assays, we see that if patient A has a true activity of 10 activity units, acceptably accurate activity assays would reveal values of 8 to 12 units, and each milligram of protein would be quantified as 0.8 to 1.2 mg. Therefore, within each assay, an identical sample could have acceptable measurement error, resulting in a 150% increase of one value relative to the other (12 units/8 units, or 1.2 mg/0.8 mg).

Normalizing enzyme activity units to milligrams of protein inflates this possible discrepancy between two identical sample results from 150% to a 220% apparent increase in one sample versus another; that is, the most extreme acceptable values from each assay, normalized against each other, would be 8 units/1.2 mg and 12 units/0.8 mg, for a final resulting range of 6.7 to 15 units/mg as possible results from replicate aliquots of the same sample.

---

this example, two assay results showing a 220% increase in one sample's final result versus another are possible when the two actually are replicates, having an identical amount of the biomarker.

This example demonstrates the compounding of only two measurement errors. Imagine the case when more measurements go into a calculation, each having its own error: for example, when using reporter gene induction in a cell-based system as a model of induction of an in vivo response. In this case there are four measurements that need to be done to control properly, each with its own measurement error range: the reporter gene expression in transfected cells challenged with the experimental treatment, its expression in transfected cells challenged with only the vehicle in which the treatment compound is dissolved, and those same two conditions in cells that were transfected with empty vectors (as a control for the effects of the vector backbone on the system). All too often researchers using models of this sort tend to "simplify" by expressing induction of reporters as simply a fold-increase over the vehicle effect, essentially pretending that the complex and compounded error does not exist.

***Noncontinuous Data*** Measurement error can also have an important effect on noncontinuous data, an effect which, as noted above, is discrete from random error and is rarely taken into account. In this case, misclassifications can occur, such as false negatives or positives. We can all imagine the personal impact of such misclassifications, such as being falsely diagnosed with a deadly disease, or conversely, not receiving appropriate treatment because of a false-negative test result. Misclassification errors also decrease the power of studies to detect significant differences and have the potential to obscure or even falsify study results, leading to equally devastating poststudy impact on a much wider scale.

For example, successful efforts to develop an effective vaccine for malaria, considered to be the deadliest pediatric disease, have been hampered by measurement errors leading to misclassifications. In this case, the fact that vaccine efficacy (VE) is a ratio of the number of infected, vaccinated persons over the number of infected, nonvaccinated persons gives rise to the misconception that false-positive or false-negative diagnosis in the numerator and denominator will cancel out or balance. Actually, since measurement error increases as the lower limit of detection is approached, a partial effect of vaccination causing lower-level signals at the time of measurement (such as a slowing of parasite growth, the biomarker for VE), can lead to systematic error in the measurement of the test group that is not matched in the control group, and an overestimation of VE [12].

The end effects of misclassification error of this sort can vary substantially, depending not only on the goal of the testing and the decisions that are dependent on it, but also on the overall frequency of the disease being studied. For instance, elevated false-positive rates for a test may have a huge impact in the case of rare diseases, and conversely, elevated false-negative rates may have

a bigger impact with very common diseases. Example 2 shows that if a test for rare disease $X$ (occurring in 1 in 1000 people) gives 99% accurate positive results in people who truly have the disease and 95% accurate negative results in people who truly do not have it, any positive test result it gives will be false in 98% of the cases.

---

**Example 2.** A diagnostic test has the following known accuracy: It returns a positive result in 99% of patients who truly have the disease and a negative result in 95% of patients who truly do not have the disease. The disease is rare, occurring in the general population at a rate of only 1/1000 people. Although you have no known predisposition or reason to think you have the disease, you decide to get tested and are shocked to learn that the result has come back positive. What are your chances that the diagnosis is actually a false positive?

$$\text{False positive probability} = 1 - \text{true positive probability}$$

$$\begin{aligned}\text{True positive probability} &= \frac{\text{probability of true positives}}{\text{total of true and false positive probabilities}} \\ &= \frac{0.99 \times 0.001}{(0.99 \times 0.001) + (0.05 \times 0.999)} \\ &= 0.019\end{aligned}$$

$$\text{False positive probability} = 1 - 0.019 = 0.981$$

Your chances are about 98% that the diagnosis of positive is false.

---

The test would have to be 99.9% accurate for negative results before it would have a zero-rated false-positive detection rate. Whereas, intuitively, one might think of 95% and 99.9% accuracy as both sounding pretty good, you can see that the seemingly small gap between them could have a surprisingly important impact.

### What to Do?

Samuel Taylor Coleridge once said that literary reviewers are usually people who would have been poets, historians, and biographers if they could, but having tried their talents at one or the other, and failed, they turn to critics. As in other areas of life, with the issue of measurement error, too, it is much easier to be a critic than to do something well oneself. What should one actually do? This is a very good question indeed, and, feeling a little uncomfortable under Samuel's gaze from the sentence above, I shall attempt to come up with some useful suggestions.

First, the appropriate approach is likely to vary according to the situation, and the most crucial issue to consider when choosing an approach to dealing

with measurement error is the potential risk of the erroneous measurement to the end user of the drug, device, or process under study. Experiments and/or data interpretation should then be designed to err on the side of caution, if risk is an issue.

Other key determinants may include practicalities such as resources and time available to approach the problem, and questions of clinical or physiological relevance of the potential error margin, thresholds, or measurement ranges of interest. If, for example, a 10- to 100-fold increase in biomarker expression would be required for clinical relevance, and the maximum estimated measurement error is in the range of less than threefold, perhaps there is no issue. The discussion above, concerning whether significant differences are clinically relevant also applies to the relevance of measurement errors.

If measurement error is of a magnitude that could be clinically relevant, its quantification becomes more important. Perhaps because of the status quo of ignoring measurement error quantification in biomedical methods (for which, after all, error quantification is much more challenging than for analysis of less complex models), this is not a widespread practice in the drug development industry. One way of acknowledging this error while avoiding complex quantifications might be to use the definition of maximum allowable measurement error that is already made routinely for the purposes of acceptance criteria as a threshold of level of difference, below which results would either not be considered to reflect anything but error, or would be more carefully scrutinized in some predefined way. In other words, if you have determined that up to a doubling in endpoint values in a pretreatment versus posttreatment sample could be attributed solely to compounded measurement error rather than to treatment per se, one might decide not to consider effects to be a result of treatment unless they surpass this measurement error threshold. Of course, the fact that a method has been validated according to across-the-board standards such as having better than 15% or 20% total error does not mean that it actually produces results that are this far from the nominal values of the reference standards, but merely that it may do so at times, and that results of assays where these limits are exceeded will be rejected from the study. Error quantifications may well indicate much less error than the maximum allowed for assay acceptance, so given the circumstances it may be worth putting the time in to determine what the actual error is rather than simply assuming that the worst possible error has always occurred. Actual error might be revealed by data mining from previously amassed quality control data and/or on a per-batch basis using internal standards, for example. In these cases it is as simple as quantifying the proportional values obtained for accuracy rather than simply noting a pass or fail.

In the case where biomarker analysis involves more complex sources of error than a single or a few analytical assays (such as for epidemiological studies or metaanalyses to qualify particular biomarkers for their intended use), more complex uncertainty quantifications would be required. Various such methods are available, including Bayesian [13] and Monte Carlo simula-

tions [14]. Bayesian approaches, which stem from an alternative view of probability compared to the more common frequentist inferential statistics, are touched upon further later. Monte Carlo simulations represent the brute-force type of computation, basically executing a huge number of possible permutations using random draws from all the potential inputs, so as to spit out a final probability function (made up of all the output values from the various random inputs) that comprises the compound error from all the unknown inputs. In a simple example, consider the film *Groundhog Day*, where a man woke every morning to find that it was the same day as the one before. The film consisted of this man going through his interminable repeated days, making different choices when the fixed events of the day occurred, as they did every day; you can think of this man as being stuck inside a Monte Carlo simulation (well, a simple one, since his own behavior choices were the only independent variable with randomly chosen modifications; the range of outcomes were consequences of his varied choices from that one variable). In questions of compound error, of course, there would be many more variables, also all being sampled randomly over and over again. One advantage of choosing a Monte Carlo approach to resolving complex error sources should your experimental paradigm present them, is that user-friendly software is readily available, due to its common use for risk analysis in financial and engineering applications.

Once measurement error has been quantified, by whatever means, error distributions could be used to express uncertainty limits around results, analogous to the random error-related confidence intervals common in statistical analyses. These would help to clarify how much of the apparent differences between sets of results might be an artifact of the measuring tool(s) and not necessarily an effect of the treatment that one wishes to measure. In many cases, such as the reporter gene model example above, literature search will reveal readily applicable statistical models for specific types of assay that incorporate compound sources of error into the calculation of statistical significance of differences [15].

Phillips and Lapole suggest a much simpler approach to partially quantify method error uncertainty without any complicated analysis [16]. Like Pablo Picasso, who once said it took him his entire life to learn how to paint like a child, these authors advocate a return to earlier ways: in this case, the grade-school lesson of rounding figures. Rounding can adjust the measurement scale to reflect the limitations of the measurement error. In their words: "Rounding does not create imprecision—the imprecision exists even if we do not accurately report it." Thus even though the measurement instrument may spit out values with several decimal places in them, samples with results in the range of 100 for a test demonstrating 2% inaccuracy (i.e., 98 to 102) might better be reported rounding to the nearest 5 rather than even to 1's, let alone the decimal points reported by the instrumentation.

Apart from attempting to quantify the error, contain it within a wider measurement scale, or set thresholds of interest that exceed it, as discussed

above, we should also not forget to implement strategies to reduce it, when it is possible to do so. For instance, internal controls such as reference standards or paired samples can reduce the impact of the between-assay variance proportion of measurement error.

For noncontinuous data, misclassification rather than numerical differences are the end effect of measurement error. In this case, a suggested approach to avoid the negative impact of measurement error might be to assume that error will happen, and design studies to have higher power from the beginning, in order to minimize the impact of error on the power of the study. Statistical tests for noncontinuous data are already inherently less powerful than those for continuous data, and a lower initial power level in the experimental design can further magnify the impact of measurement error. For example, Gordon and Finch illustrate how the same misclassification error rate causes much more power loss in a study that has lower power to start with than in a study that has higher power to start with. In their example, given exactly the same error rate, if the study is designed to have 99% power, it ends up with 95.5% after the error is taken into account, whereas if it is designed to have 80% power, it ends up with only 65.4% [17]. Therefore, specifying a higher power in the experimental design gives an advantage of robustness to the obscuring effects of misclassification error. In real terms, designing an experiment with higher power most often translates into recruiting more subjects; these authors advocate that "paying the cost up front" strategy.

## Guilty by Association (and Other Misuses of Correlation Analysis)

"Statistics show that of those who contract the habit of eating, very few ever survive." This William Wallace Irwin quote [18] may bring a small smile to your lips. Should we conclude that eating is a bad habit that causes death? Ridiculous, right? Especially for well-educated biomedical scientists. How about this one: "Greater adiposity is associated with lower everyday physical functioning, such as climbing stairs or other moderate activities, as well as lower feelings of well-being and greater burden of pain" [19]. This quote from a respected medical journal does not seem quite so ridiculous, does it? Whereas the first one is clearly a joke, we might be tempted to nod our heads at the second and experience discomfort in direct proportion to the increases we noted over the past few years in our own level of adiposity.

If we take a closer look at both quotes we will notice that neither author actually claimed a cause-and-effect relationship. As we all know, however (especially those employed in the marketing department), just putting them together in a small room can't help but to make everyone jump to conclusions. The strength of an association is indeed appropriately quantified by correlation analysis, but this analysis should not be misinterpreted to infer causation. Zou et al. hit the nail on the head when they note that rather than being misused to test hypotheses, correlation analysis should serve to *generate* them [20].

The fact remains, however, that correlation is very commonly used to suggest that A causes B, or as in the second example above, being fat causes lower physical ability to function and lower feelings of well-being. Such guilt by association certainly is one possibility, but another possibility is that B causes A, or having lower ability to function and poor self-esteem induces weight gain. A third equally plausible hypothesis is that another cause or set of causes is (are) responsible for both the poor feelings/abilities and the weight gain. A fourth is that the observer has just observed and documented the effects of coincidence; if an insufficient sample number were examined, for instance, a fluke might be documented. Even with larger sample numbers, statistical conclusions are always subject to chance, with the type II error (or *p*-value) describing the percent chances that the "significant difference" is actually just due to chance.

Some extra consideration needs to be given to this when choosing an appropriate biomarker for one's particular purposes. Biomarkers are, after all, by definition, a marker for something else, or an associative entity. If a particular biomarker has been qualified properly, its association with the disease of interest will not be due to chance, but that still does not imply causation. Whether a marker is simply associated, is a cause or an effect of disease, or is a cause or effect of something else causing a disease needs to be considered carefully in the experimental design. The issue of what considerations are important to determine causation is discussed in a landmark paper by Hill [21]. The reader is further referred to an interesting perspective commentary on this subject by Phillips and Goodman [22].

A further frequent misuse of correlation is to test agreement between methods. As pointed out by Bland and Altman [23], even poorly agreeing methods can produce quite high correlations, and for methods to be in perfect agreement their results would plot against each other along the line of equality, whereas results plotting on any straight line would give perfect correlation. Furthermore, the test of significance may show that the two methods are related, but this is irrelevant to the question of agreement, and indeed, it would be amazing if two methods designed to measure the same quantity were not related. One can almost hear their teeth gnashing as they bemoan, "Why has a totally inappropriate method, the correlation coefficient, become almost universally used for this purpose?" It seems clear that this is a case of copycat crime … perhaps a case for the remarkable Hercule Poirot?

## Checking the Dessert Menu First

In much the same way that Hercule Poirot might consider the temptations of the dessert menu before settling on his choice of main courses, we all need to consider potential data analysis issues before settling on our experimental design. Some questions need to be asked long before the data are in hand. For example:

1. Is the assay to measure the biomarker as good as it can be, or the best type of assay given the circumstances? Can you make it better? All assays have error, but the more you fine-tune the accuracy, precision, specificity/sensitivity, and robustness via appropriate optimization, the more your result will reflect for you the marker you wish to measure rather than the shortcomings of your technique. Assay optimization is beyond the scope of this chapter, but should definitely be completed before starting the study. Assay-specific parameters such as accuracy, precision, and specificity of the measurements all have a direct impact on the uncertainty of the statistical analysis outcomes.
2. Have you chosen the best biomarker, or should you rethink from the bottom up? To evaluate this, you must be well aware not only of the qualification that a biomarker is relevant, linked appropriately, and specific to the pathology of interest (important issues, beyond the scope of this chapter), but also of quantitative issues of what to expect in the sample and control groups you have chosen.

   If one wishes to use a biomarker to differentiate between normal healthy persons and those with a disease, for example, one of the many questions to be considered prior to deciding which indeed is the most informative biomarker for the job concerns the potential degrees of measurement between them. In many cases the biomarker of interest is present in normal healthy persons as well as in those with disease, but with different typical levels in the two populations. In this case it is very important to establish a priori the descriptive characteristics of the two groups in order to define what a clinically relevant treatment effect might be and how one might judge that such an effect occurred. Published literature, unpublished previous data, and pilot studies all provide potential resource material for defining these descriptive characteristics and should be evaluated carefully in advance. Since both populations will present with a range of values, relevant descriptive characteristics include not only mean values, but also the variances, ranges, and the difference between the population distributions (or extent of overlap, if that is the case). One should consider that the biomarker that is more consistent in its levels (i.e., less variance) may be a better choice than one that shows bigger mean differences between populations (but higher variability from person to person within a group). In the latter case, the variability can translate into overlap of two populations, which can be quite a complication when one wishes to discriminate between them.

   Another point for reflection in advance concerns integrating the quantitative ability of the measurement technique you intend to use with the quantitative characteristics of the populations you hope to discriminate between, or the effect size that you have determined you will need to achieve clinical relevance. For example, the smaller the differences are between expression levels of a biomarker that you wish to use to

discriminate between two populations, the more sensitive the quantitation method you choose needs to be. Consider the example of wishing to measure the treatment effect of drug *X* on a disease by assessing whether a biomarker is increased from disease-typical levels to the level seen in healthy persons. Would one be able to accomplish this with statistical confidence if the two populations showed close levels or overlap between their two distributions? If the difference between the populations is not big but the measurement error margin is, one may be doomed to an incapacity of discerning what one hopes to, regardless of the potential effects of the treatment. In this case it is wiser to rethink the fundamental premises of your study design and methodology before starting. Although this may sound like an obvious point, when one considers the common practice of ignoring measurement error and ignoring the compound error introduced by common study practices such as normalizing one assay result to another, it should perhaps give pause for thought.

3. Is your design sufficiently powered to detect what you wish to detect? Statistical power is the ability of a study to enable detection of a statistically significant difference when there truly is one; power in the statistical sense is analogous to the sensitivity of a diagnostic test [24]. Eng says: "One could argue that it is as wasteful and inappropriate to conduct a study with inadequate power as it is to obtain a diagnostic test of insufficient sensitivity to rule out a disease" [25]. In clinical trials, power often receives attention, perhaps thanks to ethics committees that have pointed out that it is frankly unethical to expose people to the risks and discomforts of research if the study has no potential for scientific gain, such as if a study was not designed to include enough people to adequately test the research hypothesis [26]. In laboratory experiments, on the other hand, statistical power is rarely estimated [1] and seems in general overlooked and poorly understood.

   To ensure sufficient power, one needs to think ahead and choose an appropriate sample size. Factors playing a role in how one would define *appropriate* include:

   a. The smallest meaningful difference between the two means being compared that the investigator would like the study to detect (this is usually based on subjective criteria such as judgment of clinical importance; additionally, sensitivity and error margins of the measurement method may influence its choice).
   b. The estimated standard deviation within each comparison group (this can be based on a pilot study, prior literature that has been carried out with the same outcome measure, or subjective criteria).
   c. The desired power level (this is usually 80% or higher; however, although high power is desirable, it is always in a trade-off with a fixed amount of time and resources for the study).

d. The significance criterion (usually set at 0.05).

e. Whether one- or two-tailed analysis is planned.

There are different formulas for computing power or sample size, depending on the experimental design and whether data are continuous or categorical; Example 3 shows an illustrative calculation using a formula suited to a standard design involving comparing means of two groups.

---

**Example 3.** You have qualified your biomarker $X$ as an efficacy marker in your disease model. In your studies qualifying $X$ as a biomarker and developing a method to assay it, you found that its expression was low in rats with your disease of interest (10 ± 3 units; mean ± SD) but higher in healthy rats (100 ± 30 units). You also noted that interventions that increased $X$ were associated with improvements in disease morbidity and that previous groups of animals studied tended to show biological variability of about 30% coefficient of variation between rats in a treatment group. You wish to design an appropriately powered study comparing untreated diseased rats to those treated with a single dose of your new treatment $Y$, to determine if $Y$ demonstrates efficacy, and have decided that a mean 5-unit change in $X$ is the lowest meaningful change that you wish to be able to detect. You will settle for standard levels of a significance level of 0.05 and a power of 80%. How many rats do you need in each group?

$$n \text{ of each group} = \frac{\left(\mathrm{SD1}^2 + \mathrm{SD2}^2\right)\left(z_{\mathrm{crit}} + z_{\mathrm{pwr}}\right)^2}{D^2}$$

where SD1 is the standard deviation of group 1 and SD2 that of group 2 to be compared to each other (for a minimum difference from 10 to 15 units, this would be 3 and 4.5, respectively); the $z$-values are standard normal deviates corresponding to selected significance criteria and statistical powers ($z_{\mathrm{crit}}$ for two-tailed $p < 0.05$ is 1.960, $z_{\mathrm{pwr}}$ for 0.80 power is 0.842; these and other values are found in standard statistical tables and software packages); and $D$ is the minimum difference that one wishes to detect (5 units).

$$\begin{aligned} n &= \frac{\left(3^2 + 4.5^2\right) \times (1.960 + 0.842)^2}{5^2} \\ &= \frac{29.25 \times 7.851}{25} \\ &= 9.2 \end{aligned}$$

Round up to 10 rats in each group.

---

Numerous Web page calculators and statistical programs are available to perform this type of calculation for you; however, it is difficult to use them properly without a working knowledge of the factors affecting power. This subject is well explained by Eng [25,26].

If your study is already complete and you found a significant difference, you do not have to be concerned as to whether the study was sufficiently powered. This is akin to wondering if you put enough dynamite under the bridge after you have blown it to smithereens. If, however, you found no significant difference and you are feeling uncomfortable about the fact that you did not consider calculating the power before starting the study, you may be tempted to do so after the fact, by plugging the values actually observed in your study for standard deviation and the difference between the populations into the power formula. This practice, dubbed *retrospective power analysis*, is problematic, because it does not tell you what the power to find the smallest meaningful difference was, only what the power was to find statistically significant the difference that actually existed between the groups in your analysis; furthermore, this "observed" power will be inversely related to the *p*-value observed, so it does not add meaningfully to the information already obtained with the *p* value [26]. Instead, approaches involving confidence intervals or chi-square tests are recommended to guide the interpretation of negative results [27]. Some practical considerations about power and sample size are discussed further on page 264.

4. Have you planned in advance which groups you are going to compare to which, for significant differences? Planned comparisons allow the researcher a greater rejection level $\alpha$ or *p*-value than do unplanned comparisons. This is because if you first look at your data, then choose your comparisons, you are introducing bias into your analysis, and the rejection level will have to be tightened to compensate for that [28]. Planning ahead also allows you clear-headed time to consider which comparisons you really need when your design includes many different groups. The more comparisons you make, the more the *p*-value will have to be adjusted against false positives, in effect making it harder for you to see a true positive. Therefore, it is important not to compare groups that are irrelevant to your experimental goals. This is discussed further on page 272 and in Example 6.

## Taking Care of Business

As touched upon previously, the challenges faced by industry-based scientists often include business-related considerations that are not considered in most statistics manuals. In particular, time and resource constraints are often significant, as well as pressures to proceed at a pace that is intended to satisfy management, investor, or profit-related concerns, but is often incompatible with sufficient depth of study or optimal experimental design. In short, between

the business and science aspects of the industry, a strong pressure front constantly flows from the business side to the science side.

At times when shortsightedness and lack of understanding of the impact of financial decisions on scientific practice threaten to nullify the potential uses of our scientific output, we should certainly consider exerting a pressure front in the opposite direction, from science to business. Since we cannot realistically hope to change the pressures of the business world, however, we need to at least consider modifying our experimental design and/or data interpretation approaches to take their effects into account.

For example, let us consider the case of a study design for analysis of a biomarker as a pharmacodynamic (PD) endpoint in a clinical trial,where each patient provides a pretreatment control sample to be compared with a time series of their posttreatment samples. Given an endpoint assay with a high between-experiment variance component, the wisest experimental design might be to study all of a given patient's posttreatment samples within the same assay, together with the matched pretreatment sample.

Instead, scientists are often pressured to run assays and provide PD results as soon as the first or first few of several posttreatment time point samples become available in clinical trials. This business-driven decision is often made despite freezer stability of samples, simply because drug development companies are anxious to get the news about their product's effects in clinical trial announced as quickly as possible; good news has an economic payoff, and similarly, the earlier that bad news is known, the better the chances for diverting precious resources before they are wasted. In the laboratory, however, the corollary of this pressure may be that precious volumes or tissue samples from a pretreatment control are used up in the first analysis, and later analyses on samples not yet collected will have to be calculated in assays that lack this internal per-patient control. As well as introducing bias between the time points that were assayed within a single run and those that incorporate between-assay variability, the added component of variability between assays may add unnecessary noise, obscuring potential treatment effects.

Certainly, as drug development scientists we cannot ignore concerns such as continued supply of cash by investors; indeed, such issues are supremely important to whether a scientific endeavor can even be continued. The point is that the effects of making less than ideal experimental design choices, too, should not be ignored or glossed over (which is, in fact, what tends to happen). Often, no consideration is given in the statistical analysis model to sample processing issues such as that described above, with raw values instead being logged into tables as if they had all been generated in a single assay of balanced design, regardless of the circumstances under which the data were generated.

Obviously, there is no one correct approach to remedy such a situation, and the researcher must decide what best fits the situation, hopefully with the help of a statistician who understands the sample processing issues. For instance, in some cases the best solution might be to convince decision makers to wait for a controlled comparison of pre- and posttreatment samples, in others it might

be including an internal standard for normalization and changing to a ratio-data design, yet others might be amenable to another analysis strategy. The point is that analysis and/or interpretation of the data should be rethought in the light of particular business decisions that have affected their generation.

In another example of rethinking assay design strategies to respond to the challenges of the business world, one might consider various strategies to increase statistical power in the face of time and resource constraints that typically limit the number of patients one can study. Browner et al. [29] and Eng [25] list a number of strategies for minimizing the sample size while maintaining power.

1. Use continuous measurements instead of categories; this is because, for a given sample size, statistical tests for continuous values are mathematically more powerful than those used for proportions.
2. Seek to decrease the variability of the measurement process (i.e., improve the precision of your processes, optimize your assays; the less standard deviation, the more power).
3. If possible, use paired measurements, matching each patient to his or her own control. Paired measurements are more powerful than unpaired measurements.
4. Add control subjects, even if case subjects are too difficult to obtain. Although it takes more complex equations to calculate power in unbalanced designs than if both groups have the same size, this will still add power.
5. Expand the minimum expected difference-perhaps that which you have specified is unnecessarily small and a larger expected difference could be justified, especially if the study is preliminary, or intended as a screening test. The smaller the expected difference you want to pick up, the greater the number of samples needed to give an equivalent power.
6. Reflect on whether your hypothesis could have merit as a one-tailed, rather than a two-tailed one (although in most cases this is not an option). You can obtain equivalent power using less samples for a one-tailed test. The sample size necessary for a one-tailed design with a 0.05 significance criterion is the same as that for a two-tailed design with 0.10 significance, if all other factors are equal.

Many researchers think of power calculations as a futile exercise, likely to specify an unrealistically high sample size, given the resources on hand for the study. In fact, proper thought to power and the factors that contribute to it could, in many instances, lead to cost and resource savings. Rather than applying an across-the-board animal toxicity experimental design including, for example, 10 animals per group, Example 4 shows a case where five animals per group would be sufficient; as this example is built on a data set similar to the one used in Example 3, which showed 10 animals per group as being

appropriate for the experimental paradigm, it also illustrates that the number of samples that need to be in a test group for appropriate power depends strongly on issues specific to each particular research question.

Thus, industry-based scientists grappling with business cases, profit-and-loss issues, how best to spread their research budget over the necessary items, and other financial-based decisions with an impact on scientific quality would do well to manage their resources appropriately by incorporating power calculations into the design of their laboratory research.

---

**Example 4.** Imagine an example similar to that described in Example 3, except that the biomarker is now a toxicity marker, which you have qualified and shown low under normal conditions in rats (10 ± 3 units; mean ± SD) but markedly increased in response to toxic reactions (10- to 100-fold). You wish to test a compound for toxicity using the same experimental design as in Example 3, but with a few different assumptions:

1. Because of being cautious about the risks of toxicity, you want to ensure higher than 80% power this time: You want to ensure 95% power, so you have more certainty that any finding of no significant difference in the toxicity marker would not be attributable to too few animals being studied.
2. Because you are not concerned if the toxicity marker decreases in association with the treatment, but only if it increases, you adopt a one-tailed design.
3. Because toxicity so markedly increases the biomarker, you decide that the smallest meaningful increase you wish to detect can be greater than in Example 3. You decide that 20 units of biomarker would be greater than 3SD over the normal condition and is thus an acceptable threshold, so you set $D$ at 20 − 10 = 10.

How many rats do you need in each group?

Using the same formula as in Example 3, but substituting a $z_{\text{pwr}}$ value of 1.645 (the value for 95% power), a $z_{\text{crit}}$ value of 1.645 (the one-tailed value for a $p < 0.05$ significance level), a $D$ value of 10, and SD1 and SD2 values of 3 and 6, respectively:

$$\begin{aligned} n &= \frac{(3^2+6^2)(1.645+1.645)^2}{10^2} \\ &= \frac{45\times 10.824}{100} \\ &= 4.87 \end{aligned}$$

Round up to 5 rats in each group.

---

## The Odd Risk

Biomarkers are often used in studies that involve nominal data, such as a positive/negative dichotomy for a test result. There are various ways of quantifying the strength of the relationship between groups when it comes to this type of data, which is sometimes also referred to as *proportional data*, in reference to the fact that one measurement is often expressed as relative (proportional) to another (i.e., a percentage of patients responding, or the accuracy or sensitivity of a diagnostic test). Proportional data is often presented in a $2 \times 2$ contingency table. The chi-squared statistic is appropriate to test the null hypothesis of independence between row and column variables. Data are often described, however, in terms of difference of proportions, relative risk, or odds ratio. These are not statistical terms, but rather, quantification terms. Often, these are misunderstood, sometime leading to serious confusion and misinterpretation of published research. When considering the use of diagnostic tests classifying disease states, based on biomarkers, the terminology of odds and risks is actually less useful than true and false positive and negative rates, since it does not capture the aspect of error in disease classification. It is nonetheless important to be able to interpret research correctly where results are described in odds/risks terminology. This subject is well explained in a review by Sistrom and Garvan [30], from which most of the following is paraphrased.

Assuming a data table comparing some outcome measure in two groups, one of which was exposed to some variable and the other of which was not, the difference in proportions is the proportion with the outcome in the exposed group minus the proportion with the outcome in the unexposed group. It is always between –1.0 and 1.0, and equals zero when the response $Y$ is completely independent of the explanatory variable $X$.

*Relative risk* is a measure of association between exposure to a factor and risk of a certain outcome, and is defined as risk in the exposed groups divided by risk in the unexposed groups. Public health impact depends not just on relative risk but also absolute risk, so just looking at relative risk can be misleading. For example, the public health impact of a vaccine that halves the risk of an infection is very different for a rare (say, two cases per million people) or a common disease (say two cases per 10 people); in both cases the relative risk for vaccinated subjects would be 0.5 compared to unvaccinated subjects, but in the former case, for every million people vaccinated, only one would be saved, whereas in the latter case it would be 100,000 people saved [31].

*Odds* are the probability that the outcome does occur divided by the probability that it does not occur, with the *odds ratio* being the ratio of the odds of an outcome in one group ("exposed") versus another ("unexposed"). A key concept about using odds to estimate risk of an event is that the relationship between the odds ratio and the risk ratio depends on the outcome frequency of the event. If the odds of an event are greater in one group than another, it does *not* mean that the risk ratio is increased by the same amount.

This is a very common misunderstanding, but in fact, odds ratios and risk ratios are similar only when the outcome being studied is rare (i.e., about 10% or less probability in the unexposed group) [32]. Odds ratios are not a good estimate of risk ratio when the outcome is common in the population being studied. In that case, to obtain correct risks one needs to apply the equation relating odds to risks, as shown in Example 5.

---

**Example 5.** The equation relating risk ratio (RR) to odds ratio (OR) is as follows:

$$\mathrm{RR} = \frac{\mathrm{OR}}{(1-\mathrm{Pr}_0)+(\mathrm{Pr}_0)(\mathrm{OR})}$$

where $\mathrm{Pr}_0$ is the probability of the outcome in the unexposed group. Schulman et al. [33] compared the frequency of referral for cardiac catheterization for white men (exposed) to that for black men (unexposed). Black men were referred 90% of the time, and the odds ratio for black to white men being referred was 0.6:1. The risk ratio for black men to white being referred was

$$\begin{aligned}\mathrm{RR} &= \frac{0.6}{(1-0.9)+(0.9\times 0.6)}\\ &= \frac{0.6}{0.64}\\ &= 0.94\end{aligned}$$

---

This example shows how an odds ratio of 0.6 for referral to cardiac catheterization for black men vs. white men (as reported by Schulman et al. [33]) does not mean that blacks were referred 40% less often than whites, as was reported in extensive media coverage following publication of that study. Once the correction for a 90% referral frequency for black men was applied, the risk ratio for black men to be referred was 0.94, or 6% less relative to white men rather than the 40% less that the media coverage of the original study had interpreted [34].

## PRACTI(STI)CAL MAGIC: WHICH WITCH IS WHICH?

During my time as a contract research organization scientist, I was astonished at the frequency of requests from pharmaceutical and biotechnology clients to perform inappropriate statistical treatment of the results; it seems that in the drug development industry at least, there is an appreciable level of confusion over when particular tests are appropriate and what underlying assumptions must be met for them to be valid. Accordingly, some refresher basic statistics background is summarized briefly below.

## Parametric vs. Nonparametric

Parametric tests are so named because they use parameters, such as mean and standard deviation, in their computations. Student's *t*-test and the analysis of variance (ANOVA) are examples of this type of test. Parametric tests assume that the data sets are normally distributed and that the various groups have roughly equivalent variance between members of the group. They are prone to falsely indicating a significant difference between the groups if these assumptions are not met. In that case, nonparametric tests provide an alternative. Nonparametric tests rely on ranking of samples relative to each other, and make fewer numerical assumptions. They are less likely to commit type I error, but usually also have less power to reveal true significant differences.

Researchers need to evaluate data with descriptive statistics in order to decide whether the assumptions of the more powerful parametric tests are met or whether nonparametric tests would be the more appropriate choice. Some assumptions apply to both types of test and must also be considered. In that regard, both parametric and nonparametric tests assume that samples included in the analysis are independent of each other. How does one judge whether these assumptions are met, and what should one do about it if they are not?

***Independence*** To judge whether samples are independent, one needs to ask if within any one group of data included in the analysis, the value of one data point is somehow related to another. If the answer is yes, these samples are not independent. For example, if biomarker expression is measured in two separate tissue samples three separate times each and you enter $n = 6$ data items into the analysis, you have violated the independence criterion. The three replicates of each sample are related to one another (from the same tissue sample); there are actually only two independent measures, not six. Since $n = 2$ is not enough samples to analyze statistically, potential plans to redress this analysis error could include either omitting statistical analysis in favor of descriptive results only, or redesigning the experiment to include additional tissue samples.

Often, experiments are designed to measure biomarker expression repeatedly in a single subject, such as in the case of pharmacodynamic measures at different time points following administration of a drug in clinical trial patients. In this case, the independence assumption is not violated, since the related measures will be analyzed as different groups rather than being included within one group. Repeated-measure statistical tests are then applicable, such as the paired *t*-test (for two measures) or repeated-measure ANOVA (for more than two measures).

To reiterate, then, if samples are not independent of one another and they do not constitute repeated measures, the correct statistical approach is to redesign your experiment. Although more complex analysis strategies do exist

for clustering correlated samples within groups, there is no manipulation when using simple tests of significance that will correct for samples in one data set being selected such that the choice of one sample depends on another sample being picked.

***Normally Distributed Data*** How do you judge if your data are normally distributed? The first step is to perform descriptive statistics and plot the data out as a histogram. In a normal distribution, about 68% of the data falls within 1 standard deviation (1 SD) of the mean, and 95% within 2 SD of the mean. Many statistics programs will compute normality for you using statistical tests. The Kolmogorov–Smirnoff test, for instance, has nothing to do with martinis, being instead one such test for the normal distribution of data. If you have around $n = 25$ or more in each group, you can safely apply the simple eyeball test to your data, accepting data as normal if points fall into a roughly bell-shaped curve when plotted as a histogram [35].

What if you are not sure whether your data are normally distributed? If you have small data sets, for instance, normality can be difficult to determine visually. Similarly, statistical tests do not have much power to make the determination in this case. At this point, one's own personal choice comes into play: If you are conservative, you might choose to play safe with a nonparametric test; if not, you might choose a parametric test. As with other life choices, there are risks associated with either: With parametric tests you risk error of type I, whereas with nonparametric ones you risk type II error. The best solution might be to collect reasonably large data sets so as to avoid this uncertainty.

Assuming you do have a large enough data set to be certain, and what you are certain of is that the data are not distributed normally, well, what then? One solution might be to transform data mathematically such that the transformed data sets become normally distributed, in which case the more powerful parametric tests can then be used. Some commonly used transformations and appropriate types of data for their use [35–38] are listed in Table 1. If data are still not normally distributed with equal variances following mathematical transformation, a nonparametric test should be chosen for analysis. Some appropriate nonparametric equivalents of parametric tests are shown in Table 2.

***Equivalent Variances Between Groups*** As with the question of whether data are normally distributed, equivalence of variance between groups needs to be evaluated prior to choosing appropriate tests of significance. Sometimes, homogeneity of variance from one group of data to another can be amply judged by plotting a frequency histogram of each group and seeing how spread out the data are, or calculating the variance for each group and qualitatively assessing if the numbers are very different or appear close. It is simple enough, however, to apply the appropriate statistical test, the $F$-test for two groups, or the $F_{max}$-test for more than two.

**TABLE 1 Common Data Transformations**

| Type of Data | Example | Transformation |
|---|---|---|
| Proportional data | 0–100% | Arcsin of the square-root transformation |
| Poisson distributed (counts of random events) | Number of cells of a particular type in a given volume of blood | Square-root transformation |
| Variance proportional to mean squared (i.e., standard deviation proportional to mean) | Serum cholesterol levels in patients | Log transformation |
| Ratio data (normalized to internal standard on a per-assay basis) | Blots normalized to a housekeeping gene | Log transformation |
| Highly variable quantities, where variance is proportional to the mean to the fourth power (i.e., standard deviation proportional to mean squared) | Serum creatinine levels in patients | Reciprocal transformation |

**TABLE 2 Nonparametric Equivalents of Parametric Tests**

| Parametric | Nonparametric |
|---|---|
| Student's *t*-test (unpaired) | Mann–Whitney test (Wilcoxon rank sum) |
| *t*-test (paired) | Wilcoxon signed-rank test |
| *t*-test (paired), in cases where data are not symmetrical around the median | Sign test |
| ANOVA (without repeated measures) | Kruskal–Wallis test |
| ANOVA (with repeated measures) | Friedman's test |

*Source:* Adapted from ref. 35.

If variances differ between groups and are not rendered equivalent by mathematical transformation (see Table 1), you may still be able to use parametric statistics. If, for example, the *t*-test is appropriate to your experimental design, and variances are unequal between groups but not extremely disparate, you may choose to use the *t*-test for unequal variances (in the case of extreme differences in variance, however, a nonparametric test is the recommended alternative).

If your intended parametric test is ANOVA, on the other hand, the assumption of equivalent variance between test groups is essential [35]. If variances are not equivalent and transformation also does not render them equivalent, you must test statistically using a nonparametric equivalent (see Table 2).

## Simple Regression and Correlation Analysis

At times there seems to be some confusion about simple linear regression vs. correlation analysis. This issue is well summed up in a review by Zou et al. [20], from which the following information is, for the most part, culled. The two are similar mathematically, but their purposes are different. Both relate to a function that describes the relationship between a given $X$ and $Y$; however, regression analysis generally focuses on the form of that relationship and correlation generally on the strength. Furthermore, regression focuses on evaluating the relative impact of a predictor variable on a particular (dependent) outcome, whereas correlation's purpose is to examine the strength and direction of the relationship between two random variables.

Correlation analysis commonly involves either the Pearson coefficient ($r$) or the Spearman coefficient ($r^2$), with the former reflecting proportional changes in one variable when the other is changed, and the latter using ranks and reflecting instead a monotonic relationship between two variables (i.e., whether one tends to take either a larger or a smaller value than the other, but not necessarily with a proportional change in one variable when the other one is changed). If data sets are skewed or contain outliers, the Spearman coefficient rather than the Pearson is the appropriate choice.

Interpretation of correlation coefficients is often rather qualitative, with the sign indicating the direction of the relationship (positive or negative). Values range from 0.0 (no correlation) to 1.0 (perfect correlation), with 0.5 generally being thought of as moderate, 0.8 as strong, and 0.2 as weak. Statistical significance can also be computed, by formulating a null hypothesis of no correlation and a one-sided alternative hypothesis that the underlying value exceeds or is less than that value, then computing the $z$-test statistic and rejecting the null hypothesis based on the $p$-value.

Simple linear regression analysis results in an $r^2$ value that is calculated on the basis of Pearson $r$ coefficient and reflects the fraction of the variability in $y$ that can be explained by the variability in $x$ through their linear relationship (or vice versa). As with correlation analysis, a finding of strong linear relationship in a regression analysis does not mean that the variable causes the outcome (as discussed in the section on misuses of correlation), and should not be interpreted that way. Student's $t$-test can be used, for example, to test if there is a linear relationship (i.e., null hypothesis of slope = 0) or whether the $y$-intercept is a particular value. As a salient point, no extrapolation outside the range of values of the independent variable in the regression analysis should be used to make any predictions.

## *t*-Test vs. ANOVA

Using inappropriate tests of significance for the circumstances, such as testing differences between more than two groups by means of repeated $t$-tests rather than by ANOVA, is exceedingly common [35]. Don't do it! Repeating

pairwise comparisons ignores the experimental effects in all but the two groups being compared and increases the number of false-positive "significant results" expected; multiple means comparisons following ANOVA correct for the increasing chances of false positives by adjusting the acceptable hypothesis rejection level.

If you have only two groups to compare to each other and they meet the assumptions of parametric tests, use Student's *t*-test. If you are testing more than two groups, however, use ANOVA. Since a standard ANOVA will tell you if there is a statistically significant difference somewhere in your data sets but will not tell you where, you need to apply a multiple means comparison test after the ANOVA, to pinpoint the culprits. There are many tests available, each suited to different types of data, so it is advised that you consult the help menu of your own statistical program to decipher which one best fits your own experimental design. As a rule of thumb, do not choose to make any more comparisons than are appropriate for the goals of the experiment, or you risk to lose the ability to detect significant differences that may truly exist. This is because multiple means comparison tests apply corrections that tighten the rejection level in proportion to an increased group comparison number (i.e., effectively lowering the *p*-value required to find a difference significant). In general, you can make one less comparison between groups than the total number of groups, without affecting this rejection level; however, if you exceed this, the rejection level is adjusted. Dunnett's test, for instance, is suited to designs such as a dose–response curve, which include a matched control sample and several different doses; this test compares each group only to the control (not to each other), to answer the question "Which dose(s) give(s) a response" but not "Was the response to one dose different than to another?" Example 6 shows how a significant difference between the control and treatment groups could be missed if more comparisons were made than necessary.

---

**Example 6.** You wish to test for whether any of several doses of your treatment has a significant effect on your outcome. You have collected samples *a*, *b*, *c*, *d*, (i.e., four groups), where *a* is the control and *b*, *c*, *d* represent responses to three different concentrations of your treatment. Assuming that your highest dose, *d*, shows a *p*-value of 0.02 with respect to control, this difference would be reported as significant by Dunnett's test; that is, three comparisons would be made (*b* vs. *a*, *c* vs. *a*, *d* vs. *a*), one less than the total number of groups, so no adjustment to the rejection level would be necessary to compensate for multiple comparisons, and the *d* vs. *a* calculated *p*-value of 0.02 ($\leq 0.05$) would be interpreted as statistically significant.

---

If instead of Dunnett's test, the Bonferroni test were chosen, six comparisons would be made (the same three as above, plus *b* vs. *c*, *b* vs. *d*, and *c* vs. *d*). Since this would be two more comparisons than the number of groups, the significance threshold of 0.05 would be adjusted to compensate for multiple

comparisons (0.05/6 = 0.008), such that a computed $p$-value of ≤0.008 would be required for significance at the so-called 0.05 rejection level (i.e., $p = 0.02$ would not be considered significant).

### One-Tailed vs. Two-Tailed

The tails referred to in your statistical package are the nonwagging variety; the term represents the tail edges of the bell-shaped normal distribution. Tests are one- or two-tailed depending on whether they compute the probability of values exceeding the range expected from members of a population in only one direction or potentially in both directions (i.e., greater than or less than).

Using a one-tailed test is justified only if there was an a priori determination that only differences in one specified direction would be assessed, not because one decides after seeing that the one-tailed $p$ is significant but the two-tailed is not, that one only really cares about one direction anyway. Furthermore, having an expectation that the difference to be found will go in a particular direction is not considered adequate justification for performing one-tailed analysis; rather, one-tailed analysis could be considered appropriate if having a large difference in one direction vs. having no difference at all would have equivalent consequences [39]. Under most circumstances a two-tailed test is the appropriate choice, and if the researcher is in any doubt about which is appropriate, the two-tailed should be chosen.

### Paired vs. Unpaired

Paired analysis is appropriate in the following circumstances:

- When you do a before and after measurement in each subject.
- When you match subjects in pairs (i.e., age, etc.), then treat one of the subjects and not the other. Pairs must be made before data are collected.
- When you compare relatives (i.e., sibling studies).
- When you perform an experiment many times, each time with the experimental and control sample treated in parallel (including log-transformed ratio data, as described in Example 7).

---

**Example 7.** You have done Northern blots of gene expression in mutant animals vs. wild-type animals. Each of your blots shows wildly different signal magnitudes from other blots, due to technical reasons such as exposure time; having mutant and wild-type samples on every blot allows you to control for this variability by expressing data as a ratio of mutant to wild-type values.

Ratios are not normally distributed, but logs of ratios are; therefore, you log-transform your data. Ratios become differences when log-transformed [i.e., log(mutant/wild type) = log(mutant) – log(wild type)], so to analyze this type of data you can take the log of each data point and perform a paired $t$-test on the transformed data.

---

In judging whether distributions are normal in paired testing, it is the differences between the two members of each pair that must be approximately normally distributed rather than the two distributions themselves. If data from two sets to be compared do not meet one of the criteria above, they should be analyzed by unpaired testing [35].

## Miscellaneous Data

***Outliers*** Within any set of observations it is not unusual to obtain occasional data points that vary substantially from others found in the same group of subjects. These *outlier data* may be the result of unknown errors or may represent genuine information, and indeed, whether such data should be included or rejected is the subject of much debate. Methods for detecting outlier data and the use of robust methods of statistical analysis to accommodate such data are complex issues, beyond the scope of this chapter; a thorough treatment of the subject is presented by Barnett and Lewis [40], to which the reader is referred.

***BLQ Data*** Statisticians sometimes warn against *left-censored distributions*, a term evocative of political intrigue and vaguely racy data sets that actually refers to the more bland reality of omitting results that are below the levels of quantitation of the assay (BLQ). BLQ results are fairly common in clinical and preclinical biomarker analyses and can cause serious problems with interpretation of the data if they are not taken into account, since unreported values are unusable in any statistical analysis of the data. Data imputation, a process to replace unreported BLQ values with valid estimates, can also affect the data interpretation, since all values in a data set affect the measures of central tendency, error, and statistical power. It seems clear that estimating a theoretical value for BLQ data points would allow for more accurate conclusions to be drawn from the data than simply removing data through left-censoring, but what value should indeed be assigned to BLQ data points?

Some researchers report BLQ values as zero, perhaps due to the influence of the sometimes black-and-white perspective of assay validation-style cutoffs. Assigning a zero value regardless of the sensitivity of a given test could be misleading, however, since zero is an absolute, whereas BLQ represents a scale from zero up to an upper limit that differs from one test to another in an assay-specific manner. Furthermore, assigning zero values creates other mathematical difficulties, such as not being able to compute fold-differences of another value with respect to the zero, or not being able to work with log transforms of data.

One common approach to address these issues in the data imputation process has been to assign arbitrary nonzero values to BLQ data, most commonly either the limit of detection (LOD) divided by 2 or divided by the square root of 2, depending on the skewness of the data set in question [41]. Both appreciable bias and loss of power have been demonstrated with this

type of approach, however [42,43]. Succop et al. compared the bias of alternative methods for imputing BLQ values and concluded that by imputing the values based on either median percentiles below the detection limit or based on predicting the BLQ value from an equation model fit to the noncensored data set, both produced a good correlation between predicted and reported low values. In their example data set, bias between the predicted and true values was only 2.9% this way, compared to 348% overestimation bias when LOD/2 was used to impute the BLQ values [44].

Standard curves of analytical assays typically provide an equation model fit that can be used to calculate values down to the zero level, but assay validation typically limits the data reported to levels that have met a cutoff level that is deemed to have accuracy and precision within proscribed acceptance limits (i.e., the lower limit of quantitation). The study above suggests that although equation-predicted values below the levels of quantitation of an assay may not meet acceptance criteria per se, they still represent more accurate estimations for BLQ data than either arbitrary imputation methods or left-censoring of data sets. Succop et al. recommended that analytical labs should provide a numerical result for all samples analyzed, with a flag of those values that are below the detection limit [44].

## ISSUES SPECIFIC TO SPECIALIZED FIELDS

### Genetic Association Studies

Genetic association studies evaluate the association between specific genetic polymorphisms and disease. They often assess relatively small effects against a noisy background of biological and social complexity, and consequently, tend to lack statistical power [45]. Power calculations are critical for genetics researchers who wish to map susceptibility genes, with chi-square being the most commonly used test statistic for association studies. The noncentrality parameter for the chi-square asymptotic distribution developed by Mitra [46] is the key to computing power and sample size for this type of study. Gordon and Finch present detailed instructions (including step-by-step use of a Web-based calculator) for this, as well as reviewing other concepts of statistical genetics and the best study/analysis design to optimize power for this type of study [17].

Synthesizing evidence from multiple studies has been used as a means of increasing power; however, it seems that metaanalyses in this field have been fraught with serious flaws. As well as more general concerns, there are genetic issues particular to molecular association studies, including checking Hardy–Weinberg equilibrium (HWE, discussed below), handling data from more than two groups while avoiding multiple comparisons, and pooling data in a way that is sensitive to genetic models [47]. As regards the latter two points, there are always at least three possible genotypes to compare, but in practice

most studies reduce the number of comparisons by assuming a specific genetic model, such as dominant or recessive, even though there are often no biological justifications for assumption of the model. Using an inappropriate genetic model or inappropriately pooled data produce misleading estimates of odds ratios. Appropriate metaanalysis methods and avoiding assumptions about genetic models are discussed further by Minelli et al. [45] and Thakkinstian et al. [48].

Bias seems to be another serious problem in this type of study. Some methods for detecting and correcting it are discussed by Sterne et al. [49]; publication bias, for example, may be detected by a simple funnel plot of the data. Not just the metaanalyses, but also single studies in this field seem excessively prone to various sources of bias, with some critics going so far as to assert that most documented associations, even those that are replicated, represent nothing other than false positives based on bias [50]. The list of prevalent bias sources is impressive and includes biological plausibility bias, ascertainment bias, publication bias, selective reporting bias, spectrum of disease bias, population stratification, biased selection of controls, lack of blinding in the genotyping process, and genotyping error.

Some of these error sources can be addressed simply through researchers being cognizant of them and striving to avoid them. For example, one can choose to blind genotyping rather than failing to do so. In another example, ascertainment bias can be avoided by scrutinizing controls with the same intensity as cases; this type of bias relates to the situation where affected individuals ("cases") have their DNA resequenced to identify rare variants. Identification of such rare variants in an affected group does not necessarily signify a role in disease (as is often assumed when only cases are ascertained by resequencing) since sequencing DNA from any group tends to turn up a few rare mutations. Instead, a strong and statistically convincing *preferential* presence of variants in the cases compared to controls (that had been scrutinized in the same manner) would support the involvement of the variants in disease [51].

In cases where being cognizant of the potential for bias does not allow us to avoid it, designing experiments and/or data analysis with an expectation of inevitable bias can dampen its influence. For example, designing studies to have higher than desired power can combat the power loss expected from genotyping misclassification errors, as discussed earlier. As another example, various statistical techniques that are robust to bias can be used, such as genomic control techniques or family-based methods to address problems of population stratification [17]. *Population stratification* refers to the situation in case–control studies where control groups are not well matched to case groups and in fact have different typical levels of what is being measured simply because they are from different populations. In this situation, association tests like chi-square may falsely indicate associations, even with as high as 100% probability.

In fact, having appropriate control groups is a key determinant of the validity of genetics association studies, and determining whether controls deviate from HWE is a standard way of evaluating this. Theoretically, disease-free control groups from outbred populations should follow HWE, as should combined cases and controls if they both have a particular disease (i.e., such as in studies where different treatments are evaluated). If they do not, it is a signal of some peculiarity, error, or problem with the data sets that could invalidate the key inferences from a genetic association study [52]. For example, if there is a recessive model and the control group has an excess or deficit of one group of homozygotes, this will directly affect calculation of the odds ratio (the control homozygotes divided by the other genotypes is the denominator of the odds ratio). Overall, if there is a significant deviation from HWE it should induce some thinking about the study; deviations could result from genotyping errors but may also arise from other sources. For example, HWE deviation may suggest that allele-based estimates of genetic effects are biased, or may give further insight into the population from which the data are derived.

A recent review of studies published in high-quality specialized genetics journals demonstrated that HWE is very commonly tested improperly or inadequately [3]. Of 776 associations tested, only 29% reported on HWE, introducing uncertainty as to if it was tested in the others. Furthermore, where HWE testing was described, the test was applied to the correct control group in only 50% of the associations tested. It is a common error to include the disease cases with the control cases in the HWE test when controls are disease-free. Combined cases and controls should be tested only when the controls have the same disease as the cases; otherwise, only controls should be tested. These authors recalculated HWE for the studies reviewed and noted that in most of the samples where HWE was actually violated, this was either not mentioned in the original article or HWE conformity was actually claimed.

Another common problem in the studies reviewed was the unjustified use of an inappropriate test. A number of different statistical tests could be used to test HWE (including chi-square, exact tests, and Bayesian methods), all based on the conditional probability that there would be the number of homozygotes that actually turned out to be in the sample. The chi-square test was the only test applied to HWE calculation in the studies reviewed, despite the fact that the chi-square asymptotic distribution is inadequate to deal with low genotype frequencies and is therefore not justified in studies involving them; in this case an exact test provides a simple and superior alternative [3,54].

A final serious concern in the studies reviewed was that only 7% of the studies had an acceptable power to detect HWE deviation, with most studies being much too underpowered to make any claim of lack of deviations. Whereas power to detect HWE is of secondary importance compared to the prime consideration of power to detect a genetic association, undetected modest HWE deviations could affect considerably the inferences of many genetic association studies [3].

## Microarray or Other "Omics"-Type Data

Relating "omics"-style biomarker data (gene/protein arrays, multiple *m*/*z* peaks from mass spectrometry, etc.) to clinical outcomes is the focus of a large body of research. Making statistically based decisions about what is expressed or produced differentially, and how this relates to outcomes is not simple for these types of studies. For one, the data are highly dimensional, with the number of variables often exceeding the number of samples by magnitudes (i.e., in contrast to the desirable format for statistical comparisons of a group of sufficient numbers of samples for each variable tested). Furthermore, data tend to be very "noisy," with widely different magnitudes of signals and the potential for large irrelevant signals to obscure lesser (but relevant) ones. This translates into an inherent tendency for algorithms fit to the data to be more influenced by noise than by salient interest points (a phenomenon known as *overfitting the noise*).

A recent review of analysis of genomics/proteomics data outlines the steps of quality control, clustering, classification, feature ranking, and validation [55]. In Joe/Josephine terms these could be described as checking and normalizing the data, grouping like things together, statistically evaluating their likely involvement with the outcome, getting rid of the dead wood, and checking the conclusions on an independent data set.

In the first stage, quality control and normalization of signals reduce the technical variability and ensure that biomarkers discovered are statistically significant. Cui and Churchill reviewed statistical tests that have been specifically adapted to cDNA microarray data analysis [56]; many of the issues discussed are similar for highly dimensional data from other study types. They conclude that fold-change is the simplest method for detecting differential expression of genes, but the arbitrary nature of assigned cutoff values, the lack of statistical confidence measures, and the potential for biased conclusions all detract from its appeal. They also stress the need to normalize data properly, such as by using intensity-specific thresholds, since otherwise an excess of low-intensity genes may be identified as being expressed differentially simply because their fold-changes have larger variance than the fold-change values of high-intensity genes.

Expression analysis steps following this generally fall into two categories, supervised or unsupervised, based on whether or not they are led by prior knowledge. Clustering is an *unsupervised method* that groups genes of similar expression profile together; clustered genes may have diverse biological functions, but this is nonetheless useful for exploring new territory and seeing who is dancing together. *Supervised classification methods*, on the other hand, train a mathematical model using prior biological knowledge of certain genes' functional involvement with each other, then use the model for predictions on other genes. This is a useful approach for building on what is already known, but in cases of limited available information it is difficult for a supervised method to achieve accurate predictions.

Flawed statistical analysis is very common in this emerging field, even in studies published in high-impact factor journals. Dupuy and Simon recently found that half of a survey of 90 microarray studies had at least one of three major analysis flaws [2]. (These authors also present a list of analysis do's and don'ts specific to this study type, to which the reader is referred for more detail.) Analysis flaws like those discussed in this review and elsewhere [53,57] can lead to completely erroneous conclusions about the utility of given biomarkers or "signatures" of multiple biomarkers as prognostic or diagnostic tools, and may underlie some of the problems reproducing data in this field.

One of the most common flaws present in the published studies was to perform multiple comparisons without making adequate adjustment for the expected increased false-positive rate. Although the standard of setting $p$ at less than or equal to 0.05 is accepted as a reasonable compromise between type I and II errors when data sets are relatively small, in the relatively large-omic data sets a $p$-value this high is not sufficiently stringent to identify disease screening markers with an acceptably low false-positive rate. For example, a microarray containing 5000 genes examined for meaningful relationships between the genes and some outcome using this $p$-value could be expected to yield 250 false-positive relationships due simply to chance.

Dupuy and Simon suggest that the simplest method to combat this type of error is to use a $p$-value of less than or equal to 0.001, which would give one false positive for every 1000 genes tested. Several other accepted methods are grouped under the term *family-wise error-rate control* (FWER), but they have been criticized for decreasing power substantially [58]; the Bonferroni correction (dividing the nominal significance level by the number of tests) is the simplest of these. The false detection rate (FDR) test (or Benjamini–Hochberg test) is a favored control statistic for microarray analysis and is included in statistical analysis packages such as SAM (statistical analysis for microarray). The FDR achieves the same control of false significant results while greatly improving the statistical power compared to FWER methods [59,60]. An FDR of 5% is not the same as a $p$-value of 0.05: $p = 0.05$ means that 5% of tested genes will be false positive, whereas 5% FDR means that 5% of positive genes detected will be false. For example, given a scenario of 5000 genes tested, with 20 of them being found positive, a $p$-value of 0.05 means that there could have been 250 false positives (due to chance alone), whereas an FDR of 0.05 means that only one false positive was attributable to chance alone.

A second common flaw was to make spurious claims of a meaningful correlation between clusters of genes and clinical outcome, when, in fact, correlation was detected after clustering using a selection of outcome-related differentially expressed genes. Therefore, genes being related to the outcome was a consequence of the selection bias, but was instead reported as independent proof of a correlation between the clinical outcomes and the gene clusters. A number of classification methods decrease noise in the data by requiring selection of the relevant and informative predictor variables before modeling

is performed [61], so users need to be aware that if they have used one of these methods there is no point in performing correlation analysis afterward.

The third major flaw concerned the mechanics of classification model validation. Once a model is constructed it is generally validated using either an external test set or cross-validation, with the end goal being to use the validated model to predict outcomes for unknown samples. In the studies reporting supervised prediction, Dupuy and Simon noted the common flaw of biased, overly optimistic estimation of the prediction accuracy through an incorrect cross-validation procedure. Different errors in cross-validation procedure were noted, but generally they were all related to failing to have test sets (data used to test the model's predictive capacity) completely uninvolved in generating the algorithm in the first place.

## Bayesian Information Integration for Personalized Medicine

A key challenge for the success of pharmacogenomics lays in defining analytical methods to integrate and interpret the increasingly complex information that it provides, so that decisions can be made about how people can best benefit from certain therapeutic interventions. Bayesian statistics approaches, often favored by decision theorists, have proven useful in classification models that integrate traditional clinical risk factors with complex genomics data. The point of doing this is to try to predict the outcome of future patients who share certain key clinicogenomic attributes with those from whom the model was derived and to use these predictions for patient-customized diagnostic, prognostic, and treatment purposes.

Most of this chapter is based on the classical frequentist approach to inferential statistics with which most of us are familiar. In this approach, a single hypothesis (the null hypothesis) is formed about the effect of (or, rather, lack thereof) a variable or several variables that are tested using a fairly rigid and well-defined format. The evidence against the hypothesis is all that is considered, and it is in the form of probabilities computed from the distribution of the outcome measure(s). A baseline assumption of randomness is assumed in defining what the probabilities are, and rejection of a null hypothesis is founded on the experimental result being too different from the random probability function to actually be random (i.e., therefore it isn't).

Bayesian statistics, on the other hand, represent a fundamentally different way of thinking about probability, with statistical inferences being made using accumulated evidence of all kinds, both for and against a hypothesis, or several alternative hypotheses. These hypotheses do not even have to be testable, let alone be tested, and certainly not by any particular format. (The name of the first publication of Bayes himself shows an example of such a hypothesis: Divine Benevolance, or an Attempt to Prove that the Principal End of Divine Providence and Government Is the Happiness of His Creatures [62].) Importantly, there is no baseline assumption of randomness: Instead, a "prior probability" is included in the paradigm. Although assigning it a value based

on random chances is one option, another is that prior knowledge is incorporated into this value; in this way, Bayesian methods build on past experience to make statistical inferences.

To compute the probability of each particular hypothesis being true, a calculation using Bayes' theorem would multiply all the conditional probabilities that each is true by their prior probabilities and weight the final values relative to each other. *Conditional probabilities* are those from each piece of evidence for or against the hypothesis, and *prior probability* is defined in a very open-ended way; it could be pretty much anything, including subjective feelings, provided that it is decided upon in a consistent manner between all the hypotheses, it sums to 1 for all the hypotheses, and the information used to produce it is not used again later as evidence. As more evidence accumulates, it can be incorporated into updated models [63]. A simple Bayes calculation has already been shown (see Example 2); in that case, known disease frequency in the population constituted the prior probability, and the predictive values of the diagnostic test for true and false positives were the conditional probabilities.

Biomarker researchers aiming to define predictive models for personalized medicine have showed how Bayesian methods can strengthen these models by integrating different types of information into them. For instance, Qi et al. noted that both supervised and unsupervised methods have different drawbacks for genomics data analysis, with the latter capturing relevant expression profiles but ignoring important a priori knowledge, and the former being limited where little prior knowledge exists. Accordingly, they used a Bayesian approach to perform accurate predictions even when scant prior knowledge is available; their integrated model achieved higher sensitivity and specificity compared to both unsupervised and supervised methods [64].

Similarly, Pittman et al. [65] used Bayesian tree modeling to integrate both clinical information and genomic biomarker status for a sample set of breast cancer patients. They showed that the integrated clinicogenomic model gave substantially greater log-model likelihood (>7) than the model based on genomic biomarker signatures alone, and blew the clinical predictors alone right out of the picture, with a weight of evidence for the clinicogenomic vs. the clinical predictors only of more than 26 log-likelihood units.

Tree models in general involve successive splitting of a given patient sample group with certain known characteristics (i.e., gene signature, clinical risk factors) and outcomes (e.g., cancer status, survival, relapse) into more and more homogeneous subgroups. At each split, the collection of evidence (e.g., clinical or gene factors) is sampled to determine which of them optimally divide the patients according to their outcome, and a split is made if significance exceeds a certain level. Multiple possible splits generate "forests" of possible trees.

Some caveats noted for those using these methods include how to choose among alternative potential models that are identified as being of similar or significant probability, and the issue of uncertainty. In that regard, whereas in

some other applications Bayesian approaches are used to choose a single best hypothesis from among several, such an approach is warned against for pharmacogenomic modeling applications [61,65,66]. In this case it is typical to see multiple plausible tree models representing the data adequately, and this is consistent with the physical reality of multiple plausible combinations of genetic and clinical factors that could lead to the same outcome measures.

Rather than choosing one of them, it is critical to define overall predictions by averaging across the multiple candidate models using appropriate weights that reflect the relative fits of the trees to the data observed. The impact of averaging is seen in greater accuracy of the model predictive capacity and, importantly, in accurate estimation of the uncertainty about the resulting prediction uncertainty (i.e., the prediction uncertainty of such a model is conceptually akin to the measurement uncertainty associated with a laboratory method, as discussed earlier). Nevins et al. underlined well the importance of prediction uncertainty when they pointed out: "A further critical aspect of prognosis is the need to provide honest assessments of the uncertainty associated with any prediction. A predicted 70% recurrence probability, for example, should be treated quite differently by clinical decision makers if its associated uncertainty is ±30% than if it were ±2% …" [66].

## SUMMARY: QUICK DO'S AND DON'TS

The field of biomarkers is a wide one, with a huge diversity of potential applications, all of which may have different and complex statistical analysis issues. This chapter has undoubtedly missed many of these and glossed over others, but summarizing concisely what it has covered is still a challenge. Instead, I present below a distillation in the form of one-liners addressing some of the more critical points ("do") and more common misconceptions ("don't").

Do:

1. Put on Hercule Poirot's hat and use your judgment when considering analytical issues and your experimental situation.
2. Think ahead: Consider data analysis issues before settling on your experimental design (and long before having the data in hand).
3. Take the limitations of your techniques and experimental error into account in your interpretations.
4. Check that your design is powered appropriately to detect what you wish to detect.
5. Use appropriate tests (i.e., ANOVA when comparing means from more than two groups).
6. Use appropriate controls and check HWE in genetic association studies.
7. Make adequate adjustment for the elevated false-positive rates when dealing with-omics-style data.

8. Average over multiple plausible candidate models with appropriate weights (rather than choosing a single one), for best predictive accuracy and uncertainty estimations in pharmacogenomic applications of Bayesian modeling strategies.

Don't:

1. Use parametric tests if data do not meet the assumptions of these tests (such as being normally distributed).
2. Use repeated *t*-tests when comparing means from more than two groups in one experimental design.
3. Compare more groups than are relevant to the goals of your experiment when applying multiple means testing.
4. Use correlation analysis to infer cause and effect or to test agreement between different methods.
5. Simply equate the odds ratio and risk ratio without considering the outcome frequency.
6. Use data to test correlation to outcomes or as a test set for validation of algorithms if they were already involved with those earlier in the process of mathematical modeling of genomics data (i.e., selected based on outcomes in the former case, or involved in algorithm generation in the latter).

## REFERENCES


1. Ludbrook J (2001). Statistics in physiology and pharmacology: a slow and erratic learning curve. *Clin Exp Pharmacol Physiol*, 28(5–6):488–492.
2. Dupuy A, Simon RM (2007). Critical review of published microarray studies for cancer outcome and guidelines on statistical analysis and reporting. *J Nat Cancer Inst*, 99(2):147–157.
3. Salanti G, Amountza G, Ntzani EE, Ioannidis JP (2005). Hardy–Weinberg equilibrium in genetic association studies: an empirical evaluation of reporting, deviations, and power. *Eur J Hum Genet*, 13:840–848.
4. Attia J, Thakkinstian A, D'Este C (2003). Meta-analyses of molecular association studies: methodological lessons for genetic epidemiology. *J Clin Epidemiol*, 56:297–303.
5. Ransohoff D (2005). Lessons from controversy: ovarian cancer screening and serum proteomics. *J Nat Cancer Inst*, 97(4):315–319.
6. Goncalves A, Borg JP, Pouyssegur J (2004). Biomarkers in cancer management: a crucial bridge towards personalized medicine. *Drug Discov Today: Ther Strategies*, 1(3):305–311.
7. Ransohoff DF (2005). Bias as a threat to validity of cancer molecular-marker research. *Nat Rev Cancer*, 5:142–149.
8. Gawrylewski A (2007). The trouble with animal models. *Scientist*, 21(7):45–51.

9. Hubert Ph, Nguyen-Huu JJ, Boulanger B, et al. (2004). Harmonization of strategies for the validation of quantitative analytical procedures: a SFSTP proposal—part 1. *J Pharm Biomed Anal*, 36(3):579–586.
10. Findlay JWA, Smith WC, Lee JW, et al. (2000). Validation of immunoassays for bioanalysis: a pharmaceutical industry perspective. *J Pharm Biomed Anal*, 21: 1249–1273.
11. Hubert Ph, Nguyen-Huu JJ, Boulanger B, et al. (2006). Validation des procédures analytiques quantitatives: harmonisation des démarches. Partie II—Statistiques. *STP Pharma Prat*, 16:30–60.
12. Prudhomme O'Meara W, Fenlon Hall B, Ellis McKenzie F (2007). Malaria vaccine efficacy: the difficulty of detecting and diagnosing malaria. *Malaria J*, 6:136.
13. Carlin BP, Louis TA (1996). *Bayes and Empirical Bayes Methods for Data Analysis*. Chapman & Hall, London.
14. Phillips CV, Maldonado G (1999). Using Monte Carlo methods to quantify the multiple sources of error in studies. *Am J Epidemiol*, 149:S17.
15. Plant N, Ogg M, Crowder M, Gibson G (2000). Control and statistical analysis of in vitro reporter gene assays. *Anal Biochem*, 278:170–174.
16. Phillips CV, LaPole LM (2003). Quantifying errors without random sampling. *BMC Med Res Methodol*, 3:9.
17. Gordon G, Finch SJ (2005). Factors affecting statistical power in the detection of genetic association. *J Clin Invest*, 115(6):1408–1418.
18. Cohen MJ (1998). *The Penguin Thesaurus of Quotations*. Penguin Books, Harmondsworth, UK.
19. Coakley EH, Kawachi I, Manson JE, Speizer FE, Willet WC, Colditz GA (1998). Lower levels of physical functioning are associated with higher body weight among middle-aged and older women. *Int J Obes Relat Metab Disord*, 22(10):958–965.
20. Zou KH, Tuncali K, Silverman SG (2003). Correlation and simple linear regression. *Radiology*, 227:617–628.
21. Hill AB (1965). The environment and disease: association or causation? *Proc R Soc Med*, 58:295–300.
22. Phillips CV, Goodman KJ (2006). Causal criteria and counterfactuals; nothing more (or less) than scientific common sense. *BMC Med Res Methodol*, 3:5.
23. Bland JM, Altman DG (1986). Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*, 1(8476):307–310.
24. Browner WS, Newman TB (1987). Are all significant *p* values created equal? The analogy between diagnostic tests and clinical research. *J Am Med Assoc*, 257:2459–2463.
25. Eng J (2003). Sample size estimation: How many individuals should be studied? *Radiology*, 227:309–313.
26. Eng J (2004). Sample size estimation: a glimpse beyond simple formulas. *Radiology*, 230:606–612.
27. Detsky AS, Sackett DL (1985). When was a "negative" clinical trial big enough? How many patients you need depends on what you found. *Arch Intern Med*, 145:709–712.
28. Sokal RR, Rohlf FJ (1981). *Biometry*, 2nd ed. W.H. Freeman, New York.

29. Browner WS, Newman TB, Cummings SR, Hulley SB (2001). Estimating sample size and power. In Hulley SB, Cummings SR, Browner WS, Grady D, Hearst N, Newman TB (eds.), *Designing Clinical Research: An Epidemiological Approach*, 2nd ed. Lippincott Williams & Wilkins, Philadelphia, pp. 65–84.
30. Sistrom CL, Garvan CW (2004). Proportions, odds, and risk. *Radiology*, 230:12–19.
31. Motulsky H (1995). *Intuitive Biostatistics*. Oxford University Press, Oxford, UK.
32. Agresti A (2002). *Categorical Data Analysis*. Wiley, Hoboken, NJ.
33. Schulman KA, Berlin JA, Harless W, et al. (1999). The effect of race and sex on physicians' recommendations for cardiac catheterization. *N Engl J Med*, 340: 618–626.
34. Schwartz LM, Woloshin S, Welch HG (1999). Misunderstandings about the effects of race and sex on physicians' referrals for cardiac catheterization. *N Engl J Med*, 341:279–283.
35. Ryder EF, Robakiewicz P (1998). Statistics for the molecular biologist: group comparisons. In Ausubel FM, Brent R, Kingston RE, et al. (eds.), *Current Protocols in Molecular Biology*. Wiley, New York, pp. A.31.1–A.31.22.
36. Bland JM, Altman DG (1996). The use of transformation when comparing two means. *BMJ*, 312(7039):1153.
37. Bland JM, Altman DG (1996). Transforming data. *BMJ*, 312(7033):770.
38. Ludbrook J (1995). Issues in biomedical statistics: comparing means under normal distribution theory. *Aust N Z J Surg*, 65(4):267–272.
39. Bland JM, Altman DG (1994). One and two sided tests of significance. *BMJ*, 309:248.
40. Barnett V, Lewis T (1994). *Outliers in Statistical Data*, 3rd ed. Wiley, New York.
41. Hornung RW, Reed DL (1990). *Appl Occup Environ Hyg*, 5:46–51.
42. Hughes MD (2000). Analysis and design issues for studies using censored biomarker measurements with an example of viral load measurements in HIV clinical trials. *Stat Med*, 19:3171–3191.
43. Thiebaut R, Guedj J, Jacqmin-Gadda H, et al. (2006). Estimation of dynamic model parameters taking into account undetectable marker values. *BMC Med Res Methodol*, 6:38.
44. Succop PA, Clark S, Chen M, Galke W (2004). Imputation of data values that are less than a detection limit. *J Occup Environ Hyg*, 1(7):436–441.
45. Minelli C, Thompson JR, Abrams KR, Thakkinstian A, Attia J (2005). The choice of a genetic model in the meta-analysis of molecular association studies. *Int J Epidemiol*, 34:1319–1328.
46. Mitra SK (1958). On the limiting power function of the frequency chi-square test. *Ann Math Stat*, 29:1221–1233.
47. Attia J, Thakkinstian A, D'Este C (2003). Meta-analyses of molecular association studies: methodological lessons for genetic epidemiology. *J Clin Epidemiol*, 56:297–303.
48. Thakkinstian A, McElduff P, D'Este C, Duffy D, Attia J (2005). A method for meta-analysis of molecular association studies. *Stat Med*, 24(9):1291–1306.
49. Sterne JAC, Egger M, Smith GD (2001). Investigating and dealing with publication and other biases in meta-analysis. *BMJ*, 323:101–105.

50. Ntzani EE, Rizos EC, Ioannidis JPA (2007). Genetic effect versus bias for candidate polymorphisms in myocardial infarction: case study and overview of large-scale evidence. *Am J Epidemiol*, 165(9):973–984.
51. Hirschhorn JN, Altshuler D (2002). Once and again: issues surrounding replication in genetic association studies. *J Clin Endocrinol Metab*, 87(10):4438–4441.
52. Khoury MJ, Beaty TH, Cohen BH (1993). *Fundamentals of Genetic Epidemiology*. Oxford University Press, New York.
53. Baggerly KA, Morris JS, Edmonson SR, Coombes KR (2005). Signal in noise: evaluating reported reproducibility of serum proteomic tests for ovarian cancer. *J Nat Cancer Inst*, 97(4):307–309.
54. Emigh T (1980). A comparison of tests for Hardy–Weinberg equilibrium. *Biometrics*, 36:627–642.
55. Phan JH, Quo CF, Wang MD (2006). Functional genomics and proteomics in the clinical neurosciences: data mining and bioinformatics. *Prog Brain Res*, 158:83–108.
56. Cui X, Churchill GA (2003). Statistical tests for differential expression in cDNA microarray experiments. *Genome Biol*, 4:210.
57. Liotta LA, Lowenthal M, Mehta A, et al. (2005). Importance of communication between producers and consumers of publicly available experimental data. *J Nat Cancer Inst*, 97(4):310–314.
58. Dudoit S, Schaffer JP, Boldrick JC (2003). Multiple hypothesis testing in microarray experiments. *Stat Sci*, 18:71–103.
59. Benjamini Y, Hochberg Y (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc B*, 57:289–300.
60. Benjamini Y, Drai D, Elmer GL, Kafkafi N, Golani I (2001). Controlling the false discovery rate in behavior genetics research. *Behav Brain Res*, 125:279–284.
61. Tong W, Xie Q, Hong H, et al. (2004). Using Decision Forest to classify prostate cancer samples on the basis of SELDI-TOF MS data: assessing chance correlation and prediction confidence. *Toxicogenomics*, 112:1622–1627.
62. Dale AI (2003). *Most Honourable Remembrance: The Life and Work of Thomas Bayes*. Springer-Verlag, New York.
63. Strachan T, Read AP (2003). *Human Molecular Genetics*. Garland Science, London.
64. Qi Y, Missiuro PE, Kapoor A, et al. (2006). Semi-supervised analysis of gene expression profiles for lineage-specific development in the *Caenorhabditis elegans* embryo. *Bioinformatics*, 22(14):e417–e423.
65. Pittman J, Huang E, Dressman H, et al. (2004). Integrated modeling of clinical and gene expression information for personalized prediction of disease outcomes. *Proc Nat Acad Sci USA*, 101(22):8431–8436.
66. Nevins JR, Huang ES, Dressman H, Pittman J, Huang AT, West M (2003). Towards integrated clinico-genomic models for personalized medicine: combining gene expression signatures and clinical factors in breast cancer outcomes prediction. *Hum Mol Genet*, 12:R153–R157.

# QUALIFICATION OF SAFETY BIOMARKERS FOR APPLICATION TO EARLY DRUG DEVELOPMENT

William B. Mattes, Ph.D., DABT
*The Critical Path Institute, Rockville, Maryland*

Frank D. Sistare, Ph.D.
*Merck Research Laboratories, West Point, Pennsylvania*

## HISTORICAL BACKGROUND TO PRECLINICAL SAFETY ASSESSMENT

It is often forgotten that the first "blockbuster" drug, sulfanilamide, was discovered and developed in an era devoid of regulatory oversight and guided only by free-market forces. Domagk discovered the antibacterial properties of Prontosil in 1932, and with the subsequent discovery in 1935 that the active moiety was the off-patent and widely available substance sulfanilamide, a number of companies rushed to make preparations for sale to the public. This explosion of therapy options was unfettered by requirements for medicines to be tested for efficacy or safety, although preparations could receive the endorsement of the American Medical Association [1]. Thus, in 1937, when the S.E. Massengill Company of Bristol, Tennessee, sought to prepare a flavored syrup, it simply identified an appropriate excipient to dissolve the drug, prepared 240 gallons of the raspberry-tasting Elixir Sulfanilamide, and marketed it across the nation. Unfortunately, the excipient chosen was diethylene glycol, also used as an antifreeze. We now know this agent to be lethal to a

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

large number of species, causing acute kidney injury at relatively modest doses. Before Elixer Sulfanilamide was identified as the causative agent and pulled from pharmacies, 34 children and 71 adults died. A year later, Congress passed the 1938 Federal Food, Drug and Cosmetic Act, requiring pharmaceutical manufacturers to show product safety before distribution [1]. Thus, society transitioned from its former free-market approach to pharmaceutical development to one of safety testing (initially in animals), careful clinical trials, and government oversight.

Safety testing as it is practiced today owes most of its form to procedures developed by the U.S. Food and Drug Administration (FDA) for testing food [2]. As early as 1949, the publication of "Procedures for the Appraisal of the Toxicity of Chemicals in Foods" began to formalize the practices the agency expected industry to follow in safety testing [3]. These practices included standard study designs and expectations as to what experimental observations would be recorded. They have evolved into the descriptive toxicity tests well known to modern toxicology [4]. Key to the value of these tests is not only their experimental design in terms of dose, route of administration, and duration, but also the endpoints evaluated, going beyond the observations of overall animal behavior and health. Thus, a battery of clinical pathology tests examining urine, hematological parameters, and serum chemistry is commonly evaluated [5]. Importantly, a number of tissues from the animal are examined both macroscopically and microscopically after sacrifice, and this histopathological examination allows for the identification of unusual and subtle lesions and changes following compound treatment [6].

In the arena of pharmaceutical product development, these studies, carried out in at least two animal species, one rodent and one nonrodent, are used to assure the safety of human subjects exposed to experimental doses of novel compounds [7]. The types and durations of studies required to support safety in various types of clinical studies have been codified by the International Conference on Harmonisation (ICH) and described in their guidelines on nonclinical safety studies [8]. Even so, human subjects need to be monitored for "adverse events and/or laboratory abnormalities identified in the protocol as critical to safety evaluations" [9].

## LIMITATIONS FACED IN PRECLINICAL SAFETY ASSESSMENT

A critical problem faced by nonclinical safety assessment groups in pharmaceutical drug development is the disparity of responses sometimes seen between the two nonclinical test species in the tools used to assess these responses. Historically, microscopic histopathology is used as a primary tool for identifying compound-induced damage. When damage is identified only at exposures far exceeding those expected in clinical studies, clinical safety is expected. However, microscopic histopathology is not a tool generally applicable to human studies, where clinical pathology measurements play the criti-

cal role in assessing adverse responses to drugs. Thus, if damage is observed in one nonclinical species at exposures close to those anticipated for human studies, the crucial question is whether the onset and reversibility of such damage could be monitored with clinical pathology or some other relatively noninvasive technology. Unfortunately, as described here, there are several types of drug-induced organ injury where current clinical pathology assays do not detect damage with sufficient certainty at early stages, and where assurances are needed that discontinuation of drug treatment would be followed by a complete and swift return to normal structure and function.

## Kidney Injury

Kidney injury may be produced by a variety of insults, including those induced by drugs or toxicants [10]. Given the known morbidity and mortality associated with acute kidney injury [11,12], evidence of drug-induced kidney injury in preclinical studies is a matter of serious concern. While on one hand the kidney is capable of recovery from mild damage [13] if the injurious agent is removed, the very real clinical problem is that traditional noninvasive measures of kidney function are insensitive and confounded by many factors [13–16]. Thus, even modest increases in serum creatinine are associated with significant mortality [12]. There is a real need for noninvasive markers that would detect kidney damage or loss of function at a stage before significant and irreversible damage has occurred. Several markers with just such a potential have been described in numerous reviews [13–20]. However, many of these are described in relatively few clinical studies, most have not been examined carefully for their performance in animal models of drug-induced kidney injury, and no consensus understanding between drug development sponsors and regulatory review authorities had been reached as to their utility for regulatory decision-making purposes. Ultimately, if a microscopic histopathological examination shows evidence of even mild drug-induced kidney injury in a preclinical study at exposures close to those anticipated for clinical use, development of that compound may be stopped, even if human relevance may be questioned, yet unproven.

## Liver Injury

The fact that medicines can cause liver injury and failure has been appreciated for some time [21], and drug-induced liver injury remains a serious public health and drug development concern [22–25]. As with other drug-induced organ damage, it may be detected in preclinical studies through microscopic histopathology and clinical chemistry measurements. Since the late 1950s, serum transaminase measurements, in particular that of alanine aminotransferase (ALT), have served as a sensitive and less-invasive measure of liver damage in both animal and human settings [26]. In conjunction with serum

cholesterol, bilirubin, alkaline phosphatase, and other factors, ALT has served as a translational biomarker for drug-induced liver injury [27–30]. However, ALT elevations are not always associated with clear evidence of liver injury [31–33], and ALT elevations cannot clearly indicate the etiology of damage [26,27,30,34]. Furthermore, ALT measurements either alone or with bilirubin cannot distinguish patients on a trajectory to severe liver disease and inability to heal or recover from injury, from patients with a full capacity to compensate and return ALT levels to normal despite continuation of drug dosing [35]. Combinations of clinical pathology changes have been used, including ALT and bilirubin, to assure safety in clinical trials [36] but there remains a need for diagnostic assays that reliably link and/or predict the histological observation of liver injury in both a preclinical and clinical setting, and discriminate the types and trajectory of apparent injury [30].

## Vascular Injury

Although injury to the vascular system is known to be caused by a variety of agents [37], many classes of therapeutic agents produce vascular lesions, in preclinical species with or without clinical signs, and with normal routine clinical pathology data [38]. Often, different preclinical species show a different level and type of response, and in many cases (e.g., minoxidil) the vascular injury reported in preclinical species is not observed in a clinical setting [39]. Drug-induced vascular injury in animals may result from altered hemodynamic forces, from a direct chemical-mediated injury to cells of the vasculature, and/or to an indirect immune-mediated injury of the endothelium and/or medial smooth muscle. The conundrum faced in drug development is that there are no specific and sensitive biomarkers of endothelial and/or vascular smooth muscle injury that are clearly linked to the histological observations in animals and could be used to monitor for injury in clinical settings. Although it is assumed that an inflammatory component may be active at some stage in this process, and biomarkers are proposed for such processes [40], biomarkers that are sufficiently sensitive at early and fully reversible stages of vascular injury have not been fully evaluated [38,41]. Furthermore, specific markers of vascular injury/inflammation are sought that can discriminate from the multitude of other more benign causes for elevations of inflammatory biomarkers.

## Drug-Induced Skeletal Myopathy

With the introduction of hydroxymethylglutaryl-coenzyme A (HMG-CoA) reductase inhibitors (statins), not only was there a successful treatment of hypercholesterolemia and dyslipidemia, but soon also a heightened awareness of the issue of drug-induced myopathy [42]. Statin-induced myotoxicity ranges from mild myopathy to serious and sometimes fatal rhabdomyolysis. Not surprisingly, a variety of drugs have been reported to induce myotoxicities

[43]. While skeletal muscle toxicity may be monitored with elevations in serum creatinine kinase (CK), urinary myoglobin, and other markers [42], these biomarkers lack the sensitivity to definitively diagnose early damage or to distinguish the various etiologies of the muscle injury [43]. Thus, there is a need for markers of drug-induced muscle injury with improved sensitivity, specificity, and general utility.

## WHY QUALIFY BIOMARKERS?

More often than not, new biomarkers are judged on the basis of whether they have been subjected to a process of validation. Strictly speaking, *validation* is a process to "establish, or illustrate the worthiness or legitimacy of something" [44]. For judgments of biomarker worthiness or legitimacy, an assessment is needed of both (1) the assay or analytical method to measure the biomarker, and (2) the performance against expectations of the biomarker response under a variety of biological or clinical testing conditions. The term *validation* reasonably applies to the first of these, the process by which the technical characteristics of an assay of a biomarker are defined and determined to be appropriate for the desired measurements [45]. Thus, Wagner has defined validation as the "fit-for-purpose process of assessing the assay and its measurement performance characteristics, determining the range of conditions under which the assay will give reproducible and accurate data" [46]. Even for assay validation, the concept of fit-for-purpose is introduced, which connotes that the process depends on context, and its level of rigor depends on the application of and purpose for the assay. Thus, a biomarker used for an exploratory purpose may not require the more rigorous analytical validation required of a biomarker used for critical decision making. The elements of biomarker assay validation that would be addressed for different categories of biomarker data and for different purposes have been discussed in this book and elsewhere, and they essentially constitute a continuum of bioanalytical method validation [45,47]. Such technical bioanalytical assay method validation is a familiar process and does not generally pose a problem for an organization embarked on assay development [45].

The term *validation* has also been applied to a process by which a new test method is confirmed to be broadly applicable to interpretation of biological meaning in a wide variety of contexts and uses, such as in the validation of alternatives to animal tests as overseen by the Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) [48]. This process involving assessment of biological performance expectations is in contrast to that of *qualification*, which Wagner defines as "the fit-for-purpose evidentiary process of linking a biomarker with biological processes and clinical endpoints" [46]. As for assay validation, this fit-for-purpose biological qualification concept marries the nature and extent of testing rigor to the intended application. In the case of biomarkers applied to predicting human

outcomes, four general phases have been proposed [49], and as the level of qualification progresses, the utility of a biomarker in clinical use increases [46].

In the case of biomarkers of safety (i.e., those that are used to predict or diagnose adverse responses to drug treatment), this qualification process will necessarily involve certain steps. As with qualification of a clinical disease or outcome marker, these steps link biomarker results with biological processes and endpoints. For example, in qualifying a biomarker for nonclinical use, the levels of a protein biomarker in urine may be correlated with certain chemically induced microscopic histopathology lesions in the kidneys of treated animals and thus serve as a noninvasive diagnostic of the appearance of that lesion. As with a clinical disease or outcome marker, qualification of such a marker could proceed in stages. For the example given, an initial stage may be the correlation described, using a variety of treatments that produce only that lesion measured. Such a correlation could establish the sensitivity of the biomarker. However, establishing the specificity of that biomarker for that particular lesion would require a number of treatments that did not produce the lesion being monitored, but instead produce no lesions in the kidney or anywhere else, produced different lesions in the kidney, and/or produced lesions in different organs. Furthermore, the diversity of chemical treatments (i.e., structural and mechanistic diversity) would also need to be considered in the qualification such that a variety of mechanisms underlying the genesis and progression of the lesion can be evaluated.

Clearly, more data could support a higher level of qualification and thus a higher level of utility. For biomarker qualification, the highest phase or level of qualification is the status *surrogate endpoint*, in which case the biomarker can, in fact, substitute for and serve as the new standard clinical endpoint of how a patient feels, functions, or will survive: for example, in efficacy determinations to support marketing approval decisions. Qualified biomarkers that fall short as surrogate endpoints are nevertheless extremely valuable for both drug development and the general practice of medicine. The key to appreciating the value that will come from opportunities to deploy such biomarkers appropriately is in revealing a thorough understanding of their inherent strengths and limitations. In the early steps of designing studies to test sensitivity and specificity aspects of a biomarker's performance to reveal that thorough understanding, the strategy may be relatively clear. For safety biomarker qualification, the first and most important attributes to benchmark are knowledge of the biomarker link to biology and outcome, sufficient test sensitivity, and minimizing false test negatives. Evaluating the response of a new proposed biomarker in animals against biomarkers in conventional use using an agreed-upon set of well-recognized toxicants known to induce the desired organ injury is a fairly straightforward strategy. Pivotal, however, to the successful execution of such studies is to provoke a sufficient number of cases where the timing of the samples taken and the choice of dose levels will yield subtle and mild treatment-related effects at the boundary between normal and abnormal. A full spectrum of histologic treatment-related lesions from very

slight, to slight, mild, moderate, marked, and severe will be important in this regard for evaluating biomarker sensitivity.

The approaches taken for qualification of a safety biomarker for clinical uses will necessarily be different from those taken for nonclinical uses. Clearly, one cannot expect to have studies with healthy subjects intentionally exposed to a variety of toxicants, nor can one regularly use microscopic histopathology as a benchmark for clinical toxicity. Nonetheless, the goal is to reproducibly link the biomarker to a clinical outcome currently recognized and widely accepted as adverse. For certain types of drug-induced injury, there are standard-of-care treatments that unfortunately are associated with a known incidence of drug-induced injury. As an example, aminoglycoside antibiotic treatment is associated with a significant incidence of nephrotoxicity [50], mirroring the effects seen in animal models. Similarly, isoniazid treatment has a known risk of hepatotoxicity [51]. Thus, one approach to safety biomarker qualification in the clinic is to monitor novel biomarker levels longitudinally over the course of such a treatment and compare these with the current gold standard commonly used clinical biomarkers and outcomes [52]. Of course, one can complement these studies with those examining biomarker levels in patients with organ injury of a disease etiology [53].

The number of known agents appropriate for testing the sensitivity of safety biomarkers for certain target organ toxicities may be limited. It is generally expected that the number of studies conducted to evaluate sensitivity should reasonably represent a high percentage of the known but limited diverse mechanisms available for testing in animal and human studies. If the mechanisms are varied for each test agent and sensitivity performance remains high, the biomarker will probably find strong use potential. Specificity tests then become very important considerations in a qualification strategy. The number of test compounds that could be deployed to assess the false-positive test rate is far more expansive than the number of known compounds for testing sensitivity. Specificity testing therefore becomes a highly individualized dimension to a biomarker biological qualification strategy for biomarkers that pass tests of sensitivity. The two critical questions to address are whether (1) there are alternative tissue sources to account for alterations in test safety biomarker levels, and (2) whether there are benign, nontoxicologic mechanisms to account for alterations in test biomarker levels. To evaluate specificity, therefore, a prioritized experimental approach should be taken using logical reasoning to avoid the testing burdens of an endless number of possible studies.

The ultimate goal of these studies is the qualification to the level of what Wagner et al. define as a *characterization biomarker* [49], a biomarker "associated with adequate preclinical sensitivity and specificity and reproducibly linked clinical outcomes in more than one prospective clinical study in humans." Such a level of qualification then supports the regulatory use of these biomarkers for safety monitoring in early clinical studies with a new pharmaceutical candidate. A strong case has been made that for safety biomarkers for regulatory decision-making purposes, where the rigor of

supporting evidence would be expected to be high, that only fully qualified or "characterization" biomarkers are appropriate, and that there is really no regulatory decision-making role for exploratory and "emerging" or "probable valid" biomarkers. In this regard, measurements of such unqualified biomarkers in animal studies used to support the safe conduct of clinical trials would not be expected to contribute unambiguously and sufficiently to study interpretation, should not require submission to regulatory authorities, and therefore the exploration of their utility in such highly regulated studies should be encouraged [54] in order to accelerate the pace of biomarker evaluations and understanding.

## COLLABORATION IN BIOMARKER QUALIFICATION

Clearly, the number of preclinical and clinical studies and resources required to qualify a biomarker as a characterization biomarker appropriate for regulatory decision making are significant. Thus, it is no surprise that in their Critical Path Opportunities List the FDA called for "collaborations among sponsors to share what is known about existing safety assays" [55]. Collaborations of this type have indeed played key roles in addressing technological problems common to a competitive industry. Thus, Sematech, a consortium formed in 1987 and made up of 14 leading U.S. semiconductor producers, addressed common issues in semiconductor manufacture and increased research and development (R&D) efficiency by avoiding duplicative research [56]. Sematech demonstrates that consortia can provide an opportunity for industry scientists to pool their expertise and experience to confront mutual questions collectively. The International Life Sciences Institute has for several years served as a forum for collaborative efforts between industry and academia [57], and for the past six years a Biomarkers Technical Committee has been pursuing assay development and evaluation of biomarkers of nephrotoxicity and cardiotoxicity [58]. Recently, the Critical Path Institute was incorporated as a "neutral, third party" to serve as a consortium organizer [59] and interface between industry members and the FDA. One of its first efforts was the Predictive Safety Testing Consortium (PSTC) [60,61], with a specific focus on qualification of biomarkers for regulatory use. The PSTC legal agreement addresses issues, such as intellectual property, antitrust concerns, and confidentiality, and thus assures open collaboration in a manner consistent with applicable legal requirements. The PSTC solicited representatives from the FDA and the European Medicines Agency (EMEA) to serve as advisors. As experts in various areas of toxicity, these advisors bring not only their expertise but also the experience of how problems of a given target-organ toxicity are confronted and could be addressed in a regulatory setting. Thus, the development of qualification data is targeted with a keen eye toward what will ultimately support safety decisions in the regulated drug development and regulatory review process.

## REFERENCES


1. Wax PM (1995). Elixirs, diluents, and the passage of the 1938 Federal Food, Drug and Cosmetic Act. *Ann Intern Med*, 122:456–461.
2. Miller SA (1993). Science, law and society: the pursuit of food safety. *J Nutr*, 123:279–284.
3. Stirling D, Junod S (2002). Arnold J. Lehman. *Toxicol Sci*, 70:159–160.
4. Eaton D, Klaassen CD (2001). Principles of toxicology. In Klaassen CD (ed.), *Casarett and Doull's Toxicology*, 6th ed. McGraw-Hill, New York, pp. 11–34.
5. Weingand K, Brown G, Hall R, et al. (1996). Harmonization of animal clinical pathology testing in toxicity and safety studies. The Joint Scientific Committee for International Harmonization of Clinical Pathology Testing. *Fundam Appl Toxicol*, 29:198–201.
6. Bregman CL, Adler RR, Morton DG, Regan KS, Yano BL (2003). Recommended tissue list for histopathologic examination in repeat-dose toxicity and carcinogenicity studies: a proposal of the Society of Toxicologic Pathology (STP). *Toxicol Pathol*, 31:252–253.
7. FDA (1997). International Conference on Harmonisation; Guidance on General Considerations for Clinical Trials. *Federal Register*, p. 66113.
8. FDA (2008). International Conference on Harmonisation; Draft Guidance on M3(R2) Nonclinical Safety Studies for the Conduct of Human Clinical Trials and Marketing Authorization for Pharmaceuticals. *Federal Register*, pp. 51491–51492.
9. FDA (1997). International Conference on Harmonisation; Good Clinical Practice: Consolidated Guideline. *Federal Register*, pp. 25691–25709.
10. Schnellmann RG (2001). Toxic responses of the kidney. In Klaassen CD (ed.), *Casarett and Doull's Toxicology*, 6th ed. McGraw-Hill, New York, pp. 491–514.
11. Hoste EA, Clermont G, Kersten A, et al. (2006). RIFLE criteria for acute kidney injury are associated with hospital mortality in critically ill patients: a cohort analysis. *Crit Care*, 10:R73.
12. Chertow GM, Burdick E, Honour M, Bonventre JV, Bates DW (2005). Acute kidney injury, mortality, length of stay, and costs in hospitalized patients. *J Am Soc Nephrol*, 16:3365–3370.
13. Vaidya VS, Ferguson MA, Bonventre JV (2008). Biomarkers of acute kidney injury. *Annu Rev Pharmacol Toxicol*, 48:463–493.
14. Trof RJ, Di Maggio F, Leemreis J, Groeneveld AB (2006). Biomarkers of acute renal injury and renal failure. *Shock*, 26:245–253.
15. Molitoris BA, Melnikov VY, Okusa MD, Himmelfarb J (2008). Technology Insight: biomarker development in acute kidney injury—what can we anticipate? *Nat Clin Pract Nephrol*, 4:154–165.
16. Ferguson MA, Vaidya VS, Bonventre JV (2008). Biomarkers of nephrotoxic acute kidney injury. *Toxicology*, 245:182–193.
17. Devarajan P (2007). Emerging biomarkers of acute kidney injury. *Contrib Nephrol*, 156:203–212.
18. Bagshaw SM, Langenberg C, Haase M, Wan L, May CN, Bellomo R (2007). Urinary biomarkers in septic acute kidney injury. *Intensive Care Med*, 33:1285–1296.

19. Nguyen MT, Devarajan P (2007). Biomarkers for the early detection of acute kidney injury. *Pediatr Nephrol*, 23:2151–2157.
20. Dieterle F, Marrer E, Suzuki E, Grenet O, Cordier A, Vonderscher J (2008). Monitoring kidney safety in drug development: emerging technologies and their implications. *Curr Opin Drug Discov Dev*, 11:60–71.
21. Zimmerman HJ (1999). *Hepatotoxicity: The Adverse Effects of Drugs and Other Chemicals on the Liver*, 2nd ed. Lippincott Williams & Wilkins, Philadelphia.
22. Maddrey WC (2005). Drug-induced hepatotoxicity: 2005. *J Clin Gastroenterol*, 39: S83–S89.
23. Arundel C, Lewis JH (2007). Drug-induced liver disease in 2006. *Curr Opin Gastroenterol*, 23:244–254.
24. Watkins PB, Seeff LB (2006). Drug-induced liver injury: summary of a single topic clinical research conference. *Hepatology*, 43:618–631.
25. Bleibel W, Kim S, D'Silva K, Lemmer ER (2007). Drug-induced liver injury: review article. *Dig Dis Sci*, 52:2463–2471.
26. Kim WR, Flamm SL, Di Bisceglie AM, Bodenheimer HC (2008). Serum activity of alanine aminotransferase (ALT) as an indicator of health and disease. *Hepatology*, 47:1363–1370.
27. Reichling JJ, Kaplan MM (1988). Clinical use of serum enzymes in liver disease. *Dig Dis Sci*, 33:1601–1614.
28. Ozer J, Ratner M, Shaw M, Bailey W, Schomaker S (2008). The current state of serum biomarkers of hepatotoxicity. *Toxicology*, 245:194–205.
29. Lock EA, Bonventre JV (2008). Biomarkers in translation; past, present and future. *Toxicology*, 245:163–166.
30. Amacher DE (2002). A toxicologist's guide to biomarkers of hepatic response. *Hum Exp Toxicol*, 21:253–262.
31. Pettersson J, Hindorf U, Persson P, et al. (2008). Muscular exercise can cause highly pathological liver function tests in healthy men. *Br J Clin Pharmacol*, 65:253–259.
32. Giboney PT (2005). Mildly elevated liver transaminase levels in the asymptomatic patient. *Am Fam Physician*, 71:1105–1110.
33. Gaskill CL, Miller LM, Mattoon JS, et al. (2005). Liver histopathology and liver and serum alanine aminotransferase and alkaline phosphatase activities in epileptic dogs receiving Phenobarbital. *Vet Pathol*, 42:147–160.
34. Shapiro MA, Lewis JH (2007). Causality assessment of drug-induced hepatotoxicity: promises and pitfalls. *Clin Liver Dis*, 11:477–505.
35. Andrade RJ, Lucena MI, Fernandez MC, et al. (2005). Drug-induced liver injury: an analysis of 461 incidences submitted to the Spanish registry over a 10-year period. *Gastroenterology*, 129:512–521.
36. Hunt CM, Papay JI, Edwards RI, et al. (2007). Monitoring liver safety in drug development: the GSK experience. *Regul Toxicol Pharmacol*, 49:90–100.
37. Ramos KS, Melchert RB, Chacon E, Acosta D Jr (2001). Toxic responses of the heart and vascular systems. In Klaassen CD (ed.), *Casarett and Doull's Toxicology*, 6th ed. McGraw-Hill, New York, pp. 597–651.

38. Kerns W, Schwartz L, Blanchard K, et al. (2005). Drug-induced vascular injury: a quest for biomarkers. *Toxicol Appl Pharmacol*, 203:62–87.
39. Mesfin GM, Higgins MJ, Robinson FG, Zhong WZ (1996). Relationship between serum concentrations, hemodynamic effects, and cardiovascular lesions in dogs treated with minoxidil. *Toxicol Appl Pharmacol*, 140:337–344.
40. Blake GJ, Ridker PM (2001). Novel clinical markers of vascular wall inflammation. *Circ Res*, 89:763–771.
41. Louden C, Brott D, Katein A, et al. (2006). Biomarkers and mechanisms of drug-induced vascular injury in non-rodents. *Toxicol Pathol*, 34:19–26.
42. Tiwari A, Bansal V, Chugh A, Mookhtiar K (2006). Statins and myotoxicity: a therapeutic limitation. *Expert Opin Drug Saf*, 5:651–666.
43. Owczarek J, Jasinska M, Orszulak-Michalak D (2005). Drug-induced myopathies: an overview of the possible mechanisms. *Pharmacol Rep*, 57:23–34.
44. Merriam-Webster Online Dictionary (2008).
45. FDA (2001). Guidance for Industry on Bioanalytical Method Validation. *Federal Register*, pp. 28526–28527.
46. Wagner JA (2008). Strategic approach to fit-for-purpose biomarkers in drug development. *Annu Rev Pharmacol Toxicol*, 48:631–651.
47. Lee JW, Devanarayan V, Barrett YC, et al. (2006). Fit-for-purpose method development and validation for successful biomarker measurement. *Pharm Res*, 23:312–328.
48. Stokes WS, Schechtman LM, Rispin A, et al. (2006). The use of test method performance standards to streamline the validation process. *Altex*, 23 (Suppl):342–345.
49. Wagner JA, Williams SA, Webster CJ (2007). Biomarkers and surrogate end points for fit-for-purpose development and regulatory evaluation of new drugs. *Clin Pharmacol Ther*, 81:104–107.
50. Wiland P, Szechcinski J (2003). Proximal tubule damage in patients treated with gentamicin or amikacin. *Pol J Pharmacol*, 55:631–637.
51. Tostmann A, Boeree MJ, Aarnoutse RE, de Lange WC, van der Ven AJ, Dekhuijzen R (2008). Antituberculosis drug-induced hepatotoxicity: concise up-to-date review. *J Gastroenterol Hepatol*, 23:192–202.
52. Mishra J, Dent C, Tarabishi R, et al. (2005). Neutrophil gelatinase-associated lipocalin (NGAL) as a biomarker for acute renal injury after cardiac surgery. *Lancet*, 365:1231–1238.
53. Han WK, Bailly V, Abichandani R, Thadhani R, Bonventre JV (2002). Kidney injury molecule-1 (KIM-1): a novel biomarker for human renal proximal tubule injury. *Kidney Int*, 62:237–244.
54. Sistare FD, DeGeorge JJ (2008). Applications of toxicogenomics to nonclinical drug development: regulatory science considerations. *Methods Mol Biol*, 460:239–261.
55. FDA (2006). *Critical Path Opportunities Report and List*, p 28.
56. Irwin DA, Klenow PJ (1996). Sematech: purpose and performance. *Proc Natl Acad Sci USA*, 93:12739–12742.
57. ILSI (2007). *ILSI About ILSI*.

58. ILSI and HESI (2007). *ILSI: Development and Application of Biomarkers of Toxicity*.
59. Woosley RL, Cossman J (2007). Drug development and the FDA's Critical Path Initiative. *Clin Pharmacol Ther*, 81:129–133.
60. Marrer E, Dieterle F (2007). Promises of biomarkers in drug development: a reality check. *Chem Biol Drug Des*, 69:381–394.
61. Mattes WB (2008). Public consortium efforts in toxicogenomics. *Methods Mol Biol*, 460:221–238.

# DEVELOPMENT OF SERUM CALCIUM AND PHOSPHORUS AS CLINICAL BIOMARKERS FOR DRUG-INDUCED SYSTEMIC MINERALIZATION: CASE STUDY WITH A MEK INHIBITOR

Alan P. Brown, Ph.D., DABT
*Pfizer Global Research and Development, Ann Arbor, Michigan*

## INTRODUCTION

The mitogen-activated protein kinase (MAPK) signal transduction pathways control key cellular processes such as growth, differentiation, and proliferation, and provide a means for transmission of signals from the cell surface to the nucleus. As a part of the RAS–RAF–MEK–MAPK pathway, MEK (MAP kinase kinase) phosphorylates the MAPK proteins ERK1 and ERK2 (extracellular signal-regulated kinases) as a means for intracellular signaling (Sebolt-Leopold, 2000). Although MEK has not been identified as having oncogenic properties, this kinase serves as a focal point in the signal transduction pathway of known oncogenes (e.g., RAS and RAF) (Mansour et al., 1994). MEK exists downstream of various receptor tyrosine kinases (such as the epidermal growth factor receptor) which have been demonstrated to be important in neoplasia (Jost et al., 2001). RAS activation occurs first, followed by recruitment of RAF (A-RAF, B-RAF, or RAF-1) proteins to the cell membrane through binding to RAS, with subsequent activation of RAF. RAF phosphorylates MEK1 and MEK2 on multiple serine residues in the activation

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

process. MEK1 and MEK2 phosphorylate tyrosine or threonine residues on ERK proteins in the signal transduction process, with phosphorylated ERK activating various transcription factors (Friday and Adjei, 2008). Aberrant activation of this pathway has been observed in a diverse group of solid tumors, along with leukemia, and is believed to play a key role in tumorigenesis (Hoshino et al., 1999; Milella et al., 2001).

Based on a significant amount of preclinical data, development of small-molecule inhibitors of MEK appears to be a rational approach for treatment of various malignancies (Sebolt-Leopold et al., 1999; Dent and Grant, 2001). The first MEK inhibitor to enter clinical trials was CI-1040 (also known as PD0184352), which was intended for oral administration. However, the level of antitumor activity in a multicenter phase II study in patients with various solid tumors was not sufficient to warrant further development of this drug (Rinehart et al., 2004; Wang et al., 2007). CI-1040 exhibited low oral bioavailability and high metabolism, which were primary factors resulting in insufficient plasma drug levels for antitumor activity.

PD0325901 [Figure 1; chemical name of *N*-((*R*)-2,3-dihydroxypropoxy)-3,4-difluoro-2-(2-fluoro-4-iodo-phenylamino)benzamide] is a highly potent and specific non-ATP competitive inhibitor of MEK ($K_i$ of 1 nM against activated MEK1 and MEK2 in vitro), and demonstrated anticancer activity against a broad spectrum of human tumors in murine models (Sebolt-Leopold et al., 2004). Preclinical studies indicate that PD0325901 has the potential to impair growth of human tumors that rely on the MEK/MAPK pathway for growth and survival. PD0325901 inhibits the phosphorylation of MAPK proteins (ERK1 and ERK2) as a biochemical mechanism of action, and assays were developed to evaluate inhibition of protein phosphorylation in normal and neoplastic tissues (Brown et al., 2007). This compound has greatly improved pharmacologic and pharmaceutical properties compared with CI-1040 (i.e.,

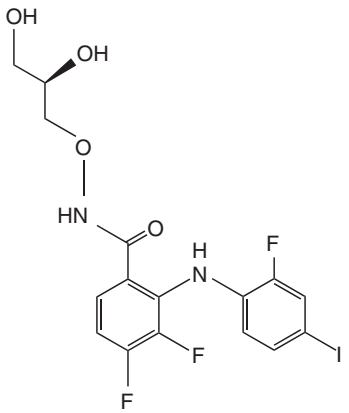


**Figure 1** Chemical structure of PD0325901.

greater potency for MEK inhibition, higher bioavailability, and increased metabolic stability) and has significant promise for determining the therapeutic potential for treating cancer with an orally active MEK inhibitor (Rinehart et al., 2004; Sebolt-Leopold et al., 2004). PD0325901 was selected for development as a clinical candidate due to its superior preclinical profile compared to CI-1040 (Wang et al., 2007). Toxicology studies were subsequently initiated to support the conduct of a phase I or II clinical trial in cancer patients with various solid tumors (advanced breast cancer, colon cancer, melanoma, non-small cell lung cancer) utilizing oral administration of the drug.

## TOXICOLOGY STUDIES

The nonclinical safety of PD0325901 was evaluated in Sprague–Dawley rats given single oral or intravenous (IV) doses, in beagle dogs given oral and IV escalating doses, and in cynomolgus monkeys given escalating oral doses to assess acute toxicity and assist in dose selection for subsequent studies. The potential effects of PD0325901 on central nervous system, cardiovascular, and pulmonary function were evaluated in single-dose safety pharmacology studies. Two-week dose-range finder (nonpivotal) oral toxicity studies were conducted in rats, dogs, and monkeys to assist in dose and species selection for the pivotal one-month oral toxicology studies. In addition, an investigative oral toxicity study was conducted in female Balb/c mice. The dog was selected as the nonrodent species for the pivotal toxicology study because of the following data. Metabolites of PD0325901 identified in human liver microsomal incubations were also present following incubation with dog liver microsomes, plasma protein binding of PD0325901 is similar in dogs and humans (>99%), and oral bioavailability in dogs is high (>90%). Finally, injury to the mucosa of the gastrointestinal tract occurred at lower doses and exposures in dogs than in monkeys (based on dose-range-finding studies), indicating the dog as the more sensitive nonrodent species. Pivotal one-month oral toxicity studies, including one-month reversal phases, were conducted in beagle dogs and Sprague–Dawley rats to support submission of an investigational new drug (IND) application to the U.S. Food and Drug Administration. A list of toxicology studies of PD0325901 conducted prior to initiation of human testing is presented in Table 1.

Upon completion of the first two-week dose-range-finding study in rats, a significant and unique toxicity was observed that involved mineralization of vasculature (Figure 2) and various soft tissues (i.e., ectopic or systemic mineralization) as determined by routine light-microscopic evaluation. In a follow-up study in rats, dysregulation of serum calcium and phosphorus homeostasis, and systemic mineralization occurred in a time- and dose-dependent manner. This toxicity was not observed in dogs or monkeys, despite systemic exposures to PD0325901 more than 10-fold higher than those associated with mineralization in rats and pharmacologic inhibition of

**TABLE 1 Summary of Toxicology Studies Conducted with PD0325901**[a]

| |
|---|
| Acute and escalating dose |
| Single dose in rats |
| Single dose in rats, IV[b] |
| Escalating dose in dogs |
| Escalating dose in dogs, IV |
| Escalating dose in monkeys |
| Safety pharmacology |
| Neurofunctional evaluation in rats |
| Neurofunctional evaluation in rats, IV |
| Cardiovascular effects in monkeys |
| Pulmonary effects in rats |
| Purkinje fiber assay |
| HERG assay |
| Nonpivotal repeated-dose studies |
| 2-Week dose-range finder in rats |
| Exploratory 2-week study in rats |
| 2-Week dose-range finder in dogs |
| 2-Week dose-range finder in monkeys |
| Pivotal repeated-dose studies |
| One month in rats (plus one-month reversal phase) |
| One month in dogs (plus one-month reversal phase) |
| Pivotal genetic toxicity studies |
| Bacterial mutagenicity |
| Structural chromosome aberration |
| In vivo micronucleus in rats |
| Special toxicity studies |
| Pharmacodynamic and toxicokinetic in rats, oral and IV |
| Time course and biomarker development in rats |
| Serum chemistry reversibility study in rats |
| Investigative study in mice |
| Enantiomer (R and S) study in rats |
| PD0325901 in combination with pamidronate or Renagel in rats |

[a]All animal studies were conducted by oral gavage unless otherwise indicated.
[b]IV, intravenous (bolus).

phosphorylated MAPK in canine or monkey tissue (demonstrating biochemical activity of PD0325901 at the target protein, i.e., MEK). Various investigative studies were conducted to examine the time course and potential mechanism of systemic mineralization in rats and to identify biomarkers that could be used to monitor for this effect in clinical trials. Next we describe the key studies conducted to investigate this toxicity, the results obtained, and how the nonclinical data were utilized to evaluate the safety risk of the compound, select a safe starting dose for a phase I trial, and provide measures to ensure patient safety during clinical evaluation of PD0325901 in cancer patients.

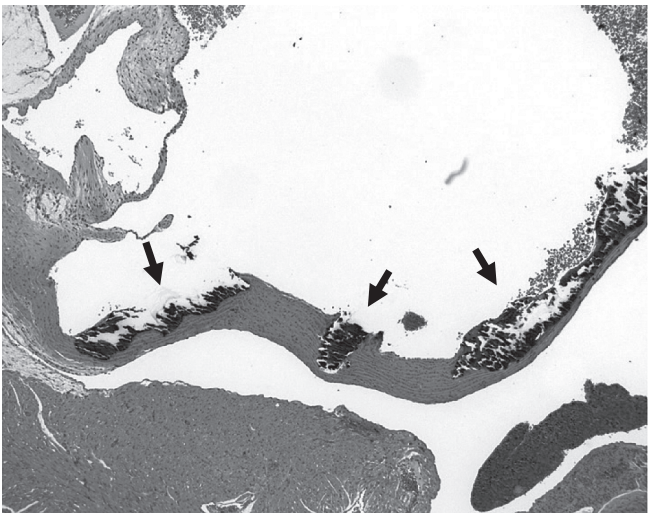

**Figure 2** Mineralization of the aorta in a male rat administered PD0325901 at 3 mg/kg in a dose-range-finding study. Arrows indicate mineral in the aorta wall. Hematoxylin and eosin–stained tissue section. (*See insert for color reproduction of the figure.*)

At the beginning of the toxicology program for PD0325901, a two-week oral-dose-range-finding study was conducted in male and female rats in which daily doses of 3, 10, and 30 mg/kg (18, 60, and 180 mg/m$^2$, respectively) were administered. Mortality occurred in males at ≥3 mg/kg and females at ≥10 mg/kg, with toxicity occurring to a greater extent in males at all dose levels. Increased serum levels of phosphorus (13 to 69%), and decreased serum total protein (12 to 33%) and albumin (28 to 58%) were seen at all doses. Light-microscopic evaluation of formalin-fixed and hematoxylin- and eosin-stained tissues was performed. Mineralization occurred in the aorta (Figure 2) and coronary, renal, mesenteric, gastric, and pulmonary vasculature of males at ≥3 mg/kg and in females at ≥10 mg/kg. Parenchymal mineralization with associated degeneration occurred in the gastric mucosa and muscularis, intestines (muscularis, mucosa, submucosa), lung, liver, renal cortical tubules, and/or myocardium at the same doses. Use of the Von Kossa histology stain indicated the presence of calcium in the mineralized lesions. Vascular/parenchymal mineralization and degeneration were generally dose related in incidence and severity. PD0325901 produced increased thickness (hypertrophy) of the femoral growth plate (physis) in both sexes at all doses, and degeneration and necrosis of the femoral metaphysis in males at ≥3 mg/kg and females at 30 mg/kg. In addition, skin ulceration, hepatocellular necrosis, decreased crypt goblet cells, reduced hematopoetic elements, and ulcers of cecum and duodenum were observed. Systemic mineralization of the vasculature and soft tissues was

the most toxicologically significant finding of this study. At this time, it was not known whether the hyperphosphatemia was due to decreased renal clearance (York and Evans, 1996) and/or related to the mineralization. However, hyperphosphatemia and elevated serum calcium–phosphorus (Ca × P) product can result in vascular and/or soft tissue mineralization (Spaulding and Walser, 1970; Block, 2000; Giachelli et al., 2001). In addition, morphologic findings similar to those seen in this study are observed in various animal species (e.g., dogs, horses, pigs, rats) with vitamin D toxicosis and altered calcium homeostasis (Grant et al., 1963; Spangler et al., 1979; Harrington and Page, 1983; Long, 1984). Tissue mineralization is observed in the aorta, various arteries, myocardium, gastric mucosa, and renal tubules, along with other soft tissues in these animals.

An exploratory two-week oral toxicity study was next conducted in male and female rats to further investigate the toxicities observed in the initial two-week dose-range finder. The objectives of this study were to identify a minimal or no-adverse-effect level and to provide toxicity, toxicokinetic, and pharmacodynamic data to aid in dose selection for future studies. In addition, an attempt was made to assess whether alterations in phosphorus and calcium homeostasis occur and whether changes can be monitored as potential biomarkers of toxicity. Doses tested in this study were 0.3, 1, or 3 mg/kg (1.8, 6, or 18 mg/m$^2$, respectively) and animals were dosed for up to 14 days. Cohorts of animals (5/sex/group) were necropsied on days 4 and 15, and hematology, serum biochemistry, plasma intact parathyroid hormone (PTH), and urinary parameters were evaluated. Urinalysis included measurement of calcium, phosphorus, and creatinine levels. Select tissues were examined microscopically, and samples of liver and lung were evaluated for total and phosphorylated MAPK (pMAPK) levels by Western blot analysis to evaluate for pharmacologic activity of PD0325901 (method described in Brown et al., 2007). Satellite animals were included for plasma drug-level analyses on day 8. In this study, systemic mineralization occurred at ≥0.3 mg/kg in a dose-dependent fashion, was first observed on day 4, and was more severe in males. By day 15, mineralization was generally more pronounced and widespread. Skeletal changes included hypertrophy of the physeal zone in males at ≥1 mg/kg and at 3 mg/kg in females, and necrosis of bony trabeculae and marrow elements with fibroplasia, fibro-osseous proliferation, and/or localized hypocellularity at ≥1 mg/kg in males and 3 mg/kg in females. The minimal plasma PD0325901 AUC(0–24) values associated with toxicity were 121 to 399 ng · h/mL, which were well below exposure levels associated with antitumor efficacy in murine models (AUC of 1180 to 1880 ng · h/mL). Pharmacologic inhibition of tissue pMAPK occurred at ≥1 mg/kg and was not observed in the absence of toxicity. The gastric fundic mucosa appeared to be the most sensitive tissue for evaluating systemic mineralization, which probably resulted from alterations in serum calcium and phosphorus homeostasis. This was based on the following observations. On day 4, serum phosphorus levels were increased 12

**TABLE 2 Mean Clinical Chemistry Changes in Male Rats Administered PD0325901 for Up to 2 Weeks**

| | | | PD0325901 | | |
|---|---|---|---|---|---|
| | Day | Control | 0.3 mg/kg | 1 mg/kg | 3 mg/kg |
| Serum phosphorus (mg/dL) | 4 | 12.90 | 13.08 | 14.48 | 16.24[a] |
| | 15 | 11.30 | 11.56 | 12.88 | 13.62[a] |
| Serum calcium (mg/dL) | 4 | 10.58 | 10.36 | 10.10 | 10.16 |
| | 15 | 10.38 | 10.36 | 10.52 | 10.36 |
| Serum albumin (g/dL) | 4 | 2.74 | 2.56 | 2.10[a] | 2.04[a] |
| | 15 | 2.56 | 2.54 | 2.36[a] | 1.98[a] |
| Plasma PTH (pg/mL)[b] | 4 | 492 | 297 | 114[a] | 155[a] |
| | 15 | 1099 | 268 | 457 | 115[a] |

[a] $p, < 0.01$ vs. control; $n = 5$/group.
[b] Intact parathyroid hormone.

to 26%, and albumin was decreased 17 to 26% at ≥1 mg/kg (Table 2, male data only). In addition, PTH levels were decreased in a dose-dependent fashion (60 to 77%) at ≥1 mg/kg. On day 15, phosphorus levels were increased 21% in males at 3 mg/kg, and albumin was decreased 8 to 32% at ≥0.3 mg/kg. PTH levels were decreased 77 to 89% at 3 mg/kg. Changes in urinary excretion of calcium and phosphorus were observed in both sexes at ≥1 mg/kg and included increased excretion of phosphorus on day 15. Although increases in excretion of calcium were observed on day 4 in females, males exhibited decreases in urinary calcium. In this study, PD0325901 administration resulted in significantly decreased levels of serum albumin without changes in serum (total) calcium levels (Payne et al., 1979; Meuten et al., 1982; Rosol and Capen, 1997). This indicates that free, non-protein-bound calcium levels were increased. Hyperphosphatemia and hypercalcemia result in an increased Ca × P product, which is associated with induction of vascular mineralization (Block, 2000; Giachelli et al., 2001). The changes observed in urinary excretion of calcium and phosphorus probably reflected the alterations in serum levels.

After completion of the two studies in rats described above it was concluded that PD0325901 produces significant multiorgan toxicities in rats with no margin between plasma drug levels associated with antitumor efficacy, pharmacologic inhibition of pMAPK (as an index of MEK inhibition), and toxicity in rats. Systemic mineralization was considered the preclinical toxicity of greatest concern, due to the severity of the changes observed and expectation of irreversibility, and the data suggested that it was related to a dysregulation in serum phosphorus and calcium homeostasis. Furthermore, skeletal lesions were seen in the rat studies that were similar to those reported with vitamin D toxicity and may be related to the calcium–phosphorus dysregulation. In concurrent toxicology studies in dogs and monkeys, neither systemic

mineralization nor skeletal changes were observed, despite higher plasma drug exposures, lethal doses, or pharmacologic inhibition of MEK. Therefore, the following questions were posed regarding PD0325901-induced systemic mineralization: (1) What is a potential mechanism? (2) Is this toxicity relevant to humans or rat-specific? and (3) Can this toxicity be monitored clinically?

The ability of an anticancer agent that modulates various signal transduction pathways to produce dysregulation in serum calcium homeostasis is not unprecedented. 8-Chloro-cAMP is an experimental compound that has been shown to modulate various protein kinase signal transduction pathways involved in neoplasia. In preclinical models, this compound produced growth inhibition and increased differentiation in cancer cells (Ally et al., 1989). In a clinical trial, 8-chloro-cAMP was administered to patients with advanced cancer via intravenous infusion and resulted in dose-limiting toxicity of reversible hypercalcemia, as serum calcium was increased by up to approximately 40% (Saunders et al., 1997). This drug produced a parathyroid hormone-like effect in these patients, resulting in increased synthesis of 1,25-dihydroxyvitamin D (up to 14 times baseline value) as a mechanism for the hypercalcemia. Intravenous administration of 8-chloro-cAMP to beagle dogs also resulted in hypercalcemia (serum calcium increased 37 to 46%), indicating similar actions across species (Brown et al., 2000). Experience with this compound was important with respect to designing investigative studies with PD0325901 in which the hormonal control of serum calcium and phosphorus were evaluated.

An investigative study was designed in rats to examine the time course for tissue mineralization in target organs and to determine whether clinical pathology changes occur prior to, or concurrent with, lesion development (Brown et al., 2005a). These clinical pathology parameters may therefore serve as biomarkers for systemic mineralization. Male rats (15/group) were used due to their increased sensitivity for this toxicity compared with females. Oral doses tested were 1, 3, or 10 mg/kg (6, 18, or 60 mg/m$^2$, respectively). Five animals per group were necropsied on days 2, 3, or 4 following 1, 2, or 3 days of treatment, respectively. Clinical laboratory tests were conducted at necropsy that included serum osteocalcin, urinalysis, and plasma intact PTH, calcitonin, and 1,25-dihydroxyvitamin D. Lung samples were evaluated for inhibition of pMAPK, and microscopic evaluations of the aorta, distal femur with proximal tibia, heart, and stomach were conducted for all animals.

Administration of PD0325901 resulted in inhibition of pMAPK in lung at all doses, demonstrating pharmacologic activity of the drug. On day 2, mineralization of gastric fundic mucosa and multifocal areas of necrosis of the ossifying zone of the physis were present only at 10 mg/kg. Necrosis of the metaphysis was present at ≥3 mg/kg. Serum phosphorus levels increased 33 to 43% and 1,25-dihydroxyvitamin D increased two- to sevenfold at all doses (Table 3). Osteocalcin increased 14 to 18%, and serum albumin decreased 8 to 14% at ≥3 mg/kg (Table 4). Osteocalcin is a major noncollagenous protein of bone matrix and synthesized by osteoblasts (Fu and Muller, 1999). Changes in serum osteocalcin can reflect alterations in bone turnover (resorption/

**TABLE 3 Mean Serum Phosphorus and Plasma 1,25-Dihydroxyvitamin D in Male Rats Administered PD0325901 for Up to 3 Days of Dosing**

| | | | PD0325901 | | |
|---|---|---|---|---|---|
| | Day | Control | 1 mg/kg | 3 mg/kg | 10 mg/kg |
| Serum phosphorus (mg/dL) | 2 | 12.06 | 16.10* | 17.22* | 16.84* M[a] |
| | 3 | 11.48 | 12.96* | 15.62* M | 19.02* M |
| | 4 | 11.34 | 13.18* M | 15.40*M | 21.70* M |
| 1,25-Dihydroxyvitamin D (pg/mL) | 2 | 309 | 856* | 1328* | 2360* M |
| | 3 | 257 | 396 | 776* M | 1390* M |
| | 4 | 191 | 236 M | 604* M | 1190* M |

[a]M, systemic mineralization observed.
*, $p < 0.01$ vs. control; $n = 5$/group.

**TABLE 4 Mean Serum Calcium and Albumin in Male Rats Administered PD0325901 for Up to 3 Days of Dosing**

| | | | PD0325901 | | |
|---|---|---|---|---|---|
| | Day | Control | 1 mg/kg | 3 mg/kg | 10 mg/kg |
| Serum calcium (mg/dL) | 2 | 10.42 | 11.04 | 11.00 | 10.66 M[a] |
| | 3 | 9.60 | 10.58 | 10.64 M | 10.58 M |
| | 4 | 10.44 | 10.44 M | 10.58 M | 7.24** M |
| Serum albumin (g/dL) | 2 | 3.08 | 2.92 | 2.82* | 2.66** M |
| | 3 | 2.88 | 2.68 | 2.62 M | 2.34** M |
| | 4 | 2.90 | 2.34** M | 2.34** M | 1.98** M |

[a]M, systemic mineralization observed.
*, $p < 0.05$ vs. control.
**, $p < 0.01$ vs. control; $n = 5$/group.

formation). Serum osteocalcin appears to reflect the excess of synthesized protein not incorporated into bone matrix, or released protein during bone resorption (Ferreira and Drueke, 2000). The increases in osteocalcin seen in this study may have been reflective of bone necrosis.

On day 3, mineralization of gastric fundic mucosa, gastric and cardiac arteries, aorta, and heart were present in all rats at 10 mg/kg. Myocardial necrosis was also seen at 10 mg/kg. Mineralization of gastric fundic mucosa was present in all rats at 3 mg/kg, and focal, minimal myocyte necrosis was present in one rat at 3 mg/kg. Thickening of the physeal zone of hypertrophying cartilage, and necrosis within the physeal zone of ossification and in the metaphyseal region in femur and tibia were seen in all animals at 10 mg/kg. Necrosis within the metaphyseal region was also present at 3 mg/kg. Serum phosphorus increased 13 to 66% at all doses and 1,25-dihydroxyvitamin D increased two- to fourfold at ≥3 mg/kg. Osteocalcin increased 12 to 28% at ≥3 mg/kg and serum albumin was decreased (7 to 19%) at all doses. Urine calcium increased

fivefold at 10 mg/kg, resulting in a fivefold increase in urine calcium/creatinine ratio. This increase may have represented an attempt to achieve mineral homeostasis in response to the hypercalcemia. In addition, hypercalciuria can occur with vitamin D intoxication (Knutson et al., 1997).

On day 4, mineralization of gastric fundic mucosa, gastric muscularis, gastric and cardiac arteries, aorta, and heart were present in the majority of animals at ≥3 mg/kg. Myocardial necrosis with accompanying neutrophilic inflammation was also seen in all rats at 10 mg/kg and in one animal at 3 mg/kg. Mineralization of gastric fundic mucosa was present at 1 mg/kg. Thickening of the physeal zone of hypertrophying cartilage, and necrosis within the physeal zone of ossification and/or in the metaphyseal region in femur and tibia, were present at ≥3 mg/kg. At 1 mg/kg, thickening of the physeal zone of hypertrophying cartilage and metaphyseal necrosis were observed. Serum phosphorus increased 16 to 91% at all doses and 1,25-dihydroxyvitamin D increased two- to fivefold at ≥3 mg/kg. Osteocalcin increased 14 to 24% at ≥3 mg/kg, and serum albumin decreased 19 to 32% at all doses. At 10 mg/kg, serum calcium was decreased 31% (possibly resulting from the hypercalciuria on day 3) and calcitonin was decreased by 71%. Calcitonin is secreted by the thyroid gland and acts to lower serum calcium levels by inhibiting bone resorption (Rosol and Capen, 1997). The decrease in calcitonin may have resulted from feedback inhibition due to low serum calcium levels at 10 mg/kg on day 4. Urine creatinine, calcium, and phosphorus were increased at 10 mg/kg. This resulted in decreases of 41% and 21% in the calcium/creatinine and phosphorus/creatinine ratios, respectively.

This four-day investigative study in rats resulted in several very important conclusions which were critical for supporting continued development of PD0325901. In the study, PD0325901 at ≥1 mg/kg resulted in systemic mineralization and skeletal changes in a dose- and time-dependent fashion. These changes were seen after a single dose at 10 mg/kg and after 3 doses at 1 mg/kg. Elevations in serum phosphorus and plasma 1,25-dihydroxyvitamin D occurred prior to tissue mineralization. Although serum albumin was decreased throughout the study, calcium remained unchanged, consistent with an increase in non-protein-bound calcium. This study set the stage for the proposal of using serum phosphorus and calcium measurements as clinical laboratory tests or biomarkers for PD0325901-induced systemic mineralization. Whereas measurement of plasma 1,25-dihydroxyvitamin D is technically complex and costly, evaluation of serum calcium and phosphorus is rapid and performed routinely in the clinical laboratory with historical reference ranges readily available. Although the data obtained with urinalysis were consistent with dysregulation of calcium and phosphorus homeostasis, concerns existed as to whether specific and reproducible urinalysis parameters could be developed for monitoring the safety of PD0325901. Based on the data obtained thus far, hyperphosphatemia appeared to be the primary factor for eliciting tissue mineralization, and serum phosphorus was proposed as the key analyte for monitoring.

An investigative study was conducted in male rats to assess the reversibility of serum chemistry changes following a single oral dose of PD0325901 (Brown et al., 2005a). The hypothesis was that serum phosphorus levels would return to control levels in the absence of drug administration. Male rats (10/group) received single oral doses at 1, 3, or 10 mg/kg, with controls receiving vehicle alone. Blood was collected on days 2, 3, 5, and 8 for serum chemistry analysis. Hyperphosphatemia (serum phosphorus increased up to 58%) and minimal increases in calcium occurred at all doses on days 2 and 3. Albumin was decreased at 10 mg/kg. These changes were completely reversible within a week. This study demonstrated that increases in serum phosphorus and calcium induced by PD0325901 are reversible following cessation of dosing. Although a single dose of 10 mg/kg produces systemic mineralization in rats, withdrawal of dosing results in normalization of serum calcium and phosphorus levels, indicating that the homeostatic mechanisms controlling these electrolytes remain intact. The results of this study were not unexpected. Oral administration to dogs of the vitamin D analogs dihydrotachysterol and Hytakerol (dihydroxyvitamin $D_2$-II) results in hypercalcemia that is reversible following termination of dosing (Chen et al., 1962). Reversal of hypercalcemia and hypercalciuria has been demonstrated in humans following cessation of dosing of various forms of vitamin D (calciferol, dihydrotachysterol, 1-α-hydroxycholecalciferol, or 1-α,25-dihydroxycholecalciferol) (Kanis and Russell, 1977).

Another investigative study was conducted in male rats to determine whether pamidronate (a bisphosphonate) or Renagel (sevelamer HCl; a phosphorus binder) would inhibit tissue mineralization induced by PD0325901 by inhibiting hyperphosphatemia. Bisphosphonates inhibit bone resorption and in so doing modulate serum calcium and phosphorus levels. Renagel is a nonabsorbable resin that contains polymers of allylamine hydrochloride, which forms ionic and hydrogen bonds with phosphate in the gut. Rats received daily oral doses of PD0325901 at 3 mg/kg for 14 days with or without co-treatment with pamidronate or Renagel. Pamidronate was given twice intravenously at 1.5 mg/kg one day prior to PD0325901 dosing and on day 6. Renagel was given daily as 5% of the diet beginning one day prior to PD0325901 dosing. Treatment groups consisted of oral vehicle alone, PD0325901 alone, pamidronate alone, Renagel alone, PD0325901 + pamidronate, and PD0325901 + Renagel. PD0325901 plasma AUC(0–24) values were 11.6, 9.17, and 4.34 μg·h/mL in the PD0325901 alone, PD0325901 + pamidronate, and PD0325901 + Renagel groups, respectively. Administration of PD0325901 alone resulted in hyperphosphatemia on days 3 and 15, which was inhibited by co-treatment with pamidronate or Renagel on day 3 only. PD0325901 alone resulted in systemic mineralization and skeletal changes consistent with changes seen in previous rat studies. Coadministration with either pamidronate or Renagel protected against systemic mineralization on day 3 only. Bone lesions were decreased with the co-treatments. Inhibition of toxicity with Renagel may have been due in part to decreased systemic drug exposure. However, the inhibition of toxic-

ity with pamidronate supports the role of a calcium–phosphorus dysregulation in PD0325901-induced systemic mineralization, because the inhibition of systemic mineralization observed on day 3 coincided with attenuation in the rise in serum phosphorus in these animals.

A two-week oral dose range-finding study was conducted in dogs in which doses tested were 0.2, 0.5, and 1.5 mg/kg (4, 10, and 30 mg/m$^2$, respectively). Also, a two-week oral-dose-range-finding study was conducted in cynomolgus monkeys at doses of 0.5, 3, and 10 mg/kg (6, 36, and 120 mg/m$^2$, respectively). In addition to standard toxicology and toxicokinetic endpoints, determination of inhibition of tissue and peripheral blood mononuclear cell pMAPK was performed to assess pharmacologic activity of PD0325901 in both studies. PTH and 1,25-dihydroxyvitamin D were evaluated in the monkey study. In both studies, mortality occurred at ≥0.5 mg/kg (dogs) and at 10 mg/kg (monkeys) due to injury to the gastrointestinal tract mucosa, inhibition of pMAPK occurred at all doses, and systemic mineralization was not observed in either study. Increases in serum phosphorus were seen in moribund animals and/or associated with renal hypoperfusion (resulting from emesis, diarrhea, and dehydration). These elevations in phosphorus were considered secondary to renal effects and were not associated with changes in serum calcium. Toxicologically significant increases in serum phosphorus or calcium were not evident at nonlethal doses in dogs or monkeys. In the two-week monkey study, a dose-related increase in 1,25-dihydroxyvitamin D was observed on day 2 only (after a single dose) at ≥3 mg/kg. This increase did not occur on days 7 or 15, and was not associated with changes in serum phosphorus or calcium, nor systemic mineralization. Therefore, there did not appear to be toxicologic significance to the day 2 increase in 1,25-dihydroxyvitamin D in monkeys.

## DISCUSSION

Mineralization of vasculature and various soft tissues (systemic mineralization) was observed in toxicology studies in rats in a time- and dose-dependent manner. This change was consistent with the presence of calcium–phosphorus deposition within the vascular wall and parenchyma of tissues such as the stomach, kidney, aorta, and heart. The stomach appeared to be the most sensitive tissue, since mineralization of gastric fundic mucosa occurred prior to the onset of mineralization in other tissues. Male rats were consistently more sensitive to this toxicity than were female rats. In the pivotal one-month toxicity study in rats, the no-effect level for systemic mineralization was 0.1 mg/kg (0.6 mg/m$^2$) in males and 0.3 mg/kg (1.8 mg/m$^2$) in females, which were associated with PD0325901 steady-state plasma AUC(0–24) values of 231 and 805 ng·h/mL, respectively. Systemic mineralization was not observed in dogs or monkeys, despite pharmacologic inhibition of tissue pMAPK levels (>70%), administration of lethal doses, and exposures greater than 10-fold of those that induced mineralization in rats (10,600 ng·h/mL in dogs and

up to 15,000 ng · h/mL in monkeys). Systemic mineralization was not observed in mice despite administration of PD0325901 at doses up to 50 mg/kg (150 mg/m$^2$).

Systemic mineralization observed in rats following administration of PD0325901 is consistent with vitamin D toxicity due to dysregulation in serum calcium and phosphorus homeostasis (Grant et al., 1963; Rosenblum et al., 1977; Kamio et al., 1979; Mortensen et al., 1996; and Morrow, 2001). A proposed hypothesis for the mechanism of this toxicity is depicted in Figure 3. Elevated serum phosphorus levels (hyperphosphatemia) and decreased serum albumin were observed consistently in rats administered PD0325901. Although serum albumin levels are decreased in rats treated with PD0325901, calcium values typically remain unchanged or slightly elevated in these animals, indicating that free, non-protein-bound calcium is increased (Rosol and Capen, 1997; Payne et al., 1979; Meuten et al., 1982). Decreased parathyroid hormone levels (PTH) were observed in the rat studies. PTH plays a central role in the hormonal control of serum calcium and phosphorus. PTH is produced by the parathyroid gland and induces conversion of 25-hydroxyvitamin D (which is produced in the liver) to 1,25-dihydroxyvitamin D (calcitriol) in the kidney. 1,25-Dihydroxyvitamin D elicits increased absorption of calcium from the gastrointestinal tract. In addition, PTH mobilizes calcium and phosphorus from bone by increasing bone resorption, increases renal absorption of calcium, and increases renal excretion of phosphorus (in order to regulate

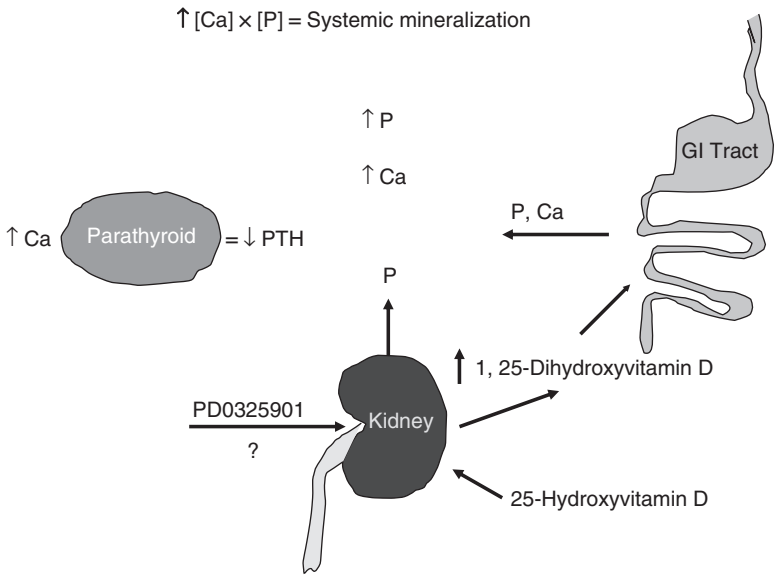


**Figure 3** Hypothesis for the mechanism for systemic mineralization in the rat following PD0325901 administration.

serum phosphorus levels). Elevations in serum calcium typically elicit decreased PTH levels as a result of the normal control (negative feedback loop) of this endocrine system (Rosol and Capen, 1997). The decreases in PTH observed in the rats were believed to be due to the elevations in serum calcium (hypercalcemia). Hyperphosphatemia in the presence of normo- or hypercalcemia can result in an increased Ca × P product, which is associated with systemic mineralization (Block, 2000). Hyperphosphatemia was also observed in rats administered PD176067, which is a reversible and selective inhibitor of fibroblast growth factor receptor tyrosine kinase. In these animals, vascular and soft tissue mineralization also occurs (aorta and other arteries, gastric fundic mucosa, myocardium, renal tubules), probably due to increased Ca × P product (Brown et al., 2005b).

Administration of PD0325901 to rats resulted in significantly increased levels of plasma 1,25-dihydroxyvitamin D. The mechanism for this action is not known but is not believed to be due to a metabolite of PD0325901. This is the most potent form of vitamin D and the primary metabolite responsible for regulating serum calcium and phosphorus. Vitamin D is converted to 25-hydroxyvitamin D in the liver and then 1-hydroxylated to 1,25-dihydroxyvitamin D in renal tubules. 1,25-Dihydroxyvitamin D acts by increasing absorption of calcium and phosphorus from the gastrointestinal tract, and can increase calcium and phosphorus reabsorption by renal tubules. Hyperphosphatemia and increased plasma 1,25-dihydroxyvitamin D levels in rats occurred 1 to 2 days prior to the detection of tissue mineralization at doses ≤3 mg/kg (18 mg/m$^2$).

Administration of PD0325901 to rats resulted in bone lesions that included necrosis of the metaphysis and the ossifying zone of the physis, and thickening of the zone of hypertrophying cartilage of the physis. The expansion of chondrocytes in the physis may be a response to the metaphyseal necrosis and loss of osteoprogenitor cells. These changes are characterized by localized injury to bone that appear to be due to local ischemia and/or necrosis. Skeletal vascular changes may be present in these animals, resulting in disruption of endochondral ossification. Skeletal lesions, including bone necrosis, can result from vitamin D intoxication (Haschek et al., 1978). The skeletal lesions observed in rats administered PD0325901 are similar to those reported with vitamin D toxicity, which provides additional evidence that toxicity occurred via induction of 1,25-dihydroxyvitamin D. Bone lesions similar to those observed in rats were not seen in dogs, monkeys, or mice administered PD0325901.

In summary, PD0325901-induced systemic mineralization in the rat results from a dysregulation in serum phosphorus and calcium homeostasis. This dysregulation appears to result from toxicologically significant elevations in plasma 1,25-dihydroxyvitamin D levels following drug administration. Based on the toxicology data, rats are uniquely sensitive to this toxicity. A summary of the primary target organ toxicities observed in the preclinical studies is presented in Table 5. Toxicity to the skin (epidermal lesions) and gastrointes-

**TABLE 5 Primary Target Organ Toxicities Observed in Preclinical Studies**

| | Species | | |
|---|---|---|---|
| Organ System | Rat | Dog | Monkey |
| Gastrointestinal tract | ×[a] | × | × |
| Skin | × | × | × |
| Systemic mineralization[b] | × | — | — |
| Bone | × | — | — |
| Liver | × | — | — |
| Gallbladder | n/a | — | × |

[a]Toxicity observed.
[b]Includes vascular (aorta, arteries) and soft tissue mineralization (e.g., stomach, heart, kidneys).

tinal tract (primarily ulcers/erosions in the mucosa) were observed across species and may have resulted from inhibition of MEK-related signal transduction pathways in these tissues (Brown et al., 2006). Gastrointestinal tract toxicity is dose-limiting in dogs and monkeys and was anticipated to be the dose-limiting toxicity of PD0325901 in the clinic. Therefore, gastrointestinal tract toxicity may preclude the development of other potential adverse events in humans, including potential dysregulation in serum phosphorus or calcium. It is not known whether systemic mineralization is relevant to humans. However, if PD0325901 does induce a dysregulation in serum calcium–phosphorus metabolism in humans, monitoring serum levels would provide an early indication of effects and guide modifications to dosing regimens.

To ensure patient safety in the phase I clinical trial with PD0325901, procedures were incorporated into the trial design to monitor for potential dysregulation in serum calcium–phosphorus homeostasis. Measurements of serum calcium, phosphorus, creatinine, albumin, and blood urea nitrogen were performed frequently during the initial treatment cycle (21 days of dosing in a 28-day cycle), with periodic measurement in subsequent cycles, and the serum Ca × P product was calculated. The serum Ca × P product has been determined to be a clinically useful value as a means for evaluating risk for tissue and/or vascular mineralization with the recommendation that the value not exceed 70 based on clinical use of vitamin D analogs such as Rocaltrol and Hectorol (Roche Laboratories, 1998; Bone Care International, Inc., 1999; Block, 2000). Serum calcium and phosphorus are readily measured in a clinical setting with well-established reference ranges available. The trial included a protocol-specific dose-limiting toxicity for a Ca × P product > 70, which required a confirmatory measurement and dose interruption for that patient. In addition, serum vitamin D, PTH, alkaline phosphatase (total and bone), osteocalcin, and urinary C- and N-terminal peptide of collagen 1 (markers of bone resorption) were included for periodic measurement. Criteria considered for exclusion of candidate patients from the clinical trial included a history of

malignancy-associated hypercalcemia, extensive bone metastasis, parathyroid disorder, hyperphosphatemia and renal insufficiency, serum calcium or phosphorus levels >1× the upper limit of normal, and/or concomitant use of calcium supplements and vitamin D in amounts exceeding normal daily allowances.

## CALCULATION OF CLINICAL STARTING DOSE

The currently accepted algorithm for calculating a starting dose in clinical trials with oncology drugs is to use one-tenth of the dose that causes severe toxicity (or death) in 10% of the rodents ($STD_{10}$) on a mg/m$^2$ basis, provided that this starting dose (i.e., 1/10 the $STD_{10}$) does not cause serious, irreversible toxicity in a nonrodent species (in this case, the dog) (DeGeorge et al., 1998). If irreversible toxicities are induced at the proposed starting dose in nonrodents or if the nonrodent (i.e., the dog) is known to be the more appropriate animal model, the starting dose would generally be one-sixth of the highest dose tested in the nonrodent (the dog) that does not cause severe, irreversible toxicity.

Calculation of the initial phase I starting dose of PD0325901 was based on the pivotal one-month toxicology studies in rats and dogs. Doses tested in the one-month rat study were 0.1, 0.3, and 1 mg/kg (0.6, 1.8, and 6 mg/m$^2$, respectively), and doses in the one-month dog study were 0.05, 0.1, and 0.3 mg/kg (1, 2, and 6 mg/m$^2$, respectively). Both studies included animals assigned to a one-month reversal phase, in the absence of dosing, to assess reversibility of any observed toxicities. In addition to standard toxicology and toxicokinetic parameters, these studies included frequent evaluation of serum chemistries, and measurement of vitamin D, osteocalcin, PTH, and inhibition of tissue pMAPK levels.

In the one-month rat study, no drug-related deaths occurred and systemic mineralization occurred in multiple tissues in both sexes at 1 mg/kg. Hypocellularity of the metaphyseal region of distal femur and/or proximal tibia occurred in males at 1 mg/kg. Toxicologic findings at lower doses included skin sores (at ≥0.1 mg/kg) and mineralization of gastric mucosa in one male at 0.3 mg/kg. The findings at ≤0.3 mg/kg were not considered to represent serious toxicity. In previous two-week dose-range-finding studies in rats, death occurred at 3 mg/kg (18 mg/m$^2$), indicating this to be the minimal lethal dose in rats. Based on these results, the $STD_{10}$ in rats was determined to be 1 mg/kg (6 mg/m$^2$)

In the one-month dog study, doses up to 0.3 mg/kg (6 mg/m$^2$) were well tolerated with minimal clinical toxicity. Primary drug-related toxicity was limited to skin sores in two animals at 0.3 mg/kg. One-tenth the $STD_{10}$ in rats is 0.6 mg/m$^2$, which is well below a minimally toxic dose (6 mg/m$^2$) in dogs. These data indicate an acceptable phase I starting dose to be 0.6 mg/m$^2$, which is equivalent to 1 mg in a 60-kg person. The relationships between the primary toxicities in rats and dogs with dose and exposure are presented in Figure 4.

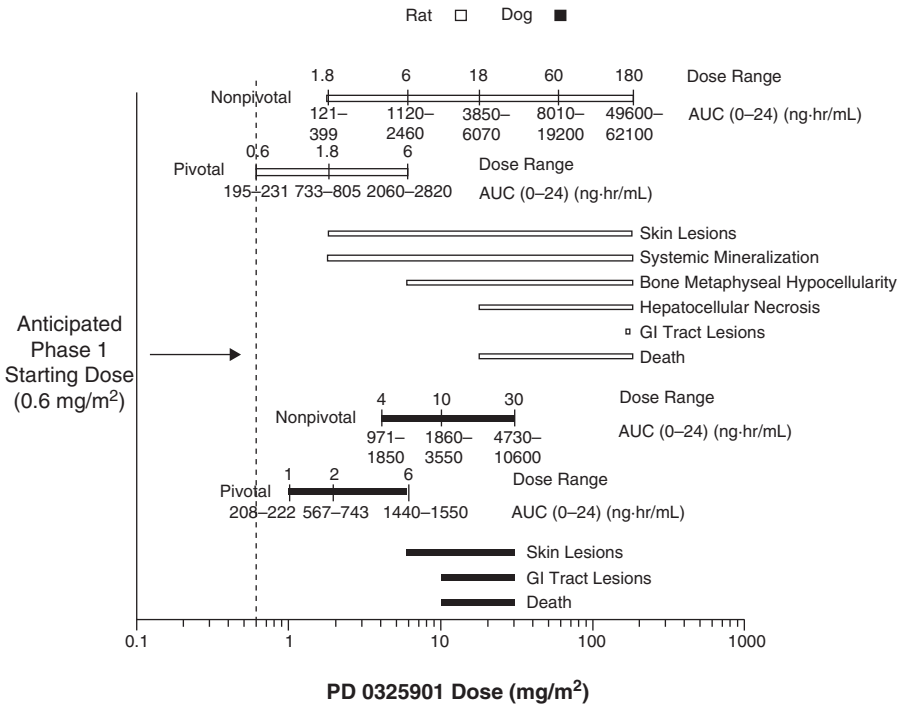


**Figure 4** Relationships between dose and exposure with the primary toxicities of PD0325901 in rats and dogs. Exposure is expressed as PD0325901 plasma AUC(0–24) in ng · hr/mL and dose in mg/m$^2$. Results from nonpivotal (dose-range-finding) studies and the pivotal one-month toxicity studies are presented.

In the phase I trial of PD0325901 in cancer patients (melanoma, breast, colon, and non-small cell lung cancer), oral doses were escalated from 1 to 30 mg twice daily (BID). Each treatment cycle consisted of 28 days, and three schedules of administration were evaluated: (1) 3 weeks of dosing with one week off, (2) dosing every day, and (3) 5 days of dosing with 2 days off per week (LoRusso et al., 2005; Menon et al., 2005; Tan et al., 2007; LoRusso et al., 2007). Doses ≥2 mg BID suppressed tumor pMAPK (indicating biochemical activity of the drug) and acneiform rash was dose limiting in one patient at 30 mg BID. There were no notable effects on serum Ca × P product and the most common toxicities included rash, fatigue, diarrhea, nausea, visual disturbances and vomiting. Acute neurotoxicity was frequent in patients receiving ≥15 mg BID (all schedules) and several patients developed optic nerve ischemia, optic neuropathy, or retinal vein occlusion (LoRusso et al., 2007). In the phase I trial, there were three partial responses (melanoma) and stable disease in 24 patients (primarily melanoma) (LoRusso et al., 2007). In a pilot phase II study of PD0325901 in heavily pretreated patients with non-

small cell lung cancer, 15 mg BID was given on various schedules over a 28-day cycle (Haura et al., 2007). The main toxicities were reversible visual disturbances, diarrhea, rash, and fatigue. The mean trough concentration of PD0325901 at 15 mg BID was 108 ng/mL (at steady state) and there were no objective responses (Haura et al., 2007).

## CONCLUSIONS

Tissue mineralization produced in rats administered the MEK inhibitor PD0325901 provides a case study of how a unique and serious toxicity observed in preclinical safety testing was effectively managed to allow progression of an experimental drug into human clinical trials. A number of key factors were critical for allowing continued development of this compound to occur, rather than early termination. PD0325901 represented a novel and targeted therapeutic agent for the treatment of various solid tumors, thereby allowing a high risk–benefit ratio to exist due to the significant unmet medical need posed by cancer. Phase I oncology trials typically occur in cancer patients with limited treatment options. Therefore, the barriers to entry for novel anticancer agents in the clinic are generally lower than for phase I trials involving healthy volunteers and therapies for non-life-threatening indications.

Early in the toxicology program with PD0325901, lesions observed in rats were similar to those seen with vitamin D toxicity, and serum chemistry data indicated changes in phosphorus and calcium. This information provided the basis for the hypotheses to be proposed regarding the mechanism for vascular and soft tissue mineralization. Because mineralization occurred in rats administered PD0325901 rather than only in dogs or monkeys, an animal model suitable for multiple investigative studies was readily available. Despite the apparent species specificity for this toxicity, it was not appropriate to discount the risks toward humans because of a "rat-specific" finding. Rather, it was important to generate experimental data that characterized the toxicity and provided a plausible mechanism as a basis for risk management. Studies conducted with PD0325901 examined the dose–response and exposure–response relationships for toxicity and pharmacologic inhibition of MEK, the time course for lesion development, whether the changes observed were reversible or not, and whether associations could be made between clinical laboratory changes and anatomic lesions. We were able to identify biomarkers for tissue mineralization that were specifically related to the mechanism, were readily available in the clinical setting, noninvasive, and had acceptable assay variability. It is important that biomarkers proposed for monitoring for drug toxicity be scientifically robust and obtainable, and meet expectations of regulatory agencies. Finally, the data generated during the preclinical safety evaluation of PD0325901 were used to design the phase I–II clinical trial to ensure patient safety. This included selection of a safe starting dose for phase I,

criteria for excluding patients from the trial, and clinical laboratory tests to be included as biomarkers for calcium–phosphorus dysregulation and tissue mineralization. In conclusion, robust data analyses, scientific hypothesis testing, and the ability to conduct investigative work were key factors in developing a biomarker for a serious preclinical toxicity, thereby allowing clinical investigation of a novel drug to occur.

**Acknowledgments**

Numerous people at the Pfizer Global Research and Development (PGRD), Ann Arbor, Michigan, Laboratories were involved in the studies performed with PD0325901, including the Departments of Cancer Pharmacology and Pharmacokinetics, Dynamics and Metabolism. In particular, the author would like to acknowledge the men and women of Drug Safety Research and Development, PGRD, Ann Arbor, who conducted the toxicology studies with this compound and made significant contributions in the disciplines of anatomic pathology and clinical laboratory testing during evaluation of this compound.

## REFERENCES

Ally S, Clair T, Katsaros D, et al. (1989). Inhibition of growth and modulation of gene expression in human lung carcinoma in athymic mice by site-selective 8-Cl-cyclic adenosine monophosphate. *Cancer Res*, 49:5650–5655.

Block GA (2000). Prevalence and clinical consequences of elevated Ca × P product on hemodialysis patients. *Clin Nephrol*, 54(4):318–324.

Bone Care International, Inc., (1999). Package insert, Hectorol™ (doxercalciferol) capsules. June 9.

Brown AP, Morrissey RL, Smith AC, Tomaszewski JE, Levine BS (2000). Comparison of 8-chloroadenosine (NSC-354258) and 8-chloro-cyclic-AMP (NSC-614491) toxicity in dogs. *Proc Am Assoc Cancer Res*, 41:491 (abstract 3132).

Brown AP, Courtney C, Carlson T, Graziano M (2005a). Administration of a MEK inhibitor results in tissue mineralization in the rat due to dysregulation of phosphorus and calcium homeostasis. *Toxicologist*, 84(S-1):108 (abstract 529).

Brown AP, Courtney CL, King LM, Groom SL, Graziano MJ (2005b). Cartilage dysplasia and tissue mineralization in the rat following administration of a FGF receptor tyrosine kinase inhibitor. *Toxicol Pathol*, 33(4):449–455.

Brown AP, Reindel JF, Grantham L, et al. (2006). Pharmacologic inhibitors of the MEK-MAP kinase pathway are associated with toxicity to the skin, stomach, intestines, and liver. *Proc Am Assoc Cancer Res*, 47:308 (abstract 1307).

Brown AP, Carlson TCG, Loi CM, Graziano MJ (2007). Pharmacodynamic and toxicokinetic evaluation of the novel MEK inhibitor, PD0325901, in the rat following oral and intravenous administration. *Cancer Chemother Pharmacol*, 59:671–679.

Chen PS, Terepka AR, Overslaugh C (1962). Hypercalcemic and hyperphosphatemic actions of dihydrotachysterol, vitamin $D_2$ and Hytakerol (AT-10) in rats and dogs. *Endocrinology*, 70:815–821.

DeGeorge JJ, Ahn CH, Andrews PA, et al. (1998). Regulatory considerations for preclinical development of anticancer drugs. *Cancer Chemother Pharmacol*, 41:173–185.

Dent P, Grant S (2001). Pharmacologic interruption of the mitogen-activated extracellular-regulated kinase/mitogen-activated protein kinase signal transduction pathway: potential role in promoting cytotoxic drug action. *Clin Cancer Res*, 7:775–783.

Ferreira A, Drueke TB (2000). Biological markers in the diagnosis of the different forms of renal osteodystrophy. *Am J Med Sci*, 320(2):85–89.

Friday BB, Adjei AA (2008). Advances in targeting the Ras/Raf/MEK/Erk mitogen-activated protein kinase cascade with MEK inhibitors for cancer therapy. *Clin Cancer Res*, 14(2):342–346.

Fu JY, Muller D (1999). Simple, rapid enzyme-linked immunosorbent assay (ELISA) for the determination of rat osteocalcin. *Calcif Tissue Int*, 64:229–233.

Giachelli CM, Jono S, Shioi A, Nishizawa Y, Mori K, Morii H (2001). Vascular calcification and inorganic phosphate. *Am J Kidney Dis*, 38(4, Suppl 1):S34–S37.

Grant RA, Gillman T, Hathorn M (1963). Prolonged chemical and histochemical changes associated with widespread calcification of soft tissues following brief calciferol intoxication. *Br J Exp Pathol*, 44(2):220–232.

Harrington DD, Page EH (1983). Acute vitamin $D_3$ toxicosis in horses: case reports and experimental studies of the comparative toxicity of vitamins $D_2$ and $D_3$. *J Am Vet Med Assoc*, 182(12):1358–1369.

Haschek WM, Krook L, Kallfelz FA, Pond WG (1978). Vitamin D toxicity, initial site and mode of action. *Cornell Vet*, 68(3):324–364.

Haura EB, Larson TG, Stella PJ, et al. (2007). A pilot phase II study of PD-0325901, an oral MEK inhibitor, in previously treated patients with advanced non-small cell lung cancer. Presented at the AACR-NCI-EORTC International Conference on Molecular Targets and Cancer Therapy, abstract B110.

Hoshino R, Chatani Y, Yamori T, et al. (1999). Constitutive activation of the 41-/43-kDa mitogen-activated protein kinase pathway in human tumors. *Oncogene*, 18:813–822.

Jost M, Huggett TM, Kari C, Boise LH, Rodeck U (2001). Epidermal growth factor receptor–dependent control of keratinocyte survival and *Bcl-XL* expression through a MEK-dependent pathway. *J Biol Chem*, 276(9):6320–6326.

Kamio A, Taguchi T, Shiraishi M, Shitama K, Fukushima K, Takebayashi S (1979). Vitamin D sclerosis in rats. *Acta Pathol Jpn*, 29(4):545–562.

Kanis JA, Russell RGG (1977). Rate of reversal of hypercalcaemia and hypercalciuria induced by vitamin D and its 1α-hydroxylated derivatives. *Br Med J*, 1:78–81.

Knutson JC, LeVan LW, Valliere CR, Bishop CW (1997). Pharmacokinetics and systemic effect on calcium homeostasis of 1α,25-dihydroxyvitamin $D_2$ in rats. *Biochem Pharm*, 53:829–837.

Long GG (1984). Acute toxicosis in swine associated with excessive dietary intake of vitamin D. *J Am Vet Med Assoc*, 184(2):164–170.

LoRusso P, Krishnamurthi S, Rinehart JR, et al. (2005). A Phase 1–2 clinical study of a second generation oral MEK inhibitor, PD0325901 in patients with advanced cancer. 2005 ASCO Annual Meeting Proceedings. *J Clin Oncol*, 23(16S), abstract 3011.

LoRusso PA, Krishnamurthi SS, Rinehart JJ, et al. (2007). Clinical aspects of a phase I study of PD-0325901, a selective oral MEK inhibitor, in patients with advanced cancer. Presented at the AACR-NCI-EORTC International Conference on Molecular Targets and Cancer Therapy, abstract B113.

Mansour SJ, Matten WT, Hermann AS, et al. (1994). Transformation of mammalian cells by constitutively active MAP kinase kinase. *Science*, 265:966–970.

Menon SS, Whitfield LR, Sadis S, et al. (2005). Pharmacokinetics (PK) and pharmacodynamics (PD) of PD0325901, a second generation MEK inhibitor after multiple oral doses of PD0325901 to advanced cancer patients. 2005 ASCO Annual Meeting Proceedings. *J Clin Oncol*, 23(16S), abstract 3066.

Meuten DJ, Chew DJ, Capen CC, Kociba GJ (1982). Relationship of serum total calcium to albumin and total protein in dogs. *J Am Vet Med Assoc*, 180:63–67.

Milella M, Kornblau SM, Estrov Z, et al. (2001). Therapeutic targeting of the MEK/MAPK signal transduction module in acute myeloid leukemia. *J Clin Invest*, 108(6):851–859.

Morrow C (2001). Cholecalciferol poisoning. *Vet Med*, 905–911.

Mortensen JT, Lichtenberg J, Binderup L (1996). Toxicity of 1,25-dihydroxyvitamin $D_3$, tacalcitol, and calcipotriol after topical treatment in rats. *J Inv Dermatol Symp Proc*, 1:60–63.

Payne RB, Carver ME, Morgan DB (1979). Interpretation of serum total calcium: effects of adjustment for albumin concentration on frequency of abnormal values and on detection of change in the individual. *J Clin Pathol*, 32:56–60.

Rinehart J, Adjei AA, LoRusso PM et al. (2004). Multicenter phase II study of the oral MEK inhibitor, CI-1040, in patients with advanced non-small-cell lung, breast, colon, and pancreatic cancer. *J Clin Oncol*, 22(22):4456–4462.

Roche Laboratories (1998). Package insert, Rocaltrol® (calcitriol) capsules and oral solution. Nov. 20.

Rosenblum IY, Black HE, Ferrell JF (1977). The effects of various diphosphonates on a rat model of cardiac calciphylaxis. *Calcif Tissue Res*, 23:151–159.

Rosol TJ, Capen CC (1997). Calcium-regulating hormones and diseases of abnormal mineral (calcium, phosphorus, magnesium) metabolism. In *Clinical Biochemistry of Domestic Animals*, 5th ed. Academic Press, San Diego, CA, pp. 619–702.

Saunders MP, Salisbury AJ, O'Byrne KJ, et al. (1997). A novel cyclic adenosine monophosphate analog induces hypercalcemia via production of 1,25-dihydroxyvitamin D in patients with solid tumors. *J Clin Endocrinol Metab*, 82(12):4044–4048.

Sebolt-Leopold JS, Dudley DT, Herrera R, et al. (1999). Blockade of the MAP kinase pathway suppresses growth of colon tumors in vivo. *Nat Med*, 5(7):810–816.

Sebolt-Leopold JS (2000). Development of anticancer drugs targeting the MAP kinase pathway. *Oncogene*, 19:6594–6599.

Sebolt-Leopold JS, Merriman R, Omer C, (2004). The biological profile of PD0325901: a second generation analog of CI-1040 with improved pharmaceutical potential. *Proc Am Assoc Cancer Res*, 45:925 (abstract 4003).

Spangler WL, Gribble DH, Lee TC (1979). Vitamin D intoxication and the pathogenesis of vitamin D nephropathy in the dog. *Am J Vet Res*, 40:73–83.

Spaulding SW, Walser M (1970). Treatment of experimental hypercalcemia with oral phosphate. *J Clin Endocrinol*, 31:531–538.

Tan W, DePrimo S, Krishnamurthi SS, et al. (2007). Pharmacokinetic (PK) and pharmacodynamic (PD) results of a phase I study of PD-0325901, a second generation oral MEK inhibitor, in patients with advanced cancer. Presented at the AACR-NCI-EORTC International Conference on Molecular Targets and Cancer Therapy, abstract B109.

Wang D, Boerner SA, Winkler JD, LoRusso PM (2007). Clinical experience of MEK inhibitors in cancer therapy. *Biochimi Biophys Acta*, 1773:1248–1255.

York MJ, Evans GO (1996). Electrolyte and fluid balance. In Evans GO (ed.), *Animal Clinical Chemistry: A Primer for Toxicologists*. Taylor & Francis, New York, pp. 163–176.

# BIOMARKERS FOR THE IMMUNOGENICITY OF THERAPEUTIC PROTEINS AND ITS CLINICAL CONSEQUENCES

Claire Cornips, B.Sc., and Huub Schellekens, M.D.
*Utrecht University, Utrecht, The Netherlands*

## INTRODUCTION

Therapeutic proteins such as growth factors, hormones, monoclonal antibodies (mAbs), and others have increased in use dramatically over the past two decades, although their first use dates back more than a century, when animal antisera were introduced for the treatment and prevention of infections. Therapeutic proteins have always been associated with immunogenicity, although the incidence differs widely [1]. Granulocyte colony-stimulating factor (G-CSF) is the only protein in clinical use that has not been reported to induce antibodies. The first proteins used in medicine around 1900 were of animal origin. As foreign proteins they induced high levels of antibodies in the majority of patients after a single or a few injections. The type of immunological response induced was identical to that seen with vaccines.

Most of the therapeutic proteins introduced in recent decades are homologs of human proteins. However, in contrast to expectations, these proteins also appear to induce antibodies, and in some cases, in the majority of patients. Given that these antibodies are directed against auto-antigens, the immunological reaction involves breaking B-cell tolerance.

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

## MECHANISMS OF ANTIBODY INDUCTION

There are two main mechanisms by which antibodies are induced by therapeutic proteins (Table 1). If the proteins are of completely of foreign origin, such as asparginase or streptokinase, and the first generation mAbs derived from murine cells or partly foreign such as chimeric or humanized mAbs, the antibody response is comparable to a vaccination reaction. Often a single injection is sufficient to induce high levels of neutralizing antibodies, which may persist for a considerable length of time.

The other mechanism is based on breaking B-cell tolerance, which normally exists, to self-antigens, such as human immunoglobulins or products such as epoetins and the interferons. For breaking B-cell tolerance, prolonged exposure to proteins is necessary. In general, it takes months before patients produce antibodies, which are mainly binding and disappear when treatment is stopped.

To induce a classical immune reaction a degree of nonself is necessary. The trigger for this type of immunogenicity is the divergence from the comparable human proteins. The triggers for breaking tolerance are quite different. The production of these auto-antibodies may occur when the self-antigens are exposed to the immune system in combination with a T-cell stimulus or danger signal such as bacterial endotoxins, microbial DNA rich in GC motifs or denatured proteins. This mechanism explains the immunogenicity of biopharmaceuticals containing impurities. When tolerance is broken with this type of mechanism, the response is often weak with low levels of antibodies with low affinity.

To induce high levels of IgG, the self-antigens should be presented to the immune system in a regular array form with a spacing of 50 to 100 Å, a supramolecular structure resembling a viral capsid [2]. Apparently the immune system has been selected to reach vigorously to these types of structures which

**TABLE 1 Main Markers of Immunogenicity of Therapeutic Proteins**

| | Product | Preclinical | Treatment | Patients |
|---|---|---|---|---|
| (Partly) foreign proteins inducing a classical immune response | Level of nonself<br>Presence of T-cell epitopes<br>Biological activity of the product | — | — | Lack of immune tolerance<br>Concomitant therapy |
| Self-protein breaking tolerance | Presence of aggregates<br>Biological activity of the product | Induction of an immune response in immune-tolerant mice | Chronic treatment | Nonimmune compromised<br>Concomitant therapy |

normally are found only in viruses and other microbial agents. The most important factor in the immunogenicity of therapeutic proteins which are human homologs has been the presence of aggregates [3]. Aggregates may be presenting the self-antigens in a repeating form which is such a potent inducer of auto-antibodies.

## FACTORS INFLUENCING IMMUNOGENICITY

So the primary factors inducing an antibody response are aggregates in the case of human proteins and the degree of nonself in (partly) nonhuman proteins. There are also cases in which the immune response cannot be explained. There are, however, a number of other factors that may influence the level of the immune response [4]:

- Product characteristics
  1. Biological function
  2. Impurities and contaminants
  3. Product modification
- Treatment characteristics
  1. Length of treatment
  2. Route of administration
  3. Dosage
  4. Concomitant therapy
- Patient characteristics
  1. Genetic background
  2. Type of disease
  3. Age
  4. Gender
- Unknown factors

### Product Characteristics

The biological activities of the product are influencing the immune response. An immune stimulating therapeutic protein is more likely to induce antibodies than an immune suppressive protein. Monoclonal antibodies targeted to cell-bound epitopes are more likely to induce an immune response than monoclonal antibodies with a target in solution. Also, the Fc-bound activities of monoclonal antibodies have an influence.

Impurities may influence immunogenicity. The immunogenicity of products as human growth hormone, insulin, and interferon α-2 have declined over the years due to improved downstream processing and formulation, reducing the level of impurities. There are studies showing the induction of antibodies by

oxidized protein which cross-reacted with the unmodified product [5] and host cell–derived endotoxin acting as an adjuvant. The probability of an immune response therefore increases with the level of impurities. Product modifications that are intended to enhance half-life potentially also increase the exposition of the protein to the immune system and may increase immunogenicity. In addition, the modification may reduce biological activity necessitating more protein for the same biological effect.

PEGylation (polyethylene glycol) is claimed to reduce the immunogenicity of therapeutic proteins by shielding. There is evidence that pegylation reduces the immunogenicity of nonhuman proteins such as bovine adenosine deamidase and asparginase. Whether pegylation also reduces the capacity of human proteins to break B-cell tolerance is less clear. There are reports of high immunogenicity of pegylated human proteins such as MDGF, but the immunogenicity of unpegylated MDGF products is unknown.

## Treatment Characteristics

Foreign proteins such as streptokinase and asparaginase often induce antibodies after a single injection. Breaking B-cell tolerance by human protein takes in general more than six months of chronic treatment. The route of administration influences the likelihood of an antibody response independent of the mechanism of induction. The probability of an immune response is highest with subcutaneous administration, less probable after intramuscular administration, and intravenous administration is the least immunogenic route. There are no studies comparing parenteral and nonparenteral routes of administration. However, intranasal and pulmonary administration of therapeutic proteins may induce an immune response.

## Patient Characteristics

Gender, age, and ethnic background have all been reported to influence the incidence of antibody response to specific therapeutic proteins. However, the only patient characteristic that has consistently been identified for a number of different products is the disease from which the patients suffer. Cancer patients are less likely to produce antibodies to therapeutic protein than other patients. The most widely accepted explanation for this difference is the immune-compromised state of cancer patients, both by the disease as by anticancer treatment. Also, the median survival of patients on treatment by therapeutic proteins may be too short to develop an antibody response. In any case, cancer reduces considerably the probability of an antibody response to a protein. As the experience in cancer patients shows, immune suppressive therapy reduces the probability to develop an immune response to proteins. In addition, immune-suppressive drugs such as methotrexate are used in conjunction with monoclonal antibodies and other protein drugs to reduce the immune reactions.

## PREDICTION OF IMMUNOGENICITY IN ANIMALS

In principle, all therapeutic proteins are immunogenic in conventional laboratory animals, and their predictive value depends on the type of proteins [6]. The immune reaction in animals to biopharmaceuticals of microbial or plant origin is comparable to humans, as they are comparably foreign for all mammalian species. Animal studies in which the reduction of immunogenicity is evaluated therefore have a high degree of predictability for immunogenicity in humans.

The development of antibodies has been observed regularly in preclinical studies in animals of biopharmaceuticals homologous to human proteins. Being considered a normal reaction to a foreign protein, it has led to the generally held assumption that immunogenicity testing and, in some cases, even preclinical testing in animals is irrelevant. However, not all antibodies interfere with the biological activity of a biopharmaceutical. And if there is a biological or clinical effect, these may help to identify the possible sequelae of immunogenicity, as has been shown with human epoetin in dogs. In the canine model human epoetin is immunogenic and it induces antibodies that neutralize the native canine epoetin, leading to pure red cell aplasia. This severe complication of antibodies to epoetin was later confirmed in humans.

Also, antibody-positive animals may provide sera for the development and validation of antibody assays; and the evaluation of an antibody response in animals is important to evaluate the safety and pharmacokinetic data in conventional laboratory animals. Nonhuman primates have been advocated as better models to predict the immunogenicity of human proteins because of a high sequence homology between the product and the monkey native molecule to which the animal is immune tolerant. Immunogenicity studies in nonhuman primates have, however, also shown mixed results. Products with a high immunogenicity in monkeys sometimes do not induce antibodies in human patients. The opposite has also been observed, although these studies may have been too limited in length of treatment or number of animals to be truly predictive. A good example of the possible use of monkeys was a study to determine the immunogenicity of different human growth hormone (hGH) preparations using a hyper-immunization protocol, including the use of adjuvant to provide a worst-case scenario of immunogenicity [7]. The monkeys were treated with cadaver-derived methionyl-hGH and natural sequence hGH. The antibody response was 81%, 69%, and 5 to 23%, respectively, which reflects the relative immunogenicity of these preparations in human patients. So in this example, rhesus monkeys predict relative immunogenicity.

Also, the immunogenicity of lys-pro biosynthetic human insulin was compared with the immunogenicity of porcine insulin and native-sequence insulin in rhesus monkeys. Neither of these proved to induce antibodies. With tissue plasminogen activators the immunogenicity in nonhuman primates was reported by the same group to reflect the immunogenicity in patients. Rhesus monkeys (*Macaca mulatta*) and cynomolgus monkeys (*M. fascicularis*) were

used to test antibody response to EPO/GM-CSF hybrids. Two of the three constructs tested produced high levels of neutralizing antibodies. In the rhesus monkeys these hybrids caused severe anemia. In cynomolgus monkeys no effect on hematological parameters was found, indicating lack of cross-reactivity of the antibodies with the native cynomolgus erythropoietin. There are also reports on immunogenicity of products in monkeys such as IFN α B/D, diaspirin cross-linked hemoglobin, and IL-3, which did not induce antibodies in patients. So monkeys cannot be used as absolute predictors of immunogenicity in humans, and even the response within the macaque family seems to differ.

Theoretically, the best predictive model for immunogenicity of human proteins is the transgenic mouse. These animals are immune tolerant for the human protein they express. The caveats are that the wild-type mouse strain used for transgenesis should be able to produce antibodies to the protein and that the transgenic should show immune tolerance for the native molecule and may be differences in the processing of antigens and epitope recognition between mouse and humans. Mice, transgenic for human insulin, showed that immunogenicity to variant insulin molecules was dependent on the number of substitutions. Mice that were made transgenic for human tissue plasminogen activator variants with a single amino acid substitution proved to be immunogenic, showing the discriminatory potential of the model [8]. The transgenic approach proved useful in finding a reason for the increased immunogenicity caused by a specific formulation. In mice transgenic for interferon α-2 only the batches immunogenic in patients produced antibodies [9]. The presence of aggregates and oxidized proteins proved the main cause of immunogenicity in the transgenic animals as in patients. In mice transgenic for hGH the immunogenicity of hGH encapsulated in microspheres for sustained release was tested and no enhanced immunogenicity was observed. We have used the transgenic mouse models to study the product characteristics capable of breaking tolerance. In these models, aggregates were shown to be the major factor. These models, however, have been used by us and others mainly in a yes-or-no fashion to study factors such as aggregates and sequence variations. To find subtle differences that may exist between products and before these models can be used to fully predict the immunogenic potential of human therapeutic proteins, much more validation is necessary. With no models yet available with sufficient predictive power, clinical studies are the only sure way to establish the induction of antibodies by therapeutic proteins.

## ASSAYS FOR ANTIBODIES

A cellular immune response to therapeutic proteins has never been established. Also, the biological and clinical consequences of an immune response to these products have always been associated with the presence of antibodies. So, by definition, a positive assay for antibodies is the biomarker for the immunogenicity of proteins. The lack of standardized assays and international

reference sera presents a major problem in assessing immunogenicity. It makes a comparison of antibody data generated by different laboratories impossible; comparisons of products based on published data (e.g., using information on package inserts) are also meaningless.

Recently, excellent reviews have been published regarding the development and validation of the various assay formats for antibodies to therapeutic proteins. There are two principles for the testing of antibodies: assays that monitor binding of antibody to the drug (EIA, RIA, BIAcore), or (neutralizing) bioassays. These assays are used in combination: a sensitive binding assay to identify all antibody-containing samples is often more practical, as bioassays are usually more difficult and time consuming. If a native molecule is modified (e.g., pegylated or truncated) to obtain a new product with different pharmacological characteristics, both the "parent" and the new molecule should be used as capture antigens for antibody assays.

A definition for the negative cutoff must be included in the validation process of this assay and is often based on the 5% false-positive rate. Such an analytical cutoff per se is not predictive of biological or clinical effect but rather, indicative of the technical limitations of the assay. And because the cutoff is set to include a relatively high number of false positives, all initial positive sera should be confirmed either by another binding assay or by a displacement assay.

The confirmed positive samples should be tested in a bioassay for neutralizing antibodies, which correlate with a potential in vivo effect in patients, as these antibodies are usually neutralizing because they interfere with receptor binding. A confirmatory step for neutralizing antibodies is not necessary because these antibodies are a subset of binding antibodies. In some cases it may be important to show the neutralization to be caused by antibodies if the presence of other inhibitory factors, such as a soluble receptor, has to be excluded. Further characterization of the neutralizing antibody response may follow, such as isotype, affinity, and specificity assays.

A positive or negative answer from the assays is not sufficient; development of antibodies is a dynamic process, and therefore the course of antibody development (kinetics) must be plotted quantitatively over time. Usually, persisting titers of neutralizing antibodies correlate with some biological effect. An analysis of the biological effect based on incidence may also be misleading. Akin to population kinetics in pharmacology, a method that compares relative immunogenicity in patient groups as "mean population antibody titers," taking into account both incidence and titers of neutralizing antibodies, has also been found more useful than a percentage of seroconversions for the purpose of comparing two products.

## CONSEQUENCES OF ANTIBODIES TO THERAPEUTIC PROTEINS

In many cases the presence of antibodies is not associated with biological or clinical consequences. The effects that antibodies may induce depend on their

level and affinity and can be the result of antigen–antibody reaction in general or of the specific interaction. Severe general immune reactions as anaphylaxis associated with the use of animal antisera have become rare because the purity of the products increased substantially. Delayed-type infusion-like reactions resembling serum sickness are more common, especially with monoclonal antibodies and other proteins administered in relative large amounts and the formation of immune complexes. Patients with a slow but steadily increasing antibody titer are reported to show more infusion-like reactions than patients with a short temporary response.

The consequences of the specific interaction between protein drugs is dependent on the affinity of the antibody translating in binding and/or neutralizing capacity. Binding antibodies may influence the pharmacokinetic behavior of the product, and both increases and reductions of half-life have been reported, resulting in enhancement or reduction in activity. Persisting levels of neutralizing antibodies in general result in a loss of activity of the protein drug. In some cases the loss of efficacy can easily be monitored by the increase of disease activity. For example, in interferon alpha treatment of hepatitis C, viral activity can be monitored by transaminase activity. Loss of efficacy is correlated directly by increased viral activity and increase in transaminase levels.

In the case of interferon beta treatment of multiple sclerosis the loss of efficacy is much more difficult to measure because the mode of action of the therapeutic protein is not known and the disease progress is unpredictable and difficult to monitor. The reduction of Mx induction which is specific for interferon activity has been used successfully to evaluate the biological effect of antibodies to interferon beta. The adverse effects of therapeutic proteins are in general the result of an exaggerated pharmacodynamic effect. So the loss of side effects may also be the result of the induction of antibodies and may be the first sign of immunogenicity. For example, in patients treated with interferon the loss of flulike symptoms is associated with the appearance of antibodies. Because by definition neutralizing antibodies interact with ligand–receptor interaction, they will inhibit the efficacy of all products in the same class with serious consequences for patients if there is no alternative treatment. The most dramatic side effects occur if the neutralizing antibodies cross-react with an endogenous factor with an essential biological function. This had been described for antibodies induced by epoetin alpha [10] and megakaryocyte growth and differentiation factor (MGDF), which led, respectively, to life-threatening anemia and thrombocytopenia, sometimes lasting for more than a year.

## Skin Reactions

Skin reactions are a common side effect of therapeutic proteins, and some of these reactions are associated with an immunogenic response. But can these skin reactions be used as a marker for the immunogenicity of therapeutic

proteins? The hypersensitivity reactions are classified as type I, II, III, and IV reactions. The type I reaction is IgE mediated. The type II reactions are caused by activated T-killer cells and macrophages and complement activation. The type III hypersensitivity reaction is caused by the disposition of immune complexes [11]. Type IV is T-cell mediated.

Type I hypersensitivity or IgE-mediated allergies are very rare, and most are related to the excipients present in formulation rather than the protein drug products. IgE-mediated reactions against human insulin have been reported, although is less common than with pork or beef insulin [12]. Theoretically, the type II hypersensitivity skin reaction may be a symptom of the immunogenicity of therapeutic proteins. During a type II hypersensitivity reaction, antibodies activate T-killer cells, macrophages, or complement factors to induce an immune response. However, the antibodies produced by most therapeutic proteins as a consequence of breaking B-cell tolerance and T-cells play only a minor role, if any.

TNF inhibitors such as etanercept cause injection-site reactions by a T-cell-mediated delayed-type hypersensitivity reaction. Antibodies against etanercept have been shown not to be correlated to adverse events. Skin reactions probably are a class effect of TNF inhibitors. Blockade of TNF can stimulate certain forms of autoimmunity by increasing T-cell reactivity to microbial and self-antigens [13].

The skin reactions that are part of the type III hypersensitivity reaction caused by the local disposition of immune complexes are seen after treatment with monoclonal antibodies, which are used in relative high and repeated doses. These immune complexes may lead to anaphylactoid reactions and serum sickness–like symptoms. Skin reactions such as urticaria and rashes are common symptoms of this complication [14]. Monoclonal antibodies may also lead to local reactions at the injection site. However, as shown in Table 2, there is no relation between immunogenicity and local skin reactions. So these local reactions cannot be used as an early marker for more serious symptoms of immunogenicity.

Some skin reactions seen after treatment with therapeutic proteins are the result of their pharmacodynamics: for example, the epidermal growth factor

**TABLE 2 Relation Between Local Skin Reactions and Immunogenicity of Monoclonal Antibodies**[a]

| | INN | Incidence of Antibodies (%) | Incidence of Local Skin Reactions (%) |
|---|---|---|---|
| Humira | adalimumab | 12 | 20 |
| Remicade | infliximab | 24 | 0 |
| Xolair | omalizumab | 0 | 45 |
| Mabthera | rituximab | 1 | >1 |

[a]Based on SPCs.

receptor (EGFR) inhibitors and beta interferon. EGFR plays a key role in normal skin function; thus, it is very likely that the rash and other skin reactions during therapy with EGFR inhibitors are caused by the inhibition of EGFR [14,15]. Interferon beta is known to induce high levels of neutralizing antibodies, but it seems that the adverse effects and skin reactions that occur during the first months of treatment have already disappeared when patients develop antibodies. Most probably, these adverse effects are a direct pharmacodynamic effect of interferon. Indeed, patients with neutralizing antibodies have a smaller risk of adverse events such as injection site reactions than do patients without [16,17]. So both local skin reactions at the site of injection and generalized reactions can be seen after the use of therapeutic proteins. These skin reactions can have different causes and cannot be used as biomarkers for immunogenicity.

## BIOMARKERS FOR THE IMMUNOGENICITY OF THERAPEUTIC PROTEINS

Two classes of biomarkers are used to indicate the clinical consequences of immunogenicity of therapeutic proteins:

1. *General:* persisting levels of neutralizing antibodies
2. *Specific:* loss of activity of endogenous homolog; increase in specific disease marker; decrease in efficiency marker

Biomarkers are also used to indicate the loss of efficiency of therapeutic proteins by antibodies:

1. *Monoclonal antibodies:* increase in side effects
2. *Other therapeutic proteins:* loss of side effects

Structural properties are the main primary factors of the induction of antibodies [18]. Therapeutic proteins of nonhuman origin will induce antibodies in the majority of patients after a limited number of applications. The degree of nonself and the presence of T-cell epitopes and the relative lack of immune tolerance are predictors of the antibody response.

Human homologs are less likely to be immunogenic. The best structural predictor of breaking tolerance is the presence of aggregates. The only animal model available for this type of immunogenicity is immune-tolerant transgenic mice. Induction of antibodies to human proteins usually occurs only after prolonged exposure. Independent of whether the protein is self or nonself, the possible immune modulating effect of the therapeutic protein, concomitant immune suppressive therapy, and the immune status of the patients are important predictors of an antibody response.

Antibody formation is by definition the marker for the immunogenicity of therapeutic proteins. The role of the cellular immunity is largely unknown and may be absent in the case of breaking B-cell tolerance. The occurrence of clinical consequences is in the majority of cases associated with relative high and persisting levels of neutralizing antibodies. In some cases the occurrence of neutralizing antibodies is preceded by binding antibodies, which may interfere with the pharmacokinetics of the proteins.

As discussed extensively, skin reactions cannot be seen as signs of the immunogenicity of proteins. Often, the loss of efficacy by neutralizing antibodies is difficult to assess because the diseases involved are chronic diseases with an unpredictable development and the proteins have only a limited effect. In these cases surrogate markers for efficacy may be monitored. Also, the effect on the side effects may be used as a marker. If the side effects are caused by the pharmacodynamic effect of the protein drugs, the loss of side effects is indicative of the development of neutralizing antibodies. If the side effect is the result of immune complexes, their appearance is associated with the induction of antibodies.

## REFERENCES


1. Schellekens H (2002). Bioequivalence and the immunogenicity of biopharmaceuticals. *Nat Rev Drug Discov*, 1(6):457–462.
2. Chackerian B, Lenz P, Lowy DR, Schiller JT (2002). Determinants of autoantibody induction by conjugated papillomavirus virus-like particles. *J Immunol*, 169:6120–6126.
3. Hermeling S, Schellekens H, Maas C, Gebbink MF, Crommelin DJ, Jiskoot W (2006). Antibody response to aggregated human interferon alpha2b in wild-type and transgenic immune tolerant mice depends on type and level of aggregation. *J Pharm Sci*, 95(5):1084–1096.
4. Kessler M, Goldsmith D, Schellekens H (2006). Immunogenicity of biopharmaceuticals. *Nephrol Dial Transplant*, 21(Suppl 5):v9–v12.
5. Hochuli E (1997). Interferon immunogenicity: technical evaluation of interferon-alpha 2a. *J Interferon Cytokine Res*, 17:S15–S21.
6. Wierda D, Smith HW, Zwickl CM (2001). Immunogenicity of biopharmaceuticals in laboratory animals. *Toxicology*, 158(1–2):71–74.
7. Zwickl CM, Cocke KS, Tamura RN, et al. (1991). Comparison of the immunogenicity of recombinant and pituitary human growth hormone in rhesus monkeys. *Fundam Appl Toxicol*, 16(2):275–287.
8. Palleroni AV, et al. (1997). Interferon Immunogenicity: Preclinical Evaluation of Interferon-a2a. *J Interferon Cytokine Res*, 19(Suppl 1):s23–s27.
9. Stewart TA, Hollingshead PG, Pitts SL, et al. (1989). Transgenic mice as a model to test the immunogenicity of proteins altered by site-specific mutagenesis. *Mol Biol Med*, 6(4):275–281.
10. Casadevall N, Nataf J, Viron B, et al. (2002). Pure red-cell aplasia and antierythropoietin antibodies in patients treated with recombinant erythropoietin. *N Engl J Med*, 346:469–475.

11. Janeway CA (2005). *Immunobiology: The Immune System in Health and Disease*, 6th ed. Garland Science, New York, pp. 517–555.
12. Frost N (2005). Antibody-mediated side effects of recombinant proteins. *Toxicology*, 209(2):155–160.
13. Thielen AM, Kuenzli S, Saurat JH (2005). Cutaneous adverse events of biological therapy for psoriasis: review of the literature. *Dermatology*, 211(3):209–217.
14. Lacatoure ME (2006). Mechanisms of cutaneous toxicities to EGFR inhibitors. *Nat Rev Cancer*, 6(10):803–812.
15. Robert C, Soria JC, Spatz A, et al. (2005). Cutaneous side-effects of kinase inhibitors and blocking antibodies. *Lancet Oncol*, 6(7):491–500.
16. Francis GS, Rice GP, Alsop JC (2005). Interferon beta-1a in MS: results following development of neutralizing antibodies in PRISMS. *Neurology*, 65(1):48–55.
17. Panitch H, Goodin D, Francis G (2005). Benefits of high-dose, high-frequency interferon beta-1a in relapsing–remitting multiple sclerosis are sustained to 16 months: final comparative results of the EVIDENCE trial. *J Neurol Sci*, 239(1):67–74.
18. Hermeling S, Crommelin DJ, Schellekens H, Jiskoot W (2004). Structure–immunogenicity relationships of therapeutic proteins. *Pharm Res*, 21(6):897–903.

# NEW MARKERS OF KIDNEY INJURY

Sven A. Beushausen, Ph.D.
*Pfizer Global Research and Development, Chesterfield, Missouri*

## INTRODUCTION

The current biomarker standards for assessing acute kidney injury (AKI), caused by disease or as a consequence of drug-induced toxicity, include blood urea nitrogen (BUN) and serum creatinine (SC). Retention of either marker in the blood is indicative of impedance in the glomerular filtration rate (GFR), which if left untreated could escalate to serious kidney injury through loss of function and ultimately, death. Although the colorimetric assays developed for SC and BUN are relatively quick, seconds as compared to hours for antibody-based analysis platforms like enzyme-linked immunosorbent assay (ELISA), they are poor *predictors* of kidney injury because they both suffer from a lack of sensitivity and specificity. For example, SC concentration is greatly influenced by other nonrenal factors, including gender, age, muscle mass, race, drugs, and protein intake [1]. Consequently, increases in BUN and SC levels *report* out injury only after serious kidney damage has occurred. These shortcomings rather limit their clinical utility to patients who are at risk of developing drug-induced AKI or where instances of AKI have already been established, require frequent monitoring, and time to treatment is critical.

Renal failure or AKI is often a direct consequence of disease, can result from complications associated with disease or postsurgical trauma like sepsis,

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

**TABLE 1 Common Medications Associated with Acute Renal Injury**[a]

| Pathoetiology | Medication | Clinical Findings | Treatment |
|---|---|---|---|
| Prerenal injury | Diuretics, NSAIDs, ACE inhibitors, ciclosporin, tacrolimus, radiocontrast media, interleukin-2, vasodilators (hydralazine, calcium-channel blockers, minoxidil, diazoxide) | Benign urine sediment, FENa < 1%, UOsm > 500 | Suspend or discontinue medication, volume replacement as clinically indicated |
| Intrinsic renal injury<br>Vascular effects<br>Thrombotic microangiopathy | Ciclosporin, tacrolimus, mitomycin C, conjugated estrogens, quinine, 5-fluorouracil, ticlopidine, clopidogrel, interferon, valaciclovir, gemcitabine, bleomycin | Fever, microangiopathic, hemolytic anemia, thrombocytopenia | Discontinue medication, supportive care, plasmapheresis if indicated |
| Cholesterol emboli | Heparin, warfarin, streptokinase | Fever, microangiopathic, hemolytic anemia, thrombocytopenia | Discontinue medication, supportive care, plasmapheresis if indicated |
| Tubular toxicity | Aminoglycosides, radiocontrast media, cisplatin, nedaplatin, methoxyflurane, outdated tetracycline, amphotericin B, cephaloridine, streptozocin, tacrolimus, carbamazepine, mithramycin, quinolones, foscarnet, pentamidine, intravenous gammaglobulin, fosfamide, zoledronate, cidofovir,adefovir,tenofovir, mannitol, dextran, hydroxyethylstarch | FENa > 2%, UOsm < 350 urinary sediment with granular casts, tubular epithelial cells | Drug discontinuation, supportive care |

| | | | |
|---|---|---|---|
| Rhabdomyolysis | Lovastatin, ethanol, codeine, barbiturates, diazepam | Elevated CPK, ATN urine sediment | Drug discontinuation, supportive care |
| Severe hemolysis | Quinine, quinidine, sulfonamides, hydralazine, triamterene, nitrofurantoin, mephenytoin | High LDH, decreased hemoglobin | Drug discontinuation, supportive care |
| Immune-mediated interstitial inflammation | Penicillin, methicillin ampicillin, rifampin, sulfonamides, thiazides, cimetidine, phenytoin, allopurinol, cephalosporins, cytosine arabinoside, furosemide, interferon, NSAIDs, ciprofloxacin, clarithromycin, telithromycin, rofecoxib, pantoprazole, omeprazole, atazanavir | Fever, rash, eosinophilia, urine sediment showing pyuria, white cell casts, eosinophiluria | Discontinue medication, supportive care |
| Glomerulopathy | Gold, penicillamine, captopril, NSAIDs, lithium, mefenamate, fenoprofen, mercury, interferon-α, pamidronate, fenclofenac, tolmetin, foscarnet | Edema, moderate to severe proteinuria, red blood cells, red blood cell casts possible | Discontinue medication, supportive care |
| Obstruction Intratubular: (crystalluria and/or renal lithiasis) | Aciclovir, methotrexate, sulfanilamide, triamterene, indinavir, foscarnet, ganciclovir | Sediment can be benign with severe obstruction, ATN might be observed | Discontinue medication, supportive care |
| Ureteral (secondary to retroperitoneal fibrosis) | Methysergide, ergotamine, dihydroergotamine, methyldopa, pindolol, hydralazine, atenolol | Benign urine sediment, hydronephrosis on ultrasound | Discontinue medication, decompress ureteral obstruction by intrarenal stenting or percutaneous nephrostomy |

*Source:* Adapted from ref. 2.

[a]ACE, angiotensin-converting enzyme; ATN, acute tubular necrosis; CPK, creatinine phosphokinase; FENa, fractional excretion of sodium; LDH, lactate dehydrogenase; NSAIDs, nonsteroidal anti-inflammatory drugs; UOsm, urine osmolality.

or is produced by drug-induced nephrotoxicity. Drug-induced renal injury is of great concern to physicians. Knowledge of toxicities associated with U.S. Food and Drug Administration (FDA)-approved compounds helps to guide product selection in an effort to manage risk and maximize patient safety. Drug-induced nephrotoxicity is of even greater concern to the pharmaceutical industry, where patient safety is the principal driver in the need to discover safer and more efficacious drugs. Because BUN and SC are such insensitive predictors of early kidney injury, many instances of subtle renal damage caused by drugs may go unrecognized. Consequently, true estimates for drug-induced nephrotoxicity are likely to be far lower than previously realized. For example, studies have indicated that the incidence of acute tubular necrosis or acute interstitial nephritis due to medication has been estimated to be as high as 18.6% [2]. In addition, renal injury attributed to treatment with aminoglycosides has been reported to approach 36% [3,4]. Not surprisingly, many common drugs have been associated with renal injury that cause site-specific damage (Table 1).

Fortunately, most instances of drug-induced nephrotoxicity are reversible if discovered early and medication is discontinued. Collectively, the combined shortcomings of BUN and SC as *predictors* of nephrotoxicity and the propensity for many classes of medicines to cause drug-induced nephrotoxicity underscore the urgent need for the development and qualification of more sensitive and specific biomarkers. The benefits such tools will provide include predictive value and earlier diagnosis of drug-induced kidney injury before changes in renal function or clinical manifestations of AKI are evident. More important, biomarkers of nephrotoxicity with increased sensitivity and specificity will be invaluable to drug development both preclinically and clinically. Preclinically, new biomarkers will aid in the development of safer drugs having fewer liabilities with an ultimate goal to considerably lower or even possibly eliminate drug-induced nephrotoxicity. Clinically, the biomarkers will be used to monitor potential nephrotoxic effects due to therapeutic intervention or the potential for new drugs to cause renal toxicity in phase I to III clinical trials.

## NEW PRECLINICAL BIOMARKERS OF NEPHROTOXICIY

In recent years, two consortia led by the nonprofit organizations ILSI-HESI (International Life Sciences Institute, Health and Environmental Sciences Institute, http://www.hesiglobal.org) and C-Path (Critical Path Institute, http://www.c-path.org) aligned with leaders in academia, industry, and the FDA with a mission to evaluate the potential utility of newly identified biomarkers of nephrotoxicity for use in preclinical safety studies and to develop a process for the acceptance of the new biomarkers in support of safety data accompanying new regulatory submissions. Several criteria for the evaluation and development of new biomarkers of nephrotoxicity were considered, including:

**TABLE 2 Biomarkers of Renal Injury by Region of Specificity, Onset, Platform, and Application**

| Biomarker | Injury Related to: | Onset | Platforms | Application |
|---|---|---|---|---|
| $\beta_2$-Microglobulin | Proximal tubular injury | Early | Luminex, ELISA | Mouse, rat, human, chicken, turkey |
| Clusterin | Tubular epithelial cells | Early | Luminex, ELISA | Mouse, rat, dog, monkey, human |
| Cystatin-C | Tubular dysfunction | Late | Luminex, ELISA | Mouse, rat, human |
| GSTα | Proximal tubular injury | Early | Luminex, ELISA | Mouse, rat, human |
| GST Yb1 | Distal tubule | Early | Luminex, ELISA | |
| KIM-1 | General kidney injury and disease | Early | Luminex, ELISA | Zebrafish, mouse, rat, dog, monkey, human |
| Microalbumin | Proximal tubular injury | Early | Luminex, ELISA | Mouse, rat, dog, monkey, human |
| Osteopontin | Tubulointerstitial fibrosis | Late | Luminex, ELISA | Mouse, rat, monkey, human |
| NGAL | Proximal tubular injury | Early | Luminex, ELISA | Mouse, rat, human |
| RPA-1 | Renal papilla and collecting ducts | Early | ELISA | Rat |

*Source:* Adapted from ref. 5.

- A preference for noninvasive sample collection.
- New biomarkers developed for preclinical use should optimally be translated to the clinic.
- Assays for new biomarkers should be robust and kits readily available for testing.
- Assays should be multiplexed to minimize cost and expedite sample analysis.
- Biomarkers should ideally predict or report out site-specific injury.
- Biomarkers must be more sensitive and specific of kidney injury than existing standards.
- Biomarkers should be predictive (prodromal) of kidney injury in the absence of histopathology.

The preference for noninvasive sample collection made urine the obvious choice of biofluid. Urine has proven to be a fertile substrate for the discovery of promising new biomarkers for the early detection of nephrotoxicity [5]. A number of these markers have been selected for further development and

qualification by the ILSI-HESI and C-Path Nephrotoxicity Working Groups in both preclinical and clinical settings, with the exception of RPA-1 and GST Yb1 (Biotrin), which are markers developed specifically for the analysis of kidney effects in rats (Table 2). The utility and limitations of each marker used in the context of early and site-specific detection are discussed below.

## $\beta_2$-Microglobulin

Human $\beta_2$-microglobulin ($\beta_2$M) was isolated and characterized in 1968 [6]. $\beta_2$M was identified as a small 11,815-Da protein found on the surface of human cells expressing the major histocompatibility class I molecule [7]. $\beta_2$M is shed into the circulation as a monomer, from which it is normally filtered by the glomerulus and subsequently reabsorbed and metabolized within proximal tubular cells [8]. Twenty-five years ago, serum $\beta_2$M was advocated for use as an index of renal function because of an observed proportional increase in serum $\beta_2$M levels in response to decreased renal function [9]. It has since been abandoned due to a number of factors complicating the interpretation of the findings. More recently, increased levels of intact urinary $\beta_2$M have been directly linked to impairment of tubular uptake. Additional work in rats and humans has demonstrated that increased urinary levels of $\beta_2$M can be used as a marker for proximal tubular function when $\beta_2$M production and glomerular filtration are normal in a setting of minimal proteinuria [10–13].

Urinary $\beta_2$M has been shown to be superior to *N*-acetyl-β-glucosaminidase as a marker in predicting prognosis in idiopathic membranous neuropathy [14]. In this context $\beta_2$M can be used to monitor and avoid unnecessary immunosuppressive therapy following renal transplantation. $\beta_2$M is being considered for evaluation as an early predictor of proximal tubular injury in preclinical models of drug-induced nephrotoxicity. Although easily detected in urine, there are several factors that may limit its value as a biomarker. For example, $\beta_2$M is readily degraded by proteolytic enzymes at room temperature and also degrades rapidly in an acidic environment at or below pH < 6.0 [15]. Therefore, great care must be taken to collect urine in an ice-cold, adequately buffered environment with the addition of stabilizers to preserve $\beta_2$M levels during the period of collection and in storage. It is unlikely that $\beta_2$M will be used as a stand-alone marker to predict or report proximal tubule injury preclinically or in the clinic. Rather, it is likely to be used in conjunction with other proximal tubule markers to support such a finding. A brief survey for commercially available antibodies used to detect $\beta_2$M indicates that most are species specific (http://www.abcam.com). Instances of cross-reactivity were noted for specific reagents between human and pig, chicken and turkey, and human and other primates. A single monoclonal reagent is reported to have cross-reactivity with bovine, chicken, rabbit, and mouse, and none were listed that specifically recognized dog $\beta_2$M. Because both commonly used preclinical species rat and dog $\beta_2$M proteins share only 69.7% and 66.7% amino acid identity with the human protein (http://www.expasy.org), it would be prudent to develop and

characterize antibody reagents specific to each species and cross-reacting antisera to specific amino acid sequences shared by all three proteins.

## Clusterin

Clusterin is a highly glycosylated and sulfated secreted glycoprotein first isolated from ram rete testes fluid in 1983 [16]. It was named clusterin because of its ability to elicit clustering of Sertoli cells in vitro [17]. Clusterin is found primarily in the epithelial cells of most organs. Tissues with the highest levels of clusterin include testis, epididymus, liver, stomach, and brain. Metabolic and cell-specific functions assigned to clusterin include sperm maturation, cell transformation, complement regulation, lipid transport, secretion, apoptosis, and metastasis [18]. Clusterin is also known by a number of synonyms as a consequence of having been identified simultaneously in many parallel lines of inquiry. Names include glycoprotein III (GPIII), sulfated glycoprotein-2 (SG-2), apolipoprotein J (apo J), testosterone-repressed message-2 (TRPM-2), complement associated protein SP-40, 40, and complement cytolysis inhibitor protein (see Table 1). Clusterin has been cloned from a number of species, including the rat [19]. The human homolog is 449 amino acids in length, coding for a protein with a molecular weight of 52,495 Da [20]. However, due to extensive posttranslational modification, the protein migrates to an apparent molecular weight of 70 to 80 kDa following sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE). Amino acid identity between species is moderate. Human clusterin shares 70.3%, 76.6%, 71.7%, and 77% with the bovine, mouse, pig, and rat homologs, respectively (http://www.expasy.org).

Clusterin is a heterodimer comprised of an α and a β subunit, each having an apparent mass of 40 kDa by SDS-PAGE. The subunits result from the proteolytic cleavage of the translated polypeptide at amino acid positions 23 and 277. This eliminates the leader sequence and produces the mature 205-amino acid β subunit and the remaining 221-amino acid α subunit. The α and β subunits are held together by five sulfhydryl bonds afforded by cysteine residues clustered within each of the subunits [21]. In addition, each subunit has three N-linked carbohydrates that are also heavily sulfated, giving rise to the higher apparent molecular weight observed following SDS-PAGE. Considerable evidence has been provided suggesting that clusterin plays an important role in development. For example, clusterin mRNA expression has been observed at 12.5 days' postgestation in mice, where it is present in all germ cell layers [22]. Furthermore, stage-specific variations of the transcript have been observed, as have changes in specific localization during development. Similarly, changes in the developmental expression of clusterin in kidney, lung, and nervous system have also been reported [23]. These observations suggest that clusterin might play a role in tissue remodeling.

In the developing murine kidney, clusterin is expressed in the tubular epithelium and later in development is diminished as tubular maturation progresses [24]. Interestingly, clusterin is observed in newly formed tubules

but appears to be absent in glomeruli. Of interest to many investigators of renal function is the reemergence of clusterin observed following induction of a variety of kidney diseases and drug-induced renal injury. Clusterin induction has been observed following ureteral obstruction [25] and ischemia reperfusion injury [26]. Elevations in the levels of clusterin have also been observed in the peri-infarct region following subtotal nephrectomy [27] and in animal models of hereditary polycystic kidney disease [28]. Marked increases of clusterin released in urine have also been recorded in animal models of aminoglycoside-induced nephrotoxicity [29–31].

Authors have opined that clusterin functions in either a protective role by scavenging cell debris or may play a role in the process of tissue remodeling following cellular injury based on these observations. Collectively, the body of work linking elevated levels of urinary clusterin to kidney damage has suggested that measurement of urinary clusterin may be useful as a marker of renal tubular injury. Indeed, an early study comparing urinary levels of clusterin against *N*-acetyl-β-glucosaminidase (NAG) following chronic administration of gentamicin over a two-month period demonstrated that while the urinary levels of both proteins rose rapidly, peaked, and then declined, clusterin levels remained significantly higher than control values over the duration of the experiment. By contrast, NAG levels dropped to within control values within 10 days of treatment even though evidence of tubulointerstitial disease persisted [30]. More recent work examining the levels of urinary clusterin in the autosomal-dominant polycystic kidney disease (cy/+) rat model compared to the FHH rat model of focal segmental glomerulosclerosis following bilateral renal ischemia demonstrated that clusterin levels correlated with the severity of tubular damage and suggested use as a marker for differentiating between tubular and glomerular damage [32]. Although the value of clusterin as an early marker of tubular epithelial injury has not yet been established clinically, preclinical findings suggest that it is an ideal candidate for translation to the clinic as an early marker of nephrotoxicity.

## Cystatin-C

Cystatin C (Cys-C) is a 13-kDa nonglycosylated protein belonging to the superfamily of cysteine protease inhibitors [33]. Cys-C is produced by all nucleated cells and, unlike SC, is unaffected by muscle mass. Serum Cys-C was suggested to be closer to the "ideal" biomarker reporting GFR because although freely filtered by the glomerulus, it is not secreted. Instead, Cys-C is adsorbed by tubular epithelial cells, where it is catabolized and is not returned to the bloodstream, thus obviating the need to calculate urinary Cys-C to measure GFR [34]. Several studies have been designed to examine the usefulness of serum Cys-C as a measure or biomarker of GFR [35]. In one such study, serum Cys-C was shown to be a useful biomarker of acute renal failure and could be detected one to two days prior to the elevation in levels of SC, the accepted clinical diagnosis of AKI [36]. Although earlier in detection than

SC, serum Cys-C levels were not predictive of kidney disease and, like SC, reported out kidney injury long after serious damage had occurred. In another study, investigators monitored and compared the levels of serum Cys-C and urinary Cys-C in patients following cardiothoracic surgery with and without complicating AKI [37]. The results clearly demonstrated that while plasma Cys-C was not a useful predictor of AKI, early and persistent increases in urinary Cys-C correlated with the development and severity of AKI. Another interesting but unexplained observation in this study was that women had significantly higher postoperative levels of urinary Cys-C than did men even though preoperative Cys-C levels were similar. These data have prompted groups like ILSI-HESI and C-Path to examine the utility of urinary Cys-C as a preclinical biomarker of drug-induced renal injury in the hope that elevated levels of Cys-C can be detected in urine prior to the emergence of overt tubular dysfunction.

## Glutathione S-Transferases

The glutathione S-transferases (GSTs) form a family of homo-and heterodimeric detoxifying enzymes [38] identified originally as a group of soluble liver proteins that play a major role in the detoxification of electrophilic compounds [39]. They have since been shown to be products of gene superfamilies [40] and are classified into alpha, mu, pi, and theta subfamilies based on sequence identity and other common properties [41]. Tissue distribution and levels of GST isoform expression has been determined by immunohistochemical localization [42], isoform-specific peptide antibody Western blotting, and mass spectrometry [40]. Analysis of GST subunit diversity and tissue distribution using peptide-specific antisera has shown GST μ isoforms to be the most widely distributed class of GSTs, with expression evident in brain, pituitary, heart, lung, adrenal gland, kidney, testis, liver, and pancreas, with the highest levels of GST μ1 observed in adrenals, testis, and liver. Isoforms of the GSTα subfamily, also known by the synonyms glutathione S-transferase-1, glutathione S-transferase Ya-1, GST Ya1, ligandin, GST 1a-1a, GST B, GST 1-1, and GST A1-1 (http://www.expasy.org/uniprot/P00502), are more limited in distribution, with highest levels of expression observed in hepatocytes and proximal tubular cells of the kidney [42].

Indeed, proximal tubular GSTα levels have been reported to approximate 2% of total cytosolic protein following exposure to xenobiotics or renal toxins [43]. In the Rowe study [40], GSTα was found to be rather evenly distributed between adrenals, kidney, and pancreas, with highest levels observed in liver, whereas isoforms of the GSTπ subclass were expressed in brain, pituitary, heart, liver, kidney, and adrenals, with highest levels of expression observed in kidney. The high levels of expression and differential distribution of GST isoforms made them attractive candidates as biomarkers that could be used to indicate site-specific drug-induced nephrotoxicity. For example, development of a radioimmunoassay to quantify leakage of ligandin (GSTα)

into the urine as a measure of nephrotoxicity in the preclinical rat model was reported as early as 1979 [44]. Subsequent work described the development of a radioimmunoassay for the quantitation of GSTπ in the urine [45a] later used as an indicator of distal tubular damage in the human kidney [45b]. Additional work described the development of a multiplexed ELISA for the simultaneous quantitation of GSTα and GSTπ to discriminate between proximal and distal tubular injury, respectively [46].

In terms of sensitivity, a study examining the nephrotoxic effects of the sevoflurane degradation product, fluoromethyl-2,2-difluoro-1-(trifluoromethyl) vinyl ether, in rats showed urinary GSTα to be the most sensitive marker of mild proximal tubular damage compared to other urinary markers measured, including protein and glucose [47]. A second study in which four human volunteers were given sevoflurane demonstrated abnormalities in urinary glucose, albumin, GSTα, and GSTπ, while levels of BUN or SC were unaffected, suggesting that the GSTs were more sensitive markers of site-specific dug-induced nephrotoxicity [48]. Immunohistochemical staining of the rat kidney with antibodies to different GST isoforms has shown that GSTα subunits are expressed selectively in the proximal tubule, whereas GSTμ and π subnits are localized to the thin loop of Henle and proximal tubules, respectively [38]. An examination of the distribution of the rat GSTμ equivalent, GSTYb1, in the kidney indicates that it is localized to the distal tubules. Simultaneous measurement of urinary GSTα and GSTYb1 has been used to discriminate between drug-induced proximal and distal tubular injury (cited by Kilty et al. [49]). The high levels of GSTs in the kidney and site-specific localization of different GST classes in addition to increased sensitivity in detecting drug-induced nephrotoxicity in humans make them ideal candidates for the development and testing of preclinical markers that predict or report early signs of nephrotoxicity to support preclinical safety studies and subsequent compound development.

## Kidney Injury Molecule 1

Rat kidney injury molecule 1 (KIM-1) was discovered as part of an effort to identify genes implicated in kidney injury and repair [50] using the polymerase chain reaction (PCR) subtractive hybridization technique of representational difference analysis originally developed to look at differences in genomic DNA [51] but adapted to examine differences in mRNA expression [52]. Complementary DNA generated from poly(A+) mRNA purified from normal and 48-hour postischemic rat kidneys was amplified to generate driver and tester amplicons, respectively. The amplicons were used as templates to drive the subtractive hybridization process to generate designated differential products, three of which were ultimately gel purified and subcloned into the pUC18 cloning vector. Two of these constructs were used to screen λZapII cDNA libraries constructed from 48-hour postischemic rat kidneys. Isolation and purification of positively hybridizing plaques resulted in the recovery of a

2.5-kb clone that contained sequence information on all three designated differential products. A BLAST search of the NCBI database revealed that the rat KIM-1 sequence had limited (59.8%) amino acid homology to HAVcr-1, identified earlier as the monkey gene coding for the hepatitis A virus receptor protein [53]. The human homolog of KIM-1 was isolated by low-stringency screening of a human embryonic liver λgt10 cDNA library using the same probe that yielded the rat clones [50]. The plaque of one of two clones purified from this exercise was shown to code for a 334-amino acid protein sharing 43.8% identity and 59.1% similarity to the rat KIM-1 protein. Comparison to the human HAVcr protein [54] revealed 85.3% identity demonstrating a clear relationship between the two proteins.

Subsequent work has demonstrated that KIM-1 and HAVcr are synonyms for the same protein, also known as T-cell immunoglobulin and mucin domain–containing protein 1 (TIMD-1) and TIM-1. The TIMD proteins are all predicted to be type I membrane proteins that share a characteristic immunoglobulin V, mucin, transmembrane, and cytoplasmic domain structure [55]. It is not clear what the function of KIM-1 (TIMD-1) is, but it is believed that TIMD-1 is involved in the preferential stimulation of Th2 cells within the immune system [56]. In the rat, KIM-1 mRNA expression is highest in liver and barely detected in kidney [50]. KIM-1 mRNA and protein expression are dramatically up-regulated following ischemic injury. Immunohistochemical examination of kidney sections using a rat-specific KIM-1 antibody showed that KIM-1 is localized to regenerating proximal tubule epithelial cells. KIM-1 was proposed as a novel biomarker for human renal proximal tubule injury in a study that demonstrated that KIM-1 could be detected in the urine of patients with biopsy-proven acute tubular necrosis [57]. Human KIM-1 occurs as two splice variants that are identical with respect to the extracellular domains but differ at the carboxy termini and are differentially distributed throughout tissues [58]. Splice-variant KIM-1b is 25 amino acids longer than the originally identified KIM-1a and is found predominantly in human kidney.

Interestingly, cell lines expressing endogenous KIM-1 or recombinant KIM-1b constitutively shed KIM-1 into the culture medium, and shedding of KIM-1 could be inhibited with metalloprotease inhibitors, suggesting a mechanism for KIM-1 release into the urine following the regeneration of proximal tubule epithelial cells as a consequence of renal injury. Evidence supporting KIM-1's potential as a biomarker for general kidney injury and repair was clearly demonstrated in another paper describing the early detection of urinary KIM-1 protein in a rat model of drug-induced renal injury. In this study increases in KIM-1 were observed before significant increases in SC levels could be detected following injury with folic acid and prior to measurable levels of SC in the case of cisplatin-treated rats [59]. In later, more comprehensive studies examining the sensitivity and specificity of KIM-1 as an early biomarker of mechanically- [60] or drug-induced renal injury [61], KIM-1 was detected earlier than any of the routinely used biomarkers of renal injury,

including BUN, SC, urinary NAG, glycosuria, and proteinuria. Certainly, the weight of evidence described above supports the notion that KIM-1 is an excellent biomarker of AKI and drug-induced renal injury. The increasing availability of antibody-based reagents and platforms to rat and human KIM-1 proteins offer convenient and much needed tools for preclinical safety assessment of drug-induced renal toxicity and for aid in diagnosing or monitoring mild to severe renal injury in the clinic. Further work is required to determine if KIM-1 is a useful marker for long-term injury and whether it can be used in combination with other makers to determine site-specific kidney injury.

## Microalbumin

The examination of proteins excreted into urine provides useful information about renal function (reviewed in [62]). Tamm–Horsfall proteins that originate from renal tubular cells comprise the largest fraction of protein excreted in normal urine. The appearance of low-molecular-weight urinary proteins normally filtered through the basement membrane of the glomerulus, including insulin, parathormone, lysozyme, trypsinogen and $\beta_2$-microglobulin indicate some form of tubular damage [63]. The detection of higher-molecular-weight (40- to 150-kDa) urinary proteins not normally filtered by the glomerulus, including albumin, transferrin, IgG, caeruloplasmin, $\alpha$1-acid glycoprotein, and HDL, indicate compromised glomerular function [64]. Albumin is by far the most abundant protein constituent of proteinuria. Although gross increases in urinary albumin measured by the traditional dipstick method with a reference interval of 150 to 300 mg/mL have been used to indicate impairment of renal function, there are many instances of subclinical increases of urinary albumin within the defined reference interval that are predictive of disease [65–67].

The term *microalbuminuria* was coined to define this phenomenon, where such increases had value in predicting the onset of nephropathy in insulin-dependant diabetes mellitus [68]. The accepted reference interval defined for microalbuminuria is between 30 to 300 mg in 24 hours [69,70]. Because microalbuminuria appears to be a sensitive indicator of renal injury, there is a growing interest in the nephrotoxicity biomarker community to evaluate this marker in the context of an early biomarker predictive of drug-induced renal injury. Although microalbuminuria has traditionally been used in preclinical drug development to assess glomerular function there is growing evidence to suggest that albuminuria is a consequence of impairment of the proximal tubule retrieval pathway [71]. Evidence that microalbuminuria might provide value in diagnosing drug-induced nephrotoxicity was reported in four of 18 patients receiving cisplatin, ifosamide, and methotrextate to treat osteosarcoma [72]. Because microalbuminuria can be influenced by other factors unrelated to nephrotoxicity, including vigorous exercise, hematuria, urinary tract infection, and dehydration [5], it may have greater predictive value for renal

injury in the context of a panel of markers with increased sensitivity and site specificity. Indeed, further evaluation of microglobulin as an early biomarker of site-specific or general nephrotoxicity is required before qualification for preclinical and clinical use.

## Osteopontin

Osteopontin (OPN) is a 44-kDa highly phosphorylated secreted glycoprotein originally isolated from bone [73]. It is an extremely acidic protein with an isoelectric point of 4.37 (http://www.expasy.org/uniprot/P10451), made even more acidic through phosphorylation on a cluster of up to 28 serine residues [74]. Osteopontin is widely distributed among different tissues, including kidney, lung, liver, bladder, pancreas, and breast [75] as well as macrophages [76], activated T-cells [77], smooth muscle cells [78], and endothelial cells [79]. Evidence has been provided demonstrating that OPN functions as a calcium oxalate crystal formation inhibitor in cultured murine kidney cortical cells [80]. Immunohistochemical and in situ hybridization examination of the expression and distribution of OPN protein and mRNA in the rat kidney clearly demonstrated that levels are highest in the descending thin loop of Henle and cells of the papillary surface epithelium [81]. Uroprontin, first described as a relative of OPN was among the first examples of OPN isolation from human urine [82].

Although normally expressed in kidney, OPN expression can be induced under a variety of experimental pathologic conditions [83,84], including tubulointerstitial nephritis [85], cyclosporine-induced neuropathy [86], hydronephrosis as a consequence of unilateral ureteral ligation [87], renal ischemia [88], nephropathy induced by cisplatin, and crescentric glomeulonephritis [89]. Up-regulation of OPN has been reported in a number of animal models of renal injury, including drug-induced nephrotoxicity by puromycin, cylcoaporine, strptozotocin, phenylephrine, and gentamicin (reviewed in [90a]). In the rat, gentamicin-induced acute tubular necrosis model OPN levels were highest in regenerating proximal and distal tubules, leading the authors to conclude that OPN is related to the proliferation and regeneration of tubular epithelial cells following tubular damage [90b]. Although osteopontin has been proposed as a selective biomarker of breast cancer [91] and a useful clinical biomarker for the diagnosis of colon cancer [92], OPN shows great promise and requires further evaluation as a clinical biomarker for renal injury. Certainly, the high levels of OPN expression following chemically or physically induced renal damage coupled with the recent availability of antibody-based reagents to examine the levels of mouse, rat, and human urinary OPN provide ample opportunity to evaluate OPN as an early marker of AKI in the clinic and a predictive marker of drug-induced nephrotoxicity preclinically. Further planned work by the ILSI-HESI and C-Path groups hope to broaden our understanding regarding the utility of OPN in either capacity as an early predictor of renal injury.

## Neutrophil Gelatinase–Associated Lipocalin

Neutrophil gelatinase–associated lipocalin (NGAL) was first identified as the small molecular-weight glycoprotein component of human gelatinase affinity purified from the supernatant of phorbol myristate acetate stimulated human neutrophils. Human gelatinase purifies as a 135-kDa complex comprised of the 92-kDa gelatinase protein and the smaller 25-kDa NGAL [93]. NGAL has subsequently been shown to exist primarily in monomeric or dimeric form free of gelatinase. A BLAST search of the 178-amino acid NGAL protein yielded a high degree of similarity to the rat α2 microglobulin-related protein and mouse protein 24p3, suggesting that NGAL is a member of the lipocalin family. Lipocalins are characterized by the ability to bind small lipophilic substances and are thought to function as modulators of inflammation [94]. More recent work has shown that NGAL, also known as siderocalin, complexes with iron and iron-binding protein to promote or accelerate recovery from proximal tubular damage (reviewed in [95]). RNA dot blot analysis of 50 human tissues revealed that NGAL expression is highest in trachea and bone tissue, moderately expressed in stomach and lung with low levels of transcript expression in the remaining tissues examined, including kidney [94]. Because NGAL is a reasonably stable small-molecular-weight protein, it is readily excreted from the kidney and can be detected in urine.

NGAL was first proposed as a novel urinary biomarker for the early prediction of acute renal injury in rat and mouse models of acute renal failure induced by bilateral ischemia [96]. Increases in the levels of urinary NGAL were detected in the first hour of postischemic urine collection and shown to be related to dose and length of exposure to ischemia. In this study the authors reported NGAL to be more sensitive than either NAG or $\beta_2$M, underscoring its usefulness as an early predictor of acute renal injury. Furthermore, the authors proposed NGAL to be an earlier marker predictive of acute renal injury than KIM-1, since the latter reports injury within 24 hours of renal injury compared to 1 hour for NGAL. Marked up-regulation of NGAL expression was observed in proximal tubule cells within 3 hours of ischemia-induced damage, suggesting that NGAL might be involved in postdamage reepithelialization. Additional work demonstrated that NGAL expression was induced following mild ischemia in cultured human proximal tubule cells. This paper also addressed the utility of NGAL as an early predictor of drug-induced renal injury by detecting increased levels of NGAL in the urine of cisplatin-treated mice.

Adaptation of the NGAL assay to address utility and relevance in a clinical setting showed that both urinary and serum levels of NGAL were sensitive, specific, and highly predictive biomarkers of acute renal injury following cardiac surgery in children [97]. In this particular study, multivariate analysis showed urinary NGAL to be the most powerful predictor in children that developed acute renal injury. Measurable increases in urinary NGAL concentrations were recorded within 2 hours of cardiac bypass surgery, whereas

increases in SC levels were not observed until 1 to 3 days postsurgery. Other examples demonstrating the value of NGAL as a predictive biomarker of early renal injury include association of NGAL with severity of renal disease in proteineuric patients [98] and NGAL as an early predictor of renal disease resulting from contrast-induced nephropathy [99]. NGAL has been one of the most thoroughly studied new biomarkers predictive of AKI as a consequence of disease or surgical intervention, and to a lesser extent, drug-induced renal injury.

Sensitive and reliable antibody-based kits have been developed for a number of platforms in both humans and rodents (Table 2) and there is considerable interest in examining both the specificity and sensitivity of NGAL for acceptance as a fit-for-purpose predictive biomarker of drug-induced renal injury to support regulatory submissions. Certainly, because NGAL is such an early marker of renal injury, it will have to be assessed as a stand-alone marker of renal injury as well as in the context of a larger panel of markers that may help define site specific and degree of kidney injury.

## Renal Papillary Antigen 1

Renal papillary antigen 1 (RPA-1) is an uncharacterized antigen that is highly expressed in the collecting ducts of the rat papilla and can be detected at high levels in rat urine following exposure to compounds that induce renal papillary necrosis [100]. RPA-1 was identified by an $IgG_1$ monoclonal antibody, designated Pap X 5C10, that was generated in mice immunized with pooled fractions of a cleared lysate of homogenized rat papillary tissue following crude DEAE anion-exchange chromatography. Immunohistochemical analysis of rat papillae shows that RPA-I is localized to the epithelial cells lining the luminal side of the collecting ducts and to a lesser extent in cortical collecting ducts. A second publication described the adaptation of three rat papilla-specific monoclonal antibodies, including Pap X 5C10 (PapA1), to an ELISA assay to examine antigen excretion in rat urine following drug-induced injury to the papillae using 2-bromoethanamine, propyleneimine, indomethicin, or ipsapirone [101]. Of the three antibodies evaluated, PapA1 was the only antigen released into the urine of rats following exposure to each of the toxicants. The authors concluded that changes in the rat renal papilla caused by xenobiotics could be detected early by urinary analysis and monitored during follow-up studies. This study also clearly demonstrated that the Pap X 5C10, PapA1, RPA-1 antigen had the potential for use as a site-specific biomarker predictive of renal papillary necrosis. Indeed, the Pap X 5C10 monoclonal antibody was adapted for commercial use as an RPA-1 ELISA kit marketed specifically to predict or monitor site-specific renal injury in the rat [49].

The specificity and sensitivity of the rat reagent has generated a great deal of interest in developing an equivalent reagent for the detection of human papillary injury. Identification of the RPA-1 antigen remains elusive. Early biochemical characterization of the antigen identified it as a large-

molecular-weight protein (150 to 200 kDa) that could be separated into two molecular-weight species with isoelectric points of 7.2 and 7.3, respectively [100]. However, purification and subsequent protein identification of the antigen were extremely challenging. A recent attempt at the biochemical purification and identification of the RPA-1 antigen has been equally frustrating, with investigators providing some evidence that the antigen may be a large glycoprotein and suggesting that the carbohydrate moiety is the specific epitope recognized by the Pap X 5C10 monoclonal antibody [102]. This would be consistent with, and help to explain why, the rat reagent does not cross-react with a human antigen in the collecting ducts, as protein glycosylation of related proteins often differs dramatically between species, thereby precluding the likelihood of presenting identical epitopes. Nevertheless, continued efforts toward identifying a human RPA-1 antigen will provide investigators with a sorely needed clinical marker for the early detection of drug-induced renal papillary injury.

## SUMMARY

A considerable amount of effort has gone into identifying, characterizing, and developing new biomarkers of renal toxicity having greater sensitivity and specificity than the traditional markers, BUN and SC. The issue of sensitivity is a critical one, as the ideal biomarker would detect renal injury before damage is clinically evident or cannot be reversed. Such prodromal biomarkers would provide great predictive value to the pharmaceutical industry in preclinical drug development, where compounds could be modified or development terminated early in response to the nephrotoxicity observed. Even greater value could be realized in the clinic, where early signs of kidney injury resulting from surgical or therapeutic intervention could be addressed immediately before serious damage to the patient has occurred.

Several of the candidate markers described above, including $\beta_2$M, GST$\alpha$, microalbumin, KIM-1, and NGAL, have demonstrated great promise as early predictors of nephrotoxicity. Continued investigation should provide ample data from which to make a determination regarding the utility of these markers for preclinical and, ultimately, clinical use. The issue of biomarker specificity is also of great value because it provides information regarding where and to what extent injury is occurring. For example, increases in levels of SC and BUN inform us that serious kidney injury has occurred but does not reveal the precise nature of that injury, whereas the appearance of increased levels of $\beta_2$M, GST$\alpha$, microalbumin, and NGAL indicate some degree of proximal tubule injury. Similarly, RPA-1 reports injury to the papilla, clusterin indicates damage to tubular epithelial cells, and GSTYb1 is specific to distal tubular damage. Low-level increases of early markers warn the investigator or clinician that subclinical damage to the kidney is occurring and provides the necessary time to alter or terminate a course of treatment or development.

Monitoring toxicity is an important aspect of achieving a positive clinical outcome and increased safety in drug development. Incorporation of many or all of these markers into a panel tied to an appropriate platform allows for the simultaneous assessment of site-specific kidney injury with some understanding of the degree of damage. Several detection kits are commercially available for many of these new biomarkers of nephrotoxicity. For example, Biotrin International provides ELISA kits for the analysis of urinary GSTs and RPA-1, while Rules Based Medicine and Meso Scale Discovery offer panels of kidney biomarkers multiplexed onto antibody-based fluorescence or chemiluminescent platforms, respectively. As interest in new biomarkers of kidney injury continues to develop, so will the technology that supports them. Presently, all of the commercial reagents and kits supporting kidney biomarker detection are antibody-based. The greatest single limitation of such platforms is how well the reagents perform with respect to target identification, nonspecific protein interaction, and species cross reactivity. Although kits are standardized and come complete with internal controls, kit-to-kit and lab-to-lab variability can be high.

Another technology being developed for the purpose of quantifying biomarkers in complex mixtures such as biofluids is mass spectrometry–based multiple reaction monitoring. This technology requires the synthesis and qualification of small peptides specific to a protein biomarker that can be included in a sample as an internal standard to which endogenous peptide can be compared and quantified. This platform is extremely sensitive (femtomolar detection sensitivity), requires very little sample volume, and offers the highest degree of specificity with very short analysis times. Limitations of the platform are related to the selection of an appropriate peptide and the expense of assay development and qualification for use. For example, peptides need to be designed that are isoform specific, being able to discriminate between two similar but not identical proteins. Peptide design is also somewhat empirical with respect to finding peptides that will "fly" in the instrument and produce a robust signal at the detector. The choice of peptides available in a particular target may be limiting given these design restrictions. Consequently, not all proteins may be amenable to this approach.

In conclusion, continued improvement in technology platforms combined with the availability of reagents to detect new biomarkers of nephrotoxicity provides both the clinician and the investigator with a variety of tools to predict and monitor early or acute kidney injury. This will be of tremendous value toward saving lives in the clinic and developing safer, more efficacious drugs without nephrotoxic side effects.

## REFERENCES

1. Bjornsson TD (1979). Use of serum creatinine concentrations to determine renal function. *Clin Pharmacokinet*, 4:200–222.

2. Choudhury D, Ziauddin A (2005). Drug-associated renal dysfunction and injury. *Nat Clin Pract Nephrol*, 2:80–91.
3. Kleinknecht D, Landais P, Goldfarb B (1987). Drug-associated renal failure: a prospective collaborative study of 81 biopsied patients. *Adv Exp Med Biol*, 212:125–128.
4. Kaloyandes GJ, et al. (2001). *Antibiotic and Renal Immunosuppression-Related Renal Failure*. Lippincott Williams & Wilkins, Philadephia.
5. Vaidya VS, Ferguson MA, Bonventre JV (2008). Biomarkers of acute kidney injury. *Annu Rev Pharmacol Toxicol*, 48:463–493.
6. Berggard I, Bearn AG (1968). Isolation and properties of a low molecular weight β2-globulin occurring in human biological fluid. *J Biol Chem*, 213: 4095–4103.
7. Harris HW, Gill TJ III (1986). Expression of class I transplantation antigens. *Transplantation*, 42:109–117.
8. Bernier GM, Conrad ME (1969). Catabolism of β2-microglobulin by the rat kidney. *Am J Physiol*, 217:1350–1362.
9. Shea PH, Mahler JF, Horak E (1981). Prediction of glomerular filtration rate by serum creatinine and β2-microglobulin. *Nephron*, 29:30–35.
10. Eddy AA, McCullich L, Liu E, Adams J (1991). A relationship between proteinuria and acute tubulointerstitial disease in rats with experimental nephritic syndrome. *Am J Pathol*, 138:1111–1123.
11. Holm J, Hemmingsen L, Nielsen NV (1993). Low-molecular-mass proteinuria as a marker of proximal renal tubular dysfunction in normo- and microalbuminuric non-insulin-dependent subjects. *Clin Chem*, 39:517–519.
12. Kabanda A, Jadoul M, Lauwerys R, Bernard A, van Ypersele de Strihou C (1995). Low molecular weight proteinuria in Chinese herbs nephropathy. *Kidney Int*, 48:1571–1576.
13. Kabanda A, Vandercam B, Bernard A, Lauwerys R, van Ypersele de Strihou C (1996). Low molecular weight proteinuria in human imminodeficiency virus–infected patients. *Am J Kidney Dis*, 27:803–808.
14. Hofstra JM, Deegans JK, Willems HL, Wetzels FM (2008). Beta-2-microglobulin is superior to *N*-acetyl-beta-glucosaminindase in predicting prognosis in idiopathic membranous neuropathy. *Nephrol Dial Transplant*, 23:2546–2551.
15. Davey PG, Gosling P (1982). Beta-2-microglobulin instability in pathological urine. *Clin Chem*, 28:1330–1333.
16. Blashuck O, Burdzy K, Fritz IB (1983). Purification and characterization of cell-aggregating factor (clusterin), the major glycoproteoin in ram rete testis fluid. *J Biol Chem*, 12:7714–7720.
17. Fritz IB, Burdzy K, Setchell B, Blashuck O (1983). Ram rete testes fluid contains a protein (clusterin) which influences cell–cell interactions in vitro. *Biol Reprod*, 28:1173–1188.
18. Rosenberg ME, Silkensen J (1995). Clusterin: physiologic and pathophysiologic considerations. *Int J Biochem Cell Biol*, 27:633–645.
19. Collard MW, Griswold MD (1987). Biosynthesis and molecular cloning of sulfated glycoprotein 2 secreted by rat Sertoli cells. *Biochemistry*, 26:3297–3303.

20. Kirszbaum L, Sharpe JA, Murphy B, et al. (1989). Molecular cloning and characterization of the novel, human complement-associated protein, SP40,40: a link between the complement and reproductive systems. *EMBO J*, 8:711–718.
21. Kirszbaum L, Bozas SE, Walker ID (1992). SP-40,40, a protein involved in the control of the complement pathway, possesses a unique array of disulfide bridges. *FEBS Lett*, 297:70–76.
22. French LE, Chonn A, Ducrest D, et al. (1993). Murine clusterin: molecular cloning and mRNA localization of a gene associated with epithelial differentiation processes during embryogenesis. *J Cell Biol*, 122:1119–1130.
23. O'Bryan MK, Cheema SS, Bartlett PF, Murphy BF, Pearse MJ (1993). Clusterin levels increase during neuronal development. *J Neurobiol I*, 24:6617–6623.
24. Harding MA, Chadwick LJ, Gattone VH II, Calvet JP (1991). The SGP-2 gene is developmentally regulated in the mouse kidney and abnormally expressed in collecting duct cysts in polycystic kidney disease. *Dev Biol*, 146:483–490.
25. Pearse MJ, O'Bryan M, Fisicaro N, Rogers L, Murphy B, d'Apice AJ (1992). Differential expression of clusterin in inducible models of apoptosis. *Int Immunol*, 4:1225–1231.
26. Witzgall R, Brown D, Schwarz C, Bonventre JV (1994). Localization of proliferating cell nuclear antigen, vimentin, c-Fos, and clusterin in the post-ischemic kidney: evidence for a heterogeneous genetic response among nephron segments, and a large pool of mitotically active and dedifferentiated cells. *J Clin Invest*, 93:2175–2188.
27. Correa-Rotter R, Hostetter TM, Manivel JC, Eddy AA, Rosenberg ME (1992). Intrarenal distribution of clusterin following reduction of renal mass. *Kidney Int*, 41:938–950.
28. Cowley BD Jr, Rupp JC (1995). Abnormal expression of epidermal growth factor and sulfated glycoprotein SGP-2 messenger RNA in a rat model of autosomal dominant polycystic kidney disease. *J Am Soc Nephrol*, 6:1679–1681.
29. Aulitzky WK, Schlegel PN, Wu D, et al. (1992). Measurement of urinary clusterin as an index of nephrotoxicity. *Proc Soc Exp Biol Med*, 199:93–96.
30. Eti S, Cheng SY, Marshall A, Reidenberg MM (1993). Urinary clusterin in chronic nephrotoxicity in the rat. *Proc Soc Exp Biol Med*, 202:487–490.
31. Rosenberg ME, Silkensen J (1995). Clusterin and the kidney. *Exp Nephrol*, 3:9–14.
32. Hidaka S, Kranzlin B, Gretz N, Witzgall R (2002). Urinary clusterin levels in the rat correlate with the severity of tubular damage and may help to differentiate between glomerular and tubular injuries. *Cell Tissue Res*, 310:289–296.
33. Abrahamson M, Olafsson I, Palsdottir A, et al. (1990). Structure and expression of the human cystatin C gene. *Biochem J*, 268:287–294.
34. Grubb A (1992). Diagnostic value of cystatin C and protein HC in biological fluids. *Clin Nephrol*, 38:S20–S27.
35. Laterza O, Price CP, Scott M (2002). Cystatin C: an improved estimator of glomerular function? *Clin Chem*, 48:699–707.
36. Herget-Rosenthal S, Marggraf G, Husing J, et al. (2004). Early detection of acute renal failure by serum cystatin C. *Kidney Int*, 66:1115–1122.

37. Koyner JL, Bennett MR, Worcester EM, et al. (2008). Urinary cystatin C as an early biomarker of acute kidney injury following adult cardiothoracic surgery. *Kidney Int*, 74:1059–1069.
38. Rozell B, Hansson H-A, Guthenberg M, Tahir G, Mannervik B (1993). Glutathione transferases of classes α, μ and π show selective expression in different regions of rat kidney. *Xenobiotica*, 23:835–849.
39. Smith GJ, Ohl VS, Litwack G (1977). Ligandin, the glutathione S-transferases, and chemically induced hepatocarcinogenesis: a review. *Cancer Res*, 37:8–14.
40. Rowe JD, Nieves E, Listowsky I (1997). Subunit diversity and tissue distribution of human glutathione S-transferases: interpretations based on electospray ionization-MS and peptide sequence–specific antisera. *Biochem J*, 325:481–486.
41. Mannervik B, Awasthi YC, Board PG, et al. (1992). Nomenclature for human glutathione transferases. *Biochem J*, 282:305–306.
42. Sundberg AG, Nilsson R, Appelkvist EL, Dallner G (1993). Immunohistochemical localization of alpha and pi class glutathione transferases in normal human tissues. *Pharmacol Toxicol*, 72:321–331.
43. Beckett GJ, Hayes JD (1993). Glutathione S-transferases: biomedical applications. *Adv Clin Chem*, 30:281–380.
44. Bass NM, Kirsch RE, Tuff SA, Campbell JA, Saunders JS (1979). Radioimmunoassay measurement of urinary ligandin excretion in nephrotoxin-treated rats. *Clin Sci*, 56:419–426.
45a. Sundberg AG, Appelkvist EL, Backman L, Dallner G (1994). Quantitation of glutathione transferase-pi in the urine by radioimmunoassay. *Nephron*, 66: 162–169.
45b. Sundberg AG, Appelkvist EL, Backman L, Dallner G (1994). Urinary pi-class glutathione transferase as an indicator of tubular damage in the human kidney. *Nephron*, 67:308–316.
46. Sundberg AG, Nilsson R, Appelkvist EL, Dallner G (1995). ELISA procedures for the quantitation of glutathione transferases in the urine. *Kidney Int*, 48: 570–575.
47. Kharasch ED, Thorning D, Garton K, Hankins DC, Kilty CG (1997). Role of renal cysteine conjugate b-lyase in the mechanism of compound A nephrotoxicity in rats. *Anesthesiology*, 86:160–171.
48. Eger EI II, Koblin DD, Bowland T, et al. (1997). Nephrotoxicity of sevofluorane versus desulfrane anesthesia in volunteers. *Anesth Analg*, 84:160–168.
49. Kilty CG, Keenan J, Shaw M (2007). Histologically defined biomarkers in toxicology. *Expert Opin Drug Saf*, 6:207–215.
50. Ichimura T, Bonventre JV, Bailly V, et al. (1998). Kidney injury molecule-1 (KIM-1), a putative epithelial cell adhesion molecule containing a novel immunoglobulin domain, is up-regulated in renal cells after injury. *J Biol Chem*, 273:4135–4142.
51. Lisitsyn N, Lisitsyn N, Wigler M (1993). Cloning the differences between two complex genomes. *Science*, 259:946–951.
52. Hubank M, Schatz DG (1994). Identifying differences in mRNA expression by representational difference analysis of cDNA. *Nucleic Acids Res*, 22:5640–5648.

53. Kaplan G, Totsuka A, Thompson P, Akatsuka T, Moritsugu Y, Feinstone SM (1996). Identification of a surface glycoprotein on African green monkey kidney cells as a receptor for hepatitis A virus. *EMBO J*, 15:4282–4296.
54. Feigelstock D, Thompson P, Mattoo P, Zhang Y, Kaplan GG (1998). The human homolog of HAVcr-1 codes for a hepatitis A virus cellular receptor. *J Virol*, 72:6621–6628.
55. Kuchroo VK, Umetsu DT, DeKruyff RH, Freeman GJ (2003). The TIM gene family: emerging roles in immunity and disease. *Nat Rev Immunol*, 3:454–462.
56. Meyers JH, Chakravarti S, Schlesinger D, et al. (2005). TIM-4 is the ligand for TIM-1 and the TIM-1-TIM-4 interaction regulates T cell proliferation. *Nat Immunol*, 6:455–464.
57. Won KH, Bailly V, Aabichandani R, Thadhani R, Bonventre JV (2002). Kidney injury molecule-1(KIM-1): a novel biomarker for human renal proximal tubule injury. *Kidney Int*, 62:237–244.
58. Bailly V, Zhang Z, Meier W, Cate R, Sanicola M, Bonventre JV (2002). Shedding of kidney injury molecule-1, a putative adhesion protein involved in renal regeneration. *J Biol Chem*, 277:39739–39748.
59. Ichimura T, Hung CC, Yang SA, Stevens JL, Bonventre JV (2004). Kidney injury molecule-1: a tissue and urinary biomarker for nephrotoxicant-induced renal injury. *Am J Renal Physiol*, 286:F552–F563.
60. Vaidya VS, Ramirez V, Ichimura T, Bobadilla NA, Bonventre JV (2006). Urinary kidney injury molecule-1: a sensitive quantitative biomarker for early detection of kidney tubular injury. *Am J Renal Physiol*, 290:F517–F529.
61. Zhou Y, Vaidya VS, Brown RP, et al. (2008). Comparison of kidney injury molecule-1 and other nephrotoxicity biomarkers in urine and kidney following acute exposure to gentamicin, mercury and chromium. *Toxicol Sci*, 101:159–170.
62. Lydakis C, Lip GYH (1998). Microalbuminuria and cardiovascular risk. *Q J Med*, 91:381–391.
63. Kaysen JA, Myers BD, Cowser DG, Rabkin R, Felts JM (1985). Mechanisms and consequences of proteinuria. *Lab Invest*, 54:479–498.
64. Noth R, Krolweski A, Kaysen G, Meyer T, Schambelan M (1989). Diabetic nephropathy: hemodynamic basis and implications for disease management. *Ann Intern Med*, 110:795–813.
65. Viberti GC (1989). Recent advances in understanding mechanisms and natural history of diabetic disease. *Diabetes Care*, 11:3–9.
66. Morgensen CE (1987). Micoalbuminuria as a predictor of clinical diabetic nephropathy. *Kidney Int*, 31:673–689.
67. Parving HH, Hommel E, Mathiesen E, et al. (1988). Prevalence of microalbuminuria, arterial hypertension, retinopathy, neuropathy, in patients with insulin-dependent diabetes. *Br Med J*, 296:156–160.
68. Viberti GC, Hill RD, Jarrett RJ, Argyropoulos A, Mahmud U, Keen H (1982). Microalbuminuria as a predictor of clinical nephropathy in insulin-dependent diabetes mellitus. *Lancet*, 319:1430–1432.
69. Morgensen CK, Schmitz O (1988) The diabetic kidney: from hyperfiltration and microalbuminuria to end-stage renal failure. *Med Clin North Am*, 72:466–492.

70. Rowe DJF, Dawnay A, Watts GF (1990). Microalbuminuria in diabetes mellitus: review and recommendations for measurement of albumin in urine. *Ann Clin Biochem*, 27:297–312.
71. Russo LM, Sandoval RM, McKee M, et al. (2007). The normal kidney filters nephritic levels of albumin retrieved by proximal tubule cells: retrieval is disrupted in nephritic states. *Kidney Int*, 71:504–513.
72. Koch Nogueira PC, Hadj-Assa A, Schell M, Dubourg L, Brunat-Metigny M, Cochat P (1998). Long-term nephrotoxicity of cisplatin, ifosamide, and methotrexate in osteosarcoma. *Pediatr Nephrol*, 12:572–575.
73. Prince CW, Oosawa T, Butler WT, et al. (1987). *J Biol Chem*, 262:2900–2907.
74. Sorensen ES, Hojrup P, Petersen TE (1995). Posttranslational modifications of bovine osteopontin: identification of twenty-eight phophorylation and three O-glycosylation sites. *Protein Sci*, 4:2040–2049.
75. Brown LF, Berse B, Van de Water L, et al. (1982). Expression and distribution of osteopontin in human tissues: widespread association with luminal epithelial surfaces. *Mol Cell Biol*, 3:1169–1180.
76. Singh PR, Patarca R, Schwartz J, Singh P, Cantor H (1990). Definition of a specific interaction between the early T lymphocyte activation 1 (Eta-1) protein and murine macrophages in vitro and its effect upon macrophages in vivo. *J Exp Med*, 171:1931–1942.
77. Weber GF, Cantor H. (1996). The immunology of eta-1/osteopontin. *Cytokine Growth Factor Rev*, 7:241–248.
78. Giachelli C, Bae N, Lombardi D, Majesky M, Schwartz S (1991). The molecular cloning and characterization of 2B7, a rat mRNA which distinguishes smooth muscle cell phenotypes in vitro and is identical to osteopontin (secreted phosphoprotein I, 2a). *Biochem Biophys Res Commun*, 177:867–873.
79. Liaw L, Lindner V, Schwartz SM, Chambers AF, Giachelli CM (1995). Osteopontin and beta 3 integrin are coordinately expressed in regenerating endothelium in vivo and stimulate ARG-GLY-ASP-dependent endothelial migration in vitro. *Circ Res*, 77:665–672.
80. Worcester EM, Blumenthal SS, Beshensky AM, Lewand DL (1992). The calcium oxalate crystal growth inhibitor protein produced by mouse kidney cortical cells in culture is osteopontin. *J Bone Miner Res*, 7:1029–1036.
81. Kleinman JG, Beshenky A, Worcester EM, Brown D (1995). Expression of osteopontin, a urinary inhibitor of stone mineral crystal growth, in rat kidney. *Kidney Int*, 47:1585–1596.
82. Shiraga H, Min W, Vandusen WJ, et al. (1992). Inhibition of calcium oxalate growth in vitro by uropontin: another member of the aspartic acid-rich protein superfamily. *Proc Natl Acad Sci USA*, 89:426–430.
83. Wuthrich RP (1998). The complex role of osteopontin in renal disease. *Nephrol Dial Transplant*, 13:2448–2450.
84. Rittling SR, Denhardt DT (1999). Osteopontin function in pathology: lessons from osteopontin-deficient mice. *Exp Nephrol*, 7:103–113.
85. Giachelli CM, Pichler R, Lombardi D, et al. (1994). Osteopontin expression in angiotensin II–induced tubulointerstitial nephritis. *Kidney Int*, 45:515–524.

86. Pichler RH, Franseschini N, Young BA, et al. (1995). Pathogenesis of cyclosporine nephropathy: roles of angiotensin II and osteopontin. *J Am Soc Nephrol*, 6:1186–1196.

87. Diamond JR, Kees-Folts D, Ricardo SD, Pruznak A, Eufemio M (1995). Early and persistent up-regulated expression of renal cortical osteopontin in experimental hydronephrosis. *Am J Pathol*, 146:1455–1466.

88. Kleinman JG, Worcester EM, Beshensky AM, Sheridan AM, Bonventre JV, Brown D (1995). Upregulation of osteopontin expression by ischemia in rat kidney. *Ann NY Acad Sci*, 760:321–323.

89. Yu XQ, Nikolic-Paterson DJ, Mu W, et al. (1998). A functional role for osteopontin expression in experimental crescentic glomerulonephritis in the rat. *Proc Assoc Am Physicians*, 110:50–64.

90a. Xie Y, Sakatsume M, Nishi S, Narita I, Arakawa M, Gejyo F (2001). Expression, roles, receptor, and regulation of osteopontin in the kidney. *Kidney Int*, 60:1645–1657.

90b. Xie Y, Nishi S, Iguchi S, et al. (2001). Expression of osteopontin in gentamicin-induced acute tubular necrosis and its recovery process. *Kidney Int*, 59: 959–974.

91. Mirza M, Shaunessy E, Hurley JK, et al. (2008). Osteopontin-C is a selective marker for breast cancer. *Int J Cancer*, 122:889–897.

92. Agrawal D, Chen T, Irby R, et al. (2002). Osteopontin identified as lead marker of colon cancer progression, using pooled sample expression profiling. *J Natl Cancer Inst*, 94:513–521.

93. Kjeldsen L, Johnson AH, Sengelov H, Borregaard N (1993). Isolation and primary structure of NGAL, a novel protein associated with human neutrophil gelatinase. *J Biol Chem*, 268:10425–10432.

94. Cowland JB, Borregaard N (1997). Molecular characterization and pattern of tissue expression of the gene for neutrophil gelatinase–associated lipocalin from humans. *Genomics*, 45:17–23.

95. De Broe M (2006). Neutrophil gelatinase–associated lipocalin in acute renal failure. *Kidney Int*, 69:647–648.

96. Mishra J, Ma Q, Prada A, et al. (2003). Identification of neutrophil gelatinase–associated protein as a novel early urinary biomarker of renal ischemic injury. *J Am Soc Nephrol*, 14:2534–2543.

97. Mishra J, Dent C, Tarabishi R, et al. (2005). Neutrophil gelatinase–associated lipocalin (NGAL) as a biomarker for acute renal injury after cardiac surgery. *Lancet*, 365:1231–1238.

98. Bolignano D, Coppolino G, Campo S, et al. (2007). Urinary neutrophil gelatinase–associated lipocalin (NGAL) is associated with severity of renal disease in proteinuric patients. *Nephrol Dial Transplant*, 23:414–416.

99. Hirsch R, Dent C, Pfriem H, et al. (2007). HNGAL as an early predictive biomarker of contrast-induced nephropathy in children. *Pediatr Nephrol*, 22: 2089–2095.

100. Falkenberg FW, Hildebrand H, Lutte L, et al. (1996). Urinary antigens as markers of papillary toxicity: I. Identification and characterization of rat kidney papillary antigens with monoclonal antibodies. *Arch Toxicol*, 71:80–92.

101. Hildebrand H, Rinke M, Schluter G, Bomhard E, Falkenberg FW (1999). Urinary antigens as markers of papillary toxicity: II. Application of monoclonal antibodies for the determination of papillary antigens in rat urine. *Arch Toxicol*, 73:233–245.
102. Price S, Betton G. Personal communication, unpublished data.

# TRANSLATIONAL MEDICINE—A PARADIGM SHIFT IN MODERN DRUG DISCOVERY AND DEVELOPMENT: THE ROLE OF BIOMARKERS

Giora Z. Feuerstein, M.D., Salvatore Alesci, M.D., Ph.D., Frank L. Walsh, Ph.D., J. Lynn Rutkowski, Ph.D., and Robert R. Ruffolo, Jr., Ph.D.

*Wyeth Research, Collegeville, Pennsylvania*

## DRUG TARGETS: HISTORICAL PERSPECTIVES

Drugs are natural or designed substances used deliberately to produce pharmacological effects in humans or animals. Drugs have been part of human civilizations for millennia. However, until the very recent modern era, drugs have been introduced to humans by empiricism and largely by serendipitous events such as encounters with natural products in search of food or by avoiding hazardous plants and animal products. The emergence of the scientific era in drug discovery evolved alongside the emergence of physical and chemical sciences at large, first as knowledge to distill, isolate, and enrich the desired substance from its natural environment, followed by deliberate attempts to modify natural substances to better serve the human needs and desires.

Scientific evolution throughout the past two centuries enabled identification of biologically active substances in humans (e.g., hormones) which were

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

manipulated chemically to improve (potency, duration of action, and exposure), or to mitigate or abrogate undesirable actions. The cumulative knowledge of human, animal, and plant biology and chemistry provided the scientific foundation and technical capabilities to alter natural substances purposely in order to improve them. Such evolution marked the era of *forward pharmacology*. The era of forward pharmacology is about drug design that emanates from primary knowledge of the action of the biological target that has clear biological action.

The exponential progress in molecular biology since the mid-twentieth century, culminating in deciphering the complete human genome in the year 2000, brought the dawn of *pharmacogenomics* and the *reverse pharmacology* era. The reverse pharmacology era is defined by the need, first, to clarify the biology and medical perspectives of the target so as to qualify it as a drugable and pharmaceutically exploitable for drug discovery and development scheme. The pharmacogenomic era provides vast opportunities for selection of new molecular targets from a gamut of approximately 30,000 primary genes, over 100,000 proteins, and multiples of their translational and metabolomics products. Thus, the permutations in respect to opportunities for pharmacological interventions are unprecedented, vast, and most promising for innovative medicines.

The pharmacogenomics era as a source for drug targets also poses unprecedented hurdles in selection, validation, and translation into effective and safe drugs. New technologies continue to drive efficiency and robustness of mining the genomic drug discovery opportunities, but physiological and integrated biology knowledge is lagging. In this perspective, translational medicine and biomarkers research have taken center stage in validation of the molecular target for pharmaceutical exploitation.

In this chapter we offer a utilitarian approach to biomarkers and target selection and validation that is driven by the translational medicine prospect of the target to become a successful drug target. We hereby offer classification and analytical process aimed to assess risk, innovation, feasibility, and predictability of success of translating novel targets into successful drugs. We provide clear definitions of the type of biomarkers that are the core of translational medicine and biomarkers research in modern pharmaceutical companies.

## BIOMARKERS: UTILITARIAN CLASSIFICATION

Biomarkers are the stepping-stones for modern drug discovery and development [1–4]. *Biomarkers* are defined as biological substances or biophysical parameters that can be monitored objectively and reproducibly and used to predict drug effect or outcome. This broad definition is, however, of little utility to the pharmaceutical process since it carries no qualification for the significance and use of the biomarker. The following classes and definitions of biomarkers are therefore offered:

1. *Target validation:* biomarkers that assess the relevance and the potential for a given target to become the subject of manipulation that will modify the disease to provide clear therapeutic benefits while securing a sufficient therapeutic index of safety and tolerability.
2. *Compound-target interaction biomarkers:* biomarkers that define the discrete parameters of the compound (or biological) interaction with the molecular target. Such parameters include binding of the compound to the target, its residency time on the target, the specific site of interaction with the target, and the physical or chemical consequences to the target induced by the compound (or biological).
3. *Pharmacodynamic biomarkers:* biomarkers that predict the consequence(s) of compound (biological) interaction with the target. The pharmacodynamic biomarkers include events that are desired therapeutically and adverse events based on mechanism of action. Pharmacodynamic biomarkers can report on discrete molecular events that are proximal to the biochemical pathway that is modified by the manipulated target or remote consequences such as in vivo or clinical outcomes (morbidity or mortality). Pharmacodynamic biomarkers are diverse and frequently nonobvious. Advanced and sophisticated bioinformatics tools are required for tracking the divergence and convergence of signaling pathways triggered by compound interaction with the target. A subset of the pharmacodynamic biomarkers are consequences induced by the compound outside its intended mechanism of action. Such pharmacodynamic effects are often termed "off-target" effects, as they are not the direct consequence of the compound interaction with the target. Usually, such pharmacodynamic events are due to unforeseen lack of selectivity or metabolic transformations that yielded metabolites not present (or detected) in the animals used for safety and metabolic studies prior to launch of the compound into human trials or into human use. These issues are not dealt with in this chapter.
4. *Disease biomarkers:* biomarkers that correlate statistically with the disease phenotype (syndrome) for which therapeutics are developed. Correlation of levels (in the circulation, other fluids or tissue) or expression patterns (gene, protein) in peripheral blood cells or tissues should signify disease initiation, progression, regression, remission, or relapse. In addition, duration of aberrantly expressed biomarkers could also be associated with risk for disease, even if the level of the biomarker does not change over time. Since disease biomarkers are defined by their statistical correlation to features of the disease, it is imperative that the clinical phenotyping is clearly defined. Stratification of all possible phenotypic variables is clearly a prerequisite for accurate assessment of the discrete relationships of the biomarker to the disease. Gender, age, lifestyle, medications, and physiological and biochemical similarities are

often not sufficiently inclusive, resulting in a plethora of disease biomarker claims that are often confusing and futile.

5. *Patient selection:* biomarkers that are used for selection of patients for clinical studies, specifically proof-of-concept studies or confirmation phase III clinical trials that are required for drug registration. These biomarkers are important in helping to select patients likely to respond (or conversely, not respond) to a particular treatment or a drug's specific mechanism of action, and potentially predict those patients who may experience adverse effects. Such biomarkers are frequently genetic (single-nucleotide polymorphism, haplotypes) or pharmacogenomic biomarkers (gene expression), but could be any of the primary pharmacodynamic biomarkers. Biomarkers for patient selection are now mainstream in exploratory clinical trials in oncology, where genotyping of tumors in view of establishing the key oncogenic "driver(s)" are critical for the prediction of potential therapeutic benefits of modern treatments with molecular targeting drugs. The success of the new era of molecular oncology (as compared to the cytotoxic era) will depend largely on the ability to define these oncogenic signaling pathways via biomarkers such as phosphorylated oncogenes, or the functional state due to mutations that cause gain or loss of function.

6. *Adaptive trial design:* The objectives of adaptive design trials are to establish an integrated process to plan, design, and implement clinical programs that leverage innovative designs and enable real-time learning. The method is based on simulation-guided clinical drug development. In a first step, the situation is being assessed, the path forward and decision criteria defined, and assumptions analyzed. Adaptive trials have become an enabler strategy, and they work to integrate competing positions and utilities into a single aligned approach and to force much clearer articulation and quantification on the path forward. Once this framework is established, a formal scenario analysis that compares the fingerprints of alternative designs through simulation is conducted. Designs that appear particularly attractive to the program are further subjected to more extensive simulation. Decision criteria steer away from doses that are either unsafe or nonefficacious and aim quickly to hone in onto the most attractive dose range. Response-adaptive dose-ranging studies deploy dynamic termination rules (i.e., as soon as a no-effective-dose scenario is established, the study is recommended for termination). Bayesian approaches are ideally suited to enable ongoing learning and dynamic decision making [5]. The integrator role of adaptive trials is particularly strong in establishing links between regulatory accepted "confirm"-type endpoints and translational medicine's efforts to develop biomarkers. Search for biomarkers that may enable early decision making need to be read out early to gain greater confidence in basing decisions on them. A biomarker can be of value even if it only

allows a pruning decision. These considerations highlight the importance of borrowing strength from indirect observations and use mathematical modeling techniques to enhance learning about the research question. For example, in a dose-ranging study, it is assumed that there should be some relationship between the response of adjacent doses, and this assumption can be used to model an algorithm. Both safety and efficacy considerations can be built into this model: ideally, integration of all efforts, from disease modeling in discovery to PK/PD modeling in early clinical development to safety/risk and business case modeling in late development [4–7].

The utility of this system is represented in Figures 1 and 2, which suggest a semiquantitative scoring system that helps assess the strength of the program

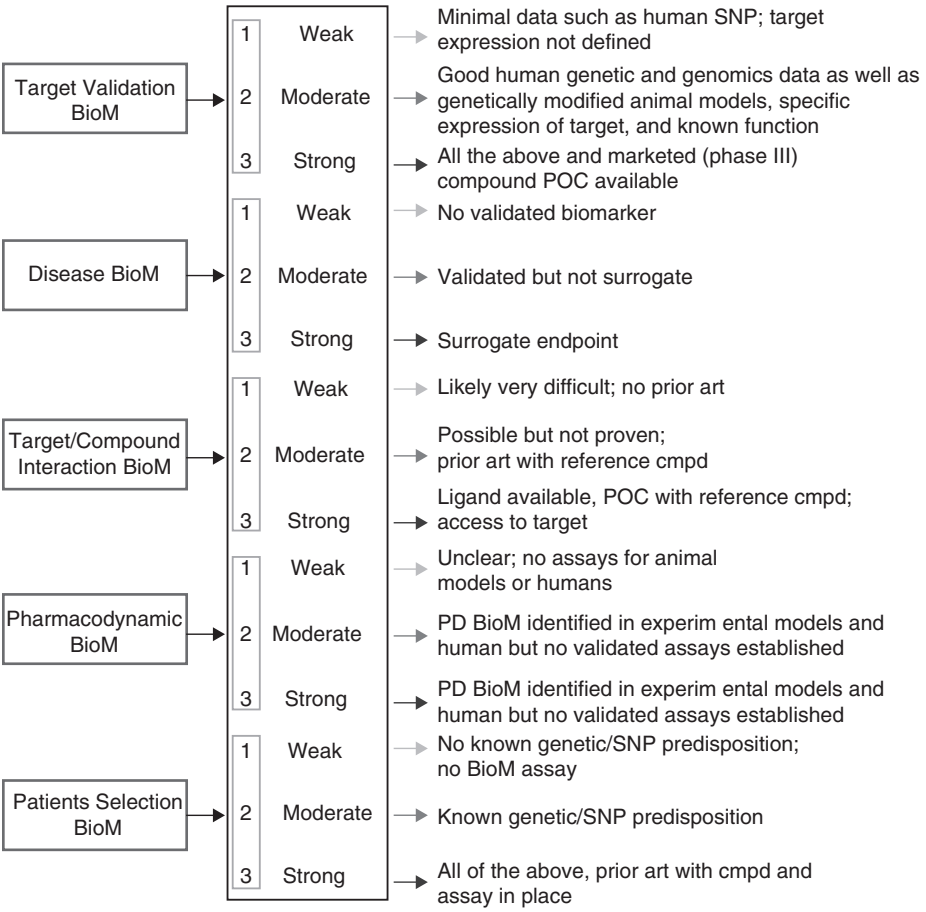


**Figure 1** Criteria for biomarker scoring.

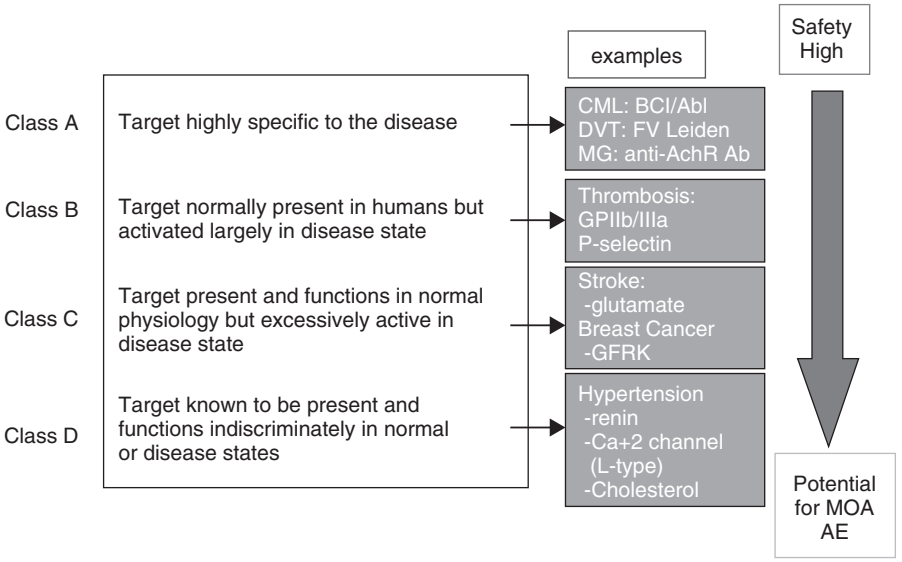


**Figure 2** Type I biomarkers: target validation translational medicine perspectives. CMI, chronic myelocytic leukemia; DVT, deep vein thrombosis; FV, factor V; MG, myasthenia gravis; AchR, acetylcholine receptor; GPIIb/IIIa, platelet integrin receptor; GFRK, growth factor receptor kinase; MOA, mechanism of action; AE, adverse effects.

overall and identification of the areas of weaknesses in each of the biomarkers needed along the compound (biological) progression path. Figure 3 illustrates the continuum of biomarkers research, validation, and implementation along the complete time line of drug discovery and development, including life-cycle management (phase IV) and new indication (phase V) when appropriate. Figure 3 illustrates the interfaces of translational medicine within the traditional drug discovery and development process, while Figure 4 represents the new model of "learn and confirm," where biomarkers figure prominently in driving the "learn" paradigm.

A program for which a STRONG score is established across all five biomarkers' specifications provides confidence of the likelihood for success from the biological and medical perspectives and is likely to result in a more promising development outcome. Similarly, it would be prudent to voice concerns regarding programs that score WEAK, especially if low scores are assigned to target validation, pharmacodynamic, and in special cases, target-compound interaction (e.g., central nervous system target). This scoring system is complementary to other definitions of biomarkers based on certain needs. For example, surrogate biomarkers as defined by the U.S. Food and Drug Administration (FDA) are markers that can be used for drug registration in

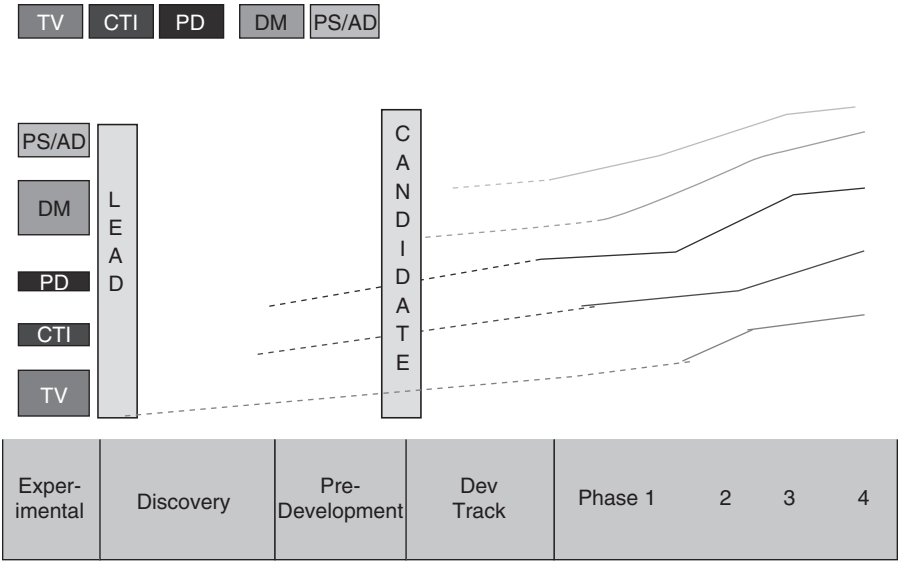


**Figure 3** Building translational medicine via biomarker research. (*See insert for color reproduction of the figure.*)

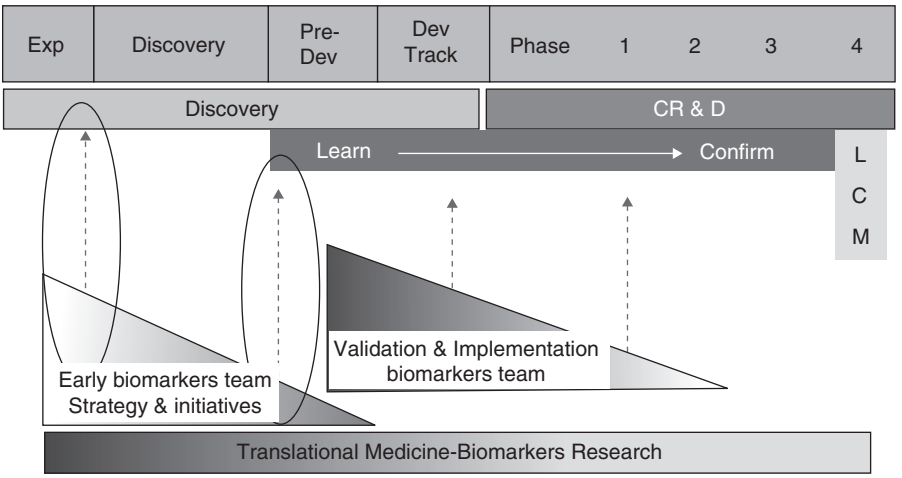


**Figure 4** Translational medicine: biomarker implementation along the pipeline. Exp, exploratory phase; Pre-Dev, predevelopment track; CR&D, clinical research and development; LCM, life-cycle management.

lieu of more definitive clinical outcome data. Surrogate biomarkers are few and difficult to establish (e.g., blood pressure and cholesterol; Figure 2).

## PRINCIPLES OF TARGET SELECTION

Two key guiding principles are essential in the early selection process of molecular targets:

1. Modulating the target carries the prospect of unequivocal medical benefit (efficacy) to patients beyond a standard of care.
2. Benefits can be garnered while maintaining a sufficient level of safety that can be realized within the attainable compound exposure.

Such a mission is frequently unachievable, and hence establishing an acceptable *therapeutic index* is the practical goal for most drug development schemes. Commonly, a therapeutic index is established by calculation of the ratio of the maximum tolerated dose (MTD) and the minimum effective dose (MED) in animal efficacy and safety studies. In this light, targets selected for drug development can be classified with respect to risk assessment based on the following categories (Figure 2):

*Class A.* The target is only present and contributing to the disease process.

*Class B.* The target is present physiologically but in a nonactive form, but then is activated and contributes to the disease.

*Class C.* The target functions physiologically but in an augmented, uncontrolled fashion that contributes to the disease.

*Class D.* The target functions in normal states and indiscriminately in disease (e.g., no difference in target expression, functions, or distribution can be identified in disease as compared to the normal physiological state).

### Class A: Disease-Specific Target

A disease-specific molecular target should be a molecule that operates only in the disease state and does not participate in physiological (normal) functions. Drug interaction with such targets should provide efficacy with the lowest chance for mechanism-based adverse effects when manipulated by drugs. Examples of such targets are genetic disorders, which result in either overactivity or loss of activity of the target. Such is the case in chronic myelogenous leukemia (CML), which results from aberrant recombination of DNA from chromosome 22 into chromosome 9 (Philadelphia chromosome), fusing the *Bcr* and *Abl* genes into an overacting tyrosine kinase, which drives oncogenic transformation. To cure the disease, potent and selective inhibitors of

this aberrant kinase had to be discovered, a task that took over a decade to accomplish [8]. Such targets have the potential for a high safety profile. It is however, important to note that this example may not necessarily represent the ultimate approach for this disease since the activity of the kinase (*Bcr*/*Abl*) is driven by the *Abl* kinase catalytic site, which is preserved in its physiological format. Thus, inhibitors of this target/kinase by drugs such as Gleevec may still carry the potential for interference in tissue and cells in which the *Abl* kinase is physiologically active.

Another example that is applicable to this category is a disease such as myasthenia gravis, where specific antibodies that block the acetylcholine receptors cause progressive muscle weakness. Specific neutralizing agents to these antibodies are likely to provide high efficacy in treating the disease, with the likelihood of fewer adverse effects [9] since such antibodies are not physiologically present in human beings. These examples are typical for type 1 class A target validation. The biomarkers that need to be established for this category should focus on validating the specificity of the target to the disease state.

### Class B: Target Present Physiologically in a Nonactive Form but Is Activated and Contributes to the Disease

This class of targets has little or no discernible physiological activity in normal states, yet in certain pathophysiological situations, the target is presented, activated, and plays a role in a pathophysiological event. Example of such targets in the type 1 class B category is the P-selectin adhesion molecule. This adhesion molecule is normally cryptic within platelets and endothelial cells. Upon activation of these cells, P-selectin is presented on the surface of the cell and mediates adhesion interaction with its ligand, a mechanism believed to play a role in thrombosis and inflammation. Inhibitors of P-selectin binding to its ligand, the P-selectin glycoprotein ligand (PSGL-1), are expected to provide clinical benefit with a lower likelihood of adverse events. To validate this situation, biomarkers that confirm the preferential role of the activated target in a pathophysiological process while maintaining little physiological function are essential.

However, one must be aware of potentially serious limitations to this approach where cryptic targets in the physiological state that are activated in pathological conditions and where inhibition of the target may not only provide for significant therapeutic benefit, but where inhibition of such a target may also expose the patient to some other risk, such as loss of host defense from injury. Such is the case of the platelet adhesion integrin molecule, GPIIb/IIIa, which serves as the final common pathway for platelet aggregation. Interfering with activated GPIIb/IIIa binding to its ligand (e.g., fibrinogen) provides effective and often lifesaving therapy in patients with acute risk for thrombosis; however, chronic treatment with GPIIB/IIIa antagonists have not been particularly effective in providing benefits, due to the

relatively high frequency of significant adverse effects due to bleeding since platelet adhesion to matrix protein is essential to seal bleeding sites in trauma and disease conditions. Thus biomarkers for this class must establish the full physiological significance of the target in order to assess the therapeutic index of tolerability (benefits as well as risks).

### Class C: Target Functions Physiologically but in an Augmented, Uncontrolled Fashion That Contributes to the Disease

This class of targets includes molecules that play an active role in normal physiological processes, some of which may be critical to health. Such is the neurotransmitter glutamate in the central nervous system, which is essential to cognition, memory, thought processes, and state of arousal. However, in ischemic stroke or asphyxia, glutamate release is uncontrolled and reaches to high levels over prolonged periods that are believed to be neurotoxic and likely contribute to neuronal death following a stroke. Inhibitors of glutamate release or antagonists of its action at various receptors are believed to carry the potential for effective treatment for stroke provided that the inhibition of excess release of the neurotransmitter can be achieved in a timely manner and only to an extent that preserves the physiological need of this transmitter, and over short periods (only that limited period where excess glutamate is neurotoxic). Such targets may be pharmaceutically exploitable when their manipulation is tuned carefully to the pathophysiological context.

Another example of a target in this category includes the human growth factor receptor kinase (hGFRK) inhibitor, Herceptin, which in certain cancers (e.g., breast cancer) is constitutively activated and participates in the oncogenic drive. Inhibition of the hGFRK, while clearly of therapeutic value in breast cancer, has also been associated with heart failure due to the physiological role of GFRK in the cardiac myocyte survival signaling pathway [10]. Thus, the biomarker challenge in modulation of class C targets of this nature is in identifying biomarkers that assess the needed "titration" for inhibition of the target activity only to an extent that is necessary for the maintenance of normal physiological function.

### Class D: Target Maintains Physiological Functions in Normal and Disease States

This class of targets encompasses the largest group of molecular targets exploited so far by modern drugs. Many members of this class have yielded highly beneficial therapies. This class consists of molecular targets that are known to have important physiological functions, which cannot be differentiated within a disease context; that is, the target is not different in its expression levels (gene, protein) or signaling pathway in normal and disease states. A priori, such targets harbor the greatest risk for mechanism-based adverse effects, as there is no apparent reason to expect that modulation of the target in the disease state will spare the normal physiological function of the target.

Examples of such targets include the coagulation factor inhibitors (e.g., FIX, FXa, and thrombin), which are critical to maintain a physiological level of homeostasis; hence, inhibition of these targets carries inherent bleeding liabilities. Similarly, all current antiarrhythmic drugs (e.g., amiodarone, lidocaine, dofetilide), while effective in treating life-threatening arrhythmias, all carry significant liability for mechanism-based pro-arrhythmic effects and the potential for sudden death. The biomarker challenges for this class are defining the fine balance needed between efficacy in the disease context and expected safety limitations. Biomarkers that define the acceptable therapeutic index are key to the successful utility of drugs that modulate such targets. However, targets in this class do not necessarily exhibit a narrow safety margin for clinically meaningful adverse effects. Significant examples are the L-type $Ca^{2+}$ channel blockers. The L-type $Ca^{2+}$ channel is an essential conduit of $Ca^{2+}$ needed for "beat by beat" $Ca^{2+}$ fluxes that secure precise rhythm and contractility of the heart, skeletal muscle, neuronal excitability, and hormone and neurotransmitter release. Yet L-type $Ca^{2+}$ channel blockers are important and sufficiently safe drugs that are used to treat hypertension, angina, and cardiac arrhythmias with undisputable medical benefits. However, inherent to the L-type $Ca^{2+}$ channel blockers in this class of targets are mechanism-based adverse effects associated with rhythm disturbances, hypotension, edema, and other liabilities.

Probably the best example for system specificity of physiological targets that provide major medical benefits with a high safety margin is the renin–angiotensin–aldosterone (RAAS) system. The RAAS is an important blood pressure, blood volume, and blood flow regulatory system, yet its manipulation by several different pharmacological agents (rennin inhibitors, angiotensin I converting enzyme inhibitors, angiotensin II receptors antagonists) has yielded highly beneficial drugs that reduce risk of morbidity and mortality from hypertension, heart failure, and renal failure, despite the fact that the system does not demonstrate significant operational selectivity between normal and disease states (especially hypertension). However, mechanism-based hypotension and electrolyte disturbances can limit the therapeutic benefit of these drugs, and elicit significant adverse effects when the RAAS is excessively inhibited [11]. The biomarker challenge for these targets is to define the relative or preferential role of the target in its various physiological activities, where minor manipulation in one organ might provide sufficient therapeutic potential while providing a low likelihood for adverse effects that result from more substantial inhibition of the same target in other organs.

## SUMMARY

The analysis and classification offered in this chapter regarding biomarkers in drug discovery and development aim to highlight the need for careful study and analysis of the significance of the target selected for therapeutic intervention as the first crossroad for success or failure in the development of effective

and safe drugs [12]. The analysis and utility of biomarkers along the process of drug discovery and development have become an integral part of the "learn and confirm" paradigm of drug discovery and development in leading pharmaceutical organizations such as Wyeth Research. Such analyses are useful to guide the "learn phase" in search for biomarkers that can better assess the benefits and risks associated with manipulation of the molecular target.

The scope of this chapter does not allow for a detailed review of the "learn and confirm" paradigm, for which the readers are directed elsewhere [13,14]. Various technological and strategic activities are needed to establish the biomarker strategies for the various targets described. The need to address these issues via biomarker research, validation, and implementation beginning at the very early stages of the drug discovery and development process is emphasized. In the pharmaceutical setting, it means beginning efforts to identify biomarkers for all five categories listed above. Such efforts could begin even before a tractable compound (biological) is in hand, a time where target validation is a clear focus of the program. As a compound becomes available, compound–target interaction, pharmacodynamic (efficacy and safety) biomarkers, and strategies for patient selection and adaptive design needs must be explored. At the onset of the first-in-human studies, all strategies, plans, and biomarker research should be as well worked out as possible. We believe that fundamental changes in the structure, function, and interfaces of pharmaceutical R&D are urgently needed to provide a key role for translational medicine and biomarkers research toward more successful discovery and development of innovative medicines.

## REFERENCES

1. Biomarker Definition Working Group (2001). Biomarkers and surrogate biomarkers endpoints: preferred definitions and conceptual framework. *Clin Pharmacol Ther*, 69:89–95.
2. Trusheim R, Berndt ER, Douglas FL (2007). Stratified medicine: strategic and economic implications of combining drugs and clinical biomarkers. *Nat Rev Drug Discov*, 6:287–293.
3. Feuerstein GZ, Rutkowski JL, Walsh FL, Stiles GL, Ruffolo RR Jr (2007). The role of translational medicine and biomarkers research in drug discovery and development. *Am Drug Discov*, 2:23–28.
4. FDA (2004). Challenge and opportunity on the critical path to new medical products. http:www.fda.gov/oc/initiatives/criticalpath/whitepaper.pdf.
5. Berry DA (2006). Bayesian clinical trials. *Nat Drug Discov Rev*, 5:27–36.
6. Gallo P, Chuang-Stein C, Dragalin V, Gaydos B, Krams M, Pinheiro J (2006). Adaptive design in clinical drug development: an executive summary of the PhRMA working group. *J Biopharm Stat*, 16:275–283.
7. Krams M, Lees KR, Hacke W, Grieve AP, Orgogozo J-M, Ford GA (2003). Acute stroke therapy by inhibition of neutrophils (ASTIN): an adaptive dose–response study of UK-279,276 in acute ischemic stroke. *Stroke*, 34:2543–2548.

8. Kurzrock R (2007). Studies in target-based treatment. *Mol Cancer Ther*, 6(9):2385.
9. Hampton T (2007). Trials assess myasthenia gravis therapies. *JAMA*, 298(1):29–30.
10. Chien KR (2006). Herceptin and the heart: a molecular modifier of cardiac failure. *N Engl J Med*, 354:789–790.
11. Hershey J, Steiner B, Fischli W, Feuerstein GZ (2005). Renin inhibitors: an anti-hypertensive strategy on the verge of reality. *Drug Dev Today*, 2:181–185.
12. Simmons D (2006). What makes a good anti-inflammatory drug target? *Drug Discov Dev*, 5–6:210–219.
13. Gombar C, Loh E (2007). Learn and confirm. *Drug Discov Dev*, 10:22–27.
14. Sheiner LB (1997). Learning versus confirming in clinical drug development. *Clin Pharmacol Ther*, 61:275–291.

# CLINICAL VALIDATION AND BIOMARKER TRANSLATION

David Lin, B.MLSc.
*University of British Columbia, Vancouver, British Columbia, Canada*

Andreas Scherer, Ph.D.
*Spheromics, Kontiolahti, Finland*

Raymond Ng, Ph.D.
*University of British Columbia, Vancouver, British Columbia, Canada*

Robert Balshaw, Ph.D., and Shawna Flynn, B.Sc.
*Syreon Corporation, Vancouver, British Columbia, Canada*

Paul Keown, M.D., D.Sc., MBA, Robert McMaster, D.Phil., and Bruce McManus, M.D., Ph.D.
*University of British Columbia, Vancouver, British Columbia, Canada*

## INTRODUCTION

Throughout history, biological and pathogenic processes have been measured to monitor health and disease. The presence of "sweetness" in urine and blood was recognized thousands of years ago as an indication of the disorder now known as diabetes. By recent example, the characterization of infections has been achieved by performing cultures for microorganisms, both for identification and to establish sensitivities to antibiotics. Any such measures, however variously quantitated, were forerunners of what are now popularly referred

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

to as biomarkers. When clinical laboratories became well established after the middle of the twentieth century, many components of body fluids and tissues were assayed. These are all biomarkers of the physiological state of being. Modern biomarkers being discovered or sought are pursued and measured based on the same principles as the first, most rudimentary of disease indicators. Hundreds of biomarkers are now used in modern medicine for the prediction, diagnosis, and prognostication about disease, as well as for monitoring preventive or therapeutic interventions. A sizable challenge now lies in selecting novel biomarkers that are most appropriate for clinical use in the face of many new candidates being suggested by various investigators.

Biomarkers are often thought of as strictly molecular in nature, but in fact they include a vast range of objectively measurable features. Any characteristic that reflects a normal biological process, pathogenic process, or a pharmacologic response to an intervention can potentially become a clinically useful biomarker [1]. Thus, there are many different types of biomarkers, including molecular, physiological, or structural features of biological systems. However, how a specific measure reaches the status of customary has not been so clearly established.

The focus on using biomarkers in clinical decision making and diagnosis has expanded, and similarly, biomarkers are playing an ever-increasing role in drug development processes and for regulatory decision making. As drug development costs continue to rise, biomarkers have become increasingly important, as they are a potential means to decrease the time for development, costs, and late-phase attrition rates in the regulatory approval process for new drugs:

- Reduce clinical trial costs:
  - Decrease late-phase attrition rates.
  - Replace time-consuming clinical endpoints.
- Reduce required clinical trial sample size by way of patient stratification:
  - Identify the population most likely to benefit.
  - Identify the population with a high level of risk for events of interest.
- Provide more robust assays than some conventional clinical endpoints.
  - Help improve models for calculating return on investment.

For example, cost and time can be reduced when biomarkers help to segment patient groups early in trials, and some new markers may be more timely than many of the more traditional clinical endpoints or outcomes used currently in assessing clinical trials, such as patient survival [2,3]. Biomarkers are also beneficial tools for selecting the best candidates for a trial or to increase safety through more effective drug monitoring. Ultimately, by reducing the required time investment to demonstrate a drug's safety and efficacy,

biomarkers may greatly reduce the costs and risks of performing a clinical trial.

## BIOMARKER DISCOVERY, DEVELOPMENT, AND TRANSLATION

Biomarker discovery can be performed using animal models, but it is now commonly carried out in humans from the very beginning stages of biomarker development. Biomarkers analyzed in preclinical animal studies are eventually transferred to the human, clinical settings. In a controlled laboratory environment the conditions are relatively constant, and "subjects" (i.e., the animals) are homogeneous and free of complicating co-morbidities. In these laboratory settings there are generally more options regarding the assays available, and the possibility of frequent and repeated testing in individual animals or groups of similar or identical animals allows for changes in the measured matrix of analytes to be detected with high precision and sensitivity. Ideally, a candidate biomarker discovered in this fashion would then be transferred into the clinical environment and evaluated further on human samples. The drawback of this approach is that many of the biomarkers discovered in animal models cannot be translated for use in humans. Animal models often do not accurately reflect human biology [4]. Further, the biomarker candidate may not achieve acceptable performance standards in heterogeneous patients with age, gender, and racial differences. This is why it has been suggested that pilot studies of biomarkers be conducted in humans first, in early phase II clinical trials that incorporate the variability in various ambient influences, and then that the marker be validated in preclinical studies, and in a later stage, clinical trials in parallel. This approach expedites the biomarker development process and minimizes the attrition rate of biomarker candidates since they are developed from the beginning under human clinical conditions.

In patients, biomarker candidate discovery is performed preliminarily in an internal primary cohort and is confirmed subsequently in an external secondary cohort. These two discovery phases are typically performed in observational or retrospective cohorts, and less and less often in preclinical studies. Some companies are avoiding animal models for discovery all together to avoid spending time and money on animal-based biomarkers that can often lead to a dead end.

It is not always practical to pursue validation of biomarker candidates identified in the discovery process. It is important to establish parameters for rational, statistically sound, and evidence-based selection and rejection of candidate biomarkers [5–7]. The decision to continue development of a biomarker candidate is largely based on its potential to contribute cost-effectively to disease management [8]. Biomarkers used in the early phases of clinical development may be useful in providing more timely proof of concept or dose-range information than a real clinical endpoint [9]. Biomarker development should also be driven by clinical need [8]. A clinically useful biomarker

must favorably affect clinical outcomes such as decreasing toxicity or increasing survival [10].

Once cost-benefit ratios are evident and more than one institution has consistently confirmed a biomarker's ability to perform at the requisite levels of sensitivity and specificity, a biomarker is ready for prospective testing in clinical trials [8]. The U.S. Food and Drug Administration (FDA) pharmacogenomic submission guidelines recommend that transferring genomic biomarker sets from microarrays to other platforms (such as quantitative real-time polymerase chain reaction) be attempted only once it has been demonstrated that the differential expression of such genes is sensitive, specific, and reproducible. Analytical quality assurance is performed continually throughout biomarker and drug development for most biomarkers; however, if a biomarker is to be used as an endpoint in a phase III efficacy study, it should be validated beforehand [11]. Assay and analytical specificity and sensitivity should be established and validated prior to clinical phases such that clinical qualification can be carried out using proven, analytically robust methods [12]. However, as biomarker and drug development are intertwined processes, they may often occur concurrently throughout the different stages of clinical trials.

## BIOMARKER VALIDITY AND VALIDATION: THE REGULATORY PERSPECTIVE

The FDA's *Guidance for Industry: Pharmacogenomic Data Submission*, published in 2005, has helped introduce and define exploratory and valid biomarkers. The FDA defines a *valid pharmacogenomic biomarker* as one that is measured in an analytical test system with well-established performance characteristics and for which there is an established scientific framework or body of evidence that elucidates the toxicological, pharmacological, or clinical significance of the test result. The FDA further classifies valid biomarkers as "probable" and "known" in terms of the level of confidence that they attain through the validation process. *Probable valid* biomarkers may not yet be widely accepted or validated externally but appear to have predictive value for clinical outcomes, whereas *known valid* status is achieved by those that have been accepted in the breadth of the scientific community.

It is important to realize that the different classes of biomarkers reflect their levels of confidence (Figure 1). This can be thought of in a hierarchical manner, with exploratory biomarkers being potential precursors of clinically useful (probable or known) valid biomarkers [13].

Integrating biomarkers into clinical trials for eventual clinical use by identifying the best or most valid biomarker candidates is not a clearcut process. The term *validity*, particularly in the field of biomarkers research, is a broad concept that has been used to describe everything from the analytical methods to the characteristics of the biomarkers identified [14]. Validity is also used across multiple industries, not only medical or health disciplines. Therefore,

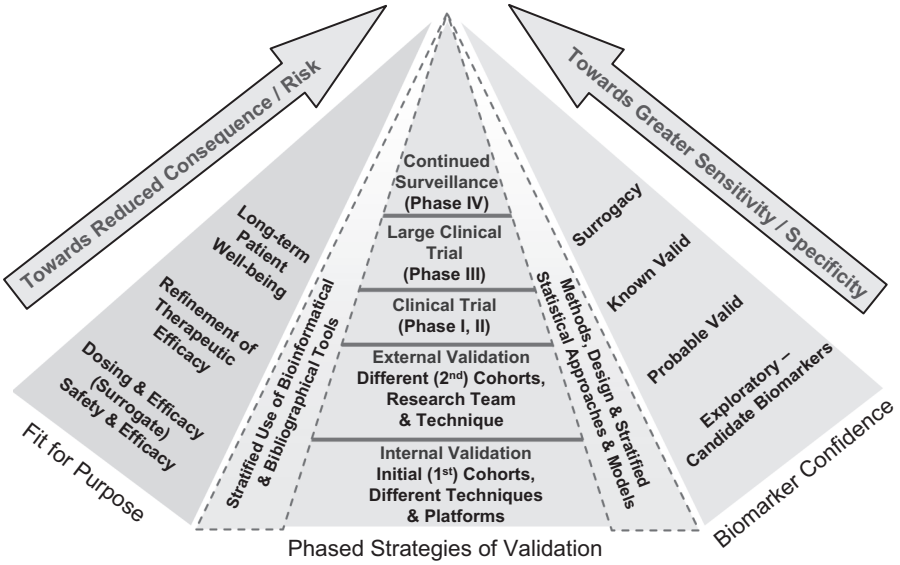


**Figure 1** Biomarker validation. Biomarker development and validation are driven by intended use or fit-for-purpose (FFP). The principle of FFP validation is that biomarkers with false-positive or false-negative indications pertaining to high patient consequences and risks necessitate many phases of validation. There are five phases of validation that can be executed with the stratified use of various bioinformatical and bibliographical tools as well as different designs, statistical approaches, and modeling: (1) internal validation, (2) external validation, (3) clinical trials (phases I and II; checking for safety and efficacy), (4) large clinical trials (phase III), and (5) continued surveillance. Sensitivity and specificity correlate with the intended purpose of the biomarker, and the level of confidence that a biomarker achieves depends on the phase of validation that has been reached. In ideal cases biomarkers reach surrogate endpoint status and can be used to substitute for a clinical endpoint. This designation requires agreement with regulatory authorities, as the consequences of an ambiguous surrogate endpoint are high.

when referring to biomarkers, validation is sometimes termed *qualification* for clarity. Biomarker qualification has been defined as a graded *fit-for-purpose* evidentiary process linking a biomarker with biology and clinical endpoints [15,16].

Traditionally, the validity of clinical biomarkers has become established in a typically lengthy process through consensus and test of time [17]. Now more than ever, clear guidelines for validation are needed, as technological advances have drastically increased biomarker discovery rates. With the recent explosion of "omics" technologies and advancements in the fields of genomics, proteomics, and metabolomics, high-throughput biomarker discovery strategies are now widely used. This has created some unforeseen issues. Biomarker

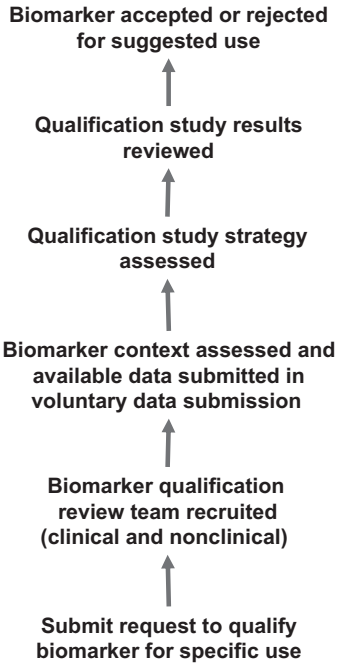


**Figure 2** FDA biomarker qualification pilot process. (Adapted from ref. 17.)

candidate discovery now commonly outruns the rate at which the candidates are being validated, creating a bottleneck in biomarker (assay) development [18,19]. Great efforts are now being undertaken to accelerate the acceptance of biomarkers from exploratory to valid, as the goal of many research teams and drug companies is to streamline the translation of biomarker from basic science and discovery to clinical use [12].

Despite the definitions provided by the FDA and the availability of FDA's *Guidance for Industry: Bioanalytical Method Validation* in 2001, there is still a lack of sufficient regulatory guidance for biomarker validation. The FDA has designed a qualification process map that sets the foundational framework toward establishing validation guidelines (Figure 2). This pilot structure, to start qualification processes for biomarkers in drug development, is designed around various FDA centers whereby the context and qualification of new biomarkers is assessed. Ultimately they are rejected or accepted for suggested use relative to current biomarkers. This may not be ideal, as it may be problematic to establish new biomarkers accurately based on the current biomarkers, which are themselves often imperfect relative to a specific endpoint [17]. Nonetheless, this pilot framework will eventually enable more detailed biomarker translation models, which address some of the remaining issues with the current guidelines, to be developed.

There remains a lack of specific guidelines on which validation process(es) are recommended or expected in order to transition effectively from exploratory to valid biomarkers, or from probable valid biomarkers into known valid biomarkers [13,20]. Confusion still exists with regard to analyses or experiments that need to be performed and data that are both appropriate and sufficient for biomarker (assay) validation [20].

The confusion and inconsistency in the validation process are contributed to partially by the diverse nature of biomarker research [3,20]. Considering the large variety of novel biomarkers, their applications, and associated analytical methods, it is unlikely that FDA regulations or other available guidelines will easily be able to address validation issues associated with all possible research objectives [16,20]. Thus, it is incredibly difficult to establish, let alone use, a specific detailed universal validation guideline [20].

## FIT-FOR-PURPOSE STRATEGY

Despite the lack of universal guidelines or agreement on the specific requirements for biomarker assay development and validation, there is a general consensus in the biomarker research community that the foundation of validation efforts is to ensure that the biomarker(s), or the assay, is "reliable for its intended use" [20]. This principle is now commonly referred to as the *fit-for-purpose validation strategy* or guideline. This approach embraces the notion that depending on the intended purpose of the biomarker, the objectives and processes or types of validation will probably be different [21,22]. Dose selection, early efficacy assessment or candidate selection, and surrogacy development are all examples of biomarker clinical purposes, each of which have differing levels of validation requirement. The risk and consequence may be different depending on the purpose, even when the same biomarker is used [4].

Further, the degree of stringency and phase of the validation, both of which are discussed later in the chapter, should be commensurate with the intended application of the validation data (Figure 1) [22]. In that sense, the term *validity* should be thought of as a continuum—an evaluation of the degree of validity—rather than an all-or-none state [14,23]. Therefore, biomarker utility is not measured dichotomously and should not be classified as simply good or bad. The level of a biomarker's worth is gauged on a continuous scale, with some being much more valuable than others, depending on what they indicate and how they can be applied [24]. Validation of biomarkers and establishment of their "worth" or "value" is a continually evolving and often iterative process (Figure 1).

### Application of the Fit-for-Purpose Strategy

The fit-for-purpose (FFP) strategy is a fluid concept that can be applied to any clinical trial to validate biomarkers of interest. As described earlier, the classification and validation of a biomarker is context-specific, and the validation

criteria required are dependent on the intended use of the biomarker. The concept of fit-for-purpose validation is sensible, as biomarkers may be used to include or exclude patients from treatment and to determine dosing, toxicity, or safety. The consequence or risk of a false-negative or false-positive biomarker indication must be considered since even the most sensitive and specific assays are not perfect. A biomarker intended as a population screen must be very sensitive and extremely specific [25]. Any tests that are to be used in isolation for decision making require particularly stringent validation, whereas drug efficacy biomarkers that are typically used in groups or involve subsequent testing may require less stringency since the consequence of a false indication of efficacy is lower [26].

### Biomarker Validation Process

Utilizing the FFP strategy, various questions can be asked at the beginning of any biomarker research and development (R&D) project: For what reason are the biomarker(s) being identified? Is the clinical validation going to validate the biomarker for specificity, sensitivity, and reproducibility for diagnostic purposes, or to serve as a surrogate endpoint? What business-critical decisions will be made based on the biomarker data? These questions not only help determine the level of confidence required for the biomarker, but also help strategize regarding the phases of validation. For example, in the case of developing surrogate biomarkers to replace clinical endpoints, candidate biomarkers would, theoretically, evolve over time toward the point of surrogacy as the research project moves through different phases of validation. For the purpose of this chapter, the validation process from initial biomarker development to postmarket surveillance has been broken down into five major phased strategies (Figure 1). It is important to note that the overall process is a continuous loop driven by the intended purpose of the biomarker data, but the flow of the overall process may be subjected to change depending on the results generated at each phase [22].

In general, the process of biomarker validation can be described as a multifaceted process that includes determining sensitivity, specificity, and reproducibility of the assay and also clinical sensitivity and specificity [27]. This applies to both methodologic and clinical validation. Method validation pertains to the process by which the assay and its performance characteristics are assessed. This assessment is performed throughout biomarker translation and is based on several fundamental parameters: accuracy, precision, selectivity, sensitivity, reproducibility, and stability [16,22]. Method validation is not discussed in detail in this chapter.

## PRECLINICAL AND CLINICAL VALIDATION PHASES

Clinical validation is the documented process demonstrating the evidentiary link of a biomarker with a biological process and clinical endpoints [15,16].

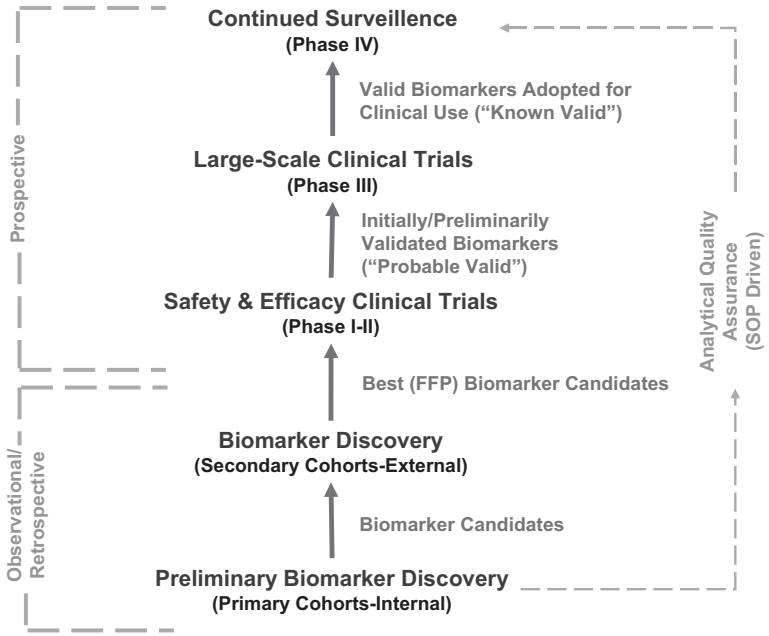


**Figure 3** Biomarker development and translation. Biomarker translation from discovery to clinical use involves five general steps or phases. Biomarker candidates are initially discovered in a primary (internal) cohort and confirmed in a secondary (external) cohort through clinical observations. Biomarker candidates that have the best fit-for-purpose and also satisfy a clinical need then enter prospective phase I and II clinical trials. Once a biomarker is used in these early clinical trials to demonstrate safety and efficacy, they may be considered "initially validated biomarkers." Following large-scale clinical trials (phase III), and once biomarkers have been used for decision making, they may be considered valid and may be adopted for clinical use. Biomarker assessment continues in postmarket phase IV trials. SOP-driven analytical quality assurance is performed throughout all processes of biomarker translation.

The validation process whereby biomarkers are translated from discovery to clinical use should be customized according to biomarker type, use, variability, and prevalence. However, the general process for validating any biomarker is the same. Before any biomarker can be applied clinically, it is subjected to analytical/method validation and also clinical validation. Biomarker validation can be performed in five general translational steps: preliminary biomarker discovery, biomarker discovery, safety and efficacy clinical trials, large-scale clinical trials, and continued surveillance (Figure 3).

Clinical biomarkers are validated in retrospective or prospective analyses and biomarker trials, or drug trials [12]. The validation process should reflect the clinical performance of the biomarker(s), based on existing clinical data, new clinical data, literature review findings, or current clinical knowledge [12]. Moreover, it should be an evidentiary and statistical process that aims to link

a biomarker to specific biological, pathological, pharmacological (i.e., drug effect), or clinical endpoints [22].

## Internal Validation

During the biomarker discovery phase, the main focus is to identify biomarkers that distinguish between the treatment and control groups or correlate with the clinical observation of interest. Prior to this process, and depending on the sample size, plans can be made to allocate the subjects into two individual cohorts for the purpose of internal validation. It is important, however, to distinguish this allocation/splitting process from that used for the purpose of internal validation of classifiers:

- Internal validation of candidate biomarkers identified in the discovery cohort
  - Use of different platforms
  - Use of different statistical methods
- Internal validation of classifiers
  - Split-sample method
  - Some form of cross-validation method

There are several different approaches to separate the initial pool of patients or samples for internal validation and creating classifiers; the traditional and alternative approaches are outlined here. Traditionally, a *discovery* and a *validation* cohort are created (Figure 4). Genomic biomarker candidates may first be identified in the discovery cohort using microarray analysis, for example. One way of identifying a panel of biomarkers from the candidates is by use of classification methods. The samples from the discovery data set can be split into a *training* set, used to find a classifier, and a *test* set, used for internal validation of classifiers. The classifier is an equation combining the expression values of some of the candidate markers that best distinguish the treatment from the control group. Once the classifier has been developed, the panel of biomarkers may be validated again in the validation cohort before the external validation phase is carried out. In this sense, the data obtained in the validation cohort serves purely as an additional test set.

Although the traditional internal validation model is simple and logical, it may not be the most applicable strategy in the real world, given the complexity of most biomarker research today. There are two potential weaknesses to this approach. First, separating available samples into discovery and validation cohorts from the outset might unintentionally restrict the use of the samples or data to their arbitrary labels. In reality, during the developmental phase of biomarker research, different statistical analyses are often carried out to identify potentially useful biomarkers. Depending on the types of comparisons being made, a sample (or portion of data) could be used for discovery pur-

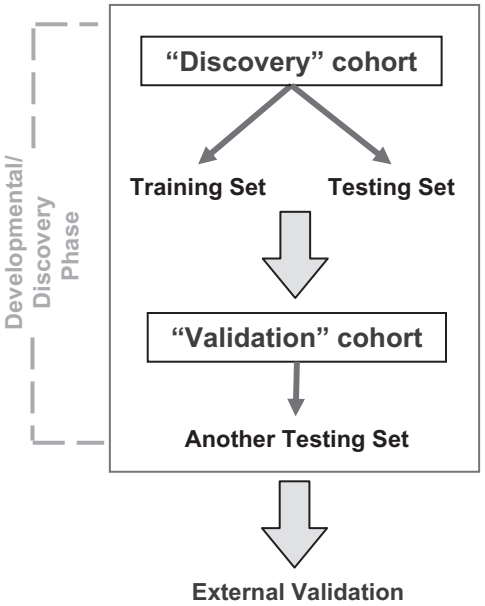


**Figure 4** Traditional internal validation approach. Traditional approaches to internal validation rely on two cohorts: discovery and validation. The discovery cohort is typically separated into training and test sets for the development of classifiers based on biomarker(s) of interest. The classifiers are then tested in a separate cohort internally before external validation.

poses in one analysis while it is used for validation in another. Second, in the case where the classifier fails to show acceptable accuracy in the initial discovery cohort testing set, one might consider incorporating additional data and reanalyzing the training set. However, collecting additional patient samples or information may not always be possible; this type of situation may warrant the reallocation of data from the validation to the discovery cohort training set, in order to redevelop the classifier.

An alternative model of internal validation may be particularly useful for smaller sample sizes (Figure 5). Similar to the traditional approach, the subjects are divided into two separate groups. However, neither cohort is marked strictly as discovery or validation, to minimize potential confusion in later analyses and maximize the utility of available data. As an example, a classifier can be generated and internally validated by creating training and testing sets from cohort 1. The same classifier can then be validated again in a separate cohort (cohort 2). Based on the outcome, the R&D team may decide that the classifier is ready for external validation or that a more robust classifier needs to be developed. In the latter case, a new classifier can be created by utilizing cohort 2 as the discovery cohort. Once the new classifier has been validated internally in cohort 2, it can be evaluated again in cohort 1, which is used

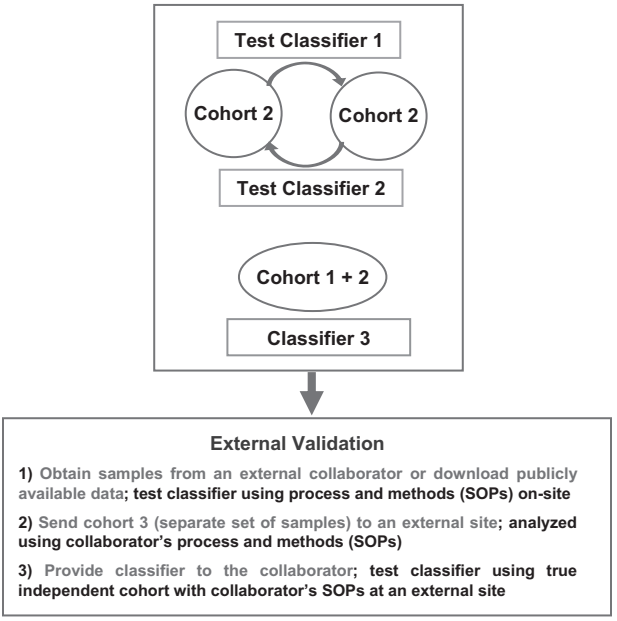


**Figure 5** Alternative internal validation approach, which may be particularly useful for smaller sample sizes. Like the traditional approach, the subjects are divided into two separate cohorts. However, in this model, a classifier can be developed from either cohort, or whenever necessary, from two cohorts combined to improve the robustness of the classifier.

entirely as another testing set. Finally, the same decision process is repeated: Is the classifier robust enough to stand up to the scrutiny of external validation? Are additional data required to create a new classifier? In the latter situation, it may be necessary to combine cohorts 1 and 2 to develop larger training and test sets for internal validation. This may, in turn, help create a new classifier that is potentially more robust and applicable to a larger intended patient population. The main advantage of this model is that it provides many possible internal validation approaches, particularly in a project where sample size is small. This gives flexibility to the overall validation process and allows decisions to be made based on the results generated at each step.

During the development and testing of a biomarker panel, concurrent literature and technical validations may also take place. Depending on the characteristics of the biomarker, various technology and analytical techniques may be applied. For example, quantitative polymerase chain reaction (qPCR) may be used to validate the microarray data generated from cohort 1. Once the initial findings have been confirmed, such as the differential expression of particular genomic biomarkers, qPCR can also be used to test the candidate biomarkers in cohort 2.

There are a number of potential advantages to cross-platform or cross-technology validations. Relative to high-throughput technologies such as microarray chips, which are typically used for identifying genomic biomarker candidates, the use of low-throughput techniques such as qPCR may help reduce the cost of the validation process. Also, by applying different platforms and analytical methods, biomarkers that are found to be statistically significant across the various cohorts are less likely to be related to or influenced by platform-specific bias. More recently, studies have also suggested that the use of combined data from multiple platforms (i.e., genomics and proteomics) to assess potential biomarkers is far superior to those generated with one technical approach alone [21].

The internal validation strategy will be largely dependent on the characteristics of the biomarkers (i.e., genomic, proteomic, metabolomic), the intended use (i.e., prognostic or diagnostic), and the sample size (i.e., traditional or alternative model). Nonetheless, the general recommendation is that the preliminary or developmental studies should be large enough so that either split-sample validation or some form of cross-validation can be performed to demonstrate the robustness of the internally validated prediction [28,29].

## External Validation

The use of internal discovery and validation cohorts has helped studies to develop biomarker panels with impressive accuracy and precision for the predicted outcome [28]. However, internal validation should not be confused with external validation, which is typically performed on a cohort from a different geographical location and is meant to simulate a broader clinical application [28]. Internal validation, even with the use of "independent" cohorts, does not guarantee "generalizability."

In principle, the aim of external validation is to verify the accuracy and reproducibility of a classifier in a truly independent cohort with similar underlying condition, given a defined clinical context [30]. External validation is a crucial and essential step before a classifier is implemented in a larger clinical setting for patient management [30]. Like internal validation, the design of external validation processes will depend on the intended use of the biomarker. For predictive biomarkers, prospective trials may be necessary, as they are considered the gold standard [28,30]. Moreover, it has been argued that a biomarker is more readily tested in a prospective clinical trial once retrospective studies conducted at external institutions have consistently shown the ability for the biomarker to perform at the required levels of sensitivity and specificity [8].

To accelerate the translational process from external validation to phase I and II clinical trials, partnerships and collaborations are often established. In some circumstances, classifiers may need to be tested in intended patients at a collaborators' site or external institution in a prospective manner. In other cases, retrospective studies using patient samples may suffice. There are

several important factors to consider when conducting an external validation, regardless of the directionality of the study (i.e., prospective or retrospective). Many of these factors are also critical in designing clinical trials.

## USE OF STATISTICS AND BIOINFORMATICS

Due to the complexity and multifaceted nature of biomarker research, it is not uncommon, especially in the recent years, to see different "omics" techniques, bioinformatics, and statistical methods incorporated into biomarker research and development [21]. Thus, there has been a large increase in the number of possible approaches to analytical and clinical validations. As illustrated in Figure 1, the use of bioinformatics and statistics to validate biomarkers can accompany any preclinical or clinical stage. More specifically, they are applied to a greater degree of freedom during the initial phases of biomarker development. Ideally, by the time a biomarker enters large clinical trials or is put on the market under continued surveillance, its robustness should already have been tested with the use of a variety of statistical and bioinformatical approaches. Regardless, these computational approaches, although not very stringent, are fundamentally useful tools for ensuring the validity of results prior to using some of the more time-consuming and costly validation methods (i.e., external validation or clinical trials).

### Statistical Approaches

High-throughput "omics" technologies such as microarray-based expression profiling generate massive amounts of data. As such, new statistical methods are continually being developed to deal with this challenging issue. The availability of a plethora of statistical techniques, when used with proper precautions, has provided a relatively quick and inexpensive way to validate biomarker candidates. The trial-and-error process with different statistical methods is especially common during early stages of biomarker development. In the exploratory phase of a biomarker project, various computational and mathematical techniques, such as multivariate analysis or machine learning, are often utilized to detect differences between treatment and control groups or between patients with different clinical presentations [18].

Statistically distinctive genomic biomarkers identified during the exploratory phase by one method may subsequently be subjected to a different technique. Similarly, given the same set of samples and expression measurements, permutation can be carried out on the data set prior to repeating an analysis. Congruency between the results generated using different methods may ultimately translate to an increase in biomarker confidence. This is especially useful during the internal and external phases of validation, when greater statistical freedom is exercised for the purpose of identifying or ranking biomarkers. Ideally, by the time a panel of biomarkers is selected for use in

a clinical trial, specific algorithms and statistical approaches should be established. It has been suggested in a recent FDA presentation [12] that the expression patterns or algorithms should be developed and confirmed in at least two independent data sets (a training set and a test set, respectively) [12].

### Bioinformatical and Bibliographical Approaches

Potential biomarker candidates should be checked continuously using bioinformatical and bibliographical tools to provide biological and clinical context. This step should be performed in parallel to the stratified statistical approaches. Although bioinformatics originated from the field of genomics, it now plays an important role in connecting and integrating biological and clinical data from a variety of sources and platforms [21].

Biomarkers that fit accepted theory are more likely to be accepted readily by the research community and the public [14]. This is especially important when selecting and transitioning biomarkers from internal and external validation phases into clinical trials. Numerous bioinformatical tools, many of which are open-source, are available to accelerate this evidentiary process. Gene and protein biomarker candidates can be processed first and then grouped through gene ontology using tools such as FatiGO and AmiGO [31,32]. More sophisticated pathway-oriented programs such as Ingenuity and MetaCore are also potentially useful in assisting the R&D team to zone in on biomarkers belonging to pathways and processes relevant to the clinical endpoints of interest [33,34].

Another advantage of the bioinformatical approach is the ability to link clinical measurements across platforms and/or identify a unique set of molecular signatures. For example, during external validation on a separate cohort of patients, results might indicate that the biomarkers identified initially as statistically significant were individually unable to strongly predict the clinical endpoint. However, linking expression data across platforms (i.e., genomic and proteomic biomarkers) may help provide a more comprehensive understanding of the biology and establish a stronger correlation between the biomarker and the clinical presentation of the patient [21].

## FROM EXTERNAL VALIDATION TO CLINICAL TRIALS: THE IMPORTANCE OF COHORT, TECHNICAL, AND COMPUTATIONAL FACTORS

As biomarker candidates continue to pour out of research laboratories, it has become increasingly evident that validation is much more difficult and complex than discovery. There are a multitude of general and specific considerations and obstacles to address in order to validate biomarker candidates clinically, some of which were discussed earlier and some of which we discuss now.

Translation of a candidate biomarker from a discovery phase into the validation phase of clinical trials imposes huge organizational, monetary, and technical hurdles that need to be considered at the onset. The team needs to meet requirements in sample number and feature distribution (*cohort factors*), sample quality, accuracy and precision in handling and processing the samples (*technical factors*), and in the analysis (*computational factors*). As we will see, each of these steps is extremely important and very challenging at the same time.

## Cohort Factors

During any phases of external or clinical validation of biomarkers, bias can be introduced unintentionally. This is a major concern in the design, conduct, and interpretation of any biomarker study [35,36]. Starting with the population selection process, variations at a biologically level can lead to discernible differences in body fluid and tissue compositions and biomarker measurements [35]. As such, basic criteria such as gender, age, hormonal status, diet, race, disease history, and severity of the underlying condition are all potential sources of variability [35]. Moreover, patient cohort characteristics of the validation phase of a candidate biomarker must be representative of all patients for which the biomarker is developed. To reduce a bias requires the inclusion of hundreds of patients per treatment or disease arm. Not every patient is willing to give biosamples for a biomarker study, however, and only 30 to 50% of the patients in a clinical trial may have signed the informed consent for the biomarker analysis, presenting an important concern for statisticians. There is a risk for population bias, since a specific subset of patients may donate biosamples, hence skewing the feature distribution. Another risk factor that needs to be dealt with in some instances is the lack of motivation of clinical centers to continue a study and collect samples. Although it is sometimes possible to compensate for a smaller sample size through the use of different statistical methods, such as the use of local pooled error for microarray analysis, the analysis team needs to ensure that there is no patient selection bias by the center selection: Affluent centers and their patients may have different characteristics from those with other social backgrounds.

The lack of available samples or patient enrolment may ultimately translate to a decrease in biomarker confidence or the generalizability of the intended biomarker. In reality, collaboration will probably make biomarker validation more robust and economically feasible than working independently [37]. Since the issue of intellectual property (IP) agreement is minimal for the biomarker validation *process*, as it is not patentable, open interactions among steering committees of large trials or cohort studies should be encouraged [37,38].

## Technical Factors

***Sample Collection, Preparation, and Processing*** Even with the establishment of a relatively homogeneous cohort for external validation or

clinical trials, results from the intended biomarker assays are valid only if sample integrity is maintained and reproducible from sample collection through analysis [22]. Major components of assay and technical validation are:

- Reference materials
- Quality controls
- Quality assurance
- Accuracy
- Precision
- Sensitivity
- Specificity

Prior to the start of a validation phase, the team needs to decide on the sampling procedure, sample quality parameters, and sample processing. Reproducibility is the key to successful biomarker validation. It is important that standardized operating protocols (SOPs) for sample collection, processing, and storage be established to provide guidance for the centers and the laboratory teams. Nurses and technicians should be trained to minimize the variability in sample collection and handling.

Most biomarkers are endogenous macromolecules which can be measured in human biological fluids or tissues [39]. The collection process for these specimens seems straightforward. However, depending on the type of biomarker (genomic, metabolomic, or proteomic) and the collection methods, various factors may need to be taken into account when designing the SOPs [22]. Several examples are given in Table 1.

Processing the samples in a core facility reduces the risk of handling bias, since the same personnel would handle all samples in the most reproducible way possible. Once the samples are collected, systematic monitoring of their quality over time should also be established. Random tests can be conducted to ensure the short-term, benchtop, or long-term stability of the samples to uphold the integrity of the biolibrary [38].

**TABLE 1 Sample Considerations**

| Collection (Biological Fluids or Tissues) | Preparation and Processing | Storage |
|---|---|---|
| Type of needle | Dilution | Type of storage containers |
| Type of collection tube or fixation | Plasma or serum isolation | Temperature of storage |
| Location of collection | Temperature of processing | Duration of storage |
| Time of collection | Reagents used | |
| Status of patient | | |

Unfortunately, during external and clinical validations, it is likely that the independent collaborators utilize a completely different set of SOPs. Even though this could ultimately contribute to the robustness of the validation process, it may be useful to establish specific guidelines that would help minimize site-to-site variability. There are several ways to achieve this. Provision to the laboratories and centers of kits that contain all chemicals needed for the processing of the sample can potentially reduce the risk of batch effects. Similarly, as pointed out in Figure 5, it may be feasible to collect a separate set of samples internally (to minimize sample collection variability) but send them to a collaborator's site for external validation using independent processing and analytical SOPs.

It is of utmost importance that any deviation from the SOPs is noted. These features can then later be accommodated by statisticians for better modeling. Any available information on bias or deviation from protocols or batch processing is useful in the computational process. Excluding them may influence the decision as to whether a biomarker was or was not validated.

***Sustained Quality Assurance, Quality Control, and Validation of the Biomarker Tests or Assays*** The aforementioned cohort (i.e., patient selection) and technical factors (i.e., sample collection, processing, and storage) can all have a significant impact on any of the phased strategies to biomarker validation shown in Figure 1. However, in reality, validation of the biomarker test is just as important as the validation of the biomarker itself. To improve the chance of successful translation from external and clinical validation results to patient care, the analytical validity of the test (does the test measure the biomarker of interest correctly and reliably?) should be closely monitored along with the clinical validity of the biomarker (does the biomarker correlate with the clinical presentation?) [8,12]. Furthermore, to sustain the quality assurance and quality control between different stages of biomarker development, it may be necessary to carry out multiple analytical or technical validations when more than one platform is used.

## Computational Factors

In addition to the statistical and bioinformatical factors in the validation process described earlier, the team must also be aware that a vast number of samples impose a huge computational burden on software and hardware. This is especially true when high-throughput and high-performance technologies or high-density arrays need to be used for the validation process. Not considering this potential issue may ultimately be costly in terms of money and time. To deal with the massive amount of data generated from these technologies, new statistical techniques are continuously being developed. Statistical methods should include those that were used successfully in the discovery phase. If the performance is not as good in prior analyses, the refinement of algorithms will be necessary ("adaptive statistics").

### Other Challenges of Clinical Biomarker Validation

There are many additional barriers and concerns for clinical validation of biomarkers:

- Choice of matrix (readily accessible, effect on biomarker concentrations)
- Variability (interindividual and intraindividual)
- Preparing calibration standards
- Implementations of quality control to assure reproducibility
- Limited availability of clinical specimens
- Heterogeneity of biomarkers (isoforms, bound states)
- IP protection (lack of collaboration)
- Lack of clear regulatory guidance

As discussed, the discovery and validation processes involve multicenter studies with large patient cohorts and technical equipment, steered by a vast number of staff members. The enormous costs for these studies can in most cases be covered only by consortia, consisting potentially of academic centers and pharmaceutical companies. The establishment of such consortia often has to overcome legal hurdles and IP issues, which may be a time-consuming process. As mentioned above, establishment of a biomarker may involve the recruitment of hundreds of patients per treatment arm. Both recruitment time and success rate are unpredictable at the beginning of the study phase.

It is during the initial phase that the team needs to determine what really constitutes a good biomarker. Key questions in this decision include: What is the threshold for decision making that needs to be established to call a biomarker "useful"? How do we achieve a "robust" biomarker: a marker that is not easily influenced by such factors as location and personnel? How economical does the biomarker test need to be for it to be used? Which biomarker matrix should be selected, since this may affect future validation possibilities? Accessible matrices such as urine or blood with limited or known concentration variability are ideal.

## BIAS AND VARIABILITY: KEY POINTS REVISITED

To summarize the sections above, variability is a major obstacle in biomarker validation, regardless of the matrix, the type of biomarker, and its use. As noted above, there are two types of variability: intraindividual variability, which is usually related to lab techniques, sample timing, drug effects, and within-subject biological processes, and interindividual variability, resulting from different individual responses involving multiple genetic factors [40]. Biological variability may be difficult to assess, but it is important to control

**TABLE 2 Fundamental Concerns in Biomarker Validation**

| | Concern | Solutions |
|---|---|---|
| Overfitting | False positives and negatives<br>High sensitivity and specificity found but fail on independent validation sets | Increase sample size<br>Statistical approaches<br>Receiver operator characteristic curves |
| Bias | Misidentification of differences between samples | Control for confounding factors |
| Generalizability | Can results be applied to appropriate clinical populations? | Representative validation cohorts |

for this factor. A statistical correlation to a clinical endpoint for a candidate biomarker cannot be determined without assessing biological variability, as the overall noise of a sample is a sum of both analytical and biological variability [6]. A biomarker with wide biological variability or time fluctuations that are difficult to control may be rejected [6]. Diurnal variability may require sample pooling or collection at the same time of day [6]. Biomarkers have diverse molecular structures, including possible bound states, which also need to be considered as influences on variability. There are specific considerations for any biomarker validation study. Overfitting, bias, and generalizability are three of the most fundamental concerns pertaining to clinical biomarker validation (Table 2) [41]. With the introduction of high-throughput discovery strategies, overfitting has become a particular fear. These discovery platforms are designed to measure countless analytes, and there is therefore a high risk of false discovery. When a large number of variables are measured on a small number of observations to produce high sensitivity and specificity, the results may not be reproducible on independent validation sets. Some biomarker candidates may be derived simply due to random sample variations, particularly with inadequate sample sizes [26]. A false positive may be thought of as critical part of a disease process, when in fact it is either associated only loosely or coincided randomly with disease diagnosis or progression [3]. As mentioned earlier in the chapter, a biomarker may correlate with a disease statistically but not prove to be useful clinically [8,42]. Increasing sample size and use of receiver operator characteristic curves may help overcome this concern of overfitting.

Bias is another major concern during biomarker validation, as there is often potential for misidentifying the cause of the differences in biomarkers between samples. Confounding variables such as age, race, and gender should be controlled for either through statistical modeling or validation study design to limit the effects of bias. Since validation cohorts require suitable diversity for widespread utility, bias may be difficult to avoid entirely through study design.

Similar to bias, most issues pertaining to the generalizability of a biomarker across clinical populations can be addressed through careful consideration of cohort selection. Cohort factors were discussed briefly above. To increase generalizability, the later phases of validation should include more rigorous testing of potential interfering endogenous components by including more diverse populations with less control of these confounding variables [22]. For example, in later-stage clinical trials there should be less control of diet and sample collection and more concomitant medications and co-morbidities [22]. This will allow a biomarker to be used in more clinically diverse situations.

## KEY MESSAGES

Many of the biomarkers in current clinical use have become accepted via debate, consensus, or merely the passage of time [13]. This rather unofficial establishment of biomarkers in the past has been very inefficient. Importantly, biomarkers can no longer become accepted in this way, as they would fail to meet the current regulatory standards of modern medicine. Contemporary biomarkers must be tested in highly regulated human clinical trials [43]. To date, the clinical trial process has not been very efficient, and a typical biomarker life cycle from discovery to clinical use may take decades. For example, the evolution of prostate-specific antigen (PSA) as a biomarker for prostate disease diagnosis and monitoring took 30 years for regulatory approval by the FDA [44]. In order to expand on the biomarker repertoire used currently in clinical practice, the acceptance process for new biomarkers needs to become much more efficient and cost-effective.

The general problem of lack of regulatory guidance is very likely to be addressed formally by regulatory bodies in the near future. Problems such as the availability of samples will also probably improve as collaborations are established and the ethical issues surrounding biobanks are clarified. Even if a biomarker candidate fails during validation, much may be learned regarding the pathophysiology of the disease and the corresponding drug effects during the process [9]. However, without successful validation and integration of biomarkers into clinical use, much of the research effort, particularly in terms of biomarkers for drug development, can be futile.

## REFERENCES

1. Biomarkers Definition Working Group (2001). Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. *Clin Pharmacol Ther*, 69(3):89–95.
2. Colburn WA (1997). Selecting and validating biologic markers for drug development. *J Clin Pharmacol*, 37(5):355–362.

3. Colburn WA (2003). Biomarkers in drug discovery and development: from target identification through drug marketing. *J Clin Pharmacol*, 43(4):329–341.
4. Boguslavsky J (2004). Biomarkers as checkpoints. *Drug Discov Dev*, Sept.
5. Hunt SM, Thomas MR, Sebastian LT, et al. (2005). Optimal replication and the importance of experimental design for gel-based quantitative proteomics. *J Proteome Res*, 4(3):809–819.
6. Lee JW, Figeys D, Vasilescu J (2007). Biomarker assay translation from discovery to clinical studies in cancer drug development: quantification of emerging protein biomarkers. *Adv Cancer Res*, 96:269–298.
7. Listgarten J, Emili A (2005). Statistical and computational methods for comparative proteomic profiling using liquid chromatography–tandem mass spectrometry. *Mol Cell Proteom*, 4(4):419–434.
8. Bast RC Jr, Lilja H, Urban N, et al. (2005). Translational crossroads for biomarkers. *Clin Cancer Res*, 11(17):6103–6108.
9. Kuhlmann J (2007). The applications of biomarkers in early clinical drug development to improve decision-making processes. *Ernst Schering Res Found Workshop*, 59:29–45.
10. Mandrekar SJ (2005). Clinical trial designs for prospective validation of biomarkers. *Am J Pharmacogenom*, 5(5):317–325.
11. Lachenbruch PA, Rosenberg AS, Bonvini E, Cavaille-Coll MW, Colvin RB (2004). Biomarkers and surrogate endpoints in renal transplantation: present status and considerations for clinical trial design. *Am J Transplant*, 4(4):451–457.
12. Harper CC (2007). *FDA Perspectives on Development and Qualification of Biomarkers, in Rediscovering Biomarkers: Detection, Development and Validation*, GTCbio; San Diego, CA.
13. Goodsaid F, Frueh F (2006). Process map proposal for the validation of genomic biomarkers. *Pharmacogenomics*, 7(5):773–782.
14. Bonassi S, Neri M, Puntoni R (2001). Validation of biomarkers as early predictors of disease. *Mutat Res*, 480–481:349–358.
15. Wagner JA (2002). Overview of biomarkers and surrogate endpoints in drug development. *Dis Markers*, 18(2):41–46.
16. Wagner JA, Williams SA, Webster CJ (2007). Biomarkers and surrogate end points for fit-for-purpose development and regulatory evaluation of new drugs. *Clin Pharmacol Ther*, 81(1):104–107.
17. Goodsaid F, Frueh F (2007). Biomarker qualification pilot process at the US Food and Drug Administration. *AAPS J*, 9(1):E105–E108.
18. Baker M (2005). In biomarkers we trust? *Nat Biotechnol*, 23(3):297–304.
19. Benowitz S (2004). Biomarker boom slowed by validation concerns. *J Natl Cancer Inst*, 96(18):1356–1357.
20. Lee JW, Weiner RS, Sailstad JM, et al. (2005). Method validation and measurement of biomarkers in nonclinical and clinical samples in drug development: a conference report. *Pharm Res*, 22(4):499–511.
21. Ilyin SE, Belkowski SM, Plata-Salaman CR (2004). Biomarker discovery and validation: technologies and integrative approaches. *Trends Biotechnol*, 22(8): 411–416.

22. Lee JW, Devanarayan V, Barrett YC, et al. (2006). Fit-for-purpose method development and validation for successful biomarker measurement. *Pharm Res*, 23(2):312–328.
23. Peck RW (2007). Driving earlier clinical attrition: If you want to find the needle, burn down the haystack. Considerations for biomarker development. *Drug Discov Today*, 12(7–8):289–294.
24. Groopman JD (2005). Validation strategies for biomarkers old and new. *AACR Educ Book*, 1:81–84.
25. Normolle D, Ruffin MT IV, Brenner D (2005). Design of early validation trials of biomarkers. *Cancer Inf*, 1(1):25–31.
26. Jarnagin K (2006). ID and Validation of biomarkers: a seven-fold path for defining quality and acceptable performance. *Genet Eng Biotech News*, 26(12).
27. O'Connell CD, Atha DH, Jakupciak JP (2005). Standards for validation of cancer biomarkers. *Cancer Biomarkers*, 1(4–5):233–239.
28. Simon R (2005). Roadmap for developing and validating therapeutically relevant genomic classifiers. *J Clin Oncol*, 23(29):7332–7341.
29. Simon R (2005). Development and validation of therapeutically relevant multi-gene biomarker classifiers. *J Natl Cancer Inst*, 97(12):866–867.
30. Bleeker SE, Moll HA, Steyerberg EW, et al. (2003). External validation is necessary in prediction research: a clinical example. *J Clin Epidemiol*, 56(9):826–832.
31. Al-Shahrour F, Diaz-Uriarte R, Dopazo J (2004), FatiGO: A Web tool for finding significant associations of gene ontology terms with groups of genes. *Bioinformatics*, 20(4):578–580.
32. AmiGO. http://amigo.geneontology.org/cgi-bin/amigo/go.cgi.
33. Ingenuity Pathways Analysis. http://www.ingenuity.com/products/pathways_analysis.html.
34. MetaCore Gene Expression and Pathway Analysis. http://www.genego.com/metacore.php.
35. Moore RE, Kirwan J, Doherty MK, Whitfield PD (2007). Biomarker discovery in animal health and disease: the application of post-genomic technologies. *Biomarker Insights*, 2:185–196.
36. Ransohoff DF (2005). Bias as a threat to the validity of cancer molecular-marker research. *Nat Rev Cancer*, 5(2):142–149.
37. McCormick T, Martin K, Hehenberger M (2007). *The evolving role of biomarkers: focusing on patients from research to clinical practice*. Presented at the IBM (Imaging) Biomarker Summit III, IBM Corporation; Nice, France.
38. Maruvada P, Srivastava S (2006). Joint National Cancer Institute–Food and Drug Administration workshop on research strategies, study designs, and statistical approaches to biomarker validation for cancer diagnosis and detection. *Cancer Epidemiol Biomarkers Prev*, 15(6):1078–1082.
39. Colburn WA, Lee JW (2003). Biomarkers, validation and pharmacokinetic–pharmacodynamic modelling. *Clin Pharmacokinet*, 42(12):997–1022.
40. Mayeux R (2004). Biomarkers: potential uses and limitations. *NeuroRx*, 1(2): 182–188.

41. Early Detection Research Network. *Request for Biomarkers*. Attachment 2: Concepts and Approach to Clinical Validation of Biomarkers: A Brief Guide. http://edrn.nci.nih.gov/colops/request-for-biomarkers.
42. Katton M (2003). Judging new markers by their ability to improve predictive accuracy. *J Natl Cancer Inst*, 95:634–635.
43. NCI (2006). *Nanotechnology-Based Assays for Validating Protein Biomarkers*. NCI Alliance for Nanotechnology in Cancer, Bethesda, MD, Nov.–Dec.
44. Bartsch G, Frauscher F, Horninger W (January 2007). *New efforts in the diagnosis of prostate cancer*. Presented at the IBM (Imaging) Biomarker Summit III, IBM Corporation; Nice, France, Jan.

# PREDICTING AND ASSESSING AN INFLAMMATORY DISEASE AND ITS COMPLICATIONS: EXAMPLE FROM RHEUMATOID ARTHRITIS

Christina Trollmo, Ph.D., and Lars Klareskog, M.D., Ph.D.
*Karolinska Institute, Stockholm, Sweden*

## INTRODUCTION

Chronic inflammatory diseases include a number of rheumatic, neurological, dermatological, and gastrointestinal diseases, which develop as a result of immune and inflammatory reactions. These perpetuating reactions ultimately cause the clinical symptoms, which are subsequently used to classify the symptoms as a "disease." Analyzing the disease course longitudinally reveals several distinct steps during disease progression, for which the presence of biomarkers is of importance, both for identification of disease status and for prediction of disease course and treatment options. However, biomarkers per se are not always available today. We have chosen here to discuss disease characteristics and potential biomarkers in one common chronic inflammatory disease, rheumatoid arthritis (RA), which affects approximately 0.5 to 1% of the population worldwide. In this chapter we focus on the following factors:

1. Onset of disease in order to discuss the questions in whom and why the disease occurs, and whether and how onset can be predicted

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

from biomarkers that are present before the occurrence of clinical symptoms

2. Progression of disease with respect to development of joint destruction, but also other with respect to complications such as extra-articular manifestations, cardiovascular events, and lymphoma development in this patient group
3. Treatment of the individual patient and of specific symptoms and disease manifestations
4. Selection of patients in clinical trials of new drugs

## RHEUMATOID ARTHRITIS DISEASE PROCESS

Rheumatoid arthritis (RA) is a disease defined by seven criteria, with four that should be fulfilled to make the diagnosis (Table 1). These criteria have been useful in harmonizing clinical trials and clinical practice. However, they are not based on what is now known about etiology or pathogenesis, and they are not too helpful in selecting treatment for the single patient. Hence, there are needs to redefine the diagnosis for RA and related diseases, first to define entities more related to distinct etiologies and pathogenetic mechanisms, then to use such new entities for stratification and selection of patients in clinical trials and clinical practice.

Basic features of the immune and inflammatory process in RA are, on the one hand, processes that can be identified in the peripheral circulation, initially

**TABLE 1 Classification Criteria for Rheumatoid Arthritis**[a]

1. Morning stiffness in and around joints lasting at least 1 hour before maximal improvement
2. Soft tissue swelling (arthritis) of three or more joint areas observed by a physician
3. Swelling (arthritis) of the proximal interphalangeal, metacarpophalangeal, or wrist joints
4. Symmetric swelling (arthritis)
5. Rheumatoid nodules
6. Presence of rheumatoid factor
7. Radiographic erosions and/or periarticular osteopenia in hand and/or wrist joints

*Source:* Arnett FC, Edworthy SM, Bloch DA, et al. (American Rheumatism Association) (1988). The American Rheumatism Association 1987 revised criteria for the classification of rheumatoid arthritis. *Arthritis Rheum*, 31:315–324.

[a]Criteria 1 through 4 must have been present for at least six weeks. Rheumatoid arthritis is defined by the presence of four or more criteria, and no further qualifications (classic, definite, or probable) or list of exclusions are required. These criteria demonstrated 91 to 94% sensitivity and 89% specificity for RA compared with non-RA rheumatic disease control subjects.

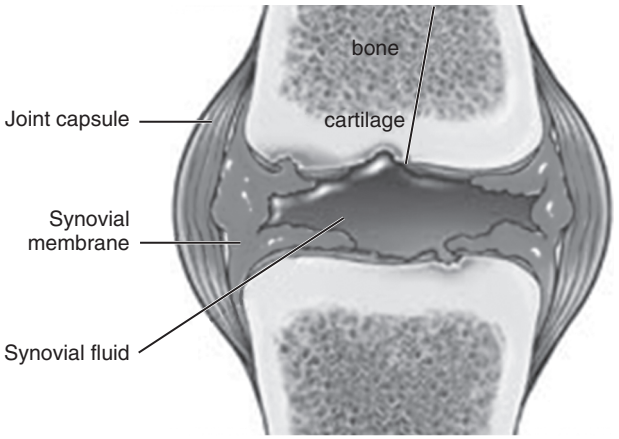


**Figure 1** Inflamed RA joint. In healthy joints a thin synovial membrane lines the joint capsule and the synovial fluid. In the RA joint both the synovial membrane and the synovial fluid are infiltrated by inflammatory cells, leading to tender and swollen joints. The synovial membrane also "grows" over the cartilage, aiding in the process of cartilage and bone desctruction.

rheumatoid factors (RFs), and on the other hand, processes in the inflamed tissue, mainly the joints (Figure 1). Rheumatoid factors, identified almost 70 years ago, are part of the diagnostic criteria for RA. Being present in some 50 to 60% on incident RA cases and increasing over time with active disease, these auto-antibodies are also seen in many non-RA conditions and are thus not very specific for the disease.

Rheumatoid factors have never been shown to be pathogenic by themselves, neither in patients nor in experimental models. They are thus seen mainly as biomarkers of importance for diagnosis and prognosis of a more severe disease course, but not necessarily directly involved in disease pathogenesis. Joint inflammation in RA is focused on synovial inflammation, in many cases associated with cartilage destruction and concomitant erosions in bone. This inflammation has been studied in large detail over the years, demonstrating that its features are common to many other types of chronic inflammation in other tissues. No real RA pathognomonic features have yet been identified, the most unique feature identified so far being the way the inflammatory cells and molecules attack and destroy bone and cartilage.

Having the major features of RA, the synovial joint inflammation and presence of RF being typical but by no means unique for RA, there is an obvious need to define more specific features of the disease. The identification of such features would enable us to search for a better understanding of the pathogenesis of RA, a more accurate diagnosis based on biomarkers, and more specific treatments.

## STUDIES ON ETIOLOGY AND PATHOGENESIS AS A BASIS FOR DEVELOPMENT OF BIOMARKERS FOR DIAGNOSIS AND PROGNOSIS IN RA

Any understanding of a complex, partly genetic disease is based on an understanding of how genes and environment interact in giving risk to immune reactions that contribute to the joint destruction and other inflammatory reactions in RA. In healthy subjects, the major role for the immune system and subsequent inflammatory reactions is to defend us against pathogens, but in RA the immune system has partly changed focus to attack our own tissues, primarily the joints, and is thus denoted as an autoimmune disease.

### Genes

There is strong evidence to support a significant genetic component to the susceptibility of RA. Twin studies clearly demonstrate an overrepresentation of disease concordance in monozygotic twins (12 to 20%, depending on study) compared to dizygotic twins (4 to 5%) and the general population (0.5 to 1%). Analysis on the gene level demonstrates the strongest genetic association with genes within the HLA region, specifically the *HLA-DRB1* gene. Its gene products, the MHC class II molecules, were described in the 1970s to be present on cells in the inflamed joint, allowing antigens (part of proteins, generally pathogens, but in autoimmune diseases self-proteins) to be presented to the immune system and subsequently to trigger inflammatory reactions (Figure 2). Serologic and later genetic typing of the various MHC molecules revealed that a few allotypes of *HLA-DRB1* were overrepresented in RA patients. A closer analysis demonstrated even identical amino acid

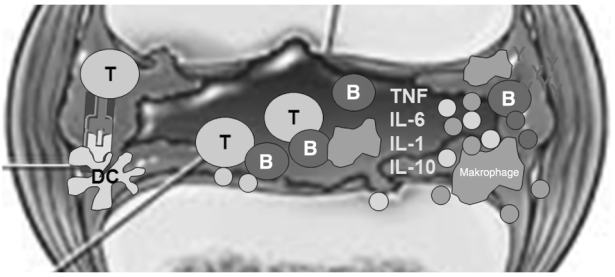


**Figure 2** Inflammatory cells in the RA joint. A number of immune cells have infiltrated the joint and local production of inflammatory mediators, including cytokines and antibodies, occurs. The synovial fluid, which functions as a cushion during joint movements, is in a healthy joint acellular. Illustrated to the left is the presentation of an antigen by the dendritic cell to a T-cell; the yellow connector is a MHC class II molecule. Cytokines, released from the immune cells, function as signaling molecules between cells. (*See insert for color reproduction of the figure.*)

sequences in those regions of the MHC class II DRβ1 chain, which are in contact with the antigen and the T-cell receptor on the T-cell mediating the specific immune responses. This amino acid sequence is termed *shared epitope*. The nature of the specific immune reactions mediated by the MHC class II molecules has, however, been surprisingly difficult to define. Identifying the specific antigens would help to identify the autoimmune trigger and make it possible to interfere specifically to break the autoimmune reactions. In total, the HLA region contributes 30 to 50% of the genetic component for RA in Caucasians. It appears that other MHC class II genes and allotypes may contribute to RA in other ethnic groups, and much is still not known about the contributions of different genes and allotypes within the MHC concerning susceptibility and disease course in RA.

Only much more recently, a second genetic risk allele has been identified in populations of European descent. This minor allele of a nonsynonomous single-nucleotide polymorphism (SNP) in the protein tyrosine phosphatase nonreceptor 22 (*PTPN22*) gene confers the second-largest genetic risk to the development of RA, with an odds ratio of about 1.8. *PTPN22* encodes the intracellular protein lymphoid tyrosine phosphatase, which plays a central role in immune responses by playing an integral part in signal transduction and T-cell receptor signaling pathway and inhibiting T-cell activation. This variant was first demonstrated for type 1 diabetes, but was soon confirmed in a number of autoimmune diseases, including RA, systemic lupus erythematosus (SLE), and Grave disease. However, other autoimmune diseases show no association with this SNP, suggesting subsets of autoimmune diseases to be defined accordingly.

The recent introduction of whole genome-wide association studies allowed, for the first time, good coverage of common variations in the human genome. Hundreds of thousands of SNPs in thousands of samples were genotyped and compared. Interestingly, studies focusing on RA again demonstrated the strongest effects for the two well-documented RA susceptibility genes *HLA-DRB1* and *PTPN22*. Other genes (Table 2) make a more modest contribution to susceptibility. Future studies have to identify remaining, probably smaller, genetic effects and how all the genetic effects interact with each other as well as with environmental factors in inducing and perpetuating the disease.

**TABLE 2 Contribution of Genetic Risk Factors in Rheumatoid Arthritis**

| Gene | Odds Ratio |
|---|---|
| *HLA* | 6.4 |
| *PTPN22* | 1.8 |
| *6q23* | 1.2 |
| *STAT4* | 1.2–1.4 |
| *TRAF1/C5* | 1.1–1.4 |

## Environment

Information on environmental factors important for the development, perpetuation, or course of RA is surprisingly scarce. Smoking is the only conventional environmental factor that has been linked reproducibly to an increased risk of developing RA. Other exposures, such as silica dust and mineral oils, have been reported in a few studies. It has not yet been possible to verify frequently hypothesized stimuli such as microbial infections with the methods used to date. Smoking was initially considered as an unspecific risk factor, of interest mainly from a public health perspective. However, newer studies indicate that smoking is a specific trigger of RA, as discussed in more detail below.

## Immunity

Studies on specific immune reactions in RA have been confined almost entirely to those involving autoantibodies. As described above, they were initially restricted to the measure of rheumatoid factors, but more recently, antibodies specific for citrullinated proteins have been shown to be of great importance. This is discussed in detail below, since their presence covariates with genetic and environmental factors, providing an important tool in subgrouping patients into different entities of RA. In the process of joint inflammation and cartilage and bone destruction, many different cells and molecules of the immune system participate. Some of them are illustrated in Figure 2. However, even if some of these inflammatory processes are revealed in detail, a specific trigger has still not been identified.

In recent years, advances within the field of cytokine regulation and cytokine-directed therapy have largely dominated the research field of RA, illustrating how therapeutic progress is possible even though the role of adaptive immunity in the disease is not fully understood. Cytokines are soluble molecules that mediate the communication between cells of the immune system but also with other cells of the body, such as the endothelium. Interestingly, the first cytokine that was targeted, tumor necrosis factor (TNF), belongs to the innate immune system. Blocking IL-1, another cytokine belonging to the innate immune system, has not proven to be as effective as TNF blockade for the majority of the RA patients. Recent clinical trails blocking a third cytokine in this family, IL-6, show promising results. This cytokine exerts effects within both the innate and the adaptive immune systems.

Temporarily eliminating B-cells, which are the producers of antibodies, has also proven a successful therapy. A third alternative, blocking the interaction between cells presenting antigens and T-cells, has resulted in an approved therapy. Together, these treatment-based data also demonstrate the significant role of the various parts of the innate as well as adaptive immune system for disease progression.

### Citrulline Immunity

The presence of antibodies specifically identifying citrullinated antigens, specifically termed *antibodies to citrullinated protein antigens* (ACPAs) are a strong predictor of developing RA. Approximately 60% of all RA patients carry such antibodies. These antibodies are highly specific for the disease; that is, they are rare in the normal population (<2%) and are also quite rare in other inflammatory conditions. Almost all patients who carry ACPAs developed them before disease onset. ACPAs are directed against proteins, which as the name implicates, contain the amino acid citrulline. Citrulline occurs only upon posttranslational modification of the amino acid arginine by enzymatic conversion. Even if these auto-antibodies are more specific for RA than for RF, it is important to ask the question if their presence is an epiphenomenon or is causally related to the disease. This helps in further dissection of the pathophysiology of the disease and in the development of new treatments. Here, animal models provide a tool to test the hypothesis of citrullination to be causative of the disease. One example is arthritis induced in the rat by immunization with collagen type II, a protein expressed almost exclusively in the joint. Immunization with the citrullinated form of collagen type II resulted in a more severe arthritis than immunization with the same noncitrullinated protein. In a mouse model, where transfer of antibodies directed toward collagen induces a mild arthritis, the addition of antibodies directed against citrullinated fibrinogen enhanced the arthritis. Thus, in experimental arthritis models citrullination can change the immunogenicity of self-antigens, and some citrullinated proteins may contribute to arthritis development.

From the data above we can conclude that immunity to citrullinated proteins may play a pathogenic role in a subset of RA patients; however, so far we have not identified the specific proteins that are targeted by the immune system. How do we then measure ACPAs? Test kits have been developed which contain cyclic citrullinated peptides (CCPs), where the amino acid citrulline is flanked by a variety of amino acids to mimic many different peptide sequences. Thus, the test measures anti-CCP antibodies, and this is what is generally reported in patient files. In summary, the presence of ACPAs in the circulation is more predictive for RA than high blood pressure is for cardiovascular disease and thus can be considered as a biomarker for a subset of RA patients.

## BETTER DIAGNOSIS BY MEANS OF BIOMARKERS AND GENETICS

By the late 1970s, the MHC class II genes were already identified as a major genetic risk factor for RA. However, only the presence of specific MHC class II alleles does not result in RA, since the disease-associated *HLA-DRB1* alleles are common also in healthy subjects. But combining genetic studies

with ACPA analysis, the genetic association of RA to *HLA-DRB1* SE has recently been shown to be confined entirely to the ACPA-positive subset of RA. In contrast, ACPA-negative RA may be associated with a unrelated *HLA-DR* allele, HLA-DR3. Also a second major genetic risk factor for RA, the polymorphism in the *PTPN22* gene was also shown to be associated only with the ACPA-positive disease.

In contrast, other genetic risk factors, in particular variations in the interferon regulating factor 5 (IRF-5), but also polymorphisms in a newly identified risk gene in the C-type lectin complex, were associated exclusively with ACPA-negative disease. Taken together, the descriptive studies of disease course and genetic linkages strongly indicate that ACPA reactivity splits RA into two major and clinically relevant subsets of disease and thus becomes an important biomarker for one subgroup of RA. Practically, ACPA-positive and ACPA-negative RA should be treated as separate entities when studying the molecular pathophysiology of RA and probably also when selecting treatment for some treatments for the single patient (Figure 3).

Including smoking habits in these combined analyses provides an even clearer risk assessment for developing RA. The relative risk of developing ACPA-positive RA was shown to be over 20 times higher for smokers carrying two copies of the *HLA-DRB1* SE alleles than for nonsmokers with no SE alleles. On the other hand, no increased risk was discerned from smoking concerning development of ACPA-negative RA. This leads directly to the question of whether smoking can trigger an ACPA immunity. Recent studies suggest that this could be the case since cells from bronchoalveolar lavage from smokers, but not from nonsmokers, contain large amounts of citrullinated proteins, and smoking can activate macrophages and thus enhance their antigen-presenting capacity.

The disease onset of RA can be gradual, and thus patients can have swollen and tender joints without being given the diagnosis RA, which is based on

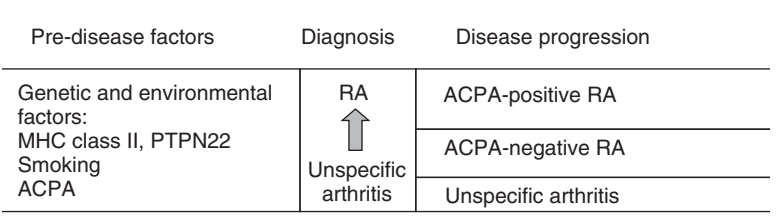


**Figure 3** Rheumatoid arthritis is a multifactorial disease. A number of predisease factors have been identified, which are of importance for the later disease course but not enough to initiate clinical disease. However, the trigger for clinical disease onset has still not been identified. The presence or absence of anticitrulline protein antibodies (ACPAs) distinguishes two forms of RA. Some patients will have swollen and tender joints but not fulfill four of the seven criteria for RA and thus carry an unspecific arthritis.

achieving four of seven specific criteria. These patients are classified as *undifferentiated arthritis* (UA), having an inflammatory arthritis for which no specific diagnosis can be made. An important question is if these patients should be treated in order to ameliorate disease progression to RA and retard radiographic joint damage or even prevent the development of RA in these patients. A recent double-blind, placebo-controlled, randomized clinical trial indicated that methotrexate (MTX) treatment as used for RA patients can postpone progression to RA and retard radiographic progression but not prevent the development of RA from UA. Subgroup analysis revealed that the beneficial outcome was most pronounced in patients with anti-CCP antibodies. In contrast, in the anti-CCP-negative subgroup the effect of MTX on the development of RA, the radiographic progression, and even the signs and symptoms was not demonstrable. Thus, this study provides evidence of distinguishing between anti CCP-positive and CCP-negative arthritis, even before the diagnosis RA is met, and therefore also becomes an important biomarker for undifferentiated arthritis.

The identification of predisease biomarkers might in the future lead to treatments which can be given to healthy subjects before the clinical diagnosis of RA and so prevent disease onset. An example already applied in medicine is the treatment of high blood pressure to prevent cardiovascular disease.

## BETTER PROGNOSTIC TOOLS WITH THE HELP OF BIOMARKERS AND GENETICS

Rheumatoid arthritis affects not only joints but also possibly, inflammation in other organs. Manifestations include rheumatoid nodules, secondary Sjögrens syndrome, rheumatoid lung disease, pleuritis, pericarditis, vasculitis, neuropathy, Felty syndrome, and severe eye disease, and are commonly summarized as extraarticular manifestations. Different co-morbidities also occur during the disease course. Cardiovascular morbidity and mortality occur at rates greater than would be expected from the profile of established risk factors. The average risk of lymphoma is increased in RA; however, a more detailed analysis indicates that the risk is increased primarily in patients with the most severe disease (Figure 4).

### Extraarticular Disease Manifestations

Extraarticular disease manifestations are more likely to occur in patients with more severe joint disease. Rheumatoid nodules often precede the onset of severe RA and thus can be considered a biomarker for severe disease. Also, their clustering with all different extraarticular manifestations indicate that vascular pathogenetic mechanisms are important in all types of extraarticular RA. This is of particular interest given the association of RA with cardiovascular co-morbidity. The different manifestations tend to cluster in specific

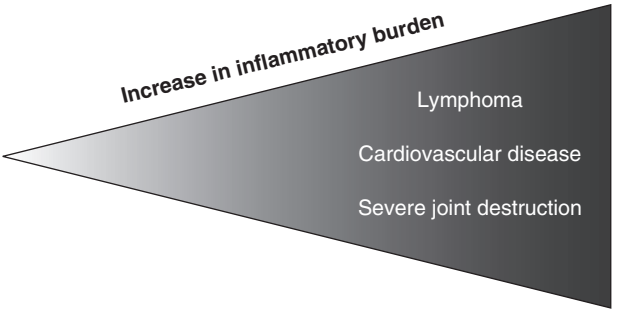


**Figure 4** Inflammatory burden. With increased accumulated inflammatory burden the risk for additional disease states and complications increases for patients with RA. It is thus of importance to control the inflammation at all time points, a difficult task when not all patients respond to the antirheumatic therapies available today.

patterns and suggest shared disease mechanisms in these systemic manifestations. There are very likely several genetic and environmental factors affecting the development of these manifestations, of which some are described below.

An association of *HLA-C3* with vasculitis has been described. *HLA-C3* is the product of a specific MHC class I gene and thus with CD8+ cytotoxic T-cells and NK-cells, in contrast to MHC class II proteins, which interact with CD4+ T-helper cells. Smoking is an independent predictor of vasculitis and is probably involved in vascular damage by antigen modification as described above. The increased number of patients with specific autoantibodies such as RF and ANAs (antinuclear antibodies) supports a role for immune complexes in the pathogenesis of vasculitis, which in turn indicates B-cell abnormalities in the development of extraarticular RA.

In summary, extraarticular manifestation indicate a more severe disease; however, this risk group is today not routinely analyzed early in the disease course. Here, specific biomarkers could possibly aid an early aggressive treatment of this patient group in order to limit the progression of extraarticular manifestations.

## Cardiovascular Disease in RA

Cardiovascular (CV) morbidity and mortality in RA occur at rates greater than would be expected from the profile of established CV risk factors. Diabetes mellitus, hypercholesterolemia, hypertension, cigarette smoking, and obesity are generally powerful classifiers of CV risk and when comparing the profiles of these risk factors in people with and without RA, they are similar. Thus, there must be RA-specific factors explaining the increased risk. Dividing the patients in younger and older patients demonstrates that such factors are more common in the younger age group. In older patients, the

proportion of atherosclerosis can be explained by established CV risk factors as described above. In contrast, in the younger age group, systemic inflammation seems to exert its atherosclerotic effects early. High-resolution carotid ultrasound is used to measure carotid intima media thickness (IMT). These measures may thus be used as a biomarker of subclinical atherosclerosis. Interaction calculations suggest that the ESR's effect on IMT varies according to the number of CV risk factors; thus, higher ESR values, a measure for inflammation, are associated with greater IMT only in the presence of CV risk factors. In patients without CV risk factors, the ESR is not associated significantly with IMT.

Taking genetic and environmental components into consideration of CV disease in RA patients demonstrates that the presence of two SE alleles predicts death from cardiovascular disease in RA patients, and this is independent from autoantibody status. However, the greatest risk of death from CV disease is associated with an interaction among smoking, SE alleles, and anti-CCP antibodies. Interestingly, no association of the *PTPN22 gene* with mortality was detected.

In summary, both established CV risk factors and RA manifestations account for a significant proportion of atherosclerosis is RA. Factors related to RA may have a greater influence on the extent of atherosclerosis in young patients. The presence of established risk factors may be necessary for systemic inflammation to promote atherosclerosis; however, more studies are needed to understand these interactions.

## Lymphoma in RA

Among RA patients there is in general a doubled risk for lymphoma. However, a closer analysis of disease severity, as measured by degree of joint destruction, ESR, number of affected joints, and accumulated inflammatory load, revealed a more pronounced relative risk in patients with the most severe disease and little or no increase in those with mild to moderate disease. This raises the question of whether lymphomas arise as a consequence of treatment or because of a lack thereof. Larger cohort studies have thus far not confirmed any treatment-related excess of lymphoma in general. However, it is premature to make firm conclusions about the true lymphoma risk in patients treated with TNF antagonists. In summary, the data so far indicate that the inflammatory burden should be kept as low as possible to possibly reduce the risk for lymphoma development.

## Cartilage and Bone Destruction

During the inflammation in the joint, cartilage and underlying bone are broken down. Matrix metalloproteinases (MMPs) are enzymes active in this process, which results in the release of a number of collagenous and noncollageneous cartilage and bone-derived molecules. Serum levels of MMP-3 and urinary

levels of c-telopepetide of collagen type II (CTX II) have been identified as two independent baseline factors that predict radiographic progression. Upon successful MTX treatment, MMP-3 levels decrease. Also, serum cartilage oligomeric matrix protein (COMP) levels are predictive of radiographic damage. Proof-of-concept studies demonstrate that targeting RANK-mediated osteoclastogenesis prevents inflammatory bone loss, and clinical application has only just begun by blocking RANKL.

Replacing radiographic measures with measures of biochemical molecules in serum and/or urinary samples might in the future become the preferred way to measure joint destruction specifically. Traditionally, high inflammatory burden and measures of CRP have been included in the prediction of joint destruction, but the concomitant measures of inflammation and joint destruction indicate that the inflammatory and destructive processes are uncoupled and therefore need to be identified and treated separately.

## BIOMARKERS PROVING LEADS CONCERNING EFFECTS OF TREATMENT OF RA

The drugs used in the treatment of RA are generally divided into three groups: nonsteroidal anti-inflammatory drugs (NSAIDs) to relieve pain, traditional disease-modifying antirheumatic drugs (DMARDs), and biologics. There are today some general guidelines for the treatment of RA; however, biomarkers guiding the most adequate selection for the single patient are missing.

In the majority of RA patients, treatment with methotrexate or other DMARDs only partly reduces clinical and radiological progression of the disease. For many patients, therefore, the introduction of the biological drugs provided a major breakthrough. By blocking the proinflammatory cytokine TNF, joint progression could be halted much more efficiently than before in a large group of patients. More recent drugs eliminating circulating B-cells or blocking co-stimulation of T-cells also indicate a slowdown in joint destruction. Measuring levels of TNF in the inflamed tissue has been suggested to be a biomarker for treatment response to TNF blocking agents. Unfortunately, it has yet been possible to describe such a marker for serum samples.

With the identification of the association of smoking, SE alleles, and anti-CCP antibodies with the greatest risk of death from CVD, it will be possible to be more active in treatment to prevent cardiovascular disease in these patients at high risk for CV disease. In the treatment of lymphoma, which in the majority of cases are B-cell derived, a specific treatment option exists for RA patients. Rituximab was established in 1998 as a drug to treat patients with B-cell lymphomas, and the indication was extended in 2006 to treat patients with moderate to severe RA, independent of lymphoma associations. Thus, in these cases both the lymphoma and the RA are treated with the same drug.

## CONCLUDING REMARKS

In summary, RA is a heterogeneous disease with regard to severity of the joint inflammation and other inflammatory manifestations. To prevent disease onset or progression of disease in the single patient, distinct biomarkers are needed, and only a few have yet been identified. The fact that the presence or absence of ACPAs divides the disease into two distinct entities will very likely influence future treatments and clinical trials. Clinical trials, if not already focusing on a subgroup of patients, need also to be adequately powered to allow subgrouping patients according to genetic setup, environmental exposures, and different nonsynovial inflammations in order to find biomarkers for individualized therapies. Today's and tomorrow's biological therapies targeting different parts of the immune and inflammatory pathways will also increase our understanding of disease etiology and progression and help identify new drugs.

## RECOMMENDED READING

Bowes J, Barton A (2008). Recent advances in the genetics of RA susceptibility. *Rheumatology*, doi10:1093.

Ekström-Smedeby K, Baecklund E, Askling J (2006). Malignant lymphomas in autoimmunity and inflammation: a review of risks, risk factors, and lymphoma characteristics. *Cancer Epidemiol Biomarkers Prev*, 15:2069–2077.

Farragher TM, Goodson NJ, Naseem H, et al. (2008). Association of the *HLA-DRB1* gene with premature death, particularly from cardiovascular disease, in patients with rheumatoid arthritis and inflammatory polyarthritis. *Arthritis Rheum*, 58:359–369.

Klareskog L, Rönnelid J, Lundberg K, Padyukov L, Alfredsson L (2008). Immunity to citrullinated proteins in rheumatoid arthritis. *Annu Rev Immunol*, 26:651–675.

Romas E, Gillespie T (2006). Inflammation-induced bone loss: Can it be prevented? *Rheum Dis Clin N Am*, 32:759–773.

Turesson C, Schaid DJ, Weyand CM, et al. (2006). Association of *HLA-C3* and smoking with vasculitis in patients with rheumatoid arthritis. *Arthritis Rheum*, 54:2776–2783.

Van der Helm-van Mil AHM, Huizinga TWJ, De Vries RRP, Toes REM (2007). Emerging patterns of risk factor make-up enable subclassification of rheumatoid arthritis. *Arthritis Rheum*, 56:1728–1735.

Van Dongen H, van Aken J, Lard LR, et al. (2007). Efficacy of methotrexate treatment in patients with probable rheumatoid arthritis: a double-blind, randomized, placebo-controlled trial. *Arthritis Rheum*, 56:1424–1432.

Young-Min S, Cawston T, Marshall N, et al. (2007). Biomarkers predict radiographic progression in early rheumatoid arthritis and perform well compared with traditional markers. *Arthritis Rheum*, 56:3236–3247.

# PHARMACOKINETIC AND PHARMACODYNAMIC BIOMARKER CORRELATIONS

J.F. Marier, Ph.D., FCP
*Pharsight, A Certara Company, Montreal, Quebec, Canada*

Keith Gallicano, Ph.D.
*Watson Laboratories, Corona, California*

## INTRODUCTION

The importance of a biomarker in drug development is based on the weight of evidence that changes in dose and/or drug concentration levels correlate strongly with the biomarker and the clinical outcome desired. The linkage between pharmacokinetics of a new chemical entity (NCE), the dynamics of a biomarker, and the ultimate clinical outcome should be based on robust theoretical considerations, prior therapeutic experience, well-understood pathophysiology, and knowledge of the drug's mechanism of action. The validation of biomarkers is therefore very important, particularly if one considers their relevance for decision-making and regulatory purposes. The elements described below are crucial for the integration of a biomarker in a drug development program.

1. *Identification, quantitation, and validation of biomarker assays and drug concentrations.* Reliable and selective assays should be validated under

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

a good laboratory practices (GLP)–like environment for quantitative methods for measuring both biomarker responses and drug concentrations. Most biomarkers are endogenous macromolecules that can be measured in biological fluids. As for drug assays, biomarker assays should provide acceptable sensitivity, specificity, precision, and accuracy. For example, intra- and intersubject variability of the biomarker should be minimal and well understood, biomarker stability and sample storage conditions should be controlled, and quantitative assay should be standardized and validated with adequate power to distinguish between treatment and control groups [1–3].

Assays should be validated to meet study objectives at various drug development stages and possess adequate performance to quantify biochemical responses specific to the target disease progression and drug intervention. In addition, biomarker measurement should be technically practical, so that measurement can be obtained by minimally invasive techniques and that the sampling of the biomarker is minimal or is taken at the same time as other physiological, safety, or efficacy measurements [4].

2. *PK/PD correlation and integration of model-based drug development.* Biomarkers typically have different time courses from clinical endpoints and often are more directly related to the time course of drug concentrations. To evaluate potential correlations, adequate information should be available on the pharmacokinetics (PK) of the new chemical entity (NCE). Similarly, adequate information should be available on the pharmacodynamics (PD) (including PK) of the biomarker, with the expected effect on clinical outcome. Following the identification of correlations between the NCE and specific biomarkers, PK/PD modeling may be performed in early stages of drug development to gain greater insight into the effect of drug concentrations on the biomarker as a function of time. PK/PD modeling involves a set of techniques using mathematical and mechanistic (as opposed to empirical) models derived from quantitative pharmacology. Biomarkers, together with PK/PD modeling and simulation, provide a continuous process to link what has been learned from today's drug development cycle to the next generation of biomarkers, assays, and models [5–7].

3. *Correlation between biomarkers and clinical endpoint.* The availability of a validated biomarker for a specific clinical endpoint may greatly facilitate the development of compounds that act via well-understood mechanisms of action. For example, biomarkers for HIV and AIDS include viral load (the number of free virus particles in the blood) and the count of CD4+ immune system cells. Fasting blood sugar and hemoglobin A1c, a protein that indicates a patient's blood sugar history, are established biomarkers for diabetes treatments. The relationship of the biomarker to the therapeutic endpoint should

be validated in early phase II trials; and, when possible, as early as phase I [8,9].

Many surrogate markers can be assessed in healthy volunteers and in the intended patient population. The ability to assess surrogate markers in healthy volunteers can bring the initial decision-making process into phase I rather than phase II or phase III of the NCE development process. This can be a significant competitive advantage to those who use the opportunity. Table 1 presents widely used biomarkers according to disease and indications. Biomarkers can be elevated to the status of a surrogate endpoint based on the

**TABLE 1 Widely Used Biomarkers for Specific Disease or Indications**

| Biomarkers | Disease or Indication |
|---|---|
| Pulmonary function tests, forced expiratory volume (FEV) | Asthma |
| PT (including INR), APTT, anti-Xa activity, fragment 1 + 2, plasminogen, plasmin inhibitor, D-dimer | Coagulation and fibrinolysis |
| Leukotrienes, cytokines, and chemokines | Chronic obstructive pulmonary disease |
| Epidermal growth factor (EGF), fibroblast growth factor (FGF), human growth hormone (HGH), neopterin | Growth modulation |
| Viral load, CD4 count | HIV |
| Cortisol, estradiol, estrone, follicle-stimulating hormone, luteinizing hormone, progesterone, testosterone | Hormonal control |
| Angiotensin-I, angiotensin-II, plasma renin, aldosterone, and angiotensin-converting enzyme (ACE) activity | Hypertension |
| Eicosanoids (prostaglandins and leukotrienes) | Inflammation |
| Interferons, interleukins, tumor necrosis factor (TNF), rheumatoid factor | Immune modulation, rheumatoid arthritis |
| Cholesterol, fatty acids, HDL, LDL, phospholipids | Lipid metabolism |
| Cytidine deaminase activity, inosine, xanthine, hypoxanthine, uric acid | Nucleotide metabolism |
| 8-Hydroxy-dG, protein carbonyls, I-κB, NF-κB | Oxidative stress (diabetes, heart disease, arthritis, obesity, and cancer) |
| Glucose, glucagon, fructosamine, glycosylated albumin, and hemoglobin A1c, insulin, C-peptide | Type 1 and 2 diabetes mellitus |
| Stomach fullness, cholecystokinin, glucagon-like peptide 1, bombesin, somatostatin, ghrelin, leptin, glucose, insulin, diet-induced thermogenesis, temperature, ventilatory parameters | Satiety and satiation |
| ProMMP-1, ProMMP-3, tissue inhibitory of MMP-1 | Tissue remodeling |

weight of evidence that demonstrates that changes in the biomarker correlate strongly with the desired clinical outcome. The best known physiologic surrogate marker is blood pressure reduction as an indicator of reduced incidence of stroke and myocardial infarction. This surrogate marker is now used as a clinical endpoint, because a strong relationship between reduction in blood pressure and a reduction in the incidence of myocardial infarctions and strokes has previously been established.

For innovative products, limited information may be available on the relationship between the PK of a drug, the effect on a candidate biomarker, and the ultimate clinical response. Although this may limit the use of predetermined biomarkers in the early development phases of a drug, the incorporation of recent advances in genomics, pharmacogenetics, and proteomics may help to qualify various candidate biomarkers. Pilot exploratory studies should therefore be conducted to qualify and validate candidate biomarkers for predictive clinical assessment of disease progression and the effect of drug intervention. Mechanistic approaches using new technological advances in target validation, functional biology, proteomics, genomics, and the monitoring of gene expression using DNA arrays, imaging, and the correlation of these effects to the disease state may hold promise for an earlier understanding of disease and toxicological processes, improve the predictiveness of preclinical pharmacology and toxicology data to humans, and find suitable biomarkers for use in the early stage of drug development of innovative products [2,10].

The appropriate combination of biomarker identification and selection, bioanalytical methods development and validation for drugs and biomarkers, and mechanism-based PK/PD models for fitting data and predicting future clinical endpoints and outcomes may provide powerful insights and guidance for effective drug development, toward safe and efficacious medicine for individual patients [6–8,11,12]. The U.S. Food and Drug Administration (FDA) is developing these concepts as part of its *critical path initiative*, where the agency recognizes that the major contributor to inefficiency in development was the absence of innovative new methods for preclinical and clinical testing of drugs. The FDA recommended new tools and various opportunities to improve drug development, including the use of validated biomarkers and the application of model-based drug development (MBDD). In the following sections we focus on the PK/PD correlation assessment of a drug and a biomarker, and the use of PK/PD modeling and simulation to optimize drug development and decision making.

## PK/PD CORRELATION OF BIOMARKERS

Plans to incorporate biomarkers for potential correlations in drug development programs should occur when a new candidate molecule is being identified and the discovery process is starting. To evaluate potential correlations between the drug and candidate biomarkers, adequate information should be

available on the PK of the drug. Preliminary PK/PD correlation analyses may be performed using noncompartmental approaches (i.e., model independent) by deriving parameters such as the area under the concentration–time curve (AUC), maximum concentration ($C_{max}$), time to maximum plasma concentrations ($t_{max}$), and terminal elimination half-life ($t_{1/2}$). Exploratory data analyses (EDAs) between drug exposure data and candidate biomarkers are typically performed by plotting individual concentration data and/or PK parameters of the drug versus the candidate biomarker using appropriate data transformations and in appropriate subgroups (e.g., by gender, dose, or route of administration) in order to identify the impact of drug exposure on the source and magnitude of variability in biomarker response. During exploratory data analyses, questions in Table 2 should be answered to perform PK/PD correlations of biomarkers.

In the simplest case, biomarkers will be related directly to the drug concentrations in the biological fluid sampled; that is, the biomarker will respond in a graded fashion so that increasing drug concentrations will result in a gradual increase or decrease in the biomarker of interest over time. A temporal relationship between the time course of drug concentrations and the biomarker is observed. In some cases, the temporal delay may only be a few minutes to an hour, whereas in other cases it may be as long as several hours to days. Exploratory analyses may also provide valuable information on the pharmacodynamics of biomarkers as well as PK/PD correlations as presented in Table 3.

Preclinical and clinical study designs should provide as much information on the nature of the PD response of candidate biomarkers. Depending on the PK of the drug and the dynamics of the biomarker, exploratory PK/PD correlations may be limited and likely to require a more mechanistic approach for an optimal assessment of PK/PD correlations. Mechanism-based PK/PD

**TABLE 2 Pharmacokinetic Properties and Impact on PK/PD Assessment**

| Pharmacokinetic Property | Impact on Correlation Assessment |
|---|---|
| Absorption | Will rate and/or extent of drug exposure correlate with biomarkers? Is a minimum threshold likely to correlate with biomarker response? |
| Distribution | Will total (unbound plus bound) or unbound drug levels contribute to the overall effect on biomarkers? Are concentration levels in peripheral tissues likely to correlate with biomarkers? |
| Metabolism | Will active metabolites contribute to the overall effect on biomarkers? Is there a temporal delay between concentrations and biomarkers (proteresis or hysteresis)? |
| Elimination | Will the rate of elimination of the drug correlate with biomarker response? Does drug elimination parallel the biomarker response? |

**TABLE 3 Pharmacodynamic Properties and Impact on PK/PD Assessment**

| Pharmacodynamic Property | Impact on Correlations Assessment |
|---|---|
| Mechanism of action | Will the drug stimulate or inhibit the formation or degradation of the biomarker? Will the drug create a reversible or irreversible effect on the biomarker? |
| Competitive ligands | Will the drug compete with endogenous, environmental, or dietary ligands for receptor binding? |
| Diurnal variation | Are placebo data available to determine a diurnal/circadian cycle of biomarkers? |
| Disease state | Will the disease state affect the biomarker response? Does drug elimination parallel the biomarker response? |

models differ from empirical descriptive models in that they contain specific expressions to characterize processes on the causal path between drug administration and effect [13–15]. In the next section we focus on important issues and concepts that apply to PK/PD modeling and the use of biomarkers to improve drug development processes.

## PK/PD Modeling of Biomarkers

PK/PD modeling and simulation tools have continued to develop to a high level of sophistication over the past 20 years. Mechanism-based PK/PD modeling involves a set of techniques using mathematical and mechanistic models derived from quantitative pharmacology. The simplest compartmental PK model can be described as a one-compartment model following intravenous administration of a drug. A schematic representation of the compartmental model is presented in Figure 1. In this example, the removal of

| PK Model | Equation | Pharmacokinetics |
|---|---|---|
| 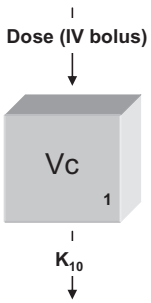  | $C_{\text{time}} = \frac{\text{dose}}{Vc} \bullet \exp^{-K_{10} \bullet \text{time}}$ | 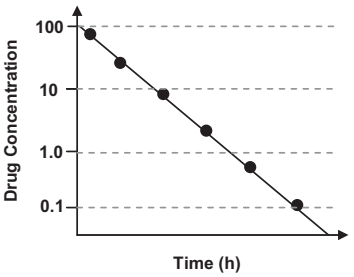  |

**Figure 1** Pharmacokinetic modeling and equations.

drug from the central compartment (i.e., systemic circulation) was described using a first-order elimination rate constant. Compartmental analyses can be performed routinely by fitting the data to a model using a PK/PD software package such as WinNonlin (Pharsight, a Certara Company, St. Louis, Missouri), Kinetica (ThermoFisher, Waltham, Massachusetts), S-ADAPT (University of Southern California,) or NONMEM (University of California at San Francisco). The PK of drugs can be described using more advanced compartmental models and can be customized to describe any type of PK behavior. Case studies in which different types of compartmental models were used in preclinical and clinical drug development are presented later in the chapter.

Modeling the concentration–effect relationship on the biomarker is very dependent on the mechanism of action of the drug. In vitro studies that may elucidate this mechanism are very important in the early stage of drug development. For example, drug-induced changes in biomarkers of interest are often modeled using classical drug receptor theory following the law of mass action. This theory predicts that as receptors in the target organ interact with the drug, the biomarker levels will increase with increasing interaction until all receptors have been occupied, at which time no further change in biomarker will occur. In its simplest form, the effect on the biomarker will be driven by the measured drug concentration over time, $C_{\text{time}}$, assuming that drug concentrations in the systemic circulation is at equilibrium with the effect site. The classic and most commonly used PD model under these conditions is the $E_{\max}$ model, which is an empirical function for describing nonlinear concentration–effect relationships. It has the general form illustrated in Figure 2, where $E_{\max}$ represents the maximum effect on the biomarker, $C_{\text{time}}$ the concentrations of a drug predicted over time with a PK model, and $EC_{50}$ the concentration of the drug that produces half of $E_{\max}$. The PD model can be

| Equation | Pharmacodynamics |
|---|---|
| $\text{Effect}_c = \text{E}_{\max} \bullet \dfrac{C^n_{\text{time}}}{C^n_{\text{time}} + EC^n_{50}}$ | 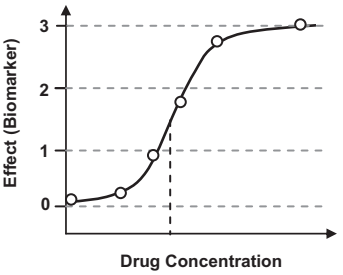 |



**Figure 2** Pharmacodynamic modeling and equations.

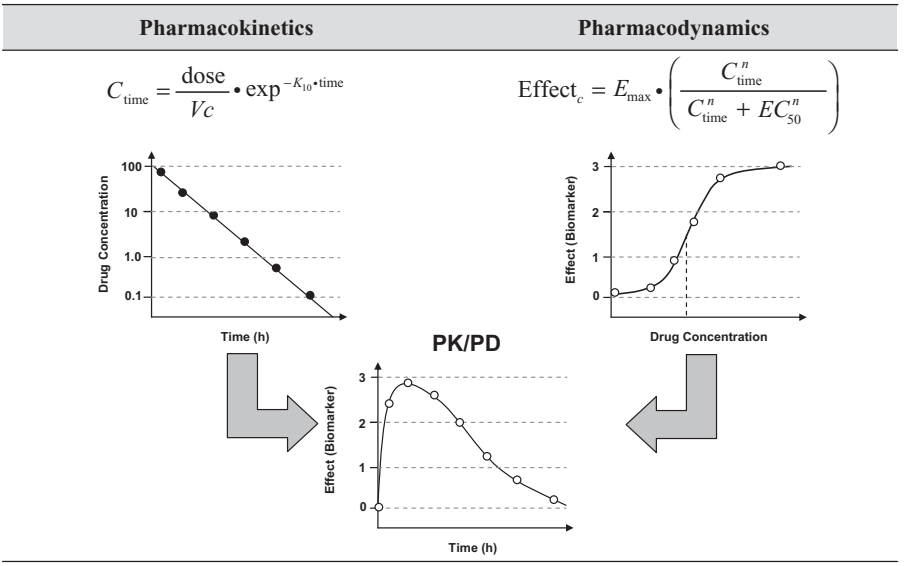


**Figure 3** Pharmacokinetic/pharmacodynamic modeling.

related to receptor theory, where $EC_{50}$ is the parameter characterizing the potency of the drug on the biomarker and $n$ is a sigmoidicity factor describing the drug concentration vs. the effect on the biomarkers or the number of molecules interacting at the biomarker receptor. Although the $E_{max}$ model is highly versatile for different situations, more sophisticated PD models have been developed over the last decades.

PK/PD modeling builds the bridge between these two classical disciplines of pharmacology, as depicted in Figure 3. PK/PD modeling is performed to gain greater insight into how drug concentrations may affect the biomarker as well as how the models can be used to perform simulations in order to predict changes in biomarker levels under a variety of experimental conditions. Models can improve the prediction and assessment of patient response with the use of simulations, and therefore increase the success of a drug development program [13–15]. For example, the models can be used to refine a dosing regimen in order to predict the drug exposure (PK modeling) and the resulting effect on the biomarker, to guide starting doses and regimens (as well as adjustments of dosage or dosing regimens in special populations), and to provide a better understanding of outcomes from clinical efficacy studies. Case studies will be presented to demonstrates how biomarkers, together with PK/PD modeling and simulation, provide a continuous process to link what has been learned from today's drug development cycle to the next generation of biomarkers, assays, and models.

## Critical Path Initiative

In the current health care environment, dominated by increasing expenses associated with drug development programs, rapidly changing technologies, generic competition, and therapeutic substitution, there is increased pressure to develop new therapies using effective drug development programs. The escalating costs and low productivity of drug development have been well documented over the past several years. In March 2004, the FDA recognized that drug development has become increasingly challenging, inefficient, and costly [16]. The agency concluded that the major contributor to the inefficiency in development was the absence of innovative new methods for preclinical and clinical testing of drugs. In its critical path initiative, the agency called upon new, publicly available scientific and technical tools to improve drug development programs. These include the use of biomarkers and clinical trial endpoints as well as PK/PD modeling to improve decision making and make the development process more effective and more likely to result in safe products to benefit patients [17–20].

The agency recommended reviewing the body of collective biomarker information to build a new framework for biomarker and surrogate endpoint use in drug development [16]. Part of this initiative is to complete an inventory of surrogate endpoints that have been used for past drug approvals and to evaluate how the decision was made to accept the respective surrogates. As part of this process there will be an attempt to identify widely used, disease-specific biomarkers and determine the gap that exists to elevate these biomarkers to surrogate endpoints. The acceptance of CD4 cell counts and HIV plasma RNA as mechanism-based surrogate endpoints are probably the best examples. The elevation of a biomarker to a surrogate endpoint is particularly attractive because biomarkers can be measured more quickly and easily and at lower cost than can true clinical endpoints, or when clinical endpoints are excessively invasive or when their use would be considered unethical.

## PK/PD Modeling of Biomarkers in the Early Phase of Drug Development (Colburn)

There are numerous stages along the development cycle where the development of biomarkers and PK/PD modeling can add value. Incorporation of PK/PD studies throughout preclinical and clinical development may lead to earlier identification of optimal dosing regimens in clinical development and may reduce the overall time of drug development. Beyond the traditional paradigm of using fractions of the nonclinical toxic doses in animals to design first-in-human dosing schedules, the use of PK/PD modeling in early preclinical development allows one to define more precisely the dose–concentration–pharmacological effects and dose–concentration–toxicity relationships with the use of key efficacy and safety biomarkers. The extrapolation of these results to humans using a combination of in vitro and in vivo data can be

particularly helpful in determining the appropriate dosing regimen for phase I studies and in guiding dose escalation in order to achieve the systemic exposure in humans that is expected to be associated with the desired effect on the biomarker and ultimately, the clinical outcome. Overall, the use of validated biomarkers and the application of PK/PD modeling and simulation principles represent an opportunity to identify optimal drug candidates and possibly to develop drugs in the shortest time frame possible [7–9].

PK/PD modeling output can be no better than the biomarkers or surrogate endpoints used for the modeling (i.e., the input). As we increase our understanding of biomarkers, surrogate markers, and disease and drug mechanisms, PK/PD modeling inputs and outputs will improve and, consequently, predictive power will improve.

## PK/PD Modeling of Biomarkers in Late Phases of Drug Development

The usefulness and validity of PK/PD models for the evaluation of dose–concentration–effect relationships are not limited to well-designed clinical studies with relatively small groups of patients and frequent measurements of concentration and effect. It has also been shown for observational data obtained from large clinical trials with sparse and imbalanced sampling schedules by applying population modeling techniques. Population PK/PD modeling is based primarily on the nonlinear mixed-effects regression models introduced by Sheiner and co-workers and makes possible the characterization of dose–concentration effect relationships in populations rather than individuals, thereby providing the opportunity to identify and account for sources of interindividual PK and PD variability [21,22].

In addition, the increased understanding of drug action derived from biomarker response and PK/PD-based drug development may lead to a definition of strategies for individualization of drug dosage regimens to ensure optimal therapeutic outcome in subpopulations of patients. The building of a PK/PD database during drug development can provide an essential framework for continued refinement and improvement during postmarketing drug use, allowing one to answer specific questions without the need to perform additional studies.

## Clinical Trial Simulation

Population PK/PD modeling has been used successfully for numerous drugs. These models may be descriptive and/or predictive of the time course of PD effects. When properly developed, predictive models may be used to guide starting doses and regimens (as well as adjustments of dosage and dosing regimens in special populations), add evidence to the certainty of the decision to allow market access, and provide a better understanding of outcomes from clinical studies [23,24]. Clinical trial simulations are usually motivated as follows:

- To improve the clinical trial process by maximizing the probability of trial objectives being attained with reduction of unnecessary costs
- To improve decision making in drug development by providing an objective framework for safety, efficacy, and commercial risk assessments
- To improve the overall efficiency of drug development in terms of informativeness, economy, and speed

Modeling and simulation may help predict the results of future drug trials and can be used at decision points in the drug development process to assess quantitatively the risk of moving forward with a candidate drug.

## CASE STUDIES

In recent years, PK/PD modeling has developed from an empirical descriptive discipline into a mechanistic science that can be applied at all stages of drug development. Preclinical PK/PD studies may prompt a series of important mechanistic studies to explore the relationship between plasma concentrations and the resulting effect on a biomarker in later stages of clinical drug development. The application of PK/PD modeling to preclinical pharmacology studies may help to provide information on drug effects and actions that would be difficult to obtain in human subjects. As such, preclinical and clinical PK/PD studies constitute a scientific basis for rational drug discovery and development. The following preclinical and clinical cases studies are presented to demonstrate how modeling techniques have improved the overall understanding of PK/PD relationship of biomarkers.

### Case Study 1 (Preclinical): Pharmacokinetic–Pharmacodynamic Modeling of the Respiratory Depressant Effect of Norbuprenorphine in Rats

The objective of this investigation was to characterize the PK/PD correlation of buprenorphine's active metabolite norbuprenorphine for the effect on respiration in rats [25]. Following intravenous administration, the time course of plasma concentrations of buprenorphine and norbuprenorphine were determined in conjunction with the effect on ventilation as determined using plethysmography, a test used to measure changes in air volume. The PK of norbuprenorphine was best described by a three-compartment PK model with nonlinear elimination. A saturable biophase distribution model with a power PD model described the PK/PD relationship best as presented in Figure 4.

No saturation of the effect at high concentrations was observed, indicating that norbuprenorphine acted as a full agonist with regard to respiratory depression. By simulation, it was shown that following intravenous administration of buprenorphine, the concentrations of norbuprenorphine reached values that were well below the values causing an effect on respiration. In

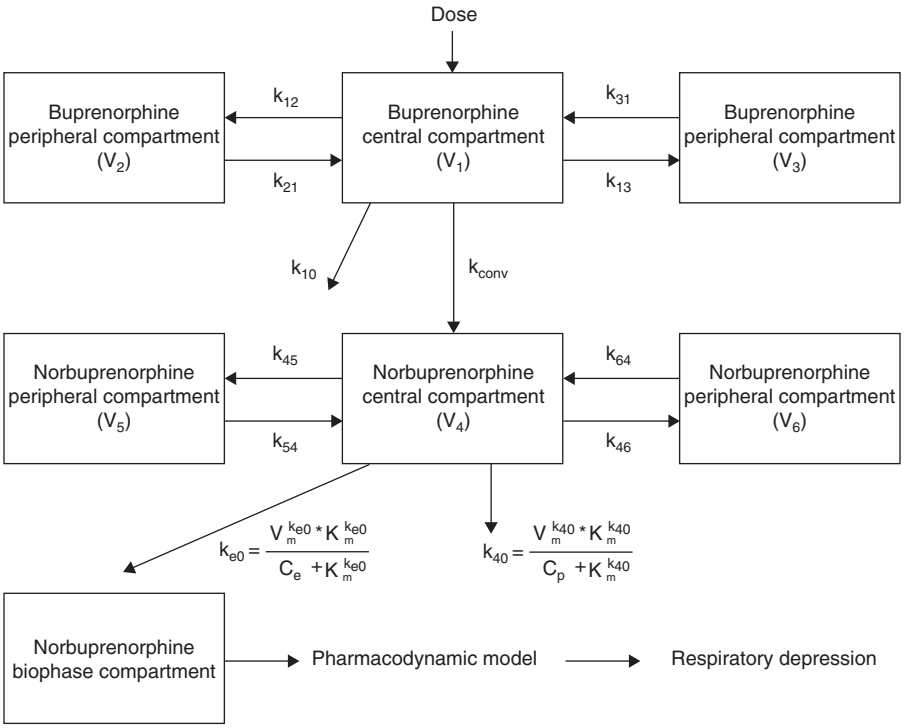


**Figure 4** PK/PD model for respiratory depressant effect of norbuprenorphine and norbuprenorphine. (From ref. 25, by permission of the American Society for Pharmacology and Experimental Therapeutics.)

conclusion, the PD of the metabolite norbuprenorphine was determined to be markedly different from the PD of the parent compound buprenorphine with regard to the receptor association–dissociation kinetics, the in vivo potency, and the intrinsic efficacy for the respiratory depressant effect. The values of these norbuprenorphine concentrations are well below values causing an effect on respiration. Therefore, it was concluded norbuprenorphine does not contribute to the overall respiratory depressant effect of buprenorphine. These results were consistent with the experience from clinical use of buprenorphine in patients.

## Case Study 2 (Preclinical): Pharmacokinetics and Pharmacodynamics of PEGylated IFN-β1a Following Subcutaneous Administration in Monkeys

The purpose of this study was to characterize the PK/PD properties of a new polyethylene glycol (PEG) conjugate formulation of interferon-β1a (IFN-β1a) following subcutaneous (SC) administration in monkeys [26]. Single SC injections of 0.3, 1, and 3 MIU/kg of PEG-IFN-β1a were administered to three

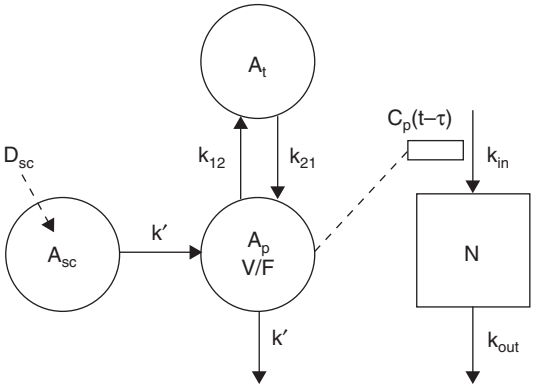


**Figure 5** PK/PD model for the effect of IFN-β1a on neopterin. (From ref. 26, by permission of Springer Science and Business Media.)

groups of cynomolgus monkeys. Plasma concentrations of drug and neopterin, a classic biomarker for IFN-β1a, were measured at various time points after dosing using an ELISA assay. PK/PD profiles were first described by noncompartmental methods, and then a pooled analysis was performed using an integrated mathematical model, where fixed and delayed concentration–time profiles were used as a driving function in an indirect stimulatory response model as presented in Figure 5.

The PK component of IFN-β1a was assessed using a standard linear two-compartment model with a first-order rate constant for absorption to and elimination from the central compartment of IFN-β1a. The PD model was a modified stimulatory indirect response model, with a zero-order rate constant of neopterin production ($k_{in}$), a first-order rate constant of neopterin elimination ($k_{out}$), where the driving function (i.e., concentrations of IFN-β1a) was delayed by a time-lag parameter ($\tau$). Neopterin concentrations followed a typical dose-dependent biphasic pattern. Pooled PD profiles were well described by the PK/PD model, and the neopterin elimination rate was consistent with previous estimates. The PEG modification of IFN-β1a provided enhanced drug exposure and comparable pharmacodynamics to unpegylated IFN-β1a.

### Case Study 3 (Clinical): Modeling of Brain D2 Receptor Occupancy–Plasma Concentration Relationships with a Novel Antipsychotic

PK/PD modeling using functional imaging methodology such as positron-emission tomography (PET) may be a very useful tool for early clinical development of antipsychotic compounds, especially in suggesting initial doses for further clinical studies. The purpose of this study was to assess PK/PD relationships of YKP1358 to guide decisions for further clinical study designs. YKP1358 is a novel serotonin (5-HT2A) and dopamine (D2) antagonist that

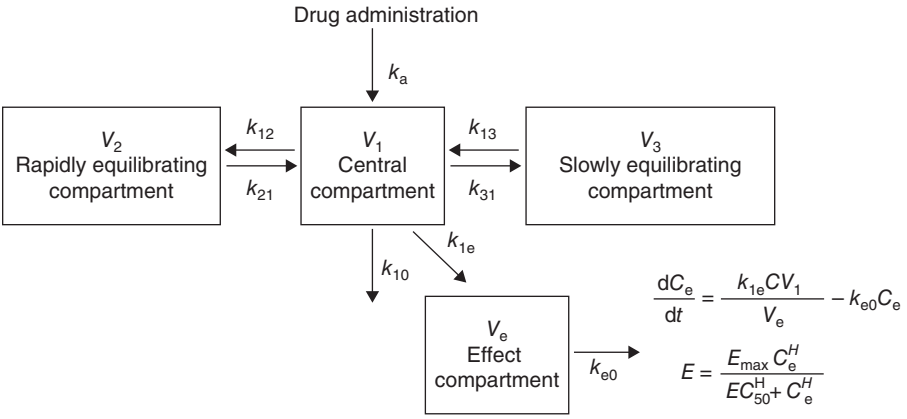


**Figure 6** PK/PD model used to assess the relationship between YKP1358 and D2 receptor occupancy. (From ref. 27, by permission of Macmillan Publishers Ltd.)

fits the general profile of an atypical antipsychotic drug [27]. A D2 receptor occupancy study was conducted in healthy volunteers using PET to measure the D2 receptor occupancy of YKP1358 and to characterize its relationship to plasma drug concentrations. A single oral dose, parallel group, dose-escalation (100, 200, and 250 mg) study was performed in 10 healthy male volunteers with the PET radiotracer [$^{11}$C]raclopride. The relationship between plasma concentration and D2 receptor occupancy was analyzed with an indirect link model that included an effect compartment, as well as an equilibrium rate constant ($k_{e0}$), as presented in Figure 6.

Results of the study demonstrated that YKP1358 possesses potential antipsychotic effects based on the D2 receptor occupancy data. Considering the D2 receptor occupancy data, effective doses in patients were predicted to be greater than 250 mg twice a day. This is the first study in which the relationship between plasma concentration and the biomarker of D2 receptor occupancy was modeled using nonlinear mixed-effects modeling. It is anticipated that these results will be useful in estimating the initial doses of YKP1358 required to achieve a therapeutically effective range of D2 receptor occupancy in subsequent studies.

## Case Study 4 (Clinical): Semimechanistic and Mechanistic Population PK/PD Model for Biomarker Response to Ibandronate

Biomarkers of bone turnover provide a rapid and accessible means of evaluating physiological responses to antiresorptive therapies. Recent evidence also suggests an association between the magnitude of biomarker suppression and the likely response in terms of bone mineral density (BMD) change and fracture risk reduction. This association suggests that these

biomarkers are on the causal pathway for prevention of fracture following administration of bisphosphonates. As such, biochemical markers of bone turnover can offer a timely indication of the likely efficacy associated with nominal dosing regimens. The aim of this project was to develop and validate a pharmacological model for ibandronate, a new bisphosphonate for the treatment of osteoporosis, capable of describing its PK in serum and urine, and the urinary excretion of the C-telopeptide of the α chain of type I collagen (uCTX), a sensitive biomarker of PD response to ibandronate [28]. A classical PK/PD model was developed that described accurately the PK of intravenously administered ibandronate and its effect on the biomarker, as presented in Figure 7.

A four-compartment PK model was used, with an indirect PD response model with a uCTX formation rate (KS) and uCTX degradation rate constants (KD). Ibandronate in the "bone compartment" was assumed to inhibit osteoclast activity and hence the rate of synthesis (KS) and the urinary excretion of CTX. The complex physiological PK/PD model adequately described the PK of intravenously administered ibandronate in serum and urine as well as the time course of uCTX. To reduce processing times, the classical PK/PD model was simplified using a *kinetics of drug action* or kinetic (K)-PD model (i.e., a dose–response model, as opposed to a dose–concentration–response model). The model was subsequently extended to consider the influence of supplemental therapy on the PD response and subjected to external validation by retrospectively simulating the time course of uCTX change reported in a phase III study and in a phase II/III study of intravenous ibandronate. In summary, a pharmacostatistical model was developed and validated that adequately described the time course of uCTX change after oral and intravenous ibandronate therapy for osteoporosis. This model is currently being used to aid in the development of novel intermittent oral and intravenous regimens for ibandronate.

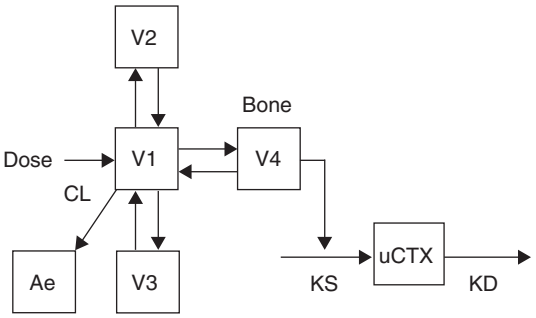


**Figure 7** PK/PD model used to assess the relationship between ibandronate in serum and urine and the urinary excretion of uCTX. (From ref. 28, by permission of Wiley-Blackwell.)

## Case Study 5 (Clinical): Receptor Theory–Based Semimechanistic PD Model for the CCR5 Noncompetitive Antagonist Maraviroc

Maraviroc (UK-427 857), a selective and reversible CCR5 co-receptor antagonist, has been shown to be active in vitro against a wide range of clinical isolates [29]. In human immunodeficiency virus type 1 (HIV-1)–infected patients, maraviroc given as monotherapy for 10 days reduced HIV-1 viral load dose dependently by up to 1.6-$\log_{10}$ copies, consistent with currently available agents that comprise the cornerstone of highly active antiretroviral therapy. The objective of this study was to develop a novel combined viral dynamics/operational model of (ant-)agonism that describes the PD effects of maraviroc, a noncompetitive CCR5 inhibitor, on viral load. A common theoretical framework based on receptor theory and the operational model of (ant-)agonism has been developed to describe the binding of maraviroc to the CCR5 receptor and the subsequent decrease in viral load. The effect of maraviroc was modeled using an inhibitory $E_{max}$ model in the PD model acting on the infection rate constant of the virus and target cells. With this parameterization, it was assumed that the antagonistic effect of maraviroc is noncompetitive, which is consistent with recent in vitro findings. In terms of receptor theory and an operational model of drug action, the viral replication process was modeled as a binding–stimulus–response cascade in which the virus acts as an agonist, as shown in the five steps in Figure 8.

In the binding–stimulus–response cascade shown, $T_{max}$ is the maximum concentration of activated target cells that can be infected, $V$ is the virus con-

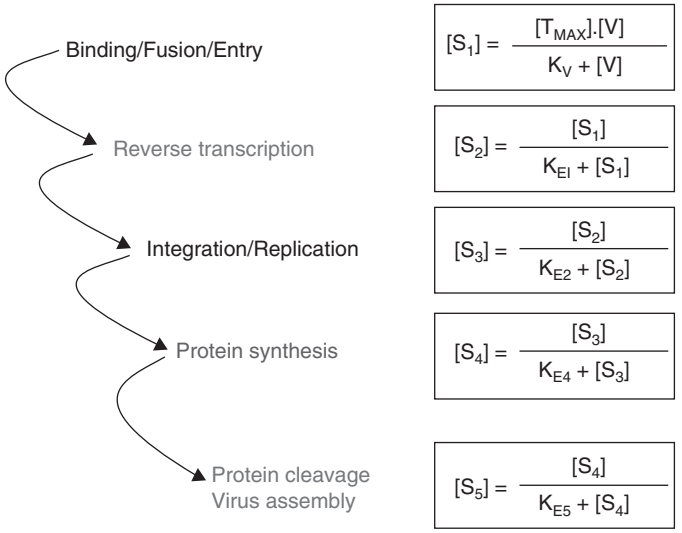


**Figure 8** Viral dynamics–operational model (binding–stimulus–response). (From ref. 29, by permission of Wiley-Blackwell.)

centration, $K_v$ is the theoretical concentration of viruses that elicits half the maximum binding to the activated receptors on the target cells, and $K_{ex}$ is the theoretical stimulus ($S_X$) at step $X$ that elicits half the maximum response ($S_{X+1}$) at step $X + 1$ ($X = 1$ to 5). This new model provided an explanation for the apparent discrepancy between the in vivo binding of maraviroc to the CCR5 receptor ($KD$ = 0.089 ng/mL) and the estimated in vivo inhibition ($IC_{50}$ = 8 ng/mL) of the infection rate. The estimated KE value of the operational model indicates that only 1.2% of free activated receptors are utilized to elicit 50% of the maximum infection rate. The model developed suggests that the target cells, when activated, express more receptors (spare receptors) than needed. In the presence of maraviroc, these spare receptors first require blocking before any decrease in the infection rate, and consequently in the viral load at equilibrium, can be detected. The model allowed the simultaneous simulation of the binding of maraviroc to the CCR5 receptor and the change in viral load after both short- and long-term treatment. The current model will be used to guide the selection of the optimal dosage regimen for maraviroc in HIV-1–infected patients.

## SUMMARY

Considering the complexity, risk, and cost involved in drug discovery and development processes, it is imperative that companies adopt strategies and technologies that will facilitate the identification and development of new therapies. The development of biomarkers and clinical trial endpoints as well as PK/PD modeling were recognized as key elements that can improve decision making and render development processes more effective and more likely to result in safe and beneficial products for patients.

A new framework for biomarker and surrogate endpoint use in drug development was recommended by the FDA. As part of this process, disease-specific biomarkers will be evaluated to determine whether some biomarkers may be elevated to surrogate endpoints. This will be particularly attractive because biomarkers can be measured more quickly and easily and at lower cost than can true clinical endpoints. Appropriate PK/PD modeling of biomarkers and surrogate endpoints will facilitate proof-of-concept demonstrations for target modulation; enhance the rational selection of an optimal drug dose and schedule; aid decision making, such as whether to continue or close a drug development project, accelerate drug approval, minimize uncertainty associated with predicting drug safety and efficacy, and decrease the overall costs of drug development.

The future of PK/PD modeling of biomarkers holds great challenges and great promises. Pharmaceutical companies that implement PK/PD modeling of biomarkers in their early drug development strategy should have a distinct competitive advantage, with powerful insights and guidance for effective and efficient rational drug development, and by providing safe and efficacious

medicines for individual patients. Although the science of PK/PD modeling and biomarkers is constantly evolving, these disciplines provide a continuous process to link what has been learned from today's drug development cycle to the next generation of biomarkers, assays, and mechanism-based models.

## REFERENCES


1. Gao J, Garulacan LA, Storm SM, et al. (2005). Biomarker discovery in biological fluids. *Methods*, 35:291–302.
2. Goodsaid F, Frueh F (2007). Biomarker qualification pilot process at the US Food and Drug Administration. *AAPS J*, 9:E105–E108.
3. Colburn WA (1997). Selecting and validating biologic markers for drug development. *J Clin Pharmacol*, 37:355–362.
4. Lee JW, Hulse JD, Colburn WA (1995). Surrogate biochemical markers: precise measurement for strategic drug and biologics development. *J Clin Pharmacol*, 35:464–470.
5. Aarons L, Karlsson MO, Mentré F, Rombout F, Steimer JL, van Peer A (2001). COST B15 experts: role of modelling and simulation in phase I drug development. *Eur J Pharm Sci*, 13:115–122.
6. Wagner JA, Williams SA, Webster CJ (2007). Biomarkers and surrogate end points for fit-for-purpose development and regulatory evaluation of new drugs. *Clin Pharmacol Ther*, 81:104–107.
7. Colburn WA, Lee JW (2003). Biomarkers, validation and pharmacokinetic–pharmacodynamic modelling. *Clin Pharmacokinet*, 42:997–1022.
8. Colburn WA (2000). Optimizing the use of biomarkers, surrogate endpoints, and clinical endpoints for more efficient drug development. *J Clin Pharmacol*, 40(12 Pt 2):1419–1427.
9. Colburn WA (2003). Biomarkers in drug discovery and development: from target identification through drug marketing. *J Clin Pharmacol*, 43:329–341.
10. Williams SA, Slavin DE, Wagner JA, Webster CJ (2006). A cost-effectiveness approach to the qualification and acceptance of biomarkers. *Nat Rev Drug Discov*, 5:897–902.
11. National Institutes of Health (1999). *Biomarkers and Surrogate Endpoints: Advancing Clinical Research and Applications*. NIH, Bethesda, MD.
12. Rolan P, Atkinson AJ Jr, Lesko LJ (2003). Use of biomarkers from drug discovery through clinical practice: report of the Ninth European Federation of Pharmaceutical Sciences Conference on Optimizing Drug Development. *Clin Pharmacol Ther*, 73:284–291.
13. Meibohm B, Derendorf H (2002). Pharmacokinetic/pharmacodynamic studies in drug product development. *J Pharm Sci*, 91:18–31.
14. Derendorf H, Meibohm B (1999). Modeling of pharmacokinetic/pharmacodynamic (PK/PD) relationships: concepts and perspectives. *Pharm Res*, 16:176–185.
15. Derendorf H, Lesko LJ, Chaikin P, et al. (2000). Pharmacokinetic/pharmacodynamic modeling in drug research and development. *J Clin Pharmacol*, 40:1399–1418.

16. U.S. Department of Health and Human Services Food and Drug Administration (2004). *Challenge and Opportunity on the Critical Path New Medical Products*. FDA, Washington, DC, pp. 1–31.
17. Lalonde RL, Kowalski KG, Hutmacher MM, et al. (2007). Model-based drug development. *Clin Pharmacol Ther*, 82:21–32.
18. Miller R, Ewy W, Corrigan BW, et al. (2005). How modeling and simulation have enhanced decision making in new drug development. *J Pharmacokinet Pharmacodyn*, 32:185–197.
19. Gieschke R, Steimer JL (2000). Pharmacometrics: modelling and simulation tools to improve decision making in clinical drug development. *Eur J Drug Metab Pharmacokinet*, 25:49–58.
20. Zhang L, Sinha V, Forgue ST, et al. (2006). Model-based drug development: the road to quantitative pharmacology. *J Pharmacokinet Pharmacodyn*, 33:369–393.
21. Sheiner LB, Steimer JL (2000). Pharmacokinetic/pharmacodynamic modeling in drug development. *Annu Rev Pharmacol Toxicol*, 40:67–95.
22. Peck CC, Barr WH, Benet LZ, et al. (1992). Opportunities for integration of pharmacokinetics, pharmacodynamics, and toxicokinetics in rational drug development. *Clin Pharmacol Ther*, 51:465–473.
23. Holford NHG, Kimko HC, Monteleone JPR, Peck CC (2000). Simulation of clinical trials. *Annu Rev Pharmacol Toxicol*, 40:209–234.
24. Lockwood P, Ewy W, Hermann D, Holford N (2006). Application of clinical trial simulation to compare proof-of-concept study designs for drugs with slow onset of effect; an example in Alzheimer's disease. *Pharm Res*, 23:2050–2059.
25. Yassen A, Kan J, Olofsen E, Suidgeest E, Dahan A, Danhof M (2007). Pharmacokinetic–pharmacodynamic modeling of the respiratory depressant effect of norbuprenorphine in rats. *J Pharmacol Exp Ther*, 321:598–607.
26. Mager DE, Neuteboom B, Jusko WJ (2005). Pharmacokinetics and pharmacodynamics of PEGylated IFN-beta 1a following subcutaneous administration in monkeys. *Pharm Res*, 22:58–61.
27. Lim KS, Kwon JS, Jang IJ, et al. (2007). Modeling of brain D2 receptor occupancy-plasma concentration relationships with a novel antipsychotic, YKP1358, using serial PET scans in healthy volunteers. *Clin Pharmacol Ther*, 81:252–258.
28. Pillai G, Gieschke R, Goggin T, Jacqmin P, Schimmer RC, Steimer JL (2004). A semimechanistic and mechanistic population PK–PD model for biomarker response to ibandronate, a new bisphosphonate for the treatment of osteoporosis. *Br J Clin Pharmacol*, 58:618–631.
29. Jacqmin P, McFadyen L, Wade JR (2008). A receptor theory–based semimechanistic PD model for the CCR5 noncompetitive antagonist maraviroc. *Br J Clin Pharmacol*, 65(Suppl 1):95–106.

# VALIDATING IN VITRO TOXICITY BIOMARKERS AGAINST CLINICAL ENDPOINTS

Calvert Louden, Ph.D.
*Johnson & Johnson Pharmaceutical, Raritan, New Jersey*
Ruth A. Roberts, Ph.D.
*AstraZeneca Research and Development, Macclesfield, UK*

## INTRODUCTION

Biomarkers have been used for many years to predict and track biological changes in tissue and organ systems. The conceptual framework for their use was captured in the output of the Biomarkers Definitions Working Group [1], which provided a valuable working definition and some challenges to the scientific community. Here, a *biomarker* was defined as "a characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes or pharmacological responses to a therapeutic intervention" [1]. The working group noted that biomarkers may have the greatest value in early efficacy studies and safety evaluations with applications that could include prediction and monitoring of clinical response to an intervention [1].

### Biomarkers in Drug Development

Biomarkers have been used for many years as indicators of biological change. In the course of drug development, biomarkers enjoy utility both as predictors

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

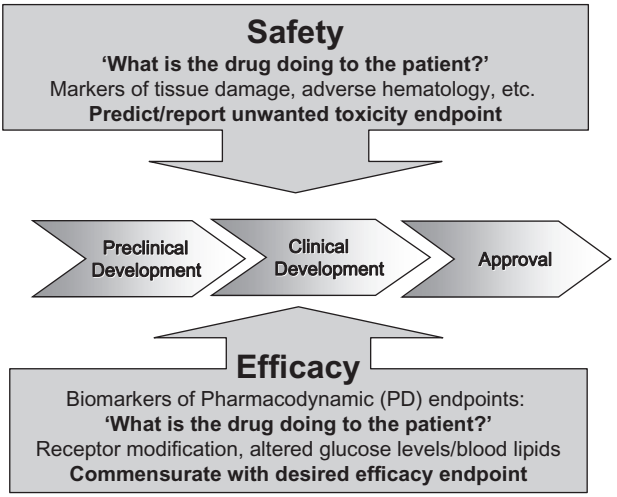


**Figure 1** Biomarkers in clinical medicine. Biomarkers of safety and efficacy can be used together as part of the overall data set to assist in decision making.

of possible efficacy and also as warning signs of potential toxicity (Figure 1). For efficacy endpoints, biomarkers are used to address the pharmacodynamic question "What is the drug doing to the patient?" commensurate with desired outcomes such as receptor modification or blood glucose changes. For safety endpoints, biomarkers may be used as indicators of unwanted endpoints such as tissue damage or adverse hematology.

Although these two types of biomarker serve diverse purposes, the parameters dictating their identification, development, and utility are very similar. However, there is one key aspect in which safety biomarkers differ; safety biomarkers are most useful when they predict rather than report damage. For example, an efficacy biomarker can provide rapid confirmation that the drug has hit its target, perhaps by confirmation of growth factor receptor inhibition in a skin biopsy. However, a safety biomarker of liver damage would ideally precede the damage (either in dose or time) to be of maximum use to the clinician. Despite this, biomarkers that report rather than predict can still be useful if they prevent further damage or permit cessation of treatment and subsequent recovery.

## Toxicity Biomarkers In Vitro

The driving force for better clinical biomarkers of potential adverse events is to monitor patient safety, and many such biomarkers will be discovered and developed within the context of human medicine. However, from a nonclinical

perspective the discovery and development of toxicological biomarkers facilitates the translation to the clinic of knowledge that will avoid or minimize the risk of occurrence of a potential unwanted outcome identified in animal studies. Thus, it is key to identify preclinical biomarkers of the adverse response of animals to potential new drugs and to evaluate fully their relevance and applicability to humans.

Potential biomarkers can be identified in vitro, and their translation to in vivo, either in animal or humans, evaluated at a later date. This concept is developed later in this review with specific examples and case histories. One argument against this approach is that findings in animals in vivo or in animal-derived tissue in vitro may not be relevant for human safety; however, the job of the clinician is to manage the best possible therapy while minimizing the risk of adverse reactions. Thus, such data would form part of an integrated assessment of safety and efficacy that can inform preclinical and clinical decisions.

Clinically useful biomarkers of toxicity can be derived from appropriately designed preclinical experiments, and such biomarkers can significantly assist safe progression into the clinic with an approach design to minimize patient risk. These biomarkers can be derived from in vitro or in vivo mode-of-action studies in preclinical species then translated to potential clinical use or, as we present here, could already be established in clinical practice. This translational science approach requires close collaboration between the clinical and the preclinical scientists.

## Characteristics of Biomarkers

An important consideration in the use of biomarkers is the practicality of subsequent analysis, which is likely to be far more successful and robust if based on an uncomplicated physiological or biochemical assay that can be carried out with minimal training and preexisting equipment (Figure 2). Nonetheless, the need for specialist techniques or expertise should not exclude a potential biomarker, especially where it contributes to advancing therapies for a serious or unmet medical need. Other practical considerations include the need for assays to be transferrable across international boundaries, supported by low background and variability of the measured parameter coupled with a robust increase or decrease signifying the biological endpoint of interest. Indeed, lower and upper limits for normal values would be better if minimally influenced by age, gender, or race within the human population. Similarly, a biomarker of drug-induced toxicity would be better if it were largely unaffected by disease states expected in the patient population of interest or, indeed, any co-medication that might be present or recently administered. However, this is often not the case, and the best that can be expected is some understanding of these relationships to mitigate against difficult or erroneous interpretation. In addition, an ideal biomarker should exhibit

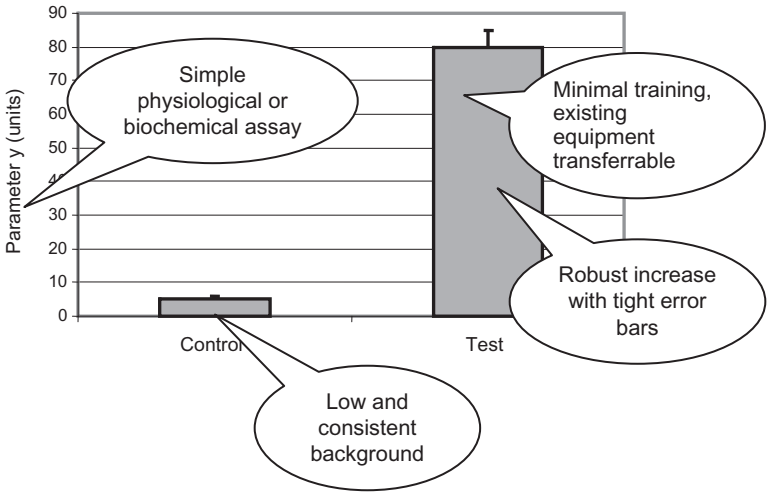


**Figure 2** Technical aspects of biomarker utility. Biomarkers must be sufficiently robust biologically and technically to transfer across patient groups and across international boundaries. This is illustrated by the different variables that need to be overcome or accommodated in the assay.

sufficient sensitivity (not too many false negatives) and specificity (not too many false positives) to underpin clinical decisions.

As well as the technical aspects of biomarker assays, there are additional practical considerations to enhance appropriateness and usability. First, a biomarker needs to be accessible and ideally could be noninvasive. If this is not possible, a biomarker could be measurable in blood and other body fluids obtained with minimal additional impact on the subject. If this cannot be achieved, biomarkers may be detectable with increasing impact on the patient via a peripheral or central biopsy. Finally, biomarker research for the advancement of medical knowledge can be conducted after autopsy. Examples of noninvasive methods would include standard physiological measures of cardiovascular parameters such as blood pressure, heart rate, and electrocardiogram, or could be ophthalmological. Fluids could include feces, saliva, and urine, in addition to standard blood sampling. Skin and hair follicles would be examples of peripheral biopsy and muscle and liver needle samples provide the most common route of more invasive biopsy.

## Biomarkers of Safety

One key aspect where biomarkers of safety differ from those of efficacy is that they preferably predict rather than report the biological effect they are flagging. Thus the biomarker should ideally precede that biological effect either in dose (i.e., be detectable at doses below where the damage occurs) or in time

(be detectable perhaps after one dose when the damage requires repeated dosing). These biomarker characteristics would provide support and confidence to the clinician during dose escalation in early clinical trials or during the transition from single to multiple doses. During later clinical development, a predictive biomarker could monitor for onset in a patient who is beginning to develop toxicity perhaps after several months of drug use. In the absence of evidence that a biomarker can predict onset, an ideal biomarker of drug-induced toxicity would characteristically show detectable changes prior to any serious or irreversible tissue injury and would also show reversibility of the biomarker signal upon removal of the toxicological insult.

## IN VITRO TOXICITY BIOMARKERS

Having considered some general points on the development and validation of toxicity biomarkers, we now present several working examples that explore the concept of in vitro toxicity biomarkers and their translation to the clinic. Opportunities and caveats are explored with an emphasis on assisting clinical progression and decision making. Toxicity biomarkers in the in vitro context are defined as quantitative measurable characteristics that serve as indicators of a pathologic process or related biochemical or molecular events.

### Endocrine and Exocrine Pancreatic Toxicity

The potential for development of clinical drug-induced diabetes is well recognized, and molecules such as cyproheptadine and structurally related compounds such as cyclizine and chlorcyclizine have been shown to cause pancreatic islet toxicity. However, this is not unique to this class of structurally related compounds because a similar finding has been observed following oral administration of some quinolone antimicrobial agents. In vivo, this effect is characterized by cytoplasmic islet vacuolation, decreased insulin secretory granules, decreased circulating insulin levels, intolerance to glucose challenge, and increased glycosylated albumin. The working hypothesis is that these compounds were associated with functional impairment of insulin synthesis and/or secretion, and an in vitro screen would be invaluable in identifying this liability early in the drug development process.

The clonal insulin-producing cell line RINm5F is widely used as a model to detect compound-induced loss and/or reduction of insulin in vitro that is comparable to in vivo findings. Using this in vitro model, it was determined that cyproheptadine and selected analogs caused inhibition of insulin biosynthesis without any effects on pre-pro-insulin mRNA. These data suggested that inhibition of transcriptional regulation of insulin was an unlikely mode of action responsible for cyproheptadine-induced pancreatic islet cell toxicity. The limitation of the RINm5F cells is that they are not glucose-sensitive and as such, cannot be used to study glucose-dependent insulin secretion. When

this effect is suspected, the use of Min 6 (mouse) and HT15 (hamster) cells are recommended.

Isolated mouse pancreatic acini provide an ex vivo tool used to assess potential exocrine pancreatic toxicity. This is achieved by evaluating digestive enzyme secretion through an assessment of basal amylase and amylase release in response to carbachol and cholecystokinin octapeptide (CCK8). In vitro, cyproheptadine at high doses causes marked release of amylase without any breaks in structural integrity, thus ruling out cytotoxicity as the likely cause of the marked amylase release. Furthermore, cyproheptadine can abolish carbachol-stimulated amylase release, and this may be related to its anti-muscarinic pharmacological property. This approach can serve as an in vitro toxicity screen to eliminate and/or reduce undesirable effects on the exocrine as well as endocrine pancreas and, additionally, the pathophysiology of these potential effects can be investigated.

## Bone Marrow Toxicity

Within the last decade, clinical hematopoietic toxicity has accounted for 10 to 20% of approved drug withdrawals and is a major cause of early candidate drug attrition. Identification of potential hematopoietic toxicity as a hazard is done only when in vivo studies are initiated even though there is strong evidence suggesting that in vitro assays could have a very good predictability. A three-tiered approach that enables a large number of compounds to be evaluated at the lead optimization phase was developed to improve the assessment of potential hematotoxicity. A key question in this approach is whether there is clinical translation and predictability of this combined in vitro, in vivo preclinical evaluation.

This novel approach uses multiple assays, cell lines representative of broad bone marrow lineages (myeloid, erythroid, lymphoid, and stromal), and a nonhematopoietic cell line (tier 1). In general, there is a good correlation between the mouse myeloid cell line and the mouse colony-forming unit/GM (CFU-GM) assays and compounds with lineage-specific in vivo bone marrow toxicity showed similar lineage specificity in the tier 1 assays. An erythrotoxicant such as chloramphenicol has an $IC_{50}$ value of 8 μM in the erythroid cell line assay and >100 μM for the other cell lines. CFU assays (tier 2) are then run to qualify and confirm the findings of tier 1. Tier 2 consists of mouse, rat, dog, and human CFU-GM assays to predict species sensitivity, but these assays are tedious and time consuming, enabling only a limited number of compounds to be evaluated. Increased throughput was achieved when a 96-well plate CFU-GM assay was developed that yielded results that correlated well with the traditional CFU-GM assay. The data generated from the tiers 1 and 2 can trigger frontloading of in vivo studies to evaluate a more comprehensive hematology profile and include use of additional technologies such as flow cytometry (tier 3). In summary, a three-tiered approach using in vitro, ex vivo, and in vivo studies allows rapid, high-throughput assessment of compounds

with potential chemistry-related bone marrow toxicity to be eliminated in the early preclinical stages of drug development.

## Liver Toxicity

The overall goal of developing an in vitro toxicity testing system is for inclusion in a toolbox as an element of investigative toxicology. This should contribute to hypothesis-based research aimed at identification of mode-of-action-based toxicity biomarkers, which can extend to preclinical and clinical phases of drug development. The current preclinical and clinical diagnostic panel used to monitor hepatocellular toxicity has been quite reliable except in cases of human drug-induced liver injury, which occurs with a low incidence and prevalence, ranging from 1 in 10,000 to 1 in 100,000. This is also referred to as *idiosyncratic hepatotoxicity*, meaning that the adverse events are a consequence of individual and unpredictable patient responses. It is well recognized that there is an urgent need to identify and develop a more successful prediction model coupled with diagnostic and prognostic biomarkers for drug-induced idiosyncratic liver injury in humans. This type of human liver toxicity is multifactorial, including individual-specific responses, and because of this, it is highly unlikely that any preclinical and/or clinical studies can be powered appropriately to unmask this risk. Furthermore, it is difficult to assess the value of our preclinical toxicology studies and their relevance to predicting drug-induced idiosyncratic liver injury in humans, because we lack understanding of the mode of action and pathophysiology of this unique type of toxicity. However, there is extensive literature related to the predictability of hepatotoxicity, and there is good concordance with in vitro and in vivo assessments.

Several different strategies have been employed to detect various forms of hepatotoxicity, including, necrosis, apoptosis, and phospholipidosis. The in vitro hepatotoxicity predictive strategy uses hepatic microsomes; immortalized cell lines; primary hepatocyte cell cultures from humans, rodents, and nonrodents; and liver slices. The latter, in particular, enables a better approximation and evaluation of metabolites that may be formed and an assessment of their hepatotoxicity potential determined and in vitro–in vivo comparisons made. For example, using this approach, studies were done to characterize the metabolic profile of troglitazone, an approved drug that has now been withdrawn from the market because of hepatotoxicity and hepatic failure requiring liver transplants. Data from a series of studies using rat and human hepatic microsomes as well as in vivo rat studies suggested that troglitazone hepatotoxicity was caused by several reactive intermediates covalently binding to hepatic proteins that may undergo redox cycling and induce oxidative stress, causing cell damage. This hypothesis was strongly supported by the data suggesting that not only is troglitazone an inducer of P450 3A, but is also responsible for metabolism of the thiazolidinedione ring, a key structural element of many of the peroxisome prolifercator–activated receptor (PPAR) agonists.

Taken together, troglitazone acts as an inducer of enzymes that catalyzes its biotransformation to chemically reactive intermediates; this autoinduction of its own metabolism is ultimately detrimental to the cells. Additionally, this combined approach was able to show that in humans these intermediates could be formed, thus identifying a potential etiology for troglitazone-induced hepatotoxicity in humans. Utilization of liver slice technology, co-cultures, and whole liver perfusion can be quite useful, but these are low-throughput assays that are not always suitable for screening. The advantage is that toxicity to the biliary and other hepatic cellular constituents can be assessed. Thus predicting the hepatotoxicity potential using in vitro cell culture systems can also employ the well-recognized preclinical and clinical biomarkers of hepatocellular damage, such as ALT (alanine aminotransferase), AST (aspartate aminotransferase), GLDH (glutamate dehydrogenase), MDH (malate dehydrogenase), PNP (purine nucleoside phosphorylase), and PON-1 (paraoxonase-1). Other enzymes, such as ALP (alkaline phosphatase), GGT (γ-glutamyl transferase) and 5′-nucleotidase (5′NT) can also be used when toxicity to the biliary tree is suspected.

Cytotoxicity (necrosis and apoptosis) in vitro is the primary endpoint, and as such, biomarkers for potential hepatotoxicity clinical monitoring should include necrosis as well as apoptosis. Recent evidence suggests that this may be possible by measuring the soluble pool of cytokeratin 18 (CK18) and the caspase-cleaved CK18 fragments in conjunction with the traditional markers. The in vitro surrogate system can also provide valuable information on the suitability of the appropriate biomarker and potential mechanism of action related to toxicity, such as mitochondrial dysfunction, metabolic pathways, and/or CYP450 induction.

In summary, in vitro hepatocyte cell culture systems and tissues derived from in vivo studies will identify the appropriate biomarker for monitoring preclinical and clinical hepatotoxicity but these data can also contribute to hypothesis-driven mode-of-action investigative toxicity studies. This will include molecular profiling, metabonomics, transcriptomics, and proteomics, all of which are useful tools that could aid in identification of novel biomarkers of hepatotoxicity.

## Phospholipidosis

Phospholipidosis (PLD) is characterized by concentric-layered multilamellar intracellular lysosomal inclusion bodies that are often composed of complex phospholipids, parent drug, and/or metabolites. Accumulation will often occur in several different cell types of the hepatobiliary, immune, and nervous systems. Typically, hepatocytes, biliary epithelial cells and macrophages of lymph nodes, and pulmonary alveoli and ganglia and nonganglia neuronal cell bodies of the central nervous system can be affected. The evidence to date suggests that phospholipidosis is a structural-related toxicity of cationic amphiphilic compounds irrespective of pharmacologic action. The finding of phos-

pholipidosis may also be associated with inflammation, severe organ damage, and possibly impairment of immune function. Although drugs are marketed that cause phospholipidosis preclinically and clinically, this is an undesirable profile for potentially new candidate drugs, and as such, this liability should be identified and avoided early in the drug discovery process. Therefore, high-throughput in vitro predictive screens can add value, particularly if phospholipidosis potency can be ranked and supported by in vivo data.

Several methodologies have been evaluated to assess neutral and phospholipid content as an index of phospholipidosis in cells growing in culture. However, accumulation of NBD-PE as a result of cytotoxicity induces false-positive results, particularly at high concentrations. Recently, a high-throughput, validated, predictive, sensitive, and selective multichannel fluorescence-based in vitro PLD assay was developed to reduce the false-positive limitation of cytotoxicity. This assay uses I-13.35 adherent mouse spleen macrophages cultured in 96-well plates with fluorescent-tagged phospholipids. Cells with an intact nucleus were differentiated from dead cells using ethidium staining and cell gating that rejects dead cells. Using this improved technique, 26 of 28 positive phospholipidogenic compounds were identified. These findings aided application of this methodology to other techniques, such as flow cytometry, which may be used in preclinical toxicology studies and clinical trials. For example, flow cytometric analysis coupled with Nile Red staining was used to detect neutral and phospholipids in a monocyte cell line, U397. Application of this methodology was utilized in (in vivo) toxicology studies, and this raised the possibility that preclinical toxicology and clinical assessment of phospholipidosis could be done using peripheral blood cells and flow cytometry.

## CONCLUSIONS

Overall, there is a compelling need for in vitro toxicity biomarkers for clinical endpoints. Toxicity biomarkers in the in vitro context are defined as quantitative measurable characteristics that serve as indicators of a pathologic process or related biochemical or molecular events. Conceptually, bridging biomarkers of toxicity should include not only the traditional parameters of biofluid and physiological measurements, but also measurable endpoints that could serve as indicators of potential adverse events preclinically and clinically even though they have been derived from in vitro, ex vivo, and in vivo studies. The use and application of this combination will undoubtedly improve the toxicologist's ability to identify human hazards so that the appropriate risk assessment and management strategy can be developed. In developing in vitro assays to identify biomarkers with potential clinical application and utility, a clear understanding and determination of what that biomarker will assess must be defined. For example, exaggerated pharmacologic action of a molecular target may be associated with an undesirable effect resulting in toxicity. In

such cases, single/multiple markers and/or assays may be required for utilization in screens during the early drug discovery phase, with continued assessment in preclinical and clinical development. If successful, these data can be invaluable in deriving the therapeutic index of a drug. In the case of a specific target organ toxicity, the ultimate goal is to identify biomarkers of toxicity that can be used as a general screen that reflects cellular damage, regardless of mode of action in a high-throughput manner. Therefore, it is imperative that in vitro assays and the appropriate platforms be developed to identify relevant toxicity biomarkers that will be useful during preclinical and clinical development.

## REFERENCE

1. Atkinson AJ Jr, Colburn WA, DeGruttola VG, et al. (2001). Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. *Clin Pharmacol Ther*, 69:89–95.

# OPPORTUNITIES AND PITFALLS ASSOCIATED WITH EARLY UTILIZATION OF BIOMARKERS: CASE STUDY IN ANTICOAGULANT DEVELOPMENT

Kay A. Criswell, Ph.D.
*Pfizer Global Research and Development, Groton, Connecticut*

## INTRODUCTION

The cost of developing new drugs continues to rise and demands an ever-increasing percentage of total research and development (R&D) expenditures [1]. Recent studies have shown that there is less than a 10% success rate between the first human trials and the launch of a new product [1]. When coupled with an attrition rate of over 70% in phase II and nearly a third of all new compounds in phase III trials [1–3], the gravity of the need for reliable, early-decision-making capability is evident. All pharmaceutical companies are focused on ways to reduce the cost of discovering and developing new drugs. In general, three areas have received the greatest attention: (1) improving target validation, (2) selecting candidates with a greater chance of success, or (3) identifying those compounds that will fail earlier in their development. Inadequate efficacy or poor safety margins are the two main reasons for compound attrition. Therefore, developing a strategy for development of safety

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

and efficacy biomarkers is key to early prediction of compound success or failure. Utilization of biomarkers in preclinical testing and ex vivo human testing may provide valuable information for compound progression and success in the clinic, but it is not without problems.

Studying the effects of markers on disease processes is not new. It has probably existed for centuries, as evidenced by such obsolete practices as tasting urine to determine the diabetic status of a patient. Utilization of biomarkers for disease diagnosis and prognosis has escalated dramatically during this century with the improved diagnostic power, reliability, and speed of routine clinical pathology, biochemical, and genomic data. However, when studying the specific effects of a drug on marker activity, utilization of biomarkers has a more recent history. One of the earliest documented cases may be the early trials of lamotrigine, which effectively utilized the electroencephalograph (EEG) as a biomarker and demonstrated decreased epileptiform activity [4]. Twenty years after the lamotrigine study the urgency for relevant and early biomarkers of safety and efficacy during drug development has been recognized, and it is not uncommon that each new drug development program is accompanied by a biomarker request.

Not only has the frequency of biomarker development exploded, but the timing of biomarker utilization has been pushed earlier in development. There is an ever-increasing demand to test and screen potential candidates before good laboratory practices (GLP) studies and commitment to human trials. Two problems inherently complicate the use of early biomarkers to preclinical outcomes: (1) truly understanding the disease process and the therapeutic intervention process, and (2) species–species differences that may alter the translation and application of the biomarker.

Understanding the disease process appears obvious, but as already noted, understanding the target is still a key area for decreasing attrition. Regardless of the expanding body of knowledge surrounding disease processes, infectious diseases may be the lone area where the disease process is fully understood. Successful therapeutic intervention and drug sensitivity can be predicted accurately with well-characterized biomarkers of causal organism growth and survival [5]. For biomarkers to be truly effective in drug development, they need to reflect a highly specific biochemical or molecular change that occurs in the disease process, is altered by the therapeutic intervention, and occurs prior to any downstream effects on clinical endpoints. This is a lofty goal, as most disease processes affect multiple pathways, and feedback mechanisms further complicate the overall biology. The further impact of nontranslatable, species-specific characteristics makes daunting the task of providing biomarkers early in development. There are, however, diseases and biological pathways that have a fairly broad acceptance regarding the essential components. Additionally, certain areas of drug development target activation or inhibition of a highly specific molecule or receptor within those pathways. These conditions provide a unique opportunity for biomarker success when conducted early in the drug discovery process.

The coagulation pathway is an example of a well-characterized model, and specific inhibitors of this pathway are being explored actively as new candidates for anticoagulant therapy. One of the documented areas is the utilization and challenge of early implementation of biomarkers to predict the clinical outcomes of factor Xa (FXa) inhibitor compounds. Venous and arterial thromboembolic disorders have a substantial impact on human health and morbidity [6]. Although these conditions have been treated for many years with the vitamin K antagonist warfarin (coumadin), or with intravenous or subcutaneous heparin administration, the opportunity to provide therapeutic intervention with a safer profile, fewer side effects, less intra- and intersubject variability, a better route of administration, and less need for continuous coagulation monitoring is attractive. Novel coagulation therapeutics have generated tremendous interest due to the medical need and opportunity to improve on current treatment strategies. FXa occupies a pivotal position within the coagulation cascade and is a highly attractive candidate as a target for novel coagulation intervention. This enzyme links the intrinsic and extrinsic coagulation pathways and is the rate-limiting step in thrombin formation [7].

Changes within the intrinsic coagulation pathway are routinely monitored by a clotting assay called the *activated partial thromboplastin time* (aPTT), whereas changes in the extrinsic pathway are assessed by increases or decreases in another clotting assay called the *prothrombin time* (PT) (7, Figure 1). Both assays are assessed spectrophotometrically or mechanically and are expressed

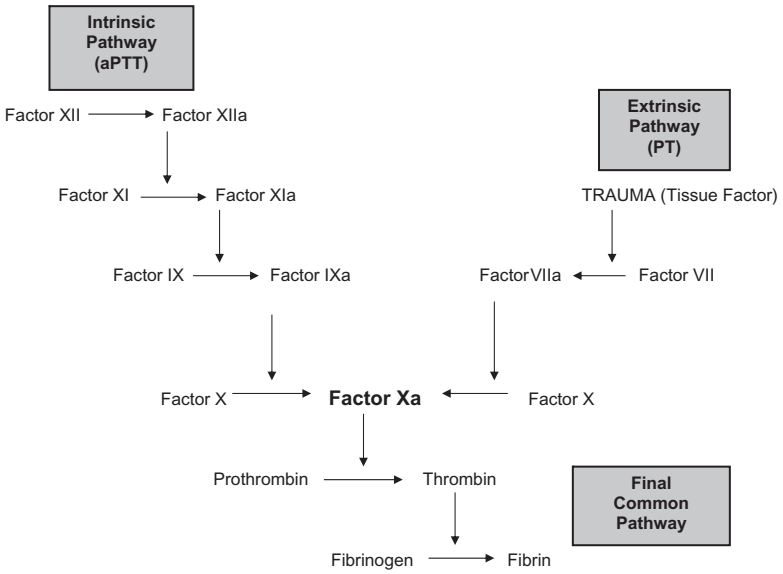


**Figure 1** Intrinsic and extrinsic coagulation pathways.

as the time to clot formation after addition of an exogenous clot-activating agent. The aPTT is used routinely to monitor the safety and level of anticoagulation of heparin therapy and coumadin in monitored via the PT assay. Safety profiles for these compounds have been established following years of use so that fold increases over the predose aPTT or PT can be used to predict therapeutically effective doses of these drugs versus levels associated with inadequate or excessive coagulation. Additionally, aPTT and PT are routine and highly standardized clinical assays. Testing reagents are standardized and instrumentation has undergone rigorous scrutiny to pass U.S. Food and Drug Administration (FDA) requirements because of its routine use in clinical diagnostics. The reliability of aPTT and PT as biomarkers of anticoagulant therapy safety is also fairly well established. Since activation of factor X is required for completion of both the intrinsic and extrinsic pathways, it appears logical that inhibition of factor X activation should prolong both PT and aPTT.

It would be anticipated that early assessment of novel anticoagulants such as FXa inhibitors could readily be assessed with the aPTT and PT assays. Furthermore, if the new compound does not require metabolic activation, ex vivo incubation of human plasma with the compound of interest followed by aPTT and PT assessment may provide a reliable method to rapidly assess efficacy. Despite their reliability and acceptability, PT and aPTT do not always reflect the anticoagulant activity of novel compounds. Rivaroxaban is an oral direct inhibitor of activated factor X [8]. In a rat venous model of the inferior vena cava, Rivaroxaban produced dose-dependent inhibition of FXa and an increase in PT [7]. An inhibition of 32% was associated with a 1.8-fold increase in PT, and nearly 100% inhibition produced a 3.2-fold increase in PT. However, in a rabbit model 92% inhibition of FX was associated with only a modest 1.2-fold increase in PT, demonstrating that species-specific sensitivity is one of the problems associated with monitoring the newer anticoagulants with standard coagulation parameters [8]. Even Coumadin provides a good example of species-specific effects that allow a useful therapeutic for humans but a lethal pest control for rodents [9]. Despite this challenge, utilization of coagulation testing for in vivo and ex vivo screening is well documented in the development of FXa inhibitors [8,10–12].

Rarely is a single biomarker considered definitive for evaluation of safety or efficacy. The drug development approach for Otamixabin and DU-176b incorporated a series of clotting parameter assays and assays to measure the effects on thrombus formation [10–12]. Although clotting parameters would be the preferred biomarker based on the ability to monitor in plasma, use of fairly simple but reproducible assays, cost, and the ability to monitor a clinical population easily, evidence is not established that clotting assays and thrombus formation assays are interchangeable.

For Otamixaban, in vitro coagulation parameters were assessed for their ability to produce a doubling of PT and aPTT. This testing allowed a rank ordering of anticoagulant effects per species of rabbit > human > monkey >

rat > dog [10]. Additionally, aPTT appeared to be the more sensitive biomarker in all species, with aPTT doubling occurring at drug concentrations that were less than half the concentration required for PT doubling. Multiple pharmacological models of thrombosis in rats, dogs, and pigs were also conducted with Otamixaban. In rats, thrombus mass was markedly reduced by nearly 95%, with a corresponding increase in aPTT of 2.5-fold and in PT of 1.6-fold [10]. In contrast, intravenous administration of 1, 5, or 15 μg/mL Otamixaban in the pig model effectively eliminated coronary flow reserves related to this stenosis model at the middle and high dose. PT was also prolonged at the middle and high dose, but aPTT was prolonged only at the high dose. Although pigs were not listed as assessed in the species-specificity model, it suggests that the clotting parameter of choice may vary per species and may not correlate well with thromobosis assays. Furthermore, clinical trial outcomes showed that at anticipated antithrombotic and therapeutic concentrations of 100 ng/mL Otamixaban, neither PT nor aPTT changed appreciably. In contrast, alternative clotting parameters such as the HepTest clotting time and the Russell viper venom clotting time, showed substantial prolongation, again suggesting that alternatives to standard PT and aPTT may be preferable [10].

Further work with the oral FXa inhibitor DU-176b provides additional evidence that selection of the right biomarker and appropriate correlation to functional assays is critical. This study was conducted in 12 healthy male volunteers [12]. The antithrombotic effect of DU-176b was assessed by measuring the difference in size of acutely formed, platelet-rich thrombus, pre- and post-drug administration using a Badimon perfusion chamber model under low and high shear force. Subjects received a single 60-mg dose of DU-176b, and pharmacokinetic and pharmacodynamic assessments were conducted at 1.5, 5, and 12 hours postdosing. Pharmacokinetic assessments included PT, international normalization ratio (INR), aPTT, thrombin generation, and anti-factor Xa activity. Drug levels were also assessed. Badimon chamber results demonstrated a strong antithrombotic effect at 1.5 hours with a progressive return toward baseline by 12 hours. All of the pharmacokinetic endpoints showed significant change from pretreatment, suggesting that any of the parameters might be an effective biomarker of DU-176b safety and/or efficacy. However, a close statistical look at this data raises some questions. A comparison of drug concentration level to anti-factor Xa activity and clotting parameters showed the strongest correlation with anti-factor Xa activity ($r^2$ = 0.85), similar correlation with PT and INR ($r^2$ = 0.795 and 0.78, respectively), but a fairly weak correlation with aPTT ($r^2$ = 0.40). This suggests that although Otamixaban and DU-176b are both FXa inhibitors, arbitrary selection of PT or aPTT as a better predictor of drug concentration is problematic. Furthermore, when the antithrombotic effects of DU-176 assessed by Badimon chamber were compared to those obtained by clotting parameters, the correlation was even more challenging. Prothrombin time showed a correlation of $r^2$ = 0.51 at both high and low stress and the correlation with aPTT was only $r^2$ = 0.39 and 0.24 [12]. This suggests that although aPTT is used for monitoring of heparin therapy and PT

is utilized for clinical safety of coumadin, their routine use as factor Xa inhibitors by themselves is insufficient.

## CASE STUDY DATA WITH A DEVELOPMENTAL FXa INHIBITOR

Beyond the published literature that is available, data and personal observations collected during the development of another FXa inhibitor at Pfizer Global Research & Development are now provided to complete this case study approach to biomarker utilization during anticoagulant development. Development of this particular FXa inhibitor ultimately was discontinued, as this compound required intravenous administration and as such lacked marketability compared to oral FXa inhibitors. However, the lessons learned provide further documentation of species-specific and interpretational complications that arise with the utilization of accepted coagulation biomarkers to monitor anticoagulant efficacy and safety for FXa inhibitors.

### Dose Selection for Ex Vivo Experiments

In designing ex vivo experiments to evaluate potential biomarkers, selection of the appropriate drug concentration is critical. Furthermore, when experiments are conducted in multiple species, selection of the same drug concentration for all species is typically not ideal, due to species-specific drug sensitivity. Factor X concentrations vary by species, and the level of FXa inhibition is also variable. Therefore, the concentrations of FXa inhibitor utilized in this particular ex vivo evaluation were selected to achieve a range of FXa inhibition that was modest to nearly complete in all species examined. Pharmacology studies predicted that this FXa inhibitor would result in species-specific factor Xa sensitivity in the order human > dog > rat. Interestingly, this species specificity was not identical to that observed with Otamixaban [10], demonstrating that extrapolating biomarker data even between compounds in the same class may be misleading. For this developmental FXa inhibitor, human plasma was spiked to obtain final drug concentrations of 0, 0.2, 0.6, 1.2, and 6.0 μg/mL. Drug concentrations of 0, 0.4, 2.0, 8.0, and 15.0 μg/mL were selected for dog assessments and 0, 1.0, 4.0, 12.0, and 24.0 μg/mL were used for ex vivo assessments in rats to achieve a comparable range of FXa inhibition compared to that observed in human samples.

Thromboplastin is the reagent that induces clot formation in the PT assay. There is ample documentation that the type and sensitivity of thromboplastin is a critical factor in the effective and safe monitoring of coumadin administration [13–17]. To minimize this variability in PT assays a calibration system was adopted by the World Health Organization (WHO) in 1982. This system converts the PT ratio observed with any thromboplastin into an international normalized ratio (INR). This value was calculated as follows: INR = observed PT ratio$^c$, where the PT ratio is subject PT/control PT and $^c$ is the power value

representing the International Sensitivity Index (ISI) of the particular thromboplastin [18]. This system has proven to be an effective means of monitoring human oral anticoagulant therapy with coumadin and has been implemented almost universally. It allows individuals to be monitored at multiple clinics using varying reagents and instrumentation, while still achieving an accurate assessment of true anticoagulation. However, there is little or no information regarding selection of thromboplastin reagents or use of the INR for monitoring of FXa inhibitors.

Typically, the higher the ISI value, the less sensitive the reagent, and the longer the PT time produced. The most commonly used thromboplastin reagents for PT evaluation are either rabbit brain thromboplastin (of variable ISI values, depending on manufacturer and product) or human recombinant thromboplastin, typically with an ISI of approximately 1.0. Use of the INR is accepted as a more relevant biomarker of anticoagulant efficacy than are absolute increases in PT alone, at least for coumadin therapy [13]. To more fully evaluate the effect of this FXa inhibitor on INR, PT was evaluated using rabbit brain thromboplastin, with ISI values of 1.24, 1.55, and 2.21 and a human recombinant thromboplastin (0.98 ISI). Although either human recombinant thromboplastin or rabbit thromboplastin are considered acceptable reagents for the conduct of PT testing, it was unclear whether these reagents would produce similar results in the presence of an FXa inhibitor or whether the sensitivity of the thromboplastin itself would affect results.

### Effect on Absolute Prothrombin Time

Prothrombin time data obtained using rabbit brain thromboplastin with the three increasing ISI values during these ex vivo studies are presented in Table 1. The source and sensitivity of thromboplastin used in the assay affected the absolute PT value in all species, clearly demonstrating the need to standardize this reagent in preclinical assessment and to be cognizant of this impact in clinical trials or postmarketing, when reagents are less likely to be standardized. As anticipated, addition of the FXa inhibitor to plasma under ex vivo conditions increased the PT in a dose-dependent manner. This increase in PT time length was observed regardless of the ISI value (sensitivity) of the thromboplastin used and occurred in all species (Table 1). Although the absolute time for clot formation generally increased with increasing ISI, this was not true for all assessments.

Table 2 summarizes the maximum change in PT and the range of variability when rabbit brain thromboplastin of varying ISI values was compared to human recombinant thromboplastin. Again, in general, the higher the PT value, the larger the deviation between reagent types. For example, although there was a 2.1-second difference between human and rabbit thromboplastin in untreated human plasma, the difference increased from 5.5, 9.2, 12.0, 13.6, and 38.2 seconds in samples containing 0.2, 0.6, 1.2, 1.8, or 6.0 μg/mL FXa inhibitor, respectively. Dogs were much less sensitive to the type of thrombo-

**TABLE 1 In Vitro Effect of an Experimental Factor Xa Inhibitor on Absolute Prothrombin Time Using a Rabbit Brain Thomboplastin**

| Concentration of Factor Xa Inhibitor (μg/mL) | International Sensitivity Index[a] | | | |
|---|---|---|---|---|
| | 0.98 | 1.24 | 1.55 | 2.21 |
| | PT (s)—Human Plasma | | | |
| 0 | 11.4 ± 0.11 | 13.4 ± 0.17* | 12.9 ± 0.13* | 10.9 ± 0.14 |
| 0.2 | 18.3 ± 0.37 | 23.8 ± 0.50* | 23.4 ± 0.44* | 16.8 ± 0.43* |
| 0.6 | 30.8 ± 0.76 | 37.3 ± 0.86* | 39.9 ± 0.70* | 27.0 ± 0.93* |
| 1.2 | 47.0 ± 1.98 | 51.4 ± 1.51* | 58.9 ± 1.08* | 38.2 ± 1.39* |
| 1.8 | 61.2 ± 2.04 | 61.8 ± 1.58 | 74.8 ± 1.54* | 48.3 ± 2.08* |
| 6.0 | 133.0 ± 3.82 | 111.4 ± 3.17* | 152.3 ± 3.08* | 94.8 ± 4.12* |
| | PT (s)—Dog Plasma | | | |
| 0 | 7.8 ± 0.14 | 8.4 ± 0.07* | 7.1 ± 0.07 | 6.7 ± 0.06* |
| 0.4 | 11.0 ± 0.26 | 11.0 ± 0.13 | 10.5 ± 0.19 | 9.2 ± 0.15* |
| 2.0 | 17.6 ± 0.42 | 16.6 ± 0.23* | 17.9 ± 0.41 | 15.0 ± 0.31* |
| 8.0 | 32.0 ± 0.98 | 27.6 ± 0.49* | 33.6 ± 0.92 | 27.1 ± 0.63* |
| 15.0 | 45.4 ± 1.52 | 36.5 ± 0.73* | 46.7 ± 1.35 | 36.8 ± 0.88* |
| | PT (s)—Rat Plasma | | | |
| 0 | 9.1 ± 0.05 | 15.1 ± 0.07* | 17.1 ± 0.15* | 13.1 ± 0.07* |
| 1.0 | 13.3 ± 0.06 | 20.7 ± 0.13* | 31.3 ± 0.27* | 23.2 ± 0.19* |
| 4.0 | 20.1 ± 0.29 | 30.8 ± 0.23* | 51.8 ± 0.52* | 36.9 ± 0.52* |
| 12.0 | 31.1 ± 0.72 | 46.2 ± 0.49* | 83.6 ± 0.84* | 56.2 ± 0.84* |
| 24.0 | 42.7 ± 1.09 | 60.6 ± 0.73* | 109.4 ± 0.79* | 75.4 ± 1.60* |

[a] *, Mean value ± S.E.M. for 10 individual subjects significantly different from 0.98 ISI thromboplastin means at 5% level by *t*-test, separately by increasing ISI value for individual rabbit thromboplastins.

plastin used and showed smaller maximum changes in PT values. In contrast, rat PT values were highly dependent on the source of thromboplastin, and samples tested with rabbit brain thromboplastin were markedly longer than with human recombinant thromboplastin. Rats showed this high level of thromboplastin dependence even in untreated control samples. Variability of PT in FXa inhibitor-treated human and dog plasma was similar to that observed in controls and did not change appreciably with increasing concentration of drug (Table 2). FXa inhibitor-treated rat plasma showed an approximately twofold increase in variability compared to control.

## Effect on PT/Control Ratio and INR

Generating a PT/control ratio by dividing the number of absolute seconds in the treated sample by the number in the control (untreated) sample provides a second method of assessing PT. If the ISI of the thromboplastin used is close to 1.0, the INR should be similar to the PT/control ratio (Table 3). The PT/

**TABLE 2 Comparison of Human Recombinant Thromboplastin and Rabbit Brain Thromboplastin on Prothrombin Time in Plasma Samples Containing Increasing Concentrations of Factor Xa Inhibitor[a]**

| Species | Intended Drug Concentration (μg/mL) | Maximum Change in PT[b] | Range of Variability[c] (%) |
|---|---|---|---|
| Human | 0 | 2.1 | –4 to +18 |
| | 0.20 | 5.5 | –8 to +30 |
| | 0.60 | 9.2 | –12 to +30 |
| | 1.20 | 12.0 | –19 to +25 |
| | 1.80 | 13.6 | –21 to +22 |
| | 6.00 | 38.2 | –29 to +14 |
| Dog | 0 | 1.1 | –15 to +7 |
| | 0.40 | 1.8 | –7 to –5 |
| | 2.00 | 2.5 | –14 to +2 |
| | 8.00 | 4.9 | –15 to +5 |
| | 15.00 | 8.9 | –20 to +3 |
| Rat | 0 | 8.0 | +43 to +88 |
| | 1.00 | 18.0 | +56 to +136 |
| | 4.00 | 31.7 | +53 to +158 |
| | 12.00 | 52.5 | +48 to +168 |
| | 24.00 | 66.7 | +42 to +156 |

[a]Samples spiked with a factor Xa inhibitor in vitro.
[b]Maximum change in prothrombin time compared to 0.98 ISI human recombinant thromboplastin.
[c]Variability of three increasing ISI levels of rabbit brain thromboplastin compared to human recombinant.

control ratio could be used effectively to normalize thromboplastin differences in untreated human, dog, or rat samples. At predicted efficacious concentrations of FXa inhibitor, the PT/control ratio effectively normalized reagent differences. However, at high concentrations of FXa inhibitor, particularly in the rat, this method lacked the ability to normalize results effectively.

Table 4 shows the corresponding INR values obtained in human, dog, and rat plasma when assessed with rabbit brain thromboplastins of increasing ISI. As anticipated, the PT/control ratio and INR were similar when the ISI was approximately 1. In contrast to the modest differences in PT when expressed as either absolute seconds or as a ratio compared to control value, the INR showed dramatic increases (Table 3). The magnitude of the INR value rose consistently with increasing ISI value and was marked. At the highest dose tested, the INR ranged from 11.1 with the 0.98 ISI reagent to 121.9 with the 2.21 ISI reagent in human samples, 5.6 to 43.4 in dogs, and 4.6 to 48.1 in rats. Assessment of PT in human, dog, or rat plasma containing this developmental FXa inhibitor was affected by the ISI of the thromboplastin selected for the

**TABLE 3 In Vitro Effect of an Experimental Factor Xa Inhibitor on Prothrombin Time/Control Ratio Using a Rabbit Brain Thomboplastin**

| Concentration of Factor Xa Inhibitor (μg/mL) | International Sensitivity Index[a] | | | |
|---|---|---|---|---|
| | 0.98 | 1.24 | 1.55 | 2.21 |
| | PT / Control Ratio (:1)—Human Plasma | | | |
| 0 | 1.0 ± 0.00 | 1.0 ± 0.00 | 1.0 ± 0.00 | 1.0 ± 0.00 |
| 0.2 | 1.6 ± 0.02 | 1.8 ± 0.12 | 1.8 ± 0.21 | 1.5 ± 0.02 |
| 0.6 | 2.7 ± 0.05 | 2.8 ± 0.04 | 2.9 ± 0.03 | 2.5 ± 0.06 |
| 1.2 | 4.1 ± 0.15 | 3.8 ± 0.08 | 4.6 ± 0.06* | 3.5 ± 0.09 |
| 1.8 | 5.4 ± 015 | 5.6 ± 0.07 | 5.8 ± 0.08 | 5.4 ± 0.15 |
| 6.0 | 11.7 ± 0.26 | 8.3 ± 0.17* | 11.9 ± 0.16 | 8.7 ± 0.29* |
| | PT / Control Ratio (:1)—Dog Plasma | | | |
| 0 | 1.0 ± 0.00 | 1.0 ± 0.00 | 1.0 ± 0.00 | 1.0 ± 0.00 |
| 0.4 | 1.4 ± 0.01 | 1.3 ± 0.01 | 1.5 ± 0.02 | 1.4 ± 0.02 |
| 2.0 | 2.2 ± 0.02 | 2.0 ± 0.02 | 2.3 ± 0.04 | 2.3 ± 0.04 |
| 8.0 | 4.1 ± 0.07 | 3.3 ± 0.05* | 4.3 ± 0.10 | 4.1 ± 0.08 |
| 15.0 | 5.8 ± 0.14 | 4.4 ± 0.08* | 6.5 ± 0.16* | 5.5 ± 0.11 |
| | PT / Control Ratio (:1)—Rat Plasma | | | |
| 0 | 1.0 ± 0.00 | 1.0 ± 0.00 | 1.0 ± 0.00 | 1.0 ± 0.00 |
| 1.0 | 1.5 ± 0.02 | 1.4 ± 0.02 | 1.6 ± 0.02 | 1.6 ± 0.01 |
| 4.0 | 2.2 ± 0.03 | 2.0 ± 0.04 | 3.0 ± 0.02* | 2.8 ± 0.03* |
| 12.0 | 3.4 ± 0.08 | 3.1 ± 0.07 | 4.9 ± 0.06* | 4.3 ± 0.05* |
| 24.0 | 5.5 ± 0.17 | 7.3 ± 0.24* | 15.3 ± 0.21* | 11.3 ± 0.28* |

[a] *, Mean value ± S.E.M. for 10 individual subjects significantly different from 0.98 ISI thromboplastin means at 5% level by *t*-test, separately by increasingly ISI value for individual rabbit thromboplastins.

assay. However, it was not affected to the same degree as was coumadin. Consequently, using the correction calculation designed for coumadin fluctuations to obtain an INR with CI-1031 grossly exaggerated the INR value. Although INR has been used clinically to monitor anticoagulant status during coumadin therapy, it probably should not be used with FXa inhibitor administration. Coumadin therapy typically produces INR values of 2, 4, and 6 as therapeutic, above therapeutic, and critical levels, respectively. INR values of 10 to 15 may be observed in acute coumadin poisoning, but INR values higher than 15 rarely occur [19]. Clearly, the magnitude of the INR obtained in this experiment (>120 in humans), combined with the incremental increase that occurred with increasing ISI value, shows that INR values in these FXa inhibitor-treated samples were an artifact of the calculation and not associated with the true anticoagulant effects of the FXa inhibitor itself. This suggests that when INR is used in clinical trials, it is important to select a thromboplastin with an ISI value close to 1.0. In this manner, the INR will

**TABLE 4 In Vitro Effect of an Experimental Factor Xa Inhibitor on International Normalization Ratio Using a Rabbit Brain Thomboplastin**

| Concentration of Factor Xa Inhibitor (μg/mL) | International Sensitivity Index[a] | | | |
|---|---|---|---|---|
| | 0.98 | 1.24 | 1.55 | 2.21 |
| International Normalized Ratio (:1)—Human Plasma | | | | |
| 0 | 1.0 ± 0.10 | 1.0 ± 0.02 | 1.0 ± 0.02 | 1.0 ± 0.03 |
| 0.2 | 1.6 ± 0.04 | 2.0 ± 0.06* | 2.6 ± 0.07* | 2.6 ± 0.16* |
| 0.6 | 2.7 ± 0.07 | 3.6 ± 0.10* | 5.7 ± 0.16* | 7.5 ± 0.57* |
| 1.2 | 4.0 ± 0.16 | 5.3 ± 0.19* | 10.5 ± 0.29* | 16.2 ± 1.31* |
| 1.8 | 5.2 ± 0.17 | 6.7 ± 0.21* | 15.3 ± 0.47* | 27.4 ± 2.618 |
| 6.0 | 11.1 ± 0.31 | 13.8 ± 0.49* | 46.0 ± 1.41* | 121.9 ± 11.54* |
| International Normalized Ratio (:1)—Dog Plasma | | | | |
| 0 | 1.0 ± 0.03 | 1.0 ± 0.01 | 0.9 ± 0.02 | 1.0 ± 0.03 |
| 0.4 | 1.4 ± 0.04 | 1.4 ± 0.01 | 1.7 ± 0.05* | 2.0 ± 0.08* |
| 2.0 | 2.2 ± 0.05 | 2.3 ± 0.04 | 3.8 ± 0.14* | 6.0 ± 0.28* |
| 8.0 | 4.0 ± 0.12 | 4.4 ± 0.10* | 10.1 ± 0.43* | 22.1 ± 1.19* |
| 15.0 | 5.6 ± 0.18 | 6.2 ± 0.16* | 16.7 ± 0.74* | 43.4 ± 2.31* |
| International Normalized Ratio (:1)—Rat Plasma | | | | |
| 0 | 1.0 ± 0.03 | 1.0 ± 0.00 | 1.0 ± 0.02 | 1.0 ± 0.01 |
| 1.0 | 1.5 ± 0.02 | 1.5 ± 0.02 | 2.6 ± 0.03* | 3.5 ± 0.07* |
| 4.0 | 2.2 ± 0.03 | 2.4 ± 0.06* | 5.6 ± 0.11* | 9.9 ± 0.31* |
| 12.0 | 3.3 ± 0.07 | 4.0 ± 0.13* | 11.7 ± 0.34* | 25.1 ± 0.82* |
| 24.0 | 4.6 ± 0.12 | 5.6 ± 0.22* | 17.8 ± 0.20* | 48.1 ± 2.27* |

[a] *, Mean value ± S.E.M. for 10 individual subjects significantly different from 0.98 ISI thromboplastin means at 5% level by *t*-test, separately by increasingly ISI value for individual rabbit thromboplastins.

closely approximate the PT/control ratio and give a true estimate of the anticoagulated state.

Table 5 indicates the maximum change in PT/control ratio and INR using thromboplastins with increasing ISI values (1.24 to 2.21). Changes in the PT/control ratio were modest at drug concentrations that produced increases of fourfold or less, the maximum targeted therapeutic PT value for clinical trials. The mean PT/control ratio in human samples increased maximally from 2.7 to 3.1 at twice the therapeutic dose (0.6 μg/mL). Absolute PT and PT ratios compared to baseline values were only modestly different using thromboplastin from various manufacturers, sources (human recombinant versus rabbit), and ISI. This finding indicates that absolute PT or PT/control ratio were more effective biomarkers of FXa inhibitor concentration than was INR.

**TABLE 5 Comparison of PT/Control Ratio and International Normalization Ratio in Plasma Samples Containing Increasing Concentrations of Factor Xa Inhibitor[a]**

| Species | Intended Drug Concentration (μg/mL) | PT/Control Ratio (0.98 ISI) | INR (0.98 ISI) | Maximum Change in PT/Control Ratio[b] | Maximum Change in INR[b] |
|---|---|---|---|---|---|
| Human | 0 | 1.00 | 0.99 | 0 | 0.02 |
| | 0.2 | 1.61 | 1.58 | 0.22 | 1.03 |
| | 0.6 | 2.70 | 2.65 | 0.38 | 4.88 |
| | 1.2 | 4.13 | 4.00 | 0.65 | 12.24 |
| | 1.8 | 5.37 | 5.19 | 0.97 | 22.21 |
| | 6.0 | 11.67 | 11.10 | 3.39 | 110.84 |
| Dog | 0 | 1.00 | 1.00 | 0 | 0.08 |
| | 0.4 | 1.42 | 1.41 | 0.10 | 0.61 |
| | 2.0 | 2.24 | 2.22 | 0.27 | 3.77 |
| | 8.0 | 4.11 | 4.00 | 0.81 | 18.06 |
| | 15.0 | 5.82 | 5.60 | 1.46 | 37.76 |
| Rat | 0 | 1.00 | 1.00 | 0 | 0.01 |
| | 1.0 | 1.45 | 1.45 | 0.38 | 2.08 |
| | 4.0 | 2.20 | 2.19 | 0.82 | 7.7 |
| | 12.0 | 3.42 | 3.34 | 1.48 | 21.74 |
| | 24.0 | 5.51 | 4.55 | 9.82 | 43.54 |

[a]Samples spiked with a factor Xa inhibitor in vitro.
[b]Results obtained by selecting the maximum result obtained with rabbit brain thromboplastin and subtracting from result obtained with human recombinant thromboplastin.

## PURSUING BIOMARKERS BEYOND PT, INR, AND aPTT

Values obtained with aPTT under ex vivo conditions were less sensitive than PT to FXa inhibitor-induced elevations and often underestimated drug concentration (data not shown). Beyond PT, INR, and aPTT, the most commonly used assay to evaluate FXa inhibitors is probably the anti-factor Xa assay (anti-FXa). It seems logical that a parameter named anti-Factor Xa assay should be the ideal biomarker for a FXa inhibitor. Additionally, this assay is used routinely in clinical settings to monitor the safety of heparin, a substance that also inhibits FXa production [20]. However, this assay is little more than a surrogate marker for drug concentration. A standard curve is prepared using the administered heparin (or other FXa inhibitor), and the chromagenic assay allows determination of the drug concentration in the plasma samples via production of FXa [21]. For heparin, the anti-FXa assay appears relevant. Years of use has allowed the development of a strong correlation between the number of international units of heparin determined via the assay and clinical safety. Reference ranges have been defined for the assay and provide a rapid

estimation of under, over, or therapeutic levels of heparin administration [22]. Still variability in the anti-FXa assay has been reported and is attributable to a number of factors, including instrumentation, assay technique, specificity of the commercially available kits, heparin preparations used in generating the standard curve, and approaches to data fitting [21]. In contrast, this experience does not exist for anti-Xa values obtained during FXa inhibitor administration. Just as the PT and INR may not be as beneficial for predicting FXa inhibitor effects as they are for coumadin, it should not be assumed that the anti-FXa assays have equivalent predictivity for heparin and other FXa inhibitors. For the Pfizer developmental FXa inhibitor, the anti-FXa assay offered little more than the PT as a monitor of drug concentration.

An additional assay called the factor X clotting (FX:C) assay was also evaluated. This assay is conducted using genetically engineered factor-deficient plasma spiked with serial dilutions of purified human factor X [23,24]. Concentrations of factor X in plasma are then determined by extrapolation from the standard curve. Since factor X must be converted to factor Xa for clot formation to occur, a functional clotting assay for factor X can also be used to assess the effects of factor Xa inhibitors. The FX:C assay provides several unique features that may make it a valuable biomarker for monitoring factor Xa inhibitor therapy: (1) the assay provides a rapid, reliable assessment of drug concentration and the percent inhibition of FXa achieved during drug inhibitor administration; (2) the assay can be performed on a high-throughput automated platform that is available in most hospital-based coagulation laboratories; and (3) individual factor X concentrations range from 60 to 150% between subjects [25]. This fairly high level of baseline intersubject variability suggests that a standard dose of drug may have a substantially different impact on total factor X inhibition. The FX:C assay defines baseline factor X activity and thereby allows continued dosing to achieve a targeted factor X concentration [4]. Literature is available concerning factor X concentrations and bleeding history in patients with either inherited or acquired factor X deficiency, so minimally there is some understanding that correlates the impact of reductions in FX:C evaluations and bleeding potential [26–28]. By determining the actual concentration of functional factor X remaining, physicians may have increased confidence in the administration of factor Xa inhibitors.

As with all the other coagulation biomarkers used for monitoring FXa inhibition, it was not immediately clear whether the FX:C assay was applicable in multiple species. Ex vivo experiments allowed this evaluation. To provide effective anticoagulant activity a 30% reduction in FX:C activity was predicted to be the minimal requirement for this compound. The FXa inhibitor concentrations in the ex vivo experiments were selected to bracket a range of factor X inhibition predicted to range from approximately 30% to 100%. Table 6 shows the intended concentrations of this FXa inhibitor in each species, the resulting FX:C activity, and the percent inhibition achieved. Assessment of these drug concentrations induced factor Xa inhibition of approximately 20% to >90%, showing that the targeted range could be predicted and achieved in

**TABLE 6 Factor X Activity and Percent Inhibition in Plasma Samples Containing Increasing Concentrations of a Factor Xa Inhibitor**[a]

| Species | Intended Drug Concentration (μg/mL) | FX:C Activity[b] (%) | Percent Inhibition[c] |
|---|---|---|---|
| Human | 0 | 106.1 ± 1.9 | NA |
| | 0.2 | 64.3 ± 1.6 | 39.4 |
| | 0.6 | 32.2 ± 1.0 | 69.7 |
| | 1.2 | 16.5 ± 0.7 | 84.4 |
| | 1.8 | 10.0 ± 0.5 | 90.6 |
| | 6.0 | 2.3 ± 0.2 | 97.8 |
| Dog | 0 | 143.0 ± 4.5 | NA |
| | 0.4 | 112.9 ± 8.6 | 21.0 |
| | 2.0 | 42.4 ± 4.3 | 70.3 |
| | 8.0 | 11.1 ± 1.4 | 92.2 |
| | 15.0 | 5.4 ± 0.8 | 96.2 |
| Rat | 0 | 84.8 ± 2.8 | NA |
| | 1.0 | 52.6 ± 1.8 | 38.0 |
| | 4.0 | 26.2 ± 0.9 | 69.1 |
| | 12.0 | 11.8 ± 0.6 | 86.0 |
| | 24.0 | 6.7 ± 0.3 | 92.1 |

[a]Samples spiked with a Factor Xa inhibitor in vitro.
[b]Mean ± SD of 10 samples/concentration.
[c]Calculated from species-specific control value. NA, not applicable.

all species. These ex vivo experiments demonstrated that the predicted efficacious dose of 0.2 to 0.3 μg/mL achieved the required 30 to 40% inhibition of FXa, providing early confidence in the dose selection process for phase I human trials. Additionally, these early ex vivo studies confirmed species-specific differences. The drug concentrations required to produce similar levels of FXa inhibition across species were markedly different. The FX:C assay was used effectively in preclinical rat and dog studies with this developmental FXa inhibitor. Knowledge of the species-specific concentration of drug required to induce the required 30% inhibition of FXa drove the selection of the low dose, whereas nearly complete inhibition of FXa drove the selection of the high dose. The FX:C assay helped determine the drug concentration required for complete inhibition of factor Xa in these species and the relative bleeding risk associated with a range of factor X concentrations. Prior knowledge of the impact of this drug on FXa inhibition through fairly simple clotting assessments helped eliminate undue risks of over-anticoagulation in preclinical studies, and there was no loss of animals due to excessive hemorrhage. It also addressed questions of whether dosing had been pushed to high enough levels when only minimal bleeding was observed at the highest dose. Since nearly 100% inhibition was achieved during the study, using higher doses was

not indicated and the lack of bleeding under conditions of complete FXa inhibition in rats and dogs suggested a strong safety profile. Inclusion of these biomarkers in preclinical studies provided greater confidence for selection of target stopping criteria for the first-in-human trial. The FX:C assay was translated and used as part of the first-in-human clinical trial with this compound. The FX:C assay provided data consistent with in vitro modeling, suggesting that it is predictive of drug concentration.

## CONCLUSIONS

One of the goals for new anticoagulant therapies is a superior safety profile compared to marketed anticoagulants, thereby minimizing or eliminating the need for clinical monitoring. Although clinical monitoring with standardized coagulation assays may appear to be a simple solution to monitoring the safety and efficacy of anticoagulants, there are inherent issues that make the elimination of clinical monitoring highly desirable. The obvious factors of cost and labor are minor in comparison to the problems associated with lack of patient compliance, delayed time to achieve therapeutic benefit, and the high degree of variability in the assay itself due to instrumentation, reagents, technique, and the inherent variability among subjects. Phase I studies using biomarkers are generally cheaper than phase II clinical endpoint studies. Additionally, new anticoagulants pose a relatively undetermined safety risk, due to the possibility of excessive bleeding. Therefore, biomarkers will continue to be essential until safety profiles can be established for this newer generation of anticoagulants. Although PT and INR are an effective reflection of drug the safety and efficacy for coumadins, they are less than ideal as biomarkers of new FXa inhibitor drugs. Assessing anti-FXa activity has been similar to drug concentration analysis for some inhibitors but variable for others [12; current case study]. FX:C clotting activity may be another alternative but remains largely unexplored. Regardless of the hope for the development of safer anticoagulants that are monitoring-free, the reality is that development of these drugs requires extensive patient monitoring to ensure safety. Compared to heparin and coumadin, which are monitored fairly effectively with aPTT and PT, respectively, development of the new FXa inhibitors are typically accompanied by a laundry list of probable biomarkers. This process is likely to continue until safety is firmly established through prolonged use and clinical experience with these agents.

It seems likely that most of these coagulation assays could just as likely be a bioassay of drug concentration as an indicator of pharmacologic response. In Stern's evaluation of biomarkers of an antithrombin agent, he concluded that "not all biomarkers are created equal" [29]. He suggested that "If a proposed biomarker measurement requires a drug and its molecular target to be combined in the same assay, it may be more a pharmacokinetic than a

pharmacodynamic assessment. Also, such assays should not be assumed to demonstrate an in vivo effect" [29]. As such, these biomarkers face a lofty hurdle to replace such pharmacodynamic endpoints as the Badimon chamber. So what does this mean for the early use of biomarkers in the development of new anticoagulants? It suggests that the greatest benefit for early utilization of coagulation biomarkers remains in allowing optimal selection of compounds, attrition of the right compounds, and the opportunity to provide an early ex vivo assessment against marketed competitors. It also demonstrates that efforts expended in understanding species-specific and reagent differences are critical in performing those early experiments.

## REFERENCES


1. CMR International (2006). *2006/7 Pharmaceutical R&D Factbook* CMR International, pp. 22–35.
2. Kola I, Landix J (2004). Can the pharmaceutical industry reduce attrition rates? *Nat Rev Drug Discov*, 3:711–715.
3. DiMasi JA (2001). Risks in new drug development: approval success rates for investigational drugs. *Clin Pharmacol Ther*, 69:297–307.
4. Jawad S, Oxley J, Yuen WC, Richens A (1986). The effect of lamotrigine, a novel anticonvulsant, on interictal spikes in patients with epilepsy. *Br J Clin Pharmacol*, 22:191–193.
5. Smith MB, Woods GL (2001). In vitro testing of antimicrobial agents. In Davey FR, Herman CJ, McPherson RA, Pincus MR, Threatte G, Woods GL (eds.), *Henry's Clinical Diagnosis and Management by Laboratory Methods*, 20th ed. W.B. Saunders, Philadelphia, pp. 1119–1143.
6. Anderson FA, Wheeler HB, Goldberg RJ, et al. (1991). A population-based perspective of the hospital incidence and case-fatality rates of deep vein thrombosis and pulmonary embolism. *Arch Intern Med*, 151:933–938.
7. Colman RW, Clowes AW, George JN, Hirsh J, Marder VJ (2001). Overview of hemostasis. In Colman RW, Hirsh J, Marder VJ, Clowes AW, George JN (eds.), *Hemostasis and Thrombosis: Basic Principles and Clinical Practice*. Lippincott Williams & Wilkins, Philadelphia, pp. 3–16.
8. Kakar P, Watson T, Gregory-Lip YH (2007). Drug evaluation: Rivaroxaban, an oral, direct inhibitor of activated factor X. *Curr Opin Invest Drugs*, 8(3):256–265.
9. Guertin KR, Choi YM (2007). The discovery of the factor Xa inhibitor Otamixaban: from lead identification to clinical development. *Curr Med Chem*, 14:2471–2481.
10. Zafar MU, Vorchheimer DA, Gaztanaga J, et al. (2007). Antithrombotic effect of factor Xa inhibition with DU-176b: phase-1 study of an oral, direct factor Xa inhibitor using an ex-vivo flow chamber. *Thromb Haemost*, 98:883–888.
11. Hylek EM (2007). Drug evaluation: DU-176b, an oral, direct factor Xa antagonist. *Curr Opin Invest Drugs*, 8(9):778–783.
12. Crowther MA, Ginsberg JS, Hirsh J (2001). Practical aspects of anticoagulant therapy. In Colman RW, Hirsh J, Marder VJ, Clowes AW, George JN, (eds.),

*Hemostasis and Thrombosis: Basic Principles and Clinical Practice*. Lippincott Williams & Wilkins, Philadelphia, pp. 1497–1516.

13. Zucker S, Cathey MH, Sox PJ, Hall EC (1970). Standardization of laboratory tests for controlling anticoagulant therapy. *Am J Clin Pathol*, 52:348–354.
14. Poller L (1987). Progress in standardization in anticoagulant control. *Hematol Rev*, 1:225–228.
15. Bailey EL, Harper TA, Pinterton PH (1971). The "therapeutic range" of the one-stage prothrombin time in the control of anticoagulant therapy: the effect of different thromboplastin preparations. *CMAJ*, 105:307–318.
16. Kirkwood TB (1983). Calibration of reference thromboplastin and standardization of the prothrombin time ratio. *Thromb Haemost*, 49:238–244.
17. Jeske W, Messmore HL, Fareed J (1998). Pharmacology of heparin and oral anticoagulants, In Loscalzo J, Schafer AI (eds.), *Thrombosis and Hemorrhage*, 2nd ed. Williams & Wilkins, Baltimore, pp. 257–283.
18. Crowther MA, Ginsberg JS, Hirsh J (2001). Practical aspects of anticoagulant therapy. In Colman RW, Hirsh J, Marder VJ, Clowes AW, George JN, (eds.), *Hemostasis and Thrombosis: Basic Principles and Clinical Practice*. Lippincott Williams & Wilkins, Philadephia, PA, pp. 1497–1516.
19. Levine MN, Hirsh J, Gent M (1994). A randominzed trial comparing activated thromboplastin time with heparin assay in patients with acute venous thromboembolism requiring large daily doses of heparin. *Arch Intern Med*, 154:49–56.
20. Kitchen S, Theaker J, Preston FE (2000). Monitoring unfractionated heparin therapy: relationship between eight anit-Xa assays and a protamine titration assay. *Blood Coagul Fibrinolysis*, 11:55–60.
21. Fifth ACCP Consensus Conference on Antithrombotic Therapy (1998). *Chest*, 119(Suppl):1S–769S.
22. Bauer KA, Kass BL, ten Cate H (1989). Detection of factor X activation in humans. *Blood*, 74:2007–2015.
23. Bauer KA, Weitz JI (2001). Laboratory markers of coagulation and fibrinolysis. In Colman RW, Hirsh J, Marder VJ, Clowes AW, George JN, (eds.), *Hemostasis and Thrombosis: Basic Principles and Clinical Practice*. Lippincott Williams & Wilkins, Philadelphia, pp. 1113–1129.
24. Fair DS, Edgington TS (1985). Heterogeneity of hereditary and acquired factor X deficiency by combined immunochemical and functional analyses. *Br J Haematol*, 59:235–242.
25. Herrmann FH, Auerswald G, Ruiz-Saez A, et al. (2006). Factor X deficiency: clinical manifestation of 102 subjects from Europe and Latin American with mutations in factor 10 gene. *Haemophilia*, 12:479–489.
26. Choufani EB, Sanchlorawala V, Ernst T, et al. (2001). Acquired factor X deficiency in patients with amyloid light-chain amyloidosis: incidence, bleeding manifestations, and response to high-dose chemotherapy. *Blood*, 97: 1885–1887.
27. Mumford AD, O'Donnell J, Gillmore JD, Manning RA, Hawkins PN, Laffan M (2000). Bleeding symptoms and coagulation abnormalities in 337 patients with AL-amyloidosis. *Br J Haematol*, 110:454–460.

28. Stern R, Chanoine F, Criswell K (2003). Are coagulation times biomarkers? Data from a phase I study of the oral thrombin inhibitor LB-30057 (CI-1028). *J Clin Pharmacol*, 43:118–121.
29. Stirling Y (1995). Warfarin-induced changes in procoagulant and anticoagulant proteins. *Blood Coagul Fibrinolysis*, 6:361–373.

# INTEGRATING MOLECULAR TESTING INTO CLINICAL APPLICATIONS

Anthony A. Killeen, M.D., Ph.D.
*University of Minnesota, Minneapolis, Minnesota*

## INTRODUCTION

The clinical laboratory plays a critical role in modern health care. It is commonly estimated that approximately 70% of all diagnoses are to some extent dependent on a laboratory finding. The clinical laboratory has various roles in the diagnosis and treatment of disease, including determining disease risks, screening for disease, establishing a diagnosis, monitoring of disease progression, and monitoring of response to therapy. Not surprisingly, the size of the market is large. A 2004 report by S.G. Cowen on the global in vitro diagnostics (IVD) industry estimated it to be $26 billion in size, and molecular diagnostics was identified as being among the more rapidly growing areas. Today, molecular testing is used in many areas of the clinical laboratory, including microbiology and virology, analysis of solid and hematologic tumors, inherited disorders, tissue typing, and identity testing (e.g., paternity testing and forensic testing). The growth has occurred rapidly over the last 20 years. In this chapter we examine the principal issues surrounding the integration of molecular testing into the clinical laboratory environment.

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

## CLINICAL LABORATORY REGULATION

The clinical laboratory environment in the United States is one of the most extensively regulated areas of medical practice and comes under the federal Clinical Laboratory Improvement Amendments (CLIA) of 1988 (http://www.cms.hhs.gov/clia/). Any implementation of molecular diagnostic is therefore governed by the provisions of the CLIA. The history of the CLIA dates back to the 1980s, when public and congressional concern was raised by reports of serious errors being made in clinical laboratories. In response to these concerns, legislation was introduced with the intention of improving laboratory testing. These regulations cover most aspects of laboratory practice. Any laboratory testing that is performed in the United States for clinical purposes such as diagnosis, monitoring, deciding appropriate treatment, and establishing prognosis must be performed in a CLIA-certified laboratory. These regulations, however, do not apply to purely research studies or to early research and development work for molecular or other testing in a non-CLIA-certified environment, but as soon as testing that has genuine clinical utility is made available, it must be performed in a certified environment.

The initial application for a CLIA certificate is usually made to the state office of the Centers for Medicare and Medicaid Services (CMS). A successful application will result in a certificate of registration, which allows a laboratory to perform clinical testing pending its first formal inspection. Depending on whether the laboratory is certified by CMS or by an accrediting organization, a successful inspection will result in a grant of either a certificate of compliance or a certificate of accreditation (Figure 1). These are essentially equivalent for

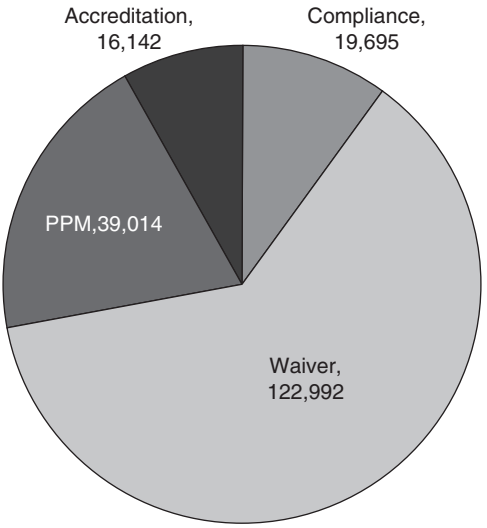


**Figure 1** Distribution of CLIA certificates by type in non-CLIA-exempt states in 2007. (Data from the CLIA database, http://www.cms.hhs.gov/CLIA/.)

the purposes of offering clinical testing. Accrediting organizations function as surrogates for CMS in the laboratory accreditation process and must be approved by CMS to accredit clinical laboratories. The accrediting organizations are the College of American Pathologists (CAP), the Council on Laboratory Accreditation (COLA), the Joint Commission, the American Association of Blood Banks (AABB), the American Society for Histocompatibility and Immunogenetics (ASHI), and the American Association of Bioanalysts (AAB). Some of these, such as the ASHI, accredit laboratories that perform only limited types of testing. Others, such as the CAP, accredit laboratories for all types of clinical testing, including molecular diagnostic testing.

Clinical tests are categorized for the purposes of the CLIA into several levels of complexity. This categorization is the function of the U.S. Food and Drug Administration (FDA). The type of CLIA certificate that a laboratory requires parallels the complexity of its test menu. The lowest level of test complexity is the waived category. Tests in this category are typically simple methods with little likelihood of error or of serious adverse consequences for patients if performed incorrectly. Commonly, such tests are performed in physician office laboratories. It should be noted that the term waived applies to a test, not to the need for the laboratory to have a CLIA certificate to perform any clinical testing. The next-highest level is the moderate-complexity test, including a category known as provider-performed microscopy. The highest level is the high-complexity test, which is applicable to most molecular tests. Laboratories that perform high-complexity testing must have a certificate to perform this type of testing. When the CLIA was written 20 years ago, there was relatively little molecular testing, and as a result, molecular diagnostics does not have specific requirements in the regulations as do most areas of clinical laboratory practice, such as clinical chemistry, microbiology, and hematology. Nevertheless, the general requirements of CLIA can be adapted to molecular testing. Accrediting organizations such as the CAP do have specific requirements for laboratories that perform molecular diagnostic testing. These are available in their laboratory inspection checklists [1]. Whereas the FDA is responsible for categorizing tests, the Centers for Medicare and Medicaid Services (CMS) are responsible for the oversight of the CLIA program, including granting certificates, approving accrediting organizations, approving proficiency testing (PT) programs, inspections, and enforcement actions.

The CLIA is a federal law and applies to all clinical testing performed in the United States and in foreign laboratories that are certified under the CLIA. There are provisions in the CLIA under which individual states can substitute their own laboratory oversight programs if it is determined that such programs are at least as stringent as the federal program. Currently, such programs exist only in New York and Washington. These are known as "CLIA-exempt" states, although CMS reserves the authority to inspect any aspect of laboratory performance in these states. The CLIA includes the

following areas of laboratory testing: proficiency testing, preanalytic testing, analytic testing, and personnel requirements.

## Proficiency Testing

Proficiency testing (PT) is one external measure by which a laboratory's performance can be judged. In a PT program, laboratories are sent samples for analysis and return their results to the PT program organizers. The correct result (or range of results) for these programs is determined by the organizers based on a comparison of participant results with results obtained by reference laboratories (accuracy-based grading), or by comparison with other laboratories that use the same analytical methods (peer-group grading). Ideally, all PT programs would use accuracy-based grading, but there are significant practical limitations to this approach. One of the major limitations is the PT material itself. For many analytes it is not possible to obtain the necessary range of concentrations to test low, normal, and high concentrations using real human samples. This necessitates the use of artificial samples that have been spiked with the analyte or from which the analyte has been removed (or at least its concentration has been lowered). Such artificial samples may behave unexpectedly when tested using some analytical equipment and give higher or lower values that would be obtained in a native specimen containing the same concentration of the analyte. This is known as the *matrix effect*. Other limitations may require peer-group grading; for example, recombinant proteins may not be detected equally in different manufacturers' immunoassays, making accuracy-based grading impossible. Enzyme concentrations may be determined by different manufacturers using different concentrations of cofactors, different temperatures, and different substrates, thus giving rise to such intermethod disagreement that accuracy-based grading is impossible.

Molecular testing poses certain challenges to PT programs. It may not be possible to obtain real human specimens such as blood from subjects known to carry mutations of interest because of the quantities required for a large PT program. This necessitates the use of cell lines or even DNA aliquots for PT programs in genetics. Such samples cannot test all phases of the analytical process, including extraction of DNA from whole blood (the normal procedure for genetic testing). The same concern applies to molecular testing for infectious diseases such as HIV-1. For these reasons, it is not uncommon that PT samples do not fully mimic patient samples.

Under the CLIA, laboratories are required to enroll in PT programs for a group of analytes specified in Subpart I of the regulations. These analytes were chosen based on clinical laboratory testing patterns that existed in 1988, and the list has not been updated since then. As a result, many newer tests, including molecular tests, are not on this list. For tests not on this list of "regulated" analytes, laboratories must verify the accuracy of their methods by some other method at least twice a year. This could include comparison of results with

those obtained by a different method: sample exchange with another laboratory, or even correlation of results with patients' clinical status. If formal PT programs exist, laboratories should consider enrolling in these. Several of the accrediting organizations do have requirements for participation in PT programs where these exist, including PT programs for molecular testing.

## Preanalytic Testing

The CLIA has requirements that cover the preanalytic phase of testing. These include the use of requisition forms with correct identification of the patient, the patient's age and gender, the test to be performed, the date and time of sample collection, the name of the ordering provider or the person to whom results should be reported, the type of specimen (e.g., blood), and any other additional information needed to produce a result. All of these are critical pieces of information that should be provided to the laboratory. Many so-called "laboratory errors" actually arise at the time of sample collection, and specimen misidentification is one of the most common types of error in the testing process.

In addition to the patient's age and gender, orders for molecular genetic testing should include relevant information about suspected diagnosis, clinical findings, and especially the family history. Many experienced clinical geneticists and genetic counselors will include a pedigree diagram on a requisition form for tests for inherited disorders. This practice is highly desirable and provides much useful information to the laboratory. As an example of the importance of this information, current practice guidelines in obstetrics and gynecology in the United States encourage the offering of prenatal testing to expectant Caucasian mothers to determine if they are carriers of mutations for cystic fibrosis. A recommended panel of mutations to be tested by clinical laboratories covers approximately 80 to 85% of all mutations in this population. In general, a negative screening test for these mutations reduces the risk of being a cystic fibrosis carrier from 1 in 30 to 1 in 141, and the laboratory would report these figures, or, if a mutation were identified, would report the specific mutation. However, these figures are based on the assumption that there is no family history of the disorder in the patient's family. If there is such a history, the risk of being a carrier (both before and after testing) is substantially higher. It is therefore essential that the ordering physician inform the laboratory if there is a family history.

## Analytic Testing

The CLIA has detailed requirements for the analytic phase of the testing process. These include the procedure manual, which is a step-by-step set of instructions on how the test should be performed, the process for method calibration, the procedures for preparation of reagents, the use of controls,

establishment of the reference range, reporting procedures, and analytical parameters such as sensitivity and specificity. There are no specific CLIA requirements that are unique to molecular testing, and therefore the molecular diagnostics laboratory has to adapt requirements from related areas such as clinical chemistry and microbiology to molecular testing. Some of the accrediting organizations have checklists that include specific requirements for molecular testing. These can provide useful guidance on procedures even for a laboratory that is not accredited by one of these organizations.

## Postanalytic Testing

Postanalytic testing refers to steps involved in reporting results to the ordering physician in a timely manner. The patient's name and identification information must be on the report, as should the name and address of the performing laboratory. In addition to the result, the report should include the reference interval and any relevant interpretive comments. The laboratory should be able to provide information on test validation and known interferences on the request of an ordering physician. Results must be released only to authorized persons.

Although certain elements of the postanalytic phase of testing can be controlled by the laboratory, there are also critical elements that are beyond its control, notably the correct interpretation of the result by the ordering physician. Molecular diagnostics (and genetics in general) is an area in which many physicians and other providers never had formal training in medical school. Concern has been expressed about the need to improve genetics education for health care professionals. Where there is a gap in provider knowledge, the laboratory should be able to offer expert consultation on the interpretation of its results to primary care providers [2]. This requires time, patience, and good communication skills on the part of the laboratory director and senior staff. Although such activity may be reimbursable under some health plans, the primary incentives for providing this kind of consultation are good patient care and customer satisfaction.

## Personnel Qualifications

Under the CLIA, requirements exist for laboratory personnel qualifications and/or experience. Perhaps the most important qualification requirements apply to the laboratory director. The director of a high-complexity laboratory such as a clinical molecular testing laboratory must hold a license in the state in which he or she works (if the state issues such licenses) and be a physician or osteopathic physician with board certification in pathology. Alternatively, the laboratory director can be a physician with at least one year of training in laboratory practice during residency, or a physician with at least two years of experience supervising or directing a clinical laboratory. A doctoral scientist holding a degree in a chemical, physical, biological, or clinical laboratory

science field with board certification from an approved board may also serve as the laboratory director. There are also provisions that allow for grandfathering of persons who were serving as laboratory directors at the time of implementation of the CLIA. Currently, there are no specific CLIA-required qualifications for the director of a molecular diagnostics laboratory. There are, however, board examinations in this field or similar fields that are offered by the American Board of Pathology, the American Board of Medical Genetics, the American Board for Clinical Chemistry, the American Board of Medical Microbiology, and the American Board of Bioanalysts. It is possible that individual states may begin to require specific qualifications in molecular diagnostics in the future or even that changes to the CLIA may require such qualifications. Other personnel and their qualifications described in the CLIA for high-complexity laboratories are technical supervisor, clinical consultant, general supervisor, cytology supervisor, cytotechnologist, and testing personnel.

## GENETIC TESTING AND PRIVACY

For many years there has been concern about the use of genetic information to discriminate against people with genetic diseases or those who are at risk of manifesting genetic disease at some time in the future. Although there are very few reported examples of such discrimination, the possibility of such misuse of genetic information by employers or insurance companies has received considerable attention by both the public and by legislative bodies [3]. A comprehensive analysis of applicable laws is beyond the scope of this chapter, but certain principles that apply to the clinical laboratory are worth mentioning. It is generally assumed, of course, that all clinical laboratory testing is performed with the consent of the patient. However, written consent is a legal requirement for genetic testing in some jurisdictions. The laboratory is generally not in a position to collect informed consent from patients, so it is usually obtained by some other health care worker, such as the ordering physician or genetics counselor. The laboratory director should be aware of applicable laws in this matter and determine, with legal advice if necessary, what testing is covered in his or her jurisdiction and ensure that appropriate consent is obtained. Genetic testing in its broadest meaning can cover more than just nucleic acid testing. For example, some laboratory methods for measuring glycohemoglobin, a test used for following diabetes control, can indicate the presence of genetic variants of hemoglobin, such as sickle-cell hemoglobin. Histopathologic examination of certain tumors can be strongly suggestive of an inherited disorder. Serum protein electrophoresis can reveal α-1 antitrypsin deficiency, an inherited disorder. The laboratory should consider how it reports such findings, which may contain genetic information that is unanticipated by both the ordering physician and the patient.

The most significant federal legislation in this area is the Genetic Information Nondiscrimination Act of 2008. This act offers protection against the use of genetic information as a basis for discrimination in employment and health insurance decisions. Under the provisions of this law, people who are healthy may not be discriminated against on the basis of any genetic predisposition to developing disease in the future. Health care insurers (but not life insurers or long-term care insurers) and employers may not require prospective clients or employees to undergo genetic testing or take any adverse action based on knowledge of a genetic trait. The benefits of this legislation are that some people may feel less trepidation about undergoing genetic testing because of fear that such information could be used by an employer or insurance company to discriminate against them.

## TESTING IN RESEARCH LABORATORIES

As research laboratories report new molecular findings in inherited and acquired diseases, it is not uncommon for clinical laboratories to receive requests to send patient samples to research laboratories for testing. This is an area in which the clinical laboratory must be careful to avoid noncompliance with CLIA regulations. One of the requirements of the CLIA is that certified laboratories must not send samples for patient testing to a non-CLIA-certified laboratory. This rule applies even if the research laboratory is the only one in the world to offer a particular test. Such samples should not be handled by a CLIA-certified laboratory, and the ordering physician should find some other means of arranging for testing if it is considered necessary. For example, it may be possible for testing to be performed under a research protocol. In this case the local institutional review board may be able to offer useful guidance on the creation and implementation of an appropriate protocol.

There are good reasons to be cautious about performing clinical testing in a research setting. The standards that one expects in a CLIA-certified laboratory are designed to promote quality and improve the accuracy of patient testing. Laboratories that do not follow these extensive requirements may not have all of the necessary protocols and procedures in place to offer the same quality of test result. Research laboratories are often part of academic institutions that may or may not carry malpractice insurance coverage in the event that a reported test result is erroneous.

## MOLECULAR TESTING FROM RESEARCH TO CLINICAL APPLICATION

The usual progression of molecular testing begins with gene and mutation discovery, typically in a research laboratory setting. Publication of these early

findings in peer-reviewed literature is the normal means of disseminating new information about a gene of clinical interest and the variations that can cause disease. It is important to document at least the most common disease-causing mutations and benign polymorphisms. It is usual to file patent applications to establish intellectual property claims, particularly if the test may have wide clinical applicability.

After a disease-causing mutation has been discovered, diagnostic testing on patients (as opposed to research subjects who have consented) requires performance in a laboratory that holds a CLIA certificate, and for molecular testing almost certainly means a certificate that allows for high-complexity testing. Because research laboratories are usually not set up to perform clinical testing or to meet the stringent criteria for clinical laboratory operations, it is usual that the rights to perform clinical testing be sold or licensed to a clinical laboratory that has the capability of offering such testing. The question of how many laboratories should be licensed is an important one. In general, it is often problematic for clinical providers when only one laboratory has a license to perform a clinical test. There is no way to verify a result in an independent laboratory, there is no competition that might lead to better test pricing for patients, there is little that can be done if the laboratory performance in areas such as turnaround time is suboptimal, and research may be inhibited [4]. For these reasons, licensing a test to multiple laboratories is generally preferable to an exclusive license to one laboratory. It has also been argued that patenting tests may inhibit their availability in clinical laboratories [5].

What should a molecular diagnostics laboratory be able to offer to meet clinical needs for molecular testing? First, the quality of the test result must be of a very high standard; that is, the results are reliable. Of course, all laboratories strive for this goal, which is implicit in the numerous regulations that govern laboratory testing. This is achieved by careful attention to the preanalytic, analytic, and postanalytic factors mentioned above and to the hiring of qualified and skilled personnel. The laboratory should offer turnaround times that are appropriate to the clinical needs of a specific test and will vary from one test to another. For example, testing for some infectious diseases is likely to require a faster turnaround time than is testing for a genetic predisposition to a chronic disease. Information should be readily available on the requirements for specimen type and the needs for special handling. The laboratory should be able to offer interpretations and consultations to ordering physicians regarding results of patient testing. If the genetic test result is a risk factor for future development of disease or for carrier status (e.g., cystic fibrosis carrier screening in pregnancy), the laboratory should be able to recalculate such risks if additional family history is provided at a later time. Many laboratories have a formal relationship with a genetic counselor who can interact with both patients and other health care workers and provide a variety of very useful services.

As clinical testing becomes more widespread, there can be significant changes to the knowledge and thinking about the relationship between disease

and underlying genetic mutation. An example of this is illustrated by the hereditary hemochromatosis gene, *HFE*. Discovery of this gene and the common mutations, *C282Y* and *H63D*, led to the view that the homozygous states, especially homozygosity for *C282Y*, would lead to chronic iron overload and hemochromatosis [6]. That view is no longer correct in light of more recent population studies of the penetrance of these mutations. Approximately one-third of patients who are homozygous for *C282Y* do not have elevated ferritin levels and appear not to be at risk of iron overload [7]. The reason for the variability of penetrance is probably related to dietary iron, blood loss, and other genetic factors that have yet to be determined. It is important for the laboratory director to be aware of such changing perspectives in thinking about diseases and to be an educator to others, making them aware of important developments, so that rational ordering patterns are encouraged.

## REIMBURSEMENT FOR MOLECULAR TESTING

In common with all areas of medical practice, reimbursement for molecular testing at the federal level (Medicare) is based on the current Common Procedural Terminology (CPT) coding system. State providers such as Medicaid and private insurance companies generally follow the same process. Under CPT coding, a charge and its payment are based on the number of individual items of service provided. Each step in a typical molecular assay ranging from extraction of DNA to performance of a polymerase chain reaction to gel electrophoresis and final result interpretation has a unique CPT code and an associated reimbursement based (in the case of Medicare) on the published fee schedule. Therefore, the Medicare reimbursement rate is calculable and is based on the individual steps in an assay. Private insurance companies may reimburse at a higher rate than federal payers. The CPT codes are updated annually by the American Medical Association, which retains copyright on the codes. Because of the rapid advances in molecular testing, it is not uncommon for laboratories to use methods that are not listed in the CPT guide. In this case, it may be necessary to seek consultation billing experts on choosing the appropriate fee codes.

Not uncommonly, genetic test prices from commercial laboratories are well above those that can be justified from published fee schedules. Although this may be perfectly legal, it can lead to significant problems for patients whose insurance companies (including Medicare) may not cover the full cost of the testing. In this situation the patient may have to pay out of pocket for part or all of the cost of the test if it is decided that the testing is essential. This situation can pose a financial risk for hospitals and clinics if they refer a sample for testing to a reference laboratory and thereby possibly incur the charges for a test. One possible option is to notify the patient and ordering physician that such tests are unlikely to be covered by insurance and determine how they propose to pay for testing. For Medicare patients, an advance beneficiary

notice (ABN) may be used to formally notify a patient that the test is considered to be a noncovered service [8]. These types of situations should be discussed with hospital management.

## SUMMARY

Molecular testing is firmly established in clinical laboratories for a wide variety of disorders. According to published reports, molecular diagnostics is and will continue to remain one of the fastest areas of growth in clinical testing. In the United States, the clinical laboratory operates under the regulations of the Clinical Laboratory Improvement Amendments of 1988, which provide the framework for producing high-quality results. The clinical laboratory differs significantly from the research laboratory both in practice and from a regulatory point of view. Careful attention should be paid to issues such as patient privacy and reimbursement for molecular testing.

## REFERENCES

1. College of American Pathologists (2008). http://www.cap.org (accessed Sept. 18, 2008).
2. Harvey EK, Fogel CE, Peyrot M, Christensen KD, Terry SF, McInerney JD (2007). Providers' knowledge of genetics: a survey of 5915 individuals and families with genetic conditions. *Genet Med*, 9:259–267.
3. Harmon A (2008). Insurance fears lead many to shun DNA tests. *New York Times*, Feb. 24.
4. Cho MK, Illangasekare S, Weaver MA, Leonard DG, Merz JF (2003). Effects of patents and licenses on the provision of clinical genetic testing services. *J Mol Diagn*, 5:3–8.
5. Merz JF, Kriss AG, Leonard DG, Cho MK (2002). Diagnostic testing fails the test. *Nature*, 415:577–579.
6. Feder JN, Gnirke A, Thomas W, et al. (1996). A novel MHC class I–like gene is mutated in patients with hereditary haemochromatosis. *Nat Genet*, 13(4):399–408.
7. Olynyk JK, Trinder D, Ramm GA, Britton RS, Bacon BR (2008). Hereditary hemochromatosis in the post-HFE era. *Hepatology*, 48:991–1001.
8. Carter D (2003). Obtaining advance beneficiary notices for Medicare physician providers. *J Med Pract Manage*, 19:10–18.

# BIOMARKERS FOR LYSOSOMAL STORAGE DISORDERS

Ari Zimran, M.D.
*Gaucher Clinic, Shaare Zedek Medical Center, Jerusalem, Israel*

Candida Fratazzi, M.D.
*Altus Pharmaceuticals, Inc., Waltham, Massachusetts*

Deborah Elstein, Ph.D.
*Gaucher Clinic, Shaare Zedek Medical Center, Jerusalem, Israel*

## INTRODUCTION

During the past few decades, because of the advances in molecular technology and improved understanding of lysosomal diseases, efforts have been made to identify appropriate prognostic and predictive factors in many of these diseases, despite the fact that each disease is a rare disorder. Indeed, what may be seen as a commonality among these diseases in terms of biochemistry or molecular underpinnings further defines the biochemical and molecular elements of each disease that make each disease unique. Thus, the situation today reflects the partial knowledge we have about this conglomerate of diseases, so that in the main, very few biomarkers are available in most of these diseases to assist the clinician to know which patients a priori will suffer from more severe manifestations or will benefit the most from specific therapies. In this context one must mention the infantile (neurological) forms that are rapidly

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

progressive, and for these it might not be ethically acceptable to offer therapies that have not yet withstood the test of time.

A biomarker should be technically feasible in many hands, easy to measure (using readily available technology, so that results are universal and standardized); useful, with a consistent relative magnitude between experimentals and controls, or treated and untreated; reliable, precise and accurate clinically, not just statistically; and classifiable as strongly predictive or prognostic. In recruiting patients with lysosomal disorders for clinical trials, the use of biomarkers is a double-edged sword: Whereas biomarkers may meet all the criteria above, they must also be clearly related to disease burden in the vast majority of patients and be capable of detection at both ends of the spectrum, from very mild to severe, and equally reactive to specific therapy within the same range. If all these prerequisites cannot be met, the use of the biomarker may be unjustified clinically. The purpose of this chapter is to review the literature and practice of biomarkers in lysosomal storage diseases and use current practices to discuss guidelines for the use of biomarkers in upcoming clinical trials.

## IDENTIFICATION OF SPECIFIC LYSOSOMAL STORAGE DISEASES

The rarer a disease, the more likely it is that biomarkers are unavailable. In terms of some of the lysosomal diseases, of which there are more than 50, there are actually no universally recognized biomarkers other than specific protein (enzyme) or substrate markers. Thus, for four mucoploysaccharidoses disorders (MPS I, MPS II, MPS III, and MPS VI), and Pompe disease, Gaucher disease, and Fabry disease there are protein markers, either enzymes or macrophage biomarkers; and for seven diseases (MPS I, MPS II, MPS IIIA, MPS IIIB, MPS IVA, MPS VII, and Fabry disease) there are substrate markers [1]. Urinary heparan sulfates can be used to differentiate MPS IIIC (Sanfilippo C syndrome) and MPS II (Hunter disease) [2], keratin sulfate to identify MPS IV and possibly other MPS disorders [3], and antibodies against gangliosides (i.e., anti-$G_{M2}$ and anti-$G_{M3}$) based on animal modeling. Antibodies, monoclonal and/or polyclonal, have been generated against some of the diseases' enzymes or substrates as well. Immunohistochemical techniques that can also be used to identify proteins [4], although not quantitative, may prove to have considerable potential as predictive markers if teamed with other techniques, such as mass spectrometry or various forms of chromatography.

In the first decade of 2000, before his untimely death, Nestor Chamoles produced filter papers for dried blood spots to identify various enzymes whose deficiencies implicated a lysosomal disorder: α-L-iduronidase (MPS I), α-galactosidase (Fabry disease), β-D-galactosidase ($G_{M1}$ gangliosidosis), and others [5–13]. Eventually, the technology was streamlined into a multiplex assay identification of enzymes for Fabry, Gaucher, Hurler, Krabbe, Niemann–

Pick A/B, and Pompe diseases [14] because of the recognition that simple and reliable diagnostic markers for patient identification were needed with the imminent and/or probable availability of specific therapies in these diseases. This methodological advance highlights the importance of using an easily obtainable patient sample (and in this case in small quantities) that can be transported without damage without special handling and which is highly reproducible. For the majority of the enzymes above, the filter paper system is reliable; however, in Pompe disease, for example, measured α-glucosidase activity may be a composite of other activities, making this specific assay less reliable. In its stead, and in similar cases where inhibition of nonspecific (substrate) activities cannot be totally suppressed, other assays, such as immunocapture, can be used [15]. There is also proof of principle for the use of diagnostic biomarkers from amniotic fluid in lysosomal disorders, with the express purpose of distinguishing normal from affected and even correlation with specific storage material in some of the disorders [16]. Within the past few years the list of diseases that can be profiled on the basis of various proteins, oligosaccharides, and glycolipids includes six MPS disorders and at least eight other diseases. Of note is a urinary measure of oligosacchirides (glycosaminoglycan derivatives) which has met the criteria of sensitivity and specificity in identifying persons with MPS disorders and, based on unique profiles, can differentiate among (all but MPS IIIB and MPS IIIC) subtypes [17].

Generally, to be useful as true biomarkers, some correlation with clinical disease expression (i.e., predictive or prognostic value) must be proven. A urinary diagnostic test that may be an appropriate marker of both disease progression and response to therapy is urinary globotriaosylceramide ($Gb_3$) in Fabry disease [18], although residual enzyme activity in the blood is a poorer marker of disease status. Similarly, in Gaucher disease, the most common lysosomal storage disorder, which has a range of clinical expression from virtually asymptomatic octogenarians to lethal neonatal forms, residual activity is a poor predictor of disease severity [19]. Therefore, in other diseases, a combination of analyses of enzyme activity with genotype or molecular phenotypes has been recommended [20] [e.g., in MPS II (Hunter disease)] to improve predictability. In summary, within the past few decades various specific assays have been developed that identify patients with enzyme deficiencies and can even quantify residual enzyme activity (relative to normal controls) based on the kinetics model of Conzelmann and Sandhoff [21] of a correlation between lipid accumulation and deficient enzyme activity [22], but residual activity is not always correlated with clinical status. Alternatively, improper or "derailed" processing of the enzyme may be indicative of disease severity since these enzymes undergo trafficking from endosome to Golgi to lysosome. This thinking has been applied in estimates of lysosomal-associated membrane proteins (LAMP-1 and LAMP-2) with the expectation of uncovering a processing defect common to all lysosomal disorders that would also be predictive of disease severity [23], but this was not proven [24]. It has also

been suggested that mutant enzyme variants may be retained in the endoplasmic reticulum and that this may be one of the factors that determine disease severity [25]. Accumulation of lipid in the endosomes or lysosomes was shown to characterize variants of type C Niemann–Pick disease because of the presence of cholesterol [26], thereby making this a good marker, but again, this was true only for a highly specific variant of this rare disorder. It should be noted that one ramification from these and similar findings is that response to therapy, even if the modality is identical for more than one lysosomal disease, may not be uniquely or sensitively monitored using non-disease-specific markers.

## IDENTIFICATION OF CLINICAL MARKERS

Clinical markers with predictive or prognostic value would be attractive if quantifiable and if assessment is non-invasive. It was hoped that the use of animal models would be illustrative of human conditions. In a most recent study of murine MPS I, although it was clearly shown that thickened aortic valves and abnormal cardiac function can be monitored from the preclinical stage, the authors note that "murine MPS I is not identical to human MPS I" [27]; each has unique clinical features. Nonetheless, a more recent study in murine MPS I employing proteomic analysis of heparin cofactor II-thrombin (HCII-T), which is a serine protease inhibitor, showed highly elevated serum levels in mice and in human patients that were correlated with disease severity and responsive to therapy [28]. HCII-T, therefore, may indeed meet the requirements of a good biomarker for MPS I, especially since it implicates a specific pathophysiology. A second good example in the MPSs is the use of accumulation of a disaccharide (HNS-UA), a marker of heparin sulfate storage in disease-specific sites of MPS IIIA [29] because the rate of accumulation is commensurate with disease severity at these sites and is appropriately reactive to disease-specific therapy.

Along these lines, therefore, it is commendable to find disease-specific parameters that lend themselves to quantification and test correlation with clinical severity and responsiveness to therapy. Biopsies and bone marrow aspirations or repeat radiological workups to stage severity should not be condoned if there is a better option: even, some might say, if that option does not exactly meet our criteria of a "good" biomarker. In Gaucher disease, because of concern about Gaucher-related skeletal involvement as associated with considerable morbidity, one study showed a reduction in osteoblast and osteoclast bone markers [30], but there was no correlation with incidence of bone pathology [31]. Biomarkers in this sense therefore might be misleading. Another example in Fabry disease showed no correlation between plasma concentrations of endothelial markers or homocysteine with response to therapy, although the endothelial and leucocyte activation are good measures of renal and cardiovascular involvement in Fabry disease [32].

## INFLAMMATORY MARKERS AS SECONDARY BIOMARKERS

Among the most prevalent hypothetical constructs used in lysosomal storage disorders is that of inflammation as a mediator, either as causative or as consequent effect, of lipid storage material. A pathway common to all MPS disorders (originally to describe MPS VI and MPS VII) has been developed based on inflammatory reactivity of connective tissues correlated with metalloproteinases in chondrocytes [33], but it is not a qualitative measure of severity or responsiveness to therapy. In $G_{M1}$ gangliosidosis because it is known that neuronal apoptosis and abnormalities in the central nervous system are secondary to storage, assessment of inflammatory cerebrospinal fluid markers showed correlation with clinical course but were not responsive to therapeutic interventions [34]. However, in a mouse model of gangliosidoses that showed disease progression with increased inflammatory cells in the microglia, the difference in $G_{M1}$ (Sandhoff disease) and $G_{M2}$ gangliosidosis (Tay–Sachs and late-onset Tay–Sachs disease) models was the timing of the onset of clinical signs [35], which is not always taken into consideration. Thus, while inflammation may be postulated to be either a primary or a secondary index of disease activity, not all markers meet the criteria of sensitivity or clinical relevance. Similarly, in an early study of Gaucher disease using macrophage-derived inflammatory markers, there were some cytokines that correlated with disease severity and clinical parameters, but the results were equivocal in many markers [36]. In a knock-out mouse model of types A and B of Niemann–Pick disease, the macrophage inflammatory cytokine MIP-1$\alpha$ was elevated in disease-specific sites and declined with therapy [37], but this marker cannot be disease-specific. On a global level, however, mouse models for the various lysosomal disorders have recently shown a connection between lipid storage in the endosome or lysosome and invariant natural killer T (iNKT)-cell function, indicative of thymic involvement, albeit these findings would conflict with the theory of elaboration of inflammatory markers in lysosomal storage disorders [38].

## MACROPHAGE SURROGATE BIOMARKERS

Of the many avenues attempted in the various common and less common lysosomal storage diseases, none is a completely satisfactory biomarker. This is a distinct disadvantage when the alternative may be invasive procedures that are more dangerous than merited by the status of the patient. A class of biomarkers has been incorporated into the evaluation initially of Gaucher disease, but now also of Fabry disease and type B Niemann–Pick disease, which are surrogate in the sense that they measure plasma levels of macrophage lipid or chemokines. Examples of this class are chitotriosidase and C-C chemokine ligand 18 (CCL 18; also called pulmonary and activation-regulated chemokine, PARC) that can be measured in plasma and in urine. Chitotriosidase

in Gaucher disease [39] was considered a specific marker of disease severity and then as a measure of response to therapy [40]. Among the methodological issues with using chitotriosidase is that it is genetically deficient in 6% of all persons, and genotyping should be done. The surrogate marker CCL18/PARC was then introduced [41] because it had the advantage of being present in everyone, yet it is not nearly as elevated in patients with Gaucher disease relative to healthy individuals. An advantage of CCL18/PARC over chitotriosidase assays is the less difficult assay of CCL18/PARC. In male patients with Fabry disease, chitotriosidase levels were found to be significantly elevated but were not correlated with disease severity, although in some cases may have normalized with therapy [42]. In two siblings with type B Niemann–Pick disease, there were elevated levels of both markers, but not commensurate with clinical severity [43]. Recently, urinary levels of chitotriosidase and CCL 18/PARC have been measured in Gaucher disease, but they do not appear to correlate with plasma levels, although there was correlation after exposure to treatment [44]. Interestingly, despite its indirect relationship to disease-specific parameters, the popularity of chitotriosidase as a putative biomarker has led to its use in testing other nonspecific markers [45]. This should not, of course, be the intention of biomarkers (i.e., that they correlate with each other) because then one is caught up in loops of correlation not one of which is related directly to a disease-specific parameter.

## BIOMARKERS AND CLINICAL TRIALS

Initiation of clinical trials is a costly and time-consuming commitment which has as a goal decreased time to market of a novel modality that will be a gold standard. This is definitely the case in rare diseases where the availability of a single therapeutic option that is safe and effective may be the only hope of affected individuals. If a pharmaceutical company undertakes the commitment to a clinical trial in an ultrarare disorder, the candidate modality must have tremendous promise to survive rigorous examination of the preclinical stages. By the time a putative therapy achieves phase II or phase III status, patients too will be highly motivated to see a successful treatment brought to market. Thus, on the one hand there is incentive for the company and for patients to get the treatment into the market, but on the other hand there is awareness that in clinical trials of patients, many hopeful candidate therapies do not meet their primary outcome measures, resulting in dismissal of that option.

Choosing outcome measures for clinical trials is both a science and an art in rare diseases because candidate patients are few, there is not always a "dream team" in terms of disease severity, and as we all know that "stuff happens." Things that happen that are unforeseen and uncontrollable may prevent a perfectly acceptable drug from getting to market. The current practice is to have secondary as well as primary outcome measures that one can

assess should the primary outcome be equivocal or difficult to interpret once the clinical trial is completed. As implied above, because they are seen as adjuncts in assessing clinical efficacy of therapy, biomarkers are popular as outcome measures in clinical trials, especially by regulatory agencies. In rare diseases, however, one must be cautious in applying biomarkers merely because they are more convenient to assess than disease-specific clinical parameters.

Importantly, all biomarkers are not equal in their predictive or prognostic value. This is a critical starting point in evaluating whether to include a biomarker as an outcome measure. Similarly, there is a difference, as implied above, between markers that measure disease-specific events and those (surrogate markers have a different use and can be confusing in this context) that measure events putatively related to a clinical event. In making clinical decisions, one should not rely on putatively related markers but be guided by clinically relevant parameters that correlate with disease severity.

In conclusion, biomarkers are a means of better prediction and follow-up, especially from the perspective of regulatory issues involving diagnostics and novel therapeutic options. This is even more cogent in cases where ancillary or additive therapies are considered to "fine tune" previously achieved therapeutic achievements. However, one should differentiate between diagnostic markers and prognostic biomarkers before choosing the latter over the former in making clinical decisions.

## REFERENCES


1. Parkinson-Lawrence E, Fuller M, Hopwood JJ, Meikle PJ, Brooks DA (2006). Immunochemistry of lysosomal storage disorders. *Clin Chem*, 52: 1660–1668.
2. Toma L, Dietrich CP, Nader HB (1996). Differences in the nonreducing ends of heparan sulfates excreted by patients with mucopolysaccharidoses revealed by bacterial heparitinases: a new tool for structural studies and differential diagnosis of Sanfilippo's and Hunter's syndromes. *Lab Invest*, 75:771–781.
3. Tomatsu S, Okamura K, Maeda H, et al. (2005). Keratan sulphate levels in mucopolysaccharidoses and mucolipidoses. *J Inherit Metab Dis*, 28:187–202.
4. Walkley SU (2004). Secondary accumulation of gangliosides in lysosomal storage disorders. *Semin Cell Dev Biol*, 15:433–444.
5. Chamoles NA, Blanco M, Gaggioli D (2001). Diagnosis of alpha-L-iduronidase deficiency in dried blood spots on filter paper: the possibility of newborn diagnosis. *Clin Chem*, 47:780–781.
6. Chamoles NA, Blanco M, Gaggioli D (2001). Fabry disease: enzymatic diagnosis in dried blood spots on filter paper. *Clin Chim Acta*, 308(1–2):195–196.
7. Chamoles NA, Blanco MB, Iorcansky S, Gaggioli D, Specola N, Casentini C (2001). Retrospective diagnosis of GM1 gangliosidosis by use of a newborn-screening card. *Clin Chem*, 47:2068.

8. Chamoles NA, Blanco MB, Gaggioli D, Casentini C (2001). Hurler-like phenotype: enzymatic diagnosis in dried blood spots on filter paper. *Clin Chem*, 47:2098–2102.
9. Chamoles NA, Blanco M, Gaggioli D, Casentini C (2002). Gaucher and Niemann–Pick diseases—enzymatic diagnosis in dried blood spots on filter paper: retrospective diagnoses in newborn-screening cards. *Clin Chim Acta*, 317: 191–197.
10. Chamoles NA, Blanco M, Gaggioli D, Casentini C (2002). Tay–Sachs and Sandhoff diseases: enzymatic diagnosis in dried blood spots on filter paper: retrospective diagnoses in newborn-screening cards. *Clin Chim Acta*, 318:133–137.
11. Chamoles NA, Niizawa G, Blanco M, Gaggioli D, Casentini C (2004). Glycogen storage disease type II: enzymatic screening in dried blood spots on filter paper. *Clin Chim Acta*, 347:97–102.
12. Wang D, Eadala B, Sadilek M, et al. (2005). Tandem mass spectrometric analysis of dried blood spots for screening of mucopolysaccharidosis I in newborns. *Clin Chem*, 51:898–900.
13. Niizawa G, Levin C, Aranda C, Blanco M, Chamoles NA (2005). Retrospective diagnosis of glycogen storage disease type II by use of a newborn-screening card. *Clin Chim Acta*, 359:205–206.
14. Gelb MH, Turecek F, Scott CR, Chamoles NA (2006). Direct multiplex assay of enzymes in dried blood spots by tandem mass spectrometry for the newborn screening of lysosomal storage disorders. *J Inherit Metab Dis*, 29:397–404.
15. Umapathysivam K, Hopwood JJ, Meikle PJ (2005). Correlation of acid alpha-glucosidase and glycogen content in skin fibroblasts with age of onset in Pompe disease. *Clin Chim Acta*, 361:191–198.
16. Ramsay SL, Maire I, Bindloss C, et al. (2004). Determination of oligosaccharides and glycolipids in amniotic fluid by electrospray ionisation tandem mass spectrometry: in utero indicators of lysosomal storage diseases. *Mol Genet Metab*, 83:231–238.
17. Fuller M, Rozaklis T, Ramsay SL, Hopwood JJ, Meikle PJ (2004). Disease-specific markers for the mucopolysaccharidoses. *Pediatr Res*, 56:733–738.
18. Whitfield PD, Calvin J, Hogg S, et al. (2005). Monitoring enzyme replacement therapy in Fabry disease: role of urine globotriaosylceramide. *J Inherit Metab Dis*, 28:21–33.
19. Fuller M, Lovejoy M, Hopwood JJ, Meikle PJ (2005). Immunoquantification of beta-glucosidase: diagnosis and prediction of severity in Gaucher disease. *Clin Chem*, 51:2200–2202.
20. Sukegawa-Hayasaka K, Kato Z, Nakamura H, et al. (2006). Effect of Hunter disease (mucopolysaccharidosis type II) mutations on molecular phenotypes of iduronate-2-sulfatase: enzymatic activity, protein processing and structural analysis. *J Inherit Metab Dis*, 29:755–761.
21. Conzelmann E, Sandhoff K (1991). Biochemical basis of late-onset neurolipidoses. *Dev Neurosci*, 13:197–204.
22. Schueler UH, Kolter T, Kaneski CR, Zirzow G, Sandhoff K, Brady RO (2004). Correlation between enzyme activity and substrate storage in a cell culture model system for Gaucher disease. *J Inherit Metab Dis*, 27:649–658.

23. Zimmer KP, le Coutre P, Aerts HM, et al. (1999). Intracellular transport of acid beta-glucosidase and lysosome-associated membrane proteins is affected in Gaucher's disease (G202R mutation). *J Pathol*, 188:407–414.
24. Meikle PJ, Ranieri E, Simonsen H, et al. (2004). Newborn screening for lysosomal storage disorders: clinical evaluation of a two-tier strategy. *Pediatrics*, 114: 909–916.
25. Ron I, Horowitz M (2005). ER retention and degradation as the molecular basis underlying Gaucher disease heterogeneity. *Hum Mol Genet*, 14:2387–2398.
26. Sun X, Marks DL, Park WD, et al. (2001). Niemann–Pick C variant detection by altered sphingolipid trafficking and correlation with mutations within a specific domain of NPC1. *Am J Hum Genet*, 68(6):1361–1372.
27. Braunlin E, Mackey-Bojack S, Panoskaltsis-Mortari A, et al. (2006). Cardiac functional and histopathologic findings in humans and mice with mucopolysaccharidosis type I: implications for assessment of therapeutic interventions in Hurler syndrome. *Pediatr Res*, 59:27–32.
28. Randall DR, Sinclair GB, Colobong KE, Hetty E, Clarke LA (2006). Heparin cofactor II-thrombin complex in MPS I: a biomarker of MPS disease. *Mol Genet Metab*, 88:235–243.
29. King B, Savas P, Fuller M, Hopwood J, Hemsley K (2006). Validation of a heparan sulfate–derived disaccharide as a marker of accumulation in murine mucopolysaccharidosis type IIIA. *Mol Genet Metab*, 87:107–112.
30. Drugan C, Jebeleanu G, Grigorescu-Sido P, Caillaud C, Craciun AM (2002). Biochemical markers of bone turnover as tools in the evaluation of skeletal involvement in patients with type 1 Gaucher disease. *Blood Cells Mol Dis*, 28:13–20.
31. Ciana G, Addobbati R, Tamaro G, et al. (2005). Gaucher disease and bone: laboratory and skeletal mineral density variations during a long period of enzyme replacement therapy. *J Inherit Metab Dis*, 28:723–732.
32. Demuth K, Germain DP (2002). Endothelial markers and homocysteine in patients with classic Fabry disease. *Acta Paediatr Suppl*, 91:57–61.
33. Simonaro CM, D'Angelo M, Haskins ME, Schuchman EH (2005). Joint and bone disease in mucopolysaccharidoses VI and VII: identification of new therapeutic targets and biomarkers using animal models. *Pediatr Res*, 57:701–707.
34. Satoh H, Yamato O, Asano T, et al. (2007). Cerebrospinal fluid biomarkers showing neurodegeneration in dogs with GM1 gangliosidosis: possible use for assessment of a therapeutic regimen. *Brain Res*, 1133:200–208.
35. Jeyakumar M, Thomas R, Elliot-Smith E, et al. (2003). Central nervous system inflammation is a hallmark of pathogenesis in mouse models of GM1 and GM2 gangliosidosis. *Brain*, 126:974–987.
36. Hollak CE, Evers L, Aerts JM, van Oers MH (1997). Elevated levels of M-CSF, sCD14 and IL8 in type 1 Gaucher disease. *Blood Cells Mol Dis*, 123:201–212.
37. Dhami R, Passini MA, Schuchman EH (2006). Identification of novel biomarkers for Niemann–Pick disease using gene expression analysis of acid sphingomyelinase knockout mice. *Mol Ther*, 13:556–564.
38. Gadola SD, Silk JD, Jeans A, et al. (2006). Impaired selection of invariant natural killer T cells in diverse mouse models of glycosphingolipid lysosomal storage diseases. *J Exp Med*, 203:2293–2303.

39. Hollak CE, van Weely S, van Oers MH, Aerts JM (1994). Marked elevation of plasma chitotriosidase activity: a novel hallmark of Gaucher disease. *J Clin Invest*, 93:1288–1292.
40. Czartoryska B, Tylki-Szymanska A, Gorska D (1998). Serum chitotriosidase activity in Gaucher patients on enzyme replacement therapy (ERT). *Clin Biochem*, 31:417–420.
41. Boot RG, Verhoek M, de Fost M, et al. (2004). Marked elevation of the chemokine CCL18/PARC in Gaucher disease: a novel surrogate marker for assessing therapeutic intervention. *Blood*, 103:33–39.
42. Vedder AC, Cox-Brinkman J, Hollak CE, et al. (2006). Plasma chitotriosidase in male Fabry patients: a marker for monitoring lipid-laden macrophages and their correction by enzyme replacement therapy. *Mol Genet Metab*, 89:239–244.
43. Brinkman J, Wijburg FA, Hollak CE, et al. (2005). Plasma chitotriosidase and CCL18: early biochemical surrogate markers in type B Niemann–Pick disease. *J Inherit Metab Dis*, 28:13–20.
44. Boot RG, Verhoek M, Langeveld M, et al. (2006). CCL18: a urinary marker of Gaucher cell burden in Gaucher patients. *J Inherit Metab Dis*, 29:564–571.
45. Moller HJ, de Fost M, Aerts H, Hollak C, Moestrup SK (2004). Plasma level of the macrophage-derived soluble CD163 is increased and positively correlates with severity in Gaucher's disease. *Eur J Haematol*, 72:135–139.

# VALUE CHAIN IN THE DEVELOPMENT OF BIOMARKERS FOR DISEASE TARGETS

Charles W. Richard, III, M.D., Ph.D.,
Arthur O. Tzianabos, Ph.D., and Whaijen Soo, M.D., Ph.D.
*Shire Human Genetic Therapies, Cambridge, Massachusetts*

## INTRODUCTION

Biomarkers have only recently come into use as an important tool in the development and clinical testing of therapeutic agents. The value of biomarker development was realized following the failure of drugs to achieve success in late-stage clinical trials in the 1990s. In many cases, small molecules and biologics being tested in clinical trials showed efficacy in preclinical testing as well as in phase I and phase II clinical trials, but failed to meet clinical endpoints once tested in expanded phase III trials.

Initially, pharmaceutical and biotechnology companies sought to improve clinical trial design through selection of more refined clinical endpoints. This was augmented by an effort to better understand the mechanism of action of the therapies being tested. These initiatives did improve the success rate of clinical trials in general and yielded an understanding in many cases of why therapies did not work in phase III trials. However, it became clear with time that these measures addressed only part of the overall shortcomings in the drug development process.

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

As more and more clinical trial results became available and were analyzed, the underlying problem with the process of drug development became evident. Clinical trials focused mainly on safety in phase I and early phase II trials. Efficacy data were only obtained in late-stage phase II trials and in expanded phase III trials. This approach squandered the opportunity to obtain meaningful efficacy data in phase I and early phase II testing. While the number of patients in phase I trials is usually smaller than in phase II trials, the opportunity to obtain data in humans was being lost.

This realization led to an effort to correlate changes in biological markers or biomarkers in humans with therapies being tested. The ability to monitor changes in the level of naturally occurring cytokines, cell surface molecules, signaling cascades, inflammatory mediators, or metabolic products that correlate with amelioration of disease following treatment gave investigators additional information about the effect of these therapies. This information could then be correlated with clinical outcomes to better understand much earlier in the drug development timeline whether a compound or biologic is effective.

## VALUE OF BIOMARKER DEVELOPMENT USING PRECLINICAL MODELS

The utility of biomarkers in clinical development is increased through the use of preclinical animal models of disease that recapitulate the major hallmarks of human disease being targeted for drug development. Therefore, biomarker development really does begin in the preclinical stage of drug development and relies heavily on selection and validation of good preclinical models of disease.

The selection of good preclinical models is an often overlooked component of drug development. This is the first step in building the value chain of a good biomarker(s) for use in clinical testing. Disease models should faithfully manifest the major aspects of human disease. This should be reflected in the development of the major and minor clinical signs that occur in humans but also in the pathogenesis of disease that leads to overt clinical signs. For example, the cytokines TNFα and IL-6 have been identified in preclinical animal models as major drivers in the pathogenesis of autoimmune diseases such as rheumatoid arthritis (RA; [1–3]). The finding that increased levels of these cytokines circulate in mice with RA and that these levels correlate with the severity of disease in these animals identified TNFα and IL-6 as potential biomarkers for human disease. Further, demonstration that that these levels decrease on immunosuppressive therapies confirmed their potential usefulness as a biomarker for the testing of novel drugs for RA in the clinic [1–3].

It is often the case that preclinical models for a given disease target do not exist or have been poorly developed. In this situation, efforts need to be directed toward understanding the underlying factors contributing to

disease pathogenesis through the development of animal models that recapitulate the hallmarks of human disease. These basic science exploratory studies can be difficult and time consuming. However, the value of these studies is realized when testing reveals that a therapy can modulate levels of the identified biomarkers *and* this correlates with a positive effect of the drug on disease endpoints in a clinically relevant animal model. Additional studies that investigate the effect of dose and regimen of a given therapy on modulation of these biomarkers are ultimately the key experiments that will inform on the dosing regimen to be used in human clinical trials. This information is critical, as it is a valuable component in the value chain of drug development.

It is now clear that the effort spent on biomarker development in preclinical animal models at the front end of the drug development process facilitates a more informed approach to clinical trial design in humans. This ultimately translates into a higher success rate for therapies tested in the clinic.

## STRATEGIES FOR DEVELOPING BIOMARKERS USING PRECLINICAL ANIMAL MODELS

There are several strategies that can be employed for the development of biomarkers using preclinical animal models. This often involves utilization of existing information about the pathogenesis of a given disease in relevant models. However, in most cases this requires basic research designed to understand what biomarkers correlate with the development of disease and if these biomarkers can be modulated by therapy in a meaningful way.

When considering animal models for biomarker development, it is important to distinguish between the types of models available. The best-case scenario is to utilize (or develop) a model that mimics the major and minor factors that drive the development of disease in humans. The identification of these factors as the cause of disease in these models allows for the ability to monitor and correlate them with disease severity. It then becomes important to determine if these factors respond to standard therapy known to reduce disease in humans. This is the best-case scenario moving forward if searching for new, improved therapeutic agents through head-to-head testing with standard therapies can be achieved.

However, it is most often the case that therapeutic agents are being developed for a disease for which there is no current treatment and/or there are no biomarkers of disease pathogenesis. In these situations, basic research is required for biomarker identification. The current approaches are varied and not uniform. Often, a good starting point for this initiative involves thorough research of the existing scientific literature to understand what serum or tissue factors are increased or decreased in patients who manifest disease and whether these factors are predictive of disease or disease severity. In addition, it is important to understand if these factors are a driving force in disease

pathogenesis. If this is the case, these factors could be good candidates as useful biomarkers to monitor the efficacy of potential therapeutic agents.

As cited above, TNFα and IL-6 are cytokines that are generally increased in sera taken from patients with RA [4]. These peptides are proinflammatory cytokines that appear early in the cascade of cytokines that ultimately lead to joint inflammation and the pathogenesis of this disease. The ability of drugs known to have a therapeutic effect on RA to decrease the levels of these cytokines in preclinical animal models demonstrated their value as biomarkers for the testing of new therapies. This highlights the usefulness of identifying serum biomarkers that play a central role in the pathogenesis of disease, as their levels typically increase in a manner that correlates with disease progression.

In cases where there is no literature to support biomarker identification, the development of genomic and proteomic techniques has created the opportunity for de novo biomarker discovery. If good animal models are available, genome-wide microarray analysis of cells or tissues obtained during the onset or progression of disease could lead to the identification of genes that are up- or down-regulated. These data need to be confirmed with additional proteomic studies to validate the potential role of the identified gene products, while additional structure–function studies in animal models need to be done to determine if these gene products correlate with disease progression and/or play a central role in disease pathogenesis. Finally, studies need to be performed to determine if these biomarkers respond to therapies known to affect the disease in animal models.

Once these basic research studies are performed in animal models, the identification of potential biomarkers needs to be validated in humans. This can be done in clinical trials in the patient target population using similar molecular techniques. The biomarkers selected need to be amenable to identification in easy-to-obtain clinical specimens such as serum or peripheral blood cells. However, once validated in humans, these biomarkers can serve as a very important tool in the value chain, leading to the testing and evaluation of novel therapeutic agents.

## VALUE OF BIOMARKER DEVELOPMENT AND USE IN CLINIC TRIALS

The sequencing of the human genome has presented a bewildering array of new targets for drug discovery that are only now being sorted out through more sophisticated systems biology approaches to biological pathway analysis. Most of these novel targets in drug discovery have not been proven pharmacologically in humans, so early readouts of potential clinical effectiveness in human clinical trials through biomarker analysis provides the much needed confidence that the drug affects the intended target in vivo. Moving beyond biological proof of principle, the ideal biomarker is one that can be measured

more easily, more frequently, and more accurately in humans and predicts early response to treatment or is an early indicator of clinical benefit. Since clinical efficacy is often apparent only after extended study in relatively large populations, identifying the most sensitive early biomarker of improvement in pathophysiology in smaller populations in shorter trials is an important research goal. Ideally, this search for the most sensitive and robust biomarker readout has been incorporated into earlier proof-of-efficacy animal studies, as it is almost too late for the investigative and experimental work to begin at activities immediately before filing the investigational new drug (IND) application and first-in-human clinical trials. Toward this end, most big pharmaceutical companies have adopted the incorporation of biomarker teams (either separate-line functions or matrixed teams) into the late discovery research process. Incorporation of potential biomarkers should be considered as part of all traditional phase IIa trials, but is considered increasingly as part of early, nonregistration, exploratory translational medicine studies.

Moving from proof of pathway perturbation to using the biomarker to establish the optimal dosing regimen is an important downstream consideration. Finely tuning dosing regimens to clinical outcome measures is rarely satisfactory, so assessing pharmacodynamic endpoints usually requires a sensitive and easily measured biomarker. Optimal biomarkers are those that can be sampled safely upon repeated measurements and are traditionally thought of as biochemical markers of bodily fluids, especially serum samples, but can include serial noninvasive imaging studies. In some instances, limited tissue biopsy material can be procured for histochemical and immunohistochemical analysis.

## USE OF BIOMARKERS FOR PERSONALIZED MEDICINE AND PATIENT STRATIFICATION

Biomarkers in clinical trials are also valuable if they can predict which subjects are most likely to respond. This is most obvious in the selection of patients for targeted chemotherapy in subgroups of cancers, but any biomarker that can potentially stratify patients into groups most likely to respond to treatment is potentially valuable. Examples from marketed products that use biomarkers for cancer responder selection based on gene expression include *Her2* (trastuzumab), *c-kit* (imatinib), epidermal growth factor receptor (EGFR; erlotinib, cetuximab), and Philadelphia chromosome (imatinib). Biomarkers are also useful if patient subpopulations at risk for toxicity can be identified for exclusion from trials of efficacy. Examples from marketed products include the screening of patients prior to therapy for glucose 6-phosphate dehydrogenase deficiency (G6PD; Dapsone), dihydropyrimidine dehydrogenase deficiency (DPD; fluorouracil) and ornithine transcarbamylase deficiency (OTC; valproic acid), since deficiency of these biomarkers leads to severe toxicity. The promise of pharmacogenetics and the discovery of genetic variants in

DNA that predicts efficacy or toxicity has gone largely unfulfilled, but some examples in marketed products do exist. Genetic variations in *N*-acetyltransferase (NAT; isoazanide), thiopurine methyltransferase (TPMT; 6-MP, azothioprine), UDP-glucuronyslytransferase-1 (UGT1A1; irinotecan), and several liver cytochrome P450–metabolizing enzymes [CYP2D6 (Strattera), CYP2C19, and CYP2C9 (warfarin)] have been proven to cause increased drug exposure, leading to toxicity, and are used for dosage adjustment [5].

Great strides have been made in recent years in automated microarray systems for surveying the entire genome for single-nucleotide polymorphisms and copy-number variation. Large numbers of patients are needed to uncover small genetic effects, so subtle differences in efficacy or rare idiosyncratic toxicology reactions will seldom be uncovered in limited phase I/II or other exploratory trials. That said, large efforts are under way by many large pharmaceutical companies to bank DNA from larger phase III and postmarketing trials to conduct genome-wide association studies with increasing sophisticated statistical genetic analysis that may identify single-nucleotide polymorphisms or copy-number variance that correlate with treatment response or toxicity.

## BIOMARKER DEVELOPMENT IN PARTNERSHIPS WITH REGULATORY AGENCIES

The most widespread use of biomarkers in current practice is for internal decision making, although the aspirational goal remains the development of surrogate biomarkers that can substitute for clinically validated endpoints for registration studies. Surrogate biomarkers are typically endpoints in therapeutic intervention trials, although surrogates are sometimes used in natural history or epidemiologic studies. To assist pharmaceutical companies in this effort, the U.S. Food and Drug Administration (FDA) has published a guidance document for pharmacogenomic data submissions that defines the concept of *valid biomarker, probably valid biomarker*, and *exploratory pharmacogenomic data* for regulatory decision making. The PG guidance document [6] defines *known biomarkers* as accepted by the scientific community at-large to predict clinical outcome, *valid biomarkers* as having predictive value but not yet replicated or widely accepted, and *exploratory biomarkers* as those found in exploratory hypothesis generation, often in the context of whole genome genomic, proteomic, or metabolomic analysis.

Perhaps the most effective use of *exploratory surrogate biomarkers* in clinical drug development occurs in the context of 21 CFR 314 and 601 Accelerated Approval Rule (1992), which allows for surrogate or nonultimate clinical registration endpoints. Fast-track designation is granted by the sponsor of the development program for a specific indication of a specific drug or biological program "to facilitate the development and expedite the review of new drugs that are intended to treat serious or life-threatening conditions and that dem-

onstrate the potential to address unmet medical conditions." Accelerated approval is often based on less well established surrogate endpoints or clinical endpoints. But postmarketing data are then required to verify and describe the drug's clinical benefit and to resolve remaining uncertainty as to the relation of the surrogate endpoint upon which approval was based to clinical benefit, or the observed clinical benefit to ultimate outcomes.

In summary, the continued development of biomarkers and true surrogate markers throughout the drug development value chain in the service of drug registration will remain an important activity to increase efficiency in the drug development process and to understand at the earliest time point in the pipeline whether a new chemical entity is likely to succeed or fail. Academic, pharmaceutical, and regulatory agency partnerships to develop biomarkers collaboratively within such initiatives as the FDA Critical Path Initiative [7] will help to "modernize the scientific process through which a potential human drug, biological product, or medical device is transformed from a discovery or proof of concept into a medical product" [8].

## REFERENCES


1. Miyata S, Ohkubo Y, Mutoh S (2005). A review of the action of tacrolimus (FK506) on experimental models of rheumatoid arthritis. *Inflamm Res*, 54(1):1–9.
2. Moller B, Villiger PM (2006). Inhibition of IL-1, IL-6, and TNF-alpha in immune-mediated inflammatory diseases. *Springer Semin Immunopathol*, 27(4):391–408.
3. Rose-John S, Waetzig GH, Scheller J, Grotzinger J, Seegert D (2007). The IL-6/sIL-6R complex as a novel target for therapeutic approaches. *Expert Opin Ther Targets*, 11(5):613–624.
4. Kremer JM, Davies JM, Rynes RI, et al. (1995). Every-other-week methotrexate in patients with rheumatoid arthritis: a double-blind, placebo-controlled prospective study. *Arthritis Rheum*, 38(5):601–607.
5. Freuh FF (2006). Qualifications of genomics biomarkers for regulatory decision making. Presented at the Annual DIA EuroMeeting, Paris, Mar. 7, 2006. http://www.fda.gov/Cder/genomics/presentations/DIA_Eur4.pdf.
6. FDA (2005). U.S. FDA Center for Drug Evaluation and Research Guidance for Industry: Pharmacogenomic Data Submissions. http://www/fda/gpv/cder/guidance/6400fnl.pdf (accessed Jan. 11, 2007).
7. FDA (2007). U.S. FDA's Critical Path Initiative. http://www.fda.gov/oc/initiatives/criticalpath (accessed Jan. 11, 2007).
8. Buckman S, Huang SM, Murphy S (2007). Medicinal product development and regulatory science for the 21st century: the Critical Path vision and its impact on health care. *Clin Pharmacol Ther*, 81(2):141–144.

# BIOMARKERS IN PHARMACEUTICAL DEVELOPMENT: THE ESSENTIAL ROLE OF PROJECT MANAGEMENT AND TEAMWORK

Lena King, Ph.D., DABT
*CanBioPharma Consulting, Inc., Guelph, Ontario, Canada*

Mallé Jurima-Romet, Ph.D.
*MDS Pharma Services, Montreal, Quebec, Canada*

Nita Ichhpurani, B.A., PMP
*MDS Pharma Services, Mississauga, Ontario, Canada*

## INTRODUCTION: PHARMACEUTICAL PROJECT TEAMS

The research-based pharmaceutical industry is one of the most complex industries in the world. Discovery and development teams constitute a well-established model to manage the complexity and integrated activities to guide projects in pharmaceutical development. Organizational models and composition of these teams vary between companies, depending on the size and business strategy of the company, but they are always multidisciplinary in nature. The discovery team is charged with discovering and developing new leads. This team may include scientists with expertise in disease models,

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

target identification, high-throughput screening, molecular biology, combinatorial chemistry, medicinal chemistry, and imaging. The development team is generally formed once a decision has been made to fund development of a new pharmaceutical lead for eventual registration. Development teams include preclinical disciplines (pharmacology, pharmacokinetics, and toxicology), pharmaceutical development (pilot and production chemists and/or biopharmaceutical expertise, formulation), regulatory affairs, clinical development, and commercial and marketing expertise.

The development team often has a formal project management structure with a project team leader and a project manager. In smaller organizations, a project manager may also serve as the team leader. Project management serves a critical role in supporting and driving forward the drug development process for the drug candidate chosen.

Particularly in recent years, the organizational structure of the discovery and development teams has been changing to adapt to internal and external demands, and the decreasing productivity and increasing costs associated with pharmaceutical development. To meet these challenges, capitalize on new technologies, and improve quality of decision making, companies are fostering collaborations between discovery and development scientists. The discovery teams increasingly include scientists with experience in DMPK (drug metabolism and pharmacokinetics), toxicology, clinical development, and project management to streamline or translate the research from discovery into development. Translational research is being proposed as the bridge between the perceived discovery–development silos and is emerging as a cross-functional discipline in its own right. As illustrated in Figure 1, some organizations have created an explicit biomarker or translational research unit that is represented on the development project team. Other organizations have adopted an implicit model in which biomarkers are part of the function of existing discovery and development units. A third option is a hybrid model that partners biomarker work in discovery and development without the creation and funding of a separate biomarker unit [1].

In addition to internal organizational restructuring, partnering between companies and outsourcing some parts of development (or in the case of virtual companies, all of development) to contract research organizations (CROs) is also becoming more common. These partnerships or alliances cover a wide spectrum of transactions and disciplines. Formalized alliance structures with contracts, governance, and team-specific guidance may not be in place for pharmaceutical development teams. However, even when a drug development team includes only partners from one company, it has been suggested that the project team is an implicit alliance and when including external partners may be an explicit alliance [2]. For small discovery startup companies, CROs may provide not only the conduct of studies necessary to help new candidates to progress through discovery and development but often also the essential development expertise and will act in implicit partnership with the sponsor. Thus, the concepts and processes developed for alliances, and their

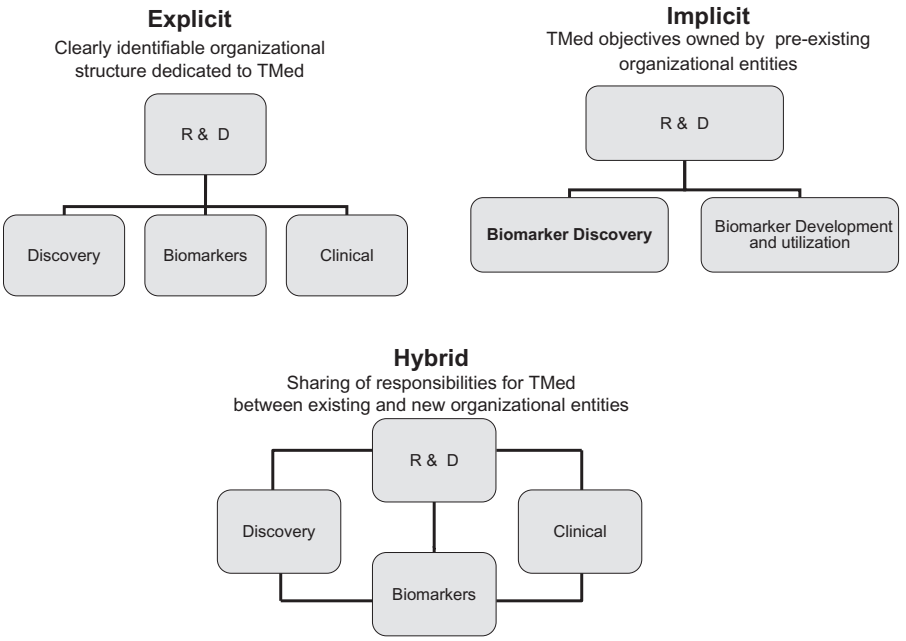


**Figure 1** Translational research (TMed) organizational models. (Adapted from ref. 1, with permission from *Drug Discovery World*.)

success stories, are instructive for drug development teams [3]. In research and development, an alliance provides a venue for access to complementary knowledge, new and different ideas, and processes with shared risk and reward. The following core principles pertain to alliances:

1. Goals and outcomes are shared with equitable sharing of risk and reward.
2. Participants have equal say in decisions, and each participant should have full management support.
3. Decision criteria must be based on what is best for the project rather than for individual participants.
4. The team operates in a culture of open and honest communication.

A recent survey of formalized research and development (R&D) alliances evaluated the contribution of alliance design (i.e., the number and type of partners, geographic proximity, R&D knowledge, and capabilities of each partner) and alliance management (governance agreements and processes) to the success of the alliance. The results showed that the alliance could generally

be designed with appropriate and complementary expertise. The number of partners and the presence of competitors among the partners had no overall effect on the success of the alliance. Effective contractual provisions and governance had a positive effect on the measures for alliance success. However, the most pronounced positive predictors of success were the frequency of communication and how ambitious a project was. The more ambitious projects were a strong predictor for success [4]. The success factors identified for other R&D alliances apply also to successful project teams involved in pharmaceutical development (Table 1).

Managing the project within budget and with appropriate resources is a major responsibility. For the pharmaceutical industry, the need for cost containment is providing compelling arguments for introducing high-value decision gates earlier in the development process. As illustrated in Table 2, biomarkers are one of the most important and tangible tools for facilitating translational research, moving data-driven decision making earlier into development, and for guiding development to the most appropriate indication and patient subpopulation. Although these additional decision gates can be helpful

**TABLE 1 Success Factors of a Drug Development Project Team**

| Predictors of Success of R&D Alliances | Successful Drug Development Team |
|---|---|
| Appropriate number, type of partner with complementary R&D knowledge and capabilities | "The right partners." |
| Effective contractual provisions and governance | "Good plan and good execution"—a well-understood and management-supported plan. |
| Excellent and transparent communication | Consultative team interactions<br>Team leader and project manager guide the team for decisions that are "on time, within scope and budget."<br>Develop trust between team members and the ability to work efficiently and effectively in a context of imperfect, incomplete, and unexpected information.<br>Solutions are sought without attribution of blame.<br>Innovative thinking and ideas are encouraged. |
| Ambitious projects | With development time lines spanning decades, these are inherently ambitious projects that require champions to obtain resources and management support. |

**TABLE 2 Biomarkers in the Pharmaceutical Development Cycle**

| Discovery/ Preclinical Stage | Phase I–IIa | Phase IIb–III | Phase IIIb–IV |
|---|---|---|---|
| Defining mechanism of action<br>Compound selection<br>PK/PD modeling<br>Candidate markers for clinical trials<br>Better prediction by animal models through translational research | Demonstrating clinical proof of concept<br>Dose and scheduling optimization<br>Optimization of patient population<br>Applications in new therapeutic indications | Minimize trial sizes through accurate inclusion and exclusion<br>Maximize success rates by early confirmation of efficacy<br>Potential for primary or secondary surrogate endpoints | Differentiation of products in marketplace through superior profiling of response<br>Differentiation in subpopulations (gender, race, genetics)<br>Personalized medicine (co-development of diagnostic) |

for the team, the inclusion of biomarkers adds complexity to the traditional linear model of drug development with a more reiterative process for the project team to manage.

## TEAM DYNAMICS: PHARMACEUTICAL PROJECT TEAMS

The development team has members with complementary technical skills. The management of the complex process of pharmaceutical development requires that these highly skilled knowledge workers engage, relate, and commit to a shared goal with defined milestones. These team interactions have to occur in a dynamic environment where (1) studies and experiments continually generate results that may fundamentally change the course of development, (2) management support and priority may be low compared to other projects, (3) team members may be geographically dispersed, and (4) resources for conduct of studies and other activities often are not controlled directly by the team. At its best, the pharmaceutical development team provides an environment that is mutually supportive, respectful, and enables discussion on controversial issues. It is open to new ideas, agile and constructive in addressing new issues, and has goals and strategy supported and understood by management. The project leader and the project manager should strive to generate an environment that is as conducive as possible to provide this ideal. These are not

features specific to biomarker development, but as mentioned below, including novel biomarkers will add to the complexity of the development project and require additional attention and management due to the increased number of communication channels. Following are some of the general principles of a productive team environment:

1. Include and plan for a project kickoff meeting and face-to-face meetings.
2. Define the roles and responsibilities of each team member.
3. Operate in a spirit of collaboration with a shared vision.
4. Practice active listening.
5. Practice transparent decision making; determine how decisions will be made and the role of each team member in the process.
6. Encourage all team members to engage in debate about strategic issues.
7. Spend time and energy to define objectives.
8. Engage and communicate actively with management.
9. Decide but revisit which communication tools are optimal.
10. Recognize and respect differences.
11. Plan for adversity.
12. Plan for the expected as well as the unexpected.

There are a number of excellent books that discuss team dynamics and team management [5–7]. Pharmaceutical scientific organizations are beginning to offer continuing education courses in program management, and dedicated pharmaceutical training courses are available [8]. However, effective drug development project leaders and managers do not come out of university programs or training centers. The understanding of how all the complex pieces of drug development come together can best be learned through hands-on experience as a team member, team leader, or project manager. Typically, it takes many years of working within the industry to gain sufficient knowledge of the drug development process to be an effective project team leader or manager.

## CONSEQUENCES OF BIOMARKERS IN PHARMACEUTICAL DEVELOPMENT STRATEGIES

Biomarkers are not new in pharmaceutical development. The interpretation, clinical significance, and normal variation of established biomarkers are generally well understood and widely accepted (discussed elsewhere in this book). Their utility, normal variation, and significance have been evaluated and corroborated in many different research and clinical studies. However, novel biomarkers that are now emerging may be available at only a single or a few vendors or laboratories. The assays may be technically and scientifically

complex, results dependent on platform, and limited data may be available on their normal variation and biological significance. Modern computational techniques allow for powerful multiplex analysis, binning of multiple parameters, and analysis of multiple biomarker on an individual animal or patient basis. These capabilities provide exciting opportunities for advancing the science; however, there are few published or marketed tools for choosing, planning, implementing, and evaluating the risk–cost benefit of biomarkers in pharmaceutical development. The risk–cost benefit for the biomarker may also be dependent on the size of the company, their portfolio, and the financing model. Large pharmaceutical companies' investment decisions for including novel biomarker strategy may be different from those of startup companies. A larger company may be able to offset the costs of biomarker development and implementation by applying the biomarkers to multiple projects and compounds. A startup company may include a biomarker throughout development despite uncertainty as to its ultimate utility; the company accepts the risk associated with new information emerging during the development process. By contrast, a larger company may require prior assessment of the value of including the biomarker in expediting development and improving development decisions. The integral role of biomarkers in decision making is discussed in Chapter Enabling Go/No Go Decisions of this book, but this aspect of biomarkers also has implications for project management and teamwork within a drug development team. Following are some of the consequences of employing novel biomarkers or a unique biomarker strategy in pharmaceutical development:

- High levels of investment in infrastructure
- Multiple technological platforms and specialized expertise
- High demands on data management
- Increased complexity in study designs, sample logistics, and study data interpretation
- Uncertainty and ambiguity for strategic decision making
  - Confidence and acceptance in translation and interpretation of results may be low
  - Lack of precedence for using biomarkers in novel regulatory alternatives such as exploratory IND
- Ethical issues: for example, tissue banking, privacy, and data integrity
- Evolving regulatory environment with changing requirements and expectations

## PROJECT MANAGEMENT

The following systematic tools and processes approach available for project management [9,10] can be applied to management of biomarker programs:

Gantt charts
Contracts, scope documents
Meeting minutes
Communication plans
RACI (responsible, accountable, consulted, informed) charts
Lessons-learned tools
Milestone charts
Decision trees
Risk analysis logs
PERT (performance, evaluation, review, tracking) charts
Work breakdown structures
Budget tracking
Lean-sigma tools

Process mapping with the team is useful to ensure that all aspects of the biomarker project management are well understood. The level of detail can range from GANTT charts (Figure 2) designed principally to track time lines to program-wide integrated biomarker strategies. An example of the latter for development of an oncology candidate is illustrated in Figure 3. The development strategy has to include open discussion about the advantages and disadvantages of including the biomarker, recognizing that this often has to occur in the absence of clear and straightforward knowledge of the value of the biomarker across species and in a specific disease or subset of patients. Increasingly, project teams are expected to analyze risk associated with different activities and develop contingency plans far in advance of the actual activity occurring. Risks associated with various biomarkers (e.g., timely assay qualification, sampling logistics, patient recruitment, regulatory acceptance) have to be part of this analysis and contingency plan development.

There are also numerous stakeholders beyond the development team who influence and may guide the development of the pharmaceutical:

- Sponsor (may include different departments with diverse goals/interests)
- Business analysts
- Investors (small companies) or shareholders
- Regulators
- CROs and/or biomarker labs
- Investigators
- Patients
- Patient support/interest groups

For the project management team, it is important to identify the stakeholders and evaluate their diverse and potentially competing priorities, particularly their perspectives on the benefit–risk effects of the new pharmaceutical. Novel targets with a drug producing an effect on a pharmacodynamic biomarker may be the last ray of hope for patients with serious or life-threatening diseases. Patients and sometimes physicians dealing with these diseases may have a different and often highly personal benefit–risk perspective compared to

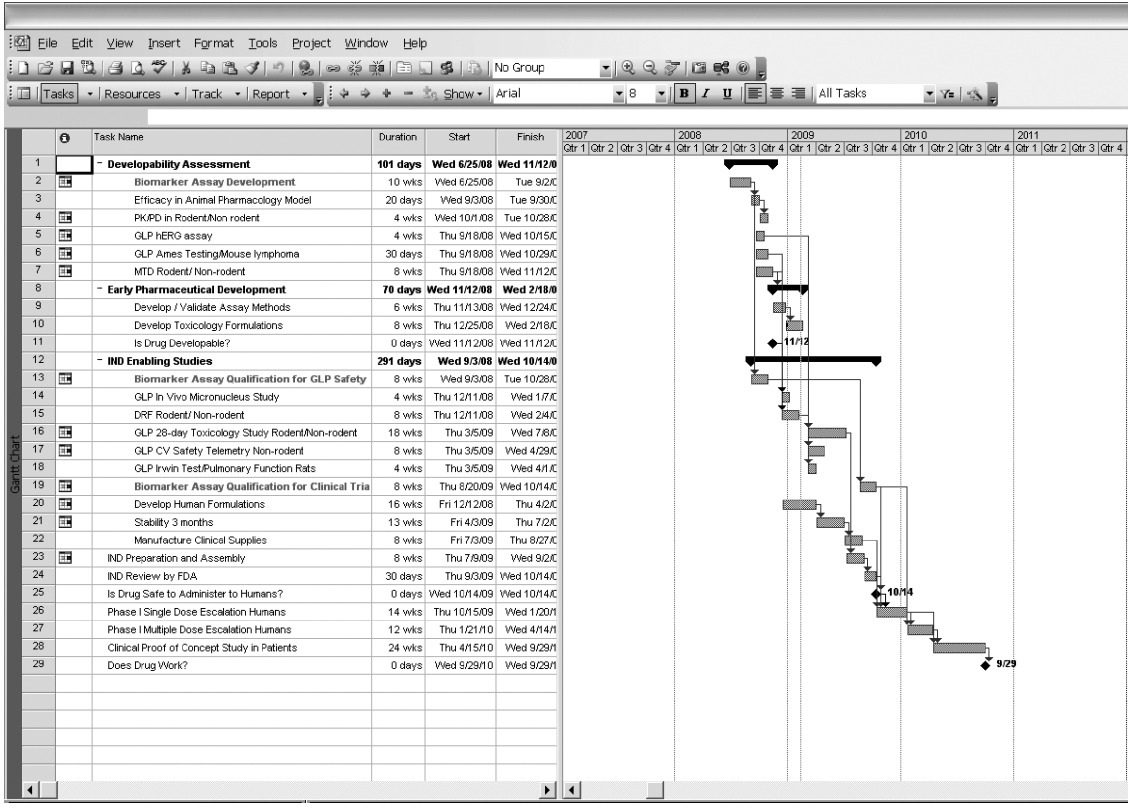


**Figure 2** Sample GANTT chart of a drug development plan incorporating biomarkers.

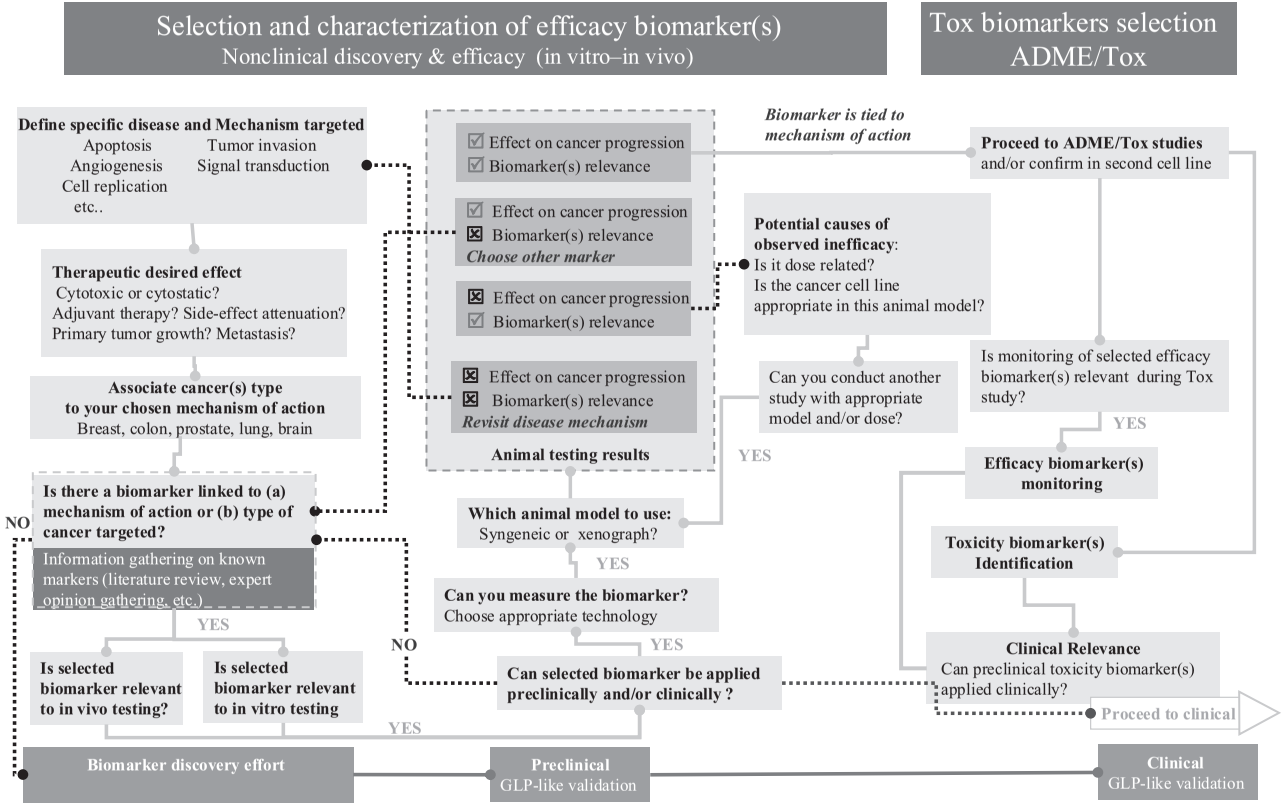


**Figure 3** Oncology biomarker map.

regulators, investors, and sponsors. Concerns about statistical significance, translation of the effect of the pharmacodynamic biomarker to clinical efficacy, and market share may carry little weight for patient advocacy groups under certain situations. Even concerns for safety biomarkers can be viewed as too restrictive: at best, perceived to delay access to potentially valuable medicines, and at worst, to stop their development.

Sponsors and investors, eager to see hints of efficacy as early as possible, can sometimes become overly confident about positive biomarker results before statistical analysis, normal variability, or relationships to other clinical efficacy markers are available. This may be more common in small emerging companies that rely on venture capital to finance their drug development programs than in larger established pharmaceutical companies. CROs or laboratories performing the assays and/or statistical analyses may be more cautious in their interpretations of biomarker data, sometimes seemingly unnecessarily so, but are motivated by the need to maintain quality standards as well as a neutral position. Consensus and communication problems are more likely to occur when these perspectives are widely disparate. Although it may be difficult at times, it is essential to achieve a common ground between stakeholders for effective communication.

## CHALLENGES ASSOCIATED WITH DIFFERENT TYPES OF BIOMARKERS

The definition and characteristic of *biomarkers* proposed by the National Institutes of Health working group: "a characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes, or pharmacological responses to a therapeutic intervention" [11], provide a categorization of biomarkers into efficacy, patient stratification, and safety biomarkers. There are different but overlapping challenges for the team, depending on the category of biomarkers.

### Efficacy Biomarkers

Efficacy biomarkers range from pharmacodynamic (PD) biomarkers, markers quantifying drug–target interaction, and markers reflecting the underlying pathology of the disease to those with established links to clinical outcomes accepted as surrogate endpoints for regulatory approval. Effect of a pharmaceutical in development on a biomarker associated with efficacy is guaranteed to generate enthusiasm and momentum in the team.

PD biomarkers have a long history in pharmaceutical development and form one of the cornerstones of hypothesis-driven approaches to drug discovery and development. These biomarkers are commonly generated as part of the discovery process. The biomarker may fulfill multiple key criteria in in vitro or animal models: (1) it may be used to characterize the pharmacology

models; (2) it may be used in knock-out or knock-in genetic models to further validate the target; and (3) it may demonstrate a characteristic PD/pharmacokinetic (PK) relationship with the drug under development. The changes reflecting underlying pathology range from largely unknown to those clearly indicative of potential market impact. An example of the latter is atrovastatin administration, resulting in decreases in serum triglycerides in normolipidemic subjects in clinical pharmacology studies [12,13]. This is more an exception than the norm. Typically, interpretation of the clinical significance and potential market impact of a biomarker is less certain, particularly if the pharmaceutical is (1) acting by a novel mechanism of action and (2) targeting a chronic progressive disease where disease modification rather than cure is the outcome anticipated.

The rationale for including PD biomarkers is generally easy to articulate to management, and particularly for smaller companies, these biomarkers may be essential for attracting investment. While enthusiasm and willingness to include these types of markers is generally not the issue, they are not without significant challenges in implementation and interpretation in the pharmaceutical development paradigm:

- Technical aspects
  - Stability of the biomarkers
  - Technical complexity of the assay
  - Assay robustness, sensitivity, specificity
  - Throughput of the assay
- Biological samples
  - Access to matrices that can or should be assayed
  - Sample collection volume or amount and timing in relation to dosing
  - Feasibility, cost, and resolution capabilities for imaging modalities for interactions with targets in the central nervous system, testis, poorly vascularized tumors, etc.
- Data interpretation
  - Normal values; inter- and intraindividual variability
  - Values in disease versus healthy conditions
  - Diurnal and environmental effects in animals
  - Effects of diet, lifestyle, concomitant medications in humans
  - Impact on development of no change or unexpected changes in biomarkers in the continuum from discovery to clinical

## Patient Stratification Biomarkers

The use of patient stratification biomarkers in pharmaceutical development and medical practice forms the foundation of what has been called *personal-*

*ized*, *individualized*, or *stratified* therapy. Patient stratification biomarkers focus on patients and/or underlying pathology rather than on the effect of the pharmaceutical on the target. For small-molecule drugs, genotyping for polymorphic drug-metabolizing enzymes responsible for elimination or activation/inactivation of the compound is now an established practice in clinical trials. Results about potential effects attributed to certain genotypes may be reflected in labeling recommendations for dose adjustments and/or precautions about drug–drug interactions [14]. A priori determination of genotype for polymorphic metabolizing enzymes are now included on the labels for irinotecan [15] and was recently added for warfarin [16] to guide selection of dosing regimen.

Targeted therapy in oncology is the best established application of patient stratification biomarkers. The development of Herceptin, the monoclonal antibody trastuzumab, with an indication restricted to breast tumors overexpressing HER2/neu protein [17], is a clinical and commercial success story for this approach. Oncology indications also include examples of the potential of using serum proteomics to classify patients according to the highest potential for clinical benefit. For example, Taguchi et al. [18] used matrix-assisted laser desorption ionization (MALDI) mass spectroscopy (MS) analysis to generate an eight-peak MALDI MS algorithm of unidentified proteins to aid in the pretreatment selection of appropriate subgroups of non-small cell lung carcinoma patients for treatment with epidermal growth factor receptor inhibitors (erlotinib or gefitinib).

As illustrated by the examples above, patient stratification biomarkers encompass a wide range of technologies, including algorithms of unknown proteins. Challenges for the development team are to understand and identify the potential for including patient stratification biomarkers either as part of or as the major thrust in the development process. This is often a major challenge, since the technologies may lie outside the core knowledge areas of the team members, making it difficult to articulate and discuss their value within the team and to communicate effectively to management. These challenges can be particularly pertinent for some of the "omics" technologies, which can be highly platform dependent and rely on complex statistical methodologies to analyze large sets of data to principal components. The results often have little intuitive inference in the underlying targeted disease pathology and may be one of the reasons that these powerful methodologies are not used more commonly. Some considerations for including patient stratification biomarkers are summarized as follows:

- Strategic issues
  - What is the purpose of including the patient stratification biomarker?
  - Will it be helpful in reaching go/no go decisions?
  - Is it required for registration purposes?
  - What will the implications be for marketing and prescribing practices?

- Practical considerations
  - Is the biomarker commercially available and accessible?
  - If a diagnostic biomarker is essential to the development of the pharmaceutical, should co-development be considered?
  - Are there IP and marketing restrictions?
  - What are the implications of the biomarker technology on the conduct of the clinical trial?

**Safety Biomarkers**

The considerations for safety during development are paramount; not surprisingly, it is one of the most regulated aspects of pharmaceutical development. Safety biomarkers have spurred on interesting and innovative regulatory and industry initiative and collaborations to develop and qualify novel biomarkers. Examples are the guidance of the U.S. Food and Drug Administration (FDA) for voluntary submissions of genomic data [19] and partnerships among government, academia, and industry for qualification of safety biomarkers [20]. Data qualifying the interpretation and significance of changes in safety biomarkers are needed to guide pharmaceutical development as well as evaluation of risk to patients or healthy volunteers in clinical trials. The purpose of safety biomarkers in clinical trials can be (1) to exclude patients at risk of developing adverse effects, (2) to increase sensitivity of the adverse event monitoring, and (3) to evaluate the clinical relevance of toxicity observed in the preclinical studies. Introducing novel or more uncommon biomarkers into a development project to address any of these aspects will not be embraced universally. There may be concerns not only about the added testing burden but also about the sensitivity and specificity of the biomarker, its relevance, and its relationship to well-established biomarkers. Nevertheless, including novel or uncommon biomarkers may be a condition for the conduct of a clinical trial as mandated by either regulatory bodies or institutional review boards. For example, there may be requirements to include sperm analysis in healthy volunteers and adapting experimental genotoxicity assays to humans to address effects observed in preclinical safety studies on the male reproductive tract and in genotoxicity evaluation, respectively. These will directly affect the conduct of the trials, the investigator, his or her comfort level with the assay, and the ability to communicate the significance of baseline values and any changes in the biomarkers to the clinical trial participant. However, novel and uncommon biomarkers will also have strategic and practical implications for the overall development program:

- Strategic issues
  - Will including the safety biomarker be a requirement for the entire pharmaceutical development program?
  - Are the efficacy and/or PK properties sufficiently promising to warrant continued development?
  - Can the identified safety concern be managed after approval?

- Practical considerations
  - What are the implications for the clinical trial program, locations of trials, linking to testing laboratories?
  - Will additional qualification of the biomarker assay be required as the development advances and for regulatory approval?

## MANAGEMENT OF LOGISTICS, PROCESSES, AND EXPECTATIONS

The logistical aspects of biomarker management are often a major undertaking, particularly if these include multisite global clinical trials (Figure 4). In clinical trials, specialized or esoteric assays, sometimes including very few samples, may require processes and systems that are not commonly in place for high-throughput analytes with standard operating procedures (SOPs) and established contract service providers. In addition, managing the logistics requires recognition and integration of the different expertise, experience, expectations, culture, and mindset in each discipline within the team. Regulations and guidelines governing the different disciplines as well as generally accepted practices have a major impact on the culture and mindset. Transitioning from the less regulated discovery process into development more accustomed to good laboratory practices (GLPs), good manufacturing practices (GMPs), and good clinical practices (GCPs) can be a major cultural shift. The regulatory requirements will vary depending on the purpose of the biomarker. Safety biomarkers will require a high degree of formalized regulatory compliance, whereas there are no requirements for GLP compliance for

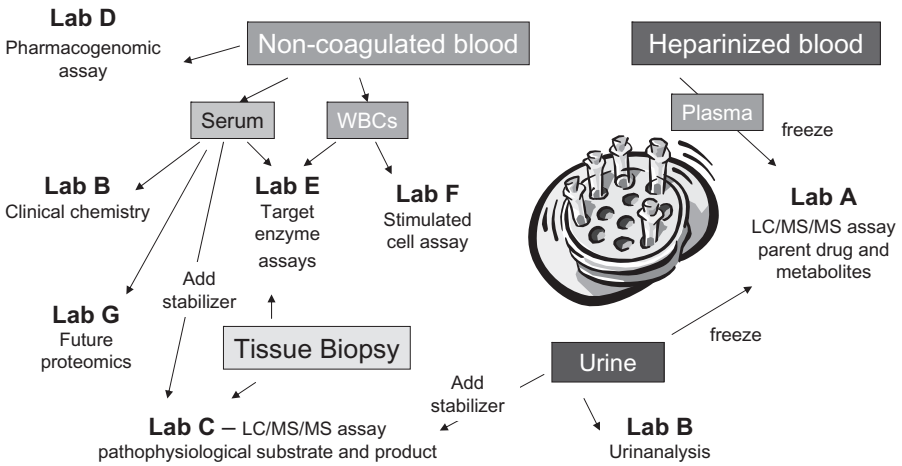


**Figure 4** Sample logistics.

pharmacodynamic biomarker assays. The question of whether or not to conduct the assay under GLP regulations will need to be considered for all types of biomarkers.

In the extensive project management coordination required to include particularly novel biomarkers in clinical trials, the long-term vision for program direction can become lost. The long-term view of the impact of the biomarker results on the pharmaceutical product under development as well as guidance for further discovery efforts should be considered. Questions about the impact of different outcomes have to be considered from both a strategic and a scientific perspective. For example, in which preclinical and clinical studies should the biomarker be included if normal values and variations are largely unknown? What will be the impact on future development of no change or unexpected effects in toxicology studies or in a first-in-human study? If there are changes to the assay or new technologies become available, should they be included to provide additional functional information about the target? Particularly when limited information is available about normal values and variation, adding additional parameters may not be of value for the decision-making process. It may be tempting to include a large number of biomarkers simply because they are available. Moreover, the increase in cost, complexity of the studies, and risk for erroneous results should be weighed carefully against the value added at each step of the development process.

The evaluation of whether to include a biomarker in a drug development program may not be straightforward. There is no doubt that biomarkers have proven valuable in pharmaceutical development to provide guidance for dose selection in early clinical studies, to enhance understanding of disease mechanisms and pathobiology, and to support decision making and strategic portfolio considerations. Success stories for the use of biomarkers in translating from discovery concept to clinical development have been published. One example is the first proteosome inhibitor bortezomib approved for treatment of multiple myeloma. The protesome is a key component in the ubiquitin–proteosome pathway involved in catabolism of protein and peptides as well as cellular signaling [21]. Ex vivo determination of proteosome inhibition was used in discovery and continued through toxicology and early- and late-stage clinical studies. Although not a clinical endpoint, proteosome inhibition provided valuable information that the drug interacted with the intended target [22].

However, in contrast to well-publicized success stories such as the example above, it is more difficult to obtain information and find examples of decisions taken when the PD biomarker did not yield the results expected. Why and where in the continuum of development did the PD marker fail, and what were the consequences of its failure? Strong confidence in the assay and the mode of action of the biomarker, as well as expectations about enhanced effects in patients compared to healthy volunteers, may result in progression of the biomarker despite lack of apparent interaction with the target. Decisions based on biomarkers require making a judgment call taking into account all

the data available. For the development team, the consequences of no effect or an unexpected effect of the drug on the PD marker should be considered and debated openly before the relevant studies are initiated.

Questions that should be discussed and understood by the team in the inclusion of biomarkers are as follows:

- Cost and logistics
  - What are the costs and logistics associated with including efficacy biomarkers?
  - What are the costs associated with not including a biomarker (i.e., progressing a compound without use of a biomarker or panel of biomarkers)?
- Confidence in the biomarker
  - Will the team/management accept go/no go decisions on the basis of the results of the efficacy biomarker?
  - How many patients are required to obtain meaningful results and/or to demonstrate response?
  - What degree of change or lack of progression of disease is considered acceptable?

These may appear to be relatively simple to answer, but it will take courage and conviction from the team to make decisions to discontinue development based, or at least partially based, on results of unproven efficacy biomarkers. There may be pressures from patient groups or specific patients for access or continuation of clinical development if the drug is perceived to be beneficial. Management may be reluctant to accept the decision if significant resources have been spent in development.

## SUMMARY

The successful launch of a novel pharmaceutical product represents the culmination of years of discovery and development work driven by knowledgeable people passionate about their project and the pharmaceutical. The development process will be challenging, require perseverance, and cannot be successful without coordination and teamwork. Novel biomarkers, organizational structures with multiple stakeholders, and a need to bring data-driven decision-making strategies earlier in development make the paradigm more complex and place higher demands on team communication and project coordination. Effective program leadership together with formalized program management and communication tools and processes facilitate this endeavor. As biomarkers in discovery and development are here to stay, more attention will be paid to best practices for project management and teamwork, as these roles are recognized increasingly to be essential for successful pharmaceutical development.

## REFERENCES


1. Hurko O (2006). Understanding the strategic importance of biomarkers for the discovery and early development phases. *Drug Discov World*, Spring, pp. 63–74.
2. Ahouse J, Fontana D (2007). Negotiating as a cross-functional project manager: lessons from alliance management. Cambridge Healthtech Institute: http://www.healthtech.com/wpapers/WP_pam.asp (accessed Oct. 4, 2007).
3. Bamford JD, Gomes-Casseres B, Robinson M (2002). *Mastering Alliance Strategy: A Comprehensive Guide to Design, Management, and Organization*. Jossey-Bass, New York.
4. Dyer JH, Powell BC, Sakakibara M, Wang AJ (2006). Determinants of success in R&D alliances. Advanced Technology Program NISTIR 7323. http://www.atp.nist.gov/eao/ir-7323/ir-7323.pdf (accessed Oct. 4, 2007).
5. Means JA, Adams T (2005). *Facilitating the Project Lifecycle: The Skills and Tools to Accelerate Progress for Project Managers, Facilitators, and Six Sigma Project Teams*. Wiley, Hoboken, NJ.
6. Parker GM (2002). *Cross-Functional Teams: Working with Allies, Enemies, and Other Strangers*. Wiley, Hoboken, NJ.
7. Wong Z (2007). *Human Factors in Project Management: Concepts, Tools, and Techniques for Inspiring Teamwork and Motivation*. Wiley, Hoboken, NJ.
8. Tufts Center for the Study of Drug Development. http://csdd.tufts.edu program.
9. Atkinson AJ, Daniels CE, Dedrick RL, Grudzinskas CV, Markey SP (2001). *Principles of Clinical Pharmacology*. Academic Press, San Diego, CA, pp. 351–364.
10. PMI Standards Committee (2004). *A Guide to the Project Management Body of Knowledge* (*PMBOK Guide*), 3rd ed. Project Management Institute, Inc., Newtown Square, PA.
11. Biomarkers Definitions Working Group (2001). Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. *Clin Pharmacol Ther*, 69:89–95.
12. Cilla DD, Gibson DM, Whitfield LR, Sedman AJ (1996). Pharmacodynamic effects and pharmacokinetics of atrovastatin after administration to normocholesterolemic subjects in the morning and evening. *J Clin Pharmacol*, 36:604–609.
13. Posvar EL, Radulovic LL, Cilla DD, Whitfield LR, Sedman AJ (1996). Tolerance and pharmacokinetics of a single-dose atrovastatin, a potent inhibitor of HMG-CoA reductase, in healthy subjects. *J Clin Pharmacol*, 36:728–731.
14. Huang S-M, Goodsaid F, Rahman A, Frueh F, Lesko LJ (2006). Application of pharmacogenomics in clinical pharmacology. *Toxicol Mechanisms Methods*, 16:89–99.
15. Pfizer Inc. (2006). Camptosar (irinotecan) label. http://www.pfizer.com/pfizer/download/uspi_camptosar.pdf (accessed Oct. 4, 2007).
16. Bristol-Myers Squibb Company (2007). Coumadin (warfarin) label. http://www.bms.com/cgi-bin/anybin.pl?sql=PI_SEQ=91 (accessed Oct. 4, 2007).
17. Genentech (2006). Herceptin (trastuzumab). http://www.gene.com/gene/products/information/oncology/herceptin/insert.jsp (accessed Oct. 4, 2007).

18. Taguchi F, Solomon B, Gregorc V, et al. (2007). Mass spectrometry to classify non–small-cell lung cancer patients for clinical outcome after treatment with epidermal growth factor receptor tyrosine kinase inhibitors: a multicohort cross-institutional study. *J Natl Cancer Inst*, 99:838–846.
19. Goodsaid F, Frueh F (2006). Process map for the validation of genomic biomarkers. *Pharmacogenomics*, 7:773–782.
20. Predictive Safety Testing Consortium. http://www.c-path.org.
21. Glickman MH, Cienchanover A (2002). The ubiquitin–proteasome proteolytic pathway: destruction for the sake of construction. *Physiol Rev*, 82:373–428.
22. EPAR (2004). Velcade (bortezomib). http://www.emea.europa.eu/humandocs/PDFs/EPAR/velcade/166104en6.pdf (accessed Oct. 4, 2007).

# INTEGRATING ACADEMIC LABORATORIES INTO PHARMACEUTICAL DEVELOPMENT

PETER A. WARD, M.D., AND KENT J. JOHNSON, M.D.
*The University of Michigan Medical School, Ann Arbor, Michigan*

## INTRODUCTION

Historically, there has been somewhat of an arm's-length relationship between researchers in academic medical centers and pharmaceutical companies. Traditionally, academic researchers often have taken an "ivory tower" approach: that their research is basic in nature and not meant necessarily to relate to the development of new drugs [1]. However, this attitude has undergone a major change over the past three decades. Academicians in California, Boston, and elsewhere have taken an entrepreneurial approach through the formation of biotech firms such as Genentech. However, it was not until fairly recently that most large pharmaceutical companies actively began developing collaborations with basic scientists in academia rather than relaying almost exclusively on their internal research and development programs. In the following discussions we look at reasons for this change, ways in which these collaborations can be fostered, and pitfalls associated with this association.

Bridging between academia and the pharmaceutical industry has increased steadily over the past 30 or more years, and this trend has accelerated

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

significantly with time. Movement in this direction has been based on fundamental scientific discoveries at the universities. These institutions generally lack the background, abilities, and resources to transform such discoveries into diagnostic and/or therapeutic modalities. Large pharma has the opposite problem: namely, more limited fundamental discovery research efforts, less emphasis on scientific research into basic disease processes, but extensive experience in drug design and development, clinical trial design and conduct, and monetary resources that ultimately result in a successful commercial product. In this chapter, we discuss the history and development of strong ties between academic scientists and the pharmaceutical industry. This will include discussions on the advantages of these collaborations and how they are usually structured. In addition, there are examples of these collaborations at both the basic science discovery stage and later in clinical development. This will include an example of scientists (Arul Chinnaiyan, George Wang, Dan Rhodes) who have been involved in fundamental discoveries regarding antigenic epitopes and autoantibodies in prostatic cancer patients, have taken these observations further with the development of spinoff companies, and have projected plans for the future.

## HISTORICAL PERSPECTIVE

Historically, academic researchers have taken a research approach that stressed National Institutes of Health (NIH)–sponsored basic research. These programs have targeted disease characterization or processes, but were not directed at a specific therapy for a given disease. In fact, many academic centers still differentiate between NIH research dollars and industry funds when ranking academic departments and consider the NIH funding to be the "gold standard" in determining the success and reputation of an investigator. However, several forces are changing this somewhat elitist perspective, as described below.

A major factor in this change has been the federal government change in funding levels. NIH dollars are much harder to come by, in particular the R01 primary research award. In fact, NIH funding for grants has diminished over the last few years, and several institutes in the NIH are now funding a much smaller percentage of research grants. Currently, only the top 8% of research applications submitted are funded (NIH Office of Extramural Research). Funding for the NIH has increased only slightly, with the fiscal year 2008 appropriation being increased only 0.9%, to $29.45 billion, compared to $29.3 billion in fiscal year 2007 (S. Forrest, *Annual Report on Research and Scholarship, FY 2007 Financial Summary*, University of Michigan). This amount has not kept up with inflation and has resulted in less funding for external investigators. For example, in fiscal year 2007 extramural funded R01 research grants decreased to 27,850 from 28,192 in 2006 (NIH Office of Extramural Research). Furthermore, the total amount of money given out in

2007 dropped to $10.04 billion from $10.12 billion in 2006. The overall impact of the reduction in NIH funding from 1995 to 2007 has resulted in a decline of real dollars available to researchers of approximately 10% (T. Mazzaschi, Research Funding Trends: Surviving an NIH Recession, presented at the 2007 AMSPC Annual Meeting). This cutback in funds has been devastating to many investigators and has even resulted in several laboratory closings. In addition, the reduction in federal funds has had a severe impact on the number of young scientists, particularly physicians, deciding on a research career.

There has also been a change in focus by the U.S. government and the NIH. Today, the NIH requires that grants have a "translational component" that provides a direct connection between the research hypothesis and a clinical disease parameter. To demonstrate this, a researcher studying a specific cytokine in an inflammatory process would have a specific aim in the grant for evidence that this cytokine was involved in a human disease. As would be expected, this has resulted in a major change in how basic investigators conduct their research projects. Additionally, it has mandated collaborations with clinicians and provided direct correlations with disease processes.

The reduction in available funding, and closer linkage to translating research findings into disease diagnosis or treatments, has lead to renewed interest in academic–pharmaceutical company collaborations. This shift has been encouraged further by the increasing need of biotechnology and small-molecule pharmaceutical companies to address more complex and chronic diseases, incorporate new technologies into their drug development programs, and enhance their molecular understanding of disease. Currently, most academics do not have a clear idea of how to solicit research support from such companies, and thus it is still relatively unusual for investigators to get significant industry funding for preclinical research. Many scientists within the pharmaceutical industry also have limited understanding of operations within universities and how to identify/access basic research programs that might be of benefit to drug development. The industry perspective will be described in greater detail below, as well as examples of research collaborations that have been successful.

## THE BIOTECH EXPERIENCE

Traditionally, biotech companies have largely been created by entrepreneurial academic scientists who want to commercialize their research findings for the development of drugs. Historically, these companies have been funded by venture capital funds or other sources rather than by the pharmaceutical industry. Such funding provides for the early-stage development of compounds but usually does not allow for the high expenses associated with full clinical development. Phase II clinical proof-of-concept and phase III pivotal safety and efficacy clinical trials involve significantly larger numbers of patients,

clinicians, institutions, and amounts of money than even well-funded startup companies have access to through normal funding sources. To further complicate the current model, large pharmaceutical companies are most interested in new molecules which have demonstrated efficacy and safety in humans (phase IIb or early phase III). Given a choice, the companies are substantially less interested in investing in early-stage preclinical development, due to the dramatically higher risk of attrition from toxicity or lack of effect against the targeted disease. Once a compound shows real promise in patients, a number of pharmaceutical companies will be interested in purchasing the rights to the compound or may invest in co-development of a drug.

This model has been quite successful in bringing new drugs to market that are not part of a pharmaceutical company's internal portfolio. However, historically, this has not allowed for the consistent support of preclinical research. Academic researchers are often placed in the unenviable situation of having a molecule with high potential for an unmet medical need but insufficient funds to develop the molecule to a stage where outside organizations see sufficient commercial potential to invest. The unfortunate outcome all too often is that the new approach either languishes or the academic institution assigns rights to the compound at a relatively low price.

## The Pharmaceutical Approach

Historically, pharmaceutical companies have relied primarily on their internal discovery scientists to develop most of their new drugs. They have also purchased the rights to compounds developed by biotech firms that show promise in the clinical trials. Thus, pharmaceutical companies previously have not been major supporters of external preclinical developmental, including those academic laboratories. Pharma often has supported targeted programs in academic medical centers, such as postdoctoral fellowships and occasional research laboratories, but this is not a widespread or reliable long-term funding mechanism for most universities. The nature and culture of academic and industrial groups can also inhibit successful interactions. The priorities, working processes, reward systems, and even ways of communicating differ significantly between the two institutional types.

Recently, however, there has been greater movement to support preclinical research programs in academic laboratories. Increased costs of drug development, high failure rates of new drug moieties, greater acceptance for outsourcing of many roles, and recognition that the next great drug discovery could come from many fields have increased the pressure on pharmaceutical companies to look beyond their own walls. This shift is further supported by mergers within the industry and the need for new drug pipelines larger than can reasonably be achieved based solely on internal resources. For a company to prosper, it is becoming essential to explore all available sources of new medical treatments. Also, competition for new drugs in late-stage development is intense and subject to great competition and cost. This last fact is

resulting in more attention to compounds in early clinical trials (phases I and IIa) or even late-stage preclinical evaluations.

Another factor supporting the movement to fund academic laboratories is that most companies now recognize that external agreements and collaborations are often more cost-effective than developing and supporting the same expertise internally. Examples of outreach by companies into academic institutions or private organizations would include companies such as Sandoz and Pfizer partnering with the Scripps Research Institute in La Jolla, California. In fact, today, many of the investigational laboratories in pharmaceutical companies have closed. There is a belief that outsourcing these studies provides expertise without the need to maintain large in-house programs.

## Advantages of Collaboration

There are a number of positive things associated with pharmaceutical company support for academic laboratories. For the company the academic laboratory provides cost-effective access to expertise and established research programs that do not have to be duplicated in-house. Since most critical observations in biology, including mechanisms of disease processes, are first elucidated in academic research laboratories, support of this process by companies provides access to this valuable basic research that cannot be duplicated in industry without tremendous internal investment. Such collaborations also allow pharmaceutical scientists access to in-residence stays in academic research laboratories. Furthermore, a pharmaceutical company investing in a faculty member's research has a deeper understanding of the benefits and limitations of the research than does the average person in industry who is looking for licensing opportunities. This often translates into earlier recognition of new therapeutic opportunities and the first right to decide on licensing a new product from the academic laboratory.

For academic researchers this funding provides an important additional source of research support in addition to the NIH. Industry funding also provides often for the purchase of expensive equipment which often would otherwise not be available to the investigator. Through academia–industry collaborations, many university scientists can gain access to specialized instruments or reagents that can be difficult or impossible to obtain within their institution. Additionally, taking an idea to actual therapy can be realized.

## Potential Pitfalls with the Collaborations

As a research investigator in a university setting, it is important to keep in mind that several conditions must be met for research collaborations with pharmaceutical companies to be successful. First, it is critical that the agreement be structured as research collaboration and not as a testing service. Although there are situations where short-term fee-for-service testing may be desirable, that is rarely the case for an individual researcher. Industry has

contracts with outside laboratories such as toxicology and medical reference laboratories where data ownership and intellectual property concerns are more efficiently defined and negotiated than within the university. This model generally does not work in academic research collaboration because in that situation the investigator is responsible for the analysis of the data and owns the data, usually with the intent of placing this information into the public domain through publications. This is a critical component of successful research collaboration since ideally the companies are interested in academic expertise in evaluating the data. The quality of research conducted will also be significantly richer when the interaction is a joint research collaboration. By structuring the agreement to allow both parties to benefit, advance the research, and look for options that neither could achieve alone, academic–industry programs are valuable tools to use to address unmet medical needs.

Hand in hand with control over the data is the issue of publishing and intellectual property. When employing a contract laboratory, the company owns all data and intellectual property pertaining thereto. This would not be acceptable to the institution in an academic environment. Most institutions require at least part of the intellectual property that comes from these collaborations, if not all of it. This is often a major sticking point between academic centers and pharmaceutical companies. Ideally, there is an agreement in place where the university has rights to patent data that come from the research collaboration, while the company has access to utilize these data in partnership with the institution.

Another major issue that goes hand in hand with intellectual property is the right to publish findings in peer-reviewed journals. In universities it is fundamental that the right to publish be part of any agreement made between research investigators and for-profit companies. For many companies this can be a difficult hurdle since they are concerned about keeping information confidential, at least until a patent is issued. Usually, what it done is to have a clause written into the contract defining when findings can be submitted for publication. This usually means a few months (but not much longer) before the academicians can submit their report for peer review.

## COLLABORATIONS BETWEEN PHARMACEUTICAL COMPANIES AND UNIVERSITIES IN PROMOTING BASIC RESEARCH

Traditionally, pharmaceutical companies have concentrated their external investments on compounds that have completed the early stages of the drug development process, and they usually already have some patient information. However, increasingly, the pharmaceutical industry is funding collaborations with universities for support for early-stage research. Historically, this role has been dominated by government and venture capital funding. The combined factors of cutbacks in NIH and venture capital funding, and risk aversion to early-stage funding by pharmaceutical companies, is leading to an emphasis

on supporting some early research for promising academic investigators and departments. This support of science at the "grassroots" level also provides companies with research capabilities that would not be possible in internal drug development laboratories and allows these companies to embrace the nimble style of biotechnology companies.

A recent example of such a collaboration is that of Pfizer forming a collaboration with the University of California at San Francisco (UCSF). This collaboration will not only fund narrowly defined research projects but will allow for early proof-of-concept funding for ideas proposed by the academic researchers (Bernadette Tansey, *San Francisco Chronicle*. June 10, 2008). Thus, this new model of collaboration between academic researchers and a major pharmaceutical company both funds specific projects and utilizes the academic scientists' expertise in developing new ideas and projects. The company can utilize the intellectual capital of these researchers, many of whom are acknowledged leaders in their particular field of study. From the perspective of the university, the individual scientists receive financial support and the ability to see accelerated progression of their discoveries toward medical implementation.

## The University of Michigan Experience with Pharma and Joint Research Projects

Like other research institutions, the University of Michigan has had long-standing collaborations with pharmaceutical and biotechnology companies. In general, these companies have provided support for lecture series, specialized symposia, specialty meetings, and consulting. In addition, there has been support for individual investigators. One example is that of the authors of this chapter and the Department of Pathology at the University of Michigan, who have had research collaborations with pharmaceutical companies: primarily Warner-Lambert and subsequently, Pfizer, as well as other companies to a lesser extent. These research collaborations have focused on areas that provided value to both institutions. One area of collaboration has been the development by the academic researchers of experimental disease models that can be utilized by pharmaceutical scientists early in the drug discovery process. Examples include models of lung, skin, and gastrointestinal injury. The use of experimental in vitro and in vivo models allows pharmaceutical scientists to evaluate new compounds of interest in models not readily available in their laboratories and to determine activity relative to a particular aspect of a disease or adverse effect. These collaborations have been used extensively to evaluate efficacy and toxicity in the discovery stage of drug development.

Another area of collaboration has involved studies comparing animal models with the human diseases that are targeted. The collaborations that these companies have with academic centers such as the University of Michigan allow comparison between the animal findings for molecules of interest and what is seen in the human disease. This is very valuable in determining how

relevant the preclinical animal models and toxicology studies are in predicting what will happen in humans. Examples include models of inflammatory vascular and lung injury, as well as sepsis.

Another area of collaboration has been in the identification of biomarkers of disease activity. Pathologists have unique training that bridges basic and clinical science dealing with mechanisms of disease, with specific interest in developing tests for clinical laboratories to diagnose specific diseases. In this regard our laboratories have been able to develop high-throughput technologies such as antibody arrays with support from these companies that would otherwise not be possible because of cost. This support allows the identification of new biomarkers of specific diseases, as well as toxicity. This information is useful not only to the pharma companies, but also to the medical community at large. Examples of these collaborative activities are cited below, where we show that this technology has great potential to identify new markers of disease in humans [2].

The fact that pathologists typically are responsible for running clinical laboratories has also been of value to the pharmaceutical companies since they may have on-site phase I clinics that evaluate human specimens. Members of our department have provided supervision of laboratory operations and support for College of American Pathology certification. Our clinical expertise has also been utilized in specific issues that arise in phase II and III testing.

Finally, and very important, our colleagues in the pharmaceutical industry have provided support for joint postdoctoral programs. This allows for the funding of postdoctoral scientists to work in the laboratories both on basic research studies at the university and on focused projects in a pharmaceutical setting. This has proven very successful, with several high-quality postdoctoral investigators trained in our laboratories going on to positions in industry and academia.

## EXAMPLES OF CORPORATE TIES BETWEEN THE UNIVERSITY OF MICHIGAN AND BIOTECH COMPANIES

Below are two examples of companies that are closely aligned with the research efforts of Arul Chinnaiyan at the University of Michigan Medical School. Ultimate commercialization of intellectual property by the two companies is closely linked to Chinnaiyan and his colleagues.

### Compendia Bioscience, Inc.

Compendia Bioscience was incorporated in January 2006. Its chief executive officer (CEO) is Dan Rhodes, who worked with Arul Chinnaiyan and obtained his Ph.D. degree through this association. Compendia Bioscience takes advantage of the research database Oncomine, which has been in place for several

years. This database takes published microarray information arising from more than 10,000 microarray studies of tumors and is made available at no charge to persons in academic settings. The information allows people to search the database and to extract information providing genomic sequences that can be used for diagnostic verification, predictions of drug sensitivity, and many other features, all of which ultimately bear on diagnostic analysis and clinical decision making in the oncology field. The Oncomine database is also available to the pharmaceutical industry on a charge basis, which involves the payment of license fees annually. Compendia was set up as a commercial entity with SPARK funding (a Michigan Economic Development Corporation entity in Ann Arbor, Michigan). In late 2007, Compendia received small business innovative research (SBIR) funding and funding from the 21st Century Job Funds (approximately $1.2 million). The company, which currently has approximately 12 employees, is seen as an important resource to the commercial world, with 14 of the top 20 pharmaceutical companies annually paying license fees to access the database. The fast-track SBIR granted to Compendia is allowing the company to increase its personnel to 17 employees, who are involved in software development and in evaluation of content and updating of the Oncomine databases and related matters. In 2008 Compendia is expected to break even financially. A decision has been made by Compendia not to become directly involved in drug discovery, personalized medicine, or other areas that depend on genomic information to make diagnoses or develop new drugs. The Compendia model is very different from that of most companies, in that in its licensing to pharmaceutical companies, no intellectual property is involved. In other words, if the pharmaceutical companies discover information in the Oncomine database that allows for development of a new drug, Compendia does not have any direct financial stake or gain. On the other hand, there is an increasing likelihood that Compendia will be involved in consultative activities with pharmaceutical companies to optimize use of its database.

### Armune BioScience, Inc.

Armune BioScience, located in Kalamazoo, Michigan, was formed in 2007 to develop and commercialize diagnostic tests for prostate, lung, and breast cancers. Its CEO is Eli Thomssen. Most of the initial work on which the company is based was done by the Chinnaiyan group at the University of Michigan Medical School. In this setting of fundamental proteomic prostate cancer research, several proteins expressed in low- or high-grade prostatic cancers and in metastatic lesions were identified and monoclonal antibodies developed in mice [3,4]. These resulting approaches are proving to be useful prognostic and diagnostic reagents. The well-known prostate-specific antigen (PSA) test has been widely used for the past decade and has a very high sensitivity (but very poor specificity), with as many as 40% of the test results being

interpreted in a manner that does not correlate with the clinical condition of the patient. This has led to great consternation for both patients and physicians as they try to select the optimal clinical intervention. With respect to lung cancer, no commercial tests are available that allow a serological diagnosis of the presence of lung cancer. The leading candidate for studies in this area is peripheral adenocarcinomas, which currently are detected by CT scans or chest x-rays, often with the diagnosis being made only at an advanced clinical state. Similarly, the accurate diagnosis of breast cancer is chiefly made based on excisional or needle biopsy, with no reliable serological test existing at this time.

The technology advanced by Armune Biosciences uses phage protein microarray methods employing beads coated with certain antigenic epitopes and the Luminex platform as the readout. In these assays, the coated beads interact with autoantibodies present in the blood of patients with prostate cancer. As reported by Wang et al. [5], the presence of these autoantibodies to prostate cancer cells represents an extremely specific humoral immune response and increasingly appears to be a reliable and early indicator of the presence of prostate cancer. It is hoped that the development of such diagnostic tests will result in high sensitivity (>90%) together with high specificity (>90%). It is expected that development of highly specific, highly sensitive, and reliable serological tests for the diagnosis of prostatic cancer will allow the diagnosis of prostatic cancer at an early stage so that a decision can be made much sooner with respect to whether surgery or radiation therapy should be employed. The technology should also reduce the number of unsuccessful or unnecessary needle biopsies of the prostate.

Developing a lung cancer diagnostic test would represent the first of its kind and might allow detection much earlier in high-risk groups such as smokers. Such success could result in a much better cure percentage than the current five-year survival rate of <15% with conventional diagnostic technology. Similarly, serological diagnostic tests for breast cancer have the potential for earlier diagnosis and treatment.

Armune Biosciences has received $900,000 of seed money, which is being used to establish a laboratory for validating the assay platform that is ultimately to be used for commercial laboratory applications. The likely test system will employ solid-phase (on beads) antigenic epitope peptides known to be reactive with autoantibodies present in the serum of patients with prostate cancer. In the future, a series A offering of $5.5 million and a B round of $5.7 million are planned. The series A funds will be used to develop a serological test that will identify aggressive prostatic cancer, followed by a similar approach to lung cancer. The series B funds will be used to complete development of the serological assay to be used in CLIA-approved laboratories, involving detection of both prostate and lung cancer. Series C funding is intended to support the conduct of prostate cancer diagnostic test market research studies to validate the product profile and to determine the optimal strategy for positioning of a commercial product.

Accordingly, Armune Bioscience currently has a well-developed plan and is being funded by venture capital funds. Although the scientific strategy appears to be sound, it will probably be several years before the development of an autoantibody-based test can be validated and ultimately applied successfully clinically.

## CONCLUSIONS

There are many areas in which academic research laboratories and pharmaceutical companies can interact and have mutually beneficial results. Industry support for investigational studies such as those described above for UCSF and the University of Michigan provide support for many research activities at academic laboratories and allow companies to access innovative research and clinical expertise that they may not have internally. We have also described two collaborations between the University of Michigan and biotechnology companies in the development of spinoff companies working to develop new ways to diagnose prostate, breast, and lung cancer based on detection of autoantibodies. The seminal discovery of autoantibodies as highly specific immune responses to tumor antigens was first described in 2005 by Wang et al. [3]. Such autoantibodies appear to be sensitive and reliable early detectors of the presence of a tumor provided by the immune system. Armune Biosciences is a relatively traditional young spinoff company supported by venture capital funding as well as by other sources, and headed ultimately for development of serological tests for the diagnosis of prostate, breast, and lung cancer. In contrast, Compendia Bioscience is a company that takes advantage of a proprietary database related to malignant tumors. This company has no commercial product development but has an enormous database that is constantly being upgraded and is available to the academic sector free of charge and on a licensing fee basis to the pharmaceutical industry. Access to a reliable and continuously updated genomic database is likely to provide pharmaceutical companies with new strategies for the development of antitumor drugs. Both companies were established in association with the University of Michigan, which holds equity in each.

## REFERENCES

1. Kozlowski RZ (1999). Industrial-academic collaboration: a bridge too far? (Editorial). *Drug Discov Today*, 4:487–499.
2. Olle E, Sreekumar A, Warner RL, et al. (2005). Development of an internally controlled antibody microarray. *Mol Cell Proteom*, 79:206–209.
3. Varambally S, Dhanasekaran SM, Zhou M, et al. (2002). The polycomb group protein EZH2 is involved in progression of prostate cancer. *Nature*, 419:624–629.

4. Tomlins SA, Laxman B, Dhanasekaran SM, et al. (2007). Distinct classes of chromosomal rearrangements create oncogenic ETS gene fusions in prostate cancer. *Nature*, 448:595–599.
5. Wang X, Yu J, Sreekumar A, et al. (2005). Autoantibody signatures in prostate cancer. *N Engl J Med*, 353:1224–1235.

# FUNDING BIOMARKER RESEARCH AND DEVELOPMENT THROUGH THE SMALL BUSINESS INNOVATIVE RESEARCH PROGRAM

James Varani, Ph.D.
*The University of Michigan Medical School, Ann Arbor, Michigan*

## INTRODUCTION

When the Small Business Innovative Research (SBIR) program was established, it was generally thought to be a less competitive road to government funding than were the traditional grant programs administered by the National Institutes of Health (NIH), the National Science Foundation (NSF), and others. There were many fewer applications relative to the pool of dollars available, and the quality of the grants was judged to be, on average, significantly lower than applications in traditional venues such as the investigator-initiated (R01) tract. The quality issue, in particular, reflected the lack of experience that most people in the private sector had with the government granting process. This no longer is the case. Obtaining funds through the SBIR program is every bit as challenging as being successful through the traditional mechanisms. Thus, there is no single formula that can guarantee success and no single approach that is necessarily better than the others. Any case study provides only a guideline. Still, one can increase the chances of being success-

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

ful by understanding the process and by carefully studying past successes as well as past failures. With that in mind, we discuss some of the key features of the SBIR program, things that may increase the likelihood of success in funding a biomarker-development project through the SBIR program.

## THE SBIR PROGRAM

Congress passed the Small Business Innovation Development Act in 1982 with a number of aims in mind. First (and foremost), the program was designed to stimulate technological innovation by funding projects initiated by high-technology companies to solve pressing research and development needs. Another aim was to provide a mechanism for encouraging participation by minority and disadvantaged persons in technological innovation. Finally, a stated goal was to increase commercialization of innovations derived from federal research and development dollars. Doing these things was thought to be necessary to increase the competitiveness of U.S. firms relative to their foreign counterparts. Largely unstated was the worry that the U.S. global lead in technology was being eroded.

Under the act, federal agencies with substantial extramural research and development budgets were required to direct a portion of their extramural funding to private enterprises that fit the federal definition of a small business. While the percentage of the total research and development budgets of the federal agencies that went into the SBIR program was small, the amount of funding available to small, high-tech companies dwarfed anything that had been readily available to them previously. The 11 federal agencies included in the program were the National Institutes of Health (NIH); National Science Foundation (NSF); Departments of Agriculture, Defense, Energy, Education, Health and Human Services, and Transportation; the Environmental Protection Agency; the Nuclear Regulatory Commission; and the National Aeronautics and Space Administration. Research in support of biomarker development efforts would probably come from the NIH, although other agencies have missions with enough overlap that limiting one's scope to the NIH might be shortsighted.

As constituted, the SBIR program is divided into three phases: phase I: feasibility, phase II: reduction to practice, and phase III: commercialization. The governmental agencies provide support for phases I and II. The third phase is expected to occur with private funding. The principal operative unit for the SBIR program is the SBIR grant. The phase I grant provides the awardee with a relatively small amount of funding: up to $100,000 currently to demonstrate the feasibility of the innovative research. The period of support is expected to be between six and 12 months. If feasibility is demonstrated, a phase II grant provides a much larger amount of support to carry the project from the feasibility phase to the point where it can be commercialized. Commercialization is to occur during the third phase with private funding.

Funding in phase II is routinely up to $1,000,000 currently, and there are mechanisms for increasing support beyond the level.

Efforts have been made by the National Research Council (NRC) to evaluate the success of the SBIR program: that is, to what extent it is fulfilling its mandate. Several publications summarizing the findings of the NRC are available [1–5]. Using successful commercialization as an endpoint, a study conducted in the early 1990s found that 12% of companies receiving funding through the SBIR program had commercialized the technology developed with program funds within four years of the phase II award, and 23% had achieved success by six years. A higher percentage of companies at each time reported optimism with regard to future commercialization.

Efforts have been made to elucidate the variables that led to success (i.e., commercialization). Commercialization was achieved in project funding through the NIH at a much higher rate than was achieved with funding from any other source. Among other interesting findings, commercialization was more likely to be achieved with product-oriented technology than with service-oriented technology. More success was achieved when the end user of the developed technology was in the private sector rather than in the government sector. Of particular importance to the present discussion, success was higher in those companies with a focus on tangible products than in companies with a research and development focus. Finally, technology that could be protected through patents, copyrights, and so on, achieved a higher level of commercialization than did technologies that were unprotectable in these ways. Other variables that improved commercialization success were a well-developed business and marketing plan for the technology and progress (at any stage) in implementing the plan. Companies that had potential partners as a source of capital, for co-development, or as customers also did better (not surprisingly) than those that did not.

Ultimately, however, there was so much variability in the individual projects initially funded and in those that achieved a measure of success that no single criterion had great predictive value by itself. Thus, one of the messages for a company seeking first-time support through the SBIR program is that while factors that have correlated with success in the past should not be discounted, success in the future will depend on the quality of the initial idea and success at the bench (i.e., turning a good idea into a proven product that meets a market need).

A subsequent, larger evaluation occurred approximately 10 years after the initial evaluation. This programmatic review came to many of the same conclusions as the initial review. The SBIR program was determine to be a success in its major mission: funding novel research in the private sector that ultimately had an economic payoff. It also arrived at the conclusion that the quality of the applications submitted (as well as those funded) had increased substantially over time. Like other funding programs within the various participating government agencies, the amount of funds available was judged to be insufficient to support all of the projects worthy of support based on

scientific merit. Summaries of these evaluations are available online at the National Research Council Web site, and many of the studies have been published as monographs.

## STRUCTURE OF THE SBIR GRANT PROPOSAL

### Phase I Application

The SBIR grant application process proceeds in two steps. The initial step is the phase I application. The application is (relatively) small and uncomplicated. The 15-page work plan introduces the overall research goal of the project, provides a rationale for doing the research, suggests feasibility, and outlines a series of experiments that (if successful) will support feasibility. Phase I funding provides up to $100,000 to demonstrate the merit of the investigation proposed. The length of support for the typical phase I application is usually six to 12 months.

The phase I grant application must provide a rationale for the study and a work plan that describes experiments that, if successful, would provide evidence of feasibility. The rationale is critical. If the reviewers are not convinced that the study is worth funding (regardless of outcome), their will be no enthusiasm for the application. Preliminary data are not required for a phase I application. Although this may be technically correct, it is very difficult to argue convincingly that the undertaking is feasible without some preliminary data. Thus, virtually all applications contain some evidence to suggest that the approaches proposed will be successful. Regarding the work plan itself, this must include criteria that define success. The criteria need to be spelled out in such a way that at the end of the phase I studies, it will be possible to justify the investment in phase II.

A detailed plan for commercialization is not required for the phase I application. However, if the reviewers of the application cannot see how the novel technology or product being developed can be commercialized, the grant will not receive a fundable score. Therefore, it behooves the investigator to link the proposed studies directly to the tangible outcomes.

After introducing the subject and providing background information, a strong rationale for the work, and whatever preliminary data are available, the detailed experimental design and methods are presented. This section needs to be concise (since probably only six to nine pages will be available), but more important, a longer work plan will stray from the mission, which is to convince review panel members that the studies proposed will (if successful) demonstrate feasibility. It is generally acknowledged that a detailed work plan, no matter how well written and how logical, cannot rescue an application with a weak rationale. On the other hand, if the rationale for the project is solid, the investigator can still "shoot himself in the foot" by failing to provide an adequate research plan.

## Phase II Application

At the end of the phase I period of support (normally six to 12 months, or when the investigator believes that feasibility has been demonstrated), a phase II application is submitted for a substantially larger amount of money (up to $1,000,000 or more). The phase II application has a work plan that is up to 25 pages in length. The phase II application has several important differences from the phase I application, in addition to size.

A successful phase II application has several critical elements. First, the value of the project goal has to be convincing. If the review panel feels that the value of the technology being developed will, ultimately, have little scientific or economic impact, there is no chance of redeeming the project, regardless of how well written and convincing the rest of the application. Often, investigators will assume that the project's value is already accepted, based on the fact that "they funded the phase I studies." The reality is that review panel members are less likely to be concerned with the overall project's value during phase I review since the amount of money awarded is small. Not so for phase II. Most reviewers are acutely aware of fund limitations and will do their best not to fund projects that seem to have no long-term payoff. Therefore, although it is critical that the project be justified at phase I, it is even more important to justify it in the phase II application. It goes almost without saying that demonstrating *both* scientific and economic merit is critical. The successful grant has both. The "Background and Significance" section is the place to provide this information.

The "Progress Report and Preliminary Studies" section is next. It is critical that the progress report justify the continued investment. Some investigators assume incorrectly that completing the experiments precisely as outlined in the initial phase I application and obtaining the hoped-for results in these studies are sufficient. The reality is that regardless of what was proposed in the initial application, the reviewers of the phase II proposal will want to be convinced that the project has merit (as assessed by them independently) and that data presented in the "Preliminary Studies" section strongly support the phase II work plan. Often, this means including a large amount of information that is above and beyond that proposed originally. The reviewers of the phase II application will not care what was proposed in phase I per se. What they will care about is the likelihood of success in phase II. Undoubtedly, much of the data that ends up in the progress report for the phase II application will have been generated with funds over and above what was allocated in the phase I grant.

The "Experimental Design and Methods" section follows. Exactly as in phase I, this is where the investigator lays out the approaches that will be used, the rationale for the approaches, and a brief description of experimental procedures. As with other grants, the rationale for the approaches to be used is more important than the actual methods. The reviewers will not want to see methodological detail unless the methods proposed are, themselves, novel. Remember—only so many pages are allowed. Pages spent describing routine

**TABLE 1 Typical Phase I and Phase II SBIR Application**[a]

| Section | Phase I | Phase II |
|---|---|---|
| Specific Aims | 1 page | 1 page |
| Background and Significance | 2–3 pages | 3–5 pages |
| Preliminary Studies (Progress Report) | 3–5 pages | 7–8 pages |
| Experimental Design and Methods | 6–9 pages | 11–14 pages |

[a]The maximum length for a phase I work plan is 15 pages and for a phase II work plan is 25 pages.

procedures are pages not available for convincing the reviewers that this proposal will provide critical answers to important questions. Table 1 lays out what is typically seen in the workplan of an SBIR application (phases I and II) submitted to NIH.

The experimental design section of a phase II SBIR application being submitted to NIH is similar to that of an investigator-initiated NIH application (R01). The most important piece of the application is the significance. If the reviewers are not convinced that the project is worthwhile, it will be virtually impossible to resurrect the application in later sections. The second most important section is the progress report and preliminary studies section. If the reviewers are not convinced that the data support the studies proposed, they will not score the application as competitive. The experimental design section is therefore the least important of the three. Having said that, however, if the experimental design section is not well laid out, the reviewers will conclude that carrying out the studies as proposed will not answer the critical questions. An application with such a work plan will not be funded. To reiterate, it is axiomatic that the experimental design section cannot save a noncompetitive application but can sink a grant that is otherwise competitive.

Everything stated above for the typical SBIR grant application can be applied to other, more traditional grant applications. Grant reviewers, particularly those who review applications for NIH, are attuned to this format. Nothing in the work plan will clearly designate this as being an SBIR application. What does distinguish an SBIR application from other grants is the business plan. Technically speaking, there is no need for a detailed business plan for a phase I application. However, somewhere in the grant, it is important to convey to the reviewers how success in the endeavor will be commercialized. Although not a major part of the phase I application, the well-developed business plan is an integral part of the phase II application. It is critical that the application describe in detail how the technology or product being developed will be brought to the marketplace. It is not a question of details but, rather, that the relevant issues have been thought through: identification of the target market, how large the market is, what actually comes to market, some idea of what it will cost to accomplish this, and so on.

For example, one recently funded application had the goal of replacing an animal-derived protein with a fully synthetic counterpart in vaccine manufacturing protocols. Key information in the business plan included an estimate of the size of the target market (i.e., how many vaccine protocols could make use of the synthetic product, how many units of vaccine this would entail, the percentage of that market that could be expected to utilize the animal-free alternative), how much of the synthetic product would be needed to meet the market need, what regulatory challenges need to be addressed, and most important, what the projected cost would be for bringing the replacement moiety to market. Estimates for each had to be convincing (not necessarily accurate or precise). If any of these were not convincing, it would not matter how interesting the science was. The project would not generate strong enthusiasm from reviewers.

In addition to addressing the specifics of any project, the reviewers want to see that the company has the necessary infrastructure to carry the project successfully beyond the end of phase II. No matter how good the technology is, there has to be a plan for commercializing it and the company requesting support must be capable not only of formulating such a plan but also for carrying it out successfully. While each business plan is unique, there are several features that most successful applications include. Some of the most important issues to address in the business plan are:

- About the project
  - Business opportunity (who or what the market is for the technology)
  - Magnitude of the market
  - Final product or service (what the commercialized product or service will look like)
  - Additional research (toxicology/clinical trials, etc.)
  - Other regulatory hurdles
  - Manufacturing cost estimate
  - Cost of delivering the service
  - Where the money will come from [plan for phase III funding (private sector)]
  - Pricing and sales
  - Marketing
  - Company commitment to project
- About the company
  - Company structure
  - Business experience and qualifications of principals
  - Capitalization
  - Strategic vision
  - Current products and markets

- Strategic alliances
- Investment strategies
- Intellectual property

In summary, virtually all (open competition) grant applications to federal government agencies are highly competitive. Although it might have been thought at one time that seeking funds through the SBIR program provided a less competitive route, this is no longer the case. Only high-quality applications are funded in today's climate. Even when the application is of the highest quality, there is a strong possibility that it will not be funded. There are many reasons for this. Needless to say, a certain percentage of the grant applications reviewed in any study section have been reviewed at least once previously.

## Revised Applications

In today's funding climate, very few applications of any kind are funded on the first submission. Anyone seeking support from the SBIR program faces the strong likelihood of having to rewrite and resubmit his or her proposal. Very soon after review of the initial grant is complete, the applicant is given a score. At this point, the applicant will know whether she or he is "in, out, or on the fence." If the score is definitely not fundable, the process of restructuring the grant begins. Most applicants will think about the resubmission even if they are in that large gray area containing the grants that may or may not be funded. The strategy is to prepare for resubmission and happily drop the project if funding comes through. In any event, there is not much in the way of concrete changes that can be made until the critique sheet arrives. In the old days, this was the infamous "pink sheet." The paper on which the critique was written was, in fact, pink. Now, like everything else, it is electronic. The critique sheet summarizes the comments of the primary and secondary reviewers and any discussion of the project in the study section. This may arrive as much as six to eight weeks after the review.

The resubmission differs from the initial application in that it contains an additional section: "Introduction to Revised Application" (one page for revised phase I proposals and up to three pages for revised phase II proposals). The Introduction summarizes the major criticisms and indicates how the investigators have altered the proposal in response. It goes without saying that responding fully to the issues raised (even if they seem trivial or incorrect) is absolutely essential. One can point out (hopefully, politely) where the reviewers are incorrect and not change the grant application's content. This may be a workable strategy in some cases. Most of the time, however, altering the application is the better choice.

The introduction summarizes how the revised application has been altered as a result of the initial review. The major changes come in the appropriate sections of the grant itself. Rewriting the background and significance section may be necessary if the reviewers do not appreciate the significance of the

work. Additional preliminary data may be added if the reviewers are not convinced that the work is feasible. New methodology is included to bolster a challenge to the approach. These changes aside, the revised application, like the initial version, has the same goal: convincing the review panel members that the proposed research has the potential of providing a novel technical solution to a pressing problem.

## FUNDING BIOMARKER RESEARCH THROUGH THE SBIR PROGRAM

The discussion above applies to just about any project for which funding through the SBIR program might be sought. The following discussion relates, more specifically, to funding of biomarker development work. Developing a biomarker is, in many respects, similar to developing a therapeutic product. Both involve a focus on a particular disease or condition. While the therapeutic agent is, obviously, a tangible material that needs to be developed and tested in the laboratory, validated in a clinical study, and then manufactured to specifications, and so on, developing a biomarker has similar issues. Included are (1) demonstrating the possible relationship between a disease state and an expression of the putative marker; (2) establishing that the putative biomarker can be identified and quantified in tissue, plasma, or urine; (3) developing reagents that can be used reliably and conveniently to assess the marker at the appropriate site; and (4) conducting preliminary and definitive clinical investigations to demonstrate the feasibility of the marker and the utility of the reagents developed. The critical question to ask is this: At what stage in the process of developing a biomarker could one expect to obtain funding through the SBIR program? Another way to phrase the question: What specific question(s) should be put forward in the grant application?

At the theoretical end of the spectrum, one might hypothesize that a particular moiety would be a good biomarker (diagnostic or prognostic) for a particular condition, and then craft an application around demonstrating that such was the case. This might be a candidate for SBIR funding since it is logical to assume that if the moiety could be shown to have biomarker potential, "the commercial ramifications would be obvious." However, reviewers of a grant with such a focus might determine this to be too basic in nature. Slightly less theoretical would be a study to demonstrate that a moiety which was already presumed to be associated with a disease state could be identified reliably (and quantified) in tissue, plasma, or urine. Phrasing the question in this way moves the project one step closer to commercialization. The problem with this approach would be, of course, that one would have to know at the start that factor X was, indeed, a biomarker candidate. Moving even closer to the commercialization end of the spectrum, one could argue that the moiety in question was a good candidate and that there was solid evidence that it could be measured in some relevant biological sample. The issue for the grant would

be developing reagents and assay procedures necessary for doing this reliably. Clearly, such a project would not seem too theoretical for the SBIR program. On the other hand, some might consider this to be "product development" and not innovative. Finally, one could propose clinical studies to confirm that demonstrating the presence of a particular moiety in a biologically relevant specimen would have significant predictive value. Such a study would not be thought of as product development, but the "product" aspect (i.e., the commercial potential) would easily be understood. Additionally, such an application would lend itself to SBIR support in that a preliminary assessment could be made in the phase I portion of the grant and a more definitive study during phase II. The downside of such an approach is that by the time that the project was funded through the SBIR grant, millions of dollars would already have been invested in it. Funding at this stage is not what most companies are looking for in an SBIR grant. Furthermore, even a phase II award would provide far less money than required for most clinical studies. Most companies that had gotten to this stage would be less interested in spending the time and effort needed for the amount of money that would be obtained if successful.

Regardless of the stage at which one approaches a government agency for SBIR funding, it is critical to remember that this a research proposal, not a contract proposal. The two are fundamentally different. The successful grant proposal must have an overall goal (i.e., a work plan tied to a testable hypothesis). Achieving the overall goal must be highly desirable, but not assured. The phase I portion of the SBIR grant is designed, a priori, to provide evidence that a novel idea has merit. If the feasibility has already been demonstrated, there is no "testable" hypothesis. Without this, there is no grant. Often, what comes to the SBIR grant reviewers is a "classical" contract proposal. There is a proposed plan of work. The plan will be followed "to the letter" regardless of findings. There is no testable hypothesis. The problem with such a proposal (from the standpoint of the SBIR program) is that, by definition, it is not innovative. The critique of such an application will undoubtedly contain comments such as "this is development rather than research" and "this should be funded through company resources." The end result is: no money.

The antithetical issue can be just as problematic. To reiterate, although the reviewers of an SBIR grant will want to see innovative research, the project will not be funded if it consists entirely of research, regardless of how innovative. The successful SBIR application must straddle a fine line between basic research and development. The successful application must not only present an interesting hypothesis and a detailed plan, but a road map of how verifying the hypothesis will translate directly into tangible benefit.

Another issue that is pertinent to biomarker development (as well as to the development of a therapeutic) is the nature of the problem for which the biomarker or therapeutic is being developed. Conditions that are primarily cosmetic or of a trivial medical nature, as opposed to serious medical conditions, will not elicit much enthusiasm, regardless of how novel the research is.

Even if the technology ultimately developed has the potential to be economically viable, interest will be low. In a like manner, interest will be low if other diagnostic or prognostic markers are already available for the condition under study. Of course, demonstrating that the new biomarker is clearly superior to existing diagnostic or prognostic markers in some way would mitigate this criticism. This is especially true if existing markers are a part of a foreign-owned technology.

Disease biomarkers are, by definition, moieties whose detection and/or quantification has predictive value in relation to the disease. In most cases, moieties with significant predictive value have such value because they are related in some way to disease pathophysiology. Thus, it is a good strategy, if possible, to combine mechanistic studies with the more technology-driven aspects of the research program. It would not be a good idea to focus entirely, or even primarily, on the mechanistic studies, as the reviewers would question the tangible benefits of the work. Nonetheless, mechanistic studies in the appropriate context always help in that they raise the overall interest level of the reviewers and provide a measure of credibility to the investigators. It goes almost without saying, of course, that weak mechanistic studies are worse than none at all. Only the investigative team can know how their proposal is likely to be helped or hindered by whatever mechanistic studies they can incorporate into their application.

There are a number of additional issues that should be considered before trying to fund a biomarker study through the SBIR program. As indicated above, the feasibility portion (phase I) of the grant may be limited to $100,000 or less. Any study that involves obtaining and processing clinical specimens quickly becomes very expensive. It is important to understand what can be accomplished with the level of funding expected and whether this will satisfy the reviewers at phase II that feasibility has been demonstrated. One can argue all he wants that what was accomplished was limited by the funding available. This reasoning won't, however, be persuasive to reviewers of the phase II application. The only thing they will want to be convinced of is that feasibility has been demonstrated. Given this reality, much of what is accomplished during phase I in many studies is supported only partially by the phase I application. Cost sharing with retained company earnings is common. This is not necessarily good or bad; the point is to be prepared to partially fund the work independent of the SBIR grant.

Another consideration is the timing. Corporate investigations, particularly at the initiation of projects, are fast-paced compared to the workings of any federal grant program. The SBIR program was designed to have a short lag between application and funding, but it is still a slow process. There are typically three deadlines spaced throughout the year. Should a company intend to meet the deadline for a phase I submission in April, much of February and March will be consumed with putting together a high-quality application. That assumes, therefore, that the idea for the project has already been generated and thought through (at least in a preliminary way). Whatever preliminary

data are to be included will also have to have already been generated. If the grant submitted for the April deadline is approved for funding without revision, the earliest time that funding would be available is the latter part of the year. Today, very few grants are funded on the first cycle. What this means is that even a high-quality application for an interesting and supportable project will probably need to be resubmitted at least once. The grant submission and review cycles are constructed so that a revised application cannot be submitted for the deadline immediately following the cycle in which it was originally submitted. The earliest that a revised application could be resubmitted for one initially put forward in April would, therefore, be December. Assuming that the applicant responded (successfully) to all of the initial criticisms and was approved for funding, June would be about the earliest that one could expect the money to arrive. Thus, a gap of one and a half years between conception of the grant and funding is not unreasonable. Often, the criticisms are such that two or more deadlines pass between the initial application and the revised submission. Often, a second resubmission is required.

When funding for the phase I application is attained, the period of support is normally up to one year. This is the period in which the studies needed to demonstrate feasibility are conducted. Often, they may not be completed within the one year of support. In that case, there is no reason why additional experiments, as necessary, cannot be carried out with company funds. It is simply a question of time and money. If the idea has merit, this is often what occurs. At whatever time it is deemed that sufficient evidence for feasibility is in hand, the phase II application is written. Again, assume one or two months for putting the application together to meet one of the three annual deadlines. Again, assume one or even two rewrites before it is approved, bringing the gap between the end of phase I and the restart of phase II at between one and a half to two and a half years. The point is that a small company should not intend to support its R&D effort on funding through the SBIR program alone. Rather, it is better to think of the program as a way to supplement ongoing research activities with new initiatives, and as a way to provide additional resources for essentially untried and inherently risky endeavors.

## SUMMARY

In summary, projects that are good bets for SBIR funding straddle a fine line between innovative research and product/technology development. If the reviewers deem the project to be too far to one side or the other, the outcome will be poor. Projects that do well in review also have a well-defined business plan: a plan to take the product or service from successful research to commercialization. Projects that actually go to completion generally have a mix of funding. Often, the SBIR-funded portion is small. Finally, the time from project initiation to successful commercialization is usually quite long when

clinical studies and regulatory approval are mandated. The SBIR program can provide the impetus for project initiation, in some cases cutting the time to successful commercialization. More important, the seed money provided through an SBIR application will, if used in accord with the way the program is designed to work, provide the impetus for certain high-risk approaches that would not be undertaken in the absence of such funding.

## REFERENCES

1. SBA (1991). Results of three-year commercialization study of the SBIR program. *SBA Small Business: Building America's Future*. U.S. Small Business Administration, Washington, DC.
2. Wessner CW (ed.) (1999). *The Small Business Innovative Research Program: Challenges and Opportunities*. Board on Science, Technology and Economic Policy. National Research Council, Washington, DC.
3. Wessner CW (ed.) (2008). *An Assessment of the SBIR Program*. Committee on Capitalizing on Science, Technology, and Innovation: An Assessment of the Small Business Innovation Research Program, National Research Council. National Academies Press, Washington, DC.
4. Wessner CW (ed.) (2007). *An Assessment of the Small Business Innovative Research Program at the National Institutes of Health*. Committee on Capitalizing on Science, Technology, and Innovation: An Assessment of the Small Business Innovation Research Program, National Research Council. National Academies Press, Washington, DC.
5. Wessner CW (ed.) (2007). *SBIR and the Phase III Challenge of Commercialization: Report of a Symposium*. Committee on Capitalizing on Science, Technology, and Innovation: An Assessment of the Small Business Innovation Research Program, National Research Council. National Academies Press, Washington, DC.

# NOVEL AND TRADITIONAL NONCLINICAL BIOMARKER UTILIZATION IN THE ESTIMATION OF PHARMACEUTICAL THERAPEUTIC INDICES

Bruce D. Car, B.V.Sc., Ph.D., Brian Gemzik, Ph.D., and William R. Foster, Ph.D.

*Bristol-Myers Squibb Co., Princeton, New Jersey*

## INTRODUCTION

Accurate projection of the safety margins of pharmaceutical agents from late discovery and early nonclinical development phase studies, including in vitro and animal studies, to humans is fundamental to the first decision to move compounds forward into the clinic. The robustness of those estimates over time is also central to the ability to conduct proof-of-concept studies in humans in phases IIa and IIb. After sufficient clinical experience, direct human information renders the nonclinical projections redundant. When nonclinical projections are discrepant with safe clinical exposures, discovery strategies for backup compound selection should be adjusted appropriately.

The essential elements of this work are well-defined no-effect levels (NOELs) or no-adverse effect levels (NOAELs) and lowest observed effect levels (LOELs) or $IC_{50}$s/$EC_{50}$s, if a molecular off-target is known, both as

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

unbound drug and plasma protein–bound drug concentrations, expressed as area under the curve (AUC), $C_{max}$, or concentration at a defined time point. A ratio is calculated from the relevant parameter ($C_{max}$, AUC, time above a

**TABLE 1 Tiers of Biomarkers**

| Toxicity Endpoint | Traditional Biomarkers | Novel or Secondary Endpoint[a] |
|---|---|---|
| Hepatocellular necrosis | Histopathology, ALT, AST, SDH, LDH | Markers of apotosis, gene signature, circulating RNA[1] |
| Renal tubular injury | BUN | Cystatin c, urinary alpha GST, urinary GGT, urine sediment |
| Renal glomerular injury | Creatinine | Urine protein electrophoresis |
| Seizure | Observational finding | Electroencephalographic seizure, repetitive sharp waves |
| Retinal degeneration | Histopathology | Electroretinography |
| Systemic phospholipidosis | Histopathology and electron microscopy | Evidence of organ dysfunction or PL storages in leukocytes, metabonomic[2] or transcriptomic[3] profiles |
| Small intestinal mucous metaplasia of gamma secretase inhibitors | Histopathology and clinical signs of diarrhea | Peripheral blood genomic biomarkers of Notch signaling inhibition[4] |
| Na channel inhibition | QRS prolongation | Decreased dP/dT determined telemetrically |
| hERG channel inhibition | QT interval prolongation | Integration of ion-channel effects in Purkinje fiber assay |
| Mutagenicity | Ames assay positivity | Genomic biomarker of DNA repair |
| Carcinogenesis | Malignant tumors in chronic toxicity studies or two-year bioassays | Gene signatures predictive for carcinogenicity[5–9] |
| Teratogenicity in rats or rabbits | Positive findings in development and reproductive segment II studies | In vitro whole embryo culture recapitulating effects at same concentrations |

[a] [1]Miyamoto et al., 2008; [2]Delaney et al., 2004; [3]Sawada et al., 2005; [4]Milano et al., 2004; [5]Ellinger-Ziegelbauer et al., 2008; [6]Fielden et al., 2007; [7]Fielden et al., 2008; [8]Nie et al., 2006; [9]Andersen et al., 2008.

certain concentration, $IC_{50}$) between the NOEL or NOAEL values and that same parameter at the efficacious human dose projected. The science of the projection of the clinical dose has become more refined (Huang et al., 2007), allowing preclinically determined therapeutic indices to have greater predictive value.

This ratio is the safety margin. A safety margin is considered a therapeutic index if a relevant efficacy or pharmacodynamic endpoint is included in the animal study. The two terms are frequently used interchangeably, although therapeutic indices by nature of their derivation should generally be considered better predictors of drug safety. Considerable skill is required for accurate human pharmacokinetic and pharmacodynamic prediction to facilitate accurate estimation of therapeutic index. Superimposed on this numerical calculation is the toxicologist's understanding of the therapeutic index that incorporates the severity of the LOELs or other higher-dose effects, its reversibility, how cumulative exposure influences the safety margin estimate over time, and the ability to monitor in clinical evaluation. A therapeutic index of greater than 1 in at least two nonclinical species indicates that a compound can generally be given to humans safely up to the efficacious concentrations projected. Therapeutic indices below 1, which frequently occur with oncologics, alert the clinician that toxicities should be expected and monitored for at exposures below those projected to have benefit for patients.

Drug discovery toxicology and development nonclinical safety groups, together with ADME (absorption, distribution, metabolism, and elimination) and discovery pharmacology groups, influence the progression of compounds through the progressive refinement of estimates of their therapeutic indices. Traditional therapeutic indices calculated from findings in toxicology studies are broad in nature, including such diverse endpoints as liver necrosis, seizure, or prolongation of the electrocardiographic QT interval. Such types of endpoints are unambiguous, and may be refined further with an additional tier of biomarker information, such as clinical chemistry, electroencephalography, or advanced electrocardiography (e.g., instrumented animals monitored by telemetry). When novel biomarkers prove to be more sensitive in detection and time relative to traditional biomarkers, they may supplant or be used to supplement traditional approaches. Examples of these differing tiers of biomarkers are provided in Table 1.

When pharmacodynamic or efficacy endpoints are available for a nonclinical species and form the denominator of the therapeutic index equation, the most predictive safety margins may be calculated, assuming that the particular species has exposure, metabolite profiles, and other ADME characteristics similar to those of humans. Creative research in applications or technologies to validate such endpoints in nonclinical species greatly enhances the predictive power of animal models of toxicity. Several examples of these are included in Table 2.

**TABLE 2 Efficacy Endpoints of Traditional and Novel Pharmacodynamic Biomarkers**

| Efficacy Endpoint | Traditional Pharmacodynamic Biomarker | Novel Pharmacodynamic Biomarker |
|---|---|---|
| Decreased anxiety, depression | Neurobehavioral change | CNS receptor occupancy |
| Cognition improvements | Learning tasks | Altered phosphorylation of proteins in key pathways |
| Improving Alzheimer dementia | Peripheral blood exposure and CNS Aβ (ex vivo) | CSF Aβ concentrations |
| Cancer xenograft regression | Size and weight of xenograft | Altered phosphorylation of proteins in key pathways |
| Immune modulation of rheumatoid arthritis | Arthritis scores (ACS 20, 50, 70) | Plasma cytokines, FACS, leukocyte transcriptomics |

## IN VITRO THERAPEUTIC INDICES

When pharmacology targets and homologous or nonhomologous secondary off-targets are molecularly well defined, the temptation to calculate ratios of $IC_{50}$s at desired efficacy endpoints to safety endpoints leads to the creation of *in vitro therapeutic indices*. Typically, these are large numbers that lull many a discovery working group into a false sense of security. Two examples are provided based on real outcomes, for which large in vitro ratios would potentially create an illusion of greater safety. For example, an oncology compound had a hERG (Ikr; repolorizing $K^+$ current) $IC_{50}$ value of 35 μM and a target receptor $IC_{50}$ value of 10 nM, ostensibly providing a 3500-fold safety window.

Caveats for the use of these simple formulas are:

- Plasma protein binding if pharmacologic or toxicologic activity relates to the free fraction
- Relative concentration in tissue may exceed plasma by many fold
- Tissue-bound drug, and thus tissue concentration at efficacy and toxicity targets can be difficult to determine and may influence the expression of efficacy or toxicity
- $C_{max}$/trough ratio.
- Typically, one uses $IC_{50}$ values for ion channels, although inhibition of hERG at an $IC_{10}$ may still produce clinically important prolongation in the QT interval
- The efficacy and toxicity of metabolites

After integrating these various considerations to the calculation of a safety margin, and considering the simple unknown, which is the biological counter-

part to in vitro activity when measured in vivo, the safety margins for QT prolongation due to hERG inhibition were in the five to tenfold range.

In a second example, a compound with $IC_{50}$ for phosphodiesterase (PDE4) intravenously of 2 μM was considered safe for a central nervous system efficacy target with $EC_{50}$ of 1000-fold less (2 nM). At the lowest dosage tested in animals, projected to provide a threefold safety multiple, portal vasculitis was observed in rodents, considered likely to be secondary to PDE4 inhibition at a plasma $C_{max}$ of only 70 nM. This toxicity is generally driven by $C_{max}$, and as peak concentrations occur in the portal vasculature during absorption of drug from the small intestine, the potential toxicity can be markedly exaggerated relative to in vitro determined safety windows.

Frequently, the exposure–response relationship for in vitro surrogates of in vivo toxicity can be quite poor. Although this makes prediction of valid safety windows difficult, in vitro–determined numbers are still valuable in permitting the rank ordering of compounds within a chemical series for selection, and for refining in vivo assessments.

## NOVEL METABONOMIC BIOMARKERS OF TOXICITY

The application of systems biology technologies, including metabonomics, proteomics, and transcriptomics to biomarker development, is a nascent science, with relatively few examples of the impactful prospective uses of these technologies (Lindon et al., 2004; Robertson, 2005; Car, 2006). The following example describes how understanding the mechanism of rat urinary bladder carcinogenesis combined with metabonomic profiling of urine, yielded a mechanism-specific biomarker that could be evaluated in studies with patients.

### Muraglitizar-Induced Urinary Bladder Transitional Cell Carcinoma

Peroxisome proliferator-actuated receptors (PPARs) are nuclear hormone receptors targeted for therapeutic modulation in diabetes. Specifically, PPARα agonism will control dyslipidemia, while PPARγ agonism affords improved glucose homeostasis. Nonclinical and clinical safety issues have prevented PPARαγ agonists from becoming drugs (Balakumar et al., 2007, Rubenstrunk et al., 2007). The results of two-year rodent carcinogenicity studies, including hemangiosarcoma, liposarcoma, and urinary bladder transitional cell carcinoma, have generally clouded a clear human risk assessment. With widespread distribution of PPARα and PPARγ receptors in tissues, including those transformed in carcinogenesis, a clear separation of the potentially beneficial role of receptor agonism from the potentially adverse contribution to tumor development is complex to research and understand.

The investigative approaches directed toward establishing a cogent human risk assessment for dual PPAR agonist–induced urinary bladder transitional cell carcinoma in rodents are described here for the PPARα/γ agonist mura-

glitazar (Dominick et al., 2006; Achanzar et al., 2007; Tannehill-Gregg et al., 2007; Waites et al., 2007). An increased incidence of ventral bladder wall transitional cell papillomas and carcinomas of the urinary bladder were noted in rats at doses as low as eight-times the projected human exposure at 5 mg/kg (Tannehill-Gregg et al., 2007). Histopathology and scanning electron microscopy revealed early microscopic injury associated with the presence of calcium phosphate crystals. Crystalluria was confirmed in studies designed to document the fragile and sometimes transient crystals in male rats dosed with muraglitazar. The crystal-induced epithelial injury was hypothesized as initiating the increased turnover as confirmed in BrdU-labeling experiments of the ventral bladder urothelium, a proliferative response strongly suspected in the genesis of tumor development. To determine the potential role of crystalluria in injury, and carcinogenesis, crystals were solubilized in rats through urinary acidification with 1% dietary ammonium chloride. Urinary acidification of male rats dosed with muraglitazar abrogated crystalluria, early urothelial injury, and cell proliferation (urothelial hyperplasia), and ultimately, urinary bladder carcinogenesis. This mode of action is recognized as a nongenotoxic mechanism of urinary bladder carcinogenesis in rats (Cohen, 1999).

To evaluate a potential role for pharmacology, the regulation of genes downstream of PPARα and PPARγ in the rat bladder urothelium were evaluated in the presence of PPARαγ agonist–treated crystalluric and acidified diet, noncrystalluric rats. No changes in gene expression or traditional endpoints were observed, suggesting that PPAR-mediated changes were not directly causative in urothelial profileration or carcinogenesis (Achanzar et al., 2007).

To investigate further the mechanism of muraglitazar-induced crystalluria, urine samples were collected from treated rats for metabonomic analysis. NMR spectroscopic evaluation of urine from treated compared to control rats revealed a striking reduction in divalent acids, including citrate and 2-oxoglutarate. Subsequent analytical-grade analyses of urinary citrate to creatinine concentrations confirmed and extended these metabonomic findings. It was hypothesized that male-rat-specific decreased urinary excretion of divalent acids, and in particular citrate, contributed to a milieu highly permissive of calcium phosphate crystal formation.

Based on the results of studies conducted in rats, a final set of experiments examined the absolute excretion of citrate in urine from humans treated with muraglitazar. No reductions in citrate concentrations were observed across many patients compared to placebo and pretest populations. Therefore, a research strategy based on determining the role of urinary crystallogenesis in rats suggested that muraglitazar was unlikely to pose any risk to humans in inducing the early procarcinogenic change observed in rats.

The muraglitazar example demonstrates how preclinical metabonomics evaluations may identify biomarkers with potential clinical impact; however, the potential for this technology to yield specific and sensitive individual or multiple-entity biomarkers is also largely unrealized (Lindon et al., 2004; Robertson, 2005; Robertson et al., 2007).

## NOVEL TRANSCRIPTOMIC BIOMARKERS

Disease- and toxicity-specific transcriptional and metabonomic biomarkers are an as yet largely untapped reservoir; however, publications investigating such biomarkers have become increasingly visible in the literature (Fielden and Kolaja, 2006; Foster et al., 2006; Robertson et al., 2007). In a retrospective review of several years of toxicogenomic analyses of drug safety studies, biomarkers of pharmacology were readily identified in 21% of studies (40% of drug targets) (Foster et al., 2006). An unvalidated version of such an mRNA signature is that of proliferation inhibition observed consistently in the liver of rats given oncologics. A set of such genes is illustrated in Table 3.

These observations can readily be adapted to transcriptomic signatures and used to determine which doses, for example in a toxicology study, demonstrate compound efficacy. When combined with traditional endpoint data, a therapeutic index may then be derived. This approach is particularly useful when the pharmacology of a compound has not been evaluated in the test species. Although early transcriptomic signatures consistent with previously identified pathology are frequently observed (in approximately 50% of studies of target tissues profiled at times preceding pathology), the target tissues involved are rarely analyzed transcriptionally such that finding valid predictive signatures will continue to be problematic (Fielden and Kolaja, 2006).

Transcriptomic signatures may also provide insight toward pharmacologic effect in distinct patient groups. The demonstration of increased expression of wild-type and mutant Kras by both immunohistochemistry and transcriptional profiling or real-time polymerase chain reaction (RT-PCR) led to the hypothesis that the selection of certain patients dosed with EGFR inhibitors based on the expression and presence of mutation could markedly increase the responder rate of patients (Lièvre et al., 2006; Di Nicolantonio et al., 2008). The ability to triage patient groups and eliminate patients for whom potentially toxic medicines offer no benefit is clearly a huge advance in the practice of oncology.

**TABLE 3 Genes Commonly Changed by Diverse Oncologic Agents in Rat Liver**

| Gene | Transcriptional Change | Gene Function | Gene Name |
|---|---|---|---|
| *Rrm2* | Repression | Proliferation | Ribonucleotide reductase M2 |
| *Cdc2a* | Repression | Proliferation | Cell division cycle 2 homolog A |
| *Cdkn1a* | Repression | Proliferation | Cyclin-dependent kinase inhibitor 1A |
| *Ccnb1* | Repression | Proliferation | Cyclin B1 |
| *Dutp* | Repression | Proliferation | Deoxyuridine triphosphatase |
| *Csnk1a1* | Induced | Cell survival | Casein kinase 1, α 1 |

## CONCLUSIONS

Accurate determination of therapeutic indices from nonclinical studies across multiple species, overlain with an understanding of the human risk associated with nonclinically identified liabilities, provides an invaluable tool for advancing compounds with reduced potential for harm and reduced likelihood of attrition for safety concerns. Novel approaches for identifying and validating biomarkers combined with highly refined clinical dose projections will allow toxicologists to predict and therefore avoid clinically adverse outcomes with increasing accuracy.

## REFERENCES

Achanzar WE, Moyer CF, Marthaler LT, et al. (2007). Urine acidification has no effect on peroxisome proliferator–activated receptor (PPAR) signaling or epidermal growth factor (EGF) expression in rat urinary bladder urothelium. *Toxicol Appl Pharmacol*, 223:246–256.

Andersen ME, Clewell H, Bermudez E, Wilson AG, Thomas RS (2008). Genomic signatures and dose-dependent transitions in nasal epithelial response to inhaled formaldehyde in the rat. *Toxicol Sci*, 105:368–383.

Balakumar P, Rose M, Ganti SS, Krishan P, Singh M (2007). PPAR dual agonists: are they opening Pandora's Box? *Pharmacol Res*, 2:91–98.

Car BD (2006). Enabling technologies in reducing drug attrition due to safety failures. *Am Drug Discov*, 1:53–56.

Cohen SM (1999). Calcium phosphate-containing urinary precipitate in rat urinary bladder carcinogenesis. *IARC Sci Publ*, 147:175–189.

Delaney J, Neville WA, Swain A, Miles A, Leonard MS, Waterfield CJ (2004). Phenylacetylglycine, a putative biomarker of phospholipidosis: its origins and relevance to phospholipid accumulation using amiodarone treated rats as a model. *Biomarkers*, 3:271–290.

Di Nicolantonio F, Martini M, Molinari F, et al. (2008). Wild-type BRAF is required for response to panitumumab or cetuximab in metastatic colorectal cancer. *J Clin Oncol*, 26:5705–5712.

Dominick MA, White MR, Sanderson TP, et al. (2006). Urothelial carcinogenesis in the urinary bladder of male rats treated with muraglitazar, a PPAR alpha/gamma agonist: evidence for urolithiasis as the inciting event in the mode of action. *Toxicol Pathol*, 34:903–920.

Ellinger-Ziegelbauer H, Gmuender H, Bandenburg A, Ahr HJ (2008). Prediction of a carcinogenic potential of rat hepatocarcinogens using toxicogenomics analysis of short-term in vivo studies. *Mutat Res*, 637(1–2):23–39.

Fielden MR, Brennan R, Gollub J (2007). A gene expression biomarker provides early prediction and mechanistic assessment of hepatic tumor induction by nongenotoxic chemicals. *Toxicol Sci*, 99:90–100.

Fielden MR, Kolaja KL (2006). The state-of-the-art in predictive toxicogenomics. *Curr Opin Drug Discov Devel*, 9:84–91.

Fielden MR, Nie A, McMillian M, et al. (2008). Interlaboratory evaluation of genomic signatures for predicting carcinogenicity in the rat: Predictive Safety Testing Consortium; Carcinogenicity Working Group. *Toxicol Sci*, 103(1):28–34.

Foster WR, Chen SJ, He A, et al. (2006). A retrospective analysis of toxicogenomics in the safety assessment of drug candidates. *Toxicol Pathol*, 35:621–635.

Huang C, Zheng M, Yang Z, Rodrigues AD, Marathe P (2007). Projection of exposure and efficacious dose prior to first-in-human studies: how successful have we been? *Pharm Res*, Sept 25.

Lièvre A, Bachet JB, Le Corre D, et al. (2006). KRAS mutation status is predictive of response to cetuximab therapy in colorectal cancer. *Cancer Res*, 66:3992–3995.

Lindon JC, Holmes E, Nicholson JK (2004). Metabonomics: systems biology in pharmaceutical research and development. *Curr Opin Mol Ther*, 6:265–722.

Miyamoto M, Yanai M, Ookubo S, Awasaki N, Takami K, Imai R (2008). Detection of cell-free, liver-specific mRNAs in peripheral blood from rats with hepatotoxicity: a potential toxicological biomarker for safety evaluation. *Toxicol Sci*, 106(2): 538–545.

Milano J, McKay J, Dagenais C, et al. (2004). Modulation of notch processing by gamma-secretase inhibitors causes intestinal goblet cell metaplasia and induction of genes known to specify gut secretory lineage differentiation. *Toxicol Sci*, 82:341–358.

Nie AY, McMillian M, Parker JB, et al. (2006). Predictive toxicogenomics approaches reveal underlying molecular mechanisms of nongenotoxic carcinogenicity. *Mol Carcinog*, 45(12):914–933.

Robertson DG (2005). Metabonomics in toxicology: a review. *Toxicol Sci*, 85:809–822.

Robertson DG, Reily MD, Baker JD (2007). Metabonomics in pharmaceutical discovery and development. *J Proteome Res*, 6:526–539.

Rubenstrunk A, Hanf R, Hum DW, Fruchart JC, Staels B (2007). Safety issues and prospects for future generations of PPAR modulators. *Biochim Biophys Acta*, 1771:1065–1081.

Sawada H, Takami K, Asahi S (2005). A toxicogenomic approach to drug-induced phospholipidosis: analysis of its induction mechanism and establishment of a novel in vitro screening system. *Toxicol Sci*, 83:282–292.

Tannehill-Gregg SH, Sanderson TP, Minnema D, et al. (2007). Rodent carcinogenicity profile of the antidiabetic dual PPAR alpha and gamma agonist muraglitazar. *Toxicol Sci*, 98:258–270.

Waites CR, Dominick MA, Sanderson TP, Schilling BE (2007). Nonclinical safety evaluation of muraglitazar, a novel PPARalpha/gamma agonist. *Toxicol Sci*, 100:248–258.

# ANTI-UNICORN PRINCIPLE: APPROPRIATE BIOMARKERS DON'T NEED TO BE RARE OR HARD TO FIND

Michael R. Bleavins, Ph.D., DABT
*Michigan Technology and Research Institute, Ann Arbor, Michigan*
Ramin Rahbari, M.S.
*Innovative Scientific Management, New York, New York*

## INTRODUCTION

Biomarkers have entered the drug development process with great fanfare, with the impression that this is the first time for a new approach. Projections of solving virtually all major issues faced in drug development, and the potential of new technologies revolutionizing medicine, have been promised by many specialists. In reality, biomarker assays represent the logical progression and integration of laboratory techniques already available within clinical pathology and biochemistry, with additional methods arising from new technologies such as molecular biology, genomics, proteomics, and metabonomics. The emphasis on biomarkers as a new approach often has lead to expectations that the ideal biomarker must be novel and/or exotic. To answer the pertinent questions during drug development, teams often embark on the quest for the "unicorn" of biomarkers, sometimes resulting in the best and most practical test being overlooked as the search progresses for elusive methods.

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

When working with drug development teams, especially those with limited hands-on experience in biomarker applications, it is imperative that the business need and scientific rationale for including a biomarker in the development plan be elaborated prospectively. By clearly defining both the question(s) to be answered and how a biomarker will advance the new drug's development, the likelihood of successful implementation to advance the compound and save time and laboratory resources can be dramatic. In most instances the most expedient biomarker approach will be identified by focusing on what is necessary to advance the compound at its particular stage of development, independent of how that parameter will be measured. Establishing precisely what is needed versus what would be nice to have may reveal opportunities not initially obvious to all participants.

Building a new biomarker on an emerging technology generally should be a last resort. Acceptance, reproducibility, quality control, standards, automation, and assessing laboratory-to-laboratory differences become exponentially more complex to characterize with the novelty of technology and lower numbers of groups using those approaches. For translation of a potential biomarker from the bench to the bedside, simpler in all aspects is preferable. Simpler tests to administer and evaluate do not mean less scientifically sound or relevant, and can provide a more solid foundation for acceptance in clinical and regulatory environments. This is not to say that the new technologies aren't braving new territory and having a significant impact on drug safety and development [1]. The emphasis of this chapter is to show how casting a wide net for ideas and approaches can expedite a compound's progress and improve decision making, keeping in mind that even less novel technologies will often be the best choice.

Under ideal conditions, and in the instances where the quest is for a decision-making biomarker, there should be solid prospective agreement to actually affect the drug's progress based on the biomarker results. To that end, scientists should keep in mind that the primary purpose of biomarkers is to enable better decisions. A better decision is one that can be made more confidently, earlier, less invasively, more efficiently, or the test is transferable to reference laboratories as the drug enters phase II or later. Therefore, it is essential that people supporting preclinical and clinical teams focus on the test(s) that aid in definitively graduating a compound to its next stage of development or to prove nonviability of the development efforts and compound attrition. This needs to be done without concern as to whether the biomarker derives from exciting new technology or is a new application of an established method using conventional science. In reality, if a team will not expedite, realign, or terminate a compound's development based on the biomarker results, inclusion of a non-decision-enabling biomarker is unlikely to serve a purpose other than to increase costs or complicate study design.

Sometimes the appropriate biomarker is a "unicorn" (exciting, rare, and exotic), but more often it is a "horse" (understood, generally accepted, and available) or a "mule" (proven performance, hybrid, and versatile). In bio-

marker selection, the primary consideration must be whether the test(s) is the best solution for the situation at hand. As a part of the biomarker development package, the biological rationale and significance (as understood at that stage), in addition to confidence in the platform, have to be evaluated. Another important factor is time; the development of a novel biomarker can be as complicated and time consuming as developing a new chemical entity. This can be essential criteria in circumstances where multiple companies are working in the same areas or patent timelines are not ideal. Identification, development, testing, characterization, and scaling of a new biomarker, for any one stage of the drug development process, generally requires at least six to nine months. By looking beyond the experience base or capabilities of one group or laboratory, it may be that your search for unicorns ends most appropriately by locating a horse residing in the stable of someone else. This chapter highlights examples of unicorn, horse, and mule biomarker approaches.

## UNICORN BIOMARKERS

Advances in technology that enable new applications can be valuable tools in addressing difficult questions of activity, efficacy, and safety for drug developers. The novelty of these approaches, and the high-tech equipment, create excitement and interest. Several new medicines actually owe their development and successful clinical utilization to tools that did not exist 25 years ago.

Gleevec (imatinib) was developed by Novartis to treat chronic myeloid leukemia (CML). This molecule, and two subsequent tyrosine kinase inhibitors for CML (Sprycel, Tasigna), were designed using imaging and molecular modeling specifically to target cells expressing the mutant BCR-ABL protein underlying the disease origin. The altered chromosome size arising from the balanced translocation between chromosomes 9 and 20 was a key feature in identifying genetic involvement in leukemia [2,3]. This formation of the Philadelphia chromosome, and the discovery that translocation resulted in the chimeric *BCR-ABL* gene, created new opportunities to target the resulting BCR-ABL oncoprotein and its role in hematologic cell proliferation [4,5]. The role of BCR-ABL and several other proteins in cell cycle progression is reviewed by Steelman et al. [6]. Cloning, mutational analysis, sequencing, and animal models [4,7–10] have proven to be valuable tools in designing effective treatments to CML. These techniques have also been important in developing the next-generation medicines for this disease since mutations in *BCR-ABL* arise in 50 to 90% of patients and result in resistance to imantinib [11–13]. Genetic testing has identified specific mutations that can be targeted by new tyrosine kinase inhibitors [12–14] and provides better clinical options for patients with these mutations. The use of specific mutational genotyping is an important consideration in developing new drugs, particularly for the T315I mutation that is currently resistant to all approved tyrosine kinase inhibitor compounds for CML. In addition to *BCR-ABL* mutational analysis, monitor-

ing the phosphorylation status of Crkl and Stat5 using flow cytometry, Western blotting, and/or enzyme-linked immunosorbent assay (ELISA) techniques provides biomarkers of compound activity [15,16].

Pharmacogenetics also was an important aspect in the target selection and development of the CCR5 antagonist class of anti-HIV drugs, including the Pfizer, Inc. compound maraviroc (Selzentry). The observation that persons homozygous for the CCR5-Δ32 mutation were resistant to the development of acquired immune deficiency syndrome (AIDS) was shown mechanistically to result from inhibiting the human immunodeficiency virus (HIV) binding to the mutated receptor and entering T-helper lymphocytes [17–20]. Maraviroc binds selectively to the human chemokine receptor CCR5 present on the cell membrane, preventing the interaction of HIV-1 gp120 and CCR5 necessary for CCR5-tropic HIV-1 to enter cells [21]. CXCR4-tropic and dual-tropic HIV-1 are not inhibited by maraviroc.

Genetic testing for the CRR5-Δ32 deletion to stratify clinical trial subjects was useful in determining whether small molecules showed differential activity in these groups, for establishing inclusion and exclusion criteria, and the characterization of safety [22]. These data were also key components in the successful registration of maraviroc in 2007. As HIV treatment has advanced, genetic assays for CCR5 mutations, as well as tropism for CXCR4 or CCR5 co-receptor, are proving useful in optimal use of these co-receptor antagonists. In fact, the Selzentry label [21] recommends tropism testing to identify appropriate candidates and states that "use of Selsentry is not recommended in patients with dual/mixed or CCR4-tropic HIV-1, as efficacy was not demonstrated in a phase II study of this patient group." The Trofile assay was used extensively in Pfizer's maraviroc phase III trials and measures the ability of the patient's specific virus envelope gene to effect entry into cells. This biomarker uses amplified RNA to establish a patient's HIV genome, followed by an assessment of that genome to infect CCR5- and CXCR4-expressing cell lines.

## HORSE BIOMARKERS

Factor Xa inhibitors are of therapeutic interest as antithrombotic agents because direct-acting antithrombin drugs often induce bleeding or deficits in fibrin [23–25]. By targeting a specific enzyme in the coagulation cascade, it is generally assumed that toxicity can be reduced and efficacy retained. Having an accurate indicator of bleeding risk is essential in this class of molecules since the major rate-limiting toxicities tend to be associated directly or indirectly with inhibited blood coagulation. During the development of factor Xa inhibitors, a logical biomarker of both safety and efficacy was available by measurement of Xa activity.

In fact, development and decision making with this class of drug were expedited by being able to monitor factor Xa activity in both preclinical effi-

cacy experiments and toxicology studies, as well as in early clinical trials. By adapting the automated human technique available in the clinical pathology coagulation laboratory to rodent and nonrodent species, direct comparisons were possible between the relative doses and plasma concentrations associated with therapeutic inhibition of clotting and exposures likely to cause unacceptably long coagulation times. At the discovery, preclinical safety, and phase I clinical stages, drug development teams had a tool that allowed rapid prioritization of molecules and decisions on dosing. Additionally, because reagents were available commercially at reasonable cost, blood volumes required were small, and instrumentation already existed for the method in clinical pathology laboratories, a reproducible means of determining a factor Xa inhibiting compound's pharmacodynamic characteristics was readily performed in a variety of research and hospital groups. This biomarker has been useful tool for developing factor Xa inhibitors, although caution must be exercised to assess compound activity beyond the results being a reflection of plasma drug concentration.

Cholesterol analysis isn't often considered very exotic or cutting edge, but has undergone significant evolution during its application in predicting cardiovascular risk. Total cholesterol has typically been measured using enzymatic, immunochemical, chemical, precipitation, ultracentrifugation, and column chromatography methods [26,27]. Since there can be significant differences in the values obtained for each of the lipoprotein classes using the various techniques, and to provide a standard for comparison, the Centers for Disease Control maintains reference methods for cholesterol, triglycerides, and high-density lipoproteins [28]. Reference methods considered the gold standards for cholesterol fractions have also been developed, validated, and credentialed [29,30]. As the relative roles and significance of very low-density lipoprotein (VLDL), high-density lipoprotein (HDL), and low-density lipoprotein (LDL) major lipid subgroups became better established in clinical practice, techniques for determining each cholesterol subcategory were developed and integrated into standard clinical laboratory use. Automation of these tests has made lipid analysis easily monitored in routine clinical practice. For many years, enzymatic and chemical methods have comprised the primary approaches to monitoring total cholesterol, triglycerides, and phospholipids. Precipitation is generally used for HDL and LDL, with ultracentrifugation considered the gold standard for assessment of new techniques [27,30]. As the fibrate and statin classes of drugs were integrated into standard clinical practice among cardiologists and general practioners, cholesterol monitoring became commonplace and continues as a leading indicator of heart disease risk.

Otvos et al. [31] published a nuclear magnetic resonance (NMR) approach to cholesterol monitoring that could be applied to clinically practical samples, as well as correlating the various lipoprotein subclasses to coronary artery disease [32–34], diabetes [35,36], and genetic polymorphisms having relevance in heart disease [37–39]. NMR has been an analytical technique since the 1950s, although its primary applications had been in research chemistry labo-

ratories, and more recently for metabonomic investigations. As cholesterol profiles and cardiovascular risk assessment have developed, many physicians have asked for increasingly detailed lipid evaluations. NMR profiling of serum lipids has demonstrated differences in the relative effects on various lipoprotein classes with available fibrate [40–42] and statin drugs [43–47].

A better understanding of particle-size distribution within lipid classes using NMR allows tailoring of patient therapeutic approaches beyond the broad goals of raising HDL and lowering LDL. Even among persons taking statin drugs, the best choice for one patient may be atorvastatin, whereas another person may achieve a more desirable lipid profile with rosuvastatin or simvastatin. This approach has been particularly useful in assessing changes in LDL particle numbers as well as total LDL content. The ability to select the optimal drug for each person provides an aspect of personalized medicine using well-accepted clinical pathology parameters determined by a technology with proven reliability in an arena outside the standard hospital laboratory, but readily accessible through specialty reference laboratories. The use of NMR to determine cholesterol and lipoprotein categories also has been useful in studying exercise [48,49], arthritis [50], and vascular responses [51].

## MULE BIOMARKERS

The anticancer drug trastuzumab (Herceptin) is a case study in stratification to identify patients most likely to respond to a drug using both conventional approaches and newer technologies. In 2007, trastuzumab achieved sales of $1.3 billion in the United States despite serious toxicity considerations and being efficacious in a limited subset of patients. Trastuzumab has primary efficacy in breast cancer patients who are overexpressing the human epidermal growth factor protein (HER2; [52–54]). The drug is a therapeutically engineered monoclonal antibody that targets the HER2 receptor protein [54]. The response rate to trastuzumab is as high as 35% in patients with overexpression of HER2, while the drug lacks a target and is ineffective at tolerated doses in patients who do not have increased HER2 levels. Pretreatment characterization of breast cancer patients to identify HER2-positive cancers and determine a patient's suitability for trastuzumab therapy is now commonly practiced. The serious adverse events that can develop following trastuzumab use (cardiac failure, pulmonary toxicity, infusion reactions) make it highly desirable to minimize exposure of those populations unlikely to benefit from the drug's therapeutic effects. The targeting of persons most likely to respond to the drug also has been a factor in reimbursement by third-party payers and acceptance by governmental health programs.

The primary and U.S. Food and Drug Administration (FDA)–approved approach for determining HER2 expression is immunohistochemistry for HER2 [55]. This "hybrid mule biomarker" combines the well-established but low-specificity technique of histology with the more recent development of

probe antibodies directed against specific proteins. Although neither of these modalities is particularly high tech or exotic, they provided the basis for developing a biomarker indicating HER2 over expression as molecular biology and proteomic techniques became available. These approaches quantified the protein resulting from the underlying gene amplification, leading to overexpression and identification and purification of the specific protein involved. As the reason for the progression to cancer became clearer, components of the overall process became suitable biomarker targets. This led, in turn, to the development of specific antibodies suitable for the identification of HER2 in tissue sections. Direct gene amplification and fluorescent in situ hybridization techniques for measuring HER2 expression also exist, but acceptance into clinical practice has been slower, due to their greater complexity and the limited number of laboratories capable of performing them in a diagnostic setting. Although not a perfect biomarker, HER2 expression is a valuable tool for the oncologist, and additional research is being conducted to refine the predictivity of this measure, either by alternative tests or by better characterization of key variables.

Osteoarthritis presents a difficult therapeutic area for pharmaceutical intervention. The progressive nature of the disease, generally late onset in life, and difficulty in reversing existing damage have made development of effective therapy challenging. Additionally, determining clinical efficacy often requires long-term clinical trial designs even in early drug development assessments where little is known about efficacy or side effects. These characteristics act as significant impediments for safe and rapid screening of new molecules. Clearly, this is an ideal place for a predictive biomarker that would allow a rapid assessment of whether a new drug had activity in the disease processes. Underlying the disease is destruction of joint cartilage by matrix metalloproteinases (MMP; [56,57]), so a mechanistic approach to identifying a biomarker was a logical approach to improving drug development in this area. As the role of collagen degradation and MMP activity became clearer in arthritis [58–61], interest focused on biomarkers that could be applied to animal models for both compound advancement and clinical monitoring as the ultimate endpoint.

Nemirovskiy et al. [62] reported development of a translatable biomarker of MMP activity with the goal of its application for MMP inhibitor compound selection and improved diagnosis of osteoarthritis. Similar approaches were under way by other groups within the pharmaceutical industry, both specific pharmaceutical companies and external contract research organizations. The approach taken by Nemirovskiy et al. [62] was to identify specific cleavage products using liquid chromatography–tandem mass spectrometry. By studying these MMP-derived peptides from human articular cartilage, they were able to show a 45-mer peptide fragment of collagen type II correlated with the pathology of human osteoarthritis and was present in urine and synovial fluid. An immunoaffinity liquid chromatography–mass spectrometry/mass spectrometry (LC-MS/MS) assay was developed to quantify collagen

type II neoepitope (TIINE) peptides as biomarkers of collagenase modulation. The resulting assay was capable of detecting TIINE peptides in the urine of healthy and afflicted human subjects and preclinical species (rat, rabbit, guinea pig, dog). This LC-MS/MS assay had excellent sensitivity, high throughput, reasonable costs, and robustness. By including immunoaffinity in the technique, a substantial improvement in assay sensitivity over traditional LC-MS/MS was achieved by eliminating much of the background noise associated with the sample matrix. ELISA methods also were developed to measure TIINE concentrations in urine, but proved to have higher intrasample variability, greater matrix sample effects, lower specificity for cleavage products, and to be more difficult to outsource to external laboratories. Although the ELISA technique remains useful as a research tool, the LC-MS/MS assay had significant advantages for clinical translation and implementation.

The TIINE biomarker was also applied to better characterize osteoarthritis in a surgically induced system using the Lewis rat [63]. This preclinical model has proven valuable to study progression and therapeutic intervention in degenerative joint disease. Using immunohistochemical staining of the joints, the authors were able to compare TIINE expression with proteoglycan loss and histological changes. This study showed that TIINE levels increased in intensity and area in lesions that co-localized with the loss of proteoglycan. From these data, Wancket et al. [63] were able to better define the medial meniscectomy surgical model of osteoarthritis, demonstrate a progressive pattern of cartilage damage similar to those seen in human lesions, and further characterize TIINE as a useful biomarker for monitoring cartilage degradation.

Clinically, TIINE has been used to evaluate the mechanism by which doxycycline slows joint space narrowing in patients with knee osteoarthritis. Urinary TIINE and radiographic determinations were conducted over a 30-month period [64]. Although the TIINE measurements were highly reproducible, the authors concluded that high visit-to-visit variability limits the sensitivity of the TIINE assay for detecting changes in clinical monitoring of osteoarthritis, and that increases in urinary TIINE concentration are unlikely to account for doxycycline reductions in joint space narrowing. The value of the TIINE biomarker for other collagen-based diseases remains to be determined. This biomarker does, however, highlight how even with mechanistic approaches, good preclinical correlations, and solid technology, a new technique may not yield a biomarker with strong clinical utility.

## CONCLUSIONS

The best biomarker for any given situation can come from a wide range of sources, so it is critical that no promising option be excluded. Matching the application and question to be answered are far more important that the platform used for analysis or if a technique is resident within a particular labo-

ratory or company. It should be recognized, however, that it is difficult to consider approaches that one has no idea exist. Nevertheless, those people providing biomarker support to drug discovery and development teams must keep in mind as wide a range of options as possible. The Internet has made specialized testing by reference laboratories, teaching hospitals, and research groups significantly more accessible. In addition, directories of tests and the groups that perform them are available [65], as well as characterized genomic biomarkers being described on the FDA Web site [66].

Genomic, proteomic, and metabonomic technologies can provide essential information when identifying new biomarkers, but have been slow to be implemented into clinical applications. Although often critical to identifying new targets or biomarker options, the extensive data sets produced, variability in sample and platform conditions, challenges of validating multiplexed measurements and algorithms, and lack of experience have limited their usefulness in clinical trials to a few diseases. The fields are rapidly progressing and hold great promise, especially when specific focused questions are defined prior to conducting the tests. To paraphrase Helmut Sterz, "use of a little grey matter at the beginning can save a lot of white powder, chips, instrumentation, and time." All too often, the quantity of information obtained from many "omics" experiments cannot be realized effectively due to limits on data mining tools and the realities of clinical trial conduct. People cannot be subjected to the same degree of environmental and genetic control possible with animal studies, and many diseases represent a constellation of effects rather than changes induced by a single cause or gene.

Our experience in developing, validating, translating, and implementing new biomarkers has emphasized repeatedly that the question to be answered must drive the technology used. It is also vital that the solution be "fit for purpose" with respect to the parameter being measured, platform selected, and level of assay definition or validation [67]. Sometimes the biomarker must utilize a cutting-edge technology and novel approaches, but more commonly the question can be answered without an exotic assay, often with a test that already exists in someone else's laboratory.

## REFERENCES


1. FDA (2008). Pharmacogenomics and its role in drug safety. *FDA Drug Saf Newsl*, 1(2). http://www.fda.gov/cder/dsn/2008_winter/pharmacogenomics.htm.
2. Nowell PC, Hungerford DAA (1960). Chromosome studies on normal and leukemic leukocytes. *J Nat Cancer Inst*, 25:85–109.
3. Rowley JD (1973). A new consistent chromosomal abnormality in chronic myelogenous leukemia identified by quinacrine fluorescence and Giemsa staining. *Nature*, 243:290–293.
4. Daley GQ, Van Etten RA, Baltimore D (1990). Induction of chronic myelogenous leukemia in mice by the $P210^{bcr/abl}$ gene of the Philadelphia chromosome. *Science*, 247:824–830.

5. Lugo TG, Pendergast AM, Juller AJ, Witte ON (1990). Tyrosine kinase activity and transformation potency of *bcr-abl* oncogenes products. *Science*, 247: 1079–1082.
6. Steelman LS, Pohnert SC, Shelton JG, Franklin RA, Bertrand FE, McCubrey JA (2004). JAK/STAT, Raf/MEK/ERK, PI3K/Akt and BCR-ABL in cell cycle progression and leukemogenesis. *Leukemia*, 18:189–218.
7. Hariharan IK, Harris AW, Crawford M, et al. (1989). A *bcr-abl* oncogene induces lymphomas in transgenic mice. *Mol Cell Biol*, 9:2798–2805.
8. Pfumio F, Izac B, Katz A, Shultz LD, Vainchenker W, Coulombel L (1996). Phenotype and function of human hematopoietic cells engrafting immune-deficient CB 17 severe combined immunodeficiency mice and nonobese diabetic-severe combined immunodeficient mice after transplantation of human cod blood mononuclear cells. *Blood*, 88:3731–3740.
9. Honda H, Oda H, Suzuki T, et al. (1998). Development of acute lymphoblastic leukemia and myeloproliferative disorder in transgenic mice expressing $P210^{bcr/abl}$: a novel transgenic model for human $Ph^1$-positive leukemias. *Blood*, 91:2067–2075.
10. Li S, Ilaria RJ, Milton RP, Daley DQ, Van Etten RA (1999). The P190, P210, and P230 forms of the *BCR/ABL* oncogene induce a similar chronic myeloid leukemia-like syndrome in mice but have different lymphoid leukemogenic activity. *J Exp Med*, 189:1399–1412.
11. Gorre ME, Mohammed M, Ellwood K, et al. (2001). Clinical resistance to STI-571 cancer therapy caused by *BCR-ABL* gene mutation or amplification. *Science*, 293:876–880.
12. Shah NP, Nicoll JM, Nagar B, et al. (2002). Multiple *BCR-ABL* kinase domain mutations confer polyclonal resistance to the tyrosine kinase inhibitor imatinib (STI571) in chronic phase and blast crisis chronic myeloid leukemia. *Cancer Cell*, 2:117–125.
13. Branford S, Rudzki Z, Walsh S, et al. (2003). Detection of *BCR-ABL* mutations in mutations in patients with CML treated with imantinib is virtually always accompanied by clinical resistance, and mutations in the ATP phosphate-binding loop (P-loop) are associated with poor prognosis. *Blood*, 102:276–283.
14. Shah NP, Tran C, Lee FY, Chen P, Norris D, Sawyers CL (2004). Overriding imatinib resistance with a novel *ABL* kinase inhibitor. *Science*, 305(5682): 399–401.
15. Klejman A, Schreiner SJ, Nieborowska-Skorska M, et al. (2002). The Src family kinase Hck couples *BCR/ABL* to STAT5 activation in myeloid leukemia cells. *EMBO J*, 21(21):5766–5774.
16. Hamilton A, Elrick L, Myssina S, et al. (2006). *BCR-ABL* activity and its response to drugs can be determined in $CD34^+$ CML stem cells by CrkL phosphorylation status using flow cytometry. *Leukemia*, 20(6):1035–1039.
17. Dean M, Carrington M, Winkler C, et al. (1996). Hemophilia Growth and Development Study, Multicenter AIDS Cohort Study, Multicenter Hemophilia Cohort Study, San Francisco City Cohort, ALIVE Study. Genetic restriction of HIV-1 infection and progression to AIDS by a deletion allele of the CKR5 structural gene. *Science*, 273:1856–1862.

18. Carrington M, Dean M, Martin MP, O'Brien SJ (1999). Genetics of HIV-1 infection: chemokine receptor CCR5 polymorphism and its consequences. *Hum Mol Genet*, 8(10):1939–1945.
19. Agrawal L, Lu X, Quingwen J, et al. (2004). Role for CCR5Δ32 protein in resistance to R5, R5X4, and X4 human immunodeficiency virus type 1 in primary CD4+ cells. *J Virol*, 78(5):2277–2287.
20. Saita Y, Kodama E, Orita M, et al. (2006). Structural basis for the interaction of CCR5 with a small molecule, functionally selective CR5 agonist. *J Immunol*, 117:3116–3122.
21. Pfizer, Inc. (2008). Selzentry label. http://www.fda.gov/cder/foi/label/2007/022128lbl.pdf.
22. Vandekerckhove L, Verhofstede C, Vogelaers D (2008). Maraviroc: integration of a new antiretroviral dug class into clinical practice. *J Antimicrob Chemother*, 61(6):1187–1190.
23. Rai R, Sprengeler PA, Elrod KC, Young WB (2001). Perspectives on factor Xa inhibition. *Curr Med Chem*, 8(2):101–119.
24. Ikeo M, Tarumi T, Nakabayashi T, Yoshida M, Naito S, Koike T (2006). Factor Xa inhibitors: new anti-thrombotic agents and their characteristics. *Front Biosci*, 11:232–248.
25. Crowther MA, Warkentin TE (2008). Bleeding risk and the management of bleeding complications in patients undergoing anticoagulant therapy: focus on new anticoagulant agents. *Blood*, 111(10):4871–4879.
26. Rifai N (1986). Lipoproteins and apolipoproteins: composition, metabolism, and association with coronary heart disease. *Arch Pathol Lab Med*, 110:694–701.
27. Bachorik PS, Denke MA, Stein EA, Rifkind BM (2001). *Lipids and dyslipoproteinemia.* In Henry JB (ed.), *Clinical Diagnosis and Management by Laboratory Methods*, 20th ed., W.B. Saunders, Philadelphia, pp. 224–248.
28. Myers GL, Cooper GR, Winn CL, Smith SJ (1989). The Centers for Disease Control–National Heart, Lung and Blood Institute Lipid Standardization Program: an approach to accurate and precise lipid measurements. *Clin Lab Med*, 9:105–135.
29. Myers GL, Cooper GR, Hassemer DJ, Kimberly MM (2000). Standardization of lipid and lipoprotein measurement. In Rifai N, Warnick GR, Dominiczak M (eds.), *Handbook of Lipoprotein Testing.* AACC Press, Washington, DC, pp. 717–748.
30. Rafai N, Warnick GR (2006). Lipids, lipoproteins, apolipoproteins, and other cardiovascular risk factors. In Burtis CA, Ashwood ER, Bruns DE (eds.), *Tietz Textbook of Clinical Chemistry and Molecular Diagnostics*, 4th ed., Elsevier Saunders, St. Louis, MO, pp. 903–981.
31. Otvos JD, Jeyarajah EJ, Bennett DW, Krauss RM (1992). Development of a proton NMR spectroscopic method for determining plasma lipoprotein concentrations and subspecies distribution from a single, rapid measure. *Clin Chem*, 38:1632–1638.
32. Freedman DS, Otvos JD, Jeyarajah EJ, Barboriak JJ, Anderson AJ, Walker J (1998). Relation of lipoprotein subclasses as measured by proton nuclear magnetic resonance spectroscopy to coronary artery disease. *Arterioscler Thromb Vasc Biol*, 18:1046–1053.

33. Kuller LH, Grandits G, Cohen JD, Neaton JD, Prineas R (2007). Lipoprotein particles, insulin, adiponectin, C-reactive protein and risk of coronary heart disease among men with metabolic syndrome. *Atherosclerosis*, 195:122–128.
34. van der Steeg WA, Holme I, Boekholdt SM, et al. (2008). High-density lipoprotein cholesterol, high-density lipoprotein particle size, and apolipoprotein A-1: significance for cardiovascular risk. *J Am Coll Cardiol*, 51:634–642.
35. MacLean PS, Bower JF, Vadlamudi S, Green T, Barakat HA (2000). Lipoprotein subpopulation distribution in lean, obese, and type 2 diabetic women: a comparison of African and White Americans. *Obes Res*, 8:62–70.
36. Berhanu P, Kipnes MS, Khan M, et al. (2006). Effects of pioglitazone on lipid and lipoprotein profiles in patients with type 2 diabetes and dyslipidaemia after treatment conversion from rosiglitazone while continuing stable statin therapy. *Diabetic Vasc Dis Res*, 3:39–44.
37. Couture P, Otvos JD, Cupples LA, et al. (2000). Association of the C-514T polymorphism in the hepatic lipase gene with variations in lipoprotein subclass profiles: the Framingham Offspring Study. *Arterioscler Thromb Vasc Biol*, 20:815–822.
38. Russo GT, Meigs JB, Cupples LA, et al. (2001). Association of the Sst-I polymorphism at the *APOC3* gene locus with variations in lipid levels, lipoprotein subclass profiles and coronary heart disease risk: the Framingham Offspring Study. *Atherosclerosis*, 158:173–181.
39. Humphries SE, Berglund L, Isasi CR, et al. (2002). Loci for CETP, LPL, LIPC, and APOC3 affect plasma lipoprotein size and sub-population distribution in Hispanic and non-Hispanic white subjects: the Columbia University BioMarkers Study. *Nutr Metab Cardiovasc Dis*, 12:163–172.
40. Ikewaki K, Tohyama J, Nakata Y, Wakikawa T, Kido T, Mochizuki S (2004). Fenofibrate effectively reduces remnants, and small dense LDL, and increases HDL particle number in hypertriglyceridemic men: a nuclear magnetic resonance study. *J Atheroscler Thromb*, 11:278–285.
41. Ikewaki K, Noma K, Tohyama J, Kido T, Mochizuki S (2005). Effects of bezafibrate on lipoprotein subclasses and inflammatory markers in patients with hypertriglyceridemia: a nuclear magnetic resonance study. *Int J Cardiol*, 101:441–447.
42. Otvos JD, Collins D, Freedman DS, et al. (2006). LDL and HDL particle subclasses predict coronary events and are changed favorably by gemfibrozil therapy in the Veterans Affairs HDL Intervention Trial (VA-HIT). *Circulation*, 113:1556–1563.
43. Rosenson RS, Shalaurova I, Freedman DS, Otvos JD (2002). Effects of pravastatin treatment on lipoprotein subclass profiles and particle size in the PLACI trial. *Atherosclerosis*, 160:41–48.
44. Schaefer EJ, McNamara JR, Taylor T, et al. (2002). Effects of atorvastatin on fasting and postprandial lipoprotein subclasses in coronary heart disease patients versus control subjects. *Am J Cardiol*, 90:689–696.
45. Blake GJ, Albert MA, Rifai N, Ridker PM (2003). Effect of pravastatin on LDL particle concentration as determined by NMR spectroscopy: a substudy of a randomized placebo controlled trial. *Eur Heart J*, 24:1843–1847.
46. Soedamah SS, Colhoun HM, Thomason MJ, et al. (2003). The effect of atorvastatin on serum lipids, lipoproteins and NMR spectroscopy defined lipoprotein sub-

classes in type 2 diabetic patients with ischemic heart disease. *Atherosclerosis*, 167:243–255.

47. Schaefer EJ, McNamara JR, Tayler T, et al. (2004). Comparisons of effects of statins (atorvastatin, fluvastatin, lovastatin, pravastatin, and simvastatin) on fasting and postprandial lipoproteins in patients with coronary heart disease versus control subjects. *Am J Cardiol*, 93:31–39.
48. Nicklas BJ, Ryan AS, Katzel LI (1999). Lipoprotein subfractions in women athletes: effects of age, visceral obesity and aerobic fitness. *Int J Obes Rel Metab Disord*, 23:41–47.
49. Yu HH, Ginsburg GS, O'Toole ML, Otvos JD, Douglas PS, Rifai N (1999). Acute changes in serum lipids and lipoprotein subclasses in triathletes as assessed by proton nuclear magnetic resonance spectroscopy. *Arterioscler Thromb Vasc Biol*, 19:1945–1949.
50. Hurt-Camejo E, Paredes S, Masana L, et al. (2001). Elevated levels of small, low-density lipoprotein with high affinity for arterial matrix components in patients with Rheumatoid Arthritis. *Arthritis Rheum*, 44:2761–2767.
51. Stein JH, Merwood MA, Bellehumeur JL, et al. (2004). Effects of pravastatin on lipoproteins and endothelial function inpatients receiving human immunodeficiency virus protease inhibitors. *Am Heart J*, 147:E18.
52. Hudelist G, Kostler W, Gschwantler-Kaulich D, et al. (2003). Serum EGFR levels and efficacy of trastuzumab-based therapy in patients with metastatic breast cancer. *Eur J Cancer*, 42(2):186–192.
53. Smith BL, Chin D, Maltzman W, Crosby K, Hortobagyi GN, Bacus SS (2004). The efficacy of Herceptin therapies is influenced by the expression of other *erbB* receptors, their ligands and the activation of downstream signalling proteins. *Br J Cancer*, 91:1190–1194.
54. Burstein HJ (2005). The distinctive nature of *HER2*-positive breast cancer. *N Engl J Med*, 353:1652–1654.
55. Genentech (2008). Herceptin label. http://www.fda.gov/cber/foi/label/2008/103792s517lbl.pdf.
56. Shinmei M, Masuda K, Kikuchi T, Shimomura Y, Okada Y (1991). Production of cytokines by chondrocytes and its role in proteoglycan degradation. *J Rheumatol Suppl*, 27:89–91.
57. Okada Y, Shinmei M, Tanaka O, et al. (1992). Localization of matrix metalloproteinase 3 (stromelysin) in osteoarthritic cartilage and synovium. *Lab Invest*, 66(6):680–690.
58. Billinghurst RC, Dahlberg L, Ionescu M, et al. (1997). Enhanced cleavage of type II collagen by collagenases in osteoarthritic articular cartilage. *J Clin Invest*, 99:1534–1545.
59. Huebner JL, Otterness IG, Freund EM, Caterson B, Kraus VB (1998). Collagenase 1 and collagenase 3 expression in a guinea pig model of osteoarthritis. *Arthritis Rheum*, 41:877–890.
60. Dahlberg L, Billinghurst RC, Manner P, et al. (2000). Selective enhancement of collagenase-mediated cleavage of resident type II collagen in cultured osteoarthritic cartilage and arrest with a synthetic inhibitor that spares collagenase 1 (matrix metalloproteinase 1). *Arthritis Rheum*, 43:673–682.

61. Wu W, Billinghurst RC, Pidoux I, et al. (2002). Sites of collagenase cleavage and denaturation of type II collagen in aging and osteoarthritic articular cartilage and their relationship to the distribution of matrix metalloproteinase 1 and matrix metalloproteinase 13. *Arthritis Rheum*, 46:2087–2094.
62. Nemirovskiy OV, Dufield DR, Sunyer T, Aggarwal P, Welsch DJ, Mathews WR (2007). Discovery and development of a type II collagen neoepitope (TIINE) biomarker for matrix metalloproteinase activity: from in vitro to in vivo. *Anal Biochem*, 361(1):93–101.
63. Wancket LM, Baragi V, Bove S, Kilgore K, Korytko PJ, Guzman RE (2005). Anatomical localization of cartilage degradation markers in a surgically induced rat osteoarthritis model. *Toxicol Pathol*, 33(4):484–489.
64. Otterness IG, Brandt KD, Le Graverand MP, Mazzuca SA (2007). Urinary TIINE concentrations in a randomized controlled trial of doxycycline in knee arthritis: Implications of the lack of association between TIINE levels and joint space narrowing. *Arthritis Rheum*, 56(11):3644–3649.
65. Hicks JM, Young DS (eds.) (2005). *DORA 05-07: Directory of Rare Analyses* AACC Press, Washington, DC.
66. FDA (2008). Table of validated genomic biomarkers in the context of approved drug labels. Center for Drug Evaluation and Research. http://www.fda.gov/cder/genomics/genomic_biomarkers_table.htm (updated Sept. 10, 2008).
67. Lee JW, Devanarayan V, Barret YC, et al. (2006). Fit-for-purpose method development and validation of biomarker measurement. *Pharm Res*, 23(2):312–328.

# BIOMARKER PATENT STRATEGIES: OPPORTUNITIES AND RISKS

Cynthia M. Bott, Ph.D.
*Honigman Miller Schwartz and Cohn LLP, Ann Arbor, Michigan*

Eric J. Baude, Ph.D.*
*Brinks Hofer Gilson & Lione, P.C., Ann Arbor, Michigan*

## INTRODUCTION

Biomarkers are used in scientific research to identify therapeutic agents and to measure the performance of a drug before and after regulatory approval. It is possible and often desirable, for a company, institution, or individual to patent a biomarker for its use in those settings. In general, as the patentee the company, institution, or individual can sue anyone who infringes the patent. If the patentee prevails in a lawsuit, the patentee may receive money damages and prevent the infringer from using the biomarker. It is equally true that others can obtain patents on biomarkers. Therefore, the use of a biomarker by a company, institution, or individual creates a risk of patent infringement for the user of the patented biomarker. Accordingly, it is important to identify patent infringement risks associated with the making and using of biomarkers. These two aspects, patenting biomarkers and identifying patent infringement risks, are discussed in this chapter.

*The views expressed herein are those of the authors and do not necessarily reflect those of Brinks Hofer Gilson & Lione, P.C., Honigman Miller Schwartz and Cohn LLP, or their clients.

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

## PATENTING BIOMARKERS

### Introduction to Patents

In exchange for fully describing an invention, a U.S. patent gives inventors the right "to exclude others from making, using, offering for sale, or selling their invention in the United States or importing the patented invention throughout the United States or importing their invention into the United States" for a limited period of time.[1] The U.S. Constitution in Article I, section 8, clause 8, provides the power under which Congress enacts patent laws. The original rationale for the provision of patents remains today: to promote the disclosure of ideas to the general public.

Patents are granted by most countries of the world and the process and requirements for getting a patent in any of these countries are essentially the same. The process for obtaining a patent on an invention begins with the inventor submitting a patent application to a government patent office, such as the U.S. Patent and Trademark Office (USPTO). After the inventor files a patent application at the USPTO, a patent examiner reviews the application and determines whether the application meets the requirements for patentability. This process is referred to as *patent application examination*.

### Patentability Requirements

The requirements for the patentability of inventions in the United States are set out in the U.S. Code, Title 35. Four of the major claims for an invention are:

1. *Utility.* To qualify for a patent, the invention must have utility; it must be useful for some purpose.[2] The invention must have a specific, substantial, and credible use (i.e., it must have a real-world use). The USPTO has promulgated utility guidelines around what is required to meet the standard of a credible, specific, and substantial utility.[3] Essentially, a biomarker must generally have an associated biological function that is demonstrated experimentally or by analogy to existing biomarkers. In general, this requirement is easily met.
2. *Novelty.* A patent claim is novel if it is different from the prior art in any aspect.[4] An examiner in the USPTO analyzes the prior art identified in a search to determine if the invention claimed in the application is different from what was known previously. If the prior art falls within the scope of a claim, the examiner will reject the claim as not being novel. Examples of prior art include patents, patent applications, Web site disclosures, public talks, scientific papers, and scientific posters presented at meetings. Therefore, it is important to file a patent application before speaking publicly about or publishing any aspect of the invention. If an invention is disclosed publicly before applying for a patent, it is not

possible to obtain a defensible patent on the invention in most countries outside the United States. In the United States, an inventor may be able to obtain a U.S. patent provided that the public disclosure is less than one year before the patent application is filed. This one-year period is referred to as a grace period.

3. *Nonobviousness.* An invention is considered nonobvious if someone in the field of technical expertise would have viewed the invention as surprising or unexpected in view of the prior art.[5] This requirement is in place to prevent patents from being granted to obvious improvements over the prior art.[6] Patent applications are often rejected as being obvious over the prior art because many inventions are combinations of what is already in the public domain. It is often the case that those rejections can be overcome by arguing that the invention offers surprising advantages over the prior art. A skilled patent attorney can offer advice on how to address potential obviousness concerns around an invention before the application is filed.
4. *Enablement.* A patent application must fully describe how to make and use the invention so that a person of ordinary skill in the technology can carry out the invention without undue experimentation.[7] In part, the enablement requirement is met from the application itself, which contains a detailed description of how to carry out the invention. In addition, examples of how actual experiments were carried out are provided in the patent application to help comply with the enablement requirement. Together, the descriptions of the invention and the examples often serve to satisfy the enablement requirement.

## The Patenting Process

The process of getting a patent granted is quite long, generally taking from three to four years and often much longer for inventions in certain technology areas. The starting point for this process is the filing of a patent application with a government patent office such as the USPTO. The patent application must include a description of the invention and one or more claims that clearly define the invention. After the patent application is filed, the invention as set forth in the patent application is reviewed by an examiner at the UPSTO to determine if it meets the requirements for patentability. To determine the patentability of an invention, the examiner performs a search to identify prior art that is relevant to novelty, obviousness, and other patentability requirements in reference to the claims of the application. If the examiner identifies prior art that renders one or more of the claims not novel or obvious, the USPTO will send the patent applicant a document termed an *office action*. Office actions set out the reasons that a claim was rejected over the prior art. The patent applicant can then respond to the USPTO with reasons why the office action is in error. The exchange of office actions and responses between the applicant and the USPTO can continue until the application is allowed by

the USPTO, or the applicant decides to abandon the application. One of the ways in which applicants can overcome the USPTO's rejections is to narrow the scope of the claims by amending the claims. Often, the claims can be amended in a manner that still provides valuable coverage for the applicant's invention and satisfies the USPTO's requirements for patentability.

Patents are territorial in scope, meaning that a patent is enforceable only in the country in which it is granted. For example, the owner of a U.S. patent cannot sue in a German court for infringement of the U.S. patent. The owner of the U.S. patent would have to obtain a German patent as well to provide protection in Germany against potential infringers. Therefore, an inventor must get a patent in every country in which protection of the invention is sought.

A process exists to simplify the filing applications in most countries. An international patent law treaty known as the Patent Cooperation Treaty (PCT) provides a unified procedure for filing a patent application to protect inventions in the 138 countries that are parties to the treaty.[8] A patent application filed under the PCT is referred to as an international patent application (or simply a PCT application). In addition to providing a unified procedure, a PCT application allows the inventor to delay the filing of patents in each of the desired countries, thereby delaying the outlay of filing fees and translation fees. After filing the PCT application, the process is much like the initial part of prosecution in the United States. An examiner of a PCT application performs a search for prior art and provides an opinion on patentability. PCT applications are published 18 months after the earlier of the filing date or priority date. The PCT application itself, however, does not result in a patent. In each country in which patent protection is desired, a county-specific patent based on the PCT application must be filed. Filing in a specific country is referred to as *entry* into the national or regional phase. National phase entry must occur at or before 30 months from the earlier of the filing date or priority date of the international application.

In most countries, the term of a patent is 20 years less the time from the date that the patent application was filed. For example, if a patent is issued four years after it was filed, the patent would expire in 16 years. The patent owner is responsible for enforcing the patent. To do so the owner files a patent infringement lawsuit in a U.S. federal district court to sue the alleged infringer.

## Examples of How to Protect a Biomarker

Examples of potentially patentable biomarkers are myriad and include DNA sequences with known function, such as genes, gene fragments or control elements, deleterious mutations, functional proteins, hormones, drug metabolites, peptides, polymorphisms such as single-nucleotide polymorphisms, and cellular antigens. Provided that patentability requirements are met, an inventor may be able to secure patent protection for a biomarker in several ways. Imagine the discovery that protein X is expressed in patients with pancreatic

cancer but not in healthy subjects. Protein X might be useful as a biomarker for the detection and diagnosis of the pancreatic cancer or for determining if a patient is responding to drug therapy. The discovery of such a biomarker would be considered an invention. The ways in which the person making this discovery can protect his or her invention depend on several factors.

If a biomarker was not known at the time of the invention, the strongest protection is generally provided by a patent covering the biomarker itself. In the example above, the biomarker would be protein X and/or the DNA sequence encoding the protein. The strongest protection of protein X would be a patent that specifically claims protein X. Such a patent is often referred to as a *composition of matter patent* because of the way that patentable inventions are defined in the patent law. A patent to the biomarker or the DNA sequence could be enforced to prevent others from making, using, offering for sale, or selling the patented biomarker.[9]

If a biomarker is already known, the biomarker itself would not be patentable because it is not novel.[10] Patent protection may still be available for new uses of that biomarker. For example, if protein X in the example above was known but its correlation with pancreatic cancer was not known, the inventor might be able to obtain patent coverage for methods of using the biomarker such as methods of diagnosing pancreatic cancer, assays for measuring the level of the protein in a pancreatic cancer patient, or methods of identifying the effectiveness of chemotherapeutic treatment against pancreatic cancer.

Another consideration is whether in a particular situation it makes sense to patent the biomarker. If the inventor is interested in excluding others from using the biomarker, either for competitive reasons or to generate income from licensing, it may be well worth the effort to patent the biomarker. If there is no intention to exploit a patent, it may make more sense to disclose the invention by publication. This disclosure would probably prevent someone else from subsequently patenting that invention.

## IDENTIFYING BIOMARKER PATENT INFRINGEMENT RISKS

Patents give the patent owner the right to exclude others from practicing the patented invention. The patent, however, does not provide the patent owner with an affirmative right to practice the invention. For example, the owner of a patent that claims a method of using a biomarker can exclude others from the use of that biomarker. If another party owns a composition of matter patent on the biomarker (i.e., a patent to the biomarker itself), the owner of the method of use patent would not be able to make or use the biomarker for any purpose without the permission of the composition of matter patent owner. This second patent is said to *dominate* the biomarker use patent. These patents are also referred to as *blocking patents*. If a blocking patent exists, a person or company could be sued for using the biomarker without permission from the owner of the biomarker composition of matter patent. It is important

to understand this distinction in order to avoid the feeling that possession of a patent on an invention protects the patentee from a charge of infringement when using that invention.

It is possible to research if a planned use of a biomarker infringes someone else's patent. This process is known as determining if there is *freedom to operate* around the use of the biomarker, sometimes referred to as *freedom to practice*. To determine if there are any patents that pose a risk of infringement, a laboratory scientist or drug development team first needs to search for patents that would cover the making and use of the biomarker. Typically, a fee-based patent search firm is employed to use text-based searching and sequence-based searching to identify potentially relevant patents. It is also possible to conduct some initial searching on one's own. There are currently several free Internet-based public databases that are available for searching patents such as the USPTO database available at http://www.uspto.gov, http://www.patentlens.net, http://www.google.com/patents, and http://www.freepatentsonline.com.

The next step is to interpret if any of the potentially relevant patents identified in the search pose a significant risk of patent infringement to the planned activities. The layperson or scientist may find the scope of patent claims to be difficult to understand. Often, a patent specialist such as a patent attorney is needed to review the search results and provide advice on potential infringement concerns. If there are any potential infringement issues, the scientist and the patent attorney can work together to address those issues. Potential ways to reduce infringement risks include licensing the problematic patent from the owner, designing around the patent, or deciding not to use the biomarker. The potential availability of certain exceptions and exemptions to infringement is also discussed below.

## Licensing a Patent

If there is concern that a planned use might infringe a patent, one way to address this issue is to obtain a license from the patent owner that would allow the licensee to practice the patented invention. Different types of licenses are common in the biomarker arena. The license may involve an up-front payment with no further obligation to the patentee. In other circumstances, during the license term the license terms might require payments such as royalty payments or payments when certain development milestones are met. For example, a license to a diagnostic kit might include a $20,000 payment upon signing the license and a 2% royalty paid annually based on that year's net sales from the diagnostic kit.

## Designing Around a Patent

For cases in which a license is not available or the terms of the license would not be reasonable, another option is to design around the claims of the patent.

A typical design around strategy involves designing a biomarker in a way that is outside the scope of the blocking patent. A skilled patent attorney can examine the scope of the claims and advise if there is a potential design-around. For example, a patent may claim only the full-length sequence of a protein or a nucleic acid biomarker. If the full-length sequence is not necessary for the use being considered, use of a modified sequence that is shorter than the full-length sequence may not infringe the problematic patent claim. Successful design-around strategies can realize cost savings in research and development costs, legal fees, and potential litigation costs, as well as minimize the delay in commercializing a product or method. Designing around a patent is one of the ways in which the patent system works to promote new advances in technology.

Another design-around strategy is sometimes called a geographic design-around. As noted above, patents are territorial and can only be enforced in the countries in which the invention is patented. Therefore, it may be possible to make and use a specific biomarker in a country where the patentee has not yet obtained patent protection or will not be obtaining patent protection. For example, it might be that a desired biomarker has been patented in the United States but not in Canada. If use of the biomarker can be limited to Canada (or to any other country where the invention is not patented), there would not be a need to license that biomarker.

### Ceasing Use of the Biomarker

One option to address a significant infringement risk is to stop using the biomarker. This may not be a preferred option, but in certain cases it may be the only appropriate course of action.

### Exemption from Infringement Under 35 U.S.C. § 271(e)(1)

Under 35 U.S.C. § 271(e)(1), the use of a biomarker to discover a drug or in clinical trials might be exempted from infringement if the use is "reasonably related to the development" of information that might be submitted to support U.S. Food and Drug Administration (FDA) approval. Although the boundaries of the exemption are not completely clear, the courts have exempted the use of a patented compound under 35 U.S.C. § 271(e)(1). The courts, however, have not addressed directly whether the use of a research tool such as a biomarker in drug development would fall under the 35 U.S.C. § 271(e)(1) exemption. Thus, the boundaries of the exemption are somewhat uncertain with regard to research tools. Once the drug is approved, however, further use of a patented biomarker will generally not be exempted from infringement.

### Experimental Use Exception

For private institutions such as companies and private universities, there is generally no immunity from a patent lawsuit under an experimental use excep-

tion.[11] This was made clear in the private university arena in the *Madey* v. *Duke* case in 2002. The federal circuit court held that the "very narrow and strictly limited experimental use defense" was applicable only if a patented invention was used "solely for amusement, to satisfy idle curiosity, or for strictly philosophical inquiry."[12] Duke's use of a patented invention was related to obtaining research funding which was in furtherance of their legitimate business. It would seem, then, to be the rare case where even a private university would not be subject to a patent lawsuit if its activities relate to obtaining research grants. Accordingly, the research use exception in the United States does not provide much protection from a patent lawsuit.

### Immunity Under the 11th Amendment

In general, a state institution such as a state university may not be sued for patent infringement. The 11th amendment of the Constitution prevents a private party from suing a state university for patent infringement.[13] Thus, an individual or a for-profit corporation would not be able to sue a public university for use of a patented biomarker. The state immunity does not transfer to a licensee of a patent from a state university. Therefore, if such a biomarker patent is licensed from a state university, the licensee could be subject to a patent lawsuit from the holder of the biomarker patent even though the university would not.

## CONCLUSIONS

The discovery and use of biomarkers presents opportunities and risks in the patent area. There are opportunities to patent biomarkers and methods of using biomarkers to further commercial ventures. There is also the risk of infringing third-party patents when making or using certain biomarkers. When discovering a biomarker or a new use of a biomarker, it is important to consider capturing biomarker patent opportunities. At the same time, it is important to be conscious that the use of a biomarker can pose risks of infringement and seek ways to minimize those risks.

The top three questions to keep in mind when working with biomarkers are:

1. Is it desirable to patent the biomarker invention?
2. Is the invention patentable?
   a. Is the biomarker unknown?
   b. Is the biomarker known but the use is not?
   c. Does the invention meet the criteria for patentability?

3. Are there risks to making and using a biomarker?
   a. Is the use "reasonably related to the development" of information that might be submitted to support FDA approval and therefore exempt from infringement?
   b. Is there a patent that could block the making and using of the biomarker?

## NOTES AND CITATIONS

1. See 35 U.S.C. § 271(a).
2. The requirement that an invention be useful is set forth in 35 U.S.C. § 101.
3. See Revised Interim Utility Guidelines Training Materials at http://www.uspto.gov/web/menu/utility.pdf and the USPTO Utility Guidelines at http://www.uspto.gov/web/offices/com/sol/notices/utilexmguide.pdf.
4. The requirements for determining whether or not an invention is new are set forth in 35 U.S.C § 102.
5. The requirements for determining obviousness are set forth in 35 U.S.C. § 103(a).
6. In Europe and Japan a similar requirement for nonobviousness is that of "inventive step."
7. The requirement enablement is set forth in 35 U.S.C. § 112, first paragraph.
8. As of December 2007, 138 countries are parties to the PCT. Separate patent applications must be filed in countries that are not signatories to the PCT to obtain patent protection in those countries. For example, Argentina and Taiwan are not members of the PCT.
9. See 35 U.S.C. § 271(a).
10. See novelty provisions at 35 U.S.C. §§ 102(a)–(g).
11. See *Madey v. Duke University*, 307 F.3d 1351 (Fed. Cir. 2002).
12. Ibid., 1362.
13. See *Fla. Prepaid Postsecondary Educ. Expense Bd. v. Coll. Sav. Bank*, 527 U.S. 627 (1999).

# IT SUPPORTING BIOMARKER-ENABLED DRUG DEVELOPMENT

Michael Hehenberger, Ph.D.
*IBM Healthcare & Life Sciences, Somers, New York*

## PARADIGM SHIFT IN BIOPHARMACEUTICAL R&D

The biopharmaceutical industry is currently undergoing a transformation driven primarily by the need to move from its proven blockbuster model to a new stratified medicine or nichebuster model [1]. This paradigm shift has been anticipated in white papers such as IBM's Pharma 2010 report [2] and will have to be accompanied by serious efforts to streamline operations and to address research and development (R&D) productivity issues as described and quantified by DiMasi [3] and DiMasi et al. [4]. In its Critical Path initiative [5], the U.S. Food and Drug Administration (FDA) is guiding industry toward use of biomarkers [6] that will address efficacy and safety issues and hold the promise of increased R&D productivity [7]. In the excellent biomarker review paper by Trusheim et al. [1], *clinical biomarkers* are defined as biomarkers that associate a medical treatment to a patient subpopulation that has historically exhibited a differential and substantial clinical response. Clinical biomarkers can be based on "genotypes, proteins, metabonomic patterns, histology, medical imaging, physician clinical observations, or even self-reported patient surveys. A clinical biomarker is not defined by its technology or biological basis, but rather by its reliable, predictive

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

correlation to differential patient responses." It is generally believed that "biomarker-enabled" drug development will lead to better and earlier decision making and that clinical biomarkers will pave the way toward targeted therapeutics combining "precision drugs" for stratified patient populations with diagnostic tests designed to identify not only "responders" (who will benefit) but also patient cohorts at risk for adverse side effects. Biomarker-enabled R&D is evolving into a new discipline with a strong patient focus. Organizations who believe in biomarker-enabled R&D are investing in tools and make the necessary organizational changes to implement the new concepts and associated processes. Among biomarkers, *imaging biomarkers* have received particular attention because of the noninvasive nature of imaging technologies and the obvious link to diagnostic procedures and clinical care. Imaging technologies are used increasingly as core technologies in biopharmaceutical R&D in both the preclinical and clinical phases of the R&D process.

Disease areas most affected by this paradigm shift are cardiology, oncology, and neurology. Below we discuss in more detail how biomarker-related data types and their increasing volumes are challenging existing information technology (IT) infrastructures, and how IT architectures have to be enhanced and modified to integrate genomic, imaging, and other biomarker data.

## PROCESSES, WORKFLOWS, IT STANDARDS, AND ARCHITECTURES

The conventional R&D process—documented extensively by most research-based biopharmaceutical companies—is sequential (Figure 1). After target identification and validation by the "biologists," the medicinal "chemists" take over and screen extensive libraries of hundreds of thousands of chemical compounds against the target of eventually finding a suitable candidate for a drug. After the IND (initial new drug) application to the FDA, the drug candidate is tested in preclinical animal studies before it is handed over to the clinical development organization for clinical trials that proceed through three phases. If the drug candidate survives through phase III, all the supporting information collected will be submitted to the FDA in form of a new drug application (NDA) dossier that has to be compiled such that FDA's rules and regulations are respected and followed. After FDA approval, phase IV trials may be conducted to collect postmarketing surveillance data about adverse drug reactions or to position the drug for new indications not yet covered by a given FDA approval.

To manage the process effectively and to terminate failed projects as early as possible, R&D organizations have created a set of disciplined processes designed to optimize project portfolios and to track the progress of individual projects by *milestones* or *decision gates* (Figure 2). Industry leaders such as

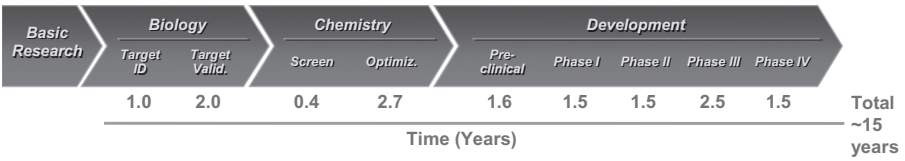


**Figure 1** Sequential R&D process. (From Peter Corr, IBM Imaging Biomarkers Summit I, Palisades, New York, December 15–17 2007.)

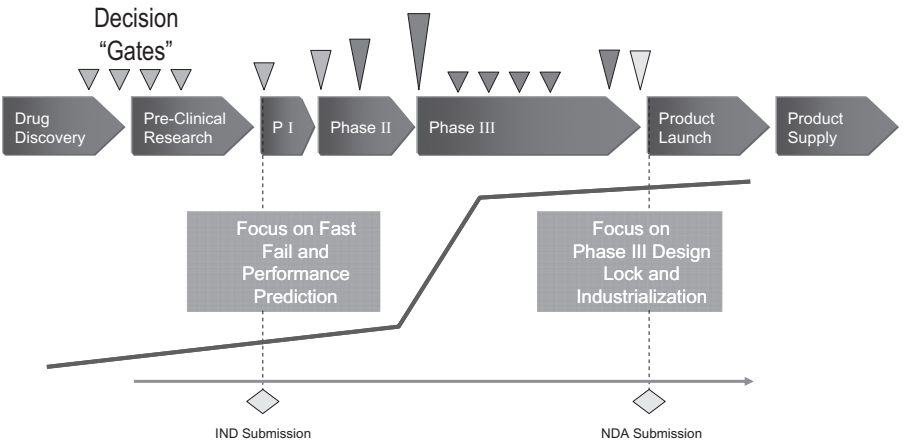


**Figure 2** Decision gates (milestones) to manage sequential R&D processes. (From Terry McCormick, Kathleen Martin, and M. Hehenberger, IBM Institute for Business Value, The evolving role of biomarkers: focusing on patients from research to clinical practice, July 2007, http://www.ibm.com/industries/healthcare/doc/jsp/resource/insight/.)

Novartis have realized that biomarker-enabled R&D has to be organized differently (personal communication from Werner Kroll, January 2007). In particular, there is an increasing emphasis on parallel processes and accelerated "proof-of-concept" in humans to "learn" quickly and to "confirm" (i.e., conduct extensive clinical trials) only if the learning is yielding promising results (Figure 3). A high-level depiction of this new approach can also be found in the IBM Pharma 2010 report [2].

It is the role of IT standards and architectures to support business strategies and to enable their implementation. IT standards relevant in this context are:

- Data standards proposed by the Clinical Data Interchange Standards Consortium (CDISC) [8] such as:
  - SEND (Standard for Exchange of Nonclinical Data) covering animal data

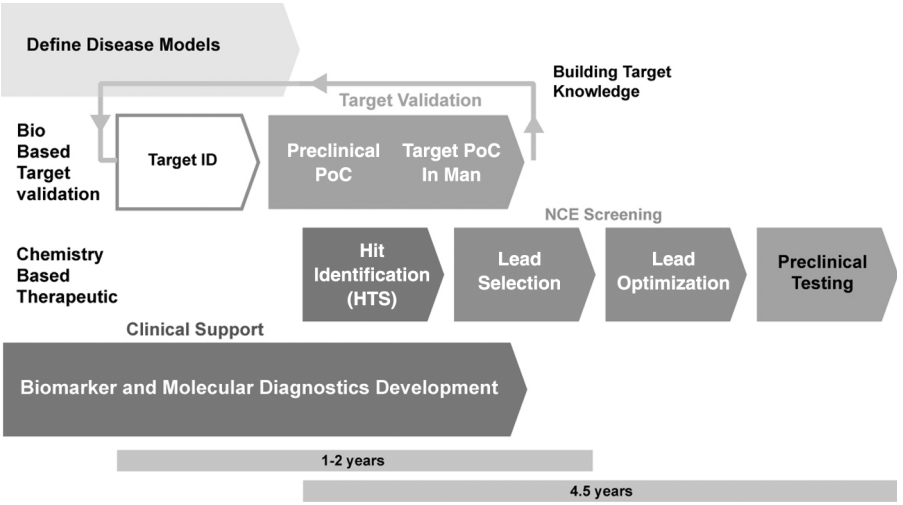


**Figure 3** Parallel biomarker-enabled processes up to preclinical development. (From Terry McCormick, Kathleen Martin, and M. Hehenberger, IBM Institute for Business Value, The evolving role of biomarkers: focusing on patients from research to clinical practice, July 2007, http://www.ibm.com/industries/healthcare/doc/jsp/resource/insight/.)

- SDTM (Study Data Tabulation Model) covering human data
- ODM (Operational Data Model) covering study data including EDC-generated data
- ADaM (Analysis Data Model) covering analysis data sets
- HL7 Clinical Document Architecture (CDA) [9]
- DICOM (Digital Imaging and Communications in Medicine) to transmit medical images [10]
- Janus Data Model [11] developed by FDA and IBM [in a collaborative research and development agreement (CRADA)] and implemented by NCI (National Cancer Institute) and the FDA

A comprehensive overview of relevant standards along with associated Web sites is shown in Figure 4.

Based on the technical challenges for integrating a diverse set of data sources for a biomarker-based clinical data submission, we propose an IT architecture (Figure 5 [12]) that addresses a majority of those requirements. Although this architecture includes software products and assets belonging to IBM, one can logically extend it to fit other vendor products as well. Our approach is to present a general-purpose platform for managing clinical submissions of patient data, enhanced with genomic and imaging data.

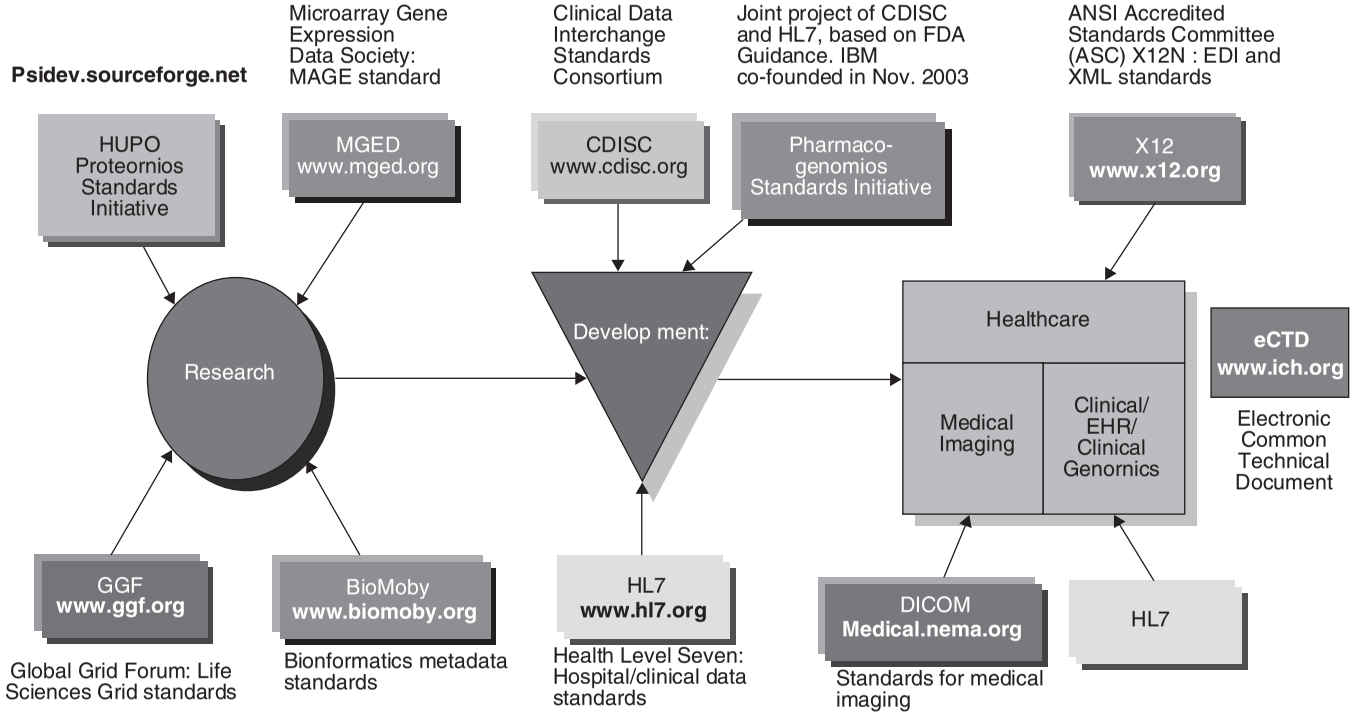


**Figure 4** Health care and life sciences standards organizations.

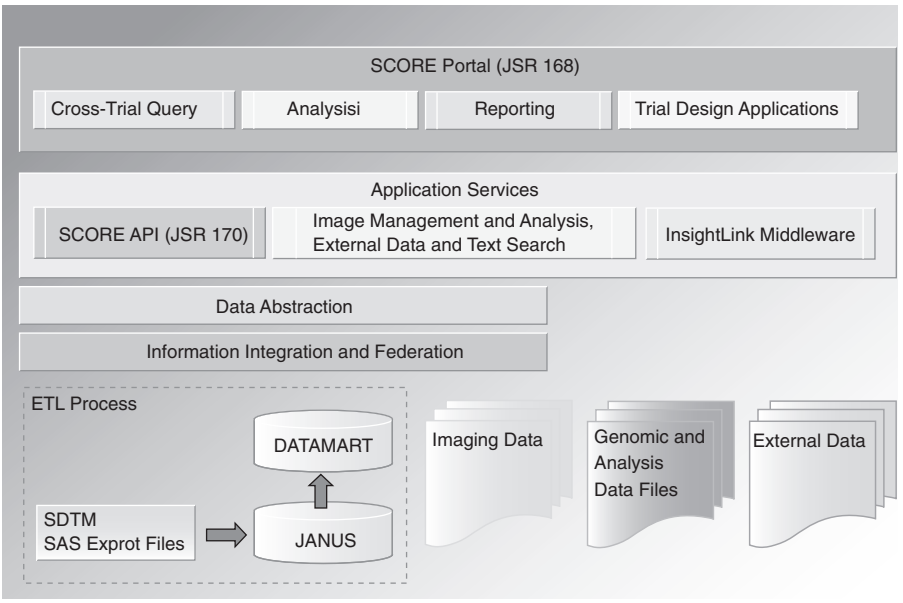


**Figure 5** Proposed IT architecture for biomarker-based clinical development. (From Terry McCormick, Kathleen Martin, and M. Hehenberger, IBM Institute for Business Value, The evolving role of biomarkers: focusing on patients from research to clinical practice, July 2007, http://www.ibm.com/industries/healthcare/doc/jsp/resource/insight/.)

At the bottom data layer, summarized clinical submission data in CDISC's SDTM format receive feeds from a clinical data management system that stores case report forms. First, the associated metadata for the SDTM submission (such as the vocabularies used for the adverse event codes, lab codes, etc.) are mapped into the tables of the Janus data model. An extract-transform-load tool is then used to improve the scalability and data validation, cleansing, and transformation needed to load the SDTM data into Janus. One may also need to build a collection of applications and use-case specific data-marts on top of Janus. These can be designed using Star-Schema-based dimensional models for optimal query performance. In addition to the clinical submission data in Janus, one would need to establish links to the imaging data that often reside in picture archiving and communications Systems. After extraction, the images can be managed centrally using a standardized imaging broker service (such as provided by IBM's content management offering) and stored along with genomic raw and analysis files in a content management repository supported by IBM's Solution for Compliance in a Regulated Environment (SCORE [13]). Finally, external reference databases such as PubMed, GenBank, dbSNP, and SwissProt are integrated using unstructured

information management technology provided, for example, by IBM's WebSphere Information Server/OmniFind middleware. All these content stores can be searched dynamically using a federated warehouse that uses a wrapper-based technology for linking diverse data sources. On top of the federation layer, we propose a data abstraction and query layer powered by the Data Discovery Query Builder (DDQB). DDQB has been designed to expose a user-centric logical data model (based on XML) mapped on top of the physical data model.

To support processes and workflows that may include the transfer of clinical biomarker data between pharmas, contract research organizations (CROs), imaging core labs, investigator sites (often academic medical research centers) and ultimately the FDA, the IT architecture should be designed as a service-oriented architecture (SOA). IBM's SCORE architecture satisfies this requirement and can therefore serve as a basis for the enablement of biomarker-based R&D.

## CLINICAL BIOMARKERS AND ASSOCIATED DATA TYPES

Among clinical biomarkers, the genomic and imaging data types are creating the greatest IT challenges. By recommending CDISC's SDTM (see above) as the standard for drug submission, the FDA has taken an important step toward the integration of genomic and imaging data with conventional clinical patient data. CDISC SDTM is an easily extendable model that incorporates the FDA submission data structures. Based on strong collaboration between the biopharmaceutical industry, clinical research organizations, clinical trial investigator sites, IT vendors, and the FDA, SDTM represents the collective thoughts from a broad group of stakeholders.

Conventional clinical data are categorized into four classes, subdivided into domains, of the SDTM data model:

1. *Events* include specific domains covering adverse events, subject disposition, and medical history.
2. *Interventions* cover exposure (to study drug), concomitant medications, and substance use.
3. *Findings* contain the assessment information, such as electrocardiogram, lab, physical exam, vital signs, and subject questionnaire data.
4. The *other* class was created to group specialized categories of information such as trial design, supplemental qualifiers, trail summary, and related records, where related records provide linkages across the different files.

To support biomarker data submission, two new SDTM domains have been added:

Parent Domain:

| STUDYID | USUBJID | PGSEQ | PGGRPID | PGREFID | PGOBJ | PGTESTCD | PGTEST |
|---|---|---|---|---|---|---|---|
| NSCLC10 | ZBI000-0007 | 1 | EGFR-KD-001 | SPEC001 | | EGFR-KD | EGFR-KD (EGFR Gene, Protein kinase domain assoc |
| NSCLC10 | ZB1000-007 | 7 | CYP1A2-00001 | SPEC002 | | CYP1A2 | CYP1A2 Mutation |
| NSCLC10 | ZB1000-008 | 1 | CYP1A2-00003 | SPEC001 | | CYP1A2 | CYP1A2 Mutation DNA Analysis (provided f |
| NSCLC10 | ZB1000-009 | 1 | CYP2D6-00001 | SPEC001 | | CYP2D6 | CYP2D6 test |
| NSCLC10 | ZB1000-009 | 11 | CYP2C19-00001 | SPEC001 | | CYP2C19 (*2, | Cytochrome P450 2C19 Test |

| PGMETHCD | PGASSAY | PGORRES | PGSTRESC | PGSTRESN |
|---|---|---|---|---|
| 88323, 88380, 83890 (X2), 838 | 12700056 | EGFR | EGFR | |
| 83891, 83892 x2, 83998 | 50-776 | CYP 1A2 | CYP 1A2 | |
| 83891, 83892 x2, 83998 | 50-777 | CYP1A2 | CYP1A2 | |
| 83891, 83892, 83901 x2, | 50-574 | CYP2D6 | CYP2D6 | |
| 83891, 83892, 83901 x2, | 50-575 | CYP2C19 | CYP2C19 | |

Child Domain:

| STUDYID | USUBJID | PGSEQ | PGGRPID | PGREFID | PGOBJ | PGTESTCD | PGTEST |
|---|---|---|---|---|---|---|---|
| NSCLC10 | ZB1000-009 | 1 | CYP2D6-00001 | SPEC001 | HGNC:2625 | CYP2D6 | CYP2D6 GENE.g.-1584C>G |
| NSCLC10 | ZB1000-009 | 2 | CYP2D6-00001 | SPEC001 | HGNC:2625 | CYP2D6 | CYP2D6 GENE.g.100C>T |
| NSCLC10 | ZB1000-009 | 3 | CYP2D6-00001 | SPEC001 | HGNC:2625 | CYP2D6 | CYP2D6 GENE.g.124G>A |
| NSCLC10 | ZB1000-009 | 4 | CYP2D6-00001 | SPEC001 | HGNC:2625 | CYP2D6 | CYP2D6 GENE.g.883G>C |
| NSCLC10 | ZB1000-009 | 5 | CYP2D6-00001 | SPEC001 | HGNC:2625 | CYP2D6 | CYP2D6 GENE.g.1023C>T |
| … | … | … | … | … | … | … | … |

| PGMETHCD | PGASSAY | PGORRES | PGSTRESC | PGSTRESN |
|---|---|---|---|---|
| MOLGEN | 50-574 | M33388:g.-1584GG | | |
| MOLGEN | 50-574 | M33388:g.100TG | | |
| MOLGEN | 50-574 | M33388:g.124GC | | |
| MOLGEN | 50-574 | M33388:g.883GC | | |
| MOLGEN | 50-574 | M33388:g.1023CG | | |
| … | … | … | | |

**Figure 6** Partial sample of the pharmacogenomics SDTM domain.

**TABLE 1 Mapping of DICOM Imaging Metadata Tags into SDTM Imaging Domain**

| CDISC STDM Imaging Domain | | DICOM Tag | | |
|---|---|---|---|---|
| Variable Name | CDISC Notes (for Domains) or Description (for General Classes) | Tag | Attribute Name | Attribute Description |
| Unique subject identifier | Unique subject identifier within the submission | (0012,0040) | Clinical trial subject ID | The assigned identifier for the clinical trial subject (see C.7.1.3.1.6; will be present if Clinical Trial Subject Reading ID (0012,0042) is absent, may be present otherwise) |
| Sequence number | Sequence number given to ensure uniqueness within a data set for a subject; can be used to join related records | (0020,0013) | Instance number | A number that identifies this image (*Note:* This attribute was named an image number in earlier versions of this standard) |
| Imaging reference ID | Internal or external Imaging identifier; example: UUID for external imaging data File | (0008,0018) | SOP instance UID | Uniquely identifies the SOP Instance (see C.12.1.1.1 for further explanation; see also PS 3.4) |
| Test or examination short name | Short name of the measurement, test, or examination described in IMTEST; can be used as a column name when converting a data set from a vertical to a horizontal format | (0008,1030) | Study description | Institution-generated description or classification of the study (component) performed |

- The *pharmacogenomic* (PG) and *pharamacogenomics results* (PR) domains will support submission of summarized genomic (genotypic data).
- A new *imaging* (IM) domain will include a mapping of the relevant DICOM metadata fields required to summarize an imaging submission.

The *PG domain* belongs to the *findings* class and is designed to store panel ordering information. The detailed test-level information (such as Genotype/SNP summarized results) is reported in the PR domain. Figure 6 shows what a typical genotype test might look like in terms of data content and use of the HUGO [14] nomenclature. The PG domain supports the hierarchical nature of pharmacogenomic results, where for a given genetic test (such as EGFR, CYP2D6, etc.) from a patient sample (listed in the parent domain), multiple Genotypes/SNPs can be reported (listed in the child domain). To support the use of imaging biomarkers, DICOM metadata tags have to be mapped into the fields of the new IM domain. Table 1 illustrates this mechanism.

While the FDA has proposed the SDTM data model for submission data, it is clear that this is only an interchange format for sponsors to submit summary clinical study data to the FDA in a standardized fashion. The FDA identified a need for an additional relational repository model to store the SDTM data sets. The requirement was to design a normalized and extensible relational repository model that would scale up to a huge collection of studies going back into the past and supporting those in the future. Under a CRADA, the FDA and IBM jointly developed this submissions repository called Janus (named after the two-headed Roman God) that can look backward to support historic retrospective trials and look forward to support prospective trials. The data classification system of CDISC such as interventions, findings, and events was leveraged in the Janus model with linkages to the subjects (for the patients enrolled in the clinical trial) to facilitate the navigation across the different tables by consolidating data in three major tables. Benefits resulting from this technique include reduced database maintenance and a simpler data structure that is easier to understand and can support cross-trial analysis scenarios. The ETL (extract–transform–load) process for loading the SDTM domain data sets instantiates the appropriate class table structure in Janus without requiring any structural changes.

## DATA INTEGRATION AND MANAGEMENT

As scientific breakthroughs in genomics and proteomics and new technologies such as biomedical and molecular imaging are incorporated into R&D processes, the associated experimental activities are producing ever-increasing volumes of data that have to be integrated and managed. There are two major approaches to solving the challenge of enterprise-wide data access. The

creation of data warehouses [15] is an effective way to manage large and complex data that have to be queried, analyzed, and mined in order to generate new knowledge. To build such warehouses, the various data sources have to be extracted, transformed, and loaded (ETL) into repositories built on the principles of relational databases [16]. Warehousing effectively addresses the separation of transactional and analysis/reporting databases and provides a data management architecture that can cope with increased data demands over time. The ETL mechanism provides a means to "clean" the data extracted from the capture databases and thereby ensures data quality. However, data warehouses require significant effort in their implementation. Alternatively, a virtual, federated model can be employed [17]. Under the federated model, operational databases and other repositories remain intact and independent. Data retrieval and other multiple-database transactions take place at query time, through an integration layer of technology that sits above the operational databases and is often referred to as *middleware* or a *metalayer*. Database federation has attractive benefits, an important one being that the individual data sources do not require modification and can continue to function independently. In addition, the architecture of the federated model allows for easy expansion when new data sources become available. Federation requires less effort to implement but may suffer in query performance compared to a centralized data warehouse.

Common to both approaches is the need for sorting, cleaning, and assessing the data, making sure they are valid, relevant, and presented in appropriate and compatible formats. The cleaning and validation process would eliminate repetitive data stores, link data sets, and classify and organize the data to enhance their utility. The two approaches can coexist, suggesting a strategy where stable and mature data types are stored in data warehouses and new, dynamic data sources are kept federated.

Genomic data are a good example of the dynamic data type. Since genomics is a relatively new field in biopharmaceutical R&D, organizations use and define data their own way. Only as the science behind genomics is better understood can the business definitions be modified to better represent these new discoveries.

The integration of external (partly unstructured) sources such as GenBank [18], SwissProt [19], and dbSNP [20], can be complicated, especially if the evolving systems use does not match the actual lab use. Standardized vocabularies (i.e., ontologies) will link these data sources for validation and analysis purposes. External data sources tend to represent the frontier of science, especially since they store genetic biomarkers associated with diseases and best methods of testing that are ever-evolving. Having a reliable link between genetic testing labs, external data sources for innovations in medical science, and clinical data greatly improves the analytical functionality, resulting in more accurate outcome analysis. These links have been designed into the CDISC PG/PR domains to facilitate the analysis and reporting of genetic factors in clinical trial outcomes.

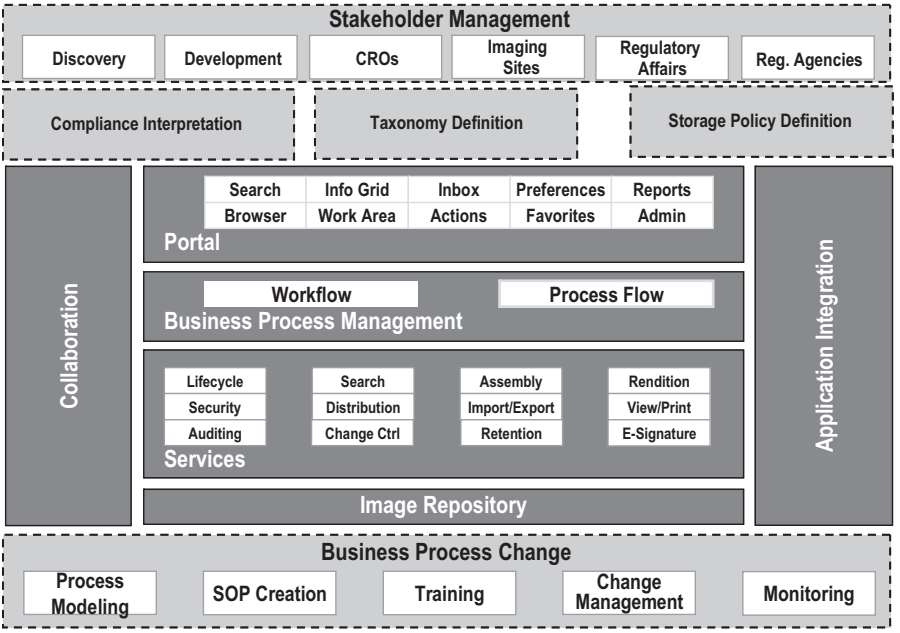


**Figure 7** IBM's medical image management solution.

As standards continue to evolve, the need for semantic interoperability is becoming increasingly clear. To use standards effectively to exchange information, there must be an agreed-upon data structure, and the stakeholders must share a common definition for the data content itself. The true benefit of standards is attained when two different groups can reach the same conclusions based on access to the same data because there is a shared understanding of the meaning of the data and the context of use of the data.

## IMAGING BIOMARKER DATA, REGULATORY COMPLIANCE, AND SOA

Under FDA's strict 21 CFR Part 11 [21] guidelines, new drug submissions must be supported by documentation that is compliant with all regulations. As explained above and illustrated in Figure 3, IBM's SCORE software asset has been designed for this purpose. SCORE's flexibility and modular design make it particularly suitable for the management of imaging biomarker data.

The FDA requires reproducibility of imaging findings so that an independent reviewer can make the same conclusion or derive the same computed measurements as that of a radiologist included in a submission. As a result, a unified architecture is required for a DICOM-based imaging data manage-

ment platform that supports heterogeneous image capture environments and modalities and allows Web-based access to independent reviewers. Automated markups and computations are recommended to promote reproducibility, but manual segmentation or annotations are often needed to compute the imaging findings. A common vocabulary is also needed for the radiological reports that spell out the diagnosis and other detailed findings as well as for the specification of the imaging protocols. Figure 7 shows how imaging biomarker data and work flows can be managed in a regulated multistakeholder environment. The solution includes a range of capabilities and services:

- *Image repository:* stores the image content and associated metadata.
- *Collaboration layer:* provides image life-cycle tasks shared across sponsors, CROs, and investigator sites.
- *Image services:* provide functionality such as security and auditing.
- *Integration layer:* provides solutions for integration and interoperability with other applications and systems.
- *Image taxonomy definition:* develops image data models, including naming, attributes, ontologies, values, and relationships.
- *Image storage policy definition:* defines and helps to manage policies and systems for image storage and retention.
- *Regulatory interpretation:* assists interpretation of regulations and guidelines for what is required for compliance.
- *Portal:* provides a role-based and personalized user interface.

In addition, the solution design incorporates the customized design, implementation, and monitoring of image management processes. It should also be pointed out that the medical image management solution architecture is fully based on principles of service-oriented architecture (SOA [22]). SOA is taking application integration to a new level. To take full advantage of the principles of SOA, it is important to start with a full understanding of business processes to be supported by an IT solution. Such a solution must be architected to support the documented business processes. The component business modeling (CBM [23]) methodology identifies the basic building blocks of a given business, leading to insights that will help overcome most business challenges. CBM allows analysis from multiple perspectives—and the intersection of those views offers improved insights for decision making. In the case of biomarker-enabled R&D, CBM will break down the transformed processes and identify the respective roles of in-house biopharmaceutical R&D functions and outside partners such as CROs, imaging core labs, investigator sites, genotyping services, and possible academic or biotech research collaborators.

After mapping work flows it is then possible to define a five-level service-oriented IT Architecture that supports the processes and work flows:

1. *Discrete:* hard-coded application
2. *Partial:* cross line of business processes
3. *Enterprise:* cross-enterprise business processes
4. *Partner:* a known partner
5. *Dynamic Partner:* any trusted partner

In its most advanced form, SOA will support a complex environment with integrated data sources, integrated applications, and a dynamic network of partners.

## CONCLUSIONS

Biomarkers are key drivers of the ongoing health care transformation toward the new paradigm of stratified and personalized medicine. In this chapter we focused on the role of IT in supporting the use of biomarkers in biopharmaceutical R&D. When doing so, we need to keep in mind that the benefits desired for patients and consumers will be realized only if the new biomedical knowledge is translated into stratified and personalized patient care. The biopharmaceutical industry will have to participate not only as a provider of drugs and medical treatments, but also as a contributor to the emerging biomedical knowledge base and to the IT infrastructures needed to enable biomarker-based R&D and clinical care. It is therefore critical to define the necessary interfaces between the respective IT environments and to agree on standards that enable data interchanges. IT standards and architectures must support the integration of new biomarker data with conventional clinical data types and the management of the integrated data in (centralized or federated) data warehouses that can be queried and analyzed Analysis and mining of biomarker and health care data are mathematically challenging but are necessary to support diagnostic and treatment decisions by providers of personalized care. Finally, service-oriented architectures are required to support the resulting processes and work flows covering the various health care stakeholders.

## REFERENCES

1. Trusheim MR, Berndt ER, Douglas FL (2007). Stratified medicine: strategic and economic implications of combining drugs and clinical biomarkers. *Nat Rev*, 6:287–293.
2. Pharma 2010: The Threshold of Innovation. http://www.ibm.com/industries/healthcare/doc/content/resource/insight/941673105.html?g_type=rhc.
3. DiMasi JA (2002). The value of improving the productivity of the drug development process: faster times and better decisions. *PharmacoEconomics*, 20(Suppl 3):1–10.

4. DiMasi JA, Hansen RW, Grabowski HG (2003). The price of innovation: new estimates of drug development costs. *J Health Econ*, 22:151–185.
5. http://www.fda.gov/oc/initiatives/criticalpath/whitepaper.html#execsummary.
6. http://www.fda.gov/cder/genomics/PGX_biomarkers.pdf.
7. Lesko LJ, Atkinson AJ Jr (2001). Use of biomarkers and surrogate endpoints in drug development and regulatory decision making: criteria, validation, strategies. *Annu Rev Pharmacol Toxicol*, 41:347–366.
8. CDISC. http://www.cdisc.org.
9. HL7. http://www.hl7.org.
10. DICOM. http://www.Medical.nema.org.
11. Janus data model. http://www.fda.gov/oc/datacouncil/, and http://crix.nci.nih.gov/projects/janus/.
12. Hehenberger M, Chatterjee A, Reddy U, Hernandez J, Sprengel J (2007). IT solutions for imaging biomarkers in bio-pharmaceutical R&D. *IBM Syst J*, 46(1): 183–198.
13. SCORE. http://www.03.ibm.com/industries/healthcare/doc/content/bin/HLS00198_USEN_02_LO.pdf.
14. http://www.gene.ucl.ac.uk/nomenclature/.
15. Kimball R, Caserta J (2004). *The Data Warehouse ETL Toolkit: Practical Techniques for Extracting, Cleaning, Conforming, and Delivering Data*. Wiley, Hoboken, NJ.
16. Codd EF (1981). The significance of the SQL/data system announcement. *Computerworld*, 15(7):27–30. See also http://www.informatik.unitrier.de/~ley/db/about/codd.html.
17. Haas L, Schwarz P, Kodali P, Kotlar E, Rice J, Swope W (2001). DiscoveryLink: a system for integrated access to life sciences data. *IBM Syst J*, 40(2):489–511.
18. GenBank. http://www.ncbi.nlm.nih.gov/Genbank/.
19. SwissProt. http://www.ebi.ac.uk/swissprot/.
20. dbSNP. http://www.ncbi.nlm.nih.gov/SNP/.
21. http://www.fda.gov/ora/compliance_ref/part11/.
22. Carter S (2007). *The New Language of Business: SOA & Web 2.0*. IBM Press, Armonk, NY.
23. http://www.ibm.com/services/us/gbs/bus/html/bcs_componentmodeling.html.

# REDEFINING DISEASE AND PHARMACEUTICAL TARGETS THROUGH MOLECULAR DEFINITIONS AND PERSONALIZED MEDICINE

Craig P. Webb, Ph.D.
*Van Andel Research Institute, Grand Rapids, Michigan*

John F. Thompson, M.D.
*Helicos BioSciences, Cambridge, Massachusetts*

Bruce H. Littman, M.D.
*Translational Medicine Associates, Stonington, Connecticut*

## INTRODUCTION

A chronic disease is really a phenotype representing the combination of symptom patterns and pathological findings that practicing clinicians have classified together as a disease. Like any other trait, each component of the disease exists because of contributions from genetic and environmental factors and the resultant modifications of biological functions that disturb the normal homeostatic state. Thus, the expression of a single chronic disease can be due to different combinations of genetic and environmental factors. Yet, when

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

physicians treat patients with a drug, they are actually modulating just single or subsets of molecular targets and their downstream pathways. In a given patient, the importance of that target will generally be more or less significant than the average importance in the disease population. Since the drug was approved based on an average clinical response, there will be patients with much better than average responses and patients who do not have a satisfactory clinical response at all.

Two important concepts for personalized medicines are hypothesized in this chapter. The first is that there is a distribution of relative expression of abnormal pathway activity for every element contributing to the pathogenesis of a disease. The second is that regardless of the proportion of subjects in a disease population, when a drug target is aligned with the most relevant abnormalities, it is likely that the therapeutic benefit of the drug for the individual patient where that alignment is found will be much greater than the average benefit for the entire disease population. This means that if the drug target and its downstream pathways are abnormally expressed to a large degree in a patient subpopulation, the therapeutic benefit of a drug targeting that pathway will also be more pronounced than that observed for the entire disease population.

These concepts are expressed graphically in Figure 1. The first is illustrated by scenarios 1, 2, and 3, where the three distributions represent the proportion of patients with different degrees of abnormal pathway expression for three different targets or pathways in the same disease population. The second concept is depicted as the solid black curve, where the degree of expression of the abnormal target or pathway is correlated with the potential therapeutic benefit of a drug targeting that pathway. The shaded areas represent the proportion of the disease population with the best clinical response.

These same two concepts will also determine the spectrum of relative toxicity (safety outcomes) of drugs across a population in a maner similar to efficacy. Thus, the therapeutic index (benefit/risk) of drugs targeting a specific pathway may become more predictable. Molecular definitions of disease and personalized medicine therefore has the potential to improve the outcomes of drug treatment provided that physicians have the appropriate diagnostic tools as well as drugs with targeted and well-understood biological activities. Then, rather than directing treatment toward the mythical average patient, physicians will be able to tailor their prescriptions to the patients who could benefit most and also customize the dose and regimen based on the pharmacogenetic profile of the patient and the absorption, distribution, metabolism, and elimination (ADME) properties of the drug. This should provide patients with safer and more efficacious treatment options.

In this chapter we use three different disease areas to illustrate these principles and the potential of personalized medicine. These concepts are most advanced in oncology, and we start with this example. Here tumors from individual patients are molecularly characterized and drug regimens are

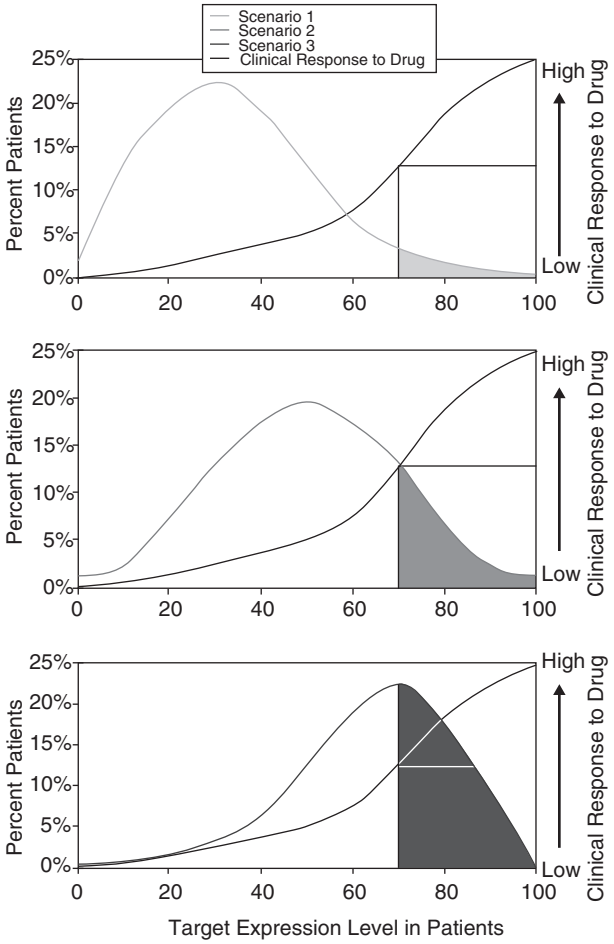


**Figure 1** Principles supporting personalized medicine strategies. Scenarios 1, 2, and 3 are three distributions representing the proportion of patients with different degrees of abnormal pathway expression for three different targets or pathways in the same disease population. The solid black curve shows the correlation between the degree of expression of the abnormal target or pathway and the potential therapeutic benefit of a drug targeting that pathway. The shaded areas represent the proportion of the disease population with increased probability of achieving the best clinical response. (*See insert for color reproduction of the figure.*)

selected that target the abnormal pathways driving the neoplastic phenotype. However, as described above, the same principles apply to all chronic diseases, and we illustrate this using type 2 diabetes and rheumatoid arthritis, where the science is just approaching a level that will enable personalized medicine strategies.

## ONCOLOGY

In the United States alone in 2007, it is estimated that nearly 1.5 million new cancer diagnoses will be reported and that over 500,000 patients will die from the disease [1]. Although early detection coupled with improved debulking procedures has led to some improvement in the survival of patients diagnosed with early-stage disease, the outcome of patients with advanced metastatic disease remains bleak. Metastatic disease will continue to burden society and the health care system due to an aging population, late-onset recurrence of microscopic metastases [2], and the fact that many tumors remain undetectable in their early stages. Long-term treatments for disseminated disease that maximize antitumor efficacy and patient survival while minimizing patient morbidity remain a primary objective in medical oncology, yet with few exceptions continue to be elusive.

The level of interest in the field of individualized molecular-based treatments has been driven by a number of factors, including public demand [3], regulatory agencies [4], and the possible financial incentives associated with biomarker–drug co-development within the pharmaceutical and biotechnology industry [5]. In oncology in particular, the lack of a statistical demonstration of agent efficacy in phase III trials is the primary reason for late-stage drug failures that logically result in increasing costs of oncology treatments that successfully attain market approval [6]. Biomarker strategies that can accurately identify the responsive tumors and/or patient population most likely to receive benefit provide a clear conduit to rescue failed or failing drugs, and also provide a realistic approach to enrich patient populations in early clinical trials to maximize the probability of drug efficacy. The concept was perhaps best illustrated during the approval of trastuzumab, a monoclonal antibody against *ERBB2* that is frequently amplified in breast tumors [7]. The co-development of a biomarker that assesses the mRNA or protein expression of *ERBB2* increased the overall response rate from approximately 10% in the overall population to 35 to 50% in the *ERBB2*-enriched subpopulation [8]. Without a means to enrich for patients with this molecular subset of tumors, trastuzumab may not have gained U.S. Food and Drug Administration (FDA) approval for the breast cancer "indication."

Today, there are an unprecedented number of available therapeutic agents that have been designed to target specific molecular entities irrespective of the selected phenotypic indications. Within our current knowledge base, there are more than 1500 drugs that with varying degrees of specificity, target defined constituents of molecular networks. Coupled with advances in technology and computer sciences, in the postgenomic era we are now presented with an unprecedented opportunity to apply our knowledge and existing and/or emerging resources to revolutionize medicine into a predictive discipline, where integration of clinical and molecular observations are used to maximize therapeutic index. We are now able to measure the molecular components of biological systems at an extraordinary density using standardized molecular

profiling technologies, which provide a portrait of the perturbed molecular systems within a diseased tissue. The lack of efficacy of targeted agents in oncology is somewhat ironically likely due in part to their specificity, and biomarker-driven selection of drug combinations will be required to target the Achilles heel of the tumor system.

Pathways involved in neoplastic transformation and in vivo tumor growth and progression are complex. Biological systems have evolved to provide the ultimate level of plasticity to allow cells to adapt to or exploit extracellular cues and ensure their long-term survival and/or survival of the organism as a whole [9]. The system is highly responsive to the cellular context, which includes both temporal (time dependent) and spatial (location dependent) factors that through a series of epigenetic events, influence the formation of the observed phenotype. The complexity of a tumor system is exacerbated due to the inherent genomic instability of a tumor cell; alterations in DNA repair mechanisms at the onset of tumorigenesis essentially instigate an accelerated microevolutionary process, where the interplay between each tumor cell and its microenvironment provides a constantly shifting context and selective milieu which naturally results in cellular heterogeneity [10,11]. The molecular network of the tumor system represents integration between the subsystems of malignant cells and their host microenvironment, which can include a multitude of proteomic and chemical constituents and other cellular systems contributed by endothelial, stromal, and inflammatory cells [12,13]. Collectively, tumor systems are highly adaptive and naturally exhibit significant variation over time, between locations, and across individuals.

The malignant phenotype results from perturbation of many pathways that regulate the tumor–host interaction and affect fundamental cellular processes such as cell division, apoptosis, invasion, and metabolism. The multistage process of tumor etiology and progression is driven by the progressive accumulation of genetic mutations and epigenetic abnormalities that include programs of gene expression. At first glance, the perturbations in individual components (DNA, RNA, protein, and metabolites) of a tumor system that have been identified in association with the various malignant phenotypes appear to reflect a somewhat stochastic process. However, while recent efforts to sequence the human genome have confirmed the large degree of redundancy within signaling networks, they have also revealed the extent that tumor cells utilize converging pathways to thrive within their selective environment [14,15]. Indeed, a relatively small number of key intracellular switches have been associated with tumorigenesis in preclinical mouse models and in vitro cell lines. This phenomenon, termed *oncogene addiction*, suggests that targeted agents against these central signaling relays may prove effective [16]. The classical oncogene target should be expanded to include any molecular aberration that is causative with respect to the etiology and/or progression of the disease at the network level. Additional intervention points within conserved tumor systems are being associated with the various phenotypes of cancer, in part through the use of high-throughput functional screens [17].

While these preserved network hubs represent obvious candidates for targeted therapies, redundancy within molecular pathways also provides a predictable path to drug resistance. Given the genetic instability within a tumor and the robustness within molecular networks, the probability that a cell within the average tumor mass will acquire resistance to a single agent with a selective molecular target during the life span of a tumor population would be expected to be very high [18]. This is well illustrated with the drug imatinib, which was developed to target the ABL kinase, which is constitutively active in patients with chronic myelogenous leukemia due to a *Bcr-Abl* gene translocation. Resistant tumors have recently emerged that utilize alternative pathways downstream of the drug target to circumvent the absolute requirement for the ABL kinase [19]. Similar results with other targeted agents used in single or minimal combinational modalities are emerging and would indeed be predicted from analysis of the target network that demonstrates the level of redundancy and robustness within signaling pathways [20]. Essentially, the plasticity within these networks reduces the dependency on single nodes within pathways. Tools such as systems biology will play a critical role in modeling the pathways to de novo and acquired resistance. Coupled with knowledge of drug–biomarker associations, logical targeted approaches that minimize the probability that a tumor will develop resistance and/or target key network nodes/hubs to reverse the resistant phenotype can be developed. The target for individualized therapy therefore becomes the perturbed tumor system as a whole, against which targeted therapeutics could be combined to maximize disruption of the networks identified.

An increasing number of technologies are available for assessment of the individual molecular components of cellular systems. Biomarkers can be genetic, genomic, proteomic, or metabolomic in nature, and have been used in various aspects of medicine to predict a phenotype of interest. While these have traditionally been developed as individual biomarkers that can readily be validated as an in vitro diagnostic, multivariate assays have recently been developed that simultaneously assess the levels of different biomarkers and provide a molecular signature in association with a phenotype or context [21]. These signature-based tests require integrated informatics, which can generate a mathematical algorithm that is trained and tested on independent sample sets [22]. Global molecular profiling offers many advantages over custom biomarkers, since a profile that accurately captures the underlying system of the disease can be attained and used as a common input for both the discovery and development of diagnostics. Indeed, a major bottleneck in the field of personalized medicine is the time required to develop a validated biomarker; a genome-scale technology that provides a standardized input of raw data coupled with computational methods that provide consistent algorithm-based predictive outputs would ultimately permit for the rapid development and testing of new molecular signatures associated with any phenotype of interest.

Despite the proliferation of new technologies that enable detection of specific biomarkers, gene expression profiling represents a relatively standardized platform that has been used extensively to create a depth of empirical data sets in association with various phenotypes. The ability to utilize gene expression profiles of human cancers to identify molecular subtypes associated with tumor progression, patient outcome, and response to therapy is increasingly evident [23]. For example, a multiplexed assay that determines the expression of a number of mRNA transcripts from breast carcinomas has been developed as a commercial test to predict the risk of tumor recurrence [21,24]. With respect to the prediction of optimal cytotoxic or targeted therapies, systematic efforts utilizing gene expression signatures to identify compounds that reverse the diseased genotype hold great promise [25]. In vitro cell line gene expression signatures associated with differential drug sensitivity have also been shown to predict tumor response to some agents in the clinic with varying degrees of accuracy [26,27]. These and other empirically based methods for predicting optimal therapeutics based on the overall genomic signature of the tumor will play a pivotal role in future personalized medicine initiatives. These and other signature-based methods are currently being evaluated within our predictive therapeutics protocol in conjunction with network analysis to determine their feasibility for broad application.

While the signature-based approaches outlined above represent a systematic approach for the logical selection of treatments based on the gene expression profile of a tumor, considerable experimentation is required to generate the predictive models. A supplementary approach is to utilize advances in network theory, systems biology, and computer modeling to reconstruct the aberrant molecular network predicted based on the same input of deregulated genes within the tumor. Although the deregulated expression of a molecular target may be associated with differential sensitivity to a targeted agent, gene or protein expression alone does not necessarily equate to target activity. For example, one of the first molecular events that occurs in some cells following stimulation with an extracellular ligand can be the down-regulation of the activated cell surface receptor (reduced protein expression) and/or reduced receptor mRNA transcription [28]. Nonetheless, successive waves of transcriptional events that occur within a tumor cell represent a hallmark of upstream chronic signaling cascades on which the tumor more likely depends. Gene expression profiling has been used successfully to identify the activation status of oncogenic pathways [29], demonstrating the feasibility of utilizing standardized gene expression signatures as a surrogate input for the prediction of network activity. Further analysis of the conceptualized network can predict convergence and divergence hubs within the tumor system, some of which can be targeted with existing therapeutics. This approach does not require comparison to an empirical data set, but rather, relies on network knowledge and graph theory to construct networks from known interactions between system components. Combinational strategies that target key nodes or hubs within

deregulated molecular networks associated with maintenance of the malignant phenotype may maximize therapeutic efficacy and may reduce the probability of a tumor cell utilizing alternative network components to achieve the "resistant" phenotype.

## PREDICTIVE THERAPEUTICS PROTOCOL

The general schema for a predictive therapeutics protocol is outlined in Figure 2. The primary objective of the protocol is to evaluate the merits of the various predictive methodologies outlined above while simultaneously providing information back to treating physicians in a real-time fashion for consideration in the design of a treatment plan. While a full description of the protocol is beyond the scope of this chapter, we have enrolled 50 patients in the first phase that was focused on the development of the critical infrastructure and logistics. From each patient, highly qualified tumor tissue (or isolated tumor cells) is processed using standard operating procedures to create a gene expression profile. The signature is used to compare to other well-annotated samples

**Figure 2** High-level review of our IRB-approved predictive therapeutics protocol in which patient tumors are processed using Affymetrix GeneChip technology after the required consenting and pathology clearance, to generate a gene expression signature that is reflective of the underlying biological context. These samples are processed using standardized procedures, to minimize confounding variables that can significantly influence the interpretation of the results. Molecular data are analyzed statistically relative to a wide variety of well-annotated samples within the database, and these intermediate results are applied further to the integrated knowledge base that includes systems biology tools. For example, enriched networks are identified and further refined to categorize significant convergence and/or divergence hubs that represent drugable targets with existing agents with known molecular mechanisms of action. Irrespective of the predictive method employed, each drug is associated with a normalized score for predicted efficacy. A report with these standardized predictions (indicated by the arrow) is provided back to the medical oncologist, who determines a treatment selection using all information available to him or her, which may include the molecular evidence. The patient's administered treatment is captured and their tumor response is assessed using standard clinical criteria. In this fashion, the association between the drug score predicted and the tumor response can be determined. In addition, a section of the patient's tumor is implanted directly into immune-compromised mice to establish a series of tumor grafts, which naturally more closely resemble the human disease at the molecular and histological level relative to established-cell-line xenograft models. These tumors are expanded in additional mouse cohorts and alternative predictive methods are tested to prioritize those with the most promise. Over time it is hoped that this approach of predictive modeling from standardized data, experimental testing, and model refinement may provide a means to identify optimal therapeutics with a high degree of confidence in a systematic fashion.

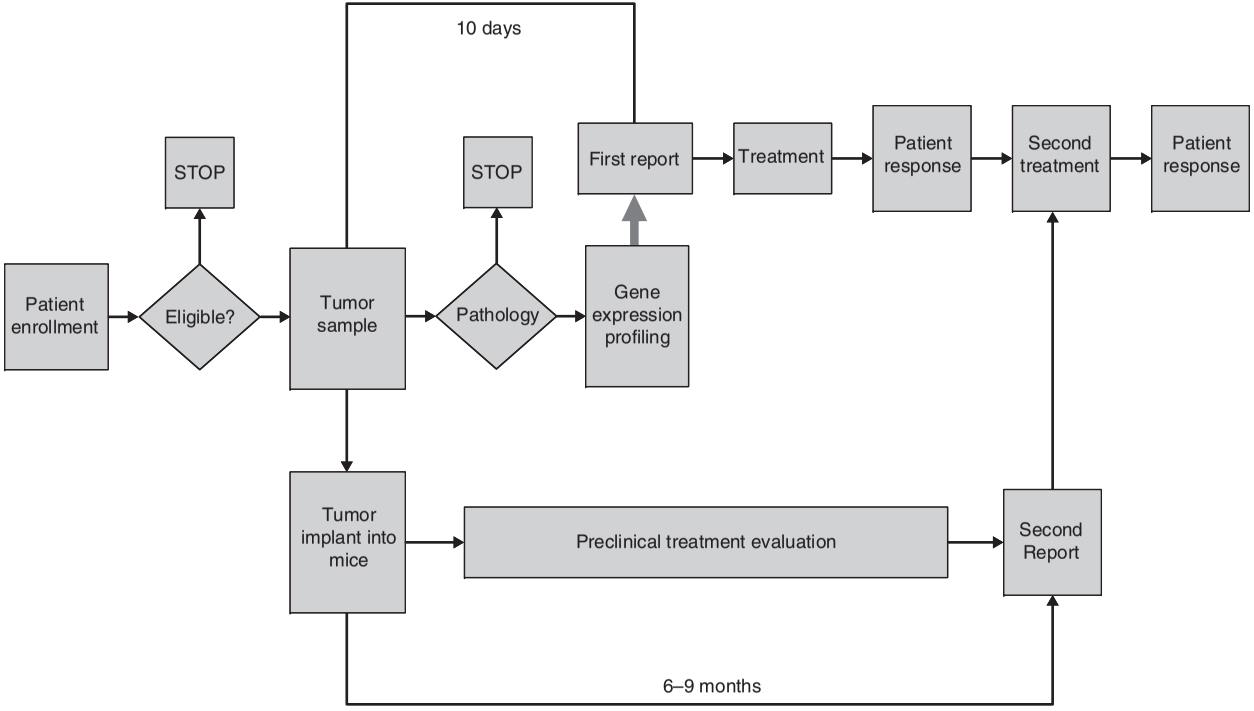
10 days
STOP
STOP
First report
Treatment
Patient response
Second treatment
Patient response
Patient enrollment
Eligible?
Tumor sample
Pathology
Gene expression profiling
Tumor implant into mice
Preclinical treatment evaluation
Second Report
6–9 months

**Figure 2**

within a large database, and deregulated patterns of gene expression used in conjunction with a knowledge base of known drug–target interactions to infer treatment strategies. We also attempt to establish a xenograft after implantation of a section of the fresh tumor into immune-compromised mice. These tumor grafts are expanded through two generations to create a large colony of mice harboring the patient's tumor, and these are then used to statistically evaluate the different predictive methodologies and their corresponding treatment recommendations. While the preclinical component of the protocol does not typically provide useful information to the treating physician, it represents an excellent resource for prioritizing predictive methodologies and for developing a biomarker strategy for novel therapeutics.

At the onset, it is apparent that this multidisciplinary protocol requires several infrastructural components as well as integrated logistics. These include the development of centralized informatics capabilities that permit full integration of clinical and molecular data, drug–biomarker knowledge, predictive modeling, and reporting. Standardized tissue procurement and pathological characterization with attention to quality control are essential to ensure consistency in the raw molecular data that are used to derive treatment predictions. Consistent feedback from the clinical and preclinical treatment outcomes is critical to assess the validity of the predictive methods. Each component of a therapeutic regimen is scored objectively based on the predictive methodology, and the ultimate success of the method is determined by comparing this standardized score with tumor response.

It is important to state that at this time, the results obtained from the clinical arm of the protocol remain anecdotal, due to the underpowered nature of the initial proof-of-feasibility experimental design. However, despite representing a nonvalidated method for drug prioritization, any molecular information that can be provided to the treating physician is deemed valuable, especially for late-stage metastatic or refractory patients who have exhausted their standard of care options. In this sense, this protocol serves as a rudimentary clearinghouse where patients are placed onto experimental protocols, including off-label protocols based on the molecular profile of their disease. With the multitude of predictive models now available to suggest optimal combinational strategies based on a standardized gene expression signature from an individual tumor, the preclinical tumor grafts provide an invaluable resource for triaging ineffective methodologies. At this time, we are exploring a range of methods that range from rudimentary target expression to the more sophisticated signature-based methods and network inference. In general, the molecular similarities between the human tumor and the derived tumor grafts are excellent and represent a significant improvement from the classical cell line xenograft models (Figure 3). This implies that the molecular network within the tumor system as a whole is generally maintained in the human and mouse hosts, although some clear exceptions are noted; for example, the expected reduction in markers of human vasculature and inflammatory cells are evident and expected. Although it is too early to claim direct equivalence

between the mouse tumor graft and patient tumor with respect to drug efficacy, early data are promising. In a handful of cases tested to date where the mouse and human harboring the same tumor are treated with the same treatment regimen, similar tumor responses have been observed. However, the tumor grafts are used predominantly to test the concept of using derived molecular data (in the mouse system) to predict optimal combinational therapies and not necessarily to define the best treatment strategy for the donating patient.

To illustrate how the molecular signature of an individual tumor can be modeled to predict target activation and sensitivity to approved drugs, we use a case study from the first phase of our protocol. A 63-year-old male presented with metastatic non-small cell lung carcinoma, and after the necessary consent, enrolled in this research protocol. A biopsy of the tumor was qualified and released by pathology, and the sample processed within a CLIA/CAP-accredited laboratory that utilizes full genome Affymetrix GeneChip technology. The standardized gene expression data were compared to a database of other tumors and normal tissues, to identify the most significantly deregulated gene transcripts within the patient's tumor. Our informatics solution uses a database of drug–biomarker knowledge that includes the reported molecular targets of more than 1500 drugs, the interaction type, and the effect of the interaction (agonistic or antagonistic). By aligning this knowledge with the deregulated signature within the tumor, potential drugs of interest are quickly identified. Of key importance, each drug is assigned a priority score (weight) based on the indication of the biomarker, which in turn depends on the predictive methodology used. For example, increased expression of a molecular target may indicate the corresponding targeted drug, and this is scored based on the normalized gene expression value. Since target expression does not necessarily equate to target activation status, we have also developed a specific network analysis tool in conjunction with GeneGo (http://www.genego.com), which systematically evaluates the topological significance within reconstructed networks.

Molecular networks are constructed based on the input provided (in this example, overexpressed genes) and an algorithm compares this to the connectivity map of the global network. The significance for each node within the identified network is calculated based on its probability of providing network connectivity, and this is used to score each drug. In the tumor of this patient, a major divergence point was identified as epidermal growth factor receptor (EGFR) (Figure 4), suggesting EGFR lie upstream of the transcriptional events observed. In this case, EGFR was also overexpressed at the transcript level relative to other tumors and, collectively, was therefore assigned a high score. Based on this inference, the medical oncologist confirmed EGFR gene amplification in the tumor using traditional FISH analysis. This patient was treated with a combination of erlotinib, cisplatin (reduced expression of the *ERCC1* gene), and bevacizumab (inferred constitutive activation of the VEGF–VEGFR pathway). The patient exhibited a partial response to this

targeted, combinational treatment, and is currently maintained on a non-cytotoxic regimen of erlotinib and bevacizumab alone (Figure 5).

While these results remain anecdotal in nature, a handful of significant tumor responses have been observed in the first phase of the protocol. Tumor regression has also been observed in several mouse models, where the various systematic methods for predicting optimal drug combinations are being evaluated and prioritized. The key to realizing the full promise of personalized medicine in oncology lies in the ability to predict combinations of agents in a systematic and objective fashion, irrespective of context and historical disease classification. The wealth of drugs currently available for combinational treatments necessitates a bold migration away from traditional diagnostics in which custom biomarkers are developed in parallel with a specific drug, typically in tumor subtypes. In conjunction with the isolation of the postulated tumor stem cell compartment of a cancer [30], standardized technologies that permit the application of genome scale network-based approaches for the prediction of any drug in various combinations may allow specific targeting of the Achilles heel of the molecular system irrespective of the observed phenotype.

## TYPE 2 DIABETES

Type 2 diabetes, in many ways like cancer, arises from a complex interplay of environmental and genetic factors that lead to a broad spectrum of conditions characterized by chronic high levels of glucose. While glucose levels are the common denominator for diagnosis of the disease, the simplicity of this mea-

**Figure 3** The preclinical arm of our predictive therapeutics protocol allows for the prioritization of methodologies based on their ability to predict optimal combinational designs derived from the networks identified within the tumorgraft system. These tumorgrafts are established directly from the patient's tumor by implantation into immune-compromised mice, and are characterized by both molecular profiling and histopathology. In this particular example, the data were restricted to include only biomarkers that represent known drug targets. In this fashion, the relative distribution of existing targets can be determined across patient tumors and their corresponding tumor grafts. (A) A heat map following unsupervised hierarchical clustering shows how the tumorgrafts in the mouse host closely resemble their donating human tumor at the genomic level even when the analysis was restricted to utilize only known drug targets. Patient tumors and their derived mouse tumorgrafts are coded with the same color and can be seen to co-cluster based on their overall genomic similarity. Probes encoding EGFR are highlighted to show the distribution of expression of this target across the various tumors. (B) The mean correlation coefficient in a direct comparison of human tumors with mouse tumor grafts is approximately 0.93, demonstrating excellent overall similarity at the biomarker level. Some notable exceptions are evident, such as reduced expression of human targets associated with angiogenesis in the murine host. (*See insert for color reproduction of the figure.*)

surement actually masks a complex set of problems that vary among individuals. Some people are beset primarily with dysfunctional pancreatic beta cells, while others may have more significant issues in muscle or liver tissues. Being able to determine the underlying nature of an individual's diabetes will help best determine the proper course of treatment.

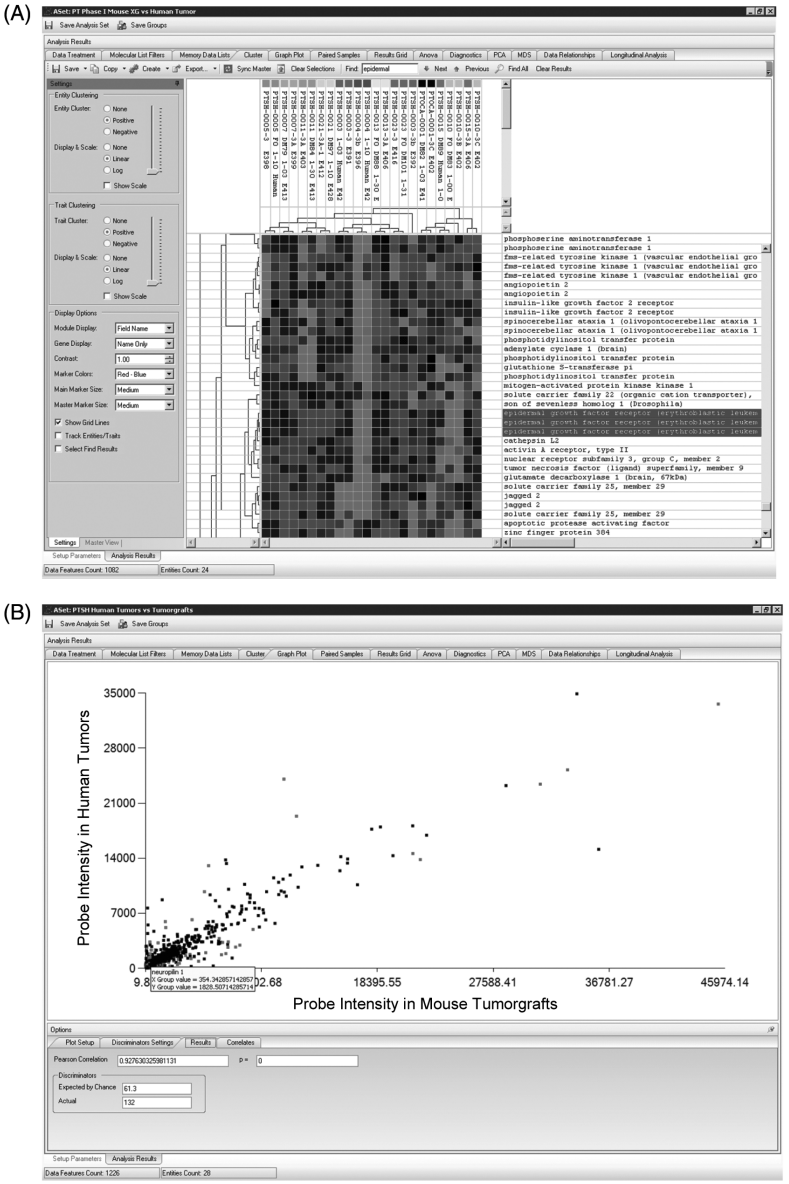


**Figure 3**

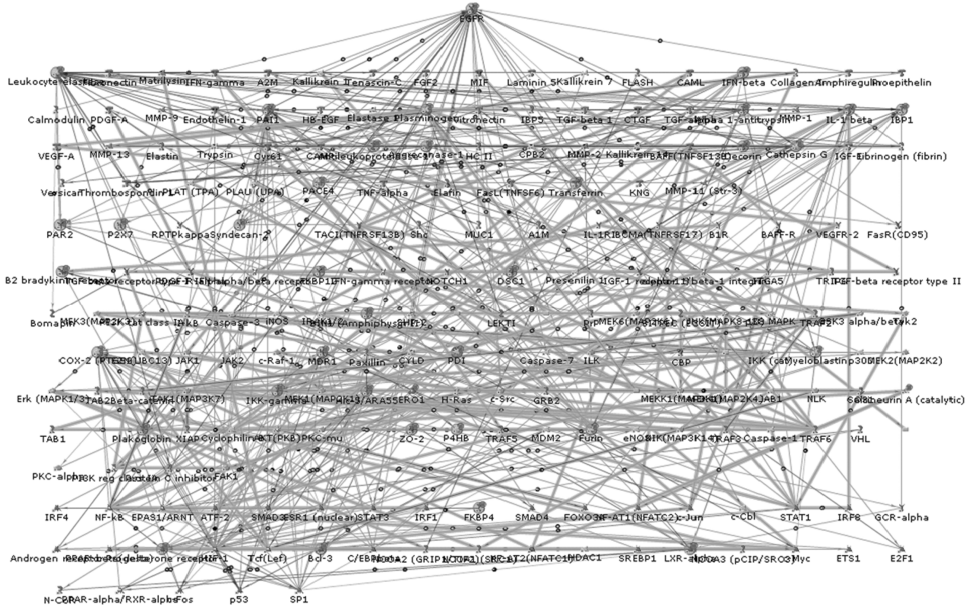


**Figure 4** Topological network analysis of the overexpressed genes from a non-small cell lung carcinoma identified a potential key input node at the level of EGFR. The results of these analyses are displayed using MetaCore, a systems biology network tool produced by GeneGo (www.genego.com). The significance of each node to confer system connectivity can be inferred after comparison with the global connectivity map and the drug–target knowledge base applied to select corresponding inhibitors. Among other applications, this type of systems approach, which does not depend on prerequisite empirical data sets, can readily be applied for the discovery of new disease targets, prioritization and/or validation of existing targets, and/or the identification of new indications for compounds that have a known or associated molecular mechanism of action. A key aspect is successful identification of the significant convergence or divergence hubs or nodes within the identified networks. (*See insert for color reproduction of the figure.*)

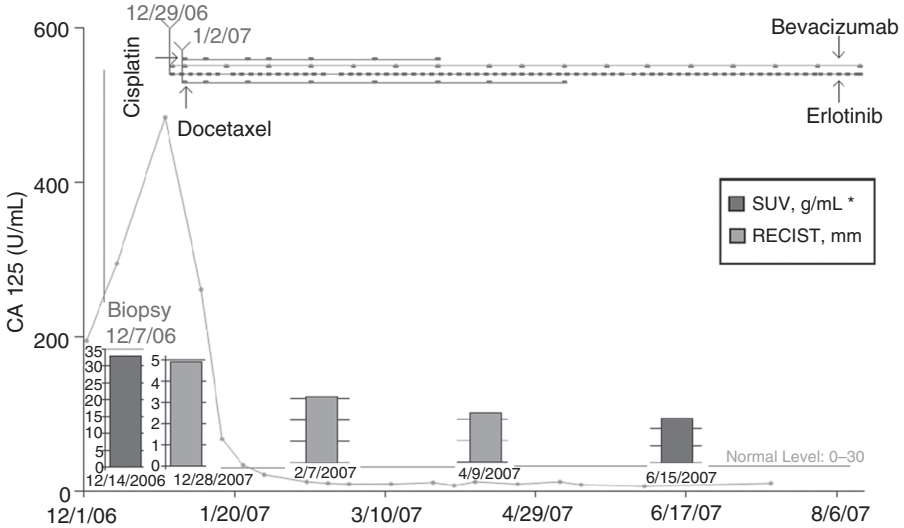


**Figure 5** Anecdotal evidence of a molecularly targeted combinational treatment in a 63-year-old man with metastatic non-small cell lung carcinoma. In this example, the patient's tumors showed a prolonged partial response to erlotinib (overexpression of EGFR and network-based inference of activated EGFR) in combination with cisplatin and bevacizumab, which were also indicated from the molecular profiling data (low ERCC1 gene expression and evidence of constitutive VEGF–VEGFR network signaling, respectively).These agents were combined with docetaxel, an approved second-line treatment for metastatic NSCLC. The levels of the serum marker CA125 together with the sum of the maximum dimensions of the target lesions (CT scan) and standard uptake value for the glucose tracer ([$^{18}$F]DG PET scan) are shown over time. The timing of the respective treatments is also shown.

Approaching diabetes from a more molecular point of view provides the opportunity to personalize treatment. Efforts to do this are still in their infancy, but significant progress is being made, with real results becoming apparent. Metabonomics focuses on analyzing a host of small molecules, generally in the urine but potentially in any bodily fluid, and determining differences among disease states. Most work thus far has been in animal models, but progress is being made with human samples. For example, a fingerprint of thiazolidinedione treatment was detected in humans, although no difference was found looking at healthy versus diseased individuals [31], a review of this field by Griffin and Nichols [32] with a focus on diabetes and related disorders highlights both the potential of the field and identifying issues.

Proteomics and transcriptomics have also been studied extensively as potential means of assessing subtypes of diabetes. For both technologies, diabetes presents a challenge, due to the inaccessibility of relevant tissues for the study of humans. However, in some cases, such as proteins derived from the kidney and other proteins secreted or leaked into the urine, it is possible to carry out such studies. Susztak and Bottinger provided examples of the advances and issues with these technologies for studying [33].

Because DNA is much more readily available than the relevant proteins and mRNA in tissues, genetic studies of diabetes have advanced more rapidly. Both type 1 and type 2 diabetes have long been known to have significant genetic components. Type 1 diabetes has generally been most highly associated with genes involved in the immune response, while type 2 has been less easily addressed, due both to its complex nature and its variable phenotype, such as age and severity of onset. Frequently, genes responsible for MODY (maturity onset diabetes of the young) have been assumed to play a role in the adult form of the disease, but that connection is weak in some cases. Genetic associations with type 1 diabetes have been strongest with genes in the MHC cluster and to a lesser extent with insulin and components of its signaling pathways. Genetic associations with type 2 diabetes have been less well replicated, but until the advent of whole genome scans had focused on genes known to be involved in obesity, lipid handling, and signaling, but many of these associations are poorly replicated. One attempt to replicate associations with 134 single-nucleotide polymorphisms (SNPs) across more than 70 genes found that only 12 SNPs in nine genes could be replicated [34]. This suggests both that there are many genes of weak effect and that there are many genes not yet discovered. Advances in selecting new therapeutic targets and the most appropriate patients for each therapy depend on identifying these genes so that the true complexity and subtypes of the disease can be determined.

For both forms of diabetes, whole genome analysis of large cohorts is beginning to make substantial inroads into understanding the etiology of the diseases. The availability of large (multi-thousands) family sets for type 1 diabetes patients and large case–control cohorts for genetic analysis, coupled with cheaper genotyping technology and an in-depth understanding of human genome structure, are allowing previously unsuspected genes to be linked with the disease and setting the stage for a much more detailed molecular understanding of the systems involved and how they may go awry in diabetes. The bottleneck in overall understanding has changed from identifying the appropriate genes to study, which used to be virtually impossible, to functionally characterizing the myriad genes that are now associated with diabetes and placing them in appropriate cellular and molecular pathways. Now that the multitude of genes that lead to diabetes are being uncovered, it will be possible to subdivide the disease into categories and determine whether the different subsets might best be treated by particular therapies.

Any review of novel genes associated with diabetes is certain to be outdated even before it is complete. With the advent of so many whole genome scans,

cross-replication of findings across cohorts has recently become a priority among research groups, and many of the new associations have already been confirmed. Those that have not been confirmed may still be real but suffer difficulties between populations due to differences with respect to disease, risk factors, ancestry, and other confounding variables that make comparisons challenging. Nevertheless, recent reviews can help make sense of the exploding databases [35]. Recent publications of whole genome scans include type 1 diabetes in 2000 British cases [36] and 563 European cases/483 European trios [37], and in type 2 diabetes, 694 French cases [38], 1464 Finnish and Swedish cases [39], 1161 Finnish cases [40], and 1399 Icelandic cases [41]. Although not completely concordant, the genes shared among these studies provide strong evidence that many are involved in the genetic basis of diabetes.

The most striking new gene to be associated with type 2 diabetes is *TCF7L2*, a transcription factor that regulates a number of genes relevant to diabetes [42]. Prior to the first publication on the association of *TCF7L2* with diabetes, its role in the regulation of proglucagon had been established [43], but its high level of importance was not apparent. After the initial report, numerous publications emerged that associated it with diabetes, glucose levels, birth weight, and other related phenotypes in many different populations. *TCF7L2* is clearly important in diabetes etiology, and its genotype may also be valuable for choosing the best mode of treatment. A retrospective analysis of patients treated with either metformin or a sulfonylurea showed no genetic difference in treatment effect with metformin but a significant genetic effect if treated with a sulfonylurea [44]. Thus, knowing a patient's *TCF7L2* genotype may help guide a physician in a drug treatment decision, but it may also affect how aggressively a patient should be handled. For example, should patients with one or two high-risk alleles be started on medication or lifestyle modification sooner than patients at lower risk? Should their glucose levels be managed more aggressively? Knowing which prediabetic patients are more likely to develop the disease based on *TCF7L2* genotype may help motivate physicians and patients to take more aggressive prophylactic approaches. The 14% of patients homozygous for the high-risk allele *rs12255372* [42], the most strongly associated SNP, might be willing to adopt more rigorous lifestyle changes than the majority who are at lower risk of disease.

Across all disease areas, pharmacogenetic studies have been plagued by a lack of replication for a variety of reasons but generally related to the study size [45]. Appropriately sized studies on the genetics of response to diabetes medications are now just beginning to emerge, but like most other areas, replication is frequently lacking. There are two broad classes of drug response studies. In many cases, individuals have variations in enzymes responsible for the metabolism, uptake, or other aspect of drug handling. A genetic analysis would inform the proper dosage level for a given drug independent of disease subtype. Alternatively, genetics may predict the subtype of diabetes within the patient population and could determine which treatment, whether lifestyle or drug, would be most efficacious for that particular subtype.

The glitazone class of therapeutics act through peroxisome proliferator-activated receptor (PPAR) gamma, and thus patients with variants in the gene, may respond differently to such drugs. One of the first such studies involved extensive resequencing of the gene in 93 Hispanic women [46]. Novel SNPs were identified that were weakly associated with troglitazone response, but the number of subjects was too small to be convincing. A subsequent study with a much larger population, 3548 individuals at high risk of diabetes, was examined for the P12A and other polymorphisms in PPARγ. This particular amino acid–changing SNP has frequently (though not always) been associated with risk of diabetes. The potential association of P12A with the efficacy of therapeutic intervention (lifestyle, metformin, or troglitazone) on development of diabetes was assessed rather than the risk of diabetes itself. Even though PPARγ is the target of troglitazone, no association with treatment effects was observed [47]. This could be caused by a multitude of possibilities. P12A may have little or no effect on troglitazone action, since it is far from the drug-binding site. The P12A polymorphism may have an early effect on diabetes progression and hence may not affect progression at later times after its early impact. Thus, even knowing that a patient is predisposed to diabetes because of variation in a particular gene may not be useful if the knowledge is not used at the appropriate time in disease progression.

In contrast to the null troglitazone effect with PPARγ, metformin appeared to substantially benefit patients with the E23K variation in *KCNJ11*, a gene known to be associated with diabetes. Those with the E23 variant of *KCNJ11* were less susceptible to progression when treated with lifestyle changes or metformin, while those with the K23 variant only benefited from lifestyle changes [48]. Further replication is required before recommendations can be made for clinical practice, but, if replicated, this information would help guide the most appropriate therapy for particular subgroups of patients. Other drugs have also been examined for associations with various candidate genes. These studies are often plagued by small numbers of people and/or varying definitions of drug response. In one study with pioglitazone, the population was relatively small ($n$ = 113) and two different measures of drug response were used, resulting in different conclusions [49]. At least one prospective study has been initiated in which diabetic patients have been selected for antioxidant therapy based on genotype [50]. Prospective studies are the gold standard for proving an effect but are not always feasible because of the high cost.

Categorizing patients for disease based on genetic or circulating markers will help choose the best therapeutic options once more data are available. In addition, choice of the most appropriate dose of a drug can be dependent on variation in ADME genes, as has been shown clearly for drugs such as warfarin [51]. Similarly, the dose of a treatment for diabetes can be affected by variation in genes completely unrelated to the underlying diabetic condition. For example, the *OCT-1* gene is not thought to be involved in diabetes but still has an apparent effect on treatment. This gene is important in the uptake of metformin into the liver, where it acts on AMPK. When people with normal

**TABLE 1 Patient Genotype as a Potential Guide for Treatment Decisions**

| TCF7L2 | PPARγ | KCNJ11 | Diabetes Risk | Lifestyle Modification | Metformin | Percent of Population |
|---|---|---|---|---|---|---|
| aa | PP | KK | Very high | Aggressive | Titrate | 1.9 |
| aa | PP | EK/EE | High | Aggressive | Normal | 10.8 |
| aa | PA/AA | KK | High | Aggressive | Titrate | 0.3 |
| aa | PA/AA | EK/EE | Moderate | Normal | Normal | 1.9 |
| AA/Aa | PP | KK | High | Aggressive | Titrate | 10.8 |
| AA/Aa | PP | EK/EE | Moderate | Normal | Normal | 61.4 |
| AA/Aa | PA/AA | KK | Moderate | Normal | Titrate | 1.9 |
| AA/Aa | PA/AA | EK/EE | Lower | Normal | Optional | 10.8 |

or variant *OCT-1* genes were subjected to an oral insulin glucose tolerance test in the presence or absence of metformin treatment, those with normal *OCT-1* were found to clear glucose much more effectively and maintain lower insulin levels [52]. Although this study and studies discussed above categorize patients as responders and nonresponders, it may be better to refine the analysis and choose the drug dose based on genetics, as is done with warfarin. As long as safety issues are not in question, patients with a nonresponder genotype may simply require a higher dose of medication, a possibility that could be tested in clinical trials. Thus, genes not directly involved in diabetes but affecting drug action can still be important to understand.

A hypothetical example of how a patient's genotype could be used in guiding treatment decisions is shown in Table 1. For simplicity, the minor allele frequency for each SNP is set at 15%, which is approximately what is observed in a prediabetic population for each of these SNPs. For *TCF7L2* and *KCNJ11*, the minor allele is high risk, whereas the minor allele is low risk for PPARγ. With just these three genotypes, the prediabetic population generally considered to be at a similar high risk can be segregated into groups containing over 10% of the population actually at low risk, 65% at moderate risk, 23% at high risk, and 2% at very high risk. Even within these categories, differential treatment paradigms may be warranted based on individual genotypes. As more information accumulates relating to circulating biomarkers and additional genetic markers, these decision trees can be made much more powerful and personalized.

## RHEUMATOID ARTHRITIS

Rheumatoid arthritis (RA) is also a complex disease phenotype defined by clinical criteria [53] and with multiple genetic and environmental factors contributing to its pathogenesis, resulting in highly variable degrees of severity

and responsiveness to therapy. Selection of an initial therapeutic regimen for RA patients is currently based on disease duration and disease severity, particularly with respect to progression or status of joint damage assessed radiographically. Most studies suggest that early aggressive treatment with disease-modifying antirheumatic drugs (DMARDs) helps to delay or prevent joint damage and leads to better long-term outcomes. All DMARDs have significant safety issues, and the cost of treatment with biological DMARDs is high, about $10,000 per year.

In the United States the most commonly used DMARDs are methotrexate (MTX), sulfasalazine, hydroxychloroquine, and the biologicals targeting TNFα. Newly diagnosed patients are generally started on a DMARD, and future changes in therapeutic regimen are based on clinical response. These include addition of combinations of DMARDs as well as changes in DMARD and use of corticosteroids. While assessment of clinical status often involves the use of biomarkers (diagnostics) as well as clinical parameters, these therapeutic choices, other than those dictated by specific safety concerns, are currently empirical and are not based on prospective genetic or biomarker factors.

Approximately 46% of RA patients achieve an ACR20 response with low-dose MTX (the most common initial DMARD) [54]. TNFα-targeted therapies are more successful at reducing or even halting the progression of joint damage, but again about 29 to 54% of patients do not achieve a satisfactory clinical (ACR20) response [54]. In addition to these agents, other biologicals are available or in development that have other specific targets. Recombinant human IL-1 receptor antagonist (anakinra) competes with IL-1 for stimulation of IL-1 receptors and has moderate efficacy in RA [54]. A recently approved biologic, abatacept (a CTLA-4$^+$ modified immunoglobulin Fc region fusion protein), targets T-lymphocyte activation by blocking co-stimulation of T-lymphocyte CD28 by antigen presenting cell CD80/86 [55]. Rituximab, an anti-CD20 monoclonal antibody previously used for treating B-cell lymphomas, has also been shown to be effective in RA [56,57]. Other biologicals targeting different cytokine pathways, such as tocilizumab, an anti-IL-6 receptor monoclonal antibody [58,59], have also reported efficacy in RA and may become available in the future. Thus, there are multiple treatment possibilities for RA patients with distinctly different molecular targets and mechanisms of action, but currently, biomarkers are not used to select those more likely to have superior efficacy in individual patients.

The heterogeneity of RA is not only apparent from the unpredictable clinical response to approved and experimental treatments; it is also confirmed by studies of RA synovial tissue histology and patterns of gene expression within inflamed joints of RA patients [60]. The Online Mendelian Inheritance in Man (OMIM) database listing for RA (http://www.ncbi.nlm.nih.gov/entrez/dispomim.cgi?id=180300) includes references to at least 19 specific genes with significant associations with RA susceptibility, disease severity, or response to therapy. Whole genome-wide scans using single-nucleotide polymorphism (SNP) maps have become very cost-effective, enabling the rapid confirmation

of genetic associations with RA [61]. In addition, RNA peripheral blood microarray-based RA studies (transcriptomic studies) are beginning to appear in the literature, and preliminary data from these suggests that patterns of gene expression may predict disease severity and response to specific therapies [62,63]. In this section we describe how biomarker, genomic, and transcriptomic data may be used in the future to help improve clinical outcomes for RA patients.

Some genetic associations with disease are really just associations with markers that tell us that a region of a particular chromosome is associated with RA or some feature of RA. However, with a greater understanding of gene function and the ability to perform whole genome-wide scans, a useful way to classify SNP associations with RA is to infer from this information whether it is likely that the genetic differences will have functional significance and influence T-lymphocyte activation, macrophage function, specific cytokine and inflammatory signaling pathways, and/or generalized inflammatory responses secondary to downstream dysregulation of these pathways. This information, together with an ever-increasing number of targeted therapeutic agents and greater understanding of biomarkers predicting response to older drugs such as MTX, may lead to a more rational basis for treating individual patients or RA subpopulations.

In this section we describe a number of genetic and biomarker associations that suggest possible strategies to design such individualized therapeutic regimens for RA patients. This is not intended to be a complete list of all such reported associations, but rather, a selection to illustrate how a path to personalized medicine in RA may be investigated further. As such, we are proposing these hypotheses partly to accelerate this type of clinical research and hopefully to improve outcomes and lower the cost of treatment for RA. We have not found any evidence in the literature that biomarkers like these have actually been used prospectively and systematically to test personalized treatment hypotheses in RA.

In Table 2 we classify a number of known genetic and biomarker associations and speculate as to how this information may lead to therapeutic decisions. Using this information, it is also possible to create hypotheses that are easily testable using samples from randomized controlled clinical trials to achieve prospective scientific confirmation. They also illustrate how the practice of personalized medicine in RA may evolve, and its potential benefits.

***Hypothesis A*** If tocilizumab is approved as a DMARD for RA patients it will likely be on the basis of results similar to those in published phase II and phase III trials. ACR20 response scores at 16 and 22 weeks in two different trials were 63% alone and 74% with MTX [58] and 59% with MTX [59], respectively. Thus, roughly one-third of patients did not achieve an ACR20 clinical response. Yet this compound had significant safety issues, including higher risk for infection, and if it becomes an approved drug, it will probably have a high cost of treatment, similar to other biologics. As described in Table

**TABLE 2 Rheumatoid Arthritis Genetic and Biomarker Associations**

| RA Mechanism | Biomarker / Gene | Function / Role | Clinical Correlation | Possible Treatment Implications |
|---|---|---|---|---|
| Cytokine and severity of disease | TNFα gene promoter: G to A SNP at position -308 [68] | TNFα response to stress or inflammatory stimuli: A/A largest TNF response, G/G lowest response [67]. | DAS response to TNF targeted biologic best in G/G (81%) vs. 42% in A/A and A/G [67]. | Higher dose of TNF targeted agents for A/A and A/G may be needed. |
| | IL-10 gene [72] | Anti-inflammatory cytokine. | -2849A/G SNP G allelle associated with higher progression rate and more joint damage; promoter SNPs -1082A, -819T, and -592A define a low IL-10 producer haplotype [73]. | IL-10 promoter genotypes or IL-10 haplotypes may correlate with response to IL-10 treatment. |
| Myeloid cell (macrophage) activation and joint inflammation | IL-1β and IL-8 mRNA in blood monocytes [78] | Macrophage activation. | Increased activated macrophages in synovium. | High levels may indicate better treatment response to anakinra, MTX. |
| Cytokine and severity of disease | HLA-G Lack of 14-bp polymorphism [71] | Soluable HLA-G is anti-inflammatory, inhibits NK cell activity, and is increased by IL-10 [71]. | Methotrexate, a folate antagonist used for the treatment of rheumatoid arthritis, induced the production of soluble HLA-G molecules by increasing IL-10 [71]. | Significant association between favorable response to MTX and lack of the 14-bp polymorphism of HLA-G, odds ratio 2.46 for methotrexate responsiveness. |

| | | | | |
|---|---|---|---|---|
| | IL-6 promoter: G/C polymorphism at position -174 [64] | C/C has low IL-6 response to IL1β, G/G and G/C higher production of IL-6. | Associated with B-cell neoplasms, Kaposi sarcoma [75,76], and systemic JRA [64] rituxamab. | Clinical benefit of IL-6 targeted therapies (R4) may be greater in carriers of G allele and may also benefit more from B-cell targeted therapies. |
| TCR response and antigen presentation | HLA-DRB1-beta-chain shared epitope (SE) [77] | May be associated with initial T-cell-driven response to disease initiation factors or auto-antigens; CD80 and CD86 candidate genes also linked to this locus. | SE predisposes to sero-positive RA and more severe disease, including extraarticular manifestations. | Response to MTX + etanercept best in RA patients with two copies of SE [77]; abatacept blocks T-cell co-stimulation through CD80/CD86. |
| Auto-antigen generation | PADI4 haplotype (peptidylarginine deiminase) | Posttranslational modification enzyme that converts arginine residues to citrulline [74]. | A functional haplotype of PADI4 is associated with susceptibility to rheumatoid arthritis [74] and was associated with levels of antibody to citrullinated peptide. | Gold inhibits myeloid differentiation (R8) and may therefore reduce PAD; estrogen or estrogen receptor antagonists modulate PADI4 activity. |
| TCR response | PTPN22 (is a lymphoid-specific intracellular phosphatase) | PTPN22 down-regulates T-cell activation mediated by TCR and CD28 co-stimulation [69]. | SNP R620W minor allele 1858T associated with RA, type 1 diabetes, SLE, and autoimmune thyroiditis is present in approximately 28% of white patients with RA [69]. | Levels of PTPN22 expression and/or presence of the 1858T SNP may predict good response to T-cell targeted therapy such as abatacept or cyclosporine A. |

2, there is a promoter SNP at position 174 of the IL-6 gene that significantly influences the amount of IL-6 produced in response to IL-1 and other inflammatory stimuli [64]. The gene frequency of the C allele is about 0.4 in healthy subjects. In vitro C/C cell constructs do not increase IL6 production in response to IL-1 stimulation compared to a 3.6-fold increase for G/G cell constructs [63]. Thus, it is likely that this SNP has functional significance. It is reasonable to hypothesize that the therapeutic benefit of tocilizumab will be different in populations that do not increase IL-6 signaling as much as a population with a robust increase in IL-6 during disease flares. If 30 to 40% of RA patients do not achieve a good clinical response with tocilizumab, could these be patients whose disease is not so dependent on this pathway (e.g., C/C genotype) or patients who produce such large amounts of IL-6 (e.g., G/G genotype) that higher doses of tocilizumab would have been needed? These hypotheses are easily tested in the clinic and could lead to a rational personalized medicine treatment regimen that would be more cost-effective and have a better efficacy and safety profile.

***Hypothesis B*** TNFα-targeted agents are very effective DMARDs in RA. Yet, on average, 40% of patients do not achieve an ACR20 response. Again these agents cost $8000 to $10,000 per year and have significant safety issues [54]. As noted in Table 2, there is a -308 promoter G to A SNP in the TNFα gene with probable functional significance since it is associated with outcomes in several infectious diseases and different clinical outcomes in septic shock [65]. The allele frequencies are reported to be 0.77 for allele G and 0.23 for allele A in a Swedish study [66]. In one published study on RA where the clinical response to infliximab was compared between -308 G/A genotypes, a disease activity score (DAS28) improvement of 1.2 occurred in 81% of G/G patients and in only 42% in A/A and A/G patients. The clinical improvement based on the DAS28 score was about twice as good in the G/G patients as in the A/A and A/G patients [67]. TNFα promoter SNPs, including the -308 SNP, are also associated with clinical outcomes in RA [68].

If these findings were replicated, how could that lead to a personalized medicine approach that improved outcomes and reduced the overall cost of therapy in a population of RA patients? Using the gene frequency and response information above, 84% of responders would be G/G, 4% would be A/A, and 12% would be A/G. Clearly, A/A and A/G patients would be far better off trying a different type of DMARD first, or perhaps they require a different dose or dose regimen. These clinical differences in response to a TNFα blocking agent could also occur if the amount of TNFα produced during disease flares is much greater in patients with the A allele and the blood level of their TNFα blockers is not high enough to neutralize TNFα activity at these times. In other words, it is possible that the dose of TNFα-targeted agents that is required for a durable clinical response is really different in these populations. Since the dose and frequency of dosing in the label for these agents is based on the average response in groups given different doses and dosing

frequencies, it is quite possible that response rates could be improved using higher doses in patients with the A allele. Patients with the G/G genotype may actually do well with lower doses. If this hypothesis is proven correct, there is an opportunity for improved patient outcomes by using higher doses in a small number of A/A and A/G patients and cost-saving with improved safety by using lower doses in a much larger number of G/G patients. Clearly, both of these personalized medicine approaches to the use of TNFα-targeted agents can be tested prospectively and could greatly influence patient outcomes.

***Hypothesis C*** Abatacept was recently approved as a DMARD in RA for patients who do not respond adequately to an earlier DMARD. It blocks T-lymphocyte co-stimulation needed to fully activate a T-lymphocyte-driven immune response through interaction between CD28 and CD80/86 (B7-1 and B7-2) mimicking natural CTLA4-mediated down-regulation of immune responses. In controlled trials with MTX background therapy about 60% of patients achieved an ACR20 response compared to about 30% on MTX alone [55]. Again, because of high cost and significant safety risks, one asks whether there is a testable personalized medicine hypothesis that could improve the probability of response.

PTPN22 is a lymphoid-specific phosphatase that down-regulates T-cell activation mediated by TCR and CD28 co-stimulation. This gene has a very strong association with RA and in particular there is a SNP associated with RA and other autoimmune diseases, such as type 1 diabetes [69]. This SNP, a 1858C-T transition, results in an arg620-to-trp amino acid change that alters the protein's function as a negative regulator of T-cell activation. This allele is present in approximately 17% of white people from the general population and in approximately 28% of white people with RA. Other variants of the *PTPN22* gene are also probably associated with RA [70]. The relationship between response to abatacept and *PTPN22* genotype has not been investigated, but it is likely that the T-lymphocyte co-stimulation pathway is more active in RA patients with a variant of the PTPN22 gene that results in reduced phosphatase function, such as the 1858T SNP. Should these patients receive a drug such as abatacept as first-line treatment for RA instead of waiting to fail another DMARD? Perhaps an alternative biomarker could be a measure of PTPN22 phosphatase activity or gene expression in lymphoid cells. Regardless of the biomarker used, this clinical question can be answered easily in appropriately designed clinical trials.

***Hypothesis D*** MTX is commonly used as the first DMARD treatment and it is used increasingly in combination with biologics. Often, the period of time to assess the therapeutic benefit of MTX is prolonged as doses are increased and other drugs are added. For those patients who do not achieve a satisfactory response to MTX, this practice often results in MTX dose escalation into ranges more likely to cause liver damage, pulmonary toxicity, or bone marrow suppression in addition to allowing further progression of disease and joint

damage. Because of its low cost and acceptable safety profile, a strategy that enriched the population of patients started on MTX with those more likely to respond and provides alternative first treatments for patients less likely to respond to MTX could lead to significant improvements in overall outcomes and treatment costs.

Several alternatively spliced *HLA-G* mRNA isoforms have been described, including a 14-bp polymorphism of the *HLA-G* gene with the 14-bp sequence deleted, and a significant association has been reported between a favorable response to MTX and a lack of the 14-bp polymorphism of the *HLA-G* gene, with an odds ratio of 2.46 for MTX responsiveness [71]. This finding, if confirmed, may enable an enrichment strategy for RA patients more likely to respond to MTX. Interestingly, in vitro MTX also induces the production of soluble *HLA-G* molecules by increasing IL-10. Promoter polymorphisms of the IL-10 gene have also been reported to have functional significance and associations with RA [72]. Using samples collected from clinical trials with MTX treatment arms, it would be possible to test the hypothesis that good prospective MTX responders can be identified by evaluating these two biomarkers.

## CONCLUSIONS

The still experimental practice of personalized medicine in cancer patients described here illustrates all of the necessary components to develop effective personalized medicine treatment strategies that are systematic in nature: diagnostics, targeted agents with well-understood mechanisms of action, an understanding of the molecular pathways important in disease progression, and ways of rapidly assessing clinical success. This has especially been enabled by the power of new genomic, biomarker, and informatics technologies. These technologies have also been applied to other chronic disease states where the potential for personalized medicine also exists. Diabetes and rheumatoid arthritis in many ways are like cancer, with genetic and environmental factors contributing to a very heterogeneous spectrum of disease. As the understanding of what drives these chronic disease phenotypes improves and more homogeneous subpopulations can be identified, treatment regimens will become more personalized. This trend will lead to safer and more efficacious treatments earlier and reduce the burden of disease to individuals and society.

## REFERENCES

1. Jemal A, Siegel R, Ward E, Murray T, Xu J, Thun MJ (2007). Cancer statistics, 2007. *CA: Cancer J Clin*, 57(1):43–66.
2. Vessella RL, Pantel K, Mohla S (2007). Tumor cell dormancy: an NCI workshop report. *Cancer Biol Ther*, 6(9):1496–1504.

3. Maron BJ, Hauser RG (2007). Perspectives on the failure of pharmaceutical and medical device industries to fully protect public health interests. *Am J Cardiol*, 100(1):147–151.
4. Goodsaid F, Frueh FW (2007). Implementing the U.S. FDA guidance on pharmacogenomic data submissions. *Environ Mol Mutagen*, 48(5):354–358.
5. Jain KK (2006). Challenges of drug discovery for personalized medicine. *Curr Opin Mol Ther*, 8(6):487–492.
6. DiMasi JA, Grabowski HG (2007). Economics of new oncology drug development. *J Clin Oncol*, 25(2):209–216.
7. O'Donovan N, Crown J (2007). EGFR and *HER-2* antagonists in breast cancer. *Anticancer Res*, 27(3A):1285–1294.
8. Vogel CL, Cobleigh MA, Tripathy D, et al. (2002). Efficacy and safety of trastuzumab as a single agent in first-line treatment of *HER2*-overexpressing metastatic breast cancer. *J Clin Oncol*, 20(3):719–726.
9. Huang S (2004). Back to the biology in systems biology: what can we learn from biomolecular networks? *Brief Funct Genom Proteom*, 2(4):279–297.
10. Wang E, Lenferink A, O'Connor-McCourt M (2007). Cancer systems biology: exploring cancer-associated genes on cellular networks. *Cell Mol Life Sci*, 64(14):1752–1762.
11. Aranda-Anzaldo A (2001). Cancer development and progression: a non-adaptive process driven by genetic drift. *Acta Biotheor*, 49(2):89–108.
12. Hanahan D, Weinberg RA (2000). The hallmarks of cancer. *Cell*, 57–70.
13. Webb CP, Vande Woude GF (2000). Genes that regulate metastasis and angiogenesis. *J Neurooncol*, 50(1–2):71–87.
14. Balakrishnan A, Bleeker FE, Lamba S, et al. (2007). Novel somatic and germline mutations in cancer candidate genes in glioblastoma, melanoma, and pancreatic carcinoma. *Cancer Res*, 67(8):3545–3550.
15. Sjoblom T, Jones S, Wood LD, et al. (2006). The consensus coding sequences of human breast and colorectal cancers. *Science*, 314(5797):268–274.
16. Weinstein IB, Joe AK (2006). Mechanisms of disease: Oncogene addiction: a rationale for molecular targeting in cancer therapy. *Nat Clin Pract*, 3(8):448–457.
17. Haney SA (2007). Increasing the robustness and validity of RNAi screens. *Pharmacogenomics*, 8(8):1037–1049.
18. Michor F, Nowak MA, Iwasa Y (2006). Evolution of resistance to cancer therapy. *Curr Pharm Des*, 12(3):261–271.
19. Hochhaus A, Erben P, Ernst T, Mueller MC (2007). Resistance to targeted therapy in chronic myelogenous leukemia. *Semin Hematol*, 44(1 Suppl 1):S15–S24.
20. Bublil EM, Yarden Y (2007). The EGF receptor family: spearheading a merger of signaling and therapeutics. *Curr Opin Cell Biol*, 19(2):124–134.
21. Kaklamani VG, Gradishar WJ (2006). Gene expression in breast cancer. *Curr Treat Options Oncol*, 7(2):123–128.
22. Webb CP, Pass HI (2004). Translation research: from accurate diagnosis to appropriate treatment. *J Transl Med* 2(1):35.
23. Rhodes DR, Kalyana-Sundaram S, Tomlins SA, et al. (2007). Molecular concepts analysis links tumors, pathways, mechanisms, and drugs. *Neoplasia*, 9(5):443–454.

24. Miller LD, Liu ET (2007). Expression genomics in breast cancer research: microarrays at the crossroads of biology and medicine. *Breast Cancer Res*, 9(2):206.
25. Lamb J, Crawford ED, Peck D, et al. (2006). The Connectivity Map: using gene expression signatures to connect small molecules, genes, and disease. *Science*, 313(5795):1929–1935.
26. Lee JK, Havaleshko DM, Cho H, et al. (2007). A strategy for predicting the chemosensitivity of human cancers and its application to drug discovery. *Proc Nat Acad Sci USA*, 104(32):13086–13091.
27. Potti A, Dressman HK, Bild A, et al. (2006). Genomic signatures to guide the use of chemotherapeutics. *Nat Med*, 12(11):1294–1300.
28. Shtiegman K, Kochupurakkal BS, Zwang Y, et al. (2007). Defective ubiquitinylation of EGFR mutants of lung cancer confers prolonged signaling. *Oncogene*, 26(49):6968–6978.
29. Bild AH, Yao G, Chang JT, et al. (2006). Oncogenic pathway signatures in human cancers as a guide to targeted therapies. *Nature*, 439(7074):353–357.
30. Schulenburg A, Ulrich-Pur H, Thurnher D, et al. (2006). Neoplastic stem cells: a novel therapeutic target in clinical oncology. *Cancer*, 107(10):2512–2520.
31. Van Doom M, Vogels J, Tas A, et al. (2006). Evaluation of metabolite profiles as biomarkers for the pharmacological effects of thiazolidinediones in type 2 diabetes mellitus patients and healthy volunteers. *Br J Clin Pharmacol*, 63:562–574.
32. Griffin JL, Nichols AW (2006). Metabolomics as a functional genomic tool for understanding lipid dysfunction in diabetes, obesity and related disorders. *Pharmacogenomics*, 7:1095–1107.
33. Susztak K, Bottinger EP (2006). Diabetic nephropathy: a frontier for personalized medicine. *J Am Soc Nephrol*, 17:361–367.
34. Willer CJ, Bonnycastle LL, Conneely KN, et al. (2007). Screening of 134 single nucleotide polymorphisms (SNPs) previously associated with type 2 diabetes replicates association with 12 SNPs in nine genes. *Diabetes*, 56:256–264.
35. Sale MM, Rich SS (2007). Genetic contributions to type 2 diabetes: recent insights. *Expert Rev Mol Diagn*, 7:207–217.
36. Wellcome Trust Case Control Consortium (2007). Genome-wide association study of 14,000 cases of seven common diseases and 3000 shared controls. *Nature*, 447:661–678.
37. Hakonarson H, Grant SFA, Bradfield JP, et al. (2007). A genome-wide association study identifies *KIAA0350* as a type 1 diabetes gene. *Nature*, 448:591–594.
38. Sladek R, Rocheleau G, Rung J, et al. (2007). A genome-wide association study identifies novel risk loci for type 2 diabetes. *Science*, 445:881–885.
39. Saxena R, Voight BF, Lyssenko V, et al. (Diabetes Genetics Initiative) (2007). Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. *Science*, 316:1331–1336.
40. Scott LJ, Mohlke KL, Bonnycastle LL, et al. (2007). A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. *Science*, 316:1341–1345.
41. Steinthorsdottir V, Thorleifsson G, Reynisdottir I, et al. (2007). A variant in *CDKAL1* influences insulin response and risk of type 2 diabetes. *Nat Genet*, 39:770–775.

42. Grant SFA, Thorleifsson G, Reynisdottir I, et al. (2006). Variant of transcription factor 7-like 2 (*TCF7L2*) gene confers risk of type 2 diabetes. *Nat Genet*, 38:320–323.
43. Yi F, Brubaker PL, Jinet T (2005). TCF-4 mediates cell type-specific regulation of proglucagon gene expression by catenin and glycogen synthase kinase-3. *J Biol Chem*, 280:1457–1464.
44. Pearson EW, Donnelly LA, Kimber C, et al. (2007). Variation in *TCF7L2* influences therapeutic response to sulfonylureas. *Diabetes*, 56:2178–2182.
45. Loannidis JPA, Trikalinos TA, Ntzani EE, Contopoulos-Ioannidis DG (2003). Genetic associations in large versus small studies: an empirical assessment. *Lancet*, 361:567–571.
46. Wolford JK, Yeatts KA, Dhanjal SK, et al. (2005). Sequence variation in PPARg may underlie differential response to troglitazone. *Diabetes*, 54:3319–3325.
47. Florez JC, Jablonski KA, Sun MW, et al. (2007). Effects of the type 2 diabetes associated PPARg P12A polymorphism on progression to diabetes and response to troglitazone. *J Clin Endocrinol Metal*, 92:1502–1509.
48. Florez JC, Jablonski KA, Kahn SE, et al. (2007). Type 2 diabetes–associated missense polymorphisms KCNJ11 E23K and ABCC8 A1369S influence progression to diabetes and response to interventions in the Diabetes Prevention Program. *Diabetes*, 56:531–536.
49. Wang G, Wang X, Zhang Q, Ma Z (2007). Response to pioglitazone treatment is associated with the lipoprotein lipase S447X variant in subjects with type 2 diabetes mellitus. *Int J Clin Pract*, 61:552–557.
50. Levy AP (2006). Application of pharmacogenomics in the prevention of diabetic cardiovascular disease: mechanistic basis and the clinical evidence for utilization of the haptoglobin genotype in determining benefit from antioxidant therapy. *Pharm Ther* 112:501–512.
51. Yin T, Miyata T (2007). Warfarin dose and the pharmacogenomics of CYP2C9 and VKORC1: rationale and perspectives. *Thromb Res*, 120:1–10.
52. Shu Y, Sheardown SA, Brown C, et al. (2007). Effect of genetic variation in the organic cation transporter 1 (OCT1) on metformin action. *J Clin Invest*, 117:1422–1431.
53. Arnett FC, Edworthy SM, Bloch DA, et al. (1988). The American Rheumatism Association 1987 revised criteria for the classification of rheumatoid arthritis. *Arthritis Rheum*, 31:315–324.
54. Olsen NJ, Stein CM (2007). New Drugs for Rheumatoid Arthritis. *N Engl J Med*, 3501:2167–2179.
55. Kremer JM, Dougados M, Emery P, et al. (2005). Treatment of rheumatoid arthritis with the selective costimulation modulator abatacept: twelve-month results of a phase IIb, double-blind, randomized, placebo-controlled trial. *Arthritis Rheum*, 52:2263–2271.
56. Cohen SB, Emery P, Greenwald MW, et al. (REFLEX Trial Group) (2006). Rituximab for rheumatoid arthritis refractory to anti-tumor necrosis factor therapy: results of a multicenter, randomized, double-blind, placebo controlled, phase III trial evaluating primary efficacy and safety at twenty-four weeks. *Arthritis Rheum*, 54:2793–2806.

57. Emery P, Fleischmann R, Filipowicz-Sosnowska A, et al. (DANCER Study Group) (2006). The efficacy and safety of rituximab in patients with active rheumatoid arthritis despite methotrexate treatment: results of a phase IIB randomized, double-blind, placebo-controlled, dose-ranging trial. *Arthritis Rheum*, 54:1390–1400.
58. Maini RN, Taylor PC, Szechinski J, et al. (CHARISMA Study Group) (2006). Double-blind randomized controlled clinical trial of the interleukin-6 receptor antagonist, tocilizumab, in European patients with rheumatoid arthritis who had an incomplete response to methotrexate. *Arthritis Rheum*, 54:2817–2829.
59. Smolen AB, Rubbert-Roth A, Alecock E, Alten R, Woodworth T (2007). Tocilizumab, A novel monoclonal antibody targeting IL-6 signalling, significantly reduces disease activity in patients with rheumatoid arthritis. *Ann Rheum Dis*, 66(Suppl II):87.
60. Glocker MO, Guthke R, Kekow J, Thiesen, H-J (2006). Rheumatoid arthritis, a complex multifactorial disease: on the way toward individualized medicine. *Med Res Rev*, 26:63–87.
61. Docherty SJ, Butcher LM, Schalkwyk LC, Plomin R (2007). Applicability of DNA pools on 500 K SNP microarrays for cost-effective initial screens in genomewide association studies. *BMC Genom*, 8:214.
62. Edwards CJ, Feldman JL, Beech J, et al. (2007). Molecular profile of peripheral blood mononuclear cells from patients with rheumatoid arthritis. *Mol Med*, 13:40–58.
63. Lindberg J, Klint E, Catrina AI, et al. (2006). Effect of infliximab on mRNA expression profiles in synovial tissue of rheumatoid arthritis patients. *Arthritis Res Ther*, 8:R179.
64. Fishman D, Faulds G, Jeffery R, et al. (1998). The effect of novel polymorphisms in the interleukin-6 (IL-6) gene on IL-6 transcription and plasma IL-6 levels, and an association with systemic-onset juvenile chronic arthritis. *J Clin Invest*, 102:1369–1376.
65. Mira J.-P, Cariou A, Grall F, et al. (1999). Association of TNF2, a TNF-alpha promoter polymorphism, with septic shock susceptibility and mortality: a multi-center study. *JAMA*, 282:561–568.
66. Rosmond R, Chagnon M, Bouchard C, Bjorntorp P (2001). G-308A polymorphism of the tumor necrosis factor alpha gene promoter and salivary cortisol secretion. *J Clin Endocrinol Metab*, 86:2178–2180.
67. Mugnier B, Balandraud N, Darque A, Roudier C, Roudier J, Reviron D (2003). Polymorphism at position -308 of the tumor necrosis factor alpha gene influences outcome of infliximab therapy in rheumatoid arthritis. *Arthritis Rheum*, 8:1849–1852.
68. Fonseca JE, Cavaleiro J, Teles J, et al. (2007). Contribution for new genetic markers of rheumatoid arthritis activity and severity: sequencing of the tumor necrosis factor-alpha gene promoter. *Arthritis Res Ther*, 9:R37.
69. Begovich AB, Carlton VEH, Honigberg LA, et al. (2004). A missense single-nucleotide polymorphism in a gene encoding a protein tyrosine phosphatase (*PTPN22*) is associated with rheumatoid arthritis. *Am J Hum Genet*, 75: 330–337.

70. Carlton VEH, Hu X, Chokkalingam AP, et al. (2005). *PTPN22* genetic variation: evidence for multiple variants associated with rheumatoid arthritis. *Am J Hum Genet*, 77:567–581.
71. Rizzo R, Rubini M, Govoni M, et al. (2006). HLA-G 14-bp polymorphism regulates the methotrexate response in rheumatoid arthritis. *Pharmacogenet Genom*, 16:615–623.
72. Lard LR, van Gaalen FA, Schonkeren JJM, et al. (2003). Association of the -2849 interleukin-10 promoter polymorphism with autoantibody production and joint destruction in rheumatoid arthritis. *Arthritis Rheum*, 48:1841–1848.
73. Summers AM, Summers CW, Drucker DB, Barson A, Hajeer AH, Hutchinson IV (2000). Association of IL-10 genotype with sudden infant death syndrome. *Hum Immunol*, 61:1270–1273.
74. Suzuki A, Yamada R, Chang X, et al. (2003). Functional haplotypes of PADI4, encoding citrullinating enzyme peptidylarginine deiminase 4, are associated with rheumatoid arthritis. *Nat Genet*, 34:395–402.
75. Kawano M, Hirano T, Matsuda T, et al. (1988). Autocrine generation and requirement of BSF-2/IL-6 for human multiple myelomas. *Nature*, 332:83–85.
76. Foster CB, Lehrnbecher T, Samuels S, et al. (2000). An IL6 promoter polymorphism is associated with a lifetime risk of development of Kaposi sarcoma in men infected with human immunodeficiency virus. *Blood*, 96:2562–2567.
77. Criswell LA, Lum RF, Turner KN, et al. (2004). The influence of genetic variation in the *HLA-DRB1* and *LTA-TNF* regions on the response to treatment of early rheumatoid arthritis with methotrexate or etanercept. *Arthritis Rheum*, 50:2750–2756.
78. Schulze-Koops MD, Davis LS, Kavanaugh AF, Lipsky PE (2005). Elevated cytokine messenger RNA levels in the peripheral blood of patients with rheumatoid arthritis suggest different degrees of myeloid activation. *Arthritis Rheum*, 40:639–647.

# ETHICS OF BIOMARKERS: THE BORDERS OF INVESTIGATIVE RESEARCH, INFORMED CONSENT, AND PATIENT PROTECTION

Heather Walmsley, M.A.
*Lancaster University, Bailrigg, UK*

Michael Burgess, Ph.D., Jacquelyn Brinkman, M.Sc., Richard Hegele, M.D., Ph.D., Janet Wilson-McManus, M.T., B.Sc., and Bruce McManus, M.D., Ph.D.
*University of British Columbia, Vancouver, British Columbia, Canada*

## INTRODUCTION

In 2000, the Icelandic Parliament (Althingi) authorized an Iceland-based subsidiary of deCODE Genetics to construct a biobank of genetic samples from the Icelandic population [1–4]. The Althingi also granted deCODE (which had a five-year commercial agreement with the Swiss pharmaceutical company Roche Holdings) a 12-year exclusive commercial license to use the country's medical records, in return for an annual 70 million kronur (approximately 1 million USD in 2007). These records were to be gathered, together with lifestyle and extensive genealogical data, into the Icelandic Health Sector Database. The resulting public outcry and academic critique has been well documented [3,5,6]. Several hundred articles appeared in newspapers [7],

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

many of them referring to the sale of the "genetic heritage" of the nation (see http://www.mannvernd.is/english/home.html for a list of media articles). A grass-roots lobby group, Mannvernd, emerged to fight the project, complaining principally about the use of "presumed consent" and the commercial aspects of the agreement [4]. Despite these critiques, Iceland was one of the first countries to discuss how to structure a biobank at the political level [8].

When a population geneticist from Stanford University announced plans for a human genome diversity project, he received a similar reception. This project aimed to challenge the ethnocentrism of the Human Genome Project by studying 722 diverse "anthropologically unique" human populations [9]. Indigenous activists were unconvinced. Debra Harry worried that "these new 'scientific findings' concerning our origins can be used to challenge aboriginal rights to territory, resources, and self-determination" [10]. The Canada-based Rural Advancement Foundation International (RAFI), now the ETC Group (Action Group on Erosion, Technology and Concentration), characterized the list of 722 as a list of peoples who had suffered most at the hands of Western "progress" and campaigned against this "bio-colonial Vampire Project." The project has since stimulated productive dialogue about the importance of race and ethnicity to health and genetic research.

In June 2007, UK Biobank opened its doors to donors in Glasgow, the fourth of about 35 planned donation points [11]. The project hopes to recruit a total of 500,000 volunteers aged between 40 and 69. This biobank is a prospective cohort study hoping to contribute to disease risk prediction through the identification of biomarkers [12]. The UK biobank has recognized the need to build public trust and knowledge. This has led to public engagement, although some critics suggest that public acceptance of this project has been carefully cultivated, with varying success, in a context of controversy and distrust [13,14]. The UK is no stranger to human tissue scandals. In 2001 it became known that the organs of deceased children were routinely kept for research purposes at Alder Hey Hospital in Liverpool and Bristol Royal Infirmary without their parents' knowledge or consent [15]. Public outrage led to a near moratorium on tissue banking and research. An expensive system of accreditation of specimen collections by the newly formed Human Tissues Authority eventually followed [16].

These three examples illustrate the increasingly visible role of large biobanking projects within biomedical research. They publicly announce the complexity of international collaborations, commercial involvement, and public–private partnerships that have become the norm in biomedical research. They also reveal major public concerns with the social and ethical implications of these projects: for privacy, indigenous identity and self-determination, ownership and control over body parts, and medical data for individuals and their families.

Traditionally, the interests of patient protection and investigative research have been served jointly by Research Ethics Boards and the guiding principles of biomedical ethics: respect for autonomy, beneficence, nonmaleficence, and

justice. These have been enacted through the process of obtaining informed consent, alongside measures to protect privacy and confidentiality of research participants and guard against discrimination. They have ensured, to a reasonable degree, the ethical enactment, legitimacy and public acceptance of research projects. Today, however, the demands of biomedical research, of the informed consent process and of patient protection, especially privacy, are beginning to jostle against each other uncomfortably. They are engaged in an increasingly public struggle and there appears to be ever-decreasing space in which to maneuver. If biomarker research is to proceed without unnecessary constraint toward improving patient care in a manner that individuals and society at large deem ethical, radical intervention is needed.

This chapter begins by outlining the diversity of social and ethical issues surrounding biomarker-related research and its applications. Focusing in on the ever more central process of banking of human biological materials and data, it then traces a recent trend toward large-scale population biobanks. Advances in genomics and computational biology have brought a whole raft of new questions and concerns to the domain of biomedical ethics. The peculiarities of these large biobanks, in the context of divergent legislative frameworks and increasing demands for international networking and collaboration, make such challenges ever starker. Privacy advocates argue that studies using DNA can never promise anonymity to their donors [17,18]. Prospective collections of human DNA and tissues seem doomed either to fail the demands of fully informed consent, or face the crippling financial and administrative burden of seeking repeated consent. Population biobanks are increasingly conceived as national resources [19]. Indigenous and wider publics are now vocal in their concerns about ownership, commercialization, privacy: essentially, about who uses their DNA, and how.

We do not set out here to design new governance frameworks for biobanking, or suggest the best ethical protocols for biomarker research, although these are sorely needed. The aim of this chapter is to suggest legitimate processes for so doing. In our search we veer outside the realm of ethics as traditionally conceived, into the domain of political science. New theories of deliberative democracy facilitate public participation in policy decision making; they aim for deliberation and communicative actions rather than strategic action; they have much to offer. Our conclusion is that ethics must embrace politics. Those involved in biomarker-related research are essential—as informers and participants in democratic public deliberation.

## BIOMARKERS, ETHICS, AND INVESTIGATIVE RESEARCH

What are the ethics of biomarkers? The application of biomarkers to assess the risks of disease, adverse effects of drugs, and organ rejection and for the development of targeted drugs and treatments is essential. Yet the search for biomarkers of exposure, effect and susceptibility to disease, toxic chemicals,

or pharmaceutical drugs raises many diverse ethical questions. Some of the most common debates surround the impact of developing predictive genetic tests as biomarkers for disease and their use in pharmacogenomics.

Neuroimaging, for example, promises much for the identification of biomarkers of diseases such as Alzheimer disease, offering earlier prediction capability than is currently available. But this technology may have unintended social or ethical consequences [20]. It could lead to reduced autonomy for patients at an earlier age if they are not allowed to work or drive. New tests may not be distributed equitably if certain health insurance plans refuse to include the test. Physicians may not be adequately prepared to counsel patients and interpret biomarker test results. Most important, the value of early prediction is questionable for a disease that as yet has no effective treatment.

Ethical concerns surrounding the use of biomonitoring in the workplace or by insurers have also been voiced within the health sciences literature and the wider media [21–23]. Biomarkers offer some hope for monitoring exposure to toxic chemicals in the workplace and protecting the health of employees. In an environment of high exposure to carcinogens, for example, a test could be developed to identify persons with an increased genetic risk of developing cancer from a specific dose, who could, for example, be excluded from the workplace. This would probably reduce the number of workers at risk of developing cancer.

There are, however, concerns about discrimination, as well as the reliability of such tests for measuring risk [22]. Is it right for an employer to exclude people from an occupation or workplace on genetic grounds rather than reducing carcinogen exposure for all employees? Some high-risk individuals could spend a lifetime working under high exposure to carcinogens and never develop cancer, whereas some low-risk co-workers might. There are also fears that insurance companies could use biomonitoring methods to exclude people from insurance opportunities on the basis of genetic risk* [21,24,25]. Confidentiality, interpretation of biomarker data, and the problem of obtaining genuinely informed consent emerge as the key *ethical tension zones* identified by occupational health stakeholders involved in one research project in Quebec [21].

The promise of pharmacogenomics and the ethical issues it raises have also been the subject of lengthy debate. The Human Genome Organization (HUGO) Ethics Committee released a statement in 2007 recognizing that "pharmacogenomics has the potential to maximize therapeutic outcomes and

*In the UK, such fears were voiced by a coalition of 46 organizations in a Joint Statement of Concern presented to a House of Commons Cross Party Group on February 14, 2006. The issue has also been the subject of much debate and policy analysis in the United States, given its system of private health insurance. The Genetic Information Nondiscrimination Act was passed in the U.S. House of Representatives on April 25, 2007. See U.S. National Institutes of Health fact sheet at http://www.genome.gov/page.cfm?pageID=10002328.

minimize adverse reactions to therapy, and that it is consistent with the traditional goals of public health and medical care to relieve human suffering and save lives" but noting many ethical concerns. These include the implications for developing countries and for those seeking access to therapy for neglected diseases, the impact on health care costs and on research priorities, and the fear that pharmacogenomics could reinforce genetic determinism and lead to stigmatization of individuals and groups [26].

Perhaps the widest range of social and ethical issues emerging from biomarker research, however, surround the process of collection, storage and use of human biological samples, and associated data for research purposes: to identify new biomarkers of exposure, effect, and susceptibility to disease and pharmacogenomic products. Many genetic and epidemiological studies require access to samples of annotated human blood, tissue, urine, or DNA and associated medical and lifestyle data. Often, they need large numbers of samples, and repeated sampling over many months or years. Often, the outcomes of the research are uncertain, technological advances in research methodologies are unpredictable, and neither can be anticipated.

This discussion focuses on ethical issues relating to the biobanking process. The development of large-scale population databases has rendered the ethics of this technology complex, controversial, and publicly visible. Debates about biobanking also reveal the increasing inadequacy of the old ethics guidelines, frameworks, and protocols that have served us for the last 50 years.

## POPULATION BIOBANKS AND THE CHALLENGE OF HARMONIZATION

The "banking" of human biological samples for research is not a twenty-first century phenomenon. Human tissue has been gathered and collected for at least 100 years. According to the U.S. National Bioethics Advisory Committee, by 1999, a total of 282 million unique tissue specimens were being held in the United States [27]. The term *biobank*, however, is relatively new. It appeared in *PubMed* for the first time in 1996 [28] and was not common nomenclature until the end of the decade. The sequencing of the human genome, advances in computational biology, and the emergence of new disciplines such as biomarker discovery, pharmacogenomics, and nutrigenomics have sparked unprecedented demand for samples of human blood, tissue, urine, DNA, and data. Three-fourths of the clinical trials that drug companies submit to the U.S. Food and Drug Administration for approval now include a provision for sampling and storing human tissue for future genetic analysis [3]. Biobanking has become deserving of its own name and has gained a dedicated society, the International Society for Biological and Environmental Repositories (ISBER), as well as two recent worldwide congresses: the WorldWide BioBank Summits (organized by IBM Healthcare and Life Sciences) and Biobanking and Biorepositories (organized by Informa Life Sciences).

The collection of human samples and data for research has not just accelerated, it has evolved. Four features differentiate biobanks today from those of 20 years ago: the emergence of large population-level biobanks, increased levels of commercial involvement, the desire for international collaborations requiring samples and data to be shared beyond national borders, and finally, the prospective nature of many emerging collections. The increased speed and scale of biobanking has contributed to the increasing public and academic concern with the ethical and social implications of this technology. The rules and practices of research and research ethics developed prior to the consolidation of these trends now inhibit the ability to construct biobanks and related research efficiently. They also provide ineffective protection for individuals and populations.

Small genetic databases containing a limited number of samples, attached to one research project focused on a specific disease were once standard. Such collections still exist: clinical collections within hospital pathology departments, and case- or family-based repositories for genetic studies of disease. Larger provincial, national, and international repositories are now increasingly common, as is the networking of existing collections. Provincial examples include the CARTaGENE project in Quebec (Canada). National disease-based biobanks and networks include the Alzheimer's Genebank, sponsored jointly by the U.S. National Institute on Ageing and the Alzheimer's Association. Examples of national or regional population-level biobanks include the Estonian Genome Project (Estonia), Biobank Japan (Japan), Icelandic Health Sector Database, UK Biobank (UK), Medical Biobank (Sweden) and the Singapore Tissue Network (Singapore). International collaborations include the European GenomEUtwin Project, a study of twins from Denmark, Finland, Italy, the Netherlands, Sweden, the UK, France, Australia, Germany, Lithuania, Poland, and the Russian Federation (http://www.genomeutwin.org/).

Levels of commercial involvement vary among these biobanks. The Icelandic Biobank was founded as a public–private partnership between the Icelandic government and deCODE Genetics. UmanGenomics was given exclusive rights to commercialize information derived from Sweden's Medical Biobank. The Singapore Tissue Network, by contrast, is publicly funded and will not be involved in commercialization. Biotechnology companies involved in biobanking include Newfound Genomics, which gathers DNA samples from volunteers across Newfoundland and Labrador.

Many of these large population databases are designed as research infrastructures. They do not focus on one specific disease or genetic characteristic, but contain samples from sick and healthy persons, often across several generations. DNA, blood, or other tissues are stored together with health and lifestyle data from medical records, examinations and questionnaires. These large population databases support research into complex gene interactions involved in multifactoral diseases and gene–gene and gene–environment interactions at the population level. There are few clinical benefits to indi-

vidual donors. Benefits are expected to be long term and often cannot be specified at the time of data and tissue collection. It is a major challenge to the requirement of informed consent that persons donating biological and data samples cannot know the specific future research purposes for which their donations will be used.

This proliferation of biobanks, and the advent of population-wide and transnational biobanking endeavors, has triggered a variety of regulatory responses. Some national biobanks have been created in association with new legislation. Estonia and Lithuania enacted the Human Genes Research Act (2000) and the Human Genome Research Law (2002), respectively, possibly motivated by the inadequacy of existing norms, a belief that genetic data and research require different regulation than traditional medicine, as well as by the need for democratic legitimacy [19]. The UK Human Tissue Act (2004), Sweden's Act on Biobanks (2002), and the Norwegian Act on Biobanks (2003) all pertain to the storage of biological samples [29]. Other national initiatives do not treat genetic data as exceptional. They remain dependent on a network of existing laws.

A series of national and international guidelines have also been produced, such as the World Medical Association's Declaration on Ethical Considerations Regarding Health Databases (2002) and guidelines from the U.S. National Bioethics Advisory Commission (1999) and the Council of Europe Committee of Ministers (2006). As with national regulation, however, the norms, systems, and recommendations for collection and processing of samples, informed consent procedures, and even the terminology for degrees of anonymization of data differ substantially between guidelines.

Anonymization terminology illustrates the confusion that can result from such diversity. European documents distinguish five levels of anonymization of samples [30]. Within European documents, *anonymized* describes samples used without identifiers but that are sometimes coded to enable reestablishing the identity of the donor. In most English Canadian and U.S. texts, however, *anonymized* means that the sample is irreversibly de-identified. Quebec follows the French system, distinguishing between reversibly and irreversibly anonymized samples. In European documents, *coded* usually refers to instances where researchers have access to the linking code. But the U.S. Office for Human Research Protection (OHRP) uses the word to refer to situations where the researcher does not have access to the linking code [30]. To add to the confusion, UNESCO has been criticized for creating new terms, such as *proportional* or *reasonable anonymity*, that do not correspond to existing categories [19].

Such confusion has led to repeated calls for harmonization of biobank regulations. The Public Population Project in Genomics consortium (P3G) is one attempt, a nonprofit consortium aiming to promote international collaboration and knowledge transfer between researchers in population genomics. With over 30 charter and associate members, P3G declares itself to have "achieved a critical mass to form the principal international body for the

harmonization of public population projects in genomics" (http://www.p3g.org).

Standardization also has its critics, notably among smaller biobanking initiatives. In 2006, the U.S. National Cancer Institute (NCI) launched guidelines spelling out best practices for the collection, storage, and dissemination of human cancer tissues and related biological specimens. These high-level guidelines are a move toward standardization of practice, following revelations in a 2004 survey of the negative impact of diverse laboratory practices on resource sharing and collaboration [31]. The intention is that NCI funding will eventually depend on compliance. The guidelines were applauded in *The Lancet* by the directors of major tissue banks such as Peter Geary of the Canadian Tumor Repository Network. They generated vocal concerns from other researchers and directors of smaller banks, many of which are already financially unsustainable. Burdensome informed consent protocols and the financial costs of infrastructural adjustments required were the key sources of concern. This is a central problem for biobanking and biomedical ethics: the centrality, the heavy moral weight, and the inadequacy of individual and voluntary informed consent.

## INFORMED CONSENT: CENTRALITY AND INADEQUACY OF THE IDEAL

Informed consent is one of the most important doctrines of bioethics. It was introduced in the 1947 Nuremberg Code, following revelations during the Nuremberg trials of Nazi medical experimentation in concentration camps. It developed through inclusion in the United Nations' Universal Declaration of Human Rights in 1948 and the World Medical Association's Declaration of Helsinki in 1964. Informed consent is incorporated in all prominent medical, research, and institutional ethics codes, and is protected by laws worldwide.

The purpose of informed consent in research can be viewed as twofold: to minimize harm to research subjects and to protect their autonomous choice. Informed consent requires researchers to ensure that research participants consent voluntarily to participation in research and that they be fully informed of the risks and benefits. The focus of informed consent has slowly shifted: from disclosure by health professionals toward the voluntary consent of the individual based on the person's understanding of the research and expression of their own values and assessments [32]. Simultaneously, health research has shifted from predominantly individual investigator-designed protocols with specific research questions to multiple investigator and institution projects that gather many forms of data and samples to understand complex phenomena and test emerging hypotheses. Further, informed consent as a protection for autonomy has become important in arguments about reproductive autonomy. Informed consent has been described as representing the dividing line between "good" genetics and "sinful" eugenics [32].

Unprecedented computational power now makes it possible to network and analyze large amounts of information, making large-scale population biobanks, and genetic epidemiology studies more promising than ever before. This research context raises the stakes of research ethics, making it more difficult to achieve individual consent and protect privacy while serving as the basis for strong claims of individualized and population health benefits. Large-scale biobanks and cohorts by their very nature cannot predict the exact uses to which samples will be put ahead of time. The ideal of voluntary participation based on knowledge of the research appears to require new informed consent for every emergent hypothesis that was not part of the original informed consent. The practicality of such an ideal approach is less than clear.

Genetic testing can also use samples that were not originally collected for genetic studies. Tissue biopsies collected for clinical diagnosis are now providing information for gene expression studies [33]. The precise nature of future technologies that will extract new information from existing samples cannot be predicted. On the other hand, seeking repeated consent from biobank donors is a costly and cumbersome process for researchers that can impede or even undermine research. Response rates for data collection (e.g., questionnaires) in any large population may vary between 50 and over 90%. The need for renewed consent could therefore reduce participation in a project and introduce selection bias [34]. Repeat consent may also be unnecessarily intrusive to the lives of donors or their next of kin.

Other forms of consent have been suggested and implemented for biobanking purposes. These include consent with several options for research use: presumed consent, broad consent, and blanket consent. Many European guidelines, including a memorandum from the Council of Europe Steering Committee on Bioethics, laws in Sweden, Iceland, and Estonia, and the European Society for Human Genetics guidelines, consider broad consent for unknown future uses to be acceptable as long as such future projects gain approval from Research Ethics Boards and people retain the right to withdraw samples at any time [30]. The U.S. Office for Human Research Protection went one step further in 2004, proposing to broaden the definition of nonidentifiable samples, upon which research is allowed under U.S. federal regulations without the requirement of informed consent.

The problem is that no informed consent mechanism—narrow or broad—can address all ethical concerns surrounding the biobanking of human DNA and data [35]. Such concerns include the aggregate effects of individual consent upon society as a whole and upon family and community members given the inherently "shared" nature of genetic material. If people are given full choice as to which diseases their samples can be used to research, and they choose only to donate for well-known diseases such as cancer, rare disease may be neglected. The discovery that Ashkenazi Jews may have particular mutations predisposing them to breast, ovarian, and colon cancer has generated fears that they could become the target of discrimination [36].

Concerns include irreconcilable trade-offs between donor desires for privacy (best achieved by unlinking samples), control over the manner in which their body parts and personal information are used (samples can be withdrawn from a biobank only if a link exists), and access to clinically relevant information discovered in the course of research. For some individuals and communities, cultural or religious beliefs dictate or restrict the research purposes for which their samples can be used. The Nuu-chah-nulth nations of Vancouver Island became angry in 2000 after discovering that their samples, collected years before for arthritis research, had been used for the entirely different purpose of migration research [37,38]. In some cases, a history of colonialism and abusive research makes a group demand that their samples be used for research that benefits their community directly.

Complete anonymization of samples containing human DNA is technically impossible, given both the unique nature of a person's DNA and its shared characteristics. Consequently, in 2003, the Icelandic Supreme Court ruled that the transfer of 18-year-old student Ragnhildur Gudmundsdottir's dead father's health data infringed her privacy rights: "The court said that including the records in the database might allow her to be identified as an individual at risk of any heritable disease her father might be found to have had—even though the data would be made anonymous and encrypted" [39]. Reasonable privacy protection in a biobanking context is tough to achieve, extending to the technological capacity to protect privacy through linked or unlinked anonymized samples without risk of error.

Informed consent cannot provide a basis for participants to evaluate the likelihood of benefit arising from their participation in a biobank when these merits are contested by experts. Critics of UK Biobank, for example, have little faith in the value and power of such prospective cohort studies, compared to traditional case–control studies, for isolating biomarkers and determining genetic risk factors. Supporters argue that the biobank will be a resource from which researchers can compile nested case–control studies. Critics claim that it will only be useful for study of the most common cancers, those that occur with enough frequency among donors. Others claim that even UK Biobank's intended 500,000 participants cannot provide reliable information about the genetic causes of a disease without a study of familial correlations [12].

Informed consent is inadequate as a solution for ensuring that the impacts of biobanking and related research will be beneficial to individuals and society, will uphold the autonomy of the individual, or will facilitate justice. Given its historical importance and bureaucratic and legal dependence [40], it is not surprising that informed consent remains central to contemporary discussions of ethical and social implications of biobanking, biomarkers, and biomedical research. Unfortunately, the substance of such debates centers upon the inadequacy of both ideal and current procedures. As Hoeyer points out in reference to Medical Biobank run by Uman Genomics in northern Sweden, informed consent offers an illusion of choice without real consideration of the

implications of such choices, "by constructing a diffuse arrangement of donors who can only be semiaccountable agents" [41].

## SCIENCE, ETHICS, AND THE CHANGING ROLE OF THE PUBLIC

Novel and innovative norms and models for biobank management have been proposed by bioethics, social science, and legal practitioners and theorists in recent years, in an attempt to deal with some of these issues. Alternative ethical frameworks based on social solidarity, equity, and altruism have been suggested [42,43]. These formed the basis for the recent Human Genome Organisation Ethics Committee statement on pharmacogenomics [26]. Onora O'Neil has argued for a two-tiered consent process in which public consent for projects is solicited prior to individual consent for donation of samples [44]. The charitable trust model has also been proposed for biobanking, as a way of recognizing DNA both as a common heritage of humanity and as uniquely individual, with implications for family members. "All information would be placed in a trust for perpetuity and the trustees overseeing the information would act on behalf of the people who had altruistically provided information to the population collection. They would be accountable to individuals but could also act as representatives for the community as a whole" [45].

It is not clear, however, whether such models could ever gain widespread public endorsement and legitimacy without direct public involvement in their design. Appeals to the need for community consultation [45] and scientific citizenship [46] may be more suited to the current mood. There is growing awareness globally, among government, policymakers, regulators, and advocacy groups alike, of the importance of public engagement, particularly in relation to emerging technologies.

In the UK, crises over bovine spongiform encephalopathy (BSE), otherwise known as "mad cow disease," and genetically modified (GM) crops have forced the government to proclaim the value of early public participation in decision making [47,48]. A statement by the UK House of Lords Select Committee in 2000 concluded that "today's public expects not merely to know what is going on, but to be consulted; science is beginning to see the wisdom of this and to move out of the laboratory and into the community to engage in dialogue aimed at mutual understanding" [49]. In Canada, the provincial government of British Columbia pioneered a Citizens' Assembly in 2003, charging 160 citizens with the task of evaluating the existing electoral system. A new BC Conversations on Health project aims to improve the health system by engaging in "genuine conversation with British Columbians" during 2007.

Indeed, public consultations have become the norm for soliciting public support for new technologies. In the UK these have included Weekends Away for a Bigger Voice, funded by the National Consumer Council in 2001 and the

highly publicized government-funded GM Nation consultation in 2002. In Canada, notable examples include the 1999 Canadian Citizen's Conference on Biotechnology and the 2001 Canadian Public Consultation on Xenotransplantation. In Denmark, more than 20 consensus conferences have been run by the Danish Board of Technology since 1989, on topics as diverse as genetically modified foods, electronic surveillance, and genetic testing [50]. In New Zealand, the government convened a series of public meetings in 2000 as part of its Royal Commission on genetic modification. UK Biobank marketing is careful to assert that the project has "undergone rigorous review and consultation at all levels" (http://www.ukbiobank.ac.uk/about/what.php).

Traditional public consultations have their limitations, however. Past examples of consultations have either been unpublicized or restricted to stakeholder involvement, undermining the claim to be representative of the full range of public interests [8]. Some critics suspect consultations of being a front to placate the public, a means of researching market strategy, and speeding product development [51] or as a mechanism for engineering consent [13]. GM Nation is one example of a consultation that has been criticized for "capture" by organized stakeholder groups and as misrepresentative of the public it aimed to consult [52].

## PROMISING FUTURE DIRECTIONS: PUBLIC CONSULTATION AND DELIBERATIVE DEMOCRACY

The use of theories and practices of deliberative democracy within such public consultations are a more recent and innovative trend. Deliberation stands in opposition to the aggregative market model of representational democracy and the strategic behavior associated with voting. It offers a model of democracy in which free and equal citizens exchange reasons through dialogue, and shape and alter their preferences collectively, and it is rapidly gaining in popularity, as evidenced by the growth of nonprofit organizations such as the Everyday Democracy (http://www.everyday-democracy.org), AmericaSpeaks (http://www.americaspeaks.org/), and National Issues Forums (http://www.nifi.org/) throughout the United States.

Origin stories of this broad deliberative democracy "movement" are as varied as its incarnations, and practice is not always as closely linked to theory as it could be. But most theorists will acknowledge a debt to the work of either (or both) Habermas and Rawls. Habermas's wider program of discourse ethics provides an overarching rationale for public deliberation [53]. This asserts that publicly binding norms can make a legitimate claim to rationality—and thus legitimacy—only if they emerge from free argument between all parties affected. Claims about what "any reasonable person" would accept as right can only be justified by putting them to the test. This is then a far cry from the heavily critiqued [13,54] model of public consultation as a tool for engendering public trust or engineering acceptance of a new technology.

By asking participants to consider the perspectives of everyone, deliberation orients individuals away from consideration of self-interest and toward consideration of the common good. Pellizzoni characterizes this *governance virtue* as one of three key virtues of deliberative democracy [55]. The second is *civic virtue*, whereby the process of deliberation produces more informed, active, responsible, cooperative, and fair citizens. The third is *cognitive virtue*, the notion that discussion oriented to understanding rather than success enhances the quality of decisions, gives rise to new or unarticulated points of view, and allows common understanding of a complex problem that no single person could understand in its entirety.

Deliberative democracy is not devoid of challenges when applied to complex issues of science and technology, rich as they can be in future uncertainties and potential societal impact. But it offers much promise as a contribution to biobanking policy that can provide legitimate challenges to rigidly structured research ethics.

## CONCLUSIONS

Biomarker research is greatly advanced by good-quality annotated collections of tissues, or biobanks. Biobanks raise issues that stretch from evaluation of the benefits and risks of research through to the complexity of informed consent for collections for which the research purposes and methods cannot be described in advance. This range of ethical and organizational challenges is not managed adequately by the rules, guidelines, and bureaucracies of research ethics. Part of the problem is that current research ethics leaves too much for the individual participant to assess before the relevant information is available. But many other aspects of biobanks have to do with how benefits and risks are defined, achieved, and shared, particularly those that are likely to apply to groups of individuals with inherited risks, or those classified as having risks or as being more amenable to treatment than others. These challenges raise important issues of equity and justice. They also highlight trade-offs between research efficiency and benefits, privacy and individual control over personal information, and tissue samples. These issues are not resolvable by appeal to an existing set of rules or ethical framework to which all reasonable people agree. Inevitably, governance decisions related to biobanks will need to find a way to create legitimate policy and institutions. The political approach of deliberative democracy may hold the most promise for well-informed and representative input into trustworthy governance of biobanks and related research into biomarkers.

### Acknowledgments

The authors thank Genome Canada, Genome British Columbia, and the Michael Smith Foundation for Health Research for their essential support.

We also appreciate the support and mutual commitment of the University of British Columbia, the British Columbia Transplant Society, Providence Health Care, and Vancouver Coastal Health, and all participants in the Biomarkers in Transplantation initiative.

## REFERENCES

1. Sigurdsson S (2001). Ying-yang genetics, or the HSD deCODE controversy. *New Genet Soc*, 20(2):103–117.
2. Sigurdsson S (2003). Decoding broken promises. Open Democracy. www.opendemocracy.net/theme-9-genes/article_1024.jsp (accessed June 1, 2004).
3. Abbott A (2003). DNA study deepens rift over Iceland's genetic heritage. *Nature*, 421:678.
4. Mannvernd, Icelanders for Ethics in Science and Medicine (2004). A landmark decision by the Icelandic Supreme Court: the Icelandic Health Sector Database Act stricken down as unconstitutional.
5. Merz JF, McGee GE, Sankar P (2004). "Iceland Inc."?: On the ethics of commercial population genomics. *Soc Sci Med*, 58:1201–1209.
6. Potts J (2002). At least give the natives glass beads: an examination of the bargain made between Iceland and deCODE Genetics with implications for global bioprospecting. *Va J Law Technol*, Fall, p. 40.
7. Pálsson G, Rabinow P (2001). The Icelandic genome debate. *Trends Biotechnol*, 19:166–171.
8. Burgess, M, Tansey J. (2009). Democratic deficit and the politics of "informed and Inclusive" consultation. In Einseidel E, Parker R (eds.), *Hindsight and Foresight on Emerging Technologies*. UBC Press, Vancouver, British Columbia, Canada.
9. Morrison Institute for Population and Resource Studies (1999). Human Genome Diversity Project: Alghero Summary Report. http://www.stanford.edu/group/morrinst/hgdp/summary93.html (accessed Aug. 2, 2007).
10. Harry D, Howard S, Shelton BL (2000). Indigenous people, genes and genetics: what indigenous people should know about biocolonialism. Indigenous Peoples Council on Biocolonialism. http://www.ipcb.org/pdf_files/ipgs.pdf.
11. BBC (2007). Volunteers join £61m health study. BBC News, July 16, 2007. http://news.bbc.co.uk/2/hi/uk_news/scotland/glasgow_and_west/6900515.stm (accessed Sept. 24, 2007).
12. Barbour V (2003). UK Biobank: a project in search of a protocol? *Lancet*, 361:1734–1738.
13. Peterson A (2007). Biobanks "engagements": engendering trust or engineering consent? *Genet Soc Policy*, 3:31–43.
14. Peterson A (2005). Securing our genetic health: engendering trust in UK Biobank. *Sociol Health Illness*, 27:271–292.
15. Redfern M, Keeling J, Powell M (2001). *The Royal Liverpool Children's Inquiry Report*. House of Commons, London.

16. Royal College of Pathologists' Human Tissue Advisory Group (2005). *Comments on the Draft Human Tissue Authority Codes of Practice 1 to 5*. The Royal College of Pathologists, London, Sept. 28.
17. Lin Z, Owen A, Altman R (2004). Genomic research and human subject privacy. *Science*, 305:183.
18. Roche P, Annas G (2001). Protecting genetic privacy. *Nat Rev Gene*, 2: 392–396.
19. Cambon-Thomsen A, Sallée C, Rial-Sebbag E, Knoppers BM (2005). Population genetic databases: Is a specific ethical and legal framework necessary? *GenEdit*, 3:1–13.
20. Illes J, Rosen A, Greicius M, Racine E (2007). Prospects for prediction: ethics analysis of neuroimaging in Alzheimer's disease. *Ann NY Acad Sci*, 1097: 278–295.
21. Caux C, Roy DJ, Guilbert L, Viau C (2007). Anticipating ethical aspects of the use of biomarkers in the workplace: a tool for stakeholders. *Soc Sci Med*, 65:344–354.
22. Viau C (2005). Biomonitoring in occupational health: scientific, socio-ethical and regulatory issues. *Toxicol Appl Pharmacol*, 207:S347–S353.
23. *The Economist* (2007). Genetics, medicine and insurance: Do not ask or do not answer? Aug. 23. http://www.economist.com/science/displaystory.cfm?story_id=9679893 (accessed Aug. 31, 2007).
24. Genewatch UK (2006). Genetic discrimination by insurers and employers: still looming on the horizon. Genewatch UK Report, Feb. 14. http://www.genewatch.org/uploads/f03c6d66a9b354535738483c1c3d49e4/GeneticTestingUpdate2006.pdf (accessed Aug. 31, 2007).
25. Rothenberg K, et al. (1997). Genetic information and the workplace: legislative approaches and policy challenges. *Science*, 275:1755–1757.
26. Human Genome Organisation Ethics Committee (2007). HUGO Statement on Pharmacogenomics (PGx): Solidarity, Equity and Governance. *Genom Soc Policy*, 3:44–47.
27. Lewis G (2004). Tissue collection and the pharmaceutical industry: investigating corporate biobanks. In Tutton R, Corrigan O. (eds.), *Genetic Databases: Socio-ethical Issues in the Collection and Use of DNA*. Routledge, London.
28. Loft S, Poulsen HE (1996). Cancer risk and oxidative DNA damage in man. *J Mol Med*, 74: 297–312.
29. Maschke KJ (2005). Navigating an ethical patchwork: human gene banks. *Nat Biotechnol*, 23:539–545.
30. Elger BS, Caplan AL (2006). Consent and anonymization in research involving biobanks. *Eur Mol Biol Rep*, 7:661–666.
31. Hede K (2006). New biorepository guidelines raise concerns. *J Nat Cancer Inst*, 98:952–954.
32. Brekke OA, Thorvald S (2006). Population biobanks: the ethical gravity of informed consent. *Biosocieties*, 1:385–398.
33. Cambon-Thomsen A (2004). The social and ethical issues of post-genomic human biobanks. *Nat Rev Genet*, 5:866–873.

34. Hansson MG, Dillner J, Bartram CR, Carlson JA, Helgesson G (2006). Should donors be allowed to give broad consent to future biobank research? *Lancet Oncol*, Mar, 7.
35. Burgess MM (2001). Beyond consent: ethical and social issues in genetic testing. *Nat Rev Genet*, 2:147–151.
36. Weijer C, Emanuel E (2000). Protecting communities in biomedical research. *Science*, 289:1142–1144.
37. Baird L, Henderson H (2001). Nuu-Chah-Nulth Case History. In Glass KC, Kaufert JM (eds.), *Continuing the Dialogue: Genetic Research with Aboriginal Individuals and Communities*, pp. 30–43. Proceedings of a workshop sponsored by the Canadian Commission for the United Nations Educational, Scientific, and Cultural Organization (UNESCO), Health Canada, and the National Council on Ethics in Human Research, pp. 26–27, Jan. 2001, Vancouver, British Columbia, Canada.
38. Tymchuk M (2000). Bad blood: management and function. Canadian Broadcasting Company, National Radio.
39. Abbott A (2004). Icelandic database shelved as court judges privacy in peril. *Nature*, 429:118.
40. Faden RR, Beauchamp TL (1986). *A History and Theory of Informed Consent*. Oxford University Press, New York.
41. Hoeyer K (2004). Ambiguous gifts: public anxiety, informed consent and biobanks. In Tutton R, Corrigan O (eds.), *Genetic Databases: Socio-ethical Issues in the Collection and Use of DNA*. Routledge, London.
42. Chadwick R, Berg K (2001). Solidarity and equity: new ethical frameworks for genetic databases. *Nat Rev Genet*, 2:318–321.
43. Lowrance W (2002). *Learning from Experience: Privacy and Secondary Use of Data in Health Research*. Nuffield Trust, London.
44. O'Neil O (2001). Informed consent and genetic information. *Stud History Philos Biol Biomed Sci*, 32:689–704.
45. Kaye J (2004). Abandoning informed consent: the case for genetic research in population collections. In Tutton R, Corrigan O (eds.), *Genetic Databases: Socio-ethical Issues in the Collection and Use of DNA*. Routledge, London.
46. Weldon S (2004). "Public consent" or "scientific citizenship"? What counts as public participation in population-based DNA collections? In Tutton R, Corrigan O (eds.), *Genetic Databases: Socio-ethical Issues in the Collection and Use of DNA*. Routledge, London.
47. Bauer MW (2002). Arenas, platforms and the biotechnology movement. *Sci Commun*, 24:144–161.
48. Irwin A (2001). Constructing the scientific citizen: science and democracy in the biosciences. *Public Understand Sci*, 10:1–18.
49. House of Lords Select Committee on Science and Technology (2000). *Science and Society*, 3rd Report. HMSO, London.
50. Anderson J (2002). Danish Participatory Models: Scenario Workshops and Consensus Conferences, Towards More Democratic Decision-Making. The Pantaneto Forum 6. http://www.pantaneto.co.uk/issue6/andersenjaeger.htm (accessed Oct. 22, 2007).

51. Myshkja B (2007). Lay expertise: Why involve the public in biobank governance? *Genet Soc Policy*, 3:1–16.
52. Rowe G, Horlick-Jones T, Walls J, Pidgeon N (2005). Difficulties in evaluating public engagement activities: reflections on an evaluation of the UK GM Nation public debate about transgenic crops. *Public Understand Sci*, 14:331–352.
53. Habermas J (1996). *Between Facts and Norms: Contributions to a Discourse Theory of Law and Democracy*. MIT Press, Cambridge, MA.
54. Wynne B (2006). Public engagement as a means of restoring public trust in science: Hitting the notes, but missing the music? *Commun Genet*, 9:211–220.
55. Pellizzoni L (2001). The myth of the best argument: power, deliberation and reason. *Br J Sociol*, 52:59–86.

# PATHODYNAMICS: IMPROVING BIOMARKER SELECTION BY GETTING MORE INFORMATION FROM CHANGES OVER TIME

Donald C. Trost, M.D., Ph.D.
*Analytic Dynamics, Niantic, Connecticut*

## INTRODUCTION

The purpose of this chapter is to introduce some approaches to thinking about biological dynamics. *Pathodynamics* is a term used by the author to describe a quantitative approach to disease that includes how the biological system changes over time. In many ways it is analogous to thermodynamics [1,2] in that it deals with the macroscopic, measurable phenotypic aspects of a biological system rather than the microscopic aspects such as those modeled in mathematical physiology [3]. For the purposes of this chapter, *macroscopic* will refer to measurements that do not involve the destruction of the biological system being studied but may vary in scale (i.e., cell, tissue, organ, and body). For example, clinical measurements would be macroscopic whether invasive or not, but histopathology, cell lysis, or genotype would be microscopic. Another way to view this is that when the dynamics of the system being studied result from something greater than the sum of the parts, usually from networks, it is macroscopic; otherwise, studying the parts of the system is microscopic. One of the problems in macroscopic biology is that many of the underlying system characteristics are immeasurable. This is why the biological

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

variation is just as important, if not more important than the mean behavior of a system because it represents the system changes that cannot be measured directly.

The term *parameter* is used rather loosely in biology and other sciences and has a very specific meaning in mathematics. For the purposes here, it will be defined in a statistical manner as a quantity which is a characteristic of the system that is not directly measurable but can be estimated from experimental data and will be represented by Greek symbols. Quantities being measured will be referred to as *variables*. The letter $y$ will designate a random, or stochastic, variable, one that has a probability distribution, and the letter $x$ will represent a deterministic, or controlled, variable, one that is measured "exactly," such as gender, temperature, and pH. The letters $s$, $t$, and $T$ will refer to time and will be either deterministic or random, depending on the experimental context. *Biomarkers* are really mathematical functions of parameters. In modeling, parameters appear as additive coefficients, multiplicative coefficients, and exponents. Sometimes parameters are deterministic, sometimes random, and sometimes functions of time or variables. Probably the most common parameter in biology is the population mean of a random variable. The sample mean is an estimate of this parameter; it is also a statistic, a quantity that is a function of the data. However, under the definitions above, some statistics do not estimate parameters and are known as *nonparametric statistics*. Nonparametric statistics are generally used for hypothesis testing and contain little or no mechanistic biological information [4–7].

## Analogy to Thermodynamics

Thermodynamics is a macroscopic view of physics that describes the flow of energy and the disorder of matter (i.e., entropy) [1,2]. The former is reflected in the first law of thermodynamics,which says that

$$\Delta\text{energy} = \Delta\text{work} + \Delta\text{heat}$$

which is basically a law of the conservation of energy. The second law has various forms but generally states that entropy changes of a system and its exterior never decrease. Aging and disease (at least some) may be examples of increasing entropy.

In chemistry, classical thermodynamics describes the behavior of molecules (particles) and changes in the states of matter. Equilibrium occurs at the state of maximum entropy [i.e., when the temperature is uniform throughout the (closed) system]. This requires constant energy and constant volume. In a system with constant entropy and constant volume, equilibrium occurs at a state of minimum energy. For example, if constant entropy occurs at a constant temperature, equilibrium occurs when the Helmholtz free energy is at its minimum. In all cases, the particles continue to move; this is a dynamics part. For nonequilibrium states, matter and energy will flow to reach an equi-

librium state. This is also a dynamics part. Modern thermodynamics is a generalization of classical thermodynamics that relates state variables for any system in equilibrium [2], such as electromagnetism and fluid dynamics. The open question is whether or not these concepts apply to biological systems.

For this chapter, the pathodynamics concept is that the states of a biological system can be measured and related in a manner similar to thermodynamics. At this time only the simplest relationships are being proposed. The best analogy seems to be viewing a biological system: in particular, a warm-blooded (constant-temperature) mammal as a particle in a high-dimensional space whose states are measured via (clinical) laboratory tests. The dimension of this space will be defined by the information contained in these biomarkers (see below). This particle is in constant motion as long as the system is alive and the microscopic behavior of the biology is generally unobservable, but the probabilitistic microscopic behavior of this particle is observable through its clinical states. The probabilistic macroscopic properties are described by the probability distributions of the particle. These distributions are inherently defined by a single biological system, and how they are related to population distributions for the same species is unknown. Hopefully, there will be some properties that are invariant among individuals so that the system behavior can be studied with large enough sample sizes to get sufficient information about the population of systems. One way to visualize this concept of pathodynamics is to imagine a single molecule moving in a homogeneous compressible fluid [8]. A drop of this fluid represents the probability distribution and is suspended in another fluid medium. The fluids are slightly miscible, so that there is no surface on the drop, but there is a cohesive (homeostasic) force that pulls the particles toward the center of the drop while the heat in the system tends to diffuse the particles out into the medium. Dynamic equilibrium occurs when the drop is at its smallest volume, as measured by levels of constant probability density. The conservation of mass is related to the total probability for the particle (mass = 1), which is analogous to the mass of the drop. The drop becomes distorted when external forces on it cause distortion of the shape and may even fragment the drop into smaller drops or change the drop so that holes appear. These external forces are due to factors such as external environment, disease, and therapies. A goal of pathodynamics is to infer the presence of an external force by observing the motion of the particle and finding the correspondence between the force and known causes.

## The Concept of Time

Since dynamics relates to the changes over time, some mention of time is appropriate. Everyone has a general concept of time, but in mathematical and physical thinking, time is a little more complicated [9,10]. There are two main classes of time: continuous (analog) and discrete (digital). Although physicists use the reversibility of time, which at least makes the mathematics easier, Prigogine [10] argues that at the microscopic level, the world is probabilistic

and that in probabilistic (stochastic) systems, time can only go forward. Furthermore, he argues that "dynamics is at the root of complexity that [is] essential for self-organization and the emergence of life."

There are many kinds of time. In the continuous category are astronomical time (ordinary time), biological time (aging), psychological time, thermodynamic time, and information time [11]. Discrete time is more difficult to imagine. Whenever digital data are collected, the observations occur only at discrete times, usually equally spaced. Even then, the process being observed runs in continuous time. This is probably the most common case. However, discrete time is really just a counting process and occurs naturally in biology. Cell divisions, heartbeats, and respirations are a few of the commonly observed discrete biological clocks. In this chapter only continuous-time processes are discussed. An example of discrete-time pathodynamics can be found elsewhere [12].

## BROWNIAN MOTION

### Diffusion

One of the most basic continuous stochastic processes is Brownian motion. The concept of Brownian motion was born when a biologist named Robert Brown observed pollen under a microscope vibrating due to interactions with water molecules [13]. This concept should be familiar to biologists. Brownian motion has been modeled extensively [14,15]. Standard Brownian motion ($B_t$) is a Gaussian stochastic process with mean and variance

$$\mu_t = 0$$
$$\sigma_t^2 = t$$

respectively. In the laboratory, Brownian motion is a good model for diffusion (e.g., immunodiffusion), where the concentration of the protein is related to the time of observation and the diffusion coefficient. In an open system where time goes forever or when the particles have no boundary, the diffusion of a particle is unbounded because the variance goes to infinity. To make it a useful concept in biology, the particle needs to be either in a closed container or stabilized by an opposing force.

### Homeostasis: Equilibrium Pathodynamics

In biological systems, diffusion occurs only on a microscopic level and is not usually measurable macroscopically. However, in a probabilistic model of pathodynamics, the particle representing person's clinical health state can be thought of as a microscopic object (e.g., a molecule in a fluid drop), while the probability distribution is the macroscopic view. The reader should note that

the concepts of microscopic biology as defined in the Introduction and this imaginary microscopic particle of pathodynamics are different uses of the *microscopic/macroscopic* dichotomy. The Ornstein–Uhlenbeck (OU) stochastic process [16] is a stationary Gaussian process with conditional mean and variance given $y_0$:

$$\mu_t = \mu + e^{-\alpha t}(y_0 - \mu)$$
$$\sigma_t^2 = \sigma^2\left(1 - e^{-2\alpha t}\right)$$

respectively, where $y_0$ is the baseline value, $\mu$ is the equilibrium point (mean of $y_t$), and $\sigma^2$ is the variance of $y_t$. The autocorrelation between two measurements of $y$ is

$$\mathrm{Corr}[y_s, y_t] = e^{-\alpha(t-s)}$$

for times $t \geq s$.

In thermodynamic terms, the average fluctuation of a particle is

$$\frac{d\mu_t}{dt} = -\alpha(\mu_t - \mu)$$
$$d\mu_t = -\alpha(\mu_t - \mu)dt$$

Read "$dx$" as a small change in $x$. This ordinary differential equation (ODE) is analogous to the stochastic differential equation (SDE) [14,15] that generates the OU process,

$$dy_t = -\alpha(y_t - \mu)dt + \sqrt{2\alpha}\,\sigma\, dB_t$$

which has a biological variation term that is driven by Brownian motion. In statistical physics, this is called the *Langevin equation*, or the equation of motion for a Brownian particle [17].

Here a link between thermodynamics and pathodynamics will be attempted. Using the Einstein theory of fluctuations [2,17], the change in entropy is

$$\Delta S = -\frac{1}{2\sigma^2}(y_t - \mu)^2$$

Since this quantity is always zero or negative, this says that in the equilibrium state the entropy is decreasing with time, which suggests that there is some organizing force acting on the system. This is the homeostatic force

$$F_H = \frac{1}{\sigma^2}(y_t - \mu)$$

In statistical terms under near-equilibrium conditions, the drag in the system increases as the autocorrelation increases or as the variance decreases.

When the time rate of change of $\Delta S$ is equal to the product $-JF$, where $J$ is the current or flow of the particles in the system, the system is in an equilibrium state. Solving this equation gives

$$J_H = \frac{1}{2}(y_t - \mu)$$

which is the entropy current. In homeostasis the average force and the average flow are zero and the average of the change in entropy is –½. This can all be generalized to multivariate biomarkers by making the measurements and their means into vectors **y** and μ, respectively, and by making α and $\sigma^2$ into symmetric positive definite matrices **A** and Σ, respectively. The usual thermodynamic parameters are embedded in these statistical parameters and can be determined as needed for specific biological uses.

To illustrate this physical analog a little further, suppose that a drop of particles were placed in a well in the center of a gel, the external medium, and allowed to diffuse for a fixed time $t_H$. If Fick's law applies, there is a diffusion coefficient $D$ and the Stokes–Einstein relation holds such that

$$D = \frac{k_B T}{\gamma}$$

where $k_B$ is Boltzmann's constant, $T$ is the absolute temperature of the system, and γ is the viscosity (friction) coefficient. Now suppose that the particles are charged and an electrical field is applied radially inward at $t_H$ so that the field strength is proportional to the distance from the center of the sample and is equal to the force of the diffusion. This means that the distribution of the particles is in steady state and that the particles experience no acceleration. This leads to the Langevin equation with the friction coefficient in the system at $\gamma = 1/\alpha\sigma^2$. It turns out that at this equilibrium state $\sigma^2 = 2Dt_H$ and then by substitution

$$\alpha = \frac{1}{2k_B t_H T} \quad \text{and} \quad \sqrt{\alpha}\,\sigma = \sqrt{\frac{D}{k_B T}}$$

As long as $T$ is constant, α is constant and inversely proportional to $T$, which may represent physical temperature or some biological analog but is assumed to be constant as well; and $\alpha\sigma^2$ is proportional to $D/T$.

## Signals of Change: Nonequilibrium Pathodynamics

Changes from equilibrium are the signals of interest in pathodynamics. In the simplest case, the signal can be an observation that occurs outside the dynamic reference interval ("normal limits"). This interval can be constructed by estimating its endpoints at time $t$ using

$$\mu + e^{-\alpha(t-s)}(y_s - \mu) \pm 1.96\sigma\sqrt{1 - e^{-2\alpha(t-s)}}$$

where $s$ is the time of a previous measurement and would be zero if it is the baseline measurement. In either case, two measurements are required to identify a dynamic signal when the parameters do not change with time. Some care needs to be exercised when estimating this interval [18]. The interval clearly will be shorter than the usual one, $\mu \pm 1.96\sigma$, will be different for each person, and will be in motion unless there is no autocorrelation. A value outside this interval would indicate a statistically significant deviation from homeostasis with a probability of a false positive being 0.05 for each pair of time points. Simultaneous control of the type I error requires multivariate methods.

Nonequilibrium states might be modeled by allowing one or more parameters to change with time. If $\mu$ is changing with time, this is called *convection*, meaning that the center of gravity of the system is changing, resulting in a flow, or trajectory. If $\alpha$ or $\sigma$ is changing, the temperature or the diffusion properties are changing. When the particle is accelerating, an acceleration term needs to be added to the Langevin equation. The Fokker–Planck equation [17] provides a way to construct the steady-state probability distributions of the particle along with the transition probability states for the general Langevin equation. It is possible for a new equilibrium distribution to occur after the occurrence of a disease or after a therapeutic intervention, which would indicate a permanent residual effect. In this paradigm, a "cure" would occur only if the particle distribution returned to the healthy normal homeostasis state. The observation of the dynamics of the particle may suggest that an external (pathological) force is acting on the system. Any changes in "thermodynamic" variables may form patterns that lead to diagnostic criteria. The construction of optimal criteria and the selection and measurement of biomarkers are discussed in the next section.

## INFORMATION FROM DATA

### Parameter Estimation and the Mean Squared Error

The *mean squared error* (MSE) is just what it says: the average of the squared difference between the estimated parameter and the true parameter. The square root of the MSE is often referred to by engineers as the *root mean square* (RMS). In general,

$$\mathrm{MSE}\left(\hat{\theta}\right) = \mathrm{Variance}\left(\hat{\theta}\right) + \mathrm{Bias}\left(\hat{\theta}\right)^2$$

The accent mark over the parameter means that it is estimated from the data; a parameter without the accent mark is a theoretical, or unknown, quantity.

The MSE shows up in statistical estimation theory in the Cramér–Rao inequality [19–21]:

$$\mathrm{MSE}\left(\hat{\theta}\right) I(\theta) \geq 1$$

The second term, $I(\cdot)$ in this inequality is called the *Fisher information*. The statistician tries to construct an estimator of the unknown parameter $\theta$ to make this product as close to 1 as possible. An estimator that makes the product equal to 1 is an unbiased, minimum-variance estimator (UMV) and derives the maximum Fisher information from the data. Unfortunately, these estimators do not always exist for a given situation. The estimation strategy is usually to minimize the MSE when the UMV estimator is not available. There may be situations where a biased estimator will produce an MSE smaller than the unbiased approach.

The maximum likelihood (ML) estimator is probably the most common estimation procedure used today, although it requires that a model for the probability can be written explicitly. It has the nice property that a function of an ML estimate is the ML estimate of the function, allowing the ML parameter estimates to be plugged directly into the function. This property may not be true for other estimation procedures. Since the mathematics is easier when a sample size ($n$) is large (a euphemism for asymptotic, i.e., when $n$ is "near" infinity), ML estimates are favored because for large samples, they are also Gaussian and UMV even when the underlying distribution of the data is not Gaussian. However, an extremely large $n$ may be needed to get these properties, a fact often overlooked in statistical applications. This is particularly true when estimating some parameter other than the mean, such as the standard deviation or the coefficient of variation (CV).

For the analysis of experiments, ordinary least squares (OLS) estimation is used most often. Analysis of variance (ANOVA) and ordinary regression are special cases of OLS estimation. OLS estimation is unbiased if the correct model is chosen. If the residuals (difference between the observation and the model) are Gaussian, independent, and have the same variance, then OLS is equivalent to ML and UMV. When the variance is not constant, such as in experiments where the analytical variation (measurement error) is related to the size of the measurement, various methods, such as a logarithm transformation, need to be used to stabilize the variance; otherwise, OLS does not have these good properties, and signals can be missed or falsely detected.

## Fisher Information

Fisher information ($I$) provides a limit on how precisely a parameter can be known from a single measurement. This is a form of the uncertainty principle seen most often in physics. In other words, what statistic gives you the best characterization of your biomarker? When measurements are independent, such as those measured on different experimental units, the information is

additive, making the total information equal to $nI$. If you can afford an unlimited number of independent measurements, which is infinite information, the parameter can be known exactly, but this is never the case. For dynamic measurements (i.e., repeated measurements over time on the same experimental unit), although the measurement errors are generally independent, the measurements are not. It is usually cheaper to measure the same subject at multiple times than to make the same number of measurements, one per subject. In addition, information about the time effects within a person cannot be obtained in the latter case. This is key information that is not available when the measurements are not repeated in time. As illustrated in the examples below, this additional information can have major effects on the characteristics of the signal and the ability to detect it.

The Fisher information is generally a theoretical quantity for the lower bound of the MSE that involves a square matrix of the expectations of pairwise second-order partial derivatives of the log-likelihood. For those interested, this theory can be found in many books [19–21] and is not covered here. However, some of the results given below for the OU model are used to illustrate the information gain. For this purpose, an ANOVA model will be used for the means at equally spaced time points $\mu_t$. In such an experimental design, time changes can be detected using specific orthogonal (independent) contrasts. These can be written as follows:

Constant effect: $\frac{1}{\sqrt{5}}(\mu_0 + \mu_1 + \mu_2 + \mu_3 + \mu_4)$

Linear effect: $\frac{1}{\sqrt{10}}(-2\mu_0 - \mu_1 + \mu_3 + 2\mu_4)$

Quadratic effect: $\frac{1}{\sqrt{14}}(2\mu_0 - \mu_1 - 2\mu_2 - \mu_3 + 2\mu_4)$

Cubic effect: $\frac{1}{\sqrt{10}}(-\mu_0 + 2\mu_1 - 2\mu_3 + \mu_4)$

Quartic effect: $\frac{1}{\sqrt{70}}(\mu_0 - 4\mu_1 + 6\mu_2 - 4\mu_3 + \mu_4)$

The Fisher information for these sums of means assuming that $\mu_0 = 0$ estimated from an OU process is compared in Figure 1 to the information when the time points are independent. To reduce the complexity in graphing the relationships, the OU information is plotted against $\lambda = \alpha \cdot \Delta t$. If the approximate value of $\alpha$ is known from previous experiments, $\Delta t$ can be chosen to achieve improved efficiency in detecting the desired time relationship of the experimental response. All efficiency appears to be lost for $\lambda$ greater than 4, while the information appears to increase exponentially when it is less then 1. The maximum relative efficiencies are 1, 1.5, 2.5, 9, and 69 for the constant, linear, quadratic, cubic, and quartic effects, respectively. Estimating a constant response is not at all efficient using dynamic measures. It is not immediately

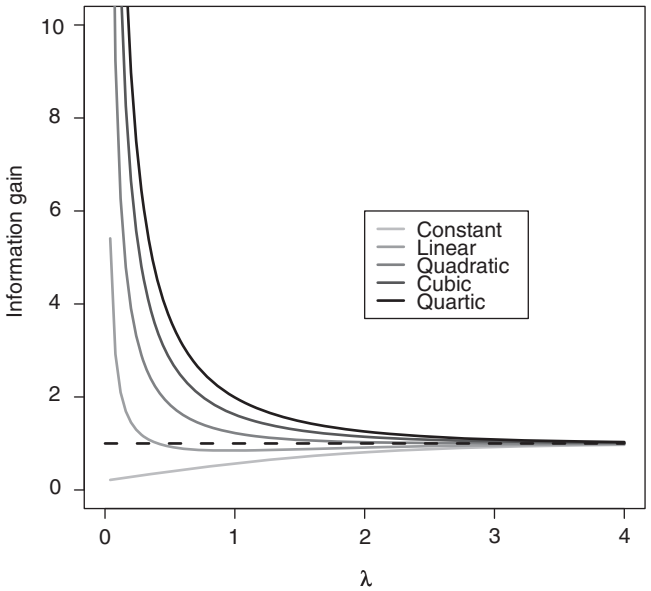


**Figure 1** Fisher information gain from dynamic measurements relative to independent measurements as a function of λ. (*See insert for color reproduction of the figure.*)

clear why a linear relationship would be less efficient than the others, although it is well known that the best design in this case is to put the observations at beginning and end, with none in the middle. Since it was assumed that the beginning was zero, only one point is needed to estimate a linear response. For in vivo biomarkers, it would be difficult to imagine a constant or linear response. For those not inclined mathematically, it should be noted that all continuous functions of time can be approximated with a polynomial in time if the degree is sufficiently large. This means that additional contrasts, extensions of those above, can be used to get a better estimate of the time relationship as long as the degree is less than the number of time points. If the change with time is not well modeled by low-degree polynomials, a regression using the specific time-dependent function of interest should be used to save error degrees of freedom. It seems apparent from Figure 1 that the Fisher information of dynamic measurements is increased as the degree of polynomial or autocorrelation increases.

Besides obtaining additional information about the curvature of the mean time effect, the autocorrelation parameter α contains information that is seldom estimated or used. For the OU process, this parameter combined with the variance provides a measure of the biological variation. It is proportional to the homeostatic force needed to maintain the dynamic equilibrium. The Fisher information for α is mathematically independent of the mean and vari-

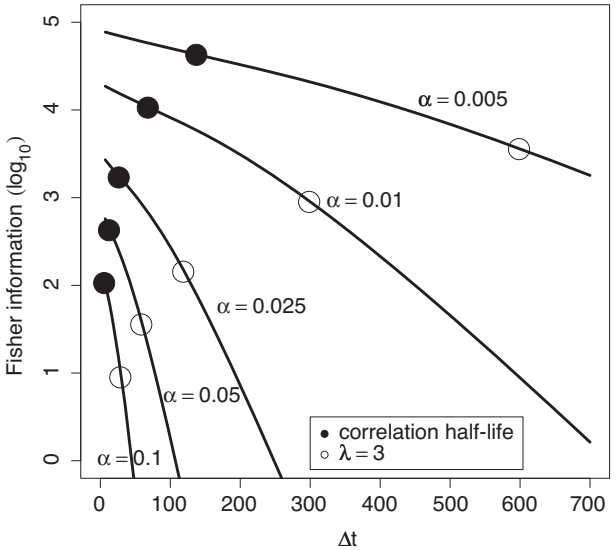


**Figure 2** Fisher information for α as a function of time between measurements ($\Delta t$).

ance. The relationship between the base 10 logarithm of α-information as a function of $\Delta t$ for various values of α is shown in Figure 2. The closed circles are the autocorrelation half-life for each α and the open circles represent a loose upper bound for $\Delta t$ at $\lambda = 3$, which is an autocorrelation of 0.05. Those serious about capturing information about α should seriously consider $\lambda < 1$. For a given α, the information approaches $2/\alpha^2$ as $\Delta t$ goes to zero. Obviously, the larger α is, the less information about the autocorrelation is available in the data, necessitating larger $m$ or smaller $\Delta t$ to get equivalent information. The time units are arbitrary in the figure but must match the time unit under which α is estimated.

## Shannon Information

Shannon information is about how well a signal is communicated through some medium or channel; in this case it is the biological medium. The measurement of variables to estimate parameters that indicate which signal was transmitted is the signal detection process. Pathological states and the body represent a discrete communication system where the disease is the signal transmitted, which may affect any number of biological subsystems that act as the communication channels, and the biomarker is the signal detected in one or more biomarker measurements. The disease is then diagnosed by partitioning the biomarker space into discrete, mutually exclusive regions ($R$) in a way that minimizes the signal misclassification. Information, or communication, theory is usually applied to electronic systems that can be designed to take

advantage of the optimal properties of the theory. In biology, it is mostly a reverse-engineering task. The signal is the health or disease state of the individual, which is transmitted through various liquid and solid tissues with interacting and redundant pathways. In this paradigm, biomarkers are the signal detection instruments and algorithms. Rules, usually oversimplified, are then constructed to determine which signal was sent based on the biomarker information. An elementary background in information theory is given by Reza [22].

Shannon information is really a measure of uncertainty or entropy and describes how well signals can be transmitted through a noisy environment. Probability models are the basis of this type of information which complements Fisher information rather than competing with it. The general framework for Shannon information follows.

As a reminder, Bayes' theorem, where $P[\cdot]$ is a probability measure and $P[A|B] = P[\text{event } A \text{ given event } B]$, states that $P[A|B] = P[A \text{ and } B]/P[B]$. In this particular representation, some elemental probabilities will be used: the probability that the signal was sent: for example, the prevalence or prior probability,

$$\pi_i = P[S_i]$$

the probability that the signal was received,

$$q_j = P[D_j]$$

and the probability that a particular diagnosis was made given that the signal was sent,

$$\begin{aligned} q_{ij} &= P[D_j | S_i] \\ &= \int_{R_j} dP_i(\mathbf{Y}) \end{aligned}$$

The last expression is rather ominous, but in words it is the probability that the multivariate biomarker $\mathbf{Y}$, a list (vector, array) of tests, is in the region $R_j$ given that signal $S_i$ was sent through a noisy channel, properly modeled by the probability function $P_i$. This $P$ can be either continuous or discrete or both and is generally a function of unknown parameters and time. It reflects both the biological variability and the analytical variability. Since this function contains the parameters, it is where the Fisher information applies. In its crudest form, $q_{ij}$ is just a proportion of counts. Although the latter is simple and tractable for a biologist, it is subject to losing the most information about the underlying signal and does not lend itself readily to the incorporation of time used in pathodynamics. Generally, any categorization of continuous data will result in information loss.

Table 1 shows the correspondence between signals ($S$) and diagnoses, or decisions ($D$), in terms of the probability structure. If the number of decision

**TABLE 1 Noisy Discrete-Signal Multichannel Communication Probability Table**

| | Decision | | | | |
|---|---|---|---|---|---|
| Signal | $D_1$ | $D_2$ | … | $D_\ell$ | Total |
| $S_1$ | $\pi_1 q_{11}$ | $\pi_1 q_{12}$ | … | $\pi_1 q_{1\ell}$ | $\pi_1$ |
| $S_2$ | $\pi_2 q_{21}$ | $\pi_2 q_{22}$ | … | $\pi_2 q_{2\ell}$ | $\pi_2$ |
| ⋮ | ⋮ | ⋮ | | ⋮ | ⋮ |
| $S_k$ | $\pi_k q_{k1}$ | $\pi_k q_{k2}$ | … | $\pi_k q_{k\ell}$ | $\pi_k$ |
| Total | $q_1$ | $q_2$ | … | $q_\ell$ | 1 |

classes is not equal to the number of underlying signals, inefficiencies are likely to occur. However, in biology it is not always possible to optimize this, especially if some of the signals are unknown. The mathematical objects under the control of the biologist are **Y**, $P$, $R$, and $D$. Since the rest of this book is mostly about **Y**, and in most cases $y$, a single biomarker, this chapter is mostly about $P$, $R$, and $D$. In other words, the biologist can optimize the $q_{ij}$'s only when $D$ is specified. The one exception is when the experiment can be designed so that the $\pi_i$'s are known and individuals can be selected to make them equally likely.

A few calculations from information theory are presented here. For applications below, $S_1$ will represent the "normal" state and $D_1$ will represent the "normal" classification. The others will represent "abnormal" or pathological states and classifications. Pharmacological or other therapeutic effects will be considered as abnormal states except when a cure results (i.e., the subject reverts back to the normal state).

The idea of entropy comes from thermodynamics, and it has been shown that thermodynamic entropy and Shannon entropy (information) are related. There are three primary types of Shannon entropy (average uncertainty): the entropy of the source,

$$H(S) = -\sum_{i=1}^{k} \pi_i \log_2 \pi_i$$

the entropy in the receiver,

$$H(D) = -\sum_{i=1}^{\ell} q_j \log_2 q_j$$

and the communication system entropy,

$$H(S, D) = -\sum_{i=1}^{k} \sum_{j=1}^{\ell} \pi_i q_{ij} \log_2 \pi_i q_{ij}$$

The base 2 logarithm is used here because the units of information are in bits (binary digits). With an ideal, noise-free communication channel, $H(S) = H(D) = H(S,D)$. This means that $q_{ii} = 1$ and $q_{ij} = 0$ for all $i \neq j$.

**TABLE 2 Biological Channel Capacity (Bits) by the Number of Possible Signals Transmitted and the Probability of an Abnormal Signal Being Sent**

| | Probability of an Abnormal Signal ($1 - \pi_1$) | | | | | | |
|---|---|---|---|---|---|---|---|
| Signals ($k$) | 0.99 | 0.95 | 0.9 | 0.8 | 0.7 | 0.6 | 0.5 |
| 2 | 0.08 | 0.29 | 0.47 | 0.72 | 0.88 | 0.97 | 1.00 |
| 3 | 1.07 | 1.24 | 1.37 | 1.52 | 1.58 | 1.57 | 1.50 |
| 5 | 2.06 | 2.19 | 2.27 | 2.32 | 2.28 | 2.17 | 2.00 |
| 10 | 3.22 | 3.30 | 3.32 | 3.26 | 3.10 | 2.87 | 2.58 |
| 25 | 4.62 | 4.64 | 4.60 | 4.39 | 4.09 | 3.72 | 3.29 |
| 50 | 5.64 | 5.62 | 5.52 | 5.21 | 4.81 | 4.34 | 3.81 |
| 75 | 6.23 | 6.19 | 6.06 | 5.69 | 5.23 | 4.70 | 4.10 |
| 100 | 6.64 | 6.58 | 6.44 | 6.03 | 5.52 | 4.95 | 4.31 |
| | 0.4 | 0.3 | 0.2 | 0.1 | 0.05 | 0.01 | 0.001 |
| 2 | 0.97 | 0.88 | 0.72 | 0.47 | 0.29 | 0.08 | 0.01 |
| 3 | 1.37 | 1.18 | 0.92 | 0.57 | 0.34 | 0.09 | 0.01 |
| 5 | 1.77 | 1.48 | 1.12 | 0.67 | 0.39 | 0.10 | 0.01 |
| 10 | 2.24 | 1.83 | 1.36 | 0.79 | 0.44 | 0.11 | 0.01 |
| 25 | 2.80 | 2.26 | 1.64 | 0.93 | 0.52 | 0.13 | 0.02 |
| 50 | 3.22 | 2.57 | 1.84 | 1.03 | 0.57 | 0.14 | 0.02 |
| 75 | 3.45 | 2.74 | 1.96 | 1.09 | 0.60 | 0.14 | 0.02 |
| 100 | 3.62 | 2.87 | 2.05 | 1.13 | 0.62 | 0.15 | 0.02 |

A measure of the information transmitted is

$$I(S, D) = \sum_{i=1}^{k} \sum_{j=1}^{\ell} \pi_i q_{ij} \log_2 \frac{q_{ij}}{q_j}$$

When $q_{ij} = q_j$, for $i \neq j$, the log term is zero; no information about $S_i$ is transmitted.

The channel capacity ($C$) is then defined as the maximum of $I(S,D)$ over all possible values of the $\pi_i$'s. For a noise-free system, this maximum occurs when all signals are equally likely (i.e., $C = \log_2 k$). In the $2 \times 2$ case, $C = 1$ bit; in general, for biological systems, $C = I(S,D)$, because the $\pi_i$'s are fixed by nature. Table 2 shows the maximum amount of information possible for the case where $\pi_i = (1 - \pi_1)/(k - 1)$ for the abnormal signals. If the biologist has some idea of the prevalence of the signal of interest, this table can give some idea of how feasible (futile) searching for biomarkers might be. This table shows that when several relatively rare signals are being sent via the same channel, very little information is available even in the best of conditions. Table 3 is a slightly different way of looking at the same question. These are the values in Table 2 divided by $\log_2 k$. Similar theory is available when $S$ and

**TABLE 3 Biological Communication Efficiency (%) by the Number of Possible Signals Transmitted and the Probability of an Abnormal Signal Being Sent**

| | Probability of an Abnormal Signal ($1 - \pi_1$) | | | | | | |
|---|---|---|---|---|---|---|---|
| Signals ($k$) | 0.99 | 0.95 | 0.9 | 0.8 | 0.7 | 0.6 | 0.5 |
| 2 | 8.1 | 28.6 | 46.9 | 72.2 | 88.1 | 97.1 | 100.0 |
| 3 | 67.6 | 78.0 | 86.4 | 96.0 | 99.8 | 99.1 | 94.6 |
| 5 | 88.8 | 94.2 | 97.7 | 100.0 | 98.2 | 93.5 | 86.1 |
| 10 | 96.9 | 99.3 | 100.0 | 98.1 | 93.3 | 86.5 | 77.8 |
| 25 | 99.5 | 100.0 | 99.0 | 94.5 | 88.1 | 80.1 | 70.9 |
| 50 | 99.9 | 99.6 | 97.8 | 92.4 | 85.3 | 76.9 | 67.5 |
| 75 | 100.0 | 99.3 | 97.2 | 91.3 | 83.9 | 75.4 | 65.9 |
| 100 | 100.0 | 99.1 | 96.9 | 90.7 | 83.1 | 74.5 | 64.9 |
| | 0.4 | 0.3 | 0.2 | 0.1 | 0.05 | 0.01 | 0.001 |
| 2 | 97.1 | 88.1 | 72.2 | 46.9 | 28.6 | 8.1 | 1.1 |
| 3 | 86.5 | 74.5 | 58.2 | 35.9 | 21.2 | 5.7 | 0.8 |
| 5 | 76.3 | 63.8 | 48.3 | 28.8 | 16.6 | 4.3 | 0.6 |
| 10 | 67.4 | 55.2 | 40.8 | 23.7 | 13.4 | 3.4 | 0.4 |
| 25 | 60.4 | 48.6 | 35.3 | 20.0 | 11.1 | 2.7 | 0.3 |
| 50 | 57.0 | 45.5 | 32.7 | 18.3 | 10.0 | 2.4 | 0.3 |
| 75 | 55.5 | 44.1 | 31.5 | 17.5 | 9.6 | 2.3 | 0.3 |
| 100 | 54.5 | 43.2 | 30.8 | 17.0 | 9.3 | 2.2 | 0.3 |

$D$ are continuous rather than discrete, but this does not seem relevant for biomarkers since diseases are treated as discrete entities [23]. Some might argue that hypertension and hypercholesterolemia are continuous diseases, but under the paradigm in this chapter, they are just biomarkers, for which some arguments can be made that they are surrogates for particular diseases.

## MEASURES OF DIAGNOSTIC PERFORMANCE

Much has been written about this topic. In this section we point out some issues with the current wisdom, and the reader can decide if changes are needed. In addition, the impact of the measurement of time changes on Shannon information is discussed here. In standard statistical hypothesis testing, there are two types of inference errors: type I ($\alpha$) is where the null hypothesis ($H_0$) is rejected when it is true, and type II ($\beta$) is where the null hypothesis ($H_0$) is accepted when it is false. With the null hypothesis symbolized by signal 1 ($S_1$) and the alternative hypothesis ($H_1$) by signal 2 ($S_2$), Table

**TABLE 4 Hypothesis-Testing Outcome Probabilities**

| | Decision | |
|---|---|---|
| Signal | $D_1$ | $D_2$ |
| $S_1$ | $1 - \alpha$ | $\alpha$ |
| $S_2$ | $\beta$ | $1 - \beta$ |

**TABLE 5 Two-Channel Signal Detection Probabilities**

| | Decision | | |
|---|---|---|---|
| Signal | $D_1(H_0)$ | $D_2(H_1)$ | $P[S_i]$ |
| $S_1$ | $\pi_1(1 - \alpha)$ | $\pi_1\alpha$ | $\pi_1$ |
| $S_2$ | $\pi_2\beta$ | $\pi_2(1 - \beta)$ | $\pi_2$ |
| $P[D_j]$ | $q_1$ | $q_2$ | 1 |

4 shows the error relationship for the decision that signal 1 was detected ($D_1$) or signal 2 was detected ($D_2$). In customary frequentist statistical practice, $\alpha$ is chosen to be fixed at 0.05, and a fixed sample size ($n$) is estimated to attain a power = $1 - \beta$ for some specified value in the interval (0.75, 0.95). The reality in biomedical science is that the power is fudged to get a sample size that the scientist can afford. It should really be set at some conventional value like $\alpha$, so that experiments are comparable and have the same probability of missing the signal. Everything in hypothesis testing is focused on controlling $\alpha$ and letting $\beta$ float. Unless a biomarker whose characteristics are fully understood is being used to prove efficacy in a phase III trial, this is probably not the best way to evaluate the biomarker. Table 4 looks very similar to Table 1 but includes only the $q_{ij}$'s.

Table 5 is the proper setup for evaluating the information. This is a Bayesian-like framework because is requires the prior probability for each hypothesis. If it is assumed that $\pi_1 = \pi_2 = 0.5$, then Tables 4 and 5 are equivalent, a state of ignorance in many cases. However, most biologists are not ignorant about the underlying system; a bad guess for these probabilities would probably be better for evaluating biomarkers than assuming equality because it can be very misleading to assume that the maximum information is transmitted if, in fact, it is not (see Tables 2 and 3). Much money can be wasted when numbers are misused.

Diagnostic tests (biomarkers) are usually evaluated in terms of true positives (TPs), true negatives (TNs), false positives (FPs), and false negatives (FNs). Table 6 is a typical setup for this evaluation. Again these are just the $q_{ij}$'s obtained by counting in most cases. The concepts of biomarker sensitivity and specificity are as follows [5]:

**TABLE 6 Outcomes for a Binary-Decision Diagnostic Test**

| | Decision | | |
|---|---|---|---|
| Signal | $D_1(-)$ | $D_2(+)$ | Total |
| $S_1$ | TN | FP | TP + FN |
| $S_2$ | FN | TP | FP + TN |
| Total | TN + FN | FP + TP | $n$ |

$$\text{sensitivity} = \frac{\text{TP}}{\text{TP}+\text{FN}} = 1-\beta$$
$$\text{specificity} = \frac{\text{TN}}{\text{TN}+\text{FP}} = 1-\alpha$$

It is clear that these are just $q_{11}$.and $q_{22}$, respectively. Obviously, a biomarker such as aspartate aminotransferase (AST, SGOT) has information about many diagnoses; therefore, looking at each diagnosis one at a time seems less than optimal, could be grossly misleading, and probably should be avoided when possible. Mathematically, there is no reason to limit the diagnostic categories to two. However, with more than two outcomes, the terms sensitivity and specificity become somewhat meaningless. Perhaps a term such as *D-specificity* would be appropriate for the generalization, where *D* is replaced by the particular disease name.

The *D*-specificity is a measure of how well the biomarkers **Y** detect the specific signal under the set of decision rules *R*. This is an important metric for the biomarker developer. However, for biomarker application the biomarker's utility must be evaluated in light of the prior probabilities of the signals being sent. A common way, preferred by clinicians, to get at this issue is through the positive and negative predictive values, PPV and NPV, respectively [5]. For the normal–abnormal case, the PPV = $P[S_2|D_2]$. Bayes' theorem can be applied to get

$$\text{PPV} = \frac{\pi_2(1-\beta)}{\pi_1\alpha + \pi_2(1-\beta)}$$

Similarly, the NPV = $P[S_1|D_1]$, which is

$$\text{NPV} = \frac{\pi_1(1-\alpha)}{\pi_1(1-\alpha) + \pi_2\beta}$$

In comparing these formulas to Table 5, these are the *D*-diagonal terms divided by the sum of the corresponding column. Both will have the value 1

in a noise-free system. For more than two outcomes, these can be generalized to $D$-predictive values (DPV), where $D_i\text{PV} = \text{P}[S_i|D_i]$ and

$$D_j PV = \frac{\pi_j q_{jj}}{\sum_{i=1}^{k} \pi_i q_{ij}}$$

This is the probability that the clinician would use if considering only one diagnosis against all others. More commonly, a clinician is considering a number of diagnoses simultaneously; this is called the *differential diagnosis*. A list of probabilities would be appropriate for this situation. Since only one $D$ is chosen in the information framework, a probability for each potential signal given $D$ can be created to arrange the differential diagnosis in order of decreasing probability. The *D-differential value* can be defined as $D_j\text{DV} = \text{P}[S_i|D_j]$ and calculated as

$$D_j DV = \frac{\pi_i q_{ij}}{\sum_{i=1}^{k} \pi_i q_{ij}}$$

which is just proportion for each cell in the $D_j$ column for each $S$. These are the numbers that a clinician would use to order the differential diagnosis according to probability. Therefore, the context for the utility of a biomarker really depends on how it will be used.

The ROC (receiver operating characteristic) curve is a common way to evaluate biomarkers. It combines specificity and sensitivity. Figure 3 illustrates how the ROC curve is constructed. These represent the probability density $p(y)$ for a continuous Gaussian biomarker $y$. $S_1$ has mean 0 and standard deviation 1 in all cases. For cases A and D, $S_2$ has mean 4 and variance 1; for cases B and C, the means are 0.01 and 2, respectively. The vertical black line represents the partition determined by some optimality rule: The $y$ values to the left of the line represent $R_1$, and the values to the right, $R_2$. If the signal observed is on the left side ($D_1$), the signal is called $S_1$; if on the right side ($D_2$), it is called $S_2$. The total channel noise ($N$) is given by

$$N = \pi_1\alpha + \pi_2\beta$$

These are the off-diagonal terms in Table 5. One optimization rule is to choose the cut point $z$ so that $N$ is minimized. This leads to the relationship

$$\frac{\pi_1 p_1(z)}{\pi_2 p_2(z)} = 1$$

which can be rewritten in a simpler form

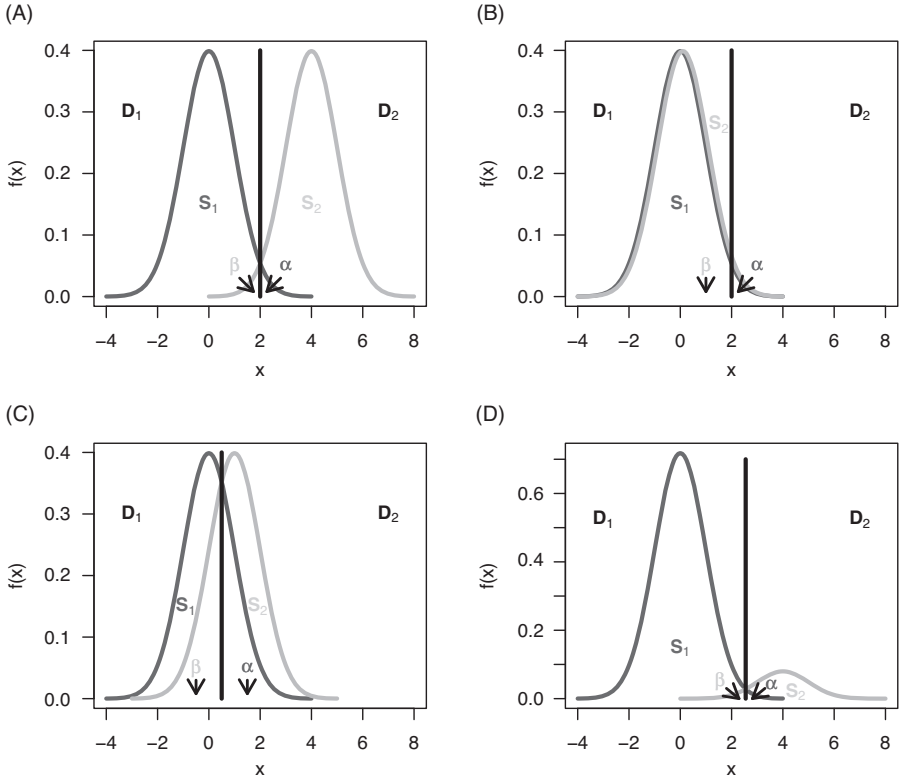


**Figure 3** Construction of diagnostic rules for various probability structures. (A) Signal $S_1$ has prior probability 0.5, mean 0, and standard deviation 1; $S_2$ has prior probability 0.5, mean 4, and standard deviation 1. (B) $S_1$ has prior probability 0.5, mean 0, and standard deviation 1; $S_2$ has prior probability 0.5, mean 0.1, and standard deviation 1. (C) $S_1$ has prior probability 0.5, mean 0, and standard deviation 1; $S_2$ has prior probability 0.5, mean 1, and standard deviation 1. (D) $S_1$ has prior probability 0.9, mean 0, and standard deviation 1; $S_2$ has prior probability 0.1, mean 4, and standard deviation 1. (*See insert for color reproduction of the figure.*)

$$\log\frac{\pi_1}{\pi_2} + \log p_1(z) - \log p_2(z) = 0$$

$$\log\frac{\pi_1}{\pi_2} - \frac{1}{2}z^2 + \frac{1}{2}(z-\mu)^2 = 0$$

The solution of this equation ($z$) is the point at which the density functions cross. In cases A, B, and D, the black line is placed at this point. For cases A, B, and C, $\pi_1 = \pi_2 = ½$; in case D, $\pi_1 = 9/10$ and $\pi_2 = 1/10$.

This approach is known in statistics as *discriminant analysis* and in computer science as *supervised learning*. Sometimes classification is done with logistic regression, which is equivalent when

$$\begin{aligned} P[S_2|z] &= \frac{\pi_2 p_2(z)}{\pi_1 p_1(z) + \pi_2 p_2(z)} \\ &= \frac{1}{1 + \pi_1 p_1(z)/\pi_2 p_2(z)} \\ &= \frac{1}{1 + \exp[\log(\pi_1/\pi_2) + \log p_1(z) - \log p_2(z)]} \end{aligned}$$

The cut point is the $z$ where $P[S_2|z] = ½$, which is the same as above as long as the logistic model has this form. Usually, the first log term is estimated from the data, which will give a result that is not optimal for the real prior probabilities.

The receiver operating characterization (ROC) curve is constructed by sweeping the black line from right to left and calculating α and β at each point. The results are shown in Figure 4 for all four cases. The positively sloped diagonal line represents the case when the distributions are exactly the same. The letters for the cases in Figure 3 are placed at the optimal points, respectively. For equally likely priors, they fall on the negatively sloped diagonal line. A statistical test can be performed to see if the area under the ROC curve

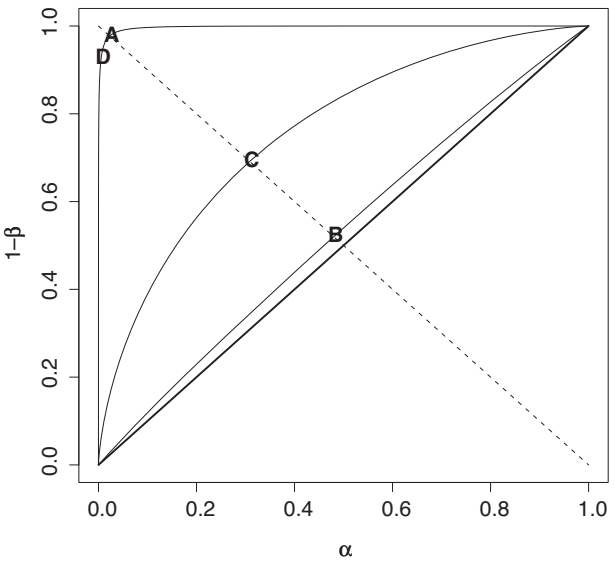


**Figure 4** ROC curves for cases A to D.

is significantly greater than ½. Unfortunately, this only indicates that the biomarker can detect one of two signals under ideal conditions. A better test might be one that can detect if the length of the vector $\left(\sqrt{P_1^2(z)+[1-P_2(z)]^2}\right)$ from the lower right corner to the optimal point extends significantly beyond the positive diagonal line. It should be noted that this length measure depends both on the data observed and on the priors. If one of the two signals is rare, it would be very hard to show significance. This is not true for the general test of ROC area.

With respect to evaluating a biomarker, Figures 3 and 4 demonstrate some interesting aspects. Case A represents a situation where a single test discriminates well between $S_1$ and $S_2$ but it is not noise-free. It would be difficult to do better than this in the real world. What is striking is that case B, which shows essentially no visible separation, actually has an ROC area greater than ½ that could be detected with sufficiently large $n$. Typical ROC curves look more like case C, perhaps slightly better. The densities in this case show only modest separation and would not make a very impressive biomarker. However, the ROC curve area test might tend to get people excited about its prospects. If one of the signals is rare, it becomes essentially undetectable, but the ROC area test gives no indication of this. Case C is a good candidate for adding another test to make it a multivariate biomarker. In higher dimensions, greater separation may occur (i.e., more information might be transmitted).

However, the ROC curve cannot handle more than two outcomes, at least not in a visually tractable way, although collapsing cells by summing in Table 1 to the $2 \times 2$ case might work in some cases. This takes us back to $I(S,D)$ as a measure of the biomarker's utility since it works for any number of signals and any number of tests. A generalization of the discriminant function would minimize the total noise, obtained by summing the off-diagonal elements of Table 1:

$$N = \sum_{i=1}^{k}\sum_{j=1}^{\ell} \pi_i q_{ij} - \sum_{i=1}^{k} \pi_i q_{ii}$$

This minimization would probably be difficult using calculus as before and would require a numerical analysis approach, too involved to describe here. This minimization would determine the $R$'s. The partitions, the edges of the $R$'s, are a point, a line, a plane, and flat surfaces of higher dimension as the number of tests increases from one, respectively, assuming that the parameters of the $P$'s differ only in the means and that the $P$'s are Gaussian. When the other parameters differ among the $P$'s, the surfaces are curved.

Without some involved mathematics, it is difficult to know if the minimal noise optimization is equivalent to the maximal information optimization. It seems entirely feasible that communication systems with the same total noise could have different diagonal elements, one of which might give the most information about all the signals in the system. The answer is unknown to the

author and is probably an open research question for biological systems. Another level of complication, but one that has more real-world relevance, is one that minimizes the cost. This is discussed at an elementary level elsewhere [24]. Multiplying each term in Table 1 by its cost, and minimizing the total expected cost, is a classical decision theory approach [25–27].

The previous discussion of diagnostic performance has said little about time effects. The time variable is contained in $P$ and will add another dimension for each time point measured, in general. This presents a severe dimensionality problem, both for the biologist and the analyst, since each measurement of the biomarker on the same person creates a new dimension. If the measurement times are not the same for all cases, the entire optimization process may depend on which times are chosen. The biggest problem with dimensionality is that it usually involves a growing number of parameters. To obtain precise (high-Fisher-information) estimates of all the parameters simultaneously, the sample size requirement grows much faster than the dimension. Here is a place where invariance plays a key role. If some parameters can be shown to be biological constants analogous to physical constants, through data pooling they can be estimated once and reused going forward. If the parameters vary with time, the time function for the parameter needs to be determined and reused similarly. Functions and parameters are very compact objects for storing such reusable knowledge.

## GENERAL STRATEGY FOR DEVELOPING AND USING DYNAMIC BIOMARKERS

To summarize, there are many types of information, only two of which were described here. In general, dynamic biomarkers will have more information than static biomarkers; multivariate biomarkers will have more information than univariate biomarkers. These ideas are foreign to most biologists and will take some time to spread among them. The standard entities used to evaluate biomarkers, such as sensitivity, specificity, and ROC curves, have questionable or limited utility but can easily be modified to fit the Shannon information framework. Some steps for biomarker development and implementation are given here as a guideline and probably contain significant gaps in knowledge, requiring further study:

1. Choose the signals ($S$) or the surrogates ($D$) that are relevant (biology).
2. Choose the best models ($P$) for the given $S$ (mathematics/statistics) and all available biomarkers ($\mathbf{Y}$) (biology).
3. Choose the best decision rules ($R$) given $P$ (mathematics/statistics).
4. Choose the best subset, or subspace, of $\mathbf{Y}$ (statistics).

The mathematical entities $P$, $R$, and $\mathbf{Y}$ determine the diagnosis of $S$. Currently, most of the effort is focused on $\mathbf{Y}$.

## MODELING APPROACHES

### Signal Types

This section deals with some very specific aspects of the models for $P(\mathbf{Y}_t)$. Most people think of signals as being electrical. This is probably because most of the terminology and use comes from electrical engineering. However, the mathematics is completely general. Signals can be static or dynamic, meaning something that is measurable with a constant value or a time-varying value, respectively. Biomarkers are just biological signals. Signals are classified as analog (continuous) or digital (discrete). Birth and death are discrete signals; blood pressure and serum glucose levels are continuous signals. Dynamic signals vary over time. Time can also be classified as continuous or discrete, as described above.

Anything that changes in time and space is a dynamic system. This generally implies that space has more than one dimension, but it does not have to be physical space. The space of biomarkers would be an example where each univariate biomarker (signal) defines a spatial dimension. A continuous system is modeled with a set of differential equations where the variables defining the space usually appear in more than one equation. A discrete system is modeled with a set of difference equations.

### Mathematical Models of Dynamic Signals

Mathematical models are usually hypothesized prior to the experiment and then verified by the experimental data. Historically, these models have been deterministic differential equations. Pharmacokinetics is an example of this. Following are some typical mathematical models (ordinary differential equations) and their solutions:

| | | |
|---|---|---|
| Linear model: | $\frac{dx}{dt} = \beta_1$ | $\Rightarrow x = \beta_0 + \beta_1 t$ |
| Quadratic model: | $\frac{dx}{dt} = \beta_1 + 2\beta_2 t$ | $\Rightarrow x = \beta_0 + \beta_1 t + \beta_2 t^2$ |
| Exponential model: | $\frac{dx}{dt} = \beta_1 x$ | $\Rightarrow x = e^{\beta_0 + \beta_1 t}$ |
| Sigmoidal model: | $\frac{dx}{dt} = \frac{\beta_1}{\beta_2} x(\beta_2 - x)$ | $\Rightarrow x = \frac{\beta_0 \beta_2}{\beta_0 + (\beta_2 - \beta_0) e^{\beta_1 t}}$ |
| Sine model: | $\frac{dx}{dt} = \beta_1 x(\beta_2^2 - x^2)$ | $\Rightarrow x = \beta_0 + \beta_2 \sin \beta_1 t$ |

The middle column is the hypothesized model of the velocity (first time derivative) of $x$, a dynamic quantity. Solving such models involves finding a method that ends up with $x$ by itself on the left of the equal sign and a

function of time on the right. More complicated types of models are used by mathematical biologists to describe and predict the behavior of microscopic biological processes such as intracellular metabolism or nerve conduction [3].

## Statistical Models

Statistical models are models that have at least one random (stochastic) variable. Going forward, mechanistic biological models will probably be a mixture of deterministic and stochastic variables. Most statistical models describe the behavior of the mean of $P$ as a function of other variables, including time. Unless there is a specific experimental design model such as analysis of variance (ANOVA), statistical models tend to be constructed after the behavior of the data is known. A serious loss of information can occur if the wrong model is chosen. Here are some typical statistical models of data varying in time, where $T$ is the total time of observation:

Mean model: $y_t = \mu_t + \varepsilon_t$

ANOVA model: $y_t = \mu_t + \varepsilon$

Straight-line model: $y_t = \beta_0 + \beta_1 t + \varepsilon$

Polynomial model: $y_t = \beta_0 + \beta_1 t + \beta_2 t^2 + \beta_3 t^3 + \beta_4 t^4 + \cdots + \varepsilon$

Trigonometric model:
$$y_t = \beta_0 + \beta_1 \cos\frac{2\pi t}{T} + \beta_2 \sin\frac{2\pi t}{T} + \beta_3 \cos\frac{4\pi t}{T} + \beta_4 \sin\frac{4\pi t}{T} + \cdots + \varepsilon$$

Log-linear model: $y_t = e^{\beta_0 + \beta_1 t + \beta_2 t^2 + \beta_3 t^3 + \beta_4 t^4 + \cdots + \varepsilon}$

Sigmoidal model: $y_t = \dfrac{\beta_0 \beta_2}{\beta_0 + (\beta_2 - \beta_0) e^{\beta_1 t}} + \varepsilon$

Sine model: $y_t = \beta_0 + \beta_2 \sin \beta_1 t + \varepsilon$

OU model:
$$y_t = e^{-\alpha(t-s)} y_s + \mu\left(1 - e^{-\alpha(t-s)}\right) + \sqrt{2\alpha}\sigma e^{-\alpha(t-s)} \int_s^t e^{\alpha u} dB_u + \varepsilon$$

In all of these models, it is generally assumed that $\varepsilon$ is the random measurement error, or residual difference between the mean of the model and the observed $y$. This error is assumed to be Gaussian with mean zero and variance $\sigma_\varepsilon^2$, and each measurement error is statistically independent of all the others. Since the rest of the model is usually deterministic, in most fields of measurement it is the error term that induces $y$ to be a random variable. However, in biology there are many sources of biological variation as well. In the OU model when the biological Brownian motion $B$ is also contributing variation in a complicated way, mechanistic interpretation is much easier through the SDE. It should be noted that all the parameters in these models are also independent of time, except the first two.

The "mean model" is not really a model; it represents simply the calculation of the mean and standard deviation at each time point, a typical biologist approach. These are descriptive statistics and do not lend themselves to efficient statistical analyses. The next five models are all special cases of the linear model; that is, they are linear in the unknown parameters, not in the variables. ANOVA is a model that parameterizes the mean for each combination of the experimental factors; there are many equivalent ways to parameterize these means. The fundamental difference between the mean model and ANOVA is that for the latter, the standard deviations are assumed to be equal, leading to a more precise estimate of the means, if the assumption is true. In the representation of ANOVA in the example below, the cell-mean model is used, meaning that a mean is estimated for each combination and any modeling of those means is done using contrasts of those means (shown above). Given the data and a linear model, a system of equations, one for each measurement, can be solved using linear algebra to get OLS estimates of the parameters.

The last three models are nonlinear and require iterative methods to solve for the parameters, which means repeated guessing using mathematical techniques such as calculus. This requires a computer and a stopping rule to determine when the guess is close enough. Sometimes it is not possible or is very difficult to get the algorithm to find the solution. This is related to the nature of the model and the nature of the data. Statisticians prefer to avoid nonlinear models for this reason. Note that the OU model is nonlinear but can be "linearized" in special cases.

## Monte Carlo Data Generation

Most software packages have random number generators. These are actually pseudorandom numbers because all of the algorithms are deterministic [28]. Monte Carlo simulations are just the generation of these numbers to suit a particular purpose. The two most common uses of Monte Carlo methods are the generation of random data to evaluate or display statistical results and the solution of deterministic equations that are too difficult to solve using mathematical methods. Commercial simulators use Monte Carlo methods as well.

All of the data used in the examples below were generated by this method. Each of the example experiments used the same data sets for a given autocorrelation and response percentage, starting with the first observation in each group. The underlying probability distribution was the OU model using different means where the control (no-effect) group ($n$ = 500) had all constant parameters and the experimental group ($n$ = 500) had the same parameters except that the mean was a quadratic model in time deviating from the controls after $t = 0$, with $\delta$ being assigned randomly as a 0 or 1, with probability $p$ for responding to the intervention:

| Parameter | Control Group | Experiment Group |
|---|---|---|
| $\mu$ | $\beta_0$ | $\beta_0 + \delta(\beta_1 t + \beta_2 t^2)$ |

The values for $\beta_0$, $\beta_1$, and $\beta_2$ were 3, 1/5, and –2/300, respectively; $\sigma$ was ½; and $\alpha$ was either 3 to simulate independence or 0.03 to simulate a value similar to those observed previously in liver tests [29,30]. The response proportion was varied among 1.0, 0.5, and 0.1 to show the effect of partial responses, which is common in biology. A measurement error was added with a CV of 10% (e.g., $\sigma_\varepsilon = 0.1\mu$), except that the added noise was based on the actual measured value, not the average measured value, where it was assumed that the raw data were log-Gaussian, like an enzyme [18,31]. The parameters of the quadratic response model were chosen so that the maximum is $3\sigma$ from the control group mean halfway between the ends of the time interval.

The first 50 samples in each group of the simulated data are shown in Figures 5 and 6 for $\alpha = 3$ and $\alpha = 0.03$, respectively. In Figure 5 at 100% response, the quadratic response is very clear and diminishes with decreasing

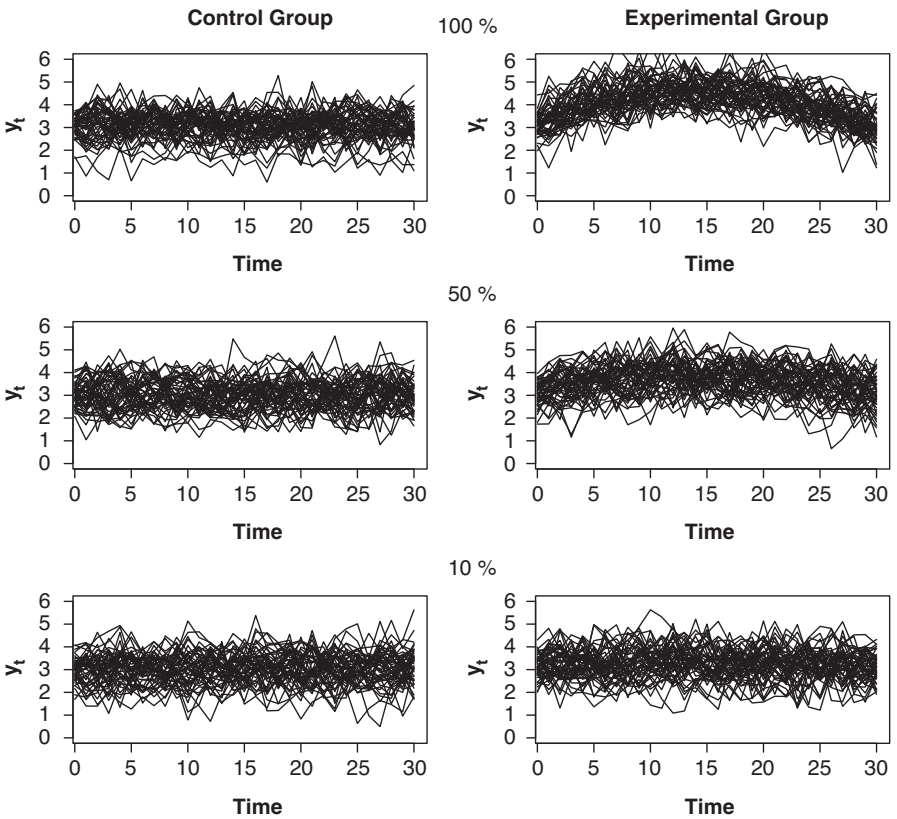


**Figure 5** Examples of the simulated data for $\alpha = 3$ and responses 100%, 50%, and 10%.

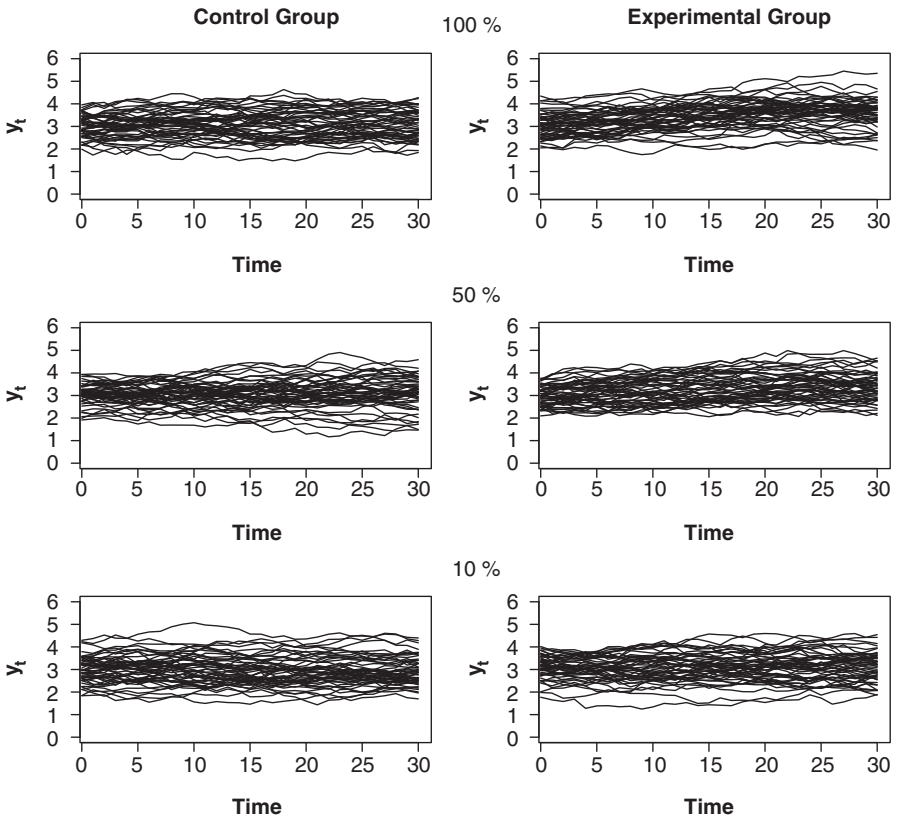


**Figure 6** Examples of the simulated data for $\alpha = 0.03$ and responses 100%, 50%, and 10%.

response. No statistician is needed here. In Figure 6 the signals have a very different appearance, even though it is only $\alpha$ that differs. The strong autocorrelation mutes the variation in the signal and causes it to lag in time. There is little visual difference among the difference responses. A look at the plots might suggest that there is no signal at all, but the mean signal is exactly the same as in Figure 5. Standard statistical methods currently used in drug development may not find this signal. The examples below are intended to illustrate some of the issues. It will obviously take more sophisticated methods, not used in this chapter, to detect this signal properly under autocorrelation.

The following statistical models were used for the experimental effect while $t$ was set to zero at all time points on the right-hand side for the control group:

Model 1: $y_t = \mu_t + \varepsilon$

Model 2: $y_t = \gamma y_s + \mu_t + \varepsilon$

Model 3: $y_t = \gamma y_0 + \mu_t + \varepsilon$

Model 4: $y_t - y_0 = \mu_t + \varepsilon$

Model 5: $y_t - y_0 = \gamma(y_s - y_0) + \mu_t + \varepsilon$

Model 6: $y_t = \beta_0 + \beta_1 t + \beta_2 t^2 + \beta_3 t^3 + \beta_4 t^4 + \varepsilon$

Model 7: $y_t = \gamma y_s + \beta_0 + \beta_1 t + \beta_2 t^2 + \beta_3 t^3 + \beta_4 t^4 + \varepsilon$

Model 8: $y_t = \gamma y_0 + \beta_0 + \beta_1 t + \beta_2 t^2 + \beta_3 t^3 + \beta_4 t^4 + \varepsilon$

Model 9: $y_t - y_0 = \beta_0 + \beta_1 t + \beta_2 t^2 + \beta_3 t^3 + \beta_4 t^4 + \varepsilon$

Model 10: $y_t - y_0 = \lambda(y_s - y_0) + \beta_0 + \beta_1 t + \beta_2 t^2 + \beta_3 t^3 + \beta_4 t^4 + \varepsilon$

These models are compared in the simulated experiments below. Models 1 to 10 used the lm(·) function in *R* [32], which is OLS. Models 11 to 20 are the same models, respectively, except that the linear mixed-effect model lme(·) in *R* with restricted maximum likelihood estimation was used, allowing random baselines (intercepts). Model 1 is ANOVA, model 2 is ANOVA covariate-adjusted for the previous observation, model 3 is ANOVA covariate-adjusted for the baseline, model 4 is ANOVA for the change from baseline, and model 5 is ANOVA for the change from baseline that is covariate-adjusted for the previous change from baseline. Models 6 to 10 replaced the means with fourth-degree polynomials in time.

These models were chosen because they are typical of what a statistician or a biologist might apply to similar data sets to study biomarkers when the underlying time-dependence structure is not known. The focus is on the mean response, which is typical; no attempt was made to extract the biological variation component $\sigma$ or the analytical variation component $\sigma_\varepsilon$. The nonconstant variance of the residuals was also ignored. If there is no measurement error and the biology is known to follow an OU model, model 2 or 7 would be the "best" to estimate $\alpha$, $\mu$, and $\sigma$, although transformations would be needed to get the correct estimates. For example, $\alpha = -\log\gamma/(t - s)$ only if $t - s$ is a constant, which was used below. When $\gamma$ is not in the model, a value of "na" is shown in the results; when it is negative, "*" is shown. The mean would have to be transformed similarly, but this was not done below in the calculation of the bias, since it is not likely that the analyst would know the proper transformation. If the underlying process is known to be OU and the baseline of both groups has the same distribution as the control group over time, covariance adjustment for the baseline loses information about $\alpha$ and increases the model bias, possibly masking the signal. Modeling the change from baseline underestimates the variance because it assumes that the variance of the baseline distribution is zero. This approach might give less biased results but is likely to have *p*-values that are too small, meaning that false signals may be detected.

Statistical tests for signal detection in the experiments below are based on the differences in parameters between the two groups: linear effect, quadratic effect (quad), cubic effect, and quartic effect. Hierarchial statistical testing was performed to learn about the mathematical form of the biological response. An overall test for the deviation from baseline was done to see if any signal was present. This is the test that counts since it controls the type I error. The subsequent tests were secondary and were not corrected for multiple testing. Second, a test for a fourth-degree polynomial (poly) was done; if it is significant, it means that at least one coefficient in the polynomial is not zero. Then, each degree of the polynomial is tested to determine the functional shape of the curve. In the simulated data, only the true quadratic term is nonzero; any other significant findings are spurious. In all cases presented below, the estimates of $\sigma$ are too low, resulting in $p$-values that are too small.

To evaluate the Fisher information for each model, the MSE was calculated using the estimated variance (Var) of the time function integrated over time from 0 to 30 and the integrated bias using the true polynomial. This represents the mean squared area between the true time function and the estimated time function. In most experimental cases, the bias can never be known, but it can be reduced using techniques such as bootstrapping [21]. When using OLS, it is assumed that the estimated parameters are unbiased. In a real statistical experiment, the behavior of the particular estimation procedure would be simulated hundreds or thousands of times and the distributions of the results would be studied. Here only a single simulation was done for illustration, and no scientific conclusion should be inferred from this presentation.

## EXAMPLE EXPERIMENTS

### In Vitro Experiments

This section is intended to explore the behavior of an OU process when only a few experiments are run but the response can be measured at many equally spaced time points. Assay development is not discussed here but can be found in many papers and textbooks, such as that of Burtis and Ashwood [33]. All the experiments that follow assume that the assay has been "optimized" according to the existing standards. One point that needs to be stressed here is the relationship between the true value of the biomarker $\mu$ and the variation of the measurement error $\sigma_\varepsilon$, assuming that no assay bias is present. Many use the concept of the CV = $\sigma_\varepsilon/\mu$ being constant. If this relationship really holds, the statistical analyst needs to know that and approximately what the constant is. Prior to using a biomarker, it is important to define the mathematical relationship between $\sigma_\varepsilon$ and $\mu$ so that the most information can be gained by the modeling. It should also be noted the since $\mu$ is a function of time, the measurement variance will also be a function of time. Most statistical procedures assume that the measurement variance is constant, including those in these examples. This assumption will tend to hide signals.

In the laboratory, it is relatively easy to obtain measurements at frequent, equally spaced time intervals. If a good biological model is available, this is the place to define the best (maximum Fisher information) mathematical model of the response. If such a model is not known, it should be explored first. This section looks at two simulation experiments comparing the size of $\alpha$ (3 vs. 0.03) with 100% response. Each has $n = 5$ experimental units per group, and each unit has $m = 30$ follow-up measurements. These could represent chemical reaction experiments, cell measurements, well measurements, or any similar in vitro biological model for a biomarker. The raw data are shown in Figures 5 and 6, and the statistics are shown in Tables 7a and 7b.

When $\alpha$ is large, the measurements are effectively independent and there should be no relationship to the previous measurements, including the baseline. However, if the baselines do not come from the same distribution, the comparability of the results is called into question. Here in Table 7a, only models 1, 5, and 7 gave reasonable estimates of $\alpha$. In the strong autocorrelation case, Table 7b shows that models 2, 5, 7, 10, 12, 15, 17, and 20 gave estimates of $\alpha$ that, although not very precise, were the correct order of magnitude.

In Table 7a, every model detected the signal and the polynomial and the quadratic effect intended. However, there are many $p$-values showing significance where there should not be any. This can lead to spurious conclusions about the nature of the biomarker. In the experiment shown in Table 7b, several models gave reasonable indications for the magnitude of the autocorrelation but did not always find the underlying signal. In the first case, all the biases show overestimates of the time effect, while in the second case, the biases are in the opposite direction and have a larger magnitude, making the MSE larger—less information. The larger and negative bias is exacerbated by the fact that the time model parameter estimates need to be divided by $1 - e^{-\alpha(t-s)}$, which is always less than 1. This correction was not done because the estimates of $\alpha$ are not good enough to make the correction reasonable to use. Additionally, the analyst would not be estimating $\alpha$, making these results more representative of current practice. When $t - s$ is not constant, software for the estimation procedure is not readily available.

## In Vivo Experiments

This section is meant to illustrate experiments of the size used in animal studies but may also apply to early clinical development. Here there are only $m = 4$ follow-up measurement, but the number ($n = 36$) per group was scaled up so that approximately the same number of observations are available. This results in comparable degrees of freedom for the model errors.

In Tables 8a and 8b, the results are analogous to those in Tables 7a and 7b. However, in Table 8a the estimates of $\alpha$ are generally all bad, leading to the conclusion that autocorrelation is present, when it really is not. The quadratic

**TABLE 7a Statistical Results for $\alpha = 3$, $p = 1$, $m = 30$, and $n = 5$**

| | | | | *p*-Value | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Model | $\alpha^a$ | Error df | $\sigma$ | All | Poly | Linear | Quad | Cubic | Quartic | Var | Bias | MSE |
| 1 | na | 240 | 0.59 | 0.00 | 0.00 | 0.08 | 0.00 | 0.54 | 0.29 | 0.0110 | 0.2746 | 0.0864 |
| 2 | 1.29 | 239 | 0.57 | 0.00 | 0.00 | 0.06 | 0.00 | 0.52 | 0.44 | 0.0111 | 0.1452 | 0.0322 |
| 3 | * | 239 | 0.50 | 0.00 | 0.00 | 0.02 | 0.00 | 0.67 | 0.34 | 0.0080 | 0.2802 | 0.0865 |
| 4 | na | 240 | 0.50 | 0.00 | 0.00 | 0.02 | 0.00 | 0.64 | 0.32 | 0.0080 | 0.2795 | 0.0861 |
| 5 | 6.39 | 239 | 0.50 | 0.00 | 0.00 | 0.02 | 0.00 | 0.64 | 0.32 | 0.0090 | 0.2787 | 0.0867 |
| 6 | na | 295 | 0.57 | 0.00 | 0.00 | 0.00 | 0.00 | 0.01 | 0.02 | 0.0040 | 0.2430 | 0.0630 |
| 7 | 1.58 | 294 | 0.56 | 0.00 | 0.00 | 0.00 | 0.00 | 0.03 | 0.04 | 0.0091 | –0.0166 | 0.0093 |
| 8 | * | 294 | 0.50 | 0.00 | 0.00 | 0.00 | 0.01 | 0.06 | 0.06 | 0.0035 | 0.0381 | 0.0050 |
| 9 | na | 295 | 0.50 | 0.00 | 0.00 | 0.00 | 0.00 | 0.04 | 0.05 | 0.0031 | 0.0626 | 0.0070 |
| 10 | * | 294 | 0.50 | 0.00 | 0.00 | 0.00 | 0.00 | 0.04 | 0.05 | 0.0070 | 0.0993 | 0.0169 |
| 11 | na | 231 | 0.50 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.0080 | 0.2746 | 0.0834 |
| 12 | * | 230 | 0.50 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.0090 | 0.2767 | 0.0856 |
| 13 | * | 9 | 0.50 | 0.00 | 0.00 | 0.00 | 0.00 | 0.23 | 0.02 | 0.0080 | 0.2802 | 0.0865 |
| 14 | na | 231 | 0.50 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.0080 | 0.2795 | 0.0861 |
| 15 | * | 230 | 0.50 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.0090 | 0.2819 | 0.0884 |
| 16 | na | 286 | 0.50 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.0233 | 0.2307 | 0.0766 |
| 17 | * | 285 | 0.50 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.0308 | 0.2905 | 0.1152 |
| 18 | * | 286 | 0.50 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.0040 | 0.0397 | 0.0056 |
| 19 | na | 286 | 0.50 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.0036 | 0.0640 | 0.0077 |
| 20 | * | 286 | 0.50 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.0079 | 0.1103 | 0.0201 |

[a]na, not available.
*, negative estimate.

**TABLE 7b Statistical Results for $\alpha = 0.03$, $p = 1$, $m = 30$, and $n = 5$**

| | | | | *p*-Value | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Model | $\alpha^a$ | Error df | $\sigma$ | All | Poly | Linear | Quad | Cubic | Quartic | Var | Bias | MSE |
| 1 | na | 240 | 0.51 | 0.00 | 0.00 | 0.00 | 0.24 | 0.77 | 0.45 | 0.0085 | –0.5978 | 0.3658 |
| 2 | 0.02 | 239 | 0.12 | 0.23 | 0.20 | 0.13 | 0.27 | 0.13 | 0.32 | 0.0005 | –0.2610 | 0.0686 |
| 3 | 0.61 | 239 | 0.44 | 0.00 | 0.00 | 0.00 | 0.19 | 0.69 | 0.41 | 0.0062 | –0.5967 | 0.3623 |
| 4 | na | 240 | 0.49 | 0.00 | 0.00 | 0.00 | 0.25 | 0.69 | 0.49 | 0.0078 | –0.5959 | 0.3629 |
| 5 | 0.01 | 239 | 0.12 | 0.32 | 0.15 | 0.08 | 0.29 | 0.13 | 0.34 | 0.0005 | –0.2578 | 0.0669 |
| 6 | na | 295 | 0.47 | 0.00 | 0.00 | 0.00 | 0.03 | 0.08 | 0.11 | 0.0028 | –0.2534 | 0.0670 |
| 7 | 0.02 | 294 | 0.12 | 0.02 | 0.02 | 0.04 | 0.10 | 0.14 | 0.15 | 0.0003 | –0.9592 | 0.9204 |
| 8 | 0.61 | 294 | 0.41 | 0.00 | 0.00 | 0.00 | 0.03 | 0.07 | 0.09 | 0.0021 | –0.2905 | 0.0865 |
| 9 | na | 295 | 0.45 | 0.00 | 0.00 | 0.00 | 0.07 | 0.13 | 0.16 | 0.0025 | –0.3215 | 0.1059 |
| 10 | 0.01 | 294 | 0.12 | 0.03 | 0.03 | 0.05 | 0.12 | 0.16 | 0.16 | 0.0003 | –0.9659 | 0.9333 |
| 11 | na | 231 | 0.26 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.0023 | –0.5978 | 0.3596 |
| 12 | 0.07 | 230 | 0.12 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.0005 | –0.2787 | 0.0781 |
| 13 | 0.61 | 9 | 0.26 | 0.00 | 0.00 | 0.00 | 0.00 | 0.12 | 0.01 | 0.0023 | –0.5967 | 0.3584 |
| 14 | na | 231 | 0.26 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.0023 | –0.5959 | 0.3573 |
| 15 | 0.08 | 230 | 0.12 | 0.00 | 0.00 | 0.22 | 0.00 | 0.00 | 0.00 | 0.0005 | –0.2809 | 0.0794 |
| 16 | na | 286 | 0.25 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.0129 | –0.5409 | 0.3055 |
| 17 | 0.10 | 285 | 0.12 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.0011 | –0.9411 | 0.8869 |
| 18 | 0.56 | 286 | 0.25 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.0122 | –0.5327 | 0.2960 |
| 19 | na | 286 | 0.25 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.0126 | –0.5461 | 0.3108 |
| 20 | 0.10 | 286 | 0.25 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.0012 | –0.9470 | 0.8980 |

[a]na, not available.

**TABLE 8a Statistical Results for $\alpha = 3$, $p = 1$, $m = 4$, and $n = 36$**

| | | | | *p*-Value | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Model | $\alpha^a$ | Error df | $\sigma$ | All | Poly | Linear | Quad | Cubic | Quartic | Var | Bias | MSE |
| 1 | na | 280 | 0.66 | 0.00 | 0.00 | 0.05 | 0.00 | 0.43 | 0.34 | 0.0024 | –0.2068 | 0.0451 |
| 2 | 0.09 | 279 | 0.58 | 0.00 | 0.00 | 0.00 | 0.00 | 0.47 | 0.05 | 0.0021 | –0.0452 | 0.0041 |
| 3 | 0.00 | 279 | 0.51 | 0.00 | 0.00 | 0.05 | 0.00 | 0.21 | 0.26 | 0.0014 | –0.2013 | 0.0420 |
| 4 | na | 280 | 0.51 | 0.00 | 0.00 | 0.05 | 0.00 | 0.21 | 0.25 | 0.0014 | –0.2013 | 0.0420 |
| 5 | 0.27 | 279 | 0.51 | 0.00 | 0.00 | 1.00 | 0.00 | 0.47 | 0.15 | 0.0018 | –0.1574 | 0.0266 |
| 6 | na | 283 | 0.66 | 0.00 | 0.00 | 0.00 | 0.11 | 0.23 | 0.22 | 0.0054 | 0.0124 | 0.0056 |
| 7 | 0.64 | 282 | 0.57 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.01 | 0.0062 | –0.4316 | 0.1925 |
| 8 | 0.00 | 282 | 0.51 | 0.00 | 0.00 | 0.00 | 0.06 | 0.16 | 0.15 | 0.0033 | –0.0717 | 0.0085 |
| 9 | na | 283 | 0.51 | 0.00 | 0.00 | 0.00 | 0.06 | 0.16 | 0.14 | 0.0033 | –0.0718 | 0.0084 |
| 10 | 0.27 | 282 | 0.51 | 0.00 | 0.00 | 0.00 | 0.02 | 0.06 | 0.06 | 0.0056 | –0.1825 | 0.0389 |
| 11 | na | 209 | 0.49 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.0013 | –0.2068 | 0.0441 |
| 12 | 0.09 | 208 | 0.58 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.0021 | –0.0452 | 0.0041 |
| 13 | 0.00 | 71 | 0.49 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.0013 | –0.2013 | 0.0418 |
| 14 | na | 209 | 0.49 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.0013 | –0.2013 | 0.0418 |
| 15 | 1.92 | 208 | 0.51 | 0.00 | 0.00 | 0.97 | 0.00 | 0.00 | 0.00 | 0.0018 | –0.1574 | 0.0266 |
| 16 | na | 212 | 0.49 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.0124 | 0.0124 | 0.0126 |
| 17 | 0.09 | 211 | 0.57 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.0062 | –0.4316 | 0.1925 |
| 18 | 0.00 | 212 | 0.49 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.0042 | –0.0717 | 0.0093 |
| 19 | na | 212 | 0.49 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.0040 | –0.0718 | 0.0092 |
| 20 | 0.27 | 212 | 0.49 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.0056 | –0.1825 | 0.0389 |

[a]na, not available.

**TABLE 8b Statistical Results for $\alpha = 0.03$, $p = 1$, $m = 4$, and $n = 36$**

| | | | | *p*-Value | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Model | $\alpha^a$ | Error df | $\sigma$ | All | Poly | Linear | Quad | Cubic | Quartic | Var | Bias | MSE |
| 1 | na | 280 | 0.62 | 0.00 | 0.00 | 0.00 | 0.26 | 0.49 | 0.71 | 0.0021 | –0.4556 | 0.2096 |
| 2 | 0.01 | 279 | 0.31 | 0.00 | 0.00 | 0.92 | 0.00 | 0.95 | 0.15 | 0.0005 | –0.3520 | 0.1244 |
| 3 | 0.02 | 279 | 0.43 | 0.00 | 0.00 | 0.00 | 0.11 | 0.33 | 0.59 | 0.0010 | –0.4556 | 0.2086 |
| 4 | na | 280 | 0.44 | 0.00 | 0.00 | 0.00 | 0.11 | 0.33 | 0.60 | 0.0010 | –0.4556 | 0.2086 |
| 5 | 0.02 | 279 | 0.31 | 0.00 | 0.00 | 1.00 | 0.00 | 0.94 | 0.16 | 0.0006 | –0.3535 | 0.1255 |
| 6 | na | 283 | 0.61 | 0.00 | 0.00 | 0.86 | 0.70 | 0.78 | 0.86 | 0.0047 | –0.6992 | 0.4935 |
| 7 | 0.01 | 282 | 0.31 | 0.00 | 0.00 | 0.93 | 0.59 | 0.52 | 0.55 | 0.0012 | –0.8789 | 0.7737 |
| 8 | 0.02 | 282 | 0.43 | 0.00 | 0.00 | 0.80 | 0.59 | 0.69 | 0.80 | 0.0023 | –0.6991 | 0.4910 |
| 9 | na | 283 | 0.44 | 0.00 | 0.00 | 0.80 | 0.59 | 0.69 | 0.80 | 0.0024 | –0.6991 | 0.4910 |
| 10 | 0.02 | 282 | 0.31 | 0.00 | 0.00 | 0.94 | 0.59 | 0.53 | 0.56 | 0.0013 | –0.8757 | 0.7682 |
| 11 | na | 209 | 0.27 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.0005 | –0.4556 | 0.2080 |
| 12 | 0.01 | 208 | 0.31 | 0.00 | 0.00 | 0.15 | 0.00 | 0.37 | 0.00 | 0.0005 | –0.3520 | 0.1244 |
| 13 | 0.02 | 71 | 0.27 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.0004 | –0.4556 | 0.2080 |
| 14 | na | 209 | 0.27 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.0004 | –0.4556 | 0.2080 |
| 15 | 0.04 | 208 | 0.29 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.0005 | –0.3704 | 0.1377 |
| 16 | na | 212 | 0.27 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.0155 | –0.6992 | 0.5043 |
| 17 | 0.01 | 211 | 0.31 | 0.00 | 0.00 | 0.21 | 0.00 | 0.00 | 0.00 | 0.0012 | –0.8789 | 0.7737 |
| 18 | 0.02 | 212 | 0.27 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.0064 | –0.6991 | 0.4951 |
| 19 | na | 212 | 0.27 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.0065 | –0.6991 | 0.4952 |
| 20 | 0.05 | 212 | 0.27 | 0.00 | 0.00 | 0.91 | 0.00 | 0.00 | 0.00 | 0.0020 | –0.8409 | 0.7092 |

[a]na, not available.

signal is missed in many models in Table 8b. A second design was analyzed where there was only one follow-up time ($m = 1$). This is typical of designs that incorporate time but want independent measurements. In comparing Tables 8e to 8h with their counterparts Tables 8a to 8d, respectively, the signals detected are much weaker and, of course, there is no information about $\alpha$ even though estimates were calculated from the regression coefficients for the baseline term.

### Clinical Trials

Clinical trials generally have much larger sample sizes, especially phase III trials. In this section the same designs and models are used, but the sample size is increased. In Tables 9a and 9b the response is 50% in the experimental treatment group and $n = 200$. In Tables 9c and 9d, the response is only 10% in the experimental group and $n = 500$. The latter case is more typical of clinical safety data. With larger sample sizes, the properties of the models would be expected to improve. The $p$-values may get smaller and the MSEs, at least the variance component, should be reduced because more samples should produce more information. Here the bias seems to be unaffected for both response categories. This generally means that if the wrong model is chosen, more measurements will not make it better.

Tables 9a and 9b look very similar to those above, but Tables 9c and 9d show some notable features. First because the models do not estimate the proportion $p$ of responders, the models for the experimental treatment group are a weighted average of 10% quadratic response and 90% no response. This should cause the bias to increase, which it generally does. When autocorrelation is present and the response rate is low (Table 9d), the signal is lost completely (i.e., no information is available). It remains to be seen if better statistical procedures can find this signal.

## DISCUSSION

### Overview of Results

Biomarker experiments with repeated measures over time do not ensure that additional information will be obtained, even though, theoretically, it is guaranteed. The experimental design has to be correct, the biomathematical model of time response has to be correct, and the statistical modeling procedure must be an efficient estimator of that model. If any one of these parts is broken, information can be lost or destroyed completely. For efficacy biomarkers, this means wasted money or missed opportunities. For safety biomarkers, this leads to late attrition or a market recall.

**TABLE 8c Statistical Results for $\alpha = 3$, $p = 0.5$, $m = 4$, and $n = 36$**

| | | | | *p*-Value | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Model | $\alpha^a$ | Error df | $\sigma$ | All | Poly | Linear | Quad | Cubic | Quartic | Var | Bias | MSE |
| 1 | na | 280 | 0.74 | 0.00 | 0.00 | 0.34 | 0.00 | 0.40 | 0.30 | 0.0030 | –0.2865 | 0.0851 |
| 2 | 0.10 | 279 | 0.66 | 0.00 | 0.00 | 0.08 | 0.00 | 0.58 | 0.09 | 0.0025 | –0.1953 | 0.0406 |
| 3 | 0.00 | 279 | 0.54 | 0.00 | 0.00 | 0.27 | 0.00 | 0.22 | 0.16 | 0.0016 | –0.2847 | 0.0826 |
| 4 | na | 280 | 0.54 | 0.00 | 0.00 | 0.27 | 0.00 | 0.22 | 0.16 | 0.0016 | –0.2847 | 0.0826 |
| 5 | * | 279 | 0.54 | 0.00 | 0.00 | 0.06 | 0.00 | 0.17 | 0.21 | 0.0017 | –0.3065 | 0.0957 |
| 6 | na | 283 | 0.74 | 0.00 | 0.00 | 0.04 | 0.17 | 0.21 | 0.19 | 0.0068 | –0.5039 | 0.2607 |
| 7 | 0.10 | 282 | 0.66 | 0.00 | 0.00 | 0.00 | 0.01 | 0.02 | 0.02 | 0.0060 | –0.7036 | 0.5010 |
| 8 | 0.00 | 282 | 0.54 | 0.00 | 0.00 | 0.01 | 0.07 | 0.10 | 0.08 | 0.0036 | –0.5322 | 0.2869 |
| 9 | na | 283 | 0.54 | 0.00 | 0.00 | 0.01 | 0.07 | 0.09 | 0.08 | 0.0036 | –0.5331 | 0.2878 |
| 10 | * | 282 | 0.54 | 0.00 | 0.00 | 0.02 | 0.15 | 0.17 | 0.15 | 0.0043 | –0.4872 | 0.2417 |
| 11 | na | 209 | 0.56 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.0017 | –0.2865 | 0.0838 |
| 12 | * | 208 | 0.55 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.0019 | –0.2904 | 0.0862 |
| 13 | 0.00 | 71 | 0.54 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.0016 | –0.2847 | 0.0826 |
| 14 | na | 209 | 0.54 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.0016 | –0.2847 | 0.0826 |
| 15 | * | 208 | 0.54 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.0017 | –0.3101 | 0.0979 |
| 16 | na | 212 | 0.55 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.0155 | –0.5039 | 0.2694 |
| 17 | * | 211 | 0.55 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.0175 | –0.4869 | 0.2546 |
| 18 | 0.00 | 212 | 0.54 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.0036 | –0.5322 | 0.2869 |
| 19 | na | 212 | 0.54 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.0036 | –0.5331 | 0.2878 |
| 20 | * | 212 | 0.54 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.0046 | –0.4772 | 0.2323 |

[a]na, not available.
*, negative estimate.

**TABLE 8d Statistical Results for $\alpha = 0.03$, $p = 0.5$, $m = 4$, and $n = 36$**

| | | | | *p*-Value | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Model | $\alpha^a$ | Error df | $\sigma$ | All | Poly | Linear | Quad | Cubic | Quartic | Var | Bias | MSE |
| 1 | na | 280 | 0.61 | 0.00 | 0.00 | 0.01 | 0.29 | 0.93 | 0.78 | 0.0020 | –0.4389 | 0.1946 |
| 2 | 0.01 | 279 | 0.30 | 0.00 | 0.00 | 0.49 | 0.01 | 0.09 | 0.71 | 0.0005 | –0.3479 | 0.1215 |
| 3 | 0.00 | 279 | 0.42 | 0.00 | 0.00 | 0.00 | 0.07 | 0.99 | 0.72 | 0.0009 | –0.4412 | 0.1956 |
| 4 | na | 280 | 0.42 | 0.00 | 0.00 | 0.00 | 0.07 | 0.99 | 0.72 | 0.0009 | –0.4413 | 0.1957 |
| 5 | 0.01 | 279 | 0.30 | 0.00 | 0.00 | 0.77 | 0.01 | 0.11 | 0.70 | 0.0005 | –0.3533 | 0.1253 |
| 6 | na | 283 | 0.61 | 0.00 | 0.00 | 0.67 | 0.93 | 0.95 | 0.94 | 0.0045 | –0.7176 | 0.5195 |
| 7 | 0.01 | 282 | 0.30 | 0.00 | 0.00 | 0.14 | 0.39 | 0.52 | 0.58 | 0.0011 | –0.8847 | 0.7838 |
| 8 | 0.00 | 282 | 0.42 | 0.00 | 0.00 | 0.40 | 0.80 | 0.86 | 0.86 | 0.0022 | –0.6812 | 0.4662 |
| 9 | na | 283 | 0.42 | 0.00 | 0.00 | 0.40 | 0.80 | 0.86 | 0.86 | 0.0021 | –0.6804 | 0.4651 |
| 10 | 0.02 | 282 | 0.30 | 0.00 | 0.00 | 0.14 | 0.40 | 0.53 | 0.59 | 0.0012 | –0.8706 | 0.7591 |
| 11 | na | 209 | 0.26 | 0.00 | 0.00 | 0.00 | 0.00 | 0.02 | 0.00 | 0.0004 | –0.4389 | 0.1930 |
| 12 | 0.01 | 208 | 0.30 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.0005 | –0.3479 | 0.1215 |
| 13 | 0.00 | 71 | 0.26 | 0.00 | 0.00 | 0.00 | 0.00 | 0.91 | 0.00 | 0.0004 | –0.4412 | 0.1951 |
| 14 | na | 209 | 0.26 | 0.00 | 0.00 | 0.00 | 0.00 | 0.80 | 0.00 | 0.0004 | –0.4413 | 0.1951 |
| 15 | 0.02 | 208 | 0.29 | 0.00 | 0.00 | 0.98 | 0.00 | 0.00 | 0.00 | 0.0005 | –0.3573 | 0.1281 |
| 16 | na | 212 | 0.26 | 0.00 | 0.00 | 0.00 | 0.01 | 0.06 | 0.02 | 0.0151 | –0.7176 | 0.5300 |
| 17 | 0.01 | 211 | 0.30 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.0011 | –0.8847 | 0.7838 |
| 18 | 0.00 | 212 | 0.26 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.0059 | –0.6812 | 0.4700 |
| 19 | na | 212 | 0.26 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.0058 | –0.6804 | 0.4688 |
| 20 | 0.03 | 212 | 0.26 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.0016 | –0.8489 | 0.7222 |

[a]na, not available.

**TABLE 8e Statistical Results for $\alpha = 3$, $p = 1$, $m = 1$, and $n = 36$**

| | | | | *p*-Value | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Model | $\alpha^a$ | Error df | $\sigma$ | All | Poly | Linear | Quad | Cubic | Quartic | Var | Bias | MSE |
| 1 | na | 64 | 0.65 | 0.00 | 0.00 | 0.28 | 0.00 | 0.27 | 0.07 | 0.0090 | –0.1718 | 0.0385 |
| 3 | 0.01 | 63 | 0.52 | 0.00 | 0.00 | 0.27 | 0.00 | 0.20 | 0.02 | 0.0058 | –0.1726 | 0.0355 |
| 4 | na | 64 | 0.52 | 0.00 | 0.00 | 0.28 | 0.00 | 0.20 | 0.02 | 0.0057 | –0.1726 | 0.0355 |
| 6 | na | 67 | 0.65 | 0.00 | 0.00 | 0.08 | 0.39 | 0.48 | 0.46 | 0.0212 | 0.0054 | 0.0212 |
| 8 | 0.02 | 66 | 0.55 | 0.00 | 0.00 | 0.07 | 0.50 | 0.69 | 0.70 | 0.0155 | –0.0788 | 0.0217 |
| 9 | na | 67 | 0.55 | 0.00 | 0.00 | 0.08 | 0.54 | 0.74 | 0.77 | 0.0152 | –0.0946 | 0.0242 |
| 11 | na | 64 | 0.23 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.0090 | –0.1718 | 0.0385 |
| 13 | 0.01 | 63 | 0.18 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.0058 | –0.1726 | 0.0355 |
| 14 | na | 64 | 0.18 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.0057 | –0.1726 | 0.0355 |
| 16 | na | 67 | 0.23 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.0212 | 0.0054 | 0.0212 |
| 18 | 0.02 | 66 | 0.19 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.0155 | –0.0788 | 0.0217 |
| 19 | na | 67 | 0.19 | 0.00 | 0.00 | 0.00 | 0.00 | 0.01 | 0.02 | 0.0152 | –0.0946 | 0.0242 |

[a]na, not available.

**TABLE 8f Statistical Results for $\alpha = 0.03$, $p = 1$, $m = 1$, and $n = 36$**

| | | | | *p*-Value | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Model | $\alpha^a$ | Error df | $\sigma$ | All | Poly | Linear | Quad | Cubic | Quartic | Var | Bias | MSE |
| 1 | na | 64 | 0.59 | 0.08 | 0.08 | 0.05 | 0.51 | 0.35 | 0.67 | 0.0074 | –0.5025 | 0.2599 |
| 3 | 0.03 | 63 | 0.42 | 0.01 | 0.01 | 0.04 | 0.88 | 0.33 | 0.77 | 0.0038 | –0.4510 | 0.2072 |
| 4 | na | 64 | 0.43 | 0.02 | 0.02 | 0.07 | 0.68 | 0.38 | 0.83 | 0.0039 | –0.4378 | 0.1956 |
| 6 | na | 67 | 0.60 | 0.16 | 0.16 | 0.97 | 0.71 | 0.64 | 0.61 | 0.0177 | –0.6966 | 0.5030 |
| 8 | 0.03 | 66 | 0.42 | 0.01 | 0.01 | 0.87 | 0.77 | 0.67 | 0.64 | 0.0086 | –0.6912 | 0.4864 |
| 9 | na | 67 | 0.42 | 0.02 | 0.02 | 0.84 | 0.81 | 0.72 | 0.68 | 0.0089 | –0.6901 | 0.4851 |
| 11 | na | 64 | 0.21 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.0074 | –0.5025 | 0.2599 |
| 13 | 0.03 | 63 | 0.15 | 0.00 | 0.00 | 0.00 | 0.24 | 0.00 | 0.02 | 0.0038 | –0.4510 | 0.2072 |
| 14 | na | 64 | 0.15 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.10 | 0.0039 | –0.4378 | 0.1956 |
| 16 | na | 67 | 0.21 | 0.00 | 0.00 | 0.80 | 0.00 | 0.00 | 0.00 | 0.0177 | –0.6966 | 0.5030 |
| 18 | 0.03 | 66 | 0.15 | 0.00 | 0.00 | 0.19 | 0.02 | 0.00 | 0.00 | 0.0086 | –0.6912 | 0.4864 |
| 19 | na | 67 | 0.15 | 0.00 | 0.00 | 0.10 | 0.05 | 0.00 | 0.00 | 0.0089 | –0.6901 | 0.4851 |

[a]na, not available.

**TABLE 8g Statistical Results for $\alpha = 3$, $p = 0.5$, $m = 1$, and $n = 36$**

| | | | | *p*-Value | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Model | $\alpha^a$ | Error df | $\sigma$ | All | Poly | Linear | Quad | Cubic | Quartic | Var | Bias | MSE |
| 1 | na | 64 | 0.88 | 0.41 | 0.41 | 0.37 | 0.74 | 0.81 | 0.27 | 0.0167 | –0.3866 | 0.1662 |
| 3 | * | 63 | 0.58 | 0.09 | 0.09 | 0.97 | 0.12 | 0.86 | 0.09 | 0.0074 | –0.2748 | 0.0829 |
| 4 | na | 64 | 0.59 | 0.10 | 0.10 | 0.77 | 0.18 | 0.83 | 0.10 | 0.0075 | –0.2965 | 0.0954 |
| 6 | na | 67 | 0.88 | 0.24 | 0.24 | 0.46 | 0.51 | 0.48 | 0.43 | 0.0384 | –0.6445 | 0.4539 |
| 8 | * | 66 | 0.59 | 0.02 | 0.02 | 0.53 | 0.72 | 0.72 | 0.67 | 0.0173 | –0.6881 | 0.4908 |
| 9 | na | 67 | 0.60 | 0.03 | 0.03 | 0.48 | 0.64 | 0.63 | 0.57 | 0.0178 | –0.6798 | 0.4799 |
| 11 | na | 64 | 0.31 | 0.00 | 0.00 | 0.00 | 0.01 | 0.05 | 0.00 | 0.0167 | –0.3866 | 0.1662 |
| 13 | * | 63 | 0.20 | 0.00 | 0.00 | 0.75 | 0.00 | 0.15 | 0.00 | 0.0074 | –0.2748 | 0.0829 |
| 14 | na | 64 | 0.21 | 0.00 | 0.00 | 0.02 | 0.00 | 0.09 | 0.00 | 0.0075 | –0.2965 | 0.0954 |
| 16 | na | 67 | 0.31 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.0384 | –0.6445 | 0.4539 |
| 18 | * | 66 | 0.21 | 0.00 | 0.00 | 0.00 | 0.01 | 0.01 | 0.00 | 0.0173 | –0.6881 | 0.4908 |
| 19 | na | 67 | 0.21 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.0178 | –0.6798 | 0.4799 |

[a]na, not available.
*, negative estimate.

**TABLE 8h Statistical Results for $\alpha = 0.03$, $p = 0.5$, $m = 0$, and $n = 36$**

| | | | | *p*-Value | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Model | $\alpha^a$ | Error df | $\sigma$ | All | Poly | Linear | Quad | Cubic | Quartic | Var | Bias | MSE |
| 1 | na | 64 | 0.62 | 0.17 | 0.17 | 0.10 | 0.67 | 0.88 | 0.21 | 0.0082 | –0.4546 | 0.2149 |
| 3 | 0.00 | 63 | 0.43 | 0.01 | 0.01 | 0.00 | 0.84 | 0.38 | 0.73 | 0.0040 | –0.4825 | 0.2368 |
| 4 | na | 64 | 0.43 | 0.01 | 0.01 | 0.00 | 0.84 | 0.38 | 0.73 | 0.0039 | –0.4823 | 0.2366 |
| 6 | na | 67 | 0.61 | 0.16 | 0.16 | 0.08 | 0.10 | 0.10 | 0.11 | 0.0184 | –0.6983 | 0.5061 |
| 8 | 0.01 | 66 | 0.44 | 0.04 | 0.04 | 0.36 | 0.59 | 0.68 | 0.73 | 0.0096 | –0.6946 | 0.4921 |
| 9 | na | 67 | 0.44 | 0.04 | 0.04 | 0.40 | 0.65 | 0.74 | 0.80 | 0.0095 | –0.6944 | 0.4917 |
| 11 | na | 64 | 0.22 | 0.00 | 0.00 | 0.00 | 0.00 | 0.23 | 0.00 | 0.0082 | –0.4546 | 0.2149 |
| 13 | 0.00 | 63 | 0.15 | 0.00 | 0.00 | 0.00 | 0.11 | 0.00 | 0.01 | 0.0040 | –0.4825 | 0.2368 |
| 14 | na | 64 | 0.15 | 0.00 | 0.00 | 0.00 | 0.11 | 0.00 | 0.01 | 0.0039 | –0.4823 | 0.2366 |
| 16 | na | 67 | 0.21 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.0184 | –0.6983 | 0.5061 |
| 18 | 0.01 | 66 | 0.15 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.01 | 0.0096 | –0.6946 | 0.4921 |
| 19 | na | 67 | 0.15 | 0.00 | 0.00 | 0.00 | 0.00 | 0.01 | 0.04 | 0.0095 | –0.6944 | 0.4917 |

[a]na, not available.

**TABLE 9a Statistical Results for $\alpha = 3$, $p = 0.5$, $m = 4$, and $n = 200$**

| | | | | *p*-Value | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Model | $\alpha^a$ | Error df | $\sigma$ | All | Poly | Linear | Quad | Cubic | Quartic | Var | Bias | MSE |
| 1 | na | 1592 | 0.72 | 0.00 | 0.00 | 0.47 | 0.00 | 0.78 | 0.64 | 0.0005 | –0.2648 | 0.0706 |
| 2 | 0.08 | 1591 | 0.61 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.07 | 0.0004 | –0.1876 | 0.0355 |
| 3 | 0.00 | 1591 | 0.51 | 0.00 | 0.00 | 0.67 | 0.00 | 0.87 | 0.57 | 0.0003 | –0.2626 | 0.0692 |
| 4 | na | 1592 | 0.51 | 0.00 | 0.00 | 0.69 | 0.00 | 0.88 | 0.57 | 0.0003 | –0.2625 | 0.0692 |
| 5 | * | 1591 | 0.51 | 0.00 | 0.00 | 0.64 | 0.00 | 0.90 | 0.58 | 0.0003 | –0.2630 | 0.0694 |
| 6 | na | 1595 | 0.72 | 0.00 | 0.00 | 0.00 | 0.07 | 0.23 | 0.26 | 0.0012 | –0.4804 | 0.2319 |
| 7 | 0.08 | 1594 | 0.61 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.0009 | –0.7330 | 0.5383 |
| 8 | 0.00 | 1594 | 0.51 | 0.00 | 0.00 | 0.00 | 0.02 | 0.12 | 0.14 | 0.0006 | –0.5147 | 0.2655 |
| 9 | na | 1595 | 0.51 | 0.00 | 0.00 | 0.00 | 0.02 | 0.12 | 0.14 | 0.0006 | –0.5156 | 0.2665 |
| 10 | * | 1594 | 0.51 | 0.00 | 0.00 | 0.00 | 0.02 | 0.13 | 0.15 | 0.0007 | –0.5142 | 0.2651 |
| 11 | na | 1193 | 0.51 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.0003 | –0.2648 | 0.0704 |
| 12 | 0.08 | 1192 | 0.61 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.0004 | –0.1876 | 0.0355 |
| 13 | 0.00 | 399 | 0.51 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.0003 | –0.2626 | 0.0692 |
| 14 | na | 1193 | 0.51 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.0003 | –0.2625 | 0.0692 |
| 15 | * | 1192 | 0.51 | 0.00 | 0.00 | 0.00 | 0.00 | 0.02 | 0.00 | 0.0003 | –0.2643 | 0.0701 |
| 16 | na | 1196 | 0.51 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.0028 | –0.4804 | 0.2336 |
| 17 | 0.08 | 1195 | 0.61 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.0009 | –0.7330 | 0.5383 |
| 18 | 0.00 | 1196 | 0.51 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.0006 | –0.5147 | 0.2655 |
| 19 | na | 1196 | 0.51 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.0006 | –0.5156 | 0.2665 |
| 20 | * | 1196 | 0.51 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.0007 | –0.5109 | 0.2617 |

[a]na, not available.
*, negative estimate.

**TABLE 9b Statistical Results for $\alpha = 0.03$, $p = 0.5$, $m = 4$, and $n = 200$**

| | | | | *p*-Value | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Model | $\alpha^a$ | Error df | $\sigma$ | All | Poly | Linear | Quad | Cubic | Quartic | Var | Bias | MSE |
| 1 | na | 1592 | 0.64 | 0.00 | 0.00 | 0.00 | 0.43 | 0.60 | 0.87 | 0.0004 | –0.4250 | 0.1810 |
| 2 | 0.01 | 1591 | 0.30 | 0.00 | 0.00 | 0.53 | 0.00 | 0.10 | 0.05 | 0.0001 | –0.3517 | 0.1238 |
| 3 | 0.00 | 1591 | 0.41 | 0.00 | 0.00 | 0.00 | 0.01 | 0.96 | 0.61 | 0.0002 | –0.4307 | 0.1856 |
| 4 | na | 1592 | 0.41 | 0.00 | 0.00 | 0.00 | 0.01 | 0.96 | 0.61 | 0.0002 | –0.4307 | 0.1856 |
| 5 | 0.02 | 1591 | 0.30 | 0.00 | 0.00 | 0.42 | 0.00 | 0.12 | 0.06 | 0.0001 | –0.3603 | 0.1299 |
| 6 | na | 1595 | 0.64 | 0.00 | 0.00 | 0.70 | 0.90 | 0.82 | 0.74 | 0.0009 | –0.8348 | 0.6977 |
| 7 | 0.01 | 1594 | 0.30 | 0.00 | 0.00 | 0.00 | 0.01 | 0.02 | 0.02 | 0.0002 | –0.9035 | 0.8165 |
| 8 | 0.00 | 1594 | 0.41 | 0.00 | 0.00 | 0.06 | 0.38 | 0.41 | 0.37 | 0.0004 | –0.7474 | 0.5589 |
| 9 | na | 1595 | 0.41 | 0.00 | 0.00 | 0.06 | 0.38 | 0.41 | 0.37 | 0.0004 | –0.7475 | 0.5591 |
| 10 | 0.02 | 1594 | 0.30 | 0.00 | 0.00 | 0.00 | 0.01 | 0.02 | 0.02 | 0.0002 | –0.8827 | 0.7793 |
| 11 | na | 1193 | 0.26 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.0001 | –0.4250 | 0.1807 |
| 12 | 0.01 | 1192 | 0.30 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.0001 | –0.3517 | 0.1238 |
| 13 | 0.00 | 399 | 0.26 | 0.00 | 0.00 | 0.00 | 0.00 | 0.20 | 0.00 | 0.0001 | –0.4307 | 0.1855 |
| 14 | na | 1193 | 0.26 | 0.00 | 0.00 | 0.00 | 0.00 | 0.02 | 0.00 | 0.0001 | –0.4307 | 0.1855 |
| 15 | 0.03 | 1192 | 0.30 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.0001 | –0.3641 | 0.1326 |
| 16 | na | 1196 | 0.26 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.0030 | –0.8348 | 0.6999 |
| 17 | 0.01 | 1195 | 0.30 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.0002 | –0.9035 | 0.8165 |
| 18 | 0.00 | 1196 | 0.26 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.0010 | –0.7474 | 0.5596 |
| 19 | na | 1196 | 0.26 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.0010 | –0.7475 | 0.5598 |
| 20 | 0.04 | 1196 | 0.26 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.0003 | –0.8708 | 0.7586 |

[a]na, not available.

**TABLE 9c Statistical Results for $\alpha = 3$, $p = 0.1$, $m = 4$, and $n = 500$**

| | | | | *p*-Value | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Model | $\alpha^a$ | Error df | $\sigma$ | All | Poly | Linear | Quad | Cubic | Quartic | Var | Bias | MSE |
| 1 | na | 3992 | 0.72 | 0.00 | 0.00 | 0.09 | 0.00 | 0.46 | 0.47 | 0.0002 | –0.3766 | 0.1420 |
| 2 | 0.08 | 3991 | 0.61 | 0.00 | 0.00 | 0.47 | 0.00 | 0.92 | 0.18 | 0.0001 | –0.3597 | 0.1296 |
| 3 | 0.00 | 3991 | 0.50 | 0.00 | 0.00 | 0.10 | 0.00 | 0.17 | 0.26 | 0.0001 | –0.3748 | 0.1406 |
| 4 | na | 3992 | 0.50 | 0.00 | 0.00 | 0.10 | 0.00 | 0.17 | 0.26 | 0.0001 | –0.3749 | 0.1406 |
| 5 | * | 3991 | 0.50 | 0.00 | 0.00 | 0.08 | 0.00 | 0.16 | 0.27 | 0.0001 | –0.3754 | 0.1411 |
| 6 | na | 3995 | 0.72 | 0.00 | 0.00 | 0.88 | 0.38 | 0.31 | 0.32 | 0.0005 | –0.8959 | 0.8031 |
| 7 | 0.08 | 3994 | 0.61 | 0.00 | 0.00 | 0.94 | 0.34 | 0.20 | 0.16 | 0.0003 | –0.9501 | 0.9030 |
| 8 | 0.00 | 3994 | 0.50 | 0.00 | 0.00 | 0.46 | 0.13 | 0.10 | 0.12 | 0.0002 | –0.9236 | 0.8532 |
| 9 | na | 3995 | 0.50 | 0.00 | 0.00 | 0.47 | 0.13 | 0.10 | 0.12 | 0.0002 | –0.9231 | 0.8523 |
| 10 | * | 3994 | 0.50 | 0.00 | 0.00 | 0.46 | 0.12 | 0.10 | 0.12 | 0.0002 | –0.9217 | 0.8497 |
| 11 | na | 2993 | 0.51 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.0001 | –0.3766 | 0.1419 |
| 12 | 0.08 | 2992 | 0.61 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.0001 | –0.3597 | 0.1296 |
| 13 | 0.00 | 999 | 0.50 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.0001 | –0.3748 | 0.1406 |
| 14 | na | 2993 | 0.50 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.0001 | –0.3749 | 0.1406 |
| 15 | * | 2992 | 0.50 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.0001 | –0.3754 | 0.1411 |
| 16 | na | 2996 | 0.51 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.0011 | –0.8959 | 0.8037 |
| 17 | 0.08 | 2995 | 0.61 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.0003 | –0.9501 | 0.9030 |
| 18 | 0.00 | 2996 | 0.50 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.0002 | –0.9236 | 0.8532 |
| 19 | na | 2996 | 0.50 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.0002 | –0.9231 | 0.8523 |
| 20 | * | 2996 | 0.50 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.0002 | –0.9217 | 0.8497 |

[a]na, not available.
*, negative estimate.

**TABLE 9d Statistical Results for $\alpha = 0.03$, $p = 0.1$, $m = 4$, and $n = 500$**

| | | | | *p*-Value | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Model | $\alpha^a$ | Error df | $\sigma$ | All | Poly | Linear | Quad | Cubic | Quartic | Var | Bias | MSE |
| 1 | na | 3992 | 0.6365 | 0.81 | 0.81 | 0.53 | 0.74 | 0.87 | 0.98 | 0.0002 | −0.3829 | 0.1468 |
| 2 | 0.01 | 3991 | 0.3033 | 0.86 | 0.86 | 0.83 | 0.69 | 0.43 | 0.87 | 0.0000 | −0.3777 | 0.1427 |
| 3 | 0.00 | 3991 | 0.3978 | 0.43 | 0.43 | 0.33 | 0.61 | 0.80 | 0.97 | 0.0001 | −0.3828 | 0.1466 |
| 4 | na | 3992 | 0.3978 | 0.43 | 0.43 | 0.33 | 0.61 | 0.80 | 0.97 | 0.0001 | −0.3828 | 0.1466 |
| 5 | 0.03 | 3991 | 0.299 | 0.81 | 0.81 | 0.99 | 0.66 | 0.46 | 0.89 | 0.0000 | −0.3784 | 0.1432 |
| 6 | na | 3995 | 0.6362 | 0.73 | 0.73 | 0.88 | 0.80 | 0.79 | 0.79 | 0.0004 | −0.9788 | 0.9583 |
| 7 | 0.01 | 3994 | 0.3033 | 0.77 | 0.77 | 0.67 | 0.49 | 0.42 | 0.39 | 0.0001 | −0.9938 | 0.9878 |
| 8 | 0.00 | 3994 | 0.3977 | 0.29 | 0.29 | 0.80 | 0.67 | 0.66 | 0.67 | 0.0001 | −0.9796 | 0.9597 |
| 9 | na | 3995 | 0.3977 | 0.29 | 0.29 | 0.80 | 0.67 | 0.66 | 0.66 | 0.0001 | −0.9796 | 0.9597 |
| 10 | 0.03 | 3994 | 0.299 | 0.71 | 0.71 | 0.67 | 0.50 | 0.44 | 0.41 | 0.0001 | −0.9919 | 0.9839 |
| 11 | na | 2993 | 0.2621 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.0000 | −0.3829 | 0.1466 |
| 12 | 0.01 | 2992 | 0.3033 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.0000 | −0.3777 | 0.1427 |
| 13 | 0.00 | 999 | 0.2621 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.04 | 0.0000 | −0.3828 | 0.1466 |
| 14 | na | 2993 | 0.2621 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.0000 | −0.3828 | 0.1466 |
| 15 | 0.03 | 2992 | 0.299 | 0.00 | 0.00 | 0.59 | 0.00 | 0.00 | 0.00 | 0.0000 | −0.3784 | 0.1432 |
| 16 | na | 2996 | 0.262 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.0012 | −0.9788 | 0.9592 |
| 17 | 0.01 | 2995 | 0.3033 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.0001 | −0.9938 | 0.9878 |
| 18 | 0.00 | 2996 | 0.262 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.0004 | −0.9796 | 0.9599 |
| 19 | na | 2996 | 0.262 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.0004 | −0.9796 | 0.9599 |
| 20 | 0.03 | 2996 | 0.262 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.0001 | −0.9919 | 0.9839 |

[a]na, not available.

## Translational Aspects of Pathodynamics

Signals are relatively easy to isolate and model in in vitro conditions, due mostly to the ability to control the environment and to the lack of exposure to the biological system network. If the biomarker cannot be characterized and modeled in the laboratory, the chance of getting meaningful information in vivo is slim to none. Once a laboratory model is established, an animal model can be chosen and the pathodynamics work starts all over again until the biomarker can be characterized and modeled in vivo. As all biologists know, just because it works in one species, it is not necessarily going to work in another species. Before going into humans, a pathodynamic model for the biomarker should be studied in several species. The similarities and differences in these interspecies results should provide guidance about the applicability in humans.

For translation across species, the pathodynamic models must have some characteristics that are invariant. Without this invariance, all the preclinical work will probably be a waste of time and money. Mathematical physics would not exist without laws of invariance such as conservation of mass and energy. The same will probably hold true in biology. In the context of this chapter, $P$ (probability structure), $\mathbf{Y}$ (biomarkers), and $D$ (disease or decision space) have to be invariant in some sense. Here having invariance in $P$ is not as strong as it seems. The simplest distribution type is the same, but the parameters of the distribution model have species-specific variation. A more complicated type of invariance is that the topologies of the interspecies probability spaces are equivalent. This just means that any "physical" distortion of the response distribution does not create or remove any holes. To a greater extreme, a type of invariance would be present if there is a one-for-one matching of probability objects between species (i.e., when a particular probability object is present in one species with disease, or response, $D$, there is always a probability object, not necessarily similar, in another species that represents $D$). The bottom line for biologists is that this will probably require more mathematics than most can presently do.

This approach to translation is relatively worthless without laboratory standards. When biomarkers are developed and applied, either the exact assay methods and standards must be applied in every experiment or there needs to be a mathematical transformation that makes them equivalent. Currently in the clinic, these standards do not exist. Therefore, preclinical experiments using the same methods may work fine, but when the clinical trial is run, variation in methods and sample handling may distort or destroy the information.

## Future Needs in Method Development

Biologists need to get involved directly in pathodynamics to get an efficient merger of the biology and the mathematics. It is a rare mathematician or

statistician that has biology training and intuition. Once the biologist gets involved in developing the models, progress will accelerate. Remember, the OU model is basically the simplest case for a pathodynamic model. As is illustrated in the examples above, simpler models will have information loss. Therefore, standard experimental design and analysis may not be sufficient.

The second issue is whether the OU model is correct. Preliminary research suggests that it is not [34], but only minor modifications may be needed for modeling homeostasis (i.e., dynamic equilibrium). The models for disease or therapeutic effects are mostly unknown. Chronic effects may be directional diffusion or slow convection, while acute effects are likely to generate trajectories such as liver injury [31]. The mathematics of statistical physics [17] is likely to be needed.

It seems clear from the examples presented here that the current statistical estimation algorithms commonly used and available are not efficient in a Fisher information sense when autocorrelation is present. This has been handled in the economic applications for equally spaced measurement times, but biology is not quite so regular, especially clinical trials, even under the strictest protocols.

The communication/information theory and decision theory that is presented here was only an introduction. Optimal information and decision algorithms need to be developed in the context of pathodynamics. Such algorithms may be synergistic with Fisher information optimization or may have some conflict. How the information will be used should determine the optimization approach.

In this chapter biomarkers have been defined as functions of parameters, as vectors of tests, and as signals. These are just aspects of the same mathematical object called a probability distribution function $P(\mathbf{Y})$. The parameters are an integral part of the probability model, the vector $\mathbf{Y}$ represents the measurements that get combined in the model, and change in these measurements with time is the signal.

## REFERENCES


1. Klotz IM, Rosenberg RM (1994). *Chemical Thermodynamics: Basic Theory and Methods*. Wiley, New York.
2. Kondepudi D, Prigogine I (1998). *Modern Thermodynamics: From Heat Engines to Dissipative Structure*. Wiley, New York.
3. Keener J, Sneyd J (1998). *Mathematical Physiology*. Springer-Verlag, New York.
4. Box GEP, Hunter WG, Hunter JS (1978). *Statistics for Experimenters: An Introduction to Design, Data Analysis, and Model Building*. Wiley, New York.
5. Woolson RF, Clarke WR (2002). *Statistical Methods for the Analysis of Biomedical Data*, 2nd ed. Wiley, Hoboken, NJ.
6. Mendenhall W, Sincich T (1995). *Statistics for Engineering and the Sciences*, 4th ed. Prentice Hall, Upper Saddle River, NJ.

7. Puri ML, Sen PK (1971). *Nonparametric Methods in Multivariate Analysis*. Wiley, New York.
8. Thompson PA (1972). *Compressible-Fluid Dynamics*. McGraw-Hill, New York.
9. Hawking SW (1988). *A Brief History of Time: From the Big Bang to Black Holes*. Bantam Books, New York.
10. Prigogine I (1996). *The End of Certainty: Time, Chaos, and the New Laws of Nature*. Free Press, New York.
11. Frieden BR (1998). *Physics from Fisher Information*. Cambridge University Press, Cambridge, UK.
12. Trost DC (2008). A method for constructing and estimating the RR-memory of the QT-interval and its inclusion in a multivariate biomarker for torsades de pointes risk. *J Biopharm Stat*, 18(4):773–796.
13. Brown R (1828). A brief account of microscopic observations made in the months of June, July, and August, 1827, on the particles contained in the pollen of plants; and on the general existence of active molecules in organic and inorganic bodies. *Philos Mag*, 4:161–173.
14. Karatzas I, Shreve SE (1991). *Brownian Motion and Stochastic Calculus*, 2nd ed. Springer-Verlag, New York.
15. Øksendal B (1998). *Stochastic Differential Equations: An Introduction with Applications*, 5th ed. Springer-Verlag, Berlin.
16. Uhlenbeck GE, Ornstein LS (1930). *Phys Rev*, 36:823–841.
17. Reichl LE (1998). *A Modern Course in Statistical Physics*, 2nd ed. Wiley, New York.
18. Trost DC (2006). Multivariate probability-based detection of drug-induced hepatic signals. *Toxicol Rev*, 25(1):37–54.
19. Hogg RV, Craig A, McKean JW (2004). *Introduction to Mathematical Statistics*, 6th ed. Prentice Hall, Upper Saddle River, NJ.
20. Bickel PJ, Doksum KA (1977). *Mathematical Statistics: Basic Ideas and Selected Topics*. Holden-Day, San Francisco.
21. Stuart A, Ord JK (1991). *Kendall's Advanced Theory of Statistics, vol. 2, Classical Inference and Relationship*, 5th ed. Oxford University Press, New York.
22. Reza FM (1994). *An Introduction to Information Theory*. Dover, Mineola, NY.
23. Kullback S (1968). *Information Theory and Statistics*. Dover, Mineola, NY.
24. Williams SA, Slavin DE, Wagner JA, Webster CJ (2006). A cost-effectiveness approach to the qualification and acceptance of biomarkers. *Nat Rev Drug Discov*, 5:897–902.
25. Wald A (1971). *Statistical Decision Functions*. Chelsea Publishing, New York.
26. Blackwell DA, Girshick MA (1979). *Theory of Games and Statistical Decisions*. Dover, Mineola, NY.
27. Chernoff H, Moses LE (1987). *Elementary Decision Theory*. Dover, Mineola, NY.
28. Knuth DE (1981). *The Art of Computer Programming,* vol. 2*, Seminumerical Algorithms*, 2nd ed. Addison-Wesley, Reading, MA.
29. Trost DC (2007). *An introduction to pathodynamics from the view of homeostasis and beyond*. Presented at the Sixth International Congress on Industrial and Applied Mathematics, Zürich, Switzerland, July 16–20.

30. Rosenkranz GK (2009). Modeling laboratory data from clinical trials. *Comput Stat Data An*, 53(3):812–819.
31. Trost DC, Freston JW (2008). Vector analysis to detect hepatotoxicity signals in drug development. *Drug Inf J*, 42(1):27–34.
32. R Development Core Team (2007). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org.
33. Burtis CA, Ashwood ER (eds.) (1999). *Tietz Textbook of Clinical Chemistry*, 3rd ed. W.B Saunders, Philadelphia.
34. Trost DC, Overman EA, Ostroff JH, Xiong W, March PD (in press). A model for liver homeostasis using a modified mean-reverting Ornstein–Uhlenbeck process.

# OPTIMIZING THE USE OF BIOMARKERS FOR DRUG DEVELOPMENT: A CLINICIAN'S PERSPECTIVE

ALBERTO GIMONA, M.D.
*Merck Serono International, Geneva, Switzerland*

## INTRODUCTION

Drug development is currently facing many challenges, from the ever-increasing costs of developing drugs to the reducing numbers of new drug approvals. Bringing a drug to market requires on average 12 years and $1 billion. Approximately 53% of compounds entering phase II fail, resulting in amortized costs of approximately $0.8 billion per registered drug (DiMasi et al., 2003). The majority of these costs are due to failures encountered during drug development; this represents approximately 75% of the costs for registration (Figure 1). According to the Pharmaceutical Research and Manufacturers of America (PhRMA), the U.S. biopharmaceutical industry spends $49.3 billion per year on drug research and development. According to the 2004 estimate of the U.S. Food and Drug Administration (FDA), only 8% of drugs entering clinical trials had a legitimate chance of reaching the market.

The reasons for drug development failure vary: from 1990 to 1999, the most relevant reason for drug development failure was related to pharmacokinetic (PK) and bioavailability issues (accounting for approximately 40% of

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

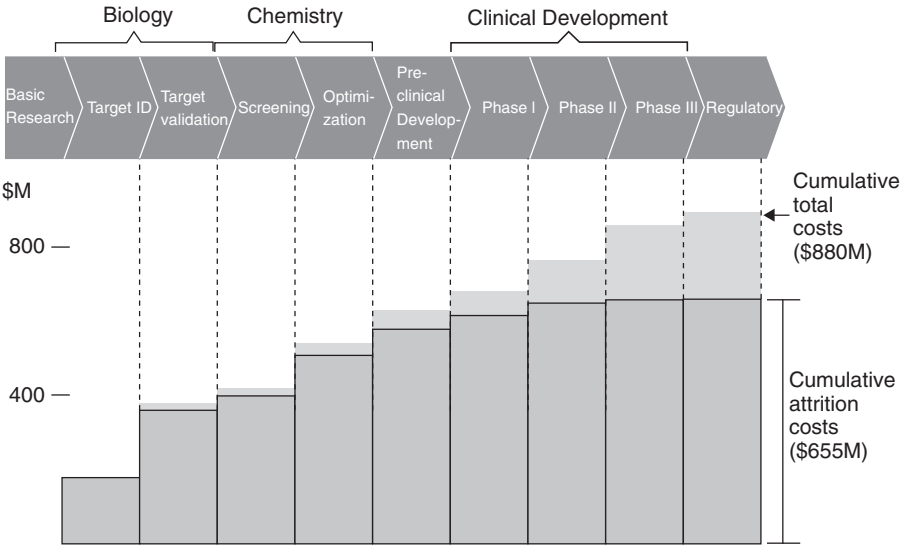


**Figure 1** Components of the overall costs for one approved drug.

failures), while from 2000 to 2008, the leading causes for failure were related to efficacy (approximately 30%), while PK and bioavailability failures diminished to approximately 10% (Figure 2). As shown in Figure 1, the late-stage failures tended to be extremely expensive.

It is now well established that tools such as biomarkers are able to predict the efficacy of a compound at the early stage of development. This can drastically increase the efficiency of the development process, resulting in an increased productivity. These lessons learned demonstrate the need to decrease the time required for drug approval as well as a decline in late-stage failures. Biomarkers, including imaging biomarkers, can address these needs. In the early drug development stage, biomarkers can be instrumental in ameliorating the decision-making process, introducing the concept of "fail early, fail fast," thus allowing drug companies to concentrate their resources on the most promising drug candidates. In addition, a biomarker strategy applied during late-stage development would allow a drug candidate to achieve regulatory approval much earlier in cases where resources are limited.

Fundamental differences do exist, depending on whether the biomarker strategy is applied to early- or late-stage drug development; implementing a biomarker strategy during early drug development may represent a risk that a company can more readily absorb. Indeed, biomarkers at the early stage are validated through the various phases of drug development and would be used solely to determine whether or not to proceed in development with the drug candidate in question. Conversely, a biomarker that is validated during the late phase of drug development could be used for drug approval purposes.

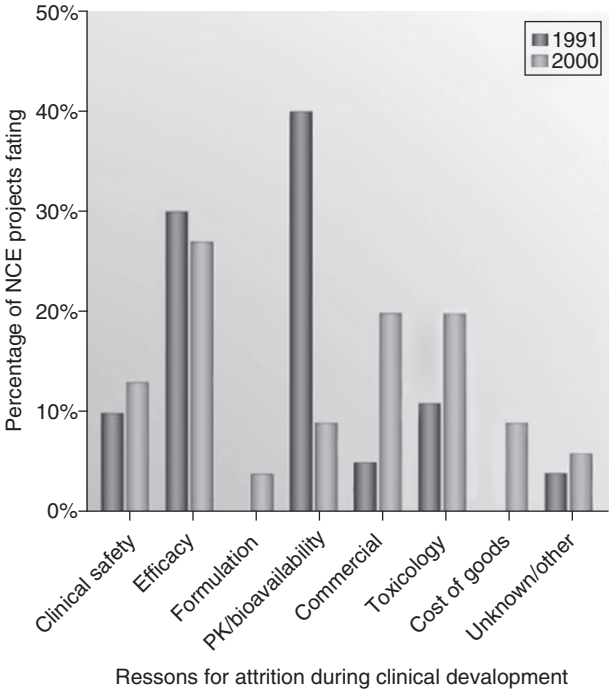


**Figure 2** Comparison of the reasons for drug failure in 1991 vs. 2000.

## DEFINITION AND CLASSIFICATION OF BIOMARKERS

In 2001, the Biomarkers Definition Working Group, sponsored by the National Institutes of Health, defined a *biomarker* as "a characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention"; a *clinical endpoint* as "a characteristic or variable that reflects how a patient feels or functions, or how long a patient survives"; and a *surrogate endpoint* as "a biomarker that is intended to substitute for a clinical endpoint."

There are important differences among biomarkers. Type 0 biomarkers are natural history markers of a disease and tend to correlate longitudinally with known clinical references, such as symptoms. Type I biomarkers capture the effect of an intervention in accordance with the mechanism of action of the drug, even though the mechanism might not be known to be associated with the clinical outcome. Type II biomarkers are considered surrogate markers since the change of a specific marker predicts a clinical benefit (Frank and Hargreaves, 2003).

A hierarchy also exists among clinical endpoints. A clinical endpoint can be defined as an *intermediate endpoint*, which is a clinical endpoint that is not

the ultimate outcome but is nonetheless of real clinical benefit; the *ultimate outcome* is a clinical endpoint such as survival, onset of serious morbidity, or symptomatic response that captures the benefits and risks of an intervention (Lesko and Atkinson, 2001).

Classification can be made according to the specificity of a biomarker versus the intended therapeutic response. A linked drug biomarker demonstrates a strict correlation between the pharmacological action of the drug and its effect on the disease. Danhof et al. (2005) have proposed to classify biomarkers in three distinct categories: (1) pharmacological and (2) toxicological markers, both observed in healthy subjects, and (3) pathological biomarkers observed in subjects affected by disease (Danhof et al., 2005). It is noteworthy to review the definition of biomarkers from the 2003 FDA "Guidance for Industry: Exposure–Response Relationships":

- Biomarkers are considered valid surrogates for clinical benefit (e.g., blood pressure, cholesterol, viral load).
- Biomarkers are thought to reflect the pathologic process and at least be candidate surrogates (e.g., brain appearance in Alzheimer's disease, brain infarct size, various radiographic/isotopic function tests).
- Biomarkers reflect drug action but of uncertain relation to a clinical outcome (e.g., inhibition of ADP-dependent platelet aggregation, ACE inhibition).
- Biomarkers may be more remote from the clinical benefit endpoint (e.g., degree of binding to a receptor or inhibition of an agonist).

### Classification Based on Mechanism of Action

The COST B15 working group 2, "Markers of Pharmacological and Toxicological Action," has proposed a conceptually similar classification. Based on the location of the biomarker in the chain of events from underlying subject genotype or phenotype to clinical scales, the following types of biomarkers have been be defined:

*Type 0:* genotype or phenotype
*Type 1:* concentration
*Type 2:* target occupancy
*Type 3:* target activation
*Type 4:* physiologic measures or laboratory tests
*Type 5:* disease processes
*Type 6:* clinical scales.

This classification is not universally accepted since the type 0 biomarker relating to a subject's genotype or phenotype can be considered a covariate

rather than a biomarker. Similarly, the type 6 biomarker, such as a clinical scale, can be regarded as a measurement of a clinical endpoint and not a biomarker. These classifications have been proposed in attempt to reconcile disagreements on the potential role of biomarkers.

### Classification Based on Clinical Applications

The paradigm is now shifting from the classical model of clinical care to development and application of biomarkers in different therapeutic areas according to their clinical application, which can be classified as follows:

- *Preventive biomarkers*, which identify people at a high risk of developing disease
- *Diagnostic biomarkers*, which identify a disease at the earliest stage, before clinical symptoms occur
- *Prognostic biomarkers*, which stratify the risk of disease progression in patients undergoing specific therapy
- *Predictive biomarkers*, which identify patients who respond to specific therapeutic interventions
- *Therapeutic biomarkers*, which provide a quantifiable measure of response in patients who undergo treatment

### Classification According to Measurement Scale

From a mathematical perspective, biomarkers can also be classified on the basis of the measurement scale it utilizes:

- *Graded response*, which is a quantifiable biomarker that is causally linked to drug treatment and temporally related to drug exposure (e.g., blood pressure, cholesterol). Usually such endpoints are chosen based on the pharmacodynamic response.
- *Challenge response*, which is a quantifiable graded response to a standardized exogenous challenge, modified by the administration of the drug (e.g., challenge test in asthma).

Usually, these markers are based on the mode of administration (MoA) of the drug and the response is a continuous variable. Other types of responses can be observed with biomarkers:

- *Categorical response* is usually a "Yes" or "No" response for a clinically relevant outcome based on the disease progression, regardless of MoA (e.g., response based on tumor size, incidence of an AE); such an event is generally not linked to the MoA of the drug.

- *Time to event response* is usually a clinically relevant outcome regardless of the MoA, such as survival time or time to relapse. It is a censored continuous clinical variable which can be measured only once for each patient.
- *Event frequency/rate of response* is the frequency of clinical events related to drug exposure (e.g., MRI lesions in multiple sclerosis); it is usually a censored continuous variable.

## DEVELOPMENT OF BIOMARKERS

The development of biomarkers can be divided artificially into two steps: (1) evaluation/qualification of the candidate biomarker and (2) validation of biomarkers to become a surrogate endpoint.

### Evaluation and Qualification of Biomarkers

Many disease biomarkers are well characterized and are used extensively in drug development. However, there is the frequently need to develop new biomarkers, especially in new therapeutic areas and/or when dealing with innovative therapeutic approaches. Development of new biomarkers should start at the preclinical stage with the intent to have a new biomarker when the lead candidate enters the human development stage. The objective of biomarker development should be clearly defined, such as the need for markers related to disease progression or the pharmacological effect of the drug or for markers indicating therapeutic activity.

In the evaluation phase, the candidate biomarker should be measured against the following attributes that define a biomarker (Lesko and Atkinson, 2001):

- *Clinical relevance*, which may theoretically reflect a physiologic or pathologic process or activity over a relatively short period of time. Ideally, this effect should be related to the MoA of the drug and to the clinical endpoint. This obviously requires an understanding of the pathophysiology of a disease and of a drug's mechanism of action, taking into consideration the fact that diseases frequently have multiple causal pathways.
- *Sensitivity and specificity to treatment effects*, defined as the ability to detect the intended measurement or change in the target patient population.
- *Reliability*, defined as the ability to measure the biomarker analytically with accuracy, precision, robustness, and reproducibility.
- *Practicality*, defined as noninvasiveness or only modest invasiveness.
- *Simplicity*, for routine utilization without the need for sophisticated equipment or operator skill, extensive time commitment, or high measurement cost.

## Validation of Biomarkers

The validation of a biomarker is a work in progress that ends when the biomarker is validated as a surrogate endpoint. During the development of a biomarker, aside from the characteristics mentioned above, the investigator should take into account the risk of a false positive or false negative result [which occurs when the value(s) of specific biomarker(s) does not reflect a positive change in the clinical endpoint(s)]. During the validation process, the assay that is used must be highly reliable and reproducible. As far as the demonstration of the predictive value of the candidate as a surrogate endpoint for the clinical outcome is concerned, regulatory guidance does not specify which methodology should be used in validating biomarkers as surrogate endpoints. It is well recognized that developing a single biomarker as a surrogate endpoint can become rather cumbersome for a pharmaceutical sponsor (Lesko, 2007). To complicate matters further, a biomarker may become a surrogate endpoint for efficacy but not for toxicity. Indeed, there are few biomarkers of toxic effects (such as the QTc prolongation) that predict *torsade de pointe*, or the increase in aminotranspherases predicting liver failure.

Biomarkers may also be misleading in areas where they may result in a short-term beneficial effect but a long-term deleterious effect. As a consequence, the benefit/risk ratio can rarely be evaluated based on a surrogate marker, hence the use of biomarkers as surrogate endpoints only in those areas with critical unmet medical needs.

In the process for biomarker validation, the following properties should be evaluated: (1) feasibility of a surrogate marker in predicting the clinical outcome, and (2) statistical relationship between a biomarker and the clinical outcome. This should first be demonstrated by the natural history of the disease, then by adequate and well-controlled clinical trials that estimate the clinical benefit by changing the specific surrogate endpoint. It should be noted that during the biomarker validation process, it is insufficient to show only that the biomarker correlates with the clinical endpoint; it is also necessary to demonstrate that the effect on the surrogate endpoint interferes with the treatment effect on the clinical endpoint. In rare cases, a biomarker is elevated to the status of surrogate endpoint based solely on the results obtained from one drug. A metaanalysis of multiple clinical trials with different drugs and different stages of disease may be required to determine the consistency of effects and strengthen the evidence that a change in the biomarker level resulted in an effect on the clinical outcome.

With the FDA Modernization Act of 1997, the U.S. Food and Drug Administration (FDA) has gained a legal basis for using surrogate endpoints in ordinary and accelerated drug approvals. Indeed, the FDA was given explicit authority to approve drugs for the "treatment of a serious or life-threatening condition … upon a determination that a product has an effect on a clinical endpoint or on a surrogate endpoint that is reasonably likely to predict clinical benefit," leading to market access of new drugs and drug

products. The standards for linking a biomarker to a clinical outcome are higher for ordinary approvals than for accelerated approvals. This difference is based on consideration of many factors, including the degree of scientific evidence needed to support biomarker surrogacy, public health needs, relative risk/benefit ratio, and the availability of alternative treatments. For ordinary approvals there are relatively few approved surrogate endpoints, such as "lower cholesterol and triglycerides" for coronary artery disease, "lower arterial blood pressure" for stroke, heart attacks and heart failure, "increase cardiac output for acute heart failure," "reduce HIV-RNA load and enhance CD4+ cells" for AIDS, "lower glycosilated hemoglobin" for diabetes, and "reduced tumor size" in solid tumors.

Oncology is an interesting example of this practice. In oncology, survival is the ultimate clinical outcome. However, approvals in the United States in the field of oncology from 1990 to 2002 highlight that tumor response was the approval basis in 26 of 57 regular approvals, supported by relief of tumor-specific symptoms in 9 of these 26 regular approvals (Table 1). Relief of tumor-specific symptoms provided critical support for approval in 13 of 57 regular approvals; approvals were based on tumor response in 12 of 14 accelerated approvals.

In Europe, regulatory awareness is increasing and some initiatives are ongoing, such as the EMEA/CHMP Biomarkers Workshop, which was held in 2006. However, while biomarker development is encouraged during early-stage development, there is significant hesitancy in accepting biomarkers as surrogate endpoints for drug approval. As an example, a review of the

**TABLE 1 Summary of Endpoints for Regular Approval of Oncology Drug Marketing Applications, January 1, 1990 to November 1, 2002**

| Parameter[a] | Endpoint |
|---|---|
| Total | 57 |
| Survival | 18 |
| RR | 26 |
| RR alone | 10 |
| RR + decreased tumor-specific symptoms | 9 |
| RR + TTP | 7 |
| Decreased tumor-specific symptoms | 4 |
| DFS | 2 |
| TTP | 1 |
| Recurrence of malignant pleural effusion | 2 |
| Occurrence of breast cancer | 2 |
| Decreased impairment of creatinine clearance | 1 |
| Decreased xerostomia | 1 |

*Source:* Johnson et al. (2003).

[a]RR, response rate; TTP, time to progression; DFS, disease-free survival.

European guidelines reveals that outside the oncology and the muscoloskeletal field, biomarkers in general and biomarker imaging in particular, are not considered surrogate endpoints. A list of imaging biomarkers endpoints accepted as primary endpoints for regulatory submissions and those suggested as endpoints in early development are listed in Table 2.

Overall, it is very complex to validate biomarkers to become surrogate endpoints, but the value of developing a biomarker resides in the information that can be obtained during such development. Among these benefits are the possibility of defining the population who may benefit from the drug candidate, the screening of patients for adverse events, the possibility to enrich the population for proof-of-concept studies, the possibility to stratify patients, the selection of doses for pivotal trials, and the potential for dose adjustments at patient level.

**TABLE 2 Imaging Biomarkers: Review of Accepted Primary or Secondary Endpoints in the CHMP Guidelines**

| Condition | Imaging Accepted as a Primary Endpoint | Imaging Accepted as a Secondary Endpoint or Early in Development |
|---|---|---|
| Fungal infections | Relevant imaging is part of the clinical outcome | |
| Cancer | — | Imaging and functional imaging |
| Osteoporosis | X-ray | DEXA for BMD |
| Juvenile idiopatic arthritis | X-ray | |
| Psoriatic arthritis | X-ray | MRI and ultrasound (US) |
| Osteoarthritis X-ray | X-ray | |
| Ankilosing spondilitis | — | X-ray/MRI/DEXA/US |
| Crohn disease | — | Endoscopy |
| Profilaxis of thromboembolic disease | Ultrasound (detection of DVT) and venography | |
| Peripheral artherial obstructive disease | — | Hemodynamic measurements |
| Ischemic stroke | — | Neuroimaging techniques |
| Treatment of venous trombotic disease | Ultrasound (detection of DVT) and venography; angiography for pulmonary embolism | |
| Incontinence | — | Urodynamic studies or x-ray videography |
| Anxiety | — | Functional neuroimaging |
| Multiple sclerosis | — | MRI |
| Panic disorders | — | Neuroimaging |
| Acute stroke | — | MRI |

## REASONS FOR FAILURE OF A BIOMARKER

There are several examples of biomarker failure. One of the most recent examples is the approval of gefitinib for non-small cell lung cancer (NSCLC). Gefitinib was originally approved based on tumor response as opposed to overall survival. In a postmarketing survival trial that included approximately 1700 patients, there was no benefit over placebo on overall survival. Other examples include bone mineral density (BMD) in osteoporosis for fluoride treatment. However, the Cardiac Arrhytmia Suppression Trial (CAST) provides the best known example of a failure of a biomarker. This study was based on the hypothesis (supported by statistical association and a plausible biological mechanism) that suppression of arrythmias would prevent sudden death after myocardial infarction. The study demonstrated a worse outcome on mortality for patients receiving active treatments compared to those receiving placebo. The theoretical background for biomarker failure is given by Frank and Hargreaves (2003). The reasons biomarkers can lead to erroneous conclusions have been divided into the following five categories (Figure 3):

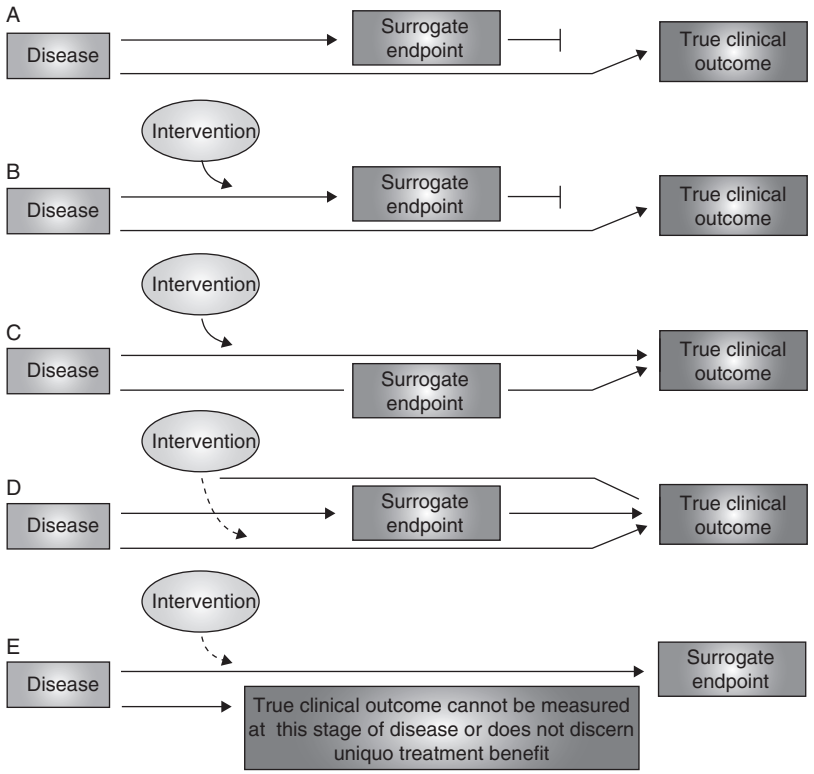


**Figure 3** Reasons why biomarkers have failed to become surrogate endpoints.

1. Changes in the biomarker reflect the effect of treatment, but these changes are irrelevant to the pathophysiology of the disease indicated (false positive).
2. Changes in the biomarker reflect an effect of treatment on an element of the pathophysiology, but this element is clinically irrelevent (false positive).
3. Changes in the biomarker reflect clinically relevant changes in pathophysiology but do not capture the mechanistic effect of the treatment (false negative).
4. Changes in the biomarker reflect one effect of the treatment, but there are other, more relevent effects on outcome that are not captured (false negative or positive).
5. The biomarker may not correlate well with classical clinical assessors because the biomarker is more sensitive or the classical assessor is irrelevant to a subset of the patient population, a novel mechanism, or a new indication.

It is important to consider this theoretical framework while developing a biomarker or considering biomarkers in decision making. Using more than one biomarker and reviewing the consistency of the data across biomarkers may decrease the risk of making an incorrect conclusion. The classical example of this comes from the osteoporosis field, where BMD alone may provide misleading results, while BMD in combination with other biomarkers of bone metabolism (such as osteocalcin and collagen cross-links) provide a much more robust basis for decision making.

## IMAGING AS A BIOMARKER TO OPTIMIZE DRUG DEVELOPMENT

Imaging (e.g., x-ray) has been used in clinical practice for over a century, but mainly for diagnostic purposes. The value of imaging as a biomarker has been recognized only recently. In clinical development, the use of imaging must take into consideration the scope of clinical studies, which are usually multicenter and multinational (at least in confirmatory development), and there are restraints such as cost, effort and resources as well as the need to standardize techniques across centers to maximize the signal-to-noise ratio. The most common techniques used in clinical development include x-ray imaging, digitalized imaging [including DEXA (dual energy x-ray absorptiometry)], computed tomography (CT) scans, nuclear imaging such as positron-emission tomography (PET) and single-photon-emission CT (SPECT), ultrasound, magnetic resonance imaging (MRI), spectrometry MR, and functional MR. Some of these tools are available only in specialized centers and are therefore adapted mainly for small studies during early development.

Imaging biomarkers can be used, for example, for the assessment of bioactivity (not only through change in anatomical shape but also through change in functional status), for the evaluation of the disposition of drugs, for the measurements of tissue concentrations of drugs, to characterize the number of receptors, for the binding efficiency and the receptor occupancy, and as a prognostic indicator as well as for assessment of molecular specificity.

## IMAGING AS A MARKER OF BIOLOGICAL ACTIVITY

Imaging techniques are used to evaluate the biological activity of a drug candidate by performing pre- and posttreatment measurements (and in many instances, performing measurements during treatment). These include:

- *Oncology:* CT scan or MRI for the measurement of the size of solid tumors
- *Neurology:* MRI for the measurement of multiple sclerosis lesions and for the evaluation and quantification of brain atrophy in Alzheimer disease (e.g., brain boundary shift integral or ROI-based MRI)
- *Musculoskeletal diseases:* x-ray and DEXA to evaluate vertebral fractures and BMD in osteoporosis; x-ray and MRI for the evaluation of erosions and joint space in rheumatoid arthritis and psoriatic arthritis; x-ray and MRI for the evaluation of joint space and cartilage volume in osteoarthritis; and the same techniques to evaluate spinal changes in spondyloarthritis

In addition to the use of imaging to capture anatomical changes, imaging can be used to perform a functional evaluation of a tissue/organ before and after treatment. Classical examples include:

- *Oncology:* the evaluation of tumor metabolism with fluorodeoxyglucose (FDG)-PET. This may be helpful as an early predictor of a later anatomical response. Indeed, the reduction in tumor size with imatinib was preceded by decreased tumor glucose intake by a median of 7 weeks (Stroobants et al., 2003). Another example is the measurement of tumor blood flow and blood volume with CT: after bevacizumab (VEGF- Ab) treatment a decreased blood flow and volume in colorectal cancer was observed as early as 12 days after initiation of treatment (Miller et al., 2005).
- *Neurology:* FDG-PET was used in Alzheimer disease (AD) to evaluate regional cerebral metabolic rate and SPECT was used to evaluate blood flow in AD.

In multiple sclerosis, PET and SPECT have shown a reduction in cerebral metabolism and blood flow. It should be noted, however, that these approaches

have not been tested and qualified to measure the potential effect of a therapeutic intervention (Bakshi et al., 2005).

## IMAGING TO EVALUATE DISPOSITION OF DRUGS

An interesting use of imaging is to estimate the concentration of a drug in different tissues or organs or to evaluate the disposition of a drug into nonaccessible compartments. An example of the first use is the antifungal drug fluconazole. By using PET imaging and [$^{18}$F]fluconazole, an in vivo pharmacokinetic profile was characterized. Concentrations were measured over time in multiple organs, such as the brain, muscle, heart, and bowel. Based on these data, it was possible to test fluconazole only for infections in organs in which adequate concentrations (to exert antifungal effect) were achieved (Pien et al., 2005).

Imaging can also be used to characterize the number of receptors, the binding efficiency, and the receptor occupancy. An interesting example is aprepitant, which is an NK1 receptor antagonist (used for the prevention of nausea and vomiting with chemotherapy). By using an $^{18}$F ligand with high affinity and specificity to NK1 receptor, PET was used to image the displacement of this ligand by aprepitant given systemically, therefore demonstrating the ability of aprepitant to cross the blood–brain barrier.

Imaging biomarkers as prognostic indicators and assessment of molecular specificity As mentioned above, FDG-PET can be used as an early indicator of bioactivity and therefore can help separate responders versus nonresponders at earlier time points compared to anatomical imaging. The case of imatimib is instructive in this sense: FDG-PET identified the biological response a median of 7 weeks earlier than did CT scans. In addition, all patients with metabolic responses were clinical responders later during the trial (Stroobants et al., 2003). Similar examples are available with chemotherapy in breast cancer (Schelling et al., 2000), in NSCLC (Weber et al., 2003), and in gastroesophageal cancer (Weber et al., 2001).

Another interesting development for use of imaging is its use to assess molecular specificity. In oncology, spectrometry MR can evaluate the molecular content of tissues and assist diagnosis. It is also used extensively in brain, breast, and prostate cancers. In neurology the brain β amyloid content can be measured in AD patients using PET technique. (Sorensen, 2006).

## DISCUSSION AND CONCLUSIONS

The need for the pharmaceutical industry to increase its productivity and, in particular, the need to concentrate the available limited resources to the most promising drug candidates lead to the current emphasis on development and use of biomarkers. Since 2001 and as a result of the Biomarkers Definition

Working Group, there is clarity on the definitions concerning biomarkers, surrogate markers, and clinical endpoints. Many classifications exist to guide the development of new biomarkers. The evaluation/qualification and eventually the validation of biomarkers is a long and complex process which needs to be supported by many stakeholders, such as the pharmaceutical industry, regulatory authorities, and academic centers. None of these entities alone can succeed easily in such an endeavor. A few biomarkers have reached the status of surrogate endpoint and can be used for registration purposes, notably in the areas of cardiovascular disease, diabetes, rheumatology, and oncology. Imaging can be particularly useful in the early stages of development in the evaluation of bioactivity and drug disposition, which are among the many possibilities. In the evaluation of biomarkers and in the decision-making process based on biomarkers, investigators and scientists should always consider the possibility that the biomarker may fail to predict the clinical outcome, therefore leading to potential false positives and false negatives. It is reassuring, however, that even when the biomarker fails, significant learning is achieved, which eventually will benefit not only the scientific community in general but the pharmaceutical industry as well.

## REFERENCES

Bakshi R, Minagar A, Jaisani Z, Wolinsky JS (2005). Imaging of multiple sclerosis: role in neurotherapeutics. *NeuroRx*, 2:277–303.

Biomarkers Definitions Working Group (2001). Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. *Clin Pharmacol Ther*, 69:89–95.

Danhof M, Alvan G, Dahl SG, Kuhlmann J, Paintaud G (2005). Mechanism-based pharmacokinetic–pharmacodynamic modeling: a new classification of biomarkers. *Pharm Res*, 22:1432–1437.

DiMasi JA, Hansen RW, Grabowskal HG (2003). The price of innovation: new estimates of drug development costs. *J Health Econ*, 22:151–185.

European Medicines Agency (2006). *Report on the EMEA/CHMP Biomarkers Workshop*. EMEA/522496/2006.

FDA Center for Drug Evaluation and Research (2003). Guidance for Industry: Exposure-response relationships—study design, data analysis, and regulatory applications. http://www.fda.gov/cder/guidance/index.htm.

FDA (2004). The "Critical Path Initiative"–innovation–stagnation: challenge and opportunity on the critical path to new medical products. http://www.fda.gov/oc/initiatives/criticalpath/whitepaper.html.

Frank R, Hargreaves R (2003). Clinical biomarkers in drug discovery and development. *Nat Rev, Drug Discov*, 2:566–580.

Johnson JR, Williams G, Pazdur R (2003). End points and United States Food and Drug Administration approval of oncology drugs. *J Clin Oncol*, 21:1404–1411.

Lesko LJ, Atkinson AJ (2001). Use of biomarkers and surrogate endpoints in drug development and regulatory decisionmaking: criteria, validation, strategies. *J Annu Rev Pharmacol Toxicol*, 41:347–366.

Lesko LJ (2007). Paving the critical path: how can clinical pharmacology help achieve the vision? *Clin Pharmacol Ther*, 81:170–177.

Miller JC, Pien HH, Sahani D, Sorensen AG, Thrall TH (2005). Imaging angiogenesis: applications and potential for drug development. *J Natl Cancer Inst*, 97:172–187.

Pien HH, Fischman AJ, Thrall JH, Sorensen AG (2005). Using imaging biomarkers to accelerate drug development and clinical trials. *Drug Discov Today*, 10:259–266.

Schelling M, Avril N, Nährig J, et al. (2000). Positron emission tomography using $^{18}$F fluorodeoxyglucose for monitoring primary chemotherapy in breast cancer. *J Clin Oncol*, 18:1689–1695.

Sorensen AG (2006). Magnetic resonance as a cancer imaging biomarker. *J Clin Oncol*, 24:3274–3281.

Stroobants S, Goeminne J, Seegers M, et al. (2003). $^{18}$FDG-positron emission tomography for early prediction of response in advanced soft tissue sarcoma treated with imatinib mesylate (Glivec). *Eur J Cancer*, 39:2012–2020.

Weber WA, Ott K, Becker K, et al. (2001). Prediction of response to preoperative chemotherapy in adenocarcinomas of the esophagogastric junction by metabolic imaging. *J Clin Oncol*, 19:3058–3065.

Weber WA, Petersen V, Schmidt B, et al. (2003). Positron emission tomography in non-small-cell lung cancer: prediction of response to chemotherapy by quantitative assessment of glucose use. *J Clin Oncol*, 21:2651–2657.

# NANOTECHNOLOGY-BASED BIOMARKER DETECTION

JOSHUA REINEKE, PH.D.
*Wayne State University, Detroit, Michigan*

## ADVANTAGES OF NANOTECHNOLOGY TO BIOMARKERS

Nanotechnology refers to the fabrication, manipulation, use, and study of phenomena of materials with at least one dimension in the nanoscale (<100 nm). The recent revolution in nanotechnology will benefit a large diversity of applications, including many consumer products, engineering, information technology, and medicine. When applied to biomarker science, nanotechnology will influence how we utilize current biomarkers and enable the discovery and development of many more new biomarkers.

Major challenges facing the detection of biomarkers that are being addressed with nanotechnology approaches include multiplexing ability, specificity, rapid or real-time detection, label-free detection, determination of early therapy response, small sample sizes, and portability. Additionally, nanodevices are likely to be suitable for reliable scale-up and massive parallelization at low cost by utilizing processes similar to those employed in the electronics industry [1,2]. A further potential asset of nanotechnology is the detection of biomarker density and distribution, particularly in cases where biomarker concentration and not the biomarker itself is specific to a pathologic condition. Recent developments in genetics and proteomics have produced a large number of biomarkers [3–5], increasing dramatically the potential applications

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

of biomarkers. However, for the potential of these recent developments to be realized, better biomarker detection systems must be engineered, and since multiple biomarkers are of particular interest in complex diseases such as cancer where single biomarker detection is insufficient [6], the detection systems must have multiplexing ability.

In this chapter we review nanoparticles and nanodevices currently being developed for the enhanced detection of biomarkers and focus on the diversity of applications to biomarker research. Specific methods of detection system fabrication are not discussed despite many exciting recent discoveries, but descriptions may be found in the relevant referenced material. Following this review is a discussion of nanotoxicology issues and other challenges that nanotechnology will present to the biomarker field.

## NANOPARTICLES

Nanoparticles can be fabricated and tailored to specific morphologies and chemistries, to incorporate various imaging agents, and to target specific ligands, allowing virtually unlimited application. This versatility has lead to intensified research into nanoparticle-based detection systems for biomarkers. Nanoparticles may also be applied to biomarkers for concentrating, amplifying, and protecting the biomarker both in vivo and ex vivo [7]. With the development of quantum dot nanoparticles, the ability to detect biomarkers via fluorescence has great potential. Additionally, biomarker science will benefit from magnetic resonance imaging (MRI) agents based on nanoparticles. These and a few other novel nanoparticle-based biomarker detection systems are discussed below.

### Quantum Dots

Several characteristics distinguish colloidal semiconductor quantum dots (qdots) from commonly used fluorophores. Qdots are single crystals 2 to 10 nm in diameter whose size and shape can be controlled precisely by the duration, temperature, and ligand molecules used during their synthesis, yielding qdots with composition- and size-dependent fluorescent emission (Figure 1) [8]. Absorption of a photon above the semiconductor band gap results in excitation. For nanocrystals smaller than the Bohr excitation radius, energy levels are quantized in an effect called *quantum confinement* (hence the name *quantum dots*). Radiative recombination of an excitiation results in the emission of a photon in a narrow symmetric energy band. A complete description of qdots can be found in a paper by Alivisatos [8]. The narrow emission wavelengths of qdots, tunability based on composition and size and long lifetime, make them far superior to commonly used fluorophores.

Qdots possess a large number of advantageous qualities over traditional fluorescent probes, including photostability and increased intensity [8,9],

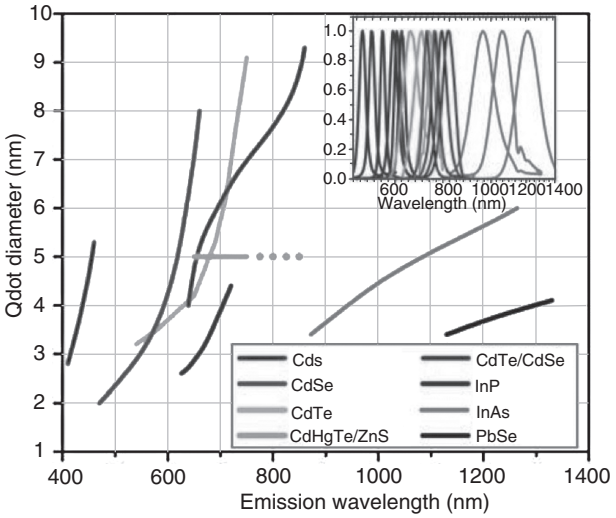


**Figure 1** Emission maxima and sizes of qdots of different composition. Qdots can be synthesized from various types of semiconductor materials characterized by different bulk bandgap energies. The curves represent experimental data from the literature on the dependence of peak emission wavelength on qdot diameter. The range of emission wavelength is 400 to 1350 nm, with size varying from 2 to 9.5 nm. All spectra are typically around 30 to 50 nm (full width at half maximum). Inset: Representative emission spectra for some materials. (From ref. 9, with permission.) (*See insert for color reproduction of the figure.*)

lower required excitation power [10], increased binding specificity [11], less binding interference and steric-hendrance [10,12–18], high surface area for multifunctionality, and the ability to escape reticuloendothelial system (RES) clearance [11]. Qdots can be surface modified to meet a diverse set of biomarker applications, as illustrated in Figure 2. The ability to conjugate qdots to antibodies specific to a single biomarker and to select qdots of a specific emission wavelength allows simultaneous detection and accurate quantification of multiple targeted biomarkers [20,21]. Figure 3 illustrates an example of in vivo imaging of qdots targeted to a tumor and imaging of three qdots with differing emission wavelengths in vivo. Several areas of untapped qdot potential are their use as customizable donor pairs for fluorescence resonance energy transfer (FRET) experiments [16] and deep tissue visualization with multiphoton imaging [10]. Recently, qdots have been developed for imaging in the near-infrared (IR) range to allow in vivo imaging without interferance of autofluorescence [9,19].

Despite the many advantages of qdots, their toxicity is problematic for in vivo and living cell applications, particularly if ultraviolet light is to be used [22]. Surface coating is often utilized to protect from toxicity and has yielded

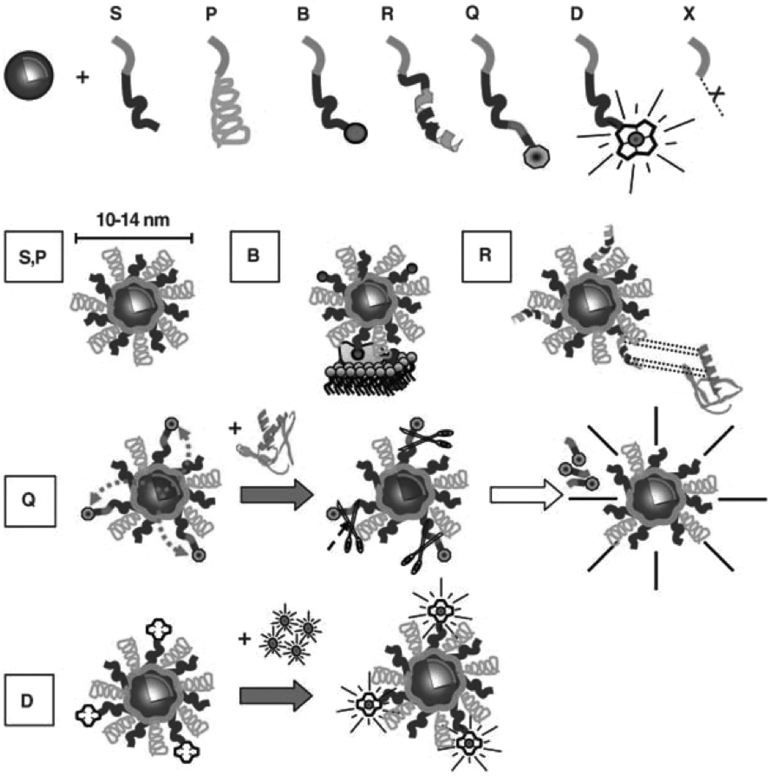


**Figure 2** Qdot peptide toolkit. The light blue segment contains cysteines and hydrophobic amino acids assuring binding to the qdot and is common to all peptides. S, solubilization sequence; P, PEG; B, boitin; R, peptide recognition sequence; Q, quencher; D, DOTA; X, any unspecified peptide-encoded function. Qdot solubilization is obtained by a mixture of S and P. Qdots can be targeted with B, R, or other chemical moieties. Qdot fluorescence can be turned on or off by attaching a Q via a cleavable peptide link. In the presence of the appropriate enzyme, the quencher is separated from the qdot, restoring the photoluminescence and reporting on the enzyme activity. For simultaneous PET and fluorescence imaging, qdots can be rendered radioactive by D chelation of radionuclides; for simultaneous MRI and fluorescence imaging, qdots can be rendered radioactive by D chelation of nuclear spin labels. (From ref. 9, with permission.) (*See insert for color reproduction of the figure.*)

many successful results. Qdots encapsulated with an ABC triblock copolymer not only resulted in less cytotoxicity, but advantageously prevented aggregation and fluorescence loss after in vivo exposure (a major and not fully understood barrier to in vivo imaging with qdots [11,23]). These coated qdots were successful in both passive and active tumor targeting, allowing imaging of multiple biomarkers [21].

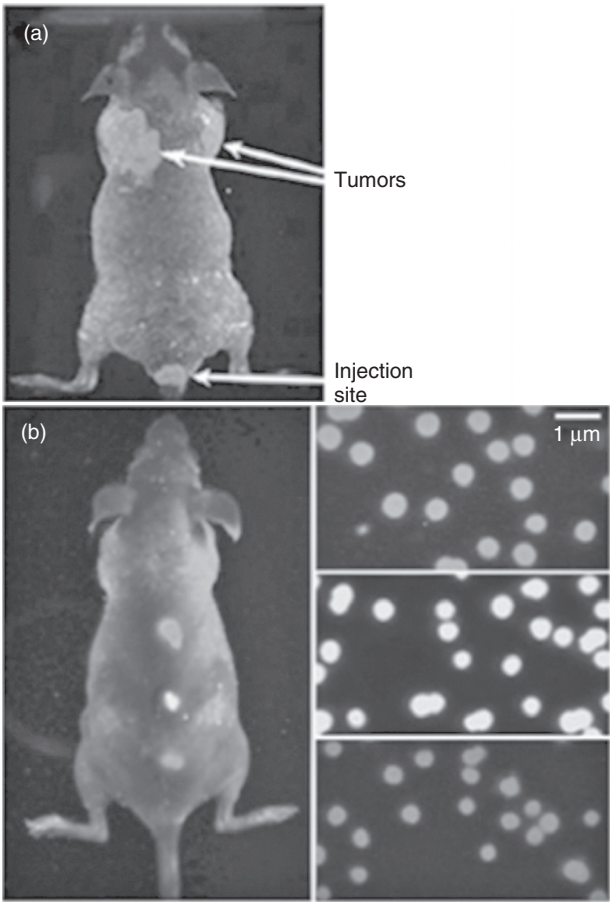


**Figure 3** In vivo imaging of qdots. (a) Spectrally resolved image of a mouse bearing C4-2 human prostate tumors following injection of qdots functionalized with antibodies for prostate-specific membrane antigen. (b) Images on the right show qdots emitting green, yellow, or red light. The image on the left illustrates the in vivo imaging of the multicolor qdots at three injection sites. (From ref. 21, with permission.) (*See insert for color reproduction of the figure.*)

The utility of qdots for in vivo imaging in live animals was demonstrated further in their use for nonspecific uptake studies and lymph node mapping [19,24]. Additionally, qdots have been used for the in vivo imaging of blood vessels with increased contrast [10]. Targeted in vivo imaging of prostate cancer was achieved following intravaneous administration of qdots functionalized with prostate-specific membrane antigen [21]. Other cancer biomarkers have been targeted and imaged by qdots, including *her2* [13] and *erbB* family receptor signaling events [25]. Recent developments in near-IR qdots lend

great potential in the use of qdots for imaging of deep tissue tumors and for applications of real-time imaging to aid in surgical procedures [9]. Qdots' largest potential to the future of biomarker science may come from the ability to give qdots an on–off switch that responds to chemical cues. This was illustrated by functionalizing qdots with a fluorescence quenching molecule that could be cleaved in the presence of a chemical species or enzyme [16]. Continued research and development of qdots should produce versatile in vivo and in vitro multiplexed biomarker detection systems.

## Nanoparticle-Based MRI Agents

There are many nanoparticle formulations utilized in the clinic and in research for the enhancement of MRI contrast, such as gadolinium nanoparticles [26] and iron oxide nanoparticles [27–32]. Nanoparticles are advantageous to MRI applications, due to their increased circulation times and high efficacy-to-safety ratio relative to iodinated imaging agents [33]. An intense area of biomarker and drug discovery research is in the detection of the progression, regression, and dormancy of tumors mediated by angiogenesis. A number of investigators have utilized integrin-targeted nanoparticles as enhanced MRI agents to visualize angiogenesis [29,34–36] and have even shown the detection of signals from very low picomolar concentrations of epitopes [37]. These studies indicate great promise for the future of clinical applications, yet present a great need for modeling software to fully interpret the biomarker information gained. As an alternative to targeted imaging agents, "smart" MRI nanoparticles have been designed to be taken up by all cells, but have increased contrast only after a specific enzymatic reaction, inherent to a particular pathology, exposes gadolinium atoms to free water enhancing image contrast [38].

Nanoparticles may also be used for enhanced interoperative imaging. The use of dextran-coated paramagnetic iron oxide nanoparticles used as MRI contrast agents during the surgical treatments of brain tumors significantly improved interoperative permanence of imaging, inflammatory targeting, and low-level detectability over conventional MRI contrast agents [39]. The use of nanoparticles as enhanced MRI agents is also advantageous for detecting biomarkers in in vitro cell culture assays. As an example, telomerase activity (a biomarker of replicative potential [40]) was detected by MRI with the use of biologically smart nanoparticles. These nanoparticles switch their magnetic state when annealed with telomerase-sythesized TTAGGG sequences [41].

## Novel Nanoparticle-Based Biomarker Detection Systems

Quantum dot–based detection systems and nanoparticle-based MRI agents hold great potential in biomarker research. However, novel nanoparticle

systems that combine some of the assets or address some of the limitations of these systems will have the greatest utility in the future advancement of biomarker research.

A major limitation of qdots is their cytotoxicity [9,22,42]. Cytotoxicity is often circumvented by surface modification of the qdots despite the fact that surface modifications result in adverse changes in imaging properties. Nanodiamonds have been proposed as a noncytotoxic alternative to qdots, due to their long-term stability without photobleaching or blinking. Nanodiamonds can be surface modified without changes in fluorescence properties since the signal originates solely from internal defects. Preliminary studies show enhanced imaging properties and the ability to track single particles over time [43]. Nanodiamonds are already mass-produced for the surface finishing industry, potentiating reproducible scale-up procedures with ease. However, the range and specificity of imaging wavelengths need to be further investigated.

Photostable nanoparticles have also been produced via the encapsulation of metal-organic luminophores within silica [44]. This extends the range and versatility of luminphores for biomarkers beyond what has been established with qdots. The silica surface of these nanoparticles allows easy covalent surface attachments for targeted imaging. Antibody-bound silica nanoparticles with an encapsulated luminophore was able to bind selectively, and therefore identify, leukemia cells [44].

The fabrication of novel polyacrylamide "nano-PEBBLE" sensors by Xu et al. [45] allows biomarker detection within single cells. The nano-PEBBLE sensors described are nontoxic polyacrylamide nanoparticles containing glucose oxidase and oxygen-sensitive fluorophores that are readily taken up by cells. Intracellular glucose concentration could be determined in real time by fluorophore intensity as a response to oxygen groups produced by the glucose oxidase. The major advantages over free sensing dyes is the lack of interference from nonspecific protein binding, protection from cytotoxic dyes (allowing a greater range of dyes to be used in biomarker applications), and the ability to allow detection of biomarkers for which no specific dyes exist (due to the synergistic effect of material selected for encapsulation). Nano-PEBBLE sensors have been developed to measure real-time intracellular concentrations of calcium, potassium, oxygen, and pH [46–49].

Additional nanoparticle systems that are relevant to biomarker detection and are not covered here include liquid perfluorocarbon MRI nanoparticles targeted to thrombi [50], neovascular integrins [35] and tumors [35], low-density lipid nanoparticles [51] and lipid- and polymer-shelled nanoparticles [52] as enhanced ultrasound agents, "bio-barcode" DNA [53] and prostate-specific antigen [54] detection with oligonuleotide-modified gold nanoparticles, silica nanoparticles containing pH-sensitive fluorescent dye for rapid, reversible pH detection [55], antibody-bound silica-coated magnetic nanoparticles for cell sorting [56], silver dendrimer nanoparticles for specific, nontoxic cell staining [57], and targeted nanoscale Raman probes [20]. Additionally, in

vivo stem cell tracking was performed by first pretreating cells with magnetic nanoparticles [58,59].

### Nanoparticles for Enhanced Biomarker Detection Devices

Most of the proposed applications of nanoparticles to biomarker science involve the enhanced imaging of pathological tissue morphologies or the specific binding to biomarkers yielding information on biomarker location, concentration, and distribution as discussed above. However, nanoparticles are readily being used to enhance current in vitro biomarker detection assays as well. Nanoparticles are being used in clinical and research settings for enhancement of cell sorting assays, immunostaining assays, proteomic and genetic microarrays, and surface plasmon resonance assays [60,61]. Nanoparticles have even been proposed as selective protein harvesting agents for serum proteomics of low-molecular-weight proteolytic fragments found in trace amounts in various cancers [62,63].

## NANODEVICES

The greatest utility of the biomarker detection nanodevices discussed below is the ability to have real-time, selective, multiplexed, and tag-free detection of biomolecules in small sample volumes. The nanoscale of these devices also increases the portability and reduces costs associated with larger reagent volumes used in conventional assays.

### Nanowires

Nanowires are nanoscale sensing wires that can be functionalized to bind proteins of interest and have great multiplexing potential [64]. When a biomolecule of interest binds a nanowire, there is a drop in its conductance. The ability of nanowires to act like nanoscale field-effect biotransistors allows their use as real-time sensors to multiple binding events, as illustrated in Figure 4. Also shown in Figure 4 is a scanning electron image of a single silicone nanowire between two electrodes.

Nanowires made from silicon are advantageous due to their ability to be easily tuned and surface modified for specific biomolecule detection. Hahm and Lieber [65] utilized silicon nanowires for the specific detection of a DNA mutation site for cystic fibrosis transmembrane receptor. The nanowire construct allowed real-time detection and was ultrasensitive, detecting the mutation in fentomole concentrations. Hahm and Lieber also found high device-to-device reproducibility.

Simultaneous multiplexed detection of four cancer biomarkers in fentomolar concentrations from undiluted serum samples was achieved by Zheng et al. [66]. Real-time detection results for a multiplexed silicon nanowire

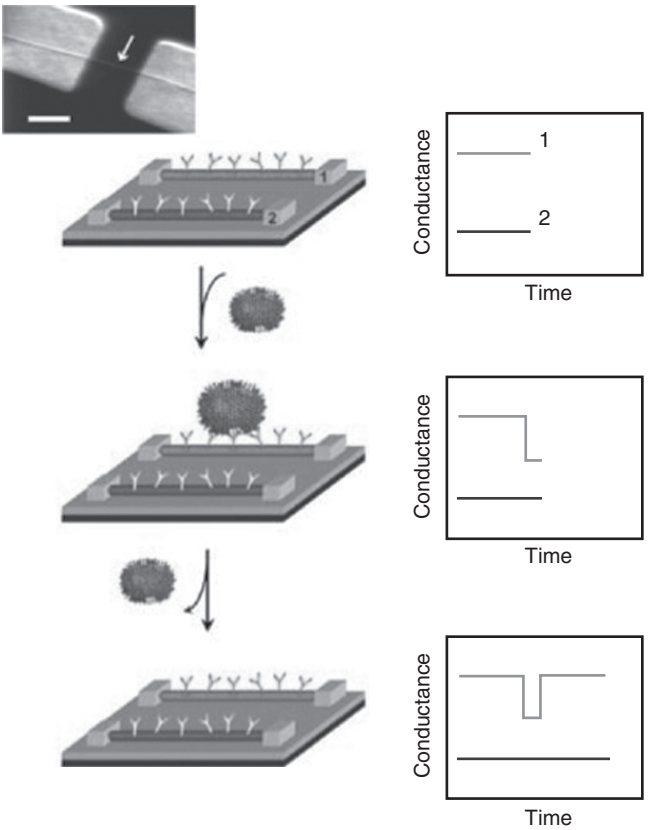


**Figure 4** Nanowire detection of a single-target biomolecule using conductive-based measurements. A biomolecule is immobilized selectively by antibody interaction, causing a change in conductance, recorded on the right. Inset: SEM image of a single silicon nanowire (scale bar = 500 nm). (From refs. 100 and 67, with permission.) (*See insert for color reproduction of the figure.*)

device designed to detect three biomarkers is shown in Figure 5. The sensitivity of the nanowire device was demonstrated further by the same group when they detected binding, activity, and inhibition of telomerase from unamplified extracts of as few as 10 cells. Detection is as sensitive as current telomere repeat application protocol (TRAP) assays but does not require the use of polymerase chain reaction (PCR) amplification and labeling.

Silicon nanowires were also used by Wang et al. [67] to measure the ATP binding (and nonbinding) to an immobilized leukemia biomarker, Abl. Although this work also demonstrates the ability of nanowires for biomarker detection, the group went a step further in utilizing the device to measure the biomarker response to known inhibitors, lending potential to drug discovery

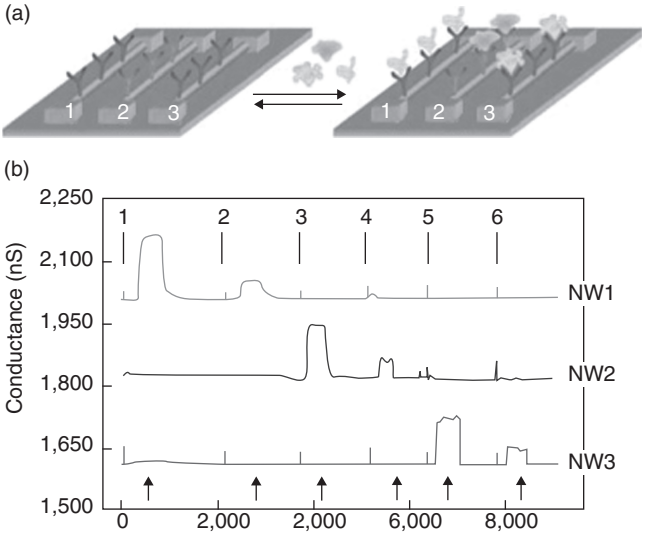


**Figure 5** Multiplexed detection of cancer marker proteins. (a) Multiplexed protein detection by three silicon-nanowire devices in an array. Devices 1, 2, and 3 are fabricated from similar nanowires and then differentiated with distinct mouse antibody receptors specific to three different cancer markers. (b) Conductance versus time data recorded for the simultaneous detection of PSA, CEA, and mucin-1 on p-type silicon-nanowire array in which NW1, NW2, and NW3 were functionalized with mouse antibodies for PSA, CEA, and mucin-1, respectively. The solutions were delivered to the nanowire array sequentially as follows: (1) 0.9 ng/mL PSA, (2) 1.4 pg/mL PSA, (3) 0.2 ng/mL CEA, (4) 2 pg/mL CEA, (5) 0.5 ng/mL mucin-1, (6) 5 pg/mL mucin-1. Buffer solutions were injected following each protein solution at points indicated by black arrows. (From ref. 66, with permission.) (*See insert for color reproduction of the figure.*)

and drug efficacy research. Detection in the nanomolar range and in a concentration-dependent manner (allowing quantification) make this sensitive method tenable for drug discovery and efficacy studies for any of the tyrosine kinases responsible for many cancers and diseases. Silicon nanowires have been developed for a number of other applications not discussed here, including pH sensing [64] and virus detection [68].

## Nanocantilevers

Nanocantilevers are a promising new approach to multimolecular sensing. Nanocantilever arrays, illustrated in Figure 6, consist of many nano-diving boardlike beams that can have antibodies attached covalently to their surface. When a biomolecule of interest binds, the beam bends and the deflection can be detected by either laser light observation or changes in resonant-vibration

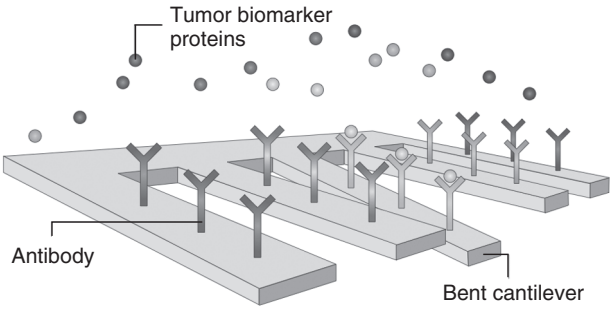


**Figure 6** Nanocantilever array. The biomarker proteins are affinity bound to the cantilevers and cause them to deflect. The deflections can be observed directly with lasers. Alternatively, the shit in resonant frequencies caused by the binding can be detected electronically. (From ref. 1, with permission.) (*See insert for color reproduction of the figure.*)

frequency [69–71]. The utility of cantilever detection has been demonstrated by multiplexed DNA assays to detect *BRCA1* mutations, indicative of early breast cancer onset [72]. Additionally, nanocantilevers were able to detect and quantitate prostate-specific antigen in clinically significant concentrations [70].

A nanomechanical cantilever array was used by Arntz et al. [2] for the multiplexed detection of the cardiac biomarkers creatin kinase and myoglobin. Both biomarkers could be detected at micromolar concentrations from plasma samples. Further development of this system is being investigated for the early and rapid diagnosis of myocardial infarction.

Nanomechanical cantilever detection systems are not limited to the detection of biomarkers due to hybridization and antibody interactions. Environmentally sensitive polymers coated onto nanocantilevers can also produce detectable cantilever stresses in response to sample chemistries. For instance, pH-sensitive poly(methacrylic acid) coated onto nanocantilevers were used as highly sensitive, small-volume pH sensors [73]. As new environmentally responsive polymers are developed, this application may grow rapidly. Similarly, analyte vapors in a gas phase resulted in a changed surface stress of nanocantilevers in a device termed an *artificial nose* [74].

## Nanoarrays

Currently, microarrays are highly utilized in research and the clinic for molecular diagnostics, genotyping, and biomarker-guided therapy. Development of nanoarrays will advantageously advance current uses by allowing higher degrees of multiplexing, higher specificity, more portability, and reduced cost associated with sample and reagent volumes. Advanced methods have already

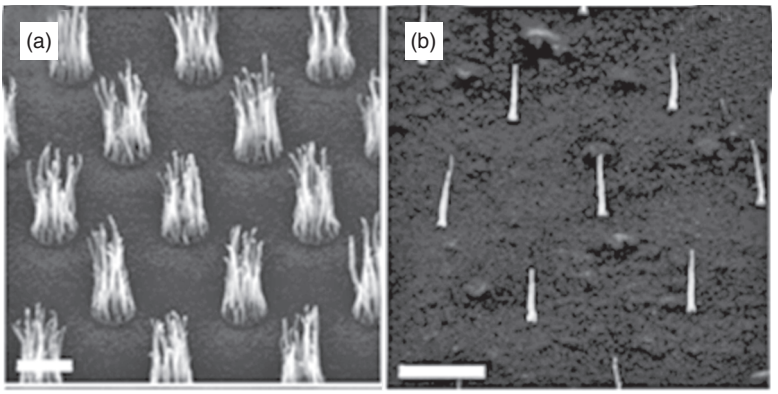


**Figure 7** SEM images of arrays of multiwalled carbon nanotubes at (a) Ultraviolet lithography and (b) e-beam patterned Ni spots. Panels are 45° perspective views with scale bars of (a) 2 and (b) 5 μm. (From ref. 79, with permission.)

been developed for the fabrication of nanoarrays for proteomic profiles and diagnostics [75–78].

Carbon nanotube electrode arrays were used by Li et al. [79] for ultrasensitive DNA detection. Scanning electron images of the arrays (Figure 7) demonstrate the precision and versatility in the fabrication of carbon nanotube electrode arrays. Nanoelectrodes are highly desirable, due to the fact that electrode performace in terms of speed and spatial resolution scale inversely with electrode radius. Oligonucleotides covalently bound to the terminus of the carbon nanotubes were able to detect target DNA sequence hybridization at concentrations of only a few attamoles. The specificity and versatility of carbon nanotube electrode arrays allow statistically significant, multiplexed detection of biomolecules in clinically relevant concentrations.

A carbon nanotube–based array sensor was employed to detect glucose by binding glucose oxidase to the terminus of the carbon nanotubes [80]. Direct electron transfer from the enzyme through the carbon nanotube and to a platinum transducer allowed real-time sensitive detection of the reaction and glucose presence. This technology may be expanded for the detection of many other biomarkers with electron-producing enzymatic reactions.

Gold nanoarrays with immobilized anti-*Escherichia coli* antibodies have been developed for the early detection of bacteria during human kidney infections [81]. A novel aspect of this work was that the antibodies were all arranged in a post configuration as opposed to a random configuration used in most studies. This allowed for the detection of bacteria in concentrations two orders of magnitude below the detection limits of current methods. Additionally, the rapid detection method eliminated the need to run bacterial cultures overnight, as is a common practice in the clinic.

### Other Novel Nanodevices for Biomarker Detection

Although the majority of research into nanodevices for biomarker detection fall into the three categories discussed above (nanowires, nanocantilevers, and nanoarrays), a plethora of other novel detection devices have been investigated, as discussed here briefly. Additionally, nanotechnology methods contribute to biomarker detection in a general sense, such as the ability to make precise nanoscale pores and channels for utilization of picoliter volumes [82–84].

In addition to their application to nanowires and nanoarrays, carbon nanotubes can be utilized as nano-biomarker probes for the precise detection of specific genes following a DNA hybridization assay [85]. The use of carbon nanotubes as probes has great potential, due to their intense Raman scattering (making it possible to detect a single nanotube) [86,87] and near-infrared fluorescence [88–91].

Nanotube-based sensors were first developed for the detection of $NO_2$ and $NH_3$ gas [92] but have since become more sophisticated to allow sensitive detection of biomarkers. Surface-modified carbon nanofiber electrodes have been developed for small-volume detection of glucose [93–95] and are being developed for glutamate and lactate [93]. Carbon nanoelectrodes coupled with oxidases may also be utilized for sensitive detection of cholesterol, alcohol, lactate, acetylcholine, choline, hypoxanthine, and xanthine from a variety of biological fluids [95]. Nanotubes have also been developed as sensors for autoimmune disease [96], single-nucleotide polymorphisms (SNPs) [97], and for reversible, small-volume pH detection [94].

Nanogap actuators have been proposed for the use of protein detection [98]. Immobilized antibodies in a nanogap actuator enable protein-binding specificity. Once bound, the target protein can be detected based on rigidity measurements yielding information on protein presence, concentration, and size.

The diverse range of nanodevices discussed above will address key issues in biomarker research, including multiplexing ability, specificity, real-time detection, and tag-free detection. As a result, they will greatly aid diagnostics, treatment monitoring, and drug discovery.

## NANOTOXICOLOGY

Many of the nanotechnology-based biomarker detection systems discussed, nanoparticle-based systems in particular, potentiate the use for real-time in vivo analysis and imaging. Although this is very exciting from the standpoint of the various applications that may benefit, the toxicology of these systems must be investigated further. The field of nanotoxicology has recently intensified and raised many more questions than answers provided [99]. Two conclusions that are agreed upon are that the biological interaction with nanomaterials

is different from bulk materials and that further research in nanotoxicology must be done. The potential toxicity of nanomaterials represents the largest barrier to in vivo, ex vivo, and in vitro cell culture biomarker assays currently being developed.

Qdots have been the most intensely researched system for biomarker detection. However, their cytotoxicity is well known, requiring surface modification that can advantageously and disadvantageously alter imaging properties. Qdots are particularly toxic when ultraviolet light sorces are used [22]. This has led to the investigation of other materials with similar imaging properties and less cytotoxicity, such as nanodiamonds [43]. Regardless, the clearance of the nanoparticles is not clearly understood, and qdots have been seen in the lymph and bone marrow months after their administration [24]. On the other hand, some advances in nanotechnology-based biomarker detection address issues of toxicity that arise from conventional biomarker assay. For example, the nano-PEBBLEs discussed protect cells from cytotoxic fluorophores and therefore extend the range of fluorophores that can be utilized [45].

## CHALLENGES FOR NANOTECHNOLOGY-BASED BIOMARKER DETECTORS

Nanotoxicology is the greatest challenge facing nanotechnology-based biomarker detection systems. However, nanotechnology-based biomarker detection systems face many other challenges, will intensify current challenges, and present new challenges to biomarker research. The second major challenge is biofouling of biomarker detection systems. Although this is not a new challenge, it is more critical with nanodevices, where biofouling could be exacerbated due to the nanoscale of the detection systems. The need for biofouling-resistant materials will therefore be intensified.

The potential of nanotechnology applications to biomarker research can be realized fully only if there is further development of mathematical models. Understanding nanocantilever beam stresses, identification of biological signatures from nanocantilever and nanowire devices, and accurate three-dimensional reconstruction of in vivo imaging data are all areas that represent a great need for further developed mathematical models. Additionally, further optimization of all the detection systems discussed is needed to ensure reproducibility, statistically relevant data, and prevention of false positives. For instance, it is not fully understood why qdots with great imaging characteristics in vitro tend to experience fluorescence loss when utilized in vivo. Finally, understanding transit mechanisms of the nanoparticle systems is needed to prevent undesired pathway susceptibility, such as endosomal entrapment.

As nanotechnology-based biomarker detection systems develop, further development will also be needed in optical detection systems and in the sensitivity and selectivity of electronic devices. Despite all the barriers discussed here, the factors that will truly determine if nanoparticles and nanodevices

become standard tools in the clinical biomarker toolbox are the cost, ease of integration with current infrastructure, and performance variability [100].

## FUTURE OF NANOTECHNOLOGY-BASED BIOMARKER DETECTORS

Nanotechnology will present many solutions and many barriers to biomarker research. For nanoparticle systems, nanotoxicology research will need to identify safe, nontoxic materials for cell culture and in vivo use. Particular emphasis should be placed on understanding nanoparticle biodistribution, clearance, and the molecular mechanisms and pathways involved in their transit. Biofouling-resistant materials and further development of mathematical models are necessary for the full potential of nanodevices in biomarker research to be realized. However, nanotechnology-based biomarker systems able to yield real-time, multiplexed results with high accuracy and specificity will change the utilization of current biomarkers and allow the discovery of new biomarkers.

## REFERENCES


1. Ferrari M (2005). Cancer nanotechnology: opportunities and challenges. *Nat Rev Cancer*, 5(3):161–171.
2. Arntz Y, Seelig J, Lang H, et al. (2003). Label-free protein assay based on a nanomechanical cantilever array. *Nanotechnology*, 14:86–90.
3. Sander C (2000). Genomic medicine and the future of health care. *Science*, 287(5460):1977–1978.
4. Etzioni R, Urban N, Ramsey S, et al. (2003). The case for early detection. *Nat Rev Cancer*, 3(4):243–252.
5. Srinivas PR, Kramer BS, Srivastava S (2001). Trends in biomarker research for cancer detection. *Lancet Oncol*, 2(11):698–704.
6. Wulfkuhle JD, Liotta LA, Petricoin EF (2003). Proteomic applications for the early detection of cancer. *Nat Rev Cancer*, 3(4):267–275.
7. Medina C, Santos-Martinez MJ, Radomski A, Corrigan OI, Radomski MW (2007). Nanoparticles: pharmacological and toxicological significance. *Br J Pharmacol*, 150(5):552–558.
8. Alivisatos A (1996). Semiconductor clusters, nanocrystals and quantum dots. *Science*, 271(5251):933–937.
9. Michalet X, Pinaud FF, Bentolila LA, et al. (2005). Quantum dots for live cells, in vivo imaging, and diagnostics. *Science*, 307(5709):538–544.
10. Larson DR, Zipfel WR, Williams RM, et al. (2003). Water-soluble quantum dots for multiphoton fluorescence imaging in vivo. *Science*, 300(5624):1434–1436.
11. Akerman ME, Chan WC, Laakkonen P, Bhatia SN, Ruoslahti E (2002). Nanocrystal targeting in vivo. *Proc Natl Acad Sci USA*, 99(20):12617–12621.

12. Dubertret B, Skourides P, Norris DJ, Noireaux V, Brivanlou AH, Libchaber A (2002). In vivo imaging of quantum dots encapsulated in phospholipid micelles. *Science*, 298(5599):1759–1762.
13. Wu X, Liu H, Liu J, et al. (2003). Immunofluorescent labeling of cancer marker Her2 and other cellular targets with semiconductor quantum dots. *Nat Biotechnol*, 21(1):41–46.
14. Jaiswal JK, Mattoussi H, Mauro JM, Simon SM (2003). Long-term multiple color imaging of live cells using quantum dot bioconjugates. *Nat Biotechnol*, 21(1):47–51.
15. Ishii D, Kinbara K, Ishida Y, et al. (2003). Chaperonin-mediated stabilization and ATP-triggered release of semiconductor nanoparticles. *Nature*, 423(6940):628–632.
16. Medintz IL, Clapp AR, Mattoussi H, Goldman ER, Fisher B, Mauro JM (2003). Self-assembled nanoscale biosensors based on quantum dot FRET donors. *Nat Mater*, 2(9):630–638.
17. Dahan M, Levi S, Luccardini C, Rostaing P, Riveau B, Triller A (2003). Diffusion dynamics of glycine receptors revealed by single-quantum dot tracking. *Science*, 302(5644):442–445.
18. Rosenthal SJ, Tomlinson I, Adkins EM, et al. (2002). Targeting cell surface receptors with ligand-conjugated nanocrystals. *J Am Chem Soc*, 124(17): 4586–4594.
19. Yezhelyev M, Gao X, Xing Y, Al-Hajj A, Nie S, O'Regan R (2006). Emerging use of nanoparticles in diagnosis and treatment of breast cancer. *Lancet Oncol*, 7:657–667.
20. Gao X, Cui Y, Levenson RM, Chung LW, Nie S (2004). In vivo cancer targeting and imaging with semiconductor quantum dots. *Nat Biotechnol*, 22(8):969–976.
21. Kim S, Lim YT, Soltesz EG, et al. (2004). Near-infrared fluorescent type II quantum dots for sentinel lymph node mapping. *Nat Biotechnol*, 22(1):93–97.
22. Derfus A, Chan W, Bhatia S (2004). Probing the cytotoxicity of semiconductor quantum dots. *Nano Lett*, 4(1):11–18.
23. Ness JM, Akhtar RS, Latham CB, Roth KA (2003). Combined tyramide signal amplification and quantum dots for sensitive and photostable immunofluorescence detection. *J Histochem Cytochem*, 51(8):981–987.
24. Ballou B, Lagerholm BC, Ernst LA, Bruchez MP, Waggoner AS (2004). Noninvasive imaging of quantum dots in mice. *Bioconjug Chem*, 15(1):79–86.
25. Lidke DS, Nagy P, Heintzmann R, et al. (2004). Quantum dot ligands provide new insights into *erbB/HER* receptor-mediated signal transduction. *Nat Biotechnol*, 22(2):198–203.
26. Oyewumi MO, Yokel RA, Jay M, Coakley T, Mumper RJ (2004). Comparison of cell uptake, biodistribution and tumor retention of folate-coated and PEG coated gadolinium nanoparticles in tumor-bearing mice. *J Control Release*, 95(3):613–626.
27. Schellenberger EA, Bogdanov A Jr, Hogemann D, Tait J, Weissleder R, Josephson L (2002). Annexin V-CLIO: a nanoparticle for detecting apoptosis by MRI. *Mol Imaging*, 1(2):102–107.

28. Harisinghani MG, Barentsz J, Hahn PF, et al. (2003). Noninvasive detection of clinically occult lymph-node metastases in prostate cancer. *N Engl J Med*, 348(25):2491–2499.
29. Winter PM, Morawski AM, Caruthers SD, et al. (2003). Molecular imaging of angiogenesis in early-stage atherosclerosis with alpha(v)beta3-integrin-targeted nanoparticles. *Circulation*, 108(18):2270–2274.
30. Perez JM, Simeone FJ, Saeki Y, Josephson L, Weissleder R (2003). Viral induced self-assembly of magnetic nanoparticles allows the detection of viral particles in biological media. *J Am Chem Soc*, 125(34):10192–10193.
31. Zhang Y, Sun C, Kohler N, Zhang M (2004). Self-assembled coatings on individual monodisperse magnetite nanoparticles for efficient intracellular uptake. *Biomed Microdevices*, 6(1):33–40.
32. Yan F, Xu H, Anker J, et al. (2004). Synthesis and characterization of silica-embedded iron oxide nanoparticles for magnetic resonance imaging. *J Nanosci Nanotechnol*, 4(1–2):72–76.
33. Rabin O, Manuel Perez J, Grimm J, Wojtkiewicz G, Weissleder R (2006). An x ray computed tomography imaging agent based on long-circulating bismuth sulphide nanoparticles. *Nat Mater*, 5(2):118–122.
34. Sipkins DA, Cheresh DA, Kazemi MR, Nevin LM, Bednarski MD, Li KC (1998). Detection of tumor angiogenesis in vivo by alphavbeta3-targeted magnetic resonance imaging. *Nat Med*, 4(5):623–626.
35. Anderson SA, Rader RK, Westlin WF, et al. (2000). Magnetic resonance contrast enhancement of neovasculature with alpha(v)beta(3)-targeted nanoparticles. *Magn Reson Med*, 44(3):433–439.
36. Winter PM, Caruthers SD, Kassner A, et al. (2003). Molecular imaging of angiogenesis in nascent Vx-2 rabbit tumors using a novel alpha(nu)beta3-targeted nanoparticle and 1.5 tesla magnetic resonance imaging. *Cancer Res*, 63(18):5838–5843.
37. Morawski AM, Winter PM, Crowder KC, et al. (2004). Targeted nanoparticles for quantitative imaging of sparse molecular epitopes with MRI. *Magn Reson Med*, 51(3):480–486.
38. Louie A, Huber M, Ahrens E, et al. (2000). In vivo visualization of gene expression using magnetic resonance imaging. *Nat Biotechnol*, 18:321–325.
39. Neuwelt EA, Varallyay P, Bago AG, Muldoon LL, Nesbit G, Nixon R (2004). Imaging of iron oxide nanoparticles by MR and light microscopy in patients with malignant brain tumours. *Neuropathol Appl Neurobiol*, 30(5): 456–471.
40. Hayflick L (1997). Mortality and immortality at the cellular level: a review. *Biochemistry* (*Mosc*), 62(11):1180–1190.
41. Grimm J, Perez JM, Josephson L, Weissleder R (2004). Novel nanosensors for rapid analysis of telomerase activity. *Cancer Res*, 64(2):639–643.
42. Kirchner C, Liedl T, Kudera S, et al. (2005). Cytotoxicity of colloidal CdSe and CdSe/ZnS nanoparticles. *Nano Lett*, 5(2):331–338.
43. Fu CC, Lee HY, Chen K, et al. (2007). Characterization and application of single fluorescent nanodiamonds as cellular biomarkers. *Proc Natl Acad Sci USA*, 104(3):727–732.

44. Santra S, Zhang P, Wang K, Tapec R, Tan W (2001). Conjugation of biomolecules with luminophore-doped silica nanoparticles for photostable biomarkers. *Anal Chem*, 73(20):4988–4993.
45. Xu H, Aylott JW, Kopelman R (2002). Fluorescent nano-PEBBLE sensors designed for intracellular glucose imaging. *Analyst*, 127(11):1471–1477.
46. Clark HA, Hoyer M, Philbert MA, Kopelman R (1999). Optical nanosensors for chemical analysis inside single living cells: 1. Fabrication, characterization, and methods for intracellular delivery of PEBBLE sensors. *Anal Chem*, 71(21):4831–4836.
47. Clark HA, Kopelman R, Tjalkens R, Philbert MA (1999). Optical nanosensors for chemical analysis inside single living cells: 2. Sensors for pH and calcium and the intracellular application of PEBBLE sensors. *Anal Chem*, 71(21):4837–4843.
48. Brasuel M, Kopelman R, Miller TJ, Tjalkens R, Philbert MA (2001). Fluorescent nanosensors for intracellular chemical analysis: decyl methacrylate liquid polymer matrix and ion-exchange-based potassium PEBBLE sensors with real-time application to viable rat C6 glioma cells. *Anal Chem*, 73(10):2221–2228.
49. Xu H, Aylott JW, Kopelman R, Miller TJ, Philbert MA (2001). A real-time ratiometric method for the determination of molecular oxygen inside living cells using sol-gel-based spherical optical nanosensors with applications to rat C6 glioma. *Anal Chem*, 73(17):4124–4133.
50. Yu X, Song S-K, Chen J, et al. (2000). High-resolution MRI characterization of human thrombus using a novel fibrin-targeted paramagnetic nanoparticle contrast agent. *Magn Reson Med*, 44:867–872.
51. May DJ, Allen JS, Ferrara KW (2002). Dynamics and fragmentation of thick shelled microbubbles. *IEEE Trans Ultrason Ferroelectr Freq Control*, 49(10):1400–1410.
52. Bloch S, Wan M, Dayton P, Ferrara KW (2004). Optical observation of lipid- and polymer-shelled ultrasound microbubble contrast agents. *Appl Phys Lett*, 84:631–633.
53. Nam J, Mirkin C (2004). Bio-barcode-based DNA detection with PCR-like sensitivity. *J Am Chem Soc*, 126:5932–5933.
54. Nam JM, Thaxton CS, Mirkin CA (2003). Nanoparticle-based bio-bar codes for the ultrasensitive detection of proteins. *Science*, 301(5641):1884–1886.
55. Gao F, Wang L, Tang L, Zhu C (2005). A Novel nano-sensor based on rhodamine-b-isothiocyanate-doped silica nanoparticle for pH measurement. *Microchim Acta*, 152:131–135.
56. Yoon TJ, Yu KN, Kim E, et al. (2006). Specific targeting, cell sorting, and bioimaging with smart magnetic silica core-shell nanomaterials. *Small*, 2(2):209–215.
57. Lesniak W, Bielinska AU, Sun K, et al. (2005). Silver/dendrimer nanocomposites as biomarkers: fabrication, characterization, in vitro toxicity, and intracellular detection. *Nano Lett*, 5(11):2123–2130.
58. Frank JA, Miller BR, Arbab AS, et al. (2003). Clinically applicable labeling of mammalian and stem cells by combining superparamagnetic iron oxides and transfection agents. *Radiology*, 228(2):480–487.

59. Kraitchman DL, Heldman AW, Atalar E, et al. (2003). In vivo magnetic resonance imaging of mesenchymal stem cells in myocardial infarction. *Circulation*, 107(18):2290–2293.
60. McFarland A, Duyne R (2003). Single silver nanoparticles as real-time optical sensors with zeptomole sensitivity. *Nano Lett*, 3:1057–1062.
61. Haes A, Van Duyne R (2003). A nanoscale optical biosensor: sensitivity and selectivity of an approach based on the localized surface plasmon resonance spectroscopy of triangular silver nanoparticles. *J Am Chem Soc*, 124:10596–10604.
62. Geho DH, Lahar N, Ferrari M, Petricoin EF, Liotta LA (2004). Opportunities for nanotechnology-based innovation in tissue proteomics. *Biomed Microdevices*, 6(3):231–239.
63. Liotta LA, Ferrari M, Petricoin E (2003). Clinical proteomics: written in blood. *Nature*, 425(6961):905.
64. Cui Y, Wei Q, Park H, Lieber CM (2001). Nanowire nanosensors for highly sensitive and selective detection of biological and chemical species. *Science*, 293(5533):1289–1292.
65. Hahm J-I, Lieber C (2004). Direct ultrasensitive electrical detection of DNA and DNA sequence variations using nanowire nanosensors. *Nano Lett*, 4(1):51–54.
66. Zheng G, Patolsky F, Cui Y, Wang WU, Lieber CM (2005). Multiplexed electrical detection of cancer markers with nanowire sensor arrays. *Nat Biotechnol*, 23(10):1294–1301.
67. Wang WU, Chen C, Lin KH, Fang Y, Lieber CM (2005). Label-free detection of small-molecule-protein interactions by using nanowire nanosensors. *Proc Natl Acad Sci USA*, 102(9):3208–3212.
68. Patolsky F, Zheng G, Hayden O, Lakadamyali M, Zhuang X, Lieber CM (2004). Electrical detection of single viruses. *Proc Natl Acad Sci USA*, 101(39):14017–14022.
69. Hansen KM, Ji HF, Wu G, et al. (2001). Cantilever-based optical deflection assay for discrimination of DNA singlenucleotide mismatches. *Anal Chem*, 73(7):1567–1571.
70. Wu G, Datar RH, Hansen KM, Thundat T, Cote RJ, Majumdar A (2001). Bioassay of prostate-specific antigen (PSA) using microcantilevers. *Nat Biotechnol*, 19(9):856–860.
71. Su M, Li S, Dravid V (2003). Microcantilever resonance-based DNA detection with nanoprobes. *Appl Phys Lett*, 82(20):3562–3564.
72. Chen H, Han J, Li J, Meyyappan M (2004). Microelectronic DNA assay for the detection of BRCA1 gene mutations. *Biomed Microdevices*, 6(1):55–60.
73. Bashir R, Hilt J, Elibol O, Gupta A, Peppas N (2002). Micromechanical cantilever as an ultrasensitive pH microsensor. *Appl Phys Lett*, 81(16):3091–3093.
74. Baller MK, Lang HP, Fritz J, et al. (2000). A cantilever array-based artificial nose. *Ultramicroscopy*, 82(1–4):1–9.
75. Demers LM, Ginger DS, Park SJ, Li Z, Chung SW, Mirkin CA (2002). Direct patterning of modified oligonucleotides on metals and insulators by dip-pen nanolithography. *Science*, 296(5574):1836–1838.

76. Lee KB, Park SJ, Mirkin CA, Smith JC, Mrksich M (2002). Protein nanoarrays generated by dip-pen nanolithography. *Science*, 295(5560):1702–1705.
77. Lee KB, Lim JH, Mirkin CA (2003). Protein nanostructures formed via direct write dip-pen nanolithography. *J Am Chem Soc*, 125(19):5588–5589.
78. Bruckbauer A, Zhou D, Kang DJ, Korchev YE, Abell C, Klenerman D (2004). An addressable antibody nanoarray produced on a nanostructured surface. *J Am Chem Soc*, 126(21):6508–6509.
79. Li J, Ng HT, Cassell A, et al. (2003). Carbon nanotube nanoelectrode array for ultrasensitive DNA detection. *Nano Lett*, 3(5):597–602.
80. Sotiropoulou S, Chaniotakis NA (2003). Carbon nanotube array-based biosensor. *Anal Bioanal Chem*, 375(1):103–105.
81. Basu M, Seggerson S, Henshaw J, et al. (2004). Nano-biosensor development for bacterial detection during human kidney infection: use of glycoconjugate-specific antibody-bound gold NanoWire arrays (GNWA). *Glycoconj J*, 21(8–9):487–496.
82. Chu W-H, Chin R, Huen T, Ferrari M (1999). Silicon membrane nanofilters from sacrificial oxide removal. *J Microeletromech Syst*, 8(1):34–42.
83. Desai T, Hansford D, Kulinski L, et al. (1999). Nanopore technology for biomedical applications. *Biomed Microdevices*, 2(1):11–40.
84. Han J, Craighead HG (2000). Separation of long DNA molecules in a microfabricated entropic trap array. *Science*, 288(5468):1026–1029.
85. Hwang E-S, Cao C, Hong S, et al. (2006). The DNA hybridization assay using single-walled carbon nanotubes as ultrasensitive, long-term optical labels. *Nanotechnology*, 17:3442–3445.
86. McCreery R (2002). Photometric standards for Raman spectroscopy. In Chalmers J, Griffiths P (eds.), *Handbook of Vibrational Spectroscopy*. Wiley, Hoboken, NJ.
87. Dresselhaus MS, Dresselhaus G, Jorio A, Souza Filho AG, Pimenta MA, Saito R (2002). Single nanotube Raman spectroscopy. *Acc Chem Res*, 35(12):1070–1078.
88. Barone PW, Baik S, Heller DA, Strano MS (2005). Near-infrared optical sensors based on single-walled carbon nanotubes. *Nat Mater*, 4(1):86–92.
89. Barone PW, Parker RS, Strano MS (2005). In vivo fluorescence detection of glucose using a single-walled carbon nanotube optical sensor: design, fluorophore properties, advantages, and disadvantages. *Anal Chem*, 77(23):7556–7562.
90. Wray S, Cope M, Delpy DT, Wyatt JS, Reynolds EO (1988). Characterization of the near infrared absorption spectra of cytochrome aa3 and haemoglobin for the non-invasive monitoring of cerebral oxygenation. *Biochim Biophys Acta*, 933(1):184–192.
91. O'Connell MJ, Bachilo SM, Huffman CB, et al. (2002). Band gap fluorescence from individual single-walled carbon nanotubes. *Science*, 297(5581):593–596.
92. Kong J, Franklin NR, Zhou C, et al. (2000). Nanotube molecular wires as chemical sensors. *Science*, 287(5453):622–625.
93. Zhang X, Wang J, Ogorevc B, Spichiger U (1999). Glucose nanosensor based on Prussian-Blue modified carbon-fiber cone nanoelectrode and an integrated reference electrode. *Electroanalysis*, 11(13):945–949.

94. Besteman K, Lee J-O, Wiertz F, Heering H, Dekker C (2003). Enzyme-coated carbon nanotubes as single-molecule biosensors. *Nano Lett*, 3(6):727–730.
95. Lin Y, Lu F, Tu Y, Ren Z (2004). Glucose biosensors based on carbon nanotube nanoelectrode ensembles. *Nano Lett*, 4(2):191–195.
96. Chen RJ, Bangsaruntip S, Drouvalakis KA, et al. (2003). Noncovalent functionalization of carbon nanotubes for highly specific electronic biosensors. *Proc Natl Acad Sci USA*, 100(9):4984–4989.
97. Woolley AT, Guillemette C, Li Cheung C, Housman DE, Lieber CM (2000). Direct haplotyping of kilobase-size DNA using carbon nanotube probes. *Nat Biotechnol*, 18(7):760–763.
98. Lee W, Cho Y-H (2004). Nanomechanical protein detectors using electrothermal nanogap actuators. *17th IEEE International Conference on Micro Electro Mechanical Systems*, Maastricht, The Netherlands. IEEE Press, Piscataway, NJ, pp. 629–632.
99. Donaldson K, Stone V, Tran CL, Kreyling W, Borm PJ (2004). Nanotoxicology. *Occup Environ Med*, 61(9):727–728.
100. Portney NG, Ozkan M (2006). Nano-oncology: drug delivery, imaging, and sensing. *Anal Bioanal Chem*, 384(3):620–630.

# FROM GENE TO PRODUCT: THE ADVANTAGE OF INTEGRATIVE BIOTECHNOLOGY

Frank-Ranier Schmidt
*Sanofi-Aventis Deutschland, Frankfurt am Main, Germany*

**Chapter Contents**

1 Introduction
2 Production of Organisms and Expression Systems
  2.1 Industrially Established Recombinant Expression Systems
  2.2 Evaluation of Secretory Expression Systems for Pharmaceutical Purposes
  2.3 Criteria for the Choice of Recombinant Expression Systems
3 Enhancement of Productivity
  3.1 Secretory Recombinant Expression Systems
  3.2 Natural Products
  3.3 Approaches and Goals for Further Strain Improvement
4 Biosynthetic Structure Modification
  4.1 Combinatorial Biosynthesis
  4.2 Precursor Directed Biosynthesis
5 Fermentation Optimization and Scale-Up
  5.1 Reduced Mixing Quality and Enhanced Stress Exposure
  5.2 Process Characterization
  5.3 Process Optimization
  5.4 Physical Scale-Up Parameters
  5.5 Development of Fermentation Models and Strategies
6 Downstream Processing
  6.1 Product Recovery and Purification
  6.2 Downstream Processing Optimization and Economization
  6.3 Downstream Processing Scale-Up
  6.4 Downstream Processing of β-Lactam Compounds

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

7 Postsynthetic Structure Modification
7.1 β-Lactam Side Chain Cleavage
8 Quality Issues
8.1 Virus and Endotoxin Removal
9 Conclusion
References

## 1 INTRODUCTION

Biotechnology and biotechnology-based methods are increasing in importance in medical therapies and diagnostics as well as in the discovery, development, and manufacture of pharmaceuticals. Biotechnologically manufactured pharmaceuticals will soon reach a market volume of more than $100 billion USD and, thus, some 20% of the total pharmaceutical market. The key step in their manufacture is the conversion of the genetic information into a product with the desired pharmacological activities by an appropriate selection, design, and cultivation of cells and microorganisms harboring the corresponding biosynthetic pathways and physiological properties.

The intention of this chapter is to give insights into the typical issues and problems encountered in the manufacture of biopharmaceuticals, to mediate general ideas and current strategies on how to proceed in the design and development of biotechnological processes, and to deliver the immediate theoretical backgrounds necessary for comprehension rather than to give detailed experimental instructions like a manual does. In focusing on gene recombinant proteins and peptidic antibiotics, the biotechnologically produced pharmaceuticals with the highest market share, representative aspects will be discussed (1) for process development and optimization approaches to increase product yield and process rentability and to ensure a consistent product quality, (2) for experimental approaches to design and to modify the molecular structure of compounds to meet specific medical needs, (3) for the replacement of chemical procedures by economically and ecologically advantageous biotechnological processes, (4) for critical issues of product purification, and (5) for specific demands in pharmaceutical production to conform to regulatory requirements. Finally, the advantage of an integrative biotechnology is emphasized, which designs the biosynthetic steps of the product in accordance with the requirements of product purification procedures already during early development stages.[1]

For further reviews comprehensively illustrating issues of biotechnological production processes, the reader is referred to further review articles [1–8]. To look up basic subjects of molecular and cellular biology, the reader is referred to textbooks [9, 10].

[1] Parts of this chapter were excerpted and modified from previously published reviews of the author [16, 86, 98, 113, 273] by courtesy of Springer-Verlag, Heidelberg.

## 2 PRODUCTION ORGANISMS AND EXPRESSION SYSTEMS

Design and development of all microbial production processes start with the selection of appropriate organisms, strains, and expression systems enabling high yields and high quality of a desired product with defined pharmacological properties.

### 2.1 Industrially Established Recombinant Expression Systems

Industrially established expressions systems for production of the marketed compounds are, besides inclusion, body-forming *Escherichia coli* strains, the yeast *Saccharomyces cerevisiae* and mammalian cells like CHO- and BHK-cells (Table 1). These systems were the genetically and physiologically most advanced and therefore mostly applied when recombinant production processes were starting to be developed in the mid-1980s and are now widely accepted by regulatory bodies. *E. coli* and *S. cerevisiae* can be grown cheaply and rapidly, are amenable to high cell density fermentations with biomasses of up to 130 g/L, possess short generation times, have high capacities to accumulate foreign proteins, are easy to handle, and are established fermentation organisms.

However, because gene recombinant pharmaceuticals continuously gain an increasing importance in medicine and are expected to help curing diseases that are not yet treatable today, new expression systems have to be exploited enabling the production of such pharmaceuticals with innovative properties that simultaneously meet key criteria like consistent product quality and cost effectiveness. Of particular interest in this regard are expression systems enabling the secretion of

**TABLE 1. Industrially Used Recombinant Expression Systems**

| Product | Company | System |
|---|---|---|
| Blood coagulation factors (VII, VIII, IX) | Novo-Nordisk/Bayer/ Centeon Genetics Baxter/ Centeon/Wyeth | BHK-Cells<br>CHO-Cells |
| Calcitonin | Unigene | *E. coli*/CHO-Cells |
| DNase (cystic fibrosis) | Roche | CHO-Cells |
| Erythropoetin | Janssen-Cilag/Amgen/ Boehringer | CHO-Cells |
| Darbepoetin | Amgen | CHO-Cells |
| Follicle stimulating hormone (follitropin) | Serono/Organon | CHO-Cells |
| Luteinisation hormone | Serono | CHO-Cells |
| Gonadotropin | Serono | CHO-Cells |
| Glucagon | Novo-Nordisk | *S. cerevisiae* |
| Glucocerebrosidase (Gaucher-disease) | Genzyme | CHO-Cells |
| Growth hormones (somatotropines) | Pharmacia & Upjohn/Lilly/ Novo-Nordisk/Ferring/ Genentech | *E. coli* |
| | Serono | Mouse Cell Line |
| | Serono/Bio-Technology General Corp | CHO-Cells |

**TABLE 1.** *Continued*

| Product | Company | System |
|---|---|---|
| Eutropin (Human growth hormone derivative) | LG Chemical | *S. cerevisiae* |
| Growth factors (GCSF u. GMCSF) | Novartis/Essex/Amgen/Roche | *E. coli* |
| | Chugai Pharmaceuticals | CHO-Cells |
| Platelet-derived growth factor (PDGF) | Janssen-Cilag | *S. cerevisiae* |
| PDGF-Agonist | ZymoGenetics | *S. cerevisiae* |
| Hepatitis B vaccine | GlaxoSmithKline | *S. cerevisiae* |
| | Rhein Biotech | *H. polymorpha* |
| Hirudin | Sanofi-Aventis/Novartis | *S. cerevisiae* |
| Insulin and muteins | Sanofi-Aventis/Lilly/Berlin-Chemie | *E. coli* |
| Insulin | Bio-Technology General Corp | *E. coli* |
| | Novo-Nordisk | *S. cerevisiae* |
| Interferon alpha and muteins | Roche/Essex/Yamanouchi | *E. coli* |
| Interferon beta | Schering | *E. coli* |
| | Biogen/Serono | CHO-Cells |
| Interferon gamma (mutein) | Amgen/Boehringer | *E. coli* |
| Interleukin 2 | Chiron | *E. coli* |
| Oprelvekin (interleukin 11-agonist) | Wyeth | Human Cell Line ROMI 8866 |
| OP-1 (osteogenic, neuroprotective factor) | Curis/Striker | *E. coli* |
| Tissue plasminogen-activator | Genentech/Roche/Boehringer | CHO-Cells |
| Recombinant plasminogen-activator | Genentech/Roche/Boehringer | *E. coli* |
| Stem cell factor | Amgen | CHO-Cells |
| Tumor necrosis factor | Boehringer | *E. coli* |

*Note*: Overview on the currently worldwide commercialized recombinant pharmaceuticals and the expression systems employed for their production. The substances are not listed strictly alphabetically but are partially grouped according to therapeutic areas. Antibodies, which are mostly manufactured by hybridoma cell line systems, are not considered. Data were extracted from the European patent database Esp@cenet (http://de.espacenet.com) and the IDdb3-database (http://www.iddb3.com). BHK = baby hamster kidney; CHO = chinese hamster ovary. (Taken from Ref. 16 © Springer-Verlag, Heidelberg)

correctly glycosylated and folded proteins into the culture broth. Such secretory systems offer advantages in terms of simple and fast product purification procedures and the avoidance of costly cell rupture, denaturation, and refolding processes (see Section 4) and thus conform to the requirements of an integrated production process.

Even though animal and plant systems (molecular pharming, [11]) and secretory plant cell culture systems have received a great deal of attention, their commercial feasibility is still under investigation, particularly with respect to their slightly different posttranslational modification modus leading to an altered pharmacological behavior and to allergenic properties [12, 13]. Established in the pharmaceutical

industry as production organisms are, besides the above-mentioned systems, further prokaryotic and yeast species as well as filamentous fungi, which are already employed for the manufacture of natural products (see Section 2.2).

The suitability of the most prominent secretory systems among these organisms from the viewpoint of an integrative process design for the manufacture of recombinant proteins will be evaluated in the next section by discussing their potential productivity and their physiological properties.

## 2.2 Evaluation of Secretory Expression Systems for Pharmaceutical Purposes

***Escherichia Coli.*** As *E. coli* lacks fundamental prerequisites for efficient secretion, the marketed pharmaceuticals (Table 1) manufactured by *E. coli*-systems are mostly produced as inclusion bodies. Due to the membrane structure, the low chaperone and foldase level and the high periplasmatic protease concentration *E. coli*-secretion systems allow only comparably low product yields, making them suitable only for compounds marketed in small quantities like orphan drugs. Genentech (San Francisco, CA), for instance, has patented a secretory *E. coli*-system for the preparation of human growth hormone [14]. The secretory potential of *E. coli* is indicated by exceptional high product titers in the range of several grams per liter, which were reached in a system developed for secretion of hirudin using the alpha-cyclodextringlykosysl-transferase signal sequence as a leader und secretor mutants deficient in their membrane structure [15]. Titers of human-insulin-like-growth-factor or human-epidermal-growth-factor were reported to be as high as 900 mg/L and 325 mg/L, respectively [references in 16]. Most of the reached and published data, however, refers to processes leading to a periplasmatic product concentration (e.g., 2 g/L of a human antibody fragment, 700 mg/L of a monoclonal antibody) or stays below 100 mg/L, a value that generally is not considered to be in a competitive and economic range. Efforts are thus undertaken to condition *E. coli*-strains to efficient secreters. The main strategies to enhance secretion efficiency [17–19] comprise (1) employment of well-characterized secretion pathways like the alpha-hemolysin system [20] or components of such pathways like efficient signal sequences from efflux proteins [21] or outer membrane proteins [22], for instance, the maltose binding protein [23] or the TolC-protein [24]; (2) variation of the signal peptide; (3) cocloning of and coexpression of chaperones and foldases [25–27]; (4) enhancement of gene expression by employment of strong promotors and efficient transcription termination sequences; (5) generation of protease deficient mutants; (6) generation of cell wall lacking or cell wall deficient mutants [28, 29]; and (7) modulation of the protein primary structure that was found to exert a strong influence on productivity and secretion efficiency by influencing protease resistance, folding efficiency, and the tendency to form inclusion bodies. Details of the strategies for the design and development of secretory *E. coli* strains as well as for the controlled soluble cytoplasmatic expression of recombinant proteins may be taken from general reviews [17, 30–32].

***Alternative Prokaryotic Expression Systems.*** In addition to conditioning *E. coli*-strains to efficient secreters, alternative species, which are considered to inherently possess a superior secretion capacity, are tried to be established as expression

systems. Comparably high product yields of 2 g/L and 1 g/L were reported for production of human calcitonin by *Staphylococcus carnosus* [33] and of proinsulin by *Bacillus subtilis* [34], an organism that is continuously characterized and improved as a cell factory for pharmaceutical proteins [8]. *Bacillus megaterium*, which is thought to be as efficient as *B. subtilis*, is currently developed as a secretory expression system by a Collaborative Research Center (SFB) of the German Research Community (DFG). For *Ralstonia eutropha* (formerly *Alcaligenes eutrophus*), employed at ICI and Monsanto for polyhydroxyalkanoate production at a scale of several 100 $m^3$ and genomically completely sequenced, 1,2 g/L of secreted organophosphohydrolase, a model enzyme proned to form inclusion bodies in *E. coli*, were reported [35]. *R. eutropha* displays a more efficient carbohydrate metabolism than *E. coli* and is easily amenable to high cell density fermentations with biomass concentration of more than 150 g/L dry weight. This permits a lower specific productivity that in turn reduces the inclinement to form inclusion bodies and thus enables a more efficient secretion. *Rhodococcus*, *Corynebacterium*, *Mycobacterium*, actinomycetes, and streptomycetes [36] are also considered to be potentially suitable for the development of efficient secretion systems. A comparative study with recombinant alpha-amylase demonstrated that final yields as well as enzyme activity were considerably higher when produced by *Streptomyces lividans*, by which it was completely secreted than by *E. coli* in which it was concentrated periplasmatically [37]. The yields reported so far, however, are still below cost-efficient ranges. A system developed by Hoechst/Aventis for insulin production yielded around 100 mg/L, and the yields of correctly folded human CD4-receptor sites are in the range of 200 mg/L. Attached as signal proteins were the prepeptide of the alpha-amylase inhibitor from *S. tendae* (tendamistat) and the signal sequence of a protease inhibitor (LTI) from *S. longisporus*. In the course of these studies, it was found that the choice of the linker and its length strongly influences secretion efficiency. The comparably low yields, however, demonstrate that still a lot of fundamental research is necessary to render streptomyces systems competitive. (Strategies and examples for enhancement of recombinant protein expression in *S. lividans* and the current status of the genetic and physiological development are given in Refs. 38 and 39.) A general focus of research will be the detailed exploration of the twin arginine translocation (TAT) pathway, which has been recently discovered in addition to the conventional prokaryotic secretory (sec) pathways and enables the export of proteins with cofactors in a fully folded conformation [40, 41]. It evidently plays a more important role in *Streptomyces* species [42] but might also be useable in other species.

As the potential and capacity of prokaryotes for prosttranslational modification appear to be quite limited and the knowledge about the pathways is quite scarce (reviews on bacterial protein glycosylation see Refs. 43 and 44), the employability of most of the known prokaryotes usually is restricted to the preparation of proteins that are naturally not glycosylated, such as insulin, hirudins, or somatotropins, or to natively glycosylated proteins that are pharmacologically also active without glycosylation, like various cytokines (tumor necrosis factor, interleukines, interferones). For production of proteins that are pharmacologically active only with an appropriate modification pattern, eukayotic cell systems are more suitable.

***Yeasts.*** Besides possessing complex posttranslational modification pathways, they offer the advantage to be neither pyrogenic nor pathogenic and to secrete more

efficiently. Species established in industrial production procedures are *Saccharomyces cerevisiae*, *Kluyveromyces lactis, Pichia pastoris*, and *Hansenula polymorpha*, which will be dealt with more in detail in this section. Whereas *S. cerevisiae* is the best genetically characterized eukaryotic organism at all and still is the prevalent yeast species in pharmaceutical production processes (Table 1), *P. pastoris*, first employed by Phillips Petroleum for single-cell-protein production, is currently the most frequently used yeast species for heterologous protein expression in general. Whereas only just a few proteins were expressed by *Pichia* species at the beginning of the last decade [45], the expression of more than 400 proteins have been meanwhile reported now [46, 47]. *P. pastoris* is considered to be superior to any other known yeast species with respect to its secretion efficiency and permits the production of recombinant proteins without intense process development. The highest yields were reported for murine collagen (15 g/L), tetanus toxin fragment C (12 g/L compared with 1 g/L in *S. cerevisiae*), human serum albumin (10 g/L compared with 3 g/L in *K. lactis* and to 90 and 150 mg/L in *S. cerevisiae*), and human interleukin 2 (10 g/L). The highest reported yields for *H. polymorpha* relate to phytase (13,5 g/L) and to hirudin (g/L range) and for *K. lactis* to human serum albumin (3 g/L, see above). Even though *S. cerevisiae* offers a high secretory potential as evidenced by some 9-g/L secreted *Aspergillus niger* glucose oxidase [references in 16], such data document a general inferior secretory capacity, the reasons for which are numerous. For the methylotrophic species *Hansenula* and *Pichia* and the lactose using *K. lactis*, natively strong promoters are available that derive from the methanol and lactose assimilating pathways and their enzymes (e.g., alcohol- and methanol oxidase, lactose permease, galactosidase). As the enzymes of these pathways account for up to 30% of the total protein content, the metabolic efficiency with respect to the secreted protein is significantly higher as documented by Buckholz and Gleeson [48]: Only one or a few gene copies are sufficient in *P. pastoris* to gain the same yields as with 50 gene copies in *S. cerevisiae*. Furthermore, proteins with a molecular mass of above 30 kD are retained in the cytoplasma of *S. cerevisiae*, whereas *H. polymorpha* efficiently secretes proteins with a molecular mass of up to 150 kD, like the glucoamylase of *Aspergillus niger*. [Detailed overviews on the physiological properties of methylotrophic yeasts and *K. lactis* with respect to their use for recombinant protein production is given in Refs. 46, 47, and 49–53]. Further reasons for the differing secretion rates among the species are the specific proteolytic activities and the specific degrees and patterns of glycosylation. Besides having an impact on the protein's final pharmacological activity, glycosylation also exercises an influence on the folding and secretion efficiency. Among the discussed species, *S. cerevisiae* was shown to possess, besides a higher enzymatic activity in the secretion vesicles that leads to a reduced portion of intact secreted proteins [54], also the highest glycosylation capacity leading to a hyperglycosylation of the protein and a reduced secretion rate. Both the degree and the pattern of glycosylation are dependent on the genetic background of the species and strains employed as well as on the sequences of the expressed protein and adjacent regions. By employment of the natively highly glycosylated alpha-mating type factor as a secretion signal, the extent of the glycosylation of the product can be diminished or completely avoided as shown for human interleukin 6 [55]. NovoNordisk reported leader sequence-dependent insulin yields in *S. cerevisiae* [56] and in *S. cerevisiae* and *P. pastoris* [57]: The sequence and therewith the degree of glycosylation of the leader influences the

efficiency of the multistage cleavage and folding processes as well as the insulin glycosylation rate and secretability. Further enhancement of the secretion efficiency can be achieved by (1) mutating secretion enhancer genes [58], (2) suppressing secretion blocking functions, and (3) reducing proteolytic activities in secretion vesicles [54]. So-called supersecreter strains of *S. cerevisiae* have, for instance, been generated by inactivation of the PMR1 (SSC1) function and suppression of the secretion blocking ypt1-1 gene: the yields of non-glycosylated human pro-urokinase [59], of human serum albumin, and of human plasminogen-activator have been augmented to a factor of up to 10 [60]. The traditional approaches pursued for enhancement of gene expression are gene amplification, employment of strong promoters, and enhancement of the transcription and translation rate. A high amplification of the gene copy number [45, 46, 61] usually is achievable with episomal vectors, which however do not reach the mitotic stability of integrative systems like the transposon (e.g., Ty-element) mediated embedment of reiterative, dispers repetitive sequences in *S. cerevisae*. Transcription rates were reported to be enhanced up to 100 times through cotransformation with transcription activators and enhancers, which evidently are limiting factors for overexpression of foreign proteins [62, 63]. Translation efficiency can be enhanced by preventing an accelerated degradation of transcripts and the yeast typical random transcription termination through modulation of the recognition sequences. Prevention of the random transcription termination led to an increase of tetanus toxin fragment C yields in *S. cerevisiae* by a factor 2000–3000 to 1 g/L and 3% of the total soluble protein fraction [64].

Despite their physiologically advantageous properties and natively high expression and secretion capacity, just one industrial application is reported for each of the alternative yeast species: *H. polymorpha* is employed for hepatitis B vaccine production at Rhein-Biotech (Düsseldorf, Germany), *K. lactis* for bovine prochymosin production in a 40-$m^3$ scale at Gist-Brocades (Delft, Netherlands), and *P. pastoris* for production of recombinant carboxypeptidase B and trypsin at Roche (Basel, Switzerland). For pharmaceutical application, it has to be considerered that the methylotrophic yeasts in contrast to *S. cerevisiae* and *K. lactis* are not used in the production of foodstuffs and therewith have no GRAS (generally regarded as safe) status according to the U.S. Food and Drug Administration (FDA) criteria and have to be grown in expensive explosion-proofed equipments when the above-mentioned native induction systems are used. The discussed properties of the mentioned yeast species are compiled in Table 2 for comparison. The employability

**TABLE 2. Typical Sequence of Biotechnological Production Process Steps**

| Step | Method/Approach |
|---|---|
| **Selection/design/engineering/ development of an appropriate species/strain/ expression system** | **Criteria:** pharmacological activity and properties of the compound, productivity, process behavior, suitability for downstream processing steps, spectrum and pharmacological activity of side products to be removed, experimental experience with the respective system, biological and medical safety, acceptance by regulatory bodies |

**TABLE 2.** *Continued*

| Step | Method/Approach |
|---|---|
| **Strain improvement** | Mutation/selection, strain recombination (e.g., breeding/protoplast fusion), directed genomic alteration/metabolic engineering/enhancement of gene expression rates |
| **Biosynthetic product structure modification** | Amino acid exchange/combinatorial biosynthesis/precursor directed biosynthesis |
| **Fermentation optimization** | Empirical optimization of culture conditions (media components, pH, oxygen supply), clarification of the influence of process parameters on growth, productivity and side product formation |
| **Fermentation scale-up** | Detailed process characterization for reduction of stress exposure, insurance of homogenized reaction conditions, and identification of suitable scale-up parameters (e.g., with the help of chemometric modeling, computational fluid dynamics) |
| **Downstream processing** | |
| Cell separation and harvest, removal of particulate matters | Centrifugation (decanter, disk-stack separator, (semi)continuous centrifuges), filtration (dead-end, tangential flow filtration) |
| Cell rupture | High-pressure homogenization, bead mills, sonication, enzymatic treatments |
| Product capture | Filtration (micro-, ultra-, nano-filtration), precipitation, solvent extraction, ion exchange, size exclusion, affinity chromatography |
| Product purification/polishing | Hydrophobic interaction/reversed-phase chromatography |
| Clearance of contaminant agents (e.g., viruses, endotoxins) | Nano-filtration, heat, pH, chemical inactivation, ultraviolet, gamma irradiation |
| Drying | Heat, freezing, vacuum |
| **Galenic preparation/filling** | Addition of galenic excipients, supplements, stabilizers, and adjuvants |

*Notes*: Comprehensive overview illustrating the typical steps of a biotechnological production process. Production starts with the selection of an appropriate species and strain and the engineering of an expression system that enable high product yields and permit easy handling during strain development, fermentation, and downstream processing. Classic measures for improvement of the selected strains and expression systems are mutation and selection runs, breeding, directed genomic alterations, metabolic engineering, and measures aimed at the enhancement of expression rates of genes involved in product biosynthesis. Cultivation of the cells occurs under suitable conditions permitting strong growth, high product yield, and high product quality and facilitating product recovery and purification during downstream processing, which usually starts with separation of the product from the biomass and the culture broth by centrifugation or filtration. Before the separation of intracellular occurring products, cells have to be ruptured by high-pressure homogenization (e.g., in a homogenizer, a Dyno mill, or a French press). For high product purity, high recovery rates and simultaneous maintenance of the pharmacological activity, product purification and polishing usually occurs by solvent and solid-phase extraction. The choice of suitable resins and solvents and their combinations is highly dependent on the chemical and physical properties of the product and the side metabolites to be removed.

The final step consists of product concentration and drying, e.g., by heat, freezing, and/or vacuum before the galenic preparation. The discrimination between product capture and polishing is somewhat arbitrary as is the assignment of the respective methods. Depending on the product and the individual product properties, the transitions between the steps are gliding and the methods can change accordingly.

of yeasts in some cases however might reach a limit, particularly when the pharmacological activity of the product is impaired by the glycosylation pattern. In *K. lactis*, for instance, which usually does not hyperglycosylate, an exceptional high glycosylation of human interleukin 1β has been observed, reducing the biological activity to 5% [65, 66]. In such cases, a postsynthetic chemical modification has to be considered or the employment of higher developed organisms.

***Filamentous Fungi.*** Filamentous fungi are higher organized than yeasts and consequently have a more complex posttranslational modification apparatus more similar to mammals. Some proteins like t-PA in *Aspergillus nidulans* are produced with the natural human glycosylation pattern. For recombinant protein production, species are prefered, which are broadly employed in industry for production of enzymes, acids and antibiotics and thus possess GRAS-status: *A. nidulans*, *A. niger*, *A. sydowii*, *A. awamori*, various *Fusarium* and *Trichoderma* species, *Penicillium chrysogenum*, and *Acremonium chrysogenum*. Their productivity and secretion potential, which is in the range of 30–40 g/L for homologous enzymes like cellulases and amylases, is considered to be superior to any other system, but unfortunately could not be converted into corresponding yields of recombinant products, even when these were fused to such homologous enzymes. The highest yields are still obtained with heterologous fungal enzymes: 4-g/L *Fusarium* protease in *A. chrysogenum* and 4.6-g/L *A. niger* glucoamylase in *A. awamori*. The highest yield of a mammalian protein was reported for human interleukin 6 in *A. sojae* in a range of 300 mg/L [references in 16]. The yields of most human proteins like t-PA and various interferons, however, were reported to be below 1 mg/L [67–71]. Possible reasons for these incompetitive yields are restrictions in posttranslational metabolic steps like intracellular transport, folding, and processing. Even though filamentous fungi are industrially used now for decades, they are not adequately characterized on the physiological and genetic level. Little is known about details of the modification and secretion metabolism, and efficient gene transformation is hampered by degradation of foreign DNA, low transformation rate, and therefore, low copy numbers of the transferred genes and random genomic integration. Expression rates are restricted by a high RNA turnover, incorrect processing of the foreign messenger, and incomplete folding and secretion, which are, as in yeasts, both influenced by the glycosylation pattern. Proteins not completely or incorrectly glycosylated and excreted are rapidly degraded as shown for human interleukin 6 [71]. Currently, filamentous fungi cannot be regarded as a serious alternative for the production of pharmaceuticals. To fully exploit the potential of filamentous fungi, their physiology, particularly the glycosylation metabolism, thus has to be investigated and clarified in more detail.

***Insect and Mammalian Cell Cultures.*** Animal cell cultures are the systems with highest similarity to human cells with respect to the pattern and capacity of posttranslational modifications. However, their cultivation is more complicated and costly and usually yields lower product titers. Among the known systems, insect cells transformed by baculovirus vectors have reached a comparable popularity as *Pichia* among yeasts because they are considered to be more stress resistant, easier to handle, and more productive compared with mammalian systems and are thus frequently employed for high-throughput protein expression. The highest reported

yields refer to human collagenase IV in *Trichoplusia* (300 mg/L). Yields reported for non-insect cultures were 80 mg/L for human apolipoprotein A1 in chinese hamster ovary (CHO)-cells and 1 mg/L for human laminin in human embryonal kidney cells [references in 16]. For commercial application scale-up, related questions have to be clarified, particularly concerning oxygen supply [72] and carbon dioxide accumulation [73], stability of the cell line, and the bioreactor type to be employed [74]. Further deficiencies of insect cells are observed in (1) an inefficient processing and an impairment of the folding and secretion capacity due to the baculovirus infection [75]; (2) the high, in part baculovirus encoded, protease activity and the resulting necessity to routinely employ protease inhibitors in the culture media or to develop protease deficient vectors [73]; (3) an insufficient expression strength; and (4) deviations of the posttranslational modification pattern, which could act immunogenic. For optimization, the construction of new innovative vectors and the coexpression of chaperones, foldases, and folding factors such as canexin [76] have been suggested as has been the broader development and application of alternative systems like *Drosophila* [75]. Preferably applied in pharmaceutical production processes are mammalian systems like CHO- and baby hamster kidney (BHK)-cells (Table 1). These systems are generally considered to be genetically more stable and easier to transform and to handle in scale-up processes, grow faster in adherent and submerged cultures, and are more similar to human cells and more consistent in their complete spectrum of modification [77], which minimizes the risk of the formation of structurally altered compounds with immunogenic properties. In some cases, mammalian systems can be the only choice for the preparation of correctly modified proteins. Studies of Tate et al. [78] comparing the expression of a rat serotonin transporter in various expression systems demonstrate that biologically active transporter was only synthesized in mammalian cells, whereas it was partially degraded in *E. coli*, not correctly folded in *Pichia* and not correctly glycosylated in insect cells. CHO- and BHK-cells further have the advantage to be recognized as safe regarding infectious and pathogenic agents and therefore to have a higher acceptance by regulatory bodies, which accelerates or at least does not delay approval procedures. [Issues related to the physiology and the technical handling of mammalian cell cultures in the manufacture of human therapeutics are discussed in detail in Refs. 74 and 79–81].

### 2.3 Criteria for the Choice of Recombinant Expression Systems

In summary, the criteria for the choice of an expression system in pharmaceutical production [82] are the existing expertise, the available physiological and genetic know-how and tools, the patent situation, and to avoid delays of product launch and commercialization, regulatory aspects like the acceptance by the approving authorities.

The overall decision criteria, however, is the pharmacological activity profile of the yielded protein in context with the posttranslational modification pattern followed by rentability.

For production of non-glycosylated proteins and proteins that are natively glycosylated but pharmacologically active also without glycosylation, prokaryotes, which usually lack metabolic pathways for glycosylation, theoretically are the most

suitable organisms, offering two alternatives: Either *E. coli*-strains are conditioned to efficient secreters, or efficient native secreters like *Bacillus*-species are accordingly developed. (The implications of the glycosylation pattern for the choice of expression systems and the physiology of glycosylation is reviewed and discussed in Ref. 83.) To fully exploit the secretion capacity of fungal species, a deeper understanding of their posttranslational modification physiology will be necessary to steer the degree and pattern of glycosylation, which influences both folding and secretion efficiency. Insect and mammalian cells display posttranslational modification patterns more similar or identical to humans, but in view of the entailed expenditures, their employment can only be justified if their modification machinery is required to ensure a desired pharmacological activity.

*E. coli*, *P. pastoris*, and Baculovirus-based systems are currently preferred in fundamental research for structural and functional analysis of proteins and are employed as high-throughput-expression systems but will certainly find their ways into production processes in the near future. None of the systems, however, can be considered to be generally superior to any other. For each product, the most suitable expression system has to be identified and optimized individually both on the genetic and on the fermentative level by taking into account the properties of the product, the organism, and the expression cassettes, which suggests to always have a set of accordingly developed expression systems available. Some key properties of the established expression systems are compiled in Table 3. A review on properties and prospects of further alternative systems is given in Ref. 84.

Once an expression system has been selected, the respective strains are continuously submitted to improvement programs to render fermentation processes more efficient by increasing strain productivity and by modifying physiological properties and process behavior to enhance process economy. This does not only apply to the fermentation process, but also to parts of downstream processing, thus requiring an integrated strain selection and process design. Strategies to improve productivity of secretory recombinant systems and of strains employed for the manufacture of natural products will be outlined in the next section.

**TABLE 3. Features and Characteristics of Expression Systems**

| Property | Bacteria | Yeast | Insect and Mammalian Cell Cultures |
|---|---|---|---|
| Growth | Fast | Fast | Slow |
| Nutrient demand | Minimal | Minimal | Complex |
| Costs of media | Low | Low | High |
| Possible product yield | High | High | Low |
| Secretory capacity | Limtited | High | Medium |
| Glycosylation capacity | Limited | High | High |
| Modification capacity | Limited | High | High |
| Risk of retroviral Contamination | Low | Low | High |
| Risk of pyrogens | High | — | — |
| Scalability | Good | Good | Low |
| Process robustness | Good | Good | Low |

## 3 ENHANCEMENT OF PRODUCTIVITY

### 3.1 Secretory Recombinant Expression Systems

In terms of downstream processing efficiency, secretory expression systems indeed offer potential advantages for production of recombinant proteins compared with inclusion body-forming cytosolic systems, but most of the potentially available secretory systems are not yet fully competitive for high-volume therapeutics like insulin and therefore still require intensive improvement efforts.

Current strategies to improve productivity and secretion efficiency comprise (1) enhancement of gene expression rates, (2) optimization of secretion signal sequences, (3) coexpression of chaperones and foldases, (4) creation of protease deficient mutants to avoid premature product degradation, and (5) subsequent breeding and mutagenesis.

### 3.2 Natural Products

Completely different is the situation faced in the production of the so-called natural products, which are mostly secondary metabolites and therewith the endproducts of complex biosynthetic pathways of filamentous bacteria and fungi and thus require different approaches. With few exceptions like the antidiabetic drug acarbose (Glucobay) [5] or the chosterol-lowering drug lovastatin (Mevicor) [85], such fermented pharmaceuticals of natural origin currently belong for the most part to the class of antibiotics [86].

Enhancement of the productivity is achieved at first by rounds of random mutation and selection and, if possible, by breeding, and/or by DNA injection or fusion techniques like protoplasting and at further stages, after a comprehensive characterization of the usually complex biosynthetic pathways, directed genomic alterations and metabolic engineering.

These approaches, developments, and states in industrial strain improvement and pathway characterization will be illustrated for some of the best characterized organisms, namely the two fungal species mainly employed in industrial β-lactam antibiotic production, the penicillin producing *Peni-cillium chrysogenum*, and the cephalosporin C producer *Acremonium chrysogenum*.

***Characterization of Biosynthetic Pathways.*** The productivity of *Penicillium chrysogenum* could have been augmented impressively during the last decades: Penicillin titers have been increased by a factor of 50,000 from a very few milligrams/Liter in the 1940s to more than several 10 g/L by now and also some several 10 g/L cephalosporin are currently gained.

This development, however, has come to a halt during the last couple of years. As these titers have been reached through conventional mutation and selection [87–90] and no further significant increase in productivity could be noted, the tools of genetic engineering were more and more included into the improvement programs [91]. To gain a profound basis for directed genomic alterations, studies were conducted at various corporations and universities to elucidate the mechanisms and genes involved in β-lactam biosynthesis to characterize the respective biosynthetic pathways. As the results and achievements of these studies have already been

reviewed extensively and in detail elsewhere [92–97], the current knowledge will only shortly be summarized.

*Cephalosporin C Synthesis in Acremonium Chrysogenum.* Biosynthesis (Figure 1) starts with the polymerization of L-α-aminoadipic acid, L-cysteine, and L-valine to the linear tripeptide L-α-aminoadipyl-L-cysteinyl-D-valine (ACV-peptide). This reaction is catalyzed by the ACV-synthase (MW about 420 kD) through the following steps: (1) the ATP-dependent activation of these amino acids to bind them as thiolesters, (2) the epimerization of L-valine, and finally (3) the condensation by a thiotemplate mechanism [99].

Cyclization of the ACV-peptide to the bicyclic isopenicillin N (IPN) occurs under oxygen-, $Fe^{2+}$-, ascorbate-, and α-ketoglutarate-dependent action of the IPN-synthase (IPNS), which has an MW of about 38 kD. Inhibitory to IPNS activity are cobalt ions and glutathione.

Further pathway reactions are as follows:

- IPN-epimerization to penicillin N (IPN-epimerase).
- Penicillin N conversion to deacetoxycephalosporin C (DAOC) by expansion of the five-membered thiazolidine ring to a 6 C-dihydrothiazine-ring (DAOC-expandase).
- Formation of deactylcephalosporin C (DAC) by dehydroxylation and oxidation of the methylgroup in C3-position (DAC-hydroxylase).
- The acylation of DAC to cephalosporin C (DAC acetyltransferase).

DAOC-expandase and DAC-hydroxylase activities in *A. chrysogenum* are exerted by the same enzyme (MW 41 kD), which like the IPNS belongs to the group of α-ketoglutarate-dependent dioxygenases.

The first two enzymes, ACV- and IPN-synthase, are encoded by the *pcbAB*- and the *pcbC* gene, respectively. Both genes are linked to each other on chromosome VI by a 1.2-kb intergenic region carrying the putative promotor sequences from which they are divergently transcribed. The *pcbC*-promotor seems to be about five times stronger than the one of the *pcbAB*-gene [100]. The *cefEF*-gene and the *cefG*-gene encoding for the bifunctional expandase/hydroxylase and the DAC-acetyltransferase, respectively, are located adjacent to each other on chromosome II. Again the genes are separated by an intergenic region of 938 bp, which is supposed to harbor the promotors from which they are transcribed in opposite directions. The *cefG*-gene has been proven to contain two introns [101]. Due to an extreme *in vitro* lability, no information is yet available on the epimerase converting IPN to penicillin N as well as its putative *cefD* gene. Until recently, the structure of the *cefD* gene could have not been elucidated. Meanwhile, data could have been obtained indicating the existence of two reading frames (*cefD1* and *cefD2*) containing all characteristic motifs of mammalian acyl-CoA ligases and α-methyl-acyl-CoA racemases [93].

Regulatory studies demonstrated that the early functions are expressed simultaneously, whereas the later pathway genes *cefD* and *cefEF* seem to be induced sequentially.

In contrast, the penicillin biosynthesis genes in *P. chrysogenum* are considered to be expressed completely concomitantly.

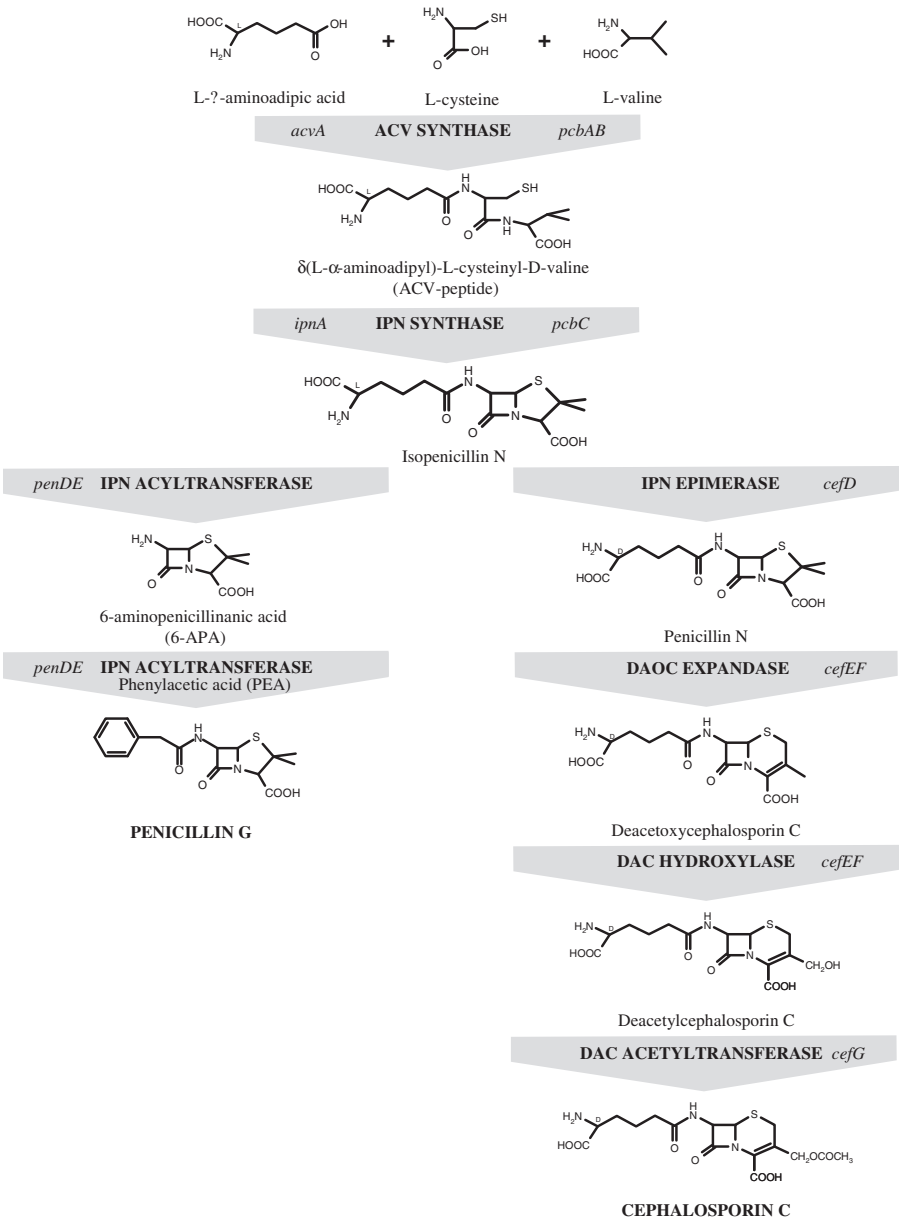


**Figure 1.** Biosynthetic pathways of penicillin G and cephalosporin C. In contrast to recombinant proteins, the structure and production of which can easily be modified by alterations of the encoding DNA sequence and the use of stronger promoters, respectively, structure modification and enhancement of expression rates of natural products by directed genomic alteration and metabolic engineering require a detailed characterization of the biosynthetic pathways as it has been performed for penicillin and cephalosporin biosynthesis in the filamentous fungi *Penicillium chrysogenum* (left) and *Cephalosporium acremonium* (right). Enzymes (bold) and genes are underlined by gray arrows. The first two reaction steps are common to both pathways. Cleavage of isopenicillin N to 6-APA and the subsequent acylation to penicllin G in *P. chrysogenum* and the expandase/hydroxylase activity in *A. chrsyogenum* are exerted by bifunctional enzymes. For details, see text. (Graph taken from Ref. 98, © Springer-Verlag, Heidelberg.)

*Penicillin Formation by Penicillium Chrysogenum.* The first reactions of the penicillin biosynthetic pathway are identical to the ones in *A. chrysogenum* (Figure 1). IPN, however, is not epimerized to penicillin N; instead it is converted to 6-aminopenicillanic acid (6-APA) by removal of the L-α-aminoadipic acid side chain, which is substituted by a hydrophobic acyl group. Both steps are catalyzed by the same enzyme, the acyl coenzyme A: IPN acyltransferase (IAT). The enzymatic activity of IAT is believed to be the result of the processing of a 40-kD monomeric precursor into a dimeric form consisting of two subunits with MWs of 11 and 29 kD. Due to the broad substrate specifity of IAT, various penicillin derivatives are synthesized naturally by attachment of different acyl-CoA derivatives to the 6-APA-core. For industrial purposes, to facilitate extraction by organic solvents, synthesis usually is directed to the less hydrophilic penicillin V or penicillin G. This is by addition of phenoxyacetic acid or phenylacetic acid, respectively, as precursors to the culture broth.

The IAT encoding *penDE*-gene (also named *aatA*) is located on chromosome I and organized as a cluster together with the ACV-synthase and IPN-synthase encoding genes *pcbAB* (also named *acvA* in *Penicillium*) and *pcbC* (*ipnA*).

The enzymes involved in penicillin biosynthesis are distributed at different sites of the cell: ACV-activity was found to be bound to vacuole membranes, IPN-synthase occurs dissolved in the cytoplasm and IAT-activity is microbody associated.

### 3.3 Approaches and Goals for Further Strain Improvement

***Analysis and Comparison of Strains.*** To get hints for more rational strain improvement approaches, (1) highly mutated production strains were genetically and physiologically compared with their less-productive ances-tors, (2) concentrations of pathway intermediates were determined to identify potential pathway bottlenecks, and (3) regulatory mechanisms were investigated.

In the course of such studies, high-performance strains of *P. chrysogenum* turned out to possess amplified copies of single genes like the *pcbC* gene or even copies of the whole cluster as well as increased steady-state transcript levels of pathway genes [102–104]. In some strains, the amplifications were shown to be organized in tandem repeats, which presumably were generated by a hot-spot TTTACA hexanukleotide [105]. Comparison of promotor strengths of these genes from high and low productive strains did not reveal any differences. This indicates the involvement of additional unknown trans-acting factors [106], as the amounts of increased mRNA did not correlate with the degree of gene amplification. Also a high specific activity of IPN synthase was reported in a more evolved *Penicillium* strain, which was independent from transcript amounts and probably due to a higher enzyme stability [102, 107].

In *A. chrysogenum*, no amplification of the relevant genes could be detected. Nevertheless, transcript amounts in production strains are significantly increased [108].

Also chromosome rearrangements could be detected in high titer strains [109–111], but their causative influence on productivity remains unclarified for the moment.

***Directed Genomic Alterations.*** *Enhancement of Gene Expression.* Con-sidering these findings, the experiments conducted so far to improve productivity mainly concentrated on enhancing the pathway gene dosages and on enhancing promotor strengths to remove presumed pathway bottlenecks. For instance, the *cefEF*-activity was generally believed to constitute a potential rate-limiting step in view of the high accumulation of penicillin N in various strains of *A. chrysogenum* and accordingly amplified.

Such approaches resulted in partially significant increases of productivity in low and medium titer strains: Improvements of the final yields by 50% have been reported [96].

However, all improvements achieved so far with high-performance production strains remained with a 5% maximum in the range of normal statistical deviations.

Also more exotic approaches like the (presumed) improvement of the intracellular oxygen supply through cloning and expressing of a bacterial hemoglobin gene from *Vitreoscilla* [112] failed to increase productivity in high titer strains.

The reasons for these failures are manifold:

1. As both direction and site of integration of the imported DNA cannot be controlled sufficiently and transformation efficiency is still quite low, the probability of finding new, higher producing mutants even in a large-scale screening is rare.
2. Most of the data available are from studies performed at academic institutions with original strains. Even when sharing the same ancestral strain, the high-performance strains employed in industries have an individual genealogy and mutation-selection history, so that their physiological behavior and properties differ drastically. Data and knowledge obtained with a particular strain thus cannot always be generalized and transfered to other strains. Even among industrial strains, significant differences have been revealed. For instance, the expression of the *cefG*-gene and the acetyl transferase activity was reported to constitute a possible rate-limiting step in Panlabs-strains [113]. Studies conducted on other industrial strains, however, could not confirm this observation.
3. The genetic instability and drift rises with the degree of genomic alterations, particularly during long-term vegetative propagation with numerous generation cycles in industrial large-scale fermentations.
4. The complex interdependence with other metabolic areas is not yet completely investigated. Consequently, after removal of an obvious β-lactam pathway bottleneck, unknown reaction steps of the linked and preceding pathways might become flux limiting. To overcome such flux limitations, one of the most exciting recent discoveries in molecular biology, namely RNA interference (RNAi), might play a major role in the future.

*The Possible Impact of RNAi in Strain Improvement.* The mechanisms of p osttranscriptional gene silencing, generally summed up as RNAi [114] offer potential approaches for the specific shut down of any gene of interest: Native genes can be silenced by triggering the RNAi cascade through introduction of homologous genetic sequences. As a result of these possibilities, RNAi has already

revolutionized fundamental research in molecular biology, particularly in functional genomics, and will substantially contribute to novel therapeutic approaches in medicine [115, 116]. For fungi, which evidently were among the first organisms in which RNAi and associated phenomena were observed [117–122], this RNAi triggered cosuppression phenomenon has been termed "quelling" and bears an enormous potential with respect to the study of biosynthetic pathways and the improvement of metabolic productivity.

On the one hand, the implication of RNAi may supply an explanation for observations made in the course of strain improvement programs of *Penicillium chrysogenum* and *Acremonium chrysogenum*, according to which antibiotic productivity can decrease drastically upon amplification of genes involved in antiobiotic biosynthesis.

The characterization and inactivation of these cosuppression mechanisms thus might help to overcome such limitations as does an application of fungal RNAi tools for silencing of pathway blocking or metabolic flux limiting genes, e.g., by the use of a novel vector system consisting of dsRNA viruses [114].

To fully realize the potential for further enhancement of productivity, which is generally assumed to be in the range of some 300% and to fully exploit the potential of RNAi, current and future studies thus aim at filling the actual gaps of knowledge and will be focused on

- The exploration of regulatory mechanisms and circuits on the transcriptional, translational, and posttranslational level.
- The determination of the specific intermediate turnover rates resulting from specific enzyme activities, enzyme titers, and enzyme stabilities.
- The detailed investigation of linked and preceding pathways.
- The investigation of product secretion mechanisms.

However, as no great leaps in productivity are currently reached, strains emerging from the development programs are also selected for beneficial alterations of genetic, physiological, and morphological properties that contribute to an enhanced process economy.

***Improvement of Process Behaviour.*** Among the desired alterations are:

- A decrease of side products, which (1) consume metabolic energy at the expense of the desired main product, (2) hamper the final product purification, or (3) even can be inhibitory to the production organism. For instance, a significant reduction of DCPC concentrations has been reached through knocking out esterases that hydrolyze cephalosporin C to DCPC during fermentation. Esterase inactivation was by conventional mutagenesis as well as through gene-disruption and by introduction of anti-sense genes.
- Increase of the tolerance toward toxic side products and fed precursors like phenyl- or phenoxyacetic acid (see above).
- Further reduction of feedback inhibition by endproducts and metabolites (see below).
- Further reduction of catabolite repression (see below).

- Enhancement of genetic stability.
- Enhancement of viability, stress resistance, and life span.
- Accelerated product formation and enhancement of time-specific productivity to shorten fermentation time.
- Enhancement of the strain-specific productivity to reduce the amount of biomass to be processed and disposed.
- Increase of product secretion rate.
- Improvement of the fermentation behavior like (1) higher efficiency of substrate/precursor consumption, (2) diminished oxygen demand, and (3) diminished shear sensitivity.
- Improvement of filterability.

The in-depth characterization of physiological properties and biosynthetic pathways not only aims at improving process behavior and process economy, but also it relates to modifications of the inherent biosynthetic pathways (pathway engineering) to create innovative, novel products with improved properties and enhanced therapeutic values, e.g., to overcome the antibiotic resistance problem, a major global health-care problem of today [86].

## 4 BIOSYNTHETIC STRUCTURE MODIFICATION

Although the metabolic engineering of beta-lactams by combining β-lactam encoding genes from various sources [95, 123] is still in an explorative state, the study of polyketide pathways and the creation of polyketide antibiotic derivatives is more advanced.

### 4.1 Combinatorial Biosynthesis

Compounds emerging from the polyketide pathways are the classic subject of metabolic engineering [124] and for creation of novel structures. To create novel antibiotics, the natural biodiversity of polyketides is trying to be enhanced artificially by modifying and newly combining the biosynthetic pathway steps and enzymes involved [125, 126]. Polyketide synthases [127, 128] are multienzyme systems, which due to their modular structure and the occurrence of their encoding genes as clusters, are considered to be most amenable for directed alterations, genetic intervention, and heterologous expression in foreign hosts with different metabolic pathways. Concepts and strategies that are currently being pursued for creating novel and un-natural polyketides [129–132] comprise (1) creating block mutants at various biosynthetic levels to accumulate intermediate metabolites; (2) elimination of unwanted groups; (3) modification of polyketide synthases to further broaden substrate range, eventually combined with a precursor-directed biosynthesis (see below); (4) formation of new structures by deleting or introducing additional modules; (5) modification, addition, or elimination of postsynthetic modification steps like glycosylation, lactonization, and amidation; (6) the cloning and combination of various polyketide synthase genes in heterologous hosts to construct hybrid pathways and antibiotics [130]; and (7) linking and combining routes of polyketides

with pathways of nonribosomal peptide synthesis to create hybrid peptide–polyketide compounds [131–135]. As the microbial nonribosomal peptide synthases share many structural and catalytic properties [136] with polyketide synthases like their organization into catalytic modules and domains [137], they are also envisaged as ideal subjects for metabolic engineering by creating large sets of modules and subsequently combining them artificially [95, 138, 139]. The rapidly growing body of patents makes evident that an increasing number of companies are working on the field of combinatorial biosynthesis. However, it also becomes evident that, despite their promising prospects, these strategies of combinatorial biosynthesis are still in an initial state and need further intense research to elucidate the required genetic and physiological details.

This is just in contrast to another approach of modifying the molecular structure of peptidic compounds on the biosynthetic level, namely the directed biosynthesis of the desired compound structures by feeding of appropriate biosynthetic precursors or stimulating agents during fermentation, as it is routinely performed for production of penicillin V und G by the addition of phenoxyacetic acid or phenylacetic acid, respectively, as precursors to facilitate penicillin extraction from the culture broth.

### 4.2 Precursor Directed Biosynthesis

Microbial peptidic compounds can be modified not only in their peptide structure by direct incorporation of amino acids supplemented during fermentation but also, if present, in their fatty acid moiety [140]. The possibilities of precursor-directed derivatization of cyclic peptides cores and fatty acid moieties by amino acids will be illustrated for three compound complexes (Figure 2) with unique modes of action and outstanding spectra of activity against multidrug resistant germs, which were developed by Aventis, Eli Lilly, and Cubist Pharmaceuticals. The complex A1437 of Aventis is synthesized by *Actinoplanes friulensis* [141], which originally produced a mixture of altogether eight lipopeptides (A–H), classified according to the exocyclic amino acid position and the type of their fatty acid chains [142] (Figure 2). Through addition of L-valine and asparagines, the biosynthesis could be directed toward the production of the desired D-peptide [143], which exhibited superior pharmacological properties in preclinical trials and is characterized by an exocyclic asparagine and an iso-C14 fatty acid moiety (Figure 2). The A54145 complex from Eli Lilly [144] exhibits the noteworthy characteristic, that a built-up resistance was immediately lost in the absence of this compound [145]. It consists of eight lipopeptides produced by *Streptomyces fradiae* containing four similar peptide nuclei in combination with three different fatty acid acyl side chains. The nuclei differ in their valine/isoleucine and glutamate/3-methylglutamate substitutions at one or both of two locations of the peptide ring. The composition of the peptide nucleus as well as the percentage of branched-chain fatty acid acyl substituents could be directed by supplementation of either valine or isoleucine [144] (Figure 2). Likewise, the complex A21978 from *Streptomyces roseosporus*, also developed at Eli Lilly, can be influenced in its composition by modification of the fatty acid chain. Precursing with valine results in an enhanced formation of the compound C2, whereas the compound C1 was preferably built upon feeding of isoleucine [146] (Figure 2). Compound C1 was inlicensed as

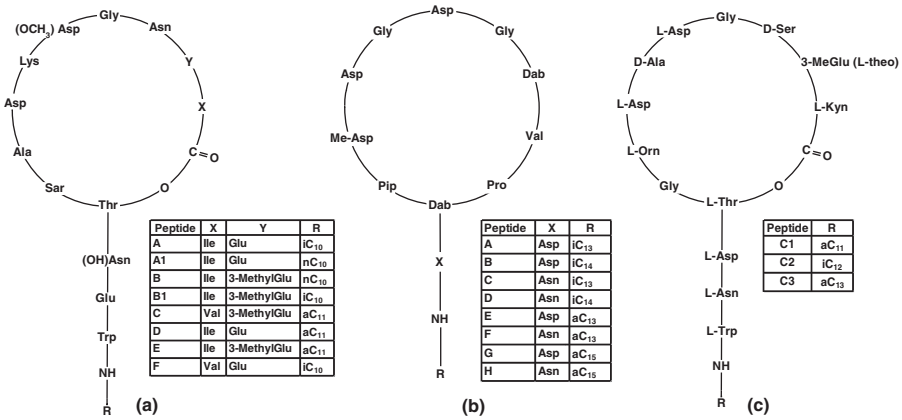


| Peptide | X | Y | R |
|---|---|---|---|
| A | Ile | Glu | $iC_{10}$ |
| A1 | Ile | Glu | $nC_{10}$ |
| B | Ile | 3-MethylGlu | $nC_{10}$ |
| B1 | Ile | 3-MethylGlu | $iC_{10}$ |
| C | Val | 3-MethylGlu | $aC_{11}$ |
| D | Ile | Glu | $aC_{11}$ |
| E | Ile | 3-MethylGlu | $aC_{11}$ |
| F | Val | Glu | $iC_{10}$ |

| Peptide | X | R |
|---|---|---|
| A | Asp | $iC_{13}$ |
| B | Asp | $iC_{14}$ |
| C | Asn | $iC_{13}$ |
| D | Asn | $iC_{14}$ |
| E | Asp | $aC_{13}$ |
| F | Asn | $aC_{13}$ |
| G | Asp | $aC_{15}$ |
| H | Asn | $aC_{15}$ |

| Peptide | R |
|---|---|
| C1 | $aC_{11}$ |
| C2 | $iC_{12}$ |
| C3 | $aC_{13}$ |

**Figure 2.** Structure modification of peptide antibiotics by precursor directed biosynthesis. Biosynthetic structure modification can be performed by metabolic engineering requiring an in-deph pathway characterization or by feeding of appropriate precursor compounds incorporated into the product. Examples are the lipopeptide complexes A 54145 from Eli Lilly (a), A 1437 from Aventis (b), and A 21978 developed by Eli Lilly and Cubist Pharmaceuticals (c) synthesized by *Streptomyces fradiae, Actinoplanes friulensis*, and *S. roseosporus*, respectively. The composition of both the peptide cores as well as the fatty acid moieties can be modulated *in vivo* by supplementation of amino acids like valine, leucine, isoleucine, and asparagine. The addition of valine leads to the preferable formation of peptides A 54145 C and F. The biosynthetic pathways of *A. friulensis* could be directed toward a more specific production of the desired peptide A 1437 D by feeding of valine and asparagines, whereas the preferred compound A 21978 C1 (daptomycin) could be stimulated by the addition of isoleucine. For the further details and references, see text. (Graph taken from Ref. 84, © Springer-Verlag, Heidelberg.)

daptomycin by Cubist Pharmaceuticals for clinical development. However, due to rhabdomyolytic activities and a slight haemolytic potential [138, 147, 148], it has been approved by the U.S. Food and Drug Administration in 2003 as Cubicin® salve for topical use only. Despite such disadvantageous properties, peptides in general, including those that are from nonmicrobial sources and ribosomally synthesized, are currently considered to be among the most promising compound classes for development of novel antibacterials [86, 149, 150].

However, it has to be taken into account that amino acid and precursor incorporation as well as regulation mechanisms of biosynthetic pathways also depend on growth phases, on environmental influences, and thus on the conduction of the fermentation process during which the cells are propagated.

A prerequisite for a successful fermentation are controlled and homogenous environmental conditions for the cellular reactions. However, the ensurance of controlled reaction conditions, particularly in large production scales, has been turned out to be one of the most critical issues of industrial biotechnology, as will be illustrated in the next section. For a detailed discussion of problems specifically related to the fermentation of β-lactam antibiotics, the reader is referred to the review by Schmidt [98].

## 5 FERMENTATION OPTIMIZATION AND SCALE-UP

The scale-up of fermentation processes as a central problem in biotechnology was first recognized and described during industrial penicillin production at the beginning of the 1940s and has been studied in more detail in *E. coli* and recombinant protein production.

### 5.1 Reduced Mixing Quality and Enhanced Stress Exposure

Fermentation scale-up is aimed at the manufacture of larger product quantities, if at all possible, with a simultaneous increase or at least consistency of specific yields and product quality. The changed geometric and physical conditions in larger scales, however, has lead to a less favorable mixing behavior and to impaired physiological reaction conditions, which in turn may lead to a decreased process constance and reproducibility, to reduced specific yields, to an increase of unwanted side products, and thus ultimately, to a diminished batch-to-batch consistency and product quality, which are all key issues in industrial production processes.

The problem of reduced mixing quality in larger scales is aggravated with increasing vessel sizes: The opposite substrate and oxygen gradients along the vessel height that are formed as a result of the conventional fermenter design, according to which substrate feed usually occurs from the top and aeration from the bottom, are more pronounced in larger reactors due to (1) longer distances to be covered leading to larger substrate and oxygen depletion zones, (2) larger volumes of culture broth to be stirred and therewith longer mixing times, and (3) stronger hydraulic pressure gradients influencing the oxygen transfer rate (OTR; Equations 14 and 15, Table 4) [151–155]. Cells at the fermenter top are exposed to excess glucose concentrations and simultaneously suffer from oxygen limitations, whereas those at the bottom are exposed to glucose starvation. Excess glucose concentrations (threshold value for *E. coli* at 30 mg/L) result in acetate overproduction (overflow-metabolism), and a simultaneous oxygen limitation further induces the formation of ethanol, hydrogen, formiate, lactate, and succinate (mixed acid fermentation: Refs. 154, 156, and 157). The produced acids can become reassimilated in oxygen-rich zones but in any case first lead to a temporary acidification of the microenvironment and later eventually of larger regions. Combined with a decreased transportation and elimination of carbon dioxide, detrimental metabolites, and surplus heat generated by agitation and metabolic processes, resulting in zonal overheating, the lower mixing rates in large scales thus lead to the formation of zones with enhanced stress conditions. The subsequent activation of stress genes (a survey on bacterial stress proteins is given in Ref. 158) only partially protects the cell against detrimental stress effects. For instance, despite activation of heat stress genes like *E. coli* dnaK and clpB, reducing misfolding and aggregation of heat-sensitive proteins [159], metabolic changes and damages such as translocation of proteins and membranes have been observed upon heat exposure [160]. The evidently unavoidable detrimental effects of the repeated and cyclic passing of different stress zones and the subsequent continuous activation and shut down of the corresponding stress genes are believed to lead to a completely altered physiology [155, 161] with metabolic shifts, which ultimately reduce growth and productivity and increase byproduct formation.

**TABLE 4. Fermentation parameters, coefficients and terms, implicated in mixing, aeration, oxygen, and heat transfer, suitable as scale up variables to be kept constant alone or combined with each other or other process relevant variables, preferably but not necessarily, as dimensionless groups as described and discussed e.g., by Hubbard [228] and Wang and Cooney [230]. For instance, the $k_La$—value as the most frequently applied physical scale up variable has been combined with [1] the vessel backpressure *p* by maintaining the product of $k_La \times p$ constant through variation of pressure, agitation, and aeration for scale up of *Bacillus thuringiensis* fermentations (Flores et al. [232]) or [2] the aeration rate *vvm* under variation of power input, working volume and aeration according to the Wang-Cooney-equation [217]. For details and further explanations see text (taken from Ref 273 © Springer-Verlag, Heidelberg).**

| Parameter / coefficient | Mathematical characterization | Symbol explanation |
|---|---|---|
| Power input (P); volumetric power input (P/V) | (1) $P = 2\,\pi\, n\, M = N_{Po}\, \rho\, n^3\, d_I^5$ [kg m$^2$ s$^{-2}$ = W] | P = power input; n = stirrer speed; M = momentum; $N_{Po}$ = dimensionless power number (impeller specific); |
| Dimensionless power number ($N_{Po}$) | (2) $N_{Po} = P / \rho\, n^3\, d_I^5$ | $\rho$ = density of the medium; n = stirrer speed; $d_I$ = impeller diameter |
| Impeller tip speed ($v_{tip}$) | (3) $v_{tip} = 2\,\pi\, n\, d_I$ [m / sec] | n = stirrer speed; $d_I$ = impeller diameter |
| Reynolds number (Re) | (4) $Re = n\, d_I^2\, \rho / \eta$ | Re = Reynolds number; n = stirrer speed; $d_I$ = impeller diameter; $\rho$ = density of the medium; $\eta$ = dynamic viscosity |
| Mixing time ($T_m$) | (5) $T_m = f\,(n, d_I, \upsilon) = V / Q = V / N_{fl}\, n\, d_I^3$ [sec] | $T_m$ = mixing time; n = stirrer speed; $d_I$ = impeller diameter; $\upsilon$ = kinematic viscosity; V = volume; Q = volumetric flow rate; $N_{fl}$ = pumping number |
| Dimensionless mixing time ($T_m n$) | (6) $T_m\, n = f\,(Re, N_{Po})$ | $T_m$ = mixing time; n = stirrer speed; $T_m\, n$ = dimensionless mixing time; Re = Reynolds number $N_{Po}$ = dimensionless power number |
| Modified dimensionless power number | (7) $N'_{Po} = N_{Po}\, Re^3\, d_F / d_I = P\, d_F\, \rho^2 / \eta^3$ | $N'_{Po}$ = modified dimensionless power number $T'_m\, n$ = modified dimensionless mixing time |
| Modified dimensionless mixing time | (8) $T'_m\, n = n\, T_m / [(d_F / d_I)^2\, Re] = T_m\, \eta / d_F^2\, \rho$ | $N_{Po}$ = dimensionless power number; Re = Reynolds number; $d_F$ = inside vessel diameter; $d_I$ = impeller diameter; P = power input; n = stirrer speed; $t_m$ = mixing time; $\rho$ = fluid density; $\eta$ = dynamic viscosity |

**TABLE 4.** *Continued*

| Parameter / coefficient | Mathematical characterisation | Symbol explanation |
|---|---|---|
| Aeration rate (volume per volume per minute, vvm) | (9) $A_R = F_G / V_R$ [$m^3/ m^3 min$] | $A_R$ = aeration rate; $F_G$ = volumetric gas flow rate $V_R$ = fermenter reaction volume |
| Superficial gas velocity ($v_s$) | (10) $v_s = F_G / A$ [m / sec] | $v_s$ = superficial gas velocity $F_G$ = volumetric gas flow rate A = fermenter cross section |
| Gas hold up ($\tau$) | (11) $\tau = V_R / F_G$ | $\tau$ = gas hold up; $V_R$ = fermenter reaction volume; $F_G$ = volumetric gas flow rate; $\varepsilon$ = fractional gas hold up; $V_G$ = dispersed gas volume; |
| Fractional gas hold up ($\varepsilon$) | (12) $\varepsilon = V_G / V_R$ | |
| Gassing number ($N_{Qg}$) | (13) $N_{Qg} = f\,(Fr, Re, S, (\rho_g / \rho_0), (\eta_g / \eta_0)) = q_g / n\, d_I^3$ | $N_{Qg}$ = gassing number; Fr = Froude number; Re = Reynolds number; S = fluid constant $\rho$ = density; $\eta$ = dynamic viscosity $_g$ = aerated; $_0$ = not aerated; $q_{g=}$ gas throughput n = stirrer speed; $d_I$ = impeller diameter |
| Oxygen transfer rate (OTR) | (14) $OTR = k_La\,(C_G - C_L) = k_La\,L_{O2}\,(pO_{2G} - pO_{2L})$ [kg $O_2$ /$m^3$h] with (15) $C_G = 0{,}526\, p_i / 36 + T$ [mg / l] | OTR = oxygen transfer rate from gas to liquid phase $k_L$ = mass transfer coefficient; a = specific interfacial surface area; $C_G$ = oxygen saturation concentration in the gas phase; $C_L$ = measured oxygen saturation concentration in the liquid phase; $L_{O2}$ = oxygen solubility in the liquid phase; $pO_{2G}$ = partial pressure of oxygen in the gas phase; $pO_{2L}$ = partial pressure of oxygen in the liquid phase; $p_i$ = vessel back pressure [bar]; T = temperature [°C] |
| Volumetric oxygen mass transfer coefficient ($k_La$) | (16) $k_La = a'\,(P / V_R)^b\, v_s^c$ [$s^{-1}$] | $k_La$ = oxygen transfer coefficient; P = power input; $V_R$ = fermenter reaction volume; $v_s$ = superficial gas velocity; a′, b and c as specific fluid constants to be determined experimentally |

| | | |
|---|---|---|
| $k_La$ scale dependent according to Wang-Cooney | (17) $k_La = k'(P_g/V_R)^a(v_s)^b(B/6)^{0,8}(j/d_I)^{0,3}$ [$s^{-1}$] | $k'$, a, b = vessel specific coefficients; $P_g$ = power input in aerated vessel; V = volume of culture broth; $V_s$ =superficial gas velocity; B = number of stirrers j = baffle width; $d_I$ = impeller diameter |
| Respiratory quotient (RQ) | (18) $RQ = CER / OUR = pCO_{2E} - pCO_{2i} / pO_{2i} - pO_{2E}$ [mol $CO_2$ / mol $O_2$] | RQ = respiratory quotient; CER = carbon dioxide emission rate; OUR = oxygen uptake rate; $pCO_2$ = partial pressure of carbon dioxide; $pO_2$ = partial pressure of oxygen; $_i$ = air feed; $_E$ = exhaust air |
| Energy balance ($Q_{ges}$) | (19) $Q_t = Q_{met} + Q_{ag} + Q_{aer} - Q_{evap} - Q_{hxch}$ [kWh = kJ] with<br>(20) $Q_{met} = Y_{ho} R_o$ [kWh = kJ] | $Q_t$ = total energy; $Q_{met}$ = energy generated by metabolic activities; $Q_{ag}$ = energy generated by agitation; $Q_{aer}$ = energy generated by aeration; $Q_{evap}$ = energy losses through evaporation; $Q_{hxch}$ = energy losses through cooling; $Y_{ho}$ = approximate metabolic constant (460 kJ/mol $O_2$); $R_o$ = molar oxygen uptake rate (mol $O_2$ / s) |
| Heat transfer coefficient | (21) $\alpha = (dQ/dt) / A\ dT$ | $\alpha$ = heat transfer coefficient; dQ/dt = heat transfer; A = surface area; dT = temperature difference |

***Altered physiology.*** Changes in respiratory states like the switch from aerobic to anaerobic conditions and the imposed stress conditions, increased by induction and overproduction of the recombinant protein, lead to shifts in the coupled amino acid biosynthesis, an altered composition of the amino acid pools and changes in the protein biosynthesis machinery (162, 162a, b, c). Even though translation processes are considered to be generally performed precisely [163], the resulting shifts in the amino acid pool in turn lead, due to the inherent inaccuracy of the translation machinery (unspecificity of aminoacyl-tRNA-synthases, "wobbling" of tRNA), to aminoacid misincorporations in freshly synthesized proteins, particularly in phases of high protein production rates, which are characterized by amino acid shortages. Muramatsu et al. [164] report the incorporation of β-methylnorleucin instead of isoleucin into recombinant *E. coli* hirudin in positions 29 and 59. A misincorporation of norvalin instead of leucin and methionine into human recombinant hemoglobin in *E. coli* has been published by Apostol et al. [165] and Kiick et al. [166]. Amino acid misincorporations are thus a cause for an increase of byproducts at the expense of the desired main product yields.

***Reduced Plasmid Stability.*** An essential prerequisite for high product yields particularly in larger scales, in which cultures pass a higher number of generations due to larger culture broth volumes and longer inoculation chains from the cell bank to the production stage, is the stable propagation of plasmids to daughter cells. Plasmid stability is influenced by the plasmid properties, including size and nucleotide sequence; by the genetic background of the host [167, 168]; as well as by process parameters like temperature, growth rates, and substrate concentrations. Lin and Neubauer [169] show that rapid glucose oscillations favor plasmid stability and recombinant protein production rate, whereas high glucose concentrations diminish plasmid stability [170]. Plasmid stability and plasmid numbers are thus negatively influenced, more difficult to control, and less easy to be maintained in larger scales. A concept to render plasmid stability and expression rates more independent from such physiological influences is the application of runaway-plasmids, the replication of which can be induced separately from growth in the desired fermentation phase [171, 172], enabling copy numbers of up to 1000 in the expression phase and, due to the separation of their replication and expression from the growth phase, simultaneously a more precise replication and amino acid incorporation. Further possible measures to enhance the accuracy of metabolic processes are a reduction of the generation time and numbers throughout the process and a deceleration of the metabolic speed, e.g., through reduction of temperature, and the development of more stress-resistant microbial strains.

A prioritized goal of process optimization and scale-up thus consists of an appropriate process design that improves the physiological conditions and the metabolic accuracy by minimizing microbial stress exposure.

### 5.2 Process Characterization

To identify process-specific stress factors and to understand the physiological responses to the vessel-specific physical conditions, the mutual influences and interactions of the various physical and physiological parameters have to be analyzed in detail. An overview on the analytical methods currently applied and yet in development is given in the following.

***Analytical Methods.*** Among the physiologically most relevant parameters are, besides the pH as discussed, biomass, cell viability, the concentrations of substrates, metabolites and products, the partial oxygen and carbon dioxide pressures in culture broth, and the composition of the exhaust gas, giving information about respiratory states as indicated, e.g., by the respiratory quotient (RQ; Equation 18, Table 2). Established analytical methods and tools are colorimetric procedures, chromatography [173], mass spectroscopy [174, 175], enzymatic and electrophoretic methods [176–178], hybridizing techniques, biochips [155, 161, 179], or flow cytometry. Multiparameter flow cytometry permits, in addition to the analysis of usual metabolites, also the analysis of the cellular DNA-, RNA-, and protein content, cell viability, membrane potential, intracellular pH, cell size, and cell development stages [180–183]. In general, the data are, if possible, preferably captured as real-time values by in situ online measurements as occurring for physical parameters [184] to avoid falsified or gappy data due to time differences between sampling and analysis. *Ex situ* methods can be designed as real-time procedures in combination with robust automated sampling and sample preparation methods like ultra-filtration [173, 175, 185] and flow injection analysis (FIA) [173, 186–190]. An online FIA glucose analysis method of the native culture broth without prior sampling and filtration has been presented by Arndt and Hitzmann [191] and Kleist et al. [192]. For their robustness, *in situ* measurements of biological parameters are preferably performed by optical probes (optodes), working either on a physical base, e.g., by refractional or spectroscopic measurements, chemically as by ion exchange reactions [193] or enzymatically by gel-embedded biocatalysts, the activity of which can be measured by means of pH-changes or by fluorophores (chemo- and biosensors) [194, 195]. The hitherto employed enzymes have restricted their application mostly to the analysis of sugar compounds [190], whereby the most frequent applications are reported for glucose [187]. However, despite intensive research, only a few concepts practically applicable for industrial purposes could have been developed. Broader applications are physical optodes enabling the simultaneous measurement and quantification of several parameters and metabolites through continuous scans or excitations and subsequent absorption and scattering measurements within defined wavelength ranges like (1) 2-D-fluorescence spectroscopy [196–198] as shown for the determination of NAD(P)H concentrations through excitation at 350 nm and the fluorescence measurements at 450 nm [190]; (2) near-infrared-spectroscopy (NIR) [199–201], which has been employed as a 700–2500-nm scan in the antibiotic fermentation at Pfizer [202, 203] for the measurement of nutrient and product as well as side product concentrations; or (3) measurements in infrared [204], ultraviolet, or visible wavelength spectra [190]. By means of capacitance measuring electrodes (Aber Instruments Ltd., U.K.), viable cell counts can be performed by seizing the electric capacitance and conductivity of cells with intact membranes in a generated electric field. To gain a picture as comprehensive as possible by enhancing the amount of analyzable substances, multichannel arrays with parallel employment of various analytical procedures monitoring several parameters simultaneously have been established and new concepts like artificial noses [205] and electronic tongues [206] currently being integrated into fermentation technology, allowing the analysis of a large number of analytes by cross-talking semi-specific sensors that act in analogy to the human olphactory sensoric system. The realization of this concept will be facilitated by a

further miniaturization of the sensors and chips employed [207] as it has been achieved with biochips developed for the pharmaceutical natural product screening, onto which the needed number of reactants is fixed in smallest space or is brought together in microchannels and the analytes are separated in microcapillars (lab-on-a-chip-technology) [208–210].

### 5.3 Process Optimization

For analysis, interpretation and correlation of the obtained signals and data through chemometric modeling [195, 197], neural network—tools (e.g., Unscrambler 9.0, Camo, Norway)—seem to have gained an increasingly important role [211]. Despite inherent limitations [212, 213], adequately trained neural networks are capable of recognizing and revealing nonlinear, highly complex, and even nonobvious, hidden relations among a multitude of various physical, biochemical, and physiological process parameters and to design the according process models. Hooked up to expert systems, they allow immediate and short-term reactions to process deviations by anticipating physiological drifts and their influences on product yield and quality and to change the according set points and process profiles. In this way, neural network tools help to optimize process control and to facilitate the identification of suitable strategies for process optimization and scale-up, e.g., with respect to an optimized glucose–oxygen equilibrium.

As the maximum glucose and oxygen uptake rates are not constant but depend on growth phases and rates and get reduced by induction of product synthesis [170], Lin et al. [214] established an integrative kinetic model combining these parameters with the aid of the simulation program SCILAB (Inria, F) enabling the determination of the maximum glucose and oxygen uptake rates through determination of the time that passes between glucose pulses and subsequent changes of partial oxygen pressure ($pO_2$) and biomass concentration. Such a model thus allows the alignment of the glucose feed meeting the maximum uptake capacity and therefore avoiding overflow metabolism. The observed pattern of glucose oscillation even can serve as a key parameter condition in later scale-up stages [169]. The equilibration of carbohydrate and oxygen supply thus constitutes a key component of process control and optimization. Further strategies aiming at the avoidance of oxygen limitation and glucose overflow metabolism consist of the employment of alternative carbon sources like glycerol [215] or galactose [216] and of a drastic glucose restriction. Wong et al. [217] report that minimal concentrations of glucose and yeast extract yielded the highest concentrations of 0.5 and 1 g/L of a recombinant K99 antigen. Yim et al. [218] achieved constant yields of 4.4 g/L of human granulocyte colony stimulating factor during scale-up from 2.5 to 30 L by limiting glucose feed and keeping the growth rate at the minimum ($\mu = 0.116\,h^{-1}$). Besides the above-mentioned parameters like the maximum uptake capacities, the various different physiological or directly measurable physical process parameters like pH, ammonium consumption, $pO_2$, OTR, or the growth parameters can be applied as a reference parameter for glucose feed in a closed-loop design. Dantigny et al. [219] describe a biomass controlling feed depending on the ethanol formation rate and RQ for *S. cerevisiae*. To achieve a more uniform glucose distribution in large-scale tanks, Larsson et al. [151] and Bylund et al. [152] suggested a glucose feed at the dynamic zones of the fermenter bottom together with the injected air. A step ahead

in this regard would be a further enhancement of the mixing quality by a reactor and process design permitting a multilevel injection of both air and substrates into high turbulence zones.

A supportive approach for comprehension of the large-scale hydrodynamic and reaction conditions to ensure homogenous reaction conditions and to reduce both the size of stress zones and the zonal residence times is the depiction of these zones through transfer into small reactors (scale-down) or through high-performance computing computational fluid dynamics (CFD) [220]. Whereas the sort and amount of parameters that can be simulated by the scale-down approach appear to be restricted, CFD is meanwhile broadly applied, e.g., the flow modeling software tools of Fluent Inc., Lebanon, NH (Fluent 4,5,6 and MixSim). A simulation of the trajectories and distributions of gas bubbles and mass transports enables the determination of zonal $pO_2$ values and oxygen transfer rates. The combination with parameters like substrate concentrations and gradients, the residence times in the respective zones, population dynamics, and metabolic fluxes (structured metabolic models) leads to integrative models (integrated fluid dynamics, IFD), which permit the prediction of physiological effects and reactions. Even though it is emphasized that the integration of CFD and structured biokinetics currently requires a deeper understanding of the dynamics of metabolic and regulatory networks and cascades of signal transduction triggered by microenvironmental fluctuations and thus further research [221], it will most likely contribute long term to a realistic modeling of the interplay between physics and physiology and thus will facilitate the identification of key parameters influencing product yield and quality the most and therewith of suitable scale-up parameters and strategies. Problems related to the scale-up of physical parameters are illustrated in the next section.

### 5.4 Physical Scale-Up Parameters

Suitable for employment as physical scale-up parameters are all known process parameters and coefficients exerting known physiological effects, particularly those affecting oxygen supply; heat transportation and mixing, such as power input (Equation 1; Table 4), aeration, and agitation rate (Equations 3, 9, and 10; [222]); mixing time (Equation 5; [223]); $pO_2$; OTR (Equations 14, 15; [224, 225]); oxygen mass transfer coefficient ($k_La$, Equations 16, 17); and biocalorimetric variables like heat fluxes and heat transfer coefficients (Equations 19, 20, and 21). Biocalorimetric measurements were found to be in definite correlation with metabolic activities and with OTR and are employable as growth phase indicators [226, 227]; the calculations, however, require a precise measurement and knowledge of heat fluxes and sources, energy inputs, and the temperature distribution in the vessel. In most cases, however, it is, depending on the scale-up factor, not, or only restrictedly, possible from the physical viewpoint to keep physical parameters constant throughout the scale-up. Constance of even a single specific parameter mostly leads to an uncontrolled and unpredictable change of other variables into dimensions that are technically not realizable. Classic examples are (1) the mixing time, which inevitably increases in larger vessels due to the larger volumes to be stirred and that cannot be unlimitedly compensated and kept constant by increasing stirrer speeds and energy inputs, and (2) the volumetric energy input P/V. A constant volumetric power input has indeed been successfully applied as a scale-up

parameter for the early industrial penicillin fermentations (1 hp per gallon, equivalent to 1.8 kW per 1 $m^3$) and in fermentations with low energy inputs [216], but it is limited in fermentations requiring high energy inputs, like recombinant *E. coli* cultures.

For this reason, mathematically driven approaches are pursued for both process and reactor scale-up by forming dimensionless coefficients (e.g., by the Buckingham-Pi method as cited by Hubbard [228]), which are kept constant by an appropriate choice and adjustment of the relevant influencing process and fermenter parameters. Well-established examples are, among many others, the dimensionless power number (Equation 2), Reynolds number (Equation 4), gassing number (Equation 13), and the modified dimensionless power number (Equation 7) and the modified dimensionless mixing time (Equation 8). The comparison of the latter two coefficients and of the corresponding curves enables the identification of appropriate types of stirrers capable of exerting the desired mixing performance at a given stirrer speed with a minimum of energy consumption and therewith to compensate the above-mentioned limitations of volumetric energy inputs at larger scales. (A quantum leap in stirrer technology in this regard seems to be the Visco-Jet of Inotec, which possesses cone-formed short tubes instead of blades and pretends to exhibit the best mixing performance with a minimum of power input and shear stress without foam generation.) The performance of further various types and designs of stirrers is presented and discussed by Junker et al. [229]. Stirrer performance coefficients and curves are thus crucial scale-up aids.

Furthermore, any of these parameters and coefficients can be combined with other variables to set up and create new, process-specific parameter correlations, coefficients, terms, groups, and characteristic curves (examples given, e.g., by Wang and Cooney [230]).

This relates also to the $k_La$—value (Equation 16), which currently is the most applied physical scale-up variable because it includes the relevant parameters influencing oxygen supply like agitation (via energy input) and aeration as superficial gas velocity (Equation 10) and that, as a component of dimensionless terms, is also frequently employed for reactor scale-up [231].

Taking into account the limitation of a $k_La$–oriented scale-up in the form of a limited power input, OTR, and tolerable shear stress, Flores et al. integrated the backpressure $p$ by maintaining the product of $k_La \times p$ constant through a variation of pressure, agitation, and aeration for scale-up of *Bacillus thuringiensis* fermentations [232]. Although the specific growth was reduced and the biomass production and sporulation efficiency remained constant, fermentation time could be shortened and toxin yields were increased. Wong et al. successfully scaled-up *E. coli* fermentations from 5 L to 200 L by keeping constant the product of $k_La$ and aeration rate (vvm) under variation of power input, working volume, and aeration according to the Wang–Cooney equation (Equation 17; [230]), which reflects these parameters in dependence of the fermentation scale [217]. Diaz and Acevedo [224] argue that the oxygen transfer capacity, indicated by the $k_La$ value, is not the most process-relevant parameter, but the effective oxygen transfer rate as a product of $k_La$ and mass transfer potential is, i.e., the difference between the $pO_2$-values in the gas and the liquid phase, and suggest an OTR-based scale-up strategy (Equations 14 and 15).

A common, simple and robust method is the maintenance of a constant $pO_2$ in the culture broth by variation of stirrer speed and aeration rate. To avoid shear

stress and to keep the energy input (P) at the lowest possible level, the $pO_2$ is steered at the minimum limit. Riesenberg et al. employ a stirrer speed and glucose-feed steered $pO_2$ of 20% for interferon α–*E. coli*-fermentations in 30-L and 450-L scales [233]. These examples demonstrate that, for each process, appropriate variables and coefficients and therewith scale-up strategies have to be identified individually.

### 5.5 Development of Fermentation Models and Strategies

Despite the central role of the scale-up issue in biotechnology and the comparably large body of literature, no common, generally applicable strategy seems to be established. For each product, process, and facility, a suitable scale-up strategy has to be elaborated.

A wholistic scale-up strategy consists of a comprehensive and detailed process characterization to identify key stress factors and key parameters influencing product yield and quality the most, and of an appropriate process control and process design ensuring optimum mixing and reaction conditions, supported by appropriate knowledge and data-driven models [234, 235] as well as computational tools.

It should be kept in mind, however, that any approach and any model will always be approximative and that a compromise in the process-related knowledge will always be gappy and that the known mathematical methods and relations cannot completely reflect the highly complex interactions and relationships of the physical conditions governing the fermentation process [236].

As a matter of fact, it seems, that, in view of the high complexicity of the fermentation parameters influencing each other and the only rudimentary and fragmentary reflection of the reality any model can deliver and finally the different layouts of vessels and facilities that rarely are designed according to strict scale-up criteria, successful scale-up in most cases will not be the result of a conclusive and straight-lined experimental strategy but will be the outcome of an independent optimization on each scale that highly depends on the experience, skill, and last, but not least, intuition of the experimentalist.

To a lesser extent, this also holds true for the design of procedures aimed at the isolation and purification of the compounds from the culture broth. As these purification steps follow upon the biosynthetic steps and the upstream processing in the bioreactor, they are usually referred as "downstream processing."

## 6 DOWNSTREAM PROCESSING

### 6.1 Product Recovery and Purification

Biopharmaceutical products consist of a multitude of compounds and structures with most having different physical and chemical properties and derive from a large variety of sources like human and animal tissues, body fluids, plant material [237], and as illustrated above, microbial fermentations. Accordingly, purification strategies have to be developed individually and empirically that reflect the physicochemical properties of the product, of the product source, and of potential contaminants by finding appropriate sorts, sequences, combinations, and operation

modes of the respective downstream processing steps to finally achieve high purities and high recovery rates while maintaining the pharmaceutical activity of the molecule.

Traditionally, downstream processing steps are roughly subgrouped into operations related to cell harvest, cell rupture in case of intracellularly occurring products, product capture, product purification, and product polishing to manufacture a drug ready for galenic preparation and consist of a sequence of solid–liquid separation and solvent and solid-phase extraction steps. Large-scale cell harvest usually occurs by preparative centrifugation, e.g., in a decanter or in disk stack separator or by filtration to simultaneously separate existing suspended particles from the surrounding broth. Fragile mammalian cells are usually separated by filtration methods only.

If the value product occurs intracellularly, the slurry is taken up in washing buffers of appropriate ionic strength for preparation of the cell rupture, which occurs, e.g., by high-pressure homogenization, sonification, bead milling, or enymatic procedures.

Product capture is defined as the first product extraction and purification step aimed at a product concentration by volume reduction and partial purification, whereas polishing constitutes the final purification step, which removes persistent residual minor impurities like denatured, aggregated, or nonfunctional isoforms of the product, and which follows the preceding (intermediate) purification steps aimed at the removal of contaminating solutes like host cell proteins, DNA, and media components. Product capture usually starts with precipitation, solvent extraction, and/or a cascade of ultra-filtration and nano-filtration steps before the first ion exchange or size exclusion chromatography. Antibodies and antibody derivatives, which account for about 20% of the biopharmaceutical products currently in development, are preferably bound to affinity matrices [238] by natural immunoglobulin-binding ligands such as the *Staphylococcus aureus* membrane and cell wall protein A [239] or the streptococci surface protein G. To broaden the availability of specific structures with suitable binding properties, synthetic ligands with enhanced stability and resistance to chemical and biochemical degradation are continuously attempting to be developed, supported, e.g., by computer-aided design. The development of synthetic ligands with enhanced selectivity and stability, tailored to specific biotechnological needs and product structures, will lead to more efficient, less expensive, and safer procedures not only for purification of antibodies but also for other proteins at manufacturing scales. Also, antibodies can be employed as a means for protein purification in the form of immunoaffinity chromatography [240]. A novel emerging technology is the use of synthetic single-stranded nucleic acid molecules (aptamers) as affinity ligands [241], which fold up into unique three-dimensional structures specifically binding to the desired target structure as it has been shown for the purification of thyroid transcription factor or selectin receptor globulin [242, 243].

The subsequent purification steps including polishing are mostly performed on hydrophobic interaction resins and/or on reversed-phase matrices.

### 6.2 Downstream Processing Optimization and Economization

Mostly due to the tremendous costs of sometimes several thousand € per liter for chromatographic resins and production scales of up to several hundred liters column

volume and due to the prices for the columns themselves of up to hundreds of thousands of U.S. dollars, downstream processing expenses account for the largest part (50–90%) of the total production costs and are therewith a focal area for process economization attempts. One of the many criteria for the choice of a resin is its shelf life. Bearing in mind that every downstream processing step coincides with a minimum product loss of 5–10%, the removal of a particular step not only helps to save the respective capital investment, but concomittantly to increase the final product yields, and thus to render the whole production process more economical in several respects. To reduce the number of required steps, a selection has to be made for resins exhibiting excellent separation, resolution, and yield with respect to a given product and the contaminants, i.e., for resins exhibiting excellent product binding selectivity and capacity. Likewise, the operation mode of the column has to be optimized with respect to the specific product load, adjustment of the physico-chemical operational conditions like pressure, temperature, pH(-shift), ionic strength, and sort of buffers and eluents (e.g., evaluation of continuous vs. step gradients).

A further approach is the combination of the solid–liquid separation and product recovery into a single step, e.g., by expanded bed adsorption [244, 245], which permits the direct initial purification from culture broths without the need of prior removal of suspended solids. In contrast to the conventional column operation, the mobile phase is pumped upward through the column bed from beneath. The bed thus starts to expand at a liquid flow rate above a critical value, and the gaps between the sorbent beads expand. The controlled distribution of bead size and bead weight results in a stable expanded bed in which the beads oscillate around a steady position and thus avoid clogging of the column. Whereas preparative liquid chromatography methods [246] have become a well-established separation and purification method in the pharmaceutical industry, techniques like two-phase system partitioning [247] and its variant reverse micellar extraction [248], which also potentially offer advantages like the possibility of a direct extraction from the culture broth through partitioning of the compounds into two immiscible phases, need further developmental activities for economic large-scale applications in the production of biopharmaceuticals and are less common.

A further criterion for the choice and design of a downstream processing procedure is thus its scalability.

### 6.3 Downstream Processing Scale-Up

Comparable with the scale-up of fermenters and fermentation processes, the scale-up of downstream processes, particularly of column and column operation, is not simply a matter of increasing size. Increasing column diameters, reducing the stabilizing wall effects, and increasing resin bed heights may lead to an altered settling behavior of the beads, to channel formation through shrinking and swelling of the packed matrix, to column clogging with a concomittant reduction of the flow rates and an increase of the column back pressure, to altered hydrodynamic behavior and residence times of the process fluids, and to an altered pattern for the product/contaminants adsorption/desorption pattern [249]. To avoid the resulting quality impairments and deviations, resins with a suitable scale-up behavior have to be selected and the resin bed height is tried to be kept constant as far as possible as

is the flow rate as key scale-up parameters, which requires an appropriate experimental design in early laboratory development stages. As this, however, is only possible to a limited extent, downstream processing scale-up also is the result of an optimization of the respective operation conditions (see above) on each scale.

A procedure that has been optimized for decades and scaled-up to dimensions of several cubed meters is the commercial β-lactam purification.

(For further general articles dealing with problems and issues of downstream processing, the reader is referred to the reviews in Refs. 249–255).

### 6.4 Downstream Processing of β-Lactam Compounds

As most of the product is secreted and is thus concentrated in the culture broth, the recovery process starts with filtration, usually with a rotary vacuum filter, followed by a cascade of solvent and solid-phase extraction steps.

Common penicillin extraction solvents are amyl- and butyl-acetate or methylisobutylketone. As penicillin is extracted as a free acid at pH 2–2.5 and the molecule is instable at this pH, extraction occurs in a counter-flow using Podbielniak- and Luwesta-centrifugal extractors to shorten the contact time with the solvent and to prevent product decomposition.

Purification is through subsequent and repeated crystallization from an aqueous solution after alkalization.

In contrast to penicillin, the hydrophilic cephalosporin is more suitable for solid-phase extraction. For high yield and purity extraction, a combination of several, different chromatographic steps is used.

Hydrophobic interaction chromatography on neutral polyaromatic resins like Amberlite XAD 4, 16, 1180, or Diaion HP20 is widely used in combination with weak basic anion exchangers like Diaion WA-30 or strong acidic cation exchangers like Amberlite XAD 2000 and Diaion SK-1B [256, 257]. The mentioned resins are recognized for their high sorption capacity and their long shelf life.

Whereas the improvements on the extraction level are less spectacular, significant progress over the last few years can be noticed with respect to side-chain cleavage to 6-aminopenicillanic acid (6-APA) and 7-aminocephalosporanic acid (7-ACA) in switching from chemical to enzymatic procedures.

## 7 POSTSYNTHETIC STRUCTURE MODIFICATION

Most antibiotics in therapeutic use are synthesized or modified exclusively by the means of chemistry and are derived from the established compound classes, which have been known for decades. Chemical derivatization thus has also been applied to substances that are of microbial origin and that therefore are termed semi-synthetic.

Among the most recent semi-synthetic antimicrobials are Aventis' streptogramin derivative Synercid and the erythromycin derivatives clarithromycin (Klacid) from Abbott (Abbott Park, Illinois) and roxithromycin (Rulid) and the brand new ketolide telithromycin (Ketek), both from Aventis (Paris, France) [86].

Despite these recent successes and increasing efforts, the yields of therapeutically useful entities emerging from such chemical derivation programs are continuously decreasing.

For this reason, large-scale derivation methods (combinatorial chemistry) are integrated in the search for innovative antibiotics as are biotechnological approaches for structure modification. Biocatalytic procedures, which have been shown to be useful for generating novel antibiotic structures like Loracarbef, a novel β-lactam compound from Eli Lilly (Indianapolis, Indiana) highly active against various β-lactamase producing species [258, 259], are also about to replace continuously chemical production processes. At several universities and corporations, screening studies have been initiated to find appropriate enzymes for compound conversions, as it has been successfully established in production processes for penicillin and cephalosporin. (A review discussing criteria like resource and energy consumption, emissions, health risk potential, area use, and environmental effects for the evaluation of the eco-efficiency of biotechnological processes is given by Saling [260]).

### 7.1 β-Lactam Side-Chain Cleavage

6-APA and 7-ACA as the key intermediates for the production of semisynthetic penicillins and cephalosporins (see Section 3.2) are obtained by removal of the acyl side chains.

There is a large body of patents existing for chemical and enzymatic splitting procedures, but enzymatic processes have been more successful for economical as well as for ecological reasons.

Enzymatic 7-ACA splitting procedures [for general review, see 261] have been developed and commercialized by companies like Asahi Chemical, Hoechst, and Novartis. The replacement of the hitherto employed chemical deacylation processes like the imino ether (Figure 3) or the nitrosyl chloride method [262] resulted in a cost reduction of 80% and a decrease of the waste volume by a factor 100 from 31 t to 0.3 tons per 1-ton 7-ACA. Chlorinated hydrocarbons like dimethyl aniline and methylene cloride as well as heavy metal ions can be completely avoided. Instead of zinc salt formation, multiple silylation, formation of the imino chloride, imino ether, and finally an imino ether hydrolysis, the side chain is removed in two enzymatic steps (Figure 3).

Cephalosporin C is first oxidized and deaminated by a D-amino acid oxidase (DAO), which can be obtained from various fungal species, like the yeasts *Trigonopsis variabilis* and *Rhodotorula gracilis* or the ascomycete *Fusarium solani*. The resulting α-keto-adipyl-7-ACA, upon decarboxylation, is converted into glutaryl-7-ACA (G-7-ACA). DAO is a flavoenzyme containing flavin adenine dinucleotide as the prosthetic group and catalyzes oxidation of D-amino acids to their corresponding keto acids. In a second step, the glutaryl side chain of G-7-ACA is deacylated by a glutarylamidase from *Pseudomonas diminuta* [263]. The molecular data of other potentially suitable enzymes and genes from various sources are given by Isogai [264]. It is noteworthy that the enzymatic splitting process could have only been rendered economical and therefore commercially employable through a significant increase of glutarylamidase yield on the fermentation level by using a gene-recombinant *E. coli*-strain.

A recombinant amidase from *E. coli* is also the most commonly employed enzyme for 6-APA production by deacylation of penicillin G [265], a process that has been established now for decades [266]. With an annual turnover of 30 tons, the *E. coli* penicillin amidase is one of the most widely used biocatalysts, despite

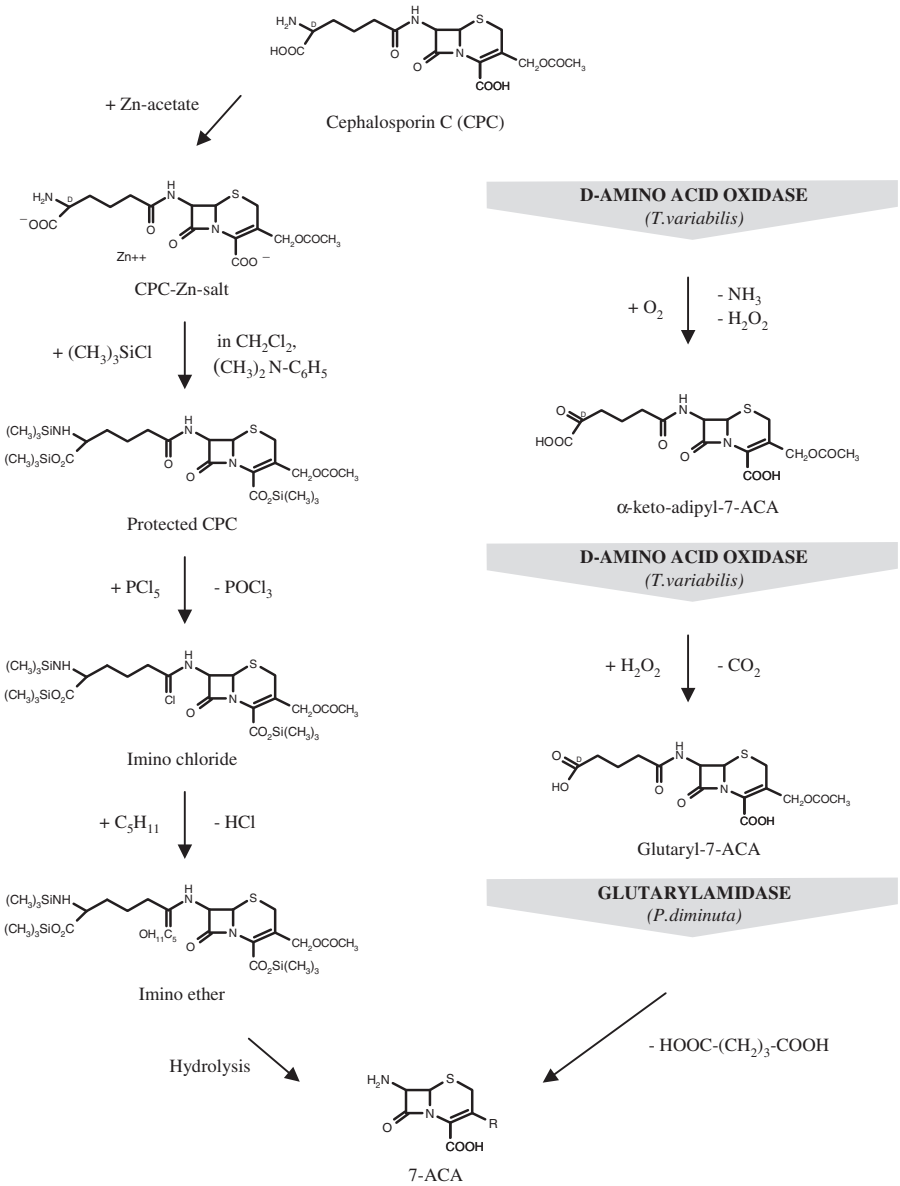


**Figure 3.** Replacement of the chemical cephalosporin C splitting procedure by a biotechnological process. Biotechnological production and derivatization procedures are continuously being developed to replace chemical processes for both economical and ecological reasons. An established industrial process is the enzymatic splitting of cephalosporin C to 7-ACA by D-amino acid oxidase of the yeast *Trigonopsis variabilis* and a glutarylamidase of *Pseudomonas diminuta* (right) replacing the chemical imino ether method (left). For further details, see text. (Graph taken from Ref. 98, © Springer-Verlag.)

the discovery of further similar acting enzymes from various microbial sources, including penicillin V acylases [267] from the basidiomycetes *Bovista plumbea* [268, 269] and *Pleurotus ostreatus* [270]; from the ascomycete *Fusarium* [271]; or from the yeast *Cryptococcus*.

A significant milestone in economization of enzymatic β-lactam production has been reached by enzyme immobilization permitting a preservation and multiple use of the cleavage enzymes. The currently most employed resins in industries for this purpose are epoxyacrylic acids (Eupergit) and silica gel derivatives (Deloxan). The general economic criterion for the preparation of a biocatalyst is production costs in relation to yield, turnover rate, storage stability, and operational/mechanical stability. Further criteria are outlined in the *Guidelines for the Characterization of Immobilized Biocatalysts* by the European Federation of Biotechnology. Application criteria are filterability, sedimentation velocity, and particle firmness. Operational and storage stability as well as the activity of this biocatalyst and the quality of 6-APA with respect to color are improved by sulfur reducing compounds [272]. As pointed out, the ensurance of product quality and drug safety is a key issue in pharmaceutical production processes.

## 8 QUALITY ISSUES

According to regulatory requirements, pharmaceutical production facilities and processes have to be proven to function properly across the entire range of process critical parameters by a qualification of the facilities and equipment, which consists of the steps design qualification (DQ), installation qualification (IQ), operational qualification (OQ), and performance qualification (PQ) and which results in a validation of the production process to substantiate that the complete system is capable of manufacturing the respective product reproducibly in a consistent quality as expected within preset specifications and in compliance with all laws and guidelines, like the cGMP (current good manufacturing practice) regulations, among others, issued by regulating authorities like the FDA and European Medicines Agency (EMEA).

The products have to be sterile and free of any contaminations that might derive from process material, residues of preceding production campaigns (cross-contaminations) in the case of product changes, and of the agents used for facility cleaning. To restrict the introduction of such contaminations into the repective production steps, the production facilities, equipments, and solutions are usually cleaned according to validated cleaning procedures (e.g., with 1 M NaOH) and sterilized by heat or filtration before their use. Product sterility is achieved by aseptical performance of the last manufacturing steps and by micro-filtration (0.2 μm) or, if possible, by terminal heating of the product before filling. Residues of chemical cleaning agents are removed according to validated procedures by rinsing with purified water.

As biopharmaceuticals are potentially contaminanted with possibly harmful viruses and allergenic and pyrogenic acting endotoxins as residual components of their host cell, guidelines and directions have been issued by the FDA's Center for Biologics Evaluation and Research (CBER) requesting validated procedures for virus and endotoxin removal and inactivation.

### 8.1 Virus and Endotoxin Removal

Endotoxins, which consist of the lipopolysaccharide fraction present in the cell wall of gram negative germs and which tend to adhere to equipment surfaces and to persist in products and product solutions, as do viral contaminants deriving from cell culture processes or from starting plasma or tissue material, can also be efficiently removed by final filtration steps. Whereas ultra-filtration with a cut-off of or below 10 kDa is sufficient for endotoxin removal, viruses need to be removed by nano-filtration steps complemented by a variety of inactivation measures like pasteurization, pH inactivation, solvent/detergent treatment, or ultraviolet and gamma ray irradiation. Usually, endotoxins and viruses are already deconcentrated to a significant extent by the chromatographic steps and the solvent exposure in the course of the normal purification procedures [251].

## 9 CONCLUSION

Marching in step with the technical and methodological progresses, biotechnology is gaining in increasing importance in pharmaceutical production processes by replacing chemical production procedures for economical and ecological reasons and in the development and commercialization of novel therapeutic principles.

To fully exploit the potential of biotechnological production methods, an integrated process design will be necessary considering the downstream processing requirement during the design of upstream operations and vice versa (A comprehensive overview illustrating the complete biotechnological production chain is given in Table 3.) As the fate of a compound as a pharmaceutical also depends on the prospective production costs already estimated in the development phase and production processes cannot be significantly changed after approval without running the risk of having to perform ad-ditional clinical trials, it is of utmost importance to choose and design production organisms, strains, vectors, expression cassettes, and fermentation procedures as early as possible in compliance with the prospective harvest and purification methods also to be chosen and designed according to cost criteria. Integrative biotechnology thus is not only a key prerequisite for the development of competitive and economical production processes to relieve the downward price pressure applied by generic drug makers after patent expiration and for the acceleration of market approval for new drugs in the highly competitive environment of the biopharmaceutical industry, but also it is a key technology to save and liberate capital for the development of novel drugs for the benefit of mankind.

## REFERENCES

1. Choi J H, Lee S Y (2004). Secretory and extracellular production of recombinant proteins using *Escherichia coli*. *Appl. Microbiol. Biotechnol.* (online).
2. Li Y, Wei G, Chen J (2004). Glutathione: a review on biotechnological production. *Appl. Microbiol. Biotechnol.* 66:233–242.
3. Leathers T D (2003). Biotechnological production and application of pullulan. *Appl. Microbiol. Biotechnol.* 62:468–473.

4. Hermann T (2003). Industrial production of amino acids by coryneform bacteria. *J. Biotechnol.* 104:155–172.
5. Wehmeier U F, Piepersberg W (2004). Biotechnology and molecular biology of the alpha-glucosidase inhibitor acarbose. *Appl. Microbial. Biotechnol.* 63:613–625.
6. Pulz O, Gross W (2004). Valuable products from biotechnology of microalgae. *Appl. Microbiol. Biotechnol.* 65:635–648.
7. Lee P C, Schmidt-Dannert C (2002). Metabolic engineering towards biotechnological production of carotenoids in microorganisms. *Appl. Microbiol. Biotechnol.* 60:1–11.
8. Westers L, Westers H, Quax W J (2004). Bacillus subtilis as cell factory for pharmaceutical proteins: A biotechnological approach to optimize the host organism. *Biochim. Biophys. Acta.* 1694:299–310.
9. Miesfeld R (1999). *Applied Molecular Genetics*, Wiley, New York.
10. Weaver R F (1994). *Molecular Biology of the Cell*, Garland Publishing, New York.
11. Faye L, Boulaflous A, Benchabane M, Gomord V, Michaud D (2005). Protein modifications in the plant secretory pathway: Current status and practical implications in molecular pharming. *Vaccine* 23:1770–1778.
12. Hellwig S, Drossard J, Twyman R M, Fischer R (2004). Plant cell cultures for the production of recombinant proteins. *Nat. Biotechnol.* 22:1415–1422.
13. Goldstein D A, Thomas J A (2004). Biopharmaceuticals derived from genetically modified plants. *QJM* 97:705–716.
14. Gregory G, Heyneker H L (1988). Secretion of correctly processed human growth hormone in *E. coli* and *Pseudomonas*. U.S. Patent 4755465.
15. Habermann P, Bender R (2001). Signal sequences for the production of leu-hirudin via secretion by *E. coli* in a culture medium. International Patent WO 01/21662.
16. Schmidt F R (2004a). Recombinant expression systems in the pharmaceutical industry. *Appl. Microbiol. Biotechnol.* 65(4):363–372.
17. Choi J H, Ryu Y W, Seo J H (2005). Biotechnological production and applications of coenzyme $Q_{10}$. *Appl. Microbiol. Biotechnol.* (epub ahead of print).
18. Weickert M J, Doherty D H, Best E A, Olins P O (1996). Optimization of heterologous protein production in *Escherichia coli*. *Curr. Opin. Biotechnol.* 7:494–499.
19. Makrides S C (1996). Strategies for achieving high-level expression of genes in *Escherichia coli*. *Microbiol. Rev.* 60:512–538.
20. Gentschev I, Dietrich G, Goebel W (2002). The *E. coli* alpha-hemolysin secretion system and its use in vaccine development. *Trends Microbiol.* 10:39–45.
21. Ward A, Hoyle C, Palmer S, O'Reilly J, Griffith J, Pos M, Morrison S, Poolman B, Gwynne M, Henderson P (2001). Prokaryote multidrug efflux proteins of the major facilitator superfamily: Amplified expression, purification and characterization. *J. Mol. Biotechnol.* 3:193–200.
22. Pines O, Inouye M (1999). Expression and secretion of proteins in *E. coli*. *Mol. Biotechnol.* 12:25–34.
23. Bassford P J, Silhavy T J, Beckwith J R (1979). Use of gene fusion to study secretion of maltose-binding protein into *Escherichia coli* periplasm. *J. Bact.* 139:19–31.
24. Yamanaka H, Nomura T, Fujii Y, Okamoto K (1998). Need for TolC, an *Escherichia coli* outer membrane protein, in the secretion of heat-stable enterotoxin I across the outer membrane. *Microb. Pathogen.* 25:111–120.
25. Lund P A (2001). Microbial molecular chaperones. *Adv. Microb. Physiol.* 44:93–140.
26. Hartl F U (1996). Molecular chaperones in cellular protein folding. *Nature* 381:571–580.

27. Collier D N (1993). SecB: A molecular chaperone of *Escherichia coli* protein secretion pathway. *Adv. Protein. Chem.* 44:151–193.
28. Kujau M J, Hoischen C, Riesenberg D, Gumpert J (1998). Expression and secretion of functional miniantibodies McPC603scFvDhlx in cell-wall-less L-form strains of *Proteus mirabilis* and *Escherichia coli*: A comparison of the synthesis capacities of L-form strains with an *E. coli* producer strain. *Appl. Microbiol. Biotechnol.* 49:51–58.
29. Rippmann J F, Klein M, Hoischen C, Brocks B, Rettig W J, Gumpert J, Pfizenmaier K, Mattes R, Moosmayer D (1998). Procaryotic expression of single-chain variable-fragment (scFv) antibodies: Secretion in L-form cells of *Proteus mirabilis* leads to active product and overcomes the limitations of periplasmic expression in *Escherichia coli*. *Appl. Environ. Microbiol.* 64:4862–4869.
30. Mergulhao F J, Summers D K, Monteiro G H (2005). Recombinant protein secretion in *Escherichia coli*. *Biotechnol. Adv.* 23:177–202.
31. Swartz J R (2001). Advances in *Escherichia coli* production of therapeutic proteins. *Curr. Opin. Biotechnol.* 12:195–201.
32. Sorensen H P, Mortensen K K (2005). Soluble expression of recombinant proteins in the cytoplasm of *Escherichia coli*. *Microb. Cell. Fact.* 4:1–8.
33. Dilsen S, Paul W, Sandgathe A, Tippe D, Freudl R, Thömmes J, Kula M R, Takors R, Wandrey C, Weuster-Botz D (2000). Fed-batch production of recombinant human calcitonin precursor fusion protein using *Staphylococcus carnosus* as an expression-secretion system. *Appl. Microbiol. Biotechnol.* 54:361–369.
34. Olmos-Soto J, Contreras-Flores R (2003). Genetic system to overproduce and secrete proinsulin in *Bacillus subtilis*. *Appl. Microbiol. Biotechnol.* 62:369–373.
35. Srinivasan S, Barnard G C, Gerngross T U (2002). A novel high-cell-density protein expression system based on *Ralstonia eutropha*. *Appl. Env. Microbiol.* 68:5925–5932.
36. Nakashima N, Mitani Y, Tamura T (2005). Actinomyctes as host cells for production of recombinant proteins. *Microb. Cell. Fact.* 4:7.
37. Pierce J J, Robinson S C, Ward J M, Keshavarz-Moore E, Dunnill P (2002). A comparison of the process issues in expressing the same recombinant enzyme periplasmatically in *Escherichia coli* and extracellularly in *Streptomyces lividans*. *J. Biotechnol.* 92:205–215.
38. Anné J, van Mellaert L (1993). *Streptomyces lividans* as host for heterologous protein production. *FEMS Microbiol. Lett.* 114:121–128.
39. Paradkar A, Trefzer A, Chakraburtty R, Stassi D (2003). *Streptomyces* genetics: a genomic perspective. *Crit. Rev. Biotechnol.* 23:1–27.
40. Berks B C, Sargent F, Palmer T (2000). The Tat protein export pathway. *Molec. Microbiol.* 35:260–274.
41. Teter S A, Klionsky D J (1999). How to get a folded protein across a membrane. *Trends Cell. Biol.* 9:428–431.
42. Schaerlaekens K, van Mellaert L, Lammertyn E, Geukens N, Anne J (2004). The importance of the Tat-dependent protein secretion pathway in Streptomyces as revealed by phenotypic changes in tat deletion mutants and genome analysis. *Microbiol.* 150:21–31.
43. Upreti R K, Kumar M, Shankar V (2003). Bacterial glycoproteins: functions, biosynthesis and applications. *Proteomics* 3:363–379.
44. Benz I, Schmidt A (2002). Never say never again: Protein glycosylation in pathogenic bacteria. *Molec. Microbiol.* 45:267–276.

45. Russel C, Mawson J, Yu P L (1991). Production of recombinant products in yeasts: A review. *Aust. J. Biotechnol.* 5:48–55.
46. Lin Cereghino J L, Cregg J M (2000). Heterologous protein expression in the methylotrophic yeast *Pichia pastoris*. *FEMS Microbiol. Rev.* 24:45–66.
47. Lin Cereghino G P, Lin Cereghino J, Ilgen C, Cregg J M (2002). Production of recombinant proteins in fermenter cultures of the yeast *Pichia pastoris*. *Curr. Opin. Biotechnol.* 13:329–332.
48. Buckholz R G, Gleeson M A (1991). Yeast systems for the commcercial production of heterologous proteins. *Bio/technol.* 9:1067–1072.
49. Gellissen G (2000). Heterologous protein production in methylotrophic yeasts. *Appl. Microbiol. Biotechnol.* 54:741–750.
50. Gellissen G, Weydemann U, Strasser A W M, Piontek M, Janowicz Z A, Hollenberg C P (1992). Progress in developing methylotrophic yeasts as expression systems. *Trends Biotechnol.* 10:413–417.
51. Cregg J M, Lin Cereghino J L, Shi J, Higgins D R (2000). Recombinant protein expression in *Pichia pastoris*. *Mol. Biotechnol.* 16:23–52.
52. Swinkels B W, van Ooyen A J J, Bonekamp F L (1993). The yeast *Kluyveromyces lactis* as an efficient host for heterologous gene expression. *Antonie van Leeuwenhoek* 64:187–201.
53. Schaffrath R, Breuning K D (2000). Genetics and molecular physiology of the yeast *Kluyveromyces lactis*. *Fungal Genet. Biol.* 30:173–190.
54. Gellissen G, Hollenberg C P (1997). Application of yeasts in gene expression studies: A comparison of *Saccharomyces cerevisiae*, *Hansenula polymorpha* and *Kluyveromyces lactis*—a review. *Gene* 190:87–97.
55. Guisez Y, Tison B, Vandekerckhove J, Demolder J, Bauw G, Haegeman G, Fiers W, Contreras R (1991). Production and purification of recombinant human interleukin-6 secreted by the yeast *Saccharomyces cerevisiae*. *Eur. J. Biochem.* 198:217–222.
56. Kjeldsen T (2000). Yeast secretory expression of insulin precursors. *Appl. Microbiol. Biotechnol.* 54:277–286.
57. Kjeldsen T, FrostPettersson A, Hach M (1999). Secretory expression and characterization of insulin in *Pichia pastoris*. *Biotechnol. Appl. Biochem.* 29:79–86.
58. Chow T Y, Ash J J, Dignard D, Thomas D Y (1992). Screening and identification of a gene, PSE-1, that affects protein secretion in *Saccharomyces cerevisiae*. *J. Cell Sci.* 101:709–719.
59. Turner B G, Avgerinos G C, Melnick L M, Moir D T (1991). Optimization of pro-urokinase secretion from recombinant *Saccharomyces cerevisiae*. *Biotechnol. Bioengin.* 37:869–875.
60. Sleep D, Belfield G P, Ballance D J, Steven J, Jones S, Evans L R, Moir P D, Goodey A R (1991). *Saccharomyces cerevisiae* strains that overexpress heterologous proteins. *Bio/technol.* 9:183–187.
61. Fleer R (1992). Engineering yeast for high level expression. *Curr. Opin. Biotechnol.* 3:486–496.
62. Poletti A, Weigel N L, McDonnell D P, Schrader W T, O'Malley B W, Conneely O M (1992). A novel, highly regulated, rapidly inducible system for the expression of chicken progesterone receptor, $cPR_A$, in *Saccharomyces cerevisiae*. *Gene* 114:51–58.
63. Purvis I J, Chotai D, Dykes C W, Lubahn D B, French F S, Wilson E M, Hobden A N (1991). An androgen-inducible expression system for *Saccharomyces cerevisiae*. *Gene* 106:35–42.
64. Romanos M A, Makoff A J, Fairweather N F, Beesley K M, Slater D E, Rayment F B, Payne M M, Clare J J (1991). Expression of tetanus toxin fragment C in yeast: Gene

synthesis is required to eliminate fortuitous polyadenylation sites in AT-rich DNA. *Nucl. Ac. Res.* 19:1461–1467.

65. Fleer R, Chen X J, Amellal N, Yeh P, Fournier A, Guinet F, Gault N, Faucher D, Folliard F, Fukuhara H, Mayaux J F (1991a). High-level secretion of correctly processed recombinant human interleukin-1β in *Kluyveromyces lactis*. *Gene* 107:285–295.
66. Fleer R, Yeh P, Amellal N, Maury I, Fournier A, Bacchetta F, Baduel P, Jung G, L'Hôte H, Becquart J (1991b). Stable multicopy vectors for high level secretion of recombinant human serum albumin by *Kluyveromyces lactis*. *Bio/technol.* 9:968–975.
67. Jeenes D J, Mackenzie D A, Roberts I N, Archer D B (1991). Heterologous protein production by filamentous fungi. *Biotechnol. Genet. Eng. Rev.* 90:327–367.
68. Radzio R, Kück U (1997). Synthesis of biotechnologically relevant heterologous proteins in filamentous fungi. *Proc. Biochem.* 32:529–539.
69. Gouka R J, Punt P J, van den Hondel C (1997). Efficient production of secreted proteins by *Aspergillus*: Progress limitations and prospects. *Appl. Microbiol. Biotechnol.* 47:1–11.
70. Conesa A, Punt P J, van Luijk N, van den Hondel C (2001). The secretion pathway in filamentous fungi: A biotechnological view. *Fungal Genet. Biol.* 33:155–171.
71. Punt P J, van Biezen N, Conesa A, Albers A, Mangnus J, van den Hondel C (2002). Filamentous fungi as cell factories for heterologous protein production. *Trends Biotechnol.* 20:200–206.
72. Reiter M, Blüml G (1994). Large-scale mammalian cell culture. *Curr. Opin. Biotechnol.* 5:175–179.
73. Ikonomou L, Schneider Y J, Agathos S N (2003). Insect culture for industrial production of recombinant proteins. *Appl. Microbiol. Biotechnol.* 62:1–20.
74. Birch J R, Froud S J (1994). Mammalian cell culture system for recombinant protein production. *Biologicals* 22:127–133.
75. McCarroll L, King L A (1997). Stable insect cultures for recombinant protein production. *Curr. Opin. Biotechnol.* 8:590–594.
76. Betenbaugh M J, Ailor E, Whiteley E, Hinderliter P, Hsu T A (1996). Chaperone and foldase coexpression in the baculovirus-insect cell expression system. *Cytotechnology* 20:149–159.
77. Gervais A, Hammel Y A, Pelloux S, Lepage P, Baer G, Carte N, Sorokine O, Strub J M, Koerner R, Leize E, Van Dorsselaer A (2003). Glycosylation of human recombinant gonadotropins: Characterization and batch-to-batch consistency. *Glycobiology* 13:179–189.
78. Tate C G, Haase J, Baker C, Boorsma M, Magnani F, Vallis Y, Williams D C (2003). Comparison of seven different heterologous protein expression systems for the production of the serotonin transporter. *Biochim. Biophys. Acta.* 1610:141–153.
79. Grampp G E, Sambanis A, Stephanopoulos G N (1992). Use of regulated secretion in protein production from animal cells: An overview. *Adv. Biochem. Engin. Biotechnol.* 46:35–62.
80. Wurm F, Bernard A (1999). Large-scale transient expression in mammalian cells for recombinant protein production. *Curr. Opin. Biotechnol.* 10:156–159.
81. Hesse F, Wagner R (2000). Developments and improvements in the manufacturing of human therapeutics with mammalian cell cultures. *Trends Biotechnol.* 18:173–180.
82. Vapnek P (1991). Choosing a production system for recombinant proteins. *Biol. Rec. Microorg. Anim. Cells* 34:1–14.

83. Brooks S A (2004). Appropriate glycosylation of recombinant proteins for human use: implications of choice of expressions system. *Mol. Biotechnol.* 28:241–255.
84. Sodoyer R (2004). Expression systems for the production of recombinant pharmaceuticals. *BioDrugs* 18:51–62.
85. Sutherland A, Auclair K, Vederas J C (2001). Recent advances in the biosynthetic studies of lovastatin. *Curr. Opin. Drug Discov. Devel.* 4:229–236.
86. Schmidt F R (2004b). The challenge of multidrug resistance: Actual strategies in the development of novel antibacterials. *Appl. Microbiol. Biotechnol.* 63:335–343.
87. Parekh S, Vinci V A, Strobel R J (2000). Improvement of microbial strains and fermentation processes. *Appl. Microbiol. Biotechnol.* 54:287–301.
88. Penalva M A, Rowlands R T, Turner G (1998). The optimization of penicillin biosynthesis in fungi. *TIBTECH* 16:483–489.
89. Rowlands R T (1992). Strain improvement and strain stability. In: Finkelstein D B, Ball C (eds.), *Biotechnology of Filamentous Fungi, Technology and Products*, CRC Press, Boca Raton, FL, pp. 41–64.
90. Rowlands R T (1984). Industrial strain improvement: mutagenesis and random screening procedures. *Enzyme Microb. Technol.* 6:3–10.
91. Lal R, Khanna R, Kaur H, Khanna M, Dhingra N, Lal S, Gartemann K H, Eichenlaub R, Ghosh P K (1996). Engineering antibiotic producers to overcome the limitations of classical strain improvement programs. *Crit. Rev. Microbiol.* 22:201–255.
92. Brakhage A A, Sprote P, Al-Abdallah Q, Gehrke A, Plattner H, Tuncher A (2004). Regulation of penicillin biosynthesis in filamentous fungi. *Adv. Biochem. Eng. Biotechnol.* 88:45–90.
93. Schmitt E K, Hoff B, Kück U (2004). Regulation of cephalosporin biosynthesis. *Adv. Biochem. Eng. Biotechnol.* 88:1–43.
94. Brakhage A A (1998). Molecular Regulation of β-lactam biosynthesis in filamentous fungi. *Microbiol. Mol. Biol. Rev.* 62:547–585.
95. Martin J F (1998). New aspects of genes and enzymes for β-lactam antibiotic synthesis. *Appl. Microbiol. Biotechnol.* 50:1–15.
96. Skatrud P L (1992). Genetic engineering of β-lactam antibiotic biosynthetic pathways in filamentous fungi. *TIBTECH* 10:324–329.
97. Brakhage A A, Turner G (1995). Biotechnical genetics of antibiotic synthesis. In: Kück U (ed.), *The Mycota II, Genetics and Biotechnology*, Springer, Berlin, pp. 263–285.
98. Schmidt F R (2002). Beta-Lactam Antibiotics: Aspects of Manufacture and Therapy. In Osiewacz H D (ed.), *The Mycota X; Industrial Applications*, Springer, Berlin, pp. 69–91.
99. Zhang J, Wolfe S, Demain A L (1988). Phosphate repressible and inhibitible β-lactam synthetases in *Cephalosporium acremonium* strain C-10. *Appl. Microbiol. Biotechnol.* 29:242–247.
100. Menne S, Walz M, Kück U (1994). Expression studies with the bidirectional pcbAB-pcbC promotor region from *Acremonium chrysogenum* using reporter gene fusions. *Appl. Microbiol. Biotechnol.* 42:57–66.
101. Guiterrez S, Velasco J, Fernandez F J, Martin J F (1992). The cefG gene of *Cephalosporium acremonium* is linked to the *cefEF* gene and encodes a deacetylcephalosporin C acetyltransferase closely related to homoserin O-acetyltransferase. *J. Bacteriol.* 174:3056–3064.
102. Smith D J, Bull J H, Edwards J, Turner G (1989). Amplification of the isopenicillin N synthase gene in a strain of *Penicillium chrysogenum* producing high levels of penicillin. *Mol. Gen. Genet.* 216:492–497.

103. Smith D J, Burnham M R K, Bull J H, Hodgson J E, Ward J M, Browne P, Brown J, Barton B, Earl A J, Turner G (1990). β-lactam antibiotic biosynthetic genes have been conserved in clusters in prokaryotes and eukaryotes. *EMBO J.* 9:741–747.
104. Barredo J L, Diez B, Alvarez E, Martin J F (1989). Large amplification of 35-kb DNA fragment carrying two penicillin biosynthetic genes in high penicillin producing strains of *P. chrysogenum*. *Curr. Genet.* 16:453–459.
105. Fierro F, Barredo J L, Diez B, Guiterrez S, Fernandez J, Martin J F (1995). The penicillin cluster is amplified in tandem repeats linked by conserved hexanucleotide sequences. *Proc. Natl. Acad. Sci. USA* 92:6200–6204.
106. Newbert R W, Barton B, Greaves P, Harper J, Turner G (1997). Analysis of a commercially improved *Penicillium chrysogenum* strain series: involvement of recombinogenic regions in amplification and deletion of the penicillin biosynthesis gene cluster. *J. Ind. Microbiol. Biotechnol.* 19:18–27.
107. Brakhage A A, Browne P, Turner G (1994). Analysis of penicillin biosynthesis and the expression of penicillin biosynthesis genes of *Aspergillus nidulans* by targeted disruption of the acvA gene. *Mol. Gen. Genet.* 242:57–64.
108. Kück U, Walz M, Mohr G, Mracek M (1989). The 5′-sequence of the isopenicillin N-synthetase gene (pcbC) from *Cephalosporium acremonium* directs the expression of the prokaryotic hygromycin B phosphotransferase gene (hph) in *Aspergillus niger*. *Appl. Microbiol. Biotechnol.* 31:358–365.
109. Walz M, Kück U (1991). Polymorphic karyotypes in related *Acremonium* strains. *Curr. Genet.* 19:73–76.
110. Skatrud P L, Queener S W (1989). An electrophoretic molecular karyotype for an industrial strain of *Cephalosporium acremonium*. *Gene* 79:331–338.
111. Smith A W, Collis K, Ramsden M, Fox H M, Peberdy J F (1991). Chromosome rearrangements in improved cephalosporin C producing strains of *Acremoinium chrysogenum*. *Curr. Genet.* 18:235–237.
112. DeModena J A, Gutierrez S, Velasco J, Fernandez F J, Fachini R A, Galazzo J L, Hughes D E, Martin J E (1993). The production of cephalosporin C by *Acremonium chrysogenum* is improved by the intracellular expression of a bacterial hemoglobin. *Bio/Technology* 11:926–929.
113. Mathison L, Soliday C H, Stepan T, Aldrich T, Rambosek J (1993). Cloning, characterization, and use in strain improvement of the *Cephalosporium acremonium* gene *cefG* encoding acetyl transferase. *Curr. Genet.* 23:33–41.
114. Schmidt F R (2005a). About the nature of RNA interference. *Appl. Microbiol. Biotechnol.* 67:429–435.
115. Hannon G J, Rossi J J (2004). Unlocking the potential of the human genome with RNA interference. *Nature* 431:371–378.
116. Wassenegger M (2002). Gene silencing-based disease resistance. *Transgenic. Res.* 11:639–653.
117. Schmidt F R, Davis N D, Diener U L, Lemke P A (1983). Cycloheximide induction of aflatoxin synthesis in a nontoxigenic strain of *Aspergillus flavus*. *Bio/technology (Nature)* 1:794–795.
118. Schmidt F R, Lemke P A, Esser K (1986). Viral influences on aflatoxin formation by *Aspergillus flavus*. *Appl. Microbiol. Biotechnol.* 24:248–252.
119. Schmidt F R (2004c). RNA interference detected 20 years ago? *Nat. Biotechnol.* 22:267–268.
120. Pandit N N, Russo V E (1992). Reversible inactivation of a foreign gene, hph, during the asexual cycle in *Neurospora crassa* transformants. *Mol. Gen. Genet.* 234:412–422.

121. Romano N, Macino G (1992). Quelling: transient inactivation of gene expression in *Neurospora crassa* by transformation with homologous sequences. *Mol. Microbiol.* 6:3343–3353.
122. Pickford A S, Catalanotto C, Cogoni C, Macino G (2002). Quelling in *Neurospora crassa*. *Adv. Genet.* 46:277–303.
123. Liras P, Rodriguez-Garcia A, Martin J F (1998). Evolution of the clusters of genes for beta-lactam antibiotics: A model for evolutive combinatorial assembly of new beta-lactams. *Int. Microbiol.* 1:271–278.
124. Nielsen J (2001). Metabolic engineering. *Appl. Microbiol. Biotechnol.* 55:263–285.
125. Weber T, Welzel K, Pelzer S, Vente A, Wohhleben W (2003). Exploiting the genetic potential of polyketide producing streptomycetes. *J. Biotechnol.* 106:221–232.
126. Kantola J, Kunnari T, Mantsala P, Ylihonkoa K (2003). Expanding the scope of aromatic polyketides by combinatorial biosynthesis. *Comb. Chem. High Throughput Screen* 6:501–512.
127. Hutchinson C R, Fujii I (1995). Polyketide synthase gene manipulation: A structure-function approach in engineering novel antibiotics. *Annu. Rev. Microbiol.* 49:201.
128. Katz L, Donadio S (1993). Polyketide synthesis: prospects for hybrid antibiotics. *Annu. Rev. Microbiol.* 47:875–912.
129. Floss H G (2001). Antibiotic synthesis: from natural to unnatural products. *J. Ind. Microbiol. Biotechnol.* 27:183–194.
130. Pfeifer B A, Khosla C (2001). Biosynthesis of polyketides in heterologous hosts. *Microbiol. Mol. Biol. Rev.* 3:106–118.
131. Staunton J, Wilkinson B (2001). Combinatorial biosynthesis of polyketides and non-ribosomal peptides. *Curr. Opin. Chem. Biol.* 5:159–164.
132. McDaniel R, Katz L (2001). Genetic engineering of novel macrolide antibiotics. In: Lohner K (ed.), *Development of Novel Antimicrobial Agents: Emerging Strategies.* Horizon Scientific Press Norwich, U.K., pp. 45–60.
133. Du L, Shen B (2001). Biosynthesis of hybrid peptide–polyketide natural products. *Curr. Opin. Drug. Discov. Dev.* 4:215–228.
134. Du L, Sanchez C, Shen B (2001). Hybrid peptide—polyketide natural products: Biosynthesis and prospects towards engineering novel molecules. *Metab. Eng.* 3:78–95.
135. Kleinkauf H, von Döhren H (1995). Linking peptide and polyketide biosynthesis. *J. Antibiot.* 48:563–567.
136. Cane D E, Walsh C T (1999). The parallel and convergent universes of polyketide synthases and non-ribosomal peptide synthases. *Chem. Biol.* 6:R319–325.
137. Marahiel M A, Stachelhaus T, Mootz H D (1997). Modular peptide synthases involved in nonribosomal peptide synthesis. *Chem. Rev.* 97:2651–2673.
138. Mankelow D P, Neilan B A (2000). Non-ribosomal peptide antibiotics. *Expert Opin. Ther. Pat.* 10:1583–1591.
139. Trauger J W, Walsh C T (2000). Heterologous expression in *Escherichia coli* of the first module of nonribosomal peptide synthase for chloroeremomycin, a vancomycin-type glycopeptide antibiotic. *Proc. Natl. Acad.* 97:3112–3117.
140. Kaneda T (1977). Fatty acids of the genus Bacillus: an example of branched chain preference. *Bacteriol. Rev.* 41:391–418.
141. Aretz W, Meiwes J, Seibert G, Vobis G, Wink J (2000). Friulimicins: Novel lipopetide antibiotics with peptidoglycan synthesis inhibiting activity from *Actinoplanes friulensis* sp. Nov. I. Taxonomic studies of the producing microorganism and fermentation. *J. Antibiotics* 53:807–815.

142. Vertesy L, Ehlers E, Kogler H, Kurz M, Meiwes J, Seibert G, Vogel M, Hammann P (2000). Friulimicins: Novel lipopetide antibiotics with peptidoglycan synthesis inhibiting activity from *Actinoplanes friulensis* sp. nov. II. Isolation and structural characterization. *J. Antibiotics* 53:816–827.

143. Vertesy L, Aretz W, Decker H, Ehlers E, Kurz M, Schmidt F R, Knauf M, Kogler H (1999). Lipopetide calcium salts: Process for their preparation and their use. *EP* 1068223.

144. Boeck L D, Wetzel R W (1990). A54145, A new lipopeptide antibiotic complex: factor control through precursor directed biosynthesis. *J. Antibiotics* 43:607–615.

145. Counter F T, Allen N E, Fukuda D S, Hobbs J N, Ott J, Ensminger P W, Mynderse J S, Preston D A, Wu C Y E (1990). A54145 A new lipopetide antibiotic complex: microbiological evaluation. *J. Antibiotics* 43:616–622.

146. Zmijewski M J, Briggs B, Occolowitz J (1986). Role of branched chain fatty acid precursosrs in regulating factor profile in the biosynthesis of A21978 complex. *J. Antibiot* 39:1483–1485.

147. Stephenson J (2001). Researchers describe latest strategies to combat antibiotic resistant microbes. *JAMA* 285:2317–2318.

148. Tally F P, DeBriun F (2000). Development of daptomycin for Gram-positive infections. *J. Antimicrob. Chemo.* 46:523–526.

149. Hancock R E W, Patrzykat A (2002). Clinical development of cationic antimicrobial peptides: From natural to novel antibiotics. *Curr. Drug Targets Infect. Disord.* 2:79–83.

150. Mor A (2000). Peptide–based antibiotics: a potential answer to raging antimicrobial resistance. *Drug Dev. Res.* 50:440–447.

151. Larsson G, Törnkvist M, Wernersson E S, Trägard C, Noorman H, Enfors S O (1996). Substrate gradients in bioreactors: origin and consequences. *Bioproc. Engin.* 14:281–289.

152. Bylund F, Collet E, Enfors E O, Larsson G (1998). Substrate gradient formation in the large scale lowers cell yield and increases byproduct formation. *Bioproc. Eng.* 18:171–180.

153. Bylund F, Guillard F, Enfors S O, Trägardh C, Larsson G (1999). Scale down of recombinant production: a comparative study of scaling performance. *Bioproc. Engin.* 20:377–289.

154. Bylund F, Castan A, Mikkola R, Veide A, Larsson G (2000). Influence of scale-up on the quality of recombinant human growth hormone. *Biotechnol. Bioengin.* 69:119–128.

155. Enfors S O, Jahic M, Rozkov A, Xu B, Hecker M, Jürgen B, Krüger B, Schweder T, Hamer G, O'Beirne D, Noisommit-Rizzi N, Reuss M, Boone L, Hewitt C, McFarlane C, Nienow A, Kovacs T, Trägardh C, Fuchs L, Revstedt J, Friberg P C, Hjertager B, Blomsten G, Skogman H, Hjort S, Hoeks F, Lin H J, Neubauer P, van der Lans R, Luyben K, Vrabel P, Manelius A (2001). Physiological responses to mixing in large scale bioreactors. *J. Biotechnol.* 85:175–185.

156. Xu B, Jahic M, Bomsten G, Enfors S O (1999). Glucose overflow metabolism and mixed acid fermentation in aerobic large-scale fed-batch processes with *Echerichia coli*. *Appl. Microbiol. Biotechnol.* 51:564–571.

157. Castan A, Enfors S O (2001). Formate accumulation due to DNA release in aerobic cultivations of *Escherichia coli*. *Biotechnol. Bioengin.* 77:324–328.

158. Kwint K, Nachin L, Diez A, Nyström T (2003). The bacterial universal stress protein: Function and regulation. *Curr. Opin. Microbiol.* 6:140–145.

159. Mogk A, Tomoyasu T, Goloubinoff P, Rudiger S, Roder D, Langen H, Bukau B (1999). Identification of thermolabile *Escherichia coli* proteins: Prevention and reversion of aggregation by DnaK and ClpB. *EMBO J.* 18:6934–6949.
160. Umakoshi H, Kuboi R, Komaswa I, Tsuchido T, Matsumura Y (1998). Heat-induced translocation of cytoplasmic beta-galactosidase across inner membrane of *Escherichia coli*. *Biotechnol. Prog.* 14:210–217.
161. Schweder T, Krüger E, Xu B, Jürgen B, Blomsten G, Enfors S O, Hecker M (1999). Monitoring genes that respond to process-related stress in large-scale bioprocesses. *Biotechnol. Bioengin.* 65:151–159.
162. Fenton D, Lai P H, Lu H, Mann M, Tsai L (1997). Control of norleucine incorporation into recombinant proteins. U.S. Patent 5599690.
162a. Martinez-Force E, Benitez T (1992). Changes in yeast amino acid pool with respiratory versus fermentative metabolism. *Biotechnol. Bioengi.n* 40:843–849.
162b. Hondorp ER, Matthews RG (2004). Oxidative stress inactivates cobalamin-independent methionine synthase (MetE) in Escherichia coli. *PLoS Biology* 2:1138–1158.
162c. Hoffmann F, Rinas U (2004). Stress induced by recombinant protein production in *Escherichia coli*. *Adv. Biochem. Eng. Biotechnol.* 89:73–92.
163. Weickert M J, Apostol I (1998). High-fidelity translation of recombinant human hemoglobin in *Escherichia coli*. *Appl. Env. Microbial.* 64:1589–1593.
164. Muramatsu R, Negishi T, Mimoto T, Miura A, Miosawa S, Hayashi H (2002). Existence of β-methylnorleucine in recombinant hirudin produced by *Escherichia coli*. *J. Biotechnol.* 93:131–142.
165. Apostol I, Levine J, Lippincott J, Leach J, Hess E, Glascock C B, Weickert M J, Blackmore R (1997). Incorporation of norvaline at leucine positions in recombinant human hemoglobin expressed in *Escherichia coli*. *J. Biol. Chem.* 272:28980–28988.
166. Kiick K L, Weberskirch R, Tirell D A (2001). Identification of an expanded set of translationally active methionine analogues in *E. coli*. *FEBS Lett.* 502:25–30.
167. Friehs K (2003). Plasmid copy number and plasmid stability. *Adv. Biochem. Engin. Biotechnol.* 86:47–82.
168. Thiry M, Cingolani D (2002). Optimizing scale-up fermentation processes. *Trends Biotechnol.* 20:103–105.
169. Lin H Y, Neubauer P (2000). Influence of controlled glucose oscillations on a fed-batch process of recombinant *Escherichia coli*. *J. Biotechnol.* 79:27–37.
170. Neubauer P, Lin H Y, Mathiszik B (2003). Metabolic load of recombinant protein production: Inhibition of cellular capacities for glucose uptake and respiration after induction of a heterologous gene in *Escherichia coli*. *Biotechnol. Bioengin.* 83:53–64.
171. Ansorge M B, Kula M R (2000). Production of recombinant L-leucine dehydrogenase from Bacillus cereus in pilot scale using the runaway replication system *E. coli* (pIET98). *Biotechnol. Bioengin.* 68:557–562.
172. Nordstrom K, Uhlin B E (1992). Runaway-replication plasmids as tools to produce large quantities of proteins from cloned genes in bacteria. *Biotechnology* 6:661–666.
173. van de Merbel N C (1997). The use of ultra-filtration and column liquid chromatography for an on-line fermentation monitoring. *Trends. Anal. Chem.* 16:162–173.
174. Buchholz A, Takors R, Wandrey C (2001). Quantification of intracellular metabolites in *Escherichia coli* K12 using liquid chromatographic electrospray ionization tandem mass spectrometric techniques. *Anal. Biochem.* 295:129–137.
175. Matz G, Lennemann F (1996). On-line monitoring of biotechnological processes by gas chromatographic—mass spectrometric analysis of fermentation suspensions. *J. Chromatogr.* 750:141–149.

176. Klyushnichenko V (2004). Capillary electrophoresis in the analysis and monitoring of biotechnological processes. *Methods Mol. Biol.* 276:77–120.
177. Nishi H, Tsumagari N, Kakimoto T, Terabe S (1989). Separation of β-lactam antibiotics by micellar electro-kinezic chromatography. *J. Chromatogr.* 477:259–270.
178. Nishi H, Fukuyama T, Matsuo M (1990). Separation and determination of aspoxicillin in human plasma by micellar electrokinetic chromatography with direct sample injection. *J. Chromatogr.* 515:245–255.
179. Gabig-Ciminska M, Holmgren A, Andresen H, Bundvik Barken K, Wumpelmann M, Albers J, Hintsche R, Breitenstein A, Neubauer P, Los M, Czyz A, Wegrzyn G, Silfversparre G, Juregn B, Schweder T, Enfors S O (2004). Electric chips for rapid detection and quantification of nucleic acids. *Biosens. Bioelectron.* 15:537–546.
180. Hewitt C J, Nebe-von Caron G (2001). An industrial application of multiparameter flow cytometry: Assessment of cell physiological state and its application to the study of microbial fermentations. *Cytometry* 44:179–187.
181. Hewitt C J, Nebe-Von-Caron G (2004). The application of multi-parameter flow cytometry to monitor individual microbial cell physiological state. *Adv. Biochem. Eng. Biotechnol.* 89:197–223.
182. Hewitt C J, Nebe-von Caron G, Axelsson B, McFarlane C M, Nienow A W (2000). Studies related to the Scale-up of high-cell-density *E. coli* fed-batch fermentations using multiparameter flow cytometry: Effect of a changing microenvironment with respect to glucose and dissolved oxygen concentration. *Biotechnol. Bioengin.* 70:381–390.
183. Kell D B, Ryder H M, Kaprelyants A S, Westerhoff H V (1991). Quantifying heterogenity—flow cytometry of bacterial cultures. *Anton. van Leeuwen. Int. J. Gen. Mol. Microbiol.* 60:145–158.
184. Harms P, Kostov Y, Rao G (2002). Bioprocess monitoring. *Curr. Opin. Biotechnol.* 13:124–127.
185. Schaefer U, Boos W, Takors R, Weuster-Bootz D (1999). Automated sample device for monitoring intracellular metabolite dynamics. *Anal. Biochem.* 270:88–96.
186. Rocha I, Ferreira E C (2002). On-line simultaneous monitoring of glucose and acetate with FIA during high cell density fermentation of recombinant *E. coli*. *Bioelectrochem.* 56:127–129.
187. Tothill I E, Newman J D, White S F, Turner A P F (1997). Monitoring of the glucose concentration during microbial fermentation using a novel mass-producible biosensor suitable for on-line use. *Enzyme Microb. Technol.* 20:590–596.
188. Schügerl K, Brandes T, Dullau K, Holzhauer-Rieger K, Hotop S, Hübner U, Wu X, Zhou W (1991). Fermentation monitoring and control by online flow injection and liquid chromatography. *Anal. Chim. Acta.* 249:87–100.
189. Huang Y L, Foellmer T J, Ang K C, Khoo S B, Yap M G S (1995). Characterization and application of an on-line flow injection analysis/wall-jet electrode system for glucose monitoring during fermentation. *Anal. Chim. Acta.* 317:223–232.
190. Vaidyanathan S, Macaloney G, Vaughan J, McNeil B, Harvey L M (1999). Monitoring of submerged bioprocesses. *Crit. Rev. Biotechnol.* 19:277–316.
191. Arndt M, Hitzmann B (2004). Kalman filter based glucose control at small set points during fed-batch cultivation of Saccharomyces cerevisiae. *Biotechnol. Prog.* 1:377–383.
192. Kleist S, Miksch G, Hitzmann B, Arndt M, Friehs K, Flaschel E (2003). Optimization of the extracellular production of bacterial phytase with *Escherichia coli* by using different fed-batch fermentation strategies. *Appl. Microbiol. Biotechnol.* 61:456–462.

193. Spichiger-Keller U E (1997). Ion- and substrate specific optode membranes and optical detection modes. *Sensors and Actuators, B: Chemical* B38:68–77.
194. Wolfbeis O S (2002). Fiber-optic chemical sensors and biosensors. *Anal. Chem.* 74:2663–2678.
195. Griffiths D, Hall G (1993). Biosensors—what real progress is being made. *Trends Biotechnol.* 11:122–130.
196. Boehl D, Solle D, Hitzmann B, Scheper T (2003). Chemometric modeling with two-dimensional fluorescence data for Claviceps purpurea bioprocess characterization. *J. Biotechnol.* 105:179–188.
197. Solle D, Geissler D, Stark E, Scheper T, Hitzmann B (2003). Chemometric modeling based on 2D-fluorescence spectra without a calibration measurement. *Bioinformatics* 19:173–177.
198. Stark E, Hitzmann B, Schügerl K, Scheper T, Fuchs C, Koster D, Märkl H (2002). In-situ-fluorescence probes: a useful tool for non-invasive bioprocess monitoring. *Adv. Biochem. Eng. Biotechnol.* 74:21–38.
199. Hall J W, McNeil B, Rollins M J, Draper I, Thompson B G, Macaloney G (1996). Near-infrared spectroscopic determination of acetate, ammonium, biomass and glycerol in an industrial *Escherichia coli* fermentation. *Appl. Spectroscopy* 50:102–108.
200. Vaccari G, Dosi E, Campi A L, Gonzalez-Vara A, Matteuzi D, Mantovani G (1994). A near infrared spectroscopy technique for the control of fermentation processes: an application to lactic acid fermentation. *Biotechnol. Bioengin.* 43:913–917.
201. Macaloney G, Hall J W, Rollins M J, Draper I, Thompson B G, McNeil B (1994). Monitoring biomass and glycerol in an *Escherichia coli* fermentation using near-infrared spectroscopy. *Biotechnol. Techn.* 8:281–286.
202. Hammond S V (1992). NIR analysis of antibiotic fermentations. In: Murray I, Cowe I A (eds.), *Making Light Work: Advances in Near Infrared Spectroscopy*, vol 1. VCH Publishers, New York, pp. 584–589.
203. Hammond S V, Brookes I K (1992). Near infrared spectroscopy—a powerful technique for at-line and on-line analysis of fermentations. In: Ladisch M R, Bose A (eds.), *Harnessing Biotechnology for the 21st Century: Proceedings of the 9th International Symposium and Exhibition*, *Am. Chem. Soc.*, Washington, DC, pp. 325–333.
204. Pollard D, Buccino R, Connors N, Kirschner T, Olewinski R, Saini K, Slamon P (2001). Real-time analyte monitoring of a fungal fermentation, at a pilot scale, using in mid-infrared spectroscopy. *Bioproc. Biosyst. Engin.* 24:13–24.
205. Dickinson T A, White J, Kauer J S, Walt D R (1998). Current trends in artificial nose technology. *TIBTECH* 16:250–258.
206. Turner C, Rudnitskaya A, Legin A (2003). Monitoring batch fermentations with an electronic tongue. *J. Biotechnol.* 103:87–91.
207. Moser I, Jobst G, Urban G A (2002). Biosensor array for simultaneous measurement of glucose, lactate, glutamate, and glutamine. *Biosens. Bioelectron.* 17:297–302.
208. Collins G E, Wu P, Lu Q, Ramsey J D, Bromund R H (2004). Compact, high voltage power supply for the lab-on-chip. *Lab. Chip.* 4:408–411.
209. Neils C, Tyree Z, Finlayson B, Folch A (2004). Combinatorial mixing of microfluidic streams. *Lab. Chip.* 4:342–350.
210. Farinas J, Chow A W, Wada H G (2001). A microfluidic device for measuring cellular membrane potential. *Anal. Biochem.* 295:138–142.
211. Ferreira L S, de Souza M B, Folly R O M (2001). Development of an alcohol fermentation control system based on biosensor measurements interpreted by neural networks. *Sens. Act B: Chem.* 75:166–171.

212. Chen V C P, Rollins D K (2000). Issues regarding artificial neural network modeling for reactors and fermenters. *Bioproc. Biosyst. Engin.* 22:85–93.
213. Trelea I C, Titica M, Landaud S, Latrille E, Corrieu G, Cheruy A (2001). Predictive modeling of brewing fermentation: From knowledge-based to black box models. *Math Comp. Sim.* 56:405–424.
214. Lin H Y, Mathiszik B, Xu B, Enfors S O, Neubauer P (2001). Determinantion of the maximum specific uptake capacities for glucose and oxygen in glucose-limited fed-batch cultivations of *Escherichia coli*. *Biotechnol. Bioengin.* 73:347–357.
215. Lee S Y (1996). High-cell density culture of *Escherichia coli*. *Trends Biotechnol.* 14:98–105.
216. Kim C H, Rao K J, Youn D J, Rhee S K (2003). Scale-up of recombinant hirudin production from *Saccharomyces cerevisiae*. *Biotechnol. Bioproc. Engin.* 8:303–305.
217. Wong I, Hernandez A, Garcia M A, Segura R, Rodriguez I (2002). Fermentation scale-up for recombinant K99 antigen production cloned in *Escherichia coli* MC1061. *Proc. Biochem.* 37:1195–1199.
218. Yim S C, Jeong K J, Chang H N, Lee S Y (2001). High-level secretory production of human granulocyte-colony stimulating factor by fed-batch culture of recombiant *E. coli*. *Bioproc. Biosyst. Engin.* 24:249–254.
219. Dantigny P, Ninow J L, Lakrori M (1991). A new control strategy for yeast production based on the L/A approach. *Appl. Microbiol. Biotechnol.* 36:352–357.
220. Cant S (2002). High-performance computing in computational fluid dynamics: Progress and challenges. *Philosoph. Transact.: Mathem. Phys. Engin. Sci.* 360:1211–1225.
221. Schmalzriedt S, Jenne M, Mauch K, Reuss M (2003). Integration of physiology and fluid dynamics. *Adv. Biochem. Eng. Biotechnol.* 80:19–68.
222. Alves S S, Vasconcelos J M T (1996). Optimization of agitation and aeration in fermenters. *Bioproc. Biosyst. Engin.* 14:119–123.
223. Oniscu C, Galaction A I, Cascaval D, Ungureanu F (2001). Modeling of mixing in stirred bioreactors 1. Mixing time for non-aerated simulated broths. *Roum. Biotechnol. Lett.* 6:119–129.
224. Diaz A, Acevedo F (1999). Scale-up strategy for bioreactors with Newtonian and non-Newtonian broths. *Bioproc. Engin.* 21:21–23.
225. Russel A B, Thomas C R, Lilly M D (1995). Oxygen transfer measurements during yeast fermentations in a pilot scale airlift fermenter. *Bioproc. Biosyst. Engin.* 12:71–79.
226. Türker M (2004). Development of biocalorimetry as a technique for process monitoring and control in technical scale fermentations. *Thermochimica Acta.*, In Press.
227. Anderson R K I, Jayaraman K, Voisard D, Marison I W, von Stockar U (2002). Heat flux as an on-line indicator of metabolic activity in pilot scale bioreactor during the production of *Bacillus thuringiensis* var *galleriae*-based pesticides. *Thermochimica Acta.* 386:127–138.
228. Hubbard D W (1987). Scaleup strategies for bioreactors containing non-newtonian broths. *Ann. NY Acad. Sci.* 506:600–607.
229. Junker B H, Stanik M, Barna C, Salmon P, Buckland B C (1998). Influence of impeller type on mass transfer in fermentation vessels. *Bioproc. Biosyst. Engin.* 19:403–413.
230. Wang D I C, Cooney C L (1997). Translation of laboratory, pilot, and plant scale data. In: Wang D I C, et al. (eds.), *Fermentation and enzyme technology.* Wiley, New York, pp. 194–211.
231. Yawalkar A A, Heesink A B M, Versteeg G F, Vishwas G P (2002). Gas-liquid mass transfer coefficient in stirred tank reactors. *Can. J. Chem. Engin.* 80:840–848.

232. Flores E R, Perez F, de la Torre M (1997). Scale-up of *Bacillus thuringiensis* fermentation based on oxygen transfer. *J. Ferm. Bioengin.* 83:561–564.
233. Riesenberg D, Menzel K, Schulz V, Schumann K, Veith G, Zuber G, Knorre W A (1990). High cell density fermentation of recombinant *Escherichia coli* expressing human interferon alpha 1. *Appl. Microbiol. Biotechnol.* 34:77–82.
234. Berkholz R, Guthke R (2001). Model based sequential experimental design for bioprocess optimization—an overview. *Focus Biotechnol.* 4:129–141.
235. Berkholz R, Rohlig D, Guthke R (2000). Data and knowledge based experimental design for fermentation process optimization. *Enzyme Microb. Technol.* 27:784–788.
236. Liden G (2002). Understanding the bioreactor. *Bioproc. Biosyst. Engin.* 24:273–279.
237. Drossard J (2004). Downstream processing of plant-derived recombinant therapeutic proteins. In: Fischer R, Schillberg S (eds.), *Molecular Farming.* Wiley-VCH, Weinheim, Germany.
238. Roque A C, Lowe C R, Taipa M A (2004). Antibodies and genetically engineered related molceules: Production and purification. *Biotechnol. Prog.* 20:639–554.
239. Jungbauer A, Hahn R (2004). Enegineering protein A affinity chromatography. *Curr. Opin. Drug Discov. Devel.* 7:248–256.
240. Burgess R R, Thompson N E (2002). Advances in gentle immunoaffinity chromatography. *Curr. Opin. Biotechnol.* 13:304–308.
241. Hage D S (1999). Affinity chromatography: A review of clinical applications. *Clin. Chem.* 45:593–615.
242. Murphy M B, Fuller S T, Richardson P M, Doyle S A (2003). An improved method for the in vitro evolution of aptamers and applications in protein detection and purification. *Nucl. Ac. Res.* 31:18–110.
243. Romig T S, Bell C, Drolet D W (1999). Aptamer affinity chromatography: combinatorial chemistry applied to protein purification. *J. Chromatogr. B Biomed. Sci. Appl.* 731:275–284.
244. Hubbuch J, Thommes J, Kula M R (2005). Biochemical engineering aspects of expanded bed adsorption. *Adv. Biochem. Eng. Biotechnol.* 92:101–123.
245. Anspach F B, Curbelo D, Hartmann R, Garke G, Deckwer W D (1999). Expanded bed chromatography in primary protein purification. *J. Chromatogr. A* 865:129–144.
246. Guiochon G (2002). Preparative liquid chromatography. *J. Chromatogr. A* 965:129–161.
247. Cunha T, Aires-Barros R (2002). Large-scale extraction of proteins. *Mol. Biotechnol.* 20:29–40.
248. Krishna S H, Srinivas N D, Raghavarao K S, Karanth N G (2002). Reverse micellar extraction for downstream processing of proteins/enzymes. *Adv. Biochem. Eng. Biotechnol.* 75:119–183.
249. Walter J (2000). Scale-up of downstream processing. *J. Chromatogr. Libr.* 61:765–783.
250. Kadir F, Hamers M (2002). Production and downstream processing of biotech compounds. In: Crommelin D J A, Sindelar R D (eds.), *Pharmaceutical Biotechnology* (2nd ed.). Taylor and Francis Ltd, London, U.K., pp. 53–71.
251. Kalyanpur M (2002). Downstream processing in the biotechnology industry. *Mol. Biotechnol.* 22:87–98.
252. Chisti Y (1998). Strategies in downstream processing. In: Subramanian G (ed.), *Bioseparation and Bioprocessing* (2nd ed.). Wiley-VCH, Weinheim, Germany, pp. 3–30.

253. Desai M A (2002). *Downstream Processing of Proteins.* Humana Press, Totowa, NJ.
254. Bhat S T, et al. (2005). *Chemistry of natural products.* Springer, Berlin, Germany.
255. Verrall M S (1996). Downstream processing of natural products. A practical handbook. Wiley, New York.
256. Gosh A C, Bora M M, Dutta N N (1996). Developments in liquid membrane separation of beta-lactam antibiotics. *Bioseparation* 6:91–105.
257. Gosh A C, Mathur R K, Dutta N N (1997). Extraction and purification of cephalosporin antibiotics. In: Scheper T (ed.), *Advances in Biochemical Engineering/ Biotechnology*, vol 56. Springer-Verlag, Berlin, pp. 111–145.
258. Bandak S I, Tumak M R, Allen B S, Bolzon L D, Preston D A (2000). Assessment of the susceptibility of Streptococcus pneumoniae to cefaclor and loracarbef in 13 countries. *J. Chemother.* 12:299–305.
259. Troxler R, von Grävenitz A, Funke G, Wiedemann B, Stock I (2000). Natural antibiotic susceptibility by *Listeria* species: *L. grayi, L. innocua, L. ivanovii, L. monocytogenes, L. seeligeri* and *L. welshimeri* strains. *Clin. Microbiol. Infect.* 6:525–535.
260. Saling P (2005). Eco-efficiency analysis of biotechnological processes. *Appl. Microbiol. Biotechnol.* 68:1–8.
261. Parmar A, Kumar H, Marwaha S S, Kennedy J F (1998). Recent trends in enzymatic conversion of cephalosporin C to 7-aminocephalosporanic acid (7-ACA). *Cri. Rev. Biotechnol.* 18:1–12.
262. Bianchi D, Bortolo R, Golini P, Cesti P (1998). Application of immobilized enzymes in the manufacture of beta-lactam antibiotics. *Chim. In. (Milan)* 80:875–885.
263. Binder R G, Numata K, Lowe D A, Murakami T, Brown J L (1993). Isolation and characterization of a *Pseudomonas* strain producing glutaryl-7-aminocephalosporanic acid acylase. *Appl. Env. Microbiol.* 59:3321–3326.
264. Isogai T (1997). New processes for production of 7-aminocephalosporanic acid from cephalosporin. *Drugs Pharm. Sci.* 82:733–751.
265. Gomez A, Rodriguez M, Ospina S, Zamora R, Merino E, Bolivar F, Ramirez O T, Quintero R, Lopez-Munguia A (1994). Strategies in the design of a penicillin acylase process. In: Galindo R, Ramirez OT (eds.), *Advances in Bioprocess Engineering.* Kluwer, Dodrecht, pp. 29–40.
266. Bruggink A, Roos E C, de Vroom E (1998). Penicillin acylase in the industrial production of beta-lactam antibiotics. *Org. Process Res. Dev.* 2:128–133.
267. Shewale J G, Sudhakaran V K (1997). Penicillin V acylase: Its potential in the production of 6-aminopenicillanic acid. *Enzyme Microb. Technol.* 20:402–410.
268. Stoppok E, Wagner F, Zadrazil F (1981). Identification of a penicillin V acylase processing fungus. *Eur. J. Appl. Microbiol. Biotechnol.* 13:60–61.
269. Schneider W J, Roehr M (1976). Purification and properties of penicillin acylase of *Bovista plumbea. Biochimica. Biophysica. Acta.* 452:177–185.
270. Brandl D, Kleiber W (1972). Verfahren zur Herstellung von 6-Aminopenicillansäure. Austrian Patent no. 297923, Vienna, Austria.
271. Singh K, Sehgal S N, Vezina C (1967). Conversion of penicillin V to 6-aminopenicillanic acid by the use of spores. U.S. Patent no. 3.305.453.
272. Wilms J, de Vroom E (1998). Preparation of β-lactam antibiotics. European Patent no. 0201688.
273. Schmidt F R (2005b). Optimization and scale-up of industrial fermentation processes. *Appl. Microbiol. Biotechnol.* 68:425–435.

# SEQUENCING THE HUMAN GENOME: WAS IT WORTH IT?

EUGENE ZABAROVSKY
*Microbiology and Tumor Biology Center, Karolinska Institute, Stockholm, Sweden*

**Chapter Contents**

1 Summary of the Chapter
2 Definitions of Keywords
3 Introduction. Short History of the Human Genome Project
4 Vectors. Description of the Main Types of Vectors Used in the HGP
  4.1 Lambda-Based Vectors and Main Approaches to Construct Genomic Libraries
  4.2 YACs
  4.3 Bacterial (BAC) and P1-Derived (PAC) Artificial Chromosome Vectors
5 Mapping of the Human Genome
6 Main Approaches to Sequence Human Genome
  6.1 Hierarchical and the Whole Genome Shotgun Sequencing Scheme (WGS)
  6.2 Massively Parallel Sequencing Approaches
  6.3 Slalom Library: A Novel Approach to Genome Mapping and Sequencing
7 Identification of Genes. How Many Genes Are in the Human Genome?
  7.1 Many Approaches Exist for Gene Identification
  7.2 CpG Islands
  7.3 SAGE and CAGE Approaches to Analyze Human Transcriptome and to Estimate the Number of Human Genes
8 Was it Worth to it Sequence Human Genome?
  Acknowledgment
  References

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

## 1 SUMMARY OF THE CHAPTER

An almost complete sequence of the human genome is available now, and progress in the sequencing of other model organisms is impressive. This sequence information provides enormous possibility for identification of disease genes. The first step for sequencing of large genomes usually is the construction of a high-resolution, long-range physical map based on sequence tagged site (STS) markers. Using this map, the contig of overlapping clones is constructed and sequenced. This strategy (hierarchical) was used by the International Human Genome Sequencing Consortium to sequence the human genome. The second approach, the whole genome sequencing (WGS) strategy, was used by researchers from Celera Genomics. This strategy, in principle, doesn't need construction of high-resolution physical map and is based mainly on shotgun sequencing of large and small insert clones.

Unfortunately, both strategies for mapping and sequencing are not appropriate for the challenge of high-throughput comparative genomics. Alternative and complementary strategies need to be developed, and the imperative now is to find cost-effective and convenient methods that allow comparative genomics projects to be performed by a wide range of smaller laboratories. Recently developed novel techniques open new perspectives for nonexpensive and fast sequencing. These novel developments can be grouped into two classes. The first type of improvement is enormous, increasing the speed and throughput of producing sequences. These techniques are based on massively parallel sequencing approaches (MPS and "494" techniques) where billions of individual DNA molecules could be sequenced simultaneously. Another type of improvements is a recently suggested completely different and efficient strategy for simultaneous genome mapping and sequencing. The approach is based on physically oriented, overlapping restriction fragment libraries called slalom libraries. Both techniques allow increasing the speed and decreasing the cost of sequencing 10–100 times. Importantly, they can be combined to increase efficiency even more.

## 2 DEFINITIONS OF KEYWORDS

**Adaptor**—an oligonucleotide that after ligation to a DNA molecule confers to it specific features.

**Allele**—is one of several alternative forms of a gene (or in the wider sense of any region of the chromosomal DNA) that occupies a given locus of chromosome.

**Cap**—the chemical modification at the 5′-end of most eukaryotic mRNA molecules.

**Centromere**—the region of a chromosome where the pairs of chromatids are held together.

**Clone contig**—a collection of clones (i.e., vectors or cells containing the same recombinant molecule) whose DNA inserts overlap.

**Coding RNA**—an RNA molecule that codes for a protein (mRNA).

**DNA sequencing**—establishing the order of nucleotides in a DNA molecule.

**Episome**—a vector or DNA sequence that can exist as an autonomous extrachromosome replicating DNA molecule.

**Exon**—a coding region within a discontinuous gene.

**Genetic selection**—in cloning, this usually means selection against parental, nonrecombinant molecules in favor of recombinant ones. SupF selection exploits vectors carrying *amber* (nonsense mutation UAG) mutations (*ochre* is a nonsense mutation UAA). These vectors cannot replicate without the supF gene, which must be present either in the host or in the cloned insert. If the insert carries the supF gene, only recombinant phages will be able to replicate in an *Esherichia coli* host without the suppressor gene.

**Intron**—a noncoding region within a discontinuous gene.

**Kilobase pairs** (kb)—1000 base pairs.

**Ligation**—formation of a 3′–5′ phosphodiester bond between nucleotides at the end of the same or two different DNA or RNA molecules.

**Marker**—a gene or just DNA fragment that was located on a genome map.

**Megabase pairs** (Mb)—1000 kb.

**Polylinker**—a short DNA fragment (in the vector) that contains recognition sites for restriction enzymes that can be used for cloning DNA fragments into this vector.

**Restriction enzyme**—an enzyme that recognizes a specific sequence in DNA and can cut at or near this sequence. In cloning procedures, the most commonly used enzymes produce specific protruding (sticky) ends at the ends of a DNA molecule. Each enzyme produces unique sticky ends. The DNA molecules possessing the same sticky ends can be efficiently joined with the aid of DNA ligase (see **Ligation**).

**STS**—sequence tagged site. This is a short (200–500 bp) sequenced fragment of genomic DNA that can be specifically amplified using polymerase chain reaction (PCR) and represent or is linked to some kind of marker (i.e., it is mapped to a specific locus on a chromosome).

**Shot-gun approach**—a sequencing strategy for when a large DNA molecule is randomly broken, cloned, and sequenced. The final sequence of a DNA molecule is assembled by a computer program.

**Telomere**—the end of a eukaryotic chromosome.

## 3 INTRODUCTION. SHORT HISTORY OF THE HUMAN GENOME PROJECT

The Human Genome Program (HGP) was suggested and widely discussed within the scientific community and public press during the mid-1980s and over the last half of that decade [1–4]. In 1985, Dr. Charles DeLisi, then Associate Director for Health and Environmental Research at the Department of Energy (DOE) began to discuss an unprecedented biology project—to sequence the complete human genome. DOE funding began in 1987. The National Institutes of Health (NIH) established the Office of Human Genome Research in September 1988. The Director of this Office (renamed the National Center for Human Genome Research, NCHGR) was Dr. James Watson. The Human Genome Initiative was proposed to the Congress of the United States in 1988 by James Watson, initiating a worldwide coordinated research activity to sequence the DNA of humans and several other test organisms. In the United States, the (DOE) initially, and the NIH soon

thereafter, were the main research agencies within the U.S. government responsible for developing and planning the project. By 1988, the two agencies were working together, a relationship that was formalized by the signing of a Memorandum of Understanding to "coordinate research and technical activities related to the human genome." The initial planning process culminated in 1990 with the publication of a joint research plan. The HGP started in 1990 as a 15-year program. One of its main coordinators was Dr. Francis Collins, Director of the National Human Genome Research Institute (NHGRI) at the NIH. Soon, the Human Genome Organization (HUGO) spread to Europe (HUGO Europe) and Japan (HUGO Pacific) to coordinate the research performed in other countries outside of the United States. The HGP was an international research program designed to construct detailed genetic and physical maps of the human genome, to determine the complete nucleotide sequence of human DNA, and to localize the estimated 50,000–100,000 genes within the human genome. The similar analyses on the genomes of several model organisms were also planned to be performed. The scientific products of the HGP were suggested to comprise a resource of detailed information about the structure, organization, and function of human DNA, information that constitutes the basic set of inherited "instructions" for the development and functioning of a human being. Achievement of the ambitious aims of HGP would demand the development of new technologies and necessitate advanced means of disseminating information widely to scientists, physicians, and others in order that the results may be rapidly used for the public good. Important for the success of the HGP was the statement that a fraction of the project money should be devoted to the ethical, legal, and social implications of human genetic research. It was clearly recognized that acquisition and use of such genetic knowledge would have momentous implications for both individuals and society and would pose several policy choices for public and professional deliberation. The HUGO program aims at a detailed understanding of the organization of the human genome, to enable a molecular definition of the basis for normal function as well as genetic disorders affecting human cells and the organism as a whole. An intermediate aim in this program was the construction of a high-resolution physical map for the human genome. At the beginning of this work, with sequencing still slow and expensive, the HGP adopted a "map–first, sequence later strategy." In the early 1990s, two Parisian laboratories, the Centre d'Etude du Polymorphisme Humain and Genethon, had an integral role in mapping of the human genome. The laboratories' driving forces were Drs. Daniel Cohen and Jean Weissenbach. Later the genome project constructed a higher resolution map that was used to sequence and assemble the human genome. In July 1995, a team led by Dr. J. Craig Venter published the first sequence of a free-living organism, *Haemophilis influenzae* (1.8 Mb) [5]. In October 1996, an international consortium publicly released the complete genome sequence of the yeast *Saccharomyces cerevisiae* (12 Mb) [6]. In May 1998, Venter announced a new company named Celera and declared that it will sequence the human genome within 3 years for $300 million. In December 1998, Drs. John Sulston, Robert Waterston, and colleagues completed the genomic sequence of *Caenorhabditis elegans* (100 Mb) [7]. The first complete human chromosome sequence—number 22 (33 Mb)—was published in December 1999 [8], and in June 2000, leaders of the public project and Celera announced completion of a working draft of the human genome sequence. Finally, in February 2001, the HGP consortium

published its working draft of the human genome sequence in *Nature* and Celera published its draft in *Science* (2.9 billion bp) [9, 10].

## 4 VECTORS. DESCRIPTION OF THE MAIN TYPES OF VECTORS USED IN THE HGP

In this section, common vectors used for everyday cloning and subcloning like plasmid and M13 vectors will not be discussed. Only vectors suitable for cloning relatively large DNA inserts will be mentioned with the exception of lambda phage-based vectors specially designed to clone cDNA.

### 4.1 Lambda-Based Vectors and Main Approaches to Construct Genomic Libraries

Numerous modifications have been made to bacteriophage lambda-based vectors to facilitate their handling and to extend their use for new types of biological experiments [11–15]. The application of each vector is usually limited to a specific task: the construction of general genomic libraries that contain all genomic DNA fragments or special genomic libraries that contain only a particular subset of genomic DNA fragments. Among these special libraries, NotI linking and jumping libraries have particular value for physical and genetic mapping of the human genome.

Despite their obvious advantages for special purposes, novel vectors (YACs, BACs, PACs; see below) cannot compete with lambda-based vectors for ease of handling, screening, amplification, and biological flexibility, due to the almost complete knowledge about phage lambda genes and biology. The easiest way to obtain a maximal proportion of recombinant molecules is to perform ligation at a high concentration of vector and insert (genomic) DNA because an elevated concentration of DNA facilitates intermolecular ligation instead of self-ligation of the vector's molecules. In this case, the main product is long DNA chains (>200 kbp) containing many copies of vector and genomic DNA fragments. Extremely efficient *in vitro* systems for packaging such DNA into lambda phage particles ($10^9$ plaque-forming units per microgram of DNA) to produce viable phages were developed and are commercially available.

Extensive modifications of lambda phage vectors were developed that combine the features of different vector systems (see Figures 1 and 2, [13–15]).

Cosmids are essentially plasmids that contain the *cos* region of phage lambda responsible for packaging of DNA into the phage particle. Any DNA molecule, containing this region and having a size between 37.7 kb and 52.9 kb, can be packaged into phage particles and introduced into the *E. coli*. The advantages of cosmids are their easy handling (as with plasmids) and big cloning capacity. As the plasmid body is usually small (3–6 kb), large DNA molecules (46–49 kb) can be cloned in these vectors.

Phasmids are lambda phages that have an inserted plasmid. They have the same basic features as lambda phage vectors, but inserted DNA fragment can be separated from the phage arms DNA and converted into a plasmid form. There are two main ways for such conversion: biological and enzymatic. In the first case, the

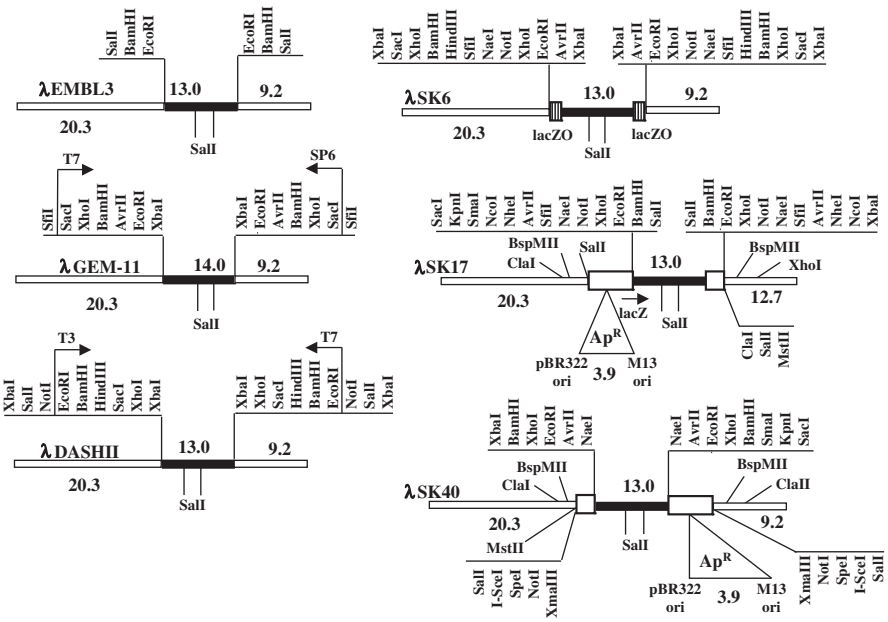


**Figure 1.** Examples of vectors used for construction of genomic libraries (maximum capacity 24 kb). Standard λ vectors (λEMBL3, λGEM11, λDASH, λEMBL3, and λSK6) and diphasmid vectors (λSK17 and λSK40) are shown. Sizes are in kilobases and are not shown in the scale. Not all restriction sites are shown. Heavy thick lines represent stuffer fragments; open boxes mark plasmid and M13 sequences; *lacZO* is the *lac* operator sequence. T3, T7, SP6-promoters for T3, T7, and SP6 RNA polymerases.

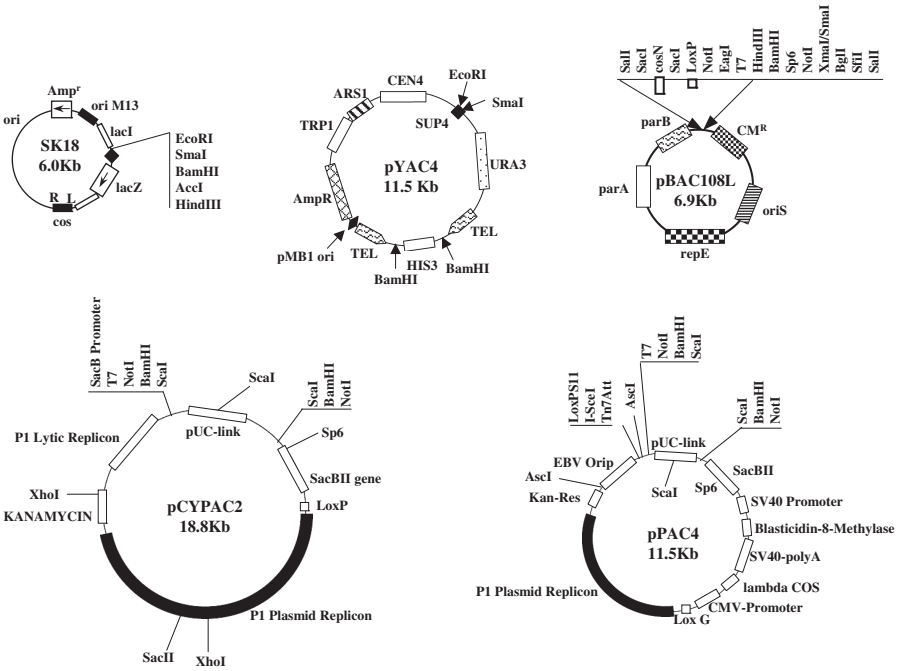


**Figure 2.** Examples of vectors used for construction of genomic libraries (capacity more than 45 kb). Sizes are in kilobases and are not shown to scale. Not all restriction sites are shown. For more details, see the text.

phasmid vector contains signal sequences bordering the insert and plasmid body in the phage vehicle. These signal sequences (e.g., from P1, lambda, M13 phages) can be recognized by specific proteins (e.g., *cre*-recombinase), and the cloned DNA fragment together with plasmid body will be cut out [16, 17].

The enzymatic conversion (see, for example, SK17 and SK40 vectors in Figure 1) can occur if the inserted DNA fragment and plasmid sequences are placed between two recognition sequences of some rare-cutting restriction endonuclease (e.g., I-SceI enzyme that recognize 18 bp and probably does not have any recognition site in the human genomic DNA). This enzyme can be used to separate the cloned DNA fragment together with the plasmid sequences from phage arms. Self-ligation of the molecules with insert and plasmid sequences will result in the production of a recombinant plasmid that can be introduced into the *E. coli* host.

Diphasmids combine the advantages of lambda and M13 phages, and plasmids, i.e., the three main types of vectors used in everyday molecular cloning. They can be divided into two classes: (1) diphasmids that can replicate as phage lambda (a further improvement of phasmids); see SK17 in Figure 1; and (2) diphasmids that are not able to replicate as phage lambda, i.e., a cosmid that can be packaged into phage M13 particles, e.g., SK18 in Figure 2.

In some cases, it is more convenient to work with a genomic library in plasmid than in lambda phage form. The construction of representative genomic library directly in a plasmid vector has several drawbacks and difficulties. However, these problems can be easily solved with the help of phasmid and diphasmid vectors. In this case, a genomic library is constructed in lambda phage (e.g., SK40), and then the whole library is converted to plasmid form.

Although the work with different lambda-based vectors is very similar, vectors for construction of genomic and cDNA libraries still have some special features. The major difference between cDNA and genomic cloning is the size of insert: short for cDNA and long for genomic fragments. Due to this difference, it is obvious that the enzymatic way of transferring inserts into the plasmid form is more dangerous (with the respect to the representativity of the library) for genomic libraries than for cDNA, because the probability that the insert contains a recognition site for the particular restriction enzyme increases with the length of the DNA insert. Most of cDNA vectors are "insertion" vectors, and genomic vectors are "substitution" vectors, which means that inserted foreign DNA fragments replace the stuffer vector fragment that does not have any important replicative function.

Full-length cDNAs are the most important material for the identification of all human genes. Cloning of full-length cDNA inserts has been hampered by problems related to both the preparation and the cloning of long cDNAs. Part of the difficulty associated with the preparation of long cDNAs has been overcome with the introduction of a thermostabilized and thermoactivated reverse transcriptase and the development of cap-based full-length cDNA selecting techniques. In contrast to standard cloning techniques, full-length cDNA cloning has the inherent risk of the under-representation or absence of clones corresponding to long mRNAs in the libraries, because the truncated cDNAs are usually not cloned. However, even in a perfect full-length cDNA library, cDNAs deriving from very long mRNAs will not be cloned if the capacity of the vector is insufficient. The most available cloning vectors show bias for short cDNAs: Shorter fragments are cloned more efficiently than the longer because of the intermolecular competition that occurs at the

ligation and library amplification steps. To solve this problem, substitution vectors for cDNA cloning were designed.

For conversion into the plasmid form, the biochemical approach for SK-vectors was used [14]. Some vectors, like SK16 and SK17, have a cloning capacity of 0.2–15.4 kb and others, for example SK23, could be used for cloning of 3.7–18.9-kb inserts only. These cDNA vectors retain the most important advantages of the genomic vectors, e.g., the possibility of biochemical selection when producing vector arms and genetic selection against nonrecombinant phages. Importantly, the libraries made in some of these SK-vectors can be used as both expressed and nonexpressed libraries. For instance, libraries constructed in SK16 would be expressed only in *E. coli* strains containing *ochre*-codon mutation, and SK17 has *amber*-codon-controlled expression of cDNA.

In lambda FLC (full-length cDNA cloning), cloning capacity is either 0.2–15.4 kb or 5.7–20.9 kb. In these vectors, *cre*-recombinase mediated excision was used [17]. The average size of the inserts from excised plasmid cDNA libraries was 2.9 kb for standard and 6.9 kb for size-selected cDNA. The average insert size of the full-length cDNA libraries was correlated to the rate of new gene discovery, suggesting that effectively cloning rarely expressed mRNAs requires vectors that can accommodate large inserts from a variety of sources without bias for cloning short cDNA inserts.

One of the most important features of a genomic library is a representativity. A representative genomic library means that every genomic DNA fragments will be present at least in one of the recombinant phages of the library. In practice, however, this is difficult to achieve. Some genomic fragments are not clonable because of the strategy used for construction of the genomic library. In other cases, genomic DNA fragments (contain poison sequences) can suppress the growth of the vector or the host cell so its cloning can be restricted to specific vector systems. The important reason for the decreased representativity is different replication potential of different recombinant vectors. Thus, amplified libraries have significantly worse representativity. Usually a library is considered to be representative if after the first plating (before amplification) it contains several recombinant clones together containing genomic DNA fragments equal to a 7–10 genome equivalent. The way in which the genomic DNA fragments are produced for cloning is also important. The more randomly the genomic DNA is broken, the more the representative library can be obtained. Clearly, the EcoRI enzyme (6-bp recognition site) will cut genomic DNA less randomly than Sau3AI (4 bp recognition). Probably, the shearing of DNA molecules using physical methods (sonication, shearing using syringe) is the most reliable way to obtain randomly broken DNA molecules.

There are many different modifications to construct libraries; however, the three ways to construct genomic libraries are the most commonly used and will be exemplified below using genomic lambda-based vectors (Figure 3, [12–14]).

The classic method includes generation of sheared genomic DNA fragments using physical or enzymatic fragmentation followed by the physical separation of fragments of a particular size using, for example, ultra-centrifugation or gel-electrophoresis. The vector DNA is digested with two (or even three) restriction enzymes; those recognition sites are located in the polylinker. The arms and the stuffer fragment are purified, and on the further steps, arms/stuffer would be not able to ligate because they have different sticky ends, preventing recreation of the original vector molecules during subsequent ligation with genomic DNA fragments.

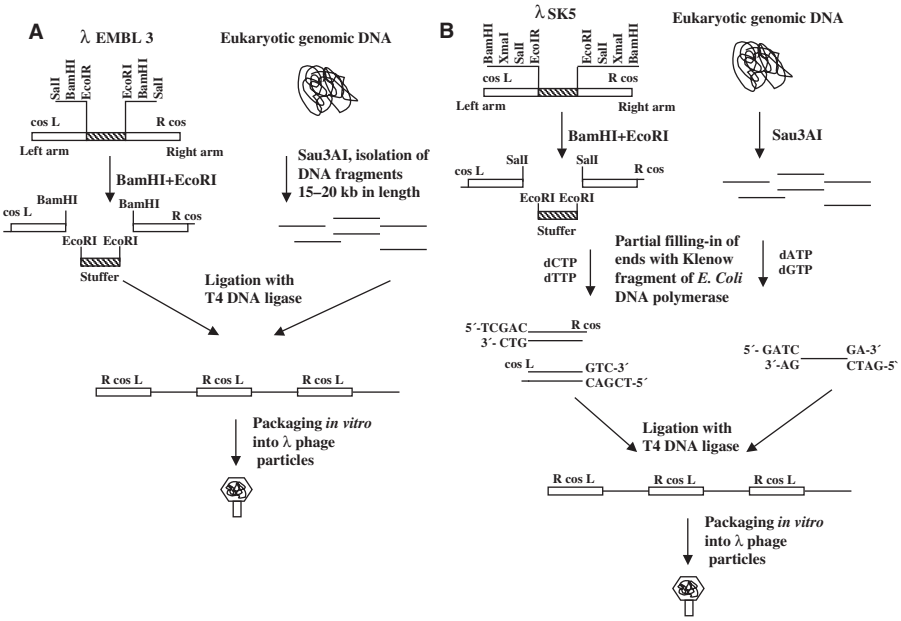


**Figure 3.** Two main approaches for construction of genomic libraries: (a) classic method and (b) partial filling-in method. cosL and Rcos—left and right parts of the *cos* site, respectively.

In the "dephosphorylation" approach [18, 19], the phage arms are prepared as described above. Genomic DNA is partially digested to obtain DNA fragments with sizes in the desired range; i.e., for the lambda vectors FIXII and SK6, it is 15–22 kb. These DNA fragments are dephosphorylated (which prevents their ligation to each other) and then ligated to the vector arms.

The third, "partial filling in" method [20], also avoids fractionation steps (Figure 3B). Phage arms are prepared by double digestion as described before (in this particular case, SalI and EcoRI are shown, but many other combinations can be used), and the sticky-ends produced after digestion are partially filled-in with the Klenow fragment of DNA polymerase I (or other DNA polymerase) in the presence of dTTP and dCTP. Genomic DNA partially digested with Sau3AI is also partially filled-in but in the presence dATP and dGTP. Under these conditions, self-ligation of vector arms or genomic DNA is impossible.

## 4.2 YACs

The mapping and analysis of complex genomes require vectors that allow cloning and work with large DNA fragments. Lambda-based vectors have many advantages as described above; however, a comparatively small size of inserts (maximum 48–49 kb) complicates the use of these vectors for the mapping of complex genomes.

Technological improvements made the cloning of large DNA pieces possible using artificially constructed chromosome vectors that carry human DNA fragments as large as 1Mb. Such vectors are maintained in yeast cells as artificial chromosomes (YACs). YAC methodology drastically reduces the number of clones to

be ordered; many YACs span entire human genes. A more detailed map of a large YAC insert can be produced by subcloning, a process in which fragments of the original insert are cloned into smaller insert vectors. YAC cloning was widespread, and many laboratories used this technique to obtain large cloned DNA fragments from different eukaryotic genomes.

One of the most widely used YAC vectors was pYAC4 (Figure 2), and it has the main features of the YAC vectors. This vector can replicate in *E. coli* and in yeast (*S. cerevisiae*). It consists of five important genetic elements inserted into *E. coli* pBR322-like plasmid:

1. Centromeric *CEN4* sequences required for centromere function.
2. Autonomous replicating sequences *ARS1* that are necessary for replication in yeast.
3. Telomere sequences from *Tetrahymena* ribosomal DNA (rDNA) that seed the formation of functional telomeres at high efficiency.
4. Selective marker for cloning, *SUP4* (*ochre*-suppressing allele of tyrosine tRNA gene). As cloning EcoRI and SmaI sites are located inside *SUP4*, insertion of any large DNA fragment into the cloning sites will inactivate *ochre* suppression function.
5. Two yeast selectable markers: wild-type genes *TRP1* and *URA3*, which are located on opposite sides of the cloning sites.

The yeast *HIS3* gene is just a stuffer fragment to separate two arms. The general scheme of the library construction is as follows (Figure 4).

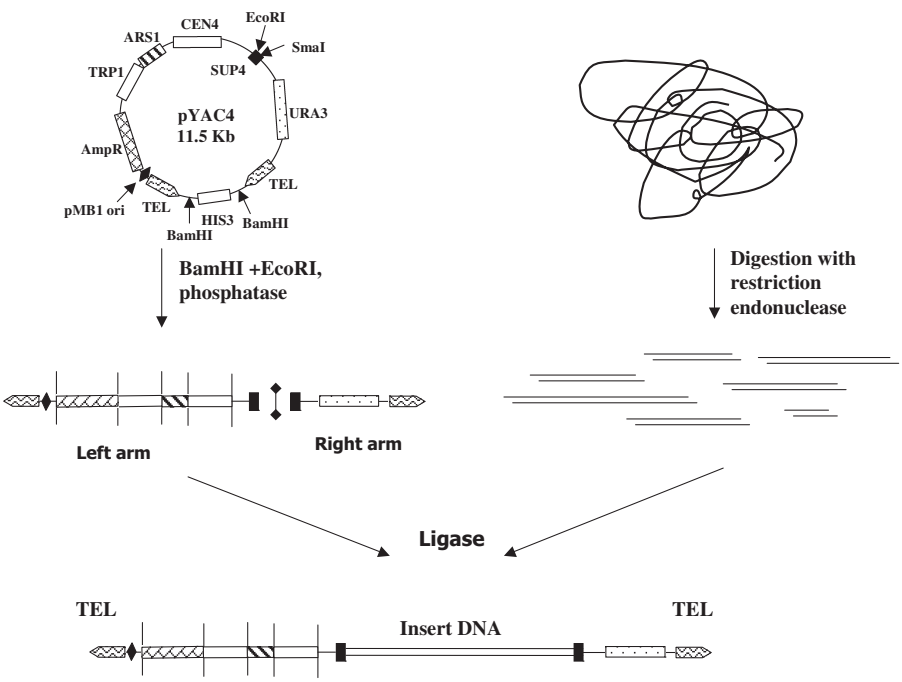


**Figure 4.** Construction of genomic library in YAC vectors. See text for more details.

The pYAC4 DNA is double digested with BamHI and SmaI (or EcoRI) to generate three fragments: left chromosomal arm containing *TEL*, *TRP1*, *ARS1*, and *CEN4* and right chromosomal arm with *TEL* and *URA3*. *HIS3* gene is discarded during the cloning procedure. The two arms are treated with alkaline phosphatase to prevent self-ligation and ligated with genomic DNA (usually) partially digested. The ligation product is transformed into yeast and selected for complementation of a mutated *ura3* gene present in yeast strain by the *URA3* gene in the vector. The grown transformants are selected for complementation of a host *trp1* mutated gene by the wild-type *TRP1* gene from the vector.

This double selection is necessary to ensure the presence in a recombinant YAC of both left and right arms and inactivation of the *SUP4* gene. *SUP4* is an especially useful cloning marker because in an *ade2-ochre* host, the cells with the functional *SUP4* gene are white and those in which the suppressor was inactivated form red colonies.

Using YAC vectors, several important genes were identified and cloned, e.g., *VHL*, *CDPX1*, *GART*, *and SON* [21–23]. Moreover, the first physical map covering the whole human genome was constructed with YAC contigs [24].

Although YACs greatly contributed to the progress in long-range mapping of the human genome, the construction of YAC contigs is not free from problems—certain regions of the human genome are difficult to clone in YACs. Moreover, repeat rich human sequences cloned in YACs are frequently unstable in yeast. Low transformation efficiency, the presence of an abundance of chimeric clones in nearly all YAC libraries, insert instability, and difficulties in DNA manipulation relative to bacterial systems are other important limitations.

Efforts to overcome the limitations of cosmids and YACs resulted in alternative large-insert cloning approaches using bacteriophage P1-based vectors and vectors based on the *E. coli* fertility plasmid (F-factor).

### 4.3 Bacterial (BAC) and P1-derived (PAC) Artificial Chromosome Vectors

BACs represent the state-of-the-art technology for such large-insert DNA library development. It has been demonstrated that BAC libraries are invaluable and desirable genetic resources for all kinds of modern structural, functional, and evolutionary genomics research. Genome-wide, as well as regional, physical maps of the human genome from BACs have been developed in different laboratories.

The BAC system is based on the well-studied *E. coli* F factor. Replication of the F factor in *E. coli* is strictly controlled. The F plasmid is maintained in low copy number (one or two copies per cell), thus reducing the potential for recombination between DNA fragments carried by the plasmid. Furthermore, F factors carrying inserted bacterial DNA are capable of maintaining fragments as large as 1 megabase pair, suggesting that the F factor is suitable for cloning of large DNA fragments. Individual clones of human DNA seem to be maintained with a high degree of structural stability in the host, even after 100 generations of serial growth.

The F factor not only codes for genes that are essential to regulate its own replication but also controls its copy number. The regulatory genes include *oriS*, *repE*, *parA*, and *parB*. The *oriS* and *repE* genes mediate the unidirectional replication of the F factor, whereas *parA* and *parB* maintain a copy number at a level of one or

two per *E. coli* genome. The BAC vector (pBAC108L) incorporates these essential genes as well as a chloramphenicol resistance marker and a cloning segment (Figure 2). The cloning segment includes (1) the bacteriophage lambda cosN and phage P1 loxP sites, (2) two cloning sites (HindIII and BamHI), and (3) several CG-rich restriction enzyme sites (Not I, Eag I, Xma I, Sma I, Bgl I, and Sfi I) for potential excision of the inserts. The cloning site is flanked by T7 and SP6 promoters for generating RNA probes for chromosome walking and for DNA sequencing of the inserted segment at the vector–insert junction. The cosN site provides a fixed position for specific cleavage with the bacteriophage lambda terminase. The loxP site can be used similarly. In this case, P1 Cre protein catalyzes the cleavage reaction in the presence of the loxP oligonucleotide. These sites (cosN and loxP) allow convenient generation of ends that can be used for restriction-site mapping to arrange the clones in an ordered array. Restriction maps of the individual clones can be determined by indirect end-labeling and subsequent partial digestion.

PAC vectors combine the best features of the P1 and BAC systems. One of the most commonly used PAC vectors is pCYPAC2 (Figure 2) that was constructed for the cloning of large DNA fragments using electroporation. Previously developed P1-based vectors (like pAd10SacBII) employed bacteriophages P1 or T4 *in vitro* packaging systems, enabling the cloning of recombinants with inserts in the 70–120-kb range.

The pCYPAC2 vector was constructed by removing the stuffer fragment from the pAd10SacBII vector and by inserting a pUC19 plasmid into the BamHI cloning site. During the cloning process, the pUC19 sequences are removed through a double-digestion scheme using BamHI and ScaI. Religation of the pUC19 and producing of nonrecombinant clones is prevented at three levels, as follows:

1. pUC19 is cleaved into two ScaI fragments. ScaI sticky ends are not compatible with BamHI.
2. Oligonucleotides BamHI–ScaI connecting vector sequences with the stuffer fragment are physically removed from the vector and stuffer fragments.
3. pCYPAC2 fragments are treated with alkaline phosphatase to inhibit self-ligation.

NotI sites are flanking inserts that make it possible to obtain an insert that is fragment free from the vector sequences.

Different modifications of original BAC and PAC vectors were constructed, e.g., capable of replicating in human cells too [25]. These vectors, e.g., pPAC4 (Figure 2) and pBACe4, facilitate the use of large-insert bacterial clones for functional analysis and contain two additional genetic elements that enable stable maintenance of the clones in mammalian cells:

(1) The Epstein–Barr virus replicon, *oriP*. It was included to ensure stable episomal propagation of the large insert clones upon transfection into mammalian cells.
(2) The blasticidin deaminase gene is placed in a eukaryotic expression cassette to enable selection for the desired mammalian clones by using the nucleoside antibiotic blasticidin.

Sequences important to select for loxP-specific genome targeting in mammalian chromosomes were also inserted into the vectors. In addition, the attTn7 sequence present on the vectors permits specific addition of selected features to the library clones. Unique sites have also been included in the vector to enable linearization of the large-insert clones, e.g., for optical mapping studies. The pPAC4 vector has been used to generate libraries from the human, mouse, and rat genomes.

Libraries constructed in lambda-based, BAC, and PAC vectors are widely used now in different laboratories, and almost nobody is currently working with libraries constructed in YAC vectors.

## 5 MAPPING OF THE HUMAN GENOME

As mentioned in the Introduction, mapping of the human genome was a major step in achieving the complete human genome sequence. Human genome maps were also important by themselves because they led to the isolation of many disease genes.

The 3 billion bp in the human genome are organized into 24 distinct, physically separate microscopic units called chromosomes. All genes are arranged linearly along the chromosomes. The nucleus of most human cells contains two sets of chromosomes, one set given by each parent. Each set has 23 single chromosomes—22 autosomes and an X or Y sex chromosome. A normal female will have a pair of X chromosomes; a male will have an X and Y pair. Chromosomes contain roughly equal parts of protein and DNA, and chromosomal DNA contains on average 150 million bases.

In general, all maps can be divided in two main classes: genetic and physical [see Refs. 3 and 4]. The genetic map is based on the frequency of recombination between two genetic markers on the chromosome. In fact, such a map just shows how these two markers are "linked" in respect to the frequency of recombination. Physical maps are based on physical distances between markers. Recombination frequencies define a genetic distance that is not the same as a physical distance. Two loci that show 1% recombination are defined as being 1 centimorgan (cM) apart on a genetic map. However, for distances above about 5 cM, human genetic map distances are not simple statements of the recombination fraction between pairs of loci. Loci that are 40 cM apart will show less than 40% recombination. This result reflects that recombination fractions never exceed 50% no matter how far apart are the loci. The female genetic map is larger than male due to a larger frequency of recombination. In general, for the whole human genome, the sex-averaged figure is 1 cM = 0.9 Mb, but the actual correspondence varies very significantly in different chromosomal regions. Usually there are more recombinations toward the telomeres of chromosomes and less toward the centromeres.

The value of the genetic map is that an the inherited disease can be located to the particular chromosomal region exploiting the inheritance of a DNA marker present in affected individuals (but absent in unaffected individuals), even though the molecular basis of the disease may not yet be understood nor the responsible gene identified. Genetic maps have been used to find the exact chromosomal location of several important disease genes, including cystic fibrosis, sickle cell disease, Tay-Sachs disease, fragile X syndrome, and myotonic dystrophy. Currently, some

genetic linkage maps contain more than 100,000 polymorphic markers (see, for example, http://www.ncbi.nlm.nih.gov/SNP/).

Different types of physical maps are based on different techniques. For instance, the cytogenetic map is based on detection chromosomal bands, and an average band has from one to several Megabase of DNA. Chromosomal breakpoint maps can be based on somatic cell hybrid panels containing human chromosome fragments derived from natural translocation or deletion chromosomes. The resolution for such maps is usually several Megabases. Monochromosomal radiation hybrid (RH) maps have a distance between breakpoints that is often several Megabases long. Whole genome RH maps have much higher resolution and can be as high as 0.5 Mb. The RH maps were very important for joining genetic and physical maps. They are based on *in vitro* radiation-induced chromosome fragmentation and cell fusion (merging human with other cells from other species to form hybrid cells) to create panels of cells with specific and varied human chromosomal components. Assessing the frequency of marker sites remaining together after radiation-induced DNA fragmentation can establish the order and distance between the markers.

Very useful were clone contig maps that could be constructed with YAC (or BAC, PAC) clones, and in this case, the average DNA insert has several hundred kilobases of DNA [24, 26]. If such a clone contig map was created using overlapping cosmid clones (average insert is 40–45 kb), the quality of such a map is significantly higher. A restriction map created with rare-cutting restriction enzymes like NotI can be very useful to check RH or clone contig maps [27].

A sequence-tagged site (STS) map requires prior sequence information and is based on mapping short known sequences using PCR amplification.

An expressed sequence tag (EST) map is based on STSs that were generated from cDNA.

DNA fragments are called DNA markers after they were mapped to particular chromosomal regions. Such DNA markers are prerequisites for physical and genetic mapping of the genome. DNA markers are also important in the diagnosis of genetic diseases. DNA markers can be divided into several different classes depending on the way in which the markers were selected among the fragments of genomic DNA. Examples of such classes are anonymous, micro- and minisatellites, restriction fragment length polymorphism (RFLP) markers, *NotI* linking clones, ESTs, STSs, and so on.

Another strategy of mapping is based on the completely different principle of NotI jumping and linking libraries [28, 29].

A big advantage to using chromosome jumping and linking clones is that they are small-insert vectors and mapping with these clones can be completely automated.

The basic principle of jumping is to clone only the ends of large DNA fragments rather than a continuous DNA segment (Figure 5). The DNA between two ends is deleted with the different biochemical techniques, and clones containing only the ends of a large DNA molecule are enriched by the mean of genetic or biochemical selection.

In general, jumping clones contain DNA sequences adjacent to neighboring NotI restriction sites, and linking clones contain DNA sequences surrounding the same NotI restriction site.

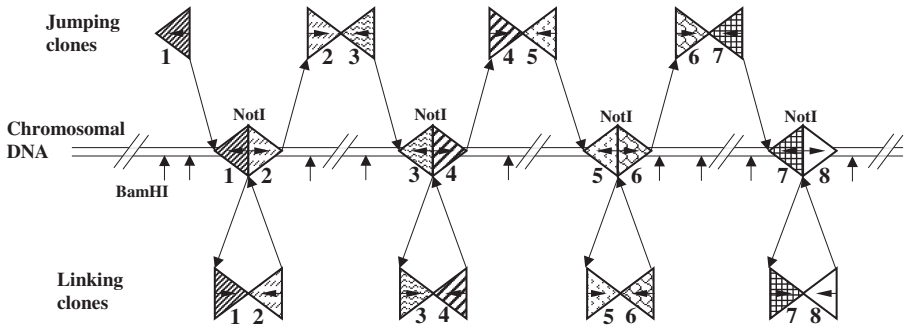


**Figure 5.** General scheme of long-range mapping using jumping and linking libraries. Small vertical arrows designate BamHI, and horizontal arrowheads show short sequences flanking NotI sites. Triangles indicate NotI–BamHI boundary fragments. See text for details.

A NotI jumping library is a collection of cloned DNA fragments that contains the ends of large NotI genomic DNA fragments. A NotI linking library is a collection of cloned genomic DNA fragments flanking the same NotI site. Basically, a linking library in combination with pulsed field gel electrophoresis (PFGE) is sufficient to construct a physical chromosomal map. When linking clones are used to probe a PFGE genomic blot, each clone should reveal two DNA fragments, and they must be adjacent in the genome. In principle, with just a single library and digest, one should be able to order the rare cutting sites, but in reality, it is not possible because many fragments could have the same size. Using these two types of libraries, it is possible to move fast along the chromosome because the human genome contains a limited number of recognition sites for this enzyme (10,000–15,000). The easiest way to establish the order of NotI sites is a shotgun sequencing of NotI linking and jumping clones (see below and Figure 5).

The most well-known and efficient way (integrated approach) to construct NotI linking and jumping libraries is shown in Figure 6 [30, 31].

To construct linking libraries (Figure 6B), genomic DNA is completely digested with BamHI. (Libraries can also be constructed with BglII, EcoRI, etc., but not with frequently cutting enzymes like MboI or infrequently cutting enzymes like SalI.) Subsequently, the DNA is self-ligated at a low concentration of DNA (without a *supF* marker), to yield circular molecules as the main product. To eliminate any remaining linear molecules, the sticky ends are partly filled-in with the Klenow fragment in the presence of dATP and dGTP (this is for the genomic DNA digested with BamHI; for other enzymes, partial filling-in with other nucleotides should be used). This way, all BamHI sticky ends will be neutralized and nearly all ends originated from linear molecules will be unavailable for ligation. Resulting DNA fragments are digested with NotI and cloned into λSK4, λSK17, and λSK22 vectors. These vectors permit the cloning of DNA fragments from 0.2 to 24 kb. The resulting phage particles are used to infect *E. coli* cells in which only recombinant phages could grow (spi- selection, [12]).

To construct the NotI jumping library, the same DNA and vectors can be used. However, the first steps are different. High-molecular-weight DNA is digested with

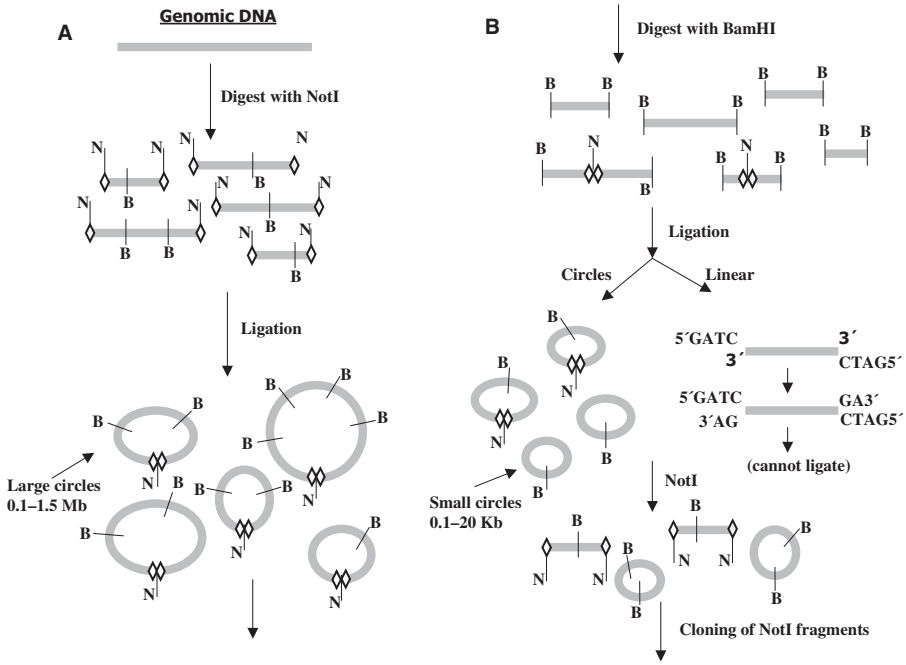


**Figure 6.** Construction of jumping (a and b) and linking libraries (b). B represents BamHI, and N denotes NotI sites. Digestion of genomic DNA with BamHI is a first step in construction of a linking library.

NotI and circularized at very low DNA concentration (0.1 μg/μL). Other steps were the same (Figure 6A).

The libraries can be converted into plasmid form, and sequences flanking NotI sites can be generated with standard sequencing primers. This sequence information can be used for generation of sequence tagged sites (STS). It was also demonstrated that the sequencing of these clones is a very efficient method for gene isolation because even short nucleotide sequences (approximately 500 bp from each side) flanking NotI sites are sufficient to detect genes [32, 33]. This approach was used for isolation of numerous new genes. The reason for this linkage of NotI sites and genes is the localization of practically all NotI sites in the CpG islands, which are tightly associated with genes.

A shotgun sequencing approach useful for the whole genome mapping was proposed for NotI linking and jumping libraries constructed in λSK17 and λSK22 vectors using an integrated approach [14, 29, 34]. This strategy was based on the fact that NotI sites in both libraries were available for the sequencing with standard sequencing primers, and sequence information about 800–1000 bp surrounding each NotI site could be easily obtained by automated sequencing. Plasmid DNA could be isolated automatically, and thousands of clones could be sequenced. Subsequently the linear order of the NotI clones can be established using a computer program. This approach is schematically illustrated in Figure 5.

NotI linking and jumping libraries were very important for the mapping of two regions in the short arm of chromosome 3 and isolation of tumor suppressor genes from the AP20 and LUCA regions (*RASSF1*, *G21*, *SEMA4B*, *RBSP3*, etc.).

## 6 MAIN APPROACHES TO SEQUENCE HUMAN GENOME

### 6.1 Hierarchical and the Whole Genome Shotgun Sequencing Scheme (WGS)

The international public (HGP) and American private (Celera) efforts to determine the sequence of the human genome exploited different approaches [9, 10, 35]. It is still not clear whether Celera's whole genome shotgun sequencing (WGC) approach was really efficient as they used publicly available data and it is not known to what extent.

The major points in the hierarchical or HGP approach [9] (Figure 7A) are as follows:

1. To construct a precise high-density STS map of the human genome. This map was integral and included not only STS but also genetic, microsatellite, FISH, and all other available mapping information.
2. To construct the clone contig of overlapping BAC and PAC clones. This contig was created and checked using a map constructed at the first step.
3. Each BAC or PAC clone was sequenced using a shotgun sequencing strategy. These sequences were assembled into the draft human genome sequence.

The shotgun sequencing approach for sequencing relatively small inserts was developed a long time ago [3, 4]. This method is based on the building of a complete

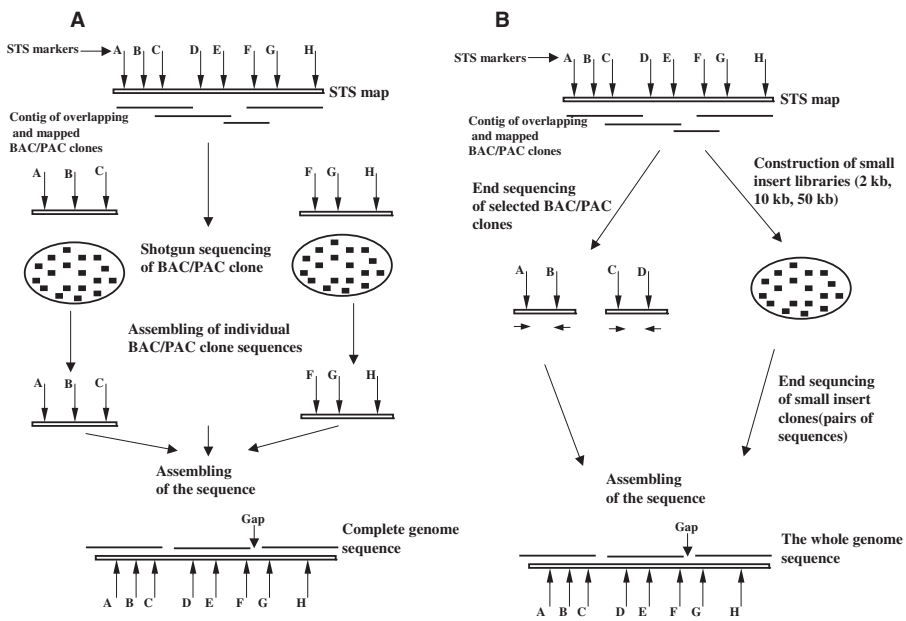


**Figure 7.** Two main approaches used to sequence the human genome. The hierarchical approach is based on creation of clone contigs covering the whole genome and shotgun sequencing of individual cosmid/BAC/PAC clones (a). The whole genome shotgun sequencing scheme (WGS) shown in (b) is based on pair-end sequencing of small-insert and selected BAC/PAC clones. See text for details.

insert sequence by randomly sequencing clones with overlapping inserts. To do this the insert DNA is subjected to partial digestion with a 4-bp cutter (like MboI) and the partially overlapping fragments are cloned at random into a suitable M13 or plasmid vector and sequenced. The sequence data obtained are fed into a computer programmed to detect overlaps between sequences and to assemble a composite sequence. Inevitably, this means a waste of efforts: The same sequence will be obtained over and over again.

Celera used a variant of this strategy that it named WGS [10] (Figure 7B). Central to the WGS strategy is preparation of high-quality plasmid libraries in a variety of insert sizes so that pairs of sequence reads (mates) are obtained; one is read from both ends of each plasmid insert. High-quality libraries have an equal representation of all parts of the genome and are constructed in three different sizes: 2 kb, 10 kb, and 50 kb. The method also involved end sequencing of some PAC clones that were carefully mapped. DNA fragments in these small-insert clones were generated by physical shearing of whole genomic DNA. According to the original predictions, this method requires the generation of sequences covering the whole genome not more than seven times.

The HGP and Celera announced that they have obtained a sequence of 92–97% of the human genome sequence and that a 7–10-fold genome sequence coverage is enough to generate an almost complete sequence for any human or mammalian genome. After 5 years, the scientific community understood that these estimations were overstated (see below), and even now when more than 10-fold genome sequence coverage was achieved, the human genome sequence is far from being complete.

An important point of both approaches is that the genome should be covered by sequences at least 10 times to yield contigs covering a significant part of the genome. If the coverage is not extensive enough, then sequences and clones will represent unconnected and unordered islands. Thus, despite impressive progress, mapping and sequencing even of small genomes, like from one bacterial strain, is still expensive and laborious. As a result, genome mapping and sequencing efforts need to be concentrated to big centers, and independent work by smaller groups is seriously hampered.

Partly in view of these logistic and economic limitations, HGP and Celera approaches may not be the optimal solution, especially for several experiments addressing biocomplexity. After completion of one sequence from an organism, in many cases there will be a great demand for comparison with other individuals, related species, pathogenic and nonpathogenic strains, and so on, in the growing field of comparative genomics. Such comparisons will be highly relevant for the better understanding of health, evolution, and ecology questions but must be performed by laboratories receiving considerably less funding than those involved in the recent high-profile sequencing efforts. Another area is a sequencing of related bacterial strains and species in order to identify the genomic basis of their biological differences and their interactions within the human intestinal flora. What is a difference between pathogenic and nonpathogenic members of the gut flora?

The whole genome sequencing will be performed only for the selected organisms. However, analysis of disease genes in other species may be extremely useful for the understanding of fundamental processes, leading to development of disease.

As the body of knowledge concerning the genome structure and genes in different organisms grows, it also becomes unnecessary to collect total sequences with 10–15-fold coverage for all organisms that would be interesting to study.

One of the most perspective solutions for this problem is sequencing using microarrays [36]. However, this approach will be not discussed here as it cannot be a strategy for the *de novo* sequencing. Two other alternative approaches that could be used for both the *de novo* and the comparative sequencing will be described below.

### 6.2 Massively Parallel Sequencing Approaches

The first publication using beads for massively parallel sequencing was published by Dr. Sydney Brenner et al. (2000) [37]. They described a novel sequencing approach that combined non-gel-based signature sequencing with *in vitro* cloning of millions of templates on separate 5-μM-diameter microbeads. After constructing a microbead library of DNA templates by *in vitro* cloning, a planar array of a million template-containing microbeads in a flow cell at a density greater than $3 \times 10^6$ microbeads/cm$^2$ was assembled. Sequences of the free ends of the cloned templates on each microbead were then simultaneously analyzed using a fluorescence-based signature sequencing method that does not require DNA fragment separation. Signature sequences of 16–20 bases were obtained by repeated cycles of enzymatic cleavage with a type IIs restriction endonuclease (BbvI), adaptor ligation, and sequence interrogation by encoded hybridization probes. The approach was validated by sequencing over 269,000 signatures from two cDNA libraries constructed from a fully sequenced strain of *S. cerevisiae*, and by measuring gene expression levels in the human cell line THP-1. This method is complicated and expensive and was used mainly for generation of SAGE tags (see below).

Recently MPSS principles were applied for long-range sequencing. Dr. George Church and his colleagues introduced a sequencer that uses a microscope and other off-the-shelf equipment [38]. With this technology, his team sequenced a strain of *E. coli* and could detect single-base-pair changes from an almost identical *E. coli* genome. The approach reduces sequencing costs by 90%. Dr. Jonathan Rothberg has demonstrated the power of another cost-cutting technology called "454" [39]. Using this approach, *Mycoplasma genitalium* was sequenced.

Both groups save money by eliminating the need for bacteria and miniaturizing the process wherever possible. Instead of bacteria, they attach DNA to aqueous beads encased in oil where PCR amplification was performed to produce the necessary amount of DNA. That change alone could reduce by two thirds the costs associated with space and personnel. Moreover, both perform many thousands of these sequencing reactions at once in miniature "reactors," decreasing the need for expensive chemicals. Once the DNA is ready, the two technologies diverge: The "454" technique uses pyrosequencing [40] to identify the bases and Church's technique (MPS, multiplex polony sequencing) uses bursts of different fluorescent colors, one each to a particular base, to distinguish the bases (sequencing by ligation, [41]). Both use high-speed charge-coupled device cameras to record the labeled bases.

The "454" technique uses a novel fiber-optic slide of individual wells and can sequence 25 million bases, at 99% or better accuracy, in one 4-hour run. To achieve

an approximately 100-fold increase in throughput over current Sanger sequencing technology, an emulsion method for DNA amplification and an instrument for sequencing by synthesis using a pyrosequencing protocol optimized for solid support and picolitre-scale volumes was developed. In a pilot experiment, 96% of the *M. genitalium* genome was covered by sequence contigs at 99.96% accuracy in one run of the machine. The method could potentially allow one individual to prepare and sequence an entire genome in a few days. The sequencer itself, equipped with a simple detection device and liquid delivery system, and housed in a casing roughly the size of a microwave oven, is actually relatively low-tech. The complexity of the system lies primarily in the sample preparation and in the microfabricated, massively parallel platform, which contains 1.6 million picoliter-sized reactors in a 6.4-$cm^2$ slide.

Sample preparation starts with fragmentation of the genomic DNA, followed by the attachment of adaptor sequences to the ends of the DNA pieces. The adaptors allow the DNA fragments to bind to tiny beads (around 28 μm in diameter). This is done under conditions that allow only one piece of DNA to bind to each bead. The beads are encased in droplets of oil that contain all reagents needed to amplify the DNA using a standard tool called the PCR. The oil droplets form part of an emulsion so that each bead is kept apart from its neighbor, ensuring the amplification is uncontaminated. Each bead ends up with roughly 10 million copies of its initial DNA fragment.

To perform the sequencing reaction, the DNA-template-carrying beads are loaded into the picoliter reactor wells—each well having space for just one bead. The technique uses a sequencing-by-synthesis method developed by Dr. Uhlen and colleagues [40], in which DNA complementary to each template strand is synthesized. The nucleotide bases used for sequencing release a chemical group as the base forms a bond with the growing DNA chain, and this group drives a light-emitting reaction in the presence of specific enzymes and luciferin. Sequential washes of each of the four possible nucleotides are run over the plate, and a detector senses which of the wells emit light with each wash to determine the sequence of the growing strand.

This new system shows great promise in several sequencing applications, including resequencing and *de novo* sequencing of smaller bacterial and viral genomes. It could potentially allow research groups with limited resources to enter the field of large-scale DNA sequencing and genomic research, as it provides a technology that is inexpensive and easy to implement and maintain. However, this technology cannot yet replace the Sanger sequencing approach for some of the more demanding applications, such as sequencing a mammalian genome, as it has several limitations [42].

First, the technique can only read comparatively short lengths of DNA, averaging 80–120 bases per read, which is approximately a tenth of the read-lengths possible using Sanger sequencing. This means not only that more reads must be done to cover the same sequence, but also that assembling short reads into longer genomic sequences is a lot more complicated. This is particularly true when dealing with genomes containing long repetitive sequences.

Second, the accuracy of each read is not as good as with Sanger sequencing—particularly in genomic regions in which single bases are constantly repeated. Third, because the DNA "library" is currently prepared in a single-stranded format,

unlike the double-stranded inserts of DNA libraries used for Sanger sequencing, the technique cannot generate paired-end reads for each DNA fragment. The paired-end information is crucial for assembling and orientating the individual sequence reads into a complete genomic map for *de novo* sequencing applications.

Finally, the sample preparation and amplification processes are still complex and will require automation and/or simplification.

In the MPS technique, sheared, size-selected genomic fragments (approximately 1 kb in size) were circularised with a linker bearing MmeI (type IIs restriction enzyme) recognition site (TCCRAC). Then all noncircularized material was destroyed with exonuclease and circular molecules were amplified using rolling circle amplification with random hexamers. Afterward, resulting material was digested with MmeI. MmeI digests at a distance of 18/20 or 19/21 from its recognition site. In the MPS protocol, MmeI digestion results in 17-bp or 18-bp tags of unique genomic sequence. A sharp band at approximately 70 bp was purified with 6% PAGE, and DNA ends were blunted and ligated to two adaptors of different lengths. The molecules with two different adaptors at their ends were again purified using 6% PAGE and PCR amplified. Then sharp bands at approximately 135 bp were purified on a 6% PAGE gel and emulsion PCR with 1-μm magnetic beads containing biotin labeled primer was performed. Millions of beads with amplified DNA fragments (paired tags) were immobilized in acrylamid-based gel system and a four-color sequencing by ligation was made [41]. This method allowed to obtain 13-bp sequencing information per tag separated by a 4- to 5-bp gap (sequencing per each tag was done from both ends). Thus, 26-bp information per amplicon was obtained (2 tags × 13 bp). In a pilot experiment, this technology was applied to resequence an evolved strain of *E. coli*, and according to the estimation, this was done on less than a one-error-per-million consensus bases. In this pilot experiment, two deletion genomic fragments and several 1-bp substitutions were detected in the analyzed *E. coli* strain.

Still, neither "454" nor the MPS method is up to speed yet. The accuracy of both should be improved by at least one order of magnitude. Also, to sequence mammalian genomes the length of sequence generated should be about 700 bases, but reads reported from these new approaches are between 26 and 110 bases.

### 6.3 Slalom Library: A Novel Approach to Genome Mapping and Sequencing

All described above sequencing approaches were based on sequencing of randomly generated DNA fragments.

We recently suggested a completely different and efficient strategy for simultaneous genome mapping and sequencing. The approach was based on physically oriented, overlapping restriction fragment libraries called slalom libraries. Slalom libraries combined features of general genomic, jumping, and linking libraries. Slalom libraries could be adapted to different applications, and two main types of slalom libraries will be discussed below. This approach was used to map and sequence (with ~46% coverage) two human PAC clones, each of ~100 kb. This model experiment demonstrates the feasibility of the approach and shows that the efficiency (cost-effectiveness and speed) of existing mapping/sequencing methods

could be improved at least 5–10-fold. Furthermore, as the efficiency of contig assembly in the slalom approach is virtually independent of length of sequence reads, even short sequences of 19–20 bp produced by rapid, high-throughput sequencing techniques would suffice to complete a physical map and a sequence scan of a small genome.

Two slalom libraries are used in the first type of approach. This approach allows us to construct contigs of plasmid clones covering a whole genome. However, these contigs will contain some gaps. Comparatively modest sequence information can be generated (20–25% of a genome). Using this technique, two (or more) bacterial strains (e.g., pathogenic and nonpathogenic) could be quickly compared and pathogenic islands could be identified.

The main principle of the type I slalom libraries is shown in Figure 8 and looks similar to the computer-assisted long-range mapping with NotI linking and jumping clones (Figure 5). In this case, the role of NotI jumping library plays a standard EcoRI genomic (slalom "R") library that is produced by complete digestion of genomic DNA with EcoRI, and the role of a NotI linking library–EcoRI linking (slalom "BR") library. Slalom "BR" library is constructed in exactly the same way as a NotI linking library using BamHI as a second enzyme and fulfills the connecting function joining EcoRI fragments. Shortly, DNA is digested with BamHI, circularized and cut with EcoRI ("BR" library). Thus, EcoRI in this case plays the role of a NotI enzyme.

Sequences from "BR" and "R" libraries are produced using standard reverse and forward sequencing primers and overlapping clones are found using a computer program. The homologies found will join the ends of EcoRI fragments in the "BR" library with the ends of EcoRI fragments in the "R" library to yield an ordered set of BamHI–EcoRI clones/STSs distributed along the genome.

Figure 8 shows the scheme where EcoRI and BamHI sites alternate. In reality, EcoRI sites and BamHI sites do not always alternate, and this will lead

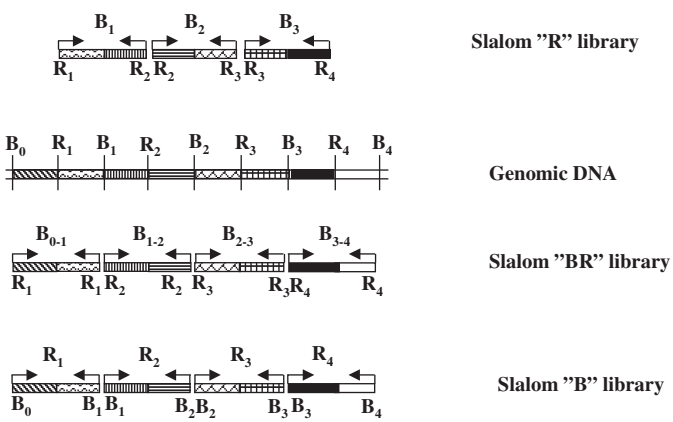


**Figure 8.** The main idea of the slalom approach. Numbers designate identical end sequences in different libraries that can be joined by a computer program in a contig of overlapping clones. B represents BamHI, and R denotes EcoRI sites. Small horizontal arrows show short sequences flanking EcoRI sites. The "BR" library plays a role of a linking library where EcoRI replaces NotI. The ordinary "R" library plays the role of a jumping library where EcoRI replaces NotI.

to the gaps in this set of overlapping clones, if several EcoRI or BamHI sites lie together. Thus, the genome will not be completely covered by clones because the information between some EcoRI and BamHI sites will be missing.

The problem with the gaps can be solved using the second variant of the slalom libraries (type II) where three libraries are used:

1. *Standard, completely EcoRI digested (slalom "R") library*
2. *Standard, completely BamHI digested (slalom "B") library*
3. *EcoRI jumping (slalom "RBR" or connecting) library*

This approach results in generation of plasmid clone contigs covering a whole genome without gaps. A large volume of sequencing information can be obtained (45–55% of a genome). The connecting library in this case is similar to a NotI jumping library. If the ultimate aim is to sequence the complete genome, then sequencing gaps could be filled-in using standard methods such as primer walking or transposon-mediated sequencing.

An "RBR" slalom library or EcoRI jumping library is prepared in the same way as a NotI jumping library using as a second enzyme BamHI. The only difference is that circularization after BamHI digestion is performed in the presence of a selective marker (e.g., $Kan^R$ gene or oligonucleotide adaptor, etc.). The circular constructs are opened with EcoRI and cloned [43].

The "RBR" library can be constructed in a simpler way. Plasmid DNA isolated *en masse* from a slalom "R" library is digested with BamHI and circularized in the presence of $Kan^R$. Then *E. coli* cells are transformed with ligated DNA and plated on agar with kanamycin. The clones obtained in this manner will be identical in structure to the clones from an EcoRI jumping library prepared using the classic method.

By comparing the end sequences of the "B" library clones with the internal BamHI (from the marker fragment) sequences of the slalom "RBR" library clones, the BamHI clones can be positioned in relation to each other. After the comparison of end sequences in "R" and "RBR" libraries, EcoRI clones will be positioned relative to each other. Finally, EcoRI and BamHI clones will be assembled into a contig representing their organization in complete genome.

The EcoRI jumping library provides the connecting function in type II slalom libraries.

The major difference between the slalom library mapping/sequencing approach and other sequencing strategies is that the clones are generated according to a specific scheme and using complete digestion. As a result, the number of variants required to cover the whole genome decreases significantly. The preparation of libraries for the slalom approach is remarkably simple: Only complete digestion with EcoRI or BamHI is used. There is no need for size separation, agarose gel purification, or establishing conditions for partial shearing/digestion. There is no need to develop a new sequenator as any existing instrument can be used. Importantly, it is not necessary to keep all slalom clones because sequencing information can be used to design PCR primers and even large fragments (up to 40–50 kb) can be amplified by long-range PCR.

The slalom library approach differs fundamentally from the shotgun sequencing approach with respect to the efficiency of assembly (EOA). The EOA for the latter method is strictly dependent on the length of sequencing reads. The longer the reads, the higher the EOA. As the slalom approach uses nonrandom fractionation of the DNA and each start site is tightly linked with the recognition site for the restriction enzyme, even very short sequences will, in principle, be enough to create a contig of the overlapping clones. The EOA of the slalom library approach is, therefore, essentially independent of read-length. Even the short sequences generated by pyrosequencing, MPS, or MPSS [37–40] should be sufficient to completely cover a genome. As one person can generate thousands of sequences a day using a pyrosequencer, the minimal set of overlapping clones covering a 4-Mb genome can be completed in a couple of days. Repeat sequences are in fact less of a problem for the slalom approach than for the shotgun sequencing approach for several reasons [43].

It is important to stress that the benefits of the slalom approach are most obvious in comparative sequencing experiments in combination with high-throughput techniques like pyrosequencing or MPSS (which are not compatible with the shotgun sequencing approach). Our particular experiments demonstrated that approximately 4% coverage would be enough to construct the contig of the overlapping clones and subsequently generate >20% ordered sequences with almost 100% efficiency, i.e., with 0.2-fold coverage.

Of course, closing of the gaps will be done with significantly lower efficiency. However, the finishing stages of the shotgun sequencing approach are also the most expensive and time-consuming part of the process. Lander et al. [9] distinguished three types of gaps: gaps within unfinished sequenced clones; and gaps between sequenced clone contigs, but within fingerprint clone contigs; and gaps between fingerprint clone contigs. The first type is the simplest, and the third is the most complicated to close because constructing a contig of overlapping clones is the most difficult procedure. With the slalom approach, we already have a contig of overlapping clones, and thus, we will only suffer from the first, simplest type of gaps. It is important to mention another difference in the finishing stages of these two approaches. With the shotgun sequencing approach, sequences from different clones must be connected, and here highly related repeats, gene families, and polymorphisms will represent a major problem. These problems are nonexistent in the slalom approach, where a single insert should be sequenced.

The slalom strategy can increase several orders of magnitude the efficiency of mapping and sequencing, i.e., decrease the cost and labor, and increase the speed of sequencing project fullfillment (see Table 1). This strategy will allow the establishment of a physical map with a minimal set of overlapping clones that will pinpoint differences in genome organization between organisms. At the same time, a considerable sequence coverage (about 50%) of the genome at a less than onefold coverage will be achieved. This will make it possible to locate virtually every gene in a genome for more detailed study. The method makes it economically and logistically possible for a wide range of laboratories to gather the detailed information about large numbers of genomes. It offers a way to use genomics techniques for the study of biocomplexity, evolution, and taxonomy in the close future.

**TABLE 1. Slalom Strategy and Shotgun Sequencing Approach**

| | Slalom | Shotgun |
|---|---|---|
| Sequence coverage 1.2-fold | A. Minimal set of overlapping clones; restriction map<br>B. 20–25% of sequences are ordered into STS islands covering the whole genome (+appr. 30% of unordered sequences) | A. Random clones, no map<br>B. Practically no ordered sequences (appr. 50–70% of unordered sequences) |
| Pyrosequencing 4% coverage | Minimal set of overlapping clones; restriction map | No use |
| Pyrosequencing plus sequencing unique clones (25–30% coverage) | 20–25% of sequences are ordered into STS islands covering the whole genome | No use |
| Sequence coverage needed to achieve 95% of complete sequence | 1 | 10–15 |
| Sequence coverage needed to compare two genomes and isolate clones/genes completely covering regions of divergence | 4% (0.04-fold) | At least 6–8-fold (600%–800%) |

## 7 IDENTIFICATION OF GENES. HOW MANY GENES ARE IN THE HUMAN GENOME?

### 7.1 Many Approaches Exist for Gene Identification

The most important for the majority of research tasks is identification of genes in a particular region of the human genome. The different tissues in the body express information across a wide range, between less than 2000 (in specialized and terminally differentiated white blood cells) and more than 20,000 (in the placenta). Several methods were suggested previously for gene identification [3, 4]. Genomic DNA can be used as a probe to screen cDNA libraries. Different approaches were developed to capture cDNA clones using genomic DNA attached to a solid substrate (e.g., magnetic beads, nylon filters) as a target to that cDNA clones can hybridize. To identify genomic sequences that could be expressed as a special approach, exon trapping was suggested. For example, in the method of Church et al. [44], the DNA is subcloned into a plasmid expression vector pSPL3 that contains an artificial minigene that can be expressed in a suitable host cell. The minigene consist of the following:

1) Segment of the simian virus 40 (SV40) genome that contains an origin of replication plus a powerful promoter sequence.

2) Two splicing-competent exons separated by an intron that contains a multiple cloning site (MCS).
3) SV40 polyadenylation site.

The recombinant DNA is transfected into a strain of monkey cells, known as COS cells. COS cells were derived from monkey CV-1 cells by artificial manipulation, leading to integration of a segment of the SV40 genome containing a defective origin of replication (*ori*). The integrated SV40 segment in COS cells allows any circular DNA that contains a functional *ori*SV40 to replicate independently of the cellular DNA. Transcription from the SV40 promoter results in an RNA transcript that normally splices to include the two exons of the minigene. If the DNA cloned into the MCS contains a functional exon, then the additional exons can be spliced to the minigene exons. After isolation of RNA and preparing cDNA, PCR reactions with primers specific for the minigene exons is made. Agarose gel electrophoresis can easily distinguish between normal splicing and splicing involving exons in the insert DNA.

Another method to identify expressed sequences is called zoo-blotting. In this method, genomic DNA fragments are hybridized at reduced hybridization stringency against a Southern blot of genomic DNA samples isolated from different species. This method is based on a fact that expressed sequences are usually well conserved, and for example, human DNA fragment with exons from *RB1* gene will hybridize to mouse and pig genomic DNA fragments containing *RB1* gene.

However, all of these methods are laborious, time consuming, and inefficient.

Currently, three approaches are the most popular for the identification of expressed sequences/genes in the genomic DNA: computer analysis of DNA sequences for possible exons, identification of CpG islands that can be done both experimentally and using bioinformatics, and large-scale sequence analysis of transcribed sequences, like sequencing of ESTs, SAGE, and CAGE.

Bioinformatic methods to identify genes will not be discussed here as special journal issues are focused on this topic, and many new programs appeared every year. In principle, these programs could be divided into two classes. One type of program predicts genes using only genomic sequences. These programs can also compare genomic DNA sequences from different organisms and search for conserved sequences. The second type of program compares different cDNA, EST, and genomic sequences to find the most possible gene sequence. This method is more reliable; however, it is important to remember that all computer programs frequently predict nonexisting genes and exons and miss really existing genes and exons. Thus, all computer-predicted genes/exons need experimental confirmation. In cases when a gene has alternative 5′ and 3′ ends and different splicing forms computer programs are not very much helpful.

### 7.2 CpG Islands

The dinucleotide CpG is notable because it is greatly under-represented in human DNA, occurring at only about one fifth of the roughly 4% frequency that would be expected by simply multiplying the typical fraction of Cs and Gs ($0.21 \times 0.21$). The deficit occurs because most CpG dinucleotides are methylated on the cytosine base, and spontaneous deamination of methyl-C residues gives rise to T residues.

Spontaneous deamination of ordinary cytosine residues gives rise to uracil residues that are readily recognized and repaired by the cell. As a result, methyl-CpG dinucleotides steadily mutate to TpG dinucleotides. However, the genome contains many CpG islands that represent stretches of unmethylated DNA with a higher frequency of CpG dinucleotides when compared with the entire genome [9, 10, 13, 32, 33, 45]. CpG islands are believed to preferentially occur at the transcriptional start of genes, and it has been observed that most housekeeping genes have CpG islands at the 5′ end of the transcript. In addition, experimental evidence indicates that CpG island methylation is correlated with gene inactivation and has been shown to be important during gene imprinting and tissue-specific gene expression. Thus, identification of CpG methylation is crucial for isolation of tumor suppressor genes and early diagnosis of cancer [46, 47].

CpG islands are usually 1–2 kb long and are dispersed in the genome. They are called CpG (rich) islands because the %(G+C) of CpG island sequences usually exceed 60% and the human genome contains on average about 40% of C+G content. The combination of G+C richness and lack of CpG suppression means that CpG islands contain 10–20 times more CpGs than an equivalent length of non-island DNA. Altogether, there are about 30,000 to 45,000 islands in the haploid genome (the average spacing is about 1 per 100 kb). It is now clear that the majority (if not all) of CpG islands are associated with genes.

As one of the main goals of the human genome project is the isolation of all genes and the construction of a transcriptional gene map, it is clear that markers located in the CpG islands have an additional value for physical mapping. It has been shown that recognition sites for many of the rare cutting enzymes are closely associated with CpG islands. For example, at least 82% of all NotI and 76% of all XmaIII sites are located in the CpG islands. More than 20% of CpG island-containing genes have at least one NotI site in their sequence, whereas about 65% of those genes have XmaIII site(s). Summarizing the data for all genes (with or without CpG islands), we can conclude that approximately 12% of all well-characterized human genes contain NotI sites, and 43% of them have XmaIII sites. For human genome mapping this means that by sequencing DNA fragments containing a NotI site, it is possible to tag up to one fifth of all expressed genes. As discussed, the recombinant clone containing a NotI (or other rare cutting enzymes) recognition site is called a linking clone.

How many genes contain the human genome? This question is difficult, and the answer is very much dependent on the definition of a gene. According to Celera and HGP, the human genome contains approximately 30,000–35,000 genes.

However, sequencing of NotI linking clones suggested that the human genome contains 15,000–20,000 *Not*I sites, of which 6000–9000 are unmethylated in any particular cell. It means that the human genome contains 45,000–60,000 genes [33].

Comparison of our database containing experimentally obtained NotI linking sequences with complete chromosome 21 and 22 sequences revealed several interesting features. First, it was shown that chromosome 22 contains greater than twofold more genes than chromosome 21 (545:225 = 2.4), and we see the same ratio within the NotI flanking sequences (119:49 = 2.4). We have demonstrated that nearly all our NotI clones contained genes and suggested that 12.5–20% of all genes contain NotI sites [32]. This correlates well with the number of genes on

chromosomes 21 and 22 (168/770 = 22%). Second, according to sequencing, the two chromosomes contain 390 NotI sites. Therefore, if we assume that each NotI site is associated with a gene, then almost half of the genes contain NotI sites. This estimate seems excessive.

The human genome sequence data cannot discriminate between methylated and unmethylated cytosines. There are several algorithms for the identification of CpG islands on the basis of primary sequence. One quantitative definition holds that CpG islands are regions of DNA >200 bp with C+G content of >50% and a ratio of the "observed vs. expected" frequency of CG dinucleotides, which exceeds 0.6 [48, 49]. The ratio for the entire genome is approximately 0.2 [9]. According to the previous data, 82% of NotI sites are located in CpG islands (see above). It is important to note that these data were obtained either using computational methods or limited experimental datasets. Using the NotI cloning method, only unmethylated NotI sites can be isolated. Comparing the experimental NotI cloning method and computer-based identification of CpG islands [9, 33], two main features are apparent: The fraction of sequences with >80% CG content is nine times higher in the NotI collection, i.e., 142 vs. 22 sequences. Another striking finding is that even NotI flanking sequences with CG content less than 50% have a very high ratio of observed versus expected frequency of CG dinucleotide: 0.71. This suggests that essentially all NotI flanking sequences generated in the study are located in CpG islands, and therefore, the computational method misses at least 8.7% of CpG islands associated with NotI sites. Thus, existing tools for computational determination of CpG islands fail to identify a significant fraction of functional CpG islands, and unmethylated DNA stretches with a high frequency of CpG dinucleotides can be found even in regions with low CG content.

### 7.3 SAGE and CAGE Approaches to Analyze Human Transcriptome and to Estimate the Number of Human Genes

Another approach to identify genes and analyze their expression pattern in human/mammalian cell is large-scale sequencing of transcribed sequences. Sequencing of ESTs and isolation of full-length cDNA clones was known a long time ago and will not be discussed here. Recently two novel high-throughput methods for the analysis of human transcriptome were developed: serial analysis of gene expression (SAGE, [50]) and cap analysis gene expression (CAGE, [51]). Both methods are based on the generation of short sequence tags. Each of these tags is assumed to identify a gene transcript.

SAGE is based on two principles. First, a specific adaptor is ligated to cDNA close to the 3′ end of the mRNA (polyA tail). This adaptor contains a recognition site for a type IIs restriction aendonuclease. Originally, it was FokI, and in a recent modification (LongSAGE), MmeI was used [50, 52]. FokI generates 13–14-bp, and MmeI generates 20–21-bp tags. The 21-bp tag consists of a constant 4-bp sequence representing the restriction site at which the transcript was cleaved, followed by a unique 17-bp sequence derived from an adjacent sequence in each transcript. Theoretical calculations show that >99.8% of 21-bp tags are expected to occur only once in genomes the size of the human genome. Similar analyses based on actual sequence information from ~16,000 genes suggest that >75% of 21-bp tags would be expected to occur only once in the human genome, with the remaining tags

matching duplicated genes or repeated sequences. Second, to optimize the quantification of transcripts, tags are ligated together to form "ditags," which are then concatenated and cloned. Sequencing tag concatemers in parallel allows the identification of up to ~30 tag sequences in each sequencing reaction. Matching tags to genome sequences identifies the gene corresponding to each tag, and the number of times a particular tag is observed provides a quantitative measure of transcript abundance in the RNA population.

CAGE is similar to SAGE in principle; however, it is based on preparation and sequencing of concatamers of DNA tags deriving from the initial 20 nucleotides from 5′ end mRNAs.

The method essentially uses cap-trapper full-length cDNAs to the 5′ ends of which linkers are attached [53, 54]. This is followed by the cleavage of the first 20 base pairs by class IIs restriction enzymes (Mme) and then another linker is ligated. After it, PCR amplification, excision of tags, their concatamerization, and cloning of the CAGE tags was performed. CAGE tags derived by sequencing these clones are mapped to the genome and used for expression analysis, as well as for the determination of the 5′ end borders of new transcriptional units. Thus, in contrast to SAGE, CAGE allows high-throughout gene expression analysis and the profiling of transcriptional start points, including promoter usage analysis.

SAGE and CAGE concatamer sequencing is more cost-effective than a full-length cDNA library sequencing because of the much higher throughput of identified tags.

In a recent study (FANTOM 3 project), full-length cDNA isolation and 5′- and 3′-end sequencing of cloned cDNAs was combined with CAGE, gene identification signature (GIS), and gene signature cloning (GSC) ditag technologies for the identification of RNA and mRNA sequences corresponding to transcription initiation and termination sites [55]. In this study, paired initiation and termination sites were identified and the boundaries for 181,047 independent transcripts in the mouse genome were established. In total, 1.32 5′ start sites for each 3′ end and 1.83 3′ ends for each 5′ end are found. Based on these data, the number of transcripts is at least one order of magnitude larger than the estimated 22,000 "genes" in the mouse genome, and the large majority of transcriptional units have alternative promoters and polyadenylation sites. To extend the mouse data, two human CAGE libraries, one constructed with random primers and the other with oligo-dT primers, were combined to produce 1,000,000 CAGE tags. Mapping of these tags to the human genome identified the likely promoters and transcriptional starting site of many genes and clearly indicates that the same level of transcriptional diversity occurs in humans as in mice and is at least 10 times as great as the number of "genes."

Of the 102,281 FANTOM3 cDNAs, 34,030 lack any protein-coding sequence (CDS) and are annotated as non-protein-coding RNA (ncRNA). The function of ncRNAs is a matter of debate [56], but it is clear that some of them are very important for transcription regulation. Some ncRNAs are highly conserved even in distant species. According to the previous data, only 1–1.5% of the human genome is spanned by exons and 75% of the genome is non-transcribed intergenic DNA. However, FANTOM 3 data demonstrated that the majority (more than 60%) of the mammalian genome is transcribed and commonly from both strands.

Analysis of the output of FANTOM 3 suggested that many more transcripts were still to be discovered and only the future can reveal the actual complexity in

the mammalian transcriptome. Very similar conclusions were made in another detailed study of human transcriptome [57, 58].

## 8 WAS IT WORTH IT TO SEQUENCE HUMAN GENOME?

The Human Genome Project was controversial from the beginning. Some criticism concerns scientific strategy and the ways in which it was executed [1–4]. Is it important to sequence the whole human genome or only coding DNA sequences (genes)? Now it is clear that DNA sequences recognized previously as a "junk" in fact are very important for understanding how human genome functions. This "junk" DNA contains important regulatory sequences, ncRNAs, and probably the most intriguing discoveries will come from areas considered as wastelands. Many ethical questions arose during the HGP. For instance, sample collection from individuals was often done without proper consultation or explanation of the possible future use. As the HGP moved forward, commercial concerns were increasingly taking a stake in the research. Large financial investments from big companies can monopolize certain research areas, for example, sequencing of cDNA and gene identification. By subsequently seeking patents to protect their financial investments, they raised a question: Who owns the human genome?

Any major scientific advance carries with it the fear of exploitation, and HGP is not an exception. For example, knowledge of disease-associated mutations can be used for disease prevention, and at the same time, it can be used for discrimination. The comprehensive knowledge of human genome and genes can lead to biological determinism and revival of eugenics with all its negative consequences.

Another objection was that funding of the HGP would be done at the expense of other scientific directions.

The human genome map and sequence still contain many uncertainties, and only careful examination by some research group focusing on a particular genomic region can produce a reliable map and sequence. For instance, we have cloned all *Not*I sites and constructed a physical map for two chromosome 3 regions containing tumor suppressor genes [26, 59, 61]. In the course of these studies, it became apparent that large-insert vectors from these regions were unstable, sensitive to deletions and rearrangements, and the original maps produced by HGP were erroneous [26, 59, 60].

It should be emphasized that the enormous efforts deployed on sequencing the human genome are extremely important; however, a critical role remains for verified, integrated maps. In sequencing, the short and long repeats spread throughout the genome are sources of numerous errors. These errors are difficult to identify with the shotgun strategy, but they become evident when mapping information is combined with the sequence. Furthermore, difficulties in sequence assembly caused by the existence of large families of recently duplicated genes and pseudogenes are easier to resolve using integrated maps.

In many cases, sequence and mapping information is duplicated, overlapping, or contradictory. One must always keep in mind that even absolutely correct and long nucleotide sequences may be localized incorrectly along the chromosomal DNA if the appropriate accompanying mapping information is ignored. For this reason, despite the vast amount of information currently available, there is an

urgent need to reconcile this information in a unified framework, to generate an integrated noncontroversial map for each chromosome.

The draft human genome sequence produced by HGP was estimated to cover at least 92% of the whole human genome. However, it contained only 55.7% of the *Not*I flanking sequences and Celera's database contained matches to 57.2% of the clones. The data suggest that the shotgun sequencing approach used to generate the draft human genome sequence resulted in a bias against cloning and sequencing of *Not*I flanks. Even in 2005 up to 5% of our NotI flanking sequences (that are also incomplete!) are not present in the human genome sequence available in public databases.

Several explanations can be offered to account for the low representation of *Not*I flanking sequences in the draft human genome sequences. One potential explanation is that the cloning of some *Not*I containing regions may be selected against experiments with large-insert cloning vectors. Our experience has also proven that even in small-insert plasmid vectors, some human sequences are more easy to clone than other. In our procedure, we directly selected clones containing *Not*I sites, and in a shotgun sequencing approach, such sequences could be under-represented. An alternative explanation, based on the observation that some *Not*I flanking sequences can have 100% identity over long DNA stretches [62], is that some *Not*I sites were incorrectly fused in the assembly process. Furthermore, our experience demonstrated that sometimes it is very difficult to read *Not*I flanking sequences because of extremely high CG content. During human genome assembling, such sequences would be eliminated as possessing low-quality data. Further experimental analysis is needed to conclusively identify the cause(s) of the bias.

HGP resulted in enormous progress in automation of sequencing and other molecular biology methods.

It gives us new tools for disease-gene discovery. In fact, the information obtained during the HGP helps rather than hinders small research groups. It is still important to remember that only the euchromatin portion of the genome was sequenced, and it is not clear to what extent it was done. Many heterochromatin regions are practically not sequenced, and they most probably contain many unknown genes and other functionally important sequences. Thus, the draft human genome sequence was made, but it is a very long way to obtain a really complete human genome sequence and to understand how many genes it contains and how it functions.

## ACKNOWLEDGEMENT

This work was supported by research grants from the Swedish Cancer Society, the Swedish Research Council, the Swedish Foundation for International Cooperation in Research and Higher Education (STINT), the Swedish Institute, the Royal Swedish Academy of Sciences, INTAS, and the Karolinska Institute.

## REFERENCES

1. The Human Genome. Special Issue. (2001). *Nature*, *409*, No.6822.
2. The Human Genome. Special Issue. (2001). *Science*, *291*, No.5507.

3. Brown T A (1999). *Genomes*, Wiley, New York co-published with BIOS Scientific Publishers, Oxford, UK.
4. Strachan T, Read A (1999). *Human Molecular Genetics*, Wiley, New York co-published with BIOS Scientific Publishers, Oxford, UK.
5. Fleischmann R D, Adams M D, White O, et al. (1995). Whole-genome random sequencing and assembly of Haemophilus influenzae Rd. *Science*. 269:496–512.
6. Goffeau A, Barrell B G, Bussey H, et al. (1996). Life with 6000 genes. *Science*. 274:546–567.
7. The *C. elegans* Sequencing Consortium. (1998). Genome sequence of the nematode C. elegans: A platform for investigating biology. *Science*. 282:2012–2018.
8. Dunham I, Hunt A R, Collins J E, et al. (1999). The DNA sequence of human chromosome 22. *Nature*. 402:489–495.
9. Lander E S, Rogers J, Waterston R H, et al. (2001). Initial sequencing and analysis of the human genome. *Nature*. 409:860–921.
10. Venter G J, Adams M D, Myers E W, et al. (2001). The sequence of the human genome. *Science*. 291:1304–1351.
11. Ausubel F M, Kingston R E, Brent R, et al. (1987–2004). *Current Topics in Molecular Biology*, Wiley, New York.
12. Sambrook J, Fritsch E F, Maniatis T (1989). *Molecular Cloning: A Laboratory Manual* (2nd ed). Cold Spring Harbour Laboratory Press, Cold Spring Harbour; N.Y.
13. Zabarovsky E R (2004) Genomic DNA Libraries, Construction and Applications. In R A Meyers (ed.), *Encyclopedia of Molecular Cell Biology and Molecular Medicine*, Vol. 5 Wiley-VCH, Weinheim, Germany, pp. 441–468.
14. Zabarovsky E R, Winberg G, Klein G (1993). The SK-diphasmids—vectors for genomic, jumping and cDNA libraries. *Gene*. 127:1–14.
15. Zabarovsky E R, Turina O V, Nurbekov M K, et al. (1989). SK18—diphasmid for constructing genomic libraries with blue/white selection, multiple cloning sites and with the M13 and lambda ori. *Nucl. Acids Res*. 17:3309.
16. Short J M, Fernandez J M, Sorge J A, et al. (1988). Lambda ZAP: A bacteriophage lambda expression vector with in vivo excision properties. *Nucleic Acids Res*. 16:7583–7600.
17. Carninci P, Shibata Y, Hayatsu N, et al. (2001). Balanced-size and long-size cloning of full-length, cap-trapped cDNAs into vectors of the novel lambda-FLC family allows enhanced gene discovery rate and functional analysis. *Genomics*. 77:79–90.
18. Frischauf A M, Lehrach H, Poustka A, et al. (1983). Lambda replacement vectors carrying polylinker sequences. *J Mol Biol*. 170:827–842.
19. Frischauf A M, Murray N, Lehrach H (1987). Lambda phage vectors–EMBL series. *Methods Enzymol*. 153:103–115.
20. Zabarovsky E R, Allikmets R L (1986). An improved technique for the efficient construction of gene library by partial filling-in of cohesive ends. *Gene*. 42:119–123.
21. Latif F, Tory K, Gnarra J, et al. (1993). Identification of the von Hippel-Lindau disease tumor suppressor gene. *Science*. 260:1317–1320.
22. Wang I, Franco B, Ferrero G B, et al. (1995). High-density physical mapping of a 3-Mb region in Xp22.3 and refined localization of the gene for X-linked recessive chondrodysplasia punctata (CDPX1). *Genomics*. 26:229–238.
23. Cheng S, Lutfalla G, Uze G, et al. (1993). GART, SON, IFNAR, and CRF2–4 genes cluster on human chromosome 21 and mouse chromosome 16. *Mamm Genome*. 4:338–342.

24. Cohen D, Chumakov I, Weissenbach J (1993). A first-generation physical map of the human genome. *Nature*. 366:698–701.
25. Frengen E, Zhao B, Howe S, et al. (2000). Modular bacterial artificial chromosome vectors for transfer of large inserts into mammalian cells. *Genomics*. 68:118–126.
26. Protopopov A, Kashuba V, Zabarovska V I, et al. (2003). An integrated physical and gene map of the 3.5-Mb chromosome 3p21.3 (AP20) region implicated in major human epithelial malignancies. *Cancer Res*. 63:404–412.
27. Ichikawa H, Hosoda F, Arai Y, et al. (1993). A NotI restriction map of the entire long arm of human chromosome 21. *Nat Genet*. 4:361–366.
28. Poustka A, Lehrach H (1988). Chromosome Jumping: A Long Range Cloning Technique. In Setlow J K (ed.), *Genetic Engineering Principles and Methods*, Vol. 10, Brookhaven National Laboratory, Upton, N.Y. and Plenum Press, New York, pp. 169–193.
29. Zabarovsky E R, Kashuba V I, Gizatullin R Z, et al. (1996). NotI jumping and linking clones as a tool for genome mapping and analysis of chromosome rearrangements in different tumors. *Cancer Detect. Prevention*. 20:1–10.
30. Zabarovsky E R, Boldog F, Thompson T, et al. (1990). Construction of a human chromosome 3 specific *Not*I linking library using a novel cloning procedure. *Nucl. Acids Res*. 18:6319–6324.
31. Zabarovsky E R, Boldog F, Erlandsson R, et al. (1991). A new strategy for mapping the human genome based on a novel procedure for constructing jumping libraries. *Genomics*. 11:1030–1039.
32. Allikmets R L, Kashuba V I, Bannikov V M, et al. (1994). NotI linking clones as tools to join physical and genetic mapping of the human genome. *Genomics*. 19:303–309.
33. Kutsenko A, Gizatullin R, Al-Amin A N, et al. (2002). NotI flanking sequences: A tool for gene discovery and verification of the human genome. *Nucleic Acids Res*. 30:3163–3170.
34. Zabarovsky E R, Kashuba V I, Zakharyev V M, et al. (1994). Shot-gun sequencing strategy for long range genome mapping: First results. *Genomics*. 21:495–500.
35. Myers E W, Sutton G G, Delcher A L, et al. (2000). A whole-genome assembly of Drosophila. *Science*. 287:2196–2204.
36. Drmanac R, Drmanac S, Chui G, et al. (2002). Sequencing by hybridization (SBH): Advantages, achievements, and opportunities. *Adv. Biochem. Eng. Biotechnol*. 77:75–101.
37. Brenner S, Johnson M, Bridgham J, et al. (2000). Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays. *Nature Biotechnol*. 18:630–634.
38. Shendure J, Porreca G J, Reppas N B, et al. (2005). Accurate multiplex polony sequencing of an evolved bacterial genome. *Science*. 309:1728–1732.
39. Margulies M, Egholm M, Altman W E, et al. (2005). Genome sequencing in microfabricated high-density picolitre reactors. *Nature*. 437:376–380.
40. Ronaghi M, Pettersson B, Uhlen M, et al. (1998). A sequencing method based on real-time pyrophosphate. *Science*. 281:363–365.
41. Jarvius J, Nilsson M, Landegren U (2003). Oligonucleotide ligation assay. *Methods Mol. Biol*. 212:215–228.
42. Rogers Y H, Venter JC (2005). Genomics: Massively parallel sequencing. *Nature*. 437:326–327.
43. Zabarovska V I, Gizatullin R G, Al-Amin A N, et al. (2002). Slalom libraries: A new approach to genome mapping and sequencing. *Nucleic Acids Res*. 30:1–8.

44. Church D M, Stotler C J, Rutter J L, et al. (1994). Isolation of genes from complex sources of mammalian genomic DNA using exon amplification. *Nat. Genet.* 6:98–105.
45. Bird A (2002). DNA methylation patterns and epigenetic memory. *Genes Dev.* 16:6–21.
46. Palmisano W A, Divine K K, Saccomanno G, et al. (2000). Predicting lung cancer by detecting aberrant promoter methylation in sputum. *Cancer Res.* 60:5954–5958.
47. Sugimura T, Ushijima T (2000). Genetic and epigenetic alterations in carcinogenesis. *Mutat. Res.* 462:235–246.
48. Gardiner-Garden M, Frommer M (1987). CpG islands in vertebrate genomes. *J. Mol. Biol.* 196:261–282.
49. Larsen F, Gundersen G, Lopez R, et al. (1992). CpG islands as gene markers in the human genome. *Genomics.* 13:1095–1107.
50. Velculescu V E, Zhang L, Vogelstein B, et al. (1995). Serial analysis of gene expression. *Science.* 270:484–487.
51. Shiraki T, Kondo S, Katayama S, et al. (2003). Cap analysis gene expression for high-throughput analysis of transcriptional starting point and identification of promoter usage. *Proc. Natl. Acad. Sci. U.S.A.* 100:15776–15781.
52. Saha S, Sparks A B, Rago C, et al. (2002). Using the transcriptome to annotate the genome. *Nat. Biotechnol.* 20:508–512.
53. Carninci P, Hayashizaki Y (1999). High-efficiency full-length cDNA cloning. *Methods Enzymol.* 303:19–44.
54. Shibata Y, Carninci P, Watahiki A, et al. (2001). Cloning full-length, cap-trapper-selected cDNAs by using the single-strand linker ligation method. *BioTechniques.* 30:1250–1254.
55. Carninci P, Kasukawa T, Katayama S, et al. (2005). The transcriptional landscape of the mammalian genome. *Science.* 309:1559–1563.
56. Szymanski M, Barciszewska M Z, Erdmann V A, et al. (2005). A new frontier for molecular medicine: Noncoding RNAs. *Biochim. Biophys. Acta.* 1756:65–75.
57. Kapranov P, Drenkow J, Cheng J, et al. (2005). Examples of the complex architecture of the human transcriptome revealed by RACE and high-density tiling arrays. *Genome Res.* 15:987–997.
58. Cheng J, Kapranov P, Drenkow J, et al. (2005). Transcriptional maps of 10 human chromosomes at 5-nucleotide resolution. *Science.* 308:1149–1154.
59. Wei M-H, Latif F, Bader F, et al. (1996). Construction of a 600 Kilobase cosmid clone contig and generation of a transcriptional map surrounding the lung cancer tumor suppressor gene (TSG) locus on human chromosome 3p21.3.: Progress toward the isolation of a lung cancer TSG. *Cancer Research.* 56:1487–1492.
60. Lerman M I, Minna J D. et al. (2000) The International Lung Cancer Chromosome 3p21.3 Tumor Suppressor Gene Consortium. The 630-kb lung cancer homozygous deletion region on human chromosome 3p21.3: Identification and evaluation of the resident candidate tumor suppressor genes. *Cancer Research.* 60:6116–6133.
61. Imreh S, Klein G, Zabarovsky E (2003). Search for unknown tumor antagonizing genes. *Genes Chromos. & Cancer.* 38:307–321.
62. Kashuba V I, Protopopov A I, Kvasha S M, et al. (2002). hUNC93B1: A novel human gene representing a new gene family and encoding an unc-93-like protein. *Gene.* 283:209–217.

## FURTHER READING

Burke D T, Carle G F, Olson M V (1987). Cloning of large segments of exogenous DNA into yeast by means of artificial chromosome vectors. *Science*. 236:806–812.

Burke D T, Olson M V (1991). Preparation of clone libraries in yeast artificial-chromosome vectors. *Methods Enzymol*. 194:251–270.

Shizuya H, Birren B, Kim U J, et al. (1992). Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in Escherichia coli using an F-factor-based vector. *Proc Natl Acad Sci USA*. 89:8794–8797.

Ioannou P A, Amemiya C T, Garnes J, et al. (1994). A new bacteriophage P1-derived vector for the propagation of large human DNA fragments. *Nat Genet*. 6:84–89.

# OVERVIEW: DIFFERENTIATING ISSUES IN THE DEVELOPMENT OF MACROMOLECULES COMPARED WITH SMALL MOLECULES

Günter Blaich,[1] Bernd Janssen,[2] Gregory Roth,[2] and Jochen Salfeld[2]

[1] *Abbott Bioresearch Center, Worcester, Massachusetts*
[2] *Abbott GmbH & Co. KG, Ludwigshafen, Germany*

**Chapter Contents**

1 Introduction
2 Target Selection and Drug Discovery Approaches
3 Target Product Profile
4 Lead Identification
  4.1 Small Molecules
  4.2 Macromolecules
5 Lead Optimization
  5.1 Small Molecules
  5.2 Macromolecules
6 Selection of Clinical Candidates and Preclinical Development
7 Production of Active Pharmaceutical Ingredients
  7.1 Small Molecules
  7.2 Macromolecules
8 Formulation and Delivery Devices
  8.1 Formulation Support in Discovery
  8.2 Formulation Activities Until Clinical Proof of Concept
  8.3 Formulation of Activities in the Commercial Phase (Clincal Phase IIB until Market)
9 Clinical Development
10 Intellectual Property
11 Generics, "Biosimilar Products," or "Follow-on Biologics"

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

12 Summary
12.1 Step 1: Target Selection and Validation
12.2 Step 2: Choice Between Small-Molecule and Macromolecule Platforms
12.3 Step 3: Definition of Target Product Profile
12.4 Step 4: Drug Discovery and Development
Acknowledgment
References

## 1 INTRODUCTION

The probability of success in drug discovery and development is highly dependent on good planning right from the beginning, a rational approach to each phase in the light of the ultimate goal, and—unsurprisingly—good science. There are many possible reasons for the recent dearth of new molecular entity or new biological entity submissions to regulatory agencies [1, 2]. One possible reason is suboptimal discovery and development strategies for novel targets with little known biology [3]. In the late 1980s and early 1990s, drug discovery was specialized to an extent, with small-molecule drug discovery the domain of pharmaceutical companies and macromolecule discovery practiced by biotechnology companies. Today, drug discovery in both areas is practiced widely in both arenas. Large pharmaceutical companies have embraced biological therapies, in particular therapeutic antibodies, and large biotechnology companies have added small-molecule discovery to complement their macromolecular portfolio with orally available medicines. Pharmaceutical companies realize this portfolio diversification either by expanding or readjusting internal research or by collaborations with biotechnology companies. In either case, preclinical development is often conducted by the larger of the partners, and a comprehensive understanding about the different approaches to either drug class is critical.

Small molecules and macromolecules differ in much more than size, and this overview will discuss the consequences for each stage of drug discovery and development.

With a few exceptions, small molecules are synthesized chemically and are smaller than 600 daltons. For the purpose of this overview, macromolecules are defined as protein molecules of approximately 150 Kd, designed to bind to and functionally affect drug targets. The majority of these macromolecules are antibodies and antibody derivatives, which constitute approximately 25% of therapeutic entities currently in development [4]. Another highly represented class includes receptor fusion proteins. This overview will not describe other large molecules, including therapeutic proteins (naturally existing human proteins used for therapeutic purposes), antisense or siRNA, gene therapy, and others. However, many of the general concepts discussed here can be applied to such drug candidates as well.

The molecular weight and domain structure of antibodies or antibody derivatives are the basis for important limitations and potential advantages when

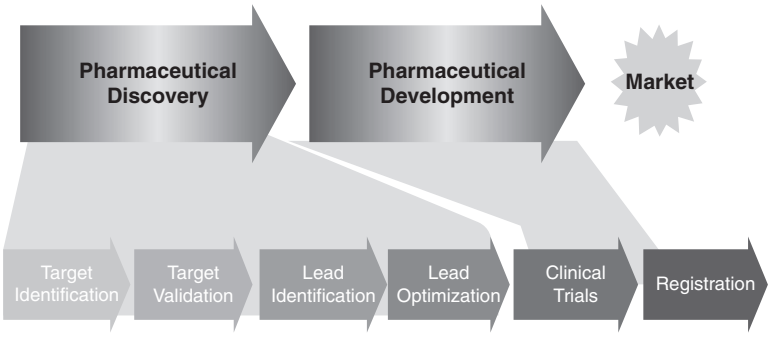


**Figure 1.** Critical path in the drug discovery process.

compared with the traditional drug discovery process. Although antibodies are at somewhat of a disadvantage because they are limited to soluble, cell-surface, and tissue matrix targets, their domain structure leads to very significant time-based advantages in the lead finding and the lead optimization phase (Figure 1).

Drug discovery today relies on the selection of specific and disease-relevant drug targets. Traditional (pre-1990) drug screening, based purely on observed pharmacological effects *in vivo*, has fallen out of favor due to regulatory pressures to conduct discovery research on well-characterized and validated targets. New screening technologies, coupled with the concept of using small molecules to complement gene-based methods of perturbing protein function, has generated renewed interest in this past paradigm and allows for the successful combination of pharmacological drug screening with the identification of specific and potentially new and novel targets in the "druggable genome" [5, 6]. The current general focus toward drug candidates with known mechanisms of action has important consequences for preclinical development. Whereas more standardized safety assessments were appropriate for drug candidates with unknown mechanisms, the understood biology of the drug action today allows and requires an appropriate adjustment at each stage. However, caution about any claim of a "known mechanism" is warranted. Indeed, although the industry worked previously on drug targets with, in some cases, more than 20 years of academic research into target biology, the Human Genome Project has led to work and focus on poorly understood targets. Therefore, drug discovery and development of the future has to undertake target validation work in parallel with drug discovery, consider the assumed mechanism in all formal pharmacology and safety assessments, and maintain the safety net of traditional drug safety studies.

## 2 TARGET SELECTION AND DRUG DISCOVERY APPROACHES

The term "target" describes a specific molecular entity, most often a protein, presumed to play a role in the pathogenesis of a disease state. A target can then be

defined precisely with molecular biology tools and can be affected by the action of a specific drug in a manner beneficial to the patient. Estimates place the number of targets modulated by currently marketed drugs at between 120 and 500, with only about 7% of medicines not having a clearly defined target [7]. Target selection starts typically with an analysis of met or unmet medical need and a comprehensive understanding of the pathophysiogical processes underlying the human disease of interest. Companies or academic researchers are typically focused on a limited number of key indications (such as cardiovascular disease, diabetes, autoimmune disease, and cancer) and have significant and relevant internal discovery expertise. This includes, to some extent, novel target discovery and will in most cases include target identification and validation [8]. Indeed, some level of confidence in mechanism-based target validation is a key component of any discovery project considering that, industry-wide, 30–40% of experimental drugs fail because an inappropriate or non-physiologically relevant biological target was pursued [9].

Once a preferred set of targets is identified, the viability and tractability of a small molecule or macromolecule approach can be assessed. In general, differences in physicochemical properties between small and macromolecules may help drive the approach strategy. For example, intracellular and nuclear proteins are often best targeted by small molecules that would more easily overcome cellular permeability barriers in order to reach the desired target. The molecular weight of approximately 150 Kd precludes antibodies from entering cells. Some initial progress has been made to overcome this limitation through the use of viral peptides as vectors for macromolecular delivery into cells [10–15]. Commercially viable therapeutic applications in this direction are not expected to materialize in the short term, because the research efforts are still in the early discovery phase. Another approach under investigation is the use of antibodies directed toward specific internalizing receptors, which can then carry their "payloads" into cells [16]. Cell-surface receptors are often ideal candidates for both small molecule and macromolecule approaches, whereas soluble or circulating proteins are most often targeted by high-affinity macromolecules. Typical selection criteria are summarized in Table 1 and may be used to determine which technology could be applicable. Additional factors that can influence target selection include those listed in Table 2. The key point is that the decision of whether to use a small molecule or macromolecule approach within a discovery project needs to be made early in the process, because intrinsic molecular differences will subsequently affect program strategy all the way through to the clinical setting.

The size and structure of small molecules and macromolecules also have very important consequences for some key steps in drug discovery. It is readily apparent that the molecular weight of a drug determines the "footprint" of the drug on the target and related molecules. Macromolecules, including antibodies and receptor-based fusion proteins, have evolved to be able to accurately distinguish between closely related target structures. They can accomplish this by using structural differences on a larger surface area than is typically accessible for small-molecule affinity. Indeed, the footprint of antibodies is estimated to be on the order of about 5 times larger than that of a typical small molecule. However, high specificity requires careful antibody selection and optimization. Desired specificity has to be accomplished while maintaining orthologue (similar gene family member in another species) binding to enable preclinical development.

**TABLE 1. Criteria for Selecting a Small Molecule or Macromolecule Approach**

| Target Features | Small Molecule | Macromolecule |
|---|---|---|
| | Extracellular Location | |
| Soluble nonenzymatic protein/peptide | – | + |
| Soluble enzyme | + | (+)[1] |
| Soluble hormones and mediators | – | + |
| | – | + |
| Receptor for protein/peptide (proteinergic GPCR) | – | + |
| Receptor for nonpeptide hormone (aminergic GPCR) | + | (+)[2] |
| Pathogen external protein, nonenzymatic | – | + |
| Pathogen, external protein, enzymatic | + | (+)[1] |
| | Intracellular Location | |
| Receptor associated enzymatic proteins | + | (+)[3] |
| Cytosolic enzymatic proteins | + | – |
| Nonenzymatic proteins | + | – |
| Enzymes of pathogenic organisms | + | – |

[1] Enzyme inhibitory antibodies are theoretically possible, but evidence is sparse [17].
[2] Agonistic or agonistic antibodies are theoretically possible, but evidence is sparse [18].
[3] Indirect effect by interaction with extracellular receptor portion feasible; no evidence for enzyme inhibitory antibodies [19].
*Abbreviation*: GPCR = G-protein-coupled receptor.

**TABLE 2. Additional Target Selection Criteria**

| Target Selection Criteria | Small Molecule | Macromolecule |
|---|---|---|
| CNS location | Requires specific compound characteristics for BBB penetration | Required dose to achieve effect *in vivo* may be too high |
| Homology within protein family | Specificity may be difficult to achieve | Specificity achievable, but orthologue binding required |
| Druggability | Assessed after high-throughput library screen, consider macromolecule if screen fails | Consider different approaches if the first fails (*in vivo*/*in vitro*) |
| Protein/protein interaction | Difficult to achieve, but examples available [20] | Often achievable |

*Abbreviation*: CNS = central nervous system.

Another key difference related to size is the difference between small molecules and macromolecules in the degree of interdependence of domains. Although careful evaluation of a structure-activity-relationship (SAR) can readily determine domains in small molecules affecting metabolism, specificity, and other alternative domains, the optimization of these domains will often influence other parameters in a positive or, more often, in a negative manner. In macromolecules, the protein domains are for the most part independent, and optimization of a single parameter, such as specificity in the antigen binding region, will typically not affect half-life, which is determined by a different domain [21–23]. Optimization of small molecules is therefore "multidimensional," whereas macromolecule optimization is, to the most extent, "one-dimensional." For this, and other reasons, drug discovery using macromolecules tends to be more rapid than lead identification and optimization of small molecules.

## 3 TARGET PRODUCT PROFILE

The critical role of an early agreement on a target product profile (TPP) for any new discovery project has been reviewed recently [3]. Briefly, a target product profile describes the exact expectations for the drug candidate in specific quantitative terms. The parameters will vary depending on whether the desired molecule would be "first in class" or "best in class" for a given target. For best & in class, the TPP is often more rigorous because the new therapeutic must be demonstrated to be superior to accepted standard-of-care entities (gold standard) currently on the market. In general, the TPP describes the main development objectives in a precise, focused manner. It can serve as a tool for planning and decision making through the course of the research and development program. Ideally it anticipates product claims that can be realistically expected based on preclinical results and can be translated to and confirmed during development. For small-molecule drugs, this includes the proposed mechanism of drug action, specificity, affinity, *in vitro* and *in vivo* potency, pharmacokinetics, tissue distribution, solubility (LogP [less hydrophobic], $LogS_w$ [more soluble]), metabolism, chemical accessibility, acceptable safety profile (both mechanistic and nonmechanistic), and biomarker requirements. For macromolecules, the criteria include mechanism of action, epitope, specificity (including reactivity with relevant preclinical toxicology species), affinity, potency, constant region isotype, effector function readout, expression rate in mammalian cells, stability in typical formulations, side effects, pharmacokinetics from relevant models, and biomarker requirements. Target product profiles should be agreed to by all functions who will be involved in the project during its lifetime, including discovery, preclinical safety, clinical, manufacturing, and others. Modifications of a target product profile should be considered carefully and only based on new data.

## 4 LEAD IDENTIFICATION

Table 3 summarizes the most prevalent key technologies available today for lead identification of small molecules and macromolecules. This landscape is ever-

**TABLE 3. Lead Identification Technologies**

| Small Molecule | | | Macromolecule (Antibody) | | |
|---|---|---|---|---|---|
| Technology | Description | Reference | Technology | Description | Reference |
| High-throughput screening | An instrumentation-driven, rapid, high-volume approach to assay single compounds or compound mixtures against a biochemical target | Archer [24]; Bajorath [25]; Kariv [26] | | ***In vitro* technologies** | |
| | | | Humanization of murine antibodies | CDR grafting with and without framework mutations | Lo [27] |
| Affinity screening by mass spectrometry | High-throughput assay technology based on ligand–protein affinity as detected by mass spectrometry | Markara [28]; Comess [29] | Resurfacing of murine antibodies | Selective replacement of surface residues of mouse antibodies | Roguska [30]; Zhang [31] |
| | | | Phage display | *De novo* antibody selection from very large and diverse phage libraries | Hoogenboom [32] Marks [33] |
| Directed and focused combinatorial libraries | Knowledge-driven combinatorial library design and synthesis around target class-related structures | Keri [34]; Miller [35] | | ***In vivo* technologies** | |
| | | | EBV transformed human B cells | Useful for natural or anti-pathogen antibodies, can be combined with phage display | Maeda [36] |

**TABLE 3.** *Continued*

| Small Molecule | | | Macromolecule (Antibody) | | |
|---|---|---|---|---|---|
| Technology | Description | Reference | Technology | Description | Reference |
| Fragment-based screening<br>Example: NMR screening/SAR by NMR techniques | A technique for identifying one or more low-molecular-weight, weak affinity molecules for a given target with synthetic optimization driven toward increasing affinity | Mitchell [37]; Verdonk [38]; Petros [39]; Sillerud [40] | Trioma technology | Selection of human antibodies to pathogen or cancer targets by fusion of patient lymphocytes with a human heteromyeoloma line and subsequent selection in microtiter plates | Vollmers [41] |
| Virtual screening | The use of knowledge-driven computational selection and scoring methods to identify molecules that have a high probability for interaction with a biochemical target | Alvarez [42]; Anderson [43] | Transgenic mice | Immunization of mice transgenic for portions of the human Ig repertoire followed by traditional hybridoma technology | Lonberg [44] |

| | | |
|---|---|---|
| Pharmacophore modeling | A complementary virtual selection or design exercise based on specific compound class SAR knowledge; uses three-dimensional conformation of molecules and spatial requirements | Guener [45] |
| QSAR | Mathematical relationships linking chemical structure and pharmacological activity in a quantitative manner for a series of compounds. Methods that can be used in QSAR include various regression and pattern recognition techniques | Lill [46] |

*Abbreviations*: CDR = complementary determining region; EBV = Epstein–Barr virus; NMR = nuclear magnetic resonance; QSAR = quantitative structure-activity relationship; SAR = structure-activity relationship.

changing, and new technological innovations occur frequently. As a caveat, fusion proteins will often use preexisting binding domains (typically receptors) and are not considered in this table.

### 4.1 Small Molecules

***Collections and Screening.*** Frequently, a small-molecule drug discovery approach identifies the initial lead molecules primarily by high-throughput screening of vast corporate compound collections against a biochemical target that is associated with a disease state. In general, most pharmaceutical companies will screen between 750,000 to up to 2 million chemical entities in search of a hit set. Although many corporate collections can exceed the 2 million sample mark, it becomes an economic exercise to reduce this, as well as a case of data management to effectively process and act on the results. A variety of innovative high-throughput screening assay formats have been developed that render this process efficient for surveying a reduced set of 500,000 to about 1.5 million entities as either single compounds or mixtures [25]. With increasing pressure to identify and optimize molecules faster, this process has been supplemented with alternative lead identification approaches such as nuclear magnetic resonance (NMR) [39, 40], fragment-based screening [37, 38], and others (*see* Table 3). Outside of the traditional microtiter plate-based biochemical assays, significant utilization of affinity selection mass spectrometry (ASMS) has become a primary screening paradigm in the past several years. Mass spectrometry-based measurement of small-molecule–protein interaction has subsequently produced results for challenging targets and has provided a cost-effective, rapid method to screen large compound collections while validating hits using plate-based biochemical techniques on smaller subsets [28, 29]. Various *in silico* hit selection techniques have been developed and applied as complementary approaches [47] to minimize the physical number of entities to be tested as well as to optimize the chances, based on appropriate training sets, to identify active chemical space within a large collection. The most common techniques involve a target-based virtual screening process whereby compounds with molecular similarity to known active small molecules are estimated to have affinity toward a biological target binding site as estimated through dock-ing and scoring functions [48, 49]. Selection techniques can be supplemented through the use of pharmacophore modeling [45] and quantitative structure-activity relationship (QSAR) screening [46]. Results with this approach can be limited because it does not initially involve direct biological assay of diverse compounds but a knowledge-based selection of a "high probability" subset of compounds to be screened. A high level of success remains to be achieved for this technique. More frequently, a diverse subset of compounds that should represent "hit-like" and "lead-like" molecules is selected based on calculated physicochemical properties [26]. These compounds are expected to provide ideal starting points for a medicinal chemistry optimization program because they are preselected to have the probability of acceptable solubility, perme-ability, and properties such as molecular weight where they can be subsequently elaborated on and optimized.

To further enhance the chances for success in the screening arena, most pharmaceutical companies have expended considerable effort in acquiring molecules

with hit-like and lead-like properties to improve the quality of their collections [50]. Besides general acquisitions of diverse arrays of molecules, many have also focused on the generation of gene family or target class directed libraries [34], and a slight resurgence in natural products and natural product-like libraries has been seen [51]. Currently on the horizon are the emerging technologies using the concepts of chemical genomics and high content screening, and a variety of industry and academic laboratories are paving the path forward [52].

***Hit to Lead.*** The overall probability that a high-throughput screening (HTS)-derived lead will reach launched drug status is about 1 in a million, shifting the focus of initial post-HTS research efforts toward identifying and characterizing good quality leads [53]. The distinction between hit (compound meeting initial screening criteria) and lead (verified on target activity and selectivity with initial exploration of SAR) is critical [54–56]. It is felt that a rigorous hit to the lead selection process will enhance the probability of overall success. Typical evaluation metrics that a compound series should meet, in order to proceed to lead status, are listed in the points below in Table 4.

**TABLE 4. Typical Assessment Parameters for Start of Lead Optimization of a Small Molecule**

| Attributes | Hit-to-Lead Assessment Parameters |
|---|---|
| Chemical | Reproducible activity between HTS assay and local laboratory assay |
| | Dose responsive activity |
| | Confirmed structural identity |
| | Purity established |
| | No evidence of compound class instability |
| | Tractable synthetic route established |
| | Favorable intellectual property and competitive assessment for compound class |
| | Demonstrated exploitable SAR |
| Biochemical | Support for interaction with molecular target (SAR due to desired mechanism) |
| | Selectivity profile established |
| Pharmacology | Assessment of druggability and potential attributes that will need optimization |
| | Secondary assay paradigm established (cell assay, mechanism assay, *in vivo* model, etc.) |

*Abbreviations*: HTS = high-throughput screening; SAR = structure-activity relationship.

With this, it is no surprise that hit to lead has emerged as a unique discipline within many companies, typically built or grown at the expense of the former technology-driven combinatorial chemistry groups. Some companies even have devoted entire departments to the activity so that consistent methods and quality metrics are applied in a rapid time frame [57].

### 4.2 Macromolecules

The lead identification process of macromolecules, and especially therapeutic antibody leads, differs for the *in vivo* and *in vitro* approaches (*see* Table 3). In the case of the *in vivo* approaches, the screening of initial leads occurs through an enzyme-linked immunosorbent assay or through functional assays drawn parallel to the HTS approaches discussed above. However, the number of antibody candidates typically screened is significantly lower because target-specific antibodies have been enriched by the immunization of transgenic mice [44] or preexist in patients. For the *in vitro* approaches, the selection is template based. As an example, the antibodies from phage or yeast libraries are isolated by a process similar to affinity chromatography [32, 33].

## 5 LEAD OPTIMIZATION

The biggest difference between small-molecule drugs and therapeutic antibodies lies in the area of lead optimization; however, the overall concepts are very similar. The original leads are optimized in iterative cycles to achieve candidates meeting the pre-agreed target product profile. The drug structure and size makes the lead optimization of small molecules multidimensional (change in one parameter often affects another) and therefore lengthy, whereas the domain structure of antibodies allows for rapid, primarily one-dimensional optimization of the binding and effector regions, where optimization of one has little or no effect on the other.

Although opportunities for macromolecule optimization were limited in the past, antibody technologies today allow one to dial in, more or less, any affinity, potency, or specificity as directed by the target product profile.

### 5.1 Small Molecules

In traditional small-molecule drug discovery, hit-to-lead selection is followed by careful, systematic exploration of SAR and *in vivo* compound properties. Iterative cycles of design and data gathering are used, often supported by generating small-molecule–target binding hypotheses using computational resources. For small molecules, this represents the research stage where chemical series with demonstrated affinity of the receptor, enzyme, or protein target of interest are optimized to build in drug-like properties. More properly defined, lead optimization is the synthetic modification of a biologically active compound, to fulfill all stereoelectronic, physicochemical, pharmacokinetic, and toxicologic properties required for clinical usefulness.

This is a joint potency, selectivity, and property optimization process that builds on data (design/analysis cycle) generated during the hit-to-lead prioritization phase and is heavily driven by absorption, distribution, metabolism, excretion, and toxicology (ADMET) and physicochemical property profiling as early as practical. Combinatorial or parallel synthesis techniques resulting in the generation of scaffold-targeted libraries are often used to understand and develop both the chemical and the intellectual property (IP) space. If amenable, specific analog design guided by QSAR or homology modeling hypotheses is incorporated. It is

preferable to have more discrete structural information either from protein NMR studies or ligand/target x-ray crystallography if tractable for the target.

In the effort to choose and accelerate programs that offer the best possible chance to move forward, a value and risk assessment at the start of lead optimization typically occurs along with project and resource prioritization (Table 5).

**TABLE 5. Typical Assessment Parameters for Start of Lead Optimization**

| Priority | Lead Optimization Assessment Parameters | Small Molecule | Macromolecule |
|---|---|---|---|
| 1st | Potential for therapeutic advance | + | + |
| | Commercial potential/market size of indication area | + | + |
| | Competitive environment | + | + |
| | Intellectual property position | + | + |
| | Relevance of target for disease | + | + |
| | Feasibility of biological predictive preclinical model | + | + |
| | Quality of lead candidates (several dimensions, *see* hit-to-lead criteria) | Specific | Specific |
| 2nd | Target related tolerability | + | + |
| | Fit to therapeutic area strategy/portfolio | + | + |
| | Clinical trial design feasibility | + | + |
| | Amenability of rational or structural drug design | + | – |

The chemistry strategy involves defining critical regions of the molecule that can be identified as interacting with the specific amino acid residues of the targets in a three-dimensional model and thus drives potency and selectivity. Areas of the molecule that are SAR "neutral" can be found, which can be used as positions to alter the net physicochemical properties. Small modifications will often affect the *in vivo* bioavailability, the metabolism profile, and the absorption properties of the molecule. Optimization of chemical entities is therefore always multidimensional, and a compromise must be reached before preclinical candidate selection can occur. Any improvement in specificity may be gained with the loss in absorption or half-life, a correction of which will again have effects on specificity. Because of the many iterations of changes and retesting, traditional drug discovery projects are typically long and very costly. Not infrequently, non-mechanism-related toxicity is observed later in development and is not predicted based on the assays available preclinically. To address some of these complications, new methods for predictions of half-life and toxicity are being developed [58]. Introduction of early experimental ADME testing has been credited with significant improvements in the probability of successful lead optimization [59, 60] and is now a widely used practice. To date, the crucial component to small-molecule discovery is the ability to assess, as early as possible, the potential for ADMET failure. The key is to "fail fast" in order to minimize the substantial total research and development investment and costs incurred in the process of learning the shortcomings of a candidate at the clinical trial stage.

### 5.2 Macromolecules

Macromolecules, like antibodies, have a defined domain structure separating the domains for specificity and affinity (Fv) from the constant region domains responsible for half-life (Fc receptor [FcRn] binding domains) [61] and effector functions (antibody-dependent cellular cytotoxicity [ADCC] and complement-dependent cytotoxicity [CDC]) [62]. This separation is important because optimization can be done in the first instance in "one dimension." Extensive variation of the complementary determining region 3 (CDR3) regions of the heavy or light chain to increase affinity will in most or all cases have no affect whatsoever on half-life or effector functions of the antibody. Similarly, changes in the half-life or effector functions typically would not affect the affinity. However, examples of interaction between the Fc domain and the Fv antigen binding site have been reported [63, 64] and can be monitored and ruled out by appropriate testing at each stage of optimization. It is easy to see that this one-dimensional optimization should be more straightforward.

Antibody optimization of the antigen binding portion should focus on hypervariable regions, primarily both the heavy- and the light-chain CDR3 regions and preferably the heavy-chain CDR3, as not to introduce undesirable epitopes. Older methods for antibody optimization include a variety of mutation and selection techniques [65], most of which are tedious. The optimal method is yeast display, combining rapid library creation and rapid selection using cell sorters, with a wide range of selection options (selection for optimal "on-rate," "off-rate," "affinity," etc.) [66, 67]. Other methods include ribosome display [68]. Fc optimization includes optimization of effector functions [62] and half-life [21]. Interestingly, the differential effects of Fc mutants on either Fc effector functions (ADCC or CDC) or half-life is further proof that the overall concept of one-dimensional antibody optimization remains correct. This notwithstanding, careful analytical follow-up at the end of the project or at appropriate intervals with meaningful and well-characterized assays will ascertain that indeed no change in functions of another domain has occurred. Because of the specificity and lack of accessibility to intracellular targets, antibodies typically are devoid of nonmechanistic toxicity, which further reduces the discovery and development risks. However, meaningful and appropriate safety testing still will be necessary to assure the safety of therapeutic antibodies [69].

## 6 SELECTION OF CLINICAL CANDIDATES AND PRECLINICAL DEVELOPMENT

Selection of clinical candidates and the identification of new chemical entities or new biological entities that best match the TPP is a multidisciplinary task that relies heavily on *in vitro–in vivo* correlations. For first-in-class targets, mechanism of action validation (target validation) research typically parallels the lead identification and lead optimization phases of any drug discovery project. For best-in-class targets, the preferred initial small-molecule validation reagents are competitor's leads or gold standard therapeutics currently on the market, their own initial screening leads, or monoclonal antibodies to the target orthologue. As the project

progresses, target validation continues with the drug leads until the *in vivo* data in relevant disease models meets the expectation set out in the TPP. For macromolecule discovery projects, target validation requires either surrogate antibodies or drug leads that cross-react with the orthologue of the preferred pharmacology model species. Usually, a small series of candidate molecules that best match the TPP set at the initiation of the project are subjects of consideration. Outliers are also considered if they can provide unique insight or datasets that might challenge the original TPP, because potential opportunities may not have been realized at the outset of the program. As the project progresses, target validation continues with the selected drug leads until the *in vivo* data in relevant disease models meets the expectations set out in the TPP. For macromolecule discovery projects, target validation requires either surrogate antibodies or drug leads that cross-react with the orthologue of the preferred pharmacology model species.

The topics in Table 6 are frequently evaluated as part of the risk assessment before starting development activities. Comprehensive *in vivo* efficacy and preliminary safety testing, preferably in multiple species, with simultaneous pharmacokinetic assessment is typically the last dataset collected before making a decision on a development candidate and possible back-up compounds. The Core Battery for safety pharmacology, which can be found in the International Conference on Harmonisation (ICH) Guideline S7A (Good Laboratory Practice [GLP]) document series [70], governs safety assessment and data generation guidelines. The suitable second (nonrodent) toxicity species should be identified in combination with drug metabolism, pharmacokinetics, and pharmacology studies in that species. An overall evaluation of the toxicity and toxicokinetic data should be undertaken to understand the risk for adverse effects, target organ issues, and systemic exposure liability. This should allow researchers to set a preliminary limit dose (mg/kg in relation to $C_{max}$) for toxicity studies and give recommendations for further

**TABLE 6. Topics Evaluated as Part of the Risk Assessment Prior to Starting Development Activities**

- CMC developability drug substance
- CMC physicochemical properties drug substance
- CMC developability formulated drug product
- Relevance of target for disease/indication
- Efficacy in experimental models
- Feasibility of clinical proof of concept
- Exposure and half-life
- Linearity of pharmacokinetics
- Drug–drug interaction potential
- Safety pharmacology
- General pharmacology
- Toxicology
- Potential for therapeutic advance
- Commercial potential/indication area/market size
- Fit to therapeutic area strategy/portfolio
- Patent status/freedom to operate

*Abbreviation*: CMC = chemical manufacturing and control.

toxicity assessments. The assessment also should include data from predevelopment lead optimization work. For small molecules, the 2-week toxicology study in rodents, including toxicokinetic data, should be suitable as dose-range finding for further repeat-dose (4 week) studies. If performed according to GLP, the safety data from the 2-week study could support phase I. Optional toxicity studies at this stage can include, but are not limited to, single-dose toxicity (GLP, rat) and genotoxicity tests (GLP, Ames, mouse lymphoma) [71, 72].

Early interaction with process research and development teams allows for intermediate-scale synthesis and production of material required for these advanced studies and will enable an assessment of any possible issues around scale-up, cost of goods, stereochemical issues, and compound solid-state crystallinity and stability. Selection criteria are often rigorous, and given the multidisciplinary nature of the profiling, collaborative, tightly scheduled efforts are required across research and development department structures. The preferred criteria set for candidate molecules include those listed in Table 7.

A discussion of regulatory requirements for the preclinical safety assessment of drug candidates intended for use in humans would require significant detail. The reader is referred to excellent resources and websites (www.ich.org, www.fda.gov, and www.emea.com) for detail and additional information.

Although the safety assessment of small-molecule drug candidates is more standardized, the preclinical program for a macromolecular biological medicine requires adaptation (i.e., case-by-case and highly tailor-made approach), simultaneously meeting the basic requirements for types of studies, duration, and exposure [73–75]. The key principles are outlined in ICH guidance S6 and Center of Biologics Evaluation and Research (CBER) 1997, and some key highlights are summarized in Table 8 [76, 77].

## 7 PRODUCTION OF ACTIVE PHARMACEUTICAL INGREDIENTS

Differences in structural complexity, continuous versus stepwise production, and the overall production techniques and equipment results in significant differences between the manufacturing of small and macromolecule bulk drug substances or active pharmaceutical ingredients (APIs). In addition, continuous improvement of analytical tools is changing the landscape of both process types over time. Small-molecule synthesis consists typically of 6 to 10 (or more) separate synthetic steps, whereas macromolecule production consists of 15 to 20 connected steps. Each step in small-molecule synthesis carries its own risk, but because they are separate or can be separated from each other, low efficiency or failed steps can be repeated, thus minimizing loss. On the macromolecule side, the process is continuous and all steps are directly linked. Failure at any step will typically result in complete batch failure. Another key difference is that small-molecule process research can and often begins during the lead optimization phase of research. This "front-loaded" model [78] is in place in many companies, and initial research toward the efficient production of a promising chemical series is initiated, at risk, early on. With a macromolecule, typically the final sequence or composition of the candidate molecule must be clearly identified so that appropriate constructs can be generated and then optimized for fermentation processes.

**TABLE 7. Criteria for Development Candidates for Internal Decision Making**

| Criteria | Small Molecule | Macromolecule |
|---|---|---|
| Chemical Manufacturing and Control | | |
| Synthetic accessibility | Synthetic steps, chiral centers, reagent cost of goods | — |
| LogP | $-4 \leq \log P \leq 4.2$ | — |
| LogSw | ≥–5 | — |
| Formulation | Project specific | ≥50 mg/mL |
| Production yields | Project specific (overall cost of goods) | 50–100 pg/cell/day |
| *In Vitro* Pharmacology | | |
| Cell assay potency with and without serum | Project specific | Project specific |
| Protein binding | Availability required | — |
| Target specificity | Specificity within close and distant target family members | ELISA specificity within diverse set of related proteins |
| Affinity | Project specific | BiaCore $>10^{-10}$ M |
| Tissue cross-reactivity | — | Binding to tissues with known target expression |
| *In Vivo* Pharmacology/PK/PD/ADMET | | |
| Efficacy in relevant model at appropriate dose, route, and schedule | Project specific | Project specific |
| Metabolites | Identified and characterized | — |
| PK/PD relationship | Project specific | Availability desired |
| Biomarker for PD effect | Availability desired | Availability desired |
| No overt toxicity in repeat dose (1 to 4 week); toxicity studies in a rodent and nonrodent species | Availability required | As follow-up for unexpected pharmacological observations |
| Drug–drug interaction | No significant interactions | — |
| Human CYP inhibition | No relevant interactions; also consider comedications | — |

*Abbreviations*: ADMET = absorption, distribution, metabolism, excretion, toxicology; CYP = cytochrome P450; ELISA = enzyme-linked immunosorbent assay; PD = pharmacodynamics; PK = pharmacokinetics; — = not applicable.

**TABLE 8. Key Issues Around the Designs of Safety Studies [76, 77]**

| Issue | Small Molecule | Macromolecules | Guidance for Macromolecules |
|---|---|---|---|
| Selection of relevant animal species | Typically 1 rodent and 1 nonrodent species | Studies in irrelevant species are of no value or might even be misleading.<br>Criteria for species selection include:<br>• Tissue cross-reactivity studies comparing human and animal tissues<br>• Pharmacologically active (i.e., expressing the appropriate epitope or receptor) | • Closest pattern to human tissues defines animal species<br>• One species (often the monkey) might be sufficient; if no species available, use of a surrogate and/or transgenic animal might be appropriate<br>• Early and close interaction with agencies highly recommended |
| Dose level selection | Based on the anticipated highest clinical dose, high dose should show toxicity | | Three dose groups plus a control group with multiples of anticipated highest clinical dose (e.g., 2, 10, and 50 times clinical dose); high dose selection often not driven by toxicity |
| Treatment frequency | Should parallel the treatment regimen in humans | | Depending on PK and immunogenicity, a higher dose frequency might be needed to achieve comparable exposure |
| Local tolerability | With oral dosing not an issue | Application route predominantly IV and/or SC; it is important to test the clinical formulation | Visual inspection and histopathology of injection site |
| Recovery period | Long enough to allow clearance of the drug | | At least 5 half-lives; can be up to 3 months |
| Immunogenicity | Not an issue | Induction of ADA, characterization (i.e., neutralization, impact on drug exposure because of increased clearance)<br>At high dose level immune response often suppressed | • If exposure decreases due to ADA formation, no adequate long-term treatment might be possible<br>• In general, animals are not predictive of immunogenic response in humans |

| | | | |
|---|---|---|---|
| Genotoxicity studies | Standard battery | Not applicable | Interactions with DNA or chromosomes is not expected; large amounts might even lead to uninterpretable results |
| Safety pharmacology | Standard core battery | For mAbs, generally not necessary. If cause of concern (e.g., cross-reactivity with CNS, CV, or respiratory tissues), specific study might be indicated | As many parameters as possible to be included in toxicology study. A study in telemetrized animal to address CV and respiratory endpoints might be an alternative approach |
| Carcinogenicity | Required | Not applicable | The standard carcinogenicity bioassays in rat and mouse are generally inappropriate |
| Reproductive toxicology | Standard package (Fertility,embryo-fetal toxicity and peripost natal toxicity) | Dependent on product, intended use and patient population. Study must be conducted in relevant species. Expectations from authorities still different for U.S., EU, and Japan regarding timing of the studies | The primate is often the only relevant species; however, because of the small numbers of animals available and the lack of a broad dataset, testing of a rodent homologue might be an alternative<br>U.S. not required to support phase I/II |
| ADME | Standard battery *in vitro* and *in vivo* | Not applicable | Not relevant as the drug enters the normal proteolytic degradation pathway |
| Biomarkers | Should be included whenever possible to demonstrate the intended pharmacology of the drug | | PK/PD relationship helps to define safety margins as well as selection of dose level and regimen |

*Abbreviations*: ADA = anti-drug antibodies; ADME = absorption, distribution, metabolism, excretion; CNS = central nervous system; CV = cardiovascular; EU = European Union; IV = intravenous; mAbs = monoclonal antibodies; PD = pharmacodynamics; PK = pharmacokinetics; SC = subcutaneous; U.S. = United States.

### 7.1 Small Molecules

As discussed, pressure to move new medicines to the market faster has compressed many aspects of the research and development cycle. Process research initiates their search for scalable routes early in the discovery programs in the hope of having an efficient or semi-efficient route to prepare up to 100 kg of API as a program nears the final preclinical stages and prepares for phase I studies in humans [79].

Although any synthetic transformation or use of a specialized reagent can be scaled up, cost and safety are two primary driving factors in the manufacturing process. In a recent survey of GMP bulk reactions run in the last several decades, a predominant subset of transformations were shown to fall into favor [80], with less than a 2% change within the categories. This result indicates that significant process research efforts are often required to advance an initial medicinal chemistry analog synthesis route to one that is amenable for use in large-scale production. Once route selection occurs at phase I, synthetic efficiency is critical and reactions must be controlled so that consistency of yield and purity is ensured, especially as a program nears phase II in the clinical setting. This is the point at which the process for generating the API must be finalized and approval packages initiated for the regulatory agencies. To control costs, many manufacturers will partner with specialty firms to transform key synthetic intermediates in the stepwise production process toward the final API (which is almost always prepared within the company's facility to meet regulatory requirements). Such external, contract synthesis often involves sensitive or specific technology-based transformations or separations for purification that would have significant cost impact if a production facility needed substantial capital improvement and/or steep licensing fees to conduct. Environmental aspects of production also have become increasingly important. Considerable steps are taken to minimize waste streams, and the advent of "green chemistry," a move to decrease usage of organic solvents in favor of water, when appropriate, is actively under investigation [81]. In addition, energy-efficient methods and technologies such as microwave-driven processes are being explored to accelerate cycle times [82] within synthesis routes.

Small-molecule manufacturing is also a system of checks and balances in which any small increase or decrease in reaction step performance can have consequences downstream. As always, patient safety is of utmost concern. Any manufacturing process changes that alter an impurity level or introduce a new impurity, even as low as 0.1%, may necessitate additional toxicity studies and documentation for review and registration. In addition, an improvement in reaction efficiency may alter bulk product crystallinity or polymorph composition that can affect formulation and human pharmacokinetics. Once process parameters are finalized, the ultimate manufacturing step involves selection of a manufacturing site, transfer of the process, and preparation of a demonstration batch followed by a minimum of three consecutive validation batches of API to demonstrate that the synthesis of material can be controlled within analytical specifications and reproducibility.

### 7.2 Macromolecules

Macromolecules, including therapeutic antibodies and fusion proteins, are complex molecules requiring extensive analytical characterization. Native antibodies com-

prise a heterodimer of two light and two heavy chains with a single N-glycosylation site in the heavy-chain constant region. Proper assembly and glycosylation is important for antibody function. Fusion proteins often consist of natural binding proteins (such as receptors) fused to the nFc (CH2 and CH3 IgG domains) region. Due to interchain disulphide bridges, Fc-fusion proteins assemble into homodimers. The complexity of fusion proteins depends on the structure of the ligand binding domain and the number of N-glycosylation sites. The single N-glycosylation site in antibodies is "buried" into the opposite heavy chain constant region, whereas the carbohydrate residues on ligand binding portions of Fc-fusion proteins are exposed and may have a pronounced effect on the *in vivo* behavior of the molecule. For these reasons, full-length antibodies and fusion proteins should be expressed in mammalian cells. Modified yeast systems capable of appropriate glycosylation and fungal systems have been proposed [83]. By contrast, antibody fragments have been expressed successfully in *Escherichia coli* and yeast. Decisions for an expression system should be based on the complexity of the molecule, the specific biological activity achieved with a given system, and the appropriate technical, IP, and regulatory considerations for manufacturing at scale. Expression of therapeutic antibodies in a variety of systems has been reviewed recently [84, 85]. Methods include transgenic animals [86], expression in plants [87], or the more standard myeloma systems [88] or CHO cell expression [89]. The recovery and purification of therapeutic antibodies also has been reviewed recently [90]. For the reasons discussed, when designing purification schemes, attention should be paid to reduction of aggregation and DNA content, especially if bacterial systems are involved, to minimize any immunogenicity risks. In each case, a defined set of assays needs to be applied to the purified final product to ascertain the purity and specific activity of the expressed antibody. This is especially important if reconstitution and refolding is part of the process [91].

Historically, macromolecule products were characterized with a limited set of tools and additional methods beyond that were defined by their specific process. This led to complications when changes in the process were desired. A chance event changed this landscape. Briefly, biotechnology company Biogen, Inc. (today Biogen Idec, Inc., Cambridge, Massachusetts) and the Rentschler daughter company, Bioferon GmbH (Laupheim, Germany), formed a joint venture in 1984 for the clinical development of interferon-beta, called BG9015. Rentschler was responsible for the creation of the cell line and production and Biogen for clinical development, which commenced in 1990 with a U.S. clinical study in patients with multiple sclerosis. When this joint venture failed in 1993, Biogen completed the initial trial with the remaining BG9015 and switched to BG9216 from their own process. After the U.S. Food and Drug Administration (FDA) determined that these products were not identical, a new process for what is today called Avonex (interferon beta-1a; Biogen Idec, Inc.) was approved and was considered sufficiently identical with the original BG9015 for the original study results to be considered as part of the Avonex biologic license application. Extensive biological, biochemical, biophysical, and human pharmacokinetic data were the basis for the FDA conclusion that BG9015 and Avonex were "biochemically and functionally equivalent," leading to Avonex approval on May 17, 1996 [92]. This decision created a path for others to follow, leading to a fundamental change in macromolecule production and ushering in the concept of "well characterized biologic" [93]. Increasingly sophisticated analytical

techniques have facilitated process changes even further, and today, macromolecules require very detailed characterization, resembling more and more the requirements for small-molecule pharmaceutical ingredients [94]. Indeed, analytical tools have become so sophisticated that much of the current dispute on biogenerics (or "follow-on biologics") focuses on the true equivalence of macromolecule products [95].

## 8 FORMULATION AND DELIVERY DEVICES

Drug product development takes place in a highly regulated environment with a clear expectation for quality, safety, and efficacy in the final product. From a formulation perspective, quality primarily translates into a defined composition, assurance of purity, potency, an acceptable stability, and availability of a robust and consistent large-scale manufacturing process. Thus, the ultimate goal of formulation activities is to achieve the systemic delivery of a drug product that can be successfully commercialized. Product life-cycle management strategies are considered at the beginning of development and often include evaluation of specialized administration devices for parenteral (injectable) products, especially for macromolecules. If these options are considered early in development, a product family can often be designed to minimize subsequent development efforts by using a common platform (e.g., prefilled syringe-based autoinjectors). One challenge in the development of devices is a thorough understanding of the limitations of the device, limitations of the formulation, and finding the right compromise between both. Also, one has to keep in mind that the regulatory path for devices and fixed-device drug combination products is different, and significant differences exist among the major markets such as the United States, the European Union, and Japan.

The scope of formulation activities is dependent on the development stage of the drug substance. In general, the development process can be divided into three major phases, namely formulation support of discovery, formulation activities until clinical proof of concept (end of clinical phase IIa), and formulation activities in the commercial phase (phase IIb until market).

### 8.1 Formulation Support in Discovery

Formulation support in discovery usually starts at the lead optimization phase with a focus on key properties of molecules to support selection of the best candidates. The preferred route of administration for small molecules is oral, which requires the development of a solid dosage form. Because only dissolved molecules can be absorbed, assessment of solubility is of prime interest. In general, macromolecules are not orally absorbed and a parenteral dosage form is required. Some proteins, like monoclonal antibodies, often require high doses, and therefore, it is important to understand the solubility range for the protein in question. Compared with small molecules in which established strategies exist for tackling solubility problems on the basis of chemical modifications, solubility issues with proteins are mechanistically poorly understood and very difficult to address on a molecular basis. Most small-molecule drug candidates are crystalline in nature, whereas macromolecules are usually dissolved in aqueous media. The composition/formulation of the solution determines the properties of the protein, and therefore, a basic understanding

about suitable pH ranges, temperature sensitivity, buffer composition, and so on, is important. These parameters have to be defined and controlled early in the discovery phase. The complex molecular structure of proteins poses a significant challenge to the analytical techniques and the ability to detect the instabilities of the molecule. Characterization of degradation products is a prerequisite to understand the major degradation pathways of the protein, which will enable a rational design of the formulation. Degradation pathways of proteins include deamidation, oxidization, hydrolysis, fragmentation, and aggregation. Formulation strategies have been developed to address these instabilities and can act as guiding principles in early development phases. Once the selection of clinical candidates has been initiated, the number of drug molecules will be narrowed down significantly and profiling of the proposed candidate for development can take place. For oral small molecules, solid-state characterization is a very relevant topic for a solid dosage form as summarized in Table 9. For a protein delivered parenterally, it is

**TABLE 9. Formulation Attributes at the Development Project Stages**

| Stage | Small Molecule | Macromolecule |
|---|---|---|
| | Discovery Stage | |
| Common dosage route (all stages) | Oral | Parenteral |
| Critical factors to optimize | Bioavailability, solubility, logP, permeability, absorption | Bioavailability, solubility, pH stability, freeze/thaw behavior |
| Physicochemical attributes | Crystalline solids or salts | Solution stability, aggregation |
| | Early Development Stage | |
| Common dosage form | Powder in bottle, tablets | Lyophilized powder in bottle |
| Critical factors to optimize | Dosage form | Temperature sensitivity and storage conditions, dose per unit and dosage form |
| Physicochemical attributes | Crystal polymorphism, solid-state stability, optimized salt selection | Conformational stability, buffer formulation, surfactant or excipient stability, aggregation |
| Device | Tablet, capsule, oral suspension or solution | Prefilled syringe |
| | Commercial Stage | |
| Develop commercial process | Yes | Yes |
| Batch scale-up | Yes | Yes |
| Batch stability | Yes | Yes |
| Process validation | Yes | Yes |
| Establish commercial supply chain | Yes | Yes |

important to define the composition of the drug substance. Ideally, the drug substance composition should be as simple as possible to allow for flexibility in later development phases. In the most simple situation, the drug substance composition consists of a protein (concentration at least half of the highest single dose for first in human studies), a buffer system (suitable for parenteral administration) in a pH range of 5 to 8, and a surfactant, if necessary.

### 8.2 Formulation Activities Until Clinical Proof of Concept

For early clinical trials, preliminary dose formulations or "clinical service forms" are developed. Small-molecule oral products can be either powder in a bottle or oral solutions/suspensions. Proteins are usually initially formulated as lyophilized powders that maximize long-term stability at the expense of a lesser convenient dosage form due to the necessity of the reconstitution step. Parenteral dosage form manufacturing requires a highly technical and sophisticated infrastructure, including highly trained personnel and well-established GMP systems in order to assure sterility of the products. These requirements limit flexibility in the development of parenterals compared with oral products. Temperature stability studies have to be conducted to justify use of material that has experienced temperature variations on storage. This is usually not an issue for small molecules, but sometimes specific safety considerations are more important and require special handling and shipment conditions.

During the course of the early clinical development program, a more detailed definition of the commercial product evolves. The most important information is the dose per unit and the specific dosage form. Oral products often require modified release dosage forms in which extensive pharmacokinetic modeling/testing of prototype formulation have to be performed. For parenteral products, self-administration and convenience is becoming more and more an expectation. Prefilled syringes that allow subcutaneous administration of liquids satisfy this need. The main issues involved with development of a liquid protein formulation in a prefilled syringe are using volumes less than 1 mL (consequently, protein concentration can be as high as 100–200 mg/mL), aggregation of proteins due to the high protein concentration, viscosity issues, interaction of the protein with silicone oil present in the prefilled syringe, and the higher chance of chemical degradation in a liquid solution. Aggregation of proteins and conformational stability is a major obstacle because reliable detection of small amounts of aggregates is analytically challenging and aggregates or conformationally altered protein molecules are suspected to be involved in undesired immunogenic responses to therapeutic proteins. Conformational changes can be induced by small changes in drug substance/drug product manufacturing. Therefore, changes in manufacturing have to be carefully evaluated on product impact and often require additional stability testing because the impact is not predictable.

### 8.3 Formulation Activities in the Commercial Phase (Clinical Phase IIb until Market)

The main activities are similar for small molecules and macromolecules at this stage. The primary focus is in the areas of scale-up, generation of primary batches

for stability testing, and development and validation of the commercial manufacturing process.

For solid dosage forms, scale-up is a major challenge and includes ensuring the bioequivalency of the clinical product to the commercial product. For proteins, shear-induced aggregation can be an issue, as well as establishment of a commercial supply chain that is robust, reliable, and able to deal with large amounts of temperature-controlled materials.

## 9 CLINICAL DEVELOPMENT

There are differences in the pharmacokinetic properties of typical small molecules and monoclonal antibodies that will affect the clinical development path. A few of the differences are listed in Table 10. Small molecules are usually administered orally, which is desired because of convenience for patients. On the other hand, the administration of monoclonal antibodies is limited to the parenteral route, i.e., intravenously (*i.v.*), intramuscularly (*i.m.*), or subcutaneously (*s.c.*). Because of the availability of liquid formulation or reconstituted lyophilized powder to support parenteral administration, measurement of absolute bioavailability after *s.c.* or *i.m.* administration is necessary for monoclonal antibodies. The relative bioavailability is usually obtained for small molecules because of the expense of developing liquid formulations. Both renal and hepatic clearance pathways are important for elimination of small molecules, and an understanding of drug disposition in patients with renal or hepatic impairment is important to support the registration documentation. Small molecules are often metabolized in the liver by cytochrome P450 enzymes; therefore, drug–drug interaction studies are an important component of the clinical data package. Therapeutic monoclonal antibodies follow the typical pathway of elimination for immunoglobulins (IgGs), involving the neonatal (FcRns). Briefly, endothelial cells lining the blood vessel are thought to be the major site of IgG homeostasis, with the FcRns acting as salvage receptors,

**TABLE 10. Differences Between Small Molecules and Therapeutic Monoclonal Antibodies that Can Affect Clinical Development**

| Small Molecules | Monoclonal Antibodies |
|---|---|
| Usually administered orally | Given parenterally |
| Transporters often important for disposition | Transporters are not important for disposition |
| Metabolism by hepatic cytochrome P450 | Proteolytic degradation |
| Renal clearance often important | Renal clearance is not important for molecular weight higher than 20 Kd |
| Half-life is shorter (usually <24 h) | Half-life is long (1–3 weeks) |
| PD usually does not affect PK | PK can be dependent on PD |
| Immunogenicity is a not an issue | Immunogenicity is an important issue |

*Abbreviations*: PD = pharmacodynamics; PK = pharmacokinetics.

protecting IgGs against lysosomal degradation. The IgGs are taken up by endothelial cells by nonspecific pinocytosis and then enter acidic endosomes. If the IgG is bound to the FcRn, then it is protected and released back into the blood stream; otherwise, it is degraded. Thus, the FcRns serve the dual function of transcytotic shuttle (i.e., placental transfer, etc.,) and homeostat [22, 61]. As long as therapeutic antibodies bind to the FcRns, they are thought to participate in this pathway. Whether modulation of that interaction by mutating the IgG Fc portion can predictably extend or shorten IgG half-life is the subject of ongoing research. Metabolic interactions such as inhibition or induction of cytochrome P450 enzymes are usually not a consideration. Transporters must be a consideration for the understanding of the disposition of small molecules, whereas except for FcRns transporters, transporters are of lesser concern for monoclonal antibodies.

Allergic reactions are a potential issue for both small and large molecules. The potential for development of significant levels of antidrug antibodies needs to be considered for any recombinant protein, including fusion proteins and monoclonal antibodies, regardless of whether the therapeutic antibody is chimeric, humanized, or fully human. Assessment of immunogenicity begins in phase I and continues throughout clinical development. Incidence of immunogenicity can differ for each patient population and should be assessed for each indication. The development of assays to assess immunogenicity begins before filing the investigational new drug application. Validated assays are necessary by phase III. In the case of fusion proteins, the cross-reactivity of antidrug antibodies with the endogenous human protein in the patient needs to be considered.

## 10 INTELLECTUAL PROPERTY

Farnley and colleagues [96] have recently reviewed the complexity of intellectual property around biotechnology and biotechnology products. The authors note that the IP issues for small molecules and macromolecules, with a few exceptions, may be more similar than commonly assumed. The most important difference relevant to this overview is that each step in the discovery process may potentially be subject to different intellectual property considerations, which the authors note as unusual. The major differences between small-molecule and macromolecule patenting are shown in Table 11.

**TABLE 11. Intellectual Property Considerations for Small Molecules and Macromolecules**

| Patents | Small Molecule | Macromolecule |
|---|---|---|
| Target | Yes | Yes |
| Drug class (“antibody to target X”) | No | Yes |
| Substance | Yes | Yes |
| Drug screening | Yes | Yes |
| Drug synthesis steps in discovery stage | No | Yes+ |

## 11 GENERICS, "BIOSIMILAR PRODUCTS," OR "FOLLOW-ON BIOLOGICS"

Although a full assessment of generics is beyond the scope of this introduction, the differences between small-molecule and macromolecule generics are highlighted briefly in Table 12. Importantly, the terminology "generics" should not be used for noninnovator developed off-patent macromolecules. They are not "generic" in the true sense, and "biosimilar products" or "follow-on biologics" are the more appropriate terms. Reasons include the independently developed manufacturing process, the resulting possible differences in product heterogeneity, and the need for preclinical and clinical development. Many companies around the world are working toward registration of off-patent recombinant proteins. Interferons, human growth hormone, and erythropoietin are some of the more popular therapeutics that are targeted.

The development path for small-molecule generics has been clear for many years. In contrast, the path for biosimilar products or follow-on biologics has only recently been clarified in Europe (October 2005), and regulatory uncertainty remains in the United States. The new European guideline 2004/27/EC clearly

**TABLE 12. Regulatory Criteria for Biosimilar Products or Follow-On Biologics**

| | Small Molecule[1] | | Macromolecule | |
|---|---|---|---|---|
| **Terminology** | Generics | | Biosimilar products or Follow-on biologics | |
| | U.S. | EU | U.S. | EU |
| **Regulatory framework** | 21 CFR Part 314 Subpart C. | Article 10 (1) of Directive 2001/83/EC | TBD | Annex I to the Directive 2001/83/EC and 2004/27/EC |
| **Application and review** | ANDA in CTD format<br>FDA moving to QbR by 1/07 | CTD format | TBD | CTD<br>Agency consultation by "scientific advice request" strongly recommended |
| **Nonclinical studies required** | No | No | TBD | Yes |
| **Clinical studies required** | No | No | TBD | Yes |

[1] Including some small proteins/peptides in the United States (e.g., insulin, human growth hormone).

*Abbreviations*: ANDA = Abbreviated new drug application; CFR = Code of Federal Regulations; CTD = Common technical document; EC = European Commission; EU = European Union; FDA = Food and Drug Administration; QbR = question-based review; TBD = to be determined; U.S. = United States.

anticipates a more or less full development program for biosimilar products or follow-on biologics, with material from the sponsor's manufacturing process compared with the innovator reference product.

Importantly, for registration in countries of the European Union, the sponsor of a follow-on biologic has to develop a process generating a protein sufficiently similar to the reference product without access to the typically trade-secret-protected innovator's manufacturing process or the manufacturing cell banks. In addition, a significant nonclinical safety and clinical program has to be conducted.

No biosimilar products/follow-on biologics have been approved in the United States, European Union, or Japan to date. Products confusingly referred to as "generic interferon alpha-2b" or "generic human growth hormone" available in some countries in Asia, Eastern Europe, and Mexico were registered after completion of clinical studies and not through the process used for small-molecule generics. In January 2006, the European Medicines Agency (EMEA) issued a positive opinion for the Novartis subsidiary, Sandoz GmbH (Holzkirchen, Germany), application for the recombinant human growth hormone Omnitrope (somatropin) as biosimilar to the already marketed reference product Genotropin.

## 12 SUMMARY

Biologics-based medicines have become an increasingly important part of the armamentarium available for physicians to therapeutically manage disease. Many promising new entities are in various stages of discovery and development. Although some conceptual differences between small-molecule and macromolecule drug discovery have decreased over recent years, the fundamental overall structural differences dictate unique approaches for each.

Previously, macromolecules were used with little or no optimization, whereas small-molecule and macromolecule discovery today are both geared toward developing the optimal molecule that meets the precise needs of the underlying biology. In mimicking natural processes, biologics, and especially antibodies, can be optimized in all binding characteristics without limitation. The overall discovery/development process can be divided into four convenient steps:

### 12.1 Step 1: Target Selection and Validation

Successful drug discovery generally starts with the definition of a well-defined and biologically understood target. Importantly, the role of the target in disease needs to be demonstrated with unequivocal target validation studies and the role in normal physiology and existence of variants or closely related family members needs to be investigated. Drug discovery today often fails because targets with uncertain biology are being pushed into drug discovery too early.

### 12.2 Step 2: Choice Between Small-molecule and Macromolecule Platforms

At this phase of the discovery process, the biology and location of the target, the disease, and the commercial goal should drive the decision of whether a small-

molecule or biologics discovery approach is optimal. Overall, the management of a mixed portfolio using both drug discovery platforms needs to be considered in light of managing a risk-balanced portfolio. Biologics benefit from a more rapid discovery process and a higher probability of success, with one factor being the lack of nonmechanistic toxicity, often leading to first-in-class commercial opportunities.

## 12.3 Step 3: Definition of Target Product Profile

Once the target is defined and the technology platform selected, the TPP needs to be defined. The TPP is an agreement between all stakeholders in drug discovery, development, manufacturing, commercial, and others on the specific goal of the project considering all phases of development. The TPP lays down in detail the expected characteristics of the development candidate considering the needs of a formal preclinical program. As an example, if the biologics discovery project does not consider cross-reactivity with an orthologue in the planned toxicology as important, the final candidate may bind to the human target only and face huge hurdles in development.

## 12.4 Step 4: Drug Discovery and Development

With key decisions from Steps 1–3 complete, drug discovery using either platform can proceed. It is critical that target validation work continues throughout this period. The TPP planning discussed under Step 3 helps to anticipate reagents needed much later, but it may be time consuming to prepare (e.g., transgenic mice with human target proteins for pharmacology studies).

In contrast to the previous decade, biologics-based medicines can be finely tuned to meet challenging affinity, potency, or specificity goals, and large advances have been made toward dialing in appropriate half-life and effector functions. Antibody affinity optimization, for example, if attempted at all, may have taken a year or more previously, but it can be accomplished in 3 months today.

The predominant message within this introductory chapter is held in the concept that the drug discovery process involves a rational path of critical steps that a pharmaceutical scientist must think through (*see* Figure 1). Decisions are made early in a program, often in the conceptual stage, that are forward thinking and with an eye toward the clinical setting and meeting crucial medical needs in the general population. Although these steps appear identical when comparing macromolecular biomolecules and small-molecule paths, many key differences exist. The stepwise nomenclature tends to be consistent, and the two processes can be compared as waves in which there are distinct nodes of intersection yet many distinct divergences between phases. Many of these divergences involve small steps that can be technology driven. For example, the exploratory phases hold similar target identification and validation steps, whereas the subsequent lead identification step may greatly differ in assay development strategies and high-throughput screening with hit-to-lead follow-up (small molecule) versus surrogate selection and characterization (macromolecule). Surrogate antibody characterization (a lead identification function) can often run simultaneously with human antibody generation (a lead optimization function). Even at this early stage, a macromolecule most

likely will be subject to *in vivo* experiments early in the lead identification phase, whereas it may take many months of even preliminary property optimization of a small molecule to probe *in vivo* pharmacology, especially for intracellular targets. As emphasized early in this overview, a great divergence is the one-dimensional versus multidimensional lead optimization dependencies when comparing small molecules and macromolecules and that much discovery speed can be gained given the distinct macromolecule domains. On the side of the drug development process, developing a large or small molecule that satisfies the TPP parameters is similar, and this can be extended to its measurement and performance in clinical trials. Not only do manufacturing processes diverge greatly, but also the ability to make small or incremental process improvements without affecting the composition and integrity of the bulk pharmaceutical product. In a small-molecule manufacturing process, a solvent or reagent can potentially be substituted to improve yield but must have no effect on the overall final product impurity profile, whereas the addition of an alternative feedstock or genetic optimization of a cell culture line can drastically alter a macromolecule's characteristics. In summary, a great appreciation can be gained for the strategic aspects of advancing small-molecule and macromolecular entities and how each step of a rational process is interdependent on the next to efficiently and successfully advance a therapeutic concept to the pharmacy shelves. Considering all aspects of small-molecule and biologics research and development, including the differences in attrition and speed [1], optimal integration of both approaches into the research and development portfolio may be the best strategy for the future.

## ACKNOWLEDGMENT

We thank our colleage, Dr. Hans-Jürgen Krause, for very helpful contributions to and discussion of the Formulation and Drug Device sections and Lori Lush, PharmD, for helpful assistance with the manuscript.

## REFERENCES

1. Kola I, Landis J (2004). Can the pharmaceutical industry reduce attrition rates? *Nat. Rev. Drug Discov.* 3:711–715.
2. Kubinyi H (2003). Drug research: myths, hype and reality. *Nat. Rev. Drug Discov.* 2:665–668.
3. Salfeld J G (2004). Use of a new biotechnology to design rational drugs against newly defined targets. *Best Pract. Res. Clin. Rheumatol.* 18:81–95.
4. Glennie M J, van de Winkel J G (2003). Renaissance of cancer therapeutic antibodies. *Drug Discov. Today* 8:503–510.
5. Mayer T U (2003). Chemical genetics: Tailoring tools for cell biology. *Trends Cell. Biol.* 13:270–277.
6. Stockwell B R (2004). Exploring biology with small organic molecules. *Nature.* 432:846–854.
7. Zambrowicz B P, Sanda A T (2003). Knockouts model the 100 best-selling drugs—Will they model the next 100? *Nat. Rev. Drug Discov.* 2:38–51.

8. Bagowski C P (2005). Target validation: An important early step in the development of novel biopharmaceuticals in the post-genomic era. In: Knäblein J (ed.), *Modern Biopharmaceuticals: Design, Development, and Optimization*. Vol. 2. Wiley-VCH, Weinheim, Germany. pp. 621–647.
9. Butcher S P (2003). Target discovery and validation in the post-genomic era. *Neurochem. Res*. 28:367–371.
10. Chen B X, Erlanger B F (2002). Intracellular delivery of monoclonal antibodies. *Immunol. Lett*. 84:63–67.
11. Hong G, Chappey O, Neil E, et al. (2000). Enhanced cellular update and transport of polyclonal immunoglobulin G and fab after their cationization. *J. Drug Target*. 8:67–77.
12. Kabouridis P S, Hasan M, Newson J, et al. (2002). Inhibition of NF-kappa B activity by a membrane-transducting mutant of 1 kappa B alpha. *J. Immunol*. 169:2587–2593.
13. Mie M, Takahashi F, Funabashi H, et al. (2003). Intracellular delivery of antibodies using TAT fusion protein A. *Biochem. Biophys. Res. Commun*. 310:730–734.
14. Temsamani J, Vidal P (2004). The use of cell-penetrating peptides for drug delivery. *Drug Discov. Today* 9:1012–1019.
15. Zhao Y, Lou D, Burkett J, et al. (2001). Chemical engineering of cell penetrating antibodies. *J. Immunol. Methods* 254:137–145.
16. Marks J D (2004). Selection of internalizing antibodies for drug delivery. *Methods Mol. Biol.* 248:201–208.
17. Prabbu R, Khalap N, Burioni R, et al. (2004). Inhibition of hepatitis C virus nonstructural protein, helicase activity, and viral replication by a recombinant human antibody clone. *Am. J. Pathol*. 165:1163–1173.
18. Goetzl E J (2004). An IgM-kappa rat monoclonal antibody specific for the type 1 sphingosine 1-phosphate G protein-coupled receptor with antagonist and agonist activities. *Immunol. Lett*. 93:63–69.
19. Bardelli C, Sala M, Cavallazzi U, et al. (2005). Agonist Met antibodies define the signalling threshold required for a full mitogenic and invasive program of Kaposi's Sarcoma cells. *Biochem. Biophys. Res. Commun*. 334:1172–1179.
20. Fry D C, Vassiley L T (2005). Targeting protein-protein interactions for cancer therapy. *J. Mol. Med*. 83:955–963.
21. Hinton P R, Xiong J M, Johlfs M G, et al. (2006). An engineered human IgG1 antibody with longer serum half-life. *J. Immunol*. 176:346–356.
22. Lencer W I, Blumberg R S (2005). A passionate kiss, then run: Exocytosis and recycling of IgG by FcRn. *Trends Cell. Biol*. 15:5–9.
23. Vaccaro C, Zhou J, Ober R J, et al. (2005). Engineering the Fc region of immunoglobulin G to modulate in vivo antibody levels. *Nat. Biotechnol*. 23:1283–1288.
24. Archer J R (2004). History, evolution, and trends in compound management for high throughput screening. *Assay Drug Dev. Technol*. 2:675–681.
25. Bajorath J (2002). Integration of virtual and high-throughput screening. *Nat. Rev. Drug Discov*. 1:882–894.
26. Kariv I, Rourick R A, Kassel D B, et al. (2002). Improvement of "hit-to-lead" optimization by integration of in vitro HTS experimental models for early determination of pharmacokinetic properties. *Comb. Chem. High Throughput Screen*. 5:459–472.
27. Lo B K C (2004). Antibody humanization by CDR grafting. *Methods Mol. Biol*. 248:135–159.
28. Makara G M, Athanasopoulos J (2005). Improving success rates for lead generation using affinity binding techniques. *Curr. Opin. Biotechnol*. 16:1–8.

29. Comess K M, Schurdak M E (2004). Affinity-based screening techniques for enhancing lead discovery. *Curr. Opin. Drug Discov. Dev.* 7:411–416.
30. Roguska M A, Pedersen J T, Henry A H, et al. (1996). A comparison of two murine monoclonal antibodies humanized by CDR-grafting and variable domain resurfacing. *Protein Eng.* 9:895–904.
31. Zhang W, Feng J, Li Y, et al. (2005). Humanization of an anti-human TNF-α antibody by variable region resurfacing with the aid of molecular modeling. *Mol. Immunol.* 42:1445–1451.
32. Hoogenboom H R (2005). Selecting and screening recombinant antibody libraries. *Nat. Biotechnol.* 23:1105–1116.
33. Marks J D, Bradbury A (2004a). Selection of human antibodies from phage display libraries. *Methods Mol. Biol.* 248:161–176.
34. Keri G, Szekelyhidi Z, Banhegyi P, et al. (2005). Drug discovery in the kinase inhibitory field using the Nested Chemical Library technology. *Assay Drug Devel. Technol.* 3:543–551.
35. Miller J L (2006). Recent developments in focused library design: Targeting gene-families. *Curr. Topics Med. Chem.* 6:19–29.
36. Maeda F, Takekoshi M, Nagatsuka Y, et al. (2005). Production and characterization of recombinant human anti-HBs Fab antibodies. *J. Virol. Methods* 127:141–147.
37. Mitchell T, Cherry M (2005). Fragment-based drug design: Fragments offer the prospect of a more efficient approach to drug discovery—resulting in the generation of high-quality leads with better chance of success in clinical development. *Innov. Pharm. Technol.* May:34–36.
38. Verdonk M L, Hartshorn M J (2004). Structure-guided fragment screening for lead discovery. *Curr. Opin. Drug Discov. Devel.* 7:404–410.
39. Petros A M, Dinges J, Augeri D J, et al. (2005). Discovery of a potent inhibitor of the antiapoptotic protein Bcl-$x_L$ from NMR and parallel synthesis. *J. Med. Chem. Web.* Release Date: December 20, 2005.
40. Sillerud L O, Larson R S (2005). Nuclear magnetic resonance-based screening methods for drug discovery. *Methods Mol. Biol.* 316:227–289.
41. Vollmers H P, Braendlein S (2005). The early birds: Natural IgM antibodies and immune surveillance. *Histol. Histopathol.* 20:927–937.
42. Alvarez J, Shoichet B, eds. (2005). *Virtual Screening in Drug Discovery.* Taylor & Francis; Boca Raton, Florida.
43. Anderson A C, Wright D L (2005). The design and docking of virtual compound libraries to structures of drug targets. *Current Computer-Aided Drug Design* 1:103–127.
44. Lonberg N (2005). Human antibodies from transgenic animals. *Nat. Biotechnol.* 23: 1117–1125.
45. Guener O F (2005). The impact of pharmacophore modeling in drug design. *IDrugs* 8:567–572.
46. Lill M A, Dobler M, Vedani A (2005). Multi-dimensional QSAR in drug discovery: Probing ligand alignment and induced fit—application to GPCRs and nuclear receptors. *Curr. Comp. Aided Drug Design* 1:307–324.
47. Lang P T, Kuntz I D, Maggiora G M, et al. (2005). Evaluating the high-throughput screening computations. *J. Biomol. Screen.* 10:649–652.
48. Jain A N (2003). Surflex: Fully automatic flexible molecular docking using a molecular similarity-based search engine. *J. Med. Chem.* 46:499–511.
49. Pirard B (2005). Knowledge-driven lead discovery. *Mini Rev. Med. Chem.* 5:1045–1052.

50. Schuffenhauer A, Popov M, Schopfer U, et al. (2004). Molecular diversity management strategies for building and enhancement of diverse and focused lead discovery compound screening collections. *Comb. Chem. High Throughput Screen*. 7:771–781.
51. Newman D J, Cragg G M, Snader K M (2003). Natural products as sources of new drugs over the period 1981–2002. *J. Nat. Prod*. 66:1022–1037.
52. Kubinyi H, Müller G, eds. (2004). *Chemogenomics in Drug Discovery: A Medicinal Chemistry Perspective*. Vol. 22. Wiley-VCH, Weinheim, Germany.
53. Oprea T I (2002). Current trends in lead discovery: Are we looking for the appropriate properties? *J. Comput. Aided Mol. Des*. 16:325–334.
54. Davis A M, Keeling D J, Steele J, et al. (2005). Components of successful lead generation. *Curr. Top Med. Chem*. 5:421–439.
55. Deprez-Poulain R, Deprez B (2004). Facts, figures and trends in lead generation. *Curr. Top. Med. Chem*. 4:569–580.
56. Olah M M, Bologa C G, Oprea T I (2004). Strategies for compound selection. *Curr. Drug Discov. Technol*. 1:211–220.
57. Alanine A, Nettekoven M, Roberts E, et al. (2003). Lead generation—enhancing the success of drug discovery by investing in the hit to lead process. *Comb. Chem. High. Throughput Screen*. 6:51–66.
58. Van De Waterbeemd H, Gifford E (2003). ADMET in silico modeling towards prediction paradise? *Nat. Rev. Drug Discov*. 2:192–204.
59. Roberts S A (2003). Drug metabolism and pharmacokinetics in drug discovery. *Curr. Opin. Drug Discov. Devel*. 6:66–80.
60. Yu H, Adedoyin A (2003). ADME-Tox in drug discovery: Integration of experimental and computational technologies. *Drug Discov. Today* 8:852–861.
61. Ghetie V, Ward E S (2000). Multiple roles for the major histocompatibility complex class I-related receptor FcRn. *Ann. Rev. Immunol*. 18:739–766.
62. Presta L G (2005). Selection, design, and engineering of therapeutic antibodies. *J. Allergy Clin. Immunol*. 116:731–736.
63. Pandley J, Astemborski J, Thomas D L (2004). Epistatic effects of immunoglobulin GM and KM allotypes on outcome of infection with hepatitis C virus. *J. Virol*. 78:4561–4565.
64. Williams R C Jr, Malone C C (1992). Heteroclitic polyclonal and monoclonal anti-Gm(a) and anti-Gm(g) human rheumatoid factors react with epitopes induced in Gm(a), Gm(g-) IgG by interaction with antigen or by nonspecific aggregation. A possible mechanism for the in vivo generation of rheumatoid factors. *J. Immunol*. 149:1817–1824.
65. Balint R, Larrick J W (1993). Antibody engineering by parsimonious mutagenesis. *Gene* 137:109–118.
66. Chlewicki L K, Holler P D, Monti B C, et al. (2005). High-affinity, peptide-specific T cell receptors can be generated by mutations in CDR1, CDR2 or CDR3. *J. Mol. Biol*. 346:223–239.
67. Feldhaus M J, Siegel R W, Opresko L K, et al. (2003). Flow-cytometric isolation of human antibodies from a nonimmune Saccharomyces cerevisiae surface display library. *Nat. Biotechnol*. 21(2):163–170.
68. Schaffitzel C, Zahnd C, Amstutz P, et al. (2005). In vitro selection and evolution of protein-ligand interactions by ribosome display. In: Golemis E A, Adams P D (eds.), *Protein-Protein Interactions, 2nd ed*. pp. 517–548.
69. Cavagnaro J A (2002). Preclinical safety evaluation of biotechnology-derived pharmaceuticals. *Nat. Rev. Drug Discov*. 1:469–475.

70. ICH Steering Committee. Safety pharmacology studies for human pharmaceuticals S7A. Available: http://www.ich.org/LOB/media/MEDIA504.pdf.
71. Gollapudi B, Krishna G (2000). Practical aspects of mutagenicity testing strategy: An industrial perspective. *Mutat. Res.* 455:21–28.
72. Krishna G, Urda G, Theiss J (1998). Principles and practices of integrating genotoxicity evaluation into routine toxicology studies: A pharmaceutical industry perspective. *Environ. Mol. Mutagen.* 32:115–120.
73. Brennan F R, Shaw L, Wing M G, et al. (2004). Preclinical safety testing of biotechnology-derived pharmaceuticals. *Mol. Biotechnol.* 27:59–74.
74. Green J D, Black L E (2000). Status of preclinical safety assessment for immunomodulatory biopharmaceuticals. *Hum. Exp. Toxicol.* 19:208–212.
75. Pilling A M (1999). The role of the toxicologic pathologist in the preclinical safety evaluation of biotechnology-derived pharmaceuticals. *Toxicol. Pathol.* 27:678–688.
76. Food and Drug Administration, Center of Biologics Evaluation and Research (1997). Points to consider in the manufacture and testing of monoclonal antibody products for human use. Fed Reg 62:9196.
77. ICH Steering Committee. Preclinical safety evaluation of biotechnology-derived pharmaceuticals S6. Available: http://www.ich.org/LOB/media/MEDIA503.pdf. Accessed January 3, 2006.
78. Federsel H J (2000). Building bridges from process R&D: From a customer-supplier relationship to full partnership. *Pharm. Sci. Technol. Today* 3:265–272.
79. Federsel H J (2003). Logistics of process R&D: Transforming laboratory methods to manufacturing scale. *Nat. Rev.* 2:654–664.
80. Dugger R W, Ragan J A, Brown Ripin D H (2005). Survey of GMP bulk reactions run in a research facility between 1985 and 2002. *Org. Process Res. Dev.* 9:253–258.
81. Andraos J (2005). Unification of reaction metrics for green chemistry II: Evaluation of named organic reactions and application to reaction discovery. *Org. Process Res. Dev.* 9:404–431.
82. Gronnow M J, White R J, Clark J H, et al. (2005). Energy efficiency in chemical reactions: A comparative study of different reaction techniques. *Org. Process. Res. Dev.* 9:516–518.
83. Lehmbeck J, Wahlbom F. Expression systems comprising heterokaryon fungus host cell and nucleic acids encoding heterologous signal peptide and cellulose-binding domain for industrial scale production of human antibodies. PCT Int. Appl. WO 2005070962 A1. Filing date: 20 January 2005.
84. Humphreys D P, Glover D J (2001). Therapeutic antibody production technologies: Molecules, applications, expression and purification. *Curr. Opin. Drug Discov. Devel.* 4:172–185.
85. Sanna P P (2002). Expression of antibody Fab fragments and whole immunoglobulin in mammalian cells. *Methods Mol. Biol.* 178:389–395.
86. Houdebine L M (2002). Antibody manufacture in transgenic animals and comparisons with other systems. *Curr. Opin. Biotechnol.* 13:625–629.
87. Schillberg S, Twyman R M, Fischer R (2005). Opportunities for recombinant antigen and antibody expression in transgenic plants-technology. *Vaccine* 23:1764–1769.
88. Yoo E M, Chintalacharuvu K R, Penichet M L, et al. (2002). Myeloma expression systems. *J. Immunol. Methods* 261:1–20.
89. Miescher S, Zahan-Zabal M, DeJesus M, et al. (2000). CHO expression of a novel human recombinant IgG1 anti-RhD antibody isolated by phage display. *Br. J. Haematol.* 111:157–166.

90. Fahrner F L, Knudsen J L, Basey C D, et al. (2001). Industrial purification of pharmaceutical antibodies: Development, operation, and validation of chromatography processes. *Biotechnol. Genet. Eng. Rev.* 18:301–327.
91. Lee M H, Kwak J W (2003). Expression and functional reconstitution of a recombinant antibody (Fab') specific for human apolipoprotein B-100. *J. Biotechnol.* 101:189–198.
92. *Berlex Laboratories Inc., Plaintiff, v. Food And Drug Administration, et al., Defendants.* Civil Action No. 96-0971 (JR), United States District Court for the District of Columbia, 942 F. Supp. 19, 1996 U.S. Dist. LEXIS 15169. Available: http://www.law.berkeley.edu/institutes/bclt/courses/fall99/berlex.html. Accessed March 15, 2006.
93. Henry C (1996). FDA, reform, and the well-characterized biologic. *Anal. Chem.* 68:674A–677A.
94. Lubarskaya Y, Houde D, Woodard J, et al. (2005). Analysis of recombinant monoclonal antibody isoforms by electrospray ionization mass spectrometry as a strategy for streamlining characterization of recombinant monoclonal antibody charge heterogeneity. *Anal. Biochem.* Available: sciencedirect.com.
95. Herrera S (2004). Biogenerics stand-off. *Nat. Biotechnol.* 22:1343–1346.
96. Farnley S, Morey-Nase P, Sternfeld D (2004). Biotechnology—a challenge to the patent system. *Curr. Opin. in Biotechnol.* 15:254–257.

# INTEGRATED DEVELOPMENT OF GLYCOBIOLOGICS: FROM DISCOVERY TO APPLICATIONS IN THE DESIGN OF NANOPARTICULAR DRUG DELIVERY SYSTEMS

Christine Vauthier, Isabelle Bertholon, and Denis Labarre
*University of Paris XI, Chatenay-Malabry, France*

**Chapter Contents**

1 Introduction
2 Carbohydrate Research: Main Steps of Development and Discovery of Their Functionalities
3 Development of Nanoparticular Drug Carriers: Current States and Limitations
4 Integration of Carbohydrate in the Design of Nanoparticle Drug Carriers With New Functionalities
  4.1 Carbohydrates to Improve Mucoadhesion of Nanoparticle Drug Carriers
  4.2 Use of Carbohydrates to Control the Fate of Drug Carrier After Intravenous Administration
  4.3 Carbohydrates and Drug Carrier Targeting Perspectives
5 Enhancement of Carbohydrate Therapeutic Activity by Association With Nanoparticle Delivery Systems
  5.1 Anti-infectious Agents
  5.2 Anticancer Activity
  5.3 Control of Vascular Smooth Muscle Cell Proliferation
  5.4 Activity on the Immune System
  5.5 Control of Blood Coagulation and Complement Activation
6 Nanoparticle Drug Carriers Made of Carbohydrates for the Delivery of Fragile Molecules
7 Conclusion
References

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

## 1 INTRODUCTION

From the three major classes of biomolecules including proteins, nucleic acids, and carbohydrates, the carbohydrates entering the constitution of glycans and polysaccharides are, at the moment, the least exploited in biology. The recent development of operating methods for the analysis of carbohydrates and for the synthesis of oligosaccharides support the efforts made to investigate the roles of carbohydrates in biological phenomena [1–4]. Different already identified functions of carbohydrates may be interesting to be integrated within drug delivery systems conferring new properties and opening avenues for innovative developments. In this view, carbohydrates may be used either as a drug due to their own biological activities or as a molecule to design nanoparticles with new functionalities [5–8]. Developing drug delivery systems, the main challenge is to design systems able to carry and target a drug to the diseased organs or cells after administration in the body by the appropriate route [6, 7, 9, 10]. This assumes that the biodistribution of the carrier system is perfectly controlled after *in vivo* administration. In the case of a soluble carrier, this can be achieved by grafting define types of carbohydrate ligands on the polymer backbone [10, 11]. Using colloidal particles like liposomes and polymer nanoparticles, it is now well accepted that their biodistribution is mainly controlled by their surface properties [9, 12–15]. Several carbohydrates and polysaccharides were identified at the surface of cells, bacteria, and viruses as having a key role in biological recognition and signaling events. Therefore, the potential of their use as coating material or as ligands to promote specific recognition between carrier and target cells was carefully considered as a possible way to develop biomimetic drug delivery systems having different functionalities [16–24]. In this chapter, we will summarize the main stages of the development of carbohydrates and polysaccharides as components of interest to be used either as a targeted drug or as a component entering the design of drug delivery systems. The first part of the chapter will focus on the principal milestones that allowed discovery of most of the carbohydrates and polysaccharides known today in terms of their chemical analysis and biological activity investigations. This part will also present status about the chemical synthesis of these compounds. The second part of the chapter will briefly discuss the current state and limitations to the developments of nanoparticle drug carriers. Integration of carbohydrate and polysaccharide developments in the design of drug delivery systems with new functionalities will constitute the third part of the chapter. The fourth and fifth parts will, respectively, review the current research about applications of nanoparticles as delivery systems for biologically active carbohydrates and the use of nanoparticle drug carriers made of polysaccharides for the delivery of fragile bioactive molecules.

## 2 CARBOHYDRATE RESEARCH: MAIN STEPS OF DEVELOPMENT AND DISCOVERY OF THEIR FUNCTIONALITIES

Forty years ago, Eylar [25] suggested for the first time that the carbohydrate units found in biologically active glycoproteins could act as a kind of “chemical label” or “chemical passport” to promote specific interactions of glycoproteins with corresponding cell-surface receptors to achieve their transportation across the cell

membrane. However, progress in the understanding of the biological role of glycan moieties included in biological structures was hampered for a long time, because glycans and polysaccharides were much more complex structures to study than were proteins and nucleic acids. Thus, it was only 30 years ago, while full chemical synthesis of peptides was already possible and that of nucleic acid was in its infant age, that various carbohydrate containing compounds came into light as important elements in the number of biological activities [26–30]. For instance, the years 1976 to 1983 were critical in the discovery of the antithrombine binding site of heparin, which now can be used to confer anticlotting and anticomplement activation properties to drug carrier systems [30–32]. The study of lectines isolated from animals that are now known to be involved in many biological recognition molecules involving carbohydrates started to grow rapidly in the 1980s thanks to the development of recombination techniques to produce them [28]. Identification of the role of these molecules in recognition phenomena as well as understanding their functionality are interesting to be able to promote specific recognition between target cells and drug delivery systems achieving drug targeting. It is also noteworthy to point out the discovery of carbohydrates and carbohydrate derivatives that are currently used as major anticancer drugs and antibiotics in clinics [33–35]. Because of the high toxicity of these compounds for the healthy tissues, they are good drug candidates to be incorporated in nanoparticle drug carrier systems improving their specific biodistribution.

For long, progresses in glycobiology were hampered due to technological difficulties in analytical order. The appearance of the high-performance anion–exchange chromatography (HPAEC) in the mid-1990s revolutionized the analytical capacity and considerably improved the accuracy of determining the carbohydrate sequences and arrangements in glycans isolated from glycosylated biological compounds [2, 4]. This was a significant milestone that boosted the resolution of glycan structures and in parallel opened new areas to further investigate their biological activity. The analytical techniques were constantly improved by the subsequent introduction of pulsed amperometry detection (PAD) and more recently by online coupling the HPAEC to mass spectrometry devices and nuclear magnetic resonance (NMR) methods using a nano-probe [3, 4, 36–38]. New methodologies continue to be developed improving performance of the existing analytical methods [39]. The introduction of screening methods using microarray and chip technologies appeared within the last 3 years accelerating discovery and progress toward the understanding of the role of carbohydrates in biology [29, 40, 41].

Research into the biological roles of carbohydrates may be possible because sufficient quantities may be isolated from natural extract. Because of their structural complexity, it is extremely difficult to isolate pure carbohydrates from natural sources. Progress in the analytical chromatographic methods has helped to improve preparative techniques providing purified fractions of defined glycans from various natural extracts [36, 42]. However, the only way to access very pure carbohydrates relies on their chemical or enzymatic synthesis, which also constitutes the more attractive route for the production of the sufficient quantities for biological studies designed to identify biological roles and eventually medicinal interests. The synthesis of polysaccharides and glycans was an important technological challenge. The intensive efforts based on innovative strategies led to major advances, making the synthesis of complex carbohydrates now possible [1, 43–45]. Stereoregular

polysaccharides, including linear, branched, amino, and deoxy polysaccharides, can be synthesized by ring opening polymerization of anhydro-sugar derivatives [1]. The technically extremely difficult synthesis of highly branched carbohydrates was made possible with the development of solid-phase synthesis strategies [44]. Another important milestone was completed thanks to the recent development of an automated oligosaccharide synthesizer leading to accelerated access to many highly branched carbohydrates for identification of their biological function [44]. Even if this methodology provides just enough glycan to study the biological role of the product, this really opens avenues for the identification and creation of important compounds for future biochemical and medicinal applications [44]. For instance, after years of research, many synthetic heparins proposed as anticoagulant drugs can be synthesized on a production scale. They are currently at different stages of clinical development [46]. The chemistry remains complicated, and some synthetic routes that have been scaled up comprised up to 65 steps.

## 3 DEVELOPMENT OF NANOPARTICULAR DRUG CARRIERS: CURRENT STATES AND LIMITATIONS

Despite several advances in chemotherapy, the real therapy of cancer and major infections remains a challenge. One problem comes from the nonspecific drug distribution resulting in low tumor or infected tissue concentrations and systemic toxicity. Another problem comes from the lack of stability of drugs in biological media and their incapacity to cross biological barriers. To improve drug concentration in the targeted tissues, several approaches have been suggested, including the association of the drug to carrier systems. A large piece of work has been done using liposomes as carriers, which has led to several marketed formulations so far. For instance, the liposomal formulation of Amphotericin B, which is the leading compound in the treatment of leishmaniasis is today the more efficient treatment against this parasite and other fungal infections [47]. The toxicity of Amphotericin B was reduced by a factor of 50- to 70-fold, which is among the most obvious benefit for the use of this liposome formulation in clinics [48]. The main drawback of liposomes remains the cost of the phospholipids entering their composition. To reduce cost, the biggest challenge is believed to come from polymer nanoparticles. Indeed polymers are a much cheaper material compared with phospholipids, and methods for producing nanoparticles are generally simpler. Other advantages of the polymer systems are their higher stability in biological media especially for the administration of drugs by mucosal routes and upon slight changes in the formulation. Additionally, they should be theoretically tailor-made thanks to the progress in colloid and polymer technology [9, 49]. However, only a couple of polymers are suitable as main components of the polymer nanoparticles designed as drug carrier systems [6, 7, 9]. As for liposomes, polymer nanoparticles were developed as drug carriers to improve the efficacy of the associated drug thanks to a better control of its biodistribution toward the diseased organ or cells and to their faculty to overcome biological barriers [9, 50, 51]. These colloidal particles have demonstrated enhanced efficacy for numerous drugs compared with conventional formulations. Indeed, the biological activity of the drug associated with the carrier is generally improved, whereas the side effects due to their toxicity are reduced thanks to a better control of the biodistribution [52–57]. The improved efficacy can also be explained by a

higher stability of the drug in biological media and ability to overcome biological barriers. However, the main limitation of these systems remains the full control of their biodistribution because they still lack precise targeting specificities. This is the main challenge that remains to be addressed. According to our current knowledge, a better control of the interactions between the drug delivery system and surrounded biological media should improve targeting efficacy as well as promote the occurrence of highly specific interactions. Thus, it was postulated that the control of the biodistribution of drug carriers could be possible by modifying and by adjusting the surface characteristics of the drug delivery systems. The modification of surface properties of drug delivery systems by coating with poly(ethylene glycol) actually rerouted the carrier from a very efficient recognition and uptake by macrophages of the mononuclear phagocyte system (MPS) toward long circulating particles in the blood stream [58, 59]. It is now admitted that the biodistribution of a drug carrier greatly depends on the nonspecific and specific interactions between its exposed surface and the biological components of the surrounded medium found either in blood, on the cell surface, or at the level of biological barriers, i.e., endothelium or mucosa. Thus, efforts are now carried on tailoring the design of the carrier surface [12].

To be used as drug carriers, different drugs should be incorporated into the nanoparticles. Most nanoparticles made of synthetic polymers, including polyesters or polyanhydrides and poly(aminoacids), can incorporate enough lipophilic drugs. However, they are totally inefficient to incorporate hydrophilic molecules such as peptides, proteins, or nucleic acids. These molecules with a high therapeutic potential hardly cross biological barriers. They need to be associated with a drug carrier enhancing their transport across the biological barriers and preserving their biological activity because they are highly un-stable in biological media. This problem could be solved using nanopar-ticles made of poly(alkylcyanoacrylate) [60]. Polysaccharides extracted from natural compound could also be used to formulate new types of drug delivery systems in which macromolecules from biotechnology could be incorporated [6–8, 61, 62].

## 4 INTEGRATION OF CARBOHYDRATE IN THE DESIGN OF NANOPARTICLE DRUG CARRIERS WITH NEW FUNCTIONALITIES

As mentioned, the interactions of nanoparticle drug carriers with the surrounded biological environment are key factors controlling the *in vivo* fate of the transported drug. Three levels of interactions may be distinguished, but all will involve surface phenomena between nanoparticle surface and compounds of the surrounding environment. Interactions occurring at the level of a mucosa may be critical to promote the absorption of the drug. They will be discussed in the first paragraph. The interactions taking place in the blood compartment that greatly affect the biodistribution of the carrier due to interactions with blood proteins are considered in the second paragraph. Finally, the more highly specific interactions between drug carrier and cells in the view of targeting and allowing delivery of the drug to precise cells will be considered in the last paragraph of this section. As will be discussed, carbohydrates seem, in all of these cases, as versatile tools to control the interactions between the drug carriers and the biological surrounding media.

### 4.1 Carbohydrates to Improve Mucoadhesion of Nanoparticle Drug Carriers

Mucosal routes, especially the oral route, are the preferred methods for the administration of drugs. However, many drugs remain poorly available when they are administered by these routes. Among other reasons, this can be due to low mucosal permeability for the drug, permeability restricted to a region of the gastrointestinal tract, low or very low solubility of the compound that results in a low dissolution rate in the mucosal fluids and in elimination of a fraction of the drug from the alimentary canal prior to absorption, and lack of stability in the gastrointestinal environment, resulting in a degradation of the compound before its absorption (e.g., peptides and oligonucleotides).

To reduce problems of low permeability and low solubility, it was proposed to retain the drug at the surface of the mucosa using bioadhesive formulations increasing the contact time between the drug and the epithelium and in turn chance to the drug to be absorbed. In this aim several authors have suggested to use polysaccharides to design bioadhesive nanoparticle drug delivery carriers. Chitosan obtained from deacetylation of chitin is the most commonly used bioadhesive polysaccharide. At the level of mucosa, chitosan is known to enhance drug absorption thanks to it bioadhesive properties and to its capacity to induce transient opening of the tight junctions without damaging the cells [63, 64]. The bioadhesive properties of chitosan were described for the first time by Lehr et al. [65]. They depend on the molecular weight [66, 67] and on the pH of the surrounded media, which affect the polysaccharide solubility. Chitosan is soluble at pH below 6, and its bioadhesive properties require that the molecule is either soluble or strongly swollen [68]. Thus, to improve the solubility properties of chitosan especially at pH higher than 6 encountered at the level of several mucosa, modified chitosan were synthesized. For instance, the partial substitution of the amino function of chitosan by methyl-5-pyrrolidone led to derivatives enhancing the absorption of drugs at the level of the bucal and vaginal mucosa [68–70]. Enhancement of the bioadhesive properties and of drug absorption was also reported after introduction of thiol groups in the structure of chitosan. In this case, the thiol groups are assumed to form disulfide bonds with the cystein-rich domains of the mucus glycoproteins standing on the gastrointestinal mucosa [71, 72]. Drug carriers that have been coated with chitosan also showed enhanced bioadhesive properties at the level of different mucosa, including the ocular, nasal, gastrointestinal, and pulmonary mucosa [8, 73–76]. However, it is worth pointing out that a simple adsorption of chitosan at the nanoparticle surface is not stable enough to resist desorption when the nanoparticles are in contact with the mucus [77]. Thus, several authors have suggested preparing nanoparticles directly with chitosan obtaining plain chitosan nanoparticles [7, 75] or using copolymers of chitosan and the polymer constituting the nanoparticle, i.e., poly(lactic acid) [61] or poly(alkylcyanoacrylate) [19, 21, 62, 78].

According to Prego et al. [8], the mucoadhesive properties of the carrier and its interaction with the mucus occurring *in vivo* are determinant to provide an efficient transport of the nanodevice carrying the drug across the epithelium, resulting in an increase of the bioavailability. The mucoadhesive properties were shown to depend on the coating material but also on the size of the nanoparticles [8, 79]. Generally, the smallest nanoparticles accumulate more because of a higher

diffusion in the mucus layer compared with bigger particles. However, according to a recent study, Bertholon et al. showed that the conformation of the chitosan chains grafted at the surface of poly(alkylcyanoacrylate) nanoparticles also influenced the bioadhesive properties of the nanoparticles on the intestinal mucosa [21]. Indeed, these authors reported differences in bioadhesion with nanoparticles onto which chitosan was grafted either on a "side-on" conformation or on the "end-on" conformation. Adhesion of the nanoparticles to the intestinal mucus layer can be viewed as a complex process, including the diffusion of the particles in the hydrogel, depending on their size, followed by their interactions with the mucus glycoproteins depending on both the nature of the polysaccharide and its conformation at the nanoparticle surface. It is likely that the molecular interactions played a key role in this latter process. Interestingly these results suggest that the bioadhesive properties of nanoparticulate systems can be modulated by a fine tuning of the surface properties of the particles.

Several other polysaccharides were used to obtain bioadhesive formulations of drugs [80, 81]. Hyaluronic acid, which is applied in ocular surgery to protect the corneal endothelium and to manipulate intraocular tissue is also a remarkable bioadhesive. It was suggested as coating material at the surface of nanoparticles made of poly(epsilon-caprolactone) for the development of bioadhesive ocular formulation of bioactive molecules to increase the retention of the drug at the level of the precorneal zone [82]. The other polysaccharides found in pharmaceutical formulations that may be interesting to use as bioadhesive coated material at the surface of nanoparticle drug carriers are alginate, dextran, starch, and cellulose derivatives, including carboxymethylcellulose and hydroxypropyl cellulose. These polysaccharides have in common high swelling properties in an aqueous environment.

### 4.2 Use of Carbohydrates to Control the Fate of Drug Carrier After Intravenous Administration

The fate of nanoparticles administered by the intravenous route is controlled by the interactions between the nanoparticle surface and the serum proteins. In the blood, nanoparticles are normally considered foreign bodies and are rapidly recognized as such by the host defense system. The primary event of the recognition phenomena is basically nonspecific and results in a massive adsorption of serum proteins at the nanoparticle surface [14, 58, 83]. Consequently, the complement system is activated, leading to the labeling of the nanoparticle surface with specific activated proteins of the complement systems. Once the nanoparticles are labeled with activated complement proteins, they are efficiently recognized by macrophages of the mononuclear phagocyte system (MPS) due to the presence of the corresponding receptors at the cell surface. The overall process is so efficient that less than 5 minutes are generally necessary to remove from the blood 80% to 90% of the injected dose of the drug carrier. After these events, the drug and its carrier are found in macrophages of the liver and of the spleen [9, 12, 84–87]. This phenomenom is highly beneficial to target drugs in the liver and spleen but hamper the distribution of the drug toward other biological territories.

Targeting territories, organs, and tumors located outside the MPS organs can be achieved by triggering the host defense systems thanks to the modification

of blood–protein interactions with the nanoparticle surface in order to obtain long circulating nanoparticles. The more general approach consists of grafting poly(ethylene glycol) chains at the nanoparticle surface to reduce both opsonization of the carrier surface by blood proteins and complement activation [14, 15, 58, 88]. An alternative route is based on a biomimetic approach using carbohydrates as material to coat the nanoparticle surface in order to modify interactions with blood proteins [19, 20, 31, 32, 89–92]. The design of biomimetic drug delivery systems using carbohydrate as surface coating material is almost infinite as much of the diversity of carbohydrates appears to be in Mother Nature as exposed at the surface of the eukaryotic cells, bacteria, and viruses [5, 40, 44]. So far, only simple polysaccharides were used to modify the surface properties of nanoparticle drug carriers. The simplest polysaccharide was dextran, which is an alpha-1-6-poly(glucose). Measurements of complement activation given by different dextran-coated nanoparticles provided contradictory results depending on the method of the nanoparticle preparation. Indeed, dextran-coated poly(alkylcyanoacrylate) nanoparticles prepared by redox radical polymerization induced only a low complement activation [19, 20, 31, 32], whereas dextran-coated nanoparticles prepared by other methods induced a strong activation of complement, as does sephadex, which is a reticulated dextran [19, 20, 32, 90, 93, 94]. The contradictory complement activation measured could not be correlated to a marked difference in the interaction of the nanoparticles with blood proteins. Indeed, each type of the nanoparticles showed a proper pattern of blood protein adsorption and no general behavior could be drawn [32]. The hypothesis suggested that explains the results of complement activation is that it may depend on the conformation of the dextran chains at the nanoparticle surface, which can be defined by the method of nanoparticle preparation. This hypothesis is supported by the results of a very recent work that showed that nanoparticles with dextran chains grafted at the nanoparticle surface on the "end-on" conformation are nonactivators when the molecular weight of dextran passes above a certain limit (40,000 g/mol), whereas the nanoparticles on which dextran are arranged on a "side-on" conformation activated complement whatever was the molecular weight of dextran [20]. The effect of the polysaccharide chain conformation was more evident with a series of chitosan-coated nanoparticles. Indeed, nanoparticles with chitosan on the "end-on" conformation were frankly nonactivators, whereas the nanoparticles with chitosan on the "side-on" conformation were strong activators [20]. Although the conformation of the polysaccharide chains at the nanoparticle surface seemed to be a key parameter to control complement activation, several authors showed that the presence of dextran at the nanoparticle surface reduced the adsorption of albumin in favor of a reduction of opsonization phenomena [92, 95]. In terms of biodistribution of the corresponding nanoparticles, the nanoparticles having dextran on the "side-on" conformation were recognized by macrophages of MPS and concentrated in the liver and in the spleen [13], whereas the nanoparticles with dextran on the "end-on" conformation remained in the blood circulation for a longer period [89].

Although chitosan-coated nanoparticles showed contradictory results in complement activation depending on the conformation of the chains at the nanoparticle surface [20], the level of complement activation induced by chitosan was also influenced by the physico-chemical properties of the polysaccharide, including its solubility, degree of deacetylation, and molecular weight [20, 96–98]. Using different

chitosan and chitosan derivatives, nanoparticles with a prolonged circulation time in the blood can be obtained. Additionally, their distribution can be modulated in a certain way because they were found to accumulate in defined tissues outside the MPS [99–104].

In Mother Nature, sialic acids play a key role in the biological mask hiding epitopes at the cell surface, which are immediately recognized by macrophages when the sialic acid is removed from the cell surface [103–107]. The most typical example is given by the erythrocytes, which are rapidly destroyed when sialic acids are removed from their surface [106, 108, 109]. Highly virulent bacteria have also adopted sialic acids as coating material to escape host defense mechanisms [110–112]. Sialic acid, which seems to inhibit complement activation, allows cells and particular material to stay in the blood without being recognized by macrophages [110]. Thus, it was suggested to coat nanoparticles with sialic acids to mask their surface and avoid recognition by macrophages. Two different approaches were followed. At first, the ad-sorption of a sialic acid-rich glycoprotein, orosomucoid, at the surface of poly(alkylcyanoacrylate) nanoparticles reduced opsonization of the nanoparticles and complement activation. However, the orosomucoide that was simply adsorbed at the nanoparticle surface was rapidly displaced from the surface by other serum proteins due to the Vroman effect, and the effect of the sialic acids could be shown on short duration [17, 113]. Second, the nanoparticles coated with sialic acid were obtained from a copolymer of poly(sialic acid) and poly(lactic acid). These nanoparticles coated with a polysialic acid remained in the blood circulation longer than the uncoated nanoparticles after intravenous administration to mice [114]. Poly(sialic acids) have the advantage of being biodegradable, and their catabolic products (i.e., neuraminic acids) are not known to be toxic. Another important advantage of poly(sialic acids) is their T-independent antigens, meaning that they do not induce immunological memory. Because of these properties, poly(sialic acids) are very promising molecules to design biomimetic drug carrier systems with an increased half-life in the blood stream of the drug carrier [16, 115]. The main limitation of the development of such drug carriers so far may be explained by the poor availability of poly(sialic acids), which remain a problem because they are produced by highly pathogenic bacteria strengths at a very high cost.

The other bioactive polysaccharide that seemed interesting to be used to produce surface-modified nanoparticles reducing their recognition by the host defense was heparin. Heparin is used as a drug for its anticoagulation properties. Additionally, it is an inhibitor of the complement activation phenomenon [116–118]. It was demonstrated that heparin-coated nanoparticles did not activate the complement system [19, 31, 32] and remained in the blood stream for a longer time compared with nanoparticles, which do not show heparin on the nanoparticle surface [89]. Other polysaccharides extracted from mushrooms were found to inhibit the activation process of the complement. They could be alternative polysaccharides to produce nanoparticles with a reduced capacity to activate the complement, such as heparin [119].

According to the current experiences, several polysaccharides can be used to modulate the interactions of nanoparticle drug carriers, with blood proteins modifying their blood clearance and consequently their biodistribution. Two categories of polysaccharides giving nanoparticles an ability to escape recognition from macrophages of the MPS can be identified. The first category, including heparin and sialic

acids, specifically interact with serum protein according to a well-defined mechanism, thanks to their specific biological activity being inhibitors of the complement activation phenomenom. The polysaccharide of the other category, including dextran, interacts with serum protein through nonspecific interactions. In addition to the specific and nonspecific interactions with blood proteins, it was clearly demonstrated that the conformation of the chains grafted at the nanoparticle surface is one key parameter that controls the complement activation phenomena and the opsonization of the nanoparticle surface. The nature of the polysaccharide and the molecular weight took on the second level of importance [20].

### 4.3 Carbohydrates and Drug Carrier Targeting Perspectives

Colloidal carriers have proved to be interesting to deliver drugs by the intravenous route because of their potential to improve the therapeutic index of the carried drug while reducing their side effects. Although with the current systems available it can be possible to enhance accumulation of the drug in diseased tissue, especially in tumors, the interactions of the carriers with diseased cells and their uptake remain insufficient. The main strategy that was proposed to further enhance drug delivery and retention at the level of diseased cells is based on the active targeting of the drug [12, 120–122]. This supposes that drugs reached target cells thanks to a highly specific biodistribution of the carrier. Our experience with drug carriers suggests that such a degree of specificity will only be obtained if the drug carrier can recognize the target cells through specific interactions.

One way to promote recognition between drug carriers and target cells is to attach ligands at the carrier surface that can bind specifically to target cells [12, 121, 122]. For this purpose, more or less complex carbohydrates can be used as ligands of endogenous receptors for carbohydrates known as lectins at the cell surfaces. Lectines are universally found in the microbial world, on vegetal and animal cells [5, 28, 123]. They can bind various types of carbohydrates from simple monosaccharides to oligo- or polysaccharides of complex structures with a very high specificity of recognition resulting in strong interactions. The binding is a very fast process, and the recognition between the carbohydrate and the corresponding lectin can occur whether the carbohydrate is free or bound on a macromolecule or on the surface of cells and of drug carriers [123, 124]. So far, the majority of the work done in the field of active targeting was performed with liposomes and soluble polymers. Currently, only a few studies were carried out on nanoparticles [12, 121, 125]. Microbial and vegetal lectins were bound to drug carriers as targeting moiety to find complementary carbohydrate epitopes at the surface of target cells [126–128]. However, with the recent progress in glycobiology, it seems that a better option would consist in grafting the carbohydrates on the drug carrier instead of lectins, which often lose activity during the grafting process. Another advantage of using carbohydrates as the targeting moieties is that the drug carrier would be targeted toward cell receptors with expression levels that are subjected to dramatic modifications upon physiological conditions of the cells. For instance, in several cancer cells, some receptors are specifically over-expressed, whereas other are downregulated and almost disappear from the cell surface. Thus, it seems better to use the ligand receptor, i.e., carbohydrate, as the targeted moiety attached on the drug carrier to take advantage of differences in the level of receptor expression

in the targeting strategy. This method, which was applied with folic acid, actually showed high targeting potential [129]. Additionally, it would probably be easier to find universal methods for grafting various types of carbohydrate moiety on nanoparticle surface in comparison with the coupling methods of proteins, which need to be finely adjusted for each protein receptor to preserve the functionality of their active site [44].

Many endogenous lectins in animals might serve as target receptors for drug carriers exhibiting complementary carbohydrates at the surface. Examples of animal lectins and principal characteristics, locations, and specificities are given in Table 1, whereas Table 2 summarizes variations of the level of expression of lectins in several diseases that were already identified and reported in the literature.

This last table may be especially useful for identifing over-expressed lectins on diseased cells and may be used as a target to direct drug delivery carriers, bearing the corresponding carbohydrate ligand specifically toward them. Such a targeting strategy was successful using soluble drug carriers [126, 137, 141, 142], and in the case of targeting DNA/polymer complexes [130, 143–147] bearing simple carbohydrate or polysaccharides. Most of the work done so far with nanoparticles included mono- and di-saccharides as the targeting moiety. Lectins can recognize galactose and mannose residues after chemical grafting at the surface of polymer nanoparticles [148, 149]. Galactose was used to target paclitaxel-loaded nanoparticles toward hepatic cells. The lectins of the hepatic cells specifically recognized the galactose residues at the surface of the nanoparticles, improving the targeting of the drug in these cells. Thus, the nanoparticles were found to be more toxic toward the cells expressing the lectin at the surface compared with cells devoid of the corresponding lectin [147]. It also seems that lactose bearing nanoparticles that are recognized by asialoglycoprotein receptors of hepatic cells of mice are internalized by endocytosis [148]. A carbohydrate of Lewis was suggested as targeting moiety to target poly(lactide-co-glycolide) nanoparticles toward cells expressing E selectins on endothelial cells or in inflammatory sites. It was found that, for this more complex carbohydrate corresponding to an oligosaccharide, the affinity of the E-selectins for the Lewis carbohydrate exposed at the nano-particle surface depended on the density of grafting [121].

According to these data, targeting strategies that involve carbohydrates are principally directed toward cell receptors involved in the mechanisms of cell regulation and communication either between cells or with the surrounding medium. They are based on specific interactions with high affinity between the ligand and the corresponding receptor. So far, they have been applied with simple carbohydrates, including monosaccharides, di-saccharides, and some Lewis carbohydrates. It provided more than satisfactory results regarding the level of targeting specificity as evaluated *in vitro* using relevant cell models. Only a few studies that went over the *in vitro* stage were performed with liposomes and soluble drug carriers, providing extremely convincing perspectives [122, 141, 152]. There is no doubt that similar targeting efficiency may be obtained with success using nanoparticles. It can also be expected that current investigations aiming to identify structures and the biological functionalities of carbohydrates at a large scale will contribute to amplify our knowledge of cell communication and therefore highlight new targets and target carbohydrate-based materials [5, 35, 40, 44, 153]. This promises very exciting

**TABLE 1. Principal Characteristics of Some Animal Lectins**

| Class | Subclass | General Properties | Localization | Specific Ligand | Non-Specific Ligand | Ref. |
|---|---|---|---|---|---|---|
| **C–Lectins Calcium dependent** | Endocytic receptors | Asialoglycoprotein receptors | Hepatocyte | Ending β D galactose<br>Ending N Acetylgalactosamine | | [120, 130] |
| | | Involved in the clearance of desialylated proteins | Liver endothelial cells | Ending N Acetylgalactosamine | | |
| | | | Kuppfer cells | Ending N Acetylgalactosamine | | |
| | | | Macrophages | Mannose | | [131, 132] |
| | | | Parenchymal liver cells | Arabinogalactane<br>Ending Galactose | Dextran | |
| | | | Non-parenchymal liver cells | Pullulan<br>Manan | Dextran<br>Pullulan<br>Arabinogalactan | [133] |
| | Selectins | Transmembrane glycoproteins involved in cell traffic to injury and inflammation sites | Platelets (P selectin) + endothelial cells | Heparin | | [121, 125] |
| | | | Endothelial cells (E selectin) | | | [121, 125, 134] |
| | | | Leukocyte (L selectin) | Sulfated and sialylated ligands<br>Heparin | | [132, 134] |
| | Collectins | Involved in innate immune phenomena | | Saccharides on pathogens or allergens | | [132] |

| | | | | | |
|---|---|---|---|---|---|
| **C–Lectins Calcium dependent** | Proteoglycans | Involved in regulation mechanisms of cell growth, differentiation, and lipid metabolism | *Hyaluronate receptor:*<br>CD44: hematopoietic cells<br>Fibroblast<br>RHAMM: Fibroblast<br>Smooth muscle cells<br>Macrophages<br>Lymphocytes<br>ICMA 1: Leukocytes<br>Macrophages | Hyaluronic acid and sodium salts | [135–138] |
| | | | *Non hyaluronate receptors:* | Fructose, Galactose | [139] |
| **S-Lectins Calcium independent** | Galectins 1 | Cell adhesion and proliferation, regulation of immune infections | Skeletal muscle | B Galactoside<br>Glucose | [132] |
| | Galectins 3 | Allergy, inflammatory process | Eosinophils<br>Neutrophils<br>Macrophages<br>Epithelial cells | β Galactoside<br>Glucose | [132, 140] |

**TABLE 2. Modification of Expression Level of Different Lectins in Several Pathologies**

| Pathologies | Over-Expressed Lectins | Under-Expressed Lectins | Ref. |
|---|---|---|---|
| Cancer | P Selectins in plasma<br>Hyaluronate receptor CD44 and RHAMM<br>Galectins | | [121, 125, 140] |
| Metastasis | E Selectins<br>Hyaluronate receptor CD44 and RHAMM | | [132]<br>[135–138] |
| Colon cancer | Galectins 1 et 3 | Galectin 8 | [126] |
| Meningeal leukaemia | L Selectins | | [121] |
| Malignant B-cells | L Selectins | | [121] |
| Angiogenesis | E Selectins | | [132] |
| Connective tissue disease | P Selectins in plasma | | [121, 125] |
| Rheumatoid arthritis | P Selectins in plasma<br>E Selectins | | [121, 125] |
| Asthma, allergy | P Selectins in plasma<br>E Selectins | L Selectins | [121, 125] |
| Multiple sclerosis | L Selectins | | [121] |
| Kawasaki disease | E Selectins | L Selectins | [121] |
| Guillain—Barre syndrome | E Selectins | | [121] |
| Intestine disease | P Selectins in plasma | | [121, 125] |
| Sever infection (malaria) | P Selectins in plasma<br>E Selectins | | [121, 125] |
| Neonatal bacterial infection | | L Selectins | [121] |
| HIV | L Selectins | | [121] |
| Insulin-dependent diabetes | L Selectins | | [121] |
| Risk factors of stroke | | L Selectins | [121] |
| Cardiopulmonary bypass surgery | | L Selectins | [121] |

developments in the future. However, as stressed by Nobs et al., one has to keep in mind that the ability to achieve targeting *in vivo* depends directly on whether the target is accessible from the vascular compartment [12]. In the first case, particles bearing targeting moiety and having long circulating properties have a high probability to reach the target cells. On the contrary, when the target is extravascular, for instance, in solid tumors, penetration of the carrier in the tumor tissue is more difficult because it needs to extravasate. In this case, the size of carriers needs also to be adjusted so that they can cross the vessel endothelial through the pores of the local discontinuous tumor microvasculature, which may vary between 100 and 780 nm. Another requirement would probably concern the orientation of the carbohydrate moiety at the nanoparticle surface. Indeed, biological recognition systems are so sensitive that they are, for example, capable of differentiating small changes in the conformation of epitopes at the surface of pathogens.

## 5 ENHANCEMENT OF CARBOHYDRATE THERAPEUTIC ACTIVITY BY ASSOCIATION WITH NANOPARTICLE DELIVERY SYSTEMS

Research to understand biological functions of carbohydrates is currently extremely active providing new data almost everyday. The literature is growing exponentially in this very fast moving area. The purpose of this chapter was not to review all the different biological activity known for carbohydrates today but was to focus on carbohydrates whose biological activity is interesting for therapeutic purposes and whose therapeutic efficacy may be improved by association with a nanoparticle drug carrier.

### 5.1 Anti-Infectious Agents

The emergence of antibiotic resistant strengths of bacteria, parasites, fungus viruses, and new pathogens are very demanding for the discovery of new anti-infectious agents. If the discovery of new molecules is undoubtedly the leading strategy in finding effective treatments, all active molecules will need to reach sites of infections where pathogens are located at sufficient concentration. Several efficient carbohydrate antibiotics were extracted from various strengths of *Streptomyces*. They are simple aminoglycoside molecules known as streptomycins, including neomycin, gentamicin, or amikacin. Other recently discovered carbohydrate antibiotics, including the family of the orthosomycins, are more complex molecules. These carbohydrate antibiotics are very attractive because they are efficient against pathogen bacteria, including those resistant to antibiotics used as a last resort (penicillin or vancomycin). Their interest is so huge that several methods for their synthesis were already suggested [154–156]. However, they are generally very toxic for healthy tissues limiting the maximal dose of antibiotic that can be administered in patients and consequently the efficacy of treatment. In general, the efficacy of antibiotics can be considerably increased when they are associated with drug carriers thanks to a reduction of the toxicity and a higher accumulation in the infected cells that, in many infections, are cells of the mononuclear phagocytes system. The improvement of the safety profile of the targeted drug allowed administration of higher doses of the antibiotic when requested to promote treatment efficacy [48, 157]. In other cases, the therapeutic index of the antibiotic was markedly increased by a factor ranging from 6 to up to 120; thus, the dose of antibiotic required to kill infectious agents can be considerably reduced [158–164]. Complete sterilization of organs can be obtained with the antibiotic targeted with nanoparticles, whereas the corresponding liposomal formulation can not completely sterilized the organs in the same experimental condition [165].

Currently, there is only one formulation of Amikacin in liposomes under clinical trials for treatment of tuberculosis. The antibiotic was found to be twofold to sixfold more active than the free drug in an acute experimental model of murine tuberculosis in which bacteria are located in macrophages. The count of viable bacteria found in the liver and spleen was reduced by a factor of 3-$\log_{10}$ compared with the untreated mice [158]. However, in human, this formulation was not as active as expected. To explain the disappointed results obtained, it was suggested that the liposomes could target the antibiotics in the macrophages but this was not enough to reach extracellular bacilli, which are clustered in cavity caseum in the human

infection and can be responsible for reinfection. It was suggested that these residual bacteria may be reached using drug carriers with long circulating properties.

So far, the more advanced studies on the targeting of antibiotics that reach the clinic stages were performed with liposomes. The only formulation marketed so far, AmBisome (Gilead, Foster City, CA), is the more efficient treatment against leishmaniasis and other fungal infections [47]. This demonstrates the high potential of targeted formulations of antibiotics to challenge devastating pathogens with toxic antibiotics for healthy tissue provided that they are targeted to the infected cells. It may be expected that research will continue considering targeting strategies of the most interesting carbohydrate antibiotics to improve the efficacy of anti-infectious treatments. In this view, it will also be interesting to pay attention to nanoparticle formulations that seemed more efficient than liposomes when they were tested in the same conditions probably because of their higher stability in the blood compartment *in vivo* [157, 161].

Some polysaccharides showed either antibacterial and antifungal activities or antiviral activity. For instance, the antibacterial and antifungal activity of chitosan was studied for a long time both *in vitro* and *in vivo*. They are used to control bacterial infections during the wound healing process [166]. The activity of chitosan against pathogens is maximal at acidic pH [166, 167]. The degree of acetylation has no influence on the activity, but high-molecular-weight chitosan was reported to be active against the Gram-negative bacteria *Escherichia coli* [168]. Depending on the authors, either high- or low-molecular-weight chitosan were reported to be active on the Gram-positive bacteria *Staphylococcus aureus* [168, 169]. To explain activity of chitosan, it is assumed that high-molecular-weight chitosan may form a film around the bacterial preventing the exchange of nutriment, whereas low-molecular-weight chitosan may enter bacteria cells to disturb metabolism [168]. Other mechanisms were suggested such as chelation of metal ions required for the synthesis of the bacterial toxin or interaction with cell membrane components inducing agglutination of bacteria or increasing of the permeability of the cell membrane [167, 170]. Chemically modified derivatives of chitosan such as diethy–methyl chitosane are more active than chitosan [171]. Chitosan has not yet been used in a targeted formulation for *in vivo* application. However, it was associated at the surface of nanoparticles to obtain fabrics with antibacterial activity.

Finally, antiviral activities were reported for polysaccharides bearing sulfonate groups. Most studies considered either the activity against the human herpes virus simplex 1 and 2 or against the Human Immunodeficiency Virus (HIV). Generally, the antiviral activity depended on the molecular weight, the amount and the distribution of the sulfonate groups, and the conformation and flexibility of the polysaccharide chains [172]. It was suggested that the anti-HIV activity can result from electrostatic interactions occurring between the negatively charged sulfonate groups of the polysaccharide and the positively charged amino-groups of the glycoprotein GP120 of the virus envelope [170]. Polysaccharides with antiviral activity can be obtained by grafting sulfonate or sulfamide groups [172, 174]. Modified dextranes containing carboxymethyl benzylamine and carboxylmethyl benzylaline sulfonates groups showed anti-HIV activity [175], whereas the anti-herpes virus activity was described for carraghennan corresponding to sulfated polygalactanes extracted from a red algae [172]. More interestingly, a sulphated polysaccharide extracted from the algae *Caulerpa racemosa* is active against strengths of the

herpes virus resistant toward acyclovir, which is one of the leading antiviral compound used in clinics [176]. These polysaccharides showing antiviral activity may be interesting compounds to target against infected cells.

### 5.2 Anticancer Activity

Several small carbohydrates derivatives showed anticancer activity and are used in clinics as antitumoral drugs [33–35]. Pharmacologically active concentrations of such an anticancer drug in the tumor tissue are often reached at the expense of massive contamination of the rest of the body. As it was discussed with antibiotics, this poor specificity creates a toxicological problem that represents a serious obstacle to effective antitumor therapy. Another obstacle to their efficient delivery is the occurrence of multidrug resistance (MDR), which may appear either as a lack of tumor size reduction or as a clinical relapse after an initial positive response to the tumor tissue; it can be either directly linked to specific mechanisms developed by the tumor cells or connected to the physiology of the tumor tissue, including a poor vasculature and unsuitable physico-chemical conditions [177]. Outside the tumor tissue, the resistance to chemotherapy can be due to the more general problem of the distribution of a drug relative to its targeted tissue [15]. To overcome drug resistance, many attempts have been made using strategies that consider the more general problem of the control of the drug biodistribution either at the cellular level or at the tissue level [178, 179]. These reasons make small carbohydrate antitumoral drugs relevant candidates to be formulated in drug delivery systems allowing their targeting in tumors as it was already demonstrated with the anthracyclin doxorubicin [50, 180–183].

Antitumor activity was also reported for several polysaccharides [35]. Nanoparticles made of chitosan either loaded with copper or unloaded showed a higher cytotoxicity against tumoral cells, whereas their cytotoxicity remained low toward healthy cells of human liver L-02. The antitumoral efficacy depended on the nanoparticle diameter being the highest with the smaller nanoparticles (40 nm) [184]. Heparin showed interesting antitumoral activity especially when it formed a complex with poly(beta-aminoesters), a positively charged polymer, which enhanced the intracellular penetration of heparin by endocytosis into cancer cells. The mechanism of the anticancer activity of heparin was explained by an induction of apoptosis in cells in which the polysaccharide was internalized. The specificity of the activity against cancer cells was explained by a favorable competition of the endocytosis toward these cells compared with the lower endocytosis activity of healthy cells. However, it may be suggested that the specificity of the delivery may further be increased by a specific delivery of the conjugate at the level of the tumor [185, 186]. Many other carbohydrates and polysaccharides have recently been isolated showing antitumoral activity. The characterization of their structure and mechanism of action are still not fully elucidated, and this is part of a very active field of investigation. Among the very recently isolated compounds showing an antitumoral activity, several were extracted from plants, mushrooms, and microorganisms [143, 187–191] or were collected from their culture medium [192]. It was shown that some extracted polysaccharides may interact with the host defense mechanisms to induce a production of interleukin-2, which in turn can stimulate either cytotoxic T lymphocytes, Natural Killer lymphocytes, or the production of an

antibody against tumoral cells [184]. Another hypothesis suggested that the polysaccharides can stimulate macrophages resulted in a production of nitrogen oxide (NO) and of tumor necrosis factor-alpha (TNF-α) [143, 187, 192–194]. A direct activity against tumoral cells was also assumed [188, 189]. Finally, certain active compounds may act on the angiogenesis reducing the formation of neovessels required for the irrigation of the tumor tissue [190, 191].

### 5.3 Control of Vascular Smooth Muscle Cell Proliferation

Several sulfonated polysaccharides showed an ability to control proliferation of vascular smooth muscle cells. Certain compounds may act on the angiogenesis reducing the formation of neovessels required for the irrigation of tumor tissues [190, 191]. They can contribute to reduce tumor size growth. Others can inhibit the proliferation of the vascular smooth muscle cells on a damaged artery to prevent and control artherosclerosis [195]. The inhibitor activity of the polysaccharides seemed to increase with their molecular weight because it is assumed that they are better internalized by cells. Among sulfonated polysaccharides, fucan showed a higher activity compared with heparin [196, 197]. With modified dextran, the inhibition of cell growth depended on the degree of substitution by sulfonate groups and on the presence of benzylamide groups. The production of collagen by the vascular smooth muscle cells was influenced by modified dextran and heparin [198]. Complexes among sulfonated polysaccharides, such as heparin, hyaluronate, and chondroitine sulphate with chitosan, were also reported to control proliferation of vascular smooth muscle cells [199].

Sulfonated polysaccharides present different biological activities. To take advantage of their antiproliferative activity toward vascular smooth muscle cells, it will be necessary to associate the polysaccharides with targeted drug carriers such as nanoparticles. The nanoparticles may also enhance the cellular endocytosis of the polysaccharides to further promote its antiproliferative activity on vascular cells.

### 5.4 Activity on the Immune System

Many antigens located at the surface of pathogens show more or less complex carbohydrate epitopes. The immune response induced by these antigens results in the production of antibodies, which is relevant in the aim of vaccination [200–203]. Thus, a few vaccines were developed using polysaccharide antigens against meningitis [203, 204], pneumonia [202], and against *Staphylococcus aureus*, which is responsible for nocosomial infections [201]. Usually, an efficient vaccine requires formulation of the antigen with an adjuvant. The more recent formulation strategy that was followed for vaccine development was based on the design of antigen presenting devices using nanotechnology [205–207]. Systems proposed so far are, on the one hand, derived from the liposome technology based on phospholipid bilayer reconstitution with incorporation of antigenic protein or glycoprotein, i.e., virosomes [205, 206], and, from the other hand, on the reconstitution of empty viral capside with viral recombinant proteins, i.e., virus like particles [207]. Finally, the immune stimulating complex (ISCOM) corresponds to a nanosized aggregate (40 nm in diameter) of saponin, lipids, and antigen held together by hydrophobic interactions occurring among the three components [208]. These systems formed spontaneously upon self-molecular assembly mechanisms. They were mainly

developed as vaccines against viral infections and are mimicking viruses showing the corresponding antigenic epitopes at the surface. Although they are efficient for parenteral vaccinations, their application in mucosal vaccines is subjected to controversy because of possible stability problems in the biological environment of mucosa. Thus, there is a need for development of antigen presenting systems for bacterial antigens including those of a polysaccharide nature and for antigen presenting devices stable enough in biological media encountered at the level of mucosa to develop mucosal vaccination. Several types of polymer nanoparticles were proposed as carriers for antigens to be used as vaccine adjuvants [209, 210]. In general, in these systems, the antigen is mostly encapsulated within the nanoparticles and is not "visible" from the outside of the carrier as it is the case of systems obtained by self-assembly molecules, i.e., in virosomes and virus-like particles. Polysaccharide-coated nanoparticles would be more suitable as possible antigen presenting devices. To this purpose, nanoparticles were specifically designed to exhibit polysaccharides at the nanoparticle surface and can easily be obtained with various polysaccharides [211]. Their versatility should offer interesting developments for new vaccines as a polysaccharide antigen presenting device.

Although carbohydrate antigens were first described a long time ago, their interactions with the immune system are not limited to the production of antibodies in the frame of vaccination. Indeed, some polysaccharide can modulate the immune system activity in a specific manner. For instance, several polysaccharides interact with interleukin 10 (IL-10) [212]. Human IL-10 is known to stimulate CD16 and CD64 receptor expression on the monocyte and macrophage population within peripheral blood mononuclear cells. Soluble heparin, heparan sulfate, chondroitin sulfate, and dermatan sulfate inhibit the human IL-10-induced expression of CD16 and CD64 in a concentration-dependent manner. The antagonistic effect of heparin on human IL-10 activity depends on N-sulfation, the de-N-sulfated heparin had little or no inhibitory effect on the IL-10- induced expression of CD16, and the effect of de-O-sulfated heparin was comparable with that of unmodified heparin. Furthermore, the inhibition of cell-bound proteoglycan sulfation reduced the human IL-10-mediated expression of CD16 molecules on monocytes and macrophages. Taken together, these findings support the hypothesis that soluble and cell-surface glucosaminoglycans and, in particular, their sulfate groups are important in binding and modulation of human IL-10 activity and, therefore, promote or inhibit several pathways of the immune system. It may be suggested that these molecules grafted at the surface of a polymer nanoparticles may regulate the expression of the receptors at the cell surface modifying the immune system response in a very definitive and precise manner.

Small oligosaccharides can inhibit immune response by reducing production of oxidative species in immune cells. These components may be interesting to control the inflammatory process and for treatment of autoimmune diseases. Activity depended on both length and conformation of the oligosaccharides. It seemed that oligosaccharides longer than six residues and adopting a helical conformation enhanced interactions with immune system components [213]. The immunomodulation properties shown by such oligosaccharides can be combined with anti-inflammatory properties of other polysaccharides extracted from *Costus spicatus* [214]. Although several oligosaccharides can inhibit immune response, others stimulate the immune system. Polysaccharides extracted from *Lycium barabrum*

promoted expression of IL-2, TNF-alpha, in the peripheral mononuclear cells in human [215]. Those extracted from *Discirea opposita* induced proliferation of T lymphocytes [216]. It was suggested that stimulation of macrophages resulted from an interaction occurring between the n-acetylglucosamine groups of the oligosaccharides with the mannose calcium-independent receptors of the macrophages [217, 218].

Presenting under the form of nanoparticles, these polysaccharides may be better targeted toward macrophages and monocytes to focus the activity of the polysaccharide in macrophages and to avoid reaching other cells that may be affected by unspecific action on metabolism or physiology. Indeed, it will be very important to target precisely polysaccharides like heparin because it has other biological activity. The various activities found for polysaccharides on the immune system may be used to finely tune the activity of the very complex defense system of the host. This could lead to the emergence of new therapeutic strategies, provided the polysaccharides or the oligosaccharides will be precisely targeted against the right cells of the immune systems.

### 5.5 Control of Blood Coagulation and Complement Activation

Polysaccharides and oligosaccharides may induce or inhibit blood coagulation depending on their structure. Chitosan and polysaccharides extracted from the Chinese laquer showed biological activity promoting the blood coagulation [173, 219]. In general, negatively charged polysaccharides, including sulphate, sulfonate, carboxylic, and sulphonamide groups, show an opposite activity having anticoagulant properties. Heparin, which is used in clinics and for storage of blood in the blood banks, is for sure the more generally known natural anticoagulant agent [134, 220]. For the other polysaccharides, including sulphated and carboxylated derivatives of dextran, galactomannan, pullulan, and chitosan, it was shown that the amount of functional groups, their distribution along the chain, and the molecular weight of the polysaccharide chains are all parameters that influence the anticoagulant activity of the polysaccharide [174, 220–226]. In clinics, the main problem encountered with treatments based on the use of heparin is the rapid elimination of the polysaccharides from the blood stream. Long circulating formulations of heparin with longer biological and therapeutic efficacy may consist of heparin associated with long circulating drug carriers. It was reported that some nanoparticles coated with heparin displayed long circulating properties and that the anticoagulant activity of the heparin was preserved after grafting on the nanoparticle surface, making these nanoparticles interesting as a potential long-acting anticoagulant agent [19, 89]. Heparin that is used in clinics must be administered via the parenteral route, which is expensive, inconvenient, unsafe, and limits use by outpatients. The development of an oral form of heparin is more than warranted because the oral route remains the preferable route of administration of drugs for patient. It is also the safer and less-expensive route for the administration of drugs so far. However, heparin is a polyanionic macromolecule and is unstable under acidic conditions of the stomach, thus exhibiting poor oral bioavailability [227]. Nanoparticles were found to enhance absorption of hydrophilic macromolecules across the intestinal epithelium [51]. They seemed to be an interesting alternative for the development of an oral formulation of heparin to promote its bioavailability by this challenging route of administration.

## 6 NANOPARTICLE DRUG CARRIERS MADE OF CARBOHYDRATES FOR THE DELIVERY OF FRAGILE MOLECULES

Among the different delivery approaches explored so far to deliver complex molecules such as peptides, proteins, and nucleic acids by mucosal routes, those based on the use of polysaccharide nanoparticles were found to be interesting alternatives toward this ambitious goal thanks to physico-chemical and biopharmaceutical properties of these polymers [6–8, 51]. Indeed, the mucosal delivery of these molecules is very challenging, and despite their increasing market value, their clinical use was hampered due to their poor natural transport across biological barriers and their extremely rapid degradation in biological media. Elaboration of the polysaccharide nanoparticles requires neither an organic solvent nor a powerful homogenization procedure, possibly damaging macromolecular drug activity [75, 228]. This is an advantage regarding all other preparation methods of nanoparticles. Indeed, the mild conditions used to prepare polysaccharide nanoparticles suits with those required to preserve the biological activity of peptides, proteins, and nucleic acids in the nanoparticle formulations [9]. So far, nanoparticles made of chitosan have opened promising alternatives for the delivery of peptides, proteins, and nucleic acids by the mucosal route, including oral and nasal administration and for topical delivery of drugs in the eye [229–231]. The system that is stable in the mucosal environment, including in the harsh conditions of the gastrointestinal tract, can interact with the negatively charged glycoproteins of the mucus, which cover the epithelium retaining the nanoparticles on the mucosa surface. This retention effect due to the bioadhesive properties of chitosan to mucus increases the chance for the drug to be absorbed. The drug-loaded nanoparticles either made of chitosan or bearing a chitosan coating enhanced drug penetration and absorption after administration by different mucosal routes. For instance, the oral absorption of salmon calcitonin was improved when it was associated with chitosan nanoparticles. Interestingly, this system elicited long-lasting hypocalcemia levels, whereas calcitonin emulsions used as control led only to a negligible response [8]. To explain the effect, it was suggested that the nanosystems were able to enter the epithelia and provided a continuous delivery of the peptide to the blood stream [8, 232, 233]. Insulin-loaded chitosan nanoparticles also improved the delivery of insulin by the nasal route [234]. This formulation may be interesting for applications requiring a rapid supply of insulin in the blood. Indeed, the hormone is absorbed in about 30 minutes, but the duration of the effect was quite short, lasting for 1.5 to 2 hours.

Chitosan nanoparticles as well as chitosan-coated nanoparticles were also found interesting for the delivery of antigens to the Nasal Associated Lymphoid tissue. The results indicated that the nanoparticles could facilitate the transport of the associated antigen across the nasal epithelium leading to an efficient antigen presentation to the immune system. This elicited both an IgA and an IgG response increasing over time. For the immune response, it was suggested that not only the size and surface properties of the nanoparticles may influence the immune response but also the polymer composition and the structural architecture of the nanosystems, which were found critical for the optimization of such antigen carriers [210]. For the delivery of antigen by the oral route, it has very recently been proposed to coat chitosan nanoparticles with alginate in order to prevent a burst released of the antigen in the intestinal medium and to further improve the stability of the chitosan nanoparticles in this very aggressive medium [235].

Chitosan nanoparticles and chitosan-coated nanoparticles showed interesting bioadhesiveness on the cornea and good ocular tolerance. Their potential for the delivery of drugs at the ocular level was investigated. Both systems showed suitable delivery properties to improve topical administration of drugs at the level of the ocular mucosa [231]. Chitosan-coated nanoparticles, which are very versatile regarding their drug-loading capacity and release properties, increased the drug levels in cornea and aqueous humor to a significantly greater extent compared with the commercial drug preparation or the drug-loaded uncoated nanoparticles. These nanoparticles were clearly able to enhance transportation of the drug across the cornea to reach the inner part of the eye. In contrast, chitosan nanoparticles accumulated into the corneal and conjunctival epithelia, where an increased concentration of drugs can be maintained for a prolonged duration. Initial experiments have indicated the adequate tolerance and low toxicity of these ocular drug carriers [73].

Several polysaccharides can form a complex with nucleic acids. This property was exploited to formulate DNA loaded nanoparticles to deliver genes and antisens oligonucleotides *in vivo*. Most polysaccharide-based carriers were obtained from chitosan and alginate [7, 236]. *In vitro* expression of gene transfected into cells using chitosan nanoparticles reached a similar level than the expression level obtained with reference transfection agents, including Lipofectin (Invitrogen, Cergy-Pontoise, France) and Surperfect (Qiagen S.A., Coutaboef, France) [237]. The nanoparticles presented high encapsulation efficiency and good protection of DNA from DNase digestion [238]. To further improve the transfection properties of chitosan nanoparticles and enhance the specificity of the carrier toward the target cells, Manssouri et al. suggested to attach folic acid residues to the nanoparticle surface [239]. Nanoparticles made of polyesters or silica coated with chitosan or modified chitosan may also serve as carriers for DNA plasmids or for antisense oligonucleotides [7, 62, 240–243]. In these nanoparticles, the nucleic acid is adsorbed onto the carrier surface thanks to the formation of a polyion complex with the polycationic polysaccharide standing at the nanoparticle surface. Another system made of alginate nanoparticles was reported to associate antisense oligonucleotides [244]. The stability of the associated oligonucleotides in the presence of serum was considerably improved, indicating that the carrier system may be suitable for the *in vivo* delivery of antisense oligonucleotides. The main objective of gene therapy is to obtain successful *in vivo* transfer of the genetic material to the targeted tissue. However, the growing potential of gene therapy will not achieve this goal until the issue of gene delivery has been resolved [245]. Naked DNA cannot be delivered efficiently. Consequently, good gene delivery formulations are needed requiring good encapsulation efficiency of DNA, protection from degradation, and specific targeting to desired cells. The optimal system is still not found. The polysaccharide nanoparticles presenting so far good complexing and transfection efficiency are interesting potential systems. It may be expected that research on these systems will continue to grow and expand in the close future.

Polyanionic polysaccharides including heparin were shown to interact strongly with hemoglobin preserving its functionality as oxygen carrier [246]. This was exploited to load functional hemoglobin on heparin-coated nanoparticles to be used as a possible oxygen carrier for *in vivo* applications [247]. The nanoparticles having a diameter of 80 nm showed the highest hemoglobin-loaded capacity compared with other carrier systems. Indeed, they can associate up to 40-mg

hemoglobin per gram of nanoparticles, whereas other nanoparticles showed one third of this loading capacity [247, 248]. The functionality of hemoglobin standing in the heparin corona at the nanoparticle surface was preserved as well as the anticoagulant activity of heparin. This system in which a drug may additionally be incorporated in the polymer core can constitute a new generation of multifunctional drug delivery system.

## 7 CONCLUSION

In the recent years, there has been a resurgence of interest in the biological role of carbohydrates. Active research that started a couple of years ago already highlighted the many biological roles and the diverse functions of carbohydrates in physiological and diseased processes. Several are useful for application as powerful drugs providing they are specifically delivered to the diseased cells. Their combination with nanoparticle drug carriers may be required improving their efficacy and specificity at the target site. From another point of view, several polysaccharides seemed to be interesting constituents of a drug carrier having remarkable mucoadhesive properties and promoting drug delivery by mucosal routes. The wide applications of carbohydrates in the design of drug delivery systems are illustrated in Figure 1.

Carbohydrate-based nanoparticles seemed to be well tolerated so far. However, full toxicological studies have not been performed yet and safety of the formulated

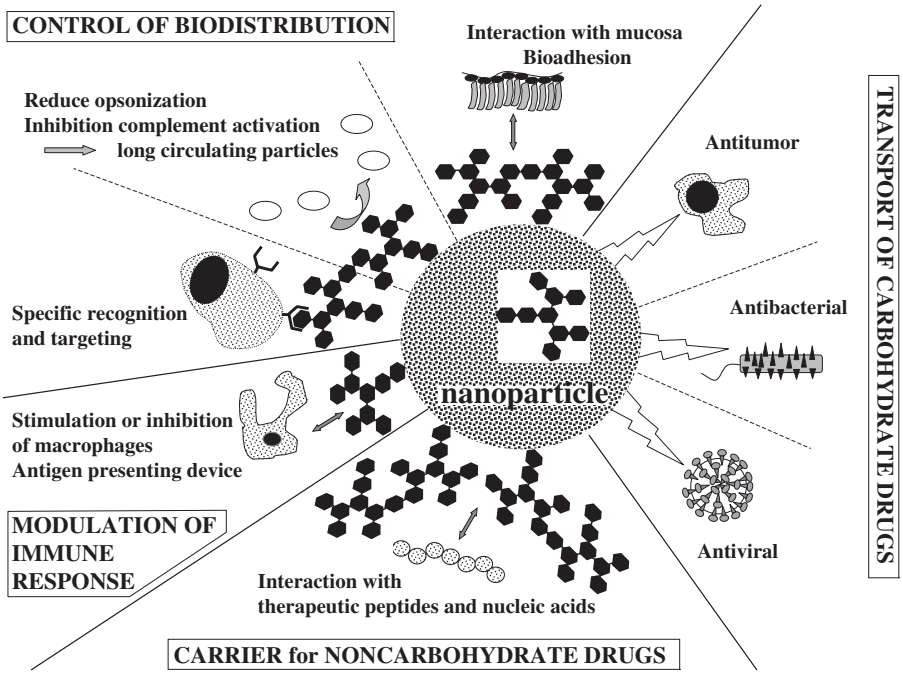


**Figure 1.** Diverse carbohydrate properties and their application in the design of a nanoparticle drug delivery system.

delivery systems will be an important point to clarify in the near future for further development for *in vivo* use. Finally, because carbohydrates are implicated in many cell recognition mechanisms and cell signaling phenomena based on high specific interactions with corresponding cell receptors, they may provide very useful tools to increase the targeting specificity of drug delivery systems toward defined cells *in vivo*. Considerable work remains to be done before the potential of carbohydrates will be depicted and will be ready to be integrated into the design of better defined drug delivery systems with more precise functions. This will require the understanding of the role of biological functions of more carbohydrates found in Mother Nature and production or synthesis of sufficient amounts of the defined carbohydrate to clarify their functionalities. From the first results of the experiences done so far on quite simple systems, it may be expected that carbohydrates will allow us to step over a new milestone in the development of drug delivery systems integrating the most recent developments in glycobiology.

## REFERENCES


1. Yoshida T (2001). Synthesis of polysaccharides having specific biological activities. *Progress in Polymer Sci.* 26:379–441.
2. Venkataraman G, Shriver Z, Raman R, et al. (1999). Sequencing complex polysaccharides. *Science.* 286:537–542.
3. Toukach F V, Shashkov A S (2001). Computer-assisted structural analysis of regular glycopolymers on the basis of 13C NMR data. *Carbohydrate Research.* 335:101–114.
4. Lee Y C (1996). Carbohydrate analyses with high-performance anion-exchange chromatography. *J. Chromato. A.* 720:137–149.
5. Gabius H-J (2004). The sugar code in drug delivery. *Adv. Drug Del. Rev.* 56:421–424.
6. Vauthier C, Couvreur P (2000). Development of Polysaccharide Nanoparticles as Novel Drug Carrier Systems. In D L Wise (ed.), *Handbook of Pharmaceutical Controlled Release Technology.* Marcel Dekker, New York, pp. 413–429.
7. Janes K A, Calvo P, Alonso M J (2001). Polysaccharide colloidal particles as delivery systems for macromolecules. *Adv. Drug Del. Rev.* 47:83–97.
8. Prego C, Garcia M, Alonso M J (2005). Transmucosal macromolecular drug delivery. *J. Control. Rel.* 101:151–162.
9. Vauthier C, Fattal E, Labarre D (2004). From polymer chemistry and physicochemistry to nanoparticulate drug carrier design and applications. In Yaszemski M J, Trantolo D J, Lewandrowski K V, et al. (eds.), *Tissue Engineering and Novel Delivery System*: Marcel Dekker, New York, pp. 562–598.
10. Nori A, Kopecek J (2005). Intracellular targeting of polymer-bound drugs for cancer chemotherapy. *Adv. Drug Del. Rev.* 57:609–636.
11. Duncan R (2003). The dawning era of polymer therapeutics. Nature Reviews. *Drug Disc.* 2:347–360.
12. Nobs L, Buchegger F, Gurny R, et al. (2004). Current methods for attaching targeting ligands to liposomes and nanoparticles. *J. Phar. Sci.* 93:1980–1992.
13. Torchilin V P (2005). Recent advances with liposomes as pharmaceutical carriers. *Nat. Rev. Drug Disc.* 4:145–160.

14. Moghimi S M, Szebeni J (2003). Stealth liposomes and long circulating nanoparticles: Critical issues in pharmacokinetics, opsonization and protein-binding properties. *Progress Lipid Res.* 42:463–478.
15. Moghimi S M, Hunter A C, Murray J C (2001). Long-circulating and target-specific nanoparticles: Theory to practice. *Pharmacol. Rev.* 53:283–318.
16. Gregoriadis G, McCormack B, Wang Z, Lifely R (1993). Polysialic acids: Potential in drug delivery. *FEBS Lett.* 315:271–276.
17. Olivier J-C, Vauthier C, Taverna M, et al. (1996). Stability of orosomucoid-coated polyisobutylcyanoacrylate nanoparticles in the presence of serum. *J. Control. Rel.* 40:157–168.
18. Passirani C, Ferrarini L, Barratt G, et al. (1999). Preparation and characterization of nanoparticles bearing heparin or dextran covalently-linked to poly(methyl methacrylate). *J. Biomater. Sci. Polym. Ed.* 10:47–62.
19. Chauvierre C, Labarre D, Couvreur P, et al. (2003). Novel polysaccharide-decorated poly(isobutyl cyanoacrylate) nanoparticles. *Pharm. Res.* 20:1786–1793.
20. Bertholon I, Vauthier C, Labarre D (2006). Core-shell poly(isobutylcyanoacrylate)-polysaccharide nanoparticles: Influences of surface morphology, length and type of polysaccharide on complement activation in vitro. *Pharm. Res.* 23:1313–1323.
21. Bertholon I, Ponchel G, Labarre D (2006). Bioadhesive properties of poly(alkylcyanoacrylate) nanoparticles coated with polysaccharide. *J. Nanosci. Nanotechnol.* 6:3102–3109.
22. Sihorkar V, Vyas S P (2001). Potential of polysaccharide anchored liposomes in drug delivery, targeting and immunization. *J. Pharm. Pharmceut. Sci.* 4:138–158.
23. Drotleff S, Lungwitz U, Breunig M, et al. (2004). Biomimetic polymers in pharmaceutical and biomedical sciences. *Euro. J. Pharmaceutics and Biopharmaceutics.* 58:385–407.
24. Bravo-Osuna I, Schmith T, Bernkop-Schnurch A (2006). Elaboration and characterization of thiolated chitosan-coated acrylic nanoparticles. *Int. J. Pharm.* 316:170–175.
25. Eylar E H (1966). On the biological role of glycoproteins. *J. Theoret. Bio.* 10:89–113.
26. Kimmerlin T, Seebach D (2005). "100 years of peptide synthesis": Ligation methods for peptide and protein synthesis with applications to beta-peptide assemblies. *J. Peptide Res.: Official J. Amer. Peptide Soc.* 65:229–260.
27. Zamecnik P (2005). From protein synthesis to genetic insertion. *Ann. Rev. Biochem.* 74:1–28.
28. Sharon N, Lis H (2004). History of lectins: From hemagglutinins to biological recognition molecules. *Glycobiology.* 14:53R–62R.
29. Khersonsky S M, Ho C M, Garcia M-A F, et al. (2003). Recent advances in glycomics and glycogenetics. *Curr. Topics Medic. Chem.* 3:617–643.
30. Petitou M, Casu B, Lindahl U (2003). 1976–1983, A critical period in the history of heparin: the discovery of the antithrombin binding site. *Biochimie.* 85:83–89.
31. Passirani C, Barratt G, Devissaguet J-P, et al. (1998). Interactions of nanoparticles bearing heparin or dextran covalently bound to poly(methyl methacrylate) with the complement system. *Life Sci.* 62:775–785.
32. Labarre D, Vauthier C, Chauvierre C, et al. (2005). Interactions of blood proteins with poly(isobutylcyanoacrylate) nanoparticles decorated with a polysaccharidic brush. *Biomaterials.* 26:5075–5084.
33. Di Marco A, Cassinelli G, Arcamone F (1981). The discovery of daunorubicin. *Cancer Treat. Rep.* 65:3–8.

34. Osborn H M, Evans P G, Gemmel N, et al. (2004). Carbohydrate-based therapeutics. *J. Pharm. Pharmacol.* 56:691–702.
35. Barchi J J (2000). Emerging roles of carbohydrates and glycomimetics in anticancer drug design. *Curr. Pharm. Des.* 6:485–501.
36. Kabel M A, Schols H A, Voragen A G J (2001). Mass determination of oligosaccharides by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry following HPLC, assisted by on-line desalting and automated sample handling. *Carboh. Polym.* 44:161–165.
37. Broberg A, Thomsen K K, Duus J O (2000). Application of nano-probe NMR for structure determination of low nanomole amounts of arabinoxylan oligosaccharides fractionated by analytical HPAEC-PAD. *Carboh. Res.* 328:375–382.
38. Morelle W, Michalski J-C (2005). Glycomics and mass spectrometry. *Curr. Pharm. Des.* 11:2615–2645.
39. Iwafune M, Kakizaki I, Nakazawa H, et al. (2004). A glycomic approach to proteoglycan with a two-dimensional polysaccharide chain map. *Analyt. Biochem.* 325:35–40.
40. Ratner D M, Adams E W, Disney M D, et al. (2004). Tools for glycomics: Mapping interactions of carbohydrates in biological systems. *Chembiochem: A Euro. J. Chem. Bio.* 5:1375–1383.
41. Howbrook D N, van der Valk A M, O'Shaughnessy M C, et al. (2003). Developments in microarray technologies. *Drug Dis. Today.* 8:642–651.
42. Vives R R, Goodger S, Pye D A (2001). Combined strong anion-exchange HPLC and PAGE approach for the purification of heparan sulphate oligosaccharides. *Biochem. J.* 354:141–147.
43. Plante O J (2005). Combinatorial chemistry in glycobiology. *Combinatorial Chem. & High Throughput Screening.* 8:153–159.
44. Holemann A, Seeberger P H (2004). Carbohydrate diversity: Synthesis of glycoconjugates and complex carbohydrates. *Curr. Opin. Biotech.* 15:615–622.
45. Köhn M, Breinbauer R (2004). The Staudinger ligation-a gift to chemical biology. *Angew. Chem. Int. Ed. Engl.* 43:3106–3116.
46. de Kort M, Buijsman R C, van Boeckel C A A (2005). Synthetic heparin derivatives as new anticoagulant drugs. *Drug Dis. Today.* 10:769–779.
47. Croft S L, Coombs G H (2003). Leishmaniasis–current chemotherapy and recent advances in the search for novel drugs. *Trends Parasitol.* 19:502–508.
48. Adler-Moore J P, Proffitt R T (1993). Development, characterization, efficacy and mode of action of ambisome®. *J. Liposome Res.* 3:429–450.
49. Couvreur P, Barratt G, Fattal E, et al. (2002). Nanocapsule technology: A review. *Crit. Rev. Ther. Drug Carrier Syst.* 19:99–134.
50. Vauthier C, Dubernet C, Chauvierre C, et al. (2003). Drug delivery to resistant tumors: The potential of poly(alkyl cyanoacrylate) nanoparticles. *J. Control. Rel.* 93:151–160.
51. Alonso M J (2004). Nanomedicines for overcoming biological barriers. *Biomed. Pharmacother.* 58:168–172.
52. Fouarge M, Dewulf M, Couvreur P, et al. (1989). Development of dehydroemetine nanoparticles for the treatment of visceral leishmaniasis. *J. Microencapsulation.* 6:29–34.
53. Gaspar R, Preat V, Opperdoes F R, et al. (1992). Macrophage activation by polymeric nanoparticles of polyalkylcyanoacrylates: activity against intracellular Leishmania donovani associated with hydrogen peroxide production. *Pharm. Res.* 9:782–787.

54. Kattan J, Droz J P, Couvreur P, et al. (1992). Phase I clinical trial and pharmacokinetic evaluation of doxorubicin carried by polyisohexylcyanoacrylate nanoparticles. *Investigat. New Drugs.* 10:191–199.
55. Barraud L, Merle P, Soma E, et al. (2005). Increase of doxorubicin sensitivity by doxorubicin-loading into nanoparticles for hepatocellular carcinoma cells in vitro and in vivo. *J. Hepatol.* 42:736–743.
56. Sangro B (2005). Refined tools for the treatment of hepatocellular carcinoma. *J. Hepatol.* 42:629–631.
57. Tyagi R, Lala S, Verma A K, et al. (2005). Targeted delivery of arjunglucoside I using surface hydrophilic and hydrophobic nanocarriers to combat experimental leishmaniasis. *J. Drug. Target.* 13:161–171.
58. Stolnik S, Dunn S E, Garnett M C, et al. (1994). Surface modification of poly(lactide-co-glycolide) nanospheres by biodegradable poly(lactide)-poly(ethylene glycol) copolymers. *Pharm. Res.* 11:1800–1808.
59. Gref R, Minamitake Y, Peracchia M T, et al. (1994). Biodegradable long-circulating polymeric nanospheres. *Science.* 263:1600–1603.
60. Vauthier C, Couvreur P (2002). Biodegradation of poly(alkylcyanoacrylates) in Matsumara J P and Steinbuchel A (eds.), *Handbook of Biopolymers*, Wiley-VHC, Weinheim, Germany, pp. 457–490.
61. Ravi Kumar M N V, Bakowsky U, Lehr C M (2004). Preparation and characterization of cationic PLGA nanospheres as DNA carriers. *Biomaterials.* 25:1771–1777.
62. Bertholon-Rajot I, Labarre D, Vauthier C (2005). Influence of the initiator system, cerium-polysaccharide, on the surface properties of poly(isobutylcyanoacrylate) nanoparticles. *Polymer.* 46:1407–1415.
63. Schipper N G M, Olsson S, Hoogstraate J A, et al. (1997). Chitosans as Absorption Enhancers for Poorly Absorbable Drugs 2: Mechanism of Absorption Enhancement. *Pharm. Res.* 14:923–929.
64. Smith J M, Dornish M, Wood E J (2005). Involvement of protein kinase C in chitosan glutamate-mediated tight junction disruption. *Biomaterials.* 26:3269–3276.
65. Lehr C-M, Bouwstra J A, Schacht E H, et al. (1992). In vitro evaluation of mucoadhesive properties of chitosan and some other natural polymers. *Int. J. Pharm.* 78:43–48.
66. Schipper N G, Varum K M, Artursson P (1996). Chitosans as absorption enhancers for poorly absorbable drugs. 1: Influence of molecular weight and degree of acetylation on drug transport across human intestinal epithelial (Caco-2) cells. *Pharm. Res.* 13:1686–1692.
67. Chae S Y, Jang M-K, Nah J-W (2005). Influence of molecular weight on oral absorption of water soluble chitosans. *J. Control. Rel.* 102:383–394.
68. Thanou M, Verhoef J C, Junginger H E (2001). Oral drug absorption enhancement by chitosan and its derivatives. *Adv. Drug Deliv. Rev.* 52:117–126.
69. Avadi M R, Jalali A, Sadeghi A M M, et al. (2005). Diethyl methyl chitosan as an intestinal paracellular enhancer: ex vivo and in vivo studies. *Int. J. Pharm.* 293:83–89.
70. Sandri G, Rossi S, Ferrari F, et al. (2004). Assessment of chitosan derivatives as buccal and vaginal penetration enhancers. *Eur. J. Pharm. Sci.* 21:351–359.
71. Kafedjiiski K, Föger F, Werle M, et al. (2005). Synthesis and in Vitro Evaluation of a Novel Chitosan-Glutathione Conjugate. *Pharm. Res.* 22:1480–1488.
72. Bernkop-Schnurch A, Pinter Y, Guggi D, et al. (2005). The use of thiolated polymers as carrier matrix in oral peptide delivery–Proof of concept. *J. Control. Rel.* 106:26–33.

73. De Campos A M, Sanchez A, Gref R, et al. (2003). The effect of a PEG versus a chitosan coating on the interaction of drug colloidal carriers with the ocular mucosa. *Eur. J. Pharm. Sci.* 20:73–81.
74. Yamamoto H, Takeuchi H, Hino T, et al. (2000). Mucoadhesive liposomes: Physicochemical properties and release behavior of water-soluble drugs from chitosan-coated liposomes. *S.T.P. Pharma. Sci.* 10:63–68.
75. Calvo P, Vila-Jato J L, Alonso M J (1997). Evaluation of cationic polymer-coated nanocapsules as ocular drug carriers. *Int. J. Pharm.* 153:41–50.
76. Yamamoto H, Kuno Y, Sugimoto S, et al. (2005). Surface-modified PLGA nanosphere with chitosan improved pulmonary delivery of calcitonin by mucoadhesion and opening of the intercellular tight junctions. *J. Control. Rel.* 102:373–381.
77. Messai I, Delair T (2005). Adsorption of chitosan onto poly(D,L-lactic acid) particles: A physico-chemical investigation. *Macromolecular Chem. Phys.* 206:1665–1674.
78. Yang S C, Ge H X, Hu Y, et al. (2000). Formation of positively charged poly(butyl cyanoacrylate) nanoparticles stabilized with chitosan. *Colloid Polym. Sci.* 278:285–292.
79. Ponchel G, Irache J (1998). Specific and non-specific bioadhesive particulate systems for oral delivery to the gastrointestinal tract. *Adv. Drug Del. Rev.* 34:191–219.
80. Valenta C (2005). The use of mucoadhesive polymers in vaginal delivery. *Adv. Drug Del. Rev.* 57:1692–1712.
81. Song Y, Wang Y, Thakur R, et al. (2004). Mucosal drug delivery: Membranes, methodologies, and applications. *Crit. Rev. Ther. Drug Carrier Syst.* 21:195–256.
82. Barbault-Foucher S, Gref R, Russo P, et al. (2002). Design of poly-e-caprolactone nanosphere coated with bioadhesive hyaluronic acid for ocular delivery. *J. Control. Rel.* 83:365–375.
83. Leroux J-C, De Jaeghere F, Anner B, et al. (1995). An investigation on the role of plasma and serum opsonins on the evternalization of biodegradable poly(D,L-lactic acid) nanoparticles by human monocytes. *Life Sciences.* 57:695–703.
84. Lenaerts V, Nagelkerke J F, Van Berkel T J, et al. (1984). In vivo uptake of polyisobutyl cyanoacrylate nanoparticles by rat liver Kupffer, endothelial, and parenchymal cells. *J. Pharm. Sci.* 73:980–982.
85. Lemarchand C, Gref R, Couvreur P (2004). Polysaccharide-decorated nanoparticles. *Eur. J. Pharm. Biopharm.* 58:327–341.
86. Torchilin V P (2000). Drug targeting. *Eur. J. Pharm. Sci.* 11:S81–S91.
87. Owens III D E, Peppas N A. Opsonization, biodistribution, and pharmacokinetics of polymeric nanoparticles. *Int. J. Pharm.* In Press, Corrected Proof.
88. Torchilin V P, Trubetskoy V S (1995). Which polymers can make nanoparticulate drug carriers long-circulating? *Adv. Drug Del. Rev.* 16:141–155.
89. Passirani C, Barratt G, Devissaguet J P, et al. (1998). Long-circulating nanoparticles bearing heparin or dextran covalently bound to poly(methyl methacrylate). *Pharm. Res.* 15:1046–1050.
90. Lemarchand C, Gref R, Passirani C, et al. (2006). Influence of polysaccharide coating on the interactions of nanoparticles with biological systems. *Biomaterials.* 27:108–118.
91. Jaulin N, Appel M, Passirani C, et al. (2000). Reduction of the uptake by a macrophagic cell line of nanoparticles bearing heparin or dextran covalently bound to poly(methyl methacrylate). *J. Drug Targeting.* 8:165–172.
92. Rouzes C, Gref R, Leonard M, et al. (2000). Surface modification of poly(lactic acid) nanospheres using hydrophobically modified dextrans as stabilizers in an o/w emulsion/evaporation technique. *J. Biomed. Mat. Res.* 50:557–565.

93. Carreno M P, Maillet F, Labarre D, et al. (1988). Specific antibodies enhance Sephadex-induced activation to the alternative pathway in human serum. *Biomaterials.* 9:514–518.
94. Olivier J-C (1995). Développement d'un vecteur biomimétique, PhD Dissertation, Université Paris 11, Faculté de Pharmacie, Chatenay-Malabry. France, p. 133.
95. Maruyama A, Ishihara T, Kim J S, et al. (1997). Nanoparticle DNA carrier with poly (L-lysine) grafted polysaccharide copolymer and poly(D,L-lactic acid). *Bioconjug. Chem.* 8:735–742.
96. Benesch J, Tengvall P (2002). Blood protein adsorption onto chitosan. *Biomaterials.* 23:2561–2568.
97. Minami S, Suzuki H, Okamoto Y (1998). Chitin and chitosan activate complement via the alternative pathway. *Carbohyd. Polym.* 36:151–155.
98. Suzuki Y, Okamoto Y, Morimoto M, et al. (2000). Influence of physico-chemical properties of chitin and chitosan on complement activation. *Carbohyd. Polym.* 42:307–310.
99. Mitra S, Gaur U, Ghosh P C, et al. (2001). Tumour targeted delivery of encapsulated dextran-doxorubicin conjugate using chitosan nanoparticles as carrier. *J. Control. Rel.* 74:317–323.
100. Banerjee T, Mitra S, Kumar Singh A, et al. (2002). Preparation, characterization and biodistribution of ultrafine chitosan nanoparticles. *Int. J. Pharm.* 243:93–105.
101. Banerjee T, Singh A K, Sharma R K, et al. (2005). Labeling efficiency and biodistribution of Technetium-99m labeled nanoparticles: Interference by colloidal tin oxide particles. *Int. J. Pharm.* 289:189–195.
102. Son Y J, Jang J-S, Cho Y W, et al. (2003). Biodistribution and anti-tumor efficacy of doxorubicin loaded glycol-chitosan nanoaggregates by EPR effect. *J. Control Rel.* 91:135–145.
103. Kato Y, Onishi H, Machida Y (2000). Evaluation of N-succinyl-chitosan as a systemic long-circulating polymer. *Biomaterials.* 21:1579–1585.
104. Hyung Park J, Kwon S, Lee M (2006). Self-assembled nanoparticles based on glycol chitosan bearing hydrophobic moieties as carriers for doxorubicin: In vivo biodistribution and anti-tumor activity. *Biomaterials.* 27:119–126.
105. Schauer R (1985). Sialic acids and their role as biological masks. *Trends Biochem. Sci.* 10:357–360.
106. Durocher J R, Payne R C, Conrad M E (1975). Role of sialic acid in erythrocyte survival. *Blood.* 45:11–20.
107. Schauer R (1992). Sialic acids regulate cellular and molecular recognition. In Ogura H, Hasegawa A, Suami T, Carbohydrates, Kodansha Ltd., Tokyo, pp. 340–354.
108. Jancik J, Schauer R (1974). Sialic acid–a determinant of the life-time of rabbit erythrocytes. *Hoppe Seylers Z Physiol. Chem.* 355:395–400.
109. Jancik J, Schauer R, Anders K H, et al. (1978). Sequestration of neuraminidase-treated erythrocytes. *Cell Tiss. Res.* 186:209–226.
110. Ourth D D, Bachinski L M (1987). Bacterial sialic acid modulates activation of the alternative complement pathway of channel catfish (Ictalurus punctatus). *Devel. Compar. Immunol.* 11:551–564.
111. Sakarya S, Oncu S (2003). Bacterial adhesins and the role of sialic acid in bacterial adhesion. *Med. Sci. Monitor: Internat. Med. J. Exper. Clin. Res.* 9:RA76–RA82.
112. Vimr E R, Kalivoda K A, Deszo E L, et al. (2004). Diversity of microbial sialic acid metabolism. *Microbiol. Mol. Biol. Rev.* 68:132–153.

113. Olivier J C, Taverna M, Vauthier C, et al. (1994). Capillary electrophoresis monitoring of the competitive adsorption of albumin onto the orosomucoid-coated polyisobutylcyanoacrylate nanoparticles. *Electrophoresis.* 15:234–239.
114. Huve P (1994). Comprendre et éviter la capture des nanoparticules de poly(acidelactique) par le système des phagocytes mononucléaires, PhD Dissertation, Université Paris Sud, Faculé de Pharmacie, Chatenay-Malabry, France, p. 193.
115. Yamauchi H, Yano T, Kato T, et al. (1995). Effects of sialic acid derivative on long circulation time and tumor concentration of liposomes. *Int. J. Pharm.* 113:141–148.
116. Kazatchkine M D, Fearon D T, Silbert J E, et al. (1979). Surface-associated heparin inhibits zymosan-induced activation of the human alternative complement pathway by augmenting the regulatory action of the control proteins on particle-bound C3b. *J. Exp. Med.* 150:1202–1215.
117. Keuren J F, Wielders S J, Willems G M (2003). Thrombogenicity of polysaccharide-coated surfaces. *Biomaterials.* 24:1917–1924.
118. Weber N, Wendel H P, Ziemer G (2002). Hemocompatibility of heparin-coated surfaces and the role of selective plasma protein adsorption. *Biomaterials.* 23:429–439.
119. Jeong S C, Yang B K, Ra K S (2004). Characteristics of anti-complementary biopolymer extracted from Coriolus versicolor. *Carbohyd. Polym.* 55:255–263.
120. Yamazaki N, Kojima S, Bovin N V, et al. (2000). Endogenous lectins as targets for drug delivery. *Adv. Drug Del. Rev.* 43:225–244.
121. Ehrhardt C, Kneuer C, Bakowsky U (2004). Selectins-an emerging target for drug delivery. *Adv. Drug Del. Rev.* 56:527–549.
122. Yamazaki N, Kojima S, Yokoyama H (2005). Biomedical nanotechnology for active drug delivery systems by applying sugar-chain molecular functions. *Curr. Appl. Phys.* 5:112–117.
123. Bies C, Lehr C-M, Woodley J F (2004). Lectin-mediated drug targeting: History and applications. *Adv. Drug Del. Rev.* 56:425–435.
124. Faivre V, Costa M d. L, Boullanger P, et al. (2003). Specific interaction of lectins with liposomes and monolayers bearing neoglycolipids. *Chem. and Phys. of Lipids.* 125:147–159.
125. Forssen E, Willis M (1998). Ligand-targeted liposomes. *Adv. Drug Del. Rev.* 29:249–271.
126. Minko T (2004). Drug targeting to the colon with lectins and neoglycoconjugates. *Adv. Drug Del. Rev.* 56:491–509.
127. Lehr C-M, Gabor F (2004). Lectins and glycoconjugates in drug delivery and targeting. *Adv. Drug Del. Rev.* 56:419–420.
128. Montisci M J, Giovannuci G, Duchene D, et al. (2001). Covalent coupling of asparagus pea and tomato lectins to poly(lactide) microspheres. *Int. J. Pharm.* 215:153–161.
129. Reddy J A, Allagadda V M, Leamon C P (2005). Targeting therapeutic and imaging agents to folate receptor positive tumors. *Curr. Pharm. Biotechnol.* 6:131–150.
130. Gao S, Chen J, Dong L, et al. (2005). Targeting delivery of oligonucleotide and plasmid DNA to hepatocyte via galactosylated chitosan vector. *European J. of Pharmaceutics and Biopharmaceutics.* 60:327–334.
131. Hubbard A L, Wilson G, Ashwell G, et al. (1979). An electron microscope autoradiographic study of the carbohydrate recognition systems in rat liver. Distribution of 125I-ligands among the liver cell types. *J. Cell. Biol.* 83:47–64.
132. Kishore U, Eggleton P, Reid K B (1997). Modular organization of carbohydrate recognition domains in animal lectins. *Matrix Biol.* 15:583–592.
133. Tanaka T, Fujishima Y, Hanano S, et al. (2004). Intracellular disposition of polysaccharides in rat liver parenchymal and nonparenchymal cells. *Int. J. Pharm.* 286:9–17.

134. Blondin C, Bataille L, Letourneur D (2000). Polysaccharides for vascular cell targeting. *Crit. Rev. Ther. Drug Carrier Syst.* 17:327–375.
135. Vercruysse K P, Prestwich G D (1998). Hyaluronate derivatives in drug delivery. *Crit. Rev. Ther. Drug Carrier Syst.* 15:513–555.
136. Sy M S, Guo Y J, Stamenkovic I (1991). Distinct effects of two CD44 isoforms on tumor growth in vivo. *J. Exp. Med.* 174:859–866.
137. Luo Y, Bernshaw N J, Lu Z-R, et al. (2002). Targeted Delivery of Doxorubicin by HPMA Copolymer-Hyaluronan Bioconjugates. *Pharm. Res.* 19:396–402.
138. Gunthert U, Hofmann M, Rudy W, et al. (1991). A new variant of glycoprotein CD44 confers metastatic potential to rat carcinoma cells. *Cell.* 65:13–24.
139. Entwistle J, Hall C L, Turley E A (1996). HA receptors: Regulators of signalling to the cytoskeleton. *J. Cell Biochem.* 61:569–577.
140. Danguy A, Camby I, Kiss R (2002). Galectins and cancer. *Biochim. Biophys. Acta.* 1572:285–293.
141. Julyan P J, Seymour L W, Ferry D R, et al. (1999). Preliminary clinical study of the distribution of HPMA copolymers bearing doxorubicin and galactosamine. *J. Control. Rel.* 57:281–290.
142. Jensen K D, Kopeckova P, Bridge J H, et al. (2001). The cytoplasmic escape and nuclear accumulation of endocytosed and microinjected HPMA copolymers and a basic kinetic study in Hep G2 cells. *AAPS PharmSci.* 3:E32.
143. Kim G-Y, Choi G-S, Lee S-H, et al. (2004). Acidic polysaccharide isolated from Phellinus linteus enhances through the up-regulation of nitric oxide and tumor necrosis factor-[alpha] from peritoneal macrophages. *J. Ethnopharm.* 95:69–76.
144. Zhang X Q, Wang X L, Zhang P C, et al. (2005). Galactosylated ternary DNA/polyphosphoramidate nanoparticles mediate high gene transfection efficiency in hepatocytes. *J. Control Rel.* 102:749–763.
145. Wu C, Wilson J, Wu G (1989). Targeting genes: delivery and persistent expression of a foreign gene driven by mammalian regulatory elements in vivo. *J. Biol. Chem.* 264:16985–16987.
146. Wilson J, Grossman M, Wu C, et al. (1992). Hepatocyte-directed gene transfer in vivo leads to transient improvement of hypercholesterolemia in low density lipoprotein receptor-deficient rabbits. *J. Biol. Chem.* 267:963–967.
147. Fajac I, Thévenot G, Bédouet L, et al. (2003). Uptake of plasmid/glycosylated polymer complexes and gene transfer efficiency in differentiated airway epithelial cells. *J. Gene Med.* 5:38–48.
148. Cade D, Ramus E, Rinaudo M, et al. (2004). Tailoring of bioresorbable polymers for elaboration of sugar-functionalized nanoparticles. *Biomacromol.* 5:922–927.
149. Racles C, Hamaide T (2005). Synthesis and Characterization of Water Soluble Saccharide Functionalized Polysiloxanes and Their Use as Polymer Surfactants for the Stabilization of Polycaprolactone Nanoparticles. *Macromol. Chem. and Phys.* 206:1757–1768.
150. Jeong Y-I, Seo S-J, Park I-K, et al. (2005). Cellular recognition of paclitaxel-loaded polymeric nanoparticles composed of poly([gamma]-benzyl l-glutamate) and poly(ethylene glycol) diblock copolymer endcapped with galactose moiety. *Int. J. Pharm.* 296:151–161.
151. Uchida T, Serizawa T, Ise H, et al. (2001). Graft copolymer having hydrophobic backbone and hydrophilic branches. 33. Interaction of hepatocytes and polystyrene nanospheres having lactose-immobilized hydrophilic polymers on their surfaces. *Biomacromol.* 2:1343–1346.

152. Sinha V R, Kumria R (2001). Polysaccharides in colon-specific drug delivery. *Int. J. Pharm.* 224:19–38.
153. Jozefowicz M, Jozefonvicz J (1997). Randomness and biospecificity: random copolymers are capable of biospecific molecular recognition in living systems. *Biomaterials.* 18:1633–1644.
154. Walker J B (2002). Enzymatic synthesis of aminoglycoside antibiotics: novel adenosylmethionine:2-deoxystreptamine N-methyltransferase activities in hygromycin B- and spectinomycin-producing Streptomyces spp. and uses of the methylated products. *Appl. Environ. Microbiol.* 68:2404–2410.
155. Macmillan D, Daines A M (2003). Recent developments in the synthesis and discovery of oligosaccharides and glycoconjugates for the treatment of disease. *Curr. Med. Chem.* 10:2733–2773.
156. Wong C H, Bryan M C, Nyffeler P T, et al. (2003). Synthesis of carbohydrate-based antibiotics. *Pure Appl. Chem.* 75:179–186.
157. Basu M K, Lala S (2004). Macrophage specific drug delivery in experimental leishmaniasis. *Curr. Mol. Med.* 4:681–689.
158. Donald P R, Sirgel F A, Venter A, et al. (2001). The early bactericidal activity of a low-clearance liposomal amikacin in pulmonary tuberculosis. *J. Antimicrob. Chemother.* 48:877–880.
159. Fielding R M, Lewis R O, Moon-McDermott L (1998). Altered tissue distribution and elimination of amikacin encapsulated in unilamellar, low-clearance liposomes (MiKasome). *Pharm. Res.* 15:1775–1781.
160. Fielding R M, Moon-McDermott L, Lewis R O, et al. (1999). Pharmacokinetics and urinary excretion of amikacin in low-clearance unilamellar liposomes after a single or repeated intravenous administration in the rhesus monkey. *Antimicrob. Agents Chemother.* 43:503–509.
161. Fattal E, Youssef M, Couvreur P, et al. (1989). Treatment of experimental salmonellosis in mice with ampicillin-bound nanoparticles. *Antimicrob. Agents Chemother.* 33:1540–1543.
162. Pinto-Alphandary H, Andremont A, Couvreur P (2000). Targeted delivery of antibiotics using liposomes and nanoparticles: research and applications. *Int. J. Antimicrob. Agents.* 13:155–168.
163. Schiffelers R, Storm G, Bakker-Woudenberg I (2001). Liposome-encapsulated aminoglycosides in pre-clinical and clinical studies. *J. Antimicrob. Chemother.* 48: 333–344.
164. Barratt G (2003). Colloidal drug carriers: achievements and perspectives. *Cell. Mol. Life Sci.* (CMLS). 60:21–37.
165. Fattal E, Balland O, Alphandary H, et al. (1993). Colloidal carriers of antibiotics as an alternative approach for the treatment of intracellular infections. In Raoult D (ed.), *Antimicrobial agents and intracellular pathogens*: CRC Press, Boca Raton, F.L., pp. 63–72.
166. Senel S, McClure S J (2004). Potential applications of chitosan in veterinary medicine. *Adv. Drug Del. Rev.* 56:1467–1480.
167. Rabea E I, Badawy M E-T, Stevens C V, et al. (2003). Chitosan as antimicrobial agent: applications and mode of action. *Biomacromol.* 4:1457–1465.
168. Zheng L Y, Zhu J F (2003). Study on antimicrobial activity of chitosan with different molecular weights. *Carbohyd. Polym.* 54:527–530.
169. Jeon Y J, Park P J, Kim S K (2001). Antimicrobial effect of chitooligosaccharides produced by bioreactor. *Carbohyd. Polym.* 44:71–76.

170. Liu H, Du Y, Wang X, et al. (2004). Chitosan kills bacteria through cell membrane damage. *Intern. J. Food Microbiol.* 95:147–155.

171. Avadi M R, Sadeghi A M M, Tahzibi A, et al. (2004). Diethylmethyl chitosan as an antimicrobial agent: Synthesis, characterization and antibacterial effects. *Euro. Polym. J.* 40:1355–1361.

172. Carlucci M J, Pujol C A, Ciancia M, et al. (1997). Antiherpetic and anticoagulant properties of carrageenans from the red seaweed Gigartina skottsbergii and their cyclized derivatives: correlation between structure and biological activity. *Int. J. Biol. Macromol.* 20:97–105.

173. Yoshida R L, Nakashima H, Premanathan M, et al. (2000). Specific biological activities of Chinese lacquer polysaccharides. *Carbohyd. Polym.* 43:47–54.

174. Hattori K, Yoshida T, Nakashima H, et al. (1998). Synthesis of sulfonated amino-polysaccharides having anti-VIH and blood anticoagulant activities. *Carbohyd. Res.* 312:1–8.

175. Seddiki N, Mbemba E, Letourneur D, et al. (1997). Antiviral activity of derivatized dextrans on HIV-1 infection of primary macrophages and blood lymphocytes. *Biochim. Biophys. Acta.* 1362:47–55.

176. Ghosh P, Adhikari U, Ghosal P K (2004). In vitro anti-herpetic activity of sulfated polysaccharide fractions from Caulerpa racemosa. *Phytochem.* 65:3151–3157.

177. Hobbs S K, Monsky W L, Yuan F, et al. (1998). Regulation of transport pathways in tumor vessels: role of tumor type and microenvironment. *Proc. Natl. Acad. Sci. USA*, 95:4607–4612.

178. Krishna R, Mayer L D (2000). Multidrug resistance (MDR) in cancer. Mechanisms, reversal using modulators of MDR and the role of MDR modulators in influencing the pharmacokinetics of anticancer drugs. *Eur. J. Pharm. Sci.* 11:265–283.

179. Jain K K (2005). Nanotechnology-based Drug Delivery for Cancer. *Technol. Cancer Res. Treat.* 4:407–416.

180. Jain R K (2001). Delivery of molecular medicine to solid tumors: lessons from in vivo imaging of gene expression and function. *J. Control. Rel.* 74:7–25.

181. Brigger I, Dubernet C, Couvreur P (2002). Nanoparticles in cancer therapy and diagnosis. *Adv. Drug Del. Rev.* 54:631–651.

182. Minko T, Dharap S S, Pakunlu R I, et al. (2004). Molecular targeting of drug delivery systems to cancer. *Curr. Drug Targets.* 5:389–406.

183. Couvreur P, Vauthier C. Nanotechnology in pharmacology. Submitted.

184. Qi L, Xu Z, Jiang X, et al. (2005). Cytotoxic activities of chitosan nanoparticles and copper-loaded nanoparticles. Bioorg. *Medic. Chem. Lett.* 15:1397–1399.

185. Berry D, Lynn D M, Sasisekharan R (2004). Poly(b-amino ester)s promote cellular uptake of heparin and cancer cell death. *Chem. Bio.* 11:487–498.

186. Linhardt R J (2004). Heparin-induced cancer cell death. *Chem. Bio.* 11:420–422.

187. Peng Y, Zhang L, Zeng F, et al. (2005). Structure and antitumor activities of the water-soluble polysaccharides from Ganoderma tsugae mycelium. *Carbohyd. Polym.* 59:385–392.

188. Zhang M, Zhang L, Cheung P C K, et al. (2004). Molecular weight and anti-tumor activity of the water-soluble polysaccharides isolated by hot water and ultrasonic treatment from the sclerotia and mycelia of Pleurotus tuber-regium. *Carbohyd. Polym.* 56:123–128.

189. Yoo S-H, Yoon E J, Cha J, et al. (2004). Antitumor activity of levan polysaccharides from selected microorganisms. *Internat. J. Biol. Macromol.* 34:37–41.

190. Ho J C K, Konerding M A, Gaumann A, et al. (2004). Fungal polysaccharopeptide inhibits tumor angiogenesis and tumor growth in mice. *Life Sci.* 75:1343–1356.
191. Cheng J-J, Huang N-K, Chang T-T, et al. (2005). Study for anti-angiogenic activities of polysaccharides isolated from Antrodia cinnamomea in endothelial cells. *Life Sci.* 76:3029–3042.
192. Peng Y, Zhang L, Zeng F, et al. (2003). Structure and antitumor activity of extracellular polysaccharides from mycellium. *Carbohyd. Polym.* 54:297–303.
193. Kabanov A V, Batrakova E V, Miller D W (2003). Pluronic block copolymers as modulators of drug efflux transporter activity in the blood-brain barrier. *Adv. Drug Del. Rev.* 55:151–164.
194. Saima Y, Sarkar K K, Sen Sr A K, et al. (2000). An antitumor pectic polysaccharide from Feronia limonia. *Int. J. Biol. Macromol.* 27:333–335.
195. Clowes A W, Karnowsky M J (1977). Supression by heparin of smooth muscle cell proliferation in injured arteries. *Nature.* 265:625–626.
196. Logeart D, Prigent-Richard S, Boisson-Vidal C, et al. (1997). Fucans, sulfated polysaccharides extracted from brown seaweeds, inhibit vascular smooth muscle cell porliferation. II. Degradation and molecular weight effect. *Euro. J. Cell Bio.* 74:385–390.
197. Logeart D, Prigent-Richard S, Jozefonvicz J, et al. (1997). Fucans, sulfated polysaccharides extracted from brown seaweeds, inhibit vascular smooth muscle cell porliferation. I. Comparison with heparin for antiproliferative activity, binding and internalization. *Euro. J. Cell Bio.* 74:376–384.
198. Mestries P, Borchiellini C, Barbaud C, et al. (1998). Chemicaly modified dextrans modulate expression of collagen phenotype by cultured smooth muscle cells in relation to the degree of carboxymethyl, benzylamide, and sulfation substitution. *J. Biomed. Mater. Res.* 42:286–294.
199. Chupa J M, Foster A M, Summer S R, et al. (2000). Vascular cell response to polysaccharide materials: in vitro and in vivo evaluations. *Biomaterials.* 21:2315–2322.
200. Sato N, Nakazawa F, Ito T (2003). The structure of the antigenic polysaccharide produced by Eubactrium saburreum T15. *Carbohyd. Res.* 338:923–930.
201. Fattom A I, Horwith G, Fuller S (2004). Development of StaphVAX(TM), a polysaccharide conjugate vaccine against S. aureus infection: From the lab bench to phase III clinical trials. *Vaccine.* 22:880–887.
202. Menzel M, Muellinger B, Weber N, et al. (2005). Inhalative vaccination with pneumococcal polysaccharide in healthy volunteers. *Vaccine.* 23:5113–5119.
203. Ruben F L, Froeschle J E, Meschievitz C, et al. (2001). Choosing a route of administration for quadrivalent menigococcal polysaccharide vaccine: intramuscular versus subcutaneous. *Clin. Infec. Dis.* 32:170–172.
204. Moe G R, Tan S, Granoff D M (1999). Molecular mimetics of polysaccharide epitopes as vaccine candidates for prevention of Neisseria meningitidis serogroup B disease. *FEMS Immunol. Med. Microbiol.* 26:209–226.
205. Westerfeld N, Zurbriggen R (2005). Peptides delivered by immunostimulating reconstituted influenza virosomes. *J. Pept. Sci.* 11:707–712.
206. Moser C, Metcalfe I C, Viret J F (2003). Virosomal adjuvanted antigen delivery systems. *Expert Rev. Vacc.* 2:189–196.
207. Garcea R L, Gissmann L (2004). Virus-like particles as vaccines and vessels for the delivery of small molecules. *Curr. Opin. Biotechnol.* 15:513–517.
208. Morein B, Hu K F, Abusugra I (2004). Current status and potential application of ISCOMs in veterinary medicine. *Adv. Drug Del. Rev.* 56:1367–1382.
209. Kreuter J (1995). Nanoparticles as adjuvants for vaccines. *Pharm. Biotechnol.* 6:463–472.

210. Koping-Hoggard M, Sanchez A, Alonso M J (2005). Nanoparticles as carriers for nasal vaccine delivery. *Expert Rev. Vacc.* 4:185–196.

211. Chauvierre C, Couvreur P, Labarre D, et al. (2002). Copolymères à structure séquencée composé d'un segment saccharidique lié à au moins un segment hydrophobe bioérodable et particules correspondantes, Patent n°WO 02/399-79.

212. Salek-Ardakani S, Arrand J R, Shaw D, et al. (2000). Heparin and heparan sulfate bind interleukin-10 and modulate its activity. *Blood.* 96:1879–1888.

213. Bland E J, Keshavarz T, Bucke C (2004). The influence of small oligosaccharides on the immune system. *Carbohyd. Res.* 339:1673–1678.

214. da Silva B P, Parente J P (2003). Bioactive polysaccharides from Costus spicatus. *Carbohyd. Polym.* 51:239–242.

215. Gan L, Zhang S H, Liu Q, et al. (2003). A polysaccahride-protein complex from Lyciul barbarum upregukates cytokine expression in human peripheral blood mononuclear cells. *Euro. J. Pharmac.* 471:217–222.

216. Zhao G, Kan J, Li Z, et al. (2005). Structural features and immunological activity of a polysaccharide from Dioscorea opposita Thunb roots. *Carbohyd. Polym.* 61:125–131.

217. Peluso G, Petillo O, Ranieri M, et al. (1994). Chitosan-mediated stimulation of macrophage function. *Biomaterials.* 15:1215–1220.

218. Han Y, Zhao L, Yu Z, et al. (2005). Role of mannose receptor in oligochitosan-mediated stimulation of macrophage function. *Int. Immunopharmacol.* 5:1533–1542.

219. Klokkevold P R, Fukayama H, Sung E C, et al. (1999). The effect of chitosan (poly-N-acetyl glucosamine) on lingual hemostasis in heparinized rabbits. *J. Oral Maxillofac. Surg.* 57:49–52.

220. Jordan R, Beeler D, Rosenberg R (1979). Fractionation of low molecular weight heparin species and their interaction with antithrombin. *J. Biolog. Chem.* 254:2902–2913.

221. Zeerleder S, Mauron T, Lämmle B, et al. (2002). Effect of low molecular weight dextran sulfate on coagulation and platelet function tests. *Thrombosis Res.* 105:441–446.

222. Mauzac M, Jozefonvicz J (1984). Anticoagulant activity of dextra derivatives. Part I: synthesis and characterization. *Biomaterials.* 5:301–304.

223. Magel-Din Hussein D, Helmy W A, Salem H M (1998). Biological activities of some galactomannans and their sulfated derivatives. *Phytochem.* 48:479–484.

224. Alban S, Schauerte A, Franz G (2002). Anticoagulant sulfated polysaccharides: Part I. Synthesis and structure-activity relationship of new pullulan sulfates. *Carbohyd. Polym.* 47:267–276.

225. Krentsel L, Chaubet F, Rebrov A, et al. (1997). Anticoagulant activity of functionalized dextrans. Structure analyses of carboxymethylated dextran and first Monte Carlo simulations. *Carbohyd. Polym.* 33:63–71.

226. Ronghua H, Yumin D, Jianhong Y (2003). Preparation and anticoagulant activity of carboxybutyrylated hydroxyethyl chitosan sulfates. *Carbohyd. Polym.* 51:431–438.

227. Siddhanta A K, Shanmugam M, Mody K H, et al. (1999). Sulphated polysaccharides of Codium dwarkense Boergs. from the west coast of India: Chemical composition and blood anticoagulant activity. *Int. J. Biol. Macromol.* 26:151–154.

228. Ross B P, Toth I (2005). Gastrointestinal absorption of heparin by lipidization or coadministration with penetration enhancers. *Curr. Drug Deliv.* 2:277–287.

229. Rajaonarivony M, Vauthier C, Couarraze G, et al. (1993). Development of a new drug carrier made from alginate. *J. Pharm. Sci.* 82:912–917.

230. Vila A, Sanchez A, Tobio M, et al. (2002). Design of biodegradable particles for protein delivery. *J. Control Rel.* 78:15–24.
231. Sakuma S, Hayashi M, Akashi M (2001). Design of nanoparticles composed of graft copolymers for oral peptide delivery. *Adv. Drug Del. Rev.* 47:21–37.
232. Alonso M J, Sanchez A (2003). The potential of chitosan in ocular drug delivery. *J. Pharm. Pharmacol.* 55:1451–1463.
233. Behrens I, Pena A I V, Alonso M J, et al. (2002). Comparative uptake studies of bioadhesive and non-bioadhesive nanoparticles in human intestinal cell lines and rats: the effect of mucus on particle adsorption and transport. *Pharm. Res.* 19:1185–1193.
234. Mao S, Germershaus O, Fischer D, et al. (2005). Uptake and Transport of PEG-Graft-Trimethyl-Chitosan Copolymer-Insulin Nanocomplexes by Epithelial Cells. *Pharmac. Res.*, In Press.
235. Fernandez-Urrusuno R, Calvo P, Remunan-Lopez C, et al. (1999). Enhancement of Nasal Absorption of Insulin Using Chitosan Nanoparticles. *Pharm. Res.* 16:1576–1581.
236. Borges O, Borchard G, Verhoef J C, et al. (2005). Preparation of coated nanoparticles for a new mucosal vaccine delivery system. *Int. J. Pharm.* 299:155–166.
237. Douglas K L, Tabrizian M (2005). Effect of experimental parameters on the formation of alginate-chitosan nanoparticles and evaluation of their potential application as DNA carrier. *J. Biomater. Sci. Polym. Ed.* 16:43–56.
238. Li X W, Lee D K L, Chan A S C, Alpar H O (2003). Sustained expression in mammalian cells with DNA complexed with chitosan nanoparticles. *Biochimica et Biophysica Acta (BBA)—Gene Struc. Expres.* 1630:7–18.
239. Peng J, Xing X, Wang K, et al. (2005). Influence of anions on the formation and properties of chitosan-DNA nanoparticles. *J. Nanosci. Nanotechnol.* 5:713–717.
240. Mansouri S, Cuie Y, Winnik F, et al. (2006). Characterization of folate-chitosan-DNA nanoparticles for gene therapy. *Biomaterials.* 27:2060–2065.
241. Ravi Kumar M N V, Samedt M, Mohapatra S S, et al. (2004). Cationic silica nanoparticles as gene carriers: synthesis, characterization and transfection efficiency in vitro and in vivo. *J. Nanosci. Nanotechnol.* 4:876–881.
242. Messai I, Lamalle D, Munier S, et al. (2005). Poly(D,L-lactic acid) and chitosan complexes: interactions with plasmid DNA. *Colloids and Surfaces A: Physicochem. Engin. Aspects.* 255:65–72.
243. Zimmer A (1999). Antisense Oligonucleotide Delivery with Polyhexylcyanoacrylate Nanoparticles as Carriers. *Methods.* 18:286–295.
244. Zobel H P, Kreuter J, Werner D, et al. (1997). Cationic polyhexylcyanoacrylate nanoparticles as carriers for antisense oligonucleotides. *Antisense & Nucleic Acid Drug Dev.* 7:483–493.
245. Aynie I, Vauthier C, Chacun H, et al. (1999). Spongelike alginate nanoparticles as a new potential system for the delivery of antisense oligonucleotides. *Antisense Nucleic Acid Drug Dev.* 9:301–312.
246. El-Aneed A (2004). An overview of current delivery systems in cancer gene therapy. *J. Control. Rel.* 94:1–14.
247. Riggs A (1981). Preparation of blood hemoglobins of vertebrates. *Methods Enzymol.* 76:5–29.
248. Chauvierre C, Marden M C, Vauthier C (2004). Heparin coated poly(alkylcyanoacrylate) nanoparticles coupled to hemoglobin: a new oxygen carrier. *Biomaterials.* 24:3081–3086.

# R&D PARADIGM SHIFT AND BILLION-DOLLAR BIOLOGICS

Krishan Maggon
*Pharma Biotech R&D Consultant, Geneva, Switzerland*

**Contents**

1 Introduction
2 Regulatory Approvals
3 Blockbuster Human Medicines
4 Biologics Gold Rush
5 R&D Overview
  5.1 R&D Productivity
6 R&D Success
  6.1 Erythropoietin
  6.2 Tumor Necrosis Factor Inhibitors
  6.3 Psoriasis
  6.4 Insulin
  6.5 Interferon
7 R&D Failures
  7.1 Alzheimer's Disease
  7.2 Anti-infective Exodus and U-turn
  7.3 RSV
  7.4 Sepsis
8 Cost Constraints
9 Public–Private Partnerships
10 Information Resources
11 Industry and Medicinal Brands in 2006
12 Conclusion
  Acknowledgment
  References

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

## 1 INTRODUCTION

There is no good measure of R&D productivity in the pharmaceutical and biotechnology industry because of the long maturation time needed to bring drugs to the market. The R&D investment has a long span of 7–15 years of payback time before drugs are approved. It takes even longer to market in major markets and turn the product into a profitable blockbuster [1–4]. Interim measures of R&D productivity like the number of patents filed and granted, scientific publications (variable impact factor), R&D expenses, and new drug approvals provide only partial measures of success. The real measure of R&D success is the market and profits from blockbuster sales after recovery of development costs. The success of a drug is based on its safety and efficacy established in clinical studies and favorable benefit-to-cost ratio and advantages over current therapy. A marketed product with limited sales and low growth, which barely recovers development costs, is an R&D failure. New projects with potential medicine fail due to lack of safety and/or efficacy. The major objective of the industrial R&D is to plan and execute the shortest development path to gain regulatory approval, strong patent position, and market the drug on a global basis in a short time frame and make it a blockbuster drug by expanding to related indications or new markets [4–8].

Company R&D productivity and efficiency can be measured by its introduction of new billion- and multibillion-dollar medicines through in-house R&D or extramural R&D, licensing, partnership, or marketing rights. Pharmaceutical and biotechnology R&D concentrates essentially on existing therapeutic areas of billion- and multibillion-dollar medicines and/or of first-in-class breakthrough products in new disease areas with double-digit growth. The R&D goal is to create, protect, and extend a company's franchise, to compete with market leaders with products with enhanced safety and efficacy or me-too follow-up for a breakthrough drug. Most current models measure past R&D performance, provide conflicting and variable results, and do not predict current or future outcome.

The modern blockbuster products era started four decades ago with Valium (Roche, diazepam), which was the most prescribed global medicine between 1969 and 1982 with 2.3 billion doses taken in 1978 and the first medicine to achieve peak sales of $200 million. It was later joined by benzodiazapine, antibiotics, and nonstroidal analgesics and by several innovative medicines like Captopril, Capoten, and Mevacor in the 1980–1990 period [9–13]. The period of billion-dollar pharmaceutical brands started 10 years ago and of biotechnology brands about 5 years ago. A commercial report estimated the total blockbuster medicine sales in 2005 at $145 billion and forecasted strong growth in biologics [14].

The profitability of the top pharmaceutical companies and their market value has increased over the years despite a decline in new drug approvals since 1990. Mergers and acquisitions, cost cutting, job elimination, and rationalization of operations achieved short-term profits. The higher R&D budgets failed to increase the number of new regulatory approvals. Moreover, patent expirations on existing blockbuster drugs, frequent and lengthy patent challenges, high prices of medications, drug supplies to poor countries and patients, product injury litigation, drug withdrawals, activists' pressure to shift R&D resources to tropical and neglected diseases, and tough new regulatory requirements for safer drugs and black box warnings hit the industry hard [15–19]. The price multiples or price earnings ratio of innovative pharma was flat or declined over the past 2 years and was converging

with that of generic companies like Forrest and Teva. It is ironic that the industry that took pride in development of new drugs to combat human diseases is now compared with big tobacco and big oil firms [1–4, 15–19].

Profits from billion-dollar (blockbuster) products invariably cover R&D and market failures, fund future R&D investments, pay financial incentives to executives and dividends to shareholders, and fund philanthropy and charity. Fast track regulatory approval, medical need, media coverage, and stories of responding patients contribute to the success of a new medicine. Reports of links to serious adverse reactions, high prices with questionable benefits or low response rates, and black box warnings or regulatory delays have negative effects on the marketing of a new or existing medicine.

The Gold Rush of biotechnology-derived blockbuster bands and the paradigm shift in innovation from pharmaceutical to biotechnology R&D, within biologics towards monoclonal antibodies and from Europe to the United States will be discussed. The successes and failures of biotechnology R&D, criteria for blockbuster brands, global sales of biologics blockbuster brands, changing regulations, cost constraints, public–private partnership to deliver and develop low-cost treatments for tropical diseases, and information resources for biologics will be discussed in this chapter.

## 2 REGULATORY APPROVALS

Regulatory approval in the United States, Europe, and Japan is the first step before marketing of a new/future blockbuster; this aspect is first covered before discussions of blockbuster drugs, global sales, and R&D performance.

Regulatory approval of a new medical product is based on the safety and efficacy of a treatment established in open-label and double-blind comparative clinical trials of an adequate number of patients. For life-saving medicines, the number of patients is low, whereas for chronic diseases, the number of patients may go up to 25,000 to define rare and unexpected toxicities or adverse drug reactions. The major markets are in the United States [Food and Drug Administration (FDA)], Europe [European Medicinal Evaluation Agency (EMEA)], and Japan, with common registration requirements for innovative and biotechnology products (International Conference of Harmonization, ICH). Regulatory approval is used to measure the success of a company R&D, but the product may fail to make an impact in the market, only generate modest sale leading to a loss, or barely recover costs of R&D. The R&D success of a company should only be linked to the profitable new blockbuster drugs introduced each year and to a sales increase of $1 billion for existing products.

Biotechnology companies contributed over 60% of the new approvals by the FDA in 2005. Despite the increasing R&D costs [2, 5, 6] and the increasing numbers of active investigational new drugs (INDs) (41% increase from 2001 to 2004), the number of new drug and new molecular entity and first-in-class FDA approvals from top pharmaceutical companies has declined (Table 1). Since 2001, over 40% of the approved products by top pharmaceutical companies were licensed from smaller biotechnology companies. FDA approvals for new drugs in 2005 included 32 products including Byetta (exenatide, Lilly), Levemir (insulin detemir,

**TABLE 1. FDA Drug Approvals 2001–2006**

| Categories | 2006 | 2005 | 2004 | 2003 | 2002 | 2001 |
|---|---|---|---|---|---|---|
| New drugs | 26 | 75 | 118 | 72 | 78 | 66 |
| New molecular entities (NMEs) | 18 | 20 | 38 | 21 | 17 | 24 |
| Priority review | 10 | 20 | 29 | 14 | 11 | 10 |
| Generics | NA | 361 | 384 | 263 | 321 | 234 |
| Biologics | 4 | 5 | 7 | 22 | 20 | 7 |
| Active INDs | 5445 | 5029 | 4827 | 4544 | 4158 | 3883 |

Data based on information available at www.fda.gov. The European Medicinal Evaluation Agency (EMEA) provides an annual review of drug approvals on its web site, *www.emea.eu.int*.

Novo Nordisk), Hydrase (hyaluronidase, Prima Pharm), Naglazyme (galsulfase, BioMarin), Copagus (peginterferon alpha-2a+ribavarin, Roche), Revlimid (lenalidomide, Celgene), Nexavar (sorafenib, Bayer, Onyx), and Orencia (abatacept, BMS). The FDA and the EMEA approved breakthrough medicines like Macugen (pegaptanib, Pfizer, OSI) for age-related macular degeneration or Avastin (bevacizumab, Roche, Genentech) and Erbitux (cetuximab, BMS, ImClone) for cancer in 2004 [4]. In 2005, several big pharma companies failed to gain approval for any compounds discovered in their laboratories. Most of the approved drugs were for rare diseases. FDA approved 26 new medicines in 2006 including Januvia (Sitagliptin, Merck) for type 2 diabetes, Sutent (sunitinib, Pfizer), Sprycel (desatinib, BMY), and Zolinga (vorinostat) for cancer. The 8 new biologics approved in 2006 by FDA included Rotateq (rotavirus vaccine, Merck) for gastroenteritis, Gardasil (human papillomas virus vaccine, Merck) for cervical cancer, Zostavax (Zoster vaccine live, Merck) for herpes Zoster in adults, Vectibix (penitumumab, Amgen) for colon cancer, Lucentis (ranibizumab, Genentech, Novartis) for age related wet macular degeneration, inhaled insulin Exubera (Pfizer), Myozyme (alglucosidase alpha, Genzyme) for Pompe disease, and Elaprase (idursulfase, Shire) for Hunter syndrome. The European regulatory agency EMEA approved 51 products in 2006, including first approvals for NCE like Accomplia (rimonabant, Sanofi Aventis) for obesity, Nexavar (sorafanib, Bayer-Schering) for renal cell carcinoma, Thelin (sitaxentan, Encysive) for pulmonary arterial hypertension, Baraclude, Sutent, Sprycel, and Tygacil. Biologics approved included 2 biosimilar human growth hormones, Atryn (r antithrombin alpha, Genzyme) for deep vein thrombosis, Byetta, Elaprase, Lucentis, and Tysabri. EMEA gave first approvals for the new vaccine Daronix (H5N1 whole inactivated antigen, GSK) for avian influenza pandemic and Proquad (measles, mumps, rubella, and varicella, Sanofi Aventis), in additional to Gardasil, Rotateq, and Zostavax. Rotarix (rotavirus vaccine, GSK) was approved in late 2005 and launched in early 2006.

In the past 5 years, at least six or seven products a year, each with peak sales potential of more than $1 billion annually, were terminated in late-stage phase III development or at the new drug application (NDA) filing stage, resulted in market value loss of tens of billion dollars for the affected companies. All big pharma companies like Merck, Lilly, BMS, GSK, and Novartis had their share of phase III failures, regulatory delays, NDA rejections, and requests for additional data [16, 17]. During the past 5 years, 15 drugs with peak sales potentials of a combined $11 billion a year were withdrawn from the market for safety reasons, some of which

had received fast-track priority approval. The resulting product injury litigation cost for just one of the withdrawn drugs, Vioxx, may reach $15 billion (Table 2) [20, 21]. Black box warnings, media coverage, and links to serious adverse effects resulted in significant sales decline for affected drugs in 2005 (Table 3). Patient safety and comfort was given priority by the black box warning for increased cardiovascular risks with COX II inhibitors and antidepressants link to suicidal behavior in children. The FDA rightly imposed the requirements for long-term safety

**TABLE 2. Major Drug Withdrawals**

| Drug | Company | Year | Adverse Effect | Patients Claims | Litigation $ |
|---|---|---|---|---|---|
| Benoxaprofen | Lilly | 1982 | Liver failure | 2,000 | $10 million |
| Redux/Pondimin (Fen-Phen) | Wyeth | 1997 | Heart valve problems<br>Pulmonary hypertension | 130 deaths<br>62,000 cases | 16 billion paid<br>5 billion reserve |
| Rezulin (Troglitazone) | Pfizer | 2000 | Liver failure | 400 deaths<br>4,000 cases | 15 billion |
| Baycol/Lipobay | Bayer | 2001 | Rhabdomyolysis<br>Kidney failure | 100 deaths<br>1,500 cases | 1 billion paid |
| Vioxx | Merck | 2004 | Heart attacks<br>Heart failure | 10,000 cases | 15–25 billion |

Two drugs initially withdrawn were allowed back by the FDA in U.S. markets. Lotronex (Alosetron, Glaxo Smith Kline) for irritable bowel syndrome was approved in early 2000 and withdrawn 9 months later due to serious gastrointestinal problems. It was reapproved with restrictions in 2002. Tysabri (Natalizumab, Biogen Idec, Elan) for multiple sclerosis was approved in late 2004 and withdrawn in 2005 due to rare fatal brain lesions and was allowed back in mid-2006.

**TABLE 3 Significant Sales Decline in 2005–2006**

| Generic Name | Brands | Companies | Indications | Sales $ billion | | |
|---|---|---|---|---|---|---|
| | | | | 2004 | 2005 | 2006 |
| Celecoxib* | Celebrex | Pfizer | Pain | 3.3 | 1.73 | 2.04 |
| Paroxitine* | Paxil | Glaxo Smith Kline | Depression | 3.9 | 2.2 | 1.15 |
| Azithromycin** | Zithromax | Pfizer | Antibiotic | 1.8 | 2.02 | 0.64 |
| Pravastatin** | Pravachol | BMS, Daiichi Sankyo | Cholesterol | 5.5 | 5.0 | 4.0 |
| Sertraline* | Zoloft | Pfizer | Depression | 3.4 | 3.25 | 2.11 |
| Simvastatin** | Zocor | Merck | Cholesterol | 5.2 | 4.4 | 2.8 |

Sales of top brands in 2005 were down due to safety concern about increased suicidal tendancies in younger children linked to Paxil and Zoloft and increased risk of cardiovascular events (heart attacks) due to COX II inhibitors and withdrawal of Vioxx and added black box warnings imposed by FDA. Medicinal brands with highest sales decrease in 2006 were mainly NCE synthetic products and the loss was due to patent expiry and generic competition.
*Black box warning **Patent Expiry and generics

data for new coxibs and carcinogenicity data before phase III trials for Peroxisome Proliferator activator receptor (PPAR) agonists [20–22].

The current regulatory requirements mandate extensive testing of all biological products even with small differences. Demonstrating that a biosimilar product is as safe and effective as the originator is a difficult task [18, 20]. Europe has taken the lead in biosimilar products by establishing a legal framework for authorization of biogenerics, and EMEA issued several final and draft guidelines in 2006 to cover erythropoietin, insulin, interferon, somatotropin, and granulocyte colony stimulating factor. Sandoz generic human growth hormone was approved in USA and Europe (EU) in early 2006 after a 3-year delay. The recent episode of an unexpected severe inflammation and multiple organ failure adverse reaction in six healthy subjects in phase I trials in England, with an immunomodulatory humanized agonistic anti-CD28 monoclonal antibody for leukemia, multiple sclerosis, and arthritis, and similar reactions in cancer patients in the United States with another anti CD28 antibody, after investigations, may result in new safety regulations for testing biologics and monoclonal antibodies.

The history of modern drug regulations is closely linked to the incidence of drug-induced injury, organ failure, and deaths. Each tragedy linked to marketed drugs resulted in increased regulations and additional testing to prevent future episodes with new drugs of the same class (Table 4). The Elixir of Sulfanilamide episode resulted in the Food, Drug and Cosmetic Act of 1938 and Thalidomide tragedy in early 1960 resulted in requirements for safety and efficacy testing for drugs in animals and humans.

The withdrawal of benoxaprofen resulted in a 10-fold increase in the number of patients (from 200 to 400 patients/year exposure to 2500 to 5000 patients/year exposure to the study drug) required for newer nonsteroidal anti-inflammatory drugs (NSAIDs) and trials in the elderly, hepatic and renal impairment, and drug interactions. Astra Zeneca total patient exposure in clinical trials submitted for approval of rosuvastatin (12,500) was considerably greater than the 2000–3000 patients submitted for most other marketed statins. Since the withdrawal of troglitazone (Rezulin), no other PPAR agonist has been approved because of the concerns about the class carcinogenicity, hepatotoxicity, and cardiotoxicity [19, 22]. Merck was the dominant R&D-driven and most admired company during the 1980s and 1990s and had one of the highest market valuations in the industry. Thus, a sale loss of $2.5 billion of rofecoxib resulted in a loss of $27 billion market value in one day and another $20 billion within the next month. Analysts have provided estimates of up to $20 billion for rofecoxib litigation. Similarly Pfizer lost $20 billion in one day in market value after discontinuation of its torcetrapib-Lipitor Phase III trials and termination of all trocetrapib (HDL promoter) development projects due to higher number of deaths in the treated group.

The cost of drug development has reached astronomic proportions. Considering the projections of previously contested estimates from 2005 and adding in marketing and manufacturing costs, a new drug for a chronic disease requires $2 billion in investment [2, 23, 24]. Industry critics have doubted earlier estimates, which included interest payments on blocked R&D funds and excluded R&D tax credits. A new GMP facility for a biological product costs an additional $500 million [2, 20]. With the new FDA safety focus, an average drug NDA file for registration for chronic diseases includes data on over 10,000–20,000 patients enrolled in 100–200

**TABLE 4. Regulatory Impact of Health Crisis/Drug Withdrawals**

| Drug | Year | Regulatory Action |
|---|---|---|
| Elixir Sulfanilamide | 1937 | Food Drug and Cosmetic Act 1938 |
| Thalidomide | 1962 | Safety and efficacy testing |
| Contraceptives Risks | 1970 | Package inserts |
| Benoxaprofen | 1982 | Studies in special populations, elderly, renal & hepatic insufficiency, genetic, gender differences |
| Rare diseases | 1983 | Orphan drug act |
| Life-threatening diseases | 1987 | Treatment IND for life saving drugs |
| Global regulations | 1990 | International Conference on Harmonization ICH 1-ICH 7(2006): GMP; GLP, GCP |
| High costs | 1992 | Generic drug act (1984 ANDA) |
| Drug safety | 1993 | MedWatch |
| AIDS | 1995 | Fast-track approval |
| Fen-Phen | 1997 | Orlistat NDA 5,000 patients, Safety for new obesity drugs |
| Rezulin | 2000 | Carcinogenicity prior to Phase III for PPAR agonists, CV safety for Murgaglitazar in 2005 |
| Tropical/neglected diseases | 2000 | Public–private partnerships: MMV, GAVI, |
| Baycol | 2001 | Safety for new statins, Rosuvastatin NDA 12,500 patients |
| Vaccines | 2002 | NDA >20,000–60,000 patients safety, Rotavirus |
| Vioxx | 2004 | CV safety in phase III for new COX II inhibitors<br>Etoricoxib & Lumiracoxib NDA >25,000 patients |
| Reduced new drug approvals | 2004 | Critical Path Initiative: 0 Phase Clinical trials, biomarkers, and clinical trial |
| Low R&D productivity | | design to speed up development |
| Black box warnings | 2005 | Reduced sales of affected medicines |
| Biogeneric guidelines | 2006 | Europe (EMEA), FDA 2008? |

NDA-New drug application; ANDA-Abbreviated new drug application; CV-cardiovascular; PPAR-Peroxisome Proliferator Activator Receptor; MMV-Medicines for Malaria Venture; GAVI-Global Alliance for Vaccines and Immunization; EMEA-European Medicinal Evaluation Agency

clinical studies over a 7–10-year period. The requirements for life-threatening diseases require only a few thousand patients. The development of a new drug typically takes 7–12 years, and clinical trials for a new vaccine require 25,000–100,000 individuals enrolled in a 2 : 1 ratio in the treatment and placebo groups. The recently approved rotavirus vaccines [Glaxo Smith Kline, Rotarix (in Europe) and Merck, Rotateq (in the USA and Europe)] to protect children from severe gastroenteritis were studied in 60,000–90,000 patients. These two NDAs established a historic record number of patients monitored for safety. With such progression both in the duration of treatment and in the number of patients, new drug development is going to be even more costly (Table 4).

The FDA's current twin objectives are fast-track approvals of life-saving drugs and long-term safety and efficacy data for chronic conditions. The FDA has taken several initiatives to speed up drug development by introducing 0 phase trials at low doses, improving clinical trials study design and finding validated biomarkers. The objective is to improve R&D productivity [4]. Fast-track approvals contribute greatly to the success of the breakthrough drugs or drugs with significant therapeutic advantage and bring along media interest and coverage of the medical meetings, expert opinion, and raised demand, hope, and awareness of health-care professionals and patients for the new medicine. The FDA and EMEA fast-track approvals have contributed to the blockbuster sales of several new anticancer and antiviral products for AIDS and RSV and new vaccines. The first-year sales for Neulasta, Pegasys, and Avastin were $1300, $760, and $550 million, respectively, due to favorable regulatory and media buzz. Flu Mist (MedImmune) failed to benefit from the U.S. shortage of flu vaccines in 2004, and despite the surge in demand due to Avian Influenza in 2005, because the FDA had limited its use in adults (who do not require flu vaccination) and excluded its use in the elderly and infants.

## 3 BLOCKBUSTER HUMAN MEDICINES

The "Gold Rush" of billion- and multibillion-dollar human medicine brands started in 1995 for the pharmaceuticals and in 2000 for the biotechnology-derived brands and will continue to expand in the future. Despite the proclaimed and much heralded "End of the Blockbuster Model" for the pharmaceutical industry, the number of blockbuster drugs, their market share, and contribution to profits for companies has steadily increased (Table 5).

**TABLE 5 Number of Blockbuster Drugs and Market Share**

| | | $ billion | |
|---|---|---|---|
| | **2005** | | **2006*** |
| Global Pharma Market | 603 | IMS | 640 |
| Global Biotech Market | 63 | Ernst & Young | 75 |
| Global Generic Market | 60 | IMS | 65 |
| R&D Expenses | 50 | PhRMA | 55 |
| Biotech R&D | 20 | Ernst & Young | 25 |

| Year | Billion dollar drug brands | | Total share of global market % |
|---|---|---|---|
| | NCE | NME | |
| 1991 | 4 | 0 | 6 |
| 1998 | 29 | 3 | 16 |
| 2000 | 55 | 5 | 25 |
| 2003 | 65 | 13 | 32 |
| 2004 | 79 | 18 | 38 |
| 2005 | 109 | 29 | 42 |
| 2006 | 125 | 35 | 45 |

NCE-New Chemical Entity (Pharmaceuticals); NME-New Molecular Entity (Biologics)
*Global 2006 market estimates by author.

**BOX 1**

The success criteria for blockbuster biologics are as follows:

- Life saving, reduction in hospitalization days, critical care, and life support systems
- Improve, maintain, restore, or reverse declining vital/organ functions
- Complete or partial disease remission
- Lifestyle changes, reverse or slow aging process, obesity, diabetes, and hypertension
- High safety and efficacy profile
- Long-term benefits
- Fast-track regulatory approvals
- Strong patent protection and repeated extension for new indications
- Global sales, marketing, and manufacture to meet any surge in demand

From 1991 to 2006, the number of billion-dollar medicines increased 30-fold and their market share by 7-fold and accelerated for biologics brands in 2004 and 2005. The share of biotech sales of the total pharma sales increased in the United States from 4% in 1993 to 15% in 2005. The total number of blockbuster medicinal brands increased to 125 in 2006 and included 35 biologic brands with global sales of biologics reaching $75 billion (Table 5).

The criteria for blockbuster drugs and diseases with unmet medical need and market success to $5 billion, $10 billion, and $15 billion are as described (Box 1) [15].

The first effective and safe treatments for Alzheimer's, Parkinson's, chronic graft rejection, permanent graft acceptance, obesity, chronic obstructive pulmonary disease (COPD), adult respiratory distress syndrome (ARDS), sepsis, and vaccines [common cold, respiratory syncytial virus (RSV), severe acute respiratory syndrome (SARS), AIDS, malaria, Avian Influenza, radiation, contraception, emerging new infections] will be multibillion-dollar products [15, 16].

Table 6 lists the R&D budgets, total sales, % contribution of blockbuster brand sales to human pharmaceutical sales and the list of blockbuster brands by different companies. Only companies with either R&D budgets of over $1 billion or with at least one blockbuster medicine in 2006 are included. The IMS sales figures for the same brand often differed from company figures by more than $500 million for several drugs. The contribution of blockbuster brand sales for top pharmaceutical companies ranged from 37% (Roche) to 74% (Johnson & Johnson). For biotechnology companies, the contribution of a blockbuster brand was even greater and ranged from 66% (Serono) to 98% (Amgen). The contribution of these drugs to the profitability and the market value of the company was even greater. Profits from billion-dollar (blockbuster) products cover R&D and marketing failures in other projects, fund future R&D investments, increase dividend payments to shareholders, fund philanthropy and charity, provide access to medicines for needy patients, and bonuses and motivation to executives. This aspect has not been included in R&D productivity evaluation by commercial reports.

**TABLE 6. Multibillion-Dollar Brand Medicines in 2006**

| Company | Sales $ Billion (% Blockbuster Sales) | | R&D Budget $ Billion | | >2 Billion | >$ 1 billion |
|---|---|---|---|---|---|---|
| | 2005 | 2006 | 2005 | 2006 | | |
| J&J | 50(74) | 53.3 | 6.3 | 7.1 | Remicade Risperdal **Procrit** ***Topamax*** | Aciphex, Duragesic, Floxin, |
| Pfizer | 51(55) | 48(55) | 7.4 | 7.6 | Lipitor, Norvasc, Zoloft, *Celebrex* ***Zithromax*** | Viagra, Xalatan, Zyrtec, *Lyrica* |
| Glaxo Smith Kline | 34(70) | 43 | 5.7 | 6.4 | Advair, Avandia ***Vaccines*** | Augmentin, Coreg, Flovent, Imigran, Lamictal, Paxil, Valtrex, Wellbutrin, Zofran Combivir |
| Novartis | 32(45) | 37 | 4.8 | 5.3 | Diovan, Glivec | Lotrel, Zometa Neoral, Lamisil |
| Sanofi Aventis | 32(49) | 36.8 | 4.8 | 5.76 | Plavix, **Lovenox** *Ambien Aprovel, Eloxatine,* ***Lantus***, *Taxotere* ***Vaccines*** | Copaxone, Tritace, Allegra, ***Fluzone*** |
| Roche | 28(37) | 33.6 | 4.6 | 5.2 | ***Avastin***, *Tamiflu,* ***Herceptin*****, Rituxan** | Cell Cept, **NeoRecormon**, **Pegasys** |
| Astra Zeneca | 24(73) | 26.5 | 3.4 | 3.9 | Nexium, Seroquel, *Crestor* | Arimidex, *Atacand*, Casodex, Prilosec, Pulmicort, Symbicort, Seloken, Zoladex |
| Merck | 22(61) | 22.6 | 3.8 | 4.8 | Fosamax, Cozaar, Singulair, Zocor | *Vytorin, Zetia,* ***Vaccines*** |
| Wyeth | 18(57) | 20.4 | 2.7 | 3.1 | Effexor, **Enbrel** | **Prevenar**, Protonix; *Premarin* |
| Abbott | 22(71) | 22.5 | 1.8 | 2.2 | Prevacid ***Humira*** | Depakote, Kaltera, Mobic *TriCor*, Biaxin |
| Bristol Myers Squibb | 19(40) | 18 | 2.7 | 3.0 | Plavix, Pravachol *Avapro* | *Abilify,* ***Erbitux*** *Pravachol* |
| Bayer Schering | 10.4 | 15.2 | 2.2 | 2.8 | | **Betaferon,** ***Kogenate,*** *Yasmin* |

| | | | | | | |
|---|---|---|---|---|---|---|
| Lilly | 14(52) | 15.6 | 2.7 | 3.1 | Zyprexa | Evista, Zemzar, **Humalog, Humulin**, *Cymbalta* |
| Amgen | 12.4(98) | 14.3 | 2.3 | 3.3 | **Aranesp, Epogen, Enbrel, Neulasta** | **Neupogen** |
| Boehringer Ingelheim | 12.3 | 13.8 | 1.5 | 2.0 | *Spiriva* | Flomax, Micardis, Mobic |
| Takeda | 10 | 10.5 | 1.3 | 1.5 | Prevacid, *Actos, Blopress* | **Lupron** |
| Schering Plough | 9 | 10.6 | 1.8 | 2.2 | | *Zetia, Vytorin,* Remicade |
| Baxter | 10 | 10.4 | 0.5 | 0.6 | | **Advate rAHF-PFM (Factor VIII)** |
| Merck-Serono | 6.9 | 9.84 | 1.2 | 1.36 | | **Rebif, *Erbitux*** |
| Astellas | 8 | 9 | 1.2 | 1.3 | | Prograf, Harnal |
| Teva | 4.7 | 8.4 | 0.37 | 0.46 | *Copaxone* | |
| DaiichiSankyo | 8 | 8 | 1.4 | 1.45 | *Benicar* | Levofloxacin, Mevalotin, Pravastatin |
| Novo Nordisk | 5 | 7.2 | 0.6 | 1.0 | **Insulins** | |
| Eisai | 5 | 5.7 | 0.83 | 0.92 | | Aricept, Aciphex |
| Nycomed Atlanta | 4 | 4.4 | 0.6 | 0.65 | Protonix | |
| Otsuka | 3.0 | 3.7 | 0.49 | 0.5 | Abilify | |
| Allergan | 2.3 | 3.45 | 0.39 | 1.0 | | ***Botox*** |
| UCB Schwarz | 3.0 | 3.3 | 0.66 | 0.78 | | *Keppra* |
| Genzyme | 2 | 3.2 | | | | ***Cerezyme*** |
| Gilead | 2.03 | 3.03 | 0.28 | 0.38 | | *Truvada* |
| Forest | 3.5 | 2.96 | 0.3 | 0.4 | | Lexapro |
| Biogen Idec | 2.4 | 2.7 | 0.74 | 0.72 | | **Avonex** |
| MedImmune | 1.2 | 1.3 | 0.4 | 0.42 | | **Synagis** |
| Purdue | NA | NA | NA | NA | | OxyContin |

NA-Not Available

Companies with one blockbuster brand or R&D budget of >$1 billion are listed. Marketing joint ventures like Takeda-Abbott Pharmaceutical, Merck-Schering Plough are not listed.

The % contribution of blockbuster human medicinal brands to the total human pharmaceutical 2005 sales of the company which increased even more in 2006. Biologics are in **bold**, new blockbuster drugs are in *italics*, brands going off blockbuster status in 2006 are underlined.

In 2006, there was 1 brand with over $12 billion dollar, 3 with over $5 billion, 9 with over $4 billion and additional 8 brands crossed $3 billion annual sales (Table 7). Although Merck has only four blockbuster drugs, all four had sales over $3 billion each, which is equal to 12 blockbuster drugs. Lipitor alone with sales of over $12 billion is equivalent to 12 billion-dollar drugs. Table 7 lists all new chemical entities with annual sales over $3 billion in 2006. Lipitor, Plavix, and Advair each had explosive sales growth during the 2002–2005 periods. An increase of over $1 billion a year was recorded with Lipitor during 2000–2005 and for Plavix during the 2001–2005 periods. For Pfizer and Sanofi Aventis, it is like adding a new blockbuster drug each year during the above period.

The U.S. patents on 35 drugs with global sales totaling more than $22 billion expired in the year 2006, resulting in the loss of market to generics. Blockbuster products coming off patent are valued at $27 billion in 2007 and $29 billion in 2008 [17]. Blockbuster brands nearing patent expiration in therapeutic areas like ulcers, hypertension, lipid lowering, depression, schizophrenia, and cancer are replaced by new molecules under patent cover with improved safety and efficacy. Thus, Esmoprazole and two other proton pump inhibitors have replaced Prilosec (omeprazole). Similarly the angiotensin converting enzyme inhibitors class has been taken over by angiotensin receptors blockers with four brands in the multibillion-dollar sales category.

**TABLE 7. Top Brand Medicines with Over $ 3 Billion Sales in 2006**

| Generic Name | Brands | Companies | Indications | Sales $ billion | | |
|---|---|---|---|---|---|---|
| | | | | 2004 | 2005 | 2006 |
| Atorvastatin | Lipitor | Pfizer | Cholesterol | 10.8 | 12.2 | 12.9 |
| Fluticasone Salmetrol | Advair | Glaxo Smith Kline | Asthma | 4.5 | 5.5 | 6.13 |
| Clopidrogel | Plavix | Bristol Myers Squibb, Sanofi Aventis | Atherosclerosis | 5.2 | 6.2 | 5.55 |
| Esomaprazole | Nexium | AstraZeneca | Ulcers | 3.88 | 4.63 | 5.2 |
| Amlodipine | Norvasc | Pfizer | Hypertension | 4.46 | 4.76 | 4.85 |
| Glanzapine | Zyprexa | Lilly | Schizophrenia | 4.42 | 4.2 | 4.36 |
| Valsartan | Diovan | Novartis | Hypertension | 3.1 | 3.67 | 4.22 |
| Risperidone | Risperdal | J&J | Schizophrenia | 3.0 | 3.55 | 4.18 |
| Pravastatin | Pravachol | BMS, Daiichi Sankyo | Cholesterol | 5.7 | 5.0 | 4.0 |
| Lansoprazole | Prevacid | Takeda, Abbott | Ulcers | 3.1 | 3.8 | 3.8 |
| Venlafaxine | Effexor | Wyeth | Depression | 3.3 | 3.5 | 3.7 |
| Montelukast | Singulair | Merck | Asthma | 2.6 | 3.0 | 3.6 |
| Glatiramer | Copaxone | Teva, Sanofi Aventis | Multiple Sclerosis | 1.8 | 2.4 | 3.6 |
| Pentoprazole | Protonix | Atlanta, Wyeth | Ulcers | 3.3 | 4.0 | 3.4 |
| Losartan | Cozaar | Merck | Hypertension | 2.8 | 3.0 | 3.2 |
| Alendronate | Fosamax | Merck | Osteoporosis | 3.1 | 3.2 | 3.1 |

Sales as reported by companies.
Zocor which was 5th in 2005 failed to make the list in 2006.

New-patented products in the same class replaced blockbuster anticancer products in the class taxol, camptothecin, and platinum after patent expiry. In biotechnology-derived biologics, second- and third-generation products with improved properties, like improved insulin, erythropoietin, and peg-interferon, replace the first-generation products. Some companies failed to switch to new patent-covered formulations or analoges like Lilly with Prozac and BMS with Glucophage (no follow-up), and Schering Plough from Claritin to Clarinex. The pharmaceutical industry is very creative in producing me-too blockbuster drugs with extended patent life. There is no reason to believe that this trend will stop in the next 5 years due to looming patent expiration of several blockbuster drugs as some commercial reports have suggested.

## 4 BIOLOGICS GOLD RUSH

IMS [12] data estimated the global pharmaceutical market at $600 billion and biotechnology products at $52 billion in 2005. Ernst and Young [1] estimated the total biotechnology product sales at $54.6 billion in 2004 and $63 billion in 2005. Datamonitor estimated the 2005 global biotechnology market at only $40 billion. All sales forecasts are on the low side as the total sales of the listed biotechnology products in Tables 8 and 9 alone were $60.4 billion in 2005 and $71.25 billion in 2006. There are 300 approved biotech products in the United States and Europe, 100 under regulatory review, and another 400 in advanced clinical trials for 200 diseases [1].

Several brands of EPO, TNF inhibitors, interferon, insulin, and GCSF had sales of over $1, $2, and $3 billion each in 2004 and 2005. The period 2002 to 2005 recorded a sales increase of over $1 billion per year for TNF inhibitors. The top five best-selling human medications were all biotechnology drugs by sales in 2005, apart from Lipitor. There were five biologic brands with over $3 billion and six brands with over $2 billion in sales in 2005. As several companies market erythropoietins, interferons, insulins, GCSF, and human growth hormone under different brand names, these proteins rarely make bestsellers lists in commercial databases, unless taken together (Tables 8, 9) [22, 25].

A review of several therapeutic areas and brands indicates intense market competition for market leadership and change of market leadership among different brands and fast growth areas (Tables 8, 9). This has been previously observed with me-too or follow-on pharmaceuticals. Losec overtook Zantac, which had overtaken the first-in-class Tagamet; Enalapril over Captotril (Capoten); and Lipitor over Zocor and the first-in-class Mevacor. Lipitor has retained market dominance despite the arrival of superstatin Crestor and newer combination products. Diovan overtook the first-in-class Cozaar in 2004, for the market leadership of the angiotensin II receptor blocker ARB antihypertensive agents [22]. Astra Zeneca retained its dominance of the antiulcer market by successful switching patients to Nexium and limiting the decline in Losec sales after patent expiration. The success of Atlanta to launch a blockbuster antiulcer Protonix shows that follow-on patented me-too drugs carry low risk and high rewards (Table 7). Monoclonal antibodies for cancer and immunoinflammatory conditions had global sales of $20.5 billion in 2006 with tumor necrosis factor blockers accounting for $10.6 billion. Vaccines

**TABLE 8. Best Selling Biologics (Top 1–5)**

| Protein | Indication | Total sales $ Billion | | | Company | Brands | Sales $ Billion | Sales $ Billion | |
|---|---|---|---|---|---|---|---|---|---|
| | | 2004 | 2005 | 2006 | | | 2004 | 2005 | 2006 |
| Erythropoietin PEG, α, β, Darbepoetin | Anemia | 11.8 | 12.3 | 12.1 | Amgen | Epogen | 2.6 | 2.45 | 2.51 |
| | | | | | | Epogin | | | |
| | | | | | | Aranesp | 2.4 | 3.3 | 4.12 |
| | | | | | J&J | Procrit | 3.6 | 3.3 | 3.2 |
| | | | | | | Eprex | | | |
| | | | | | Roche | NeoRecormon | 1.8 | 1.7 | 1.8 |
| | | | | | | Epogin | | | |
| | | | | | Kirin | ESPO | 0.50 | 0.55 | 0.50 |
| Tumor Necrosis Factor Inhibitors | RA, JRA, Ps, PsA, AS | 5.55 | 8.6 | 10.6 | Amgen Wyeth | Enbrel | 2.6 | 3.7 | 4.4 |
| | RA, UC, CD, Ps, PsA, AS | | | | J&J Schering Plough | Remicade | 2.1 | 3.5 | 4.2 |
| | RA, AS, PsA, CD | | | | Abbott | Humira | 0.85 | 1.4 | 2.0 |
| Human Insulin | Diabetes | 6.5 | 7.2 | 9.05 | Novo | Novolin | 2.1 | 2.4 | 2.5 |
| | | | | | Nordisk | Analogs | 0.9 | 1.2 | 1.7 |
| | | | | | Lilly | Humulin | 1.0 | 1.12 | 0.92 |
| | | | | | | Humalog | 1.1 | 1.0 | 1.3 |
| | | | | | | Byetta | | 0.43 | |
| | | | | | SanofiAventis | Lantus | 0.96 | 1.4 | 2.2 |
| | | | | | | Apidra | | | |
| Interferons α, β | Hepatitis C | 6.8 | 6.8 | 6.7 | Schering Plough | PegIntron | 1.8 | 1.4 | 1.04 |
| | | | | | Roche | Pegasys | 1.4 | 1.4 | 1.2 |
| | Multiple sclerosis | | | | Biogen Idec | Avonex | 1.4 | 1.5 | 1.7 |
| | | | | | Merck Serono | Rebif | 1.1 | 1.3 | 1.45 |
| | | | | | Bayer Schering | Betaseron | 1.0 | 1.2 | 1.3 |
| Granulocytes-Colony Stimulating Factor, G-CSF | Granulocytes stimulator | 2.9 | 3.5 | 3.9 | Amgen | Neulasta | 1.7 | 2.3 | 2.7 |
| | | | | | | Neupogen | 1.2 | 1.2 | 1.2 |

RA-Rheumatoid Arthritis; JRA-Juvenile Rheumatoid Arthritis; Ps-Psoriasis; PsA-Psoriatic arthritis; CD-Crohn's Disease; UC-Ulcerative Colitis; AS-Ankylosing Spondylitis.

**TABLE 9. Best Selling Biologics (6–20)**

| Generic Name | Brands | Companies | Indications | Sales $ billion | | |
|---|---|---|---|---|---|---|
| | | | | 2004 | 2005 | 2006 |
| Rituximab | Rituxan | Roche, Biogen Idec | Leukemia, Lymphoma Rheumatoid Arthritis | 2.8 | 3.2 | 4.7 |
| Trastuzumab | Herceptin | Roche | Breast Cancer | 1.3 | 1.65 | 3.14 |
| Factor VIII | Advate, Helixate, Kogenate, ReFacto Recombinate | Baxter, Bayer Schering Sanofi Aventis Wyeth | Hemophilia | 1.9 | 2.1 | 2.5 |
| Bevacizumab | Avastin | Genentech, Roche | Colon cancer | 0.55 | 1.3 | 2.4 |
| Human Growth Hormone | Serostim, Saizen, Humatrope, Protopin, Neutropin, Genotropin | Merck Serono, Biogen Idec, Roche, Novo Nordisk, Akzo Nobel, Lilly, Pfizer | Dwarfism | 1.8 | 2.2 | 2.4 |
| Influenza Vaccines | Fluvin, Flumist, Fluzone, Fluvirin, Fluarix, Influvac | Novartis, GSK, Merck, Sanofi Aventis, Crucell Berna, Tanabe | Influenza | 1.2 | 1.9 | 2.3 |
| Pneumococcal ConjugateVaccine | Prevanar | Wyeth | Pneumococcal Dis | 1.2 | 1.5 | 1.9 |
| Luprorelin | Lupron | Takeda, Abbott | Prostate Cancer | 1.2 | 1.5 | 1.7 |
| FSH | Gonal F, Follistim | Merck-Serono, Akzo Nobel | Infertility | 0.95 | 0.93 | 1.3 |
| Hepatitis Vaccines | | GSK, Merck, SanofiAventis | Hepatitis A, B | 1.0 | 1.1 | 1.2 |
| Pediatric Vaccines | | GSK, Sanofi Aventis Merck | Measles, Mumps, Rubella Chickenpox, tetanus, diphtheria, acellular pertussis | 1.0 | 1.1 | 1.2 |
| Palivizumab | Synagis | MedImmune | RSV | 0.95 | 1.1 | 1.1 |
| Botulin toxin | Botox | Allergan | Wrinkles | 0.70 | 0.83 | 1.1 |
| Glucocerebrosidase | Cerezyme, Ceradase | Genzyme | Gaucher's disease | 0.84 | 0.93 | 1.0 |
| Factor VII | Novo Seven | Novo Nordisk | Hemophilia | 0.76 | 0.81 | 0.95 |

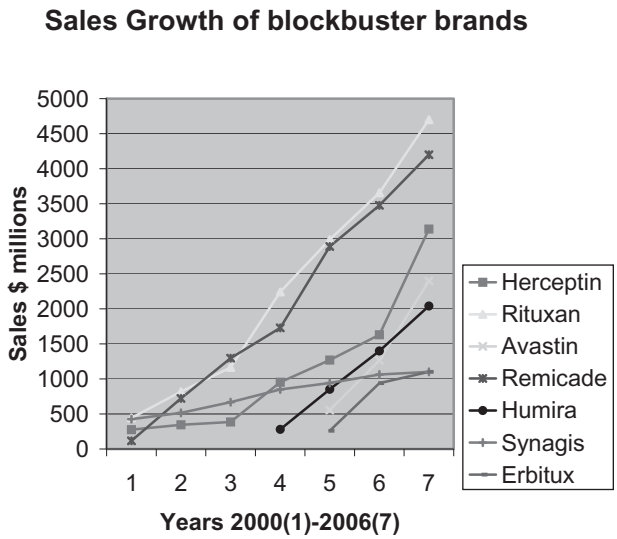


**Figure 1.** Blockbuster monoclonal antibodies sales growth.

were second with global sales reaching $15 billion. The other major categories of biologic brands with global sales were erythropoietins for anemia with $12 billion, traditional and newer insulins like long acting at $9 billion and interferon for hepatitis and multiple scelrosis at $7 billion.

Erythropoietin has remained the best-selling human medicine for the past 3 years, followed by Lipitor (Atorvastatin), and both of these had over $12 billion in sales [20–22]. The sales of erythropoietin in Japan were over $1 billion. Tumor necrosis factor inhibitors, interferons, and insulins with, respectively, sales of $8.6 billion, $7.3 billion, and $6.5 billion in 2005 were the third, fourth, and fifth best-selling products after EPO and Lipitor (Table 8) [20–22, 24, 25]. The time taken to reach $1 billion annual sales for monoclonal antibodies has varied and was much faster for follow up and cancer products (Fig. 1). Rituxan is the leading antibody since 2003 by sales. The growth of Avastin and Erbitux has been much faster as compared to Herceptin and took only two years to reach $1 billion sales.

## 5 R&D OVERVIEW

An overview of the blockbuster medications during the past few years and a comparison of R&D productivity clearly show an ongoing paradigm shift of innovation from Europe to the United States and from pharmaceutical to biotechnology R&D. A comparative analysis of R&D productivity of pharmaceutical and biotechnology firms is provided.

Pharmaprojects lists over 7300 drugs in active R&D, whereas the PHRMA site using Kluwer Health Adis R&D Insight lists over 10,000 active R&D projects. The

number of new drugs in R&D has almost doubled to 1970 drugs in 2004 from 1010 drugs in 1995 [1]. There are over 300 approved biotech products in Europe and the United States, 100 under regulatory review, 200 in late phase II–III clinical trials, and 800 in early clinical trials [1, 16, 20]. Monoclonal antibodies outnumber all other biologics combined in all stages of preclinical and clinical development. The biotechnology industry spent over $23 billion on R&D and generated over 7000 patents in the year 2005, with patents in cancer leading over other therapeutic categories. The highest number of patents was issued for cancer, and the success of cancer vaccines is attracting a lot of R&D funds. Only a strong R&D pipeline contributes to long-term growth of the company.

Several pharmaceutical companies spent over $3 billion on R&D in 2005 for a portfolio of 80–150 projects, of which 50–80 were in clinical development. In general, only about 30% NCE are truly innovative, the rest are me-too products or line extensions. About 30–40% of the projects of big pharma are licensed in from other sources [2, 3]. The share of the top 10 global R&D companies in the innovative first-in-class type of new molecular entities has steadily decreased over the past decade. It now takes 10–15 years to bring a product from discovery to markets. The success rate has decreased over the years, and out of a million compounds screened in highly automated computerized systems High Throughput Screening (HTS), only 250 advanced to become development candidates for additional testing [16].

All major pharmaceutical companies run their own discovery, development, licensing, partnership, manufacturing, marketing, and sales for the majority of their product pipelines and portfolios with a relatively high risk and investment [1, 3]. Strategic review, analysis, and consulting firms have prompted big pharmaceutical firms to shift focus to targeting specific populations or disease areas like the biotechnology companies and shift R&D resources to vaccines, biologics, and monoclonal antibodies that have higher rates of success, form partnerships in all stages of drug discovery and development to reduce risks and share rewards and focus on disease and needs of the patients, and operate as strategic business units [5, 9].

There has been a paradigm shift in innovation from the pharmaceutical to biotechnology R&D based on the number of new billion-dollar biologics introduced. There is a clear indication from Table 6 that biotechnology companies with smaller R&D budgets were more efficient in creating new blockbuster brands. [10, 11, 14, 16]. With the large-scale production of antibiotics, vaccines, and insulin during the post-World War II period, the pharmaceutical industry acquired and developed skills to manufacture, market, and supply medicines worldwide and retains its edge over biotechnology companies. An earlier paradigm shift from academic to industrial R&D happened in the 1970–1980 period. Pharmaceutical firms' R&D laboratory scientists won Nobel prizes and made important discoveries leading to new drug classes like ACE inhibitors, proton pump inhibitors, NSAIDs, and immunosuppressive agents for organ transplantation. The recent paradigm shift during the past 5 years has been the transfer of innovative R&D from Europe to the United States in parallel with the shift from pharmaceutical to the biotechnology R&D [25–28]. The strength of the biotechnology industry, university industry relationship, groundbreaking discoveries, National Institutes of Health (NIH), absence of price controls, and rapid uptake of high-price new treatments to avoid potential

litigation from patients have attracted R&D investment from European and Japanese companies.

### 5.1 R&D Productivity

There was no clear-cut relationship between a higher R&D budget and creation of new blockbuster medicines. If one only goes by the number of blockbuster medicines, the top four companies did not increase the number of such drugs in the year 2005 from 2004. Several companies like Pfizer, Merck, BMS, and Wyeth had decreased numbers, whereas Sanofi Aventis, Astra, Roche, and Amgen had increased numbers. The international R&D expansion activities, research capabilities, and patent output of 65 Japanese pharmaceutical firms from 1980 to 1991 were studied. It was observed that firms benefited from international R&D only when they had in-house existing research capabilities in the underlying technologies [29]. Another study of the Japanese pharmaceutical industry innovative capabilities from 1975 to 1995 found that the unattractive home market pushed these firms into trivial innovations [30]. The global expansion of the Japanese companies and their creation of blockbuster drugs do not support the above conclusion. German majors like Bayer, Schering, and mostly generic Merck failed to create any new blockbuster medicine during the past few years, indicating the decline of the German industry from global to regional levels. Boehringer Ingelheim and mid-sized Atlanta were able to add new billion dollar molecules. The German industry moved strongly in 2006 towards consolidation and mergers. Bayer purchased Schering and Merck made a move into biotechnology by acquiring Swiss Serono. Atlanta was overtaken by Danish Nycomed and another mid-size, Schwartz, was absorbed by Belgian UCB. These moves reflect a desire to consolidate marketing and R&D.

Drews [31, 32] raised concerns about the low R&D productivity gap despite the mergers in the early and late 1990s. These mergers in fact resulted in reduced R&D productivity to produce new approvals and blockbusters. Glaxo Smith Kline of today was formed from Glaxo, Wellcome, Smith Kline French, Beecham, Beckman, Affymatrix, Sterling, and a host of other smaller companies. Similarly Sanofi-Aventis was a merger of the following component companies: Hoechst, Rhone Poulenc, Marion Merell Dow, Roussel, Rorer, Sanofi, Connaught Labs, Merieux, and Synthelabo. Pfizer acquired Pharmacia, Werner Lambert, Parke Davis, Searle, Upjohn, and a host of other smaller companies. This applies to most top pharmaceutical companies except Merck. Mergers have started within the biotechnology industry, like Amgen taking over Immunex and Tularik, and Biogen-Idec and pharma companies like Roche taking over Genentech, Novartis-Chiron, and J&J—Alza, Centocor. Astellas was formed from the merger of Fujisawa and Yamanouchi, and Daichi and Sankyo merged in 2005.

As several competitive pharma R&D units were merged, several projects were terminated or given low priority and funds, and R&D staff was reduced or shifted, resulting in high turnover and low morale. The bigger company centralized R&D units became more risk averse. Several once promising areas of research like combinatorial chemistry produced huge chemical libraries with minor structural variations; HTS, gene therapy, proteomics, antisense, vaccines for AIDS, sepsis, RSV, and malaria, Alzheimer's disease, and Parkinson's disease have increased our

knowledge and understanding and the number of targets. However, no breakthrough blockbuster medicine has emerged out of these new high-speed and costly technologies. Demain [33] attributes such failure to the elimination of natural products and extracts from the HTS screening and recommends including natural products in HTS screens. The partnerships between different companies for R&D and marketing alliances/joint ventures to develop products have a higher probability of blockbuster success [17, 34]. Historically, blockbuster drugs were aimed at the larger markets of depression and ulcers. Other large-market indications like cholesterol and lipid lowering agents, antihypertensives, osteoporosis, asthma, and schizophrenia emerged due to the availability of effective and safe drugs (Table 7). These therapeutic areas have dominated the pharmaceutical markets for the past two decades. Several of these drugs were discovered and initially developed by smaller companies and licensed to big pharma for marketing and late development.

The successful biological products were developed to treat anemia, cancers, infections, and inflammatory joint diseases. The biologic blockbuster medicines have significantly contributed to the recent increase in the number of such drugs and made several previously ignored indications like multiple sclerosis, pulmonary arterial hypertension, anemia, cystic fibrosis, vaccines, and age-related macular degeneration commercially attractive and highly competitive areas for new drug development (Tables 6, 7) [1–4, 16, 20, 34–37]. The demand for new and better medications remains strong as currently only about one third of the over 35,000 human diseases have some treatment options available.

Based on new product approvals and creation of blockbuster brands, the success rates of biotech products are higher as compared with traditional chemically derived small-molecule new chemical entities. The biotechnology industry growth and market valuation has outpaced the pharmaceutical industry during the past 2 years. The increasing dependence of big pharma on small biotech is evident by the increasing number of licensing and R&D funding deals. The bulk of new medications approved in the years 2003–2005 originated in small biotech companies [35–39].

The higher R&D productivity of biotech companies is linked to their close relationship with academic biomedical research, higher risk taking by boldly testing first-in-class, and new concepts and designs in clinical trials, which are driven by unmet medical needs even for diseases for which the pharmaceutical industry has shown no interest. The R&D failure rate of biotech companies is much higher as compared with big pharma testing of follow-on or me-too drugs [1–3, 16, 20, 32–35]. For blockbuster drugs, the biotechnology companies have followed the standard pharmaceutical company strategy. They have followed fast-track approval for the first indication in an orphan or niche disease with high unmet medical need, market the drug at high prices, and expand approvals in other related diseases and new markets.

The risk associated with the development of human antibodies and biologics is lower than that with an NCE with a probability of success at 8% in phase I. Chimeric or humanized antibodies have a much higher success rate at the phase I stage (25%), due to recent advances in target selection and recombinant technology. New developments include techniques for converting murine monoclonal antibodies to mouse–human chimeric antibodies, for humanizing the antibodies, and producing

human antibodies using transgenic mice. These approaches minimize or eliminate the immune responses and rejection associated with the earlier murine antibodies [37–40]. The strong sales and demand from cancer and arthritis patients have outdated almost all commercial market research reports and sales forecasts for monoclonal antibodies. The higher success rate, low cost of entry and IND filing, and use of human/humanized monoclonal antibodies—with 6 blockbuster brands, half of which arrived in 2005, and the potential for more billion-dollar monoclonal antibodies to reach the market—has fueled the current Monoclonal Antibody Gold Rush. The bandwagon has been joined by 200 companies with hundreds of new projects and targets and has attracted billions of dollars in R&D investment. Factors contributing to the success are unmet medical needs of patients, lower cost of entry into R&D and production, the overcoming of technical hurdles in large scale manufacturing, and availability of humanized or fully human monoclonal antibodies.

The R&D productivity of the big pharma with biotech companies has been compared for the period 1998–2004 in commercial reports [1–3]. The R&D expenses and the FDA approvals of the new molecular entities were linked [16, 20, 23, 24]. The biotech approvals were less than 10, and the pharma NCE approvals declined from 25 to 15 from 1998 to 2005. The decline of the first-in-class medications from big pharma was even more dramatic. The biotech approvals were more numerous than pharma approvals in 2003 and 2004, and the trend reversed in 2005. The big pharma R&D expenses were $50 billion in 2004, resulting in 10 NME approvals, whereas biotech approvals were 20 for an R&D budget of $20 billion. Tables 6 and 7 list the top-selling major therapeutic proteins and biological brands created by the biotechnology industry since 1997. The list provides therapeutic areas of blockbuster brands, marketing wars between different brands for market leadership, and shift to newer brands with improved properties.

## 6 R&D SUCCESS

The biotechnology industry had its share of major R&D and clinical trial success. The big gains for biotech products are mainly in anemia; cancer; TNF alpha inhibitors; human insulin with short, mixed, and long duration of action; interferon for multiple sclerosis and hepatitis C; monoclonal antibodies for cancer; RSV; and immuno-inflammatory diseases. This success has resulted in attracting top pharmaceutical companies and R&D funds for biosimilar products with improved properties. It is easier for start-up companies in the above areas to attract capital and angel investors or to conclude licensing deals and find marketing partners.

As EPO, insulin, interferon, and monoclonal antibodies are covered in other chapters, only a brief resume will follow of these major success stories of biotechnology R&D.

### 6.1 Erythropoietin

Erythropoietin, Lipitor, and TNF inhibitors are likely to cross the $15 billion mark within the next few years. Amgen commercial success with its EPO brands has attracted several competitors with which Amgen has been involved in costly and protracted patent and marketing litigation to protect its franchise. Erythropoietin

is the top litigated biotech product. In 2005, over 20 erythropoietins were in development either as follow-on or improved biosimilar products [20, 41]. Numerous players are striving to dominate this market, including, among others, Amgen, J&J, Roche, Chugai, Shire, Kirin and Elanex, Teva, Sandoz, Stada, Cangene, Microbix and Gene Medix. Amgen patents expired in Europe in 2004 but are valid in the United States until 2015, and it has shifted patients to its new Aranesp brand. In Europe, two forms of rEPO are available: Eprex (J&J) and NeoRecormon (Roche). Two other forms Dynepo (Shire, Sanofi Aventis) and CERA (Roche, PEG-EPO) are likely to be approved in 2007. Amgen has filed a preemptive patent infringement against Roche for its PEG-EPO (CERA), thereby delaying its market launch in the United States. The FDA's recent alert about erythropoietin (EPO) use in cancer patients with over 12 gm/dL of hemoglobin and the black box warning linking EPO agents to increased mortality, cardiovascular events, tumor growth, and hyptertension are likely to result in market and sales decline in 2007. The FDA advised doctors to use the lowest dose of EPO for patients to achieve hemoglobin levels of 12 gm/dL.

### 6.2 Tumor Necrosis Factor Inhibitors

The commercial success of antitumor necrosis factor monoclonal antibodies like Remicade (Infliximab), Humira (Adalimumab), and Enbrel (Etanercept) in rheumatoid arthritis is due to better tolerance and superior efficacy as compared with older disease modifying antirheumatic drugs (DMARDs) with variable and short response rates but higher toxicity leading to either their low clinical use or market withdrawal (Table 6) [42–44]. TNF inhibitors with methotrexate provide a response rate of 50–60% by the FDA, American and European Rheumatology remission criteria. Risks associated with TNF-blockers are serious infections, congestive heart failure, demyelinating diseases, and systemic lupus erythematosus, but in most cases, they can be identified and managed. Biological agents have revolutionized the treatment of autoimmune/inflammatory diseases like rheumatoid arthritis and Crohn's Disease (CD). Enbrel was the first TNF inhibitors approved for treatment of rheumatoid arthritis patients in 1998. The FDA (fast track) initially approved Remicade for Crohn's Disease the same year. Both agents were tried and failed to show any efficacy in early trials in cancer, sepsis, and congestive heart failure. Both Enbrel and Remicade are approved for rheumatoid arthritis (RA), psoriatic arthritis (PsA), and ankylosing spondylitis (AS). There is some differentiation as Enbrel is effective in juvenile rheumatoid arthritis (JRA) and psoriasis (Ps). Remicade is approved for CD and ulcerative colitis (UC) and has shown efficacy in sarcoidosis and Wagener's granulomatosis. The third agent, Humira, is approved for use in RA, PsA, AS and CD.

After the launch of Enbrel, the demand for patients during the first 3 years of marketing was so strong that Immunex and its marketing partner Wyeth (American Home Product) were unable to increase good manufacturing practice (GMP) production to meet the surge in demand. There was a 30% shortfall, and these patients drifted to Remicade. Only Amgen's takeover of Immunex in 2002 resolved the supply problems to meet demand. Remicade, with its approval in UC and CD, was the market leader and overtook Enbrel in sales in the 2002–2003 period. Enbrel

has been the current market leader in its class since 2004. Kineret (anakinra, Amgen) is a recombinant, nonglycosylated synthetic form of the human interleukin-1 receptor antagonist (IL-1Ra) that mimics the body's endogenous mechanism for regulating IL-1 and is approved for treatment of RA patients not responding to DMARDs and TNF inhibitors. Kineret (anakinra, Amgen) has failed to gain a significant market share of the RA patients because it cannot be used in combination with TNF inhibitors. Orencia (Bristol-Myers Squibb abatacept) is the first T-cell costimulation modulator approved for the treatment of RA and was marketed in early 2006. Orencia in combination with TNF inhibitors or other biological agents like Anakirna is not allowed. Roche Rituxan, a monoclonal B-cell-specific antibody to CD20 (MabThera), was approved for use in RA patients along with methotrexate.

Recent R&D failures include Biogen Idec, Lilly-ICOS, and Alexion products in clinical trials. The Roche monoclonal antibody to the IL-6 receptor for RA (MRA) has shown efficacy in RA patients not responding to methotrexate and TNF inhibitors. If long-term safety, efficacy, and remission of new biologics are established in osteoarthritis trials, it will be an important therapeutic advance and create a large market.

### 6.3 Psoriasis

Alefacept (Amevive, Biogen Idec) and Efalizumab (Reptiva, Xoma, Genentech, Serono), the two monoclonal antibodies, which block the activation of T cells, were hailed as major advances in the treatment of psoriasis. The sales of Reptiva were $112 million and $160 million and of Amevive were $48 and $15 million in 2005 and 2006 below initial sales forecasts mainly due to competition from TNF inhibitors [45]. The technical success of R&D of non TNF biologicals for treatment of psoriasis has not translated into commercial success due to the large clinical data base and dominance of TNF antagonist products.

### 6.4 Insulin

A review of the insulin blockbuster brands in Table 10 indicated the market entry of a new company Sanofi Aventis to challenge the market dominance by Lilly and Novo Nordisk and the rapid growth of the newer insulin analogs with improved properties. The aim of the R&D is to move away from several daily insulin injections. There has been intense R&D activity during the past few years to provide patients with improved insulin that is fast acting, of a long duration, and dual action. This approval and launch of the first inhaled insulin Exubera (Pfizer) should expand the market, and initial peak sales projections are $3 billion per year depending on its long-term effects on lung functions. The UK cost effectiveness agency has determined that the high costs of Exubera are not justified as it does not offer any advantages over insulin injections, and concerns about long term toxicity to lung cells has delayed marketing by Pfizer in USA despite FDA approval. Table 8 lists new improved marketed and approved insulin analogs and those in development [46–48].

**TABLE 10. Improved Insulin Analogs**

| Company | Product | Description | 2006 Sales Status $ Million |
|---|---|---|---|
| Sanofi Aventis | Lantus Apidra | Insulin glargine Insulin glulisine | 2200 |
| Novo Nordisk | Insulin Aspart Levemir | Rapid insulin Insulin detemir | 1700 |
| Lilly/Amylin | Byetta | Exenatide incretin mimetics | 430 |
| Pfizer | Exubera | Inhaled insulin powder | Approved |
| Lilly/Alkermes | AIR Insulin | Insulin powder | Phase III |
| Mannkind | Technosphere Insulin | Powder liquifies in lungs | Phase III |
| Novo Nordisk Aradigm | NN 1998 | Aerosol Mist | Phase III |
| Kos | KI 02 212 | Liquid suspension | Phase II |
| Novo Nordisk | NN 344 NN 5401 | Long-acting insulin, improved analog | Phase I |

### 6.5 Interferon

Interferon alpha in combination with ribavarin for the treatment of hepatitis C is a commercial success. The biotech industry had its once promising drugs failing in clinical trials, and drug withdrawal by Biogen and Elan of Tysbari (natalizumab) for multiple sclerosis due to deaths of some patients from a rare brain infection was a setback. Multiple sclerosis (MS) is a neurological disorder affecting 2 million patients worldwide. Teva/Sanofi-Aventis Copaxone joined the three interferon blockbuster brands for the multiple sclerosis market led by Avonex in 2005. The total MS market sales reached $7 billion in 2006. Copaxone and return to market of Tysbari will expand the market and replace older interferon brands. Interferon sales slowed in 2005 after 2 years of rapid growth due to the new use of a poly ethylene glycol (PEG)–interferon combination with ribavarin in hepatitis C and withdrawal of Tysabri.

Biologics are important for new emerging threats, bioterrorism, and infectious diseases such as pandemic influenza, West Nile Virus (WNV), SARS, antibiotic-resistant tuberculosis, and the prion agent that causes the human form of mad cow disease, variant CJD. New vaccines to protect against HIV, RSV, anthrax, smallpox, hepatitis, malaria, WNV, and bird (avian) flu are in clinical testing [49–51]. Monoclonal antibodies and peptides have made a great contribution to the treatment of cancer patients, and several antibodies and peptides have become blockbuster drugs. Many of these antibodies carry black box warnings. Several new anticancer agent brands are blockbuster drugs, and the list grows each year [52–54].

The advent of Macugen and Lucentis to treat age-related macular degeneration and endothelin antagonists (Tracleer, Thelin, Ambrisentan) and prostacyclin analogs (Epoprostenol, Treprostinil, Beraprost, Iloprost) to treat pulmonary arte-

rial hypertension are important treatment advances. Both of these therapeutic areas have several active R&D projects at the approval and the clinical trial stage.

Roche has been the most efficient and productive industrial R&D during the past 2 years. The association of Roche with 5 of the top 10 biotech products is a tribute to its success in acquiring Genentech, an investment in biotechnology and its alliance with biotech companies. Amgen and Genentech market cap was in the $90–$100 billion range in early 2006 but has come down recently, and was still higher than Wyeth, BMS, or Schering Plough. Other big companies like J&J and Lilly have a strong pipeline of biological products in development.

## 7 R&D FAILURES

The biotechnology industry had its share of major R&D and clinical trial failures. The big late-stage setbacks for biotech products are mainly in Alzheimer's disease, sepsis, and vaccines for RSV, malaria, and AIDS. Despite such failures, the R&D funding by big pharma continues in Alzheimer's and sepsis but has bailed out in vaccines and infectious diseases. The long history of failures has made it difficult for a biotechnology company working in the above areas to attract venture capital and angel investors or to conclude licensing deals or find marketing partners.

### 7.1 Alzheimer's disease

Alzheimer's disease (AD), the most common cause of cognitive impairment in older patients, is characterized by the development of senile plaques and neurofibrillary tangles, which are associated with neuronal loss affecting to a greater extent cholinergic neurons. A cascade of pathophysiological events is triggered in AD that ultimately involves common cellular signaling pathways and leads to cellular and neural networks disfunction, failure of neurotransmission, cell death, and a common clinical outcome. The viable neurons remain an important target for therapeutic intervention at each stage of disease evolution. Currently, symptomatic drugs inhibiting the degradation of acetylcholine within synapses and more recently glutamate receptor antagonists represent the mainstay of therapy. However, interventions able to halt or slow disease progression (i.e., disease-modifying agents) are necessary. Although much progress has been made in this area, there are currently no clinically approved interventions for AD classed as disease-modifying or neuroprotective drugs [55].

The first four approved and marketed drugs Aricept (Pfizer, Eisai, donepezil), Reminyl (J&J, galantamine), Exelon (Novartis, Rivastigmine), and Cognex (Pfizer, Tacrine) are reversible cholinesterase inhibitors [55]. Memantine is an *N*-methyl-D-aspartate (NMDA)-type glutamate receptors antagonist, which is the first in a new class of AD medications. Glutamate is the most common excitatory neurotransmitter in the central nervous system, and the modulation of glutamatergic neurotransmission is a major target for the treatment of Alzheimer's disease. Memantine was developed by Merz and licensed to Forest for the U.S. and Lundbeck for the European and international markets. Memantine is marketed under the brands Axura and Akatinol by Merz, Namenda by Forest, and Ebixa by Lundbeck. Clinical studies of the safety and efficacy of memantine for other neurological disorders, including glaucoma and other forms of dementia, are currently

ongoing [56]. A series of second-generation memantine derivatives are currently in development and may have better neuroprotective properties than memantine. Recent failures have included Axonyx, Forest Labs, Elan, Myriad, and Neurocrine. Elan and Wyeth (American Home Products) ended trials of their Alzheimer's vaccine (AN-1792) in 2002, after patients suffered swelling of the central nervous system or inflammation of the brain and spinal cord. The vaccine activated the patient's own immune system to attack a protein called beta-amyloid that is implicated in Alzheimer's. Several tested drugs provided good results in phase II but failed in phase III. Some drugs showed a biphasic response curve. Several muscarinic receptors agonists failed in phase II–III trials.

### 7.2 Anti-infective Exodus and U-turn

During the 2001–2003 period, several companies, namely Aventis, Bristol Myers Squibb, Eli Lilly, Roche, GlaxoSmithKline, Proctor & Gamble, and Wyeth, eliminated or greatly curtailed their anti-infective R&D [57]. The reasons for pharmaceutical companies leaving the anti-infective area were slow growth in the West, unpredictable and seasonal demand, pricing pressure to donate supply or at cost to poor countries and patients, the vaccine unprofitability, and low investment in vaccine companies. Vaccine companies were scared and hesitated to launch new vaccines with sales in the tens to hundreds of millions of dollars because of the threat of billion-dollar punitive damage settlements. The liability issues and current regulatory requirements for new vaccines have a negative impact on vaccine R&D. Public–private partnerships, along with NGO and public initiatives, have resulted in increased funding of newer vaccines for tropical, orphan, and neglected diseases like malaria, tuberculosis, and leishmaniasis.

Vaccines were back in favor and were highlighted in R&D portfolios of major companies (Sanofi Aventis, Glaxo Smith Kline, Merck, Novartis, and Wyeth). Global sales in 2006 were $15 billion because of scare about avian influenza, bioterror organisms, and new emerging infections like SARS. Introduction of new cancer and rotavirus vaccines will greatly expand the vaccine market in the next 2 years. During the past few years, the spotlight on SARS, Avian Influenza, new emerging infections, and bioterror agents have seen a few big companies like GSK, Merck, Novartis, and Sanofi Aventis declaring vaccine production as a high priority area of R&D. It is ironic that Roche, which had closed its anti-infectives R&D unit, was in the news to supply its antiviral Tamiflu to stop Avian Influenza. The success of Prevenar (Wyeth), the explosive demand for influenza vaccines, and the approval of two new Rotavirus vaccines with huge safety databases have brought back the interest of the industry in vaccines. Pediatric vaccines, influenza, and hepatitis B vaccines have been the leading categories. Although the demand for influenza vaccine was strong, MedImmune FluMist 2005 sales of $21 million remained below forecast, due to high cost and its use limited to younger patients (5 to 17 years of age, and healthy adults, 18 to 49 years of age). A new refrigerator-stable formulation was 55% more effective in children in the age range of 6 and 59 months. New vaccines for cancer, RSV, AIDS, and malaria will expand the market [57–61]. Vaccines have been the most cost-effective treatment and probably were undervalued, but the higher prices of newer vaccines in the $100–$400 range are likely to keep even these out of the reach of poor patients [49–51].

### 7.3 RSV

The commercial success of a new product in most indications or additional use in new indications results in much higher R&D activity and funding by competitors. However, the success of Synagis (Palivizumab) in RSV and Drotrecogin alpha (activated protein C) in sepsis has resulted in industry bail-out due to the failure of several promising projects during the past decades [57, 62, 63]. The new drugs in development and currently approved drugs for RSV were reviewed recently and are listed in an updated Table 11 [57], which confirms a significant decline from over 25 to only 4 active RSV projects. Several RSV vaccines and antivirals, which were in phase III testing during the last review, are considered terminated either due to low efficacy or safety concerns due to adverse effects. These projects are not listed in the companies' current R&D pipeline, but many commercial databases still list these as active projects.

### 7.4 Sepsis

Severe sepsis or septic shock occurs when an infection (bacterial, viral, fungal, or parasitic)—often the result of trauma, surgery, burns, or cancer—triggers a cascade of immune system responses that can lead to acute organ disfunction and often death [62]. The death rate from severe sepsis ranges from 30% to 50%. Every day 1400 people worldwide die from severe sepsis. The International Sepsis Forum estimates that more than 750,000 individuals develop severe sepsis in North America each year, with similar estimates for Europe, and all need to be actively treated in the hospital. Sepsis kills 215,000 people annually in the United States and is the most common cause of death in noncoronary intensive care units. The annual costs associated with the treatment of patients with severe sepsis are estimated to be US$17 billion. The incidence of sepsis is expected to increase to 2.2 million per annum in the leading markets over the next 10 years as the elderly population grows. Thus, a safe and effective treatment, which saves lives, can be a blockbuster drug [62, 63].

Activated protein C (Drotrecogin alpha, Xigris) is the first and only drug shown to cure those suffering from sepsis. It had sales of $214 million and $192 million

**TABLE 11. Marketed and R&D Development Products for Prevention of RSV**

| Product | Characteristics | Company | Sales 2006 Status |
|---|---|---|---|
| Palivizumab Synagis | hAnti F-glycoprotein | MedImmune | $1100 million |
| RespiGam | Immunoglobulin | MedImmune | negligible |
| Ribavarin | Antiviral | Valeant | Generics |
| Motavizumab Numax | Monoclonal antibody | Medimmune | Phase III |
| A 60444 | Antiviral | Arrow | II |
| Live attenuated | RB/HPIV3-RSV-A/B | MedImmune | I |
| ALN RSV01 | RNAi | Alnylam | I |

*Notes*: Several commercial databases still list RSV vaccines and antiviral in clinical development by companies, e.g., Wyeth, Sanofi Aventis, and Glaxo Smith Kline. These projects are considered discontinued as companies' websites do not list any active RSV projects or there has been no new information for the past 3 years after completion of phase II or III trials.

in 2005 and 2006, respectively. In Lilly's PROWESS clinical trial, drotrecogin alpha only reduced the absolute risk of death by 6%. It seems that, with a $6800 price tag, this was insufficient to persuade physicians to prescribe the drug outside the most severe cases. In 3.5% of drotrecogin alpha-treated patients (compared with 2% of placebo treated patients), the drug also increased the risk of serious bleeding events. Another obstacle to drotrecogin alpha's widespread adoption is that it can be hard to identify candidates for treatment [62–64].

Drug development for sepsis and septic shock has been disappointing: Numerous agents have demonstrated promising activity in early clinical phase I/II and then failed to show significant efficacy in phase III trials. In recent years, several biotechnology firms have gone bankrupt or were taken over trying to develop a drug to combat sepsis. Billions of dollars have been wiped out in market values of the biotechnology companies due to failures of their drugs in sepsis.

In the early 1990s, two promising drugs named Centotoxin and E5 from Centocor and Xoma were in a race to be the first effective treatment. The two drugs had shown efficacy in early phase II clinical studies and were rushed into large-scale phase III clinical studies. Both drugs failed to show improvements in mortality in phase III and were rejected by the FDA and its advisory committee in 1992 [62].

Randomized trials of anti-inflammatory therapies agents, including NSAIDs and corticosteroids, failed to show improvement in survival from sepsis and septic shock. Three monoclonal antibodies to endotoxin were tested in clinical trials, and all failed to improve survival from sepsis and septic shock. The Centocor project went down in 1992, and drugs from Chiron, Xoma, and Synergen failed in 1994. Many anticytokine therapies have not improved survival from septic shock. These trials include the use of naturally occurring antagonists to tumor necrosis factor and interleukin-1, antibodies to tumor necrosis factor, bradykinin antagonists, ibuprofen, and a platelet-activating-factor antagonist. Additional endotoxin-directed therapies being studied were the use of bactericidal permeability-increasing factor, soluble CD14 receptors, and reconstituted high-density lipoproteins [62].

Anti-inflammatory interventions under study include efforts to modulate nitric-oxide activity, therapies directed at neutrophil adhesion molecules, the infusion of antithrombin III, tissue-factor pathway inhibitors, and pentoxyfylline. Alternatively, GCSF is being studied in an effort to enhance phagocytic-cell function in patients with severe sepsis [62]. During the past 5 years, Chiron and Pharmacia's tifacogin failed late-stage clinical trials at the end of 2001 and endotoxin from Baxter/Xoma (Neuprex) has failed to show clear benefits in large-scale studies and requires extensive additional studies (2002). ICOS and Suntory, Pafase, a platelet-activating factor acetylhydrolase; Lilly's LY-315920 phospholipase A2 inhibitor; and Knoll-Abbott's monoclonal murine anti-tumor necrosis factor-alpha antibody, Afelimomab, are some recent phase III trial failures.

Despite past failures, there are still several active R&D projects in late-stage clinical trials (Box 2).

## 8 COST CONSTRAINTS

The start-up cost of development, manufacture, and distribution of biological products is very high, and only a few companies can manufacture and test the product to current regulatory standards of GMP, GLP, and GCP (Good Manufacturing,

**BOX 2**

**Phase III**

- Eisai Eritroran, an antagonist of lipopolysaccharide
- Takeda TAK 242, which inhibits signal transduction through Toll Like receptor 4 Phase
- Prosthetics' CytoTAb target anti-tumor necrosis factor alpha polyclonal antibody (licensed in late 2005 to Astra Zeneca)

**Phase II**

- Novartis (PMX 622) covalent conjugate of Polymyxin B and Dextran-70
- GSK 270773 phospholipid anti-endotoxin emulsion
- Behring Complement CI inhibitor

**Phase I**

- Idun (acquired by Pfizer in 2005) IDN 6556 caspase inhibitor

**TABLE 12. High-priced Biologic Medications 2005–2006**

| Company | Trade Name | Generic Name | Indication | Price per Month $ |
|---|---|---|---|---|
| Biogen Idec | Zevalin® | Ibritumomab | Lymphoma NHL | 24,000 |
| BMS | Erbitux® | Cetuximab | Colon Cancer | 17,000 |
| Genzyme | Fabrazyme® | Agalsidase Beta | Fabry Disease | 15,000 |
| Genzyme | Aldurazyme® | Laronidase | MPS 1 | 16,600 |
| Genzyme | Cerezyme® | Imiglucerase | Gaucher disease | 16,600 |
| Serono | Serostim | HGH | Dwarfism | 7,000 |
| Lilly | Xigris® | Activated Protein C | Sepsis | 6,800 |
| MedImmune | Synagis® J | Palivizumab | RSV | 5,600 |
| Genentech | Avastin® | Bevacizumab | Colon Cancer | 4,400 |
| Roche | Herceptin | Trastuzumab | Breast Cancer | |

Laboratory, and Clinical Practice). The biologics for cancer and rare and other life-threatening diseases also rank as the most expensive treatments, and these products achieve blockbuster sales sooner than traditional pharmaceuticals (Table 12, Fig. 1). The price of biosimilar or follow-on products will be higher as compared with traditional generic drugs as very few companies can invest in manufacturing technology.

With the introduction of biologics for the treatment of rheumatoid arthritis and psoriasis, the annual cost of treatment has multiplied 10-fold to $25,000. Only about 10% of the eligible RA patients in the United States and Europe are currently on TNFα therapy. If all RA patients were to receive TNFα therapy, the total cost of therapy would reach $100 billion just for one disease only. The current or future health-care systems are not likely to fund such cost increases for anemia,

cancers, infectious, cardiac, metabolic, neurological, and rare diseases as well as increasing acute care, diagnostic, and device costs [65, 66].

Treatment of psoriasis with new biologics costs in the range of $25,000–$45,000 per year as compared with $2200 for methotrexate and $3000–$5000 per year for phototherapy. The cost of erythropoietin for renal dialysis is $10,000 per year and for cancer patients $1000 per month of treatment. New vaccines are priced in the $200–$400 range. Xigris is priced at $6800 per injection and drug-coated stents at over $3000. The European government-funded healthcare systems and insurance companies try to delay and limit access to costly new medicines and try to negotiate lower prices.

The average price of new cancer treatments and for rare diseases has increased from the $20,000–$25,000 range to the $100,000–$280,000 range per year of treatment [51–54, 66]. If two or three new drugs were to be combined for treatment of colon cancer like Eloxatin, Erbitux, and Avastin, the yearly cost could easily reach $0.5–1 million per patient. If Erbitux was used for all eligible colon cancer patients in the United States, it would cost $1.2 billion. Several European health insurance systems, oncologists, and patient organizations have raised questions about the cost–benefit ratio and called for regulatory rejection of anticancer medications with short-term limited benefits. In the future, about 80% of cancer patients will be from poor countries or without insurance coverage. Colon cancer survival doubled in the past decade, but the costs multiplied by 340 for the first 2 months of treatment [66].

The clinical testing of these high-priced drugs in poor and uneducated patients in the developing countries is problematic due to ethical and moral grounds. These drugs are not likely to be marketed in poor countries or made available to responding patients at the end of the trial. The high cost of these life-saving medicines raises several critical issues concerning access, payment, and the legal, moral, and ethical aspects of supply and demand to poor countries and patients [67, 68].

Initial reluctance and resistance by industry to reduce the prices of AIDS medications for poor countries/patients, let aside patent protection to increase supplies, or reduce prices in national emergencies like the Anthrax scare in the United States (Bayer Ciprofloxacin), SARS, and Avian Influenza (Roche Tamiflu) has eroded the image of the research-based innovative pharmaceutical industry.

Biotechnology products have so far escaped pricing pressure to reduce costs as the treatments were for life-saving or rare diseases and were marketed by big pharmaceutical firms. As biological medications prove effective in treating chronic diseases like arthritis, psoriasis, macular diseases, and diabetes, big biotechnology firms will come under pressure to reduce costs and face class-action litigation for product injury. Biotechnology industry leaders should find an appropriate balance between the current "high cost high profits, low volume" model to "affordable cost, high volume, reasonable profits" model to increase the availability of life-saving medications to poor patients and countries.

## 9 PUBLIC–PRIVATE PARTNERSHIPS

Several public–private partnerships support development of drugs, vaccines, and diagnostics to address diseases that predominantly afflict the poor or to make drugs

available to patients in poor countries and operate on the virtual company model [57, 69, 70]. As most pharmaceutical and biotech R&D has been directed toward diseases of the Western World, there has been no interest in developing vaccines and treatments for the tropical and neglected diseases like AIDS, Chagas' Disease, contraception, dengue, diarrhea, leprosy, leishmaniasis, lymphatic filariasis, malaria, onchocerciasis, schistosomiasis, trypanosomiasis, tuberculosis, and sexually transmitted diseases. None of the major companies has an R&D laboratory in Africa, and only two, Glaxo Smith Kline and Astra Zeneca, have R&D laboratories in Asia (outside Japan). International and national organizations like the World Health Organization (WHO) failed to develop new medications for tropical diseases and lacked sufficient resources. Universities or research institutes were not able to attract interest from the industry for promising leads, and companies terminated their own R&D leads because of lack of profits or commercial success.

Developments in the area of neglected diseases have opened up new opportunities for licensing by universities and companies. Over the past decade, private foundations and donors have provided social venture capital to launch several nonprofit "companies" that have now collectively raised over a billion dollars from philanthropic and government donors to support product development.

The price of AIDS drug tritherapy is $10,000 a year per patient in the United States and Europe. Indian companies have reduced the cost to $130 per patient per year and introduced a single tritherapy combination. When Roche was initially reluctant to increase production or reduce the price of its Tamiflu to treat Avian Influenza for developing countries claiming a difficult synthetic process, some Indian companies made the active ingredient from Shikimic acid within one week

The Foundation for the National Institutes of Health identifies and develops opportunities for innovative public–private partnerships involving industry, academia, and the philanthropic community (www.fnih.org). There were over 100 public–private partnerships funded with over $1 billion, with 60 active R&D projects for vaccines and drugs for tropical and neglected diseases by the end of 2005 (Table 13).

**TABLE 13. Public–Private Partnership for Tropical and Neglected Diseases**

| |
|---|
| Aeras, Global TB Vaccine Foundation (Aeras) (http://aeras.org) |
| Bill and Melinda Gates Foundation (www.gatesfoundation.org/GlobalHealth/InfectiousDiseases/) |
| Drugs for Neglected Diseases Initiative (www.dndi.org) |
| Foundation for Innovative New Diagnostics http://www.finddiagnostics.org |
| Global Alliance for TB Drug Development (www.tballiance.org) |
| Global Alliance for Vaccines and Immunization GAVI at WHO (www.who.int) |
| Global Forum for Health Research (www.globalforumhealth.org) |
| Global Fund to fight AIDS, Tuberculosis and Malaria (www.theglobalfund.org) |
| Institute for One World Health http://www.oneworldhealth.org |
| International AIDS Vaccine Initiative (www.iavi.org) |
| International Partnership for Microbicides http://www.ipm-microbicides.org |
| Malaria Vaccine Initiative (MVI). (http://www.malariavaccine.org) |
| Medicine for Malaria Venture (www.mmv.org). |
| Pediatric Dengue Vaccine Initiative http://www.pdvi.org |
| Program for Appropriate Technology in Health http://www.path.org |

## 10 INFORMATION RESOURCES

Traditional and new media play a vital role in projecting breakthrough and blockbuster medications and raising awareness among the general population (Table 14). The FDA/EMEA website provides detailed summaries and assessments of the actual data submitted in support of the NDAs. The transcripts of the FDA Advisory Committee meetings are valuable sources of information for approved and

**TABLE 14. Web Information Resources for Biotechnology Products**

| | |
|---|---|
| FDA | http://www.fda.gov |
| EMEA | http://www.emea.eu.int |
| WHO | http://www.who.int/en/ |
| Canada | http://www.hc-sc.gc.ca/ |
| UK | http://www.mhra.gov.uk/ |
| NICE | http://www.nice.org.uk/ |
| NIH | http://www.nih.gov/ |
| CDC | http://www.cdc.gov |
| BIO | http://www.bio.org/ |
| EuropaBio | http://www.europabio.org/ |
| PHRMA | http://www.phrma.org/ |
| IFPMA | http://www.phrma.org/ |
| GPHA | http://www.gphaonline.org/ |
| DIA | http://www.diahome.org/ |
| IMS | http://www.ims-global.com/ |
| DataMonitor | http://www.datamonitor.com/ |
| Prous | http://www.prous.com/ |
| ScienceDirect | http://www.sciencedirect.com/ |
| Thompson | http://www.thompsonpharma.com/ |
| PharmaProjects | http://www.pjbpubs.com/pharmaprojects/ |
| BioCentury | http://www.biocentury.com/ |
| BioSpace | http://www.biospace.com/ |
| Mayo | http://www.mayoclinic.com/ |
| CenterWatch | http://www.centerwatch.com/ |
| ClinTrialsGov | http://www.clinicaltrials.gov/ |
| Cochrane | http://www.cochrane.org/index0.htm |
| http://www.blackwell-synergy.com | |
| http://www.springerlink.com/ | |
| http://www3.interscience.wiley.com/ | |
| http://www.biomedcentral.com/ | |
| http://gateway-di.ovid.com/ | |
| http://www.isinet.com/ | |
| http://www.centerwatch.com/ | |
| Accenture http://www.accenture.com/ | |
| Bains, http://www.bain.com/bainweb/home.asp | |
| Boston Consultancy Group http://www.bcg.com/ | |
| IBM. http://www.ibm.com/us/ | |
| Frost & Sullivan. http://www.frost.com/prod/servlet/frost-home.pag | |
| Reuters. http://www.reutershealth.com/en/index.html | |
| Ernst & Young http://www.ey.com/global/content.nsf/International/Home | |

new drugs under review, concerns about safety and efficacy and about prescribing information, and requirements for additional studies (http://www.fda.gov/foi/electrr.htm). Similar information is available on the Internet for public scrutiny, by the FDA/EMEA for the accepted IND files of new medications. The NIH (NIAID), Centers for Disease Control, Institute of Medicine, Mayo Clinic, Center Watch, WHO, and Public Citizen websites offer valuable information about the disease and its treatment and the list of ongoing trials. Company websites offer prescription information and other details about the currently approved drugs, clinical trials results, sales data, and R&D portfolio, patent litigation, and expenses. The commercial websites of Pharma Projects, Prous Science, Thompson, and BioCentury are other important sources for drug R&D. BiomedCentral and PubMed offer free access to many full text papers and abstracts. The ISI web of Science, Science Direct and BiomedCentral, and PubMed were consulted for reviews. To keep the number of references reasonable, original research papers have often been replaced by very recent reviews. Cochrane Reviews is a good source of new drug clinical safety and efficacy. For commercial reports, only information available in the public domain was used and cited. Negative media stories about high prices, profitability, quality/ethical problems, and links to serious adverse reactions may lead to declining sales or regulatory action.

## 11 INDUSTRY AND MEDICINAL BRANDS IN 2006

Most of the top companies had reported 2006 annual earnings and sales of top selling brands within the first two months of the new year. The Japanese companies had reported annual sales or forecasts based on actual results of 1–3 Q 2006, which was used. The new data was incorporated in the tables and text at the last-minute stage.

A review of the 2006 sales, market, earnings, and R&D showed several new trends and surprises in the top ten company listing as well as best selling medicinal brands [71,72]. Pfizer, which has been the most valuable company during the past five years, was overtaken by Johnson & Johnson in sales last year and briefly for a few months in market value but has now regained its top position in market valuation. Pfizer retained the top spot as the most profitable and with the highest R&D budget. Glaxo Smith Kline (GSK) retained the third position by total sales, market cap, and profitability, and topped the list with highest sales increase in 2006. Roche consolidated its return to the top ten groups by overtaking Novartis in market cap and closing the gap with Amgen in sales of biologics. Roche and Sanofi Aventis ranked second and third by sales of biologics as in 2005. Overall, increased R&D spending by the pharmaceutical and biotechnology industry failed to increase the number of new drugs approved. Only Merck and Amgen increased their R&D budget in 2006 by $1 billion over previous year. Roche, with no blockbuster going off patent in 2006, came out as the most efficient R&D company. Sanofi Aventis, despite loss of 3 blockbuster drugs to generic competition, came out second with creation and growth of several blockbuster brands. Merck and BMS ranked in the third position, while Takeda and Astra Zeneca took the fourth and fifth positions (Tables 8, 15) [71,72].

**TABLE 15. Pharmaceutical Biotechnology Industry Performance Rankings 2006**

| Sales Increase $ Billion | | Biologics Sales $ Billion | | Profits $ Billion | | Market Cap $ billion 7 March 2007 | | R&D Efficiency* |
|---|---|---|---|---|---|---|---|---|
| GSK | 9 | Amgen | 14.3 | Pfizer | 19.3 | J&J | 179 | Roche |
| Roche | 5.6 | Roche | 13.5 | J&J | 14.5 | Pfizer | 176 | Sanofi Aventis |
| Novartis | 5 | Sanofi Aventis | 8.7 | GSK | 14.43 | GSK | 151 | BMS, Merck |
| Sanofi Aventis | 4.8 | Novo Nordisk | 7.25 | Sanofi Aventis | 12.6 | Roche (Genentech 86) | 142 | Takeda |
| Teva | 3.7 | J&J | 6.2 | Astra Zeneca | 8.54 | Novartis | 128 | Astra Zeneca |
| J&J | 3.3 | Baxter | 4.4 | Roche | 7.34 | Sanofi Aventis | 114 | Abbott |
| Wyeth | 2.4 | Allergan | 3.5 | Novartis | 7.20 | Merck | 96 | J&J, Teva |
| Novo Nordisk | 2.2 | Wyeth | 3.46 | Merck | 6.22 | Astra Zeneca | 82 | Daiichi Sankyo Pfizer |
| Amgen | 1.9 | Genzyme | 3.2 | Takeda | 5.4 | Abbott | 81 | Genzyme, Merck-Serono, |
| Lilly Schering Plough | 1.6 | Merck Serono | 3.2 | Lilly | 3.42 | Amgen | 72 | Wyeth Lilly |

*R&D productivity is linked to the introduction of new blockbuster medicinal brands and maintenance or sales growth of existing billion dollar brands and R&D resources shift to biologics and monoclonal antibodies (Maggon K. 2007).

**TABLE 16. Top Selling Medicinal Brands in 2006 ($ billion)**

| NCE | | NME | | Sales Increase | | Sales Decline | |
|---|---|---|---|---|---|---|---|
| Lipitor | 12.9 | Rituxan | 4.7 | Herceptin | 1.59 | Zocor | 1.6 |
| Advair | 6.13 | Enbrel | 4.4 | Rituxan | 1.5 | Zithromax | 1.38 |
| Plavix | 5.55 | Remicade | 4.2 | Copaxone | 1.2 | Zoloft | 1.14 |
| Nexium | 5.2 | Aranesp | 4.12 | Avastin | 1.1 | Paxil | 1.05 |
| Norvasc | 4.8 | Procrit | 3.2 | Tamiflu | 0.9 | Pravachol | 1.0 |
| Zyprexa | 4.36 | Herceptin | 3.14 | Aranesp | 0.82 | Allegra | 0.87 |
| Diovan | 4.22 | Neulasta | 2.7 | Lantus | 0.80 | Plavix | 0.6 |
| Risperdal | 4.18 | Epogen | 2.5 | Enbrel | 0.70 | Protonix | 0.6 |
| Pravachol | 4.0 | Novulin | 2.5 | Lipitor | 0.70 | Losec | 0.3 |
| Lansoprazole | 3.8 | Factor VIII | 2.5 | Remicade | 0.70 | Amaryl | 0.29 |

NCE-New chemical entity; NME-New molecular entity

Several pharmaceutical companies highlighted their R&D in biologic, deals and links with biotechnology companies, and sales of biologic brands. Roche is likely to overtake Amgen in biologic sales in 2007 to emerge as the top biotechnology company. One of Roche's unit, Genentech, was the most valued biotechnology company in 2006. The big pharmaceutical companies made significant strides in biologic sales and Roche, Sanofi Aventis, Merck-Serono, J&J, and Wyeth made the list of the top ten biotechnology companies, leaving only 5 pure biotech companies in the top ten [71,72]. Teva, the world's top generic company, took the fourth position in increased net sales (Table 15).

Lipitor remains the top selling medicinal brand with $12.9 billion in sales, followed by Advair and Plavix. Rituxan, with sales of $4.7 billion, was the best selling biologic brand followed by Enbrel, Remicade, and Darbepoetin. Medicinal brands with the highest sales increase in 2006 were mostly biologic and all the brands with highest sales decline due to patent expiry and introduction of generics were synthetic chemicals. Anticancer, tumor necrosis factor inhibitors, monoclonal antibodies, and vaccines had significant sales growth. The monoclonal antibody Rituxan for cancer and arthritis emerged as the top selling biologic brand followed by TNF inhibitors (Enbrel, Remicade) and Darbepoetin, each with sales above $4 billion. All the top sales decliners in 2006 were new chemical entities in patent expiry phase in major markets. There was no biologic brand in this list. Sanofi Aventis with 3 blockbuster brands lost to patent expiry was followed by Bristol Myers Squibb and Pfizer with two each. Five top products each had sales drops of over $ 1 billion each (Table 16) [71,72].

## 12 CONCLUSION

R&D success has been clearly linked to the development and creation of billion-dollar human medication. Current regulations favor fast-track approval of life-saving medication and long-term safety and efficacy data for chronic diseases. The blockbuster drugs are the main driver of industrial R&D and contribute a major share of the total pharmaceutical sales, profits, and market value of the successful companies. The criteria for blockbuster drugs are defined, and output of blockbuster

drugs by pharmaceutical and biotech R&D are compared. Biotech R&D was more productive and successful than pharmaceutical R&D in creating new blockbuster brands. The current Gold Rush of biologics indicates a paradigm shift in innovation from pharmaceutical to biotechnology R&D. Another perceptible shift is the industrial R&D bail-out from Europe toward the United States, once again based on the success of the U.S. biomedical science and biotechnology firms. Biotechnology R&D success includes several blockbuster brands of erythropoietin, tumor necrosis factor alpha blockers, insulin, interferon, and GCSF, monoclonal antibodies, vaccines, and anticancer, which had sales of over $1 billion during the 2004–2006 period. Biotechnology R&D failure includes Alzheimer's Disease, sepsis, and vaccines for RSV, malaria, and AIDS. Several public–private partnerships, operating on the virtual drug development company model, have taken up the void left by industry bail-out of tropical and neglected diseases. The high cost of biologics and link to serious adverse reactions will become important as biologics are used to treat chronic conditions like rheumatoid arthritis and psoriasis. New and old media have played an important role in raising awareness about new blockbuster biologics brands.

There were no indications or signs during the past 3 years from the big consultancy firms predicting the "End of the Blockbuster Era." On the contrary, the number of blockbuster brands has shown steady and continuous growth and now accounts for a major share of market, sales, and profits of the pharmaceutical and biotechnology industry. Pharmaceutical companies' outright acquisition of biotechnology companies and the licensing of technology/late stage projects in development increased significantly. The market and sales data once again provide strong support for the R&D paradigm shift to biologics and within biologics towards human monoclonal antibodies with a message: "It's biologics, Stupid."

## ACKNOWLEDGMENT

The author wishes to thank the following persons for their review and critical comments on the draft: Dr. Daniel Brandt, Debiopharm, Lausanne; Mr. Mohammad Akram, Business Consultant, Geneva; and my wife Dr. Sunita Maggon for editing, data collection, and literature search. I wish to thank Mr. Robert Posey (PDP, Geneva) for providing office facilities and administrative support.

## REFERENCES

1. Ernst & Young. (2005 & 2006). Global Biotech Report and Global Pharmaceutical Trends. Levinson AD (2005). What distinguishes biotech from big pharma? Global Industry Perspectives.
2. Bains & Company Reports. Rebuilding Big Pharma's Business Model. (2003) Addressing the Innovation Divide: Imbalanced innovation (2004).
3. Boston Consultancy Group. The gentle art of licensing. (2004) Rising to the productivity challenge. (2004).
4. FDA. (2004, 2006). Innovation or stagnation: Challenge and opportunity on the critical path to new medicinal products. CDER (2004). Report to the nation. Improving public health through human drugs.

5. Cockburn I M (2004). The changing structure of the pharmaceutical industry, *Health Affairs*. 23:10–22.
6. Ng R (2003). *Drugs: From Discovery to Approval*. John Wiley, New York, pp. 368.
7. Pfizer (2005). Annual Report. Available: www.pfizer.com.
8. Trombetta B (2005). Industry Audit, *Pharm. Exec*. 10:68–80.
9. Sellers L J (2004). Top 50 Pharma Companies. *Pharm. Exec*. 24:60–70. Available: www.pharmexec.com.
10. Parker M G, Amar D (2005). Building blockbusters. *Pharm. Exec*. 10:82–88.
11. Forbes (2004). Potential blockbuster drugs. Available: www.forbes.com/2003/10/08/cz_fmc_1008sf.html.
12. IMS Top Line Global Industry Data (2005). Available: www.imshealth.com, http://www.ims-global.com/insight/insight.htm.
13. Reuters Health. Reports (2004). The Pharmaceutical Market Outlook to 2010. Available: http://www.reutersbusinessinsight.com/content/rbhc0091t.pdf.
14. Datamonitor Potential blockbuster drugs (2004). Available: www.datamonitor.com.
15. Maggon K (2003). The ten billion dollar molecule. *Pharm. Exec*. 23:60–68.
16. Maggon K (2005). Best Selling Human Medicines 2002–2004. *Drug Disc. Today*. 10:738–742.
17. Maggon K (2004). The future of the research based pharmaceutical industry. *Express Pharma. Pulse*. January 29, 2004. Available: http://www.expresspharmapulse.com/20040129/oped01.shtml.
18. Maggon K (2004). Impact of Vioxx withdrawal on pharma R&D. *Medicine and Biotech.com*. November, 15, 2004. Available: http://www.medicineandbiotech.com/therapy.html1.html.
19. Maggon K (2002). Ups and down in drug development. *Express Pharma. Pulse*. Mumbai. October 3, 2002. Express Healthcare Management Mumbai, October 16, 2002. http://www.expresshealthcaremgmt.com/20021031/edit2.shtml.
20. Niazi S K (2005). *Handbook of Biogeneric Therapeutic Proteins: Regulatory, Manufacturing, Testing, and Patent Issues*. Taylor & Francis CRC Press, Boca Raton, F.L., p. 584.
21. Rader R A (2005). *Biopharmaceutical Products in the USA and European Markets*. Bioplan Associates, Rockville, M. D., p. 1234.
22. Maggon K (2004). Setting new standards about drug safety. *Express Pharma. Pulse*. 22 May 2004. Available: http://www.expresspharmapulse.com/20030522/edit2.shtml.
23. DiMasi J A, Hansen R W, Grabowski H G (2003). The price of innovation: New estimates for drug development costs. *J. Health Econ*. 22:151–185.
24. Grabowski H, Vernon J (2000). The determinants of pharmaceutical research and development expenditures. *J. Evo. Econ*. 10:201–215.
25. Biobusiness (2005). Sales of recombinant drugs. Available: www.i-s-b.net/business/recombinant_f.htm.
26. Danzon P M, Nicholson S, Pereira N S (2005). Productivity in pharmaceutical–biotechnology R&D: The role of experience and alliances. *J. Health Econ*. 24:317–339.
27. Higgins M J, Rodriguez D (2005). The outsourcing of R&D through acquisitions in the pharmaceutical industry. *J. Financial Econ.*, In Press, Corrected Proof, Available online 2 November.
28. Schmid E F, Smith D A (2004). Is pharmaceutical R&D just a game of chance or can strategy make a difference? *Drug Disc. Today*, 9:18–26.

29. Thomas L G (2004). Are we all global now? Local vs. foreign sources of corporate competence: the case of the Japanese pharmaceutical industry. *Strat. Manag. J.* 25:865–886.
30. Penner-Hahn J, Shaver J M (2005). Does international research and development increase patent output? An analysis of Japanese pharmaceutical firms. *Strat. Manag. J.* 26:121–140.
31. Drews J, Ryser S (1996). Innovation deficit in the pharmaceutical industry. *Drug Inform. J.* 30:97–107.
32. Drews J (1998). Innovation deficit revisited: Reflections on the productivity of pharmaceutical R&D. *Drug Disc. Today.* 11:491–494.
33. Demain A L (2002). Prescription for an ailing pharmaceutical industry. *Nature Biotechn.* 20:331–334.
34. Ho R J Y, Gibaldi M (2003). *Biotechnology and Biopharmaceuticals: Transforming Proteins and Genes into Drugs.* Wiley, New York.
35. Maggon K (2004). Rise of biologicals. *Express Pharma. Pulse.* February 5, 2004. Available: http://www.expresspharmapulse.com/20040205/editorial02.shtml.
36. Maggon K (2005). Biotech segment: Moving ahead of pharmaceutical industry. *Express Pharma. Pulse.* April 28, 2005. Available: http://www.expresspharmapulse.com/20050428/edit02.shtml.
37. Collins S W (2004). *The Race to Commercialize Biotechnology: Molecules, Market and the State in Japan and the US.* p. 224 Routledge, Oxford, U.K., p. 224.
38. Liossis S N C, Tsokos G C (2005). Monoclonal antibodies and fusion proteins in medicine. *J. Allergy Clin. Immunol.* In Press, Corrected Proof, Available online 1 September 2005.
39. Qu Z, Griffiths G L, Wegener W A, et al. (2005). Development of humanized antibodies as cancer therapeutics. *Methods.* 36:84–95.
40. Pavlou A K, Belsey M J (2005). The therapeutic antibodies market to 2008. *Eur. J. Pharmaceut. Biopharmaceut.* 59:389–396.
41. Charles S A (2005). Super Generics: A better alternative for biogenerics. *Drug Disc. Today.* 10:533–535.
42. Sesin C A, Bingham C O (2005). Remission in rheumatoid arthritis: Wishful thinking or clinical reality? *Seminars in Arthritis and Rheumatism.* 35(3):185–196.
43. Hochberg M C, Lebwohl M G, Plevy S E, et al. (2005). The benefit/risk profile of TNF-Blocking agents: Findings of a consensus panel. *Seminars in Arthritis and Rheumatism.* 34:819–836.
44. Miossec P (2005). Therapeutic targets in rheumatoid arthritis: More to come but which one(s) to select? *Drug Disc. Today: Disease Mech.* 2:327–330.
45. Thomas V D, Yang F C, Kvedar J C (2005). Biologics in psoriasis: A quick reference guide. *J. Amer. Acad. Dermatol.* 53:346–351.
46. Gómez-Pérez F J, Rull J A (2005). Insulin therapy: Current alternatives. *Arch. Med. Res.* 36:258–276.
47. Gerich J E (2002). Novel insulins: Expanding options in diabetes management. *American J. Med.* 113:308–316.
48. Feemester R (2006). Holding their breath: Pipeline watch. *Pharmceut. Exec.* 26:58–62.
49. Girard M P, Cherian T, Pervikov Y, et al. (2005). A review of vaccine research and development: Human acute respiratory infections. *Vaccine.* 23:5708–5724.

50. Manoj S, Babiuk L A, van Drunen Littel-van den Hurk S (2004). Approaches to enhance the efficacy of DNA vaccines. *Critical Reviews in Clinical Laboratory Sciences*. 41:1–39.
51. Plotkin S A (2005). Vaccines: Past, present and future. *Nature Med. Suppl*. 11:s5–s11.
52. Nahta R, Esteva F J (2005). Herceptin: Mechanisms of action and resistance. *Cancer Letters*. In Press, Corrected Proof, Available online 30 March 2005.
53. Ferrara N, Hillan K J, Novotny W (2005). Bevacizumab (Avastin), A humanized anti-VEGF monoclonal antibody for cancer therapy. *Biochem. Biophy. Res. Commun*. 333:328–335.
54. Erlichman C, Sargent D J (2004). New Treatment Options for Colorectal Cancer. *N. Engl. J. Med*. 351:391–392.
55. Silvestrelli G, Lanari A, Parnetti L, et al. (2006). Treatment of Alzheimer's disease: From pharmacology to a better understanding of disease pathophysiology. *Mech. of Ageing and Devel*. 127:148–157.
56. Lipton S A (2004). Failures and successes of NMDA receptor antagonists: Molecular basis for the use of open-channel blockers like memantine in the treatment of acute and chronic neurologic insults. *NeuroRX*. 1:101–110.
57. Maggon K, Barik S (2004). New Drugs and treatment for respiratory syncytial virus. *Rev. Molec. Virology*. 14:149–168.
58. Tongren J E, Zavala F, Roos D S, et al. (2004). Malaria vaccines: If at first you don't succeed . . . *Trends in Parasitology*. 20:604–610.
59. Girard M P, Osmanov S K, Kieny M P (2006). A review of vaccine research and development: The human immunodeficiency virus (HIV). *Vaccine*. In Press, Corrected Proof, Available online 28 February 2006.
60. Lowrie D B (2006). DNA vaccines for therapy of tuberculosis: Where are we now? *Vaccine*. 24:1983–1989.
61. Okaji Y, Tsuno N H, Saito S, et al. (2006). Vaccines targeting tumour angiogenesis—A novel strategy for cancer immunotherapy. *Euro. J. Surgical Oncology*. In Press, Corrected Proof.
62. Maggon K (2003). Activated Protein C in treating sepsis. *Express Pharma. Pulse*. Mumbai February 27, 2003, Express Healthcare Management Mumbai, March 1, 2003.
63. Maggon K (2003). Risk benefit assessment of APC. *Express Pharma. Pulse*. Mumbai. March 6, 2003. Express Healthcare Management Mumbai, April 1, 2003.
64. Griffin J H, Fernández J A, Mosnier L O, et al. (2006). The promise of protein C. *Blood Cells, Mol. Dis*. In Press, Corrected Proof, Available online 7 February 2006.
65. Cacciotti J, Shew B (2006). Pharma's next top model. *Pharm. Exec*. 26:82–86.
66. Schrag D (2004). The price tag on progress—Chemotherapy for colorectal cancer. *N. Engl. J. Med*. 351:317–319.
67. Angell M (2004). The truth about drug companies. Random House, New York, p. 336.
68. Maggon K (2004). Regulatory reforms and GCP Clinical Trials with New Drugs in India. *Clinical Trials*. 1:1–7.
69. Gardner C, Gardner C (2004). Technology licensing to nontraditional partners: Non-profit health product development organizations for better global health. Available: http://www.mihr.org/?q=taxonomy_menu/1/6.
70. Moran M (2005). The new landscape of neglected disease drug development. *Wellcome Trust, LSE and OECD Report*. Available: http://www.wellcome.ac.V

71. Maggon K (2007). Market intelligence and pharma data mining. http://www.medicineandbiotech.com. e-Published, Feb 1, 2007.
72. Maggon K, Maggon S (2007). Overview of the pharmaceutical and biotechnology industry performance in 2006. http://www.medicineandbiotech.com. e-Published, March 1, 2007.

# FROM DEFINING BIOINFORMATICS AND PHARMACOGENOMICS TO DEVELOPING INFORMATION-BASED MEDICINE AND PHARMACOTYPING IN HEALTH CARE

Ioannis S. Vizirianakis
*Aristotle University of Thessaloniki, Thessaloniki, Greece*

**Chapter Contents**

1 Introduction
2 Pharmacogenomics and Pharmacotyping for Complex and Polygenic Diseases
3 Pharmacotyping Concepts and CYP-mediated Metabolism of Drugs
4 Pharmacotyping Concepts for Modulating the Pharmacodynamics and Pharmacokinetics of Drugs
5 Cancer Pharmacogenomics and Pharmacotyping
6 The Challenges of Pharmaceutical and Medical Education to Meet the Central Dogma of Pharmacotyping as well as of Personalized and Information-Based Medicine
7 The Development of Information-Based Platforms to Ensure Major Benefits for Pharmacotyping in Clinical Practice
8 Concluding Remarks
References

## 1 INTRODUCTION

Over the last three decades, the scientific efforts mainly focused on molecular biology, genetics, recombinant DNA (rDNA) technology, and lastly genomics and

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

bioinformatics have had a huge impact on our understanding of disease pathophysiology and of drug actions at the molecular level. In addition, the accumulated genomic knowledge has provided the ability to predict drug–drug interactions and the emergence of adverse drug reactions (ADRs) in clinical practice, a fact of great importance in modern pharmacology and therapeutics. The elucidation of specific clinical- as well as genetic-related factors predisposing to individual drug–drug interactions, ADRs, or drug disposition also gives new dimensions to new drug development and pharmacotherapy [1–5]. Furthermore, the evolution of genomic technologies offered new tools for extracting drug-related information especially through the emergence of pharmacogenomics in recent years [6–9] (for term definition in pharmacogenetics, pharmacogenomics, and other drug-related areas, see the website: http://www.genomicglossaries.com/). Pharmacogenomics aims to strengthen drug efficacy and to prevent most, if not all, ADRs from emerging during pharmacotherapy. Furthermore, pharmacogenomics provides the methodology and the knowledge to medical practitioners to genetically test each patient before the administration of the drug of interest, a practice that obviously can lead to personalized drug therapy. To this end, improved pharmacotherapy outcomes are expected by anticipating and minimizing toxicity related to drug delivery in everyday patient care. By integrating the genomic drug-related data into drug delivery, pharmacogenomics is moving toward the application of pharmacotyping in drug prescriptions, where clinical and genomic information will be used to ensure maximum efficacy and safety in clinical practice, as it has been recently proposed [10].

Interestingly, this is happening in parallel with the efforts focusing on the creation and establishment of unified information-based platforms in medicine, a fact that obviously will accelerate the technological advancement for the integration of genomic data in clinical practice. The latter has been recently facilitated through the technological developments that now allow the application of molecular diagnostics and devices at the nanoscale level. Such a direction, however, stresses the need for a multidisciplinary approach in future health care by permitting nanomedicine to apply molecular diagnostics and new innovative drugs in a way that is beneficial in routine medical practice both in disease diagnosis and in therapeutics. This is a major advantage, because the development of nanotools is in high demand in information-based medicine and the application of suitable platforms for the exploitation of genomic information in medicine and pharmacy. Thus, the intercorrelation of the molecular knowledge extracted from the clinical, genomic, and technological health-related areas as well as the transformation into a form readily applicable in health care can be achieved. As a matter of fact, the evolvement of nanomedine seems to go in parallel with the capability of genomic technologies to transform medicine and pharmacy, and in this regard, their use depends on well- and properly educated practitioners able to work in a well-organized and information-based clinical infrastructure. Collectively, this approach in pharmacotherapy belongs to the broader concept of personalized medicine because its major goals are considered the anticipation and minimization of the individual risk of disease onset and progression as well as the individualized drug therapy for improved patient care [11].

Overall, by using the principles of pharmacogenomics, personalized medicine, and pharmacotyping in patient care, and by monitoring their impact on current drug development and delivery, a major benefit is expected upon co-evaluation of

genomic with clinical data to improve pharmacotherapy outcomes. In this regard, the clinical validation of genomic data for routine drug prescription, the training of future health-care professionals, and the integration of molecular medicine into clinical trials for new drug development have to be defined and clearly addressed. Also, the need for extensive discussions about the ethical, societal, and economic impacts arising from the application and use of genomic data in patient care is now, more than ever, a necessity. The latter stresses the need for the proper education of health-care professionals in order for them to actively be involved in, and positively influence, all these issues in health care. The experience already gained has shown that this can be achieved in the near future, because the first pharmacogenomics test (AmpliChip Cytochrome P450) approved by the U.S. Food and Drug Administration (FDA), recently, to improve the delivery of drugs that are substrates or inhibitors of cytochrome P450 (CYP)2D6 and/or CYP2C19 has been introduced in everyday clinical use. Also, recently released on March 22, 2005 by the FDA, the guidelines for industry concerning the submission of pharmacogenomics data upon the process of application of new drug development, a fact expected to have a major impact on the flow and generation of data in clinical pharmacogenomics and the application of pharmacotyping in drug prescription. This means that the detection of single nucleotide polymorphisms (SNPs) or other genetic variations and the establishment of specific genotypes and haplotypes for several genes must be first defined and clinically validated at such a level as to explain why one patient responds well in drug therapy, another does not, or even another experiences serious ADRs. In fact, this will be a major transition in pharmacotherapy because pharmacotyping will be finally achieved, and, to this end, the medical practitioner in collaboration with other specialties will also be based on patient's genotyping–haplotyping analysis data for initiating and maintaining drug dosage therapeutic profiles for individual patients [5, 10]. However, in order for these new directions in therapeutics to succeed with maximum benefits to patients and society, the future advances in pharmacogenomics and bioinformatics have to first allow the smooth digital integration and second the easy end-user utilization of genomic and clinical data in unified information-based platforms within the healthcare system.

## 2 PHARMACOGENOMICS AND PHARMACOTYPING FOR COMPLEX AND POLYGENIC DISEASES

The concept of ensuring maximum efficacy and safety upon drug delivery is the long-desirable target of pharmaceutical care that implies improved pharmacotherapy profiles and outcomes in clinical practice and coincides chronologically with the establishment of pharmacology as a basic and clinical discipline. Also, the ability to lower the incidence of drug toxicity and the emergence of ADRs is a fundamental issue for the empowerment of health care in terms of patient's quality of life and cost (Box 1). Nowadays, this situation in pharmacotherapy has been, more or less, closer to reality. This happens due to the rapid development of genomic and information-based technologies and their application in pharmacotherapy and clinical practice (Figure 1). For example, it is now well accepted that, in terms of drug toxicity, many ADRs arise because of inter-individual genetic differences in drug metabolizing enzymes, drug transporters, ion channels,

**BOX 1 MAJOR POINTS RAISED IN HEALTH CARE UPON THE INCIDENCE AND THE SEVERITY OF ADRS IN CLINICAL PRACTICE**

- In a meta-analysis from 39 prospective studies in U.S. hospitals, it has been proposed that 6.7% of hospitalized patients developed ADRs during pharmacotherapy and 0.32% showed fatal ADRs, the latter causing about 100,000 deaths per year in the United States [12].
- The report released from the Institute of Medicine in December 1999 has shown that nearly 98,000 deaths in the United States annually were attributed to medication errors including ADRs [13].
- Furthermore, 10–17% of patient hospitalizations are directly related to the emergence of ADRs, an effect that poses patient pharmaceutical care and safety as a fundamental issue in each healthcare system [14].
- It has been estimated in the United States that the cost attributed to the emergence of ADRs is approximately $100 billion, a fact that further stresses for the improvement of drug delivery in clinical practice [15].
- In a 2-year period of study conducted in by the Department of Internal Medicine in a Norwegian hospital and from 732 deaths reported amongt the 13,993 in-patients, 133 of them were directly attributed to ADRs observed during pharmacotherapy [16]. This result gives almost 10 deaths per 1000 hospitalized patients, a number that urgently demands an improvement of drug delivery by strengthening drug efficacy and minimizing toxicity in the healthcare system.
- It has also been estimated that about 7% of all hospital admissions in the United Kingdom and Sweden are due to ADRs developed in clinical practice [15].

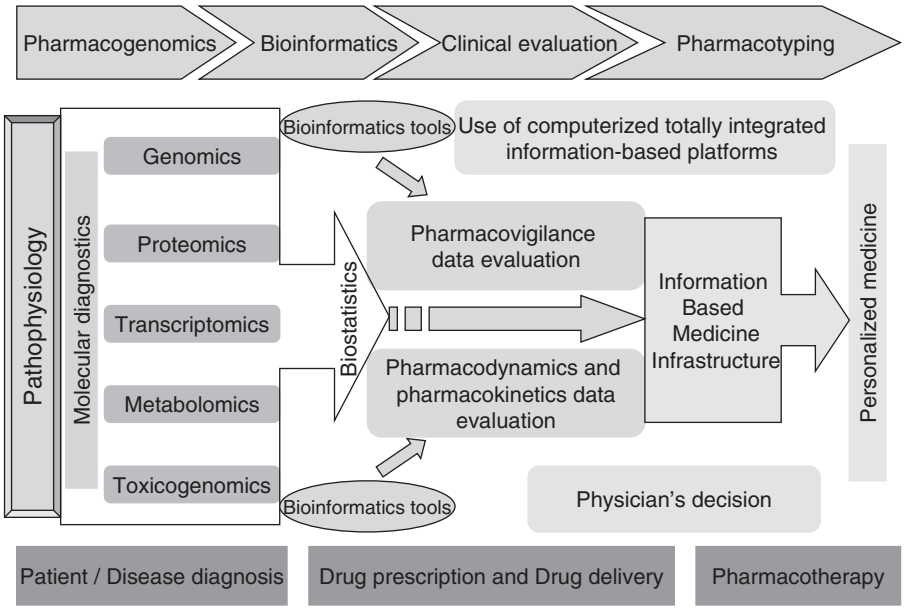


**Figure 1.** Pharmacotyping and information-based medicine in postgenomic drug delivery and health care (see the text for details).

receptors, and other drug targets [1, 2, 5, 10, 17, 18]. Furthermore, several pharmacogenomics studies have revealed various examples of clinically relevant genetic polymorphisms that are associated with altered drug response in patient care. Representative examples of such polymorphic gene variants of pharmacogenomics relevance are shown in Box 2. Although these intriguing data are very interesting in trying to understand the correlation between genetic make-up and various drug effects in the body, much work is needed for this type of drug-related genomic information to be transformed, integrated in everyday clinical practice, and, finally be available on a broader basis in health care.

Through the sophisticated technological advances achieved thus far, the genomic knowledge is entering clinical practice and being used to analyze complex diseases including cardiovascular diseases, asthma, cancer, and degenerated disorders [38–42]. Furthermore, the broad use of automated DNA sequencing techniques and the application of DNA microarray (DNA chip) systems permit the rapid and reliable assessment of a large number of genes implicated in drug response and/or disease phathogenesis [43–45]. Alternatively, and of equal importance, specific gene variants have been found either to predispose for, to contribute to, or even to increase the susceptibility of individual persons for disease development and progression. Unanimously, the availability of DNA microarray technology helped genomics research to achieve widespread interest particularly in clinical sciences and pharmacotherapy. As a matter of fact, it is a valuable technique for the improvement of clinical diagnosis and the understanding of molecular etiology and pathophysiology of polygenic and multifactorial conditions, e.g., cardiovascular diseases and degenerating disorders, because it gives researchers the ability to assess and evaluate each time the expression of nearly all genes presented in the human genome. For example, there have been several attempts, in the past few years, to establish specific polymorphic gene variants at pharmacodynamic loci predisposing to cardiovascular diseases or to affect drug response such as angiotensin-converting enzyme (ACE), angiotensin II type-1 receptor (AT-1), apolipoprotein E (APOE), α-adducin, and adrenergic receptors that could alternatively be used to improve the pharmacotherapy profiles of drugs exerting their effects through binding to these receptors [46–49]. It is also very interesting for anyone to see now, that genomic studies, throughout the last period of 10–15 years, have revealed several chromosomal loci and genes implicated for example in cardiomyopathies, arrhythmias, heart failure, and hypertension. The identification of several genes whose function is associated with diseases like hypertrophic or dilated cardiomyopathy, human long-QT syndrome, and essential hypertension give new insights on a molecular basis to heart disease progression and may help in orienting the proper drug therapies in a given population. Among them are included genes encoded for β-myosin heavy-chain, cardiac troponins and α-tropomyosin in hypertrophic cardiomyopathy; dystrophin, tafazzin, and actin in dilated cardiomyopathy; cardiac sodium channel (SCN5A) and potassium channel (HERG) in long-QT syndrome; as well as ACE, α-adducin, angiotensinogen, and $\beta_2$-adrenergic receptor ($\beta_2$AR) in essential hypertension [33, 39, 50]. Furthermore, even more impressive is the fact that, in the case of cardiomyopathies, genomic analysis has been able to establish specific molecular gene expression profiles that clearly define different molecular classes of cardiomyopathy such as of sarcomyopathy, cytoskeletalopathy, and cytokinopathy [51–54]. However, although these represent intriguing results in genomic

**BOX 2 GENETIC POLYMORPHISMS IN SPECIFIC GENES ASSOCIATED WITH VARIOUS DRUG RESPONSE AND EMERGENCE OF ADRS IN CLINICAL PRACTICE***

Genetic polymorphisms of pharmacogenomics importance that can modulate the pharmacological response of specific drugs implicate the following genes:

1) *CYP2C9* that has been implicated in warfarin, phenytoin, tolbutamide, and/or glipizide has had various clinical responses [1].
2) *CYP2C19* that has been involved in proton pump inhibitors (e.g., omeprazole) has had an altered clinical response [19].
3) *ABCB1* gene that has been associated with the development of molecular resistance of anti-epileptic drugs [20, 21].
4) UDP-glucuronosyltransferase 1A1 (*UGT1A1*) that has been involved in irinotecan toxicity [20, 22].
5) Human leukocyte antigen (*HLA-B*) that has been associated with abacavir hypersensitivity [23, 24].
6) Glutathione-S-transferases (*GSTs*) that have been correlated with platinum chemotherapeutic agents and D-penicillamine altered response [25, 26].
7) N-acetyltransferases (*NATs*) that have been implicated with the emergence of sulfonamides hypersensitivity and the toxicity observed in some cases of isoniazid, procainamide, and hydralazine administration [27].
8) HERG potassium channel (*KCNH2*) that has been found to modulate quinidine and/or cisapride safety in clinical practice as well as KvLQT1 potassium channel (*KCNQ1*) that causes a similar effect on terfenadine and disopyramide response [28, 29].
9) Thiopurine methyltransferase (TPMT) activity has been well correlated with 6-mercaptopurine, 6-thioguanine, and azathioprine toxicity seen in some patients [1, 30, 31].
10) Several genetic polymorphisms of serotonin *5-HT2A* and *2C* receptors have been implicated with various pharmacological responses to clozapin delivery [32].
11) The angiotensin II type I-receptor (*AT-1*) gene has been associated with the altered drug response of losartan seen in certain patients [33].
12) The apolipoprotein E4 (*APOE4*) allele has been correlated with the variability of tacrine pharmacotherapy [34].
13) The gene of 5-lipoxygenase (*ALOX5*) has been studied in relation to zileuton and montelukast altered drug response [28].
14) Several genetic polymorphisms of $\beta_2$-adrenergic receptors ($\beta_2$*AR*) have been shown to modulate the $\beta_2$-agonist (e.g., albuterol) effect in clinical practice [35, 36].
15) The ABCG2 transporter gene, a member of the ATP-binding cassette transporters, in which an Arg to Gly mutation at amino acid 482 (G482 mutation) has been shown to confer high-level resistance to various antifolate chemotherapeutic agents [37].

* Modified from Ref. 5.

medicine, the recognition of the complexity of such diseases clearly demands the development and clinical validation of suitable genetic diagnostic markers; the application and use of more sophisticated, clinically validated genomic and bioinformatic tools; as well as the design of prospective pharmacogenomics studies, in order to further advance the molecular understanding of the disease pathophysiology and to improve the therapeutic intervention in clinical practice.

Regarding the degenerating diseases, the structural and functional genomics analysis of patients with Alzheimer's disease (AD) has recently identified several genetic loci that potentially contribute to disease pathophysiology in cooperation with both environmental and epigenetic factors [40]. In particular, genetic predisposition to AD development is demonstrated either for genes whose variants exhibit a mendelian inheritance pattern, e.g., the amyloid precursor protein (APP), and the presenilin-1 (PS1) and -2 (PS2), or for those considered to be potential susceptibility genes contributing to AD predisposition, e.g., APOE, ACE, interleukin-1α (IL1α), nitric oxide syntase-3 (NOS3), low-density lipoprotein-related protein 1 (LRP1), and α-2-macroglobulin (A2M). Furthermore, recent results have suggested that specific genotypes and haplotypes may cause different drug response rates in individual AD patients receiving cholinesterase inhibitors and/or non-cholinergic agents used for their therapeutic intervention [34, 40]. This progress in AD genomics clearly shows that pharmacogenomics analysis of complex and polygenic disorders has started to give valuable clinical results. As a matter of fact, the broader application of pharmacogenomics can be an alternative and suitable platform for further assessing the pathophysiology and molecular diagnosis of complex diseases, as well as for the identification of new drug targets to improve healthcare outcomes and patient's quality of life. To this end, however, as in the case of cardiovascular diseases, the establishment of clinically validated specific molecular diagnostic biomarkers and pharmacotyping profiles for individual patients is still a demand, in order for these efforts to be finally achieved in clinical practice.

## 3 PHARMACOTYPING CONCEPTS AND CYP-MEDIATED METABOLISM OF DRUGS

By considering the example of CYPs that are involved in the phase I drug metabolism in the body, the importance of pharmacogenomics studies and their clinical relevance in drug disposition will be further defined. Nowadays, 57 putative functional genes encoding different CYP enzymes and 58 pseudogenes have been characterized [55, 56] (see also the website: http://drnelson.utmem.edu/CytochromeP450.html). The major human CYP enzymes involved in drug metabolism include CYP1A2, CYP2A6, CYP2B6, CYP2C8, CYP2C9, CYP2C19, CYP2D6, CYP2E1, CYP3A4, and CYP3A5 that exhibit several polymorphic sites of clinical significance (see the website: http://www.imm.ki.se/cypalleles). Indeed, several experimental studies of applied pharmacogenomics research have already been published that correlate specific CYP gene variants with altered drug efficacy and safety and clearly shed light on the elucidation of the molecular mechanisms involved in inter-individual drug response observed among patients suffering from the same disease. The existed complexity at the genomic level in drug metabolism is additionally complicated by considering the high number of polymorphic gene

variants encoding CYP enzymes discovered, especially those belonging into CYP families 1, 2, and 3. It has been estimated that enzymes belonging to these families mediate about 70–80% of all phase I-dependent metabolism of clinically used drugs. Interestingly, polymorphic enzymes that mainly account for ~40% of CYP-mediated drug metabolism are those of CYP2C9, CYP2C19, and CYP2D6 and, for this reason, are considered very important in pharmacotherapy by making problematic the application of a general drug dosing scheme for the drugs whose disposition is determined by these enzymes [15]. Similarly, by evaluating the incidence of ADRs, it has been shown that ~56% of drugs cited in ADR-related studies are metabolized by polymorphic enzymes of phase I, in which ~86% account for the CYP-mediated metabolism. On the contrary, only 20% of drugs that are substrates of non-polymorphic enzymes are cited in the ADRs reports [15, 57].

It is now well known that the genetic variability observed in CYP enzymes corresponds into four phenotypes of clinical relevance in drug delivery. The definitions used to identify these four phenotypes of CYP-related drug metabolism are as follows:

1. Poor metabolizers (PMs) for those individuals that have shown no enzymatic activity in diagnostic tests. Therefore, this is a phenotype that refers to people who lack the functional CYP enzyme and slowly metabolize the drug substrates, allowing their concentration to reach high plasma levels.
2. Intermediary metabolizers (IMs) for those individuals with reduced enzymatic activity. These people are heterozygous for one deficient allele or carry two alleles that cause lower CYP enzyme activity.
3. Extensive metabolizers (EMs) for individuals with physiological enzymatic activity, a fact that suggests the existence of two wild-type (normal) alleles allowing normal drug metabolism.
4. Ultrarapid metabolizers (UMs) for individuals having higher than normal enzymatic activity. Therefore, this is the phenotype applied to people carrying multiple gene copies of the respected enzyme. That trait is dominantly inherited and allows these individuals to rapidly metabolize the drug substrates and to achieve low, and sometimes inefficient, plasma levels.

Several examples of drugs whose dosage is related to CYP phenotypes have already been reported that further envisage on the importance of CYP polymorphisms for routine drug prescription and the modulation of pharmacotherapy outcomes. For example, CYP2D6 and CYP2C19 are responsible for the metabolism of most psychoactive drugs, including antidepressants. In particular, specific dosage recommendations for patients with EM, IM, and/or PM phenotypes of either CYP2D6 or CYP2C19 have been suggested for 14 antidepressant drugs [58]. Furthermore, the kinetics of nortriptyline is dependent on the number of active CYP2D6 genes, and the dosage required to reach the same plasma levels varies from 30–50 mg in CYP2D6 PMs to 500 mg in UMs [15]. Another example applies to sertraline, a substrate of CYP2C19 (and an inhibitor of CYP2D6), because CYP2C19 PMs develop ADRs (nausea and dizziness), a fact that might be attributed to increased sertraline plasma concentrations. In addition, the metabolism of valproate (commonly used to treat bipolar disorders) has been found to be influenced by CYP2C9

polymorphism, an effect of clinical relevance. Thus, the broader application of CYP-related pharmacogenomics knowledge to implement clinical experience can improve pharmacotherapy outcomes through the establishment of individualized dosage-regimens that take the difference in drug metabolic capacity into account.

Similarly, the fundamental role of CYP-related knowledge in minimizing the incidence of drug–drug interactions in clinical pharmacology and the application of such information in improving pharmacotherapy outcomes is now well appreciated in drug delivery and can be further defined by the example of cholinesterase inhibitors (rivastigmine, tacrine, donepezil, and galantamine), which are agents used to treat symptoms of patients suffering from AD. By assessing the drug–drug interactions of this group of agents published in several clinical studies, it can be suggested that the individual CYP-isoform implicated in the metabolism of each cholinesterase inhibitor also specifies the type of interacting drugs referred to modulate its pharmacokinetic parameters and clinical outcome [59, 60]. In particular, because the metabolism of rivastigmine is mediated predominantly by esterases rather than by hepatic CYP enzymes, clinically relevant drug interactions implicated by this phenomenon are unlikely. This fact for rivastigmine was indeed verified by *in vitro* assays and pharmacokinetic studies conducted in humans. To this regard, in a retrospective analysis including four major clinical trials with 2459 patients, it has been shown that no increase in adverse events occurred among those patients who received 22 different therapeutic classes of drugs commonly used in elderly populations, including antihypertensive, anti-histamine, anti-inflammatory, or anxiolytic agents [59].

Tacrine is primarily metabolized by the hepatic CYP1A2 (and to some extent by CYP2D6), so substrates, inhibitors, or inducers mainly of CYP1A2 are expected to alter its bioavailability. This is indeed the case: (1) with fluvoxamine (50 or 100 mg/day), which is a potent CYP1A2 inhibitor that caused ~85% reduction of tacrine metabolism; (2) with cimetidine, also a CYP1A2 inhibitor, where it has been observed that plasma concentrations of tacrine were increased by ~30%, so a reduction of its dosage has been advised when co-administered with cimetidine; (3) with estradiol also metabolized by CYP1A2 that caused an increase of ~60% in AUC and of ~46% in the mean $C_{max}$ of tacrine. Following these data, it has also been reported that tacrine decreased by ~50% the clearance of theophylline, which is also a substrate of CYP1A2, so reduction of theophylline dosage has been suggested upon its co-administration with tacrine. In addition, tobacco smoking that induces CYP1A2 metabolism significantly accelerates tacrine metabolism, a result indicating that clinicians should be aware that smokers may require higher doses of tacrine than that would be given to nonsmokers [60].

Similarly, in the case of donepezil that is primarily metabolized by the hepatic CYP3A4 and CYP2D6, it is expected that substrates, inhibitors, or inducers of CYP3A4 and of CYP2D6 can alter its bioavailability. This prediction is indeed verified in certain pharmacotherapy dosage regimens: (1) with ketoconazole (200 mg/day), a potent inhibitor of CYP3A4, where a significant increase (23–30% at steady state) in the plasma concentrations of donepezil was observed; and (2) with paroxetine and sertraline, which are both of them CYP2D6 inhibitors, in which a careful patient monitoring has been recommended upon co-administration, due to possible interaction of donepezil with these drugs through inhibition of CYP2D6. On the contrary, upon the delivery of donepezil with

risperidone and/or thioridazine, the latter two are substrates, but not inhibitors, of CYP2D6, and no significant interaction was observed. The same has been shown with digoxin (no significant CYP metabolism), theophylline (substrate of CYP1A2), warfarin (substrate of CYP2C9), and/or cimetidine (inhibitor of CYP1A2 and CYP2C19) [60].

By proceeding through the analysis to galantamine, similar results have been obtained. Galantamine is also metabolized by CYP3A4 and CYP2D6, so again the potent inhibitors of these isoenzymes are expected to result in significant drug–drug interactions. Such interaction was indeed observed: (1) with paroxetine (an inhibitor of CYP2D6 and CYP2C19) that caused an increase in the bioavailability of galantamine by 40%, as it has been reported upon their co-administration; and (2) with ketoconazole (an inhibitor of CYP3A4 and CYP2C19) that exerted a significant increase in the bioavailability of galantamine by 40%. However, no significant interaction of galantamine with warfarin and digoxin has been observed in clinical studies [60].

The above-mentioned example of cholinesterase inhibitors clearly shows the predictive value of CYP-related information during everyday drug prescription, because the avoidance of co-administration of drugs implicated with the same CYP isoform can ultimately improve pharmacotherapy outcomes, at least for drugs whose metabolism is the rate-limiting step upon the establishment of their therapeutic concentrations in the body. However, and more importantly, this knowledge has already started to implement routine clinical practice, since the first pharmacogenomics test was approved by the FDA on December 23, 2004. The "AmpliChip Cytochrome P450 Genotyping test" permits the identification of 31 different polymorphisms of CYP2D6 (29 polymorphisms) and CYP2C19 (2 polymorphisms), and it is used along with clinical evaluation and other tools to determine the best treatment options for patients taking drugs that are either substrates or inhibitors of these metabolizing enzymes. The development of this test is of great importance in pharmacotherapy, because it represents the first step toward the broader application of pharmacotyping concepts in drug prescription.

## 4 PHARMACOTYPING CONCEPTS FOR MODULATING THE PHARMACODYNAMICS AND PHARMACOKINETICS OF DRUGS

Interestingly, and in addition to CYP gene variants already mentioned in the previous section, specific genetic polymorphisms were also analyzed in genes encoding P-glycoprotein (P-gp), G-protein coupled receptors (GPCRs), thiopurine methyltransferase (TPMT), ACE, arylamine N-acetyltransferases (NATs), and/or UDP-glucuronosyltransferases (UGTs), just to mention a few, which have also been shown to correlate with altered drug response and incidence of ADRs in humans [20,22, 27, 30, 35, 46, 61–63], (see also Box 2). To better clarify the correlation of such gene variants with the pharmacotherapy outcome, experimental data related to either the ABC-type transporters function or to the variability of anti-asthmatic drug delivery will be discussed. P-gp is an ATP-dependent transporter (efflux pump) belonging to the superfamily of ABC transporters that includes 48 already identified members in humans. The role of P-gp in limiting intestinal, brain, and placental transport, as well as in biliary and urinary excretion of its substrates, is

now well recognized. Thus, variations in its gene (*ABCB1* or *MDR1*) that existed in several individuals may influence drug disposition and alter drug response [20, 63]. These genetic variations can affect the function of P-gp as transporter by altering its binding with drugs and other xenobiotics [20]. Confirmatory to this notion is the mutant variant of *ABCB1* designated as ABCB1-C3435T, which has been characterized recently and has been shown to be well associated with the variability of antiepileptic drug response [21]. At the same time, further functional characterization of specific P-gp gene variants, or other ABC-type transporters, and elucidation of the molecular mechanisms that underlined their involvement in interindividual drug variability are expected to give new insights into the selective modulation of the P-gp function in the blood-brain barrier for optimizing therapeutic intervention and improving patient pharmaceutical care [64]. The clinical relevance of the mutant alleles of ABC-type transporters in pharmacotherapy was further confirmed in another member of this class. Recent studies have established that an Arg to Gly mutation at amino acid 482 (G482 mutation) in the ABCG2 transporter confers high-level resistance to various antifolate chemotherapeutic agents (e.g., methotrexate) [37].

Nowadays, specific polymorphisms in individual genes have been studied to explain the variability seen upon anti-asthmatic drug delivery, like $\beta_2$-agonists (see also Box 2). For example, it has been estimated that as much as 60% of interindividual variability in anti-asthmatic drug response observed with $\beta_2$-agonist delivery might involve genetic factors modulating the pharmacotherapy outcome [35]. Thirteen SNPs identified in the promoter and the coding region of $\beta_2$AR gene were found to be organized into 12 haplotypes. Some of these haplotypes have shown significant divergence in different ethnic populations, and this fact is associated with the bronchodilator response to $\beta_2$-agonists in patients with asthma [36]. As asthma is considered a complex and polygenic disease, several other gene variants have also been shown to contribute to the pharmacotherapy outcome in asthmatics in addition to $\beta_2$*AR* polymorphisms. Such genes include those of 5-lipoxygenase (*ALOX5*) for zileuton, muscarinic receptor-2 ($M_2$) and 3 ($M_3$) for muscarinic antagonists (e.g., ipratropium bromide), glucocorticoids receptor for glucocordicoids (e.g., prednisolone, beclomethasone), and phosphodiesterase-4A (*PDE4A*) and 4D (*PDE4D*) for theophylline delivery [65]. However, although these correlations add valuable knowledge to basic and clinical pharmacology, more systematic efforts at the molecular level must be undertaken toward thorough elucidation of the underlined mechanisms of genetic drug response variability, in order for the anti-asthmatic therapy and outcome to be finally enriched and improved.

## 5 CANCER PHARMACOGENOMICS AND PHARMACOTYPING

Recent progress in cancer pharmacogenomics suggests that improvement of chemotherapy outcomes for specific anticancer drugs, like 5-fluorouracil (5-FU), irinotecan, platinum agents, and thiopurine-like drugs, can be achieved. In particular, specific polymorphisms in genes encoding the enzymes dihydropyrimidine dehydrogenase and thymidylate synthase have been shown to influence the efficacy and toxicity of 5-FU in certain individuals. The applicability, however, of such an approach by selecting patients who are likely to tolerate and respond to 5-FU

therapy remains very complicated, and it is not easily attainable for everyday care [20, 66]. Furthermore, it has been proposed that polymorphisms in the promoter region of the gene encoded UDP-glucuronosyltransferase 1A1 (*UGT1A1*) may be clinically useful for predicting severe patients' toxicity to irinotecan, thus making this drug an excellent candidate for individualized therapy, as it has been recently proposed [67]. Irinotecan belongs to the camptothecin class of topoisomerase I inhibitors and has been shown to have potent activity against many types of solid human tumors, in particular, gastrointestinal and pulmonary malignancies [22, 67]. It is a prodrug that is converted in the body by carboxyesterase-2 to SN-38 metabolite with more than 1000-fold enhancement of cytotoxic activity. The rate-limiting step of SN-38 pharmacokinetics is considered the UGT glucuronidation reaction *via* hepatic UGT1A1, whereas the presence of the homozygous *UCT1A1*28* genotype in patients as a molecular predictor of toxicity risk (the promoter region contains seven TA repeats instead of six in the normal wild-type genotype) seems to be associated with high risk of grade 4 neutropenia in clinical practice [67]. The predictive value of the *UCT1A1*28* genotype, in order clinicians to address more definitive guidelines regarding the effective and safe dose of irinotecan in these patients, however, remains to be established.

Also, in the case of platinum chemotherapeutic agents (cisplatin, carboplatin, and oxaliplatin), several polymorphic genes have been implicated in their efficacy and toxicity. These genes are the *XPD* and *XRCC1* that encode proteins involved in the cellular DNA excision repair system and that of glutathione-S-transferases (GSTs) [66]. Although these results can be con-sidered as positive indicators for the improvement of cancer patient care, additional work must be done in order for the general applicability and acceptance of genetic information related to platinum agents response to be established in clinical practice.

However, one of the most studied examples of applied pharmacogenomics that clearly shows how the translation of genomic information is being transferred to guide patient therapeutics is related to TPMT gene polymorphisms. Interestingly, specific TPMT gene variants have been isolated and clinically validated in studies showing that their presence is associated with high-risk toxicity in individuals taking thiopurine drugs (6-mercaptopurine, 6-thioguanine, azathioprine) [30, 31, 68]. Currently, eight TPMT alleles have been identified, including three variant alleles designated as TPMT*2 (with G238C mutation), TPMT*3A (with G460A and A719G mutations), and TPMT*3C (with A719G mutation) that account for about 80–95% of intermediate or low TPMT enzymatic activity, and for this reason, they are analyzed during genotyping and haplotyping population studies [30]. The pharmacogenomics data regarding the TPMT deficiency correlated well with thiopurine drug toxicity and were so clear and convincing that a TPMT genetic test was developed for clinical use before the initiation of 6-mercaptopurine delivery to children suffering from acute lymphoblastic leukemia (ALL) [69, 70]. The clinical usefulness of this genetic test was based on pharmacogenomics data suggesting that the calculation of a drug dosage scheme could be achieved though the identification of the specific TPMT genotype in each child. However, the difficulty of applying the genetic test in clinical practice was further demonstrated by the discussions after the attempts to apply TPMT pharmacogenomics data upon thiopurine drug prescription. Such discussions have been extended even to the need and usefulness of the genetic testing as compared with the traditional biochemical tests,

e.g., the use of a red-cell assay for the detection of TPMT enzyme deficiency [11, 70]. The example of a TPMT genetic test clearly shows the difficulties that these new concepts of drug delivery will meet before their general acceptance in routine everyday use. Undoubtedly, to ascertain for applied pharmacogenomics a final success, more systematic work and advancement in information-based technologies are needed. In turn, this will allow better transformation of genomic data into clinical forms to finally lead into the wide use of genetic testing with high-quality and validity outcomes, thus ensuring specific dosage recommendations of individual drugs in routine pharmaceutical care [71, 72].

Of special interest, however, related to the application of pharmacotyping in anticancer therapeutics are issues concerning the development of drug resistance, as this phenomenon consistently affects the pharmacotherapy outcomes and patient's quality of life. Over the past several years, novel exploitable targets for cancer drug development have been revealed that recently resulted in the development of some innovative drugs already applied to clinical practice. Such targets include enzymes like tyrosine kinases involved in signal transduction pathways, genes regulating apoptosis and differentiation of malignant cells, cell-lineage transcriptional factors, angiogenesis factors, and unique proteins driving the cell cycle machinery [73–78].

The development of imatinib (Gleevec) as an effective agent capable of inducing cellular apoptosis for the treatment of patients suffering from chronic myelogenous leukemia (CML) represents such a bright example of molecularly targeted anticancer therapeutics [79]. Gleevec is a selective inhibitor of Bcr-Abl fusion tyrosine kinase specifically present in CML patients that results from a reciprocal translocation between chromosomes 9 and 22. This translocation causes the genetic fusion of *bcr* with exon-1 of *c-abl* resulting in a defect known as Philadelphia chromosome. Although the use of imatinib give very impressive and promising results for the therapy of CML as well as for other tumors by also inhibiting the c-kit kinase, the development of drug resistance causes a real restriction point into its clinical use for all CML patients [80]. This restriction is partially attributed to the occurrence of specific mutations in the gene encoding the Bcr-Abl fusion tyrosine kinase, a result that poses the drug resistance development of major concern in modern cancer pharmacotherapy.

The difficulties in reaching improved pharmacotherapy outcomes have also been reported upon the clinical use of gefitinib (Iressa) developed for solid tumor therapeutic intervention, which has had limited, success. This agent, indeed, has been shown to be effective and beneficial only in non-small cell lung cancer (NSCLC) patients whose tumor cells exhibit mutations in the epidermal growth factor receptor (EGFR) gene leading to activated forms of the extracellular domain of EGFR kinase [81–84]. Furthermore, the failure of gefitinib to prolong patient's lives limited its use in the United States during 2005, whereas at the same time caused a withdrawal of its marketing application in Europe [85]. This effect, however, has not been seen in the case of erlotinib (Tarceva), which is another EGFR tyrosine kinase inhibitor in clinical use, where a survival benefit in lung cancer patients has been measured in prospective clinical studies. At the same time, discussions regarding the development of multitargeted tyrosine kinase inhibitors to overcome the limited efficiency of gefitinib, like zactima that inhibits both tyrosine kinases of EGFR and vascular endothelial growth factor (VEGF) receptor-2, have sparked

new research efforts in the field of developing tyrosine kinase inhibitors as new anticancer drugs. Overall, the experience already gained from the development and use in clinical practice of drugs of the new class of tyrosine kinase inhibitors (imatinib/Gleevec; gefitinib/Iressa; erlotinib/Tarceva) has clearly shown that the path from the initial enthusiasm of innovative drug discovery to improving pharmacotherapy outcomes *via* the identification of the molecular basis of responders and nonresponders has not been straightforward and easily attainable [77].

The problems confronted in clinical practice upon the delivery of tyrosine kinase inhibitors as drugs obviously support the notion for the exploitation of alternative and more sophisticated ways that must be undertaken in order to better cure cancer patients. Also, it stresses the need for thorough understanding of the molecular mechanisms that underline vital cellular processes before pharmacotyping can be generally applied in clinical practice. To this end, and to avoid such undesirable effects, the development and use of drugs that can induce simultaneously differentiation and apoptosis of leukemia cells have also been suggested [78]. Both processes result in restricted cell growth and thus represent a desirable target for the development of new anticancer therapeutics. To this end, the process of cell renewal in leukemia versus that of cell differentiation and/or apoptosis seems to be linked somehow to each other, thus giving hope that both differentiation and apoptosis can be simultaneously activated in order to serve as potential platforms for developing combined therapeutic approaches [78, 86, 87]. Toward this direction, the use of model cellular systems for more detailed analysis of the function of genes and the elucidation of their specific interactions in normal and disease states represents a useful alternative experimental design of clinical significance, as it has been recently shown [78, 87].

Interestingly, great progress has recently occurred in breast cancer biology and pharmacotherapy by revealing the structural interactions of the drug Herceptin and its receptor HER2. Herceptin is a humanized monoclonal antibody (known as trastuzumab) that was approved by the FDA in 1998 for breast cancer treatment and was the first genomic-research-based targeted anticancer therapeutic entered into clinical practice [77]. Its effects are achieved through the binding in the HER2 receptor (also known as Neu or ErbB2), a member of the epidermal growth factor receptor (EGFR; also known as ErbB) family of tyrosine kinases, and its prescription is based on the expression level of HER2 in the patient's tumor tissue (individualized therapy) [88]. The recent elucidation of the crystal structure of the entire extracellular regions of HER2 complexed with the Herceptin antigen-binding fragment (Fas) gives new information on the drug-receptor structural interactions, and this knowledge may lead to better drug design and development of new therapeutics of this type [89]. Furthermore, such studies envisage a better understanding of cancer biology at the molecular level and permit the design of more specific pharmacogenomics studies to help establish pharmacotyping in cancer chemotherapy. In particular, such an interesting step related to human breast cancer progression and metastasis is considered in the recently published work on the gene expression profiles of the premalignant, pre-invasive, and invasive stages of this type of tumor, as well as the identification of a "gene signature" for breast cancer metastasis to lungs [90, 91]. Through this approach of genomic analysis, the better understanding of disease pathogenesis and metastatic potential can be achieved and the identification of specific genomic targets for better drug development could

also be attained. Moreover, by analyzing gene expression profiles in cancer pharmacotherapy treatments, it would be possible to characterize molecular gene networks that in turn might have important clinical cosequences for improving patient care. This happened in the case of ALL where 124 genes were identified to discriminate between different treatments of chemotherapeutic agents (methotrexate and mercaptopurine), a result that can be reversely applied afterward in order for the individualized therapy to be based on the expression level of such genes. Such a direction implies the application of genomic analysis of the identified genes for each ALL patient before initiating therapy with the referred drugs [92]. This procedure that uses transcriptional profiling analysis has been recently proposed as one way for applying pharmacogenomics in cancer pharmacotherapy [93]. Overall, however, the accumulated knowledge at the molecular level is expected to help toward elucidating the underlined complexity of both cancer biology and anticancer drug response variability. In any case, however enriching our knowledge of the molecular biology of cancer will ensure that better pharmacological interventions will be achieved in the near future.

## 6 THE CHALLENGES IN PHARMACEUTICAL AND MEDICAL EDUCATION TO MEET THE CENTRAL DOGMA OF PHARMACOTYPING AS WELL AS OF PERSONALIZED AND INFORMATION-BASED MEDICINE

The achievements and challenges in pharmaceutical research that have already been discussed in previous sections have added new and innovative knowledge in basic pharmacology and therapeutics and now stress for the appropriate adjustments in education and curricula in both medicine and pharmacy faculties [94–96]. Also, new roles for healthcare providers have emerged in the changing environment of pharmacotherapy and clinical practice that clearly coincide with their need for better education in pharmacogenomics and personalized medicine. It is evident that better training will be achieved by the development of new curricula based on the integration of drug-related genomic knowledge and bioinformatics technologies into the teaching process. However, healthcare educators already face difficulties in teaching pharmacogenomics and personalized medicine concepts in order to give students the skills and the knowledge to keep them up to date with recent advances in relevant drug-related issues [94]. The classic background bridge of physiology with chemistry and pharmacology needed in order for students to understand drug-related actions and effects in the body has now been expanded to include the knowledge coming from biochemistry, molecular biology, genetics, and finally genomics (Figure 2). The latter is now considered crucial for more thorough understanding of several aspects of drug delivery in order for future health and pharmaceutical care practitioners to improve their skills for better drug selection, dosage regimens, and co-administration of drugs in individual patients. Undoubtedly, this need for better training in pharmacogenomics, for both educators and healthcare providers, will be in increasing demands in the years to come, because developments are achieved by genomic technologies in all drug-related areas and rapidly introduced into health care and clinical practice [5, 10, 25, 94]. However, the successful integration of such knowledge in the educational process is a difficult

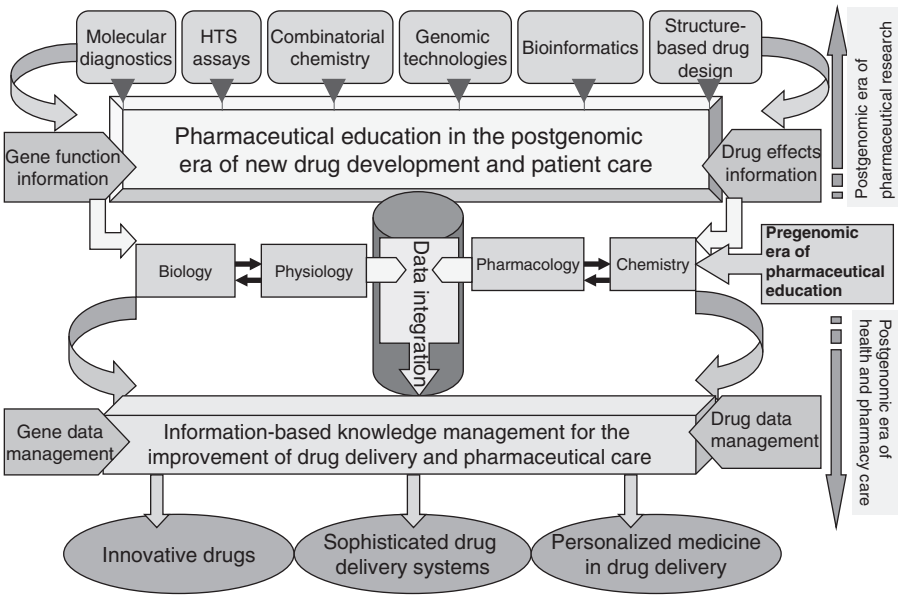


**Figure 2.** Convergence of genomic and clinical knowledge in postgenomic new drug development and health care (see the text for details).

task. At first, this means the development and use of new interactive computer-assisted learning methods capable of combining drug-related information with the function of genes and proteins being expanded through the genome and proteome. Second, it implies that the proper introduction of these programs into the teaching process organized in a tight timetable for completing both theory and clinical practice is needed [25, 97–100]. Another challenging issue in the design of new curricula has to deal with students' future demands to keep pace with the new trends in clinical and pharmacy practice rapidly entering and enriching the drug-related era, even after their graduation, because such profound changes in pharmacotherapy are continuously entering into the clinics, thus affecting the practice of modern medicine and pharmacy.

Pharmaceutical education also has to address some ethical aspects related to the application of genomic technologies in pharmacotherapy. In other words, by integrating genomics data into the clinical practice and the drug delivery process, several societal- and clinical-related issues are raised and must be taken into account during the educational process. It is evident that the application of pharmacotyping requires the health and pharmaceutical care practitioner to look into the patient's genetic makeup and to correlate the extracted information with clinical knowledge in standardized pharmacogenomics protocols in order to validate the compatibility of the drug dosage profile for the individual patient examined. In addition, pharmacogenomics data impinge on people's genetic variations and traits implicated in altered drug response that are usually different among populations from various parts of the world [15, 101, 102]. So, an obvious question is raised: How can pharmacogenomics-based molecular diagnostic tests be clinically tested and validated in order to be generally applied to all populations around the world,

by taking apart the patient's ethnic or racial origin? The same question seems to apply to the case of the drug developmental process and the criticism developed regarding the use of pharmacogenomics criteria for the stratification of patients recruited in clinical trials in order to gain better benefits for the efficacy and safety of the drug being tested [103–106]. As recently claimed, at least 29 drugs (or combination of drugs) have been shown to produce different effects in terms of efficacy and safety among ethnic or racial groups [102]. Although there is enough skepticism on these issues, one cannot easily preclude the discussion of such phenomena from the drug delivery process and must group patients according to the existed drug-related information that correlates with variability observed in pharmacological response. The latter is strengthened by the fact that several individuals show a genetic etiology implicated in their altered pharmacotherapy outcome. In any case, the categorization in pharmacotherapy profiling, i.e., pharmacotyping, must be focused and always be referred to specific drugs and not to specific ethnic populations [5, 25]. In addition, the development of DNA-based molecular diagnostics by no mean must dismiss the utilization of classic biochemical tests such as measurement of enzymatic activity, but they have to work in parallel. As molecular diagnosis is getting simpler, cheaper, and readily automated, then the application of pharmacogenomics will become more easily integrated into the clinical practice [107–110]. Undoubtedly, as the field of pharmacogenomics and personalized medicine continues to expand toward the development of relevant clinical research data, more issues will be raised for society that will need to be addressed. However, until pharmacogenomics research is generally accepted in patient care, its beneficial clinical impact together with the ethical and societal aspects raised must be taken into consideration by all scientists involved in its application to clinical practice. In particular, clinicians along with pharmacists and bioinformaticians must work together in order for pharmacogenomics clinical use to be compatible with good pharmaceutical and clinical practice guidelines [5, 25]. Furthermore, issues related either to quality or validation of the genetic tests that could be applied in pharmacotherapy and therapeutics must be clearly addressed, before the usage of genomic information in routine patient care. In any case, however, the positive consequences in the clinical practice will be greatly enhanced by imperatively incorporating the fundamental principles of pharmacogenomics, personalized medicine, and pharmacotyping into the core curricula of pharmacy and medical schools. And for sure, this will be beneficial for both health and pharmaceutical care, as well as for the society and the public in general.

## 7 THE DEVELOPMENT OF INFORMATION-BASED PLATFORMS TO ENSURE MAJOR BENEFITS FOR PHARMACOTYPING IN CLINICAL PRACTICE

As mentioned, genetic variations detected in genes encoding proteins implicated in drug action could affect their function and thus might result in an altered drug response (e.g., decreased metabolism, inadequate intracellular transport, impaired function on target site, etc.). Also, genomic technologies provide valuable techniques for uncovering the molecular mechanisms leading to pathogenesis and phenotype of specific human diseases. The accumulated data are organized into specific

databases; most of them can be accessed through the Web. Some of these functional genomics databases related to genetic disorders, gene polymorphisms, ADRs, and drug response variability are shown in Box 3. In this frame of genetic research, new genes, proteins, and genetic variations of clinical importance can be detected in individuals, because the new technology allows the screening of thousands of

**BOX 3 FUNCTIONAL GENOMICS DATABASES RELATED TO GENETIC DISORDERS, GENE POLYMORPHISMS, ADVERSE DRUG REACTION, AND DRUG RESPONSE VARIABILITY (MODIFIED FROM REF. 25)**

1. David Nelson's Cytochrome P450 Homepage
   http://drnelson.utmem.edu/CytochromeP450.html
2. Human Cytochrome P450 Nomenclature Website
   http://www.imm.ki.se/CYPalleles
3. The SNP Consortium-High-density maps of SNPs in the human genome
   http://snp.cshl.org
4. Adverse Drug Reactions Unit (Australia)
   http://www.health.gov.au/tga/adr/index.htm
5. Australian Adverse Drug Reactions Bulletin
   http://www.health.gov.au/tga/adr/aadrb.htm
6. Database of cytochrome P450-mediated drug interactions
   http://www.drug-interactions.com
7. Database of drug-induced arrhythmia
   http://www.torsades.org
8. Entrez SNP
   http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=snp
9. Pharmacological Targets Database (PTbase)
   http://research.bmn.com/ptbase
10. The "Online Mendelian Inheritance in Man" (OMIM) database (catalog of human genes and genetic disorders)
    http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=OMIM
11. The International HapMap Research Project with haplotyping data of the human genome
    http://www.hapmap.org/
12. The official database of the International Union of Pharmacology (IUPHAR) on Receptor Nomenclature and Drug Classification
    http://www.iuphar-db.org/iuphar-rd/index.html
13. The Pharmacogenetics and Pharmacogenomics Knowledge Base (PharmGKB) (It is an integrated resource database that contains genomic data and molecular, cellular, as well as clinical phenotype data related to drug response.)
    http://www.pharmgkb.org/

14. The Endogenous GPCR List
    http://www.tumor-gene.org/GPCR/gpcr.html
15. GPCRDB: An Information System for G Protein-Coupled Receptors (GPCRs) (Version 8)
    http://www.gpcr.org/
16. NucleaRDB: An Information System for Nuclear Receptors (Version 4)
    http://receptors.ucsf.edu/NR/
17. The Wnt Gene Homepage (Wnt proteins form a family of highly conserved secreted signaling molecules that regulate cell-to-cell interactions during embryogenesis. Wnt genes and Wnt signaling are also implicated in cancer.)
    http://www.stanford.edu/~rnusse/wntwindow.html
18. Genomic glossaries and Taxonomies—Evolving Terminology for Emerging Technologies (Cambridge Healthtech Institute)
    http://www.genomicglossaries.com/
19. Genome Programs of the U.S. Department of Energy Office of Research
    http://www.doegenomes.org/
20. Human Genome Project Information
    http://www.ornl.gov/sci/techresources/Human_Genome/home.shtml
21. The "GeneTests" website is a publicly funded medical genetics information resource developed for physicians, other healthcare providers, and researchers in order to provide authoritative information on genetic testing and its use in diagnosis, management, and genetic counseling.
    http://www.geneclinics.org

genes simultaneously. Genotyping and haplotyping analysis of individuals can identify specific gene variations and correlate them with drug effects, an approach that can obviously lead to individualized drug therapy. Furthermore, the conventional concept applied in drug therapy of considering patients suffering from the same disease as the homogenous population and using a similar drug therapy of all patients is changing in the genomic era. Under the new bioinformatic and genomic technologies used in personalized medicine, each patient is being handled as an individual, a practice that applies differentiation in pharmacotherapy [5, 25]. As a matter of fact, the detection of SNPs or other variations and the establishment of specific genotypes and haplotypes for several genes could explain why one patient responds well in drug therapy, another does not, or even why some people experience serious ADRs. Overall, if our understanding of these idiosyncratic pharmacological effects is finally achieved, then drug prescription would be greatly influenced. In fact, this will be a major transition in pharmacotherapy because pharmacotyping will be finally achieved, meaning that the physician could be based on the patient's genotyping–haplotyping analysis data for initiating and maintaining drug dosage therapeutic regimens for individual patients [5, 10, 25]. As the future prescription process is expected to be done by the physician in a fully

computerized environment with the aid of genomic and bioinformatic information concerning patient status, pharmacotyping could better ensure drug efficacy and safety. Such a case also means a transition from a drug-selection process mainly based on the physician's own experience into a more, highly integrated, information and computer-aided pharmacotherapy, thus making drug delivery digitized, more efficient, and safer. It is evident, however, that the genetic information derived from pharmacogenomics studies must be incorporated into clinical practice in such a way to assess simultaneously, besides drug efficacy, the risk of ADRs attributed to specific drug specimens. This fact can be further exploited through the recent scientific achievements of nanotechology and nanomedicine that obviously facilitate the development, validation, and establishment of information-based infrastructure in health care. As the advances in health sciences, either at a basic or a clinical level, are incorporated into clinical practice, nanotechnology can create the suitable diagnostic and drug delivery platforms in the healthcare system for better exploitation of such information at the molecular level. However, the path toward this goal is difficult enough, especially considering the existing complexity in physiology of individual cells and even more in the whole organism. The latter coincides with the high-level degree of cellular dynamics and homeostatic capability to respond to intrinsic or environmental factors by extremely complex structural and functional diversity. Toward this direction, the successful development of suitable nanodevices and the construction of biocompatible nanomaterials in nanomedicine implies that several factors ranging from the level of physiology and molecular biology to that of physics and chemistry must be taken into consideration. Furthermore, the identification of functional SNPs and the establishment of specific haplotypes for genes implicated in drug response, or ADRs, could also be used as a marker for drug prediction effects in a given individual, or even for a specific group of patients [25]. In this way, a standard framework for information-based medicine must first be developed, in order for this later approach to support an infrastructure for personalized medicine to achieve pharmacotyping in routine pharmaceutical care.

Another crucial aspect toward the successful application of pharmacogenomics and personalized medicine is related to the way in which the unprecedented load of the genomic information available at the molecular level has to be easily and rapidly transformed into a type-form applicable for everyday use by the medical practitioner or other specialties in health care. Indeed, the fast growing accumulation of genomic data concerning drug action is a real challenge that needs to be clarified in terms of the development of suitable tools capable of analyzing such information in a manner to be readily compatible for clinical use. This is of great importance in current drug delivery and biomedicine. In such a case, the validation of the genomic data mining processes will totally be based on the development of suitable computerized and data integration systems in order for these then to support the clinical transformation of the gathered genotyping and haplotyping profiling data from individual patients, before making such information applicable for routine patient care. Moving toward personalized medicine also means the existence of tools and molecular diagnostics capable of assessing genome-related clinical information in laboratory medicine to make the extracted information easily applicable by the physician. This is task very difficult, because before the transfer of techniques from genomic-related research laboratories to those used for the routine analysis of

clinical samples in diagnostic laboratories, these techniques must be extensively assessed for their ethical, social, and cost–benefit consequences [104]. As a consequence, the gradual integration of technologies transferring genomic information into the clinic needs the development of carefully selected and evaluated specific genetic biomarkers for the diagnosis of disease and prediction of drug response, as well as the use and application of genome-wide linkage analysis capable of genotyping, gene array, proteomics, transcriptomics, and metabolomics profiling [5, 105–107]. These developments of laboratory medicine should also highlight the way that must be undertaken in order to assure the level of quality and validity needed in genetic and molecular diagnostics tests. If this is confirmed in practice, then the individualized drug delivery profiling based on a patient's clinical, genomic, and bioinformatics data will be achieved in pharmacotherapy, e.g., pharmacotyping [5, 10]. Furthermore, the advent of specialized techniques coming from the advances of functional genomics will greatly influence the application of genomic information in laboratory and clinical practice. This also means that unified platforms must be developed to permit compatibility in handling different data gathered from unrelated sources like those of drug databases, clinical trials, DNA sequencing, and functional genomic analysis, in a way to ultimately support the clinical application of pharmacogenomics, personalized medicine, and pharmacotyping in health care.

The use of computational and bioinformatics approaches to predict the pharmacokinetics (absorption, distribution, metabolism, excretion; ADME) and pharmacodynamics properties of a drug are well appreciated throughout the process of drug development and delivery. Moving forward, the application of *in silico* methods and technology to evaluate safety and efficacy issues in pharmacotherapy by predicting mainly the emergence of drug interactions and ADRs in clinical practice are now considered major advancements to ultimately improve the success of new drug discovery and delivery outcomes [111]. Unfortunately, the data and the information generated up to now through the application of genomic and high-throughput technologies are impressive in scale but limited in clinical usefulness due to different database system formats and organization. For such an effort to finally succeed, information-based platforms in the drug development era and health care have to be developed to evaluate and integrate knowledge from different genomic and clinical sources in a manner that is simple for the end user. To this end, the application of semantic technologies with ontologies able to integrate the proper knowledge in a way that will be reusable by several applications and in different scientific areas can be a more beneficial approach toward better and quicker exploitation of genomic information in health care and therapeutics [112]. In any case, however, a crucial aspect in health care and pharmacotherapy to achieve major benefits from the technological and scientific advances in genomics and bioinformatics has been to create infrastructure and utilities that easily integrate each time the knowledge generated from different disciplines and facilities.

The clinical integration of genomic data in information-based platforms will also be greatly advanced by careful design and experimentation of pharmacogenomic studies, by ensuring genetic test quality, and evaluating and validating the data in routine patient care. This direction in medical and pharmacy practice will positively affect the rate by which genotyping data are transformed into specialized types applicable in personalized medicine, thus allowing pharmacotyping concepts to be

generally accepted as the new dimension in drug delivery to ensure drug efficacy and safety. The future advancement of patient care is directly related to the ability of medical practitioners to collaborate with other specialties to apply genomic technologies in clinical practice through the development of suitable tools and information-based platforms compliant with the good pharmaceutical and clinical practice guidelines [5]. Although this direction needs a multidisciplinary approach in order to finally be achieved, a recently published study has clearly shown that it is an affordable task. In particular, the maintenance warfarin dose was estimated through an algorithm generated by evaluating at the same time several clinical, pharmacogenetic, and demographic factors that were obtained before the warfarin initiation therapy was applied to patients [113]. The development of such a pharmacogenomics-based algorithm to predict an individual's response to coumarin therapy and the patient's maintenance dose has been recently discussed, showing that the better elucidation of genetic variants that can affect the coumarin drug response is an efficient and beneficial way to improve pharmacotherapy outcomes [114]. However, before the generalized application of such pharmacotyping/pharmacogenomics concepts in drug delivery, the complexity of personalized prescribing in terms of raised ethnicity issues has clearly to be addressed and dismissed in clinical practice [115]. By means of suitable, cost-affordable, and precious pharmacogenomics methodology, the *in vivo* selection, structural genomics analysis, functional verification, and clinical validation of biological samples can be safely performed in laboratory medicine in a manner that is compatible with information-based platforms. Overall, pharmacogenomics- and informatics-based approaches are expected to be finally used, giving hope, in the years to come, that the association of specific genetic markers with drug delivery dosage regimens, even for the therapy of complex diseases, is a realistic approach and can be achieved in everyday health care.

## 8 CONCLUDING REMARKS

By knowing how to better use drugs in clinical practice with enhanced efficacy and safety through the exploitation of genomic knowledge, the drug delivery era will certainly be advanced and medical practitioners will gain further roles in this challenging health and pharmaceutical care environment. But in order for this attempt to finally succeed, the appropriate education must be given to healthcare professionals through the development of new curricula and educational approaches. The training has to be focused on pharmacogenomics, personalized medicine, and pharmacotyping concepts as well as on bioinformatics and information-based medical practice. Also, the differences and peculiarities in health care and education found among several countries all over the world must be seriously taken into account, when the profound changes in post-genomic drug delivery and clinical practice are organized through the development of information-based platforms to implement and improve patient care. The trends in patient care-related issues or even the changes happening in one part of the world must be carefully examined and then adjusted before their use in other regions and vice versa. And for sure, this will be beneficial for both health and pharmaceutical care, as well as for the society and the public in general.

## REFERENCES


1. Evans W E, McLeod H L (2003). Pharmacogenomics—drug disposition, drug targets, and side effects. *N. Engl. J. Med.* 348:538–549.
2. Pirmohamed M, Park B K (2001). Genetic susceptibility to adverse drug reactions. *Trends Pharmacol. Sci.* 22:298–305.
3. Müller M (2003). Pharmacogenomics and the complexity of drug response. *Int. J. Clin. Pharmacol. Ther.* 41:231–240.
4. Ginsburg G S, McCarthy J J (2001). Personalized medicine: Revolutionizing drug discovery and patient care. *Trends Biotechnol.* 19:491–496.
5. Vizirianakis I S (2005). Improving pharmacotherapy outcomes by pharmacogenomics: From expectation to reality? *Pharmacogenomics*, In Press.
6. Housman D, Ledley F D (1998). Why pharmacogenomics? Why now? *Nature Biotechnol.* 16:2–3.
7. Evans W E, Relling M V (1999). Pharmacogenomics: Translating functional genomics into rational therapeutics. *Science.* 286:487–491.
8. McLeod H L, Evans W E (2001). Pharmacogenomics: Unlocking the human genome for better drug therapy. *Annu. Rev. Pharmacol. Toxicol.* 41;101–121.
9. Roses A D (2000). Pharmacogenetics and the practice of medicine. *Nature.* 405: 857–865.
10. Vizirianakis I S (2004). Challenges in current drug delivery from the potential application of pharmacogenomics and personalized medicine in clinical practice. *Curr. Drug Deliv.* 1:73–80.
11. Nebert D W, Nebert L J, Vesell E S (2003). Pharmacogenomics and "individualized drug therapy": High expectations and disappointing achievements. *Am. J. Pharmacogenomics.* 3:361–370.
12. Lazarou J, Pomeranz B H, Corey P N (1998). Incidence of adverse drug reactions in hospitalized patients: A meta-analysis of prospective studies. *J. Am. Med. Assoc.* 279:1200–1205.
13. Kohn L T, Corrigan J M, Donaldson M S (Eds.) (2000). *To Err is human: Building a safer health system. Committee on Quality of Health Care in America.* Institute of Medicine, The National Academy Press, Washington, D.C.
14. Flockhart D A, Tanus-Santos J E (2002). Implications of cytochrome P450 interactions when prescribing medication for hypertension. *Arch. Intern. Med.* 162:405–412.
15. Ingelman-Sundberg M (2004). Pharmacogenetics of cytochrome P450 and its applications in drug therapy: The past, present and future. *Trends Pharmacol. Sci.* 25: 193–200.
16. Ebbesen J, Buajordet I, Erikssen J, et al. (2001). Drug-related deaths in a department of internal medicine. *Arch. Intern. Med.* 161:2317–2323.
17. Ingelman-Saundberg M, Oscarson M, McLellan R A (1999). Polymorphic human cytochrome P450 enzymes: An opportunity for individualized drug treatment. *Trends Pharmacol. Sci.* 20:342–349.
18. Meyer U A (2000). Pharmacogenetics and adverse drug reactions. *Lancet.* 356:1667–1671.
19. Furuta T, Shirai N, Sugimoto M, et al. (2004). Pharmacogenomics of proton pump inhibitors. *Pharmacogenomics.* 5:181–202.
20. Schwab M, Eichelbaum M, Fromm M F (2003). Genetic polymorphisms of the human MDR1 drug transporter. *Annu. Rev. Pharmacol. Toxicol.* 43:285–307.

21. Siddiqui A, Kerb R, Weale M E, et al. (2003). Association of multidrug resistance in epilepsy with a polymorphism in the drug-transporter gene ABCB1. *N. Engl. J. Med.* 348:1442–1448.
22. Guillemette C (2003) Pharmacogenomics of human UDP-glucuronosyltransferase enzymes. *Pharmacogenomics J.* 3:136–158.
23. Hetherington S, Hughes A R, Mosteller M, et al. (2002). Genetic variations in HLA-B region and hypersensitivity reactions to abacavir. *Lancet.* 359:1121–1122.
24. Mallal S, Nolan D, Witt C, et al. (2002). Association between presence of HLA-B*5701, HLA-DR7, and HLA-DQ3 and hypersensitivity to HIV-1 reverse-transcriptase inhibitor abacavir. *Lancet.* 359:727–732.
25. Vizirianakis I S (2006). The transformation of pharmacogenetics into pharmacogenomics reinforces personalized medicine towards pharmacotyping for improved patient care. In Barnes L P (ed.), *New Research on Pharmacogenetics*: Nova Science Publishers, Inc., New York., In Press.
26. Watters J W, McLeod H L (2003). Cancer pharmacogenomics: Current and future applications. *Biochim. Biophys. Acta.* 1603:99–111.
27. Meisel P (2002). Arylamine N-acetyltransferases and drug response. *Pharmacogenomics.* 3:349–366.
28. Anantharam A, Markowitz S M, Abbott G W (2003). Pharmacogenetic considerations in diseases of cardiac ion channels. *J. Pharmacol. Exp. Ther.* 307:831–838.
29. Tsai Y J, Hoyme H E (2002). Pharmacogenomics: The future of drug therapy. *Clin. Genet.* 62:257–264.
30. McLeod H L, Siva C (2002). The thiopurine S-methyltransferase gene locus—implications for clinical pharmacogenomics. *Pharmacogenomics.* 3:89–98.
31. Coulthard S A, Hall A G (2001). Recent advances in the pharmacogenomics of thiopurine methyltransferase. *Pharmacogenomics J.* 1:254–261.
32. Arranz M J, Munro J, Birkett J, et al. (2000). Pharmacogenetic prediction of clozapine response. *Lancet.* 355:1615–1616.
33. Cadman P E, O'Connor D T (2003). Pharmacogenomics of hypertension. *Curr. Opin. Nephrol. Hypertens.* 12:61–70.
34. Manasco P K, Rieser P, Renegar G (2002). Pharmacogenetics and the genetic basis of ADRs. In: Mann, R D, Andrews, E B (eds.), *Pharmacovigilance*: Wiley, West Sussex, pp. 516–553.
35. Small K M, Tanguay D A, Nandabalan K, et al. (2003). Gene and protein domain-specific patterns of genetic variability within the G-protein coupled receptor superfamily. *Am. J. Pharmacogenomics.* 3:65–71.
36. Drysdale C M, McGraw D W, Stack C B, et al. (2000). Complex promoter and coding region beta 2-adrenergic receptor haplotypes alter receptor expression and predict in vivo responsiveness. *Proc. Natl. Acad. Sci. U.S.A.* 97:10483–10488.
37. Shafran A, Ifergan I, Bram E, et al. (2005). ABCG2 harboring the Gly482 mutation confers high-level resistance to various hydrophilic antifolates. *Cancer Res.* 65:8414–8422.
38. Loktionov A (2004). Common gene polymorphisms, cancer progression and prognosis. *Cancer Let.* 208:1–33.
39. Anderson J L, Carlquist J F, Horne B D, et al. (2003). Cardiovascular pharmacogenomics: Current status, future prospects. *J. Cardiovasc. Pharmacol. Ther.* 8:71–83.
40. Cacabelos R (2004). The application of functional genomics to Alzheimer's disease. *Pharmacogenomics.* 4:597–621.

41. Ulrich C M, Robien K, Mcleod H L (2003). Cancer pharmacogenetics: Polymorphisms, pathways and beyond. *Nat. Rev. Cancer.* 3:912–920.
42. Sayers I, Hall I P (2005). Pharmacogenetic approaches in the treatment of asthma. *Curr. Allergy Asthma Rep.* 5:101–108.
43. Rusnak J M, Kisabeth R M, Herbert D P, et al. (2001). Pharmacogenomics: A clinician's primer on emerging technologies for improved patient care. *Mayo Clin. Proc.* 76:299–309.
44. McCarthy J J, Hilfiker R (2000). The use of single-nucleotide polymorphism in pharmacogenomics. *Nature Biotechnol.* 18:505–508.
45. Carpenter A E, Sabatini D M (2004). Systematic genome-wide screens of gene function. *Nature Rev. Genet.* 5:11–22.
46. Rudnicki M, Mayer G (2003). Pharmacogenomics of angiotensin converting enzyme inhibitors in renal disease—pathophysiological considerations. *Pharmacogenomics.* 4:153–162.
47. Humma L M, Terra S G (2002). Pharmacogenetics and cardiovascular disease: Impact on drug response and applications to disease management. *Am. J. Health-Syst. Pharm.* 59:1241–1252.
48. Ferrari P, Bianchi G (2000). The genomics of cardiovascular disorders. *Drugs.* 59:1025–1042.
49. Nakagawa K, Ishizaki T (2000). Therapeutic relevance of pharmacogenetic factors in cardiovascular medicine. *Pharmacol. Ther.* 86:1–28.
50. Ferrari P (1998). Pharmacogenomics: A new approach to individual therapy of hypertension? *Curr. Opin. Nephrol. Hypertens.* 7:217–222.
51. Liew C-C, Dzau V J (2004). Molecular genetics and genomics of heart failure. *Nat. Rev. Genet.* 5:811–825.
52. Roberts R, Brugada R (2003). Genetics and arrhythmias. *Annu. Rev. Med.* 54:257–267.
53. Saavedra J M (2005). Studies on genes and hypertension: A daunting task. *J. Hypertension.* 23:929–932.
54. Kurland L, Lind L, Melhus H (2005). Using genotyping to predict responses to antihypertensive treatment. *Trends Pharmacol. Sci.* 26:443–447.
55. Wrighton S A, Schuetz E G, Thummel K E, et al. (2000). The human CYP3A subfamily: Practical considerations. *Drug Metab. Rev.* 32:339–361.
56. Nelson D R, Zeldin D C, Hoffman S M, et al. (2004). Comparison of cytochrome P450 (CYP) genes from the mouse and human genomes, including nomenclature recommendations for genes, pseudogenes and alternative-splice variants. *Pharmacogenetics.* 14:1–18.
57. Phillips K A, Veenstra D L, Oren E, et al. (2001). Potential role of pharmacogenomics in reducing adverse drug reactions: A systematic review. *J. Am. Med. Assoc.* 286:2270–2279.
58. Kirchheiner J, Brosen K, Dahl M L, et al. (2001). CYP2D6 and CYP2C19 genotype-based dose recommendations for antidepressants: A first step towards subpopulation-specific dosages. *Acta Psychiatr. Scand.* 104:173–192.
59. Gabelli C (2003). Rivastigmine: An update on therapeutic efficacy in Alzheimer's disease and other conditions. Curr. Med. Res. Opin. 19:69–82.
60. Bentue-Ferrer D, Tribut O, Polard E, et al. (2003). Clinically significant drug interactions with cholinesterase inhibitors: A guide for neurologists. *CNS Drugs.* 17:947–963.

61. Sadee W, Hoeg E, Lukas J, et al. (2001). Genetic variations in human G protein-coupled receptors: Implications for drug therapy. *AAPS PharmSci.* 3(3):E22. Available: http://www.pharmsci.org.
62. Desta Z, Zhao X, Shin J-G, et al. (2002). Clinical significance of the cytochrome P450 2C19 genetic polymorphism. *Clin. Pharmacokinet.* 41:913–958.
63. Sakaeda T, Nakamura T, Okumura K (2003). Pharmacogenetics of MDR1 and its impact on the pharmacokinetics and pharmacodynamics of drugs. *Pharmacogenomics.* 4:397–410.
64. Bredel M, Zentner J (2002). Brain-tumour drug resistance: The bare essentials. *Lancet Oncol.* 3:397–406.
65. Hall I P (2002). Pharmacogenetics, pharmacogenomics and airway disease. *Respir. Res.* 3:10.
66. Watters J W, McLeod H L (2003). Cancer pharmacogenomics: Current and future applications. *Biochim. Biophys. Acta.* 1603:99–111.
67. McLeod H L, Watters J W (2004). Irinotecan pharmacogenetics: Is it time to intervene? *J. Clin. Oncol.* 22:1356–1359.
68. Krynetski E Y, Evans W E (2000). Genetic polymorphism of thiopurine S-methyltransferase: Molecular mechanisms and clinical importance. *Pharmacology.* 61: 136–146.
69. O'Kane D J, Weinshilboum R M, Moyer T P (2003). Pharmacogenomics and reducing the frequency of adverse drug events. *Pharmacogenomics.* 4:1–4.
70. Marshall E (2003). Preventing toxicity with a gene test. *Science.* 302:588–590.
71. Coulthard S A, Hogarth L A, Little M, et al. (2002). The effect of thiopurine methyltransferase expression on sensitivity to thiopurine drugs. *Mol. Pharmacol.* 62:102–109.
72. Goldstein D B (2003). Pharmacogenetics in the laboratory and the clinic. *N. Engl. J. Med.* 348:553–556.
73. Traxler P (2003). Tyrosine kinases as targets in cancer therapy—successes and failures. *Expert Opin. Ther. Targets.* 7:215–234.
74. Tenen D G (2003). Disruption of differentiation in human cancer: AML shows the way. *Nature Rev. Cancer.* 3:89–101.
75. Graf T (2002). Differentiation plasticity of hematopoietic cells. *Blood.* 99:3089–3101.
76. Hockenbery D M, Giedt C D, O'Neill J W, et al. (2002). Mitochondria and apoptosis: New therapeutic targets. *Adv. Cancer Res.* 85:203–242.
77. Fischer O M, Streit S, Hart S, et al. (2003). Beyond Herceptin and Gleevec. *Curr. Opin. Chem. Biol.* 7:490–495.
78. Tsiftsoglou A S, Pappas I S, Vizirianakis I S (2003). Mechanisms involved in the induced differentiation of leukemia cells. *Pharmacol Ther.* 100:257–290.
79. Capdeville R, Buchdunger E, Zimmermann J, et al. (2002). Clivec (STI-571, imatinib), a rationally developed, targeted anticancer drug. *Nature Rev. Drug Discov.* 1:493–502.
80. Cowan-Jacob S W, Guez V, Fendrich G, et al. (2004). Imatinib (STI571) resistance in chronic myelogenous leukemia: molecular basis of the underlying mechanisms and potential strategies for treatment. *Mini Rev. Med. Chem.* 4:285–299.
81. Suzuki T, Mitsudomi T, Hida T (2004). The impact of *EGFR* mutations on gefitinib sensitivity in non-small cell lung cancer. *Personalized Med.* 1:27–34.
82. Golsteyn R M (2004). The story of gefitinib, an EGFR kinase that works in lung cancer. *Drug Discov. Today.* 9:587.

83. Paez J G, Janne P A, Lee J C, et al. (2004). EGFR mutations in lung cancer: Correlation with clinical response to gefitinib therapy. *Science.* 304:1497–1500.
84. Lynch T J, Bell D W, Sordella R, et al. (2004). Activating mutations in the epidermal growth factor receptor underlying responsiveness of non-small-cell lung cancer to gefitinib. *N. Engl. J. Med.* 350:2129–2139.
85. Golsteyn R M (2005). Gefitinib does not increase survival in lung cancer patients. *Drug Discov. Today.* 10:381
86. Domen J (2001). The role of apoptosis in regulating hematopoietic stem cell numbers. *Apoptosis.* 6:239–252.
87. Tsiftsoglou A S, Pappas I S, Vizirianakis I S (2003). The developmental program of murine erythroleukemia cells. *Oncol. Res.* 13:339–346.
88. Lindpaintner K (2002). The impact of pharmacogenetics and pharmacoge-nomics on drug discovery. *Nature Rev. Drug Discov.* 1:463–469.
89. Cho H-S, Mason K, Ramyar K X, et al. (2003). Structure of the extracellular region of HER2 alone and in complex with the Herceptin Fab. *Nature.* 421:756–760.
90. Ma X J, Salunga R, Tuggle J T, et al. (2003). Gene expression profiles of human breast cancer progression. *Proc. Natl. Acad. Sci. U.S.A.* 100:5974–5979.
91. Minn A J, Gupta G P, Siegel P M, et al. (2005). Genes that mediate breast cancer metastasis to lung. *Nature.* 436:518–524.
92. Cheok M H, Yang W, Pui C H, et al. (2003). Treatment-specific changes in gene expression discriminate in vivo drug response in human leukemia cells. *Nature Genet.* 34:85–90.
93. Slonim D K (2001). Transcriptional profiling in cancer: The path to clinical pharmacogenomics. *Pharmacogenomics.* 2:123–136.
94. Vizirianakis I S (2002). Pharmaceutical education in the wake of genomic technologies for drug development and personalized medicine. *Eur. J. Pharm. Sci.* 15:243–250.
95. Gurwitz D, Weizman A, Rehavi M (2003). Education: Teaching pharmacogenomics to prepare future physicians and researchers for personalized medicine. *Trends Pharmacol. Sci.* 24:122–125.
96. Brock T P, Valgus J M, Smith S R, et al. (2003). Pharmacogenomics: Implications and considerations for pharmacists. *Pharmacogenomics.* 4:321–330.
97. Wieczorek S J, Tsongalis G J (2001). Pharmacogenomics: Will it change the field of medicine? *Clin. Chim. Acta.* 308:1–8.
98. Hughes I, Hollingsworth M, Jones S J, et al. (1997). Knowledge and skills needs of pharmacology graduates in first employment: How do pharmacology courses measure up? *Trends Pharmacol. Sci.* 18:111–116.
99. Landro J A, Taylor I C, Stirtan W G, et al. (2000). HTS in the new millennium: The role of pharmacology and flexibility. *J. Pharmacol. Toxicol. Methods.* 44:273–289.
100. Rodriguez R, Vidrio H, Lopez-Martinez E, et al. (1997). Changing the countenance of pharmacology courses in medical schools. *Trends Pharmacol. Sci.* 18:314–318.
101. Varmus H (2002). Getting ready for gene-based medicine. *N. Engl. J. Med.* 347:1526–1527.
102. Tate S K, Goldstein D B (2004). Will tomorrow's medicines work for everyone? *Nat. Genet.* 36:S34-S42.
103. Lindpaintner K (2002). The impact of pharmacogenetics and pharmacogenomics on drug discovery. *Nature Rev. Drug Discov.* 1:463–469.
104. Issa A M (2000). Ethical considerations in clinical pharmacogenomics research. *Trends Pharmacol. Sci.* 21:247–250.

105. Issa A M (2002). Ethical perspectives on pharmacogenomic profiling in the drug development process. *Nat. Rev. Drug Discov.* 1:300–308.
106. Weber W W, Caldwell M D, Kurth J H (2003). Edging toward personalized medicine. *Curr. Pharmacogenomics.* 1:193–202.
107. Debouck C, Goodfellow P N (1999). DNA microarrays in drug discovery and development. *Nature Genet.* 21(Suppl):48–50.
108. Sander C (2000). Genomic medicine and the future of health care. *Science.* 287: 1977–1978.
109. Schmitz G, Aslanidis C, Lackner K J (2001). Pharmacogenomics: Implications for laboratory medicine. *Clin. Chim. Acta.* 308:43–53.
110. Hardiman G (2004). Microarray platforms-comparisons and contrasts. *Pharmacogenomics.* 5:487–502.
111. Ekins S (2004). Predicting undesirable drug interactions with promiscuous proteins in silico. *Drug Discov. Today.* 9:276–285.
112. Gardner S P (2005). Ontologies and semantic data integration. *Drug Discov. Today.* 10:1001–1007.
113. Gage B F, Eby C, Milligan P E, et al. (2004). Use of pharmacogenetics and clinical factors to predict the maintenance dose of warfarin. *Thromb. Haemost.* 91:87–94.
114. Voora D, McLeod H L, Eby C, et al. (2005). The pharmacogenetics of coumarin therapy. *Pharmacogenomics.* 6:503–513.
115. Hall I P (2005). Pharmacogenetics and ethnicity: More complexities of personalized prescribing. *Am. J. Respir. Crit. Care Med.* 171:535–536.

# TOXICOGENOMICS

Mary Jane Cunningham and Mrinal Shah
*Houston Advanced Research Center (HARC), The Woodlands, Texas*

**Chapter Contents**

1 Introduction
2 Genomics in Toxicology
3 Proteomics in Toxicology
4 Metabonomics in Toxicology
5 Pharmacogenomics in Toxicology
6 Systems Biology
7 Future Technologies in Toxicogenomics
References

## 1 INTRODUCTION

Toxicogenomics is a relatively new discipline within the field of toxicology. The phrase was first coined in 1998 at the first Toxicogenomics Workshop held as part of the U.S. and European Community Consortium on Molecular Toxicology in Palo Alto, California. In its broadest sense, it is defined as the use of OMICS technologies to investigate issues of toxicity [1]. In its narrowest sense, it is defined as investigating the safety of compounds by using only cutting-edge gene expression technologies [2, 3]. In this chapter, the broader definition will be applied. Toxicogenomics is the use of OMICS technologies to assess the safety of new chemical entities or other compounds used in diagnostics or therapeutics [4–6].

OMICS technologies encompass genomics, proteomics, metabonomics, and pharmacogenomics [5]. Other descriptive terms (for example, transcriptomics, toxicoproteomics, and toxicogenetics) have been used intermittently, but most

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

references have centered on those listed above. Genomics is the study of gene expression through the use of high-throughput screening techniques, such as gene expression microarrays. Other gene expression techniques have been added to this category, and these methods include gene reporter assays, branched DNA amplification assay, scintillation proximity assay (SPA), rapid analysis of gene expression (RAGE), serial analysis of gene expression (SAGE), and various polymerase chain reaction PCR-based assays (e.g., real-time, quantitative, representational difference analysis, differential display) as has already been reviewed [7]. However, in some cases, it may be a stretch to qualify the method as "high-throughput" because several of these technologies either address gene expression more indirectly, such as SAGE, or can only address a few genes at a time, such as PCR. Proteomics, on the other hand, is the study of protein expression using either two-dimensional polyacrylamide gel electrophoresis (2DE) annotated by mass spectrometry (MS) or protein expression microarrays [7, 8]. Metabonomics is the use of high-resolution combinations of nuclear magnetic resonance (NMR), chromatography, or mass spectroscopy to evaluate metabolite profiles of body fluids or cells [9]. Finally, pharmacogenomics is the study of genetic variability to explain the adverse effects caused by compound–cell interactions [10–12]. Most pharmacogenomics studies have focused on single nucleotide polymorphisms (SNPs). This chapter will describe all of these OMICS methods and show how they can be applied to studying toxicity in pharmaceutical development. Case studies will be presented from very early citations to examples of current applications. Data analysis methods will not be discussed but are discussed in Chapter "From defining bioinformatics and pharmacogenomics to developing information-based medicine and pharmacotyping in health care". This chapter will conclude with the authors' views on what the future may hold.

## 2 GENOMICS IN TOXICOLOGY

The applications of genomics to toxicology will be discussed with the focus on gene expression microarrays and how they have been used to investigate toxicity. Microarrays are covered in detail in Chapter "Microarrays in drug discovery and development". Briefly, they are manufactured by attaching pieces of DNA or RNA molecules to a nitrocellulose filter, a glass slide, or a silicon wafer [13]. Although early forms of arrays used nitrocellulose filters (now referred to as "macroarrays"), the most common substrate used today is a glass slide. Once the arrays have been printed with portion(s) of DNA or RNA (referred to as "probes"), they can be stored for further processing. The next step is to make the "targets" or isolated RNAs from both control cells and treated cells. These RNAs are labeled with a tag (most commonly fluorescent) and hybridized to complementary probes on the printed arrays. The microarrays are washed, scanned, and image analyzed to derive quantitative values for each probe signal. The remaining steps include data analysis of this raw information yielding (1) statistical analysis of replicate arrays and a measure of the variability in the processing steps, (2) similarities between gene expression patterns of control and treated samples, (3) identification of significantly expressed genes that may be biomarkers, and (4) patterns of gene–gene and gene–cell interactions.

Microarrays have been used as a screening tool in drug discovery and development. However, the focus and application of arrays in toxicology differs from

discovery in several aspects. The emphasis in discovery is feasibility: to ascertain the interaction of the compound and drug target rapidly. The emphasis in toxicology is development and validation: to understand this interaction more fully and, more importantly, to assess if any adverse effects are occurring. The focus here is on (1) validation of the compound–target interaction, (2) the prediction of adverse effects, and (3) the discovery of any alternative mechanisms of action resulting in "off-target" effects and aiding in the redesign of the candidate to a less toxic substance.

One of the most significant factors in microarray experiments is the experimental design. The overall design is dictated by whether a one-color (one label) or two-color (two labels) array system is being used [14, 15]. A one-color system requires that the control and treated samples are run on separate arrays, and therefore, intra-array as well as interarray variability needs to be considered. A two-color system requires that both the control target and the treated target compete for a complementary probe molecule on the same array. In this system, the importance of the intra-array variability is lessened. This aspect is the same whether a discovery approach or a toxicology approach is being used.

However, several aspects of the experimental design do differ between microarray screens used in discovery or toxicology. In toxicity assessments using microarrays, the emphasis on validation is an essential requirement of the experimental design. Particular attention needs to be paid to (1) types of species; (2) gender; (3) cell or tissue system; (4) treatment scheduling, dose, and route; (5) appropriate controls; (6) numbers of replicates, and (7) correlation with independent morphological and pathological toxicity assays [4, 5, 14]. Making these choices is critical for minimizing the process variability in order to maximize and visualize the biological variability. In 2001, The Microarray Gene Expression Data (MGED) Society proposed a list of guidelines for microarray work, known as the Minimum Information About a Microarray Experiment (MIAME) guidelines [16]. These guidelines are recommended requirements for how experiments are designed and reported. A set of guidelines specifically for toxicogenomics applications was drafted, but this draft has not yet been agreed to [17, 18].

Another aspect unique to toxicology research is the emphasis on analyzing the activity of annotated genes as opposed to unannotated genes. Annotated genes are genes whose sequence and function is already known. Unannotated genes are genes whose sequence is known but have not been assigned a function. As the focus for toxicology is validation, one priority is correlating significant expression of annotated genes observed in the data with their previously published values. A secondary priority is discovering which annotated genes or proteins are involved in any adverse or "off-target" effects. The final priority is discovering whether any unannotated genes are involved in these adverse effects and what their function or role may be. Therefore, when selecting microarrays to use, arrays that have the most annotated genes are preferred. Two exceptions to this recommendation are (1) whole genome arrays and (2) focus arrays. The availability of whole genome arrays, or arrays containing the entire complement of genes for that organism, make the selection of arrays biased toward annnotated genes a mute point. Focus arrays are sets of a smaller number of genes printed several times on one substrate. These arrays allow for profiling different samples using one array and, conceivably, hundreds of compounds could be screened in a high-throughput manner.

A third aspect of the experimental design that differs is the use of control or untreated samples. Discovery-focused experiments have emphasized the use of pooled samples to maximize consistency and minimize cost [19–25]. Pooled samples are individual samples from a cell or tissue system within a species and from the same treatment group that are pooled together to make one sample. In some cases, the control samples (from normal, non-diseased individuals) are pooled to make a "universal" reference [19]. However, the concerns are that (1) the resulting gene expression is an average of the group of samples, (2) the whole range of expression for the individuals within the group is not observed, and (3) pooling can only take place if the biological variation is greater than the process variation. [21–24]. One proposed method to address some of these concerns is "sub-pooling" where only subsets of samples are pooled and there are still replicates run for each group [25].

Toxicology-focused experiments may also include different sets of control samples. One set of control samples includes untreated samples both at the initial time point as well as at subsequent time points. If sufficient activity is found in the untreated control curve, this activity may need to be subtracted from the activity of the test curve. Another set of control samples includes samples from treatments using both positive and negative control compounds. Investigating the gene expression of the test compound is done by comparing it with the gene expression of the known control compounds. All of these controls need to be woven into the data analysis scheme.

As the focus is on validation, part of the evaluation is to determine how much the individual samples from a group vary in their gene expression. If the samples are pooled, the differences between responders and nonresponders will not be seen [24].

To understand toxicogenomics, it may be important to understand how microarrays evolved. An ongoing effort for developing more efficient sequencing by hybridization methods led to various initial microarray efforts where large numbers of genes could be screened at a time [26–30]. The first use of a microarray was in 1987 by Augenlicht et al., where macroarrays were used to differentiate a disease state [31]. In this case, cDNAs from a reference polyA mRNA library of the human colon carcinoma cell line, HT-29, were inserted into bacterial plasmids. Over 4000 clones were isolated and gridded onto several nitrocellulose filters. Radiolabeled cDNA probes from several biopsy samples (ranging from patients at low risk for colon cancer to familial adenomatous polyposis [FAP] patients to colon cancer patients) were then hybridized to the filters. The amount of radiolabel was scanned and analyzed. The amazing fact from this experiment was that a high percentage (20%) of FAP biopsies (in which the cells had not yet accumulated into adenomas) was upregulated as compared with the low-risk biopsies. These results suggested that increased gene expression seemed to correlate with early stages of the disease.

Initial toxicology studies using microarrays focused on screening and prioritizing lead compounds. Gray et al. designed a combinatorial library of protein kinase inhibitors used in cancer therapy and initially screened the compounds through an *in vitro* toxicity activity assay [32]. Then, three compounds were selected, and their interactions with yeast were investigated with oligonucleotide microarrays. The microarray results confirmed the diminished activity of one compound observed

in the *in vitro* assay. Another collaborative study between Incyte and Tularik screened several lead compounds for efficacy and toxicity using two-color microarrays [33]. The result was that the optimized lead compound had a similar profile to a known toxin, which led it to be redesigned to a better nontoxic lead compound.

One early study focused on discovering alternative modes of action. Karpf et al. used microarrays to broadly screen for genes expressed when a colon adenocarcinoma cell line was exposed to 5-aza-2′-deoxycytidine [34]. Several genes were significantly expressed. When investigated further, an alternative pathway was discovered and linked to the signal transducer and activator of transcription (STAT) genes.

The first comprehensive gene expression profiling time-course study using microarrays was first cited in 1998 by Cunningham et al. [35–38]. Using an acute short-term exposure regimen, rats were treated with toxic doses of three known hepatotoxins: benzo(a)pyrene (BP), acetaminophen (APAP), and clofibrate (CLO). mRNA isolated from the livers were analyzed using a cDNA microarray containing 7400 rat genes. Significant gene activity was observed at the early time points (12 hours, 1 day, and 3 days), and less activity was observed at the later time points (7, 14, and 28 days). All three compounds resulted in different expression profiles. Several cytochrome P450 genes and genes involved in phase II reactions were expressed over all six time points; however, the genes were induced at different times depending on the compound used (Figure 1 of chapter "Toxicogenomics"). Three different data analysis methods were compared. It was hypothesized that both non-genotoxic (causing damage by a non-DNA mechanism) APAP and CLO would have the most significantly expressed genes in common compared with genotoxic (damage caused by a DNA mechanism) BP. However, with all three analysis methods, APAP and BP showed the most overlap of significantly expressed genes. Interestingly, both APAP and BP have their primary metabolic pathway involving cytochrome P450, whereas the metabolism of CLO involves both cytochrome P450 and β-oxidation.

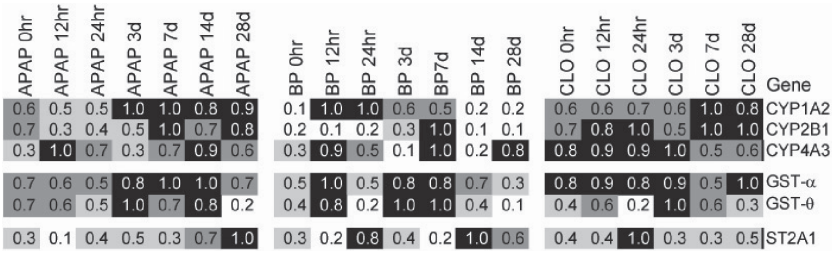


**Figure 1.** Normalized differential expression values for each gene across all time points (0 h, 12 h, 24 h, day 3, day 7, day 14, and day 28) for the following treatments: acetaminophen (APAP), benzo(a)pyrene (BP), and clofibrate (CLO). The highest relative expression is 1.0, and the lowest is 0.1. Three genes from the cytochrome P450 family (CYP1A2, CYP2B1, and CYP4A3) and two genes from the glutathione-S-transferase family (GST-α and GST-θ) give expression profiles that are similar to other members of the family but not identical. Gene expression profiles differ among all three gene families: CYP, GST, and sulfotransferase (ST).

Since these early citations, the use of gene expression microarrays as a tool for predictive toxicology has progressed. An early emphasis was to use microarrays to classify compounds based on their resulting gene expression profiles [39–46]. More recently, specific "sets of genes" giving rise to a gene expression profile have been used to distinguish classes of compounds [47, 48]. For example, Hu et al. showed through hierarchical clustering that distinct differences were observed between the gene expression profiles of direct- and indirect-acting genotoxins. When an array of over 9900 probes was used, a set of 58 genes could distinguish between these classes of compounds.

Another emphasis was to use gene expression microarrays to investigate mixtures. The profiles of the cellular interactions of mixtures were compared to the profiles of each component of the mixtures [49, 50]. Nadadur et al. compared the profile of residual oil fly ash (ROFA) in rat lung to the single profiles of vanadium sulfate and nickel sulfate, which are known components of ROFA. ROFA is derived from the exhaust plume of a power plant and is used as a model mixture to investigate the effects of air pollution. Using a focus macroarray, 12 genes were found to be significantly expressed in all three treatments. In a more recent study, arrays were shown to be able to distinguish toxic components of another type of mixture, wastewater effluent. From this study, Wang et al. found 11 genes that may ascertain the presence of endocrine-disrupting compounds in the environment [51].

A third emphasis was to match gene expression profiles to phenotypic changes. This correlation is the current goal for toxicogenomics. It will help toxicogenomics to be established as a tool in the regulatory arena. Early investigations focused on correlating gene expression from *in vivo* biological systems to gene expression from *in vitro* systems [52, 53]. Taking this concept further, attempts were made to correlate gene expression with specific organ regions exhibiting toxicity [54]. More recent studies have focused on anchoring specific cellular histopathological changes to gene expression [55–61]. To identify these changes, investigators have used laser capture microdissection (LCM) to extract specific cells exhibiting the phenotypic changes and then comparing the gene expression profiles of these cell populations with "normal" cell populations [62–66].

Finally, the most pressing issue in toxicogenomics may be the relevance of gene expression to toxicity. In other words, how does both the qualitative and the quantitative expression of a gene relate to toxicity? Several studies, including most of the papers already cited, have attempted to find this answer. Investigations by Heinloth et al. and Zhou et al. reported seeing gene expression changes before the observance of the overt toxicity [63, 67]. In each paper, *subchronic* doses of a toxicant were used and links of their expression with phenotypic changes were attempted. Today, research is ongoing, with the ultimate goal, to define a set of biomarkers that can predict toxicity earlier than the occurrence of phenotypic changes.

## 3 PROTEOMICS IN TOXICOLOGY

Two new words, *proteome* and *proteomics*, were coined a decade after classic terms such as *genome* and *genomics*. Here *proteome*, the complement of a *genome*, refers to the expression of all proteins in an organism, organ, or cell line at any given

time. The study of the proteome, its interactions, and its post-translational modifications (such as glycosylation, phosphorylation, etc.) that may be associated with a specific disease or toxicant is defined as *proteomics* [68, 69]. One advantage of proteomics is the ability to profile several thousands of proteins on a single platform and provide insights into post-translational modifications. Another advantage is that because proteins govern the normal physiology and disease processes, a more accessible endpoint of any changes in gene expression can be determined by direct analysis of proteins. Also, unlike genomics, proteomics offers the advantage of readily assaying both tissues and body fluids (e.g., plasma and serum) to investigate the molecules correlating with disease and drug action [70]. The proteomic analysis of body fluids is of particular interest as it can provide a noninvasive method for biomarker identification [71].

In the realm of toxicology, proteomics is one of many new technologies being used today to identify novel protein biomarkers and signature patterns in protein profiles that measure sensitive cellular changes in response to xenobiotic exposure [72–75]. The most widely used technique to date for proteomic science is high-resolution 2DE coupled with MS. First developed in 1975, it separates proteins in the first dimension based on their isoelectric points and in the second dimension by molecular weight [76]. Typically, 2DE can resolve hundreds to thousands of proteins from a single sample [77, 78]. To avoid overcrowding on a single gel, researchers now run multiple "zoom" gels that cover narrow pH ranges (e.g., covering 1 pH unit) [79]. In the mid-1990s, 2DE was coupled to MS allowing for annotation of the resolved spots. Proteins resolved using 2DE are analyzed by MS to give mass spectra. Subsequently, individual spots are excised from the gel and trypsin-digested to give shorter peptide fragments for sequencing. With the help of software programs and sequence databases, the peptide mass spectrum is compared with known partial gene and protein sequences and is then assigned an annotation. Recent developments in MS techniques, such as "tandem MS" (MS/MS), allow for better analysis of protein sequences, by avoiding ambiguity arising when MS data match more than one protein sequence.

The use of proteomics in toxicology can be divided into two classes: (1) investigative studies and (2) screening/predictive toxicology [71]. The mechanism of toxic damage in several model systems has been elucidated with proteomics [72, 79, 80]. But proteomics offers the prospect of identifying new toxic mechanisms along with traditional toxicology methods [74, 79, 81]. Such insights may provide valuable information about specific effects caused by a specific group or class of compounds. In addition to supporting investigative studies, proteomics promises to provide insight into screening and predictive toxicology. The screening of new compounds for toxicity can be possible by following specific signatures of protein expression, thus identifying them as predictive biomarkers of toxicity [82]. The sensitive nature of proteomics offers the detection of toxic effects at low dosage levels, which escape conventional screening methods, such as histopathology and clinical chemistry [83]. In addition, proteomics provides a method for the early detection of potential toxicity and the opportunity of ranking compounds during drug development. A recent review lists publications that use proteomics in toxicology studies, with the most popular application being biomarker identification [84]. The issues surrounding the experimental design and the emphasis on development and validation covered in the previous section for genomics applies to proteomics as well.

As an investigative tool, proteomics has been used to determine the failure of several drug candidates due to adverse drug reactions [72, 74]. In an early study, Anderson et al. developed a rodent liver proteomic toxicity database [85]. This database, containing profiles of effects of 43 compounds on liver, was among the first available for the detection, classification, and characterization of a wide range of hepatotoxins. The development of such a database has considerable implications for predictive toxicology. The potential hepatotoxicity of a new and unknown compound can be determined by comparing its effects on the proteome against known toxins from the database [74, 85].

New biomarker discovery from such studies can lead to early detection of deleterious effects. In a more recent study, researchers validated toxicological protein markers from an *in vivo* system (rat liver) as well as from an *in vitro* system (human HepG2 cell line) [86]. They reported a total of 11 protein markers with reactivity toward multiple toxic compounds and no reactivity toward nontoxic compounds. An important conclusion from this work is that cells in culture can be used as an *in vitro* toxicity testing system to assess hepatotoxicity. However, in the future, a much more extensive study may be required to identify a larger group of toxicology markers to detect more diverse types of toxic reactions.

Toward understanding the mechanistic action of toxicology using 2DE, Fountoulakis et al. showed the hepatotoxic effects of paracetamol (acetaminophen) overdose in human and rodent liver [74]. They found that the expression of 35 proteins was altered after treatment with paracetamol or its nontoxic regioisomer 3-acetamidophenol. Paracetamol selectively increases or decreases the phosphorylation state of proteins. This effect translates to a decrease in protein phosphatase activity [80]. The control of cellular functions of cells may be lost as a result of the dephosphorylation of certain regulatory proteins due to paracetamol overdose.

A dose- and time-dependent proteomics study with gentamicin confirmed the histopathological findings of renal toxicity at high doses and renal regeneration during the recovery phase [72]. Gentamicin, a known renal toxicant, represents a class of aminoglycosides. The effects of gentamicin treatment were observed after proteomic evaluation of rat kidney cortex samples. Examination of rat serum samples exposed to varying doses of gentamicin identified a protein being consistently overexpressed. Surprisingly, this particular protein marker was present both at the low treatment dose as well as at early time points before changes were observerd by routine clinical pathology. Intriguingly, during the recovery phase, the marker returned to control levels, thereby highlighting the sensitive nature of proteomics. Such protein markers are of great interest as they provide a noninvasive means of monitoring the onset of toxicity before any evident cellular damage.

Although 2DE offers a high-quality approach to proteomic research, it has certain limitations: Only major components of protein mixtures are visualized, the detection of low and high molecular mass of basic and hydrophobic proteins is inefficient, and it is a highly laborious technique. Alternative technologies, such as the protein microarrays, allows investigators to control conditions, such as pH, temperature, ionic strength, and different stages of protein modification, while monitoring the protein–protein interactions of thousands of proteins spotted on an array [87]. Protein microarrays not only provide a technique for high-throughput screening but also a method to study interactions of proteins with non-proteinaceous molecules. Zhu et al. have developed a protein microarray of the yeast

proteome from 5800 open reading frames [88]. The expressed proteins are purified and printed on glass slides with high spatial density to form a yeast proteome microarray. They were used to screen for interactions with proteins and phospholipids. The researchers identified many new calmodulin- and phospholipid-interacting proteins. Thus, protein microarrays can serve to screen for several hundreds to thousands of biochemical activities on a single chip. In toxicology, these arrays can be used to screen for adverse protein–drug interactions, to detect post-translational modifications leading to disease, and to identify biomarkers for drug targets and early disease detection [89].

A recent advancement is the coupling of MS with LCM for identification of tumor markers specific for cancer [89]. Specific cells of pathological interest are selected with LCM under direct microscopic visualization, laser captured, and removed from the tissues. The extracted lysate from the cells is applied directly to spots on the substrate of a surface-enhanced laser desorption/ionization (SELDI) MS, ionized, and desorbed from the surface, and a time-of-flight (ToF) proteomic profile of the entire cellular system is observed. This method has been recently applied to toxicology [90].

Several efforts are attempting to improve proteomic technologies to map and measure proteomes and subproteomes. However, no single proteomic platform seems to ideally suit and quantify the broad range of protein expression in a given cell/tissue system. More than one proteomic platform may likely be needed to distinguish the multiple forms and post-translational modifications of proteins, to address the inadequate annotation of proteomes, or to accomplish integration of proteomic data with genomic and metabonomic data. Additionally, advances in genomics data through investigations of pathways and subcellular structures induced by toxicity may guide studies in proteomics.

## 4 METABONOMICS IN TOXICOLOGY

The term metabonomics was derived from early work in Dr. Jeremy Nicholson's laboratory [9]. This term along with metabolomics has been used in several citations interchangeably [91–94]. Attempts have been made to clarify the definition of each term [95–99]. It seems that agreement has been reached where metabonomics is the broader encompassing term of metabolite profiling and metabolomics refers specifically to profiling in cells [9, 95, 98].

This technology uses combinations of NMR, MS, and chromatography to profile multiple components in biofluids, tissues, or cells. The main advantage of metabonomics is the serial sampling of a biofluid noninvasively. The earliest papers used high-resolution $^{1}$H-NMR alone for profiling [100, 101]. The technique became more sophisticated when bioinformatics tools, such as pattern recognition and expert systems, were applied to the data [102, 103]. Today, the following combinations of techniques are used: liquid chromatography–nuclear magnetic resonance (LC–NMR), magic angle spinning–nuclear magnetic resonance (MAS–NMR), gas chromatography–mass spectroscopy (GC–MS), LC–MS, capillary electrophoresis–mass spectroscopy (CE–MS), and LC–NMR–MS [97, 98]. The first technique listed in the combinations above refers to the method used to separate the multiple components of the sample, and these components are identified using the second technique [96].

The advantages and limitations for each combination of techniques have already been reviewed [94, 97].

Two of the earliest papers using metabonomics in toxicology were studies by Nicholson et al., which investigated mercury and cadmium toxicity in the rat [101, 104]. Each of these papers used an experimental design that reiteratively sampled a biological system over either dose or time using high-resolution $^{1}$H NMR. The changes in the urinary excretion patterns correlated with the expected histological changes. However, an interesting note was that in the case of cadmium, a urinary metabolite related to nephrotoxicity was found that was undetectable using traditional methods. A more extensive study that included a-naphthylisothiocyanate and 2-bromoethylamine showed changes in the metabolite profiles ahead of the time point where the clinical chemistry or microscopic changes would be detected [105].

As with the previously described OMICS technologies, one emphasis has been to profile groups of compounds using metabonomics [103, 105, 106]. Each compound gave a unique metabonomic profile and was successfully classified using pattern recognition techniques. A second emphasis was to be able to monitor not only biofluids, such as urine or plasma, but also tissues and cells. Moka et al. performed the first study to classify carcinoma biopsy samples by MAS–NMR [107]. Since then, two studies have used metabonomics to investigate toxicity using a wider approach where urine, plasma, and tissue samples are all analyzed [108, 109].

To collaborate and share information, a proposal for a users group in metabonomics was put forth [110]. Since then, the Consortium for Metabonomic Toxicology (COMET) was formed and the first report from this group was given in 2003 [111]. The final report was just published [112]. This report describes what methodologies have been assessed, the development of a curated database for metabonomic profiles, and computer-based experts systems for data analysis.

## 5 PHARMACOGENOMICS IN TOXICOLOGY

Pharmacogenomics and pharmacogenetics are two terms that have created confusion because they have been used interchangeably and defined differently by several authors [113]. The most widely accepted term, pharmacogenetics, is the study of an individual's response to a drug as determined by their genetic makeup [114]. Pharmacogenetics as a field of study has been around for several decades [113]. Early studies focused on observations of patient variability in metabolizing various drugs such as primaquine, succinylcholine, and dibucaine [115–118]. Comprehensive reviews have been written, including one by Werner Kalow, a pioneer in this field, and more details are covered in Chapter "Genetic markers and genotyping analyses for genetic disease studies" [11, 113, 114, 119, 120]. In this chapter, the term "pharmacogenomics" will be used to refer to the use of high-throughput screening techniques for the detection of the genetic variation of an individual and its role in toxicity, especially in causing adverse effects.

In toxicology, most pharmacogenomics studies have focused on variations within the drug metabolizing enzymes (DMEs) [11]. These DMEs catalyze the phase I and phase II metabolic reactions. Several examples have been cited for arylamine N-acetyltransferase (NAT1 and NAT2), thiopurine-S-methyltransferase (TPMT),

glutathione-S-transferase (GST), as well as the various cytochrome P450 enzymes (e.g., CYP2A6, CYP2C8, CYP2C9, CYP2C19, CYP2D6, and CYP3A4) [10, 11, 120–122]. Early studies showed that individuals with these genetic variations can be classified as "poor," "extensive," and "ultrarapid" metabolizers [114, 121, 123]. From a toxicity standpoint, this variability is extremely important in that poor metabolizers could accumulate the drug within their tissues causing a toxic response, whereas in extensive metabolizers, the drug has an extremely short half-life and may not stay around the tissues long enough to exert its desired effect. For example, an individual with the gene variant CYP2D6*2, which represents an "ultrarapid metabolizer," can possess a much higher activity for cytochrome P450 isozyme 2D6, whereas an individual with the gene variant CYP2D6*10 possesses a defective enzyme and therefore is a "poor metabolizer" [121, 123]. Individuals who have slow or fast metabolism rates have been observed in several mammalian species [124–128].

Early examples of polymorphisms being detected in a high-throughput manner are studies that used oligonucleotide microarrays containing perfect-match and mismatch sequences. These studies investigated polymorphic changes in breast cancer, HIV infection, and cystic fibrosis [129–131]. A more recent study used a chip containing over 11,000 SNPs to detect genetic variabilities in over 100 DNA samples—a high-throughput variation on association studies [132]. Another array format blended nanotechnology by using multiplexed nanoparticle probes with printed oligonucleotides to distinguish genetic polymorphic variants [133].

## 6 SYSTEMS BIOLOGY

Systems biology is a recent term that encompasses looking at the entire biological system from the molecular, cellular, tissue, organism, population, and finally, ecosystem levels. As Witkamp describes it, "systems biology is a realization that organisms do not consist of isolated subsets of genes, proteins and metabolites" [134]. It is also the application of network biology or comprehensive computational methods to analyze the data produced by the combination of OMICS technologies [134–137]. The ultimate goal of toxicogenomics is to predict compound–cell and cell–cell interactions *in silico*, and the application of systems biology brings this one step closer [5, 138, 139].

In attempts to obtain this goal, several studies have correlated datasets from combined genomics and proteomic studies and combined genomics and metabonomics studies. An early toxicogenomics study by Cunningham et al. combined the datasets from the rat toxicity study with three hepatotoxins evaluated by gene expression microarrays described earlier with a matched proteomics arm (Figure 2 of chapter "Toxicogenomics") [8, 35–38, 140, 141]. In this study, a portion of the gene and protein expression profiles matched but the remaining did not. Four other studies also tried correlating gene and protein expression in evaluating the effects of zinc, carbon tetrachloride, lithium, and bromobenzene [142–145]. The experimental design for these studies used only one or two time points with one dose, and in all three studies, only a portion of overlap between gene and protein expression was observed. Finally, Coen et al. attempted to correlate metabonomics with genomics [146]. They observed some overlap between the two technologies and,

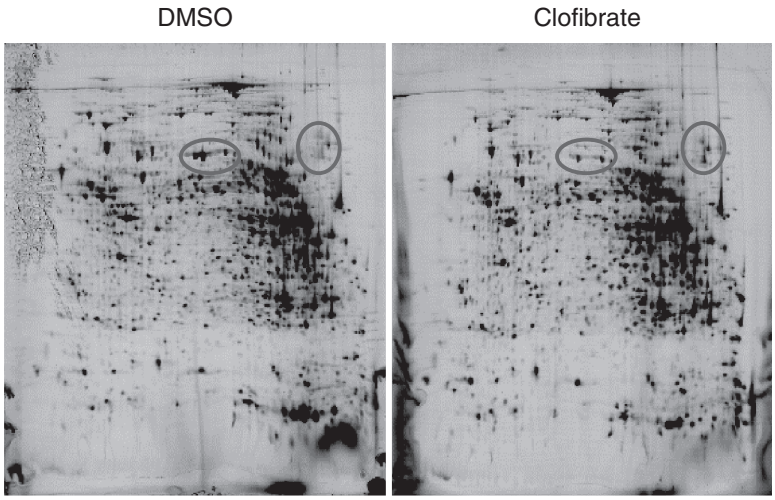


**Figure 2.** Proteome changes in livers at day 3 from rats treated with the vehicle DMSO (left) and clofibrate (right). The dose of clofibrate was a toxic dose at 250 mg/kg i.p. Two sets of specific changes are observed within the outlined areas.

hence, concluded that these technologies are complementary in providing a view into toxicity.

A fully integrated approach using genomics, proteomics, and metabonomics was attempted by two groups of investigators. Schnackenberg et al. studied the hepatotoxicity effects of valproic acid [147]. Correlative changes in glucose metabolism were shown by both metabonomics and proteomics methods. However, no significant changes were observed with the gene expression method. Kleno et al. used genomics, proteomics, and metabonomics to investigate the hepatotoxicity of hydrazine [148]. Disruption in both glucose and lipid metabolism were present in all three datasets, and therefore, once again, these three technologies were shown to be complementary.

## 7 FUTURE TECHNOLOGIES IN TOXICOGENOMICS

The field of toxicogenomics continues to evolve and develop. For all the technologies listed above, research is continuing to develop more efficient, more high-throughput versions. The field of nanotechnology may aid in this effort. The various engineered nanomaterials may enable the current microarray format to be redeveloped on a nanoscale level and provide more rapid and possibly instaneous results. Some of these engineered nanomaterials, such as single-walled carbon nanotubes, are being developed for applications in the communications and information technologies field and may lead to increased capacity and advances in bioinformatics. The entire field of biosensors is being redeveloped on the nanoscale to nanosensors—from microelectronic membranes (MEMs) to nanoelectronic membranes (NEMs). A preliminary study done by Cunningham et al. assesses the toxicity of these nanomaterials using genomics (Figure 3 of chapter "Toxicogenomics") [5, 149, 150]. In this study, the gene expression profiles of three known nanoscale

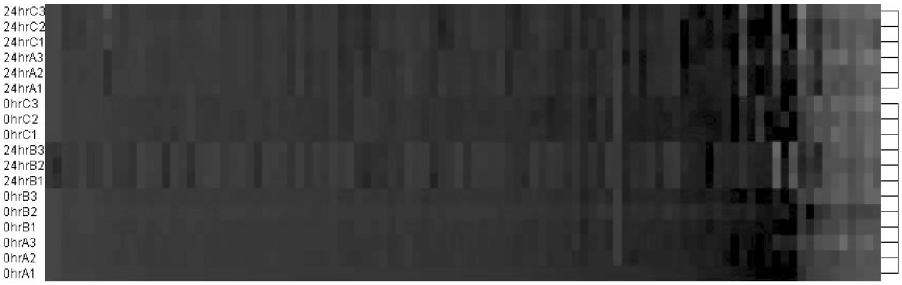


**Figure 3.** K-means clustering of gene expression profiles from human skin cells treated with $TiO_2$ (Compound A: A1, A2, A3), carbon black (Compound B: B1, B2, B3), and $SiO_2$ (Compound C: C1, C2, C3). Columns are the top 100 significantly expressed genes; rows are the samples representing triplicate arrays for each biological sample. The red color denotes upregulation of the gene, and the blue color denotes downregulation of the gene. Insignificantly-expressed genes are denoted by the black color. The expression profiles for $TiO_2$ and $SiO_2$ are more similar than the profile for carbon black. (This figure is available in full color at ftp://ftp.wiley.com/public/sci_tech_med/pharmaceutical_biotech/.)

materials were compared and two were found to be more similar with each other. This similarity correlated with the similarity in their chemical structures. As more and more engineered nanomaterials are being made, the applications for these materials in the life sciences will only grow. As a result, their safety will also need to be assessed.

In addition to the microarray formats discussed above, different substances are now being printed as reviewed by Cunningham [13]. Arrays are being manufactured with chemical ligands [151, 152], whole cells [153–155], tissue slices [156–158], peptide nucleic acids [159], and nanomaterials [160–166]. These new formats will increase the ability to investigate cell and tissue interactions more extensively than previously known.

Facing the future, other new technological fields are making their way to the forefront of research and could have pertinent applications for investigating toxicity. RNA interference (RNAi) is a burgeoning area (as covered in Chapter "Rna interference: the next gene-targeted medicine"). This field was first thought to include only investigations of naturally-occurring microRNA (miRNA) molecules and synthesized short-interfering RNA (siRNA) molecules. An early paper used siRNAs to the aryl hydrocarbon receptor (AhR) and the AhR nuclear translocator (ARNT) to look at gene silencing [167]. Since this 2003 paper, many references have now reported using siRNA molecules as knockout molecules to investigate mechanisms of toxicity. miRNA molecules, on the other hand, have been sequenced and printed on macroarrays and microarrays [168–169]. These array tools may allow the expression of these molecules, and consequently, their function to be investigated. Recently, other types of short non-coding RNA molecules have been reported as well as naturally-occurring siRNAs [170–172]. It will be interesting to see what roles, if any, each of these new classes of short RNA molecules play in toxicity.

Another technological field is the exciting area of stem cells. Although work has been ongoing with adult stem cells for many years, research involving embryonic stem cells continues to hold great promise. Cell lines for mouse and rat have

been used in several research areas to develop new animal models and devise new therapies [173, 174]. The next front is human embryonic stem (HES) cells. Currently, this tool is in the midst of an ethics and morality debate [175]. However, the same debate was waged when fetal tissue was first used in the laboratory setting over two decades ago. If an investigator could direct the differentiation pathway of HES to more mature tissue-specific cells, the need for primary human tissue, which is already difficult to get, would be lessened. This application would increase dramatically the capability of *in vitro* toxicity assays.

Toxicogenomics is a growing and ever-changing field within toxicology. All of these advances in research and development could completely change the safety assessment landscape within just a few years. As the regulatory assays become more predictable, the time from discovery of a pharmaceutical agent to its market release will be lessened.

## REFERENCES


1. Vojta P J, Kazmer V P, Kier L D (2001/2). Toxicogenomics: Insights into the present and future. *Drug Disc. World*. 3:18–22.
2. Farr S, Dunn R T (1999). Concise review: Gene expression applied to toxicology. *Toxicol. Sci.* 50:1–9.
3. Nuwaysir E F, Bittner M, Trent J, et al. (1999). Microarrays and toxicology: The advent of toxicogenomics. *Molecular Carcinogen*. 24:153–159.
4. Cunningham M J, Zacharewski T, Somogyi R, et al. (2002). Two-stepping through toxicogenomics: A basic primer. *Toxicol. Sci.* 66:1.
5. Nadadur S, Cunningham M J (2006). Predictive power of novel technologies (cells to "Omics"): Promises, pitfalls and potential applications. *Toxicol. Sci.* 90:1.
6. Mohrenweiser H W (2004). Genetic variation and exposure related risk estimation: will toxicology enter a new era? DNA repair and cancer as a paradigm. *Toxicol. Pathol.* 32*:*(1):136–145.
7. Cunningham M J (2000). Genomics and proteomics: the new millennium of drug development and discovery. *Journal of Pharmacol. Toxicolog. Methods.* 44:291–300.
8. Page M J, Moyses C, Cunningham M J, et al. (2003). Proteomics. *EXS*. 93:19–30.
9. Nicholson J K, Lindon J C, Holmes E (1999). "Metabonomics": Understanding the metabolic responses of living systems to pathophysiological stimuli via multivariate statistical analysis of biological NMR spectroscopic data. *Xenobiotica*. 29:1181–1189.
10. Sengupta L K, Sengupta S, Sarkar M (2002). Pharmacogenetic applications of the post genomic era. *Current Pharmac. Biotechnol.* 3:141–150.
11. Regnstrom K, Burgess D J (2005). Pharmacogenomics and its potential impact on drug and formulation development. *Critical Rev. Therap. Drug Carrier Sys.* 22:465–492.
12. Severino G, Del Zompo M (2004). Adverse drug reactions: Role of pharmacogenomics. *Pharmacological Res.* 49:363–373.
13. Cunningham M J, Dat D D (2005). Microarrays-Fabricating, Fuchs J and Podda M, (ed.), Marcel Dekker, *Encyclopedia of Diagnostic Genomics and Proteomics.* New York, pp. 819–823.
14. Lee K M, Kim J H, Kang D (2005). Design issues in toxicogenomics using DNA microarray experiment. *Toxicol. Appl. Pharm.* 207:S200–S208.

15. Yauk C L, Berndt M L, Williams A, et al. (2004). Comprehensive comparison of six microarray technologies. *Nucleic Acids Res.* 32:e124.
16. Brazma A, Hingamp P, Quackenbush J, et al. (2001). Minimum information about a microarray experiment (MIAME)-toward standards for microarray data. *Nature Genetics.* 29:365–371.
17. Microarray Gene Expression Data (MGED) Society. A MIAME for Toxicogenomics-MIAME/Tox, Based on Tox 1.1., Draft Publication, August, 2003. Available: http://www.mged.org.
18. Mattes W B, Pettit S D, Sansone S A, et al. (2004). Database development in toxicogenomics: Issues and efforts. *Environ. Health Persp.* 112:495–505.
19. Novoradovskaya N, Whitfield M L, Basehore L S, et al. (2004). Universal reference RNA as a standard for microarray experiments. *BMC Genomics.* 5:20.
20. Shih J H, Michalowska A M, Dobbin K, et al. (2004). Effects of pooling mRNA in microarray class comparisons. *Bioinformatics.* 20:3318–3325.
21. Zhang S D, Gant T W (2005). Effect of pooling samples on the efficiency of comparative studies using microarrays. *Bioinformatics.* 21:4378–4383.
22. Glass A, Henning J, Karopka T, et al. (2005). Representation of individual gene expression in completely pooled mRNA samples. *Biosci. Biotechnol. Biochem.* 69:1098–1103.
23. Kendziorski C M, Zhang Y, Lan H, et al. (2003). The efficiency of pooling mRNA in microarray experiments. *Biostatistics.* 4:465–477.
24. Jolly R A, Goldstein K M, Wei T, et al. (2005). Pooling samples within microarray studies: A comparative analysis of rat liver transcription response to prototypical toxicants. *Physiol. Genomics.* 22:346–355.
25. Peng X, Wood C L, Blalock E M, et al. (2003). Statisical implications of pooling RNA samples for microarray experiments. *BMC Bioinformatics.* 4:26.
26. Khrapko K R, Lysov Y P, Khorlyn A A, et al. (1989). An oligonucleotide hybridization approach to DNA sequencing. *FEBS Lett.* 256:118–122.
27. Southern E M, Maskos U, Elder J K (1992). Analyzing and comparing nucleic acid sequences by hybridization to arrays of oligonucleotides: Evaluation using experimental models. *Genomics.* 13:1008–1017.
28. Maskos U, Southern E M (1992). Oligonucleotide hybridisations on glass supports: A novel linker for oligonucleotide synthesis and hybridisation properties of oligonucleotides synthesised *in situ*. *Nucleic Acids Res.* 20:1679–1684.
29. Drmanac R, Drmanac S, Strezoska Z, et al. (1993). DNA sequence determination by hybridization: A strategy for efficient large-scale sequencing. *Science.* 260:1649–1652.
30. Beattie W G, Meng L, Turner S L, et al. (1995). Hybridization of DNA targets to glass-tethered oligonucleotide probes. *Molecular Biotechnol.* 4:213–225.
31. Augenlicht L H, Wahrman M Z, Halsey H, et al. (1987). Expression of cloned sequences in biopsies of human colonic tissue and in colonic carcinoma cells induced to differentiate *in vitro*. *Cancer Res.* 47:6017–6021.
32. Gray N S, Wodicka L, Thunnissen A M W H, et al. (1998). Exploiting chemical libraries, structure and genomics in the search for kinase inhibitors. *Science.* 281:533–538.
33. Braxton S, Bedilion T (1998). The integration of microarray information in the drug development process. *Curr. Opin. Biotech.* 9:643–649.
34. Karpf A R, Peterson P W, Rawlins J T, et al. (1999). Inhibition of DNA methyltransferase stimulates the expression of signal transducer and activator of transcription 1, 2, and 3 genes in colon tumor cells. *Proc. of the National Academy of Sciences USA.* 96:14007–14012.

35. Cunningham M J, Zweiger G, Bailey D, et al. (1998). Application of a genome-based screen. *Archives of Pharmacol.* 358:R564.
36. Cunningham M J, Liang S, Fuhrman S, et al. (2000). Gene expression microarray data analysis for toxicity profiling. *Annals of the New York Acad. Sci.* 919:52–67.
37. Fuhrman S, Cunningham M J, Wen X, et al. (2000). The application of Shannon entropy in the identification of putative drug targets. *BioSys.* 55:5–14.
38. Fuhrman S, Cunningham M J, Liang S, et al. (2000). Making sense of large-scale gene expression data with simple computational techniques. *Amer. Biotechnol. Lab.* 18:68–70.
39. Burczynski M E, McMillian M, Ciervo J, et al. (2000). Toxicogenomics-based discrimination of toxic mechanism in HepG2 human hepatoma cells. *Toxicol. Sci.* 58:399–415.
40. Waring J F, Jolly R A, Ciurlionis R, et al. (2001). Clustering of hepatotoxins based on mechanism of toxicity using gene expression profiles. *Toxicol. Appl. Pharm.* 175:28–42.
41. Thomas R S, Rank D R, Penn S G, et al. (2001). Identification of toxicologically predictive gene sets using cDNA microarrays. *Molec. Pharmacol.* 60:1189–1194.
42. Hamadeh H K, Bushel P R, Jayadev S, et al. (2002). Gene expression analysis reveals chemical-specific profiles. *Toxicol. Sci.* 67:219–231.
43. Hamadeh H K, Bushel P R, Jayadev S, et al. (2002). Prediction of compound signature using high density gene expression profiling. *Toxicol. Sci.* 67:232–240.
44. de Longueville F, Atienzar F A, Marcq L, et al. (2003). Use of a low-density microarray for studying gene expression patterns induced by hepatotoxicants on primary cultures of rat hepatocytes. *Toxicol. Sci.* 75:378–392.
45. Ishida S, Shigemoto-Mogami Y, Kagechika H, et al. (2003). Clinically potential subclasses of retinoid synergists revealed by gene expression profiling. *Molec. Cancer Therapeutics.* 2:49–58.
46. Kier L D, Neft R, Tang L, et al. (2004). Applications of microarrays with toxicologically relevant genes (tox genes) for the evaluation of chemical toxicants in Sprague Dawley rats in vivo and human hepatocytes in vitro. *Mutation Res.* 549:101–113.
47. Hu T, Gibson D P, Carr G J, et al. (2004). Identification of a gene expression profile that discriminates indirect-acting genotoxins from direct-acting genotoxins. *Mutation Res.* 549:5–27.
48. McMillian M, Nie A Y, Parker J B, et al. (2004). A gene expression signature for oxidant stress/reactive metabolites in rat liver. *Biochem. Pharmacol.* 68:2249–2261.
49. Nadadur S S, Schladweiler M C J, Kodavanti U P (2000). A pulmonary rat gene array for screening altered expression profiles in air pollutant-induced lung injury. *Inhalation Toxicol.* 12:1239–1254.
50. Bae D S, Hanneman W H, Yang R S H, et al. (2002). Characterization of gene expression changes associated with MNNG, arsenic, or metal mixture treatment in human keratinocytes: Application of cDNA microarray technology. *Environ. Health Persp.* 110(Suppl. 6):931–941.
51. Wang D Y, McKague B, Liss S N, et al. (2004). Gene expression profiles for detecting and distinguishing potential endocrine-disrupting compounds in environmental samples. *Environ. Sci. Technol.* 38:6396–6406.
52. Jessen B A, Mullins J S, de Peyster A, et al. (2003). Assessment of hepatocytes and liver slices as *in vitro* test systems to predict *in vivo* gene expression. *Toxicol. Sci.* 75:208–222.
53. Boess F, Kamber M, Romer S, et al. (2003). Gene expression in two hepatic cell lines, cultured primary hepatocytes, and liver slices compared to the *in vivo* liver gene

expression in rats: Possible implications for toxicogenomics use of *in vitro* systems. *Toxicol. Sci.* 73:386–402.

54. Konu O, Kane J K, Barrett T, et al. (2001). Region-specific transcriptional response to chronic nicotine in rat brain. *Brain Res.* 909:194–203.
55. Hamadeh H K, Knight B L, Haugen A C, et al. (2002). Methapyrilene toxicity: Anchorage of pathologic observations to gene expression alterations. *Toxicol. Pathol.* 30:470–482.
56. Kriete A, Anderson M K, Love B, et al. (2003). Combined histomorphometric and gene-expression profiling applied to toxicology. *Genome Biol.* 4:R32.1–R32.9.
57. Sadlier D M, Connolly S B, Kieran N E, et al. (2004). Sequential extracellular matrix-focused and baited-global cluster analysis of serial transcriptomic profiles identifies candidate modulators of renal tubulointerstitial fibrosis in murine adriamycin-induced nephropathy. *J. Biolog. Chem.* 279:29670–29680.
58. Kisby G E, Standley M, Lu X, et al. (2005). Molecular networks perturbed in a developmental animal model of brain injury. *Neurobiol. Dis.* 19:108–118.
59. Luo W, Fan W, Xie H, et al. (2005). Phenotypic anchoring of global gene expression profiles induced by N-hydroxy-4-acetylaminobiphenyl and benzo[*a*]pyrene diol epoxide reveals correlations between expression profiles and mechanism of toxicity. *Chem. Res. Toxicol.* 18:619–629.
60. Boverhof D R, Burgoon L D, Tashiro C, et al. (2005). Temporal and dose-dependent hepatic gene expression patterns in mice provide new insights into TCDD-mediated hepatotoxicity. *Toxicol. Sci.* 85:1048–1063.
61. Huang Q, Jin X, Gaillard E T, et al. (2004). Gene expression profiling reveals multiple toxicity endpoints induced by hepatotoxicants. *Mutation Res.* 549:147–168.
62. Patel S K, Ma N, Monks T J, et al. (2003). Changes in gene expression during chemical-induced nephrocarcinogenicity in the Eker rat. *Molecular Carcinogen.* 38:141–154.
63. Heinloth A N, Irwin R D, Boorman G A, et al. (2004). Gene expression profiling of rat livers reveals indicators of potential adverse effects. *Toxicol. Sci.* 80:193–202.
64. Ogawa K, Asamoto M, Suzuki S, et al. (2005). Downregulation of apoptosis revealed by laser microdissection and cDNA microarray analysis of related genes in rat liver preneoplastic lesions. *Medical Molec. Morphol.* 38:23–29.
65. Chung C Y, Seo H, Sonntag K C, et al. (2005). Cell type-specific gene expression of midbrain dopaminergic neurons reveals molecules involved in their vulnerability and protection. *Human Molec. Genetics.* 14:1709–1725.
66. Lefebvre d'Hellencourt C, Harry G J (2005). Molecular profiles of mRNA levels in laser capture microdissected murine hippocampal regions differentially responsive to TMT-induced cell death. *J. Neurochem.* 93:206–220.
67. Zhou T, Jia X, Chapin R E, et al. (2004). Cadmium at a non-toxic dose alters gene expression in mouse testes. *Toxicol. Lett.* 154:191–200.
68. Anderson N L, Anderson N G (1998). Proteome and proteomics: New technologies, new concepts, and new words. *Electrophoresis.* 19:1853–1861.
69. Yarmush M L, Jayaraman A (2002). Advances in proteomic technologies. *Ann. Rev. Biomed. Engin.* 4:349–373.
70. Anderson N L, Anderson N G (1991). A two-dimensional gel database of human plasma proteins. *Electrophoresis.* 12:883–906.
71. Kennedy S (2002). The role of proteomics in toxicology: Identification of biomarkers of toxicity by protein expression analysis. *Biomarkers.* 7:269–290.

72. Kennedy S (2001). Proteomic profiling from human samples: The body fluid alternative. *Toxicol. Lett.* 120:379–384.
73. Fountoulakis M (2001). Proteomics: Current technologies and applications in neurological disorders and toxicology. *Amino Acids.* 21:363–381.
74. Fountoulakis M, Berndt P, Boelsterli U A, et al. (2000). Two-dimensional database of mouse liver proteins: Changes in hepatic protein levels following treatment with acetaminophen or its nontoxic regioisomer 3-acetamidophenol. *Electrophoresis.* 21:2148–2161.
75. Ruepp S U, Tonge R P, Shaw J, et al. (2002). Genomics and proteomics analysis of acetaminophen toxicity in mouse liver. *Toxicol. Sci.* 65:135–150.
76. O'Farrell P H (1975). High resolution two-dimensional electrophoresis of proteins. *J. Biolog. Chem.* 250:4007–4021.
77. McGregor E, Dunn M J (2003). Proteomics of heart disease. *Human Molec. Genetics.* 12:R135–R144.
78. James P (2001). Protein expression analysis: From "tip of the iceberg" to a global method. *Disease Markers.* 17:235–246.
79. Anderson N L, Esquer-Blasco R, Richardson F, et al. (1996). The effects of peroxisome proliferators on protein abundances in mouse liver. *Toxicol. Appl. Pharm.* 137:75–89.
80. Bruno M K, Khairallah E A, Cohen S D (1998). Inhibition of protein phosphatase activity and changes in protein phosphorylation following acetaminophen exposure in cultured mouse hepatocytes. *Toxicol. Appl. Pharm.* 153:119–132.
81. Witzmann F A, Carpenter R L, Ritchie G D, et al. (2000). Toxicity of chemical mixtures: Proteomic analysis of persisting liver and kidney protein alterations induced by repeated exposure of rats to JP-8 jet fuel vapor. *Electrophoresis.* 21:2138–2147.
82. Steiner S, Aicher L, Raymackers J, et al. (1996). Cyclosporine A decreases the protein level of the calcium-binding protein calbindin-D 28kDa in rat kidney. *Biochem. Pharmacol.* 51:253–258.
83. Bandara L R, Kelly M D, Lock E A, et al. (2003). A correlation between a proteomic evaluation and conventional measurements in the assessment of renal proximal tubular toxicity. *Toxicol. Sci.* 73:195–206.
84. Wetmore B A, Merrick B A (2004). Toxicoproteomics: Proteomics applied to toxicology and pathology. *Toxicol. Pathol.* 32:619–642.
85. Anderson N L, Taylor J, Hofmann H-P, et al. (1996). Simultaneous measurement of hundreds of liver proteins: Application in assessment of liver function. *Toxicol. Pathol.* 24:72–76.
86. Thome-Kromer B, Bonk I, Klatt M, et al. (2003). Toward the identification of liver toxicity markers: A proteome study in human cell culture and rats. *Proteomics.* 3:1835–1862.
87. MacBeath G, Schreiber S L (2000). Printing proteins as microarrays for high-throughput function determination. *Science.* 289:1760–1763.
88. Zhu H, Bilgin M, Bangham R, et al. (2001). Global analysis of protein activities using proteome chips. *Science.* 293:2101–2105.
89. Petricoin E F, Zoon K C, Kohn E C, et al. (2002). Clinical Proteomics: Translating benchside promise into bedside reality. *Nature Rev. Drug Disc.* 1:683–695.
90. Petricoin E F, Rajapaske V, Herman E, et al. (2004). Toxicoproteomics: Serum proteomic pattern diagnostics for early detection of drug induced cardiac toxicities and cardioprotection. *Toxicol. Pathol.* 32:122–130.

91. Kell D B (2002). Metabolomics and machine learning: Explanatory analysis of complex metabolome data using genetic programming to produce simple, robust rules. *Molecular Biol. Reports.* 29:237–241.
92. Watkins S M, German J B (2002). Toward the implementation of metabolomics assessments of human health and nutrition. *Curr. Opin. Biotech.* 13:512–516.
93. Griffin J L, Bollard M E (2004). Metabonomics: Its potential as a tool in toxicology for safety assessment and data integration. *Curr. Drug Metabol.* 5:389–398.
94. Robertson D G (2005). Metabonomics in toxicology: A review. *Toxicol. Sci.* 85:809–822.
95. Adams A (2003). Metabolomics: Small molecule "Omics". *The Scientist.* 17:38–40.
96. Griffin J L (2004). Metabolic profiles to define the genome: Can we hear the phenotypes? *Philos. T. Roy. Soc. B.* 35, 857–871.
97. Weckwerth W, Morganthal K (2005). Metabolomics: From pattern recognition to biological interpretation. *Drug Disc. Today.* 10:1551–1558.
98. Keun H C (2006). Metabonomic modeling of drug toxicity. *Pharmacol. Therap.* 109:92–106.
99. Reo N V (2002). NMR-based metabolomics. *Drug and Chem. Toxicol.* 25:375–382.
100. Nicholson J K, Buckingham M J, Sadler P J (1983). High resolution $^{1}$H NMR studies of vertebrate blood and plasma. *Biochem. J.* 211:605–615.
101. Nicholson J K, Timbrell J A, Sadler P J (1985). Proton NMR spectra of urine as indicators of renal damage: Mercury-induced nephrotoxicity in rats. *Molec. Pharmacol.* 27:644–651.
102. Gartland K P R, Beddell C R, Lindon J C, et al. (1991). Application of pattern recognition methods to the analysis and classification of toxicological data derived from proton nuclear magnetic resonance spectroscopy of urine. *Molec. Pharmacol.* 39:629–642.
103. Holmes E, Nicholls A W, Lindon J C, et al. (1998). Development of a model for classification of toxin-induced lesions using 1H NMR spectroscopy of urine combined with pattern recognition. *NMR in Biomedicine.* 11:235–244.
104. Nicholson J K, Higham D P, Timbrell J A, et al. (1989). Quantitative high resolution $^{1}$H NMR urinalysis studies on the biochemical effects of cadmium in the rat. *Molec. Pharmacol.* 36:398–404.
105. Robertson D G, Reily M D, Sigler R E, et al. (2000). Metabonomics: Evaluation of nuclear magnetic resonance (NMR) and pattern recognition technology for rapid *in vivo* screening of liver and kidney toxicants. *Toxicol. Sci.* 57:326–337.
106. Beckwith-Hall B M, Nicholson J K, Nicholls A W, et al. (1998). Nuclear magnetic resonance spectroscopic and principal components analysis investigations into biochemical effects of three model hepatotoxins. *Chem. Res. Toxicol.* 11:260–272.
107. Moka D, Vorreuther R, Schicha H, et al. (1998). Biochemical classification of kidney carcinoma biopsy samples using magic-angle-spinning $^{1}$H nuclear magnetic resonance spectroscopy. *J. Pharmac. Biomed. Anal.* 17:125–132.
108. Waters N J, Holmes E, Williams A, et al. (2001). NMR and pattern recognition studies on the time-related metabolic effects of α-naphthylisothiocyanate on liver, urine, and plasma in the rat: An integrative metabonomic approach. *Chem. Res. Toxicol.* 14:1401–1412.
109. Coen M, Lenz E M, Nicholson J K, et al. (2003). An integrated metabonomic investigation of acetaminophen toxicity in the mouse using NMR spectroscopy. *Chem. Res. Toxicol.* 16:295–303.

110. Car B D, Robertson R T (1999). Commentary: Discovery toxicology-a nascent science. *Toxicol. Pathol.* 27:481–483.
111. Lindon J C, Nicholson J K, Holmes E, et al. (2003). Contemporary issues in toxicology: The role of metabonomics in toxicology and its evaluation by the COMET project. *Toxicol. Appl. Pharm.* 187:137–146.
112. Lindon J C (2005). The consortium for metabonomic toxicology (COMET): Aims, activities and achievements. *Pharmacogenomics.* 6:691–699.
113. Goldstein D B, Tate S K, Sisodiya S M (2003). Pharmacogenetics goes genomic. *Nature Rev. Genetics.* 4:937–947.
114. Shastry B S (2005). Genetic diversity and new therapeutic concepts. *J. Human Genetics.* 50:321–328.
115. Kalow W (1956). Familial incidence of low pseudocholinesterase level. *Lancet.* 2:576–577.
116. Carsen P R, Flanagan C L, Iokes C E (1956). Enzymatic deficiency in primaquine-sensitive erythrocytes. *Science.* 124:484–485.
117. Kalow W, Staron N (1957). On distribution and inheritance of atypical forms of human serum cholinesterase, as indicated by dibucaine numbers. *Canadian J. Biochem. Physiol.* 35:1305–1320.
118. Motulsky A G (1957). Drug reactions, enzymes and biochemical genetics. *J. Amer. Med. Assoc.* 165:835–837.
119. Kalow W (1997). Pharmacogenetics in biological perspective. *Pharmacol. Rev.* 49:369–379.
120. Ingelman-Sundberg M, Rodriguez-Antona C (2005). Pharmacogenetics of drug-metabolizing enzymes: Implications for a safer and more effective drug therapy. *Philos. T. Roy. Soc. B.* 360:1563–1570.
121. Danielson P B (2002). The cytochrome P450 superfamily: Biochemistry, evolution and drug metabolism in humans. *Curr. Drug Metabol.* 3:561–597.
122. Cascorbi I (2006). Genetic basis of toxic reactions to drugs and chemicals. *Toxicol. Lett.* 162:16–28.
123. Ingelman-Sundberg M (2005). Genetic polymorphisms of cytochrome P450 2D6 (CYP2D6): Clinical consequences, evolutionary aspects and functional diversity. *Pharmacogenomics J.* 5:6–13.
124. Barham H M, Lennard M S, Tucker G T (1994). An evaluation of cytochrome P450 isoform activities in the female dark agouti (DA) rat: Relevance to its use as a model of the CYP2D6 poor metaboliser phenotype. *Biochem. Pharmacol.* 47: 1295–1307.
125. Vorhees C V, Reed T M, Schilling M A, et al. (1998). CYP2D1 polymorphism in methamphetamine-treated rats: Genetic differences in neonatal mortality and effects on spatial learning and acoustic startle. *Neurotoxicol. Teratol.* 20:265–273.
126. Paulson S K, Zhang J Y, Jessen S M, et al. (2000). Comparison of celecoxib metabolism and excretion in mouse, rabbit, dog, cynomolgus monkey and rhesus monkey. *Xenobiotica.* 30:731–744.
127. Mise M, Yadera S, Matsuda M, et al. (2004). Polymorphic expression of CYP1A2 leading to interindividual variability in metabolism of a novel benzodiazepine receptor partial inverse agonist in dogs. *Drug Metabol. Dispos.* 32:240–245.
128. Kusano K, Seko T, Tanaka S, et al. (1996). Purfication and characterization of Rhesus monkey liver amido hydrolases and their roles in the metabolic polymorphism for E6123, a platelet-activating factor receptor antagonist. *Drug Metabol. Dispos.* 24:1186–1191.

129. Cronin M T, Fucini R V, Kim S M, et al. (1996). Cystic fibrosis mutation detection by hybridization to light-generated DNA probe arrays. *Human Mutation*. 7:244–255.

130. Kozal M J, Shah N, Shen N, et al. (1996). Extensive polymorphisms observed in HIV-1 clade B protease gene using high-density oligonucleotide arrays. *Nature Med*. 2:753–759.

131. Hacia J G, Brody L C, Chee M S, et al. (1996). Detection of heterozygous mutations in BRCA1 using high density oligonucleotide arrays and two-colour fluorescence analysis. *Nature Genetics*. 14:441–447.

132. Kulle B, Schirmer M, Toller M R, et al. (2005). Application of genomewide SNP arrays for detection of simulated susceptibility loci. *Human Mutation*. 25:557–565.

133. Bao Y P, Huber M, Wei T F, et al. (2005). SNP identification in unamplified human genomic DNA with gold nanoparticle probes. *Nucleic Acids Res*. 33:e15.

134. Witkamp R F (2005). Genomics and systems biology—how relevant are the developments to veterinary pharmacology, toxicology and therapeutics? *J. Veterinary Pharmacol. Therapeut*. 28:235–245.

135. Barabasi A L, Oltvai Z N (2004). Network biology: Understanding the cell's functional organization. *Nature Rev. Genetics*. 5:101–113.

136. Heijne W H, Kienhuis A S, van Ommen B, et al. (2005). Systems toxicology: Applications of toxicogenomics, transcriptomics, proteomics and metabolomics in toxicology. *Expert Rev. Proteomics*. 2:767–780.

137. Ekins S, Nikolsky Y, Nikolskaya T (2005). Techniques: application of systems biology to absorption, distribution, metabolism, excretion and toxicity. *Trends Pharmaceut. Sci*. 26:202–209.

138. Cunningham M J (2000). Toxicological database mining in the twenty-first century. *Toxicol. Sci*. 54(1-S):412.

139. Tomita M (2001). Whole-cell simulation: A grand challenge of the 21st century. *Trends in Biotechnol*. 19:205–210.

140. Cunningham M J, Holt G, Moyes C, et al. (2000). The use of proteomics in molecular toxicology. *Toxicol. Sci*. 54(1-S):62.

141. Keitt S K, Fagan T F, Marts S A (2004). Understanding sex differences in environmental health: A thought leaders' roundtable. *Environ. Health Perspec*. 112:604–609.

142. Hogstrand C, Balesaria S, Glover C N (2002). Application of genomics and proteomics for study of the integrated response to zinc exposure in a non-model fish species, the rainbow trout. *Compar. Biochem. Physiol. Part B, Biochem. Molec. Biol*. 133:523–535.

143. Fountoulakis M, de Vera M C, Crameri F, et al. (2002). Modulation of gene and protein expression by carbon tetrachloride in the rat liver. *Toxicol. Applied Pharmacol*. 183:71–80.

144. Bro C, Regenberg B, Lagniel G, et al. (2003). Transcriptional, proteomic and metabolic response to lithium in galactose-grown yeast cells. *J. Biolog. Chem*. 278:32141–32149.

145. Heijne W H M, Stierum R H, Slijper M, et al. (2003). Toxicogenomics of bromobenzene hepatotoxicity: A combined transcriptomics and proteomics approach. *Biochem. Pharmacol*. 65:857–875.

146. Coen M, Ruepp S U, Lindon J C, et al. (2004). Integrated application of transcriptomics and metabonomics yields new insight into the toxicity due to paracetamol in the mouse. *J. Pharmac. Biomed. Anal*. 35:93–105.

147. Schnackenberg L, Jones R C, Thyparambil S, et al. (2006). An integrated study of acute effects of valproic acid in the liver using metabonomics, proteomics, and transcriptomics platforms. *OMICS: A J. Integ. Boil.* 10:1–14.
148. Kleno T G, Kiehr B, Baunsgaard D, et al. (2004). Combination of "omics" data to investigate the mechanism(s) of hydrazine-induced hepatotoxicity in rats and to identify potential biomarkers. *Biomarkers.* 9:116–138.
149. Cunningham M J, Magnuson S R, Falduto M T (2005). Gene expression profiling of nanoscale materials using a systems biology approach. *Toxicol. Sci.* 84(1-S):9.
150. Monteiro-Riviere N, Cunningham M J (2006). Screening methods for assessing skin toxicity of nanomaterials. *Toxicol. Sci.* 90(1-S):5.
151. You A J, Jackman R J, Whitesides G M, et al. (1997). A miniaturized arrayed assay format for detecting small molecule-protein interactions in cells. *Chem. & Biol.* 4:969–975.
152. Kuruvilla F G, Shamji A F, Sternson S M, et al. (2002). Dissecting glucose signalling with diversity-oriented synthesis and small-molecule microarrays. *Nature.* 416:653–657.
153. Xia Y, Whitesides G M (1998). Soft lithography. *Angewandte Chemie (International Edition in English).* 37:550–575.
154. Kane R S, Takayama S, Ostuni E, et al. (1999). Patterning proteins and cells using soft lithography. *Biomaterials.* 20:2363–2376.
155. Ziauddin J, Sabatini D M (2001). Microarrays of cells expressing defined cDNAs. *Nature.* 411:107–110.
156. Kononen J, Bubendorf L, Kallioniemi A, et al. (1998). Tissue microarrays for high-throughput molecular profiling of tumor specimens. *Nature Med.* 4:844–847.
157. Moch H, Schraml P, Bubendorf L, et al. (1999). High-throughput tissue microarray analysis to evaluate genes uncovered by cDNA microarray screening in renal cell carcinoma. *Amer. J. Pathol.* 154:981–986.
158. Bubendorf L, Kolmer M, Kononen J, et al. (1999). Hormone therapy failure in human prostate cancer: Analysis by complementary DNA and tissue microarrays. *J. National Cancer Institute.* 91:1758–1764.
159. Matysiak S, Reuthner F, Hoheisel J D (2001). Automating parallel peptide synthesis for the production of PNA library arrays. *Biotechniques.* 31:896–904.
160. Fritz J, Baller M K, Lang H P, et al. (2000). Translating biomolecular recognition into nanomechanics. *Science.* 288:316–318.
161. McKendry R, Zhang J, Arntz Y, et al. (2002). Multiple label-free biodetection and quantitative DNA-binding assays on a nanomechanical cantilever array. *Proc. of the National Academy of Sciences.* 99:9783–9788.
162. Choi J W, Park K W, Lee D B, et al. (2005). Cell immobilization using self-assembled synthetic oligopeptide and its application to biological toxicity detection using surface plasmon resonance. *Biosens. and Bioelectron.* 20:2300–2305.
163. How S E, Yingyongnarongkul B, Fara M A, et al. (2004). Polyplexes and lipoplexes for mammalian gene delivery: From traditional to microarray screening. *Combinat. Chem. & High Throughput Screen.* 7:423–430.
164. Piner R D, Zhu J, Xu F, et al. (1999). "Dip-pen" nanolithography. *Science.* 283:661–663.
165. Park S J, Taton A T, Mirkin C A (2002). Array-based electrical detection of DNA with nanoparticle probes. *Science.* 295:1503–1506.
166. Taton A T, Mirkin C A, Letsinger R L (2000). Scanometric DNA array detection with nanoparticle probes. *Science.* 289:1757–1760.

167. Abdelrahim M, Smith R, Safe S (2003). Aryl hydrocarbon receptor gene silencing with small inhibitory RNA differentially modulates Ah-responsiveness in MCF-7 and HepG2 cancer cells. *Mol. Pharmacol.* 63:1373–1381.
168. Krichevsky A M, King K S, Donahue C P, Khrapko K, Kosik K S (2003). A microRNA array reveals extensive regulation of microRNAs during brain development. *RNA* 9:1274–1281.
169. Miska E A, Alvarez-Saavedra E, Townsend M, Yoshii A, Sestan N, Rakic P, Constantine-Paton M, Horvitz H R (2004). Microarray analysis of microRNA expression in the developing mammalian brain. *Genome Biol.* 5:R68.
170. Dalmay T (2006). Short RNAs in environmental adaptation. *Proc. Biol. Soc.* 273:1579–1585.
171. Yang Z, Wu J (2006). Small RNAs and development. *Med. Sci. Monit.* 12:RA125–RA129.
172. Ying S Y, Chang D C, Miller J D, Lin S L (2006). The microRNA: Overview of the RNA gene that modulates gene function. *Methods Mol. Biol.* 342:1–18.
173. Downing G J, Battey J F (2004). Technical assessment of the first 20 years of research using mouse embryonic stem cell lines. *Stem Cells.* 22:1168–1180.
174. Ruhnke M, Ungefroren H, Zehle G, et al. (2003). Long-term culture and differentiation of rat embryonic stem cell-like cells into neuronal, glial, endothelial, and hepatic lineages. *Stem Cells.* 21:428–436.
175. Bremer S, Hartung T (2004). The use of embryonic stem cells for regulatory developmental toxicity testing in vitro—the current status of test development. *Curr. Pharmac. Design.* 10:2733–2747.

# PRECLINICAL PHARMACOKINETICS

Donald E. Mager and Murali Ramanathan
*University at Buffalo, The State University of New York, Buffalo, New York*

**Chapter Contents**

1 Introduction
2 Disposition of Macromolecules
  2.1 Absorption
  2.2 Distribution
  2.3 Metabolism
  2.4 Excretion
3 Pharmacokinetic Data Analysis
  3.1 Noncompartmental Analysis
  3.2 Pharmacokinetic Models
4 Predicting Macromolecule Pharmacokinetics in Humans
  4.1 Allometric Scaling
  4.2 Physiologically Based Pharmacokinetic Modeling
5 Conclusions
  Acknowledgment
  References

## 1 INTRODUCTION

The preclinical pharmacokinetics of protein pharmaceuticals is distinctively different from small-molecule drugs. The phase I and phase II metabolizing enzymes, for example, the cytochrome P450 and glucoronosyl transferases, which play dominant roles in the metabolism of small-molecule drugs, are not significantly involved in protein metabolism. Instead, because all protein pharmaceuticals invariably share a common polypeptide backbone, they are susceptible to proteolysis as a clearance mechanism. However, the disposition, pharmacokinetics,

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

and pharmacodynamics of proteins can be modulated by a variety of additional factors, including glycosylation and receptor binding, binding proteins, and neutralizing antibodies.

## 2 DISPOSITION OF MACROMOLECULES

### 2.1 Absorption

During drug development, the oral route is often targeted as a preferred route of administration for small molecules. However, the oral route is not usually an option for protein pharmaceuticals. Each organ in the gastrointestinal tract contains a variety of digestive proteases that enable it to break down nutritional proteins to small peptides and amino acids that can be absorbed in the intestine via secondarily active transport mechanisms. These digestive proteases also cause the extensive hydrolysis of the polypeptide bond in protein drugs in the gastrointestinal tract resulting in poor bioavailability. Protein pharmaceuticals are usually, therefore, administered via injection, such as the intravenous, subcutaneous, or intramuscular routes.

The absorption of protein drugs from extravascular sites is dependent on the size of the molecule. Upon subcutaneous administration, small molecules with molecular weights less than 1 kD predominantly enter the systemic circulation via blood capillaries, whereas subcutaneously administered proteins with molecular weights greater than 16 kD enter the systemic circulation predominantly via the lymphatic system. The extent of the lymphatic recovery after subcutaneous administration is a linear function of molecular weight (Figure 1) [1–5].

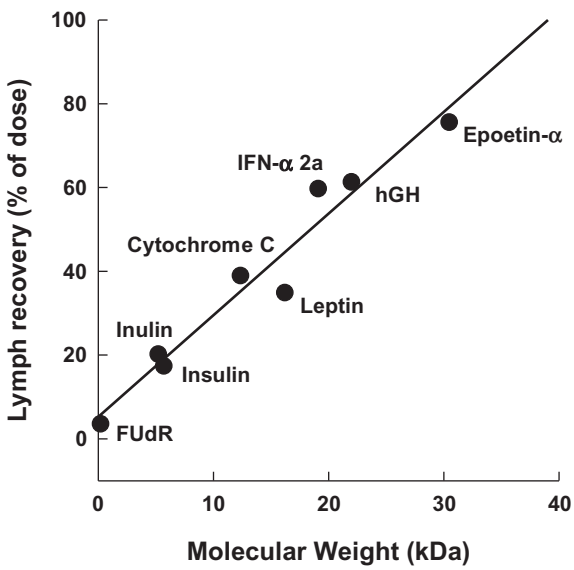


**Figure 1.** Linear relationship between cumulative lymphatic recovery and molecular weight of various therapeutic peptides and proteins. Data were extracted from the literature [1–5].

Nonparenteral routes of administration for protein drugs in general are not extensively used; however, recombinant human deoxyribonuclease I solution (dornase alfa or PULMOZYME) and, more recently, insulin powder by inhalation (EXUBERA) have been approved for clinical use in the United States [6]. The approval of inhaled insulin may provide the stimulus for greater interest in the inhalation route for other protein drugs in the near future. The lung presents a large surface area for transport and high permeability compared with other barriers. Inhaled insulin is administered as a novel powder formulation that is metered from a propellant-free metered dose inhaler. However, delivery to the lungs, particularly to the alveolar regions deep in the lung, presents several technological challenges and the overall system bioavailability (which is based on the amount of drug loaded into the inhaler device) of inhaled insulin is 9–22% that of the subcutaneous route. The reduced bioavailability is managed by dose adjustment and compared with subcutaneously administered regular insulin, the serum insulin concentrations after inhalation peak earlier (at 7–80 minutes compared with 42–274 minutes for subcutaneous regular insulin) and decay more rapidly; the pharmacodynamics of glucose changes are also correspondingly more rapid [7]. The duration of inhaled insulin action are comparable with regular insulin [6].

## 2.2 Distribution

The distribution of protein drugs is limited compared with that of small-molecule drugs and is primarily a function of the physico-chemical properties of the macromolecule and the physiological characteristics of the vasculature and cellular membranes [8]. Primary molecular descriptors, such as size, charge, and lipophilicity, will determine or influence the extent of specific and nonspecific protein binding and passage through various types of capillary endothelia [8]. The relatively large size of peptide and protein drugs alone can often account for the limited volume of distribution of these compounds, which is generally restricted to the extracellular space. Following intravenous injection, the initial volume of distribution for most macromolecules is confined approximately to plasma volume (3–8 L), and the total volume at steady state is typically only up to two fold greater.

Active transport and receptor-mediated uptake also may influence the biodistribution of macromolecules. For example, the internalization of monoclonal antibodies into cells may occur via Fcγ receptors, found on various immune and hematopoietic cells [9], or binding to membrane-localized antigens [10]. Thus, active uptake processes can serve to increase the volume of distribution to a large extent relative to that which would be expected based only on physico-chemical properties. Receptor-mediated endocytosis may represent a significant elimination pathway as well (see the Receptor-mediated elimination subsection) and is recognized as a major process influencing the overall disposition of macromolecules [11].

One must carefully interpret the volumes of distribution of peptides and proteins reported in the literature. Most studies rely on a so-called noncompartmental analysis to estimate primary pharmacokinetic parameters (see Section 3.1). However, this method is only valid for linear systems, assuming that the site of drug elimination is in rapid equilibrium with the sampling site (plasma). The former

condition is often assumed for compartmental models as well, including those with nonlinear mechanisms of elimination. Therefore, uptake processes and relatively slow or complex tissue catabolism may confound reported distribution volumes of select macromolecules estimated by these common methods.

***Binding Proteins.*** For some therapeutic proteins, specific and nonspecific protein interactions can modulate distribution and activity. Soluble receptors are the best-characterized specific binding proteins and $\alpha_2$ macroglobulin, an acute phase protein, has been implicated in the binding of several cytokines. Binding and neutralizing antibodies are another class of binding protein that will be discussed separately in the section on Immunogenicity, Binding, and Neutralizing Antibodies.

Soluble cytokine receptors (SR) occur naturally in human serum and urine and play a role in determining the disposition and biological activities of cytokines in health and in pathological states. Additionally, bioengineered soluble receptors and their immunoglobulin derivatives have been investigated as specific cytokine-inhibiting therapeutic agents [12, 13]. However, the physiological role(s) of naturally occurring SR is controversial and not well understood.

As expected, the SR can inhibit cytokine bioactivity by competing with the receptor and, when these antagonist-like effects are observed, the dose-response curve of the cytokine effect as a function of SR dose usually decreases monotonically, and almost complete inhibition is observed at the highest SR doses. In animal models, soluble receptors for interleukin-4 [14], interleukin-7 [15], interleukin-6 [15, 16], and tumor necrosis factor-$\alpha$ [17] have been shown to enhance rather than inhibit cytokine responses under certain conditions [17–19]. When these agonist-like effects are observed, the dose-response curve of the cytokine effect as a function of SR dose is bell-shaped, with no inhibition of cytokine bioactivity at low SR doses, maximal cytokine bioactivity at intermediate SR doses, followed by increasing inhibition of cytokine activity at higher SR doses. As the SR effects of the cytokines biological response is dependent on experimental conditions and can be either agonist-like or antagonist-like, the physiological roles of SR have been difficult to assess. Cytokines generally have short half-lives consistent with their physiological roles in cell-cell communication and the apparent SR-mediated agonism of cytokine activity occurs because the SR stabilizes the cytokine in vivo either by reducing the rate of spontaneous activity decay [17] or by increasing the half-life of cytokine [20]. The formation of SR-cytokine complex reduces cytokine-free concentrations, but the increases in circulation time can offset the effect of reduced receptor occupancy and contribute to overall increases in exposure [20]. The binding half-life of insulin-like growth factor-1 (IGF-1) in rats deficient in IGF-1 binding protein is 20–30 min; but in rats with normal levels of the binding protein, the half-life of IGF-1 increases 8–12-fold to 4 hours [21]. In the plasma, nearly 40% of human growth hormone (GH) is also bound to binding proteins; the high-affinity component of the GH binding activity consists of the extracellular domain of the GH receptor [22, 23]. The metabolic clearance of the SR-bound GH fraction is 10-fold slower than free GH because the bound GH is not filtered by the kidney and the competitive inhibition of receptor binding reduces elimination via receptor-mediated pathways [24]. Indirect evidence for the physiological role of SR binding is provided by the data from short-stature Mountain Ok people of Papua, New

Guinea who have SR levels 50% that of controls of normal stature; the levels of GH, IGF-1, and low-affinity GH-binding protein levels are comparable in the two groups [25].

***Immunogenicity, Binding, and Neutralizing Antibodies.*** The bioactivity and distribution of therapeutically administered proteins can be modulated by their immunogenicity; humoral responses to therapeutic proteins can result in binding and neutralizing antibodies. Generally, chronic or intermittently administered proteins are more likely to be affected by these pathways than proteins administered for acute conditions. The underlying protein structure, aggregation state, and protein formulation can contribute to the risk of binding and neutralizing antibodies. Factors related to the dosing regimen such as dose, frequency, and route of administration are also important determinants. The role of disease-related and individual-specific variables to the development of antibodies is poorly understood. The time to development of antibody depends on the dosing regimen and immunogenicity of the preparation. For some proteins, the development of antibodies to endogenous proteins can cause serious adverse events (e.g., red cell aplastic anemia can occur with anti-erythropoeitin antibodies). The impact of antibodies on drug efficacy in clinical trials should be assessed during the period the antibodies are present.

As a case study, consider interferon-β, an endogenous cytokine, recombinant forms of which are used chronically in multiple sclerosis (MS). Two types of IFN-β exist: IFN-β-1a and IFN-β-1b. IFN-β-1a is glycosylated and has the same amino acid sequence as endogenous IFN-β. IFN-β-1b is not glycosylated, lacks the N-terminal methionine, and is engineered with cysteine replaced by serine. Two approved IFN-β-1a products exist that are dosed by the once-weekly, intramuscular, 30-μg and three-times-weekly, subcutaneous, 44-μg dosing regimens; the IFN-β-1b product is an every-other-day, subcutaneous, 250-μg dosing regimen. As with other protein drugs, some patients develop neutralizing antibodies (NAB) upon chronic administration of all IFN-β products. The proportion of patients developing NAB ranges from 25% for three-times-weekly, subcutaneous IFN-β-1a regimen to 2% for the once-weekly, IM IFN-β-1a regimen [26]. In a clinical trial where the weekly IM dose in the once-weekly, IM IFN-β-1a regimen was increased from 30 μg to 60 μg, the percentage of patients developing NAB titers greater than 20 after 3 years of treatment increased from 2.2% to 5.8% and suggests that, for a given route of administration, the frequency of IFN-β-1a administration is probably the major factor influencing the immunogenicity of an IFN-β product and the increase in dosage has relatively less impact. Published data from large randomized clinical trials demonstrate that disease activity in NAB-positive patients became similar to placebo-treated patients [27]. Interestingly, skin necrosis, an adverse effect that occurs with SC IFN-β administration, occurred only in the NAB-negative patients [27]. Despite the evidence from large clinical trials, some controversy exists regarding the appropriate clinical decisions for treatment of NAB-positive MS patients, especially those with a clinically stable disease, because (1) many NAB-positive patients test negative over the course of treatment [28], (2) no accepted IFN-β treatment biomarkers exist that are MS-relevant [29, 30] and, (3) market and competitive factors occur because of the availability of multiple IFN-β products for MS.

## 2.3 Metabolism

The cytochrome P450 enzymes, expressed in the endoplasmic reticulum of the liver, kidney, and intestine, are the predominant metabolizing pathways for small-molecule drugs. After the phase I oxidation reactions, small molecules typically are further modified by phase II conjugation reactions mediated by glucoronosyl transferases that render them hydrophilic for renal elimination. However, as noted earlier, the metabolism of protein drugs does not involve these small-molecule pathways: Endogenous pathways involved in protein homeostasis mediate protein drug metabolism.

As indicated earlier, proteolysis is the predominant mechanism for protein drug metabolism, but diverse endogenous proteolytic pathways exist in humans. Further, the half-lives of protein *in vivo* are, not surprisingly, regulated by a variety of intracellular mechanisms. The half-life of intracellular proteins is regulated via the ubiquitin pathway, which targets proteins to degradation in a large multimeric protease complex referred to as the proteasome. Ubiquitin is a widely expressed globular protein with a protruding carboxyl-terminus that tags proteins for degradation. Ubiquitin-activating enzymes activate through a process that requires ATP, and the activated ubiquitin is then transferred to a lysine on the target protein by another enzyme, the ubiquitin ligase. Multiple ubiquitin chains can be ligated on a single target protein lysine by conjugating glycine-lysine bonds between successive ubiquitin chains. Some target proteins are also tagged by binding proteins referred to as "recognin" that bind sequence elements called "degrons" that specify half-life. Sufficiently ubquitinated proteins are recognized and degraded in the proteasome. Although the ubiquitin pathway is critical for cellular protein homeostasis, its involvement in the metabolism of exogenously administered proteins is probably minor.

Some circulating proteins, most notably albumin and immunoglobulin G (IgG), have anomalously long half-lives because they are protected from degradation by specific mechanisms after cellular uptake. The best characterized of these protein-stabilizing mechanisms is the neonatal Fc receptor or FcRn, which is present on a variety of cell types and tissues including the kidney, liver, and lung endothelial cells, hepatocytes, monocytes, and intestinal macrophases. It confers a long half-life to IgG (mean half-life in humans = 23 days) relative to other types of immunoglobulins (half-lives for IgA, IgD, IgE, and IgM are 6, 3, 2.5, and 5 days, respectively). Most proteins generally undergo degradation in lysosome after pinocytosis and receptor-mediated endocytosis from the cell surface. However, the constant Fc region of immunoglobulin G and antibody-based drugs are internalized by pinocytosis or, when they bind, the cell surface antibody-binding Fcγ receptors. However, upon internalization into the endosomes, immunoglobulin G is bound by FcRn and the complex, upon endocytosis, is protected from proteolysis and recycled to the cell surface instead of being processed for degradation in lysosomes. The FcRn receptor for IgG is often referred to as the Brambell receptor after the authors [31] who postulated their existence to explain the relatively long half-life and the increased rates of IgG degradation at higher IgG concentration. Mice that are $\beta_2$ microglobulin-deficient do not express FcRn and have IgG half-lives that are 10–15-fold shorter than mice with intact FcRn [32–34].

Overall, 3 major mechanisms for clearing protein drugs exist: (1) renal, (2) hepatic, and (3) receptor-modulated.

### 2.4 Excretion

The rate of elimination of proteins is strongly dependent on the size, structure, and, in the case of chronically administered protein drugs, the presence or absence of antidrug antibodies. Renal clearance is relatively uncommon for protein drugs with molecular weights greater than 50 kD. Specific primary structural features such as the presence of a humanized immunoglobulin constant (Fc) region can confer prolonged stability via endogenous recycling pathways. Likewise, glycosylation and the addition of polyethylene glycol moieties (a synthetic modification often referred to as PEGylation) can also prolong circulation times.

Protein drugs with molecular weights less than 20 kD are filtered out of circulation in the Bowman's capsule of glomerulus of the kidney and enter the luminal space of the nephron. Even under pathological conditions, urinary excretion of intact protein is unusual. The filtered protein drugs are reabsorbed from the lumen and released into the circulation either as intact drugs or degraded into peptides. The proximal tubule is the predominant site for reabsorption, which transfers proteins into apical vacuoles that fuse with lysosomes. Lysosomal proteases degrade proteins into amino acids that re-enter the circulation. Small peptides can also be directly degraded by proteases on the luminal side of the renal tubule and the resulting amino acids can be reabsorbed by transport mechanisms.

Hepatic metabolism is important to the clearance of therapeutic proteins with molecular weights greater than 20 kD, although renal clearance can contribute to proteins as large as 50 kD. The Kupfer cell and hepatocytes of the liver are involved in clearing protein drugs.

***Pegylation.*** Pegylation refers to the covalent modification of proteins by polyethylene glycol (PEG) moieties. As PEG is a hydrophilic, degradation-resistant polymer, pegylation can increase the circulating half-life of proteins, reduce renal elimination, alter the distribution, and reduce the immunogenicity of proteins [35]. The addition of PEG increases the molecular weight of the protein, which is instrumental in reducing renal filtration, and it provides a hydrophilic shield containing tightly associated water molecules that protects against proteolysis and some nonspecific interactions that mitigate therapeutic protein immunogenicity. However, pegylation can reduce receptor-binding affinity. When engineering pegylation during drug development, the effects of reduced receptor binding can be offset by the increased drug exposures; the improvements in circulating half-life and protein disposition effects and reduced immunogenicity benefit of this modification. The chemistry of PEG conjugation involves chemical activation of PEG followed by covalent attachment at the reactive amino groups of lysine or the N-terminal amino acid or the sulfhydryl groups of cysteine. As a rule of thumb, the molecular weight of the PEG moiety has to be comparable with that of the therapeutic protein for pegylation to be effective. Pegylation chemistry is also capable of conjugating branched PEG polymers that can increase the mass of PEG added at a each site and can better protect the protein therapeutic from degradation and immune recognition [36]. The pegylation approach is quite general and has

also been used for liposomes: It prolongs the circulation times of liposomes by reducing the rapid elimination by the reticuloendothelial system.

Interferon-α-2b (IFN-α-2b, PEGINTRON), which is used for treating hepatitis, is a case study on the impact possible with the pegylation approach because considerable preclinical and clinical data are also available for the IFN-α-2b unmodified by pegylation (INTRON-A). As a result of pegylation, the molecular weight of IFN-α-2b is increased from 19 kilodaltons to approximately 31 kilodaltons and the half-life increases from 3–8 hours to 65 hours. Other pharmacokinetic variables altered include a decreased volume of distribution (from 31–73 L to 8–12 L) and decreased systemic clearance (from 6.6–29.2 L/hour to 0.06–0.1 L/hour). The pharmacodynamic effects of antiviral and antitumor activities of pegylated IFN-α-2b are 12–136-fold and 18-fold, respectively, greater than the unpegylated form. The net result of these pharmacokinetic and pharmacodynamic improvements is that the frequency of pegylated IFN-α-2b administration for treating chronic hepatitis C is once weekly compared with three times weekly for the unpegylated form.

***Hyperglycosylation.*** Hyperglycosylation is the addition of sugar moieties to proteins and can be engineered using molecular biology techniques to specific locations on a protein. Many cellular proteins have N-linked or O-linked glycan modification that are added in posttranslational and cotranslational processes in the endoplasmic reticulum and in the Golgi apparatus [37–39]. The N-linked glycosylations occur at amino groups of the asparagine in Asn-X-Ser/Thr sequences, although not all potential glycosylation sites may be modified and heterogeneity can occur within given protein molecules. The O-linked glycosylations occur at the hydroxyl groups of serine and threonine.

***Receptor-Mediated Elimination.*** For some proteins, pharmacokinetic nonlinearities can occur because of receptor-mediated elimination. When therapeutic proteins bind their cognate cell-surface targets, the bound therapeutic protein can be internalized by receptor-mediated or fluid-phase endocytosis and traffic to the lysosomes where it may be degraded (Figure 2) [40]. The capacity-limited nature of these systems can manifest as saturable dose-dependent drug clearance and may be susceptible to further modulation through functional adaptation mechanisms such as receptor upregulation or downregulation. The saturable internalization of erythropoietin (EPO) in erythroid progenitor cells is a classic example [41]. Chapel et al. have shown that the bone marrow, a major site for erythropoietic cells, represents a significant EPO elimination pathway [42]. Busulfan or 5-fluoruracil-induced bone marrow ablation in sheep was associated with a decrease in EPO clearance, revealing the *in vivo* significance of target-mediated uptake and elimination for this therapeutic protein.

Complex target-mediated clearance mechanisms suggest that both pharmacokinetics and pharmacodynamics might have to be considered concurrently to understand the disposition of select macromolecules. Peptide and protein drugs that alter the expression of target cells may have the ability to modulate their pharmacokinetic properties. For example, thrombopoietin (TPO) is an endogenous hematopoietic growth factor that stimulates the proliferation and differentiation of megakaryocytes and thus the production of new platelets, which, in turn, would yield a net increase in the density of available TPO receptors (c-Mpl) and reconcile

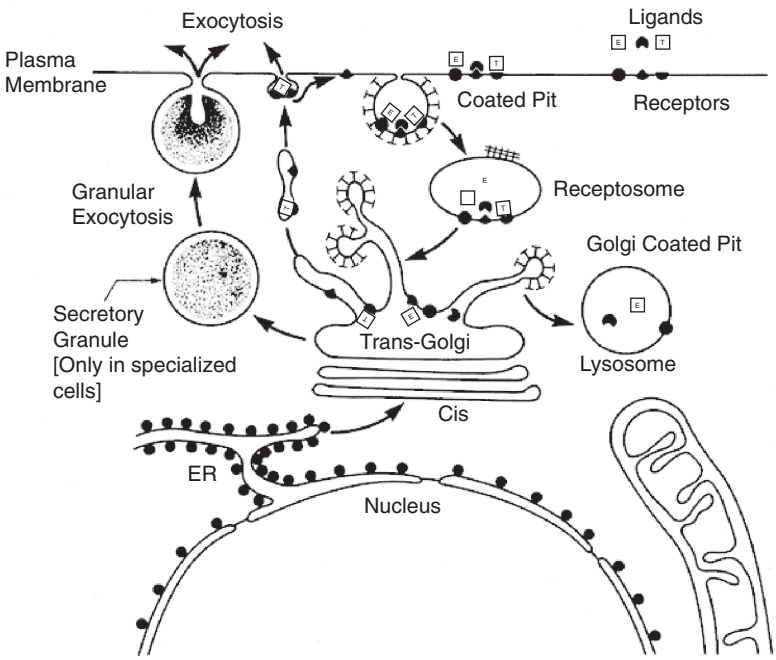


**Figure 2.** Schematic for the pathway of receptor-mediated endocytosis. Example ligands include EGF (E), transferrin (T), and $\alpha_2$-macroglobuilin (◠), and they are shown along with their respective receptors (●, ▬, and ▲). These ligands are examples of systems in which both ligand and receptor are transferred to lysosomes (EGF), both ligand and receptor are recycled to the cell surface (transferrin), and the ligand is delivered to lysosomes but the receptor is efficiently recycled ($\alpha_2$-macroglobuilin). For some systems, receptors may undergo significant turnover in the absence of ligand. (Figure and legend reproduced with permission [from Ref. 40].)

the reciprocal relationship between platelet count and circulating plasma TPO concentrations [43].

## 3 PHARMACOKINETIC DATA ANALYSIS

The primary objective of pharmacokinetic data analysis is to characterize quantitatively the drug disposition processes discussed in Section 2. Sufficient knowledge of the preclinical pharmacokinetics of macromolecules, which can only be obtained via mathematical formalisms, may be used to anticipate relevant pharmacokinetic properties in humans (Section 4), thereby serving to guide dose selection in subsequent clinical trials. The two most commonly used methods for characterizing pharmacokinetic data are noncompartmental analysis and compartmental modeling. Basic assumptions and techniques are presented in this section and are placed within the context of biotechnology pharmaceuticals, which may provide guidance for selecting the most appropriate and desirable method for a given dataset. It is important to recognize that no method of data analysis can offset shortcomings in

experimental designs and assays, and that careful consideration of such details is a sine qua non in pharmacokinetics [44].

### 3.1 Noncompartmental Analysis

***Intravenous Drug Disposition.*** The estimation of primary pharmacokinetic parameters using noncompartmental analysis is based on statistical moment theory [45, 46]. The relationships defined by this theory are valid under the assumption that the system is linear and time-invariant. For simplicity, we further assume that drug is irreversibly removed only from a single accessible pool (e.g., plasma space). Regardless of the route of administration, the temporal profile of plasma drug concentrations, $C_p(t)$, can represent a statistical distribution curve. As such, the zeroth and first statistical moments ($M_0$ and $M_1$) are defined as:

$$M_0 = \int_0^\infty C_p(t)\,dt = AUC \tag{1}$$

$$M_1 = \int_0^\infty t \cdot C_p(t)\,dt = AUMC \tag{2}$$

where $AUC$ and $AUMC$ are the so-called area under the plasma concentration-time curve and area under the first-moment curve. Several methods have been described for estimating the values of these moments, some of which will be presented in this section, and may be used to calculate relevant pharmacokinetic parameters.

Under linear conditions, the time-course of plasma drug concentrations following a single IV bolus dose can be characterized by a sum of exponentials:

$$C_p(t) = \sum C_i \cdot e^{-\lambda_i \cdot t} \tag{3}$$

where the coefficients ($C_i$) and exponentials ($\lambda_i$) may be obtained via curve-stripping, or more appropriately, from curve-fitting using nonlinear regression analysis. One advantage of fitting Equation (3) to pharmacokinetic data is the relative simplicity in calculating $AUC$ and $AUMC$ values [47]:

$$AUC = \sum C_i / \lambda_i \tag{4}$$

$$AUMC = \sum C_i / \lambda_i^2 \tag{5}$$

Primary pharmacokinetic parameters such as total systemic clearance ($CL$) and the volume of distribution at steady state ($V_{ss}$) can be defined as:

$$CL = D_{iv} / AUC \tag{6}$$

and

$$V_{ss} = \frac{D_{iv} \cdot AUMC}{AUC^2} = CL \cdot MRT_{iv} \tag{7}$$

where $D_{iv}$ represents the IV bolus dose and $MRT_{iv}$, or mean residence time, is the average time a drug molecule resides in the system before being irreversibly removed. Also, the central volume of distribution ($V_c$) can be calculated as:

$$V_c = D_{iv} \Big/ \sum C_i \tag{8}$$

Additional advantages to resolving *AUC* and *AUMC* from curve-fitting the sums of exponentials are the stability of the estimates and that nonlinear regression procedures provide measures of error about the pharmacokinetic parameter estimates.

Peptide and protein drugs requiring an IV route of administration may also be given as a constant rate infusion. Pharmacokinetic data from experiments designed to achieve a steady-state plasma concentration ($C_{ss}$) allow for the direct calculation of drug clearance according to the equation:

$$CL = k_0 / C_{ss} \tag{9}$$

where $k_0$ is the constant zero-order rate of drug infusion. The steady-state volume of distribution can be obtained from:

$$V_{ss} = \frac{k_0 \cdot T - CL \cdot AUC_0^T}{C_{ss}} \tag{10}$$

where $T$ is the duration of the IV infusion. For short-term constant rate infusions, when plasma concentrations do not attain $C_{ss}$, Equation (6) is valid for calculating clearance by substituting $k_0 \cdot T$ for $D_{iv}$. Moment analysis also may be used to generate $V_{ss}$ under these conditions from the equation:

$$V_{ss} = \frac{k_0 \cdot T \cdot AUMC}{AUC^2} - \frac{k_0 \cdot T^2}{2 \cdot AUC} \tag{11}$$

***Extravascular Drug Disposition.*** In general, peptide and protein drugs exhibit poor systemic bioavailability following oral administration owing primarily to gastrointestinal enzymatic degradation. On the other hand, IV drug administration may not result in a desirable time-course of drug exposure and certainly lacks appeal for patients requiring chronic therapy. One of the most common routes of extravascular peptide and protein drug administration is subcutaneous injection, with additional routes, such as inhalation, intranasal, and transdermal delivery representing the targets of current investigation [48].

The time-course of plasma drug concentrations following extravascular administration is often irregular, and simple analytical solutions for calculating *AUC* and *AUMC*, such as Equations (4) and (5), are not readily available. For these cases, numerical integration methods may be used to evaluate the integrals in Equations (1) and (2), such that:

$$AUC = \int_0^{t^*} C_p(t)dt + \frac{C_p^*}{\lambda_z} \tag{12}$$

and

$$AUMC = \int_0^{t^*} t \cdot C_p(t)dt + \frac{t^* \cdot C_p^*}{\lambda_z} + \frac{C_p^*}{\lambda_z^2} \tag{13}$$

where $C_p^*$ and $t^*$ are the last observed plasma concentration and time values, respectively, $\lambda_z$ is the terminal slope of the curve, and the quotients provide estimates of the area extrapolating from $t^*$ to infinity. The trapezoidal and log-trapezoidal rules are the most common methods for numerically evaluating the integrals in Equations (12) and (13) [49]. The trapezoidal rule for calculating the area under a curve within an interval $t_i$ and $t_{i-1}$ is given as:

$$\int_{i-1}^{i} y \cdot dy = 0.5 \cdot [(y_i + y_{i-1})(t_i - t_{i-1})] \tag{14}$$

where $y$ is $C_p(t)$ for $AUC$ and $t \cdot C_p(t)$ for $AUMC$. For the log-trapezoidal rule, the equation is:

$$\int_{i-1}^{i} y \cdot dy = \frac{1}{\ln\left(\frac{y_i}{y_{i-1}}\right)} \cdot [(y_i + y_{i-1})(t_i - t_{i-1})] \tag{15}$$

The cumulative sum of the interval areas thus provides an estimate of the $AUC$ and $AUMC$ from time zero to $t^*$. Yu and Tse have evaluated the performance of several numerical integration algorithms for calculating $AUC$ [49]. In practice, it is most common to employ the trapezoidal rule for ascending ($y_i > y_{i-1}$) and plateau ($y_i \approx y_{i-1}$) regions of the curve and the log-trapezoidal rule when measured values are decreasing ($y_i < y_{i-1}$). Such an approach is used for noncompartmental analysis in the computer program WinNonlin (Pharsight, Mountain View, CA), a nonlinear regression analysis program specifically designed for pharmacokinetic and pharmacodynamic data analysis.

The noncompartmental analysis of pharmacokinetic data after extravascular drug administration, when coupled with that of IV dosing, can yield additional relevant pharmacokinetic parameters, particularly regarding absorption processes. For example, the systemic availability ($F$), which represents the net fraction of the drug dose reaching the systemic circulation after extravascular administration, is defined as:

$$F = \frac{D_{iv} \cdot AUC_{ev}}{D_{ev} \cdot AUC_{iv}} \tag{16}$$

where catabolism at the extravascular site of administration and other processes may result in incomplete drug absorption. Also, the mean absorption time ($MAT$), or the average time a drug molecule remains unabsorbed, can be calculated from the relationship:

$$MAT = MRT_{ev} - MRT_{iv} \tag{17}$$

where the mean residence times derive from the ratios of *AUMC* and *AUC* [e.g., Equation (7)].

Overall, the relative simplicity of noncompartmental analysis, as compared with the methodological challenges of developing and validating structural mathematical models [50], forms the basis for it being the most common method for ascertaining the major pharmacokinetic properties of drugs. It is beyond the scope of this chapter to contrast the merits of these two methods, and the reader is referred to more complete treatments and the references therein [44, 51]. However, as discussed in Section 2, various pharmacokinetic processes of many peptide and protein pharmaceuticals are inherently nonlinear, which violate the fundamental assumptions of linearity and time-invariance in statistical moment analysis. For such compounds, the construction of mechanism-based mathematical models is required and essential for characterizing the *in vivo* pharmacokinetics of the drug (Section 3.2). In any event, noncompartmental analysis represents a highly useful starting point for any pharmacokinetic data analysis [44]. For simple linear systems, this technique may be sufficient for calculating primary pharmacokinetic parameters from preclinical experiments and anticipating dose-concentration relationships in humans (Section 4.1). Additionally, moment analysis of plasma concentration-time profiles resulting from a suitable range of dose levels may be used initially to identify whether nonlinearities exist, such as disproportionate *AUC* values (relative to dose) indicating dose-dependent clearance or bioavailability, thus necessitating the implementation of more sophisticated methods of pharmacokinetic systems analysis.

### 3.2 Pharmacokinetic Models

Compartmental modeling involves the specification of a structural mathematical model (commonly using either explicit or ordinary differential equations) and system parameters are estimated from fitting the model to pharmacokinetic data via nonlinear regression analysis or population mixed effects modeling. The methodological issues involved in the development, application, and interpretation of pharmacokinetic models are beyond the scope of this chapter and are discussed in detail elsewhere [50, 52]. The most popular structural model applied to macromolecule pharmacokinetic data is the open two-compartment model shown in Figure 3. This system is described, in general, by the following differential equations:

$$\frac{dA_p}{dt} = \frac{Input}{V_c} - \left(\frac{CL}{V_c} + \frac{CL_D}{V_c}\right) \cdot A_p + \frac{CL_D}{V_t} \cdot A_t \tag{18}$$

$$\frac{dA_t}{dt} = \frac{CL_D}{V_c} \cdot A_p - \frac{CL_D}{V_t} \cdot A_t \tag{19}$$

where $A_i$ represents the amount of drug in the plasma ($i = p$) or tissue ($i = t$) compartments, *Input* represents a function describing the appearance of drug based on the route of drug administration, $CL_D$ is the distributional clearance between compartments, and $V_t$ is the volume of the second or tissue compartment. Plasma drug concentrations are thus specified as $C_p = A_p/V_c$.

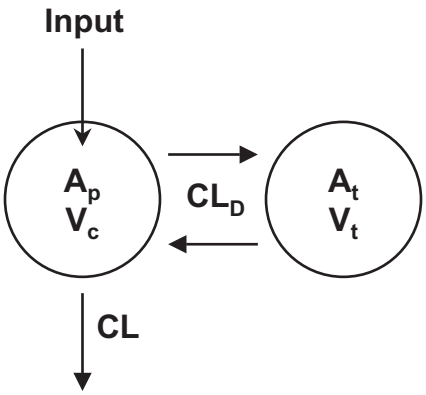


**Figure 3.** Schematic of a two-compartment mammillary model commonly used to characterize macromolecule pharmacokinetics. Drug input and elimination ($CL$) occur to and from a central compartment ($A_p$) and may distribute ($CL_D$) to a peripheral site ($A_t$, $V_t$). Plasma drug concentrations reflect the amount in the central compartment relative the volume of distribution of that site ($V_c$).

**TABLE 1. Input Functions for Specific Pharmacokinetic Applications**

$D$, dose; $F$, bioavailability; $fr$, fraction of dose undergoing first-order input in a dual absorption model; $K_0$, zero-order input rate; $k_a$ and $k'_a$, first-order absorption rate constants; $MAT$, mean absorption time; $NV^2$, normalized variance of Gaussian density function, $t$, nominal time; $t_{lag}$, duration of time-lag; $\tau$, duration of rapid input in dual absorption model; $T_{inf}$, duration of zero-order IV infusion; $T$, modulus time.

| Condition | Input Function | | Equation |
|---|---|---|---|
| Rapid intravenous injection | 0 | | — |
| Intravenous infusion | $K_0$ | $t \le T_{inf}$ | — |
| | 0 | $t > T_{inf}$ | |
| Linear first-order absorption with a time-lag | 0 | $t \le t_{lag}$ | 20 |
| | $F \cdot D \cdot k_a \cdot e[-k_a \cdot (t - t_{lag})]$ | $t > t_{lag}$ | |
| Two consecutive first-order processes | $F \cdot D \cdot \frac{k_a \cdot k'_a}{(k'_a - k_a)} \cdot [e(-k_a \cdot t) - e(-k'_a \cdot t)]$ | | 21 |
| Rapid zero-order input followed by linear first-order absorption | $F \cdot D \cdot (1 - fr)/\tau$ | $t \le \tau$ | 22 |
| | $F \cdot D \cdot fr \cdot k_a \cdot e[-k_a \cdot (t - \tau)]$ | $t > \tau$ | 23 |
| Inverse Gaussian | $F \cdot D \cdot \sqrt{\frac{MAT}{2\pi \cdot NV^2 \cdot T^3}} \cdot e\left[-\frac{(T - MAT)^2}{2 \cdot NV^2 \cdot MAT \cdot T}\right]$ | | 24 |

The drug input function in Equation (18) will depend on the experimental design and several common functions are listed in Table 1. For rapid IV injection, the input function is set equal to zero and the dose ($D$) is reflected in the initial condition for Equation (18) [$A_p(0) = D$]. A constant rate IV infusion also is easily handled by setting *Input* equal to the zero-order infusion rate ($K_0$) for time ($t$) less

than or equal to the infusion time ($T_{inf}$), otherwise $Input = 0$. The remaining equations in Table 1 typically are used to describe extravascular drug administration. The simplest function [Equation (20)] assumes a linear first-order absorption process ($k_a$) from an extravascular site following a short time-lag ($t_{lag}$). If a time-lag is not observed, $t_{lag}$ may be set to zero, resulting in a simple linear absorption model, as exemplified in a pharmacokinetic model for recombinant human growth hormone following SC dosing in monkeys [53]. As mentioned in the previous section, IV pharmacokinetic data must be analyzed simultaneously with SC data to estimate drug bioavailability ($F$).

Owing to the complexities of drug absorption from extravascular sites (see Section 2.1), peptide and protein drugs may require an additional delay compartment [Equation (21)] comprising two consecutive first-order processes [54, 55], which may be accompanied by an initial rapid input of a fraction of the dose as defined by Equations (22) and (23). Such a dual absorption model was used to describe the input of recombinant erythropoietin after SC administration in monkeys [56]. Interestingly, the bioavailability was shown to be dose-dependent and ranged from 0.268 to 1.0, increasing with increased dose level. This property may be caused by saturation of catabolism at the site of drug administration; however, further research is needed to elucidate the specific processes involved in drug disposition following extravascular dosing. Many other complex functions are available, such as an inverse Gaussian model [Equation (24)], which may be substituted into pharmacokinetic models to describe the absorption of macromolecules.

The clearance of several macromolecules may be characterized using simple linear first-order elimination as shown in Equation (18) [53, 57], but may depend on the animal species and range of dose levels being evaluated. In contrast, many peptide and protein drugs demonstrate saturable or capacity-limited clearance, and $CL$ in Equation (18) may be defined with a concentration-dependent function, such as:

$$CL = \frac{V_{\max} \cdot V_c}{K_m \cdot V_c + A_p} + CL_{ns} \tag{25}$$

where $V_{\max}$ and $K_m$ are the traditional Michaelis–Menten parameters and $CL_{ns}$ represents a linear nonsaturable clearance pathway. The nonlinear Michaelis–Menten function [quotient in Equation (25)] may be specified alone [56] or in parallel with $CL_{ns}$ [58, 59]. Thus, at high plasma drug concentrations (i.e., $C_p >> K_m$), the saturable clearance function approaches a limiting value ($V_{\max}$), and at relatively low concentrations will represent a first-order elimination rate ($V_{\max}/K_m$).

Receptor-mediated clearance is an example of target-mediated drug disposition (TMDD), where binding of a compound to a pharmacological target (e.g., enzyme, receptor, or other proteins) influences the pharmacokinetics of the drug [60]. Such systems often exhibit a dose-dependent decrease in the apparent volume of distribution (approaching a limiting value). The general pharmacological expectations of TMDD and techniques for characterizing this phenomenon have been reviewed recently [61]. A general pharmacokinetic model of TMDD has been described [62], and the operative equations defining the rate of change of plasma drug concentrations typically include:

$$\frac{dC_p}{dt} = -k_{on} \cdot R_f \cdot C_p + k_{off} \cdot RC \tag{26}$$

$$\frac{dRC}{dt} = k_{on} \cdot R_f \cdot C_p - (k_{off} + k_{\text{int}}) \cdot RC \tag{27}$$

where $k_{on}$ and $k_{off}$ are second- and first-order rates of association and dissociation, $R_f$ and $RC$ represent free and bound receptor (or pharmacological target) concentrations, and $k_{\text{int}}$ is a first-order internalization rate constant of $RC$. Under certain conditions, the total receptor concentration ($R_{tot}$) may be assumed to be time-invariant [11, 62], and thus $R_f = R_{tot} - RC$. Therefore, at high drug concentrations, $RC \rightarrow R_{tot}$ and effectively limits the distribution and elimination of drug from the central compartment. This approach has been used to characterize the complex nonlinear pharmacokinetics and pharmacodynamics of IFN-β 1a in monkeys [54]. Alternatively, an additional equation may be introduced that describes the rate of change of $R_f$ directly [55]. Also, an equilibrium solution to the general TMDD model is available, where the drug-binding microconstants ($k_{on}$ and $k_{off}$), which are often difficult to estimate from routine pharmacokinetic data, are replaced with an equilibrium dissociation constant ($K_D = k_{off}/k_{on}$) [63].

## 4 PREDICTING MACROMOLECULE PHARMACOKINETICS IN HUMANS

One of the major goals of preclinical pharmacokinetics is to obtain relevant information on the *in vivo* disposition of drugs early in the drug development process and to apply such knowledge to anticipate probable pharmacokinetic properties in humans for initial dose selection in phase I clinical trials. Allometric scaling and physiologically based pharmacokinetic (PBPK) models are the most popular techniques for predicting pharmacokinetic properties of drugs in one species from data collected in one or more other species. Although these methods have been applied to small-molecular-weight compounds for decades with variable and often limited success, it is hypothesized that these techniques are more accurate for peptide and protein drugs owing to the relative species conservation of mechanisms that control the biodistribution and elimination of such compounds [64, 65].

### 4.1 Allometric Scaling

Many physiological processes and organ sizes scale across species according to the well-known power-law relationship or allometric equation [66]:

$$Y = a \cdot W^b \tag{28}$$

where $Y$ is a physiological parameter of interest, $W$ represents body weight, and $a$ and $b$ are the allometric coefficient and exponent. The linearized form of Equation (28) is obtained by taking its logorithm:

$$\log Y = \log a + b \cdot \log W \tag{29}$$

where log $a$ and $b$ represent the intercept and slope of the physiological data presented on a log-log plot. The units of $a$ will depend on the units of $Y$ and $W$, and its absolute value will depend on the specific property of interest. In contrast, $b$ is a unitless parameter characterizing the rate of change of $Y$ with respect to $W$. According to the power-law relationship, $Y$ may increase more rapidly than $W$ ($b > 1$), less rapidly than $W$ ($0 < b < 1$), or be directly proportional ($b = 1$). For rates of metabolism and clearance processes, the allometric exponent tends to be around 0.75, whereas organ sizes or physiological volumes tend to be proportional to body weight [66]. Physiological times or the duration of physiological events (e.g., heartbeat and breath duration, lifespan, or turnover times of endogenous substances or processes) scale across species with $b$ values around 0.25 [66, 67]. These interspecies relationships appear to manifest from the fractal nature of biological systems [68, 69].

Retrospective analysis suggests that primary pharmacokinetic parameters of many small synthetic molecules scale across species according to allometric principles, primarily reflecting the physiological processes controlling such characteristics [67, 70]. However, the use of allometric scaling for prospective predictions of human pharmacokinetic parameters and dose selection of these compounds is controversial and is the topic of considerable debate [71, 72]. In addition, several correction factors and techniques have been suggested for improving the prediction of drug clearance in humans from preclinical values [73, 74]. Few studies have been conducted that specifically address the role of parameter uncertainty and species selection; however, those that have been reported suggest careful study designs and caution for prospective allometric scaling predictions [75, 76]. Research directed toward improving the understanding of interspecies pharmacokinetic relationships is likely to continue and necessary for resolving key methodological issues.

Notwithstanding concerns associated with prospective allometric scaling, primary pharmacokinetic parameters of select macromolecules appear to follow allometric relationships [65, 77–81], possibly resulting from a conservation of processes controlling peptide and protein disposition [64, 65]. Mahmood reviewed the prediction of total clearance of 15 therapeutic proteins from interspecies scaling, confirming the need for a suitable number of animals and that simple allometry may be used for most proteins [82]. As an example, Figure 4 shows log-log plots of total systemic clearance and the volume of distribution at steady state for interferon-α in mice, rats, rabbits, dogs, monkeys, and humans [78]. The lines of regression were generated without considering human values and the relationships were as follows: $CL = 3.7W^{0.71}$ ($r^2 = 0.98$) and $V_{ss} = 0.2W^{0.94}$ ($r^2 = 0.99$). Typical values in man (open symbols) were slightly underestimated using these equations, and Lave et al. suggest that transformations of plasma concentration-time profiles make more use of preclinical data and may provide improved predictions of human pharmacokinetic parameters [78]. Building on this concept, Cosson et al. described a method of fitting preclinical pharmacokinetic profiles and estimating allometric relationships simultaneously using a population mixed effects modeling technique [83]. Such an approach has been applied to characterize the interspecies pharmacokinetic relationships of pegylated human erythropoietin [84]. The final model was used to predict typical pharmacokinetic parameters in humans, as well as to simulate expected concentration-time profiles from IV and SC administration.

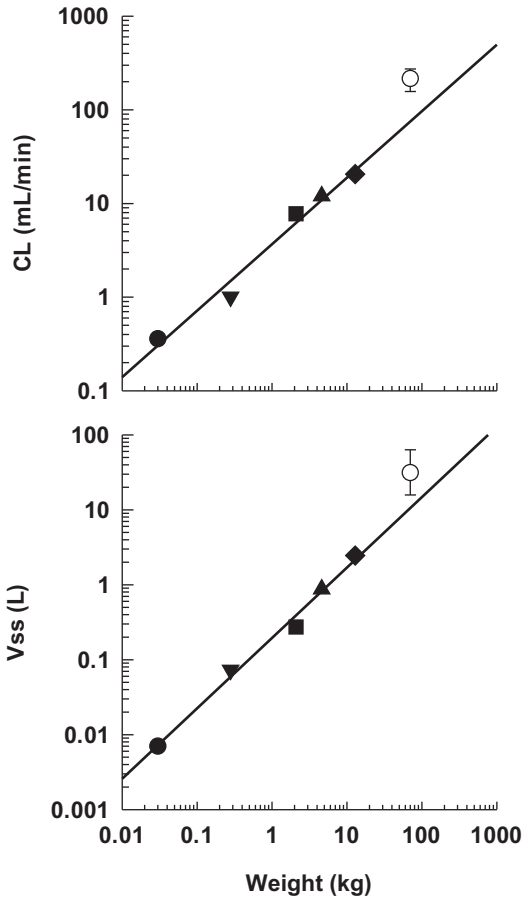


**Figure 4.** Interspecies scaling of total systemic clearance (*CL*; top panel) and steady-state volume of distribution ($V_{ss}$; bottom panel) of interferon-α as a function of total body weight. Species include mouse (●), rat (▼), rabbit (■), African green monkey (▲), dog (◆), and human (○). Human values were not included in the line of regression, and data were obtained from Lave et al. [78].

Despite successful applications of interspecies scaling to macromolecule pharmacokinetics, potential complicating factors exist, the most significant of which relate to saturable or dose-dependent pharmacokinetic processes. As discussed in Section 2, nonlinear behavior may manifest from several sources including receptor capacity and binding kinetics, macromolecule physico-chemical properties, and major mechanisms of peptide and protein transport and elimination. Thus, evidence of nonlinear pharmacokinetics in preclinical experiments would raise concerns as to the utility of allometric scaling for predicting such properties in humans. It remains to be determined whether a combination of *in vivo* preclinical data, compartmental modeling, and *in vitro* uptake studies might reliably anticipate macromolecule pharmacokinetics in humans under such nonlinear conditions [85].

### 4.2 Physiologically Based Pharmacokinetic Modeling

Classic pharmacokinetic modeling attempts to characterize drug disposition with systems comprised of a relatively small number of compartments (see Section 3.2). In contrast, PBPK models seek to mimic physiological pathways and processes controlling the time-course of plasma drug concentrations and represent the state-of-the-art in advanced pharmacokinetic systems analysis. As Dedrick has noted, "Physiologic modeling enables us to examine the joint effect of a number of complex inter-related processes and assess the relative significance of each" [86]. Basic concepts of PBPK modeling and applications to macromolecule pharmacokinetics are described in this section, and several reviews may be consulted for further details regarding the specification, implementation, and evaluation of PBPK models [87, 88].

The compartments in PBPK models represent organs and tissues of interest and are arranged and connected according to anatomical and physiological relationships (Fig. 5). Most PBPK models contain principal components, such as arterial and venous blood pools, liver, and kidney, along with additional tissues depending on its role in specific disposition processes (e.g., absorption, transport, and elimination), whether it represents a potential site of action (biophase) or is likely to account for a significant proportion of the administered dose [88]. A series of mass-balance equations are specified that define the time-course of drug concentrations within each compartment. In the simplest case, it is assumed that mass transfer is

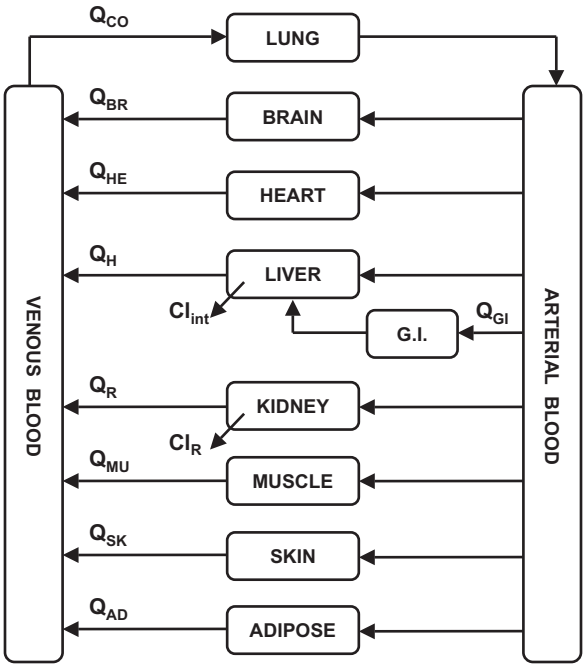


**Figure 5.** Model structure of a hypothetical physiologically based pharmacokinetic model. Tissue blood flows are shown (Q) along with renal ($Cl_R$) and nonrenal ($Cl_{int}$) drug clearance pathways.

blood-flow or perfusion-rate limited and compartments are well-stirred spaces. Under this assumption, the rate of change of drug concentration ($C$) in a noneliminating tissue ($T$) may be described by the following ordinary differential equation:

$$V_T \cdot dC_T / dt = Q_T \cdot (C_A - C_T / K_{pT}) \tag{30}$$

where $V$ and $Q$ represent the volume and blood flow associated with a particular tissue, $K_p$ is the tissue:plasma partition coefficient, and $C_A$ is arterial drug concentration. Drug elimination (via excretion or metabolism) in specific organs often is described using a first-order approximation or a Michaelis–Menten function, separately or in parallel [Equation (25)]. Physiological parameters (e.g., $V$ and $Q$) are frequently fixed to experimentally-measured values or literature-reported estimates [89, 90]; however, these terms have also been estimated during model fitting with a Maximum *a Posteriori* Bayesian method [91]. Several *in vitro* and *in vivo* experimental methods have been described for determining drug-specific parameters [92, 93]. Again, most contemporary applications of PBPK modeling involve estimating such terms from fitting the model to pharmacokinetic data.

Sugiyama and Hanano constructed a simple "minimal" PBPK model, consisting of plasma, liver, and kidney compartments, to evaluate the role of receptor-mediated clearance in the overall disposition of epidermal growth factor (EGF) in rats [11]. A remainder compartment was included to account for the rest of the body of the animal. The kinetics of the liver compartment incorporated receptor turnover processes (i.e., synthesis and degradation), binding of drug to free cell-surface receptors, and internalization of the drug-receptor complex (endocytosis and degradation). Computer simulations were conducted to evaluate this model, revealing that nonlinear kinetics of EGF may result from capacity limitation produced by either a saturation of free receptors or dose-dependent receptor downregulation.

The simple assumptions that constitute blood-flow-limited PBPK models often are inadequate for characterizing the pharmacokinetics of macromolecules. Instead, a membrane- or permeability-rate-limited model is more common, where it is assumed that mass transfer across the cell membrane is rate-limiting. For these models, organ compartments are subdivided into at least two well-stirred spaces representing vascular ($C_{T,V}$) and extravascular ($C_{T,EV}$) compartments. Such a system might be described by the following equations for a noneliminating tissue:

$$V_{T,V}\, dC_{T,V} / dt = Q_T \cdot (C_A - C_{T,V}) - f_{uB} \cdot PS_T \cdot (C_{T,V} - C_{T,EV} / K_{pT}) \tag{31}$$

$$V_{T,EV}\, dC_{T,EV} / dt = f_{uB} \cdot PS_T \cdot (C_{T,V} - C_{T,EV} / K_{pT}) \tag{32}$$

where $f_{uB}$ is the free fraction in blood and $PS_T$ is the permeability surface area product. Examples of membrane-limited models developed for protein molecules include β-endorphin, a 31 amino acid straight-chain polypeptide [94], and ISIS 1082, a 21-mer phosphorothioate oligonucleotide [95].

Monoclonal antibodies represent a diverse group of therapeutic proteins typically presenting with complex pharmacokinetic properties [10]. For these compounds, more comprehensive, mechanism-based PBPK models have been described [96, 97]. Baxter et al. developed and evaluated a bifunctional antibody PBPK model in mice and scaled the model to predict its pharmacokinetics in humans [96]. A membrane-

limited model was used; however, additional equations were required to describe up to nine molecular species in each tissue (resulting from specific and nonspecific binding), and mass transport across the membrane included convective and diffusive components (so-called two-pore system). The latter model, by Friedrich et al., further incorporated lymphatic circulation and identified critical properties requiring optimization for this treatment modality to be successful [97].

Although the PBPK approach has existed for several decades, reports of its use in characterizing macromolecule pharmacokinetics have been relatively sparse. The technological requirements for PBPK modeling are often seen as formidable; however, advances in computer hardware and software are bringing this methodology within common reach. The significant potential for scaling preclinical models to anticipate human pharmacokinetics in various relevant tissues (including target sites) represents a strategic advantage of PBPK modeling in drug development. Furthermore, this systems approach provides a means of understanding the pharmacological implications of the complexities associated with the disposition of macromolecules.

## 5 CONCLUSIONS

The pharmacokinetic properties of biotechnology pharmaceuticals are complex relative to small-molecule drugs, and preclinical studies are critical to the discovery and development of these therapeutic agents [98]. Understanding the mechanisms and implications of *in vivo* disposition of macromolecules provides a means for investigating and designing novel drug candidates and delivery systems to optimize exposure-response relationships or therapeutic efficacy. Mathematical modeling of preclinical pharmacokinetic data is necessary for characterizing such properties, and techniques like allometric scaling and PBPK modeling may be used to anticipate human pharmacokinetics under certain conditions. The utility of these approaches may be further enhanced by coupling pharmacokinetic data with mechanism-based pharmacodynamic modeling of biomarkers and surrogates of drug effects [8, 99, 100].

## ACKNOWLEDGMENT

Support from the National Science Foundation (Research Grant 0234895) and the National Institutes of Health (P20-GM 067650) is gratefully acknowledged.

## REFERENCES

1. Supersaxo A, Hein W, Steffen H (1990). Effect of molecular weight on the lymphatic absorption of water-soluble compounds following subcutaneous administration. *Pharm. Res.* 7:167–169.
2. Charman S A, McLennan D N, Edwards G A, et al. (2001). Lymphatic absorption is a significant contributor to the subcutaneous bioavailability of insulin in a sheep model. *Pharm. Res.* 18:1620–1626.

3. Charman S A, Segrave A M, Edwards G A, et al. (2000). Systemic availability and lymphatic transport of human growth hormone administered by subcutaneous injection. *J. Pharm. Sci.* 89:168–177.
4. McLennan D N, Porter C J, Edwards, et al. (2003). Pharmacokinetic model to describe the lymphatic absorption of r-metHu-leptin after subcutaneous injection to sheep. *Pharm Res.* 20:1156–1162.
5. McLennan D N, Porter C J, Edwards, et al. (2005). Lymphatic absorption is the primary contributor to the systemic availability of epoetin Alfa following subcutaneous administration to sheep. *J. Pharmacol. Exp. Ther.* 313:345–351.
6. Odegard P S, Capoccia K L (2005). Inhaled insulin: Exubera. *Ann. Pharmacother.* 39:843–853.
7. Patton J S, Bukar J G, Eldon M A (2004). Clinical pharmacokinetics and pharmacodynamics of inhaled insulin. *Clin. Pharmacokinet.* 43:781–801.
8. Braeckman R (1999). Pharmacokinetics and Pharmacodynamics of Protein Therapeutics. In: Rereid (ed.), *Peptide and Protein Drug Analysis*, University of British Columbia, Vancouver, B.C., pp. 633–639.
9. Ravetch J V, Kinet J P (1991). Fc receptors. *Annu. Rev. Immunol.* 9:457–492.
10. Lobo E D, Hansen R J, Balthasar J P (2004). Antibody pharmacokinetics and pharmacodynamics. *J. Pharm. Sci.* 93:2645–2668.
11. Sugiyama Y, Hanano M (1989). Receptor-mediated transport of peptide hormones and its importance in the overall hormone disposition in the body. *Pharm. Res.* 6:192–202.
12. Kurschner C, Ozmen L, Garotta G, et al. (1992). IFN-gamma receptor-Ig fusion proteins: Half-life, immunogenicity, and in vivo activity. *J. Immunol.* 149:4096–4100.
13. Teng M N, Turksen K, Jacobs C A, et al. (1993). Prevention of runting and cachexia by a chimeric TNF receptor-Fc protein. *Clin. Immunol. Immunopathol.* 69:215–222.
14. Sato T A, Widmer M B, Finkelman F D, et al. (1993). Recombinant soluble murine IL-4 receptor can inhibit or enhance IgE responses in vivo. *J. Immunol.* 150:2717–2723.
15. Finkelman F D, Madden K B, Morris S C, et al. (1993). Anti-cytokine antibodies as carrier proteins. Prolongation of in vivo effects of exogenous cytokines by injection of cytokine-anti-cytokine antibody complexes. *J. Immunol.* 151:1235–1244.
16. Schobitz B, Pezeshki G, Pohl T, et al. (1995). Soluble interleukin-6 (IL-6) receptor augments central effects of IL-6 in vivo. *Faseb. J.* 9:659–664.
17. Aderka D, Engelmann H, Maor Y, et al. (1992). Stabilization of the bioactivity of tumor necrosis factor by its soluble receptors. *J. Exp. Med.* 175:323–329.
18. Debets R, Savelkoul H F (1994). Cytokine antagonists and their potential therapeutic use. *Immunol. Today.* 15:455–458.
19. Klein B, Brailly H (1995). Cytokine-binding proteins: stimulating antagonists. *Immunol. Today.* 16:216–220.
20. Ramanathan M (1997). A physicochemical modelling approach for estimating the stability of soluble receptor-bound tumour necrosis factor-alpha. *Cytokine.* 9:19–26.
21. Zapf J, Hauri C, Waldvogel M, et al. (1986). Acute metabolic effects and half-lives of intravenously administered insulinlike growth factors I and II in normal and hypophysectomized rats. *J. Clin. Invest.* 77:1768–1775.
22. Baumann G, Shaw M A, Amburn K (1994). Circulating growth hormone binding proteins. *J. Endocrinol. Invest.* 17:67–81.
23. Baumann G, Vance M L, Shaw M A, et al. (1990). Plasma transport of human growth hormone in vivo. *J. Clin. Endocrinol. Metab.* 71:470–473.

24. Baumann G, Amburn K D, Buchanan T A (1987). The effect of circulating growth hormone-binding protein on metabolic clearance, distribution, and degradation of human growth hormone. *J. Clin. Endocrinol. Metab.* 64:657–660.
25. Baumann G, Shaw M A, Brumbaugh R C, et al. (1991). Short stature and decreased serum growth hormone-binding protein in the Mountain Ok people of Papua New Guinea. *J. Clin. Endocrinol. Metab.* 72:1346–1349.
26. Panitch H, Goodin D S, Francis G, et al. (2002). Randomized, comparative study of interferon beta-1a treatment regimens in MS: The EVIDENCE trial. *Neurology.* 59:1496–1506.
27. The IFNB Multiple Sclerosis Study Group and the University of British Columbia MS/MRI Analysis Group (1996). Neutralizing antibodies during treatment of multiple sclerosis with interferon beta-1b: experience during the first three years. *Neurology.* 47:889–894.
28. Rice G P, Paszner B, Oger J, et al. (1999). The evolution of neutralizing antibodies in multiple sclerosis patients treated with interferon beta-1b. *Neurology.* 52:1277–1279.
29. Pachner A R (2001). Measurement of antibodies to interferon beta in patients with multiple sclerosis. *Arch. Neurol.* 58:1299–1300.
30. Rice G (2001). The significance of neutralizing antibodies in patients with multiple sclerosis treated with interferon beta. *Arch. Neurol.* 58:1297–1298.
31. Brambell F W, Hemmings W A, Morris I G (1964). A theoretical model of Gamma-Globulin Catabolism. *Nature.* 203:1352–1354.
32. Ghetie V, Hubbard J G, Kim J K, et al. (1996). Abnormally short serum half-lives of IgG in beta 2-microglobulin-deficient mice. *Eur. J. Immunol.* 26:690–696.
33. Israel E J, Wilsker D F, Hayes K C, et al. (1996). Increased clearance of IgG in mice that lack beta 2-microglobulin: Possible protective role of FcRn. *Immunology.* 89:573–578.
34. Junghans R P, Anderson C L (1996). The protection receptor for IgG catabolism is the beta2-microglobulin-containing neonatal intestinal transport receptor. *Proc. Natl. Acad. Sci. USA* 93:5512–5516.
35. Harris J M, Chess R B (2003). Effect of pegylation on pharmaceuticals. *Nat. Rev. Drug Discov.* 2:214–221.
36. Monfardini C, Schiavon O, Caliceti P, et al. (1995). A branched monomethoxypoly(ethylene glycol) for protein modification. *Bioconjug. Chem.* 6:62–69.
37. Hart G W (1992). Glycosylation. *Curr. Opin. Cell Biol.* 4:1017–1023.
38. Jenkins N, Parekh R B, James D C (1996). Getting the glycosylation right: Implications for the biotechnology industry. *Nat. Biotechnol.* 14:975–981.
39. Jones J, Krag S S, Betenbaugh M J (2005). Controlling N-linked glycan site occupancy. *Biochim. Biophys. Acta.* 1726:121–137.
40. Pastan I, Willingham M C (1985). The pathway of endocytosis. In I Pastan, M C Willingham (eds.), *Endocytosis*: Plenum Press, New York, pp. 1–44.
41. Mufson R A, Gesner T G (1987). Binding and internalization of recombinant human erythropoietin in murine erythroid precursor cells. *Blood.* 69:1485–1490.
42. Chapel S, Veng-Pedersen P, Hohl R J, et al. (2001). Changes in erythropoietin pharmacokinetics following busulfan-induced bone marrow ablation in sheep: Evidence for bone marrow as a major erythropoietin elimination pathway. *J. Pharmacol. Exp. Ther.* 298:820–824.
43. Kuter D J, Rosenberg R D (1995). The reciprocal relationship of thrombopoietin (c-Mpl ligand) to changes in the platelet mass during busulfan-induced thrombocytopenia in the rabbit. *Blood.* 85:2720–2730.

44. Jusko W J (2005). Guidelines for collection and analysis of pharmacokinetic data. In M Leslie, J J Schentag, W E Evans, et al. (eds.), *Applied Pharmacokinetics & Pharmacodynamics: Principles of Therapeutic Drug Monitoring*: Applied Therapeutics, Inc., Vancouver, Canada, Chapter 2.
45. Yamaoka K, Nakagawa T, Uno T (1978). Statistical moments in pharmacokinetics. *J. Pharmacokinet. Biopharm.* 6:547–558.
46. Gibaldi M, Perrier D (1982). *Pharmacokinetics*, 2nd ed.: Marcel Dekker, Inc., New York.
47. Caprani O, Sveinsdottir E, Lassen N (1975). SHAM, A method for biexponential curve resolution using initial slope, height, area and moment of the experimental decay type curve. *J. Theor. Biol.* 52:299–315.
48. Tang L, Persky A M, Hochhaus G, et al. (2004). Pharmacokinetic aspects of biotechnology products. *J. Pharm. Sci.* 93:2184–2204.
49. Yu Z, Tse F L (1995). An evaluation of numerical integration algorithms for the estimation of the area under the curve (AUC) in pharmacokinetic studies. *Biopharm. Drug Dispos.* 16:37–58.
50. Bellissant E, Sebille V, Paintaud G (1998). Methodological issues in pharmacokinetic-pharmacodynamic modeling. *Clin. Pharmacokinet.* 35:151–166.
51. Foster D M (2001). Noncompartmental vs. compartmental approaches to pharmacokinetic analysis. In A J Atkinson, C E Daniels, R L Dedrick, et al. (eds.), *Principles of Clinical Pharmacology*: Academic Press, New York, pp. 75–92.
52. Gabrielsson J, Weiner D (2000). *Pharmacokinetic and Pharmacodynamic Data Analysis: Concepts and Applications*, 3rd ed.: Swedish Pharmaceutical Press, Stockholm, Sweden.
53. Sun Y N, Lee H J, Almon R R, et al. (1999). A pharmacokinetic/pharmacodynamic model for recombinant human growth hormone effects on induction of insulin-like growth factor I in monkeys. *J. Pharmacol. Exp. Ther.* 289:1523–1532.
54. Mager D E, Neuteboom B, Efthymiopoulos C, et al. (2003). Receptor-mediated pharmacokinetics and pharmacodynamics of interferon-beta1a in monkeys. *J. Pharmacol. Exp. Ther.* 306:262–270.
55. Segrave A M, Mager D E, Charman S A, et al. (2004). Pharmacokinetics of recombinant human leukemia inhibitory factor in sheep. *J. Pharmacol. Exp. Ther.* 309:1085–1092.
56. Ramakrishnan R, Cheung W K, Farrell F, et al. (2003). Pharmacokinetic and pharmacodynamic modeling of recombinant human erythropoietin after intravenous and subcutaneous dose administration in cynomolgus monkeys. *J. Pharmacol. Exp. Ther.* 306:324–331.
57. Benincosa L J, Chow F S, Tobia L P, et al. (2000). Pharmacokinetics and pharmacodynamics of a humanized monoclonal antibody to factor IX in cynomolgus monkeys. *J. Pharmacol. Exp. Ther.* 292:810–816.
58. Bauer R J, Dedrick R L, White M L, et al. (1999). Population pharmacokinetics and pharmacodynamics of the anti-CD11a antibody hu1124 in human subjects with psoriasis. *J. Pharmacokinet. Biopharm.* 27:397–420.
59. Veng-Pedersen P, Chapel S, Al-Huniti, et al. (2003). A differential pharmacokinetic analysis of the erythropoietin receptor population in newborn and adult sheep. *J. Pharmacol. Exp. Ther.* 306:532–537.
60. Levy G (1994). Pharmacologic target-mediated drug disposition. *Clin. Pharmacol. Ther.* 56:248–252.
61. Mager D E (2006). Target-mediated drug disposition and dynamics. *Biochem. Pharmacol.* 72:1–10.

62. Mager D E, Jusko W J (2001). General pharmacokinetic model for drugs exhibiting target-mediated drug disposition. *J. Pharmacokinet. Pharmacodyn.* 28:507–532.
63. Mager D E, Krzyzanski W (2005). Quasi-equilibrium pharmacokinetic model for drugs exhibiting target-mediated drug disposition. *Pharm. Res.* 22:1589–1596.
64. Ferraiolo B L, Mohler M A, Gloff C A (1992). *Protein Pharmacokinetics and Metabolism*. Plenum Press, New York.
65. Mordenti J, Chen S A, Moore J A, et al. (1991). Interspecies scaling of clearance and volume of distribution data for five therapeutic proteins. *Pharm. Res.* 8:1351–1359.
66. Adolph E F (1949). Quantitative relations in the physiological constitutions of mammals. *Science.* 109:579–585.
67. Boxenbaum H (1982). Interspecies scaling, allometry, physiological time, and the ground plan of pharmacokinetics. *J. Pharmacokinet. Biopharm.* 10:201–227.
68. West G B, Brown J H (2005). The origin of allometric scaling laws in biology from genomes to ecosystems: Towards a quantitative unifying theory of biological structure and organization. *J. Exp. Biol.* 208:1575–1592.
69. West G B, Brown J H, Enquist B J (1997). A general model for the origin of allometric scaling laws in biology. *Science.* 276:122–126.
70. Boxenbaum H, Ronfeld R (1983). Interspecies pharmacokinetic scaling and the Dedrick plots. *Am. J. Physiol.* 245:R768–R775.
71. Bonate P L, Howard D (2000). Prospective allometric scaling: Does the emperor have clothes? *J. Clin. Pharmacol.* 40:335–340.
72. Mahmood I (2000). Critique of prospective allometric scaling: does the emperor have clothes? *J. Clin. Pharmacol.* 40:341–346.
73. Tang H, Mayersohn M (2005). A mathematical description of the functionality of correction factors used in allometry for predicting human drug clearance. *Drug Metab. Dispos.* 33:1294–1296.
74. Tang H, Mayersohn M (2005). A novel model for prediction of human drug clearance by allometric scaling. *Drug Metab. Dispos.* 33:1297–1303.
75. Hu, T M, Hayton W L (2001). Allometric scaling of xenobiotic clearance: Uncertainty versus universality. *AAPS PharmSci.* 3:E29.
76. Tang H, Mayersohn M (2005). Accuracy of allometrically predicted pharmacokinetic parameters in humans: role of species selection. *Drug Metab. Dispos.* 33:1288–1293.
77. Mordenti J, Osaka G, Garcia K, et al. (1996). Pharmacokinetics and interspecies scaling of recombinant human factor VIII. *Toxicol. Appl. Pharmacol.* 136:75–78.
78. Lave T, Levet-Trafit B, Schmitt-Hoffmann A H, et al. (1995). Interspecies scal-ing of interferon disposition and comparison of allometric scaling with concentration-time transformations. *J. Pharm. Sci.* 84:1285–1290.
79. Bazin-Redureau M, Pepin S, Hong G, et al. (1998). Interspecies scaling of clearance and volume of distribution for horse antivenom F(ab')2. *Toxicol. Appl. Pharmacol.* 150:295–300.
80. Richter W F, Gallati H, Schiller C D (1999). Animal pharmacokinetics of the tumor necrosis factor receptor-immunoglobulin fusion protein lenercept and their extrapolation to humans. *Drug Metab. Dispos.* 27:21–25.
81. Lin Y S, Nguyen C, Mendoza J L, et al. (1999). Preclinical pharmacokinetics, interspecies scaling, and tissue distribution of a humanized monoclonal antibody against vascular endothelial growth factor. *J. Pharmacol. Exp. Ther.* 288:371–378.
82. Mahmood I (2004). Interspecies scaling of protein drugs: prediction of clearance from animals to humans. *J. Pharm. Sci.* 93:177–185.

83. Cosson V F, Fuseau E, Efthymiopoulos C, et al. (1997). Mixed effect modeling of sumatriptan pharmacokinetics during drug development. I: Interspecies allometric scaling. *J. Pharmacokinet. Biopharm.* 25:149–167.
84. Jolling K, Perez Ruixo J J, Hemeryck A, et al. (2005). Mixed-effects modelling of the interspecies pharmacokinetic scaling of pegylated human erythropoietin. *Eur. J. Pharm. Sci.* 24:465–475.
85. Proost J H, Beljaars L, Olinga P, et al. (2006). Prediction of the pharmacokinetics of succinylated human serum albumin in man from in vivo disposition data in animals and in vitro liver slice incubations. *Eur. J. Pharm. Sci.* 27:123–132.
86. Dedrick R L (1973). Animal scale-up. *J. Pharmacokinet. Biopharm.* 1:435–461.
87. Gerlowski L E, Jain R K (1983). Physiologically based pharmacokinetic modeling: Principles and applications. *J. Pharm. Sci.* 72:1103–1127.
88. Nestorov I (2003). Whole body pharmacokinetic models. *Clin. Pharmacokinet.* 42:883–908.
89. Davies B, Morris T (1993). Physiological parameters in laboratory animals and humans. *Pharm. Res.* 10:1093–1095.
90. Brown R P, Delp M D, Lindstedt S L, et al. (1997). Physiological parameter values for physiologically based pharmacokinetic models. *Toxicol. Ind. Health.* 13:407–484.
91. Xu L, Eiseman J L, Egorin M J, et al. (2003). Physiologically-based pharmacokinetics and molecular pharmacodynamics of 17-(allylamino)-17-demethoxygeldanamycin and its active metabolite in tumor-bearing mice. *J. Pharmacokinet. Pharmacodyn.* 30:185–219.
92. Lin J H, Sugiyama Y, Awazu S, et al. (1982). In vitro and in vivo evaluation of the tissue-to-blood partition coefficient for physiological pharmacokinetic models. *J. Pharmacokinet. Biopharm.* 10:637–647.
93. Houston J B (1994). Utility of in vitro drug metabolism data in predicting in vivo metabolic clearance. *Biochem. Pharmacol.* 47:1469–1479.
94. Sato H, Sugiyama Y, Sawada Y, et al. (1987). Physiologically based pharmacokinetics of radioiodinated human beta-endorphin in rats: An application of the capillary membrane-limited model. *Drug Metab. Dispos.* 15:540–550.
95. Peng B, Andrews J, Nestorov I, et al. (2001). Tissue distribution and physiologically based pharmacokinetics of antisense phosphorothioate oligonucleotide ISIS 1082 in rat. *Antisense Nucleic Acid Drug Dev.* 11:15–27.
96. Baxter L T, Zhu H, Mackensen D G, et al. (1995). Biodistribution of monoclonal antibodies: Scale-up from mouse to human using a physiologically based pharmacokinetic model. *Cancer Res.* 55:4611–4622.
97. Friedrich S W, Lin S C, Stoll B R, et al. (2002). Antibody-directed effector cell therapy of tumors: Analysis and optimization using a physiologically based pharmacokinetic model. *Neoplasia.* 4:449–463.
98. Galluppi G R, Rogge M C, Roskos L K, et al. (2001). Integration of pharmacokinetic and pharmacodynamic studies in the discovery, development, and review of protein therapeutic agents: A conference report. *Clin. Pharmacol. Ther.* 69:387–399.
99. Toutain P L (2002). Pharmacokinetic/pharmacodynamic integration in drug development and dosage-regimen optimization for veterinary medicine. *AAPS PharmSci.* 4: E38.
100. Mager D E, Wyska E, Jusko W J (2003). Diversity of mechanism-based pharmacodynamic models. *Drug Metab. Dispos.* 31:510–518.

# STRATEGIES FOR THE CYTOSOLIC DELIVERY OF MACROMOLECULES: AN OVERVIEW

Aravind Asokan, Roland Cheung, and Moo J. Cho
*University of North Carolina at Chapel Hill, Chapel Hill, North Carolina*

**Chapter contents**

1 Introduction
2 Endosome-to-Cytosol Transfer
  2.1 pH-Sensitive Drug Carriers
  2.2 Acid-Labile Linkers and Drug Carriers
  2.3 Redox/Enzyme-Triggered Systems
  2.4 Caveolae-Mediated Cytosolic Delivery
  2.5 Cell-Penetrating Peptides: HIV-TAT
3 Direct Cytosolic Delivery Across the Plasma Membrane
  3.1 Cell-Penetrating Peptides: Penetratin and Other Peptides
  3.2 Molecular Umbrellas
4 Assays for Monitoring Cytosolic Delivery
  References

## 1 INTRODUCTION

The advent of molecular biology and genetics has identified numerous biotechnology-derived macromolecular drugs as potential targets for development. The factors required for the development of this new class of drugs are, however, significantly different from those required for conventional drug compounds with molecular weight less than 1 kDa or so [1, 2]. Efforts directed toward the delivery of small

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

molecules have largely sought a compromise between two major factors, i.e., aqueous solubility and membrane partition properties. On the other hand, not only does the primary structure of proteins and genes require protection from enzymatic degradation but also higher order structures from denaturation [3]. In addition, for many macromolecules with therapeutic potential, be it peptide or protein-based drugs or nucleic acid-based drugs such as antisense oligonucleotides, ribozymes, siRNA, aptamers, and plasmid DNA, delivery to the cytosol becomes a necessary, if not sufficient, condition for their full pharmacological activity. For example, antisense oligomers or siRNA, which are extensively sequestered in endosomes after cellular uptake, require cytosolic delivery to bind ribosomal mRNA/nuclear pre-mRNA in order to inhibit protein translation or correct splicing [4–6]. As a result, several strategies to promote the cytosolic delivery of macromolecular therapeutics have been developed over the past few years. We have classified these diverse strategies into (1) those that exploit the biochemical aspects of the endosomal lumen to promote endosome-to-cytosol transfer and (2) approaches that bypass the endocytic pathway by facilitating direct transport of macromolecules across the plasma membrane.

## 2 ENDOSOME-TO-CYTOSOL TRANSFER

The plasma membrane is a mosaic network of lipids and proteins decorated with complex carbohydrates. This 5–10-nm-thick lipid bilayer is burdened with the dual role of serving as a barrier to external agents, while regulating the entry of nutrients into the cell. Conventional small molecules, such as amino acids, sugars, and ions, generally traverse the plasma membrane through a con-certed function of integral membrane protein carriers or channels. However, macromolecules, by virtue of their polarity and large hydrodynamic diameter, are taken up in membrane-bound vesicles derived by the invagination and pinching-off of sections of the plasma membrane in a process termed endocytosis. Endocytosis can occur via multiple mechanisms that can be broadly classified into (1) phagocytosis or the uptake of large particles and (2) pinocytosis or the uptake of fluid and solutes. Phagocytosis is typically restricted to specialized mammalian cells, such as macrophages and other antigen presenting cells, whereas pinocytosis occurs in all cells by at least four known mechanisms: macropinocytosis, clathrin-mediated endocytosis, caveolae-mediated endocytosis, and clathrin- and caveolae-independent endocytosis [7].

Internalization of macromolecules and a variety of pathogens via caveolae is another endocytic mechanism alternative to the clathrin-mediated process. Originally discovered in the 1950s, caveolae are uncoated, 50–100-nm flask-shaped invaginations of the plasma membrane. The process shares several similarities with clathrin-mediated typical endocytosis in that docking and fusion proteins involved in the latter are also found in caveolae. However, an important difference, especially from the perspective of delivery of intact therapeutic macromolecules, is the finding that the resulting "caveosomes" can avoid the degradative route to lysosomes [8]. Mechanistic aspects of such diverse and highly regulated endocytic pathways are beyond the scope of this review and discussed elsewhere. Nevertheless, three biochemical traits of endocytic vesicles or endosomes are particularly relevant in the context of macromolecular drug delivery.

First and foremost, it is well documented that the endosomal compartments maintain an acidic internal pH ranging from pH 5.5 to 6.5 and seem to mature or fuse with lysosomes, which maintain a lower pH (~5.0 to 5.5). This is vital for sorting of ligand-receptor complexes and the activation of hydrolytic enzymes such as cathepsins. Although the $H^+$-ATPase pumps protons into the endosome, the $Na^+/K^+$-ATPase acts to limit proton pumping into the endosomal compartment. In its normal orientation, the pump generates a positive interior membrane potential by pumping three $Na^+$ ions into the endosome for every two $K^+$ ions pumped out, thereby regulating the proton-translocating activity of the membrane bound $H^+$-ATPase. Evidence supporting these events has been provided by the treatment of isolated endosomes with inhibitors of $Na^+/K^+$-ATPase such as ouabain, which causes a fall in the pH within the endosomal compartments. Lysosomes, on the other hand, seem to lack $Na^+/K^+$-ATPase, and their luminal acidification is mainly regulated by $H^+$-ATPases.

Second, of several proteases that are known to participate in peptide/protein hydrolysis, the cysteine and aspartic proteases, also known as cathepsins, constitute a majority of the endosomal/lysosomal proteolytic machinery. These enzymes are often synthesized in an inactive form requiring excision of the pro-domain facilitated by other proteases or autocatalytic mechanisms activated by acidic pH. Cathepsins seem to have originally evolved to catabolize both internalized and endogenous proteins crucial for cellular homeostasis, autophagy, apoptosis, and antigen presentation. The optimal pH range for enzymatic activity of cathepsins ranges from 5.0 to 6.5, as is the case with endosomes/lysosomes.

Last, several redox pathways are thought to function within endosomes and lysosomes, although the exact mechanism(s) of disulfide reduction within the endosomal lumen are not well understood. Primary evidence stems from antigen processing within the endocytic pathway, which involves partial denaturation of internalized proteins facilitated in part by the acidic pH. Disulfide bonds, however, are not susceptible to lysosomal proteolysis and remain chemically stable at acidic pH. Disulfide reduction within endosomes/lysosomes has been attributed to active transport of cysteine (free thiol form) into lysosomes and redox enzymes such as γ-Interferon-inducible lysosomal thiol reductase (GILT) that function at an acidic pH optimum [9]. However, a recent report disputes the long-held view that endosomes and lysosomes are reductive environment. Rather, the experimental data support that they may exhibit an oxidizing environment suggesting that disulfide-based systems may function via a nonreductive mechanism or that they may be reduced in the cytosol after discharged from the endosome [10]. Overall, strategies for the endosome-to-cytosol transfer of macromolecular drugs can broadly be classified on the basis of the three aforementioned characteristics of the endosomal lumen, namely, acidic pH, proteolytic activity of cathepsins, and disulfide bond reduction.

### 2.1 pH-Sensitive Drug Carriers

Viruses have evolved proteins that perturb the host cell membrane by fusion. Stemming from this knowledge is the creation and use of pH-dependent fusogenic peptides. For example, diINF-7 is one such peptide that resembles the amino-terminal domain of the influenza virus hemagglutinin HA-2 subunit. At an endosomal pH of 5.0, the peptide undergoes a conformational change allowing it to penetrate

deeper into the endosomal membrane resulting in membrane destabilization [11]. The pH-sensitive behavior of viral peptides and bacterial proteins originates from that of weak acids with pKa values in the range of 4.5 to 7.0, primarily Glu residues with a pKa of approximately 5.0. Thus, the pH-dependent membrane partitioning or the partition coefficient of weak acids (with elevated membrane-bound pKa values) depends critically on the endosomal pH. Although the negatively charged side chains of the Glu residues in viral peptides such as INF-7, initially repel each other, they collapse into a hydrophobic α-helix due to protonation of the side-chain carboxylate ions at a pH close to their pKa. Subsequently, the helical peptide inserts itself into the endosomal membrane and leads to fusion or pore formation and eventually endosomal release. Based on these principles, Li et al. designed the GALA peptide, which consists of repeats of the peptide motif Glu-Ala-Leu-Ala and adapts a α-helical conformation at acidic pH [12]. The peptide was subsequently found to efficiently release calcein from phosphatidylcholine liposomes in a pH-specific manner, presumably via formation of pores or perturbation-mediated flip-flop of lipids.

Weak acid-based homopolymers such as poly(Glu) and poly(alkylacrylic acid) have been shown to disrupt membranes in a pH-dependent manner [13, 14]. Poly(ethylacrylic acid) is thought to form cation-selective aqueous pores in lipid bilayers at relatively low concentrations, providing a potential application in promoting endosome-to-cytosol transfer of macromolecules. Mechanistically, the carboxylate head groups become neutral at a pH < pKa triggering polymer contact with endosomal membranes, resulting in membrane permeabilization. Polycationic molecules have been extensively used to promote the cellular delivery of macromolecules, due to their potent membrane disruptive, destabilizing, or fusogenic activity. A proton gradient such as the one found in endosomes and lysosomes can be exploited to protonate weakly basic drug carriers with pKa values between 5.0 and 7.0. The cationic charge thus generated can induce dramatic changes in the physico-chemical properties of the carrier.

Poly(His) was perhaps the first pH-sensitive polymer shown to promote fusion between lipid bilayers [15]. The imidazole functionality in the side chain of the His residue is protonated at low pH, and it interacts with negatively charged head groups of lipids. Although proximity-driven lipid and content mixing is observed at lower concentrations, a higher concentration of the polymer causes disruption of the bilayer and leakage of contents. Imidazole-containing polymers and their endosomolytic properties have received much attention lately, as exemplified by a pH-sensitive His-containing peptide analogous to the influenza HA-2. Histidylated polylysine and gluconic acid modified poly(His) have also been shown to promote efficient gene transfer [16]. Lee et al. constructed a core-shell-type micelle from the two block copolymer components poly(L-histidine)-*b*-poly(ethylene glycol) and poly(L-lactic acid)-*b*-PEG-*b*-polyHis-biotin and showed that after biotin receptor-mediated endocytosis, the micelle exhibited pH-dependent dissociation, thus allowing for the release of the model drug doxorubicin to be released in the early endosomal pH [17]. Furthermore, the collapsed micellar components could subsequently destabilize the endosomal membrane allowing for doxorubicin to enter the cytosol.

Polyethyleneimine (PEI) and starburst dendrimers are other examples of pH-sensitive cationic polymers, which become protonated incrementally over a wide

pH range [18, 19]. The application of these pH-sensitive cationic polymers to oligonucleotide and gene delivery has been studied extensively. Other examples of the use of pH-sensitive materials are hydrogels made of chitosan and copolymers of chitosan [20]. These represent a class of polymeric delivery systems used for controlled drug release. Hydrodynamic swelling or shrinking of these gel-like polymers at low pH (due to protonation of the polymer side-chain moieties) is thought to release the entrapped contents. The proton sponge hypothesis suggests a mechanism by which the polymers such as PEI containing protonatable moieties lead to chloride influx from the cytosol and subsequent osmosis into the endosome, resulting in endosomal swelling and lysis. Endosomal lysis allows the polymer and its cargo to be released into the cytosol [21].

Acid-sensitive liposomes have been extensively used to deliver macromolecular therapeutics agents. The first example of pH-sensitive liposomes was outlined by Yatvin et al. [22]. The mechanism of acid-induced fusion between liposomes composed of DOPE and palmitoylhomocysteine was further established by Connor et al. [23]. As outlined in the review by Simões et al. [24], three hypothetical mechanisms of cytosolic delivery from liposomes have been proposed: (1) Destabilization of pH-sensitive liposomes destabilizes the endosomal membrane possibly through pore formation, (2) destabilization of pH-sensitive liposomes allows cargo to diffuse through the endosomal membrane, and (3) liposomal membrane fusion with the endosomal membrane (illustrated in Figure 1). Several pH-sensitive lipids have been designed to promote endosome-to-cytosol transfer of macromolecules using liposomal systems. Fusogenic lipids such as oleic acid and cholesteryl hemisuccinate contain weak acids as head groups. These lipids, when incorporated into liposomes, induce pH-dependent fusion. Wang and Huang demonstrated the utility of oleic acid-containing pH-sensitive liposomes in mediating targeted delivery of transgenes to lymphoma tumors *in vivo* [25]. Other examples for weak acid-based pH-sensitive lipids include the titratable, double-chain amphiphiles 1,2-dipalmitoyl-sn-3-succinylglycerol (1,2-DPSG), 1,2-dioleoyl-sn-3-succinylglycerol (1,2-DOSG), and 1,3-dipalmitoylsuccinylglycerol (1,3-DPSG) developed by Collins et al. [26]. The fusogenicity of these lipids in the liposomal formulation is greatly enhanced by the incorporation of dioleoylphosphatidyl ethanolamine (DOPE), a neutral fusogenic lipid.

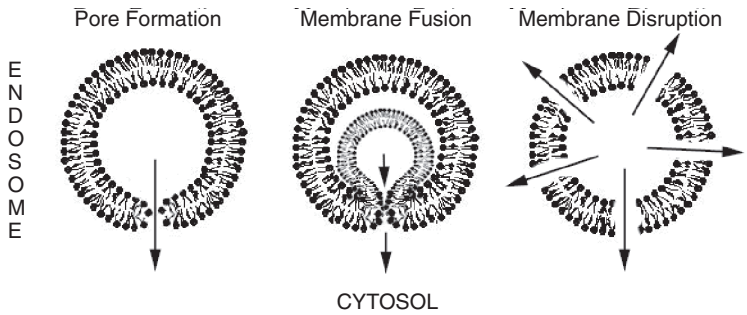


**Figure 1.** Mechanisms of endosome-to-cytosol transfer facilitated by various drug carriers.

The neutral dioleoyldimethylaminopropane (DODAP) free base tends to be fusogenic; however, protonation of the tertiary amino head group results in a bilayer-forming cationic lipid. It has been shown that interaction of the cationic lipid head group with anionic membrane-associated moieties such as proteoglycans results in a charge neutralization, which in turn, triggers the lamellar-to-hexagonal phase transition of such cationic lipid-based carriers. This nonbilayer phase transition is thought to promote extensive fusion between the liposomal formulation and the cellular membrane. Another example of weak base carriers is a series of amphiphilic imidazole-containing lipids. These lipids have been shown to efficiently promote the delivery of plasmid DNA to the cytoplasm [27]. As observed in the case of weak acid lipids, the fusogenic activity of the weak base lipids is potentiated by the presence of DOPE, a "helper" lipid that prefers the hexagonal phase above a temperature of 10°C.

Chen et al. synthesized a series of acyloxyalkyl imidazole lipids, with membrane-bound pKa values in the range of 5.0 to 5.5 [28]. It was expected that these molecules, once protonated in the acidic environment of the endosome, would become surface active and permeabilize the endosomal membrane. Eventually these lipids are degraded by cytosolic esterases into imidazole, formaldehyde, and the corresponding fatty acid. However, a serious limitation that precludes utilization of single-tailed, pH-sensitive detergents for the cytosolic delivery of macromolecules is their low limit of incorporation in stable liposomal formulations. Limited incorporation of detergent molecules resulted in a lack of liposomal collapse that leads to minimal endosomal membrane destabilization [29]. However, over-incorporation of such detergents destabilizes the liposomal membrane and would result in premature leakage of entrapped cargo. To address this issue, two Gemini surfactants or "bis-detergents" have been prepared by cross-linking the headgroups of single-tailed, tertiary amine detergents through oxyethylene (BD1) or acid-labile acetal (BD2) moieties [30]. The ability of BD2-containing liposomes to promote efficient cytosolic delivery of antisense oligonucleotides was confirmed by (1) their diffuse intracellular distribution observed in fluorescence micrographs, and (2) the upregulation of luciferase in an antisense functional assay. The low pH-responsive, bis-detergent constructs described herein are suitable for triggered release strategies targeted to acidic intracellular or interstitial environments.

## 2.2 Acid-Labile Linkers and Drug Carriers

Acid-labile linkers that are cleaved off in the endosomal milieu have been used in conjugating chemotherapeutic agents such as doxorubicin to targeting moieties such as folic acid, antibodies, and antibody fragments. Some common examples are acid-labile ketals, hydrazones, diorthoesters, and the *cis*-dicarboxylic acid linkers, which undergo intramolecular catalysis at low pH. Additionally, such moieties can also be conjugated to carrier molecules such as lipids to create an inactive pro-carrier. Hydrolysis of the pro-moiety at a low pH activates the carrier, which disrupts, destabilizes, or fuses with the endosomal membrane and subsequently promotes the release of entrapped molecules.

Recently, Murthy et al. have synthesized an acid-sensitive microgel material for the development of protein-based vaccines. The chemical design of these microgels is such that they degrade under the mildly acidic conditions found in the phago-

somes of antigen-presenting cells (APCs). The rapid cleavage of the microgels leads to phagosomal disruption through a colloid osmotic mechanism, releasing protein antigens into the APC cytoplasm for class I antigen presentation [31, 32]. Other examples of such acid-labile hydrogels include poly(orthoester) microspheres [33] and poly(ketal) nanoparticles [34] that undergo acid-catalyzed hydrolysis into low-molecular-weight hydrophilic compounds and release encapsulated therapeutics at an accelerated rate in acidic environments.

With respect to lipid-based drug carriers, an attractive approach for triggered release is to design cleavable surfactants with uncharged functional groups whose hydrolysis is catalyzed by acidic conditions. The chemical principles behind this approach have been described in an outstanding review published almost three decades ago [35]. For example, Boomer and Thompson reported several mono- and diplasmenyl lipids with an acid-sensitive vinyl ether linkage between the headgroup and one or both of the hydrocarbon side chains [36]. Upon exposure to low pH or photo-oxidation, the vinyl ether side chains are cleaved from the lipids, leading to structure defects in the bilayer and the release of liposomal contents. In another example, Guo and Szoka used a diorthoester bridge to link PEG to distearoylglycerol. This lipid efficiently stabilizes DOPE bilayer vesicles at a total lipid concentration as low as 10%. Release of the PEG by cleavage of the acid-labile diorthoester linker disrupts the lamellar organization and promotes the fusogenic lipid configuration of DOPE within the endosome. Subsequent fusion with the endosomal membrane resulted in cytoplasmic delivery of liposomal contents. These strategies have been reviewed extensively by Guo and Szoka [37, 38].

### 2.3 Redox/Enzyme-Triggered Drug Carriers

In addition to lower pH, the endosome also exhibits other characteristics that can be used in cytosolic delivery. The endosome houses many enzymes, including redox enzymes [9], and may exhibit a slightly reductive environment, although the latter portion of this statement currently remains a subject of controversy [10]. Using a listeriolysin O-protamine conjugate, the reducing potential of the endosome was exploited for plasmid DNA delivery into the cytosol [39]. The conjugate contained a redox-labile disulfide bond between the cysteine of listeriolysin O and the polycationic peptide protamine, which could be reduced in the reductive environment of the endosome. Upon disulfide cleavage, listeriolysin O exerted its endosomolytic activity, resulting in the cytosolic release of condensed DNA polyplexes.

Liposomes prepared from DOPE and a disulfide-containing cationic lipid showed a greater release of plasmid DNA than a non-disulfide-containing analog [40]. The disulfide bond of the lipid, 1-2-dioleoyl-sn-glycero-3-succinyl-2-hydroxyethyl disulfide (DOGSDSO), was shown to be cleaved in the reductive environment leading to the destabilization of the liposome-DNA complex (lipoplex) and increased transfection activity. A novel set of cationic lipids was designed to confer reduction-sensitivity for lipopolyamines [41]. These lipids contain a disulfide bond at different sites on the lipid backbone. Compared with nondegradable lipopolyamine analogs, increased transfection efficiency is observed when inserting the disulfide between one lipid chain and the rest of the molecule. However, loss or a substantial decrease in transfection activity occurs when the disulfide bridge is placed in the linker position between the polyamine and the lipid chains.

Chittimalla et al. have designed cationic thiol-based detergents [42]. Subsequent detergent polymerization by formation of intermolecular disulfide bonds within the condensed plasmid DNA leads to 32-nm-large neutral particles that promote efficient gene delivery.

Enzymatic cleavage of a synthetic peptide-lipid conjugate has been shown to trigger liposomal fusion to cellular membranes, thus allowing liposomal cargo to be distributed to the cytosol [43, 44]. The conjugate, formed from the covalent linkage of dioleoyl phosphatidylethanolamine (DOPE) with an elastase substrate N-acetyl-ala-ala-, having no intrinsic fusogenic activity, required cleavage by elastase or proteinase K and subsequent conversion to the fusogenic lipid DOPE. Only a few examples of enzymatic triggering have been described here although many enzymatic triggers have been studied. A few of these have explored the use of alkaline phosphatase [45], matrix metalloproteinases [46], and elastases [47].

### 2.4 Caveolae-Mediated Cytosolic Delivery

A variety of lipid-anchored membrane proteins are associated with caveolae that exist as a subdomain in a biochemically defined glycolipid raft. Not surprisingly, the intracellular trafficking of a given ligand is largely determined by what protein in caveolae they bind. It ranges from ultimate transcytosis as often found in endothelial cells, direct discharge to cytosol of folates, or endoplasmic reticulum and Golgi as in the case of simian virus 40 and cholera toxin, respectively [48–50]. The glycosylphosphatidylinositol (GPI)-anchored folate receptor-α (FR-α) internalizes and subsequently discharges 5-methyltetrahydrofolate and other folates directly to cytoplasm upon dissociation with FR-α in response to an acidic environment [48]. Although clustering of GPI-anchored proteins such as FR-α in lipid raft remains controversial [51], the current literature is rich in its application to cellular delivery of macromolecules and submicron drug carriers [52, 53]. Another unresolved issue is whether chemical modification of α-Glu of folate affects recognition of the conjugate by FR-α. A recent study reported no difference in liposomal delivery of antisense oligonucleotides when the surface was grafted with α- or γ-modified of folic acid [54].

### 2.5 Cell-Penetrating Peptides: HIV-TAT

The HIV TAT peptide is a short polybasic peptide sequence derived from the HIV-1 TAT protein, which plays a critical role in viral replication. The 11 amino acid TAT peptide is capable of penetrating cell membranes without causing membrane disruption. The TAT peptide was originally thought to enter cells directly at the plasma membrane via a nonsaturable, nonenergy-dependent mechanism [55, 56], possibly via transient formation of non-bilayer structures during the membrane penetration [57]. Recently, however, this picture has changed dramatically to a view that CPPs enter cells via saturable, energy-dependent processes involving endocytosis. Some early studies indicating nonendocytic uptake are now believed to reflect fixation artifacts [58]. Furthermore, cell-surface receptors for TAT protein chimeras were identified as being heparan sulfate proteoglycans [59, 60]. Additional studies showed that TAT and several other cationic CPPs clearly entered the cell by endocytosis, possibly via lipid raft [61, 62], and their release into the cytoplasm involves endosome acidification [58, 63–65].

## 3 DIRECT CYTOSOLIC DELIVERY ACROSS THE PLASMA MEMBRANE

Until recently, strategies for cytosolic delivery of macromolecules have largely been formulated by mimicking the cellular entry pathways of viruses and bacteria. Among these strategies, most have focused on promoting endosomal escape of macromolecules into the cytosol exploiting features such as acidic pH and redox/enzyme activity within endocytic vesicles. Nevertheless, simpler methods to deliver large macromolecules directly across the plasma membrane, while bypassing the degradative endocytic pathway, have been pursued. In particular, the seminal discoveries that peptides derived from glycoproteins embedded in the lipid envelope of viruses such as HIV and Herpes virus as well as those derived from the homeodomain of certain transcription factors can directly translocate across the plasma membrane of mammalian cells have opened up new avenues in macromolecular drug delivery. In general, these "cell-penetrating peptides (CPPs)" are approximately 30-amino-acid-long amphipathic peptides with a net positive charge [66]. The most prominent examples of CPPs are the HIV-TAT, discussed in the previous section, penetratins, and synthetic peptides derived thereof. Less-studied examples such as the peptides derived from HIV gp41 and the Herpes VP22 are also discussed (Table 1).

**TABLE 1. Examples of Peptide Sequences of Common Cell-Penetrating Peptides**

| | |
|---|---|
| HIV-TAT | $H_2N$-**GRKKRRQRRRPPQ**-$CONH_2$ |
| PENETRATIN | $H_2N$-**RQIKIWFQNRRMKWKK**-$CONH_2$ |
| TRANSPORTAN | $H_2N$-**GWTLNSAGYLLGKINLKALAALAKISIL**-$CONH_2$ |
| SN50 | $H_2N$-**AAVALLPAVLLALLAP**-$CONH_2$ |

The main application of CPPs is the attachment of biologically active cargo for translocation into cells. Conjugating cargo to CPP can be achieved through several means. When the cargo is a peptide or protein, CPP and cargo are most often synthesized or expressed in tandem as a fusion protein. Alternatively, a suitable amino acid side-chain or bifunctional spacer molecule can be used to couple cargo to CPPs through maleimide-thiol, disulfide, or amine-carboxylate chemistry. Attachment of cargo to CPP can also be achieved via noncovalent bonds, employing for instance the interaction between streptavidin and a biotinylated CPP conjugate. When chemically conjugated, TAT has been shown to enhance intracellular delivery of oligonucleotides [67–71], peptides and proteins [72, 73], and particulates [74, 75]. The strategy has met with moderate success *in vivo* [76–79]. Several strategies involving target proteins fused to the TAT peptide have also been described [80, 81]. The technology requires the synthesis of a fusion protein, linking the TAT transduction domain to the molecule of interest using a bacterial expression vector, followed by the purification of this fusion protein under either soluble or denaturing conditions. The original 11-residue TAT sequence (47–57) has been extensively modified to improved versions of TAT [82], poly-arginine peptides [83, 84], and guanidinium-rich peptoids [84].

### 3.1 Cell-Penetrating Peptides: Penetratin and Other Peptides

Penetratins are a class of CPPs derived from transcription factor proteins of Antennapedia (a *Drosophila* homeoprotein). Using site-directed mutagenesis, it was discovered that the third helix (amino acids 43–58) of the latter homeodomain was necessary and sufficient for translocation. Structure–function studies of penetratin yielded several active and inactive analogs that allowed formulation of the inverted micelle internalization model. The cellular internalization of penetratins occurs at 37°C and 4°C through a nonsaturable mechanism. The demonstration that retro-inverso penetratin analogs and similar peptides derived from other homeodomains are capable of membrane translocation suggests that the number of member peptides comprising the penetratin family is high.

Transportans are 27-amino-acid-long chimeric peptides designed and synthesized from other unrelated peptides. The first 13 amino acids are derived from the highly conserved amino-terminal part of galanin, whereas the 14-residue C-terminus is derived from the wasp venom peptide toxin, mastoparan. Although the N-terminus portion is the smallest highly active galanin receptor ligand with agonist properties, the C-terminus end is thought to penetrate the cell membrane by creating transient pores by translocating into the inner leaflet of the membrane bilayer. Applications of transportans have been similar to that of HIV-TAT and penetratins and have been discussed in detail elsewhere. Morris et al. designed a synthetic CPP that contains a hydrophobic domain derived from the HIV fusion protein, gp41, and a polycationic sequence derived from the nuclear localization sequence of the SV40 T-antigen [85]. The peptide vector exhibits a high affinity for single- and double-stranded DNA in the nanomolar range and has been used for the nuclear delivery of oligonucleotides, siRNA, and plasmid DNA. Although unclear, the peptide vector seems to facilitate direct cytosolic and nuclear delivery of cargo through nonendocytic means.

Elliott and O'Hare demonstrated that the Herpes simplex virus protein VP22 exhibits the remarkable property of intercellular trafficking, whereby the protein spreads from the cell in which it is synthesized to many surrounding cells [86]. This function of VP22 has been exploited to construct fusion proteins similar to aforementioned strategies that retain their ability to spread between cells and accumulate in recipient cell nuclei. The extreme C-terminus (residues 267–301) of VP22 is thought to be required for membrane translocation. However, further studies by Aints et al. indicate that although this region facilitates VP22 transport, a core region in the protein (residues 81–195) also contributes to transport activity [87]. VP22 has effectively delivered proteins such as GFP, p53, and the SV40 T antigen in both actively dividing and terminally differentiated cell lines. It has been successfully employed to enhance the effectiveness of thymidine kinase and cytosine deamidase suicide gene therapy systems for cancer therapy both *in vitro* and *in vivo*. VP22 fusion proteins have also been successful in enhancing the efficacy of lentivirus and adenovirus-based gene therapy systems. Additionally, when combined with small oligonucleotides, VP22 forms spherical particles, named Vectosomes, that are taken up by an energy-dependent process when applied to cells in culture.

Hawiger has developed CPPs derived from the membrane-translocating hydrophobic sequence of the h-region of the signal peptide sequence of K-FGF [88].

These alanine and leucine-rich peptides adopt a predominantly alpha-helical conformation in solution. A particularly striking characteristic of these CPPs (SN50) is the presence of a kink in the center of the alpha-helix due to a single proline, which is thought to play a key role in the membrane translocating activity of these peptides. Several studies testing the ability of these CPPs to carry functional cargo have been applied to control signal transduction-dependent subcellular traffic of transcription factors such as NF kappa-B mediating the cellular responses to different agonists. More recently, the authors have used the SN50 peptide sequence for intracellular protein therapy with SOCS3 to inhibit inflammation and apoptosis induced by bacterial enterotoxins [89].

### 3.2 Molecular Umbrellas

A conventional strategy to enhance transbilayer transport of several drugs, at least in the case of small molecules, is to simply modify the lipophilicity of the molecule by chemical conjugation to hydrophobic species such as fatty acids or cholesterol. However, such strategies seem merely to facilitate the adsorptive pinocytosis of macromolecules owing to their highly polar nature. Steven Regen et al. have extended this principle beyond simply altering the hydrophilic/hydrophobic balance of macromolecules. Their strategy hinges on an "umbrella" mechanism involving chemical conjugates that adopt a shielded conformation to traverse the plasma membrane (Illustrated in Figure 2, [90]).

Molecular umbrellas are a unique class of amphiphiles that are capable of shielding an attached agent from an incompatible environment. Typically, such molecules are composed of a central scaffold that contains two or more facially amphiphilic units. When a hydrophilic agent is attached to an umbrella molecule, contact with water favors an exposed conformation such that intramolecular hydrophobic interactions are maximized. Alternatively, when immersed in a hydrophobic environment, such as the lipid bilayer, the molecular umbrella can shield the agent by providing a hydrophobic exterior. Recent mechanistic studies have provided strong evidence that facial amphiphilicity plays a major role in umbrella transport across lipid bilayers. The fact that facial amphiphilicity is more important than hydrophobic/hydrophilic balance in promoting transbilayer transport lends strong support for

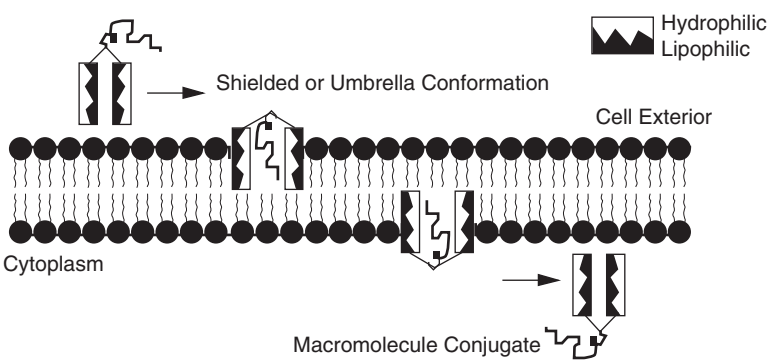


**Figure 2.** Proposed mechanism of cystolic delivery of macromolecules conjugated to molecular umbrellas.

an umbrella mechanism. In essence, the permeation pathway involves monolayer insertion of a shielded conformer, transbilayer diffusion, and entry to the opposite membrane/water interface. These molecular umbrellas were demonstrated to promote transport and release of small peptides by using a detachable handle based on disulfide cleavage. More recently, a series of molecular umbrellas derived from cholic acid, deoxycholic acid, spermidine, lysine, and cleavable disulfide linkages were found capable of transporting an attached 16-mer oligonucleotide across liposomal membranes. The ability of these novel agents in promoting the membrane translocation of macromolecules in a cellular setting remains to be seen [91].

## 4 ASSAYS FOR MONITORING CYTOSOLIC DELIVERY

In principle, the drug carrier or cargo molecules or both could be labeled with a variety of different radionuclides such as [$^{125}$I] or [$^{3}$H] for quantitative assessment. However, the disadvantage associated with such assays is the need for separating the cytosolic fraction from organelles such as endosomes, lysosomes, and nuclei using differential centrifugation techniques. Zaro and Shen have developed a novel method to quantitate the membrane transduction efficiency of CPPs and other drug carriers [92]. In essence, the amount of oligopeptide located in the cytoplasmic compartment was separated from the amount located in the intracellular vesicles using size exclusion chromatography. This separation allowed for the measurement of the membrane transduction efficiency of various [$^{125}$I]-labeled oligopeptides.

Most assays to determine the efficiency of cellular uptake and cytosolic delivery of macromolecules have involved the use of fluorescence microscopy or flow cytometry techniques. However, a major drawback associated with such qualitative fluorescence-based assays is their propensity for generating false-positives. For example, in the abundant literature dealing with CPPs, optimistic reviews highlighting their promising properties coexist with articles questioning their actual efficacy. The tight binding of most CPPs to negatively charged cell surfaces is responsible for microscopy misinterpretations. Methods such as methanol fixation of cells before microscopic observations have been shown to redistribute CPPs. Thus, the experimental assessment of the cytosolic delivery of proteins and other macromolecules is hindered by these artifacts. In fact such artifacts, which affect conventional as well as confocal microscopy analyses, has led to the false conclusion that CPPs, such as TAT and Penetratin, enter cells through an endocytosis-independent fashion [58, 93]. Recent studies suggest that the interaction of HIV-TAT with the cell surface leads to internalization of HIV-TAT fusion proteins by lipid raft-dependent macropinocytosis.

Owing to aforementioned issues, assays that reflect the functional activity of the cargo rather than the carrier have been developed. These assays can broadly be categorized on the basis of protein-based drugs and nucleic acid therapeutics. Prerequisite conditions for presentation of antigenic peptides through the MHC class I pathway are endosomal escape, followed by ubiquitin conjugation in the cytosol, which in turn targets the protein for proteasomal degradation. Conventional antigen presentation assays based on activation of cytotoxic T lymphocytes by monitoring cytokine levels and [$^{51}$Cr] release from target cells have been described elsewhere. Loison et al. have devised an ubiquitin-based assay for monitoring the cytosolic

delivery using CPPs [94]. The authors developed a stringent enzymatic assay based on the processing of a CPP-Ubiquitin-Cargo fusion protein by exclusively cytosolic deubiquitinating enzymes. Based on this strategy, insertion of a ubiquitin moiety between the CPP and the protein cargo serves as a stringent test of cytosolic delivery by ensuring the cytosolic release of the cargo, followed by its proper folding and localization, which is required for interaction with intracellular targets.

Another relatively simple approach to assess cytosolic delivery is to monitor direct enzymatic activity of the protein cargo, such as β-galactosidase or horse radish peroxidase. However, as such enzyme activity could be recovered from within endosomes after fixation and staining, it is not possible to quantitate the cytosolic fraction exclusively. To address this question, a functional assay using a conjugate of TAT-RNase A was tested for its ability to inhibit protein synthesis through the nonspecific degradation of cellular RNA. In this model, if TAT-RNase A were entering the cell and remaining within endosomes, there should be no effect on protein synthesis. However, cytosolic entry of TAT-RNase A was found to be sufficient to decrease cellular protein synthesis as monitored by [$^{35}$S]-methionine incorporation. Another functional assay involved the use of TAT-Cre-Recombinase to mediate recombination of a *lox*P-STOP-*lox*P enhanced green fluorescent protein (EGFP) reporter gene in live mouse reporter T cells (tex.loxP.EG) as a measure of cellular uptake [81]. This system requires that exogenous TAT-Cre protein enter the cytosol, translocate to the nucleus, and excise the transcriptional STOP DNA seg-ment in live cells in a nontoxic fashion before scoring positive for EGFP expression.

Nucleic acid-based therapeutics such as antisense oligonucleotides, siRNA, and plasmid DNA act by interaction with translational or transcriptional protein machinery to downregulate a specific protein or promote transgene expression. In contrast to protein cargo, the interpretation and quantitation of cytosolic delivery of RNA/DNA-based therapeutics is relatively straightforward. Within this framework, reporter proteins such as luciferase and EGFP have been commonly used. Among several functional assays, the upregulation of reporter proteins by correction of aberrant splicing using antisense oligomers is notable [95]. This assay has recently been extended into an *in vivo* mouse model. In this model, the expressed gene (*EGFP-654*) encoding enhanced green fluorescence protein is interrupted by an aberrantly spliced mutated intron of the human β-globin gene. Aberrant splicing of this intron prevents expression of EGFP in all tissues. However, cytosolic entry of the splice site-targeted antisense oligomer restores corrective splicing and results in upregulation of EGFP in several tissues and organs.

## REFERENCES

1. Cho M J, Juliano R (1996). Macromolecular versus small-molecule therapeutics: Drug discovery, development and clinical considerations. *Trends Biotechnol*. 14:153–158.
2. Jain R K (1998). The next frontier of molecular medicine: delivery of therapeutics. *Nat. Med*. 4:655–657.
3. Brockman S, Murphy R (1993). Endosomal and lysosomal hydrolases. In T J Raub, K Audus (eds.), *Biological Barriers to Protein Delivery*: Plenum Press, New York, pp. 51–65.

4. Gray G D, Basu S, Wickstrom E (1997). Transformed and immortalized cellular uptake of oligodeoxynucleoside phosphorothioates, 3′-alkylamino oligonucleotides, 2′-O-methyl oligoribonucleotides, oligodeoxynucleoside methylphosphonates, and peptide nucleic acids. *Biochem. Pharmacol.* 53:1465–1476.
5. Juliano R L, Alahari S, Yoo H, et al. (1999). Antisense pharmacodynamics: Critical issues in the transport and delivery of antisense oligonucleotides. *Pharmaceut. Res.* 16:494–502.
6. Braasch D A, Corey D R (2002). Novel antisense and peptide nucleic acid strategies for controlling gene expression. *Biochem.* 41:4503–4510.
7. Conner S D, Schmid S L (2003). Regulated portals of entry into the cell. *Nature.* 422:37–44.
8. Cohen A W, Hnasko R, Schubert W, et al. (2004). Role of caveolae and caveolins in health and disease. *Physiol. Rev.* 84:1341–1379.
9. Saito G, Swanson J A, Lee K D (2003). Drug delivery strategy utilizing conjugation via reversible disulfide linkages: role and site of cellular reducing activities. *Adv. Drug Del. Rev.* 55:199–215.
10. Austin C D, Wen X, Gazzard L, et al. (2005). Oxidizing potential of endosomes and lysosomes limits intracellular cleavage of disulfide-based antibody-drug conjugates, *Proc. of the National Academy of Sciences of the United States of America.* 102: 17987–17992.
11. Mastrobattista E, Koning G A, van Bloois L, et al. (2002). Functional characterization of an endosome-disruptive peptide and its application in cytosolic delivery of immunoliposome-entrapped proteins. *J. Bioli. Chem.* 277:27135–27143.
12. Li W, Nicol F, Szoka F C, Jr (2004). GALA: A designed synthetic pH-responsive amphipathic peptide with applications in drug and gene delivery. *Adv. Drug Deli. Rev.* 56:967–985.
13. Murthy N, Robichaud J R, Tirrell D A, et al. (1999). The design and synthesis of polymers for eukaryotic membrane disruption. *J. Control. Rel.* 61:137–143.
14. Stayton P S, El-Sayed M E, Murthy N, et al. (2005). "Smart" delivery systems for biomolecular therapeutics. *Orthodn. Craniofacial Res.* 8:219–225.
15. Midoux P, LeCam E, Coulaud D, et al. (2002). Histidine containing peptides and polypeptides as nucleic acid vectors. *Somat Cell Molec. Genetics.* 27:27–47.
16. Pack D W, Putnam D, Langer R (2000). Design of imidazole-containing endosomolytic biopolymers for gene delivery. *Biotechnol. Bioengineer.* 67:217–223.
17. Lee E S, Na K, Bae Y H (2005). Super pH-sensitive multifunctional polymeric micelle. *Nano Letters.* 5:325–329.
18. Boussif O, Lezoualc'h F, Zanta M A, et al. (1995). A versatile vector for gene and oligonucleotide transfer into cells in culture and in vivo: polyethylenimine, *Proc. of the National Academy of Sciences of the United States of America.* 92:7297–7301.
19. Kukowska-Latallo J F, Bielinska A U, Johnson J, et al. (1996). Efficient transfer of genetic material into mammalian cells using starburst poolyamidoamine dendrimers, *Proc. of the National Academy of Sciences of the United States of America.* 93:4897–4902.
20. Shi L, Yang L, Chen J, et al. (2004). Preparation and characterization of pH-sensitive hydrogel of chitosan/poly(acrylic acid) co-polymer. *J. Biomat. Sci. Polymer Ed.* 15:465–474.
21. Sonawane N D, Thiagarajah J R, Verkman A S (2002). Chloride concentration in endosomes measured using a ratioable fluorescent Cl- indicator: Evidence for chloride accumulation during acidification. *J. Biolog. Chem.* 277:5506–5513.

22. Yatvin M B, Kreutz W, Horwitz B A, et al. (1980). pH-sensitive liposomes: Possible clinical implications. *Science.* 210:1253–1255.
23. Connor J M, Yatvin B, Huang L (1984). pH-sensitive liposomes: Acid-induced liposome fusion, *Proc. of the National Academy of Sciences of the United States of America.* 81:1715–1718.
24. Simões S, Slepushkin V, Duzgunes N, et al. (2001). On the mechanisms of internalization and intracellular delivery mediated by pH-sensitive liposomes. *Biochim. Biophys. Acta.* 1515:23–37.
25. Wang C Y, Huang L (1987). pH-sensitive immunoliposomes mediate target-cell-specific delivery and controlled expression of a foreign gene in mouse, *Proc. of the National Academy of Sciences of the United States of America.* 84:7851–7855.
26. Collins D, Litzinger D C, Huang L (1990). Structural and functional comparisons of pH-sensitive liposomes composed of phosphatidylethanolamine and three different diacylsuccinylglycerols. *Biochim. Biophys. Acta.* 1025:234–242.
27. Semple S C, Klimuk S K, Harasym T O, et al. (2001). Efficient encapsulation of antisense oligonucleotides in lipid vesicles using ionizable aminolipids: formation of novel small multilamellar vesicle structures. *Biochim. Biophys. Acta.* 1510:152–166.
28. Chen F J, Asokan A, Cho M J (2003). Cytosolic delivery of macromolecules: I. Synthesis and characterization of pH-sensitive acyloxyalkylimidazoles. *Biochim. Biophys. Acta.* 1611:140–150.
29. Asokan A, Cho M J (2003). Cytosolic delivery of macromolecules. II. Mechanistic studies with pH-sensitive morpholine lipids. *Biochim. Biophys. Acta.* 1611:151–160.
30. Asokan A, Cho M J (2004). Cytosolic delivery of macromolecules. 3. Synthesis and characterization of acid-sensitive bis-detergents. *Bioconjug. Chem.* 15:1166–1173.
31. Murthy, N, Thng Y X, Schuck S, et al. (2002). A novel strategy for encapsulation and release of proteins: Hydrogels and microgels with acid-labile acetal cross-linkers. *J. Amer. Chem. Soc.* 124:12398–12399.
32. Murthy N, Xu M, Schuck S, et al. (2003). A macromolecular delivery vehicle for protein-based vaccines: acid-degradable protein-loaded microgels, *Proc. of the National Academy of Sciences of the United States of America.* 100:4995–5000.
33. Wang C, Ge Q, Ting D, et al. (2004). Molecularly engineered poly(ortho ester) microspheres for enhanced delivery of DNA vaccines. *Nat. Mat.* 3:190–196.
34. Heffernan M J, Murthy N (2005). Polyketal nanoparticles: A new pH-sensitive biodegradable drug delivery vehicle. *Bioconjug. Chem.* 16:1340–1342.
35. Cordes E H, Bull H G (1974). Mechanism and catalysis for hydrolysis of acetals, acetals and orthoesters. *Chem. Rev.* 74:581–603.
36. Boomer J A, Thompson D H (1999). Synthesis of acid-labile diplasmenyl lipids for drug and gene delivery applications. *Chem. Phys. Lipids.* 99:145–153.
37. Guo X, Szoka F C, Jr (2001). Steric stabilization of fusogenic liposomes by a low-pH sensitive PEG–diortho ester–lipid conjugate. *Bioconjug. Chem.* 12:291–300.
38. Guo X, Szoka F C, Jr (2003). Chemical approaches to triggerable lipid vesicles for drug and gene delivery. *Accounts of Chem. Res.* 36:335–341.
39. Saito G, Amidon G L, Lee K D (2003). Enhanced cytosolic delivery of plasmid DNA by a sulfhydryl-activatable listeriolysin O/protamine conjugate utilizing cellular reducing potential. *Gene Ther.* 10:72–83.
40. Tang F, Hughes J A (1998). Introduction of a disulfide bond into a cationic lipid enhances transgene expression of plasmid DNA. *Biochem. Biophys. Res. Communic.* 242:141–145.

41. Byk G, Wetzer B, Frederic M, et al. (2000). Reduction-sensitive lipopolyamines as a novel nonviral gene delivery system for modulated release of DNA with improved transgene expression. *J. Med. Chem.* 43:4377–4387.
42. Chittimalla C, Zammut-ltaliano L, Zuber G, et al. (2005). Monomolecular DNA nanoparticles for intravenous delivery of genes. *J. Amer. Chem. Soc.* 127:11436–11441.
43. Pak C C, Ali S, Janoff A S, et al. (1998). Triggerable liposomal fusion by enzyme cleavage of a novel peptide-lipid conjugate. *Biochim. Biophys. Acta.* 1372:13–27.
44. Pak C C, Erukulla R K, Ahl P L, et al. (1999). Elastase activated liposomal delivery to nucleated cells. *Biochim. Biophys. Acta.* 1419:111–126.
45. Davis S C, Szoka F C, Jr (1998). Cholesterol phosphate derivatives: Synthesis and incorporation into a phosphatase and calcium-sensitive triggered release liposome. *Bioconjug. Chem.* 9:783–792.
46. Sarkar N R, Rosendahl T, Krueger A B, et al. (2005). "Uncorking" of liposomes by matrix metalloproteinase-9. *Chem. Communicat.* 999–1001.
47. Meers P (2001). Enzyme-activated targeting of liposomes. *Adv. Drug Del. Rev.* 53:265–272.
48. Anderson R G (1998). The caveolae membrane system. *Ann. Rev. Biochem.* 67: 199–225.
49. Nabi I R, Le P U (2003). Caveolae/raft-dependent endocytosis. *J. Cell Biol.* 161:673–677.
50. Pelkmans L, Helenius A (2002). Endocytosis via caveolae. *Traffic.* 3:311–320.
51. Sabharanjak S, Mayor S (2004). Folate receptor endocytosis and trafficking. *Adv. Drug Del. Rev.* 56:1099–1109.
52. Drummond D C, Hong K, Park J W, et al. (2000). Liposome targeting to tumors using vitamin and growth factor receptors. *Vit. Hormones.* 60:285–332.
53. Low P S, Antony A C (2004). Folate receptor-targeted drugs for cancer and inflammatory diseases. *Adv. Drug Del. Rev.* 56:1055–1058.
54. Leamon C P, Cooper S R, Hardee G E (2003). Folate-liposome-mediated antisense oligodeoxynucleotide targeting to cancer cells: Evaluation in vitro and in vivo. *Bioconjug. Chem.* 14:738–747.
55. Derossi D, Calvet S, Trembleau A, et al. (1996). Cell internalization of the third helix of the Antennapedia homeodomain is receptor-independent. *J. Biolog. Chem.* 271: 18188–18193.
56. Vives E, Brodin P, Lebleu B (1997). A truncated HIV-1 Tat protein basic domain rapidly translocates through the plasma membrane and accumulates in the cell nucleus. *J. Biolog. Chem.* 272:16010–16017.
57. Lindgren M, Hallbrink M, Prochiantz A, et al. (2000). Cell-penetrating peptides. *Trends Pharmacol. Sci.* 21:99–103.
58. Richard J P, Melikov K, Vives E, et al. (2003). Cell-penetrating peptides: A reevaluation of the mechanism of cellular uptake. *J. Biolog. Chem.* 278:585–590.
59. Ghibaudi E, Boscolo B, Inserra G, et al. (2004). The interaction of the cell-penetrating peptide penetratin with heparin, heparansulfates and phospholipid vesices investigated by ESR spectroscopy. *J. Peptide Sci.* 11:401–409.
60. Tyagi M, Rusnati M, Presta M, et al. (2001). Internalization of HIV-1 tat requires cell surface heparan sulfate proteoglycans. *J. Biolog. Chem.* 276:3254–3261.
61. Foerg C, Ziegler U, Fernandez-Carneado J, et al. (2005). Decoding the entry of two novel cell-penetrating peptides in HeLa cells: lipid raft-mediated endocytosis and endosomal escape. *Biochem.* 44:72–81.

62. Jones S W, Christison R, Bundell K, et al. (2005). Characterisation of cell-penetrating peptide-mediated peptide delivery. *British J. Pharmacol.* 145:1093–1102.
63. Fischer R, Kohler K, Fotin-Mleczek M, et al. (2004). A stepwise dissection of the intracellular fate of cationic cell-penetrating peptides. *J. Biolog. Chem.* 279:12625–12635.
64. Potocky T B, Menon A K, Gellman S H (2003). Cytoplasmic and nuclear delivery of TAT-derived peptide and a beta-peptide after endocytic uptake into HeLa cells. *J. Biolog. Chem.* 278:50188–50194.
65. Wadia J S, Stan R V, Dowdy S F (2004). Transducible TAT-HA fusogenic peptide enhances escape of TAT-fusion proteins after lipid raft macropinocytosis. *Nat. Med.* 10:310–315.
66. Langel U (2005). Cell-penetrating peptides, mechanisms and applications. *Curr. Pharmaceut. Des.* 11:3595.
67. Astriab-Fisher A, Sergueev D, Fisher M, et al. (2002). Conjugates of antisense oligonucleotides with the Tat and antennapedia cell-penetrating peptides: Effects on cellular uptake, binding to target sequences, and biologic actions. *Pharmaceut. Res.* 19:744–754.
68. Koppelhus U, Awasthi S K, Zachar V, et al. (2002). Cell-dependent differential cellular uptake of PNA, peptides, and PNA-peptide conjugates. *Antisense and Nucleic Acid Drug Devel.* 12:51–63.
69. Moulton H M, Hase M C, Smith K M, et al. (2003). HIV Tat peptide enhances cellular delivery of antisense morpholino oligomers. *Antisense Nucleic Acid Drug Dev.* 13:31–43.
70. Oehlke J, Birth P, Klauschenz E, et al. (2002). Cellular uptake of antisense oligonucleotides after complexing or conjugation with cell-penetrating model peptides. *Euro. J. Biochem.* 269:4025–4032.
71. Xia H, Mao Q, Davidson, B L (2001). The HIV Tat Protein Transduction Domain Improves the Biodistribution of β-Glucuronidase Expressed from Recombinant Viral Vectors. *Nat. Biotechnol.* 19:640–644.
72. Bucci M, Gratton J P, Rudic R D, et al. (2000). In vivo delivery of the caveolin-1 scaffolding domain inhibits nitric oxide synthesis and reduces inflammation. *Nat. Med.* 6:1362–1367.
73. May M J, D'Acquisto F, Madge L A, et al. (2000). Selective inhibition of NF-kB activation by a peptide that blocks the interaction of NEMO with the IkB kinase complex. *Science.* 289:1550–1554.
74. Kaufman C L, Williams M, Ryle L M, et al. (2003). Superparamagnetic iron oxide particles transactivator protein-fluorescein isothiocyanate particle labeling for in vivo magnetic resonance imaging detection of cell migration: uptake and durability. *Transplantation.* 76:1043–1046.
75. Torchilin V P, Levchenko T S, Rammohan R, et al. (2003). Cell transfection in vitro and in vivo with nontoxic TAT peptide-liposome-DNA complexes, *Proc. of the National Academy of Sciences of the United States of America.* 100:1972–1977.
76. Lee H J, Pardridge WM (2001). Pharmacokinetics and delivery of tat and tat-protein conjugates to tissues in vivo. *Bioconjug. Chem.* 12:995–999.
77. Pooga M, Soomets U, Hallbrink M, et al. (1998). Cell penetrating PNA constructs regulate galanin receptor levels and modify pain transmission in vivo. *Nat. Biotechnol.* 16:857–861.
78. Leifert J A, Harkins S, Whitton JL (2002). Full-length proteins attached to the HIV tat protein transduction domain are neither transduced between cells, nor exhibit enhanced immunogenicity. *Gene Therapy.* 9:1422–1428.

79. Ye D, Xu D, Singer A U, et al. (2002). Evaluation of strategies for the intracellular delivery of proteins. *Pharmac. Res.* 19:1302–1309.
80. Becker-Hapak M, McAllister S S, Dowdy S F (2001). TAT-mediated protein transduction into mammalian cells. *Methods.* 24:247–256.
81. Wadia J S, Stan R V, Dowdy S F (2004). Transducible TAT-HA fusogenic peptide enhances escape of TAT-fusion proteins after lipid raft macropinocytosis. *Nat. Med.* 10:310–315.
82. Ho A, Schwarze S R, Mermelstein S J, et al. (2001). Synthetic protein transduction domains: enhanced transduction potential in vitro and in vivo. *Cancer Res.* 61:474–477.
83. Futaki S (2002). Arginine-Rich Peptides: Potential for Intracellular Delivery of Macromolecules and the Mystery of the Translocation Mechanisms. *Internat. J. Pharmac.* 245:1–7.
84. Wender P A, Mitchell D J, Pattabiraman K, et al. (2000). The design, synthesis, and evaluation of molecules that enable or enhance cellular uptake: Peptoid molecular transporters, *Proc. of the National Academy of Sciences of the United States of America.* 97:13003–13008.
85. Morris M C, Vidal P, Chaloin L, et al. (1997). A new peptide vector for efficient delivery of oligonucleotides into mammalian cells. *Nucleic Acids Res.* 25:2730–2736.
86. Elliott G, O'Hare P (1997). Intercellular trafficking and protein delivery by a herpesvirus structural protein. *Cell.* 88:223–233.
87. Aints A, Guven H, Gahrton G, et al. (2001). Mapping of herpes simplex virus-1 VP22 functional domains for inter- and subcellular protein targeting. *Gene Therapy.* 8:1051–1056.
88. Hawiger J (1999). Noninvasive intracellular delivery of functional peptides and proteins. *Curr. Opin. Chem. Biol.* 3:89–94.
89. Jo D, Liu D, Yao S, et al. (2005). Intracellular protein therapy with SOCS3 inhibits inflammation and apoptosis. *Nat. Med.* 11:892–898.
90. Janout V, Staina I V, Bandyopadhyay P, Regen S L (2001). Evidence for an umbrella mechanism of bilayer transport. *J. Amer. Chem. Soc.* 123:9926–9927.
91. Janout V, Jing B, Regen S L (2005). Molecular Umbrella-Assisted Transport of an Oligonucleotide across Cholesterol-Rich Phospholipid Bilayers. *J. Amer. Chem. Soc.* 127:15862–15870.
92. Zaro J L, Shen W C (2003). Quantitative comparison of membrane transduction and endocytosis of oligopeptides. *Biochem. Biophys. Res. Communicat.* 307:241–247.
93. Lundberg M, Johansson M (2001). Is VP22 nuclear homing an artifact? *Nat. Biotechnol.* 19:713–714.
94. Loison F, Nizard P, Sourisseau T, et al. (2005). A ubiquitin-based assay for the cytosolic uptake of protein transduction domains. *Molec. Ther.* 11:205–214.
95. Kang S H, Cho M J, Kole R (1998). Up-regulation of luciferase gene expression with antisense oligonucleotides: Implications and applications in functional assay development. *Biochem.* 37:6235–6239.

# BASIC ISSUES IN THE MANUFACTURE OF MACROMOLECULES

Corinne Lengsfeld and Thomas Anchordoquy
*University of Denver, Denver, Colorado*

**Chapter Contents**

1 Introduction to Bioprocessing
  1.1 General Method for Manufacturing Therapeutic Proteins
  1.2 General Methods for Manufacturing Plasmid DNA
2 Review of Molecular Structure
  2.1 Comparison of Protein and DNA Primary Structure
  2.2 Comparison of Protein and DNA Higher Order Structure
3 Manufacturing Concerns
  3.1 pH Sensitivity
  3.2 Metal Catalyzed and Enzymatic Degradation
  3.3 Hydrodynamic Degradation
  3.4 Cavitation
  3.5 Surface Interaction
  3.6 Freeze–Thaw Cycling
4 Example Process—Atomization
  4.1 Common Techniques Employed in Processing and Delivery
  4.2 Degradation Behavior of Linear Genomic DNA
  4.3 Degradation Behavior of Plasmid DNA
  4.4 Applications to siRNA Molecules
  4.5 Degradation Behavior of Proteins
5 Separation Techniques
  5.1 Classic Approaches
  5.2 Novel Molecular-Scale Probing Concepts
6 Chapter Summary
  References

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

# 1 INTRODUCTION TO BIOPROCESSING

In general the manufacturing methods for proteins and plasmid DNA require a cellular medium to either express a protein or replicate genetic material, which is later harvested and separated from undesirable cellular material. The type of cell culture will play a large role in the efficiencies of macromolecular production as well as in the separation process. Often forgotten are the manufacturing steps that bridge the raw purified product and that which is delivered. Lyophilization, encapsulation, delivery, and formulation steps suffer many of the same issues as those in the initial growth and harvesting steps. In the following sections, we briefly describe the initial manufacturing steps involved in the production of therapeutic proteins and plasmid DNA so that later sections specific manufacturing issues can be placed in context. However, it is in these later sections we will draw connections to the later manufacturing stages.

## 1.1 General Method for Manufacturing Therapeutic Proteins

The production of therapeutic protein starts with the design of a recombinant DNA that is typically transfected into mammalian cells, *Escherichia coli*, or yeast, although some transgenic plants and animals are being employed. Conditions, especially in the medium, that elicit the excretion of the desired proteins into extracellular medium are typically preferred over harvesting through cell disruption because of the reduction in contaminants and subsequent purification steps required. The recovery of the product in active form is complicated by protein aggregation that can be promoted by separation and purification processes. It should be recognized that the medium that generated the best protein extrusion may now hinder recovery because of high viscosity or a tendency to promote colloids due to the elevated protein concentrations. The isolation stage normally involves high-pressure homogenization, bead milling, centrifugation, filtration, precipitation, and/or ultrafilration. This stage is followed by a purification stage involving chromatographic sequence (i.e., polishing, desalting, buffer exchange, ion exchange, and reverse phase) that balances yield and purity. An in-depth discussion of all unit-based application can be found in Ref. 1.

## 1.2 General Methods for Manufacturing Plasmid DNA

The production of plasmid DNA begins with the replication of the desired therapeutic gene by bacterial cells, most often *E. coli* [2]. The cells are allowed to grow in a nutritious broth before they are harvested, typically by centrifugation. In a process called cell lysis, the harvested cells are broken open either mechanically (e.g., high-pressure homogenizer), enzymatically, or chemically (e.g., alkaline lysis). During this process the contents of the cells, including plasmid and genomic DNA, are most susceptible to shearing forces as they are no longer protected by the bacterial membrane. There are several ways of removing and purifying the plasmid DNA from the rest of the cell debris. These ways include centrifugation, filtration, and chromatography methods, which are briefly reviewed by Durland and Eastman [3].

## 2 REVIEW OF MOLECULAR STRUCTURE

Like all polymers, proteins and DNA are composed of a sequence of repeating units that interact at various levels to produce the active structure. Although the primary structure of DNA does largely determine its physical and biological properties, that of proteins is of minimal use in determining the desired active structure. This section is not intended to replace a biochemistry textbook, but it provides the basic information necessary to understand how the molecular structure of proteins and plasmid DNA differ, and how this impacts macromolecular stability in a general sense.

### 2.1 Comparison of Protein and DNA Primary Structure

Proteins, peptides, and polypeptides are single-stranded polymer chains composed of repeating amino acids (20 different amino acids are available) and only differ in the number of amino acids that comprise the polymer chain. Although the definitions of these terms are not strictly defined, peptides typically have less than 50 amino acids, whereas polypeptides possess up to 100 amino acids, and proteins generally have more than 100 amino acids covalently linked (through "peptide" bonds) into a polymer chain. Analogous terms for small polymers of nucleic acids, "oligonucleotide" and "polynucleotide," generally refer to chains of <50 and >50 nucleotides (adenosine, guanosine, cytosine, or thymidine), respectively. There is no term for nucleic acids that is analogous to "protein," but long strands of nucleic acids are typically produced as a plasmid, and they are generally >3000 nucleic acids in length. A critical difference between proteins and plasmid DNA is that proteins are a single polymer chain, whereas plasmid DNA consists of two strands that are associated in the classic double helix. Although DNA used for antisense therapy is typically single stranded, thereby designed to associate with a specific sequence to exert a therapeutic effect, these molecules are typically 10–30 nucleotides in length (oligonucleotides) and are chemically synthesized instead of being produced in bacteria as described above. Although the chemical synthesis of oligonucleotides presents its own set of purification issues that are reviewed by Sanghvi and Schulte [4], our discussion will focus on issues involved with the bioprocessing of plasmid DNA.

Simply based on the number of atoms participating in the primary structure of these biopolymers, proteins have a smaller hydrodynamic diameter (<30 nm), whereas typical plasmid DNA molecules range from 100 to 400 nm. As a result, proteins are less susceptible to mechanical forces, especially hydrodynamic forces. This resistance is simply due to a size range that is well below the continuum limit of the fluid. In other words, when particle sizes are well below 100 nm (Figure 1) in aqueous solutions, they "slip" between water molecules rather than having a drag force applied continuously over the particle surface.

### 2.2 Comparison of Protein and DNA Higher Order Structure

Unlike most polymers, these biopolymers have extraordinary secondary and tertiary structures that can dramatically affect therapeutic efficacy. For example, proteins form complex geometries based on hydrophobic interactions and hydrogen

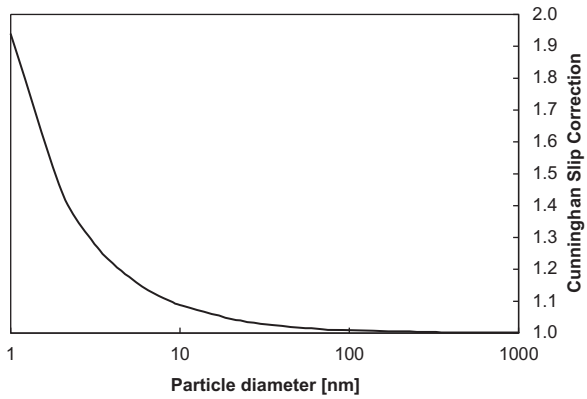


**Figure 1.** Cunningham slip correction vs. particle diameter for water fluid flow.

bonding, which results in well-defined structures, e.g., beta sheet and alpha helix. Because these structures rely on relatively weak bonding interactions (i.e., noncovalent), protein molecules are acutely sensitive to changes in pH and hydrophobic/hydrophilic interfaces. In contrast, plasmid DNA is a double-stranded helix that is produced as a circular form. In addition, this circular form is coiled in a direction opposite that of the double helix, resulting in "negative supercoils" that create torsional stress in the supercoiled form of the molecule. The added stress due to supercoiling is stored in the plasmid and cannot be relieved unless a break occurs in the covalent bonds comprising the phosphate backbone. However, the cleavage of a single covalent bond in the backbone of either strand composing the double helix results in the release of this torsional stress and conversion of the supercoiled form to the relaxed, "open circle" form. Another cleavage in the phosphate backbone of the opposite strand that is in the same vicinity (within a few nucleotides) of the first break will cause the circular integrity to be broken, resulting in a linear form of the plasmid. Because the supercoiled form requires an intact backbone in both DNA strands, supercoil content (the percentage of plasmids that are in the supercoiled form) is used as a measure of quality for a plasmid preparation. Typically, plasmid preparations used in clinical trials have supercoil contents of >90%. Not surprisingly, it is critical that stresses during bioprocessing be low enough to ensure that the integrity of the phosphate backbone is retained so that preparations with sufficiently high supercoil contents are produced.

## 3 MANUFACTURING CONCERNS

Basic manufacturing concerns fall into three categories: those associated with chemical interaction, mechanical phenomena, or thermodynamic energy states. In the following sections, we attempt to provide a general overview of the most common problems as well as the most commonly perceived problems: whether real or imaginary. A clear understanding of the source of degradation during a specific manufacturing process is essential to developing successful mitigation strategies. Numerous mitigation techniques have been described in literature, but it should be

recognized that the best option is dependent on the ultimate form in which the molecule will be administered, e.g., free or in a delivery system.

### 3.1 pH Sensitivity

Extreme pH has been shown to have a role in the break down of macromolecular structure in aqueous solution via hydrolysis [5–7] and aggregation. For example, DNA can disintegrate by depurination reactions resulting in the loss of a purine base and the creation of apurinic (AP) sites. This absence of a base weakens the molecular chain, thereby leading to an increased susceptibility to hydrolytic attack [6]. The destabilized DNA can then undergo beta-elimination of a phosphodiester bond in the backbone, resulting in strand breaks and loss of the supercoiled form. These reactions can be accelerated if pH levels fall below 7.5 or rise above 9.0 [8].

In contrast, proteins will not denature in the same way when exposed to varying pH solutions. Therapeutic proteins are often only stable over a narrow pH range; outside these limits, the aggregation rate is strongly influenced by pH. A solution pH governs the total charge on a molecule; thus, it can affect intra- and intermolecular interactions. Increasing the acidity or alkalinity of the solution beyond the stable range will highly charge nonspecific groups within the protein generating strong repulsive forces, thus destabilizing the protein [9].

pH sensitivity is well known by process engineers, and solution pH is usually carefully controlled, but manufacturing steps that include a gas–liquid interface can unintentionally lower pH if solutions are not properly buffered. The diffusion of carbon dioxide into the aqueous solution could be facilitated by the use of a driving gas that produces a large gas–liquid interface across which substantial gas exchange can occur. Localized nonequilibrium supersaturation of carbon dioxide can lead to large transient changes in pH because carbon dioxide in water can form carbonic acid that can deprotonate and acidify the solution [10].

$$CO_2 + H_2O \rightarrow H_2CO_3$$

$$H_2CO_3 \rightarrow H^+ + HCO_3^-$$

### 3.2 Metal Catalyzed and Enzymatic Degradation

Trace amounts of reduced transition metal ions such as iron and copper are known to generate hydroxyl radicals resulting in oxidized nucleotide bases and strand breaks by means of Fenton-type reactions [11–15]. Diethylenetriaminepentaacetic acid (DTPA) and ethylenediamine tetraacetic acid (EDTA) are compounds that chelate metal ions present in solution or bound to macromolecules. These chelators have also been shown to inhibit DNA degradation via DNase by binding various divalent cations (e.g., $Mg^{2+}$) necessary for enzymatic activity [16].

### 3.3 Hydrodynamic Degradation

Observations of hydrodynamic shear-induced degradation of long polymer molecules first appeared in the literature approximately 60 years ago when Frenkel [17]

mathematically demonstrated that hydrodynamic forces along a long polymer molecule could be sufficient to break covalent bonds. Early experimental investigations into the degradation of long linear DNA molecules blamed hydrodynamic shear forces for the observed fragmentation [18, 19]. These studies varied shear rates through needles and measured degradation through sedimentation rates. This technique did not allow for the clear distinction between molecular scission and structural changes, but these early experiments demonstrated the sensitivity of DNA to hydrodynamic shear stress encountered during stirring and pipetting. But as experimental techniques evolved, it become clear that the shear *rate* instead of the shear *force* was the dominant contributor. In their classic paper, Bowman and Davidson [20] provided the groundwork for relating the probability of fracture to the strain rate. These authors realized that the experimentally measured intact DNA concentration divided by the original concentration represented the total probability ($\langle P \rangle$); an integrated average over each radial position depending on the tube radius ($R$), volumetric flow rate ($Q$), tube length ($l$), and rate constant ($k$), which depends on strain rate ($G$). A complete review of the classic theory is available in Lengsfeld and Anchordoquy [21]:

$$\langle P \rangle = \frac{2\pi l}{Q} \int_0^R krdr \tag{1}$$

When even more conditions (i.e., length and flexibility of the molecule, strain rates, and shear forces) where explored, it became clear that the original classic theory needed modification to be applicable to bioprocessing. Recent work indicates that the degradation of very large biopolymers is similar to cell fragmentation, which has been linked to unsteady flow and turbulent intensity [22, 23]. When energy is transferred to a fluid, some is imparted to the motion of the bulk in a downstream direction, whereas some is consumed by turbulent mixing. This energy is transmitted from eddies on the order of the characteristic length through a cascade of sizes that terminate at the Kolmogorov length scale. At this length scale, energy is dissipated via viscosity.

Several studies investigating the dynamics of individual polymer molecules in flow demonstrated the transition of polymers in their natural coiled state to a stretched configuration in extensional flows [24–27]. Investigations that focus on simple shearing flow have demonstrated that the tumbling of the molecules prevents the elongation of the molecule [27–29]. In these cases, simple shear flow is defined as flow in which the strain rate is equal to the magnitude of the vorticity [28]. Random flows, however, have demonstrated that polymer molecules such as DNA can be effectively stretched [30–32]. In these cases, where the viscous scale of the flow exceeds the length of the polymer, constant strain is applied locally by the turbulent flow [32].

Theoretical models suggest that the transition from a coil-stretch transition within random flows depends on the principle Lyapunov exponent $\lambda 1$ and the longest relaxation time for the molecule $\tau_R$. To stretch the molecules, $\lambda_1 \tau_R$ must be >1 [31, 32]. The Lyapunov exponent can be approximated by taking the fluctuating velocity $u'$, divided by the characteristic length scale, which in turbulence, is determined by the eddies at the viscous scale; i.e., $\lambda = u'/\eta$ [31].

Yim and Shamlou [22] provide a comprehensive list of energy dissipation rates in each unit process used in the manufacture of biomolecules (Table 1). Worden [23] showed that degradation commences when the size of this small-scale turbulent structure becomes equal to or less than the molecule's hydrodynamic diameter (Figure 2). The scale of these smallest eddies η can be calculated based on the energy dissipation rate ε and the diffusivity ν using the relationship derived by Kolmogorov in 1941 [33]:

$$\eta = \left(\frac{\nu^3}{\varepsilon}\right)^{1/4} \tag{2}$$

Estimating the size of the smallest length scale is relatively simple. One could use computational fluid dynamic modeling techniques or estimate them based on the power input to the system (head loss) and the mass of the fluid being powered. For example, in pipe flow, the energy dissipation rate is a function of the total head loss in the flow $h_{ll}$, the volumetric flow rate $Q$, the density of the solution ρ, and the mass of the solution $m$, which in this case is the mass of fluid contained within the pipe [Equation (3)]

**TABLE 1. Energy Dissipation Rate Estimates by Yim and Shamlou for Common Bioprocessing Equipment [22]**

| System | Energy Dissipation Rates [W kg$^{-1}$] |
|---|---|
| Homogenizer | $10^7$–$10^9$ |
| Liquid whistle | $10^7$ |
| Centrifuge | $10^6$–$10^7$ |
| Collid mills | $10^5$ |
| Pumps, jets | $10^3$–$10^5$ |
| Agitated vessels | $10^0$–$10^2$ |
| Static mixers | $10^0$–$10^1$ |
| Bubble columns | $<10^1$ |

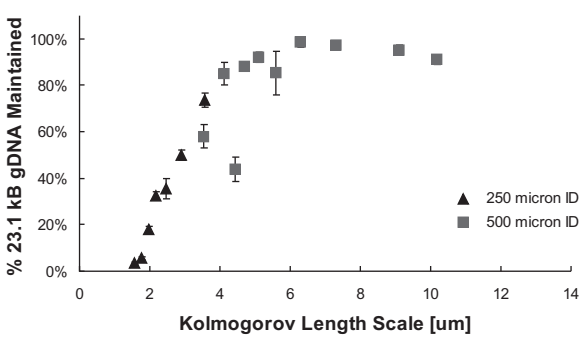


**Figure 2.** Degradation behavior of 23.1-kb genomic DNA during simple pipe flow with an initial contraction. Kolmogorov length scales were varied through changes in contraction ratio (i.e., pipe diameter) and pressure drop.

$$\varepsilon = \frac{h_{lt} Q \rho}{m} \tag{3}$$

Note that to create length scales lower than those of proteins in water requires energy inputs of 1012 W/kg, which is well above that achieved in bioprocessing. Thus, hydrodynamic degradation is only an issue in the production of large macromolecules, e.g., plasmid DNA.

### 3.4 Cavitation

A phenomena present in most pumping systems, but until recently ignored by process engineers is the impact of cavitation on macromolecular stability. For example, the pressure waves induced by the piezoelectric element in an ultrasonic nebulizer have compressive crests and tensile troughs (Figure 3). If sufficient tensile stress is applied at the fluid during the low-pressure cycle, a void of vapor will be formed. Dissolved gases can diffuse to this surface causing the bubble to grow. In stable cavitaion, the bubble oscillates in size about a mean diameter with the passing of the pressure waves. However, in transient cavitation, once the bubble has reached a critical size, the compressive pressure at the crest is sufficient to cause implosion. The bubble collapse releases significant energy in the form of a shock exhibiting high temperatures and pressures, which can generate free radicals.

The factors influencing cavitation are the solution's viscosity, surface tension, vapor pressure, and the presence of contaminants. In addition, the applied forcing frequency and amplitude can contribute to the efficiency of cavitation.

Manufacturing engineers can easily track transient cavitation using a well-established method. When water molecules are exposed to a shock wave (intense temperature and pressures), it becomes ionized into hydrogen and hydroxyl radicals, and two hydroxyl radicals can combine to form hydrogen peroxide. If potassium iodide is present in the solution, the peroxide will react and form iodide and potassium hydroxide. If the solution contains excess potassium iodide, the iodide will form triiodide, which can be measured spectroscopically using ultraviolet absorption at 350 nm. This is essential for plasmid DNA systems where we have determined that the DNA molecule serves as a nucleation site for cavitation. Lentz et al. [34] observed a reduction in nucleation efficiency with size, suggesting that proteins are too small (<30 nm) to nucleate cavitation:

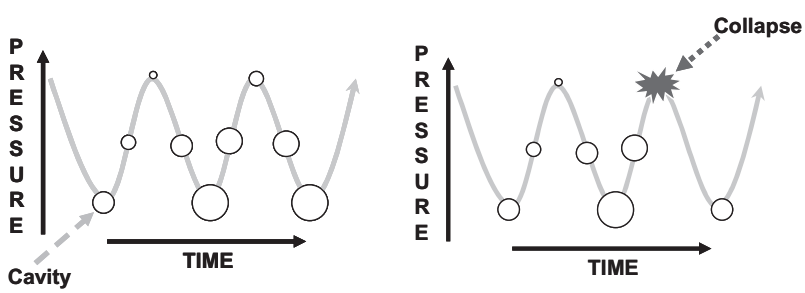


**Figure 3.** Schematic for the evolution of cavitation in (a) stable and (b) transient processes.

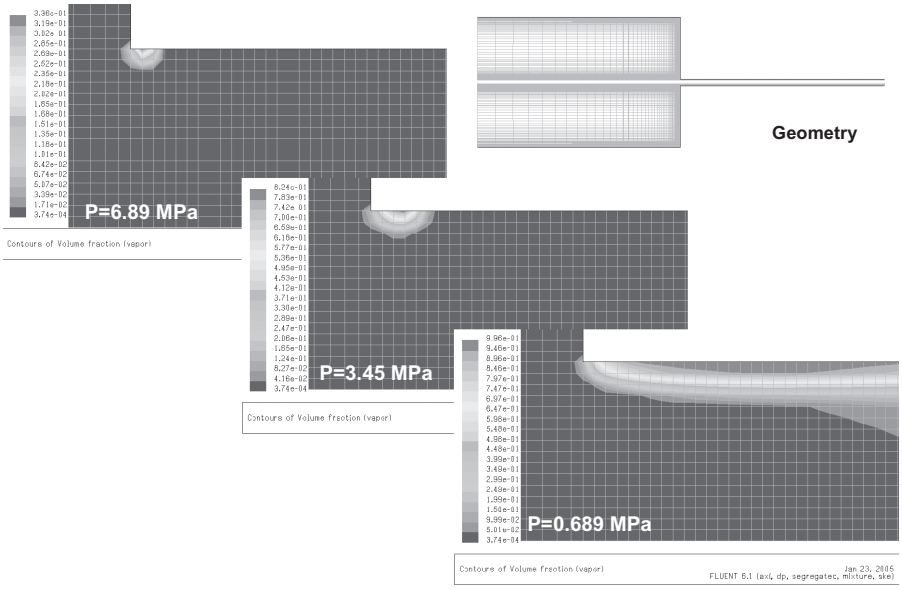


**Figure 4.** Computational fluid dynamics simulation of vapor fraction formation from a tube contraction as a function of tube length (diameter = 500 μm, pressure drop = 3.45 MPa). Red areas indicate 100% vapor regions within the pipe. (This figure is available in full color at ftp://ftp.wiley.com/public/sci_tech_med/pharmaceutical_biotech/.)

$$\mathrm{H-O-H} \xrightarrow{\text{Cavitaton}} \mathrm{H^{\bullet}} + \mathrm{H^{\bullet}}$$

$$2\mathrm{OH^{\bullet}} \longrightarrow \mathrm{H_2O_2}$$

$$\mathrm{KI} + \mathrm{H_2O_2} \longrightarrow \mathrm{I_2} + 2\mathrm{KOH}$$

$$\mathrm{I_2} \xrightarrow{\text{Excess KI}} \mathrm{I_3^{\bullet}}$$

However, unlike DNA, proteins will degrade at the bubble interface. Therefore, modeling stable cavitation may be important in the design phase of protein bioprocessing systems. Figure 4 demonstrates this process using a commercial computational fluid dynamics (CFD) package to observe the formation of vapor phases in low-pressure regions around a tube contraction. As anticipated, the volume of vapor formed increases with decreasing tube length.

### 3.5 Surface Interaction

Gas–liquid interfaces are problematic when processing macromolecules [35, 36]. Proteins, due to their amphipathic nature, undergo conformational changes by adsorbing to the gas–liquid interface [36, 37]. The Gibbs free energy available at the interface allows the protein to rearrange such that hydrophilic components remain in solution while hydrophobic moieties associate with the gas phase. The addition of mild shear stresses can expedite this unfolding process. It was Maa and Hsu [37] who clearly showed the correlation between protein denaturation and surface-to-volume ratio under hydrodynamic shear conditions.

In addition, some research has hypothesized that an attack by molecular oxygen could accelerate the degradation rates of DNA [38]. For example, DNA bases and sugars would be oxidized, resulting in base alterations or strand breaks [5, 39]. However, the investigations by Lentz et al. [40] into the aerosolization of genetic therapeutics showed not evidence of this effect. Although proteins act as surface-active agents and diffuse to hydrophilic–hydrophobic interfaces, DNA is not surface active and remains relatively homogenous throughout the bulk of the fluid (Figure 5). As a result, oxidative attack is controlled by diffusion and solubility limits in aqueous solutions.

It is important to note that these molecules are highly charged and may preferentially diffuse to, or away from, the electric double layer formed at the interface of a metal surface and water. Adsorption to solid surfaces is controlled by electrostatic interaction, dispersion, hydration state, and structural rearrangement because the molecule must be transported from the bulk to the surface, attach, and then spread (complete review in Ref. 41). For example, in rigid proteins (e.g., those in poor solvents), the surface energy increases necessary to spread along the surface must be overcome. Flexible proteins (e.g., those in good solvents) will spread more readily. Most proteins are relatively stable in solution when processed under appropriate buffer conditions. However, at high concentrations, proteins are more likely to adhere to solid surfaces, especially glass and plastics.

What can be done to inhibit molecular adsorption to solid surfaces? Several techniques exist for the protection against surface-induced damage during fermentation, purification, freeze-drying, shipping, and storage. Surfactants and osmolytes (e.g., small organic solutes like sugars, methylamines, and amino acids) all have demonstrated protection against protein aggregation. Surfactants can enhance

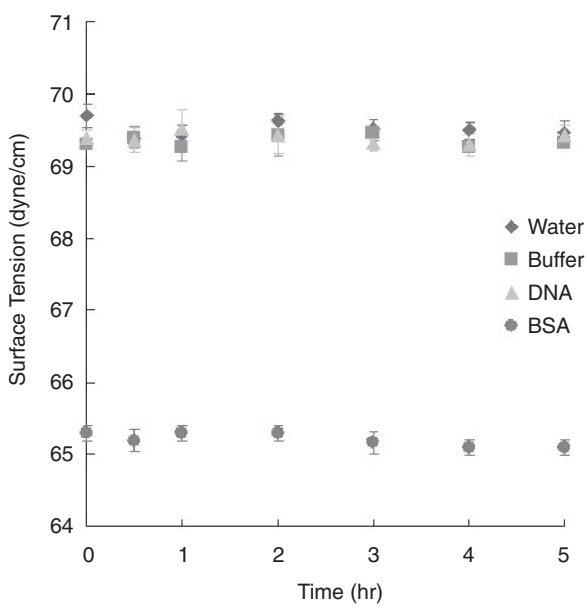


**Figure 5.** Magnitude of surface tension as a function of buffer, DNA, and protein (BSA) concentration [40].

protein stability in four ways by competing with proteins for adsorption sites on surfaces, binding to hydrophobic regions inhibiting intermolecular interactions, increasing the free energy of the unfolding pathway, and acting as a simple chaperone [42]. On the other hand, osmolytes, such as sucrose, are preferentially excluded from the protein surface, thereby increasing the potential energy and subsequently compacting the protein [43, 44]. Instead of manipulating the solution formulation, one can also alter the free energy of the surface. Coating the surfaces with materials that are not porous (i.e., large surface area) or those that are more hydrophilic will reduce adhesion. Exposure to gamma radiation or pulsed plasma [45] can alter the hydrophilic/hydrophobic state of a polymer surface by changing the functionalization of the surface molecules. Other groups have demonstrated that using polyDuramid or poly-N-hydroxyethacrylamid (PHEA) inhibits adhesion, whereas others in the microfluidic industry use polyethylene glycol (PEG) (plastics evaluation in Ref. 46).

### 3.6 Freeze–Thaw Cycling

The formation of discrete ice crystals during freezing causes boundaries to form between different crystal orientations. These boundaries have high surface energies that are known to contribute to fragmentation. As crystal size decreases, the probability that protein or DNA molecules will span one of these high-energy interfaces increases. When storing the therapeutic (plasmid DNA or protein) before use in secondary processing, it is essential to consider the freezing rate because high rates lead to smaller crystal sizes. These small differences in crystal morphology can cause significant batch-to-batch variability.

Embedding a macromolecular therapeutic in a highly crystalline polymer may experience a similar decrease in activity with decreasing particle size. Control over particle size during product formation may therefore be extremely important during fabrication steps.

## 4 EXAMPLE PROCESS—ATOMIZATION

Many issues that plague bioprocessing steps make atomization of biologically active macromolecules difficult. This section will begin with a review of nearly all the classic atomization techniques invented; we will compare and contrast these methods in regard to their ability to sustain molecular conformation from large micron-sized linear polymers down to less than 30 nm for proteins. The goal here is to establish the mechanism for degradation in each technique and to suggest a variety of procedures to help mitigate these issues.

### 4.1 Common Techniques Employed in Processing and Delivery

The generation of an aerosol is as varied as the methods for manufacturing macromolecules. All employ some sort of external force to divide the liquid bulk into individual droplets. Some methods employ aerodynamic forces, mechanical vibration, or pressure oscillations, whereas others harness electrical attraction and repulsion.

**Figure 6.** Photo of a Spraying Systems, Inc. hollow cone misting nozzle.

**Figure 7.** Photo of various commercial jet nebulizers.

Pressure-driven devices (Figure 6), such as pressure swirl atomizers, use aerodynamic drag at the gas–liquid interface to amplify natural disturbance to pinch droplets off liquid sheets and columns. Finer mists can be generated by adding a centrifugal force to the process that must overcome viscous damping and surface tension.

Jet nebulization (Figure 7), the cheapest and most common aerosol device in the pulmonary delivery arena, uses driving gas to accelerate liquid from a reservoir and achieve primary atomization. These primary droplets then hit an obstruction that causes splashing and spreading, which generates secondary droplets in the 1- to 10-μm range.

Ultrasonic nozzles use a metal horn geometry to amplify a small peizoelectric vibration. These vibrations drive surface instabilities along a thin film and generate droplets of very uniform size, but at low flow rates.

The details of how ultrasonic nebulizers create an aerosol is limited. A piezoelectric ceramic drives a pressure disturbance within the fluid at a rate of 15 to

**Figure 8.** Photo of an Aeroneb Pro vibrating mesh atomizer.

19,000 kHz. These pressure waves can form small capillary wave crests at the surface of the liquid or on a central fountain that ejects small liquid droplets from the crest. A third hypothesis is that cavitation events within the liquid cause large pressure singularities that eject liquid mass out from the liquid pool. Most likely, combinations of all three of these phenomena work to generate a fine aerosol suitable for inhalation.

Vibrating mesh technology (Figure 8) uses microfluidic manufacturing techniques to create arrays of small orifices that act as nozzles. Liquid enters these pores via capillary action but are ejected when a high-frequency vibration is imparted to the mesh using a piezoelectric crystal. The droplet uniformity is high and velocity low, making them excellent for respiratory applications.

Fine aerosols can also be generated at low flow rates by creating a potential force between a charged fluid and a ground plate creating a charged jet stream. The charged jet stream breaks into droplets when electrical stresses overcome the surface tension of the fluid. The final size of these droplets are a function of fluid properties (i.e., electrical conductivity, surface tension, density, viscosity, and dielectric constant), fluid flow rate, and applied voltage.

### 4.2 Degradation Behavior of Linear Genomic DNA

The degradation behavior of large molecules in atomizers can be correlated with turbulent length scales when it is the only mechanism acting on the system. Figure 9 demonstrates that when large DNA (7 μm) is atomized by various system; only those with very large turbulent length scale maintain structural integrity.

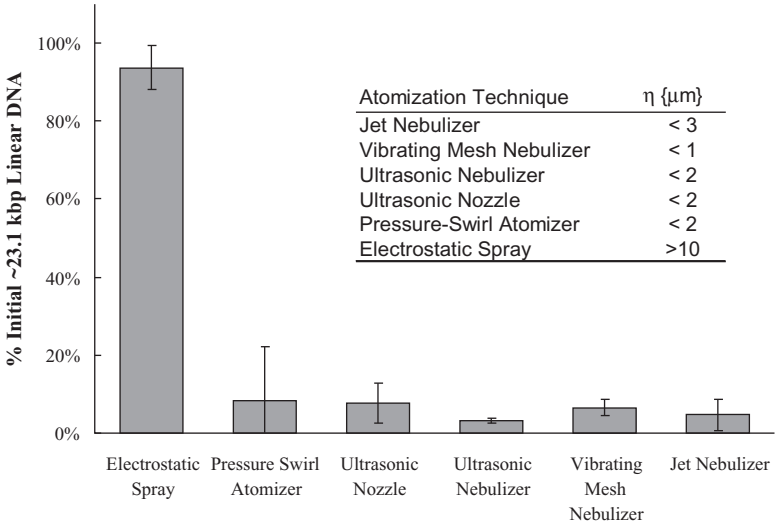


**Figure 9.** Degradation behavior of 23.1 kb genomic DNA after atomization. [23].

### 4.3 Degradation Behavior of Plasmid DNA

The structural degradation behavior of 5-, 10-, and 38-kbp plasmid/cosmid DNA in all atomizers presented so far has also been thoroughly studied. Lentz et al. [47] found that the jet nebulizer, vibrating mesh, and ultrasonic nebulizer are ineffective for all sizes. This is due to the extremely small turbulent length scales featured in the jet and vibrating mesh nebulization, and due to cavitation events nucleated at the DNA during ultrasonic nebulization. The ultrasonic nozzle is only suitable for very small molecular sizes based on length scales, whereas the pressure atomizer and electro spray seem to perform well over all size ranges.

### 4.4 Applications to siRNA Molecules

This chapter is focused on the effect of bioprocessing on proteins and plasmid DNA, but the use of small interfering RNA molecules to silence specific genes is gaining much attention and is currently being tested in clinical trials. Therefore, we thought it would be useful to briefly discuss how the principles described above for DNA might apply to small (≈21 bp), double-stranded RNA molecules. Our laboratory recently undertook an investigation to determine whether these small biopolymers would be degraded during aerosolization. Previous work in DNA indicated that jet nebulization is one of the worst processing techniques; thus, the degradation behavior of a 21-bp siRNA with a calculated melting point of 46°C was subjected to aerosolization by a jet nebulizer (in contrast, plasmid DNA typically has melting temperatures above 100°C). Because these biopolymers are much too small to experience significant shear stress, they are more likely to denature (separation of the two strands in the double helix) than break a phosphodiester bond in their backbone. Degradation can be observed by using molecular probes that bind specifically to double-stranded oligonucleotides (not single-stranded) and

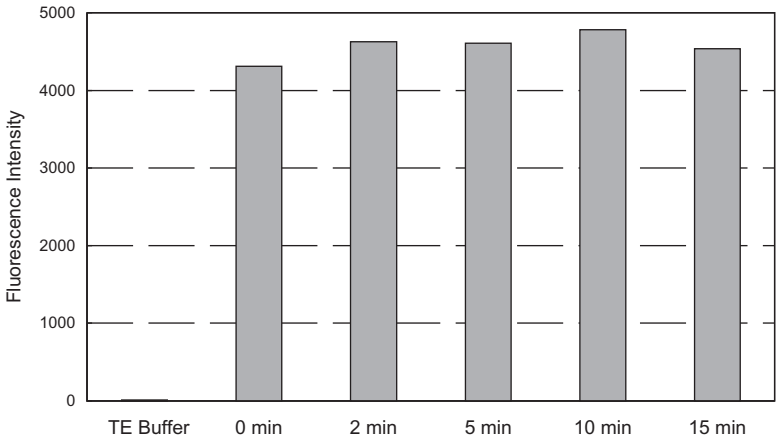


**Figure 10.** Changes in fluorescence intensity during nebulization of 21 bp siRNA with Pico Green fluorescent probes [Zelles-Hahn, Anchordoqury, and Lengsfeld, unpublished data].

fluoresce upon binding. In our experiments, we verified that Pico Green fluorescence was practically eliminated when the siRNA molecule was incubated above its melting point. Upon cooling to temperatures that allow annealing, Pico Green fluorescence was restored. When the siRNA was aerosolized for progressively longer times at room temperature, no change in fluorescence was observed (Figure 10). This indicates that even siRNA molecules with relatively low melting points remain double-stranded upon aerosolization with a jet nebulizer. Although it is possible that siRNA molecules with lower melting points might exhibit denaturation, sufficient stability can usually be achieved by increasing salinity to enhance the double-stranded form.

### 4.5 Degradation Behavior of Proteins

Work by Maa and Hsu [37] and Maa et al. [36] definitively demonstrated that protein degradation during atomization processing steps is due to the large interface to which the molecule readily diffuses (Figure 11). Here the application of a small shear stress is all that is needed to facilitate the unfolding of the amphipathic molecule to preferentially expose hydrophilic moieties to the gas phase (i.e., remove them from water). Thus, increased protein degradation is observed at smaller droplet sizes (Figure 11). In less-stable therapeutic proteins, we have seen batch-to-batch variability, whereas others have had success with using excipients, e.g., sugars and buffers. However, our group and others (Gomez et al. [48], Dunn-Rankin) have observed little to no degradation of protein during electrostatic spraying. This maybe because we all used relatively stable proteins or because our droplet sizes were not well controlled at sizes below 10 μm, but we have not experienced the severe difficulty found with numerous other techniques. Electrosprays are difficult to control with an aqueous solution, so if droplet uniformity and performance is needed over a broad range of conditions (or worse variable conditions), this is not the technique for you.

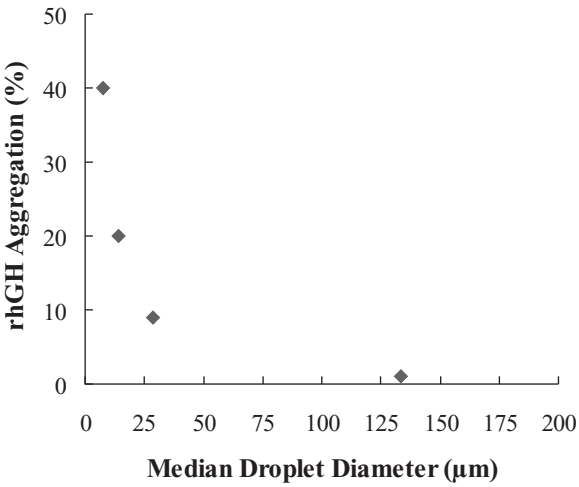


**Figure 11.** Data from Maa demonstrating protein aggregation of rhGH was linked to the surface-to-volume ratio of the liquid droplet produced during atomization [36, 37].

## 5 SEPARATION TECHNIQUES

The separation and purification steps in macromolecular bioprocessing typically cause the largest loss of product. The sources for molecular degradation presented above apply to these techniques, but it is informative to understand the classic approaches used today as well as methods under development for the future. To this end, the following sections provide a general overview of the subject followed by a brief review of molecular scale technologies currently under investigation.

### 5.1 Classic Approaches

Structural changes and contamination must be limited for macromolecular therapies to become a clinical reality. Current techniques for assessment, detection, and separation continue to struggle to meet industrial needs; a novel high-throughput, integrated system capable of qualifying and separating contamination from the bulk is highly desirable.

Conventional DNA purification techniques have been adapted from principles governing protein purification, but the copurification of *E. coli* chromosomal DNA with the desired plasmid presents a unique challenge. Cell lysis, differential precipitation, enzymatic digestion, column chromatography, desalting, buffer exchange, and sterile filtration represent the typical processing steps. The large size of genomic DNA makes it especially susceptible to shearing, which results in fragments that are similar in size to the desired plasmid. Due to their similar size and virtually identical chemical composition, it is difficult to separate the fragmented genomic DNA from the plasmid product. As a result, contemporary plasmid purification protocols sacrifice approximately 30–50% of the therapeutic plasmid in an attempt to minimize genomic DNA contamination. The most common large-scale purification method is column chromatography where separation is achieved largely based on molecule size using anion exchange, size exclusion, and reverse phase. Although

a single pass is relatively effective for proteins (i.e., >95% purity), plasmid purification usually requires more than one of these columns to achieve acceptable levels of purity. In-process quality assessment is essential to biopharmaceutical development as the quality of the product is dependent on the initial stages of growth through every step of purification. Currently, in-process quality assurance tests are conducted off-line using electrophoresis and high-preformance liquid chromatography (HPLC) to assess plasmid concentration, hybridization to measure *E. coli* genomic DNA contamination levels, immunological assays for protein contaminations, and a standard *Limulus* amebocyte lysate assay to assess endotoxin levels. However, the major challenge is the detection of small quantities of bacterial genomic DNA in the large bulk quantity of plasmid DNA. This can be achieved with amplification techniques, but as with many of the other assessment techniques, this must be conducted off-line in a time-consuming manner. As a result, it is not determined whether batches are satisfactory until processing is completed. Clearly, it would be advantageous to develop techniques that allow process modification in real time to improve quality essential to sustaining the batch. For example, these sensors could be used to optimize production rates. High throughput usually requires higher flow velocities, which are known to fragment large macromolecules, but conservatively low velocities lead to long batch production times and revenue limitations. Real-time detection of contaminating DNA fragments could signal the operating system to reduce the energy per mass input to the system, lowering turbulence-induced fragmentation and thereby continuously optimizing throughput and product quality regardless of processing scale. A technique that probes individual molecules would provide an opportunity to overcome this multimillion-dollar problem. Because probing individual molecules requires a device with length scales at or below the molecular length scale, the answer must lie in a nanotechnology-based solution.

### 5.2 Novel Molecular-Scale Probing Concepts

Some work already exists in the area of using micron or submicron fluidic devices to enhance DNA separation techniques. Most of these techniques use electric fields or electrosmotic flow [49] to induce movement in these macromolecules. Electric field-induced diffusion through cylindrical or rectangular obstacles set at an angle of 45° to the field has demonstrated that small molecules move faster than large ones, making continuous sorting possible [50, 51]. Switching the electric field in 10-s intervals can provide enhanced separation rates by slowing the larger molecules to a greater degree [52, 53]. Instead of rectangular post arrays, as Han and Craighead [54] demonstrated, requiring the macromolecule to change conformation by altering the channel dimension from thick to thin, can create entropic traps that collect a particular size range. These confinement channels (post array limited or trap geometries) separate DNA fragments of a different size using the fact that as channel walls prohibit some molecular conformations as channel size approaches the molecular scale or smaller [55–57]. These limitations result in excess free energy (or an entropy gradient), which may impart a measurable force. For example, Turner et al. [55] report forces on the order of femto-Newtons. If entropy is proportional to molecular length, matrices can be designed such that only a specific length scale will have the necessary inherent force to move through the system. Yi

and Lengsfeld [58] have built off these concepts and computationally explored the development of converging channels to alter the speed and compression of molecules at the molecular scale. Their new approach to analyzing molecular scale behavior is computationally efficient, and it will enable extension to circular and supercoiled DNA conformations where the excess energy generated by confinement might not only be governed by molecular length, but also by structurally restricted degrees of freedom inherent to the supercoiled and open circle conformation.

## 6 CHAPTER SUMMARY

Remember when dealing with proteins that (1) the small physical size of the molecule is below the continuum limit and thus the molecule is not susceptible to shear induced degradation; (2) structures that control activity are a result of relatively weak hydrophobic interactions and hydrogen bonds that result in an unstable active structure; and finally (3) proteins are surface active due to the amphipathic nature, and thus large surface-to-volume ratios should be avoided.

Double-stranded RNA molecules, although only discussed in a limited fashion, hold the best traits of DNA and proteins, i.e., siRNA's structure does not lend itself to high surface activity, and thus degradation by absorption to solid– or gas–liquid interfaces is not an issue. Furthermore, siRNA's small size is well below the continuum limit, and therefore, the molecule is not susceptible to shear-induced degradation.

When dealing with plasmid DNA, it is important to remember that the size straddles the continuum limit; thus, the molecule may be susceptible to hydrodynamic forces, especially turbulence. Reductions in power inputs to the fluid volume must be regulated to keep the Kolmogorov length scale larger than the molecule; alternatively, mitigation techniques using condensing cationic molecules must be used. Overall, the best-performing aerosolization devices are those with low residence times, such as pressure swirl devices, and low turbulence systems free of cavitation, e.g., electrostatic sprays.

## REFERENCES

1. Hatti-Kaul R, Mattiasson B, eds. (2003). Isolation and Purification of Protiens. 1st ed.: Marcel Dekker Inc, New York, pp. ix–xvii.
2. Prazeres D M, Ferreira G N, Monteiro G A, et al. (1999). Large-scale production of pharmaceutical-grade plasmid DNA for gene therapy: problems and bottlenecks. *Trends in Biotechnol.* 17:169–174.
3. Durland R H, Eastman E M (1998). Manufacturing and quality control of plasmid-based gene expression systems. *Adv. Drug. Deliv. Rev.* 30:33–48.
4. Sanghvi Y S, Schulte M (2004). Therapeutic oligonucleotides: The state-of-the-art in purification technologies. *Curr. Opin. Drug. Discov. Devel.* 7(6):765–776.
5. Lindahl T (1993). Instability and decay of the primary structure of DNA. *Nature.* 362:709–715.
6. Lindahl T, Andersson A (1972). Rate of chain breakage at apurinic sites in double-stranded deoxyribonucleic acid. *Biochem.* 11:3618–3623.

7. Lindahl T, Nyberg B (1972). Rate of depurination of native deoxyribonucleic acid. *Biochem.* 11:3610–3618.
8. Evans R K, Xu Z, Bohannon K E, et al. (2002). Evaluation of degradation pathways for plasmid DNA in pharmaceutical formulations via accelerated stability studies. *J. Pharmac. Sci.* 89:76–87.
9. Chi E Y, Krishnan S, Randolph T W, et al. (2003). Physical stability of proteins in aqueous solution: Mechanism and driving forces in nonnative protein aggregation. *Pharmac. Res.* 20(9):1325–1336.
10. Tservistas M, Levy M S, Lo-Yim M Y A, et al. (2001). The formation of plasmid DNA loaded pharmaceutical powders using supercritical fluid technology. *Biotechnol. Bioengin.* 72:12–18.
11. Breen A P, Murphy J A (1995). Reactions of oxyl radicals with DNA. *Free Radical Biol. Med.* 18:1033–1077.
12. Graf E, Mahoney J R, Bryant R G, et al. (1984). Iron-catalyzed hydroxyl radical formation. Stringent requirement for free iron coordination site. *J. Biol. Chem.* 259:3620–3624.
13. Keyer K, Imlay J A (1996). Superoxide accelerates DNA damage by elevating free-iron levels, *Proc. of the National Academy of Sciences USA.* 93:13635–13640.
14. Luo Y, Han Z, Chin S M, et al. (1994). Three chemically distinct types of oxidants formed by iron-mediated Fenton reactions in the presence of DNA, *Proc. of the National Academy of Sciences USA.* 91:12438–12442.
15. Middaugh C R, Evans R K, Montgomery D L, et al. (1999). Analysis of plasmid DNA from a pharmaceutical perspective. *J. Pharmac. Sci.* 87:130–146.
16. Cowan J A (1998). Metal activation of enzymes in nucleic acid biochemistry. *Chem. Rev.* 98:1067–1087.
17. Frenkel J (1994). Orientation and rupture of linear macromolecules in dilute solutions under the influence of viscous flow. *Acta Physicochim. USSR.* 19:51–76.
18. Davison P F (1959). The effect of hydrodynamic shear on the deoxyribonucleic acid from T2 and T4 bacteriophages, *Proc. of the National Academy of Science.* 45:1560–1568.
19. Levinthal C, Davison P F (1961). Degradation of deoxyribonucleic acid hydrodynamic shearing forces. J. Molec. Biol. 3:674–683.
20. Bowman R D, Davidson N (1972). Hydrodynamic shear breakage of DNA. *Biopolymers.* 11:2601–2624.
21. Lengsfeld C S, Anchordoquy T J (2002). Shear-induced degradation of plasmid DNA. *J. Pharmac. Sci.* 91:1581–1589.
22. Yim S S, Shamlou P A (2000). The engineering effects of fluids flow on freely suspended biological macro-materials and macromolecules. Adv. Biochem. *Engin./Biotechnol.* 67:83–122.
23. Worden L (2005). Degradation behavior of genomic DNA during bioprocessing. Masters Thesis, University of Denver, Denver, C.O., pp. 69–73.
24. De Gennes P G (1974). Coil-stretch transition of dilute flexible polymers under ultra-high velocity gradients. *J. Chem. Phys.* 60(12):5030–5042.
25. Larson R G, Magda J J (1989). Coil-stretch transitions in mixed shear and extensional flows of dilute polymer solutions. *Macromo.* 22(7):3004–3010.
26. Perkins T T, Smith D E, Chu S (1977). Single polymer dynamics in an elongation flow. *Science.* 276:2016–2021.
27. Smith D E, Chu S (1998). Response of flexible polymers to a sudden elongation flow. *Science.* 281:1335–1340.

28. LeDuc P, Haber C, Bao G, et al. (1999). Dynamics of individual flexible polymers in a shear flow. *Nature*. 399:564–566.
29. Hur J S, Shaqfeh E S G, Babcock H P, et al. (2002). Dynamics and configurational fluctuations of single DNA molecules in linear mixed flows. *Phys. Rev. E*. 66(1):Pt 1, 011915.
30. Lumley J L (1973). Drag reduction in turbulent flow by polymer additives. *J. Polymer Sci.* Part D: Macromolecular Review, 7(1):263–290.
31. Balkovsky E, Fouxon A, Lebedev V (2001). Turbulence of polymer solutions. *Phys. Rev. E*. 64:056301:1–14.
32. Thiffeault J (2003). Finite extension of polymers in turbulent flow. *Phys. Lett.* A:308:445–450.
33. Kundu P K, Cohen I M (2002). Fluid Mechanics, 2nd ed.: Academic Press, San Diego, C.A., pp. 438–440.
34. Lentz Y, Anchordoquy T A, Lengsfeld C S (2006). DNA acts as a nucleation site for transient cavitation in ultrasonic nebulizers. *J. Pharmac. Sci.* 95(3):607–619.
35. Levy M S, O'Kennedy R D, Ayazi-Shamlou P, et al. (2000). Biochemical engineering approaches to the challenges of producing pure plasmid DNA. *Trends in Biotechnol.* 18:296–305.
36. Maa Y F, Nguyen P A, Hsu S W (1998). Spray-drying of air-liquid interface sensitive recombinant human growth hormone. *J. Pharmac. Sci.* 87:152–159.
37. Maa Y F, Hsu C C (1997). Protein denaturation by combined effect of shear and air-liquid interface. *Biotechnol. Bioengin.* 54:503–512.
38. Levy M S, Collins I J, Yim S S, et al. (1999). Effect of shear on plasmid DNA in solution. *Bioproc. Engin.* 20:7–13.
39. Imlay J A, Linn S (1988). DNA damage and oxygen radical toxicity. *Science*. 240:1302–1309.
40. Lentz Y, Worden L, Anchordoquy T A, et al. (2005). Impact of jet nebulization of plasmid DNA: degradation mechanism and mitigation methods. *J. Aerosol Sci.* 36:973–990.
41. Malmsten M ed. (2003). Biopolymers at Interfaces, 2nd ed.: Marcel Dekker Inc, New York.
42. Chou D K, Krishnamurthy R, Randolph T W, et al. (2005). Effects of Tween 20 and Tween 80 on the stability of albutropin during agitation. *J. Pharmac. Sci.* 94(6): 1368–1381.
43. Kim Y S, Jones L S, Dong A, et al. (2003). Effects of sucrose on conformational equilibria and fluctuations within the native-state ensemble of proteins. *Protein Sci.* 12:1251–1261.
44. Kendrick B S, Carpenter J P, Cleland J L, et al. (1998). A transient expansion of the native state precedes aggregation of recombinant human interferon-γ, *Proc. of the National Academy of Science USA*. 95:14142–14146.
45. Teare D O H, Barwick D C, Schofield W C E (2005). Functionalization of Solid Surfaces with Thermoresponsive Protein-Resistant Films. *J. Phys. Chem. B*. 109:22407–22412.
46. DeFife K M, Hagen K M, Clapper D L, et al. (1999). Photochemically immobilized polymer coatings: effects on protein adsorption, cell adhesion, and leukocyte activation. Journal of Biomaterials Science. *Polymer Edition*. 10(10):1063–1074.
47. Lentz Y, Anchordoquy T J, Lengsfeld C S (2006). Impact of Aerosolization Devices on plasmid DNA. *J. Aerosol Med.* In Press.

48. Gomez A, Bingham D, de Juan L, et al. (1998). Production of protein nanoparticles by electrospray drying. *J. Aerosol Sci.* 29(5–6):561–574.
49. Huang M F, Kuo Y C, Huang C C, et al. (2004). Separation of long double-stranded DNA by nanoparticle filled capillary electrophoresis. Anal. Chem. 76(1):192–196.
50. Chou C F, Bakajin O, Turner S W, et al. (1999). Sorting by diffusion: An asymmetric obstacle course for continuous molecular separation. *Proc. Natl. Acad. Sci. USA.* 96(24):13762–13765.
51. Kaji N, Tezuka Y, Takamura Y, et al. (2004). Separation of long DNA molecules by quartz nanopillar chips under a direct current electric field. *Anal. Chem.* 76(1): 15–22.
52. Cole K D, Tellez C M (2002). Separation of large circular DNA by electrophoresis in agarose gels. *Biotechnol. Prog.* 18(1):82–87.
53. Bakajin O, Duke T A I, Tegenfeldt J, et al. (2001). Separation of 100-kilobase DNA molecules in 10 seconds. *Anal. Chem.* 73(24):6053–6056.
54. Han J, Craighead H G (2000). Separation of long DNA molecules in a microfabricated entropic trap array. *Science.* 288(5468):1026–1029.
55. Turner S W P, Cabodi M, Craighead H G (2002). Confinement-induced entropic recoil of single DNA molecules in nanofluidic structure. *Phys. Rev. Let.* 88(12):128103.
56. Cabodi M, Turner S W P, Craighead H G (2002). Entropic recoil separation of long DNA molecules. *Anal. Chem.* 74:5169–5174.
57. Foquet M, Korlach J, Zipfel W, et al. (2002). DNA fragment sizing by single molecule detection in submicrometer sized closed fluidic channels. *Anal. Chem.* 74:1415–1422.
58. Yi Y B, Lengsfeld C S (2006). Mechanical modeling of a CA molecule motion in simple channels. *J. Applied Phy.* In Press.

# PROCESS VALIDATION FOR BIOPHARMACEUTICALS

BETH JUNKER
*Merck Research Laboratories, Rahway, New Jersey*

**Chapter Contents**

1 Biopharmaceuticals
2 Process Validation Overview
  2.1 Definitions
  2.2 Purpose and Benefits
  2.3 Scope
  2.4 Regulatory
3 Evaluating Measures of Product Quality and Consistent Process Performance
  3.1 Critical Process Input Variables/Parameters
  3.2 Noncritical Process Input Variables/Parameters
  3.3 Process Control/Input Parameters (Xs)
  3.4 Performance Uncontrolled Output/Variables/Attributes (Ys)
4 Background and Requirements
  4.1 Types of Process Validation
  4.2 Aspects of Process Validations
  4.3 Process Validation Execution Timing
5 Development Support and Technology Transfer Activities
  5.1 Process Development
  5.2 Assessment of Process Risks
  5.3 Validation of Assays
  5.4 Change Control and Ongoing Process Validation
  5.5 Process Validation Documentation
6 Process Validation Strategies
  6.1 Historical Data Review
  6.2 Statistical Design of Experiments (DOE) and Analysis for Process Characterization
  6.3 Family/Matrix Validation and Platform Technologies

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

6.4 Scale-Down Models
6.5 Well-characterized Biotechnology Products
7 Future Trends
7.1 Process Analytical Technologies (PAT)
7.2 Process Reliability
7.3 Biogenerics
7.4 Outsourcing
7.5 Risk-Based GMPs
8 Summary
Nomenclature
References

## 1 BIOPHARMACEUTICALS

Biopharmaceuticals are drugs, vaccines, and pharmaceutical agents (typically therapeutic proteins or polypeptides) developed or produced using techniques of biotechnology [1]. In many cases they are complex proteins exhibiting varying levels of heterogeneity [2], derived from recombinant gene expres-sion or hybridoma production systems [3]. These products of biotechnology generally are classified as biologicals by regulatory agencies [4]. Specific categories of products include monoclonal antibodies and other recombinant proteins produced in cell culture, enzymes, hormones, soluble receptors, growth factors, imunogens, blood and plasma products, and animal-derived products [3, 5]. Biopharmaceuticals range in molecular weight from a low value of 22,000 for human growth hormone to about 150,000 for an intact immunoglobulin [6]. These agents are among the most expensive products to manufacture, with typical inventory costs of hundreds to thousands of dollars per gram [3]. There is also intense pressure to introduce a product to the market in a timely manner [2, 7] to recover research and facil-ity investments. Consequently, the reliable production of quality biopharmaceuticals is of the highest importance, translating into an intense focus on the creation of an effective and economical production process. Process validation is a key mechanism to successfully accomplish the task of process creation.

## 2 PROCESS VALIDATION OVERVIEW

### 2.1 Definitions

Validation derives from the Latin word *valere* meaning "to have the power" [8], presumably referring to the power to prevent an unexpected and undesirable occurrence. It establishes that the process used is reliable and reproducible [9]. The level of control implemented is supported by scientific knowledge of the culture and product, along with equipment capabilities [10]. The U.S. Food and Drug Administration (FDA) definition of process validation (PV) is documented evidence to

provide a high degree of assurance that a specific process consistently produces a product meeting predetermined specifications and quality characteristics [11]. The International Conference on Harmonization (ICH) definition is similar, stating that process validation is documented evidence that the process, when operated within established parameters, performs reproducibly and effectively to produce an intermediate or active pharmaceutical ingredient (API) meeting predetermined specifications and quality attributes [11]. Most simply stated, process validation provides a high degree of assurance that process will routinely achieve the stated goals [12]. It accomplishes this goal by verifying product and process specifications to determine their weaknesses and reproducibility [13] and by demonstrating at which steps contaminants are removed and the product stream achieves acceptance criteria [14]. A verification of process consistency is conducted, highlighting the ability to produce essentially the same product (with the same safety and efficacy) over time, despite normal variability in operating conditions [12]. Consistency of multiple batches is demonstrated at full scale, with specific focus on the process operating in a consistent manner, reproducible removal of contaminants to acceptable levels, and monitoring parameters that demonstrate this consistency [14].

Process validation for biopharmaceuticals is somewhat different in application from traditional pharmaceuticals because the technology is different, but the theory and principles are similar [15]. Biotechnology quality control is defined by the individual production process and the product itself [16] because quality cannot be tested into the product [4]. The manner in which the biopharmaceutical is produced is part of its description; thus, biopharmaceuticals are characterized by chemical structure as well as by operational definition [17]. Few doubt the need for process validation, but there is disagreement and confusion about what needs to be validated and how to perform it [18]. Operational definitions of process validation have been developed to suit the philosophies of individual companies [3]. Thus, there is increased interest in developing industry guidelines for the design and performance of validation studies [19].

The term *validation* has different implications under different circumstances, and thus, its use can be confusing [19]. Furthermore, the term *validation* has been used interchangeably with other terms such as *performance qualification* and *process characterization*, when discussing process as well as equipment testing. Recognizing that individual firms distinguish among these terms differently, there has not been an attempt in this chapter to create strict definitions that artificially introduce restrictions where none exist currently. Rather, the terminology employed has remained flexible regardless of the specific terminology used in the source material, such that it reads as internally consistent. In addition, the focus of this chapter primarily is on process validation for steps required to produce bulk biological substances, but the principles discussed also can be applied to the final processing steps involved in formulation and packaging.

## 2.2 Purpose and Benefits

Validation has been a critical success factor in product approval and ongoing commercialization [14]. It serves as a formal link of the process to the product [10], focusing on quality but often concurrently addressing critical business concerns. Each Good Manufacturing Practice (GMP) step must be controlled to maximize

the probability that the finished product meets all quality and design specifications [14]. In addition, each step must be evaluated for its function and effect on product identity, purity, potency, safety, or stability [12, 20]. An unvalidatable process is unacceptable and unreliable from both regulatory and manufacturing (i.e., business) perspectives [12].

There are two goals of process validation: (1) to support the safety, efficacy, and quality of the product; and (2) to identify major sources of process variability so appropriate controls can be implemented to provide consistency [10]. Achievement of these goals demonstrates that potential risks have been adequately reduced and assures products consistently possess the required established quality attributes [12]. Weaker points of the manufacturing process are uncovered and strengthened during the validation effort [12]. Thus, process validation deepens process understanding, which in turn decreases the risk of processing problems/failures, defect costs, lot rejections, reworking/reprocessing, product shortages, or regulatory noncompliance [3, 21]. Assurance of process consistency maximizes productivity and increases cost efficiency. It influences process economics in many ways; one example is through the reduction of in-process control and endproduct testing [21] by reducing the number and scope of quality control tests per lot [12].

### 2.3 Scope

Process validation is the culmination of all qualification and validation activities, representing process testing in its entirety [22]. It is a dynamic process varying in depth according to the product life-cycle timeline, spanning from initial process design through ongoing commercial operation [14]. Its execution can be envisioned as creating a "design space," a multidimensional space that defines how different variables interact [23]. This design space (1) establishes the process ranges demonstrated to provide quality assurance, and (2) defines areas in which process adjustments and refinements may be undertaken without regulatory involvement. The dimensions of this design space are based on the scientific understanding of critical process parameters that has been created through development studies [23, 24].

For a product, the overall list of subcomponent systems to be validated includes equipment, facilities, utilities, computers (including software and controls), processing environment, analytical methods, and operating procedures (e.g., cleaning and sterilization), as well as the process itself [14, 25]. Before process validation begins, these subcomponent systems are qualified [11]. As process validation rests on the fidelity of the qualification of the subcomponent systems required for process operation, the undetected shortcomings of these efforts can compromise subsequent process validation [10]. As with equipment qualification, process validation is most effective when conducted in combination with effective process and facility design efforts [12]. Also, as with equipment qualification, predefined acceptance criteria and action plans in case of test failure are devised [26]. The use of predetermined specifications serves to reduce the goals of a validation program to actual practice [12], assisting in implementation.

Process validation for the production process is only one means to assure product consistency. Other factors include quality control for the final product, in-process assays, raw material controls, appropriately designed facilities, personnel monitor-

ing and training, and environmental monitoring [4, 10, 27]. Additional factors that become concerns for multiproduct biopharmaceutical facilities include (1) segregation of materials, personnel, and equipment; (2) cleaning and product changeover; and (3) product cross-contamination [4]. Characteristics of biopharmaceutical processing that generate specific areas of quality control concern are large, complex product molecules, complicated manufacturing systems, long production times, and pooling of fermentation lots before isolation [4]. Other product quality concerns include genetic stability, product yield and stability, host and nonhost cell component contamination, and posttranslational processing, including glycosylation and folding [16]. As biopharmaceutical manufacturing processes are complicated with multiple steps and extended production times, adequate in-process testing, thorough process validation, and appropriate measures to eliminate adventitious agents are necessary [28].

### 2.4 Regulatory

The completion of process validation represents the point at which the science of the process is ready to be explained to regulatory agencies [14]. Process validation data are initially presented to the regulatory agency during application submission [11]. These data set the bar by which the process is to be judged during future inspections, and consequently, one of most common FDA 483 observations is the lack of process validation [14]. As manufacturing is a key factor in achieving the required quality and safety of biotechnology products [27], process validation is a regulatory requirement for product licensure, and limited process validation is considered essential even for product destined for clinical trials [3]. Process validation is required in license submissions for all products regulated by either CBER or CDER [14]. There are added challenges for biopharmaceuticals, however, as process consistency often replaces detailed product analysis as a key method of quality assurance, and biologically based production systems contribute biological contaminants that pose safety risks [29].

It has been stated that validation should be written as if the FDA were the customer [14], and this approach still is valid because worldwide harmonization efforts are well underway [24]. Several guidance documents have been published in draft and finalized forms from various regulatory agencies and industry associations worldwide. Key documents are listed in Table 1. New guidance, as well as interpretations of prior guidance, is continuously published, suggesting that there is not yet one consistent approach to process validation. Additional guidance leading to the improved definition of process validation requirements and reduction in marginally helpful studies is welcomed by industry [3]. Several comprehensive publications describing the interpretation and application of process validation by industry are available, but terminology and definitions have become blurred and over-lapping. A summary of various terms used is shown in Table 2. Consequently, the use of the term selected must be fully and consistently understood by authors, reviewers, approvers, executors, and auditors in the context of its specific application.

***Pitfalls.*** Validation (consisting of process, equipment, and facility systems) is a costly and time-consuming investment with more resources believed devoted to its

**TABLE 1. List of Selected Relevant Guidance Documents**

| Topic | Year of Publication | Authoring Organization | Title |
|---|---|---|---|
| Process validation | 1987, revised 1996 | FDA | Guideline on the general principles for process validation |
| | 2005 (draft) | PDA | Process validation of protein manufacturing |
| Clinical materials | 2001 | EMEA | EU GMP Guide: Manufacture of investigational medicinal products |
| Inspections/ manufacturing | 1985 | FDA | Points to consider in the production and testing of new drugs and biologicals produced by recombinant DNA technology |
| | 1991 | FDA | Biotechnology inspection guide |
| | 1987, 1997 | FDA | Points to consider in manufacture and testing of monoclonal antibody products for human use |
| | 2000, 2003 | ICH | Q7A Good Manufacturing Practice guidance for active pharmaceutical ingredients |
| | 2005 (draft) | ICH | Q8, Pharmaceutical development |
| Cell substrates | 1997 | ICH | Derivation and characterization of cell substrates used for production of biotechnological/ biological products |
| | 1998 | ICH | Q5A Viral safety evaluation of biotechnology products derived from cell lines of human or animal origin |
| | 1985, 1993 | FDA | Points to consider in the characterization of cell lines used to produce biologicals |
| Assay validation | 1995 | ICH | Guideline on validation of analytical procedures |
| | 1996 | ICH | Q2B Validation of analytical procedures: Methodology |
| | 1999 | ICH | Specifications: Test procedures and acceptance criteria for biotechnology/biological products |
| | 2001 | FDA | Guidance for Industry—Bioanalytical method validation |
| PAT | 2003 (draft), 2004 (final) | FDA | Guidance for Industry—PAT—A framework for innovative pharmaceutical manufacture and quality assurance |
| Risk | 2003 | WHO | Application of hazard analysis and critical control point (HACCP) methodology to pharmaceuticals |
| | 2005 (draft) | ICH | Q9, Risk management guidance |

**TABLE 2. Selected Process Validation Terminology**

| Aspect | Classification Level | | |
|---|---|---|---|
| | A | B | C |
| Process | Validation<br>(Performance) qualification<br>Characterization<br>Robustness<br>Consistency | Variability<br>Reproducibility<br>Optimization<br>Stability | Definition<br>Understanding<br>Capability |
| Product | Biologicals<br>Biopharmaceuticals<br>Biotechnology products<br>Recombinant DNA products | API<br>Bulk<br>Drug substance | Protein |
| Scale | Manufacturing<br>Production<br>Commercial<br>Marketable<br>Full scale<br>Large scale | Clinical<br>Development<br>Pilot scale | Small scale<br>Laboratory scale |
| Exercise | Test<br>Study<br>Run<br>Experiment<br>Evaluation<br>Assessment<br>Investigation | Protocol<br>Procedure<br>Plan<br>Document | Variable<br>Parameter<br>Factors<br>Inputs/outputs |
| Result | Quality<br>Specifications<br>Requirements<br>Standards<br>Acceptance criteria | Change<br>Amendment<br>Corrective action | Fault<br>Failure<br>Hazard<br>Nonconformance<br>Excursion<br>Deviation<br>Atypical |

completion than to the initial process development and production of clinical supplies [3]. The actual cost of process validation is thought by some to be uncontrolled and excessive because there is no consensus on how much is enough [30]. One key risk associated with biopharmaceutical product manufacturing is the failure to gather sufficient information to support process decisions when faced with issues arising in process validation, failure investigations, product specifications, analytical methods, and equipment qualification [30].

Common pitfalls associated with process validation are numerous. They include a lack of overall strategy, insufficient planning/management, delay in early consultation with regulatory agencies, inadequate product definition, failure to observe current GMPs (cGMPs) and follow standard operating procedures (SOPs), poorly defined cell bank genealogy, inadequate analytical procedures, confusion over participant responsibilities, excessive process changes after process validation completion, insufficient prior process characterization or qualification, starting process validation too early and before process finalization, poor validation study design/

definition, disagreement on evaluation/interpretation of results, and inappropriate acceptance criteria [8, 14]. When overall validation activities begin too late, the work required to validate methods, facilities, and/or processes typically has been underestimated [4], placing product launch at risk. When processes perform erratically during process validation, further investigation to uncover inadequate development or lack of control of one or more process variables is required, followed by additional validation studies [12].

Validation must be science based [31]. Sound science during process validation is the primary vehicle to reduce manufacturing costs while remaining compliant with regulatory expectations [30]. Overly complex plans contradict desires to be efficient, lean, and cost-effective. More intensely planned projects produce more efficient, lower cost, and higher quality products in a faster time frame [30]. Resulting risks for biopharmaceutical companies from ineffective process validation include patient (safety and efficacy), operational (safety, contamination, variability), financial (product loss, reputation, legal costs), and regulatory (FDA 483s, warning letters, recalls) factors [11].

## 3 EVALUATING MEASURES OF PRODUCT QUALITY AND CONSISTENT PROCESS PERFORMANCE

A streamlined, integrated, and comprehensive validation process is desired [32]. The key to achieving a manageable process validation design is determining which process variables can be relegated to a noncritical status or maintained under tight control. Only when absolutely necessary should parameter ranges be validated. Thus development studies are required to identify critical parameters, limits to their operating ranges, and those process characteristics and equipment that enable their tight control [32]. It is necessary to distinguish between measures of product quality and process performance (consistency) and to differentiate between release specifications (such as critical quality attributes, CQA) that are measures of product quality and process attributes that demonstrate consistent process performance. Validation studies are required to examine process parameter variations that cause yield losses by affecting product quality (i.e., critical parameters); they are not needed for process parameter variations that cause yield losses that do not affect product quality even though these losses can cause economic/business concerns [10]. A comparison of process validation terminology for variables is shown by Table 3. A recent document has attempted to clarify these definitions further [33].

### 3.1 Critical Process Input Variables/Parameters

The basis of all process validation studies is the demonstration of control of critical process input parameters for reproducible operation of a commercial-scale production process [11, 32]. Critical parameters vary among processes so they must be separately assessed for each new process [10]. As a typical process has hundreds of variables, critical process input parameters must be identified first, and then the extended range and normal (target) operating range for each critical parameter must be determined during process development and/or using historical data [11,

**TABLE 3. Comparison of Process Validation Terminology for Variables (bold entries according to Ref. 33)**

| Term | Criticality | | Process Position | | Control | |
|---|---|---|---|---|---|---|
| | Critical | Noncritical | Input | Output | Controlled | Uncontrolled |
| **Operational parameters** | **X** | **X** | **X** | | **X** | |
| Input variables/ parameters | X | X | X | | X | X |
| Output variables/ parameters | X | X | | X | | X |
| Control parameters | X | X | X | | X | |
| Critical variables/ parameters | X | | X | | X | |
| Critical control points, CCP | X | | X | | X | |
| Key process input variables, KPIV | X | | X | | X | |
| **Key/non-key operational parameter** | | **X** | **X** | | **X** | |
| Xs (dependent) | X | X | X | | X | |
| Ys (independent) | X | X | | X | | X |
| Quality/performance attributes, CQA | X | | | X | | X |
| Acceptance criteria | X | | | X | | X |
| **Performance parameters** | **X** | **X** | | **X** | | **X** |
| Process attributes | | X | | X | | X |

34]. Critical variables are identified using defined procedures combining elements of data analysis and scientific judgment and making maximum use of available data gathered during process development [20]. Various companies have developed differing decision matrices to determine critical parameters, some of which are based on formal risk analysis exercises. A critical variable is defined as a parameter whose operating range lies near the edge of failure [34]. Using this definition, in practice, most identified critical variables can be engineered out of the study. Critical steps are those steps that are difficult to control because they usually contain at least one or more critical variables that cannot be removed [34].

Critical control points (CCPs) are those locations within a process where control can be applied to protect product quality, specifically to prevent and eliminate quality hazards or to reduce them to acceptable levels [11]. For controlling hazards, the appropriate CCP may be positioned in the sub-sequent step from the actual hazard location. In addition, more than one CCP can be used to control a single hazard and more than one hazard can be controlled by a single CCP [11].

Critical parameters are studied next to determine the effects of their outer limit values on important process characteristics [10]. These critical outer limits must be attainable, accurate, robust, and scientifically based [11]. The control parameter range is the span of values that lies between two outer limits or control levels of a

parameter, and it represents the highest and/or lowest values of a given control parameter that actually are evaluated during validation [25]. It is not usually necessary to test both ends of the control parameter range for all parameters because scientific judgment can help determine which end of the range (high or low) is more likely to generate an adverse effect [10]. In other cases, testing of both the upper and the lower control limits for a large number of variables is required to identify worst-case conditions [35]. Acceptable ranges are narrowed, and experiments are repeated until the acceptance criteria are met using an iterative process that requires substantial execution [3]. The edge of failure is reached when an exceeded control parameter value adversely affects the process or the product in that process performance degrades or product fails [20]; however, it is not necessary to validate the edge of failure when validating a process [25].

One strategy is to select worst-case conditions that encompass the upper and lower processing limits of critical input parameters, along with situations and circumstances over the entire process (rather than only a single step) that together result in the greatest chance of process or product failure compared with the ideal or target conditions [31, 34]. Worst-case runs are long and complex because the entire process must be run completely, usually at large scale or full scale [10, 20]. Because the worst-case approach comprehensively in-cludes the direct effects and actions of all variables, if all experiments are successful, only a few studies are required [20]. To mitigate the high likelihood of failure of a worst-case study that results in repeating lengthy runs using revised conditions, worst-case runs are performed after factorial experiments, usually conducted at a smaller scale and of a narrower scope, have confirmed a satisfactory outcome under the proposed worst-case conditions [20]. A key philosophical weak point of the worst-case run strategy is that it is less probable in practice that all critical parameters experience simultaneous excursions to outer limits [10]. However, because all critical parameters are tested at their extreme range, the worst-case approach can identify interactions that design of experiment (DOE) methods cannot detect [10].

Regardless of the strategy employed, worst-case conditions for critical parameters are established and process validation assures that the process yields an acceptable product when challenged under these worst-case conditions [30], even though full-scale validation runs may not have been conducted at many of these worst-case limits. Then operating ranges and control limits are set to establish a process robust enough to produce acceptable product [14, 35], and full-scale validation runs are executed to demonstrate performance consistency. The assumption is that if the product is of acceptable quality when parameters are controlled at the outer limits of control ranges, then the quality also will be acceptable when these parameters are held within those control ranges. This strategy is analogous to the approach of using fractional cycles for validation and full cycles for actual operation when testing at cleaning and sterilization procedures [36].

For a controlled variable, several types of ranges exist around the set point (Table 4). The normal operating range (NOR), or the alert limit, is established by trending performance during clinical material or qualification lot production, and then assigning usually two or three standard deviations to the set point [20]. The proven acceptable range (PAR) encompasses all values of a given control parameter that fall between proven high and low worst-case conditions [25] and constitutes the proposed operating range [3]. The PAR should be greater than or equal to the MOR [3]. The maximum operating range (MOR), or the action limit, is the

**TABLE 4. Ranges of Controlled Variables/Parameters in Order of Increasing Breadth**

| Controlled Variable/Parameter | Definition |
|---|---|
| Set point | Target |
| Normal operating range (NOR) | Alert limit, variations typically +/– 2 or 3 SDs from set point |
| Maximum operating range (MOR) | Action limit, product quality acceptable |
| Proven acceptable range (PAR) | Upper/lower worst-case conditions; outer or control limits/control parameter range; characterization, tolerance, or design space range; product quality acceptable |
| Limit/edge of failure | Unacceptable product quality, critical limit |

range of the controlled variable within which product quality is acceptable, but not necessarily at the limit of failure; thus, if the variable value falls outside the MOR, it is not necessarily a failure [20]. As MORs can lead to a decline in process performance even if product quality remains acceptable, their frequent occurrence can become an economic or operational issue. The MOR should be sufficiently broader than the NOR to reduce the frequency of unplanned deviations [20]. The MOR can be extended based on manufacturing experience when excursions outside the MOR occur without adverse quality effects [20]. Overall, a demonstrated state of control is the condition when all operating variables that can affect performance remain within such ranges that performance is consistent and as intended [25].

### 3.2 Noncritical Process Input Variables/Parameters

Noncritical or operational parameters are process control set points that define the process recipe and show that the process can be executed consistently [14], but for which there is no established evidence of product quality impact [11]. If the process performance is consistent when noncritical variables are controlled within their ranges and if minor deviations (e.g., instrumentation inaccuracy) from these ranges do not affect product quality, then the variable or step is classified as noncritical [34]. Typical noncritical control ranges are wide and/or the variable/parameter is easy to control within the range [33]. Although noncritical parameters do not require further study [10], they also are controlled and monitored during process validation to reduce variability and operator error [11]. Recently, noncritical process parameters have been designated as key (ensuring operational reliability and desired process performance when maintained within a narrow range, significant impact but not failure within range) and nonkey (well-controlled within a wide range but a potential process or quality impact when outside range, no significant impact within range) [33].

### 3.3 Process Control/Input Parameters (Xs)

Process control parameters are operating variables that have assigned values to be used for control [25]. They usually are readily maintainable away from a point of

failure and within a tested manufacturing range by a feedback control mechanism [15, 35]. Control parameters may also be thought of interchangeably as input parameters, defined as operational parameters that can be controlled [32]. Control/input variables are those operating parameters with set points or ranges that define process execution, e.g., pH, flowrate, temperature, raw material amounts and specifications, properties, and hold times for in-process streams of prior steps [3, 14]; thus, these parameters can be validated as a range or as set points [32]. They are a subset of operating variables, all factors that potentially affect process control or product quality [25], that reasonably are expected to affect operation of a process step [15]. Control/input parameters also are referred to as Xs, analogous to independent variables [37].

All input parameters must be identified and demonstrated to be in control for the product to be in a validated state and the more tightly the range is controlled, the less critical the parameter is considered [32]. Validation steps are (1) identification; (2) classification as critical or noncritical, subclassification of critical parameters into those to be validated over a broad range and those to be validated as a set point with a narrow tolerance range; (3) validation execution for those critical input parameters with a broad range; and (4) validation execution using target set points for all process input parameters both critical and noncritical. The complexity of process validation is dictated by several critical process input parameters requiring either a broad range or a broad set point. As fewer range testing studies are required if more parameters can be tightly controlled, one strategy is to design tight control set points for as many critical process input parameters as possible (e.g., pH and temperature) [32].

### 3.4 Performance Uncontrolled Output Variables/Attributes (Ys)

Performance parameters reflect the outcome of a given step and indicate that the process gave the desired result [14] or quality attribute. They are uncontrolled performance variables [15] without a control action [35]. Their natural variation is defined by operating history; specifically, their variability is characterized from known historical data or estimated based on similar process performance [35]. Similarly, output variables reflect the step outcome and indicate performance was acceptable in terms of performance attributes for the step (e.g., titer and yield) or properties of the product stream (e.g., product homogeneity, purity, contaminant levels, and chromatography peak shape) [3, 14]. Still another term used is critical Ys (analogous to dependent variables), defined as product and process output variables that relate to critical quality attributes (CQAs), which are measurable outputs of each process step that are used to provide evidence that the step performed correctly [37].

An input parameter value is directly related to an output parameter or quality attribute; specifically, the output parameter of one unit operation often is the input parameter for the subsequent unit operation [32]. Equivalently stated, critical Ys (outputs or responses) from one stage often are inputs for next the stage [37]. Input parameters are assessed as to whether they significantly impact critical output parameters (e.g., product safety, identity, and efficacy), although the definition of significant varies among applications [32]. As all product quality attributes are not routinely measured during process development, some output variables are selected

based on a hypothesis that they are linked to desirable product characteristics [20]. An input parameter is determined as critical based on whether its operating point is located near the edge of failure and how well it can be controlled [32]. Input parameters then are prioritized based on the (1) output sensitivity to input variations, (2) step proximity to the final product, and (3) technical difficulty controlling the input within the specified range [32].

A more formal approach involves listing all key process input variables (KPIVs), factors [both process (physical) and product (chemistry)] that could potentially affect the mean and standard deviation of critical Ys (outputs) [37]. KPIVs are identified using brainstorming to generate initial information. Then the identified factors are reviewed to select those that should be maintained constant based on scientific judgment. Next, a review of process development data and additional experiments using DOE are used to identify those KPIVs which influence critical Ys [37]. Experimental results are inserted into a cause and effect diagram which statistically correlates input and output variables by constructing scatter plots [20].

## 4 BACKGROUND AND REQUIREMENTS

### 4.1 Types of Process Validation

All three types of validation, prospective, concurrent, and retrospective, are part of most comprehensive process validation programs that occur throughout a product's lifetime [3]. As a process approaches commercial status, validation approaches completion [25].

***Prospective.*** Prospective validation is validation conducted before the distribution of either a new product or a product made under a revised manufacturing process where revisions may affect the product characteristics [37]. Prospective validation assures that process quality attributes are met before manufacturing operations have commenced [26]. It is based on a preplanned protocol [25]. Usually process validation is prospective [18] because a science-based, prospective validation is integrated as a continuous part of the development process [36]. Validation performed prospectively is relatively less dependent on in-process or endproduct testing [25]. Manufacturing lots can be released faster if prospective rather than concurrent validation is performed [26].

***Concurrent.*** Concurrent validation is based on information generated during the actual implementation of the process [25], at the same time the marketed product is being manufactured [26]. Concurrent validation reduces the load at the time of regulatory submission because experiments are delayed until the manufacturing process is operating, but it places marketable lots at risk if a failure is detected. Until process validation (e.g., useful resin lifetime) can be established at the manufacturing scale via concurrent validation, a thorough performance analysis of the step (e.g., chromatography) must be undertaken for each manufactured lot, which is quarantined until the results are evaluated [26]. Thus, concurrent validation is based on batchwise process control and requires the use of validated in-process sampling and methods [25].

***Retrospective.*** Retrospective validation is based on the review and analysis of historical information (e.g., batch records, in-process control testing, and stability testing) to reconfirm formally that control parameter ranges are appropriate [25]. Validation protocols sometimes are not needed, but the final results are formally approved. As with concurrent validation, retrospective validation requires batchwise control of the process and usually depends on in-process testing [25]. Retrospective validation has been found useful to augment (but not replace) premarket prospective validation for new products or changed processes [3]. The proven acceptable range (PAR) can be developed using the principles of retrospective validation to link acceptable product lots to normal/permissible values for process operating parameters [12].

Retrospective analysis applies to a large number of batches that have been prepared similarly by evaluating input/output parameter variation and relating it directly to batch product quality [10]. Specifically, it is most effective when many similar batches (e.g., tens to hundreds) exist. However, the observed range of some input parameters may be insufficient if they were tightly controlled during processing to effectively determine the impact of wider variations [10]. In cases where disparate historical data sources may not be directly comparable or missing, a screening DOE might be appropriate to establish that existing specifications maintain CQAs at desirable levels [37].

### 4.2 Aspects of Process Validations

Process validation starts with the identification of product quality attributes and justification of acceptance criteria, followed by a review of the risk analysis, execution of process development runs, and compilation of clinical material manufacturing data to set specifications considering process variability [11]. There is a greater focus on process validation for downstream steps rather than for upstream steps because downstream steps are associated with virus removal. Process validation is just one approach used to control virus contamination, however; others include cell bank characterization, in-process testing, inactivation procedures, control of raw materials, containment, and postmarket surveillance [6].

One significant source of variability in downstream process steps is changes in input streams between research/pilot-scale to full-scale [38]. Thus, input streams of high and low quality are used when evaluating the performance of the subsequent step in a pair-wise validation strategy [3]. Specifically, forward linkage studies are performed to understand downstream consequences for critical parameters [10]. When process steps are studied individually in a factorial approach, forward linkage variables are the output variables of given step that affect the performance of the next step [20].

***Fermentation.*** A validatable fermentation process demonstrates a controllable method of growing cells that reliably express the biopharmaceutical product that can be reproducibly recovered if broth is harvested within a set of specified conditions that allow the product to meet its quality specifications [10]. The goals of fermentation process validation are to provide documented evidence that all aspects of the process perform as intended to generate consistent fermentation broths at harvest; this consistency in turn permits downstream steps to yield a

purified product consistently meeting its predetermined specifications [39]. Allowable ranges for culture conditions must be justified with scientific data [30]. Typical components of fermentation process validation include the fermenter sterile medium hold, inoculum train, fermenter inoculation, sampling, feeding, fermenter growth and production phases, and harvesting [9]. Specifically, the trigger for fermentation harvest (e.g., cultivation time, cell density, and viability) should be demonstrated to yield acceptable product [30]. CQAs defined for fermentation include product yield, cell density, and culture purity; critical process parameters include inoculum state, nutrient feeding scheme, and harvest time [40]. To ensure consistent broth quality, critical process parameters are best defined based on a culture physiological event rather than on elapsed cultivation time and are preferably those that exhibit distinctive online monitoring characteristics [40].

To link fermentation process input parameters to output parameters, specifically behavior in downstream isolation steps and ultimate product quality, usually a partial or total purification is necessary, which requires substantial labor to process samples [10]. Sample evaluation also is limited depending on the level of product characterization. Consequently, fermentation process validation has been viewed as less rigorous compared with other forms of process validation [10].

### *Viral Clearance Studies*

*Requirements.* Any biopharmaceutical product using animal-derived materials during the manufacturing process has the potential for virus contamination [5]. The risk varies depending on the animal species and country of origin of the raw material [5] and the safety concerns associated with expression and amplification of the production system [13]. Process validation is conducted to determine the process capacity to remove and/or inactivate a virus using a wide variety of viruses in virus clearance studies. These studies complement analytical viral testing of product (at various stages of production, including the final product), starting materials (e.g., cell banks), and raw materials [5, 6]. Each process validation viral clearance study is reviewed individually to determine the log reduction performance versus requirements, which are set based on experimental limitations as well as on risk factors (i.e., number of patients and dosage) [5, 41]. In practice, viral inactivation testing is less rigorous than microbial sterilization testing, viral test methods generally are cumbersome with variable results (although PCR technology has improved technical robustness), and viral inactivating agents less robust than saturated steam [10]. Consequently, viral clearance studies are a substantial component of process validation.

Viral clearance can be demonstrated by two distinct mechanisms that both generate high clearance values. One mechanism is inactivation to reduce infectivity by chemical or physical means without affecting protein product stability. Methods include using pH (acidic, low pH conditions for pH inactivation for protein A, basic, high pH conditions for ion exchange resins), chaotropic agents (urea), ultraviolet irradiation, and solvent, detergent, or heat treatment [42–45]. The second mechanism is removal or partitioning to effect a physical separation of virus from the product. Methods include using chromatography (adsorbing protein and leaving virus in the column flowthrough or vice versa) and filtration (depth filters with a 20–40-nm pore size or ultra-filtration membranes with a cutoff of <300 kDa) [5, 7, 46–49].

Processing conditions significantly affect virus inactivation/removal performance. Specifically pH, buffer composition, protein concentration, flow rate, and product binding strength to resin influence chromatography steps [5]. Membrane chemistry and pore size, geometry, velocity and pressure of process streams, transmembrane pressure, and product load per area (volume/area, concentration/area) influence filtration steps [47]. When virus-inactivating buffers are used in chromatography, virus reduction is effected in the eluted product from a combination of removal and inactivation [46]. Analytical methods that can evaluate infectious as well as inactivated virus distribution in process streams can assist in understanding virus removal mechanisms [46].

*Steps to be Validated for Viral Clearance.* Process validation for viral clearance is conducted only on robust steps that can (1) be scaled down accurately and (2) reproducibly and effectively remove and/or inactivate a wide variety of potential viral contaminants under a wide variety of process conditions [5, 41]. The number of steps selected for validation depends on estimated viral clearance effectiveness based on historical data and target clearance values [5]. The FDA demands at least two different steps for virus reduction to guarantee safety and efficacy [7]. Potentially only two steps are required for antibody processes that use serum-free medium, but additional steps might be required if viral contamination risk is increased by using serum-containing medium [7]. Due to the use of live viruses to perform clearance studies, this work usually is outsourced to reduce cross-contamination issues [14].

The viral clearance reduction factor, the common logarithm of the ratio between the total virus loads before and after clearance, is established for viruses known to contaminate the production process [50]. Individual step clearances are combined to obtain the total clearance reduction factor. This reduction factor is used in combination with an assessment of step robustness to classify the step as effective ($\geq 4$ reduction factor and unaffected by small changes in process variables), moderately effective ($4 >$ reduction factor $> 1$), or ineffective ($\leq 1$ reduction factor) with respect to virus clearance [50]. Clearance factors are usually multiplied if the mechanism is different for two separate steps and sometimes are added if the mechanism is same [51]. In other cases, if two independent steps have similar mechanisms of clearance, only one step is included in the summation because virus particles removed via that mechanism would only be expected to be removed in the first step [3, 5]. A total clearance of 12–15 logs is desired for lipid-enveloped viruses and fewer logs for nonenveloped viruses (e.g., polio) [30].

Virus loads in unprocessed, prepurified bulk (harvest samples) are quantified, typically via transmission by electron microscopy or infectivity, to estimate actual virus load versus the expected virus removal capacity of the process [47, 50]. Total virus clearance should exceed measured levels by at least 3–5 logs [47]. An example of a typical virus load is $10^5$ to $10^7$ RVLPs (retrovirus-like particles) per milliliter of unprocessed bulk from CHO cells [3]. Clearance factors of 15–20 logs are typical for murine retroviruses [3] and slightly more than for lipid-enveloped viruses. For most viruses besides endogenous retroviruses, however, there is no reasonable way to establish virus load in the bulk because these types of viruses should not normally be present. In addition direct testing methods are limited inherently because they are designed to detect only known and not known contaminants [5]. Thus, the

capacity of the downstream purification process to remove/inactivate known viruses should they be inadvertently introduced is defined for an extra degree of assurance [3, 5].

*Spiking Studies.* Often, if viruses are present in process streams, they are present at levels below analytical detection, so spiking-clearance studies are required for purification steps [3]. Spiking studies are repeated so that their reproducibility and variability can be assessed [5]. There are concerns that large spikes may not be representative of actual low-level virus behavior, however. Virus spikes are desired to contain as little protein as possible to reduce the impact of their addition on purification procedures [5]. Addition of a virus spike can alter protein concentration and viscosity, which can change elution characteristics and prevent the small-scale model from accurately predicting full-scale performance [52]. In practice, this goal is more readily achievable for non-lipid-enveloped viruses because enveloped viruses need a certain amount of protein to maintain stability [5]. Maintaining the salt concentration constant or evaluating product resolution and purity are two ways to demonstrate that the virus spike addition had no significant effect [52]. Consequently, there is a compromise between adding the highest amount of virus possible to more rigorously challenge and maximize the step clearance and maintaining the spike volume low (<10%) to avoid appreciable alteration of product composition and thus purification performance [5, 50].

*Virus Selection.* The goal of model virus selection is to demonstrate inactivation or removal of as many types of viruses as possible without compromising product activity and quality [50]. One virus can act as a model for a group of viruses with similar physio-chemical characteristics [5]. Model viruses should be possible to grow in cell culture to high titers and detectable using a simple but sensitive assay [47, 53]. Some viruses that are known to potentially contaminate starting materials grow quite poorly in cell culture; thus, some model viruses have been tissue-culture adapted and consequently may not be truly representative of wild-type viruses [41]. Another limitation of using model virus clearance studies is that model viruses may not behave similarly to actual viruses [5], but in many cases, model viruses are the only practical alternative.

Viral selection is based on (1) relevant viruses that are actual viruses (or of the same species as actual viruses and relevant to the host cell) that have been identified as contaminants (or potential contaminants) of the process, (2) specific model viruses that are closely related to actual viruses (e.g., same genus or family) and have similar physico-chemical properties, and (3) nonspecific model viruses believed to be representative of the spectrum of different virus physio-chemical characteristics [5, 50]. Nonspecific model viruses are used to show inactivation/removal of viruses in general and to characterize purification robustness [50]. Virus clearance studies should cover emerging viruses and viruses currently believed to be absent in raw materials. These concerns are not addressed when relying on direct testing to ensure safety, specifically consideration of future virus removal requirements in anticipation of future regulatory changes [5, 41].

Model viruses are selected that cover the range of physico-chemical properties of different virus species such as size (i.e., diameter and geometry), enveloped or nonenveloped, RNA or DNA, and single- or double-stranded genome [47, 51, 53].

Examples of model viruses selected for various characteristics are medium-to-large DNA/enveloped (Herpes Simplex I, pseudorabies), small/DNA/nonenveloped (Simian virus SV-40), small/RNA/nonenveloped (Sabin Type I Polio, animal parvovirus), medium-to-large RNA/enveloped (Influenza Type A, parainfluenza virus, Sinbis virus 1), and retrovirus/RNA/enveloped for murine hybridomas (Moloney murine leukemia) [3, 51]. Preference in the selection of specific model viruses is given to those viruses with significant resistance to physical removal/chemical agents [47, 49, 51].

The number and type of viruses selected for spiking studies depends on the process materials (e.g., host cell) and product development stage [5]. Before phase I, virus validation must be conducted. Typical viruses studied are the murine leukemia virus as a model virus for retrovirus-like particles produced by murine cell expression systems (if applicable to the process) and parvovirus [50]. Before phase II, two to four additional viruses are selected that depend on the host cell and media origin [50]. It is advised to solicit feedback on virus selection from the relevant regulatory agencies depending on where the trials are to be conducted before conducting viral clearance studies for product destined for initial clinical trials [14].

***Starting and Raw Materials.*** Starting and raw materials are chemical, biochemical, or biological components used in a biopharmaceutical manu-facturing process [53]. Quality control strategies for these materials have been developed based on the assumption that these materials may be contaminated with a virus [54]. Direct testing is limited by the fact that viral contamination can be nonhomogenously distributed in a raw material lot; thus, additional sampling has little value. In addition to being effective only if contaminants are homogeneously distributed, direct viral testing also only is effective if the contaminant is present at high titer, false positives can be investigated promptly, a broad spectrum of viruses are detectable by available assays, assays are adequately sensitive, and raw materials do not inhibit assay performance [54].

*Cell Substrate.* Since the genetically altered cell substrate is considered the main component in the production process, complete characterization plus evaluation of genetic stability are primary concerns [27]. Cell bank characterization in conjunction with product characterization (e.g., peptide mapping and amino acid sequencing) demonstrates the stability of the production process [55]. Cell bank characterization in conjunction with process validation and endproduct testing demonstrate product safety [55]. The most likely source of viral contaminations in a large-scale cell culture is the cell substrate [54]. Process validation plus cell bank validation provides assurance that final product is free of contaminating viruses and other adventitious agents [55]. There is little concern about viruses that do not replicate in the cell substrate [54], although nonreplication can be hard to demonstrate.

As one of the most critical raw materials in production is the cell bank, defined acceptance/rejection criteria and parameters under which the bank will be used are required [55]. Assays are performed on master and working cell banks to demonstrate stability and freedom from adventitious agents. Cells are analyzed both from master/working bank vials as well as from the end of the production run or a few doublings later [55].

During cell bank characterization, phenotypic characteristics of the master cell bank are established such as morphology, doubling time, product expression rates, karyology (for diploid cell lines), and isoenzyme analysis [55]. Genotypic characteristics also are established, including a description of the vector and inserted gene, copy number, restriction maps, and the nucleotide sequence of a cloned gene. Tests are conducted for adventitious agents such as bacteria, fungi, mycoplasmas, and a wide range of potential viruses (e.g., retroviruses) [55]. Routine stability programs are undertaken to monitor genetic stability with respect to phenotype, expression, and nucleic acid integrity [4]. A passage history (e.g., split ratio, storage conditions, and media used) is documented as well as the cell line origin [55]. As a result of these regulatory requirements, cell banks must be prepared by those well versed in the technical processing steps, but also well aware of regulatory requirements.

*Media.* Basal medium and other supplemental components are tested in a scaled-down mimic of the production process to determine cell growth and product yield, and the product characteristics are compared against reference materials [16]. It is necessary to assure that multiple lots of raw materials meet user specifications for performance and quality before selecting a vendor [52] and that at least one back-up vendor is available.

***Step Hold Times.*** Documentation of bioburden and endotoxin control at the manufacturing scale over the entire hold time is accomplished via monitoring of intermediates using in-process samples [33]. In addition, small-scale studies can provide relevant worst-case data. The cumulative worst-case hold times are obtainable from the summation of the maximum individual unit operation hold times. However, this worst case is not required to be validated because such a series of unexpected events is likely to cause adverse drug quality as well as high endotoxin or bioburden. Instead, hold time studies are conducted for process solutions over the maximum durations that they will be held in tanks used for manufacturing [33].

Real-time stability data are required for reliably assigning expiry periods [4]. Stability data are obtained on the drug product stored in the proposed container under recommended storage temperatures [56], and this strategy has been applied for in-process streams. Although elevated temperature stability studies are used to accelerate product degradation rate to predict shelf-life [56], this acceleration methodology is usually not used to determine step hold times, which typically are a few days, weeks, or in rarer cases months. Additional stability and degradation concerns arise from the possible interaction of protein with stoppers and container closure systems and interaction of proteins with residual cleaning agents [4].

***Contaminant/Impurity Clearance.*** Classes of contaminant removal include process-related, host-cell related, and product-related (e.g., aggregates, deaminations, and oxidations) substances [14]. Contaminant clearance studies can be used to avoid lot-to-lot acceptance criteria because safety margins of several logs of excess clearance are demonstrated [33]. In these studies, a contaminant is added to the input feed stream at the small scale, its recovery is measured at each stage of the

process step (e.g., column flowthrough, product pool, and regeneration fractions), and mass balance measurements are performed [51]. In a similar manner as viral clearance studies, each step's impurity clearance is challenged separately; then overall clearance is calculated by the product of individual clearances [51]. Impurity clearance studies are similar in concept to viral clearance studies without many of the virus handling and personnel contact safety issues.

Process-related contaminants include cell substrate-derived (e.g., host cell protein, DNA, retroviral-like particles, and endotoxin for *E. coli* cultures), cell culture-derived (e.g., inducers, antibiotics, and serum/media components such as antifoam), downstream-derived (e.g., enzymes, processing reagents such as glycerol and guanidine, inorganic salts, solvents, leachables, and ligands), and adventitious agents (e.g., virus, bioburden, and endotoxin) [33]. Cell substrate-derived and adventitious agents have been discussed. For many process-related contaminants (e.g., antifoams and enzymes), the key to tracking and demonstrating their removal is the development of a robust, sensitive, and yet straightforward assay.

A documented methodology is required for how product-contact materials are selected and purchased [57] for specific process use. Leachate process validation demonstrates that leachates are below limits of detection in each eluent pool; insignificant leaching occurs under severe chemical, time, and temperature exposure, and if significant leaching does occur, it is removed by the first few bed volumes of equilibration buffer [58]. Product extractable testing evaluates product contact surfaces in terms of materials of construction, process solutions or solvents, duration of product contact with various surfaces, product contact surface areas, and potential for extractables at various process stages [57, 59]. Materials are prioritized to direct studies at the most critical, high-risk materials. High-risk materials are based on the proximity to the final API, extraction capability of solvent, length of contact, product contact surface area, cytotoxicity of extractables, temperature, and inherent material resistance to extraction [57]. Example high-priority materials to assess are containers (e.g., stainless steel cans and plastic bags) for unfrozen and frozen storage of fermentation media/purification buffers/product intermediate streams, chromatography resins, valve diaphragms/transfer hoses, elastomers/pumps especially on the final purification skid, filtration membranes, final sterilizing filter, gaskets/o-rings especially on the final bulk tank, and silicone boot/tubing for the filling machine [57]. Material extractable data can be obtained in-house, or often it is available from the vendor.

Product-related impurities include truncated product forms, modified product forms (e.g., deaminated, isomerized, disulfide-linked, and oxidized), and product aggregates [33]. One way these undesired forms arise is during freezing and thawing owing to product concentration and pH gradients [60]; thus, the number of freeze/thaw cycles is minimized. Another way is from mixing liquids and dissoluting powders [33]. The minimum mixing time until homogeneity is determined for which results vary only within assay variability; maximum mixing times are established to avoid shear, oxidation, and/or product degradation.

***Reuse of Chromatography Resin.*** It is necessary to demonstrate that chromatographic media can be reused, cleaned, and sanitized [61]. Used resins receive additional analysis not required with new resins such as the titration of small ion binding capacity, measurement of total protein capacity, comparison of flow versus

pressure plots to indicate particle size (and attrition), particle size distribution, total organic carbon (TOC) removal by cleaning solutions, and microbial contamination/endotoxin (LAL) analysis [58]. Exposure to cleaning/regeneration solutions, rather than contact with mild buffers and protein solutions during normal processing, is most likely to cause chemical degradation [58]. Three types of studies thus apply to resins: Column reuse studies verify continued resin performance after multiple uses followed by small-scale carry-over runs to confirm cleanability with multiple uses, characterization studies prove the process is robust and defines acceptable operational limits, and another third study type demonstrates that regeneration sufficiently sanitizes the resin and that the storage solution is bacteriostatic and fungistatic [61].

Columns reused for many cycles may not provide the same virus clearance as new columns [41]. Thus, virus spiking studies can be performed on new and aged resins to ensure clearance does not vary [5, 26]. Alternatively, virus clearance can be studied for new resins only, and performance attributes expected to decay before virus clearance ability can be monitored subsequently during actual reuse [62, 63]. Validation of the column resin cleaning and sanitization regime for virus inactivation also is necessary to demonstrate that bound virus does not accumulate and subsequently elute [41]. The appropriate number of reuses is selected by balancing resin cost with the time required to conduct the reuse study and the expected annual number of production runs [61]. Using resin to its functional lifetime maximizes its cost basis, which helps economics if the resin cost is high or the number of cycles is low [26]. Virus removal filters [64] also have the potential to be reused, and although less, if anything, has been published about their reuse, the performance characteristics necessary to demonstrate are expected to be similar to resin reuse.

### 4.3 Process Validation Execution Timing

The execution timing of key process validation elements varies according to the clinical phase of the product. As validation starts well before the last three lots before actual production [31], it is essential to consider the question of "how much process validation to do and when to do it" as early as possible [14]. Creating the validation package too early might result in redoing substantial portions [14]; creating it too late may delay product launch owing to an incomplete regulatory package. Process validation is a continual process and should be monitored closely for adherence to timelines, scope, and cost [4]. It evolves along with the development of the process and should be complete at the end of phase III [14].

At the preclinical product phase, critical and noncritical classification of process input parameters should be initiated [32]. Critical components of facility subsystem validation need to be essentially complete before phase I product manufacture [15]. For phase I, it is necessary to validate aspects of the process related to product safety (e.g., sterility, mycoplasma, viral clearance, impurity removal, and stability) [14]. Abbreviated viral clearance studies for model viruses/retroviruses and impurity clearance studies for host cell DNA often are acceptable, resulting in fewer downstream steps validated at this product stage [3, 5]. If viral clearance results are available in sufficient time, the results can be applied to developing the phase I process steps. All assays do not have to be validated at this stage, but some (especially product-specific ones) should be at least qualified [14].

For phase II, no specific process validation activity is required unless process improvement changes potentially affect prior phase I validation work [14]. Assay validation efforts should continue, however, so assays are ready for phase III validation and process development should be finalized before phase III [14]. Most process and assay validation activities begin in phase II and often extend far into phase III [4], occurring in parallel with phase II and phase III clinical material production [3]. Intensive process characterization often is delayed until after completion of phase II studies to conserve resources [65]. About 12–15 months are allowed for its completion before the process validation runs [65]. To minimize risk, if all phase II clinical data are not available, these activities can be ramped up slowly over the first few months. Raw material vendor audits should be performed between phases II and III before the manufacturing process is fixed [47].

Before phase III and process finalization, process robustness is evaluated via several scaled-down runs to quantify process variability [14]. These studies form the basis for the process validation program if they are well documented, use operational range extremes, and test different resin/raw material lots. Based on the resulting data, additional optimization to improve robustness is conducted. Fully developed analytical assays are employed during these pre-phase III scale-down runs to ensure the process meets the quality acceptance criteria. Then during phase III, any remaining assays are validated [14].

The heavy-duty process validation occurs in phase III when the process is better understood, a commercial process is in place, and there are likely to be fewer changes, adjustments, and mistakes that can cause deviations [14]. First equipment and other subsystem validation are completed and followed by process validation during three or more consecutive batches [14]. The master validation plan usually is approved during phase III clinical material manufacture [32], and it sometimes includes manufacturing equipment validation in addition to process validation [47]. Viral clearance studies are conducted with a greater number of model viruses possessing a greater range of virus characteristics, and the full range of impurity clearance and operating range validations is undertaken [3]. Mass balances are attempted, duplicate runs are tested, and resin lifetime studies are conducted [14]. Range testing validations typically are completed during phase III clinical production before the manufacture of qualification lots, except for viral clearance studies that are done earlier [32]. Thus, the pivotal phase III clinical studies are conducted using a fixed process with refinements already incorporated to achieve a robust and cost-effective process [15]. At this stage, operational parameters are defined and performance parameters are specified within wide limits until more extensive data are accumulated [15].

Process performance qualification typically is conducted during the production of qualification lots using target set points for all process input parameters to demonstrate consistency and reproducibility of input and output parameters [32]. This process consistency is demonstrated by multiple full-scale batches, for example, three-to-five consecutive purification lots from three consecutive fermentation lots, during which yield and product concentration are monitored at each step [14]. A failed lot can occur without resetting the count if it is due to a known assignable cause that is not process-related (e.g., operator error or equipment failure) [33]. The cause of noncompliance must be determined and corrected, the noncompliant product disposition designated, and any corrective actions taken recorded [11]. At

this stage, all assays are validated, all equipment and support systems are qualified, all personnel are trained, all critical input parameters with ranges are validated, and all critical input parameters with set points and noncritical input parameters are validated as a by-product of range-finding studies [32].

## 5 DEVELOPMENT SUPPORT AND TECHNOLOGY TRANSFER ACTIVITIES

### 5.1 Process Development

***Role, Stages, and Plans.*** Process validation is an integral part of process development [13]. At the early stage of process development, the process is basic and unpolished [14] with purity ill defined and poorly controlled [2]. At the mid-stage, an optimized and more robust process exists, and by the end stage, the process is fully characterized, validated, and qualified [14]. During process development and scale-up, critical variables either are removed or else highly understood to develop a robust process [34]. Identification should start early and be based on literature knowledge, prior experience, or specific product information [32]. Some critical variables are identified via close in-process monitoring and thorough investigation when unexpected results are obtained [34]. Early, reliable, and accurate measurement of process variability increases the probability of operating and maintaining a robust, well-understood, and controlled process [66].

Successful process development establishes a clear logic path that streamlines subsequent validation design and testing [22, 32]. Process development serves to establish evidence that all key process control parameters and all control parameter ranges are validated and optimized to meet acceptance criteria before process validation [14, 25]. These acceptance criteria, specifically the critical process output parameters for scalability studies and range testing validations, in turn are derived from data obtained during development, preclinical, and clinical material production [32].

The greatest validation problems surface when insufficient resources and time are budgeted for understanding and optimizing production [14]. Development personnel require sufficient time and financial support for thorough validation and technical transfer activities, specifically remaining involved in process validation activities along with manufacturing personnel [2]. Consequently, development staff should understand compliance issues [13] to accurately anticipate the appropriate level of quality assurance applicable to process development studies [30]. Actual formal product development procedures have been recommended to ensure consistency [13], as well as capture best practices.

***Process Development Documentation.*** As a considerable amount of relevant data is generated during process development, its documentation must be sufficient to support the selection of manufacturing steps [2]. Development work and decisions made during development are captured in development reports that support future regulatory inquiries [14] and demonstrate that regulatory requirements have been considered and addressed [30]. Development reports should highlight potential

weaknesses in the product and process design with the goal of suggesting which ones should undergo subsequent development and validation [13]. They also include an analysis of critical process parameters [13]. The first reports can be generated as early as before the initial product transfer to the process development group [32]. Additional reports are generated after the end of process development and before the start of process validation [32]. All of these reports need to be compiled and indexed so they are readily retrievable [13], perhaps stored in a database for easy access [14]. They serve as a resource for the product's process development history if an employee leaves (or is reassigned) or as training for new employees [14].

An upfront listing of the titles and scope of expected reports felt to be necessary to support process scale-up, validation, and regulatory filings is created and reviewed within the multidisciplinary validation team [30]. This listing helps identify required studies and form a process development plan that can facilitate future tracking of executed studies. Once the required component tasks and their execution order are established, this plan should be reviewed semiannually, adding accuracy and sophistication over time. Formal evaluation of the status of the process development plan demonstrates whether the project is on track and identifies deviations in time and cost. The interim goals, resources, and milestones listed in the basic plan focus efforts on priorities, permit staff to see how tasks are integrated, and reduce the likelihood of running out of time [30].

***Process Characterization.*** The terms *process characterization*, *robustness*, and *validation* are sometimes used interchangeably, and consequently, the distinction among these terms consequently is blurred. Process characteriza-tion is more encompassing than process validation because process validation uses only a subset of the data generated from process characterization [15]. Characterization studies are similar in nature to validation studies except with less formal compliance to regulatory guidelines [12]. It is the part of the process development examining manufacturing ranges, robustness, and the edge of failure for a limited number of critical parameters [67]. For characterization studies, established scientific and engineering principles are used, along with methods that are not validated in the regulatory sense but are qualified scientifically with appropriate controls and written procedures. [12]. Typical causes of unexpected results in process characterization are inadequate procedures, or inadequate following of adequate procedures, and uncontrolled variables [37]. Process robustness provides assurance that the process will not fail when operating within defined control limits for key process variables [35]. This definition of robustness has been adapted from the ICH definition of analytical assay robustness (ICH-Validation of analytical methods, 1993), not from the guidelines for process validation [35].

Process characterization defines process capability and facilitates prospective process validation at the production scale [40]. Full process characterization is valuable in maintaining smooth manufacturing operations and minimizing lost batches, and it provides supporting information for lot release justification for atypical batches [15]. Its goals are to identify key operating and performance parameters, define control limits for key process parameters, demonstrate robustness of the commercial process, and provide technical information about the process [68]. The steps involved in process characterization include risk assess-

ments, scale-down model qualification, single-parameter ranging studies, parameter interaction studies, and worst-case studies [68]. Process characterization is conducted initially during the preclinical through phase III product stages and continues during the ongoing commercial manufacturing product stage [65].

The overall goals of process characterization are to (1) understand the roles of each process step, (2) become aware of the effect of process inputs (operating parameters) on process outputs (performance parameters) and identify key operating and performance parameters, (3) assure that the process delivers consistent product yields and purity in all operating ranges, and (4) develop acceptance criteria for in-process performance parameters [65]. Example operating and performance parameters for various process steps are shown in Table 5. Before beginning process characterization activities, advance preparation consists of examining available data and assessing risk, qualifying scale-down models, and developing process characterization plans. Execution involves four steps to achieving these process characterization goals: (1) characterize process performance ("process fingerprinting data") using screening experiments to eliminate additional parameters from further characterization studies, (2) identify interactions between key parameters, (3) establish process redundancy by determining the effect of feed quality on each unit operation, and (4) develop unit operation reports to document key operating and performance parameters and their respective ranges [65].

One recommended strategy for process characterization is to (1) measure as much as possible early in process development, (2) determine if changes in the

**TABLE 5. Example Operating and Performance Parameters for Various Process Steps**

| Process Stage | Operating Parameter | Performance Parameter |
|---|---|---|
| Seed (shake flasks or fermenters) | Agitation rate | Time to reach transfer cell density |
| | Working volume | |
| | Vial thawing conditions | |
| | Growth temperature | |
| | Number of generations | |
| Production fermentation | $DO/dCO_2$ | Titer |
| | Broth pH | Specific productivity |
| | Growth temperature | Plasmid loss/genetic stability |
| | Agitation rate | Growth rate/doubling time |
| | Aeration rate | Peak growth cell density |
| | Back-pressure | Harvest cell density/viability |
| | Working volume | |
| | Percentage inoculum | Product quality (glycoform and/or sialylation distribution) |
| | Transfer cell density for inoculum | |
| | | Bioburden/adventitious agents |
| | Broth pH | Feed/acid/base usage |
| | Point of nutrient feed initiation | Nutrient and waste by-product metabolism |
| | Nutrient feeding rate | |
| | Osmolarity | |
| | Point of induction | Time to induction |

**TABLE 5.** *Continued*

| Process Stage | Operating Parameter | Performance Parameter |
|---|---|---|
| | Concentration of inducer | |
| Filtration | Transmembrane pressure | Process flux and water flux recovery |
| | Feed/recyle/permeate flowrate | Cell lysis |
| | Membrane area/liquid volume or mass | Permeate turbidity |
| | Membrane geometry | |
| | Volume concentration factor/diafiltration volumes | |
| | Protein concentration | Yield |
| | Cell density/viability | Purity |
| Centrifugation | Centrifugal force | Cell lysis |
| | Feed/centrate flowrate | Impurity clearance (host cell proteins, nucleic acids, endotoxins) |
| | Feed/centrate solids content | |
| | Back-pressure | Centrate turbidity |
| Chromatography | Column pressure | Yield |
| | Column flowrate | Product concentration |
| | Column loading level | Purity |
| | Column load/wash/elution conditions (pH, conductivity, temperature, gradient slope, volume) | Impurity clearance (host cell proteins, nucleic acids, endotoxins) |
| | | Viral clearance |
| | Buffer volumes | Ligand leakage |
| | | Column HETP |
| | | Peak shape |
| Precipitation | pH | Yield |
| | Conductivity | Product concentration |
| | Precipitating agent concentration | Purity |
| | Precipitate agent addition rate | Impurity |
| | Temperature | Precipitate yield |
| | Protein concentration | Supernatant turbidity |
| General | Pre- and/or poststerilization media/solution pH | Growth rate/doubling time |
| | | Specific productivity |
| | Media/solution sterilization conditions (time, temperature, filtration) | |

magnitude of each variable/parameter are relevant to process (or process model) performance, (3) select the relevant variables to be controlled and documented, and (4) collect, organize, and archive all raw data [69]. As the total number of factors to be tested can be very large, factors are combined and treated as single variables based on the experience gained during process development [60]. Variables expected to have similar effects are linked so that their cumulative effect will be additive [60].

### 5.2 Assessment of Process Risks

Quality assurance (QA) groups often desire a 100% guarantee of quality, leading to risk avoidance and promoting "over-validation" in the form of additional testing

and repetitive documentation that increases production costs [31]. Although absolute safety is a theoretical concept just like absolute sterility or product purity, it can be approximated using established scientific and engineering principles [12]. Risk analysis identifies the steps, critical factors, and parameters that affect product quality [11]. A documented risk analysis determines the extent of process robustness, summarizes results in a form readily presentable to internal and external regulatory personnel, meets regulatory requirements, identifies critical parameters for monitoring, increases process understanding, assists in troubleshooting, and, most appropriately, focuses process validation efforts where they are most warranted and beneficial [11]. It is first necessary to determine what level of each putative risk factor is considered safe, considering the intended clinical use of the biopharmaceutical (e.g., administration route, dose size and frequency, and total amount consumed) [12].

The process design must include steps that reliably remove or inactivate potential risk factors [12]. Rare, significant, and real risk factors must be distinguished from frequent, less significant, theoretical, or putative risk factors. In some cases, targets are more stringent for real and present risks than for putative ones, and reliance on removal/inactivation steps is less important if risks are purely theoretical. The list of putative risk factors is developed early in the process development cycle and then updated periodically as additional information and experience are obtained. Each potential risk is defined in experimental terms by assessing its biological, biophysical, and biochemical aspects. Two categories of risks are (1) potential pharmacological and toxicological effects of the product's active ingredient(s), and (2) other product components arising from the preparation process. Examples of specific risk factors are the presence of intact cells, adventitious agents (e.g., bacteria, fungi, mycoplasma, and viruses), endogenous retroviruses, residual cellular nucleic acid/proteins, microbial contaminants (e.g., endotoxin and proteins), and process chemicals (e.g., antibiotics, ligands, solvents, cleaning chemicals, inducers, and nutrients) [12].

Risk assessment provides a systematic estimation and evaluation of the risk potential for all components used in the manufacturing process, including the origin and fate of each raw material [53]. Risk assessment also estimates potential risks associated with the application of certain procedures during the manufacturing process (e.g., isolation, validation, and testing) by quantifying a procedure's robustness and susceptibility to errors. It assesses which contaminants have the highest probability of being present and then provides a basis to establish a reasonable and economical testing scheme [53].

Each risk factor must be identified and quantified in terms of its likelihood of occurrence and its severity [12]. After the initial first key steps of risk identification and quantification, the greatest risks then can be reduced [29]. Reducing the probability of a risk's occurrence is achieved via implementations such as process automation, tighter controls, or removal of open processing steps [70]. Reducing a risk's severity factor is achieved by process changes or product redesign [70]. In contrast, raising the probability of failure detection by increasing sampling and monitoring activities has been found to be the least productive and most time-consuming solution with little opportunity for automation [70]. If a significant risk cannot be reduced sufficiently, then the associated process step can be validated to provide additional performance assurances.

***Risk Assessment Tools.*** Risk assessment tools (FTA, FMEA, and HACCP) can be used alone or in combination [11]. They are tools for problem prevention, requiring failure anticipation rather than analyzing failures that have already occurred [71]. After process characterization is completed, formal risk analysis exercises can be repeated so future focus can shift to newly uncovered unit operations and operating parameters requiring attention [65].

*FTA.* Fault tree analysis (FTA) is a deductive, top-down approach to failure mode analysis [11]. Its open-ended structure is readily expandable, making it helpful for complex problems [71]. The failure or hazard is assumed; then the combination of conditions or probable causes necessary for that event to occur is systematically defined by identifying how that high-level failure is caused by lower level, primary failures, specifically the failure of an individual component or subsystem. This information is presented in the form of a fault tree diagram with the assumed failure/hazard listed first at the top and the associated conditions/probable causes listed as branches with successive levels listed as further branches [11, 71]. Events at different levels of the tree are listed below the main failure event and connected by event statements and logic gates defined by Boolean logic [11, 71]. "And" is used if events need to occur together to cause the failure/hazard; "or" is used if the event alone would cause the failure/hazard [11]. As a fully developed FTA can be complex, special symbols are used to describe events and gates to simplify graphical representation of casual relationships between branches [71].

*FMEA.* Failure mode and effect [criticality] analysis (FME[C]A) is a preventative, bottom-up approach to identifying potential failures [11] that has been recently adapted to quantifying GMP risks [70]. FMEA has been applied to audit preparation, batch record review, production validation, change control evaluation, retrospective validation, trend analysis, and validation master plan development [72]. Although it includes identifying the effects or consequences of failure modes, it is a more practical tool and a structured guide to analyze major areas identified by the FTA [11, 71]. However, it is not considered as effective as FTA in identifying root causes for intractable process failures [71]. FMEA uses its structured methodology to facilitate identification of critical variables, specifically reducing 20 or more to 2–3 to permit efficient further optimization [72]. FMEA progresses over time, starting in the early stages of process development [11], and it is most successful if the process to be examined is clearly defined at the outset [71]. Overall, FMEA serves to (1) readily identify potential problem areas where operators must be especially alert and where intensified test procedures may be required, (2) define operational constraints and preventative maintenance actions for equipment, (3) provide information for planning performance qualification, (4) prioritize corrective actions and improvements, (5) foster interdisciplinary teamwork and communication, and (5) facilitate decision making throughout the development process [71].

The first step in FMEA is to develop a working matrix that identifies all information categories that are intended to be studied [67, 71]. A flowchart describing process steps serves as the left column of the matrix. An accurate description of each step helps stimulate thinking about how the step might fail. A list of reasons is generated describing why each step might fail to complete its intended function

(potential failure modes) and what would happen if failure occurs (potential downstream effects) [73]. All possible causes, however improbable, are identified that might result in failure; then the risk attached to each failure is evaluated using an established scale, for example, values from 1 to 5 [71, 73].

The next step is to rate the importance of the identified failure in terms of its probability of occurrence based on the controls in place (P), severity (S), and ability to detected (D), resulting in a risk priority number (RPN) [11, 70]. It asks the following questions: What could go wrong, what is the probability of its going wrong, what are the consequences of it going wrong, and can in-process or final product testing detect it when it does go wrong [72]? RPNs, also known as the criticality index, then are used to sort the identified causes/effects of failures, which enables FMEA to identify those failures posing the greatest threat to the manufacturing process [71]. P, S, and D are all rated between the low and high values; P and S directly increase with numerical rating, and D inversely increases with numerical rating. The RPN value is calculated by the product, $P \times S \times D$ [71–73]. If P, S, or D is completely unknown, then a high value rating (e.g., 5) is assigned [72]. Thus, in the worst case RPN = 125, and in the best case RPN = 1. In practice, RPNs only are calculated for the worst aspects of each failure to reduce workload and to permit identification of the most critical failures first without requiring detailed calculations for each identified cause [71]. Frequently, the highest RPN steps are those steps involving human intervention because P is typically 3 or more [72].

The output of an FMEA study is the list of preventative and corrective actions necessary to improve the manufacturing process that now can be prioritized based on RPN values [71]. An acceptable RPN should depend on the severity rating of the failure, specifically if S = 5 (high value), then the RPN should be ≤5, but if S is below 5, the acceptable RPN can be higher [72]. The RPN then can be recalculated after corrective actions have been implemented providing measurable proof of impact [71].

*HACCP.* Hazard analysis and critical control points (HACCP) is a systematic approach that is system-based to determine high-risk steps [11]. The definition of hazard includes both safety and quality concerns for biopharmaceuticals. An effective HACCP system reduces endproduct testing because sufficient validated safeguards are instituted [11]. HACCP provides detail and documentation to show process/product understanding through identifying parameters to control and monitor. Its stages are to (1) conduct the hazard analysis, (2) determine critical control points, (3) establish critical limits, (4) establish monitoring procedures, (5) establish corrective actions, (6) establish verification procedures, and (7) establish record-keeping and documentation procedures [11, 74].

### 5.3 Validation of Assays

*Purpose.* As biopharmaceutical products are large and complex molecules and no single test method is sufficient, product analysis uses multiple, different analytical methods strategically designed to be complementary with respect to selectivity and specificity [4, 27]. More sensitive test methods allow proof of greater removal of a putative risk factor [12], and the more key parameters that can be measured assure that process variability is understood and controlled [66]. However, it is important

to avoid setting product specifications at the assay's limit of detection [2]. Assays provide adequate in-process testing via process validation and a comprehensive assessment of the final product using a variety of analytical methodologies for identity, purity, potency, strength, safety, and stability [4].

Assay validation characterizes the assay performance so that the significance of the measured assay values obtained is readily understood [75]. Test methods should be validated when important decisions are to be based on the data generated [30]. Thus, the extent of method validation depends on the stage of clinical supply manufacture. The key elements of assay validation for a method are to establish reliability, the intra- and interlaboratory test variation, and relevance, the meaning of the results for a specific purpose [19]. The robustness of an analytical procedure (according to the ICH-Validation of Analytical Methods, 1993) is its measured capacity to be unaffected by small variations in controlled parameters and reliability under normal usage [35].

***Scope.*** The scope of assay validation is to assess the essential test method performance characteristics of accuracy, reproducibility, repeatability, linear-ity, and limit of quantitation/detection [41]. A test method is evaluated for its readiness for assay validation against the following criteria: (1) description of the test basis and scientific purpose, (2) case for relevance, (3) proposed practical application, (4) statement of limitations, and (5) acceptable intra- and interlaboratory precision [8]. Before actual validation, the test method should be optimized and standardized [8], specifically controls necessary to invalidate inappropriate out-of-specification results should be incorporated [30]. Assays must be validated for each matrix in which samples will be run [75] and for its capability on a protein-by-protein basis [56].

Assay ruggedness also is measured in terms of the similarity of results from different laboratories, manufacturer's equipment, and analysts [2]. Interoperator, interday, and intraassay variations can be quantified by comprehensive validation [5]. Interassay (intermediate) precision is the precision of multiple determinations of a single sample analyzed in various runs [75]. Intra-assay precision or repeatability is the precision of multiple determinations of a single sample within one assay run. Interlot precision tests analyze values of aliquots from a single sample run using lots of different assay components. Finally, lot-to-lot precision is the precision of multiple determinations of a single sample analyzed using different lots of assay components [75]. These multiple types of variations must be considered so that release criteria ranges are not set too tight.

The limit of quantitation (LOQ) is the lowest concentration that produces a signal 10-fold above background, whereas the limit of detection (LOD) is the lowest concentration that produces a signal threefold above background [75]. The assay sensitivity and achievable concentration limits of additives define a "window of clearance," which is the difference (on a log scale) between the highest attainable initial contaminant/impurity concentration and the lowest detectable concentration (LOD) of that additive [12]. This difference is the amount that can be measured and thus the amount that can be claimed to have been removed by the process [12].

A low-level sensitivity is required for viral test methods to maximize reduction factors [5]. However, the ability to detect low virus concentrations is limited by

statistical sampling [5]. The probability of detection (POD) for low-level virus concentrations depends on test volumes, batch volumes, and test sensitivity [40]. The POD is high when sample virus concentrations are well above the test method sensitivity, the POD is lower when concentrations are at or below the test method sensitivity, and the POD is even lower for very low concentrations due to the low probability of obtaining a virus in the sample as described by the Poisson distribution. Test method sensitivity should be based on a POD ≥95% [40].

In-process and release (endproduct) parameters, both critical and operational controls, are controlled and monitored during production using assays [11]. In-process controls generate data at critical process steps, which serve as a basis for specification development for product intermediates [2]. Specific in-process controls, routine sampling, and monitoring at specific process steps are established to monitor key process outputs [32]. Information only tests are used postvalidation without preestablished acceptance criteria until sufficient data have been collected to statistically establish clear limits or else omit the testing [32]. Thus, sampling and obtaining data for information only is an important part of process validation that must be recognized.

### *Assay Types*

*Potency.* Binding assays for quantitation use the binding of at least two molecules to form an interaction that withstands multiple washing steps, and they are the least variable [75]. Cell-based bioassays for potency use a cell line mixed with a sample to generate a cellular response that is quantified [75]. They have more variability [75] and are tricky to use to measure potency because the CV often is >30% [2]. Whole animal assays for potency are time-consuming and highly variable [75], and they have the typical disadvantages of *in vivo* tests in that they are expensive, inaccurate, slow, and raise animal ethics issues [19].

The development and validation of *in vitro* tests that correlate with potency, typically desirable for vaccines, is tedious and rarely successful, requiring thorough knowledge of the immune response after infection or vaccination [19]. An alternative approach is to demonstrate batch comparability and then assume identical properties *in vivo* [19]. For an increasingly greater number of vaccines, no routine animal potency testing is being conducted because the product is well characterized, critical production process steps are controlled, immunology knowledge permits functional test development, analytical knowledge permits antigen characterization, and test procedures are becoming harmonized and validated [19].

*Impurities.* Analytical methods for impurity clearance require sensitivities capable of demonstrating sufficient clearance to well below safety margins [33]. rDNA technology has stimulated the development of a wide variety of assay types [16]. Specifically, PCR is used to understand how manufacturing steps impact retroviral clearance due to its high precision, high throughput, and low cost [76]. Typical testing for in-process impurities includes endotoxin, host cell and other proteins, DNA, viruses, and altered protein forms (e.g., proteolytic clips, deamination, oxidized methionines, and amino acid substitutions) [16]. Clearance of host cell proteins is measured using immunoassay [50]. Host-cell DNA is radiolabeled by nick-translation and then measured in column fractions (also radiolabeled host-cell

proteins), and it is measured directly in production streams [51]. Cytotoxicity and interference testing using non-virus-containing samples are performed to ensure that sample matrices in a viral clearance study do not adversely affect the ability to titrate virus [5]. These tests are performed well in advance of the actual validation so that conditions may be found for which dilution is not required, and thus, assay sensitivity is not reduced by dilution [5].

There is a lack of practical laboratory assays to detect bovine spongiform encephalopathy (BSE)-like agents so the potential introduction of these agents into the process must be controlled by avoiding bovine-based materials, using materials from low-risk bovine tissues, and/or sourcing materials from countries with good BSE control measures [21].

Typical testing on final purified bulk includes those tests conducted with in-process testing plus total protein content and potency [16]. The final product (sterile filtered, filled and sometimes lyophilized, then packaged into vials and ampoules) is tested for sterility, pyrogenicity, particulates, content uniformity, identity, excipient chemical content, potency, and protein content [16].

***Standards.*** Key to the success of any assay method is the availability of a reliable standard. A primary standard is one recognized by a national or international agency, whose characteristics have been established by collaborative effort [77]. Secondary standards are calibrated against primary standards. Reference material is used for the evaluation of other materials to generate comparative data. Primary reference material has been thoroughly characterized, and it is the purest material available in quantities large enough for use. Biochemical reference standards are used for all product tests and for process-specific tests both for material release as well as for upstream/downstream process monitoring and assay development. They are unique to a specific product in its amino acid sequence and glycosylation but different for each manufacturer. To ensure compatibility with the matrix of the intended assay, different standards can be required for different process steps. It can be challenging to provide sufficient quantities of the various types of reference material early in the product development cycle [77]. A working reference material is required to validate analytical procedures [2]. It should be stable for at least 2–4 years, and its stability should be monitored regularly [77].

### 5.4 Change Control and Ongoing Process Validation

Validation maintenance, an ongoing activity for manufacturing processes [11], continues through the life cycle of the process with a changing focus as the process matures [20]. There is a need to address the process life cycle as a whole and not to suspend process validation after three production scale runs are completed [14]. To support this life-cycle approach, process expert teams are created to rapidly resolve process deviations, determine trends toward loss of control, comply with regulatory requirements, assess process change impact, and identify areas for process improvement [11].

A validation program is specific for a product manufactured by a given process [12]. A strong change control program needs to be established once the process is validated [15]. Changes to process equipment, operation, or facilities should be reviewed carefully before the change is approved to determine the potential valida-

tion impact and if repeating validation studies are required [12, 25]. Such changes include process improvements, raw materials changes, step substitutions, and scale-up [3]. Changes that are expected to shift the value of a performance parameter require revalidation [15]. Specifically, if the purity is decreased at a specific step, then recovery of this purity must be demonstrated downstream [15]. In general the potential quality impact of process changes is higher with glycosylated rather than with nonglycosylated proteins [13].

The process change is documented in a proposal with its justification, along with the number of batches necessary and strategy (e.g., samples and data needed, extent of validation testing, and acceptance criteria) to evaluate the change [15]. In some cases, the original risk assessment (e.g., FMEA) is revisited or a new one conducted to ensure an unintentional new failure mode has not been created [71]. A formal change control report is written and approved [15].

Revalidation is the repetition of the validation process or specific part of it [25], and it has some conceptual overlap with retrospective validation. Periodic revalidation has been conducted in some cases even if nothing has changed [25]. Specific elements of most validation programs are repeated (at least partially) at regular, appropriate intervals (e.g., every 6–24 months) to verify that the original parameters are still effective [3, 12]. In other cases, if there have been no changes, process revalidation can consist of a formal review of operating and performance parameters, along with the nature and frequency of excursions, which can be executed as part of the annual product review [15]. When existing processes are reassessed, some input parameters may be more tightly controlled than previously permitted or a noncritical parameter may be found to be critical [32]. Product failures occurring within validated ranges of a process might be due to simultaneous excursions of several variables not studied during the initial process validation or from the use of an inadequate scale-down model [20].

Trending of process variables is used to refine range limits [3]. Manufacturing data can lead to superseded NORs and/or MORs, which are justified by the associated operational trends and product quality achieved [20]. Extreme parameter values subsequently can become acceptable in process validation if the acceptability of final product has been confirmed by multiple observations [12].

### 5.5 Process Validation Documentation

Documentation of process validation serves to establish that its goals are reasonable and achievable, as well as to ensure that the study design is based on solid scientific and engineering principles [12]. The process validation master plan defines the scope and rationale of process validation, giving an overview of the validation, its philosophy, and general testing strategy [33]. It describes how the validation is to be done, including the schedule, studies to be performed, systems to be validated, responsibilities, and approvals required [3, 15, 33].

The process validation protocol is a preapproved written plan stating how the validation will be conducted and identifying acceptance criteria as well as sampling and assay requirements [33] and other testing details [3]. It is a prospective experimental plan that when executed produces documented evidence that the system has been validated [25]. The protocol defines the system to be validated, identifies operating variables and probable control parameters, and indicates the number of

replications required to provide statistical significance [25]. There is a major savings of cost and time realized by adopting potential streamlining measures, for example, combining resin reuse and chromatography step characterization studies into a single protocol [61]. The entire process validation study protocol is reviewed and approved by manufacturing and quality personnel [12] as well as by process development representatives. The process validation report includes the validation protocol, deviations from protocol, validation results, and conclusions about validation status of the step [3]. Copies of most critical validation reports and summaries of other reports are included in the license application [3], so their careful assembly, clarity, and accuracy is critical. Common validation submission deficiencies include the lack of data or protocols, inadequate data summaries, missing acceptance criteria, and inadequate monitoring or sampling [78].

## 6 PROCESS VALIDATION STRATEGIES

Several value-added, science-based, best practices are evolving for process validation, which are appropriate from both business and regulatory viewpoints. Selected industrial applications of process validation are summarized in Table 6. For a biopharmaceutical process, the potential number of variables and possible interactions among variables is large [12]. One possible approach is to conduct a worst-case test to evaluate each known critical variable separately, and if necessary to run steps more than once if they possess more than one critical variable [34]. Although the use of simple experimental designs to develop the relationship and describe the link between input and output parameters is straightforward, it can result in high numbers of experimental runs, which are impractical [10]. Potential streamlined approaches include retrospective analysis, qualified scale-down studies to identify critical parameters and important interactions, and worst-case runs [10]. In addition, more than one variable can be coupled together into a single variable if they are believed, based on science, to have a predictable effect on output variables [35]. Scouting studies then are used if process development data are limited regarding one or more variables or if the ranges previously studied were insufficiently large [20].

### 6.1 Historical Data Review

Retrospective analysis of historical data can be used to assess process reliability [88]. The initial process capability specifications are established based on a few runs (usually less than around 5–10) with acceptable results [30]. This number of runs often is too small to accurately define and characterize actual process variability [30, 31]. It can be difficult to claim that the previously set process capability specifications were flawed after validation runs fail to meet them [30]. Thus, the selection of acceptance criteria must consider the benefits of broad versus narrow tolerance ranges.

It is best to develop acceptance limits for parameters such as yield and purity using statistics applied to historical data [89]. The difficulty in setting acceptance criteria depends on the availability and quality of historical data. Small datasets tend to underestimate process variance because they contain only a limited number

**TABLE 6. Selected Industrial Applications of Process Validation**

| Process Stage | Specific Description | Strategy | Reference (Company) |
|---|---|---|---|
| Fermentation | Recombinant *E. coli* protein production (defined medium, induced fed-batch process) | Bridged process changes to former process used for earlier clinical material | [38] (Amgen) |
| | Monoclonal NSO (GS) humanized antibody production(fed-batch process with concentrated medium additions) | Statistical DOE used random combinations of selected control parameters at high/low values in scale-down model | [79] (GlaxoWellcome) |
| | Recombinant yeast vaccine production (defined medium, fed-batch process) | Scale-down model to identify critical parameters | [40] (Merck) |
| Chromatography | Chromatography of recombinant *E. coli* protein | Parameters varied over narrow range and effect on step yield/pool purity determined | [80, 81] (Amgen) |
| | Reuse of column with affinity ligand for chromatography of antibody | Automation used to perform repeat cycles to evaluate elution profiles, yield, and purity | [82] (Celltech) |
| | Affinity chromatography of recombinant protein produced by CHO cells | Validated scale-down model used to test robustness (via fractional factorial design), reuse and virus removal | [83] (Wyeth/Genetics Institute) |

**TABLE 6.** *Continued*

| Process Stage | Specific Description | Strategy | Reference (Company) |
|---|---|---|---|
| | Chromatography steps for purification of monoclonal antibody produced by CHO cells | Scale-down studies for resin use and viral clearance fully predictive of larger scale | [26] (Genentech) |
| | Cation exchange chromatography of nerve growth factor produced by CHO cells | Process variables combined and linked, then statistically analyzed | [60, 84] (Genentech) |
| | Retroviral clearance in chromatography of antibody produced by hydrodomas | Moloney murine leukemia virus (model retrovirus) used in scaled-down chromatography column | [85] (Celltech) |
| | Viral reduction during manufacture of Factor IX | Scale-down models used to demonstrate log reduction. | [86] (CSL) |
| Overall purification | Purification of a multivalent bacterial vaccine | Matrix/family approach to reduce number of full scale validation lots from 69 to 32 | [36] (Merck) |
| | Purification of a therapeutic monoclonal antibody | Scale-down characterization studies for various filtration and chromatography steps (including viral clearance) to predetermine acceptance criteria | [87] (Biogen Idec) |

of historical production runs [89]. If historical data are used, excessively large ranges are difficult to defend.

Process variability is determined using available tools such as preparing historical trends of the mean ±3 standard deviations SDs to demonstrate agreement of manufacturing scale runs with clinical manufacturing/development runs [11]. After 15 or preferably 30 data points are obtained, a control chart based on a moving range is the best tool for monitoring process stability [89]. Statistical process control (SPC) and multivariate analysis can be applied to historical data to identify strategies for improving yield through investigation of cause-and-effect relationships using correlation tools and process knowledge [88].

### 6.2 Statistical Design of Experiments (DOE) and Analysis for Process Characterization

Experimental designs are used to efficiently identify those input variables that significantly affect output [20]. The use of DOE to easily screen a variety of operating parameters provides a framework for the design and interpretation of experiments, identifies those factors with the greatest effects, segregates key from nonkey parameters, identifies interactions and synergistic effects, and models how outputs relate to inputs [37, 65]. DOE is a powerful, prospective method compared with retrospective analysis [10]. Its ability to determine interactions between key parameters leads to the identification of other weak spots in the process not observed during initial screen experiments [65]. DOE is limited to assessing interactions among input parameters that can be tested within the same experiment [10]. Usually there are more variables than can be tested in single experiment, however [10]. To prepare for successful characterization via DOE, there is a need to design for simplicity by removing non-value-added experiments and to remove as much experimental variation as possible [37].

DOE modeling can be used early in process characterization to set guard bands, ranges for key process input variables where no statistical change to output is observed [31]. These guard bands are set to locate optimal settings that maximize output and minimize variation. Subsequent validation consists first of experimental design to screen key process input variables set at the guard bands previously determined and then to confirm that the output is within specification [31]. Modeling DOE helps understand the relationship between key input factors and critical Ys, relates input factors to responses (critical Ys) via a model (transfer function) that essentially is process characterization, determines input factor settings that optimize responses, and uses a model to set a limit on input variation such that no statistical changes in KPIVs are observed [37]. This approach is consistent with the PAT initiative.

The maximum information per experiment can be obtained by combining studies and varying factors (i.e., variables) using fractional factorial design [61]. A set of fractional factorial runs provides an estimate of each variable's effect and interactions, but no single run can be compared with a control run [35]. This design permits a large number of variables to be tested with lower number of experiments. Often variables are tested at a low and a high level that permits only linear interpolation (i.e., assumes a linear process response) [35]. If center points (i.e., midpoint of the tested range) are added to the experimental design, nonlinearity or

curvature can be detected if present. Replication of low- and high-level test conditions increases the precision of the estimate of that factor's effect. Two-way variable interactions are tested with factorial design, but additional experiments may be necessary to deconvolute the effects of two factors that are confounded (i.e., indistinguishable effects from one another). Fractional fac-torial designs test the extremes of all variables and then generate the best-fit model with an evaluation of the statistical confidence of the model response predictions. If no process failures are predicted by the best-fit model, then the process step is considered robust over the ranges of control parameters tested [35].

Statistical analysis demonstrates which input parameters are important contributors to output parameters [10]. JMP statistical software (SAS, Cary, NC) can be used to analyze data and generate separate analyses for each response variable of interest [61]. A difference of 3 SDs often is selected as the criterion for significance corresponding to a 95% confidence level [61], and the test acceptance criteria range typically are set as the mean plus/minus 3 SDs (industry standard) [91].

### 6.3 Family/Matrix Validation and Platform Technologies

Common validation studies are required across different processes to demonstrate capabilities for viral clearance, resin and membrane reuse/sanitization/storage, buffer stability, filter extractables, and resin leachables [14]. Furthermore, many biopharmaceuticals are being developed using common platform technologies (e.g., cell line, medium, unit operations) to reduce development time [7]. Thus, templates can be developed that identify key input and output parameters for a certain unit operation and then customized to address unique process or product specifics [14]. Generic validated assays can be used for impurity removal, if a platform cell line and similar culture conditions are used that then are qualified for each new product [14]. In addition, generic validation can be developed in which the supplier and potential industry users define generic operating conditions under which the technology is to be used, and the equipment manufacturer conducts much of the testing for every industry user [90]. This reduces additional user validation testing to only product- or user-specific applications (e.g., product compatibility) [90].

To significantly reduce the number and extent of new studies associated with new products processed using similar or identical process operations (i.e., platform technologies), family and matrix validation approaches are applied to reduce repetitive work [33]. In a family approach, one piece of equipment is validated by three consecutive runs and other similar equipment is validated by one confirming run in each unit [33]. This approach can be applied to the process validation for a product with several distinct, different but similar, components (e.g., antigens for a multivalent vaccine) by (1) treating different product components as a product family, (2) classifying product components into groups based on their similarities, and (3) selecting one component from each group to validate, possibly based on worst-case experience during process development [36]. This strategy is facilitated when (1) the final product can be fully analytically characterized, (2) there is significant prior manufacturing experience with similar production processes and significant process development experience with the new production process, and (3) clinical trial data show comparable safety and efficacy profiles for some product components [36].

In a matrix approach, validation is conducted at the full range or extremes of process parameters, and then values in between are assumed validated [33]. This generic strategy is used to support a wide range of bracketed conditions that include the most common operating parameters for more than one product, as long as process streams and key step components (e.g., resin type) are similar [14, 91]. The matrix approach reduces the extent of process validation required for successive products, at the expense of only a limited impact on the initial product's process validation, as long as the platform technology is reasonably maintained. Example applications have been retrovirus inactivation via low pH treatment [42] and viral clearance using anion exchange chromatography [91].

### 6.4 Scale-Down Models

Although the most reliable data are obtained from actual production runs [12], determination of operational ranges at the manufacturing scale is laborious, time-consuming, and costly [40]. Thus, typical manufacturing scale validation is limited to impurity clearance (e.g., proteins, DNA, and small molecules) and resin/membrane cleaning [14]. If the process step cannot be scaled down successfully, then range definition validation studies are required at the large scale [32].

Small-scale process models are an essential and valuable tool for process validation and ongoing troubleshooting [32]. Scale-down factors ranging from 10- to 10,000-fold are employed [3] with the extent of scale-down de-pending on the actual production scale and smallest scale that can reliably reproduce the process [51]. Small-scale equipment is permissible only if the scalability of the unit operation is demonstrated [12]. The same technical principles for scale-down are used as for scale-up for the step of interest [3].

The scale-down model should represent the large-scale model with respect to input parameters and typical values for output parameters [10]. Data are collected under conditions relevant to full-scale manufacturing GMP process [12]. Qualified scale-down models are used in combination with strong statistical tools like DOE [10], but sample handling at the small scale is easier than during GMP manufacturing runs [26]. Multiple validation experiments can be conducted in parallel employing several identical scale-down units, thereby shortening the experimental time expended [3]. Also a hybrid approach can be instituted where samples from full-scale process streams are used as feed streams to scaled-down steps [3]. Overall the amount of re-sources used in small-scale validation is likely to be less than that needed for manufacturing scale validation, although in manufacturing scale validation, resource utilization is spread over a longer period of time [26].

Scale-down studies have been used for a wide variety of process validation studies, including resin lifetimes, in-process stream hold times, buffer stability, virus clearance, harvest criteria, filter extractables, resin leachables, and cell age at harvest [14, 91, 92]. The ease of scale-down differs depending on step and should be considered in selecting those steps to be validated [5]. In fact, certain validation issues can be addressed only via small-scale models (e.g., virus clearance evaluation, nucleic acid and other impurities/additives removal, cleaning and storage procedure evaluation, and column lifetime estimation) because their use increases worker safety, reduces costs, and permits use of higher titer samples for improved

accuracy of prediction [52]. Specifically, viral clearance studies can only be performed at the small scale because virus introduction into cGMP manufacturing scale equipment is impractical and inappropriate [5, 26]. For viral clearance studies, the ideal scale-down factor is 10- to 100-fold [6].

Although scale-down models have been used largely in purification, they also have been used in fermentation. Fermenter scale-down models use the same set points for volume-independent operating parameters (e.g., temperature, pH, and DO) and adjust the set points for volume-dependent parameters such as mixing and gassing (e.g., $dCO_2$ stripping) [68]. The inoculum split ratio for cell culture scale-down is maintained constant [3, 68]. Characterization studies are performed in small-scale fermenters, and then it is demonstrated that the ranges established are insensitive to scale-up [40]. Only the critical process parameters found to be sensitive to scale-up require range studies at the manufacturing scale [40].

There is increased regulatory concern over the accuracy of results obtained from small-scale models [52]. Consequently, the scale-down model must be validated to be consistent with full-scale manufacturing before validation studies begin. For chromatography steps, it is necessary to determine that product purity and yield are the same [52] or demonstrate comparable HETP, peak asymmetry, and retention time [5]. Any product streams tested should be representative of a commercial-scale process feed stream [52].

For chromatography steps, the scale-down model consists of a column with system and auxiliary components (e.g., distributor, monitors, and fraction collectors) [52]. Scale equivalence for chromatography columns maintains constant column bed height and linear flow rate, while varying column diameter, buffer, and load amounts, which are scaled to column volume [61]. Specifically, the column diameter is decreased, but the bed height is maintained to maintain residence time, and the volumetric flow rate is reduced, but the linear flow rate is maintained [5, 52]. Constant residence time is important if viral clearance is due to inactivation rather than removal [52]. Configuration, transport distances (tubing diameters as well as tubing lengths), and materials of construction are similar to the commercial-scale system. The same methods of preparing buffers are used, including the same quality of water, buffer, and salts. The chromatography medium has the same base matrix, functional groups, and degree of substitution, as well as the lot used should meet approved specifications [52]. Finally, the product mass loading and relative buffer/load volumes between scale-down and production systems are similar [33].

Although chromatography scale-down is extensively studied owing to its use for viral clearance validation, the scale-down of other purification steps also is conducted. Virus filtration scale-down maintains the same linear or filtration pressures, product feed concentration, ratio of feed volume to filter area, and temperature as the production system [33]. Other steps such as solution inactivation (e.g., pH and heat treatment) are relatively easy to scale down because it is simply necessary only to maintain buffer composition, pH, protein concentration, and temperature consistent with production conditions [5]. The time course of the inactivation is quantified along with the upper tolerable range of inactivation agent concentration and minimum exposure time. Residence time differences between the production and small scales might change parameters such as buffer

temperature [52], which might affect solution as well as chromatography inactivation steps.

### 6.5 Well-Characterized Biotechnology Products

A well-characterized biological (biopharmaceutical) is a chemical entity whose identity purity, impurity, safety, potency, and quality can be determined and controlled [19, 93]. To be well characterized, the drug substance must be >95% of the main component and/or related isoforms [93]. Well-characterized molecules are evaluating using sensitive and discriminating tests that are quantitative and relevant *in vivo* and *in vitro* potency assays [93]. The recent explosion in the availability of sophisticated, reliable, analytical instrumentation, combined with extensive, sensitive product analysis methods has resulted in great advancements in defining "well-characterized" biopharmaceuticals [28].

Although for a well-characterized biological the FDA trend is to rely on reduced final product testing, process validation must be present to assure product quality and consistency [15]. There has been an elimination of regulatory agency lot release for specific well-characterized biopharmaceuticals [19, 93]. In addition, process-related impurities (e.g., host cell protein and DNA) are controlled by process validation *in lieu* of lot-release testing [27]. Furthermore, many process changes are justifiable using analytical data and process validation studies without a clinical trial [3]. The following principles can be applied to the manufacturing and testing of a well-characterized product: (1) The manufacturing process should be robust, reproducible, validated, and designed to produce active drug substance; (2) the safety of the final product is assured through process validation using validated analytical testing to demonstrate removal of impurities to safe levels; (3) tests for identity, purity, and potency should be sensitive, quantitative, and validated; and (4) specifications should be quantitative and based on historical manufacturing data and clinical experience [27].

The more characterizable the biopharmaceutical product, the more emphasis is placed on validating that the correct molecule is being produced than assuring that the cell biology is being controlled [10]. However, the difficulty of characterization increases in direct proportion to the complexity of molecule [93]. When less is known about the product, a greater reliance is placed on product and process consistency [10]. Most vaccines used today are not well-characterized biologicals, so batch release must be performed by the manufacturer and by the appropriate national control laboratory [19]. The quality of vaccines is increasingly guaranteed by the use of robust and reproducible production processes [19], and thus, a high reliance is placed on process validation.

## 7 FUTURE TRENDS

Future developments in the biopharmaceutical industry are likely to impact the scope and execution of process validation activities. The nature of process validation is continuously changing to meet the requirements of biopharmaceutical development and manufacturing. The understanding and implementation of new strategies based on future developments is key to maintaining an effective process validation methodology.

### 7.1 Process Analytical Technologies (PATs)

PAT is a system for designing, analyzing, and controlling manufacturing through timely in-process measurements of critical quality and performance attributes, leading to the goal of ensuring final product quality [94]. The implementation of PAT potentially minimizes the number of tests that need to be performed at the time of lot release, or in the future for certain products, it might eliminate lot-release testing by the manufacturer entirely. Its precedent was the dropping of lot-release testing for certain impurities when validation of their removal was able to be assured [94].

Online and automatic methods have been demonstrated to be more accurate and precise compared with manual methods because (1) sampling and sample pretreatment artifacts are eliminated, minimized, or reproduced; (2) measurement frequency is increased; and (3) results are independent of personnel availability [69]. Continuous, real-time monitoring is preferred when feasible [11]. PAT is used to implement in-line controls to increase productivity and reduce costs [31]. In addition, accumulated data from numerous commercial batches are used to generate trend analysis profiles showing high and low control limits for process parameters [25]. These limits then are used to predict when processes may be deviating from their validated control ranges, although they are currently operating within them [15].

Relationships among manipulated (controlled) variables, online measured variables, and product (uncontrolled) variables in most biosystems are nonlinear to some extent [95]. A forward model is when parameters, starting conditions, and relevant equations governing behavior are known, readily measurable inputs and the outputs are variables; an inverse model is when the inputs are readily measurable variables and the outputs are difficult to measure parameters [69]. The forward model is most applicable to process validation, whereas the inverse model is most applicable to metabolic pathway analysis. Modeling systems such as neural networks have been used to describe the characteristics of extremely complex bioprocess systems [95].

Mathematical models can be used to simulate the controlled process variables and thus predict the effect of variations on the uncontrolled process variables [35]. Linking online sensing to modeling is effective when (1) there is product yield and biomass variation among optimal values, (3) optimal values may vary over time for processes that are not stationary, (4) parameters can be measured directly or otherwise, (5) the cost of online control is a small fraction of capital and expense costs to run a process, and (5) historical data combined with mathematical methods provide identification of key parameters and their optimal values to achieve rational process improvement [69]. Achieving good modeling is not always easy and involves correct identification of parameters and variables, and selection of the type of model (forward or inverse) along with key inputs and outputs, showing the relationship between predictions and actual data to assess model effectiveness [69].

A solid understanding of a product's critical quality attributes and what aspects of the manufacturing process control them is necessary to fully gain the benefits of PAT [94]. Furthermore, the inherent complexity of protein-based biopharmaceutical products makes it difficult to use PAT to assess product characteristics critical to safety, efficacy, and stability. In addition, it is difficult, compared with

small molecules, to directly monitor the desired product concentration during upstream/early downstream processes in a background of protein impurities deriving from medium components or host cells [94]. However, PAT can be used for process monitoring and control that then indirectly results in improved productivity, process consistency, and product quality.

The pharmaceutical industry lags behind other automated industries in manufacturing quality analysis because it is mostly lab-oriented with little closed-loop, real-time control, and limited enterprise-wide data availability [96]. There is a fear of potential regulatory agency reprisals if PAT is implemented for existing processes, and its use detects behavior that would not have been detected using conventional process monitoring [96]. Before PAT is widely implemented, the risks associated with data variability need to be evaluated and reliable methods need to be instituted to distinguish among signal, process, and product variability [94].

### 7.2 Process Reliability

More increasingly, the supply capability of a commercial organization depends not only on manufacturing but also on process reliability, GMP compliance, and operational metrics [79]. Process reliability lies at the heart of lean manufacturing and Six Sigma programs because it ensures that the manufacturing process can be maintained in a compliant, validated state [97]. Performance factors are measured and monitored on an ongoing basis so that improvement efforts can be targeted at the most significant opportunities leading to the removal of variation from the manufacturing process [97]. The greatest threats to process reliability are production variances and associated investigations [79]. Variance analysis identifies systems prone to problems, and then it uncovers the root causes of these variances, strengthening the mechanical and operational aspects of the process [79].

Process reliability is the ability of the process to consistently produce required results as measured through the primary dimensions of uptime, dependability, and first run yield [97]. Uptime is the time the process is in operation compared with the available time the process is scheduled to be in operation. Dependability is the repeatability of a process' actual run rate compared with its scheduled run rate. First run yield is the ability to produce quality outputs the first time through production run without any rework. Reliability losses are measured in terms of unscheduled downtime, run rate losses, yield losses, and reprocessing [97].

Process reliability is improved most effectively using cross-functional teams by the following process: (1) identification of the opportunity; (2) collection of reliability loss data regarding uptime, dependability, and first run yield; (3) analysis of loss data and determination of the root cause of the problem; (4) brainstorming of solution alternatives; (5) selection and implementation of solutions considering solutions that are safe, feasible, implementable, and cost-effective; and (6) monitoring of results and repetitions to assess the impact of solutions and to prioritize remaining opportunities [97]. Six Sigma states these concepts for continuous process improvement as DMAIC: define, measure, analyze, improve, and control [37], which are readily applicable to process validation. Measurable benefits of improved process reliability include increased customer service (e.g., timely delivery), increased schedule adherence, reduced lead time, reduced inventory levels, increased effective capacity (i.e., greater process uptime and reduced unplanned

downtime), decreased cost, creation of time available for predictive maintenance, incremental capacity for new products or increased volume without additional capital investment, increased process dependability, improved quality and first run yield, and reductions in setup and changeover time [97].

Process flexibility is needed for rapid turnover and reduced inventories that are key to lean manufacturing, which in turn are crucial to making more customized drug products and remaining competitive [97]. Flexibility losses are when time, people, and material waste are consumed by product changeovers. Flexibility is improved by mapping the changeover process, analyzing steps, and then eliminating, resequencing, or improving activities. In addition, standardizing procedures and setup locations and advance preparation for changeovers improve efficiency [97].

### 7.3 Biogenerics

Biogenerics, also known as "follow-on" biologics [98], are pharmaceutical preparations, using a generic name, that are based on drug substances arising from recombinant technology (complex structure and high molecular weight) that have been demonstrated to be essentially similar to an original pharmaceutical with an expired patent [99]. The main features of biogenerics processes (such as reproducibility, validation, and controls) need to be maintained similar to the patented biopharmaceutical product and should include quality, safety, and efficacy [100]. Key scientific issues for biogenerics are product characterization, demonstration of bioequivalence, and manufacturing issues (e.g., lack of ownership or access to a contracted facility, incomplete disclosure of patented product processes, and difficulty obtaining a suitable cell line and raw materials) [100]. The process as well as the strain for biogenerics are different from those of the original product, and the impact of even small changes on biopharmaceuticals is not clear [99].

The potential world market for biogenerics has been estimated at $5.4 billion/year by 2008, assuming the evolution of a favorable regulatory environment [101], compared with a total biopharmaceutical market of $30 billion/year [99]. These estimates may be too high for the market potential for biogenerics because patients often are switched to the next-generation versions of branded products due to rapid innovations in these products [101]. In fact, the main barriers to entry are the fast pace of existing product enhancement or new product introduction, and an ill-defined regulatory framework [99]. In addition, due to the expense of biomanufacturing and developing a biomanufacturing process, the cost reduction of biogenerics may not be as significant as with non-biopharmaceuticals [1]. One estimate is that biogenerics will sell for 10–20% less than branded counterparts rather than 40–80% less as with small-molecule drugs [101].

Generic biopharmaceutical products, possessing a similar complex structure and high molecular weight as the original product, are harder to characterize than small-molecule generic drugs; thus, they are harder to prove equivalent [101]. To show bioequivalence, that the biogeneric is "essentially similar" to the patented biopharmaceutical, sophisticated analytical analysis is required. Comparability studies are required to prove the similar nature of the generic product, and the nature of these studies is product-class specific [101]. Prior experience is that even small changes in the product can result in significant safety or efficacy differences.

Biogenerics are currently in a state of "regulatory purgatory" [98]. Key regulatory issues are that the approval process is not well defined and is uncertain, particularly in the United States [100]. The FDA says creating a regulatory pathway for biogenerics is a top priority (Europe already has a biogenerics regulatory framework), but little resolution has been reached [98].

Regulatory authorities are requesting extensive demonstrations of biological activity, which translate to lengthy and costly clinical trials [99]. Proof of identity is difficult for biopharmaceuticals because they are difficult to analytically characterize. "The process defines the product" notion still exists when there is limited knowledge [100] and experience regarding cause-and-effect relationships [99]. Extensive oversight of biogenerics manufacturers is expected because the entire manufacturing process is as important as the final product and needs to be carefully regulated [1]. These requirements make development timelines and costs substantially higher than for small-molecule generics [99].

### 7.4 Outsourcing

As biopharmaceutical companies redefine their goals and priorities, outsourcing of workload can maximize investment [28] and assure an ability to meet capacity demands [102]. Previously, outsourcing has been done in clinical research development and fill/finish manufacturing steps [28, 103]. Now it is being done for QC/QA functions because every quality function except final release responsibility for biopharmaceutical drug product for human clinical trials can be outsourced [28]. Outsourcing also is occurring in many areas such as process development, clinical material production, and marketable production [102, 104]. Specific areas such as cell line development (high producer screening) and scale-up also can be outsourced [104]. Risk management systems are used to proactively identify the risks and liability between the organizations [105].

Contract companies can provide more than simply additional capacity. They can provide translation of bench-scale operations to cGMP compliant manufacturing, technology transfer, and process validation [104]. Contract services also are provided by equipment suppliers to optimize/develop processes using their equipment [90]. These services can be more cost-effective than third-party laboratories, contract manufacturing, or in-house.

### 7.5 Risk-Based GMPs

In the future, when manufacturing processes are developed and understood, there will be increased innovation and regulatory oversight proportional to risk [23]. The major goals of risk-based GMPs are to focus GMP attention on potential risk areas, ensure regulations do not impede innovations, enhance GMP inspection consistency, and encourage the use of the latest scientific advances in manufacturing [106]. To accomplish these goals, work is pro-gressing based on several major principles, including risk-based orientation, science-based policies/standards, integrated quality systems orientation, international cooperation, and strong public health protection. The advent of risk-based GMPs provides a timely opportunity for industry to prospectively evaluate ways to streamline the complexity of existing and proposed processes using sound and defendable methodologies for risk assessment [106].

## 8 SUMMARY

Overall, process validation is a methodology to permit organizations to take advantage of process improvement opportunities presenting themselves over the course of a product's lifetime. Process validation continues to be a formidable undertaking, requiring extensive resources and time for full execution. Improvements in efficiency are likely to arise as technologies for high-throughput scale-down units and fast-turnaround, sensitive analytical assays become more readily available. Focused guidance on process development, characterization, and validation also should be beneficial to evaluate and align varied industry practices.

## NOMENCLATURE

API—active pharmaceutical ingredient
BSE—bovine spongiform encephalopathy
CBER—Center for Biologics Evaluation and Research, US FDA
CCP—critical control points
CQA—critical quality attributes
CDER—Center for Drug Evaluation and Research, US FDA
CV—coefficient of variation
DO—dissolved oxygen
dCO2—dissolved carbon dioxide
DOE—design of experiments
EMEA—European Medicines Evaluation Agency
EU—European Union
FDA—(United States) Food and Drug Administration
FME(C)A—failure mode and (critical) effect analysis
FTA—fault tree analysis
(c)GMP—(current) Good Manufacturing Practice
HACCP—Hazard analysis and critical control point
ICH—International Conference on Harmonization
KPIV—key process input variables
LAL—limulus-amoebocytes-lysate
LOD—limit of detection
LOQ—limit of quantitation
MOR—maximum operating range
NOR—normal operating range
PAR—proven acceptable range
PAT—process analytical technology
PCR—polymerase chain reaction
PDA—Parenteral Drug Association
POD—probability of detection
PV—process validation

QA—Quality Assurance
QC—Quality Control
RPN—risk priority number
SD(s)—standard deviation(s)
SOP(s)—standard operating procedure(s)
SPC—statistical process control
TOC—total organic carbon
WHO—World Heath Organization

## REFERENCES


1. Dibner M D (2005). Biopharmaceuticals in 2005 and beyond. *Bioexec. Internat.* 1: 2–8.
2. Akers J, McEntire J, Sofer G (1994). Biotechnology product validation, part 1: Identifying the pitfalls. *BioPharm.* 7:40–43.
3. Zabriskie D W, Smith T M, Gardner A R (1999). Process validation. In M C Flickinger, S W Drew (eds.), *Encyclopedia of Bioprocess Technology: Vol. 4, Fermentation, Biocatalysis and Bioseparation*: Wiley, New York, pp. 2070–2079.
4. Federici M M (1994). The quality control of biotechnology products. *Biologicals.* 22:151–159.
5. Darling A (2002). Validation of biopharmaceutical purification processes for virus clearance evaluation. *Molec. Biotechnol.* 21(1):57–83.
6. Hill C R (1990). The manufacture of proteins for human therapeutic use. In D L Pyle (ed.), *Separations For Biotechnology Vol. 2*, Elsevier Applied Science, London, pp. 431–443.
7. Sommerfeld S, Strube J (2005). Challenges in biotechnology production- generic processes and process optimization for monoclonal antibodies. *Chem. Engin. Process.* 44:1123–1137.
8. Hendriksen C F M (1999). Validation of tests methods in the quality control of biologicals. In F Brown, C F M Hendriksen, D Sesardic (eds.), *Alternatives to Animals in the Development and Control of Biological Products for Human and Veterinary Use. Dev. Biol. Stand.* Basel Switzerland, Karger, 101:217–221.
9. Naglak T J, Keith M G, Omstead D R (1994). Validation of Fermentation Processes. *BioPharm.* 7:28–36.
10. Lubiniecki A S, Gardner A R, Smith T M, et al. (2003). Validation of fermentation processes. In F Brown, A S Lubineck (eds.), *Process Validation for Manufacturing of Biologics and Biotechnology Products, Devel. Biolog.* pp. 37–44, 111–112.
11. Mollah A H (2004). Risk analysis and process validation. *BioProcess. Internat.* 2(9):28–35.
12. Lubiniecki A S, Wiebe M E, Builder S E (1990). Process validation for cell culture-derived pharmaceutical proteins. *Bioproc. Technol.* 10:515–541.
13. Bliem R F (1995). Impact of research and development on validation, GMP and registration of biopharmaceuticals. *Pharmac. Engin.* 15(3):48–54.
14. Rathore A S, Noferi J F, Arling E R, et al. (2002). Process validation: How much to do and when to do it. *BioPharm.* 15:18–28.
15. Murphy R, Seeley R J (2001). Validation of biotechnology active pharmaceutical ingredients. In I R Berry, D Harpaz (eds.), *Validation of Active Pharmaceuticals Ingredients*: IHS Health Group, Englewood C. O., pp. 451–470.

16. Garnick R L, Solli N J, Papa P A (1988). The role of quality control in biotechnology: An analytical perspective. *Anal. Chem.* 60:2456–2557.
17. Jungbauer A, Boschetti E (1994). Manufacture of recombinant proteins with safe and validated chromatographic sorbents. *J. Chromatog. B: Biomed. Applic.* 662:143–179.
18. Ferreira J, Cooper D (1996). Process validation and drug cGMPs: A detailed analysis of the proposed revisions. *J. Valid. Technol.* 3:80–86.
19. Metz B, Hendriksen C F M, Jiskoot W, Kersten G F A (2002). Reduction of animal use in human vaccine quality control: Opportunities and problems. *Vaccine.* 20:2411–2430.
20. Gardner A R, Smith T M (2000). Identification and establishment of operating ranges of critical process variables. In G Sofer, D W Zabriskie (eds.), *Biopharmaceutical Process Validation*, Marcel Dekker, Inc.: Basel Switzerland, pp. 61–76.
21. Dellepiane N, Griffiths E, Milstien J B (2000). New challenges in assuring vaccine quality. *Bulletin of the W.H.O.*, 78(2):155–162.
22. Martin-Moe S, Kelsey W H, Ellis J, et al. (2000). Process validation in biopharmaceutical manufacturing. In G Sofer, D W Zabriskie (eds.), *Biopharmaceutical Process Validation*, Marcel Dekker, Inc.: Basel Switzerland, pp. 287–298.
23. FDC Reports (2005). Design space to facilitate quality initiative implementation. *The Gold Sheet.* 39(4):1.
24. Graffner C (2005). PAT-European regulatory perspective. *Proc. Analyt. Technol.* 2(26):8–11.
25. Chapman K G (1993). Validation terminology. In I R Berry, R A Nash (eds.), Pharmaceutical Process Validation, 2nd ed.: Marcel Dekker, Inc., New York, pp. 587–597.
26. O'Leary R M, Feuerhelm D, Peers D, et al. (2001). Determining the useful lifetime of chromatography resins. *BioPharm.* 14:10–18.
27. Zoon K, Garnick R (1998). Definition of a well-characterized biotechnology product. In F Brown, A Lubiniecki, G Murano (eds.), *Characterization of Biotechnology Pharmaceutical Products, Vol. 96, Dev. Biol. Stand.* Karger, Basel Switzerland, pp. 191–197.
28. Geigert J (1997). The challenge of managing quality control/quality assurance for loosely structured and virtual biopharmaceutical companies. *Drug Inform. J.* 31(1):97–100.
29. Lubiniecki A S (1987). Safety considerations for cell culture-derived biologicals. In B K Lydersen (ed.), *Large Scale Cell Culture Technology*: Hanser Publishers, New York, pp. 232–247.
30. Bobrowicz G (1999). The compliance costs of hasty process development. *BioPharm.* 12:35–38.
31. Welch K A (2003). Validation of checkweighers: A case study. *J. Validat. Technol.* 9(3):246–252.
32. Martin-Moe S, Ellis J, Coan M, et al. (2000). Validation of critical process input parameters in the production of protein pharmaceutical products: A strategy for validating new processes or revalidating existing processes. *J. Pharmac. Sci. Technol.* 54(4):315–319 and 54(5):423.
33. PDA (2005b). Process validation of protein manufacturing. PDA Technical Report No. 52, *PDA J. Pharmac. Sci. Technol.* 59(S-4):1–28.
34. Seely R J, Hutchins H V, Luscher M P, et al. (1999). Defining critical variables in well-characterized biotechnology processes. *BioPharm.* 12(4):33–36.
35. Kelley B D (2000). Establishing process robustness using designed experiments. In G Sofer, D W Zabriskie (eds.), *Biopharmaceutical Process Validation*: Marcel Dekker, New York, pp. 29–59.

36. Pujar N S, Gayton M G, Herber W, et al. (2005). Process validation of a multivalent bacterial vaccine: A novel matrix approach. In A S Rathore, G Sofer (eds.), *Process Validation*, Taylor and Francis, New York, pp. 529–550.
37. Welch K A, Fung C A, Schmidt S R (2004). Risk/science-based approach to validation: A win-win-win for patients, regulators, and industry. *PDA J. Pharmac. Sci. Technol.* 58(1):15–23.
38. Sofer G (1994), Validation of chromatography processes. In L Alberghina, L Frontali, P Sensi (eds.), *Proc. of the 6th European Congress on Biotechnology*, Elsevier Science B. V., Florence, Italy, pp. 977–980.
39. Sniff K S, Garcia L A, Hassler R A, et al. (1996). Process development, scale-up and validation of a recombinant *E. coli*-based fermentation process. *ASTM Spec. Tech. Publ., STP 1260*: 97–106.
40. Zhang J, Reddy J, Salmon P, et al. (1998). Process characterization studies to facilitate validation of a recombinant protein fermentation. In B D Kelley, R A Ramelmeier (eds.), *Validation of Biopharmaceutical Manufacturing Processes. ACS Symposium Series.* 698:12–27.
41. Darling A, Spaltro J J (1996). Process validation for virus removal: Considerations for design of process studies and viral assays. *BioPharm.* 9(9):42–50.
42. Brorson K, Krejci S, Lee K, et al. (2003b). Bracketed generic inactivation of rodent retroviruses by low pH treatment for monoclonal antibodies and recombinant proteins. *Biotechnol. Bioeng.* 82:321–329.
43. Sofer G (2002a). Virus inactivation in the 1990s- and into the 21st century, Part 2: Red blood cells and platelets. *BioPharm.* 15:42–49.
44. Sofer G (2002b). Virus inactivation in the 1990s- and into the 21st century, Part 3a: Plasma and plasma products (heat and solvent detergent treatments). *BioPharm.* 15:28–42.
45. Sofer G (2003b). Virus inactivation in the 1990s- and into the 21st century, Part 4: Culture media, biotechnology products, and vaccines. *BioPharm.* 16:50–57.
46. Lau A S L, Lie Y S, Norling L A, et al. (1999). Quantitative competitive reverse transcription-PCR as a method to evaluate retrovirus removal during chromatography procedures. *J. Biotechnol.* 75:105–115.
47. Walter J, Allgaier H (1997). Validation of downstream processes. In R Wagner, H Hauserv (eds.), *Mammalian Cell Biotechnology in Protein Production*: Walter De Gruyter Inc., New York, pp. 453–482.
48. Hirasaki T, Tsuboi T, Noda T, (1992). Removability of virus particles and permeability of protein from cell culture medium using cuprammonium regenerated cellulose hollow fiber (BMM™). In H Murakami, S Shirahata, H Tachibana (eds.), *Animal Cell Technology: Basic & Applied Aspects*, Kluwer Academic Publishers, The Netherlands, pp. 49–55.
49. PDA (2005a). Virus filtration- Introduction. *PDA J. Pharm. Sci. Technol.* 59 (S-2): 8–42.
50. Siklosi T (2005). Manufacturing of biopharmaceutical proteins. *Genetic Engin. News.* 25(17):58–59.
51. Levine H L, Tarnowski S J, Dosmar M, (1992). Industry perspective on the validation of column-based separation processes for the purification of proteins. *J. Parental Sci. Tech.* 46(3):87–97.
52. Sofer G (1996) Validation: Ensuring the accuracy of scaled-down chromatography models. *BioPharm.* 9(9):51–55.
53. Hesse F, Wagner R (2000). Developments and improvements in the manufacturing of human therapeutics with mammalian cell cultures. *Tibtech.* 18:173–180.

54. Garnick R L (1998). Raw materials as a source of contamination in large-scale cell culture. In F Brown, E Griffiths, F Horaud, J C Petricciani (eds.), *Safety of Biological Products Prepared from Mammalian Cell Culture, Vol. 93, Dev. Biol. Stand.* Karger, Basel Switzerland, pp. 21–29.
55. Facklam T J, Geyer S. The preparation and validation of stock cultures of mammalian cells. In Y H Chui, J L Gueriguian (eds.), Drug Biotechnology Regulation: Scientific Basis and Practices. Marcel Dekker, Inc., New York, 1991, pp. 54–85.
56. Geigert J (1989). Overview of the stability and handling of recombinant protein drugs. *J. Parenteral Sci. Technol.* 43(5):220–224.
57. Bennan J, Bing F, Boone H, et al. (2002). Evaluation of extractables from product-contact surfaces. *BioPharm Internat.* 22–34.
58. Seely R J, Wight H D, Fry H H, et al. (1994). Biotechnology product validation, Part 7: Validation of chromatography resin useful life. *BioPharm.* 15:7:41–48.
59. Depalma A (2006). Bioprocessing: Bright sky of single use bioprocess products. *GEN.* 26(3):1.
60. Webb S D, Webb J N, Hughes T G, et al. (2002). Freezing biopharmaceuticals using common techniques—and the magnitude of bulk-scale freeze-concentration. *BioPharm.* 15:22–34.
61. Breece T N, Gilkerson E, Schmelzer C (2002a). Validation of large-scale chromatographic processes, Part 1: Case study of neuleze capture on macroprep high-S. *BioPharm.* 15:16–20.
62. Brorson K, Brown J, Hamilton E, et al. (2003a). Identification of protein A media performance attributes that can be monitored as surrogates for retroviral clearance during extended use. *J. Chrom. A.* 989:155–163.
63. Norling L, Lute S, Emery R, et al. (2005). Impact of multiple re-use of anion exchange chromatography media on virus removal. *J. Chrom. A.* 1069:79–89.
64. Brough H, Antoniou C, Carter J, et al. (2002). Performance of a novel Viresolve NFR virus filter. *Biotechnol. Prog.* 18:782–795.
65. Seely J E, Seely R J (2003). A rationale, step-wise approach to process characterization. *BioPharm Internat.* 16: 24–34.
66. Krause S O (2005). How to maintain suitable analytical test methods: Tools for ensuring a validation continuum. *BioPharm Internat.* 18(10):52–60.
67. Seely R J, Haury J (2005). Applications of failure modes and effects analysis to biotechnology manufacturing processes. In A S Rathore, G Sofer (eds.), *Process Validation in Manufacturing of Biopharmaceuticals*: Taylor and Francis, New York, pp. 13–29.
68. Seely J E (2005). Process characterization. In A S Rathore, G Sofer (eds.), *Process Validation in Manufacturing of Biopharmaceuticals*: Taylor and Francis, New York, pp. 31–67.
69. Kell D B, Sonnleitner B (1995). GMP- good modelling practice: An essential component of good manufacturing practice. *Tibtech.* 13:481–492.
70. Noble P T (2001). Reduction of risk and the evaluation of quality assurance. *PDA J. Pharmac. Sci. Technol.* 55(4):235–239.
71. Sahni A (1993). Using failure mode and effects analysis to improve manufacturing processes. *Med. Dev. Diagn. Ind.* 15:47–51.
72. Kieffer R G, Bureau S, Borgmann A (1997). Applications of failure mode effect analysis in the pharmaceutical industry. *Pharma. Technol. Europe.* 10:36–49.
73. Mollah A H (2005). Application of failure made and effect analysis (FMEA) for process risk assessment. *BioProcess Internat.* 3:12–21.

74. Armbruster D, Feldsien T (2000). Applying HACCP to pharmaceutical process validation. *Pharmac. Technol.* 13:170–178.
75. Little L E (1995). Validation of immunological and biological assays. *BioPharm.* 36–42.
76. Xu Y, Brorson K (2003). An overview of quantitative PCR assays for biologicals: Quality and safety evaluation. In F Brown, A S Lubiniecki (eds.), *Process Validation for Manufacturing of Biologics and Biotechnology Products*, Basel, Karger Switzerland, pp. 89–98.
77. Federici M M, Garnick R L, Geigert J (1991). A perspective on reference standard and reference material requirements in biotechnology-derived pharmaceutical protein products. *Pharmacopeial Forum.* 17:2683–2687.
78. Bozzo T (1995). Biotechnology product validation, Part 9: A former FDAer's view. *BioPharm.* 8:35–38.
79. Moran E B, McGowan S T, McGuire J M, et al. (2000). A systematic approach to the validation of process control parameters for monoclonal antibody production in fed-batch culture of a murine myeloma. *Biotechnol. Bioengin.* 69(3):242–255.
80. Rathore A S, Johnson G V, Buckley J J, et al. (2003). Process characterization of the chromatographic steps in the purification process of a recombinant *Escherichia coli*-expressed protein. *Biotechnol. Appl. Biochem.* 37:51–61.
81. Rathore A S (2005). Process optimization and characterization studies for purification of an *E.coli*-expressed protein product. In A S Rathore, G Sofer (eds.). *Process Validation in Manufacturing of Biomanufacturing of Biopharmaceuticals*, Taylor and Francis, New York, pp. 451–468.
82. Francis R, Bonnerjea J, Hill C R (1990). Validation of the re-use of protein a sepharose for the purification of monoclonal antibodies. In D L Pyle (ed.), *Separations for Biotechnology 2*, Elsevier Applied Science, New York, pp. 491–498.
83. Kelley B D, Tannatt M, Magnusson R, et al. (2004). Development and validation of an affinity chromatography step using a peptide ligand for cGMP production of factor VIII. *Biotechnol. Bioengin.* 87(3):400–412.
84. Breece T N, Gilkerson E, Schmelzer C (2002b). Validation of large-scale chromatographic processes, Part 2: Results from the case study of neuleze capture on macroprep high-S. *BioPharm.* 15:35–42.
85. Brady D, Bonnerjea J, Hill C R (1990). Purification of monoclonal antibodies for human clinical use: Validation of DNA and retroviral clearance. In D L Pyle (ed.), *Separations for Biotechnology 2*, Elsevier Applied Science, New York, pp. 472–480.
86. Johnston A, MacGregor A, Borovec S, et al. (2000). Inactivation and clearance of viruses during the manufacture of high purity factor IX. *Biologicals.* 28:129–136.
87. Conley L, McPherson J, Thommes J (2005). Validation of the ZEVALIN® purification process- A case study. In A S Rathore, G Sofer (eds.), *Process Validation*, Taylor and Francis, New York, pp. 469–521.
88. McGurk T L (2004). Ramping up and ensuring supply capability for biopharmaceuticals. *BioPharm. Internat.* 17(1):38–44.
89. Seely R J, Munyakazi L, Haury J (2001). Statistical tools for setting in-process acceptance criteria. *BioPharm.* 14:28–34.
90. Martin J (2005). Turning compliance challenges into competitive advantages. *Chem. Engin.* 112(8):23–26.
91. Curtis S, Lee K, Blank G S, et al. (2003). Generic/matrix evaluation of SV40 clearance by anion exchange chromatography in flow-through mode. *Biotechnol. Bioengin.* 84(2):179–186.

92. Godarvarti R, Petrone J, Robinson J, et al. (2005). Scale-down models for purification processes: Approaches and applications. In A S Rathore, G Sofer (eds.), *Process Validation*: Taylor and Francis, New York, pp. 69–142.
93. Zoon K C (1998). Well-characterized biotechnology products: Evolving to meet the needs of the 21st century. In F Brown, A Lubiniecki, G Murano (eds.). *Characterization of Biotechnology Pharmaceutical Products, Dev. Biol. Stand.*, Basel Switzerland, Karger, 96:3–8.
94. Webber K (2005). FDA Update: Process analytical technology for biotechnology products. *J. Process Analytical Technol.* 2(4):12–14.
95. Montague G, Morris J (1994). Neural-network contributions in biotechnology. *Tibtech.* 12:312–324.
96. Rockwell Automation (2004). PAT initiative expected to invigorate pharmaceutical industry with improved quality, better efficiency, and improved profits. White Paper. Publication LIFE-WP001A-EN-P, Nov.
97. Geismar D, White G (2004). Process reliability and flexibility—A tool to improve pharmaceutical plan floor operations. *Pharm-Ind*, 11a:1399–1403.
98. Katsnelson A (2005). Teva to acquire Chinese biogenerics manufacturer. *Nat. Biotechnol.* 23(7):765.
99. Polastro E T, Tulcinsky S (2004). Are biopharmaceuticals immune to the generics threat? *J. Gene. Med.* 1(3):259–265.
100. Henry Steward Publications (2005). An overview of the scientific and regulatory issues facing biogenerics in the USA today. *J. Gener. Med.* 2(2):169–172.
101. Ainsworth S J (2005). Biopharmaceuticals: Patent expirations are beckoning generic drug companies, but numerous hurdles remain to a profitable business. *Chem. Engin. News.* 83(23):21–29.
102. Hempel J C, Hess P N (2005). Contract manufacturing of biopharmaceuticals including antibodies or antibody fragments. In J Knablein (ed.), *Modern Biopharmaceuticals*, Wiley-VCH, Weinheim, Germany, 1083–1107.
103. Kulkarni M (2005). Outsourcing QA: Perspectives on future strategies. *Contract Pharma*. 7(9):58–61.
104. Fox S (2004). Maximizing outsourced biopharma production. *Contract Pharma.* 6: 72–78.
105. Ryan C S, Wan M (2000). Risk management and liability for biologics. *BioPharm.* 13:40–42.
106. Shenoy P (2004). Risk based cGMP—A new initiative by USFDA *PharmaTimes.* 36(2):23–24.

## FURTHER READING

Brown F, Lubiniecki A S (eds.) (2003). *Process Validation for Manufacturing of Biologics and Biotechnology Products*, Basel, Karger Switzerland.

Kelley B D, Ramelmeier R A (eds.) (1998) Validation of Biopharmaceutical Manufacturing Processes. ACS Symposium Series 698, San Francisco, California.

Rathore A S, Sofer G (eds.), (2005). *Process Validation in Manufacturing of Biopharmaceuticals*, Taylor and Francis, New York.

Sofer G, Zabriskie D W (eds.) (2000). *Biopharmaceutical Process Validation*, Marcel Dekker, Inc., Basel Switzerland.

# STABILITY ASSESSMENT AND FORMULATION CHARACTERIZATION

Albert H.L. Chow,[1] Henry H.Y. Tong,[2] and Ying Zheng[3]

[1] *The Chinese University of Hong Kong, Shatin, N.T., Hong Kong, China*
[2] *Macao Polytechnic Institute, Macao, China*
[3] *University of Macao, Macao, China*

**Chapter Contents**

1 Chemical Instability and Potential Degradation Pathways
- 1.1 Hydrolytic Reactions in Proteins and Peptides
- 1.2 Hydrolytic Reactions in Nucleic Acids
- 1.3 Oxidation
- 1.4 β-Elimination in Nucleic Acids, Proteins and Peptides
- 1.5 Maillard Reaction in Nucleic Acids, Proteins, and Peptides
- 1.6 Covalent Dimerization and Polymerization in Proteins

2 Physical Instability
- 2.1 Denaturation
- 2.2 Aggregation and Precipitation
- 2.3 Adsorption

3 Considerations in the Stability Testing of Biopharmaceutcals
- 3.1 Temperature
- 3.2 Moisture
- 3.3 Light and Ionizing Radiation
- 3.4 Physical Stress
- 3.5 Freeze-Drying and Spray-Drying
- 3.6 Excipients

4 Regulatory Requirements in the Stability Testing of Biopharmaceuticals

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

5 Formulation and Characterization of Biopharmaceuticals
5.1 Protein and Peptide Drugs
5.2 Nucleic Acid-Based Drugs
6 Summary and Conclusions
References

A drug must be formulated in a safe, consistent, and stable dosage form with proven clinical efficacy before it can be approved for specific therapeutic indications by drug regulatory authorities. Although biotechnological/biological products or biopharmaceuticals (therapeutic proteins, peptides, and nucleic acids) are no exception to this requirement, the inherently complex primary structures of this important class of therapeutic agents coupled with their ability to form higher order structures have posed additional challenges and constraints in their formulation and assessment.

The current chapter is intended to provide an overview of the stability problems of biopharmaceuticals and their assessment as well as the formulation approaches that can be used to circumvent these problems. Presented next is a literature review on the known degradation/denaturation mechanisms of various biopharmaceuticals, followed by a general discussion on the formulation strategies using specific excipients/additives for improving the product shelf-life.

## 1 CHEMICAL INSTABILITY AND POTENTIAL DEGRADATION PATHWAYS

### 1.1 Hydrolytic Reactions in Proteins and Peptides

***Deamidation.*** Deamidation is a common degradation pathway for proteins and peptides, involving the conversion of the side chains of the asparagine and glutamine residues to the respective carboxylates of aspartate and glutamate. The reaction is highly pH dependent. Under acidic conditions, the reaction is initiated by the direct intermolecular attack of water at the side-chain amide linkage, yielding ammonia and the corresponding acids (i.e., aspartic acid for asparagine residue or glutamic acid for glutamine residue) [1–3]. At neutral or basic pH, the degradation is predominated by intramolecular deamidation with a relatively minor contribution from direct hydrolysis [2].

Mechanistically, deamidation of asparagine proceeds via the formation of a succinimide intermediate, involving the intramolecular attack by the nitrogen in the peptide bond on the C-terminus (Scheme 1) [2, 4]. The resulting cyclicimide tends to racemize at α-carbon. Hydrolysis of the cyclicimide yields a mixture of peptide products differing in the type of amide linkage in the polypeptide backbone; for the α-carboxyl linkage, the hydrolytic products are a mixture of *d*- and *l*-normal aspartyl peptides, whereas the β-carboxyl linkage affords a mixture of *d*- and *l*-isoaspartyl transpeptidation products [3, 5]. Spontaneous deamidation of the polypeptide chain at the asparagine residue can also occur via the attack by the nitrogen

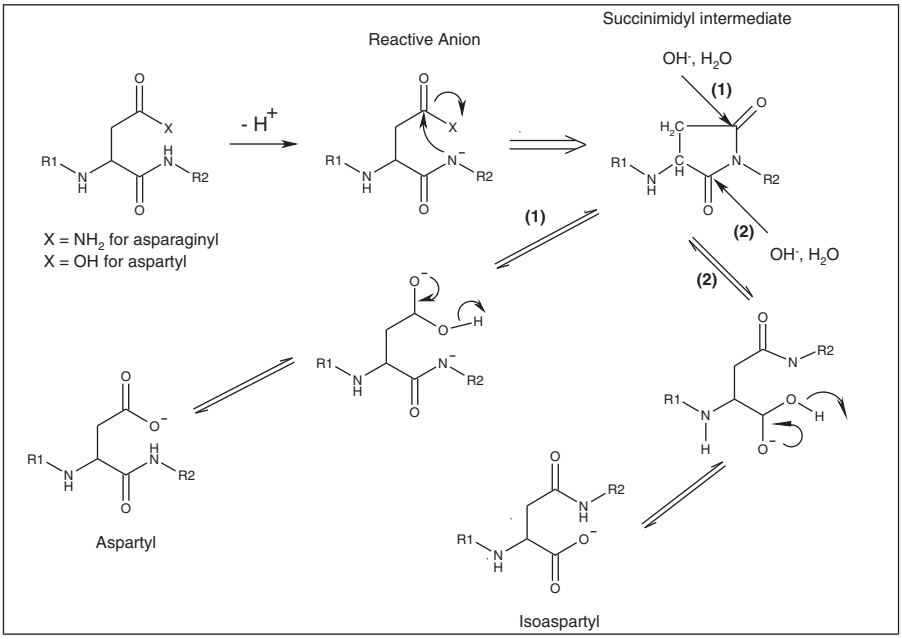


**Scheme 1.**

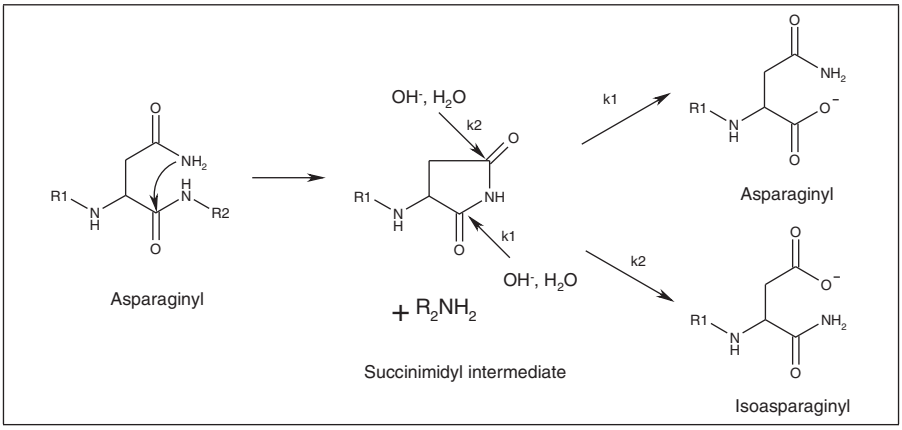


**Scheme 2.**

in the side chain of the amide on the following peptide-bond carbonyl to form two polypeptides, one of which contains a C-terminal succinimide residue (Scheme 2) [3]. Most biopharmaceuticals are formulated at physiological pH (7.4) for parenteral administration and thus are prone to decomposition by deamidation via the aforementioned mechanisms. These products encompass recombinant human parathyroid hormone [6], orthoclone OKT3 [7], human epidermal growth factor [8], insulin [9], antiflammin 2 [10], recombinant tissue plasminogen activator [11], monoclonal immunoglobulin h1B4 [12], and murine monoclonal antibody MMA383 [13]. Adenovirus, a therapeutic viral gene carrier, is also known to degrade by

deamidation and isoaspartate formation at four susceptible asparagine residues, and the stability of a recombinant adenovirus AV3760 has demonstrated considerable improvement after site-directed mutagenesis of the key asparagine residues in the hexon protein [14].

Similar to the asparagine residue (Scheme 1), glutamine residue undergoes deamidation at neutral or alkaline pH, where intramolecular attack results in glutarimide formation. Having one more methylene group than the asparagine residue, the glutamine residue yields a six-membered ring intermediate structure, which is formed at a much slower rate than that of the five-membered succinimide intermediate for the asparagine residue because of entropic factors [1, 11]. Although the asparagine residue is widely known to be more susceptible to deamidation than the glutamine residue, more facile deamidation for the glutamine residue under acidic conditions has been observed with the glucagon fragment 22–29 [15]. The lower reactivity of the asparagine residue in this case has been attributed to a decreased electrophilicity of its side-chain carbonyl carbon imparted by a parallel cleavage pathway at this residue.

Primary sequence has an important bearing on the spontaneous deamidation in peptides. It has been demonstrated using model peptides that the structure of amino acid residue immediately following the asparagine residue N + 1 exerts a significant impact on the overall rate of succinimide formation. For example, glycine residue in the N + 1 position results in a deamidation rate 70- to 100-fold faster than that found for amino acid residues with branched hydrophobic side chains [2]. In another study, the bulky side chain of N + 1 histidine residue has been found to limit the flexibility of the model peptide around the reaction site, thereby reducing the reaction rate [3]. As isoaspartate formation depends on both primary sequence and chain flexibility, two important generalizations can be made [11]. First, isoaspartate tends to form preferentially at the asparagine-glycine, asparagine-serine, aspartic acid-glycine, and possibly asparagine-histidine sites. Second, isoaspartate formation is favored in the highly flexible region of a polypeptide chain. However, it must be noted that these generalizations, which are based on solution-state data, may not be applicable to solid peptides because the mechanisms of deamidation involved in the solution and solid states are intrinsically different. As has been demonstrated in the solution state, a glutamine residue in the N + 1 position can either hinder asparagine deamidation by electrostatic and inductive effects or enhance the reaction though general acid/base catalysis, whereas in the solid form, a glutamine residue in the N + 1 position can increase the hydration of the peptide and afford asparagine deamidation rates similar to those in solution [16].

Higher order structures, viz. secondary, tertiary, and quaternary structures, also influence deamidation in proteins by positioning the asparagine or glutamine residue in spatial proximity to a functional group elsewhere in the primary sequence, which is capable of accelerating or inhibiting the deamidation. In addition, the conformational structure of the asparagine or glutamine residue and the nearby peptide region containing it may be strongly conditioned by the nature of the structural environment that incorporates the residue [17]. In general, α-helices, β-turns, and possibly β-sheets can stabilize asparagine residues against deamidation and drastically reduce the rate of succinimide formation for a given sequence relative to the same sequence in a random coil [11]. However, the effect is relatively weak with α-helices and destabilization of the asparagine residues may even occur

[17]. Conformational restrictions and reduced nucleophilic reactivity of the backbone NH centers have been suggested as probable stabilizing mechanisms with the higher order structures. In recombinant human lymphotoxin, asparagine deamidation can be inhibited by secondary and tertiary structures due to reduced polypeptide flexibility [18].

Apart from protein sequence and structure, temperature, pH, and the type of buffers can all influence the deamidation and isoaspartate formation. The effect of temperature on the deamidation reaction rate generally follows the Arrhenius law. Deamidation activation energies around 21~22 kcal $mol^{-1}$ have been reported for two model peptides under alkaline conditions [19, 20]. The deamidation is also subject to general acid/base catalysis, as evidenced by an increase in deamidation rate with an increase in buffer concentration [2].

In stability assessment, chromatographic methods of high sensitivity are normally required to analyze the degradation products formed from parent drugs in trace amounts. Biological potency is not a sensitive marker for detecting deamidation and isoaspartate formation, which may or may not influence the biological activities of protein and peptides. If the asparagine and glutamine residues are not involved in receptor binding, there may not be any effect on biological activity upon deamidation, as in the case of human growth hormone [21], insulin [22, 23], recombinant human flial cell line derived neurotrophic factor [24], interleukin-1α [25], and recombinant human interleukin-2 [26]. However, activity loss associated with deamidation is probably the norm rather than the exception, particularly if the asparagine or glutamine residues take part in receptor binding, as exemplified by calmodulin, CD4, human and murine epidermal growth factors, growth hormone releasing factor, methionyl murine interleukin-1β, parathyroid hormone, and bovine pancreatic ribonucleases A [25]. It is important to note that in addition to the direct effect of receptor binding on activity, modification of a site outside the binding region may affect the biological activity indirectly through a local conformational change, and isoaspartate formation can also elicit autoimmunity against the proteins/peptides [11].

***Peptide Bond Cleavage at Aspartic Acid Residues.*** The breakdown of proteins and peptides may occur by cleavage at the amide bonds. Although all amide bonds are theoretically susceptible to hydrolysis, they normally exhibit different hydrolytic rates. For instance, the rate of acid-catalyzed hydrolysis in polypeptides containing the aspartic acid residues are at least 100 times higher than those of the other peptide bonds [3, 4, 27]. The cleavage proceeds via either the *N*- or *C*-terminal bond adjacent to the aspartic acid residue (Scheme 3) [4]. For example, in acidic solutions, antiflammin 2 undergoes both *C*- and *N*-terminal hydrolyses across the aspartyl peptide bonds, and the peptide bonds at C-termini of the aspartyl residues are most susceptible to hydrolysis [28]. In addition, isoaspartate formation with aspartic acid residue is possible via the cyclicimide intermediate, as shown in Schemes 1 and 2. The influence of primary sequence on aspartic acid degradation is significant, but probably not as obvious as that observed for the asparagine residue deamidation [3]. The N + 1 glycine, which confers the greatest flexibility on the polypeptide chain, has been shown to be most vulnerable to hydrolysis at the aspartic acid residue in model peptides, which are structurally analogous to adrenocorticotrophin fragment 22–27 [29]. In a similar series of hexapeptides, it

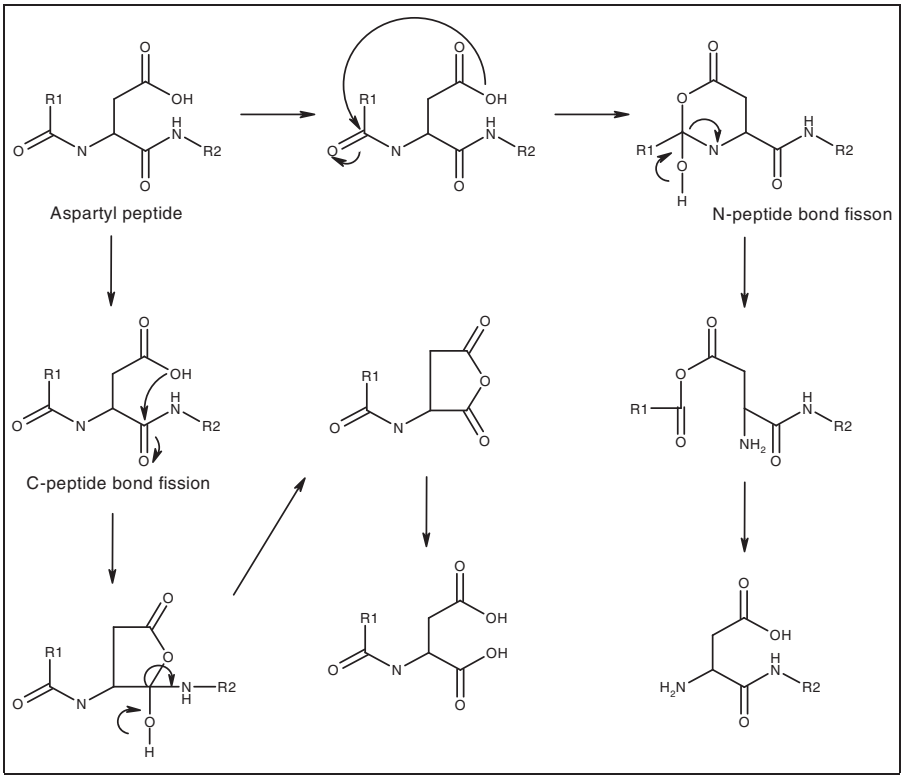


**Scheme 3.**

has been observed that the rate of hydrolysis decreases with an increase in steric hindrance of the N + 1 residue [30].

It is well established that aspartic acid-proline peptide bonds are particularly labile and undergo hydrolysis under conditions where other aspartic acid peptide bonds are stable, e.g., at low pHs [4, 31]. The hydrolysis proceeds via intramolecular catalysis by carboxylate anion displacement of the protonated nitrogen of the peptide bond [31]. This mechanism has been demonstrated for recombinant human interleukin 11, which undergoes peptide cleavage specifically at aspartic acid[133]-proline[134] residues in acidic solution [32]. The increase in reaction rate, by approximately 8- to 20-fold, may be ascribed to the weakened amide bond resulting from the increased electronegativity of the protonated proline nitrogen [4, 27].

As with aspartic acid containing amide bonds, peptide bonds carrying serine or threonine are generally more labile than other amide bonds. The degradation involves the nucleophilic attack of the serine or threonine hydroxyl group on the neighboring amide bond via the formation of a cyclic tetrahedral intermediate [27]. This mechanism exhibits predominance in luteinizing hormone-releasing hormone antagonist RS-26306 at pH ~5 [33].

It is worth noting that hydrolysis of the peptide bond can also occur with other amino acids. For example, salmon calcitonin undergoes hydrolysis at the cysteine[1]-serine[2] peptide bond where the cysteine remains linked to the peptide by a disulfide bond [34].

**N-*terminal Degradation via Diketopiperazine and Pyroglutamic Acid Formation.*** If the proteins and peptides possess a penultimate proline residue in an *N*-terminal sequence, the first amino acid in the primary sequence may be lysed from the polypeptide skeleton via the formation of diketopiperazine (DKP). This reaction is commonly termed nonenzymatic aminolysis, which results from the intramolecular nucleophilic attack of the *N*-terminal nitrogen on the carbonyl carbon of the peptide bond between the second and the third amino acid residues (Scheme 4) [3]. Pharmaceutical materials known to follow this degradation pathway include substance P [35], recombinant human growth hormone [36], histrelin [37], RMP-7 (a bradykinin analog) [38], as well as aspartame and aspartylphenylalanine in the solid state [39].

The rate of DKP formation generally increases with increasing temperature, as shown for RMP-7 [38] and phenylalanine-proline-*p*-nitroaniline (Phe-Pro-*p*-NA) [40]. Avoidance of general base catalysis or decreasing the buffer concentration at high pHs can minimize this cyclization reaction [38, 40]. The degree of ionization of the *N*-terminal group, which is governed by the $pK_a$ of the amino group and the pH of the medium, also affects the rate of DKP formation because the first protonated amino acid in the sequence will not be available to effect base-catalyzed reaction. However, with Phe-Pro-*p*-NA and its analog, X-Pro-*p*-NA (where X is the other amino acid), the influence of primary sequence on the rate of DKP formation appears to be mediated through the ability of the X-Pro peptide to undergo *cis-trans* isomerization rather than through the difference in extent of ionization or in steric hindrance of the *N*-terminal amino acids [3, 40].

Pyroglutamic acid formation is a result of the condensation of either the amino group with γ–carboxylic acid group of glutamic acids or the amino group with γ–carboxamide group of glutamines at the *N*-terminal positions (Scheme 5) [27]. It

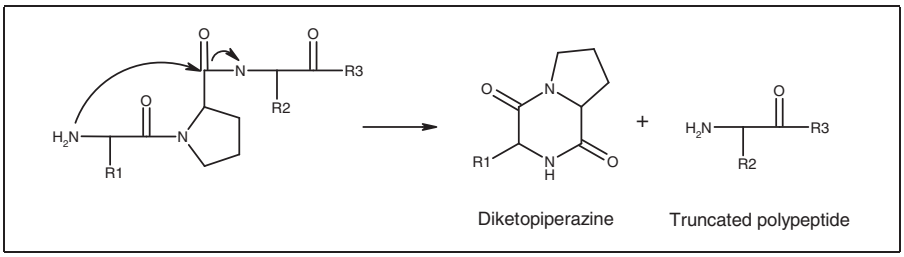


**Scheme 4.**

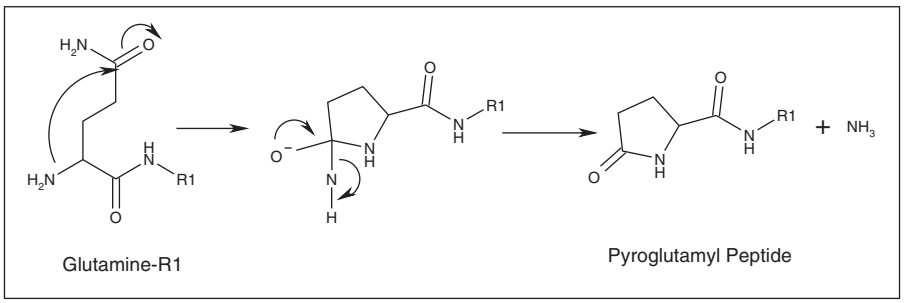


**Scheme 5.**

has been shown with model peptides that the rate of pyroglutamic acid formation is the fastest when the N + 1 amino acid residue next to glutamine is asparagine, leucine, or glycine [41]. Pyroglutamic acid formation in proteins and peptides is usually accompanied by a loss of biological activity. For example, decapeptide ELA, a Melan-A/MART-1 antigen immunodominant peptide analog, exhibits reduced binding to the specific class I major histocompatibility complex and a loss of cytotoxic T-lymphocyte activity after conversion of its *N*-terminal glutamic acid to pyroglutamic acid [42]. However, several peptides of pharmaceutical interest (including buserelin and gonadorelin) have been developed with the chemically more stable pyroglutamic acid instead of N-terminal glutamine or glutamic acid without loss of pharmacological activities.

## 1.2 Hydrolytic Reactions in Nucleic Acids

***Depurination and Depyrimidation.*** Depurination, which refers to the hydrolysis of the *N*-glycosidic bond, is an important degradation pathway for nuclei-acid-derived drugs. By inhibiting free radical oxidation, depurination and β-elimination (Section 1.4) represent the most significant pathways for the degradation of DNA [43]. Due to the inductive effect of the 2′-OH group, deoxyribonucleosides and deoxyribonucleotides in DNA are more susceptible to depurination than are the ribonucleosides and ribonucleotides in RNA. The hydrolysis of *N*-glycosidic bond in DNA can be acid- (Scheme 6) or base-catalyzed (Scheme 7) [44]. The acid-catalyzed reaction pathway involves the protonation of the respective base, followed

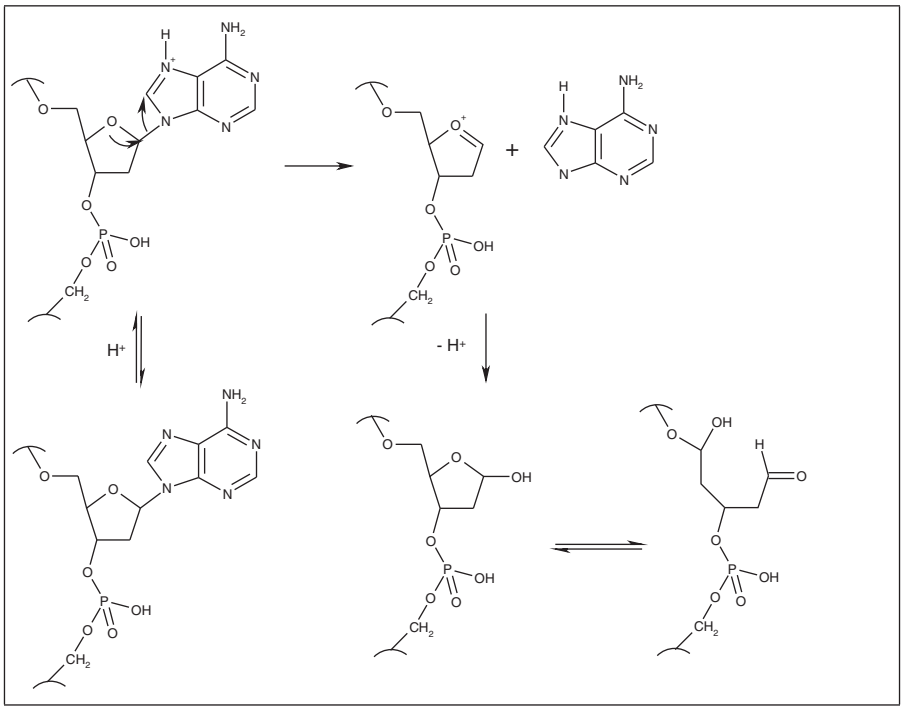


**Scheme 6.**

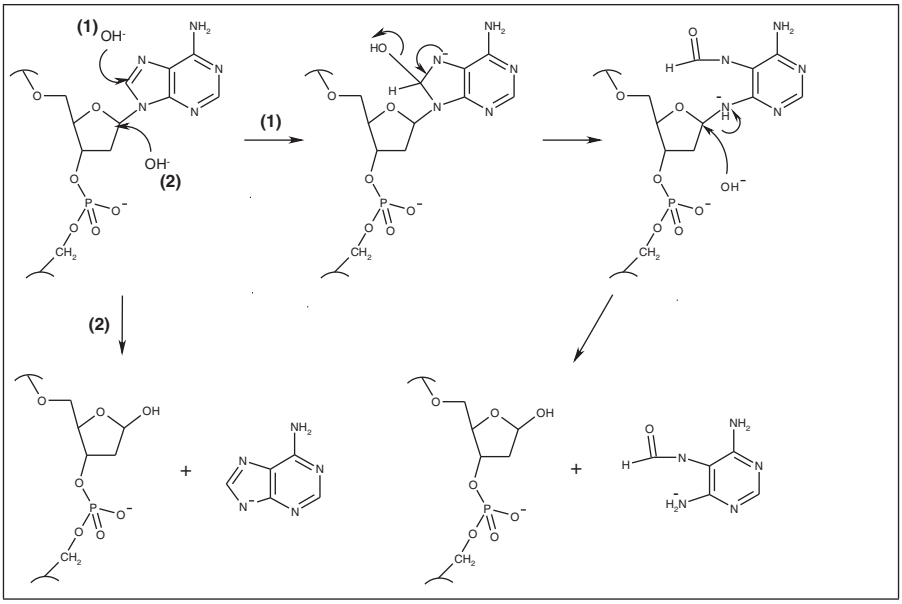


**Scheme 7.**

by a rate-limiting cleavage step to yield the C-1′ carbonium ion (oxocarbonium ion) (Scheme 6). Hence, the stability of the *N*-glycosidic bond can be enhanced or reduced by destabilizing or stabilizing the C-1′ carbonium ion being formed. The base-catalyzed hydrolytic reaction pathway (Scheme 7) proceeds by two $S_N2$ mechanisms at either C-8 or C-1′, which are dependent on temperature and the hydroxide ion concentration [44]. Depurination is more rapid with guanine and adenine, whereas the loss of cytosine and thymine (depyrimidation) proceeds at 5% of the depurination rate [45]. The double helical structure in DNA can offer only limited protection against depurination, as suggested by the mere 4-fold difference in depurination velocity observed between single-stranded and double-stranded DNAs. Introduction of a single break in the DNA backbone is capable of converting supercoiled plasmid DNA to the open circular form, and thus monitoring of this conversion affords a convenient and sensitive means of determining DNA stability [46]. For instance, V1JpHA plasmid DNA at 50°C and pH 7.4 underwent depurination and β-elimination with respective rate constants of $5.6 \times 10^{-11}\,s^{-1}$ and $1.4 \times 10^{-6}\,s^{-1}$, reflecting an accumulation of ~0.5 apurinic sites per plasmid after 2 years of storage at room temperature and a corresponding decrease of ~20% in supercoiled DNA content [43].

***Hydrolysis of the Phosphoester Bond.*** The hydrophilic surface of multistranded oligonucleotides on the sugar-phosphate backbones is a prominent target for hydrolysis. The nucleophilic phosphodiester cleavage can be mediated through both intermolecular and intramolecular reactions. DNA, being devoid of the 2′-OH group, is much less susceptible to base-catalyzed hydrolysis than RNA [44]. As shown in Scheme 9, the intramolecular displacement of the 5′-linked nucleoside by the 2′-OH group results in the phosphodiester cleavage. In acid-catalyzed

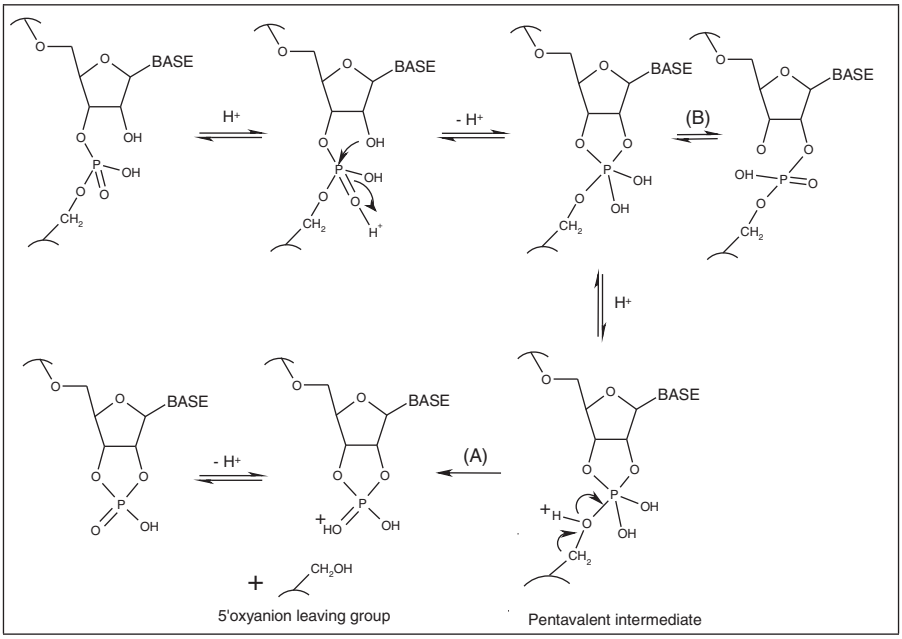


**Scheme 8.**

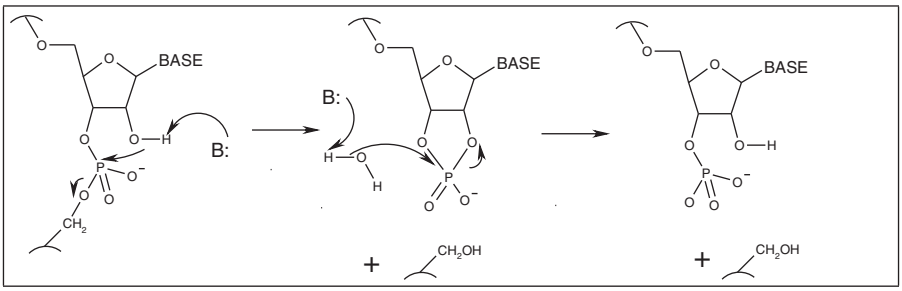


**Scheme 9.**

hydrolysis, isomeri-zation may also be involved (Scheme 8) [44]. Generally, pyrimidine nucleoside 3′-phosphodiesters react faster than their purine counterparts.

The hydrolysis of phosphoester bond can be accelerated by several catalytic factors, including enhancement of the nucleophilicity of the 2′-oxygen (Scheme 9), stabilization of the leaving 5′ oxyanion group (Scheme 8), stabilization of the pentavalent intermediate (Scheme 8), and the presence of buffers (e.g., imidazole, morpholine, and carboxylates) and divalent or trivalent metal ions (e.g., $Mg^{2+}$, $Zn^{2+}$, $Ca^{2+}$, $Cu^{2+}$, and $Fe^{3+}$) [44].

***Deamination of Pyrimidine and Purine Nucleosides.*** Deamination of pyrimidine nucleosides into uridines is a common hydrolytic degradation for nucleic acid-derived drugs under both acidic and basic conditions. The main targets of this

hydrolytic reaction are cytosine and its homologue 5-methylcytosine [45], whereas adenine and guanine are more prone to undergo deamination via the dediazoniation pathways [47]. In DNA, cytosine can be converted to uracil via direct deamination by base-catalyzed hydrolysis and parallel attack of a water molecule on the protonated base [44]. Consequently, the hydrolysis shows no strong pH dependence around pH 7.4. The reaction for single-stranded DNA is 150- to 200-fold more rapid than that for double-stranded DNA, reflecting the considerably better protection offered by the double helical structure [44, 47].

Purine nucleosides behave similarly to their pyrimidine counterparts, but their contribution to the overall degradation is insignificant [44]. For instance, adenine is converted to hypoxanthine in single-stranded DNA at only 2–3% of the rate of cytosine deamination [48].

### 1.3 Oxidation

Oxidation is one major chemical degradation pathway for biopharmaceuticals. The reaction involves reactive oxygen species such as hydroxyl radical (OH), hydrogen peroxide ($H_2O_2$), superoxide ($O_2$), and singlet oxygen ($^1O_2$), which can be generated by autooxidation, photoactivation, and metal catalyzed oxidation. Autooxidation refers to the oxidation in the sheer absence of oxidants. True autooxidation, being characterized by extremely slow reaction rate, is normally not regarded as an important degradation pathway for proteins [5, 49]. Oxidation can be catalyzed by various external factors, such as the presence of transition metal ions and ultraviolet (UV) irradiation [5, 49]. Among them, Fenton-type metal catalyzed oxidation is the most relevant generating source of reactive oxygen species in pharmaceutical formulations where contaminating peroxides may react with traces of redox-active transition metals such as iron or copper.

***Metal Catalyzed Oxidation.*** Metal catalyzed oxidation is known to adversely influence the stability of proteins, peptides, plasmid DNAs, and nucleic acid-derived drugs by inducing conformational and functional changes in these biopharmaceuticals [43, 44, 49, 50]. In proteins and peptides where metal catalyzed oxidation is extremely common, the presence of trace metal ions can catalyze oxidation by either directly reacting with the side chains of certain amino acid residues to produce free radicals or complexing with oxygen to produce various reactive oxygen species. Due to site-specificity, a faster oxidation rate is commonly observed for amino acid residue having closer proximity to the metal-binding sites [51]. However, site-specific oxidation can be greatly reduced by replacement of the existing amino acid residues with their larger, bulkier, or charged counterparts, as in the case of condon-222 mutant subtilisin [52].

Generally, an electron donor or reducing agent known as prooxidant is also required to reduce the transition metal ions. The prooxidant may be present as contaminant in buffers or formulations. For instance, the presence of reducing sugar impurities in mannitol has been suggested as a possible cause for the oxidative degradation of L-367073, a potent fibrinogen receptor antagonist, which involves the formation of a Schiff base intermediate [53]. The role of prooxidants in the oxidative degradation of proteins and peptides has been experimentally verified

under controlled conditions. For instance, in the presence of oxygen, $Fe^{3+}$ and an appropriate electron donor (e.g., ascorbic acid and dithiothreitol), the methionine residues in proteins are rapidly oxidized to methionine sulfoxides [49].

Although fewer stability studies have been reported for plasmid DNA and other nucleic acids than for proteins and peptides, the ability of trace metal ions to catalyze many oxidative processes (including Fenton reaction) even in the absence of hydrogen peroxide has been clearly demonstrated with the nucleic-acid-derived drugs [46]. Free radical oxidation is the major degradation process exhibited by plasmid DNA in pharmaceutical formulations in the absence of free radical scavengers and/or specific metal ion chelators [43]. Plasmid DNA formulations with demetalated phosphate buffers have significantly higher stability than those with non-demetalated buffers, strongly reflecting the important role of trace metal ions in the oxidative degradation of DNA [43].

***Specific Oxidation Reactions in Proteins and Peptides.*** In general, the amino acid residues exposed on the surface of proteins are more susceptible to oxidation than those buried within the hydrophobic core. For instance, different methionine residues in proteins can exhibit substantial differences in oxidation rate, as observed for human growth hormone [54], granulocyte colony-stimulating factor [55], and human parathyroid hormone [56]. Such differences in methionine reactivity are consistent with the location and solvent accessibility of the methionine residues.

Oxidation in proteins and peptides are highly amino acid specific. The amino acid residues that are susceptible to oxidation generally fall into two groups, viz. those containing a sulphur atom (methionine, cysteine) and those with an aromatic side chain (histidine, tryptophan and tyrosine) [49].

*Cysteine.* The free thiol groups present in cysteine residue can undergo spontaneous oxidation to form either intra-/intermolecular disulfide bonds and cross-links or monomolecular by-products such as sulfenic acid, sulfinic acid and sulfonic acid [49]. The mechanisms of cysteine oxidation by hydrogen peroxide have been investigated in detail. In the absence of metal ions and at low concentration of $H_2O_2$, the oxidation of cysteine is a two-step nucleophilic reaction that generates cystine as the final product, the rate-determining step being the reaction of thiolate anion with the neutral $H_2O_2$ to form cysteine sulfenic acid as an intermediate [57]. At high concentration of $H_2O_2$, the percentage of disulfide formation declines due to cysteine depletion or operation of alternative reaction pathways for consumption of cysteine sulfenic acid, leading to increased formation of cysteine sulfonic acid [57].

The oxidation of cysteine to cystine disulfide is strongly catalyzed by transition metal ions. As the formation of disulfide bonds and cross-links can bring about not only changes in tertiary structure but also covalent dimerization and aggregation, oxidation of cysteine has been of particular concern with many biopharmaceuticals, including interleukin-1B and interleukin-2 [54]. The spatial orientation of thiol groups is an important factor affecting the oxidation of cysteine to cystine. As two cysteine residues are required for the reaction, the oxidation rate is usually inversely related to the distance between the thiol groups [49]. In 28 hen lysozyme variants with random coil structure, the extent of disulfide bond formation in each variant has been shown to be proportional to the distance between the cysteine residues [58]. As cysteine oxidation is favored by deprotonation of the thiol group, the local

structure governing the ionization or $pK_a$ of cysteine residue will also influence the cysteine oxidation rate [49, 59]. As demonstrated using seminal ribonucleases segment 29–34 [60] and α-lactalbumin [61], introduction of an electronegative environment into the structures raises the $pK_a$ of the thiol groups and reduces the cysteine oxidation rate. In contrast, cysteine$^{25}$ residue in papain, which displays a lower $pK_a$ due to the presence of a nearby ionizable histidine group, is chemically more reactive than expected [59].

*Methionine.* The major oxidative degradation product of methionine is methionine sulfoxide. In acidic medium, methionine can be readily oxidized to methionine sulfoxide through a nucleophilic substitution reaction by alkyl hydrogen peroxides present in a solvent complex intermediate form [49]. In the presence of peracid, methionine can be oxidized to its sulfone derivative [54].

Light exposure, storage temperature, and oxygen level can all significantly influence the oxidation rates of protein pharmaceuticals [62, 63]. Local structure in the immediate vicinity of an amino acid residue in peptides may also affect the oxidation pattern of the residue. For instance, oxidation of threonine-methionine promotes the cleavage of the side chain of threonine residue [64]. Histidine facilitates the oxidation of neighboring methionine, and *C*-terminal methionine is oxidized more readily than the *N*-terminal or intra-chain methionine [54]. In addition, the formation of methionine sulfoxide is pH dependent, being favored at low pHs [54] and different methionine residues in the same protein molecule exhibit different oxidation rates, which also vary with pH [55, 56]. In human parathyroid hormone, the oxidation of the two methionine residues is virtually independent of pH between 4 and 8, but the rate of oxidation increases at pH 2 [56]. In contrast, the four methionine residues in granulocyte colony-stimulating factor show much larger differences in oxidation rate with $H_2O_2$ at various pHs [55]. Such pH dependence of the methonine oxidation has been attributed to the changes in local environment of the methonine residues in response to pH changes, i.e., the protonation or deprotonation of the nitrogen atoms in the neighboring histidine groups.

*Aromatic Amino Acids (Histidine, Tryptophan and Tyrosine).* Histidine residues are highly susceptible to oxidation, as shown for human growth hormone [65] and relaxin [66]. The resulting degradation products are aspartic acid, asparagines, and 2-oxo-histidine [65, 67]. Metal catalyzed oxidation of histidine may alter the secondary/tertiary structures of proteins. As has been demonstrated, oxidation of the human relaxin histidine$^{A(12)}$, which exists in an extended loop that joins two α-helices, alters the protein conformation, resulting in pH-dependent protein aggregation and precipitation [66, 68].

Tryptophan is also sensitive to oxidation, yielding mostly N′-formylkynurenine and 3-hydroxykynurenine [69]. The formation of N′-formylkynurenine can be activated by photoionization of tryptophan residues arising from the formation of solvated electrons [70]. The reaction is also catalyzed by the presence of metal ions [69]. Tryptophan residue can also mediate photooxidation of the disulfide bond in bovine somatotropin without being oxidized itself [70].

Tyrosine is also prone to oxidation, yielding 3,4-dihydrophenylalanine and dityrosine cross-link as the major reaction products [69]. For example, dityrosine-linked dimer was formed upon photoactivation of γB-crystallin at 445 nm [71].

*Other Amino Acid Residues.* N-terminal glycine residue is also a potential site for oxidation. Oxidation of Pexiganan has been reported to generate N-glyoxylyl-des-gly$_1$-pexiganan, which can be further decomposed to yield other products, as verified by high-performance liquid chromatography (HPLC) and mass spectrometry [72].

***Specific Oxidation Reactions in Nucleic Acids.*** Nucleobases and sugar moieties are the oxidation sites of nucleic acids. Similar to metal catalyzed oxidation in proteins and peptides, DNA damage through oxidative stress can also be induced by Fenton-type reactions involving transition metal ions, e.g., $Fe^{2+}$ and $Cu^{2+}$ [73]. The electron-rich purines and pyrimidines are the prime targets for the reaction with electrophilic oxidizing and photooxidizing agents [74, 75], and hydrogen abstraction and radical addition are the two known mechanisms responsible for the oxidation of aromatic nucleobases [44]. Such oxidation reactions will result in base cleavage and subsequent strand scission [76, 78]. As nucleobase oxidation removes electron density from the heterocycle, the oxidized nucleobase is a better leaving group, leading to an enhancement of hydrolytic depurination and depyrimidation [78]. Guanine, being the most readily oxidized site among the nucleobases, can be activated by light and catalyzed by trace metal ions to yield a wide range of oxidation products such as 8-hydroxyguanine, xanthine [76], 8-oxoguanine (8-oxoG), and 2,6-diamino-4-hydroxy-5-formamidopyrimidine [78]. Of all oxidation products of guanine, 8-oxoG is the most common oxidative lesion observed in duplex DNA, and is generally regarded as a key lesion in toxicological assessment [78]. Another well-characterized guanine lesion is 7,8-dihydro-8-oxo2′-deoxyguanosine (8-OH-dG) [79], which has been used as a degradation marker for a plasmid DNA in pharmaceutical formulation [43]. Adenine can also be oxidized readily, although significantly fewer oxidative lesions are formed because of its higher one-electron redox potential [78].

For the sugar moieties, the only oxidation mechanism is radical-mediated hydrogen abstraction, which can be induced by Fenton- or radiation-generated hydroxyl radicals [80]. Various hydrogen atoms in sugar moieties, such as H-1′, H-2′, H-3′, H-4′, and H-5′, are all potential sites for abstraction, although not all hydrogen atoms of deoxyribose in B-DNA have equal probability of being abstracted from duplex DNA [80]. The reaction depends on the helical structure of the nucleic acid, the orientation of the oxidant relative to the sugar as well as the stability of the carbon-centered radical being formed [44, 80]. As resonance stabilization is more difficult with sugar moieties than with nucleobases, the sugar moieties are less-favorable targets than nucleobases for radical attacks. The accessibility of the C–H bond also affects the rate and selectivity of hydrogen abstraction. For instance, oxidation of sugar moieties in the B-form of DNA mainly occurs at C-4′ due to its accessibility [73, 78, 80].

### 1.4 β-Elimination in Nucleic Acids, Proteins and Peptides

The two-step process of depurination (see the "Depurination and depyrimidation" Subsection in Section 1.2) and β-elimination is an important DNA degradation pathway in aqueous medium [45]. Upon depurination, the apurinic site contains a chemically altered sugar that alternates between a cyclic furanose form and an

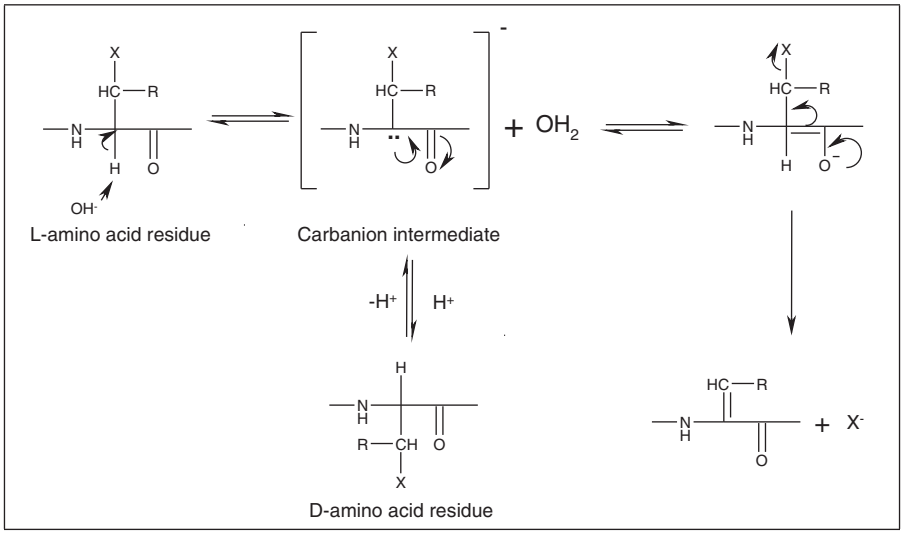


**Scheme 10.**

acyclic form containing an aldehyde functional group on the 1′ carbon. The aldehyde form of the sugar then undergoes β-elimination, which commences with the abstraction of a proton from the 2′ carbon by the $OH^-$, leading to strand breakage on the 3′ carbon of the abasic site [81].

In proteins and peptides, the β-elimination step involves the abstraction of a proton from the α-carbon of an amino acid residue in the peptide chain to form a carbanion intermediate (Scheme 10) [3]. Racemization can then occur via the addition of a proton to the opposite plane of the intermediate. Alternatively, the carbanion intermediate undergoes further reaction to form a dehydroalanine residue (Scheme 10) [3]. Cysteine, serine, threonine, phenylalanine, and lysine are the amino acid residues known to undergo β-elimination in peptides and proteins. Protein pharmaceuticals that can decompose by this reaction route include leuprolide [82, 83], salmon calcitonin [34], and insulin [84]. Leuprolide, which is formulated as a highly concentrated solution in DMSO for delivery in DUROS osmotic implant, shows increased β-elimination for the serine[4] residue in nonaqueous media [82, 83]. For insulin, the covalent and noncovalent aggregation of lyophilized preparation has been attributed to the intermolecular thiol-catalyzed disulfide interchange following β-elimination of an intact disulfide bridge in the insulin molecule [84]. This degradation process can be accelerated by elevated temperature; increased moisture content of the insulin, lyophilization, and/or dissolution of the insulin in alkaline media; as well as the presence of $Cu^{2+}$ (catalyst for the oxidation of free thiols).

## 1.5 Maillard Reaction in Nucleic Acids, Proteins, and Peptides

Maillard reaction, which is responsible for the nonenzymatic browning of materials, has been extensively investigated with food materials, but less so with pharmaceutical substances. The reaction consists of an initial reaction of reducing sugars with

amino or free amine groups in proteins, peptides, or nucleic acids, followed by Schiff base conversion and the formation of brown pigments via Amadori and Heyns rearrangement [85, 86]. For proteins and peptides, the amino acid residue involved in the reaction is usually lysine because of its free ε-amino group, but other bases such as arginine, asparagine, and glutamine can also react with the sugars [86]. In nucleic acids, the reaction may occur between the aldehyde of the sugar at the apurinic sites and the amino groups on the DNA bases in another plasmid molecule to produce concatamers [46]. Biopharmaceuticals documented to undergo Maillard reaction include glucagon [85], human relaxin [87], β-lactoglobulin variant A [88], lysozyme [89, 90], recombinant human serum albumin [90], calf thymus DNA [91], and plasmid pBR322 DNA [92].

Maillard reaction can give rise to significant changes/losses in physico-chemical properties and pharmacological activities of therapeutic proteins. Of particular concern in this regard is the modification of protein antigenicity, which is associated with the formation of advanced glycation endproduct (AGE) [93]. Glucose, a monosaccharide with reducing property, is particularly reactive toward the amino acid residues in proteins and peptides [87–90], whereas other sugars such as trehalose and mannitol have very little effect, as attested by the findings of the formulation studies with these sugars on glucagons [85] and relaxin [87]. Sucrose can also cause glycation when it is hydrolyzed to glucose and fructose by heat treatment, as observed for β-lactoglobulin variant A, lysozyme, and rHSA during antiviral heat bioprocessing [88–90].

As with proteins and peptides, nucleic acids can suffer substantial losses of biological activities as a result of Maillard reaction upon aging [91, 92]. Age-related dysfunction in gene expression has been demonstrated in DNA modified by reducing agents [91]. In addition, the amino groups of DNA have been shown to react nonenzymatically with reducing sugars (i.e., glucose) to afford a yellow-brown coloration, and the adducts formed *in vitro* have been found to decrease the transfection ability of f1 phage DNA [91, 92].

### 1.6 Covalent Dimerization and Polymerization in Proteins

Precipitation is a macroscopic process characterized by a visible change of an otherwise clear solution in the form of an increase in solution viscosity, clouding of solution, or phase separation of solid materials. The formation of precipitates may be due to covalent polymerization and/or noncovalent aggregation (Section 2.2). In covalent polymerization, dimer formation usually precedes the precipitation, and the phenomenon is best illustrated by insulin, which has been extensively reported to form covalent dimers.

Covalent dimer is the main degradation product of insulin upon storage [94]. Covalent trimer, tetramer, oligo-, and polymer formation is also possible with insulin when stored at ambient or higher temperatures. In general, the rate of insulin polymerization is virtually independent of the protein source (e.g., bovine, porcine, and human), but it varies with the composition and formulation of the protein and, for the isophane preparation, with the strength of preparation [94]. The rate of polymerization is one order of magnitude slower than insulin hydrolysis except for the NPH preparations [94, 95]. The covalent aggregation of insulin can be inhibited by self-association at high insulin concentration [9]. Both deamidation

and covalent dimer formation in human insulin solution at low pH seem to originate from a common cyclic anhydride intermediate [96]. Covalent dimer formation of human insulin is also evident in lyophilized solid state, accounting for no more than 15% of the total degradation [97]. Similar to the effect observed in solution, an increase in pH favors dimerization in lyophilized insulin, which can be ascribed to the increased nucleophilicity of the terminal amino groups after deprotonation. However, covalent dimerization can be almost completely suppressed by incorporating trehalose into the lyophilized matrix, which restricts the molecular mobility of insulin and keeps the protein molecules in a fully dispersed state.

Besides insulin, other biopharmaceuticals known to exhibit covalent polymerization include recombinant tumor necrosis factor-alpha, which forms nonreducible dimers and oligomers [98], and human insulin-like growth factor I, which is prone to covalent aggregation [99].

## 2 PHYSICAL INSTABILITY

### 2.1 Denaturation

Denaturation refers to an alteration of the global fold of a biopharmaceutical with higher order structures without an accompanying change in the primary structure. For proteins, these higher order structures refer to the secondary, tertiary, and quaternary structures with different levels of complexity. Although nucleic acids are structurally less complex than proteins, more elaborate structures such as double helix and supercoiling exist in DNAs, and many RNAs also have well-defined tertiary structures. Any conformational changes in these structures can lead to denaturation of these biopharmaceuticals.

The conformational integrity of native protein is maintained by a proper balance of weak nonbonding forces. Conceptually, an equilibrium exists between the native-folded and the denatured-unfolded forms of the protein, with or without the involvement of unfolding intermediate(s), or molten globule(s), in definable thermodynamic state(s). As the free energy of protein unfolding $\Delta G_{\mathrm{unf}}$, is typically in the range of 10–20 kcal/mol, the native structure of the protein is readily disrupted by a change of environment (e.g., application of heat, change of pH, and addition of detergents or chaotropes) [100]. Such structural perturbation, which is either permanent or temporary, leads to a change of conformationally sensitive physical properties (e.g., optical rotation, viscosity, and UV absorption) and, most seriously, to a concomitant loss of biological activities. If the denaturation process is reversible, the loss of native structure can be recovered once the stress is removed. However, no such structure recovery is possible with irreversible denaturation involving covalent bond formation such as disulfide exchange or cross-linking. Irreversible denaturation also leads to aggregation and precipitation with accompanying changes in the secondary and/or tertiary structures of the proteins. The soluble aggregates derived from denatured-unfolded protein and unfolding protein intermediates are prone to aggregation and precipitation in pharmaceutical formulation, resulting in a severe activity loss, as illustrated by recombinant human granulocyte colony stimulating factor [101]. Addition of sucrose to this protein brings about an increase in $\Delta G_{\mathrm{unf}}$ and a consequential decrease in aggregation and an increase in conformational stability.

Almost all water-soluble proteins are denatured by heat treatment. Although some denaturation processes follow the Arrhenius kinetics, the $\Delta G_{\text{unf}}$ in most cases follow a parabolic dependence on temperature, as depicted by the following equation [100]:

$$\Delta G_{\text{unf}} = \Delta H_m (1 - T/T_m) - \Delta C_p (T_m - T) + T \ln(T/T_m) \tag{1}$$

where $\Delta H_m$ is enthalpy of melting, $\Delta C_p$ is the specific heat capacity, and $T_m$ is the melting temperature. In an actual situation, once the temperature is raised and a significant fraction of the protein ensemble is unfolded, aggregation and/or precipitation can proceed rapidly, resulting in irreversible denaturation. In addition to heat denaturation, cold denaturation of proteins can occur during freeze-drying. In general, freeze-drying causes considerable conformational changes and aggregation in protein molecules, thereby promoting the formation of the β-sheet structure at the expense of the α-helix and random structures [102].

Protein stability is highly sensitive to pH changes and decreases sharply at acidic or basic pH. Because of the differences in $pK_a$ of certain amino acid groups between the native and unfolded proteins, protons released in response to pH changes can play an important role in the unfolding process [100]. It has been shown that protein conformational fluctuations generally agree with the predictions based on the $pK_a$ values of the titrating groups in the protein [103]. Aside from pH manipulation, protein unfolding can be achieved with chaotropic agents (e.g., urea and guanidine hydrochloride). However, in contrast to the less-than-complete protein unfolding induced by thermal denaturation, the denatured state produced by such chaotropes is virtually devoid of structure, resembling a random coil conformation [104].

Subjection of duplex DNA to various stress factors such as heat, pH, and ionic strength can lead to a collapse of the native DNA structure. The two complementary strands separate and assume a random coil conformation with concomitant changes in physical properties (e.g., viscosity and UV absorbance) and biological activities. The thermodynamic stability of a nucleic acid duplex is a complex function of temperature and pressure, being strongly dependent on the denaturation temperature $T_m$ [105, 106]. Interestingly, the relationship between the free energy associated with helix-to-coil transitions of nucleic acid duplexes $\Delta G_m$ and the corresponding $\Delta H_m$, $T_m$, and $\Delta C_p$ conforms to that predicted by Equation (1) for the protein unfolding process, where a parabolic curve is expected [106]. Cold denaturation of nucleic acid duplex can also occur when the temperature is held below –120°C at atmospheric pressure [105].

Duplex stability is dependent on DNA conformation. Classic B conformation has a lower free energy of helix-to-coil transition and lower $T_m$ than B′ conformation, whereas B′ conformation has a higher thermal stability than A conformation [106]. Base pairs also influence duplex stability.

## 2.2 Aggregation and Precipitation

Aggregation and precipitation are common problems of instability with protein biopharmaceuticals, e.g., keratinocyte growth factor [107], human immunoglobulin [108], and human serum albumin [109]. In clinical practice, an aggregation level of

as low as 1% over a period of 2 years is deemed unacceptable [110]. Aggregation differs from self-association in that it is confined only to non-native proteins, and it enamates from a conformational intermediate produced during folding or denaturation of the proteins, whereas self-association occurs in native proteins in response to deliberate changes of solvent environment (e.g., pH, ionic strength, and solvent composition) to effect crystallization or isoelectric precipitation.

Protein aggregation can be described as a two-step process. The first step entails protein unfolding and exposure of buried hydrophobic residues to the aqueous solvent. The second step involves intermolecular interactions among the exposed hydrophobic residues of the unfolded protein intermediates, which are thermodynamically favored by the tendency of the exposed hydrophobic amino acid residues to minimize their contact with the aqueous medium. The reaction is site specific, as the initial stage of aggregation involves the interaction of one protein molecule with another at specific sites. Two such sites per protein molecule are sufficient for aggregation to proceed longitudinally, leading to the formation of long fibers. As more of these stable aggregates comprising a few molecules are formed (nucleation), they will develop into larger aggregates (growth) and eventually into floccules, which will precipitate from the solution once their solubility limit is exceeded.

In aqueous solution, non-native protein aggregation is dependent on various factors, such as temperature, solution pH, salt type, and concentration as well as the presence of ligands, cosolutes, preservatives, and surfactants [111]. Intrinsic conformational stability of native proteins as well as the colloidal stability between protein molecules may play a crucial role in the aggregation process. As aggregation requires the presence of two or more protein molecules, the aggregation reaction rate is expected to be higher because of increased probability of molecular collision [112]. However, the problem of protein aggregation in solution cannot be solved simply by preparing the protein in solid form alone, as solid phase aggregation is also common with lyophilized or freeze-dried formulations. There are a wide variety of mechanisms responsible for moisture-mediated solid-phase aggregation, encompassing thiol-disulfide interchange, thiol-catalyzed disulfide exchange after β-elimination, non-covalent aggregation, and non-disulfide covalent aggregation [113].

Aggregation can lower protein potency simply by decreasing the effective protein concentration. Moreover, it can increase the immunogenicity of proteins, as observed for insulin, human growth hormone, and recombinant human interferon-α2a [114]. Various low-molecular-weight excipients have been investigated for their potential to minimize protein aggregation or stabilize protein conformation [112]. Simple sugars, polyhydric alcohols, and amino acids have all been shown to effectively prevent freeze/thaw-induced aggregation in chimeric L6, a mouse-human monoclonal antibody [115]. Sucrose is known to inhibit the aggregation of recombinant human granulocyte colony stimulating factor [116]. Sorbitol has also been demonstrated to reduce moisture-induced aggregation in tetanus and diphtheria toxoids [117]. β-Cyclodextrins have also been found to inhibit the aggregation of recombinant human growth hormone [118]. However, there are also excipients that can adversely impact the protein aggregation and precipitation. For example, benzyl alcohol can induce precipitation of human insulin-like growth factor I from solution [119] as well as aggregation of human interleukin-1 receptor antagonist (a predominantly β-sheet protein) in aqueous solution [120] and reconstituted lyophilized formulation [121].

## 2.3 Adsorption

Interfaces between two separate phases, such as air/water, oil/water, and solid/water, are potential adsorption sites of biopharmaceuticals. Adsorption often involves simple diffusion of surface-active solute molecules in the bulk to the interface, and hence the rate of adsorption is generally dependent on the solute concentration. At saturation, a close packed monolayer of protein molecules corresponding to 0.1 to 0.5 $\mu g/cm^2$ is normally formed at the interface [122], and this adsorption behavior is of particular concern for high-potency therapeutic proteins. However, certain proteins do not conform to such saturation-limited adsorption behavior and tend to show increased adsorption with increasing protein concentration, attaining a local protein concentration at the interface 1000 times higher than the initial concentration in the bulk solution [123].

The surface activity resulting from different amino acid compositions of polypeptides has an important bearing on the extent of surface adsorption. As amino acids can be hydrophobic (e.g., tryptophan, phenylalanine, and isoleucine), hydrophilic (e.g., serine and threonine), negatively charged (e.g., aspartic acid and glutamic acid), or positively charged (e.g., histidine), the proteins involved can potentially adsorb on to solids such as plastics and glasses with different affinities. Owing to their intrinsic polyelectrolyte nature, the adsorption of proteins on solids is highly pH dependent, reaching a maximum at their respective isoelectric points (p*I*s) [122]. It has been demonstrated that recombinant human interleukin 11, which is monomeric and highly basic ($pI > 10.5$), displays a nonspecific loss to container in alkaline solution [32]. Surface adsorption alone can cause more than 40% activity reduction of interleukin 11 in solution after 3 hours of storage at room temperature [124].

Upon adsorption to container, conformational changes can occur in the higher order structures of proteins. As hydrophobic amino acid residues tend to stay inside a protein (i.e., away from the aqueous environment), surface interaction may unfold the protein to expose the inside structure, thereby leading to structural rearrangement of the protein. For instance, it has been shown that the native structure of α-helices and extended loop bands of recombinant interleukin 2 disappeared, whereas the β-sheet appeared upon contact with a pump-based delivery system (borosilicate glass pump reservoir and catheter tubing), resulting in irreversible structural changes and major biological activity loss of the protein [125].

Apart from solid surfaces, liquid surfaces are potential sites for protein adsorption. By creating a large surface area, agitation may induce protein adsorption at the liquid surface where the area occupied per unit mass of protein in the monolayer region is high (~1 $m^2/mg$) compared with that at the solid–liquid interface (~0.1 $m^2/mg$) [122]. The adsorption process coupled with interfacial and shear forces may lead to protein denaturation through unfolding as well as aggregation and precipitation (Section 2.2), as observed for recombinant factor VIII SQ [126].

Addition of surface-active agents (e.g., polysorbates) can prevent protein adsorption and stabilize the protein preparation. Numerous reports have documented the stabilizing effect of surface-active agents against adsorption of biotechnological products. For example, polysorbate-80 exerts a primary stabilizing effect on adenovirus type-5 based vaccine by inhibiting adsorption of the vaccine to glass vials

[127]. The presence of 0.12% polysorbate-20 reduces the adsorption of recombinant human interferon-γ to both air–liquid and ice–liquid interfaces [128]. Similarly, the presence of polysorbate-80 or polysorbate-20 reduces surface adsorption of recombinant factor VIII SQ and, hence, the agitation-induced denaturation of the protein [126]. Inclusion of albumin, which minimizes the contact of proteins with glass surface, may also help to stabilize the proteins against glass adsorption, as demonstrated for salmon calcitonin [129].

Paradoxically, the adsorption process can be used to stabilize biopharmaceuticals, particularly vaccines [130]. Aluminum salts have been widely used as adjuvants in intramuscular vaccine injections. Novel vaccine formulations containing aluminum are available; for example, alum-adsorbed hepatitis B surface antigen and diphtheria-tetanus toxoid vaccines have been optimized for epidermal immunization in powder form [131].

## 3 CONSIDERATIONS IN THE STABILITY TESTING OF BIOPHARMACEUTICALS

### 3.1 Temperature

The stability of pharmaceuticals, including biopharmaceuticals, is highly temperature dependent. For most small organic molecules, thermal degradation generally follows Arrhenius kinetics. However, many biomaterials, e.g., proteins, do not exhibit Arrhenius degradation kinetics even over a narrow temperature range, which can be attributed to phase transitions, pH shifts, poor relative humidity control at elevated or subambient temperature, and involvement of complex reaction mechanisms [132]. Multistep reaction pathways are particularly common in biopharmaceuticals because of the existence of higher order structures. For instance, with recombinant bovine granulocyte colony stimulating factor, a reversible equilibrium between the native protein and an intermediate state is established, followed by irreversible aggregation [133]. Although individual denaturation steps seem to obey the Arrhenius law, the overall kinetic behavior for product formation may not [132, 133]. Thus, for protein pharmaceuticals, extrapolated shelf-life prediction from accelerated stability test must be viewed with caution and real-time stability monitoring should also be conducted whenever feasible.

As many therapeutic proteins are formulated in aqueous solution and stored at 2–8°C, they are liable to cold denaturation at such low temperatures and possibly to freezing/thawing inactivation arising from temperature fluctuation during storage or handling. When a protein solution freezes, water crystallizes out as solid ice and the solution becomes more concentrated in the protein and salt components present, causing dramatic changes in pH and ionic strength and hence protein inactivation [134]. For instance, during freezing of sodium phosphate solutions, crystallization of the disodium salt can reduce the pH by as much as 3 units [135]. The effects of freezing and thawing rates on the stability of model proteins have been investigated in the absence of cryoprotectants, and higher activity recovery at a slower freezing rate and a faster thawing rate has been observed [136]. During fast freezing, small ice crystals are preferentially formed and the associated increase in total surface area of the ice–liquid interface will enhance the exposure of protein molecules to

the interface and increase the protein damage. Upon thawing, additional damage to proteins may be caused by recrystallization [136]. Many biopharmaceuticals, including recombinant human interferon-γ [135], tetanus toxoid conjugate vaccines [137], and pCMVβ-gal plamid DNA/lipid complexes [138] show a reduction in activity upon freezing/thawing. To circumvent the aforementioned protein inactivation problems, various excipients have been used in biopharmaceutical formulations. For example, Tween 80, a polysorbate surfactant, has proved useful for reducing surface-induced protein denaturation during freezing [139], and glycine can also stabilize proteins through the preferential exclusion mechanism [135].

### 3.2 Moisture

Moisture is well documented to adversely affect the chemical stability of solid biopharmaceuticals. The presence of water enhances the mobility and flexibility of the bioactive macromolecules in solid excipient matrix, thereby facilitating their rearrangement and increasing their susceptibility to chemical degradation. For instance, the deamidation rate of an asparagine containing hexapeptide in lyophilized poly(vinyl alcohol) (PVA) and poly(vinyl pyrrolidone) (PVP) increases with increasing moisture uptake and water activity of the system [140]. The asparagine deamidation seems to correlate closely with the extent of water-induced plasticization of PVA and PVP matrices (as determined by the respective glass transition temperature $T_g$), suggesting that the chemical stability of the peptide may be predicted by the physical state of the formulation [140]. Similarly, aspartate isomerization in humanized monoclonal antibody, which can be accelerated at elevated temperatures, has been observed to follow the Arrhenius law above the $T_g$ of a lyophilized formulation [141].

The presence of moisture can also reduce physical stability. The mechanisms of moisture-induced aggregation are well documented and have been exemplified by a good number of solid biopharmaceuticals, including salmon calcitonin spray-dried powders for inhalation [142], human serum albumin [143], and recombinant human albumin [109]. To prevent such protein aggregation, excipients are often employed in protein formulation. The excipients, which comprise mostly sugars such as mannitol, lactose, trehalose, and cellobiose, prevent protein aggregation by occupying the water-binding sites of protein in the dried state [144]. However, these sugar excipients may also present formulation problems, notably sugar crystallization at high excipient-to-protein ratios [144] and phase separation due to prevalent sugar–sugar interactions [145].

The conventional wisdom of moisture content control for biopharmaceuticals is "the drier, the better." As has been demonstrated with a lyophilized humanized monoclonal antibody formulation, moist cakes tend to have higher aggregation rates than drier samples if stored above their $T_g$ [141]. However, excessive moisture removal can destabilize the protein, as observed with bovine immunoglobulin [146].

### 3.3 Light and Ionizing Radiation

Eletromagnetic radiations of different wavelengths, i.e., UV-visible light and x-ray, are known to induce degradation in proteins [147]. Radiation, which is commonly employed for sterilization purpose, can generate free radicals and lead to peptide

chain cleavage and aggregation in solid protein pharmaceuticals [147]. In aqueous solution, the protein may also degrade through the destruction of the amino acid residues by the hydroxyl radicals and electrons produced from water molecules [147].

As biopharmaceuticals are formulated mostly for parenteral administration, complete shielding of light is impossible. Although amber glass containers can be used for shielding against UV light, their utility in light protection is often limited by the glass thickness. Light-induced degradation or photodegradation can be minimized by formulation into solid dosage forms. By formulating into a lyophilized form, recombinant human factor VIII showed no significant loss of activity after accelerated photostability testing, whereas the reconstituted preparation displayed partial activity loss after similar light exposure, and the observed photodegradation could be effectively prevented by packaging the protein in tinfoil wrap [148].

Nucleic acids are also susceptible to UV light. DNA molecules absorb photon energy at wavelengths below 320 nm, which gives rise to the widely reported mutagenic DNA damage. DNA photocleavage can occur at both deoxyriboses and nucleobases [75]. The presence of photosensitizers, such as psoralen, chlorpromazine, and fluoroquinolones, may promote the light-induced degradation by generating reactive oxygen radicals or singlet oxygens, which then interact with DNA to produce damage [149].

### 3.4 Physical Stress

Biopharmaceuticals may be inactivated by physical stress associated with vial filling, shipping, storage, and handling. Insulin is a well-known example of agitation-induced instability, and patients are advised to avoid vigorous shaking of the insulin preparation. Shaking is known to accelerate the degradation of insulin by way of covalent dimerization [150]. To assess the adverse impact of agitation on the stability of insulin preparations, two automated physical stress tests, viz., the Temperature Cycling and Resuspension Test (TCRT) and the High Temperature and Extreme Agitation Test (HTEAT), have been applied to a commercial portable insulin pen-cartridge device [151]. TCRT involves temperature cycling (25–37°C) in an incubator unit combined with resuspension (three sets of 10 rolls plus 10 inversions) conducted twice daily, whereas HTEAT requires continuous high-temperature (37°C) exposure in an incubator unit combined with periodic daily agitation for 4 hours at 30 rpm. These two tests afford a rapid means for screening potential aggregation during the formulation development of proteins [151].

Apart from insulin, monomeric recombinant human growth hormone has been shown to aggregate rapidly within 10 hours of shaking [152]. The presence of Tween 20, a nonionic surfactant, at an excipient : protein molar ratio larger than 4 effectively inhibited this aggregation. Similar aggregation was observed with IgG1-antibody upon the application of physical stress. However, differences in aggregation mechanism exist among different mechanical stress methods (e.g., shaking and stirring) [153]. Moreover, the aggregation kinetics is dependent on the shear rate applied, and trimer formation is particularly evident at a high shear rate, as observed for human serum albumin [154]. Even simple vortexing for just 1 minute can cause substantial aggregation and precipitation, as shown for recombinant human growth hormone [155].

Agitation-induced instability is not limited to protein pharmaceuticals. For instance, agitation has been shown to significantly reduce transfection rates in pCMVβ-gal, a plasmid DNA, and in complexes prepared with three different commercially available lipid formulations, namely, DMRIE-C, lipofectAMINE, and DOTAP:DOPE [138].

### 3.5 Freeze-Drying and Spray-Drying

Lyophilization, the most widely used process for preparing solid proteins, consists of two major steps: freezing of a protein solution and drying of the frozen solid under vacuum. The drying step can be further divided into two phases, namely, primary drying (removal of frozen water) and secondary drying (removal of the non-frozen bound water) [156]. As alluded to, freezing can cause cold denaturation and introduce additional stresses to the proteins through ice crystallization (Section 3.1). Likewise, drying can bring about protein degradation. By removing the hydration shell covering the protein surface, the native state of protein may be disrupted, resulting in denaturation [157]. Water molecules may also be an integral part of protein active sites. Removal of these functional water molecules can inactivate proteins during dehydration [157]. Thus, it is perhaps not surprising that freeze-drying-induced degradation is common with protein pharmaceuticals, e.g., *Humicola lanuginose* lipase [158], lysozyme [159], and recombinant human factor XIII [160], as well as nucleic-acid-derived drugs, e.g., poly((2-dimethylamino)ethyl methacrylate)-based pCMV-lacZ plasmid gene delivery system [161] and cationic lipid-protamine-DNA complexes [162]. Aggregate formation is also apparent during storage of freeze-dried biotechnological products, e.g., ribonuclease A [163], and sugars, polyols, and polymers are the excipients commonly used for stabilizing them [157].

Spray-drying, a single step operation that converts a liquid feed to a dried particulate form, has proved useful for producing powders within the respirable range for pulmonary delivery of biopharmaceuticals. Many of these products, including recombinant human growth hormone [164], recombinant human deoxyribonucleases [165], and recombinant humanized anti-immunoglobulin E monoclonal antibody (rhuMAbE25), have been prepared by the spray-drying process [166]. As spray-drying of these proteins alone in solutions can result in unfolding, aggregation, and inactivation because of temperature and interfacial effects, stabilizers such as polysorbates [164, 167], mannitol [166], sucrose, and trehalose [168], are usually added to improve the processing and storage stability of these preparations [169].

### 3.6 Excipients

The inclusion of chemical excipients/additives can inhibit or slow degradation reaction in biotechnological products. In the case of metal catalyzed oxidation, chelating agents such as ethylene diamine tetra acetic acid (EDTA) can suppress the reaction by removing trace transition metals ions from the bulk solution. Some sugars and polyols (mannitol, glucose, glycerol, ethylene glycol) can inhibit metal catalyzed oxidation by complexation with the transition metal ions, whereas polymeric additives such as dextran may function as a hydroxyl radical scavenger [170]. Sodium thiosulfate has been reported to reduce the formation of a disulfide-linked high-molecular-weight species of OKT3 antibody in solution [171]. However,

precautions must be taken when using these excipients for stabilization purposes. As has been reported, inclusion of aliphatic alcohols as inhibitors of metal catalyzed oxidation can give rise to cosolvent-induced perturbation of the metal binding site in recombinant human growth hormone [172].

Although several excipients have been made available for prolonging the shelf-life of biopharmaceuticals, potential reactions between excipients and actives may adversely influence the overall stability of the formulations. For instance, excipients containing carbonyl functional groups, such as PVP, can form a covalent adduct with peptides and proteins containing primary amines (i.e., N-terminus and lysine side chain) at high temperatures (70°C) and low RH (~0%) [173]. The presence of sucrose can promote methionine oxidation in recombinant factor VIIa, possibly due to the increase of chemical potential or a small conformational change in microenvironment [174]. Furthermore, excipients may be contaminated with traces of impurities, such as transition metals and reducing agents, which can compromise the stability of the protein products. For instance, both food-grade and pharmaceutical-grade polysorbates often contain traces of alkylhydroperoxide and hydrogen peroxide arising from the bleaching step in the synthetic process [175]. During storage, polysorbates may also decompose to peroxides, and the decomposition is accelerated by increased exposure to air and light, elevated temperature, and the trace presence of metal catalysts [54, 175].

## 4 REGULATORY REQUIREMENTS IN THE STABILITY TESTING OF BIOPHARMACEUTICALS

In the development of stability testing protocols for biotechnological/biological products, all relevant regulatory guidelines must be closely followed. The International Conference on Harmonization (ICH) guidelines for stability testing are widely adopted and are available online (http://www.ich.org). It is important to check the logistics and requirements of the appropriate stability programs before initiation of the stability testing. These guidelines have defined the storage conditions for long-term, intermediate, and accelerated stability testing for different climatic zones [Q1A(R2); Q1C; Q1F]. Photostability is also considered under the relevant section [Q1B]. Experimental design and data analysis with specific requirements also constitute an important part of the stability testing programs [Q1D; Q1E]. Given the complexity of biotechnological products and their sensitivity to environmental factors, such as temperature, humidity, light, oxygen, ionic strength, and shear stress, additional guidelines are available for assuring their quality and Q5B aims to regulate the stability testing of these products. Apart from the method of batch selection for the testing, Q5B guidelines have a particular emphasis on the establishment of stability-indicating profiles for the active drug substances. Potency testing should be included in the stability assessment, and the purity of biotechnological products should normally be determined by more than one method.

## 5 FORMULATION AND CHARACTERIZATION OF BIOPHARMACEUTICALS

As discussed, biopharmaceuticals are characterized by large molecular sizes and various levels of structural complexity (i.e., primary, secondary, tertiary, or

quaternary), which render them particularly vulnerable to various forms of degradation or denaturation. Most current strategies for formulating biopharmaceuticals into appropriate dosage forms or delivery systems are aimed specifically at preserving the structural integrity and biological activities of these complex macromolecules as well as at improving their bioavailability or targeting them to specific sites in the body. Presented below is a general discussion on the formulation of biopharmaceuticals with particular emphasis on the techniques using pharmaceutical excipients or additives to achieve these aims. Biotechnological techniques that use viral vectors to deliver genes or nucleic acid-based drugs are outside the scope of the current discussion.

### 5.1 Protein and Peptide Drugs

With the aid of newer production techniques, such as recombinant DNA technology and transgenic protein production, coupled with cost-efficient protein purification techniques, proteins of high purity can now be mass-produced for therapeutic uses.

In protein production, selection of the host cells (e.g., bacterial, yeast, or mammalian cells) for the expression of recombinant proteins is critical, as this will determine the chemical properties (e.g., amino acid sequence/composition and glycosylation) as well as the pharmacokinetic behavior, biological activities, and potential toxicities of the resulting proteins [176]. For instance, granulocyte macrophage colony stimulating factor (GM-CSF) expressed in *Escherichia coli* bacteria is a glycosylated form with six fewer amino acids and one more methionine residue than the non-glycosylated native protein [177, 178]. Although both yeast and mammalian expression systems can produce glycosylated proteins, only the mammalian system can generate the more complex carbohydrate moieties, which may be associated with higher immunogenicity. In addition, various types of contaminants (e.g., viruses, toxins, and nucleic acids) may arise during the production, and thus establishment of the protein purity at the initial stage of formulation development is of paramount importance [179, 180].

A challenge to develop a high-concentration protein formulation is perhaps the limited solubility of the protein in aqueous solution. The aqueous solubility of a protein is governed by the interaction of its polar residues with water, whereas the non-polar side chains and peptide groups are normally buried in the folded protein to maintain the native state. In general, minimum protein solubility occurs at the isoelectric point (*p*I), where the protein molecule acquires a net-zero charge [181]. Thus, to ensure adequate solubility in water, proteins should not be buffered at or near their *p*Is. However, when selecting an appropriate pH for the vehicle, considerations should also be given to the chemical integrity and activity of the proteins, as these are also pH-dependent.

Proteins are amino acid chains that fold into unique three-dimensional structures. The primary structure of proteins is the peptide sequence of the whole molecule, which can be characterized by the amino acid composition, N-terminal amino acid sequence, C-terminal amino acid sequence, and peptide mapping.

The secondary structure refers to the general three-dimensional local regions of the protein, encompassing regions of α-helices, β-sheets, and supersecondary

structures. Such local structures can be analyzed by circular dichroism, Fourier transform infrared spectroscopy, and fluorescence spectroscopy.

The tertiary structure refers to the overall three-dimensional structure of the polypeptide units of a given protein. Just as the secondary structure may be conceptually linked to distinct domains within a protein, the tertiary structure can be associated with the way by which these domains are related to one another. Interactions among such domains are governed by various physical forces, including hydrogen bonding as well as hydrophobic, electrostatic, and van der Waals interactions. The tertiary structure of proteins can be determined using a combination of analytical and complementary computer-aided modeling techniques, including solution nuclear magnetic resonance (NMR) spectroscopy, single-crystal x-ray crystallography, neutron diffraction, and molecular dynamics simulation.

In general, no single analytical technique can be used to fully characterize biopharmaceuticals. Aside from structural characterization, functional analyses, including immunoassay and bioactivity assay, are also essential for ascertaining the quality of biopharmaceuticals. Although immunoassay is more specific than bioassay in the presence of interfering factors, the data derived from the two techniques may not show any direct correlation with each other, mainly because the former detects not only biologically active proteins, but also denatured proteins and fragments as well as inactive protein-receptor complexes [182]. Some commonly used protein characterization techniques are summarized in Table 1. More specific reviews on protein and peptide analysis are presented elsewhere [183, 184].

An appropriate choice of excipients is also an important consideration in protein/peptide formulation, as it can profoundly influence the shelf-life, pharmacokinetic behavior, and efficacy of the final products. Table 2 shows a list of excipients commonly used in protein/peptide formulation [185]. Extensive reviews on excipient selection are available in literature [186, 187]. In general, excipients are experimentally selected by screening [188, 189]. The selection is guided by their abilities to produce the desired outcomes for the final dosage forms. For example, the stability of hybrid interferon-α in solution, which is sensitive to both pH and buffer composition, can be improved by adding mannitol or lactose at pH 4.0 [190]. However, at pH 7.6, lactose exerts an accelerating effect on the degradation of interferon-α, suggesting that an optimal pH exists for the excipient to function as a stabilizer. In another study, hydroxypropyl-β-cyclodextrin has been shown to effectively protect β-galactosidase against degradation during spray-drying [191]. Apart from the type of excipients used, the physical form of the excipients is also critical for the stabilization of proteins. For instance, mannitol in the amorphous state can preserve the structure and prevent the aggregation of recombinant humanized anti-IgE monoclonal antibody [166].

As many excipients are capable of stabilizing proteins under different conditions, it is unlikely that the stabilization effect is excipient-specific. It is widely believed that the stabilizing mechanism of excipients involves the preferential exclusion of the added excipient from the protein in aqueous solution and during freeze–thawing [192]. Briefly, a protein in an aqueous environment is in dynamic equilibrium between the native and the denatured (unfolded) forms. As the exclusion of excipient from the unfolded form is greater than that from the native form, the equilibrium will shift toward increasing concentration of the native form. Thus, the unfolding of protein in the presence of excipient is a thermodynamically

**TABLE 1. Summary of the Commonly Used Analytical Techniques for Protein/Peptide Characterization**

| Technique | Principle | Application | References |
|---|---|---|---|
| *Colorimetric analysis* | Based on specific chemical reaction and associated color change | Quantitative analysis; BCA assay | [207] |
| *High pressure liquid chromatography (HPLC)* | | | [226] |
| Reversed-phase chromatography (RPC) | Separation by hydrophobicity | Quantitative, stability (aggregation), purity analysis | |
| Size exclusion chromatography (SEC) | Separation by size (larger molecule elutes faster) | | |
| Ion-exchange chromatography (IEC) | Separation by electrostatic interaction | | |
| *Electrophoresis* | | | [227] |
| SDS-PAGE | Separation based on molecular sieving by gel matrix | Size, purity, stability analysis | |
| Capillary electrophoresis (CE) | Electrophoretic separation on capillary | Peptide analysis, purity, stability analysis | |
| Isoelectric focusing (IEF) | Separation based on continuous pH gradient (proteins migrate according to charge properties until the point of zero charge, i.e., isoelectric point or pI, is reached.) | Purity analysis, pI determination | |
| *Mass spectrometry (MalDI-TOF)* | Separation based on mass-to-charge ratio (m/z) in magnetic or electric field | Quantitative analysis (high sensitivity) Identification and integrity | [228, 229] |
| *Immunochemical analyses* | | | |
| Enzyme-linked immunosorbent assay (ELISA) | The first antibody is coated onto plastic surface and conjugated with antigen, whereas the second antibody is coupled to an enzyme; detection by color change of a chromogen when it is cleaved by the enzyme | Quantitative analysis; epitope quality | [207] |
| *Bioassay* | Based on biological effects of proteins (e.g., cell proliferation, cytotoxicity, and antiviral activity.) | Quantitative analysis; potency test | [230–231] |

**TABLE 2. Excipients Commonly Used to Stabilize Biopharmaceuticals (from Ref. 185)**

| Excipient Type | Examples |
|---|---|
| Amino acids | Glycine, arginine, alanine, proline, aspartic acid, glutamic acid, lysine |
| Sugars | Trehalose, sucrose, maltose, fructose, raffinose, lactose, glucose |
| Surfactants | Poloxamer 407, Poloxamer 188, polysorbate 80, polysorbate 20, octoxynol-9, polyoxyethylene-(23) lauryl alcohol, polyxyethylene-(20) oleyl alcohol, sodium lauryl sulphate |
| Salts | Sodium sulphate, ammonium sulphate, magnesium sulphate, sodium acetate, sodium lactate, sodium succinate, sodium proprionate, potassium phosphate |
| Polyols | Cyclodextrins, mannitol, sorbitol, glycerol, xylitol, inositol |
| Antioxidants | Ascorbic acid, glutathione |
| Polymers | Polyethylene glycol, dextran, polyvinylpyrrolidone |
| Chelating agents | EDTA, tris(hydroxymethyl)aminomethane (TRIS), diethylenetriaminepentaacetic acid, inositol, hexaphosphate, ethylenediaminebis (O-hydroxyphenylacetic acid), desferal |

unfavorable process. This is to be expected because the excipient would denature the protein if it were to bind favorably to the exposed hydrophobic sites of the unfolded protein.

Two nonexclusive mechanisms, namely, glass dynamics hypothesis and water substitution concept, have also been proposed to explain the stabilization effects of excipients during the dehydration stage in freeze-drying [193]. According to the glass dynamics hypothesis, freeze-dried excipients form a rigid, amorphous, inert matrix where the dispersed protein molecules have limited mobility, being restricted from undergoing arrangement and eventual phase separation. Therefore, stabilization is via a kinetic mechanism and would be expected to correlate with the molecular mobility in the rigid matrix. The water substitution theory suggests that the stabilization effect is due to the substitution by the excipient for water to maintain hydrogen bonds with surface-accessible polar sites on the protein. Such interaction facilitates the preservation of the native-like solid-state protein structure and may help reduce the tendency of the protein molecules to form aggregates during storage. These stabilization mechanisms have been reviewed in depth by Pikal [194].

Most commercial protein/peptide products are administered by the parenteral route (e.g., subcutaneous, intramuscular, or intravenous) in the form of injections. The major disadvantages of parenteral injections are the inconvenience to the patient, high cost, and poor patient compliance. To overcome these problems, alternative administration routes such as peroral [195, 196], buccal [197], pulmonary, nasal [198], and transdermal [199] routes have been considered. However, nonparenteral protein delivery also suffers certain drawbacks, most of which are related to the inherent physicochemical characteristics of the proteins, including their large molecular size, low lipid membrane permeability, susceptibility to enzymatic degradation, potential immunogenicity, and bio-incompatibility as well as the propensity to undergo aggregation, adsorption, and denaturation. Not all

nonparenteral routes share these limitatons. For instance, absorption via the buccal mucosa is free from degradation by protease activity and allows the absorbed protein to bypass the liver to gain direct entry into the systemic circulation, thus avoiding the first-pass hepatic metabolism. Pulmonary delivery also offers a similar advantage of bypassing the first-pass effect.

The feasibility of protein delivery via different routes of administration is best illustrated by insulin, a well-characterized and extensively studied therapeutic protein. To date, attempts to exploit the nasal, oral, and transdermal routes for insulin delivery have not been very successful for reasons related mostly to the stability and efficiency of delivery [200]. For instance, oral insulin delivery is subject to activity loss due to the acidity in the stomach and the presence of proteases and peptidases in the gut. Moreover, the intestinal absorption of the protein is limited by its large molecular size and hydrophilic nature. Various specific strategies to overcome these delivery barriers have been attempted, including the use of permeation enhancers, protease inhibitors, and enteric-coated microsphere formulations [201]. As an alternative solution to these problems, the pulmonary route has also been extensively investigated for its utility in protein delivery [202]. As has been well documented, lungs have a large surface area, which allows rapid drug absorption due to the close contact between the alveoli in the deep lung and the systemic circulation. As a result of intense formulation research, a novel inhalation insulin product by the commercial name, Exubera, was developed and subsequently approved by the Food and Drug Administration (FDA) Advisory Committee Panel in September 2005 for the treatment of adult type 1 and type 2 diabetes. The product is a rapid-acting, dry powder form of insulin derived through rDNA. The main concern with inhaled insulin products is the potential long-term adverse effects resulting from the intra-alveolar deposition of insulin within the lung, because insulin has growth-promoting properties. It can be envisaged that once long-term safety and efficacy is established, inhalation may become the first nonparenteral route used clinically for insulin delivery.

Apart from the aforementioned problems with oral delivery, certain undesirable pharmacokinetic properties of proteins, notably the short biological half-life, have presented additional hurdles in their formulation development. To lengthen the biological half-life and prolong the action of proteins in the body, several novel drug delivery strategies have been developed through chemical modification of the protein molecule (e.g., pegylation) or new formulation development (e.g., controlled or sustained release formulation).

Pegylation, defined as the chemical process by which polyethylene glycol chains are attached to protein and peptide drugs, have been successfully applied to produce peginterferon alfa-2a (PEGASYS) and peginterferon alfa-2b (PEG-Intron), both of which have been approved by the FDA for the treatment of chronic hepatitis C [203].

Another approach to extend the action of proteins is to develop biodegradable polymeric microspheres. Owing to their excellent biocompatibility, the biodegradable polyesters, poly(lactic acid) (PLA) and poly(lactic-co-glycolic acid) (PLGA), are the most frequently used biomaterials to achieve sustained action [204, 205]. Polymeric microspheres are commonly prepared by the solvent extraction/evaporation methods [206]. In brief, protein in a solid or liquid form is mixed with a polymer (dissolved in an organic solvent, e.g., dichloromethane) to prepare a solid-

in-oil or water-in-oil emulsion. The resulting emulsion is further mixed with water containing an emulsifier (e.g., polyvinyl alcohol) to effect protein loading into the PLGA microspheres [207]. Release of protein from such microspheres usually occurs in three phases, namely, initial burst release, diffusion-controlled release, and erosion-controlled release.

The intrinsic drawbacks with these microencapsulation methods are that the organic solvents used and the low pH generated inside the matrix due to PLGA degradation may denature the protein. As an alternative for these biodegradable polyesters, hydrogels, which are prepared from natural or synthetic hydrophilic polymers, have been evaluated for use as protein-releasing matrices either alone or in combination with PLGA (i.e., as composite microspheres). Using the latter approach, a novel sustained-release insulin formulation has been developed by incorporating insulin into the hydrogel to yield microparticles, followed by encapsulation in a PLGA matrix as microspheres [208]. The swelling behavior of hydrogels, which governs drug release, is dependent on the external environmental conditions, such as temperature and pH. Hydrogel-based delivery devices can be used for oral, rectal, ocular, epidermal, and subcutaneous application. Excellent reviews on this topic can be found in literature [209, 210].

In general, proteins are more stable in the solid state than in the liquid state. A variety of methods are available for preparation of protein powders, including precipitation, lyophilization, and spray-drying. The main concern with these methods is the potential protein denaturation or inactivation during the preparation. More specialized techniques for preparing protein materials with specific applications have also been developed, such as supercritical fluid processing for human growth hormone [211], spray-drying for producing inhaled insulin [212], spray–freeze-drying into liquid to produce stable bovine serum albumin (BSA) nanostructured microparticles [213], supercritical fluid-based coating technology to produce lipid coated BSA particles [214], and emulsion precipitation [207]. Maa and Prestrelski [215] have presented a systematic review on protein powder production techniques. Table 3 shows a comparison of the advantages and disadvantages of these powder preparation methods. For assessment of the utility of these techniques, it is important that the resulting powders be subjected to rigorous physical and chemical characterization, including crystal form, particle morphology, particle size distribution, surface properties, and moisture content [212].

### 5.2 Nucleic Acid-based Drugs

Nucleic acid-based drugs also represent an important group of biopharmaceuticals that are still at their very early stage of development. Nucleic acids are complex, and high-molecular-weight macromolecules composed of nucleotide chains that convey genetic information. The most common nucleic acids are DNA and RNA. RNA is usually single stranded, whereas DNA is normally double stranded and can form a double helix via hydrogen bonding. Single-stranded RNA molecules tend to associate with one another via hydrogen bonding to attain the minimum energy state. DNA and RNA can be characterized for primary sequence by biochemical techniques [e.g., Southern blotting and polymerase chain reaction (PCR) for DNA and Northern blotting and reverse transcription-PCR (RT-PCR) for RNA]. They can also be analyzed in very much the same way as proteins by spectroscopic and

**TABLE 3. Advantages and Disadvantages of Different Powder Preparation Methods (from Ref. 215)**

| Method | Advantage | Disadvantage |
|---|---|---|
| Freeze-drying | High yield; convenient operation; aseptic process; expensive process | Poor particle size control; irregular-shaped particles; broad size distribution |
| Spray-drying | Good particle size control; spherical-shaped particles; easy and convenient operation | Yield dependent on formulation; heat inactivation; surface denaturation |
| Spray–freeze-drying | Good particle size control; spherical-shaped particles; high yield; good aerosol properties | Poor density control |
| Supercritical fluid | Good particle size control; spherical-shaped particles | Complex process; limited solubility in organic solvents |
| Pulverization | Easy operation; high yield | Poor particle size control; irregular-shaped particles; broad size distribution |
| Precipitation | Simple and convenient operation; high yield | Poor particle size control; irregular-shaped particles; broad size distribution; protein denaturation; difficult to handle multicomponents |
| Emulsification followed by precipitation | Easy to control; particle size <3 μm; spherical-shaped particles; applicable to a variety of materials | Yield limited by solubility; protein denaturation by organic solvent; difficult to handle multicomponents |

chromatographic techniques, including ultraviolet, infrared, Raman, fluorescence spectroscopy, circular dichroism, NMR, mass spectrometry, and HPLC [46].

Nucleic acid-based therapeutics aims to deliver nucleic acid-derived products into the nuclear compartment, which may act at the molecular genetic level either to replace a defective gene or to modulate a specific genetic function. Six types of nucleic acid-based therapeutics have been proposed, namely, antisense (molecules that interact with complementary strands of nucleic acids and modify expression of genes), ribonucleic acid inhibition (RNAi), gene therapy (insertion of genes into an individual's cells and tissues to treat a disease, particularly hereditary diseases), nucleoside analogs, ribozymes (RNA enzymes or catalytic RNAs that catalyze chemical reactions), and aptamers (DNAs or RNAs that are selected from random pools based on their ability to bind nucleic acids, proteins, or small organic compounds). Only one antisense drug named fomivirsen (Vitravene® developed by Isis Pharmaceuticals, Carlsbad, California) has been approved by the FDA for the treatment of cytomegalovirus retinitis and is currently available in the clinic, whereas others are still under development.

Compared with other organic drugs, nucleic acids are relatively large, hydrophilic, and negatively charged molecules. Systemic delivery of nucleic acids targeted at the nuclear compartment is limited by the hostile extracellular environment, including extreme pH, degradation enzymes (proteases and nucleases), as well as the immune defense and scavenger systems. Lechardeur et al. [216] have summarized the major cellular, metabolic, and physico-chemical barriers to the delivery of plasmid DNA into nucleus. A series of strategies employing physical forces (e.g., microinjection, naked DNA injection, electroporation, and gene gun), nonviral gene carriers (e.g., cationic liposome or cationic polymers), and/or biological carriers (e.g., viral gene carrier) have been developed to deliver nucleic-acid-derived drugs into cells [217]. Viral gene carriers, which rely on cell infection by live (nonpathogenic) viruses (e.g., adenovirus) to achieve specific site targeting, may cause severe immune response in the patient, and their delivery is limited by the gene size, which can be accommodated by the viral genome. Nonviral gene medicine is composed of three elements: a gene encoding a therapeutic protein, a plasmid-based gene expression system that controls the function of a gene within a target cell, and a synthetic gene carrier that controls the stability and gene delivery within the body. The most extensively investigated synthetic gene carriers involve the combination of plasmid DNA with either cationic lipids (lipoplexes) or polymers (polyplexes), or in association with encapsulation into microspheres or preparation of nanoparticulate systems.

Lipoplexes are prepared by the interaction of anionic nucleic acids with the surface of cationic lipid to afford multilamellar lipid–nucleic acid complexes. Cationic liposomes can protect nucleic acids from serum nucleases and facilitate the cellular uptake and release of nucleic acids into the cytosol [218]. Pedroso et al. [219] have extensively discussed the structure-activity relationships of cationic liposome/DNA complexes and the key formulation parameters influencing the properties of lipoplexes. In addition, optimization of the cationic liposomal complexes for *in vivo* application has been reviewed by Smyth [220].

Polyplexes are formulated by condensation of negatively charged nucleic acids via electrostatic interactions with polycationic condensing polymers to yield compact particles, which can then protect the nucleic acids and mask the negative DNA charges. The polycations used include natural DNA binding proteins such as histones, synthetic polymers such as poly[dimethylaminoethyl-methacrylates], linear or branched polyethylenimine, diethylaminoethyl modified dextran, or modified chitosan. The physico-chemical and biopharmaceutical behaviors of these polymers in nucleic acid delivery have been extensively reviewed elsewhere [221, 222]. Besides these polymers, polyethylene glycol-conjugated copolymers have also been introduced as an alternative drug carrier material [223]. However, these polyplexes are incapable of affording sustained release for nucleic acids and prolonging gene transfer. To achieve prolonged action of nucleic-acid-based drugs, the feasibility of encapsulating polyplexes into certain matrices to obtain sustained-release nanoparticles or microparticles is currently under investigation. Yun et al. [224] have prepared microspheres by physically combining poly(ethylene glycol)-grafted chitosan with poly(lactide-co-glycolide) using a modified conventional in-emulsion solvent evaporation method to achieve prolonged delivery of DNA/chitosan polyplexes.

Nanoparticles are submicronic (less than 1 μm) colloidal systems, which can be classified into two main categories depending on whether the formation of

nanoparticles requires a polymerization reaction. Nanoparticulate systems have been developed to deliver antisense oligonucleotides with enhanced stability and delivery efficiency [225].

## 6 SUMMARY AND CONCLUSIONS

Biopharmaceuticals constitute a unique class of therapeutic agents that are difficult to handle and process because of their structural complexity and inherent chemical/physical instability. Depending on the modes of chemical degradation or physical denaturation, a wide variety of formulation excipients or additives, including buffers, antioxidants, sugars, and polysaccharides, can be employed to preserve the structural integrity and biological activity of such vulnerable macromolecules. In addition, an appropriate choice of administration route is critical for maximizing their bioavailability and efficacy. Sustained or prolonged action for specific therapeutic indications has also been made possible by formulating them in synthetic polymeric carriers, including biodegradable polymers, hydrogels, and polycationic polymers. It can be envisioned that as more is known about the relationship between the structures and the bioactivities of these vital macromolecules, the existing formulation strategies can be further improved or fine-tuned for particular biopharmaceuticals to achieve more promising therapeutic outcomes.

## REFERENCES

1. Clarke S, Stephenson R C, Lowenson J D (1992). Lability of asparagine and aspartic acid residues in proteins and peptides: Spontaneous deamidation and isomerization reactions. In T J Ahern, M C Manning, (eds.), *Stability of Protein Pharmaceuticals: Part A Chemical and Physical Pathways of Protein Degradation*, Plenum Press, New York, pp. 1–29.
2. Brennan T V, Clarke S (1995). Deamidation and isoaspartate formation in model synthetic peptides: The effects of sequence and solution environment. In D W Aswad, (eds.), *Deamidation and Isoaspartate Formation in Peptides and Proteins*, CRC Press, Boca Raton F.L., pp. 65–90.
3. Goolcharran C, Khossravi M, Borchardt R T (2000). Chemical pathways of peptide and protein degradation. In S Frokjaer, L Hovgaard, (eds.), *Pharmaceutical Formulation Development of Peptides and Proteins*, Taylor & Francis, London, pp. 70–88.
4. Manning M C, Patel K, Borchardt R T (1989). Stability of protein pharmaceuticals. *Pharm. Res.* 6:903–918.
5. Schoneich C, Hageman M J, Borchardt R T (1997). Stability of peptides and proteins. In K Park, (ed.), *Controlled Drug Delivery*, American Chemical Society, Washington, D.C., pp. 205–228.
6. Nabuchi Y, Fujiwara E, Kuboniwa H, et al. (1997). The stability and degradation pathway of recombinant human parathyroid hormone: Deamidation of asparaginyl residue and peptide bond cleavage at aspartyl and asparaginyl residues. *Pharm. Res.* 14:1685–1690.
7. Kroon D J, Baldwin-Ferro A, Lalan P (1992). Identification of sites of degradation in a therapeutic monoclonal antibody by peptide mapping. *Pharm. Res.* 9:1386–1393.
8. Son K, Kwon C (1995). Stabilization of human epidermal growth factor (hEGF) in aqueous formulation. *Pharm. Res.* 12:451–454.

9. Darrington R T, Anderson B D (1995). Effects of insulin concentration and self-association on the partitioning of its A-21 cyclic anhydride intermediate to desamido insulin and covalent dimer. *Pharm. Res.* 12:1077–1084.
10. Wolfe J L, Lee G E, Potti G K, et al. (1994). Degradation of antiflammin 2 in aqueous solution. *J. Pharm. Sci.* 83:1762–1764.
11. Aswad D W, Paranandi M V, Schurter B T (2000). Isoaspartate in peptides and proteins: Formation, significance and analysis. *J. Pharm. Biomed. Anal.* 21:1129–1136.
12. Tsai P K, Bruner M W, Irwin J I, et al. (1993). Origin of the isoelectric heterogeneity of monoclonal immunoglobulin h1B4. *Pharm. Res.* 10:1580–1586.
13. Perkins M, Theiler R, Lunte S, et al. (2000). Determination of the origin of charge heterogeneity in a murine monoclonal antibody. *Pharm. Res.* 17:1110–1117.
14. Blanche F, Cameron B, Somarriba S, et al. (2001). Stabilization of recombinant adenovirus: Site-directed mutagenesis of key asparagine residues in the hexon protein. *Anal. Biochem.* 297:1–9.
15. Joshi A B, Kirsch L E (2002). The relative rates of glutamine and asparagine deamidation in glucagons fragment 22–29 under acidic conditions. *J. Pharm. Sci.* 91:2332–2345.
16. Li B, Gorman E M, Moore K D, et al. (2005). Effects of acidic N + 1 residues on asparagine deamidation rates in solution and in the solid state. *J. Pharm. Sci.* 94:666–675.
17. Xie M, Schowen R L (1999). Secondary structure and protein deamidation. *J. Pharm. Sci.* 88:8–13.
18. Xie M, Shahrokh Z, Kadkhodayan M, et al. (2003). Asparagine deamidation in recombinant human lymphotoxin: Hindrance by three-dimensional structures. *J. Pharm. Sci.* 92:869–880.
19. Geiger T, Clarke S (1987). Deamidation, isomerization, and racemization at asparaginyl and aspartyl residues in peptides. *J. Biol. Chem.* 262:785–794.
20. Patel K, Borchardt R T (1990). Chemical pathways of peptide degradation. III. Effect of primary sequence on the pathways of deamidation of asparaginyl residues in hexapeptides. *Pharm. Res.* 7:787–793.
21. Pearlman R, Bewley T A (1993). Stability and characterization of human growth hormone. In Y J Wang, R Pearlman, (eds.), *Stability and Characterization of Protein and Peptide Drugs: Case Histories*, Plenum Press, New York, pp. 1–58.
22. Brange J (1994). *Stability of Insulin: Studies on the Physical and Chemical Stability of Insulin in Pharmaceutical Formulation*, Kluwer Academic Publishers, Dordrecht, pp. 6–59.
23. Jars M U, Hvass A, Waaben D (2002). Insulin aspart ($Asp^{B28}$ human insulin) derivatives formed in pharmaceutical solutions. *Pharm. Res.* 19:621–628.
24. Markell D, Hui J, Narhi L, et al. (2001). Pharmaceutical significance of the cyclic imide form of recombinant human glial cell line derived neurotrophic factor. *Pharm. Res.* 18:1361–1366.
25. Teshima G, Hancock W S, Canova-Davis E (1995). Effect of deamidation and isoaspartate formation on the activity of proteins. In D W Aswad, (ed.), *Deamidation and Isoaspartate Formation in Peptides and Proteins*, CRC Press, Boca Raton F.L., pp. 167–191.
26. Sasaoki K, Hiroshima T, Kusumoto S, et al. (1992). Deamidation at asparagine-88 in recombinant human interleukin 2. *Chem. Pharm. Bull.* 40:976–980.
27. Powell M F (1994). Peptide stability in aqueous parenteral formulations. In J L Cleland, R Langer, (eds.), *Formulation and Delivery of Proteins and Peptides*, American Chemical Society, Washington D.C., pp. 100–117.

28. Ye J M, Lee G E, Potti G K, et al. (1996). Degradation of antiflammin 2 under acidic conditions. *J. Pharm. Sci.* 85:695–699.
29. Stephenson R C, Clarke S (1989). Succinimide formation from aspartyl and asparaginyl peptides as a model for the spontaneous degradation of proteins. *J. Biol. Chem.* 264:6164–6170.
30. Oliyai C, Borchardt R T (1994). Chemical pathways of peptide degradation. VI. Effect of primary sequence on the pathways of degradation of aspartyl residues in model hexapeptides. *Pharm. Res.* 11:751–758.
31. Piszkiewicz D, Landon M, Smith E L (1970). Anomalous cleavage of aspartyl-proline peptide bonds during amino acid sequence determinations. *Biochem. Biophys. Res. Commun.* 40:1173–1178.
32. Kenley R A, Warne N W (1994). Acid-catalyzed peptide bond hydrolysis of recombinant human interleukin 11. *Pharm. Res.* 11:72–76.
33. Strickley R G, Brandl M, Chan K W, et al. (1990). High-performance liquid chromatographic (HPLC) and HPLC-mass spectrometric (MS) analysis of the degradation of the luteinizing hormone-releasing hormone (LH-RH) antagonist RS-26306 in aqueous solution. *Pharm. Res.* 7:530–536.
34. Windisch V, DeLuccia F, Duhau L, et al. (1997). Degradation pathways of salmon calcitonin in aqueous solution. *J. Pharm. Sci.* 86:359–364.
35. Kertscher U, Bienert M, Krause E, et al. (1993). Spontaneous chemical degradation of substance P in the solid phase and in solution. *Int. J. Peptide Protein Res.* 41:207–211.
36. Battersby J E, Hancock W S, Canova-Davis E, et al. (1994). Diketopiperazine formation and *N*-terminal degradation in recombinant human growth hormone. *Int. J. Peptide Protein Res.* 44:215–222.
37. Oyler A R, Naldi R E, Lloyd J R, et al. (1991). Characterization of the solution degradation products of histrelin, a gonadotrophin releasing hormone (LH/RH) agonist. *J. Pharm. Sci.* 80:271–275.
38. Straub J A, Akiyama A, Parmar P, et al. (1995). Chemical pathways of degradation of the bradykinin analog, RMP-7. *Pharm. Res.* 12:305–308.
39. Leung S S, Grant D J W (1997). Solid state stability studies of model dipeptides: Aspartame and aspartylphenylalanine. *J. Pharm. Sci.* 86:64–71.
40. Goolcharran C, Borchardt R T (1998). Kinetics of diketopiperazine formation using model peptides. *J. Pharm. Sci.* 87:283–288.
41. Orlowska A, Witkowska E, Izdebski J (1987). Sequence dependence in the formation of pyroglutamyl peptides in solid phase peptide synthesis. *Int. J. Peptide Protein Res.* 30:141–144.
42. Beck A, Bussat M C, Klinguer-Hamour C, et al. (2001). Stability and CTL activity of *N*-terminal glutamic acid containing peptides. *J. Pept. Res.* 57:528–538.
43. Evans R K, Xu Z, Bohannon K E, et al. (2000). Evaluation of degradation pathways for plasmid DNA in pharmaceutical formulations via accelerated stability studies. *J. Pharm. Sci.* 89:76–87.
44. Pogocki D, Schoneich C (2000). Chemical stability of nucleic acid-derived drugs. *J. Pharm. Sci.* 89:443–456.
45. Lindahl T (1993). Instability and decay of the primary structure of DNA. *Nature.* 362:709–715.
46. Middaugh C R, Evans R K, Montgomery D L, et al. (1998). Analysis of plasmid DNA from a pharmaceutical perspective. *J. Pharm. Sci.* 87:130–146.

47. Glaser R, Rayat S, Lewis M, et al. (1999). Theoretical studies of DNA base deamination. 2. Ab initio study of DNA base diazonium ions and of their linear, unimolecular dediazoniation paths. *J. Am. Chem. Soc.* 121:6108–6119.
48. Karran P, Lindahl T (1980). Hypoxanthine in deoxyribonucleic acid: Generation by heat-induced hydrolysis of adenine residues and release in free form by a deoxyribonucleic acid glycosylase from calf thymus. *Biochem.* 19:6005–6011.
49. Li S, Schoneich C, Borchardt R T (1995). Chemical instability of protein pharmaceuticals: Mechanisms of oxidation and strategies for stabilization. *Biotechnol. Bioeng.* 48:490–500.
50. Meucci E, Mordente A, Martorana G E (1991). Metal-catalyzed oxidation of human serum albumin: Conformational and functional changes. *J. Biol. Chem.* 266: 4692–4699.
51. Li S, Nguyen T H, Schoneich C, et al. (1995). Aggregation and precipitation of human relaxin induced by metal-catalyzed oxidation. *Biochem.* 34:5762–5772.
52. Wells J A, Powers D B, Bott R R (1987). Protein engineering of subtilisin. In D L Oxender, C F Fox, (eds.), *Protein Engineering*: Alan R. Liss, Inc. New York, pp. 279–287.
53. Dubost D C, Kaufman M J, Zimmerman J A, et al. (1996). Characterization of a solid state reaction product from a lyophilized formulation of a cyclic heptapeptide. A novel example of an excipient-induced oxidation. *Pharm. Res.* 13:1811–1814.
54. Nguyen T H (1994). Oxidation degradation of protein pharmaceuticals. In J L Cleland, R Langer, (eds.), *Formulation and Delivery of Proteins and Peptides*, American Chemical Society, Washington, D.C., pp. 59–71.
55. Chu J W, Yin J, Wang D I C, et al. (2004a). Molecular dynamics simulations and oxidation rates of methionine residues of granulocyte colony-stimulating factor at different pH values. *Biochem.* 43:1019–1029.
56. Chu J W, Yin J, Wang D I C, et al. (2004b). A structural and mechanistic study of the oxidation of methionine residues in hPTH(1-34) via experiments and simulations. *Biochem.* 43:14139–14148.
57. Luo D, Smith S W, Anderson B D (2005). Kinetics and mechanism of the reaction of cysteine and hydrogen peroxide in aqueous solution. *J. Pharm. Sci.* 94:304–316.
58. Shioi S, Imoto T, Ueda T (2004). Analysis of the early stage of the folding process of reduced lysozyme using all lysozyme variants containing a pair of cysteines. *Biochem.* 43:5488–5493.
59. Shaked Z, Szajewski R P, Whitesides G M (1980). Rates of thiol-disulfide interchange reactions involving proteins and kinetic measurements of thiol pKa values. *Biochem.* 19:4156–4166.
60. Parente A, Merrifield B, Geraci G, et al. (1985). Molecular basis of superreactivity of cysteine residues 31 and 32 of seminal ribonucleases. *Biochem.* 24:1098–1104.
61. Kuwajima K, Ikeguchi M, Sugawara T, et al. (1990). Kinetics of disulfide bond reduction in α-lactalbumin by dithiothreitol and molecular basis of superreactivity of the cys6–cys120 disulfide bond. *Biochem.* 29:8240–8249.
62. Fransson J, Florin-Robertsson E, Axelsson K, et al. (1996). Oxidation of human insulin-like growth factor I in formulation studies: Kinetics of methionine oxidation in aqueous solution and in solid state. *Pharm. Res.* 13:1252–1257.
63. Fransson J, Hagman A (1996). Oxidation of human insulin-like growth factor I in formulation studies, II. Effects of oxygen, visible light, and phosphate on methionine oxidation in aqueous solution and evaluation of possible mechanisms. *Pharm. Res.* 13:1476–1481.

64. Schoneich C, Zhao F, Yang J (1997). Mechanisms of methionine oxidation in peptides. In Z Shahrokh, V Sluzky, J L Cleland, et al. (eds.), *Therapeutic Protein and Peptide Formulation and Delivery, ACS Symp. Ser. 675*, American Chemical Society, Washington, D.C., pp. 79–89.
65. Schoneich C (2000). Mechanisms of metal-catalyzed oxidation of histidine to 2-oxo-histidine in peptides and proteins. *J. Pharm. Biomed. Anal.* 21:1093–1097.
66. Khossravi M, Shire S J, Borchardt R T (2000). Evidence for the involvement of histidine A(12) in the aggregation and precipitation of human relaxin induced by metal-catalyzed oxidation. *Biochem.* 39:5876–5885.
67. Tomita M, Irie M, Ukita T (1969). Sensitized photooxidation of histidine and its derivatives. Products and mechanism of the reaction. *Biochem.* 8:5149–5160.
68. Khossravi M, Borchardt R T (2000). Chemical pathways of peptide degradation. X: Effect of metal catalyzed oxidation on the solution structure of a histidine-containing peptide fragment of human relaxin. *Pharm. Res.* 17:851–858.
69. Bummer P M, Koppenol S (2000). Chemical and physical considerations in protein and peptide stability. In E J McNally, (eds.), *Protein Formulation and Delivery*, Marcel Dekker, Inc. New York, pp. 5–69.
70. Miller B L, Hageman M J, Thamann T J, et al. (2003). Solid-state photodegradation of bovine somatotropin (bovine growth hormone): evidence for tryptophan-mediated photooxidation of disulfide bonds. *J. Pharm. Sci.* 92:1698–1709.
71. Kanwar R, Balasubramanian D (1999). Structure and stability of the dityrosine-linked dimer of γB-crystallin. *Exper. Eye Res.* 68:773–784.
72. Feibush B, Snyder B (2000). Oxidation of the N-terminal Gly-residue of peptides: Stress study of pexiganan acetate in a drug formulation. *Pharm. Res.* 17:197–204.
73. Meneghini R, Martins E A L, Calderaro M (1993). DNA damage by reactive oxygen species. The role of metals. In G Poli, E Albano, M U Dianzani, (eds.), *Free Radicals: From Basic Science to Medicine*, Basel, Birlhauser Verlag, pp.102–113.
74. Beckman K B, Ames B N (1997). Oxidative decay of DNA. *J. Biol. Chem.* 272: 19633–19636.
75. Armitage B (1998). Photocleavage of nucleic acids. *Chem. Rev.* 98:1171–1200.
76. Jaruga P, Dizdaroglu M (1996). Repair of products of oxidative DNA base damage in human cells. *Nucleic Acids Res.* 24:1389–1394.
77. Henle E S, Linn S (1997). Formation, prevention and repair of DNA damage by iron/hydrogen peroxide. *J. Biol. Chem.* 272:19095–19098.
78. Burrows C J, Muller J G (1998). Oxidative nucleobase modifications leading to strand scission. *Chem. Rev.* 98:1109–1151.
79. Evans M D, Dizdaroglu M, Cooke M S (2004). Oxidative DNA damage and disease: induction, repair and significance. *Mutation Res.* 567:1–61.
80. Pogozelski W K, Tullius T D (1998). Oxidative strand scission of nucleic acids: Routes initiated by hydrogen abstraction from the sugar moiety. *Chem. Rev.* 98:1089–1107.
81. Suzuki T, Ohsumi S, Makino K (1994). Mechanistic studies on depurination and apurinic site chain breakage in oligodeoxyribonucleotides. *Nucleic Acids Res.* 22:4997–5003.
82. Hall S C, Tan M M, Leonard J J, et al. (1998). Characterization and comparison of leuprolide degradation profiles in water and dimethyl sulfoxide. *J. Pept. Res.* 53:432–441.
83. Stevenson C L, Leonard J J, Hall S C (1999). Effect of peptide concentration and temperature on leuprolide stability in dimethyl sulfoxide. *Int. J. Pharm.* 191:115–129.

84. Costantino H R, Langer R, Klibanov A M (1994). Moisture-induced aggregation of lyophilized insulin. *Pharm. Res.* 11:21–29.
85. Colaco C A L S, Smith C J S, Sen S, et al. (1994). Chemistry of protein stabilization by trehalose. In J L Cleland, R Langer, (eds.), *Formulation and Delivery of Proteins and Peptides*, American Chemical Society, Washington, D.C., pp. 222–240.
86. Lai M C, Topp E M (1999). Solid-state chemical stability of proteins and peptides. *J. Pharm. Sci.* 88:489–500.
87. Li S, Patapoff T W, Overcashier D, et al. (1996). Effects of reducing sugars on the chemical stability of human relaxin in the lyophilized state. *J. Pharm. Sci.* 85: 873–877.
88. Smales C M, Pepper D S, James D C (2000a). Mechanisms of protein modification during model anti-viral heat-treatment bioprocessing of β-lactoglobulin variant A in the presence of sucrose. *Biotechnol. Appl. Biochem.* 32:109–119.
89. Smales C M, Pepper D S, James D C (2000b). Protein modification during antiviral heat bioprocessing. *Biotechnol. Bioeng.* 67:177–188.
90. Smales C M, Pepper D S, James D C (2002). Protein modification during anti-viral heat-treatment bioprocessing of factor VIII concentrates, factor IX concentrates, and model proteins in the presence of sucrose. *Biotechnol. Bioeng.* 77:37–48.
91. Bucala R, Model P, Cerami A (1984). Modification of DNA by reducing sugars: A possible mechanism for nucleic acid aging and age-related dysfunction in gene expression. *Proc. of the National Academy of Sciences of the United States of America*, 81:105–109.
92. Lee A T, Cerami A (1987). Elevated glucose 6-phosphate levels are associated with plasmid mutations *in vivo*. *Proc. of the National Academy of Sciences of the United States of America*, 84:8311–8314.
93. Davis P J, Smales C M, James D C (2001). How can thermal processing modify the antigenicity of proteins? *Allergy.* 56(Suppl 67):56–60.
94. Brange J, Havelund S, Hougaard P (1992). Chemical stability of insulin. 2. Formation of higher molecular weight transformation products during storage of pharmaceutical preparations. *Pharm. Res.* 9:727–734.
95. Brange J, Langkjaer L, Havelund S, et al. (1992). Chemical stability of insulin. 1. Hydrolytic degradation during storage of pharmaceutical preparations. *Pharm. Res.* 9:715–726.
96. Darrington R T, Anderson B D (1995). Evidence for a common intermediate in insulin deamidation and covalent dimer formation: Effects of pH and aniline trapping in dilute acidic solutions. *J. Pharm. Sci.* 84:275–282.
97. Strickley R G, Anderson B D (1996). Solid-state stability of human insulin I. Mechanism and the effect of water on the kinetics of degradation in lyophiles from pH 2–5 solutions. *Pharm. Res.* 13:1142–1153.
98. Hora M S, Rana R K, Smith F W (1992). Lyophilized formulations of recombinant tumor necrosis factor. *Pharm. Res.* 9:33–36.
99. Fransson J R (1997). Oxidation of human insulin-like growth factor I in formulation studies. 3. Factorial experiments of the effects of ferric ions, EDTA, and visible light on methionine oxidation and covalent aggregation in aqueous solution. *J. Pharm. Sci.* 86:1046–1050.
100. Baldwin R L, Eisenberg D (1987). Protein stability. In D L Oxender, C F Fox, (eds.), *Protein Engineering*: Alan R. Liss, Inc., New York, pp. 127–148.
101. Chi E Y, Krishnan S, Kendrick B S, et al. (2003). Roles of conformational stability and colloidal stability in the aggregation of recombinant human granulocyte colony-stimulating factor. *Protein Sci.* 12:903–913.

102. Roy I, Gupta M N (2004). Freeze-drying of proteins: Some emerging concerns. *Biotechnol. Appl. Biochem.* 39:165–177.
103. Zhou H X, Vijayakumar M (1997). Modeling of protein conformational fluctuations in $pK_a$ predictions. *J. Mol. Biol.* 267:1002–1011.
104. Shirley B A (1992). Protein conformational stability estimated from urea, guanidine hydrochloride, and thermal denaturation curves. In T J Ahern, M C Manning, (eds.), *Stability of Protein Pharmaceuticals: Part A: Chemical and Physical Pathways of Protein Degradation*: Plenum Press, New York, pp. 167–194.
105. Dubin D N, Lee A, Macgregor R B, et al. (2001). On the stability of double stranded nucleic acids. *J. Amer. Chem. Soc.* 123:9254–9259.
106. Tikhomirova A, Taulier N, Chalikian T V (2004). Energetics of nucleic acid stability: The effect of $\Delta C_P$. *J. Amer. Chem. Soc.* 126:16387–16394.
107. Chen B L, Arakawa T, Morris C F, et al. (1994). Aggregation pathway of recombinant human keratinocyte growth factor and its stabilization. *Pharm. Res.* 11:1581–1587.
108. Bermudez O, Forciniti D (2004). Aggregation and denaturation of antibodies: a capillary electrophoresis, dynamic light scattering, and aqueous two-phase partitioning study. *J. Chromatogr. B.* 807:17–24.
109. Klibanov A M, Schefiliti J A (2004). On the relationship between conformation and stability in solid pharmaceutical protein formulations. *Biotechnol. Lett.* 26:1103–1106.
110. Kendrick B S, Carpenter J F, Cleland J L, et al. (1998). A transient expansion of the native state precedes aggregation of recombinant human interferon-γ. *Proc. of the National Academy of Sciences of the United States of America*, 95:14142–14146.
111. Chi E Y, Krishnan S, Randolph T W, et al. (2003). Physical stability of proteins in aqueous solution: Mechanism and driving forces in nonnative protein aggregation. *Pharm. Res.* 20:1325–1336.
112. Krishnamurthy R, Manning M C (2002). The stability factor: Importance in formulation development. *Curr. Pharm. Biotechnol.* 3:361–371.
113. Costantino H R, Langer R, Klibanov A M (1994). Solid-phase aggregation of proteins under pharmaceutically relevant conditions. *J. Pharm. Sci.* 83:1662–1669.
114. Hermeling S, Crommelin D J A, Schellekens H, et al. (2004). Structure-immunogenicity relationships of therapeutic proteins. *Pharm. Res.* 21:897–903.
115. Paborji M, Pochopin N L, Coppola W P, et al. (1994). Chemical and physical stability of chimeric L6, a mouse-human monoclonal antibody. *Pharm. Res.* 11:764–771.
116. Krishnan S, Chi E Y, Webb J N, et al. (2002). Aggregation of granulocyte colony stimulating factor under physiological conditions: characterization and thermodynamic inhibition. *Biochem.* 41:6422–6431.
117. Schwendeman S P, Costantino H R, Gupta R K, et al. (1995). Stabilization of tetanus and diphtheria toxoids against moisture-induced aggregation. *Proc. of the National Academy of Sciences of the United States of America.* 92:11234–11238.
118. Tavornvipas S, Tajiri S, Hirayama F, et al. (2004). Effects of hydrophilic cyclodextrins on aggregation of recombinant human growth hormone. *Pharm. Res.* 21:2369–2376.
119. Fransson J, Hallen D, Florin-Robertsson E (1997). Solvent effects on the solubility and physical stability of human insulin-like growth factor I. *Pharm. Res.* 14:606–612.
120. Zhang Y, Roy S, Jones L S, et al. (2004). Mechanism for benzyl alcohol-induced aggregation of recombinant human interleukin-1 receptor antagonist in aqueous solution. *J. Pharm. Sci.* 93:3076–3089.

121. Roy S, Jung R, Kerwin B A, et al. (2005). Effects of benzyl alcohol on aggregation of recombinant human interleukin-1-receptor antagonist in reconstituted lyophilized formulations. *J. Pharm. Sci.* 94:382–396.
122. Horbett T A (1992). Adsorption of proteins and peptides at interfaces. In T J Ahern, M C Manning, (eds.), *Stability of Protein Pharmaceuticals: Part A: Chemical and Physical Pathways of Protein Degradation*: Plenum Publishing Corporation, New York, pp. 195–214.
123. Brange J (2000). Physical stability of proteins. In S Frokjaer, L Hovgaard, (eds.), *Pharmaceutical Formulation Development of Peptides and Proteins*: Taylor & Francis, London, pp. 89–112.
124. Page C, Dawson P, Woollacott D, et al. (2000). Development of a lyophilization formulation that preserves the biological activity of the platelet-inducing cytokine interleukin-11 at low concentrations. *J. Pharm. Pharmacol.* 52:19–26.
125. Tzannis S T, Prestrelski S J (1999). Moisture effects on protein-excipient interactions in spray-dried powders. Nature of destabilizing effects of sucrose. *J. Pharm. Sci.* 88:360–370.
126. Fatouros A, Sjostrom B (2000). Recombinant factor VIII SQ—the influence of formulation parameters on structure and surface adsorption. *Int. J. Pharm.* 194:69–79.
127. Evans R K, Nawrocki D K, Isopi L A, et al. (2004). Development of stable liquid formulations for adenovirus-based vaccines. *J. Pharm. Sci.* 93:2458–2475.
128. Webb S D, Golledge S L, Cleland J L, et al. (2002). Surface adsorption of recombinant human interferon-γ in lyophilized and spray-lyophilized formulations. *J. Pharm. Sci.* 91:1474–1487.
129. Chang S L, Hofmann G A, Zhang L, et al. (2003). Stability of a transdermal salmon calcitonin formulation. *Drug Del.* 10:41–45.
130. Matheis W, Zott A, Schwanig M (2001). The role of the adsorption process for production and control combined adsorbed vaccines. *Vaccine.* 20:67–73.
131. Maa Y F, Shu C, Ameri M, et al. (2003). Optimization of an alum-adsorbed vaccine power formulation for epidermal powder immunization. *Pharm. Res.* 20:969–977.
132. Waterman K C, Adami R C (2005). Accelerated aging: Prediction of chemical stability of pharmaceuticals. *Int. J. Pharm.* 293:101–125.
133. Roberts C, Darrington R T, Whitley M B (2003). Irreversible aggregation of recombinant bovine granulocyte-colony stimulating factor (bG-CSF) and implications for predicting protein shelf life. *J. Pharm. Sci.* 92:1095–1111.
134. Volkin D B, Middaugh C R (1992). The effect of temperature on protein structure. In T J Ahern, M C Manning, *Stability of Protein Pharmaceuticals: Part A: Chemical and Physical Pathways of Protein Degradation*: Plenum Press, New York, pp. 215–247.
135. Pikal-Cleland K A, Cleland J L, Anchordoquy T J, et al. (2002). Effect of glycine on pH changes and protein stability during freeze-thawing in phosphate buffer systems. *J. Pharm. Sci.* 91:1969–1979.
136. Cao E, Chen Y, Cui Z, et al. (2003). Effect of freezing and thawing rates on denaturation of proteins in aqueous solutions. *Biotechnol. Bioeng.* 82:684–690.
137. Ho M M, Mawas F, Bolgiano B, et al. (2002). Physico-chemical and immunological examination of the thermal stability of tetanus toxoid conjugate vaccines. *Vaccines.* 20:3509–3522.
138. Anchordoquy T J, Girouard L G, Carpenter J F, et al. (1998). Stability of lipid/DNA complexes during agitation and freeze-thawing. *J. Pharm. Sci.* 87:1046–1051.

139. Chang B S, Kendrick B S, Carpenter J F (1996). Surface-induced denaturation of proteins during freezing and its inhibition by surfactants. *J. Pharm. Sci.* 85:1325–1330.

140. Lai M C, Hageman M J, Schowen R L, et al. (1999). Chemical stability of peptides in polymers. 1. Effect of water on peptide deamidation in poly(vinyl alcohol) and poly(vinyl pyrrolidone) matrices. *J. Pharm. Sci.* 88:1073–1080.

141. Breen E D, Curley J G, Overcashier D E, et al. (2001). Effect of moisture on the stability of a lyophilized humanized monoclonal antibody formulation. *Pharm. Res.* 18:1345–1353.

142. Chan H K, Clark A R, Feeley J C, et al. (2004). Physical stability of salmon calcitonin spray-dried powders for inhalation. *J. Pharm. Sci.* 93:792–804.

143. Costantino H R, Langer R, Klibanov A M (1995). Aggregation of a lyophilized pharmaceutical protein, recombinant human albumin: Effect of moisture and stabilization by excipients. *Biotechnol.* 13:493–496.

144. Costantino H R, Carrasquillo K G, Cordero R A, et al. (1998). Effect of excipients on the stability and structure of lyophilized recombinant human growth hormone. *J. Pharm. Sci.* 87:1412–1420.

145. Tzannis S T, Hrushesky W J M, Wood P A, et al. (1996). Irreversible inactivation of interleukin 2 in a pump-based delivery environment. *Proc. of the National Academy of Sciences of the United States of America.* 93:5460–5465.

146. Sarciaux J M, Mansour S, Hageman M J, et al. (1999). Effects of buffer composition and processing conditions on aggregation of bovine IgG during freeze-drying. *J. Pharm. Sci.* 88:1354–1361.

147. Yamamoto O (1992). Effect of radiation on protein stability. In T J Ahern, M C Manning, *Stability of Protein Pharmaceuticals: Part A: Chemical and Physical Pathways for Protein Degradation*: Plenum Press, New York, pp. 361–421.

148. Parti R, Ardosa J, Yang L, et al. (2000). *In vitro* stability of recombinant human factor VIII (Recombinate™). *Haemophilia.* 6:513–522.

149. Gocke E (2001). Photochemical mutagenesis: Examples and toxicological relevance. *J. Environ. Pathol. Toxicol. Oncol.* 20:285–292.

150. Oliva A, Farina J B, Llabres M (1996). Influence of temperature and shaking on stability of insulin preparations: Degradation kinetics. *Int. J. Pharm.* 143:163–170.

151. Shnek D R, Hostettler D L, Bell M A, et al. (1998). Physical stress testing of insulin suspensions and solutions. *J. Pharm. Sci.* 87:1459–1465.

152. Bam N B, Cleland J L, Yang J, et al. (1998). Tween protects recombinant human growth hormone against agitation-induced damage via hydrophobic interactions. *J. Pharm. Sci.* 87:1554–1559.

153. Mahler H C, Muller R, Frieb W, et al. (2005). Induction and analysis of aggregates in a liquid IgG1-antibody formulation. *Euro. J. Pharm. Biopharm.* 59:407–417.

154. Oliva A, Santovena A, Farina J, et al. (2003). Effect of high shear rate on stability of proteins: kinetic study. *J. Pharm. Biomed. Anal.* 33:145–155.

155. Katakam M, Bell L N, Banga A K (1995). Effect of surfactants on the physical stability of recombinant human growth hormone. *J. Pharm. Sci.* 84:713–716.

156. Nail S L, Jiang S, Chongprasert S, et al. (2002). Fundamentals of freeze-drying. In S L Nail, M J Akers, (eds.), *Development and Manufacture of Protein Pharmaceuticals*: Kluwer Academic, New York, pp. 281–360.

157. Wang W (2000). Lyophilization and development of solid protein pharmaceuticals. *Int. J. Pharm.* 203:1–60.

158. Kreilgaard L, Frokjaer S, Flink J M, et al. (1999). Effect of additives on the stability of *Humicola lanuginose* lipase during freeze-drying and storage in the dried solid. *J. Pharm. Sci.* 88:281–290.
159. Liao Y H, Brown M B, Martin G P (2004). Investigation of the stabilization of freeze-dried lysozyme and the physical properties of the formulations. *Euro. J. Pharm. Biopharm.* 58:15–24.
160. Kreilgaard L, Frokjaer S, Flink J M, et al. (1998). Effects of additives on the stability of recombinant human factor XIII during freeze-drying and storage in the dried solid. *Arch. Biochem. Biophys.* 360:121–134.
161. Cherng J Y, Talsma H, Crommelin D J A, et al. (1999). Long term stability of poly((2-dimethylamino)ethyl methacrylate)-based gene delivery systems. *Pharm. Res.* 16:1417–1423.
162. Li B, Li S, Tan Y, et al. (2000). Lyophilization of cationic lipid-protamine-DNA (LPD) complexes. *J. Pharm. Sci.* 89:355–364.
163. Townsend M W, Deluca P P (1991). Nature of aggregates formed during storage of freeze-dried ribonuclease A. *J. Pharm. Sci.* 80:63–66.
164. Maa Y F, Nguyen P A T, Hsu S W (1998). Spray-drying of air-liquid interface sensitive recombinant human growth hormone. *J. Pharm. Sci.* 87:152–159.
165. Maa Y F, Nugyen P A, Andya J D, et al. (1998). Effect of spray drying and subsequent processing conditions on residual moisture content and physical/biochemical stability of protein inhalation powders. *Pharm. Res.* 15:768–775.
166. Costantino H R, Andya J D, Nguyen P A, et al. (1998). Effect of mannitol crystallization on the stability and aerosol performance of a spray-dried pharmaceutical protein, recombinant humanized anti-IgE monocolonal antibody. *J. Pharm. Sci.* 87:1406–1411.
167. Adler M, Lee G (1999). Stability and surface activity of lactate dehydrogenase in spray-dried trehalose. *J. Pharm. Sci.* 88:199–208.
168. Liao Y H, Brown M B, Nazir T, et al. (2002). Effects of sucrose and trehalose on the preservation of the native structure of spray-dried lysozyme. *Pharm. Res.* 19:1847–1853.
169. Lee G (2002). Spray-drying of proteins. In J F Carpenter, M C Manning, (eds.), *Rational Design of Stable Protein Formulations: Theory and Practice*: Kluwer Academic, New York, pp. 137–158.
170. Li S, Patapoff T W, Nguyen T H, Borchardt R T (1996). Inhibitory effect of sugars and polyols on the metal-catalyzed oxidation of human relaxin. *J. Pharm. Sci.* 85: 868–872.
171. Rao P E, Kroon D J (1993). Orthoclone OKT3: Chemical mechanisms and functional effects of degradation of a therapeutic monoclonal antibody. In Y J Wang, R Pearlman, (eds.), *Stability and Characterization of Protein and Peptide Drugs: Case Histories*: Plenum Publishing Corporation, New York, pp. 135–158.
172. Hovorka S W, Hong J Y, Cleland J L, et al. (2001). Metal-catalyzed oxidation of human growth hormone: Modulation by solvent-induced changes of protein conformation. *J. Pharm. Sci.* 90:58–69.
173. D'Souza A J M, Schowen R L, Borchardt R T, et al. (2003). Reaction of a peptide with polyvinylpyrrolidone in the solid state. *J. Pharm. Sci.* 92:585–593.
174. Soenderkaer S, Carpenter J F, van de Weert M, et al. (2004). Effects of sucrose on rFVIIa aggregation and methionine oxidation. *Euro. J. Pharm. Sci.* 21:597–606.
175. Donbrow M, Azaz E, Pillersdorf A (1978). Autooxidation of polysorbates. *J. Pharm. Sci.* 67:1676–1681.

176. Robert T (1993). Clinical properties of yeast-derived versus *Escherichia coli*-derived granulocyte-macrophage colony-stimulating factor. *Clin. Therap.* 15:19–29.
177. Schrimsher J L, Rose K, Simona G, et al. (1987). Characterization of human and mouse granulocyte-macrophage colony-stimulating factors derived from *Escherichia coli*. *Biochem. J.* 247:195–199.
178. Burgess A W, Begley C G, Johnson G R, et al. (1987). Purification and properties of bacterially synthesized human granulocyte-macrophage colony-stimulating factor. *Blood.* 69:43–51.
179. Briggs J, Panfili P R (1991). Quantitation of DNA and protein impurities in biopharmaceuticals. *Anal. Chem.* 63:850–859.
180. Berthold W, Walter J (1994). Protein purification: Aspects of processes for pharmaceutical products. *Biologicals.* 22:135–150.
181. Tan J, Leung W, Moradian-Oldak J, et al. (1998). The pH dependent amelogenin solubility and its biological significance. *Conn. Tiss. Res.* 38:215–221.
182. Piscitelli S C, Reiss W G, Figg W D, et al. (1997). Pharmacokinetic studies with recombinant cytokines. Scientific issues and practical considerations. *Clin. Pharmacokinet.* 32:368–381.
183. Metz B, Hendriksen C F M, Jiskoot W, et al. (2002). Reduction of animal use in human vaccine quality control: opportunities and problems. *Vaccine.* 20:2411–2430.
184. Reid R E (2000). *Peptide and Protein Drug Analysis*: Marcel Dekker. New York, pp. 309–859.
185. Parkins D A, Lashmar U T (2000). The formulation of biopharmaceutical products. *Pharm. Sci. Technol. Today.* 3:129–137.
186. Liao Y H, Brown M B, Quader A, et al. (2002). Protective mechanism of stabilizing excipients against dehydration in the freeze-drying of proteins. *Pharm. Res.* 19:1854–1861.
187. Bilati U, Allémann E, Doelker E (2005). Strategic approaches for overcoming peptide and protein instability within biodegradable nano- and microparticles. *Euro. J. Pharm. Biopharm.* 59:375–388.
188. Ruiz L, Reyes N, Duany L, et al. (2003). Long-term stabilization of recombinant human interferon α-2b in aqueous solution without serum albumin. *Int. J. Pharm.* 264:57–72.
189. Broadhead J, Rouan S K E, Hau I, et al. (1994). The effect of process and formulation variables on the properties of spray-dried β-galactosidase. *J. Pharm. Pharmacol.* 46:458–467.
190. Allen J D, Bentley D, Stringer R A, et al. (1999). Hybrid (BDBB) interferon-α: preformulation studies. *Int. J. Pharm.* 187:259–272.
191. Branchu S, Forbes R T, York P, et al. (1999). Hydroxypropyl-β-cyclodextrin inhibits spray-drying-induced inactivation of β-galactosidase. *J. Pharm. Sci.* 88:905–911.
192. Arakawa T, Kita Y, Carpenter J F (1991). Protein-solvent interactions in pharmaceutical formulations. *Pharm. Res.* 8:285–291.
193. Chang L Q, Shepherd D, Sun J, et al. (2005). Mechanism of protein stabilization by sugars during freeze-drying and storage: Native structure preservation, specific interaction, and/or immobilization in a glassy matrix? *J. Pharm. Sci.* 94:1427–1444.
194. Pikal M J (2004). Mechanisms of protein stabilization during freeze-drying and storage: The relative importance of thermodynamic stabilization and glassy state relaxation dynamics. In L Rey, J C May, (eds.), *Freeze-drying/Lyophilization of Pharmaceutical and Biological Products*: Marcel Dekker. New York, pp. 63–108.

195. Shah R B, Ahsan F, Khan M A (2002). Oral delivery of proteins: progress and prognostication. *Crit. Rev. Therapeut. Drug Carr. Sys.* 19:135–169.
196. Mahato R I, Narang A S, Thoma L, et al. (2003). Emerging trends in oral delivery of peptide and protein drugs. *Crit. Rev. Therapeut. Drug Carr. Sys.* 20:153–214.
197. Senel S, Kremer M, Nagy K, et al. (2001). Delivery of bioactive peptides and proteins across oral (buccal) mucosa. *Curr. Pharm. Biotechnol.* 2:175–186.
198. Johnson K A (1997). Preparation of peptide and protein powders for inhalation. *Adv. Drug Del. Rev.* 26:3–15.
199. Banga A K, Prausnitz M R (1998). Assessing the potential of skin electroporation for the delivery of protein- and gene-based drugs. *Trends Biotechnol.* 16:408–412.
200. Owens D R, Zinman B, Bolli G (2003). Alternative routes of insulin delivery. *Diabetes Med.* 20:886–898.
201. Carino G P, Mathiowitz E (1999). Oral insulin delivery. *Adv. Drug Del. Rev.* 35:249–257.
202. Mandal T K (2005). Inhaled insulin for diabetes mellitus. *Am. J. Health-Sys. Pharm.* 62:1359–1364.
203. Pedder S C (2003). Pegylation of interferon alfa: structural and pharmacokinetic properties. *Seminars in Liver Disease*. 23(suppl 1):19–22.
204. Jiang W, Cupta R K, Deshpande M C, et al. (2005). Biodegradable poly(lactic-co-glycolic acid) microparticles for injectable delivery of vaccine antigens. *Adv. Drug Del. Rev.* 57:391–410.
205. Schwendeman S P (2002). Recent advances in the stabilization of proteins encapsulated in injectable PLGA delivery systems. *Adv. Drug Del. Rev.* 19:73–98.
206. Freitas S, Merkle H P, Gander B (2005). Microencapsulation by solvent extraction/evaporation: reviewing the state of the art of microsphere preparation process technology. *J. Control. Rel.* 102:313–332.
207. Yang J, Cleland J L (1997). Factors affecting the *in vitro* release of recombinant human interferon-γ (rhIFN-γ) from PLGA microspheres. *J. Pharm. Sci.* 86:908–914.
208. Jiang G, Qiu W, DeLuca P P (2003). Preparation and *in vitro/in vivo* evaluation of insulin-loaded poly(acryloyl-hydroxyethyl starch)-PLGA composite microspheres. *Pharm. Res.* 20:452–459.
209. Peppas N A, Bures P, Leobandung W, Ichikawa H (2000). Hydrogels in pharmaceutical formulations. *Euro. J. Pharm. Biopharm.* 50:27–46.
210. Hennink W E, De Jong S J, Bos G W, et al. (2004). Biodegradable dextran hydrogels crosslinked by stereocomplex formation for the controlled release of pharmaceutical proteins. *Int. J. Pharm.* 277:99–104.
211. Velega S P, Carlfors J (2005). Supercritical fluids processing of recombinant human growth hormone. *Drug Dev. Ind. Pharm.* 31:135–149.
212. Stahl K, Claesson M, Lilliehorn P, et al. (2002). The effect of process variables on the degradation and physical properties of spray dried insulin intended for inhalation. *Int. J. Pharm.* 233:227–237.
213. Yu Z, Garcia A S, Johnston K P, et al. (2004). Spray freezing into liquid nitrogen for highly stable protein nanostructured microparticles. *Euro. J. Pharm. Biopharm.* 58:529–537.
214. Ribeiro D S I, Richard J, Thies C, et al. (2003). A supercritical fluid-based coating technology. 3: Preparation and characterization of bovine serum albumin particles coated with lipids. *J. Microencap.* 20:110–128.

215. Maa Y F, Prestrelski S J (2000). Biopharmaceutical powders: particle formulation and formulation considerations. *Curr. Pharm. Biotechnol.* 1:283–302.
216. Lechardeur D, Verkman A S, Lukacs G I (2005). Intracellular routing of plasmid DNA during non-viral gene transfer. *Adv. Drug Del. Rev.* 57:755–767.
217. Wagner E, Kircheis R, Walker G F (2004). Targeted nucleic acid delivery into tumors: new avenues for cancer therapy. *Biomed. Pharmacother.* 58:152–161.
218. Tranchant I, Thompson B, Nicolazzi C, et al. (2004). Physicochemical optimisation of plasmid delivery by cationic lipids. *J. Gene Med.* 6(suppl 1):S24–S35.
219. Pedroso de L M C, Neves S, Filipe A, et al. (2003). Cationic liposomes for gene delivery: From biophysics to biological applications. *Curr. Med. Chem.* 10:1221–1231.
220. Smyth T N (2003). Cationic liposomes as *in vivo* delivery vehicles. *Curr. Med. Chem.* 10:1279–1287.
221. Smedt S C, Demeester J, Hennink W E (2000). Cationic polymer based gene delivery systems. *Pharm. Res.* 17:113–126.
222. Prabaharan M, Mano J F (2005). Chitosan-based particles as controlled drug delivery systems. *Drug Del.* 12:41–57.
223. Lee M, Kim S W (2005). Polyethylene glycol-conjugated copolymers for plasmid DNA delivery. *Pharm. Res.* 22:1–10.
224. Yun Y H, Jiang H, Chan R, et al. (2005). Sustained release of PEG-g-chitosan complexed DNA from poly(lactide-co-glycolide). *J. Biomat. Sci., Polym. Ed.* 16:1359–1378.
225. Lambert G, Fattal E, Couvreur P (2001). Nanoparticulate systems for the delivery of antisense oligonucleotides. *Adv. Drug Del. Rev.* 47:99–112.
226. Luykx D M A M, Goerdayal S S, Dingemanse P J (2005). High-performance anion-exchange chromatography combined with intrinsic fluorescence detection to determine erythropoietin in pharmaceutical products. *J. Chromatogr. B.* 1078:113–119.
227. Runkel L, Meier W, Pepinsky R B, et al. (1998). Structural and functional differences between glycosylated and non-glycosylated forms of human interferon-beta (IFN-beta). *Pharm. Res.* 15:641–649.
228. Chan J H, Timperman A T, Qin D, et al. (1999). Microfabricated polymer devices for automated sample delivery of peptides for analysis by electrospray ionization tandem mass spectrometry. *Anal. Chem.* 71:4437–4444.
229. Diwan M, Park T G (2003). Stabilization of recombinant interferon-alpha by pegylation for encapsulation in PLGA microspheres. *Int. J. Pharm.* 252:111–122.
230. Mire-Sluis A R, Thorpe R (1998). Laboratory protocols for the quantitation of cytokines by bioassay using cytokine responsive cell lines. *J. Immunol. Meth.* 211: 199–210.
231. Rosendahl M S, Doherty D H, Smith D J, et al. (2005). A long-acting, highly potent interferon α-2 conjugate created using site-specific PEGylation. *Bioconjug. Chem.* 16:200–207.

# PROTEIN POSTTRANSLATIONAL MODIFICATION: A POTENTIAL TARGET IN PHARMACEUTICAL DEVELOPMENT

M.D. Mostaqul Huq and Li-Na Wei
*University of Minnesota Medical School, Minneapolis, Minnesota*

**Chapter Contents**

1 Introduction
2 What Is Protein Posttranslational Modification?
3 Overview of Protein Posttranslational Modification
4 Why Proteins Need to be Modified?
5 Prokaryotic versus Eukaryotic PTM
6 Studies of PTM
  6.1 Mass Spectrometry-Based PTMs
7 Mapping of PTM Sites on RIP140 by LC–ESI–MS/MS Analysis
8 The role of PTM in Cellular Function and Signaling
9 Protein Posttranslational Modification and Diseases
10 PTM as a Target for Therapeutic Development
11 Conclusion
  Acknowledgment
  References

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

## 1 INTRODUCTION

Protein posttranslational modification (PTM) is a hallmark of signal transduction and it allows the cells to rapidly respond to extracellular signals. In most cells, only a fraction of the genes are turned on at any given time, producing no more than 6000 primary proteins in a process called translation. However, several hundred types of PTMs could occur, greatly amplifying the diversity of proteins present in the cells [1]. These modifications, which involve stable and, sometimes, irreversible additions of nonamino acid chemical groups to primary translation products, occur in a large proportion of the cellular protein pool. In some instances, the types of PTMs of a protein can be predicted from its primary sequence. In many cases, however, such predictions are not possible. These modifications can result in an enormous amplification of the diversity of a protein pool. For example, glycosylation, the addition of sugar chains of varying lengths and compositions of one unmodified protein at three sites, can generate 11,520 protein variants. PTM of a protein is predicted to generate a combinatorial effect on protein structure and functions [2, 3]. The known PTMs include phosphorylation, acetylation, methylation, glycosylation, oxidation, sumoylation, ubiquitination, and nitration [4–10]. PTM of core histone proteins has been widely studied in the context of chromatin structure and gene transcription [11]. However, PTM of nonhistone proteins has only recently begun to receive attention [12–15].

This chapter will be focusing on several important issues in PTM studies such as:

- Why the proteins need to be modified and how they are modified.
- The relationship between prokaryotic and eukaryotic PTMs.
- Basic tools for studying PTM and the recent advancement in facilities and resources.
- PTM and signal transduction by nuclear receptors, co-regulators, and transcription factors.
- PTM as a target in rational design and development of therapeutic agents.

## 2 WHAT IS PROTEIN POSTTRANSLATIONAL MODIFICATION?

*Simple Definition.* The primary structure of a protein is dictated by its genetic code. All the proteins from the genome constitute the "proteome." After translation, the proteome is covalently modified by a variety of small molecules of exogenous or endogenous origin, and also by other proteins or peptides. The event is defined as the "protein posttranslational modification."

Studies of proteins present in a cell revealed a rich repertoire of expressed proteins way beyond what is expected from direct translation of the messages produced by a genome. Proteins can be modified posttranslationally by covalent attachment of one or more of several classes of molecules, by the formation of intramolecular or intermolecular linkages, by proteolytic processing of the newly synthesized polypeptide chain, or by any combination of these events. These modifications endow a protein with various properties that may be specifically required under a particular condition.

## 3 OVERVIEW OF PROTEIN POSTTRANSLATIONAL MODIFICATION

The most common PTMs include:

- *Glycosylation*—the addition of one or more sugar molecules to asparagine (N-linked) and Ser/Thr (O-linked) residues, which may involve more than one type of sugar, as shown in estrogen receptor (ER) [16] and androgen receptor (AR) [17].
- *Phosphorylation*—the addition of phosphate groups to Ser/Thr and Tyr residues, as shown in receptor interacting protein 140 (RIP140) [4], orphan testicular receptor 2 (TR2) [18], retinoic acid receptor (RAR) [19], and retinoid X receptor (RXR) [19].
- *Myristoylation and Prenylation*—the addition of certain fatty acids as shown in G-protein coupled receptor (GPCR) [20] and estrogen receptor (ER) [21].
- *Acetylation*—the addition of acetate groups to lysine residue as shown in nuclear receptors [22, 23], p53 [6], NF-κB [24], and histone proteins.
- *Methylation*—the addition of methyl groups to Arg/Lys/Gln/Asn (N-methyl) and Ser/Thr (O-methyl) as shown in p53 [25], PGC-1α [13], CBP/p300 [26], and YY1 [27].
- *Ubiquitination*—the addition of one or more ubiquitin molecules to proteins, which marks the protein for degradation as shown in p53 [28].
- Addition of a prosthetic group (e.g., heme in hemeproteins, such as hemoglobin), which is required for the protein's function).
- Addition of a certain chemical bond (i.e., a disulfide bond) between two sulfur-containing amino acids.
- Addition of a target leader sequence (a small removable peptide) at the beginning of the protein chain to allow the protein to be targeted to or from subcellular organelles (e.g., nuclei and mitochondria).

## 4 WHY THE PROTEINS NEED TO BE MODIFIED?

Proteins sometimes need to be modified to become functional. Secondly, the functional diversity of a protein can be greatly amplified by different posttranslational modifications. Modification of protein can dramatically alter its physico-chemical properties (i.e., hydrophobicity, stereochemical structure and conformation, and protein stability). Phosphorylation and glycosylation make a protein more hydrophilic, whereas acetylation, methylation, and prenylation are generally known to increase hydrophobicity of the protein. Ubiquitination typically functions as a tag for the proteins, which could be recognized by proteases (proteosomes) for degradation. Reactive oxygen species can oxidize cystein and methionine, whereas nitric oxide can react with cystein and tyrosine residues. These modifications can also modulate the function of the protein by influencing processes such as protein–protein interaction, transport, and stability.

## 5 PROKARYOTIC VERSUS EUKARYOTIC PTM

It is generally believed that proteins are not posttranslationally modified in prokaryotic cells with some exception for glycosylation of the membrane protein, which is essential for the membrane integrity and virulence property of many bacteria. However, recent investigation showed that prokaryotic proteins could also be modified, such as by phosphorylation. Some prokaryotic protein kinases, which were evolutionarily related to kinases of the eukaryotes, were discovered [29]. It was also speculated that bacterial protein could be modified by methylation, as shown by metabolic labeling using radioactive S-adenosylmethionine. Methylation was shown to enhance the chemotaxic response of the bacteria as well as the sensory transduction in *E. coli* [30]. Thus, prokaryotic proteins could also be modified to enhance their functional diversity. Our recent investigation using mass spectrometry-based proteome analyses of nuclear co-regulator RIP140 expressed from bacteria also revealed that proteins expressed in bacteria can be modified by methylation at asparagine (N) and arginine (R) (Huq and Wei, unpublished result).

It has become increasingly clear that PTM can also occur in *Archaea*. *Archaea* express proteins that enable them to strive in harsh habitats. Archaeal proteins are able to remain properly folded and functional under extremely harsh conditions such as high salinity and temperature, and other adverse physical conditions that would normally cause protein denaturation, loss of solubility, and aggregation [31].

## 6 STUDIES OF PTM

Studies of PTM have been benefited from the recent advancements in mass spectrometry, the introduction of new software and Internet-based MS data search facilities, computer-assisted topology prediction for a variety of PTMs (visit http://ca.expasy.org/), chemical synthesis of modified peptides and proteins, development of modified peptide specific antibody, *in vitro* modification techniques, exploitation of other eukaryotic cells such as insect cells for protein expression [4, 5, 16, 32], and progress in affinity purification of modified proteins.

### 6.1 Mass Spectrometry-based PTMs

Mass spectrometry has revolutionized the idetinification of PTM. In the past, the metabolic radio-labeling technique was the principal methodology used to determine whether the protein is modified. However, identification of a specific residue that is modified by PTM using metabolic labeling relies on point mutation of the protein, which is practically time- and effort-consuming and sometimes unreliable because of existance of other similar sites of PTM [33]. The mass spectrometric analysis can identify all kinds of modification on a protein in a single experiment. However, sometimes it can be a laborious and lengthy process because of the need to purify the protein. Data analysis itself can also be time-consuming. We will use several MS-based PTM analyses in our own experiences as examples just to illustrate the design and execute the entire experiment.

- *Sample.* The type of protein (i.e., membarne, cytosolic or neuclear protein) the pI value, the molecular weight, the solubility and stability. The amino acid sequence of the protein needs to be considered, which is necessary for the selection of proper proteases for proteolytic digestion.
- *Sample Preparation.* Sample preparation is the most critical step in mass analysis. The prurifacation procedure and the presence of potentially interfering substance for MS should be evaluated. For instance, the type and the concentartion of detergent used for solubilizing the protein may not be compatibe with mass spectrometry.
- *Type of Modifications.* The specific chemistry of the type of modification and its stability during the procedure.
- *Sites of Modification (residues).* The residues (amino acid) where the modification could occur, such as Arg/Lys for methylation, Ser/Thr or Tyr for phosphorylation, Tyr or Cys for nitration. The sites of modification is also important because it may interefer with protease digestion of the protein. Selection of the protease sometimes depends on the location of the sites of modification.
- *Mass Spectrometry Facilities.* The facilites of mass spectrometry (i.e., MALDI–TOF, MALDI–TOF–MS/MS, LC–MS, LC–ESI–MS/MS) and other resources for data search and management. In most cases, each machine is designed for a particular purpose.
- *Knowledge and Training.* It is crucial to acquire adequate working knowledge about mass spectrometry in addition to understanding the principle. It is also important to receive hands-on training in the operation of MS database search and manual data interpretation skills to verify the PTM and PTM sites, which is particularly critical to accurately interpret the data and to avoid any false-positive or fasle-negative result.

  These points are presented in Scheme 1.

A practical, step-by-step guideline for successful studies of PTM is provided based on our own experience in studying PTMs of several transcription regulatory proteins including RIP140, RAR, RXR, TR2, TR4, and NF-κB.

*Step 1.* Determine the types of modification to be examined. It would be useful to gather information on whether this type of modification can be enriched. For example, enzymatic induction, *in vitro* chemical reaction, and expression of protein in suitable eukaryotic (i.e., insect cells) or prokaryotic systems. Some commercial products for enrichment of modified peptides such as phosphopeptides are currently available (visit Ionsource website at http://www.ionsource.com/ for more information about this option).

*Step 2.* Purify the protein in a system in which the modification is stable. Avoid any kind of external contamination during purification, for example, keratin, which is the most common impurity present in MS analysis. If purification is not possible, it can be enriched, separated, and recovered from an SDS–PAGE. The amount of sample is also important. A very sensitive machine can identify the protein at the famtomole level.

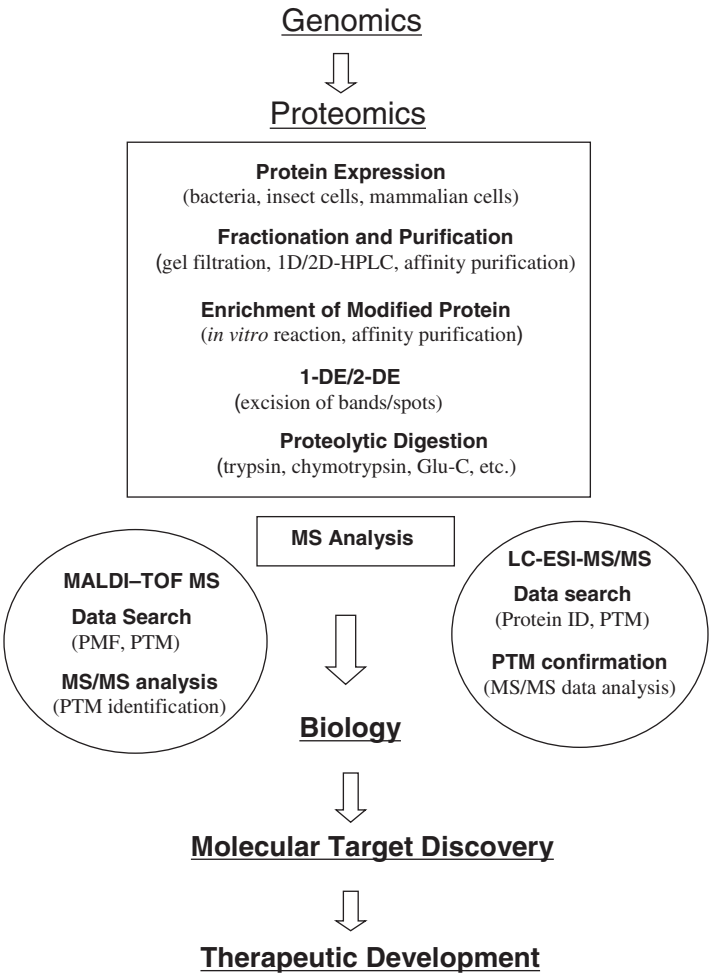


**Scheme 1.** Strategies and goal of proteome analysis.

*Step 3.* Carry out theoretical digestion of the protein to select proteases (i.e., trypsin, chymotrypsin, and Glu-C) using software available online (http://ca.expasy.org/ and MS–DIGEST at http://prospector.ucsf.edu/). Analyze the theoretical mass of each peptide and select the proper protease for digestion that will give you the maximum coverage of the protein with the molecular weight of each peptide no more than 3000 Da. However, most of the researchers select trypsin initially.

*Step 4.* Run a 1-DE/2-DE gel and stain the gel by commassie blue. Silver-stained gels need extra steps in destaining. All operation should be carried out in a dust-free envrionment to avoid contaminants like keratin. Staining or destaining solution and PAGE running buffer should not be recycled. Excise the protein band of interest and conduct in gel digestion. In-solution tryptic digestion can also be done. However, most investigators prefer in-gel digestion. Several protocols for in-gel digestion on commassie-stained and silver-stained gels are available online at the

University of Minnesota proteomics core facilities website (http://www.cbs.umn.edu/msp/).

*Step 5.* Desalt the tryptic peptides using ZipTip (Millipore). The ZipTip protocol is also available online (http://www.cbs.umn.edu/msp/) and spotted on a MALDI target over the crystal of matrix molecule (i.e., α-cyano-4-hydroxycinnamic acid). The full-scan spectrum is recorded and the peak lists are generated by software. The experimental mass value (monoisotopic) from the peak lists could be copied and pasted to the Mascot peptide mass fingerprint data search (http://www.matrixscience.com/) or MS-fit data serach (http://prospector.ucsf.edu/) for peptide mass fingerprinting (PMF) to identify the protein of interest or verify the protein identity. The search result will show the identified protein according to the match of mass of a number of peptides, their scores, and the sequence coverage of the total protein. An overview about PMF is described below, which is also available at Ionsource website as tutorials (http://www.ionsource.com/). When the protein is properly identified, the investigator can look at the original mass spectrum and compare with the mass data of theoretical digestion and subtract the peak manually from the full-scan MALDI spectrum one by one. The remaing peaks in the spectrum could originate from the PTM-modified trypsin autolysis peaks or contaminants. If the spectrum is of high quality, the investigator can analyze each individual peak and the neighboring peaks to determine the mass shift between the neighboring peaks, providing some insights into whether the protein is modified. An important consideration is that the modification should be stable or suitable for ionization by MALDI. If the mass shift between the two neighboring peaks give a mass shift equal to the mass shift caused by modification (i.e., 80 amu for phosphorylation and 42 amu for acetylation), that particular ion peak is likely to be originated from the modified peptide. To confirm this modification, the MS/MS spectrum of that parent ion can be recorded from the same sample spotted on a different target.

***Peptide Mass Fingerprinting.*** PMF is an important tool to identify proteins by matching their constituent fragment ions (peptide masses) to the theoretical peptide masses generated from a protein or DNA database. The conceptual basis of PMF is that every unique protein generates a unique set of peptides and hence unique peptide masses. An intact unknown protein is first cleaved with a proteolytic enzyme to generate peptides. A PMF database search is conducted following MALDI–TOF mass analysis of the protein sample. Identification is accomplished by matching the observed peptide masses to the theoretical masses derived from a sequence database. PMF identification relies on observing a large number of peptides (5 or more) from the same protein at high mass accuracy. This technique is particularly suited for 1D/2D gel spots where the protein purity is high. However, it is less efficient for complex mixtures of proteins.

Although this technique was introduced in early 1990s, it was the introduction of a MALDI–TOF instrument capable of 50 ppm mass accuracy that made PMF a routine procedure. In MALDI–TOF mass spectrometry, peptides appear as singly charged species in the mass spectrum (Figure 1). Unlike an electrospray (ESI) mass spectrum, which displays multiply charged species, the MALDI–TOF spectrum is simple to interpret. PMF can also be used to identify proteins in ESI spectra, but it is seldom used because the peptide masses would need to be deconvoluted for each search.

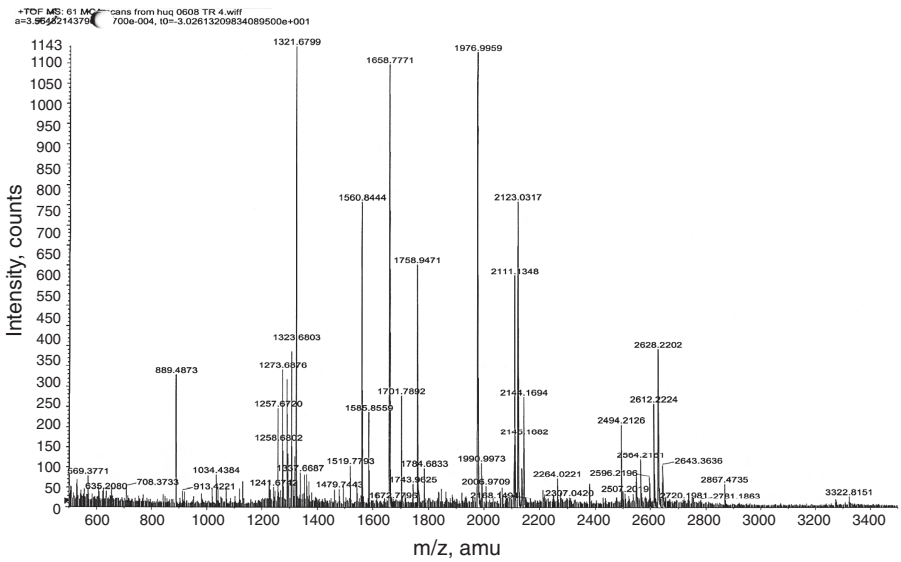


**Figure 1. MALDI–TOF mass spectrum of orphan nuclear receptor TR4.** This was collected on a QSTAR-XL MALDI–TOF mass spectrometer and was an average of 240 scans; peptide peaks appear as $[M+H]^{1+}$ ions.

*LC–ESI–MS/MS analysis.* To examine the potential PTM of a protein after MADLI–TOF mass, a liquid chromatography-tandom mass spectrometry (LC–ESI–MS/MS) is run to obtain information-dependent acquisition (IDA) data. Alternatively, an LC–ESI–MS/MS analysis can be performed without prior MALDI–TOF. However, it is the most common practice to first conduct a MALDI–TOF mass to determine the amount of sample to be loaded on a subsequent LC–ESI–MS/MS. The IDA data could then be searched at Mascot MS/MS data search (available online at http://www.matrixscience.com/) by selecting several parameters, such as protein data bank (i.e., NCBI and MSDB), the species, the fixed modification (i.e., carbamidomethyl cystein), the variable modifications, and the mass tolerance. The search result will generate information about the protein. The reconstructed MS/MS spectrum along with the calculated mass value of product ions matched in the original MS/MS spectrum can be obtained, which will show the residue modified by PTM. To confirm the modification, manual analysis of the MS/MS data is essential.

***MS/MS Fragmentation and Nomenclature.*** The fragmentation pattern and the nomenclature commonly used in MS/MS studies are discussed below. This information was taken from Mascot website (http://www.matrixscience.com/).

*Sequence Ions.* The types of fragment ions observed in an MS/MS spectrum depend on many factors including primary sequence, the amount of internal energy, how the energy was introduced, and the charge state. The MS/MS fragmentation pattern is shown in Chart 1.

Fragments will only be detected if they carry at least one charge. If this charge is carried by the N-terminal fragment, the ion is called either *a*, *b*, or *c*. If the charge

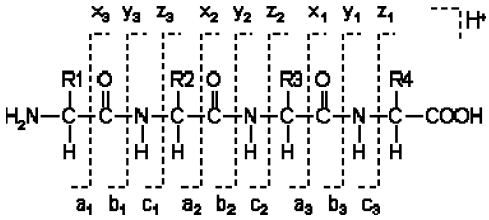


**Chart 1.**

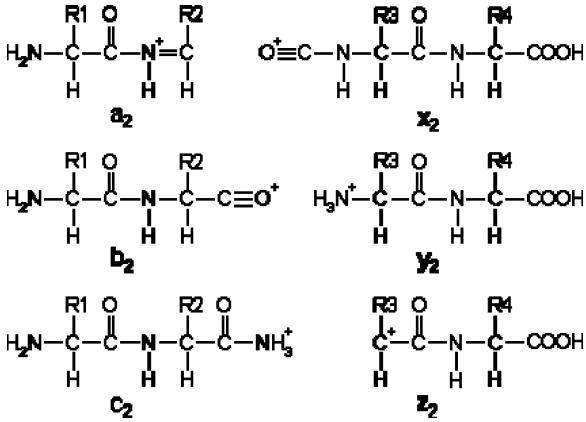


**Chart 2.**

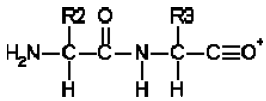

**Chart 3.**

is retained on the C-terminal fragment, the ion is classified as either *x*, *y*, or *z*. A subscript indicates the number of residues in the fragment. In addition to the proton(s) carrying the charge, *c* ions and *y* ions abstract an additional proton from the precursor peptide. The structures of the six singly charged sequence ions are shown in Chart 2.

The above structures show only a single charge carrying proton. However, in electrospray ionization, tryptic peptides generally carry two or more charges, so fragment ions may carry more than one proton.

*Internal Cleavage Ions.* Double-backbone cleavage gives rise to internal fragments (Chart 3), which are usually formed by a combination of *b*-type and *y*-type cleavage to produce the illustrated structure, an amino-acylium ion. Sometimes, internal cleavage ions can form by a combination of *a*-type and *y*-type cleavage, an amino-immonium ion.

*Immonium Ions.* An internal fragment with a single side chain formed by a combination of *a*-type and *y*-type cleavage is called an immonium ion (Chart 4).

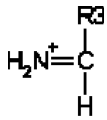

**Chart 4.**

***Protein Identification with MS/MS Data.*** The previously described PMF uses the intact masses of the peptides to fit a protein in a sequence database. The MS/MS spectral matching employs the uninterpreted peaks in a peptide fragment spectrum to match to a theoretical fragment spectrum in a sequence database. This MS/MS spectrum is acquired from a collision-induced dissociation (CID) within a mass spectrometer. The fragmentations are produced either in a collision cell in a tandem mass spectrometer or within an ion trap. In an MS/MS-based protein ID experiment, multiple peptides are usually found and all of their fragment spectrums are used to identify the protein. In both PMF and MS/MS ID, the larger the number of peptides identified, the greater the confidence in the protein correlation. The correlation made with multiple MS/MS spectra are usually superior to the identification made in a PMF experiment. It is impossible to rely on a single peptide mass in a PMF experiment to correlate a protein. However, the heterogeneity imparted to a single peptide in an MS/MS experiment can generate enough amino acid sequence to correlate a protein. Tutorials from Ionsource website (http://www.ionsource.com/) for a step-by-step simplification of how sequence database search programs accomplish peptide and protein ID are provided in the following.

### *A Step-by-Step General Scheme for MS/MS Protein ID*

*Step 1. Enzyme specificity constraint:* Some programs preindex the sequence database based on the enzyme specificity, which facilitates the search much faster because tryptic peptides can be indexed and mass lists can be premade. The downside is the necessity of a separate database to be indexed for each enzyme or each time a potential modification is changed.

*Step 2. Peptide matching:* It involves matching the parent mass of the intact peptide to the peptides in the database. Generally, the narrower parent mass constraint, the faster the search will proceed because fewer peptides will need to be correlated in the next step.

*Step 3. Peptide listing and comparison:* It takes the list of peptides identified by parent mass in step two and compares the theoretical fragment masses of these peptides to the experimentally derived fragment spectra.

*Step 4. Peptide ranking:* Hits are ranked by how many of the fragment masses match the theoretical fragment masses in the sequence database.

*Step 5. Significance of hit*: If more than one peptide is searched, all the peptides found are correlated to their prospective proteins. The protein with the greatest number of well-correlated peptides is usually the most significant hit.

*Step 6. Probability factor and validation:* Some programs show the probability number to back-up the proposed match, which gives the users some comfort in the designation. Select the most important individual protein hit and manually validate the spectral match.

*In the following, we describe data management of PTM analysis of RIP140, where we detected 11 phosphorylation and eight acetylation sites.*

## 7 MAPPING OF PTM SITES ON RIP140 BY LC–ESI–MS/MS ANALYSIS

*Expression and Purification of RIP140 from Sf21 Insect Cells.* Sf21 insect cells ($1 \times 10^6$) were infected with recombinant Baculovirus vector. Using affinity chromatography under a denaturing condition, we were able to purify the recombinant protein to over 95% homogeneity [4]. The eluted protein from the affinity column was further resolved by a SDS–PAGE and a distinctly separated single band of RIP140 was obtained. The gel bands were excised for trypsin digestion and the tryptic peptides were analyzed by MALDI–TOF and LC–ESI–MS/MS.

*SDS-PAGE, In-gel Tryptic Digestion of RIP 140.* The purified His6-RIP140 protein (500 ng) from insect cells and GST-RIP140 (1 mg) from *E. coli* were resolved by an 8% SDS–PAGE. The bands were visualized by commassie staining. Gel slices containing RIP140 were subjected to overnight in-gel tryptic digestion The tryptic peptides were extracted from the gel with 5% acetic acid, followed by 5% acetic acid in 50% ACN. The digests and the extracted solution were combined and dried. The sample was re-dissolved in 5% acetic acid and desalted using ZipTip C18 reverse-phase desalting Eppendorf tips (Millipore). The peptides were eluted with 2% ACN containing 0.1% TFA to a volume of 50 μL.

*Mass Spectrometric Analysis of RIP140.* For identification of the proteins, the samples were analyzed by MALDI–TOF MS in a positive ion reflection mode (Qstar XL, Applied Biosystems) using α-cyano-4-hydroxycinnamic acid as a matrix. For LC-MS, the tryptic peptides were subjected to a 20-fold dilution using 2% acetonitrile in water containing 0.1% trifluoroacetic acid. An LC packings (LCP, a Dionex Company, Sunnyvale, CA) Famos autosampler aspirated 27.5 μL of the peptide solution into a 100-μL sample loop using the Famos μl-pick-up injection mode and 98 : 2 water : ACN, 0.1% formic acid (load buffer) as the transfer reagent. An LCP Switchos pump was used to concentrate and desalt the sample on an LCP C18 nano-precolumn (0.3 mm internal diameter × 5 mm length). The precolumn was switched in-line with a capillary column and peptides were eluted at 350 nL/min using an LCP Ultimate LC system. The capillary column (100 mm internal diameter) was packed in-house to 12 cm length with 5-μm, 200-Å pore size C18 particles. Peptides were eluted with a linear gradient of solvent B (5 : 95 water : ACN, 0.1% formic acid) from 0% (100% solvent A: 95 : 5, water : ACN, 0.1% formic acid) to 40% over 60 minutes and 40% to 90% for 5 minutes. The LC system was online with an Applied Biosystems, Inc. (ABI, Foster City, CA) QSTAR Pulsar quadrupole time-of-flight (TOF) mass spectrometer (MS), which was equipped with Protana's nano-electrospray source. An electrospray voltage of 2250 V was applied distal to the

analytical column. The TOF region acceleration voltage was 4 kV and the injection pulse repetition rate was 6.0 kHz. The $[M + 3H]^{3+}$ monoisotopic peak at 586.9830 *m/z* and $[M + 2H]^{2+}$ monoisotopic peak at 879.9705 *m/z* from human renin substrate tetradecapeptide (Sigma-Alrdich, St. Louis, MO) were used for external calibration. As peptides eluted from the column, they were focused into the MS. The IDA was used to acquire MS/MS data with experiments designed such that the three most abundant peptides were subjected to collision-induced dissociation, using argon as the collision gas every 15 seconds. Collision energies were varied as a function of the *m/z* and the charge state of each peptide. To avoid continued MS/MS of peptides that had already undergone collision-induced dissociation, a dynamic exclusion was incorporated for a further 45 seconds. IDA mode settings included continuous cycles of three full-scan TOF MS of 400–550 *m/z*, 550–750 *m/z*, and 750–1200 *m/z* (1.5 seconds) plus three product ion scans of 50–4000 *m/z* (3 seconds each). Precursor *m/z* values were selected from a peak list automatically generated by Analyst QS software (ABI) from the TOF MS scans during acquisition, starting with the most intense ion. Peptide mass software MS–Digest from the ProteinProspector (available online http://prospector.ucsf.edu) was used to generate a theoretical tryptic digest of RIP140 by considering serine-, threonine-, and tyrosine-containing peptides to account for phosphorylation and lysine for acetylation. From the LCMS data, the molecular weights of the detected peptides were calculated using the LCMS reconstruct feature of Analyst QS. Experimentally measured peptide masses were compared with the theoretical digest. To confirm the sequence of the peptides and the sites of modification, MS/MS spectra were examined. Peaks with a minimum height of 3% relative to the base peak were considered and a 100-ppm tolerance was used to establish matches with the theoretical b and y ions that were predicted using Bioanalyst software (Applied Biosystems). In addition, in all cases, the data from IDA experiments were searched using MASCOT (http://www.matrixscience.com) MS/MS data search.

*Mapping of Phosphorylation and Acetylation Sites on RIP140.* The tryptic digested samples were first analyzed by MALDI–TOF mass to identify the RIP140 and posttranslational modification by phosphorylation at serine, threonine, and tyrosine residues and acetylation sites on lysine residues. The MALDI–TOF MS data were subjected to a MASCOT search at the NCBI data bank. The mass tolerance of both precursor ions was set at 175 ppm, carbamidomethyl cystein was specified as a static modification, and phosphorylated serine, threonine, and tyrosine and acetylation on lysine residues were specified as variable modifications. The search result confirmed the identification of the protein with significant sequence coverage to over 60% of the total protein [4]. However, only a few phosphopeptides were predicted from the MALDI–TOF mass data. The MS/MS analysis was carried out for the precursor ions of the predicted phosphopeptides and acetylated peptides to identify the peptide sequence and phosphorylation/acetylation sites. Unfortunately, in most cases, high-quality MS/MS spectra for the identification of the phosphorylation sites by MALDI–TOF mass were not obtained. However, sequences of the unmodified peptides were identified properly from the MS/MS analysis. Based on this information, we speculated that the poor quality of MS/MS spectra of the phosphopeptides/acetylated peptides could be caused by improper ionization of the RIP140 tryptic peptides by MALDI. Therefore, LC–ESI–MS/MS was applied

using nanoelectrospray source for ionization to solve this problem. We recorded three independent full-scan ion chromatograms of 400–550 *m/z*, 550–750 *m/z*, and 750–1200 *m/z*. The IDA was used to acquire MS/MS data. IDA analyses were performed on the tryptic digests of RIP140 expressed in *E. coli* and insect cells. The data from IDA experiments were searched using a MASCOT search. The mass tolerance of both precursor ion and the MS/MS fragment ions was set at ±0.1 Da and carbamidomethyl cystein was specified as a static modification. Phosphorylated serine, threonine, and tyrosine and acetyaled lysine were specified as variable modifications. The result revealed 11 tryptic phosphopeptides (Table 1) and eight acetylated peptides (Table 2) from RIP140 expressed in insect cells. For a control, no phosphorylated/acetylated peptides were detected from RIP140 expressed in *E. coli* [4, 5, 34].

*Mapping of Phosphorylation Sites.* The MS/MS data of the phosphopeptides in comparison with the unmodified form were analyzed manually to map the phosphorylation sites on RIP 140. As an example, the MS/MS spectrum of tryptic peptide spanning residues (100–111) is presented to explain how phosphorylation was assigned (Figure 2). The MS/MS fragmentation patterns for both the precursor ions of the unmodified (top panel) and modified peptide (bottom panel) residues 100–111 were identical, especially from *b*1 to *b*4 ions and *y*1 to *y*4 ions. The *b*5 ion of the unmodified form appeared as a singly charged ion at 559.26 *m/z*, whereas the *b*5 ion in the phosphopeptide appeared at 541.26 *m/z* because of a –18 amu delta-mass shift. This delta-mass shift was considered to be the β-eliminated product of *b*5 ion caused by loss of either $H_3PO_4$ or $H_2O$ from Ser-104. Similar β-elimination was also shown at *b*6 ion of the modified peptide. The *b*8 ion of the modified peptide showed an 80-amu ($H_3PO_4$) delta-mass shift (965.39 *m/z*) along with its β-eliminated ion signal at 867.41 *m/z* due to loss of $H_3PO_4$ (98 amu). These data asserted that the β-eliminated *b*5 and *b*6 ions were from the loss of $H_3PO_4$ from Ser-104. Thus, the phosphorylation site was assigned to Ser-104. This phosphorylation site was also confirmed by an independent MS/MS analysis (data not shown) of the phosphopeptide residues 101–111 and 101–112 (Table 1).

*Mapping of Acetylation Sites on RIP140.* The MS/MS data of the acetylated peptides were also analyzed manually to identify the acetylation sites on RIP140. To identify the acetylated peptides, a 42-amu mass shift was considered as an indication of acetylation. To distinguish acetylation from trimethylation, both rendering the same integral mass shift, some marker ion signals, such as an immonium ion at 126 *m/z* specific for acetylated lysine was monitored [35, 36]. In addition, the absence of a marker ion for trimethylated lysine generated from loss of trimethylamine from b or y ions as $[M\text{-}59]^+$ was also confirmed. Significant differences in mass shift were caused by loss of acetyl moiety (42 *m/z*) and trimethylamine (59 *m/z*). Therefore, a 42-amu net loss for acetyl group from y or b ions was also accounted to assign the acetylation sites [33, 34]. Initial analysis of the mass data showed some peptides displaying a 43-unit mass shift instead of 42 units for acetylation, which raised a possibility of carbamylation of lysine residue rather than acetylation in the modified peptides. But careful analysis of MS/MS data finally confirmed that this 1-unit positive delta mass shift was originated from the deamidation of either glutamine or asparagine residue present in the acetylated peptides. All of these

**TABLE 1. LC–ESI–MS Profile of the Tryptic Phosphopeptides of RIP140 [4]**

| Residues | Sequence | Unmodified | | | Modified | | |
|---|---|---|---|---|---|---|---|
| | | *m/z* (z) | $M^{+2}$ | RT[1] | *m/z* (z) | $M^{+2}$ | RT[1] |
| 100–111 | RLSD**S**IVNLNVK | 453.26(3) | 1356.76 | 40.24 | 719.36(2) | 1436.72 | 41.08 |
| | | 679.37(2) | | | | | |
| 101–111 | LSD**S**IVNNLNVK | 601.33(2) | 1200.66 | 41.74 | 641.31(3) | 1280.63 | 42.87 |
| 101–112 | LSD**S**IVNLNVKK | 443.92(3) | 1328.75 | 37.86 | 705.35(2) | 1408.72 | 39.45 |
| | | 665.36(2) | | | | | |
| 199–212 | SGPTLPDV**T**PNLIR | 740.40(2) | 1478.80 | 47.06 | 780.38(2) | 1558.76 | 47.36 |
| 343–360 | NNAATFQSPMGVVPS**S**PK | 916.45(2) | 1830.90 | 41.44 | 956.43(2) | 1910.86 | 42.03 |
| 375–387 | QAANN**S**LLLHLLK | 478.94(3) | 1433.82 | 49.17 | 505.60(3) | 1513.80 | 53.03 |
| | | 717.90(2) | | 49.05 | 757.90(2) | | 52.90 |
| 476–492 | IPGVDIKEDQDT**S**TNSK | 923.94(2) | 1845.87 | 34.61 | 963.94(2) | 1925.88 | 35.20 |
| | | 616.30(3) | | | | | |
| 517–537 | NA**S**PQDIHSDGTKF**S**PQNYTR | 591.53(4) | 2362.12 | 35.31 | 815.02(2) | 2442.03 | 35.91 |
| | | | | | 841.68(2) | 2522.02 | 36.55 |
| 538–548 | TSVIE**S**PSTNR | 595.80(2) | 1189.59 | 30.83 | 635.78(2) | 1269.56 | 30.75 |
| 670–678 | LN**S**PLLSNK | 493.28(2) | 984.55 | 35.92 | 533.25(2) | 1064.52 | 33.89 |
| 1001–09 | TF**S**YPGMVK | 515.24(2) | 1028.49 | 39.21 | 555.24(2) | 1108.46 | 41.23 |

[1] RT, retention time in minute.
[2] $M^{+}$, precursor mass.

**TABLE 2. LC–ESI–MS Profile of the ACTEYLATED Tryptic Peptides of RIP140 [5]**

Tryptic digests of RIP140 protein was subjected to LC–ESI–MS/MS. Three independent full-scan ion chromatograms from *m/z* 400–550, 550–750, and 750–1200 were recorded in an information-dependent acquisition (IDA) mode to acquire MS/MS data. The IDA data were searched online at MASCOT (http://www.matrixscience.com) MS/MS data search at the NCBI data bank. The MS/MS data were analyzed manually to confirm the sequence of the modified and the unmodified forms of the same peptide identified by the data bank search. The full scan chromatograms were analyzed to assign the charged state, retention time, and intensities of the peptides.

| Residue | Sequence[1] | Unmodified | Modified |
|---|---|---|---|
| | | [*m/z* (z)], $M^{+2}$, RT[3] (min) | *m/z* (z), $M^{+}$, RT(min) |
| 101–112 | LSDSINNLNV**K**K | [443.92 (3)], 1328.75, 38.09 | [686.88 (2)], 11371.75, 41.37 |
| 155–170 | QSL**K**EQGYALSHESLK | [606.65 (3)], 1816.93, 35.80 | [620.98 (3)], 1859.91, 40.67 |
| 283–298 | EHAL**K**TQNAHQVASER | Not detected | [621.31 (3)], 1860.91, 23.80 |
| 305–320 | LQENGQ**K**DVGSSQLSK | [859.92 (2)], 1717.85, 29.34 | [880.93(2)], 1759.84, 32.56 |
| 476–492 | IPGVDI**K**EDQDTSTNSK | [923.95 (2)], 1845.90, 34.61 | [945.43 (2)], 1888.88, 36.96 |
| 517–537 | NASPQDIHSDGT**K**FSPQNYTR | [788.36 (3)], 2362.09, 35.35 | [802.69 (3)], 2405.11, 37.65 |
| 606–630 | G**K**ESQAEKPAPSEGAQNSATFSK | [836.40 (3)], 2506.17, 30.82 | [850.74 (3)], 2459.19, 32.38 |
| 930–938 | ES**K**SFNVLK | [526.27 (2)], 1050.58, 36.99 | [547.79 (2)], 1093.56, 38.22 |

[1] The 43-unit mass difference between modified and unmodified peptides appeared because of acetylation of lysine along with deamidation of either asparagine (N) or glutamine (Q) residues present in the peptides.
[2] RT, retention time in minutes.
[3] $M^{+}$, precursor ion.

parameters were evaluated carefully while MS/MS data were analyzed to sequence the modified peptides. In addition, the MS/MS spectra of the modified peptide were always compared with the unmodified peptide (data not shown) for fingerprinting purposes.

We provide an example of how an acetylation site can be identified by manual analysis of MS/MS spectrum. The acetylated peptide-spanning residues 101–112 displayed a doubly charged ion at 686.88 *m/z* (precursor 1371.75 *m/z*) in total ion chromatogram (TIC). The precursor ion showed a 43-unit mass shift caused by acetylation of a lysine and deamidation from an asparagine residue. The MS/MS spectrum of the modified peptide showed *y*1 ion at 147.1 *m/z* and *y*1° ion at 129.1 *m/z* caused by loss of $H_2O$, which were identical to the unmodified peptide (Figure 3). The *y*2 ion appeared at 317.2 *m/z* instead of 275.2 *m/z*, resulting in a net

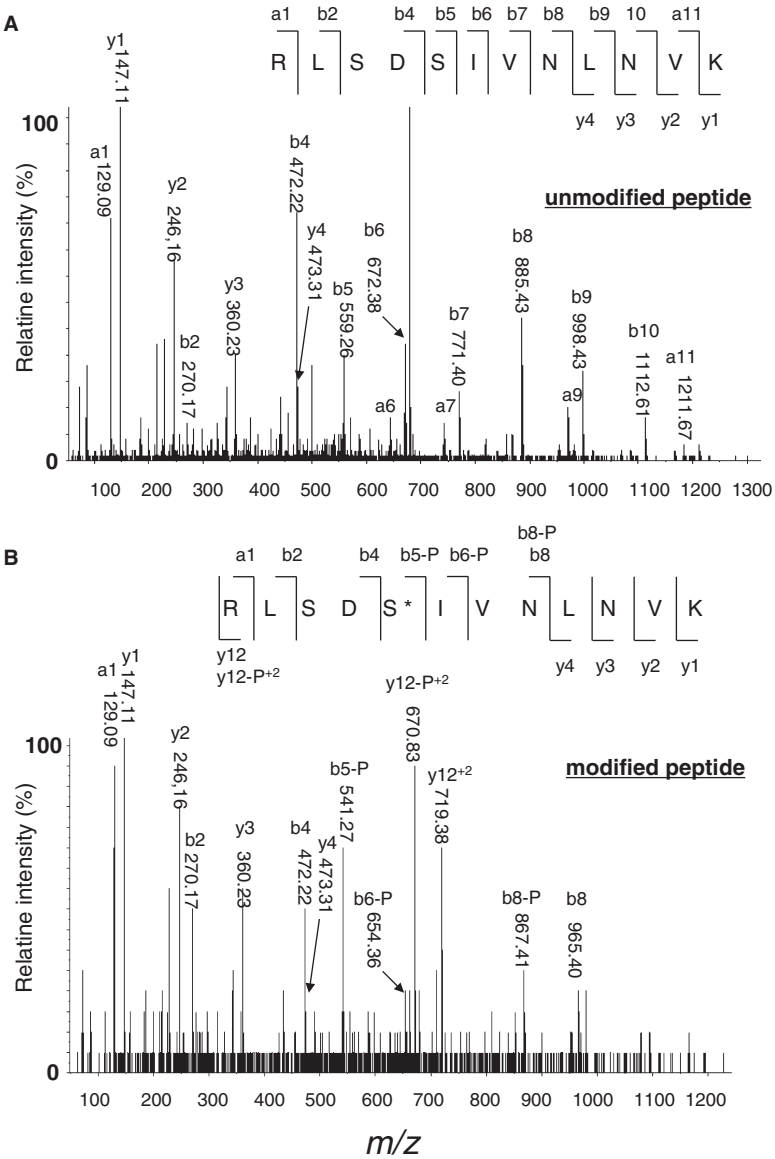


**Figure 2. CID–MS/MS spectra peptide residues 100–111 form RIP140.** Both spectra (unmodified, top and phosphorylated, bottom) showed identical b4 ions at *m/z* 472.22, but the b8 and b5 ions of the phosphopeptide showed 80 and –98 units mass shift, respectively, suggesting Ser-104 phosphorylation. (Courtesy: *Proteomics*, 5:2157–2166.)

mass shift of 42 Da, which confirmed the modification of Lys-111 by acetylation. The presence of a marker ion at 126 *m/z* specific for lysine acetylation confirmed this modification originated from acetylation rather than trimethylation [35, 36], which was further substantiated by the presence of consecutive *y* ions from *y*2 to *y*8 due to modified lysine residue. The *y* ions starting from *y*4 showed a 43-amu shift instead of 42, suggesting deamidation of Asn-109.

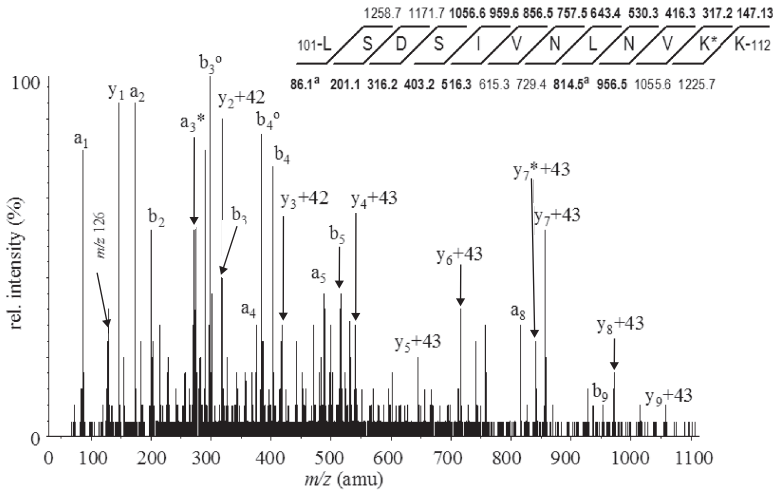


**Figure 3. CID–MS/MS spectra of acetylated peptides spanning residues 101–112 from RIP140.** The y1 ion at *m/z* 147.1 was identical to the native peptide, but the y2 and y3 ions showed a 42-unit mass shift, suggesting Lys-111 acetylation. (Courtesy: *Molecular & Cellular Proteomics,* 4:975–983.)

## 8 THE ROLE OF PTM IN CELLULAR FUNCTION AND CELL SIGNALING

*Phosphorylation.* Phosphorlation is the most well-studied PTM, and it occurs on a large number of proteins. We examined PTM of RIP140, which primarily functions as a lignad-dependent co-repressor for many nuclear receptors by recruiting HDACs in the transcription complex [34]. RIP140-null mice exhibited reproductive defects as well as abnormal energy homeostasis. We identified 11 phosphorylated sites including $Ser^{104}$, $Thr^{202}$, $Thr^{207}$, $Ser^{358}$, $Ser^{380}$, $Ser^{488}$, $Ser^{519}$, $Ser^{531}$, $Ser^{543}$, $Ser^{672}$, and $Ser^{1003}$ [4, 34]. The MAP kinase-mediated phosphorylation of $Thr^{202}$, $Thr^{207}$, and $Ser^{358}$ enhances the repressive potential of RIP140 by facilitating the recruitment of HDAC3. We generated constitutive dephosphorylated mutant by replacing the phosphorylated Ser/Thr with Ala and a constitutive positive mutant by replacing phospho-Ser/Thr with glutamic acid. The Ala-null mutant was unable to exert the repressive activity, whereas the constitutive positive mutant became more repressive [34], which demonstrates that phosphorylation of RIP140 critically modulates its biological activity.

In another study, we asked whether and how the orphan nuclear receptor 2 (TR2) can activate its target genes by phosphorylation without ligands [18]. We hypothesized that PTM of TR2 can modulate TR2 function without a putative ligand. We showed that, indeed, PKC-mediated phosphorylation of TR2 enhanced its stability and activation potential for its target RARβ2. We identified PKC-mediated phosphorylation on the ligand-binding domain of TR2 at Ser-461 and Ser-568, which enhanced its protein stability. We further identified phosphorylation of the DNA-binding domain (DBD) of TR2 at Ser-170 and Ser-185, which facilitated its DNA-binding ability and recruitment of coactivator p300/CBP-associated factor (P/CAF) [39]. Ser-185 is required for DNA binding, whereas both Ser-170

and Ser-185 are necessary for receptor interaction with P/CAF [18]. For a comprehensive review of phosphorylation of nuclear receptors, see the review by Rochette-Egly [14].

*Acetylation.* Acetylation is another widely known PTM. Many nuclear receptors (NRs) and transcription factors are known be modified by acetylation. We have also shown that RIP140 can be acetylated on eight lysine residues, including $Lys^{111}$, $Lys^{158}$, $Lys^{287}$, $Lys^{311}$, $Lys^{482}$, $Lys^{529}$, $Lys^{607}$, and $Lys^{932}$. The amino-terminal region [amino acids (aa) 1–495] was more repressive and accumulated more in the nuclei under a hyperacetylated condition, whereas hyperacetylation reduced the repressive activity and nuclear translocation of the central region (aa 336–1006). Hyperacetylation also enhanced the repressive activity of the full-length protein and triggered its export into the cytosol. This study revealed differential effects of PTM on various domains of RIP140. Recent reviews described nuclear receptors and transcription factors that can be modified by acetylation [22, 40].

*Tyrosine Nitration.* Tyrosine nitration is one of the protein modifications induced by reactive oxygen species. NO is an important factor that induces posttranslational modifications of proteins by cellular reduction and oxidation mechanism: cysteinyl-nitrosylation or Tyr nitration. Nuclear factor (NF)-kappaB activity can be rapidly suppressed by sodium nitroprusside, an NO donor. This effect was effectively reversed by peroxynitrite scavenger deferoxamine, suggesting a Tyr nitration-mediated mechanism. Tyr nitration of p65 induced its dissociation from p50, its association with IκBα, and subsequent sequestration of p65 in the cytoplasm by IkBa-mediated export. LCMS revealed specific nitration on Tyr-66 and Tyr-152 residues of p65, which were confirmed by mutation studies. These residues are important for the direct effects of NO on p65, which resulted in more p65 export and inactivation of NF-κB activity. This study identified a novel and efficient means in which NO rapidly inactivated NF-κB activity by inducing Tyr nitration on p65 [12].

## 9 PROTEIN POSTTRANSLATIONAL MODIFICATION AND DISEASES

The establishment of species- and tissue-specific protein databases provides a foundation for proteomics studies of diseases. Continual development will lead to functional proteomics studies, in which identification of protein modification in conjunction with functional data from established biochemical and physiological methods enables the examination of interplay between changes in a proteome and the progression of diseases. Recently, many investigations provided direct evidence for PTM in the pathophysiological progression of many diseases like diabetes, Alzheimer's diseases (AD), atherosclerosis, and oncogenesis.

*Cardiovascular Diseases.* In a comprehensive cardiovascular proteome analysis, various PTMs were documented in dilated cardiomyopathy [41]. A 2-DE analysis of dilated cardiomyopathy-diseased human myocardial tissue revealed more than 50 HSP27 proteins by iummunoblotting, illustrating a large number of PTMs potentially occur on a single protein [42].

*Neurodegenerative Diseases.* Mircotubule-associasted protein *tau* is documented to undergo several PTMs and aggregates into paired helical filaments (PHF) in AD and other taupathies [43]. PTMs of *tau* include hyperphosphorylation, glycosylation, ubiquitination, glycation, polyamination, nitration, and proteolysis. Hyperphosphorylation and glycosylation are crucial to the molecular pathogenesis of neurofibrillary degeneration of AD [43]. The others appeared to represent a failed mechanism for neurons to remove damaged, misfolded, and aggregated proteins. Therefore, it was proposed that modified *tau* can serve as a biomarker for the diagnosis of AD.

*Diabetes.* Very recent investigation showed methylglyoxal modification of mSIN3 protein and its linkage to diabetes retinopathy [44]. The report showed that in diabetes, because of impaired metabolism of glucose, the level of glyoxal concentration increased. This glyoxal can modify proteins and finally can modulate gene transcription.

*Aging and Age-related Diseases.* Oxidatively modified proteins increase as a function of age. Studies revealed an age-related increase in the level of protein carbonyl content, oxidized methionine, protein hydrophobicity, and cross-linked and glycated proteins [9]. Factors reducing protein oxidation increase the life span of experimental animals and vice versa. Furthermore, a number of age-related diseases are shown to associate with an elevated level of oxidized proteins. The accumulation of oxidatively modified protein can be attributed to a multitude factors that govern (1) the rate of formation of various kinds of reactive oxygen species, (2) the level of antioxidant defense that guards proteins against oxidative modification of proteins, (3) the sensitivity of proteins to oxidative attack, and (4) the repair and elimination of damaged proteins. The ROS are formed by ionizing radiation, activation of neutrophils and macrophages, oxidase catalyzed reaction, lipid peroxidation, and glycation/glycoxidation reactions. It is well established that the accumulation of oxidized protein is associated with a number of diseases. Elevated levels of protein carbonyls have been found in AD diseases, amyotrophic lateral scelerosis, cataractogenesis, systemic amyloidosis, muscular dystrophy, Parkinson's disease, progeria, Warner's syndrome, rheumatoid arthritis, and respiratory distress syndrome. Elevated levels of proteins modified by lipid peroxidation products are associated with Parkinson's diseases, cardiovascular diseases, iron-induced renal carcinogenesis, and experimental pancreatitis and atherosclerosis. Elevated levels of protein glycation/glycoxidation endproducts (AGEs) are associated with diabetes mellitus, ADs, atherosclerosis, Parkinson's diseases, and Down's syndrome. Elevated levels of protein nitrotyrosine damaged are associated with atherosclerosis, AD, lung injury, multiple sclerosis, and endotoxemia.

*Cataractogenesis.* Cataractogenesis has been noted in the human lens, due to PTM of lens protein αβ-crystallin by carbamylation and acetylation at Lys-92 [45].

*Autoimmune Diseases.* PTM can dictate an antigen in eliciting autoimmune diseases. For example, isoaspartyl posttranslational modification (conversion of aspartic acid to isoaspartic acid) has been shown to trigger autoimmune response to self-proteins. The presence of isoaspartyl proteins has been observed as a major component of the amyloid containing brain plaques of patients with AD [46].

*Alcoholism.* Many investigations showed that alcoholism modulates the level of PTM-like phosphorylation [47]. Alcohol consumption was shown to decrease the sialic acid conjugation to transferrin, an important carrier protein secreted from the liver to blood and other glycoproteins. This observation led to the development of a laboratory test for chronic alcohol use. Similar studies showed that direct production of alcohol metabolism (alpha-hydroxyethyl radicals, acetaldehyde and lipid peroxides) causes PTM that correlates with alcohol consumption in animal models and human subjects [48].

## 10 PTM AS A TARGET FOR THERAPEUTIC DEVELOPMENT

Histone deacetylase (HDAC) inhibitors are the new class of agents that modulate gene expression by altering chromatin structure and gene transcription. Several classes of HDAC inhibitors are known as therapeutics for tumors. The depsipeptides (FR901228 or FK228) are under clinical trial [49]. Three other classes including short-chain fatty acids (phenylbutyrate and valproic acid), benzamides (Cl-994 and MS-27-275), and hydroxamic acid (suberoylanilide hydroxamic acid) are being developed. An attractive model is that the increase in histone acetylation leads to transcriptional activation of a few genes that can inhibit tumor growth. Ten structurally related HDACs have been described and fall into two classes [50, 51]. Class I HDACs consist of HDAC1, 2, 3, and 8; whereas class II HDACs consist of HDAC4, 5, 6, 7, 9, and 10. Members of a third class of HDACs (class III) are structurally unrelated to the human class I and class II HDACs, and they consist of homologues of the yeast Sir2 proteins [52]. The activity of class I and class II HDACs is inhibited by short-chain fatty acids and hydroxamic acids, but class III HDACs are not inhibited by these agents. Therefore, one major challenge is to identify specific HDACs inhibitors.

Another report showed curcumin (a component of many tropical spices) as a novel histone acetyl transferase (HAT) inhibitor [53]. The report showed that curcumin could block acetylation of histone and p53 *in vivo*, which led to apoptosis of Hela cells. Therefore, potential in the future exists to develop HAT inhibitors for managing cancers.

Similarly, histone arginine/lysine methylation can be the target for the development of therapeutics. The status of histone arginine methylation is intimately involved in gene transcription. Recently, several compounds were found to inhibit protein arginine methyltransferases (PRMTs) [54]. It can also be interesting to explore how this small molecule could be exploited for developing therapeutics.

Protein farnesylation is a lipid-conjugate-type posttranslational modification required for the cancer-causing activity of the GTPase Ras [55]. Although farnesyltransferase inhibitors (FTIs) are in clinical trials, their mechanism of action and the role of protein farnesylation in normal physiology are poorly understood. Protein farnesylation was found to be essential for early embryogenesis, dispensable for adult homeostasis, and critical for progression, but not initiation, of tumorigenesis. Preclinical work has revealed FTIs' ability to effectively inhibit tumor growth in vitro and in animal models across a wide range of malignant phenotypes. Acute myeloid leukemias (AMLs) are appropriate disease targets in that they express relevant biologic targets such as Ras, MEK, AKT, and others that may

depend on farnesyl protein transferase activity to promote cell proliferation and survival. Phase I trials in AML and myelodysplasia have demonstrated biologic and clinical activities as determined by target enzyme inhibition, low toxicity, and both complete and partial responses. As a result, phase II trials have been initiated to further validate the clinical efficacy and to identify downstream signal transduction targets that may be modified by these agents [56].

We have recently reported rapid inactivation of NF-kappa B by tyrosine nitration [12]. As NF-kappaB is involved in many cellular functions in response to inflammatory responses [57], it can be a very important target in antioxidation therapy.

Nuclear receptors (NRs) orchestrate the transcription of specific gene networks in response to binding of their cognate ligands. They also act as mediators in a variety of signaling pathways by integrating diverse PTMs. NR phosphorylation concerns exist on all three major domains. Often, phosphorylation of NRs by kinases that are associated with general transcription factors (e.g., cdk7 within TFIIH) or activated in response to a variety of signals (MAPKs, Akt, PKA, PKC) facilitates the recruitment of coactivators and, therefore, cooperates with the ligand to enhance transcription activation. But phosphorylation can also contribute to the termination of the ligand response through inducing DNA dissociation, triggering NR degradation or decreasing ligand affinity. These different modes of regulation reveal an unexpected complexity of the dynamics of NR-mediated transcription. Therefore, small molecules that can modulate the phosphorylation status of NR can also be developed as therapeutics [14, 19]. For comprehensive reviews on PTM as targets in therapeutic development, the readers are referred to review Refs. 57–59.

## FURTHER READING

*Proteomic Resources*

The readers are referred to the following websites, journals, and suppliers to acquire comprehensive knowledge about the proteomics resources.

*World Wide Web (WWW) Resources (short listed)*

http://www.matrixscience.com/: Theoretical digestion, database search option (intact protein) for PMF by MALDI–TOF MS data, LC–ESI–MS/MS data search for PMF and PTM identification. Other information, like MS/MS fragmentation and PTM, is also available.

http://proteinprospector.uscf.edu: Theoretical digestion using MS-Digest. One of the unique options in using MS–Digest is that one can generate theoretical MS of the modified peptide selecting a wide variety of PTMs. Database search for PMF can be done using MS-Fit and also mass peak list information from the common contaminants, for example, keratin and trypsin autolysis peaks in MS are available.

http://ca.expasy.org/: Protein database, calculation of protein MS and pI values, theoretical digestion. Prediction of varieties of PTMs (glucosylation, phosphorylation, sumoylation, etc.) can be done from the protein sequence.

http://www.ionsource.com/: Very helpful tutorials about PMF using MALDI–TOF MS and ESI–MS/MS, *de novo* protein sequencing and related information. This website provides

very helpful tips and information for beginners in proteomics. One can also get many other website links for proteomics resources from this website.

*Literature and Journals:*

(1) Proteomics, (2) Journal of Proteome Research, (3) Molecular & Cellular Proteomics, (4) Electrophoresis, and (5) BBA protein-proteomics.

*Vendors and Suppliers:*

A comprehensive lists of vendors and suppliers related to proteomics resources are available on Ionsoure website (http://www.ionsource.com/).

## 11 CONCLUSION

The understanding of PTM will advance as more new PTMs are explored as biomarkers and more powerful techniques and tools are developed. It can be predicted that, within a few years, PTM analysis can become a routine procedure like HPLC in chemistry or PCR in biology laboratories. Currently, research of PTM is focusing on the basic principle and the studies of biological problems. Predictably, it will gradually be applied to the discovery of molecular targets for disease intervention and the development of therapeutics.

## ACKNOWLEDGMENT

This work was supported by National Institutes of Health Grants DA11190, DA11806, DK54733, DK60521, and K02-DA13926 to L.-N. W.

## REFERENCES

1. Gooley A A, Packer N H (1997). The importance of protein co- and post-translational modifications in proteome projects. In M.R. Wilkins; K.L. Williams; R.D. Appel; et al. (eds.), Springer–Verlag, *Proteome Research: New Frontiers in Functional Genomics: Berlin*. pp. 65–91.
2. Pandey A, Mann M (2000). Proteomics to study genes and genomes. *Nature*. 405: 837–846.
3. Liebler D C (2001). *Introduction to Proteomics: Tools for the New Biology*. Humana Press, Totowa, NJ.
4. Huq M D, Khan S A, Park S W, et al. (2005). Mapping of phosphorylation sites of nuclear corepressor receptor interacting protein 140 by liquid chromatography-tandem mass spectrometry. *Proteomics*. 5:2157–2166.
5. Huq M D, Wei L N (2005). Post-translational modification of nuclear corepressor receptor-interacting protein 140 by acetylation. *Mol. Cell. Proteomics*. 4:975–983.
6. Barlev N A, Liu L, Chehab N H, et al. (2001). Acetylation of p53 activates transcription through recruitment of coactivators/histone acetyltransferases. *Mol. Cell*. 8:1243–1254.
7. Bedford M T, Richard S (2005). Arginine methylation an emerging regulator of protein function. *Mol. Cell*. 18:263–272.

8. Wells L, Vosseller K, Hart G W (2001). Glycosylation of nucleocytoplasmic proteins: signal transduction and O-GlcNAc. *Science*. 291:2376–2378.
9. Stadman E R (2001). Protein oxidation in aging and age-related diseases. *Ann. N.Y. Acad. Sci.* 928:22–38.
10. Johnson E S (2004). Protein modification by SUMO. *Ann. Rev. Biochem.* 73:355–382.
11. Lee D Y, Teyssier C, Strahl B D, et al. (2005). Role of protein methylation in regulation of transcription. *Endocr. Rev.* 26:147–170.
12. Park S W, Huq M D, Hu, X, et al. (2005). Tyrosine nitration on p65: A novel mechanism to rapidly inactivate nuclear factor-kappaB. *Mol. Cell. Proteomics.* 4:300–309.
13. Teyssier C, Ma, H, Emter R, et al. (2005). Activation of nuclear receptor coactivator PGC-1alpha by arginine methylation. *Genes Dev.* 19:1466–1473.
14. Rochette-Egly C (2003). Nuclear receptors: integration of multiple signalling pathways through phosphorylation. *Cell. Signal.* 15:355–366.
15. Cote J, Boisvert F M, Boulanger M C, et al. (2003). Sam68 RNA binding protein is an in vivo substrate for protein arginine N-methyltransferase 1. *Mol. Biol. Cell.* 14:274–287.
16. Cheng X, Hart G W (2001). Alternative O-glycosylation/O-phosphorylation of serine-16 in murine estrogen receptor beta: post-translational regulation of turnover and transactivation activity. *J. Biol. Chem.* 276:10570–10575.
17. McCann J P, Mayes J S, Hendricks G R, et al. (2001). Subcellular distribution and glycosylation pattern of androgen receptor from sheep omental adipose tissue. *J. Endocrinol.* 169:587–593.
18. Khan S A, Park S W, Huq M, et al. (2005). Protein kinase C-mediated phosphorylation of orphan nuclear receptor TR2: effects on receptor stability and activity. *Proteomics.* 5:3885–3894.
19. Gronemeyer H, Gustafsson J A, Laudet V (2004). Principles for modulation of the nuclear receptor superfamily. *Nature Rev. Drug Disc.* 3:950–964.
20. Goody R S, Durek T, Waldmann H, et al. (2005). Application of protein semisynthesis for the construction of functionalized posttranslationally modified rab GTPases. *Methods Enzymol.* 403:29–42.
21. Acconcia F, Ascenzi P, Bocedi A, et al. (2004). Palmitoylation-dependent estrogen receptor alpha membrane localization: regulation by 17beta-estradiol. *Mol. Biol. Cell.* 16:231–237.
22. Xu W, Cho H, Evans R M (2003). Acetylation and methylation in nuclear receptor gene activation. *Methods Enzymol.* 364:205–223.
23. Fu M, Wang C, Zhang X, et al. (2004). Acetylation of nuclear receptors in cellular growth and apoptosis. *Biochem. Pharmacol.* 68:1199–1208.
24. Chen L F, Williams S A, Mu, Y, et al. (2005). NF-kappaB RelA phosphorylation regulates RelA acetylation. *Mol. Cell. Biol.* 25:7966–7975.
25. Chuikov S, Kurash J K, Wilson J R, et al. (2004). Regulation of p53 activity through lysine methylation. *Nature.* 432:353–360.
26. Chevillard-Briet M, Trouche D, Vandel L (2002). Control of CBP co-activating activity by arginine methylation. *EMBO J.* 21:5457–5466.
27. Rezai-Zadeh N, Zhang X, Namour F, et al. (2003). Targeted recruitment of a histone H4-specific methyltransferase by the transcription factor YY1. *Genes Dev.* 17:1019–1029.
28. Chan W M, Mak M C, Fung T K, et al. (2006). Ubiquitination of p53 at multiple sites in the DNA-binding domain. *Mol. Cancer Res.* 4:15–25.

29. Wurgler-Murphy S M, King D M, Kennelly P J (2004). The Phosphorylation Site Database: A guide to the serine-, threonine-, and/or tyrosine-phosphorylated proteins in prokaryotic organisms. *Proteomics.* 4:1562–1570.
30. Goy M F, Springer M S, Adler J (1977). Sensory transduction in Escherichia coli: role of a protein methylation reaction in sensory adaptation. *Proc. Nat. Acad. Sci. USA.* 74:4964–4968.
31. Eichler J, Adams M W (2005). Posttranslational protein modification in Archaea. *Microbiol. Mol. Biol. Rev.* 69:393–425.
32. Wu R C, Qin J, Yi P, et al. (2004). Selective phosphorylations of the SRC-3/AIB1 coactivator integrate genomic reponses to multiple cellular signaling pathways. *Mol. Cell.* 15:937–949.
33. Vo N, Fjeld C, Goodman R H (2001). Acetylation of nuclear hormone receptor-interacting protein RIP140 regulates binding of the transcriptional corepressor CtBP. *Mol. Cell. Biol.* 21:6181–6188.
34. Gupta P, Huq M D, Khan S A, et al. (2005). Regulation of co-repressive activity of and HDAC recruitment to RIP140 by site-specific phosphorylation. *Mol. Cell. Proteomics.* 4:1776–1784.
35. Zhang K, Yau, P M, Chandrashekhar B, et al. (2004). Differentiation between peptides containing acetylated or tri-methylated lysines by mass spectrometry: An application for determining lysine 9 acetylation and methylation of histone H3. *Proteomics.* 4:1–10.
36. Hirota J, Satomi Y, Yoshikawa K, et al. (2003). ε-*N*,*N*,*N*-trimethyllysine-specific ions in matrix-assisted laser desorption/ionization-tandem mass spectrometry. *Rapid Commun. Mass Spectrom.* 17:371–376.
37. Johnstone R W (2004). Deamidation of Bcl-X(L): a new twist in a genotoxic murder mystery. *Mol. Cell.* 10:695–697.
38. Ramasamy R, Yan S F, Schmidt A M (2006). Methylglyoxal comes of AGE. *Cell.* 124:258–260.
39. Khan S A, Park S W, Huq M D, et al. (2006). Ligand-independent orphan receptor TR2 activation by phosphorylation at the DNA-binding domain. *Proteomics.* 6:123–130.
40. Sterner D E, Berger S L (2000). Acetylation of histones and transcription-related factors. *Microbiol. Mol. Biol. Rev.* 64:435–459.
41. Arrell D K, Neverova I, Van Eyk J E (2001). Cardiovascular proteomics: evolution and potential. *Circ. Res.* 88:763–773.
42. Scheler C, Muller E C, Stahl J, et al. (1997). Identification and characterization of heat shock protein 27 protein species in human myocardial two-dimensional electrophoresis patterns. *Electrophoresis.* 18:2823–2831.
43. Gong C X, Liu F, Grundke-Iqbal I, et al. (2005). Post-translational modifications of tau protein in Alzheimer's disease. *J. Neural Transmis. (Vienna, Austria),* 112:813–838.
44. Yao D, Taguchi T, Matsumura T, et al. (2006). Methylglyoxal modification of mSin3A links glycolysis to angiopoietin-2 transcription. *Cell.* 124:275–286.
45. Lapko V N, Smith D L, Smith J B (2001). In vivo carbamylation and acetylation of water-soluble human lens alphaB-crystallin lysine 92. *Protein Sci.* 10:1130–1136.
46. Mamula M J, Gee R J, Elliott J I, et al. (1999). Isoaspartyl post-translational modification triggers autoimmune responses to self-proteins. *J. Biol. Chem.* 274:22321–22327.
47. Anni H, Israel Y (2002). Proteomics in alcohol research. *Alcohol Res. Health.* 26:219–232.

48. Anni H, Pristatsky P, Israel Y (2003). Binding of acetaldehyde to a glutathione metabolite: mass spectrometric characterization of an acetaldehyde-cysteinylglycine conjugate. *Alcoholism, Clin. Exper. Res.* 27:1613–1621.
49. Richon V M, O'Brien J P (2002). Histone deacetylase inhibitors: a new class of potential therapeutic agents for cancer treatment. *Clin. Cancer Res.* 8:662–664.
50. Marks P A, Rifkind R A, Richon V M, et al. (2001). Histone deacetylases and cancer: causes and therapies. *Nat. Rev. Cancer.* 1:194–202.
51. Kao H Y, Lee C H, Komarov A, et al. (2002). Isolation and characterization of mammalian HDAC10, a novel histone deacetylase. *J. Biol. Chem.* 277:187–193.
52. Landry J, Sutton A, Tafrov S T, et al. (2000). The silencing protein SIR2 and its homologs are NAD-dependent protein deacetylases. *Proc. Nat. Acad. Sci. USA.* 97:5807–5811.
53. Balasubramanyam K, Varier R A, Altaf M (2004). Curcumin, a novel p300/CREB-binding protein-specific inhibitor of acetyltransferase, represses the acetylation of histone/nonhistone proteins and histone acetyltransferase-dependent chromatin transcription. *J. Biol. Chem.* 279:51163–51171.
54. Cheng D, Yadav N, King R W, et al. (2004). Small molecule regulators of protein arginine methyltransferases. *J. Biol. Chem.* 279:23892–23899.
55. Sebti S M (2005). Protein farnesylation: implications for normal physiology, malignant transformation, and cancer therapy. *Cancer Cell.* 7:297–300.
56. Thomas X, Elhamri M (2005). Farnesyltransferase inhibitors: preliminary results in acute myeloid leukemia. *Bulletin du Cancer.* 92:227–238.
57. Garg A, Aggarwal B B (2002). Nuclear transcription factor-kappaB as a target for cancer drug development. *Leukemia.* 16:1053–1068.
58. Rohlff C (2000). Proteomics in molecular medicine: applications in central nervous systems disorders. *Electrophoresis.* 21:1227–1234.
59. Valaskovic G A, Kelleher N L (2002). Miniaturized formats for efficient mass spectrometry-based proteomics and therapeutic development. *Curr. Topics Med. Chem.* 2:1–12.

# PEGYLATION: CAMOUFLAGE OF PROTEINS, CELLS, AND NANOPARTICLES AGAINST RECOGNITION BY THE BODY'S DEFENSE MECHANISM

Yong Woo Cho,[1] Jae Hyung Park,[2] Ji Sun Park,[1] and Kinam Park[3]

[1] *Asan Institute for Life Sciences, University of Ulsan College of Medicine, Seoul, Korea*
[2] *College of Environment and Applied Chemistry, Kyung Hee University, Gyeonggi-do, Korea*
[3] *Purdue University, West Lafayette, Indiana*

**Chapter Contents**

1 Introduction
2 PEGylated Proteins
  2.1 Chemical Modification
  2.2 Immunogenicity
  2.3 Clinical PEGylated Proteins
3 PEGylated Cells
  3.1 Immunocamouflage of Cells
  3.2 Red Blood Cells (RBCs)
  3.3 White Blood Cells
  3.4 Pancreatic Islets
4 Colloidal Nanoparticles with PEG-Exposed Surfaces
  4.1 Stealth Liposomes
  4.2 Polymeric Nanoparticles
5 Conclusions
  References

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

## 1 INTRODUCTION

Poly(ethylene glycol) (PEG) is a synthetic neutral polymer that is soluble in water as well as in many organic solvents, such as methylene chloride, ethanol, acetone, and chloroform. In aqueous solutions, PEG is highly hydrated due to the presence of the ether oxygen linkage as a hydrogen bond acceptor. The hydrated PEG has a high degree of segmental flexibility and large excluded volume. These properties enable PEG to behave as if 5–10 times larger than proteins of comparable molecular weights and to effectively exclude other polymers (both natural and synthetic) from PEG-grafted surfaces [1, 2]. As PEG is demonstrated to be nontoxic, the U.S. Food and Drug Administration (FDA) has approved it for use as a vehicle or base in foods, cosmetics, and pharmaceuticals, including injectable, topical, rectal, and nasal formulations. When injected into the bloodstream, PEG is rapidly removed from the body with clearance rates inversely proportional to polymer molecular weights [3]. PEG tends to accumulate in muscle, skin, bone, and liver to a higher extent than other organs irrespective of the molecular weights, and its elimination occurs either in the kidneys (for PEGs < 30 kDa) or in the feces (for PEGs > 20 kDa) [3, 4]. There have been no reports on the generation of antibodies to PEG under routine clinical administration of PEG–protein conjugates. A few reports have indicated production of antibodies to PEG under extreme experimental conditions using animals [5, 6]. It is important to note, however, that the minute quantities of PEG in therapeutic conjugates are generally considered to provide no chance of eliciting unwanted host responses.

PEGylation is the process of attaching PEG to various objects, such as peptides, proteins, cells, and other biologically active materials. As PEG has a unique set of properties in a biological environment, PEGylation of the target molecules has provided a wide range of biomedical applications. It has been demonstrated that PEGylated proteins have less toxicity [7, 8] and less immunogenicity [9, 10]. Increase in the molecular weights of proteins by PEGylation slows down kidney ultrafiltration, resulting in prolonged circulation in the bloodstream. The shielding effect of PEG reduces protein degradation by the enzymatic reaction and minimizes recognition of the immune system. In addition to proteins, recent efforts have made significant progress in designing PEGylated cells that show low immunogenic recognition, thus allowing transplantation. PEGylation has also been used to modify surfaces of drug and gene carriers to improve the blood circulation time and targeting efficiency.

## 2 PEGYLATED PROTEINS

### 2.1 Chemical Modification

Although therapeutic proteins are highly potent endogenous substances, they are generally susceptible to proteolytic degradation, aggregation, polymerization, and adsorption onto foreign surfaces. In addition, most proteins have short biological half-lives and often induce immunogenic responses when administrated into the body. In an attempt to stabilize proteins toward proteolysis and to increase blood circulation time in the body, various water-soluble polymers, such as polysaccharides [11, 12] and albumin [13], have been used for protein conjugation. Of the numerous natural and synthetic polymers, PEG is becoming a standard polymer

for chemical modification of proteins and peptides to enhance their pharmaceutical properties. The pioneering first steps in PEGylation of proteins were taken in the late 1970s by Professor Frank Davis and his colleagues at Rutgers University [4].

PEG is a linear or branched polyether terminated with hydroxyl groups that is synthesized by anionic ring opening polymerization of ethylene oxide. The molecular weights of linear PEGs available for biomedical applications usually range between a few hundred to several tens of thousands of daltons. For PEGylation of proteins, the reactive functional groups at the chain end of PEG need to be introduced. The functional group is selected based on the type of available reactive groups on proteins, originating from lysine, cysteine, histidine, arginine, aspartic acid, glutamic acid, serine, threonine, tyrosine, N-terminal amino group, and C-terminal carboxylic acid.

***PEGs for Carboxyl Conjugation.*** The nucleophilic PEG derivatives, such as PEG-amine and PEG-*p*-aminobenzylether [14], have been attached by the reaction with carbodiimide-activated carboxylic groups of proteins. For PEG-amine, this approach has been limited because the reactivity of PEG-amine is similar to that of the amino groups of proteins, resulting in poor selectivity. On the other hand, PEG-*p*-aminobenzylether is useful for selective reaction at a slightly acidic solution (pH 4.8–6.0) because it has the lower pKa value than the primary amino groups of the protein. Also, PEG-hydrazide (pKa ~ 3) can specifically react with the carbodiimide-activated carboxyl groups of proteins under acidic conditions (pH 4.5–5) at which primary amino groups of proteins are unreactive due to the protonation. It should be considered for protein modification, however, that the nucleophilic PEG derivatives often involve side reactions, including modification of tyrosyl and cysteinyl side groups and formation of *N*-acylurea derivatives [15, 16].

***PEGs for Amine Conjugation.*** The most common reactive sites on proteins for PEGylation are the alpha or epsilon groups of lysine or the *N*-terminal amino groups of other amino acids [17]. To react with amino groups of proteins, PEG has been activated by a variety of chemical modifications. Figure 1 shows various chemical structures of activated PEGs.

Activated PEGs can be prepared by reacting the primary alcohol group of PEG with the chlorine group of trichloro-s-triazine (cyanuric chloride), by which two remaining chlorine groups are available for protein modification [18]. The resulting PEG-dichlorotriazine (Figure 1a) can react with nucleophilic functional groups on the surface of proteins such as lysine, serine, tyrosine, cysteine, and histidine [19]. One chloride is displaced during the reaction, whereas the other remains less reactive and may provide the possibility of the side reaction causing chemical cross-linking of protein molecules. PEG-tresylates (Figure 1b) were used to nonspecifically modify amino groups to form secondary amine linkages that are stable and does not change the total charge of the conjugate, compared with the native protein [20]. PEG-tresylates have more specific reactivity to amino groups than PEG-dichlorotriazine, whereas they may involve the side reactions such as formation of a degradable sulfamate linkage [21].

PEG-aldehydes (Figure 1c) are a useful electrophilic reagent capable of specifically reacting with amino groups of proteins without interference from other nucleophiles belonging to proteins [22]. The reaction of PEG-aldehyde with protein proceeds through a reversible Schiff base intermediate that is reduced in the

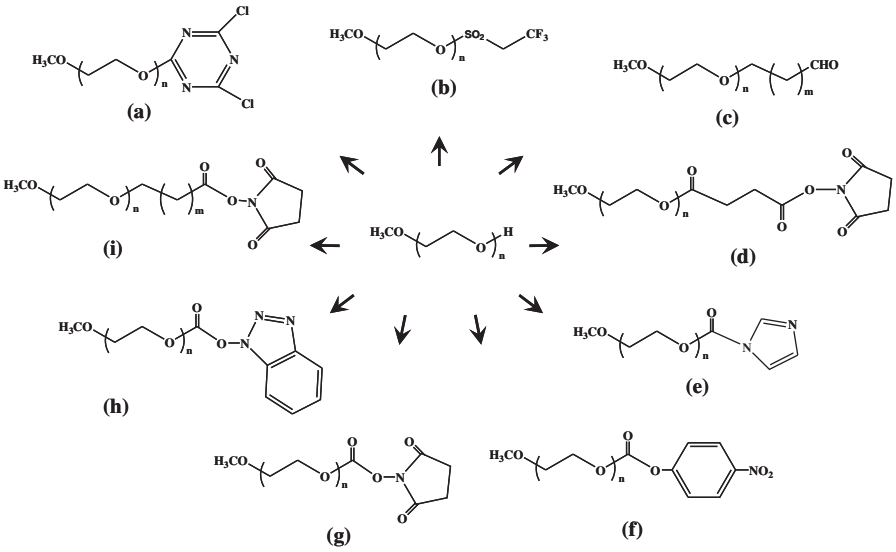


**Figure 1.** Chemical structures of activated PEGs for amine conjugation: (a) PEG-dichlorotriazine; (b) PEG-tresylate; (c) PEG-aldehydes; (d) PEG-succinimidyl succinate; (e) PEG-imidazole carbonate; (f) PEG-phenyl carbonate; (g) PEG-succinimidyl carbonate; (h) PEG-benzotriazole carbonate; and (i) PEG-active ester.

presence of sodium cyanoborohydride or sodium borohydride. Also, PEG-aldehyde has the ability to achieve selective modification between N-terminal α-amines and lysine ε-amines. When the reaction is performed under mild acidic conditions (pH 5–6), the α-amino groups are preferentially reacted with PEG-aldehyde. PEG-succinimidyl succinate (Figure 1d) is prepared by reaction of PEG with succinic anhydride, followed by carbodiimide-mediated condensation with N-hydroxy succinimide (NHS) [23]. This activated PEG reacts with proteins under mild conditions (pH < 7.8, room temperature), and the conjugates show comparable biological activities with native proteins [23, 24]. As the conjugates contain the ester linkage between PEG and protein, exposure to the physiological condition may induce hydrolysis, leading to loss of the benefits of PEG grafting. Katre et al. [25] demonstrated that the hydrolysis rate can be slowed by substituting the succinate group with a glutarate. PEG-imidazole carbonates (Figure 1e) and PEG-phenyl carbonates (Figure 1f) are prepared by reacting carbonylimidazole [17, 26] or chloroformate [27] with the terminal hydroxyl group of PEG, respectively. These derivatives have been used to produce conjugates via the formation of a carbamate bond between PEG and proteins. In general, they show much lower reactivity than PEG-succinimidyl succinate. The reaction time takes more than 24 h. The slower reaction rate is beneficial for selective modification of amino groups in proteins, thus producing conjugates with preserved biological activity.

PEG-succinimidyl carbonate (Figure 1g) is another activated PEG that specifically reacts with amino groups of proteins, thus forming a carbamate linkage [28]. It is found to be more selective than PEG-succinimidyl succinate for modification of the amino groups of proteins. In particular, PEG-succinimidyl carbonate is

highly useful to prepare a minimally toxic conjugate because the liberated NHS produced by the reaction is known to be nontoxic. PEG-benzotriazole carbonate (Figure 1h) is similar to PEG-succinimidyl carbonate. PEG-benzotriazole carbonate forms the carbonate bonds via modification of proteins and is more selective than PEG-succinimidyl succinate in reacting with amino groups of proteins [29]. Furthermore, PEG-benzotriazole carbonate can react under mild conditions within a short reaction time (e.g., pH 8.5 at 25°C for 30 min). Despite their high selectivity, both PEG-succinimidyl carbonate and PEG-benzotriazole carbonate have been recently reported to react with histidine and tyrosine residues of proteins, which forms hydrolytically unstable linkages such as imidazolecarbamate [30].

In recent years, the most popular PEG derivatives for protein modification have been active esters of PEG carboxylic acids (Figure 1i) [17]. As the active esters react with primary amines to form the amide linkage under physiological conditions, they are useful for preparing stable conjugates. PEG carboxylic acids are readily activated by reacting them with NHS and carbodiimide. Of various PEG-active esters, the succinimidyl ester of carboxymethylated PEG was the first derivative developed for protein modification [31]. This derivative, however, is extremely reactive toward hydrolysis that limits the extended applications. The reactivity of PEG-active esters was found to be controlled by changing the distance between the active ester and the PEG backbone [32]. For example, the hydrolysis half-life of carboxymethylated PEG is 0.75 min at pH 8 and 25°C, whereas those of PEG-active esters bearing one and two additional methylene groups between PEG and active esters are 16.5 and 23 min, respectively, under the same condition.

***PEGs for Cysteine Conjugation.*** Several PEG derivatives have been developed to specifically react with the thiol group of cysteines, including PEG-maleimide [33], PEG-orthopyridyl-disulfide [34], PEG-iodoacetamide [33], and PEG-vinyl sulfone [35]. These cysteine-targeting PEG derivatives are primarily designed for site-specific PEGylation of proteins because cysteine residues are rarely found on the surface of proteins, compared with lysine residues. If necessary, however, it is possible to add cysteine residues on the protein surfaces by genetic engineering, which may allow site-specific modification at areas that do not affect the biological activity of protein. Each derivative has its own characteristics. For example, the reaction of PEG-maleimide with the free thiol group is known to form a hydrolytically stable thioether linkage throughout the Michael-addition reaction. PEG-orthopyridyl-disulfide can react with protein sulfhydryl groups to form a disulfide bond capable of being cleaved to regenerate the native protein under reducing conditions. PEG-iodoacetamide reacts slowly with protein thiols by nucleophilic substitution via formation of a stable thioether linkage. For this polymer, the reaction should be carried out in a dark condition to prevent the generation of free iodine that may react with other amino acids such as tyrosine. PEG-vinyl sulfone is another activated PEG reacting with protein thiols to form a stable thioether bond.

***PEGs for Selective Modification.*** In an attempt to produce PEGylated proteins with preserved biological activity, selective modifications are often required. In most cases, the successful protein modification depends on the balance between reactivity and selectivity. It is generally accepted that the selective modification is

achieved by using less-reactive PEG derivatives, in which the reaction is performed at the low temperature for the long time period (>24 h).

Based on the reversible protection of interfering nucleophilic groups, selected amino groups in proteins have been modified with PEG. The protected intermediates of proteins are first prepared by chemically attaching the protective reagent to the amino groups that play an important role in biological activity. PEG is then conjugated to the preferred sites on the surfaces of proteins. The subsequent deprotection of PEGylated intermediates affords the final products that may show comparable activities with native proteins. This approach allowed selective modifications with PEG for insulin [36, 37], vapreotide [38], and tumor necrosis factor-alpha (TNF-α) [39].

The site-specific modifications of proteins with PEG have also been achieved by the enzyme-mediated coupling reaction [40]. One representative enzyme for this approach is a genetically engineered glutamine aminotransferase that can be used to selectively conjugate PEG to glutamine residues in a recognized substrate sequence. Interleukin-2, site-specifically PEGylated by this method, showed good *in vitro* activities and increased half-lives, compared with the native protein [40].

Thiol-specific PEG derivatives are useful for selective modification of proteins because a few cysteine sulfhydryl groups are generally available on the surface of proteins. The addition of cysteine residues at specific sites in a protein sequence is possible by genetic engineering, allowing site-specific PEGylation of proteins. Chapman et al. [41] have developed PEGylated antibody fragments by site-specific modification using cysteine-selective PEG derivatives. The conjugates showed long *in vivo* half-lives and full retention of antigen-binding properties. A cysteine residue was also inserted into anti-HIV protein cyanovirin-N by site-directed mutagenesis, followed by chemical reaction with thiol-specific PEG derivatives [42]. The resulting conjugates were highly active, whereas nonspecific modification with PEG resulted in a total loss of antiviral activity.

PEG-aldehyde derivatives have been used to selectively modify N-terminal amino groups of proteins [2, 17, 43]. For example, PEGylated granulocyte colony stimulating factor (GCSF) was prepared using PEG-aldehyde derivatives that were attached to the N-terminal methionine residue [44, 45]. The reaction mixture contained more than 80% of the intended conjugates with preserved biological activity. PEGylated GCSF is currently approved by the FDA for clinical use.

## 2.2 Immunogenicity

Although recent advances in the recombinant DNA technology have allowed production of several proteins in large quantities, their applications are frequently limited due to immunogenic responses by the body. For example, recombinant human interleukin-2 (rhIL-2), expressed and purified from *Escherichia coli*, has received attention as an anticancer therapeutic agent. However, some patients undergoing clinical trials with rIL-2 have reported to develop antibodies to the protein [9]. It is also known that repeated administrations of human insulin, produced by genetic engineering, resulted in generation of antibody [46]. Immunogenicity of recombinant human proteins can lead to serious clinical consequences. Antibody binding to a protein may change its pharmacokinetics in the body [47]. Antibody formation can inhibit the therapeutic effect of proteins and/or neutralize the essential endogenous proteins. Long-term administration of recombinant

human erythropoietin to patients induced formation of antibody capable of neutralizing endogenous erythropoietin [48, 49]. PEGylation has shown to suppress such adverse effects of recombinant proteins [18, 50]. In their pioneering studies on PEGylation, Abuchowski et al. demonstrated that administration of native albumin to rabbits resulted in rapid removal of protein by immuno-complex formation, whereas PEGylated albumin showed long circulation in the bloodstream [18]. Although a few references indicate generation of antibodies to PEG [5, 51], it is generally accepted that generation of anti-PEG antibodies is practically negligible, mainly due to its weak immunogenicity and the low amounts of PEGylated conjugates in clinical trials.

Suppression of immunogenicity by PEGylation is primarily based on shielding of the epitope (immunoreactive site) of a protein by flexible PEG chains. Improved immunological properties alter the pharmacokinetic profile of the modified protein. For example, when IgG was intravenously administrated to monkeys, it was cleared rapidly because of an antiprotein immune response [52]. On the other hand, PEGylated IgG did not elicit an immune response, and their dose levels were 125% as compared with the native antibody. This indicated that PEGylation of a protein increases the circulation time in the bloodstream by reducing clearance by the immune system.

The extent of suppression of immunogenicity is known to depend on the PEG molecular weight, number of grafted PEG chains, conjugated sites, and PEG architecture (linear or branched) [46, 53]. Protein immunogenicity is generally reduced as the number of PEG, and its molecular weight increases. Conjugation of branched PEG is reported to be more effective than linear PEG in improving immunological and pharmacokinetic properties of the modified proteins. Asparaginase modified with 10-kDa PEG reduced the antigenic character by 10-fold, as compared with that of the same protein modified with 5-kDa PEG [46]. Branched PEG could reduce immunogenicity of uricase more efficiently than linear PEG [54]. The author suggested that the "umbrella-like" structure of branched PEG efficiently prevented the approach of antiprotein antibodies and immuno-competent cells (Figure 2). It would

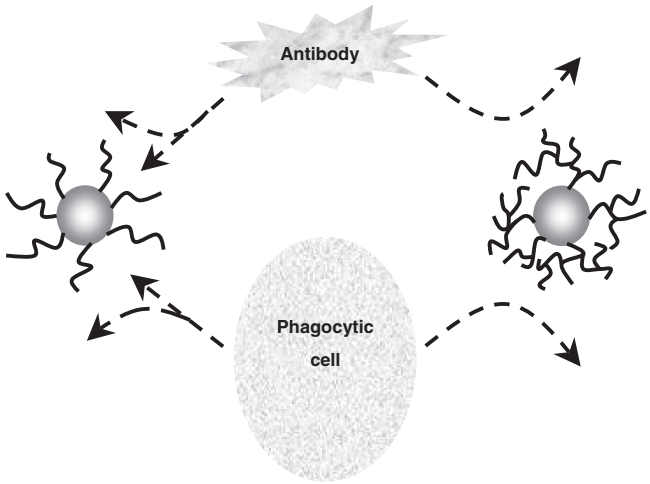


**Figure 2.** Molecular architectures of PEG grafted on proteins. The branched PEG protects more effectively from antibodies, phagocytic cells, and enzymes.

be preferred to attach the PEG chains around the antigenic epitopes, if this does not deteriorate the biological activity of the protein [50].

## 2.3 Clinical PEGylated Proteins

In the early 1990s, two PEGylated proteins, PEG-adenosine deaminase (Pegademase) and PEG-L-asparaginase (Pegaspargase), were approved by the FDA as therapeutic agents replacing the parent compounds. Recently approved PEGylated protein drugs include Pegfilgrastim, Peginterferon alfa-2a, and Peginterferon alfa-2b. More than 20 PEGylated proteins are on the market or in clinical trials, as shown in Figure 3 [4, 17].

Pegademase was developed for the treatment of severe combined immunodeficiency (SCI), which is associated with an inherited deficiency of adenosine deaminase [55]. SCI patients are conventionally treated by transfusing red blood cells containing adenosine deaminase. However, the patients often suffer from iron overload and transfusion-associated viral infections [56, 57]. Use of Pegademas does not need to consider such adverse effects and allows minimizing the dose of therapeutic agents. L-Asparaginase has been used for treatment of acute lymphocytic leukemia, acute lymphoblastic leukemia, and chronic myelogenous leukemia. L-Asparaginase, however, requires frequent intramuscular injections and often induces allergic reactions [58]. Compared with unpegylated L-asparaginase,

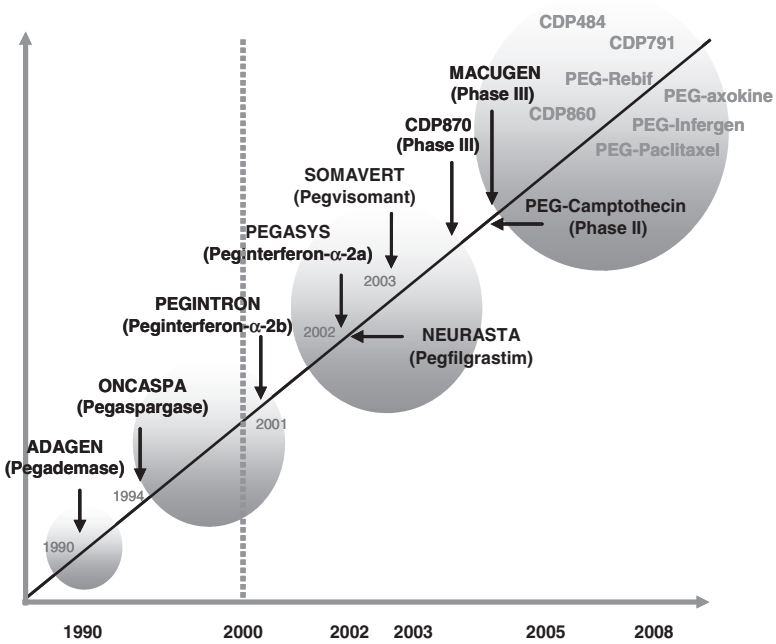


**Figure 3.** PEGylated products on the market or under clinical development. Since the first PEGylated protein was introduced, Pegademase (Adagen) entered the market in 1990 followed by a growing stream of PEGylated pharmaceuticals, as symbolized by circles in the figure. In 2004, a total of 26 PEGylated pharmaceuticals were on the market or under clinical trials.

Pegaspargase showed a substantially prolonged half-life. The half-lives of L-asparaginase and Pegaspargase were 20 and 357 h, respectively [59]. PEGylation of L-asparaginase also reduced adverse immune responses [58]. Although the actual cost of Pegaspargase is higher than that of multiple injections of unpegylated L-asparaginase, en-hanced therapeutic effects and minimal complications may make overall the cost considerably less than that of conventional therapy.

Interferon-based therapeutic drugs have been approved for treatment of various diseases, such as chronic hepatitis B and C, renal cell carcinoma, chronic myelogenous leukemia, and malignant melanoma. However, the clinical use of interferon has been limited because of rapid absorption from the subcutaneous injection site, large volume of distribution, rapid clearance via the kidney, short serum half-life (~6 h), and significant side effects such as depression and flu-like symptoms [60]. Interferons were initially PEGylated using a small, linear PEG (MW = 5 kDa) at multiple sites via the urea linkage [61]. These conjugates were not suitable for further clinical applications because of pharmacokinetics needed to be improved. Interferon alfa-2b was then modified with longer PEG (MW = 12 kDa) by forming a degradable linkage to improve the therapeutic effect [62]. The result was Peginterferon alfa-2b, which showed a long half-life with clinical efficacy, and is now marketed for treatment of chronic hepatitis C. The next generation of PEGylated interferon was interferon alfa-2a modified with a branched 40-kDa PEG via formation of an amide linkage [63]. The resulting Peginterferon alfa-2a allowed sustained delivery of the protein and showed the ability to be detected in the serum for longer than 1 week. It is interesting to note that the different chemical linkages of the two PEGylated interferons led to different formulations for the market. Peginterferon alfa-2b is available as a lyophilized powder, thus requiring reconstruction before use. On the other hand, Peginterferon alfa-2a is provided as a ready-to-use solution because the conjugates are linked by a stable amide bond.

Several PEGylated proteins are currently undergoing clinical trials. PEGylated human growth hormone antagonist (Pegvisomant) is being tested for treatment of acromegaly. Pegvisomant has been approved in Europe and is waiting for FDA approval in the United States. The PEGylated inhibitor of tumor necrosis factor-alpha is being tested for the treatment of rheumatoid arthritis and Crohn's disease [64]. Patients treated with the PEGylated protein experienced improvement in physical function, pain, vitality, and mental health [65].

## 3 PEGYLATED CELLS

### 3.1 Immunocamouflage of Cells

Immunological recognition of allogeneic tissues is a primary concern in transfusion and transplantation medicine. Graft-versus-host-disease (GVHD) is induced by the transfer of allogeneic lymphocytes into immunocompromised individuals [66]. GVHD can cause significant morbidity and mortality among patients with tissue transplantation. Several pharmacological agents that inhibit the T-cell signaling and activation have been developed such as cyclosporine, azathioprine, and methotrexate [67]. However, these drugs are less than optimal due to their substantial toxicity.

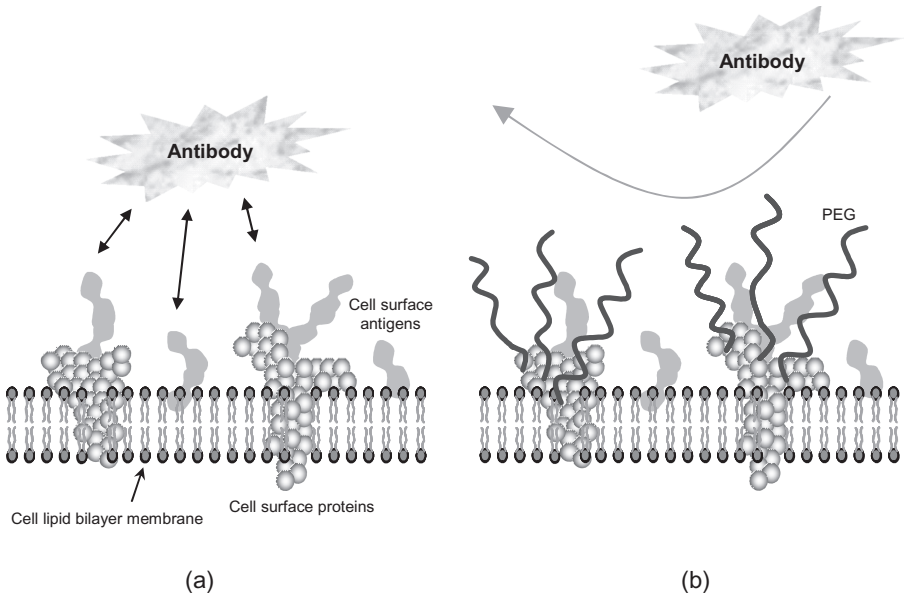


**Figure 4.** Immunocamouflage of cells by PEGylation: (A) Cell membrane containing surface antigens causes immune reactions, which destroy the transplanted cells. (B) PEGs grafted on the cell surface cytoprotect against antibodies.

Scott et al. [68–74] suggested a new method that can significantly diminish rejection episodes and may further enhance the induction of tolerance. They have demonstrated that the surface engineering of cells by PEGylation yielded antigenically and immunogically inert cells. Figure 4 schematically illustrates immunocamouflage of cells by surface PEGylation. The mechanisms underlying immunocamouflage by PEGylation are the loss of antigen recognition, impaired cell–cell interaction, and an inability of endogenous antibodies to effectively recognize and bind foreign epitopes. As a result of the high molecular flexibility and intense hydration, the grafted PEG can sterically block off a large three-dimensional volume thereby giving immunoprotective effects to membrane proteins and carbohydrates.

### 3.2 Red Blood Cells (RBCs)

Blood transfusion is a widely used therapy in medicine, particularly under surgery and trauma conditions. Surface engineering of red blood cells (RBCs) by PEGylation has been studied to camouflage the blood group antigens from their antibodies, thereby to generate universal RBCs. Several studies demonstrated that the surface engineering of RBCs with PEG significantly attenuated immunologic recognitions of surface antigens [68–71, 73–75]. PEGylation significantly decreased binding of human RBC blood group antibody and phagocytic destruction by heterologous phagocytes [68], PEGylation also dramatically diminished the immunogenicity of allogeneic and even xenogeneic RBCs in mouse models [71]. Importantly, this protective immunologic effect can be accomplished without adversely

affecting the structure, function, or viability of the modified RBCs. PEGylation was found to have no significant detrimental effects on the RBC structure or function. PEGylation did not affect RBC lysis, morphology, and hemoglobin oxidation state. The functional aspects of the PEGylated RBCs were also maintained, as evidenced by normal oxygen binding and cellular deformability [71].

The immunocamouflage efficacy is largely dependent on the PEG grafting density, PEG length, and tomography of the membrane surface. Differently functionalized PEGs have been used to chemically graft PEG onto the membrane of cells. Typical functionalized PEGs include PEG-dichlorotriazine [68], PEG-benzotriazole carbonate [73], and PEG-succinimidyl succinate [73]. They all preferentially target the ε-amino group of exposed lysine residues on membrane proteins. PEG-dichlorotriazine is the most common reagent. PEGylation of RBCs with PEG-dichlorotriazine has been shown to mask Rh antigens to a reasonable extent, but masking of antigen A or B was partial. Recently, Nacharaju et al. suggested a new approach for PEGylation of RBC proteins using thiolation-mediated maleimide chemistry [75]. In the study, masking of blood group antigens was not directly correlated to the PEG-grafting density or PEG chain length. For optimizing the masking of A, B, D, and CE antigens, a combination of different size PEG chains was used. PEG-5000 (molecular weight of PEG 5 kDa) alone masked the most important antigens of the Rh system (C, c, E, e, and D) from their antibodies. However, the masking of the A and B antigens needed a combination of PEG-20000 and PEG-5000 to inhibit agglutination of RBCs completely with anti-A or anti-B [75]. Recently, another strategy for RBC PEGylation was suggested by Chung et al. [76]. RBC PEGylation was carried out by incorporation of lipid-PEG into RBC membranes. The incorporation was rapid and spontaneous at room temperature. No change was observed in the membrane shape of the lipid-PEG incorporated RBC. The incorporation of lipid-PEG dramatically attenuated the antigenic recognition of RBC. As many as 35% of individuals with sickle cell disease or thalassemias exhibit clinically significant alloimmunization to non-ABO antigens [77, 78]. Immunocamouflage of RBCs by PEGylation may significantly attenuate the risk of alloimmunization for patients requiring chronic RBC transfusions such as patients with sickle cell anemia and thalassemias.

### 3.3 White Blood Cells

T-lymphocytes play a central role in immune responses. Transfusion or transplantation of T lymphocytes into an allogeneic recipient can evoke potent immune responses. Extensive attention has been paid to inactivation of T cells as well as other leukocytes (B lymphocytes, mononuclear cells, and granulocytes) from blood products. PEGylation dramatically blocked allorecognition and proliferation in human peripheral blood mononuclear cells (PBMCs) and isolated murine splenocytes [69, 72, 79–81]. Loss of cellular proliferation was not due to cytotoxicity of grafted PEG but to the global camouflage of the cell surface, which inhibited cell–cell interactions. T-cell and antigen-presenting cell adhesion molecules (CD2, CD11a), signaling molecules (CD3ε, T-cell receptor), and costimulatory molecules (CD28, CD80) were efficiently camouflaged by PEGylation [81]. Immunocamouflage was clearly observed in murine models of transfusion-associated graft versus host disease (TA-GVHD) [81]. Mice receiving saline or the PEGylated splenocytes

exhibited an identical survival curve, whereas mice receiving control splenocytes (unmodified) showed significant TA-GVHD. This improved survival was directly associated with the donor cell proliferation. Unmodified control splenocytes showed significant proliferation due to allorecognition of the host mice. In contrast, PEGylated cells failed to proliferate due to immunocamouflage and lack of allorecognition, indicating that PEGylation of donor splenocytes significantly decreased donor T-cell proliferation, thereby suppressing GVHD.

### 3.4 Pancreatic Islets

Transplantation of pancreatic islets has been one of the most challenging examples of tissue transplantation. Insulin-dependent diabetes mellitus (IDDM) is a disease from the autoimmune destruction of β cells of the islets of Langerhans in the pancreas. Tranplantation of islets into diabetic patients can be the most attractive mode of treatment, but the islets must be protected from the host's immune system to prevent graft rejection. Transplanted islets are recognized by host as antigens that activate immune cells, such as macrophages, granulocytes, and lymphocytes. Several studies demonstrated that camouflage of pancreatic islets by PEGylation could be an attractive approach to prevent activation of immune cells and secretion of cytokines, thereby improving transplantation of pancreatic islets [82–87].

PEG chains were covalently conjugated to the amino groups of collagen capsules of islet at physiological conditions through a stable amide bond [84]. PEG was not grafted onto the inner islet cell membrane but directly onto the collagen matrix of islet. Therefore, PEGylation did not damage the inner islet cells. Cell viability and functional activity remained intact after PEGylation. When free islets were cultured with lymphocytes, they completely lost the integrity of the collagen capsule [86]. The co-cultured lymphocytes were activated by free islets, thereby secreting a large amount of IL-2 and TNF-α cytokines. In contrast, the PEGylated islets were not damaged by co-culture with lymphocytes. The co-cultured lymphocytes were not activated by the PEGylated islets and rarely secreted IL-2 and TNF-α cytokines, suggesting that the grafted PEG effectively prevented cell–cell interaction necessary for the allogeneic rejection. The PEGylated islets were not affected by macrophoages, either. However, when co-cultured with lipopolysaccharide (LPS)-stimulated macrophages (activated macrophages), the PEGylated islets lost the integrity of the collagen capsule and were completely destroyed. Once macrophages are activated, they secrete several kinds of cytotoxic molecules, such as IL-1β, TNF-α, and nitric oxide (NO). Complete destruction of PEGylated islets indicates that the grafted PEG could not prevent infiltration of the cytotoxic molecules into the islets [84, 86].

## 4 COLLOIDAL NANOPARTICLES WITH PEG-EXPOSED SURFACES

At the interface of nanotechnology and biotechnology, nanosized colloidal particles (nanoparticles) have emerged as a new entity with great potential for target-specific therapy and imaging [88–90]. The fundamental advantages of colloidal nanoparticles for drug delivery result from their two main basic properties. First,

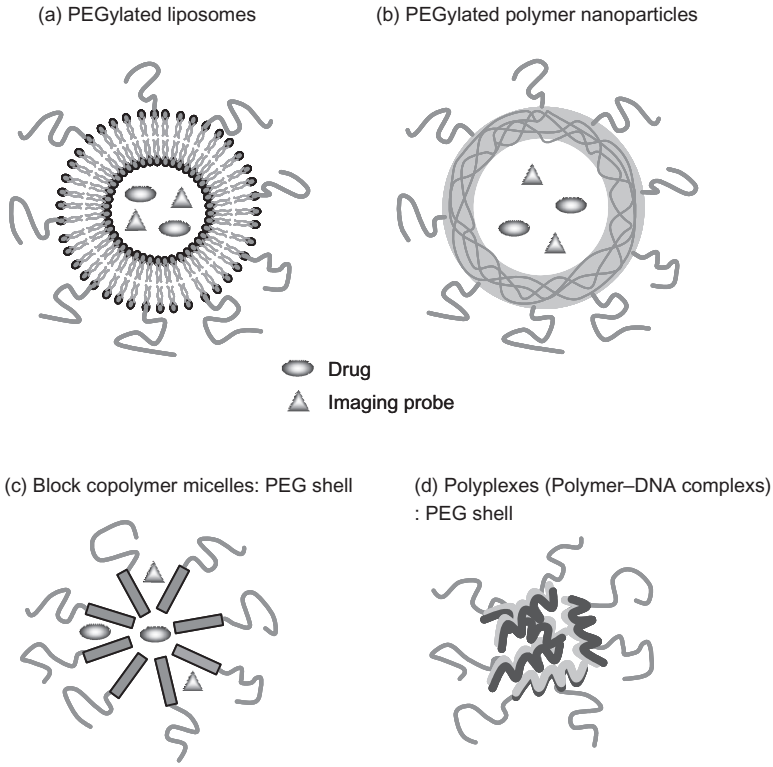


**Figure 5.** Colloidal nanoparticles with PEG-exposed surfaces.

nanoparticles due to their small size can penetrate through capillaries and are taken up by cells, and this allows efficient drug accumulation at the target sites. Second, nanoparticles allow sustained drug release within the target site over a period of days or even weeks. Thus, nanoparticles could be effective delivery systems for therapeutic and imaging payloads. However, intravenously injected nanoparticles are also treated as foreign by the host, just like exogenously delivered proteins and cells. The nanoparticles should avoid recognition by the cells of the reticuloendothelial system (RES), namely macrophages of the liver, spleen, lungs, bone marrow, and lymphnodes. Adsorption of plasma proteins onto the surface of nanoparticles is also highly problematic for efficient delivery of drugs and diagnostic agents. As a result, there has been a growing interest in the surface engineering of colloidal nanoparticles to avoid rapid recognition by RES and to reduce adsorption of plasma proteins. These systems are often called "Stealth" nanoparticles, which are "invisible" to macrophages and have a prolonged half-life in the blood circulation. PEGylation has been increasingly exploited to camouflage colloidal nanoparticles against recognition by the body's defenses. Figure 5 schematically represents different colloidal nanopaticles with the PEG-exposed surface.

### 4.1 Stealth Liposomes

Development of PEGylated liposomes, also known as "stealth" liposomes, greatly broadened the application of liposomes. For incorporation of PEG into the

liposomal bilayer, various functionalized lipids have been developed that contain a primary amino group, an epoxy group, or a diacylglycerol moiety [91, 92]. Alternatively, activated PEGs were anchored to reactive phospholipids groups of preformed liposomes [93]. In most cases, PEGylated liposomes with prolonged circulation time are in the size range of 70 to 200 nm and contain 3–7 mol% of PEG in addition to various amounts of phospholipids and cholesterol. The circulation half-lives of such stealth liposomes are 15–24 h in rodents and as high as 45 h in humans. In addition to PEG, various hydrophilic polymers have been grafted onto liposomes for prolonging the circulation time. They include polyvinylpyrrolidone, poly(vinyl alcohol), poly(2-methyl-2-oxazoline), poly(2-ethyl-2-oxazoline), polyacrylamide, poly(acryloyl morphine), and N-(2-hydroxylpropyl)methacrylamide copolymers [92, 94–98].

### 4.2 Polymeric Nanoparticles

Surface engineering of polymeric nanoparticles with PEG has been performed by physical adsorption, by incorporation during production of nanoparticles, or by covalent attachment to the surface. Typical examples of PEG-conjugated polymeric nanoparticles are PEG-poly(L-lactic acid) (PLA), PEG-poly(l-lactic-co-glycolic acid) (PLGA), PEG-poly(ε-caprolactone) copolymers, and PEG-polyphosphazene [92]. Although *in vitro* physico-chemical characterization studies, such as measurement of hydrodynamic volume and protein adsorption, have confirmed the steric barrier effects of PEG, the circulation half-lives of these nanoparticles were not as long as expected, generally less than a few hours. Many polymeric nanoparticles with a promising outlook based on *in vitro* results have faced the same problems when tested *in vivo*. Furthermore, little information is known with regard to the colloidal stability of nanoparticulate systems *in vivo* and their extent of interaction with blood and cellular components. This is an area that needs more understanding for effective use of nanoparticles.

## 5 CONCLUSIONS

PEGylation offers significant pharmacological advantages to proteins, cells, and colloidal nanoparticles. PEGylation improves protein solubility and molecular stability, and it reduces protein immunogenicity. By preventing rapid renal clearance, PEGylation extends the circulating life of proteins, thus allowing longer duration of action. PEGylation greatly enhances the clinical value of therapeutic proteins. Indeed, many PEGylated proteins have entered the market or are under clinical trials. PEGylation provides a unique strategy in avoiding immunological recognition of allogeneic cells and tissues. The global camouflage of the cell surface by PEGylation effectively blocks adhesion, recognition, and stimulatory pathways involved in the immune response. More importantly, unlike most therapeutics, PEGylation is nontoxic to modified cells as well as to the animal as a whole. PEGylation provides colloidal nanoparticles with a way to avoid rapid recognition by RES. PEGylation reduces deposition of plasma proteins to the surface of nanoparticles and thereby prolongs the circulating time in the bloodstream. Indeed, development of the PEGylation technology has been a cornerstone of the advances

in protein therapeutics. Not only has this technology brought numerous clinically useful products, but also triggered explosive interest in polymers in the field of pharmaceutical biotechnology.

## REFERENCES


1. Harris J M (1992). *Poly(ethylene glycol) chemistry: Biotechnical and biomedical applications*, Plenum Press, New York.
2. Harris J M, Zalipsky S (1997). *Polyethylene Glycol Chemistry and Biological Applications*, American Chemical Society, Washington, D.C.
3. Yamaoka T, Tabata Y, Ikada Y (1994). Distribution and tissue uptake of poly(ethylene glycol) with different molecular weights after intravenous administration to mice. *J. Pharm. Sci.* 83:601–606.
4. Harris J M, Chess R B (2003). Effect of pegylation on pharmaceuticals. *Nat. Rev. Drug Discov.* 2:214–221.
5. Richter A W, Akerblom E (1983). Antibodies against polyethylene glycol produced in animals by immunization with monomethoxy polyethylene glycol modified proteins. *Int. Arch. Allergy Appl. Immunol.* 70:124–131.
6. Cheng T L, Wu P Y, Wu M F (1999). Accelerated clearance of polyethylene glycol-modified proteins by anti-polyethylene glycol IgM. *Bioconjug. Chem.* 10:520–528.
7. Talpaz M (2001). Interferon-alfa-based treatment of chronic myeloid leukemia and implications of signal transduction inhibition. *Semin. Hematol.* 38:22–27.
8. Santhanam S, Decatris M, O'Byrne K (2002). Potential of interferon-alpha in solid tumours: Part 2. *BioDrugs.* 16:349–372.
9. Katre N V (1990). Immunogenicity of recombinant IL-2 modified by covalent attachment of polyethylene glycol. *J. Immunol.* 144:209–213.
10. Yang Z, Wang J, Lu Q, et al. (2004). PEGylation confers greatly extended half-life and attenuated immunogenicity to recombinant methioninase in primates. *Cancer Res.* 64:6673–6678.
11. Melton R G, Wiblin C N, Foster R L (1987). Covalent linkage of carboxypeptidase G2 to soluble dextrans–I. Properties of conjugates and effects on plasma persistence in mice. *Biochem. Pharmacol.* 36:105–112.
12. Melton R G, Wiblin C N, Baskerville A, et al. (1987). Covalent linkage of carboxypeptidase G2 to soluble dextrans–II. In vivo distribution and fate of conjugates. *Biochem. Pharmacol.* 36:113–121.
13. Wong K, Cleland L G, Poznansky M J (1980). Enhanced anti-inflammatory effect and reduced immunogenicity of bovine liver superoxide dismutase by conjugation with homologous albumin. *Agents Actions.* 10:231–239.
14. Pollak A, Whitesides G M (1976). Organic synthesis using enzymes in two-phase aqueous ternary polymer systems. *J. Am. Chem. Soc.* 98:289–291.
15. Zalipsky S, Gilon C, Zilkha A (1983). Attachment of drugs to polyethylene glycols. *Eur. Polym. J.* 19:1177–1183.
16. Zalipsky S (1995). Functionalized poly(ethylene glycol) for preparation of biologically relevant conjugates. *Bioconjug. Chem.* 6:150–165.
17. Roberts M J, Bentley M D, Harris J M (2002). Chemistry for peptide and protein PEGylation. *Adv. Drug Deliv. Rev.* 54:459–476.
18. Abuchowski A, van Es T, Palczuk N C, et al. (1977). Alteration of immunological properties of bovine serum albumin by covalent attachment of polyethylene glycol. *J. Biol. Chem.* 252:3578–3581.

19. Zalipsky S, Lee C (1992). Use of functionalized poly(ethylene glycol)s for modification of polypeptides. In J M Harris, S Zalipsky, (eds.), *Polyethylene Glycol Chemistry: Biotechnical and Biomedical Applications*: Plenum, New York, pp. 347–370.
20. Nilsson K, Mosbach K (1984). Immobilization of ligands with organic sulfonyl chlorides. *Methods Enzymol.* 104:56–69.
21. Gais H J, Ruppert S (1995). Modification and immobilization of proteins with polyethylene glycol tresylates and polysaccharide tresylates: Evidence suggesting a revision of the coupling mechanism and the structure of the polymer-polymer linkage. *Tetrahedron. Lett.* 36:3837–3838.
22. Harris J M, Herati R M (1993). Preparation and use of polyethylene glycol propionaldehyde. U.S. Patent 5,252,714.
23. Abuchowski A, Kazo G M, Verhoest, Jr. C R, et al. (1984). Cancer therapy with chemically modified enzymes. I. Antitumor properties of polyethylene glycol-asparaginase conjugates. *Cancer Biochem. Biophys.* 7:175–186.
24. Pasta P, Riva S, Carrea G (1988). Circular dichroism and fluorescence of polyethylene glycol-subtilisin in organic solvents. *FEBS Lett.* 236:329–332.
25. Katre N V, Knauf M J, Laird W J (1987). Chemical modification of recombinant interleukin 2 by polyethylene glycol increases its potency in the murine Meth A sarcoma model. *Proc. Natl. Acad. Sci. U.S.A.* 84:1487–1491.
26. Tondelli L, Laus M, Ferruti A S A (1985). Poly(ethylene glycol) imidazolyl formates as oligomeric drug-binding matrices. *J. Control Release.* 1:251–257.
27. Veronese F M, Largajolli R, Boccu E, et al. (1985). Surface modification of proteins. Activation of monomethoxy-polyethylene glycols by phenylchloroformates and modification of ribonuclease and superoxide dismutase. *Appl. Biochem. Biotechnol.* 11:141–152.
28. Miron T, Wilchek M (1993). A simplified method for the preparation of succinimidyl carbonate polyethylene glycol for coupling to proteins. *Bioconjug. Chem.* 4:568–569.
29. Dolence E K, Hu C, Tsang R, et al. (1997). Electrophilic polyethylene oxides for the modification of polysaccharides, polypeptides (proteins) and surfaces. U.S. Patent 5,985,263.
30. Wang Y S, Youngster S, Bausch J, et al. (2000). Identification of the major positional isomer of pegylated interferon alpha-2b, *Biochemistry* 39:10634–10640.
31. Zalipsky S, Barany G (1990). Facile synthesis of alpha-hydroxy-omega-carboxymethylpolyethylene oxide. *J. Bioact. Compat. Polym.* 227–231.
32. Harris J M, Kozlowski A (1997). Polyethylene glycol and related polymers monosubstituted with propionic or butanoic acids and fuctional derivatives thereof for biotechnical applications. U.S. Patent 5,672,662.
33. Kogan T P (1992). The synthesis of substituted methoxy-poly(ethylene glycol) derivatives suitable for selective protein modification. *Synth. Commun.* 22:2417–2424.
34. Woghiren C, Sharma B, Stein S (1993). Protected thiol-polyethylene glycol: a new activated polymer for reversible protein modification. *Bioconjug. Chem.* 4:314–318.
35. Morpurgo M, Veronese F M, Kachensky D (1996). Preparation of characterization of poly(ethylene glycol) vinyl sulfone. *Bioconjug. Chem.* 7:363–368.
36. Uchio T, Baudys M, Liu F, et al. (1999). Site-specific insulin conjugates with enhanced stability and extended action profile. *Adv. Drug Deliv. Rev.* 35:289–306.
37. Hinds K D, Kim S W (2002). Effects of PEG conjugation on insulin properties. *Adv. Drug Deliv. Rev.* 54:505–530.

38. Morpurgo M, Monfardini C, Hofland L J, et al. (2002). Selective alkylation and acylation of alpha and epsilon amino groups with PEG in a somatostatin analogue: tailored chemistry for optimized bioconjugates. *Bioconjug. Chem.* 13:1238–1243.
39. Tsunoda S, Ishikawa T, Yamamoto Y (1999). Enhanced antitumor potency of polyethylene glycolylated tumor necrosis factor-alpha: A novel polymerconjugation technique with a reversible amino-protective reagent. *J. Pharmacol. Exp. Ther.* 290: 368–372.
40. Sato H (2002). Enzymatic procedure for site-specific pegylation of proteins. *Adv. Drug Deliv. Rev.* 54:487–504.
41. Chapman A P, Antoniw P, Spitali M, et al. (1999). Therapeutic antibody fragments with prolonged in vivo half-lives. *Nat. Biotechnol.* 17:780–783.
42. Hinds K D (2005). Protein conjugation, cross-linking, and PEGylation. In R I Mahato, *Biomaterials for Delivery and Targeting of Proteins and Nucleic Acids.* CRC Press, Boca Raton, F. L., pp. 120–185.
43. Pepinsky R B, LePage D J, Gill A, et al. (2001). Improved pharmacokinetic properties of a polyethylene glycol-modified form of interferon-beta-1a with preserved in vitro bioactivity. *J. Pharmacol. Exp. Ther.* 297:1059–1066.
44. Kinstler O B, Brems D N, Lauren S L, et al. (1996). Characterization and stability of N-terminally PEGylated rhG-CSF. *Pharm. Res.* 13:996–1002.
45. Kinstler O B, Gabriel N E, Farrar C E, et al. (1999). N-terminally chemically modified protein compositions and methods. U.S. Patent 5,985,265.
46. Caliceti P, Veronese F M (2003). Pharmacokinetic and biodistribution properties of poly(ethylene glycol)-protein conjugates. *Adv. Drug Deliv. Rev.* 55:1261–1277.
47. Schernthaner G (1993). Immunogenicity and allergenic potential of animal and human insulins. *Diabetes Care.* 16(suppl 3):155–165.
48. Casadevall N, Nataf J, Viron B, et al. (2002). Pure red-cell aplasia and antierythropoietin antibodies in patients treated with recombinant erythropoietin. *N. Engl. J. Med.* 346:469–475.
49. Gershon S K, Luksenburg H, Cote T R, et al. (2002). Pure red-cell aplasia and recombinant erythropoietin. *N. Engl. J. Med.* 346:1584–1586.
50. Hermeling S, Crommelin D J, Schellekens H, et al. (2004). Structure-immunogenicity relationships of therapeutic proteins. *Pharm. Res.* 21:897–903.
51. Richter A W, Akerblom E (1984). Polyethylene glycol reactive antibodies in man: titer distribution in allergic patients treated with monomethoxy polyethylene glycol modified allergens or placebo, and in healthy blood donors. *Int. Arch. Allergy Appl. Immunol.* 74:36–39.
52. Chapman A P (2002). PEGylated antibodies and antibody fragments for improved therapy: A review. *Adv. Drug Deliv. Rev.* 54:531–545.
53. Veronese F M (2001). Peptide and protein PEGylation: a review of problems and solutions. *Biomaterials.* 22:405–417.
54. Caliceti P, Schiavon O, Veronese F M (2001). Immunological properties of uricase conjugated to neutral soluble polymers. *Bioconjug. Chem.* 12:515–522.
55. Hershfield M S (1995). PEG-ADA replacement therapy for adenosine deaminase deficiency: an update after 8.5 years. *Clin. Immunol. Immunopathol.* 76:S228–232.
56. Burnham N L (1994). Polymers for delivering peptides and proteins. *Am. J. Hosp. Pharm.* 51:210–218; 228–229.
57. Hilman B C, Sorensen R U (1994). Management options: SCIDS with adenosine deaminase deficiency. *Ann. Allergy.* 72:395–403; quiz 403–394, 407.

58. Graham M L (2003). Pegaspargase: A review of clinical studies. *Adv. Drug Deliv. Rev.* 55:1293–1302.
59. Holle L M (1997). Pegaspargase: An alternative? *Ann. Pharmacother.* 31:616–624.
60. Wills R J (1990). Clinical pharmacokinetics of interferons. *Clin. Pharmacokinet.* 19:390–399.
61. Monkarsh S P, Ma Y, Aglione A, et al. (1997). Positional isomers of monopegylated interferon alpha-2a: isolation, characterization, and biological activity. *Anal. Biochem.* 247:434–440.
62. Kozlowski A, Harris J M (2001). Improvements in protein PEGylation: Pegylated interferons for treatment of hepatitis C. *J. Control Rel.* 72:217–224.
63. Bailon P, Palleroni A, Schaffer C A, et al. (2001). Rational design of a potent, long-lasting form of interferon: a 40 kDa branched polyethylene glycol-conjugated interferon alpha-2a for the treatment of hepatitis C. *Bioconj. Chem.* 12:195–202.
64. Edwards C K 3rd (1999). PEGylated recombinant human soluble tumour necrosis factor receptor type I (r-Hu-sTNF-RI): Novel high affinity TNF receptor designed for chronic inflammatory diseases. *Ann. Rheum. Dis.* 58(suppl 1):I73–81.
65. Moreland L W, McCabe D P, Caldwell J R, et al. (2000). Phase I/II trial of recombinant methionyl human tumor necrosis factor binding protein PEGylated dimer in patients with active refractory rheumatoid arthritis. *J. Rheumatol.* 27:601–609.
66. Brubaker D B (1993). Immunopathogenic mechanisms of posttransfusion graft-vs-host disease. *Proc. Soc. Exp. Biol. Med.* 202:122–147.
67. Blazer B R, Korngold R, Vallera D A (1997). Recent advances in graft-versus-host disease (GVHD) prevention. *Immunol. Rev.* 157:79–109.
68. Scott M D, Murad K L, Koumpouras F, et al. (1997). Chemical camouflage of antigenic determinants: Stealth erythrocytes. *Proc. Natl. Acad. Sci. U.S.A.* 94:7566–7571.
69. Scott M D, Murad K L (1998). Cellular camouflage: Fooling the immune system with polymers. *Curr. Pharmeceut. Des.* 4:423–438.
70. Bradley A J, Test T T, Murad K L, et al. (2001). Complement interactions with methoxypoly(ethylene glycol)-modified human erythrocytes. *Transfusion* 41:1225–1233.
71. Murad K L, Mahany K L, Kuypers F A, et al. (1999). Structural and functional consequences of antigenic modulation of red blood cells with methoxypoly(ethylene glycol). *Blood* 93:2121–2127.
72. Murad K L, Gosselin E J, Eaton J W, et al. (1999). Stealth cells: Prevention of MHC class II mediated T cell activation by cell surface modification. *Blood.* 94:2135–2141.
73. Bradley A J, Murad K L, Regan K L, et al. (2002). Biophysical consequences of linker chemistry and polymer size on stealth erythrocytes: Size dose matter. *Biochim. Biophys. Acta.* 1561:147–158.
74. Scott M D, Chen A M (2004). Beyond the red cell: pegylation of other blood cells and tissues. *Transfus. Clin. Biol.* 11:40–46.
75. Nacharaju P, Boctor F N, Manjula B N, et al. (2005). Surface decoration of red blood cells with maleimidophenyl-polyethylene glycol facilitated by thiolation with iminothiolane: an approach to mask A, B, and D antigens to generate universal red blood cells. *Transfusion.* 45:374–383.
76. Chung H A, Kato K, Itoh C, et al. (2004). Casual cell surface remodeling using biocompatible lipid-poly(ethylene glycol) (n): Development of steal cells and monitoring of cell membrane behavior in serum-supplemented conditions. *J. Biomed. Mater. Res.* 70A:179–185.
77. Sirchia G, Zanella A, Parravicini A, et al. (1985). Red cell alloantibodies in thalassemia major. Results of an Italian cooperative study. *Transfusion.* 25:110–112.

78. Vichinsky E P, Earles A, Johnson R A, et al. (1990). Alloimmunization in sickle cell anemia and transfusion of racially unmatched blood. *N. Engl. J. Med.* 322.
79. Scott M D, Bradley A J, Murad K L (2000). Camouflaged red cells: low technology bioengineering for tranfusion medicine? *Transfus. Med. Rev.* 14:53–63.
80. Chen A M, Scott M D (2001). Current and future applications of immunologic attenuation via pegylation of cells and tissues. *BioDrugs.* 15:833–847.
81. Chen A M, Scott M D (2003). Immunocamouflage: Prevention of transfusion-induced graft-versus-host disease via polymer grafting of donor cells. *J. Biomed. Mater. Res.* 67A:626–636.
82. Panza J L, Wagner W R, Rilo H L R, et al. (2000). Treatment of rat pancreatic islets with reactive PEG. *Biomaterials.* 21:1155–1164.
83. Lee D Y, Yang K, Lee S, et al. (2002). Optimization of monomethoxy-polyethylene glycol grafting on the pancreatic islet capsules. *J. Biomed. Mater. Res.* 62:372–377.
84. Lee D Y, Nam J H, Byun Y (2004). Effect of polyethylene glycol grafted onto islet capsules on prevention of splenocyte and cytokine attacks. *J. Biomater. Sci. Polym. Ed.* 15:753–766.
85. Contreras J L, Xie D, Mays J, et al. (2004). A novel approach to xenotransplantation combining surface engineering and genetic modification of isloated adult porcine islets. *Surgery.* 136:537–547.
86. Jang J Y, Lee D Y, Park S J, et al. (2004). Immune reactions of lymphocytes and macrophages against PEG-grafted pancreatic islets. *Biomaterials.* 25:3663–3669.
87. Xie D, Smyth C A, Eckstein C, et al. (2005). Cytoprotection of PEG-modified adult porcine pancreatic islets for improved xenotransplantation. *Biomaterials.* 26:403–412.
88. Sahoo S K, Labhasetwar V (2003). Nanotech approaches to drug delivery and imaging. *Drug Disc. Today.* 8:1112–1120.
89. Barrat G M (2000). Therapeutic applications of colloidal drug carriers. *Pharm. Sci. Technol. Today.* 3:163–171.
90. Ferrari M (2005). Cancer nanotechnology: Opportunities and challenges. *Nature Rev. Cancer.* 5:161–171.
91. Drummond D C, Meyer O, Hong K, et al. (1999). Optimizing liposomes for delivery of chemotherapeutic agents to solid tumors. *Pharmacol. Rev.* 51:691–743.
92. Moghimi S M, Hunter A C, Murray J C (2001). Long-circulating and target-specific nanoparticles: Theory to practice. *Pharmacol. Rev.* 53:283–318.
93. Senior J, Delgado C, Fisher D, et al. (1991). Influence of surface hydrophilicity of liposomes on their interaction with plasma protein and clearance from the circulation: studies with poly(ethylene glycol)-coated vesicles. *Biochim. Biophys. Acta.* 1062:77–82.
94. Torchilin V P, Shtilman M I, Trubetskoy V S, et al. (1994). Amphiphilic vinyl polymers effectively prolong liposome circulation time in vivo. *Biochim. Biophys. Acta.* 1195:181–184.
95. Takeuchi H, Toyoda T, Toyobuku H, et al. (1996). Improved stability of doxorubicin-loaded liposomes by polymer coating. *Proc. Int. Symp. Control Rel. Bioact. Mater.* 23:409–410.
96. Torchilin V P, Trubetskoy V S, Whiteman K R, et al. (1995). New synthetic amphiphilic polymers for steric protection of liposomes in vivo. *J. Pharm. Sci.* 84:1049–1053.
97. Woodle M C, Engbers C M, Zalipsky S (1994). New amphiphathic polymer-lipid conjugates forming long-circulating reticuloendothelial system-evading liposomes. *Bioconjug. Chem.* 5:493–496.
98. Zalipsky S, Hansen C B, Oaks J M, et al. (1996). Evaluation of blood clearance rates and biodistribution of poly(2-oxazoline)-grafted liposomes. *J. Pharm. Sci.* 85:133–137.

# UNEXPECTED BENEFITS OF A FORMULATION: CASE STUDY WITH ERYTHROPOETIN

DAVID KEAST

*Parkwood Hospital, St. Joseph's Health Care, London, Ontario, Canada; University of Western Ontario, London, Ontario, Canada*

**Chapter Contents**

1 Introduction
2 Erythropoietin as a Potential Growth Factor in Wound Healing
3 Clinical Experience with Recombinant Human Erythropoietin and Patients with Pressure Ulcers
4 Conclusions and Future Research
References

## 1 INTRODUCTION

Recombinant human erythropoietin (rHuEPO) is a purified glycoprotein produced by recombinant DNA technology. It is indistinguishable from human urinary erythropoietin in its biologic and immunologic activity. It is commercially available and has absolute indications for the treatment of the anemia associated with chronic renal failure, prematurity, and platinum-based chemotherapy. Other uses have included the anemia associated with multiple myeloma, cancer, myelodysplasia, HIV infection, and other forms of chemotherapy. It has been used to potentiate preoperative autologous blood donation. Less common uses have been in the treatment of the anemia of chronic disease, for perisurgical augmentation and after allogenic bone marrow transplantation [1].

Endogenous erythropoietin (EPO) is produced primarily by the peritubular capillary endothelium of the kidney in response to tissue hypoxia or anemia. EPO is a

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

hormone that plays a crucial role in the regulation of hematopoiesis and induces the proliferation, maturation, and differentiation of erythroid (red blood cell) precursors. In renal failure, the nonfunctioning kidneys fail to produce adequate amounts of endogenous EPO to support erythropoiesis with resultant anemia. Before the introduction of exogenous erythropoietin in the 1980s, patients in renal failure required repeated blood transfusions.

After 1998 there was a significant increase in the number of cases of pure red cell aplasia (PRCA) associated with subcutaneous injections of rHuEPO in the renal dialysis population. PRCA is a relatively rare condition. It presents as a sudden onset of severe isolated anemia characterized by near-complete absence of erythroid precursors in the bone marrow [2]. Most of the initial reports in the literature before 1998 involved isolated cases, of which 50% had no known cause. After 1998, the PRCA was shown to be related to neutralizing antibodies to erythropoietin that cross-reacted with endogenous EPO [3]. Most of these reported antibody-mediated cases were for patients using Eprex (Ortho Biotech, a division of Janssen-Cilag, Bridgewater, NJ). In total between January 1, 1989 and June 30, 2004, the company reported a total of 206 cases, of which 155 occurred in the peak years of 2001–2003 [4]. The mean time from initial exposure to onset of PRCA was 9.1 months [5].

The rise in cases of PRCA corresponded with the introduction of polysorbate 80 stabilized Eprex in prefilled syringes with uncoated rubber stoppers, which was administered subcutaneously. Previously human serum albumin (HSA) was used to stabilize the product, but this practice was discontinued in several countries due to fears of disease transmission by HSA. Epidemiologic and immunologic data reported by Boven et al. [4] demonstrated that the rise in PRCA was most likely related to leachates from the uncoated stoppers in the presence of the polysorbate 80 stabilizer. These leachates may have potentiated the immunogenicity of the product when administered subcutaneously.

No cases for intravenous administration were reported. Since the introduction of coated stoppers in the prefilled syringes in April 2003 and a concomitant switch to intravenous administration in many dialysis units, the number of reported cases fell from 71 in 2003 to 2 reported cases in the first 4 months of 2004. Even at its peak, the incidence rate for PRCA was relatively low at 4.61/10,000 patient years of exposure to the product.

## 2 ERYTHROPOIETIN AS A POTENTIAL GROWTH FACTOR IN WOUND HEALING

In addition to its ability to stimulate erythropoiesis, EPO enhances cell phagocytosis and reduces macrophage activation, therefore modulating the inflammatory process. These anti-inflammatory effects of EPO may be able to reverse the chronic inflammatory condition that is believed to underlie chronic skin ulcers. By interfering with the chronic inflammatory process EPO may help reduce inflammatory cytokines and degradative enzymes that interfere with many wound healing processes and limit new tissue growth. Recent literature suggests that many chronic wounds persist because of inflammatory processes that alter the wound environment. Analysis of wound fluid taken from healing and chronic wounds revealed

that chronic wounds have elevated levels of inflammatory mediators and degradative enzymes that interfere with the healing process [6]. Healing wounds are characterized by high mitogenic potential, rapid cellular migration, balanced inflammatory cytokines, low proteases, and good cellular response to growth factors. Chronic wounds, however, are characterized by poor mitogenic potential and cell migration, high proteases and inflammatory cytokines, and senescent cells unresponsive to growth factors [7]. EPO's effects on the wound healing process not only include restoring the normal wound environment, but it also has recently been found to interact with vascular endothelial growth factor (VEGF). Together EPO and VEGF stimulate endothelial cell mitosis and motility important in new vessel growth and wound healing [8]. Galeano et al. [9] demonstrated in an artificial wound model in genetically diabetic mice that erythropoietin injections increased VEGF mRNA expression, wound protein content, wound healing, and wound breaking strength. The additional benefits of EPO are its neurotrophic and neuroprotective functions that may ameliorate neurological damage after spinal cord and brain injury [10]. A specific EPO/EPO receptor system has been found in the central nervous system and in the cerebrospinal fluid, which is independent of the heatopoietic system [11].

Chronic skin ulcers experienced by patients who have low concentrations of circulating hemoglobin (≤100 g/L) may be difficult to heal because of impairment in tissue oxygenation. There are various types of anemias with differing underlying causes, including nutrient deficiencies, such as iron deficiency anemia, and anemia as the result of chronic disease or chronic inflammatory processes. Iron deficiency anemia results from inadequate intake, absorption, or utilization of iron and/or acute or chronic blood loss [12, 13]. It is characterized by low hemoglobin and other hematological changes combined with microcytic, hypochromic red blood cells. There are low levels of stored iron as indicated by low serum ferritin values. Anemia of chronic disease (ACD) is characterized by low hemoglobin concentrations and other hematological changes with normocytic, normochromic red blood cells. In anemia of chronic disease, red blood cell production is impaired and there may be a shortened red blood cell life span. ACD is also characterized by a normal or elevated serum ferritin level. ACD is generally not an irondeficiency anemia, and it is refractory to an iron-enhanced diet, iron replacement therapy, either oral or intravenous, and transfusion. In fact, iron supplementation and transfusion are generally contraindicated in ACD because of the risk of iatrogenic hemochromatosis. ACD occurs in individuals with a chronic inflammatory process such as arthritis, critical illness, or chronic skin ulcers. It is thought to be the result of impaired responsiveness of erythroid progenitor cells due to a persistent elevated level of circulating inflammatory cytokines that are known to occur in patients with chronic inflammatory conditions [14]. Ferrucci et al. [15] demonstrated in a study of inflammatory serum markers in a sample of the residents of Chianti Italy that the anemia of inflammation evolved from a pre-anemic state in which normal hemoglobin was maintained in persons with high levels of inflammatory markers by increasing levels of erythropoietin to a clinical anemia in which erythropoietin levels were suppressed possibly through the inhibitory effect of inflammation on erythropoietin production.

However, despite the numerous documented effects of EPO on cellular and physiological events known to be important in the healing process and the

management of ACD, few clinical studies are investigating the use of EPO to promote the healing of chronic wounds. In 1992 Turba et al. reported the successful treatment of ACD related to stage IV pressure ulcers [16]. In another hematological disorder, treatment with EPO resulted in the rapid and complete healing of chronic leg ulcer, further improvement in hematological parameters, and relief from chronic pain [17].

## 3 CLINICAL EXPERIENCE WITH RECOMBINANT HUMAN ERYTHROPOIETIN AND PATIENTS WITH PRESSURE ULCERS

Spinal cord-injured patients are at high risk of developing pressure ulcers, many of which often become associated with anemia of chronic disease. These ulcers are challenging to heal. Pressure redistribution, nutritional support, management of incontinence, and good local wound care are key components for healing these ulcers. Despite best-practice care, patients with anemia of chronic disease remain difficult to heal.

In an attempt to reverse the anemia of chronic disease, several patients in the Spinal Cord Injury rehabilitation unit at Parkwood Hospital, St. Joseph's Health Care, London, Canada, were treated with subcutaneous erythropoietin. A retrospective chart audit was conducted to review the effectiveness of 6 weeks of subcutaneous erythropoietin 75 IU/kg subcutaneously 3 times weekly in resolving refractory anemia of chronic disease and healing stage IV pressure ulcers [18]. The mean age of the patients was 59, and all had a stage IV pressure ulcer.

All patients received pressure off-loading, nutritional assessment and supplementation, and best-practice local wound care. Comorbid conditions such as diabetes were medically managed. All patients received rHuEPO 75 IU/kg subcutaneously three times weekly for 6 weeks. Iron supple-mentation with oral ferrous gluconate as indicated by ferritin levels was administered.

Mean initial hemoglobin was 88 g/L. After 6 weeks of rHuEPO injections, the mean hemoglobin for the four patients rose to 110 g/L. The individual results are shown in Figure 1. The mean number of ulcers decreased from 3 to 2.3, and the

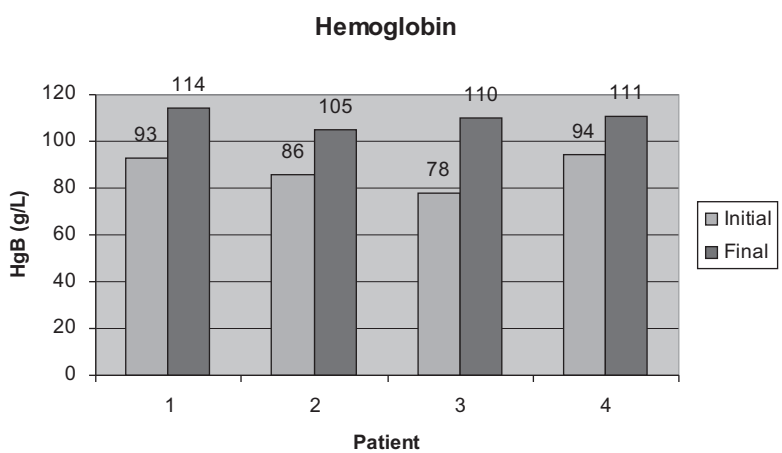


**Figure 1.** Comparison of initial and final hemoglobin after 6 weeks rHuEPO.

mean surface area of the largest ulcer decreased from 42.3 $cm^2$ to 38.4 $cm^2$. Ulcer depth was decreased by half. The extent of undermining improved in all ulcers. Some patients showed an increased ability to fight intercurrent infections. All patients felt more energetic and were better able to participate in their rehabilitation activities. No adverse effects were observed.

The results were promising enough to suggest that a prospective study is warranted. Such a study would carefully collect all hematologic factors as well as use validated tools for determining ulcer size and appearance. In addition, collecting wound fluid before, during, and after treatment to determine the effect on chronic inflammatory mediators would be useful. Research into effects at the molecular level may be useful.

## 4 CONCLUSIONS AND FUTURE RESEARCH

The introduction of human recombinant erythropoietin significantly improved the quality of life for patients on real dialysis by managing the concomitant anemia, which results from end-stage renal failure. It has also been successfully employed in several other conditions in which erythropoiesis is suppressed either through disease processes or as a result of treatments.

The emergence of pure red cell aplasia as a complication of treatment with subcutaneous injection in the late 1990s was probably the result of the use of prefilled syringes with uncoated rubber stoppers in which the product was stabilized with polysorbate 80. Leachates from the stoppers most likely potentiated the immunogenicity of the product leading to the production of neutralizing antibodies. Since the reformulation of the product, the incidence of PRCA has been steadily dropping.

Human recombinant erythropoietin shows promise in resolving the refractory anemia of chronic disease associated with stage IV pressure ulcers, and further study is suggested. The results may also suggest that rHuEPO acts as a growth factor either alone or in conjunction with intrinsic factors in the wound. Studies of its role at the molecular level are indicated. Further studies of the neurotrophic and neuroprotective properties observed in rat models hold promise in the area of spinal cord injury research.

## REFERENCES

1. Hoffman R (2005). Erythropetic growth factors. In R Hoffman, et al. (eds.), *Hematology Basic Principles and Practice*, 4th ed.: Churchill Livingstone, New York.
2. Dessypris E N (2005). Pure red cell aphasia. In R Hoffman, et al. (eds.), *Hematology Basic Principles and Practice*, 4th ed.: Churchill Livingstone, New York.
3. Casadevall N (2002). Antibodies against rHuRPO: Native and recombinant. *Nephrol. Dial. Transplant.* 17(suppl 5):42–47.
4. Boven K, Stryker S, Knight J, et al. (2005). The increased incidence of pure red call aphasia with Eprex formulation in uncoated stopper syringes. *Kidney Internat.* 67:2346–2353.

5. Bennett C, Luminari S, Nissenson A, et al. (2004). Pure red-cell aphasia and recombinant epoetin therapy—a follow-up report from the research on adverse drug events and reports project. *N. Engl. J. Med.* 351:1403–1408.
6. Mast B, Schultz G (1996). Interactions of cytokines, growth factors, and proteases in acute and chronic wounds. *Wound Rep. Reg.* 4:411–420.
7. Schultz G S, Sibbald R G, Falanga V, et al. (2003). Wound bed preparation: A systematic approach to wound management. *Wound Rep. Reg.* 11(suppl 2):1–28.
8. Buemi M, Aloisi C, Cavallaro E, et al. (2002). Recombinant human erythropoietin (rHuEPO): More than just the correction of uremic anemia. *J. Nephrol.* 15(2): 97–103.
9. Galeano M, Altavilla D, Cucinotta D, et al. (2004). Recombinant human erythropoietin stimulates angiogenesis and wound healing in the genetically diabetic mouse. *Diabetes.* 53(9):2509–2517.
10. Gorio A, Gokmen N, Erbayraktar S, et al. (2002). Recombinant human erythropoietin conteracts secondary injury and markedly enhances neurological recovery from experiment spinal cord injury. *Proc. Natl. Acad. Sci. U.S.A.* 99(14):9450–9455.
11. Buemi M, Cavallaro E, Floccari F, et al. (2002). Erythropoietin and the brain: from neurodevelopment to neuroprotection. *Clin. Sci. (London)* 103(3):275–282.
12. Woloschuk D M M (1997). A practical guide to anemias. *Pharm. Practice.* Aug 1–8.
13. Simmons-Holcomb S (2001). Anemia—pointing the way to a deeper problem. *Nursing* 31(7):36–42.
14. Spivak J L (2002). Iron and anemia of chronic disease. *Oncology.* 9(suppl 10):25–33.
15. Ferricci L, Guaralnik J M, Woodman R C, et al. (2005). Proinflammatory state and circulating erythropoietin in persons without anemia. *Amer. J. Med.* 118(11):1288e11–1288e19.
16. Turba R M, Lewis V L, Green D (1992). Pressure sore anemia: Response to erythropoietin. *Arch. Phys. Rehab. Rehabil.* 73(5):498–500.
17. Al-Momen A K (1991). Recombinant erythropoietin induced rapid healing of a chronic leg ulcer in a patient with sickle cell disease. *Acta. Haematol.* 86(1):46–48.
18. Keast D H, Fraser C (2004). The treatment of Chronic Skin Ulcers in individuals with anemia of chronic disease using recombinant human erythropoietin (EPO): A review of four cases. *Ostomy/Wound Managem.* 50(10):38–47.

# CAPILLARY SEPARATION TECHNIQUES

Dong Hee Na[1] and Kang Choon Lee[2]

[1] *College of Pharmacy, Kyungsung University, Nam-ku, Busan, Korea*
[2] *Drug Targeting Laboratory, College of Pharmacy, SungKyunKwan University, Jangan-ku, Suwon, Korea*

**Chapter Contents**

1 Introduction
2 Principles of Capillary Electrophoresis
  2.1 Instrumentation
  2.2 Theory
  2.3 Separation Modes
  2.4 Detection
3 Applications
  3.1 Identity Determination
  3.2 Purity Control
  3.3 Heterogeneity Characterization of Glycoproteins
  3.4 Stability Study
  3.5 Characterization of PEGylated Peptides and Proteins
  3.6 Proteome Analysis
4 Conclusions
  References

## 1 INTRODUCTION

Capillary electrophoresis (CE) is a modern analytical method that is being extensively applied to the characterization of biotechnology-derived products like peptides and proteins [1]. Due to its ease of automation and facilitating the development of reproducible routine analysis, CE seems to be well suited for the quality control of biotechnological products, including process monitoring, purity assessments,

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

and stability studies [2]. Moreover, the high-resolution capacity of CE offers great potential for the analysis of heterogeneous protein products such as glycoproteins and polymer-conjugated proteins [3, 4]. Recently, the importance of CE in protein analysis is enhanced, with efforts being made to learn more about the compositions and functions of proteins [5]. As the field of proteomics becomes more important, the thousands of new proteins and peptides will be discovered, which will lead to the developments of many new pharmaceutical drugs.

Recently, CE has emerged as a powerful tool because it has many advantages over the conventional protein separation techniques such as polyacrylamide gel electrophoresis (PAGE) and high-performance liquid chromatography (HPLC). Compared with PAGE, CE is faster, easier, simpler, quantitative, and has automation capability. Like HPLC, CE has various different separation modes and is applicable to a wide range of analytes. CE requires a very small sample amount (nanoliters) and limited quantities of reagents (microliters of buffer), compared with HPLC that requires microliters of sample and milliliters of solvent. The main advantage of CE over PAGE and HPLC is the ability to produce a higher number of theoretical plates. The efficient separations result from the application of high electric fields (100–500 V/cm) and flat flow generated by the electro-osmotic flow (HPLC generates the pressure-driven parabolic flow). The high electrical resistance of a capillary and its favorable surface area-to-volume ratio, permitting efficient heat dissipation, enables the application of high electric fields without causing detrimental heat generation. Furthermore, the mild separation conditions of CE, which use mostly aqueous buffer solutions, present the compatibility to protein studies in native state, which can be important especially in the characterization of pharmaceutical proteins. The main disadvantage of CE is the low concentration limits of detection due to the short path length and the limited introduction of sample volume. The enhancements in this issue have been achieved by the advent of several preconcentration methods [6] and highly sensitive detection methods such as laser-induced fluorescence and mass spectrometry [7].

This chapter introduces the fundamental principles of CE and its application to the separation and characterization of peptides and proteins. One major characteristic of CE is the availability of various separation modes with different separation mechanisms based on the differences in charge-to-mass ratio (capillary zone electrophoresis), molecular size (capillary gel electrophoresis), isoelectric point (capillary isoelectric focusing), and hydrophobicity (micellar electrokinetic capillary chromatography and capillary electrochromatography). In the principles of CE, the instrumentation, theory, separation modes, and detection methods are described focusing on the analysis of peptides and proteins. In applications, CE methods for identification, purity assessment, heterogeneity characterization, and stability of peptide and protein products are highlighted with the characterization of Poly(ethylene glycol) (PEG)ylated biomolecules. Finally, the role of CE in proteomics research is discussed with the recent approaches.

## 2 PRINCIPLES OF CAPILLARY ELECTROPHORESIS

### 2.1 Instrumentation

The main components of a CE system are a high-voltage power supply, a capillary, inlet and outlet buffer vials, a detector, and a data output and handling device such

as an integrator or computer (Figure 1A). A variety of commercial instruments are available with different capabilities, including injection methods, detectors, capillary cooling systems, and software [8]. CE is performed by filling inlet and outlet vials and the capillary with an electrolyte, usually an aqueous buffer solution. Electrodes connected to a high-voltage power supply are then immersed in the vials. Very small amounts of sample are introduced into one end of the capillary, and these are driven electrophoretically down the lumen of the capillary toward the opposite electrode. The species that migrate through the capillary are detected by an online optical detector near the capillary outlet, and these data are displayed as an electropherogram, in which separated compounds appear as peaks with different migration times.

Fused silica capillaries of length of 20–100 cm and inner diameters of 50–100 μm are typically used. The outer surface of capillary is coated with polyimide, which is strongly ultraviolet (UV) absorbent. A detection window can be made by simply burning or scraping off a small section of the polyimide outer capillary coating. This section of the capillary is then placed in the light path of the detector, and thus solutes are detected while in the capillary. Most commercial instruments have a capillary cartridge, which retains the capillary, provides mechanical support, and allows the capillary to be consistently aligned at the optical center of the detector. Cartridges are used in conjunction with cooling systems, which maintain capillary temperatures using coolant or cooled air.

Only a few microliters of sample are required for CE because usually only 1–50 nanoliters of sample are injected into a capillary. For example, a 50-μm inner diameter capillary 50 cm long has a volume of only ca. 1 μL [9]. Samples are injected

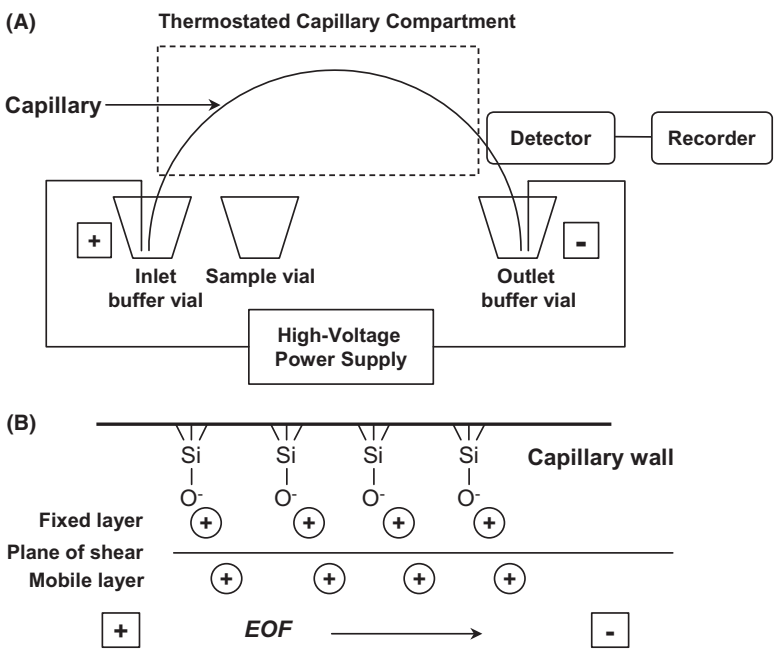


**Figure 1.** Schematic diagram of a capillary electrophoresis system (A) and electro-osmotic flow (EOF) (B).

into capillaries using different techniques, such as hydrodynamically or electrokinetically. Hydrodynamic injection can be performed by pressure or siphoning. Pressure injection is performed by either pressurizing the sample vial or by applying a vacuum to the outlet vial. Siphon injection is performed by raising the sample vial, which causes the sample to be siphoned into the capillary. In electrokinetic injection, an electric field is applied to the sample vial, causing the sample components to migrate into the capillary. Most commercially available instruments have autosamplers into which several sample vials can be loaded. The samples are then automatically injected into capillaries using one or more of the techniques above.

A variety of detectors have been used for CE, including UV absorbance, fluorescence, laser-induced fluorescence, and mass spectrometry. The most widely used is the UV absorbance detector. In some cases, two or more detectors are connected in series. A more detailed the description of the detection system is given in Section 2.4.

The power supply provides an electric field across the capillary with voltages up to 30 kV, currents up to 300 μA, and power up to 6 W. Most instruments can be operated in either constant voltage, constant current, or constant power mode and have a reverse applied polarity facility. The constant voltage mode is most commonly used. It is necessary to have a stable voltage, as any variations in voltage will cause changes in migration times.

Most CE systems are controlled by an external computer that controls all instrumental functions. The operating parameters for each analysis are programmed by computer. The electropherograms obtained are plots of detector response versus time; thus, they resemble familiar HPLC or gas chromatography (GC) chromatograms, which means that familiar data handling systems can be used. In addition, electronic integrators or computers are used for qualitative identification and quantitation, as in HPLC or GC.

### 2.2.2 Theory

Electrophoresis is the phenomenon whereby ionic species in a conductive aqueous medium moves under the influence of an electric field. Thus, ionic molecules are separated due to differences in their electrophoretic mobilities. The electrophoretic mobility ($\mu$) of a spherical ion is given by

$$\mu = q/6\pi\eta r$$

where $q$ is the charge of the ion, $\eta$ is the viscosity of the solution, and $r$ is the hydrodynamic radius of the ion. This equation demonstrates the positive relation between charge-to-mass ratio ($q/r$) and electrophoretic mobility. The electrophoretic migration velocity ($\nu$) depends on the electrophoretic mobility ($\mu$) and the applied electric field ($E$):

$$\nu = \mu E$$

Besides the electrophoretic migration, a fundamental electrophoretic phenomenon occurring in CE is the electro-osmotic flow (EOF), which essentially is an electrical field-driven bulk solution flow from the anode to the cathode (Figure 1B). This flow occurs because acidic silanol groups on the inside of fused-silica

capillary are ionized when in contact with buffer solution. At pH above 3, these silanol groups are deprotonated and form an electric double layer. When an electric field is applied, the net positively charged solution in a capillary moves toward the cathode. EOF is highly dependent on buffer pH; i.e., it increases on raising pH and plateaus at about pH 8, but it is not significant below pH 4. At neutral and higher pH, the EOF is sufficiently strong to allow all molecules, regardless of charge, to move toward the outlet and pass the detector. EOF can be effectively controlled by changing several experimental conditions, including separation buffer pH, ionic strength, addition of organic solvents, and buffer additives. The apparent migration velocity of the analyte depends on their electrophoretic mobility ($\mu_e$), electroosmotic mobility ($\mu_{eo}$), the applied voltage ($V$), and capillary length ($L$):

$$\nu = (\mu_e + \mu_{eo})V/L$$

The electrophoretic mobility of peptides and proteins was first described by Offord [10], who proposed the following equation, which relates the mobility ($\mu$) with valence (Z) and molecular mass (M):

$$\mu = kZM^{-2/3}$$

where $k$ is an empirical constant. This equation states that frictional forces opposing electrophoretic migration are proportional to the surface areas of the species concerned. This implies that the electrophoretic mobility would be proportional to $1/r^2$ ($1/M^{2/3}$), rather than $1/r$ ($1/M^{1/3}$), which is suggested by the Stoke's model. This relationship was later confirmed and modified by several researchers in the CE field [11, 12]. Janini et al. presented the electrophoretic mobilities of 58 peptides that varied in size from 2 to 39 amino acids and in charge from 0.65 to 7.82 [13]. Figure 2 shows CE separation of several peptides in a 50-m*M* phosphate buffer at

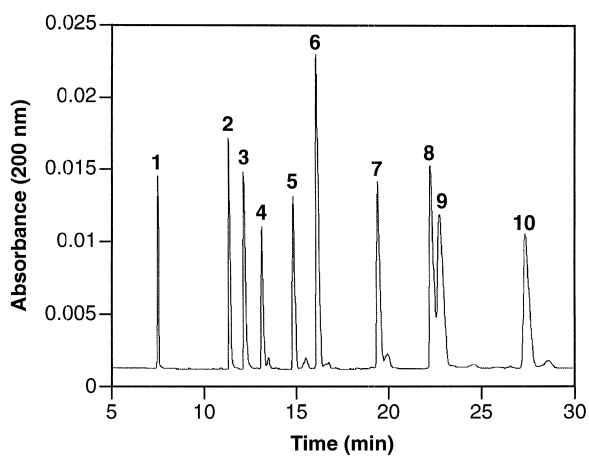


**Figure 2.** Electropherogram of a mixture of bioactive peptides separated at pH 2.5. Peptides: (1) reference; (2) bradykinin; (3) bradykinin fragment 1–5; (4)substance P; (5) [arg]–vasopressin; (6) luteinizing hormone releasing hormone; (7) bombesin; (8) leucine enkephalin; (9) methionine enkephalin; (10) oxytocin. (Reprinted from Ref. 13, with permission.)

pH 2.5. The electrophoretic data were used to test existing theoretical models that correlate electrophoretic mobility with physical parameters. The best fit of the experimental data was obtained with the Offord model that correlates electrophoretic mobility with the charge-to-size parameter $q/M^{2/3}$.

## 2.3 Separation Modes

As HPLC has different separation modes of chromatography, including adsorption, partition (normal and reversed phase), ion-exchange, size-exclusion, and affinity, CE also has various separation modes, e.g., capillary zone electrophoresis (CZE), capillary gel electrophoresis (CGE), capillary isoelectric focusing (CIEF), micellar electrokinetic capillary chromatography (MEKC), and capillary electrochromatography (CEC) (Table 1). Moreover, these different CE separation modes can be used to complement each other, and thus, they greatly enhance the versatility of the technique. In many cases, it seems that no single method can separate peptides and proteins because these biomolecules are diverse and complex in terms of structure and composition. Therefore, different optimal strategies involving different CE separation modes are often used to solve the separation problems.

***Capillary Zone Electrophoresis (CZE).*** CZE is a widely used CE technique and separates peptides and proteins based on differences in their charge-to-mass ratios. Separations occur in a capillary filled with a buffer of constant composition. For CZE, the run buffer choice is extremely important because it determines the charge on the analyte molecule and its migration rate. Thus, the type of buffer, its ionic strength, and its pH are optimized for particular separation problems. Buffers based on sodium phosphate, citrate, acetate, or combinations thereof with concentrations ranging from 10 to 200 mM are frequently used [14].

**TABLE 1. Separation Modes of CE for the Characterization of Peptides and Proteins**

| Technique | Separation Mechanism | Applications | Conventional Technique |
|---|---|---|---|
| CZE | Charge-to-mass ratios | Peptide mapping, purity and stability of peptides and proteins, and glycoprotein heterogeneity | — |
| SDS–CGE | Molecular size | Estimation of relative molecular mass of protein (10–200 kDa) and purity and stability of peptides and proteins | SDS–PAGE |
| CIEF | Isoelectric point (p*I*) | p*I* determination of proteins, purity of proteins, and glycoprotein heterogeneity | Gel IEF |
| MEKC | Hydrophobicity | Purity of peptides and glycoprotein heterogeneity | RP–HPLC |
| CEC | Hydrophobicity | Peptide mapping and purity of peptides | RP–HPLC |

Proteins contain both positively and negatively charged functional groups. The positively charged moieties in proteins, such as the guanidinium group of arginine residues, the amino groups of lysine residues and N-termini, and histidine residues, interact with negatively charged silanol groups on capillary walls at pH values above 3. These interactions between proteins and capillary walls (protein adsorption) lead to sample loss, peak broadening, poor resolution, and longer migration times.

Many strategies devised to overcome protein–wall interactions have focused either on selecting separation buffer conditions that reduce protein binding sites or on treating capillary surfaces to reduce interaction sites [15]. By using separation buffers with extreme pH values (e.g., below pH 2~3 or above pH 10), protein adsorption can sometimes be reduced. At pH values below around 2, the silanol groups of the capillary are fully protonated and surface charge approaches zero, whereas at pH values above 10, capillary surfaces are completely deprotonated and highly anionic. In these situations, protein adsorption is significantly reduced because of electrostatic repulsion. However, operations at extreme pH conditions can be problematic in finding suitable separation conditions due to reduction of differences of analytes in effective electrophoretic mobility. At low pH, EOF is negligible, making separations slow, and at high pH, EOF is generally high and resolutions are reduced. Additional problems encountered under extreme pH conditions include the effect on protein conformational changes. Unfolding or aggregation of proteins may occur, resulting in irreproducible data, often reduced efficiency, and a loss of biological activity. Therefore, the use of this approach is limited to a few selected applications.

To optimize separation conditions at moderate pH conditions, surface modifications of capillary can be helpful [16]. Such modifications can decrease protein–surface interactions by reducing the surface charge of the silica material. Several approaches have been proposed for modifying the inner walls of capillaries, such as chemical coating with bonded/cross-linked polymers, physical adsorption coating with cationic polymers, and dynamic coating with adsorbed surfactants. The most popular coating method is to covalently bind polymers such as polyacrylamide or poly(vinyl alcohol) to capillary walls [17]. Chemically coated capillaries are widely used and are available commercially. The preparations of this capillary type often involve several chemical reactions that may cause large variances between capillaries. In addition, these coatings are usually stable in restricted pH ranges and have limited lifetimes. Noncovalent capillary coatings are simply formed by treating the capillary with one or more charged polymers that are physically adsorbed onto internal walls. Initially, positively charged monolayer coatings of polybrene [18], polyethyleneimine [19], or poly(diallyldimethylammonium chloride) [20] were used. More recently, noncovalently bilayer-coated capillaries prepared by add anionic polymer to cationic polymer-coated capillary have been reported to be long-lived and to have an EOF that is stable over wide pH ranges (2–11) [21]. Katayama et al. reported on the use of a polymeric bilayer of polybrene and dextran sulfate for the separation of some model proteins [22]. Catai et al. demonstrated the use of capillaries coated with a bilayer of polybrene and poly(vinyl sulfonate) for the fast, highly reproducible, and efficient analysis of peptides and proteins [23, 24]. Another way of coating capillaries is to add the coating agents (usually low-molecular-weight compounds such as amines and surfactants) to the separation buffer for the dynamic coating of the inner capillary surface [25].

***Sodium Dodecyl Sulfate-Capillary Gel Electrophoresis (SDS–CGE).*** SDS–CGE is a capillary-based version of SDS–polyacrylamide gel electrophoresis (SDS–PAGE) in the slab gel format, with advantages of shorter analysis times, ease of automation, and online detection and quantitation [26, 27]. In SDS–CGE, replaceable sieving polymers, such as linear polyacrylamide, poly(ethylene oxide), dextran, or pullulan, are used to achieve reproducible separations. These polymers permit the replacement of a separation matrix for each sample, thereby eliminating cross-contamination between samples and improving reproducibility. Best results are often obtained using chemically or dynamically coated capillaries.

For SDS–CGE, proteins are denatured and complexed with SDS before analysis. These SDS–protein complexes are then separated based on their sizes. The molecular mass of an unknown protein can be estimated by running protein standards as well. This assumes that migration is dependent on relative mass, on the basis of the following two assumptions. First, all SDS–protein complexes have the same charge-to-mass ratio in the presence of excess SDS (~1.4 g of SDS/1 g of protein). Second, the SDS–protein complexes have similar shapes, and thus, their sizes are linearly related to their molecular masses. However, these assumptions are not valid for some classes of proteins, e.g., basic proteins, hydrophobic membrane proteins, and glycoproteins, because deviations from predicted charge-to-mass ratio are often observed for their SDS–protein complexes. For example, basic proteins have lower charge-to-mass ratios because of the presence of positively charged amino acids, whereas hydrophobic membrane proteins have larger charge-to-mass ratios. Moreover, in the case of glycoproteins, the carbohydrate moieties do not bind with SDS, and this lowers the charge-to-mass ratio, leading to a decreased migration and overestimation of molecular mass [3, 28]. Figure 3 shows the SDS–CGE result for ricin toxin purified from the seeds of the castor bean (*Ricinus communis*) [29], in which two peaks of ricin toxin were partially separated, while only a single band appeared in the SDS–PAGE gel. Two peaks were identified to be ricin glycoform isomers with identical protein sequences but with different carbohydrate contents. When the molecular masses of two peaks were determined with a calibration curve made by protein standards (MW 14–200 kDa), they were significantly higher than those measured by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI–TOF MS). This shows the slower electrophoretic migration of glycoproteins in SDS-based gel electrophoresis compared with that of standard proteins. However, the presence of carbohydrate may improve the resolution between glycoforms in SDS–CGE [29].

In addition to the estimation of protein molecular mass, SDS–CGE is also the common choice for checking the presence of other proteins, protein aggregates, or protein degradation products during characterization, purity, or stability studies.

***Capillary Isoelectric Focusing (CIEF).*** CIEF separates peptides and proteins according to isoelectric point (p*I*) differences. Proteins that differ by 0.005 p*I* units or even less have successfully been separated by CIEF. The sample is normally mixed with ampholytes (zwitterionic compounds), which have slightly different p*I* values spanning the desired pH range and act as a strong buffer at their p*I* values and then the capillary is filled with this mixture. The capillary inlet is placed in a vial containing acidic solution (anolyte) and the capillary outlet in a basic solution (catholyte), before applying an electric field. A pH gradient is quickly formed by

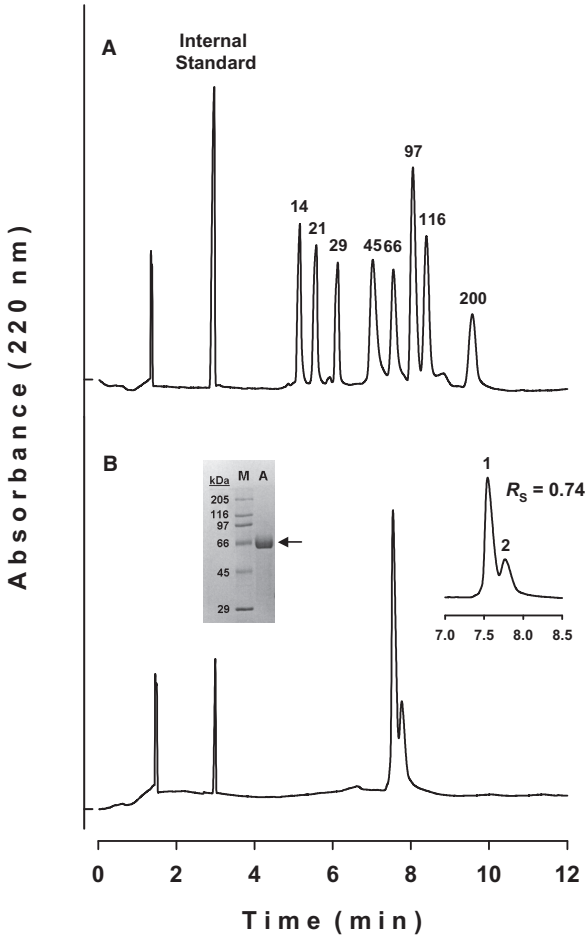


**Figure 3.** SDS–CGE electropherograms of molecular weight standard marker (14–200 kDa) (A) and ricin toxin glycoprotein under nonreducing conditions (B). The right insert is an enlargement of ricin peaks in SDS–CGE, presenting the partially separated two peaks ($R_s$ = 0.74) compared with only a single band shown in the SDS–PAGE gel. (Reprinted from Ref. 29, with permission.)

the ampholytes ranging from low to high pH along the entire capillary length. According to their electrophoretic mobilities, proteins migrate to their p*I*s in the capillary where they are focused into very sharp zones. The overall focusing process can be monitored by observing current changes, which approach zero when ion movement inside a capillary stops. The CIEF process can be performed in one or two steps. One-step CIEF is performed in an uncoated capillary and usually employs polymers such as hydroxypropyl methyl cellulose or hydroxyl ethyl cellulose in order to reduce but not eliminate EOF [30]. At this point, either migration past a detection window or whole-column imaging is needed to visualize the separation [31]. In two-step CIEF, proteins are first focused in a coated capillary to eliminate EOF and then mobilized by adding a salt (sodium chloride) to the catholyte or by applying pressure to the capillary inlet (anolyte) [32, 33]. In general, the two-step

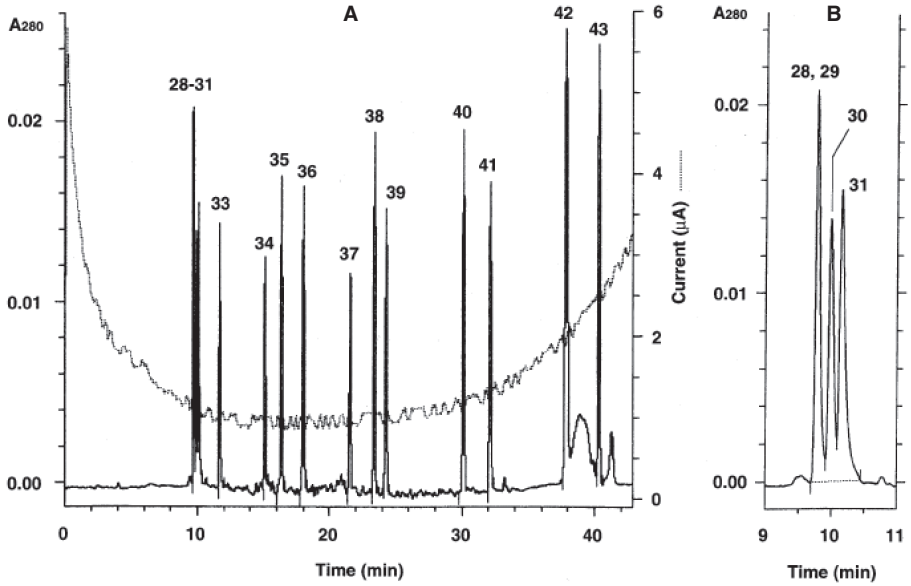


**Figure 4.** CIEF separation of the peptide p*I* markers ranging from pH 3.38 (peptide #43) to 10.17 (peptide #28). (B) Enlargement of the portion from 9 to 11 min of (A). (Reprinted from Ref. 35, with permission.)

method is preferred because it provides a much higher reproducibility [34]. As CIEF is usually run from positive to negative polarity, the most basic proteins pass the detector first. Absorbance must be monitored at UV 280 nm to avoid the strong absorption of ampholytes at lower wavelengths. The use of a coated capillary is required for high resolution because EOF interferes with maintenance of focused zones.

The estimation of p*I* in CIEF needs appropriate p*I* markers. Synthetic peptide and oligopeptide p*I* markers for CIEF with UV absorption detection have been developed [34]. Shimura et al. proposed a set of 16 synthetic oligopeptides as p*I* markers for CIEF that are fully compatible with UV detection [35]. Figure 4A shows the CIEF separation of 16 p*I* markers ranging from pH 3.38 (peptide number 43) to 10.17 (peptide number 28). Figure 4B is an enlargement of the region between 9 and 11 min, showing partial resolution of high p*I* markers and the overlap between peptide numbers 28 and 29 due to a lack of extension of the pH gradient toward more alkaline pH values. CIEF produces the best resolution for proteins and peptides among the CE separation modes, and it is a powerful tool for resolving modified proteins, for characterizing microheterogeneity, and for identifying glycoforms [36]. CIEF has also been found to be a powerful tool in proteomics studies [37].

***Micellar Electrokinetic Capillary Chromatography (MEKC).*** MEKC separation is based on the partition of analytes between micelles and the surrounding aqueous phase [38]. This technique can be considered a type of chromatography where the stationary phase is essentially mobile and the mobile phase is electro-osmotically

pumped. The micellar phase is composed of a surfactant added to the buffer above its critical micellar concentration (CMC). The most commonly used surfactants are SDS, bile salts, and hydrophobic chain quaternary ammonium salts. In general, neutral or alkaline buffer solutions are used to create a strong EOF that moves even anionic micelles in the capillary toward the cathode.

MEKC has been applied with great success to the analysis of a variety of small molecules. However, relatively few protein applications have been found for MEKC [39, 40], which may be because most proteins (>MW 5000) are too large to partition into the hydrophobic core of micelles. However, proteins can associate with micelles through hydrophobic, hydrophilic, and electrostatic mechanisms, and these interactions have been exploited to manipulate protein separation using MEKC [41]. By manipulating these protein–micelle interactions using the variables, such as surfactant concentration, pH, ionic strength, and the addition of an organic modifier, MEKC has been demonstrated to resolve proteins with minor structural variations and to allow the quantitative analysis of proteins present in complex matrices [3].

***Capillary Electrochromatography (CEC).*** Capillary electrochromatography (CEC) is a hybrid technique of CE and HPLC that is generally carried out using packed capillary columns by the electro-osmostically driven mobile phase at high electric field strength. CEC has been rapidly developed in recent years as it has combined advantages of both techniques, i.e., the high selectivity of HPLC and the high efficiency of CE, with minimum consumption of both reagents and samples, and good compatibility with mass spectrometry [42]. The separation mechanism involved is based on chromatographic retention, and for charged analytes, on a combination of this and electrophoretic mobility. Initially, CEC was mainly used to separate neutral compounds such as polyaromatic hydrocarbons. Recently, the successful development of CEC column technology and improvements in instrumentation have encouraged research on CEC for peptide and protein separations [43]. Various capillary columns have been used for peptide and protein analysis, including packed capillaries, open-tubular capillaries, and monolithic columns.

In addition to the use of octadecylsilane (ODS) as a stationary phase, sulfonated poly(styrene-divinylbenzene), poly(2-sulfoethylaspartamide)-silica, open tubular silica, porous styrenic sorbents and octadecylsilica, pentofluorophenylsilica, trycontysilica, octadecylsilica, acrylate-based porous monoliths, and wide-pore stationary phases have been recently investigated [44, 45]. CEC separations using monolithic columns are being actively investigated because the stationary phases are easily prepared and supporting frits are unnecessary. In monolithic columns, the stationary phase is covalently bound to inner capillary walls. CEC separations of peptides and proteins in monolithic columns are usually performed in counter-directional mode; i.e., peptides and proteins migrate electrophoretically in a direction opposite to the EOF [46, 47].

Recently, a great deal of interest has been shown in the combination of CEC and mass spectrometry (MS) because CEC is considered to overcome many of the limitations of other CE techniques [48–50]. When combined with electrospray ionization (ESI)–MS, CZE is limited to a small number of buffer systems and suffers from low sample loading capacity. MEKC has relatively poor selectivity and

difficult compatibility with MS because it requires the use of high surfactant concentrations. In contrast, CEC has good compatibility with MS, selectivity, sample loading capacity, and general applicability.

## 2.4 Detection

In the CE system, peptides and proteins are typically detected by UV absorbance, laser-induced fluorescence (LIF), or MS. UV absorbance is the most widely used detection method, but when higher sensitivity is required, LIF or MS are preferred. Recently, MS is being increasingly used to detect peptides and proteins separated by CE with the improvement of proteomics research.

***UV Absorbance.*** UV detection of peptides and proteins is mostly performed at 200–220 nm, where the absorption is proportional to the number of peptide bonds, but sometimes around 254 or 280 nm, where the detection is based on the absorbance of the aromatic residues such as tryptophan, tyrosine, and phenylalanine. This detection method is most commonly used, but it has low sensitivity, which is a major disadvantage for the detection of analytes present at low concentrations. In CZE, UV detection is limited to micromolar or submicromolar concentrations for peptides and proteins. Several capillaries have been designed to improve the sensitivity in CE: (1) a rectangular capillary extended in the direction of the light path [51], (2) a Z-shaped capillary [52], and (3) a bubble cell capillary with an locally enlarged diameter in the detection region [53]. These capillaries help somewhat to improve sensitivity, but practical difficulties of availability and implementation into existing instruments remain.

***Laser-induced Fluorescence (LIF).*** Fluorescence is inherently a sensitive detection method with lower detection limits than UV absorbance detection. In addition, fluorescence detectors are selective because only fluorescing molecules are detected. Typical detection limits of fluorescence detection lie in the range $10^{-7}$–$10^{-9}$ M. The light sources used are usually deuterium, tungsten, or xenon lamps. Lasers are also good sources of high-intensity radiation and are used for LIF detection. LIF is the most sensitive detection method in CE. Detection limits of LIF have been reported to fall in the range $10^{-18}$–$10^{-21}$ M [54–56]. A variety of lasers have been employed for the detection of peptides and proteins, which include Nd:YAG, argon ion, helium–cadmium, and helium–neon lasers. As proteins containing aromatic amino acid residues (tryptophan, tyrosine, and phenylalanine) have native fluorescence around 310 nm, the intrinsic fluorescence has been directly measured using Nd: YAG at 266 nm or helium-cadmium at 320 nm as a laser source [57, 58]. However, these systems are expensive and the helium–cadmium lasers have a problem of short lifetimes [59]. The most popular lasers for the detection of peptides and proteins are argon lasers at 488 and 514 nm and helium–neon lasers at 544, 593, and 633 nm that are relatively inexpensive, stable, and compact. However, they are often not suitable for exciting most proteins to induce native fluorescence. To overcome this disadvantage, the labeling of the molecules with a suitable fluorescent dye is required via pre-column, on-column, or post-column. Several fluorescence-labeling reagents are commercially available and conveniently used for peptides and proteins. Rhodamine, solvatochromic dyes (SYPRO Red, Nile Red), and

Albumin Blue 580 have been used to form stable and highly fluorescent complexes with proteins [59].

***Mass Spectrometry (MS).*** MS is an important and powerful detection tool for the characterization of peptides and proteins by CE, as evidenced by recent developments of CE–MS methodology and its applications [60, 61]. MS detection considerably enhances the utility of CE by providing information about the identity of the separated molecules. The availability of CE–MS is also highly desirable for purity and stability studies of peptide and protein drugs. Especially, in proteomics, peptidomics, and peptide mapping, the importance of both on- and offline coupling of MS with CE separations of peptides and proteins is greatly growing [62].

The introduction of soft ionization techniques such as ESI and MALDI brought tremendous progress in on- and offline characterization of electrophoretically separated peptides and proteins by MS [50]. Combination of CE with MS techniques allows not only high-accuracy molecular mass determination of peptides and proteins separated by CE, but also it provides important structural data on amino acid sequence, the sites of posttranslational modifications, peptide mapping, and the noncovalent interactions of peptides and proteins.

ESI forms gaseous ions by applying a strong electric field to a fine spray of their liquid solutions. It allows the generation of multiply charged ions, which enables the detection of very large molecules even with instruments having a low mass range. The multiply charged peaks can be transformed into a singly charged peak by mathematical deconvolution, which enables the determination of the molecular weight of the original species. ESI is the preferred mode for online coupling CE with MS because molecules can be transferred directly from the capillary to the mass spectrometer via an interface. Three types of interfaces, i.e., sheathless, liquid–junction, and coaxial liquid sheath-flow, have been constructed for CE–ESI–MS coupling [63].

In MALDI, the analyte is cocrystalized with the matrix (small organic compounds having strong absorbance in the laser wavelength) and a laser beam is directed onto this crystal, causing vaporization of the matrix and desorption of the ions. The main advantage of MALDI is the generation of predominantly single-charged molecular ions of even macromolecules with a molecular mass up to 300 kDa with the exact determination of molecular mass with accuracy of ±0.1%. Therefore, the MALDI–MS is a more attractive option for the characterization of heterogeneous samples [7]. The MALDI–MS is combined with CE separations of peptides and proteins more in an offline mode than with an online liquid sample delivery connection [64]. In offline mode, fractions separated from the CE are collected and deposited on a MALDI target in the form of spots for MS analysis. Recently, approaches for the online coupling of CE with MALDI–MS have been developed to minimize sample handling and potential losses [65, 66].

## 3 APPLICATIONS

In pharmaceutical biotechnology, the development of analytical procedures to validate identity, strength, quality, and purity of biopharmaceutical products is one of the most important issues for the production of highly specialized biotechnology

products, as described in the U.S. Food and Drug Administration's (FDA's) current Good Manufacturing Procedures (cGMP) requirements for drugs (Title 21 of the Code of Federal Regulations, Parts 210 and 211). In this chapter, we discuss the use of CE to assess the identity, purity, heterogeneity, and stability of peptides and proteins, and its applications of CE in proteomics research.

### 3.1 Identity Determination

The identity of peptides and proteins can be determined by specific activity assays, determination of amino acid composition and sequence, and assessment of such physico-chemical parameters as molecular mass and p*I*. Several CE techniques have been used for the identity of peptides and proteins, which include peptide mapping by CZE or CEC, CIEF for the determination of protein's p*I*, SDS–CGE for the determination of relative molecular masses of proteins, and CE–MS for direct molecular mass assignment of peaks separated by CE.

Traditionally, IEF and SDS–PAGE in slab gel systems have been routinely used to confirm the identity of the proteins. CE formats, CIEF and SDS–CGE, offer fast and reproducible separations and direct online detection without the need of staining and destaining procedures used in slab gel techniques. As a typical example, Hunt et al. showed CIEF and SDS–CGE methods for the qualitative analysis of recombinant humanized monoclonal antibody HER2 (rhuMAbHER2) [67]. The CIEF separated five charged isoforms of rhuMAbHER2 with estimated p*I* values in the range of 8.6–9.1. These results agreed well with the p*I* values determined on slab gel IEF. In SDS–CGE, the expected molecular masses of intact rhuMAbHER2 under nonreduced conditions and its heavy-chain and light-chain fragments under reduced conditions were identified with good correlation with SDS–PAGE analysis. This study demonstrated the feasibility of replacing the slab gel techniques with CE methods in a quality control environment.

***CE–MS.*** In SDS–PAGE, the molecular mass of proteins is estimated based on comparison with reference proteins with relatively poor accuracy of ±5–10% [68]. When proteins are extracted from slab gels and subsequently transferred to ESI–MS or MALDI–MS, the mass accuracy somewhat improves, but the methodology still suffers from the interferences originated from gel [50]. Frequently, accurate molecular mass determination is necessary especially for the identification of unambiguous peptides and proteins. Since its introduction in 1987 by Olivares et al. [69], CE coupled to MS has gained increasing attention in peptide and protein research because of the CE ability to separate analyte from complex mixtures with high efficiency and minimal sample consumption. In CZE–ESI–MS, mass accuracies down to 0.0008% have been realized for model mixtures of proteins [70].

Tsuji et al. reported CZE–ESI–MS for the analysis of recombinant bovine and porcine somatotropins (rbSt and rpSt) [71]. The average molecular masses of rbSt and rpSt were determined to be 21,812.6 and 21,798.3, which were nearly identical to the theoretical values of 21,812.0 and 21,797.9, respectively. Yeung et al. reported on the CE–ESI–MS method for the analysis of high-mannose glycoproteins, ribonuclease B, and recombinant human bone morphogenetic protein-2 [72]. CE separations were performed with 50-mM beta-alanine buffer (pH 3.5) in a

polyacrylamide-coated capillary and a coaxial sheath–liquid interface was used for CE–MS coupling. The identities of glycoform peaks separated by CE were determined by the combination of UV and MS detection data without the need of oligosaccharide release or derivatization treatments.

Na et al. reported on an offline combination of CZE and MALDI–TOF MS for the identification of salmon calcitonin (sCT) acylation products formed in the degrading poly(lactic-co-glycolic acid) (PLGA) microsphere formulations [64]. As shown in Figure 5, the peptides extracted from sCT microspheres incubated in drug release medium were analyzed by both CZE with UV detection (200 nm) and MALDI–TOF MS. After the incubation, the additional peak was observed in CE (Figure 5c) and mass peaks of m/z 3491.15 and 3549.16 except the intact sCT were presented in MALDI–TOF MS spectrum (Figure 5d). The identities of two peaks separated in CE (Figure 5c) were confirmed by reanalysis of each peak fraction using MALDI–TOF MS, which was also used for the quantitation of peptide loaded into microsphere formulations [73].

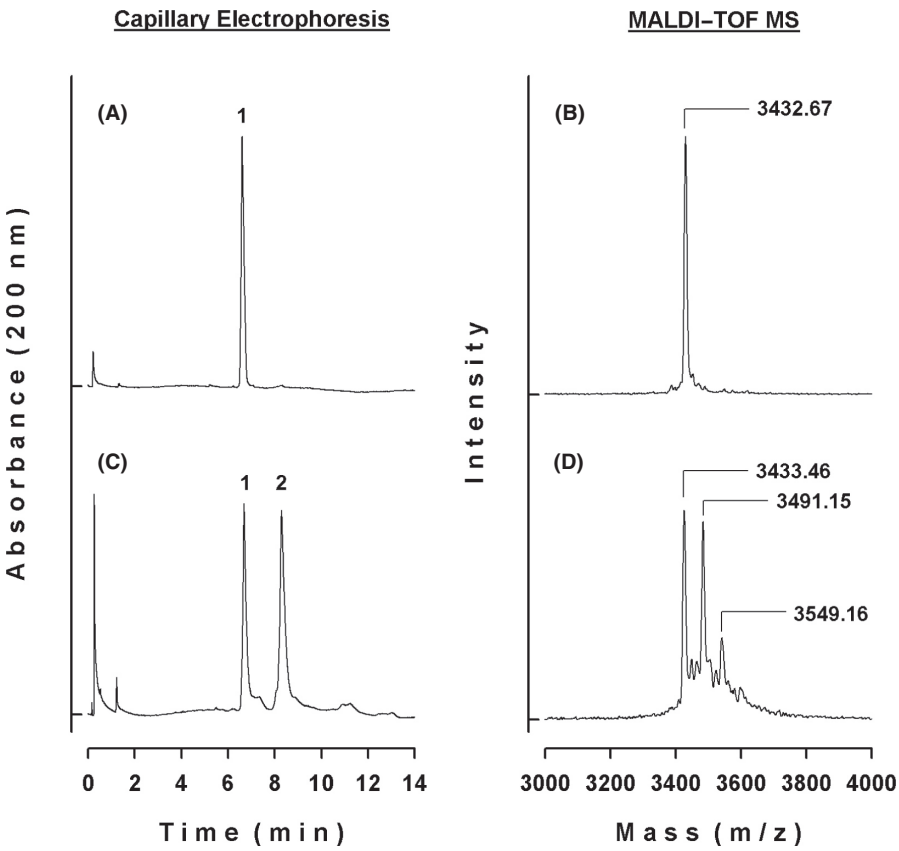


**Figure 5.** CZE and MALDI–TOF MS analysis of sCTs extracted from microsphere at 0 day (A and B) and 21 days (C and D) of incubation in release medium at 37°C. In the CE electropherogram, peak 1 is native sCT and 2 represents the degradation products. (Reprinted from Ref. 64, with permission.)

Recently, a variety of peptides and proteins were analyzed by CE–MS techniques, which include CZE–MALDI–TOF–MS, CZE–ESI–MS, CIEF–ESI–MS, and CEC–ESI–MS. Their applications are well summarized in a recent review paper [50].

***Peptide Mapping.*** Peptide mapping is an important tool for protein identif-ication, primary structure determination, the detection of posttranslational modifications, the identification of genetic variants, and the determination of glycosylation and/or disulfide sites. For these reasons, peptide mapping is widely used for quality control and for the characterization of recombinant DNA-derived products. Moreover, the high resolution of CE makes it a powerful peptide mapping technique.

Rush et al. (1993) reported on the CE-based peptide mapping of recombinant human erythropoietin (EPO) derived from Chinese hamster ovary (CHO) cells. Using 100-mM heptanesulfonic acid in 40-mM sodium phosphate buffer (pH 2.5) as a separation buffer, this technique showed a baseline separation of 16 tryptic-digested peptides with one partial separation of two peptides [74]. This peptide map demonstrated the structural differences between the glycosylated form expressed in CHO cells and the non-glycosylated form expressed in *E. coli*, and it provided information on the heterogeneity in glycoforms of EPO. Zhou et al. presented the results of the CE–ESI–MS analysis of tryptic digests of EPO, and they found that the technique is complementary to HPLC–ESI–MS [75]. Boss et al. reported on an evaluation of CE–MS for the peptide mapping of EPO [76]. In this study, tryptic-digested peptides were first separated by reversed-phase (RP)–HPLC and collected fractions were analyzed by offline and online CZE–MS experiments employing a capillary coated with Polybrene in the presence of polyethylene glycol and 0.67-M formic acid as a separation buffer.

Despite the impressive ability of CE to separate peptides, a complete separation of all digested peptides is unlikely to be achieved because of the highly complex natures of the peptide maps of large proteins. Recently, multidimensional separations, such as two-dimensional (2D) electrophoresis, HPLC–CZE, HPLC–MS, CZE–MS, and HPLC–CZE–MS, have been applied for complete resolution of highly complex peptide mixtures [77]. A sequential offline combination of RP–HPLC and CZE was used for the peptide mapping of pepsin isoenzymes, recombinant human tissue plasminogen activator, cytochrome c, and myoglobulin [78–80]. Kang et al. presented unique "fingerprint" peptide maps of two closely related proteins, β-lactoglobulins A and B, by combining CZE separations performed in four different channels and MEKC separations performed in two different channels in a 96-capillary array [81]. He et al. developed a novel multiplex CE system for the high-throughput comprehensive peptide mapping of proteins [82]. By using multiple separation conditions using six CZE buffers and two MEKC buffers in a 20-capillary array, the peptide fragments of proteins digested by three different enzymes were readily resolved and showed unique fingerprints. These 20 capillaries were monitored simultaneously at 214 nm using a single photodiode array detector, and the overall analysis time from reaction to detection was about 40 min.

### 3.2 Purity Control

Purity control is necessary on peptide and protein products for process and quality control purposes [83–85]. Minor amounts of impurities and degradation products

must be determined in the presence of much larger quantities of primary components. These components may be structurally very similar to the main component. CE can be applied as a sensitive control method for such determinations because it provides rapid and accurate qualitative and quantitative data on peptide and protein preparations.

***Peptide Purity.*** Peptide purity checks have been routinely performed by RP–HPLC. Recently, CE has increasingly been used as a versatile technique for the analysis of peptide drugs in the pharmaceutical industry [84]. Because of the different separation mechanisms, CZE is recognized as an excellent complementary tool to RP–HPLC. CZE separates the peptide based on differences in mass-to-charge ratios, whereas RP–HPLC separation is per-formed based on hydrophobicity differences. Peptides with similar hydro-phobicities, which are difficult to be separated by RP–HPLC, can often be resolved based on charge-to-mass ratio differences. Therefore, peaks that appear pure by RP–HPLC are often resolved in multiple peaks by CZE. Moreover, CZE requires only nanoliter quantities of sample so that almost all of the sample can be used for subsequent sequence analysis.

Ridge and Hettiarachchi reported on CZE separations of bradykinin and its impurities using phosphate buffers at various pH values (pH 2.5, 3.5, and 4.5), which were not fully resolved by HPLC [86]. When buffer pH was increased from 2.5 to 4.5, several impurities were well separated from the major peak of bradykinin. In purity checks of peptide, the similar superiority of CZE over RP–HPLC is also found in publications by Chen et al. [87], Hettiarachchi et al. [88], and Moumakwa et al. [89].

***Protein Purity.*** During the production of recombinant proteins, process monitoring is required to assure purity levels required in every step. The recovery and purification of product from fermentation broth typically involve various procedures, such as filtration, centrifugation, and chromatography. After each purification and final step, constituent levels must be determined to ensure that the desired levels of purity have been achieved. In addition to its control function, this purity information is also frequently used to further optimize purification processes. CZE and SDS–CGE can be mostly used for the purity checks of protein products.

Several CZE methods were reported for the purity determinations of recombinant human insulin. Human insulin consists of two peptide chains, A (21 amino acids) and B (30 amino acids), which are connected by two disulfide bonds. To determine the purity of insulin, two main degradation products, acidic and neutral desamido-insulin (desamidated at positions A-21 or A-21 and B-3, respectively), should be separated from the main compound [90]. According to U.S. Pharmacopoeia (USP), the relative amount of desamido insulin must not exceed 3% of the total amount of insulin and desamido insulin [91].

A CZE method for the separation of insulin and its deamidation products with untreated fused-silica capillaries using a run buffer containing 2-(*N*-cyclohexylamino) ethanesulfonic acid (CHES), triethylamine, and 10% acetonitrile has been reported [90]. This system separated acidic and neutral desamido-insulin in formulated human insulin with the relative standard deviation (RSD) of the migration times ≤1%, whereas RP–HPLC coeluted the neutral desamido-insulin with insulin. Sergeev et al. described an analytical scheme for monitoring recombinant human

insulin during the various steps of its production [92]. The CZE method has been included, together with narrow-bore HPLC and MALDI–TOF MS, in the analytical scheme for monitoring the production of recombinant human insulin. Combinations of these complementary techniques allowed us to obtain unambiguous information about the purity and primary structure of all intermediates of recombinant human insulin production, and they enabled the optimization of some process parameters. Figure 6 shows the result of a CZE analysis performed to check the purity of insulin purified by ion-exchange chromatography. This CZE analysis showed the presence of several impurities from the preparation (Arg-insulin) and degradation (unidentified desamido-insulins) in the isolated insulin.

SDS–CGE provides an overall composition of a protein sample based on the molecular mass, even if it is less suited for the analysis of subtle changes in a protein. Thus, SDS–CGE is mainly used to check for the presence of other proteins, protein aggregates, or protein degradation fragments in the purity checks.

Hunt and Nashabeh showed the SDS–CGE method with LIF detection for the analysis of humanized recombinant $IgG_1$-monoclonal antibody (MAb) [93]. For LIF detection, the MAb was first derivatized with a 5-carboxytetramethylrhoda-

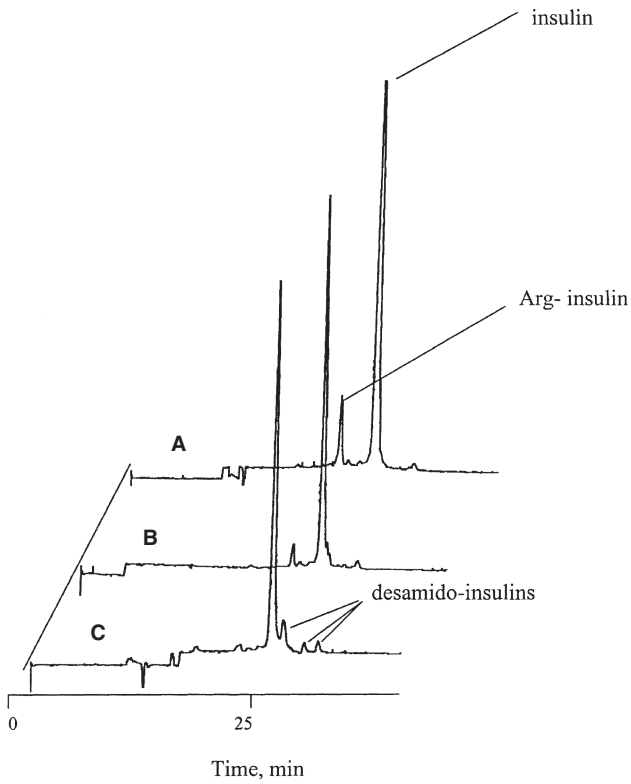


**Figure 6.** CZE analysis for insulin purity checks in three fractions obtained by ion-exchange chromatography. This CZE analysis demonstrated the presence of several impurities (Arg-insulin and desamido-insulins) in the isolated fractions. (Reprinted from Ref. 92, with permission.)

mine succinimidyl ester. The derivatized sample was then incubated with SDS, and the SDS–MAb complexes were separated by SDS–CGE using a hydrophilic polymer as a sieving matrix. The capabilities of SDS–CGE using UV detection or LIF detection after derivatization were compared. Figure 7 shows the results of the SDS–CGE analysis by UV detection at 220 nm and LIF detection using argon laser

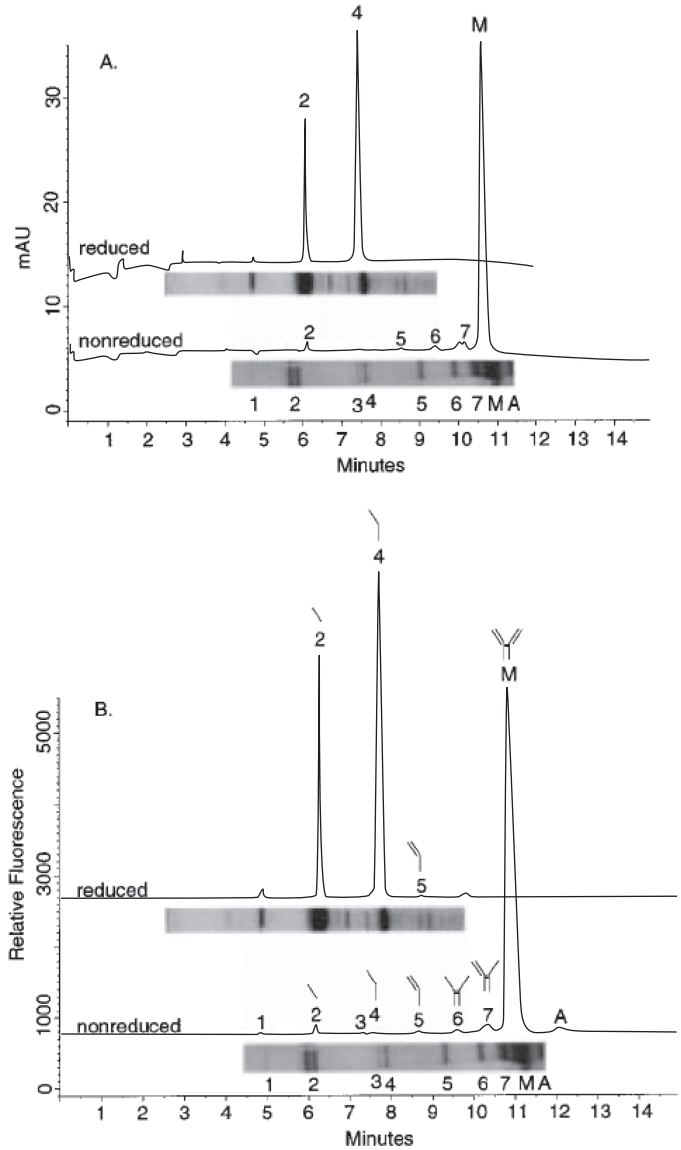


**Figure 7.** SDS–CGE separations of nonreduced and reduced preparations of a therapeutic recombinant MAb. Insets show silver-stained SDS–PAGE traces of the same sample preparations (M: monomer, A: aggregate). (A) UV detection of unlabeled samples at 220 nm. (B) LIF detection of labeled samples at 488-nm excitation and 560-nm emission. (Reprinted from Ref. 93, with permission.)

with excitation wavelength of 488 nm and emission wavelength of 560 nm. Under reducing conditions, the light and heavy chains originating from the original MAb were well separated, and the results of SDS–CGE under both reducing and non-reducing conditions generally showed good agreement with band profiles obtained by silver-stained SDS–PAGE. SDS–CGE using LIF detection allowed the detection of MAbs at a low-nanomolar concentration (~9 ng/mL), which was comparable with silver-stained SDS–PAGE and was a 140-fold increase over SDS–CGE using UV detection. This improved sensitivity also allowed the detection of low-level impurity peaks, which were not detected by UV detection. Repeated analysis showed that RSD values for the migration time were below 1%, and the RSD of the area of the main peak was lower than 0.6%. These applications demonstrate the usefulness of SDS–CGE–LIF for detecting manufacturing inconsistencies in recombinant protein production.

Several validated CZE and SDS–CGE methods for the purity analysis of peptides and proteins are provided in Table 2.

### 3.3 Heterogeneity Characterization of Glycoproteins

Glycoproteins produced from mammalian expression systems consist of a population of glycosylated variants (glycoforms). Glycosylation is one type of posttranslational modification that requires monitoring because variation in the carbohydrate composition may significantly alter the properties of a protein such as biological activity, pharmacokinetic properties, solubility, and stability. Several CE methods have been used for analyzing the intact glycoproteins and monitoring their production [3, 101]. Because of simplicity and high resolving capacity based on the charge-to-mass ratio, CZE has been routinely used for characterizing the heterogeneity of glycoproteins. In addition, CIEF has been increasingly used for separation of the glycoforms.

***Erythropoietin (EPO).*** EPO is a glycoprotein produced primarily by the kidney and the main regulator of red blood cell production [102]. Human EPO consists of 165 amino acids that are heavily glycosylated (one O-linked and three N-linked carbohydrate chains corresponding to 40% of the molecular mass) [103]. Several glycoforms of human EPO exist with different degrees of glycosylation and sialic acid residue numbers [104]. Several reports are available on the application of CE for the separation of EPO glycoforms, involving different CE modes [105]. Among them, CZE and CIEF are the most successful CE modes in this respect. Table 3 provides an overview of the CE analyses of EPO glycoforms.

A CZE method for the characterization of the glycoform patterns of pharmaceutical EPO is included in the 2002 European Pharmacopoeia, as a substitute for the conventional isoelectric focusing test [113]. As shown in Figure 8, EPO was resolved with high resolution into its glycoforms using a separation buffer containing putrescine (1,4-diaminobutane) and urea [110]. Individual peaks correspond to multiple glycoforms with similar overall mass-to-charge ratios [3, 109, 114]. At pH 5.5, sialic acid residues are negatively charged and migrate against the EOF, and thus the glycoforms are considered to elute in the order of increasing number of sialic acid residues [109, 114]. The addition of putrescine to the separation buffer leads to a reduction in EOF and solute interaction with the capillary wall. Urea is

**TABLE 2. CE Applications for the Purity Checks of Peptides and Proteins**

| Peptides and Proteins | CE Modes | Separation Buffers | Capillary | Detection | Ref. |
|---|---|---|---|---|---|
| Bradykinin | CZE | 50-mM phosphate buffer, pH 2.5, 3.5, 4.5 | Polyacrylamide-coated | UV (215 nm) | [86] |
| Biphalin | CZE | 50-mM phosphate buffer, pH 2.5 | Polyacrylamide-coated | UV (215 nm) | [88] |
| Enkephalin and dalagin analogs | CZE | (1) 100-mM phosphoric acid, 50-mM Tris, pH 2.25<br>(2) 100-mM iminodiacetic acid, pH 2.30<br>(3) 40-mM Tris, 40-mM tricine, pH 8.10 | Uncoated | UV (206 nm) | [94] |
| Placental alkaline phosphatase | CZE | 50-mM boric acid, 2-mM putrescine (pH 8.5) | Uncoated | UV (200 nm) | [95] |
| Placental alkaline phosphatase | SDS–CGE | CE–SDS protein kit (Bio-Rad) | Uncoated | UV (220 nm) | [95] |
| Protegrin IB-367 | CZE | 100-mM phosphate buffer, pH 2.6 | Uncoated | UV (200 nm) | [87] |
| Recombinant acidic fibroblast growth factor | CZE | 50-mM phosphate buffer (pH 2.5) with 0.25% hydroxypropyl-methylcellulose (HPMC) | Uncoated | UV (214 nm) | [96] |
| Recombinant hirudin | CZE | 50-mM sodium phosphate, 40-mM aminoethanesulfonic acid, 25-mM hexanesulfonic acid, 10-mM sodium borate, 2-mM tetramethylammonium sulfate, 1-mM diaminopentane (pH 7) | Uncoated | UV (200 nm) | [97] |

**TABLE 2.** *Continued*

| Peptides and Proteins | CE Modes | Separation Buffers | Capillary | Detection | Ref. |
|---|---|---|---|---|---|
| Recombinant human insulin | CZE | 50-mM acetate, 850-mM CHES, 10% acetonitrile (pH 7.8) | Uncoated | UV (214 nm) | [90] |
| Recombinant human insulin | CZE | 100-mM sodium borate (pH 9.3) | Uncoated | UV (214 nm) | [98] |
| Recombinant human insulin | CZE | 0.1-M tricine, 0.02-M NaCl, 1-M urea, 0.45-M ethylene glycol (pH 8.1) | Uncoated | UV (214 nm) | [91] |
| Recombinant human insulin | CZE | 50-mM sodium phosphate (pH 10.6); 100-mM sodium phosphate-30% MeOH (pH 9) | Uncoated | UV (210 nm) | [92] |
| Recombinant human interleukin-6 | ITP–CZE | 20-mM ammonium acetate (pH 4.2) | Uncoated | UV (200 nm) ESI–MS | [99] |
| Recombinant monoclonal antibodies | SDS–CGE | CE–SDS protein kit (Bio-Rad) | Uncoated | UV (220 nm) LIF (Ex 488 nm, Em 560 nm) | [93] |
| Recombinant $NADP^+$-dependent formate dehydrogenase | SDS–CGE | eCAP SDS 14-200 kit (Beckman) | Uncoated | UV (214 nm) | [100] |
| Synthetic porcine secretin | CZE | 50-mM phosphate buffer (pH 7.4 or 3.5) | Uncoated | UV (210 nm) | [89] |

**TABLE 3. CE Applications for the Heterogeneity Characterization of Recombinant Erythropoietin**

| Objective | CE Modes | Separation Buffers | Capillary | Detection | Ref. |
|---|---|---|---|---|---|
| Separation in drug formulations | CZE | 200-mM sodium phosphate, 1-mM nickel chloride (pH 4.0) | eCAP amine (Beckman) | UV (200 nm) | [106] |
| Separation of glycoforms | CZE | 100-mM acetate-phosphate (pH 4.0) | Uncoated | UV (214 nm) | [75] |
| Separation of glycoforms | CZE | 10-mM sodium acetate, 0.5% HPMC (pH 5.7) | DB-1 | UV (214 nm) | [107] |
| Discrimination of recombinant and urinary origin | CZE | 10-mM tricine, 10-mM NaCl, 10-mM sodium acetate, 7-M urea, 3.9-mM putrescine (pH 5.5) | Uncoated | UV (214 nm) | [108] |
| Separation of glycoforms | CZE | 300-mM acetic acid-ammonium acetate (pH 4.0–5.5) | Polybrene-coated | UV (214 nm) | [109] |
| Separation of NESP (novel EPO analog) | CZE | 10-mM tricine, 10-mM NaCl, 10-mM sodium acetate, 7-M urea, 2.5-mM putrescine (pH 4.5–5.5) | Uncoated | UV (214 nm) | [110] |
| Separation of glycoforms | CIEF | Catholyte: 20-mM NaOH<br>Anolyte: 91-mM phosphoric acid<br>Ampholyte: Beckman ampholyte 3–10, Pharmalyte 2.5–5 | polyacrylamide-coated | UV (280 nm) | [111] |
| Separation of glycoforms | CIEF | Catholyte: 20-mM NaOH<br>Anolyte: 91-mM phosphoric acid<br>Ampholyte: Beckman ampholyte 3–10, Pharmalyte 2.5–5 | eCAP neutral (Beckman) | UV (280 nm) | [112] |

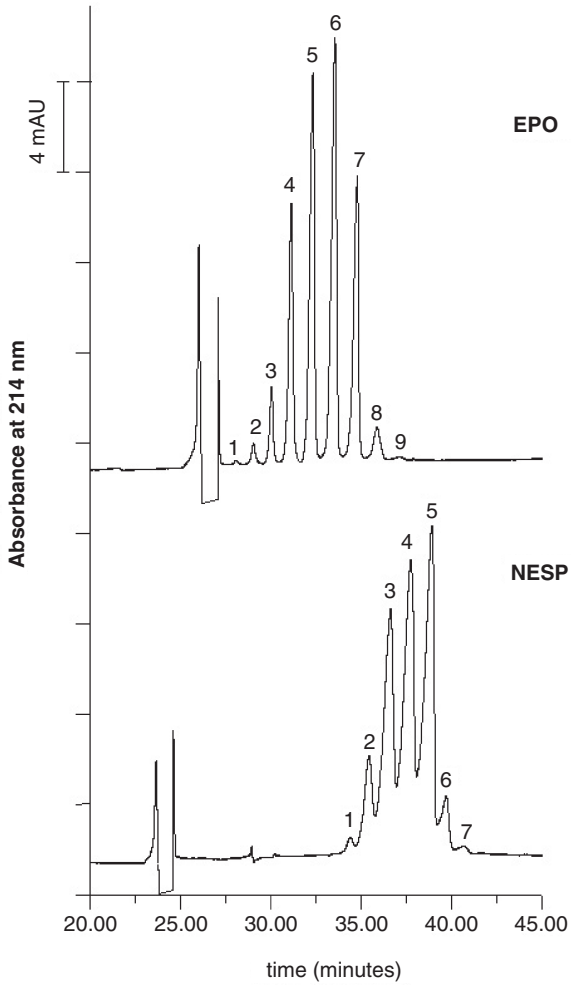


**Figure 8.** CZE separation of EPO and NESP glycoforms using a separation buffer containing 0.01-M Tricine, 0.01-M NaCl, 0.01-M NaAc, 7-M urea, and 2.5-mM putrescine (pH 5.5 adjusted with 2-M acetic acid) with bare fused-silica capillary (68.5 cm × 50 μm). (Reprinted from Ref. 110, with permission.)

used to inhibit protein aggregation by preventing the formation of intermolecular hydrogen bonds and disrupting hydrophobic and noncovalent interactions [105]. Sanz-Nebot et al. described the CZE separation of a novel erythropoiesis-stimulating protein (NESP), which is a recently approved hyperglycosylated analog of human EPO with a long-lasting effect [110]. As NESP, due to its two extra *N*-linked oligosaccharide chains, has a higher sialic acid content than EPO, its glycoforms are expected to carry a higher overall negative charge at pH 5.5. As shown in Figure 8, NESP glycoforms migrated slower than EPO glycoforms at pH 5.5 and the peaks corresponding to NESP glycoforms were only partially resolved. Figure 9 shows the influence of separation buffer pH on the separation of NESP glycoforms. The peak resolution was improved at lower pH condition, which may be due to the lower

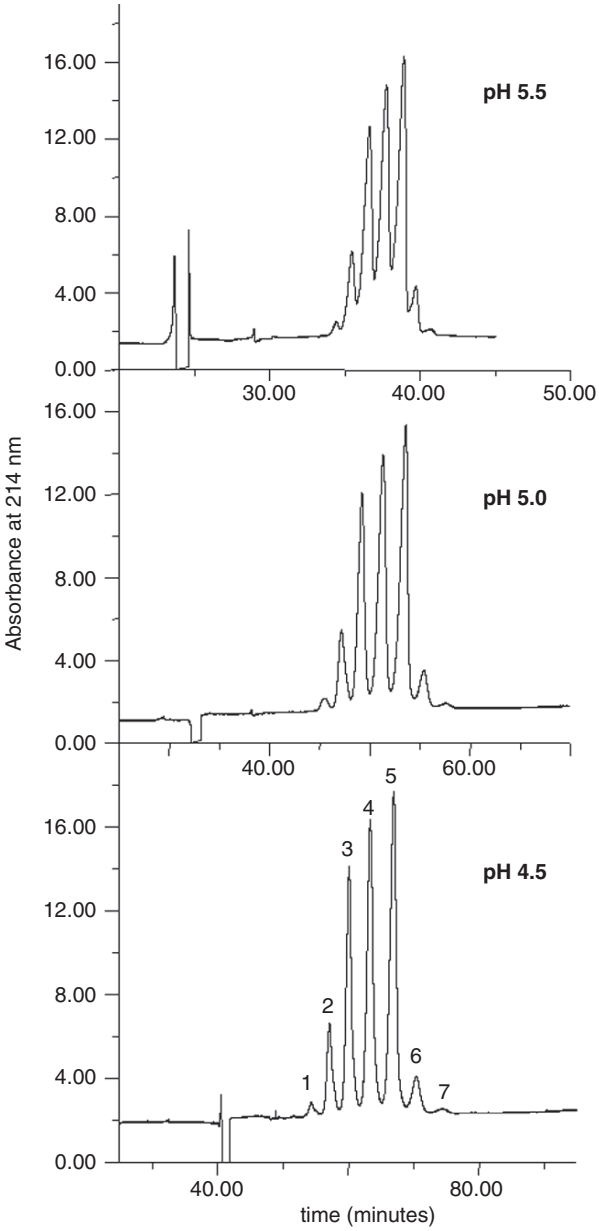


**Figure 9.** Effect of separation buffer pH on the separation of NESP glycoforms. The compositions of separation buffers are the same as Fig. 8, and the pH value was adjusted with 2-M acetic acid. (Reprinted from Ref. 110, with permission.)

p*I* of NESP compared with EPO. At pH 4.5, the seven peaks of NESP glycoforms were resolved to baseline.

CIEF is an important technique for analyzing the charged variants of EPO glycoforms on the basis of sialic acid content differences. Cifuentes et al. showed that EPO glycoforms could be resolved by using a mixture of broad and narrow

pH-range ampholytes and by adding urea to the EPO sample [111]. The separation and quantification of seven EPO glycoforms with apparent p*I* values of 3.78–4.69 was accomplished using an optimized hydrodynamic mobilization method. Compared with conventional gle-IEF, better resolution was obtained in a shorter time. Lopez-Soto-Yarritu et al. described an improved CIEF method for the separation of EPO glycoforms with neutral-coated capillaries using ampholytes in the pH range 2–10 and bovine β-lactoglobulin-A as an internal standard [112]. CIEF analysis of EPO glycoforms was achieved with a migration time reproducibility of 0.05% and peak area reproducibility better than 3.4%.

***Monoclonal Antibodies.*** Monoclonal antibodies (MAbs) are being in-creasingly used for a variety of diagnostic and therapeutic applications. As is the case for most recombinant proteins, preparations of MAbs often show considerable heterogeneity due to posttranslational modifications involving glycosylation. CIEF has been widely used for the separation of charged isoforms of several MAbs [67, 115–117].

Hunt et al. reported validation of CIEF method for recombinant MAb C2B8 [115]. As shown in Figure 10, CIEF separated four isoforms of MAb C2B8. Among them, three peaks were identified as charge isoforms present due to incomplete C-terminal lysine processing and represent MAb C2B8 with 2 (peak 1), 1 (peak 2), and 0 (peak 3) C-terminal lysines, respectively. The CIEF method was validated in accordance with International Conference on Harmonization (ICH) guidelines, and the result demonstrated that it is accurate, precise, linear, and highly specific for the determination of identity and charge distribution of MAb C2B8.

Tang et al. performed a systematic study on the routine analysis of a recombinant immunoglobulin G (IgG) by CIEF [116]. The method used a dimethyl siloxane-coated capillary (DB-1) and a separation matrix of 2% ampholytes in 0.4%

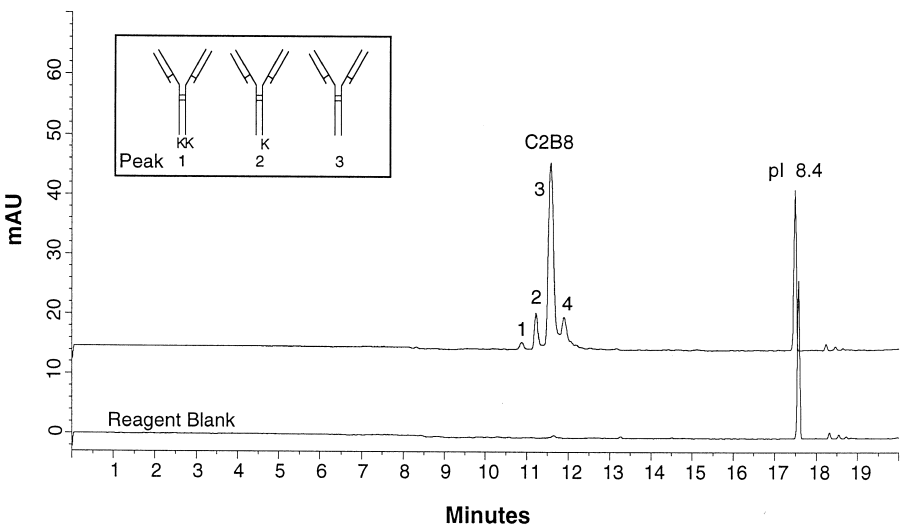


**Figure 10.** CIEF analysis of MAb C2B8. Box inset is a schematic illustration on the three C-terminal lysine variants. (Reprinted from Ref. 115, with permission.)

methylcellulose. The composition of various IgG samples with respect to isoform content was analyzed and compared quantitatively with gel-IEF. The reproducibility of the method was examined with the quantitative analysis of IgG sample in replicate over 3 days. The RSD of peak areas was below 2% intraday and 8% interday. The RSD for the mobilization times was below 1% intraday and 3% interday. Separation variability was examined over 150 runs. During this period, reproducible migration times and good resolution without peak shape deterioration were observed. This type of robustness demonstrates the potential of the CIEF method as a useful routine analysis tool.

Kats et al. employed the MEKC method to separate isoforms of chimeric MAb BR96 [118]. In the range from pH 2 to 12, MAb BR96 was separated into one to five isoforms by MEKC using a 12-mM borate buffer (pH 9.4) containing 25-mM SDS in uncoated capillary. Kats et al. further applied MEKC to separate structurally similar isoforms of a single-chain immunotoxin fusion protein BR96-sFV-PE40 using a 12-mM borate buffer (pH 9.6) containing 16-mM cholic acid [119].

Table 4 summarizes CE applications for the characterization of charge-based heterogeneity of several MAbs.

### 3.4 Stability Study

Peptides and proteins are complex molecules containing many functional groups, which can undergo a variety of degradation reactions, such as oxidation, reduction, deamidation, hydrolysis, arginine conversion, β-elimination, and racemization of amino acids [120]. Proteins further undergo physical changes in the secondary, tertiary, and quaternary structures [121]. These degradation reactions can alter the conformation, size, charge, and hydrophobicity of the peptides and proteins, thereby affecting the biological activity or even leading to formation of toxic compounds. As is recognized by the ICH guidelines for stability testing of biotechnological products, the monitoring of the stability of pharmaceutical peptides and proteins is of the utmost importance in the aspects of quality control and safety. CE has been used as a powerful tool for monitoring of stability and degradation of peptides and proteins in various conditions [122].

The high resolving power of CZE could be used for monitoring the stability and degradation reactions of peptide and protein products. Hoitink et al. applied CZE–MS for a stability study of goserelin, a luteinizing hormone-releasing hormone analog [123]. Using a 10% acetic acid as running buffer, high-resolution separation was obtained. In a stability study of goserelin at pH 5 and 9, the degradation of the C-terminal semi-carbazide group was observed. Lai et al. used CZE for the separation of asparagines-containing hexapeptide and its deamidation products and applied this method to a peptide stability study in the presence of polymers [124]. The offline CZE/MALDI–TOF MS combination has been used to monitor the stability of sCT, human parathyroid hormone 1-34 (PTH), and leuprolide in biodegradable PLGA microsphere formulations [64]. Figure 11 shows CZE monitoring of the sCT stability in the PLGA microspheres incubated in a 10-mM phosphate buffer saline (pH 7.4) containing 0.02% Tween 80 and 0.02% sodium azide at 37°C. CZE separations were performed using a 100-mM phosphate buffer (pH 2.5) in polyacylamide-coated capillary with UV detection at 200 nm. During the *in vitro* drug release study, intact sCT peak rapidly decreased, whereas the additional peaks

**TABLE 4. CE Applications for the Heterogeneity Characterization of Monoclonal Antibodies**

| MAbs | CE Modes | Separation Buffers | Capillary | Detection | Ref. |
|---|---|---|---|---|---|
| MAb HER2 | CIEF | Catholyte: 40-mM NaOH<br>Anolyte: 20-mM phosphoric acid<br>Ampholyte: Pharmalyte 8–10.5, Biolyte 7–9, Biolyte 3–10 (8:1:1) containing 0.5% TEMED, 0.2% HPMC | BioCAP LPA (Bio-Rad) | UV (280 nm) | [67] |
| MAb C2B8 | CIEF | Catholyte: 40-mM NaOH<br>Anolyte: 20-mM phosphoric acid<br>Ampholyte: Pharmalyte 8–10.5, Biolyte 7–9, Biolyte 3–10 (8:1:1) containing 0.5% TEMED, 0.2% HPMC | BioCAP LPA (Bio-Rad) | UV (280 nm) | [115] |
| Recombinant IgGs | CIEF | Catholyte: 20-mM NaOH in 0.4% MC<br>Anolyte: 120-mM phosphoric acid in 0.4% MC<br>Ampholyte: 2% Pharmalyte 3–10 in 0.4% MC | DB-1 | UV (280 nm) | [116] |
| Anti-TNF MAb D2E7 | CIEF | Catholyte: 40-mM NaOH<br>Anolyte: 20-mM phosphoric acid<br>Ampholyte: Pharmalyte 8–10.5, Biolyte 7–9, Biolyte 3–10 (8:1:1) containing 0.5% TEMED, 0.2% HPMC | Neutral-coated (Beckman) | UV (280 nm) | [117] |
| MAb BR96 | MEKC | 12-mM borate buffer (pH 9.4) containing 25-mM SDS | Uncoated | UV (214 nm) | [118] |
| BR96 sFV-PE40 immunotoxin | MEKC | 12-mM borate buffer (pH 9.6) containing 16-mM cholic acid | Uncoated | UV (214 nm) | [119] |

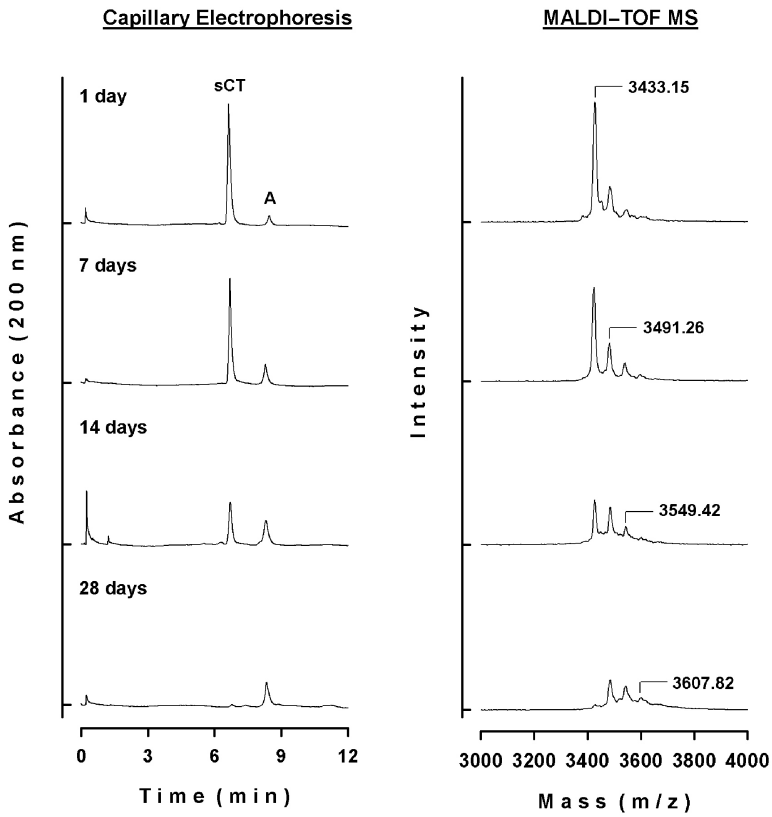


**Figure 11.** Monitoring of sCT stability in degrading PLGA microsphere formulations by CZE and MALDI–TOF MS. In CE, "A" represents the acylation products of sCT. (Reprinted from Ref. 64, with permission.)

predominantly appeared after 28 days of incubation. MALDI–TOF MS data supported CE results with good correlation.

SDS–CGE can be also a useful stability-indicating method because it detects changes in fragmentation of proteins. Hunt et al. demonstrated that SDS–CGE could detect peaks resulting from the fragmention and aggregation of MAb HER2 when incubated at 37°C for 27 days [67]. Hunt and Nashabeh showed the potential of SDS-CGE for the separation and detection of proteolytic and deglycosylated fragments generated from MAbs [93].

### 3.5 Characterization of PEGylated Peptides and Proteins

The covalent attachment of PEG, PEGylation, is a well-established technique of overcoming several problems associated with the therapeutic uses of proteins. This technique has demonstrated reduced immunogenicity, extended circulating half-life, and improved stability for these therapeutic agents [125–127].

In general, the PEGylation process results in molecular heterogeneity with respect to the distribution in terms of number and positions of attached PEG molecules as well as the inherent polydispersity of PEG itself [128]. Therefore,

PEGylated molecules are among the most challenging products in analytical biochemistry, and their characterizations are becoming more important because these heterogeneities may confer different biological properties [129, 130]. When developing PEGylated biomolecules as therapeutic agents, various points should be considered, including the characterization of starting materials (i.e., proteins and PEG molecules), the determination of PEGylation sites, and the stoichiometry of PEG attachment [131]. The consistency of each preparation should also be established. For well-characterized PEGylated biomolecules, it is important to develop and validate suitable analytical techniques that enable consistent manufacturing processes to be established. Table 5 provides an overview of the applications of CE methods for the characterization of PEGylated peptides and proteins.

Cunico et al. described a charge-reversed CZE method for the analysis of PEGylated proteins [132]. A removable coating using ethylene glycol was applied to allow the negative capillary surface charge to be made positive, to prevent the protein adsorption to the capillary wall. Using this coated capillary, six different PEGylated molecules were characterized.

Bullock et al. characterized PEGylated superoxide dismutase (SOD) by CZE using a low-pH separation buffer (pH 2.05) [133]. At low pH, protein molecules are completely protonated and the charge differences between individual PEGmers (i.e., mono-, di-, or tri-PEGmers) are minimized due to neutralization of lysine residues modified by PEG attachment. This results in a predominantly size-based separation of PEGylated proteins. Under low pH conditions, PEG–SOD conjugates were size-dependently separated and the separation pattern was very similar to the MALDI–TOF MS spectrum.

Li et al. used a semi-aqueous phosphate buffer (pH 2.5) of acetonitrile-water (1 : 1, v/v) to improve the resolution of PEGylated proteins [134]. They found that acetonitrile was mainly responsible for the good resolution observed, and this was attributed to reduced protein adsorption to the inner wall of the silica capillary used.

Na et al. determined PEGylation sites in three positional isomers of mono-PEGylated salmon calcitonins (mono-PEG-sCTs) using CZE-based peptide mapping analysis [135]. The resistance of PEGylation sites to proteolytic degradation resulted in different CE electropherogram patterns for tryptic digested mono-PEG–sCT isomers, and PEGylation sites were assigned accordingly and confirmed by MALDI–TOF MS.

Na and Lee reported on the characterization of PEGylated human parathyroid hormone (PEG–PTH) using CZE [136]. As shown in Figure 12, the CZE was used to optimize reaction conditions by monitoring the effects of reaction pH and the molar ratios of reactants on the PEGylation of PTH. The CZE method also allowed for determination of the extent of positional isomers in mono-PEG–PTH as well as for the identification of PEGylation sites.

Na et al. applied SDS–CGE to characterize PEGylated interferon alpha (PEG-IFN) [4]. The method well resolved the PEG–IFN species as well as the native IFN. As shown in Figure 13, four PEGylated IFNs (i.e., mono-, di-, tri-, and tetra-PEGylated IFNs) were detected and completely separated by SDS–CGE. The distribution of PEGylation reaction mixture in the SDS–CGE was found to be almost identical with that obtained by SDS–PAGE with Coomassie blue staining, although no band corresponding to tetra-PEGylated IFN was visualized.

**TABLE 5. CE Applications for the Characterization of PEGylated Peptides and Proteins**

| Objective | CE Modes | Separation Buffers | Capillary | Detection | Ref. |
|---|---|---|---|---|---|
| Separation using charge-reversed mechanism | CZE | 10-mM phosphate, 10% ethylene glycol (pH 2.0) | Uncoated | UV (215 nm) | [132] |
| Heterogeneity of PEG–SOD | CZE | 15-mM boric acid, 15-mM 1,3-diaminopropane, 10-mM NaCl (pH 10.1) | Uncoated | UV (200 nm) | [137] |
| Separation of PEG–SOD and comparison with MALDI–TOF MS | CZE | 40-mM phosphoric acid, 0.1-mg/mL PEG-114000, 4-mg/mL Jeffamine ED-600 (pH 2.05) | Uncoated | UV (200 nm) | [133] |
| Separation of PEG-lysozyme and peptide mapping | CZE | 25-mM phosphate, 0.1-mg/mL PEG 600K (pH 2.7) | Uncoated | UV (200 nm) | [138] |
| Separation of semi-aqueous CE method | CZE | 25-mM sodium phosphate (pH 2.5), acetonitrile-water (1:1, v/v) | Uncoated | UV (200 nm) | [134] |
| Identification of PEGylation sites in mono-PEG-salmon calcitonin | CZE | 100-mM sodium phosphate (pH 2.5) | Polyacrylamide-coated | UV (200 nm) | [135] |
| Identification of PEGylation sites in mono-PEG-human parathyroid hormone (1–34) | CZE | 100-mM sodium phosphate (pH 2.5) | Polyacrylamide-coated | UV (200 nm) | [136] |
| Separation of PEG-interferon alpha | SDS–CGE | CE–SDS protein kit (Bio-Rad) | Uncoated | UV (220 nm) | [4] |

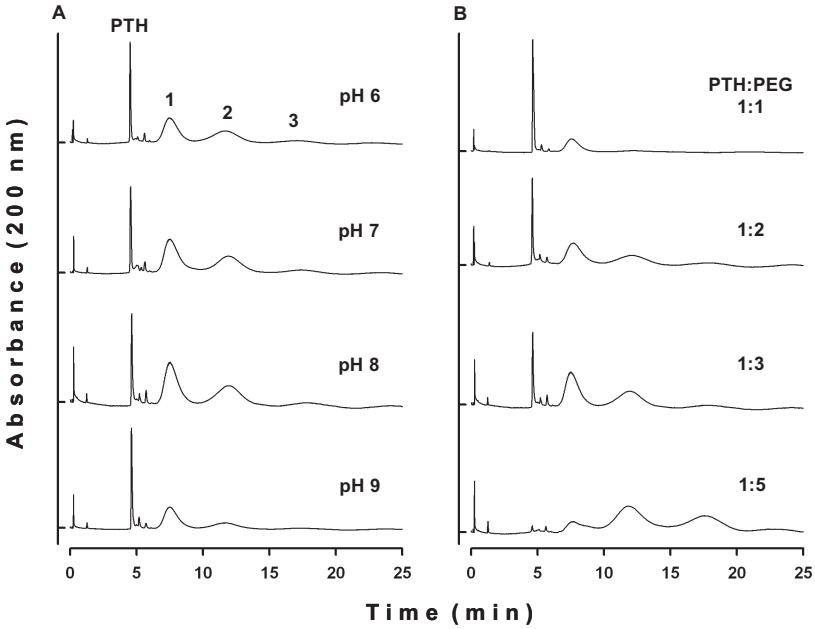


**Figure 12.** CZE monitoring of the PEGylation reaction of PTH through control of reaction pH (A) and molar ratios of reactants (B). The numbers above peaks represent the number of PEG molecules conjugated to PTH (i.e., 1: mono-, 2: di-, and 3: tri-PEGylated PTHs). (Reprinted from Ref. 136, with permission.)

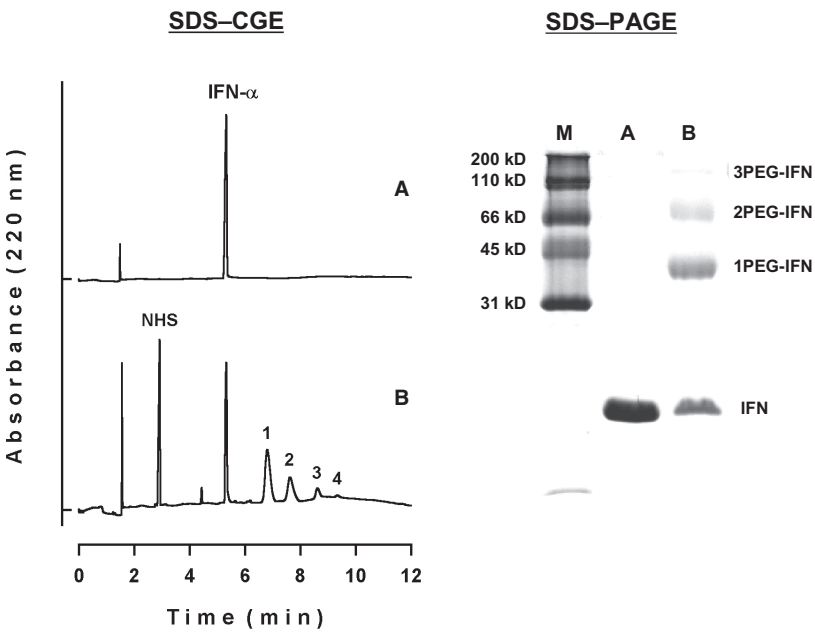


**Figure 13.** SDS–CGE and SDS–PAGE of IFNs before (A) and after (B) PEGylation reaction. In the SDS–CGE electropherogram, the numbers above the peaks represent the number of PEG molecules conjugated to IFN (i.e., 1: mono-, 2: di-, 3: tri-, and 4: tetra-PEGylated IFNs). (Reprinted from Ref. 4, with permission.) (This figure is available in full color at ftp://ftp.wiley.com/public/sci_ tech_med/pharmaceutical_biotech/.)

SDS–CGE was also useful for monitoring PEGylation reaction to optimize the reaction conditions such as the reaction molar ratio. This study demonstrated the potential of SDS–CGE for the characterization of PEGylated proteins with the advantages of speed, minimal sample consumption, and high resolution.

### 3.6 Proteome Analysis

Analysis of the proteome through the detection and identification of proteins from biological samples is gaining importance in the postgenomic era. The most traditional way to analyze proteome consists of protein separations by 2D gel electrophoresis (2-DE) and identifications by MS employing ESI or MALDI ionization [139]. However, the 2-DE–MS approach remains difficult for the detection of proteins expressed at extremely low levels or having extreme p*I* values or molecular masses and sensitivity [140]. In view of the limitations of 2-DE-based techniques, a considerable effort has been focused on the development of liquid phase-based analytical techniques, such as HPLC and CE, enabling the rapid, broad, and sensitive analysis of complex proteomic samples through the combination with MS or tandem MS analysis [5, 62, 141].

Various CE-based techniques have been widely used for proteome analysis, as has been described in numerous recent review articles [5, 62, 142, 143]. In addition to the complexity of protein samples, the large variation of the protein relative is one of the great challenges to the proteome analysis. For the broad proteome analysis, including the identification of lower abundance proteins, developments in capillary separations capable of providing extremely high resolving power and selective analyte enrichment are being highlighted [5]. As the resolving power of one-dimensional separation systems is generally not adequate for highly complex mixtures, multidimensional separations are becoming increasingly important for the comprehensive proteome analysis [7, 77].

Chen et al. developed an online combination system of CIEF with capillary RP–HPLC (CRPLC) using a microinjector as the interface for 2D separations of complex protein mixtures [144]. The resolving power of a combined CIEF–CRPLC system was demonstrated with tryptic digests of proteins from *Drosophila* salivary glands. The overall peak capacity was estimated to be around approximately 1800 over a run time of less than 8 hours.

Mohan and Lee developed a 2D CE separation system for proteomics by combining CIEF with transient capillary isotachophoresis (CITP)–CZE [145]. 2D separation of proteolytic peptides was applied to analyze tryptic digests of model proteins, including cytochrome c, ribonuclease A, and carbonic anhydrase II. Maximum peak capacity was estimated to be around 1600.

Ramsey et al. have combined MEKC with CZE, using switching of applied voltages to control fraction transport between the separation dimensions [146]. In analysis times of less than 15 min, 2D separation of tryptic digests of bovine serum albumin produced a peak capacity of 4200 (110 in the first dimension and 38 in the second dimension). The system was used to identify peptides from a tryptic digest of ovalbumin using standard addition and to distinguish between tryptic digests of human and bovine hemoglobin.

## 4 CONCLUSIONS

In biopharmaceutical analysis, CE is becoming the method of choice for the separation and characterization of peptide and protein pharmaceuticals. As shown in this chapter, the application of CE in this area has significantly increased during the past decade. The high efficient separation capability of CE provides an alternative and complementary technique to conventional analytical methods that are currently used to characterize biopharmaceuticals. Diverse separation modes of CE have been applied for the identity determination, purity control, heterogeneity characterization, stability study, and process consistency of biopharmaceuticals. In particular, CE has been recognized as a highly valuable tool for the efficient separation and characterization of the heterogeneous and complex protein products such as glycoproteins and PEGylated proteins. Recently, as the field of proteomics is becoming increasingly important, the role of CE in the strategy of protein analysis increases, with an emphasis on multidimensional separation and the combination of mass spectrometric techniques.

## REFERENCES

1. Quigley W W, Dovichi N J (2004). Capillary electrophoresis for the analysis of biopolymers. *Analyt. Chem.* 76:4645–4658.
2. Patrick J S, Lagu A L (2001). Review applications of capillary electrophoresis to the analysis of biotechnology-derived therapeutic proteins. *Electrophoresis.* 22:4179–4196.
3. Pantazaki A, Taverna M, Vidal-Madjar C (1999). Recent advances in the capillary electrophoresis of recombinant glycoproteins. *Analyt. Chim. Acta.* 383:137–156.
4. Na D H, Park E J, Youn Y S, et al. (2004). Sodium dodecyl sulfate-capillary gel electrophoresis of polyethylene glycolylated interferon alpha. *Electrophoresis.* 25:476–479.
5. Cooper J W, Wang Y, Lee C S (2004). Recent advances in capillary separations for proteomics. *Electrophoresis.* 25:3913–3926.
6. Guzman N A, Majors R E (2001). New directions for concentration sensitivity enhancement in CE and microchip technology. *LC-GC.* 19:14–30.
7. Monton M R N, Terabe S (2005). Recent development in capillary electrophoresis-mass spectrometry of proteins and peptides. *Analyt. Sci.* 21:5–13.
8. Oda R P, Spelsberg T C, Landers J P (1994). Commercial capillary electrophoresis instrumentation. *LC-GC.* 12:50–51.
9. Altria K D (1996). *Capillary Electrophoresis Guidebook*, Humana Press Inc., Totowa, N.J., pp. 15–19.
10. Offord R E (1966). Electrophoretic mobilities of peptides on paper and their use in the determination of amide groups. *Nature.* 211:591–593.
11. Compton B J (1991). Electrophoretic mobility modeling of proteins in free zone capillary electrophoresis and its application to monoclonal antibody microheterogeneity analysis. *J. Chromatog.* 559:357–366.
12. Cifuentes A, Poppe H (1994). Simulation and optimization of peptide separation by capillary electrophoresis. *J. Chromatog. A.* 680:321–340.

13. Janini G M, Metral C J, Issaq H J, et al. (1999). Peptide mobility and peptide mapping in capillary zone electrophoresis. Experimental determination and theoretical simulation. *J. Chromatog. A*. 848:417–433.
14. Righetti P G, Gelfi C, Bossi A, et al. (2000). Capillary electrophoresis of peptides and proteins in isoelectric buffers: An update. *Electrophoresis*. 21:4046–4053.
15. Janini G M, Issaq H J (2001). Selection of buffer in capillary zone electrophoresis: Application to peptide and protein analysis. *Chromatographia*. 53:S18–S26.
16. Horvath J, Dolnik V (2001). Polymer wall coatings for capillary electrophoresis. *Electrophoresis*. 22:644–655.
17. Belder D, Deege A, Husmann H, et al. (2001). Cross-linked poly(vinyl alcohol) as permanent hydrophilic column coating for capillary electrophoresis. *Electrophoresis*. 22:3813–3818.
18. Li M X, Liu L, Wu J T, et al. (1997). Use of a polybrene capillary coating in capillary electrophoresis for rapid analysis of hemoglobin variants with on-line detection via an ion trap storage/reflectron time-of-flight mass spectrometer. *Analyt. Chem*. 69:2451–2456.
19. Erim F B, Cifuentes A, Poppe H, et al. (1995). Performance of a physically adsorbed high-molecular-mass polyethyleneimine layer as coating for the separation of basic proteins and peptides by capillary electrophoresis. *J. Chromatog. A*. 708:356–361.
20. Wang Y, Dubin P L (1999). Capillary modification by noncovalent polycation adsorption: Effects of polymer molecular weight and adsorption ionic strength. *Analyt. Chem*. 71:3463–3468.
21. Bendahl L, Hansen S H, Gammelgaard B (2001). Capillaries modified by noncovalent anionic polymer adsorption for capillary zone electrophoresis, micellar electrokinetic capillary chromatography and capillary electrophoresis mass spectrometry. *Electrophoresis*. 22:2565–2573.
22. Katayama H, Ishihama Y, Asakawa N (1998). Stable capillary coating with successive multiple ionic polymer layers. *Analyt. Chem*. 70:2254–2260.
23. Catai J R, Somsen G W, de Jong G J (2004). Efficient and reproducible analysis of peptides by capillary electrophoresis using noncovalently bilayer-coated capillaries. *Electrophoresis*. 25:817–824.
24. Catai J R, Tervahauta H A, de Jong G J, et al. (2005). Noncovalently bilayer-coated capillaries for efficient and reproducible analysis of proteins by capillary electrophoresis. *J. Chromatog. A*. 1083:185–192.
25. Righetti P G, Gelfi C, Verzola B, et al. (2001). The state of the art of dynamic coatings. *Electrophoresis*. 22:603–611.
26. Guttman A (1996). Capillary sodium dodecyl sulfate-gel electrophoresis of proteins. *Electrophoresis*. 17:1333–1341.
27. Takagi T (1997). Capillary electrophoresis in presence of sodium dodecyl sulfate and a sieving medium. *Electrophoresis*. 18:2239–2242.
28. Wu D, Regnier F E (1992). Sodium dodecyl sulfate-capillary gel electrophoresis of proteins using non-cross-linked polyacrylamide. *J. Chromatog*. 608:349–356.
29. Na D H, Cho C K, Youn Y S, et al. (2004). Capillary electrophoresis to characterize ricin and its subunits with matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Toxicon*. 43:329–335.
30. Mazzeo J R, Krull I S (1991). Capillary isoelectric focusing of proteins in uncoated fused silica capillaries using polymeric additives. *Analyt. Chem*. 63:2852–2857.
31. Fang X, Tragas C, Wu J, et al. (1998). Recent developments in capillary isoelectric focusing with whole-column imaging detection. *Electrophoresis*. 19:2290–2295.

32. Hjerten S, Liao J L, Yao K Q (1987). Theoretical and experimental study of high-performance electrophoretic mobilization of isoelectrically focused protein zones. *J. Chromatog.* 387:127–138.
33. Schwer C (1995). Capillary isoelectric focusing: a routine method for protein analysis? *Electrophoresis*. 16:2121–2126.
34. Righetti P G (2004). Determination of the isoelectric point of proteins by capillary isoelectric focusing. *J. Chromatog. A*. 1037:491–499.
35. Shimura K, Wang Z, Matsumoto H, et al. (2000). Synthetic oligopeptides as isoelectric point markers for capillary isoelectric focusing with ultraviolet absorption detection. *Electrophoresis*. 21:603–610.
36. Kilar F (2003). Recent applications of capillary isoelectric focusing. *Electrophoresis*. 24:3908–3916.
37. Shimura K (2002). Recent advances in capillary isoelectric focusing: 1997–2001. *Electrophoresis*. 23:3847–3857.
38. Terabe S (2004). Micellar electrokinetic chromatography. *Analyt. Chem.* 76:241A–246A.
39. Kats M, Richberg P C, Hughes D E (1995). Conformational diversity and conformational transitions of a monoclonal antibody monitored by circular dichroism and capillary electrophoresis. *Analyt. Chem.* 67:2943–2948.
40. Jing P, Kaneta T, Imasaka T (2002). Micellar electrokinetic chromatography with diode laser-induced fluorescence detection as a tool for investigating the fluorescence labeling of proteins. *Electrophoresis*. 23:550–555.
41. Strege M A, Lagu A L (1997). Capillary electrophoresis of biotechnology-derived proteins. *Electrophoresis*. 18:2343–2352.
42. Fu H, Huang X, Jin W, et al. (2003). The separation of biomolecules using capillary electrochromatrography. *Curr. Opin. Biotechnol.* 14:96–100.
43. Krull I, Sebag A, Stevenson R (2000). Specific applications of capillary electrochromatography to biopolymers, including proteins, nucleic acids, peptide mapping, antibodies, and so forth. *J. Chromatog. A*. 887:137–163.
44. Bandilla D, Skinner C D (2004). Capillary electrochromatography of peptides and proteins. *J. Chromatog. A*. 1044:113–129.
45. Dearie H S, Smith N W, Moffatt F, et al. (2002). Effect of ionic strength on perfusive flow in capillary electrochromatography columns packed with wide-pore stationary phases. *J. Chromatog. A*. 945:231–238.
46. Ericson C, Hjerten S (1999). Reversed-phase electrochromatography of proteins on modified continuous beds using normal-flow and counterflow gradients. Theoretical and practical considerations. *Analyt. Chem.* 71:1621–1627.
47. Zhang S, Huang X, Zhang J, et al. (2000). Capillary electrochromatography of proteins and peptides with a cationic acrylic monolith. *J. Chromatog. A*. 887:465–477.
48. Choudhary G, Apffel A, Yin H, et al. (2000). Use of on-line mass spectrometric detection in capillary electrochromatography. *J. Chromatog. A*. 887:85–101.
49. Klampfl C W (2004). Review coupling of capillary electrochromatography to mass spectrometry. *J. Chromatog. A*. 1044:131–144.
50. Stutz H (2005). Advances in the analysis of proteins and peptides by capillary electrophoresis with matrix-assisted laser desorption/ionization and electrospray-mass spectrometry detection. *Electrophoresis*. 26:1254–1290.
51. Tsuda T, Sweedler J V, Zare R N (1990). Rectangular capillaries for capillary zone electrophoresis. *Analyt. Chem.* 62:2149–2152.

52. Moring S E, Reel R T, van Soest R E J (1993). Optical improvements of a Z-shaped cell for high-sensitivity UV absorbance detection in capillary electrophoresis. *Analyt. Chem.* 65:3454–3459.
53. Heiger D N, Kaltenbach P, Sievert H J (1994). Diode array detection in capillary electrophoresis. *Electrophoresis.* 15:1234–1247.
54. Drossman H, Luckey J A, Kostichka A J, et al. (1990). High-speed separations of DNA sequencing reactions by capillary electrophoresis. *Analyt. Chem.* 62:900–903.
55. Liu J, Shirota O, Novotny M V (1992). Sensitive, laser-assisted determination of complex oligosaccharide mixtures separated by capillary gel electrophoresis at high resolution. *Analyt. Chem.* 64:973–975.
56. Swerdlow H, Zhang J Z, Chen D Y, et al. (1991). Three DNA sequencing methods using capillary gel electrophoresis and laser-induced fluorescence. *Analyt. Chem.* 63:2835–2841.
57. Tseng W L, Chang H T (2000). On-line concentration and separation of proteins by capillary electrophoresis using polymer solutions. *Analyt. Chem.* 72:4805–4811.
58. Yeung E S (1999). Study of single cells by using capillary electrophoresis and native fluorescence detection. *J. Chromatog. A.* 830:243–262.
59. Lin Y W, Chiu T C, Chang H T (2003). Laser-induced fluorescence technique for DNA and proteins separated by capillary electrophoresis. *J. Chromatog. B.* 793:37–48.
60. Schmitt-Kopplin P, Frommberger M (2003). Capillary electrophoresis-mass spectrometry: 15 years of developments and applications. *Electrophoresis.* 24:3837–3867.
61. Schmitt-Kopplin P, Englmann M (2005). Capillary electrophoresis-mass spectrometry: survey on developments and applications 2003–2004. *Electrophoresis.* 26:1209–1220.
62. Simpson D C, Smith R D (2005). Combining capillary electrophoresis with mass spectrometry for applications in proteomics. *Electrophoresis.* 26:1291–1305.
63. von Brocke A, Nicholson G, Bayer E (2001). Recent advances in capillary electrophoresis/electrospray-mass spectrometry. *Electrophoresis.* 22:1251–1266.
64. Na D H, Youn Y S, Lee S D, et al. (2003). Monitoring of peptide acylation inside degrading PLGA microspheres by capillary electrophoresis and MALDI-TOF mass spectrometry. *J. Control. Rel.* 92:291–299.
65. Preisler J, Hu P, Rejtar T, et al. (2000). Capillary electrophoresis-matrix-assisted laser desorption/ionization time-of-flight mass spectrometry using a vacuum deposition interface. *Analyt. Chem.* 72:4785–4795.
66. Tegeler T J, Mechref Y, Boraas K, et al. (2004). Microdeposition device interfacing capillary electrochromatography and microcolumn liquid chromatography with matrix-assisted laser desorption/ionization mass spectrometry. *Analyt. Chem.* 76:6698–6706.
67. Hunt G, Moorhouse K G, Chen A B (1996). Capillary isoelectric focusing and sodium dodecyl sulfate-capillary gel electrophoresis of recombinant humanized monoclonal antibody HER2. *J. Chromatog. A.* 744:295–301.
68. Michalski W P, Shiell B J (1999). Strategies for analysis of electrophoretically separated proteins and peptides. *Analyt. Chim. Acta.* 383:27–46.
69. Olivares J A, Nguyen N T, Yonker C R, et al. (1987). On-line mass spectrometric detection for capillary zone electrophoresis. *Analyt. Chem.* 59:1230–1232.
70. Stutz H, Bordin G, Rodriguez A R (2004). Capillary zone electrophoresis of metal-binding proteins in formic acid with UV- and mass spectrometric detection using cationic transient capillary isotachophoresis for preconcentration. *Electrophoresis.* 25:1071–1089.

71. Tsuji K, Baczynskyj L, Bronson G E (1992). Capillary electrophoresis-electrospray mass spectrometry for the analysis of recombinant bovine and porcine somatotropins. *Analyt. Chem.* 64:1864–1870.
72. Yeung B, Porter T J, Vath J E (1997). Direct isoform analysis of high-mannose-containing glycoproteins by on-line capillary electrophoresis electrospray mass spectrometry. *Analyt. Chem.* 69:2510–2516.
73. Na D H, DeLuca P P, Lee K C (2004). Direct determination of the peptide content in microspheres by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Analyt. Chem.* 76:2669–2673.
74. Rush R S, Derby P L, Strickland T W, et al. (1993). Peptide mapping and evaluation of glycopeptide microheterogeneity derived from endoproteinase digestion of erythropoietin by affinity high-performance capillary electrophoresis. *Analyt. Chem.* 65:1834–1842.
75. Zhou G H, Luo G A, Zhou Y, et al. (1998). Application of capillary electrophoresis, liquid chromatography, electrospray-mass spectrometry and matrix-assisted laser desorption/ionization-time of flight-mass spectrometry to the characterization of recombinant human erythropoietin. *Electrophoresis.* 19:2348–2355.
76. Boss H J, Watson D B, Rush R S (1998). Peptide capillary zone electrophoresis mass spectrometry of recombinant human erythropoietin: an evaluation of the analytical method. *Electrophoresis.* 19:2654–2664.
77. Issaq H J, Chan K C, Janini G M, et al. (2005). Multidimensional separation of peptides for effective proteomic analysis. *J. Chromatog. B.* 817:35–47.
78. Hynek R, Kasicka V, Kucerova Z, et al. (1996). Application of reversed-phase high-performance liquid chromatography and capillary zone electrophoresis to the peptide mapping of pepsin isoenzymes. *J. Chromatog. B.* 681:37–45.
79. Wu S L (1997). The use of sequential high-performance liquid chromatography and capillary zone electrophoresis to separate the glycosylated peptides from recombinant tissue plasminogen activator to a detailed level of microheterogeneity. *Analyt. Biochem.* 253:85–97.
80. Issaq H J, Chan K C, Janini G M, et al. (1999). A simple two-dimensional high performance liquid chromatography/high performance capillary electrophoresis set-up for the separation of complex mixtures. *Electrophoresis.* 20:1533–1537.
81. Kang S H, Gong X, Yeung E S (2000). High-throughput comprehensive peptide mapping of proteins by multiplexed capillary electrophoresis. *Analyt. Chem.* 72:3014–3021.
82. He Y, Zhong W W, Yeung E S (2002). Multiplexed on-column protein digestion and capillary electrophoresis for high-throughput comprehensive peptide mapping. *J. Chromatog. B.* 782:331–341.
83. Catai J R, De Jong G J, Somsen G W (2005). *Methods for Structural Analysis of Protein Pharmaceuticals*, AAPS Press, Arlington, V.A. pp. 331–377.
84. Kasicka V (2003). Recent advances in capillary electrophoresis and capillary electrochromatography of peptides. *Electrophoresis.* 24:4013–4046.
85. Hutterer K, Dolnik V (2003). Capillary electrophoresis of proteins 2001–2003. *Electrophoresis.* 24:3998–4012.
86. Ridge S, Hettiarachchi K (1998). Peptide purity and counter ion determination of bradykinin by high-performance liquid chromatography and capillary electrophoresis. *J. Chromatog. A.* 817:215–222.
87. Chen J, Fausnaugh-Pollitt J, Gu L (1999). Development and validation of a capillary electrophoresis method for the characterization of protegrin IB-367. *J. Chromatog. A.* 853:197–206.

88. Hettiarachchi K, Ridge S, Thomas D W, et al. (2001). Characterization and analysis of biphalin: an opioid peptide with a palindromic sequence. *J. Peptide Res.* 57:151–161.
89. Moumakwa B A, Crawley C D, Purich E, et al. (2005). Use of capillary electrophoresis in drug quality assessment of synthetic porcine secretin. *Biomed. Chromatog.* 19:68–79.
90. Mandrup G (1992). Rugged method for the determination of deamidation products in insulin solutions by free zone capillary electrophoresis using an untreated fused-silica capillary. *J. Chromatog.* 604:267–281.
91. Yomota C, Matsumoto Y, Okada S, et al. (1997). Discrimination limit for purity test of human insulin by capillary electrophoresis. *J. Chromatog. B*. 703:139–145.
92. Sergeev N V, Gloukhova N S, Nazimov I V, et al. (2001). Monitoring of recombinant human insulin production by narrow-bore reversed-phase high-performance liquid chromatography, high-performance capillary electrophoresis and matrix-assisted laser desorption ionization time-of-flight mass spectrometry. *J. Chromatog. A*. 907:131–144.
93. Hunt G, Nashabeh W (1999). Capillary electrophoresis sodium dodecyl sulfate nongel sieving analysis of a therapeutic recombinant monoclonal antibody: a biotechnology perspective. *Analyt. Chem.* 71:2390–2397.
94. Solinova V, Kasicka V, Barth T, et al. (2005). Analysis and separation of enkephalin and dalargin analogues and fragments by capillary zone electrophoresis. *J. Chromatog. A*. 1081:9–18.
95. Eriksson H J, Somsen G W, Hinrichs W L, et al. (2001). Characterization of human placental alkaline phosphatase by activity and protein assays, capillary electrophoresis and matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *J. Chromatog. B*. 755:311–319.
96. Roddy T P, Molnar T E, McKean R E, et al. (1997). Method of analysis of recombinant acidic fibroblast growth factor by capillary electrophoresis. *J. Chromatog. B*. 695:49–58.
97. Donges R, Brazel D (2002). Determination of recombinant hirudin structural deviants by capillary zone electrophoresis augmented with buffer additives. *J. Chromatog. A*. 979:217–226.
98. Klyushnichenko V E, Koulich D M, Yakimov S A, et al. (1994). Recombinant human insulin: III. High-performance liquid chromatography and high-performance capillary electrophoresis control in the analysis of step-by-step production of recombinant human insulin. *J. Chromatog. A*. 661:83–92.
99. Gysler J, Mazereeuw M, Helk B, et al. (1999). Utility of isotachophoresis-capillary zone electrophoresis, mass spectrometry and high-performance size-exclusion chromatography for monitoring of interleukin-6 dimer formation. *J. Chromatog. B*. 841:63–73.
100. Klyushnichenko V, Tishkov V, Kula M R (1997). Rapid SDS-Gel capillary electrophoresis for the analysis of recombinant NADP(+)-dependent formate dehydrogenase during expression in Escherichia coli cells and its purification. *J. Biotechnol.* 58:187–195.
101. Kakehi K, Honda S (1996). Analysis of glycoproteins, glycopeptides and glycoprotein-derived oligosaccharides by high-performance capillary electrophoresis. *J. Chromatog. A*. 720:377–393.
102. Krantz S B (1991). Erythropoietin. *Blood*. 77:419–434.
103. Lai P H, Everett R, Wang F F, et al. (1986). Structural characterization of human erythropoietin. *J. Biolog. Chem.* 261:3116–3121.

104. Davis J M, Arakawa T, Strickland T W, et al. (1987). Characterization of recombinant human erythropoietin produced in Chinese hamster ovary cells. *Biochem.* 26:2633–2638.

105. Yu B, Cong H, Liu H, et al. (2005). Separation and detection of erythropoietin by CE and CE–MS. *Trends Analyt. Chem.* 24:350–357.

106. Bietlot H P, Girard M (1997). Analysis of recombinant human erythropoietin in drug formulations by high-performance capillary electrophoresis. *J. Chromatog. A*. 759:177–184.

107. Kinoshita M, Murakami E, Oda Y, et al. (2000). Comparative studies on the analysis of glycosylation heterogeneity of sialic acid-containing glycoproteins using capillary electrophoresis. *J. Chromatog. A*. 866:261–271.

108. De Frutos M, Cifuentes A, Diez-Masa J C (2003). Differences in capillary electrophoresis profiles of urinary and recombinant erythropoietin. *Electrophoresis.* 24:678–680.

109. Sanz-Nebot V, Benavente F, Vallverdu A, et al. (2003). Separation of recombinant human erythropoietin glycoforms by capillary electrophoresis using volatile electrolytes. Assessment of mass spectrometry for the characterization of erythropoietin glycoforms. *Analyt. Chem.* 75:5220–5229.

110. Sanz-Nebot V, Benavente F, Gimenez E, et al. (2005). Capillary electrophoresis and matrix-assisted laser desorption/ionization-time of flight-mass spectrometry for analysis of the novel erythropoiesis-stimulating protein (NESP). *Electrophoresis.* 26:1451–1456.

111. Cifuentes A, Moreno-Arribas M V, de Frutos M, et al. (1999). Capillary isoelectric focusing of erythropoietin glycoforms and its comparison with flat-bed isoelectric focusing and capillary zone electrophoresis. *J. Chromatog. A*. 830:453–463.

112. Lopez-Soto-Yarritu P, Diez-Masa J C, Cifuentes A, et al. (2002). Improved capillary isoelectric focusing method for recombinant erythropoietin analysis. *J. Chromatog. A*. 968:221–228.

113. European Pharmacopoeia (2002). 4:1123–1128, 1316.

114. Watson E, Yao F (1993). Capillary electrophoretic separation of human recombinant erythropoietin (r-HuEPO) glycoforms. *Analyt. Biochem.* 210:389–393.

115. Hunt G, Hotaling T, Chen A B (1998). Validation of a capillary isoelectric focusing method for the recombinant monoclonal antibody C2B8. *J. Chromatog. A*. 800: 355–367.

116. Tang S, Nesta D P, Maneri L R, et al. (1999). A method for routine analysis of recombinant immunoglobulins (rIgGs) by capillary isoelectric focusing (cIEF). *J. Pharmac. Biomed. Anal.* 19:569–583.

117. Santora L C, Krull I S, Grant K (1999). Characterization of recombinant human monoclonal tissue necrosis factor-α antibody using cation-exchange HPLC and capillary isoelectric focusing. *Analyt. Biochem.* 275:98–108.

118. Kats M, Richberg P C, Hughes D E (1997). pH-dependent isoform transitions of a monoclonal antibody monitored by micellar electrokinetic capillary chromatography. *Analyt. Chem.* 69:338–343.

119. Kats M, Richberg P C, Hughes D E (1997). Conformational transformations of the fusion protein BR96 sFv-PE40 as monitored by micellar electrokinetic capillary chromatography and circular dichroism. *J. Chromatog. A*. 766:205–213.

120. Reubsaet J L, Beijnen J H, Bult A, et al. (1998). Analytical techniques used to study the degradation of proteins and peptides: chemical instability. *J. Pharmac. Biomed. Anal.* 17:955–978.

121. Reubsaet J L, Beijnen J H, Bult A, et al. (1998). Analytical techniques used to study the degradation of proteins and peptides: physical instability. *J. Pharmac. Biomed. Anal.* 17:979–984.

122. International conferences on harmonization of technical requirements for registration of pharmaceuticals for human use. In Q5C, ICH Harmonised Tripartite Guideline, Quality of Biotechnological Products: Stability Testing Biotechnological/Biological Products. Available at: http://www.ich.org.

123. Hoitink M A, Hop E, Beijnen J H, et al. (1997). Capillary zone electrophoresis–mass spectrometry as a tool in the stability research of the luteinising hormone-releasing hormone analogue goserelin. *J. Chromatog. A*. 776:319–327.

124. Lai M, Skanchy D, Stobaugh J, et al. (1998). Capillary electrophoresis separation of an asparagine containing hexapeptide and its deamidation products. *J. Pharmac. Biomed. Anal.* 18:421–427.

125. Harris J M, Chess R B (2003). Effect of pegylation on pharmaceuticals. *Nature Rev. Drug Disc.* 2:214–221.

126. Na D H, Youn Y S, Park E J, et al. (2004). Stability of PEGylated salmon calcitonin in nasal mucosa. *J. Pharmac. Sci.* 93:256–261.

127. Na D H, DeLuca P P (2005). PEGylation of octreotide: I. Separation of positional isomers and stability against acylation by poly(D,L-lactide-co-glycolide). *Pharmac. Res.* 22:736–742.

128. Lee K C, Moon S C, Park M O, et al. (1999). Isolation, characterization, and stability of positional isomers of mono-PEGylated salmon calcitonins. *Pharmac. Res.* 16:813–818.

129. Na D H, Youn Y S, Lee K C (2003). Optimization of the PEGylation process of a peptide by monitoring with matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Rapid Comm. Mass Spectrom.* 17:2241–2244.

130. Na D H, Lee K C, DeLuca P P (2005). PEGylation of octreotide: II. Effect of N-terminal mono-PEGylation on biological activity and pharmacokinetics. *Pharmac. Res.* 22:743–749.

131. Bailon P, Berthold W (1998). Polyethylene glycol-conjugated pharmaceutical proteins. *Pharmac. Sci. Technol. Today.* 1:352–356.

132. Cunico R L, Gruhn V, Kresin L, et al. (1991). Characterization of polyethylene glycol modified proteins using charge-reversed capillary electrophoresis. *J. Chromatog.* 559:467–477.

133. Bullock J, Chowdhury S, Johnston D (1996). Characterization of poly(ethylene glycol)-modified superoxide dismutase: Comparison of capillary electrophoresis and matrix-assisted laser desorption/ionization mass spectrometry. *Analyt. Chem.* 68:3258–3264.

134. Li W, Zhong Y, Lin B, Su Z (2001). Characterization of polyethylene glycol-modified proteins by semi-aqueous capillary electrophoresis. *J. Chromatog. A*. 905:299–307.

135. Na D H, Park M O, Choi S Y, et al. (2001). Identification of the modifying sites of mono-PEGylated salmon calcitonins by capillary electrophoresis and MALDI-TOF mass spectrometry. *J. Chromatog. B*. 754:259–263.

136. Na D H, Lee K C (2004). Capillary electrophoretic characterization of PEGylated human parathyroid hormone with matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Analyt. Biochem.* 331:322–328.

137. Snider J, Neville C, Yuan L C, et al. (1992). Characterization of the heterogeneity of polyethylene glycol-modified superoxide dismutase by chromatographic and electrophoretic techniques. *J. Chromatog.* 599:141–155.

138. Roberts M J, Harris J M (1998). Attachment of degradable poly(ethylene glycol) to proteins has the potential to increase therapeutic efficacy. *J. Pharmac. Sci.* 87:1440–1445.

139. Aebersold R, Mann M (2003). Mass spectrometry-based proteomics. *Nature*. 422: 198–207.

140. Gygi S P, Corthals G L, Zhang Y, et al. (2000). Evaluation of two-dimensional gel electrophoresis-based proteome analysis technology. *Proc. of the National Academy of Sciences of the United States of America*. 97:9390–9395.

141. Washburn M P, Wolters D, Yates III J R (2001). Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nature Biotechnol.* 19:242–247.

142. Issaq H J (2001). The role of separation science in proteomics research. *Electrophoresis*. 22:3629–3638.

143. Shen Y, Smith R D (2002). Proteomics based on high-efficiency capillary separations. *Electrophoresis*. 23:3106–3124.

144. Chen J, Lee C S, Shen Y, et al. (2002). Integration of capillary isoelectric focusing with capillary reversed-phase liquid chromatography for two-dimensional proteomics separation. *Electrophoresis*. 23:3143–3148.

145. Mohan D, Lee C S (2002). On-line coupling of capillary isoelectric focusing with transient isotachophoresis-zone electrophoresis: A two-dimensional separation system for proteomics. *Electrophoresis*. 23:3160–3167.

146. Ramsey J D, Jacobson S C, Culbertson C T, et al. (2003). High-efficiency, two-dimensional separations of protein digests on microfluidic devices. *Analyt. Chem.* 75:3758–3764.

# PHARMACEUTICAL BIOASSAY

Jun Li, Ming Zhao, and Shiqi Peng
*School of Chemical Biology and Pharmaceutical Sciences, Capital Medical University, Beijing, China*

**Chapter Contents**

1 Bioassay for Detecting Toxicity of Toxins from a Microorganism
- 1.1 Colorimetric Yeast Assay for Detecting Trichothecene Mycotoxins
- 1.2 Ciliate *T. thermophila* Assay for Detecting Trichothecene Mycotoxins
- 1.3 MTT Assay for Detecting Fusarium Mycotoxins
- 1.4 Mortality and Frass Production Assay for Detecting Toxicity of Bacterial Strains
- 1.5 *L. sativum* Assay for Detecting Microcystin Toxicity

2 Toxicity Bioassay for Chemicals
- 2.1 Lux-Fluoro Assay for Detecting Combined Genotoxicity and Cytotoxicity of Chemicals
- 2.2 ALIC-Based Assay for Detecting Toxicity of Chemicals in SPM
- 2.3 PWM-Induced IgM Assay for Detecting Toxicity of Chemicals
- 2.4 GFP-Based Cell Assay for Detecting Toxicity of TCDD and Related Chemicals
- 2.5 RACB Protocol for Detecting Ovarian Toxicity of Xenobiotics
- 2.6 Uroepithelial Cell Assay for Detecting Urotoxicity of Chemicals
- 2.7 Mating Efficiency Assay for Detecting Toxicity of Endocrine Disruptors
- 2.8 Pnar-gfp Assay for Detecting Carcinogenic Toxicity of Nitrate

3 Antiviral and Anticancer Assay
- 3.1 Human Breast Cancer Cell Assay for Detecting Anticancer Activity
- 3.2 Ishikawa Cell and Rat Assay for Detecting Antiestrogens
- 3.3 RT-PCR and Swine Assay for Detecting Anti-HEV Antibody
- 3.4 Antitubercular and Cytotoxic Assay
- 3.5 DNA Polymerase β Lyase Assay
- 3.6 HIV-1 Protease and Reverse Transcriptase Kinetic Assay

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

3.7 P-Glycoprotein Pump Assay
3.8 KIRA Assay

4 Hepatoxicity and Hepatoprotective Assay

4.1 GST-P Enzyme-Altered Foci Assay
4.2 Partial Hepatectomy Assay
4.3 Hepatoprotective Assay
4.4 Immunological Assay
4.5 Body Weight-Based Liver Tumor Incidence Assay
4.6 Hepatocyte Primary Culture Assay
4.7 CCL-64 Cell Growth Assay

5 Anti-Inflammatory Assay

5.1 Adhesion Formation Assay
5.2 PBMC Proliferation Assay
5.3 COX-1, COX-2, and 5-LOX Assay
5.4 CCR5 Receptor Binding Assay
5.5 Tissue Binding Affinity Assay
5.6 G93A-SOD1 Transgenic Mice Assay
5.7 MAPK p44 (ERK1) and p42 (ERK2) Assay
5.8 ELA4.NOB-1/CTLL Cell Assay
5.9 FKBP51mRNA Assay
5.10 Xylene-Induced Ear Edema Assay

6 Thrombus-Related Assay

6.1 Antiplatelet Aggregation Assay
6.2 Ferric Chloride-Induced Thrombosis Assay
6.3 Electrical Stimulation-Induced Arterial Thrombosis Assay
6.4 Thread-Induced Thrombosis Assay
6.5 Euglobulin Clot Lysis Time (ECLT) Assay
6.6 Fibrinolytic Area Assay
6.7 Thrombolytic Assay

7 Immunomodulating Assay

7.1 Rat Mast Cell and Rabbit Aortic Assay
7.2 Dendritic Cell (DC) Assay
7.3 Lymphoid Organ Assay
7.4 IFN-γ Assay
7.5 Anti-rHuEPO NAb Assay
7.6 Radial Assay of Chemotaxis

8 Estrogen Assay

8.1 Yeast Oestrogen Assay
8.2 HELN α and HELN β Transfected Cell Assay
8.3 Luciferase Assay
8.4 Green Fluorescent Protein Expression Assay

9 Antimalarial Assay

9.1 Plasmodium Falciparum and Murine P388 Leukemia Cell Assay
9.2 Histidine-Rich Protein II Assay
9.3 Antimalarial Activity Assay
9.4 Survival of *Anopheles gambiae* Assay
9.5 Chloroquine Assay
9.6 Plasmodium Falciparum Clone and Dihydroartemisinin (DHA) Assay

10 Blood Pressure-Related Assay
10.1 Human Plasma Assay
10.2 Pulmonary Hypertension Assay
10.3 Coronary Arteries (CA) Constriction Assay
10.4 Middle Cerebral Artery Occlusion (MCAO) Assay and Vascular Heparan Sulfate Proteoglycans (HSPG) Perlecan Assay
10.5 Temperature Assay in Awake Subjects
References

## 1 BIOASSAY FOR DETECTING TOXICITY OF TOXINS FROM A MICROORGANISM

Some *Fusarium* mycotoxins and the metabolites of filamentous fungi and molds occur naturally in a variety of feeds, foodstuffs, crops, vegetables, fruits, water, and plants. Ingestion of the contaminative feeds, foodstuffs, crops, vegetables, fruits, water, and plants can produce health problems, such as reduced growth, anorexia, induced emesis, enhanced proliferation of estrogen-responsive tumor cells, human cervical cancer, and premature initial breast development in humans and animals. Among the common toxins, verrucarin A, T2-toxin, roridin A, deoxynivalenol (DON), zearalenone (ZEN), fumonisin B1 (FB1), moniliformin (MON), microcystin-LR (MCLR), and the toxins from *Bacillus thuringiensis* provide a worldwide threat to humans and animals. Based on these facts that the lactose-utilizing yeast *Kluyveromyces marxianus* exhibits well-characterized β-galactosidase activity and has chromogenic substrates, the toxicants interfering with any cellular function required for induction and expression of β-galactosidase gene can suppress β-galactosidase activity, DON and fumonisin B can inhibit cell growth, the dehydrogenase ensurveyed enzymes involved in mitochondrial functions can be inhibited by the toxicants, in axenic culture the melanin precursor overproducer mutant strain of the ciliate *Tetrahymena thermophila* can be grown to high concentration, in protecting plants from the hazardous effects of xenobiotics glutathione S-transferase (GST) and glutathione peroxidase (GPX) can play essential roles, colorimetric yeast assay [1], ciliate *T. thermophila* assay [2], MTT assay [3], mortality and frass production assay [4], and *Lepidium sativum* assay [5] are established.

### 1.1 Colorimetric Yeast Assay for Detecting Trichothecene Mycotoxins

To 50 mL of liquid medium, a single-cell colony of *K. marxianus* GK1005 (maintained routinely and grown on 1% (w/v) yeast extract, 1% (w/v) bacteriological peptone, and 2% (w/v) glucose (YPD), solidified when required with 2% (w/v) agar) on agar plate is transfered to prepare cultures that are incubated for 16 h at 35°C and 200 rpm in a rotary incubator. At 560 nm, the absorbance is measured to determine cell density, which is calibrated by direct hemocytometer counts. Mycotoxins containing sample is dissolved in spectroscopic grade methanol to make stock solutions typically at 0.1 mg/mL. Cetyl trimethyl ammonium bromide, polymyxin B sulphate, and polymyxin B nonapeptide are dissolved in

water, filter-sterilized, and kept no more than a day as stock solutions. 5-Bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-gal) is dissolved in dimethylformamide (DMF) at 100 mg/mL and stored at –20°C in the dark, and before each assay, it is immediately diluted in aqueous DMF (2 parts water : 3 parts DMF) to prepare a working solution of 20-mg/mL X-gal assay. Before use *o*-nitrophenyl-β-D-galactopyranoside (ONPGal) is immediately dissolved in water at 4 mg/mL and subsequently discarded. Before use, MTT is dissolved in phosphate-buffered solution (PBS) and filter-sterilized to prepare a 5-mg/mL stock solution.

In the wells of a microtiter plate (sterile, flat-bottomed), 136 μL of growth medium consisted of 1% (w/v) yeast extract, 1% (w/v) bacteriological peptone, and 50-mM glucose—"YPD-50" are mixed with polymyxin B sulphate to give a final assay concentration of 15 μg/mL, to which 8 μL of stock solution of mycotoxin containing sample or methanol (controls) and 16 μL of yeast inoculum are successively added to give an initial assay cell density of approximately $2 \times 10^8$ cells/mL. Blank wells contain 152 μL of medium and 8 μL of methanol. The plate is mixed, sealed with a plate sealer, and incubated at 35°C for the duration of the assay, during which cell density is regularly monitored at 560 nm. After about 10 h, the control (mycotoxin-free) cultures reach the stationary phase; the activity of β-galactosidase or mitochondrial (MTT-cleavage) in the cultures is assayed.

To each well of the microtiter plate, 5 μL of sodium dodecyl sulfate (SDS) (0.1%, w/v) and 3 μL of chloroform are added to permeabilize the cells. For *in vivo* experiments and experiments examining the effects of different carbon sources, 1 μL of 100-mg/mL X-gal in DMF are added to each well. For examining the effects of altering glucose concentration and inoculum cell density, 5 μL of 20-mg/mL X-gal in DMF are added to each well. For methanol and ethanol toxicity experiments and the standardized bioassay, 8 μL of 20-mg/mL X-gal in aqueous DMF are added. After mixing the contents in the wells, the plates are incubated at 35°C for a maximum of 30 min. Using a test filter at 666 nm and a reference filter at 560 nm, the plates are read. β-Galactosidase activity is expressed as product formation ($A_{666}$–$A_{560}$), as a function of cell density ($A_{560}$) as well.

After the addition of 16 μL of MTT in PBS to each well and statically incubated at 35°C for 4 h, the medium is displaced by 200 μL of dimethyl sulfate (DMSO). Thorough mixing and dissolution the MTT-cleavage product is pipeted repeatedly. Using a test filter at 560 nm and a reference filter at 666 nm, the plates are read.

Into 50 mL of the solution consisting of 1% yeast extract, 1% bacteriological peptone, and 50-mM lactose, a single colony of *K. marxianus* GK1005 is inoculated and then incubated for 16 h at 35°C and 200 rpm in an orbital shaker. Into each of two universal bottles 10 mL of the samples of the culture are transferred, and then 0.2 mL of chloroform and 0.1 mL of SDS (0.1%, w/v) are added. To permeabilize the cells, to one bottle 20 μL of X-gal (100 mg/mL in DMF) are added, and to the other bottle 20 μL of DMF are added. In an orbital shaker, both bottles are incubated at 35°C and 200 rpm until in the bottle containing X-gal visible indigo precipitate is formed, of which the absorption spectrum is determined with DMF control sample as a reference.

### 1.2 Ciliate *T. thermophila* Assay for Detecting Trichothecene Mycotoxins

The phenotype characteristics of *T. thermophila* strain BI3840 are amicronucleated (amc), pigment producing (pig), and resistant to 25 μg/mL of cycloheximide (cy-r)

and mating type IV [6]. The PP210 medium consists of aqueous solution of proteose peptone (2%, w/v), supplemented with 10 μM of $FeCl_3$, 250 μg/mL of streptomycin sulphate, and penicillin G. Mycotoxins containing samples are dissolved in propyleneglycol or acetonitrile to make stock solutions at 5 mg/mL. *T. thermophila* strain BI3840 are grown axenically in PP210 medium for 24 h and maintained constantly at 28 ± 1°C. Then the exponential phase cellular suspensions are distributed in microtiter plates (100 μL/well). According to the required final concentration, different dilutions of the stock solutions of mycotoxins containing samples in PP210 medium are added into the cultures (100 μL/well). To each 100 μl/well of *T. thermophila* strain BI3840 culture, 100 μl/well of PP210 medium without a mycotoxin containing sample is added to obtain one type of control, and 100 μl/well of PP210 medium plus the highest concentration of the corresponding solvent added to obtain another type of control. In all cases, the microtiter plates are incubated at 28 ± 1°C for 48 h and the wells with a dark-brown pigmentation are considered as the positive growth wells.

### 1.3 MTT Assay for Detecting Fusarium Mycotoxins

Caco-2, C5-O, V79 cells (in Dulbecco's Modified Eagle Medium, DMEM), and Chinese hamster ovary (CHO)-K1 cells (in DMEM/F-12) at passage numbers between 30 and 50, and HepG2 cells (in Minimum Essential Medium) at passage numbers between 80 and 100, are grown as monolayers in 80-$cm^2$ culture flasks, which are harvested when they reach 80% confluence to maintain exponential growth.

Using trypsin ethylenediaminetetraacetic acid (EDTA) (0.25% trypsin and 1-mM EDTA·4Na), the cell monolayers in exponential growth are harvested and after repeated pipetting the single-cell suspensions are obtained. To 96 well plates, the single cell suspensions ($1 \times 10^2$ to $5 \times 10^4$ cells per 200 μL medium/well) are added by serial dilution. The medium are supplemented with 1.5-g/L sodium bicarbonate, 0.11-g/L sodium pyruvate, 1% nonessential amino acid, 25-mM HEPES, 100 U of penicillin/mL, 100 μg of streptomycin/mL, 25 ng of amphotericin B/mL, and 10% fetal bovine serum). Caco-2, C5-O, HepG2 ($1 \times 10^4$ cells/100-μL medium), CHO-K1, and V79 ($5 \times 10^3$ cells/100-μL medium) are seeded to each well of the 96 well plates and incubated in a humidified atmosphere of 5% $CO_2$ at 37°C for 24 h. A mycotoxin containing sample in 100 μL of medium is added from high to low concentration to the wells and adjusted to final concentration needed. MTT is dissolved in PBS (5-mg/mL final concentration), filtered (0.22-μm filter), and stored in the amber vials at 4°C for a month. After 48-h and 72-h incubation, to each well, 25 μL of MTT solution is added and incubated in a humidified atmosphere of 5% $CO_2$ at 37°C for 4 h. At the end of the incubation period, the media are discarded using a suction pump. The extraction buffer of 20% (w/v) SDS in a solution of 50% DMF in demineralized water (50:50, v/v) is prepared at pH 4.7 and filtered (0.22-μm filter). The buffer in 100 μL of 20% (w/v) of SDS [7] is extracted and added into each well to solubilize formazan crystals, and the plates are shaked at 37°C overnight. The absorbance is measured at 570 nm with 690 nm as a reference wavelength. The positive control contains an adjusted seeding cell number in the log phase, of which the culture medium contains 0.1% ethanol.

### 1.4 Mortality and Frass Production Assay for Detecting Toxicity of Bacterial Strains

*B. thuringiensis* subsp. *tenebrionis* and the six unidentified strains of *B. thuringiensis,* labeled A30, A299, A311, A409, A410, and A429, are grown on nutrient agar buffered with an equimolar concentration of $KH_2PO_4$ (50 mM, pH 7.0) at 30°C for 5 days. The bacterial culture consists of vegetative cells, sporangia, spores, and crystals. These stages are lyophilized and stored at –20°C. The spore-crystal suspensions of each strain with 1 : 1 in the spore : crystal ratio are used in the toxicity assays [8]. The concentration of extractable proteins (the final protein solution is referred to as extractable protein) is determined and adjusted to a given concentration of protein in the crystal-spore mixture. Spore-crystal mixtures are suspended in Triton X-100 solution (0.01%, v/v) and washed three times by centrifugation at 22°C and 11750 g for 3 min. To extract the proteins, the final pellet in a putatively selective crystal solubilizing buffer (40.5-mM $Na_2CO_3$, 0.5-mM phenylmethylsulfonylfluoride, 0.1-mM dithiothreitol, pH 10.0) is solubilized at 42°C by incubating for 2 h and vortexing every 30 min [9]. The suspension is centrifuged (22°C, 11750 g, 3 min), and the supernatant is assayed for total protein with bovine serum albumin as the standard [10]. The laboratory cultures of plants are periodically supplemented with pest insects to maintain hybrid vigor. The colony is reared in an incubator at 24°C under the light for 16 h and at 16°C in the darkness for 8 h. Abundant food is placed in plastic petri dishes (100 mm in diameter) with moist filter paper that contains 10 adults, the plastic petri dishes are incubated, and the number of frass pellets on filter paper is counted daily for 3 days.

Adult mortality and modified frass production assays are used to measure the toxicity of the extractable proteins of the bacterial strains against plants [11]. The modified frass production assay is regarded as a rapid method to determine the specificity of numerous bacterial strains toxins against pest insect. The food cylinders (6 mm in length, 36 mm in diameter, 10 cylinders/plate) of optimum medium for insect feeding are cut from foliage homogenate supplemented with agar. The cylinders are dipped into distilled water containing different concentration of the extractable proteins, agitated for 5 s, air-dried, and then coated with test spore-crystal suspensions. The concentration of extractable proteins is 75-, 150-, 225-, and 300-μg protein/mL with distilled water as a control. Three to eight days after eclosion, starved adults are collected and added to the diet in petri dishes (15 × 100 mm diameter) containing moistened filter paper (15 adult insects per plate per dose). In the mortality assay, insects are fed at 25°C in darkness for 8 days. Mortality is monitored daily, and the $LC_{50}$ for each test strain is calculated by probit analysis. In the frass assay, insects are fed for 6 days. The total number of frass pellets is calculated.

### 1.5 *L. sativum* Assay for Detecting Microcystin Toxicity

Toxic *Microcystis aeruginosa* PCC 7806 is cultured, and sufficient aeration is provided for culture mixing and air supply. From the freeze-dried cells, toxin is extracted by 70% aqueous methanol and determined using the protein phosphatase inhibition assay from the difference in the change in absorbance at 410 nm [12, 13]. With 5% $H_2O_2$ for 5 min, the seeds of *L. sativum* are surface sterilized, washed three times with sterile water for 10 min, left in fresh water overnight to germinate,

and then transferred to nutrient medium containing 6.5% N, 2.7% P, 13% K, 7% Ca, 2.2% Mg, 7.5% S, 0.15% Fe, 0.024% Mn, 0.0024% B, 0.005% Zn, 0.002% Cu, and 0.001% Mo in 0.8% agar. The container (12 seeds/container) is cultured at 27°C under continuous light (20 photons/$m^2$/s). Before pouring, either 1- or 10-μg/L microcystin-LR (MCLR) toxin extracts are added to the medium. The fresh weights, root, and leaf lengths are measured daily in 2 days. GST and GPX activities are determined in 3 days.

The stems and roots of 10 plants from each of three containers from each group are prepared for 30% homogenate of plant material in buffer (0.01-M Tris-HCl pH 7.8, 7.5-mM PMSF, 2.5-mM EDTA, 325-μM bestatin, 3.5-μM E-64, 2.5-μM leupeptin, and 0.75-μM aprotinin). The homogenate is centrifuged at 4°C and 12,000 g in a benchtop centrifuge for 2 h, and the supernatant is stored at −80°C for the GST and GPX assay.

In the GST assay, to the supernatant fluid, a cocktail containing 0.1-M potassium phosphate (pH 6.5), 30-mM 1-chloro-2,4-dinitrobenzene and 20-mM glutathione are added and measured immediately at 340 nm for 4 min [14]. In the GPX assay, to the supernatant fluid, a cocktail of 0.1-mM potassium phosphate buffer (pH 7), glutathione reductase (4 ng/μL final), 10-mM glutathione, 30-mM EDTA, and 1-mM NADPH is added. The reaction is initated by adding cumene hydroperoxide and measured immediately at 340 nm for 3 min.

## 2 TOXICITY BIOASSAY FOR CHEMICALS

Environmental chemical contaminants as toxic agents may be suspended in air, absorbed in suspended particulate matter (SPM) and foods, and dissolved in water, and thus, they may result in several toxic effects such as DNA damage, decline in lung function, imbalance of immune function, activation of aryl hydrocarbon receptor (AhR), abnormal reproduction, and urinary bladder injuries. Lux-Fluoro assay is developed on specific lesions for cells induced by DNA damaging chemicals may lead to cell death or induce an error-prone repair pathway leading to mutagenesis and cancer induction [15]. Based on the air-liquid interface culture (ALIC) of the human alveolar type II cell line (A459) and culture of spleen cells obtained from BALB/c mice producing IgM after pokeweed mitogen (PWM) stimulation, ALIC-based assay and PWM-induced IgM assay are established [16, 17]. Exposure to extremely low concentrations of 2,3,7,8-tetrachlorodibenzo-p-dioxin (TCDD) results in activation of AhR and the induction of expression of a battery of genes, green fluorescent protein (GFP)-based cell assay is established [18]. In Reproductive Assessment by Continuous Breeding (RACB) protocol, differential follicle counts can provide a sensitive means of estimating the extent of ovarian toxicity in females exposed to xenobiotics [19]. Based on these facts that urinary bladder injuries induced by repeated oral administration of pharmaceuticals may be epithelial ulceration with edema and hemorrhage in the lamina propria and muscle layer of the urinary bladder, environmental chemical contaminants disrupt mammalian peptide signal transduction pathway via the interaction with the endocrine system in humans and various wildlife species and the cellular production of GFP is a function of nitrate concentration, uroepithelial cell assay [20], mating efficiency assay [21], and Pnar-gfp assay [22] are established and can be used in bioassays.

### 2.1 Lux-Fluoro Assay for Detecting Combined Genotoxicity and Cytotoxicity of Chemicals

The plasmid pPLS-1 (DSM10333), the genotoxicity sensing reporter component of the test panel, which carries the luxCDABFE genes downstream of a strong SOS-dependent promoter is constructed. The plasmid pGFPuv that carries the optimized "cycle 3" variant of GFP in frame with the lacZ initiation codon is the cytotoxicity sensing reporter component of the test panel. According to the modified "Hanahan" procedure, pPLS-1 and pGFPuv are used to transform the strain *S. typhimurium* TA1535 (ATCC: *S. choleraesuis* subsp. *choleraesuis* strain TA1535) [23]. For selecting plasmid carrying cells, the bacteria are cultured at 37°C in NB-medium supplemented with 50-μg/mL ampicillin. The DNA damaging agents containing sample are dissolved in distilled water at high concentration and diluted for the test. Each well of White LB96P-CMP Mikro Lumat Plates with a transparent bottom contains 10 μL of the solvent or different concentration of the test sample. The agar plate containing a single colony of bacteria, 10 mL of NB-medium with 50-μg/mL ampicillin, is shaken on a rotary shaker at 37°C for 16 h. Then the dilution (1 : 50) of bacteria in fresh warm NB-medium containing 50-μg/mL ampicillin are incubated at 37°C until the absorbance at 600 nm ($A_{600}$) reaches 0.2–0.3. Aliquots of 90 μL of this culture are added to each well, the microplate is covered with a gas-permeable self-adhesive seal and placed into the temperature-controlled microplate reader. The absorbance, luminescence, and fluorescence of the culture is successively and repeatedly determined by the programmed measurement cycle at 30°C for up to 8 h of continuous incubation with a duration of about 10 min per measurement cycle (total of 50 cycles). The measurement cycle of each well includes 120-s shaking, 0.2-s luminescence measurement without filter, 0.1-s absorbance measurement (490 nm, 20 nm bandwidth), and 0.1-s fluorescence measurement (excitation at 405 nm, emission at 510 nm).

### 2.2 ALIC-Based Assay for Detecting Toxicity of Chemicals in SPM

A549 cells and Hep G2 cells cultured in DMEM supplemented with 10% fetal bovine serum, 20-mM hydroxyethylpiperazine-N′-2-ethanesulfonic acid, penicillin (100 U/mL), streptomycin (100 μg/mL), amphotericin B (0.25 μg/mL), and 1.0% non-essential amino acid solution (only for Hep G2 cells) are subcultivated using 0.25% trypsin and 0.02% EDTA in PBS. After collection by trypsinization A549 cells are seeded at $1.0 \times 10^5$ cells/cm$^2$ onto the membrane culture insert precoated with a 0.03% type collagen solution. Initially, to both the apical (Ap) and the BL sides, the medium is added. Using a phase contrast microscope, the formation of the cell sheet is assessed. The evaluation of the development of tight junctions in the cell layers with a Millicell-ERS is performed based on the measurement of trans-epithelium electrical resistance (TEER). Until the TEER shows constant and saturated values (45–50 Ωcm$^2$ without the blank value of the membrane itself), the culture medium of A549 cells in the Ap side is removed and the ALIC is started. The chemical-free SPM samples are sterilized by ethylene oxide gas (EOG) performed with a commercially available sterilization system. After being measured carefully with a digital weighting machine, the EOG-sterilized SPM samples are loaded directly to the Ap side of an A549 cell layer in ALIC. In 12 well plates, A549 and Hep G2 cells are seeded at $1.0 \times 10^5$ cells/cm$^2$, cultured, and used after

they reach confluence. In the preparation of the extracts of the SPM samples, the suspension of 24 mg of each sample and 200 μL of DMSO is sonicated for 15 min and centrifuged for 10 min at 15,000 rpm, and the supernatant is diluted with culture medium to adjust the final concentration of DMSO in the culture medium 0.5% and the highest concentration of SPM in the culture medium 600-μg-SPM/mL-culture medium, which corresponds to the cell-surface-area-based load of the 158-μg-SPM/$cm^2$-cell surface.

The suspension of an SPM sample in 100 mL of $CH_2Cl_2$ is sonicated for 15 min and filtered with a membrane that has a 0.1-μm pore size, the filtrate is evaporated, and the residue is resolved in acetonitrile [24]. The concentration of the chemicals in the SPM extracts is determined using a multicolumn high-performance liquid chromatography (HPLC)/spectrofluorometer/computer system. According to $EQ = S([PAH]_1EF_1 + \ldots + [PAH]_nEF_n)$, the chemically derived induction equivalent (EQ) is calculated to standardize the biological effect of the chemical on the EROD capacity of the Hep G2 cells [25].

After 48 h of exposure of SPM, each culture is rinsed twice with PBS, and to the Ap and BL sides, 500 μL and 1.5 mL of the substrate solution of acid phosphatase is added, respectively. For the monolayer culture, 1 mL of AP solution is added, the 12 well plates are incubated at 37°C for 2 h, the absorbance is measured at 405 nm with the value of cell-free well as control, and the cell survivability is calculated by referencing a predetermined calibration curve.

A549 and Hep G2 cells are cultured in 12 well plates, washed twice with PBS, and loaded with 10-μM 7-ethoxyresorufin in the presence of 10-μM dicumarol in culture medium. After 1 h of incubation, the fluorescence intensity is measured (530-nm excitation wavelength, 585-nm emission wavelength) to detect the formed resorufin. The intensity is calibrated to the resorufin concentration using a standard curve.

### 2.3 PWM-Induced IgM Assay for Detecting the Toxicity of Chemicals

The toxicants containing sample are dissolved in Ca, Mg-free PBS or DMSO, or the vehicle solvent (10 μL each). Spleen cells are aseptically removed from BALB/cAnN mice (8–9 weeks old) [26], washed in Hanks' balanced salt solution supplemented with 10-mM HEPES buffer (pH 7.4), and diluted by RPMI 1640 medium containing 10% heat-inactivated fetal bovine serum, 2-mM glutamine, 0.4-M sodium pyruvate, 20-mM mercaptoethanol, non-essential amino acids, and antibiotics to prepare suspensions of $10^6$ cells/mL. Cells are cultivated in a 24-well culture plate under 5% $CO_2$-air and the presence of PWM at 37°C for 2 to 6 days, supplemented with chemical toxicants containing a sample at the beginning of the culture. A total of 700 μL of culture media and residual cells in 300 μL of media are used for total IgM assay and cell proliferation assay, respectively.

The 96-well microtiter immunoplates are precoated overnight at 4°C with 100 μL of rabbit anti-mouse IgM antibody (IgG) in PBS (pH 7.4), washed with PBS containing 0.05% Tween20, and blocked overnight at 4°C with 150 μL of PBS containing 1% BSA. Each well is washed with PBS; into each well, 100-μL aliquots of culture media mentioned above (usually 11-fold diluted media) or reference mouse serum or standard mouse IgM diluted with 1% BSA-PBS are added, left at room temperature for 60 min, washed, and incubated; and 100-μL aliquots of horseraddish peroxidase-conjugated goat antimouse IgG in PBS are added. Sixty minutes

later, to each well, 150-μL aliquots of substrate solutions containing 0.4% o-phenylenediamine and 0.01% hydrogen peroxide in 0.1-M citrate–0.2-M phosphate buffer are added and incubated. Fifteen minutes later, 50 μl of $H_2SO_4$ (0.5 M) is added to terminate the enzyme reaction. Using a microplate reader the optical density at 490 nm of each well is read.

After the addition of 30 μL of TetraColar One reagent, 300 μL of media mentioned above are incubated under 5% $CO_2$-air at 37°C for 2 h. The optical density at 450 nm of the media containing the water-soluble tetrazolium salt produced by the enzyme reaction is read, and the cell number in cultures is measured by the modification (WST-8 method) of the MTT assay for mitochondrial dehydrogenase activity in viable cells [27].

### 2.4 GFP-Based Cell Assay for Detecting the Toxicity of TCDD and Related Chemicals

By excising the 1846 base-pair (bp) Hind III fragment from the plasmid pGudLuc1.1, pGreen1.1 is created [28]. pGreen1.1 contains the 480-bp dioxin-responsive domain from the mouse CYP1A1 gene inserted upstream of the mouse mammary tumor virus (MMTV) promoter, confers dioxin responsiveness upon the MMTV promoter and adjacent reporter gene, and is inserted into the Hind III site immediately upstream of the enhanced GFP reporter gene in the plasmid pEGFP-1.

Plates of the mouse hepatoma (Hepa1c1c7, approximately 80% confluent) cells maintained in αMEM containing 10% fetal bovine serum are transfected with the construct pGreen1.1 (20 μg) using polybrene, grown for 24 h, split 1 to 10, and replated into selective media. After growth for 4 weeks, resistant clones are isolated and screened for the EGFP assay. H1G1.1c3 cells and DMSO (1% maximum final concentration) or TCDD (1 nM in DMSO) are cultured in 6-well culture plates at 37°C for 24 h, harvested by scraping into lysis buffer (50-mM $NaH_2PO_4$, 10-mM Tris-HCl pH 8, 200-mM NaCl), and lysed by repeated passage through a 27-gauge needle. The medium is centrifuged, and the fluorescence of an aliquot of the supernatant is determined (460-nm excitation wavelength, 510-nm emission wavelength). In the microtiter plate analysis of EGFP, the intact cells are plated into 96-well tissue culture dishes at $7.5 \times 10^4$ cells/well and allowed to attach for 24 h. The selective media are then changed to 100 μL of nonselective media containing the test chemical or DMSO (1% final solvent concentration). EGFP levels of the intact cells in the nonselective media are measured (at the indicated time points, 485-nm excitation wavelength, 515-nm emission wavelength). To normalize results between experiments, the instrument fluorescence gain setting is adjusted so that the level of EGFP induction by 1-nM TCDD produces a relative fluorescence of 9000 relative fluorescence units (RFUs). To photograph the cells, after growth on 25-mm round cover slips for 24 h and treated with DMSO or a 1-nM TCDD containing sample for 48 h, H1G1.1c3 cells are replaced into PBS. Cell fluorescence is visualized at a 490-nm excitation filter and 535-nm emission filter.

Recombinant mouse hepatoma (H1L1.1c2) cells grown in 24-well microplates are incubated with DMSO (10 μL/mL) and a TCDD containing sample in DMSO or its related chemical containing sample in DMSO at 37°C for 4 h; luciferase activity of cells in each well is determined and normalized to sample protein concentration using fluorescamine with bovine serum albumin as the standard [29, 30].

The complementary pair of 5′-GAT-CTG-GCTCTTCTCACGCAACTCCG-3′ and 5′-GATCCGGAGTTGCGTGAGAAGAGCCA-3′ (corr-esponding to the AhR binding site of DRE3 and designated as the DRE oligonucleotide) is radiolabeled with [$\gamma$ $^{12}$P]ATP (6000 Ci/mmol). Gel retardation analysis of cytosolic AhR complexes transformed *in vitro* with a TCDD containing sample (20 nM) or related chemical containing sample is performed and protein–DNA complexes are visualized. The amount of [$^{32}$P]-labeled DRE in the induced protein–DNA complex is determined.

## 2.5 RACB Protocol for Detecting Ovarian Toxicity of Xenobiotics

RACB protocol consists of Task 1 (initial dose-setting study), Task 2 (continuous breeding phase), Task 3 (crossover mating trial), and Task 4 ($F_1$ offspring dose). In Task 2, control (n = 40 animals/sex) and up to three treatment groups (n = 20 animals/sex) are implicated. In Task 2, $F_0$ (parental) rodents are exposed to a chemical containing sample during a 7-day premating period, randomly assigned to a mating pair, and treated with the same chemical containing sample and dose throughout a 98-day period of continual cohabitation (during which multiple litters are born). To encourage immediate remating, in early pregnancies, neonates are removed from the dam within 12 h. The principal toxicity in Task 2 is aberrant reproductive performance in $F_0$ rodents as indicated by alterations in the number of litters per breeding pair or neonatal body weight and sex ratio. If a positive toxicity is observed in Task 2, Task 3 should be performed to determine whether males or females are more sensitive. For this phase, both male and female high-dose $F_0$ mice are paired to control $F_0$ mice of the opposite sex, and reproductive performance is compared to determine the affected sex (es). If a negative toxicity is observed in Task 2, Task 3 is omitted. After 98 days, breeding pairs are separated and continuously treated until the $F_1$ generation is delivered and weaned. In Task 4 the $F_1$ offspring of Task 2 parents are dosed until 74 ± 10 days of age, at which time male and female animals from the same treatment group but different litters are mated (n = 20 /sex/group) to generate $F_2$ litters. The $F_1$ offspring have been exposed to the chemical as gametes and as young adults during prenatal and postnatal development.

Ovaries from Task 1 (approximately 50 days old), $F_0$ parents from negative Task 2 (approximately 215 days old) or from Task 3 (approximately 240 days old), or $F_1$ offspring from Task 4 (approximately 120 days old) are removed at necropsy, trimmed of fat, fixed by immersion in Bouin's solution for 12 to 24 h, and transferred to 70% ethanol for storage and transport. An intact ovary from each animal is dehydrated in graded alcohols and xylene, embedded in a longitudinal orientation in separate paraffin blocks, and sectioned serially at 6 μm (approximately 400 sections per ovary). To each slide (approximately 40 slides per ovary), two rows of five sections retained in sequence are applied and stained with hematoxylin and eosin. Counts of ovaries from 10 mice per group are gathered. Ovaries are available from only 5 to 6 treated animals in $F_1$ offspring from two Task 4 EGME assays and from 9 treated animals from Task 2 oxalic acid study. To produce an equal number of control tissues for analysis from untreated animals of the same studies, the ovaries are also chosen randomly. Beginning with the first section of the first slide, sections from each ovary are examined. At least from the third or fourth section

(a distance of 18 to 24 μm into the ovary), follicles are encountered and then categorized and enumerated. Differential counts from every tenth section or approximately 40 sections per ovary are made.

The ovarian follicles are categorized to small, growing, and antral ones by (1) the relative cross-sectional diameter of the follicle as measured from the outer margins of the granulosa cell layers, (2) the number of granulosa cell layers, and (3) the nature of the antral space. Small follicles, approximately 20 μm in mean diameters, consist of an isolated oocyte or an oocyte surrounded by a partial or unbroken, single layer of granulosa cells. Growing follicles, 20 to 70 μm, have an oocyte surrounded by a multilayered, solid mantle of granulosa cells. Antral follicles, more than 70 μm, are characterized by a central oocyte and fluid-filled space bordered by hundreds of layered granulosa cells [31]. Each counting session is limited to no more than 3 h to limit fatigue.

### 2.6 Uroepithelial Cell Assay for Detecting Urotoxicity of Chemicals

Male beagle dogs (31 months old) are individually housed at 23 ± 2°C and a relative humidity of 60 ± 20% with a 12-h light/dark cycle. The dogs from a control group are anesthetized by sodium pentobarbital (25 mg/kg, i.v.) and killed by exsanguination. The bladder is aseptically excised, then lengthwise incised, and washed three times with the Krebs solution containing 110-mM NaCl, 5.8-mM KCl, 25-mM $NaHCO_3$, 1.2-mM $KH_2PO_4$, 2.0-mM $CaCl_2$, 1.2-mM $MgSO_4$, and 11.1-mM glucose (pH 7.4) at 4°C. After removal of excess fatty tissues, the bladder is transferred, mucosal side down, to a metal rack with 10 sharp metal pins along each edge [32], placed in 4°C Krebs solution, and the smooth muscle layer is carefully removed. The tissue is stretched, mucosal side up, across the metal pins on a $10 \times 10/cm^2$ plate, incubated in the minimum essential medium (MEM) containing 1% (v/v) penicillin/streptomycin/fungizon (PSF), 2.5-mg/mL dispase, and 20-mM MEM/PSF/dispase solution (pH 7.4) at 4°C for 24 h, the MEM/PSF/dispase solution is aspirated, the stripped mucosa is transferred to a sterile 150-mm culture dish, and uroepithelial cells are scraped from the connective tissues with cell scrapers, which are suspended in 20 mL of trypsin-EDTA (0.25% trypsin and 1-mM EDTA·4Na) and incubated at 37°C for 30 min. Later, the single cell suspension is diluted with MEM containing 1% PSF, 5% FBS, and 20-mM MEM/PSF/FBS solution (pH 7.4) to 50 mL, spun down at 4°C, and centrifuged (1000 g, 5 min). The supernatant is aspirated to prepare the suspension of the cells in 50 mL of MEM/PSF/FBS solution. The cells are then rewashed in 50 mL of the keratinocyte-SFM medium and resuspended at $6.0–7.0 \times 10^5$ cells/mL.

By mixing 5 mg of type IV collagen, 100 μL of glacial acetic acid, and 50 mL of distilled water, the collagen solution is prepared, kept overnight at 4°C, sterilized with a 0.22-μm bottle top filter, and stored at 4°C. The keratinocyte medium is added to both chambers and incubated at 37°C for 2 h. Before use, the collagen solution is diluted 1 : 9 with 10-mM $Na_2CO_3$-HCl (pH 9.0), of which 500 μL is added to each apical chamber after aspirating the keratinocyte medium and incubated at 37°C for 1 h. Before plating, the collagen solution is aspirated, 0.5 mL of the cell suspension is added to the apical chamber, and 1.5 mL of keratinocyte medium is added to the basal chamber. After incubation at 37°C for 3 days, when the transepithelial electric resistance (TER) reaches levels of approximately $1000\,\Omega cm^2$ or

higher, the apical and basal media are aspirated and replaced with 0.5 and 1.5 mL of the keratinocyte medium containing 1-mM $CaCl_2$ (KM/Ca solution), respectively, and TER is measured. Before use, by immersing in 70% ethanol, the electrodes are sterilized and washed with sterile PBS, and $\Omega cm^2$ are calculated.

Cultured cells on a 12-mm transwell filter are fixed in 2% glutaraldehyde of 0.1-M phosphate buffer (pH 7.4), rinsed with 8.2% sucrose, postfixed in 1% $OsO_4$ of the buffer, dehydrated with alcohol, and embedded in epoxy resin. For light microscopic examination, the semi-thin sections are stained with 1% toluidine blue. To observe cells with a transmission electron microscope, ultrathin sections are stained with uranyl acetate and lead citrate. For immunofluorescence staining, the cultured cells on a 12-mm transwell filter are successively fixed with HISTOCHOICE at room temperature for 1 min and acidic methanol (95% methanol and 5% glacial acetic acid) at –20°C for 15 min, then incubated successively with anti-ZO-1 rabbit polyclonal antibody diluted 1:50 with PBS for 1 h and FITC-conjugated donkey antirabbit IgG diluted 1:40 with PBS for 30 min, or incubated successively with anti-E-cadherin mouse monoclonal antibody 1:100 with PBS and FITC-conjugated goat antimouse IgG diluted 1:10 with PBS. To observe cells with a confocal laser scanning microscope, the cells are washed three times for 5 min with PBS, the cell-grown transwell filters are cut, transferred to a slide glass, and mounted with the mounting medium.

After plating on a 12-mm transwell filter, TERs are monitored for 20 days, during which the medium is replaced every 3 days. When TER reaches 10,000 $\Omega cm^2$ or more, the cultured cells are observed and their ZO-1 and E-cadherin are checked. Validating this culture system in the examination of the effects of cytochalasin-B on TER and ZO-1, cytochalasin-B is first dissolved in DMSO, diluted with the KM/Ca solution to make concentration of 1.6 μM (final DMSO concentration, 0.08%), 4 μM (final DMSO concentration, 0.2%), and 10 μM (final DMSO concentration, 0.5%), and sterilized with a 0.22-μm membrane filter. TER is measured 0, 1, 2, 4, 8, 24, and 48 h after exposure, and immunofluorescence for ZO-1 is observed 48 h after exposure, with DMSO diluted 1:200 with KM/Ca solution serving as a negative control solution.

### 2.7 Mating Efficiency Assay for Detecting Toxicity of Endocrine Disruptors

In the pheromone signaling pathway experiment, yeast *Saccharomyces* cerevisiae strains W303A (MATa; ade2, his3, trp1, ura3) with pSL307 plasmid fused to the FUS1-lacZ are cultured in SD medium (0.67% yeast nitrogen base and 2% glucose/L) supplemented with adenine, l-histidine, and l-tryptophan at 30°C overnight. To fresh YPD medium (10 g of Bacto-yeast extract, 20 g of Bacto-peptone/L, and 2% glucose) or SD medium one tenth of the overnight culture is transferred, incubated at 30°C for 3 h, and the cultures that are adjusted at the logarithmic phase (OD = 0.8–1.0) of growth are harvested and subsequently used at 30°C [33].

W303A and W303A with fused plasmid pSL307 are incubated in YPD or SD medium for 60 min with or without test compound, harvested, washed three times with distilled water, and incubated in fresh YPD or SD medium with 5-mM α factor for 120 min to observe shmoo formation. Using a β-galactosidase activity assay, the pheromone response pathway in the W303A strain with pSL307 fused to the

FUS1-lacZ reporter gene is quantified [34]. Yeast cells are centrifuged at 3000 g, pellets are washed twice in distilled water, resuspended in 1.0 mL of Z buffer (21.5-g $Na_2HPO_4 \cdot 12H_2O$, 6.2-g $NaH_2PO_4 \cdot 2H_2O$, 0.75-g KCl, 0.246-g $MgSO_4 \cdot 7H_2O$, 2.7-mL β-mercaptoethanol/l, pH 7.0) with glass beads, with which three drops of chloroform and two drops of 0.1% SDS are mixed by ortex mixing, and incubated at 28°C for 5 min. After the addition of 0.2 mL of 4-mg/mL *p*-nitrophenyl-β-D-galactopyranoside at 28°C, the reaction is started, and by the addition of 0.5 mL of 1-M $Na_2CO_3$, the reaction is stopped. On removal of the cells by centrifugation, the superant is measured at 420 nm. In the mating efficiency assay, the diluted W303A strain is treated with the solution of test chemicals in YPD medium for 60 min and spread on both SD and YPD media. The colony-forming units (CFUs) of W303A strain on YPD medium are incubated for 2–3 days to give the cardinal number of mating efficiency. Simultaneously, the 144-3A strain (MATa; ura3, leu2, his4) is incubated on YPD medium for 60 min and streaked with diluted W303A strain on SD medium. Conjugated diploid 144-3A strain/W303A strain on SD medium complements the auxotrophy of each haploid cell.

### 2.8 Pnar-gfp Assay for Detecting Carcinogenic Toxicity of Nitrate

*Escherichia coli* DH5α (Φ80dlacZΔM15 recA endA gyrA thi hsdR supE relA deoRΔ [lacZYA-argF]), pUC18 (for routine cloning procedures), pCRII (for TA cloning of PCR products), pGreen-TIR containing the gfp gene, and pMV4 containing the narG promoter are used [35, 36]. Minimal medium (pH 7.0) contains 3.9-mM $KH_2PO_4$, 6.1-mM $K_2HPO_4$, 1.5-mM $(NH_4)_2SO_4$, 0.2-mM $MgSO_4 \cdot 7H_2O$, 22.4-μM $MnSO_4 \cdot 4H_2O$, 0.9-μM $FeSO_4 \cdot 4H_2O$, 4.4-μM $CaCl_2 \cdot 2H_2O$, 400-mg/L casamino acids, 14.8-μM thiamine hydrochloride, and 22.2-mM glucose. LB medium (pH 7.5) contains 10-g/L bacto-tryptone, 5-g/L yeast extract, and 10-g/L NaCl. *E. coli* DH5α (pPNARGFP) cultures are supplemented with 100-μg/mL ampicillin. Primers J1 (5′-CATCGAATTCTCCTGTGGGAGCCT-3′) and J2 (5′-CTGGCATGCATTCACTTGCCGCCTT-3′) are designated for the amplification of the nar promoter [37, 38]. J1 contains an artificial GAATTC (EcoRI site) and anneals to nucleotides +3 to −23 in the nar region, where +1 is the first nucleotide of the start codon. J2 contains an artificial GCATGC (SphI site) and anneals to nucleotides −438 to −414. After purification using Qiagen Plasmid Midi Kits (Qiagen, USA) or Wizard Plus SV miniprep kits (Promega, USA), the DNA sequence is determined by MWG Biotech (Milton Keynes, UK). The whole bacterial cells are measured with an excitation wavelength of 480 nm (10-nm bandwidth) and an emission wavelength of 510 nm (5-nm bandwidth) on a fluorescence spectrometer. The nitrate-induced fluorescence image of *E. coli* DH5α (pPNARGFP) is obtained with a BP465-495 excitation filter and a Ba 520 emission filter on a fluorescence microscope equipped with epifluorescence optics. The measurements of optical density are performed on a spectrophotometer at 600 nm.

## 3 ANTIVIRAL AND ANTICANCER ASSAY

The elucidation of the mechanisms of special enzymes provides the bases for highly specific and sensitive bioassays focused on the target enzymes and cells. The

alkaline phosphatase (AP) enzyme in the human endometrial adenocarcinoma cell line is sensitive to estrogen stimulation. Selective inhibition of the lyase activity of DNA polymerase sensitizes cancer cells to DNA-damaging agents. The human immunodeficiency virus type-1 reverse transcriptase (HIV-1 RT) and protease play key roles in HIV replication. The overexpression of the P-glycoprotein can produce cancer cell multidrug resistance (MDR). The receptor phosphorylation may induce ligand-induced receptor tyrosine kinase activation. Based on these facts, the human breast cancer cell assay [39], the Ishikawa cell and rat assay [40], the RT–PCR and swine assay [41], the antitubercular and cytotoxic assay [42], the DNA polymerase β lyase assay [43], the HIV-1 protease and reverse transcriptase kinetic assay [44], the P-glycoprotein pump assay [45], and the kinase receptor activation (KIRA) assay [46] are established and can be used in bioassays.

### 3.1 Human Breast Cancer Cell Assay for Detecting Anticancer Activity

Toxicity against brine shrimp *Artemia salina* nauplii is carried out in 96-well microplates using emetine hydrochloride as a positive control [47]. The selection of human breast cancer cell lines (MCF-7, MCF-7/ADR, MT-1, and MDA-MB-435) provides cells with a variety of receptor types; for instance, MCF-7 cells express high levels of estrogen receptors, MCF-7/ADR cells are resistant to doxorubicin, MT-1 cells have low estrogen receptor levels [48], the MDA-MB435 cell line has no estrogen receptors. In the determination of inhibition of cell growth in 96-well microtitre plates, the MTT assay is used [49]. The samples of test compounds in DMSO are diluted in medium so that the concentration of DMSO is more than 1% and does not affect cell growth. Overall, 180-μL aliquots of a cell suspension (1–2 $\times 10^4$ cells/mL) are plated into microplate wells and incubated overnight at 37°C in air containing 5% $CO_2$, and 150 μL of the culture medium is replaced with 150 μL of fresh medium, to which 20 μL of the solution of the test sample in medium is added to give final concentrations in the plate of 10, 1, 0.1, 0.01, and 0.001 μg/mL. Control wells with medium only and a positive control with 5-fluorouracil are used in each test. After 96 h of the incubation, 150 μL of medium is replaced with 150 μL of fresh medium, to which 20 μL of the 5-mg/mL MTT solution is added. After incubation for 4 h, 180 μL of medium is replaced with 180 μL of DMSO and carefully mixed and then determined at 540 nm using an enzyme-linked immunosorbent assay (ELISA) multiwell spectrophotometer.

### 3.2 Ishikawa Cell and Rat Assay for Detecting Antiestrogens

The cells are grown in 96-well plates in estrogen-free medium (phenol red free, with charcoal-stripped calf serum) and contain test compounds and antiestrogens at concentrations that are varied over several log orders. For the antiestrogen assay, to the cells, a range of concentrations of samples are added concurrently with 1-nM antiestrogens. After grown for 3 days, to determine AP activity, the cells are frozen, defrosted, and incubated with p-nitrophenylphosphate at room temperature, and the hydrolysis product p-nitrophenol is measured kinetically at 405 nm [50].

The specificity of the antiestrogenic activity is determined by using ERα and ERβ in ER element (ERE)-transfected human choriocarcinoma JAR cells, which are transfected with plasmids containing a consensus ERE fused to a firefly

luciferase reporter gene and separately with the expression vectors for either human ERα or human ERβ. JAR cells are routinely cultured in RPMI 1640 supplemented with 10% fetal bovine serum, 0.5% non-essential amino acids, and 1% PEST (100-U penicillin/mL and 100-μg streptomycin/mL). After seeded in six-well plates for 24 h, cells are transfected using the Mirus Trans IT reagent in a serum- and antibiotic-free mixture of phenol-red free OptiMEM with 0.1–0.4-μg pC × N2 human ERα or pC × N2 h-ERβ, and a 0.75-μg 3 × ERE-TATA-Luc reporter constructed by introducing an HpaI/BglII fragment containing 3 × ERE-TATA into SmaI/BglII of the pGL3-Luc basic vector [51–53]. Medium is replaced with a phenol red-free RPMI containing 10% dextran-coated charcoal-treated calf serum and 0.5% non-essential amino acids. Twenty-four hours later, antiestrogens are added. Cells are incubated at 37°C in 5% $CO_2$ for 12 h, harvested in 10-mM Tris-HCl/10-mM EDTA/150-mM NaCl, and centrifuged at 4000 g for 4 min. The supernatant is removed, the cell pellets are lysed in Lysis Buffer 2, and luciferase activity is measured.

After the AP assay and washing three times with PBS, the Ishikawa cells are lysed using 1% Nonidet P-40 and 0.1% sodium dodecyl sulfate in the presence of protease inhibitors. Before the antibody incubation, proteins (25 μg/well) separated by SDS–PAGE on ice using 10% polyacrylamide gel are transferred to nitrocellulose membranes stained with Ponceau Red to ensure proper transfer. After blocking the membranes with 5% powdered milk in water, immunoblotting is performed. The blots are incubated with the ERα monoclonal antibody clone 6F11 overnight at 4°C. Using peroxidase-labeled horse antimouse secondary antibody and Chemiluminescence Reagent Plus, ERα is detected by Western blotting. Using a digital imaging analysis system, the intensity of the signal is analyzed. To normalize the amount of protein loaded in the gels, β-actin is used as an internal control.

The uterotrophic assay is performed with stimulation in immature rats [54]. Female SD rats (22 days old) are injected s.c. daily for 3 days with antiestrogens, control animals receive the vehicle (0.1-mL sesame oil), then animals are killed, and uteri are removed, dissected, blotted, and weighed. To determine whether antiestrogens have tissue-selective effects in cholesterol, uteri, and bone, ovariectomized female SD rats (250 g) are injected with antiestrogens s.c. for 35 days, and then killed by exsanguination under ether anesthesia. The total cholesterol concentration of the serum is determined by a commercial chromogenic assay. The uteri are dissected, weighed, fixed in formalin, and imbedded in paraffin to prepare for 5-μm sections. Using the Openlab image analysis system, endometrial luminal epithelium and glandular cell height are measured. The tibia free of extraneous tissue are histomorphometrically analyzed. The bones fixed in 70% ethanol are dehydrated in graded ethanol and cleared in toluene. Then the specimens are infiltrated with increasing concentrations of methymethacrylate (MMA) and embedded in MMA [55]. After polymerization, MMA blocks are cut to size, sanded, and polished to the appropriate level to prepare 4–5-μm sections. The sections are mounted on gelatin-coated slides, stained with toluidine blue, and measured [56].

### 3.3 RT–PCR and Swine Assay for Detecting Anti-HEV Antibody

To choose seronegative pigs for inoculation, 75 SPF pigs (2 weeks old) are tested by ELISA for swine Hepatitis E virus (HEV) IgG antibodies. Before inoculation,

the pigs are allowed to acclimate to the research facilities for 1 week. Tissues (liver, heart, pancreas, or skeletal muscle) and feces of HEV-infected SPF pigs are collected at 3–7, 14–20, and 27–55 days postinoculation (DPI) and are pooled and stored at –70°C for inoculation [57]. Each inoculum is prepared as a tissue homogenate or fecal suspension (10%, w/v) in PBS and tested by a semiquantitative RT–PCR for swine HEV RNA. The positive control inoculum is a standard infectious pool of swine feces with a $10^{4.5}$ 50% pig infectious doses ($PID_{50}$) of swine HEV/mL.

Total RNA is extracted from 100 μL of each sample by TriZol reagent and tested by a nested PCR with primers located in the putative capsid gene (ORF2) [58, 59]. In the first round PCR, the fragment of 404 base pairs (bps) with the forward primer F1 (5′-AGCTCCTGTACCTGATGTTGACTC-3′) and the reverse primer R1 (5′-CTACAGAGCGCCAGCCTTGATTGC-3′) is formed. In the second round PCR, the fragment of 266 bps with the forward primer F2 (5′-GCTCACGTCATCT GTCGCTGCTGG-3′) and the reverse primer R2 (5′-GGGCTGAAC-CAAAA TCCTGACATC-3′) is formed. With R1 reverse primer and SuperScript II reverse transcriptase (GIBCO/BRL), total RNA is reverse transcribed at 42°C for 1 h and the resulting cDNA is amplified by PCR with ampliTaq Gold DNA polymerase. The PCR reaction is carried out for 39 cycles of denaturation at 94°C for 1 min, annealing at 52°C for 1 min, extension at 72°C for 1.5 min, and incubation at 72°C for 7 min. Overall, 10 μL of each round of PCR are mixed as the template. By gel electrophoresis, the amplified PCR products are examined. The virus titer of inocula is calculated and expressed as a genome equivalent (GE)/mL. After 7 to 21 days, serum samples of inoculated pigs are collected and tested by RT–PCR for swine HEV RNA.

Anti-HEV IgG antibodies in swine sera are detected using a standardized ELISA. A purified 55-kDa truncated recombinant putative capsid protein of human HEV strain Sar-55 is used as the antigen that cross-reacts well with the swine HEV. Peroxidase-labeled goat antiswine IgG is used as the secondary antibody. Duplicates per serum sample are used.

### 3.4 Antitubercular and Cytotoxic Assay

Using the microtiter plate alamarblue technique, the antitubercular activity is assessed against a nonvirulent strain of *Mycobacterium tuberculosis* ($H_{37}Ra$). As positive controls, isoniazid and kanamycin sulfate exhibit MICs in the ranges of 0.29–0.66 and 3.5–8.5 μM, respectively. According to the sulforhodamine B procedure, cytotoxicity is determined. MCF-7 (human breast adenocarcinoma), HeLa (human cervical carcinoma), KB (human oral epidermoid carcinoma), and HT-29 (colorectal carcinoma) are used as the target cell lines. After exposure to test samples, cell viability is determined colorimetrically at 515 nm. Dose-response evaluations yield a concentration mediating a 50% cytotoxic response ($IC_{50}$).

### 3.5 DNA Polymerase β Lyase Assay

Using terminal deoxynucleotidyltransferase + [$\alpha$-$^{32}$P]ddATP, a 36-nucleotide oligodeoxyribonucleotide containing uridine at position 21 is labeled at its 3′-end and the product is subjected to 20% denaturing polyacrylamide gel electrophoresis for

purification. Using autoradiography, the interesting band is visualized and excised. By heating to 70°C for 3 min and then slow cooling to 25°C, the DNA substrate is annealed to its complementary strand. To 200 μL of 354-nM [α-$^{32}$P]-labeled double-stranded oligodeoxynucleotide with uridine at position 21, 10-mM Hepes-KOH (pH 7.4), 5-mM $MgCl_2$, 50-mM KCl, 10-mg/mL bovine serum albumin, 2.4 U of uracil-DNA glycosylase, and 3 U of AP endonuclease are added, incubated at 37°C for 20 min, and an apurinic (AP) site is created in the [α-$^{32}$P]-labeled double-stranded oligodeoxynucleotide. To 5 μL of the above reaction mixture, the test samples and 0.172 U of rat DNA polymerase β are added. After incubation at room temperature for 30 min, the reaction is terminated by adding 0.5-M $NaBH_4$ (50-mM final concentration) and incubated at room temperature for 10 min. The reaction products are incubated at 70–80°C for another 20 min, separated on a 20% denaturing polyacrylamide gel, and visualized by autoradiography.

A total of 200 μL of culture samples containing approximately $1.0 \times 10^4$ A549 cells (maintained in Kaighn's modification of Ham's F12 medium (F12K) with 2-mM L-glutamine supplemented with 1.5-g/L sodium bicarbonate and 10% fetal bovine serum at 37 °C in air containing 5% $CO_2$) are placed in each well of 96-well culture plates, treated with test samples, and incubated at 37°C for 48 h in air containing 5% $CO_2$. In the determination of cytotoxicity, the culture medium is replaced with 15 μL of 5-mg/mL MTT per well, the samples are incubated at 37°C for 4 h in air containing 5% $CO_2$, 200 μL of DMSO is added, and the $OD_{570}$ value is obtained using a microplate reader.

### 3.6 HIV-1 Protease and Reverse Transcriptase Kinetic Assay

To the wells of a streptavidinecoated microtiter plate containing 20 μL of test sample solution and 4 ng of the HIV-1 RT in 20 μL of lysis buffer, 20 μL of the reaction mixture containing a homogenous template/primer hybrid ($(rA)_n(dT)_{15}$, 750-$mA_{260nm}$/mL final concentration) and a triphosphate substrate (dUTP/dTTP, 10-μM final concentration) are added. After the reaction is carried out at 37°C for 1 h, 200 μL of anti-digoxigenin-peroxide solution and ABTS [2,2–azino-bis-(3-ethyl benzothiazoline-6-sulfonic acid) diammonium salt] substrate are added for the coloring reaction. The absorbance of each well is measured at 405 nm with the reference wavelength at 490 nm and nevirapine as the reference compound. Using the same procedure mentioned above, except for the concentration of the enzyme solution (1 ng of the HIV-1 RT), the incubation time (30, 52, 80, 105, 130 min), and various concentrations of either the template/primer (1500, 750, 187.5 $mA_{260nm}$/mL) or the triphosphate substrate (20, 15, 10, 5, 2.5 mM) in the presence of the inhibitor, the enzyme kinetic assay is also performed. With a slight modified procedure to that mentioned above, the HIV-1 protease assay is performed. To the mixture of 1 μL of the solution of test compound in DMSO and 10.5 μL of the substr-ate solution (His-Lys-Ala-Arg-Val-Leu-(p-$NO_2$)-Phe-Glu-Ala-Nle-Ser-$NH_2$, 0.1 mg/mL in HIV-1 protease assay buffer), 0.5 μL of the recombinant protease solution (0.3 mg/mL) is added and incubated at 37°C for 15 min, the reaction is stopped by the addition of 1.2 μL of 10% trifluoroacetic acid, and the reaction mixture is diluted with 20 μL of water. The hydrolysate and the remaining substrate are quantitatively analyzed by HPLC [column, Inertsil ODS-3 (4.6 × 150 mm), a linear gradient of $CH_3CN$ (15% to 40%) in 0.1% TFA, an injection volume of 20 μL, and a

flow rate of 1.0 mL/min, detection at 280 nm]. The hydrolysate and substrate are eluted at 8.61 and 10.84 min, respectively.

### 3.7 P-Glycoprotein Pump Assay

MCF-7R cells (human breast adenocarcinoma cell line, resistant to adriamycin) maintained at 37°C in humidified atmospheric air containing 5% $CO_2$ in nutrient mixture (F10/HAM, contains L-glutamine, with 10% fetal calf serum, 1% non-essential amino acids, 60-μg/mL tylosin, and 1% antibiotic/antimycotic solution) are seeded in the wells of 96-well tissue microtiter culture plates at $3 \times 10^4$ cells in 200 μL of medium per well, incubated at 37°C for 24 h in a humidified atmospheric air containing 5% $CO_2$, and the medium is replaced with fresh medium containing 0.3-μM rhodamine 6G together with samples. To thorough mixing of 50 μL of each of these prepared mixtures, 450 μL of fresh medium containing 0.3-μM rhodamine 6G is added. After incubation of the test compound, reserpine (positive control) or rhodamine alone (negative control) is added and the plates are incubated at 37°C for another 3 h. Then the cells are washed twice with 200 μL of ice-cold PBS and trypsinized with 100 μL of phenol-red free trypsin solution for 15 min, which is then transferred onto empty wells. Overall, 100 μL of 4% SDS in PBS is added and the plates are shaken for 2 h to solubilize the cells and release rhodamine 6G, which is determined by measuring the fluorescence of the dye at excitation and emission wavelengths of 530/25 and 590/35 nm, respectively.

On Lab-Tek chamber slides, MCF-7R cells are seeded at $5 \times 10^4$ cells per chamber and incubated at 37°C for 24 h in a humidified atmospheric air containing 5% $CO_2$, and the medium is replaced with fresh medium containing 0.3-μM rhodamine 6G with or without 50-μM reserpine, with 100-μg/mL samples in separate chambers of the slide. The cells are incubated at 37°C for another 3 h and imaged using fluorescence microscopy with a 510–560-nm band-pass excitation filter and a 590-nm long-pass emission filter set.

### 3.8 KIRA Assay

There are two different approaches using the KIRA format. The first approach, IGF-I KIRA, uses adherent MCF-7 cells derived from a human breast adenocarcinoma expressing an endogenous IGF-I receptor to measure the bioactivity of IGF-I [60]. MCF-7 cells are seeded in each well of a flat-bottomed 96-well culture plate and incubated overnight, the supernatants are decanted, and the plates are lightly blotted on paper towels. To each well, the medium containing either experimental samples or the recombinant IGF-I standards are added. The cells are exposed to ligand for 15 min, the supernatants are decanted, and the plates are blotted. After addition of lysis buffer containing Triton X-100, sodium orthovanadate, and a cocktail of protease inhibitors, the crude lysates are generated and then transferred to an ELISA plate coated overnight with a 3B7 antibody (5.0 mg/mL) and blocked with 0.5% BSA. After removal of unbound material, the degree of receptor tyrosine phosphorylation is quantified with biotinylated anti-phosphotyrosine monoclonal antibody followed by HRP-conjugated dextranstreptavidin, and visualized with the development of a tetramethyl benzidine (TMB) substrate. The absorbance is determined at 450 nm with a 650-nm reference wavelength.

The second approach, N-terminal polypeptide D (gD )·trkA KIRA using CHO cells stably transfected with a recombinant human trkA receptor with an gD flag, is developed to measure the bioactivity of NGF. A polypeptide flag is cloned onto the N-terminus or C-terminus of the full-length recombinant human receptor stably transfected into CHO cells. A 26-amino-acid polypeptide derived from HSV gD as a capture reagent in the ELISA phase of the KIRA and a mAb 3C8 antibody as the capture antibody are used.

## 4 HEPATOXICITY AND HEPATOPROTECTIVE ASSAY

Mitochondrial proliferation of the chemicals related to hepatocellular carcinomas and their hepatocarcinogenecity are important factors for health. The renal adenocarcinomagenecity, hepatic adenoma, and carcinomagenecity induced by chemicals can promote the formation of preneoplastic foci in rats or mice. The clonal growth of glutathione-S-transferase (GST-P) enzyme-altered foci and the histopathological change in the liver of rats or mice may relate to liver carcinogenesis of chemicals. In the liver injury induced by chemicals, such as $CCl_4$, Bacillus Calmette Guerin (BCG), and lipopolysaccharide (LPS), the activities of protein, glucose 6-phosphatase, amidopyrine *N*-demethylase, and aniline hydroxylase; the levels of hepatic triglycerides and lipid peroxidation; the concentrations of nitric oxide (NO), content of malondialdehyde (MDA), and superoxide dismutate (SOD); and the viability of thymocytes can be changed significantly. In the hepatocarcinogenesis induced by chemicals such as chloral hydrate, the body weight of B6C3F1 mice can be decreased significantly. Proliferation factor (PF) and hepatocyte growth factor (HGF) are involved in liver regeneration cascade and TGF-β-induced growth inhibition of CCL-64 cells; thus, PFs present in the partially hepatectomized rat serum can serve as an index of liver regeneration cascade and two cytokines (HGF and HGF) may result in an additive effect on proliferation. Based on these facts, the GST-P enzyme-altered foci assay [61], partial hepatectomy assay [62], hepatoprotective assay [63], immunological assay [64], body-weight-based liver tumor incidence assay [65], hepatocyte primary culture assay [66], and CCL-64 cell growth assay [67] are established.

### 4.1 GST-P Enzyme-Altered Foci Assay

According to the required levels (0%, 0.03%, 0.1%, and 0.3%), the test chemical is incorporated into an irradiated (6.0 kGy) powder diet MF. The stability of 0.1%, 0.03%, 0.5%, and 5.0% chemical in the prepared diets is previously confirmed for 6 weeks at room temperature. Male F344/DuCrj rats (5 week old) are given an approximately 1-week quarantine/acclimation period to monitor health conditions and body weights. The normal rats (6 week old) are randomly divided into 8 groups (18 rats each for groups 1–5, 9 rats each for groups 6–8). The rats of groups 1–5 receive an injection of the initiator N-nitrosodiethylamine (200-mg/kg body weight, i.p.) [68]. The rats of groups 6–8 receive an injection of the vehicle. Two weeks later, the rats receive an injection of the test chemical at the desirable dose for a suitable treating period. Three weeks after beginning the experiment, all rats are given two-thirds partial hepatectomy [69]. At week 8, all surviving rats are killed

under ether anesthesia and their organs in the thoracic and abdominal cavities are examined macroscopically. Their livers are immediately excised and weighed to calculate the liver-to-body-weight ratio. For the 4–5-mm thick sections, the cranial and caudal parts of the right lateral lobe and the caudal part of the caudate lobe of all surviving rats are fixed in 10% buffered formalin solution, embedded in paraffin wax, sectioned, and stained immunohistochemically for glutathione S-transferase analysis (GST-P, ABC method) [70]. All GST-P positive hepatocytic foci larger than 0.2 mm in diameter (the lowest limit for reliable evaluation) are measured using a color image processor, and the numbers and areas (foci/$cm^2$) of the liver section are calculated. By microscopic analysis, BrdU-positive labeling indices (LIs) are quantified by randomly counting the number of positive nuclei per 1000 hepatocytes or number per unit area ($mm^2$) in sections stained immunohistochemically for BrdU [71].

### 4.2 Partial Hepatectomy Assay

Male F344 rats (30 days old) acclimated for 4 weeks before the experiment are randomly divided into 3 groups. At week 0, the rats receive a single i.p. injection of the solution of N-nitrosodiethylamine (200 mg/kg) in 0.9% saline. Two weeks later, the rats receive daily gavage administration of corn oil or 0.1-mmol/kg test chemical in a corn oil vehicle through the remainder of the 8-week study. At week 3, all rats receive a partial hepatectomy. The rats are given food and water, and lighting is set on a 12-h light/dark cycle. On days 23, 26, 28, 47, and 56, at least five rats from each group are sacrificed by aortic exsanguination. Whole livers are removed, and the tissues are fixed in 10% neutral-buffered formalin, embedded in paraffin, serially sectioned at 5 μm, and mounted on microscope slides. Formalin-fixed sections are stained with hematoxylin and eosin to perform histopathological examination.

### 4.3 Hepatoprotective Assay

To induce liver injury, CF rats (150–200 g) and Swiss albino mice (20–30 g) of either sex are administered orally $CCl_4$ diluted with liquid paraffin. The animals of the vehicle control group are orally administered an equal volume of liquid paraffin. In hepatoprotective studies, four suitable doses of test compound and two standard doses of silymarin (25 and 50 mg/kg, p.o.) are fed to a respective group of rats 48 h, 24 h, and 2 h before and 6 h after $CCl_4$ (0.5 μL/kg, p.o.) intoxication. From the orbital sinus of all animals, blood is collected 18 h after $CCl_4$ intoxication for glutamic-pyruvic transaminase (GPT), glutamic oxaloacetic transaminase (GOT), and bilirubin analysis. In the posttreatment studies, the same doses of test compound and silymarin as mentioned for hepatoprotective studies are fed to rats 6 h, 24 h, and 48 h after $CCl_4$ (0.5 mL/kg, p.o.) intoxication. From all animals blood is collected 2 h after the last dose of test compound administration for GOT, GPT, and bilirubin analysis. The test compound and silymarin (50 mg/kg, p.o.) are fed to two different groups of rats at 48 h, 24 h, and 2 h before and 6 h after $CCl_4$ (100 μl/kg, p.o.) intoxication, with the remaining two groups served as $CCl_4$ and vehicle control. Eighteen hours after $CCl_4$ intoxication, all animals of the four groups are fasted overnight and killed by decapitation. Livers are immediately excised and divided into two

parts for preparing homogenate (10%, w/v), among which one part is homogenized in isotonic sucrose (0.25 M) for determining protein and glucose 6-phosphatase activity, and for preparing microsomes by calcium precipitation to determine amidopyrine N-demethylase and aniline hydroxylase activities with spectrophotometric methods, in which NADPH is used instead of the NADPH generating system, and another part is homogenized in isotonic PBS (0.01 M, pH 7.2, 0.15-M NaCl) for determining hepatic triglycerides and lipid peroxidation. The hepatoprotective activity is expressed as hepatoprotective percentage H and calculated by $H = [1 - (T - V)/(C - V)] \times 100$, wherein T is mean value of drug and $CCl_4$, C is mean value of $CCl_4$ alone, and V is the mean value of control-treated animals. In the determination of acute toxicity, different groups of mice (each ten) are fed with different doses of test compound, with one group with the same number of mice served as control. The animals are observed continuously for 1 h and then hourly for 4 h for any gross behavioral changes and further up to 72 h for any mortality.

### 4.4 Immunological Assay

A suspension of 2.5 mg of BCG (viable bacilli) in 0.2 mL of saline is injected via the tail vein into each mouse, and 10 days later a solution of 7.5 μg of LPS in 0.2 mL of saline is injected. The mice are anesthetized with ether, sacrificed by cervical dislocation 16 h after LPS injection, and trunk blood is collected into heparinized tubes (50 U/mL) and centrifuged at 4°C and 1500 rpm for 10 min. Serum is aspirated and stored at –70°C until assayed as described below. The liver is also removed and stored at –70°C until required.

For the *in vivo* experiment, the mice are equally divided into 5 groups randomly, including normal, model control, and test compound groups (3 different doses). Mice in test compound groups receive suitable doses using an 18-gauge stainless steel animal feeding needle for 10 days before LPS injection. Mice in the normal and model control groups are fed the same volume of vehicle only. For the *in vitro* experiment, the Kupffer cells isolated from normal and BCG priming rats are divided into 7 groups randomly, including control cells, cells added with LPS (5 μg/mL) alone, and cells added with LPS (5 μg/mL) and the test compound.

The liver of normal rat is initially perfused through the portal vein with D-Hank's until blood free and finally by recirculation with Hank's containing 0.5-g/L collagenase IV until the vessels are digested (up to 20 min). The liver is scraped using a cell scraper, filtered by a 100-μm filter, and stirred in Hank's containing 2.5-g/L pronase and 0.05-g/L DNase at 37°C for 20 min. After three times of centrifugation and washing at 4°C and 300 g for 10 min in Gey's balanced salt solution (GBSS), the cells are centrifuged in an 180-g/L Nycodenz gradient at 2500 g for 20 min. Kupffer cells are carefully sucked by cusp-straws at the pearl layer inderphase. The purified Kupffer cell fractions are finally collected by centrifugal elutriation. The Kupffer cells are washed with Hanks' and resuspended in RPMI 1640 medium containing antibiotics (100-U/mL penicillin, 100-mg/mL streptomycin), 2-mM glutamine, and 10% fetal calf serum. Overall, 1-mL aliquots containing $1 \times 10^6$ cells are added to 24-well culture plates. The cells are incubated for 60 min in a humidified atmosphere containing 5% $CO_2$ at 37°C. Nonadherent cells are removed, and adherent cells are washed twice with PBS. To observe the direct effect, the cells at a density of $1 \times 10^6$/mL are incubated with different concentrations of test compound. The cells

($1 \times 10^6$/well) are cultured for 48 h with 5-μg/mL LPS, the supernatants are collected, and the concentrations of TNF-α and NO are measured.

According to the *in vitro* liver injury model, BCG-induced Kupffer cells are isolated from the livers of the rats injected via the tail vein with 3 mg of BCG 10 days before, and hepatocytes are isolated from the normal rats. The hepatocytes ($1 \times 10^9$/mL), different concentrations of test compound, and BCG-induced Kupffer cells ($1 \times 10^6$/well) are cocultured for 48 h with 5-μg/mL LPS and the supernatants are collected. In a 96-well plate, 100 μL of cell culture supernatant and 100 μL of Griess reagent (10-g/L sulfanilamide and 1-g/L N-1-naphthylethylenediamine dihydrochloride in 2.5% phosphoric acid) are incubated at room temperature for 10 min. Absorbance at 540 nm is measured. The nitrite concentration is calculated by comparing samples with standard solutions of sodium nitrite produced in the culture medium.

Livers are thawed, weighed, homogenized with Tris-HCl (5 mM containing 2-mM EDTA, pH 7.4), and centrifuged (1000 g, 10 min, 4°C), and MDA and SOD in the supernatant are immediately analyzed. MDA in liver tissue is determined by the thiobarbituric acid method [72]. The assay for total SOD is based on its ability to inhibit the oxidation of oxyamine by the xanthine–xanthine oxidase system. The absorbance of the red nitrite produced by the oxidation of oxyamine is determined at 550 nm.

Thymocytes ($2 \times 10^6$/well) from mice are cultured for 48 h in 96-well plates containing RPMI 1640 medium supplemented with 5-μg/mL concanavalin A and 0.1 mL of collected supernatant. Three hours before the termination of the culture, cells are pulsed with MTT stock (in sterile PBS, 5 mg/mL, stored in the dark at 4°C for up to 1 week, before use immediately filtered, 0.22 μm, to remove any formazan precipitate, 20 μL/well), returned to 37°C, and incubated for another 3 h. The plates are centrifuged at 1000 g for 10 min to form cell pellets and MTT formazan products. The supernatant is carefully aspirated without disturbing the pellets, and formazan is solubilized by adding isopropanol (100 μL of isopropanol : 200 μL of supernatant). Insoluble material is then removed by centrifugation at 1000 g for 10 min. The solubilized formazan in isopropanol is collected and distributed into 12-well flat-bottom ELASA plates at a final volume of 100 μL/well. Plates are read at 570 nm within 1 h of adding isopropanol.

### 4.5 Body Weight-Based Liver Tumor Incidence Assay

Addition to extra groups for high and low outliers, the B6C3F1 mice (at the 9-week age point) are assigned to 1 of 17 consecutive weight groups ranging from 20 to 57.5 g in 2.5-g intervals. Each weight group of mice developing a liver tumor is sorted for the calculation of the relative tumor risk. The mice with hepatocellular adenoma, carcinoma, or hepatoblastoma are designated as positive or otherwise designated as negative. The process is repeated for each age point to 68 weeks.

Approximately 4-, 5-, and 6-week-old male B6C3F1/Nctr BR mice are used at receipt, the initiation of controlled feeding, and on the first day of dosing, respectively. During the studies, the health of the mice is monitored. The dose for groups of 120 male mice receiving the solution of chloral hydrate in distilled water is 0, 25, 50, or 100 mg/kg (all at 5 mL/kg), 5 days per week for 104 to 105 weeks, each of which is divided into two dietary groups of 60 mice, and the vehicle controls

received distilled water only. The *ad libitum* fed mice are fed available *ad libitum*, and the dietary controlled mice are fed in measured daily amounts. All mice have water available *ad libitum*. At the 71-week age point, 12 mice of each diet/dose group are euthanized for pathological and biochemical evaluation. At the 110-week age point, the remaining 48 mice of each diet/dose group are fasted overnight before necropsy. Among them the mice with liver tumors refer to mice bearing single or multiple hepatocellular adenomas, hepatocellular carcinomas, and/or hepatoblastomas.

From week 21, the weight dataset of the first experimental group is imported with the preceding week's dataset as a reference. According to the predefined sort criteria (including whether the weight value for the previous week is outside either the 5% or the 12% confidence limits of the idealized weight curve or whether the mouse gains or loses weight in the preceding week), each mouse weight from the dataset is sorted into its appropriate weight group and assigned either the corresponding *ad libitum* or the weight-reduced tumor risk value from the tumor risk tables. The resulting tumor risk dataset in the appropriate week's column of a new table is stored and proceeded to the next week's dataset. According to all weekly datasets from 21 to 68 weeks, the mean tumor risk for each mouse in the dataset is calculated, the survival time for each mouse is imported, and the Poly-3 weighting time at risk factor ($a_{ij}$) for each mouse is calculated. The mean (cumulative) tumor risk dataset is sorted into 2% incremental percentage tumor risk groups, and the corresponding $a_{ij}$ value for each mouse is assigned into the appropriate tumor risk group. This procedure can be adapted to sort individual data for each week rather than the means of all 48 weeks. The resulting tumor risk group table is used to calculate the mean tumor risk estimate and standard deviation of the experimental group to predict the survival-adjusted liver tumor rate, over all liver tumor rates and the number of animals bearing tumors.

### 4.6 Hepatocyte Primary Culture Assay

On anesthetic male SD rats (225–250 g) with sodium pentobarbital (65 mg/kg i.p), tracheotomy is performed. A catheter is flushed with heparin (200 U/mL) to prevent blood clotting of PE 240 polyethylene tubing that is inserted into the trachea to facilitate the respiration. By inserting PE 50 polyethylene tubing, the right femoral artery and vein of the rats are cannulated, and the rats are infused with 0.5-mL saline containing 1-mg/mL sodium pentobarbital/100-g body weight/h through the vein by a Syringe Infusion Pump to supplement fluid loss and maintain the anesthetic level. On the abdomen posteriorly from the xiphoid process of the sternum, a 3-cm median-line incision, which is sutured with two layers of the muscle and skin, is made. Using surgical silk (size 0), left lateral and median liver lobes are ligated and given the partial hepatectomy (2/3 PHX). At various time points after PHX, the diaphragm is cut, and within 2 min from the right ventricle of the heart, a blood sample (5–10 mL per rat) is drawn and centrifuged (2500 g, 20 min) for collecting plasma. In the preparation of serum, the blood is allowed to clot on ice for 10 min, then spun down, with the serum of non-PHX rats as 0 h control. For sham-operated rats, the liver is manipulated without PHX.

In the liver perfusion, the male SD rat (300 g) is anesthetized with sodium pentobarbital (65 mg/kg, i.p.). By surgical silk an 18 G, 1.25-in (32 mm) i.v. catheter is

used as the portal vein catheter tie to the vein. Through a catheter inserted into the inferior vena cava via the right atrium of the heart, the perfusion buffer is drained. The liver is perfused with 400 mL of nonrecycled oxygenated $Ca^{2+}$-$Mg^{2+}$ free Dulbecco's PBS (DPBS, 2.68-mM KCl, 1.47-mM $KH_2PO_4$, 136.9-mM NaCl, 8.1-mM $Na_2HPO_4$, pH 7.4) containing 0.49-mM EDTA to reduce $Ca^{2+}$ in the liver tissue, with 100 mL of DPBS without EDTA to wash out the EDTA in the system and with 100 mL of recycled oxygenated collagenase/Swim's 77 (0.25 mg/mL) with 5-mM $Ca^{2+}$ for 10–15 min. Upon finishing perfusion, the liver is rinsed thoroughly with 10–20 mL of sterile DMEM. The liver is transferred into a sterile petri dish containing fresh medium. Using two forceps, the liver capsule membrane is slit, which is gently shaken and the cells are released into the medium. Two Spectra/Mesh N filters (70 μm, 40 μm, Spectrum) are used to filter the isolated liver cells to remove tissue chunks and cell debris. The cell suspension is filtered by low-speed differential centrifugation (300 g, 3 min, 4°C) for further purification. By aspiration the supernatant is discarded, the cell pellet is gently mixed with fresh medium, and the centrifugation procedure is repeated three times. Using the trypan blue exclusion method, viable cell concentration and percentage of nonhepatocytes is determined. In the purification of hepatocyte, 20 μL of trypan blue (0.4%) are added into 20 μL of cell suspension (dilution factor = 2), which is mixed and loaded onto the hemocytometer. All corner squares (64 squares/chamber) of the two chambers are counted (128 squares total; each square is 1 nL, 128 nL × 7.813 = 1 μL). Viable cells/μL = viable cells per 128 squares × 7.813 × 2. The final concentration is the average value of three separate countings. Viability (%) = viable cells/(viable + dead cells) × 100%. Nonhepatocyte (%) = nonhepatocyte/total cells × 100%.

Each well of six-well plates is coated with 20 μL of the solution of rat tail collagen (0.8-mg/mL double distilled $H_2O$ with 0.1% acetic acid) and allowed to dry. To the well approximately 2 mL of DMEM/F-12 (supplemented with 25-mM sodium bicarbonate, 10-mM HEPES, 100-U/mL penicillin G, and 0.1-mg/mL streptomycin) is added. To each well the suspension of viable cells (final cell concentration, 100,000 cells/mL) are seeded and constantly mixed by agitation. The plates are incubated at 37°C overnight. When the attachment reaches the end, the medium is changed, mildly shaken, and aspirated. The wells are rinsed with 1 mL of medium and aspirated. From three randomly selected wells in the culture, the counts of the viable tarting-cell-number are calculated. Serum or plasma samples in 2 mL of fresh medium (10%) are added to wells, to which heparin (35-U/mL final) is added to prevent the medium from clotting during the culture. Cells grown in serum-free medium and in serum-free medium plus 100-ng/mL EGF and 20-mU/mL insulin are used as the negative and positive control, respectively. The medium is changed by aspiration at 24 h with no rinse. To each well, fresh medium with serum or plasma and other additives are added. At the end of 48 h, culture-attached hepatocytes are harvested and counted. The medium in each well is aspirated, serum or plasma-containing wells are rinsed once with saline and aspirated to wash off various plasma proteins. Overall, 400 μL of trypsin-EDTA is added to each well. The plate is covered and in the incubator for approximately 1 min and checked. The trypsin digestion is assessed. The plate is immediately placed on ice, and the serum of SD rats (40 μL, final concentration 10%) is added into each well to inhibit further trypsin digestion of the cells (440 μL total volume per well). On ice and using a pasteur pipet, the cells are gently blown from the bottom of the well. After

all cells are lifted, the cell suspension is transferred into a 1.5-mL Eppendorf tube and kept on ice for an immediate counting with no centrifugation or trypsin removal. In a plate (dilution factor = 1.25) 40 μL of cells and 10 μL of 0.4% trypan blue are mixed and loaded onto the hemocytometer.

### 4.7 CCL-64 Cell Growth Assay

Serial dilutions of test samples are prepared on 96-well flat-bottomed microplates with 100 μL of CM per well. Before the assay, to activate latent TGF-B, the samples are transiently acidified by first adding HCl to pH 2 and subsequently neutralizing with NaOH. CCL-644 cells (Mv-1-Lu) grown in RPMI 1640 supplemented with 10% FCS, 2-mM L-glutamine, and 40-μg/mL gentamicin (Complete medium, CM) are seeded at $1 \times 10^4$ cells/well and grown for 24 h in a total volume of 0.2-mL CM. After 20 h, the cells are pulsed for 4 h with 1 μCi/well of [methyl-$^3$H]thymidine and harvested with a Micromate 196 cell harvester. The concentration of TGF-β in the sample is determined by the growth inhibition caused by the sample, compared with a standard curve obtained by testing serial dilutions of porcine TGF-β. When the assay is used to measure HGF, the concentration is determined by this cytokine's ability to reverse the growth inhibitory effect of 350 pg/mL of TGF-β.

## 5 ANTI-INFLAMMATORY ASSAY

In the development of serious diseases, inflammatory is frequently involved. The development of inflammatory events may be induced by chemicals such as glucocorticoid; are regulated by endogenous factors such as tumor necrosis factor alpha (TNFα), enzymes and proteins such as copper and zinc-superoxide dismutase (SOD1), proinflammatory peptide substance P (SP), RGD peptides, interleukin-4 (IL-4), IL-10, interferon-γ (IFN-γ), cyclooxygenase-1 (COX-1), cyclooxygenase-2 (COX-2), 5-lipoxygenase (5-LOX), macrophage inflammatory protein (MIP)-1R, glucocorticoid regulated protein CD163, FK506 binding protein 51 (FKBP51), and monocyte chemoattractant protein-1 (MCP-1); implicated by cytokine such as interleukin-1 (IL-1); and mediated by adhesions such as fibrous adhesions and cell adhesions resulting from the receptor, such as $\alpha_{II}\beta_{III}$, human glucocorticoid receptor and chemokine receptors (CCR5), and banding. Based on these facts, the adhesion formation assay [73]; human peripheral blood mononuclear cell (PBMC) proliferation assay [74]; COX-1, COX-2, and 5-LOX assay [75]; CCR5 re-ceptor binding assay [76]; tissue binding affinity assay [77]; G93A-SOD1 transgenic mice assay [78]; mitogen-activated protein kinases (MAPK) p44 (ERK1) and p42 (ERK2) assay [79]; ELA4.NOB-1/cytotoxic T lymphocyte line (CTLL) cell assay [80]; and FKBP51 mRNA assay [81]; and Xylene-induced ear edema assay [82] are established.

### 5.1 Adhesion Formation Assay

For assessment of the neurokinin 1 receptor antagonist (NK-1RA) on peritoneal adhesion formation, a laparotomy is performed through a midline incision, and four ischemic buttons, spaced 1 cm apart, are created on both sides of the parietal

peritoneum by grasping 5 mm of peritoneum with a hemostat and ligating the base of the segment with a 4-0 silk suture. To assess the effects of NK-1RA on adhesion formation, peritoneal adhesions are induced in 42 male Wistar rats (200–250 g) that are randomized to experimental groups receiving the specific non-peptide NK-1RA, the test compound, or vehicle. In the initial study, the rats in the experimental group receive a 0.2-mL i.p. injection of 25-mg/kg NK-1RA twice a day for 2 days. At the time of surgery, 1 ml of a 0.75-mg/mL test compound is given as a peritoneal lavage, and the rats then received i.p. injection for 7 days. Control rats are similarly injected/lavaged with sterile vehicle. This experiment is repeated with 10-mg/kg NK-1RA per day. At day 7, all the rats are killed and the adhesions are quantified in a blinded fashion. Each rat receives a percent adhesion score based on the number of ischemic buttons with attached adhesions.

For assessment of NK-1RA on peritoneal tissue plasminogen activator (tPA) and PAI-1 expression and activity, the temporal expression pattern of tPA and PAI-l mRNA in peritoneal tissue collected from a 0.5-cm radius of the ischemic buttons are determined by RT–PCR analysis at days 0, 1, 3, and 7 after surgery. Based on the results, the effects of NK-1RA administration on tPA and PAI-1 mRNA and protein levels are determined at postoperative day 1 in pedtoneal tissue and fluid by RT–PCR ana1ysis and bioassay, respectively. The rats receive 5.0-mg/kg NK-1RA or vehicle per day. Control samples are collected from 6 nonoperated rats. All samples are immediately frozen in liquid nitrogen and stored at −80°C until used.

Total RNA is isolated from 50 mg of peritoneal tissue with the SV Total RNA Isolation System, and RT–PCR is conducted with the Gene-Amp RNA PCR System. To amplify tPA and PAI-1, the 28 cycles of 95°C, 60°C, and 72°C for 30 s each are used. In the amplification of tPA, the primer sets of 5′-TCTGACTTCGTCTGCCAGTG-3′ (sense) and 5′-GAG-GCCTTGGATGTGG TAAA-3′ (antisense) are used. In the amplification of PAI-1, the primer sets of 5′-ATCAACGACTGGGTGGAGAG-3′ (sense) and 5′-AGCCTGGTCATGTT GCTCTT-3′ (antisense) are used. Overall, 15 μL of PCR products are subjected to electrophoresis on 2% agarose gels containing 0.03-μg/mL ethidium bromide, and the quantitative level of the transcript is determined by scanned photographs of gels. Levels of mRNA expression are normalized to GAPDH, a constitutively expressed gene that does not vary among treatment groups.

Using the corresponding kits, the total levels of tPA and PAI-1 in peritoneal fluid samples are measured. From 12 vehicle-administered, 12 NK-1RA-administered (10.0 mg/kg per day), and 12 nonoperated control rats, peritoneal fluid is collected in 5-mM citrate and 0.1-M acetate for assessment of fibrinolytic activity caused by tPA activation of plasminogen. Overall, 1.0 μL of peritoneal fluid samples are run on 10% SDS polyaerylamide gels containing 0.1% gelatin and 0.002% plasminogen. After electrophoresis the gels are washed twice in 2%Triton X-100 and incubated overnight at 37°C in 0.1-M glycine (pH 8.3). The gels are stained with a 0.25% Coomassie blue solution, and PA activity is visualized as clear bands produced by plasmin lysis of gelatin. Determining the contribution of serine proteases, a 10-mM serine protease inhibitor, PMSF, is added to the developing buffer. In the identification of the zones of lysis corresponding to tPA activity, tPA and/or uPA are immunoprecipitated from peritonea1 fluid samples. To the mixture of 10 μL of peritoneal fluid and 10 μL of buffer containing 40-mM phosphate, 1-M

NaCl, 0.2% SDS, 2% Igepal CA-630, and 1% deoxycholate (pH 7.5), 1 μg of tPA and/or uPA antibodies are added and incubated at 4°C overnight. To each sample, 10 μL of a 50% UltraLink protein A/G slurry is added and incubated overnight at 4°C. Samples are centrifuged at 4°C and 16,000 g for 1 min, and the supematant (50%) is analyzed by zymography for comparison with human recombinant tPA and uPA standards.

### 5.2 PBMC Proliferation Assay

On removal of plasma the heparinized human peripheral blood from 60 mL of healthy donors is centrifuged (4°C, 2000 g, 10 min), and the blood cells are diluted with PBS and centrifuged (1500 g, 30 min) [83]. After removal of red blood cells, the PBMC cell layers are collected, washed with cold distilled water and 10 × Hanks' buffer saline solution, and resuspended ($2 \times 10^6$ cells/mL) in RPMI-1640 medium supplemented with 2% fetal calf serum, 100-U/mL penicillin, and 100-μg/mL streptomycin. In the determination of the lymphoproliferation, 100 μL of the PBMC suspension is deposited into a 96-well flat-bottomed plate with or without 5-μg/mL phytohemagglutinin (PHA, with cyclosporin as a positive control) [84], into which solutions of various concentrations of test compounds are added, the plates are incubated at 37°C for 3 days in humidified atmospheric air containing 5% $CO_2$, tritiated thymidine is added, incubated for another 16 h, and the cells are harvested on glass fiber filters using an automatic harvester and measured with a scintillation counting. In the determination of cell viability, approximately $2 \times 10^5$ T cells with or without PHA are cultured with 0.1% DMSO, solutions of various concentrations of test compounds are added, the plates are incubated for 3 days, the viable cell numbers are counted using a microscope with a hemocytometer following staining by trypan blue, and the percentage of viable cells is calculated. To analyze the cell cycle, 1 mL of the PBMC suspension is added into a 6-well flat-bottomed plate with or without 5-μg/mL PHA [85], 25 μg/mL of each test compound is added, the plates are incubated at 37°C for 3 days in humidified atmospheric air containing 5% $CO_2$, the medium is centrifuged, and the cells are harvested, washed with PBS, and fixed in 70% ethanol at –20°C for 30 min. DNA is then stained with 4-μg/mL propidium iodide containing 100-μg/mL ribonuclease A. Flow cytometric analysis is conducted.

After incubation with PHA alone or in combination with solutions of varying concentrations of test compounds for 3 days, the medium is centrifuged, and PBMC ($2 \times 10^5$ cells/well) supernatants are collected to quantify the concentrations of IL-2, IL-4, IL-10, and IFN-γ by means of enzyme immunoassays.

After stimulation with or without PHA and coculture with 25-μg/mL, each compound for 18 h, the collected PBMC cells ($5 \times 10^6$) are subjected to lysis with RNA-Beek and centrifuged, the supernatants are extracted with a phenol-chloroform mixture, and the extracted RNA is precipitated with isopropanol, pelleted by centrifugation, and redissolved in diethyl pyrocarbonate (DEPC)-treated $H_2O$. The concentration of the extracted RNA is measured using the optical density at 260 nm.

In the synthesis of the first-strand cDNA, 1-μg aliquots of RNA are reverse-transcribed. The mixture of 1-μg RNA in 12.5 μL of DEPC-treated $H_2O$ and 20-μM oligodeoxythymidine (oligo dT) 18 is heated at 72°C for 2 min, and

then quick-chilled on ice. To this mixture, 5.5 μL of concentrated synthesis buffer (50-mM Tris-HCl, pH 8.3, 75-mM KCl, 3-mM $MgCl_2$, 0.5-mM deoxynucleotides triphosphates (dNTPs), and 1 U ribonuclease inhibitor), and 10 U of moloney murine leukemia virus reverse transcriptase are added. The reaction mixture is incubated at 42°C for 1 h and then at 94°C for 5 min. To the reaction mixture, 80 μL of DEPC-treated $H_2O$ is added and the mixture is stored at –20°C for use in the PCR. To 5 μL of the first-strand cDNA, 0.6-μM primers, 1.25-U Taq polymerase, 10-μL reaction buffer (2-mM Tri-HCl, pH 8.0, 0.01-mM EDTA, 0.1-mM dithiothreitol, 0.1% Triton X-100, 5% glycerol, and 1.5-mM $MgCl_2$) and 15.75 μL of water are added and the total volume is 25 μL. According to the air thermocycler, a denaturing temperature of 94°C for 45 s, annealing temperature of 58°C for 45 s, and elongation temperature of 72°C for 1 min for the first 25 cycles, and finally 72°C for 10 min, PCR is carried out and the amplified product is run on 1.8% agarose gels.

### 5.3 COX-1, COX-2, and 5-LOX Assay

For the COX-1 and COX-2 assay, the diluted solution of the test compound is incubated with COX-1 or COX-2 according to the standard method and then the enzyme reaction is initiated by the addition of [1-$^{14}$C]-arachidonic acid. From the incubation mediate, prostaglandin $E_2$ and prostaglandin $D_2$ ($PED_2$) are extracted to measure radioactivity. Indomethacin is used as a positive control. For the 5-LOX assay, the inhibitory activity of the test compound is evaluated. The test compound is incubated with 5-LOX (46 μg of protein) at 24°C for 10 min, and then the enzyme reaction is initiated by the addition of [1-$^{14}$C]-arachidonic acid (50 nCi). After 5 min, 4-M formic acid is added to terminate the enzyme reaction. Indomethacin is used as a positive control.

### 5.4 CCR5 Receptor Binding Assay

In the determination of chemokine receptor (CCR5) binding activity, a 96-well scintillation proximity assay (SPA) format and CHO cells overexpress-ing the human CCR5 receptor are used. From the CHO cells, MIP-1α and membranes are obtained, and after $^{125}$I labeling, the former is converted into [$^{125}$I]-human MIP-1α. The solution of the test compound, 12.5% aqueous DMSO, 12 μg of membranes, 0.17-nM [$^{125}$I]-MIP-1α, 0.25 mg of Wheat Germ Agglutinin-SPA beads, and assay buffer (50-mM HEPES, 1-mM $CaCl_2$, 1-mM $MgCl_2$, 1% BSA, and a protease inhibitor cocktail), is incubated at room temperature for 5 h with shaking. The beads are settled for 2 h, and the total binding is measured via the radioactivity test. In the presence of 1-μM recombinant human MIP-1α, the nonspecific binding is defined. A $IC_{50}$ of 2.7 nM of human MIP-1α is used as a reference.

### 5.5 Tissue Binding Affinity Assay

In hydrophobic teflon bags, the test compound and blood monocytes isolated from pooled buffycoats at a density of $2 \times 10^8$-cells/mL McCoy's medium supplemented with 20% fetal calf serum are cultured for 2 days. The monocytes are washed with

cold PBS (pH 7.4), incubated with BSA (1%) at 4°C for 30 min, washed with PBS, incubated with monoclonal antibody anti-CD163 (5–10 μg/mL) at 4°C for 45 min, washed with PBS, and incubated with FITC-labeled secondary antibody goat-antimouse $IgG_1$ in 1% BSA at 4°C for 30 min. At the last 3 min of the incubation, Propidium iodide is added to determine cell viability and to exclude dead cells using activated cell sorter (FACS) analysis (488 nm, 250 mW, logarithmic amplification).

Human lung cancer-free tissue from patients with bronchial carcinomas and receiving no glucocorticoids for the last 4 weeks before surgery is immediately washed with Krebs-Ringer-HEPES buffer (118-mM NaCl, 4.84-mM KCl, 1.2-mM $KH_2PO_4$, 2.43-mM $MgSO_4$, 2.44-mM $CaCl_2 \cdot 2H_2O$, and 10-mM HEPES, pH 7.4) and sliced into pieces of 1 $mm^3$ or deeply frozen immediately in liquid nitrogen and stored at –70°C.

The frozen human lung tissue is pulverized and homogenized in three aliquots buffer solution A (10-mM TRIS, 10-mM $NaMoO_4$, 30-mM NaCl, 10% glycerol, 4-mM DTT, 5-mM dichlorvos, and 1-mM Complete) with an Ultra Turrax mixer in an ice bath and then centrifuged at 4°C and 105,000 g for 1 h to prepare glucocorticoid receptors (30–60 fmol/mg cytosol [86]). The receptor binding experiments are performed according to the standard method [87]. The radiochemical purity of the labeled glucocorticoid is determined by HPLC, TLC, and scintillation counting.

### 5.6 G93A-SOD1 Transgenic Mice Assay

EOC-20 cells are grown in 75-$cm^2$ cell culture flasks with DMEM supplemented with 10% fetal calf serum and 20% L292 fibroblast-conditioned medium, transferred into 24-well cell culture plates, treated with the solution of test compound in DMSO or DMSO vehicle alone (1% final volume) for 30 min, and challenged with TNFα for 24 h. The culture medium is collected and assayed for determining the $IC_{50}$ of the test compound that suppresses TNFα-induced $NO_2^-$ by 50% via the determination of $NO_2$ [88]. The medium is removed, and DMEM lacking the phenol red indicator is added. To each well, MTS reagent [3-(4,5-dimethylthiazol-2-yl)-5-(3-carboxymethonyphenol)-2-(4-sulfophenyl)-2H-tetrazolium, inner salt] is added and incubated at 37°C for 30–45 min. The aliquots of media are collected and evaluated spectrophotometrically at 540 nm. The MTS solution prepared for the blank is incubated in the absence of cells. Viability is calculated as the ratio of $OD_{540}$ in test compound-treated wells, relative to the same variable measured in wells treated with approximately 25-μM (typically 20 ng/mL) TNFα alone, after subtraction of the blank.

Transgenic mice expressing high copy numbers of human mutant G93A-SOD1 are maintained in the hemizygous state by mating G93A males with B6SJL-TGN females. The mice are fed *ad libitum* standard AIN93G diets or the same diets formulated with nordihydroguaiaretic acid at 2500 ppm; at 90 days of age, drug administration is started to demonstrate motor weakness and fine limb tremors [89]. At 90 days, 100 days of age, and subsequent 5-day intervals, mice are placed on a horizontal rod rotating at 1 rpm every 10 s until they fall from the rod; the mice that are no longer able to right themselves within 10 s of being placed on their sides are killed.

Using TRI reagent, total RNA is collected from the spinal cords of non-transgenic control and G93A+ transgenic mice. In the presence of avian myeloblastosis virus (AMV) RT using oligo$(dT)_{15}$ to prime the reaction, 5 μg of RNA samples are reverse transcribed. On completion, each reaction mixture is diluted with a TE buffer (10-mM Tris, 1-mM EDTA, pH 8.0) to a final volume of 50 μL. PCR amplification of a 309-bp 5LOX gene product is accomplished with Taq DNA polymerase, using the buffer and final concentrations of 1.5-mM $MgCl_2$, 0.2-mM dNTP, and 0.3-μM primer (5′-GGCACCGACGACTACATCTAC-3′, forward and 5′-CAATTTTG-CACGTCCATCCC-3′, reverse), of which the final volumes are 50 μL. A 353-bp PCR product of the β-actin primer (5′-CGGCCAGGTCATCACTATTG-3′, forward and ACT-CCTGCTTGCTGATCCAC-3′, reverse) is used as the normalization control. The optimal cycling conditions of PCR amplification of a 309-bp 5LOX gene product are at 94°C for 2 min, one cycle; at 94°C for 1 min, at 56°C for 1 min, and at 72°C for 1 min for 27 cycles; and at 72°C for 7 min, one cycle. The conditions for β-actin primers are the same as mentioned above, except that the annealing temperature is 54°C and performing 24 cycles. From each reaction, 25 μL of samples are collected, electrophoresed in 2% agarose/TBE [tris/borate/EDTA buffer (0.09-M Tris, 0.09-M borate, 0.002-M EDTA)] gels for 1.5 h, stained with ethidium bromide, and photographed with a NucleoVision imaging system. SL-29 fibroblast lysate is used as the positive control for 5LOX Western blots. On 4–20% gradient polyacrylamide gels, the electrophoresis is performed and the bands are visualized with chemiluminescence detection reagents. In the solution of 50-mM Tris-HCl (pH 6.8), 0.3-M NaCl, 1% β-mercaptoethanol, 1-mM phenylmethylsulfonyl fluoride, and 5-μM leupeptin, the spinal cord samples are homogenized and centrifuged (4°C, 30,000 g, 5 min). The supernanants are boiled for 10 min and centrifuged at 30,000 g and 4°C for 30 min, and the supernatants are dialyzed overnight against 50-mM Tris-HCl. From the heat-soluble fraction of total spinal cord lysate, microtubule-associated tau protein (C-tau) is cleaved and measured by sandwich ELISA, using affinity-purified monoclonal antibody 12B2 [90–92]. In immunohistochemical experiments, the terminally anesthetized mice are successively perfused transcardially with PBS (pH 7.4) and 4% paraformaldehyde in 0.1-M phosphate buffer. The spinal cord is collected, and the lumbar L5 region is processed for paraffin embedding. Serial cross sections of 5 μM thickness of the L5 spinal cord region are prepared for immunostain with antibodies to glial GFAP. In the absence of primary antibody, negative immunohistochemical controls are treated in the same way. GFAP-labeled sections are counterstained with hematoxylin.

### 5.7 MAPK p44 (ERK1) and p42 (ERK2) Assay

By RT–PCR from PMA-treated THP-1 cells, cDNA encoding CCR2B is isolated and subcloned into the expression vector pXMT3-neo [93], which contains an adenovirus late promoter and neo- and DHFR-selectable markers. After transfection with pXMT3-neo-CCR2B, selection with G418, amplification with methotrexate, and clone of CHO DUK-X cells, the resultant CHO-CCR2B cell line is cultured with modified MEMα (without ribonucleosides and deoxyribonucleosides) containing Glutamax-I, 100-U/mL penicillin, 100-μg/mL streptomycin, 10% (v/v) dialyzed FBS, and 80-nM methotrexate. Parental CHO cells are cultured with original MEMα containing 100-U/mL penicillin, 100-μg/mL streptomycin, and 10% (v/v)

FBS, split twice per week by incubation for 5 min with an enzyme-free cell dissociation buffer, and reseeded at a density of 1–2 × $10^4$ cells/$cm^2$. Before an assay, the cells are seeded in methotrexate at 2 × $10^4$ cells/well in flat-bottomed 96-well tissue culture plates, allowed to attach for 6 h, washed with 250-μL/well PBS, and incubated overnight in methotrexate containing endotoxin-free BSA (0.1%, w/v).

After preincubation of 60 μL of 20-nM MCP-1 with 60 μL of 0.4–1.8-μg/mL mAb at 37°C for 30 min, 50 μL of the MCP-1-mAb solution is added. The cells are preincubated at 37°C for 45 min with the solution of test compound in 50 μL of serum-free medium. After the addition of 50 μL of 20-nM MCP-1, the cells are incubated at 37°C for 5 min, the medium is replaced by 100-μL/well methanol equilibrated at –20°C, incubated at –20°C for another 20 min, and then washed three times with PBS containing Triton X-100 (0.1%, w/v). The cells are incubated with $H_2O_2$ (0.6%, v/v) in washing buffer for 20 min and washed as above, by which the activity of endogenous peroxidase is quenched. By adding a 250 μL/well assay buffer (washing buffer containing fraction V BSA, 5%, w/v) and incubating at room temperature for 1 h, the nonspecific binding sites are blocked. Then, the solution is replaced by a 100-μL/well monoclonal anti-phospho-ERK antibody (1 μg/mL in assay buffer). After incubating at 37°C for 1 h and washing three times with washing buffer, to the plate, a 100-μL/well goat-antimouse-IgG coupled with horseradish peroxidase (0.5 μg/mL in assay buffer) is added. After incubation at room temperature for 1 h and the plate is washed six times with washing buffer, 100 μL/well of tetramethylbenzidine is added. After color development for 10–30 min, the reaction is stopped by adding 50 μL of 2-M $H_2SO_4$ and the plates are read at 450 nm.

### 5.8 ELA4.NOB-1/CTLL Cell Assay

The ELA4.NOB-1 cells and CTLL cells maintained in RPMI 1640 containing 2-mM glutamine, 40-μg/mL gentamicin, and 100-U/mL penicillin, are transferred into a tissue culture medium (TCM) containing 5% heatinactivated fetal calf serum (FCS) and 10% FCS and recombinant human L-2 (100 U/mL), respectively. Cell lines are centrifuged (room temperature, 400 g, 1 min) and washed twice in fresh TCM supplemented with 10% FCS by resuspension and centrifugation. By enumeration using trypan blue (0.4%, w/v) in saline, ELA4.NOB-1 and CTLL cells are adjusted to 1 × $10^6$ and 4 × $10^4$ viable cells/mL, respectively. To 96-well round-bottom microtiter plates containing 100 μL of cytokine (control) or test compound, 50 μL aliquots of each cell suspension are placed. After incubation in air containing 5% $CO_2$ at 37°C for 24 h, the cells are treated with 50 μL of 2-μCi/mL tritiated thymidine, incubated for another 24 h, harvested onto printed filtermats that are dried at 30°C for 1 h and heat sealed into sample bags containing 4.5 mL of β-scintillant, and the radioactivity (cpm) is measured.

To each well containing 100-μL aliquots of the standard dilutions, 50 μL of ELA4.NOB-1 and 50 μL of CTLL cell suspensions are added. Using human recombinant IL-2 (0.025–1.6 U/mL) or test compound, the integrity of CTLL responses in the absence of ELA4.NOB-1 cells is measured. Human recombinant IL-1ra (0.1–100 ng/mL), ultrapure natural human TGF-β1 (0.01–10 ng/mL), and test compound are tested on ELA4.NOB-1 and CTLL cocultures in a similar way, generating standard dose response curves for these cytokines. In the tests of the effects of

IL-1ra and TGF-β1 on IL-1 induced activity in the coculture bioassay and on IL-2 induced CTLL proliferation, serial dilutions of IL-1β (0.78–50 pg/mL), IL-2 (0.025–1.6 U/mL), IL-1ra (1–100 ng/mL), TGF-β1 (0.01–10 ng/mL), and test compound in TCM supplemented with 10% FCS are used. Overall, 50-μL aliquots of each dilution of IL-1ra and IL-1β are dispensed and pipetted into 96-well microtiter plates, respectively, and 50 μL of the suspension of ELA4.NOB-1 and CTLL cells are aliquoted into each well, incubated, and treated with $^3$H-thymidine as described above; similar assays are set up for IL-1ra and IL-2, for TGF-β1 and IL-1β, and for TGF-β1 and IL-2.

### 5.9 FKBP51 mRNA Assay

Within a 2-week period between 9 and 10 h by venapuncture, the blood of nine healthy controls without a history of GC medication (37 years old) and one GC hyposensitive patient harboring one nonfunctional GC receptor allele and resulting in a net functional GC receptor expression of approximately 50% is collected in sodium heparin tubes [94], and centrifuged, and the isolated PBMC are washed twice with RPMI 1640 and resuspended in assay medium consisting of RPMI 1640 supplemented with 4-mM L-glutamine, 100-U/mL penicillin, 100-μg/mL streptomycin, and 10% dextran-coated charcoal steroid-stripped fetal calf serum at $0.5 \times 10^6$ cells/mL. Overall, 1-mL aliquots of the PBMC suspension are added onto 24-well plates and incubated overnight at 37°C in humidified atmospheric air containing 5% $CO_2$. To the medium, dexamethasone (DXM) or the test compound is added and incubated for another 24 h to isolate RNA. For activation of PBMC, 0.1-mL aliquots of $0.5 \times 10^5$ cells are added onto a 96-well round-bottom plate and incubated with 1.5-μg/mL etanus toxoid for 9 h. Seventy two hours later, DXM is added. Then 96 h later, RNA is isolated. In a simultaneous control experiment, 78 h later, 1 μCi 3H-thymidine is added, the cells are harvested, and incorporated radioactivity is determined.

For purification of PBMC subsets, cells are isolated, washed twice, resuspended in PBS at $1 \times 10^7$ cells/mL, incubated at room temperature for 30 min with phycoerythrin labeled primary antibodies (mouse-antihuman CD20, CD4, CD8, or CD14), and washed twice with PBS. The PBMC is resuspended in PBS at $1 \times 10^8$ cells/mL, incubated at 4°C for 30 min and labeled with goat-anti-mouse IgG, and washed and resuspended in 0.5 mL of PBS. The labeled and unlabeled PBMC are separated, resuspended in assay medium, divided into 0.1-mL aliquots onto a 96-well round-bottom plate, and incubated overnight. To the wells, DXM is added and incubated for another 24 h to isolate RNA.

For real-time PCR performed on the LightCycler, after the isolation by use of the TriPure reagent and reverse transcription by use of MMLV-RT RNase H Minus of RNA, the synthesized cDNA is diluted 20–40 times in $ddH_2O$ [95]. To calculate the relative expression level of the genes of interest, the relative expression level of the housekeeping gene $\beta_2$-microglobulin ($\beta_2$m) in each sample is determined. Real-time PCR is performed in 10-μL medium containing 5.0 μL of diluted cDNA (0.5 pmol/μL each primer), 10% either DNA master SYBR-green I solution (for the FKBP51 and $\beta_2$m PCR) or DNA hotstart mixture (for the FKBP12, FKBP13, FKBP22, FKBP25, FKBP52, and Cyp40 PCR), and an optimal concentration of $MgCl_2$ (4-mM $\beta_2$m PCR, 3-mM FKBP51 PCR). The mixture is denatured at 95°C

for 30 s and 10 min for DNA Master I kit and Fast Start kit, respectively, and subjected to up to 40 amplification cycles: denaturing 15 s at 95°C, annealing 10 s at 56°C for β2m or at 67°C for FKBP51, and elongation 30 s at 72°C. A single measurement of fluorescence is taken after each elongation step at 82°C.

### 5.10 Xylene-Induced Ear Edema Assay

Male Kunming mice (about 25 g) are randomly divided into three groups of 12 mice, namely the test group, vehicle control group, and positive control group. The mice in the vehicle control group are administrated orally a suspension of Aspirin in CMC at a dosage of 20 mg/kg, and a concentration of 0.3 mg/mL, whereas the mice in the test group are administrated orally a suspension of test compound in CMC at a dosage of 20 mg/kg, 4.0 mg/kg, and 0.8 mg/kg, and a concentration of 2.0 mg/mL, 0.4 mg/mL, and 0.08 mg/mL. Thirty minutes later, 0.03 mL of xylene is applied to both the anterior and the posterior surfaces of the right ear. The left ear is considered a control. Two hours after xylene application, the mice are killed and both ears are removed. Using a cork borer with a diameter of 7 mm, several circular sections are taken and weighed. The increase in weight caused by the irritant is measured through subtracting the weight of the untreated left ear section from that of the treated right ear section.

## 6 THROMBUS-RELATED ASSAY

A series of events such as the platelet aggregation, fibrinogenesis, fibrin adhesion, fibrin aggregation, and vascular inner wall injury are implicated in the thrombosis. Thus the thrombosis, antithrombosis, and thrombolysis may relate to antiplatelet aggregation [96], chemical- and electrical-induced blood vessel injury [97, 98], thread-induced fibrin or platelet adhesion [99], euglobulin clot lysis time [100, 101], fibrinolytic area [102], and reduction of thrombus mass[103, 104].

### 6.1 Antiplatelet Aggregation Assay

Platelet-rich plasma is prepared by centrifugation of normal rabbit blood anticoagulation with sodium citrate at a final concentration of 3.8%. The platelet counts are adjusted to $2 \times 10^5/\mu L$ by the addition of autologous plasma. Platelet aggregation tests are conducted in an aggregometer using the standard turbidimetric technique. The agonists used may be either the usual platelet-activating factor (PAF, final concentration $10^{-5}$–$10^{-7}$ M) and adenosine diphosphate (ADP, final concentration $10^{-5}$–$10^{-7}$ M) or thrombin and collagen. The effects of the tested compound on PAF or ADP-induced platelet aggregation are observed. The maximal rate of the platelet aggregation (Am%) is represented by the peak height of the aggregation curve.

### 6.2 Ferric Chloride-Induced Thrombosis Assay

After overnight fasting, male SD rats (320–380 g) are administered orally water (blank control), test compound (15, 30, 60 mg/kg), and aspirin (positive control, 30 mg/kg), and they are anesthetized with urethane (1.5 g/kg i.p.) at 90 min after

the oral administration. The experiments are carried out according to the modified method described by Kurz et al. [97]. The left common carotid artery is isolated carefully, and a plastic sheet is placed under the vessel to separate it from the surrounding tissue. The surface of carotid artery is covered with a $4 \times 0.5$-cm cotton sheet saturated with 300 μL of $FeCl_3$ solution (25%, w/v) for 15 min. Then, the injured carotid artery segment (0.6 cm) is cut off, from which the formed thrombus is taken out. After drying for 24 h at room temperature in a dehumidifier, the dried weight of the thrombus is measured.

### 6.3 Electrical Stimulation-Induced Arterial Thrombosis Assay

After overnight fasting, male SD rats (250–350 g) are anesthetized with urethane (1.5 g/kg, i.p.). The left carotid artery is isolated carefully. A plastic sheet is placed under the vessel to separate it from the surrounding tissue. Thrombus formation is induced with the modified Hladovec method [98]. The holder incorporates two electrodes, and the temperature sensor is fixed under the exposed carotid artery. Then, the rats are administered intravenously by NS (blank control), test compound (5, 10, 20 mg/kg), and aspirin (4 mg/kg), respectively. After 5 min, a current of 3 mA is delivered for 3 min. With thrombus formation, the temperature of carotid blood is abruptly decreased. The occlusion time (OT) is measured through the temperature sensor and timer on the electric thrombosis stimulator. The rate of thrombosis inhibition is calculated according to as follows: thrombosis inhibition (%) = $(A_1 - A)/A \times 100\%$, where A is the OT of the control group and $A_1$ is the OT of agent groups.

### 6.4 Thread-Induced Thrombosis Assay

Male SD rats (320–380 g) are treated with water (blank control), test compound (15, 30, 60, 120 mg/kg), and aspirin (positive control, 30 mg/kg, b.i.d. for 2.5 d i.g.) and anesthetized by intraperitoneal injection of urethane (1.5 g/kg) at 75 min after the last dose. The arteriovenous shunt operation is carried out as the method described by Umetsu and Sanai [99]. The left jugular vein and the right carotid artery are annulated with a 4-cm-long polyethylene tube (o.d. 1 mm) with heparin (50 U/kg) injection intravenously for anticoagulation. These catheters are connected to the ends of a 15-cm-long polyethylene tube (o.d. 2 mm) containing a 5-cm-long suture silk thread (no. 4) measured wet weight. At 120 min after the last treatment, the blood flowing through the shunt is confirmed for 15 min. The silk thread with thrombus is gently removed and measured immediately for the wet gross weight. The thrombus wet weight is determined by subtracting the premeasured silk wet weight from the gross weight. The rate of thrombosis inhibition is calculated as follows: thrombosis inhibition (%) = $(A - A_1)/A \times 100\%$, where A is the thrombus wet weight of the control (water) and $A_1$ is the weight after treatment with the agents.

### 6.5 Euglobulin Clot Lysis Time (ECLT) Assay

The rabbit euglobulin clots are prepared according to a published method [100, 101]. Plasma diluted at 1:20 in distilled water is precipitated at pH 5.7 with acetic

acid (0.25%). After 30 min at 4°C, the suspension is centrifuged at 2000 g for 15 min and the precipitate is resuspended to the initial plasma volume with 50-mM sodium barbiturate buffer (pH 7.8, containing 1.66-mM $CaCl_2$, 0.68-mM $MgCl_2$, and 93.96-mM NaCl). To the rabbit euglobulin clots, NS (blank control), UK (positive control), or test compound is added and the ECLT or time to clot lysis is determined in a 96-well microtiter plate.

### 6.6 Fibrinolytic Area Assay

The fibrinogen–agarose mixture is prepared and coagulated with thrombin in plastic dishes according to a published procedure [102]. The fibrinogen–agarose mixture is prepared by mixing equal volumes of 0.3% of rabbit fibrinogen and 0.95% of agarose solutions, both dissolved in 50-mM sodium barbiturate buffer (pH 7.8). The fibrinogen–agarose mixture is coagulated with 100 mL of thrombin (100 U/mL) in plastic dishes (90 mm diameter × 1 mm depth). After 30 min at 4°C, an adequate number of wells, 5 mm in diameter, are perforated. To determine fibrinolytic activity, 30 μL of NS (blank control), UK (positive control), or test compound is added to the corresponding well. The plate is incubated, and areas of lysis are quantified by lysis area.

### 6.7 Thrombolytic Assay

Male Wistar rats (200–300 g) are anesthetized with pentobarbital sodium (80.0 mg/kg, i.p.). The right carotid artery and left jugular vein of the animals are separated. To the glass tube containing 1.0 mL of blood obtained from the right carotid artery of the rat, a stainless steel filament helix (15 circles; 15 mm × 1.0 mm) is added immediately. Fifteen minutes later, the helix with thrombus is carefully taken out and weighed. It is then put into a polyethylene tube that is filled with heparin sodium (50-U/mL NS), and one end is inserted into the left jugular vein [103, 104]. Heparin sodium is injected via the other end of the polyethylene tube as the anticoagulant, after which the test compound is injected. The blood is circulated through the polyethylene tube for 90 min, after which the helix is taken out and weighed. The reduction of thrombus mass is recorded.

## 7 IMMUNOMODULATING ASSAY

Immunomodulation not only relates to a series of physiologic and pathologic phenomena, but also it can be estimated by related reaction cells, such as mast cells, DCs and NK cells, and level of cell factors such as interleukin, interferon, tumor necrosis factor, and transforming growth factor. Based on those facts, the rat mast cell and rabbit aortic assay [105], dendritic cell assay [106], lymphoid organ assay [107], IFN-γ assay [108], anti-rHuEPO NAb assay [109], and chemotaxic assay [110] are established.

### 7.1 Rat Mast Cell and Rabbit Aortic Assay

The ice-cold PBS (10 mL, 137-mM NaCl, 2.68-mM KCl, 0.91-mM $CaCl_2$, 8.1-mM $Na_2PO_4$, 1.47-mM $KH_2PO_4$, 0.91-mM $MgCl_2$, 5.6-mM glucose, and 20.0-mM

HEPES) is injected (i.p.) into male Wistar rats (200–250 g). Approximately 90–120 s later, PBS is collected and peritoneal is subsequently washed by 5 mL and 10 mL of PBS. The washing PBS are also collected and combined with first PBS, followed by centrifugation (200 g, 5 min, 4°C). After washing twice with PBS, the pellet is resuspended into 10 mL of PBS. From the rat, peritoneal lavage mast cells are isolated.

Preincubation of 1.8 mL of mast cell suspensions are carried out at 37°C for 10 min, and 0.1 mL of test compound is added to stimulate the mast cells, which are cooled in ice to terminate histamine release. After centrifugation (100 g, 10 min, 4°C), to the supernatant and the pellet, an equal volume of 0.8-M $HClO_4$ and twice volume of 0.4-M $HClO_4$ are added, respectively. A mixture of 125 μL of 5-M NaOH, 0.4 g of NaCl, and 2.5 mL of n-butanol is added to 1-mL solution of test compound, the mixture is centrifuged (200 g, 1 min, room temperature) and separated. The upper organic phase is mixed with 2 mL of 0.1-M NaOH saturated with NaCl and centrifuged (200 g, 1 min, room temperature), and the procedure is repeated, of which the upper organic phase is mixed with 2 mL of 0.1-M HCl and 7.6 mL of n-heptane. The lower aqueous phase (1 mL) is mixed with 0.1 mL of 10-M NaOH. By incubation with 0.1 mL of o-phthalaldehyde (10-mg/mL methanol) at room temperature for 4 min to perform the histamine-o-phthalaldehyde conjugation, and by addition of 0.6 mL of 3-M HCl to terminate the conjugation. The fluorescence of histamine-o-phthalaldehyde conjugate is assessed at 450 nm with emission excited at 360 nm.

The thoracic aorta of male Japanese White rabbits (3–3.5 kg) is cut into helical strips (approximately 4 mm wide and 20 mm long), and the endothelium is removed by gently rubbing the endothelial surface with cotton pellets. In 1 mL of organ bath containing the modified Krebs–Ringer–bicarbonate solution (120-mM NaCl, 4.8-mM KCl, 1.2-mM $CaCl_2$, 1.3-mM $MgSO_4$, 25.2-mM $NaHCO_3$, 1.2-mM $KH_2PO_4$, 5.8-mM glucose), the strips are mounted and suspended. With a force-displacement transducer connected to a polygraph muscle, the tensions of the strips are recorded isometrically. A passive tension of 1 g is initially applied and the strips are equilibrated for 60 min, after which, 60-mM NaCl in the modified Krebs–Ringer–bicarbonate solution is replaced by equimolar KCl for precontraction of the strips. Reaching a steady level of the response, the experiment is started. Contractile response to histamine is normalized with that of high $K^+$.

### 7.2 Dendritic Cell (DC) Assay

Before being used either for DC culture or for preparing T cells, human peripheral blood mononuclear cells (PBMCs) isolated from freshly leukapheresed blood are centrifuged and cryopreserved in 90% autologous serum and 10% DMSO. By negative depletion and using anti-HLA-DR monoclonal antibody-conjugated paramagnetic beads from allogeneic PBMCs, enriched T cells are prepared. The cells possessing potential costimulatory function are removed from the PBMCs, as suspension each batch of which is tested for the presence of any remaining B cells, monocytes, and 80–90% T cells. Using cell-specific paramagnetic bead preparations and by biomagnetic separation of the stimulators, B cells and T cells are purified and the purity of each cell type is greater than 90%, as determined by flow cytometry.

The cryopreserved PBMCs are thawed in warm AIM-V medium, washed with PBS, and resuspended in Opti-MEM medium supplemented with 1% heat-inactivated autologous plasma to prepare (5–10) $\times 10^6$ cells/mL suspension. The suspension of $1 \times 10^9$ cells is transferred into T-75 culture flasks and cultured for 1 h, the nonadherent cells are resuspended, aspirated out, and stringently washed with cold PBS to remove loosely adherent cells. To each flask, $1.5 \times 10^9$ Opti-MEM medium containing 5% heat-inactivated autologous plasma, 500-U/mL rhGM-CSF, and 500-U/mL rhIL-4 are added and the adherent cells are incubated for 6 days. These DCs are then treated either with BCG alone or BCG plus IFN-γ for 24 h.

In the COSTIM bioassay, allogeneic T cells are thawed in warm AIM-V culture media, and washed with and resuspended in PBS at $1 \times 10^5$ DCs or $1 \times 10^6$ T cells/mL. To each triplicate well of a U-bottom 96-well plate, 100 μL of $1 \times 10^4$ DCs and 100 μL of $1 \times 10^5$ allogeneic T cells are successively added, and with or without 0.005-μg/mL anti-CD3 monoclonal antibody, which is incubated at 37°C for 44 h in humidified atmospheric air containing 5% $CO_2$. A total of 0.5 μCi Tritiated ($^3$H)-thymidine in 50 μL of AIM-V is added to each well, cultured for the last 18 h, the cells are harvested, and the incorporated radioactivity is quantified. Before 1 h of adding T cells and anti-CD3, the sterile, azide-free, and low-endotoxin IgG1 monoclonal antibodies specific for CD54, CD80, CD86, and an isotype control (BD Pharmingen) are added to the DCs at 1 μg/well to observe the costimulatory molecule block. In the mixed lymphocyte reaction (MLR), the cryopreserved DCs (stimulators) and T cells (responders) are thawed in warm AIM-V media, washed with PBS, and resuspended in PBS at $1 \times 10^5$ DCs or $1 \times 10^6$ T cells/mL. To each well of a U-bottom 96-well plate, 100 μL of $1 \times 10^4$ DCs and 100 μL of $1 \times 10^5$ allogeneic T cells are successively added and incubated at 37°C for 6 days in humidified atmospheric air containing 5% $CO_2$. A total of 0.5 μCi Tritiated ($^3$H)-thymidine in 50 μL of AIM-V is added to each well, cultured for the last 18 h, the cells are harvested, and the incorporated radioactivity is quantified.

### 7.3 Lymphoid Organ Assay

After a 4-week acclimation period, male Wistar rats (4 weeks old, 225–250 g, housed in polypropylene cages in a environment-controlled room maintained at 22°C, 55% relative humidity, and a 12:12 h light/dark cycle) are allocated to six groups (n = 15–20 rats each). The untreated group is used as control (maintained on basal diet and sacrificed at weeks 4 and 30). During weeks 1 and 2, the DMBDD, DMBDD/PB, and DMBDD/2-AAF groups are sequentially treated with initiators DEN (N-nitrosodiethylamine, 100 mg/kg, i.p.), MNU (N-methyl-N-nitrosourea, 20 mg/kg, i.p., four times, two doses per week), and BBN (0.05% in drinking water during 2 weeks); during weeks 3 and 4, DHPN and DMH groups are sequentially treated with DHPN (0.1% in drinking water during 2 weeks) and DMH (1,2-dimethylhydrazine, 40 mg/kg, s.c., four times, two doses per week). At the end of the week 4, some animals of the DMBDD group are killed and the remainder is maintained on a basal diet until week 30. After the initiation, the DMBDD/PB and DMBDD/2-AAF groups are supplied with phenobarbital (PB, 0.05%) and 2-acetylaminofluorene (2-AAF, 0.01%) in the diet for 25 weeks, respectively. From the sixth week until week 30, two noninitiated groups receive PB or 2-AAF in the diet. At the week 4 or 30, all animals are killed under pentobarbital (45 mg/kg)

anesthesia. The liver, kidneys, spleen, thymus, mesenteric lymph nodes, and bone marrow removed from all animals are fixed in buffered formalin for 48 h for tissue processing and histological analysis. Only at week 30 are the lung, small and large intestine, and Zymbal's gland examined. After removal of the liver, the kidneys, spleen, and thymus are weighed immediately. The spleen is cut in two halves for evaluation of cytokines and histological analysis, respectively. All removed organs are embedded in paraffin and stained with hematoxylin and eosin for histological analysis [111]. The suspensions of spleen cells dispersed in a Petri dish containing RPMI-1640 culture medium are centrifuged; resuspended in RPMI-1640 culture medium supplemented with 20-mg/mL gentamycin, 2-mM glutamine, and 10% inactivated fetal calf serum; and washed twice by centrifugation at 1500 g for 10 min. To aliquots of $2 \times 10^6$ cells/mL (500 μL/well) in 24-well flat-bottom microtiter plates, RPMI-1640 (500 μL/well) or Concanavalin A (CON A-2.5 μAg/mL, 500 μL/well) or Staphylococcus aureus Cowan's strain 1 (SAC-1 : 5000, 500 μL/well) is added and the plates are incubated at 37°C for 72 h in humidified atmospheric air containing 5% $CO_2$. After incubation, the collected supernatants are stocked at –70°C for quantification of cytokines. CONA *in vitro* stimulated samples are used to quantify IL-2, IFN-γ, IL-10, and TGF-β1. SAC *in vitro* stimulated samples are used to quantify TNF-α and IL-12. Using ELISA kits, cytokine production, IL-2, IL-12, TNF-α, IFN-γ, and IL-10 levels are measured. Samples are acidified by 1-M HCl and measured by Quantikine antihuman TGF-h1 kit for detection of the TGF-h1 immunoreactive form.

### 7.4 IFN-γ Assay

Human myelomonocytic KG-1 cells (ATCCCCL246) are incubated with RPMI-1640 medium supplemented with 10% FCS, 100-μg/mL penicillin, and 100-μg/mL streptomycin at 37°C in air containing 5% $CO_2$. By expression of the corresponding cDNA in *E. coli* HuIL-18 and MuIL-18 are prepared and purified to homogeneity. Using Pfu DNA polymerase (at 95°C for 45 s; at 72°C for 3.5 min; 10 cycles and at 95°C for 45 s; at 68°C for 3.5 min; 35 cycles) MuIL-18R cDNA (1.7 kbp) are amplified from murine liver RNA by RT–PCR and cloned into pCRScript Cam SK (+) to synthesize 5′-AGAGGAACCACCCACAACGATCCT-3′ and 5′-TGAATAGGCACACGCAT-XGACCTCT-3′ [112]. With the EF-1 promoter of pEF-BOS vector, the dihydrofolate reductase unit of pSV2dhfr (ATCC 37146) and the backbone of pRc/CMV vector pREF-XN is constructed. IL-18R cDNA is ligated into *Xho*I/*Not*I sites of the vector to form MuIL-18R expression vector pRcEFM18R. KG-1 cells ($1 \times 10^7$) are washed twice with RPMI-1640 medium and transfected with 50 mg of pRcEFM18R by electroporation. In the presence of 400-μg/mL G-418, the transformed cells are selected and cloned. The suspension of $2 \times 10^6$ cells, on which the receptor binding of $^{125}$I-labeled MuIL-18 or HuIL-18 has been examined, in RPMI-1640 containing 0.1% $NaN_3$ and 100-mM HEPES (pH 7.2) is incubated at 4°C for 1 h with approximately 4 ng of $^{125}$I-labeled MuIL-18 or HuIL-18. After the separation of unbound IL-18, the cell-bound $^{125}$I count is determined. Subtracting the nonspecific binding measured from 3 μg of unlabeled cognate ligand, the specific binding of IL-18 is obtained. To prepare mice serum containing endogenous MuIL-18, C57BL/6 mice are treated with 500 μg of heat-killed *Propionibacterium acnes* for 1 week and challenged with 1 μg of lipopolysaccharide for 2 h to induce endotoxic shock. From

the heart, under proper anesthetization, blood samples are taken to prepare sera. The MuIL-18R-expressing KG-1 cells are washed and resuspended at $5 \times 10^5$ cells/mL with RPMI-1640 medium for 2 days. The cells are adjusted to $1 \times 10^6$ cells/mL with RPMI-1640 containing 10% FCS, to which the indicated amounts of MuIL-18, HuIL-18, or serum sample are added. One day later, the culture supernatants are recovered and the quantitative analysis of the produced IFN-γ is performed by ELISA. The culture supernatant is incubated with mAb-IFN-γ-15. The bound IFN-γ is further incubated with mAb-IFN-γ-6 and detected with hydrogen peroxide and o-phenylenediamine.

### 7.5 Anti-rHuEPO NAb Assay

32D-EPOR cells are incubated at 37°C in humidified atmospheric air containing 5% $CO_2$ with RPMI 1640 supplemented with 15% heat-inactivated FBS, 2-mM l-glutamine, and penicillin/streptomycin (1%, v/v) mixture. The EPO-dependent cells are incubated in RPMI 1640 supplemented with 10 U/mL of rHuEPO. Via two to three subcultures a week, cell densities are maintained between $3 \times 10^4$ and $1 \times 10^6$ cells/mL. Up to 30 days after thawing, the old cells in the cryopreserved cells are discarded and a vial of frozen cells is thawed and expanded in culture [113].

In the cell proliferation assay, 32D-EPOR cells are incubated overnight, harvested, and washed twice with RPMI 1640 lacking rHuEPO by centrifugation (200–300 g). The supernatant is discarded, the cell pellet is resuspended in RPMI 1640, and the suspention is recentrifuged. The formed cell pellet is resuspended in RPMI 1640 and adjusted to $5 \times 10^5$ cells/mL. Cells are incubated at 37°C (humidified atmospheric air containing 5% $CO_2$, without rHuEPO) for 16–24 h, centrifuged, resuspended in fresh RPMI 1640, and counted. Overall, 100 μL of $2 \times 10^4$ staged 32D-EPOR cells and 100 μL of prepared testing sample are incubated at 37°C for 44 h in humidified atmospheric air containing 5% $CO_2$. The solution of 2-μCi [methyl-3H] thymidine diluted in 50 μL of RPMI 1640 is added to each well and incubated for 4 h. The contents of the plate are harvested, 25 μL of scintillation fluid are added, and the cells are counted.

All assay controls are prepared in a mixture of 5% human serum and 15% pooled rat serum. In anti-rHuEPO NAb assay, the background control (N) consisting of cells only is prepared by mixing 40 μL of pooled human serum with 240 μL of RPMI 1640 and 120 μL of pooled rat serum. The maximum growth control (M) consisting of cells and 1-ng/mL rHuEPO is prepared by mixing 40 μL of pooled human serum with 120 μL of rHuEPO at 6.67-ng/ml, 120 μL of RPMI 1640 and 120 μL of pooled rat serum. The neutralizing antibody positive control (P) consisting of cells, 1-ng/mL rHuEPO and 500-ng/mL positive control antibody is prepared by mixing 120 μL of RPMI 1640 with 120 μL of pooled rat serum. Samples are prepared by mixing 40 μL of individual donor serum with 120 μL of rHuEPO at 6.67 ng/mL, 120 μL of RPMI 1640, and 120 μL of pooled rat serum. Before addition to the cells, all controls and samples are preincubated at room temperature for at least 30 min.

### 7.6 Radial Assay of Chemotaxis

From stock sorocarp cultures of *D. discoideum* grown on agar plates, spores from strain v12 are harvested and heat-shocked at 45°C for 30 min. The suspension of

the spores is mixed with a full loop of *E. coli* B/r, the 200-μL aliquots in new SM-agar plates are cultured at 22°C for 24 h, the cells are harvested and maintained vegetatively by shanking (200 rpm in Lpp medium, approximately $10^6$ cells/mL), or the cells are starved by washing four times by centrifugation (3000 g, 30 s). The cells are shaken in 15-mM Tris-HCl (pH 7.0) for 2 h or 4 h; the latter is for routine chemotaxis assays. Chemotaxis of *D. discoideum* amoebae is performed on thin agar plates. Overall, 1 mL of 1.0% agarose in 15-mM Tris-HCl (pH 7.0) is added to each Petri plate and agitated to allow even spreading. The chemoattractants cAMP or folic acid are added to the agar with or without agonists before pouring plates. Chemotactically competent cells are centrifuged to form a viscous suspension and spotted on the agar plates. Each plate containing six aliquots of cells is covered, incubated at 22°C for 3 h, uncovered and placed on a heater to fix cells and dessicate the agar. As a measure of chemotactic efficiency, the initial spot diameters are subtracted from the diameters of visible rings or "halos" formed by outwardly migrating cells. During a period of 60 min with a playback time of 3 min, the individual halos on the agar plates are monitored. Proteins (20 μg/lane) are separated from whole cell lysates by polyacrylamide gel electrophoresis in 12% sodium dodecyl-sulphate polyacrylamide gels and transferred to nitrocellulose membranes in a mini-trans-blot cell using a buffer containing 25-mM Tris, 192-mM glycine (pH 8.3), and 20% methanol. Transfer is with 100 V for 40–60 min employing a frozen cooling unit. Using standard curves produced by running prestained Rainbow markers in parallel lanes, relative molecular weights are estimated. Nitrocellulose blots are blocked in 4% BSA at 4°C for 16 h, incubated for 1 h in 1/500 monoclonal anti-phosphotyrosine PT-66, and diluted by a 1/1000 dilution of peroxidase conjugated goat-antimouse IgG at room temperature for 1 h to detect the proteins containing phosphorylated tyrosine residues. Between each incubation, blots are washed three times for 5 min and immunoreactive bands are visualized using the chromogenic substrate diaminobenzidine HCl and 0.025% peroxide.

## 8 ESTROGEN ASSAY

After transfection, some cells become special fused cells such as the GAIA–DNA domain binding estrogen receptor yeast [114], the HELNα and HELNβ transfected cells [115], the yeast stably expresses human estrogen receptor β (hERβ) [116], the yeast stably expresses human estrogen receptor α (hERα), and enhanced green fluorescent protein (yEGFP) in response to estrogens α (hERα) [117]. With the special characteristics and high response to estrogen stimulation, the mentioned cells and related substances are used for bioassays.

### 8.1 Yeast Oestrogen Assay

The yeast, of which the steroid-binding domain is fused by GAL4-VP16, is incubated with SC-medium without histidine at 30°C and 130 rpm overnight. By adding DMSO up to a final concentration of 15% (v/v), stock cultures are prepared from exponentially growing cultures and stored in 0.5-mL aliquots at –80°C. Exponentially growing overnight, cultures are diluted with SC-medium to an $OD_{600nm}$ of 0.75. To 10-mL aliquots 100 μL of DMSO (negative controls), 100 μL of 17β-estradiol in DMSO (positive controls), or 100 μL of test compound in DMSO are added, incubated at 30°C and 130 rpm for 2 h, diluted to five-fold volume and

determined to get $OD_{600nm}$. To 200 μL of the test culture, 600 μL of Z-buffer (60-mM $Na_2HPO_4 \cdot 7H_2O$, 40-mM $NaH_2PO_4 \cdot H_2O$, 10-mM KCI, 1-mM $MgSO_4 \cdot 7H_2O$, 35-mM β-mercaptoethanol), 20-μL SDS solution (3.5 mM), and 50 μL of chloroform are added, carefully mixed, and pre-incubated at 28°C for 5 min, and then 200 μL of o-nitrophenyl-3-D-galactopyranoside in Z-buffer (13.3 mM) are added to initiate the enzyme reaction. The cultures are incubated at 28°C until a significant yellow color develops. For 17β-estradiol-induced positive controls, the yellow color occurs within 20 min. For test chemical-induced assays, the yellow color occurs after 120 min. To the cultures, 500 μL of $Na_2CO_3$ (1 M) is added to stop the reaction, the cell debris is pelleted by centrifugation (25,500 g, 15 min), and the $Ex_{420nm}$ of the supernatants is determined. The β-galactosidase activity of the test cultures is calculated according to u[μ mol/min] = Cs/t·V·ODs, wherein t = incubation time (min) of the enzyme reaction, V = volume (0.2 $cm^3$) of the used test culture aliquot, ODs = $OD_{600nm}$ of test culture, Cs = concentration of o-nitrophenyl-β-D-galactopyranoside (μM) in the reaction supermatant calculated according to $Cs(\mu M) = 10^6 \cdot [Exs\text{-}Ex_B]/\epsilon_N \cdot d$, wherein Exs = $Ex_{420nm}$ of the enzyme reaction supernatant of test compound, ExB = $Ex_{420nm}$ of the enzyme reaction supernatant of the blank control, $\epsilon_N = \epsilon$ for o-nitrophenyl-β-D-galactopyranoside in the enzyme assay reaction mixture ($4666 \times 10^3 cm^2$/mole), and d = diameter of the cuvette (1 cm). $EC_{50}$ are calculated from dose response curves obtained by fitting the data by $Y = D + [A - D]/[1 + (C/X)^B]$, wherein X = estrogen concentration in the test, Y = β-galactosidase activity, B = relative slope of the middle region of the curve as estimated from a linear/log regression of the linear part of the dose response curve, C = estrogen concentration at half maximal response, and D = minimum β-galactosidase activity. The relative β-galactosidase induction factor defined as the ratio of β-galactosidase activity in a chemical sample and in the corresponding negative control is used for test chemical sample screening results.

### 8.2 HELN α and HELN β Transfected Cell Assay

After incubation with 30 mg of C18 Oasis HLB batch for 10 min, 800 mL of every patient's serum are desteroided and centrifuged (13,000 g, 10 min) to separate the supernatant. To this stripped serum, a known amount of estradiol is added to obtain a set of estradiol standards with concentrations from $10^{-9}$ to $10^{-4}$ mM, and incubated for 4 h at 37°C to permit an equilibrium bet-ween free estrogens and estrogens bound to sex hormone-binding globulin. HeLa cells transfected with ERE-β Glob-Luc-SVNeo and pSG5-Puro-hERα or β, so-called HELNα and HELNβ, are selected by geneticin and puromycin at 1 mg/mL and 0.5 μg/mL, respectively. Luminescent and inducible clones are identified using photon-counting cameras. HELN Erα is cultured in 150-$cm^2$ plastic flasks in DMEM without phenol red, supplemented with 5% dextran-coated and charcoal-treated FCS, 0.25-μg/mL puromycin, and 0.5-mg/mL geneticin. To the culture medium, the aromatase inhibitor aminogluthetimide (AG) is added at a concentration of 50 μM. The cells are seeded in 96-well plates ($3 \times 10^4$ cells per well) in the presence of DMEM without phenol red, supplemented with 3% dextran-coated and charcoal-treated FCS, 50-μM AG, and incubated for 8 h at 37°C. Culture medium is replaced by 100 μL of DMEM without phenol red, supplemented with 50-μM AG, to which 20 μL of human serum in triplicate is added.

### 8.3 Luciferase Assay

The p403- and p405-GPD yeast expression vectors and the p406-CYC1 yeast expression vector are used to express the human estrogen receptor α and β and construct the reporter plasmid, respectively. On the isolated mRNA of T47D human breast cancer cells and intestinal Caco-2 cells, cDNA is synthesized. Using the T47D cDNA, marathon uterus cDNA, and the human intestine cDNA by PCRs, full-length human estrogen receptor β (ERβ) cDNA is obtained. To perform the first PCR, 34.2 μL of ultra-pure water, 5 μL of 25-mM $MgCl_2$, 5 μL of Expand HF 10 × concentrated buffer (without $MgCl_2$), 0.8 μL of 25-mM dNTP mix, 1 μL of the enzyme mixture, 2 μL of the different cDNAs, and 2 μL of a primer mixture containing 10 μM of each primer are pipetted into a thin-walled PCR tube and (1) denature template 3 min at 95°C, (2) denature template 30 s at 94°C, (3) anneal primers 1 min at 60°C, (4) elongation 2 min at 72°C, (5) go to step (2) and repeat 35 times, (6) elongation 7 min at 72°C, and (7) for over 10°C. The second PCR is performed with the same conditions, but 2 μL of the first PCR mixture is used instead of 2 μL of the different cDNAs. The 5′- and 3′-primer, 5′-CGTCTAGAGCT GTTATCTCAAGACATGGATATAA-3′ and 5′-TAGGATCCGTCACTGAGA CTGTGGGTTCTG-3′, contains a restriction site for *Xba*I just before the ATG start codon and for *Bam*HI just after the TGA stop codon, respectively. The full-length ERβ PCR product is isolated and ligated into a pGEM-T Easy Vector. The uterus/intestine human ERβ cDNA clone that fully corresponded to the ERβ sequence is cut out of the pGEM-T Easy plasmid with *Xba*I and *Bam*HI and cloned into the corresponding *Xba*I–*Bam*HI site of the p405-GPD-ERβ vector, which is used to transform Epicurian Coli XL-2 Blue Cells.

Yeast K20 is transfected with the p406-ERE2s2-CYC1-yEGFP reporter vector and integrated at the chromosomal location of the Uracil gene via homologous recombination to construct yeast hERα and hERβ cytosensors. The transformants are grown on MM/LH plates using PCR and Southern blot hybridization to select clones, in which the integration has occurred at the desired URA3 site with only a single copy of p406-ERE2s2-CYC1-yEGFP. This strain is transformed with the p403-GPD-Erα and the p405-GPD-ERβ, or both expression vectors and transformants are grown on MM/L, MM/H, or MM plates, respectively.

The plate with selective MM/L, MM/H, or MM medium is inoculated with –80°C stock yeast ERα, ERβ, or ERα/β cytosensor (20% glycerol, v/v), respectively; incubated at 30°C for 24–48 h; and stored at 4°C. The day before the assay, a single colony of the yeast cytosensor is inoculated with 10 mL of the corresponding selective medium overnight at 30°C with vigorous orbital shaking at 225 g. At the late log phase, the yeast ERα, ERβ, and ERα/β cytosensor culture is diluted (1:10) in MM/L, (1:20) in MM/H, and (1:20) in MM, respectively. This minimal medium consists of a yeast nitrogen base without amino acids or ammonium sulphate (1.7 g/L), dextrose (20 g/L), and ammonium sulphate (5 g/L). The MM/L and MM/H medium are supplemented with L-leucine (60 mg/L) or L-histidine (2 mg/L), respectively.

A total of 200-μL aliquots of the yeast culture are pipetted into each well of 96 well plates, and then 1 μL of ethanol or DMSO stock solution of test chemical is added to result in 0.5% final concentration. The plates are included for 4 h and 24 h, fluorescence at 485 nm and emission at 530 nm are measured directly, and the OD

of the yeast culture at 630 nm is determined to check the toxicity of test chemical for yeast.

### 8.4 Green Fluorescent Protein Expression Assay

Aliquots of 2 mL of blank calf urine and 17β-estradiol (E2β, 1 ng/mL), diethylstilbestrol (DES, 1 ng/mL) and 17α-ethynylestradiol (EE2, 1 ng/mL), α-zearalanol (30 or 50 ng/mL) and mestranol (10 ng/mL), and spiked calf urine samples are adjusted to pH 4.8 and 20 μL of β-glucuronidase/arylsulfatase (3 U/mL) are added. The samples are deconjugated by enzyme at 37°C overnight, treated with 2 mL of 0.25-M sodium acetate buffer (pH 4.8) and subjected to solid phase extraction (SPE) on a C18 column conditioned previously with 2.5 mL of methanol and 2.5 mL of sodium acetate buffer. The column is successively washed with 1.5 mL of 10% (w/v) sodium carbonate solution, 3.0 mL of water, 1.5 mL of sodium acetate buffer (pH 4.8), 3.0 mL of water, and 2 mL of methanol/water (50/50 v/v). The column is air-dried and eluted with 4 mL of acetonitrile. The eluate is applied to an $NH_2$-column conditioned previously with 3.0 mL of acetonitrile. The acetonitrile eluate is evaporated to 2 mL by nitrogen gas stream, of which a 100-μL part (equivalent to 100-μL urine) is transferred to a 96-well plate and mixed with 50 μL of water and 2 μL of DMSO. The plate is dried overnight in a fume cupboard to remove the acetonitrile and screened on estrogenic activities with the yeast estrogen bioassay. In the same way, a reagent blank is prepared, using 2 mL of the 0.25-M sodium acetate buffer pH 4.8 instead of urine.

The yeast cytosensor (20% glycerol v/v) expressing the human estrogen receptor α (hERα) and yEGFP in response to estrogens is incubated in an agar plate containing the selective MM/L medium at 30°C for 24–48 h and then stored at 4°C. The day before the assay, a single colony of the yeast cytosensor is inoculated in 10 mL of selective MM/L medium overnight at 30°C with vigorous orbital shaking at 225 rpm. At the late log phase, the yeast ERα cytosensor is diluted in MM/L, giving an OD at 604 nm in the range of 0.07–0.13. For exposure in 96-well plates, aliquots of 200 μL of this diluted yeast culture are pipetted into each well, already containing the extracts of the urine samples. A E2β dose-response curve is included in each exposure experiment. Aliquots of 200 μL of the diluted yeast culture are pipetted into each well of a 96-well plate and exposed to different doses of E2βs performed through the addition of 2 μL of E2β stock solutions in DMSO. Each urine sample extract and each E2β stock is assayed. Exposure is performed for 0 h and 24 h. Fluorescence at these time intervals is measured directly using excitation at 485 nm and measuring emission at 530 nm. The densities of the yeast culture at these time intervals are also determined by measuring the OD at 630 nm to check whether a urine sample is toxic for yeast.

In the determination of the decision limit (CCα) and detection capability (CCβ) of the yeast estrogen bioassay, extracts of 20 blank calf urine, 20 spiked calf urine samples (E2β, DES and EE2 at 1 ng/mL, α-zearalanol at 30 or 50 ng/mL, and mestranol at 10 ng/mL), and a reagent blank are analyzed. After 24 h of exposure, the obtained fluorescence signals of the 20 blank urine samples, the 20 spiked urine samples, and a reagent blank are corrected for the signals obtained at 0 h ($t_{24}$–$t_0$). On three different days, the extracts of the blank urines and their corresponding spikes are prepared and analyzed in the yeast estrogen bioassay in three separate

exposures. Within a time period of 10 days, these three sample treatments and exposures are performed. In another experiment, extracts of 20 blank urines and spikes of 50-ng zearalanol per mL urine are prepared in one day and are analyzed in the yeast estrogen bioassay in one exposure. In the context of EC Decision 2002/657, the mean signal of 20 blank calf urine samples plus 2.33 times the corresponding standard deviation is defined as the decision limit CCα ($\alpha = 1\%$), and the decision limit CCα plus 1.64 times the standard deviation of the signal of the spiked calf urine sample is defined as the detection capability CCβ ($\beta = 5\%$).

In the determination of the specificity of the yeast estrogen bioassay, three blank calf urine samples are spiked with a high dose of testosterone or progesterone (1000 ng/mL) and extracts are analyzed. In the determination of interference, these blank calf urine samples are spiked with a high dose of either testosterone or progesterone in combination with a low dose of estrogens (E2β, DES, and EE2 at 1 ng/mL, α-zearalanol at 50 ng/mL, and mestranol at 10 ng/mL).

## 9 ANTIMALARIAL ASSAY

The tremendous progress in the biology and the biochemistry of malaria parasites has led to the identification of drug targets that are both parasite specific and essential for parasite growth and survival, and some of them are now exploited in the mechanisms-based assays of various screening programs and determination of antimalarials' concentration such as plasmodium falciparum and murine P388 leukemia cell assay [118], histidine-rich protein II assay [119], anti-plasmodium activity assay [120], survival of Anopheles gambiae assay [121], chloroquine plus doxycycline assay[122], and plasmodium falciparum clone and dihydroartemisinin assay [123].

### 9.1 Plasmodium Falciparum and Murine P388 Leukemia Cell Assay

The required quantity of *P. falciparum* parasites maintained in a complete medium (RPMI-1640, 25-mM HEPES, 25-mM $NaHCO_3$, and 10% pooled human serum, with uninfected human red blood cells at 2.5% haematocrit) is introduced into flat-bottomed 96-well plates. The cell suspension (1% parasitaemia) is added to the plates (0.2 mL/well), which contains test compound in triplicate alongside untreated controls. The plates are shaken vigorously using a microculture plate shaker, and they are incubated at 37°C for 18 h under microaerophilic conditions [124]. To each well, tritiating hypoxanthine with a specific activity of 14.1 Ci/mmol is added (0.5 μCi/well) and incubated at 37°C for another 24 h. The contents of the well are frozen at –30°C, unfrozen at 50°C, filtrated, and washed several times with water. The disks are dried and added to toluene scintillator in vials, and the radioactivity incorporated into parasites is estimated.

Test compounds inhibiting 75% or more of the parasite growth are systematically submitted to cytotoxicity tests. Murine P388 leukemia cells grown in RPMI 1640 medium containing 0.01-nM β-mercaptoethanol, 10-mM L-glutamine, 100-U/mL G-penicillin, 100-μg/mL streptomycin, 50-μg/mL gentamycin, and 50-μg/mL nystatine, supplemented with 10% fetal calf serum, are incubated in humidified atmospheric air containing 5% $CO_2$ at 37°C. The inoculum seed at $10^4$

cells/mL (0.1 mL/well) is transferred into flat-bottomed 96-well plates containing serial concentrations of test compound, and the plates are incubated in the required atmosphere at 37°C for 72 h. Thereafter, cells are incubated at 37°C with a 0.02% solution of neutral red in 1/9 methanol/water (0.1 mL/well) for 1 h, washed with 1N PBS, and lyzed with 1% SDS. After a brief agitation, the plates are read at 540 nm to measure the absorbance of the extracted dye, and cell viability is expressed as the percentage of cells incorporating dye relative to the untreated controls and $IC_{50}$ values are determined by the linear regression method.

### 9.2 Histidine-Rich Protein II Assay

The serial dilutions of complement-inactivated plasma samples, which are from uncomplicated falciparum malaria patients and a healthy volunteer treated with oral sodium artesunate (100 mg followed by 50 mg orally every 12 h for 5 days), are applied in two columns to 96-well microculture plates at 50 μL/well. The serial dilutions of spiked plasma are applied in two columns and added to each plate as controls. On each 96-well plate, five unknown samples plus the controls are tested; in addition, one plate with six serial dilutions of known drug concentrations covering the whole test range is also tested. According to the literature, the culture and ELISA procedures are performed [125]. A total of 175-μL aliquots of the dilutions (0.1% parasite density, 1.7% hematocrit) of synchronized parasitized (clone $W_2$) blood samples from continuous culture are added to the plates (resulting in a total of 225 μL/well), with the frozen mixture of 1.05 mL of the remaining cell medium and 0.3 mL of drug-free plasma per plate as negative controls. Before being frozen-thawed, the plates are incubated at 37°C for 72 h in a candle jar or gas mixture (5% $CO_2$, 5% $O_2$, 90% $N_2$).

### 9.3 Antimalarial Activity Assay

The chloroquine resistant strain FcB1 of *P. falciparum* (50% inhibitory concentration [$IC_{50}$] of chloroquine = 62 ng/mL) maintained continuously in culture on human erythrocytes is used for the evaluation of antimalarial activities of test compounds [126]. The semiautomated microdilution technique [127] is modified to determine the *in vitro* antiplasmodial activity of test compounds. The serial medium dilutions of the stock solutions of test compounds in DMSO 10 mg/mL, of which the concentration never exceeds 0.1% and does not inhibit the parasite growth, are incubated with asynchronous parasite cultures (0.5% parasitemia and 1% final hematocrit) on 96-well plates at 37°C for 24 h and treated with [$^3$H] hypoxanthine (0.5 μCi/well) for 24 h. By comparison of the radioactivity incorporated into the treated culture with that in the control culture maintained on the same plate, the growth inhibition concentration for each test compound is defined. From the drug concentration-response curve, the $IC_{50}$ is obtained and expressed as the mean determined from three independent experiments.

### 9.4 Survival of *Anopheles Gambiae Assay*

Adults of laboratory-reared *An. gambiae* mosquitoes (from a colony established from wild gravid females, all mosquito life stages are maintained under semi-field conditions) are given 6% glucose solution on filter-paper wicks, water, and routine

human bloodmeals three times per week. Three days after each bloodmeal inside the cages, which are kept in screen houses under ambient conditions, oviposition cups are used to collect eggs on the following day. Under semi-field conditions, the larvae for both the colony and the experiments are maintained on fish food and the plastic pans (25 × 20 × 14 cm) are filled with fresh water [128].

In 30-cm cubic metal frame cages covered with mesh netting, approximately 100 pupae/cage and the test compound are for direct observation. *An. gambiae* mosquitoes are allowed to emerge, and evening observations of test compound feeding are conducted from 20:00 to 22:00 h, for which at intervals of 30 min, a flashlight is used. On a second night of observation, a mosquito net is placed over *Ricinus communis* in a screenhouse, under which a cup of water containing approximately 500 pupae is placed and emerging adults are allowed to feed on the test compound for 24 h. From 20:00 to 22:00 h in intervals of 30 min, on the following night, mosquitoes are observed until they are observed probing or feeding on the test compound.

For survival assay, mosquitoes are divided into regime groups. From the main mosquito colony, approximately 100 pupae per regime are harvested, transferred to plastic cups, and placed in cages and held in a screenhouse under ambient conditions. Each cage of mosquitoes has access to test compound sources; one group of mosquitoes has access to a cotton pad moistened with distilled water, representing negative controls; and another group of mosquitoes with neither test compound source nor water is used as an additional negative control. Dead mosquitoes are removed at 4-h intervals from 7:00 to 23:00 h until all die.

### 9.5 Chloroquine Assay

The blood samples (10 mL) from healthy volunteers who are on chloroquine (2 × 150-mg base weekly) and doxycycline (50 mg or 100 mg daily) malaria prophylaxis are centrifuged (1200 g, 10 min), and the plasma is separated and stored at –20°C. The chloroquine-sensitive FC27 strain of *P. falciparum* is used for the determination of antimalarial activity of each plasma sample [129]. On microculture plates, the plasma samples are diluted twofold with drug-free serum and inoculated with a suspension of parasitized erthrocytes in culture medium. The drug susceptibility and minimum concentration of spiked drug that inhibits parasites, relative to control, from developing to schizonts is recorded. The maximum inhibitory dilution (MID) of the volunteers' samples containing a choroquine concentration and the minimum inhibitory concentration (MIC) observed in the sensitivity test are recorded. The chloroquine equivalent concentration in the prediluted volunteers' specimen is estimated by multiplying the MID by the MIC.

### 9.6 Plasmodium Falciparum Clone and Dihydroartemisinin (DHA) Assay

By centrifugation for 5 s, 50 μL of Affigel protein A (binding capacity of 20 mg of purified human immunoglobulin G[IgG]/mg of gel) is washed twice with PBS (pH 7.2, 50 μL), and the RPMI 1640 medium (50 μL) that is supplemented with 5.94-g/L HEPES and 2.1-g/L sodium bicarbonate, centrifuged to discarde the supernatant, and Affigel protein A gel is incubated with the plasma and serum (250 μL) spiked with dihydroartemisinin (DHA) or plasma from patients at room temperature for

30 min [130]. The stock solution of DHA in 70% ethanol (1 mg/mL) is kept at –10°C for up to 1 month before use, which is further diluted in plasma or serum to give a working concentration (1.25 to 100 ng/mL).

Affigel protein A-treated plasma (100 μL) and heat-inactivated serum (50 μL/well) are transferred to row A and B of a flat-bottom plate, respectively. Using the heat-inactivated serum added through G into row B as the diluent for plasma or serum containing DHA, twofold serial dilutions are prepared. To row H, 50 μL of heat-inactivated control plasma or serum is added for parasitizing and nonparasitizing erythrocyte controls. To all wells for rows A to G and eight wells for row H, the suspension of 175 μL of malaria parasite-infected erythrocytes (W2 clone; 0.5% parasitemia with 80% young rings at a 1.7% hematocrit) is added. To the remaining four wells of row H for nonparasitized controls, the similar suspension of uninfected erythrocytes is added. The microtiter plates are incubated at 37°C for 24 h and treated with 1 Ci/mmol [$^3$H] hypoxanthine by the addition of 25 μL of 0.5 μCi isotope solution to each well. The microtiter plates are incubated for another 18 to 20 h. Harvesting the contents of the plates, the particulate material of the water-lysed cell suspension is collected on glass-fiber filter paper by filtration, and the filter paper is dried in an oven and placed in a plastic bag, to which 10 mL of scintillation fluid is added and the plastic bag is then sealed. By counting in a liquid scintillation counter, the level of incorporation of [$^3$H] hypoxanthine by the malaria parasites is determined.

Fresh whole blood is collected to heparinized tubes, and the plasma and buffy coat are removed, and the erythrocytes are lysed and serially diluted (2- to 64-fold) with sterile water, of which 50-μL aliquots are diluted in 250 μL of plasma or medium spiked with DHA to give a final concentration of 50 ng/mL and a range of hemoglobin concentrations from 1.5 to 0.05 g%. The effect of hemolysate in both plasma and culture medium (RPMI 1640 medium with 15% heat-inactivated human serum) on the detection of DHA activity is examined. The final concentration of plasma and serum in each well is 30% and 15%, respectively, which is similar to that in the normal drug susceptibility test. In addition to testing hemolysates of normal erythrocytes, the plasma samples collected from patients with severe falciparum malaria before treatment with any antimalarial drugs can also be spiked with 50 ng of DHA/mL and subjected to the same bioassay.

## 10 BLOOD PRESSURE-RELATED ASSAY

Blood pressure can be regulated by receptor, enzyme, endogerous substance, and cell, such as bradykinin B2-receptor, ACE, 5-Hydroxytryptamine, and endothelia cell. Based on these facts, the human plasma assay [131], pulmonary hypertension assay [132], coronary arteries (CA) constriction assay [133], MCAO and HSPG assay [134], and temperature assay in awake subjects [135] are established.

### 10.1 Human Plasma Assay

The solution of test peptide is injected (i.v.) into adult male Wistar rats (housed in pairs, under a 12:12 h light/dark cycle, 0.4% NaCl rat chow, and water available *ad libitum*, after 3–5 days of adaptation to the facility, about 300 g) at a proper dose.

The inactin (100 mg/kg, i.p.; sodium ethyl-(1-methyl-propyl)-malonylthio-urea) anesthetized rats are treated with atropine (2.4 mg/kg, s.c.), and their ganglions are blockaded with pentolinium (19.2 mg/kg, s.c.). Overall, 10-mg/kg captopril is administered i.v. acutely to the rats in the captopril group, and it inhibits fully the pressor effect of 60-ng/kg angiotensin I.

To assist ventilation, the trachea is intubated and both vagi are severed to block reflex bradycardia. Into the right femoral vein and carotid artery, catheters are inserted for the injection of test compound and other agonists and antagonists, and for monitoring arterial blood pressure and heart rate and for blood sampling, respectively [136, 137], and blood pressure is monitored.

Responses, in terms of the increases in systolic (SBP) and diastolic (DBP) blood pressures, to test peptide and bradykinin are determined. After ganglion blockade with pentolinium (group GB) without or with added captopril (group GB+Cap), heart rate and plasma adrenaline and noradrenaline concentrations are determined. In some cases, the rats are not subjected to ganglion blockade and served as controls for animals in the GB group (control group).

To evaluate the contribution of bradykinin B2-receptor-mediated mechanisms, the selective bradykinin B2-receptor antagonist, HOE-140, is given to ganglion-blocked, the rats are captopril treated, and the increases in SBP, DBP, heart rate, and plasma adrenaline/noradrenaline induced by test peptide and bradykinin (group GB+Cap+HOE) are observed. To evaluate the contribution of AT1-receptor-mediated mechanisms, the selective AT1 receptor antagonist, losartan, is given to group GB rats before and after treatment with captopril (1 mg/kg, i.v.) and the increases in SBP, DBP, heart rate, and plasma adrenaline/noradrenaline induced by test peptide are observed.

For determining plasma adrenaline and noradrenaline arterially approximately 1 mL of blood is withdrawn via the carotid cannula at baseline (before injection of test peptide and bradykinin), at the peak of the SBP response, and after recovery from the response [138, 139]. Before each withdrawal, the rat is given a transfusion of approximately 1 mL of blood from a similarly prepared donor rat. The SBP peak is observed at 2–3 min after the injection of test peptide and at 1–2 min after the injection of bradykinin; blood sampling coincides with these intervals. Using HPLC and fluorimetric detection, plasma catecholamines are determined [140].

### 10.2 Pulmonary Hypertension Assay

The chest cavity of a mouse (10–20 weeks old, 22–30 g) anesthetized by pentobarbital sodium (80 mg/kg, i.p.) is opened, and the lungs are removed rapidly and placed in Krebs–Ringer bicarbonate solution (KRBS) containing 118.3-mM NaCl, 4.7-mM KCl, 1.2-mM $MgSO_4$, 1.2-mM $KH_2PO_4$, 2.5-mM $CaCl_2$, 25.0-mM $Na_2CO_3$, and 10.0-mM glucose, bubbled with 21% $O_2$. From the intrapulmonary artery (third to fourth generation, PA), rings (50–100 μm internal diameter, 2–3 mm long) are isolated using a dissection microscope. PA rings are mounted as ring preparations by threading two steel wires into the lumen and securing the wires to two supports, and they are placed in a small vessel wire myograph chamber. The support is attached to a micrometer for the control of ring circumference and to a force transducer for measurement of isometric tension, respectively. After removal of endothelial cells by gently rubbing the intraluminal surface with a steel wire,

some vessels are successively perfused with 2-mL air bubbles and 2-mL of KRBS (perfusion pressure < 5 mm Hg), before being mounted in the chamber. In the chamber filled with KRBS (pH 7.35–7.45), bubbled with 21% $O_2$–5% $CO_2$–balance $N_2$, the whole preparation is kept at 37°C. To control oxygen tension over the superfusate, Plexiglas is used as a cover over the chamber. The temperature and PA tension are recorded. At initial tension of 0 mN (1 g = 4.905 mN), isolated murine PA rings in the chamber are allowed to equilibrate for 10–15 min, which are increased to 5 mN in 2.5-mN steps at 4- to 5-min intervals and held constant thereafter. Through preliminary experiments, the resting tension for maximal constrictor response is optimized, in which the resting tension levels of 2.5, 5, and 7.5 mN are compared. In the evaluation of vascular viability, after the treatment of PA with 60-mM KCl, the tension is determined. The PA is washed extensively with KRBS, successively exposed to U-46619 (0.01 μM, thromboxane $A_2$ agonist) and 1-μM Ach, the resulting tension is recorded and stabilized for 5–10 min, and the agonists are washed out of the myograph chamber with KRBS.

Murine PA is isolated, placed in a confocal pressure myograph chamber, cannulated at both ends with glass micropipettes, and secured with a 12-0 nylon monofilament suture. To control transmural pressure, both cannulas are connected to a reservoir that can be raised or loared. In the chamber, PA is superfused at 37°C, constantly with KRBS, and gassed with 21% $O_2$–5% $CO_2$–balance $N_2$. To control oxygen tension over the superfusate, Plexiglas is used as a cover over the chamber. To the chamber, 0.1-mM dihydroethidum (DHE) is added, in which murine PA is incubated at 37°C for 45 min and then washed for 30 min. The reaction of ntracellular DHE and ROS produces a fluorescent oxidized product [141, 142] and PA images are scanned (480-nm-line argon laser, fluorescence 620 nm). Superoxide anion levels in isolated murine PA are measured by a lucigenin-enhanced chemiluminescence technique [143, 144] and scintillation counter, using a solution of 5-μM lucigenin in Krebs–HEPES buffer (10.0-mM HEPES acid, 135.3-mM NaCl, 4.7-mM KCl, 1.2-mM $MgSO_4$, 1.2-mM $KH_2PO_4$, 1.8-mM $CaCl_2$, 0.026-mM Na-EDTA, and 11.1-mM glucose), 1-mL total volume, after background chemiluminescence activity has stabilized for 5 min.

### 10.3 Coronary Arteries (CA) Constriction Assay

The left main coronary arteries (CA, 70–90 μm in diameter, 1-mm length) of a mouse are placed into a microvascular chamber, cannulated at both ends with glass micropipettes, and pressurized. By gently rubbing the intraluminal surface with a steel wire, the endothelial cells of some vessels are removed. The vascular intraluminal pressure (Ptm) is measured by using a pressure transducer positioned at the level of vessel lumen. The vessels in the chamber are superfused constantly with recirculating Krebs–Ringer bicarbonate solution containing 118.3-mM NaCl, 4.7-mM KCl, 1.2-mM $MgSO_4$, 1.2-mM $KH_2PO_4$, 2.5-mM $CaCl_2$, 25.0-mM $NaHCO_3$, and 11.1-mM glucose, which is gassed with 16% $O_2$, 5% $CO_2$, and balance $N_2$ (pH 7.35–7.45), and maintained at 37°C. To control oxygen tension over the superfusate the chamber is covered by Plexiglas, through which port an oxygen electrode is passed into the superfusate and positioned near the vessel to provide continuous measurement of oxygen tension. The vascular intraluminal diameter (ID) is measured continuously. The oxygen tension, vascular ID, and Ptm are recorded. The

isolated CA in the chamber is allowed to equilibrate 30 min at a Ptm of 10 mm Hg, which is increased to 60 mm Hg in 10-mm Hg steps at 5–7-min intervals and held constant thereafter. When Ptm is increased to 60 mm Hg (time 0), the ID is recorded (appointed to ID60, namely ID at Ptm = 60 mm Hg) and the measurement is continued throughout the experiment.

The superoxide anions produced in isolated murine CA are measured using a lucigenin (bis-*N*-methyacidinium nitrate)-enhanced chemiluminescence technique. After background has stabilized chemiluminescence activity for 5 min, CA is placed in the chemiluminometer containing 1 mL of 5-μM lucigenin buffer solution, photon emission is recorded continuously, and the chemiluminescence signal is recorded as relative light units per second (RLU/s).

### 10.4 Middle Cerebral Artery Occlusion (MCAO) Assay and Vascular Heparan Sulfate Proteoglycans (HSPG) Perlecan Assay

Male baboons are randomly divided into treatment and control groups. In the treatment groups the animals undergo MCAO for 1 h or 2 h, or 3 h MCAO with subsequent reperfusion for 1 h or 4 h, or 1.5 h MCAO with 24 h reperfusion. In the control groups, the animals do not undergo any preparation procedure or suffer from MCAO at surgical implantation and has sustained hemiparesis for 7 days. The animals are transcardially perfused with isosmotic heparinized perfusate, and their brain tissues are removed under thiopental Na and prepared for frozen and paraffin sections [145, 146]. The samples of normal and 50 μL of ischemic brain tissue or 100 μL of purified reagents are added to the recipient tissues consisting of unfixed 10-μm-thick normal basal ganglia sections mounted on microscope slides.

When proteases are added to the normal recipient tissues, PBS (pH 7.0) is the incubation buffer for type-7 bacterial collagenase. The solution of matrix metalloproteinases (MMPs) and urokinase/plasminogen in a buffer containing 90-mM NaCl, 5-mM KCl, 1.5-mM $MgCl_2$, 23-mM Na gluconate, 27-mM NaAc, 10-mM $CaCl_2$, and 40-mM $ZnCl_2$ (pH 7.4) is prepared. The solution of cathepsin B and L in 200-mM NaAc containing 8-mM dithiothreitol, 1-mM EDTA, and 2.7-mM L-cysteine (pH 6.0) is prepared. The mixture of 5-U/mL plasminogen and 5-U/mL uPA is incubated at 37°C for 1 h; 1-μg/mL Pro-MMP-2 or 1-μg/mL pro-MMP-9 is added and incubated for 1 h. The recipient sections are incubated with 100 μL of each mixture or buffer alone at 37°C for 5 h (collagenase) or 18 h (MMPs and cathepsins) and washed with PBS and fixed.

From approximately 1.0-cm × 1.0-cm × 10-μm frozen sections, donor samples are derived. By centrifugation (300 g, 30 s) and repeated gentle pipetting, the mixture of 10 consecutive 10-μm sections of ischemic or normal basal ganglia is prepared. Overall, 50 μL of sample or PBS is added to the recipient sections, the mixture is incubated at 37°C for 18 h, and the recipient sections are washed with PBS and fixed in acetone or paraformaldehyde (PFA).

In the development of the specific antigens, the immunoperoxidase methods are used [147]. At 4°C, the overnight incubation of acetone-fixed frozen sections in the presence of the primary antibody is followed by developing immunoperoxidase with 3,3′-diaminobenzidine tetrahydrochloride. At 4°C, the overnight deparaffination of acetone-fixed frozen sections in the presence of the primary antibody is also followed by developing immunoperoxidase with 3,3′-diaminobenzidine

tetrahydrochloride. Incorporating dUTP into nuclear DNA is used as the evidence of nuclear DNA scission/repair, and at 2 h, MCAO defines the ischemic core and peripheral regions of cellular neuronal injury. PFA-fixed cryosections are subject to the DNA polymerase I method [148]. All ischemic samples include ischemic core and ischemic peripheral regions. In the measurement of MMP-related activities, Gelatin zymography is performed with a modification to increase sensitivity [149]. Protease activities are identified by incubation of the gels in buffer containing GM6001 or APMSF.

## 10.5 Temperature Assay in Awake Subjects

The male SD rats (250–325 g) are caused reversible ischemia with the intraluminal filament occlusion of the middle cerebral artery (MCA) and maintained anesthesia with 2% halothane and mixed gas of nitrous oxide and oxygen (60 : 40) by face mask [150]. To expose the left carotid artery, a midline neck incision is made. The external carotid and pterygopalatine arteries are ligated with 5-0 silk. In the arterial wall, an incision is made. From the bifurcation of external and internal carotid arteries, a 4-0 heat blunted nylon suture of 18 mm is advanced. This distance reliably produces blockage of the origin of the MCA. The duration of occlusion for individual rats is varied to generate a range of ischemia durations for each experimental group. During surgery, the temporalis and brain temperature are monitored and maintained.

During the period (4–5 h) of post-anesthetic recovery, the awake rats are kept on a pad maintained at 37.5°C. To verify the effect of this procedure on brain temperature, an indwelling radio telemetry thermister is implanted into the rat of the nonischemic group. After induction of anesthesia, the rat is placed in a stereotaxic head frame. A thermister probe is placed in the left parietal cortex (3 mm deep to dura and 1 mm anterior, 3 mm lateral to bregma). To secure the probe to the skull, a plastic cap is fitted using methylmethacrylate [151]. After full recovery from the anesthesia, telemetered brain temperature is measured every 15 min. Temperatures are measured every 5 min for 24 h during a baseline period and 6 h after drug administration. After a drug treatment, some animals are placed on a heating pad for 2 h and their temperature is measured every 15 min. At 48 h and 72 h, the behavior of the rats is evaluated and classified to normal/abnormal according to a modified rodent examination scale [152]. All obtundation/reduced exploration, forepaw retraction on tail lifting, asymmetric forepaw grasp, axial twist, forced circling, or death are defined as abnormal.

To correctly position the filament, 0.2 mL of 2% Evan's Blue in saline is injected into each rat via tail vein and circulated for 30 to 60 min, after which the rat is perfused transcardially with 100-mL normal saline, and the brain is removed and immediately examined. Each day, two rats of the control group and four or five rats of the drug treatment group are injected test compound dissolved in saline via tail vein at 30, 60, 75, 120, 240, or 360 min after initiating occlusion to assess therapeutic efficacy.

To determine the effects of the test compound on blood gases, heart rate, and arterial blood pressure in separate groups, physiological determinations are performed. To observe the interaction between test compound and anesthesia, the determinations in awake or in anesthetized rats are performed. The ventral tail

artery is cannulated with PE50 tubing for recording the arterial pressures and sampling arterial blood. Test compound is administered after the rat is stable for 30 min with no further adjustments needed to maintain homeostasis under 75–90 mmHg steady-state mean arterial pressure (MAP) and minimal halothane. To obtain similar data from awake subject, tail artery catheters are implanted under halothane inhaled anesthesia. After implantation, the rat is allowed to recover and is lightly restrained. One hour after stopping the anesthesia, the cannula is attached to the transducer and vital signs are recorded for a 30-min baseline period. After the subject is shown to be stable for 30 min, the test compound is given. Blood pressures and blood gases are recorded every 10 min for 1 h and then hourly.

## REFERENCES


1. Engler K H, Coker R, Evans I H (1999). A novel colorimetric yeast bioassay for detecting trichothecene mycotoxins. *J. Microbiol. Meth.* 35:207–218.
2. Gonzaulez A M, Beniutez L, Soto T, et al. (1997). A rapid bioassay to detect mycotoxins using a melanin precursor overproducer mutant of the ciliate *Tetrahymena thermophila*. *Cell Biol. Int.* 21:213–216.
3. Cetin Y, Bullerman L B (2005). Cytotoxicity of *Fusarium* mycotoxins to mammalian cell cultures as determined by the MTT bioassay. *Food Chem. Toxicol.* 43:755–764.
4. Saadé F E, Dunphy G B, Bernier R L (1996). Response of the carrot weevil, *Listronotus oregonensis* (Coleoptera: Curculionidae), to strains of *Bacillus thuringiensis*. *Biol. Control.* 7:293–298.
5. Gehringer M M, Vijayne K, Nadya C (2003). The use of *Lepidium sativum* in a plant bioassay system for the detection of microcystin-LR. *Toxicon.* 41:871–876.
6. Kaney A R, Speare V J (1983). An amicronuleate mutant of *Tetrahymena thermophila*. *Exp. Cell Res.* 143:461–467.
7. Hansen M B, Nielsen S E, Berg K (1989). Re-examination and further development of precise and rapid dye method for measuring cell growth/cell kill. *J. Immunol. Methods.* 119:203–210.
8. Saadé F E (1993). Evaluation of strains of *Bacillus thuringiensis* as biological control agents of the adult stages of the carrot weevil, *Listronotus oregonensis* (Coleoptera; Curculionidae). M.Sc. Thesis. National Library, Ottawa, Canada.
9. Gringorten J L, Witt D P, Milne R E, et al. (1990). An *in vitro* system for testing *Bacillus thuringiensis* toxins: The lawn assay. *J. Invertebr. Pathol.* 56:237–242.
10. Bradford M (1976). A rapid and sensitive method for the quantification of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* 72:248–254.
11. van Frankenhuyzen K, Gringorten J L (1991). Frass failure and pupation failure as quantal measurements of *Bacillus thuringiensis* toxicity to Lepidoptera. *J. Invertebr. Pathol.* 58:465–467.
12. Gehringer M M, Downs K S, Downing T G, et al. (2003). An investigation into the effect of selenium supplementation on microcystin hepatotoxicity. *Toxicon.* 41:451–458.
13. Ward C J, Beattie K A, Lee E Y C, et al. (1997). Colorimetric protein phosphatase inhibition assay of laboratory strains and natural blooms of cyanobacteria: comparisons with high performance liquid chromatographic analysis for microcystins. *FEMS Microbiol. Lett.* 153:465–473.

14. Habig W H, Pabst M J, Jakoby W B (1974). Glutathione Stransferases: the first step in mercapturic acid formation. *J. Biol. Chem.* 249:7130–7139.
15. Baumstark-Khan C, Rode A, Rettberg P, et al. (2001). Application of the *Lux-Fluoro* test as bioassay for combined genotoxicity and cytotoxicity measurements by means of recombinant *Salmonella typhimurium* TA1535 cells. *Anal. Chim. Acta.* 437:23–30.
16. Shimizu K, Endo O, Goto S, et al. (2001). Bioassay-based evaluation of toxicity of suspended particulate matter in humans: integrated uses of alveolar cells (A549) in air liquid interface culture and hepatocarcinoma cells (Hep G2). *Biochem. Eng. J.* 22:1–9.
17. Takahashi K, Ohsawa M, Utsumib H (2002). A simple bioassay for evaluating immunotoxic properties of chemicals by use of *in vitro* antibody production system. *J. Health Sci.* 48:161–167.
18. Nagy S R, Sanborn J R, Hammock B D (2002). Development of a green fluorescent protein-based cell bioassay for the rapid and inexpensive detection and characterization of Ah receptor agonists. *Toxicol. Sci.* 65:200–210.
19. Bolon B, Bucci T J, Warbritton A R, et al. (1997). Differential follicle counts as a screen for chemically induced ovarian toxicity in mice: results from continuous breeding bioassays. *Fundam. Appl. Toxicol.* 39:1–10.
20. Yoshikazu I, Toshimasa J, Kazuhisa F (2003). Effect of nefiracetam, a neurotransmission enhancer, on primary uroepithelial cells of the canine urinary bladder. *Toxicol. Sci.* 72:164–170.
21. Fujita K, Nagaoka M, Komatsu Y, et al. (2003). Yeast pheromone signaling pathway as a bioassay to assess the effect of chemicals on mammalian peptide hormones. *Ecotox. Environ. Safe.* 56:358–366.
22. Taylor C J, Bain L A, Richardson D J, et al. (2004). Construction of a whole-cell gene reporter for the fluorescent bioassay of nitrate. *Anal. Biochem.* 328:60–66.
23. Sambrook J, Fritsch E F, Maniatis T (1989). *Molecular Cloning, A Laboratory Manual.* Cold Spring Harbor Laboratory Press, New York.
24. Koyano M, Endo O, Goto S, et al. (1998). Carcinogenic polynuclear aromatic hydrocarbons in the atmosphere in Chiang Mai, Thailand. *Jpn. J. Toxicol. Environ. Health.* 44.
25. Willett K L, Gradinali R P, Sericano J L, et al. (1996). Characterization of the H4IIE rat hepatoma cell bioassay for evaluation of environmental samples containing polynuclear aromatic hydrocarbons (PAHs). *Arch. Environ. Contam. Toxicol.* 32:442–448.
26. Ohsawa M, Masuko-Sato K, Takahashi K, et al. (1986). Strain differences in cadmium-mediated suppression of lymphocyte proliferation in mice. *Toxicol. Appl. Pharmacol.* 84:379–388.
27. Mosmann T (1983). Rapid colorimetric assay for cellular growth and survival: Application to proliferation and cytotoxicity assays. *J. Immunol. Meth.* 65:55–63.
28. Garrison P M, Tullis K, Aarts J M, et al. (1996). Species-specific recombinant cell lines as bioassay systems for the detection of 2,3,7,8-tetrachlorodibenzo-*p*-dioxin-like chemicals. *Fundam. Appl. Toxicol.* 30:194–203.
29. Denison M S, Phelan D, Winter G M, et al. (1998). Carbaryl, a Carbamate Insecticide, is a ligand for the Hepatic Ah (Dioxin) receptor. *Toxicol. Appl. Pharmacol.* 152: 406–414.
30. Ziccardi M H, Gardner I A, Denison M S (2000). Development and modification of a recombinant cell bioassay to directly detect halogenated and polycyclic aromatic hydrocarbons in serum. *Toxicol. Sci.* 54:183–193.

31. Pedersen T, Peters H (1968). Proposal for a classification of oocytes and follicles in the mouse ovary. *J. Reprod. Fertil.* 17:555–557.
32. Lewis S A, Hanrahan J W (1990). Physiological approaches for studying mammalian urinary bladder epithelium. In S. Fleischer and B. Fleischer, (eds.), *Methods in Enzymology* 192:632–650. Academic Press.
33. De Virgilio C, Burckert N, Bell W, et al. (1993). Disruption of TPS2, the gene encoding the 100-kDa subunit of the trehalose-6-phosphate synthase/phosphatase complex in *Saccharomyces cerevisiae*, causes accumulation of trehalose-6-phosphate and loss of trehalose-6-phosphate phosphatase activity. *Eur. J. Biochem.* 212:315–323.
34. Miller J H (1972). *Experiments in Molecular Genetics.* Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, pp. 352–355.
35. Miller W G, Lindow S E (1997). An improved GFP cloning cassette designed for prokaryotic transcriptional fusions. *Gene.* 191:149–153.
36. Sodergren E J, Hsu P Y, DeMoss J A (1998). Roles of the narJ and narI gene products in the expression of nitrate reductase in *Escherichia coli*. *J. Biol. Chem.* 263:16156–16162.
37. Sambrook J, Fritsch E F, Maniatis T (1989). *Molecular Cloning: A Laboratory Manual, 2nd* ed.: Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
38. Li S F, DeMoss J A (1987). Promoter region of the nar operon of *Escherichia coli*: Nucleotide sequence and transcription initiation signals. *J. Bacteriol.* 169:4614–4620.
39. Khamis S, Bibby M C, Brown J E, et al. (2004). Phytochemistry and preliminary biological evaluation of *Cyathostemma argenteum*, a malaysian plant used traditionally for the treatment of breast cancer. *Phytother. Res.* 18:507–510.
40. Zhang J X, Labaree D C, Mor G, et al. (2004). Estrogen to antiestrogen with a single methylene group resulting in an unusual steroidal selective estrogen receptor modulator. *J. Clin. Endocrinol. Metab.* 89:3527–3535.
41. Kasorndorkbua C, Halbur P G, Thomas P J, et al. (2002). Use of a swine bioassay and a RT-PCR assay to assess the risk of transmission of swine hepatitis E virus in pigs. *J. Virol. Meth.* 101:71–78.
42. Wonganuchitmeta S N, Yuenyongsawad S, Keawpradub N, et al. (2004). Antitubercular sesterterpenes from the Thai sponge *Brachiaster* sp. *J. Nat. Prod.* 67:1767–1770.
43. Feng X, Gao Z, Li S, et al. (2004). DNA polymerase β lyase inhibitors from *Maytenus putterlickoides*. *J. Nat. Prod.* 67:1744–1747.
44. Ahn M J, Yoon K D, Min S Y, et al. (2004). Inhibition of HIV-1 reverse transcriptase and protease by phlorotannins from the brown alga *Ecklonia cava*. *Biol. Pharm. Bull.* 27:544–547.
45. Owusu M K, Kamuhabwa A R, Nshimo C (2004). Investigation of fractions present in the stem bark of *Annickia kummeriae* on their P-glycoprotein inhibitory pump activity. *Phytother. Res.* 18:652–657.
46. Sadick M D, Intintoli A, Quarmby V, et al. (1999). Kinase receptor activation (KIRA): A rapid and accurate alternative to end-point bioassays. *J. Pharm. Biomed. Anal.* 19:883–891.
47. Solis P N, Wright C W, Anderson M M, et al. (1993). A microwell cytotoxicity assay using *Artemia salina* (brine shrimp). *Planta Med.* 59:250–252.
48. Hambly R J, Double J A, Thompson M J, et al. (1997). Establishment and characterisation of new cell lines from human breast tumours initially established as tumour xenografts in NMRI nude mice. *Breast Cancer Res. Tr.* 43:247–258.
49. Mosmann T (1983). Rapid colorimetric assay for cellular growth and survival: Application to proliferation and cytotoxicity assays. *J. Immunol. Meth.* 65:55–63.

50. Littlefield B A, Gurpide E, Markiewicz L, et al. (1990). A simple and sensitive microtiter plate estrogen bioassay based on stimulation of alkaline phosphatase in Ishikawa cells: Estrogenic action of $\Delta^5$ adrenal steroids. *Endocrinol.* 127:2757–2762.
51. Kuiper G G, Lemmen J G, Carlsson B, et al. (1998). Interaction of estrogenic chemicals and phytoestrogens with estrogen receptor. *Endocrinol.* 139:4252–4263.
52. Mor G, Sapi E, Abrahams V M, et al. (2003). Interaction of the estrogen receptors with the Fas ligand promoter in human monocytes. *J. Immunol.* 170:114–122.
53. Niwa H, Yamamura K, Miyazaki J (1991). Efficient selection for high-expression transfectants with a novel eukaryotic vector. *Gene.* 108:193–199.
54. Emmens C W (1962). Estrogens. In RI, Dorfman (ed.), *Methods in Hormone Research*. Academic Press, New York, pp. 59–111.
55. Baron R, Vignery A, Neff L, et al. (1983). Processing of undecalcified bone specimens for bone histomorphometry. In RR, Recker (ed.), *Bone Histomorphometry: Techniques and Interpretation*, CRC Press. Boca Raton, FL.
56. Parfitt A. M, Drezner M. K, Glorieux F. H, et al. (1987). Bone histomorphometry: standardization of nomenclature, symbols, and units. Report of the ASBMR Histomorphometry Nomenclature Committee. *J. Bone Miner. Res.* 2:595–610.
57. Halbur P G, Kasorndorkbua C, Gilbert C, et al. (2001). Comparative pathogenesis of infection of pigs with hepatitis E virus recovered from a pig and human. *J. Clin. Microbiol.* 39:918–923.
58. Meng X J, Halbur P G, Haynes J S, et al. (1998). Experimental infection of pigs with the newly identified swine hepatitis E virus (swine HEV), but not with human strains of HEV. *Arch. Virol.* 143:1405–1415.
59. Meng X J, Halbur P G, Shapiro M S, et al. (1999). Genetic and experimental evidence for crossspecies infection by swine hepatitis E virus. *J. Virol.* 72:9714–9721.
60. Sadick M D, Sliwkowski M X, Nuijens A, et al. (1996). Analysis of heregulin-Induced ErbB2 phosphorylation with a high-throughput kinase receptor activation enzyme-linked immunosorbant assay. *Anal. Biochem.* 235:207–214.
61. Hagiwara A, Imai N, Doi Y, et al. (2003). Absence of liver tumor promoting effects of annatto extract (norbixin), a natural carotenoid food color, in a medium-term liver carcinogenesis bioassay using male F344 rats. *Cancer Lett.* 199:9–17.
62. Daniel L, Gustafson A L, Coulson L F, et al. (1998). Use of a medium-term liver focus bioassay to assess the hepatocarcinogenicity of 1,2,4,5-tetrachlorobenzene and 1,4-dichlorobenzene. *Cancer Lett.* 129:39–44.
63. Anand K K, Singh B, Saxena A K, et al. (1997). 3,4,5-trihydroxy benzoic acid (gallic acid), the hepatoprotective principle in the fruits of *Terminalia belerica*-bioassay guided activity. *Pharmacol. Res.* 36:315–321.
64. Wang H, Wei W, Shen Y X, et al. (2004). Protective effect of melatonin against liver injury in mice induced by *Bacillus Calmette-Guerin* plus lipopolysaccharide. *World J. Gastroentero.* 10:2690–2696.
65. Leakey J E A, Seng J E, Allaben W T (2003). Body weight considerations in the $B6C3F_1$ mouse and the use of dietary control to standardize background tumor incidence in chronic bioassays. *Toxicol. Appl. Pharmacol.* 193:237–265.
66. Wang H H, Lautt W W (1997). Hepatocyte primary culture bioassay: a simple tool to assess the initiation of the liver regeneration cascade. *J. Pharmacol. Toxicol. Meth.* 38:141–150.
67. Borset M, Waage A, Sundan A (1996). Hepatocyte growth factor reverses the TDF-β-induced growth inhibition of CCL-64 cells: a novel bioassay for HGF and implication for the TDF-β bioassay. *J. Immunol. Meth.* 189:59–64.

68. Ito N, Tsuda H, Tatematsu M, et al. (1988). Enhancing effect of various hepatocarcinogens on induction of preneoplastic glutathione S-transferase placental form positive foci in rats—an approach for a new medium-term bioassay system. *Carcinogen.* 9:387–394.
69. Hasegawa R, Tsuda H, Shirai T, et al. (1986). Effect of timing of partial hepatectomy on the induction of preneoplastic liver foci in rats given hepatocarcinogens. *Cancer Lett.* 32:15–23.
70. Hasegawa R, Ito N (1992). Liver medium-term bioassay in rats for screening of carcinogens and modifying factors in hepatocarcinogenesis. *Food Chem. Toxicol.* 30:979–992.
71. Ward J M, Hagiwara A, Anderson L M, et al. (1988). The chronic hepatic and renal toxicity of di(2-ethylhexyl)phthalate, acetaminophen, sodium barbital, and phenobarbital in male B6C3F1 mice: Autoradiographic, immunohistochemical, and biochemical evidence for levels of DNA synthesis not associated with carcinogenesis or tumor promotion. *Toxicol. Appl. Pharmacol.* 96:494–506.
72. Gavino V C, Miller J S, Ikharebha S O, et al. (1981). Effect of polyunsaturated fatty acids and antioxidants on lipid peroxidation in tissue cultures. *J. Lipid Res.* 22:763–769.
73. Reed K L, Fruin A B, Gower A C, et al. (2004). A neurokinin 1 receptor antagonist decreases postoperative peritoneal adhesion formation and increases peritoneal fibrinolytic activity. *P. Natl. Acad. Sci. USA.* 101:9115–9120.
74. Liu C P, Tsai W J, Lin Y L, et al. (2004). The extracts from *Nelumbo nucifera* suppress cell cycle progression, cytokine genes expression, and cell proliferation in human peripheral blood mononuclear cells. *Life Sci.* 75:699–716.
75. Li R W, Leach D N, Myers S P, et al. (2004). A new anti-inflammatory glucoside from *Ficus racemosa* L. *Planta Med.* 70:421–426.
76. Yoganathan K, Rossant C, Glover R P, et al. (2004). Inhibition of the human chemokine receptor CCR5 by variecolin and variecolol and isolation of four new variecolin analogues, emericolins A–D, from *Emericella aurantiobrunnea*. *J. Nat. Prod.* 67:1681–1684.
77. Valotis A, Neukam K, Elert O, et al. (2004). Human receptor kinetics, tissue binding affinity, and stability of mometasone furoate. *J. Pharm. Sci.* 93:1337–1350.
78. West M, Mhatre M, Ceballos A, et al. (2004). The arachidonic acid 5-lipoxygenase inhibitor nordihydroguaiaretic acid inhibits tumor necrosis factor α activation of microglia and extends survival of G93A-SOD1 transgenic mice. *J. Neurochem.* 91:133–143.
79. Terra J H, Montaño I, Schilb A, et al. (2004). Development of a microplate bioassay for monocyte chemoattractant protein-1 based on activation of p44/42 mitogen-activated protein kinase. *Anal. Biochem.* 327:119–125.
80. Ashwood P, Harvey R, Verjee T, et al. (2004). Competition between IL-1, IL-1ra and TGF-β1 modulates the response of the ELA4.NOB-1/CTLL bioassay: Implications for clinical investigations. *Inflamm. Res.* 53:60–65.
81. Vermeer H, Hendriks-Stegemana B I, van Suylekoma D, et al. (2004). An *in vitro* bioassay to determine individual sensitivity to glucocorticoids: induction of FKBP51 mRNA in peripheral blood mononuclear cells. *Mol. Cell. Endocrinol.* 218:49–55.
82. Bi L, Zhang Y, Zhao M, et al. (2005). Novel synthesis and anti-inflammatory activities of 2,5-disubstituted-dioxacycloalkanes. *Bioorg. Med. Chem.* 13:5640–5646.
83. Kuo Y C, Huang Y L, Chen C C, et al. (2002). Cell cycle progression and cytokine genes expression of human blood mononuclear cells modulated by *Agaricus blazei*. *J. Lab. Clin. Med.* 140:176–187.

84. Kuo Y C, Meng H C, Tsai W J (2001). Regulation of cell proliferation, inflammatory cytokines production and calcium mobilization in primary human T lymphocytes by emodin from *Polygonum hypoleucum Ohwi*. *Inflamm. Res.* 50:73–82.
85. Kuo Y C, Yang N S, Chou C J, et al. (2000). Regulation of cell proliferation, gene expression, production of cytokines, and cell cycle progression in primary human T lymphocytes by piperlactam S isolated from *Piper kadsura*. *Mol. Pharmacol.* 58:1057–1066.
86. Lowry O H, Rosebrough N J, Farr A L, et al. (1951). Protein measurement with the Folin phenol reagent. *J. Biol. Chem.* 193:265.
87. Högger P, Rohdewald P (1994). Binding kinetics of fluticasone propionate to the human glucocorticoid receptor. *Steroids.* 59:597–602.
88. Marzinzig M, Nussler A K, Stadler J, et al. (1997). Improved methods to measure end products of nitric oxide in biological fluids: nitrite, nitrate and S-nitrosothiols. *Nitric Oxide Biol. Chem.* 1:177–189.
89. Guégan C, Przedborski S (2003). Programmed cell death in amyotrophic lateral sclerosis. *J. Clin. Invest.* 111:153–161.
90. Zemlan F P (1999). Antibodies to CSF Tau: Characterization of antibodies recognizing cerebrospinal fluid structural proteins. *J. Neurochem.* 73:437–438.
91. Zemlan F P, Jauch E C (1999). Serum and CSF biomarker of neuronal damage in traumatic brain injury. *J. Neurotrauma.* 16:690.
92. Zemlan F P, Rosenberg W S, Luebbe P A, et al. (1999). Quantification of axonal damage in traumatic brain injury: affinity purification and characterization of cerebrospinal fluid tau proteins. *J. Neurochem.* 72:741–750.
93. Binkert C, Landwehr J, Mary J L, et al. (1989). Cloning, sequence analysis and expression of a cDNA encoding a novel insulin-like growth factor binding protein (IGFBP-2). *EMBO J.* 8:2497–2502.
94. Blackhurst G, McElroy P K, et al. (2001). Seasonal variation in glucocorticoid receptor binding characteristics in human mononuclear leukocytes. *Clin. Endocrinol.* (Oxf.) 55:683–686.
95. Vermeer H, Hendriks-Stegeman B I, van der Burg B, et al. (2003). Glucocorticoid-induced increase in lymphocytic FKBP51 mRNA expression: A potential marker for glucocorticoid sensitivity, potency and bioavailability. *J. Clin. Endocrinol. Metab.* 88:277–284.
96. Kluft C (1979). Studies on the fibrinolytic system in human plasma: quantitative determination of plasminogen activators and proactivators. *Thromb. Haemost.* 41:365–383.
97. Kurz K D, Main B W, Sandusky G E (1990). Rat model of arterial thrombosis induced by ferric chloride. *Thromb. Res.* 60:269–280.
98. Hladovec J (1974). Experimental arterial thrombosis in rats with continuous registration. *Thromb. Diath. Haemorrh.* 26:407–410.
99. Umetsu T, Sanai K (1978). Effect of 1-methyl-2-mercapto-5-(3-pyridyl)-imidazol(KC-6141), an antiaggregation compound, on experimental thrombosis in rats. *Thromb. Haemost.* 39:74–83.
100. Urano T, Sakakibara K, Rydzewski A (1990). Relationships between euglobulin clot lysis time and the plasma levels of tissue plasminogen activator and plasminogen activator inhibitor 1. *Thromb. Haemost.* 63:82–86.
101. Wardlaw J M, Warlow C P, Counsell C (1997). Systematic review of evidence on thrombolytic therapy for acute ischaemic stroke. *Lancet.* 350:607–614.

102. Kluft C (1979). Studies on the fibrinolytic system in human plasma: Quantitative determination of plasminogen activators and proactivators. *Thromb. Haemost.* 41:365–383.
103. Zhao M, Lin N, Wang C, et al. (2003). Synthesis and thrombolytic activity of fibrinogen fragment related cyclopeptides. *Bioorg. Med. Chem. Lett.* 9:961–964.
104. Zhao M, Liu J, Wang C, et al. (2005). Synthesis and biological activity of nitronyl nitroxide containing peptides. *J. Med. Chem.* 48:4285–4292.
105. Saito S, Tanaka M, Matsunaga K, et al. (2004). The combination of rat mast cell and rabbit aortic smooth muscle is the simple bioassay for the screening of antiallergic ingreditent from methanolic extract of corydalis tuber. *Biol. Pharm. Bull.* 27:1270–1274.
106. Shankar Gopi, Fourrier M S, Grevenkamp M A, et al. (2004). Validation of the COSTIM bioassay for dendritic cell potency. *J. Pharm. Biomed. Anal.* 36:285–294.
107. Spinardi-Barbisan A L T, Kaneno R, Barbisan L F, et al. (2004). Chemically induced immunotoxicity in a medium-term multiorgan bioassay for carcinogenesis with Wistar rats. *Toxicol. Appl. Pharmacol.* 194:132–140.
108. Taniguchi M, Nagaoka K, Ushio S, et al. (1998). Establishment of the cells useful for murine interleukin-18 bioassay by introducing murine interleukin-18 receptor cDNA into human myelomonocytic KG-1 cells. *J. Immunol. Methods.* 217:97–102.
109. Wei X, Swanson S J, Gupta S (2004). Development and validation of a cell-based bioassay for the detection of neutralizing antibodies against recombinant human erythropoietin in clinical studies. *J. Immunol. Methods.* 293:115–126.
110. Browning D D, The T, O'Day D H (1995). Comparative analysis of chemotaxis in dictyostelium using a radial bioassay method: Protein tyrosine kinase activity is required for chemotaxix to folate but not to camp. *Cell. Signal.* 7:481–489.
111. Kuper C F, Harlerman J H, Richter-Reichelm H B, et al. (2000). Histopathologic approaches to detect changes indicative of immunotoxicity. *Toxicol. Pathol.* 28:454–466.
112. Parnet P, Garka K E, Bonner T P, et al. (1996). IL-1 Rrp is a novel receptor-like molecule similar to the type I interleukin-1 receptor and its homologues T1 r ST2 and IL-1R Acp. *J. Biol. Chem.* 271:3967.
113. Mason S, La S, Mytych D, et al. (2003). Validation of the BIACORE 3000 platform for detection of antibodies against erythropoietic agents in human serum samples. *Curr. Med. Res. Opin.* 19:651.
114. Rehmann K, Schramm K W, Kettrup A A (1999). Applicability of a yeast oestrogen screen for the detection of oestrogen-linke activity in environmental samples. *Chemosphere.* 38:3303–3312.
115. Paris F, Servant N, Térouanne B, et al. (2002). A new recombinant cell bioassay for ultrasensitive determination of serum estrogenic bioactivity in children. *J. Clin. Endocr. Metab.* 87:791–797.
116. Bovee T F H, Helsdingen R J R, Rietjens I M C M, et al. (2004). Rapid yeast estrogen bioassays stably expressing human estrogen receptors α and β, and green fluorescent protein: a comparison of different compounds with both receptor types. *J. Steroid Biochem. Mol. Biol.* 91:99–109.
117. Bovee T F H, Heskamp H H, Hamers A R M, et al. (2005). Validation of a rapid yeast estrogen bioassay, based on the expression of green fluorescent protein, for the screening of estrogenic activity in calf urine. *Anal. Chim. Acta.* 529:57–64.
118. Rasoanaivo P, Ramanitrahasimbola D, Rafatro H (2004). Screening extracts of Madagascan plants in search of antiplasmodial compounds. *Phytother Res.* 18:742–747.

119. Noedl H, Teja-Isavadharm P, Miller R S (2004). Nonisotopic, semiautomated plasmodium falciparum bioassay for measurement of antimalarial drug levels in serum or plasma. *Antimicrob. Agents Chemother.* 48:4485–4487.
120. Krief S, Martin M T, Grellier P, et al. (2004). Novel antimalarial compounds isolated in a survey of self-medicative behavior of wild chimpanzees in Uganda. *Antimicrob. Agents Chemother.* 48:3196–3199.
121. Impoinvil D E, Kongere J O, Foster W A, et al. (2004). Feeding and survival of the malaria vector *Anopheles gambiae* on plants growing in Kenya. *Med. Vet. Entomol.* 18:108–115.
122. Kotecka B, Edstein M D, Rieckmann K H (1996). Chloroquine bioassay of plasma specimens obtained from soldiers on cholo quine plus doxycycline for malaria priphylaxis. *Int. J. Parasitol.* 26:1325–1329.
123. Isavadharm T, Peggins J O, Brewer T G, et al. (2004). Plasmodium falciparum-based bioassay for measurement of artemisinin derivatives in plasma or serum paktiya. *Antimicrob Agents Chemother.* 48:954–960.
124. Trager W, Jensen J B (1977). Plasmodium falciparum in culture: use of outdated erythrocytes and description of the candle jar method. *J. Parasitol.* 63:883–887.
125. Noedl H, Wernsdorfer W H, Miller R S, et al. (2002). Histidine-rich protein II, a novel approach to antimalarial drug susceptibility testing. Antimicrob. *Agents Chemother.* 46:1658–1664.
126. Trager W, Jensen J B (1976). Human malarial parasites in continuous culture. *Science.* 193:673–675.
127. Desjardin R E, Canfield C J, Haynes J D, et al. (1979). Quantitative assessment of antimalarial activity in vitro by a semiautomated microdilution technique. *Antimicrob. Agents Chemother.* 16:710–718.
128. Knols B G J, Njiru B N, Mathenge E M, et al. (2002). MalariaSphere: a greenhouse-enclosed simulation of a natural *Anopheles gambiae* (Diptera: Culicidae) ecosystemin western Kenya. *Malaria J.* 1:19.
129. Kotecka B, Rieckmann K H (1993). Chloroquine bioassay using malaria microcultures. *Am. J. Trop. Med. Hyg.* 49:460–464.
130. Webster HK, Boundreau E F, Pavanand K, et al. (1985). Antimalarial drug susceptiblity testing of *Plasmodium falciparum* in Thailand using a microdilution radio-isotope method. *Am J. Trop. Med. Hyg.* 34:228–235.
131. Amfilochiadis A A, Papageorgiou P C, Kogan N, et al. (2004). Role of bradykinin B2-receptor in the sympathoadrenal effects of "new pressor protein" related to human blood coagulation factor XII fragment. *J. Hypertens.* 22:1173–1181.
132. Liu J Q, Folz R J (2004). Extracellular superoxide enhances 5-HT-induced murine pulmonary artery vasoconstriction. *Am. J. Physiol. Lung. Cell. Mol. Physiol.* 287: L111–118.
133. Liu J Q, Zelko I N, Folz R J (2004). Reoxygenation-induced constriction in murine coronary arteries: the role of endothelial NADPH oxidase (gp91phox) and intracellular superoxide. *J. Biol. Chem.* 279:24493–24497.
134. Fukuda S, Fini C A, Mabuchi T, et al. (2004). Focal cerebral ischemia induces active proteases that degrade microvascular matrix. *Stroke.* 35:998–1004.
135. Lyden P D, Jackson-Friedman C, Shin C, et al. (2000). Synergistic combinatorial stroke therapy: A quantal bioassay of a GABA agonist and a glutamate antagonist. *Exp. Neurol.* 163:477–489.
136. Mavrogiannis L, Trambakoulos D M, Boomsma F, et al. (2002). The sympathoadrenal system mediates the blood pressure and cardiac effects of human coagulation factor XII-related "new pressor protein." *Can. J. Cardiol.* 18:1077–1086.

137. Mavrogiannis L, Kariyaisam K P, Osmond D H (1997). Potent blood pressure raising effects of activated coagulation factor XII. *Can. J. Physiol. Pharmacol.* 75:1398–1403.

138. Papageorgiou P C, Pourdjabbar A, Amfilochiadis A A, et al. (2004). Are the cardiovascular and sympathoadrenal effects of human "new pressor protein" preparations attributable to human coagulation β-FXIIa? *Am. J. Physiol. Heart. Circ. Physiol.* 286: H837–H846.

139. Simos D, Boomsma F, Osmond D H (2002). Human coagulation factor XII-related "new pressor protein": role of PACAP in its cardiovascular and sympathoadrenal effects. *Can. J. Cardiol.* 18:1093–1103.

140. van der Hoorn F A, Boomsma F, Man in 't Veld AJ, et al. (1989). Determination of catecholamines in human plasma by high-performance liquid chromatography: Comparison between a new method with fluorescence detection and an established method with electrochemical detection. *J. Chromatogr.* 487:17–28.

141. Athar M, Elmets C A, Bickers D R, et al. (1989). A novel mechanism for the generation of superoxide anions in hematoporphyrin derivativemediated cutaneous photosensitization. Activation of the xanthine oxidase pathway. *J. Clin. Invest.* 83:1137–1143.

142. Becker L B, vanden Hoek T L, Shao Z H, et al. (1999). Generation of superoxide in cardiomyocytes during ischemia before reperfusion. *Am. J. Physiol. Heart Circ. Physiol.* 277:H2240-H2246.

143. Brandes R P, Barton M, Philippens K M, et al. (1997). Endothelial-derived superoxide anions in pig coronary arteries: Evidence from lucigenin chemiluminescence and histochemical techniques. *J. Physiol.* 500:331–342.

144. Guzik T J, Mussa S, Gastaldi D, et al. (2002). Mechanisms of increased vascular superoxide production in human diabetes mellitus: Role of NAD(P)H oxidase and endothelial nitric oxide synthase. *Circulation.* 105:1656–1662.

145. Hamann G F, Okada Y, Fitridge R, et al. (1995). Microvascular basal lamina antigens disappear during cerebral ischemia and reperfusion. *Stroke.* 26:2120–2126.

146. del Zoppo G J, Copeland B R, Harker L A, et al. (1986). Experimental acute thrombotic stroke in baboons. *Stroke.* 17:1254–1265.

147. Tagaya M, Haring H P, Stuiver I, et al. (2001). Rapid loss of microvascular integrin expression during focal brain ischemia reflects neuron injury. *J. Cereb. Blood Flow. Metab.* 21:835–846.

148. Tagaya M, Liu K F, Copeland B, et al. (1997). DNA scission after focal brain ischemia: temporal differences in two species. *Stroke.* 28:1245–1254.

149. Heo J H, Lucero J, Abumiya T, et al. (1999). Matrix metalloproteinases increase very early during experimental focal cerebral ischemia. *J. Cereb. Blood Flow Metab.* 19:624–633.

150. Longa E Z, Weinstein P R, Carlson S, et al. (1989). Reversible middle cerebral artery occlusion without craniectomy in rats. *Stroke.* 20:84–91.

151. Colbourne F, Sutherland G R, Auer R N (1996). An automated system for regulating brain temperature in awake and freely moving rodents. *J. Neurosci. Methods.* 67: 185–190.

152. Bederson J B. Pitts L H, Tsuji M, et al. (1986). Rat middle cerebral artery occlusion: Evaluation of the model and development of a neurologic examination. *Stroke.* 17:472–476.

# ANALYTICAL CONSIDERATIONS FOR IMMUNOASSAYS FOR MACROMOLECULES

Marian Kelley[1] and Binodh DeSilva[2]

[1] *Centocor R&D, Inc. Randor, Pennsylvania*
[2] *Amgen, Inc. Thousand Oaks, California*

**Chapter Contents**

1 Introduction
2 Assessment Parameters
  2.1 Reference Material
  2.2 Assay Format/Reagent Selection
  2.3 Specificity and Selectivity
  2.4 Sample Collection Process/Matrix/Sample Preparation/Minimum Required Dilution
  2.5 Standard Calibrators and Standard Curves
  2.6 Precision/Accuracy: Role of QCs
  2.7 Range of Quantification
  2.8 Sample Stability (Bench Top, 4°C, –70°C)
  2.9 Dilutional Linearity and Parallelism
  2.10 Robustness and Ruggedness
  2.11 Run Acceptance Criteria
3 Partial Validation, Method Transfer, and Cross-Validation
  3.1 Partial Validation
  3.2 Method Transfer
  3.3 Cross-Validation
4 Total Error Versus 4–6–20 Rule
  4.1 Total Error
5 Documentation
6 Conclusion
  References

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

## 1 INTRODUCTION

Current guidance documents for the assessment of analytes and their metabolites in biological fluids are focused primarily on the quantification of small molecules using chromatographic and mass spectrometry platforms [1–7]. Specific recommendations coming out of these guidances fall short in the immunoassay realm. The antibody binding of analytes directly from biological fluids confers a practicality and specificity unique to these assays. These assays are typically nonlinear and experience specific issues associated with the protein matrix, e.g., the presence of endogenous materials, to name just one. Until the authorities propose a guideline targeted at best practices in developing and validating ligand binding assays (antibody-based immunoassays) to support pharmacokinetics/toxicokinetics (PK/TK), the following recommendations, based on the paper generated by a 10-member panel from Biotech, Pharma, and CROs [8], and supported by several previous publications on the topic [9, 10], are proposed.

Assay development and implementation are dynamic processes. The life-cycle steps of development, validation, and in-study assay monitoring can be reassessed throughout the process as the analyst benefits from the experience of extended use, adds stability data, or updates current precision and accuracy data.

During early immunoassay development, it is important to assess the "intended use" of the assay. When the intended use is to support pharmacokinetics in either nonclinical PK/TK or clinical bioequivalence (BE) studies, developing a validatable assay is of utmost importance. A well-thought-out validation plan and fully documented assay methods and results that support the validation are crucial. To withstand the scrutiny of the authorities, and to guarantee the optimal method is applied to your drug development process, the following parameters should be assessed critically, validated to specific criteria, and fully documented.

## 2 ASSESSMENT PARAMETERS

The goal of the development phase for immunoassays is to establish a method that can consistently produce a reliable result culminating in a validation plan with established target acceptance criteria for accuracy and precision. The following recommended parameters are assessed critically during early development, evaluated in late development, and finally, confirmed during validation.

### 2.1 Reference Material

Inherent to macromolecules is the associated variability (i.e., glycosylation and deamidation) of the material from different preparations. Therefore, it is not usually feasible to obtain a "reference standard" for a macromolecule drug for use in bioanalysis. Typically, what is obtained is a well-characterized product. Therefore, it is important to clearly state the source of the material and to refer to any documentation describing characteristics of that material. Whenever possible, the standards, the validation samples, and quality control (QC) samples should be prepared from separate vials of the same source material. In the case of lyophilized material, it is preferred to reconstitute two separate vials to prepare standards and QCs. In

circumstances of limited availability of reference material (e.g., drug interaction studies where the pure standard is not available except from a limited source of commercial kits), standards and QC samples can be prepared from single aliquots after checking the comparability between lots or other commercial sources. In any circumstance, the lot numbers, batch numbers, and supporting documentation should be carefully monitored. This is a critical parameter that should be evaluated during method transfers and cross-validation studies.

### 2.2 Assay Format/Reagent Selection

Assay formats more recently have been defined as much by the platform (ELISA, RIA, Bioveris, Luminex) as by the type of immunoassay (direct or indirect binding, sandwich, competition or inhibition). Considering several different formats before selecting one is optimal, if time allows, because there are pros and cons to each format. Solid phase formats include glass, treated plastics, and coated plates (e.g., streptavidin). As antibody binding to the coated surface is highly dependent on pH, buffer content, salt concentration, detergents, nonspecific carrier protein, temperature, and shaking, any relevant conditions should be tested during the development of an immunoassay. When solid phase formats prove difficult, solution phase binding, for instance using coated beads, can result in a practical and sensitive assay.

The antibody or pair of antibodies selected for the assay format is the primary critical reagent for an immunoassay because it will ultimately define the specificity and sensitivity of the assay. Therefore, whenever possible, as many antibodies or pairs should be tested hand-in-hand with the selection of the assay format before moving into further development. This may include whole or fragmented monoclonals, polyclonals, and combinations as the assay format demands. The time it takes to develop and purify unique reagent antibodies demands that this start as early in the drug development process as possible.

Other critical reagents are those that impact the sensitivity, ruggedness, and consistency of the assay and especially include the detection system. It is a recommendation that a list of critical reagents and criteria for their acceptance be identified before initiating sample analysis. These must also be sourced, characterized, documented, and stored under conditions defined by stability investigation. When in-house reagents make up any part of the immunoassay, it is important to define and document the source, preparation, expiration, storage conditions, and acceptance criteria for their use to ensure a reliable result.

Once the final conditions are established, it is important to understand the assay's accuracy and precision capabilities because target criteria should be set for both before entering into validation. It is also good practice during validation to "stress the system" in a manner representing a true sample analysis run. Therefore, if the "batch" size is expected to be unusually large, a comparable batch, run during validation, should demonstrate that it could achieve the same acceptance criteria. It is particularly useful to evaluate the variability of QCs on every plate to predict the outcome expected during sample analysis.

The established/validated format should be used during sample analysis. Changes to the format must be revalidated to the extent that a consistent format is confirmed. The acceptability of changes to critical reagents should be documented before placing them into the sample analysis assays.

### 2.3 Specificity and Selectivity

One challenge of interpreting results derived from immunoassays is understanding exactly what the assay is detecting. Fundamental to this is knowing to what the antibody binds, that is, determining the antibody's specificity. A related parameter, selectivity, refers to the assay's ability to solely measure the target analyte in the presence of other protein constituents in the sample, some in very high concentrations.

Typically, experiments designed to determine the antibody's specificity involve assessing their cross-reactivity with similar compounds. When feasible, it is suggested that the sample matrix is spiked with physio-chemically similar compounds or variants of the target analyte.

Selectivity is the easier of the two parameters to assess experimentally. After the assay format is relatively fixed, multiple lots (at least 10 normals and 10 of the intended patient population in the case of clinical samples) of the sample matrix should be spiked near the lower limit of quantification and the recovery of the analyte should be calculated. The percent recovery of spike (expressed as relative error or %RE) is calculated as: 100× (observed mean conc – nominal concentration)/nominal concentration. This will aid in finally selecting a lower limit of quantification, before initiating assay validation.

An additional experiment to understand selectivity involves spiking both the diluent and the matrix with standard curve calibrators to detect matrix effects at all concentrations across the range of calibration.

Interfering substances, especially in patient populations, that require some evaluation may include prescription and over-the-counter medication, herbal medications, dietary substances, and even the presence of antidrug antibodies and associated immune complexes. During sample analysis, new matrix conditions may develop that would require further investigation and could include hemolysis, lipemia, rheumatoid factor, and so on.

Both specificity and selectivity experiments, if performed during the late development phase, after the assay format and critical reagents are fixed, may be cited in the validation report. Otherwise the experiments can be performed during validation. In both cases, the *a priori* acceptance criteria should be applied and acceptable recovery should be realized in at least 80% of the samples. Additionally, as lots of critical reagents are replenished, these types of assessments must reconfirm their acceptability.

### 2.4 Sample Collection Process/Matrix/Sample Preparation/Minimum Required Dilution

As mentioned, the ability of the antibody to bind to its target is the basis for assay sensitivity; therefore, maintaining the integrity of the analyte is vital. Frequently overlooked, the selection of sample matrix type and process for collection are crucial elements in bioanalysis. Rigorous investigation should go into the stability of the sample as it is collected, the anti-coagulant requirements, protease inhibitors, and other possible additives, and it should include the postprocessing storage. Furthermore, a process must be defined that provides for sample tracking "from cradle

to grave" documenting the chain of custody, the sample ID, and the time spent at each location and temperature through analysis. After analysis, the archival and destruction process should be identified.

For practical reasons, the matrix of choice for immunoassays is typically serum. The plate washers or automated systems used during sample analysis easily clog when plasma is used, which then has the potential to impact the validity of the result. However, protein analytes, including cytokines that are identified as labile and prone to degradation in serum, would preferentially be collected in plasma. When collecting citrated plasma, it is wise to clearly identify the salt form of citrate (Na, K, Li) so that it is transparent what the comparable matrix should be for QC preparation.

Whenever possible, the matrix expected in the samples should be used in the preparation of the standard calibrators and QCs. Individual donors (found acceptable when assayed as blank and spiked) are pooled to produce the standard curve and QC matrix; then aliquots are spiked to produce the standard curve and QCs. Some matrices are difficult to obtain. To accommodate this, as long as the QCs prepared in the intended matrix demonstrate the accuracy and precision expected, non-matrix standard curve calibrators are acceptable.

QC accuracy assessments can be challenging when the therapeutic drug shares an endogenous protein. In this case, the endogenous protein concentration can either be added to or subtracted from the final value. Before establishing the calculations, the necessary experiments must be performed, and the consistency of the endogenous value contribution must be documented.

In cases where the endogenous interferents do not generate a linear signal, then a minimum required dilution (MRD) may be applied. A MRD is the dilution of a matrix that finally permits an accurate and precise measurement of analyte. Although the diluting sample reduces the sensitivity of the assay, it may be required to develop and validate an accurate and precise method.

### 2.5 Standard Calibrators and Standard Curves

As described in the Guidance for the Industry, May 2001 [7], the "calibration curve is the relationship between instrument response and known concentrations of the analyte" in the matrix of interest. In immunoassays, the affinity and avidity of the antibody(ies) used to develop the standard curve are fundamental to the range because they define the sensitivity of the assay. The antibody (or pairs) once selected should be used throughout validation and sample analysis. During the development phase, it is a good idea to use many more calibrators and replicates evenly across the range to thoroughly describe the concentration-response relationship and to investigate the "best fit." Thereafter, select at least six to eight calibrators, over the entire range of the curve before the start of validation. Once validation begins, the standard curve calibrators should not be adjusted. Spiked standard curve points may be used outside the quantification range of the curve to help fit the curve, but they are not used in the assessment.

"Best-fit" can be evaluated in a straightforward manner by fitting several regression methods to the data generated during the development phase. The relative error of the back-calculated standard calibrators should be calculated. As expected,

the regression that consistently generates the most accurate and precise data is selected for confirmation in the validation phase. Usually a four- or five-parameter logistic with/without weighting describes the s-shaped immunoassay curve optimally; however, the smoothed spline has also been used with success with some platforms. It is important to critically assess whether the curves fall consistently above or below the expected concentration. This "trending" can invalidate a regression model.

### 2.6 Precision/Accuracy: Role of QCs

One function of the QC sample (or validation samples as they are sometimes called when used during the validation phase) is to mimic the study sample. Therefore, the therapeutic drug is spiked into a neat sample matrix, and not into a diluted or otherwise modified matrix. The QCs should be stored preferably along with the samples in the same freezer or at least in a comparable freezer. Documentation of freezer conditions will be critical to provide evidence that the conditions of samples and QCs are similar.

A second function of the QC sample is to measure the degree to which the assay is under control as determined by the assessment of accuracy and precision in multiple runs. Accuracy is described as the % relative error (%RE) or the amount the assayed value deviates from the nominal value. The precision is described as the % coefficient of variation (%CV) by dividing the standard deviation by the mean value recovered value.

During development, preliminary assessments are conducted to predict not only the QC concentration values to be used but also the target acceptance criteria. In this case, it is prudent to include more QC levels than will finally be used during validation. For instance, three levels prepared at the LLOQ will likely provide a clear indication of which QC is acceptable and where the assay cannot support the target acceptance. The same can be performed for the ULOQ. At least three levels between the two asymptotes are necessary because these will be used during sample analysis. For the validation process, five QC levels should be selected for testing, based on the accurate and precise readout: the final LLOQ to be confirmed; the LQC; less than three times the LLOQ; the MQC, placed about the middle of the curve; and the HQC, placed no more than 75% below the ULOQ and the ULOQ.

During validation, the five QCs should be tested in a minimum of six runs, at least in duplicate. It is not acceptable to discard any runs during this testing. The inter- and intrabatch precision and accuracy from all runs performed should be reported. It is generally expected that the precision and accuracy acceptance be at least 20–25% of the target range with a Total Error not exceeding 30%. (Total Error acceptance criterion is discussed later in this report.)

During sample analysis, only the LQC, MQC, and HQC defined during validation need to be run in duplicate with every assay. The precision and accuracy should conform to that established during validation. Typically, run acceptance criteria also require that at least one QC at each concentration level is acceptable, and that four out of the six QCs assayed must be acceptable. It is recommended that method acceptance criteria that are consistent with the validation be used during sample analysis [11, 12].

### 2.7 Range of Quantification

It is important to understand that the LLOQ and ULOQ have been selected because they define the true, accepted limits of the standard curve. Standard curve points below or above the LLOQ and ULOQ, respectively, are considered extrapolated and cannot be used to report sample concentrations. The LLOQ and the ULOQ validation samples and not the standard curve samples spiked into neat matrix define the standard curve range of quantification.

With certain toxicokinetic studies or therapeutic drugs dosed in high concentrations, the expected concentration in the sample may be well outside the quantification range as defined by the ULOQ. These samples must be diluted into the accepted range of the standard curve to determine an initial concentration and then multiplied by the dilution factor to determine the final concentration of the analyte in the sample. Although it is preferable to dilute the sample in matrix, in many cases, the process is performed in a buffer diluent.

### 2.8 Sample Stability (Bench Top, 4°C, –70°C)

The underlying reason that samples are tracked from "cradle to grave" is to document the integrity of the sample. Stability testing under all conditions to which a set of samples in a study is exposed supports the statement of integrity. Therefore, such testing of QC samples in matrix comprises evaluating stability during sample collection and processing: –20°C or –70°C long-term storage to cover the time from when the sample is processed to its final analysis, bench top under conditions similar to analysis, resident time in laboratory refrigerators and freezers, and freeze-thaw to cover the expected or possible cycles. Another consideration is to test the stability of the analyte in intermediate stock solutions if stored for later analysis.

The samples should be thawed under the same conditions used to thaw study samples. Similarly, refreeze the samples for a length of time that would ensure complete freezing, usually 12–24 hours. Samples should be assayed using a freshly prepared standard curve, and acceptable QCs should be used to monitor assay performance. The stability samples should meet the criteria for acceptance as defined for in the validation plan.

### 2.9 Dilutional Linearity and Parallelism

Dilutional linearity is the ability of a QC sample spiked at a high concentration (at least as high as the expected $C_{max}$ in a study) to be acceptably diluted into the range of the standard curve so that the initial concentration determined off the regression and multiplied by the dilution factor produces a final concentration with an accuracy and precision expected for that assay. Usually, serial dilutions are employed to ensure that at least two to three fall on the acceptable range of the curve so that linearity can be demonstrated. When a study design produces samples that require a dilution greater than that established, it is advisable to employ an equivalent dilutional QC that is run in the assay.

To evaluate the "hook effect," dilute a very high QC (100- to 1000-fold greater than the ULOQ) so that the expected concentration read off the regression reads above the ULOQ.

Parallelism is determined in a manner similar to dilutional linearity but with incurred samples. This is usually not possible until after the initiation of a study. When incurred samples are available, one strategy is to pool the $C_{max}$ samples to create parallelism QCs. Serially dilute the sample so that at least two to three fall on the acceptable range of the curve to demonstrate linearity.

### 2.10 Robustness and Ruggedness

Robustness and ruggedness experiments are used to demonstrate how reproducible a method is when conditions vary. An assay method protocol defines the exact steps to be followed to ensure an accurate and precise result. Unfortunately, slight-to-large deviations occur in the everyday process, which may or may not impact the result. Assay conditions like incubation temperature or exposure to light define the robustness of the assay, whereas changes to the routine, for instance, multiple analysts or different instruments, define the ruggedness. The actual batch size for a routine sample analysis run is frequently overlooked but is often an impactful ruggedness measurement that should be assessed.

### 2.11 Run Acceptance Criteria

Because validation QCs are used to assess the overall accuracy and precision of the method during the validation phase, no runs initiated for validation may be eliminated due to QC failure. Therefore, the acceptance criteria for the standard curves are used to accept each assay. At the conclusion of the validation, run acceptance criteria for QCs to be used in-study are summarized and run acceptance criteria are established based on this data.

During sample analysis, the standard curve should be evaluated for acceptance before evaluating the QCs. Once the curve has been accepted, then the performance of the QCs should be evaluated. The acceptance criteria established during the precision and accuracy (4–6–20 rule) should be followed as described.

## 3 PARTIAL VALIDATION, METHOD TRANSFER, AND CROSS-VALIDATION

Method validations can be classified into three broad categories, full, partial, and cross. A full validation is done for any new methods and involves method development, prestudy, and in-study validation. It is essential that a full validation be conducted for changes in species (e.g., rat to mouse), change in matrix within a species (e.g., rat serum to rat urine). However, a partial validation is sufficient when a minor change is made to an already fully validated method as described below.

### 3.1 Partial Validation

A partial validation is conducted when minor changes to a method are made. These may include method transfer, changes to anticoagulant (e.g., EDTA, heparin, citrate), changes in the reagents used in the method (especially pivotal reagents

such as the primary antibody or secondary antibody), sample processing changes (e.g., how fast a clot needs to be spun, collection vessels, and storage condition), changes to sample volume (e.g., if the volume of sample is changed from 100 to 200 μL per well), extension of the concentration range, selectivity issues (e.g., different disease population and concomitant medication), conversion of a manual to an automated method, qualification of an analyst, and so on. Partial validations can range from a single intra-assay accuracy and precision run to a nearly full validation. Although sample-processing changes may only require one run, several runs would typically be expected for changes to lots of reagents. Transferring an analytical method may require substantially more experimentation. Typically, for partial validations, three accuracy and precision runs are conducted and the inter-assay accuracy and precision criteria are compared against the original validated assay. If these criteria are met, the change to the method will be accepted and the proper documentation should be made to the method or method standard operating procedure (SOP). Other validation parameters have to be compared depending on the situation where these experiments are conducted to scientifically justify the change; e.g., if the method is to be used with a diseased population different to the original intended use of this method, then additional selectivity experiments must be conducted to evaluate the matrix effect.

### 3.2 Method Transfer

Method transfer is the situation in which the method is fully validated in one laboratory (sending laboratory) and transferred to another laboratory (receiving laboratory) and requires at least a partial validation. The parameters to evaluate during this partial validation are accuracy and precision of the method, selectivity, bench-top stability, and so on.

In addition to the required documentation (e.g., method description, validation report, and certificate of analysis), the sending laboratory should provide information on those factors that may affect the ruggedness of the assay (e.g., identifying pivotal reagents and material). The method transfer process can be conducted in a multiphase manner, a transfer phase in which the sending laboratory participates in the transfer of the method at the receiving laboratory; the feasibility evaluation phase, in which the receiving laboratory runs the assay independently; the validation phase, which could be the phase where the partial validation is conducted; and finally a qualification phase in which both the sending and the receiving laboratory assays blinded samples. The method transfers require a plan or protocol that defines the process (e.g., experiments to be conducted) and the acceptance criteria. It is not uncommon for the sending laboratory to send personnel to physically demonstrate and train the personnel at the receiving laboratory.

Once the transferred method is validated at the receiving laboratory, an ideal scenario is to have both the sending and the receiving laboratories analyze blinded, 30 spiked samples covering the standard curve range, and 30 pooled incurred samples. The two sets of data may be compared by using a predefined statistical equivalence test [12, 13]. Alternatively, the differences between the two sets of data may be compared using an agreed to range of acceptability. Any acceptance criteria must be set *a priori* and documented in a plan or protocol before executing this phase of the method transfer process.

### 3.3 Cross-Validation

Cross-validation is conducted when two validated bioanalytical methods are used within the same study or submission, for example, ELISA assay to Biacore and ELISA to a liquid chromatography/mass spectrometry. It is recommended that test samples (spiked and/or pooled incurred samples) be used to cross-validate the bioanalytical methods. Data should be evaluated using an appropriate predefined acceptance criteria or statistical method [12, 13]. It should be cautioned that many times the methods that are being cross-validated may not have the same range of quantification. In these situations, it is necessary to prepare spiked samples within the range that are common to both methods for comparison.

## 4 TOTAL ERROR VERSUS 4–6–20 RULE

### 4.1 Total Error

The total error of a measurement takes into account both the systematic error (bias) and the random error components. Any measurement that is made during an experiment consists of both of these error parameters, and it is not possible to separate these two. Therefore, it is scientifically correct to use the total error criteria to assess the acceptability of a quality control result during a run.

As described in this chapter, prestudy validation runs are accepted based on the standard-curve acceptance criteria. No run acceptance criteria are applicable for prestudy validation sample assessments; i.e., no run can be rejected due to poor validation sample performance during accuracy and precision evaluation, and all data from the prestudy validation runs are reported without exceptions. In some cases, there may be assignable cause (e.g., technical issues) for removal of a validation sample data point before the calculation of the cumulative mean. Exclusion is applied at the end of the validation study period and must be documented as described in the documentation section.

For each in-study run, the standard curve must satisfy criteria described in the standard-curve section; however, run acceptance is based primarily on the performance of the QC samples. When using total error for ligand binding assays of macromolecules, the run acceptance criteria recommended in the precision and accuracy section requires that at least four of six (67%) QC results must be within 30% of their nominal values, with at least 50% of the values for each QC level satisfying the 30% limit. The recommended 4–6–30 rule imposes limits simultaneously on the allowable random error (imprecision) and systematic error (mean bias). If the application of an assay requires a QC target acceptance limit different than the 30% deviation from the nominal value, then prestudy acceptance criteria for precision and accuracy should be adjusted so that the limit for the sum of the interbatch imprecision and absolute mean RE is equal to the revised QC acceptance limit.

## 5 DOCUMENTATION

It is critical to document the information generated during the development of a method in a laboratory notebook or other acceptable format. Information on the

following assessments should be generated during assay development: critical assay reagent selection and stability, assay format selection (antibody[ies], diluents, plates, detection system, etc.), standard-curve model selection, matrix selection, specificity of the reagents, sample preparation, preliminary stability, and a preliminary assessment of assay robustness. At the end of the development portion of the assay life cycle, a draft method or an assay worksheet should be generated to be referred to during the prestudy validation.

A validation plan should be written before the initiation of the prestudy validation experiments. Alternatively, reference to an appropriate SOP can be made to ensure that a documented outline exists for the experiments required for prestudy validation. This plan can be a stand-alone document or can be contained in a laboratory notebook or some comparable format. The documentation should include a description of the intended use of the method under consideration and a summary of the performance parameters to be validated that should include, but may not be limited to, standard curve, precision and accuracy, range of quantification, specificity and selectivity, stability, dilutional linearity, robustness, batch size, and run acceptance criteria. The plan should include a summary of the proposed experiments and the target acceptance criteria for each performance parameter studied.

After completion of the validation experiments, a comprehensive report should be written. The format of the report may be dictated by internal policies of the laboratory; however, such reports should summarize the assay performance data and deviations from the method SOP or validation plan and any other relevant information related to the conditions under which the assay can be used without infringing the acceptance criteria.

Cumulative standard curve and QC data tables containing appropriate statistical parameters should be generated and included with the study sample values in the final study report. Unlike the prestudy validation, failed runs are not included in these tables. Additional information to be included in the final report is a description of deviations that occurred during the study, a table of the samples that underwent repeat analysis and the reasons, and a table with details on all failed runs.

## 6 CONCLUSION

The primary focus in the analytical considerations of immunoassays for macromolecules should be in the development phase. By thoroughly defining each component of the immunoassay in the first phase of the assay life cycle, the resulting validation plan should be concise and the validation process should be straightforward. Several facets of the assay defined in the late development stage, such as specificity and dilutional linearity, can appropriately be included in the final validation report.

A typical validation will include at least six precision and accuracy assays to define the consistency of the assay. Within those assays, several parameters can be defined, including early stability, specificity, selectivity, and range of quantification. No assay run should be eliminated except for a true and documented analyst's error.

The validation QC samples define the range of the assay, and no values below the LLOQ or above the ULOQ may be reported. Within the six validation assay

runs, the validation samples are used to define the cumulative precision and accuracy. During validation, no validation sample may be eliminated to show the true profile of the assay.

During sample analysis before assessing the QC samples for acceptance, the standard curve must be deemed appropriate by predetermined criteria. Only after the curve is accepted may the assessment of QC samples continue. QC sample results determine whether the assay run is valid. Acceptance criteria can be based on 4–6–20 rule or on Total Error and should be predicated on the criteria used in both the development and the prestudy validation phase. Overall, the immunoassay is a highly sensitive assay that can be used to quantify protein and peptide drugs in a biological matrix, often routinely in the pg/mL range.

## REFERENCES


1. Shah V P, Midha K K, Dighe S V, et al. (1992). Analytical methods validation: Bioavailability, bioequivalence, and pharmacokinetic studies. *J. Pharm. Sci.* 81:309–312.
2. Shah V P, Midha K K, Dighe S, et al. (1992). Analytical methods validation: Bioavailability, bioequivalence, and pharmacokinetic studies. *Pharm. Res.* 9:588–592.
3. Riley C M, Rosanke T W (1996). *Development of validation of analytical methods: progress in pharmaceutical and biomedical analysis*, vol 3. Ed. Elsevier, Pergamon, NY.
4. Shah V P, Midha K K, Dighe S, et al. (1991). Analytical methods validation: Bioavailability, bioequivalence and pharmacokinetics studies. Conference Report. *Eur. J. Drug Metabol. Pharmacokinet.* 16:249–255.
5. Guideline on validation of analytical procedures: Definitions and terminology. *International Conference of Harmonization (ICH) of Technical Requirements for the Registration of Pharmaceuticals for Human Use.* Geneva, 1995. Switzerland, (1996).
6. Shah V P, Midha K K, Findlay J W A, et al. (2000). Bioanalytical method validation. A revisit with a decade of progress. *Pharm. Res.* 17:1551–1557.
7. Guidance for the Industry: Bioanalytical Method Validation (2001). US department of Health and Human Services FDA (CDER) and (CVM).
8. DeSilva B, Smith W, Weiner R, et al. (2003). Recommendations for the Bioanalytical Method Validation of Ligand-binding Assays to Support Pharmacokinetic Assessments of Macromolecules. *Pharmac. Res.* 20(11):1885–1900.
9. Findlay J W A, Smith W C, Lee J W, et al. (2000). Validation of Immunoassays for bioanalysis: A pharmaceutical industry perspective. *J. Pharm. Biomed. Anal.* 21:1249–1273.
10. Searle S R, Casella G, McCulloch C E, (1992). *Variance Components, Chapter 3.* Wiley, New York.
11. Mee R W (1984). β-expectation and β-content tolerance limits for balanced one-way ANOVA random model. *Technometrics.* 26:251–254.
12. Bland J M, Altman D G (1999). Measuring agreement in method comparison studies. *Stat. Meth. Med. Res.* 8:135–160.
13. Hartmann C, Smeyers-Verbeke J, Penninckx W, et al. (1995). Reappraisal of hypothesis testing for method validation: Detection of systematic error by comparing the means of two methods or two laboratories. *Analyt. Chem.* 67:4491–4499.

# CHROMATOGRAPHY-BASED SEPARATION OF PROTEINS, PEPTIDES, AND AMINO ACIDS

JAN MOEBIUS,[1] REMCO SWART,[2] AND ALBERT SICKMANN[1]

[1]*Rudolf-Virchow-Center for Experimental Biomedicine, Würzburg, Germany*
[2]*LC-Packings—A Dionex Company, Amsterdam, The Netherlands*

**Chapter Contents**

1 Introduction
2 Basics of Chromatography
  2.1 Hardware Requirements
  2.2 HPLC Columns
  2.3 Miniaturization
3 Separation of Peptides
  3.1 Short Introduction
  3.2 Reversed Phase Separation of Peptides
  3.3 Column Media and Resins
  3.4 Multidimensional Separation of Peptides
4 Separation of Proteins
  4.1 Short Introduction
  4.2 Reversed Phase–Liquid Chromatography (RP–HPLC) Separation of Proteins
  4.3 Ion-exchange Separation of Proteins
  4.4 Multidimensional Liquid Chromatography of Proteins
5 Separation of Amino Acids—Amino Acid Analysis
  References

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

# 1 INTRODUCTION

Since the first commercial introduction of high-performance liquid chromatography (HPLC) in the 1970s, this versatile method has gained wide popularity, resulting in a phenomenal increase of successful applications described in literature. HPLC is a highly efficacious method in peptide and protein analysis especially for the preparation of samples for further analysis by electrospray mass spectrometry or quantification using amino acid analysis. This story of success is related with the continuous development and miniaturization of flow rates and column diameters and the high level of automation combined with the reliability of the instruments. Today HPLC can separate minute amounts of peptides and proteins and find those areas of biochemistry and biotechnology where only small amounts of sample are accessible. The following chapter is focused on the analytical aspects of liquid chromatography of biomolecules from amino acids to proteins.

# 2 BASICS OF CHROMATOGRAPHY

## 2.1 Hardware Requirements

The chromatography-based separation of amino acids, peptides, or proteins has to be divided into the different applications of preparative and analytical chromatography. For analytical HPLC, the identification and quantification of single components from a low complex mixture is the most important task. Therefore, the analytes of interest have to be separated as much as possible.

The preparative biochromatography is focused on the isolation and enrichment of a single protein species. Typically, the HPLC separation is performed with bigger columns providing higher loading capacities but relative short column lengths in order to keep separation time short. However, in this chapter, we will focus on the analytical aspects of HPLC separation.

Generally, solvent pumps with high pressure capability are required for HPLC due to the backpressure generated by the dense package of the column resin. The typical specifications are pressure limits of about 400 bar and 5800 psi, respectively, although this limit is normally not required. Recently, new column materials with small particle sizes of about 1.3 μm have been introduced that require HPLC pumps and tubings that resist pressures of up to 1000 bar (approximately 14,000 psi) [1–3]. For the separation of biomolecules, two different HPLC-pump types have been established: piston pumps and syringe pumps.

The basic principle of a syringe pump principle is simple: A motor activates a piston that pressurizes the solvent within a syringe. Although this type of a pump was widely used in the past, the syringe pump has some limitations; e.g., it usually cannot deliver the same high-pressure limits as stainless steel piston pumps. In addition, the amount of solvent for a single chromatographic separation is restricted to the syringe volume. Moreover, these pump systems are most often designed for binary gradients because for each solvent, an syringe pump is required. Recently, syringes have experienced a revival due to their scalability to smaller dimensions, which enables access to native nanoscale flow rates [4, 5]. Nevertheless, piston pumps are the pump systems that are provided by all major vendor companies

because of several imminent advantages in flexibility and performance. A small piston pressurizes the solvent continually within the flow system. Check-valves prevent the backflush of the solvent. The refilling of the piston causes pulsation, which can be eliminated by using a second pump unit that performs the pump stroke while the first piston refills. The gradient is formed by a mixing device that can be located at the high- or low-pressure side, i.e., before or after the pump unit. Syringe pumps normally use a mixing-tee located directly after the two syringes. The gradient is formed by different movement speeds of the syringes.

Piston pumps can be designed for low-pressure mixing as well as for high-pressure gradient formation. The low-pressure mixing is realized by solenoids that open the flow path of a solvent for a discrete period of time. Subsequently, the solvents are mixed and pressurized. This setup applies only to a single pump unit, whereas high-pressure mixing in piston-pump systems requires a single pump unit for each solvent. The major advantage of this setup is the enhanced response time to changes in the gradient, which is required for steep and short gradients. Another aspect of HPLC pumps for biochromatography is the availability of bio-inert versions, which can be of importance for several applications covering extreme pH-values as well as several modified peptides that are retarded by stainless steel [6]. In contrast to stainless steel, all components of the flow-path are composed of inert materials like PEEK, Titanium, or Saphire (for the pistons).

Besides the pump as a solvent delivery device, the HPLC valves are important components for switching solvent lines (e.g., injection loops and parallel column setups). They are designed to resist the applied pressure and feature a small dead volume to prevent peak dispersion. HPLC valves are available in different switching modes and port numbers. Additionally, valves are also fashioned as bio-inert versions. For several applications, the HPLC column should be used in a column-oven for ensuring constant temperature during separation.

To introduce samples into the flow path, the easiest way is the usage of a sample loop that is connected to an HPLC valve. After loading and switching, the loop is integrated into the flow path and the sample is delivered to the separation column. Today, this task is done by autosamplers that can temper samples and provide 24 hours of fully automated and reproducible sample injection.

A further important component of each HPLC system is the detection device. Often, Ultraviolet (UV)-detectors are used for this purpose because many analytes absorb light in the range of 200–350 nm. UV detectors provide reasonable sensitivity and dynamic range as well as a low noise level.

Recently, multi-wavelength detectors that can switch between different wavelengths have been used. Diode array detectors (DADs) use a light source that is dispersed by a holographic grating after transmission through the flow cell. The light of the single wavelengths is reflected onto a diode array with several hundreds of diodes acquiring in real time a full-spectrum chromatogram. This leads to a better characterization of each eluted component.

Electrochemical detectors are applied for analytes that are capable of redox-reactions. The application field of biochromatography for PADs covers only a small spectrum of amino acids, whereas a broad range of applications is established for smaller molecules.

One of the most important detectors for proteins and peptides is a mass spectrometer that is necessary for structural analysis of peptides. The high sensitivity

and capability of structural information are the two major reasons for the success of HPLC coupled to mass spectrometry in proteomics.

Often, a pressure is applied to generate the mobile phase flow though the column. Alternatively, in capillary electrochromatography (CEC), an electric field is applied along the column to drive the mobile phase through packed capillary columns with internal diameters (IDs) of 50–100 μm. This principle was proposed by Pretorius et al. [7] and further developed by other groups [8, 9]. In these electrically driven systems, the mobile phase flow is characterized by a flat velocity profile, which results in higher separation efficiencies. However, ineffective heat dissipation restricts the use of high electric field strengths for large ID columns and thereby limits the applicability of the technique.

## 2.2 HPLC Columns

The efficiency of an HPLC column is determined by the type of stationary phase particle and the quality of the column packing. The internal diameter of the column is not affecting the performance characteristics [10], although it has been deliberated that columns with low aspect ratio, i.e., ratio between internal diameter and particle diameter, have higher separation efficiency. A recent and extensive experimental study did not show convincing evidence for higher column efficiency for columns with an aspect ratio from 7.5 to 30 [11].

A more effective way to increase the efficiency in HPLC is by reducing the stationary phase particle size, as predicted by the well-known van Deemter equation [12], which describes the plate height in HPLC as the sum of different contributions. In this respect, submicron particles have been applied and efficiencies up to 730,000 plates/m have been reported for small molecules [13, 14].

However, as a result of the increased flow resistance of these submicron particle packed columns, extremely high pressures of 200–300 MPa are necessary to drive the mobile phase. Working at these high pressures, several phenomena should be taken into account: (1) increased viscosity of the mobile phase affecting the efficiency and pressure drop over the column, (2) Joule heating effects that may cause radial temperature gradients in analytical and narrow-bore columns, and (3) linear velocity surge caused by mobile-phase compression responsible for band broadening [15].

Recently ultra-performance liquid chromatography (UPLC) has been introduced commercially as a new direction in liquid chromatography. In UPLC columns packed with 1.7-μm, particles are operated at pressures up to 100 MPa. The main advantage is obviously the reduction in analysis time that has been demonstrated for pharmaceutical analysis [16] and metabolomics [17].

The use of monolithic columns in LC has advanced rapidly since their first introduction in the 1990s [18–21]. In contrast to capillary columns packed with particulate stationary phases, monolithic columns consist of a single continuous support. Monolithic stationary phases can be subdivided in two classes, i.e., polymer- and silica-based materials.

Polymer monoliths are prepared by *in situ* polymerization of monomers such as styrene, acrylate, methacrylate, acrylamide, or cyclic compounds and have been reviewed thoroughly [19]. The selectivity of the monolithic column can be changed by incorporation of functional monomers during the polymerization. A

disadvantage of this procedure is that for each polymerization mixture, the conditions have to be optimized. Alternatively the functionality can be modified at the surface of the monolithic structure by (photo)grafting [22].

Silica monoliths are prepared by sol–gel technology starting from alkoxide precursors, such as tetraethoxy-silane or tetramethoxy-silane [23]. The structure of a silica monolith is characterized by an interconnected silica skeleton and bimodal distribution of 1–3-μm macropores (flow through pores) and 10–20 nm mesopores.

The main advantages of monolithic columns are the superior separation performance and low flow resistance. In addition due to their continuous nature, frits are not required to retain the stationary phase. The production process of monolithic columns is more flexible than that of packed columns; e.g., photo-polymerization can be applied to prepare monolithic structures or add selectivity locally. Both polymer- and silica-based monolithic capillary columns have been used for highly efficient separations in LC–mass spectrometry (MS) applications for proteomic research [24, 25].

### 2.3 Miniaturization of HPLC

Miniaturization in HPLC was started more than 25 years ago and was mainly driven by the need for analyzing minute samples [26–29]. The developments were accelerated by the widespread utilization of electrospray ionization (ESI) for interfacing LC and MS.

Reducing the ID of the LC column offers several advantages, among which are most importantly: (1) increased mass sensitivity for concentration sensitive detectors due to reduced chromatographic dilution, (2) easier coupling to MS interfaces [ESI and matrix-assisted laser desorption/ionization (MALDI)] due to the lower flow rates, and (3) a lower stationary phase and solvent consumption.

The increased mass sensitivity for concentration-sensitive detectors is the direct result of reduced dilution of the sample in small ID columns and is inversely proportional with the square of the reduction in column diameter. As a result of miniaturization in LC, a variety of new column formats were developed, which can be subdivided according to the ID. In Table 1, the ID, typical flow rate, and sample loading are listed for various HPLC techniques.

Analytical and narrow-bore HPLC columns are most commonly used for routine analysis, which require the highest possible robustness. Capillary- and nano-LC columns find application in various fields of biotechnology research, mainly for sample limited analysis of peptides and proteins [30–32].

**TABLE 1. Typical Dimensions and Flow Rates for Analytical RP Chromatography**

| | Inner Diameter | Flow Rate | Relative Gain in Sensitivity |
|---|---|---|---|
| Conventional LC | 4.6 mm | 1000 μl/min | Baseline = 1 |
| Narrowbore LC | 2.1 mm | 200 μL/min | 5 |
| Microbore LC | 0.5–1.0 mm | 50 μL/min | 35 |
| Capillary LC | 180–300 μm | 2–5 μL/min | 200 |
| Nano LC | 50–100 μm | 200 nL/min | 3500 |

A significant reduction of the column ID put stringent demands on the instrumentation for micro-LC as all volumetric extra-column dispersion contributions must be down scaled accordingly. Initial developments in this area were achieved on modified standard LC instruments. For approximately 10 years, dedicated instrumentation for micro-, capillary-, and nano-LC has become commercially available and was recently discussed [33].

The reduction or downscaling factor in going from an analytical to a capillary column equals the ratio of the square of the column diameters. Downscaling from a 4.6-mm ID standard HPLC column to a 300-μm ID packed capillary column equals a theoretical factor of 200 in sensitivity (see Table 1). This factor applies to all components of the LC system, like flow rate, injection and detection volume, and connecting capillaries [34]. Likewise, for a 300-μm ID capillary column, the flow cell volume must be reduced to ~50 nL to prevent extra-column peak broadening. For injection of the sample, the downscaling factor implies that an injection volume of 20 μL on a 4.6-mm ID column equals an injection volume of ~90 nL on a 300-μm ID capillary column. Consequently autosamplers used for capillary LC must be able to inject nanoliter volumes in a reproducible manner. However, in gradient elution, the preferred mode of operation for separation of peptides and proteins, sample preconcentration can be applied. Relatively large volumes, often exceeding the column volume several times, are injected while the compounds are concentrated at the head of the column or trap column. In nano-LC with typical flow rates of 100–300 nL/min, column switching is a widespread technique, e.g., for peptide mapping when the sample amount is limited. A trap column is mounted onto a switching valve (see Figure 1) to reduce the sample volume from several microliters present in the sample loop to a few nanoliters (basically the volume of the trap column). After sample loading, the trap column is switched in line with the separation column before starting the HPLC gradient. An additional advantage of column switching is the ability of online sample desalting. A common setup for

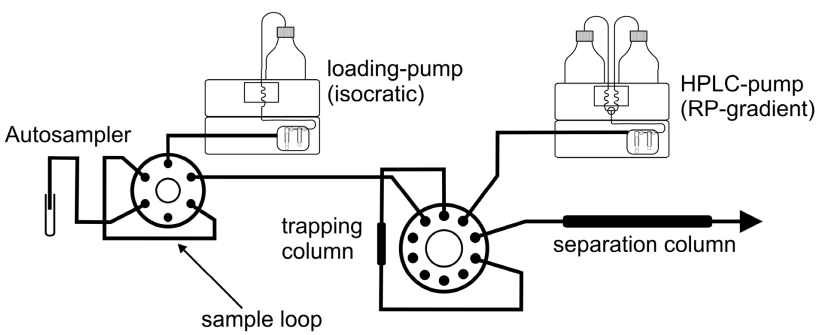


**Figure 1.** Schematic overview of typical reversed phase-HPLC with precolumn and column loading for peptide separation. The central switching valve has two positions: In position one, the sample is loaded onto the precolumn with the loading pump. After trapping and desalting of the sample, the valve is switched and the precolumn is integrated into the flow of the gradient pump. By increasing the amount of solvent B, the single peptides are separated on the separation column and afterward detected with a UV-detector and/or by mass spectrometry.

column switching in nano-LC is shown in Figure 1, although other configurations have been described as well [35, 36].

Typical flow rates applied in nano- and capillary-LC are in the range of 50 nL–10 μL/min according to Table 1. HPLC systems that can generate mobile phase gradients at these flow rates are available and are based on either flow splitting or dedicated micro- or nano-pumps. The generation of gradient profiles should be characterized by small delay volumes, efficient mixing, high accuracy, and acceptable reproducibility. Significant improvements in this respect have been achieved during the last couple of years, partly due to the incorporation of liquid flow sensors that can measure flow rates down to several nanoliters/minute. With current state-of-the art instrumentation, it is now possible to generate solvent gradients with a reproducibility better than 0.3% RSD at flow rates suitable for nano-LC.

## 3 SEPARATION OF PEPTIDES

### 3.1 Short Introduction

Reversed phase (RP) chromatography has been established as the superior separation technique for peptides. The chromatographic resolution of this method is higher compared with other methods like ion-exchange (IEX) or size exclusion chromatography (SEC).

### 3.2 Reversed Phase Separation of Peptides

RP–HPLC is well suited for direct online coupling with electrospray ionization–mass spectrometry (ESI–MS) due to solvent compatibility with the electrospray ionization technique.

The common gradient for peptide separation consists of a binary solvent system containing in solvent A: 0.1% formic acid and in solvent B: 0.1% formic acid and 84% acetonitrile. If no high-sensitive MS detection is required, formic acid can be replaced by TFA. For solvent B, the amount of TFA has to be reduced to 0.08% to prevent baseline shift during the gradient run time.

Gradients normally begin at 5% B with a linear increase to 50% B within 45 minutes. For rinsing the column, the percentage of B is increased to 95% and held for 5 minutes. Afterward the column is re-equilibrated to 5% B. Therefore, the complete gradient is finished within 1 hour. Extension to longer and thus shallower gradients is suggested if complex samples have to be separated.

The formic acid acts as an ion pair reagent that masks the peptide-charge and increases the hydrophobicity for better retention on reversed phase supports. Trifluoroacetic acid (TFA) has better characteristics for HPLC separation because of less UV absorption at 215 nm and superior properties as an ion pair reagent [37, 38]. Therefore, it is preferably used in solvents for loading on precolumns or separation without the need for MS detection. In this context, the usage of formic acid as an ion pair reagent instead of trifluoroacetic acid, which has less absorption in the UV range, is preferable because of its better ionization efficiency in MS ion sources. In contrast, TFA offers less UV absorption and superior separation results at the cost of inferior ionization rates.

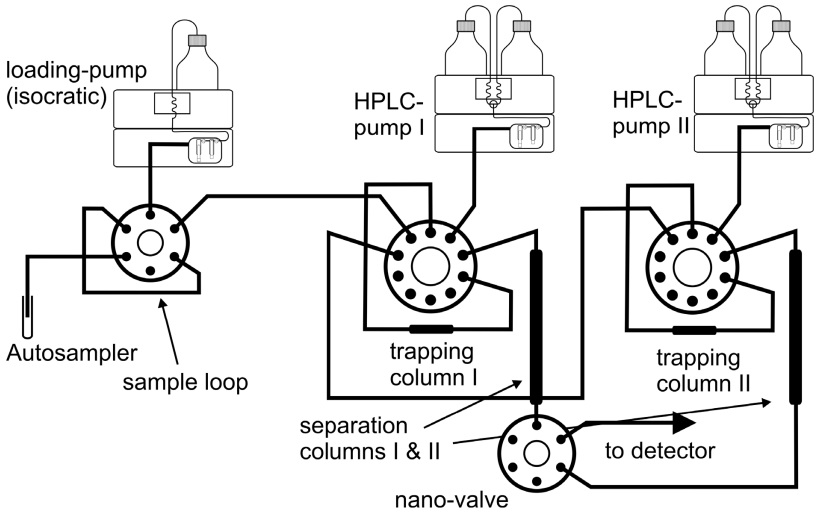


**Figure 2.** Set up of a dual-gradient HPLC system. This HPLC-system consist of a normal reversed phase separation system with an additional gradient pump and HPLC-valve. Two complete sets of pre- and separation columns are required. This setup enables the equilibration and sample preconcentration on the one column pair, whereas on the second column pair, the separation is performed. This reduces the dwell time of a connected mass spectrometer and thereby increases the sample throughput per day significantly. The final nano-valve directs the flow of the active separation column to the connected detector.

MS represents the most expensive part of a nano-LC–MS systems. Therefore, dwell times are expensive and should be reduced to a minimum. Parallel LC-systems consist of a more complex setup with two pairs of pre- and separation columns (see Figure 2).

***CARLA Test.*** Dead volume is a serious obstacle for high-resolution RP separation using nano-HPLC. A dead volume acts as a passive mixing chamber in the flow path that levels the preformed gradient, thereby leading to peak broadening and reduced sensitivity. The definition of a dead volume is a relative subject because a 0.5-μL dead volume in a preparative HPLC with a 1-mL/min flow rate and a step-gradient is not recognized as such. But the chromatographic separation would fail in the case of a nano-HPLC with typical flow rates of 200 nL/min. For this reason, the detection of dead volumes that cannot be visually located is a very important task for the highest performance and sensitivity. The CARLA Test provides a very simple approach for online detection of dead volumes in HPLC systems [39]. For the test, a UV detector, a precolumn system with a loading pump using 0.1% TFA and the nano-HPLC-system using 0.1% FA, is required.

While the sample is loaded onto the trapping column and desalted by flushing with 0.1% TFA at an increased flow rate, the nano-HPLC branch is equilibrated with 0.1% FA. The UV baseline is set on the FA containing solution. By switching the HPLC valve, the trapping-column loop is then integrated into the nano-flow. Now, the nano-flow pushes the small volume of TFA solution through the

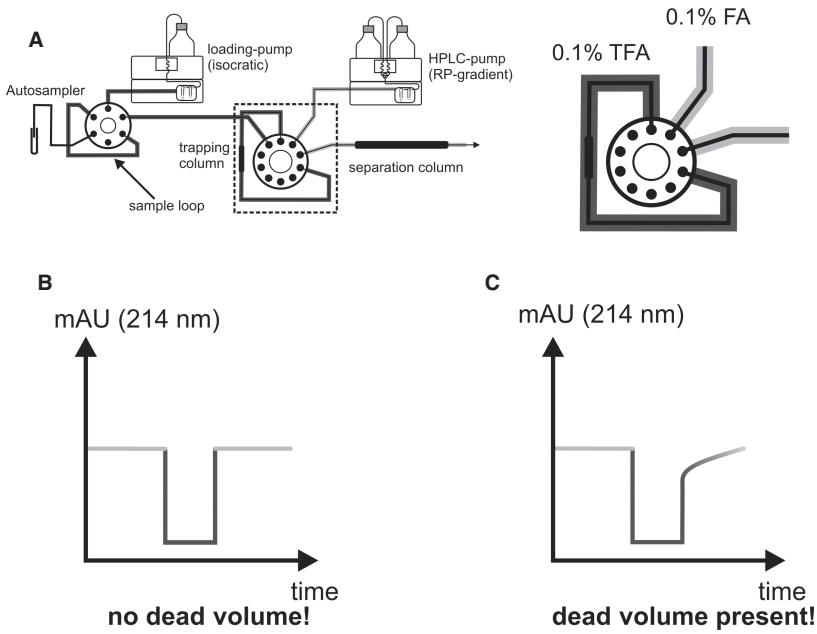


**Figure 3.** Carla test. (A) The Carla test requires the detection of the UV-trace at 214 nm and a precolumn setup with loading pump that delivers 0.1% TFA for peptide trapping and desalting and a nano-flow system working with 0.1% FA for peptide separation. The close-up view of the HPLC valve depicts the localization of TFA and formic acid. (B) Schematic UV-chromatogram after the switching event. The baseline is set to UV-absorption level of formic acid. The elution of TFA from the precolumn loop decreases the UV-signal to a significant level. If no dead volume is present, the signal has a rectangular shape. (C) If a dead volume acts as a small mixing chamber, the discrete volumes of TFA and FA are partially mixed, leading to long transitions back to baseline. In this case, eluted peptide peaks are broader and sensitivity level of such a HPLC-system is decreased.

trapping-column loop and the separation column to the UV detector that is reported by a signal drop. If no mixing occurs, the pattern should be shown as a sharp rectangular signal (see Figure 3B). However, if a dead volume is present, the two different solutions are mixed during passage through the flow path (see Figure 3C). This results in a smooth transition back to the FA baseline after an increased amount of time.

### 3.3 Column Media and Resins

The performance of HPLC separations depends on the quality of the applied column. Column dimensions as well as the used resins are the main features that have to be regarded. Although column systems with beads as chromatographic material have been extensively used, recently available monolithic columns represent a completely new technology.

Bead materials are normally made of silica spheres or synthetics like polystyrenedivinylbenzene (PS-DVB). The beads are then functionalized with the appropriate chemical group for interaction with analytes. In the case of peptides, C18- or

C8-phases are coupled to the silanol groups on the bead-surface with chlorotrialkylsilane reagents. Often, not all available silanol groups are modified and are often blocked by end-capping to avoid ion exchange functionality interfering with the RP mode. PS-DVB materials are hydrophobic in nature and need only surface modification if other separation properties are required. The silica-bead packed columns provide a very good resolution for HPLC, but they are becoming unstable above pH 8. Under basic conditions, polymeric materials are therefore preferable. However, several vendors have developed chemical modifications on silica beads that slightly increase the stability of the material up to pH 9–10.

The pore diameter of the beads has a decisive influence on the binding capacity of the column, which means that small pore sizes increase the effective surface of the material and thereby the binding capacity. Nevertheless, if the pores are too small, the peptides cannot diffuse and therefore the interaction of the analyte and the stationary phase is suppressed. A pore diameter of 100 Å has proven to be very suitable for tryptic peptides, whereas high-molecular-weight peptides should be separated using 300-Å pore material. The bead size and the bead size distribution have a direct influence on the separation efficiency, which is expressed in the van Deemter-equation by the Height Equivalent to the Theoretical Plate (HETP). The smaller the bead size, the smaller the plate height and the higher the resolution. But with smaller bead sizes, the required flow rate has to be increased and therefore the backpressure rises significantly. Typically, analytical RP columns for peptides use bead sizes between 3 and 5 μm. The column diameter and the column length depend mainly on the required resolution and the available sample amount, respectively.

### 3.4 Multidimensional Separation of Peptides

Proteins differ in molecular weight, hydrophobicity, charge, and structure over a wide range, which renders it impossible to apply a single chromatographic separation for complex protein mixtures. Therefore, proteins are digested to peptides to reduce the huge differences in their physico-chemical properties that enables the application of the peptide mixture to chromatographic separation approaches. However, the digest increases the number of analyte species from the protein to the peptide level significantly. For instance, the expected number of proteins in a cell differs from 5 to 20,000 species. After digestion the expected number of peptides ranges between 50,000 and 200,000 species. Such complex samples can only be handled by multiple separation according different separation parameters. Generally, the concept of this approach is the subsequent separation and prefractionation of peptides with different chromatographic media. In most cases, the first dimension represents a strong IEX Chromatography where the peptides are fractionated during fixed time intervals. Afterward, each ion-exchange fraction is applied to an RP column as the final high-resolution separation dimension that is usually online-coupled to an MS representing an additional gas phase separation dimension with concurrent peptide analysis. Several different applications have been developed so far: offline or online MDLC, MudPIT, and the COFRADIC [40] technology.

The classic MDLC technique separates the peptides using a strong cation exchange (SCX) column for the first and C-18 RP material for the second

dimension. Offline and online separation, respectively, differs in peptide transfer from the first to the second HPLC-separation dimension. For the online mode, eluted fractions are directly transferred to the next dimension, whereas during offline MDLC, the collected fractions are prepared for injection onto the RP column. The usage of fraction collectors and liquid handlers enables a semi-automated sample preparation before reinjection for RP separation and subsequent analysis. The advantages of the offline mode are the application of continuous gradients and more flexibility in solvent selection, thereby leading to a more effective elution and better resolved separation (see Figure 4). Additionally, both separation dimensions can be performed subsequently on a single HPLC system with less hardware requirements in contrast to the online system.

Online MDLC works in full automation mode and integrates first- and second-dimension separation into a single system, providing high sample throughput capabilities and short dwell times. The elution of the SCX column is performed with salt-plugs (see Figure 5) or step-gradients (see Figure 6) with increasing salt amounts. Subsequently, the eluted fraction is trapped onto an RP-precolumn where the peptides are desalted and concentrated. Then the HPLC valve is switched, and the precolumn is introduced into the flow path of the RP gradient. Peptides are separated and eluted similar to common RP separations. After the complete RP run, the precolumn is switched back and traps the next fraction eluting from the IEX column. In more complex MDLC systems, two or more precolumns are alternately used for trapping and elution to reduce the complete run time as well as to enable the application of continuous IEX gradients. A general disadvantage of the online system is the restriction to non-organic solvents for IEX elution. It has been shown that little amounts of organic modifiers increase the elution efficiency and prevent memory effects during the first dimension [41]. In online systems, this

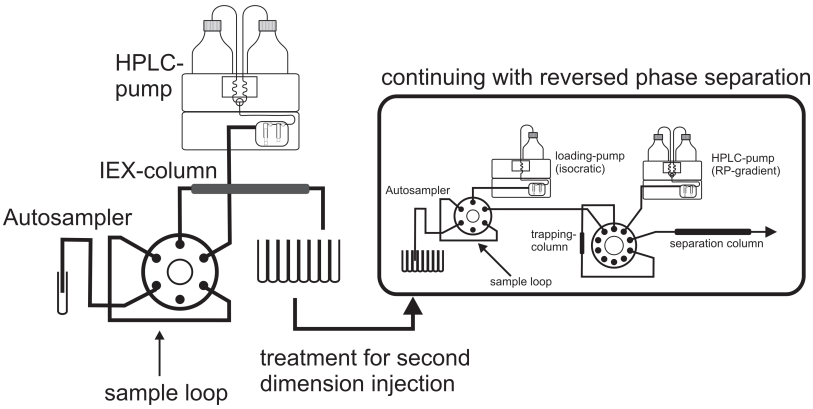


**Figure 4.** Scheme of an offline MDLC setup. The separation with this two-dimensional HPLC separation is divided into two steps. At the first step, the analyte is separated by an ion-exchange separation (IEX) and the eluted fractions are collected. Before the second step, the samples are concentrated and desalted and organic modifiers are removed if present. The second dimension is performed by a normal reversed phase-HPLC separation (see Figure 1). If the required capacities are not available, this type of MDLC run enables the separation of both dimensions on the same HPLC after a setup change of column and solvents.

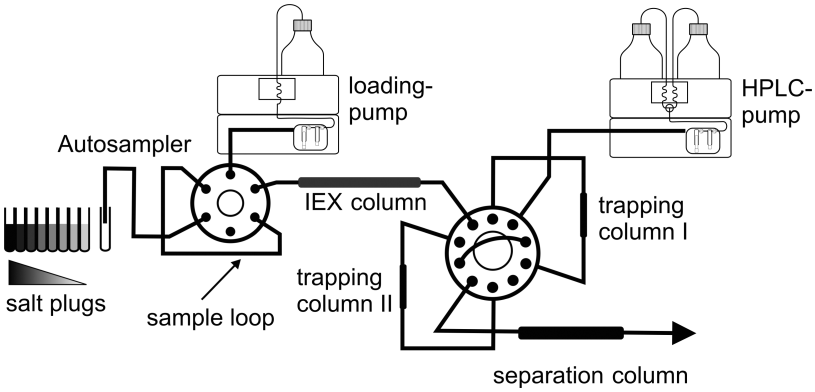


**Figure 5.** Online MDLC using salt-plugs. This HPLC separation mode uses salt-plugs with increasing salt concentrations that are injected by the autosampler onto the IEX-column to elute stepwise stronger bound peptides.

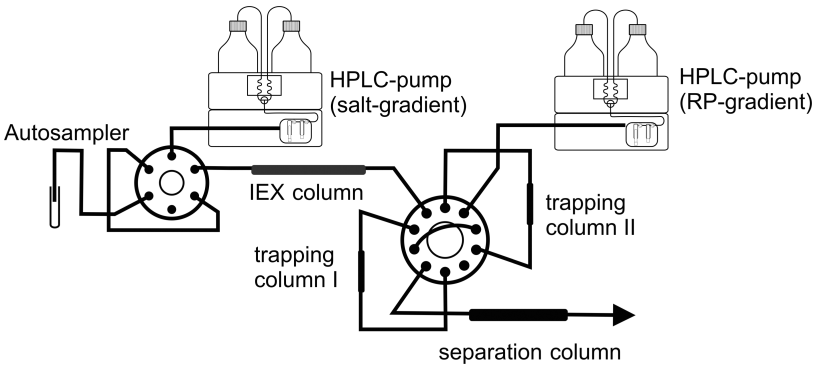


**Figure 6.** Online MDLC with gradient elution. This type of MDLC system is similar to the setup with salt-plug elution, but in this case, the elution is performed with a step- or linear-gradient that requires a second gradient capable HPLC-pump.

would result in inferior peptide recovery because the SCX-peptide fraction would not be trapped quantitatively on the precolumn due to the organic solvent.

This type of elution has the further disadvantage that peptides are present in different fractions due to memory effects. Additionally, the online systems consist of multiple HPLC pumps and valves, several connections, and a complex control program that gives rise to many potential sources of error as well as to difficulties in obtaining acceptable sensitivity. The multidimensional protein identification technology (MuDPIT) was designed to provide a high-resolution peptide separation technique for electrospray ionization mass spectrometry (ESI–MS) with a very good sensitivity [42] (see Figure 7). In contrast to the classic MDLC, the two separation phases are located in a single biphasic column. Thereby, the eluted peptides from the SCX phase are directly trapped on the RP resin without any dead volume. Additionally, the system is simplified due to the location of the two phases in the ESI-tip emitter. The sample is bound on the SCX phase, and fractions are eluted by salt-plug injection. The fractions are trapped on the RP phase and then

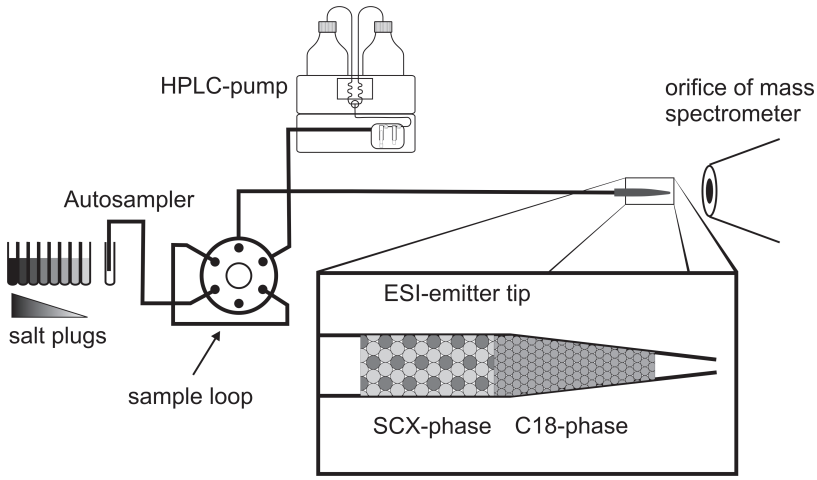


**Figure 7.** Schematic workflow of MuDPIT. The MuDPIT technique represents a relative simple setup compared with the other MDLC approaches. For the complete system, a single nano-gradient pump and an optional autosampler is required. The two-dimensional separation takes place in the biphasic ESI-Tip emitter, which results in a near dead-volume-free solution. Peptides are eluted from the strong cation exchange phase (SCX) with salt-plugs that are injected by the autosampler. Afterward the trapped peptides are separated with a linear gradient and directly sprayed into the mass spectrometer.

separated by common RP gradients that are delivered by a single nano-HPLC pump. Therefore, the technical requirements are reduced in comparison with other online MDLC systems, and the complete separation is performed fully automated. The major disadvantage of this approach is the design of the biphasic separation column, because the two phases are in contact with a high salt concentration and high concentration of organic modifiers, respectively. This may lead to poor sample recovery and imminent loss of peptide species.

One major disadvantage of all presented approaches is the limited resolution in a two-dimensional separation. The discrete separation with tens of thousands of peptide species is still not possible. The combined fractional diagonal chromatography (COFRADIC) circumvents this problem by discarding most peptides [43]. The protein identification by MS requires only a single peptide that is specific for this particular protein. Therefore, the increase of sample complexity after enzymatic digest produces a lot of redundancy that is again reduced by the COFRADIC technique. The proteins are blocked at the cysteine and free amino-residues before proteolytic digestion. The resulting peptides are separated by RP chromatography and collected in minute fractions. Each fraction is then treated with TNBS to modify the newly generated free N-terminal amino groups and increase the hydrophobicity of these peptide species. Afterward, each fraction is rerun under the same conditions. All peptides that elute at the same retention time as before are original N-terminal peptides because they were blocked prior to TNBS-modification and are therefore not susceptible to retention time increase. These fractions can be collected for later MS analysis. The COFRADIC approach has also been extended to the separation of cysteine or methionine containing peptides [44]. Other targets

like phosphorylations or other moieties that can be chemical modified may be targeted as well. Therefore, the COFRADIC approach is very suitable for directed global analysis of specific peptide subsets in complex peptide mixtures.

## 4 SEPARATION OF PROTEINS

### 4.1 Short Introduction

RP–HPLC separates compounds based on the reversible interaction with the hydrophobic stationary phase. RP–HPLC has become the most extensively practiced type of chromatography for proteins due to the unparalleled resolution with applications in purification, desalting, and analysis.

RP–HPLC separation of proteins has been the topic of several reviews and book chapters, e.g., Refs. 45 and 46.

RP chromatography of proteins is almost exclusively dominated by gradient elution. Besides the ability of gradient elution to separate a complex mixture of compounds with a broad range of hydrophobicity, there is another reason for the almost exclusive use of gradient elution.

In contrast to small molecules, proteins are amphiphilic and interact with the stationary phase only via a small part of the molecule, whereas the hydrophilic parts of the protein are in contact with the mobile phase. The mobile phase at the start of the gradient is typically an aqueous solution. During development of the gradient, a protein will almost completely desorb from the stationary phase at a specific organic solvent concentration. Before this concentration, the protein has an almost infinite retention on the stationary phase and will therefore not migrate through the column. This principle, often referred to as "on/off" retention mechanism, is also indicated by the steep adsorption isotherms of proteins. As a result, closely related proteins can be separated by small adjustments in the mobile phase composition. Hence, isocratic elution of proteins is not an attractive technique.

### 4.2 Reversed Phase–Liquid Chromatography (RP–HPLC) Separation of Proteins

Liquid chromatography is an essential analytical tool for the separation of proteins in biotechnology and pharmaceutical product development. Applications are found throughout the development process of biological products, from drug discovery to quality control. Molecular properties of proteins, like molecular mass, net charge, hydrophobicity, and structure give ample opportunities to use HPLC for their separation.

A variety of chromatography techniques is available for separation and analysis of proteins, i.e., RP, IEX, size exclusion, hydrophobic interaction, and affinity chromatography. Selection of the type of chromatography, scale of operation, and applied detection technique will be based on the sample availability and complexity and, on the other hand, the required information.

***Stationary Phases.*** The first step in the optimization of a chromatographic method for proteins is the selection of the stationary phase. Many stationary phases are

available and are characterized by the material, particle shape and size, pore diameter, and functional group. For analytical purposes, regular-shaped, silica-based stationary phase particles are mostly employed. These particles are made of alternating silicon and oxygen atoms that form an incompressible three-dimensional network. The surface of the silica particle is covered with hydroxyl groups, called silanol groups, which are modified with a reactive alkyl silane to yield a siloxane bond. The length of the alkyl chain, which can vary from methyl (C1) to octadecyl (C18), determines the hydrophobicity of the RP sorbent and the binding strength of the protein. Unreacted, residual silanol groups are present after functionalization of the silica bead and have acidic properties that impair the RP separation. Ionic interaction between these silanol groups and the protein result in low protein recovery and peak tailing. To overcome this effect, end-capping of the stationary phase can be performed, i.e., the reaction of residual silanol groups with a small molecule.

Another limitation of silica-based stationary phases is the limited pH stability, which extends from around 3.0 to 7.0. Acidic solutions hydrolyze the siloxane bond between the bonded alkyl silane and the silica particle; at high pH, the silica substrate is readily dissolved. The hydrolytic stability of RP silica particles can be improved by shielding the siloxane bond with bulky alkyl groups or by surface coverage of the silica with a highly cross-linked aromatic layer [47]. Organic polymer-based supports have been developed as an alternative for the inorganic, silica-based stationary phases and have a high pH stability.

Resolution of the protein separation is affected by several stationary phase properties. Reduction of the particle diameter is obviously the most straightforward way to improve the efficiency. For analytical purposes, 3- to 10-μm particles are a good compromise between chromatographic performance and the required pressure. The pore size of the particles should be a minimum of 300 Å to provide accessibility for large proteins.

The large size of the proteins implies that diffusion coefficients are low and that mass transfer between the mobile and the stationary phase is slow. Several stationary phases have been designed that prevent excessive band broadening as a result of the low diffusivity of proteins. Perfusion beads are highly porous particles made from polystyrene divinylbenzene with pore sizes in the range of 100–8000 Å. The structure allows the mobile phase to perfuse through the particles and a convective transport of the molecules, not limited by diffusion. The analysis time in perfusion chromatography can be reduced significantly compared with conventional stationary phases, due to the low flow resistance and independence of flow rate on the resolution [48].

On the contrary, particles without pores enable fast, efficient separation of proteins due to the absence of intraparticle diffusion resistance. The absence of pores reduces the available surface area and thereby the loading capacity. Silica- [49] and polymer-based [50] nonporous particles have been applied for biomolecule separations. Also hybrid particles have been described that have a non/porous core and a 0.25-μm porous layer, composed of colloidal silica particles [51].

The application of monolithic columns for efficient RP–HPLC separation of proteins is rapidly developing. The advantage of monolithic columns for protein separations is the high mass transfer efficiency due to convective flow in the macroporous structure. As a result, very fast separations can be achieved.

Several polymer-based monolithic columns have become available commercially in different formats, e.g., disks, membranes, and LC columns with diameters in the range of 0.1–100 mm ID [52]. The preparation and application of organic polymer-based monoliths have been reviewed recently [20].

Extremely rapid RP separations of proteins have been demonstrated on poly(styrene-co-divinylbenzene) monolithic columns. A fast gradient and a flow rate of 10 mL/min were used to separate five model proteins in less than 20 s on a 50 × 4.6-mm ID column [53].The same type of monolithic structure has also been prepared in capillary columns of 200 μm ID and showed excellent performance in the identification of proteins by LC–MS through peptide mass fingerprinting and accurate intact molecular mass determination [24].

Acrylate monomers are used for the preparation of monolithic columns and are available in a broad variety of chemistries. For the RP type of monolithic columns, acrylates with an alkyl chain are employed.

Poly(butyl methacrylate-co-ethylene dimethacrylate monolithic columns have been prepared by photoinitiated polymerization and used for RP–HPLC of proteins [54]. The flow resistance of these monolithic columns was low, which enabled the use of high flow velocities of 87 mm/s. A mixture of four model proteins ribonuclease A, cytochrome c, myoglobin, and ovalbumin was baseline separated in around one minute and using flow rates of 100 μL/min. In Figure 8, chromatograms are shown that were obtained with three different gradients.

No effect of the pore size in the range of 0.7–2.2 μm was measured on the retention of proteins.

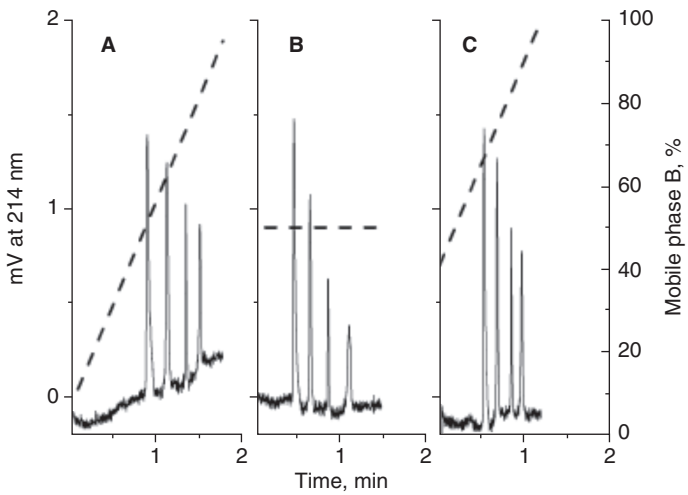


**Figure 8.** Effect of gradient on separation of a protein mixture using monolithic column I at 100 μL/min. Conditions: column size 100 mm × 0.2 mm. Mobile phase A, 0.1% TFA in water–acetonitrile (90:10, v/v); mobile phase B, 0.1% TFA in water–acetonitrile (10:90, v/v). Gradient in (a) 5–95% B from 0 to 1.8 min; gradient in (b) 5–50–50% B from 0 to 0.01 to 1.5 min, gradient in (c) 5–40–100% B from 0 to 0.01 to 1.21 min. Peaks: ribonuclease A, cytochrome *c*, myoglobin, and ovalbumin (elution order). (Figure and caption reprinted from Ref. 54, with permission from Elsevier.)

***Mobile Phases and Gradients.*** Ion-pairing RP–HPLC is an ideal technique for the separation of proteins. A charged ion-pairing reagent with an opposite charge to that of the protein is added to the mobile phase to form a noncharged complex. The ion-pairing reagent increases the hydrophobicity and consequently the retention of the protein. TFA is the most commonly used ion-pairing reagent for proteins. Added in a concentration of 0.05% (v/v) to the mobile phase, the pH will be around 2.5. At this pH, the carboxylic groups of proteins are protonated and the trifluoroacetic anion binds to the protonated, positively charged amino groups of proteins.

Several ion-pairing reagents, including acids, bases, and salts, have been applied in RP liquid chromatography of proteins. TFA is preferred as it generally yields the highest separation performance. However, a disadvantage of TFA is that it can significantly lower the detection sensitivity in EIS–MS detection. The strong ion pairs are not easily broken apart under conditions used for EIS, thereby preventing ionization of the analyte. This effect is not easy predictable and has been found dependent on the ion-pairing reagent concentration, instrumental conditions, and the analyte. The influence of ion-pairing reagents on the sensitivity of detection in the coupling of RP–HPLC and ESI–MS for protein analysis has been reviewed [55]. In practice, a compromise between detection sensitivity and chromatographic resolution has to be found in the use of ion-pairing reagents for HPLC–ESI–MS applications. A more than 100-fold increase in detection sensitivity of proteins was reported when using 0.1% formic acid instead of 0.1% TFA at the cost of larger peak widths [56].

Several solvents can be chosen for the development of the gradient in RP chromatography. Acetonitril has a high volatility, low viscosity, and high UV transparency and is often preferred in LC/MS applications. For the analysis of large, hydrophobic proteins, alcohols with higher elution strength such as isopropanol can be employed. The gradient must be optimized for each application. Steep gradients are useful for protein desalting, whereas a shallow gradient should be used for complex samples.

***Applications in Biotechnology.*** Biopharmaceutical products must be well characterized by means of analytical techniques. RP–HPLC with UV detection is suitable for the content assay and impurity profiling because of its high resolution. However, the bioactivity of protein-based pharmaceuticals not only relies on the primary structure but also on the higher order structure, i.e., the molecular confirmation, which is not revealed by UV absorbance. Circular dichroism and intrinsic fluorescence detection have been combined with UV detection in RP–HPLC to monitor the conformational properties of active compounds and impurities in biopharmaceutical products [57].

Characterization of (recombinant) proteins at the molecular level requires the use of MS detection.

RP–HPLC–MS can be applied to separate and identify proteins modifications such as deamidation, isomerization [58], and protein folding [59]. Even intact monoclonal IgG1 antibodies have been characterized by RP–HPLC–MS using a C8 stationary phase, elevated temperature of 65–70°C, and a combination of trifluoroacetic acid and heptafluorobutyric acid as ion-pairing agents.

Characterization of C-terminal lysine variants, glycosylation heterogeneity and degradation products of recombinant antibodies demonstrates the capability of this method [60].

The development of MS instrumentation has been accelerated during the last decade by proteomics research, the large-scale identification of the proteins in biological samples or tissues. Although nowadays the precise mass determination of even large proteins is possible on standard MS instruments, identification of proteins is still done at the peptide level after enzymatic digestion of the protein. An interesting development is internal gas phase protein fragmentation in the MS enabling accurate protein identification. In combination with a chromatographic separation, this development holds great promise for top-down proteomics [61].

### 4.3 Ion-exchange Separation of Proteins

In ion-exchange chromatography (IEC), retention is based on the interaction of charged amino acid residues of the protein with the counter-charge on the stationary phase media. In addition, protein modifications such as acetylation, phosphorylation, and glycosylation alter the charge of the protein and therefore the retention on the IEX column. IEX stationary phases carry either a positive charge (anion exchanger) or a negative charge (cation exchanger) and can be subdivided into weak and strong ion-exchanger resins. Strong ion-exchangers are ionized over a much wider pH range compared with their weak ion-exchanger and can be used over a broad pH range (~pH 2–12) Weak IEX columns lose their charge at pH > 9 (anion) and <6 (cation). The terms *strong* and *weak* do not refer to the strength of interaction; the interaction of the protein with the column is determined by the stationary phase, the mobile phase composition, and the protein. Elution of the protein is accomplished by competition with small counter-ions for the adsorption sites on the ion exchanger. Most often, linear salt gradients are used. Other means for protein elution are a pH gradient or a combined salt/pH gradient.

Method development is started with the selection of the type of ion-exchange resin and a buffer with appropriate pH to assure retention of the proteins to be separated. IEC separations are usually started with a low buffer and salt concentration with sufficient buffer capacity.

An advantage of IEC over RP–HPLC is that chromatographic conditions are very mild, thereby preserving the protein structure and bioactivity. Protein purification by means of IEC generally gives high yields and is an excellent technique for large-scale purification of pharmaceutical proteins in biotechnology.

***Monolithic Columns for Ion-Exchange Chromatography of Proteins.*** Ionizable groups can be incorporated into monolithic structures to make them suitable for IEC in different ways. Starting with functional monomers has the advantage that the monolithic structure can be prepared in a single step. However, the process requires optimization for each new set of functional monomers and cross-linkers to obtain the required properties. Alternatively, the ionizable groups are introduced after the monolithic structure has been created in an additional step through reaction with the epoxide groups [62].

The epoxide groups of a polyglycidyl methacrylate-co-ethylene dimethacrylate monolith readily react with diethylamine to create a weak anion exchange column. This monolithic column is characterized by fast mass transport kinetics leading to excellent resolution. The columns seem to be stable, and the specific capacity is 40-mg/g medium for ovalbumin [60, 62]. More recently, graft polymerization within

the large pores of the monoliths has been used to attach chains of reactive polymer. The advantage of the grafting procedure is the high density of surface groups and consequently high binding capacities that can be achieved, which are favorable for the separation of biopolymers [20].

***Application in Biotechnology.*** A high-performance anion-exchange chromatography method was used for the determination of erythropoietin (EPO) in pharmaceutical products [63]. EPO is a glycoprotein with a molecular mass of 30–34 kDa. Besides being the main factor regulating red blood cell production, it is known for extensive misuse as a performance enhancer in endurance sports. The content of EPO in pharmaceutical preparations is currently measured by complex *in vivo* potency assays.

It was shown that the chromatography method was suitable to separate recombinant EPO from amounts of human serum albumin commonly present as a stabilizer in various pharmaceutical preparations. In addition, it was possible to obtain different elution profiles for EPO products with variations in the glycoforms. Fluorescence detection was applied for quantification and showed linear signals over the range of 10–200-μg/mL EPO.

### 4.4 Multidimensional Liquid Chromatography of Proteins

In proteomics the sample complexity is often high and requires high-resolution analytical separation techniques. Two-dimensional polyacrylamide gel electrophoresis (2D PAGE) fulfills this requirement and has been established as a core technique for complex protein mixtures. However, the small dynamic range and the inability to resolve very small and large proteins and proteins with extreme pI values have been serious limitations.

Multidimensional liquid chromatography is becoming an attractive alternative to gel-based separations and has been reviewed recently [64]. The power of multidimensional techniques is the ability to increase the peak capacity significantly by adding selectivity to the system. From a theoretical point of view, the peak capacity of a truly orthogonal 2D-HPLC system, i.e., the nature of the separation principle of both dimensions are independent, is given by multiplication of the peak capacity of each dimension [65]. If the peak capacity of both columns is 50, the total peak capacity for the 2D-LC system will be 2500. However in practice, the resolving power of multidimensional LC techniques is impaired by limited orthogonality and the slowness of the second dimension separation.

A wide variety of separation mechanisms have been combined in 2D-HPLC to separate complex protein mixtures. Size exclusion [66], affinity [67], IEX [68–70], and chromatofocusing [71] have been used as a first-dimension separation in combination with RP as a second-dimension separation. IEX–RP, the most extensively studied combination, has been coupled to EIS time-of-flight MS for separation and identification of ribosomal proteins from yeast [72]. Approximately 70% of the potential ribosomal subunits isoforms could be identified with an average mass error of ~50 ppm. In the experimental setup, two parallel RP columns were alternately switched in series with a strong anion-exchange column to reduce the analysis time.

Distribution of proteins over multiple second-dimension fractions is unwanted and influenced by several aspects. Peak splitting is unavoidable and depends on the

elution profile and fraction size. In general, peak splitting can be worse for proteins with a molecular heterogeneity as a result of different modification forms or structural conformations. Also carryover or peak "ghosting" results in dispersion into multiple fractions and is generally more profound for high abundant and hydrophobic proteins.

The first dimension separation can be run with linear and step gradients.

An extensive comparison between these two gradient modes in IEX–RP separation of cytosolic proteins from *Escherichia coli* revealed a more effective fractionation for a linear salt gradient [70]. Fewer major proteins were distributed to multiple second-dimension fractions.

In the second RP dimension, gradient elution is required to elute proteins with a wide range in hydrophobicity. As gradient elution is typically a slow process, several approaches have been proposed to speed up the second-dimension separation, such as the use of short columns, high flow rates, and a parallel column setup [68].

Polystyrene–divinylbenzene monolithic columns have shown excellent separation performances for intact proteins [24]. A capillary PS–DVB monolithic column with 500 μm ID has been used in combination with IEX columns for offline 2D-LC of complex proteins samples. The power of this technique is illustrated with chromatograms of the separation of soluble proteins from *Salmonella typhimurium* in Figure 9. The ion-exchange separation has been developed on a strong cation- and anion-exchange column, coupled in series. The combination of these two IEX columns allow retention of both acidic and basic proteins. The resolution of the multidimensional separation is high. The PS–DVB monolithic column and IEX column have peak capacities of, respectively, 80 and 60. Giving the fact that 30 fractions were taken in the first dimension, the total peak capacity can be estimated as 2400 (80 × 30 fractions). The offline column configuration makes the method flexible; i.e., both dimensions can be optimized independently with respect to mobile phases, column dimensions, and sample loading.

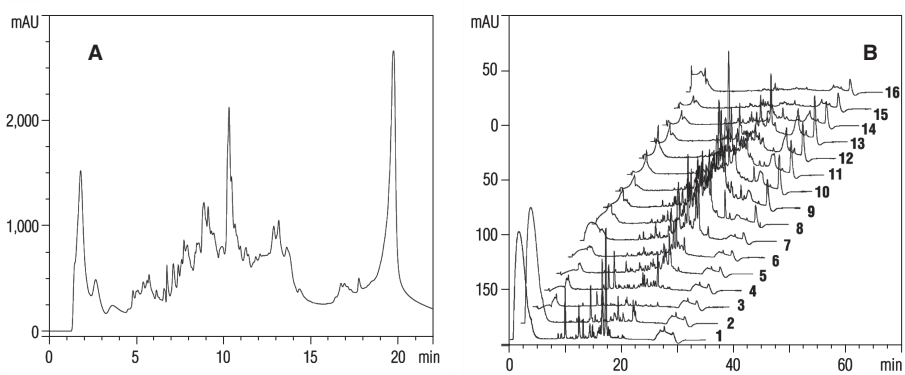


**Figure 9.** Offline 2D-LC of intact *S. typhimurium* proteins. (A) Separation of *S. typhimurium* protein extract on a ion-exchange ProPac SAX column. (B) Overview of the RP separations of Salmonella proteins on PS–DVB monolithic column 500 μm i.d. × 5 cm.

***SEC.*** Another liquid chromatographic technique applied in biopharmaceutical characterization is SEC. In this technique, compounds are separated based on size and shape. In contrast to RP–HPLC, SEC has a low resolving power that has limited the application of the technique. The main application of SEC in pharmaceutical analysis is the determination of the native size of a protein and the quantification of protein aggregates. Typically protein-friendly conditions are employed in SEC that are expected not to affect the higher order structure of the protein.

## 5 SEPARATION OF AMINO ACIDS—AMINO ACID ANALYSIS

Amino acid analysis has become common in biotechnology, biomedical, and food analysis laboratories. In proteomics laboratories, this technique becomes now more important because quantitative proteomics requires knowledge about the exact sample amounts used in various kinds of experiments. Amino acid analysis is an exact, fast, and quantitative method for determining free and/or protein bound amino acids. Especially compared with Bradfort or Biuret, assays amino acid analysis is a more reliable technique with about an order-of-magnitude lower standard deviation. It is used to control protein yields from various protein preparations and to determine sample amounts prior to 2D-PAGE and 1D- and 2D-HPLC. Furthermore, it is necessary to measure exact protein amounts before stable isotope labeling with reagents like iTRAQ, ICPL, or ICAT [73–75]. If the ratio between chemical reagent and reactive side chain is chosen in the wrong way, a quantitative incorporation cannot be obtained and will lead to false results. Using relative quantification, the same protein amounts should be used. In the case of comparing sample A and B, the standard deviation will increase depending on the ratio of A to B and additional computation and data analysis are required to do normalization on the results.

Using the AQUA technology for absolute quantification of peptides from complex mixtures, it is mandatory to quantify the used stable isotope-labeled peptides before analysis, because peptides in solution undergo unspecific absorption to surfaces and will therefore alter their concentration [76]. However, even dried stable isotope-labeled peptides usually contain an amount of TFA salts from purification.

The method of choice to overcome these problems with the chemical quantification of proteins, peptides, or free amino acids are the classic amino acid analysis methods.

Amino acid analysis is an old method in protein chemistry (the first experiments were already described in 1820 by Braconnot). In 1972, Moore and Stein were awarded the Nobel Prize for establishing automated amino acid analysis. Roughly only two major methods for amino acid analysis exist: precolum derivatization and postcolumn derivatization. Today the sensitivity of amino acid analysis is at the femtomole level for fluorescence detection and at the picomole level for UV light detection methods. An overview about commonly used methods and reagents for amino acid analysis is shown in Table 2.

**TABLE 2. Commonly Used Reagents for Amino Acid Analysis**

| Reagent | Method | Detection |
|---|---|---|
| Ninhydrin | Postcolumn derivatization | λ = 570 nm (prim. AA)<br>λ = 440 nm (sec. AA) |
| Phenylisothiocyanate | Precolumn derivatization | λ = 245 nm |
| Orthophthaldialdehyde | Precolumn derivatization<br>Postcolumn derivatization | λ = 338 nm<br>λex = 230 nm, 335 nm<br>λem = 455 nm |
| Fluorescamine | Postcolumn derivatization | λex = 390 nm<br>λem = 475 nm |
| Fluorenylmethylchloroformate (FMOC) | Precolumn derivatization | λ = 260 nm<br>λex = 266 nm<br>λem = 305 nm |
| 6-Aminoquinolyl-N-Hydroxy succinimidyl Carbamate (AQC) | Precolumn derivatization | λex = 248 nm<br>λem = 395 nm |
| Dabsylchloride | Precolumn derivatization | λ = 436 nm |
| Dansylchloride | Precolumn derivatization | λex = 310 nm<br>λem = 540 nm |

*Abbreviations:* λex = exitation wavelength; λem = emission wavelength.

## REFERENCES


1. Churchwell M I, Twaddle N C, Meeker L R, et al. (2005). Improving LC-MS sensitivity through increases in chromatographic performance: Comparisons of UPLC-ES/MS/MS to HPLC-ES/MS/MS. *J. Chromatogr. B Analyt. Technol. Biomed. Life Sci.* 825(2):134–143.
2. Sherma J (2005). UPLC: Ultra-performance liquid chromatography. *J. AOAC Int.* 88(3):63A–67A.
3. Wren S A (2005). Peak capacity in gradient ultra performance liquid chromatography (UPLC). *J. Pharm. Biomed. Anal.* 38(2):337–343.
4. Exsigent Available: www.eksigent.com.
5. Proxeon Available: www.proxeon.com/proxeon/scienceEasyLC.do.
6. Tuytten R, Lemiere F, Witters E, Van Dongen W, Slegers H, Newton R P, Van Onckelen H, Esmans E L (2006). Stainless steel electrospray probe: A dead end for phosphorylated organic compounds? *J. Chromatography A*. 1104:209–221.
7. Pretorius V, Hopkins B J, Schieke, J D (1974). *J. Chromatogr.* 99:23.
8. Jorgenson J W, Lukacs K D (1981). *J. Chromatogr.* 218:209.
9. Eeltink S, Rozing G P, Kok W T (2003). Recent applications in capillary electrochromatography. *Electrophoresis*. 24(22–23):3935–3961.
10. Knox J H (1980). *J. Separation Sci*. 18:453.
11. Eeltink S, Rozing G P, Schoenmakers P J, et al. (2004). Study of the influence of the aspect ratio on efficiency, flow resistance and retention factors of packed capillary columns in pressure- and electrically-driven liquid chromatography. *J. Chromatogr. A*. 1044(1–2):311–316.

12. Meyer V R (2004). *Practical High-Performance Liquid Chromatography,* 4th ed. Wiley-VCH, Weinheim, Germany, p. 374.
13. Patel K D, Jerkovich A D, Link J C, et al. (2004). In-depth characterization of slurry packed capillary columns with 1.0-microm nonporous particles using reversed-phase isocratic ultrahigh-pressure liquid chromatography. *Anal. Chem.* 76(19):5777–5786.
14. Mellors J S, Jorgenson J W (2004). Use of 1.5-microm porous ethyl-bridged hybrid particles as a stationary-phase support for reversed-phase ultrahigh-pressure liquid chromatography. *Anal. Chem.* 76(18):5441–5450.
15. Jerkovich A D, Mellors J S, Thompson J W, et al. (2005). Linear velocity surge caused by mobile-phase compression as a source of band broadening in isocratic ultrahigh-pressure liquid chromatography. *Anal. Chem.* 77(19):6292–6299.
16. Novakova L, Matysova L, Solich P (2005). Advantages of application of UPLC in pharmaceutical analysis. *Talanta.* 68(3):908–918.
17. Plumb R, Castro-Perez J, Granger J, et al. (2004). Ultra-performance liquid chromatography coupled to quadrupole-orthogonal time-of-flight mass spectrometry. *Rapid Commun. Mass Spectrom.* 18(19):2331–2337.
18. Svec F, Frechet J M J (1992). Continuous Rods of Macroporous Polymer as High-Performance Liquid Chromatography Separation Media. *Anal. Chem.* 64: 820–822.
19. Ikegami T, Tanaka N (2004). Monolithic columns for high-efficiency HPLC separations. *Curr. Opin. Chem. Biol.* 8(5):527–533.
20. Svec F (2004). Preparation and HPLC applications of rigid macroporous organic polymer monoliths. *J. Sep. Sci.* 27(10–11):747–766.
21. Hjerten S, Liao J L, Zhang R (1989). *J. Chromatogr.* 473:273.
22. Rohr T, Hilder E F, Donovan J J, et al. (2003). Photografting and the control of surface chemistry in three-dimensional porous polymer monoliths. *Macromol.* 36:1677–1684.
23. Minakuchi H, Nakanishi K, Soga N, et al. (1996). Octadecylsilylated porous silica rods as separation media for reversed-phase liquid chromatography. *Anal. Chem.* 68: 3498–3501.
24. Walcher W, Oberacher H, Troiani S, et al. (2002). Monolithic capillary columns for liquid chromatography-electrospray ionization mass spectrometry in proteomic and genomic research. *J. Chromatogr. B Analyt. Technol. Biomed. Life Sci.* 782(1–2): 111–125.
25. Wienkoop S, Glinski M, Tanaka N, et al. (2004). Linking protein fractionation with multidimensional monolithic reversed-phase peptide chromatography/mass spectrometry enhances protein identification from complex mixtures even in the presence of abundant proteins. *Rapid Commun. Mass Spectrom.* 18(6):643–650.
26. Ishii D (1988). *Introduction to Mircoscale High-performance Liquid Chromatography.* VCH, Weinheim, Germany.
27. Scott R P W, Kucera P (1976). *J. Chromatogr.* 125:251.
28. Novotny M (1981). Microcolumns in liquid chromatography. *Anal. Chem.* 53(12): 1294A–1308A.
29. Ishii D, Asai K, Hibi K, et al. (1977). *J. Chromatogr.* 144:157.
30. Henzel W J, Bourell J H, Stults, J T (1990). Analysis of protein digests by capillary high-performance liquid chromatography and on-line fast atom bombardment mass spectrometry. *Anal. Biochem.* 187(2):228–233.
31. Moritz R L, Simpson R J (1992). Application of capillary reversed-phase high-performance liquid chromatography to high-sensitivity protein sequence analysis. *J. Chromatogr.* 599(1–2):119–130.

32. Griffin P R, Coffman J A, Hood L E, et al. (1991). Structural analysis of proteins by capillary HPLC electrospray tandem mass spectrometry. *Int. J. Mass Spectrom.* 111:131–149.
33. Rapp E, Tallarek U (2003). Liquid flow in capillary (electro)chromatography: Generation and control of micro- and nanoliter volumes. *J. Sep. Sci.* 26:453–470.
34. Chervet J P, Ursem M (1996). Instrumental requirements for nanoscale liquid chromatography. *Anal. Chem.* 68:1507–1512.
35. Meiring H D, van der Heeft E, ten Hove G J, et al. (2002). Nanoscale LC–MS(n): technical design and applications to peptide and protein analysis. *J. Sep. Sci.* 25: 557–568.
36. Licklider L J, Thoreen C C, Peng J, et al. (2002). Automation of nanoscale microcapillary liquid chromatography-tandem mass spectrometry with a vented column. *Anal. Chem.* 74(13):3076–3083.
37. Hearn M T W (1985). Ion-pair chromatography of amino acids, peptides, and proteins. In MTW Hearn (ed.), Ion-pair chromatography. *Theory and Biological and Pharmaceutical Applications.* Marcel Dekker, New York, p. 207.
38. Winkler G, Briza P, Kunz C (1986). Spectral properties of some ion-pairing reagents commonly used in reversed-phase high-performance liquid chromatography of proteins and peptides in acetonitrile gradient systems. *J. Chromatogr. A.* 361:191–198.
39. Mitulovic G, Smoluch M, Chervet J P, et al. (2003). An improved method for tracking and reducing the void volume in nano HPLC-MS with micro trapping columns. *Anal. Bioanal. Chem.* 376(7):946–951.
40. Gevaert K, Van Damme J, Goethals M, et al. (2002). Chromatographic isolation of methionine-containing peptides for gel-free proteome analysis: Identification of more than 800 Escherichia coli proteins. *Mol. Cell Proteomics.* 1(11):896–903.
41. Burke T W, Mant C T, Black J A, et al. (1989). Strong cation-exchange high-performance liquid chromatography of peptides. Effect of non-specific hydrophobic interactions and linearization of peptide retention behaviour. *J. Chromatogr.* 476:377–389.
42. Wolters D A, Washburn M P, Yates J R, III (2001). An automated multidimensional protein identification technology for shotgun proteomics. *Anal. Chem.* 73(23): 5683–5690.
43. Gevaert K, Goethals M, Martens L, et al. (2003). Exploring proteomes and analyzing protein processing by mass spectrometric identification of sorted N-terminal peptides. *Nat. Biotechnol.* 21(5):566–569.
44. Martens L, Van Damme P, Van Damme J, et al. (2005). The human platelet proteome mapped by peptide-centric proteomics: a functional protein profile. *Proteomics.* 5(12):3193–5204.
45. Kastner M (2000). *Protein liquid chromatography*, Vol. 61. Elsevier, Berlin, Germany, p. 976.
46. Aguilar M-I (2004). *Hplc of Peptides and Proteins: Methods and Protocols*, Vol. 251. Humana Press, Totowa, NJ.
47. Yang X, Ma L, Carr P W (2005). High temperature fast chromatography of proteins using a silica-based stationary phase with greatly enhanced low pH stability. *J. Chromatogr. A.* 24(1–2):213–220.
48. Afeyan N B, Gordon N F, Mazsaroff I, et al. (1990). Flow-through particles for the high-performance liquid chromatographic separation of biomolecules: Perfusion chromatography. *J. Chromatogr.* 519(1):1–29.
49. Janzen R, Unger K K, Giesche H, et al. (1987). Evaluation of advanced silica packings for the separation of biopolymers by high-performance liquid chromatography, IV. Mobile phase and surface-mediated effects on recovery of native proteins in gradient

elution on non-porous, monodisperse 1.5-microns reversed-phase silicas. *J. Chromatogr.* 397:81–89.

50. Maa Y F, Horvath C (1988). Rapid analysis of proteins and peptides by reversed-phase chromatography with polymeric micropellicular sorbents. *J. Chromatogr.* 445(1):71–86.
51. Kirkland J J, Truszkowski F A, Dilks C H Jr, et al. (2000). Superficially porous silica microspheres for fast high-performance liquid chromatography of macromolecules. *J. Chromatogr. A*. 890(1):3–13.
52. Jungbauer A, Hahn R (2004). Monoliths for fast bioseparation and bioconversion and their applications in biotechnology. *J. Sep. Sci.* 27(10–11):767–778.
53. Xie S, Allington R W, Svec F, et al. (1999). Rapid reversed-phase separation of proteins and peptides using optimized "moulded" monolithic poly(styrene-co-divinylbenzene) columns. *J. Chromatogr. A*. 865(1–2):169–174.
54. Lee D, Svec F, Frechet J M (2004). Photopolymerized monolithic capillary columns for rapid micro high-performance liquid chromatographic separation of proteins. *J. Chromatogr. A*. 1051(1–2):53–60.
55. Garcia M C (2005). The effect of the mobile phase additives on sensitivity in the analysis of peptides and proteins by high-performance liquid chromatography-electrospray mass spectrometry. *J. Chromatogr. B. Analyt. Technol. Biomed. Life Sci.* 825(2):111–123.
56. Huber C G, Premstaller A (1999). Evaluation of volatile eluents and electrolytes for high-performance liquid chromatography-electrospray ionization mass spectrometry and capillary electrophoresis-electrospray ionization mass spectrometry of proteins. I. Liquid chromatography. *J. Chromatogr. A*. 849(1):161–173.
57. Luykx D M, Goerdayal S S, Dingemanse P J, et al. (2005). HPLC and tandem detection to monitor conformational properties of biopharmaceuticals. *J. Chromatogr. B. Analyt. Technol. Biomed. Life Sci.* 821(1):45–52.
58. Zhang W, Czupryn J M, Boyle P T, Jr, et al. (2002). Characterization of asparagine deamidation and aspartate isomerization in recombinant human interleukin-11. *Pharm. Res.* 19(8):1223–1231.
59. Happersberger H P, Cowgill C, Glocker M O (2002). Structural characterization of monomeric folding intermediates of recombinant human macrophage-colony stimulating factor b (rhM-CSFb) by chemical trapping, chromatographic separation and mass spectrometric peptide mapping. *J. Chromatogr. B. Analyt. Technol. Biomed. Life. Sci.* 782(1–2):393–404.
60. Dillon T M, Bondarenko P V, Speed Ricci M (2004). Development of an analytical reversed-phase high-performance liquid chromatography-electrospray ionization mass spectrometry method for characterization of recombinant antibodies. *J. Chromatogr. A*. 1053(1–2):299–305.
61. Zhai H, Han X, Breuker K, et al. (2005). Consecutive ion activation for top down mass spectrometry: improved protein sequencing by nozzle-skimmer dissociation. *Anal. Chem.* 77(18):5777–5784.
62. Svec F, Frechet J M (1995). Modified poly(glycidyl methacrylate-co-ethylene dimethacrylate) continuous rod columns for preparative-scale ion-exchange chromatography of proteins. *J. Chromatogr. A*. 702(1–2):89–95.
63. Luykx D M, Dingemanse P J, Goerdayal S S, et al. (2005). High-performance anion-exchange chromatography combined with intrinsic fluorescence detection to determine erythropoietin in pharmaceutical products. *J. Chromatogr. A*. 1078(1–2):113–119.
64. Wang H, Hanash S (2003). Multi-dimensional liquid phase based separations in proteomics. *J. Chromatogr. B. Analyt. Technol. Biomed. Life Sci.* 787(1):11–18.

65. Giddings J C (1990). *The use of multiple dimensions in analytical separations.* In H J Cortes (ed.), *Multidimensional Chromatography: Techniques and Applications*, Marcel Dekker, New York.
66. Opiteck G J, Ramirez S M, Jorgenson J W, et al. (1998). Comprehensive two-dimensional high-performance liquid chromatography for the isolation of overexpressed proteins and proteome mapping. *Anal. Biochem.* 258(2):349–361.
67. le Coutre J, Whitelegge J P, Gross A, et al. (2000). Proteomics on full-length membrane proteins using mass spectrometry. *Biochem.* 39(15):4237–4342.
68. Wagner K, Racaityte K, Unger K K, et al. (2000). Protein mapping by two-dimensional high performance liquid chromatography. *J. Chromatogr. A.* 893(2):293–305.
69. Opiteck G J, Lewis K C, Jorgenson J W, et al. (1997). Comprehensive on-line LC/LC/MS of proteins. *Anal. Chem.* 69(8):1518–1524.
70. Millea K M, Kass I J, Cohen S A, et al. (2005). Evaluation of multidimensional (ion-exchange/reversed-phase) protein separations using linear and step gradients in the first dimension. *J. Chromatogr. A.* 1079(1–2):287–298.
71. Wall D B, Kachman M T, Gong S, et al. (2000). Isoelectric focusing nonporous RP HPLC: a two-dimensional liquid-phase separation method for mapping of cellular proteins with identification using MALDI-TOF mass spectrometry. *Anal. Chem.* 72(6):1099–1111.
72. Liu H, Berger S J, Chakraborty A B, et al. (2002). Multidimensional chromatography coupled to electrospray ionization time-of-flight mass spectrometry as an alternative to two-dimensional gels for the identification and analysis of complex mixtures of intact proteins. *J. Chromatogr. B Analyt. Technol. Biomed. Life Sci.* 782(1–2):267–289.
73. Ross P L, Huang Y N, Marchese J N, et al. (2004). Multiplexed protein quantitation in saccharomyces cerevisiae using amine-reactive isobaric tagging reagents. *Mol. Cell Proteomics.* 3(12):1154–1169.
74. Schmidt A, Kellermann J, Lottspeich F (2004). A novel strategy for quantitative proteomics using isotope-coded protein labels. *Proteomics.* 5(1):4–15.
75. Gygi S P, Rist B, Gerber S A, et al. (1999). Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat. Biotechnol.* 17(10):994–999.
76. Gerber S A, Rush J, Stemman O, et al. (2003). Absolute quantification of proteins and phosphoproteins from cell lysates by tandem MS. *Proc. Natl. Acad. Sci. USA.* 100(12): 6940–6945.

# BIOANALYTICAL METHOD VALIDATION FOR MACROMOLECULES

Gerhard K. E. Scriba
*Friedrich-Schiller-University Jena, School of Pharmacy, Jena, Germany*

**Chapter Contents**

1 Introduction
2 Method Validation
  2.1 Ligand Binding Assays
  2.2 Validation of Cell-based Bioassays
3 Conclusions and Future Considerations
4 Glossary of Terms
  References

## 1 INTRODUCTION

Before implementation of analytical methods for routine use, careful validation is required to demonstrate that the method is suitable for the intended purpose. Analytical methods employed for the quantitative determination of drug substances and their metabolites in biological media play a significant role in the evaluation and interpretation of pharmacokinetic data. To define the requirements and procedures for the validation of bioanalytical methods, a conference was held in 1990 in Crystal City in the Washington, DC, area, which was co-sponsored by the U.S. Food and Drug Administration (FDA), The Canadian Health Protection Branch, the American Association of Pharmaceutical Scientists (AAPS), and the Association of Official Analytical Chemists (AOAC), bringing together scientists from

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

regulatory authorities, industry, and academia [1]. Upon evaluation of the results of the first meeting, another conference was held 10 years later [2] that ultimately led to the publication of the FDA guideline "Guidance for Industry: Bioanalytical Method Validation" [3] complementing the general guidelines on method validation such as guidelines Q2A and Q2B by the International Conference on Harmonization (ICH) [4, 5] or pharmacopeial regulations [6]. From the beginning, it was realized that each analytical technique has its own characteristics, which will vary from analyte to analyte. Despite the fact that some similarities exist, a general difference between chemical methods, such as chromatography, and biological assays, such as immunoassays and microbiological assays, was acknowledged [1, 2].

Small, "conventional" drug molecules are preferentially analyzed by chromatographic techniques, specifically by liquid chromatography-mass spectrometry (LC-MS) and liquid chromatography-tandem mass spectrometry (LC-MS/MS), and, to date, most emphasis has been on the validation of bioanalytical methods for such molecules, which is also reflected in the FDA guideline [3]. However, because of the progress in recombinant DNA technology, the number of protein pharmaceuticals has increased dramatically. From 1996 through December 2005, the FDA has approved 253 so-called "biologics" for 384 indications [7]. Most of these products are proteins. As a result of their high potency and subsequent low applied doses resulting in extremely low concentrations in biological media, chromatographic techniques are not sensitive enough for bioanalysis. Thus, immunoassays are primarily used for the bioanalysis of protein drugs.

This divergence in analytical techniques for small molecules and macromolecules has triggered workshops and conferences on bioanalytical method validation of macromolecules focusing on issues such as quantitative immunoassays for therapeutic proteins, biomarkers, and drug neutralizing antibodies as well as bioassays. The results are documented in several publications of meeting reports [8–15], and these publications currently serve as quasi-guidance as no complete official document by the regulatory authorities of the United States or Europe on bioanalysis method validation of macromolecules exist to date. The FDA guidance on bioanalysis [3] acknowledges differences between the assay formats but does not cover all necessary topics for bioassays. A guideline by the European Agency for the Evaluation of Medical Products (EMEA) on pharmacokinetics of pharmaceutical proteins [16] also states specifics of immunoassays and lists points that should be addressed during method validation, but it provides no general guideline. Chromatographic assays are commonly applied to the analysis of protein drugs, but almost exclusively during product characterization release. Method validation of chromatographic techniques is basically identical for small molecules and macromolecules and is addressed in the ICH guidelines [4, 5], FDA guidelines [17], as well as in publications [18] and books [19, 20]. This chapter focuses on validation issues for macromolecule bioanalysis summarizing the current opinion according to the meeting reports [8–15] and further publications on method validation for macromolecules [21–25]. If possible, all terms related to assay validation are used in the sense of the ICH guideline Q2A [4]. A glossary of the most important terms is provided at the end of the chapter. A general guideline for method development of all different kinds of biological assays including validation aspects can be found on the Internet [26].

## 2 METHOD VALIDATION

Relevant macromolecular analytes in biological media can be classified into three categories: (1) pharmaceutical proteins administered as therapeutic agents; (2) biomarkers (i.e., endogenous substances that reflect physiological or pathophysiological processes or pharmacological responses to a therapeutic intervention); and (3) drug neutralizing antibodies that are generated as the response of the human organism to the application of a therapeutic protein. The primary assay formats for these molecules are ligand-binding assays (i.e., immunoassays and receptor-binding assays). Method validation of immunoassays will be the focus of the following discussion whereas cell-based assays will be only briefly addressed. Immunoassays can be roughly divided into competitive assays and sandwich assays, details are discussed in Chapter "Analytical considerations for immunoassays for macromolecules" of this volume. Validation will be discussed for immunoassays for the determination of pharmaceutical proteins in detail and, in subsequent sections, differences with regard to biomarkers and anti-drug antibodies will be mentioned. Not specifically addressed here, but evident in a good laboratory practice (GLP) environment, is the fact that proper documentation and standard operation procedures (SOPs) have to be written.

### 2.1 Ligand-Binding Assays

The term "ligand-binding assay" refers to methods that depend on the specific binding of an analyte to a biomolecule. Generally acknowledged inherent differences exist between bioanalytical chromatographic techniques and ligand-binding assays. The most relevant differences are summarized in Table 1. Whereas chromatography is based on physico-chemical principles, ligand-binding assays are based on a biological response because of the interaction of a ligand with an antibody or a receptor. Consequently, the reagents are derived from living organisms with the attendant variation typical for such reagents (i.e., poor batch-to-batch reproducibility). High-purity, well-characterized reference standards are most

**TABLE 1. Comparison of the Characteristics of Chromatographic Assays and Immunoassays for Bioanalysis**

| Parameter | Chromatographic Assay | Immunoassay |
|---|---|---|
| Principle of measurement | physico-chemical properties of the analytes | antigen-antibody reaction |
| Detection | direct | indirect |
| Reagents, analytical standard | well characterized, high purity, widely available | not always completely characterized, unique, not widely available |
| Analytes | small molecules | small molecules and macromolecules |
| Sample pretreatment | typically pretreatment | usually no pretreatment |
| Calibration model | linear | nonlinear |
| Intermediate precision | high (<10%) | moderate (<20%) |
| Range | broad (several orders of magnitude) | limited |

often not commercially available. Although specificity in chromatography is obtained by isolation of the typically small molecules from the matrix combined with analytical separation from other sample components and probably even detection by mass spectrometry, the isolation of macromolecules from biological media is impractical in most cases because of the low concentrations and the structural or physico-chemical similarity between the analytes and endogenous compounds. Thus, detection of the macromolecule analytes occurs in a complex physiological milieu and, therefore, highly depends on the specificity of the reagents and detection systems. Finally, whereas chromatographic assays display linear relationships between analyte concentration and detector response over a broad range covering 2–3 orders of magnitude, calibration curves of ligand-binding assays are typically nonlinear covering a rather limited range requiring dilution of very concentrated samples. On the time scale, the development of an immunoassay will be longer compared with high-performance liquid chromatography (HPLC) methods because of the need of the generation of antibodies. Once these antibodies have been obtained, the time frame for method validation between the assays is comparable. Sample throughput of immunoassays is excellent.

***Standard Immunoassays.*** Validation is a continuing process through the whole life cycle of an assay. After selection of the assay format, preliminary data are obtained during method development, which are confirmed during prestudy validation and consequently applied during in-study validation (Table 2). This scenario is considered a "full validation." Partial validation is conducted when changes of a fully validated method occur that are considered minor, such as changes in the anticoagulant or changes in the used reagents. Partial validation can range from a single intra-assay accuracy and precision determination to nearly full validation. Method transfer from the developing laboratory to another laboratory or a production site requires at least partial validation. Cross-validation is conducted when two or more validated bioanalytical methods are used within the same study. Test samples (spiked samples or incurred test samples) should be used and the data should be evaluated using appropriate predefined acceptance criteria. Some, but not all, validation issues have been addressed in the FDA guideline on bioanalytical method validation [3].

*Assay Format.* Assay format selection is the first step in method development. Assay formats include competition, sandwich, direct and indirect binding, inhibition, solid-phase, and solution phase assays. Reagents, first of all, include the antibodies; but diluents and additives such as detergents also have to be considered. For solid-phase assays, selection of the solid support as well as the chemistry used for the immobilization of the antibody may be critical. Consideration should also be given to the selection of the assay detection system to provide good signal-to-noise ratio. Detection may be improved by switching from colorimetric detection to fluorimetric or chemiliminescesence or, seldom, to radiometric detection. All these variables are evaluated during method development and confirmed during prestudy validation. In addition to the individual components, the manner in which an array is set up and run should already be considered during method development. The assay configuration (i.e., the number and placement of standards) of quality control (QC) samples, and study samples on a plate in an attempt to mimic the anticipated size of the run batches during routine use should be established as

**TABLE 2. Validation Parameters**

| Parameter | Method Development | Prestudy Validation | In-Study Validation |
|---|---|---|---|
| Assay format | establish | confirm | apply |
| Reagent selection, stability | identify and establish | confirm | monitor, lot change requires confirmation of performance |
| Specificity and selectivity | establish | confirm | may not apply, confirm |
| Matrix selection | establish | confirm | monitor, changes in matrix require demonstration of comparability |
| Minimum required dilution | establish | confirm | monitor |
| Standard curve, calibrators | select model, establish | confirm | monitor |
| Assay range | evaluate | establish | apply |
| Dilutional linearity | evaluate | establish | monitor, establish for dilutions not covered in prestudy validation |
| Parallelism | not applicable | investigate when possible | establish with incurred samples |
| Precision and accuracy | evaluate imprecision and bias | establish imprecision and bias | total error (4–6–30 rule) |
| Sample stability | initiate | establish | continuing assessment |
| Robustness and ruggedness | evaluate | establish | monitor |
| Run acceptance criteria | not applicable | based on standard curve acceptance criteria | standard curve and QC acceptance criteria, apply 4–6–30 rule |

early as possible and confirmed during prestudy validation. It is recommended that at least 5% of the total samples of a given batch consist of QC samples.

*Reagent Selection and Stability.* Probably the most critical components of the assay are the antibodies used. These antibodies are produced by living organisms with the inherent variability of such reagents. They must be acquired in adequate amounts and sufficiently characterized. As antibodies are prone to lot-to-lot variations, ideally, different lots are evaluated during method development and prestudy validation. When an antibody (reagent) lot has to be changed during routine application, in-study validation must demonstrate comparable performance of the lots.

Assay performance and sensitivity will deteriorate upon degradation of the antibodies and other reagents. Therefore, it is important to investigate storage conditions to ensure the integrity of key reagents for the estimated period of time that they will be used. Stability testing is not addressed in the regulatory guidance documents [3–5], but it is an important aspect of method validation. If

manufacturers provide expiration dates of reagents, they may be used instead of in-house stability determination. Reagent and antibody stability does not only concern storage at low temperatures (refrigerated or frozen), but standing times in laboratory equipment have to be considered. Thus, storage and handling conditions usually include bench-top stability, short-term and long-term storage, and stability to multiple freeze–thaw cycles. As it may also be desirable to store batches of assay microtiter plates for later use, the performance of a stored plate batch should be evaluated testing positive and negative samples. Stability is typically assessed during method development and confirmed during prestudy validation. It should be monitored during routine use.

*Reference Material.* In contrast to small molecules where reference standards are well characterized and certified standards are often commercially available from sources such as the U.S. Pharmacopeia, the European Pharmacopoeia, or the World Health Organization (WHO), proteins are often not as rigorously characterized and their purity may vary from supplier to supplier. Variation in posttranslational modifications such as glycosylation and deamidation may be present. Thus, the proteins can vary in their potency and immunoreactivity. As the reference compounds are used as standard calibrators, validation sample and QC sample variation of the reference will have a profound impact on the assay performance. Therefore, it is important to clearly document the source of the material and to characterize the proteins as thoroughly as possible. Comparability between lots or sources should be evaluated if possible. If the analyte is not a new drug entity, the innovator company is typically the most reliable source of authentic material. As stated for the reagents, stability of the reference compounds is an issue that has to be ensured.

*Specificity and Selectivity.* With regard to antibodies, specificity is the ability to specifically bind to the protein of interest in the presence of related endogenous and exogenous components (i.e., without cross reactivity). Selectivity, a related concept, describes the ability to determine an analyte in the presence of other constituents in a sample. Chromatographic methods are selective because they separate and detect analytes in a complex sample, typically after sample preparation steps for compound isolation. In contrast, ligand-binding assays measure analytes in biological matrices without prior isolation. Thus, high specificity may be desirable. With regard to interferences with other sample and matrix components and nonspecific binding of an antibody to such components, the terms "specific nonspecificity" and "nonspecific nonspecificity" are also used [9] sometimes as synonyms for the ICH terms. Specific nonspecificity is the interference caused by compounds with structural similarity to the analyte of interest. Such compounds may be metabolites, degradation products of the analyte, isoforms and variants with posttranslational modifications, as well as endogenous substances. Causes for nonspecific nonspecificity may be unrelated matrix components, (patho) physiological factors interfering with ligand binding such as serum proteins and lipids, hemolysis, or anti-IgG antibodies. In addition, nonspecific adsorption to the microtiter plate may occur.

As macromolecular analytes often have structural elements in common with endogenous compounds, specificity may be difficult to achieve. Variant forms of proteins may not be available at the time of method development. In such cases,

retrospective assessment may be acquired for the assay as more data become available over time. Rather than investigating specificity during method development, it is recommended to focus on reliable quantification of the analyte against a background of interfering matrix components [11].

Assay selectivity (nonspecific nonspecificity) is evaluated during method development by assaying spiked samples. Multiple lots of matrix should be evaluated, at least six [9] to 10 lots [11] are recommended. As selectivity problems occur, mostly at low concentrations, spiking should be performed at or near the lower limit of quantification (LLOQ). However, it may also be advisable to investigate higher concentrations. In case of interferences, it may be necessary to adjust the LLOQ before validation.

During prestudy validation, specificity and selectivity are confirmed. Selectivity may be expressed as acceptable recovery, applying the same principles as for the assessment of accuracy (see below). Acceptance criteria should be predefined. Acceptance criteria for selectivity and specificity typically do not exist for in-study validation. If potential interference may become a problem caused by the matrix from persons with the disease, specificity and selectivity should be confirmed by repeating suitable experiments once those disease-state matrices become available.

*Matrix Selection*. Matrix selection occurs in the development stage. Typically, biological fluids such as urine, plasma, serum, or cerebrospinal fluid as well as tissue samples may be collected. Additives such as anticoagulants, protease inhibitors, antioxidants, and so on may be present. It is necessary to document collection, processing, and storage conditions. The assay format may influence the choice of the matrix. For example, automated pipetting systems may be clogged by fibrin clots from plasma. In the absence of an endogenous signal, simple spike recovery experiments will determine the suitability of the matrix. In contrast, when the therapeutic protein is a recombinant version of an endogenous constituent, quantifiable amounts of the endogenous protein will be present in the matrix. Various strategies may be employed to limit or eliminate such interferences [9]. If the endogenous concentration is very low, and subsequently the percentage of the area under the curve obtained in pharmacokinetic studies caused by the endogenous compound is below 5%, the endogenous protein will introduce only a small bias and can be neglected. Alternatively, the endogenous level is determined from blank samples (no spike added) of a number of subjects and subsequently subtracted from the spiked samples. As already stated, 6–10 lots of the matrix should be evaluated. Further strategies include "stripping" of the matrix from the endogenous analyte either by nonspecific adsorption on charcoal or specific removal by affinity chromatography, the use of surrogate matrix from other species, or the use of protein-containing buffers. However, one has to keep in mind that the QC samples have to be prepared in the "original," unprocessed matrix. The presence or absence of matrix effects should be demonstrated by analyzing spiked samples at least at one concentration level. Differences between matrices obtained from healthy and diseased individuals may have to be considered.

In prestudy validation, the matrix selected during method development will be used to construct the calibration curves and validation samples. Once the effect of the matrix has been determined in method development, no further validation in later stages is required unless changes in the matrix occur.

*Minimum Required Dilution.* The minimum required dilution for an assay is the minimum magnitude of dilution of a sample with a defined diluent to optimize accuracy and precision in an assay. In many cases, dilution may not be necessary when analyzing plasma, serum, or other body fluids. Calibrator and validation samples are directly prepared in the matrix. For example, dilution may be required for samples with a background signal that is not caused by the endogenous version of the analyte. In the interest of a high signal-to-noise ratio (i.e., good accuracy and precision), sample dilution should be kept at a minimum. Once established during method development, this parameter should not be changed in later stages.

*Assay Range.* The assay range is defined as the validated concentration range between the LLOQ and the upper limit of quantitation (ULOQ), for which the results have an (predefined) acceptable level of precision and accuracy. The LLOQ and ULOQ are determined by the lowest and highest validation samples that show precision and accuracy of at least 25% expressed as relative error. The range is estimated during method development and validated by samples in the area of the anticipated LLOQ and ULOQ during prestudy validation. During routine use, samples that are above the ULOQ must be diluted (see "minimum required dilution" above). Samples that are below the LLOQ must be reported as "below LLOQ." The LLOQ during a batch run must be revised to higher concentrations if the QC samples at the lowest concentration fail to meet the 25% precision criteria.

*Calibrators and Standard Curve.* Standard calibrators are prepared by spiking known amounts of the (ideally, well-characterized) reference material into the matrix to obtain a standard curve from which the sample concentrations will be calculated. One of the major differences between chromatographic methods and immunoassays is that immunoassays display nonlinear relationships between the concentration and the measured response, which makes selection of the mathematical calibration curve model more complicated. Selection of the optimum calibration function is important to define the correct quantification range, to maximize accuracy and precision, and to achieve the quality control criteria. The mathematical model most widely used for immunoassay calibration curves is the four-parameter logistic function. If the curve is asymmetric, inclusion of a fifth parameter (i.e., using a five-parameter logistic function) may improve the fit to the data [25]. Algorithms that linearize the function such as logit-log may be used if goodness of fit is demonstrated, but such models are not recommended. Proper weighting of the data points in a calibration curve is also important to minimize bias and imprecision of the interpolated values near the LLOQ and ULOQ. Replicates with smaller variances are given greater weight compared with those with larger variances. The latter are normally found at the asymptotic ends of the curve [25].

As the standard concentrations should not be changed once validation has started, detailed investigation during method development is recommended using a greater number of data points and replicates compared with later validation stages [9, 11]. In the method development phase, the calibration curve should be constructed from a minimum of 10 non-zero standard points in duplicate spanning the anticipated concentration range about equally spaced on a logarithmic scale. Curve

fit is achieved by a four-to-five-parameter logistic function. Whether to weight or not weight the responses should be supported by an evaluation of the relationship between the standard deviations of replicate values and the mean values at different concentration levels [25]. A minimum of three independent runs should be used to establish the calibration model with duplicate curves included in each run to estimated intrabatch standard curve repeatability. The absolute relative error for back-calculated standard point should be ≤20% ("intracurve"). Acceptability of a model is verified by evaluating the relative error (relative bias) between back-calculated and nominal concentrations of the calibration samples. A model is considered acceptable if the relative error for all back-calculated standard points does not exceed 10% ("intercurve") and the precision (coefficient of variation) for each calibrator is ≤15% [11]. Lack of fit may be caused by the use of an inappropriate mathematical function such as applying a four-parameter logistic function to asymmetric curves or inappropriate weighting of calibrators.

During prestudy validation, a minimum of six non-zero standards in duplicate are spaced evenly on the logarithmic concentration scale to fit the four-to-five-parameter logistic function. At this point, additional calibrators outside the range of quantification (so-called "anchor points") may be included to facilitate curve fitting. This approach is in agreement with the FDA guidance on bioanalytical method validation [3]. The regression model should be confirmed using at least six independent runs. Typically, the same runs are used to determine accuracy and precision (see below). For acceptable curves, the back-calculated values for at least 75% of the standard points not including the anchor points should be within 20% of the theoretical value (except at the LLOQ, where 25% are acceptable), and upon completion of the validation, the cumulative relative error and coefficient of variation for each calibrator should be ≤15% and ≤20% at the LLOQ.

The standard curve should be monitored during in-study validation with at least one set of calibrators per patch run. As for prestudy validation, the curve should be constructed from six concentrations in duplicate. Anchor points may be used. The final number of points used for curve fit must be either 75% of the total number or a minimum of six calibrator samples not including the anchor points. The relative error of the back-calculated samples should be ≤20% (≤25% at the LLOQ). If either the high or low calibrator standards have to be deleted, the range for this particular run must be limited to the next standard point. Samples out of range must be repeated.

*Precision and Accuracy.* Precision and accuracy are assay performance characteristics that describe the random (statistical) errors and systematic errors (bias) associated with repeated measurements of the same sample under specified conditions [3–5]. Precision is typically estimated by the percent coefficient of variation (% CV, also referred to as relative standard deviation or RSD) but may certainly also be reported as standard deviations. Method accuracy is expressed as the percent relative error (% RE) and is determined by the percent deviation of the weighted samples mean from samples with nominal reference values. A collection of validation sample statistics can be found in References 9, 11, and 25.

Spiked samples are analyzed over multiple runs with replicate determinations during method development and prestudy validation. QC samples are used during in-study validation to monitor the performance of the assay. Limits for minimum

acceptable precision and accuracy should be established before or during method development and used throughout the life cycle of the assay.

It is recommended to determine accuracy and precision during method development with at least three batch runs using a minimum of eight sample concentrations analyzed in duplicate. The concentrations should span the whole validation concentration range of the assay. Recommended target limits for intrabatch and interbatch precision (% CV) are ≤20%, bias (% RE) should not exceed ±20%. At the LLOQ, a maximum of 25% for % CV and % RE is acceptable. These values are more lenient compared with the typical target values of chromatographic assays (% CV; 15% and 20% at LLOQ [3–5]) because immunoassays are inherently less precise than chromatographic assays [27]. In prestudy validation, at least six batch runs with a minimum of five different concentrations, one at the LLOQ, one at the ULOQ, and three concentrations in the lower, medium, and high range, analyzed in duplicate should be used for accuracy and precision determination. For each validation sample, the repeated measurements from all runs have to be statistically analyzed together. The target values for intrabatch and interbatch precision as well as accuracy (bias) are the same as the values applied in method development (20%, except for LLOQ, where 25% is acceptable). The total error of the method (i.e., the sum of % RE and % CV) should not exceed 30% (40% at the LLOQ). Further, for each in-study run, precision and accuracy are monitored by QC samples. Recommended run acceptance criteria are based on the total error, specifically on the deviation of the measured values from the nominal values, and not on statistical calculations such as the means or standard deviations. As QC samples, one set of at least three concentrations, one each in the low, medium, and high concentration range, are analyzed in duplicate in each batch run. As for small molecules, at least two thirds of the measured values of the QC samples must fall within a certain percentage of the corresponding nominal values and at least 50% of the results for each concentration of the QC samples must be within the specified limit. Thus, not all samples of a specific concentration are allowed outside the specifics. For chromatographic assays for small molecules, 15% has been adopted as limit (i.e., the "4–6–15 rule") [2]. At the LLOQ, 20% are acceptable. As a result of the lower precision of immunoassays, the error margin has been widened, a 6–4–25 rule is recommended by Findley et al. [9], whereas a 4–6–30 rule has been proposed at the macromolecule bioanalysis workshop [8]. The 30% margin is identical with the maximal acceptable total error of the method (sum of % RE and % CV) of the prestudy validation. Other statistical methods as acceptance criteria have been published [9, 28, 29] and can also be applied.

*Dilutional Linearity.* As the range of an immunoassay is usually limited, it may be necessary to dilute concentrated samples. Therefore, it has to be demonstrated that the analyte can still be reliably quantified upon dilution of high concentration samples so that they fall within the validated range of the assay. Moreover, a so-called prozone or "hook effect" can be identified. A hook effect is present when high concentration samples above the ULOQ display a lower response than ULOQ samples because of signal suppression caused by the high concentration of the analyte [30]. Dilutional linearity should not be confused with parallelism (see below).

Dilutional linearity is evaluated during method development, typically with spiking 100-fold to 1000-fold greater concentrations into the sample matrix

followed by dilution with the assay matrix. Dilutions should be made so that several dilutions fall within the standard curve in the lower, middle, and upper parts of the curve. Dilution samples above the ULOQ can be included to evaluate a hook effect. The dilutions are further confirmed during prestudy validation. The back-calculated concentration for each diluted sample should be within 20% of the nominal or expected value. The precision of the cumulative back-calculated concentration should be ≤20%. If a sample has to be diluted during routine use of the method to a higher extend than assessed during prestudy validation, dilutional linearity should either be repeated or a dilutional QC sample should be included in the assay.

*Parallelism.* Parallelism is a characteristic that is typically assessed during in-study validation. It is conceptually similar to dilutional linearity with the difference that it is determined by dilutions of actual study samples (incurred samples). In some cases, samples from a preclinical pilot study may be available during prestudy validation so that assessment of parallelism may be performed at that time. Moreover, when an assay is validated with the aim to replace another assay, incurred samples from a previous study may be available for evaluation of this performance characteristic.

Parallelism is assessed using $C_{max}$ samples from an actual study. Commonly, samples from several individuals are pooled to create a suitable validation sample. Using pooled samples eliminates the need for generation of multiple values for individual study samples. It is recommended that the relative standard deviation between samples of a dilutional series should be ≤30%.

*Sample Stability.* Experiments demonstrating sufficient stability of the analyte in the sample matrix must be included in prestudy validation. Such experiments should mimic as closely as possible the conditions under which study samples will be collected, stored, and processed. Assessment should include bench-top stability, refrigerator stability, whole-blood stability, freeze–thaw stability, and long-term freezer stability. Stability samples can be prepared by spiking the analyte reference at high and low concentrations into the sample matrix.

Bench-top stability refers to the conditions under which the samples are handled during processing of the samples at the analytical site and should be examined at room temperature for at least 2 h and at 2–8°C (refrigerator temperature) for a minimum of 24 h [31]. The stability in whole blood can be determined by spiking the analyte into freshly collected whole blood followed by incubation for up to 2 h and processing to obtain plasma or serum samples at certain intervals. The samples are subsequently processed and analyzed [31]. Typically, freeze–thaw stability evaluation includes three freeze–thaw cycles with at least 12 h between the thaws. The rate of freezing and thawing should mimic the manner in which samples will be handled as they are thawed before the assay. Long-term stability should demonstrate that the samples are stable throughout the lifetime of the study. The necessity to conduct studies on samples stored at –20°C and –70°C to –80°C may depend on the duration of the study.

Assessment of the stability is typically performed during prestudy validation and continued during in-study validation. If changes in sample handling or storage occur, additional stability evaluations must be carried out to reflect the altered conditions. The acceptance criteria for the stability evaluations, with the exception of the whole-blood stability, will be the same acceptance criteria applied for

accuracy and precision of QC samples. If the measured value is within the acceptance criteria for accuracy, the sample is considered stable.

*Robustness and Ruggedness.* Robustness is the ability of an assay to withstand small but deliberate changes that may affect the assay. Such factors may, for example, include incubation temperature and duration, number of washes, light exposure, and lot-to-lot differences in key assay reagents or microtiter plates. Changes that may have an impact on the assay have to be clearly identified in the method description (SOP). The term ruggedness is not mentioned in the IHC guidelines [4, 5], but it is included in the USP monograph on validation [6]. The term describes the consistent performance of an assay under routine changes (i.e., different analysts, instruments, batch size). Such changes should have no significant impact on the consistency of an assay.

The extent of the assessment of robustness and ruggedness depends on the anticipated application of the method, the current status of the assay's life cycle, industry standards, and, last but not least, common sense. The majority of robustness testing will be conducted during method development to facilitate the early identification of factors that affect assay performance. Prestudy robustness and ruggedness validation may be limited to the conditions demonstrating acceptable performance under the anticipated in-study conditions, such as incubation temperature and time tolerances, batch sizes, and so on, but more formal evaluation can also be applied. Acceptable robustness and ruggedness during routine use are assumed when monitoring in-study QC samples demonstrating acceptable intra-assay and interassay precision.

*Run Acceptance Criteria.* Run acceptance criteria are used to accept or reject a run because of its performance. As a consequence, no defined run acceptance criteria are applicable during method development. Prestudy validation runs are accepted based on the standard curve acceptance criteria. No run can be rejected because of poor performance of a sample during precision and accuracy evaluation; all data from prestudy validation runs are reported unless there has been a clearly recognizable error during sample preparation or measurement. Despite the fact that the standard curve must satisfy the criteria described for standard curves above, in-study runs are accepted based primarily on the performance of the QC samples. As stated above in the paragraph on precision and accuracy, the 4–6–30 rule is recommended (i.e., at least four of six QC samples must be within 30% of their theoretical values and at least 50% of the values for each level must satisfy the 30% limit.

***Biomarker Assays.*** A biomarker is defined as an endogenous substance that reflects physiological or pathophysiological processes or pharmacological responses to a therapeutic intervention [14], with a few exceptions such as viral load biomarkers are endogenous substances. It is a diverse class ranging from electrolytes to small molecules and macromolecules and a wide variety of analytical methodologies is used to quantify such substances. Although many assay characteristics apply to other analytical techniques and analytes as well, the following discussion will only touch on ligand-binding assays for the determination of macromolecules. As for the bioanalysis of therapeutic proteins, no official guidance documents are currently

available for biomarker analysis. Moreover, differences exist between validation approaches according to GLP regulations, which are the basis for documents of the FDA and other regulatory agencies and the National Committee for Clinical Laboratory Standards (NCCLS) guidelines for diagnostic assays. Details can be found in References 13–15; the present discussion will only include the parameters used in a GLP regulatory environment.

Besides, in clinical diagnostics especially, novel biomarkers play an increasing role in drug development for the investigation of the pharmacologic response to drug treatment or as surrogate markers for clinical endpoints. Clinical and drug development decisions will depend on the quality of biomarker data. Thus, the utility and value of such data is ultimately determined by the validity of the assay and requires demonstration and documentation of performance characteristics, as mentioned in the previous subsection, for immunoassays such as accuracy, precision, specificity, range, stability, and so on. However, in contrast to bioassays for drug compounds, where quantitative results are achieved by calibration typically using well-defined reference standards, biomarker assays may differ considerably depending on the type of analytical measurement, the type of the analytical data that develop from the assay, or the intended use of the assay. Subsequently, different assay types and validation levels may apply. Rigorous method validation for a novel biomarker is not necessary for drug discovery-phase work. However, the design of the validation must change when the assay is transferred from screening to quantitative determination in later phases of drug development.

Biomarker assays (as other bioassays) may be classified into "definitive quantitative assays," "relative quantitative assays," "quasi-quantitative assays," and "qualitative assays" with varying degrees of validation requirements (Table 3) [14, 15]. For definitive quantitative assays, a well-defined or characterized standard of the biomarker is available. In the case of relative quantitative assays, calibration is performed with a standard that is not well characterized, not available in pure form, or not representative of the endogenous biomarker. Results from these assays are

**TABLE 3. Assay Categories and Validation Characteristics**

| Parameter | Definitive Quantitative Assay | Relative Quantitative Assay | Quasi-Quantitative Assay | Qualitative Assay |
|---|---|---|---|---|
| Accuracy | ✓ | ✓ | — | — |
| Precision | ✓ | ✓ | ✓ | — |
| Sensitivity | ✓ (LLOQ) | ✓ (LLOQ) | ✓ | ✓ |
| Specificity | ✓ | ✓ | ✓ | ✓ |
| Range | ✓ (LLOQ–ULOQ) | ✓ (LLOQ–ULOQ) | ✓ | — |
| Dilutional linearity | ✓ | ✓ | — | — |
| Parallelism | ✓ | ✓ | — | — |
| Standard and reagent stability | ✓ | ✓ | — | — |
| Matrix stability | ✓ | ✓ | ✓ | ✓ |

expressed as numerical values. Currently, most biomarker assays fall into the relative quantitative category. In quasi-quantitative assays, no reference material is available for the construction of a calibration curve. Nevertheless, the analytical result is expressed in numerical units. Examples are the measurement of enzymatic activity (expressed as units per volume) and vaccine or anti-drug antibodies where the response is reported as percent binding or titer. Qualitative assays use no standard either, and the results are reported non-numerically (i.e., "low, medium, and high" or "+, ++, and +++").

Although with respect to validation criteria a lot of similarities exist between ligand-binding assays for pharmaceutical proteins and biomarker assays, significant differences have to be noted, especially for novel biomarkers, in which a suitable reference standard is not present. If the intended reference is a recombinant protein, one should keep in mind that such proteins often have glycosylation patterns different from the endogenous equivalents. The glycosylation pattern of endogenous proteins is often heterogeneous so that it is virtually impossible to prepare glycoprotein standards that are identical to the natural circulating proteins. In the ideal case, a purified endogenous protein from the target species is used as a reference. This standard should be characterized in terms of analytical purity as thoroughly as possible. If a well-characterized standard is not available, the assay results will provide rather "relative" than "true" numerical values. As a result of the presence of endogenous biomarkers, no analyte-free matrix exists for the preparation of calibrator standards, which makes the establishment of the LLOQ particularly challenging. Standard curves may be prepared using pooled matrix from individuals with low baseline concentrations of the compound. Alternatively, (affinity) stripped matrix, a protein-containing buffer or matrix from an alternate species with negligible concentrations of the analyte, may be considered. Further strategies for minimization of interference from endogenous biomarkers can be used [9]. The preparation of calibrator standards not in the actual sample matrix is one major difference of biomarker assays from assays of protein drugs. When using such approaches, it has to be ensured that the use of the surrogate matrices does not introduce a bias in the assay. In contrast to calibration samples, QC samples must be prepared in an authentic matrix. In this case, matrix samples containing low basal levels can be selected and the middle and upper QC concentrations can be prepared by the addition of known biomarker amounts. Moreover, the disease state can have an impact not only on endogenous biomarker levels but also on the composition of the matrix itself. High concentrations of the marker produced during disease can cause a hook effect. A disease state may also alter the heterogeneity of the biomarker with altered cross reactivity to the antibody relative to the standard. The modified matrix composition can result in increased nonspecific binding (i.e., nonspecific nonspecificity).

The stages of validation of biomarker assays include establishment of the biomarker (development), so-called prevalidation, prestudy validation, and in-study validation [13–15]. The following short discussion will focus on the "GLP-like" definitive and relative quantitative assays. As the development and validation of an assay for novel biomarkers is quite diverse, the application of strict validation procedures appears problematic. Therefore, upon establishment of the prototype assay in the development phase, a formalized validation plan should be developed that

defines the scope and purpose of the assay. Further activities during method development include the establishment of the reference standard, selection of antibodies and the assay format, as well as evaluation of key reagents. As mentioned above, selection of the matrix for the preparation of the calibrator concentrations may be challenging because of the presence of endogenous biomarker. Subsequently, the calibration curve model will be established. As for other immunoassays, nonlinear calibration using the four-to-five-parameter logistic model is the commonly acknowledged model for data fitting.

The prevalidation phase may be regarded as a method optimization phase where the calibration model is confirmed; range, LLOQ, and ULOQ are defined; and matrix interference is evaluated. The use of anchor points for the calibration curve may be feasible. As mentioned above, determination of the LLOQ may be difficult because of the presence of the endogenous analyte. Dilutional linearity may be evaluated as well. It is also considered useful to assess the biomarker in healthy and diseased individuals. For this purpose, at least 25 individuals should be tested [9] to account for intrasubject variability caused by circadian and seasonal fluctuations and intersubject variability.

Prestudy validation will additionally evaluate standard and reagent stability as well as matrix stability during collection processing and storage of the samples, will determine dilutional linearity and parallelism, and will confirm assay range and the calibration model. These criteria can be determined as described for immunoassays for pharmaceutical proteins above. In addition, accuracy and precision have to be determined. It is recommended to use validation samples at five different concentration levels analyzed at least in duplicate and in a minimum of six runs. One concentration should be at the anticipated LLOQ, one about 2–4 times the LLOQ, one in midrange on the log scale, one about 70–80% of the anticipated ULOQ, and one at the anticipated ULOQ. The following method acceptance criteria have been recommended [14]: Both accuracy (% RE) and precision (% CV) should be within 25%, except at the LLOQ where 30% is acceptable. Even more lenient criteria may be required in some cases, depending on the analyte or the type of assay and its limitations. It is important to note that such QC samples must be prepared in the actual matrix. The use of buffer, surrogate, or stripped matrix is not feasible, except for rare matrices such as cerebrospinal fluid or tears, where a surrogate matrix may be the only practical option. Spike recovery experiments should be performed on multiple individual lots of matrix to assess the accuracy, matrix effects, and interference. It may also be favorable to include a pilot study into prestudy validation runs.

As for immunoassays for pharmaceutical proteins, in-study validation of biomarker assays should include one set of calibrators to monitor the standard curve as well as a set of QC samples at three concentrations analyzed in duplicate for the decision to accept or reject a specific run. Recommended acceptance criterion is the 6–4–30 rule, but even more lenient acceptance criteria may be justified based on statistical rationale developed from experimental data [14].

***Anti-drug Antibody Assays.*** Basically, the protein sequence of biophar-maceutical therapeutics can be nonhuman, chimeric, humanized, or fully human. Most such therapeutics elicit some level of antibody response against the product leading to

potentially serious side effects or loss of drug efficacy. Thus, the immunogenicity of therapeutic proteins is a concern for clinicians, manufacturers, and regulatory agencies. For the detection of anti-drug antibodies, a number of assay methods including enzyme-linked immuno-sorbert assays (ELISA), immunoblotting, surface plasmon resonance, and bioassays are available, each technique having its own advantages and disadvantages [24]. Whereas binding assays identify the antibodies, immunoblotting provides information on the specificity. Surface plasmon resonance can show the antigen-antibody interaction in real time, whereas bioassays demonstrate the neutralizing potential of the antibodies. To date, microtiter plate-based ELISAs are still the most widely used format for testing for anti-drug antibodies because of their simplicity, sensitivity, and high-throughput capability. Generally, the validation parameters outlined for standard immunoassays as required by the regulatory authorities apply to immunoassay-based anti-drug antibody assays as well. These parameters include specificity, selectivity, accuracy, sensitivity, precision, robustness, ruggedness, and stability of reagents, analyte, and matrix. However, some differences exist because of the nature of the antibodies. These differences will be briefly addressed below. Further details can be found in References 22–24.

As a result of the lack of reference material, anti-drug antibody assays are generally quasi-quantitative assays (Table 3). In addition, the target analyte is generally polyclonal, consisting of antibodies of various isotype classes, specificities, and affinities (i.e., the analyte is not a defined molecule). Thus, no single positive control exists that accurately represents the target analyte. During the early validation phase, no clinical studies have typically been performed, so it may be a challenge to obtain a representative positive sample for method development and validation. When establishing specificity, accuracy, and sensitivity, several control analytes from different individuals or sources representing the test population should be investigated. Specificity of analyte binding can be assessed using immunodepleted samples.

Selectivity is a critical parameter determining the reliability of an antidrug antibody assay. In this context, one has to keep in mind that selectivity can vary between test samples because of the heterogenous nature of the antibodies. For the evaluation of selectivity, immunoglobulins and other potential interfering substances can be spiked into positive and negative samples at high but physiologically relevant concentrations. No substantial interference can be concluded if the recovery is 80–125%. The influence of different sample matrices, typically serum and plasma, should also be evaluated if both matrices may be analyzed with the same assay. The type of the matrix should not change the outcome of the assay. Furthermore, a comparison of specificity between normal- and disease-state matrices should be conducted to detect interfering substances that may be present in certain populations or disease states. Another unique property of anti-drug antibody assays is that the drug itself can act as an interfering substance, which can be mimicked with addition of the drug in varying concentrations to positive controls.

As a result of the quasi-quantitative nature of anti-drug antibody immunoassays and the lack of a reference standard a threshold value, the so-called "cutoff" or "cutpoint" is used to identify positive samples from nonspecific background noise. The assay cutoff is preferably determined by analyzing samples from healthy individuals and those affected by the disease. The data are subsequently used to

calculate the cutoff value yielding 5% of false positives [23]. Moreover, one should consider that a low optical readout increases imprecision so that the cutoff level should not be too low for optical assays.

The sensitivity of an immunoassay is typically defined by its LLOQ, which can only be determined when a reference standard is available. Alternatively, a detection limit of qualitative assays is used where a distinction is essentially made between positive and negative results only. Typically, the antibody data are reported as "titer," the titer being the reciprocal of the highest dilution of a sample in which the instrument response is greater than the cutoff response [22]. At least two positive control analytes should be used. If a positive control antibody is available, a pseudo-calibration curve can be generated by a series of dilutions. However, one should keep in mind that "true" quantitation is impossible because of the lack of a true reference compound. Determination of the dilutional linearity is not so important when the result is reported as a titer. However, if the determination of positive samples is based on the interpolation from a reference standard curve, it is essential to demonstrate that the QC samples fall within the (limited) linear range of the calibration curve and not on a plateau or a region that may include a hook effect.

Precision should be assayed using positive controls, negative controls, and a diluent sample. Positive controls should be prepared at a high and a low concentration to demonstrate precision within this assay range. Typically, precision (% CV) of ≤30% is considered acceptable [22]. In-study monitoring of a batch run using QC samples consisting of at least one positive control, a matrix negative control, and a diluent negative control should be used to estimate assay performance.

### 2.2 Validation of Cell-based Bioassays

Bioassays use living systems that measure the biological activity of a therapeutic agent. Such assays may be used to study the effects of hormones or growth factors, but such systems can also address drug toxicity and side effects. Moreover, bioassays may also be applied to bioanalysis of biopharmaceutical proteins. Only *in vitro* assays using cell culture systems measuring a discrete response such as cell proliferation, differentiation, or survival will be briefly addressed. Bioassays may generate quantitative or quasi-quantitative data. Only a few considerations for bioassay method validation will be mentioned, as this type of assay is not frequently applied. Most validation procedures for quantitative assays, such as calibration and reagent and matrix stability, accuracy, and precision, are essentially identical to the procedures described for standard immunoassays above.

For a validated assay, an established immortal cell line is typically used. Thus, during method development, not only must a suitable reference standard be established, but also documentation of the cell line with respect to characteristics such as origin of the cell line, culture and passage history, morphology, surface markers, and receptors is required. It is advisable to study the effect of cell age (number of passages) on the measured response. Specificity may be another issue as cell lines proliferate, differentiate, or senesce and die in response to a large number of biomolecules that may be present in the samples obtained for bioanalytical studies. In addition, macromolecules can be metabolized leading to metabolites that may also be biologically active, which is not relevant when a pure compound is applied to the cell culture to study its effect on the cells. However, bioanalytical methods based on

bioassays may not be specific for the analyte of interest. In this case, extensive study of interference because of nonspecificity should be conducted during prestudy validation using samples from a number of representative individuals.

## 3 CONCLUSIONS AND FUTURE CONSIDERATIONS

Despite the wide availability of chromatographic techniques hyphenated to mass spectrometry, immunoassays remain the most important methods for bioanalytical applications for monitoring macromolecules such as therapeutic proteins, biomarkers, or drug-induced antibodies. Current guidelines of regulatory authorities focus on chromatographic techniques with no or little reference to the specifics of ligand-binding assays. As some differences with respect to assay validation exist between chromatographic and ligand-binding assays, a number of issues require special attention. Reference material of the target analyte(s) is not always available in pure form, which is especially true for biomarkers and anti-drug antibodies. In addition, key reagents such as the antibodies for the immunoassay are frequently not commercially available. Stability issues of those key reagents, the reference material, and the biological matrix have to be considered.

The calibration curves of immunoassays are nonlinear, so that special attention should be paid to the selection of the correct calibration model. Anchoring points out of the validated range may optimize the curve fit. Selectivity of an immunoassay depends on the specificity of the antibody directed toward the analyte. Nonspecific interferences from the matrix as well as specific interferences (cross reactivity) from related compounds have to be considered. Special challenges occur if the analyte is an endogenous compound. In this case, analyte-free matrix may not exist, so that alternative strategies for the preparation of validation samples have to be explored. However, QC samples should be prepared in the original matrix if possible. Pathological states may alter the composition of the matrix or, in the case of biomarkers, the respective concentration, which has to be considered for an appropriate selection of the calibration standards and QC samples. Assays for anti-drug antibodies are quasi-quantitative so that complete GMP-like validation is normally not possible. Finally, as ligand-binding assays are inherently less precise than chromatographic assays, more lenient acceptance criteria for accuracy and precision as well as for run acceptance should be applied. The current opinion according to conference reports recommends as target values for precision (expressed as % CV) and accuracy (expressed as % RER) a maximum of 20% (25% at the LLOQ). Despite known disadvantages, application the 6–4–30 rule as run acceptance criteria during in-study validation has been adopted.

The current gap between the need for validated immunoassays according to GLP compliance and the lack of official guidance documents will certainly be closed in the near future because ongoing efforts at conferences between scientists from regulatory authorities, pharmaceutical companies, and scientific organizations will ultimately result in such guidance documents. However, technological advances in instrument automation and new technologies will continue to create new issues that have to be considered when developing and validating methods for bioanalysis. As each technique has and will have unique features, the challenge is the implementation of a dynamic, yet standardized and systematic approach for analytical method validation.

## 4 GLOSSARY OF VALIDATION TERMS

The terms are used according to the ICH guidelines [4, 5], where indicated, or according to the FDA guideline [3].

***Acceptance criteria***: Numerical limits, ranges, or other suitable measures for the acceptance of the results of analytical procedures.

***Accuracy* (ICH)**: The closeness of agreement between a measured value and the theoretical true value. In statistics, accuracy is typically reported as % relative error (% RE).

***Batch***: A set of standard curve calibrators, validation samples or QC samples, or study samples that are analyzed in a single group of measurements; it is synonymous with run.

***Bias***: Systematic difference between measured test results and the theoretical true value. Bias is expressed as relative error (% RE) or as a ratio (% recovery).

***Calibration curve***: The functional relationship between the analyte concentration in standards (calibrators) and the measured response; it is synonymous with standard curve.

***Calibrator***: A solution of a biological matrix spiked with the analyte of interest for the construction of calibration curves; it is synonymous with standard.

***Dilutional linearity***: A condition where a spiked sample is serially diluted to result in a set of samples containing analyte concentrations that fall within the quantitative range of the assay and the whole set of dilutions can be measured with acceptable accuracy. Dilution of the sample does not result in biased measurement of the analyte concentration.

***Intermediate precision* (ICH)**: Precision within a laboratory: different days, analysts, equipment, and so on.

***Lower limit of quantitation (LLOQ)***: The lowest concentration of an analyte in a test sample that can be determined quantitatively with suitable accuracy and precision (ICH term: limit of quantitation).

***Parallelism***: A condition where spiked (calibrator) samples are serially diluted to result in a set of samples having analyte concentrations that fall within the quantitative range of the assay and do not exhibit an apparent trend toward increasing or decreasing estimates of the analyte over the range of dilutions.

***Precision* (ICH)**: The closeness of agreement (degree of scatter) between a series of measurements obtained from multiple sampling of the same homogenous sample under defined conditions. Precision is considered at three levels: repeatability, intermediate precision, and reproducibility. In statistics, precision is typically reported as % coefficient of variation (% CV), also referred to as relative standard deviation (RSD).

***Quality control (QC) sample***: Prestudy or in-study samples with a known (nominal) concentration that are treated as unknowns in the assay. During in-study runs, QC samples are used as the basis for run acceptance or rejection.

***Range* (ICH)**: The interval between the lower and upper concentrations of an analyte in samples for which it has been demonstrated that the analytical procedure has an acceptable level of accuracy and precision.

***Repeatability* (ICH)**: The precision of repeated measurement under the same operating conditions over a short interval of time (within one run), also termed intrabatch or intrarun precision.

***Reproducibility* (ICH)**: Precision of repeated measurements between analytical laboratories; also termed intralaboratory precision.

***Robustness* (ICH)**: A measure of the capacity of analytical methods to remain unaffected by small but deliberate variations of the experimental parameters; indicates the reliability of the method during routine use.

***Run***: A set of standard curve calibrators, validation samples or QC samples, or study samples that are analyzed in a single group of measurements; it is synonymous with batch.

***Selectivity* (ICH)**: Ability of a bioanalytical method to measure particular analyte(s) in a complex medium without the interference from other components of the mixture.

***Specificity* (ICH)**: The ability to unequivocally measure the analyte in the presence of other components including related compounds or matrix components.

***Upper limit of quantitation (ULOQ)***: The highest concentration of an analyte in a test sample that can be determined quantitatively with suitable accuracy.

***Validation***: Demonstration that a method is suitable for the intended use; it is typically demonstrated by establishing validation parameters that are applied to sample analysis.

***Validation samples***: Biological matrix samples spiked with the analyte at predetermined concentrations; it is used during prestudy validation to assess accuracy and precision.

## REFERENCES

1. Shah V P, Midha K K, Dighe S, et al. (1992). Analytical methods validation: Bioavailability, bioequivalence and pharmacokinetic studies. *Pharm. Res.* 9:588–592.
2. Shah V P, Midha K K, Findlay J W A, et al. (2000). Bioanalytical method validation—a revisit with a decade of progress. *Pharm. Res.* 17:1551–1557.
3. Food and Drug Administration (2001). *Guidance for Industry: Bioanalytical Method Validation*. Available: http://www.fda.gov/cder/guidance/4252fnl.pdf.
4. International Conference on Harmonization (1994). *Guideline Q2A, Text on Validation of Analytical Procedures*. CPMP/ICH/381/95. Available: http://www.ich.org/LOB/media/MEDIA417.pdf.
5. International Conference on Harmonization (1996). *Guideline Q2B: Validation of Analytical Procedures : Methodology*. CPMP/ICH281/95. Available: http://www.ich.org/LOB/media/MEDIA418.pdf.
6. United States Pharmacopoeia 29 (2006). Section 1125: Validation of Compendial Methods, United States Pharmacopeia Convention, Rockville, MD.

7. Approved Biotechnology Drugs. Available: www.bio.org/speeches/pubs/er/approveddrugs.asp.
8. Findley J W A, Das I (1998). Validation of Immunoassay for Macromeolcules from Biotechnology. *J. Clin. Ligand Assay*. 21:249–253.
9. Findley J W A, Smith W C, Lee J W, et al. (2000). Validation of immunoassays for bioanalysis: A pharmaceutical industry perspective. *J. Pharm. Biomed. Anal.* 21:1249–1273.
10. Miller K J, Bowsher R R, Celniker A, et al. (2001). Workshop on Bioanalytical Methods Validation for Macromolecules: Summary Report. *Pharm. Res.* 18:1373–1383.
11. DeSilva B, Smith W, Weiner R, et al. (2003). Recommendations for the Bioanalytical Method Validation of Ligand-Binding Assays to Support Pharmacokinetic Assessments of Macromolecules. *Pharm. Res.* 20:1885–1900.
12. Smolec J, DeSilva B, Smith W, et al. (2005). Bioanalytical method validation for macromolecules in support of pharmacokinetic studies. *Pharm. Res.* 22:1425–1431.
13. Lee J W, Devanarayan V, Barrett Y C, et al. (2006). Fit-for-purpose method development and validation for successful biomarker validation. *Pharm. Res.* 23:312–328.
14. Lee J W, Weiner R S, Sailstad J M, et al. (2005). Method validation and measurement of biomarkers in nonclinical and clinical samples in drug development: A conference report. *Pharm. Res.* 22:499–511.
15. Lee J W, Nordbloom G D, Smith W C, et al. (2003). Validation of bioanalytical assays for novel biomarkers: Practical recommendations for clinical investigation of new drug entities. In J. Bloom (ed.), *Biomarkers in Clinical Drug Development*, Marcel Dekker, New York, pp. 119–149.
16. Dantus M M, Wells M L (2004). Regulatory Issues in Chromatographic Analysis in the Pharmaceutical Industry. *J. Liquid Chromatogr. Rel. Technol.* 27:1413–1442.
17. European Medicines Agency (2005). *Draft Guideline on the Clinical Investigation of the Pharmacokinetics of Therapeutic Proteins*. EMEA/CHMP/89249/2004. Available: http://www.emea.eu.int/pdfs/human/ewp/8924904en.pdf.
18. Food and Drug Administration (2001).*Guidance for Industry: IND Meetings for Human Drugs and Biologics. Chemistry, Manufacturing, and Controls Information*. Available: http://www.fda.gov/cder/guidance/3683fnl.pdf.
19. Deyl Z (ed.) (1997). *Quality Control in Pharmaceutical Analysis*, Elsevier, Amsterdam, the Netherlands.
20. Ermer J, Miller J H M (eds.) (2005). *Method Validation in Pharmaceutical Analysis*. Wiley-VCH, Weinheim, Germany.
21. Colburn W A, Lee J W (2003). Biomarkers, Validation and Pharmacokinetic-Pharmacodynamic Modeling. *Clin. Pharmacokin.* 42:997–1022.
22. Geng D, Shankar G, Schantz A, et al. (2005). Validation of immunoassays used to assess immunogenicity to therapeutic monoclonal antibodies. *J. Pharm. Biomed. Anal.* 39:364–375.
23. Mire-Sluis A R, Barrett Y C, Devanarayan V, et al. (2004). Recommendations for the design and optimization of immunoassays used in the detection of host antibodies against biotechnology products. *J. Immunol. Meth.* 289:1–16.
24. Wadhwa M, Bird C, Dilger P, et al. (2003). Strategies for detection, measurement and characterization of unwanted antibodies induced by therapeutic biologicals. *J. Immunol. Meth.* 278:1–17.
25. Smith W C, Sittampalam G S (1998). Conceptual and Statistical issues in the validation of analytic dilution assays for pharmaceutical applications. *J. Biopharm. Stat.* 8:509–532.

26. Assay Guidance Manual, version 4.1. Available: www.ncgc.nih.gov/guidance.
27. Braggio S, Barnaby R J, Grossi P, et al. (1996). A strategy for validation of bioanalytical methods. *J. Pharm. Biomed. Anal.* 14:375–388.
28. Hubert P, Chiap P, Crommen J, et al. (1999). The SFSTP guide on the validation of chromatographic methods for drug bioanalysis: From the Washington conference to the laboratory. *Anal. Chim. Acta*. 391:135–148.
29. Kringle R O (1994). An assessment of the 4–6-20 rule for acceptance of analytical runs in bioavailability, bioequivalence, and pharmacokinetic studies. *Pharm. Res.* 11:556–560.
30. Rodbard D, Feldman Y, Jaffe M L, et al. (1978). Kinetics of Two-Site Immunoradiometric (Sandwich) Assays. *Immunochemistry.* 15:77–82.
31. Kringle D, Hoffman D (2001). Stability methods for assessing stability of compounds in whole blood for clinical bioanalysis. *Drug Inf. J.* 35:1261–1270.

# MICROARRAYS IN DRUG DISCOVERY AND DEVELOPMENT

Krishnanand D. Kumble[1] and Sarita Kumble[2]

[1] *Genesis Research, Parnell, Auckland, New Zealand*
[2] *Pictor Limited, Glendowie, Auckland, New Zealand*

**Chapter Contents**

1 Introduction
2 Array Formats
  2.1 Planar Arrays
  2.2 Solution Arrays
3 Factors for Consideration
  3.1 Surfaces
  3.2 Probes
  3.3 Sample Processing
  3.4 Signal Detection
  3.5 Data Analysis
4 Types of Arrays
  4.1 DNA Arrays
  4.2 Protein Arrays
  4.3 Tissue Arrays
  4.4 Other Arrays
5 Implications of Array Technology
  5.1 Experimental Design
  5.2 Molecular Network Analysis
  5.3 Drug Discovery and Development
References

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

## 1 INTRODUCTION

Microarrays are grids of biomolecules (DNA, proteins, carbohydrates, small molecules) in which each element of the grid performs a specific assay such as identification of a specific binding molecule or measuring a particular enzymatic activity. A single sample is analyzed on each microarray providing simultaneous measurements for multiple parameters. Typically a complete experiment involves a large number of microarrays resulting in huge datasets. This deluge in data has led to approaches for integration of diverse datasets to enhance our understanding of both an individual molecular function and elaborate biological processes.

This potential of microarrays to facilitate global analysis of genes and gene products has led to its adoption in nearly every area of biological science—from basic research to clinical diagnostics. They are being used to tackle one of the major challenges in the postgenomic era—to understand the regulation, expression, and function of entire sets of genes, transcripts, and proteins. This information will help in characterizing complex biological processes at the molecular level in various cell types, both in normal and in disease states.

Since their conception in the mid-1990s, microarrays have seen widespread use in almost all aspects of biological research and their range of applications continues to expand. These include (1) biomarker discovery, to find genes that can be used for measuring and following disease progression; (2) target selectivity, to identify genes that can distinguish one disease from another as well as subclasses of disease; (3) pharmacology and toxicogenomics, to help in identification of poor compounds and to optimize the selection of promising leads; and (4) drug screening, identification of protein activity modulators. As the cost of microarrays continues to drop, it is clear that microarrays are becoming a more integral part of the drug discovery process. In short, microarrays are redefining the drug discovery and development process by providing greater knowledge at each step and by helping to understand the complex workings of biological systems.

This chapter provides a technical overview of microarrays, including information on array formats, production and use of microarrays for interrogating various molecular species—DNA, RNA, and proteins—and analysis of data. The challenges in experimental design to maximize microarray data quality, and the use of microarrays in understanding intracellular molecular networks as well as in drug discovery and development, are also discussed.

## 2 ARRAY FORMATS

### 2.1 Planar Arrays

***High-Density Microarrays.*** High-density arrays are ideally suited for analyzing a small number of samples against thousands of genes, proteins, or small molecules. Early transcript profiling studies used high-density arrays in the context of target discovery with the objective of obtaining a short list of high-priority candidates that showed interesting expression patterns in disease states [1]. Arrays of oligonucleotides representing the entire human genome have been used for discovering polymorphic loci as well as the presence of thousands of alternative alleles [2, 3]. Other uses of high-density DNA arrays include determination of methylation patterns [4] and identification of transcription factor binding sequences [5].

Manufacture of high-density protein arrays presents a greater challenge due to the inherent heterogeneity in the physico-chemical properties and stability of proteins. However, arrays of recombinant proteins representing all yeast open reading frames (ORFs) have been manufactured and used for identification of novel binding activities [6]. *Escherichia coli* transformed with cDNA libraries to express mammalian proteins have been arrayed to identify autoantibodies in serum from the mouse model of systemic lupus erythematosus [7].

***Low-density Microarrays.*** Low-density arrays can be used for simultaneously performing less complex analysis on many different samples [8]. The ability to rapidly screen thousands of biological samples for multiple parameters in the same assay makes the process cost-effective and easy to perform. This format is based on either using slides with hydrophobic barriers to create wells [9] or the 96-well microtiter plate formats, in which each well contains replicate mini-arrays of spots [10].

Low-density arrays are becoming increasingly popular for cytokine measurements due to the relative ease with which preexisting conventional enzyme-linked immunosorbent assay (ELISA) assays can be adapted to a miniaturized format [11]. In addition, low-density DNA arrays are being used for measuring polymorphisms to determine drug toxicity [12], and expression signatures obtained from high-density arrays are being used to guide the design of arrays with small sets of genes for use in cancer subclassification and prognosis [13].

Over the last few years, several different types of microarrays of various probe densities and content—DNA or protein—have become commercially available (Table 1). These provide microarray tools for analyzing samples derived not only from human tissues, but also from a wide spectrum of other species ranging from bacterial pathogens to Arabidopsis to mouse.

### 2.2 Solution Arrays

An alternative approach to planar arrays is three-dimensional arrays in solution. These have the advantage of parallel analysis and high throughput, in addition to better kinetics of binding in solution compared with planar arrays [14]. Positional information in these arrays is retained by a variety of methods such as fluorescence-encoded beads (two dyes with varying ratios incorporated into the bead set) and bar-coded nanoparticles (self-encoded with submicron metal stripes). The independence of each element offers the flexibility to multiplex either a few or thousands of measurements without the need to customize each assay. The high degree of reliability and reproducibility has resulted in the development of fluorescence-encoded beads for multiplex diagnostic assays [15].

This article will focus on technical issues, applications, and future directions in the use of planar arrays.

## 3 FACTORS FOR CONSIDERATION

### 3.1 Surfaces

***Solid Supports.*** A wide variety of support materials have been used for manufacturing microarrays. Glass is the most commonly used support due to its

**TABLE 1. A Partial List of Commercially Available Ready-to-Use Microarrays**

| Array Type | Company | URL | Brief Product Description |
|---|---|---|---|
| High-density oligonucleotide | Affymetrix | www.affymetrix.com | *In situ* synthesized oligonucleotides with multiple perfect match and mismatch probes per target gene. |
| High-density oligonucleotide | Agilent | www.agilent.com | 60-mer probes in predesigned printed arrays; custom microarray design service is also available. |
| High-density oligonucleotide | Applied Biosytems | www.appliedbiosystems.com | 60-mer probes designed to be within 1500 bases from the 3′ end of the transcript spotted on each chip. |
| Low density oligonucleotide | Mergen | www.mergen.com | 60-mer probes printed in 16 replicate arrays per slide with 1000 probes in each array. |
| Low-density DNA | Superarray | www.superarray.com | Pathway-specific probes (up to 500 target genes) arrayed as 400-bp cDNA or 60-mer oligonucleotides. |
| Multiple content | Pamgene | www.pamgene.com | Printed oligonucleotides, peptides, or proteins as 4 or 96 replicate arrays on chips or plates, respectively. |
| Peptide | JPT Peptide Technologies | www.jpt.com | Printed peptide (12 to 15-mer) arrays for profiling of protease, kinase and phosphatase substrates. |
| Peptide | Pepscan systems | www.pepscan.nl | Arrays printed with 1200 protein kinase substrate peptides. |
| Protein | Invitrogen | www.invitrogen.com | Slides printed with recombinant proteins (yeast or human). |
| Antibody | Pierce | www.piercenet.com | Four to sixteen antibodies for ELISA assays printed in replicate arrays in 96-well format. |
| Antibody | BD Biosciences | www.clontech.com | Over 500 antibodies spotted for comparative profiling of protein levels from two samples. |
| Tissue | Biochain | www.biochain.com | Arrays or frozen tissue from 15 different species for determination of species cross-reactivity of antibodies. |
| Tissue | Chemicon | www.chemicon.com | Normal and cancer tissues arrayed in duplicate from paraffin-embedded tissue blocks. |

low background fluorescence and amenability to automation [14]. Polystyrene with fluorescence and binding properties similar to glass is beginning to be increasingly adapted for microarray use [16]. These surfaces are usually modified by coating with one-, two-, or three-dimensional structures that provide either covalent attachment chemistry or enable binding of molecules through adsorption. Other microarray surfaces include membranes such as nitrocellulose and nylon and gold or silicon films [17–19]. Synthesis of oligonucleotides has been performed *in situ* on fused silica wafers [20], and peptide arrays have been synthesized on cellulose and polypropylene membranes [21].

***Chemistry.*** Microarray surfaces must maintain the stability and activity of attached biomolecules, while remaining surface-bound through all processing steps. Spot morphology and background noise, either due to nonspecific sample binding or from the detection system used, are important issues to be considered in choosing the right immobilization chemistry. A variety of surfaces have been derivatized to expose various active groups; these determine the type of attachment—ionic or covalent—of the biomolecule to the surface [22].

Polylysine, which results in an amine surface carrying a positive charge, was among the early microarray surfaces used for binding to negatively charged biomolecules [23]. Membranes such as nitrocellulose and nylon that have traditionally been used for a variety of blotting applications have been modified for microarrays by application onto glass backing [17]. Table 2 provides commercial sources of commonly used surfaces for printing nucleic acids and proteins.

Oligonulceotides modified to carry an amino group or proteins through their lysine residues can be covalently attached to aldehyde- and epoxy-derivatized surfaces. Other reactive surfaces used to covalently link both DNA and proteins include N-hydroxy succinimide [20] and maleimide [24]. Proteins expressed with polyhistidine and biotin tags have been attached to surfaces coated with nickel chelate and streptavidin, respectively [25, 26]. This strategy is likely to result in the proper orientation of the displayed molecules on the surface, thereby improving their functionality and stability.

Heterobifunctional cross-linkers have been used to attach activated biomolecules on surfaces functionalized with either aminosilane [27] or mercaptopropylsilane [28]. These strategies have resulted in up to fourfold higher signal-to-noise ratios compared with the polylysine surface for arrayed antibodies [29]. Specific immobilization via interaction of surfaces functionalized with salicylhydroxamic acid derivatives and phenyldiboronic acid-labeled nucleic acids and proteins has also been demonstrated [30].

A new generation of chemistries has introduced poly(ethylene glycol)-functionalized surfaces that prevent nonspecific protein binding, thereby removing the need for a blocking step during processing [31]. This commercially available chemistry also offers a wide variety of additional functional groups. A large spacer between the support and the biomolecule results in higher analyte binding by avoiding steric hindrance [29]. Three-dimensional surfaces used for arraying biomolecules were shown to increase binding capacity and reduce denaturation of immobilized molecules, but they slowed down reaction kinetics due to reduction in diffusion rates [32]. These same chemistries are applied for biomolecule attachment to beads in the manufacture of suspension arrays.

**TABLE 2. A Partial List of Commercially Available Microarray Surfaces**

| Company | Surface Chemistry | Application | URL |
|---|---|---|---|
| Telechem | Primary amine | Oligonucleotides and PCR products | www.arrayit.com |
| | Primary aldehyde | Amino-modified oligonucleotides and PCR products, or proteins | |
| Corning | Gamma aminopropyl silane | Long (>50-mer) oligonucleotides and cDNA | www.corning.com |
| Erie Scientific | Aluminized Aminopropyl silane aldehyde silane or epoxide | PCR products | www.eriemicroarray.com |
| | | Amino-modified oligonucleotides and PCR products, or proteins | |
| | Poly-L-lysine | PCR products or proteins | |
| Schleicher and Schuell | Nitrocellulose polymer | Proteins | www.schleicher-schuell.com |
| Pall Corporation | Modified nylon membrane | mRNA, PCR products and proteins | www.pall.com |
| Perkin Elmer | Specially formulated acrylamide-based polymer | Proteins | http://las.perkinelmer.com |
| NUNC | Aminopropylsilane, | Oligonucleotides and PCR products | www.nuncbrand.com |
| | Aldehyde and Epoxy | Amino-modified oligonucleotides and PCR products, or proteins | |
| | Maxisorp | Proteins | |
| Sigma | Poly-L-lysine | PCR products or proteins | www.sigma-aldrich.com |
| | Streptavidin | Biotin-tagged oligonucleotides or proteins | |
| | Maleimide | Amino-modified oligonucleotides and PCR products, or proteins | |
| Greiner | Aminosilane | Oligonucleotides and PCR products | www.greinerbiooneinc.com |
| | Aldehyde and Epoxy | Amino-modified oligonucleotides and PCR products, or proteins | |
| | Nickel | Histidine-tagged proteins | |

### 3.2 Probes

***Printing.*** Printing refers to spotting arrays of presynthesized biomolecules directly onto a solid surface. This versatile approach can be applied to generate microarrays of almost any biomolecule in conjunction with a proper im-mobilization method. Printing can generate many copies of the same array more efficiently than *in situ* synthesis because the immobilized molecules need to be synthesized only once. Several kinds of microarray printing technologies have been developed (Table 3), but the most commonly used method uses contact printing. In this method, pins are used to transfer samples from a source to the solid support by direct surface contact. Noncontact piezoelectric printers, which are also often used, can print more spots per unit time than contact printers by employing an electric current to accurately and rapidly dispense samples onto the solid support. Although the reproducibility of spot volumes delivered by contact printers typically has a

**TABLE 3. A Partial List of Commercially Available Microarray Printers**

| Company/Arrayer | Features of the Printing Technology | URL |
|---|---|---|
| Affymetrix<br>417 Arrayer<br>427 Arrayer | Microcontact printing using a pin and ring system.<br>Spot size of 150 to 200 μ with 50-pL delivery volume. | www.affymetrix.com |
| Cartesian Dispensing<br>SynQuad<br>Hummingbird | High-speed microsolenoid valve with a high-resolution syringe pump dispense system.<br>Delivery volumes from 20 nL to μl. | www.cartesiantech.com |
| Genomic Solutions<br>MicroGrid<br>OmniGrid | Microcontact printing with solid or quill pins.<br>Spot size and dispense volume dependent upon pin type. | www.genomicsolutions.com |
| Telechem<br>SpotBot | Microcontact printing with solid or quill pins.<br>Spot size and dispense volume dependent upon pin type.<br>Low-cost arrayer with a footprint for bench-top use. | www.arrayit.com |
| LabNext<br>Xact | Microcontact printing with solid or quill pins.<br>Spot size and dispense volume dependent upon pin type.<br>Lowest cost arrayer on the market. | www.labnext.com |
| Packard Bioscience<br>SpotArray<br>BioChip Arrayer | Non-contact printing with a piezo-electric printhead.<br>Spot size of 75 to 200 μ dependent upon the pin type. | www.packardbioscience.com |

coefficient of variation (*cv*) below 20%, piezoelectric printers have greater control over the volume dispensed resulting in *cv*s of under 5%.

"Dropouts" or missing spots can be a problem, especially during high-throughput manufacture, and are primarily due to clogging of the dispensing components in the arrayer. To identify dropouts, microarrayers can be fitted with optical devices that can monitor dispensing and provide data on dropouts that can then be used to "fill in" missing spots. This system has vastly improved the quality of microarrays, reducing the need to print replicate spots of the same probe.

Other less-common deposition methods include electrospraying in a stable cone-jet mode to generate highly reproducible spots of as little as 50-pL volumes [33] and a laser transfer technique that allows the accurate deposition of picoliter volumes of proteins onto the solid surface [34].

***In Situ Synthesis.*** Two typical *in situ* synthesis approaches are light-directed parallel synthesis and peptide synthesis on membrane (SPOT) synthesis. Although the former approach was initially developed for peptide synthesis [35], it has been adapted for the synthesis of DNA microarrays. This method uses a combination of photo-lithographic masks and combinatorial chemistry to synthesize oligonucleotides on fused silica wafers [36]. High-density arrays have also been manufactured by maskless methods using either digitally controlled aluminium mirrors to fabricate oligonucleotides by photodeposition [37] or by phos-phoramidite chemistry in microfluidic chips with three-dimensional nano-chambers [38].

The maskless method is also being used to create peptide arrays in microfluidic chips by parallel synthesis using digital photolithography and photogenerated acid during the deprotection step [39]. Peptide arrays have also been manufactured by SPOT synthesis using a combination of novel polymeric surfaces, linker and cleavage chemistries, as well as robotic liquid-handling systems [40]. These methods involve alternate synthesis and washing steps, increasing the possibility of contamination by reagents from adjacent spots, and thereby limiting array density [41].

### 3.3 Sample Processing

***DNA Analysis.*** Detection of single nucleotide polymorphisms (SNPs) was one of the early applications of high-density oligonucleotide arrays [42, 43]. More recently these arrays have also been used to assess DNA copy numbers [44]. Each of these measurements requires the initial extraction of genomic DNA, which is then processed using several different methods each designed to incorporate labels for visualization of binding to specific probes on the array. Fragments from restriction digested genomic DNA were ligated with adaptors, amplified, and enzymatically end-labeled with biotin followed by hybridization to oligonucleotide arrays [42, 44]. Alternatively, labeling was done by incorporation of a biotin-tag during primer extension after hybridization of unlabeled amplified DNA fragments [43]. The biotin was then detected using Streptavidin conjugated to a fluorescent dye.

Genome-wide DNA–protein interactions have been mapped using intergenic oligonucleotide microarrays [45]. In this approach, called chromatin immunoprecipitation (ChIP), epitope-tagged proteins of interest are allowed to bind specific regions in genomic DNA and then cross-linked. The DNA-bound epitope-tagged proteins are immunoprecipitated using a tag-specific antibody. The DNA is then

delinked from the protein and, after fluorescent labeling, hybridized to arrays for identification of regions that bind to the protein of interest. A similar approach has been used for identification of methylation sites in genomic DNA. In this method, oligonucleotide primer adaptors are attached to restriction-digested DNA, followed by a secondary digestion with a methylation-sensitive enzyme, labeling with a fluorescent dye, and hybridization to arrays [46]. These methods are providing a wealth of information on the factors that regulate gene expression and helping to understand these processes at a global level.

***Transcript Analysis.*** Microarrays permit the rapid analysis of complex gene expression changes in cells and tissues during development, both normal and disease. These changes give distinct patterns that can be used for discovering new disease-specific therapeutic targets, and molecular diagnostics, as well as for following treatment efficacy and disease prognosis.

Typically for transcript profiling, a labeled nucleotide is incorporated during reverse transcription of the total cellular mRNA pools. Not only does this approach require a large amount of RNA (50 to 200 μg of total RNA) for hybridization, but RNA purity is also a critical factor in array performance due to potential nonspecific binding by other labeled macromolecules [47]. Arrays have been widely used for obtaining relative mRNA concentrations between two samples, test and reference, in which each sample is labeled with a different dye (typically Cy3 and Cy5), followed by mixing of the samples before hybridization [48]. An alternative method avoids incorporation of bulky dyes during reverse transcription by incorporating amino allyl labels to which N-hydroxyl succinimidyl dyes are chemically coupled in a later step [49].

The use of a common reference RNA allows for the comparison of data between various array experiments. Typically a pool of RNA derived from a variety of tissues is used as a reference sample. Efforts are being made to implement common standards (MIAME, minimal information about a microarray experiment) for transcript profiling through the MGED Society to enable data sharing between different groups [50].

Methods in which the amount of mRNA has been amplified to produce labeled cRNA, by incorporating a T7 RNA polymerase promoter site into one end of the cDNA followed by *in vitro* transcription are also widely used [51]. In this method, quantitative estimates of the amount of each transcript in a given sample can be calculated. In single-sample labeling experiments, a reference RNA may be spiked in during sample labeling and hybridization to facilitate quality control of the process as well as for comparison of data between different arrays.

Despite its current widespread use and the drive toward ensuring high-quality microarray data by the implementation of MIAME guidelines, very little is known about the extent to which application of different technologies influences the outcome of transcriptional profiles and differential expression. However, microarray users should be aware of studies that have attempted to present a comprehensive evaluation encompassing different probe types (oligonucleotides and cDNAs), labeling techniques and hybridization protocols [52].

***Protein Analysis.*** The early adaptation of protein microarrays is the reformatting of already available ELISAs. In these assays, which were first described by Roger

Ekins [53], a protein antigen is identified and its quantity is measured using two antigen-specific antibodies—one surface-immobilized to capture the antigen, and the other chemically labeled to bind the captured antigen in the solution phase. The second antibody produces a detectable signal through the label. This method, in which the protein sample does not have to be labeled, has been widely used for measuring cytokine levels in a variety of samples such as serum and tissue culture supernatants [54, 55]. Protein expression profiles have also been measured by capturing dye-labeled protein lysates derived from cells and tissues on arrays of antibodies. These measurements have largely been ratiometric in which two samples each labeled with a different dye using protocols similar to those used for transcript analysis are mixed and incubated with arrays of antibodies [56, 57]. Single-sample analyses could be performed by spiking known amounts of control proteins into the samples before labeling.

### 3.4 Signal Detection

***Detection Chemistries.*** The most prevalent method for detecting binding of target to immobilized probes on microarrays are fluorescent dyes, of which Cy3 and Cy5 are the most widely used especially for differential transcript and protein profiling [58], whereas single-target hybridizations are commonly performed with fluorescein isothiocyanate [51]. Although Alexa dyes allow use of more than two colors in a single experiment, they are less commonly used [59]. Enzyme-labeled fluorescence (ELF), a phosphatase substrate that results in a precipitable product, has been used for both DNA and protein arrays [60].

Chemiluminescent detection is also used, mainly to detect antigen capture on antibody arrays by sequential incubations with biotinylated secondary antibodies and Streptavidin-conjugated horseradish peroxidase (HRP) [61]. Phosphorylation of arrayed kinase substrates in the presence of radiolabeled adenosine triphosphate (ATP) has been detected by autoradiography [62].

To increase sensitivity of detection, several signal amplification methods have been applied to the various types of molecular targets. Tyramide signal amplification (TSA) uses biotinyl-tyramide, an HRP substrate, to accumulate biotin at the reaction site. The "amplified" biotin is then detected using Streptavidin-HRP in conjunction with substrates that result in products detectable either by fluorescence, chemiluminescence, or colorimetry [63]. Other signal amplification methods include rolling circle amplification (RCA) in which an oligonucleotide-conjugated antibody binds to biotin on the target, followed by hybridization of a circular DNA molecule to the oligonucleotide. The circular DNA is then replicated using a DNA polymerase in the presence of a fluorescently labeled nucleotide [64]. Preformed branched DNA (dendrimers), each of which is attached to over 200 fluorescent dye molecules, have also been used for labeling array-bound targets [65].

***Imaging Systems.*** Assay format and cost are the primary determinants of the kind of devices used for imaging arrays. Table 4 lists some commercially available imaging systems used for microarray normalization. Fluorescence-based scanners that use lasers to excite fluorescent dyes attached to the target molecule (for transcript profiling) or to a specific secondary detection reagent (for ELISA assays) are the most widely used for signal detection and imaging. These instruments

**TABLE 4. A Partial List of Commercially Available Scanners**

| Company/ Scanner | Scanner features | URL |
|---|---|---|
| Alpha Innotech NovaRay | CCD camera with a broadband excitation source with up to eight excitation and emission filters for imaging slides or microtiter plates. | www.alphainnotech.com |
| Kodak ImageStation 2000MM | CCD camera with selectable multi-wavelength illumination source for imaging microtiter plates; can also be used for radiography and whole animal imaging. | www.kodak.com |
| Tecan LS Reloaded | Scanning laser imaging system can be used for imaging slides or microtiter plates with a sensitivity of less than 0.1 fluor/$\mu m^2$. | www.tecan.com |
| Axon Instruments GenePix 4000B | Scanning laser imaging system with multiple filter options can be used for imaging slides with a sensitivity of less than 0.1 fluor/$\mu m^2$. | www.axon.com |
| Affymetrix 418 Array Scanner | Laser scanning microscope with two excitation filters can read Affymetrix' GeneChips with a sensitivity of less than 1 fluor/$\mu m^2$. | www.affymetrix.com |
| ChromaVision ACIS | Digital microscope with software for tissue scoring and quantitative analysis suited for tissue microarrays. | www.chromavision.com |
| Molecular Devices ImageXpress Micro | Inverted epifluorescent microscope with CCD camera and optional laser designed for scanning slides and microtiter plates. | www.moleculardevices.com |

enable user-defined scanning resolutions and photomultiplier tube settings to adjust detection sensitivity. In addition, the availability of a wide range of fluorescent dyes makes this the most flexi-ble system for use in microarray image detection. Chemiluminescence-based detection systems using a charge-coupled device (CCD) together with conventional camera optics are also commonly used. Although the latter systems are flexible for applications in a wide range of assay formats, fluorescence scanners provide a higher range of sensitivity and dynamic range [66].

Label-independent methods such as surface plasmon resonance (SPR) can overcome variations caused by inconsistencies in labeling chemistries that are often seen in label-dependent detection systems. SPR has been used to measure affinities

in binding reactions, especially between antigens and their cognate antibodies. These highly sensitive systems are beginning to be adapted for microarray-based measurements [67]. Another recently described method relies on the change in fluorescence decay times of tryptophan and tyrosine residues in proteins to identify protein–protein interactions. In this method, a change in binding behavior of proteins in solution toward protein partners arrayed on a solid support is detected with regard to binding specificity and protein amount [68].

### 3.5 Data Analysis

Most commercial microarray imagers supply data extraction software that can accommodate the unique parameters of scanned images. In addition, printing precision has simplified the process of detecting spot boundaries and measurement of inter-spot distances. A useful consideration is the format for storage of primary scanned images (usually as tiff files) so as to be able to take advantage of future developments in image analysis software [69]. Storage of raw image files retains maximum information, allowing the use of different normalization, image extraction, and quality metrics.

Several commercially available software packages are available that interpret and transform an array image into a dataset (Table 5). Software programs grid the elements of the array as a first step in processing the image. Background-corrected intensity values for each spots are obtained using one of several options: local (area around individual spots) or global (average signal in area outside of the grid) background corrections, or user-defined values such as those from negative control data points contained within the array. After background-corrected intensity values have been calculated for each spot, the data are normalized with respect to sources of systematic and biological variation [70]. The choice of the normalization method is critical because it impacts precision of data comparison between arrays.

Several methods for statistical analysis of microarray data are available depending on the experimental setup and the kind of biological question that needs to be addressed. Initial approaches to analyzing microarray data focused on

**TABLE 5. A Partial List of Microarray Data Analysis Software**

| Name | Software Features | URL |
|---|---|---|
| Spotfinder | Software tool designed for microarray image processing using the TIFF image files generated by most microarray scanners. | www.tigr.org/software/ |
| MIDAS | Tool for microarray data quality filtering and normalization that allows raw experimental data to be processed through various data normalizations, filters, and transformations via a user-designed analysis pipeline. | |

**TABLE 5.** *Continued*

| Name | Software Features | URL |
|---|---|---|
| ArrayViewer | Software tool designed to facilitate the presentation and analysis of microarray expression data, leading to the identification of genes that are differentially expressed. | |
| GeneTraffic | The application allows users to import, visualize, analyze, interpret, and store microarray data in a complete, secure data management environment. Software allows easy development of custom interface to relational database. | www.iobion.com |
| Genespring | The platform is designed to identify genes/pathways that are relevant to specific biological questions by comparing analysis results from expression, genotyping, protein, metabolite, and other data types. | www.agilent.com/chem/ |
| BASE | Comprehensive database server manages biomaterial information, raw data, and images, and provides integrated and "plug-in"-able normalization, data viewing, and analysis tools. The system also has array production LIMS features that can be integrated with the data analysis. | http://base.thep.lu.se/ |
| Image Pro Plus | Software features automated microscope control, advanced fluorescence acquisition, image deconvolution, and three-dimensional reconstruction and measurement. | www.mediacy.com |
| Rosetta Resolver | Comprehensive customizable gene expression analysis package that incorporates analysis tools with a scalable database. Software has capabilities to associate expression profiles to molecular pathways. | www.rosettabio.com |
| Pathiam | Hardware-independent, Web-enabled software for viewing and analyzing immunohistochemically stained slides. Pathologists can automatically generate reports containing a complete quantitative analysis of each sample. | www.bioimagene.com |

unsupervised hierarchical clustering because these are simple ways in which data can be organized [71]. It is still the most commonly used analysis tool especially if the experiment has been designed to obtain a bird's eye view of transcript or protein profiles during a particular process. Machine learning techniques such as neural networks [72] and support vector machines [73] should be used for more advanced analyses when preliminary information can be used to guide data interpretation. Most software packages have a user-friendly graphical user interface (GUI) that enables performing a number of simple analysis, including data normalization, various kinds of clustering, and principal component analysis.

## 4 TYPES OF ARRAYS

### 4.1 DNA Arrays

***Transcript Profiling.*** Transcript profiling was one of the earliest applications of microarray technology. In this application, fluorescently labeled cellular transcripts are hybridized either to arrays of spotted cDNA [19] or *in situ* synthesized oligonucleotides [36]. The former have generally been used for hybridization of a mixture of two transcript pools, each of which are derived from a different biological sample (such as diseased and normal tissue), and labeled with a different fluorescent dye (Cy3 and Cy5) before mixing [19]. A ratiometric analysis of fluorescence then helps to determine the relative expression levels of each transcript in the two samples. cDNA arrays have received wide acceptance within the academic community due to their low cost and ease of manufacture, as well as the ability to customize rapidly as new genomic sequence information becomes available. Customization of array content also easily accommodates the research interests of individual laboratories.

Oligonucleotide arrays, used for measuring transcript amounts from single samples, have the advantage of displaying a much larger number of very small spots (5 μ), enabling the interrogation of multiple probes for each transcript, including mismatch probes for determination of hybridization specificity [36, 74]. Currently, oligonucleotide arrays that display probes covering all predicted genes in the human genome are commercially available from several vendors (Table 1).

The ability to obtain global transcript profiles has accelerated the process of discovery in all areas of biology ranging from basic discovery to drug development and diagnostics. Correlations between gene-expression patterns and disease states have led to the selection of the best drug targets for pharmaceutical development as well as for monitoring therapeutic outcomes [75]. One area in which DNA arrays are making a critical impact is cancer, where expression profiles from hundreds of cancer tissue samples have allowed subclassification of cancers based on the identification of specific transcript expression signatures. These are likely to guide therapy and improve prognosis for cancer patients in the near future [76–79].

***Physical Characterization of Genes.*** Oligonucleotide arrays have found use in the physical characterization of the genome by helping to map transcription factor binding sites and methylation sites [80]. These arrays, with their ability to package hundreds of thousands of spots on each chip, have been widely used for SNP discovery [81]. Clinically significant SNPs have then been rearrayed in a low-density

format for analyzing large numbers of samples to determine clinical validity. A U.S. Food and Drug Administration (FDA)-approved SNP genotyping array for CYP450 has recently been used to identify patients with a decreased ability to metabolize risperidone [10]. More recently, bacterial artificial chromosome (BAC) arrays have provided high-resolution maps of genetic changes, including gains and losses, as well as amplifications in chromosomal DNA by comparative genomic hybridization [82]. Other applications of oligonucleotide arrays have been to analyze splice variants by hybridization of labeled transcripts with arrays that combine exon and junction-derived probes, which are either specific or nonspecific to a splice event [83].

***Cell Arrays.*** Immobilized arrays of plasmid DNA used to transfect cells in the presence of a transfection reagent are called cell arrays or reverse transfection arrays. These arrays are overlayed with cultured cells in medium and incubated so that cells superimposed on the DNA spot are transfected. The desired phenotypic change is visualized by staining the cells after a brief incubation period [84]. This format allows a variety of readouts, including cytoskeletal changes, apoptosis, and DNA replication, which can be either visualized with a laser scanner or a fluorescence microscope. Similar approaches have been used for increasing the throughput of loss-of-function studies using RNAi to monitor effects on cell phenotype [85, 86]. These studies enable functional whole genome screens without the need for expensive screening facilities.

## 4.2 Protein Arrays

### *Protein Profiling*

*Antibody Arrays*. Antibodies are the most commonly arrayed protein class due to their structural similarity and stability. Many monoclonal antibodies generated over the years for binding to various epitopes on proteins offer a diverse source of capture and detection antibodies. These are being used in antibody arrays, especially for determination of growth factor and cytokine levels from a variety of biological samples [87]. Arrays have been multiplexed with antibodies that were developed and used for conventional ELISAs to measure levels of more than 50 cytokines with a high degree of sensitivity (pg/mL) from low sample volumes (less than 50 μL) [64]. However, each set of arrayed antibodies needs to be characterized for specificity of binding to the target molecule to ensure data quality. Recent studies have shown significant cross-reactivity of arrayed antibodies to proteins other than their intended targets [88].

Identification of disease biomarkers is a rapidly growing area of proteomic research. These biomarkers can enable better predictive capabilities in disease diagnosis and prognosis, as well as in drug development by identifying potential drug toxicities and side-effects. Although it has been suggested that antibody arrays can be used to profile serum and cell lysates, these measurements are complicated by the fact that protein concentrations in these samples cover 10 to 14 orders of magnitude [87], requiring systems that can simultaneously detect both low- and high-abundance proteins within a single array. The sample complexity issue may be alleviated by reduction in sample complexity before profiling. This can be accomplished using either liquid phase protein separation systems or by profiling

the protein complement of various cell organelles separately [89]. Samples can be labeled with either fluorescent tags followed by direct detection after capture or with other haptens such as biotin followed by detection with Streptavidin conjugated to a reporter molecule [90].

As well-characterized antibodies are available to only a small subset of total cellular proteins, profiling complex protein samples could be performed using arrays of *in vitro* synthesized antibody libraries [91]. This format has the advantage of being able to directly array bacterial cells rather than purified recombinant antibodies [92]. Once disease-relevant proteins have been identified through screening for global protein changes, arrays containing the "diagnostic" subset of antibodies that recognize disease-specific proteins can be used for high-throughput, large-sample analysis [66].

***Antigen Arrays.*** Several studies have been performed to demonstrate disease identification by analysis of serum antibodies to clinically relevant proteins [93–96]. One study used arrays of autoantigen diagnostic markers to measure autoantibody titers in serum of patients with autoimmune disease [95]. In an extension of this study, autoimmune patient sera screened for binding to arrayed antigens identified specificity of autoantibody responses defining autoantigens relevant in human disease [97]. More recently, the diversity of B-cell response as a function of disease severity was measured using an array of over 225 distinct epitopes derived from several myelin-associated proteins. These arrays could identify distinct sets of epitopes and could demonstrate a correlation between reduced epitope spreading and improved disease outcome. This information is now being used to guide the development of a tolerizing DNA vaccine, which could potentially limit epitope spreading and improve disease outcome [98]. These antigen arrays have been shown to have a much higher sensitivity than conventional ELISAs with a three-log linear dynamic range [97].

***Protein Function Arrays.*** The development of high-throughput expression and purification methods have made a large number of proteins available for arraying. A small amount of material (pg to ng) is sufficient for printing numerous arrays that can then be used to perform rapid, functional screens. The first genome-wide protein display, an array of all yeast ORFs, was used to screen for binders to calmodulin and phospholipids [6]. In this study, 5800 different yeast proteins with hexahistidine tags were arrayed on nickel-coated glass slides. The immobilized proteins were then probed with various labeled phospholipids resulting in the identification of more than 150 proteins that were shown to bind phospholipids for the first time. Other protein binding measurements include identification of protein–protein and protein–DNA interactions [99].

In addition to binding assays, functional properties are also being measured on arrays. Protein kinase activity has been determined either by immobilizing fluorogenic substrates followed by the addition of specific kinases [100] or on kinase arrays incubated with peptide substrates in the presence of radiolabeled ATP [101]. Protein kinase arrays have also been used for identification of inhibitors by incubating small-molecule binders in the presence of substrates demonstrating the ability of arrays to screen for kinase inhibitors [102].

G-protein-coupled receptors (GPCRs) currently make up the largest class of therapeutic drug targets and are an obvious choice for microarray-based drug

screening. An early study demonstrating the capability of microarrayed GPCRs for agonist and antagonist screening showed retention of activity through several activation/ deactivation cycles after ligand binding on arrayed rhodopsin [103]. Competitive binding assays in which mixtures of fluorescently labeled ligands and unlabeled inhibitors were used to demonstrate binding selectivity to different receptor subtypes have also been described [104].

***Other Types of Protein Arrays.*** Peptides, arrayed either by deposition on the surface [105] or by *in situ* synthesis [106], have been used for epitope-mapping [107], measuring enzyme activity [108], and cell capture [105]. They have also been used for identification of peptide antagonists to proteins that are likely drug targets [109]. In addition, protease specificities were identified using peptidyl coumarin substrate arrays created by immobilizing the fluorescent coumarin leaving group via the C-terminus to the solid support, enabling synthesis of a high-diversity peptide library at the N-terminal end [110].

Arrays of peptide–major histocompatibility complex (MHC) complexes have been created on acrylamide gel-coated glass surfaces and used for ligand-specific capture of CD4(+) and CD8(+) lymphocytes. This approach should be useful to characterize multiple epitope-specific T-cell populations during immune responses associated with infection, cancer, autoimmunity, and vaccination [111].

Covalent mRNA–protein fusions were displayed on single-stranded DNA arrays through nucleic acid hybridization of the mRNA in the fusion molecule [112]. Similarly, capture agents such as nucleic acid aptamers have also been proposed for use in protein binding [113, 114].

### 4.3 Tissue Arrays

Tissue microarrays (TMAs) are displays of several tens to hundreds of tissue specimens on a single slide for parallel analysis [115]. Paraffin embedded tissue are generally used for arraying [116]; however, arrays have also been constructed from frozen tissue [117]. The ability to array archival paraffin embedded tissue opens up vast archives of patient samples for medical research. Arrayed tissue sections processed either for cytological staining or *in situ* hybridization, combined with automated image analysis systems, provides a powerful molecular profiling tool. TMAs are commonly used to confirm results obtained from expression microarrays, as well as in the development of diagnostic and prognostic markers for clinical applications.

### 4.4 Other Arrays

Although DNA and protein arrays continue to be widely used, other types of molecules are being arrayed and used for a variety of applications. These include small-molecule arrays in which individual members of a chemical library are deposited on a surface and used for performing binding assays with target proteins [102, 118, 119]. Several strategies are being applied for displaying small molecules on surfaces. The simplest strategy is to spot molecules on glass after mixing with glycerol to maintain "wetness' after deposition. This strategy was used to create and screen arrays by spraying the target protein, human caspase 6, along with its fluorigenic substrate and screening for dark spots of enzyme inhibitors against noninhibitor fluorescent spots [120]. An elegant method that takes advantage of

DNA hybridization specificities to provide an address for arrays of protein nucleic acid (PNA)-tagged small molecules has been described [121]. More sophisticated libraries of molecules that selectively modulate activities of mutated kinases have been used to identify chemical inhibitors that switch off individual kinases [122].

Carbohydrate arrays have been developed and used as novel high-throughput analytical tools for monitoring carbohydrate–protein interactions such as profiling protein binding to various sugars and to measure enzymatic activity on sugar substrates [123]. These microarrays have also been used to discover anti-adhesion therapeutics by identifying carbohydrate binding specificities of pathogenic bacteria. The display of carbohydrate ligands on a surface in a manner that mimics interactions at the cell–cell interface is ideal for whole-cell applications [124]. Similarly, monosaccharide-based arrays of N-acetyl galactosamine and N-acetylneuraminic acid derivatives were used to identify binders to cholera and tetanus toxins [125].

## 5 IMPLICATIONS OF ARRAY TECHNOLOGY

### 5.1 Experimental Design

Microarrays are observed as tools for "descriptive" research and not for "hypothesis-driven" research, which is primarily driven to answer specific questions related to a known set of molecules. However, most microarray-based studies do have clear objectives and are designed to answer well-defined questions. The plan for specimen selection should follow from the objectives of the microarray study. Studies may either be exploratory, whose results should be confirmed, or designed to test various models; in which case, the experiments have to be designed in a focused manner [126].

Appropriate design of microarray studies is critical to draw valid and useful conclusions from the large amount of data that are invariably obtained from each experiment. Criteria for design will vary depending on the type of array used—DNA, protein, and small molecule—as well as the experimental objectives, and sources of variability within the system. Issues such as selection of samples, number of replicates needed, allocation of samples to dyes for ratiometric profiling, and sample size are important considerations in these studies [127].

Three sources of variation should be considered in the design of a microarray experiment: (1) biological, at the level of setting up of the cells, cell lines or animals, and sample treatment; (2) technical, which is introduced during sample processing; and (3) analytical, which highlights signal bias and readout [128]. To minimize each of these variations, the experiment should be set up such that the biological material to be used for analysis should be handled separately from the initiation of the experiment providing biological replicates. If this results in an unwieldy experimental setup, pooling the samples at some stage of processing may still allow a reduction in biological variance. This approach is risky in the event of the presence of an extreme outlier in the pool, because this can unduly influence the expression values obtained from the pool. Samples from every stage of the experiment with a likelihood of introduction of variation should be treated separately. Replicate measurements from the same biological sample helps to reduce technical variation.

**TABLE 6. Experimental Design for Two Color Microarray-Based Comparisons of Transcript Profiles Between Disease and Healthy Tissue Samples**

| Design | Array 1 | | Array 2 | | Array 3 | |
|---|---|---|---|---|---|---|
| | Dye 1 (e.g., Cy3) | Dye 2 (e.g., Cy5) | Dye 1 | Dye 2 | Dye 1 | Dye 2 |
| Reference | Disease Sample 1 | Healthy sample | Disease Sample 2 | Healthy sample | Disease Sample 3 | Healthy sample |
| Loop | Disease Sample 1 | Healthy Sample 1 | Healthy Sample 1 | Disease Sample 2 | Disease Sample 2 | Healthy Sample 2 |
| Balanced Block | Disease Sample 1 | Healthy Sample 1 | Healthy Sample 2 | Disease Sample 2 | | |

Ratiometric analysis of signal by comparing signal intensities between two samples from the same spot reduces errors due to printing differences. This has led to the use of a reference sample in transcript profiling experiments, and it is an integral part of differential profiling, making the choice of the reference sample critical. The most important considerations in choosing the appropriate standard are that it is readily available, homogeneous, and stable over a period of time. The reference sample does not need to have any biological relevance, because it merely serves to compare expression profiles between different experiments performed in different laboratories. Among the standards that have been used are mRNA populations obtained from a wide variety of cells or tissues with the aim of lighting up every element on the array. Alternatively, a reference sample is prepared from the samples that will be assayed in the experiment to ensure that every transcript present in the test samples will be represented in the reference standard [126]. Care has to be taken, however, that no transcript in the reference sample saturates the signal in the detection system.

If a direct comparison needs to be made between two samples, for example, samples from pooled healthy tissue with samples from disease tissue, it would be necessary to perform forward and reverse labeling to remove any dye bias that may alter the results [129]. Swapping dyes between samples reduces systematic bias in the data that can be corrected during the normalization step [130]. Alternative design strategies that have been suggested are the balanced block design [131] and the loop design [132]. The same considerations can be applied to two-color antibody array-based protein profiling [133]. These experimental design strategies have been briefly illustrated in Table 6.

### 5.2 Molecular Network Analysis

Computational analysis is essential to transform the large amount of data generated by microarrays into a mechanistic understanding of various modules connecting at hubs and nodes to form molecular networks. Methods that identify the regulatory mechanisms underlying these modules and processes by an integrative analyses in the context of other data sources are often capable of extracting deeper biological insight from the data. Such integrative computational approaches include meta-analysis, functional enrichment analysis, interactome analysis, transcriptional

network analysis, and integrative model system analysis. In addition, comparative analysis, combining human data with model organisms, can lead to more robust findings [134, 135].

Such methods that analyze biological processes in terms of higher level modules can identify robust signatures of disease mechanisms. Application of these methods to understand human cancers have delineated molecular subtypes of cancer associated with disease progression and treatment response [76–78, 136, 137].

A closely integrated data warehouse has been constructed to link different kinds and numbers of biological networks to experimental results such as those coming from microarrays. The basic idea is to consider and store biological networks as graphs in which the nodes represent promoters, proteins, genes, and transcripts, and the edges specify the relationship between the various nodes. Direct links to underlying sequences (exons, introns, promoters, amino acid sequences) in a systematic way enable close interoperability to sequence analysis methods. This approach allows us to store, query, and update a wide variety of biological information in a way that is semantically compact without requiring changes at the database schema level when new kinds of biological information are added. Such a system can be set up using software available from the public domain [138].

### 5.3 Drug Discovery and Development

***Target Discovery.*** Microarrays are increasingly being used in drug discovery for a wide range of applications, including target discovery and selection, pharmacology, and toxicogenomics [139]. The objective of using transcript profiling in target discovery is to identify a shortlist of candidate genes with distinct expression patterns during disease. Typical selection criteria include disease-specific changes in expression levels and tissue or cell-type selectivity [140]. A secondary screen is then typically used to examine the role(s) of short-listed genes by analyzing additional samples, and identifying correlations with disease progression. Validation experiments are performed to follow up the candidate genes using animal disease models or mice in which the target gene has been knocked out.

Network analysis is especially relevant to antimicrobial drug discovery. The relatively small size of microbial genomes enables data from profiling experiments to be used for modeling gene networks because these would be several orders of magnitude lower in complexity compared with human cellular networks. In addition, the availability of whole-genome nucleotide sequence data from a growing list of microbial genomes allows designing of probes for DNA arrays. Such microarrays are already being used to obtain a global perspective of host–pathogen interactions. This is beginning to make an impact on our understanding of pathogenesis and the strategies taken to combat infectious diseases, including the identification of novel drug targets [141].

***Target Validation.*** Several approaches, including those that involve the use of arrays, can be taken to validate short-listed therapeutic candidates. Genes whose transcript profiles suggest additional analysis can be profiled using TMAs in tens to hundreds of tissues by immunohistochemistry [142]. TMAs can be used to identify heterogeneities between primary tumors and their metastases, as demonstrated in the analyses of erbB2 in primary and metastatic breast cancers

[143]. The expression of a target gene in normal tissue from various vital organs can also be assessed using TMA panels. It is well known that mRNA levels do not reflect protein quantities and function, so proteomic analysis may help in making better decisions on targets to be developed for therapeutics [144].

Cell-based gene knockouts can closely mimic the actions of potential drugs to identify phenotypic changes and potential side effects. The successful adaptation of RNA interference in a microarray format to facilitate efficient suppression of gene expression has enabled high-throughput target validation for a wide variety of cell types [85, 86].

***Pharmacology.*** A study that monitored the expression patterns of the entire complement of yeast genes under a variety of experimental conditions and genetic backgrounds was used to categorize drugs into various classes based on the expression pattern changes they induced. This database is being used to stratify novel drugs based on their mechanism of action by monitoring the changes in expression induced by them [1].

Drug safety in animals is tested by monitoring a diverse spectrum of events, most commonly liver and kidney toxicity, and fatty liver. Transcriptional activation of drug metabolizing enzymes in livers of drug-treated animals or primary human hepatocytes has been widely used to identify drug metabolizing pathways. This helps to identify drugs early in development that might have toxicity issues due to inadequate metabolism [145–147]. In human populations, certain polymorphisms in the cytochrome P450 genes result in slow drug metabolism leading to toxicity. Additionally, the identification of SNPs in the cytochrome P450 genes of patients during clinical trials can help to stratify patients in whom the treatment is likely to be effective [148].

## REFERENCES


1. Hughes T R, Marton M J, Jones A R, et al. (2000). Functional discovery via a compendium of expression profiles. *Cell*. 102:109–126.
2. Lipshutz R J, Fodor S P, Gingeras T R, et al. (1999). High density synthetic oligonucleotide arrays. *Nat. Genet*. 21:20–24.
3. Kidgell C, Winzeler E A (2005) Elucidating genetic diversity with oligonucleotide arrays. *Chromos. Res*. 13:225–235.
4. Gitan R S, Shi H, Chen C M, et al. (2002). Methylation-specific oligonucleotide microarray: a new potential for high-throughput methylation analysis. *Genome Res*. 12:158–164.
5. Lieb J D, Liu X, Botstein D, et al. (2001). Promoter-specific binding of Rap1 revealed by genome-wide maps of protein-DNA association. *Nature Genet*. 28:327–334.
6. Zhu H, Bilgin M, Bangham R (2001). Global analysis of protein activities using proteome chips. *Science*. 293:2101–2105.
7. Gutjahr C, Murphy D, Lueking A, et al. (2005). Mouse protein arrays from a TH1 cell cDNA library for antibody screening and serum profiling. *Genomics*. 85:285–296.
8. Huang J X, Mehrens D, Wiese R, et al. (2001). High-throughput Genomic and Proteomic Analysis Using Microarray Technology. *Clini. Chem*. 4:1912–1916.
9. Haab B B, Geierstanger B H, Michailidis G, et al. (2005). Immunoassay and antibody microarray analysis of the HUPO Plasma Proteome Project reference specimens:

Systematic variation between sample types and calibration of mass spectrometry data. *Proteomics*. 5:3278–3291.

10. Wiese R, Belosludtsev Y, Powdrill T, et al. (2001). Simultaneous multianalyte ELISA performed on a microarray platform. *Clin. Chem*. 47:1451–1457.
11. Kumble K D (2003). Protein microarrays: new tools for pharmaceutical development. *Analytical and Bioanalytical Chemistry*. 377:812–819.
12. de Leon J, Susce M T, Pan R M, et al. (2005). The CYP2D6 poor metabolizer phenotype may be associated with risperidone adverse drug reactions and discontinuation. *J. Clin. Psychia*. 66:15–27.
13. Perez E A, Pusztai L, Van de Vijver M (2004). Improving patient care through molecular diagnostics. *Semin Oncol*. 31:14–20.
14. Zhou H, Roy S, Schulman H, et al. (2001). Solution and chip arrays in protein profiling. *Trends Biotechnol*. 19:S34–S39.
15. Martins T B, Burlingame R, von Muhlen C A, et al. (2004). Evaluation of multiplexed fluorescent microsphere immunoassay for detection of autoantibodies to nuclear antigens. *Clin. Diagn. Lab. Immunol*. 11:1054–1059.
16. Available: NUNC Brand Products, Denmark. http://www.nuncbrand.com
17. Stillman B A, Tonkinson J L (2000). FAST slides: A novel surface for microarrays. *Biotechniques*. 29:630–635.
18. Kanda V, Kariuki J K, Harrison D J, et al. (2004). Label-free reading of microarray-based immunoassays with surface plasmon resonance imaging. *Analyt. Chem*. 76:7257–7262.
19. Manning M, Redmond G (2005). Formation and characterization of DNA microarrays at silicon nitride substrates. *Langmuir*. 21:395–402.
20. Venkatasubbarao S (2004). Microarrays-status and Prospects. *Trends Biotechnol*. 22:630–637.
21. Zander N (2004). New planar substrates for the *in situ* synthesis of peptide arrays. *Molec. Divers*. 8:189–195.
22. Angenendt P, Glokler J (2004). Evaluation of antibodies and microarray coatings as a prerequisite for the generation of optimized antibody microarrays. *Methods Molec. Bio*. 264:123–134.
23. Schena M, Shalon D, Davis R W, et al. (1995). Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science*. 270:467–470.
24. Vaidya A A, Norton M L (2004). DNA attachment chemistry at the flexible silicone elastomer surface: toward disposable microarrays. *Langmuir*. 20:11100–11107.
25. Meredith G D, Wu H Y, Allbritton N L (2004). Targeted protein functionalization using His-tags. *Bioconjug. Chem*. 15:969–982.
26. Wacker R, Schroder H, Niemeyer C M (2004). Performance of antibody microarrays fabricated by either DNA-directed immobilization, direct spotting, or streptavidin-biotin attachment: A comparative study. *Analyt. Biochem*. 330:281–287.
27. Chrisey L A, Lee G U, O'Ferrall C E (1996). Covalent attachment of synthetic DNA to self-assembled monolayer films. *Nucleic Acids Res*. 24:3031–3039.
28. Rogers Y H, Jiang-Baucom P, Huang Z J, et al. (1999). Immobilization of oligonucleotides onto a glass support via disulfide bonds: A method for preparation of DNA microarrays. *Analyt. Biochem*. 266:23–30.
29. Kusnezow W, Hoheisel J D (2003) Solid supports for microarray immunoassays. *J. Biomolec. Recog*. 16:165–176.

30. Springer A L, Gall A S, Hughes K A, et al. (2003). Salicylhydroxamic acid functionalized affinity membranes for specific immobilization of proteins and oligonucleotides. *J. Biomolec. Technol.* 14:183–190.
31. Angenendt P, Glokler J, Sobek J, et al. (2003). Next generation of protein microarray support materials: evaluation for protein and antibody microarray applications. *J. Chromatog. A.* 1009:97–104.
32. Guschin D, Yershov G, Zaslavsky A, et al. (1997). Manual manufacturing of oligonucleotide, DNA, and protein microchips. *Analyt. Biochem.* 250:203–211.
33. Moerman R, Frank J, Marijnissen J C M, et al. (2001). Miniaturized electrospraying as a technique for the production of microarrays of reproducible micrometer-sized protein spots. *Analyt. Chem.* 73:2183–2189.
34. Ringeisen B R, Wu P K, Kim H, et al. (2002). Picoliter-scale protein microarrays by laser direct write. *Biotechnol. Prog.* 18:1126–1129.
35. Fodor S P, Read J L, Pirrung M C, et al. (1991). Light-directed, spatially addressable parallel chemical synthesis. *Science.* 251:767–773.
36. Lockhart D J, Dong H, Byrne M C, et al. (1996). Expression monitoring by hybridization to high-density oligonucleotide arrays. *Nature Biotechnol.* 14:1675–1680.
37. Singh-Gasson S, Green R D, Yue Y, et al. (1999). Maskless fabrication of light-directed oligonucleotide microarrays using a digital micromirror array. *Nature Biotechnol.* 17:974–980.
38. Gao X, LeProust E, Zhang H, et al. (2001). A flexible light-directed DNA chip synthesis gated by deprotection using solution photogenerated acids. *Nucleic Acids Res.* 29:4744–4750.
39. Pellois J P, Zhou X, Srivannavit O, et al. (2002) Individually addressable parallel peptide synthesis on microchips. *Nature Biotechnol.* 20:922–926.
40. Reimer U, Reineke U, Schneider-Mergener J (2002). Peptide arrays: from macro to micro. *Cur. Opin. Biotechnol.* 13:315–320.
41. Xu Q, Lam K S (2003) Protein and Chemical Microarrays—Powerful Tools for Proteomics. *J. Biomed. Biotechnol.* 3:257–266.
42. Matsuzaki H, Loi H, Dong S, et al. (2004). Parallel genotyping of over 10,000 SNPs using a one-primer assay on a high-density oligonucleotide array. *Genome Res.* 14:414–425.
43. Gunderson K L, Steemers F J, Lee G, et al. (2005). A genome-wide scalable SNP genotyping assay using microarray technology. *Nature Gen.* 37:549–554.
44. Bignell G R, Huang J, Greshock J, et al. (2004). High-resolution analysis of DNA copy number using oligonucleotide microarrays. *Genome Res.* 14:287–295.
45. Iyer V R, Horak C E, Scafe C S, et al. (2001). Genomic binding sites of the yeast cell-cycle transcription factors SBF and MBF. *Nature.* 409:533–538.
46. Huang T H, Perry M R, Laux D E (1999). Methylation profiling of CpG islands in human breast cancer cells. *Human Molec. Gen.* 8:459–470.
47. Duggan D J, Bittner M, Chen Y, et al. (1999). Expression profiling using cDNA microarrays. *Nature Gen.* 21:10–14.
48. Grunenfelder B, Winzeler E A (2002). Treasures and traps in genome-wide data sets: case examples from yeast. *Nature Rev. Gen.* 3:653–661.
49. Manduchi E, Scearce L M, Brestelli J E, et al. (2002). Comparison of different labeling methods for two-channel high-density microarray experiments. *Physiol. Genom.* 10:169–179.
50. Brazma A, Hingamp P, Quackenbush J, et al. (2001). Minimum information about a microarray experiment (MIAME)-toward standards for microarray data. *Nat Genet.* 29:365–371.

51. Lockhart D J, Dong H, Byrne M C, et al. (1996). Expression monitoring by hybridization to high-density oligonucleotide arrays. *Nature Biotechnol*. 14:1675–1680.
52. Yauk C L, Berndt M L, Williams A, et al. (2004). Comprehensive comparison of six microarray technologies. *Nucleic Acids Res*. 32:e124.
53. Ekins R P. (1998). Ligand assays: from electrophoresis to miniaturized microarrays. *Clin. Chem*. 44:2015–2030.
54. Tam S W, Wiese R, Lee S, et al. (2002). Simultaneous analysis of eight human Th1/Th2 cytokines using microarrays. *J. Immunolog. Meth*. 261:157–165.
55. Kastenbauer S, Angele B, Sporer B, (2005). Patterns of protein expression in infectious meningitis: A cerebrospinal fluid protein array analysis. *J. Neuroimmunol*. 164:134–139.
56. Nielsen U B, Cardone M H, Sinskey A J, et al. (2003). Profiling receptor tyrosine kinase activation by using Ab microarrays. *Proc. of the National Academy of Sciences U S A*. 100:9330–9335.
57. Sreekumar A, Nyati M K, Varambally S, et al. (2001). Profiling of cancer cells using protein microarrays: discovery of novel radiation-regulated proteins. *Cancer Res*. 61:7585–7593.
58. Holloway A J, van Laar R K, Tothill R W, et al. (2002). Options available—from start to finish—for obtaining data from DNA microarrays II. *Nature Gen*. 32:481–489.
59. Forster T, Costa Y, Roy D, et al. (2004). Triple-target microarray experiments: a novel experimental strategy. *BMC Genomics*. 5:13.
60. Mendoza L G, McQuary P, Mongan A, et al. (1999). High-throughput microarray-based enzyme-linked immunosorbent assay (ELISA). *Biotechniques*. 27:778–780.
61. Paweletz C P, Charboneau L, Bichsel V E, et al. (2001). Reverse phase protein microarrays which capture disease progression show activation of pro-survival pathways at the cancer invasion front. *Oncogene*. 20:1981–1989.
62. Falsey J R, Renil M, Park S, et al. (2001). Peptide and small molecule microarray for high throughput cell adhesion and functional assays. *Bioconjug. Chem*. 12:346–353.
63. Bobrow M N, Shaughnessy K J, Litt G J (1991). Catalyzed reporter deposition, a novel method of signal amplification. II. Application to membrane immunoassays. *J. Immunolog. Meth*. 137:103–112.
64. Schweitzer B, Roberts S, Grimwade B, et al. (2002). Multiplexed protein profiling on microarrays by rolling-circle amplification. *Nature Biotechnol*. 20:359–365.
65. Stears R L, Getts R C, Gullans S R (2000). A novel, sensitive detection system for high-density microarrays using dendrimer technology. *Physiolog. Genom*. 3:93–99.
66. Ault-Riche D, Atkinson B, Kumble K D (2005). *Methods for Structural analysis of Protein Pharmaceuticals*. AAPS Press, Arlington, V.A. pp: 545–571.
67 Kanda V, Kariuki J K, Harrison D J, et al. (2004). Label-free reading of microarray-based immunoassays with surface plasmon resonance imaging. *Analyt. Chem*. 76:7257–7262.
68. Striebel H M, Schellenberg P, Grigaravicius P, et al. (2004). Readout of protein microarrays using intrinsic time resolved UV fluorescence for label-free detection. *Proteomics*. 4:1703–1711.
69. Geschwind D H (2001). Sharing gene expression data: An array of options. *Nature Rev. Neurosci*. 2:435–438.
70. Bilban M, Buehler L K, Head S, et al. (2002). Defining signal thresholds in DNA microarrays: exemplary application for invasive cancer. *BMC Genomics*. 3:19.
71. Slonim D K. (2002). From patterns to pathways: gene expression data analysis comes of age. *Nature Genetics*. 32:502–508.

72. Xu Y, Selaru F M, Yin J, et al. (2002). Artificial neural networks and gene filtering distinguish between global gene expression profiles of Barrett's esophagus and esophageal cancer. *Cancer Res*. 62:3493–3497.
73. Brown M P, Grundy W N, Lin D, et al. (2002). Knowledge-based analysis of microarray gene expression data by using support vector machines. *Proc. of the National Academy of Sciences U S A*. 97:262–267.
74. Harbig J, Sprinkle R, Enkemann S A (2005). A sequence-based identification of the genes detected by probesets on the Affymetrix U133 plus 2.0 array.*Nucleic Acids Res*. 33:e31.
75. Petricoin E F III, Hackett J L, Lesko L J, et al. (2002). Medical applications of microarray technologies: a regulatory science perspective. *Nature Gen*. 32:474–479.
76. Golub T R, Slonim D K, Tamayo P (1999). Molecular classification of cancer: Class discovery and class prediction by gene expression monitoring. *Science*. 286:531–537.
77. Alizadeh A A, Eisen M B, Davis R E, et al. (2000). Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature*. 403:503–511.
78. van de Vijver M J, He Y D, van't Veer L J, et al. (2002). A gene-expression signature as a predictor of survival in breast cancer. *New Eng. J. Med*. 347:1999–2009.
79. van't Veer L J, Paik S, Hayes D F (2005). Gene expression profiling of breast cancer: a new tumor marker. *J. Clin. Oncol*. 23:1631–1635.
80. Pollack J R, Iyer V R (2002). Characterizing the physical genome. *Nature Genetics*. 32:515–521.
81. Shi M M (2001). Enabling large-scale pharmacogenetic studies by high-throughput mutation detection and genotyping technologies. *Clin. Chem*. 47:164–172.
82. Cowell J K (2005). High throughput determination of gains and losses of genetic material using high resolution BAC arrays and comparative genomic hybridization. *Combin. Chem. High Throughput Screen*. 7:587–596.
83. Fehlbaum P, Guihal C, Bracco L, et al. (2005). A microarray configuration to quantify expression levels and relative abundance of splice variants. *Nucl. Acids Res*. 33:e47.
84. Ziauddin J, Sabatini D M (2001). Microarrays of cells expressing defined cDNAs. *Nature*. 411:107–110.
85. Wheeler D B, Carpenter A E, Sabatini D M (2005). Cell microarrays and RNA interference chip away at gene function. *Nature Gen*. 37:S25–S30.
86. Vanhecke D, Janitz M (2004). High-throughput gene silencing using cell arrays. *Oncogene*. 23:8353–8358.
87. Patterson S D, Aebersold R H (2003). Proteomics: the first decade and beyond. *Nature Gen*. 33:311–323.
88. Michaud G A, Salcius M, Zhou F, et al. (2003). Analyzing antibody specificity with whole proteome microarrays. *Nature Biotechnol*. 21:1509–1512.
89. Madoz-Gurpide J, Wang H, Misek D E, et al. (2001). Protein based microarrays: a tool for probing the proteome of cancer cells and tissues. *Proteom*. 1:1279–1287.
90. Haab B B (2005). Antibody arrays in cancer research. *Molec. Cell. Proteom*. 4:377–383.
91. Wingren C, Ingvarsson J, Lindstedt M, et al. (2003). Recombinant antibody microarrays—a viable option? *Nature Biotechnol*. 21:223.
92. de Wildt R M, Mundy C R, Gorick B D, et al. (2000). Antibody arrays for high-throughput screening of antibody-antigen interactions. *Nature Biotechnol*. 18:989–994.
93. Bacarese-Hamilton T, Mezzasoma L, Ingham C, et al. (2002). Detection of allergen-specific IgE on microarrays by use of signal amplification techniques. *Clin. Chem*. 48:1367–1370.

94. Mezzasoma L, Bacarese-Hamilton T, Di Cristina M (2002). Antigen microarrays for serodiagnosis of infectious diseases. *Clin. Chem.* 48:121–130.
95. Joos T O, Schrenk M, Hopfl P (2000). A microarray enzyme-linked immunosorbent assay for autoimmune diagnostics. *Electrophor.* 21:2641–2650.
96. Jain K K (2004). Applications of biochips: from diagnostics to personalized medicine. *Curr. Opin. Drug Disc. Devel.* 7:285–289.
97. Robinson W H, DiGennaro C, Hueber W (2002). Autoantigen microarrays for multiplex characterization of autoantibody responses. *Nature Med.* 8:295–301.
98. Robinson W H, Fontoura P, Lee B J (2003). Protein microarrays guide tolerizing DNA vaccine treatment of autoimmune encephalomyelitis. *Nature Biotechnol.* 21:1033–1039.
99. Ge H (2000). UPA, a universal protein array system for quantitative detection of protein-protein, protein-DNA, protein-RNA and protein-ligand interactions. *Nucl. Acids Res.* 28:e3.
100. Houseman B T, Huh J H, Kron S J, et al. (2002). Peptide chips for the quantitative evaluation of protein kinase activity. *Nature Biotechnol.* 20:270–274.
101. Zhu H, Klemic J F, Chang S (2000). Analysis of yeast protein kinases using protein chips. *Nature Gen.* 26:283–289.
102. MacBeath G, Schreiber S L (2000). Printing proteins as microarrays for high-throughput function determination. *Science.* 289:1760–1763.
103. Bieri C, Ernst O P, Heyse S, et al. (1999). Micropatterned immobilization of a G protein-coupled receptor and direct detection of G protein activation. *Nature Biotechnol.* 17:1105–1108.
104. Fang Y, Webb B, Hong Y (2004). Fabrication and application of G protein-coupled receptor microarrays. *Methods Molec. Biol.* 264:233–243.
105. Falsey J R, Renil M, Park S, et al. (2001). Peptide and small molecule microarray for high throughput cell adhesion and functional assays. *Bioconjug. Chem.* 12:346–353.
106. Toepert F, Knaute T, Guffler S (2003). Combining SPOT synthesis and native peptide ligation to create large arrays of WW protein domains. *Angewandte Chemie Internat. Ed.* 42:1136–1140.
107. Reineke U, Kramer A, Schneider-Mergener J (1999). Antigen sequence- and library-based mapping of linear and discontinuous protein-protein-interaction sites by spot synthesis. *Curr. Topics Microbiol. Immunol.* 243:23–36.
108. Houseman B T, Huh J H, Kron S J, et al. (2002). Peptide chips for the quantitative evaluation of protein kinase activity. *Nature Biotechnol.* 20:270–274.
109. Reineke U, Sabat R, Kramer A, et al. (1996). Mapping protein-protein contact sites using cellulose-bound peptide scans. *Molec. Divers.* 1:141–148.
110. Salisbury C M, Maly D J, Ellman J A (2002). Peptide microarrays for the determination of protease substrate specificity. *J. Amer. Chem. Soc.* 124:14868–14870.
111. Soen Y, Chen D S, Kraft D L, et al. (2003). Detection and characterization of cellular immune responses using peptide-MHC microarrays. *PLoS Biol.* 1:E65.
112. Weng S, Gu K, Hammond P W, et al. (2002). Generating addressable protein microarrays with PROfusion covalent mRNA-protein fusion technology. *Proteomics.* 2:48–57.
113. Brody E N, Willis M C, Smith J D, et al. (1999). The use of aptamers in large arrays for molecular diagnostics. *Molec. Diagnos.* 4:381–388.
114. Collett J R, Cho E J, Lee J F (2005). Functional RNA microarrays for high-throughput screening of antiprotein aptamers. *Analyt. Biochem.* 338:113–123.
115. Braunschweig T, Chung J Y, Hewitt S M (2004). Perspectives in tissue microarrays. *Combinat. Chem. High Throughput Screen.* 7:575–585.

116. Callagy G, Cattaneo E, Daigo Y (2003). Molecular classification of breast carcinomas using tissue microarrays. *Diagnos. Molec. Pathol.* 12:27–34.
117. Miyaji T, Hewitt S M, Liotta L A, et al. (2002). Frozen protein arrays: a new method for arraying and detecting recombinant and native tissue proteins. *Proteomics.* 2:1489–1493.
118. Barnes-Seeman D, Park S B, Koehler A N, et al. (2003). Expanding the functional group compatibility of small-molecule microarrays: discovery of novel calmodulin ligands. *Angewandte Chemie Internat. Ed.* 42:2376–2379.
119. Koehler A N, Shamji A F, Schreiber S L (2003). Discovery of an inhibitor of a transcription factor using small molecule microarrays and diversity-oriented synthesis. *J. Amer. Chem. Soc.* 125:8420–8421.
120. Gosalia D N, Diamond S L (2003). Printing chemical libraries on microarrays for fluid phase nanoliter reactions. *Proc. of the National Academy of Sciences U S A.* 100:8721–8726.
121. Winssinger N, Ficarro S, Schultz P G, et al. (2002). Profiling protein function with small molecule microarrays. *Proc. of the National Academy of Sciences U S A.* 99:11139–11144.
122. Bishop A C, Ubersax J A, Petsch D T, et al. (2000). A chemical switch for inhibitor-sensitive alleles of any protein kinase. *Nature.* 407:395–401.
123. Shin I, Cho J W, Boo D W (2004). Carbohydrate arrays for functional studies of carbohydrates. *Combinat. Chem. High Throughput Screen.* 7:565–574.
124. Disney M D, Seeberger P H (2004). The use of carbohydrate microarrays to study carbohydrate-cell interactions and to detect pathogens. *Chem. Biol.* 11:1701–1707.
125. Ngundi M M, Taitt C R, McMurry S A (2005). Detection of bacterial toxins with monosaccharide arrays. *Biosens. Bioelectr.* 21:1195–1201.
126. Simon R, Radmacher M D, Dobbin K (2002). Design of studies using DNA microarrays. *Genet. Epidemiol.* 23:21–36.
127. Dobbin K, Simon R (2005). Sample size determination in microarray experiments for class comparison and prognostic classification. *Biostatist.* 6:27–38.
128. Churchill G A (2002). Fundamentals of experimental design for cDNA microarrays. *Nature Genet.* 32:490–495.
129. Yang Y H, Speed T (2002). Design issues for cDNA microarray experiments. *Nature Rev. Genet.* 3:579–588.
130. Kerr M K, Churchill G A (2001). Statistical design and the analysis of gene expression microarray data. *Genet. Res.* 77:123–128.
131. Dobbin K, Shih J H, Simon R (2003). Questions and answers on design of dual-label microarrays for identifying differentially expressed genes. *J. Nation. Cancer Instit.* 95:1362–1369.
132. Kerr M K, Churchill G A (2001). Experimental design for gene expression microarrays. *Biostatist.* 2:183–201.
133. Eckel-Passow J E, Hoering A, Therneau T M, et al. (2005). Experimental design and analysis of antibody microarrays: Applying methods from cDNA arrays. *Cancer Res.* 65:2985–2989.
134. Rhodes D R, Chinnaiyan A M (2005). Integrative analysis of the cancer transcriptome. *Nature Genet.* 37:S31–S37.
135. Segal E, Friedman N, Kaminski N (2005). From signatures to models: understanding cancer using microarrays. *Nature Genet.* 37:S38–S45.
136. Sorlie T (2004). Molecular portraits of breast cancer: tumour subtypes as distinct disease entities. *Euro. J. Cancer.* 40:2667–2675.

137. Greer B T, Khan J (2004). Diagnostic classification of cancer using DNA microarrays and artificial intelligence. *Annals N. Y. Acad. Sci.* 1020:49–66.
138. Koehler J, Rawlings C, Verrier P (2005). Linking experimental results, biological networks and sequence analysis methods using Ontologies and Generalised Data Structures. *In Silico Biol.* 5:33–44.
139. Butte A (2002). The use and analysis of microarray data. *Nature Rev. Drug Disc.* 1:951–960.
140. Gerhold D L, Jensen R V, Gullans S R (2002). Better therapeutics through microarrays. *Nature Genet.* 32:547–551.
141. Shaw K J, Morrow B J (2003). Transcriptional profiling and drug discovery. *Curr. Opin. Pharmacol.* 3:508–512.
142. Sauter G, Simon R, Hillan K (2003). Tissue microarrays in drug discovery. *Nature Rev. Drug Disc.* 2:962–972.
143. Simon R, Nocito A, Hubscher T (2001). Patterns of her-2/neu amplification and overexpression in primary and metastatic breast cancer. *J. Nat. Cancer Inst.* 93:1141–1146.
144. Griffin T J, Gygi S P, Ideker T (2002). Complementary profiling of gene expression at the transcriptome and proteome levels in Saccharomyces cerevisiae. *Molec. Cell. Proteom.* 1:323–333.
145. Kato N, Shibutani M, Takagi H (2004). Gene expression profile in the livers of rats orally administered ethinylestradiol for 28 days using a microarray technique. *Toxicol.* 200:179–192.
146. Meneses-Lorente G, de Longueville F, Dos Santos-Mendes S (2003). An evaluation of a low-density DNA microarray using cytochrome P450 inducers. *Chem. Res. Toxicol.* 16:1070–1077.
147. Harris A J, Dial S L, Casciano D A (2004). Comparison of basal gene expression profiles and effects of hepatocarcinogens on gene expression in cultured primary human hepatocytes and HepG2 cells. *Mutat. Res.* 549:79–99.
148. Rodrigues A D, Rushmore T H (2002). Cytochrome P450 pharmacogenetics in drug development: in vitro studies and clinical consequences. *Curr. Drug Metab.* 3:289–309.

# GENETIC MARKERS AND GENOTYPING ANALYSES FOR GENETIC DISEASE STUDIES

Anne E. Kwitek and Michael Olivier
*Medical College of Wisconsin, Milwaukee, Wisconsin*

**Chapter Contents**

1 Introduction
2 Simple Sequence Length Polymorphisms (SSLPs)
  2.1 General Description
  2.2 Discovery
  2.3 Methods for Genotyping
  2.4 Tools and Resources
3 Single Nucleotide Polymorphisms (SNPs)
  3.1 General Description
  3.2 Discovery
  3.3 Methods for Genotyping
  3.4 Linkage Disequilibrium and Haplotypes
4 Application for Genetic Analyses
  4.1 Linkage Analysis
  4.2 Association Analysis
  Internet Resources
  References

## 1 INTRODUCTION

Over the past decade, major research efforts have resulted in a wealth of sequence information for many genomes. In addition to the human genome, several other mammalian genomes have also been completely sequenced, providing a tremendous resource for biomedical researchers to examine gene sequences, their

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

regulation, and their role in health and disease. However, although we commonly talk about "the genome" when we refer to the genome sequence, every individual of a species has a slightly different genome sequence. It is this variation in the genome sequence that accounts for phenotypic differences, and for different susceptibilities to a large number of diseases. In humans, if we knew which sequence differences predisposed individuals to common human disorders such as hypertension or Alzheimer's disease, we could develop preventive treatments that could prevent disease, despite the genetic risk. It is this vision that has driven researchers involved in the Human Genome Project over the past years. Likewise, if we understood the genetic variation in model organisms, such as rats or mice, responsible for disease susceptibility or variability in drug metabolism and efficacy, we could use this information to test novel drugs more efficiently and target specialized drugs toward individuals with specific genetic predispositions.

Although there has been significant progress in biomedical research to suggest that this vision may one day become reality, we do not yet have a good understanding of the genetic differences in the genome sequence, and how they influence the development of disease. One reason why progress has been slow is the fact that there are so many differences. The sequence of any two human genomes differs in at least three million bases. Most differences in mammalian genomes are single base pair differences in the DNA sequence, but others include length variation in repeated sequences and chromosomal rearrangements. This presents a tremendous challenge to geneticists and biomedical researchers: How do you distinguish between DNA sequence variation that has no functional consequences and the differences that affect gene function and possibly the well-being of the organism? After all, only about 5% of any mammalian genome encodes for genes, and only small portions of the remaining 95%, have discernible functions. As a consequence, it is likely that most DNA sequence variation has no effect on the organism because it is located in these regions without (known) function.

To further identify the determinants of genetic disorders, different approaches have been used to zoom in on regions of the genome that contain the mutation(s) that are responsible for the genetic defect. Both commonly used approaches, family-based linkage studies and association analyses, require genetic markers, i.e., sequences that are variable between individuals and can be used to distinguish the DNA from different individuals. Even before the completion of the various genome sequencing projects, linkage analyses identified numerous loci in the genomes of both humans and model organisms that harbor genes involved in the development of a variety of common disorders. These studies used genetic polymorphisms, i.e., differences in the DNA sequence, to distinguish individuals with and without the disease and to identify regions shared by affected individuals. It is this initial use of DNA sequence variation as markers for genetic studies that this chapter will focus on. We will describe two types of genetic variation commonly used as markers in genetic studies, describe how they were discovered and can be interrogated by genotyping in large numbers of individuals, and will end by briefly describing the approaches of linkage and association studies that are aided by and rely on these genetic markers. Numerous review articles have summarized different aspects of this chapter, and we will focus primarily on the methodologies underlying the different genotyping approaches commonly used today in genetic laboratories.

## 2 SIMPLE SEQUENCE LENGTH POLYMORPHISMS (SSLPs)

### 2.1 General Description

Eukaryotic genomes contain interspersed repetitive sequence elements, many of which have no known function. Incredibly, these "nonfunctional" sequences comprise about 20% of the human genome [1]. It is thought that these repetitive motifs were generated by transposons or retrotransposons that replicate and then "jump" to other locations to propagate their genome, in the form of either DNA (in the case of transposons) or RNA (in the case of retrotransposons). Some repetitive sequence elements are long, about 1000–5000 base pairs (bp) in length, and they are known as LINES (long interspersed elements), whereas others are shorter, about 200–500 bp, and are known as SINES (short interspersed elements). The Alu element, approximately 300 bp in length, is the most abundant SINE in the human genome [2] and comprises about 3% of the human genome.

Another type of repetitive element consists of very short sequences (2 to 5 bp), arrayed in tandem. These sequences, called microsatellites, simple sequence length polymorphisms (SSLPs), simple sequence repeats (SSRs), or short tandem repeat polymorphisms (STRPs), tend to vary in unit size (from 5 to 30 units). The number of these tandem elements varies, and as a result, so does their length, making these sequences *polymorphic* (having different forms). The most abundant class of microsatellite repeat, is the dinucleotide repeat of cytosine and adenine, or (CA) · (GT) repeats [3, 4]. In the human genome, these dinucleotide repeat microsatellites occur every 30 to 60 kbp, corresponding to approximately 50,000 to 100,000 microsatellites in the 3 billion base pair human genome [5]. Other common microsatellites include trinucleotide and tetranucleotide repeats (3 and 4 bp repeats, respectively) [6].

The length of the microsatellite repeat is thought to be due to errors in DNA replication [7]. During DNA replication, the polymerase machinery is thought to "slip," either eliminating or adding repeat units. The deletion or addition of a single repeat unit occurs most frequently (91%) [8]. The frequency of this type of mutation, on average, is $1.2 \times 10^{-3}$ per microsatellite, translating to roughly 1 in 1000 meioses (as determined by examining how often a parent passes an expanded or contracted microsatellite repeat to his/her offspring) [8]. The range of mutation frequency for any given microsatellite is 0 to $8 \times 10^{-3}$ and seems to occur more frequently in the male vs. the female germline.

With a few exceptions, microsatellites do not have functional consequences on clinical outcomes. They are most commonly found in intergenic or intronic portions of a eukaryotic genome. However, it is important to keep in mind that, although the microsatellite is repeated throughout the genome, there is a unique sequence flanking it, corresponding to the location within the genome in which it integrated. Therefore, this class of polymorphic repeat is ideal for an assay that can tag a specific genome location for use in genetic studies, which will be discussed in detail in this chapter.

### 2.2 Discovery

The microsatellite was first discovered in 1981 by Miesfeld et al. [9] by screening a human gene clone library with a mouse 6-kbp (6 kilobase pair) ribosomal gene

nontranscribed spacer probe (rDNA NTS). They found this sequence was interspersed not only throughout the human and mouse genomes, but also in every other eukaryotic organism they studied—pigeon, frog, slime mold, and yeast. However, it was not identified in *Escherichia coli*. Therefore, the rDNA NTS has been conserved throughout all eukaryotes, but not down to prokaryotes. Miesfeld et al. then narrowed down the conserved element, which they called ECl (evolutionarily conserved), to a 251-bp fragment of the rDNA NTS that contained a tandem array of the (TG) · (CA) dinucleotide. However, the reason for its dramatic evolutionary conservation remained a mystery.

In 1989, two back-to-back manuscripts were published, by Weber and May [4] and Litt and Luty [3], reporting the use of microsatellites as genetic markers; i.e., they could be used to identify inheritance patterns from parent to offspring. By cloning and sequencing several microsatellites, it was found that each tandem repeat was flanked by a unique genomic sequence, thereby providing the ability to identify an individual's "genotype" at a specific DNA location (locus) containing a microsatellite. By sequencing 59 of these microsatellites, Weber and May found that more than half of them contained repeats of 13 or greater dinucleotides, with the longest being 30 units in length [4]. They went on to test 10 microsatellites for Mendelian inheritance. All microsatellites tested displayed expected Mendelian inheritance in three-generation families.

Figure 1 displays a cartoon example of the inheritance of a single SSLP marker. Each band represents the length of the microsatellite fragment, separated according to its length by gel electrophoresis. Each parent, being diploid, has two alleles for the microsatellite, each having a different number of dinucleotide repeats. The mother and father each contribute one of these alleles to their offspring, with a 50% probability of a child inheriting either allele. The inheritance of each allele from their parents is shown in the children, in the expected Mendelian ratios.

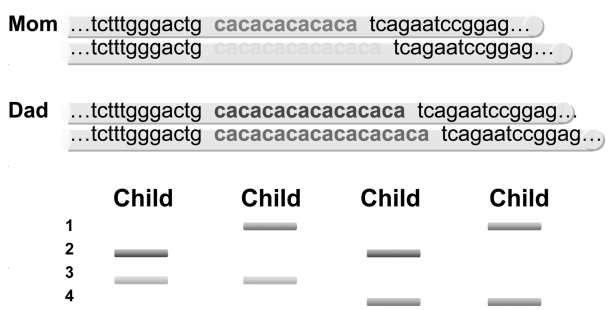


**Figure 1.** Schematic of SSLP genotypes. The top two bars represent a mother's chromosomal fragments containing two different alleles of a $(CA)_n$ repeat, the first having six tandem repeats (red allele) and the second having seven tandem repeats (green allele). The bottom two bars represent a father's alleles (eight repeats = blue; nine repeats = purple) for the same SSLP. The bottom represents genotypes from four children of the above parents, as they might appear on an electrophoretic gel. The four columns indicate four different children's genotypes for the SSLP, with the numbers indicating the four possible alleles. The colored bars correspond to the alleles inherited from the parents, representing all possible allele combinations according to Mendelian inheritance. (This figure is available in full color at ftp://ftp.wiley.com/public/sci_ tech_med/pharmaceutical_biotech/.)

The discovery of the microsatellite as a highly polymorphic genetic marker, and its ability to generate genotypes in a high-throughput manner, revolutionized genetic analyses of disease and greatly facilitated the mapping and sequencing of not only the human genome but also several model organisms, as discussed in the following sections.

### 2.3 Methods for Genotyping

The major advantage of genotyping using SSLPs over previously used genetic markers, e.g., restriction fragment length polymorphisms (RFLPs), is twofold. First, SSLPs are highly polymorphic, with alleles ranging anywhere from just a few to over 30 nucleotide repeat units. This greatly increases the likelihood that an individual will have alleles of differing sizes for any assayed SSLP, i.e., the marker is informative, and can be used across multiple families for genetic analyses [10]. In fact, many tens of thousands of SSLP assays have been developed for organisms ranging from human to mouse, rat, dog, cat, cow, chicken, pig, fish, frog, fruit fly, corn, rice, poplar, and yeast.

Second, the relatively short stretch of DNA containing a microsatellite allows it to be readily assayed using the polymerase chain reaction (PCR). PCR is a chemical reaction that can exponentially amplify DNA fragments [11]. In the case of SSLP assays, short oligonucleotide primers (~18–25 nucleotides in length) are generated that are complimentary to sequences flanking either side of the tandem repeat. The double-stranded genomic DNA is denatured (unraveled) by heating to 94°C, after which the reaction is cooled to approximately 55°C to 60°C. This temperature allows the oligonucleotide primers to anneal to the denatured genomic DNA. Finally the reaction is heated to 72°C, which allows a temperature stable polymerase, generally *Taq* polymerase isolated from a thermostable bacterium (*Thermus aquaticus*), to extend the oligonucleotide chain, using the complimentary genomic DNA strand as its template and the four dNTPs (dATP, dTTP, dCTP, and dGTP) as building blocks to extend the DNA fragment. This procedure is repeated multiple times, doubling the copy of the sequence contained between the primers with each cycle. Therefore, at the end of 30 cycles of consecutive denaturing, annealing, and extension, there are $2^{30}$ copies of the desired fragment, which is a sufficient amount of product for detection using radioactive isotopes or fluorescent dyes, as described below. Because PCR can be carried out in high-density microtiter plates (up to 1536 wells/plate) on commercial thermocyclers (machines that heat and cool the reaction samples required for PCR), genotyping of thousands of individuals can be performed in a single day.

***Radioactive Detection.*** Initially, PCR amplified product (amplicon) containing an SSLP, was most commonly detected using gel electrophoresis and a radioactively labeled nucleotide. This was done by either directly incorporating a radioactive tag in the PCR amplicon or by labeling the oligonucleotide primer. For direct incorporation, a radioactively tagged nucleotide, either $^{32}P$ or $^{35}S$ attached to the $\alpha$ position of one of the dNTPs (i.e., $\alpha^{32}P$-dCTP or $\alpha^{35}S$-dATP), is added to the PCR reaction (for detailed protocol, see Ref. 12). During each extension step of the reaction, labeled nucleotides are directly incorporated into the amplicon. An alternative approach used to detect the PCR amplicon is to label one primer used

in PCR. In this approach, $^{32}$P attached to the γ position of a dNTP (generally $\gamma^{32}$P-dATP or $\gamma^{32}$P-dCTP) is added to the 3′ end of one PCR primer, using T4 DNA kinase (for detailed protocol, see Ref. 12) before the PCR reaction. During PCR, this labeled primer is annealed to the target sequence and extended, thereby labeling the PCR amplicon. The advantage of direct incorporation is that only a single step amplifies and labels the desired PCR fragment; however, this method can also lead to labeling of nonspecific PCR product, making results more difficult to interpret. End-labeling of the PCR primer can eliminate the detection of nonspecific product, but also it requires high levels of radioactivity and an extra experimental step. However, both approaches have been used extensively and successfully, and which approach is used depends on individual preference.

To detect the radio-labeled SSLP amplicon, one must be able to separate the labeled fragments, based on their length. With radioactively labeled amplicons, this is done using denaturing gel electrophoresis. Reaction samples are mixed with a loading buffer containing blue tracking dye and formamide, and then heated to 94°C to denature the product. The formamide keeps the product denatured. Samples are then loaded onto a standard denaturing polyacrylamide sequencing gel. As the fragments migrate through the gel, the smaller sized fragments (containing fewer tandem repeats) migrate more quickly, whereas the larger fragments move more slowly, separating the two alleles. The blue dye allows tracking of the migration so that the product does not travel off the end of the gel. The polyacrylamide gel is then transferred to a piece of paper, dried, and exposed to autoradiographic film for detection and genotype determination. Figure 2 shows an example of an autoradiograph result of the radioactive genotyping method.

***Fluorescence Detection.*** The use of radioactivity has been largely supplanted by fluorescent dyes for genotyping, due to the obvious reduction of health hazards and a greater flexibility (as discussed later). As with radioactive labeling, the PCR amplification of the specific SSLP is largely the same, except that one dNTP is labeled with a fluorescent dye (fluorophore). Also, rather than detecting the electrophoresed product by autoradiography, the fluorescently labeled amplicon is

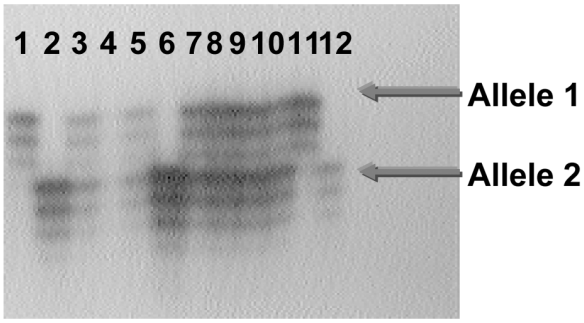


**Figure 2.** X-ray autoradiographic film from radioactive genotyping. As in Figure 1 each column (labeled lanes 1–12) is the genotype of a single SSLP marker in 12 individuals. The two possible alleles are indicated by the red arrows. (This figure is available in full color at ftp://ftp.wiley.com/public/sci_ tech_med/pharmaceutical_biotech/.)

detected as it passes a laser detector, either on a gel-based (e.g., an ABI 377 automated slab gel sequencer) or capillary-based DNA sequencer (e.g., an ABI 3700 or 3730 capillary sequencer). A major advantage of using fluorescent detection lies in the automation of genotype determination. As the labeled product passes by the laser detector, the signal is transferred to a computer algorithm that automatically converts the signal detection to a genotype. This greatly reduces manual genotype interpretation, reducing technical time and error as well as increasing overall genotyping throughput. Another important advantage of fluorescent geno-typing lies in the ability to label different SSLP amplicons with different fluorophores. Up to three different fluorophores are commonly used, 6-FAM, HEX, and NED, each of which has a distinct excitation and emission profile. A fourth dye, ROX, is used to label a common internal size standard, allowing accurate determination of product size [13]. The ability to distinguish multiple SSLP amplicons, each labeled with a different fluorophore, allows for simultaneous detection of multiple SSLPs in a single assay, increasing genotyping throughput. This approach is also amenable to high-throughput robotic sample preparation, as it eliminates the precautions necessary for handling radioactivity.

Again, as with radioactive detection, the fluorescent tags can either be directly incorporated into the amplicon or added to the PCR primer. A disadvantage of the labeled primer is that each SSLP must be synthesized with the fluorescent tag attached, which greatly increases the cost of the primer. Oetting et al. [14] devised a method to overcome this cost issue by combining unlabeled template-specific primers and a labeled, common M13 primer. By this method, one template-specific primer for each SSLP is synthesized with an 18-bp tail specific to the M13 phage on its 5′ end. This primer is combined with its template-specific mate, as well as a common M13 primer labeled with a particular fluorophore, and subject to PCR amplification. Figure 3 shows the general strategy for this method. The amplification begins with the tailed primer incorporating into the template sequence in the initial cycles. In subsequent reactions, the incorporated sequence becomes the docking site for the M13 primer-dye conjugate. The single M13 tailed primer-dye conjugate is common for all SSLPs, greatly reducing the cost of genotyping.

***Multiplex Analysis.*** As discussed, using fluorescent detection of SSLPs greatly improves genotyping throughput because of the availability of multiple fluorophores and the ability to multiplex (determine genotypes of multiple SSLPs simultaneously). The amplicon size of SSLPs varies not only within a single SSLP but also from one SSLP to another, depending on the distance from the repeat the flanking primers are chosen. This allows for the ability to detect multiple SSLPs within a single sample on a gel or capillary, as again each SSLP migrates according to size. For example, the allele size range for one SSLP amplicon may be 120–140 bp, whereas another may range from 200 to 250 bp, and a third may range from 325 to 341 bp. Therefore, all three amplicons could be combined in a single reaction, as the laser can detect the size differences and their ranges are sufficiently different to differentiate the genotypes of the three SSLPs. Furthermore, because the fluorophores have distinct excitation/emission profiles, multiple fluorophores can also be combined in a single sample. Therefore, by combining amplicons of distinct sizes labeled with one fluorophore with those labeled with other fluorophores, one

**Figure 3.** Schematic of fluorescent genotyping using an M13-tailed unique PCR primer (green with pink 5′ tail), an untagged unique PCR primer (green), and a common M13-dye conjugated primer (yellow). In the first PCR cycles (top), the unique primers hybridize to the DNA flanking either side of the $(CA)_n$ repeat. The M13 tag is incorporated into the PCR product (middle), creating a site for the fluorescently tagged complementary M13 primer to anneal to and label the PCR product with a particular fluorescent tag (FAM; dark blue star). This product is then detected by either a gel-based or capillary-based DNA sequencer. (This figure is available in full color at ftp://ftp.wiley.com/public/sci_tech_med/pharmaceutical_biotech/.)

can genotype up to nine SSLPs in a single sample. Figure 4 displays the results of multiplexing six SSLPs of different sizes and colors (indicating different fluorophores) on a single electrophoretic gel.

Two means of multiplexing are possible, either at the PCR or the electrophoresis steps. It is possible to combine PCR primers specific for multiple SSLPs in a single PCR reaction, whereby each SSLP is amplified simultaneously. The advantage of this approach is the reduction in PCR costs, as the overall reaction size is not modified. However, this approach requires up-front knowledge of the allele-size range of each SSLP in the samples being genotyped. Furthermore, some level of up-front optimization of the PCR reaction may be required so that each SSLP amplicon is generated with the same efficiency within the reaction. Once a set of two or three SSLPs are optimized for PCR multiplexing, these combinations are generally not modified. This approach is most efficient when the same set of SSLPs is used time and time again with varying DNA samples. The other multiplexing approach involves independent PCR amplification of each SSLP, followed by pooling of the amplicons before electrophoresis (gel or capillary). The advantage of this approach lies in the freedom to mix and match different marker sets. Marker sets may be slightly different depending on the population being genotyped. In some cases, the entire genome may be screened for initial linkage results. In other cases, an

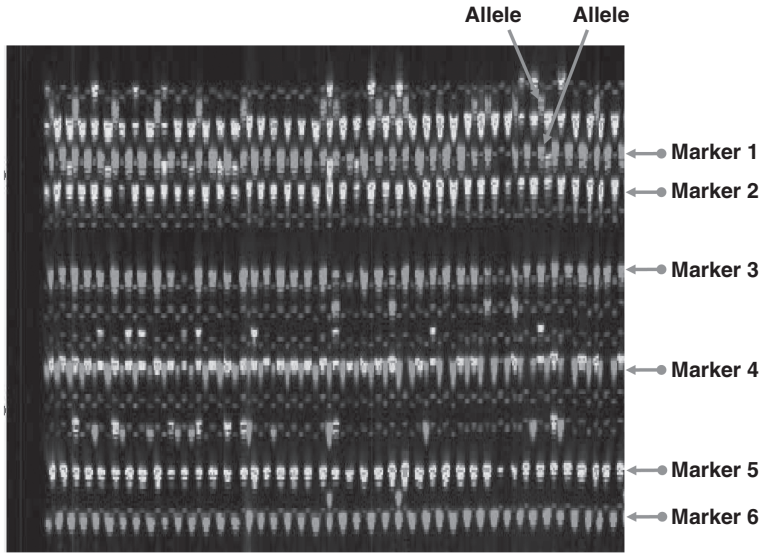


**Figure 4.** Multiplex genotyping results from an ABI 377 DNA sequencer. PCR product from six different SSLPs in the same DNA sample were pooled and electrophoresed on the ABI 377 slab gel sequencer. Three SSLPs of independent size ranges were labeled with FAM (blue), and three SSLPs of independent size ranges were labeled with NED (yellow) according to the method outlined in Figure 3. The red product is a ROX-labeled internal size standard that allows automated determination of allele sizes and conversion to individual genotypes. The laser detection can distinguish between the overlapping FAM-labeled and the NED-labeled SSLP products. (This figure is available in full color at ftp://ftp.wiley.com/public/sci_ tech_med/pharmaceutical_biotech/.)

individual locus may be the focus of higher density mapping, for instance, to narrow a disease gene interval by identifying critical recombinant individuals. This assay allows the individual to mix and match markers from a characterized set to do these specific scans. By this approach, multiple pooled amplicons, regardless of primer sequence, have a high success rate.

The use of fluorescent detection, combined with robotics and automated data analysis pipelines, facilitate highly automated, high-throughput genotyping platforms, resulting in thousands of genotypes per day at a relatively low cost.

## 2.4 Tools and Resources

To increase the efficiency of genotyping, major efforts have been made to generate a well-characterized and highly efficient marker set for genetic linkage analysis. One such effort was the Cooperative Human Linkage Center (CHLC). This National Institutes of Health (NIH)-funded project characterized thousands of di-, tri-, and tetranucleotide repeat polymorphisms to determine their informativeness (i.e., the number and frequency of SSLP alleles) [6, 10] and to array them according to their genetic position in the human genome (i.e., to develop high-quality genetic maps) [15]. Dr James Weber, a member of this Center, and his colleagues developed several genome-wide marker sets (http://research.marshfieldclinic.org/genetics/

Genotyping_Service/mgsver2.htm). Furthermore, the NIH went on to fund a centralized genotyping center, allowing investigators to submit research proposals for genome-wide linkage scans of their study populations (http://www.nhlbi.nih.gov/resources/medres/genotype.htm). A European group, Genethon, also led a major effort to develop human SSLPs and generate high-density genetic maps for use in genetic linkage studies [16]. These efforts have also reached the commercial arena, where companies such as DeCode will generate genotypes on a genome-wide or region-specific basis on a fee-for-service basis (http://www.decode.com/).

The discovery and use of microsatellites have revolutionized the ability to perform genome-wide genetic linkage studies, resulting in the positional cloning of many important genetic diseases ranging from mental retardation, glaucoma, heart defects, to neurological disease. They are particularly powerful for the localization and identification of single-gene disorders. Furthermore, the high-density genetic maps generated using microsatellite markers were instrumental to laying the groundwork for the eventual sequencing of not only the human genome, but also several other model organisms.

## 3 SINGLE NUCLEOTIDE POLYMORPHISMS (SNPs)

### 3.1 General Description

DNA is constantly modifying and mutating. During DNA replication, the duplication of the genetic material during cell divisions and inaccurate copying of the existing DNA leads to minor sequence changes in the copied version of the genome. Likewise, environmental influences such as mutagenic chemicals, ultraviolet (UV), or other radiation can also alter the DNA. On the basis of these constantly occurring changes in the DNA sequence of any organism, it becomes obvious that all individuals of a species are slightly different based on their DNA sequence. Most of these differences are single base changes in the DNA. Essentially one nucleotide in the sequence of one individual is replaced by a different nucleotide in another individual. These differences, illustrated in Figure 5, are called *single nucleotide*

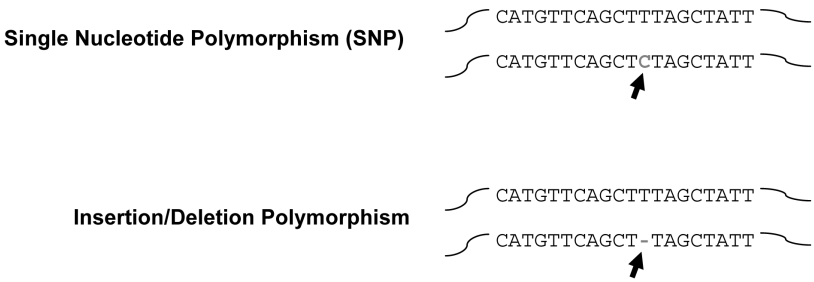


**Figure 5.** Single nucleotide polymorphisms and insertion/deletion polymorphisms. Two sections of a DNA strand are shown. The top of the figure shows a SNP, a single nucleotide difference between the two DNA strands (highlighted by a red nucleotide and the arrow). In the bottom diagram, one base is missing in the lower DNA strand (deletion polymorphism), leading to a frame shift in the DNA sequence. (This figure is available in full color at ftp://ftp.wiley.com/public/sci_ tech_med/pharmaceutical_biotech/.)

*polymorphisms* (SNPs). SNPs are the most common genome sequence variation. For the human genome, a comparison of two individuals reveals on average one SNP every 800 bases [17]. As a consequence, two humans differ in their DNA sequence at over 3 million positions, or approximately 0.1%. Similarly, other mammalian species show significant sequence variation as well. In mice, one SNP can be found every 250 bases across the genome when 13 inbred (i.e., genetically identical) and two wild strains are compared [18]. Inbred rat strains show a similar rate of SNPs [19]. These sequence differences are responsible for all genetically caused differences between individuals, such as eye color, hair or coat color, or disease susceptibility. If we could unravel the DNA sequence of every human individual, we would be able to decipher all genetic mutations that cause the most common diseases such as diabetes, stroke, or Alzheimer's disease. Unfortunately, though, the cost of sequencing every patient far exceeds the available budgets for these studies. Therefore, scientists have focused on discovering the most important and most common SNPs in the human genome, and now use this subset of SNPs in the human genome to investigate the genetic causes of these disorders [20]. Similar efforts are under way in other mammalian (and nonmammalian) species that serve as model organisms for specific diseases or traits in humans. Below we will describe and discuss the methods commonly used to discover SNPs, and the methodologies used to interrogate these SNPs in large numbers of individuals simultaneously by genotyping.

### 3.2 Discovery

Often, SNPs are discovered by sequencing the same genomic sequence in a number of unrelated samples. For this approach, a region of genomic DNA is amplified by PCR from different DNA samples. The resulting amplicons are resequenced using common sequencing technology. Approaches using gel-based resequencing and hybridization-based oligonucleotide array sequencing have successfully been used for large-scale SNP discovery efforts in the human and mouse genome as well as in other species. Analysis of the resulting sequences reveals base pair differences, or SNPs. Software tools have been developed to automate the SNP discovery from sequencing data [21], and these tools are routinely used today in a variety of organisms for SNP discovery. An example of an SNP as it would appear in sequencing data is shown in Figure 6.

Although sequencing is commonly used for SNP discovery today, other indirect methods exist to uncover genomic sequences containing SNPs. Given the cost of sequencing, it may be desirable to preselect these DNA fragments for subsequent sequencing, and to eliminate DNA sequences that do not contain polymorphisms.

One of the first reported methodologies for SNP discovery was the PCR-based single-stranded conformational polymorphism method [22]. In this approach, a region of DNA is amplified using PCR. The resulting amplicons are denatured, and they are allowed to re-anneal under conditions that favor the formation of single-stranded secondary structures. As the formation of secondary structures is sequence-dependent, any difference in the sequence of the PCR amplicon would lead to different folding. These differently folded products can be separated through non-denaturing polyacrylamide gel electrophoresis. Samples with different

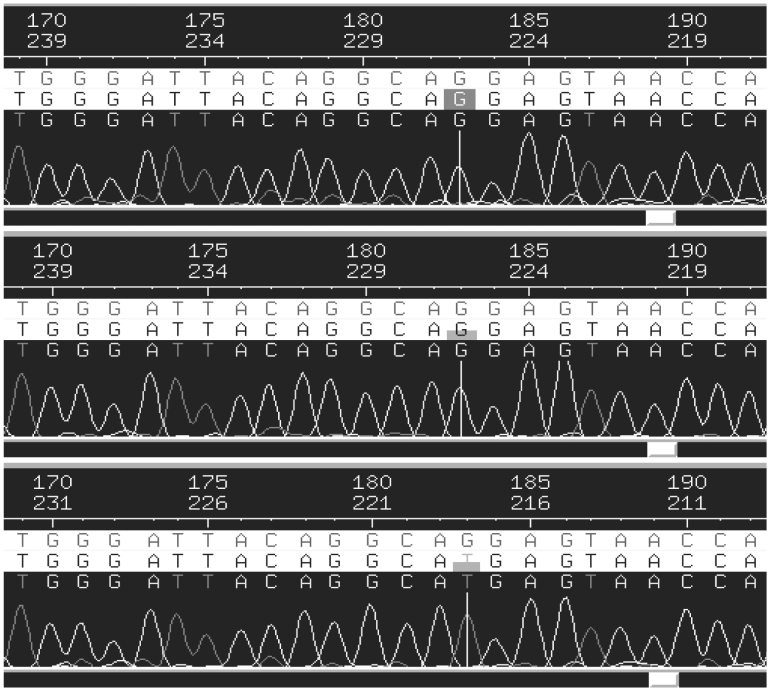


**Figure 6.** SNP discovery using the consed and polyphred software tools. The figure depicts three DNA traces from different individuals. Sequencing was performed using BigDye Terminator Chemistry. The white vertical bar highlights a variant nucleotide in the sequence. The top two traces have a "G" at this position, and the bottom trace shows a "T". (This figure is available in full color at ftp://ftp.wiley.com/public/sci_tech_med/pharmaceutical_biotech/.)

sequences will fold differently and, as a result, migrate differently in the gel, allowing for the detection of SNPs. This method has been used extensively for targeted screening of genes for potential disease-causing mutations in a variety of species. Recently, it has been adapted for capillary electrophoresis to facilitate large-scale screening (for a recent review, see Ref. 23).

Another approach to identify PCR amplicons that contain sequence variants uses denaturing high-performance liquid chromatography (DHPLC) [24]. In this approach, PCR amplicons are first denatured and then reannealed. In contrast to SSCP analysis, products are then allowed to form double-stranded duplexes. If the PCR product contains an SNP, this re-annealing process will generate homoduplexes (matching sequences reanneal) and heteroduplexes (strands containing the different two alleles of an SNP will anneal and form a single base pair mismatch). When these duplexes are separated on a liquid chromatography column under partially denaturing conditions (separation at temperatures close to the melting temperature of the duplex), the elution times of homo- and heteroduplexes will differ because the melting temperature of the heteroduplex will be slightly lower than the homoduplex due to the base pair mismatch in the sequence. The method and instrumentation has been automated, and it is commercially available. DHPLC has been used successfully for the identification of mutations in a large number of

disease genes, and it has been adapted for the analysis of cancer samples and chromosomal abnormalities [25–32]. Other modifications of the heteroduplex analysis exist, and they have been used by laboratories for a variety of studies. These include mobility analysis [33] or enzymatic cleavage of heteroduplexes [34].

One additional methodology for SNP discovery should be mentioned because it is based on an entirely different experimental approach. In 1995, Faham and Cox [35] described a method called mismatch repair detection (MRD) that used the bacterial mismatch repair system in *E. coli* to enrich for DNA fragments containing SNPs. The template for mismatch repair in *E. coli* is hemi-methylated double-stranded DNA (only one strand of the DNA is methylated), which is formed by mixing and annealing different single-stranded DNA samples grown in methylation-competent and -incompetent *E. coli* strains. The method has been developed into a high-throughput screening platform that can simultaneously analyze pooled PCR amplicons from large numbers of individuals [36, 37]. The method has been shown to enhance the discovery of rare SNPs that would be missed in pooled sequencing (or would require large numbers of individual samples to be sequenced).

In summary, several methods exist to identify SNPs in genomic DNA. Although some of them may be useful to enrich for SNP-containing PCR amplicons before traditional DNA sequencing, the commonly used approach of discovering SNPs from aligned sequence data from different individuals remains the "gold standard" and the basis of the majority of SNPs publicly available in databases.

### 3.3 Methods for Genotyping

Numerous methods have been developed and are being marketed for SNP genotyping, i.e., the specific determination of the nucleotide present at a polymorphic site in a genome. Theoretically, this information could be obtained by resequencing each and every sample using the same approach as for SNP discovery. However, this process is labor- and cost-intensive and provides significantly more information than is needed; i.e., it delineates the complete flanking sequence of the SNP, which, in most cases, will not differ between samples, and therefore is redundant. Instead, genotyping methods provide a cost-efficient alternative that can quickly generate genotype information on large numbers of samples. No attention is paid to the remainder of the genome sequence and its potential variation; the only information obtained is the nucleotide present in the sequence at the specific site of a known SNP. To develop methodologies for this genotyping procedure, the SNP in question, its possible nucleotides (alleles) that can be present at the site, and the sequence adjacent to the SNP need to be known. All methods discussed below use this information to obtain genotyping information for a large number of DNA samples. As outlined below, the different methods allow the analysis (genotyping) of individual SNPs in large numbers of samples, or varying numbers of SNPs for one sample at a time. All methods have been shown to result in reliable genotyping results with low error rates. The choice of the appropriate methodology primarily depends on the number of SNPs and the number of samples that need to be interrogated.

Essentially all genotyping approaches used to date fall into one of four categories: allele-specific hybridization, primer extension, oligonucleotide ligation, or invasive cleavage. Figure 7 illustrates these different approaches. We will first

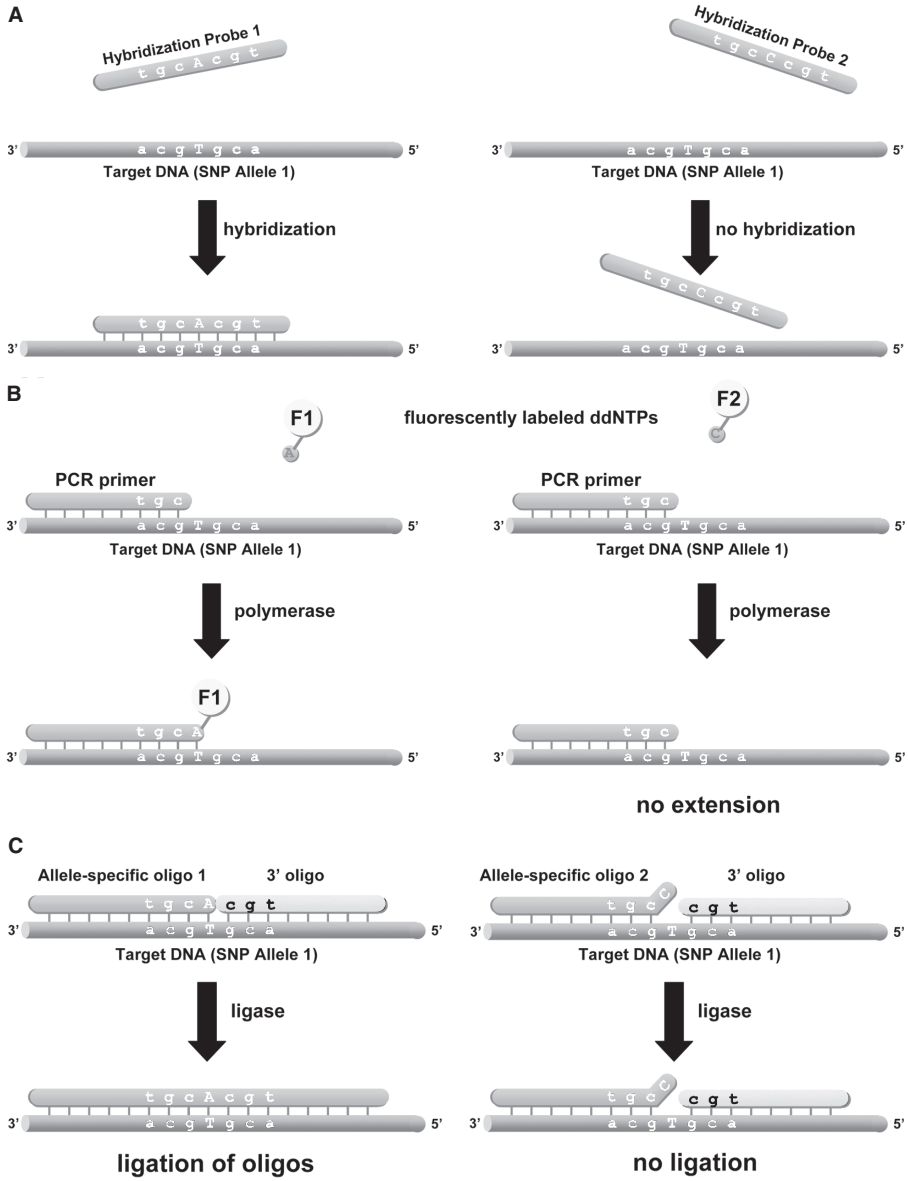


**Figure 7.** Schematic overview of different SNP genotyping approaches. (**A**) Allelic hybridization. (**B**) Primer extension. (**C**) Oligonucleotide ligation. (**D**) Invasive cleavage. (This figure is available in full color at ftp://ftp.wiley.com/public/sci_tech_med/pharmaceutical_biotech/.)

briefly describe the principles of each approach and then outline commercially available genotyping platforms based on these approaches for different study applications.

***Allele-specific Hybridization.*** Allele-specific hybridization distinguishes the allele present at an SNP using two different oligonucleotide probes. Each probe

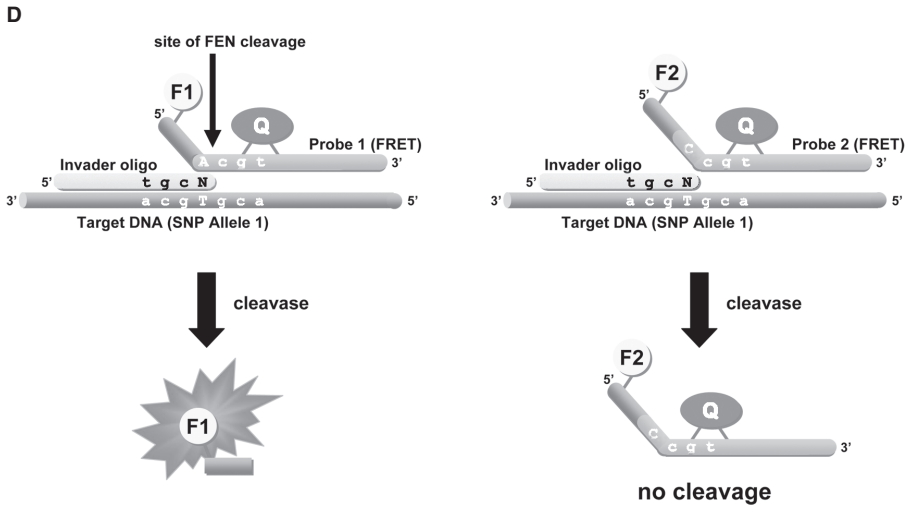


**Figure 7.** *Continued*

contains the SNP position in the middle and is complementary to one allele of the SNP. For hybridization to genomic DNA or PCR amplicons, conditions are chosen that allow the hybridization of the matching probe but not the mismatch probe. The successful hybrids between the target DNA and the oligonucleotide probes can be detected using a variety of approaches. Methods include the use of fluorescence resonance energy transfer (FRET) probes (e.g., in the 5′ exonuclease TaqMan assay, described below), molecular beacon probes, or oligonucleotide arrays (Affymetrix GeneChip system, see below).

***Primer Extension.*** Primer extension uses the standard principle of PCR. Here, an oligonucleotide primer extends up to the SNP site. The assay then determines what nucleotide is incorporated into the extended PCR product by the polymerase. Often, the assay is performed in two steps. First, the genomic interval is amplified by PCR. The amplified product is subsequently mixed with another primer that anneals adjacent to the SNP site. Then individual dideoxynucleotides are added to the reaction. Upon incorporation of the matching nucleotide at the SNP site, the resulting extension product can no longer be extended further because it lacks a 3′–OH group required for nucleotide incorporation. Thus, the reaction essentially stops once the nucleotide has been incorporated.

Dideoxynucleotides are commercially available with different fluorescent tags; therefore, the resulting extension products can be examined using laser scanners to see which fluorophore has been incorporated. This method directly reveals the genotype of the template DNA. The method can be multiplexed and is commercially available (SNaPshot, see below). Alternatively to the use of dideoxynucleotides, regular deoxynucleotides can be used sequentially in the extension reaction. Here, the primer is essentially extended as in a regular PCR reaction. The only difference is the sequential addition of one nucleotide at a time. This method is commercially available (see pyrosequencing, described below).

***Oligonucleotide Ligation.*** Similar to DNA polymerase, DNA ligase is a highly specific enzyme that repairs nicks in the DNA. Nicks are missing phosphodiester

bonds in one of the two strands of the DNA molecule. DNA ligase detects and repairs these nicks, provided the nucleotides on both sides of the missing bond are complementary to the nucleotides in the intact DNA strand. Landergren et al. [38] first described a novel approach for SNP genotyping that uses specific oligonucleotides that will hybridize to single-stranded DNA templates. One oligonucleotide will align with the sequence upstream of an SNP, including the actual nucleotide at the SNP site. A second oligonucleotide is complementary to the downstream template sequence, starting with the first nucleotide adjacent to the SNP. Although the latter oligonucleotide can be used to interrogate both alleles of the SNP, the first oligonucleotide is specific for one allele of the SNP, because it will end with the specific SNP nucleotide. Only the oligonucleotides perfectly matching the template sequence will be ligated by the ligase, whereas a mismatch will prevent the ligation of the alternate oligonucleotide. The method has been incorporated in a variety of commercial platforms such as the SNPlex platform or the Illumina GoldenGate technology, which is described below in more detail.

***Invasive Cleavage.*** The last SNP genotyping method we describe is invasive cleavage, which is commercially available as the Invader assay. This method depends on the ability of flap endonucleases to recognize specific three-dimensional structures that are formed when two overlapping oligonucleotides hybridize perfectly to a target DNA molecule [39]. The overlap occurs at the SNP site, and only the oligonucleotide combination that is complementary to the SNP allele will be cleaved by the enzyme. The reaction is highly specific, and a mismatch to the target DNA will prohibit the enzymatic reaction. This methodology has been adapted for SNP genotyping assays, as described in more detail below.

*Analysis of Individual SNPs.* Historically, investigators have used a variety of PCR-based methods to genotype individual SNPs. For all of these methods, the genomic sequence around the SNP was known and amplified using specific PCR primers that selectively amplified the genomic interval, including the SNP. SNPs that alter a restriction enzyme recognition site could then be genotyped by performing a restriction digest of the PCR product and by separating the resulting DNA segments by agarose gel electrophoresis. PCR products that contain the restriction enzyme recognition site will be cleaved into two products and will result in two bands of different size on the agarose gel. In contrast, PCR products from samples that contain an SNP in the recognition site will no longer be cleaved by the enzyme, and will result in a single band of larger molecular weight upon gel electrophoresis. Finally, PCR products from samples heterozygous for the SNP would result in a total of three bands, one band for the uncleaved product and two additional bands from the portion of the PCR product that can be cleaved. A schematic of the procedure is shown in Figure 8. The method enjoys widespread use to this day because it is robust and uses basic molecular biology technology available to most laboratories. With the ever increasing number of commercially available restriction enzymes uncovered from various microbes, the repertoire of SNPs that can be genotyped by this approach has been steadily increasing.

Unfortunately, many SNPs do not lead to an alteration of a restriction enzyme recognition site in the genome sequence. For these SNPs, alternative approaches have been developed that use PCR and gel electrophoresis. In one approach, one

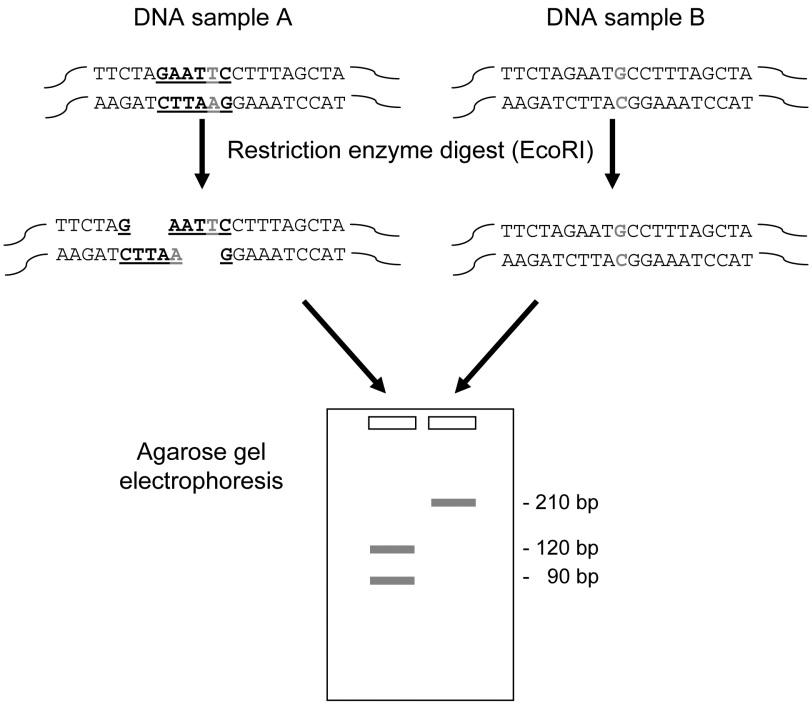


**Figure 8.** Schematic overview of PCR–RFLP. Double-stranded DNA molecules are shown for DNA sample A and B. These two samples are PCR amplicons from genomic DNA. The two sequences differ for the base pair highlighted in red. Sample A contains a restriction enzyme recognition site for *Eco*RI (underlined), and sample B does not. After restriction enzyme digestion, agarose gel electrophoresis results in different DNA fragments for the two samples. (This figure is available in full color at ftp://ftp.wiley.com/public/sci_tech_med/pharmaceutical_biotech/.)

initial PCR primer is designed so that it includes the polymorphic SNP site at its 3′ end. Two separate primers are used in two PCR reactions. One reaction includes the forward primer ending with one nucleotide present at the SNP, and the other reaction includes a primer ending with the alternative nucleotide. As the thermostable DNA polymerase used in PCR requires a perfect match of the primer at the 3′ end to attach additional nucleotides and amplify the DNA segment, the PCR reaction will only yield a PCR product when the forward primer perfectly matches the DNA template. Accordingly, the presence or absence of PCR product depending on the primer used indicates the genotype for the template DNA. Although this method works well for established assays, each assay needs to be optimized individually to ensure that both alleles of the SNP are successfully amplified when present. Furthermore, the design of the PCR primers may be difficult for some SNP sequences.

As an alternative to the allele-specific PCR and the restriction digest of PCR products (often called PCR–RFLP), several commercial platforms are available to genotype individual SNPs in large numbers of samples. In our discussion here, we will focus on three commonly used platforms: the 5′-exonuclease TaqMan assay,

based on allele-specific hybridization; pyrosequencing, based on the primer extension principle; and the Invader assay, based on the invasive cleavage reaction of a flap endonuclease.

***TaqMan Assay.*** The TaqMan assay uses the 5′ exonuclease activity of Taq polymerase [40]. Two FRET oligonucleotide probes complementary to the sequence around an SNP are hybridized to the target DNA. Each contains a different fluorescent dye that is quenched in the FRET probe. Each probe is specific for one of the two SNP alleles; i.e., it contains the complementary base to the SNP allele in the middle of the probe. The hybridization temperature is chosen such that only the matching probe hybridizes efficiently. After hybridization, a normal PCR extension reaction begins starting from an oligonucleotide primer annealed upstream of the FRET probe. The Taq polymerase extends the primer and, upon reaching the hybridized FRET probe, begins to remove it from the template DNA by cleaving one nucleotide at a time. The reaction is mediated by the specific 5′ exonuclease activity of the *Taq* polymerase. Once the entire oligonucleotide FRET probe has been cleaved, the fluorophore and the quencher molecule have been separated, and the energy transfer can no longer take place because the two molecules are no longer in close physical proximity. As a consequence, the fluorophore will now emit light upon excitation. As the two FRET probes contain different fluorophores depending on the allele, the resulting fluorescent color unambiguously determines the allele of the SNP in the target DNA.

TaqMan assays are commercially available for many SNPs and can be designed on demand by Applied Biosystems. The method requires a real-time PCR instrument or a fluorescent plate reader to measure the fluorescence of the two dyes at the end of the PCR reaction.

***Pyrosequencing.*** Pyrosequencing [41] has been described as a sequencing-by-synthesis method. In this approach, four different enzymes and specific substrates are used in a sequential reaction to produce light whenever a nucleotide complementary to the template DNA strand is incorporated. If the added nucleotide is not complementary to the base in the template DNA, no light is generated. The light-generating reaction is depended on the pyrophosphate that is released during the DNA polymerase reaction. If the added nucleotide forms a base pair, the DNA polymerase incorporates the nucleotide and pyrophosphate will be released. The released pyrophosphate will be converted to ATP by another enzyme (ATP sulfurylase). Luciferase then uses the ATP to generate a detectable light signal. The light intensity is proportional to the number of the ATP molecules, which in turn is proportional to the number of the incorporated nucleotides. After the completion of the reaction, the excess of each nucleotide is degraded by apyrase.

The reaction is repeated in the same order, with a different nucleotide added. If the added nucleotide does not form an incorporated base pair with DNA template, no light will be produced. The four nucleotides are added to the mixture in a defined order (e.g., ACGT), and thus, the sequence of the template DNA can be deduced by matching the nucleotide with the light signal generation.

Pyrosequencing provides rapid real-time determination of 20–30 nucleotides of target DNA sequence. With this technology, not only the SNP alleles but also the

adjacent nucleotides are determined, providing greater confidence in the correct SNP detection. The reaction is automated and requires a dedicated instrument.

***Invader Assay.*** The Invader assay [39] uses two oligonucleotide probes that are hybridized to target DNA containing an SNP site. The two oligonucleotides hybridize to the single-stranded target and form an overlapping invader structure at the site of the SNP, as illustrated in Figure 7. One oligonucleo-tide, the Invader oligo, is complementary to the target sequence 3′ of the polymorphic site, and it ends with a nonmatching base overlapping the SNP nucleotide. The second oligonucleotide, the allele-specific probe, contains the complementary base of the SNP allele, and extends to the sequence 5′ of the polymorphic site. This probe can also extend on its 5′ site with additional noncomplementary nucleotides. Once the two oligonucleotides anneal to the target DNA, they form a three-dimensional invader structure over the SNP site that can be recognized by a specific enzyme, cleavase, or flap endonuclease. The enzyme cleaves the probe 3′ of the base complementary to the polymorphic site (i.e., 3′ of the overlapping invader structure). If the probe is designed as a fluorescence resonance energy transfer (FRET) molecule containing a fluorophore at the 5′ end and an internal quencher molecule, the cleavage reaction will separate the fluorophore from the quencher and generate a fluorescent signal. If, in contrast, the probe oligonucleotide does not match the SNP allele present in the target DNA (i.e., the probe is complementary to the alternate SNP allele), then no overlapping invader structure is formed, and the probe is not cleaved. This distinction is highly specific, with only minimal unspecific cleavage of the mismatch probe.

Often, the Invader oligonucleotide is designed to permanently anneal to the target DNA at the assay temperature. In contrast, the probe oligonucleotide is designed to have a melting temperature close to the assay temperature. As a consequence, the probe constantly anneals and detaches. During the annealing, cleavase can cleave the oligonucleotide, the remnant detaches, and a new uncleaved oligonucleotide probe can reanneal to the same site. This design ensures the cleavage of a large number of probes, resulting in a strong signal even from small amounts of template DNA. Therefore, SNPs can be genotyped directly from genomic DNA using Invader assays, eliminating the need for additional PCR amplification. In addition to the FRET probes, other methods of detection (polarized light, mass spectrometry) have been used [42–44]. However, the benefit of the fluorescence detection is the ability to read the assay after isothermic incubation in a standard fluorescence plate reader.

*Multiplex Analysis.* Although the methods described above work well and reliably, they are not suited for genotyping multiple SNPs in large sets of samples. As each SNP has to be assayed individually, all SNPs have to be interrogated sequentially rather than in parallel. However, numerous other platforms exist that allow the genotyping of several SNPs at the same time. Here, we will discuss methods that allow up to 10 SNPs to be genotyped in one reaction. In addition, we will discuss high-throughput approaches in the next section.

Both methods described here are based on the primer extension reaction. In both platforms, oligonucleotide primers are designed that extend up to the nucleotide adjacent to the SNP site. The following primer extension reaction will use ddNTPs

to incorporate the matching base for the SNP allele present in the target DNA. The resulting products are then separated and detected by different methods.

The first detection method, as implemented in the MassEXTEND platform from Sequenom, uses mass spectrometry to determine the accurate molecular weight of the resulting extension reaction product, using matrix-assisted laser desorption/ionization time-of-flight (MALDI–TOF) mass spectrometry. For this approach, the reaction products are mixed with a matrix and deposited on the surface of a metal plate. The matrix and the reaction products are hit with a pulse from a laser beam, resulting in the vaporization of a small number of DNA molecules from the reaction product. The vaporized molecules are transferred into a vacuum flight tube and charged with an electrical field pulse. This accelerates the resulting ions toward the detector at the end of the tube. The time between the application of the electrical field pulse and the collision of the ions with the detector is referred to as the time of flight. As the time it takes for the ion to pass through the flight tube to the detector is directly proportional to the mass of the ion (with larger molecules flying slower), this is a very precise measure of the molecular weight of the DNA products. The detectable mass differences are small, and the difference between an incorporated ddA and ddT of 9 mass units can be clearly distinguished. The assay allows further multiplexing even for reactions resulting in very similar mass for the products by adding noncomplementary tail sequences to the 5′ end of the primer, thus increasing the mass of the extension product. A 12-plex has been described by Ross et al. [45] and even a 20-plex by Kim et al. [46].

Alternatively, the single base extension products can be analyzed using the SNaPShot technology from Applied Biosystems. In this approach, fluorescently labeled ddNTPs are used in the primer extension reaction, and the resulting extension products are separated and detected using commercially available capillary sequencers. Each ddNTP is labeled with a different fluorescent dye; thus, the different alleles of a SNP can be distinguished by the different color. Similar to the MassEXTEND approach, multiplexing can be improved by adding noncomplementary tail sequences to the primer. According to the manufacturer, 10-plex reactions can be performed with high accuracy.

*High-Throughput Approaches.* Currently, three high-throughput platforms for SNP genotyping are commercially available. All platforms allow the genotyping of large numbers of SNPs (1000–500,000) in a single experiment for one DNA sample. Although this significantly reduces the genotyping cost per SNP, it results in a tremendous cost for a project when 1000–2000 DNA samples need to be genotyped for thousands of SNPs. In addition, all approaches require specialized equipment and therefore are only of interest to laboratories that intend to perform large numbers of genotyping projects.

The first assay platform, developed by Illumina, combines allele-specific ligation with an extension reaction. The method, called the GoldenGate assay, uses the BeadArray technology, a method using fiberoptic substrates with randomly assembled arrays of beads. Detailed information and flow charts of the methodology can be found at the Illumina website (http://www.illumina.com/products/prod_snp.ilmn).

In the assay, three oligonucleotides are designed for each SNP locus. Two oligos are specific to each allele of the SNP site, called the Allele-Specific Oligos (ASOs).

A third oligo that hybridizes downstream from the SNP site is the Locus-Specific Oligo (LSO). All three oligonucleotide sequences are complementary to the target sequence and contain universal PCR primer sites. During the primer hybridization process, the assay oligonucleotides hybridize to the genomic DNA sample bound to paramagnetic particles. Hybridization occurs before any PCR amplification step; therefore, no amplification bias is introduced into the assay. Extension of the appropriate ASO and ligation of the extended product to the LSO joins information about the genotype present at the SNP site to a unique address sequence on the LSO. These joined, full-length products provide a template for PCR using universal PCR primers, two of which are fluorescently labeled. As a last step, the single-stranded, dye-labeled amplification products are hybridized to the complement bead type through their unique address sequences, and the labeled beads can be analyzed using arrayed optical fibers. Up to 1536 SNPs may be interrogated simultaneously in this manner using this Golden-Gate technology.

Similar degrees of multiplexing can be achieved using oligonucleotide array technology from Affymetrix. These arrays contain 25-mers that are synthesized directly on chip surfaces. In recent iterations, Affymetrix has developed arrays that allow the interrogation of 10,000, 100,000, or even 500,000 SNPs per reaction. The 500K array methodology uses an approach called whole-genome sampling analysis to selectively amplify regions of the genome containing SNPs, and subsequently interrogate the alleles present in the amplification products by hybridization to allele-specific oligonucleotides.

In the first step of the analysis, a single genomic DNA sample is digested with a restriction enzyme. After complete digestion, adaptors are ligated to the digestion products that allow the amplification of a subset of the fragments using a universal primer pair complementary to the adaptor sequence. Any SNP located in close proximity to the restriction site will be amplified by PCR and can be interrogated on the chip. The resolution is primarily driven by the restriction enzyme used. The more it cuts the genomic DNA, the more small amplification products can be obtained and hybridized to the chip. Accordingly, the different chips, (10K, 100K, 500K) use different enzymes in the initial step of the reaction, but subsequent steps are identical.

Although the assay allows the interrogation of a large number of SNPs simultaneously, the user has no influence on the composition of the chip, i.e., the SNPs that will be interrogated. The complete panel of 500,000 SNPs has been preselected during the design of the chip, primarily driven by the technical requirements of the restriction digest and amplification described above. However, a novel methodology developed by Parallele Biosciences (now Affymetrix) uses the chip-based hybridization approach for the ultimate analysis, but it uses a different approach for the SNP interrogation. A more detailed description and illustrations of the method can be found at http://www.affymetrix.com/technology/mip_technology.affx. The approach, termed molecular inversion probe (MIP) technology, uses a long oligonucleotide that will hybridize with its end sequences in inverted fashion to the flanking sequences of an SNP [47, 48]. Using one dNTP in four different reactions, the SNP site can be filled in to form a circular padlock probe around the SNP site. The probe will only circularize when the dNTP added is complementary to the SNP allele present. The circular probe is resistant to DNAse digestion. Once the nonfilled probes have been digested using DNAse, the padlock probe is opened,

using a cleavage site within the MIP to form a linear template for a PCR reaction using universal primers. The amplification product contains a unique tag sequence for each MIP that can be detected by hybridization to an Affymetrix TrueTag array. Assays can be designed and genotyped in parallel for up to 10,000 SNPs. This platform offers a new tool to design custom panels of SNPs for a specific study, rather than depend on a preselected set of SNPs on other Affymetrix chips.

Overall, all three approaches allow the efficient genotyping of large numbers of SNPs. All require specialized equipment, and although the cost of genotyping per SNP is low, the total cost of any genotyping project is high due to the large number of SNPs. These technologies are only suitable for large-scale studies, and they are best performed in collaboration with or support from institutional core facilities.

### 3.4 Linkage Disequilibrium and Haplotypes

Most SNPs in a genome probably arose from individual single mutation events at an early time during the history of the species hundreds of generations ago. From the time of each mutation event, the new allele (i.e., the new "mutant" nucleotide) is located on an individual chromosome that has specific alleles of other SNPs that arose previously on the same chromosome. This physical arrangement of SNP alleles along a chromosome is called a *haplotype*. Over multiple successive generations, recombination and novel mutation events will lead to a rearrangement or modification of this ancestral haplotype around the new SNP allele. As a consequence, the new allele only remains on the same portion of the ancestral haplotype with other SNP alleles that are on a short distance away; i.e., recombination events are less likely to separate the two alleles. This nonrandom arrangement of SNPs along the chromosome, i.e., the maintenance of a small segment of the ancestral haplotype, is called *linkage disequilibrium* (LD) or *allelic association*. This arrangement is used in disease association analyses that are based on the idea that it should be possible to identify the effect of any common disease-causing SNP allele by determining the ancient haplotype segment on which it is located. By using a sufficient number of SNP markers in a genetic study, any common variant, even if it was not assayed directly, should display significant LD with a neighboring marker SNP. This approach has been used successfully in human genetic studies to identify genes responsible for Mendelian disorders such as Hirschprung's Disease or cystic fibrosis [49, 50].

The LD structure has been primarily studied in the human genome [51–53], although initial analyses have been performed for the mouse, rat, and dog genome as well. In the human genome, it has become clear over the past several years that the extent of LD and haplotypes in the human genome is not simply a function of distance between SNPs. Rather, the size of regions of significant LD is highly variable in different regions of the genome and reflects the complex history and interplay of recombination and mutation on different regions of the genome. If SNPs are not in LD, the alleles of the SNPs occur in seemingly random combination on individual chromosomes. As a consequence, the alleles of neighboring SNPs, in the absence of LD between them, can form a large number of different haplotypes ($2^n$ for n SNPs). In contrast, regions where neighboring SNPs are in significant LD, only a small number of resulting haplotypes is observed. These haplotypes are representative of the ancient haplotypes that have not been broken up by

recombination. Therefore, an analysis of haplotype patterns would not only identify regions of significant LD between SNPs, but it would also identify those common (presumably ancient) haplotype patterns that represent the majority of chromosomes in that particular genomic interval. Knowledge of these common haplotypes would then permit the identification of "tag SNPs," individual representative non-redundant SNPs that would unambiguously differentiate all major haplotypes without analyzing all SNPs in that particular region.

To facilitate the selection of these tag SNPs, the International HapMap Consortium, a publicly funded effort by the National Institutes of Health, has analyzed the genome-wide LD structure comprehensively in four different human populations: A cohort of North Americans of Northern European descent, an African population, a cohort of Han Chinese, and Japanese individuals [20]. In total, over 5.8 million SNPs have been genotyped in each of these populations, and the data of this effort are publicly available. It is the hope that selecting an informative subset of SNPs based on LD information from this study will provide a genome-wide coverage in whole genome disease association studies [54–57]. Here, SNPs would be selected so that all other SNPs from the HapMap study not included in the representative set would be in LD with one or more of the selected SNPs. It has been estimated that this comprehensive coverage of the human genome will require around 500,000 SNPs across all chromosomes. Undoubtedly, commercial providers of genotyping platforms will offer high-throughput SNP panels based on the HapMap information in the near future.

## 4 APPLICATION FOR GENETIC ANALYSES

As mentioned in Section 1, genotyping of SSLPs and SNPs is essential for the analysis of the genetic basis of common disorders. In the following section, we will briefly discuss the use of these polymorphisms in linkage and association analysis.

### 4.1 Linkage Analysis

A powerful means of identifying the location of genes causing disease is by genetic linkage mapping. During meiosis, when germ cells are being generated, homologous chromosomes can recombine, or cross over. As a result, offspring will have different combinations of parental alleles. We can use the percentage of recombinant offspring as a measure of physical distance, making a genetic linkage map. Linkage maps are based on probability. For instance, if two genes or loci are on different chromosomes, the probability of cosegregation of these loci is 50%. However, the closer loci are on the genome, the lower the probability of recombination between them. The measure of genetic distance is approximately the recombination frequency between loci. Although linkage is defined as <50% recombination, 47% recombination, for example, is not a reassuring or reliable measure of distance. Therefore, a genetic map needs many genetic markers to be reliable. Because of their relatively high abundance, their high informativeness, and their representation across all eukaryotes, SSLPs are the backbone of nearly every eukaryotic genetic map. To generate a genetic linkage map, large families, such as the

three-generation families from The Foundation Jean Dausset-Centre d'Etude du Polymorphisme Humain (CEPH), are typically genotyped with hundreds or thousands of SSLPs. Analytical computer programs determine the frequency of recombination between parents and offspring, thus determining the distance between markers, and arrange the SSLPs according to their order and distance on each chromosome.

High-density genetic maps spanning the entire human genome have been generated by several groups and are commonly used for mapping disease genes and quantitative trait loci (QTL) [15, 16]. Just as recombination determines the genetic distance between SSLPs, it also determines the distance between a disease gene and an SSLP. If a particular gene determines a specific phenotype or disease, then one can genotype families with SSLPs spanning the genome and identify cosegregation of an SSLP allele and a clinical phenotype. If recombination is significantly lower than expected between an SSLP allele and a clinical outcome, e.g., hemophilia, then the gene causing the disease is *linked* to that marker. Because the genome location of that marker is known, the location of that disease gene is also known. The first disease linked by the use of microsatellite genotyping was facioscapulohumeral muscular dystrophy in 1990 [58]. Since that time, thousands of loci have been mapped using this approach. Genetic linkage studies are most powerful when studying large, multigenerational families affected by a single gene (Mendelian) disease. However, linkage has proven more of a challenge with common complex diseases. Because multiple genes with relatively small effects are thought to be involved in common diseases such as hypertension or diabetes, combined with the fact that these diseases often affect older patients, it is often impossible to find large families with multiple generations. To address these issues, association studies offer the advantages of being able to study cases and controls or small families, which are much more readily ascertained. Furthermore, the large number of SNPs and high-throughput SNP genotyping technologies now makes genome-wide association studies feasible, as discussed below.

### 4.2 Association Analysis

When it proves difficult to recruit a sufficient number of families for a genetic study, a popular alternative is an association study to elucidate the genetic basis of common diseases. In contrast to linkage studies where SSLP markers are used to identify a region of the genome that is inherited by affected offspring from affected parents, association studies aim to identify alleles that are found more often in individuals affected by a disease than in a control group. In the perfect association study, one would test the causal mutation for association with disease. However, in almost all cases, the causal mutation is unknown. As a consequence, association studies resort to indirect approaches relying on linkage disequilibrium (LD) between the causal mutation and another SNP that is being genotyped. If the SNP was associated with the disease, one would expect to find a higher frequency of the associated allele of the SNP in individuals affected by the disease (cases) when compared with a healthy control group.

Association studies can be performed in family-based cohorts. Numerous statistical approaches have been developed to test for association of SNP alleles with disease in small trios, nuclear, or even extended families. However, most

association studies use unrelated patients in their analysis. In these cohorts, it is assumed that individuals are affected because they share a similar genetic susceptibility. However, as it is possible that some individuals may be affected for reasons other than genetic susceptibility (most common human disorders are also significantly influenced by environmental factors, e.g., lifestyle), it is imperative that a large number of patients is recruited for the study so that most individuals share the same genetic susceptibility genes and mutations. This cohort of affected individuals is then matched with a cohort of unaffected, healthy controls. Here, it is important that these control individuals match the patient cohort in the distribution of age, sex, ethnicity, and potential environmental factors that may influence the development of the disease.

Once the case and control cohorts have been collected, the DNA samples can be used for SNP genotyping. Depending on the scope of the study, samples are either genotyped for SNPs in candidate genes that were preselected for the study or a genome-wide association analysis using a comprehensive SNP set across the entire genome will be performed. Published data to date primarily include candidate gene association analyses. A recent review by Newton-Cheh and Hirschhorn [59] discusses in detail the important considerations for the design of these association studies and the potential problems that can be encountered. The number of SNPs to be analyzed would determine the best genotyping platform to be used in the study. Clearly, candidate gene SNPs can be genotyped with common approaches for individual SNP genotyping or moderate multiplexing platforms. However, genome-wide association studies will require high-throughput approaches.

Although the recent efforts to elucidate the LD structure of the human genome (HapMap Consortium) promise to facilitate the selection of informative subsets of SNPs for these types of association studies, it remains to be seen whether whole genome association studies will help uncover the genetic determinants for such common diseases as diabetes, asthma, or cardiovascular disorders. Clearly, candidate gene studies have yielded several well-replicated associations, but they account only for a small portion of the overall genetic susceptibility. Hopefully, comprehensive association studies of these disorders will help uncover additional genes responsible for the disease susceptibility. These discoveries would significantly advance our ability to predict, diagnose, and hopefully treat these disorders in affected individuals.

## INTERNET RESOURCES

*SSLP Genotyping Resources*

*DeCode*: fee-for-service Genotyping—http://www.decode.com/

*Marshfield Medical Research Foundation*: Genetic markers and maps—http://research.marshfieldclinic.org/genetics/Genotyping_Service/mgsver2.htm

*Fondation Jean Dausset—CEPH*: DNA from multigenerational families, Human Diversity Panel, Genotype database, genetic maps—http://www.cephb.fr/

*CIDR* (Center for Inherited Disease Research): NIH-funded genotyping (Multi-instutional)—http://www.cidr.jhmi.edu/

*RSnG* (Resequencing and Genotyping Service): NIH-funded resequencing and genotyping (NHLBI)—http://rsng.nhlbi.nih.gov/scripts/index.cfm

*SNP Genotyping Resources*

*dbSNP*: NCBI-maintained central database and repository for SNPs and genotyping information—http://www.ncbi.nlm.nih.gov/SNP/

*HapMap* homepage: http://www.hapmap.org/

*Affymetrix*: Information on GeneChip arrays—http://www.affymetrix.com/index.affx

*Applied Biosystems*: Information on SNaPshot, TaqMan, fluorescently labeled ddNTPs—http://myscience.appliedbiosystems.com/navigation/mysciApplications.jsp?tabNameAttribute=applSnp

*Illumina*: Information on GoldenGate SNP genotyping technology—http://www.illumina.com/products/prod_snp.ilmn

*Parallele*: Information on MIP genotyping platform—http://www.affymetrix.com/technology/mip_technology.affx

*Pyrosequencing*: http://www.pyrosequencing.com/

*Sequenom*: Information on MassEXTEND MALDI-TOF-based SNP genotyping approaches, fiberoptic bead arrays—http://www.sequenom.com/

*Third Wave Technologies*: Information on Invader assay technology—http://www.twt.com/

## REFERENCES


1. Griffiths A J F (2000). *An Introduction to Genetic Analysis*, 7th ed. W.H. Freeman, New York, p. 860.
2. Houck C M, Rinehart F P, Schmid C W (1979). A ubiquitous family of repeated DNA sequences in the human genome. *J. Mol. Biol.* 132(3):289–306.
3. Litt M, Luty J A (1989). A hypervariable microsatellite revealed by in vitro amplification of a dinucleotide repeat within the cardiac muscle actin gene. *Am. J. Hum. Genet.* 44(3):397–401.
4. Weber J L, May P E (1989). Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. *Am. J. Hum. Genet.* 44(3):388–396.
5. Hamada H, et al. (1984). Characterization of genomic poly(dT-dG).poly(dC-dA) sequences: Structure, organization, and conformation. *Mol. Cell Biol.* 4(12):2610–2621.
6. Sheffield V C, et al. (1995). A collection of tri- and tetranucleotide repeat markers used to generate high quality, high resolution human genome-wide linkage maps. *Hum. Mol. Genet.* 4(10):1837–1844.
7. Jeffreys A J, Wilson V, Thein S L (1985). Hypervariable "minisatellite" regions in human DNA. *Nature.* 314(6006):67–73.
8. Weber J L, Wong C (1993). Mutation of human short tandem repeats. *Hum. Mol. Genet.* 2(8):1123–1128.
9. Miesfeld R, Krystal M, Arnheim N (1981). A member of a new repeated sequence family which is conserved throughout eucaryotic evolution is found between the human delta and beta globin genes. *Nucleic Acids Res.* 9(22):5931–5947.
10. Weber J L (1990). Informativeness of human (dC-dA)n.(dG-dT)n polymorphisms. *Genomics.* 7(4):524–530.
11. Mullis K B, Faloona F A (1987). Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction. *Methods Enzymol.* 155:335–350.
12. Ausubel F M (1992). *Short protocols in molecular biology : A compendium of methods from Current protocols in molecular biology*, 2nd ed. Greene Pub. Associates, New York.

13. Moreno C, et al. (2004). Genome-Wide Scanning With SSLPs in the Rat. *Methods. Mol. Med.* 108:131–138.
14. Oetting W S, et al. (1995). Linkage analysis with multiplexed short tandem repeat polymorphisms using infrared fluorescence and M13 tailed primers. *Genomics.* 30(3):450–458.
15. Murray J C, et al. (1994). A comprehensive human linkage map with centi-morgan density. Cooperative Human Linkage Center (CHLC). *Science.* 265(5181):2049–2054.
16. Dib C, et al. (1996). A comprehensive genetic map of the human genome based on 5264 microsatellites. *Nature.* 380(6570):152–154.
17. Sachidanandam R, et al. (2001). A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature.* 409(6822):928–933.
18. Frazer K A, et al. (2004). Segmental phylogenetic relationships of inbred mouse strains revealed by fine-scale analysis of sequence variation across 4.6 mb of mouse genome. *Genome Res.* 14(8):1493–1500.
19. Smits B M, et al. (2004). Genetic variation in coding regions between and within commonly used inbred rat strains. *Genome Res.* 14(7):1285–1290.
20. The international HapMap project. *Nature.* 2003. 426(6968):789–796.
21. Nickerson D A, Tobe V O, Taylor S L (1997). PolyPhred: Automating the detection and genotyping of single nucleotide substitutions using fluorescence-based resequencing. *Nucleic Acids Res.* 25(14):2745–2751.
22. Hayashi, K (1991). PCR-SSCP: A simple and sensitive method for detection of mutations in the genomic DNA. *PCR Methods Appl.* 1(1):34–38.
23. Andersen P S, et al. (2003). Capillary electrophoresis-based single strand DNA conformation analysis in high-throughput mutation screening. *Hum. Mutat.* 21(5):455–465.
24. Xiao W, Oefner P J (2001). Denaturing high-performance liquid chromatography: A review. *Hum. Mutat.* 17(6):439–474.
25. Yu B, et al. (2005). Denaturing high performance liquid chromatography: High throughput mutation screening in familial hypertrophic cardiomyopathy and SNP genotyping in motor neurone disease. *J. Clin. Pathol.* 58(5):479–485.
26. Han W, et al. (2004). Using denaturing HPLC for SNP discovery and genotyping, and establishing the linkage disequilibrium pattern for the all-trans-retinol dehydrogenase (RDH8) gene. *J. Hum. Genet.* 49(1):16–23.
27. Mellai M, et al. (2003). Prolactin and prolactin receptor gene polymorphisms in multiple sclerosis and systemic lupus erythematosus. *Hum. Immunol.* 64(2):274–284.
28. Huang H, et al. (2002). Screening of single nucleotide polymorphisms in nasopharyngeal carcinoma associated genes by denaturing high-performance liquid chromatography. *Di Yi Jun Yi Da Xue Xue Bao.* 22(7):602–604.
29. Shi J, et al. (2002). Sequence variations in the mu-opioid receptor gene (OPRM1) associated with human addiction to heroin. *Hum. Mutat.* 19(4):459–460.
30. Ribas G, Neville M J, Campbell R D (2001). Single-nucleotide polymorphism detection by denaturing high-performance liquid chromatography and direct sequencing in genes in the MHC class III region encoding novel cell surface molecules. *Immunogenet.* 53(5):369–381.
31. Hecker K H, et al. (2000). Mutation detection in the human HSP7OB' gene by denaturing high-performance liquid chromatography. *Cell Stress Chaper.* 5(5):415–424.
32. Giordano M, et al. (1999). Identification by denaturing high-performance liquid chromatography of numerous polymorphisms in a candidate region for multiple sclerosis susceptibility. *Genomics.* 56(3):247–253.

33. Upchurch D A, Shankarappa R, Mullins J I (2000). Position and degree of mismatches and the mobility of DNA heteroduplexes. *Nucleic Acids Res.* 28(12):E69.
34. Youil R, Kemper B W, Cotton R G (1995). Screening for mutations by enzyme mismatch cleavage with T4 endonuclease VII. *Proc. Natl. Acad. Sci. U.S.A.* 92(1):87–91.
35. Faham M, Cox D R (1995). A novel in vivo method to detect DNA sequence variation. *Genome Res.* 5(5):474–482.
36. Faham M, et al. (2005). Multiplexed variation scanning for 1000 amplicons in hundreds of patients using mismatch repair detection (MRD) on tag arrays. *Proc. Natl. Acad. Sci. U.S.A.* 102(41):14717–14722.
37. Faham M, et al. (2001). Mismatch repair detection (MRD): High-throughput scanning for DNA variations. *Hum. Mol. Genet.* 10(16):1657–1664.
38. Landegren U, et al. (1988). A ligase-mediated gene detection technique. *Science.* 241(4869):1077–1080.
39. Lyamichev V, et al. (1999). Polymorphism identification and quantitative detection of genomic DNA by invasive cleavage of oligonucleotide probes. *Nat. Biotechnol.* 17(3):292–296.
40. Livak K J (1999). Allelic discrimination using fluorogenic probes and the 5′ nuclease assay. *Genet. Anal.* 14(5–6):143–149.
41. Fakhrai-Rad H, Pourmand N, Ronaghi M (2002). Pyrosequencing: An accurate detection platform for single nucleotide polymorphisms. *Hum. Mutat.* 19(5):479–485.
42. Kwok P Y (2002). SNP genotyping with fluorescence polarization detection. *Hum. Mutat.* 19(4):315–323.
43. Hsu T M, et al. (2001). Genotyping single-nucleotide polymorphisms by the invader assay with dual-color fluorescence polarization detection. *Clin. Chem.* 47(8):1373–1377.
44. Berggren W T, et al. (2002). Multiplexed gene expression analysis using the invader RNA assay with MALDI-TOF mass spectrometry detection. *Anal. Chem.* 74(8):1745–1750.
45. Ross P, et al. (1998). High level multiplex genotyping by MALDI-TOF mass spectrometry. *Nat. Biotechnol.* 16(13):1347–1351.
46. Kim S, et al. (2003). Multiplex genotyping of the human beta2-adrenergic receptor gene using solid-phase capturable dideoxynucleotides and mass spectrometry. *Anal. Biochem.* 316(2):251–258.
47. Hardenbol P, et al. (2005). Highly multiplexed molecular inversion probe genotyping: Over 10,000 targeted SNPs genotyped in a single tube assay. *Genome Res.* 15(2):269–275.
48. Hardenbol P, et al. (2003). Multiplexed genotyping with sequence-tagged molecular inversion probes. *Nat. Biotechnol.* 21(6):673–678.
49. Kerem B, et al. (1989). Identification of the cystic fibrosis gene: Genetic analysis. *Science.* 245(4922):1073–1080.
50. Puffenberger E G, et al. (1994). A missense mutation of the endothelin-B receptor gene in multigenic Hirschsprung's disease. *Cell.* 79(7):1257–1266.
51. Reich D E, et al. (2001). Linkage disequilibrium in the human genome. *Nature.* 411(6834):199–204.
52. Stephens J C, et al. (2001). Haplotype variation and linkage disequilibrium in 313 human genes. *Science.* 293(5529):489–493.
53. Daly M J, et al. (2001). High-resolution haplotype structure in the human genome. *Nat. Genet.* 29(2):229–232.

54. Halldorsson B V, Istrail S, De La Vega F M (2004). Optimal selection of SNP markers for disease association studies. *Hum. Hered.* 58(3–4):190–202.
55. Stram D O (2004). Tag SNP selection for association studies. *Genet. Epidemiol.* 27(4):365–374.
56. Sebastiani P, et al. (2003). Minimal haplotype tagging. *Proc. Natl. Acad. Sci. U.S.A.* 100(17):9900–9905.
57. Ke X, Cardon L R (2003). Efficient selective screening of haplotype tag SNPs. *Bioinform.* 19(2):287–288.
58. Wijmenga C, et al. (1990). Facioscapulohumeral muscular dystrophy gene in Dutch families is not linked to markers for familial adenomatous polyposis on the long arm of chromosome 5. *J. Neurol. Sci.* 95(2):225–229.
59. Newton-Cheh C, Hirschhorn J N (2005). Genetic association studies of complex traits: Design and analysis issues. *Mutat. Res.* 573(1–2):54–69.

# PROTEINS: HORMONES, ENZYMES, AND MONOCLONAL ANTIBODIES—BACKGROUND

Wayne Materi, Zoltan Gombos, and David S. Wishart
*National Institute for Nanotechnology, University of Alberta, Edmonton, Alberta, Canada*

**Chapter Contents**

1 Introduction
2 Protein Pharmaceuticals
3 Peptide and Protein Hormone Biopharmaceuticals
  3.1 Examples of Hormone Biopharmaceuticals
  3.2 The Future of Hormone Biopharmaceuticals
4 Antibody Biopharmaceuticals
  4.1 Examples of Antibody Biopharmaceuticals
  4.3 The Future of Antibody Biopharmaceuticals
5 Enzyme Biopharmaceuticals
  5.1 Examples of Enzyme Biopharmaceuticals
  5.2 The Future of Enzyme Biopharmaceuticals
6 Conclusions
  References

## 1 INTRODUCTION

The human body is a veritable pharmacopia of useful protein drugs. After all, it is the collection of our own, endogenously produced enzymes, hormones, and antibodies that are responsible for maintaining homeostasis, stabilizing wounds, fighting infections, neutralizing toxins, keeping cancerous cells in check, and generally keeping us alive. It is only when we lose the function of certain enzymes, hormones,

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

or antibodies—or when our bodies are overwhelmed with some kind of trauma (blood loss, stroke, heart attack, massive infection, or heavy tumor burden)—that we need some supplementation to our natural protein drugs. The fact that our bodies or body fluids contain some kind of miraculous healing or "vital" substances was recognized even in ancient times by the Chinese, Moche Indians, Maasai, ancient Scythians, Gypsies, and early Christians when would-be warriors or the sick and sinful were encouraged to drink such body fluids as blood or urine to imbue health, healing, or strength. However, the general failure of these primitive protein cocktails was probably evident to most patients and many physicians especially by the Middle ages. Given the abysmal failure of oral protein delivery, Italian physicians, in the late 1400s began experimenting with a different approach, namely intravenous delivery via blood transfusions. The first recorded instance of a therapeutic blood transfusion was for the treatment of Pope Innocent VIII who fell into a stroke-induced coma in 1492 [1]. Although this effort ultimately failed, efforts to refine or improve blood transfusions from animals to humans or humans to humans as a way to treat wounds and other illnesses continued for another 300 years. The first therapeutically successful blood transfusion, and perhaps the first example of the successful use of a protein drug (albeit impure albumin), was performed in England by James Blundell in 1825 to treat a woman suffering from postpartum hemorrhaging [2]. By 1844 Blundell and his colleague Samual Armstrong Lane became the first to use blood (i.e., factor VIII) to treat a genetic disease—hemophilia [3]. Throughout the 1800s and early 1900s, blood transfusions became more refined (through blood typing and the use of aseptic conditions) and blood became a therapeutic protein product to treat wounds (albumin), infections (antibodies, antiglobulins), and other genetic disorders.

Obviously blood is not the only source of protein drugs. As far back as 1796 other sources of protein pharmaceuticals were beginning to be recognized. In particular, the work of Edward Jenner in developing an effective small-pox vaccine [4] could possibly be described as the first example of using protein drugs for prophylactic purposes. Keeping with the ancient theme of using "vital" body fluids for medical applications, Jenner used the fluid from cow-pox pustules as the source of his protein pharmaceutical (cow-pox viral coat proteins). In the 1870s Louis Pasteur extended Jenner's ideas by developing vaccines for cholera, anthrax, and rabies using deactivated bacterial cells or dried nerve tissues. Pasteur formulated the dried bacterial or viral protein components into injectable solutions. In this regard Pasteur was the first to introduce the concept of "synthetic" or *ex vivo* biological products as potential drugs. However the idea of working with a purified, active ingredient, especially with biological material, was still many years away. The principle of prophylactic vaccines, developed by Jenner and refined by Pasteur, was also extended to the development of "therapeutic vaccines" or antivenoms in the late 1800s. Antivenoms are generated using a method developed by Albert Calmette in 1895, which he devised to treat victims of cobra bites. Antivenoms are created by injecting a small amount of the targeted venom into animals such as horses, sheep, or rabbits. This leads to the production of different antibodies against the toxin. The mixture of antibodies is then harvested from the animal's blood and serves as the antivenom for a specified source venom.

Up until the mid-1800s the true source of the therapeutic powers of blood or the prophylactic powers of pustules or bacterial extracts was not known. Then in 1838

the concept of proteins was introduced by Jons Jakob Berzelius [5]. Berzelius argued that proteins (he is credited for creating the name) found in blood and bacteria were special organic substances that behaved or functioned as chemical compounds. This led to the realization that many different proteins existed in blood and other biological tissues and that they could potentially be isolated and treated as pure substances, just like all other chemical or small-molecule drug entities. However, it was not until 1922 that the technology to isolate therapeutically useful proteins was put to good use. Thanks largely to the work of four Canadian scientists (Banting, Best, Collip, and MacLeod), insulin became the first "pure" protein therapeutic to be used in the successful treatment of a human disease—diabetes. Before 1922, type I diabetes was a near-certain death sentence characterized by an agonizingly long and painful wasting process. The fact that it frequently afflicted adolescent children made it particularly devastating. Although the cause of diabetes was partially known by 1910, early efforts to use unpurified pancreatic extracts met with repeated failure and severe allergic reactions. The use of purified insulin, on the other hand, had absolutely stunning results. It is hard not to overstate the kind of spectacular recoveries observed among diabetic patients nor the effect that insulin, as a drug, had on the public as well as on the medical and pharmaceutical community. Indeed, the insulin "effect" not only saved the lives of millions of diabetics, but it essentially launched the modern concept of "pure" protein pharmaceuticals and laid the foundation to today's modern biotech industry.

Most of today's protein pharmaceuticals fall into 5 general classes: (1) hormones, (2) vaccines, (3) antibiotics, (4) antibodies, and (5) enzymes. Three of these classes, hormones, antibodies, and enzymes, are primarily used to treat noninfectious or endogenous diseases (i.e., genetic diseases or diseases of aging), whereas the remaining two classes, vaccines and antibiotics, are used to treat or prevent infectious or exogenous diseases (i.e., bacterial or viral infections). Interestingly all five drug classes were essentially identified or defined only in the past 80 years. Insulin, a hormone, was the first purified protein drug to be marketed (1922). Shortly after, in 1923, purified diptheria toxoid (a 58-kDa protein) was introduced, making it the first purified vaccine product [6]. The first purified peptide antibiotic, gramicidin S, was introduced in Russia in 1943 as a topical wound-healing agent. The first purified antivenoms or antitoxin antibodies were introduced in the 1950s and the first monoclonal antibody—OKT3 (muronomab or Orthoclone)—was approved for human use in 1986 [7]. The first purified enzyme to be approved for therapeutic use, Activase, was marketed in 1987. Interestingly, the first human recombinant protein used in disease management was also insulin (Humulin) [8], which was introduced in 1982 by Eli Lilly.

Up until 1982 all peptide and protein pharmaceuticals were isolated from "natural" sources, meaning they were purified from animal (or human) biofluids and tissues. Working with natural sources is exceedingly difficult, expensive, and can lead to the transmittance of undetected infectious materials (viruses, bacteria, prions) to patients. Because most proteins of pharmaceutical interest are relatively scarce, it was often necessary to process tens or hundreds of kilograms of tissue or fluid to produce a few milligrams of the desired product in pure form. For instance, up until the 1980s the preparation of sufficient human growth hormone (somatotropin) for a single treatment to combat dwarfism would require the extraction and homogenization of several adult human (cadaver) brains [6].

For the more abundant proteins (such as albumin) it is possible to find them in concentrations of about 35 g/L in selected tissues or biofluids [9]. This makes their isolation and purification relatively simple. However, many important protein pharmaceuticals are only found in concentrations of 1 or 2 pg/L. At such low concentrations, it is almost impossible to isolate and purify these compounds from the thousands of other proteins that are in the body. Furthermore, even if one wanted to produce enough material for commercial purposes, it would probably require the processing or homogenization of the entire world supply of cattle, pigs, or humans. Today most protein pharmaceuticals are produced through recombinant methods, including insulin, growth hormone, and most monoclonal antibodies. This allows large quantities of even the rarest protein to be purified and prepared under tightly controlled manufacturing conditions without the concerns over contamination with host viruses (HIV, HCV), toxins, or prions (that cause diseases such as CJD or BSE). Although mast protein drugs are biosynthetically produced (via cell culture), a small number of peptide pharmaceuticals (<20 residues) are produced chemically via automated peptide synthesis. These include Leuprolide (Eligard), Oxytocin (Oxytocin), Calcitonin (Miacalcin), and Abarelix (Plenaxis), for example. Peptides and proteins that continue to be isolated from natural sources include Menotropin (Repronex) and Hyaluronidase (Vitrase), for example.

Since the introduction of insulin in 1922, more than 110 distinct protein or peptide drugs have been approved for human use by the U.S. Food and Drug Administration (FDA) [10], with 50 being hormones, 2 being vaccines, 2 being peptide antibiotics, 26 being antibodies or antibody mixtures, and 17 being enzymes (see Tables 1–3). These peptide/protein pharmaceuticals range in size from less than 10 amino acids to over 1000 and may contain a variety of covalent modifications such as carbohydrate chains, D-amino acids, or amino acids with unusual side chains. More details about the structure and chemical composition of most protein pharmaceuticals can be found in subsequent chapters in this book as well as in the DrugBank database under its collection of biotech drugs [10]. According to the Biotechnology Industry Organization (www.bio.org), there are now more than 300 peptide and protein pharmaceuticals currently in clinical trials or under review. Just as with the FDA-approved biotech drugs, most of these newer products or potential products fall into the three major classes of endogenous or human-derived proteins: (1) hormones, (2) antibodies, and (3) enzymes. It is these three classes of pharmaceutically important proteins that are the focus of this chapter.

Beyond providing a historical perspective to proteins and protein pharmaceuticals, this chapter is intended to provide the reader with an overview of the general features and characteristics of pharmaceutically important hormones, antibodies, and enzymes. Together these three classes of protein therapies account for nearly three quarters of all approved biotech drugs and more than 90% of all biotech drug sales. In addition to providing a general outline about what hormones, enzymes, and antibodies are, this chapter will also provide the reader with specific examples and detailed descriptions of a few of the more important or historically interesting hormones, enzymes, and antibodies used today. The chapter is divided into five sections, an introduction to the history of protein pharmaceuticals, a general discussion about the advantages and disadvantages of protein pharmaceuticals, a detailed discussion on hormones, a detailed description of enzymes, and a detailed discussion on antibodies. Each section on hormones, enzymes, and antibodies

provides a working description or definition of class of these protein/peptide products; a general discussion on certain unique aspects of their formulation or delivery; several specific examples of important FDA-approved hormones, enzymes, or antibodies; and a brief discussion of some of the new enzymes, hormones, and antibodies that are under development or in clinical trials. A complete discussion concerning all aspects of hormones, enzymes, and antibodies is far beyond the scope of this chapter. Readers who are interested in learning more might want to consider reading several excellent books [11–13] or referring to the DrugBank database (http://redpoll.pharmacy.ualberta.ca/drugbank/) for more extensive biological, pharmaceutical, and biochemical data.

## 2 PROTEIN PHARMACEUTICALS

For a peptide or protein to be a useful drug, it must generally be water soluble, relatively stable, nonimmunogenic (i.e., identical or near identical to a human homologue), mostly monomeric, and causal or preventative to some disease process. Not all polypeptides match these requirements, and therefore, not all polypeptides are potentially suitable as drugs. Typically those that are not (yet) useful drugs are highly polymeric structural proteins such as tubulin, actin, or myosin. Equally impractical or improbable drugs are membrane-bound receptor proteins (G-protein-coupled receptors, laminins, etc.). These classes of proteins, which may account for as much as 60% of the human proteome, are usually most suitable as drug targets—not drugs. Certain other proteins, such as polymerases, histones, and ribosomal proteins, perform such vital functions that they are "essential" to life and so cannot generally serve as useful protein drugs either, although DNaseI (Pulmozyme) is a notable exception. Using these relatively simple criteria about what constitutes a viable drug versus a viable drug targets, it is possible to scan the human proteome and identify likely classes of potential protein drugs. When this is done one is typically left with three major groups of peptides or proteins: (1) hormones, (2) enzymes, and (3) immunological molecules (antibodies).

Peptide and protein pharmaceuticals are different from small-molecule drugs. For one, polypeptide drugs are generally much larger, ranging in size from 1000 daltons to more than 200,000 daltons (compared with <600 daltons for most small-molecule drugs). For another, the noncovalent structure or three-dimensional fold (i.e., the tertiary structure) of most polypeptides is absolutely critical to their function. This is in distinct contrast to small-molecule drugs, wherein their covalent or "primary" structure defines their function. Smaller polypeptides have relatively little tertiary structure or exhibit substantial tertiary structure only when bound to a target receptor. Larger (>40 residues) polypeptides generally have a stable, well-defined tertiary structure. As a general rule, polypeptides with less than 40 amino acids are called peptides, whereas those with more than 40 residues are called proteins. Typically most hormones are peptides (<40 residues), although some hormones such as somatotropin (human growth hormone) can be up to 200 residues in length. Many peptide hormones are actually fragments of larger precursor proteins or pro-proteins. Although most hormones are relatively small polypeptides, enzymes and antibodies are generally much larger. As a rule, most enzymes are between 200 and 800 residues in length, whereas most antibodies are typically

1200 residues in length, although smaller antibody fragments (Fabs, single-chain antibodies) can also be used as protein drugs.

Perhaps the most impressive and appealing quality about peptide and protein pharmaceuticals is their exquisite selectivity and specificity. Bioactive proteins target their receptors or bind their small-molecule ligands with a precision that matches a cruise missile. Almost no cross-reactivity to other targets or other receptor molecules is observed with most known hormones, enzymes, or antibodies. Additionally most protein drugs also exhibit very appealing pharmacokinetics and excellent bioavailability. These favorable pharmaceutical properties are not entirely unexpected. Indeed, the process of natural selection over hundreds of millions of years has evolved bioactive proteins such as hormones, enzymes, and antibodies to perform their functions in a fashion that is optimal to the health and viability of each organism. In contrast to so-called large-molecule (i.e., peptide and protein) drugs, almost no known small-molecule drug exhibits comparable selectivity or sensitivity. Perhaps the only small molecules that do are those drugs that are identical to the naturally occurring chemicals in the body (i.e., estrogen, testosterone, thyroxine, and epinephrine). Indeed, these small molecules likely coevolved with their target proteins.

The fact that polypeptides are large molecules means that far more functional and operational information can be chemically encoded into them than small molecules. In other words, there is much more room on a protein scaffold to add, remove, or substitute chemical moieties to change a given protein's specificity, bioavailability, kinetics, or function. This chemical flexibility can allow a protein chemist to engineer a potential or promising new protein drug into a more useful product. Indeed, many peptide and protein pharmaceuticals on the market today are engineered, via site-directed mutagenesis, or chemically modified to enhance their stability, selectivity, or efficacy. For example, the cytokine hormone Infergen is a "designed" interferon derived from a consensus alignment of bioactive mammalian interferon alpha sequences. This designed molecule was found to give the protein greater bioactivity. Another example is pegademase, a PEGylated version of the enzyme adenosine deaminase [14]. This protein is decorated with poly(ethylene glycol) (PEG) subunits covalently attached to exposed lysine side chains. This covalent chemical modification extends the *in vivo* half-life of the protein by a factor of 10 or more. Yet another example of an engineered or rationally designed protein is the hormone LisPro Insulin, a derivative of insulin in which Lysine 28 and Proline 29 have been interchanged [15]. This chemical change was found to enhance the activity and extend the lifetime of this form of insulin relative to natural insulin. Beyond these specific examples of engineered hormones and enzymes is a much larger collection of chimeric murine-human antibodies such as Infliximab, Rituximab, and Abciximab that have been designed to have murine variable domains and human constant domains. These chimeric features reduce the antigenicity of monoclonal murine antibodies and thereby greatly enhance the tolerance and half-life of these molecules.

Selectivity, specificity, and "programmability" are what make protein pharmaceuticals so attractive to today's pharmaceutical and biotech industry. As a result, enormous efforts are now going into the discovery, production, and FDA approval of dozens of new protein drugs. In 2005, the market capitalization of the biotechnology industry in the United States was estimated to be $311 billion. In 2004, 40

new protein pharmaceuticals (drug types, indications, or formulations) were approved by the (FDA)—more than in any prior year. With the completion of the Human Genome Project and the launch of many large-scale proteomic efforts [16, 17], the potential to find even more pharmaceutically useful hormones, enzymes, and antibodies will likely grow even further. How much further can it grow? The "universe" of endogenous protein drugs can be estimated by considering the size of the human proteome and the current diversity of functions or roles that have been identified to date. In particular, the approximately 25,000 genes identified in the human body probably code for around 40,000 different proteins (isozymes or splice variants) and up to 1,000,000 different posttranslationally modified variants, including chemically modified or cleaved proteins [18]. Current annotations suggest that approximately 3000 different human enzymes exist, along with over 150 different hormones. Given the number of enzymes and hormones available (which may be used for diseases requiring replacement therapy), this suggests that there could be at least 3500 antibody targets (which may be needed for antagonizing the unwanted effects of certain enzymes or hormones). If we include the number of known receptors (estimated to be 4000) for which antibodies may be targeted, then the number of potential, pharmaceutically important antibodies may be closer to 8000 molecules. Given that there are less than 400 protein drugs that are either approved or in the drug development pipeline, this rough calculation suggests we still have a long way to go before we exhaust the supply of potential endogenous protein drugs.

Although there is much to be optimistic about the future of protein pharmaceuticals, there are still many unique problems with their development, production, and delivery. Among the more obvious problems with protein drugs is the fact that they are much more delicate than small-molecule drugs. Proteins such as hormones, antibodies, and enzymes cannot normally be "compounded" or pressed into dry pills or emulsified or concentrated into tinctures. This type of conventional pharmaceutical manufacturing and formulation would destroy the activity of most protein pharmaceuticals. Similarly most peptide hormones, antibodies, and enzymes cannot be stored indefinitely at room temperatures in nonsterile containers; instead they must be kept in a cool, dark, aqueous, sterile environment for no more than a few weeks. These limitations to protein preparation and formulation have created a significant challenge to pharmaceutical chemists.

Yet another challenge in working with or producing peptide and protein drugs is their cost. Relative to small-molecule drugs, manufacturing costs for peptide and protein pharmaceuticals are enormous, with many protein drugs being priced at 10 to 1000 times the cost of a small-molecule drug (on a per-milligram basis). This price difference exists because most polypeptide drugs cannot be synthesized using classicl synthetic organic methods—they are simply too large and too complex for today's technology. Currently the only cost-effective method for producing most polypeptide hormones, antibodies, and enzymes is through recombinant gene expression in bacterial or mammalian cell culture. This is a time-consuming, low-yield (100 mg—10 g per liter) process that requires the use of large fermentor systems, carefully monitored growth conditions, and multiple purification steps.

Beyond the obvious problems of storage and production of protein pharmaceuticals lies one of the more discouraging facts about protein drugs: The body is not

a particularly friendly place for foreign proteins. Indeed, our bodies have developed a huge arsenal of tricks to quickly and efficiently dispose of proteins that are swallowed, ingested, or injected. For instance, proteins that are swallowed are immediately attacked by enzymes (amylases [19]) in the mouth, which are designed to remove the protective sugars that cover many proteins. Moving down the throat, these "naked" proteins are adsorbed and denatured by surface interactions with the mucosal cells. As the surviving proteins enter the stomach and small intestine, they are exposed to extremely acidic conditions (which instantly denature most proteins) and a host of acid-stable proteases that digest these denatured molecules into short, inactive peptides. Similarly, if a peptide or protein pharmaceutical is injected into the bloodstream, it may be attacked by antibodies or swallowed by T cells and cut into fragments. Alternatively, it may be adsorbed by albumins or lipoproteins and permanently removed from the circulation. Even if an injected protein survives these insults, a foreign protein may still be attacked by plasma proteases and demolished into tiny fragments. Foreign proteins are "ill treated" due to the design of our bodies to use proteins as a source of food (if ingested) or to identify pathogens (such as bacteria, viruses, and parasites) that have invaded our circulatory system. This makes most proteins "enemies" of the body, something to be digested or eliminated at every opportunity. Even for its own "friendly" proteins (i.e., the proteins needed for cellular functions), the body has established several mechanisms to turn over or eliminate these proteins every few days. This regular turnover prevents old proteins from malfunctioning and keeps tight control over regulatory proteins that need only be around for minutes or hours at a time. These finely tuned mechanisms, which have evolved over the past two billion years, make it very difficult for pharmaceutical companies to use protein drugs as injectables or oral consumables.

## 3 PEPTIDE AND PROTEIN HORMONE BIOPHARMACEUTICALS

A hormone is classically defined as a chemical messenger, either a small molecule or a large macromolecule, which is synthesized in an organ or tissue and then secreted into the circulatory system where it subsequently affects separate target organs whose cells bear an appropriate receptor. Hormone actions include the stimulation or inhibition of growth, the induction or suppression of programmed cell death (apoptosis), the modulation of the immune system, the regulation of metabolism, and the preparation for a new activity (e.g., fighting, fleeing, and mating) or phase of life (e.g., puberty, childbirth, and menopause). In many cases, one hormone may regulate the production and release of other hormones. Many hormones can be described as messengers serving to regulate the metabolic activity of an organ or tissue. Hormones also control the reproductive cycle of virtually all multicellular organisms.

Animal hormones essentially fall into four different chemical classes: (1) amine-derived hormones, (2) peptide/protein hormones, (3) steroid hormones, and (4) lipid or phospholipid hormones. Amine-derived hormones such as catecholamines and thyroxine are derivatives of the amino acids tyrosine and tryptophan. Peptide hormones including insulin, growth hormone, and vasopressin consist of polypeptides ranging in length from 5 to 200 residues. Steroid hormones such as estrogen

and testosterone are derived from cholesterol. The adrenal cortex and the gonads are primary sources for these hormones. Lipid and phospholipid hormones are derived from lipids such as linoleic acid and phospholipids such as arachidonic acid. For this chapter we are primarily concerned with peptide hormones.

The functional diversity of peptide/protein hormones is enormous. There are as many potential classes of these hormones as there are classes of hormonally regulated biological functions. If we also include potential antagonists to natural hormonal function (i.e., peptides that may bind competitively or noncompetitively to hormone receptors), we could double the number of classes. However we can generally group subcategories so that pharmaceutically important peptide hormones might be classified into the following broad conceptual categories:

1. Homeostatic regulators control basic physiological conditions such as blood glucose levels (insulin), water retention (vasopressin, desmopressin), or uterine muscle contractions (oxytocin).
2. Fertility regulators can be used to either inhibit or promote fertility. Although luteinizing hormone (LH) and follicle stimulating hormone (FSH) are regulated indirectly through the steroid hormone (estrogen and progesterone) levels to suppress fertility, direct introduction of the peptide hormones themselves is used to induce development and release of multiple ova.
3. Growth/division regulators may either include hormones regulating growth of the entire body (e.g., human growth hormone) or of a specific tissue type (e.g., erythropoeitin for hematopoeisis, palifermin for mucositis, or aldesleukin for lymphocytes). Some fertility hormones (e.g., leuprolide) may act indirectly as general or specific growth inhibitors.
4. Immunomodulatory hormones modulate the normal immune or inflammatory responses. This class can include hormones that upregulate the immune response (e.g., interferon) or that suppress it (e.g., cyclosporine).

If the definition of a hormone were expanded to include localized secreted or cell-surface signaling molecules, many other kinds of regulators would join this list, including molecules with roles in developmental fate specification and differentiation (e.g., netrins, semaphorins, and morphogens). Because the effects of such signals are highly localized, their use as effective therapeutic agents would require greater precision in administration than current systemic delivery methods. Table 1 provides a complete list of the peptide/protein hormones that are currently approved by the FDA. More details about some of these are provided below.

### 3.1 Examples of Hormone Biopharmaceuticals

***Homeostatic Hormones.*** Insulin is produced by beta cells found in the islets of Langerhans that are dispersed throughout the pancreas. The islet cells constitute only 1% of total pancreatic tissue. Each islet contains about 3000 endocrine cells with a core of beta cells surrounded by other endocrine cells such as the glucagon-secreting alpha cells, the somatostatin-secreting delta cells, and the pancreatic polypeptide-secreting PP cells. The beta cells first synthesize a 109-amino-acid preproinsulin that is subsequently processed to an 86-amino-acid proinsulin. The

**TABLE 1. FDA-Approved Protein/Peptide Hormone Pharmaceuticals**

| Generic Name | Chemical Name | Brand Name(s) | Description | Indication |
|---|---|---|---|---|
| Abarelix | acetyl-D-β-naphthylAla-D-4-chloroPhe-D-3-pyridylAla-L-Ser-L-N-methylTyr-D-Asp-L-Leu-L-N(e )-isopropyl-Lys-L-Pro-D-Ala | Plenaxis | Anti-Testosterone, Antineoplastic. Synthetic decapeptide antagonist to gonadotropin releasing hormone (GnRH) | Palliative treatment of advanced prostate cancer |
| Aldesleukin | Human interleukin 2 (modified) | Proleukin | Antineoplastic. Produced in recombinant *E. coli. N*onglycosylated. N-terminal Ala deleted. Cys125Ser substitution. | adults with metastatic renal cell carcinoma |
| Anakinra | Human interleukin-1 receptor antagonist protein | Kineret | Antirheumatic, Immunomodulatory. Recombinant human interleukin-1 receptor antagonist. Produced in *E. coli.* 153 residues. | adult rheumatoid arthritis |
| Becaplermin | Human platelet-derived growth factor | Regranex | Topical anti-ulcer. Recombinant protein produced in the yeast, *Saccharomyces cerevisiae.* Homodimer. | topical treatment of skin ulcers (from diabetes) |
| Calcitonin, salmon | Salmon calcitonin | Miacalcin | Antihypocalcemic, Antiosteporotic. Synthetic peptide, 32 residues long formulated as a nasal spray. | osteoporitis |
| Cetrorelix | acetyl-D-3-(2′-naphtyl)-ala-D-4-chlorophe-D-3-(3′-pyridyl)-ala-L-ser-L-tyr-D-citruline-L-leu-L-arg-L-pro-D-ala-amide | Cetrorelix, Cetrorelixum, Cetrotide | Infertility Agent. Synthetic hormone blocks premature ovulation by antogonizing GnRH-stimulated LH release. | inhibition of premature LH surges in women undergoing controlled ovarian stimulation |

| | | | | |
|---|---|---|---|---|
| Choriogonadotropin alfa | Human chorionic gonadotropin | Novarel, Ovidrel,; Pregnyl, Profasi | Fertility Agent. Has 2 subunits, alpha = 92 residues, beta = 145 residues, each with N-and O-linked carbohydrates. | female infertility |
| Darbepoetin alfa | Human erythropoietin | Aranesp | Antianemic. Human erythropoietin with 2 aa substitutions to enhance glycosylation. Produced in recombinant CHO cells. | anemia (from renal transplants or certain HIV treatment) |
| Denileukin diftitox | Diptheria toxin-Interleukin-2 fusion protein | Ontak | Antineoplastic. Recombinant diphtheria toxin fragments A and B (1–387) fused with interleukin-2 (1–133). Produced in *E. coli*. | cutaneous T-cell lymphoma |
| Desmopressin | 1-desamino-8-D-arginine vasopressin | Concentraid; DDAVP, Stimate | Renal Agent. Synthetic modification of vasopressin. Deaminated and Arg8 is D not L. | primary nocturnal enuresis, diabetes insipidus, temporary polyuria and polydipsia |
| Desmopressin acetate | Desmopressin acetate | Stimate | Antidiuretic. Synthetic analogue of the natural pituitary hormone 8-arginine vasopressin (ADH). 3-mercaptopropionic acid replaces N-terminal Cys and D-Arg replaces Arg. | diabetes insipidus, prevention of polydipsia, polyuria and dehydration |
| Epoetin alfa | Human erythropoietin | Epogen, Epogin, Epomax, Eprex, NeoRecormon | Antianemic. Produced in recombinant CHO cells. | anemia (from renal transplants or certain HIV treatment) |

**TABLE 1.** *Continued*

| Generic Name | Chemical Name | Brand Name(s) | Description | Indication |
|---|---|---|---|---|
| Eptifibatide | N 6-(aminoiminomethyl)-N 2-(3-mercapto-1-oxo Pro-L-Lys-L-Gly-L-Asp-L-Try-L-Pro-L-Cys | Integrilin | Anticoagulant, Antiplatelet. Synthetic cyclic hexapeptide binds to platelet receptor glycoprotein and inhibits platelet aggregation. | myocardial infarction, acute coronary syndrome |
| Felypressin | 2-(L-Phenylalanine)-8-L-lysine-vasopressin | Felipresina, Felipressina, Felypressin, Felypressine, Felypressinum | Renal Agent, Vasoconstrictor. Hemostatic. Synthetic analog of Lypressin (Lys-vasopressin) with more vasoconstrictor than antidiuretic action. | localizing agent for anesthetics |
| Filgrastim | Human granulocyte colony stimulating factor | Neupogen | Antineutropenic, Antiinfective, Immunomodulatory. 175 residues, expressed in *E. coli.* | Increases leukocyte production. Treatment of non-myeloid cancer, neutropenia, and bone marrow transplant. |
| Follitropin alfa/beta | Human follicle stimulating hormone | Gonal-F | Fertility Agent. Fertility Agent. Recombinant.Two non- covalently linked, non-identical glycoproteins with 92 and 111 amino acids. | female infertility |
| Follitropin beta | Human follicle stimulating hormone | Follistim | Fertility Agent. Recombinant. Two non-covalently linked, non-identical glycoproteins with 92 and 111 amino acids. | female infertility |
| Glucagon recombinant | Human Glucagon | GlucaGen, Glucagon | Antihypoglycemic Agent. 29 residue peptide hormone. Produced in *E. coli.* | severe hypoglycemia, also used in gastrointestinal imaging |

| | | | | |
|---|---|---|---|---|
| Goserelin | pyro-Glu-His-Trp-Ser-Tyr-D-Ser($Bu^t$)-Leu-Arg-Pro-Azgly-$NH_2$ acetate | Zoladex | Antineoplastic Agent. Synthetic hormone. Stops production of testosterone in men. Decreases production of estradiol in women. | breast cancer, prostate cancer, endometriosis |
| Insulin Glargine recombinant | Recombinant human insulin | Lantus | Hypoglycemic agent. Produced in *E. coli.* IDiffers from human insulin in that Asn21 is replaced by Gly and two Arg are added to the C-terminus of B chain. | diabetes (type I and II) |
| Insulin Lyspro recombinant | Human Insulin | Humalog, Insulin Lispro | Hypoglycemic Agent. Monomeric form of regular insulin hexamer. Absorbed more rapidly than regular insulin. Lsy28 and Pro29 swapped in B chain | diabetes (type I and II) |
| Insulin recombinant | Human insulin | Novolin R | Hypoglycemic Agent. 51 residues. Produced as 2 chains in recombinant *E. coli.* Chains linked chemically *in vitro.* | diabetes (type I and II) |
| Insulin, porcine | Porcine insulin | Iletin II | Hypoglycemic agent, Isolated from pro-insulin of pig pancreas. Forms hexameric structure. | diabetes (type I and type II) |
| Interferon alfa-2a | Human interferon alpha 2a | Roferon A | Antineoplastic, Antiviral. 165 residues. Produced in recombinant *E. coli.* | chronic hepatitis C, hairy cell leukemia, AIDS-related Kaposi's carcoma, chronic myelogenous leukemia |

**TABLE 1.** *Continued*

| Generic Name | Chemical Name | Brand Name(s) | Description | Indication |
|---|---|---|---|---|
| Interferon alfacon-1 | Consensus sequence interferon alpha | Infergen | Antineoplastic, Antiviral, Immunomodulatory. Consensus sequence derived Interferon alpha, differs from human interferon alpha by 20 residues (88% identical to IFN alpha 2a) | hairy cell leukemia, malignant melanoma, and AIDS-related Kaposi's sarcoma |
| Interferon alfa-n1 | Human interferon alpha | Wellferon | Antiviral, Immunomodulatory. Purified, natural glycosylated human interferon alpha. 166 residues. | venereal or genital warts caused by the Human Papiloma Virus |
| Interferon alfa-n3 | Human interferon alpha (2a, 2b and 2c) | Alferon | Antineoplastic, Antiviral, Immunomodulatory. Purified, natural human interferon alpha. proteins (includes 2a, 2b and 2c) | hairy cell leukemia, malignant melanoma, AIDS-related Kaposi's sarcoma, laryngeal papillomatosis in children, some types of hepatitis |
| Interferon beta-1a | Human interferon beta | Avonex | Antiviral, Immunomodulatory. 166 residues, glycosylated. Produced in recombinant CHO cells. Primary sequence identical to natural human interferon beta. | relapsing/remitting multiple sclerosis. Also condyloma acuminatum. |
| Interferon beta-1b | Human interferon beta | Betaseron | Antiviral, Immunomodulatory. 165 residues. Cysteine 17 substituted with serine. Produced in *E. coli*, nonglycosylated. | relapsing/remitting multiple sclerosis |
| Interferon gamma-1b | Human interferon gamma-1b | Actimmune | Antiviral, Immunomodulatory. Produced from recombinant *E coli*. Purification by column chromatography. | chronic granulomatous disease, osteopetrosis |

| | | | | |
|---|---|---|---|---|
| Lepirudin | Hirudo medicinalis hirudin (variant 1) | Refludan | Anticoagulants, Antithrombotic. Identical to natural hirudin except for substitution of leu for ile at N-terminal and absence of a sulfate group on tyr63. Produced in yeast cells. | heparin-induced thrombocytopenia |
| Leuprolide | 5-oxo-pro-his-tryp-ser-tyr-leu-leu-arg-N-ethyl-pro | Eligard | Antineoplastic, analog of gonadotropin releasing hormone. | prostate cancer, endometriosis, uterine fibroids, premature puberty |
| Lutropin alfa | Human luteinizing hormone | Luveris | Fertility Agent. Recombinant human LH produced in yeast with 2 subunits, alpha = 92 residues, beta = 121 residues. | female infertility |
| menotropins | mixture of LH and FSH | Gengraf, Neoral, Restasis, Sandimmune | Immunomodulatory. Mixture of LH and FSH extracted from urine of postmeonopausal women. | transplant rejection, rheumatoid arthritis, severe psoriasis |
| Octreotide acetate | Octreotide acetate | Sandostatin LAR | Anabolic Agent, Hormone replacement. Acetate salt of cyclic octapeptide. Pharmacologic properties mimick somatostatin. | acromegaly and side effects from cancer chemotherapy |
| Oprelvekin | Human interleukin 11 | Neumega | Coagulant, Thrombotic. | Increases reduced platelet levels due to chemotherapy |
| Oral interferon alfa | Human interferon alpha | Veldona | Antiinfective; Antineoplastic, Interferon alpha, 165 amino acids. | oral warts arising from HIV infection |
| Oxytocin | Oxytocin | Oxytocin, Pitocin, Syntocinon | Antitocolytic, Synthetic 9 residue cyclic peptide. Prepared synthetically to avoid possible contamination with vasopressin. | assist in labor, labor induction, uterine contraction induction |

**TABLE 1.** *Continued*

| Generic Name | Chemical Name | Brand Name(s) | Description | Indication |
|---|---|---|---|---|
| Palifermin | Human keratinocyte growth factor | Kepivance | Antimucositis. 140 residues long, expressed in *E. coli*. First 23 residues removed to improve stability. | mucositis (mouth sores) |
| Pegfilgrastim | PEGylated human granulocyte colony stimulating factor | Neulasta | Anti-infective, Antineutropenic, Immunomodulatory. 175 residues. Produced in recombinant *E. coli*. | non-myeloid cancer, neutropenia, bone marrow transplant |
| Peginterferon alfa-2a | PEGylated human interferon alpha 2a | Pegasys | Antineoplastic, Antiviral, Immunomodulatory. Covalent conjugate of recombinant alfa-2a interferon with single polyethylene glycol (PEG) chain. Produced in *E. coli*. 165 aa. | hairy cell leukemia, malignant melanoma, and AIDS-related Kaposi's sarcoma |
| Peginterferon alfa-2b | PEGylated human interferon alpha 2b | PEG-Intron | Antineoplastic, Antiviral, Immunomodulatory. 165 amino acids. Has a single amino acid substitution (Lys23Arg). | hairy cell leukemia, malignant melanoma, and AIDS-related Kaposi's sarcoma |
| Pegvisomant | PEGylated human growth hormone | Somavert | Anabolic Agent, Hormone replacement. Selectively binds to GH receptors on cell surfaces and blocks binding of endogenous GH. | acromegaly |
| Sargramostim | Human granulocyte macrophage colony stimulating factor | Leucomax, Leucine | Antineoplastic, anti-infective. Recombinant GM-CSF is expressed in yeast. 127 residue glycoprotein. | cancer and bone marrow transplant |
| Secretin, synthetic | Human secretin | Secremax, SecreFlo | Diagnostic Agent. Produced in the S cells of the duodenum in response to low local pH. Stimulates secretion of bicarbonate from liver and pancreas. | diagnosis of pancreatic exocrine dysfunction and gastrinoma |

| | | | | |
|---|---|---|---|---|
| Sermorelin | Human growth hormone-releasing hormone fragment | Geref | Anabolic Hormone replacement. Acetate salt of an amidated synthetic 29-amino acid peptide from amino-terminal segment of the naturally occurring growth hormone-releasing hormone. | dwarfism, HIV-induced weight loss |
| Somatropin recombinant | Human growth hormone | BioTropin, Genotropin, Humatrope, Norditropin, Nutropin, NutropinAQ, Protropin, Saizen, Serostim, Tev-Tropin | Anabolic Agent, Hormone replacement. 191 residues. Produced in recombinant *E. coli*. | dwarfism, acromegaly, and prevention on HIV-induced weight loss |
| Thyrotropin Alfa | Human alpha thyrotropin | Thyrogen | Diagnostic. Heterodimeric glycoprotein with two noncovalently linked subunits identical to human pituitary TSH. | detection of residueal or recurrent thyroid cancer |
| Urofollitropin | Human follicle stimulating hormone | Fertinex, Metrodin | Fertility Agent. Purified from urine of postmenopausal women. | female infertility |
| Vasopressin | Cys-Tyr-Phe-Gln-Asn-Cys-Pro-Arg-Gly-NH2 (Cys linked by- S-S-bridge) | Pitressin; Pressyn | Antidiuretic, Vasoconstrictor. Nine amino acid peptide secreted from posterior pituitary. Binds receptors in collecting tubules of the kidney and promotes reabsorbtion of water. | enuresis, polyuria, diabetes insipidus, polydipsia, oesophageal varices with bleeding |

A-chain and B-chain regions are connected by the C-peptide residues and by two disulfide bridges. Proinsulin is further cleaved by the Golgi apparatus before secretion, releasing the C-peptide (35 residues) from the two insulin chains (51 residues total) joined by disulfide bonds. Insulin is secreted into the blood by the beta cells and carried to its major target tissues in the liver, adipocytes, and muscles. Insulin binds to the insulin receptor tyrosine kinase localized in the plasma membrane of target cells leading to receptor autophosphorylation and a signal transduction cascade that is known to regulate more than 100 downstream genes.

Banting, Best, Collip, and MacLeod first isolated insulin in 1921 and demonstrated its therapeutic potential in treating type I diabetes (diabetes mellitus) in 1922 [20]. Before the development of recombinant forms of the hormone, it was isolated from bovine or porcine pancreata. Recombinant human insulin is produced in either *Escherichia coli* bacteria or *Saccharonlyces cerivisiae* yeast strains that have been genetically engineered. Active insulin consists of two peptide chains (called A and B) joined by disulfide linkages (Figure 1). The A chain is 21 amino acids long, whereas the B chain is 30 amino acids long. *E. coli* lack the protease necessary to cleave proinsulin, so A and B chains are either produced in separate cells, purified, and oxidized chemically to form the disulfide linkages, or full-length proinsulin is produced and cleaved in a separate *in vitro* proteolytic reaction (reviewed in Ref. 21). Insulin produced in *S. cerivisiae* is genetically modified so that A and B chains are linked directly or by a short synthetic sequence. The final product is converted into human insulin by a trypsin-mediated transpeptidation reaction carried out in an aqueous–organic solvent media, in the presence of excess threonine ester. Human insulin was the first animal protein to have been made in bacteria in such a way that its structure is absolutely identical to that of the natural molecule. Modifications to the original recombinant insulin have included reversing selected amino acid positions (insulin Lispro) and replacing and adding selected

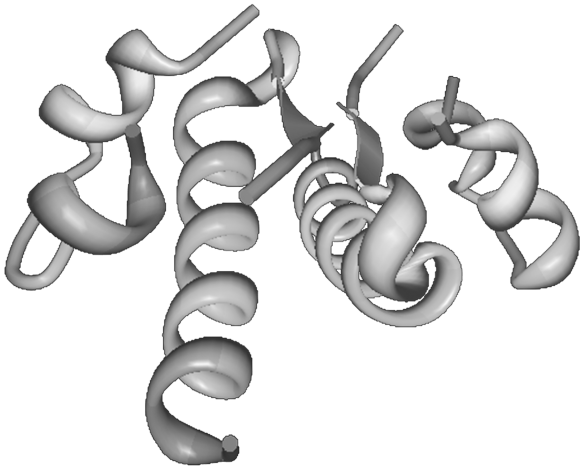

**Figure 1.** The x-ray crystal structure of a human insulin homodimer (A/B/C/D) at 1.5-Å resolution (PDB ID: *1ZEH*). Each component of the dimer is composed of a smaller chain (A/C of 21 residues) linked by disulfide bridges to a larger chain (B/D of 30 residues).

amino acids (insulin Glargine and Aspart). The purpose of these modifications is to improve the kinetics or duration of drug action.

Insulin is generally administered parenterally (subcutaneous injection), although many researchers and companies are developing alternative administration methods, i.e., oral, buccal, (Oral-lyn) or pulmonary (Exubera) [10]. These methods require greatly increased dosage in the case of buccal or pulmonary administration or additives to improve the bioavailability of orally administered insulin (including encapsulation, permeabilization, or chemical stabilization additives). No intestinally absorbed (oral) version of insulin has yet been proven effective and received FDA approval.

Another group of important homeostatic regulatory hormones are oxytocin and vasopressin. These two hormones are structurally related 9-mer oligopeptides produced in the hypothalamus and then transported along axons to the posterior pituitary where they are stored until they are secreted. Oxytocin and vassopressin are identical in sequence and structure except for residues three and eight; yet they have very different physiological effects. Oxytocin stimulates mammary milk secretion and contraction of uterine smooth muscle and likely plays a role in labor initiation. Vasopressin (also known as antidiuretic hormone—ADH) is a vasoconstrictor and reduces the water content of urine by increasing the reclamation of urinary water by the renal collecting duct. However, due to their structural similarity, there is a slight overlap of physiological response to these hormones; oxytocin has about 1% the antidiuretic potency of vasopressin, whereas vasopressin has about 15% the mammary secretion potency of oxytocin [11]. Both proteins are cyclic peptides with cysteines at positions one and six being joined by disulfide bonds, forming a six-residue ring structure with a flexible three-residue amidated carboxyl-terminal tail. Both hormones act through binding to specific receptors that are coupled to G-protein signal transduction pathways. G-protein activation leads to activation of inositol triphosphate and diacylglycerol, resulting in internal $Ca^{2+}$ release and a MAP kinase cascade. The end result is phosphorylation of the myosin light chain leading to smooth muscle contraction. Oxytocin receptors are found in mammary myoepithelial and uterine smooth muscle cells, whereas vasopressin receptors are found predominantly in the cortical glomerular mesangial cells and medullary vascular smooth muscle cells of the kidney. In addition to stimulating contraction of vascular cells, vasopressin also causes translocation of aquaporin water channels to the apical surface of cells in the renal collecting duct. Traumatic or metastatic damage to the hypothalamus as well as rare hereditary disease can result in a lack of vasopressin, leading to the development of diabetes insipidus [11]. Those inflicted with this disease produce large amounts of dilute urine and must drink large amounts of fluid to replace what is lost.

Vasopressin is administered parenterally, but its synthetic analog, desmopressin acetate, can be administered by rhinal tube or orally, by tablet, for the treatment of diabetes insipidus [10]. Oxytocin and Vasopressin are manufactured through solid-phase synthesis [22, 23]. Antidiuretics are in fairly widespread use, having been prescribed in about 15% of all physician visits in the United Sates in 2002–2003 up from about 10% only 7 years before [24]. Oxytocin is used to induce labor by parenteral administration, either through subcutaneous or intravenous injection, or can be administered nasally to increase milk production in nursing mothers. Oxytocin induction of labor occurred in approximately 20% of births in the United States in 2003 [24].

***Fertility Hormones.*** LH and FSH are heavily N-glycosylated glycoprotein dimers containing a common alpha chain of 92 amino acids and a unique 121 (LH) or 111 (FSH) residue beta chain. These hormones regulate human sex steroid production (LH) and gametogenesis (LH and FSH) in both males and females. They are expressed in a cyclic fashion by the pituitary gonadotrope cells. In males, for example, there is one LH pulse of consistent amplitude approximately every 90 minutes, but in females the amplitude and frequency of LH pulses varies throughout the ovarian cycle [11]. LH activates a seven-transmembrane G-protein-coupled receptor that is found primarily in the Leydig cells of the testes of males and in granulosa and theca cells in the female ovarian follicle. The FSH seven-transmembrane G-protein-coupled receptors are found only in Sertoli cells in the male testes or in granulosa cells in the female ovarian follicle. In males, LH induces the formation of testosterone in its target cells, whereas FSH induces production of androgen. In females, the complex cooperative activation of LH and FSH receptors regulates the levels of the primary sex steroids, progesterone, and estradiol. LH and FSH can be isolated from the urine of postmenopausal women, or they can be made in recombinant form by expression of transgenes in Chinese hamster (CHO) cells. LH and FSH are usually administered together to treat female infertility by stimulating ovulation and follicular maturation of the ovulated follicle into a corpus luteum. LH and FS (Lutropin—LH—or Repronex—combined LH and FSH) are administered parenterally. Of the approximately 6.7 million women with fertility problems in the United States in 1995, 35% were treated with some form of ovulation-inducing drug [25].

***Growth Regulatory Hormones.*** Growth hormone (also known as somatotropin or somatropin) was first isolated in 1944 [11]. It is an anabolic hormone responsible for inducing cell division and controlling an individual's height. Growth hormone also increases calcium retention, strengthens and increases the mineralization of bone, increases muscle mass, induces protein synthesis, and stimulates the immune system. Mature growth hormone is a single-chain, nonglycosylated 191 residue polypeptide formed in the anterior pituitary. It is derived from a larger prohormone that includes a cleaved secretory signal peptide. In addition to the pituitary, growth hormone is also expressed at a much lower level in other tissues, including the central nervous system, mammary glands, gonads of both sexes, and hematopoietic and immune system cells. The growth hormone receptor was the first identified cytokine receptor and is an approximately 620-residue single-pass transmembrane protein that is highly glycosylated and ubiquitinated. A proteolytic fragment of the receptor, growth hormone binding protein, circulates throughout the bloodstream and increases the half-life of growth hormone by reducing its rate of degradation and clearing. The growth hormone receptor has been observed throughout the body in tissues as diverse as the gastrointestinal tract; the musculoskeletal system; the cardiorespiratory system; hematopoietic and immune systems; the central nervous system; integument, renal, and urinary systems; and the endocrine system. It is also found on cells of all major lineages in the developing fetus. However, in adults, it achieves its highest level of expression in the liver. A single growth hormone ligand binds to two receptors causing dimerization. This trimer, lacking its own kinase function, then recruits a member of the Janus tyrosine kinase family (JAK2) leading to activation of STAT-mediated gene transcription along with

activation of the Ras-Raf-MEK, phosphatidylinositol 3-kinase (PI 3-kinase) and protein kinase C pathways [26, 27]. As might be expected, this has diverse and widespread effects on physiology and morphology. The main effect is to stimulate both chondrocyte and osteoblast proliferation leading to longitudinal bone growth. In addition, muscle mass increase is favored as an increase in lipolytic activity in adipose tissue leads to a reduction in total fat. Growth hormone also affects differentiation within the central nervous system with consequent cognitive effects such as enhancement of long-term memory, learning, and rapid eye movement sleep, although only when administered at supraphysiological levels.

A reduction or loss of growth hormone production, an autosomal recessive disorder, leads to obvious signs of dwarfism as early as 6 months of age. Laron syndrome is a defect in the growth hormone receptor also leading to sharply reduced stature but is resistant to exogenous treatment with growth hormone [29]. By contrast, chronic overproduction of growth hormone can lead to acromegaly, a sometimes fatal condition due to resulting cardiovascular, cerebrovascular, respiratory, and metabolic diseases.

Recombinant human growth hormone for injection is produced in *E. coli*. The mature product is identical to the natural human hormone, although it is initially translated with a signal peptide that is cleaved by the cell as the peptide is synthesized and exported into the periplasm. Growth hormone production can also be increased through administration of semorelin acetate, the acetate salt of an amidated synthetic 29-amino acid peptide corresponding to the amino-terminal residues of human growth hormone releasing hormone [11]. By contrast overproduction of growth hormone leading to acromegaly can be treated by a PEGylated version of growth hormone, known as Pegvisomant [30]. This chemically altered version of growth hormone competitively binds to the receptors without activating them, thus interfering with the action of the elevated endogenous hormone.

Another example of a growth regulatory hormone is erythropoietin. Erythropoeitin (EPO) is a 166-residue cytokine, produced primarily by interstitial cells near the proximal tubular cells of the kidney. EPO stimulates erythropoiesis, the formation of red blood cells. The hormone was initially partially purified by Goldwasser and colleagues from the urine of anemic patients in 1977, and the cDNA was cloned by Lin et al. and Jacobs et al. in 1985 [31, 32]. The protein is highly N- and O-glycosylated, with up to 40% of its weight being carbohydrate. Each glycosylation chain is terminated with a sialic acid, which may also be incorporated into the chain, and the total sialylation affects both receptor-binding affinity and serum half-life (the latter is of greater importance in total EPO biological activity). Thus, the kidney sets the hematocrit (the percent of whole blood that is composed of red blood cells) at a normal value of 45% by regulating red cell mass through EPO production and plasma volume through excretion of salt and water [11, 33]. EPO released into the blood travels to the bone marrow and binds to its cell-surface receptor on erythroid-specific precursors to stimulate red blood cell production. Like the related cytokine, human growth hormone, the binding of erythropoietin to its cognate receptor leads to dimerization and subsequent recruitment of the JAK2 tyrosine kinase signal transduction pathways to stimulate erythrocyte proliferation and differentiation. EPO has been approved for the treatment of anemia associated with chronic renal failure, with chemotherapy in cancer patients, and anemia associated with surgery. EPO is also beneficial for treatment of anemia

associated with Zidovudine (e.g., AZT) therapy for patients infected with HIV. Because the functional protein must be correctly glycosylated posttranslationally, it is prepared from recombinant CHO cells expressing the erythropoietin precursor, including the 27-amino-acid signal peptide. A variant form (darbepoietin alpha) includes 2 amino acid substitutes that add N-glycosylation sites. The drug is currently only available for parenteral administration.

***Immunomodulatory Hormones.*** Interferons constitute a family of species-specific vertebrate proteins (type I: IFN-α, IFN-β, IFN-ε, IFN-κ, and IFN-ω and type II: IFN-γ in humans) that confer general resistance to a broad range of viral infections, affect cell proliferation, and modulate immune responses [34]. Interferons were originally characterized in 1957 by Isaacs and Lindenmann as soluble proteins that induce antiviral activity in chicken cells. IFN-β and IFN-ω have only one functional copy each, but IFN-α is found in 13 different alleles producing 12 functional protein isoforms. The various IFN-α members each exhibit a distinct profile of antiviral, antiproliferative, and immunostimulatory effects, but variants of the IFNA2 gene (IFN-α2a, IFN-α2b, and IFN-α2c) provide the basis for most interferon therapeutics. Although the type I interferons invoke different cellular responses, they share a common dimeric receptor, IFNAR, which is found in low levels on the plasma membrane of all cell types. Each IFNAR subunit is composed of a single-pass transmembrane protein associated with a member of the JAK tyrosine kinase family (TYK2 for IFNAR-1 and JAK1 for IFNAR-2c). Upon binding of the ligand, the receptor dimerizes and cross-autophosphorylates the cytoplasmic receptor tails along with TYK2 and JAK1. This induces the phosphorylation and subsequent nuclear translocation of a STAT signaling complex and activation of gene-specific transcription [36]. It is not yet known how the many different species of IFN-α can lead to different cellular responses while acting through the same receptor, although data have indicated different ligands may have slightly different binding sites and may involve other receptors in the activated complex.

Parenteral administration of natural and recombinant interferons for the treatment of hairy cell leukemia, malignant melanoma, AIDS-related Kaposi's sarcoma, and a variety of viral diseases has been approved by the FDA. The type I IFNs exhibit a wide breadth of biological activities, including antiviral and antiproliferative effects as well as stimulation of MHC I antigen presentation and NK-cell activation. This makes them ideal for the treatment of many serious illnesses but can also result in undesirable side effects, especially at the high dosages normally employed. Oral administration of low-dose interferons has been studied in clinical trials but has not yet received approval for general use. Interferon-alpha was approved by the FDA on February 25, 1991 as a treatment for hepatitis C [37, 38]. In January 2001, the FDA approved PEGylated interferon-alpha. The PEGylated form is injected once weekly, rather than three times per week for conventional interferon-alpha.

### 3.2 The Future of Hormone Biopharmaceuticals

Research into new peptide hormone pharmaceuticals continues along several lines. Probably the most active area of research is into alternative delivery methods such as oral, nasal, buccal, transdermal, or pulmonary. All of these alternatives share issues of barrier penetration (whether mucosal or epidermal), protein stability, rate of clearance, and bioavailability.

The epithelial mucosa lining intestinal, buccal, nasal, and pulmonary tracts present significant barriers to protein or peptide hormone access to the circulatory system. Increasing the mucosal adhesiveness of therapeutic proteins delivered through the mucosa leads to an increase in the duration of residence at absorption sites. This is typically accomplished through encapsulation of the protein in micro- or nanoparticles made of mucoadhesive polymers [39, 40]. Increased duration in the mucosa also exposes the protein drug to increased extracellular proteolytic activity. Therefore therapeutic proteins and peptides can be chemically altered to reduce their susceptibility to proteases (e.g., in the intestine). Oral desmopressin acetate, for example, is partly protected against proteolytic degradation through amidation of the carboxyl-terminus, replacement of L-Arg with D-Arg and replacement of the amino-terminal residue with mercaptopropionic acid [41].

In addition to the mucosa, the epithelial cells also present a formidable barrier to the uptake of therapeutic proteins. A variety of strategies have been used to overcome this barrier, most notably through changing the hydrophobicity of the protein by the covalent attachment of lipophilic moieties. For example, the oral hexyl insulin monoconjugate 2 (HIM2) includes seven to nine units of PEG and a hexyl alkyl chain covalently attached to Lys29 and has proven safe and somewhat effective in controlling postprandial hyperglycemia in patients with type 1 diabetes mellitus during phase I/II clinical trials [42]. More recently, epithelial permeation enhancers, such as cationic or polymeric derivatives of the hydrophilic polysaccharide, chitosan, have proven effective in enhancing the effectiveness of oral [43], nasal [44], or pulmonary [39] administration of insulin or calcitonin in experimental settings. The precise mechanism by which these various permeant enhancers work is not yet understood, although, at least in the intestine, the modification of the tight junctions between cells in the intestinal epithelia is thought to be crucial [40].

In addition to novel delivery methods, several new peptide hormones with therapeutic potential have been recently identified. Ghrelin is a 28-residue, posttranslationally modified peptide that is secreted primarily from cells in the stomach and that has been implicated in the regulation of feeding behavior [11]. There is evidence that its orexigenic (appetite stimulating) activity is mediated, in part, by the hypothalamic hormone neuropeptide Y. Ghrelin is also the endogenous ligand for the growth hormone secretagogue receptor; it stimulates release of growth hormone from the pituitary independently of hypothalamic growth hormone releasing hormone. Repeated administration of Ghrelin has recently been shown to improve muscle mass and functional capacity in cachectic patients with heart failure or chronic obstructive pulmonary disease [45] and to alleviate chemotherapy-induced dyspepsia in rodents [46]. Leptin is an adipocyte-derived cytokine with an appetite-suppressing effect, likely mediated, in part, through neuropepetide Y. Along with adiponectin, another adipocyte-derived hormone, leptin may increase fat oxidation and promote insulin sensitivity via the AMP-activated protein kinase, thus reducing circulating fatty acids and triacylglycerol [47].

## 4 ANTIBODY BIOPHARMACEUTICALS

Antibodies or immunoglobulins are a class of disease-fighting proteins produced by B cells found in the lymphatic system or bone marrow. Antibodies are produced to

bind a specific antigen that has stimulated the immune system. Their selectivity and tunability to recognize and bind an almost infinite range of large- and small-molecule substrates is what makes antibodies so important to the immune system. Once bound, the antigen can be ingested and destroyed by other cells of the immune system, primarily macrophages. Each antibody consists of four polypeptides—two heavy chains and two light chains joined by disulfide linkages in a hinge region to form a "Y"-shaped molecule (Figure 2a). The arms of the Y are known as the antigen-binding fragments (Fab). The region of the heavy chain associated with the light chains in the Fab contains one variable domain and one constant domain. The light chains each contain one constant domain and one variable domain. The variable domains of both the light and the heavy chain are known collectively as the variable fragment (Fv), and each contains hypervariable regions (known as complementarity-determining regions–CDRs) that are primarily responsible for determining antigen recognition. The lower portion of the "Y," consisting primarily of the heavy chains, is called the constant region. The constant region determines the mechanism used to destroy antigen. There are five classes of immunoglobulins (antibodies) in humans: IgM, IgG, IgE, IgA, and IgD. Most biopharmaceuticals are based on IgG.

Antibodies not only recognize and bind to antigens, but they also direct the process by which the antigens are eliminated or killed. These are called the effector function. The base of the heavy chain (called the constant fragment, Fc) is species-

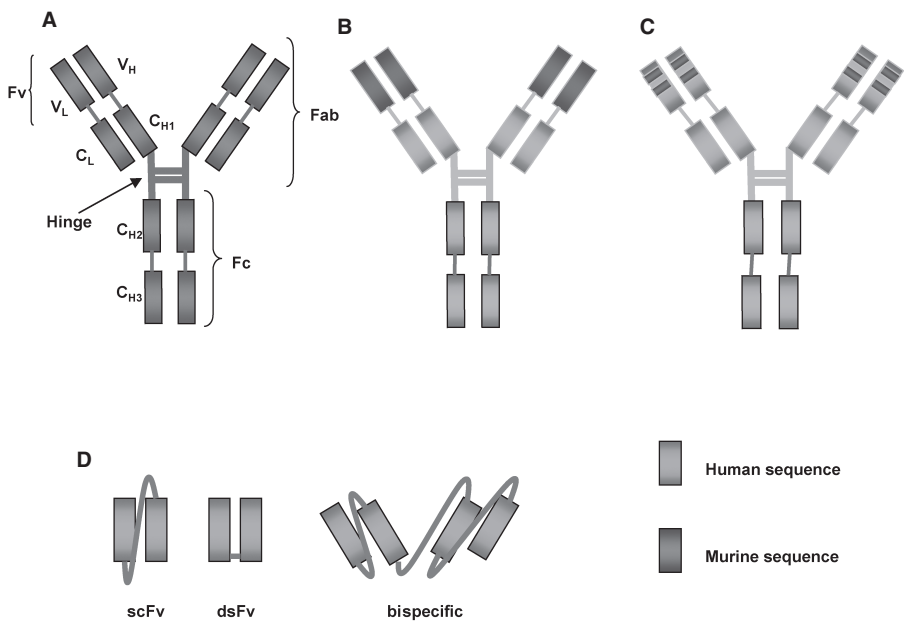


**Figure 2.** IgG antibodies and antibody fragments. (A) Murine IgG antibody showing variable (Fv), antigen-binding (Fab), and constant (Fc) fragments. Heavy and light chains are indicated by suffix H and L, respectively. (B) Chimeric IgG with murine Fv regions fused to human constant regions ($C_L$, $C_{H1}$, and Fc). (C) Humanized IgG with fused murine CDRs in Fv regions. (D) Antibody fragments. scFv—single-chain Fv; dsFc—disulfide-stabilized Fv; bispecific—two linked scFvs with different antigen recognition.

specific and mediates effector functions, including antibody-dependent cellular cytotoxicity (ADCC) and complement-dependent cytotoxicity (CDC). In ADCC, antibodies bind to Fc receptors on the surface of natural killer (NK) cells or macrophages, and trigger phagocytosis or lysis of the targeted cells. In CDC, antibodies kill the targeted cells by triggering the complement cascade at the cell surface. IgG is most efficient in inducing both ADCC and CDC and, therefore, is most suitable for use as a therapeutic molecule.

The first monoclonal antibody (mAb) tested in 1986 as a therapeutic in humans was a murine antibody, Muromonab (OKT3) [6]. Despite the high expectations of mAb therapy, OKT3 failed as a good treatment for transplantation rejection, primarily as a result of severe human anti-murine antibody response in patients. The second mAb marketed for therapeutic use in humans, Abciximab (ReoPro), required nearly 9 years of development to produce a chimeric Ab with mouse Fab and human Fc regions [48]. Since that time, more than 27 recombinant antibody therapeutics have been approved by the FDA, accounting for 24% of all approved protein/peptide therapeutics (Table 2) The current scheme for classifying mAbs is based on variation in their structure (Figure 2b–d).

1) Standard mAbs are produced by fusing murine cultured melanoma cells with isolated immune-reactive spleen cells to produce hybridoma fusions. These hybridomas are assayed *in vitro* for reactivity to the selected antigen and clones of single reactive cells are grown by intraperitoneal injection into naïve mice. Thus, all reactive cells from a single mouse result from a single reactive clone, and hence, the resulting antibodies are called monoclonal.
2) Fc-humanized mAbs are the result of genetic engineering to replace the murine Fc with human Fc in the melanoma cell lines [49].
3) An alternative approach to producing humanized Abs is to graft murine CDRs and a few structural residues from the heavy chain into a human IgG heavy chain [50]. In combination with the normal mouse light chain produced by the hybridoma, this CDR-grafting leads to the development of Daclizumab (Zenapax) in 1997. Instead of generating CDRs in whole animals by antigen immunization, they can also be selected by panning combinatorial phage libraries with recombinant Fabs [51]. Adalimumab (Humira) was the first therapeutic antibody developed with this technique in 2002. Yet another strategy for producing humanized Abs immunizes transgenic mice engineered to express human IgG against the selected antigen [52, 53]. Several therapeutic mAbs derived from such transgenic "humanized" mice are in late-phase clinical trial, including Panitumumab, an anti-EGFR Ab for the treatment of advanced metastatic colorectal cancer.
4) Recombinant bacterial techniques, coupled with the recognition that the entire antibody is not always required, have led to the formation of many therapeutic antibody fragments (Figure 3d). Whole IgG molecules are about 150 kDa, and subsequently, they diffuse into tissue slowly and are cleared from the body slowly as well. These characteristics make them poor candidates in diagnostic imaging or radiotherapy applications. The Fv region is the smallest portion of the Ab that maintains its antigen specificity, and it is small enough to be easily expressed in *E. coli*. Single-chain Fvs (scFvs) are fusion proteins containing both heavy and light variable regions connected through

**TABLE 2. FDA-Approved Antibody Pharmaceuticals**

| Generic Name | Chemical Name | Brand Name(s) | Drug Category | Indication |
|---|---|---|---|---|
| Abciximab | Chimeric human-murine (anti CD41) Ab | ReoPro | Anticoagulant, Antiplatelet. Fab fragment of chimeric human-murine monoclonal antibody binds to the glycoprotein lIb/IIIa receptor of human platelets and inhibits aggregation. | myocardial infarction, adjunct to percutaneous coronory intervention, unstable angina |
| Adalimumab | Humanized anti-TNF antibody | Humira | Antirheumatic, Immunomodulatory. Created using phage display technology. Has human-derived heavy- and light-chain variable regions and human constant regions. It is produced by recombinant DNA technology in a mammalian cell expression system. | for treatment of rheumatoid arthritis |
| Alefacept | Human LFA-3/IgG1 fusion protein | Amevive | Immunomodulatory, Immunosuppresant. Dimeric fusion of extracellular CD2-binding portion of leukocyte function antigen-3 linked to the Fc (hinge, CH2 and CH3 domains) portion of human IgG1. Produced in recombinant CHO cells. | moderate-to-severe chronic plaque psoraisis |
| Alemtuzumab | Humanized anti-CD52 antibody | Campath | Antineoplastic. Humanized mAb specific to lymphocyte antigen (cell surface glycoprotein,CD52). IgG1 with human variable framework and constant regions and murine CDRs. Produced in CHO cells. | B-cell chronic lymphocytic leukemia |

| | | | | |
|---|---|---|---|---|
| Antithymocyte globulin | Rabbit anti-thymocyte globulin | Atgam, Thymoglobulin | Immunomodulatory. Polyclonal antibody depletes T cells responsible for acute organ rejection in transplant patients. | renal transplant rejection |
| Arcitumomab | AntiCEA antibody | CEA-Scan | Diagnostic, Imaging Agent. Fab fragment of murine IgG1 mAb IMMU-4 (also called NP-4 with specificity for carcinoembryonic antigen (CEA) covalently labeled with Technitium 99. | imaging colorectal tumors |
| Basiliximab | Chimeric murine/human anti-CD25 antibody | Simulect | Immunomodulatory, Immunosupressant. Blocks IL-2R alpha on surface of activated T lymphocytes. Produced by mouse myeloma cell line. Contains human heavy- and light-chain constant regions and murine heavy- and light-chain variable regions. | kidney transplant rejection |
| Bevacizumab | Humanized anti-VEGF antibody | Avastin | Antiangiogenesis, Antineoplastic. Contains human framework regions and the murine anti-VEGF CDRs. Produced in recombinant CHO cells. | metastatic colorectal cancer |
| Capromab | Murine anti-7E11 antibody | ProstaScint | Imaging Agent. Murine IgG1 mAb conjugated to linker chelator peptide group for indium attachment. | diagnosis of prostate cancer and detection of intra-pelvic |
| Cetuximab | Humanized anti-EGFR antibody | Erbitux | Antineoplastic. Epidermal growth factor receptor binding Fab. Has variable regions of murine EGFR mAb for the N-terminal portion of human EGFR with human IgG1 heavy- and kappa light-chain constant (framework) regions. | metastases metastatic colorectal cancer |

**TABLE 2.** *Continued*

| Generic Name | Chemical Name | Brand Name(s) | Drug Category | Indication |
|---|---|---|---|---|
| Daclizumab | Humanized anti-CD25 antibody | Zenapax | Immunomodulatory, Immunosupressant. Humanized IgG1 Mab that binds to the human IL-2R (anti-CD25). Composite of 90% human (IgG1 constant plus variable framework regions of the Eu myeloma Ab) and 10% murine Ab (CDR) sequences. | renal transplant rejection |
| Digibind | Anti-digoxin antibody fragment | Digibind | Sheep antibody Fab fragment specific to Digoxin. | digitoxin overdose or digitalis glycoside toxicity |
| Efalizumab | Humanized anti-CD11a antibody | Raptiva | Immunomodulatory, Immunosupressant. Humanized IgG1 anti-CD11a mAb. Produced in recombinant CHO cells. | psoriasis |
| Etanercept | Human tumor necrosis factor receptor fusion protein | Enbrel | Antirheumatic, Immunomodulatory. Dimeric fusion protein consisting of extracellular ligand-binding portion of human p75 TNFR linked to IgG1 Fc portion lacking CH1 domain. Produced by recombinant DNA technology in CHO cells. | severe adult and juvenile rheumatoid arthritis |
| Gemtuzumab ozogamicin | Humanized chimeric anti-CD33 antibody | Mylotarg | Antineoplastic. Recombinant humanized IgG4 conjugated with cytotoxic antitumor antibiotic, calicheamicin. Binds to cell-surface CD33 antigen. Produced in mammalian cell culture. | acute myeloid leukemia |
| Ibritumomab | Murine anti-CD20 antibody | Zevalin | Antineoplastic. Indium conjugated murine IgG1 anti-CD20 mAb. CD20 antigen is found on the surface of normal and malignant B lymphocytes. Produced in recombinant CHO cells. | non-Hodgkin's lymphoma |

| | | | | |
|---|---|---|---|---|
| Immune globulin | Human antibodies | Flebogamma, Gamunex | Anti-infective, Immunomodulatory. Mixture of IgG1 and other Abs derived from human plasma via Cohn fractionation. 70.3% IgG1, 24.7% IgG2, 3.1% IgG3, and 1.9% IgG4. | immunodeficiencies, thrombocytopenic purpura, Kawasaki disease, gammablobulinemia, leukemia, bone transplant |
| Infliximab | Chimeric mouse/human anti-TNF-alpha antibody | Remicade | Immunomodulatory, Immunosuppresants. TNF-alpha binding antibody. Chimeric Ab composed of human constant and murine variable regions. Produced by a recombinant cell line cultured by continuous perfusion | Crohn's disease, psoriasis, rheumatoid arthitis and ankylosing spondylitis |
| Muromonab | Murine anti-CD3 antibody | Orthoclone OKT3 | Immunomodulatory, Immunosuppressant. Purified murine (mouse) mAb, directed against CD3 (T3) receptor on the surface of T lymphocytes. Cultured from murine ascites. 93% monomeric IgG2a. | prevention of organ rejection |
| Natalizumab | Humanized anti a4 integrin antibody | Tysabri, Antegren | Immunomodulatory. Humanized IgG4 mAb produced in murine myeloma cells. Contains human framework regions and murine anti-a4 itegrin CDR. | multiple sclerosis |
| Omalizumab | Humanized anti-IgE antibody | Xolair | Anti-Asthmatic, Immunomodulatory. Recombinant humanized IgG1 mAb binds to human IgE. Produced by recombinant CHO cells. | allergy-induced asthma |

**TABLE 2.** *Continued*

| Generic Name | Chemical Name | Brand Name(s) | Drug Category | Indication |
|---|---|---|---|---|
| Palivizumab | Humanized anti-RSV antibody | Synagis | Antiviral. Humanized mAb (5% murine) directed to an epitope in the A antigenic site of the F protein of respiratory syncytial virus (RSV). Produced by a stable murine myeloma cell line. | respiratory diseases caused by respiratory syncytial virus |
| Rituximab | Chimeric murine/human monoclonal anti-CD20 Ab | Rituxan | Antineoplastic. Chimeric murine/human monoclonal anti-CD20 Ab. Contains murine light- and heavy-chain variable region sequences and human constant region sequences. C20 is found on the surface of normal and malignant B lymphocytes. | B-cell non-Hodgkins lymphoma (CD20 positive) |
| Satumomab Pendetide | Murine anti-TAG72 antibody conjugated with Indium 111 | OncoScint | Imaging Agent. Tumor associated glycoprotein mAb conjugated with Indium 111 for radioimaging colon tumors. | diagnosis of extrahepatic malignant cancers |
| Tositumomab | Murine anti-CD20 antibody | Bexxar | Antineoplastic. Murine IgG2 anti-CD20 mAb. Produced in cultured mammalian cells. Can be covalently linked to radioactive Iodine 131. | non-Hodgkin's lymphoma |
| Trastuzumab | Humanized anti-HER2 antibody | Herceptin | Antineoplastic. A recombinant humanized mAbbinds to the extracellular domain of human EGFR. Produced in CHO cell culture. | HER2-positive metastatic breast cancer |

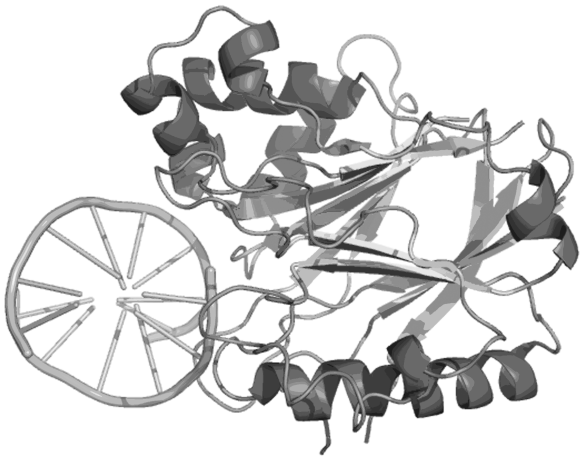

**Figure 3.** The x-ray crystal structure of DNase I in the presence of the palindromic d(GCGATCGC) DNA fragment at 2.0-Å resolution (PDB ID: *2DNJ*).

a short peptide linker. Disulfide-stabilized Fvs (dsFvs) connect the two variable regions through engineered cysteine disulfide linkages. Diabodies are homodimers of scFvs that are covalently linked by a short peptide linker so that the V domains are forced to make intermolecular complexes. Bispecific Fvs are fusions of two complete Fv regions with a different antigen specificity. Variations on these themes incorporating part of the constant region are also possible.

### 4.1 Examples of Antibody Biopharmaceuticals

Muromonab (Orthoclone OKT3) is a purified murine monoclonal antibody specific to the CD3 receptor on the surface of human T cells (T lymphocytes). Used in organ transplant prophylaxis, Muromonab is administered parenterally. It has been effective in reversing corticosteroid-resistant acute rejection in renal, liver, and cardiac transplant recipients. Muromonab binds to the epsilon subunit of the surface glycoprotein CD3 that is associated with the T-cell receptor. It seems to kill CD3 positive cells by inducing Fc mediated apoptosis, antibody mediated cytotoxicity, and complement-dependent cytotoxicity. Immediately after administration CD3-positive T lymphocytes are abruptly removed from circulation, most likely by NK-mediated cytoxicity directed toward the T cells themselves [54] or by inducing T-cell apoptosis [55]. The OKT3-dependent reduction in alloreactive T lymphocytes severely downmodulates transplant rejection. However, because OKT3 contains the mouse Fc regions, its application led to several unfortunate side effects, including acute cytokine overstimulation, flu-like symptoms, and, in rare cases, pulmonary edema, central nervous system disorders, or an anaphylactic reaction to the murine Fc (reviewed in Ref. 56.)

Abciximab (ReoPro) was the first partially humanized antibody approved for therapeutic use. It was produced from transgenic murine melanoma cells containing human heavy-chain genes so that the resulting Fab fragment is a chimera of human and mouse Ab fragments. Abciximab binds to the glycoprotein lIb/IlIa

receptor (a member of the integrin family of adhesion receptors and the major platelet surface receptor involved in platelet aggregation) and competitively inhibits fibrinogen-mediated platelet aggregation and clot formation [57]. Thus it is used in the treatment of myocardial infarction, unstable angina, and as an adjunct to percutaneous coronary intervention. Because the constant regions of the Fab fragment are derived from a human gene, side effects such as those observed with Muromonab are dramatically reduced. Nevertheless, in rare cases, acute profound thrombocytopenia can occur after Abciximab administration leading to a recommendation to monitor platelet count in patients within 11 to 21 hours after treatment [58].

Daclizumab is a humanized IgG mAb that binds to the human IL-2 receptor (anti-CD25). Daclizumab is a composite of human (90%) and murine (10%) antibody sequences. The human sequences were derived from the constant domains of human IgG1 and the variable framework regions of the Eu myeloma antibody [59]. The antibody is produced by expression in a mammalian cell line. The murine sequences were derived from the complementarity-determining regions of a murine anti-CD25 antibody [50]. Daclizumab competitively inhibits the cytokine IL-2 from binding to its CD25 receptor and phosphorylating the IL-2R beta and gamma-chains. This leads to an inhibition of downstream STAT activator genes and a subsequent immunosuppressive effect [60]. Early evaluations of clinical effectiveness of Daclizumab have been promising with reduced side effects relative to OKT3.

Adalimumab (Humira) is an anti-TNF antibody used in the treatment of rheumatoid arthritis. Elevated levels of TNF are found in the synovial fluid of rheumatoid arthritis patients and play an important role in both the pathologic inflammation and the joint destruction that are hallmarks of the disease. Adalimumab binds specifically to TNF-alpha and blocks its general cytokine effects through its interaction with the p55 and p75, thereby reducing TNF-induced inflammation and halting joint destruction. Adalimumab also lyses surface TNF expressing cells *in vitro* in the presence of complement. It was the first FDA-approved antibody therapeutic created using phage display technology [48] resulting in an antibody with human-derived heavy- and light-chain variable regions and human IgG constant regions. Adalimumab is produced by recombinant DNA technology in a mammalian cell expression system from which it is purified and administered by subcutaneous injection. Like other drugs that block TNF, use of Adalimumab has been associated with reactivation of previous serious infections such as tuberculosis, sepsis, and fungal infections. Patients with active infections should not be treated with Adalimumab.

### 4.3 The Future of Antibody Biopharmaceuticals

The development of future therapeutic antibodies revolves around several central issues: (1) further humanization to reduce unwanted Fc-based immune responses, (2) generation of novel CDRs against selected targets, and (3) selection of novel therapeutic Ab targets. Humanization of mAbs has been extensively reviewed elsewhere [61, 62]. A leading focus of recent activity is the use of transgenic mice that have been engineered by a combination of modified traditional transgenic

techniques, yeast artificial chromosome (YAC) technology, and microcell-mediated chromosome transfer. The use of these techniques is required due to the large size of the immunoglobulin gene clusters; the human heavy-, λ-light-, and κ-light-chain loci (located on chromosomes 14, 22, and 2, respectively) span over a megabase each. In addition, murine native light- and heavy-chain regions are deleted in these heavily engineered mice to reduce the production of murine antibodies. Several mAbs produced from these mice are currently under development, including Zanolimumab (anti-CD4 for treatment of cutaneous and peripheral T-cell lymphoma [63]), Panitumumab (anti-epidermal growth factor receptor (EGFR) for treatment of advanced metastatic colorectal and other cancers [64]), and Denosumab (anti-receptor activator of nuclear factor kappa B ligand (RANKL)—a key mediator of the resorptive phase of bone remodeling—for treatment of osteoporosis, rheumatoid arthritis, and cancer with skeletal metastasis [65]).

We have discussed techniques using combinatorial phage libraries to pan for CDRs responding to therapeutic targets. In addition, libraries of novel CDRs can be displayed on the surface of *E. coli* or *S. cerevisiae* or produced *in vitro*, using ribosome display or semi- or fully synthetic antibody libraries (reviewed in Ref. 48). Libraries can be based on CDRs of inoculated or näive individuals and extract variable domains from IgG or IgM regions by polymerase chain reaction (PCR) and subsequent cloning into expression vectors for display (e.g., in phage or *E. coli*). Although naïve libraries (those that have been derived from nonimmunized individuals) have greater complexity than libraries from inoculated individuals, they also tend to have reduced expression and increased toxicity in *E. coli*. *In vitro* methods circumvent these issues and permit harsher selection conditions that may lead to the production of higher affinity antibodies. Phage display is the only technology that has been developed sufficiently that it can be used commercially, and several therapeutic antibodies produced using this approach are in various stages of development, including Mapatumumab (an agonistic mAB that activates the TNF receptor apoptosis induced ligand (TRAIL-R1) to induce apoptosis in multiple types of cancer cells), Metelimumab (anti-TGF-β1 for treatment of sclerodoma), and AbThrax (anti-*Bacillus anthracis* Protective Antigen prevents anthrax from inducing toxin-mediated cell death).

Although novel therapeutic targets continue to be sought routinely, antibody development has mainly been focused on traditional cell-surface and extracellular antigens. Development of novel potential therapeutic targets is dependent on discoveries from fields as diverse as cell biology, cancer biology, microbiology, and developmental genetics. Recently, attention has been focused on the development of antibody derivatives, called intrabodies, that are directed toward intracellular targets [66]. The reducing environment found within the cells makes it difficult for most antibodies to form the disulfide bonds necessary for molecular stability. Intrabodies are formed from a subset of naturally occurring immunoglobulin frameworks that are capable of supporting a variety of antigen-binding structures inside this reducing environment. Alternative methods for discovering intrabodies involve screening scFvs, yeast surface display, phage display, intracellular antibody capture, and directed modifications to the Fc region to improve molecular stability in the cell. Experimental intrabodies that bind H-RAS and block oncogenic transformation of NIH 3T3 cells have recently been developed from scFvs [67] as have anti-angiogenic adenoviral-delivered scFvs [68].

## 5 ENZYME BIOPHARMACEUTICALS

Enzymes are produced by living organisms and function as biological catalysts in specific biochemical reactions of substrates. Enzymes accelerate reactions usually through stabilization of a reactive intermediate at their active site, the place where the reaction occurs. The enzymes, themselves, may be chemically altered during the reaction but are restored to their initial state by its completion. Enzymes are critical to the functioning of most cellular systems as most essential metabolic or catabolic reactions would occur too slowly, or would lead to different products without enzymes. Because of their importance, the mutation, overproduction, underproduction, or deletion of a single critical enzyme can lead to severe disease. For example, mucopolysaccharidosis type VI (or Maroteaux–Lamy syndrome), is a lethal syndrome that develops when patients are deficient in an enzyme called N-acetylgalactosamine 4-sulfatase (galsulfase) and, thus, cannot degrade glycosaminoglycans sufficiently. These compounds accumulate in various tissues leading to permanent, progressive cellular damage affecting an individual's appearance, physical abilities, organ and system functioning, and, in most cases, mental development [69].

More than 3000 human enzymes have been identified and named. Typically the suffix-*ase* is added to the name of the substrate (e.g., lactase is the enzyme that catalyzes the breakdown of lactose) or the type of reaction (e.g., DNA polymerase catalyzes the assembly of DNA). However, this is not always the case, especially when enzymes modify multiple substrates. For this reason, enzymes are formally named and classified based on the recommendations of the Nomenclature Committee of the International Union of Biochemistry, which assigns an Enzyme Commission (EC) number and a formal name. There are six main classes of enzymes: oxidoreductases (EC 1), transferases (EC 2), hydrolases (EC 3), lyases (EC 4), isomerases (EC 5), and ligases (EC 6). Most enzyme pharmaceuticals in current use belong to the hydrolase class such as galsulfase and agalsidase. These enzymes are involved in the hydrolysis (i.e., cleaving) of various bonds. However, there are also examples of pharmaceutical enzymes that belong to oxidoreductases, for example, rasburicase. Some enzymes are highly specific to certain substrates, such as adenosine deamidase, making them ideal drug candidates. On the other hand, some enzymes have wide specificities, such as vitamin A esterase, making these less desirable for drug development.

The potential utility of enzymes as pharmaceuticals was noted many decades ago, and since then, nearly two dozen enzymes have been developed to treat a variety of diseases. Almost all enzyme therapies developed to date are used to deal with a loss of function defect (a mutation that diminishes activity, a low level of production, a deletion). Hence, most enzyme drugs are used as enzyme replacement therapies (ERT) for relatively rare, inborn errors of metabolism (IEMs). As a result, many enzyme therapeutics fall under the FDA's Orphan Drug Designation. However, a few enzyme therapies can also be used to treat much more common conditions such as cancer, heart attacks, and stroke. In the United States, the first enzyme to receive FDA approval was a tissue plasminogen activator called alteplase. This protein, which is now commonly used to treat strokes, was introduced in 1987 as Activase. Since then at least 16 other enzyme drugs have been introduced into the marketplace. Some of these are described in more detail below and in Table 3.

**TABLE 3. FDA-Approved Enzyme Pharmaceuticals**

| Generic Name | Brand Name(s) | Indication | Description |
|---|---|---|---|
| Agalsidase beta | Fabrazyme | Fabry's disease | Recombinant human α-galactosidase A. The mature protein is composed of 2 subunits of 398 aa. Protein is glycosylated and produced by CHO cells. |
| Alglucerase<br>Imiglucerase | Ceredase<br>Cerezyme | Gaucher's disease | Human β-glucocerebrosidase or β-D-glucosyl-N-acylsphingosine glucohydrolase. 497 aa protein (59.3 kDa) with N-linked carbohydrates. Alglucerase is prepared by modification of the oligosaccharide chains of human β-glucocerebrosidase. The modification alters the sugar residues at the nonreducing ends of the oligosaccharide chains of the glycoprotein so that they are predominantly terminated with mannose residues. |
| Alteplase | Activase | Acute myocardial infarction, acute ischemic stroke, and acute pulmonary emboli | Human tissue plasminogen activator, purified, glycosylated, 527 aa purified from CHO cells. |
| Anistreplase | Eminase | Acute pulmonary emboli, intracoronary emboli, and myocardial infarction | Human tissue plasminogen activator, purified, glycosylated, 527 residues purified from CHO cells. Eminase is a lyophilized formulation of anistreplase, the p-anisoyl derivative of the primary Lys-plasminogen-streptokinase activator complex (a complex of Lys-plasminogen and streptokinase). A p-anisoyl group is chemically conjugated to a complex of bacterial-derived streptokinase and human plasma-derived Lys-plasminogen proteins. |
| Asparaginase | Elspar | Acute lymphocytic leukemia & non-Hodgkin's lymphoma | L-asparagine amidohydrolase from *Escherichia coli*. |
| Collagenase | Santyl | Chronic dermal ulcers & severe skin burns | The enzyme collagenase is derived from fermentation of *Clostridium histolyticum*. |

**TABLE 3.** *Continued*

| Generic Name | Brand Name(s) | Indication | Description |
|---|---|---|---|
| Dornase alfa DNase | Dilor<br>Lufyllin<br>Neothylline<br>Pulmozyme | Cystic fibrosis | Recombinant human deoxyribonuclease I. Delivered by aerosol mist. The protein is produced by genetically engineered CHO cells containing DNA encoding for the native human protein, DNase I. 260 aa. |
| Hyaluronidase | Vitrase | For increase of absorption and distribution of other injected drugs, for rehydration and diabetic retinopathy | Highly purified sheep hyaluronidase for injection into the vitreous cavity of the eye. |
| Laronidase | Aldurazyme | Mucopolysaccharidosis | Human recombinant α-L-iduronidase, 628 aa (mature form), produced by recombinant DNA technology in CHO cells. Laronidase is an 83 kDa glycoprotein. The nucleotide sequence is identical to a polymorphic form of human α-L-iduronidase. It contains 6 N-linked oligosaccharide modification sites. |
| Pancrelipase | Cotazym<br>Pancrease<br>Ultrase<br>Zymase | Cystic fibrosis, chronic pancreatitis, and pancreatic duct blockage | Protein mixture isolated from porcine or bovine pancreas, sometimes called pancreatin. Contains 3 enzymes: amylase, lipase, and protease (chymotrypsin). |
| Pegademase | Adagen | Severe combined immunodeficiency disease | Bovine adenosine deaminase derived from bovine intestine that has been extensively PEGylated for extended serum half-life. |
| Pegaspargase | Oncaspar | Acute lymphoblastic leukemia | PEGylated L-asparagine amidohydrolase from *Escherichia coli*. PEGylation substantially (factor of 4) extends the protein half-life. |
| Rasburicase | Elitek | Hyperuricemia | Rasburicase (34.1 kDa) reduces elevated plasma uric acid levels due to tumor lysis (from chemotherapy). It catalyzes the oxidation of uric acid into an inactive and soluble metabolite (allantoin). |

| | | | |
|---|---|---|---|
| Reteplase | Retavase | Acute pulmonary emboli, intracoronary emboli, and myocardial infarction | Human tissue plasminogen activator, purified from CHO cells, glycosylated, 355 aa. Retavase is considered a "third-generation" thrombolytic agent, genetically engineered to retain and delete certain portions of human tPA. Retavase is a mutant of human tPA formed by deleting various aa's present in endogenous human tPA. Retavase contains 355 of the 527 aa of native human tPA (aa 1–3 and 176–527), and retains the activity-related kringle-2 and serine protease domains of human tPA. Three domains are deleted from retavase—kringle-1, finger, and EGF. |
| Streptokinase | Streptase | Acute evolving transmural myocardial infarction, pulmonary embolism, deep vein thrombosis, arterial thrombosis or embolism, and occlusion of arteriovenous cannulae | A purified preparation of a bacterial protein from group C (β)-hemolytic *Streptococcus*. |
| Tenecteplase | TNKase | Myocardial infarction and intracoronary emboli | Tenecteplase is a 527 aa glycoprotein developed by introducing the following modifications to the cDNA for natural human tPA: $Thr_{103}$ to $Asp_{103}$, $Asp_{117}$ to $Gln_{117}$ (both within the kringle 1 domain), and the tetra-Ala substitution at aa 296–299 in the protease domain. |
| Urokinase | Abbokinase | Pulmonary embolism, coronary artery thrombosis, and *IV* catheter clearance | Low-molecular-weight form of human urokinase that consists of an A chain of 2 kDa linked by a SH bond to a B chain of 30.4 KDa. Recombinant urokinase plasminogen activator. |

### 5.1 Examples of Enzyme Biopharmaceuticals

Dornase alfa (DNase I; EC 3.1.21.1) was introduced over a decade ago as an optional drug to treat cystic fibrosis (CF). Cystic fibrosis is an autosomal, recessive, hereditary disease that affects the lungs, sweat glands, and the digestive system, causing chronic respiratory and digestive problems. It is caused by mutations in the cystic fibrosis transmembrane conductance regulator (CFTR) protein and is the most common fatal autosomal recessive disease in Caucasians. Approximately 1 in 25 individuals of European descent is a carrier of a CF mutation. Dornase breaks down the built-up sputum in the lungs of affected individuals through the digestion of the accumulated extracellular DNA from the breakdown of neutrophils that collect due to the local inflammation of the lungs [70]. As the drug has a short duration of action, it is administered continuously through daily inhalation by a nebulizer or a compressor system. This treatment benefits about 40% of patients. The drug is usually well tolerated, although some side effects such as chest pain and skin rashes, have been reported [71]. The three-dimensional structure of DNase I in the presence of a short DNA segment is illustrated in Figure 3.

Galsulfase (Aryplase) or N-acetylgalactosamine 4-sulfatase (ARSB; EC 3.1.6.12) is a hydrolase drug used in the treatment of mucopolysaccharidosis (MPS) type IV (or Maroteaux–Lamy syndrome), a lysosomal storage disease. Patients suffering from this lethal syndrome are deficient in Galsulfase expression leading to a block of the degradation of glycosaminoglycans, which causes an accumulation of these compounds in various tissues [69]. Galsulfase treatment is well tolerated, although some patients develop antibodies against the enzyme. However, these enzyme-specific antibodies generally decrease significantly in these patients after about a year and a half of treatment [72]. Another drug has been available for the treatment of MPS type I since 2003. This disease is caused by a deficiency in α-L-iduronidase (laronidase; Aldurazyme; EC 3.2.1.76) [73]. Treatment with laronidase for the milder forms of MPS type I has been shown to be safe, well tolerated, and effective. No significant side effects are associated with the administration of this enzyme-drug, although most patients develop antibodies against the enzyme; however, with time the antibody titers subside [74]. Injection of laronidase must be done on a weekly basis [73].

Agalsidase (Fabrazyme) is used in the treatment of Fabry disease, a lethal inborn error of metabolism disorder in which patients lack lysosomal hydrolase α-galactosidase A (αGalA; EC 3.2.1.22). This enzyme is involved in the breakdown of glycosphingolipids leading to the accumulation of α-galactose in tissues and plasma [75]. This accumulation can lead to skin lesions, febrile episodes, and burning sensations in the extremities. Death in early adulthood is usually due to renal failure because of proteinuria induced hypertension. The bimonthly injection of agalsidase-α (Replagal) or agalsidase-β (Fabrazyme) alleviates the severe pain associated with this condition. The cost of these drugs can exceed $170,000 per year.

Pegademase (adenosine deaminase; EC 3.5.4.4) is a PEGylated enzyme used for treating patients with adenosine deaminase (ADA) deficiency [76] and suffering from a form of severe combined immunodeficiency syndrome (SCID), a disease that occurs in about 1/50,000 births. SCID patients have a severely crippled immune system and cannot clear or recover from the mildest microbial or viral infections.

SCID patients, if untreated, usually die within 1 year due to severe, recurrent infections such as chronic diarrhea, ear infections, recurrent pneumonia, and profuse oral candidiasis. The ADA deficiency leads to an accumulation of purine metabolites, in particular dGTP, which are cytotoxic to lymphoid stem cells. Restoration of normal ADA levels removes the dGTP and allows the lymphoid stem cells to propagate. Pegademase is administered intravenously, stored at 4°C, and should never be frozen [77].

Imiglucerase (alglucerase) is used in the treatment of type 1 Gaucher disease (GD), a lysosomal storage disorder due to a deficiency of glucocerebrosidase (EC 3.2.1.45). This enzyme deficiency leads to the accumulation of glucosylceramide in cells [78]. Imiglucerase has been used to treat Gaucher disease for the past decade and is given intravenously. Its application leads to a marked improvement in the clinical manifestation of the disease [79]. However, this drug cannot reverse all effects of Gaucher disease, such as fibrotic and necrotic tissue damage [80]. The drug is well tolerated and has few side effects. The most common side effect is antibody generation to imiglucerase [81]. However, the excellent safety profile of imiglucerase enables patient self-administration and prevents the morbidity associated with the disease.

Although most enzyme therapies are used to treat relatively rare metabolic or genetic disorders, some enzymes are used to treat much more common afflictions such as heart attacks and stroke. Tissue plasminogen activators (tPA; EC 3.4.21.68) are intravenously administered fibrinolytic agents that are used in reperfusion for the treatment of acute ischemic stroke [82]. It has been shown that the use of tPAs such as reteplase (Retavase), tenecteplase (TNKase), alteplase (Activase), and anistreplase (Eminase) within 3 hours post-stroke-onset affords at least an order of magnitude more benefit to patients than aspirin, the only other beneficial drug for this condition. About one third of patients treated with intravenous fibrinolytic therapy exhibit an improvement poststroke [83]. Furthermore, tPAs are also used in the treatment of acute myocardial infarction [84]. Tenecteplase is known to have a long half-life; therefore, a single bolus injection is sufficient for treatment [85]. The half-life of reteplase is even longer allowing its administration as a double-bolus injection. However, this increased half-life seems to be only useful in reducing the number of injections, as it provides no additional health benefits to the patient [86]. Streptokinase (Streptase; EC 3.4.99.0) is another thrombolytic agent that, in some instances, has been shown to be more effective than tPAs [87]. Lastly, urokinase (Abbokinase; EC 3.4.21.73) is another type of thrombolytic agent that activates plasminogen through another mechanism and is used in the treatment of pulmonary embolisms.

Several enzyme therapies are also used in the treatment of cancer. For instance, rasburicase (uricase; Elitek; EC 1.7.3.3) is used in the treatment of hyperuricemia due to tumor lysis syndrome (TLS). This is a potentially life-threatening condition that is associated with rapidly developing tumors, such as those found in lymphoma and leukemia, in patients undergoing chemotherapy. The incidence of hyperuricemia in these patients is close to 20% [88]. Uricase is an interesting enzyme, because it is found in most mammals, with the exception of humans, where the gene contains a nonsense mutation [89]. Rasburicase is well tolerated, has a very fast onset of action, and is administered intravenously once a day. However, the enzyme cannot be used on patients with glucose-6-phosphate dehydrogenase deficiency [90]. The

PEGylated formulation of uricase is less immunogenic and has a longer half-life than its "naked" form [89].

Asparaginase (EC 3.5.1.1) is an example of another enzyme-based anticancer therapy. Asparaginase, derived from *E. coli*, is frequently used in the treatment of acute lymphoblastic leukemia [91]. Asparaginase works by depriving leukemic cells of L-asparagine. Asparaginase hydrolyzes the amino acid L-asparagine to L-aspartic acid and ammonia. Most human tissues do not require L-asparagine because it can be synthesized from L-glutamine by the action of asparagine synthase (AS). However, acute lymphoblastic leukemia cells generally have very low levels of AS and cannot upregulate the AS gene during conditions of L-asparagine depletion. Asparaginase is an effective way of depleting L-asparagine from the body, and this depletion results in selective cell death of those cells that cannot make L-asparagine. Asparaginase has a relatively short half-life (10–20 hours). However, PEGylated asparaginase has a half-life that is 10–15 times longer. PEGylated-asparaginase (Pegaspargase) was developed in the 1970s and 1980s. The drug has undergone extensive testing and seems to retain its antileukemic effectiveness while allowing less-frequent administration than the native compound. Although the actual cost to patients for Pegaspargase is greater than that of multiple injections of other Asparaginases, the reduced need for physician visits and treatment of complications of therapy may make overall treatment costs considerably less than that of the conventional Asparaginases.

Also some drug formulations contain more than one enzyme. For example, pancrelipase or Zymase contains amylase (for processing starch), lipase (for breaking down fats), and chymotrypsin (a protease) in various combinations. This protein mixture is isolated from porcine or bovine pancreatic tissues and is sometimes called pancreatin. This enzyme cocktail is sold to help digestion in people who do not produce enough of their own digestive enzymes. Such a condition can arise due to pancreatic disfunction caused by pancreatitis, cystic fibrosis, or pancreatic cancer.

### 5.2 The Future of Enzyme Biopharmaceuticals

The field of enzyme therapy is not static. New enzymes are being developed and tested to treat a variety of diseases or conditions all the time. Some of the most active work has been in the area of lysosomal storage diseases. Indeed there are now several enzyme pharmaceuticals that are entering phase II and phase III trials to treat several fatal or debilitating syndromes associated with sugar metabolism or sugar catabolism. One example is acid maltase, which is associated with a rare, fatal condition called Pompe's disease. Pompe's disease, also known as infantile glycogen storage disease type 2 (GSD-II), or acid maltase deficiency, results from a genetic deficiency of the acid maltase enzyme, a protein that normally breaks down glycogen (sugar stored in cells). In the absence of acid maltase, glycogen accumulates to toxic levels in cardiac and skeletal muscles, causing the muscles to waste away. Early experiments with acid maltase on several newborns with Pompe's disease indicated good success, particularly with those with residual acid maltase activity [92]. Several other lysosomal storage diseases are also being investigated with enzyme replacement therapy, including Hurler's disease and Maroteaux–Lamy syndrome. Hurler's disease is characterized by a deficiency of α-iduronidase activ-

ity that causes accumulation of dermatan sulfate and heparin sulfate in patients. This leads to stiff joints, reduced range of motion, restrictive pulmonary disease, sleep apnea, and other problems with the eyes, liver, joints, and skeleton. Recent clinical trials with α-iduronidase ERT have shown significant improvement in some patients, especially with their range of motion, flexibility, and reduced sleep apnea (reviewed in Ref. 93). Maroteaux–Lamy syndrome is caused by the accumulation of excessive quantities of dermatan sulfate because of a deficiency of the enzyme N-acetylgalactosamine-4-sulphatase (arylsulfatase B). An ERT trial has recently been conducted with recombinant arylsulfatase B given once every week. In this trial, urinary glycosaminoglycans were reduced, walking ability improved, the range of shoulder motion increased, and joint pain lessened [72]. However, several adverse reactions were also noted and all patients developed antibodies to arylsulfatase.

Not all enzyme therapies are likely to be successful, at least not until better methods for formulating and delivering enzymes are developed. One central challenge to almost all ERTs is the delivery of the proteins to the right tissues. For example, among the lysosomal storage diseases, new and improved formulation and delivery technologies need to be developed to deliver therapeutic enzymes effectively to cardiac muscle and kidney in Fabry's disease, skeletal muscle in patients with Pompe's disease, and to joint tissues and structures in patients with Hurler's disease and Maroteaux–Lamy syndrome. Improved tissue targeting would no doubt help many other enzyme therapeutics used to treat cancer, heart attacks, and strokes.

Another challenge with ERT is the enormous cost and the sometimes questionable benefits or small improvements to patient quality of life. For some IEMs, the rarity of the disease means that production costs will always remain high. For Gaucher's disease, in which ERT has been the most effective, the annual cost of ERT ranges from $40,000 to 320,000 per patient. The cost of ERT for Fabry's disease is about $160,000 per year. Even for more common disorders such as cancer, heart attacks, and stroke, the costs of enzyme-based drugs are sometimes observed as too high. In fact, similar cost–benefit debates have swirled around the issue of tPA versus aspirin and heparin treatments for stroke for more than 15 years [94]. In some cases, the enzyme drug is determined to be very cost-effective. In other cases it is not. Certainly as the market size grows for some enzyme therapies, the costs will inevitably drop. Likewise as formulations are improved and delivery mechanisms are enhanced, it is likely that the cost–benefit balance will change markedly.

## 6 CONCLUSIONS

Pharmaceuticals based on natural and recombinant proteins have proven their therapeutic value. Whether for hormone or enzyme replacement therapies or to block or inhibit other protein signals, pharmaceutical proteins exhibit exquisite selectivity and specificity in their targets. Over 100 proteins have been approved by the U.S. Food and Drug Administration since their inception and many more are currently in various phases of development. Despite their potency, many issues hinder the widespread use of protein pharmaceuticals. Proteins require special

(expensive) preparation techniques, usually involving recombinant engineering and expression in cell culture followed by lengthy and costly purifications. They must be stored and shipped under special conditions as they are susceptible to denaturation and degradation by proteolysis. The barriers to easy oral administration and bioavailability are formidable, as the body has many natural defenses to exogenous proteins. Nevertheless, although there are many hurdles to be overcome, protein biopharmaceuticals are certain to see continued rapid development and growth because their potential for providing a more specific and potent therapeutic approach is unparalleled.

## REFERENCES


1. Lindeboom G A (1954). The story of a blood transfusion to a Pope. *J. Hist. Med. Allied Sci.* 9(4):455–459.
2. Baskett T F (2002). James Blundell: The first transfusion of human blood. *Resuscita.* 52(3):229–233.
3. Farr A D (1981). Treatment of haemophilia by transfusion: the first recorded case. *J. Royal Soc. Med.* 74(4):301–305.
4. Jenner E (1801). *An inquiry into the causes and effects of the Variolae Vaccinae, A Disease Discovered in Some of the Western Counties of England, Particularly Gloucestershire, and Known by the Name of the Cow-pox*, 3rd ed.: D.N. Shury, London.
5. Dunsch L (1986). *Jèons Jacob Berzelius,* 1. Aufl. ed, Biographien hervorragender Naturwissenschaftler, Techniker und Mediziner., Bd. 85. BSB B.G. Teubner Leipzig.
6. Raben M S (1957). Preparation of growth hormone from pituitaries of man and monkey. *Science.* 125(3253):883–884.
7. du Toit D F, Heydenrych J J (1986). The application, mechanism of action and side-effects of immunosuppressive agents in clinical transplantation. *S. Afr. Med. J.* 70(11):687–691.
8. Federlin K, Laube H, Velcovsky H G (1981). Biologic and immunologic *in vivo* and *in vitro* studies with biosynthetic human insulin. *Diabet. Care.* 4(2):170–174.
9. Don B R, Kaysen G (2004). Serum albumin: relationship to inflammation and nutrition. *Semin. Dial.* 17(6):432–437.
10. Wishart D S, et al (2006). DrugBank: a comprehensive resource for *in silico* drug discovery and exploration. *Nucleic Acids Res.* 34(Database iss.):D668–D672.
11. Henry H L, Norman A W (2003). *Encyclopedia of Hormones.* Academic Press, Amsterdam, the Netherlands.
12. Bugg T (2004). *Introduction to Enzyme and Coenzyme Chemistry.* Blackwell Publishers, Malden, MA.
13. Lo B K C (2004). *Antibody engineering: Methods and Protocols*, *Methods in molecular biology*, vol. 248. Humana Press, Clifton, N.J.
14. Hershfield M S, et al (1987). Treatment of adenosine deaminase deficiency with polyethylene glycol-modified adenosine deaminase. *New Eng. J. Med.* 316(10):589–596.
15. Howey D C, et al (1994). [Lys(B28), Pro(B29)]-human insulin. A rapidly absorbed analogue of human insulin. *Diabetes.* 43(3):396–402.
16. Venter J C, et al (2001). The sequence of the human genome. *Science.* 291(5507): 1304–1351.
17. Lander E S, et al (2001). Initial sequencing and analysis of the human genome. *Nature.* 409(6822):860–921.

18. Christopher S (2004). Has the yo-yo stopped? An assessment of human protein-coding gene number. *PROTEOMICS*. 4(6):1712–1726.
19. Svensson B, et al (2004). Proteinaceous α-amylase inhibitors. *Biochim. Biophy. Acta*. 1696(2):145–156.
20. Rosenfeld L (2002). Insulin: discovery and controversy. *Clin. Chem*. 48(12):2270–2288.
21. Walsh G (2005). Therapeutic insulins and their large-scale manufacture. *Appl. Microbiol. Biotechnol*. 67(2):151–159.
22. Meienhofer J, Sano Y (1968). A solid-phase synthesis of [lysine]-vasopressin through a crystalline protected nonapeptide intermediate. *J. Amer. Chem. Soc*. 90(11):2996–2997.
23. Bayer E, Hagenmaier H (1968). Solid phase synthesis of oxytocin. *Tetrahed. Lett.*, 172037–172039.
24. Martin J, et al (2005). *Births: Final Data for 2003. National Vital Statistics Reports*, vol. 54. National Center for Health Statistics. Hyattsville, MD.
25. Stephen E H, Chandra A (2000). Use of infertility services in the United States: 1995. *Family Plan. Perspect*. 32(3):132–137.
26. Vanderkuur J A, et al (1997). Signaling molecules involved in coupling growth hormone receptor to mitogen-activated protein kinase activation. *Endocrinol*. 138(10): 4301–4307.
27. Frank S J (2002). Receptor dimerization in GH and erythropoietin action—it takes two to tango, but how? *Endocrinol*. 143(1):2–10.
28. Schneider H J, Pagotto U, Stalla G K (2003). Central effects of the somatotropic system. *Euro. J. Endocrinol*. 149(5):377–392.
29. Hull K L, Harvey S (1999). Growth hormone resistance: Clinical states and animal models. *J. Endocrinol*. 163(2):165–172.
30. Kopchick J J (2003). Discovery and mechanism of action of pegvisomant. *Euro. J. Endocrinol*. 148(Suppl 2):S21–S25.
31. Lin F-K, et al (1985). *Cloning and Expression of the Human Erythropoietin Gene*, Proc. National Academies of Sciences, vol. 82. pp. 7580–7584.
32. Jacobs K, et al (1985). Isolation and characterization of genomic and cDNA clones of human erythropoietin. *Nature*. 313(6005):806–810.
33. Binley K, et al (2002). Long-term reversal of chronic anemia using a hypoxia-regulated erythropoietin gene therapy. *Blood*. 100(7):2406–2413.
34. Pestka S, Krause C D, Walter M R (2004). Interferons, interferon-like cytokines, and their receptors. *Immunol. Rev*. 202(1):8–32.
35. Isaacs A, Lindenmann J (1957). Virus interference. I. The interferon. *Proc. of the Royal Society of London, Series B*. 147(927):258–267.
36. Stark G R, et al (1998). How cells respond to interferons. *Ann. Rev. Biochem*. 67227–67264.
37. Thomson CenterWatch, *CenterWatch Clinical Trials Listing Service*, 2006.
38. Wikipedia contributors. *Interferon*. Wikipedia, The Free Encyclopedia 2006, Available: http://en.wikipedia.org/w/index.php?title=Interferon&oldid=40341679.
39. Takeuchi H, Yamamoto H, Kawashima Y (2001). Mucoadhesive nanoparticulate systems for peptide drug delivery. *Adv. Drug Del. Rev*. 47(1):39–54.
40. Salamat-Miller N, Johnston T P (2005). Current strategies used to enhance the paracellular transport of therapeutic polypeptides across the intestinal epithelium. *Internat. J. Pharm.*, 294(1–2):201–216.

41. Lundin S, et al (1989). Degradation of [mercaptopropionic acid1, D-arginine8]-vasopressin (dDAVP) in pancreatic juice and intestinal mucosa homogenate. *Pharmacol. Toxicol.* 65(2):92–95.
42. Clement S, et al (2004). Oral modified insulin (HIM2) in patients with type 1 diabetes mellitus: results from a phase I/II clinical trial. *Metabolism*. 53(1):54–58.
43. Prego C, et al (2005). Transmucosal macromolecular drug delivery. *J. Control. Rel.* 101(1–3):151–162.
44. Varshosaz J, Sadrai H, Heidari A (2006). Nasal delivery of insulin using bioadhesive chitosan gels. *Drug Del.* 13(1):31–38.
45. Nagaya N, Kojima M, Kangawa K (2006). Ghrelin, a novel growth hormone-releasing peptide, in the treatment of cardiopulmonary-associated cachexia. *Inter. Med.* 45(3): 127–134.
46. Liu Y L, et al (2006). Ghrelin alleviates cancer chemotherapy-associated dyspepsia in rodents. *Cancer Chemother. Pharmacol.* 58(3):326–333.
47. Greenberg A S, Obin M S (2006). Obesity and the role of adipose tissue in inflammation and metabolism. *Amer. J. Clin. Nutrit.* 83(2):461S–465S.
48. Kim S J, Park Y, Hong H J (2005). Antibody engineering for the development of therapeutic antibodies. *Molec. Cells.* 20(1):17–29.
49. Morrison S L, et al (1984). Chimeric human antibody molecules: mouse antigen-binding domains with human constant region domains. *Proc. Natl. Acad. Sci. USA*. 81(21): 6851–6855.
50. Jones P T, et al (1986). Replacing the complementarity-determining regions in a human antibody with those from a mouse. *Nature*. 321(6069):522–525.
51. Zebedee S L, et al (1992). Human combinatorial antibody libraries to hepatitis B surface antigen. *Proc. Nat. Acad. Sci. USA*. 89(8):3175–3179.
52. Green L L, et al (1994). Antigen-specific human monoclonal antibodies from mice engineered with human Ig heavy and light chain YACs. *Nature Genetics*. 7(1):13–21.
53. Fishwild D M, et al (1996). High-avidity human IgG kappa monoclonal antibodies from a novel strain of minilocus transgenic mice. *Nature Biotechnol.* 14(7):845–851.
54. Nizet Y, et al (2000). The experimental (in vitro) and clinical (in vivo) immunosuppressive effects of a rat IgG2b anti-human CD2 mAb, LO-CD2a/BTI-322. *Transplantat.* 69(7):1420–1428.
55. Popma S H, Griswold D E, Li L (2005). Anti-CD3 antibodies OKT3 and hOKT3gamma1(Ala-Ala) induce proliferation of T cells but impair expansion of alloreactive T cells; aspecifc T cell proliferation induced by anti-CD3 antibodies correlates with impaired expansion of alloreactive T cells. *Internat. Immunopharmacol.* 5(1): 155–162.
56. Sgro C (1995). Side-effects of a monoclonal antibody, muromonab CD3/orthoclone OKT3: Bibliographic review. *Toxicol.* 105(1):23–29.
57. Coller B S (1997). Platelet GPIIb/IIIa antagonists: the first anti-integrin receptor therapeutics. *J. Clin. Investigat.* 100(suppl 11):S57–S60.
58. Berkowitz S D, et al (1997). Acute profound thrombocytopenia after C7E3 Fab (abciximab) therapy. *Circulation*. 95(4):809–813.
59. Morgan E L, Hugli T E, Weigle W O (1982). Isolation and identification of a biologically active peptide derived from the CH3 domain of human IgG1. *Proc. Nat. Acad. Sci. USA*. 79(17):5388–5391.
60. Goebel J, et al (2000). Daclizumab (Zenapax) inhibits early interleukin-2 receptor signal transduction events. *Transpl. Immunol.* 8(3):153–159.

61. Weiner L M (2006). Fully human therapeutic monoclonal antibodies. *J. Immunother.* 29(1):1–9.
62. Lonberg N (2005). Human antibodies from transgenic animals. *Nature Biotechnol.* 23(9):1117–1125.
63. Bayes M, Rabasseda X, Prous J R (2005). Gateways to clinical trials. *Meth. Findings Experim. Clin. Pharmacol.* 27(5):331–372.
64. Bayes M, Rabasseda X, Prous J R (2005). Gateways to clinical trials. *Meth. Findings Experim. Clin. Pharmacol.* 27(8):569–612.
65. Bayes M, Rabasseda X, Prous J R (2005). Gateways to clinical trials. *Meth. Findings Experim. Clin. Pharmacol.* 27(10):711–738.
66. Stocks M (2005). Intrabodies as drug discovery tools and therapeutics. *Curr. Opin. Chem. Biol.* 9(4):359–365.
67. Tanaka T, Rabbitts T H (2003). Intrabodies based on intracellular capture frameworks that bind the RAS protein with high affinity and impair oncogenic transformation. *Embo J.* 22(5):1025–1035.
68. Popkov M, et al (2005). Targeting tumor angiogenesis with adenovirus-delivered anti-Tie-2 intrabody. *Cancer Res.* 65(3):972–981.
69. Litjens T, Hopwood J J (2001). Mucopolysaccharidosis type VI: Structural and clinical implications of mutations in N-acetylgalactosamine-4-sulfatase. *Human Mutat.* 18(4): 282–295.
70. Shah P L, et al (1996). *In vivo* effects of recombinant human DNase I on sputum in patients with cystic fibrosis. *Thorax*. 51(2):119–125.
71. Davies J, et al (1997). Retrospective review of the effects of rhDNase in children with cystic fibrosis. *Pediat. Pulmonol.* 23(4):243–248.
72. Harmatz P, et al (2005). Direct comparison of measures of endurance, mobility, and joint function during enzyme-replacement therapy of mucopolysaccharidosis VI (Maroteaux-Lamy syndrome): Results after 48 weeks in a phase 2 open-label clinical study of recombinant human N-acetylgalactosamine 4-sulfatase. *Pediatr.* 115(6):e681–e689.
73. Miebach E (2005). Enzyme replacement therapy in mucopolysaccharidosis type I. *Acta Paediatr. Supplem.* 94(447):58–60.
74. Kakavanos R, et al (2003). Immune tolerance after long-term enzyme-replacement therapy among patients who have mucopolysaccharidosis I. *Lancet.* 361(9369): 1608–1613.
75. Desnick R J, et al (2003). Fabry disease, an under-recognized multisystemic disorder: Expert recommendations for diagnosis, management, and enzyme replacement therapy. *Ann. Intern. Med.* 138(4):338–346.
76. Blackburn M R, Kellems R E (2005). Adenosine deaminase deficiency: Metabolic basis of immune deficiency and pulmonary inflammation. *Adv. Immunol.* 86:1–41.
77. Hershfield M S (1995). PEG-ADA: An alternative to haploidentical bone marrow transplantation and an adjunct to gene therapy for adenosine deaminase deficiency. *Human Mutat.* 5(2):107–112.
78. Beutler E (1995). Gaucher disease. *Adv. Genet.* 3217–3249.
79. Pastores G M, et al (2004). Therapeutic goals in the treatment of Gaucher disease. *Semin. Hematol.* 41(4 Suppl 5):4–14.
80. Weinreb N J, et al (2005). Guidance on the use of miglustat for treating patients with type 1 Gaucher disease. *Amer. J. Hematol.* 80(3):223–229.
81. Rosenberg M, et al (1999). Immunosurveillance of alglucerase enzyme therapy for Gaucher patients: Induction of humoral tolerance in seroconverted patients after repeat administration. *Blood*. 93(6):2081–2088.

82. Hacke W, et al (2004). Association of outcome with early stroke treatment: pooled analysis of ATLANTIS, ECASS, and NINDS rt-PA stroke trials. *Lancet*. 363(9411): 768–774.
83. Saver J L (2004). Number needed to treat estimates incorporating effects over the entire range of clinical outcomes: Novel derivation method and application to thrombolytic therapy for acute stroke. *Archives of Neurol*. 61(7):1066–1070.
84. Menon V, et al (2004). Thrombolysis and adjunctive therapy in acute myocardial infarction: the Seventh ACCP Conference on Antithrombotic and Thrombolytic Therapy. *Chest*. 126(suppl 3):549S–575S.
85. Davydov L, Cheng J W (2001). Tenecteplase: A review. *Clin. Therapeu*. 23(7):982–997.
86. Noble S, McTavish D (1995). Levocabastine. An update of its pharmacology, clinical efficacy and tolerability in the topical treatment of allergic rhinitis and conjunctivitis. *Drugs*. 50(6):1032–1049.
87. Capstick T, Henry M T (2005). Efficacy of thrombolytic agents in the treatment of pulmonary embolism. *Euro. Respirat. J*. 26(5):864–874.
88. Annemans L, et al (2003). Incidence, medical resource utilisation and costs of hyperuricemia and tumour lysis syndrome in patients with acute leukaemia and non-Hodgkin's lymphoma in four European countries. *Leukemia and Lymphoma*. 44(1):77–83.
89. Bomalaski J S, et al (2002). Uricase formulated with polyethylene glycol (uricase-PEG 20): biochemical rationale and preclinical studies. *J. Rheumatol*. 29(9):1942–1949.
90. Pui C H (2002). Rasburicase: a potent uricolytic agent. *Expert Opin. Pharmacother*. 3(4):433–442.
91. Pui C H, Evans W E (2006). Treatment of acute lymphoblastic leukemia. *New Engl. J. Med*. 354(2):166–178.
92. Amato A A (2000). Acid maltase deficiency and related myopathies. *Neurol. Clin*. 18(1):151–165.
93. Wraith J E (2005). The first 5 years of clinical experience with laronidase enzyme replacement therapy for mucopolysaccharidosis I. *Expert Opin. Pharmacother*. 6(3):489–506.
94. Schneck M J, Biller J (2005). New treatments in acute ischemic stroke. *Curr. Treat. Opt. Neurol*. 7(6):499–511.

# FORMULATION AND DELIVERY ISSUES OF THERAPEUTIC PROTEINS

Sathy V. Balu-Iyer,[1,2] Razvan D. Miclea,[2] and Vivek S. Purohit[3]

[1] *University at Buffalo, State University of New York, Amherst, New York*
[2] *Roswell Park Cancer Institute, Buffalo, New York*
[3] *R&D, Eurand, Inc., Vandalia, Ohio*

**Chapter Contents**

1 Introduction
- 1.1 Recombinant DNA Technology
- 1.2 Specificity and Potency
- 1.3 Molecular Medicine
- 1.4 Therapeutic, Subunit Vaccines

2 Stability of Therapeutic Proteins
- 2.1 Chemical Instability
- 2.2 Physical Instability
- 2.3 Formulation and Delivery Approaches to Overcome Instability

3 Immunological Properties of Protein Therapeutics
- 3.1 Immnogenicity of Proteins
- 3.2 Formulation and Delivery Considerations to Reduce Immunogenicity

4 Pharmacokinetics and Pharmacodynamics
- 4.1 Formulation and Delivery Approaches to Overcome PK Issues

References

## 1 INTRODUCTION

The protein pharmaceuticals such as human growth hormone and insulin have been known for years. However, in the past decade, an increased number of protein

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

products have come on the market and into the development pipeline as a result of the following reasons.

### 1.1 Recombinant DNA Technology

Advances in recombinant technology alleviated the problem of protein production and its purity. Through the process of recombinant DNA technology and use of nonhuman cell lines, human proteins can be manufactured free of viral contamination. This process enables production of large quantities of proteins previously difficult to obtain from human sources. Further isolation of protein from human sources is associated with a high risk of viral contamination. One such example, plasma derived clotting factor VIII isolated from human blood, resulted in transmission of viral diseases such as hepatitis and AIDS [1, 2].

### 1.2 Specificity and Potency

Protein products are specific as they are identified based on their mechanism of action. Further, protein products are attractive because of their potency, as only μg quantities or a few thousand IU are required to exert the pharmacological effect.

### 1.3 Molecular Medicine

The term molecular medicine is defined as understanding disease at the molecular level. For example, the deficiency of factor VIII causes a bleeding disorder, Hemophilia A [3]. Given the accomplishments made in the area of human genome and the emergence of proteomics, functional proteins that have important therapeutic value have been developed, and more will be identified in the future.

### 1.4 Therapeutic, Subunit Vaccines

Protein and peptide antigen-based vaccines are safe. The search for safe, single-shot vaccines against infectious and other diseases is a major focus of current research. Several peptide-and protein-based antigens are at different stages of development.

However, the protein therapeutics have unique handling and development requirements relative to the small organic molecule-based drugs and pose exclusive challenges for the formulation scientist. The formulation and delivery issues of therapeutic proteins can contribute to difficulties in pharmaceutical development such as

- Stability (in vivo and in vitro)
- Immune toxicity
- Pharmacokinetics / pharmacodynamics-related problems

Formulation strategies, such as addition of specific functional excipients, freeze-drying, protein engineering, and delivery strategies, have been used to improve the stability and circulation half-life and reduce immunogenicity, which in turn improve

therapeutic efficacy of protein drugs. This chapter will focus on formulation and delivery strategies of therapeutic proteins.

## 2 STABILITY OF THERAPEUTIC PROTEINS

As a result of the complex structure of the proteins, formulation of protein therapeutics pose unique difficulties as it is susceptible to physical and chemical instabilities. The complexity develops from the hierarchical nature of its structure: primary, secondary, tertiary, and quaternary structures. Primary structure is the amino acid sequence of the polypeptide chains; secondary structure refers to local-ordered conformation; tertiary structure deals with the spatial arrangement of secondary structural elements (often referred as global fold); and the quaternary structure is the spatial arrangement of subunits. In general, chemical instability is related to primary structure of the protein, whereas physical instability is associated with the global fold or 3D structure of the molecule. The common problems encountered for protein products are listed in Table 1.

### 2.1 Chemical Instability

Chemical instability refers to covalent modifications of the molecule that result in a new chemical entity. Some amino acids in the protein are chemically labile and undergo degradation to form a new chemical entity. The chemical reactions that affect the proteins are as follows:

***Deamidation.*** The side chain of amide linkage in a Gln or Asn residue is hydrolysed to form a free carboxylic acid that gives mixtures of peptides in which the polypeptide backbone is attached via an alpha-carboxyl linkage (Asp/Glu) or is attached via beta carboxyl linkage (iso-Asp/Glu) [4]. The key step in the reaction is the formation of a five- or six-membered cyclic imide structure [5], and this degradative reaction most often occurs at a higher rate at the sequence Asn-Gly.

***Oxidation.*** Oxidation is an important chemical degradation pathway for several proteins and can occur during all steps of its pharmaceutical development. The amino acids that are labile to oxidation are Met, His, Cys, Trp, Phe, Tyr, and Pro. Met has been identified as one of the most easily oxidized amino acids in proteins

**TABLE 1.**

| Chemical Instability | Physical Instability |
|---|---|
| Hydrolysis | Denaturation |
| Deamidation (Gln and Asn) | Adsorption to surfaces |
| Oxidation (His, Met, Cys, Trp, and Tyr) | Aggregation |
| Proteolysis | Precipitation |
| Incorrect disulfide formation (Cysteine) | |
| Racemization | |
| Beta elimination | |

[6, 7]. Free radicals generated by metal ion catalysis and photo or thermal degradation of buffer or other excipients can accelerate the oxidative degradation reactions.

***Beta Elimination.*** The beta elimination can occur in amino acids that have beta side chains. The amino acids that are labile for beta elimination are Cys, Ser, Thr, Phe, and Lys [6]. The rate of this degradative reaction is influenced by pH, temperature, and presence of metal ions. The beta elimination of Cys leads to incorrect disulfide formation, as this reaction produces thiols [6].

***Incorrect Disulfide Formation.*** This reaction occurs in neutral and alkaline pH conditions. Under such conditions, deamidation also occurs. The thiols (formed by beta elimination) can promote the disufide bond formation [8]. In an acidic environment, sulfenium ion promotes the incorrect disulfide bonds [9]. Proteins with scrambled disulfide bridges have been shown to yield native structure by incubating the protein with small amounts of mercaptoethanol or Cys.

***Racemization.*** All amino acids except Gly are chiral and are subjected to base catalyzed racemization. This reaction leads to the formation of nonmetabolizable D-enantiomers and is often associated with loss of activity.

The chemical instability studies are often performed in simulated pharmaceutical stress conditions such as pH shifts or thermal stress conditions in the presence and in the absence of metal ions or buffer salts. The degradative products are analyzed by sequencing techniques such as mass spectrometry, HPLC techniques, and by employing specific enzymatic assays.

## 2.2 Physical Instability

Physical stability refers to changes in the higher-order structure (secondary and above) and is manifested as aggregation, precipitation, and surface adsorption [10]. Proteins, because of their complex structure and global fold, undergo physical instability [11]. In the global fold, the exposure of the hydrophobic groups is minimized, and this conformation is required for optimal biological activity of the protein. Denaturation is a molecular process that disrupts the global fold of the molecule.

***Denaturation.*** Denaturation refers to an alteration in the global fold of the molecule (i.e, a disruption in tertiary and frequently secondary). The global fold of the molecule can be altered by environmental conditions that are encountered in pharmaceutical development of proteins, including changes in temperature, pH, pressure, and concentration of denaturing agents such as surfactants. This molecular process is generally associated with inactivation of the protein. Denaturation of the protein may also be followed by additional phenomena of physical degradation such as surface adsorption, aggregation, or precipitation.

***Aggregation.*** The existence of I (folding intermediate) states in several proteins leads to the inactivation of the protein by aggregation. Moderate amounts of stress can generate I states, which retains secondary structure but tertiary structural

features such as intramolecular contacts are lost [12]. These intermediates associate to form large aggregates.

***Precipitation.*** Precipitation is the macroscopic manifestation of aggregation. It has been observed that the formation of finely divided precipitate of insulin occurs on the walls of the container [13]. The precipitation of proteins is also observed during the manufacturing stages. Upon expression in recombinant systems, many proteins form aggregates called inclusion bodies [14].

***Surface Adsorption.*** Adhesion of proteins to surfaces is a well-known phenomenon. The denatured or I states of the protein expose hydrophobic domains. This exposure promotes the binding of the protein to the walls of the container and other surfaces it comes in contact with. It has been well documented that insulin binds to surfaces of delivery pumps, to glass and plastic containers, and to the inside of the intravenous bags [15].

### 2.3 Formulation and Delivery Approaches to Overcome Instability

The risk of chemical instability can be assessed from the primary sequence of the protein. The sequence containing labile amino acids such as Asn-Gly and Met would be indicative of potential instability issues. The rate of chemical reactions that alter the primary sequence of the protein is higher in solution conditions and can limit the shelf-life of protein therapeutics. As the mobility of reactants is minimized in the solid state, freeze-drying is often attempted to improve the stability [16]. In such instances, physical instability is a major issue to be dealt with. Freeze-drying, also termed as lyophilization, is a dessication process in which the solvent (usually water) is first frozen and then is removed by sublimation in a vacuum [17]. In other words, the protein in solution is frozen, producing discrete ice and solute crystals. The solid ice is sublimed. Controlled heating desorbs any of the tightly bound water.

Three steps constitute the freeze-drying process [17].

- Freezing
- Primary drying (sublimation) to remove bulk, free water
- Secondary drying (desorption) to remove structured, more tightly bound water

To prevent denaturation/unfolding of the protein during the harsh freeze-drying process, excipients are added, including bulking agents, suitable buffers for pH and osmotic adjustment, cryoprotectants, protein structure stabilizers, and phase-state modifiers (glass transition temperature modifiers) [17, 18]. As a result of the potency of therapeutic proteins, a low concentration of the protein is commonly used, which may lead to loss of therapeutic ingredients during freeze-drying [17]. To prevent protein loss, bulking agents such as mannitol, glycine, lactose, and sucrose are added to increase the bulk mass. These bulking agents can also be used as cryoprotectants and to increase the collapse temperature, thereby improving product stability [18]. Cryoprotectants stabilize the proteins during the freezing and drying/storage process and improve the collapse temperature. The collapse temperature

is closely linked to product temperature, which in turn, determines the storage temperature [16]. It is generally desired to store the product at room temperature, or 4–8°C, and, thus, excipients combined with a bulking agent that increases the glass (product) temperature are desired. Many informative review articles on freeze-drying are available [16–18].

The screening of protein structure stabilizers for long-term stability is generally achieved by performing folding/unfolding studies in solution conditions. As a first step to perform solution studies, a reasonable quantity of the protein is required. As protein folding is a complex but interesting area of research, this section of formulation development involves basic research. It is essential that suitable analytical, spectroscopic, and thermal methods are developed to investigate the protein folding [19, 20]. In addition, solubility, buffer and pH conditions, and biochemical properties such as size, charge, and PI are determined. Acceptable buffers can be obtained from FDA guidelines and generally regarded as safe (GRAS) list. Commonly used buffers include phosphates, acetates and carbonates, lactates, ascorbates, and citrates. Addition of sodium chloride and other salts should also be investigated as these buffers and salts can increase the stability of a protein.

It is a common practice to employ thermal stress and pH changes to unfold the protein [21, 22], as these are often encountered during the pharmaceutical development of protein drugs. A pH range of 5 to 7 is desired for formulation development [23]. Temperature-dependent studies yield thermodynamic information such as Tm and $\Delta G$ and the stability of the protein. To determine the Tm and $\Delta G$, a simple two-state model for unfolding is assumed [24, 25].

$$N \Leftrightarrow D$$

The equilibrium constant for the unfolding is given by

$$K = [D]/[N]$$

The Gibbs free energy changes can be deduced from the equilibrium constant using the formula $\Delta G = (G_D - G_N) = -RT\ln K$, where R is the gas constant and T is the absolute temperature. The Tm and $\Delta G$ are determined by spectroscopic methods [circular dichroism (CD), fluorescence, second derivative ultraviolet absorption spectroscopy] or by thermal methods (Differential Scanning Calorimetry, DSC) [20]. The aggregation of the protein is monitored using gel filtration, by light-scattering measurements, analytical ultracentrifugation, and field flow fractionation techniques.

The spectroscopic methods allow the estimation of [D] and [N] at any temperature. At the melting temperature Tm, the $\Delta G$ is zero and $\Delta S$ is equal to $\Delta H/Tm$ where [D] and [N] are equal. The measured free energy is typically 5–20 Kcal/mol. Thus, the unfolding of the protein decreases the stability. A strategy to screen for excipients involves compounds that increase the Tm and prevent aggregation and surface adsorption. The above rational strategy has been used to identity excipients for acidic fibroblast growth factor [21].

The existence of intermediate states complicates this simple approach. The reversible denaturation or unfolding step can be followed by the formation of irreversible inactivated forms of the protein (I). The inactivated forms are chemical

reaction (intermolecular disulfide bridges, intermolecular beta strands), aggregation, dissociation of oligomers to monomers, adsorption to surfaces, and incorrect refolding.

$$N \Leftrightarrow D \rightarrow I$$

K is the equilibrium constant that defines $N \Leftrightarrow D$ transition and thermodynamic stability. k is the rate constant for the irreversible transition from $D \rightarrow I$ and provides information regarding the long-term stability. Although the equation shows the inactivation after denaturation, the inactivation can proceed from the native state through the formation of intermediate states [26].

It has been observed for several proteins that the intermediate structures are formed as the protein unfolds from N state to D state [26]. As the protein unfolds, protein loses tertiary structure and, frequently, secondary structure. In some instances, the secondary structure remains intact while the tertiary structure is lost [12], which is clear from spectral studies that measure loss of secondary and tertiary structural changes. One spectroscopic technique that is sensitive to tertiary structure (e.g., fluorescence) would detect changes, whereas other techniques that are sensitive to secondary structures (e.g., far UV CD) do not show any spectral changes. This molecular property is defined as molten globule or structured intermediate [12]. These intermediates expose hydrophobic domains, and thus promote aggregation or surface adsorption.

The mechanisms of protein aggregation and stability are further complicated by the existence of aggregation-prone states in the unfolding of the protein that mimics the native state. Proteins subjected to thermal stress can aggregate from the fully unfolded state, but partially unfolded states have more frequently been implicated in aggregation processes; an example of the latter is the transition to molten globule states, in which proteins retain the majority of their secondary structural features while losing most of their tertiary structure. Such partially unfolded states have more exposed apolar regions and are more prone to aggregation than native or unfolded conformations. For proteins such as interferon gamma (IFN-γ) [27] and granulocyte colony stimulating factor (GCSF) [28], it has been clearly demonstrated that extensive unfolding or formation of molten globule states is not a prerequisite for aggregation and can occur even under solution conditions that are considered to be physiological. Rather, aggregation seems to occur from an expanded state that more resembles the native protein than substantially disrupted structures. In many such cases, the rate-limiting step is not the generation of extensively unfolded states. In the case of rhIFN-γ, aggregation has been shown to occur through the formation of a transiently expanded conformational species, whose surface area is greater than the native state by only 9% [27]. Similarly, native rhGCSF has been shown to aggregate through the formation of a monomeric transition state in which the surface area increase was only 15% [28]. The role of such small conformational fluctuations and their effect on long-term stability has received considerable attention recently.

The large therapeutic proteins, such as factor VIII, are often organized into distinct domains that play a critical role in their structure and function [29, 30]. The folding/unfolding behavior of these individual domains, and the interactions between them, may coordinate various functions in multidomain proteins.

Furthermore, the modular assembly of multidomain proteins often results in complex unfolding characteristics and the existence of intermediate species, which may lead to physical instability including a tendency to aggregate. In the case of rFVIII, neither complete unfolding of the protein nor the generation of partially unfolded states of the protein appears to be a prerequisite for the protein to aggregate. Rather, aggregation appears to involve subtle conformational changes in the C2 domain encompassing the lipid-binding region [31]. The excipients that bind to this domain have been developed as a formulation excipient that promotes the stability [31, 32].

Another complication that can impact the protein folding and stability is the irreversibility and "kinetic" aspects of protein folding. If kinetic components exist, they can interfere with the excipient screening process. It has been observed that the unfolding of rFVIII after thermal perturbation of the protein leads to irreversible aggregation [32, 33]. Based on thermal denaturation studies of multidomain proteins like phosphoglycerate kinase (PGK) and thermolysin, Sanchez-Ruiz et al. demonstrated that irreversible denaturation defies analysis by standard equilibrium methods and suggested that the unfolding may be a kinetically controlled process [34, 35]. It is a general practice to acquire the unfolding profile at different heating rates to determine the Tm. If Tm is found to be a function of the heating rate, it is important to develop excipients to minimize this kinetic contribution to protein instability. Such heating rate dependency on unfolding transitions has also been observed for other proteins including IFN-γ, Immunoglobulin G (IgG), and streptokinase [30, 36, 37].

In general, the heating rate dependence of protein thermal structural perturbations can be understood in terms of the Lumry–Eyring model [38], as shown below. According to this framework, a reversible unfolding step is followed by an irreversible event:

$$\mathrm{N} \underset{k2}{\overset{k1}{\rightleftarrows}} \mathrm{U} \xrightarrow{k3} \mathrm{A} \qquad \text{(scheme 1)}$$

where A is the final (aggregated) state of the native protein N, irreversibly arrived at from the reversible altered form U. In this scheme, the first-order kinetic constants for each reaction j are represented by $k_j$, which varies with temperature according to the Arrhenius equation. If $k_3 >> k_2$, then all of U will be converted into A, and the process can be represented by the simpler kinetic scheme as discussed in detail by Sanchez-Ruiz et al. [34]:

$$\mathrm{N} \xrightarrow{k} \mathrm{A} \qquad \text{(scheme 2)}$$

This scheme represents the limiting case of the Lumry–Eyring model consisting of only two populated states, the native (N) and the final aggregated state (A). This scheme predicts that k is a first-order kinetic constant that varies with temperature, as given by the Arrhenius equation. Based on this simple kinetic model, $T_m$ should vary with heating rate (ν) according to the equation:

$$\ln(\nu/T_m^2) = \ln(AR/E_a) - E_a/RT_m$$

where A is the frequency factor, $E_a$ is the activation energy of the unfolding step, and R is the gas constant. Therefore, a linear plot of ln $(\nu/T_m^2)$ versus $1/T_m$ should

indicate a first-order reaction with slope of $-E_a/R$. The activation energy ($E_a$) for factor VIII was found to be ~535 KJ/Mole (~128 Kcal/Mole) and compares well with the $E_a$ associated with the two transitions observed for the multidomain protein IgG [$E_a$ for the two transitions were reported to be 456 KJ/mole (~109 Kcal/mole) and 692 KJ/mole (~165 Kcal/mole)] [30]. Thus, analysis of the data in terms of kinetic scheme 2 suggests that, under the experimental conditions used in this study, the rate-controlling step in the aggregation of rFVIII may be a unimolecular reaction involving protein conformational changes. Based on our antibody binding assays, we believe that the conformational changes involve at least the lipid-binding region, 2303–2332 of the C2 domain [31]. O-phospho-L-serine (OPLS) is known to bind to the lipid-binding region in the C2 domain [31]. The addition of ligand capable of binding to this region inhibited conformational changes and interfered with the aggregation process.

It is important to study the protein folding at different concentrations of the protein once the selection of excipients is narrowed [39], which is critical for scale up and for proteins such as monoclonal antibody products where higher doses are desired. The kinetic aspects of protein folding and concentration-dependent aggregation could potentially lead to instability. Overall, equilibrium unfolding is often used to screen excipients for the development of protein formulations. However, minor conformational changes that resemble the native state can lead to aggregation interfering with the equilibrium unfolding measurements. In addition, use of higher protein concentrations can accelerate the kinetically controlled aggregation process [39]. Thus, formulation development requires the characterization of minor conformational changes or aggregation prone states and their kinetics. Previous findings on the molecular details of the aggregation prone state of rFVIII and their kinetics [32] have contributed toward the rational design of a rFVIII-phosphoserine (PS) complex with improved physical characteristics [31].

## 3 IMMUNOLOGICAL PROPERTIES OF PROTEIN THERAPEUTICS

### 3.1 Immunogenicity of Proteins

Proteins and peptides are more immunogenic than conventional xenobiotics, and they are more prone to evoke immunotoxicity reactions (Table 2). The immunotoxicity is usually manifested in the form of a humoral immune response involving the formation of antibodies [40–42]. The antibodies can abrogate the efficacy of the administered protein by neutralizing its activity [41–43] or by altering its pharmacokinetics, thus complicating the dose-response relationships. The risk of generating an immune response is dictated by a host of product-specific factors such as product origin, molecular structure, impurities/degradation products, and formulation [44]. Proteins of animal or microbial origin can invariably be expected to elicit an immune response as they will be recognized as foreign proteins (neoantigen) triggering a classic immune response involving the B and T cells [45]. Human proteins, expressed in microbes such as *Escherichia Coli* can be immunotoxic because of lack of glycosylation [46], which can expose immunogenic neoepitopes with the possibility of epitope spreading leading to immune responses against conserved regions in the long term. Even proteins of human origin, when

**TABLE 2.**

| Therapeutic Protein | Detected Incidence of Anti-product Antibody |
|---|---|
| OKT3 | ~80% |
| Remicade | 10–57% |
| Roferon | 20–50% |
| Proleukin (interleukin-2) | 47–74% |
| Betaferon | 44% |
| Advate | 15–30% |
| Bone morphogenic protein-7 | 13–38% |
| Enbrel (TNF receptor / anti-TNF IgG) | 16% |
| Thrombopoietin | ≤10% |
| PEGylated-MGDF | 0.5–1.6% |

administered to humans, genetically deficient of the respective protein, will trigger a classic immune response as the patients are not immunotolerant to the protein. The above is illustrated by the observations in patients having severe forms (negligible endogenous protein production) of Fabry's disease [47], an X-linked deficiency of α-galactosidase A enzyme or Hemophilia A, where almost 30% of the patients develop antibodies upon therapy with the respective protein [48, 49]. Proteins used as therapeutic agents to supplement endogenous proteins (e.g., erythropoeitin, IL-2) have the potential of immunotoxicity because of the breaking of immune tolerance to self proteins. The breaking of immunotolerance to self proteins results more because of product-related factors (impurities, degradation products, and formulation excipients) [50] and dosing regimens employed for therapy [51].

A major challenge faced by a protein formulation scientist is the inability to predict or asses the risk of immunotoxicity of the final product. The prediction of immune response requires a comprehensive consideration of protein structure, putative genetic background of the patient population, the formulation excipients, as well as frequency and route of administration [44]. The ability to asses the potential of immunotoxicity of a protein therapeutic is mainly hampered by the lack of predictive *in vitro* or *in vivo* models. Animal models, widely used in preclinical drug testing, cannot predict immune responses in humans because of issues of immunotolerance and basic differences in the structure and function of the immune system that may not be readily translated across species. Mouse models have been the foundation for the majority of the work done toward understanding immune responses. However, when it comes to prediction of immunotoxicity of therapeutic proteins, significant discrepancies in adaptive and innate immunity preclude any extrapolation to humans. Nevertheless, animal models can still be used to conduct preclinical immunotoxicity studies, provided one appreciates all the caveats associated with use of such animal models. Evidence suggests that nonhuman primates and transgenic/knockout mice models may be useful for predicting the relative immunogenicity of protein therapeutics and assessing the risk of immunotoxicity [52, 53].

As mentioned above, several product-related factors can influence the immunogenicity of protein therapeutics. The impact of trace impurities and excipients used

in the formulation on the immunogenicity of the protein relative to the pure native protein should be evaluated in relevant animal models (nonhuman primates or transgenic/knockout mice) very early during the development cycle. Close attention should be paid to the impact of excipients and impurities on the structure of the protein. Structural changes can promote formation of neo-epitopes, which can initiate immune response in tolerant individuals followed by epitope spreading and thus resulting in immunotoxicity. The methods used to detect immune responses should be extremely sensitive to detect even small amounts of antibodies in the animal models, as even small amounts of inhibitory antibodies observed in animal models may indicate the potential for significant immunotoxicity in humans. Further, cellular basis for generation of the immune response should be thoroughly investigated in the animal models with the emphasis on mechanism. The above information can greatly enhance our ability to anticipate and devise strategies to prevent or minimize immunotoxicity issues in humans early on during the product development cycle. Aggregates are one of the major degradation products formed in protein therapeutics. Both chemical and physical degradation can give rise to the formation of aggregates. It is generally believed that the aggregates can enhance the immunogenicity of therapeutic protein preparations [44]. Although the presence of aggregates in protein preparations can influence its immunotoxicity, the generalization of the above statement may not be appropriate. Several studies with aggregates of various proteins have shown that aggregates do not always enhance the development of antibodies [54]. Aggregation can be associated with structural changes of the protein. The mechanism (physical or chemical) causing the aggregation of the protein dictates the structural features of the aggregated protein. These structural changes can significantly alter the epitope repertoire of the aggregated protein relative to the monomeric protein, thereby altering its properties as an immunogen [55]. Additionally, the altered structure can also change the processing of the immunogen by the endosomal proteases in the antigen-presenting cells, which is essential for presentation by the major histocompatibility complex [55]. Hence, to understand the risk of immunotoxicity associated with the administration of aggregates, one needs to have the explicit understanding of the correlation between structure and immunogenicity. Aggregates that retain the structure of the native protein are more likely to cause immunotoxicity. However, aggregates that do not retain the structure of the protein should not be considered safe as the exact immunotoxicological consequences are not known and should be investigated.

Individuals having a genetic predisposition for generating an immune response against a therapeutic agent are unlikely to benefit from the prior knowledge generated during the preclinical immunotoxicity testing. However, methods and techniques identifying their genetic predisposition will greatly benefit this subset of patients as alternatives can be explored early or before initiating therapy. *In vitro* systems using dendritic cells and B-cells isolated from patients need to be developed and hold promise as an individualized immunotoxicity screening method before initiating therapy.

### 3.2 Formulation and Delivery Considerations to Reduce Immunogenicity

Although the mechanism is not clear, immunodominant epitopes in the protein sequence and aggregates or impurities in the protein product can facilitate the

immunogenicity of the protein. Immunodominant epitopes (linear or conformational) are subunits of the protein that are most easily recognized by the immune system and can evoke a specific immune response [56, 57]. Proteins having these immunodominant epitopes when administered in humans do run the risk of breaking tolerance. Additionally, in humans who are intolerant to a given protein, the immunodominant epitopes are most likely to evoke the immune response. A strategy used to reduce immunogenicity of a protein is to identify the immunodominant regions in the protein and replace them with innocuous sequences while retaining the original structure and function of the protein. An example of the above approach is the development of hybrid porcine/human FVIII molecule in which the immunodominant regions in the human molecule were replaced with less immunogenic regions of the porcine FVIII [58]. The above strategy has the potential for development of protein molecules that are less immunogenic by virtue of their design.

Another strategy used frequently to reduce immunogenicity of protein molecules by tampering with its basic chemistry is the use of "pegnology." The term signifies the science of attaching polyethylene glycol polymer chains to specific residues in the protein molecule, thereby influencing its properties of immunogenicity and circulation half-lives of the parent protein molecule. The mechanism by which the pegylation of the protein molecule can reduce immunogenicity is the shielding of specific immunodominant epitopes. The shielding phenomenon prevents the recognition of specific epitopes by B-cells, thereby preventing initiation of an immune response. Adagen (pegylated form of adenosine deaminase) and Oncaspar (Pegylated asparaginase) are examples of pegylated proteins currently in clinical use where immunogenicity has been reduced relative to the parent protein by pegylation. Another desirable property of pegylation is that it usually reduces the clearance of the protein molecule, which serves a dual purpose as it increases the circulation half-life of the protein molecule, thereby reducing the frequency of administration. Hence, the overall affect of pegylation is an improved therapeutic protein.

## 4 PHARMACOKINETICS AND PHARMACODYNAMICS

It is generally accepted that the successful development of new therapeutic entities is strictly dependent on their pharmacokinetic/pharmacodynamic properties. The classic PK concepts can be successfully applied to large molecules, leading to optimization of the dosing regimen to achieve the desired pharmacological effect. However, the pharmacokinetic scientists usually encounter challenging difficulties in interpreting the concentration-time profiles for the protein drugs, partly because of the limitations of the detection systems in terms of lower accuracy and precision of the bioassays. Moreover, the presence of endogenous proteins in some instances and the low bioavailability and the rapid elimination from the systemic circulation in others will further decrease the ability to determine the true levels of therapeutic proteins. The recent advances in immunoassays and mass spectrometry analysis improved our ability for quantification. However, caution should be taken when analyzing data from immuno-/radio-labeled assays or LC/MS, because of the inability of those methods to discern between active and inactive protein molecules (native state vs. proteolytic fragments/denaturated states). Extensive review arti-

cles on the subject of validation of bioanalytical methods are available in the literature [59].

### 4.1 Formulation and Delivery Approaches to Overcome PK Issues

The lack of protein stability within the gastrointestinal tract leads [60] to the administration of therapeutic biomolecules through the parenteral routes. Among those, the i.v. route grants complete bioavailability, a rapid distribution within the blood with achievement of high drug concentrations [61]. On the other hand, the s.c. and i.m. administration of therapeutics protein is very appealing based on the possibility of an increase in residence time because of slow absorption from the injection site [44]. Hence, efforts are being made to develop new strategies for alternative parenteral routes. The extent of absorption from the interstitial space varies from case to case and is dependent on the extent of proteolytic degradation at the injection site. In many instances, the proteolytic degradation is a saturable process and an increase in the therapeutic dose usually translates to increased bioavailability [62]. However, a higher dose might break the tolerance level and lead to undesired immunological properties, as discussed in the previous section. Hence, to exploit the advantages offered by parenteral routes other than i.v., a need exists to design improved, novel methodology or delivery vehicles that would limit the proteolytic degradation and prevent the rapid elimination from the systemic circulation.

***PEGylation.*** Poly (ethylene glycol) (PEG) is a nontoxic, highly soluble polymer [63] that can be linked to a protein via several chemical modifications. This approach includes attaching PEG to ε-amino (lysine), thiol (cysteine), and hydroxyl (tyrosine, serine) groups or carbohydrate moieties (glycoproteins) as well as at the N- or the C-terminal of peptides and proteins [64]. The foundation of this technology resides in the steric hindrance factor induced by the bulky covalently attached polymer. For instance, the covalent linkage of PEG to protein surfaces leads to shielding of putative immunodominant epitopes that will directly affect the antibody recognition process [65]. Apart from the improved immunological properties, the shielding factor exerted by the PEG alters the accessibility of proteolytic enzymes [66]. It is also responsible for a decrease in receptor-mediated endocyotsis with a subsequent decrease in clearance. In addition, the net increase in hydrodynamic radius (two or three water molecules—tightly bound per ethylene glycol monomer) will lead to a decrease in renal clearance, further improving the pharmacokinetic properties of the therapeutic proteins and peptides [67].

The prolonged circulation time, in turn, reduces the frequency of administration that can further reduce the adverse immune response and can maximize the pharmacological effects [44].

PEGylation of asparaginase (Oncaspar) resulted in greater than 96% reduction in the incidence of detectable immunogenicity [68]. In a phase III study involving 2000 patients, a native interferon induced antibodies in 14.5% of recipients in contrast to only 1.5% of patients receiving PEGlyated derivative. In addition, the renal filtering was reduced by ~100-fold for PEG-interferon relative to the unpegylated one. Thus, PEGylation seems to be a promising strategy for reducing immunogenecity and for reducing the frequency of dosing.

Historically, the first generation of PEG derivatives targeted the labile, solvent-accessible amino groups present either at the N-terminal or on the side chain of positively charged amino acid residues [68]. Lysine is one of the amino acids commonly found on protein surfaces. Hence, it was relatively easy to covalently couple PEG or other molecules at this position. However, the high occurrence of lysine in proteins was also responsible for a high degree of heterogeneity in the first generation of PEGylated proteins. Therefore, both Oncaspar [68] and Adagen [69] (PEG adenosine deamidase) suffered from this design flaw. However, the use of different PEGylated isomers received FDA approval as long as data was provided to support batch-to-batch reproducibility.

The second generation of PEG derivatives were merely designed to overcome the heterogeneity concerns present in the first generation of PEGylated proteins. To accomplish PEGylation at specific positions, the novel PEG reagents took advantage of the subtle differences in the pKa of $\alpha$ and $\varepsilon$ amino groups present on the side chains of amino acid residues [70]. Another successful approach was to target the thiol group of cysteine residues that are rarely occurring on proteins surfaces and, hence, should lead to a more homogeneous population of PEG derivatives [71].

However, despite those advances in PEG technology, several issues still plague the formation of PEGylated derivatives. The PEGylation approaches are limited by substantial loss of specific activity of the protein, leading to inconsistent therapeutic effects. It is worthwhile to mention that new strategies are being developed to address the loss of biological activity and function. In this line, Peleg-Shulman et al. [72] adopted a reversible PEGylation approach in the case of interferon $\alpha 2$, in which the PEG derivative undergoes limited hydrolysis under physiological conditions. The advantage of the reversible PEG-interferon lies in the fact that the released interferon can penetrate the tissue effortlessly and therefore exert its biological function. On the other hand, the PEG derivative retains its PEGylation advantages in terms of PK parameters such as the long circulation half-life.

To maintain the protein bioactivity, other valuable technologies incorporate a blocking step of the active site before PEGylation through the use of ligands/enzyme inhibitors. Those inhibitors can be free in solution or bound to a solid phase, facilitating the subsequent removal procedure of free PEG.

Altogether, the PEGylation approaches presented herein have been successful to small proteins that, overall, have a limited number of active sites on the surface. Such approaches might not be applicable to all proteins, particularly for the larger ones. It is too early to speculate whether those limitations might be responsible for the low anti-PEG antibody response evoked following administration in humans of PEG-uricase [73].

***Lipid Structures as Drug Delivery Vehicles.*** Liposomes and lipid-based nanoparticles can act as a sustained release system for protein drugs. The formulation of Interferon Gamma in negatively charged liposomes [74] increases the residence time of the protein by approximately 19 times upon s.c. administration. Further, liposomal formulation of factor VIII improves physical stability, reduces immunogenicity, and prolongs the circulation time of factor VIII (unpublished results). Liposomes can be PEGylated to further extend their blood circulation time. Compared with classic liposomes, PEGylated vesicles show increased half-

life, decreased plasma clearance [75], and a shift in distribution in favor of diseased tissues. PEG is incorporated into the lipid bilayer of the liposome, forming a hydrated shell that protects it from destruction by proteins. Pegylated liposomes are also less extensively taken up by the reticuloendothelial system [76] and are less likely to leak drug while in circulation. This approach of PEGylation avoids the covalent modification of the protein and the complication developing from these modifications.

In addition, liposomes can act as protein stabilizers that reduce the protease degradation and improve the physical stability (unpublished data, [62]). The aggregation promoted by the existence of intermediate states is a common problem. As these intermediate structures expose hydrophobic domains, the addition of liposomes at critical stages of the protein unfolding stabilize the protein against aggregation and help to refold the protein to native conformation. This property is referred to as chaperones, and the use of chaperones as a formulation excipient has not received considerable attention. Another mechanism, lipidic particles, can stabilize the protein by competing with the aggregation kinetics. For example, protein inactivation after the unfolding step is controlled by a kinetic process. Lipidic particles can interfere with this aggregation step. We have shown in our laboratory that the aggregation of factor VIII is kinetically controlled and undergoes irreversible denaturation [32]. However, in the presence of lipidic particles, the recovery of factor VIII upon refolding is much higher compared to free factor VIII.

***PLGA Microspheres.*** Polymers such as PLGA microspheres have been used as delivery vehicles for peptides. In this delivery system, the peptide is embedded in a polymer matrix [77]. Two biodegradable polymeric systems are currently approved by FDA for use, Lupron [78] and Zoladex [79]. Several peptide drugs have been formulated in the biodegradable polymers. These PLGA polymers have been shown to release the peptide over the period of 1 to 3 months, which alleviates the pain caused by daily injections and improves the patient's compliance [78]. The monomers lactic and glycolic connected by ester bonds undergo ester hydrolysis that releases the peptide [77]. The peptide encapsulation is performed in methylene chloride. Such harsh conditions denature protein molecules but are acceptable for peptides. The microsphere encapsulation provides sustained release of IL-12 at the tumor site and avoids systemic toxicity associated with the administration of IL-12 [80]. However, there is substantial loss of activity.

## REFERENCES

1. Suiter T M (2002). First and next generation native rFVIII in the treatment of hemophilia A. What has been achieved? Can patients be switched safely? Semin. Thromb. Hemost. 28(3):277–284.
2. Rosendaal F R, Smit C, Briet E (1991). Hemophilia treatment in historical perspective: A review of medical and social developments. Ann. Hematol. 62(1):5–15.
3. Larner A J (1987). The molecular pathology of haemophilia. *Q. J. Med.* 63(242):473–491.

4. Xie M, Schowen R L (1999). Secondary structure and protein deamidation. *J. Pharm. Sci.* 88(1): 8–13.
5. Meinwald Y C, Stimson E R, Scheraga H A (1986). Deamidation of the asparaginylglycyl sequence. *Int. J. Pept. Protein Res.* 28(1):79–84.
6. Manning M C, Patel K, Borchardt R T (1989). Stability of protein pharmaceuticals. *Pharm. Res.* 6(11):903–918.
7. Fransson J, et al. (1996). Oxidation of human insulin-like growth factor I in formulation studies: Kinetics of methionine oxidation in aqueous solution and in solid state. *Pharm. Res.* 13(8):1252–1257.
8. Costantino H R, Langer R, Klibanov A M (1995). Aggregation of a lyophilized pharmaceutical protein, recombinant human albumin: Effect of moisture and stabilization by excipients. *Biotechnol.* 13(5):493–496.
9. Robert Liu W, R.L.A.M.K. (1991). Moisture-induced aggregation of lyophilized proteins in the solid state. *Biotechnol. Bioengine.* 37(2):177–184.
10. Wang W (1999). Instability, stabilization, and formulation of liquid protein pharmaceuticals. *Int. J. Pharm.* 185(2):129–188.
11. Chi E Y, et al. (2003). Physical stability of proteins in aqueous solution: Mechanism and driving forces in nonnative protein aggregation. *Pharm. Res.* 20(9):1325–1336.
12. Kuwajima K (1996). The molten globule state of alpha-lactalbumin. *Faseb. J.* 10(1): 102–109.
13. Benson E A, et al. (1988). Flocculation and loss of potency of human NPH insulin. *Diabetes Care.* 11(7):563–596.
14. Baneyx F, Mujacic M (2004). Recombinant protein folding and misfolding in Escherichia coli. *Nat. Biotechnol.* 22(11):1399–1408.
15. Sefton M V, Antonacci G M (1984). Adsorption isotherms of insulin onto various materials. *Diabetes.* 33(7):674–680.
16. Carpenter J F, et al. (1997). Rational design of stable lyophilized protein formulations: Some practical advice. *Pharm. Res.* 14(8):969–975.
17. Tang X, Pikal M J (2004). Design of freeze-drying processes for pharmaceuticals: Practical advice. *Pharm. Res.* 21(2):191–200.
18. Chang L L, et al. (2005). Mechanism of protein stabilization by sugars during freeze-drying and storage: Native structure preservation, specific interaction, and/or immobilization in a glassy matrix? *J. Pharm. Sci.* 94(7):1427–1444.
19. Pelton J T, McLean L R (2000). Spectroscopic methods for analysis of protein secondary structure. *Anal. Biochem.* 277(2):167–176.
20. Bruylants G, Wouters J, Michaux C (2005). Differential scanning calorimetry in life science: Thermodynamics, stability, molecular recognition and application in drug design. *Curr. Med. Chem.* 12(17):2011–2020.
21. Tsai P K, et al. (1993). Formulation design of acidic fibroblast growth factor. *Pharm. Res.* 10(5):649–659.
22. Remmele R L, Jr., et al. (1998). Interleukin-1 receptor (IL-1R) liquid formulation development using differential scanning calorimetry. *Pharm. Res.* 15(2):200–208.
23. Osterberg T, Fatouros A, Mikaelsson M (1997). Development of freeze-dried albumin-free formulation of recombinant factor VIII SQ. *Pharm. Res.* 14(7):892–898.
24. Baldwin R L, Rose G D (1999). Is protein folding hierarchic? I. Local structure and peptide folding. *Trends. Biochem. Sci.* 24(1):26–33.
25. Privalov P L (1979). Stability of proteins: Small globular proteins. *Adv. Protein Chem.* 33:167–241.

26. Fink A L (1998). Protein aggregation: Folding aggregates, inclusion bodies and amyloid. *Fold. Des.* 3(1):R9–23.
27. Kendrick B S, et al. (1998). A transient expansion of the native state precedes aggregation of recombinant human interferon-gamma. *Proc. Natl. Acad. Sci. USA*. 95(24): 14142–14146.
28. Krishnan S, et al. (2002). Aggregation of granulocyte colony stimulating factor under physiological conditions: Characterization and thermodynamic inhibition. *Biochemi.* 41(20):6422–6431.
29. Jaenicke R (1999). Stability and folding of domain proteins. *Prog. Biophys. Mol. Biol.* 71(2):155–241.
30. Vermeer A W, Norde W (2000). The thermal stability of immunoglobulin: Unfolding and aggregation of a multi-domain protein. *Biophys. J.* 78(1):394–404.
31. Ramani K, et al. (2005). Lipid binding region (2303–2332) is involved in aggregation of recombinant human FVIII (rFVIII). *J. Pharm. Sci.* 94(6):1288–1299.
32. Ramani K, et al. (2005). Aggregation kinetics of recombinant human FVIII (rFVIII). *J. Pharm. Sci.* 94(9):2023–2029.
33. Grillo A O, et al. (2001). Conformational origin of the aggregation of recombinant human factor VIII. *Biochem.* 40(2):586–595.
34. Sanchez-Ruiz J M, et al. (1988). Differential scanning calorimetry of the irreversible thermal denaturation of thermolysin. *Biochem.* 27(5):1648–1652.
35. Galisteo M L, Mateo P L, Sanchez-Ruiz J M (1991). Kinetic study on the irreversible thermal denaturation of yeast phosphoglycerate kinase. *Biochem.* 30(8):2061–2066.
36. Kendrick B S, et al. (1998). Aggregation of recombinant human interferon gamma: Kinetics and structural transitions. *J. Pharm. Sci.* 87(9):1069–1076.
37. Azuaga A I, et al. (2002). Unfolding and aggregation during the thermal denaturation of streptokinase. *Eur. J. Biochem.* 269(16):4121–4133.
38. Lumry R, Eyring H (1954). Conformation changes of proteins. *J. Phys. Chem.* 58:110–120.
39. Kim Y S, et al. (2003). Congo red populates partially unfolded states of an amyloidogenic protein to enhance aggregation and amyloid fibril formation. *J. Biol. Chem.* 278(12):10842–10850.
40. Kulkarni R, et al. (2001). Therapeutic choices for patients with hemophilia and high-titer inhibitors. *Am. J. Hematol.* 67(4):240–246.
41. Grauer A, Ziegler R, Raue F (1995). Clinical significance of antibodies against calcitonin. *Exp. Clin. Endocrinol. Diabetes.* 103(6):345–351.
42. Casadevall N, et al. (2002). Pure red-cell aplasia and antierythropoietin antibodies in patients treated with recombinant erythropoietin. *N. Engl. J. Med.* 346(7):469–475.
43. Scandella D H, et al. (2001). In hemophilia A and autoantibody inhibitor patients: The factor VIII A2 domain and light chain are most immunogenic. *Thromb. Res.* 101(5):377–385.
44. Braun A, et al. (1997). Protein aggregates seem to play a key role among the parameters influencing the antigenicity of interferon alpha (IFN-alpha) in normal and transgenic mice. *Pharm. Res.* 14(10):1472–1478.
45. Chance R E, Root M A, Galloway J A (1976). The immunogenicity of insulin preparations. *Acta. Endocrinol. Suppl.* 205:185–198.
46. Gribben J G, et al. (1990). Development of antibodies to unprotected glycosylation sites on recombinant human GM-CSF. *Lancet.* 335(8687):434–437.
47. Linthorst G E, et al. (2004). Enzyme therapy for Fabry disease: Neutralizing antibodies toward agalsidase alpha and beta. *Kidney Int.* 66(4):1589–1595.

48. Jacquemin M G, Saint-Remy J M (1998). Factor VIII immunogenicity. *Haemophil.* 4(4):552–557.
49. Lollar P, et al. (2001). Factor VIII inhibitors. *Adv. Exp. Med. Biol.* 489:65–73.
50. Schellekens H (2002). Bioequivalence and the immunogenicity of biopharmaceuticals. *Nat. Rev. Drug. Discov.* 1(6):457–462.
51. Ross C, et al. (2000). Immunogenicity of interferon-beta in multiple sclerosis patients: Influence of preparation, dosage, dose frequency, and route of administration. Danish Multiple Sclerosis Study Group. *Ann. Neurol.* 48(5):706–712.
52. Palleroni A V, et al. (1997). Interferon immunogenicity: Preclinical evaluation of interferon-alpha 2a. *J. Interferon. Cytokine Res.* 17 (Suppl 1):S23–27.
53. Wierda D, Smith H W, Zwickl C M (2001). Immunogenicity of biopharmaceuticals in laboratory animals. *Toxicol.* 158(1–2):71–74.
54. Koch C, et al. (1996). A comparison of the immunogenicity of the native and denatured forms of a protein. *Apmis.* 104(2):115–125.
55. Purohit V S, Middaugh C R, Balasubramanian S V (2006). Influence of aggregation on immunogenicity of recombinant human Factor VIII in hemophilia A mice. *J. Pharm. Sci.* 95(2):358–371.
56. Reding M T, et al. (2003). Human CD4+ T-cell epitope repertoire on the C2 domain of coagulation factor VIII. *J. Thromb. Haemost.* 1(8):1777–1784.
57. Reding M T, et al. (2004). Epitope repertoire of human CD4(+) T cells on the A3 domain of coagulation factor VIII. *J. Thromb. Haemost.* 2(8):1385–1394.
58. Barrow R T, et al. (2000). Reduction of the antigenicity of factor VIII toward complex inhibitory antibody plasmas using multiply-substituted hybrid human/porcine factor VIII molecules. *Blood.* 95(2):564–568.
59. Miller K J, et al. (2001). Workshop on bioanalytical methods validation for macromolecules: Summary report. *Pharm. Res.* 18(9):1373–1383.
60. Blum P M, et al. (1981). Survival of oral human immune serum globulin in the gastrointestinal tract of low birth weight infants. *Pediatr. Res.* 15(9):1256–1260.
61. Tang L, et al. (2004). Pharmacokinetic aspects of biotechnology products. *J. Pharm. Sci.* 93(9):2184–2204.
62. Fatouros A, Liden Y, Sjostrom B (2000). Recombinant factor VIII SQ–stability of VIII: C in homogenates from porcine, monkey and human subcutaneous tissue. *J. Pharm. Pharmacol.* 52(7):797–805.
63. Yamaoka T, Tabata Y, Ikada Y (1994). Distribution and tissue uptake of poly(ethylene glycol) with different molecular weights after intravenous administration to mice. *J. Pharm. Sci.* 83(4):601–606.
64. Veronese F M, Pasut G, (2005). PEGylation, successful approach to drug delivery. *Drug Discov. Today.* 10(21):1451–1458.
65. Hershfield M S, et al. (1991). Use of site-directed mutagenesis to enhance the epitope-shielding effect of covalent modification of proteins with polyethylene glycol. *Proc. Natl. Acad. Sci. U S A.* 88(16):7185–7189.
66. Delgado C, Francis G E, Fisher D (1992). The uses and properties of PEG-linked proteins. *Crit. Rev. Ther. Drug Carrier Syst.* 9(3–4):249–304.
67. Molineux G (2004). The design and development of pegfilgrastim (PEG-rmetHuG-CSF, Neulasta). *Curr. Pharm. Des.* 10(11):1235–1244.
68. Graham M L (2003). Pegaspargase: A review of clinical studies. *Adv. Drug Deliv. Rev.* 55(10):1293–1302.

69. Levy Y, et al. (1988). Adenosine deaminase deficiency with late onset of recurrent infections: Response to treatment with polyethylene glycol-modified adenosine deaminase. *J. Pediatr.* 113(2):312–317.
70. Kinstler O B, et al. (1996). Characterization and stability of N-terminally PEGylated rhG-CSF. *Pharm. Res.* 13(7):996–1002.
71. Goodson R J, Katre N V (1990). Site-directed pegylation of recombinant interleukin-2 at its glycosylation site. *Biotechno.* 8(4):343–346.
72. Peleg-Shulman T, et al. (2004). Reversible PEGylation: A novel technology to release native interferon alpha2 over a prolonged time period. *J. Med. Chem.* 47(20):4897–4904.
73. Ganson N J, et al. (2005). Control of hyperuricemia in subjects with refractory gout, and induction of antibody against poly(ethylene) glycol (PEG), in a phase I trial of subcutaneous PEGylated urate oxidase. *Arthritis. Res. Ther.* 8(1):R12.
74. Van Slooten M L, et al. (2001). Liposomes as sustained release system for human interferon-gamma: Biopharmaceutical aspects. *Biochim. Biophys. Acta.* 1530(2–3):134–145.
75. Allen C, et al. (2002). Controlling the physical behavior and biological performance of liposome formulations through use of surface grafted poly(ethylene glycol). *Biosci. Rep.* 22(2):225–250.
76. Allen T M, et al. (1991). Liposomes containing synthetic lipid derivatives of poly(ethylene glycol) show prolonged circulation half-lives in vivo. *Biochim. Biophys. Acta.* 1066(1): 29–36.
77. Cohen S, et al. (1991). Controlled delivery systems for proteins based on poly(lactic/glycolic acid) microspheres. *Pharm. Res.* 8(6):713–720.
78. Cox M C, Scripture C D, Figg W D (2005). Leuprolide acetate given by a subcutaneous extended-release injection: Less of a pain? *Expert Rev. Anticancer. Ther.* 5(4):605–611.
79. Perry C M, Brogden R N (1996). Goserelin. A review of its pharmacodynamic and pharmacokinetic properties, and therapeutic use in benign gynaecological disorders. *Drugs.* 51(2):319–346.
80. Hill H C, et al. (2002). Cancer immunotherapy with interleukin 12 and granulocyte-macrophage colony-stimulating factor-encapsulated microspheres: Coinduction of innate and adaptive antitumor immunity and cure of disseminated disease. *Cancer Res.* 62(24):7254–7263.

# PHARMACOKINETICS

Nobuhito Shibata,[1] Yukako Ito,[2] and Kanji Takada[2]
[1]*Doshisha Women's College of Liberal Arts, Kyoto, Japan*
[2]*Kyoto Pharmaceutical University, Kyoto, Japan*

**Chapter Contents**

1 Introduction
2 Insulin
  2.1 Background
  2.2 Pharmacokinetics of Human Insulin after Intravenous Administration
  2.3 Pharmacokinetics of Human Insulin after Subcutaneous Administration
  2.4 Pharmacokinetics of Rapid-acting Insulin Analog after Subcutaneous Administration
  2.5 Other Formulations of Human Insulin
3 Erythropoietin
  3.1 Background
  3.2 Pharmacokinetics of Erythropoietin in Healthy Subjects
  3.3 Pharmacokinetics of Erythropoietin in Patients
  3.4 Erythropoietin Formulation
4 Granulocyte-colony Stimulating Factor (G-CSF)
  4.1 Background
  4.2 Pharmacokinetics of G-CSF in Animals
  4.3 Pharmacokinetics of G-CSF in Humans
  4.4 G-CSF Formulation
5 Interferon
  5.1 Background
  5.2 Pharmacokinetics of Interferon in Animals
  5.3 Pharmacokinetics of Interferon in Humans
  5.4 Interferon Formulation

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

6 Growth Hormone
6.1 Background
6.2 Pharmacokinetics of Growth Hormone in Animals
6.3 Pharmacokinetics of Growth Hormone in Humans
6.4 Growth Hormone Formulation
7 Leuprolide
7.1 Background
7.2 *In vitro* Permeation Study of Leuprolide
7.3 Clinical Pharmacokinetics of Depot Leuprolide
7.4 Transdermal Iontophoretic Delivery of Leuprolide
8 Desmopressin
8.1 Background
8.2 Pharmacokinetics of Desmopressin in Humans
8.3 Renal and Biliary Excretion, and Sex Differences in Desmopressin Pharmacokinetics
8.4 Desmopressin Formulation
9 Antibodies
9.1 Background
9.2 Pharmacodynamics and Pharmacokinetics of mAbs
9.3 Muromonab CD-3 Pharmacokinetics
9.4 Basiliximab Pharmacokinetics
9.5 Trastuzumab Pharmacokinetics
9.6 Infliximab Pharmacokinetics
9.7 Rituximab Pharmacokinetics
9.8 Palivizumab Pharmacokinetics
10 Recommendations for Future Study
References

## 1 INTRODUCTION

Recombinant DNA technology produced many protein drugs in large scale and many important protein drugs like erythropoietin (EPO) for anemia-associated diseases, granulocyte-colony stimulating factor (G-CSF) for leucopoenia, and insulin for diabetes are now clinically available [1–3]. In addition, peptide hormones like leuprolide [an agonist of luteinizing hormone-releasing hormone (LHRH)] are also clinically used. However, these peptide/protein drugs are unstable because of hydrolytic degradation in the body [4]. Peptide/protein drugs also have a barrier for their administration because of low absorption efficiency. Therefore, their is parenteral preparation, *main dosage form* in which case intravenous (i.v.) or subcutaneous (s.c.) injection is used. After i.v. injection of these drugs, the administered drug molecules are completely available to the systemic circulation (i.e., bioavailability (BA) is 100%). On the other hand, BA becomes lower when these drugs are injected s.c. or i.m. In some protein drugs like interferon gamma, BAs after s.c. and i.m. injection are greater than 100% (i.e., 200% and 300%, respectively [5–7]. Furthermore, it is recognized that protein drugs whose mole-

cular weight is larger than about 20 kDa are absorbed into the systemic circulation via lymphatic route. As a result, a large $T_{max}$ value is obtained, when peak serum or plasma drug level appears after administration; for example, $T_{max}$ is approximately 8 h for s.c.-injected EPO. After the serum drug level reached its maximum level, $C_{max}$, serum drug concentration decreases. In the case of organic compound drugs, the elimination process is composed of hepatic metabolism and renal excretion pathways. However, for peptide/protein drugs, we must consider the hydrolytic degradation process because of poor tolerability of these drugs to endogenous hydrolytic enzymes. Therefore, the elimination half-life of peptide/protein drugs is generally shorter than that of organic compound drugs. To increase the stability of peptide/protein drugs, several approaches have been performed and second-generation peptide/protein drugs have been developed and launched on the market [4], including PEGylated protein drugs (i.e., pegylation) and attachment of polyethylene glycol (PEG) to protein molecules. In the case of PEG-EPO, the administration frequency has decreased to once per week as compared with three times per week for the conventional EPO preparations; although almost the same serum EPO levels are obtained in the two types of EPO preparations. Most of the current peptide/protein drugs are injectable, and there is great commercial potential for delivery systems that use alternative administration routes like the pulmonary route and the nasal route, as suggested from insulin new dosage forms. In these cases, the pharmacokinetic profile of the peptide/protein drugs shows a completely different pattern as compared with that obtained after i.v. injection.

In this section, the pharmacokinetics of clinically important peptide/protein drugs, such as insulin, EPO, G-CSF, interferon, growth hormone, leuprolide, desmopressin, and antibodies, are described in relation to their administration routes and formulations (i.e., dosage forms).

## 2 INSULIN

### 2.1 Background

The finding of insulin in 1922 by Ranting and Best is one of the greatest discoveries in the history of medicine [8], and now it has been already proven that adequate control of blood glucose delays or prevents the progression of diabetic complications. To achieve the suggested targets for glycemic control necessary to reduce the incidence of diabetic complications, it has been established that a more intensive insulin regimen requiring multiple insulin injections is required for patients with type 1 diabetes mellitus. From a patient's point of view, however, noninvasive insulin delivery systems have been required for a long time. Indeed, a major limitation for advancing to intensive insulin therapy is that the only viable way to administer insulin is through injection. Several other methods of insulin delivery are now available or in development, including: continuous subcutaneous insulin infusion by a wearable infusion pump; total or segmented transplantation of a pancreas; transplantation of isolated islet cells; implantation of a programmable insulin pump; oral, nasal, rectal, and transdermal mechanisms of insulin delivery; insulin analogs; implantation of polymeric capsules that give continuous or time-pulsed release of insulin; and implantation of a biohybrid artificial pancreas, which uses

encapsulated islets [9]. Many of these methods of insulin delivery are aimed at achieving a more physiological means of delivery of the insulin, at improving glycemic control, and, hopefully, at minimizing the secondary complications of diabetes.

On the other hand, during the nearly 80 years of insulin therapy, insulin preparations have undergone dramatic changes in their purification, production, and formulation. Despite these changes, and despite the development of different injection regimens and delivery systems, normalizing glycemic control still has not been adequately achieved. Recently, specific amino acid substitutions of the β-chain of the insulin molecule by recombinant DNA techniques have resulted in insulin analogs that have different pharmacokinetic properties than currently available insulins [10]. Some of these insulin analogs appear to have less variation in absorption than conventional insulins [11], and analogs have been developed with both very rapid [12] and delayed absorption [11]. Those analogs with a very rapid absorption may be useful in better mimicking the physiological insulin profiles seen postprandially than currently available native insulins [11]. All of these advances have been accomplished through protein engineering to produce insulin analogs with a decreased tendency for self-association (monomeric insulin), which in turn increases the subcutaneous absorption of these insulins [13]. These monomeric insulin analogs have been found to have a faster onset of action and are better able to reduce postprandial glucose excursions than conventional human insulin preparations, which subcutaneously become hexamers and slow absorption [14]. These insulin analogs, especially the monomeric types, appear to be less immunogenic and antigenic [13] where the pharmacokinetic aspects of human insulin and an analog insulin of aspart and lispro are described. In addition, other delivery options for insulin besides subcutaneous administration, such as transdermal and nasal approaches, are described.

## 2.2 Pharmacokinetics of Human Insulin after Intravenous Administration

In the 1970s, pork insulin had been widely used for the treatment of diabetics. In the 1980s, biosynthetic and recombinant DNA technology came to be introduced. In 1981, Halban et al. [15] compared the biologic potency of pancreatic human insulin with that of biosynthetic human insulin, which was manufactured by recombinant DNA techniques in bacteria, after intravenous injection in rats. The metabolic clearance rates of A14-mono-$^{125}$I-insulin (pork) were found to be similar to those observed for semi-synthetic [$^{3}$H]insulin (pork), being 20.8 ± 0.8 mL/min/kg and 23.6 ± 1.0 mL/min/kg, respectively. On the other hand, the metabolic clearance rates for A14-mono-$^{125}$I-biosynthetic human insulin was 24.6 ± 2.2 mL/min/kg, which was significantly higher than the value for A14-mono-$^{125}$I-pork insulin. In addition, the specific activity of the pancreatic human insulin was 32 U/mg, whereas that for biosynthetic human insulin was 27 U/mg. These values were not significantly different but were higher than the value for pancreatic pork insulin (25 U/mg). Biosynthetic human insulin is thus no less active than pork insulin in rats, and possibly somewhat more active. Moreover, in 1987, Brogden and Heel [16] reviewed that human insulin, whether produced by recombinant DNA techniques (biosynthetic, insulin crb) or enzymatic modification of porcine insulin (semi-synthetic, insulin emp) is equivalent in biological activity to porcine insulin after intravenous administration. There were no differences in the pharmacokinetic profiles among

them. However, slight differences between human and porcine insulin in hypoglycemic activity after subcutaneous injection appear to be related to differences in absorption and are unlikely to be of major clinical importance.

Turnheim and Waldhausl [17] characterized pharmacokinetic aspects of recombinant human insurin in human. After intravenous injection, the plasma concentration of insulin declines with at least two exponentials. Inslin distributes rapidly from the intravascular space to tissue compartments with a distribution half-life of 2.4 min, whereas insulin disappears slowly with an elimination half-life of 50–130 min, which is reflecting the elimination from the interstitial fluid and the tissues that use insulin. Total body clearance of insulin, which is the result of metabolic degradation, ranges between 700 and 800 mL/min. The sites of degradation are primarily the liver (hepatic insulin clearance: 320–400 mL/min) and the kidneys (renal insulin clearance: 190–270 mL/min); hence, insulin disposal depends on the function of these organs. The apparent volume of distribution for insulin is approximately equal to the extracellular space. Insulin kinetics appears not to be altered in diabetes mellitus, except in cases with insulin antibodies or in insulin-resistant patients, in which insulin removal may be retarded.

### 2.3 Pharmacokinetics of Human Insulin after Subcutaneous Administration

There is a continuing search for improved insulin formulations to imitate the physiological pattern of insulin secretion as closely as possible, and thereby to minimize the complications of diabetes mellitus. The major advances achieved to date are in the area of human insulin analog synthesis resulting from the introduction of recombinant DNA techniques and in improved delivery systems that use noninvasive or minimally invasive modes of administration. Usually, insulin is administered subcutaneously to control blood sugar levels in type 1 and type 2 diabetics. There are plenty of insulin pharmaceutics, and insulin's physico-chemical differences provide different profiles in the insulin pharmacokinetics after subcutaneous injection. Kang et al. [18], investigated absorption mechanisms of insulins with different physico-chemical properties after subcutaneous administration in humans. Based on a single-blind randomized comparison study of seven healthy male volunteers 22–43 years of age, equimolar dosages of $^{125}$I-labeled forms of soluble hexameric 2 $Zn^{2+}$ human insulin and human insulin analogs with differing association states at pharmaceutical concentrations (AspB10, dimeric; AspB28, mixture of monomers and dimers; AspB9, GluB27, monomeric) were administred. The initial fractional disappearance rates for the four insulin preparations were 20.7 ± 1.9 (hexameric soluble human insulin), 44.4 ± 2.5 (dimeric analog AspB10), 50.6 ± 3.9 (analog AspB28), and 67.4 ± 7.4%/h (monomeric analog AspB9, GluB27). The absorption of the dimeric analog was significantly faster than that of hexameric human insulin, and the absorption of monomeric insulin analog AspB9, GluB27 was significantly faster than that of dimeric analog AspB10. There was a negative linear correlation between association state and the initial fractional disappearance rates ($r = -0.98$, $p < 0.02$). A log-linear scale analysis for the elimination of insulin showed that only the monomeric analog had a monoexponential elimination. On the other hand, two phases in the rates of absorption were identified for the dimer and three for hexameric human insulin. The fractional disappearance rates calculated by log-linear

regression analysis were monomer 73.3 ± 6.8%/h; dimer 44.4 ± 2.5%/h from 0 to 2 h and 68.9 ± 3.5%/h from 2.5 h onward; and hexameric insulin 20.7 ± 1.9%/h from 0 to 2 h, 45.6 ± 5.0%/h from 2.5 to 5 h, and 70.6 ± 6.3%/h from 5 h onward. The author concluded that the lag phase and the subsequent increasing rate of subcutaneous soluble insulin absorption can be explained by the associated state of native insulin in pharmaceutical formulation and its progressive dissociation into smaller units during the absorption process.

## 2.4 Pharmacokinetics of Rapid-acting Insulin Analog after Subcutaneous Administration

By the end of the 1980s, to accommodate postprandial hyperglycemia, monomeric insulin formulations, insulin lispro (the Lys-Pro analog) and insulin aspart (the Asp-Pro analog) had been developed for clinical use [19, 20] (Figure 1).

Formulations including insulin lispro have a rapid rate of absorption after subcutaneous injection and, therefore, have to be administered at meal time. Their residence time is also about two fold shorter than regular human insulin, minimizing the risk of the excessive hypoglycemic effect that characterizes regular human insulin formulations. Becker et al. compared the pharmacokinetics and pharmacodynamics of insulin lispro with regular human insulin after subcutaneous injection in obese subjects in a single-dose, randomized, double-blind, cross-over euglycemic clamp study [21]. Insulin lispro had more rapid-acting profiles than regular human insulin. Fractional glucose infusion rate (GIR)-area under the curve (AUC) of the GIR curve and maximum GIR were greater for insulin lispro compared with regular human insulin. Time to 20% (early glucose disposal) and 80% (bulk of activity) of total GIR-AUC were shorter for insulin lispro than regular human

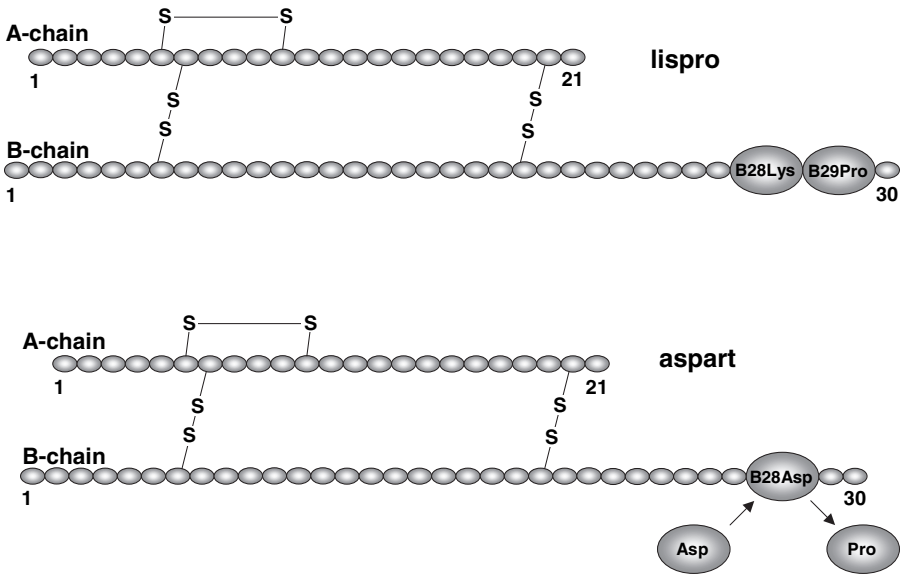


**Figure 1.** Schematic presentation of primary structure of human insulin lispro and aspart.

insulin, which was corroborated by more rapid and shorter residing pharmacokinetic profiles of insulin lispro than regular human insulin, evidenced by shorter times to 20% of total AUC of insulin, $C_{max}$ of insulin, and mean residence time. On the other hand, Osterberg et al. compared the pharmacokinetics and pharmacodynamics of insulin aspart with regular human insulin in healthy subjects in a cross-over trial. There were statistically significant differences between most of the human insulin and insulin aspart pharmacokinetic parameters, including the sigmoidicity of the time-course of absorption (1.5 for human insulin vs. 2.1 for insulin aspart), elimination rate constant ($0.010\,min^{-1}$ vs. $0.016\,min^{-1}$, respectively). In addition, the pharmacodynamic model parameters were mostly not different, except for the rate of insulin action ($0.012\,min^{-1}$ vs. $0.017\,min^{-1}$, respectively) [20]. Moreover, Homko et al. compared insulin levels and actions in patients with type 1 diabetes after subcutaneous injection of insulin analogs, aspart and lispro [22]. Both insulin analogs produced similar serum insulin levels (250–300 pmol/l) at approximately 30 min and disappeared from serum after approximately 4 h. Insulin aspart and lispro had similar effects on glucose and fat metabolism. Effects on carbohydrate metabolism (glucose uptake, glucose oxidation, and endogenous glucose production) peaked after approximately 2–3 h and disappeared after approximately 5–6 h. Effects on lipid metabolism (plasma free fatty acid, ketone body levels, and free fatty acid oxidation) appeared to peak earlier (at approximately 2 h) and disappeared earlier (after approximately 4 h) than the effects on carbohydrate metabolism. The author concluded that both insulin aspart and lispro are indistinguishable from each other with respect to blood levels and that they are equally effective in correcting abnormalities in carbohydrate and fat metabolism in patients with type 1 diabetes.

### 2.5 Other Formulations of Human Insulin

***Inhaled Insulin.*** Subcutaneous insulin has been used to treat diabetes since the 1920s; however, despite a number of different formulations, intensive insulin therapy with multiple daily injections has not gained widespread clinical acceptance. Great efforts have been made in searching for noninvasive administration modes of insulin that will avoid the need for parenteral administration of subcutaneous injection. Attempts to find effective, well-tolerated, nonenteral routes for delivering insulin began in the 1920s, and, over the years, have included ocular [23], buccal [24], rectal [25], vaginal [26], oral [27], nasal [28] and uterine [29] delivery systems, where various enhancers or formulations have been tested to increase the bioavailability of insulin from each route. Until recently, it was believed that insulin delivered noninvasively was associated with a bioavailability that was too low to offer a realistic clinical approach. However, progress in the pharmaceutical approach of insulin suggests that inhaled insulin is an effective, well-tolerated, noninvasive alternative to subcutaneous regular insulin.

Oral inhalation of insulin potentially offers noninvasive treatment and better glycemic control in diabetes by virtue of its apparently faster absorption into the systemic circulation compared with subcutaneous injection. However, the pharmacokinetics of inhaled insulin in the human lung has yet to be fully clarified because of the complexity of insulin-glucose physiology and the difficulty in approximating the inhaled dose. In 2004, Sakagami conducted a meta-analysis of

insulin pharmacokinetics in the lung after oral inhalation of insulin in humans [30], where, based on the data from well-controlled clinical studies of inhaled insulin published in the literature, a physiologically realistic insulin-glucose kinetic model was derived. The model defined the first-order absorption ($k_{a,L}$) and parallel non-absorptive loss ($k_{mm,L}$), the latter primarily occurring via metabolism and mucociliary clearance in the lung with two systemic compartments. The significant estimate for $k_{a,L}$ was found to be from 0.020 to 0.032 $h^{-1}$, effectively unchanged across doses (from 0.3 to 1.8 IU/kg), formulations (powder vs. liquid), and subjects (healthy vs. diabetic), suggesting passive diffusive absorption of insulin from the lung. In contrast, the values for $k_{mm,L}$ were much larger (from 0.5 to 1.6 $h^{-1}$) and decreased with increasing inhaled dose. From this perspective, the author considered that a dose-dependent saturable lung metabolism regulates the value of $k_{mm,L}$, alongside mucociliary clearance. As a result, the absolute bioavailability of insulin through the oral inhalation ranged from 1.5% to 4.8%.

Henry et al. [31] evaluated the AERx insulin Diabetes Management System (AERx iDMS) to the deep lung for systemic absorption, which was an aerosol of liquid human insulin produced by Aradigm (Hayward, CA). This study examined the effects on pulmonary function, pharmacokinetics, and pharmacodynamics of inhaled insulin in asthmatic and nonasthmatic subjects without diabetes, where a total of 28 healthy and 17 asthmatic subjects were enrolled in a two-part, open-label trial. The area under the time versus concentration curve (AUC) of insulin after inhalation showed significantly greater value for healthy subjects than for asthmatic subjects, whereas no difference was observed for maximum concentration in the two groups. In addition, a greater reduction of serum glucose level was observed in healthy subjects. Moreover, asthmatic subjects had greater intra-individual variations in insulin AUC, maximum concentration values, and glucose levels than healthy subjects. In this report, the author concluded that, after inhaling insulin using the AERx iDMS, asthmatic subjects absorbed less insulin than healthy subjects, resulting in less reduction of serum glucose.

The rationale behind developing a pulmonary drug delivery system is to ensure that insulin powder is delivered deep into the lungs, where it is easily absorbed into the bloodstream, in a handheld inhalation device. More recently, Nektar Therapeutics (formerly Inhale Therapeutic Systems) has developed a pulmonary drug delivery system for insulin [32]. This device converts the insulin powder particles into an aerosol cloud for the patient to inhale. The inhaler requires no power source, and the clear chamber ensures that the patient knows immediately when all the insulin has been inhaled. In addition, Nektar Therapeutics is using its Advanced PEGylation technology to develop a dry powder-inhaled polyethylene glycol (PEG) formulation for delivering peptides efficiently across the lungs and to promote prolonged serum concentration of the peptide. PEGylation is designed to increase the size of the active molecule and ultimately improve drug performance by optimizing pharmacokinetics, increasing bioavailability, and decreasing immunogenicity and dosing frequency. The investigation has begun with inhaled, long-acting (PEGylated) insulin, and is funded by Pfizer.

***Intranasal Insulin.*** Insulin administered nasally has considerable potential for the treatment of both type 1 and type 2 diabetes. Using enhancers or novel delivery systems such as adequate bioadhesive microspheres, it is believed that the low bioavailability of simple formulations of insulin can be greatly improved [33].

Therefore, several efforts on developing the nasal delivery system of insulin with nontoxic and nonirritant properties have been done.

In 1992, Drejer et al. [31] investigated the pharmacokinetics of intranasal insulin containing a medium-chain phospholipid (didecanoyl-L-alpha-phosphatidylcholine) as absorption enhancer in 11 normal volunteers. Intranasal insulin was absorbed in a dose-dependent manner with a mean plasma insulin peak $23 \pm 7$ min after administration. Mean plasma glucose nadir was seen after $44 \pm 6$ min, 20 min after intravenous injection. Moreover, intranasal administration of insulin resulted in a faster time-course of absorption than subcutaneous injection, and the bioavailability for the nasal formulation was 8.3% relative to an intravenous bolus injection when plasma insulin was corrected for endogenous insulin production estimated by C-peptide.

On the other hand, there are negative opinions on the nasal delivery of insulin. In 1995, Hilsted et al. evaluated metabolic control and safety parameters (hypoglycemia frequency and nasal mucosa physiology) in 31 insulin-dependent diabetic patients treated with intranasal insulin in an open, cross-over randomized trial [35]. Serum insulin concentrations increased more rapidly and decreased more quickly during intranasal as compared with subcutaneous insulin administration. Metabolic control deteriorated, as assessed by hemoglobin A1c concentrations, slightly but significantly after intranasal as compared with subcutaneous insulin therapy. The bioavailability of intranasally applied insulin was low; therefore, intranasal insulin doses were approximately 20 times higher than subcutaneous doses. The frequency of hypoglycemia was similar during intranasal and subcutaneous insulin therapy, and nasal mucosa physiology was unaffected after intranasal insulin. The author concluded that because of low bioavailability and a high rate of therapeutic failure, intranasal insulin treatment was not a realistic alternative to subcutaneous insulin injections.

In the early twenty-first century, nanoparticles based on nanotechnology have altered the direction of investigation in the drug delivery system. Nanoparticles, an evolvement of nanotechnology, are increasingly considered as a potential candidate to carry therapeutic agents safely into a targeted compartment in an organ, particularly a tissue or a cell [36]. In 2002, Dyer et al. investigated effects of chitosan-formulated nanoparticles on the insulin pharmacokinetics and pharmacodynamics as a nasal drug delivery system [37]. The nasal absorption of insulin after administration in chitosan nanoparticle formulations and in chitosan solution and powder formulations was evaluated in rats and in conscious sheep. Insulin-chitosan nanoparticle formulations produced a pharmacological response in the two animal models, although in both cases the response in terms of lowering the blood glucose levels was less than that of the nasal insulin-chitosan solution formulation. The insulin-chitosan solution formulation was found to be significantly more effective than the complex and nanoparticle formulations. The authors concluded that chitosan nanoparticles did not improve the nasal absorption.

## 3 ERYTHROPOIETIN

### 3.1 Background

Erythropoietin (EPO), a glycoprotein hormone with a molecular mass of 30 KDa, is mainly synthesized by the kidney in response to hypoxia. Serum EPO

concentrations in humans normally range from 6 to 32 U/L [38]. The plasma half-life of EPO is reported to range from 2 to 13 hours, with a volume of distribution close to plasma volume. As expected for a large sialoglycoprotein, less than 10% of EPO is excreted in the urine [39]. EPO is known to stimulate the proliferation and differentiation of erythrocytic progenitors in the bone marrow, leading to reticulocytosis and increased erythrocyte numbers in the blood [38].

Recombinant human EPO (rHuEPO) has proven beneficial for treating renal anemia as well as anemia of other chronic disorders, including autoimmune diseases, malignancies, and AIDS. Evidence from clinical trials indicates that anemia associated with cancer chemotherapy generally responds well to a standard rHuEPO regimen of 150 IU/kg subcutaneous (s.c.) three times per week [40, 41]. Several dosage recommendations have been made for surgery patients depending on the type of surgery and practicality of an autologous blood donation program. In patients scheduled for major orthopedic surgery, three weekly 600 IU/kg s.c. doses of rHuEPO, with the last one on the day of the surgery, have been suggested [42]. Goldberg et al. present here pharmacokinetic/pharmacodynamic (PK/PD) modeling from four comparable, placebo-controlled, parallel-group studies in healthy subjects who were administered single rHuEPO doses of 300 to 2400 IU/kg as well as multiple doses of 600 IU/kg/wk and 150 IU/kg/tiw for 4 weeks.

Clinical trials of recombinant product began in Seatle and London in 1986, and its efficacy was established at an early stage [43–46]. The early studies confirmed that intravenously administered EPO was highly effective in correcting the anemia of end-stage renal failure in hemodialysis. The intravenous route was, however, clearly impractical for patients maintained on continuous ambulatory peritoneal dialysis (CAPD) who had no ready vascular access, which necessitated an examination of the intraperitoneal and subcutaneous routes of administration, and several pharmacokinetic studies have compared these in CAPD patients [47, 48]. These three routes of erythropoietin administration are the only ones to have been investigated either from a pharmacokinetic or a therapeutic point of view. Although Storring et al. [49] and Charles et al. [50] compared the two forms of rHuEPO (i.e., EPO-alpha and EPO-beta), they reported that the volume of distribution after intravenous EPO-beta compared with EPO-alpha and that delayed subcutaneous absorption of EPO-beta compared with EPO-alpha. They believed the pharmacokinetic differences were caused by their glycosylation.

### 3.2 Pharmacokinetics of Erythropoietin in Healthy Subjects

Intravenous administration in the dialysis center has been practical and effective for patients undergoing hemodialysis [51, 52], but impractical for patients without vascular access who are receiving peritoneal dialysis. Studies of subcutaneous administration showed lower but more sustained peak plasma EPO concentrations ($C_{max}$) than occurred using the intravenous route. The efficacy of subcutaneous administration supports the concept of a sustained minimal effective EPO concentration. High $C_{max}$ values after intravenous administration appear unnecessary; the dose response relationship seen with intravenous administration most likely correlates with serum EPO concentrations maintained above a minimal effective level.

Flaharty et al. presented data on the pharmacokinetics of EPO in healthy normal subjects [53]. To assess phamacokinetics to human recombinant erythropoietin

(EPO-beta), single intravenous doses (10, 50, 150, and 500 IU/kg) were administered at monthly intervals to 16 healthy subjects in a two-panel, placebo-controlled, double-blind ascending-dose trial. A 1000-IU/kg dose was subsequently administered in an open manner. EPO concentrations were detemined in serum by radioimmunoassay. Mean EPO apparent half-lives ranged fiom 4.42 to 11.02 hours. The apparent volume of distribution was between 40 and 90 mL/kg, consistent with plasma water, and the apparent clearance values ranged from 4 to 15 mL/kg/hr, with both parameters having the highest values at the 10-IU/kg dose level. Clearance tended to decrease as a function of dose. The use of EPO in healthy individuals will have clinical relevance if EPO is found to be effective for use in autologous transfusion programs. Although EPO was well tolerated in this study, clinical trials in patients with end-stage renal disease have shown adverse effects, including hypertension, seizures, and thrombocytosis, related to the increase in erythropoiesis. These effects are not expected in normal subjects, but careful monitoring should be performed if EPO is used in repeated doses in healthy subjects.

Salmonson et al. [54] also described the pharmacokinetics of rHuEPO after intravenous and subcutaneous administration of 50 U/kg to six healthy male volunteers. The calculated mean values for volume of distribution at steady state and clearance after an i.v. dose were 76 ± 33 mL/kg and 12.0 ± 3.0 mL/h/kg, respectively. Serum concentrations of rHuEPO peaked at 13.0 ± 6.0 h after the s.c. dose and the bioavailability over 72 h was 36 ± 23%. The mean residence time and half-life of rHuEPO were 6.2 ± 1.0 and 4.5 ± 0.9 h after i.v. and 46 ± 18 and 25 ± 12 h after s.c. administration. They found that the serum concentration time profiles after i.v. adminisitaration followed a mono- or bi-exponential decline, and the elimination after s.c. dose was described by a mono-exponential decline.

McMahon et al. [55] reported a double-blind, placebo-controlled study of the pharmacokinetics and safety of multiple doses of rHuEPO 150 or 300 U/kg, either by i.v. bolus or s.c. in normal male subjects, which demonstrated that rHuEPO had a dose-related effect on the hematocrit independent of the route of administration, and that multiple doses of rHuEPO had no direct pressor effects. When rHuEPO was injected by i.v. route, a mono-exponential decrease in serum EPO level was observed for 18 to 24 hours postdose. Absorption of s.c.-injected rHuEPO occurred more slowly with relatively low serum EPO levels being maintained for 48 hours. As a result of the prolonged elevation in serum EPO concentrations with s.c. injection, with injections every 48 hours, s.c. resulted in higher preinjection serum EPO concentrations than with i.v. injection of the same dose.

Moreover, Ramakrishnan et al. [56] described the pharmacokinetic and pharmacodynamic modeling of rHuEPO after single and multiple doses in healthy volunteers. A one-compartment model with limited distribution and nonlinear elimination characterizes the data from the i.v. dosing. Most of the rHuEPO after s.c. dosing was rapidly absorbed within 2 to 3 days by zero-order rate constant. A dual-absorption model (fast zero-order and slow first-order inputs) with nonlinear disposition characterized the pharmacokinetics of s.c. rHuEPO. A similar pharmacokinetic model describes the time-course of plasma concentrations of rHuEPO on monkeys [57]. On the other hand, Cheung et al. [58] demonstrated the pharmacokinetics of rHuEPO after single s.c. doses of 300, 450, 600, 900, 1200, 1350, 1800, and 2400 IU/kg and in multiple s.c. dose regimens: 150 IU/kg three times a week for four weeks and 600 IU/kg once per week for four weeks in open-label,

randomized placebo-controlled studies in healthy volunteers. The absorption rate of rHuEPO after s.c. administration was independent of dose, whereas clearance was dose-dependent, in which it decreased with increasing dose. In an advanced investigation, Cheung et al. [59] further described the comparison of the pharmacokinetics administered s.c. once weekly (qw), 150 IU/kg, and three times weekly (tiw), 40,000 IU. The $C_{max}$ values for serum rHuEPO qw were six times and AUC values three times that of the tiw regimen. Subsequent study reported by Krzyzanski et al. [60] showed that a phamacokinetic and pharmacodynamic (PK/PD) model for rHuEPO in healthy subjects was developed to describe the time profiles of changes in serum rHuEPO and the pharmacological responses of percent reticulocytes, total red blood cell counts, and hemoglobin after single and multiple subcutaneous administration of rHuEPO. Data used is the same data as Wing Cheung et al. [59]. A dual-absorption model (fast zero-order and slow first-order inputs) with linear disposition kinetics was used to characterize the pharmacokinetics of erythropoietin after subcutaneous administration. Flip-flop kinetics was apparent in the pharmacokinetics as the absorption rate was slower (ka = 0.7 $day^{-1}$) than the elimination rate (CL/Vd = 1.2 – 9.2 $day^{-1}$).

### 3.3 Pharmacokinetics of Erythropoietin in Patients

rHuEPO is established as an effective therapeutic agent for anemia in patients undergoing long-term hemodialysis. After an intravenous dose of EPO 24 to 240 U/kg, serum clearance was 0.66 ± 0.42 L/h; terminal elimination half-life was 6 to 9 h and did not change after repeated doses given for more than two months [61]. In hemodialysis patients, the pharmacokinetics of rHuEPO was studied by Brockmöller et al. [62]. They investigated in 12 patients under chronic hemodialysis on a thrice weekly intravenous rHuEPO treatment. The pharmacokinetics of rHuEPO was also assessed after a subcutaneous injection during the initial period and during maintenance treatment. After the first i.v. dose, plasma EPO concentrations were best described by a mono-exponential disposition function with a mean (±s.d.) elimination half-life of 5.4 ± 1.7 h. The volume of distribution was 70 ± 5.2 mL/kg, and the clearance was administered intravenously as 10.1 ± 3.5 mL/h/kg. After three months of continuous therapy, the plasma half-life of rHuEPO decreased by 15% (mean half-life during steady state: 4.6 ± 2.8 h), while mean clearance and volume of distribution remained constant. On the other hand, the absorption of rHuEPO after s.c. injection was prolonged and flip-flop kinetics applied. The extent of systemic availability of the s.c. dose ranged from 28% to 100% (mean 44%).

Gladziwa et al. [63] described the pharmacokinetics of rHuEPO after long-term therapy in patients undergoing hemodialysis and hemofiltration. In 17 patients with end-stage renal disease (ESRD), studies were performed in three groups to assess pharmacokinetics during the intertreatment interval and during hemofiltration and dialysis treatment. After an intravenous bolus injection of EPO 150 U/kg, the half-life was 7.7 h, steady-state volume of distribution was 0.066 L/kg, and total plasma clearance was 5.4 mL/min. The mean steady-state serum concentration during multiple-dose administration was 656 U/L. The drug was not eliminated by hemofiltration or dialysis. Therefore, long-term treatment of ESRD patients with EPO does not significantly alter the pharmacokinetic profile of the drug. EPO dosage adjustment or substitution after hemofiltration and dialysis is not necessary. On the

other hand, peritoneal dialysis (PD) has gained recognition worldwide as an alternative to hemodialysis in the management of patients with end-stage renal disease. The most common is continuous ambulatory peritoneal dialysis (CAPD) in which most patients perform four to five exchanges daily. As the peritoneal catheter provides direct access to the peritoneum, intraperitoneal drug administration has become widely used for the administration of certain drugs. The instillation and drainage of PD fluids contribute to the total body clearance of drugs, in which peritoneal clearance is low. Taylor III et al. [67] reported the pharmacokinetic studies after rHuEPO administration to patients receiving PD. Intraperitoneal (i.p.) drug administration, via a PD catheter, produces a concentration gradient between the peritoneal and vascular compartments, favoring drug absorption. Absorption is enhanced if the plasma concentration of drug is low relative to the concentration of drug in the peritoneal fluid. Lymphatics draining the diaphragm contribute to the absorption of albumin and dextrans from the peritoneum and may be important in the absorption of protein drugs or other large-molecular-weight compounds like rHuEPO. The pharmacokinetic parameters observed in patients undergoing CAPD following i.v. and s.c. administration are similar to those determined for hemodialysis patients [65, 66]. s.c. adminisitration was found efficacious in several studies of PD patients [67–71], and allows reduction of the dose of EPO to 50–60% of the i.v. dose. i.p. administration is convenient and circumvents any pain associated with s.c. administration. In several studies, EPO was administered i.p. 1 L to 2 L of dialysate and allowed to dwell for varying time periods [72–75]. Low absorption of EPO from the peritoneal cavity is seen if dialysate is present. Hematological response similar to those observed with either s.c. or i.v. administration were seen following i.p. administration using this technique, but substantially higher doses were required [76–78]. Other investigators have reported poor efficacy, although this may result partly from the low doses administered [79, 80]. Based on these observations, the i.p. administration of EPO in large volumes of dialysate appears prohibitively costly to achieve the desired results. An improved bioavailability was observed in humans after administration of undiluted EPO 400 U/kg followed by an 8-hour dry dwell [81]. A greater than nine fold increase in bioavailability was observed with dry administration, compared with diluting the dose in 2-L dialysate followed by an 8-hour dwell. The dry-dose administration area under the concentration-time curve (AUC) approached that seen with s.c. dosing. Ateshkadi et al. [82] administered EPO 100 U/kg IP to eight patients undergoing CAPD using a 4-hour dry dwell followed by an infusion of 2-L dialysate that was allowed to dwell for 6 hours. The i.p. EPO serum-concentration time profiles were similar to those observed with s.c. administration, with a mean bioavailability of 11.4% compared with 22.8% for s.c. administration. Ongoing studies are examining the pharmacokinetics and efficacy of administration using an 8-hour dry dwell. Single-dose pharmacokinetics of EPO were studied in pediatric dialysis patients [83, 84]. EPO appears to be better absorbed and more rapidly cleared in children than it is in adults. Differences in the volume of distribution of EPO have also been noted between children and adults. These differences may be result of methodological differences between studies or true physiological differences relating to the plasma volume or tissue distribution of the drug. A recent study of i.p. administration in five children on continuous cycling peritoneal dialysis (CCPD) suggests that this route is safe and efficacious [85].

### 3.4 Erythropoietin Formulation

Humoral factors and cytokines are glycosylated with variable numbers of carbohydrate chains, each of which has variable carbohydrate structures. Although the biological function is typically determined by the protein component, carbohydrate can play a role in molecular stability, solubility, *in vivo* activity, serum half-life, and immunogenicity. The sialic component of carbohydrate, in particular, can extend the serum half-life of protein therapeutics. One aspect of glycoengineering is to introduce N-linked glycosylation consensus sequences into desirable positions in the peptide backbone to generate proteins with increased sialic acid-containing carbohydrate, thereby increasing *in vivo* activity because of a longer serum half-life [86].

DarbEPO alfa (DA) is a hyperglycosylated analog of recombinant human erythropoietin (rHuEPO) generated by glycoengineering. This rHuEPO is used to treat the anemia associated with chronic kidney disease and cancer. The discovery that the sialic acid-containing carbohydrate content of rHuEPO was directly proportional to the serum half-life and *in vivo* bioactivity [87] led to the hypothesis that additional carbohydrate on rHuEPO might further enhance *in vivo* activity. DA contains two additional N-linked glycosylation sites and has increased *in vivo* activity and prolonged duration of action because serum half-life is increased three fold above rHuEPO [88, 89], which allows for less frequent dosing with subsequent increased convenience for the patient and caregiver and improved patient compliance.

Allon et al. [90] investigated the pharmacokinetics of darbEPO alfa and rHuEPO after repeated intravenous dosing in patients with chronic kidney disease receiving hemodialysis. Overall, 47 patients were randomized to receive darbEPO alfa administered once weekly or three times weekly or EPO administered three times weekly for up to 52 weeks. Pharmacokinetic profiles were measured during weeks 1 and 12. The terminal half-life of darbepoietin alfa was 17.8 hours, which was approximately two to three times longer than that of EPO (6.3 hours). The clearance of darbepoietin alfa (2.00 mL/h/kg) was approximately four times slower than those of EPO (8.58 mL/h/kg). They conclude that darbepoietin alfa can be administered-less frequently than EPO in patients with chronic kidney disease receiving hemodaialysis.

Lerner et al. [91] also studied the pharmacokinetics of darbepoietin alfa in pedisatric patients with chronic kidney disease (CKD). Twelve patients 3–16 years of age with CKD were randomized and received a single 0.5 μg/kg dose of darbepoietin alfa administered i.v. or s.c. After a 14- to 16- day washout period, patients received an identical dose of darbepoietin alfa by the alternate route. After i.v. administration, the mean clearance of darbepoietin alfa was 2.3 mL/h/kg with a mean terminal half-life of 22.1 h. After s.c. administration, absorption was rate limiting, with a mean terminal half-life of 42.8 h and a mean bioavailability of 54%. Comparison of these results with those from a previous study of darbepoietin alfa in adult patients [89] indicated that the disposition of darbepoietin alfa administered i.v. or s.c. is similar in adult and pediatric patients, although absorption may be slightly more rapid in pediatric patients after s.c. dosing.

## 4 GRANULOCYTE-COLONY STIMULATING FACTOR (G-CSF)

### 4.1 Background

Colony stimulating factors (CSFs) as well as hematopoietic growth factors are a group of glycoproteins regulating the survival, proliferation, and differentiation of hematopoietic progenitor cells as well as the function and activation of the mature cells. The CSFs are produced by a variety of cells and range in molecular mass from 14 to 90 kDa. Recently, these factors have been purified, cloned, and produced through recombinant DNA techniques. The recombinant factors have been shown to have biologic properties and actions that are similar to the naturally occurring factors. The availability of quantities of the recombinant factors has resulted in their introduction into clinical trials and into the market.

The pharmacokinetics of CSFs has been evaluated widely in animals and in man. Early studies, however, used impure or poorly characterized fractions and the interpretation of these results is difficult. With the availability of recombinant proteins, these limitations were overcome and definitive studies have been completed. In general, the biologic effects of these factors can be quantitated by measuring the effects on the target hematopoietic cell population, which has provided a basis for relating the pharmacokinetic profile of the factor to the pharmacodynamic response. As clinical trials and the therapeutic applications of these factors are expanded, the knowledge and understanding of the pharmacokinetics of the drug will be critical to guide the clinician in the choice of dosing routes and schedule.

The generation of granulocytes from immature hematopoietic progenitor cells depends on the presence of several hormone-like glycoproteins. CSFs are acidic glycoproteins required for the survival, proliferation, and differentiation of hematopoietic progenitor cells [92–95]. Granulocyte-colony stimulating factor (G-CSF) was identified because it induced differentiation in a murine myelomonocytic leukemic cell line and stimulated granulocyte colony formation by normal progenitor cells [96]. The hematopoietic actions of this regulator are restricted exclusively to cells of neutrophilic granulocyte lineage [97–99]. Murine and human G-CSFs have been purified from various sources [96–98] and the genes encoding human and murine G-CSF have been isolated [100–102]. This factor supports the growth and proliferation of relatively late granulocyte progenitor cells already committed to the neutrophil lineage and enhances the function and activation of mature neutrophils [103, 104]. This factor has also been cloned and expressed in both bacterial cells and mammalian cells [104, 105].

Characterization of the gene encoding G-CSF has led to the production of this protein by recombinant DNA techniques. The recombinant form of human G-CSF, rhG-CSF, is also capable of supporting the formation of granulocytic colonies from committed precursor cells. The availability of large quantities of molecularly homogeneous and biologically active human G-CSF by recombinant DNA technology has made it possible to explore the use of human G-CSF as a therapeutic agent. rhG-CSF shows broad species cross-reactivity with respect to its characteristic biologic response, a rapid and dose-dependent neutrophilia. The potential clinical applications for rhG-CSF include myelorestoration after bone marrow

transplantation, mitigating chemotherapy-induced neutropenia, boosting host defense against infections, and augmenting effecter cell purpose [106].

### 4.2 Pharmacokinetics of G-CSF in Animals

Cohen et al. [107] also reported biexponential clearance of *E. coli*-derived rhG-CSF administrated intravenously to hamsters; the α and β serum half-lives were 0.5 and 3.8 hr, respectively (all data were normalized to the amount recovered at 5 min after injection). Tanaka and Kaneo [108] reported the pharmacokinetics of rhG-CSF (KRN8601) in male Sprague-Dawley rats at a dose of 5 μg/kg as a preclinical investigation. The serum concentrations of rhG-CSF were monitored using a specific sandwich enzyme immunoassay. For i.v. administration, the serum concentration-time curve showed a rapid disappearance of rhG-CSF from the systemic blood with a mean residence time (MRT) of 1.34 hr. For s.c., i.m., and i.p. administration, lower peak serum levels were observed; but after 2 to 3 hr, rhG-CSF levels were higher than those for i.v. administration with the MRTs of 3.92, 2.90, and 3.54 hr, respectively. Moreover, Tanaka et al. [109, 110] examined the influence of dose level and route of administration on the pharmacokinetics of *E. coli*-derived G-CSF after i.v. administration to Sprague-Dawley rats at doses of 10 or 100 μg/kg. Biexponential plasma clearance was noted after an i.v. dose of 100 μg/kg; the half-lives at distribution phase (α) and elimination phase (β) were 25 and 102 min, respectively. In addition, the total body clearance of rhG-CSF in the rat after i.v. administration showed an approximate 30% decrease from 80 to 56 mL/h/kg as the dose was increased from 1 to 10 μg/kg; no further decreases in the total body clearance were noted up to 100 μg/kg. However, the volume of distribution calculated was equivalent at each dose. For the serum concentration profiles of rhG-CSF after extravascular administration, an i.m. injection gave slightly faster absorption kinetics of rhG-CSF from the injection site into systemic blood than did s.c. and i.p. injections [108]. After s.c. administration, biological effects were noted at 2 h and were sustained for 36 h, suggesting rapid absorption. Peak serum concentrations were seen at approximately 2–3 h postdose. The calculated bioavailability of the subcutaneous dose was approximately 78%. There were nonproportional increases in the AUC value; a 100-fold increase in the dose resulted in a 164-fold increase in the AUC. Bioavailability after subcutaneous dosing ranged from approximately 50% to 70%. The results of this study suggested that two mechanisms may be involved in the clearance of G-CSF. One pathway is saturated at low concentrations and represents high-affinity, low-capacity receptor-mediated uptake by target cells (neutrophils). The other mechanism represents clearance by nonspecific pathways such as renal excretion.

In multiple-dose studies, rhG-CSF was injected into animals by i.v. and s.c. at 5 μg/kg/day for 7 days [108]. When the serum concentration-time profiles on day 7 after multiple dosing of rhG-CSF were compared with single dosing, the AUC after i.v. multiple dosing decreased by 17.4%. However, the half-lives and the volume of distribution were not significantly different between single and multiple dosing groups. In addition, the AUC after s.c. multiple dosing decreased by 25.6% without significant differences in the bioavailability and observed maximum serum concentration of rhG-CSF. These results showed that the clearance of rhG-CSF increased after multiple dosing, although the mechanism of increased clearance was not apparent.

Moreover, Tanaka et al. reported the effect of sex differences and hepatic or renal failure on the pharmacokinetics of rhG-CSF in the rats. After i.v. and s.c. administration of rhG-CSF to male and female Sprague-Dawley rats at a dose of 5 and 100 μg/kg, area under the concentration versus time curve and elimination half-lives of rhG-CSF in female rats were smaller than those for male rats. The volume of distribution of rhG-CSF in female rats was not significantly different from that in male rats. After s.c. administration, the values of AUC, MRT, and half-lives of elimination phase in female rats were smaller than those for male rats. In the *in vitro* biological activities of rhG-CSF using [$^3$H]thymidine uptake assay in cultures of bone marrow cells obtained from male and female rat femur, female rat bone marrow cells showed a similar dose-response profile to rhG-CSF to that of male rat bone marrow cells. The effect of rhG-CSF administration in rats was a specific activity on the neutrophil lineage with an increase of neutrophils in peripheral blood. On the other hand, the *in vivo* effects of rhG-CSF after i.v. and s.c. adrninistration to male and female rats at 5 and 100 μg/kg doses, namely the neutrophil count in female rats, was similar to that in male rats in the early period; however, the neutrophil count in female rats was lower than that in male rats 24 h after administration. These results indicate the sex differences in pharmacokinetics and the *in vivo* effects of rhG-CSF as well as the close relationship between pharmacokinetics and the *in vivo* effects of rhG-CSF [111]. In rats with experimental renal and hepatic failure, the total body clearance of rhG-CSF after i.v. administration at a dose of 10 μg/kg was 44.48 mL/h/kg in sham-operated rats compared with 9.429 mL/h/kg in bilaterally nephrectomized rats. In sham-operated rats, the elimination half-life of rhG-CSF at β phase was 1.51 h, and it increased to 5.33 h after nephrectomy. In contrast, the volumes of distribution were identical in both rats. In rats with acute renal failure induced by uranyl nitrate, the total body clearance and volume of distribution were identical to those of control rats, but the elimination half-life at β phase was slightly shorter. In 70% hepatectomized rats, the clearance of rhG-CSF decreased from 42.08 mL/h/kg to 31.93 mL/h/kg, and similar half-lives were observed in rats in both the sham-operated and hepatectomized groups. However, the volume of distribution decreased after hepatectomy. In rats with hepatic failure induced by $CCl_4$, the pharmacokinetic changes were similar to those observed in hepatectomized rats. These results suggest that renal clearance makes a major contribution to total body clearance compared with hepatic clearance [112].

### 4.3 Pharmacokinetics of G-CSF in Humans

In the early period, the pharmacokinetics of *E. coli*-derived rhG-CSF as part of phase I and II clinical trial was evaluated by Layton et al. [113]. The protein was administered by i.v. or s.c. routes at dose levels of 3 to 30 μg/kg. The elimination half-life after i.v. dosing was 1.4 h at a dose of 1 μg/kg, and increased to 4.2 h at a dose of 60 μg/kg. These results indicated saturation mechanisms in the total body clearance at doses greater than 10 μg/kg. When a continuous s.c. infusion was performed for 5 days, steady state was not achieved and there was a rapid reduction in serum G-CSF levels during the last 2 days of administration. Neutrophil levels increased during this period, and these results were interpreted as neutrophil receptor-mediated regulation of G-CSF uptake and degradation. The peak levels

of G-CSF after initiation of s.c. dosing were achieved at 6 h, and these levels were maintained for up to 16 h. When high doses of melphalan were coadministered with G-CSF, the neutrophil levels did not increase and serum G-CSF levels remained constant.

For the pharmacokinetics of rhG-CSF in humans, it has been reported that the absorption and clearance of rhG-CSF follow first-order kinetics without any apparent concentration dependence [114]. When rhG-CSF was administered by 24-h constant i.v. infusion at a dose level of 20 μg/kg, the mean serum concentration achieved 48 ng/mL. Constant i.v. infusion for 11 to 20 days produced steady-state serum concentrations over the infusion period. Subcutaneous administration of rhG-CSF at doses of 3.45 and 11.5 μg/kg resulted in peak serum concentrations of 4 and 49 ng/mL, respectively. The mean value of volume of distribution was 150 mL/kg. The elimination half-life was 3.5 h after either i.v. routes or s.c. routes, with a clearance rate of 0.5–0.7 mL/min/kg. The administration of a daily dose for 14 consecutive days did not affect the half-life.

More recently, Hayashi et al. reported the results of pharmacokinetic and pharmacodynamic (PK/PD) analysis for rhG-CSF (lenograstim) after s.c. administration [115]. A total of 72 adult healthy volunteers were administered 1 μg/kg of rhG-CSF. There was no correlation between the maximal plasma concentration ($C_{max}$) and an increase in peripheral neutrophil count, whereas a negative correlation between AUC and this increase was observed. It is considered that the mechanism of this phenomenon is probably based on the correlation between the elimination rate constant (ke) and neutrophil increase, because the ke probably has a close relationship with uptake by neutrophil and its progenitor the G-CSF receptor. Indeed, a volunteer with higher ke showed a greater increase in neutrophil count. Therefore, AUC is proportional to the rhG-CSF remainder, that is, the proportion that is not consumed in the course of increasing the neutrophil count. From this perspective, the bioavailability calculated from the AUC is unlikely to indicate the absorbed amount. A two-compartment pharmocokinetic model with zero-order absorption and first-order elimination provided a good curve fit, suggesting that the absorption process is not a first-order process but is a zero-order process.

As the clearance of rhG-CSF is known to decrease with a rise in dose and is known to be saturable, the average clearance after i.v. administration will be lower than that after s.c. administration. Therefore, the apparent absolute bioavailability with subcutaneous administration calculated from the AUC ratio is expected to be a conservative estimate. In a second study by Hayashi et al., the estimation of rhG-CSF absorption kinetics after s.c. administration with a nonlinear elimination pharmacokinetic model using a modified Wagner–Nelson method was studied [116], and the results of the bioequivalency study between two rhG-CSF formulations with a dose of 2 μg/kg were described. The apparent absolute bioavailability for s.c. administration was 56.9% and 67.5% for each formulation, and the ratio between them was approximately 120%. The true absolute bioavailability was, however, estimated to be 89.8% and 96.9%, respectively, and the ratio was approximately 108%. The absorption pattern was applied to other doses, and the values of predicted clearance for s.c. and i.v. administrations were then similar to the values for several doses reported in the literature. Using this pharmacokinetic model, the underestimation of bioavailability was around 30%, and the amplification of difference was 2.5 times, from 8% to 20%, because of the

nonlinear pharmacokinetics. The neutrophil increases for each formulation were identical, despite the different bioavailabilities. Therefore, it is considered that there is an upper limit to the transfer rate of rhG-CSF from subcutaneous tissue to blood, and that the amount eliminated through the saturable process, which indicates the amount consumed by the G-CSF receptor, was identical for each formulation.

Wang et al. also reported the PK/PD relationship of the granulopoietic effects of r-metHuG-CSF (filgrastim) in healthy volunteers through a population pharmacokinetic approach [117]. Healthy male volunteers were enrolled into a four-way crossover clinical trial. Subjects received four single doses of r-metHuG-CSF (375 and 750 μg for i.v. and s.c., respectively) with an intervening washout period of 7 days, and an absolute neutrophil count (ANC) was determined. Data analysis was performed using mixed-effects modeling as implemented in the NONMEM software package. The final PK/PD model incorporates a two-compartment PK model with bisegmental absorption from the s.c. site, first-order and saturable elimination pathways, and an indirect PD model. A sigmoidal $E_{max}$ model for the stimulation of ANC input rate ($k_{in}$) was superior to the conventional $E_{max}$ model ($E_{max}$ = 12.7 ± 1.7; $EC_{50}$ = 4.72 ± 0.72 ng/mL; Hill = 1.34 ± 0.19). In addition, a time-variant scaling factor for ANC observations was introduced to account for the early transient depression of ANC after r-metHuG-CSF administration. The absolute bioavailability after s.c. administration of r-metHuG-CSF was estimated to be 0.619 ± 0.058 and 0.717 ± 0.028 for 375 μg and 750 μg s.c. doses, respectively. The time profiles of concentration and ANC, as well as the concentration approximate ANC relationship of r-metHuG-CSF in healthy volunteers, were well described by the developed population PK/PD model.

On the other hand, Hernandez-Bernal et al. evaluated the equivalence of the pharmacokinetic, pharmacodynamic, and safety properties of two commercially available recombinant G-CSF formulations [Hebervital (Heber Biotec, Havana, formulation A) and Neupogen (Hofman-La Roche S.A, formulation B)] in 24 healthy male volunteers using a standard two-way randomized crossover double-blind study, with a 3-week washout period [118]. A single 300-μg G-CSF dose was administered subcutaneously. Absolute neutrophils (ANC), white blood cells (WBC), and $CD3^{4+}$ cell counts were the pharmacodynamic variables measured up to 120 h. Other clinical and laboratory determinations were used as safety criteria. The pharmacokinetic parameters for formulation A and B were very close to each other (i.e., AUC, 235.9 vs. 270.0 ngh/mL; $C_{max}$, 29.2 vs. 33.4 ng/mL; $T_{max}$, 4.2 vs. 4.7 h; half-life, 3.2 vs. 2.8 h; CL, 260.9 vs. 277.2 mL/h; Vd, 1.2 vs. 1.1 L; and MRT, 7.58 vs. 7.38 h). The pharmacodynamics showed high similarity because ANC and WBC had the same profiles for both products and no differences were detected for the estimated parameters. The $CD3^{4+}$ cell count increments were evident for both formulations in a similar way as well. According to the overall results, these formulations could be considered as clinically comparable.

### 4.4 G-CSF Formulation

Generally, peptide/protein drugs undergo extensive hydrolysis in the gastrointestinal tract and have short circulating half-lives in the blood. For optimal clinical efficacy, therefore, they must be given by daily injections. Recently, it has been

found that attaching a polyethylene glycol (PEG) moiety (PEGylation) improves the pharmacokinetic and pharmacodynamic profiles of peptide/protein drugs [119]. Conjugating peptide/protein drugs with PEG, a process known as pegylation, is now an established method for increasing the circulating half-life of protein and liposomal pharmaceuticals. PEGs are nontoxic water-soluble polymers that, owing to their large hydrodynamic volume, create a shield around the pegylated drug, thus protecting it from renal clearance, enzymatic degradation, and recognition by cells of the immune system [120].

For the formulation of G-CSF, PEGylated filgrastim (pegfilgrastim) has a longer half-life than unmodified filgrastim and, when administered as a single dose, has been shown to be at least as efficacious as daily filgrastim. As a result of its reduced renal clearance, the elimination of pegfilgrastim is predominantly neutrophil-mediated, so its clearance is self-regulated. In addition, pegfilgrastim can be administered at a fixed dose instead of in weight-based doses. For these reasons, pegfilgrastim can potentially increase patient adherence and acceptance of treatment, thus having a beneficial effect on their quality of life. Furthermore, treatment is likely to cost less because of the reduced need for medical interventions. Similarly, pegylated liposomal doxorubicin has a longer half-life than unmodified doxorubicin and has been shown, because of its reduced reticuloendothelial system clearance, to produce higher concentrations of doxorubicin in tumors and to have greater clinical efficacy than doxorubicin in the treatment of some solid tumors. Pegylated liposomal doxorubicin is also associated with less myelosuppression and febrile neutropenia.

For the pegylated formulation for subcutaneous route, van Der Auwera et al. produced a pegylated G-CSF formulation, Ro25-8315, which is produced by conjugation of rhG-CSF mutant with PEG [121], where the pharmacodynamics and pharmacokinetics of Ro25-8315 was evaluated in comparison with rhG-CSF (filgrastim). Subjects received single subcutaneous doses of Ro25-8315 ranging from 10 to 150 μg under a double-blind, randomized, placebo-controlled design. Filgrastim was administered as a single dose (5 or 10 μg/kg) and, following a 14-day washout period, daily for 7 days. Ro25-8315 increased absolute neutrophil count (ANC) by six- to eight-fold and $CD3^{4+}$ cell count more than 30-fold at the highest doses tested. Single doses (60–150 μg) of Ro25-8315 and multiple doses of filgrastim had similar effects on ANC and $CD34^{+}$. The phamacokinetic profile of Ro25-8315 was dose-dependent, with peak concentrations and the AUC increasing 100-fold over the range of doses studied. Time to reach peak concentration ($T_{max}$) and half-life of Ro25-8315 averaged 20–30 h at all doses, approximately three times longer than with filgrastim. Adverse events were not serious and occurred with similar frequency with both products. Pegylation of rhG-CSF mutant results in more desirable pharmacokinetic properties and a longer duration of action with effective increases in ANC and measures of peripheral blood progenitor cell mobilization for at least 1 week.

More recently, Maeda et al. developed a new sustained-release formulation of rhG-CSF [122]. The formulation is a double-layer minipellet (DL-MP) that is designed to maintain a sustained release of rhG-CSF without an initial burst. The DL-MP is composed of a core of collagen matrix and a coating layer with collagen. This formulation of rhG-CSF was then prepared, and its characteristics were determined, in normal rats. After subcutaneous administration, it was found that blood rhG-CSF concentration was maintained for about 1 week. Moreover, after admin-

istration of the DL-MP rhG-CSF formulation with additional condroitin sulfate, a persistent increase in white cell count was found. Therefore, the DL-MP formulation system was useful as a long-acting formulation of rhG-CSF characterized by excellent long-acting properties without an initial burst.

In general, an oral route is most popular form of administration; however, peptide/protein drugs without some pharmaceutical modifications should not be applicable because of poor absorption based mainly on extensive hydrolysis in the gastrointestinal tract. Taking this background into consideration, Eiamtrakarn et al. designed a new formulation of G-CSF for oral administration [123], where a new gastrointestinal mucoadhesive patch system (GI-MAPS) has been designed for the oral delivery of G-CSF in beagle dogs. The system consists of four layered films, $3.0 \times 3.0\,mm^2$, contained in an enteric capsule. The 40-μm backing layer is made of a water-insoluble polymer, ethyl cellulose (EC). The surface layer is made of an enteric pH-sensitive polymer such as hydroxypropylmethylcellulose phthalate (HP-55), Eudragit L100 or S100, and was coated with an adhesive layer. The middle layer, a drug-containing layer, made of cellulose membrane is attached to the EC backing layer by a heating press method. Both drug and pharmaceutical additives including an organic acid, citric acid, and a non-ionic surfactant, polyoxyethylated castor oil derivative (HCO-60), were formulated in the middle layer. The surface layer was attached to the middle layer by an adhesive layer made of carboxyvinyl polymer (Hiviswako103). The same three kinds of GI-MAPSs containing 125 μg of rhG-CSF were prepared and orally administered to dogs, and the increase in total white blood cell (WBC) counts were measured as the pharmacological index for G-CSF. Comparison with the total increase of WBCs after i.v. injection of the same amount of G-CSF (125 μg) indicated the pharmacological availabilities (PA) of G-CSF were 23%, 5.5%, and 6.0% for Eudragit L100, HP-55, and Eudragit S100 systems. By decreasing the amount of HCO-60 and citric acid, the PA of G-CSF decreased. These results suggest the usefulness of GI-MAPS for the oral administration of protein.

For pulmonary drug delivery, the opportunity for systemic administration of peptides and proteins that are, at present, usually administered parenterally has been available [124]. As a result of the large surface area, good vascularization, immense capacity for solute exchange, and ultra-thinness of the alveolar epithelium are unique features of the lung, which can facilitate systemic delivery via pulmonary administration of peptide/protein drugs. Physical and biochemical barriers, lack of optimal dosage forms, and delivery devices limit the systemic delivery of biotherapeutic agents by inhalation. Current efforts exist to overcome these difficulties to deliver metabolic hormones. Systemic delivery of rhG-CSF can readily induce an increase in circulating levels of natural GCSF to approximately three to five times greater than baseline. Niven et al. [125] showed that rhG-CSF induced systemic response after delivery by aerosol in hamsters. The absorption from the lung was rapid, with a concomitant increase in white blood cells to four times baseline. The bioavailability was 45.9% of the administered dose and 62.0% of the dose reached the lung lobes. In a study that compared pulmonary administration of rhG-CSF powder with solution [126], a normal systemic response was obtained, indicating that rhG-CSF retains its activity in the solid state after formulation. Dissolution and absorption of rhG-CSF from powders were not rate limiting because the plasma concentration versus time profiles peaked at similar times in both powder and solution administration.

More recently, Miyamoto et al. developed the *in vivo* nasal absorption system of the FITC-dextrans with a mean molecular weight ranging from 4.3 to 167 kDa and rhG-CSF [127], in which the effect of poly-L-arginine (poly-L-Arg) on the *in vivo* nasal absorption in rats were studied. When FITC-dextrans were coadministered intranasally with 1.0 w/v% poly-L-Args of different molecular weight [MW, ca. 45.5 and 92 kDa, poly-L-Arg (50) and poly-L-Arg (100)], the bioavailability increased markedly compared with that after administration of FITC-dextran alone. However, the bioavailability decreased exponentially with the increasing molecular weight of FITC-dextrans. There was no significant difference between the enhanced nasal absorption of FITC-dextrans achieved by the coadministration of poly-L-Arg (50) and poly-L-Arg (100). Moreover, the relationship between the bioavailability and the molecular weight of FITC-dextrans indicated that the molecular weight of protein drugs, which exhibited efficient absorption with poly-L-Arg, was about 20 kDa, whereas the lower limit of bioavailability for developing a potent transnasal delivery system was assumed to be about 10%. Indeed, the nasal absorption of rhG-CSF, which has a molecular weight of 18.8 kDa, was also increased after coadministration of 1.0 w/v% poly-L-Arg (50) and the bioavailability was about 11%. It seems likely that poly-L-Arg can be used to provide adequate nasal absorption of various protein drugs, which have a molecular weight of about 20 kDa, thereby allowing the successful development of a variety of transnasal drug delivery systems.

## 5 INTERFERON

### 5.1 Background

Interferons (IFNs) are classified into three major groups that reflect antigenic and structural differentiation. IFN-α (leukocytic interferon) is produced by B lymphocytes, null lymphocytes, and macrophages; its production can be induced by foreign, virus-infected, tumor or bacterial cells [128]. Human IFN-α is a family of more than 15 subtypes, with molecular masses of 17.5–23 kDa. These proteins are composed of 165–166 amino acid residues and show approximately 80% sequence homology. The secondary structure of human IFN-α consists of 55–70% α-helix and less than 16% β-sheet. In general, human IFN-α does not contain N-linked glycosylation sites. There are 13 species of subtype IFN-α, and each IFN-α subtype has a different affinity for IFN receptor or a different antiviral activity [129]. These IFNs tend to be acidic, with isoelectric points of 5.7–7.0. In commercial settings, native IFN-α (Ly-IFN-Ly) and recombinant IFNs (IFN-α 2a, IFN-α 2b, and IFN-α con) are available [130]. IFN-β (fibroblast interferon) is a product of fibroblasts, epithelial cells, and macrophages that can be induced by viral and other nucleic acids. Human IFN-β is a single species with the molecular size of the natural glycoprotein 23 kDa with 166 amino acid residues. IFN-β shows 29% structural homology to IFN-α. The secondary structure consists of 36% α-helix and 33%-β sheet. The isoelectric point falls between 6.8 and 7.8. Now, native IFN-β and recombinant IFN-β, IFN-β 1b are available [131]. IFN-γ (immune interferon) is a product of activated T lymphocytes stimulated by foreign antigens and is composed of 143 amino acids. Human IFN-γ is heterogeneous, the molecular size depending on glycosylation and possibly oligomerization. Higher molecular sizes of 50–70 kDa

have been reported, suggesting oligomerization. It shows no statistically significant sequence homology to IFN-α or IFN-β. IFN-γ is acid-labile and highly basic, with an isoelectric point of 8.6–8.7 [132].

A lot of works involving IFNs has been done with natural or cloned human IFN preparations. In the 1970s, human testing of IFNs was restricted by their limited availability, purity, and specific activity; however, recombinant DNA technology has provided high-purity (>99%) and high-specific-activity preparations for large-scale human testing [133]. Human leukocyte interferon was the first preparation available. Early phase I studies, intending to establish the safety, tolerance, pharmacokinetics, and efficacy of interferons, targeted a host of viral and oncologic diseases using animal species [134–136]. In clinical use, the results were often disappointing. However, the importance of the route and rate of administration as treatment variables had been identified. On the other hand, the pharmacokinetics of conventional drugs commonly demonstrates species specificity. In case of IFNs, however, the catabolism and excretion of IFNs are similar across most species. Therefore, animals may serve as suitable models for the pharmacokinetics of IFNs in humans.

### 5.2 Pharmacokinetics of Interferon in Animals

Absorption of IFN-α from intramuscular [137, 138], subcutaneous [139, 140], intraperitonea1 [141], intradermal [142], duodenal, and rectal [143] sites has been reported. In general, IFN-α absorption from these sites is prolonged, and maximum serum concentrations occur 1–6 h post injection, followed by measurable concentrations through 8–24 h post injection. The concentration-time profile appears to be independent of the purity or source of IFN-α, namely, partially purified, natural, or recombinant. In addition, the disposition profiles were similar across the tested species, including mice, rabbits, dogs, and monkeys. Several of these studies have determined the absolute bioavailability from the intramuscular site to be 42% in dogs [139] and 93% [144] and 56% [136] in monkeys.

The absorption of IFN-β from intramuscular, subcutaneous, and intraperitneal sites is similar to that observed for IFN-α. The concentration-time curves are also similar to those seen with IFN-α, namely, prolonged absorption with concentrations persisting from 9 to 24 h after the injection. Estimates of absolute bioavailability from the intramuscular site can be obtained from published data: 60% in rats [141], 33% [145] and 60% in rabbits, and 43% in monkeys [145]. Recently, a bioavailability of 2.2% was reported for IFN-β after intranasal administration to rabbits [146]. However, absorption appears to be independent of species and interferon source. Contraly to this appearance, the absorption of IFN-γ has not been studied as extensively as that of IFN-α or -β. According to the results of Weck et al., serum concentrations of IFN-γ were very low or nondetectable after intramuscular injection of nonglycosylated IFN-γ in monkeys [147].

After bolus injection, the initial decline in serum concentrations following intravenous administration is rapid for IFN-α, -β, and -γ for all species tested with the half-lives of initial distribution on the order of minutes. Serum concentrations initially decrease rapidly, then decline more slowly, with terminal elimination half-lives ranging from minutes (30–40 min for IFN-α in mice) [148] to hours (7–12 h

for IFN-β in rabbits) [145] depending on the species selected and the type of interferon.

Both IFN-α and IFN-β have a volume of distribution after intravenous administration that ranges from 20% to 100% of body weight in mice [148], rats [149], rabbits [145], dogs [139], and monkeys [138], suggesting distribution into a volume that approximates total body water. Similar data for IFN-γ are not available. The tissue content of the interferons generally parallels that found in serum or plasma. Measurable concentrations or titers of interferons have been demonstrated in the brain, spleen, lung, liver, and kidney. For IFN-β and -γ, the amount of interferon determined in specific organs or tissues reflected the amount found in the serum or plasma, suggesting no uptake of these interferons into the sample organs or tissues. For IFN-α, the uptake results on tissue distribution studies have been mixed. Heremans et al. [141] reported the uptake of IFN-α by liver, lung, and spleen after intraperitoneal administration of homologous mouse IFN-α to mice. In contrast, Bohoslawec et al. [148] showed no appreciable uptake of murine IFN-α into brain, liver, lung, or spleen after intravenous administration to mice. However, the kidney showed an appreciable uptake of murine IFN-α. Two human IFN-α subtypes, A and D, and one human hybrid, A/D, displayed a tissue distribution profile similar to that of biologically active mouse IFN-α [148].

The catabolism of interferons has been the most widely researched area of preclinical IFN pharmacokinetics. Undoubtedly, the gastrointestinal tract is the most effective and efficient system for catabolizing proteins. The liver has also been shown to play a role in the catabolism of dycosylated proteins, and the kidney is the major catabolic organ for circulating low-molecular-mass (<60 kDa) proteins. In addition, other organs and tissues, such as the lungs and muscles, are thought to play minor roles in protein catabolism. For the catabolism of interferons, IFN-α has been the most extensively studied of the three types of interferons. In general, IFN-α is filtered through the glomeruli of the kidney followed by luminal endocytosis and proximal tubular reabsorption [150]. During reabsorption, IFN-α undergoes proteolytic degradation by lysosomal enzymes; therefore, negligible amounts of intact IFN-α are excreted in the urine. In nephrectomized animals (rats and rabbits), therefore, the clearance was significantly decreased [151, 152]. In addition to the above findings, Bocci et al. demonstrated that the liver has been shown to play a small role in the catabolism of IFN-α [153]. On the other hand, both IFN-β and IFN-γ undergo renal catabolism [153], but to a much smaller extent than IFN-α. Natural IFN-β and -γ are glycoproteins that contain sialic acid groups; they are therefore subject to the liver catabolism characteristic of this class of proteins. For recombinant IFN-α, there is an evidence that it reduces hepatic drug metabolism activity in mice [154]. Moreover, it has been demonstrated that the greater the antiviral activity of a given IFN in murine cells, the greater its ability to depress hepatic microsomal cytochrome P450 content in mice [154].

### 5.3 Pharmacokinetics of Interferon in Humans

The pharmacokinetics of interferons has been evaluated using a variety of dosing regimens varying from intermittent to continuous administration. Although definitive pharmacokinetics has not been reported in all cases, enough information is available to define their basic disposition characteristics. Serum interferon concen-

trations after intravenous administration decline rapidly in a biexponential manner for IFN-α and IFN-β [155, 156]. On the other hand, monoexponential decline has been observed for IFN-γ [157]. IFN concentrations fall several orders of magnitude over the measurable serum concentration-time course. Terminal half-lives range from 4 to 16 h for IFN-α, 1 to 2 h for IFN-β, and 25 to 35 min for IFN-γ. At the doses tested, serum concentrations are generally measurable for between 8 and 24 h after dosing of IFN-α and up to 4 h after injection for IFN-β and -γ. The total body clearance of IFN-α has been reported to range from 4.9 to 21 liters/h or 24 liters/h per/m$^2$. For IFN-β, the total body clearance is much higher, 19–38 liters/h per/m$^2$, whereas that of IFN-γ falls between the two at 12.66–32.4 liters/h.

Oral absorption of intact proteins has met with limited success because of the natural proteolytic capabilities of the gastrointestinal tract. The systemic absorption of interferons from sites other than the gastrointestinal tract has been remarkably good, considering the size of these molecules. Intramuscularly and subcutaneously administered IFN-α [158, 159] and nonglycosylated IFN-γ [160] are well absorbed, >80% for IFN-α and 30–70% for IFN-γ. These routes exhibit protracted absorption, which results in maximum serum or plasma concentrations occurring after 1 to 8 h. Concentrations are measurable for 4 to 24 h after injection for both IFN-α and IFN-γ. Maximum serum concentrations following these routes of administration are at least an order of magnitude less than the highest concentration observed after intravenous administration of an equal dose. The absorption of IFN-β from muscle or skin has not been sufficient to produce much greater serum concentrations because of the limits of assay detection.

The volume of distribution is similar for both IFN-α and IFN-γ, ranging from 12 to 40 liters [158, 161]. Although this volume is not physiological, it is approximately 20% to 60% of body weight. Information regarding IFN-β has not been established. On the other hand, the penetration of interferons across the human blood-brain barrier has been actively studied. Partially purined [162], natural [163], and recombinant [162] interferons do not readily cross the blood-brain barrier intact after intravenous, intramuscular, or subcutaneous administration. These data suggest that direct effect of parent interferon for the nervous system is not responsible for the occurrence of neurotoxicity, a common side effect of interferon therapy.

### 5.4 Interferon Formulation

Other routes of administration have been evaluated, such as inhalation, intralesional, intranasal, intraperitoneal, intraventricular, and intraocular rates, in clinical studies. For the most part, these alternative routes were attempts to improve the delivery of interferon to sites not easily accessible via the systemic circulation. These dosing strategies have provided adequate concentrations of interferon in cerebrospinal fluid, lymph, nasal mucosa, and peritoneal fluid, but have not led to clinical success, undoubtedly reflecting the lack of understanding of the inherent mechanism of interferon action [164].

The recommended twice-weekly administration of IFN-α does not maintain sustained plasma concentrations of the drug, thereby adversely affecting the potential biological response [165]. During the last few years, "pegylation technology" has been used to develop long-acting forms of IFN-α that help to avoid most of these

problems [166]. Polyethylene glycols are nontoxic water-soluble polymers that, owing to their large hydrodynamic volume, create a shield around the pegylated drug, thus protecting it from renal clearance, enzymatic degradation, and recognition by cells of the immune system. Two separate compounds, peginterferon α-2a (PEGASYS) and peginterferon α-2b (PEG-Intron), are now both approved for use alone or in combination with ribavirin for the treatment of chronic hepatitis C [167]. Interferon therapy for chronic hepatitis C is not a cure, but it is able to decrease the viral load and may decrease the risk of complications (e.g., cirrhosis, liver failure, and liver cancer) [168]. Pegylation of the interferon increases the amount of time the interferon remains in the body by increasing the size of the interferon molecule. Increasing molecule size slows the absorption, prolongs the half-life, and decreases the rate of interferon clearance [169]. Thus the duration of biological activity is increased with pegylated INF over nonpegylated INF. The pegINF-α products offer an advantage over nonpegylated INF-α products because of less frequent administration. Tolerability of the pegylated INF is comparable with the nonpegylated formulations. Monotherapy with these agents produces a better response in some patients than monotherapy with the nonpegylated formulation [170]. Combination therapy with ribavirin is more effective than monotherapy. In 1995, Nalin et al. demonstrated that conjugation of IFN-α 2a to a linear polyethylene glycol (PEG) (M, 5000) via a urea linkage increased the half-life, and plasma residence time parameters of IFN-α 2a were one and one-half fold and two fold greater than those corresponding to the native IFN-α 2a in rats [171], with reduced immunogenicity in mice [172]. In humans, a twice-weekly dosing of $PEG_{5K}$-IFN-α 2a is required to achieve antiviral levels similar to those attained with unconjugated IFN, as shown in a phase II trial involving chronic hepatitis C patients [173]. Ramon et al. reported that the conjugation of INF-α 2b (IFN-α 2b) to a branched-chain (40,000) polyethylene glycol ($PEG_{2,40K}$-IFN-α 2b) markedly improved the resistance to tryptic degradation and the thermal stability of IFN-α 2b. The serum half-life of $PEG_{2,40K}$-IFN-α 2b showed 330-fold longer, whereas plasma residence time was increased 708 times compared with native IFN [174]. Pegylation alters the pharmacokinetic properties of IFN-α and allows for once-weekly administration. Pegylated IFNs contain either linear PEG chains of small molecular weight, as in pegylated IFN-α-2b (12 kD), or larger branched moieties, as in pegylated IFN-α-2a (40 kD). There are pharmacokinetic and pharmacodynamic differences between these two IFNs. The much-increased sustained virological response rates observed with pegylated IFN-α-2a (40 kD) and pegylated IFN-α-2b (12 kD) support the rationale for pegylation of IFN [175]. The absorption half-life of standard IFN-α is 2.3 h, whereas absorption half-lives for pegylated IFN-α-2a and α-2b are 50 h and 4.6 h, respectively. The volume of distribution for pegylated IFN-α-2a is considerably restricted, whereas the volume of distribution for pegylated IFN-α-2b is only approximately 30% lower than that for conventional IFN. As a result of its large size, the 40-kD pegylated IFN-α-2a has a more than 100-fold reduction in renal clearance compared with conventional IFN-α. Clearance of pegylated IFN-α-2b is about one tenth that of unmodified IFN-α. Although data are limited, both drugs appear to show differences in the initial viral decay pattern in patients with chronic hepatitis C. However, it remains unknown whether these differences in the initial viral decline predict differences in the primary clinical endpoint, sustained virological response.

The development of new IFN-α delivery strategies is a key issue to simplify its administration and improve its therapeutic effects while reducing its dose-related

side effects. One of the most attractive approaches toward this purpose is the encapsulation of IFN-α into microspheres by a biodegradable polymer, poly(lactic-glycolic acid) (PLGA). Sanchez et al. developed the method of encapsulation of IFN-α into biodegradable microparticles and nanoparticles using PLGA and poloxamer 188, which is a stabilizing agent, and encapsulated IFN-α within PLGA/poloxamer blend microspheres or nanospheres [176] where their preparation techniques led to the efficient encapsulation of IFN-α and the modulation of their particle size ranging from 280 nm to 40 μm. The antiproliferative activity of the IFN-α varied depending on the formulation. Moreover, PLGA/poloxamer blend microspheres were able to provide significant amounts of active IFN-α for up to 96 days.

In 2004, Peleg-Shulman et al. developed a new type of pegylated INF-α-2 conjugate, $peg_{40}$-FMS-IFN-α-2, which is capable of regenerating native INF-α-2 at a slow rate under physiological conditions [177]. A 2-sulfo-9-fluorenylmethoxycarbonyl (FMS)-containing bifunctional reagent, MAL-FMS-NHS, has been synthesized, enabling the linkage of a 40-kDa peg-SH to IFN-α-2 through a slowly hydrolyzable bond. The *in vitro* rate of regeneration of native interferon from this preparation was estimated to have a half-life of 65 h. Following subcutaneous administration to rats and monitoring of circulating antiviral activity, active IFN-α-2 levels peaked at 50 h, with substantial levels still being detected 200 h after administration. This value contrasts with a half-life of about 1 h measured for unmodified interferon. The concentration of active IFN-α-2 scaled linearly with the quantity injected. Comparing subcutaneous to intravenous administration of peg40-FMS-IFN-α-2, the long circulatory lifetime of IFN-α-2 was affected both by the slow rate of absorption of the pegylated protein from the subcutaneous volume and by the slow rate of discharge from the peg in circulation.

More recently, Ito et al. demonstrated the usefulness of the oral patch preparations for INF-α [178]. The patch preparations were composed of three layers; a water-insoluble backing layer, a drug-containing layer with an absorption enhancer, and a surface layer containing pH-dependent polymer were prepared. As absorption enhancer, three surfactants, Gelucire44/14 (lauroyl macrogol-32 glycerides), Labrasol (caprylocaproyl macrogol-8 glycerides), and HCO-60 (polyoxyethylated hydrogenerated castor oil) were used in preparing the IFN-α patch preparations, an in vitro release study showed that the time when half of the formulated IFN-α is released from the patches were 3.4 ± 0.1 min for HCO-60, 7.8 ± 0.1 min for Gelucire44/14, and 11.4 ± 0.1 min for Labrasol preparations. An in vivo absorption study showed that Gelucire44/14 preparation showed a higher $C_{max}$, 7.66 ± 0.82 IU/mL, and AUC, 12.85 ± 1.49 IU h/mL, than Labrasol (6.51 ± 0.89 and 8.30 ± 1.34 IU h/mL) and HCO-60 (6.02 ± 1.14, 7.53 ± 1.84 IU h/mL) preparations, respectively. By comparing the AUC obtained after s.c. injection of the same dose of IFN-α with rats, bioavailability was estimated to be 7.8% in the Gelucire44/14 preparation.

## 6 GROWTH HORMONE

### 6.1 Background

Growth hormone (GH) is an endocrine hormone produced and stored in the anterior pituitary gland, which is secreted into the blood circulation in a pulsation

pattern. Deficiency of GH results in abnormally poor growth. This disease state, namely hypopituitary dwarfism, has been recognized for several years. By the end of 1950s, it was found that a patient with hypopituitary dwarfism given injections of extracts of human pituitary tissue grew remarkably well. However, the supply of human GH (hGH) was extremely limited, because the only source of hGH was from cadaver pituitaries. Specific criteria were established for identifying patients who were clearly GH-deficient and who stood the best chance of benefiting from therapy. If identified, a child would receive intramuscular injections of these pituitary extracts three times per week. The dosing levels were limited and the continuing therapy was always threatened by shortages of material. In addition to being in thin supply, the problem of purifying the material provided a further challenge. As with any therapeutic drug that is derived from human tissue, contaminants of other protein hormones as well as those of viral origin had to be removed. However, more recently, the availability of synthesized hGH, produced through recombinant DNA technology, greatly reduced these concerns and provided a plentiful supply of hormone. As part of the process for proving the safety and efficacy of the use of hGH in humans, pharmacokinetic studies have been conducted in a number of species including mice, rats, monkeys, and humans.

### 6.2 Pharmacokinetics of Growth Hormone in Animals

In 1982, Sigel et al. studied that the disappearance rates of hGH and a 20-kDa variant over 10 min after intravenous injection of iodinated material in male and female mice [179]. The half-lives of hGH and the 20-kDa variant in males were 4.1 and 3.9 min, respectively, whereas the half-lives in females were 4.7 and 4.0 min. There were no significant differences in the protein half-lives between male and female mice. In addition, there was no apparent relationship between the half-lives, the dissimilar binding characteristics for the target sites, and growth-promoting activities. In 1988, the pharmacokinetics of biosynthetic human growth hormone (B-hGH) and pit-hGH were compared in rats [180]. Normal and hypophysectomized male and female rats were injected subcutaneously and intramuscularly with 0.1 mg/kg B-hGH and pit-hGH and intravenously with 0.06 mg/kg B-hGH and pit-hGH. A smaller distribution volume and a slower metabolic clearance rate were found for B-hGH compared with pit-hGH. Moreover, a smaller distribution volume and a slower metabolic clearance rate were found for hypophysectomized rats compared with normal rats for both proteins. The plasma half-lives was estimated to be about 3–7 min in the initial elimination phase and 29 min in the terminal elimination phase. The plasma levels of hGH were higher after subcutaneous compared with intramuscular administration for both proteins. From these results, it was suggested that an extensive local degradation took place at the subcutaneous and intramuscular injection sites. In addition, comparative tissue distribution studies with radioiodinated B-hGH and pit-hGH showed no differences between the growth hormones. The females accumulated relatively more label in the liver than the males, whereas the males accumulated relatively more label in the kidneys.

On the other hand, the total body clearance (CL) of synthetic hGH was studied in eight adult rhesus monkeys [181], where four monkeys were lean (<20% body fat) and the other four were obese (>20% body fat). The monkeys were given a

single i.v. bolus injection of hGH (2.5 mg/kg), followed by a constant infusion of hGH (250 mg/hr) for 2.5 h. Venous blood samples were collected before the infusion and every 10 min during the infusion. In both groups, steady state was reached 70 min after the start of the infusion. The mean CL of synthetic hGH was 12.7 ± 1.7 liters/24 h in the lean group and 19.5 ± 2.9 liters/24 h in the obese group ($p < 0.007$). However, the CL values that standardized by body weight were the same in both groups. These observations suggest that the CL of hGH is directly proportional to body weight, and the lower plasma GH levels in obesity may be caused by an increase in its CL not compensated for by an appropriate increase in the rate of GH secretion.

As for the metabolism, the metabolic fate of GH in peripheral tissues has not been clearly defined as with other protein hormones. The clearance of GH from the systemic circulation is thought to occur by receptor-mediated uptake at the liver and filtration at the kidney [182]. After intravenous injection of [$^{125}$I]hGH to female rats, hGH was speciacally taken up by the liver with radioactivity ultimately becoming associated with the lysosomal subfraction. The specific uptake of [$^{125}$I]hGH into rat liver was associated with accumulation in the Golgi apparatus [183, 184]. Interestingly, in each of these studies, radioactivity in the liver was characterized by TCA solubility and membrane binding as an intact hormone. These findings suggest that receptor-mediated uptake of GH in the liver is not directly linked to hydrolytic inactivation of the hormone.

### 6.3 Pharmacokinetics of Growth Hormone in Humans

In the early phases of product development, data from laboratory animals can sometimes be used to predict the phamacokinetic profile in humans [183]. Dosing levels for initial clinical trials can thus be selected. In the later stages of development, when human data were available, allometeric analysis is useful for evaluating the clinical relevance of pharmacological data collected in laboratory animals. Actually, the phamacokinetic parameters in humans can often be predicted by using pharmacokinetic data collected in laboratory animals by extrapolation by following equations based on body weight. Application of this method assumes that the pharmacokinetics are dose-proportional over the dosage range under condition. The allometric approach uses the following equation: $Y = aW^b$, where $Y$ represents a pharmacokinetic parameter and $W$ represents body weight in kg; constants, $a$ and $b$ represent the allometric constants and the allometric exponent. Using the allometric equations, that pharmacokinetic parameters (total body clearance, initial volume of distribution, and volume of distribution at steady state for rhGH in mice, rats, monkeys, and humans) show a strong association among these pharmacokinetic parameters as a function of body weight. This good interspecies agreement supports the use of pharmacokinetic data collected in laboratory animals to predict the distribution of rhGH and other proteins in humans [183].

### 6.4 Growth Hormone Formulation

GH is a pituitary hormone responsible for postnatal growth. Like all polypeptide hormones, GH was long thought to be circulating exclusively in the free form rather than being bound to plasma proteins, which was in contrast to some other

hormones of hydrophobic nature (e.g., steroids), which required protein binding for solubilization in the aqueous plasma environment. It is a 20–22-kDa polypeptide that exists in several molecular forms. GH acts by binding to GH receptors in multiple tissues; it initiates a cascade of biochemical events, among which the generation of insulin-like growth factor-I (IGF-I) is important for the growth process at the local tissue level. Other important actions of GH are those involved in anabolism, such as intracellular transport of amino acids and other cellular constituents, nitrogen, and protein synthesis. The biochemical events immediately following GH binding to receptors (intracellular signaling) are only incompletely understood. In 1987, the hepatic GH receptor was cloned and structurally characterized; it is a single-chain glycoprotein with a 620-amino acid backbone, of which 246 are extracellular, 24 form a hydrophobic transmembrane domain, and 350 are intracellular [185]. Like all polypeptide hormones, GH was long thought to be circulating exclusively in the free form rather than being bound to plasma proteins. At the end of 1980s, however, it was found that GH also circulates by binding to specific GH-binding proteins in plasma [186]. At least two GH-binding proteins (GH-BP) have been identified in human plasma; one binds GH with high affinity, the other with lower affinity.

hGH is currently administered by injection daily or three times a week. To improve patient compliance, there have been a lot of research efforts for the development of long-acting formulations of protein drugs. The parenteral delivery route often requires frequent dosing because of the short half-lives of these drugs. A few clinical studies have shown that a continuous infusion of hGH via a pump resulted in growth velocity and insulin-like growth factor-I (IGF-I) levels comparable with those of daily injections. The results indicate that pulsatile hGH release may not be required for clinical efficacy, and the sustained-release formulation may improve patient compliance and efficacy with a suitable pharmacokinetics release profile [187]. For a development of sustained-release formulation of protein drugs, one of the key requirements may be the integrity of the protein drugs within microparticles during and after formulation processes contrary to small-molecular-weight drugs; proteins have three-dimensional structures to retain their bioactivity. Some stresses (e.g., temperature, pH, or organic solvents) associated with the preparation of sustained-release formulation can cause deamidation, oxidation, denaturation, or aggregation of the protein, which can turn the protein biologically inactive and even immunogenic.

From the end of the 1990s to the early 2000s, rapid development in molecular biology and recent advancement in recombinant technology increased identification and commercialization of potential protein drugs. It is well known that traditional forms of administration for the peptide and protein drugs often rely on their parenteral injection, as the bioavailability of these therapeutic agents is poor when administered nonparenterally. Tremendous efforts by numerous investigators across the world have occured to improve protein formulations and, as a result, a few successful formulations have been developed including sustained-release hGH. At the end of 1990s, a sustained-release system was developed for rhGH using biodegradable microsphere formulations composed of polymers of lactic and glycolic acid (PLGA). Lee et al. reported that long-acting formulations of rhGH were prepared by stabilizing and encapsulating the protein into three different injectable, biodegradable microsphere formulations composed of PLGA [188]. In this study, the microsphere

formulations were compared in juvenile rhesus monkeys by measuring the serum levels of rhGH and two proteins induced by hGH, insulin-like growth factor-I (IGF-I) and IGF binding protein-3 (IGFBP-3) after single s.c. administration. The results showed that PLGA formulations induced a higher level of IGFBP-3 than was induced by injections of the same amount of rhGH in solution. Moreover, Brodbeck et al. evaluated the effects of altering dynamics of PLGA solution depot on the sustained-release delivery of hGH by subcutaneous injection in rats, and obtained about 28 days of long-term sustained-release profiles from the PLGA solution depot of hGH in normal rats at serum levels of 10–200 ng/mL [189].

In the early 2000s, several other formulations of hGH have been developed. In 2001, Maeda et al. developed a sustained-release collagen film of hGH and evaluated its effect on wound healing in db/db mice [190]. The release profiles of hGH from the collagen films varied with composition and preparation conditions; however, the film prepared by air-drying of the mixture of hGH and collagen solution released hGH continuously over 3 days both in vitro and in vivo. By application of collagen film containing 3 mg of hGH twice at an interval of 6 days to wounds, the area of wounds on day 21 was significantly reduced compared with that of nontreated wounds in mice. In 2002, Kim et al. reported the usefulness of the synthesized HA/Pluronic composite hydrogels [temperature-sensitive hyaluronic acid (HA) hydrogels, which were synthesized by photopolymerization of vinyl group modified HA in combination with acrylate group end-capped poly(ethylene glycol)-poly(propylene glycol)-poly(ethylene glycol) tri-block copolymer (Pluronic F127)] [191]. This HA/Pluronic composite hydrogels gradually collapsed with increasing temperature over the range of 5 to 40°C suggesting that the Pluronic component formed self-associating micelles in the hydrogel structure. The mass erosion occurred much faster at 37°C than at 13°C, indicating that at the higher temperature, the ester linkage between the Pluronic and acrylate group might be more exposed to an aqueous environment and thus be more readily hydrolyzed because of Pluronic micellization. This HA/Pluronic composite hydrogels is applicable for a formulation of rhGH in sustained-release profiles that followed a mass erosion pattern.

More recently, aiming at once-a-week injection, a novel sustained-release formulation of rhGH using sodium hyaluronate (SR-rhGH) was developed for the treatment of children who have growth failure caused by lack of adequate secretion of endogenous GH [192, 193]. SR-rhGH was produced in the form of solid microparticles by spray-drying technology. A single administration of a prototype formulation of SR-rhGH with a ratio of hGH:HA-1:1 to cymomolgus monkeys through a fine 26-gauge needle induced continuous elevation of serum IGF-I level for 6 days, demonstrating the bioactivity of hGH released from the prototype formulation. When the ratio of hGH to HA changed from 1:1 to 1:3, hGH released more slowly in vitro from SR-rhGH with almost complete release of hGH loaded. According to phamacokinetic and pharmacodynamic studies in beagle dogs, sustained release of hGH from the optimized formulation of SR-rhGH continued for a more extended period longer than 72 h with a lower $C_{max}$ than those of prototype formulations. The single administration resulted in an elevation of serum insulin-like growth factor-I (IGF-I) level for 6 days, with a maximum value that was higher than the baseline level by 350 ng/mL, which supported the possibility of SR-rhGH as a once-a-week injection formulation of hGH. The bioavailability of both

formulations was comparable with that of hGH daily injection formulation. Finally, toxicity studies revealed no evidence of adverse effect in both cymomolgus monkeys and beagle dogs [193].

As for a formulation using another administration route, Leitner et al. developed a nasal delivery system of hGH [194] based on the thiomer polycarbophil-cysteine (PCP-Cys) in combination with the permeation mediator glutathione (GSH). Microparticles were prepared by dissolving PCP-Cys/GSH/hGH (7.5 : 1 : 1.5), PCP/hGH (8.5 : 1.5), and mannitol/hGH (8.5 : 1.5) in demineralized water, followed by lyophilization and micronization. PCP-Cys/GSH/hGH and PCP/hGH microparticles showed a comparable size distribution (80% in the range of 4.8 to 23 μm) and swelled to almost four fold size in phosphate-buffered saline. Both formulations exhibited almost identical sustained drug release profiles. The intranasal administration of the PCP-Cys/GSH/hGH microparticulate formulation resulted in a relative bioavailability of 8.11%, which represents a three fold and a 3.3-fold improvement compared with that of PCP/hGH microparticles and mannitol/hGH powder, respectively. The nasal microparticulate formulation based on PCP-Cys/GSH/hGH might represent a promising novel tool for the systemic delivery of hGH.

## 7 LEUPROLIDE

### 7.1 Background

Leuprolide acetate is a synthetic and superpotent agonist of luteinizing hormone-releasing hormone (LHRH) with a biphasic effect on the pituitary (Figure 2).

It is a decapeptide hormone that is effective in the treatment of hormone-dependent diseases such as prostate and mammary tumors and endometriosis [195, 196]. It initially stimulates the secretion of both luteinizing hormone (LH) and follicle stimulating hormone (FSH), and on long-term continuous administration it inhibits secretion of LH and FSH, and the concentration of testosterone in men and estrogen in women drop sharply. It is administered pulsatively for adjuvant hormonal therapy and continuously for hormone-related anti-neoplastic therapy comprising prostatic and breast cancer and endometriosis [197]. As leuprolide is a peptide, it is orally inactive and generally given subcutaneously or intramuscularly. Recently, however, several attempts are considered to improve bioavailability and effectiveness of leuprolide as well as a patient's convenience for taking the medicine.

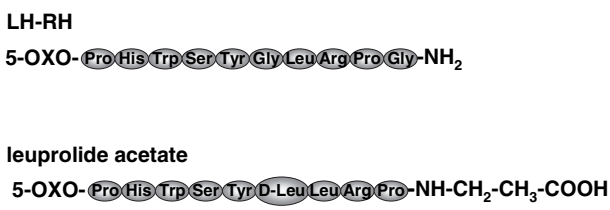


**Figure 2.** Schematic presentation of primary structure of luteinizing hormone-releasing hormone (LH-RH) and leuprolide acetate.

### 7.2 *In vitro* Permeation Study of Leuprolide

For most polypeptides, oral administration is often limited by enzyme instability in the gastrointestinal tract. Furthermore, because peptides are highly ionized, and thus are mostly hydrophilic, their potential for permeating gastrointestinal absorption barriers is poor. Attempts to understand the degradation and permeation behavior in the intestine are beneficial to improve oral efficacy. Intramuscular and intranasal routes are the common routes of administration. Undoubtedly, the oral route is the most convenient. However, plasma levels of leuprolide after oral administration of an aqueous solution to rats and humans are mostly below the limit of detection (0.2 ng/mL) [198]. Therefore, it is important to understand and elucidate the absorption and degradation mechanisms of leuprolide in the intestinal tract. In 2004, Guo et al. reported the transport mechanism of leuprolide across rat intestine, rabbit intestine, and Caco-2 cell monolayer [199]. Using techniques based on everted gut sac and Caco-2 cell monolayer, it was found that flux of leuprolide increased with increasing concentration of drug, showing a passive diffusion pathway. At a low concentration of leuprolide, trypsin inhibitor had a strong enhancement effect on the permeability of drug by protecting enough drug for permeation. Chitosan had no affect on the activity of α-chymotrypsin. In addition, the authors found that the increase in permeation of leuprolide was caused by opening of the tight junctions and interaction with cells. Therefore, the authors concluded that both inhibition of proteolytic enzymes and opening the tight junctions to allow for paracellular transport of leuprolide improved the intestinal absorption.

### 7.3 Clinical Pharmacokinetics of Depot Leuprolide

In 1984, for cancer treatment, a sustained-release intramascular formulation based on biodegradable microspheres had been developed by Saunders et al. and thereafter marketed (Lupron@ Depot, TAP Pharmaceuticals) [200]. This injectable microsphereiscapable of sustaining required drug levels over a period of 1, 3, and 4 months after a single administration. These systems, in addition to improving the patient convenience and compliance, reduce the needed dose to one-quarter to one-eighth of the injected aqueous solution by sustaining therapeutic drug levels at target receptor sites [201]. More recently, sustained-release depot formulations, in which the hydrophilic leuprolide is entrapped in biodegradable highly lipophilic synthetic polymer microspheres based on polylactic/glycolic acid (PLGA) or polylactic acid (PLA), have been developed to avoid daily injections [202], where leuprolide is released from these depot formulations at a functionally constant daily rate for 1, 3, or 4 months, depending on the polymer type (PLGA for a 1-month depot and PLA for depot of >2 months), with doses ranging between 3.75 and 30 mg. The mean peak plasma leuprolide concentrations of 13.1, 20.8–21.8, 47.4, 54.5, and 53 μg/L occur within 1–3 h of depot subcutaneous administration of 3.75, 7.5, 11.25, 15, and 30 mg, respectively, compared with 32–35 μg/L at 36–60 min after a subcutaneous injection of 1 mg of a nondepot formulation. Sustained drug release from the PLGA microspheres maintains plasma concentrations between 0.4 and 1.4 mg/L over 28 days after single 3.75-, 7.5-, or 15-mg depot injections. The

mean areas under the concentration-time curve are similar for subcutaneous or intravenous injection of short-acting leuprolide 1mg; a significant dose-related increase in the AUC from 0 to 35 days is noted after depot injection of leuprolide 3.75, 7.5, and 15 mg. The mean volume of distribution of leuprolide is 37 L after a single subcutaneous injection of 1 mg, and 36, 33, and 27 L after depot administration of 3.75, 7.5, and 15 mg, respectively. The total body clearance is 9.1 L/h and the elimination half-life is 3.6 h after a subcutaneous 1-mg injection; corresponding values after intravenous injection are 8.3 L/h and 2.9 h. A 3-month depot PLA formulation of leuprolide 11.25 mg ensures a mean plasma peak concentration of around 20 μg/L at 3 h after subcutaneous injection, and continuous drug concentrations of 0.43 to 0.19 μg/L from day 7 until before the next injection. In addition, an implant administration of these sustained-release systems of leuprolide provided a lasting delivery of leuprolide for 1 year. Serum leuprolide concentrations remained at a steady mean of 0.93 mg/L until week 52, suggesting zero-order drug release from the implant.

### 7.4 Transdermal Iontophoretic Delivery of Leuprolide

As for another route of administration of leuprolide, Kochhar et al. demonstrated an application of transdermal delivery of leuprolide using heat-separated human epidermal membrane from cadaver skin of multiple donors. The author used an iontophoresis technique to deliver the drug and examined the effect of constant voltage or constant current application [203, 204]. When the heat-separated human epidermal membrane was subjected to constant voltage within the range of 250 to 1000 mV during the iontophoretic phase, iontophoretic enhancement of leuprolide permeation at pH 7.2 was greater than at 4.5. A developed model to account for iontophoretic enhancement yielded that, first, the porosity increased with the applied voltage to as much as three times the original at 1000 mV. Second, the lipid pathway scontributed approximately 20% to the total permeation during the passive phase. Third, the electro-osmotic flow contributed significantly to the enhancement and its direction was from anode to cathode at pH 7.2 and the opposite at pH 4.5. The magnitude of the electro-osmotic flow was at pH 4.5 somewhat lower than at pH 7.2. Repeated iontophoretic applications of 250 mV on the same skin specimen resulted in the same enhancement every time and did not cause any barrier alterations when applied for 1 h every 24 h, which was not the case if the duration between the two iontophoretic applications was only 3 h. In contrast, keeping the current densities at 0.5 to 2.3 $\mu A/cm^2$, the permeation rate of leuprolide increased linearly with the current density for the universal buffer and at pH 7.2 was almost double that at pH 4.5 despite the greater ionic valence of the drug at pH 4.5 compared with pH 7.2. The drug transference number at both pH values was approximately 0.5%. Replacement of the universal buffer with polymaleic acid yielded higher drug permeation rates and increased its transference number at a comparable pH. Transference number of the drug with polymaleic acid appeared to increase with current density. From these in vitro investigations, the authors concluded that the fluxes obtained for both electrolyte systems with the present experimental arrangement could be extrapolated to deliver therapeutically relevant doses of the drug.

# 8 DESMOPRESSIN

## 8.1 Background

Desmopressin (the vasopressin analog 1-deamino-8-D-arginine vasopressin, dDAVP), a potent synthetic peptide hormone of the natural pituitary hormone vasopressin, is used chiefly for treatment of enuretic disorders, such as central idiopathic or secondary diabetes insipidus [205, 206] and nocturia and nocturnal enuresis in young children [207]. Desmopressin has also proven to be useful in the treatment of bleeding diseases by activating and mobilizing factor VIII and von Willebrand's factor reserves in hemophilia A and von Willebrand–Jurgens syndrome [208]. The 1100-Da molecule, which is typically taken in doses of 1–20 μg, shows variable and low oral and nasal bioavailability (0.1% and 3.4%, respectively) [1]. Although injectable formulations demonstrate better bioavailability, they are poorly suited for routine use in young children. For the treatment of these disorders, desmopressin (Desmopressin®) in usually administered intranasally by use of sprays or drops. This route of desmopressin administration was shown to be more efficacious than the oral route, because bypassing the gastrointestinal tract increases the absolute bioavailability from less than 1% to approximately 5% [209]. A more acceptable route of administration with potentially good bioavailability could be offered by transdermal delivery. It is well documented that the stratum comeum, the outermost layer of the skin, constitutes an impermeable barrier to hydrophilic or high-molecular-weight drugs. These molecules can only be delivered into or through the skin if the barrier function of the stratum corneum is disrupted by any of a number of available methods. In this field, ultrasound and iontophoresis, which are known to enhance transdermal delivery of small molecules, have also been proposed for use with peptides and macromolecules.

## 8.2 Pharmacokinetics of Desmopressin in Humans

As a basic study, Rembratt et al. investigated the pharmacokinetic and pharmacodynamic profiles of desmopressin after oral and intravenous ad-mini-stration in men with a high incidence of nocturia [210]. Based on an open-randomized and four-way crossover design, desmopressin was administered orally (0.2 mg) and intravenously (2 μg) at daytime and nighttime. The concentration-time curve after 2 μg intravenous desmopressin was best fitted by a biexponential equation. The mean (95% CI) AUC at night was 302 (272–335) pg h/mL and in the day was 281 (253–312) pg h/mL. No statistically significant differences were detected between night and day except for terminal half-life, which was 3.1 h at night and 2.8 h in the daytime ($p = 0.02$). Peak plasma concentration ($C_{max}$) was 6.2 (5.1–7.5) pg/mL at night and 6.6 (5.5–7.9) pg/mL in the daytime, respectively. Median time to reach $C_{max}$ ($t_{max}$) was 1.5 (range 1.0–4.1) h at night and 1.5 (range 0.5–0.3) h in the day, and the oral bioavailability was 0.08%. The pharmacodynamic effects of oral and intravenous desmopressin given in the daytime were similar during the first 6 h after dosing. The nighttime dosing and daytime intravenous dose resulted in antidiuresis throughout the measuring period, whereas the effect of the daytime peroral dose receded after 6 h. The authors concluded that the terminal half-life was longer at

night than in the daytime, but the difference is considered too small to be of clinical importance. Moreover, they also concluded that, despite low bioavailability, the pharmacodynamic effects of oral desmopressin were similar in magnitude to those after intravenous dose at night and during the first 6 h after daytime administration.

More precisely, d'Agay-Abensour et al. investigated the absolute bioavailability of an aqueous solution of desmopressin from different regions of the gastrointestinal tract (stomach, duodenum, jejunum, ileum, colon, rectum) in 6 healthy male volunteers aged 24 to 35 years [211]. For intravenous administration, the subjects received 4 μg of desmopressin. For intestinal administration, 400 μg of desmopressin was directly applied to six distinct sites in the GI tract via two- or four-channel tubes with or without a distal occlusive balloon. After intravenous administration, the elimination half-life, total body clearance, amount excreted in urine, and renal clearance of desmopressin were 60.0 min, 1.7 mL/min/kg, 2.0 μg, and 0.8 mL/min/kg, respectively. The mean bioavailability of desmopressin after gastric, duodenal, jejunal, distal ileal, proximal colonic, and rectal application was 0.19%, 0.24%, 0.19%, 0.03%, 0.04%, and 0.04%, respectively. Thus, absorption was significantly higher in the three upper GI regions in comparison with the three lower regions.

### 8.3 Renal and Biliary Excretion and Sex Differences in Desmopressin Pharmacokinetics

Agerso et al. investigated the differences in pharmacokinetic profiles of desmopressin between healthy subjects and renally impaired patients after intravenous administration [212]. Overall, 18 renal impairment and 6 healthy volunteers were enrolled in the study. Each subject received a single intravenous dose of 2-μg desmopressin, and blood and urine samples were collected for 24 h, where plasma concentrations and the amounts of desmopressin excreted in the urine were analyzed simultaneously by means of a technique of mixed effects modeling. Their results showed that both renal and nonrenal clearances of desmopressin were found to vary with the creatinine clearance (CrCL). A decrease of 1.67% in the CrCL was found to cause a 1.74% decrease in the renal clearance and a 0.95% decrease in the nonrenal clearance. The fall in renal failure caused the amount of desmopressin excreted in urine to decrease from 47% in healthy subjects to 21% in the patients with severe renal impairment. The mean total body clearance of desmopressin in healthy subjects and patients with severe renal impairment was 10 L/h and 2.9 L/h, respectively. Correspondingly, the mean terminal half-life was 5.7 h in healthy subjects and 10 h in patients with severe renal impairment. The authors concluded that great caution should be taken when deciding on an efficient dosage regimen if patients with moderately or severely impaired renal function are to be treated with desmopressin at all, despite the drug appearing to be safe and well tolerated by patients with impaired renal function.

The pharmacokinetic profiles of desmopressin in the liver were also investigated. Lundin et al. demonsted the biliary excretion of drug after intrajugular venous, intraportal venous, and intraduodenal venous administration in conscious pigs [213]. In all cases of administration routes, the biliary excretion was less than 1% of the administered dose with the plasma/bile concentration ratio of less than 1.0. A significant first-pass effect was found when the liver was exposed to a high intra-

portal dose of desmopressin. The uptake study of [$^3$H]desmopressin by incubating with liver tissue slices showed that desmopressin was rapidly removed from the incubation medium, indicating desmopressin was markedly metabolized in the liver.

As another parameter to alter the pharmacodynamic profile of desmopressin, a sex difference is most important. In 2004, Odeberg et al. demonstrated that there was a sex difference in a human pharmacokinetic (PK) and pharmacodynamic (PD) study [214], where desmopressin was administered intravenously as a single dose (for the PK study, a 2-μg dose; for the PD study, a 0.2-μg dose), and parameters for urine flow and urine osmolality were estimated. The pharmacokinetics of desmopressin after a fixed bolus injection were influenced neither by piroxicam nor by sex of subjects. However, the pharmacodynamics of desmopressin showed a sex difference where females exhibited a more pronounced antidiuretic effect than males, which was statistically significant when the effects were submaximal (>4.5 h after dose). The sex differences were diminished after pretreatment with an NSAID, piroxicam, indicating a prostaglandin $PGE_2$-mediated mechanism.

## 8.4 Desmopressin Formulation

Usually, desmopressin is administered intranasally by use of sprays or drops. This administration route of desmopressin was considered to be more efficacious than the oral route, because bypassing the gastrointestinal tract increases the absolute bioavailability from less than 1% to approximately 5%. However, nasal application of desmopressin is accompanied by high intersubject and intrasubject variability in plasma pharmacokinetics [215]. Therefore, there have been several pharmaceutical research efforts to improve nasal delivery of desmopressin.

In 1987, Lethagen et al. demonstrated a comparison study, which was made of intranasal administration of 300-μg desmopressin by spray, with intravenous administration of 0.2-, 0.3-, and 0.4-μg/kg desmopressin in 10 healthy volunteers [216], where the effect of desmopressin was measured on the antihemophilia factor, factor VIII (F VIII), and on plasminogen activator release. Plasma levels of desmopressin showed a clear dose response with the maximum levels at 0.4 μg/kg intravenous dose. The affect of the spray approximated the 0.2-μg/kg response. However, the maximum response in both F VIII complex and plasminogen activator release was obtained after a 0.3-μg/kg i.v. dose. The response at a 0.4-μg/kg i.v. dose was not significantly different from the response at a 0.3-μg/kg dose, indicating that maximum stimulation was reached with 0.3 μg/kg. There was no correlation between plasma levels of F VIII and desmopressin, indicating that the biological response to the drug is subject to saturation kinetics. The reproducibility of the effect of the spray dose on F VIII was 21% (c.v.) and 27% for the intraindividual and interindividual variation, respectively, and compared favorably with intravenous administration. Therefore, they concluded that intranasal desmopressin (300 μg) is as effective as 0.2 μg/kg intravenously and provides an accurate, reproducible, and convenient alternative to parenteral administration. Moreover, in 1988, Harris et al. reported the effects of concentration and dose volume on the nasal bioavailability and biological response to desmopressin in humans [217]. A nasal formulation of 300 μg of desmopressin was administered using a premetered spray device in doses of either $1 \times 50$-, $2 \times 50$-, or $1 \times 100$-μL actuations to both nostrils.

Intravenous administration of 0.2 μg/kg was also given as a reference for bioavailability calculations. The biological response was determined by measuring circulating levels of F VIII. Peak plasma levels of desmopressin were greatest after the 2 × 50-μL dose, followed by the 1 × 50- and 1 × 100-μL doses. The bioavailability of desmopressin from the 2 × 50-μL dose was 20%, which was significantly greater than the 11% after the 1 × 50-μL ($p < 0.01$) and 9% after the 1 × 100-μL ($p < 0.001$) doses. The biological response was clearly enhanced after the 2 × 50-μL dose compared with the 1 × 50- and 1 × 100-μL doses. The interindividual response in F VIII levels to nasal desmopressin ranged from 20% (CV) to 30%, which compared favorably with the 36% variation after intravenous administration. These results show that by optimizing concentration, volume, and splay apparatus, a significant enhancement can be obtained in bioavailability and clinical efficacy of desmopressin.

More recently, as a new apparatus for desmopressin delivery, Cormir et al. reported the usefulness of a coated microneedle array patch system (Macronux technology) to deliver desmopressin through the skin [218]. Although it is available in injectable, intranasal, and oral formulations, intranasal and oral administration result in low and variable bioavailability. However, administering desmopressin transdermally using Macronux technology, which uses a microneedle array to overcome the skin barrier, is markedly useful to improve the bioavailability of desmopressin. In their reports, the tips of microneedles in 2-$cm^2$ arrays were covered with a solid coating of various amounts of desmopressin and applied to the skin of hairless guinea pigs for 5 or 15 min. Pharmacologically relevant amounts of desmopressin were delivered after 5 min. Bioavailability was as high as 85% and showed acceptable variability (30%). Immunoreactive serum desmopressin reached peak levels after a $T_{max}$ of 60 min. Elimination kinetics for serum desmopressin was similar after other transdermal and intravenous delivery, suggesting the absence of a skin depot. These results suggest that transdermal delivery of desmopressin by Macroflux technology is a safe and efficient alternative to currently available routes of administration.

## 9 ANTIBODIES

### 9.1 Background

Antibodies serve two important functions: They bind and modulate antigens and they bind complement and immune effector cells, such as natural killer cells and monocytes. Recently, targeted therapies using monoclonal anti-bodies (mAbs) have gained increased therapeutic application. As shown in Table 1, overall, 19 mAbs are currently approved by FDA to treat various disease states, such as oncology, inflammation, infectious disease, and cardiovascular disease [219]. All approved mAbs are of the IgG class. Each IgG molecule contains two identical heavy chains and two identical light chains, 13 of which are intact mAbs, three are conjugated, and one is an mAb fragment (Fab). A feature of mAb therapeutics is the high specificity conferred by the antibody interaction with a specific region on the targeted antigen. The antigen-binding site is formed by the intertwining of the light-chain variable domain ($V_L$) and the heavy-chain variable domain ($V_H$). Each V domain contains three short stretches of peptide known as the complementarily

**TABLE 1. Therapeutic Antibodies Approved by FDA**

| Approved | Generic name | Trade mame | Disease indication |
|---|---|---|---|
| 1986 | muromanab–CD3 | 0KT3 | Allograft regction |
| 1994 | abciximab | ReoPro | Adjunct to PTCA |
| 1995 | edrecolomab | Panorex | Colorectal cancer |
| 1997 | rituximab | Rituxan | Non-Hodgkins lymphoma |
| 1997 | daclizumab | Zenapax | Prevention of kidney Transplantr rejection |
| 1998 | trastuzumab | Herceptin | Metastatic breast cancer |
| 1998 | palivizumab | Synagis | RSV propylaxis |
| 1998 | basiliximab | Simulect | Prevention of kidney Transplantr rejection |
| 1998 | infliximab | Remicade | Rheumatoid arthrists, Crohn's disease |
| 2000 | gemtuzumab ozogamicin | Mylotarg | CD33-acute myeloid leukemia |
| 2001 | alemtuzumab | Campath | B-cell chronic lymphocyte leukemia |
| 2002 | ibritumomab tiuxetan | Zevalin | Non-Hodgkings lymphoma |
| 2002 | adalimumab | Humira | Rheumatoid arthrists, Crohn's disease |
| 2003 | omalizumab | Xolair | Asthma |
| 2003 | efalizumab | Raptiva | Prorisis |
| 2003 | tosirtumomab | Bexxar | Non-Hodgkings lymphoma |
| 2004 | cetuximab | Erbitux | Colorctal cancer |
| 2004 | bevacizumab | Avastin | NSCL cancer |

determining regions (CDRs); the CDRs are the prominent determinants of antigen-binding affinity and specificity. The light chain contains one constant domain, $C_L$. The heavy chain contains three constant domains, $C_H1$, $C_H2$, and $C_H3$. The $C_H2$ and $C_H3$ domains allow interactions of the IgG molecule with various components of the immune system by either binding C1q, which activates the complement cascade and elicits complement-dependent cytotoxicity, or by binding to Fcγ receptors on immune effector cells, which elicits antibody-dependent cellular cytotoxicity (Figure 3).

These same variable and constant domains of the molecule also affect IgG catabolism and elimination [220]. In addition to their contributions to mAb pharmacological activities, the FcyRs could also regulate elimination and pharmacokinetics of mAbs. Other factors, such as mAb structure and engineering, host factors, concurrent medications, and immunogenicity, can alter the pharmacokinetic and pharmacodynamic profiles. Hence, understanding the factors that affect the pharmacokinetics of mAb is of high importance for effective therapeutic application.

### 9.2 Pharmacodynamics and Pharmacokinetics of mAbs

mAbs can exert their pharmacological effects via several different mechanisms, such as neutralizing antigen function, activating receptors by mimicking

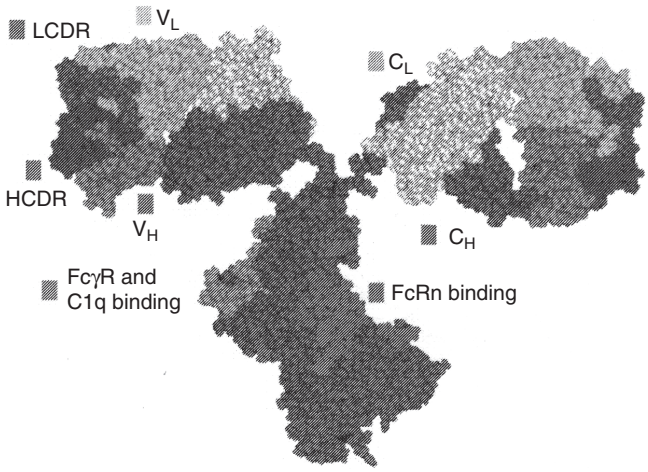


**Figure 3.** Representation of the space-filling model of an IgG molecule.

endogenous receptor ligands, delivering toxins to specific cells (targeted delivery), and eliciting effector functions in conjunction with antigen modulation [221]. Under some circumstances, where antigen expression is high in crucial organs (e.g., heart, lung, and vasculature), the effector function might not be desirable and could be deleterious; however, in other instances, such as applications in hematological malignancies, effector functions might be a significant part of the mechanism of action and maximizing the antibody effector functions might be highly desirable. Clearance of antibodies through the cells of the reticuloendothelial system (RES) can be regulated through the interaction with various Fc receptors. FcRns are expressed on phagocytic cells of the RES and protect IgG from rapid clearance. Similarly, cells of the RES express various types of Fcγ receptors. Interactions of IgG antibodies with this subset of receptors could potentially impact antibody clearance. As discussed, polymorphism in Fcγ receptors impacted the therapeutic response to rituximab [222]. Although the favorable clinical outcome was attributed to a more efficient Fcγ RIIIa-dependent cytotoxicity, it is not clear if polymorphism in Fcγ receptors could have any potential impact on antibody clearance. Polymorphism in Fcγ RIIIa was shown to impact in vivo clearance of red blood cells (RBC) coated with an anti-D IgG3 antibody in humans [223]. The faster clearance was attributed to a more efficient elimination of opsonized RBC by phagocytic cells in the spleen of the subjects homozygous for Fcγ RIIIa-F/F158 [223]. These results are consistent with previous data demonstrating the impaired clearance of IgG-sensitized RBC in the presence of an anti-Fcγ RIIIa antibody [224]. In addition, downregulation of Feγ receptor on cells of monocyte lineage by immunomodulatory drugs, such as methotrexate (MTX), might potentially impact the clearance of mAbs such as adalimumab.

As the site on IgG that is responsible for binding to FcRn has been mapped and well characterized, the trend in the field of antibody engineering is to mutate the FcRn-binding site. Therefore, an increased FcRn binding provides altering the pharmacokinetics of mAbs. In 2002, Dall'Acqua et al. found that major improvement in FcRn binding occurred when mutations were introduced at positions 252,

254, 256, 433, 434, and 436, which are at the interface of the Fc-FcRn-binding region [225]. More recently, Hinton et al. discovered two mutations, T250Q and M428L, that caused, respectively, a three fold and seven fold increase in FcRn binding and, when combined together, caused a 28-fold increase in FcRn binding [226]. After intravenous administration of the M428L or T250Q/M428L mutant, about a two fold increase in plasma half-life was observed in rhesus monkeys compared with the parental molecule. As these positions are conserved among different human IgG isotypes, the authors proposed that the beneficial effect of these mutations should hold good for other isotypes as well.

As another way to alter the pharmacokinetics of mAbs, there is a PEGylation technology (the process of conjugating polyethylene glycol chains to the antibody fragments) of mAbs. PEGylation of proteins and liposomes has been a time-tested and successful technique that offered the advantage of reducing immunogenicity, increasing the plasma half-life, increasing solubility, and reducing protease sensitivity [227]. Therefore, the science of antibody PEGylation has two primary aims, which are (1) to preserve the antigen-binding activity completely and (2) to link the PEG molecule to the antibody in a stable manner. These aims are achieved by performing site-specific PEGylation using maleimide chemistry. Site-specific PEGylation was done by introducing a free cysteine to the end of the hinge region in a Fab [228] or by incorporating the hinge region on the C-terminus of a Fab and scFv [229]. The increase in half-life observed with PEGylated antibody fragments is usually caused by a prolongation of the distribution phase that represents the redistribution of a molecule in the extravascular environment. It, therefore, appears that PEGylation slows the redistribution of the mAbs from the systemic circulation to the interstitial compartment. In the following section, representative mAbs will be described.

### 9.3 Muromonab CD-3 Pharmacokinetics

In 1986, Muromonab CD-3 (OKT-3), which is a monoclonal antibody was first introduced to clinical practice, and the agent has been effective in reversing corticosteroid-resistant acute rejection in renal, liver, and cardiac transplant recipients. OKT-3 is administered only by intravenous injection and has a harmonic half-life of approximately 18 h. It binds specifically to the CD-3 complex, which is involved in antigen recognition and cell stimulation, on the surface of T lymphocytes. Immediately after administration, CD-3-positive T lymphocytes are abruptly removed from the circulation. OKT-3 may be removed by opsonization by the reticuloendothelial system when bound to T lymphocytes, or by human antimurine antibody production [230].

### 9.4 Basiliximab Pharmacokinetics

Basiliximab (Simulect) is a chimeric monoclonal antibody for immunoprophylaxis against acute rejection in renal transplantation. Basiliximab is a glycoprotein produced by recombinant technology. It is used to prevent white blood cells from acute renal transplantation rejection. It specifically binds to and blocks the alpha chain of interleukin-2 receptors (IL-2R alpha), also known ascD25 antigen, on the surface of activated T lymphocytes. In 1997, to identify a single-dose regimen providing IL-2R-saturating serum concentrations in the critical first posttransplant month,

Kovarki et al. examined the pharmacokinetic aspects of basiliximab in a multicenter open-label, prospective dose-escalation study [231]. Thirty-two recipients (20 men and 12 women, 47 ± 11 years old, 65 ± 12 kg body weight) of primary mismatched cadaver kidneys were enrolled. The immunosuppression regimen consisted of steroids and azathioprine from day 0 and cyclosporine from day 10. Basiliximab was infused over 30 min as a single dose (40 mg or 60 mg). Peak concentration and area under the concentration versus time curves increased proportionally with dose. Postinfusion concentrations declined in a biphasic manner with a terminal half-life of 6.5 ± 2.1 days. Widely dispersed correlations were noted between body weight and distribution volume (r = 0.29) or between it and total body clearance (r = 0.45), suggesting no clinical relevance for weight-adjusted dosing. Moreover, there were no apparent gender-related differences in basiliximab disposition. As serum concentrations in excess of 0.2 μg/mL are sufficient to saturate IL-2R epitopes on circulating T lymphocytes, the authors concluded that 40 mg and 60 mg of infusion dose was sufficient to keep this threshold level for 26 ± 8 days and for 32 ± 11 days, respectively.

Moreover, Kovarki et al. reported the results of a randomized, open-label prospective study with recipients of primary cadaveric liver allografts to characterize the disposition and immunodynamics of basiliximab for immunoprophylaxis of acute rejection [232]. Patients received a total intravenous dose of 4-mg basiliximab in addition to baseline dual immunosuppression consisting of cyclosporine and steroids. The central volume of distribution was 5.6 ± 1.7 L, with a steady-state volume of 7.5 ± 2.5 L. The total body clearance and elimination half-life were 75 ± 24 mL/h and 4.1 ± 2.1 days, respectively. Basiliximab was measurable in drained ascites fluid, and clearance by this route was an average of 20% of total clearance. Total body clearance correlated positively with the volume of postoperative blood loss (r = 0.5253, p = 0.0101), suggesting that bleeding may represent an additional route of drug removal. More recently, they conducted a population pharmacokinetic screen to identify demographic-clinical covariates of basiliximab in kidney and liver transplantation [233, 234]. In kidney transplantation (169 basiliximab-treated patients), basiliximab clearance was 36.7 ± 15.2 mL/h, distribution volume 8.0 ± 2.4 L, and half life 7.4 ± 3.0 days. Body weight (range, 44–131 kg) and age (range, 20–69 years) each contributed to the variability in clearance and volume. However, gender, ethnic group, and the presence of proteinuria had no clinically relevant influences on basiliximab disposition. Receptor-saturating basiliximab concentrations were maintained for 36 ± 14 days (range, 12–91) when patients received 4-mg basiliximab [233]. On the other hand, in liver transplantation (184 basiliximab-treated patients), basiliximab clearance was 55 ± 26 mL/h, the distribution volume was 9.7 ± 4.2 L, and the half-life was 8.7 ± 6.7 days. Patient weight, age, sex, ethnicity, history of alcoholism, hepatitis C seropositivity, and notable postoperative bleeding had no clinically relevant influences on basiliximab disposition; however, the cumulative volume of drained ascites fluid in the first week was positively correlated with the clearance. Receptor-saturating basiliximab concentrations (≧0.1 μg/mL) were maintained for 38 ± 16 days, which was negatively correlated with the cumulative volume of drained ascites fluid in a week [234]. However, there was no apparent relationship between the incidence or day of onset of acute rejection episodes during CD25 saturation and basiliximab concentration in kidney and liver transplants.

### 9.5 Trastuzumab Pharmacokinetics

Trastuzumab (Herceptin) is a monoclonal antibody for human epidermal growth factor receptor 2 (HER2)-positive metastatic breast cancer (MBC). A phase II study by Baselga et al. demonstrated the safety, efficacy, and pharmacokinetics of trastuzumab monotherapy given as first-line treatment once every 3 weeks in women with HER2-positive MBC [235]. Patients received a loading dose of trastuzumab, 8 mg/kg intravenously and then 6 mg/kg at 3-week intervals until disease progression or patient withdrawal. In total, 105 patients received a median of five cycles of therapy. The overall response rate was 19% and the clinical benefit rate was 33%. Median time to progression was 3.4 months. The most common treatment-related adverse events were rigors, pyrexia, headache, nausea, and fatigue. Median baseline left ventricular ejection fraction was 63%. The mean trough levels of trastuzumab in plasma were lower and the peak levels were higher with 3-weekly trastuzumab compared with weekly treatments. The author concluded that administering higher doses on a three-times-per-week schedule did not compromise the efficacy and safety of trastuzumab in women with HER2-positive MBC, and average exposure was similar to that observed with weekly therapy. In addition, a population pharmacokinetic analysis was performed by Bruno et al. to investigate precise aspects of trastuzumab pharmacokinetics [236]. A nonlinear mixed-effect model was applied for the pharmacokinetic data from phase I, II, and III studies of 476 patients. The phase I study enrolled patients with advanced solid tumors. The phase II and III studies enrolled patients with HER2-positive MBC. Patients in the pivotal phase II and III studies were treated with a 4-mg/kg loading dose of trastuzumab followed by 2 mg/kg weekly for up to 840 days. A two-compartment linear pharmacokinetic model best described the data and accounted for the long-term accumulation observed after weekly administration of trastuzumab. Population estimates for total body clearance, volume of distribution of the central compartment, and terminal half-life of trastuzumab were 0.225 L/day, 2.95 L, and 28.5 days, respectively. Interpatient variabilities in the total body clearance and volume of distribution were 43% and 29%, respectively. The number of metastatic sites, plasma level of extracellular domain of the HER2 receptor, and patient weight were significant baseline covariates for total body clearance, volume of distribution, or both. However, these covariate effects on trastuzumab exposure were modest and not clinically important in comparison with the large interpatient variability of CL. Concomitant chemotherapy (anthracycline plus cyclophosphamide, or paclitaxel) did not appear to influence clearance.

### 9.6 Infliximab Pharmacokinetics

Infliximab (Remicade) is a chimeric monoclonal antibody against tumor necrosis factor (TNF)-alpha that has shown efficacy in Crohn's disease and rheumatoid arthritis with a disease-modifying activity and rapid onset of action [237]. It is administered intravenously, generally in a schedule with initial infusions at 0, 2, and 6 weeks, followed by administration once every 8 weeks. Infliximab is effective in the treatment of patients with moderately to severely active Crohn's disease with an inadequate response to other treatment options or those with fistulizing disease. In combination with methotrexate, infliximab reduced signs and symptoms and

delayed disease progression in patients with active, methotrexate-refractory rheumatoid arthritis and in those with early disease. However, direct pharmacokinetic trials of infliximab in healthy volunteers or patients are not yet available. Infliximab is predominantly distributed to the vascular compartment and eliminated with a t1/2beta between 10 and 14 days. No accumulation was observed when it was administered at intervals of 4 or 8 weeks Methotrexate may reduce the clearance of infliximab from serum [238].

### 9.7 Rituximab Pharmacokinetics

Rituximab (Rituxan) is a B-cell-depleting anti-CD20 chimeric IgG κ monoclonal antibody being investigated for the treatment of rheumatoid arthritis and is approved for the treatment of B-cell lymphoma [239]. Kim et al. investigated pharmacokinetic profiles of rituximab in rabbits following intravitreal administration [240]. After intravitreal injections of 1-mg rituximab, the elimination half-life of rituximab, estimated based on the two compartments, was 4.7 days. Fitting the data to a two-compartment model yielded a clearance from the aqueous humor of 1.2 μL/min. The clearance was less than the reported rate of aqueous humor outflow, indicating that elimination by this route could have been sufficient to account for the disappearance of the drug from the eye. The duration of time over which sustained levels of rituximab were achieved suggests that intravitreal administration warrants further investigation as an approach to treating vitreous and anterior chamber infiltrates in patients with primary intraocular lymphoma. In addition, a population pharmacokinetic analysis of rituximab using a total of 102 patients by Ng et al. [239] demonstrated that body surface area and gender were the most significant covariates for both a total body clearance and a volume of distribution in central compartment. Body surface area alone only explained about 19.7% of the total interindividual variability of the total body clearance. In a simulation study, body surface area-based dosing normalized drug exposure over a wide range of body surface area but did not seem to improve the predictability of rituximab AUC in rheumatoid arthritis patients. Therefore, no rationale for body surface area-based dosing for rituximab in rheumatoid arthritis patients was found.

### 9.8 Palivizumab Pharmacokinetics

Respiratory syncytial virus (RSV) infection represents a major cause of pediatric respiratory hospitalizations. RSV is the most common cause of lower respiratory tract infection in infants. Palivizumab (Synagis) is a humanized monoclonal antibody to the fusion protein of RSV, and is active in animal models for prevention of pulmonary RSV replication [241]. Based on a phase I/II multicenter, randomized, double-blind, placebo-controlled, dose escalation trial by Subramanian et al. in premature infants or infants with bronchopulmonary dysplasia, it was found that the mean half-life of 20 days was comparable with that of other immunoglobulin G preparations. Mean trough serum concentrations 30 days after infusion were 6.8, 36.1, and 60.6 μg/mL for the 3-, 10-, and 15-mg/kg dose groups, respectively. Mean serum concentrations of palivizumab that have been shown to produce a 2-log reduction in pulmonary RSV titer in cotton rats were maintained when 10- or 15-mg/kg palivizumab was given every 30 days to pediatric patients at high risk

for serious RSV disease. In addition, based on another phase I/II multicenter, randomized, double-blind, placebo-controlled, dose escalation trial in 99 healthy children hospitalized with acute RSV infection by Saez-Llornes et al., it was found that mean serum concentrations of palivizumab in the 5- and 15-mg/kg groups, respectively, were 61.2 and 303.4 μg/mL at 60 min and 11.2 and 38.4 μg/mL after 30 days. There were no significant differences in clinical outcomes between placebo and palivizumab groups for either dose. A single 15-mg/kg dose achieved serum palivizumab concentrations above the 25–30-μg/mL concentration associated with 2-log reduction of pulmonary RSV titer in the cotton rat model [242].

## 10 RECOMMENDATIONS FOR FUTURE STUDY

Recombinant DNA technology has produced many proteins and peptides drugs in large scale, and they are becoming increasingly important as therapeutic agents. The successful development of recombinant proteins depends on the proper characterization of their pharmacokinetics and an understanding of the relationship between drug exposure or dose and the pharmacological response. Although pharmacokinetic characterization of proteins is often complicated because of the absence of a specific assay, difficulties exist in identifying metabolites, endogenous circulating concentrations, and the presence of binding proteins. However, nanobiotechnologically derived formulations for peptide/protein drugs also show some new challenges. Numerous methods have been explored for increasing the residence time of proteins in the systemic circulation or oral bioavailability and enhancing the pharmacological effect. Some of the techniques have been employed to include site-specific mutagenesis, polymer modifications, and fusion with immunoglobulins, targeted through the use of liposomes or conjugation to toxins, radionuclides, and other proteins. In the future, these methods should provide adequate formulations of peptide drugs to increase the successful therapeutic efficacy when formulations are administered through the alternative administration routes.

## REFERENCES

1. Walsh G (1998). *Biopharmaceuticals: Biochemistry and Biotechnology*, John Wiley & Sons, Ltd, England, 1998, pp. 1, 41.
2. Reichert J M, Paquette C (2003). Clinical development of therapeutic recombinant proteins. *Biotechn.* 35:176–185.
3. Reichert J M, Paquette C (2003). Therapeutic recombinant proteins: Trends in US approvals 1982 to 2002. *Curr. Opin. Molec. Therapeu.* 5:139–147.
4. Fletcher K (1998). Drug delivery: Strategies and technologies. *Pharmaceut. Sci. Technol.* 1:49–51.
5. Kurzrock R, Rosenblum M G, Sherwin S A, et al. (1985). Pharmacokinetics, single-dose tolerance, and biological activity of recombinant gamma-interferon in cancer patients. *Cancer Res.* 45:2866–2872.
6. Radwanski E, Perentesis G, Jacobs S, et al. (1987). Pharmacokinetics of interferon α-2b in healthy volunteers. *J. Clin. Pharmacol.* 27:432–435.

7. McLennan D N, Christopher J H, Porter E, et al. (2004). Lymphatic absorption is the primary contributor to the systemic availability of epoetin alfa following subcutaneous administration to sheep. *J. Pharmacol. Experim. Therapeut.* 313:345–351.
8. Pillai O, Panchagnula R (2001). Insulin therapies-past, present and future. *Drug Deliv. Ther.* 6:1056–1061.
9. Kennedy F K (1991). Recent developments in insulin delivery techniquescurrent status and future potential. *Drugs.* 42:213–227.
10. Brange J, Ribel U, Hansen J F, et al. (1988). Monomeric insulins obtained by protein engineering and their medical omplications. *Nature.* 333:679–682.
11. Jorgensen S, Vaag A, Langkjaer L, et al. (1989). NovoSol basal: Pharmaco-kinetics of a novel soluble long-acting insulin analogue. *Br. Med. J.* 299:415–419.
12. Vola J P, Owens D R, Dolben J, et al. (1988). Recombinant DNAderived monomeric insulin analogue: Comparison with soluble human insulin in normal subjects. *Br. Med. J.* 297:1236–1239.
13. Brange J, Owens D R, Kany S, et al. (1990). Monomeric insulins and their experimental and clinical implications. *Diabetescare.* 13:923–954.
14. Kang S, Owens D R, Vola J P, et al. (1990). Comparison of insulin analogue B9Asp B27Glu and soluble human insulin in insulin treated diabetes. *Lancet.* 335:303–306.
15. Halban P A, Berger M, Gjinovci A, et al. (1981). Biologic activity and pharmacokinetics of biosynthetic human insulin in the rat. *Diabetescare.* 4:238–243.
16. Brogden R N, Heel R C (1987). Human insulin: A review of its biological activity, pharmacokinetics and therapeutic use. *Drugs.* 34:350–371.
17. Turnheim K, Waldhausl W K (1988). Essentials of insulin pharmacokinetics. *Wien. Klin. Wochenschr.* 100:65–72.
18. Kang S, Brange J, Burch A, et al. (1991). Subcutaneous insulin absorption explained by insulin's pharmacochemical properties. Evidence from absorption studies of soluble human insulin and insulin analogues in humans. *Diabetescare.* 14:942–948.
19. Hoffman A, Ziv E (1997). Pharmacokinetic considerations of new insulin formulations and routes of administration. *Clin. Pharmacokinet.* 33:258–301.
20. Osterberg O, Erichsen L, Ingwersen S H, et al. (2003). Pharmacokinetic and pharmacodynamic properties of insulin aspart and human insulin. *J. Pharmacokinet. Pharmacodyn.* 30:221–235.
21. Becker R H, Frick A D, Burger F, et al. (2005). Insulin gulisine, a new rapid-acting insulin analogue, displays a rapid time-action profile in obese nondiabetic subjects. *Exp. Clin. Endocrinol. Diabetes.* 113:435–443.
22. Homko C, Deluzio A, Jimenz C, et al. (2003). Comparison of insulin aspart and lispro: Pharmacokinetic and metabolic effects. *Diabetescare.* 26:2027–2031.
23. Chiou G C Y, Chuang C Y, Chang M S (1989). Systemic delivery of insulin through eyes to lower the glucose concentration. *J. Ocular. Pharmacol.* 5:81–91.
24. Starokadomskyy P L, Dubey I Y (2006). New absorption promorter for the buccal delively: preparation and characterization of lysalbinic acid. *Int. J. Pharm.* 308:149–154.
25. Aungst B J, Roger N J, Shefter T S (1988). Comparison of nasal, rectal buccal, sublingual, and intramuscular insulin efficacy and the effect of a bile salt absorption promoter. *J. Pharmacol. Exp. Ther.* 244:23–27.
26. Degim Z, Degim T, Acarturk F, et al. (2005). Rectal and vaginal administration of insulin-chitosan formulations: An experimental study in rabbits. *J. Drug. Target.* 13:563–572.

27. Spangler R S (1990). Insulin administration via liposomes. *Diabetescare.* 13:911–922.
28. Kennedy F K (1991). Recent developments in insulin delivery thechniquescurrent states and future. *Drugs.* 42:213–227.
29. Chetty D J, Chien Y W (1998). Novel methods of insulin delively: An update. *Crit. Rev. Ther. Drug. Carrier. Syst.* 15:629–670.
30. Sakagami M (2004). Insulin disposition in the lung following oral inhalation in humans. A meta-analysis of its pharmacokineticsclin. *Pharmacokinet.* 43:539–552.
31. Henry R R, Mudaliar S R, Howland W C III, et al. (2003). Inhaled insulin using the AERx Insulin diabetes management system in healthy and asthmatic subjects. *Diabetescare.* 26:764–769.
32. (2004). Insulin inhalation—Pfizer/Nektar therapeutic: HMR 4006, inhaled PEG-insulin—Nektar, PEGylated insurin—Nektar. *Drugs. R.D.* 5:166–170.
33. Illum L, Davis S S (1992). Intranasal insulin. Clinical pharmacokineticsclin. *Pharmacokinet.* 23:30–41.
34. Drejer K, Vagg A, Bech K, et al. (1992) Intranasal administration of insulin with phospholipid as absorption enhancer: Pharmacokinetics in normal subjects. *Diabet. Med.* 9:335–340.
35. Hilsted J, Madsbad S, Hvidberg A, et al. (1995) Intranasal insulin therapy: the clinical realities. *Diabetologia.* 38:680–684.
36. Yih T C, Al-Fandi M (2006). Engineered nanoparticles as precice drug delivery system. *J. Cell. Biochem.* 11:107–112.
37. Dyer A M, Hinchcliffe M, Watts P, et al. (2002). Nasal delivery of insulin using novel chitosan based formulations: A comparative study in two animal models between simple chitosan formulations and chitosan nanoparticles. *Pharm. Res.* 19:998–1008.
38. Jelkmann W (1992). Erythropoietin: structure, control of production, and function. *Physiol. Rev.* 72:449–489.
39. Lappin T R, Rich I N (1996). Erythropoietin—the first 90 years. *Clin. Lab. Haematol.* 18:137–145.
40. Markham A, Bayson H M (1995). Epoetin alfa: A review of its pharmacodynamic and pharmacokinetic properties and therapeutic use in nonrenal applications. *Drugs.* 49:232–254.
41. Thatcher N (1998). Management of chemotherapy-induced anemia in solid tumors. *Semi. Oncol.* 25:23–26.
42. Goldberg M A, McCutchen J W, Jove W, et al. (1996). A safty and efficacy comparison study of two dosing regimens of epoetin alfa in patients undergoing major orthopedic surgery. *Am. J. Orthop.* 25:544–552.
43. Bommer J, Alexiou U, Muller-Buhl E, et al. (1987). Recombinant human erythropoietin therapy in haemodialysis patients-dose determination and clinical experience. *Nephrol. Dial. Transpl.* 2:238–242.
44. Casati S, Passerini P, Campise M R, et al. (1987). Benefits and risks of protracted treatment with human recombinant erythropoietin in patients having haemodialysis. *Br. Med. J.* 295:1017–1020.
45. Eshbach J W, Egrie J C, Downing M R, et al. (1987). Correction of the anemia of end-stage renal disease with recombinant human erythropoietin. *N. Engl. J. Med.* 316:73–78.
46. Winearls C G, Oliver D O, Pippard M J, et al. (1986). Effect of human erythropoietin derived from recombinant DNA on the anaemia of patients maintained by chronic haemodialysis. *Lancet.* 2:1175–1178.

47. Boelaert J R, Schurgers M L, Matthys E G, et al. (1989). Comparative pharmacokinetics of recombinant erythropoietin administered by the intravenous, subcutaneous, and intraperitoneal routs in continuous ambulatory peritoneal dialysis patients. *Perit. Dial. Int.* 9:95–98.
48. Hughes R T, Cotes P M, Oliver D O, et al. (1989). Correction of the anaemia of chronic renal failure with erythropoietin: pharmacokinetic studies in patients on haemodialysis and CAPD. *Contrib. Nephrol.* 76:122–130.
49. Storring P L, Tiplady R J, Gaines Das R E, et al. (1998). Epoetin alfa and beta differ in their erythropoietin isoform compositions and biological properties. *Br. J. Haematol.* 100:79–89.
50. Halstenson C E, Marcres M, Katz S A, et al. (1991). Comparative pharmacokinetics and pharmacodynamics of epoetin alfa and epoetin beta. *Clin. Pharmacol. Ther.* 50:702–712.
51. Casati S, Passerini P, Campise M R, et al. (1987). Benefits and risks of protracted treatment with human recombinant erythropoietin in patients having haemodialysis. *Br. Med. J.* 295:1017–1020.
52. Eschbach J W, Egrie J C, Downung M R, et al. (1987). Correction of the anemia of end-stage renal disease with recombinant human erythropoietin: Results of a combined phase I and Iiclinical trial. *N. Engl. J. Med.* 316:73–78.
53. Flaharty K K, Caro J, Erslev A, et al. (1990). Pharmacoldnetics and erythropoietic response to human recombinant erythropoietin in healthy men. *Clin. Pharmacol. Ther.* 47:557–564.
54. Salmonson T, Danielson B G, Wikström B (1990). The pharmacokinetics of recombinant human erythropoietin after intravenous and subcutaneous administration to healthy subjects. *Br. J. Clin. Pharmacol.* 29:709–713.
55. McMahon F G, Vargas R, Ryan M, et al. (1990). Pharmacokinetics and effects of recombinant human erythropoietin after intravenous and subcutaneous injections in healthy volunteers. *Blood*. 76:1718–1722.
56. Ramakrishnan R, Cheung W K, Wacholtz M C (2004). Pharmacokinetic and pharmacodynamic modeling of recombinant human erythropoietin after single and multiple doses in healthy volunteers. *J. Clin. Pharmacol.* 44:991–1002.
57. Ramakrishnan R, Cheung W K, Farrell F, et al. (2003). Pharmacokinetic and pharmacodynamic modeling of recombinant humanerythropoietin after intravenous and subcutaneous dose administration in cynomolgus monkeys. *J. Pharmacol. Exp. Ther.* 306:324–331.
58. Cheung W K, Goon B L, Guilfoyle M C, et al. (1998). Pharmacokinetics and pharmacodynamics of recombinant human erythropoietin after single and multiple subcutaneous doses to healthy subjectsclin. *Pharmacol. Ther.* 64:412–423.
59. Cheung W, Minton N, Gunawardena K (2001). Pharmacokinetics and pharmacodynamics of epoetin alfa once weekly and three times weekly. *Eur. J. Clin. Pharmacol.* 57:411–418.
60. Krzyanski W, Jusko W J, Wacholtz M C, et al. (2005). Pharmacokinetic and pharmacodynamic modeling of recombinant human erythropoietin after multiple subcutaneous doses in healthy subjects. *Eur. J. Pharm. Sci.* 26:295–306.
61. Kndler J, Eckardt K-U, Ehmer B, et al. (1989). Single-dose pharmacokinetics of recombinant human erythropoietin in patients with various degrees of renal failure. *Nephrol. Dial. Transplant.* 4:345–349.
62. Brockmöllerm J, Köchling J, Weber W, et al. (1992). The pharmacokinetics and pharmacodynamics of recombinant human erythropoietin in haemodialysis patients. *Br. J. Clin. Pharmacol.* 34:499–508.

63. Gladziwa U, Klotz U, Bäumer K, et al. (1993). Pharmacokinetics of epoetin (recombinant human erythropoietin) after long term therapy in patients undergoing haemodialysis and haemofiltration. *Clin. Pharmacokinet.* 25:145–153.
64. Taylor III C A, Abdel-Rahman E, Zimmerman S W, et al. (1996). Clinical pharmacokinetics during continuous ambulatory peritoneal dialysis. *Clin. Pharmacokinet.* 31:293–308.
65. Macdougall I C, Roberts D E, Coles G A, et al. (1991). Clinical pharmacokinetics of epoetin (recombinant human erythropoietin). *Clin. Pharmacokinet.* 20:99–113.
66. Korbet S M (1993). Anemia and erythropoietin in hemodialysis and continuous ambulatory peritoneal dialysis. *Kidney. Int.* 43:S111–S119.
67. Bommer J, Barth H P, Schwöbel B (1990). RhEPO treatment of anemia in uremic patientscontrib. *Nephrol.* 87:59–67.
68. Lubrich-Birkner I, Schollmeyer P, Steinhauer H B (1990). One year experience with subcutaneous human erythropoietin in CAPD: Correction of renal anemia and increased ultrafiltration. *Adv. Perit. Dial.* 6:302–307.
69. Miranda B, Selgasc–Riňon R, et al. (1990). Treatment of the anemia with human recomnbinant erythropoietin in CAPD patients. *Adv. Perit. Dial.* 6:296–301.
70. Piraino B, Johnston J R (1990). The use of subcutaneous erythropoietin in CAPD patientsclin. *Nephrol.* 33:200–202.
71. Cheng I K P, Cy C, Chan M K, et al. (1991). Correction of anemia on patients on continuous ambulatory peritoneal dialysis with subcutaneous recombinant erythropoietin twice a week: A long-term study. *Clin. Nephrol.* 35:207–212.
72. Boelaert J R, Schurgers M L, Matthys E G, et al. (1989). Comparative pharmacokinetics of recombinant erythropoietin administered by the intravenous, subcutaneous, and intraperitoneal routes in continuous ambulatory peritoneal dialyasis patients. *Perit. Dial. Int.* 9:95–98.
73. Kampf D, Kahl A, Passlick J, et al. (1989). Single-dose kinetics of recombinant human erythropoietin after intravenous, subcutaneous and intraperitoneal administration. *Contrib. Nephrol.* 76:106–111.
74. Macdougall I C, Roberts D E, Nuebert P, et al. (1989). Pharmac kinetics of recombinant human erythropoietin in patients on continuous ambulatory peritoneal dialysis. *Lancet. I.* 8635:425–427.
75. Stockenhuber F, Loibl U, Gottsauner-Wolf M, et al. (1991). Pharmcokinetics and dose response after intravenous and subcutaneous administration of recombinant erythropoietin in patients on regular haemodialysis treatment or continuous ambulatory peritoneal dialysis. *Nephron.* 59:399–402.
76. Icardi A, Paoletti E, Molinelli G (1990). Efficacy of recombinant erythropoietin after subcutaneous or intraperitoneal administration to patients on CAPD. *Adv. Peril. Dial.* 6:292–295.
77. Frenken L A M, Struijk D G, Coppens P J W, et al. (1992). Intraperitoneal administration of recombinant human erythropoietin. *Perit. Dial. Int.* 12:378–383.
78. Nasu T, Mitui H, Shinohara Y, et al. (1992). Effect of erythropoietin in continuous ambulatory peritoneal dialysis patientscomparison between intravenous and intraperitoneal administration. *Perit. Dial. Int.* 12:373–377.
79. Lui S F, Chung W W M, Leung C B, et al. (1990). Pharmacokinetics and pharmacodynamics of subcutaneous and intraperitoneal administration of recombinant human erythropoietin in patients on continuous ambulatory peritoneal dialysisclin. *Nephrol.* 33(1):47–51.

80. Huang T P, Lin C Y (1995). Intraperitoneal recombinant human erythropoietin therapy: Influence of the duration of continuous ambulatory peritoneal dialysis treatment and peritonitis. *Am. J. Nephrol.* 15:312–317.
81. Bargman J M, Jones J E, Petro J M (1992). The pharmacokinetics of intraperitoneal erythropoietin administered undiluted or diluted in dialysate. *Perit. Dial. Int.* 12:369–372.
82. Ateshkadi A, Johnson C A, Oxton L L, et al. (1993). Pharmacokinetics of intraperitoneal, intravenous, and subcutaneous recombinant human erythtopoietin in patients on continuous ambulatory peritoneal dialysis. *Am. J. Kidney. Dis.* 21:635–642.
83. Evans J H C, Brocklebank J T, Bowmer C J, et al. (1991). Pharmacokinetics of recombinant human erythropoietin in children with renal failure. *Nephrol. Dial. Transpl.* 6:709–714.
84. Montini G, Zacchello G, Perfurno F, et al. (1993). Pharmacokinetics and hematologic response to subcutaneous administration of recombinant human erythropoietin in children undergoing long-term peritoneal dialysis: A multicenter study. *J. Pediatr.* 122:297–302.
85. Valentini R P, Sedman A B, Smoyer W E, et al. (1995). The safety and efficacy of intraperitoneal (IP) erythropoiethl (EPO) in children on CCPD. *J. Am. Soc. Nephrol.* 6:514.
86. Elliott S, Lorenzini T, Asher S, et al. (2003). Enhancement of therapeutic protein in vivo activities through glycoengineering. *Nat. Biotehnol.* 21:414–421.
87. Egrie J C, Brownw J K (2001). Development and characterization of novel erythropoiesis stimulating protein (NESP). *Br. J. Cancer.* 84:3–10.
88. Egrie J C, Dwyer E, Brown J K, et al. (2003). Darbepoetin alfa has a longer circulating half-life and greater in vivo potency than recombinant human erythropoietin. *Exp. Hematol.* 31:290–299.
89. Macdougall I C, Gray S J, Elsto O, et al. (1999). Pharmacokinetics of novel erythropoiesis stimulating protein compared with epoetin alfa in dialysis patients. *J. Am. Soc. Nephrol.* 10:2392–2395.
90. Allon M, Kleinman K, Walczyk M, et al. (2002). Pharmacokinetics and pharmacodynamics of darbepoetin alfa and epoetin in patients undergoing dialysisclin. *Pharmacol. Ther.* 72:546–555.
91. Lerner G, Kale A S, Wardy B A, et al. (2002). Pharmacokinetics of darbepoetin alfa in pediatric patients with chronic kidney disease. *Pediatr. Nephrol.* 17:933–937.
92. Burgess A W, Metcalf D (1980). The nature and action of granulocytemacrophage colony stimulating factors. *Blood.* 56:947–958.
93. Metcalf D (1984). *The Hemopoietic Colony Stimulationg Factors.* Elsevier, Amsterdam, The Netherlands.
94. Begley C G, Metcalf D, Lopez A F, et al. (1985). Fractionated populations of normal human marrow cells respond to both human coloney-stimulating factors with granulocyte-macrophage activity. *Exp. Hematol.* 13:956–962.
95. Metcalf D (1986). The molecular biology and functions of the granulocytemacrophage coloney-stimulating factors. *Blood.* 67:257–267.
96. Nicola N A, Metcalf D, Matsumoto M, et al. (1983). Purification of a factor indicating differentiation in murine myelomomocyte leukemia cells: Identification as granulocyte colony stimulating factor (G-CSF). *J. Biol. Chem.* 258:9017–9023.
97. Nomura H, Imazeki I, Oheda M, et al. (1986). Purification and characterization of human granulocyte colony-stimulating factor (G-CSF). *EMBO J.* 5:871–876.

98. Zsebo K, Cohen A M, Murdock D C, et al. (1986). Reconbinant human granulocyte colony syimulating factor: molecular and biological characterization. *Immunobiol.* 172:175–184.

99. Welte K, Platzer E, Lu L, et al. (1985). Prification and biochemical characterization of human pluripotent hematopoietic coloney-stimulating factor. *Proc. Natl. Acad. Sci. U.S.A.* 82:1526–1530.

100. Souza L M, Boone T C, Gabrilove J, et al. (1986). Recombinant human granulocyte colony-stimulating factor: effects on normal and leukemic myeloid cells. *Science.* 232:61–65.

101. Tsutiya M, Asano S, Kaziro Y, et al. (1986). Isolation and characterization of the cDNA for murine granulocyte coloney-stimulating factor. *Proc. Natl. Acad. Sci. U.S.A.* 83:7633–7637.

102. Nagata S, Tsuchiya M, Asano S, et al. (1986). Molecular cloning and expression of cDNA for human granulocyte colony-stimulating factor. *Nature.* 319:415–418.

103. Burgess A W, Mrtcalf D (1977). Serum half-life and organ distribution of radiolabeled colony stimulating factor in mice. *Exp. Hematol.* 5:456–464.

104. Nicola N A, Metcalf D, Matsumoto M, et al. (1983). Purification of a factor inducing differentiation in murine myelomonocytic leukemia cells: Identification as granulocyte colony-stimulating factor (G-CSF). *J. Biol. Chem.* 258:9017–9023.

105. Souza L M, Boone T C, Gabrilove J, et al. (1986). Recombinant human granulocyte colony-stimulating factor: Effects of normal and leukemic myeloid cells. *Science.* 232:61–65.

106. Morstyn G, Lieschke G J, Sheridan W, et al. (1989). Clinical experience with recombinant human granulocyte colony-stimulating factor and granulocyte macrophage coloney-stimulating factor. *Semin. Hematol.* 26:9–13.

107. Choen A M, Zsebo K M, Inoue H, et al. (1987). In vivo stimulation of granulopoiesis by recombinant human granulocyte colony-stimulating factor. *Proc. Natl. Acad. Sci. U.S.A.* 84:2484–2488.

108. Tanaka H, Kaneko T (1990). Pharmacokinetics of recombinant human granulocyte colony-stimulating factor in the rat. *Drug. Metabol. Dispos.* 19:200–203.

109. Tanaka H, Tokiwa T (1989). Pharmacokinetics of recombinant human granulocyte colony stimulating factor studied in the rat by a sandwich enzyme-linked immunosorbent assay. *J. Pharmacol. Exp. Ther.* 255:724–729.

110. Tanaka H, Okada Y, Mayumi K, et al. (1989). Pharmacokinetics and pharmacodynamics of recombinant human granulocyte-coloney stimulating factor after intravenous and subcutaneous administration in the rat. *J. Pharmacol. Exp. Ther.* 251:1199–1989.

111. Tanaka H, Kaneo T (1991). Sex differences in the pharmacokinetics of recombinant human granulocyte colony-stimulating factor in the rat. *Drug. Metabol. Dispos.* 19:1034–1039.

112. Tanaka H, Tokiwa T (1990). Influence of renal and hepatic failure on the pharmacokinetics of recombinant human granulocyte colony-stimulating factor (KRN8601) in the rat. *Cancer Res.* 50:6615–6619.

113. Layton J E, Hockman H, Sheridan W P, et al. (1989). Evidence for a novel in vivo control mechanism of granulopoiesis: Maturee cell-related control of a regulatory growth factor. *Blood.* 4:1303–1307.

114. Ferraiolo B L, Mohler M A, Gloff C A (1992). *Protein Pharmacokinetics and Metabolism.* Plenum, New York.

115. Hayashi N, Kinoshita H, Yukawa E, et al. (1999). Pharmacokinetic and pharmacodynamic analysis of subcutaneous recombinant human granulocyte coloney stimulating factor (lenograstim) administration. *J. Clin. Pharmacol.* 39:583–592.
116. Hayashi N, Aso H, Higashida M, et al. (2001). Estimation of rhG-CSF absorption kinetics after subcutaneous administration using a modified Wagner-Nelson method with a nonlinear elimination model. *Eur. J. Pharmaceu. Sci.* 13:151–158.
117. Wang B, Ludden T M, Cheung E N, et al. (2001). Population pharmacokinetic-pharmacodynamic modeling of filgrastim (r-metHuG-CSF) in healthy volunteers. *J. Pharmacokinet. Pharmacodyn.* 28:321–342.
118. Hernandez-Bernal F, Garacia-Garcia I, Gonzalez-Delgado C A, et al. (2005). Bioequivalence of two recombinant granulocyte colony-stimulating factor: Formulations in healthy male volunteers. *Biopharmac. Drug. Dispos.* 26:151–159.
119. Yowell S L, Blackwell S (2002). Novel effects with polyethylene glycol modified pharmaceuticals. *Cancer Treat. Rev.* 28(suppl A):2–6.
120. Molineux G (2002). Pegylation: Engineering improved pharmaceutical for enhanced therapy. *Cancer Treat. Rev.* 28(suppl A):13–16.
121. van Der Auwera P, Platzer E, Xu Z X, et al. (2001). Pharmacodynamics and pharmacokinetics of single doses of subcutaneous pegylated human G-CSF mutant (Ro 25-8315) in healthy volunteerscomparison with single and multiple daily doses of filgrastim. *Am. J. Hematol.* 66:245–251.
122. Maeda H, Nakagawa T, Adachi N, et al. (2003). Design of long-acting formulation of protein drugs with a double-layer structure and its application to rhG-CSF. *J. Control. Rel.* 91:281–297.
123. Eiamtrakarn S, Itoh Y, Kishimoto J, et al. (2002). Gastrointestinal mucoadhesive patch system (GI-MAPS) for oral administration of G-CSF, a model protein. *Biomat.* 23:145–152.
124. Agu R U, Ugwoke M I, Armand M, et al. (2001). The lung as a route for systemic delivery of the therapeutic proteins. *Respir. Res.* 2:198–209.
125. Niven R W, Loft F D, Cribbs J M (1993). Pulmonary absorption of recombinant methionyl human granulocyte coloney stimulating factor (r-huG-CSF) after intratracheal instillation to the hamster. *Pharm. Res.* 10:1604–1610.
126. Niven R W, Loft F D, Ip A Y, et al. (1994). Pulmonary delivery of powders and solutionscontaining granulocyte coloney stimulating factor (r-huG-CSF) to the rabbit. *Pharm. Res.* 11:1101–1109.
127. Miyamoto M, Natsume H, Satoh I, et al. (2001). Effect of poly-L-arginine on the nasal absorption of FITC-dextran of different molecular weights and recombinant human granulocyte coloney-stimulating factor (rhG-CSF) in rats. *Int. J. Pharmaceu.* 226:127–138.
128. Cantell K, Pyhala L (1976). Pharmacokinetics of human leukocyte interferon. *J. Infect. Dis.* 133(suppl A):A6–A12.
129. Blatt L M, Davis J M, Klein S B, et al. (1996). The biologic activity and molecular characterization of a novel synthetic interferon-alpha speciesconsensus interferon. *J. Interferon Cytokine Res.* 16:489–499.
130. Kohase M (2003). New IFN products for clinical use. *Sogo. Rinsyo.* 52:2540–2545.
131. Higuchi Y, Hashida M (2003). Pharmacokinetics of interferon. *Sougo. Rinsyo.* 52: 2449–2505.
132. Anderson P M, Hanson D C, Hasz D E, et al. (1994). Cytokins in liposomes: Preliminary studies with IL-1, IL2, IL6, GM-CSF, and interferon-γ. *Cytokine.* 6:92–101.

133. Goddel D V, Yelverton E, Ullrick A, et al. (1980). Human leukocyte interferon produced by E. coli is biologically active. *Nature*. 287:411–416.

134. Gutterman J U, Fine S, Quesada J, et al. (1984). Recombinant leukocyte A interferon: Pharmacokinetics, single-dose tolerance, and biologic effects in cancer patients. *Cancer Res.* 44:4164–4171.

135. Hawkins M J, Borden E C, Merritt J A, et al. (1984). Comparison of the biological effects of two recombinant human interferon alpha (rA and rD) in humans. *J. Clin. Oncol.* 2:221–226.

136. Krown S E, Real F X, Cunningham-Rundles S, et al. (1983). Preliminary observations on the effect of recombinant leukocyte A interferon in homosexual men with Kaposi's sarcoma. *N. Engl. J. Med.* 308:1071–1076.

137. Cantell K, Fiers W, Hirvones S, et al. (1984). Circulating interferon in rabbits after simultaneous intramuscular administration of human alpha and gamma interferon. *J. Interferon. Res.* 4:291–292.

138. Collins J M, Riccardi R, Trown P, et al. (1985). Plasma and cerebrospinal fluid pharmacokinetics of recombinant interferon alpha A in monkeyscomparison of intravenous, intermuscular and intraventricular delivery. *Cancer Drug. Del.* 2:247–253.

139. Gibson D M, Cotler S, Speigel H E, et al. (1985). Pharmacokinetics of recombinant leukocyte A interferon following various routes and modes of administration to the dog. *J. Interferon Res.* 5:403–408.

140. Sarkar F H (1982). Pharmacokinetic comparison of leukocyte and Escherrichia coli-derived human interferon type alpha. *Antiviral Res.* 2:103–106.

141. Heremans H, Billiau A, DeSorner P (1980). Interferon in experimental viral interferon in mice: tissue interferon levels resulting from virus infection and from exogenous interferon therapy. *Infect. Immun.* 30:513–522.

142. Bocci V, Muscettola M, Naldini A (1986). The lymphatic route IV: Pharmacokinetics of human recombinant interferon alpha and natural interferon beta administered intradermally in rabbits. *Gen. Pharmacol.* 17:93–96.

143. Naito S, Tanaka S, Mizuno M, et al. (1984). Concentrations of human interferons alpha and beta in rabbit body fluids. *Int. J. Pharm.* 18:117–125.

144. Wills R J, Spiegel H E, Soike K F (1984). Pharmacokinetics of recombinant leukocyte A interferon following iv infusion and bolus, im, and po administration to African green monkeys. *J. Interferon Res.* 4:399–409.

145. Gomi K, Morimoto M, Inoue A, et al. (1984). Pharmacokinetics human recombinant interferon-beta in monkeys and rabbits. *Gann.* 75:292–300.

146. Maitani Y, Igawa T, Machida Y, et al. (1986). Intranasal administration of beta-interferon in rabbits. *Drug. Design Del.* 1:65–70.

147. Weck P K, Shalaby M R, Apperson S, et al. (1982). Comparative biological properties of human alpha, beta and gamma INF's derived from bacteria. *Third International Congress for Interferon Research*, Miami, FL.

148. Bohoslawec O, Trown P W, Wills R J (1986). Pharmacokinetics and tissue distribution of recombinant human A,D,A/D(Bgl) and I interferons and mouse alpha-interferon in mice. *J. Interferon Res.* 6:207–213.

149. Abreu S L (1983). Pharmacokinetics of rat fibroblast interferon. *J. Pharmacol. Exp. Ther.* 226:197–220.

150. Bino T, Edery H, Gertler A, et al. (1982). Involvement of the kidney in catabolism of human leukocyte interferon. *J. Gen. Virol.* 59:39–45.

151. Bocci V, Pacini A, Muscettola M, et al. (1981). Renal metabolism of rabbit serum interferon. *J. Gen. Virol.* 55:297–304.
152. Tokazewski-Chen S A, Marafino B J Jr, Stebbing N (1983). Effects of nephrectomy on the pharmacokinetics of variousclosed human interferons in rats. *J. Pharmacol. Exp. Ther.* 227:9–15.
153. Bocci V, Pacini A, Babdinelli L, et al. (1982). The role of liver in the catabolism of human alpha-and beta-interferon. *Gen. Virol.* 60:397–400.
154. Perkinsdon A, Lasker J, Kramer M J, et al. (1982). Effects of three recombinant human leukocyte interferons on drug metabolism in mice. *Drug. Metab. Dispos.* 10:579–585.
155. Wells R J, Weck P K, Baehner R L, et al. (1988). Interferon in children with recurrent acute lymphocytic leukemia: A phase I study of pharmacokinetics and tolerance. *J. Interferon Res.* 8:309–318.
156. Liberati A M, Biscottini B, Fizzotti M, et al. (1989). A phase I study of human natural interferon-beta in cancer patients. *J. Interferon Res.* 9:339–348.
157. Vadhan-Raj S, Nathan C F, Sherwin S A, et al. (1986). Phase I trail of recombinant interferon gamma by 1-hr iv infusion. *Cancer Treat Rep.* 70:609–614.
158. Bornemann L D, Speigel H E, Dziewanowska Z E, et al. (1985). Intravenous and intramuscular pharmacokinetics of recombinant leukocyte A interferon. *Eur. J. Clin. Pharmacol.* 28:469–471.
159. Hawkins M J, Borden E C, Merritt J A, et al. (1984). Comparison of the biological effects of two recombinant human interferon alpha (rA and rD) in humans. *J. Clin. Oncol.* 2:221–226.
160. Thompson J A, Cox W W, Lindgren C G, et al. (1987). Subcutaneous recombinant gamma interferon in cancer patients: toxicity, pharmacokinetics and immunomodulatory effects. *Cancer Immunol. Immunother.* 25:47–53.
161. Gutterman J U, Fine S, Quesada J, et al. (1982). Recombinat leukocyte A interferon: Pharmakokinetics, single-dose tolerance, and biologic effects in cancer patients. *Ann. Intern. Med.* 96:549–556.
162. Smith R A, Kingsubry D, Alksne J, et al. (1985). Distribution of alpha interferon in serum and cerebrospinal fluid after systemic administration. *Clin. Pharmacol. Ther.* 37:85–88.
163. Priestman T J, Johnston M, Whiteman P D (1982). Peameability observations on the pharmacokinetics of human lymphoblastoid interferon given by intramuscular injection. *Clin. Oncol.* 8:265–269.
164. Ferraiolo B L, Mohler M A, Gloff C A (1992). Protein pharmacokinetics and metabolism, in Pharmaceutical Biotechnology, vol. 1. Plenum Press, New York, p. 136.
165. Nakamura H (2005). Pegylated interferon alfa-2b formulation. *Farmacia.* 41:769–761.
166. Reddy K R (2004). Development and pharmacokinetics and pharmacodynamics of pegylated interferon alfa-2a (40 kDa). *Semin. Liver Dis.* 24:33–38.
167. Pedder S C (2003). Pegylation of interferon alpha: Structural and pharmacokinetic properties. *Semin. Liver Dis.* 23:19–22.
168. Molineux G (2002). Pegylation: Engineering improved pharmaceuticals for enhanced therapy. *Cancer Treat Rev.* 28:13–16.
169. Baker D E (2001). Pegylated interferons. *Rev. Gastroenterol. Disord.* 1:87–99.
170. Crawford J (2002). Clinical uses of pegylated pharmaceuticals in oncology. 28:7–11.
171. Nalin C, Rosen P (1995). U.S. Patent 5, 382,657.

172. Bailon P, Palleroni A, Schaffer C A, et al. (2001). Rational design of a potent, long-lasting form of interferon: A 40 kDa balanced polyethylene glycol-conjugated interferon alpha-2a for the treatment of hepatitisc. *Bioconjug. Chem.* 12:195–202.
173. Caliceti P (2004). Pharmacokinetics of pegilated interferons: What is misleading? *Dig. Liver Dis.* 36:S334–S339.
174. Ramon J, Saez V, Baez R, et al. (2005). PEGylated interferon-α2b: A branced 40 K polyethylene glycol derivative. 22:1374–1386.
175. Zeuzem S, Welsch C, Herrmann E (2003). Pharmacokinetics of peginterferons. *Semin. Liver Dis.* 23:23–28.
176. Sanchez A, Tobio M, Gonzalez L, et al. (2003). Biodegradable micro- and nanoparticles as long termdelively vehicles for interferon-alpha. *Eur. J. Pharmaceu. Sci.* 18: 221–229.
177. Peleg-Shulman T, Tsubery H, Mironchik M, et al. (2004). Reversible PEGylation: A novel technology to release native interferon α2 over a prolonged time period. *J. Med. Chem.* 47:4897–4904.
178. Ito Y, Tosh B, Togashi Y, et al. (2005). Absorption of interferon alpha from patches in rats. *J. Drug. Target.* 13:383–390.
179. Sigel M B, Vanderlaan W P, Kobrin M S, et al. (1982). The biological half-life of human grouth hormone and a biologically active 20,000-dalton variant in mouse blood. *Endocrin. Rescommun.* 9:67–77.
180. Jorgensen K D, Monrad J D, Brondum L, et al. (1988). Pharmacokinetics of biosynthetic and pituitary human growth hormones in rats. *Pharmacol. Toxicol.* 63:129–134.
181. Duby A K, Hanukoglu A, Hansen B C, et al. (1988). Metabolic clearance rates of synthetic human growth hormone in lean and obese male rhesus monkeys. *J. Clin. Endoc. Metab.* 67:1064–1067.
182. Postel-Vinay M C, Kayseer C, Desbuquios B (1982). Fate of injected human growth hormone in the female rat in vivo. *Endocrinol.* 111:244–251.
183. Mordenti J, Chen S A, Moore J A, et al. (1991). Interspecies scaling og clearance and volume of distribution and for five therapeutic proteins. *Pharm. Res.* 8:1351–1359.
184. Josefsberg Z, Posner B I, Patel B, et al. (1979). The uptake of prolactin into female rat liver. Concentration of intact hormone in the Golgi apparatus. *J. Biol. Chem.* 254:209–214.
185. Leung D W, Spencer S A, Cachianes G, et al. (1987). Grouth hormone receptor and serum binding protein: Purification, cloning and expression. *Nature.* 330:537–543.
186. Baumann G, Shaw M A (1990). Plasma transport of the 20,000 dalton variant of human growth hormone (20 K): Evidence for a 20 K-specific binding site. *J. Clin. Endocrinal. Metab.* 71:1339–1343.
187. Cook D M, Biller B M K, Vance M L, et al. (2002). The pharmacokinetic and pharmacodynamic characteristics of a long-acting growth horomone (GH) preparation (Nutropin Depot) in GH-deficient adults. *J. Clin. Endocrinol. Metabol.* 87:4508–4514.
188. Lee H J, Rriley G, Johnson O, et al. (1997). In vivo characterization of sustained-release formulations of Human Growth Hormone. *J. Pharmacol. Exp. Ther.* 281:1431–1439.
189. Brodbeck K J, Pushpala S, McHugh A J (1999). Sustained release of human growth hormone from PLGA solution. *Depos. Pharmaceu. Res.* 16:1825–1829.
190. Maeda M, Kadota K, Kajihara M, et al. (2001). Sustained release of human growth hormone (hGH) from collagen film and evaluation of effect on wound healing in db/db mice. *J. Control. Rel.* 77:261–272.

191. Kim M R, Park T G (2002). Temperature-responsive and degradable hyaluronic acid/pluroic composite hydrogels for controlled release of human growth hormone. *J. Control Rel.* 80:69–77.
192. Hahn S K, Kim S J, Kim M J, et al. (2004). Characterization and in vivo study of sustained-release formation of human growth hormone using sodium hyaluronate. *Pharmaceu. Res.* 21:1374–1391.
193. Kim S J, Hahn S K, Kim M J, et al. (2005). Development of a novel sustained release formulation of recombinant human growth hormone using sodium hyaluronate microparticles. *J. Controlled Rel.* 104:323–335.
194. Leitner V M, Guggi D, Krauland A H, et al. (2004). Nasal delivery of human growth hormone: In vitro and in vivo evaluation of a thiomer/glutathione microparticulate delivery system. *J. Controlled Rel.* 100:87–95.
195. Adjei A, Love S, Jhonson E, et al. (1993). Effect of formation adjuvants on gastrointestinal adsorption of leuprolide acetate. *J. Drug. Target.* 1:251–258.
196. Fu lu M, Reiland T (1994). Compositions and method for the sublingual or buccal administration of therapeutic agents. U.S. Patent 5, 248,657.
197. Kuret J A, Murad F (1992). Adenohypophyseal hormones and related sub-stances. In A Goodman, T W Gilman, A S Rall, et al. (eds.), *The Pharmacological Basis of Therapeutics*: McGraw-Hill, Singapore, pp. 1335–1360.
198. Zheng Y, Lu F, Qui Y, et al. (1999). Enzymatic degradation of leuprolide in rat intestinal mucosal homogenates. *Pharm. Dev. Technol.* 4:539–544.
199. Guo J, Ping Q, Jiang G, et al. (2004). Transport of leuprolide across rat intestine, rabbit intestine, and Caco-2 cell monolayer. *Int. J. Pharmeceu.* 278:415–422.
200. Sanders L M, Kent J K, McRae G I, et al. (1984) Controlled release of a leutenising hormone releasing hormone analogue from (D,L-lactide, glycolide) microspheres. *J. Pharm. Sci.* 73:1294–1297.
201. Okada H, Toguchi H (1995). Biodegradable microspheres in drug delivery. *Clin. Rev. Ther. Drug. Syst.* 12:1–99.
202. Periti P, Mazzei T, Mini E (2002). Clinical pharmacokinetics of depot leuprorein. *Clin. Pharmacokinet.* 41:458–504.
203. Kochhar C, Imanidis G (2002). In vitro transdermal iontophoretic deliveru of leuprolide-mechanism under constant voltage application. *J. Pharm. Sci.* 92:84–96.
204. Kochhar C, Imanidis G (2004). In vitro transdermal iontophoretic delivery of leuprolide-mechanism under constant current application. *J. Control. Rel.* 98:25–35.
205. Andersson K E, Arrner B (1972). Effects of DDAVP, a tynthetic analogue of vasopressin, in patients with cranial diabetes in sipidus. *Acta. Med. Scand.* 192:21–27.
206. Becker D J, Foley T P (1978). 1-Deamino-8-arginine vasopressin in the treatment of central diabetes insipidus in childhood. *J. Pediatr.* 92:1011–1015.
207. Dimson S B (1977). Desmopressin as a treatment for enuresis. *Lancet.* 1:1260–1261.
208. Mannuicci P M, Ruggeri Z M, Pareti F I, et al. (1977). D.D.A.V.P. in haemophillia. *Lancet.* 2:1171–1172.
209. Fjellestad-Paulsen A, Lundin H P, Paulsen O (1993). Pharmacokinetics of 1-Deamino-8-arginine vasopressin after various routes of administration in healthy volunteers. *Clin. Endocortinol.* 38:177–182.
210. Rembratt A, Graugaard-Jensen C, Senderovitz T, et al. (2004). Pharmacokinetics and pharmacodynamics of desmopressin administred orally versus intravenously at daytime versus night-time in healthy men aged 55–70 years. *Eur. J. Clin. Pharmacol.* 60:397–402.

211. Agay-Abensour L, Ajellestad-Paulsen A, Hoglund P, et al. (1993). Absolute bioavailability of an aqueous solution of 1-Deamino-8-arginine vasopressin from different regions of gastrointestinal tract in man. *Eur. J. Clin. Pharmacol.* 44:473–476.

212. Agerso H, Laesen L S, Riis A, et al. (2004). Pharmacokinetics and renal excretion of desmopressin after intravenous administration to healthy subjects and renally impaired patients. *Br. J. Clin. Pharmacol.* 58:352–358.

213. Lundin S, Pierzynovski S G, Westrom B R, et al. (1991). Biliary excretion of the vasopressin analogue DDAVP after intraduodenal, intrajugular and intraportal administration in the conscious pig. *Pharmacol. Toxicol.* 68:177–180.

214. Odeberg J M, Callreus T, Lundin S, et al. (2004). A pharmacokinetics and pharmacodynamics study of desmopressin: Evaluationg sex differences and the effect of pretreatment with piroxicam, and further validation of an indirect response model. *J. Pharm. Pharmacol.* 56:1389–1398.

215. Lethagen S, Harris A S, Sjorin E, et al. (1987). Intranasal and intravenous administration of desmopressin: Effect on F VIII/vW, pharmacokinetics and reproducibility. *Thromb. Haematost.* 58:1033–1036.

216. Harris A S, Ohlin M, Lethagen S, et al. (1988). Effects of concentration and volume on nasal baioavailability and biological response to desmopressin. *J. Pharm. Sci.* 77:337–339.

217. Joukhader C, Schenk B, Kaehler S T, et al. (2003). A replicate study design for testing bioequivglence: A case study on two desmopressin nasal spray preparations. *Pharmacokinet. Dispos.* 59:631–636.

218. Cormier M, Bjohnson M, Ameri K, et al. (2004). Transdermal delivery of desmopressin using a coated microneedle array patch system. *J. Controlled Rel.* 97:503–511.

219. Kim S J, Park Y, Hong H J (2005). Autibody engineering for the development of therapeutic antibodies. *Mol. Cells.* 20:17–29.

220. Tabrizi M A, Tseng C M, Roskos L K (2006) Elimination mechanisms of therapeutic monoclinal antibodies. *Drug Discov. Today.* 11:81–88.

221. Houghton A N, Scheinberg D A (2000). Monoclonal antibody therapies—a constant treatment to cancer. *Nat. Med.* 6:373–374.

222. Cartron G, Dacheux L, Salles G, et al. (2002). Therapeutic activity of humanized anti-CD20 monoclonal antibody and polymorphism in IgG Fc receptor FcgammaRIIIa gene. *Blood.* 99:754–758.

223. Kumpel B M, De Haas M, Koene H R, et al. (2003). Clearance of red cell by monoclonal IgG3 anti-D in vivo is affected by the VF polymorphism of Fcgamma RIIIa(CD16). *Clin. Exp. Immunol.* 132:81–86.

224. Clarkson S B, et al. (1986). Treatment of refractory immune thrombocytopenic purpura with anti-Fc gamma-receptor antibody. *N. Engl. J. Med.* 314:1236–1239.

225. Dall'Acqua W F, Woods R M, Ward E S, et al. (2002). Increasing the affinity of a human IG1 for the neonatal Fc receptor: Biological consequences. *J. Immunol.* 169:5171–5180.

226. Hinton P R, Johlfs M G, Xiong J M, et al. (2003). Engineered human IgG antibodies with longer serum half-lives in primates. *J. Biol. Chem.* 279:6213–6216.

227. Chapman A P (2002). PEGylated antibodies and antibody fragments for improved therapy: A review. *Adv. Drug. Deliv. Rev.* 54:531–545.

228. Chapman A P, Antoniw A, Spitali M, et al. (1999). Therapeutic antibody fragments with prolonged in vivo half-lives. *Nat. Biotechnol.* 17:780–783.

229. King D J, Turner A, Farnsworth A P, et al. (1994). Improved tumor targeting with chemically cross-linked recombinant antibody fragments. *Cancer Res.* 54:6176–6185.

230. Hooksc M A, Wade S, Millikan W J Jr (1991). Muromonab CD-3: a review of its pharmacology, pharmacokinetics, and clinical use in transplantation. *Pharmacother.* 11:26–37.

231. Kovarik J, Wolf P, Isterne J M, et al. (1997). Disposition of Basiliximab, an interleukin-2 receptor monoclonal antibody, in recipients of mismatched cadaver renal allografts. *Transplant.* 64:1701–1705.

232. Kovarik J, Breidenbach T, Gerbeau C, et al. (1998). Disposition and immunodynamics of basiliximab in liver allograft recipients. *Clin. Pharmacol. Ther.* 64:66–72.

233. Kovarki J M, Kahan B D, Rajagopalan P R, et al. (1999). Population pharmacokinetics and exposure-response relationships for basiliximab in kidney transplantation. *Transplant.* 68:1288–1294.

234. Kovarki J M, Nashan B, Neuhaus P, et al. (2001). A population pharmacokinetic screen to identify demographic-clinical covariates of basiliximab in liver transplantation. *Clin. Pharmacol. Ther.* 69:201–209.

235. Baselga J, Carbonell X, Castaneda-Soto N J, et al. (2005). PhaseII study of efficacy, safety, and pharmacokinetics of trastuzumab monotherapy administered on a 3-weekly schedule. *J. Clin. Oncol.* 23:2162–2171.

236. Bruno R, Washington C B, Lu J F, et al. (2005). Population pharmacokinetics of trastuzumab in patients with HER2+mesenteric breast cancer. *Cancer Chemother Pharmacol.* 56:361–369.

237. Siddiqui M A, Scott L J (2005). Infliximab: A review of its use in Crohn's disease and rheumatoid arthritis. *Drugs.* 65:2179–2208.

238. Schwab M, Klotz U (2001). Pharmacokinetic considerations in the treatment of inflammatory bowel disease. *Clin. Pharmacokinet.* 40:723–751.

239. Ng C M, Bruno R, Combs D, et al. (2005). Population pharmacokinetics of rituximab (anti-CD20 monoclonal antibody) in rheumatoid arthritis patients during a phase II clinical trial. *Clin. Pharmacol.* 45:792–801.

240. Kim H, Csaky K G, Chan C C, et al. (2006). The pharmacokinetics of rituximab following an intraviteral injection. *Exp. Eye Res.* 82:760–766.

241. Subramanian K N, Weisman L E, Rhodes T, et al. (1999). Safety, tolerance and pharmacokinetics of a humanized monoclonal antibody to respiratory syncytail virus in premature, infants and infants with bronchopulmonary dysplasia. MEDI-493 study group. *Pediatr. Infect. Dis. J.* 17:110–115.

242. Saez-Llorens X, Moreno M T, Ramilo O, et al. (2004). MEDI-493 study group: Safety and pharmacokinetics of palivizumab therapy in children hospitalized with respiratory syncytail virus infection. *Pediatr. Infect. Dis. J.* 23:707–712.

# IMMUNOGENICITY OF THERAPEUTIC PROTEINS

Suzanne Hermeling,[1,2] Daan J.A. Crommelin,[1]
Huub Schellekens,[1,2] and Wim Jiskoot[1,3]

[1] *Utrecht Institute for Pharmaceutical Sciences (UIPS), Utrecht University, The Netherlands.*
[2] *Central Laboratory Animal Institute, Utrecht University, The Netherlands.*
[3] *Leiden/Amsterdam Center for Drug Research (LACDR), Leiden University, Leiden, The Netherlands.*

**Chapter Contents**

1 Introduction
2 Immune Mechanisms
  2.1 Classic Immune Response
  2.2 Breaking of Immune Tolerance
3 Factors Influencing Immunogenicity
  3.1 Product Characteristics
  3.2 Patient Characteristics
  3.3 Treatment Characteristics
  3.4 Other Factors
4 Detection
  4.1 Assays
  4.2 The Design of an Assay Strategy
  4.3 Predictive Models
  4.4 Prediction on Basis of Structure and Sequence
5 Biological and Clinical Consequences
6 A Risk-based Approach of Immunogenicity
7 Examples
  7.1 Epoetin: How a Formulation Change Increased the Immunogenicity
  7.2 RhIFNβ: Immunogenicity of Glycosylated and Nonglycosylated Variants

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

7.3 RhIFNα2: Reduction of the Immunogenicity by Optimizing Production and Formulation
8 Conclusions
References

## 1 INTRODUCTION

Since the introduction of recombinant DNA techniques, it is possible to produce proteins on a large scale. These include proteins that are identical or nearly identical to endogenous human proteins, but also foreign proteins like streptokinase and asparaginase. Not only foreign proteins, but also recombinant human protein therapeutics are potentially immunogenic, and indeed most of these products induce antibodies in some patients after a certain period of treatment (see Table 1 for examples). Patients will first develop binding antibodies (BAbs), and this can be followed by the formation of neutralizing antibodies (NAbs). BAbs usually do not cause major complications, but NAbs bind to the active site of the protein and may interfere with the therapeutic effect. NAbs sometimes also cross-react with the endogenous protein, which can lead to serious complications [3, 4].

This chapter gives an overview of the current knowledge about the immunogenicity of therapeutic proteins. First the mechanisms by which antibodies are formed will be discussed, followed by the factors that can influence the immune response. An assay strategy for the detection of antibodies in human sera is discussed. The

**TABLE 1. Examples of Recombinant Therapeutic Proteins with Reported Immunogenicity**[1]

| Type of Protein | Protein |
|---|---|
| Hormones | Insulin |
| | Growth hormone |
| Cytokines | Interferon alpha |
| | Interferon beta |
| | Interleukin 2, 3 and 12 |
| Enzymes | Factor VIII |
| | DNase |
| | Tissue plasminogen activator |
| Antibodies | Anti-CD3 (murine antibody) |
| | Anti-Her2 (humanized antibody) |
| | Anti-IgE (humanized antibody) |
| | Anti-respiratory syncytial virus (humanized antibody) |
| | Anti-IL-2 receptor (humanized antibody) |
| Growth factors | G–CSF |
| | GM–CSF |
| | Erythropoietin |
| | Thrombopoietin |

[1] Adapted from Refs. 1 and 2.

biological and clinical consequences are summarized, and the U.S. Food and Drug Administration (FDA) procedure to handle immunogenicity concerns of novel products is explained. Next, we give some examples of therapeutic proteins with immunogenicity problems, followed by the final conclusions.

## 2 IMMUNE MECHANISMS

Depending on the type of therapeutic protein, antibodies can be formed via two pathways (Table 2): a classic immune response to foreign proteins and breaking of immune tolerance to self-proteins [1].

### 2.1 Classic Immune Response

A classic immune response occurs after administration of foreign proteins. Antigen-presenting cells (APCs) will take up the protein, digest it, and present peptides on major histocompatibility complex (MHC)-molecules on their surface. T cells will recognize these peptides in combination with the MHC-molecules and will activate B cells to produce antibodies against the foreign protein. This type of reaction is observed when proteins of animal, microbial, or plant origin are administered to patients. The antibody formation is usually fast, within days to weeks, and often occurs after a single injection. The antibodies are mostly NAbs and persist for a long time.

A classic immune response can also occur on administration of recombinant human proteins to patients with an innate deficiency, e.g., children lacking growth hormone. These children lack the immune tolerance that normally exists for self-proteins.

**TABLE 2. Types of Immune Reaction Against Therapeutic Proteins**[1]

| | Classic Immune Response | Breaking of Immune Tolerance |
|---|---|---|
| Properties of Product | Microbial or Plant Origin | Human Homologue |
| Characteristics of antibody formation | Fast<br>Often after a single injection<br>High incidence<br>Neutralizing antibodies<br>Long duration | Slow<br>After prolonged treatment<br>Low incidence<br>Mainly binding antibodies<br>Disappear after stopping of treatment and sometimes during treatment |
| Cause of immunogenicity | Presence of non-self-antigens | Impurities and presence of aggregates Mechanism of antibody formation is unknown |
| Consequences | Loss of efficacy in most cases | In most patients, no consequences |

[1] Taken from Ref. 5.

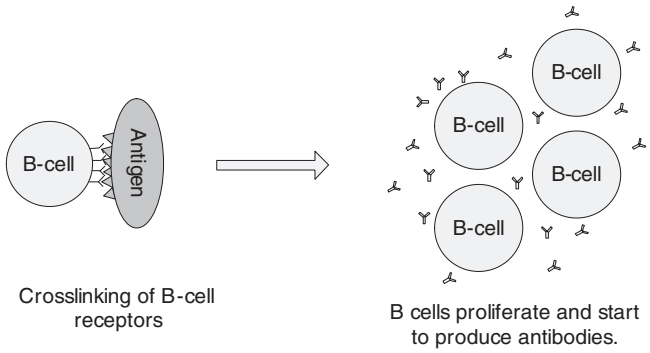


**Figure 1.** Schematic representation of a possible mechanism of breaking B-cell tolerance. The antigen contains multiple epitopes that cross-link B-cell receptors. The cross-linking of the receptors is observed as a danger signal because of which the B cells start to proliferate to plasma cells and produce antibodies.

### 2.2 Breaking of Immune Tolerance

Most recombinant human proteins are homologous to their endogenous counterparts. The immune system of the patients will therefore recognize the protein as a self-protein, for which the patient is tolerant. Antibody formation against self-proteins can occur through breaking of this immune tolerance. The antibody formation via this process is slow and often becomes apparent only after months of chronic administration of the protein to the patient. Usually the antibodies disappear when treatment is stopped. The exact mechanism of breaking of immune tolerance is not known. Multimeric antigen presentation with a narrow (~5–10 nm) spacing, however, is known to break immune tolerance, possibly by cross-linking of B-cell receptors (Figure 1). The only natural proteins showing this closely spaced multimeric antigen presentation are microbial antigens, and apparently during evolution, there was a strong selective pressure to react vigorously to this type of antigen presentation [6–8].

## 3 FACTORS INFLUENCING IMMUNOGENICITY

Many factors influence the immune response induced by therapeutic proteins. These can be divided into product, patient, and treatment characteristics, and several factors are still unknown.

### 3.1 Product Characteristics

Proteins are complex structures consisting of primary, secondary, tertiary, and sometimes quaternary structures. Changes in one of these structural levels might influence the immune response. How a protein is formulated influences the chemical and physical stability of the protein. Therefore the formulation can also influence the immunogenicity. Moreover, contaminants and impurities play an important role.

***Primary Sequence.*** Based on amino acid sequence, no definite predictions about the immunogenicity of a product can be made. A single amino acid change in insulin was enough to elicit a strong immune response [9]. Consensus interferon alpha (co-IFNα), on the other hand, has several amino acid changes as compared with the naturally occurring human IFNαs (10–23 amino acid differences and on average 89% homology with naturally occurring IFNαs), but it has no increased immunogenicity [10]. Foreign proteins, such as streptokinase, staphylokinase, bovine adenosine deaminase, and salmon calcitonin will elicit a classic immune response in patients.

Chemical changes of the primary structure, such as oxidation or deamidation of amino acids, can lead to novel epitopes, which can induce a classic immune response [11]. Moreover, the chemical change can induce aggregation leading to multimeric antigen presentation and, thus, breaking of immune tolerance.

***Glycosylation.*** Glycosylation is a common posttranslational modification and is cell and species specific. The glycans can differ in chain length, sequence, and linkage position to the peptide backbone [12]. Endogenous proteins often consist of several glycosylated isoforms [13]. The carbohydrates can play a role in molecular stability, *in vivo* activity, serum half-life, and immunogenicity. Carbohydrates can decrease the immunogenicity of therapeutic proteins, by shielding immunogenic epitopes [12, 14] or by increasing the solubility of the protein and thereby preventing aggregation [15].

Production of therapeutic proteins in plants is being developed, but concerns are raised about the immunogenicity of plant glycans. Glycosylation of proteins in plants and humans differ in fine detail [16]. Especially the plant-specific α(1,3)-fucose and β(1,2)-xylose groups are considered immunogenic in humans [12]. Metabolic engineering of the plant N-glycan biosynthesis pathway is being pursued to prevent the insertion of these sugars in the glycan chains [16].

***PEGylation.*** The most successful approach to increase the mean plasma half-life of proteins is by chemically coupling poly(ethylene glycol) (PEG) moieties to the protein. The attachment of PEG (PEGylation) decreases the overall rate of clearance of the protein, shields the protein from proteolytic enzymes, and masks immunogenic sites [17, 18]. The PEG molecules can differ in conjugation type, molecular weight, and conformation (i.e., linear, branched, or multiple-branched) [19]. PEGylation can decrease the immunogenicity of a therapeutic protein by blocking antibody binding sites, promoting solubility, and permitting less frequent dosing [20]. Branched PEG is more effective than linear PEG in improving the immunological properties of the protein [21]. Although in most cases PEGylation decreases the immunogenicity of a therapeutic protein, two examples of therapeutic proteins are known in which the PEGylated protein was more immunogenic than the non-PEGylated variant: PEGylated recombinant human megakaryocyte growth and development factor (PEG-rhMGDF) [4] and recombinant methionyl human tumor necrosis factor binding protein PEGylated dimer [22]. The reason for this increased immunogenicity is not known, but it might be related to the increased plasma half-lives, which leads to prolonged exposure to the immune system.

***Impurities and Contaminants.*** Recombinant proteins can be produced in different expression systems, each having advantages and disadvantages [16]. Impurities from the expression systems can be of influence on the immunogenicity of the final product. They can act as adjuvants or be immunogenic themselves [11]. Patients receiving *Escherichia coli*-derived granulocyte-macrophage colony-stimulating factor (GM–CSF) were shown to develop antibodies to GM–CSF as well as *E. coli* proteins [23].

***Secondary Structure.*** Many studies have shown the importance of aggregation in inducing and/or increasing an immune response [24–27]. Aggregation usually occurs after partial unfolding of the protein due to, e.g., shear/shaking or high temperature [28]. Aggregation may expose new epitopes on the surface of the protein for which the immune system is not tolerant. This will lead to a classic immune response. Aggregation can also lead to multimeric antigen presentation, which is known to break B-cell tolerance. Therefore, assessing the presence of aggregates in protein formulations is considered to be very important, although not all aggregates will induce an immune response as was shown in our laboratory [29]. The analytical tools for characterizing the aggregates in a formulation are described elsewhere in this book.

### 3.2 Patient Characteristics

Patient characteristics, such as genetic background and the disease status, are known to influence rate and type of immune reactions [11]. Hemophilia patients with severe genetic defects are more prone to antibody formation than patients with minor genetic defects [30, 31]. Eprex (Ortho Biotech, Tilburg, The Netherlands) associated pure red cell aplasia (PRCA) was only observed in patients with renal failure and not in cancer patients [32].

### 3.3 Treatment Characteristics

Also the treatment characteristics can have an influence on the immunogenicity of therapeutic proteins. Usually antibodies are only induced after prolonged treatment of the protein. The intramuscular (i.m.) route of administration is less immunogenic than subcutaneous (s.c.) administration. Intravenous (i.v.) administration usually is the least immunogenic route of administration.

### 3.4 Other Factors

When evaluating the immunogenicity of therapeutic proteins, one also has to consider the immunomodulatory effects of the protein. For instance, IFNα may increase the antibody production because of its innate immunomodulatory effects [33]. The timing of blood sampling may influence the amount of antibody present in the sera. If a blood sample is taken too early after the administration of the protein, antibodies can still be complexed to the antigen and the sample may prove to be false-negative. If the sample is taken too long after the last injection, antibodies may already have disappeared [34].

## 4 DETECTION

Detection of antibodies can be done by several methods (Table 3). As there are no standardized assays for antibodies to most products and laboratories mostly use their own in-house methods, comparison of results from different studies is in principle impossible. Ideally, a combination of standardized assays should be performed to ensure that all possible types (low affinity and high affinity, classes and isotypes, binding and neutralizing) of antibodies are detected.

### 4.1 Assays

Antibody assays can be divided into two main categories: binding assays and bioassays. The binding assays detect in principle any antibodies with sufficient affinity

**TABLE 3. Advantages and Disadvantages of the Most Commonly Used Assays to Detect Binding Anti-Drug Antibodies in Patient Sera**[1]

| Type of Assay/ Method | Advantages | Disadvantages |
|---|---|---|
| ELISA (bridging format; coating with drug and detecting with labeled drug) | • Detection of all isotype responses<br>• Species independent<br>• High throughput<br>• Easy to use<br>• Good sensitivity | • Low-affinity antibodies may be missed<br>• Epitopes may be masked by immobilization<br>• Protein conformation may becompromised by immobilization<br>• Antibody classes/isotypes are not discriminated |
| ELISA (direct format; coating with drug and detecting with labeled anti-Ig | • High throughput<br>• Easy to use<br>• Good sensitivity<br>• Isotype detection possible (by using isotype specific conjugates) | • Immobilization<br>• Protein conformation may be compromised by immobilization<br>• Low-affinity antibodies may be missed |
| ELISA (indirect format; coating with a specific Mab or biotin and then drug) | • Coating plate with specific MAb keeps drug in oriented position<br>• Consistent coating and epitope exposure | • MAb should not alter epitopes' accessibility<br>• MAb epitope should be clinically irrelevant<br>• Isotype detection determined by conjugate<br>• Species specificity determined by conjugate<br>• Low-affinity antibodies may be missed |
| Radioimmuno (precipitation) assay (RIP/RIA) | • Solution phase, which retains the conformation of the antigen<br>• Moderate-to-high throughput<br>• Generally good sensitivity | • Uses radioactivity<br>• Conjugation chemistry might degrade or alter molecule<br>• May trap labeled drug nonspecifically in precipitate |

**TABLE 3.** *Continued*

| Type of Assay/ Method | Advantages | Disadvantages |
|---|---|---|
| Immunoblotting | • Dissects the specificity of the antibodies<br>• Profile of reactivity against subcomponents of the product<br>• Sensitivity is good | • Nonquantitative<br>• Low throughput<br>• Conformational epitopes are missed |
| Surface plasmon resonance | • Solution phase, which mostly retains the conformation of the antigen<br>• Automated<br>• No detection conjugate required<br>• Not species-specific<br>• Detection of all isotypes<br>• Determination of association and dissociation constants<br>• Enables detection of low-affinity antibodies | • Linking chemistry may affect molecule or mask epitopes<br>• Regeneration step may degrade molecule<br>• Low throughput<br>• Requires dedicated equipment<br>• Sensitivity usually lower than ELISA or RIA (except for low-affinity antibodies) |

[1] Adapted from Refs. 35 and 36.

for the therapeutic protein. A bioassay tests whether the antibodies detected by the binding assay can neutralize the biological effects of the therapeutic protein. In most cases, BAbs have no clinical effect, but the binding assay is used as a screening tool to identify samples with possible neutralizing activity.

***Binding Assays.*** Several binding assays are available that can detect antibodies against therapeutic proteins, with solid phase binding assays being the most commonly used for antibody detection. Table 3 summarizes the advantages and disadvantages of the most commonly used techniques. The advantage of surface plasmon resonance over the other methods is the possibility to measure the binding of the antibodies in real time, which gives information about the affinity of the antibodies. Figure 2 shows a sensorgram of sera containing antibodies with different association and dissociation profiles. A Western blot gives information about the specificity of the antibodies in the serum and may show antibodies to impurities as was shown in patients receiving GM–CSF [23]. For a more detailed description of these methods, we refer to specialized literature [35–37].

***Neutralizing Assays.*** The detection of neutralizing antibodies is performed in assays that can show inhibition of the biological activity of the therapeutic protein. In general this will be a bioassay. The therapeutic protein is incubated with the serum to be tested, and the mixture is tested in the bioassay. If the antibodies in the serum have neutralizing capacities, the biological activity of the mixture is reduced, as compared with the pure therapeutic protein.

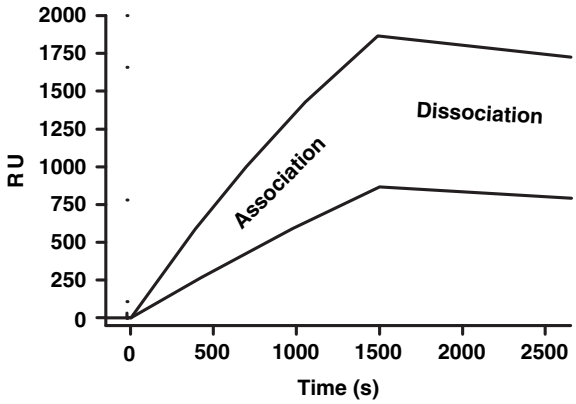


**Figure 2.** Sensorgram obtained with a BIAcore 3000 using anti-rhIFNα2b sera with different association and dissociation profiles.

### 4.2 The Design of an Assay Strategy

Long-term studies are needed to fully assess the immunogenicity of a product. Also, the timing of sampling is important. A pretreatment blood sample should be taken, and several time points should be considered because antibodies may be induced only after prolonged treatment. If a sample is taken too early after administration of the therapeutic protein, the circulating protein may interfere with the antibody assay [9]. Conversely, if the sampling is too long after the last administration of the protein, antibody levels might already have dropped below the detection limit of the assays. Both situations will lead to an underestimation of the incidence of antibody positive patients.

To fully assess the immunogenicity of a therapeutic protein, a combination of assays is necessary. In general, sera are screened first by an enzyme-linked immunosorbent assay (ELISA) or radioimmunoassay (RIA) type of binding assay, which have a high through-put and are easy to perform. A screening assay is optimized for sensitivity and therefore suffers from a relative high number of false-positive results. So, all initial positive sera should be confirmed, e.g., by showing a reduced binding after adding the product to the serum. The confirmed positive sera are then tested for neutralizing activity. Further characterization of the antibody response may follow to determine isotype, affinity, and so on. Whatever approach is used, it is important to validate the assays to show reproducibility, precision, robustness, and so on [35, 38]. International reference preparations are only available for a limited number of assays, but they are crucial for making comparison between different laboratories possible [39].

### 4.3 Predictive Models

Most of the data about immunogenicity of therapeutic proteins come from clinical trials or postmarketing surveillance. It would of course be better if the immunogenicity of a product could be predicted before the clinical phase of development.

***Conventional Animals.*** Conventional animal models may be useful to predict the immunogenicity of microbial products as staphylokinase and streptokinase, which are foreign proteins for humans as well as animals [40]. Conventional animals can also be used if the therapeutic protein shows a high degree of intraspecies sequence homology. An example is human thrombopoietin, which induced neutralizing antibodies in animal models as well as in patients [41]. However, most recombinant human therapeutic proteins are foreign proteins for animals that will elicit an immune response in all cases. So their predictive value for the immunogenicity of human proteins in humans is low. Nonhuman primates that share a higher level of sequence homology with humans have been shown to be excellent models for some products such as human growth hormone and lys-pro insulin [42], but for other proteins, their predictive value has shown to be limited.

***Immune Tolerant Animals.*** In most cases, the induction of antibodies to human therapeutic proteins in patients is based on breaking B-cell tolerance. This process makes animals tolerant for the protein the best model for predicting immunogenicity in humans. Mice are the species of choice to induce tolerance, which can be achieved by chronic administration of the protein in large quantities [43] or by making the animals transgenic for the gene expressing the protein [44]. Obvious drawbacks of the first method are the large quantities of protein required and the need for testing the tolerance of each animal. Mouse lines expressing the transgene only need to be evaluated for immune tolerance once and only need testing by polymerase chain reaction (PCR) to identify transgene positivity after breeding. Transgenic animals, immune tolerant for insulin [45], tissue plasminogen activator (tPA) [46], IFNα2 [26], and IFNβ [47] are available. The insulin and tPA models were used to evaluate whether amino acid substitutions in the protein introduced new epitopes [43, 44]. The interferon models were used to study the effect of aggregates and other degradation products on immunogenicity [26, 29, 47].

Some therapeutic proteins might have immune modulatory effects that influence the induction of antibodies. If the human protein lacks this effect in mice, their predictive value may suffer. This deficiency can be prevented by challenging both the human as well as the murine homologue. Braun et al. showed that mixtures of murine and human IFNα2a caused an increase in immune response in wild-type mice to the human protein [26]. Another point of consideration when using transgenic mice as predictors for immunogenicity is the difference between mouse and human MHC and their presentation of T-cell epitopes [39, 41]. Also, important individual patient characteristics, such as disease burden and concomitant therapy influencing the antibody response, are difficult to reproduce in the transgenic immune-tolerant animal model. Although these models will never become predictors of the individual patient's antibody response, they will be very useful in the development phase of a new protein therapy or after a formulation or production change. The transgenic immune-tolerant mice can also be used to study the mechanism of antibody induction, because they share tolerance for the protein with patients.

### 4.4 Prediction on Basis of Structure and Sequence

Other strategies, based on structural analysis, have been described to predict the immunogenicity of therapeutic proteins. Immunogenic epitopes can be predicted

based on amino acid sequence analysis and MHC-binding. Modifications in an immuno-dominant T-cell epitope by a single amino acid was reported to reduce the immunogenicity of rhIFNβ-1b in BALB/cByJ mice [48]. Analysis of B-cell epitopes is more difficult, but it is also ongoing [49]. These types of analyses may help to reduce the immunogenicity, in case the product is a foreign protein inducing a classic immune response. As discussed, in most of cases, the induction of antibodies in patients is based on breaking B-cell tolerance. Although the exact mechanism is still unknown, we know that the product quality, independent of the amino acid sequence, plays an important role.

## 5 BIOLOGICAL AND CLINICAL CONSEQUENCES

The clinical effects of antibody formation vary from no effect to severe life-threatening situations [50]. Binding antibodies can alter the pharmacokinetic properties of the therapeutic protein [51, 52]. Neutralizing antibodies cannot only affect the pharmacokinetics but also block the pharmacological effect of the therapeutic protein [51–53]. Severe side effects occur if NAbs cross-react with an endogenous protein, as was observed for PEG–rhMGDF and recombinant human erythropoietin (epoetin). Patients receiving PEG–rhMGDF developed NAbs cross-reacting with endogenous thrombopoietin, which resulted in a severe thrombocytopenia [4]. Patients receiving Eprex (an epoetin formulation marketed in Europe, Canada, and Australia) developed NAbs that cross-reacted with endogenous erythropoietin, resulting in PRCA [3]. The number of patients with antibody-mediated PRCA rapidly increased after a formulation change of Eprex, as will be discussed below. Besides the formation of antibodies, allergic reactions can occur.

## 6 A RISK-BASED APPROACH OF IMMUNOGENICITY

The FDA evaluates immunogenicity concerns of novel products in development and for major changes in manufacturing or clinical uses via a risk-based approach. A schematic representation of this approach is given in Figure 3. Three elements are considered important for the risk assessment strategy: the severity of consequences of the immune response to a therapeutic protein, host-specific factors that impact the immunogenicity positively or negatively, and product-specific factors that impact the immunogenicity positively or negatively. Based on their expertise and literature data from marketed products, the FDA decides whether immunogenicity testing should be performed before or during clinical trials.

## 7 EXAMPLES

### 7.1 Epoetin: How a Formulation Change Increased the Immunogenicity

Erythropoietin is a protein produced by the kidneys that stimulates the production of erythrocytes. Epoetin is mainly used in patients suffering from anemia associated with chronic renal failure or cancer. Three forms of epoetin are commercially

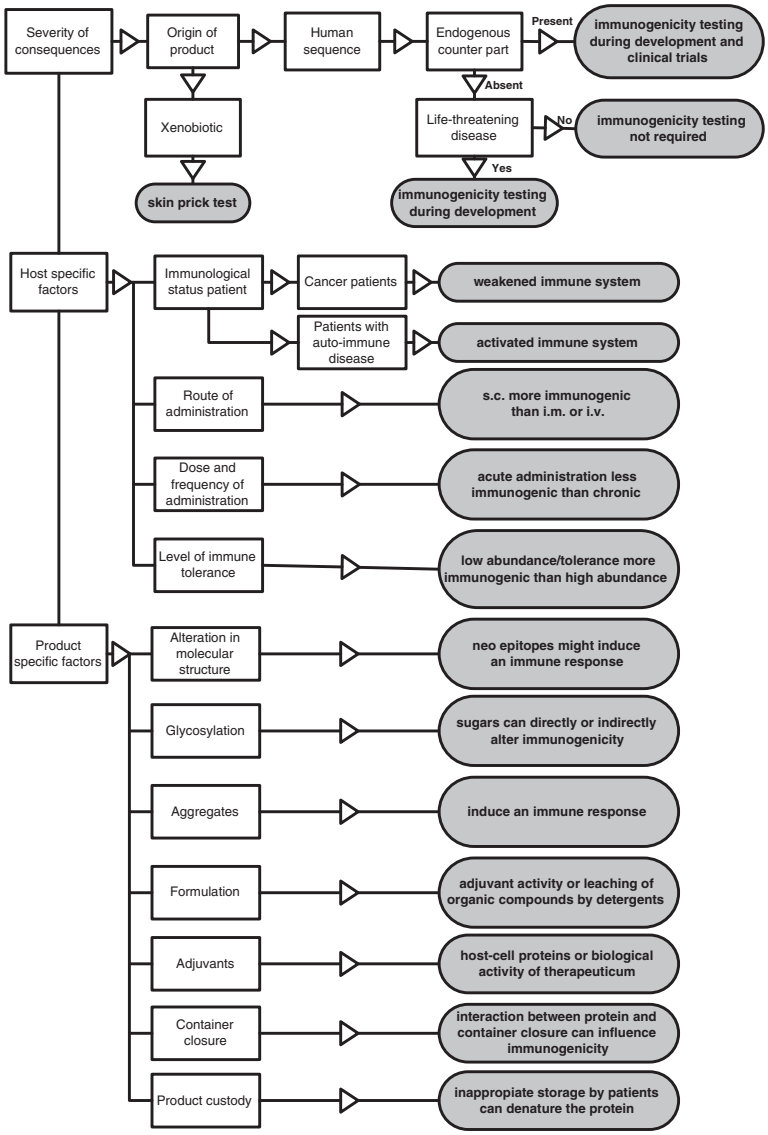


**Figure 3.** Schematic representation of the risk-based approach that the FDA uses for novel products in development and for major changes in manufacture and clinical use of licensed products. Adapted from Refs. 54 and 55.

available, epoetin alfa, epoetin beta, and darbepoetin alfa, which is a hyperglycosylated analog [32]. In 1997 the first patient with antibody-mediated PRCA was reported [56]. Antibody-mediated PRCA is caused by antibodies induced by exogenous epoetin that cross-react with endogenous erythropoietin. In 2002, 13 patients were reported that had developed antibody-mediated PRCA, after administration of Eprex [3]. These cases occurred after 1998. Since then, more patients with antibody-mediated PRCA were identified. In 1998 the formulation of Eprex was changed: Human serum albumin (HSA) was replaced by glycine and polysor-

bate 80. Several hypotheses have been postulated to explain the increase in PRCA cases. Hermeling et al. hypothesized that some epoetin molecules were solubilized in micelles, which might have led to multimeric antigen presentation [57]. Others claim that the polysorbate 80 extracted leachates out of the uncoated rubber stoppers, which had an adjuvant effect, leading to an increase in antibodies [58]. An adjuvant, however, can only increase an existing immune response and cannot break B-cell tolerance by itself. It will be very diffcult to pinpoint the exact cause of the increased number of PRCA cases. In 2002, actions were taken to decrease the number of PRCA cases: emphasizing strict adherence to storage and handling conditions, the introduction of a contraindication for s.c. administration of Eprex in chronic renal failure in many countries and the introduction of coated rubber stoppers. These actions decreased the number of patients with antibody-mediated PRCA, but the question remains of what exactly caused the increased immunogenicity.

### 7.2 RhIFNβ: Immunogenicity of Glycosylated and Nonglycosylated Variants

Interferon beta is a cytokine with anti-inflammatory, antitumor, antiviral, and cell-growth-regulatory effects. It is mainly produced by macrophages and epithelial and fibroblast cells [59, 60]. Administration of rhIFNβ has been established as a treatment for relapsing-remitting multiple sclerosis [60]. Natural hIFNβ is glycosylated, contains 166 amino acids, and has a molecular weight of approximately 25 kDa [38]. Three forms of rhIFNβ are available on the market (Table 4). RhIFNβ-1a (Rebif and Avonex) has an amino acid sequence similar to that of endogenous hIFNβ, is produced in Chinese hamster cells, and is glycosylated. RhIFNβ-1b (Betaseron) is produced in *E. coli* and thus not glycosylated. Moreover, Cys-17 has been mutated to Ser-17 and the N-terminal methionine is lacking.

All three formulations induce neutralizing antibodies of the IgG-type, reducing the efficacy of the therapy [24, 61–63]. The antibodies are usually detectable after the first 6 to 12 months of the treatment [61, 63], but the clinical effects do not appear until after 18–24 months after the start of the treatment [64]. RhIFNβ-1b induces antibodies in more patients than does rhIFNβ-1a [59, 65]. This increased incidence of immunogenicity of rhIFNβ-1b as compared with rhIFNβ-1a is probably due to the lack of glycosylation, which makes the protein more prone to aggregation. Size-exclusion chromatography showed that 60% of rhIFNβ-1b in Betaseron is heavily aggregated [15]. It was shown that Betaseron can break the tolerance of transgenic immune tolerant mice [47], which suggests that these animals may be predictive for the immunogenicity observed in patients.

**TABLE 4. Currently Marketed rhIFNβ Formulations**

| rhIFNβ-1a | rhIFNβ-1b[1] |
|---|---|
| Avonex (Biogen-Idec, Cambridge, MA) | Betaseron (Berlex Laboratories, Montville, NJ) |
| Rebif (Serono, Rockland, MA) | Betaferon (Schering, Berlin, Germany) |

[1] Betaseron and Betaferon are the same formulations, marketed by two companies.

### 7.3 RhIFNα2: Reduction of the Immunogenicity by Optimizing Production and Formulation

RhIFNα2 is used for the treatment of a variety of malignancies and viral diseases [66]. It inhibits viral replication, increases class I MHC expression, stimulates Th1 cells, and inhibits proliferation of many cell types [2]. Two regular and two PEGylated forms of rhIFNα2 are on the market (Table 5). The PEGylated forms are less immunogenic than the non-PEGylated forms. RhIFNα2a and rhIFNα2b only differ in 1 amino acid. RhIFNα2b is produced in *E. coli* cells, which implies that the protein is nonglycosylated. Natural hIFNα2 is O-glycosylated [67]. Patients receiving rhIFNα2a initially produced higher levels of antibodies and at higher incidences than rhIFNα2b. This difference was shown not to be related to the difference in amino acid sequence. The antibodies induced by rhIFNα2a fully cross-react with rhIFNα2b and vice versa [68]. Co-IFNα is a non-naturally occurring synthetic recombinant type I interferon. Its amino acid sequence (166 amino acids) is created by taking for each position the amino acid most commonly observed in 13 IFNα subtypes. Despite the nonsimilarity of co-IFNα with human IFNα species (10–23 amino acid differences and on average 89% homology with naturally occurring IFNαs) [69], the number of patients producing antibodies against co-IFNα is similar to the number of patients producing antibodies after rhIFNα2b treatment [70].

Table 6 shows that optimization of the production and formulation process decreases the immunogenicity [34]. Storing the formulation at room temperature

**TABLE 5. Currently Marketed rhIFNα2 Formulations**

| rhIFNα2a | rhIFNα2b |
|---|---|
| Roferon (Roche, Woerden, The Netherlands) | Intron A (Schering-Plough, Kenilworth, NJ) |
| Pegasys (PEGylated; Roche, Woerden, The Netherlands) | PEG-Intron A (PEGylated; Schering-Plough, Kenilworth, NJ) |

**TABLE 6. Immunogenicity of rhIFNα2a Formulations After Production and Formulation Optimization**[1]

| | Patients with NAbs | Peak Titers (INUs[2]) |
|---|---|---|
| A[3] HSA-containing, lyophilized formulation stored at 25°C | 24% (n = 190) | 435–113,100 |
| B HSA-containing, lyophilized formulation stored at 4°C | 29% (n = 86) | 400–19,200 |
| C HSA-containing, liquid formulation stored at 4°C | 13% (n = 110) | 400–19,200 |
| D HSA-free liquid formulation stored at 4°C | 11% (n = 81) | 588–3191 |
| E HSA-containing lyophilized formulation stored at 4°C | 12% (n = 74) | 672–2441 |

[1] Adapted from Ref. 34.
[2] INUs: IFN neutralizing units.
[3] Formulations A–C were made before the production process was fully optimized and therefore also contained oxidized rhIFNα2a [72]. Formulations D and E contained ultra-pure material.

leads to oxidation and increases the immunogenicity of a rhIFNα2a formulation (formulation A vs. formulation B). Better purification methods further reduce the immunogenicity (formulation C vs. formulations D and E). Formulations A–C were shown to contain aggregates responsible for the increased immune response [26].

Recently it has been shown that not all aggregates, but only aggregates of rhIFNα2b with a native-like structure, can elicit an immune response in transgenic mice immune tolerant for hIFNα2 [71]. It was also shown that not the oxidation, but the aggregation, accompanying the oxidation was the reason for the immunogenicity.

## 8 CONCLUSIONS

Many factors influence the immune response against therapeutic proteins. Unfortunately, it is still impossible to fully predict the immunogenicity of a therapeutic protein before going into clinical trials. The presence of (native-like) aggregates in a formulation is one main factor known to increase the immune response. Formulation changes might have an effect on the immunogenicity as was observed for epoetin and rhIFNα2a. The effect of a formulation change is also difficult to predict, but it can be evaluated in immune-tolerant transgenic mice. Although these tests will not fully predict the immunogenicity of a product, it can give information about the immunogenicity of a new formulation/product as compared with previous formulations/products. As most therapeutic proteins only induce antibodies in a small number of patients, postmarketing surveillance is important. Also basic studies to link physical–chemical properties with immunogenicity are important, and these studies may avoid animal testing in the future.

## REFERENCES

1. Porter S (2001). Human immune response to recombinant human proteins. *J. Pharm. Sci.* 90:1–11.
2. Frost H (2005). Antibody-mediated side effects of recombinant proteins. *Toxicol.* 209:155–160.
3. Casadevall N, Nataf J, Viron B, et al. (2002). Pure red-cell aplasia and antierythropoietin antibodies in patients treated with recombinant erythropoietin. *N. Engl. J. Med.* 346:469–475.
4. Li J, Yang C, Xia Y, et al. (2001). Thrombocytopenia caused by the development of antibodies to thrombopoietin. *Blood.* 98:3241–3248.
5. Schellekens H (2005). Factors influencing the immunogenicity of therapeutic proteins. *Nephrol. Dial. Transplant* 20(Suppl 6):vi3–vi9.
6. Chackerian B, Lowy D R, Schiller J T (2001). Conjugation of a self-antigen to papillomavirus-like particles allows for efficient induction of protective autoantibodies. *J. Clin. Invest.* 108:415–423.
7. Fehr T, Bachmann M F, Bucher E, et al. (1997). Role of repetitive antigen patterns for induction of antibodies against antibodies. *J. Exp. Med.* 185:1785–1792.
8. Bachmann M F, Rohrer U H, Kundig T M, et al. (1993). The influence of antigen organization on B cell responsiveness. *Science.* 262:1448–1451.

9. Hermeling S, Crommelin D J A, Schellekens H, et al. (2004). Structure-immunogenicity relationships of therapeutic proteins. *Pharm. Res.* 21:897–903.
10. Schellekens H (2002). Bioequivalence and the immunogenicity of biopharmaceuticals. *Nat. Rev. Drug. Discov.* 1:457–462.
11. Chirino A J, Mire-Sluis A (2004). Characterizing biological products and assessing comparability following manufacturing changes. *Nat. Biotechnol.* 22:1383–1391.
12. Lis H, Sharon N (1993). Protein glycosylation. Structural and functional aspects. *Eur. J. Biochem.* 218:1–27.
13. Skibeli V, Nissen-Lie G, Torjesen P (2001). Sugar profiling proves that human serum erythropoietin differs from recombinant human erythropoietin. *Blood.* 98:3626–3634.
14. Sinclair A M, Elliott S (2005). Glycoengineering: The effect of glycosylation on the properties of therapeutic proteins. *J. Pharm. Sci.* 94:1626–1635.
15. Runkel L, Meier W, Pepinsky R B, et al. (1998). Structural and functional differences between glycosylated and non-glycosylated forms of human interferon-beta (IFN-beta). *Pharm. Res.* 15:641–649.
16. Chen M, Liu X, Wang Z, et al. (2005). Modification of plant N-glycans processing: the future of producing therapeutic protein by transgenic plants. *Med. Res. Rev.* 25: 343–360.
17. Chaffee S, Mary A, Stiehm E R, et al. (1992). IgG antibody response to polyethylene glycol-modified adenosine deaminase in patients with adenosine deaminase deficiency. *J. Clin. Invest.* 89:1643–1651.
18. Hinds K D, Kim S W (2002). Effects of PEG conjugation on insulin properties. *Adv. Drug Deliv. Rev.* 54:505–530.
19. Chapman A P (2002). PEGylated antibodies and antibody fragments for improved therapy: A review. *Adv. Drug Deliv. Rev.* 54:531–545.
20. Chirino A J, Ary M L, Marshall S A (2004). Minimizing the immunogenicity of protein therapeutics. *Drug Discov. Today* 9:82–90.
21. Caliceti P, Veronese F M (2003). Pharmacokinetic and biodistribution properties of poly(ethylene glycol)-protein conjugates. *Adv. Drug Deliv. Rev.* 55:1261–1277.
22. Moreland L W, McCabe D P, Caldwell J R, et al. (2000). Phase I/II trial of recombinant methionyl human tumor necrosis factor binding protein PEGylated dimer in patients with active refractory rheumatoid arthritis. *J. Rheumatol.* 27:601–609.
23. Wadhwa M, Skog A L, Bird C, et al. (1999). Immunogenicity of granulocyte-macrophage colony-stimulating factor (GM-CSF) products in patients undergoing combination therapy with GM-CSF. *Clin. Cancer Res.* 5:1353–1361.
24. Bertolotto A, Malucchi S, Sala A, et al. (2002). Differential effects of three interferon betas on neutralising antibodies in patients with multiple sclerosis: a follow up study in an independent laboratory. *J. Neurol. Neurosurg. Psychiatry* 73:148–153.
25. Bertolotto A, Malucchi S, Milano E, et al. (2000). Interferon beta neutralizing antibodies in multiple sclerosis: neutralizing activity and cross-reactivity with three different preparations. *Immunopharmacol.* 48:95–100.
26. Braun A, Kwee L, Labow M A, et al. (1997). Protein aggregates seem to play a key role among the parameters influencing the antigenicity of interferon alpha (IFN-alpha) in normal and transgenic mice. *Pharm. Res.* 14:1472–1478.
27. Moore W V, Leppert P (1980). Role of aggregated human growth hormone (hGH) in development of antibodies to hGH. *J. Clin. Endocrinol. Metab.* 51:691–697.
28. Wang W (2005). Protein aggregation and its inhibition in biopharmaceutics. *Int. J. Pharm.* 289:1–30.

29. Hermeling S (2005). Structural Aspects of the Immunogenicity of Therapeutic Proteins; Transgenic Animals as Predictors for Breaking Immune Tolerance. Utrecht University, Utrecht, The Netherlands, pp. 208.
30. Prescott R, Nakai H, Saenko E L, et al. (1997). The inhibitor antibody response is more complex in hemophilia A patients than in most nonhemophiliacs with factor VIII autoantibodies. Recombinate and Kogenate Study Groups. *Blood*. 89:3663–3671.
31. Gilles J G, Jacquemin M G, Saint-Remy J M (1997). Factor VIII inhibitors. *Thromb. Haemost.* 78:641–646.
32. Schellekens H (2005). Immunologic mechanisms of EPO-associated pure red cell aplasia. *Best Pract. Res. Clin. Haematol.* 18:473–480.
33. Palleroni A V, Aglione A, Labow M, et al. (1997). Interferon immunogenicity: Preclinical evaluation of interferon-alpha 2a. *J. Interferon Cytokine Res.* 17(Suppl 1): S23–27.
34. Ryff J C (1997). Clinical investigation of the immunogenicity of interferon-alpha 2a. *J. Interferon Cytokine Res.* 17(Suppl 1):S29–33.
35. Mire-Sluis A R, Barrett Y C, Devanarayan V, et al. (2004). Recommendations for the design and optimization of immunoassays used in the detection of host antibodies against biotechnology products. *J. Immunol. Methods.* 289:1–16.
36. Wadhwa M, Bird C, Dilger P, et al. (2003). Strategies for detection, measurement and characterization of unwanted antibodies induced by therapeutic biologicals. *J. Immunol. Methods.* 278:1–17.
37. Kersten G F, Westdijk J (2005). Immunoassays. In W Jiskoot and D J A Crommelin (eds.), *Methods for Structural Analysis of Protein Pharmaeuticals*, Vol. III, Biotechnology: Pharmaceutical Aspects, AAPS Press, Arlington, V.A., pp. 501–526.
38. Brickelmaier M, Hochman P S, Baciu R, et al. (1999). ELISA methods for the analysis of antibody responses induced in multiple sclerosis patients treated with recombinant interferon-beta. *J. Immunol. Methods.* 227:121–135.
39. Schellekens H (2002). Immunogenicity of therapeutic proteins: Clinical implications and future prospects. *Clin. Ther.* 24:1720–1740.
40. Bugelski P J, Treacy G (2004). Predictive power of preclinical studies in animals for the immunogenicity of recombinant therapeutic proteins in humans. *Curr. Opin. Mol. Ther.* 6:10–16.
41. Rosenberg A S, Worobec A (2005). A risk-based approach to immunogenicity concerns of therapeutic protein products, Part 3: Effects of manufacturing changes in imuunogenicity and the utility of animal immunogenicity studies. *Biopharm. Int.* Jan. 1, 2005:1–3.
42. Wierda D, Smith H W, Zwickl C M (2001). Immunogenicity of biopharmaceuticals in laboratory animals. *Toxicol.* 158:71–74.
43. Simioni P U, Fernandes L G, Gabriel D L, et al. (2004). Induction of systemic tolerance in normal but not in transgenic mice through continuous feeding of ovalbumin. *Scand. J. Immunol.* 60:257–266.
44. Zhang J, Xu L, Haskins M E, et al. (2004). Neonatal gene transfer with a retroviral vector results in tolerance to human factor IX in mice and dogs. *Blood.* 103:143–151.
45. Ottesen J L, Nilsson P, Jami J, et al. (1994). The potential immunogenicity of human insulin and insulin analogues evaluated in a transgenic mouse model. *Diabetologia.* 37:1178–1185.
46. Stewart T A, Hollingshead P G, Pitts S L, et al. (1989). Transgenic mice as a model to test the immunogenicity of proteins altered by site-specific mutagenesis. *Mol. Biol. Med.* 6:275–281.

47. Hermeling S, Jiskoot W, Crommelin D J A, et al. (2005). Development of a transgenic mouse model immune tolerant for human interferon beta. *Pharm. Res.* 22:847–851.
48. Yeung V P, Chang J, Miller J, et al. (2004). Elimination of an immunodominant CD4+ T cell epitope in human IFN-beta does not result in an in vivo response directed at the subdominant epitope. *J. Immunol.* 172:6658–6665.
49. Saha S, Bhasin M, Raghava G P (2005). Bcipep: A database of B-cell epitopes. *BMC Genomics.* 6:79.
50. Koren E, Zuckerman L A, Mire-Sluis A R (2002). Immune responses to therapeutic proteins in humans-clinical significance, assessment and prediction. *Curr. Pharm. Biotechnol.* 3:349–360.
51. Antonelli G (1997). In vivo development of antibody to interferons: an update to 1996. *J. Interferon Cytokine Res.* 17(Suppl 1):S39–46.
52. Thorpe R, Swanson S J (2005). Current methods for detecting antibodies against erythropoietin and other recombinant proteins. *Clin. Diagn. Lab Immunol.* 12:28–39.
53. Quesada J R, Rios A, Swanson D, et al. (1985). Antitumor activity of recombinant-derived interferon alpha in metastatic renal cell carcinoma. *J. Clin. Oncol.* 3:1522–1528.
54. Rosenberg A S, Worobec A (2004). A risk-based approach to immunogenicity concerns of therapeutic protein products, Part 1: Considering consequences of the immune response to a protein. *Biopharm. Int.* Nov. 1, 2004:1–4.
55. Rosenberg A S, Worobec A (2004). A risk-based approach to immunogenicity concerns of therapeutic protein products, Part 2: Considering host-specific and product-specific factors impacting immunogenicity. *Biopharm. Int.* Dec. 1, 2004:1–7.
56. Prabhakar S S, Muhlfelder T (1997). Antibodies to recombinant human erythropoietin causing pure red cell aplasia. *Clin. Nephrol.* 47:331–335.
57. Hermeling S, Schellekens H, Crommelin D J A, et al. (2003). Micelle-associated protein in epoetin formulations: A risk factor for immunogenicity? *Pharm. Res.* 20:1903–1907.
58. Sharma B, Bader F, Templeman T, et al. (2004). Technical investigations into the cause of the increased incidence of antibody-mediated pure-red cell aplasia associated with Eprex®. *Eur. J. Hosp. Pharm.* 5:86–91.
59. Antonelli G, Dianzani F (1999). Development of antibodies to interferon beta in patients: technical and biological aspects. *Eur. Cytokine Netw.* 10:413–422.
60. Mayorga C, Luque G, Romero F, et al. (1999). Antibodies to commercially available interferon-beta molecules in multiple sclerosis patients treated with natural interferon-beta. *Int. Arch. Allergy Immunol.* 118:368–371.
61. Malucchi S, Sala A, Gilli F, et al. (2004). Neutralizing antibodies reduce the efficacy of betaIFN during treatment of multiple sclerosis. *Neurol.* 62:2031–2037.
62. Sorensen P S, Ross C, Clemmesen K M, et al. (2003). Clinical importance of neutralising antibodies against interferon beta in patients with relapsing-remitting multiple sclerosis. *Lancet.* 362:1184–1191.
63. Pachner A R (2003). Anti-IFNbeta antibodies in IFNbeta treated MS patients. *Neurol.* 61:S1–S5.
64. Bertolotto A (2004). Neutralizing antibodies to interferon beta: implications for the management of multiple sclerosis. *Curr. Opin. Neurol.* 17:241–246.
65. Cook S D, Quinless J R, Jotkowitz A, et al. (2001). Serum IFN neutralizing antibodies and neopterin levels in a cross-section of MS patients. *Neurol.* 57:1080–1084.

66. Pfeffer L M, Dinarello C A, Herberman R B, et al. (1998). Biological properties of recombinant alpha-interferons: 40th anniversary of the discovery of interferons. *Cancer Res.* 58:2489–2499.
67. Adolf G R, Kalsner I, Ahorn H, et al. (1991). Natural human interferon-alpha 2 is O-glycosylated. *Biochem. J.* 276 (Pt 2):511–518.
68. Antonelli G, Currenti M, Turriziani O, et al. (1991). Neutralizing antibodies to interferon-alpha: relative frequency in patients treated with different interferon preparations. *J. Infect. Dis.* 163:882–885.
69. Intermune. *Infergen (Interferon alfacon-1).* Available: http://www.infergen.com/pdf/infergen_pi.pdf. Accessed July 26, 2005.
70. Tong M J, Reddy K R, Lee W M, et al. (1997). Treatment of chronic hepatitis C with consensus interferon: A multicenter, randomized, controlled trial. Consensus Interferon Study Group. *Hepatol.* 26:747–754.
71. Hermeling S, Aranha L, Damen J M A, et al. (2005). Structural characterization and immunogenicity in wildtype and immune tolerant mice of degraded recombinant human interferon alpha2b. *Pharm. Res.* In Press.
72. Hochuli E (1997). Interferon immunogenicity: technical evaluation of interferon-alpha 2a. *J. Interferon Cytokine Res.* 17(Suppl 1):S15–21.

# DEVELOPMENT AND CHARACTERIZATION OF HIGH-AFFINITY ANTI-TOPOTECAN IgG AND FAB FRAGMENTS

JIN CHEN AND JOSEPH P. BALTHASAR
*University at Buffalo, The State University of New York, Buffalo, New York*

**Chapter Contents**

1 Introduction
2 Materials and Methods
  2.1 Development of Hybridomas Secreting Anti-TPT Antibodies
  2.2 Characterization of Anti-Topotecan IgG (ATI)
  2.3 Production and Purification of ATI and ATF
  2.4 Development of ELISA for Quantification and ATF in Rat Plasma
  2.5 Pharmacokinetic Studies of ATI and ATF in Rats
3 Results
  3.1 Development of Hybridomas Secreting Anti-TPT Antibodies
  3.2 Characterization of ATI
  3.3 Production and Purification of ATI and ATF
  3.4 ELISA Validation
  3.5 Pharmacokinetics of ATI and ATF
4 Discussion
  References

## 1 INTRODUCTION

This laboratory is investigating a drug targeting strategy that employs anti-drug antibodies to increase drug binding in blood, minimize systemic unbound drug exposure, and reduce the systemic drug toxicity after intraperitoneal (i.p.) chemotherapy [1, 2].

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

The approach uses i.p. administration of an anticancer drug with simultaneous intravenous (i.v.) administration of anti-drug antibodies, where the antibodies are employed to impart regio-specific alterations in drug disposition.

To test the feasibility of this approach, earlier studies in this laboratory used methotrexate (MTX) as a model drug [2, 3]. Pharmacokinetic investigations demonstrated that systemic administration of polyclonal anti-MTX Fab with i.p. administration of MTX led to a decrease in the systemic exposure of unbound MTX, while not altering peritoneal exposure [2]. To facilitate further evaluation of anti-MTX antibodies on MTX induced toxicity, monoclonal anti-MTX IgG and Fab were produced and purified [4]. Subsequent studies demonstrated that after this combination therapy, anti-MTX Fab could allow a fivefold increase of the maximum tolerated dose of i.p. MTX and enhance the median survival time of mice bearing peritoneal tumors [5].

Based on these promising results, we have proposed to apply this targeting strategy to enhance the selectivity of i.p. topotecan (TPT) chemotherapy, as TPT has been shown to be the most active second-line therapy for refractory, metastatic ovarian cancer [6–9]. To facilitate testing hypotheses related to i.p. TPT chemotherapy, it was necessary to develop anti-TPT antibodies. Relative to polyclonal antibodies, monoclonal antibodies would demonstrate more reproducible properties (i.e., affinity and specificity) and may be more feasibly produced in large quantities. In this chapter, we describe the development of hybridomas secreting monoclonal antibodies against TPT, and the production, purification, characterization, and pharmacokinetics of an anti-TPT monoclonal antibody and of monoclonal anti-TPT Fab fragments.

## 2 MATERIALS AND METHODS

### 2.1 Development of Hybridomas Secreting Anti-TPT Antibodies

***Preparation and Characterization of KLH-TPT Immunogen.*** Topotecan was provided by Drug Synthesis & Chemistry Branch, Developmental Therapeutics Program, Division of Cancer Treatment and Diagnosis, National Cancer Institute, National Institutes of Health (Bethesda, MD). TPT was conjugated to keyhole limpet hemocyanin (KLH) via the Mannich reaction using Pierce PharmLink Immunogen Kit (Pierce, Rockford, IL). Briefly, 0.9-mg TPT was linked with 2-mg KLH by condensation with formaldehyde. The reaction was incubated at 50°C overnight, followed by dialysis against PBS (pH 7.4) for 2 h, and then at 4°C overnight. The synthesized immunogen was further diluted by PBS to a final concentration of 1 mg/mL. The degree of conjugation was estimated by ultraviolet (UV) spectroscopy at 411 nm for topotecan and at 280 nm for KLH.

***Development of Hybridomas that Secrete Anti-TPT Antibodies.*** Approximately 100-μg KLH–TPT was dissolved in PBS and emulsified with an equal volume of Freund's incomplete adjuvant (Sigma, St. Louis, MO). Female Balb/c mice (Harlan, Indianapolis, IN) were immunized with 200-μL emulsion by i.p. injection every 3 weeks. Animals were bled via the saphenous vein 7 days after each dose of KLH–TPT. Whole blood was collected and centrifuged. The resulting plasma was used to assess anti-TPT activity via a previously validated HPLC assay [10]. Briefly, plasma was diluted with PBS and incubated with 250-ng/mL TPT for 2 h. Free TPT

was then separated from bound TPT through ultra-filtration using Centrifree tubes (10-KD molecular-weight cutoff; Bellerica, MA). The high-performance liquid chromatography (HPLC) assay employed an isocratic mobile phase of a mixture of 10-mM $KH_2PO_4$ water–methanol–triethylamine (72:26:2, v/v/v, pH 3.5), with fluorescence detection at an excitation wavelength of 361 nm and an emission wavelength of 527 nm. Blank mouse plasma was used as the control. The mouse with the highest activity (i.e., the lowest ratio of free TPT to that of control) was sacrificed, and the spleen was aseptically removed.

Spleen cells were obtained by teasing the spleen apart with forceps and by passing tissue fragments through a stainless steel screen. The spleen cells were washed in RPMI media (Invitrogen, Grand Island, NY) twice by centrifugation at 500*g* in tabletop centrifuge (International Equipment Co., Needham, MS). Myeloma cells, SP2/0-Ag 14 (ATCC, Manassas, VA), were washed once in RPMI. Spleen cells were mixed with myeloma cells and centrifuged. The cell pellets were carefully suspended in 1-mL 50% poly(ethylene glycol) (PEG, Sigma) for 1 min by gently swirling in 37°C water. Cells were diluted and washed again in RPMI media by centrifugation and then resuspended in RPMI HAT (hypoxanthine, aminopterin, and thymidine) medium (Invitrogen) with 10% fetal bovine serum (Invitrogen). Cells suspended in HAT medium were dispensed into the wells of several 96-well plates, which were then incubated at 37°C. Clones, which became visible in 2–3 weeks, were then transferred into the wells of 24-well plates.

***Screening of Anti-Topotecan Activity by HPLC Assay.*** Anti-TPT activities within cell culture supernatants were assessed via HPLC, using similar methods as described above. Initial screening was conducted by pooling six hybridoma supernatants together, and the pooled culture supernatants (195 μL) were incubated with 5 μL of TPT solution (in saline) to a final concentration of 25 ng/mL for 2 h. Unbound TPT was separated from bound TPT through ultra-filtration and then quantified by HPLC. The lower limit of quantitation (LOQ) for TPT was 0.02 ng on column (corresponding to 1 ng/mL in 20-μL samples). Blank RPMI HAT media spiked with TPT was used as a control. After demonstration of binding from pooled supernatants, the individual wells from that pool were further tested as described above.

## 2.2 Characterization of Anti-Topotecan IgG (ATI)

***Isotype and Binding Affinity/Capacity of ATI.*** The isotypes of secreted antibodies were determined through the use of a mouse monoclonal antibody isotyping kit (Cell Sciences, Norwood, MA). An anti-TPT IgG1 antibody (ATI), secreted by hybridoma 8C2, was selected for further characterization. The 8C2 supernatant (with a constant ATI con-centration) was incubated with increasing concentrations of TPT (100, 125, 150, 175, 200, 225, and 250 ng/mL) for 2 h to allow binding between TPT and ATI. Free TPT was then separated from antibody-bound TPT by ultra-filtration and assayed via HPLC.

The concentration of bound TPT was calculated by subtracting the measured unbound TPT concentration from the known total TPT concentration. TPT–ATI binding was characterized through the construction of Woolf and Rosenthal binding plots, according to the relationships:

$$\frac{F}{B} = \frac{1}{K_A \times nPt} + \frac{F}{nPt} \text{ and } \frac{B}{F} = nPt \times K_A - K_B \times B$$

respectively. In each equation, B represents the molar concentration of TPT bound to ATI, F is the molar concentration of unbound TPT, $K_A$ is the equilibrium association constant (i.e., a measure of binding affinity), and nPt is the total antibody binding capacity. Initial estimates for $K_A$ were determined as the inverse of the intercept times the slope of the Woolf plot or as the negative slope of the Rosenthal plot. The TPT binding capacity (nPt) was calculated as the inverse of the slope of the Woolf plot or as the negative ratio of intercept over slope of the Rosenthal plot. Final parameter estimates were determined by fitting the parameters of the equation:

$$B = \frac{nPt \times F}{1/K_A + F}$$

to the unbound and bound TPT concentration data using WinNonLin v.2.1 software (Pharsight, CA).

### 2.3 Production and Purification of ATI and ATF

***Production of ATI.*** The hybridoma cells secreting ATI were grown in serum-free medium supplemented with 0.5% gentamicin (Hybridoma SFM, Invitrogen). Large quantities of ATI were produced using a Unisyn Cell Pharm hollow fiber bioreactor (10 $ft^2$, 10-KDa molecular-weight cutoff, BioVest International, Minneapolis, MN). Briefly, seed hybridoma cells were first grown in culture within 1-L spinner flasks kept in a $CO_2$ incubator (Model 2100, VWR, West Chester, PA). Approximately $1 \times 10^9$ hybridoma cells in 60 mL of serum-free medium were loaded aseptically into the extracapillary space of the bioreactor cartridge. The temperature was maintained at 37°C, and RPMI 1640 medium (Fisher, Pittsburgh, PA), supplemented with 5-g/L glucose and 5-mg/L gentamicin, was fed into the intracapillary space. Medium containing the anti-TPT monoclonal antibody was harvested from the extracapillary space at a rate of 30–75 mL/day and was replaced with fresh serum-free medium. The collected media were centrifuged for 10 min at 7000 rpm and filtered with a sterile 0.22-μm cellulose acetate bottle-top filter (Corning) before purification.

***Purification of ATI.*** ATI was purified from culture medium via protein-G affinity chromatography (HiTrap Protein-G, Pharmacia, Piscataway, NJ) using an automated BioLogic medium-pressure chromatography system (Bio-Rad, Hercules, CA). Briefly, culture medium was loaded onto the column, which was then washed with 20-mM $Na_2HPO_4$ (pH 7.0), and antibody was eluted using 100-mM glycine buffer (pH 2.8). The ATI concentrations were assessed by UV absorbance at 280 nm, with the assumption that 1-mg/mL ATI corresponds to 1.35 AU [11].

***Production and Purification of Anti-TPT Fab (ATF).*** ATF was prepared by papain digestion of ATI. Briefly, ATI at a concentration of 5 mg/mL in sodium acetate buffer (pH 5.5) was incubated in 37°C water in the presence of papain, cysteine, and ethylene diamine tetraacetic acid (EDTA) for 6 h. The reaction was

stopped by the addition of incubation of 305-mg iodoacetamide, and subsequent incubation with the reaction mixture for 30 min. The digested mixture was dialyzed against 5-mM $KH_2PO_4$ (pH 6.0) overnight. ATF fragments were then purified from the dialysate by hydroxyapatite chromatography (BioRad) with 5-mM $KH_2PO_4$ (pH 6.0) as loading buffer and 400-mM $K_2HPO_4$ (pH 6.0) as elution buffer. The concentration of ATF was also assessed by UV absorbance as described above.

***Assessment of Antibody Purity by SDS-PAGE.*** The purity of ATI and ATF was assessed through sodium dodecyl sulphate–polyacrylamide gel electrophoresis (SDS–PAGE) using a 10% polyacrylamide gel with a 4% stacking gel. Purified ATI (2 μg), purified ATF (2 μg), and the ATI digest mixture were prepared in a nonreducing sample buffer and boiled for 5 min. Samples were then loaded on the gel, and electrophoresed at 140 V for 15 min and 200 V for 35 min. The gel was stained with Coomasie blue R-250 (Bio-Rad) and destained with 40% methanol/10% acetic acid overnight, and the migration distances of ATI and ATF were compared with the migration distances of molecular-weight standards (MW-SDS-200, Sigma).

## 2.4 Development of ELISA for Quantification of ATI and ATF in Rat Plasma

Using a similar approach as described by Hansen and Balthasar [12], a species-specific ELISA was developed to quantify ATI and ATF in rat plasma. Briefly, goat anti-mouse IgG (250 μL, 1 : 500 dilution with 20-mM $Na_2HPO_4$ buffer, no pH adjustment, Sigma) was immobilized on Nunc Maxisorp 96 well microplates (Roskilde, Denmark) overnight at 4°C. On the next day, wells were washed with phosphate buffer containing 0.05% Tween (Sigma) to remove unbound protein, 250 μL of standards or quality control (QC) samples were then added, and the plates were incubated for 2 h. The wells were washed again and incubated with 250 μL of 1 : 500 diluted Fab-specific goat anti-mouse antibody-alkaline phosphatase conjugate (in 20-mM $Na_2HPO_4$ buffer, no pH adjustment) for 1 h. After washing, 200 μL of p-nitro phenyl phosphate (Pierce), 4 mg/mL in diethanolamine buffer (pH 9.8), was added and the change of absorbance at 405 nm was monitored with a micro-plate reader (Spectra Max 250; Molecular Devices, Sunnyvale, CA).

Standard curves were constructed by plotting the change in absorbance with time (dA/dt) versus concentration of ATI or ATF. Standards for the assay were prepared by dilution of stock solutions of ATI or ATF to the appropriate concentrations (0, 25, 50, 100, 175, and 250 ng/mL for ATI; 0, 5, 10, 20, 50, and 100 ng/mL for ATF) with PBS (pH 7.4), with the addition of 1% (v/v) blank rat plasma. Assays were validated with respect to precision and accuracy, by analysis of QC samples at 25, 100, 250 ng/ml for ATI and 5, 20, 100 ng/ml for ATF, respectively. Intra-assay and inter-assay variability were determined through the analysis of QC samples.

## 2.5 Pharmacokinetic Studies of ATI and ATF in Rats

Female Sprague–Dawley rats (Harlan, Indianapolis, IN), 180–200 g, were instrumented with jugular vein cannulas under ketamine/xylazine (90/10 mg/kg) anesthesia. Phar-macokinetic studies for ATI were conducted 2–3 days after surgery by i.v. administration of 1-, 15-, or 100-mg/kg ATI (n = 4/group) via jugular vein cannula. Blood samples (200 μL) were collected through cannula at 1, 3, 6, 12, 24, 48, 96, and 168 h. For ATF studies, the antibody fragments were administered i.v.

at doses of 1, 15, and 60 mg/kg (n = 4/group) via the jugular vein cannula, and blood samples were collected at 0.25, 0.5, 1, 3, 6, 12, and 24 h. After the collection of blood, plasma was immediately isolated by centrifugation at 10,000 *g* for 2 min. Plasma samples were stored at –20°C before to analysis by ELISA.

***Pharmacokinetic Analysis.*** Standard noncompartmental analyses were conducted to assess ATI and ATF pharmacokinetics using WinNonlin software (v. 2.1) (Pharsight, Mountain View, CA). The areas under the plasma concentration versus time curve from time zero to infinity ($AUC_{inf}$) were determined via the log-linear trapezoidal method. The terminal half-life was determined from the relationship of $t_{1/2} = \ln 2/\lambda$, where $\lambda$ is the negative slope of the terminal phase of the lnC versus time plot. Systemic clearance (CL) was estimated by dividing the administered dose by $AUC_{inf}$. The volume of distribution at steady state ($V_{ss}$) was determined by the product of clearance and the mean residence time.

***Statistics.*** One-way analysis of variance and Bonferroni's post hoc tests were used to compare values of half-life and systemic clearance for ATI and ATF. Differences were considered to be significance when $p < 0.05$.

## 3 RESULTS

### 3.1 Development of Hybridomas Secreting Anti-TPT Antibodies

TPT is a small hapten with an MW of 421 D, and to stimulate an immune response, TPT was coupled to keyhole limpet hemocyanin via the Mannich reaction. The resultant immunoconjugate preparation was found to contain 90 ± 42 molecules of TPT per molecule of KLH.

After the immunization of mice with KLH–TPT, each mouse was found to show high TPT binding activity in plasma. Fusion of spleen cells with myeloma cells led to the generation of >500 hybridoma cells. Hybridoma 8C2 was identified as showing significant anti-TPT activity via HPLC. As shown in Figure 1, the ultra-filtrate of control media (blank RPMI HAT media) that was spiked with TPT demonstrated a TPT peak area of $2.5 \times 10^6$ units. The response from culture supernatant pool #17, which consisted of media pooled from six individual hybridomas, showed a >50% decrease in the TPT peak area. Further analysis of each hybridoma contributing to this pooled mixture revealed that the media in the well of 8C2 contained antibody against TPT, as the unbound TPT peak area decreased to 5.2% of control (Figure 1C).

### 3.2 Characterization of ATI

The antibody secreted by 8C2 was identified as an IgG1 (data not shown). Good linearity was obtained in the Woolf plot ($R^2 = 0.9999$, Figure 2) and in the Rosenthal plot ($R^2 = 0.9997$, Figure 3). Additionally, both plots provided very close estimates of the apparent binding affinity ($K_A = 4.28 \times 10^8 M^{-1}$ from Woolf plot and $5.01 \times 10^8 M^{-1}$ from Rosenthal plot) and identical values for binding capacity ($nPt = 3.2 \times 10^{-7} M$ from both plots). The curves were consistent with the existence of a single "type" of binding site, as expected for a monoclonal antibody. Final

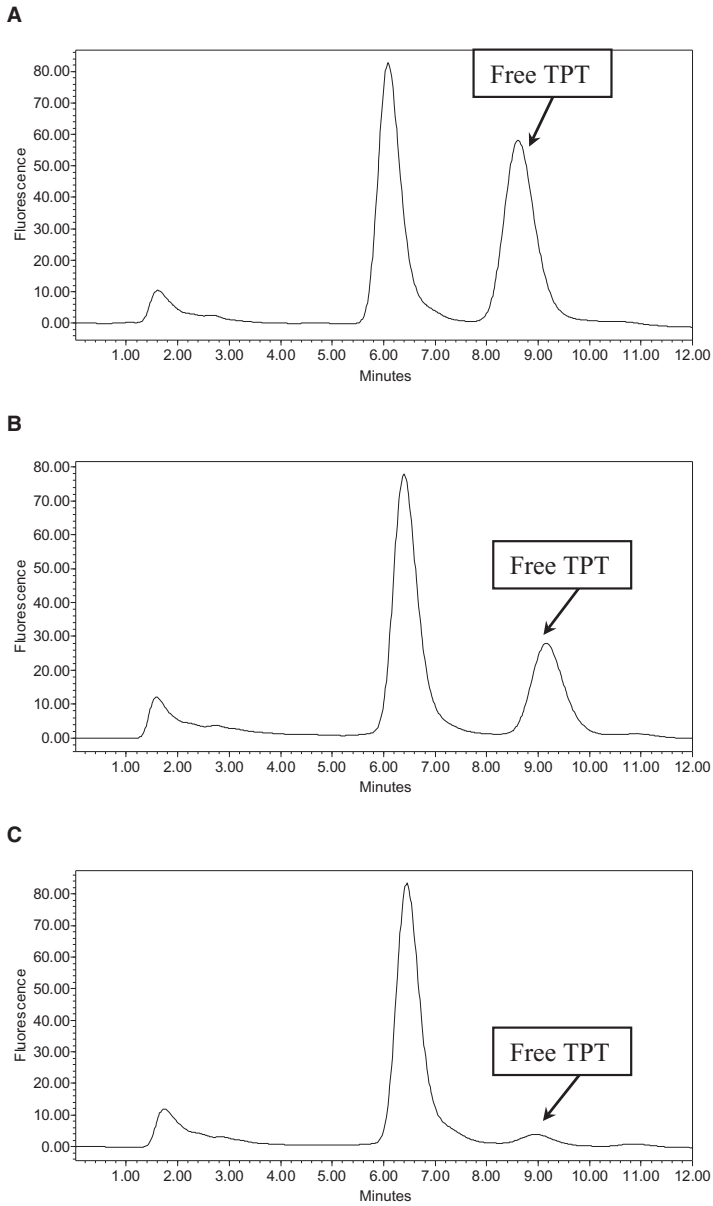


**Figure 1.** HPLC screening of positive hybridomas secreting ATAb (see detailed descriptions in text). (A) Negative control: blank HAT media spiked with TPT. (B) Response from the plate # 17 column 1 (well A1, B1, C1, D1, E1, and F1). (C) Response from 17A1 (8C2).

estimates of binding parameters obtained via nonlinear least-squares fitting were $K_A = 4.8 \times 10^8 M^{-1}$ (CV = 5.2%) and nPt = $3.2 \times 10^{-7}$ M (CV = 0.46%, Figure 4). We could not detect unbound TPT concentrations below the dissociation constant $K_D$ value ($K_D = 1/K_A = 1.01$ ng/L) due to limitations in the sensitivity of the HPLC assay; nonetheless, the accuracy of the estimated value of $K_A$ is supported by the

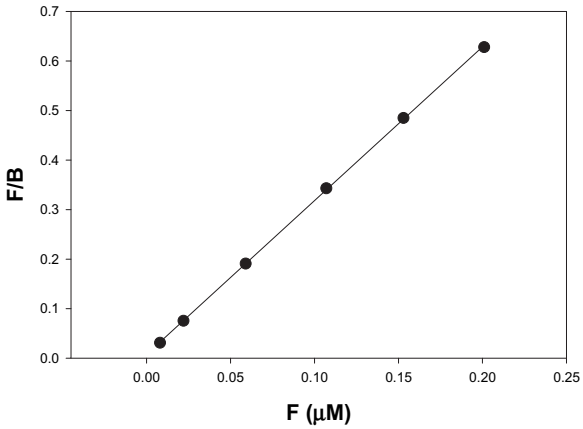


**Figure 2.** The Woolf plot was constructed according to the relationship: $\frac{F}{B} = \frac{1}{K_A \times Pt} + \frac{F}{nPt}$, where B is the molar concentration of TPT bound to ATI and F is the molar concentration of free TPT. $K_A$ is the binding affinity of the anti-TPT antibody, and nPt is the total antibody capacity within the media. The slope is $3.13\,\mu M^{-1}$, and the intercept is 0.0073.

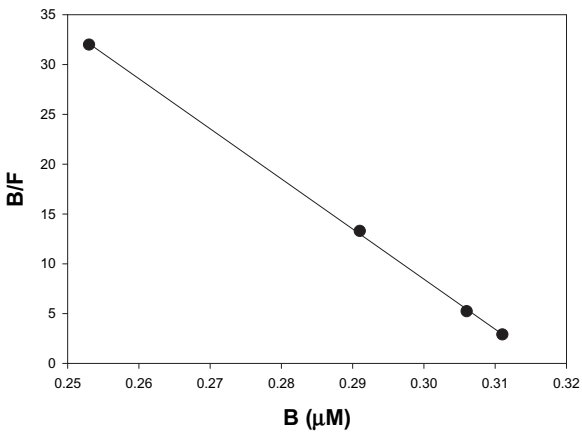


**Figure 3.** The Rosenthal plot was constructed according to the relationship: $\frac{B}{F} = nPt \times K_A - K_A \times B$, where B is the molar concentration of TPT bound to ATI and F is the molar concentration of free TPT. $K_A$ is the binding affinity, of the anti-TPT antibody and nPt is the total antibody, capacity within the media. The slope is $-501.42\,\mu M^{-1}$, and the intercept is 158.85.

excellent linearity obtained in both Woolf and Rosenthal plots and the low coefficient of variance associated with the fitted value.

### 3.3 Production and Purification of ATI and ATF

Antibody production from 8C2 cells growing within the hollow fiber bioreactor reached 240 mg/wk by the third week. ATI was purified from the media through protein G affinity chromatography (Figure 5). Final ATI concentrations in the

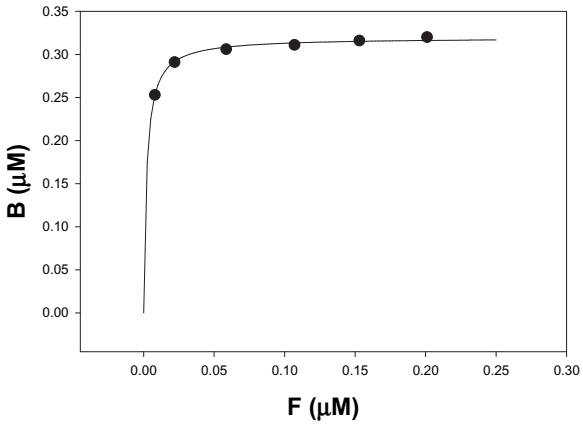


**Figure 4.** Bound vs. free TPT plot for 8C2 by fitting the free and bound TPT concentration data to the following equation: $B = \frac{nPt \times F}{1/K_A + F}$, where B is the molar concentration of TPT bound to ATI, F is the molar concentration of free TPT, $K_A$ is the binding affinity of the anti-TPT antibody, and nPt is the total antibody capacity within the media.

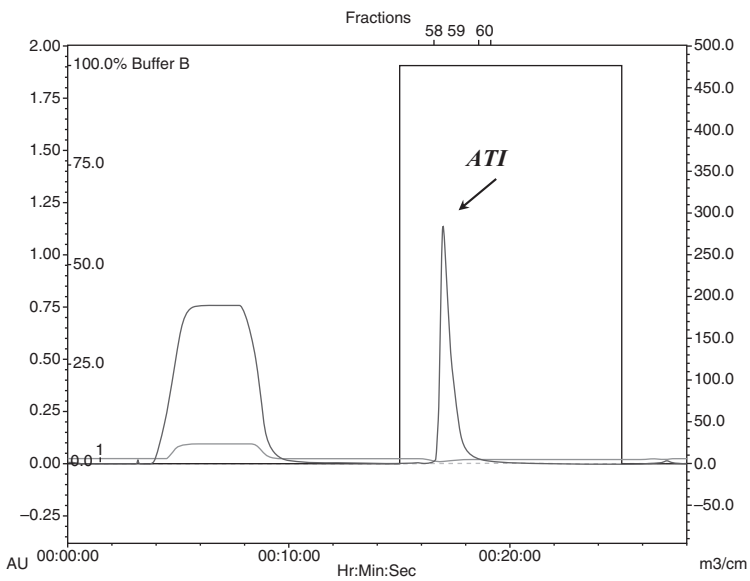


**Figure 5.** Purification of anti-topotecan IgG by a protein G column chromatograph. ATI containing media was loaded onto the column using 20-mM $Na_2HPO_4$ (pH 7.0) and eluted using 100-mM glycine buffer (pH 2.8) at a flow rate of 4 mg/mL (the line represents the percentage of elute buffer onto the column, indicated by buffer B on the y-axis).

elution buffer were typically 0.5–1.2 mg/mL. ATF was produced from the digestion of ATI with papain, and ATF was then purified from undigested ATI, Fc fragments, and secondary digestion products via hydroxyapatite chromatography (Figure 6). The purity of purified ATI and ATF were assessed by SDS–PAGE, and both ATI and ATF migrated as clear, single bands, apparently free from high- and low-molecular-weight contaminants (Figure 7). Estimated molecular weights were consistent with expectations for IgG (~150 KD) and Fab (~50 KD).

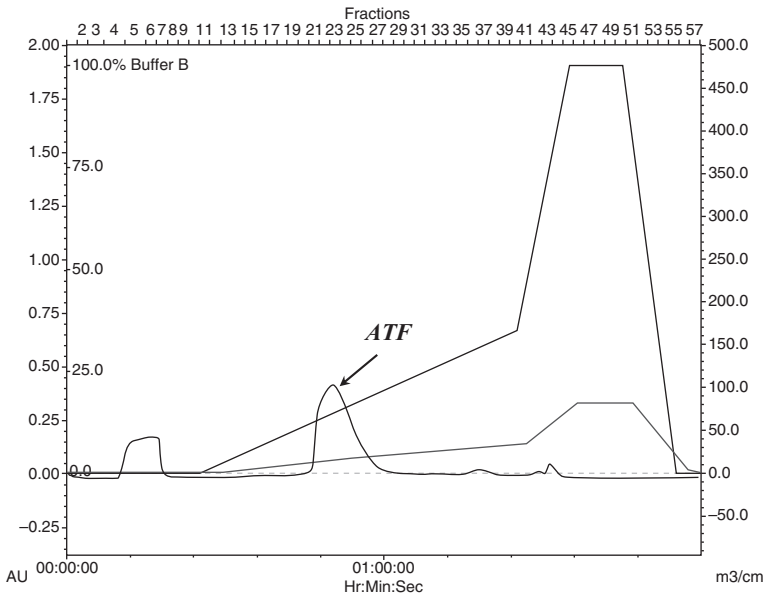


**Figure 6.** Purification of anti-topotecan Fab (ATF) by a hydroxyapatite chromatograph. ATF fragments were then purified from the dialysate by hydroxyapatite chromatograph with 5-mM $KH_2PO_4$ (pH 6.0) as loading buffer and 400-mM $K_2HPO_4$ (pH 6.0) as elution buffer (the line represents the percentage of elute buffer onto the column, indicated by buffer B on the y-axis).

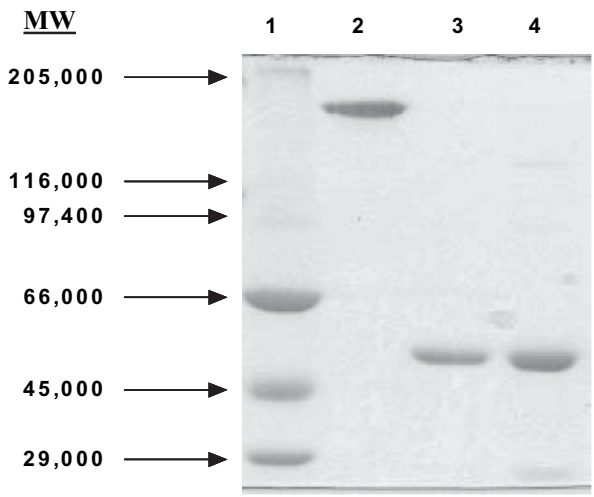


**Figure 7.** SDS–PAGE analysis of the purity of ATI and ATF. Lane 1 shows the molecular weight standards; lane 2 was the purified ATI (2 μg) collected from a protein G chromatograph of 8C2 serum-free media; lane 3 was the purified ATF (2 μg) collected from a hydroxyapatite chromatograph from a digested ATI mixed solution; lane 4 was an ATI digested mixed solution.

### 3.4 ELISA Validation

Representative standard curves are shown in Figure 8 (ATI) and Figure 9 (ATF). The curves were linear over the range of 0–250 ng/mL for ATI and 0–100 ng/mL for ATF. Intra-assay recovery for ATI ranged from 88.3% to 103%, with CVs less than 12%. The interassay recovery for ATI ranged from 94.4% to 102%, with CVs less than 8% (Table 1A). Similarly, the assay for ATF demonstrated an intra-assay recovery of 94.2% to 108%, with CVs less than 11%, and interassay recoveries ranged from 95.1% to 95.8%, with CVs less than 13% (Table 1B).

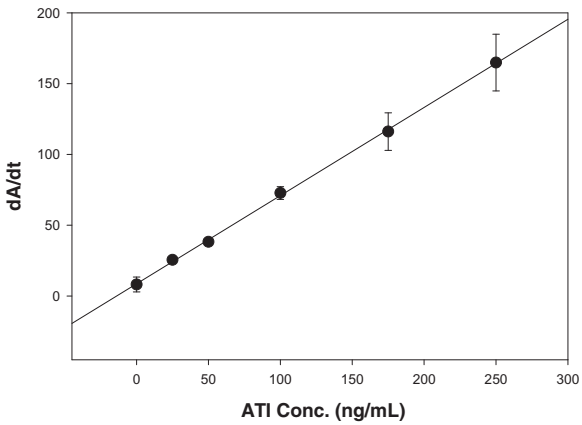


**Figure 8.** Representative ELISA standard curve for ATI over the range of 0–250 ng/mL ($r^2$ = 0.999). Error bars represent the standard deviation of the mean of the three replicates.

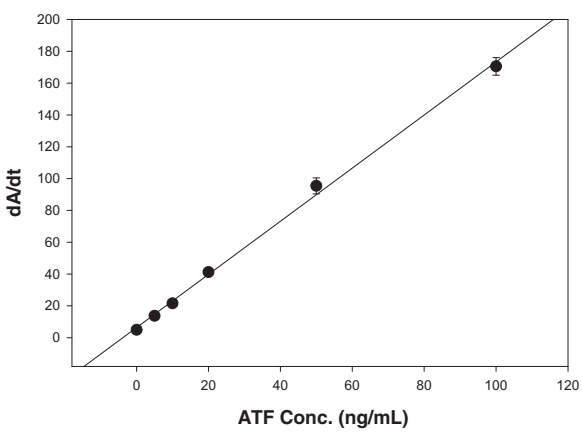


**Figure 9.** Representative ELISA standard curve for ATF over the range of 0–100 ng/mL. Analysis of linear regression shows an $r^2$ = 0.998. Error bars represent the standard deviation of the mean of the three replicates.

**TABLE 1. The Accuracy and Precision (CV%) Of ELISA Associated with Assayed Concentrations of Quality Control Samples for (A) ATI and (B) ATF in Rat Plasma Samples (n = 3)**

| (A) ATI | | | | |
|---|---|---|---|---|
| | QC Conc. (ng/mL) | Assayed Conc. (ng/mL) | CV (%) | Recovery (%) |
| Intra-assay | 25 | 22.1 | 11.9 | 88.3 |
| | 100 | 103 | 3.0 | 103 |
| | 250 | 252 | 5.0 | 101 |
| Interassay | 25 | 23.9 | 7.1 | 95.7 |
| | 100 | 102 | 4.0 | 102 |
| | 250 | 236 | 8.0 | 94.4 |
| **(B) ATF** | | | | |
| | QC Conc. (ng/mL) | Assayed Conc. (ng/mL) | CV (%) | Recovery (%) |
| Intra-assay | 5 | 5.4 | 9.7 | 108 |
| | 20 | 18.9 | 10.6 | 94.4 |
| | 100 | 94.2 | 4.5 | 94.2 |
| Interassay | 5 | 4.8 | 12.5 | 95.8 |
| | 20 | 19.0 | 9.9 | 95.1 |
| | 100 | 95.1 | 3.2 | 95.1 |

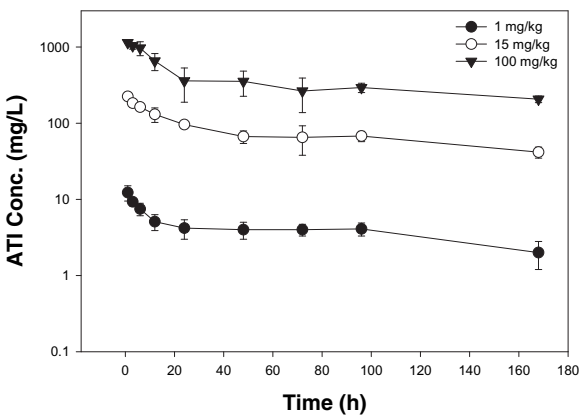


**Figure 10.** Plasma concentration-time profile after i.v. bolus of ATI at 1, 15, and 100 mg/kg to rats (n = 4/group). Blood samples were collected at 1, 3, 6, 12, 24, 48, 96, and 168 h. Plasma concentrations of ATI were determined by ELISA.

### 3.5 Pharmacokinetics of ATI and ATF

ATI plasma concentration versus time profiles after i.v. administration of 1, 15, and 100 mg/kg are shown in Figure 10, whereas Figure 11 shows ATF plasma concentration versus time profiles after i.v. bolus administration of 1, 15, and 60 mg/kg to rats. Non-compartmental analyses of the parameters for ATI and ATF are summarized in Table 2A and in Table 2B, respectively. ATI systemic clearance and terminal half-life seem to be dose independent, as no statistical differences were found over the dose range of

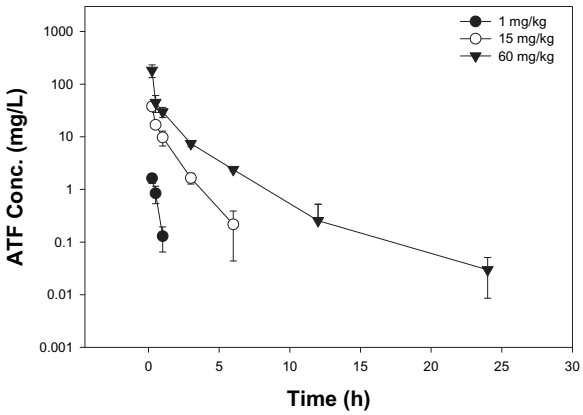


**Figure 11.** Plasma concentration-time profile after i.v. administration of ATF at 1, 15, and 60 mg/kg in rats (n = 4/group). Blood samples were collected at 0.25, 0.5, 1, 3, 6, 12, and 24 h. Plasma concentrations for ATF were determined by ELISA.

**TABLE 2. Pharmacokinetic Parameters Estimated by Noncompartmental Analysis for (A) ATI and (B) ATF**

| (A) ATI | | | | |
|---|---|---|---|---|
| Parameter | Unit | 1 mg/kg | 15 mg/kg | 100 mg/kg |
| $AUC_{inf}$ | mg/L*h | 689 ± 82.8 | $(1.46 \pm 0.227) \times 10^4$ | $(1.04 \pm 0.292) \times 10^5$ |
| $t_{1/2}$ | $h^{-1}$ | 105 ± 92.9 | 118 ± 18.1 | 154 ± 53.5 |
| $V_{ss}$ | L/kg | 0.120 ± 0.030 | 0.110 ± 0.020 | 0.200 ± 0.030* |
| CL | mL/h/kg | 1.08 ± 0.330 | 0.750 ± 0.100 | 1.00 ± 0.250 |
| (B) ATF | | | | |
| Parameter | Unit | 1 mg/kg | 15 mg/kg | 60mg/kg |
| $AUC_{inf}$ | mg/L*h | 1.17 ± 0.280 | 38.8 ± 6.80 | 189 ± 39.7 |
| $t_{1/2}$ | $h^{-1}$ | 0.360 ± 0.320 | 0.850 ± 0.24 | 1.97 ± 1.32 |
| $V_{ss}$ | L/kg | 0.330 ± 0.060 | 0.340 ± 0.030 | 0.310 ± 0.160 |
| CL | L/h/kg | 0.890 ± 0.220* | 0.400 ± 0.070 | 0.330 ± 0.060 |

*$p < 0.01$

1 to 100 mg/kg ($p > 0.05$). No differences were found for the estimated steady-state volume of distribution of ATI after dosing with 1 and 15 mg/kg, but the value of $V_{ss}$ estimated for the 100-mg/kg group was significantly higher ($p < 0.01$). ATF terminal half-life and steady-state volume of distribution were not dependent on dose ($p > 0.05$ in the dose range from 1 to 60 mg/kg). No differences were found for the systemic clearance of ATF between 15 and 60 mg/kg, but ATF clearance estimated from 1 mg/kg was found to be significantly higher than that observed for the other doses ($p < 0.01$).

## 4 DISCUSSION

This laboratory is investigating the use of anti-drug antibodies within an inverse targeting strategy that attempts to increase the pharmacokinetic and therapeutic selectivity of intraperitoneal chemotherapy. The approach combines i.p. chemotherapy

with systemic administration of anti-drug antibodies or antibody fragments (collectively named ADAb) [1, 2]. It is anticipated that ADAb in the blood may rapidly bind drug that diffuses out of the peritoneum, effectively reducing peak plasma-free (i.e., unbound) drug concentrations, the cumulative systemic exposure to unbound drug, and the extravascular distribution of the anticancer drug. In addition, it is anticipated that ADAb will likely show very slow entry into the peritoneal cavity, leading to little change in the cumulative peritoneal exposure to unbound drug. As such, the proposed targeting strategy may produce desirable regio-specific alterations in drug disposition, which may enhance to therapeutic selectivity of i.p. chemotherapy.

These hypotheses have been supported by a series of preclinical studies that employed MTX and anti-MTX Fab as model drugs [2, 4, 5]. However, MTX is not an ideal drug for future clinical applications of the inverse targeting strategy, due to its limited ability to penetrate solid tumors [13] and due to its relatively low potency. TPT may be an ideal drug for use in the proposed approach; this agent is a preferred therapy for metastatic, refractory ovarian cancer, and TPT is very potent, with a maximum tolerated dose of ~5 mg in humans (via i.p. instillation) [14].

To allow evaluation of TPT within the inverse targeting strategy, we have developed anti-TPT monoclonal antibodies. We applied traditional fusion techniques [11], which led to the generatation of >500 clones per fusion (in our hands). In our initial attempts to screen for hybridomas secreting anti-TPT antibodies, we first developed an antibody capture ELISA [11]. However, the antibody capture assay produced high background responses to a variety of proteins (e.g., bovine serum albumin and ovalbumin), even under stringent blocking and washing conditions (e.g., with various blocking agents such as 3% albumin and 20% milk). Displacement ELISAs also showed high background values when incubated with saline relative to TPT (data not shown). Due to the lack of selectivity associated with these immunoassays, we elected to screen for anti-TPT activity via HPLC.

For screening purposes, HPLC is not an ideal method due to its relatively low sensitivity and due to the low throughput of HPLC (i.e., relative to ELISA). In most cases, the identification of antibody binding via HPLC would require that the assay limit of quantification (LOQ) be substantially lower than the concentration of antibody in the supernatant (nPt), and in most cases, identification of antibody by HPLC would require that the product of LOQ and $K_A$ be approximately equal to or greater than 1. Given that the LOQ of our HPLC assay for TPT was ~$2 \times 10^{-9}$ M, we anticipated that we would only identify those hybridomas with high antibody production and those that secrete high-affinity anti-TPT antibodies ($K_A \sim 5 \times 10^8\,M^{-1}$). From our studies with ~1000 hybridomas generated, two hybridomas were found to secrete antibodies in sufficient concentration and with sufficient affinity to allow detection by HPLC. One hybridoma was found to secrete an IgA anti-TPT antibody, and the other hybridoma, 8C2, was found to secrete an IgG1 anti-TPT antibody. Hybridoma 8C2 was selected for further use as, relative to IgA, it is generally easier to produce, purify, and digest IgG into Fab fragments. Consistent with our expectations, nonlinear least-squares regression revealed that 8C2 secreted an anti-TPT IgG with a high binding affinity ($K_A = 4.8 \times 10^8\,M^{-1}$).

To facilitate pharmacokinetic studies for ATI and ATF, a species-specific ELISA was developed and validated to quantify murine ATI and ATF in rat plasma, based

on the method of Hansen and Balthasar [12]. The current assay demonstrated similar CV% values for inter- and intra-assay variability for IgG (Hansen and Balthasar reported less than 8.5% of CV% value for intra-assay variability and less than 25.1% for inter assay variability) [12].

Noncompartmental analysis of ATI plasma data indicated that ATI distributed into a steady-state volume of distribution of 110–200 mL/kg, and that ATI was cleared at a rate of 0.75–1.08 mL/h/kg, with a terminal half-life of 105–154 h. These values are similar to those reported by Bazin-Redureau et al. after single-dose (0.7 mg/kg) i.v. administration of a monoclonal murine IgG1 to rats [15]. In their study, they reported $V_{ss}$ = 125 ± 4.0 mL/kg, systemic clearance = 0.48 ± 0.05 mL/h/kg, and a terminal half-life = 194 ± 19 h.

After i.v. administration, ATF distributed into a steady-state volume of distribution of 0.31–0.33 L/kg, similar to the value (0.38 ± 0.03 L/kg) reported by Bazin-Redureau et al. [15] and similar to the value that reported by Pentel (0.43 ± 0.06 L/kg) in a study of high-dose (7.5 g/kg) human Fab administered to rats [16]. The terminal half-life of ATF was determined to be 0.36–1.97 h, which is comparable with a reported value of 1.71 ± 0.25 h, after administration of radio-iodinated Fab to rats [17], and to the half-life of an anti-MTX Fab (1.36 h) [18], but considerably shorter than the values reported by Bazin-Redureau et al. (9.84 ± 0.74 h) [15] and by Pentel et al. (16.3 ± 2.4 h) [16]. The systemic clearance of ATF was 0.33–0.89 L/h/kg, which is higher than the value reported by McClurkan et al. [19] (0.16 ± 0.05 L/h/kg, after i.v. administration of 120-mg/kg Fab in rat) and markedly higher than the value found by Pentel et al. (for high dose of Fab to rats, 0.03 ± 0.004 L/h/kg) [16] and the value reported by Bazin-Redureau et al. (0.03 ± 0.0004 L/h/kg) [15]. Some differences in IgG or Fab pharmacokinetic parameters may be due to the differences in the species of origin of IgG or Fab (human [16] or rat [17] versus mouse in our study), differences in the types of antibodies preparation (polyclonal versus monoclonal), differences in the dose level studied (possible dose-dependent kinetics), or differences in the sensitivity of the analytical techniques used for quantifying Fab concentrations (radiolabeled antibodies quantification versus ELISA). Additional experiments are underway to investigate the pharmacokinetics of anti-TPT Fab more completely (i.e., over a wider dose range, and using more extensive sampling) in animals bearing peritoneal tumors.

In summary, we describe the production, purification, characterization, and pharmacokinetics of a monoclonal anti-TPT IgG and Fab fragments. The anti-TPT antibodies have demonstrated high affinity and sufficient purity for use in evaluating their use in "inverse targeting" strategies to optimize TPT i.p. chemotherapy.

## REFERENCES

1. Balthasar J, Fung H L (1994). Utilization of antidrug antibody fragments for the optimization of intraperitoneal drug therapy: studies using digoxin as a model drug. *J. Pharmacol. Experim. Therapeu*. 268:734–739.
2. Balthasar J P, Fung H L (1996). Inverse targeting of peritoneal tumors: Selective alteration of the disposition of methotrexate through the use of anti-methotrexate antibodies and antibody fragments. *J. Pharm. Sci*. 85:1035–1043.

3. Balthasar J P, Fung H L (1995). High-affinity rabbit antibodies directed against methotrexate: production, purification, characterization, and pharmacokinetics in the rat. *J. Pharm. Sci.* 84:2–6.
4. Lobo E D, Soda D M, Balthasar J P (2003). Application of pharmacokinetic-pharmacodynamic modeling to predict the kinetic and dynamic effects of anti-methotrexate antibodies in mice. *J. Pharm. Sci.* 92:1665–1676.
5. Lobo E D, Balthasar J P (2005). Application of anti-methotrexate Fab fragments for the optimization of intraperitoneal methotrexate therapy in a murine model of peritoneal cancer. *J. Pharm. Sci.* 94:1957–1964.
6. Markman M (1997). Topotecan: An important new drug in the management of ovarian cancer. *Semin. Oncol.* 24(1 Suppl 5):S5–1.
7. Herben V M, ten Bokkel Huinink W W, Beijnen J H (1996). Clinical pharmacokinetics of topotecan. *Clin. Pharmacokinet.* 31:85–102.
8. Hsiang Y H, Liu L F (1988). Identification of mammalian DNA topoisomerase I as an intracellular target of the anticancer drug camptothecin. *Cancer Res.* 48:1722–1726.
9. Hsiang Y H et al. (1985). Camptothecin induces protein-linked DNA breaks via mammalian DNA topoisomerase I. *J. Biol. Chem.* 260:14873–14878.
10. Chen J, Balthasar J P (2005). High-performance liquid chromatographic assay for the determination of total and free topotecan in the presence and absence of anti-topotecan antibodies in mouse plasma. *J. Chromatogr. B. Analyt. Technol. Biomed. Life Sci.* 816:183–192.
11. Harlow E, Lane D (1988). *Antibodies, A Laboratory Manual.* Cold Spring Harbor Laboratory, Cold Spring Harbor, M.I., p. 276.
12. Hansen R J, Balthasar J P (1999). An ELISA for quantification of murine IgG in rat plasma: An application to the pharmacokinetic characterization of AP-3, a murine anti-glycoprotein IIIa monoclonal antibody, in the rat. *J. Pharmaceu. Biomed. Anal.* 21:1011–1016.
13. West G W, Weichselbaum R, Little J B (1980). Limited penetration of methotrexate into human osteosarcoma spheroids as a proposed model for solid tumor resistance to adjuvant chemotherapy. *Cancer Res.* 40:3665–3668.
14. Plaxe S C et al. (1998). Phase I and pharmacokinetic study of intraperitoneal topotecan. *Invest. New Drugs.* 16:147–153.
15. Bazin-Redureau M et al. (1998). Interspecies scaling of clearance and volume of distribution for horse antivenom F(ab′)2. *Toxicol. Appl. Pharmacol.* 150:295–300.
16. Pentel P R et al. (1988). Pharmacokinetics and toxicity of high doeses of antibody Fab fragments in rats. *Drug Metabol. Dispos.* 16:141–145.
17. Arend W P, Silverblatt F J (1975). Serum disappearance and catabolism of homologous immunoglobulin fragments in rats. *Clin. Exp. Immunol.* 22:502–513.
18. Tayab Z R, Balthasar J P (2004). Development and validation of enzyme-linked immunosorbent assays for quantification of anti-methotrexate IgG and Fab in mouse and rat plasma. *J. Immunoassay Immunochem.* 25:335–344.
19. McClurkan M B et al. (1993). Disposition of a monoclonal anti-phencyclidine Fab fragment of immunoglobulin G in rats. *J. Pharmacol. Experim. Therapeu.* 266:1439–1445.

# RECOMBINANT ANTIBODIES FOR PATHOGEN DETECTION AND IMMUNOTHERAPY

Nicholas J. Pokorny, Jeanine I. Boulter-Bitzer, J. Chris Hall, Jack T. Trevors, and Hung Lee

*University of Guelph, Guelph, Ontario, Canada*

**Chapter Contents**

1 Introduction
2 General Properties of Antibodies
  2.1 Binding Properties
  2.2 Monoclonal Antibodies Produced from Hybridoma Cell Lines
  2.3 Recombinant Antibody (rAb) Technology
  2.4 Antibody Libraries
  2.5 Soluble Antibody Production
3 Immunodetection of Pathogens
  3.1 Formats of Detection
4 Immunotherapy
  4.1 Passive Immunotherapy
  4.2 Vaccine/Active Immunotherapy
5 Conclusions
  Acknowledgments
  References

## 1 INTRODUCTION

Antibody technology has existed since the establishment of the hybridoma monoclonal system [1]. In recent years, there has been increasing interest in developing recombinant antibodies (rAb) for immunodiagnostics and immunotherapeutical

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

applications, with the expectation that recombinant antibody fragments will capture a significant share of the US$6 billion per year diagnostic market [2]. Recombinant antibodies have reduced conventional monoclonal antibodies into smaller capture molecules, while holding the promise of improving the sensitivity and specificity of many diagnostic assays. This chapter summarizes some new developments in rAb with a particular emphasis in pathogen detection and immunotherapy.

## 2 GENERAL PROPERTIES OF ANTIBODIES

Within the mammalian immune system, there are five general classes of immunoglobulin molecules, IgA, IgD, IgE, IgG, and IgM, differentiated by the structure of their constant regions and their immune function [3]. Immunoglobulins are antigen binding glycoproteins present on the B-lymphocyte (or B-cell) membrane and secreted by plasma cells. IgG is the most abundant class of immunoglobulins in sera. As most antigens offer multiple epitopes, a variety of B-cell clones are induced, each derived from a B-cell that recognizes a particular epitope. The resulting polyclonal antibody serum comprises a heterogeneous mixture of antibodies, each specific to one epitope. The antibodies produced from a single B-cell clone are monoclonal in that they are all specific to a single epitope [3].

The general immunoglobulin structure is a 4-chain "Y"-shaped molecule of two heavy-chain and two light-chain polypeptides joined together by disulfide bridges [3]. The greatest variability of immunoglobulins occurs at the N-terminal, or the tips, of the "Y" molecule. This is where the antibody interacts with the antigen. This variable region includes the ends of both the heavy ($V_H$) and the light ($V_L$) chains. Within the variable regions are three discontinuous regions of the greatest variability, known as the hypervariable (HV) or complementarity-determining region (CDR), which directly contacts a portion of the antigen's surface [2]. Separating the HV regions are four highly conserved framework regions that serves as a scaffold to hold the HV regions together and that forms a unique β-barrel antigen-recognition site. In general, more amino acid residues of the heavy-chain CDRs interact with the antigen than in the light-chain CDRs. In many cases, binding of the antigen induces a conformational change in the antibody, antigen, or both [4].

Antibodies have been used as research tools for cell, antigen, or pathogen identification; pathogenesis studies; ligands for column chromatography and molecule purification; diagnostic reagents; therapeutic antibody preparations; and the identification of protective antigens/epitopes in vaccine development.

It is important that antibodies being used for immunodetection of pathogens or for the immunotherapeutic purposes be able to bind strongly and specifically to the pathogen of interest. The most desired antibodies should possess high affinity and specificity for a given pathogen, with little-to-no cross-reactivity to other antigenic targets.

### 2.1 Binding Properties

The binding properties observed during antibody interactions with their target antigens are generally described in terms of affinity, avidity, and specificity.

***Affinity.*** Affinity is the strength of interaction between an antibody and an antigen at a single antigenic site. Within each antigenic site, the variable region interacts through weak noncovalent bonds with antigen at numerous sites; the more interactions, the stronger the affinity. Antibodies with more available antigen-binding domains will possess a higher intrinsic antigen affinity.

***Avidity.*** Avidity is a measure of the overall stability or strength of the antibody–antigen complex. It is controlled by three major factors: antibody-epitope affinity; the valence of both the antigen and the antibody; and the structural arrangement of the interacting ligands.

***Specificity.*** Specificity is based on the combination of both affinity and avidity to determine the likelihood that the particular antibody is binding to a precise antigen epitope. The matrix on which the antigen is tested may influence specificity and sensitivity by interfering with the antigen–antibody complex. The more specific an antibody, the less chance there is for cross-reactivity with a nontarget antigen and subsequent false-positive results in immunodiagnostic assays. In immunotherapeutic applications, the desired antibody should only target the harmful pathogen and not any indigenous microorganisms.

***Cross-Reactivity.*** Cross-reactivity refers to an antibody or population of antibodies binding to epitopes on other antigens. This can be caused either by low avidity or specificity of the antibody or by multiple distinct antigens having identical or very similar epitopes.

### 2.2 Monoclonal Antibodies Produced from Hybridoma Cell Lines

The fusion of an antibody-producing B-cell with a myeloma cell results in a hybrid cell, called a hybridoma. These cells possess the immortal growth properties of the myeloma cell and secrete the antibody produced by the B-cell. All antibodies produced by the individual hybridoma cell lines have the same antigenic specificity. Monoclonal antibodies (mAbs) require a great investment of resources, as their production is labor intensive and costs much more to develop, produce, validate, and stockpile than polyclonal antibodies [5]. However, once a hybridoma cell line is developed, mAbs can be easily made in large quantities without the lot-to-lot variation observed in polyclonal antibody preparations [6]. Lack of reliable myeloma partners has prevented the hybridoma fusion technique from widespread use in non-rodent species [6]. Monoclonal antibodies are generally of the IgG class. Because of their bivalency, their functional affinity and avidity are greatly increased as compared with monovalent recombinant antibodies [2]. Monoclonal antibodies have been traditionally used in immunoassays for the detection of pathogenic microorganisms.

### 2.3 Recombinant Antibody (rAb) Technology

With the advent of recombinant DNA technology, antibody genes can be amplified and selected through phage display, cell-surface display, or cell-free display systems (i.e., ribosome display). A major advantage shared by these systems is the direct

link between the genotype and the phenotype of displayed antibodies during selection. This allows for simultaneous coselection of the desired antibodies and their encoding genes as a result of the binding characteristics of the displayed antibodies. The gene encoding the desired antibody can be further manipulated for improvements in affinity and/or specificity, increased expression, posttranslation modification, or fusion of a secondary protein [7].

Recombinant DNA technology and antibody engineering have increased the available avenues of antibody use, particularly in the application of immunotherapies. Recombinant monoclonal antibodies from various animals can now be produced without a myeloma cell line as a fusion partner. Animals can be immunized by exposure to a desired antigen, and the antibody genes from these animals are amplified by polymerase chain reaction (PCR) and expressed in different expression systems in various formats, such as the Fab, single-chain variable fragment (scFv), and single-domain antibodies ($V_H$, $V_{HH}$, and $V_{NAR}$). Figure 1 illustrates some different recombinant antibody fragments and the parent antibodies from which they were derived. Recombinant Fab [8], scFv [9], $V_H$ [10], $V_{HH}$ [11], and

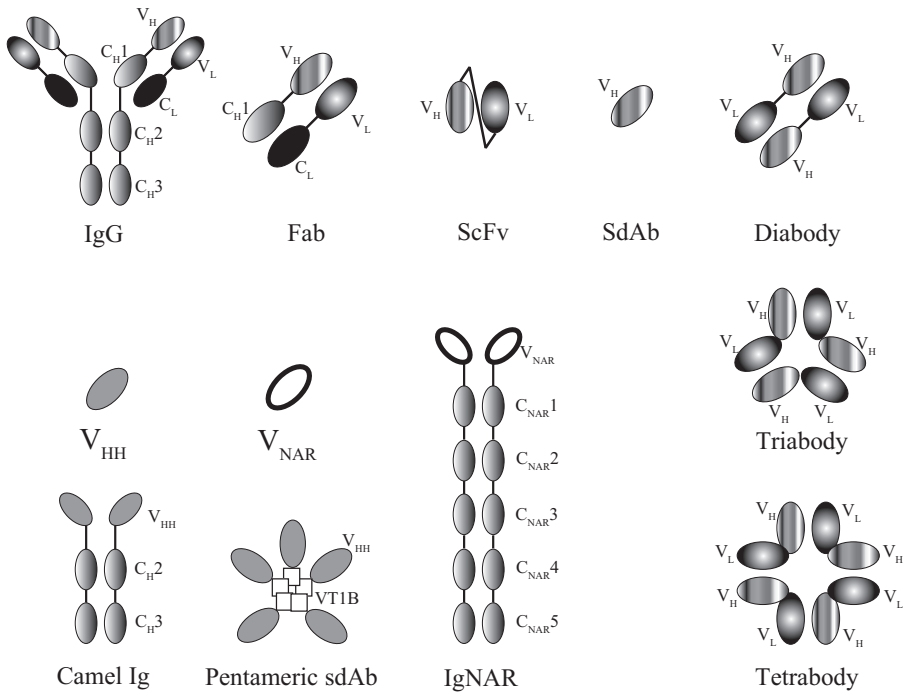


**Figure 1.** Schematic representations of common antibody formats and recombinant antibodies. The structure of the immunoglobulin IgG molecule is shown as the "gold-standard" for antibody structure. Homodimeric Camelid Ig and shark IgNAR are indicated along with their representative single-domain antibodies, $V_{HH}$ and $V_{NAR}$. A variety of common antibody fragments are depicted, as well as some of their multimeric formats. $V_H$, variable domain of heavy chain; $C_H$, constant domain of heavy chain; $V_L$, variable domain of light chain; $C_L$, constant domain of light chain; $V_{HH}$; variable domain of Camelid immunoglobulin; $V_{NAR}$, variable domain of IgNAR; $C_{NAR}$, constant domain of IgNAR; VT1B, B subunit of Verotoxin 1.

$V_{NAR}$ [12] have been synthesized and found to possess binding characteristics comparable with their respective parental monoclonal antibodies.

The major advantage of these recombinant antibody fragments is that they have increased the available avenues of antibody use, particularly in immunotherapies. They are smaller in size and therefore easier to manipulate genetically and can be expressed in bacterial systems in large quantities with little lot-to-lot variation. The rAbs can be expressed in various surface-display systems (i.e., phage display or cell-surface display) while retaining their functionality during the panning process. As the development of rAbs does not depend on species-specific cell fusion partners, rAbs can be developed from any animal for which one can design an appropriate set of primers for immunoglobulin amplification. Expression and purification of rAbs by bacterial fermentation is less expensive, easier to perform, and less time consuming than production of either polyclonal or monoclonal antibodies [13].

Recombinant antibodies have three key disadvantages when compared with conventional monoclonal antibodies: (1) they are less stable *in vivo* than natural antibodies; (2) they cannot cross-link antigens; and (3) they may lack critical domains necessary for certain biological functions [14].

Numerous differing types of rAbs exist that vary depending on their size, conformation, and source. The following describes the general features of some rAb types.

***Fab.*** Fab fragments, as the classic monovalent antibody fragment, can be produced from an IgG by treatment with papain. The Fab fragment is composed of the entire light antibody chain and part of the heavy chain (Figure 1). This configuration retains the antigen-binding domain. The interchain disulfide bond remains intact to provide proper tertiary protein conformation. Fab fragments have been shown to have an *in vivo* half-life on the order of hours [15], making them attractive for use in immunotherapeutic applications, as compared with some other rAbs.

***Scfv.*** This antibody fragment is composed of the variable segments of both the heavy and the light antibody chains. The two fragments are typically joined by a flexible peptide linker, most commonly the 15 amino acids of $(Gly_4Ser)_3$, to improve scFv folding and stability. However, the linker between $V_H$ and $V_L$ can interfere with folding in some cases, and this renders the scFv susceptible to proteolysis [16]. Due to rapid *in vivo* blood clearance of scFv, the half-life is in the range of minutes to tens of minutes [17]. The smaller size of scFv antibodies allows them to more easily penetrate tissues.

***$V_H$.*** These single-domain antibodies are composed entirely of a single-variable antibody domain. Despite early excitement surrounding the functional activity of these, in those that have been studied, they rarely retained the affinity of the parent antibody and were poorly soluble and often prone to aggregation. Exposure of the hydrophobic surface of the $V_H$ to the solvent, which normally interacts with the $V_L$, causes a sticky behavior of the isolated $V_H$s [18]. Because of their small size, it is possible for $V_H$ antibodies to potentially target some epitopes that are inaccessible to larger antibody fragments [10].

***$V_{HH}$.*** Members of the *Camelidae* family have evolved high-affinity single V-like domains that are mounted on an Fc-equivalent constant domain framework. Their

binding domains consist only of the heavy-chain variable domains, referred to as $V_{HH}$, to distinguish it from the conventional $V_H$ antibody fragment. The variable domain $V_{HH}$ is immediately followed by the hinge region consisting of the $C_H2$ and $C_H3$ domains. Affinities of $V_{HH}$s are in the nanomolar range and are comparable with those of Fab and scFv fragments, and they can be stable under high temperatures, with specific binding being retained at temperatures up to 90°C [11, 18].

***$V_{NAR}$.*** The immunoglobulin isotype novel antigen receptor (IgNAR) is a homodimeric heavy-chain complex found in the serum of the nurse shark (*Ginglymostoma cirratum*). IgNAR do not contain light chains [19]. Each molecule consists of a single-variable domain ($V_{NAR}$) and five constant domains ($C_{NAR}$). Similarities between the $V_{NAR}$ and the $V_{HH}$ include the presence of disulfide bonds and binding affinities in the nanomolar range [20].

***Antibody Variants.*** Fab and scFv fragments have been engineered into dimeric, trimeric, or tetrameric conjugates using either chemical or genetic cross-links. By reducing the scFv linker length to between zero and five residues, self-assembly is directed into multimers [2]. Bivalent diabodies are the results of joining the $V_H$ of one scFv to the $V_L$ of a second scFv molecule, and vice versa, with stabilization introduced by disulfide bridges [21]. Single-domain antibodies can be fused to the B-subunit of *Escherichia coli* verotoxin, or shiga-like toxin, which self-assembles to form a homopentamer, resulting in simultaneous antibody pentamerization and introduction of increased avidity [22]. Fusion antibodies have also been produced in which secondary proteins, such as green fluorescent protein, are expressed along with the rAb. This is done as a means to facilitate monitoring of the antibodies, making antibody-based pathogen detection easier [23].

## 2.4 Antibody Libraries

Three general types of libraries are constructed for selection of rAbs: (1) immune, (2) nonimmune or naïve, and (3) or synthetic/semisynthetic. Immune libraries are generated by cloning antibody genes from B-cells of immunized animals. This approach takes advantage of the fact that antibodies directed against the antigen are enriched in such donors, and that they have been subjected to an *in vivo* affinity maturation by the host immune system. These libraries have the added advantage that many isolated antibodies are already antigen specific [24]. Hybridoma cell lines can also be used for the formation of immune libraries. As each hybridoma cell line theoretically produces only a single type of functional antibody, only one set of immunoglobulin genes is available. However, depending on the myeloma cell line used in the fusion process, rearranged but nonfunctional heavy- and light-chain genes in the hybridoma may be present and preferentially amplified and cloned [25]. To select for functional immunoglobulin genes from hybridoma cell lines, intensive screening of cloned genes through panning of phage-display antibodies is necessary [25].

Nonimmune or naïve libraries are constructed in the same manner as an immune library, but B-cells from nonimmunized donors are used as sources for antibody genes. The diversity of isolated antibodies increases with the library size, and

increasing the library size improves the chance of obtaining high-affinity antibodies [24]. Theoretically, a single universal naïve library can be constructed and used to provide rAbs with any desired specificity. In practice, however, such antibodies have affinities that may be suboptimal [26].

Synthetic/semisynthetic libraries are artificially constructed by using PCR to randomly assemble the genes encoding the HV regions from a naïve B-cell. By employing this process, a unique set of $V_H$ and $V_L$ genes are recombined. These are sometimes referred to as single-pot libraries, because each such library can be source of antibodies directed against multiple antigens. This technology can also be used to display the immunoglobulin genes of conventional hybridomas for the purpose of rescuing unstable clones or improving their binding specificities and/or affinities through genetic manipulations such as *in vitro* mutagenesis or heavy/light-chain shuffling [26].

The use of antibody libraries has numerous advantages over hybridoma antibody technology: (1) They allow for affinity-selection of binding clones and not just to screen for binding activity; (2) recombinant antibody libraries can be rescreened against many different antigens; (3) negative subtraction screening can be done against irrevelent nontarget background proteins, thus reducing the number of nonspecific binders; (4) rAbs can be developed from any species for which primers exist; (5) rAb libraries can be applied to high-throughput screening technologies; and (6) they are well suited to many downstream molecular technologies.

***Panning Antibody Libraries.*** Panning is the methodology by which desirable rAbs are selected from an antibody library. During panning, recombinant antibody libraries are incubated with the target antigen to bind clones bearing specificities of interest. Unbound or weak clones are washed away, and the rAbs that remain bound to the antigen are eluted and amplified for further propagation. After two to four rounds of panning, all binders should be specific to the antigen of interest. Antibody-displaying complexes bearing expressed rAb can be used directly in diagnostic assays or can be genetically modified to express soluble antibody molecules [7, 27].

Because the screening for positive clones takes places in one step, millions of clones are evaluated during each round of selection. Thus, this process requires only a few hours instead of days to months required for screening conventional hybridoma cells for monoclonal antibodies. The clones are indefinitely stable and capable of self-replication when supplies are low. They can be stored as bacterial stocks, phage stocks, or plasmid DNA.

***Phage Display.*** Phage display is often used to screen for recombinant antibodies. It is based on the functional expression of peptides and proteins on the surfaces of bacteriophages. The filamentous bacteriophages are a group of viruses that contain a circular single-stranded DNA genome encased in a long cylinder protein capsule. The infection process is facilitated by attachment to the bacterial pilus. The Ff class of filamentous phages has been most extensively studied. They use the F pilus specific to *E. coli* cells containing the F plasmid [27]. These bacteriophages do not kill their host during reproduction. Rather, the bacterial host is used by the phage for DNA replication as well as for synthesizing phage proteins. There are 11 genes associated with the f1 bacteriophage genome [27]. DNA replication is performed

by the cooperation between three phage genes and the host system. Three phage gene products are used in the assembly of phage particles. There is a single phage gene encoding a major capsid protein and four genes encoding minor capsid proteins. Expression of phage proteins is performed by the bacterial host system. The assembly of the filamentous phage is a membrane-associated event requiring the five capsid proteins, the three assembly proteins, adenosine triphosphate (ATP), a proton motive force, and at least one bacterial protein, thioredoxin [28].

Specific peptides and proteins can be displayed on the phage particles as a result of the fusion of specific proteins to the periplasmic portion of the capsid proteins before bacteriophage assembly. Most phage display libraries of peptides and antibodies have been made using the pIII minor capsid protein as the fusion vector, resulting in a chimeric pIII fusion protein. It has the disadvantage that a maximum of five expressed antibody molecules can be displayed per phage host [29]. The main advantage is that pIII fusion proteins containing large inserts can be packaged into the phage. The gene encoding the displayed peptide is inserted between the signal sequence and the beginning of the first domain of the pIII capsid protein. The insertion of the molecule here places the displayed protein at a single polar end of the phage particle. The pVIII major coat protein can also be used for peptide expression, but it is limited in that only peptides of six or eight amino acids in length can be displayed. The advantage is that many copies of the peptide (up to 2500 copies) are displayed along the entire surface of the bacteriophage [30].

Peptides can also be displayed on smaller filamentous particles referred to as phagemids. The phagemid genome contains the phage origin of replication and an independent gene encoding antibiotic resistance. Chimeric genes encoding peptide-phage protein fusions can be placed under the control of a specific promoter in these phagemid genomes. Superinfection of the bacterial host with a filamentous helper phage activates the phage origin of replication. Subsequent transcription and translation provide all phage-derived proteins and enzymes required for phagemid replication. Using a helper phage containing a defective packaging signal allows most phage particles produced to contain the phagemid ss-DNA. When compared with phage vectors, higher yields of phagemid vectors are obtained by plasmid preparation, and larger inserts are more readily maintained [27]. Bacterial transformation efficiency is higher with phagemid than phage vector, and this may result in greater library diversity [31, 32]. Generally, one to three copies of the pIII fusion protein will be expressed by each phage, whereas up to 200 copies of the pVIII fusion peptides can be expressed on the phage capsid [29]. The main disadvantage of using a phagemid vector is the high background during selection due to the presence of wild-type phages, which have been shown to exhibit nonspecific binding during affinity selection. As a result, the conventional M13K07 helper phage has been modified. The modified helper phage either requires that the host bacterial strain provides the pIII proteins [33] or possesses a trypsin cleavage site within the pIII gene that is absent in the fusion protein [34]. Trypsin treatment of phage after selection can eliminate most phages that do not carry the recombinant coat protein displaying the peptide of interest. Because of the need for a helper phage to rescue phagemid vectors, they are less efficient than phage vectors, which allow higher display levels and make panning more efficient, with more binders being isolated in fewer rounds [35]. The lower valency of phagemid vectors allows the selection of binders with higher affinity. Recombinant antibodies from phagemid vectors

generally have affinities that are 5–10 times higher than those from phage vectors, on average [35]. The higher avidity associated with phage display may result in the selection of lower affinity clones.

***Bacterial Display.*** The display of recombinant peptides on the surface of bacteria such as *E. coli* was first described more than a decade ago. Several peptides have been successfully displayed on bacterial surfaces [36], including antigenic determinants, heterologous enzymes, antibody fragments, and peptide libraries. Numerous reviews detail the various classes of proteins that have been displayed on both Gram-positive and Gram-negative bacteria [24, 29, 37].

The major obstacle for export of foreign peptides to the surface of a Gram-negative bacterial cell is the presence of an extensive network of macromolecules found within the bilayered cell envelope. It is necessary to select a leader sequence that translocates the peptide through the membrane, as well as an anchor protein that displays the rAb without disrupting the membrane integrity. Initially, short gene fragments were inserted into genes for the *E. coli* outer membrane proteins: the maltoporin LamB [38]; outer membrane protein OmpA [39]; and the phosphate-inducible porin PhoE [40], with the gene products being displayed on the surface of the recombinant bacteria. Lipoproteins [41, 42], fimbriae [43] and flagellar proteins [44], as well as dedicated systems with coupled translocation and surface anchoring have also been used for surface display of peptides [45]. Most current studies of bacterial display of antibodies involve the lipoprotein–porin (Lpp–OmpA') fusion system, with expression of recombinant proteins fused to the sorting signal that directs their incorporation on the cell surface [46]. The expressed antibodies retain their binding ability and approximately 60,000 scFv molecules can be displayed per cell [47]. Selection of good binders can be done by fluorescence-assisted cell sorting (FACS) flow cytometry, using fluorescently labeled antigens, thus avoiding the immobilization of the antigen, elution of bound phages, and reinfection of bacteria with eluted phages.

Expression of scFv on the surfaces of Gram-positive bacteria has been investigated. ScFv antibodies have been displayed on the surfaces of *Staphylococcus xylosus* and *Staphylococcus carnosus* through the fusion to the C-terminus of protein A [48]. Recently, spores of *Bacillus subtilis*, a dormant form of living cells known for their resistance to harsh environments, have been used for display of recombinant antibodies. ScFv antibodies have been fused to CotB, a surface-exposed coat component of the *B. subtilis* spore coat [49]. Spores represent a unique opportunity for surface display because of the potential of providing a durable supporting matrix similar to that of chemical polymer beads. There are three advantages of using Gram-positive over Gram-negative bacteria for surface display. First, larger molecules may be displayed on Gram-positive than Gram-negative bacteria. Second, the single-layered cell wall of Gram-positive bacteria may have better efficiency of fusion protein secretion and folding than the double-layered Gram-negative bacterial cell surface. Third, the rigid Gram-positive cell wall may allow the cells to withstand harsh conditions during affinity selection [7].

***Yeast Display.*** The eukaryotic system of yeast offers an advantage of posttranslational modification and processing of mammalian proteins and is,

therefore, better suited for expression and secretion of human-derived antibody fragments than a host such as *E. coli*. Yeasts displaying antibodies share many of the same advantages as bacterial display. Yeast have a rigid, thick cell wall that enables the stable maintenance of surface-displayed proteins [50]. As in bacterial display, the large size of yeast cells allows for efficient selection using FACS flow cytometry [51]. Several systems can be used for expression of rAb on yeast cell surfaces. Yeast display primarily uses the α-agglutinin yeast adhesion receptor to display recombinant proteins on the surface of *S. cerevisiae* [52]. The α-agglutinin receptor acts as an adhesion molecule to stabilize cell–cell interactions and facilitate fusion between yeast cells of opposite mating types. The receptor consists of two proteins, Aga1 and Aga2. The Aga1 protein is secreted from the cell and becomes covalently attached to the extracellular matrix of the yeast cell wall. The rAb–Aga2 fusion proteins bind to Aga1 after secretion and remain attached to the cell via Aga1, displaying the recombinant antibody on the yeast cell surface [53, 54]. The Aga2–fusion rAb can be displayed on the cell surface at 10,000 to 100,000 copies/cell [53].

Glycosylated-inositol (GPI)-anchored proteins have also been used to anchor foreign peptides for cell-surface display in *S. cerevisiae* [55] and *Hansenula polymorpha* [56]. Four different *gpi* genes and their encoded proteins were characterized for the purpose of displaying fusion protein on the cell surface. The *FLO1* gene encodes the Flo1p protein, a lectin-like cell-wall protein thought to form stem-like structures that cover the distance of the cell wall. This protein has also been used to express peptides on the yeast surface [57].

In bacterial or yeast surface display, the phenotype and genotype are linked through the display of the protein on the surface of a cell, in which the DNA can also be propagated and expressed. No subcloning step is required after each selection. Yeast display is particularly beneficial as yeast cells are generally more robust and can express large quantities of heterologous proteins with appropriate post-translational modifications [52]. However, the display of multiple antibody fragments on bacterial or yeast surfaces contributes to an avidity effect, potentially contributing to the selection of rAbs with lower binding affinities [58].

***Ribosome Display.*** This cell-free display system uses the same principle as the phage-display systems in bacteria. The phenotype and genotype of a peptide are linked together and, therefore, can be simultaneously selected based on the function of the peptide. However, the linkage is a physical one in that the genetic material is covalently linked to its encoded product in the formation of an antibody–ribosome–mRNA (ARM) complex through *in vitro* transcription/translation [59].

Ribosome display is an *in vitro* method, with all steps being nucleic acid-based. This allows the assembly of a large-size library with up to $10^{15}$ members. Mutations can be incorporated in each cycle of selection using PCR-based mutagenesis, thereby allowing for the introduction of diversification during selection. Antibodies with equilibrium dissociation constants improved from the nanomolar to picomolar range using ribosome display have been reported [60, 61].

The key advantage of ribosome display over phage display systems is the freedom from any bacterial transformation step, which typically limits the size of antibody libraries. The library size in phage, bacterial, or yeast display is determined by the

transformation efficiency. Generally, $10^{10}$ to $10^{11}$ members represent the upper possible limit of the size of an *E. coli* library. Due to the generally lower transformation efficiency in yeasts, large yeast-display libraries are composed of up to $10^9$ clones [53]. Another advantage of ribosome display systems is that the expression of the target peptides is not limited by the host. With *in vivo* display systems, proteins that are toxic or detrimental to cell growth will inhibit or kill the host cell, and this can potentially limit the library size for screening. Furthermore, bacteria are prone to introduce natural mutations into foreign DNA to evade selection pressure. Mutations and/or recombinations within the plasmid are frequently found that may provide benefits to the cells. Undesirable selection pressure on individuals can easily be avoided in a cell-free system. Ribosome display is especially useful for construction of a naïve library, which provides a better library complexity and diversity than its phage display counterpart [7]. The use of a ribosome display cannot only select and enrich high-affinity binders from a library, but also it provides *in vitro* hypermutation for the maturation of selected clones to improve binding characteristics [61, 62].

### 2.5 Soluble Antibody Production

Numerous systems are available for the expression of soluble recombinant antibodies. Bacteria, yeasts, plants [63], insects and mammalian cell lines [64], and cloned transgenic animals have all been used for rAb expression [65]. Bacteria, in particular *E. coli*, are favored for expression of small, nonglycosylated Fab, scFv, and $V_H$ fragments and diabodies. ScFv fragments are generally expressed to a higher level over Fab in bacteria, although yields may vary for each rAb fragment [2]. Bacterial expression of rAbs is typically achieved by fusion to N-terminal signal peptides, which targets the protein to the periplasmic space of *E. coli*. This periplasmic expression may render a large fraction of the produced rAb insoluble. Soluble bacterial antibody production can also be directed into the culture medium. The risk is that the production of inclusion bodies often results in insoluble protein aggregates. This necessitates laborious renaturation steps and subsequent *in vitro* folding, resulting in a low final yield [18]. Secretion of rAbs can be associated with cell lysis and subsequent product loss. Also, if the rAb being produced is toxic to the bacteria, production levels will be drastically deceased. Because bacteria cannot carry out eukaryotic posttranslational modification, their use is not suitable when glycosylation of the rAb is required.

***Purification of Recombinant Antibodies.*** Purification of bacterial expressed rAb is assisted by the presence of fusion polypeptides such as c-Myc, poly-His, and the FLAG epitope, which have been added to allow for affinity purification. Subsequent cleavage of the fusion protein yields the rAb of interest.

## 3 IMMUNODETECTION OF PATHOGENS

Diagnostic tests for viruses, bacteria, fungi, and their spores are often based on culture-dependent methods. While these methods are very sensitive and generally simple to perform, they can be time consuming, requiring several days, and possibly

weeks, to complete. If microorganisms are nonculturable, microscopy can be used for identification, as is the case for numerous protozoan parasites. Microscopy requires specific training and may be subject to user bias and error. Also, it may be hard to differentiate between microorganisms at a species level. Epifluorescent microscopy, with fluorochrome-conjugated antibodies specific to the target organism, can make this method much easier to perform. Molecular-based methods such as PCR can be used for quick identification of specific target pathogens, but can be expensive and, like microscopy, requires specific training for the operator. However, PCR cannot distinguish between those detected organisms that are viable and infective from those that are dead. Isolation of DNA from some matrices can be problematic because of the presence of inhibitors that can adversely affect PCR [66].

An antibody to be used in immunodiagnostics should ideally have high binding specificity and affinity to the target antigen of interest. As such, many commercially available immunodiagnostic antibodies are produced as monoclonal antibodies. As the production of rAbs becomes more routine, their availability is sure to increase. Table 1 lists some developed rAbs that could be used in immunodiagnostics. One major advantage of recombinant antibodies is that they are more easily selected for when the target pathogen may be harmful to an animal being immunized for the production of monoclonal antibodies.

### 3.1 Formats of Detection

Successful immunodiagnostic applications are based on the availability of antibodies with high affinity and specificity. Initially, polyclonal antibodies from hyperimmunized animals or monoclonal antibodies from hybridoma cell lines were used. More recently, rAbs are being constructed that possess affinities rivaling those of monoclonal antibodies. Therefore, it is anticipated that rAbs will begin to be deployed in greater frequency in immunodiagnostics in the short to medium term.

The efficiency of immunoassays relies on four components: (1) the target antigen to be detected; (2) the antibody used for detection; (3) the method to separate bound antigen and antibody complexes form unbound reactants (if a heterogenous mixture is used); and (4) the detection method. At the most fundamental level, the efficacy of any given immunoassay is dependent on two major factors: the efficiency of antigen–antibody complex formation and the ability to detect these complexes. It is desirable that the antibodies used have high affinity for the target antigen with low cross-reactivity to minimize false positives. Greater sensitivity of immunoassays is generally associated with higher affinity antibodies.

The use of polyclonal antibodies in sera has inherent disadvantages. Laboratory animals are required, and results can vary drastically in their ability to respond to different antigens, resulting in significant variation among different lots of antibodies. The antibody response to a given antigen tends to be broad. This covers both specific and cross-specific epitopes, causing problems such as high-background or unwanted reactivity for contaminants in the antigen preparation. Polyclonal immunoassays can be highly sensitive but may not be very specific [6, 100].

There is an inherent disadvantage when using highly specific antibodies. Antigenic drift or shift can make using monoclonal antibodies difficult. The evolution

**TABLE 1. A List of Some Pathogens Detected by Recombinant Antibodies**

| Pathogen | rAb Type | Source Library | Reference |
|---|---|---|---|
| **Bacterial** | | | |
| *Bacillus subtilis* | scFv | Human Naïve | [67] |
| *Burkholderia pseudomallei* exotoxin | scFv | Mouse Immune | [68, 69] |
| *Chlamydia psittaci* | scFv | Human Naïve | [70] |
| *Chlamydia trachomatis* | scFv | Human Naïve | [70] |
| Cyanobacterial Hepatoxin Microcystins | scFv | Human Semi-Synthetic | [71] |
| *Clostridium difficile* Toxin B | scFv | Mouse Immune | [72] |
| *Escherichia coli* O157:H7 | scFv | Human Naïve | [73] |
| *Moraxella catarrhalis* | scFv | Human Semi-Synthetic | [74] |
| *Streptococcus mutans* | scFv | $V_H$ of Mouse Immune, $V_L$ of Human Naïve | [75] |
| *Streptococcus suis* | scFv | Human Semi-Synthetic | [76] |
| **Fungal** | | | |
| *Candida albicans* | scFv | Human Naïve | [77] |
| *Fusarium* Mycotoxin Deoxynivalenol | scFv | Mouse Immune | [78] |
| *Phytophthora infestans* | scFv | Naïve | [79] |
| **Protozoan** | | | |
| *Cryptosporidium parvum* | Fab | Mouse Immune | [80] |
| *Entamoeba histolytica* | Fab | Human Immune | [81] |
| *Plasmodium chabaudi* | scFv | Mouse Immune | [82] |
| *Plasmodium falciparum* | scFv | Mouse Immune | [81] |
| **Viral** | | | |
| Beet Necrotic Yellow Vein Virus | scFv | Unknown | [83] |
| Bluetongue Virus | scFv | Mouse Immune | [84] |
| Cucumber Mosaic Virus | scFv | Mouse Immune | [85] |
| Ebola Virus | Fab | Human Immune | [86] |
| Foot and Mouth | scFv | Mouse Immune | [87] |
| Grapevine Virus B | scFv | Human Semi-synthetic | [88] |
| Herpes Simplex Virus 1 | Fab | Human Immune | [89] |
| Herpes Simplex Virus 2 | Fab | Human Immune | [90] |
| Human cytomegalovirus | Fab | Human Immune | [91] |
| Infectious Bursal Disease Virus | scFv | Chicken Immune | [92] |
| Influenza A | Fab | Human Naïve | [93] |
| Severe Acute Respiratory Syndrome Coronavirus (SARS-CoA) | scFv | Mouse Semi-synthetic | [94] |
| Vaccinia Virus | Fab | Human Immune | [95] |
| Venezuelan Equine Encephalomyelitis Virus | scFv | Mouse Immune | [96, 97] |
| **Other** | | | |
| Prion – $PrP^{Sc}$ | scFv | Mouse Immune | [98] |
| *Androctonus australis hector* (Scorpion) Venom Toxin Aah1 | scFv | Mouse Immune | [99] |

*Abbreviation*: $PrP^{Sc}$ = The modified form of $PrP^{c}$ (a normal cellular protein thought to be involved in synaptic function), which may cause disease; i.e., the prion is known as $PrP^{Sc}$ (for scrapie).

of an antigen by either genetic modification or active selection can produce targets capable of evading detection systems. Also monoclonal antibodies rarely possess affinities that match the avidities of good polyclonal preparations [100].

***Immunoassays.*** A variety of methods are available to detect antibody–antigen interactions. Some are well established, whereas others have only been developed recently.

A common way to detect antibody–antigen binding in an immunoassay is the enzyme-linked immunosorbant assay (ELISA). ELISA is based on the target antigen being captured onto a plastic multiwell or tube by a capture antibody previously bound to the solid matrix. Bound antigen is then detected using a secondary antibody. The detector antibody can be directly labeled with a signal-generating molecule, or it can be detected with another antibody that is labeled with an enzyme. These enzymes catalyze the chemical reaction with a substrate that results in a colorimetric change. The color intensity can be measured by a spectrophotometer. The ELISA immunoassay is commonly used by clinical laboratories, can be adapted for high-throughput screening, and is relatively inexpensive. Electrochemiluminescence (ECL) is similar to ELISA except that the detector antibody is directly labeled with a chemiluminescence label. Time-resolved fluorescence (TRF) immunoassay is a sandwich-type assay similar to those used for ELISA except that the detector antibodies are directly labeled with lanthanide chelates. TRF exploits the differential fluorescence life span of lanthanide chelate labels compared with background fluorescence [100, 101].

Lateral flow immunoassays are commonly recognized as dipstick or handheld immunochromatographic assays. Detection of target antigens depends on formation of an antigen–antibody complex that migrates along the matrix by capillary action. Assessment of result is qualitative and subject to operator bias. The sensitivity is generally 1-log worse than that of ELISA [101]. Novel rapid immunoassays have been recently developed for the detection of HIV-1 antibodies in clinical samples [102].

Flow cytometry is based on the use of a fluorescent probe being bound by the target antigen, which is streamed past a laser beam where the probe is excited. A detector analyzes the fluorescent properties of the antigen–antibody complex as it passes through the laser beam. This immunoassay can be used in multiplex formats with little or no loss of sensitivity. However, flow cytometry lacks the sensitivity of ELISA and associated immunoassays, and the system requires training and expertise to operate [100]. *Cyrptosporidium parvum* oocysts have been detected in various matrices by flow cytometry and, recently, in soil samples by using antibodies conjugated to fluorescein isothiocyanate [103].

One of the more popular immunassays for pathogen detection involved the use of antibodies in combination with epifluorescent microscopy, usually referred to as an immunofluorescence assay (IFA). Monoclonal and polyclonal antibodies are conjugated to fluorochromes, so that the pathogen can be more easily detected. However, a lack of specificity sometimes results in these IFA assays, indicating the presence of more than one type of pathogenic organism, especially in species that are closely related [104]. Antibodies developed against *Eimeria acervulina* have been found to cross-react with several different *Eimeria* spp., *Neospora caninum*, *Toxoplasma gindii*, and *C. parvum* [104].

## 4 IMMUNOTHERAPY

Immunotherapy is a form of medical treatment intended to stimulate or restore the ability of the immune system to fight infection and disease. It can be either nonspecific or specific. Cytokines, growth factors, and leukocytes can enhance host defencse against a variety of pathogens and are examples of nonspecific immunotherapies. In contrast, the administration of antibodies is a form of specific immunotherapy that, similar to antimicrobial therapy, is directed specifically at the pathogen or toxin [105]. Immunotherapy can also be either replacement or reconstitution therapy, which is intended to correct the underlying immunological defects that predispose people to infection, or augmentative therapy, which is intended to enhance the body's immune function against the pathogen [105].

The first use of antibodies for treatment of ailments was performed in 1890 by Behring and Kitasato. They demonstrated that immunity could be transferred to naïve animals by serum of animals that had been challenged with nonlethal doses of a crude toxin preparation. As all original antibody preparations were derived from the serum of immunized animals or immune human donors, the treatment was called serum therapy. Even though the therapy was effective, the administration of large amounts of foreign animal proteins was often associated with side effects that ranged from immediate hypersensitivity reactions to serum sickness [106]. The use of serum therapy was abandoned due to several drawbacks, such as the occurrence of serum sickness, the risk of disease transmission, and lot-to-lot variations of different serum preparations. In the mid-1930s, the introduction of antibiotics replaced the use of antibodies for therapy against most infectious pathogens. Antibodies retained a niche as a treatment for venoms, toxins, and certain viral infections [106]. With the establishment of hybridoma monoclonal antibody production, there was hope for the rapid development of many therapeutic applications, especially in field of oncology [106].

A renewed interest in the production of antibodies for therapeutic use has occurred as a result of numerous driving forces: (1) the increased incidence of pathogen resistance to antibiotics; (2) the lack of effective chemotherapeutic agents; (3) the toxicity of available drugs; (4) the emergence of new microorganisms and viral diseases; (5) the epidemiological evidence of an important role for the humoral immune system in host control of infection; (6) the increased incidence of chronic disease in individuals with various antibody deficiencies; and (7) the use of antibodies is currently the only means to provide immediate immunological protection against biological agents [105, 107–110].

The action of antibodies in immunotherapy is directed against pathogen stages that are susceptible to antibodies. By blocking the pathogen's interaction with the host cell, antibodies can prevent pathogen attachment and/or invasion [109]. Direct lysis of pathogens can occur, by interfering with microbial gene expression and growth, and by enhancing and activating immune cells or a combination thereof [111]. In bacterial disease, antibodies neutralize toxins, facilitate opsonization, and with complement, promote bacterial lysis [112]. In viral disease, antibodies are the best option for therapy and prophylaxis [113]. Antibodies block viral entry into uninfected cells, promote antibody-directed cell-mediated cytotoxicity by natural killer cells, and neutralize virus alone or with the participant of complement [112]. That a microbe inside a host cell is separated from serum antibody has led to the

belief that serum antibody cannot be effective against intracellular pathogens. However, at some point in the infectious cycle, most intracellular pathogens reside in the extracellular space, where they are susceptible to antibody action [114]. A hallmark of protozoan parasites is their ability to evade the host's immune system. In many cases, this is accomplished by stage-specific expression of unique antigens, causing immune responses to stages that are subsequently lost during development [115, 116].

From the point of view of antibody production for immunotherapy, careful target selection is of the utmost importance. Although it is important that the antibodies are pathogen specific, monospecific antibodies, such as monoclonal antibodies, may not be a good choice because they are too specific. Narrow specificity is advantageous because it prevents development of resistance among nontargeted microorganisms and avoids disturbance of the normal microflora. However, narrow specificity becomes a disadvantage in cases of mixed infections that cannot be treated with a single antibody preparation. It also decreases the potential market for the drug [113]. Moreover, the high specificity is at a disadvantage with the emergence of pathogen variants that lack the determinant that the antibody recognizes, such as viral escape mutants [106]. The use of cocktails of antibodies that are specific for several antigens could obviate this concern. However, this approach would also have the drawback of increasing the cost of production and the complexity of regulatory issues involving efficacy and safety [106]. Therapeutically relevant targets can also be difficult to identifiy.

Most antibodies currently being used for immunotherapy are monoclonal immunoglobulins (i.e., IgG and IgM). Their multivalency and high specificity and affinity make them attractive for therapies. Brekke and Løset [117] and Casadevall [118] have reviewed the various monoclonal antibodies currently available for immunotherapies as well as those undergoing clinical trials. Most antibodies slated for therapeutic and clinical applications are being developed for use in cancer treatment [118]. However, it has been demonstrated that in cancer immunotherapy, the fraction of injected monoclonal antibody that effectively reaches the tumor is usually less than 1% of the injected dose, which is caused by poor tumor penetration and limited target specificity. By reducing the antibody size and increasing its affinity for the target, it should be easier to obtain better targeting molecules [119]. Recombinant antibodies are smaller and can be matured for increased affinity. Therefore, it is not surprising that their use for therapy against viruses and intracellular pathogens is becoming more attractive. Currently, relatively few rAbs are being used for immunotherapy, as compared with hybridoma monoclonals. Table 2 lists some rAbs currently being studied for immunotherapeutic activities. Because of the need for therapeutic antibodies to be specific, it is feasible that any antibody used for therapy could also be used for immunodiagnostic purposes.

### 4.1 Passive Immunotherapy

Passive immunotherapy involves administering preformed antibodies from external sources to the recipient patient, as opposed to therapies like vaccines, which encourage the body to make its own antibodies. Numerous rAbs are specific to various pathogens (see Table 2). Pathogens that affect humans, plants, and commercial/domestic animals are all being looked at for their susceptibility to antibodies. For

**TABLE 2. A List of Some Recombinant Antibodies that May Be Theoretically Effective in Immunotherapy**

| Antibody Target | rAb Type | Therapeutic Status | Reference |
|---|---|---|---|
| **Bacterial** | | | |
| *Bacillus anthracis* Toxin | Fab | Rat; *In Vivo* | [120] |
| *Neisseria meningitidis* | scFv | Mouse; *In Vivo* | [121] |
| *Staphlococcus aureus* – Methicillin Resistant | Aurograb | Human; Clincal Trial | NeuTec Pharma plc |
| **Fungal** | | | |
| *Candida albicans* | scFv | Rat; *In Vivo* | [122, 123] |
| *Candida* spp. | Mycograb | Human; Clincal Trial | NeuTec Pharma plc; [124] |
| *Cryptococcus neoformas* | scFv | Mouse; *In Vivo* | [108] |
| *Fusarium oxysporum* | scFv | In Planta | [125] |
| **Protozoan** | | | |
| *Plasmdodium falciparum* | Recombinant Peptide | Cell Culture; *In Vitro* | [126] |
| *Plasmodium yoelii* | scFv | Mouse; *In Vivo* | [127] |
| *Plasmodium berghei* | scFv | Mosquito; *In Vivo* | [128] |
| **Viral** | | | |
| Cucumber Mosaic Virus | scFv | In Planta | [129] |
| Foot and Mouth Disease | Fab | PRNT; *In Vitro* | [130] |
| Herpes Simplex Virus 1,2 | Fab | Cell Culture; *In Vivo* | [131] |
| HIV-1 Reverse Transcriptase | scFv | Cell Culture; *In Vitro* | [132] |
| HIV-1 Vif Protein | scFv | Cell Culture; *In Vitro* | [133] |
| Japanese Encephalitis Virus | Fab | PRNT; *In Vitro* | [134] |
| Measles Virus | Fab | PRNT; *In Vitro* | [135] |
| Rabies Virus | scFv-Fc | Mouse; *In Vivo* | [136] |
| Rotavirus | Fab | Cell Culture; *In Vitro* | [137] |
| Varicella-zoster Virus | scFv | Cell Culture; *In Vitro* | [138] |
| **Other** | | | |
| Prion $PcP^{Sc}$ | scFv | Cell Culture; *In Vitro* | [139] |
| Prion $PcP^{Sc}$ | Fab | Cell Culture; *In Vitro* | [140] |

*Abbreviation*: PRNT = Plaque Reduction Neutralization Test. $PcP^{Sc}$ = The modified form of $PrP^{c}$ (a normal cellular protein thought to be involved in synaptic function), which may cause disease; i.e., the prion is known as $PrP^{Sc}$ (for scrapie). Aurograb = Commercially available recombinant antibody from NeuTec Pharma plc. Mycograb = Commercially available recombinant antibody from NeuTec Pharma plc.

passive immunotherapy to be effective, antibodies must be present at the site of infection long enough to interact with the appropriate infective stage of the organism's life cycle, bind to the appropriate antigen(s), and activate the appropriate effector mechanisms (if any) [109]. For intestinal pathogens, considerations for the effective therapy include dosage schedule, gastrointestinal transit time and residence of the antibody in the small intestine (the primary site of infection), gastrointestinal degradation of the antibody and ways of protecting the antibody activity (formulation), as well as the antigenic specificity and antigenic relatedness of strains or isolates of the pathogen [109]. Passive antibody therapy for infectious diseases is aided by the large antigenic differences between the microorganism and the host and can be optimized by identifying and targeting molecules essential for infectivity and by understanding mechanisms of antibody-mediated neutralization [106, 141]. In some gastrointestinal diseases caused by microorganisms, prophylactic uses of specific antibodies have proven to be much more effective than conventional supportive treatment, such as in human studies of enterotoxigenic *E. coli*-induced "travelers diarrhea" [142] and *Shigella* infections [143]. Unfortunately, rAbs have been found to possess a weaker neutralization effect on pathogens when compared with the corresponding monoclonal antibody [8].

Numerous variables can lead to a negative result from passive immunotherapy. Too little antibody can result in a lack of protective efficiency. Too much antibody can produce a prozone-like effect, whereby more antibody provides less protection than less antibody. The antibodies used in therapy can also be affected by affinity/avidity, specificity, isotype, idiotype, preparation, and route of administration. The genetic background and immune competency of the host can also affect treatment as can route of pathogen infection, genetic background, and inoculum size of the pathogen [114]. A dramatic example of the limitations of passive antibody transfer experiments is provided by the observation that transfer of either too little or too much antibody can result in no protection [114].

Passive antibody therapy and prophylaxis have a natural place in animal health because mammalian species such as pigs, horses, sheep, and cows do not transmit maternal immunity prenatally (antibodies do not cross the barrier of the placenta, as is the case in humans) but postnatally through colostral antibodies [113].

***Humanized Antibodies.*** For human immunotherapeutic applications, the optimal antibody should be of human origin. Immune libraries are perhaps the best way to produce neutralizing antibody responses to protective epitopes on infectious pathogens [26]. For therapeutic purposes, the efficacy of nonhuman-derived monoclonal antibodies is limited by the induction of anti-mouse immune response. Using hybridoma methods to immortalize human B-cells for the construction of human monoclonal antibodies would avoid these issues. However, the absence of a suitable fusion partner and other technical issues has made methods that rely on human B-cell immortalization problematic, forcing reliance on nonhuman organisms [6].

The development of human(ized) antibody molecules is mostly aimed at reduction of unwanted immunological properties in medical applications. Useful antibodies have been isolated from mice and murine-based antibody libraries. However, these antibodies are immunogenic in humans. To use these molecules as successful therapeutic agents, they must be redesigned to reduce their immunogenicity, thereby

reducing the human–anti-mouse (HAMA) immune response [144]. The antigenicity of therapeutic antibodies can be reduced by chimerization, in which nonhuman domains are replaced by human domains, though fusion of unmodified variable domains to the constant regions of the light and heavy chains of a human antibody [144]. Humanization involves not only fusion of the (usually murine) variable domains to human constant regions, but also the alteration of selected murine variable domain framework residues to more closely match the most related available human variable template sequence [18]. CDR grafting involves the substitution of nonhuman CDR from an antibody into the most closely related human antibody sequence available, so that only the CDR domains are nonhuman in origin [144]. Veneering involves changing only exposed residues so that they reflect the human consensus sequences, whereas those residues that are buried are left unmodified [144]. These strategies may often preserve the binding affinity of the murine antibody while significantly reducing immunogenicity [145].

Antibody-based therapies that use human or humanized antibodies have low toxicities and high specificities. The high specificity of antibodies is both an advantage and a disadvantage. The advantage of high specificity is that antibody-based therapies target only the pathogenic organism that causes disease and, therefore, should not alter the host flora or select for resistance among nontargeted microorganisms. However, high specificity also means more than one antibody preparation might be required to target microorganisms with high antigenic variation [106].

Human naïve or semisynthetic libraries, such as the Tomlinson or Griffin libraries, can be screened for pathogen specific rAbs [146]. This eliminates the need for antibody humanization, making them readily available for immunotherapeutic applications in humans.

***Delivery of Antibodies.*** Antibodies can be administered as human or animal plasma or serum, as pooled human immunoglobulin for intravenous (IVIG) or intramuscular (IG) use, as high-titer human IVIG or IG from immunized or convalescing donors, and as monoclonal antibodies [112]. Oral administration of large doses of antibodies resulting in high local concentrations can be given without the serious safety concerns related to systemic administration of foreign proteins (as would be the case for xenogeneic systemic administration). However, oral administration of antibodies is only feasible for treatment of certain types of mucosal infectious diseases, such as *C. parvum*-associated infections. There is a restriction to oral antibody delivery, specifically the possibility of gastrointestinal destruction of the antibody given the low gastric pH and the presence of gastric and intestinal proteases. As a result, duration of the antibody availability due to peristalsis is limited to the order of 1–2 hours [109]. These problems could be remedied by coadministration of antacids, enteric coatings (pH dependant release), bioerodable coating (delayed release, pH independent), or liposomal formulation (enzyme induced release). In addition to protective coatings, formulations must be able to disperse rapidly when the coating is shed [109]. Otherwise, antibodies intended for therapeutic applications need to be administered systemically [113] to provide a much longer duration of activity, on the order of days rather than hours [109]. An important potential advantage of antibody therapies is that they can be synergistic or additive when combined with conventional antimicrobial chemotherapy against bacterial and viral diseases [106].

As antibodies are natural products, they must be produced in cell lines or other live expression systems. This raises the theoretical concern that there could be contamination of antibody preparations by infectious agents such as prions or viruses. Although tight regulation and regulatory vigilance and surveillance can reduce this concern, the need for ongoing monitoring and testing for contamination contributes to the high cost of developing and administering antibody therapies [106].

### 4.2 Vaccine/Active Immunotherapy

The advent of recombinant DNA technologies has increased the available avenues for vaccine development. Although not related to rAbs, active immunotherapy is important for the prevention of illness due to infection by pathogenic microorganisms. Active immunity is the response generated by the immune system after exposure to a vaccine or microorganism. Both antibodies and cell-mediated immunity can be generated by active immunization. Because antibody therapy can be protective against an infectious disease, a vaccine that elicits similar antibodies should be protective against the relevant pathogen. Ideal vaccines should be safe, elicit the appropriate immune response, and provide long-term protection at low doses [147]. Vaccines against infectious organisms have been traditionally developed by administration of whole live attenuated or inactivated microorganisms [148]. Although live attenuated viral vaccines can elicit responses from cytotoxic T lymphocytes (CTLs) and from the humoral system, inactivated viruses and recombinant proteins are generally not potent inducers of CTLs [147]. However, subunit and recombinant vaccines hold promise for agents of bioterror because they deliver defined antigens, perhaps for multiple agents, and have the potential to reduce the immunogenicity of whole cell preparations [110]. Table 3 indicates a few infectious pathogens whose disease can be prevented, or minimized, by the use of recombinant antigens, recombinant viruses, or DNA vaccines.

***Recombinant Antigens.*** Recombinant antigens are synthesized to represent the most immunogenic antigens from a particular pathogen. However, when purified

**TABLE 3. Representative List of Pathogens Being Treated by Active Immunotherapy**

| Antibody Target | Type of Vaccine | Therapeutic Status | Reference |
|---|---|---|---|
| *Chlamydia pneumoniae* | DNA/Recomb Virus | Mouse; *In Vivo* | [149] |
| *Cryptosporidium parvum* | Recomb Antigen | Bovine; *In Vivo* | [150] |
| *Entamoeba histolytica* | Recomb Antigen | Mouse; *In Vivo* | [151] |
| HIV-1 Envelop Protein | Recomb Virus | Mouse: *In Vivo* | [152] |
| *Listeria monocytogenes* | DNA Vaccine | Mouse; *In Vivo* | [153] |
| *Mycobacterium avium* | Recomb Antigen | Lamb; *In Vivo* | [154] |
| *Plasmdodium falciparum* | Recomb Virus | Human; *In Vivo* | [155] |
| *Vibrio anguillarum* | Recomb Peptide | Piscine; *In Vivo* | [156] |
| Virulent infectious Bursal Disease Virus | DNA Vaccine | Chicken; *In Vivo* | [157] |

recombinant antigen is administered using traditional immunization protocols, the levels of cellular immunity induced are generally low and not capable of eliciting complete protection against diseases caused by intracellular microbes [148]. Most injected proteins are not efficiently presented by major histocompatibility complex (MHC) class I, making this a major limitation for diseases in which cellular immune responses play a key role [147]. Recombinant antigens that have been used as vaccines are considered safe, but they do have limitations in their use. For example, a recombinant protein can have posttranslational modifications from the pro-tein present in the natural infection and may not elicit optimal responses. Immunization with recombinant antigens requires that parasite epitopes involved in invasion and attachment are known and available as well as amenable to re-combinant DNA technology, and they can be produced in a pure form. Molecular techniques for producing recombinant antigens that elicit neutralizing antibodies have been developed that may serve as a means to provide vaccination against pathogens.

***Recombinant Viruses.*** This type of vaccination is based on the use of recombinant viruses encoding specific antigens as immunization tools. Recombinant viruses can express the foreign antigens directly inside cells of the host organism, as would happen in natural infection. Antigens so expressed are made available to the intracellular antigen-processing machinery. Displayed antigens can be bound by MHC molecules, thus favoring their presentation to T lymphocytes [148]. The poxviruses, adenoviruses, and influenza viruses are three of the most attractive and efficient vectors that can be used for vaccination purposes [148]. However, re-exposure of the immune system to the same recombinant antigens expressed by the viral vectors may not recall the memory T-cell responses induced during the primary exposure, as subsequent administrations of the same virus were cleared faster from the organism due to immune responses induced against the virus itself [158]. In cases where immunization with one viral vector does not elicit sufficient levels of immunity, researchers have successfully elicited immunization by the sequential administration of a different recombinant virus expressing the same antigen [158].

***DNA Vaccines.*** DNA vaccines were introduced less than a decade ago, but they have already been applied to a wide range of infectious and malignant diseases, including those that are parasitic and difficult to control [159]. DNA vaccines involve the use of plasmids containing viral, parasitic, or bacterial genes that are expressing the protective antigen in mammalian cells after their introduction [160]. Uptake and expression of the DNA has been demonstrated in several species, for several parasitic and bacterial infections (reviewed in Ref. 147). DNA vaccination may also be useful to define key genes that play a role as antigens for diseases that are not well understood. Vaccination with DNA have been administered via (1) direct injection of naked DNA, (2) administration of DNA complexed with liposome formulations or other compounds that target specific receptors, or (3) compact DNA associated with gold particle bombardment [147].

The main safety issues for DNA vaccines include integration of the plasmid into the host genomic DNA, production of pathogenic anti-DNA antibodies to the plasmid that could induce or exacerbate diseases such as systemic lupus erythematosus, or immunologic tolerance to the antigen being expressed [147]. As a result,

to produce vaccines safely and efficiently, the use of plasmids that do not replicate in mammalian cells is desirable for gene expression. The risk of chromosomal integration and subsequent activation of an oncogene or disruption of a tumor suppressor gene or other regulatory gene is another safety issue [147].

## 5 CONCLUSIONS

With increased concern about the pathogen status in food and water, the application of antibodies in the detection of these pathogens also continues to increase. Design of new rAbs that are more specific and have a higher affinity for their target is becoming more commonplace. These new rAbs are making immunoassays more sensitive. When combined with advances in our understanding of the pathogenic nature of infectious diseases, rAbs are making immunotherapy more relevant in the prevention and treatment of diseases caused by pathogens. With the increased number of available antibody libraries and ease with which rAbs can be obtained, a rapid process exists for generation of specific, high-affinity rAbs against virtually any target. As the parameters for improving *in vivo* efficacy of rAbs becomes more understood, and bacterial fermentation provides a cost effective route to scale up production of many formats of engineered rAbs, their presence in pathogen diagnostic and therapies will be enhanced.

## ACKNOWLEDGMENTS

We would like to thank the Canadian Water Network (CWN), the Natural Sciences and Engineering Research Council (NSERC) of Canada, Ontario Ministry of Agriculture and Food (OMAF), and Ontario Ministry of Environment for their support of our research on recombinant antibodies and *Cryptosporidium parvum*. We would also like to acknowledge GAP EnviroMicrobial Services and the NIH AIDS Research & Reference Reagent Program for their cooperation in our studies.

## REFERENCES

1. Kohler G, Milstein C (1975). Continuous cultures of fused cells secreting antibody of predefined specificity. *Nature.* 256:495–497.
2. Holliger P, Hudson P J (2005). Engineered antibody fragments and the rise of single domains. *Nature Biotechnol.* 23:1126–1136.
3. Goldsby R A, Kindt T J, Osborne B A, et al. (2003). *Immunology.* W.H. Freeman, New York.
4. Stanfield R L, Takimoto-Kamimura M, Rini J M, et al. (1993). Major antigen-induced domain rearrangements in an antibody. *Structure.* 1:83–93.
5. Hoogenboom H R, Winter G (1992). By-passing immunisation. Human antibodies from synthetic repertoires of germline VH gene segments rearranged in vitro. *J. Mol. Biol.* 227:381–388.
6. Berry J D (2005). Rational monoclonal antibody development to emerging pathogens, biothreat agents and agents of foreign animal disease: The antigen scale. *Vet. J.* 170:193–211.

7. Yau K Y, Lee H, Hall J C (2003). Emerging trends in the synthesis and improvement of hapten-specific recombinant antibodies. *Biotechnol. Adv.* 21:599–637.
8. Thullier P, Lafaye P, Megret F, et al. (1999). A recombinant Fab neutralizes dengue virus in vitro. *J. Biotechnol.* 69:183–190.
9. Park K J, Park D W, Kim C H, et al. (2005). Development and characterization of a recombinant chicken single-chain Fv antibody detecting *Eimeria acervulina* sporozoite antigen. *Biotechnol. Lett.* 27:289–295.
10. Ward E S, Gussow D, Griffiths A D, et al. (1989). Binding activities of a repertoire of single immunoglobulin variable domains secreted from *Escherichia coli*. *Nature.* 341:544–546.
11. van der Linden R H, Frenken L G, de Geus B, et al. (1999). Comparison of physical chemical properties of llama VHH antibody fragments and mouse monoclonal antibodies. *Biochim. Biophys. Acta.* 1431:37–46.
12. Nuttall S D, Humberstone K S, Krishnan U V, et al. (2004). Selection and affinity maturation of IgNAR variable domains targeting *Plasmodium falciparum* AMA1. *Proteins.* 55:187–197.
13. Emanuel P A, Dang J, Gebhardt J S, et al. (2000). Recombinant antibodies: a new reagent for biological agent detection. *Biosens. Bioelectron.* 14:751–759.
14. Crowe J E, Jr, Cheung P Y, Wallace E F, et al. (1994). Isolation and characterization of a chimpanzee monoclonal antibody to the G glycoprotein of human respiratory syncytial virus. *Clin. Diagn. Lab. Immunol.* 1:701–706.
15. Lamarre A, Talbot P J (1995). Protection from lethal coronavirus infection by immunoglobulin fragments. *J. Immunol.* 154:3975–3984.
16. Whitlow M, Bell B A, Feng S L, et al. (1993). An improved linker for single-chain Fv with reduced aggregation and enhanced proteolytic stability. *Protein Eng.* 6: 989–995.
17. Lamarre A, Yu M W, Chagnon F, et al. (1997). A recombinant single chain antibody neutralizes coronavirus infectivity but only slightly delays lethal infection of mice. *Eur. J. Immunol.* 27:3447–3455.
18. Joosten V, Lokman C, Van Den Hondel C A, et al. (2003). The production of antibody fragments and antibody fusion proteins by yeasts and filamentous fungi. *Microb. Cell Fact.* 2:1–15.
19. Greenberg A S, Avila D, Hughes M, et al. (1995). A new antigen receptor gene family that undergoes rearrangement and extensive somatic diversification in sharks. *Nature.* 374:168–173.
20. Nuttall S D, Krishnan U V, Doughty L, et al. (2003). Isolation and characterization of an IgNAR variable domain specific for the human mitochondrial translocase receptor Tom70. *Eur. J. Biochem.* 270:3543–3554.
21. Hudson P J, Kortt A A (1999). High avidity scFv multimers; diabodies and triabodies. *J. Immunol. Meth.* 231:177–189.
22. Zhang J, Tanha J, Hirama T, et al. (2004). Pentamerization of single-domain antibodies from phage libraries: a novel strategy for the rapid generation of high-avidity antibody reagents. *J. Mol. Biol.* 335:49–56.
23. Kim I S, Shim J H, Suh Y T, et al. (2002). Green fluorescent protein-labeled recombinant fluobody for detecting the picloram herbicide. *Biosci. Biotechnol. Biochem.* 66:1148–1151.
24. Benhar I (2001). Biotechnological applications of phage and cell display. *Biotechnol. Adv.* 19:1–33.

25. Krebber A, Bornhauser S, Burmester J, et al. (1997). Reliable cloning of functional antibody variable domains from hybridomas and spleen cell repertoires employing a reengineered phage display system. *J. Immunol. Meth.* 201:35–55.
26. Siegel D L (2002). Recombinant monoclonal antibody technology. *Transfus. Clin. Biol.* 9:15–22.
27. Barbos C F III, Burton D R, Scott J K, et al. (2001). Phage Display: A Laboratory Manual. Cold Sprng Harbor Laboratory, Cold Spring Harbor, N.Y.
28. Feng J N, Russel M, Model P (1997). A permeabilized cell system that assembles filamentous bacteriophage. *Proc. Natl. Acad. Sci. (USA).* 94:4068–4073.
29. Fernandez L A (2004). Prokaryotic expression of antibodies and affibodies. *Curr. Opin. Biotechnol.* 15:364–373.
30. Iannolo G, Minenkova O, Petruzzelli R, et al. (1995). Modifying filamentous phage capsid: limits in the size of the major capsid protein. *J. Mol. Biol.* 248:835–844.
31. MacKenzie R, To R (1998). The role of valency in the selection of anti-carbohydrate single-chain Fvs from phage display libraries. *J. Immunol. Meth.* 220:39–49.
32. Tout N L, Yau K Y, Trevors J T, et al. (2001). Synthesis of ligand-specific phage-display ScFv against the herbicide picloram by direct cloning from hyperimmunized mouse. *J. Agric. Food Chem.* 49:3628–3637.
33. Rondot S, Koch J, Breitling F, et al. (2001). A helper phage to improve single-chain antibody presentation in phage display. *Nature Biotechnol.* 19:75–78.
34. Goletz S, Christensen P A, Kristensen P, et al. (2002). Selection of large diversities of antiidiotypic antibody fragments by phage display. *J. Mol. Biol.* 315:1087–1097.
35. O'Connell D, Becerril B, Roy-Burman A, et al. (2002). Phage versus phagemid libraries for generation of human monoclonal antibodies. *J. Mol. Biol.* 321:49–56.
36. Georgiou G, Poetschke H L, Stathopoulos C, et al. (1993). Practical applications of engineering gram-negative bacterial cell surfaces. *Trends Biotechnol.* 11:6–10.
37. Samuelson P, Gunneriusson E, Nygren P A, et al. (2002). Display of proteins on bacteria. *J. Biotechnol.* 96:129–154.
38. Charbit A, Wang J, Michel V, et al. (1998). A cluster of charged and aromatic residues in the C-terminal portion of maltoporin participates in sugar binding and uptake. *Mol. Gen. Genet.* 260:185–192.
39. Mejare M, Ljung S, Bulow L (1998). Selection of cadmium specific hexapeptides and their expression as OmpA fusion proteins in *Escherichia coli*. *Protein Eng.* 11:489–494.
40. Janssen R, Tommassen J (1994). PhoE protein as a carrier for foreign epitopes. *Int. Rev. Immunol.* 11:113–121.
41. Fuchs P, Breitling F, Dubel S, et al. (1991). Targeting recombinant antibodies to the surface of *Escherichia coli*: fusion to a peptidoglycan associated lipoprotein. *BioTechnol.* 9:1369–1372.
42. Fuchs P, Weichel W, Dubel S, et al. (1996). Separation of E. coli expressing functional cell-wall bound antibody fragments by FACS. *Immunotechnol.* 2:97–102.
43. Hedegaard L, Klemm P (1989). Type 1 fimbriae of *Escherichia coli* as carriers of heterologous antigenic sequences. *Gene.* 85:115–124.
44. Lu Z, Murray K S, Van Cleave V, et al. (1995). Expression of thioredoxin random peptide libraries on the *Escherichia coli* cell surface as functional fusions to flagellin: A system designed for exploring protein-protein interactions. *BioTechnol.* 13: 366–372.
45. Kornacker M G, Pugsley A P (1990). The normally periplasmic enzyme beta-lactamase is specifically and efficiently translocated through the *Escherichia coli* outer

membrane when it is fused to the cell-surface enzyme pullulanase. *Mol. Microbiol.* 4:1101–1109.

46. Francisco J A, Stathopoulos C, Warren R A, et al. (1993). Specific adhesion and hydrolysis of cellulose by intact *Escherichia coli* expressing surface anchored cellulase or cellulose binding domains. *BioTechnol.* 11:491–495.
47. Chen G, Cloud J, Georgiou G, et al. (1996). A quantitative immunoassay utilizing *Escherichia coli* cells possessing surface-expressed single chain Fv molecules. *Biotechnol. Prog.* 12:572–574.
48. Gunneriusson E, Samuelson P, Uhlen M, et al. (1996). Surface display of a functional single-chain Fv antibody on staphylococci. *J. Bacteriol.* 178:1341–1346.
49. Du C, Chan W C, McKeithan T W, et al. (2005). Surface display of recombinant proteins on *Bacillus thuringiensis* spores. *Appl. Environ. Microbiol.* 71:3337–3341.
50. Georgiou G, Stathopoulos C, Daugherty P S, et al. (1997). Display of heterologous proteins on the surface of microorganisms: from the screening of combinatorial libraries to live recombinant vaccines. *Nature Biotechnol.* 15:29–34.
51. Colby D W, Kellogg B A, Graff C P, et al. (2004). Engineering antibody affinity by yeast surface display. *Methods Enzymol.* 388:348–358.
52. Boder E T, Wittrup K D (2000). Yeast surface display for directed evolution of protein expression, affinity, and stability. *Meth. Enzymol.* 328:430–444.
53. Feldhaus M J, Siegel R W (2004). Yeast display of antibody fragments: a discovery and characterization platform. *J. Immunol. Meth.* 290:69–80.
54. Kieke M C, Cho B K, Boder E T, et al. (1997). Isolation of anti-T cell receptor scFv mutants by yeast surface display. *Protein Eng.* 10:1303–1310.
55. Washida M, Takahashi S, Ueda M, et al. (2001). Spacer-mediated display of active lipase on the yeast cell surface. *Appl. Microbiol. Biotechnol.* 56:681–686.
56. Kim S Y, Sohn J H, Pyun Y R, et al. (2002). A cell surface display system using novel GPI-anchored proteins in *Hansenula polymorpha*. *Yeast.* 19:1153–1163.
57. Matsumoto T, Fukuda H, Ueda M, et al. (2002). Construction of yeast strains with high cell surface lipase activity by using novel display systems based on the Flo1p flocculation functional domain. *Appl. Environ. Microbiol.* 68:4517–4522.
58. VanAntwerp J J, Wittrup K D (2000). Fine affinity discrimination by yeast surface display and flow cytometry. *Biotechnol. Prog.* 16:31–37.
59. He M, Taussig M J (1997). Antibody-ribosome-mRNA (ARM) complexes as efficient selection particles for in vitro display and evolution of antibody combining sites. *Nucl. Acids. Res.* 25:5132–5134.
60. Hanes J, Jermutus L, Weber-Bornhauser S, et al. (1998). Ribosome display efficiently selects and evolves high-affinity antibodies in vitro from immune libraries. *Proc. Natl. Acad. Sci. (USA).* 95:14130–14135.
61. Hanes J, Schaffitzel C, Knappik A, et al. (2000). Picomolar affinity antibodies from a fully synthetic naive library selected and evolved by ribosome display. *Nature Biotechnol.* 18:1287–1292.
62. Jermutus L, Honegger A, Schwesinger F, et al. (2001). Tailoring in vitro evolution for protein affinity or stability. *Proc. Natl. Acad. Sci. (USA).* 98:75–80.
63. Olea-Popelka F, McLean M D, Horsman J, et al. (2005). Increasing expression of an anti-picloram single-chain variable fragment (ScFv) antibody and resistance to picloram in transgenic tobacco (*Nicotiana tabacum*). *J. Agric. Food Chem.* 53:6683–6690.

64. Peipp M, Saul D, Barbin K, et al. (2004). Efficient eukaryotic expression of fluorescent scFv fusion proteins directed against CD antigens for FACS applications. *J. Immunol. Meth.* 285:265–280.
65. Kamihira M, Ono K, Esaka K, et al. (2005). High-level expression of single-chain Fv-Fc fusion protein in serum and egg white of genetically manipulated chickens by using a retroviral vector. *J. Virol.* 79:10864–10874.
66. Miller D N, Bryant J E, Madsen E L, et al. (1999). Evaluation and optimization of DNA extraction and purification procedures for soil and sediment samples. *Appl. Environ. Microbiol.* 65:4715–4724.
67. Zhou B, Wirsching P, Janda K D (2002). Human antibodies against spores of the genus *Bacillus*: a model study for detection of and protection against anthrax and the bioterrorist threat. *Proc. Natl. Acad. Sci. (USA).* 99:5241–5246.
68. Lim K P, Li H, Nathan S (2004). Expression and purification of a recombinant scFv towards the exotoxin of the pathogen, *Burkholderia pseudomallei*. *J. Microbiol.* 42:126–132.
69. Nathan S, Li H, Mohamed R, Embi N (2002). Phage display of recombinant antibodies toward *Burkholderia pseudomallei* exotoxin. *J. Biochem. Mol. Biol. Biophys.* 6:45–53.
70. Lindquist E A, Marks J D, Kleba B J, et al. (2002). Phage-display antibody detection of *Chlamydia trachomatis*-associated antigens. *Microbiol.* 148:443–451.
71. McElhiney J, Drever M, Lawton L A, et al. (2002). Rapid isolation of a single-chain antibody against the cyanobacterial toxin microcystin-LR by phage display and its use in the immunoaffinity concentration of microcystins from water. *Appl. Environ. Microbiol.* 68:5288–5295.
72. Deng X K, Nesbit L A, Morrow K J, Jr (2003). Recombinant single-chain variable fragment antibodies directed against *Clostridium difficile* toxin B produced by use of an optimized phage display system. *Clin. Diagn. Lab. Immunol.* 10:587–595.
73. Kanitpun R, Wagner G G, Waghela S D (2004). Characterization of recombinant antibodies developed for capturing enterohemorrhagic *Escherichia coli* O157:H7. *Southeast Asian J. Trop. Med. Public Health.* 35:902–912.
74. Boel E, Bootsma H, de Kruif J, et al. (1998). Phage antibodies obtained by competitive selection on complement-resistant *Moraxella (Branhamella) catarrhalis* recognize the high-molecular-weight outer membrane protein. *Infect. Immun.* 66:83–88.
75. Kuepper M B, Huhn M, Spiegel H, et al. (2005). Generation of human antibody fragments against *Streptococcus mutans* using a phage display chain shuffling approach. *BMC Biotechnol.* 5:1–12.
76. de Greeff A, van Alphen L, Smith H E (2000). Selection of recombinant antibodies specific for pathogenic *Streptococcus suis* by subtractive phage display. *Infect. Immun.* 68:3949–3955.
77. Haidaris C G, Malone J, Sherrill L A, et al. (2001). Recombinant human antibody single chain variable fragments reactive with *Candida albicans* surface antigens. *J. Immunol. Meth.* 257:185–202.
78. Choi G H, Lee D H, Min W K, et al. (2004). Cloning, expression, and characterization of single-chain variable fragment antibody against mycotoxin deoxynivalenol in recombinant *Escherichia coli*. *Protein Expr. Purif.* 35:84–92.
79. Gough K C, Li Y, Vaughan T J, et al. (1999). Selection of phage antibodies to surface epitopes of *Phytophthora infestans*. *J. Immunol. Methods.* 228:97–108.
80. Chen L, Williams B R, Yang C Y, et al. (2003). Polyclonal Fab phage display libraries with a high percentage of diverse clones to *Cryptosporidium parvum* glycoproteins. *Int. J. Parasitol.* 33:281–291.

81. Tachibana H, Cheng X J, Watanabe K, et al. (1999). Preparation of recombinant human monoclonal antibody Fab fragments specific for *Entamoeba histolytica*. *Clin. Diagn. Lab. Immunol.* 6:383–387.
82. Fu Y, Shearing L N, Haynes S, et al. (1997). Isolation from phage display libraries of single chain variable fragment antibodies that recognize conformational epitopes in the malaria vaccine candidate, apical membrane antigen-1. *J. Biol. Chem.* 272: 25678–25684.
83. Kerschbaumer R J, Hirschl S, Kaufmann A, et al. (1997). Single-chain Fv fusion proteins suitable as coating and detecting reagents in a double antibody sandwich enzyme-linked immunosorbent assay. *Anal. Biochem.* 249:219–227.
84. Nagesha H S, Wang L F, Shiell B, et al. (2001). A single chain Fv antibody displayed on phage surface recognises conformational group-specific epitope of bluetongue virus. *J. Virol. Meth.* 91:203–207.
85. Chae J S, Choi J K, Lim H T, et al. (2001). Generation of a murine single chain Fv (scFv) antibody specific for cucumber mosaic virus (CMV) using a phage display library. *Mol. Cells.* 11:7–12.
86. Maruyama T, Parren P W, Sanchez A, et al. (1999). Recombinant human monoclonal antibodies to Ebola virus. *J. Infect. Dis.* 179(suppl 1):S235–239.
87. ShengFeng C, Ping L, Tao S, et al. (2003). Construction, expression, purification, refold and activity assay of a specific scFv fragment against foot and mouth disease virus. *Vet. Res. Commun.* 27:243–256.
88. Saldarelli P, Keller H, Dell'Orco M, et al. (2005). Isolation of recombinant antibodies (scFvs) to grapevine virus B. *J. Virol. Methods.* 124:191–195.
89. Cattani P, Rossolini G M, Cresti S, et al. (1997). Detection and typing of herpes simplex viruses by using recombinant immunoglobulin fragments produced in bacteria. *J. Clin. Microbiol.* 35:1504–1509.
90. Bugli F, Manzara S, Torelli R, et al. (2004) Human monoclonal antibody fragment specific for glycoprotein G in herpes simplex virus type 2 with applications for serotype-specific diagnosis. *J. Clin. Microbiol.* 42:1250–1253.
91. Williamson R A, Lazzarotto T, Sanna P P, et al. (1997). Use of recombinant human antibody fragments for detection of cytomegalovirus antigenemia. *J. Clin. Microbiol.* 35:2047–2050.
92. Sapats S I, Heine H G, Trinidad L, et al. (2003). Generation of chicken single chain antibody variable fragments (scFv) that differentiate and neutralize infectious bursal disease virus (IBDV). *Arch. Virol.* 148:497–515.
93. Desogus A, Burioni R, Ingianni A, et al. (2003). Production and characterization of a human recombinant monoclonal Fab fragment specific for influenza A viruses. *Clin. Diagn. Lab. Immunol.* 10:680–685.
94. Liu H, Ding Y L, Han W, Liu M Y, et al. (2004). Recombinant scFv antibodies against E protein and N protein of severe acute respiratory syndrome virus. *Acta Biochim. Biophys. Sin.* 36:541–547.
95. Schmaljohn C, Cui Y, Kerby S, et al. (1999). Production and characterization of human monoclonal antibody Fab fragments to vaccinia virus from a phage-display combinatorial library. *Virol.* 258:189–200.
96. Duggan J M, Coates D M, Ulaeto D O (2001). Isolation of single-chain antibody fragments against Venezuelan equine encephalomyelitis virus from two different immune sources. *Viral Immunol.* 14:263–273.
97. Hu W G, Alvi A Z, Chau D, et al. (2003). Development and characterization of a novel fusion protein composed of a human IgG1 heavy chain constant region and a

single-chain fragment variable antibody against Venezuelan equine encephalitis virus. *J. Biochem.* 133:59–66.

98. Cardinale A, Filesi I, Vetrugno V, et al. (2005). Trapping prion protein in the endoplasmic reticulum impairs PrPC maturation and prevents PrPSc accumulation. *J. Biol. Chem.* 280:685–694.
99. Devaux C, Moreau E, Goyffon M, et al. (2001). Construction and functional evaluation of a single-chain antibody fragment that neutralizes toxin AahI from the venom of the scorpion *Androctonus australis hector. Eur. J. Biochem.* 268:694–702.
100. Andreotti P E, Ludwig G V, Peruski A H, et al. (2003). Immunoassay of infectious agents. *Biotechn.* 35:850–859.
101. Payne W J, Jr, Marshall D L, Shockley R K, et al. (1988). Clinical laboratory applications of monoclonal antibodies. *Clin. Microbiol. Rev.* 1:313–329.
102. Arens M Q, Mundy L M, Amsterdam D, et al. (2005). Preclinical and clinical performance of the Efoora test, a rapid test for detection of human immunodeficiency virus-specific antibodies. *J. Clin. Microbiol.* 43:2399–2406.
103. Walker M, Redelman D (2004). Detection of *Cryptosporidium parvum* in soil extracts. *Appl. Environ. Microbiol.* 70:1827–1829.
104. Matsubayashi M, Kimata I, Iseki M, et al. (2005). Cross-reactivities with *Cryptosporidium* spp. by chicken monoclonal antibodies that recognize avian *Eimeria* spp. *Vet. Parasitol.* 10:47–57.
105. Casadevall A, Pirofski L A (2001). Adjunctive immune therapy for fungal infections. *Clin. Infect. Dis.* 33:1048–1056.
106. Casadevall A, Dadachova E, Pirofski L A (2004). Passive antibody therapy for infectious diseases. *Nature Rev. Microbiol.* 2:695–703.
107. Casadevall A, Pirofski L A (2004). New concepts in antibody-mediated immunity. *Infect. Immun.* 72:6191–6196.
108. Cenci E, Bistoni F, Mencacci A, et al. (2004). A synthetic peptide as a novel anticryptococcal agent. *Cell Microbiol.* 6:953–961.
109. Crabb J H (1998). Antibody-based immunotherapy of cryptosporidiosis. *Adv. Parasitol.* 40:121–149.
110. Hassani M, Patel M C, Pirofski L A (2004). Vaccines for the prevention of diseases caused by potential bioweapons. *Clin. Immunol.* 111:1–15.
111. Hengel H, Masihi K N (2003). Combinatorial immunotherapies for infectious diseases. *Int. Immunopharmacol.* 3:1159–1167.
112. Keller M A, Stiehm E R (2000). Passive immunity in prevention and treatment of infectious diseases. *Clin. Microbiol. Rev.* 13:602–614.
113. Berghman L R, Abi-Ghanem D, Waghela S D, et al. (2005). Antibodies: an alternative for antibiotics? *Poult. Sci.* 84:660–666.
114. Casadevall A (2003). Antibody-mediated immunity against intracellular pathogens: two-dimensional thinking comes full circle. *Infect. Immun.* 71:4225–4228.
115. Joiner K A, Dubremetz JF (1993). *Toxoplasma gondii*: a protozoan for the nineties. *Infect. Immun.* 61:1169–1172.
116. Nardin E H, Nussenzweig R S (1993). T cell responses to pre-erythrocytic stages of malaria: role in protection and vaccine development against pre-erythrocytic stages. *Annu. Rev. Immunol.* 11:687–727.
117. Brekke O H, Løset G A (2003). New technologies in therapeutic antibody development. *Curr. Opin. Pharmacol.* 3:544–550.
118. Casadevall A (2004). The methodology for determining the efficacy of antibody-mediated immunity. *J. Immunol. Meth.* 291:1–10.

119. Chester K A, Hawkins R E (1995). Clinical issues in antibody design. *Trends Biotechnol.* 13:294–300.
120. Wild M A, Xin H, Maruyama T, et al. (2003). Human antibodies from immunized donors are protective against anthrax toxin in vivo. *Nature Biotechnol.* 21:1305–1306.
121. Beninati C, Arseni S, Mancuso G, et al. (2004). Protective immunization against group B meningococci using anti-idiotypic mimics of the capsular polysaccharide. *J. Immunol.* 172:2461–2468.
122. Magliani W, Conti S, Salati A, et al. (2003). Biotechnological approaches to the production of idiotypic vaccines and antiidiotypic antibiotics. *Curr. Pharm. Biotechnol.* 4:91–97.
123. Polonelli L, Magliani W, Conti S, et al. (2003). Therapeutic activity of an engineered synthetic killer antiidiotypic antibody fragment against experimental mucosal and systemic candidiasis. *Infect. Immun.* 71:6205–6212.
124. Matthews R C, Rigg G, Hodgetts S, et al. (2003). Preclinical assessment of the efficacy of mycograb, a human recombinant antibody against fungal HSP90. *Antimicrob. Agents Chemother.* 47:2208–2216.
125. Peschen D, Li H P, Fischer R, et al. (2004). Fusion proteins comprising a *Fusarium*-specific antibody linked to antifungal peptides protect plants against a fungal pathogen. *Nature Biotechnol.* 22:732–738.
126. Li F, Dluzewski A, Coley A M, et al. (2002). Phage-displayed peptides bind to the malarial protein apical membrane antigen-1 and inhibit the merozoite invasion of host erythrocytes. *J. Biol. Chem.* 277:50303–50310.
127. Vukovic P, Chen K, Qin Liu X, et al. (2002). Single-chain antibodies produced by phage display against the C-terminal 19 kDa region of merozoite surface protein-1 of *Plasmodium yoelii* reduce parasite growth following challenge. *Vaccine.* 20:2826–2835.
128. Yoshida S, Matsuoka H, Luo E, et al. (1999). A single-chain antibody fragment specific for the *Plasmodium berghei* ookinete protein Pbs21 confers transmission blockade in the mosquito midgut. *Mol. Biochem. Parasitol.* 104:195–204.
129. Villani M E, Roggero P, Bitti O, et al. (2005). Immunomodulation of cucumber mosaic virus infection by intrabodies selected in vitro from a stable single-framework phage display library. *Plant Mol. Biol.* 58:305–316.
130. Kim Y J, Lebreton F, Kaiser C, et al. (2004). Isolation of foot-and-mouth disease virus specific bovine antibody fragments from phage display libraries. *J. Immunol. Meth.* 286:155–166.
131. De Logu A, Williamson R A, Rozenshteyn R, et al. (1998). Characterization of a type-common human recombinant monoclonal antibody to herpes simplex virus with high therapeutic potential. *J. Clin. Microbiol.* 36:3198–3204.
132. Shaheen F, Duan L, Zhu M, et al. (1996). Targeting human immunodeficiency virus type 1 reverse transcriptase by intracellular expression of single-chain variable fragments to inhibit early stages of the viral life cycle. *J. Virol.* 70:3392–3400.
133. Goncalves J, Silva F, Freitas-Vieira A, et al. (2002). Functional neutralization of HIV-1 Vif protein by intracellular immunization inhibits reverse transcription and viral replication. *J. Biol. Chem.* 277:32036–32045.
134. Wu S C, Lin Y J, Chou J W, et al. (2004). Construction and characterization of a Fab recombinant protein for Japanese encephalitis virus neutralization. *Vaccine.* 23:163–171.
135. de Carvalho Nicacio C, Williamson R A, Parren P W, et al. (2002). Neutralizing human Fab fragments against measles virus recovered by phage display. *J. Virol.* 76:251–258.

136. Ray K, Embleton M J, Jailkhani B L, et al. (2001). Selection of single chain variable fragments (scFv) against the glycoprotein antigen of the rabies virus from a human synthetic scFv phage display library and their fusion with the Fc region of human IgG1. *Clin. Exp. Immunol.* 125:94–101.
137. Higo-Moriguchi K, Akahori Y, Iba Y, et al. (2004). Isolation of human monoclonal antibodies that neutralize human rotavirus. *J. Virol.* 78:3325–3332.
138. Kausmally L, Waalen K, Lobersli I, et al. (2004). Neutralizing human antibodies to varicella-zoster virus (VZV) derived from a VZV patient recombinant antibody library. *J. Gen. Virol.* 85:3493–3500.
139. Donofrio G, Heppner F L, Polymenidou M, et al. (2005). Paracrine inhibition of prion propagation by anti-PrP single-chain Fv miniantibodies. *J. Virol.* 79:8330–8338.
140. Peretz D, Williamson R A, Kaneko K, et al. (2001). Antibodies inhibit prion propagation and clear cell cultures of prion infectivity. *Nature.* 412:739–743.
141. Riggs M W, Stone A L, Yount P A, et al. (1997). Protective monoclonal antibody defines a circumsporozoite-like glycoprotein exoantigen of *Cryptosporidium parvum* sporozoites and merozoites. *J. Immunol.* 158:1787–1795.
142. Freedman D J, Tacket C O, Delehanty A, et al. (1998). Milk immunoglobulin with specific activity against purified colonization factor antigens can protect against oral challenge with enterotoxigenic *Escherichia coli*. *J. Infect. Dis.* 177:662–667.
143. Tacket C O, Binion S B, Bostwick E, et al. (1992). Efficacy of bovine milk immunoglobulin concentrate in preventing illness after *Shigella flexneri* challenge. *Am. J. Trop. Med. Hyg.* 47:276–283.
144. Hayden M S, Gilliland L K, Ledbetter J A (1997). Antibody engineering. *Curr. Opin. Immunol.* 9:201–212.
145. Nabel GJ (2004). Genetic, cellular and immune approaches to disease therapy: past and future. *Nature Med.* 10:135–141.
146. Griffiths A D, Williams S C, Hartley O, et al. (1994). Isolation of high affinity human antibodies directly from large synthetic repertoires. *Embo. J.* 13:3245–3260.
147. Montgomery D L, Ulmer J B, Donnelly J J, et al. (1997). DNA vaccines. *Pharmacol. Ther.* 74:195–205.
148. Rocha C D, Caetano B C, Machado A V, et al. (2004). Recombinant viruses as tools to induce protective cellular immunity against infectious diseases. *Int. Microbiol.* 7:83–94.
149. Penttila T, Tammiruusu A, Liljestrom P, et al. (2004). DNA immunization followed by a viral vector booster in a *Chlamydia pneumoniae* mouse model. *Vaccine.* 22: 3386–3394.
150. Hong-Xuan H, Lei C, Cheng-Min W, et al. (2005). Expression of the recombinant fusion protein CP15-23 of *Cryptosporidium parvum* and its protective test. *J. Nanosci. Nanotechnol.* 5:1292–1296.
151. Melzer H, Baier K, Felici F, et al. (2003). Humoral immune response against proteophosphoglycan surface antigens of *Entamoeba histolytica* elicited by immunization with synthetic mimotope peptides. *FEMS Immunol. Med. Microbiol.* 37:179–183.
152. Gonzalo R M, Rodriguez D, Garcia-Sastre A, et al. (1999). Enhanced CD8+ T cell response to HIV-1 env by combined immunization with influenza and vaccinia virus recombinants. *Vaccine.* 17:887–892.
153. Simon B E, Cornell K A, Clark T R, et al. (2003). DNA vaccination protects mice against challenge with *Listeria monocytogenes* expressing the hepatitis C virus NS3 protein. *Infect. Immun.* 71:6372–6380.
154. Sechi L A, Mara L, Cappai P, et al. (2006). Immunization with DNA vaccines encoding different mycobacterial antigens elicits a Th1 type immune response in lambs and

protects against *Mycobacterium avium* subspecies paratuberculosis infection. *Vaccine*. 3:229–235.

155. Moorthy V S, Pinder M, Reece W H, et al. (2003). Safety and immunogenicity of DNA/modified vaccinia virus ankara malaria vaccination in African adults. *J. Infect. Dis.* 188:1239–1244.
156. Xia Y J, Wen W H, Huang W Q, et al. (2005). Development of a phage displayed disulfide-stabilized Fv fragment vaccine against *Vibrio anguillarum*. *Vaccine*. 23:3174–3180.
157. Sun J H, Yan Y X, Jiang J, et al. (2005). DNA immunization against very virulent infectious bursal disease virus with VP2-4-3 gene and chicken IL-6 gene. *J. Vet. Med. B Infect. Dis. Vet. Public. Health*. 52:1–7.
158. Murata K, Garcia-Sastre A, Tsuji M, et al. (1996). Characterization of in vivo primary and secondary CD8+ T cell responses induced by recombinant influenza and vaccinia viruses. *Cell Immunol.* 173:96–107.
159. He H, Zhao B, Liu L, et al. (2004). The humoral and cellular immune responses in mice induced by DNA vaccine expressing the sporozoite surface protein of *Cryptosporidium parvum*. *DNA Cell Biol.* 23:335–339.
160. Ivory C, Chadee K (2004). DNA vaccines: designing strategies against parasitic infections. *Genet. Vaccines Ther.* 2:17.

# THE RADIOPHARMACEUTICAL SCIENCE OF MONOCLONAL ANTIBODIES AND PEPTIDES FOR IMAGING AND TARGETED *IN SITU* RADIOTHERAPY OF MALIGNANCIES

RAYMOND M. REILLY
*University of Toronto, Toronto, Ontario, Canada*

**Chapter Contents**

1 Introduction
2 Tumor and Normal Tissue Targeting Properties of mAbs and Peptides
3 Selection of a Radionuclide for Tumor Imaging
4 Selection of a Radionuclide for Targeted *in situ* Radiotherapy
5 Labeling mAbs and Peptides with Radiohalogens
  5.1 Iodine Radionuclides
  5.2 Bromine Radionuclides
  5.3 Fluorine Radionuclides
  5.4 Astatine Radionuclides
6 Labeling mAbs and Peptides with Radiometals
  6.1 Technetium Radionuclides
  6.2 Rhenium Radionuclides
  6.3 Indium Radionuclides
  6.4 Yttrium Radionuclides
  6.5 Gallium Radionuclides
  6.6 Copper Radionuclides
  6.7 Lutetium Radionuclides
  6.8 Lead, Bismuth, and Actinium Radionuclides
7 Characterization of Radiolabeled mAbs and Peptides
  7.1 Evaluation of the Homogeneity of Radiopharmaceutical Bioconjugates

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

7.2 Measurement of Radiochemical Purity (RCP)
7.3 Measurement of Immunoreactivity or Receptor-Binding Properties
7.4 Evaluation of *In Vitro* and *In Vivo* Stability
7.5 Preclinical Biodistribution, Tumor Imaging, and Dosimetry Studies
7.6 Preclinical Studies to Evaluate AntiTumor Effects and Normal Tissue Toxicity
7.7 Kit Formulation and Associated Pharmaceutical Testing
8 Summary
Acknowledgment
References

## 1 INTRODUCTION

Cancer is a major health problem worldwide. It is estimated that more than 10 million new cases of cancer are diagnosed annually around the globe and that more than 4 million individuals die each year from the disease [1]. Intense research over the past few decades has led to important discoveries by cancer biologists that are just now stimulating the development of potentially more effective and safer biologically targeted therapies for the disease. One promising strategy for treating malignancies that exploits their biological phenotype is targeted *in situ* radiotherapy [2]. In this approach, monoclonal antibodies (mAbs) that recognize tumor-associated antigens or peptide ligands that specifically bind to cell-surface growth factor receptors are used as targeting vehicles to selectively deliver radionuclides to cancer cells for *in situ* radiation therapy. These two approaches are known as radioimmunotherapy (RIT) or peptide-directed radiotherapy (PDRT), respectively. An extension of RIT or PDRT is to employ radiolabeled mAbs or peptides for imaging metastatic deposits for detection or to noninvasively characterize their phenotype *in situ*, known as molecular imaging [3]. Phenotypic characterization of tumors will be critical in the future to be able to appropriately select patients for treatment with new biologically targeted anticancer agents, including targeted radiotherapeutics. For example, it is known that breast cancers that exhibit high levels of amplification of the human epidermal growth factor receptor-2 (HER2/neu) gene respond best to treatment with the humanized anti-HER2/neu mAb, trastuzumab (Herceptin; Roche Pharmaceuticals Ltd., Nutley, NJ) [4]. HER2/neu gene amplification is currently evaluated in a primary breast cancer biopsy by immunohistochemical staining for the HER2/neu protein or by fluorescence *in situ* hybridization (FISH) for the gene copy number. A recent report, however, suggests that molecular imaging of HER2/neu expression in breast tumors using indium-111 ($^{111}$In)-labeled trastuzumab may reliably predict which patients respond to Herceptin as well as identify those who may be at risk for toxicity from the drug [5]. Moreover, imaging studies using radiolabeled forms of biopharmaceuticals could allow a noninvasive *in situ* assessment of their pharmacokinetic properties and, in particular, their tumor and normal tissue uptake and elimination. This may provide insight into their effectiveness as antitumor agents as well as their potential sites of normal organ toxicity. There have been many comprehensive reviews on the topics of imaging and targeted *in situ* radiotherapy of malignancies with radiolabeled mAbs and peptides [2, 3, 6]. In this chapter, the radiopharmaceutical science

that provides the foundation for these biomolecules for targeting radionuclides to tumors is discussed.

## 2 TUMOR AND NORMAL TISSUE TARGETING PROPERTIES OF MABS AND PEPTIDES

The tumor and normal tissue targeting properties of mAbs and their fragments and peptides must be considered when designing a radiopharmaceutical for molecular imaging or targeted *in situ* radiotherapy of cancer. Intact IgG mAbs ($M_r$ = 150 kDa; Figure 1) are macromolecules that are cleared slowly from the blood with an elimination half-life of 2–3 days for murine forms and 4 days for chimeric and humanized mAbs [7]. This slow elimination provides multiple passes through a tumor at which extravasation and interaction with tumor cells may occur. Thus, the tumor accumulation [percent injected dose/g (% i.d./g)] of radiolabeled intact IgG mAbs is much greater than that of smaller and more rapidly cleared antibody fragments (e.g., Fab or scFv; Figure 1) or that of small peptides. This property renders intact IgG mAbs more attractive for RIT, because it maximizes the amount of radioactivity that is delivered to the tumor and, thus, the radiation absorbed dose delivered. However, the macromolecular properties of intact IgG mAbs severely restricts their tumor penetration. It has been shown that radiolabeled mAbs remain close to tumor blood vessels, whereas Fab and single-chain Fv (scFv) fragments penetrate more deeply into tumors [8]. This differential depth of penetration for intact IgG mAbs compared with their fragments may be related in part to antigen-binding affinity and avidity, because it has been hypothesized that a "binding-site barrier" restricts the penetration of antibodies into tumors as a consequence of interaction with antigens on cancer cells proximal to the blood vessels [9, 10]. Antibody fragments (Fab and scFv) are monovalent, which diminishes their avidity and affinity compared with divalent IgG mAbs. The poor tumor penetration

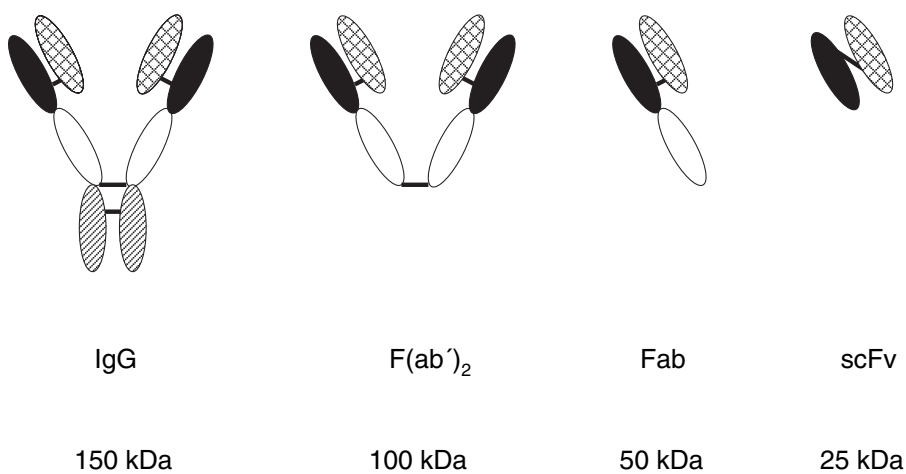


**Figure 1.** Structures and corresponding molecular weights of various antibody forms used for imaging or targeted *in situ* radiotherapy of cancer.

properties of intact IgG mAbs produces heterogeneous distribution of radioactivity within a tumor. Moreover, the Fc-domain of intact mAbs is recognized by asialoglycoprotein receptors on hepatocytes, which promotes liver accumulation of radioactivity [7]. Retention of radioactivity in the liver interferes with detection of hepatic metastases, especially when radiometal-labeled mAbs and peptides are used for imaging. This is less problematic for radioiodinated biomolecules due to their intracellular catabolism and release of radioiodine from hepatocytes.

For tumor imaging, antibody fragments [$F(ab')_2$, Fab, or scFv] or peptides are the most useful because they are rapidly eliminated from the blood and most normal tissues (except the kidneys), which minimizes circulating background radioactivity and yields high tumor/blood (T/B) and tumor/normal tissue (T/NT) ratios that occur at early times after injection. The major challenge with radiolabeled mAb fragments and peptides is their high accumulation in the kidneys. Kidney uptake is thought to be due to glomerular filtration of the fragments or peptides followed by charge interactions between cationic amino acid residues (e.g., lysine and arginine) in the proteins and the negatively charged renal tubular cell membrane [11]. High renal uptake of the somatostatin octapeptide analog, DOTATOC labeled with the β-emitter, $^{90}Y$ is associated with serious kidney toxicity in patients when used for PDRT of neuroendocrine malignancies [2]. However, renal toxicity can be avoided by co-administering intravenous solutions of lysine or arginine that competitively inhibit the interaction between renal tubular cells and radiolabeled mAb fragments or peptides [11, 12]. Furthermore, renal toxicity is not associated with the use of these radiopharmaceuticals for tumor imaging, because they are administered at much lower doses and labeled with low linear energy transfer (LET) γ-emitting radionuclides.

## 3 SELECTION OF A RADIONUCLIDE FOR TUMOR IMAGING

Radionuclides suitable for labeling mAbs or peptides for tumor imaging may be single γ-photon emitters (Table 1) or positron-emitters (Table 2). Images using single γ-photons are usually acquired in three-dimensional mode, known as single-photon emission computerized tomography (SPECT; Figure 2). The single γ-photons are detected by thallium-doped sodium iodide [NaI(Tl)] scintillation crystals that are housed in two opposing heads of a γ-camera. The heads are rotated 180° around the subject to obtain a series of images. Images of the body, organ, or tissue of interest are reconstructed by back-extrapolation of the lines of detection acquired by each camera head. This technique allows the reconstructed images to be "sliced" to visualize a single image plane separate from any overlying or underlying interfering planes. Optimal γ-energies for SPECT imaging are 100–300 keV. The corresponding imaging technique for positron-emitters is called positron-emission tomography (PET). Positrons are $\beta^+$ particles that are emitted when a proton is converted to a neutron in the decay scheme of certain radionuclides to stable elements. The positrons travel a distance (0.7 to 5 mm) that is directly proportional to their energy before they are annihilated through interaction with an electron in tissues. Positron annihilation creates two 511-keV γ-photons that travel out at approximately 180° from one another at the site of annihilation. PET relies on the simultaneous detection of these two γ-photons within a narrow

**TABLE 1. Radionuclides Suitable for Labeling Biomolecules for SPECT Imaging of Tumors**

| Radionuclide | Production Method | Eγ (abundence) | $T_{1/2}$phys | Labeling Methods |
|---|---|---|---|---|
| $^{99m}$Tc | $^{99}$Mo/$^{99m}$Tc generator | 140.5 keV (98.9%) | 6.0 hours | Binding to thiols; chelation by tetradentate complexes; HYNIC; interaction of carbonyl complex with histidine residues |
| $^{111}$In | $^{112}$Cd(p,2n)$^{111}$In | 171.3 keV (90.2%), 245.4 keV (94.0%) | 2.8 days | Chelation by DTPA, SCN-Bz-DTPA or DOTA |
| $^{67}$Ga | $^{68}$Zn(p,2n)$^{67}$Ga | 93.3 keV (35.7%), 184.6 keV (19.7%), 300.2 keV (16.0%) | 3.3 days | Chelation by desferrioxamine (DFO) or DOTA |
| $^{123}$I | $^{124}$Xe(p,2n)$^{123}$Cs→$^{123}$Xe →$^{123}$I; $^{124}$Xe(p,pn)$^{123}$Xe→$^{123}$I | 159 keV (83.4%) | 13.2 hours | Direct radio-iodination with chloramine-T or Iodogen; indirect conjugation using ATE, SIPC, SGMIB; indirect conjugation using carbohydrate adducts such as TCB or dextran |
| $^{131}$I | $^{130}$Te(n,γ)$^{131}$Te→$^{131}$I | 364 keV (81.2%) | 8.0 days | Same as for $^{123}$I |

**TABLE 2. Radionuclides Suitable for Labeling Biomolecules for PET Imaging of Tumors**

| Radionuclide | Production Method | $\beta^+$ Energy (Abundence) | Intrinsic Spatial Resolution | $T_{1/2}$phys | Labeling Methods |
|---|---|---|---|---|---|
| $^{124}I$ | $^{124}Te(p,n)^{124}I$ or $^{124}Te(d,2n)^{124}I$ | 0.8–2.1 MeV (23%) | 2.3 mm | 4.2 days | Direct radio-iodination with chloramine-T or Iodogen |
| $^{76}Br$ | $^{76}Se(p,n)^{76}Br$ | 3.4 MeV (54%) | 5.0 mm | 16.2 hours | Direct radiobromination with chloramine-T; indirect conjugation using SPBrB, HPEM, brom-3-pyridine-carboxylate, *closo*-dodecaborate, *nido*-undecaborate |
| $^{18}F$ | $^{18}O(p,n)^{18}F$ | 0.6 MeV (100%) | 0.7 mm | 110 minutes | Indirect conjugation using SFBS, SFB, NPFP, FB-CHO, AFP |
| $^{94m}Tc$ | $^{94}Mo(p,n)^{94m}Tc$ | 2.5 MeV (72%) | 3.3 mm | 52 minutes | $N_4$-tetradentate chelator |
| $^{110m}In$ | $^{110}Cd(p,n)^{110m}In$ | 2.3 MeV (62%) | 3.0 mm | 69 minutes | Chelation by DTPA |
| $^{86}Y$ | $^{86}Sr(p,n)^{86}Y$ | 1.2–1.5 MeV (33%) | 1.8 mm | 14.7 hours | Chelation by DOTA |
| $^{68}Ga$ | $^{68}Ge/^{68}Ga$ generator | 1.9 MeV (89%) | 2.4 mm | 68 minutes | Chelation by DOTA |
| $^{64}Cu$ | $^{64}Ni(p,n)^{64}Cu$ | 0.7 MeV (19%) | 0.7 mm | 12.7 hours | Chelation by DOTA, BAT, or TETA |

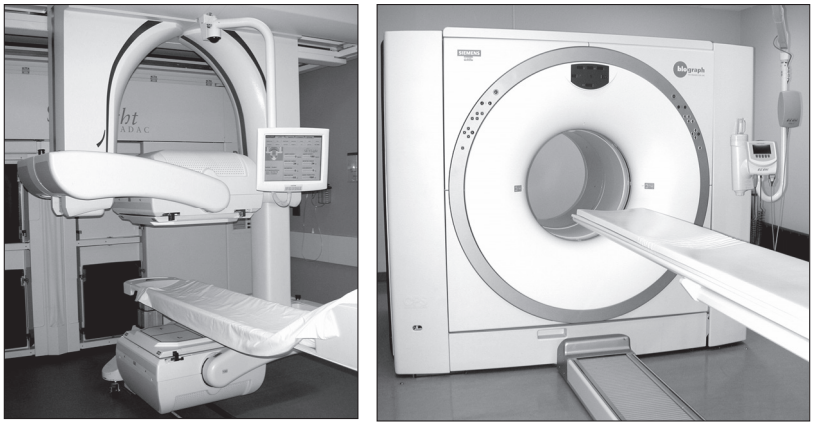

SPECT PET

**Figure 2.** Clinical imaging systems used in nuclear medicine to visualize the distribution of radiopharmaceuticals labeled with single photon-emitting radionuclides or with positron-emitting radionuclides. *Left Panel*: A gamma camera used for single photon-emission computed tomography (SPECT). *Right Panel*: A positron-emission tomograph used for PET scanning. (Photographs generously provided by Deborah Scollard.) (This figure is available in full color at ftp://ftp.wiley.com/public/sci_ tech_med/pharmaceutical_ biotech/.)

time window ("time of flight") by a ring of lutetium silicate (LSO) or bismuth subgerminate (BGO) scintillation crystals surrounding the subject (Figure 2). Back-extrapolation of the lines of detection allows accurate identification of the site of positron annihilation and yields high spatial resolution images. However, it is important to recognize that the spatial resolution of PET is impacted by the finite distance that the positron travels before its annihilation. This distance ranges from 0.7 mm for the low-energy, 0.6-MeV $\beta^+$ particles of $^{18}F$ to 5.3 mm for the higher energy 1.7-MeV $\beta^+$-particles emitted by $^{76}Br$ (Table 2) [13].

It is important to consider the pharmacokinetic properties of the vehicle in selecting an appropriate radionuclide for tumor imaging. Intact IgG mAbs require several days to reach maximum tumor uptake and to be cleared from the blood and normal tissues. Therefore, single $\gamma$-photon-emitters such as $^{111}In$ or $^{131}I$ with a half-life of 2.8 or 8 days, respectively (Table 1), or longer lived positron-emitters such as $^{124}I$ (half-life of 4 days; Table 2) are most suitable for labeling these vehicles for tumor imaging. Antibody fragments such as Fab or scFv are cleared more rapidly from the blood and most normal tissues (except the kidneys), yielding high T/B and T/NT ratios within 24 hours. Therefore, these vehicles may be labeled with short-to-intermediate half-life single $\gamma$-photon emitters such as $^{99m}Tc$, $^{123}I$, or $^{67}Ga$ for SPECT (Table 1) or with $^{64}Cu$ or $^{86}Y$ for PET (Table 2). The shorter half-life positron-emitters such as $^{18}F$ and $^{68}Ga$ are reserved for labeling peptides and scFv fragments of mAbs that exhibit very rapid tumor uptake and elimination from the blood and normal tissues. Ultra-short-lived positron-emitters such as $^{15}O$, $^{13}N$, or $^{11}C$ or (half-lives of 2, 10 or 20 minutes, respectively) that have been used for labeling small molecules for PET, are not feasible for imaging using mAbs or peptides.

## 4 SELECTION OF A RADIONUCLIDE FOR TARGETED *IN SITU* RADIOTHERAPY

Radionuclides suitable for targeted *in situ* radiotherapy of tumors (Table 3) emit either α-particles, β-particles, or Auger and conversion electrons [2]. The important differences between these different forms of radiation are their range in tissues and their LET. α-Particles consist of two protons and two neutrons and carry a $2^+$ charge. These particles have the highest LET (100 keV/μm), are densely ionizing, and travel 50–100 μm (5–10 cell diameters) in tissues. α-emitters such as $^{211}$At, $^{212}$Bi, or $^{225}$Ac are most useful for eradicating small clusters of cancer cells or micrometastases. β-particles are high-energy electrons that have a $1^-$ charge and travel 2–12 mm in tissues (200–1200 cell diameters). β-particles deposit most of their energy at the end of their track length in tissues. However, for comparison with α-particles, the average LET of the β-particles emitted by $^{131}$I (Eβ = 0.6 MeV) over their 2-mm track length is 0.3 keV/μm. The average LET of the β-particles emitted by $^{90}$Y (Eβ = 2.3 MeV) over their 10–12-mm track length is 0.2 keV/μm. Due to the long range of the β-particles emitted by $^{131}$I, $^{90}$Y, $^{186}$Re/$^{188}$Re, or $^{175}$Lu conjugated to mAbs or peptides that are targeted to tumor cells, it is possible to irradiate and kill more distant nontargeted tumor cells ("cross-fire" effect). This is advantageous for larger lesions (i.e., 2–10 mm in diameter) in which there is likely to be incomplete targeting of tumor cells. Inadequate targeting could allow some cells to survive. However, the "cross-fire" effect of the β-particles also contributes to dose-limiting bone marrow toxicity in RIT, due to nonspecific irradiation of hematopoietic stem cells by circulating radioactivity perfusing the bone marrow [2].

Auger electrons are very low-energy electrons emitted by radionuclides that decay by electron capture (EC). In EC, a proton in the nucleus captures an electron from an inner orbital shell, creating a vacancy in the shell. This vacancy is filled by the decay of an electron from a higher shell. The excess energy released is transferred to an outer orbital electron that is then ejected from the atom, creating a $2^+$ atomic species. Because of their very low energy (<30 keV), Auger electrons travel only a few nanometers to at most a few micrometers in tissues (less than one cell diameter). Their LET approaches that of α-emitters (100 keV/μm). However, an antibody or peptide carrying an Auger electron-emitting radionuclide such as $^{125}$I, $^{123}$I, $^{111}$In, or $^{67}$Ga must be internalized and, ideally, translocated to the nucleus for the electrons to be the most damaging to DNA. Biomolecules that recognize peptide growth factor receptors [e.g., epidermal growth factor receptors (EGFRs) or somatostatin receptors (SMSRs)] are frequently internalized into cells and in some cases translocated to the nucleus, which allows them to be employed for targeted Auger electron radiotherapy of malignancies. The advantage of Auger electron-emitting radionuclides is that it is possible to restrict killing to cells that specifically bind and internalize radiolabeled mAbs or peptides. There is no "cross-fire" effect from Auger electron-emitting radionuclides, which should obviate any major nonspecific radiotoxicity to bone marrow stem cells, particularly if these cells do not express the target epitopes/receptors. On the other hand, in contrast to the more energetic β-emitters, the lack of a "cross-fire" effect from Auger electron-emitters does not permit killing of nontargeted cancer cells, although a "bystander effect" has been reported [14]. Auger electron-emitting radionuclides are therefore most useful for treating small tumor deposits or micrometastases for which delivery of radiolabeled biomolecules is more homogeneous.

**TABLE 3. Radionuclides Suitable for Labeling Biomolecules for Targeted *in situ* Radiotherapy of Tumors**

| Radionuclide | Production Method | Particulate Emissions (Energy) | Maximum Range in Tissues | $T_{1/2}$phys | Labeling Methods |
|---|---|---|---|---|---|
| $^{125}I$ | $^{124}Xe(n,\gamma)^{125}Xe \rightarrow ^{125}I$ | Auger electrons (<30 keV) | <10 μm | 59.4 days | Direct radio-iodination with chloramine-T or Iodogen; indirect conjugation using ATE, SIPC, SGMIB |
| $^{123}I$ | $^{124}Xe(p,2n)^{123}Cs \rightarrow ^{123}Xe \rightarrow ^{123}I$ | Auger electrons | <10 μm | 13.2 hours | Same as for $^{125}I$ |
| $^{131}I$ | Neutron irradiation of $^{130}Te$ | β-particles (0.6 MeV) | 2 mm | 8.0 days | Same as for $^{125}I$ |
| $^{211}At$ | $^{209}Bi(\alpha,2n)^{211}At$ | α-particles (5.9–7.4 MeV) | 50–100 μm | 7.2 hours | Indirect conjugation of ATE, SAB, SAPC |
| $^{186}Re$ | $^{185}Re(n,\gamma)^{186}Re$ | β-particles (1.1 MeV) | 3 mm | 3.7 days | Binding to thiols; chelation by tetradentate complexes; HYNIC; interaction of carbonyl complex with histidine residues; trisuccin |
| $^{188}Re$ | $^{188}W/^{188}Re$ generator | β-particles (2.1 MeV) | 8 mm | 17 hours | Same as for $^{186}Re$ |
| $^{90}Y$ | $^{90}Sr/^{90}Y$ generator | β-particles (2.3 MeV) | 12 mm | 2.7 days | Chelation by DOTA |
| $^{67}Ga$ | $^{68}Zn(p,2n)^{67}Ga$ | Auger electrons (<30 keV) | <10 μm | 3.3 days | Chelation by DFO and DOTA |
| $^{64}Cu$ | $^{68}Zn(p,\alpha n)^{64}Cu$ | β-particles (0.6 MeV) | 2 mm | 12.7 hours | Chelation by DOTA, BAT, or TETA |
| $^{67}Cu$ | $^{nat}Zn(p,2p)^{67}Cu$ or $^{68}Zn(p,2p)^{67}Cu$ | β-particles (0.4–0.6 MeV) | 2 mm | 2.6 days | Same as for $^{64}Cu$ |
| $^{177}Lu$ | $^{176}Yb(n,\gamma) \rightarrow ^{177}Yb \rightarrow ^{177}Lu$ | β-particles (0.5 MeV) | 2 mm | 6.6 days | Chelation by DOTA or CHX–DTPA |
| $^{213}Bi$ | $^{225}Ac/^{213}Bi$ generator | α-particles (8 MeV) | 50–100 μm | 46 minutes | Chelation by DOTA or CHX–DTPA |
| $^{225}Ac$ | $^{233}U \rightarrow ^{225}Ac$ | α-particles (several duaghter radionuclides with different energies) | 50–100 μm | 10 days | Chelation by DOTA |

## 5 LABELING MABS AND PEPTIDES WITH RADIOHALOGENS

### 5.1 Iodine Radionuclides

Radioiodination is one of the simplest ways to radiolabel a biomolecule. Several radionuclides of iodine are available for SPECT or PET ($^{123}$I, $^{124}$I, $^{131}$I; Tables 1 and 2) or for targeted *in situ* radiotherapy of cancer ($^{125}$I and $^{131}$I; Table 3). Iodine is present in a valence state of $1^-$ in the alkaline solution in which it is supplied. It requires oxidation to a $1^+$ valence state for electrophilic substitution into tyrosine amino acids in antibodies and peptides [15]. Reaction of radioiodine with histidine, cysteine, methionine, phenylalanine, and tryptophan residues is possible but less likely. The most commonly used oxidizing agents are chloramine-T (N-chloro-4-methylbenzene sulfonamide) [16] and Iodogen (1,3,4,6-tetrachloro-3α, 6α-diphenylglycouril; Pierce Chemical Co. Rockford, IL) [17]. Chloramine-T provides higher radiolabeling yields than Iodogen (70–90% vs. 40–60%) but because it is a stronger oxidizing agent, it may damage mAbs, especially if the conditions are not carefully controlled. A typical chloramine-T-mediated radio-iodination involves diluting the antibody or peptide in a slightly alkaline buffer and incubating it with radio-iodine and 10–20 μg of chloramine-T in a glass tube for 30–60 seconds at room temperature [18]. The radio-iodination reaction is stopped by adding 20–40 μg of the reducing agent, sodium metabisulfite. Radio-iodinated antibodies and peptides can be purified from free radioiodide by size-exclusion chromatography (SEC); alternatively, peptides may be purified by reversed-phase chromatography. Iodogen is a water-insoluble oxidizing agent that is dissolved in chloroform; 10–20 μg are then aliquoted into a glass tube and the chloroform is evaporated using a gentle stream of nitrogen to leave a coating of Iodogen on the inside surface of the tube. The biomolecule contained in a suitable buffer and radioiodide are then added to the tube and incubated for 1–2 minutes at room temperature. The reaction is stopped simply by transferring the radio-iodination mixture to a chromatography column for purification, leaving the water-insoluble Iodogen remaining in the tube. Iodogen is a more gentle oxidizing agent than chloramine-T and is preferred for radio-iodinating mAbs and their fragments.

The main challenge for radio-iodination of mAbs and peptides is their instability *in vivo* to proteolysis, deiodination, and loss of radioactivity from tumor cells. It is widely recognized that radio-iodinated biomolecules are proteolytically degraded in cells to radio-iodotyrosine, which is efficiently exported from the cells by membrane amino acid transporters. Released radio-iodotyrosine is deiodinated by deiodinases found in tissues, and the free radio-iodine redistributes and accumulates in organs with sodium iodide symporter expression, particularly the thyroid, stomach, and salivary glands. For tumor imaging applications, these catabolic processes diminish the tumor signal and increase normal tissue uptake of radioactivity. Moreover, in radiotherapeutic applications, these processes diminish the radiation absorbed dose to the tumor and increase the dose to normal tissues, thus narrowing the therapeutic index. To address this issue, several new radio-iodination methods have been developed that retain the radioactive catabolites within cells; these methods are known as "residualizing" techniques.

Zalutsky and Narula [19] first reported a method for residualizing radio-iodine in cells in 1987. This group synthesized the N-succinimidyl ester of 3-(tri-n-

butylstannyl) benzoate (ATE), a precursor that could be radio-iodinated in anhydrous chloroform by iodo-destannylation using anhydrous t-butylhydroperoxide (TBHP) as an oxidant. Radio-iodinated ATE, termed N-succinimidyl-3-iodobenzoate (SIB; Figure 3), was purified on a silica gel Sep-Pak (Waters) and conjugated to the N-terminus or ε-amino groups of lysine residues on IgGs through its reactive N-succinimidyl ester side chain. In this method, radio-iodine is situated at a position on the aromatic ring of ATE that is not ortho- to a phenolic hydroxyl group (as in the case of radio-iodotyrosine); this renders the molecule resistant to *in vivo* deiodination. In mice, thyroid uptake of radioactivity was substantially decreased compared with IgGs labeled with radio-iodine using Iodogen. A paired-label experiment comparing the tumor and normal tissue uptake of an $F(ab')_2$ fragment of the OC125 mAb labeled with $^{125}I$ using SIB or $^{131}I$ using Iodogen in mice implanted with OVCAR-3 ovarian carcinoma xenografts, revealed that thyroid uptake of radioiodine was reduced 100-fold using ATE [20]. There was also more rapid elimination of radioactivity from normal tissues, and at 96 hours post-injection (p.i.), T/NT ratios were fourfold higher. Similar promising results were observed in mice implanted with s.c. D-54 MG glioblastoma xenografts administered anti-tenascin mAb 81C6 labeled with $^{125}I$ using ATE compared with labeling with $^{131}I$ using Iodogen [21]. Zalutsky et al. [22] extended this residualizing strategy with an alternative radio-iodination agent, N-succinimidyl 5-[$^{131}I$]iodo-3-pyridinecarboxylate (SIPC; Figure 3). SIPC was used to radio-iodinate 81C6 mAb IgG and the

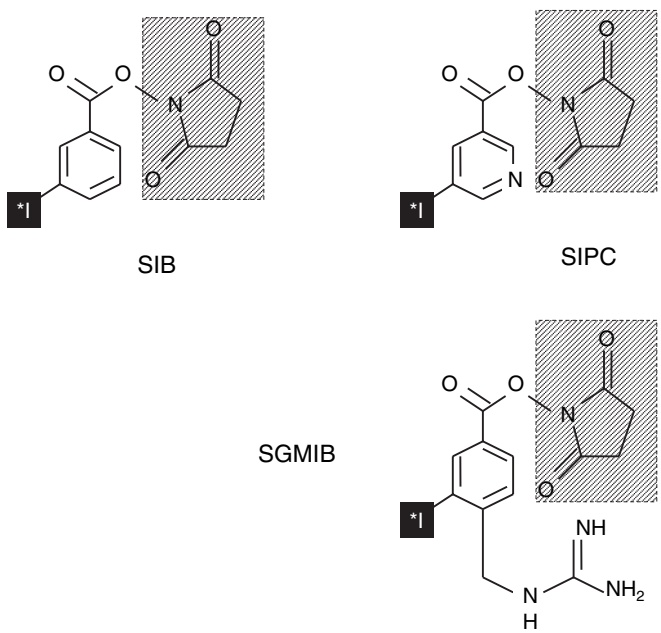


**Figure 3.** Precursors used for residualizing radio-iodination of monoclonal antibodies and peptides. Chemically reactive groups on the precursors are shaded, and site for radio-iodination of the precursor is indicated in black and with an asterisk. SIB: N-succinimidyl-3-iodobenzoate; SIPC: N-succinimidyl 5-iodo-3-pyridinecarboxylate; SGMIB: N-succinimidyl 4-guanidinomethyl-3-iodobenzoate.

$F(ab')_2$ fragment of anti-melanoma mAb Mel-14. It was hypothesized that SIPC, an iodopyridine, would be more dissimilar than SIB to iodotyrosine, and thus more resistant *in vivo* to de-iodination. The normal tissue distribution in mice of the 81C6 mAb and Mel-14 $F(ab')_2$ fragment radio-iodinated using SIPC or ATE were similar and there was very low thyroid uptake of radioactivity [<0.2–0.3% injected dose (% i.d.) at 7 days p.i.] using either reaagent. The tumor-specific anti-EGFRvIII mAb L8A4, which internalizes into receptor-positive cells and is highly susceptible to de-iodination, was radio-iodinated using SIPC resulting in improved tumor cell retention of radioactivity *in vitro* and increased accumulation in EGFRvIII positive tumor xenografts in mice providing enhanced T/NT ratios in comparison with L8A4 antibodies radio-iodinated using Iodogen [23, 24]. Radio-iodination using SIPC has also been applied to two 13-mer peptides: α-melanocyte-stimulating hormone (α-MSH) and its analog, [$Nle^4$, D-$Phe^7$]-α-MSH, resulting in preservation of receptor-binding properties *in vitro* and stability against de-iodination *in vivo* in mice [25]. However, conjugation of the peptides with SIPC increased their hydrophobicity. The catabolite of mAbs and peptides radio-iodinated with SIPC was found to be lysine-iodobenzoic acid (Lys-IBA).

One strategy that has been investigated to further enhance the retention of radio-iodine in cells is to use labeling techniques that generate a charged catabolite after intracellular proteolysis, which cannot traverse the lysosomal membrane and is thus resistant to exocytosis. Vaidyanathan et al. [26] synthesized N-succinimidyl 4-guanidinomethyl-3-[$^{131}$I]iodobenzoate (SGMIB; Figure 3), a radio-iodinating reagent that was expected to generate a positively charged catabolite at the acidic pH in lysosomes. They used this reagent to radio-iodinate anti-EGFRvIII mAb L8A4. There was three to fourfold greater retention of radioactivity in receptor-positive U87MG glioblastoma cells when these antibodies were radio-iodinated using SGMIB than when Iodogen was used. Analysis confirmed that the final catabolites were cationic. SGMIB has advantages over SIPC as a residualizing radio-iodination reagent, in that a twofold improvement in tumor retention of radioactivity was observed for L8A4 mAbs in mice implanted s.c. with D-256 glioblastoma xenografts expressing EGFRvIII. Again, similar to the SIPC reagent, thyroid radioactivity was very low (<0.35% i.d.) [27]. An analogous approach was reported by Shankar et al. [28] who used N-succinimidyl 3-[$^{131}$I]iodo-4-phosphonomethylbenzoate ([$^{131}$I]SIPMB) to radio-iodinate mAb L8A4; this reagent gen-erates an anionic catabolite that is retained within cells.

A different strategy for residualizing radio-iodine in cells uses radio-iodinated diethylenetriaminepentaacetic acid (DTPA)-containing peptides composed of one or more D-amino acids including D-tyrosine that are conjugated through a maleimide functional group to chemically reduced IgG mAbs [29]. Two such peptides: R–Gly–D–Tyr–D–Lys[1-(p-thiocarbonylaminobenzyl)DTPA], termed IMP-R1, and [R–D–Ala–D–Tyr–D–Tyr–D–Lys]$_2$(CA-DTPA), termed IMP-R2, were described by Govindan et al. [29]. The BOC-protected peptides were radio-iodinated at the D-tyrosine amino acid using chloramine-T and derivatized at their N-terminus with sulfo-SMCC to introduce maleimide groups. After deprotection, the maleimide-derivatized and radio-iodinated peptides were conjugated to thiol groups generated by dithiothreitol (DTT) reduction of some disulfide linkages on the anti-CD20 mAb LL2 or the anti-EGP-1 mAb RS7. The premise of including DTPA in the peptides was that as it had been shown that mAbs labeled with $^{111}$In

through introduction of a DTPA metal chelator (see section 3) were catabolized to $^{111}$In–DTPA–lysine, which was retained within the cells [30], it would be expected that radio-iodinated peptides containing this group would be trapped. The DTPA groups in the peptides are not used for radiolabeling. D-amino acids were included because these are more resistant to proteolysis than L-amino acids, and in particular, D-iodotyrosine is more resistant to *in vivo* de-iodination than L-tyrosine [31]. The immunoreactivity of the RS7 antibody was maintained using this labeling strategy (radio-iodinated LL2 was not tested for immunoreactivity), and both antibodies exhibited greater retention of radioactivity in tumor cells *in vitro* than antibodies radio-iodinated using chloramine-T. A series of radio-iodinated DTPA–D-amino acid peptides (IMP-R1 to IMP-R8) that differed in their hydrophobicity and charge was synthesized and conjugated to a DTT-reduced RS7 mAb [32]. These peptides varied in their DTPA content and extent of maleimide derivatization. Increasing the maleimide substitution in the peptides from one to two functional groups per molecule increased the conjugation efficiency with the R27 mAb from 30% to >80%. In mice implanted s.c. with Calu-3 lung carcinoma xenografts, tumor radioactivity was enhanced for mAb RS7 radio-iodinated using any of the DTPA–D-amino acid peptides compared with chloramine-T radioiodinated mAbs. However, normal tissue uptake was highest for the most hydrophobic peptides. IMP-R4: MCC–Lys(MCC)–Lys(1–((p-CSNH)benzyl)DTPA)–D–Tyr–D–Lys(3–((p-CSNH)benzyl)DTPA)–OH, where MCC represents the maleimide groups, provided the greatest retention of radio-iodine in tumors and the lowest normal tissue accumulation. This approach has been optimized using a one-vial kit labeling procedure under GMP conditions to yield at least 100 mCi of highly pure (>95%) $^{131}$I-labeled humanized anti-CEA mAb hMN-14 that exhibits preserved immunoreactivity (>95%) for targeted *in situ* radiotherapy of malignancies [33]. A similar strategy was recently described by Foulon et al. [34]. This group used a polycationic peptide composed of D-amino acids: D–Lys–D–Arg–D–Tyr–D–Arg–D–Arg to radio-iodinate anti-EGFRvIII mAb L8A4. These peptides were first radio-iodinated using Iodogen and then conjugated in 60% overall yield via a maleimide group to mAb L8A4, which was thiolated using 2-iminothiolane. Paired label experiments in mice implanted s.c. with U87 glioblastoma xenografts expressing EGFRvIII and administered mAb L8A4 labeled with $^{125}$I using this approach showed up to a fivefold higher tumor accumulation compared with L8A4 labeled directly with $^{131}$I using Iodogen. However, use of the peptides for radio-iodinating L8A4 increased accumulation in the kidneys, perhaps because of binding of the positively charged catabolites to renal tubular cells [34].

Radio-iodine may also be residualized in cells by substitution onto an aromatic residue that is linked to a carbohydrate moiety attached to the biomolecule [35]. A sevenfold increase in tumor retention of radioactivity was observed in mice implanted with Calu-3 lung carcinoma xenografts at 7 days p.i. of radio-iodinated mAb RS7 using a dilactitol-tyramine (DLT) carbohydrate adduct compared with chloramine-T radio-iodination (38.0% vs. 5.5% i.d./g, respectively) [36]. Similar results were achieved using tyramine cellobiose (TCB)-radio-iodinated anti-EGFRvIII mAb L8A4 in mice bearing receptor-positive tumor xenografts [37]. However, a limitation of this approach seems to be the higher liver and spleen uptake of radioactivity of mAbs radio-iodinated using TCB [38]. Epidermal growth factor (EGF), a 53-amino acid peptide ligand for the EGFR present on many

epithelial malignancies, has been labeled using an analogous approach by radio-iodinating a tyrosine-modified dextran and then conjugating this dextran molecule to EGF [39]. Glioblastoma cells expressing EGFR showed significantly greater retention of radioactivity when incubated with dextran-EGF in which the radio-iodine was present on the dextran moiety compared with EGF or dextran-EGF where the radio-iodine was substituted into tyrosine amino acids on the EGF molecule itself.

Since the advent of high-resolution PET/computed tomography (CT) [40] and small animal PET tomographs [41] and building on the success of $^{18}$F-fluorodeoxyglucose ($^{18}$F-FDG) for PET of malignancies [42], there has been a growing interest in labeling mAbs and peptides with positron-emitters. $^{124}$I is a positron-emitter with a sufficiently long half-life of 4.2 days that is feasible for labeling mAbs and peptides for PET. Nevertheless, $^{124}$I has limitations compared with $^{18}$F. These limitations include its lower abundance of positron decay compared with $^{18}$F (23% versus 100%); the higher $\beta^+$ energies for $^{124}$I compared with $^{18}$F (0.8–2.1 MeV vs. 0.6 MeV), which result in poorer spatial resolution (2.3 mm vs. 0.7 mm); and the emission of high-energy $\gamma$-photons by $^{124}$I that degrade the PET image and contribute along with the higher positron energy and longer half-life to higher radiation absorbed doses [13]. Nevertheless, intact IgG mAbs [43–46], genetically engineered antibody fragments (e.g., minibodies and diabodies) [47, 48], and peptides [49] have been labeled with $^{124}$I, in most cases using chloramine-T or Iodogen. It is important to add a small amount of ascorbic acid to the formulation after labeling biomolecules with $^{124}$I in order to protect them from radiolytic decomposition caused by the high-energy positrons [46]. $^{124}$I-labeled antibodies have been successfully used for PET imaging of receptor/antigen-positive tumors in mouse xenograft models [43, 45–48] and in one study in a child with neuroblastoma [44]. One advantage of PET using $^{124}$I-labeled antibodies is that it allows accurate quantification of radioactivity uptake in tumors and normal tissues. This could be useful for predicting the radiation absorbed doses for subsequent RIT using the corresponding $^{131}$I-labeled antibodies [47, 50].

### 5.2 Bromine Radionuclides

Although there has been interest in labeling mAbs and peptides with $^{124}$I for PET, another attractive positron-emitter is bromine-76 ($^{76}$Br). $^{76}$Br decays with a half-life of 16.2 hours, emitting positrons in 54% abundance with energy of 3.4 MeV. The almost sixfold higher positron energy of $^{76}$Br compared with that of $^{18}$F (0.6 MeV) provides poorer spatial resolution (>5 mm vs. 0.7 mm, respectively) and a higher radiation absorbed dose [13]. Nevertheless, $^{76}$Br is more practical for labeling biomolecules for PET than $^{124}$I due to its greater abundance of positron emissions, and because it can be produced using low-energy biomedical cyclotrons by the $^{76}$Se(p,n)$^{76}$Br reaction [51]. Chloramine-T was initially studied for radiobromination of mAbs and peptides using $^{76}$Br. The weaker oxidizing agent, Iodogen, does not seem to be useful for $^{76}$Br labeling because of the greater resistance to oxidation compared with radio-iodine [52]. The colon cancer mAbs A33, 3S193, and 38S1 as well as EGF have been labeled with $^{76}$Br using chloramine-T in yields of 63–77% and with good preservation of antigen/receptor-binding characteristics [53]. Nevertheless, in one study, the immunoreactivity of $^{76}$Br-mAb 38S1 was diminished to

45% and only radiobromination using bromoperoxidase was found to generate radiolabeled antibodies with preserved immunoreactivity [52]. Human colon cancer xenografts implanted into nude rats have been visualized by PET using $^{76}$Br-mAb 38S1 [54, 55].

Despite the ability to directly radiobrominate some antibodies using chloramine-T, many investigators obtained inconsistent and low yields and as mentioned, in some cases, decreased immunoreactivity using this approach. Therefore, newer strategies were explored for labeling mAbs with $^{76}$Br using N-succinimidyl para-[$^{76}$Br]bromobenzoate ($^{76}$Br-SPBrB; Figure 4) [56]. $^{76}$Br-SPBrB was generated by bromodestannylation of N-succinimidyl-para-tri-n-butylstannylbenzoate (SPMB) using chloramine-T [56, 57]. The resulting $^{76}$Br-SPBrB was purified by high-performance liquid chromatography (HPLC) and conjugated through the activated N-succinimidyl moiety to ε-amino groups of lysines on the antibodies. Using this approach, mAb 38S1 was labeled with $^{76}$Br in 49% yield and with good preservation of immunoreactivity (69–76%) [56]. Building on the strategy described by Vaidyanathan et al. [26] to radio-iodinate antibodies using reagents (e.g., SGMIB) that yield positively charged catabolites that are retained in cells, Mume et al. [58] synthesized N-succinimidyl 5-[$^{76}$Br]bromo-3-pyridinecarboxylate (Figure 4) by bromodestannylation of N-succinimidyl-5-(tributylstannyl)3-pyridinecarboxylate and conjugated it to the HER-2/neu mAb trastuzumab (Herceptin). The labeling yield was 45%, but after purification, the immunoreactivity with HER2/neu positive SKOV-3 ovarian carcinoma cells was >75%. However, this method of radiobromination did not improve cellular retention of radioactivity compared with trastuzumab labeled with $^{76}$Br–SPBrB. A site-specific radiobromination technique was recently reported for labeling the affibody molecule $(Z_{HER2\text{-}4})_2$–Cys with $^{76}$Br at a cysteine residue [59]. $(Z_{HER2\text{-}4})_2$–Cys affibodies are small recombinant proteins [$M_r$ = 7 kDa (monomer) or 15 kDa (dimer)] that bind with high specificity to HER2/neu receptors, similar to antibody fragments but produced by phage display techniques [60]. A bifunctional reagent, ((4-hydroxy-phenyl)ethyl)maleimide (HPEM; Figure 4), reactive with thiols was synthesized and radiobrominated with $^{76}$Br using chloramine-T. The radiobrominated HPEM was conjugated through its maleimide group to the free thiol on the cysteine residue of the $(Z_{HER2\text{-}4})_2$–Cys affibodies. $^{76}$Br–$(Z_{HER2\text{-}4})_2$–Cys exhibited preserved binding to HER2/neu-positive SKOV-3 cells *in vitro* and achieved high tumor uptake (5% i.d./g) and T/NT ratios (up to 31 : 1) at 4 hours p.i. *in vivo* in mice implanted with SKOV-3 tumor xenografts.

Finally, a totally new approach to labeling antibodies with $^{76}$Br employed derivatives of polyhedral boron clusters such as *closo*-dodecaborate ($2^-$) or *nido*-undecaborate ($1^-$) anions (Figure 4) [61, 62]. These structures form strong boron–bromine bonds and were labeled with $^{76}$Br using chloramine-T. Once labeled with $^{76}$Br, they were conjugated through their benzylisothiocyanato side chain to ε-amino groups of lysine residues on mAbs. The idea was that, as $^{76}$Br is attached through a bromine–boron bond to a charged molecular structure that is completely foreign to the body, enzymatic debromination and exocytosis of radioactivity from cells would be substantially reduced. Bruskin et al. [61] labeled trastuzumab (Herceptin) with $^{76}$Br in >80% yield using this approach, resulting in a preparation that was immunoreactive with HER2/neu-positive SKBR-3 human breast cancer cells. Trastuzumab has also been labeled with $^{76}$Br using the related *nido*-undecaborate ($1^-$) anion [62].

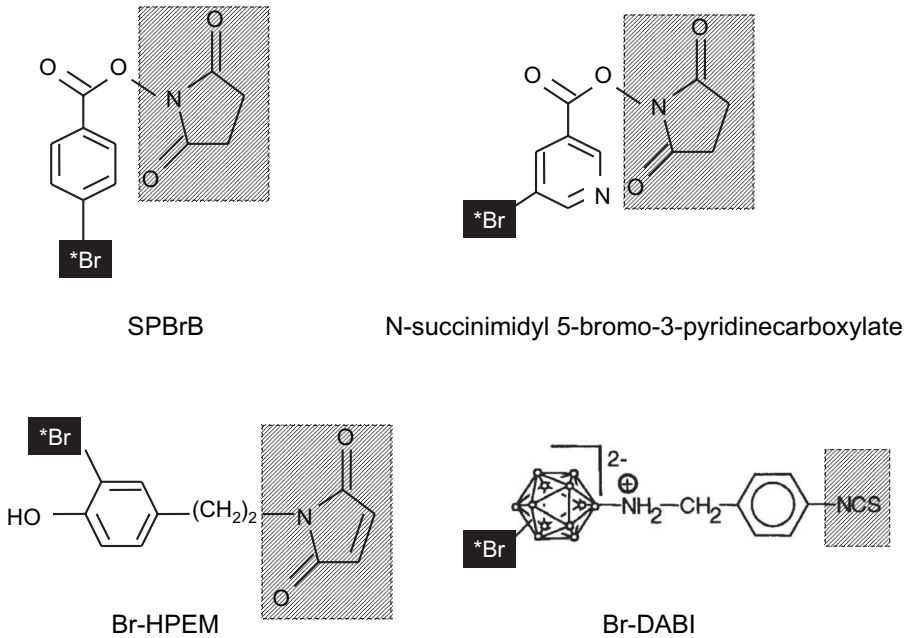


**Figure 4.** Precursors used for residualizing radiobromination of monoclonal antibodies and peptides. Chemically reactive groups on the precursors are shaded, and site for radiobromination of the precursor is indicated in black and with an asterisk. SPBrB: N-succinimidyl para-bromobenzoate; Br-HPEM: bromo-((4-hydroxy-phenyl)ethyl) maleimide; Br-DABI: bromo-(4 isothiocyanatobenzyl-ammonio)-bromo-decahydro-closo-dodecaborate.

### 5.3 Fluorine Radionuclides

Fluorine-18 ($^{18}$F) is the most widely used radionuclide for PET of tumors, usually in the form of $^{18}$F-FDG [42]. Due to its relatively low positron energy (0.6 MeV), $^{18}$F provides excellent spatial resolution (0.7 mm); it also is associated with lower radiation absorbed doses compared with $^{124}$I or $^{76}$Br [13]. There has been considerable interest in labeling mAb fragments and peptides with $^{18}$F [63]. The pharmacokinetics of intact IgG mAbs, in particular their slow elimination from the blood, do not lend themselves to the use of $^{18}$F, due to its short half-life of 110 minutes. In contrast, antibody fragments (e.g., F(ab')$_2$, Fab, and scFv) and peptides are accumulated rapidly in tumors and eliminated quickly from the blood and most normal tissues. Antibody fragments and peptides labeled with $^{18}$F could therefore be useful for PET. The principal challenge in labeling biomolecules with $^{18}$F is that nucleophilic substitution reactions are required; direct electrophilic radiofluorination of biomolecules requires carrier fluoride that yields low specific activity that is unsuitable for imaging epitopes/receptors on tumor cells [63]. To solve this problem, biomolecules are labeled with $^{18}$F by first substituting $^{18}$F onto a prosthetic group and then chemically linking this group through a reactive side chain to the biomolecules [63]. One primary consideration in designing methods for labeling antibody fragments and peptides with $^{18}$F is the amount of time required to complete the labeling and conjugation procedures due to the short 110-minute half-life of $^{18}$F. It

is desirable that the labeling procedure, including quality control testing, be completed in 1–1.5 hours, to minimize losses in yield, simply due to the physical decay of $^{18}$F.

Garg et al. [64] described an approach for labeling F(ab')$_2$ and Fab fragments of the antimyosin antibody R11D10 using 4-[$^{18}$F]fluorobenzylamine succinimidyl ester ([$^{18}$F]SFBS; Figure 5). Although this method was rapid with a total labeling time of 1.5 hours, synthesis of [$^{18}$F]SFBS required three separate reactions followed by HPLC purification of the reagent; the yield of [$^{18}$F]SFBS was 25–40%. Conjugation of [$^{18}$F]SFBS to antibody fragments through reaction of the succinimidyl ester with ε-amino groups on lysines was performed in a fourth step, which required purification of the $^{18}$F-labeled fragments from excess [$^{18}$F]SFBS by SEC. The immunoreactivity of the $^{18}$F-labeled F(ab')$_2$ and Fab fragments of R11D10 were relatively preserved (89% and 75%, respectively). Using [$^{18}$F]SFBS as a prosthetic group, Garg et al. [65] labeled F(ab')$_2$ fragments of the anti-glioma mAb Mel-14 and obtained high T/NT ratios (up to 14:1 at 4 hours p.i.) in a s.c. glioblastoma mouse tumor xenograft model. They also labeled Fab fragments of the TP-3 antibody with [$^{18}$F]SFBS, which permitted PET of osteosarcoma tumors in dogs [66]. An analogous approach for labeling biomolecules with $^{18}$F uses the prosthetic group, [$^{18}$F]-N-succinimidyl 4-(fluoromethyl)benzoate ([$^{18}$F]SFB; Figure 5) [67].

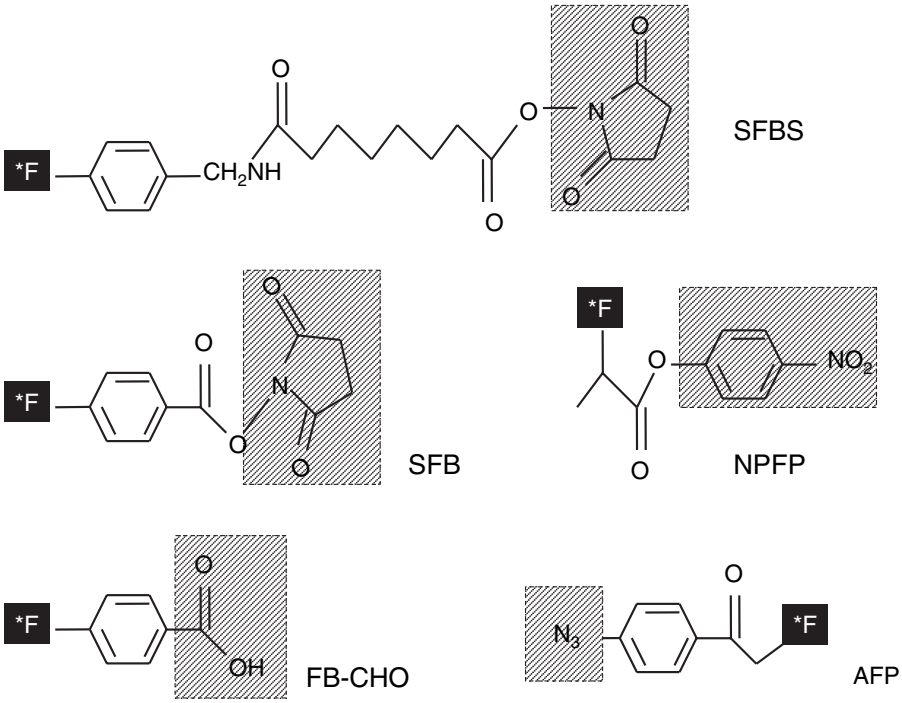


**Figure 5.** Prosthetic precursors used for radiofluorination of monoclonal antibody fragments and peptides. Chemically reactive groups on the precursors are shaded, and site for radiofluorination of the precursor is indicated in black and with an asterisk. SFBS: 4-fluorobenzylamine succinimidyl ester; SFB: N-succinimidyl 4-(fluoromethyl)benzoate; NPFP: 4-nitrophenyl 2-fluoropropionate; FB-CHO: 4-fluorobenzaldehyde; AFP: 4-azidofluorophenacyl.

The synthesis of [$^{18}$F]SFB was optimized [68], and the reagent was used for labeling F(ab')$_2$ fragments of the anti-glioma mAb Mel-14. The immunoreactivity of $^{18}$F-labeled Mel-14 F(ab')$_2$ fragments using [$^{18}$F]SFBS or [$^{18}$F]SFB was indistinguishable (65% vs. 64%, respectively), and the radiolabeled fragments demonstrated similar tissue distribution and pharmacokinetics in dogs [69]. The 13-amino acid α-MSH peptide analog, N-acetyl-Ser-Tyr-Ser-NorLeu-Glu-His-D-Phe-Arg-Trp-Gly-Lys-Pro-Val-NH$_2$, has been labeled with [$^{18}$F]SFB with good preservation of receptor-binding affinity [70].

Another reagent that has been used for labeling biomolecules with $^{18}$F is 4-nitrophenyl 2-[$^{18}$F]fluoropropionate ([$^{18}$F]NPFP; Figure 5) [71]. [$^{18}$F]NPFP is a smaller prosthetic group that may have less impact on the receptor-binding properties and/or physico-chemical properties of peptides than [$^{18}$F]SFBS or [$^{18}$F]SFB. Accordingly, [$^{18}$F]NPFP was used for labeling the octreotide peptide analog, SDZ 223-228, with $^{18}$F resulting in full retention of biological activity [72]. It was necessary to Boc-protect the ε-amino group on lysine-5 in SDZ 223-228 during labeling with [$^{18}$F]NPFP and then deprotect afterward, because this residue is critical for SMSR-binding. Nevertheless, a radiofluorination technique using 4-[$^{18}$F]fluorobenzaldehyde ([$^{18}$F]FB–CHO; Figure 5) that allows site-specific labeling of aminoxyacetic acid functionalized octreotide analogs without the need to protect/deprotect the ε-amino group on lysine-5 has been recently reported [73]. This method allowed PET imaging of s.c. transplantable AR42J rat pancreatic tumors in athymic mice at 1 hour p.i. of the $^{18}$F-labeled octreotide derivatives. In another study, D–Phe$^1$–octreotide labeled with $^{18}$F using [$^{18}$F]NPFP exhibited higher liver accumulation and hepatobiliary clearance of radioactivity than the corresponding radio-iodinated derivative in Lewis rats bearing pancreatic islet cell tumors, suggesting that [$^{18}$F]NPFP may increase the hydrophobicity of peptides [74]. Nevertheless, T/B ratios were 5.2 : 1 at 1 hour p.i. and 4.2 : 1 at 2 hours p.i., due to the rapid blood clearance of the $^{18}$F-D-Phe$^1$-octreotide. [$^{18}$F]NPFP and [$^{18}$F]SFB have also been employed for $^{18}$F labeling of peptides that contain the arginine-glycine-aspartic acid (RGD) sequence that binds to the αvβ3 integrin implicated in tumor angiogenesis [75, 76]. In a study comparing labeling of proteins with [$^{18}$F]NPFP, [$^{18}$F]SFB, and another fluorinating reagent, 4-azidofluorophenacyl-[$^{18}$F]AFP (Figure 5), it was found that [$^{18}$F]NPFP-conjugated human serum albumin was partially unstable under slightly basic conditions [77]. It was concluded that [$^{18}$F]SFB may be the most suitable radiofluorinating agent. Finally, [$^{18}$F]FB-CHO (Figure 5) has been used to site-specifically label human serum albumin with $^{18}$F at hydrazinenicotinamide (HYNIC) functional groups introduced into the protein using HYNIC N-hydroxysuccinimide ester [78]. This method has similarities with the use of HYNIC for labeling biomolecules with $^{99m}$Tc (section 6.1).

### 5.4 Astatine Radionuclides

Astatine-211 ($^{211}$At; Table 3) is an α-particle emitter with a half-life of 7.2 hours that is useful for labeling mAbs and peptides for targeted *in situ* radiotherapy of malignancies [2, 79]. $^{211}$At is produced in a cyclotron using the $^{209}$Bi(α,2n)$^{211}$At reaction. However, the astatine–carbon bond is unstable [80] and direct electrophilic astatination of biomolecules at tyrosine residues using oxidizing agents such as chloramine-T or hydrogen peroxide results in low yield [81] and susceptibility *in vivo* to de-astatination [82]. Zalutsky and Narula [83] addressed this problem

by labeling N-succinimidyl 3-(tri-n-butylstannyl) benzoate (ATE) with $^{211}$At. The N-succinimidyl 3-[$^{211}$At]astatobenzoate (SAB; Figure 6) was conjugated to ε-amino groups on lysine residues in antibodies, which is similar to the strategy used for radio-iodination. Using SAB, intact mAb 81C6 IgG and F(ab')$_2$ fragments of mAb Mel-14 directed against glioblastoma were labeled with $^{211}$At with preservation of immunoreactivity *in vitro* and excellent targeting *in vivo* to s.c. D-54 MG human glioblastoma xenografts in mice [84]. Due to its short half-life (7.2 hours) and the short range of the α-particles (50–100 μm), F(ab')$_2$ fragments are more suitable for labeling with $^{211}$At than intact mAb IgG's because they penetrate deeper into tumors and are cleared more quickly from the blood [7]. Despite their more rapid blood clearance than F(ab')$_2$, Fab fragments are not appropriate for labeling with $^{211}$At because they accumulate to high levels in the kidneys, thus posing a potential radiotoxicity hazard. Paired label experiments in normal mice of the anti-CEA mAb C110 or its F(ab')$_2$ fragment labeled with $^{211}$At or $^{131}$I using the ATE method revealed that there was greater tissue retention of $^{211}$At than for $^{131}$I [85]. In another study, mice implanted s.c. with TK-82 human renal cell carcinoma xenografts showed similar tumor and normal tissue uptake after administration of $^{211}$At or $^{125}$I-labeled A6H F(ab')$_2$ [86]. Tumor uptake in this mouse xenograft model was 30% i.d./g for $^{211}$At-A6H F(ab')$_2$ versus 19% i.d./g for $^{125}$I-A6H F(ab')$_2$ at 19 hours p.i. T/B ratios were 3:1 for both $^{211}$At and $^{125}$I. $^{211}$At-labeled anti-tenascin mAb 81C6 has been produced using the SAB reagent in sufficient quantities (2–10 mCi) under GMP conditions for use in a phase I clinical trial in glioblastoma patients [87]. Radiolytic decomposition of SAB due to the α-particles emitted by $^{211}$At and subsequent low radiolabeling yields is nevertheless a major challenge in producing clinical quality $^{211}$At-labeled mAbs [88, 89]. It seems that radiolytic decomposition is especially problematic when chloroform is used as a solvent for synthesizing SAB; benzene and methanol are alternatives that are less susceptible to the radiolytic effects of $^{211}$At.

Reist et al. [90] extended their previous residualizing radio-iodination approach using N-succinimidyl 5-[$^{131}$I]iodo-3-pyridinecarboxylate (SIPC) to $^{211}$At. Anti-EGFRvIII mAb L8A4 was labeled by conjugation with N-succinimidyl 5-[$^{211}$At]astato-3-pyridinecarboxylate (SAPC; Figure 6). Again, the premise is

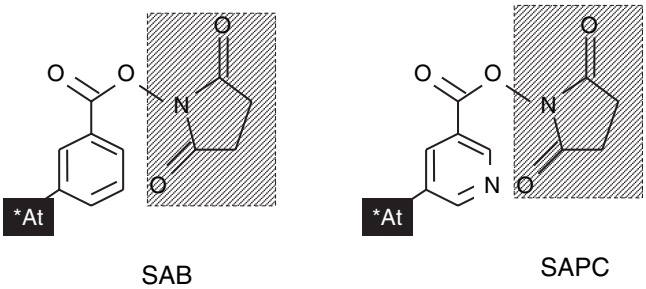


**Figure 6.** Precursors used for residualizing radio-astatination of monoclonal antibodies and peptides. Chemically reactive groups on the precursors are shaded, and site for radio-astatination of the precursor is indicated in black and with an asterisk. SAB: N-succinimidyl 3-astatobenzoate; SAPC: N-succinimidyl 5-astato-3-pyridinecarboxylate.

that proteolysis of the $^{211}$At-labeled antibodies would yield a positively charged $^{211}$At catabolite that would be unable to traverse the lysosomal membrane or be exocytosed. Immunoreactivity was maintained, and $^{211}$At- and $^{131}$I-labeled L18A4 showed similar tumor and normal tissue accumulation in mice implanted s.c. with EGFRvIII-positive U87MG glioblastoma xenografts. Other biomolecules that have been labeled with $^{211}$At include octreotide [91] and the anti-CD20 mAb rituximab (Rituxan; Roche Pharmaceuticals Ltd.) directed against non-Hodgkin's B-cell lymphomas [92].

## 6 LABELING MABS AND PEPTIDES WITH RADIOMETALS

### 6.1 Technetium Radionuclides

Technetium-99m ($^{99m}$Tc; Table 1) is the most widely available, least expensive, and most commonly used radionuclide in nuclear medicine and is thus an attractive candidate for labeling mAbs, their fragments, as well as peptides for tumor imaging. $^{99m}$Tc is a metastable form of $^{99}$Tc, which decays by internal conversion to its ground state ($^{99}$Tc) with a half-life of 6 hours, emitting a γ-photon of 140 keV that is easily imaged by gamma cameras available in all nuclear medicine facilities. $^{99m}$Tc is produced from the decay of molybdenum-99 ($^{99}$Mo) using a commercially available $^{99}$Mo/$^{99m}$Tc generator at very low cost (less than $0.50 per mCi). Its relatively short half-life minimizes the radiation exposure to patients undergoing imaging procedures, but it necessitates rapid labeling procedures for biomolecules. There are several approaches to labeling antibodies and peptides with $^{99m}$Tc [93, 94]. These approaches include (1) direct methods that rely on the binding of $^{99m}$Tc to endogenous thiols generated by reduction of disulfide linkages in mAbs or introduced chemically by reaction with thiolating agents such as 2-iminothiolane, and (2) indirect methods that involve binding of $^{99m}$Tc to a chelating agent that is then conjugated to the mAb or peptide ("pre-formed chelator" approach) or to a chelator already incorporated into the biomolecule. There are advantages and disadvantages of each of these strategies with respect to ease of use, *in vivo* stability, and amenability to kit formulation.

Intact mAb IgGs can be labeled simply and directly with $^{99m}$Tc by reduction of a small proportion (<5%) of the inter- or intrachain disulfide bonds to free thiols by treatment with a 2000-fold molar excess of 2-mercaptoethanol (2-ME) for 30 minutes (Figure 7) [94]. The reduced IgGs are purified from excess 2-ME by SEC and labeled to high efficiency (>90%) by transchelation of $^{99m}$Tc from $^{99m}$Tc-glucoheptonate or $^{99m}$Tc-methylene diphosphonate ($^{99m}$Tc–MDP). $^{99m}$Tc-glucoheptonate or $^{99m}$Tc–MDP are prepared from commercial kits and $^{99m}$Tc pertechnetate ($^{99m}TcO_4^-$) eluted from a $^{99}$Mo/$^{99m}$Tc generator. This method of labeling mAbs with $^{99m}$Tc was introduced by Schwartz and Steinstrasser in 1987 [95] and later optimized by Mather and Ellison in 1990 [96]. A recent modification of the Schwartz method employed exposure of the anti-CD20 mAb rituximab (MabThera; Roche) to ultraviolet (UV) light of wavelength 320 nm for 20 minutes to photoreduce some disulfide linkages to thiols, which were then labeled wth $^{99m}$Tc by transchelation from $^{99m}$Tc–MDP [97]. This photoreduction method was originally described by Stalteri and Mather in 1996 [98]; it does not require removal of any chemical reducing agents and is somewhat simpler and potentially more controllable than chemical

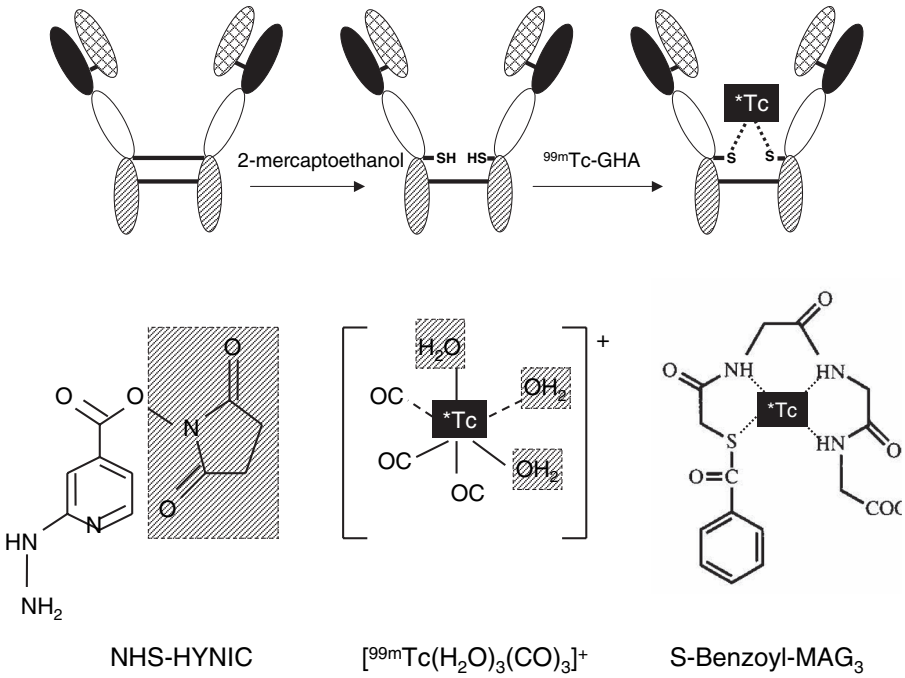


**Figure 7.** *Top Panel*: Method for direct labeling of intact monoclonal antibodies with $^{99m}$Tc. Some disulfide linkages between the antibody chains are chemically reduced with 2-mercaptoethanol. The free thiols produced are sites for binding $^{99m}$Tc-glucoheptonate ($^{99m}$Tc-GHA). *Bottom Panel*: Metal chelators that can be conjugated to monoclonal antibodies or peptides for labeling with $^{99m}$Tc. Chemically reactive groups on the precursors are shaded, and complexation of $^{99m}$Tc is indicated in black and with an asterisk. NHS-HYNIC: N-hydroxysuccinimide ester of hydrazinenicotinamide; $[^{99m}Tc(H_2O)_3(CO)_3]^+$: $^{99m}$Tc(I)-carbonyl complex; S-benzoyl-MAG$_3$: S-benzoyl-mercaptoacetyl-glycyl-glycyl-glycine.

reduction with 2-ME. The advantage of the Schwartz technique is that purified reduced IgGs can be dispensed into vials and stably maintained either frozen or lyophilized until required for labeling with $^{99m}$Tc, thus generating a kit formulation [99]. The major disadvantage of the method is that if too great an amount of reducing agent is used, too many disulfide linkages are reduced, increasing the risk for disrupting key linkages needed to maintain protein folding, integrity, and immunoreactivity. This is especially a problem for mAb fragments such as F(ab')$_2$, Fab, and scFv. Likewise the method cannot be used for labeling small peptides with $^{99m}$Tc that harbor disulfide bonds required for maintaining biological activity (e.g., somatostatin analogs). Careful control and optimization of reduction conditions through measurement of the number of free thiols generated using Ellman's reagent [100, 101] is needed to successfully apply this method for labeling biomolecules with $^{99m}$Tc. Another disadvantage is that IgGs labeled with $^{99m}$Tc through direct binding to thiols are subject to loss of the radiolabel *in vivo* over time by exchange with endogenous thiol containing molecules such as cysteine and glutathione [94]. Nevertheless, the Schwartz method has been used for labeling several mAbs with $^{99m}$Tc providing preserved immunoreactivity and the ability to target and image human cancers in mouse xenograft models [102, 103] and in patients [104, 105].

Due to the limitations of the Schwartz method for labeling mAb fragments (e.g., $F(ab')_2$ or Fab) and peptides with $^{99m}Tc$, alternative strategies have been investigated. One method uses hydrazinenicotinamide (HYNIC) and coligands such as ethylenediaminodiacetic acid (EDDA), ethlenediaminetetraacetic acid (EDTA), tricine, or glucoheptonate to form a stable $^{99m}Tc$ complex with the biomolecule (Figure 7) [106]. This method was described by Abrams et al. in 1990 [107]. HYNIC is conjugated to the mAb (or its fragments) or peptides using an excess of HYNIC N-hydroxysuccinimide ester (or other chemically reactive form), which forms an amide linkage with the N-terminus or ε-amino groups in lysines. After chromatographic purification, the HYNIC-derivatized biomolecule is labeled with $^{99m}Tc$ by incubation with $^{99m}TcO_4^-$ in the presence of $SnCl_2$ and coligand. Coligands are needed because HYNIC occupies only one or two of the six coordination sites of $^{99m}Tc$ [106]. $SnCl_2$ is included to reduce $^{99m}Tc$ from its $7^+$ valence state in $^{99m}TcO_4^-$ to a valence of $4^+$ or $5^+$ for complexation with HYNIC-coligands. Our group has labeled Fab fragments of the HER2/neu mAb trastuzumab (Herceptin) with $^{99m}Tc$ using HYNIC [108]. $^{99m}Tc$-HYNIC-trastuzumab Fab showed preserved binding affinity for HER2/neu receptors on SKBR-3 human breast cancer cells *in vitro* ($K_d = 1.6 \times 10^{-8} M^{-1}$) and demonstrated good tumor targeting *in vivo* in mice implanted s.c. with HER2/neu-positive BT-474 breast cancer xenografts (T/B ratio = 3:1 at 24 hours p.i.). BT-474 tumors were visualized by imaging. Others have labeled mAb Fab fragments with $^{99m}Tc$ using HYNIC achieving preserved immunoreactivity and good tumor targeting [109, 110]. As mentioned, the Schwartz method is not amenable to labeling peptides that harbor key intramolecular disulfide linkages. However, the somatostatin analog, D-Phe$^1$-Tyr$^3$-octreotide (TOC), which contains a disulfide bond essential for somatostatin receptor (SMSR)-binding, has been labeled with $^{99m}Tc$ by conjugation to HYNIC using a variety of coligands [111–113]. Similarly, the 53-amino-acid peptide, EGF, which harbors three disufide bonds required for maintenance of its tertiary structure and receptor-binding properties has been labeled with $^{99m}Tc$ using HYNIC and tricine as a coligand [114]. In each case, receptor-binding was preserved *in vitro*, allowing tumor imaging in mouse xenograft models [113, 114] or in patients [111, 112].

An interesting finding is that the choice of coligand for HYNIC drastically affects the plasma protein binding, elimination from the blood, and tissue distribution of the $^{99m}Tc$-labeled biomolecules. This effect was first reported by Babich and Fischman in 1995 [115] who noted differences in radioactivity accumulation in the lungs, kidneys, liver, spleen, and gastrointestinal tract of rats for the chemotactic peptide (N–For–Met–Leu–Phe–Lys) conjugated to HYNIC and labeled with $^{99m}Tc$ using glucarate, glucoheptonate, mannitol, or glucamine as coligands. Use of EDDA as a coligand instead of tricine for $^{99m}Tc$ labeling of the somastatin analogs HYNIC–TOC or RC160 produced a more hydrophilic complex that cleared more quickly from the blood in rats and favored renal over hepatobiliary clearance [116]. However, substitution of the backbone of EDDA with dimethyl, diethyl, or dibenzyl moieties increased the hydrophobicity of $^{99m}Tc$–HYNIC–RC160, resulting in a fivefold to sixfold increased plasma protein binding *in vitro* and substantially increased liver uptake and hepatobiliary elimination in rats [117]. Tricine as a coligand for $^{99m}Tc$–HYNIC–RC160 produced liver and intestinal radioactivity uptake intermediate between that of the EDDA and substituted EDDA derivatives. HYNIC–TOC labeled with $^{99m}Tc$ using EDDA as a coligand showed excellent

tumor localization in a study of 10 patients with SMSR-positive tumors allowing imaging of lesions at 4 hours p.i. [112]. Similarly, in a study of 13 patients, 12 of whom had SMSR-positive malignancies, administered $^{99m}$Tc–HYNIC–TOC labeled using tricine as a coligand, tumors were imaged as early as 10 minutes p.i. [111]. A comparison between these two studies revealed that the circulating blood background radioactivity was higher and that urinary excretion of radioactivity was lower in patients receiving $^{99m}$Tc–tricine–HYNIC–TOC than those administered $^{99m}$Tc–EDDA–HYNIC–TOC, likely due to the increased plasma protein binding of $^{99m}$Tc–tricine–HYNIC–TOC [118]. In almost all cases, the $^{99m}$Tc–HYNIC-labeled analogs detected the same lesions as $^{111}$In–DTPA–D–Phe$^1$–octreotide, the "gold standard" for SMSR tumor imaging, but the images were clearer and lesions were detected earlier using the $^{99m}$Tc analogs. $^{99m}$Tc–tricine–HYNIC–TOC missed one liver lesion in a patient, which was detected with $^{111}$In–DTPA–D–Phe$^1$–octreotide, probably due to its higher liver uptake [111]. The effect of a coligand on plasma protein binding and blood clearance does not seem to be restricted to small peptides such as $^{99m}$Tc–HYNIC–TOC. Ono et al. [119] found that 3-benzoylpyridine (BP) ternary complexes of $^{99m}$Tc–HYNIC-conjugated Fab fragments [$^{99m}$Tc–HYNIC–Fab(tricine)(BP)] exhibited lower plasma protein binding *in vitro* than the corresponding binary complexes [$^{99m}$Tc–HYNIC–Fab(tricine)$_2$] and were cleared more quickly from the blood in mice (0.35% vs. 0.98% i.d./g at 24 hours p.i., respectively). $^{99m}$Tc–HYNIC–Fab(tricine)(BP) also exhibited lower liver uptake than $^{99m}$Tc–HYNIC–Fab(tricine)$_2$. They concluded that the $^{99m}$Tc–tricine coligand exchanges *in vivo* with plasma proteins as well as lysosomal proteins in tissues, causing slow blood clearance and high liver retention of radioactivity. This loss of radioactivity was decreased using the more stable ternary $^{99m}$HYNIC–Fab(tricine)(BP) complexes. Our group similarly found a slower blood clearance in mice for trastuzumab Fab labeled with $^{99m}$Tc through HYNIC using glucoheptonate as a coligand than for $^{111}$In-labeled trastuzumab Fab (3.2% vs. 1.4% i.d./g at 24 hours p.i., respectively) [108, 120]. There was also higher liver uptake for $^{99m}$Tc–HYNIC–trastuzumab Fab than for $^{111}$In–trastuzumab Fab at 24 hours p.i. (3.8% vs. 2.4% i.d./g, respectively). As $^{99m}$Tc labeling efficiencies for antibodies and peptides are high (>90%) using HYNIC, kit formulation is possible, and indeed kits have been created for labeling HYNIC–TOC with $^{99m}$Tc [111].

A particularly useful method for labeling biomolecules with $^{99m}$Tc, especially those produced by protein engineering techniques and that contain polyhistidine affinity tags, employs a $^{99m}$Tc(I)–carbonyl complex (Figure 7) [121, 122]. This method, described by Waibel et al. in 1999 [122], involves a simple one-vial synthesis of the organometallic aqua–ion complex [$^{99m}$Tc$(H_2O)_3(CO)_3]^+$, which then efficiently complexes through release of its water molecules with imidazole nitrogens in histidine-containing biomolecules. The [$^{99m}$Tc$(H_2O)_3(CO)_3]^+$ complex is formed by heating $^{99m}TcO_4^-$, $Na_2CO_3$, and $NaBH_4$ flushed with carbon monoxide (CO) at 75°C for 30 minutes. The yield of [$^{99m}$Tc$(H_2O)_3(CO)_3]^+$ is >95%. A commercial kit for producing the [$^{99m}$Tc$(H_2O)_3(CO)_3]^+$ complex (Iso-Link; Mallinckrodt, Petten, Belgium) is available. Labeling with $^{99m}$Tc is achieved simply by mixing the [$^{99m}$Tc$(H_2O)_3(CO)_3]^+$ complex with the histidine-containing protein or peptide and heating for 20–30 minutes at 37°C. An scFv containing a polyhistidine tag was labeled with the $^{99m}$Tc(I)–carbonyl complex resulting in a stable complex *in vitro* in human serum and retaining 87% of its radioactivity over 24

hours at 37°C [122]. It was also stable to challenge *in vitro* with a 5000-fold molar excess of histidine, retaining 94% of its radioactivity [122]. This technique is attractive because recombinant antibodies and peptides often incorporate polyhistidine affinity tags for their purification, thus allowing them to be directly and easily labeled using the $^{99m}$Tc(I)–carbonyl complex. Moreover, Re(I)–carbonyl complexes have also been prepared for this method, allowing extension to radionuclides of rhenium (e.g., $^{186}$Re or $^{188}$Re; Table 3) for targeted *in situ* radiotherapy of malignancies [123].

Finally, tetradentate chelators such as $N_3S$ (triamidothiols), $N_2S_2$ (diamidodithiols), or $N_2S_4$ (diaminotetrathiols) have been used for labeling mAbs and peptides with $^{99m}$Tc (Figure 7) [106]. The advantage of these chelators is that they form stable, well-defined $^{99m}$Tc complexes, but they require introduction into biomolecules through reaction of a chemically reactive ester in a side chain with ε-amine groups on lysines. This conjugation step is nonspecific and has the potential to target a key lysine required for maintenance of antigen or receptor-binding. This problem can be solved in the case of small peptides by incorporating the tetradentate chelator at a specific position into the biomolecule during its solid phase synthesis using sequences of amino acids such as cysteine-glycine-glycine-glycine ($N_3S$) [124]. Moreover, there is the possibility that $^{99m}$Tc will bind to low-affinity endogenous metal-binding sites in antibodies in addition to the high-affinity sites introduced with the tetradentate chelator. Therefore, to avoid this possibility, a "preformed chelator" approach was reported by Fritzberg et al. [125]. In this preformed chelator approach, a stable $^{99m}$Tc–$N_2S_2$ complex is synthesized first and then conjugated to the mAbs through a reactive tetrafluorophenyl ester side chain. However, this strategy is complex and time consuming; the conjugation step is inefficient; and it does not lend itself easily to kit formulation. Despite these limitations, tetradentate chelators have been employed for labeling various mAbs and peptides with $^{99m}$Tc [94, 106]. An interesting application of a tetradentate chelator was described for $^{94m}$Tc labeling of the somatostatin analog, demotate [126]. $^{94m}$Tc is a positron-emitter ($\beta^+$ 72%; E$\beta^+$ max 2.5 MeV; Table 2) with a half-life of 52 minutes. Demotate incorporating the 1,4,8,11-tetraazaundecane ($N_4$) tetradentate chelator was labeled with $^{94m}$Tc in the presence of $SnCl_2$. $^{94m}$Tc-demotate exhibited preserved receptor-binding to A-427 non-small cell lung cancer cells infected with the AdHASSTR2 adenovirus encoding the SMSR subtype 2. A-427 xenografts infected with AdHASSTR2 implanted s.c. into athymic mice were visualized by PET at 1 hour p.i. of $^{94m}$Tc-demotate.

### 6.2 Rhenium Radionuclides

Methods for labeling antibodies and peptides with radionuclides of rhenium have been developed from techniques used for labeling with $^{99m}$Tc, due to the similarity in the chemistry of technetium and rhenium [127]. Two radionuclides of rhenium ($^{186}$Re and $^{188}$Re) are useful for targeted *in situ* radiotherapy of cancer (Table 3). $^{186}$Re decays with a half-life of 3.7 days emitting a β-particle with maximum energy ($E_{max}$) of 1.07 MeV as well as a low abundence (9%) γ-photon of 137 keV, which is useful for imaging. $^{188}$Re decays with a 17-hour half-life emitting a β-particle with $E_{max}$ of 2.12 MeV and a low abundance (15%) γ-photon of 155 keV that can be

imaged. The twofold higher β-particle energy of $^{188}$Re compared with $^{186}$Re provides a threefold longer range in tissues (8 vs. 3 mm, respectively), making it more useful for treating larger tumors (i.e., >1 cm in diameter) or tumors in which there is incomplete targeting of radiolabeled mAbs or peptides to cancer cells. On the other hand, the relatively short half-life of $^{188}$Re (17 hours) may limit its feasibility for labeling intact IgG mAbs for radiotherapeutic applications due to their slow kinetics of tumor uptake and clearance from the blood and normal tissues. Antibody fragments and peptides may be more suitable for labeling with $^{188}$Re. A major advantage of $^{188}$Re is that it can be produced carrier-free and in high purity using a $^{188}$W/$^{188}$Re generator system [128], which would make the radionuclide available at low cost in any nuclear medicine facility.

Both direct and indirect methods have been used for labeling antibodies with $^{186}$Re/$^{188}$Re [127]. Direct methods rely on the binding of reduced $^{186}$Re/$^{188}$Re to free thiols on the antibodies generated by reduction with 2-ME as described for $^{99m}$Tc (Section 6.1). However, the chemistries of technetium and rhenium are not identical. In particular, rhenium is more difficult to reduce from its $7^+$ valence state to its lower valence states of $4^+$ or $5^+$ for labeling biomolecules, and it is more easily re-oxidized [129, 130]. Thus, a greater amount of $SnCl_2$ reducing agent is required. The optimal pH for labeling with $^{186}$Re/$^{188}$Re is also lower than that for $^{99m}$Tc-labeling (pH 4.5–5.0 vs. pH 7.0–7.5, respectively) and pH values >5.0 can result in re-oxidation of rhenium [129]. Moreover, radiolysis of antibodies and peptides caused by the high-energy β-particles emitted by $^{186}$Re/$^{188}$Re can be a problem, and thus, radioprotectants such as ascorbic acid, gentisic acid or human serum albumin are often incorporated into the formulations. In one study, ascorbic acid was employed as both a radioprotectant and a reducing agent for the disulfide bonds for labeling IgG with $^{186}$Re [130]. Much longer incubation times (17–24 hours) have been used for direct labeling of mAbs with $^{186}$Re/$^{188}$Re [130, 131] than for $^{99m}$Tc labeling (30 minutes) [132]. However, in one study [133], a kit formulation was developed for labeling the anti-CD20 mAb ritiximab (Rituxan; Roche) with $^{188}$Re. A labeling efficiency >97% was obtained in only 1–1.5 hours after the addition of $^{188}$Re perrhenate eluted from a $^{188}$W/$^{188}$Re generator to the kit.

Tetradentate chelators (e.g., $N_3S$; Figure 7) have also been used for labeling mAbs and their $F(ab')_2$ fragments [134, 135] as well as the somatostatin peptide RC-160 [136] with $^{186}$Re/$^{188}$Re. In the case of antibodies, the "preformed chelate" approach was used, whereas for RC-160, the $N_3S$ chelator was introduced during solid phase peptide synthesis. One method that has been used for direct labeling of antibodies using $^{188}$Re employs the hydroxamic acid, trisuccin. Trisuccin differs from most tetradentate ligands (e.g., $N_3S$ structures) in that it does not contain a free thiol. Therefore, it can be conjugated directly to a mAb and used for labeling with $^{186}$Re/$^{188}$Re without the risk of the free thiol reacting with disulfides on the antibodies. Safavy et al. [137] labeled two different humanized forms of the tumor-associated glycoprotein-72 (TAG-72) mAb CC49 with $^{188}$Re using trisuccin conjugated through a 6-oxoheptanoic acid linker molecule. The labeling efficiency ranged from 80% to 98%, and the immunoreactivity ranged from 69% to 77%. High tumor uptake of radioactivity (18%–23% injected dose/g) was found in mice implanted with s.c. TAG-72-positive LS174T human colon cancer xenografts. Finally, bombesin, a 14-amino acid analog of gastrin-releasing peptide (GRP) that binds to GRP receptors on prostate, breast, lung, and pancreatic cancers, has been

labeled using a $^{188}Re(H_2O)(CO)_3$–carbonyl complex similar to that described for labeling biomolecules with $^{99m}Tc$ (Section 6.1) [138]. Targeting of PC-3 prostate cancer xenografts in mice was achieved with $^{188}Re$-labeled bombesin.

## 6.3 Indium Radionuclides

Indium-111 ($^{111}In$; Table 1) is a γ-emitting radionuclide (Eγ = 172 and 245 keV) with a half-life of 67 hours that is routinely used for labeling mAbs and peptides for SPECT imaging of tumors. The positron-emitter, $^{110m}In$ (Table 2), has a half-life of 69 minutes and is useful for labeling peptides for PET. The $\beta^+$ energy of $^{110m}In$ (2.26 MeV) provides a spatial resolution of 3.0 mm compared with 0.7 mm for $^{18}F$ [13]. Octreotide was labeled with $^{110m}In$ and used for PET of a patient with a SMSR-positive intestinal carcinoma metastasis [139]. $^{114m}In$ and its daughter product, $^{114}In$, are long-lived radionuclide impurities in $^{111}In$, but they have potential for targeted *in situ* radiotherapy of malignancies. $^{114m}In$ decays to $^{114}In$ with a half-life of 49.5 days emitting an imageable γ-photon of 190 keV, as well as Auger and conversion electrons that can kill cancer cells. $^{114}In$ decays to $^{114}Cd$ (0.5%) or $^{114}Sn$ (99.5%) with a half-life of 72 seconds emitting $\beta^-$ particles ($E_{max}$ = 1.98 MeV) that are also useful for treatment of tumors. Octreotide has been labeled with $^{114m}In$ [140]. Antibodies and peptides are labeled with $^{111}In$ by introducing the chelator, diethylenetriaminepentaacetic acid (DTPA), by reaction of ε-amino groups on lysine residues or the N-terminal amine of the biomolecules with reactive forms of DTPA (Figure 8), such as DTPA dianhydride (cDTPAA) [141], DTPA mixed anhydride [142], or DTPA p-benzylisothiocyanate (SCN-Bz-DTPA) [143, 144]. Conjugation with cDTPAA involves suspending cDTPAA in anhydrous chloroform, dispensing an

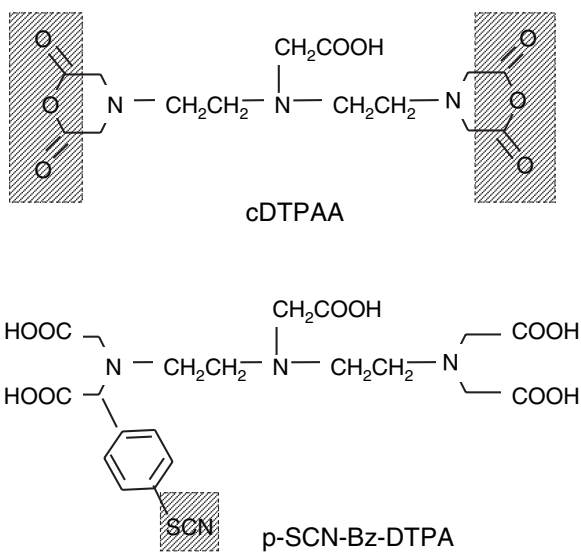


**Figure 8.** Metal chelators that can be conjugated to monoclonal antibodies or peptides for labeling with $^{111}In$. cDTPAA: Bicyclic anhydride of diethylenetriaminepentaacetic acid; p-SCN-Bz-DTPA: p-isothiocyanatobenzyl-diethylenetriaminepentaacetic acid. Chemically reactive groups for conjugation to the antibodies or peptides are shaded.

aliquot of the suspension containing a known amount of cDTPAA into a clean, dry glass tube, and evaporating the chloroform to dryness using a gentle stream of nitrogen to leave a film of cDTPAA on the inside surface of the tube. The antibody or peptide dissolved in 50-mM sodium bicarbonate buffer, pH 7.5, is then added to the tube. Reaction with cDTPAA occurs rapidly within 1–2 minutes, although 15–30 minutes are often allowed. DTPA conjugation efficiency is dependent on protein concentration, pH, and the molar ratio of cDTPAA: antibody/peptide. The DTPA-derivatized biomolecule is separated from excess DTPA by SEC and/or by ultrafiltration. Pure DTPA-conjugated antibodies or peptides can be dispensed into unit-dose vials to produce kits that can be labeled to high radiochemical purity (RCP > 90%) simply by adding $^{111}$In to the vial and incubating at room temperature for 15–30 minutes [145]. Radiolabeling is achieved by transchelation of $^{111}$In from acetate or citrate complexes to DTPA. The acetate or citrate counterions are used to maintain the solubility of $^{111}$In at pH 5–7.5 used for labeling. $^{111}$In–acetate or –citrate complexes are formed by mixing $^{111}InCl_3$ with 0.5–1-M sodium acetate or citrate buffer, pH 5.0. One critical parameter that must be considered for labeling biomolecules with $^{111}$In is the presence of trace amounts of divalent or trivalent metal ions (e.g., Fe, Al, Cd, and Zn) in the labeling reaction. These trace metals may exist at even higher levels than $^{111}$In, which is in commercial $^{111}InCl_3$ solutions [146], and can interfere with labeling by occupying the small number of DTPA metal-binding sites present in biomolecules. Trace metal contamination is minimized by acid-washing of glassware, by using trace-metal-free plasticware, through storing DTPA-conjugated proteins frozen (to avoid leaching of metals into the solution from the glass container), and by purification of conjugation buffers on a cation-exchange column (e.g., Chelex-100) [147]. A method has been reported for ultra-purification of $^{111}$In from trace metals in $^{111}InCl_3$ solution by selective extraction of $^{111}$In as an iodide complex into anhydrous diethyl ether [146].

Hnatowich et al. [141] found that the reaction between cDTPAA and IgG is dependent on pH (optimum between pH 7.5 and 8.5) and protein concentration (optimum >15 mg/mL). There was an inverse relationship between the molar ratio of cDTPAA : IgG and the conjugation efficiency with the greatest efficiency (>70%) found at a 1 : 1 ratio using a protein concentration of 15 mg/mL. However, it may not always be possible to achieve these high protein concentrations, and thus, greater molar ratios of cDTPAA : antibody/peptide (e.g., 5 : 1 to 20 : 1) are used to compensate for the lower conjugation efficiency at lower concentrations (e.g., 2–5 mg/mL). One limitation of the cDTPAA method is that the reagent contains two anhydride moieties that can react with ε-amino groups on lysines or the N-terminal amine of biomolecules. Reaction of cDTPAA with lysines on two biomolecules causes intermolecular cross-linking, leading to the formation of dimers and higher molecular weight polymers. Reaction of cDTPAA with two lysines on the same biomolecule generates intramolecularly cross-linked species. This latter possibility is especially troublesome because in contrast to intermolecularly linked species, intramolecularly cross-linked biomolecules are not easily detected or measured by chromatographic techniques but may significantly diminish immunoreactivity or receptor-binding properties due to protein misfolding [148]. The proportion of intermolecularly linked molecules is directly proportional to the molar ratio of cDTPAA : IgG and protein concentration in the conjugation reaction, and the resulting substitution level (moles DTPA/mole biomolecule) of the conjugate.

Hnatowich et al. [141] reported that the proportion of polymers increased from 0.3% when IgG was conjugated with cDTPAA at a 1:1 molar ratio (cDTPAA:IgG) to as much as 40% at a molar ratio of 10:1 when the IgG concentration was 15 mg/mL. Our group found less than 11% IgG dimers when the HER-2/neu mAb trastuzumab (Herceptin) was modified with a fourfold molar excess of cDTPAA using a protein concentration of 5 mg/mL. In this instance, there were two DTPA molecules/trastuzumab IgG [149]. Another limitation of the cDTPAA method is that one of the five carboxylic acid groups of DTPA is used to form an amide linkage with the antibody or peptide. This converts DTPA to diethylentriaminetetraacetic acid (DTTA), which forms a much less stable heptadentate complex with $^{111}$In. This instability is manifested by a moderate rate of transchelation *in vivo* (7–10% per day) of $^{111}$In from DTPA-conjugated antibodies to transferrin, which then causes deposition of radioactivity in the liver and bone marrow [150, 151]. In contrast, the transchelation rate from an octadentate $^{111}$In–DTPA complex to transferrin is 1–2% per day [151, 152]. Despite these limitations of cDTPAA as a bifunctional chelator, its has been used for $^{111}$In labeling of intact IgG mAbs [18, 120, 147, 151], single-chain Fv fragments [153], and peptides [18, 154, 155]. In the case of peptides that harbor critical lysine residues necessary for receptor-binding, these amino acids must be Boc-protected during conjugation of the N-terminal amine with DTPA, and then deprotected afterward [155].

To address the limitations of cDTPAA, other forms of DTPA have been synthesized that position a reactive group on the methylene carbon of one carboxymethyl arm instead of using one carboxylic acid group to link with an antibody or peptide [143, 144]. One of these DTPA bifunctional chelators is the p-benzylisothiocyanate derivative of DTPA (p–SCN–Bz–DTPA; Figure 8). The p–SCN–Bz–DTPA chelator reacts with ε-amino groups or the N-terminus of biomolecules to form a thiourea linkage. As there is only one reactive moiety on the p–SCN–Bz–DTPA chelator, there is no possibility of inter- or intramolecular cross-linking, which helps to preserve immunoreactivity or receptor-binding affinity [143, 144]. Our group found that the immunoreactive fraction of mAb 2G3 reactive with a 330-kDa glycoprotein on breast and ovarian cancer was 0.52 when conjugated with 0.5–1.5 moles of DTPA using cDTPAA, but it was 0.77 when conjugated with p–SCN–Bz–DTPA [151]. Importantly, the retention of all five carboxylic acid groups in the DTPA molecule for binding $^{111}$In preserves the highly stable octadentate $^{111}$In–DTPA complex, which diminishes the loss of $^{111}$In to transferrin in plasma. The B-cell lymphoma antibody Lym-1 conjugated with p–SCN–Bz–EDTA (a chelator similar to p–SCN–Bz–DTPA) and labeled with $^{111}$In showed a lower rate of loss of $^{111}$In in serum than Lym-1 conjugated with DTPA using cDTPAA (<1% vs. 14% over 5 days, respectively) [156]. Similar results were obtained *in vitro* in serum and *in vivo* in plasma in mice for other $^{111}$In–SCN–Bz–EDTA or SCN–Bz–DTPA immunoconjugates [157]. However, our group did not observe increased stability of $^{111}$In-labeled mAb 2G3 conjugated to SCN–Bz–DTPA *in vitro* in serum compared with $^{111}$In–DTPA–mAb 2G3 (both had transchelation rates of 7% per day), but we did find a lower rate of loss of $^{111}$In from $^{111}$In–SCB–Bz–DTPA–2G3 to ascites fluid from ovarian cancer patients (5% vs. 11% per day) [151]. Avoidance of polymerization and increased serum stability of p–SCN–Bz–DTPA conjugated biomolecules decreases liver uptake and retention of radioactivity, a common problem with $^{111}$In-labeled mAbs and peptides [158].

Methods have been developed to site-specifically conjugate DTPA to the Fc domain of antibodies for labeling with $^{111}$In in order to better preserve their immunoreactivity [159]. Site-specific conjugation of DTPA was achieved by reaction of the N-terminal amine of the tripeptide, glycine–tyrosine–lysine–DTPA (GTK–DTPA) with aldehydes generated in the Fc-domain by sodium periodate oxidation [160]. Stabilization of the resulting Schiff base linkage between GTK–DTPA and the antibodies was achieved by sodium borohydride reduction. This approach positions the DTPA chelator at a position that is remote from the Fab antigen-binding region. Nevertheless, the sodium periodate oxidation step can diminish antibody immunoreactivity if not controlled [161]. The number of aldehydes generated by sodium periodate oxidation can be measured by a spectrophotometric assay that relies on the reaction of the aldehyde groups with dinitrophenylhydrazine (DNPH), generating a derivative that absorbs at 360 nm [162]. Our group recently reported a novel strategy for site-specific labeling of recombinant biomolecules with $^{111}$In that takes advantage of its high affinity for transferrin ($K_a = 10^{28}$ L/mole). We fused the gene for the n-lobe of human transferrin (hn-Tf) through a DNA sequence that encoded a flexible polypeptide linker [$(GGGGS)_3$] to the gene for vascular endothelial growth factor ($VEGF_{165}$) and expressed the recombinant fusion protein in *Pichia pastoris* [163]. The hnTf–VEGF protein bound $^{111}$In directly through its hnTf moiety and retained its binding affinity for VEGF receptors on human umbilical vascular endothelial cells (HUVECs). The protein did not bind to transferrin receptors on cells, because such binding requires both the n- and the c-lobes of transferrin. $^{111}$In–hnTf–VEGF localized specifically in angiogenic U87MG glioblastoma xenografts implanted s.c. in athymic mice permitting tumor imaging at 72 hours p.i. However, $^{111}$In–hnTf–VEGF exhibited relatively high liver uptake and there was a moderately rapid loss of $^{111}$In *in vitro* from the protein to transferrin in plasma (21% per day). This transchelation was not likely due to a lower affinity of the hnTf moiety for $^{111}$In, but it was caused by competition with the higher concentrations of transferrin present in plasma.

Limiting the metal chelator substitution of a biomolecule to one to two DTPA or EDTA groups per molecule in order to preserve its immunoreactivity or receptor-binding affinity restricts the specific activity that can be achieved for labeling with $^{111}$In. For example, monosubstitution of EGF with DTPA restricts the maximum specific activity that can be practically achieved with $^{111}$In to 40 MBq/μg ($2.4 \times 10^5$ MBq/μmole). At this low specific activity, only one in eight EGF molecules carries an $^{111}$In atom and almost 90% of EGFRs in a tumor would be targeted by non-radiolabeled EGF. To address this issue, Remy et al. [164] conjugated maleimide-derivatized EGF with a thiol-containing multibranched peptide containing four EDTA-like metal chelators for $^{111}$In. However, the MCP–4–EDTA–S–MB–EGF conjugate showed a 40-fold decrease in receptor-binding affinity in a competition assay with MDA–MB–468 human breast cancer cells compared with unmodified EGF. In contrast, our group found that derivatization of EGF directly with one to two DTPA metal chelators and labeling with $^{111}$In yielded a radiopharmaceutical with receptor-binding affinity identical to that of $^{125}$I-labeled EGF ($K_a = 7.3 \times 10^8$ L/mole) [18]. Greater success was achieved by our group in maximizing the specific activity of $^{111}$In–EGF by conjugating maleimide-derivatized EGF to thiolated human serum albumin (HSA), which presents 60 lysine residues for DTPA conjugation [165]. Conjugation of EGF to HSA diminished its receptor-

binding affinity 15-fold ($K_a = 5.1 \times 10^7$ L/mole), but there were no further decreases in affinity when up to 23 DTPA chelators were preferentially substituted into the HSA moiety. The specific activity of $^{111}$In–DTPA–HSA–EGF was increased 10-fold compared with $^{111}$In–DTPA–hEGF. $^{111}$In–DTPA–HSA–EGF retained its receptor-mediated internalization and nuclear translocation properties in MDA–MB–468 cells and was fourfold more growth-inhibitory toward the cells. An analogous strategy was reported by Manabe et al. [166] that used a polylysine peptide carrier to conjugate as many as 42 DTPA molecules to the anti-HLA mAb H-1 with > 90% retention of immunoreactivity. Similarly, starburst dendrimers that display 64 amine groups on their surface were derivatized with up to 43 1B4M DTPA-like chelators, and then were conjugated through a maleimido bond to mAb OST7 [167]. The maximum specific activity achieved for labeling these starburst dendrimer immunoconjugates with $^{111}$In (8.4 MBq/μg; $1.3 \times 10^6$ MBq/μmole) was 48-fold higher than directly conjugated 1B4M-OST7. In another study, poly(ethylene glycol) (PEG) containing an amine functional group was used to conjugate DTPA chelators to the anti-EGFR mAb C225 with good preservation of immunoreactivity [168]. This approach provides a method of labeling with $^{111}$In while minimizing nonspecific retention of radioactivity in the liver due to conjugation with PEG [169].

High retention of radioactivity in the liver with $^{111}$In–DTPA-conjugated mAbs is thought to be due to intracellular trapping of $^{111}$In-catabolites. $^{111}$In–DTPA-labeled IgG antibodies interact with glycoprotein receptors on hepatocytes through their Fc-domain and are internalized into endosomes. The antibodies are routed to lysosomes for proteolytic degradation. The ultimate catabolite of proteolysis has been identified as $^{111}$In–DTPA–ε–Lys [30, 170]. This catabolite cannot easily cross the lysosomal membrane due to its positive charges or be exocytosed due to its poor recognition by cell membrane amino acid transporters and, thus, becomes trapped. An analogous mechanism has been proposed for retention of $^{111}$In-labeled mAb fragments and peptides by renal tubular cells. $^{111}$In-labeled mAb fragments or peptides are filtered by the glomerulus and reabsorbed by renal tubular cells. Proteolytic catabolism within renal tubular cells of $^{111}$In–DTPA–F(ab′)$_2$ fragments yields $^{111}$In–DTPA–ε–Lys, which is trapped within the cells [171]. Proteolysis of $^{111}$In–DTPA–D–Phe$^1$–octreotide or $^{111}$In–DTPA–L–Phe$^1$–octreotide similarly results in retention of $^{111}$In–DTPA–D–Phe$^1$–OH or $^{111}$In–DTPA–L–Phe$^1$–OH catabolites, respectively, in renal tubular cells [172, 173].

The macrocyclic chelator, 1,4,7,10-tetraazacyclododecane N,N′,N″,N‴-tetraacetic acid (DOTA) (Section 6.4) has been less commonly used for labeling mAbs and peptides with $^{111}$In because the kinetics of binding $^{111}$In by DOTA are much slower than those of DTPA. Elevated temperatures of 37–43°C for mAbs [174] and heating at 100°C for peptides [175] combined with longer incubation times (30–45 minutes) are required to obtain complete incorporation of $^{111}$In into DOTA-conjugated biomolecules. Nevertheless, intact IgG mAbs [174, 176], Fab fragments [177] and diabodies [178], as well as octreotide [175] have been labeled with $^{111}$In using DOTA. The stability of $^{111}$In–DOTA-conjugated mAbs in serum is greater than that of $^{111}$In–DTPA-conjugated antibodies. Lewis et al. [174] found that there was less than 0.8% loss of $^{111}$In in serum from $^{111}$In–DOTA–cT84.66 mAb over 10 days compared with 13.0% for $^{111}$In–DTPA–cT84.66. The advantage of conjugating mAbs or peptides with DOTA for labeling with $^{111}$In is that it provides an analog

for the corresponding yttrium-90 ($^{90}$Y)–DOTA–biomolecule used for targeted *in situ* radiotherapy. Imaging with the $^{111}$In–DOTA–biomolecule can be used to predict radiation dosimetry estimates to tissues for the $^{90}$Y–DOTA–biomolecule [179, 180]. Use of an $^{111}$In–DTPA–conjugated mAb or peptide for this purpose may yield inaccurate dosimetry estimates because of major differences in stability *in vivo* between $^{111}$In– and $^{90}$Y–DTPA complexes.

### 6.4 Yttrium Radionuclides

Two radionuclides of yttrium, $^{86}$Y and $^{90}$Y, are available for labeling biomolecules for imaging or radiotherapeutic purposes, respectively. $^{86}$Y (Table 2) is a positron-emitter with a half-life of 14.7 hours that is produced in a cyclotron using the $^{86}$Sr(p,n)$^{86}$Y reaction [181]. The spatial resolution for detection of the positron annihilation from $^{86}$Y is 1.8 mm compared with 0.7 mm for $^{18}$F [13]. $^{90}$Y (Table 3) is a pure β-emitter with a half-life of 2.7 days that is produced by a strontium-90 ($^{90}$Sr)/$^{90}$Y generator [182]. $^{90}$Y is an attractive radionuclide for *in situ* radiotherapy of cancer because it does not emit γ-radiation, which minimizes the radiation exposure to health-care personnel and family members from a patient-administered $^{90}$Y-labeled mAb or peptide. For example, patients with non-Hodgkin's B-cell lymphoma treated with $^{90}$Y-ibritumomab tiuxetan (Zevalin; Biogen Idec, Cambridge, MA) do not need to be isolated for radiation safety reasons and can be treated as outpatients, whereas patients receiving $^{131}$I-labeled tositumomab (Bexxar; GlaxoSmithKline, Uxbridge, Middlesex, UK) require special radiation safety precautions to be taken [183]. However, as $^{90}$Y does not emit γ-radiation, the tissue distribution of $^{90}$Y-labeled mAbs and peptides cannot be easily imaged (only poor-resolution Bremstrahlung images can be acquired). Therefore, the radiation absorbed doses to tissues in patients from $^{90}$Y-labeled biomolecules are often estimated by imaging studies with $^{111}$In-labeled analogs [179, 180]. However, as mentioned (see Section 3), differences in the *in vivo* stability of $^{111}$In- and $^{90}$Y-labeled biomolecules can produce discrepancies in their pharmacokinetics and normal organ distribution, yielding inaccuracies in the dosimetry estimates. PET imaging using $^{86}$Y-labeled analogs has been suggested as a means of more accurately estimating the radiation absorbed doses from $^{90}$Y-labeled biomolecules [184].

The macrocyclic chelator, DOTA (Figure 9), is the chelating agent of choice for labeling mAbs and peptides with yttrium-90 ($^{90}$Y) due to the high stability of $^{90}$Y-DOTA complexes ($K_a = 10^{24}$ L/mole) [185]. Initially, DTPA was used as a chelator for $^{90}$Y-labeling [186], but it was discovered that $^{90}$Y–DTPA complexes were unstable *in vivo*, which caused bone accumulation of free $^{90}$Y released from the immunoconjugates. The stability of $^{90}$Y-labeled biomolecules is of paramount importance because even a small amount of free $^{90}$Y sequestered in the bone can contribute significantly to bone marrow toxicity due to the long range (10–12 mm) of the β-particles ("cross-fire" effect). This is one reason that the maximum tolerated dose of $^{90}$Y-labeled mAbs (e.g., Zevalin) is 32 mCi, whereas up to 150 mCi of $^{131}$I-labeled mAbs can be safely administered [2]. Bz–SCN–DTPA (Figure 8) yields more stable $^{90}$Y complexes than those formed with DTPA, but the complexes are not as stable as $^{90}$Y–DOTA complexes. In one study [187], the bone uptake of radioactivity in mice implanted s.c. with SK–RC–52 renal cell carcinoma xenografts at 7 days p.i. of $^{90}$Y–DOTA-conjugated chimeric G250 mAbs was only 0.4% i.d./g,

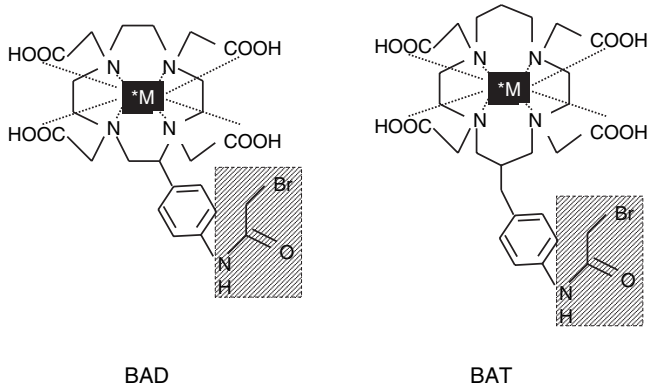


**Figure 9.** Chemically reactive forms of two macrocyclic chelators that can be conjugated to monoclonal antibodies or peptides for labeling with radiometals. Groups for conjugation to the antibodies or peptides are shaded. BAD: p-bromoacetamidobenzyl-DOTA complexes $^{90}$Y, $^{67}$Ga, $^{177}$Lu, $^{212}$Pb, $^{212}$Bi or $^{225}$Ac. BAT: bromoacetamidobenzyl-TETA complexes $^{64}$Cu or $^{67}$Cu.

compared with 1.2% i.d./g for Bz–SCN–DTPA- and 10.7% i.d./g for DTPA-conjugated antibodies. Similarly, a threefold lower bone uptake of radioactivity at 10 days p.i. in mice was observed for $^{88}$Y–DOTA–hLL2 anti-CD22 mAbs compared with $^{88}$Y–Mx–DTPA–hLL2 mAbs [188].

DOTA, which was first described by McCall et al. in 1990 [189], can be conjugated to mAbs by conversion to its reactive form, p-bromoacetamidobenzyl–DOTA (BAD; Figure 9) followed by reaction of BAD with free thiols introduced into the antibodies by reaction with 2-iminothiolane (2-IT) [190]. 2-IT also creates a spacer between the mAbs and the DOTA chelator that allows more efficient labeling with $^{90}$Y. An alternative approach involves reaction of an N-hydroxysuccinimidyl ester of DOTA with ε-amino groups on lysine residues or the N-terminus of biomolecules [191]. Labeling of DOTA-conjugated mAbs is achieved by incubation with $^{90}$Y chloride ($^{90}YCl_3$) mixed with 0.5-M ammonium acetate buffer, pH 7.0–7.5, for 30 minutes at 37°C [192]. Similar conditions have been employed for labeling DOTA-conjugated peptides with $^{90}$Y, except that a temperature of 80–100°C was used to accelerate the incorporation of $^{90}$Y [193]. Labeling of biomolecules with $^{90}$Y using DOTA is highly susceptible to the effects of divalent trace metal ion contamination, especially $Ca^{2+}$, $Fe^{2+}$, and $Zn^{2+}$. In fact, the affinity constant ($K_a$) for binding of $Fe^{2+}$ by DOTA is 100,000 times higher than that for binding $^{90}$Y [185]. It is critical to exclude trace-metals as much as possible from the $^{90}$Y labeling reaction. One recent Letter to the Editor even suggests that minor trace-metal contamination of pipette tips can seriously diminish the labeling efficiency of DOTA-conjugated peptides with $^{90}$Y [194]. Another important issue in labeling biomolecules with $^{90}$Y is the potential for radiolysis. Interaction of the moderate energy ($E_{max}$ 2.2 MeV) β-particles emitted by $^{90}$Y with water molecules in the buffers used to formulate the radiopharmaceuticals generates highly reactive free radicals that degrade the metal chelator as well as the biomolecules themselves. The radiolysis effect is dependent on specific activity and on radioactivity concentration. For example, $^{90}$Y–BAD–2–IT–Lym–1 antibodies formulated at a specific

activity of 1–2 mCi/mg remained pure and relatively immunoreactive (>75%) in storage over a 3-day period [195]. However, at a specific activity of 4 mCi/mg, the radiochemical purity (RCP) of $^{90}$Y–BAD–2–IT–Lym–1 dropped to 65% and the immunoreactivity decreased to 28%. At a specific activity of 9.4 mCi/mg, the RCP of $^{90}$Y–BAD–2–IT–Lym–1 decreased to 21% and the immunoreactivity was virtually abolished (3%). Radiolysis can be minimized by inclusion of radioprotectants such as ascorbic acid, gentisic acid, or human serum albumin in the formulation and by freezing the radiolabeled biomolecules to minimize diffusion of free radicals in the aqueous solutions [195, 196]. In one study, it was shown that ascorbic acid may also act as a suitable buffering agent, thereby removing the need for the ammonium acetate buffer [197].

A controversial issue with respect to the use of DOTA as a metal chelator for $^{90}$Y and other radiometals is its potential immunogenicity in humans. In a phase I/II clinical trial in which six ovarian cancer patients received $^{90}$Y–DOTA–HMFG1 murine mAbs intraperitoneally, all patients developed anti-DOTA antibodies and three patients developed serum sickness [198]. Four of eight patients who received $^{111}$In–DOTA-conjugated mAbs intravenously developed anti-DOTA antibodies. The immune response seems to be directed against the DOTA ring structure and not the benzyl-containing side chain. The immune response to DOTA is also dependent on the antigenicity of the biomolecule to which it is conjugated, because rabbit IgG conjugated to DOTA did not induce anti-DOTA antibodies in rabbits, whereas mouse IgG–DOTA immunoconjugates administered to rabbits stimulated an immune response toward the chelators [199]. Moreover, an immune response to DOTA has not been found with all mAbs. Anti-DOTA antibodies were not detected in the serum of 18 lymphoma patients administered multiple doses of $^{111}$In–BAD–2–IT–Lym–1 mAbs, although these patients typically do not mount a strong immune response to radioimmunoconjugates [200]. Similarly, $^{90}$Y–DOTATOC, an octapeptide somatostatin analog, does not seem to be immunogenic, again suggesting that the biomolecule carrier must be considered in assessing the potential immunogenicity of DOTA [201].

### 6.5 Gallium Radionuclides

Gallium-67 ($^{67}$Ga; Table 1) is a cyclotron-produced radionuclide with a half-life of 78.2 hours that has been used for many years in nuclear medicine as $^{67}$Ga citrate for tumor imaging [202]. Despite its history of use, the properties of $^{67}$Ga are not ideal for imaging, however, due to the high energy of two of its γ-photons ($E\gamma = 300$ and 393 keV), which make it difficult to collimate. Nevertheless, $^{67}$Ga has recently received attention as a radiolabel for mAbs for *in situ* radiotherapy of cancer, exploiting its abundent Auger and conversion electron emissions [203]. Most of the recent interest in gallium radionuclides has been focused on labeling peptides with $^{68}$Ga for PET (Table 2) [204]. Antibodies have not been labeled with $^{68}$Ga for PET, because their kinetics of tumor uptake and elimination from the blood and normal tissues is too slow. $^{68}$Ga is conveniently produced using a germanium-68 ($^{68}$Ge)/$^{68}$Ga generator system that could allow production of the radionuclide for up to one year in a nuclear medicine facility using a single generator, due to the long half-life of $^{68}$Ge (270 days) [205]. $^{68}$Ga decays with a half-life of 68 minutes to $^{68}$Zn. Its positron energy is 1.92 MeV, which provides a spatial resolution of 2.4 mm

versus 0.7 mm for $^{18}F$ [13]. $^{66}Ga$ is a longer lived positron-emitting radionuclide (half-life of 9.5 hours), which has also been studied for labeling peptides for PET [206].

Initially, desferrioxamine (DFO) was used as a chelator for labeling peptides with $^{68}Ga$ [74, 207], but more recently DOTA has been used (Figure 9) [204, 208]. Most PET tumor imaging studies with $^{68}Ga$-labeled peptides have focused on DOTATOC, a synthetic octapeptide analog of somatostatin. Labeling DOTATOC involves adjusting the pH of $^{68}GaCl_3$, which is eluted in 0.5-M HCl from the $^{68}Ge/^{68}Ga$ generator to pH 4.8 with 50-mM sodium acetate buffer. The required amount of $^{68}Ga$ acetate complex is then mixed with DOTATOC (10–20 nmols) and heated at 95°C for 15 minutes in a heating block. Early studies in which DOTATOC was labeled with $^{68}Ga$ using this technique yielded labeling efficiencies that were about 50%, thus requiring postlabeling purification on a C-18 Sep-Pak cartridge [208, 209]. The final radiochemical purity was >95%. The large volumes of $^{68}Ga$ eluates eluted from the $^{68}Ge/^{68}Ga$ generator system as well as contamination with trace metals were believed to be responsible for the low labeling efficiencies for DOTATOC with $^{68}Ga$. Improvements in the concentration of $^{68}Ga$ eluates using an ion-exchange cartridge combined with more homogeneous microwave heating of the labeling reaction for 10–20 minutes have recently yielded almost complete incorporation of $^{68}Ga$ in to extremely small quantities (<1 nmole) of DOTATOC, thus increasing the specific activity by almost 100-fold [210].

Interestingly, DOTATOC labeled with $^{68}Ga$ has a fivefold higher binding affinity for SMSR subtype 2 ($IC_{50}$ = 2.5 nmol/L) compared with $^{90}Y$–DOTATOC ($IC_{50}$ = 11 nmol/L) and a ninefold higher affinity than that of $^{111}In$–DTPA–octreotide ($IC_{50}$ 22 nmol/L) [211]. Combined with the high spatial resolution and sensitivity of PET, this provided a greater sensitivity for imaging small meningioma lesions (100% vs. 85%) in patients compared with SPECT imaging with $^{111}In$–DTPA–octreotide [209, 212]. The somatostatin analog DOTANOC, which binds to SMSR subtypes 2 and 5, has been labeled with $^{68}Ga$ and used for PET of a patient with metastases from a neuroendocrine tumor [213]. Preclinical studies have been performed in mouse tumor xenograft models using $^{68}Ga$–DOTA–α–melanocyte-stimulating hormone (α-MSH) for imaging melanoma [214] or $^{68}Ga$–DOTA–EGF to image gliomas or epidermoid carcinoma xenografts [215]. Due to the convenient availability of a generator system for $^{68}Ga$, it is not necessary for a nuclear medicine facility to have access to a cyclotron, and therefore, it is likely that more studies ultimately leading to clinical application of many different $^{68}Ga$-labeled peptides for PET imaging of tumors will be performed in the near future.

### 6.6 Copper Radionuclides

Several radionuclides of copper are suitable for labeling mAbs, their fragments, or peptides for PET (e.g., $^{60}Cu$, $^{61}Cu$, or $^{64}Cu$; Table 2) [216]. $^{64}Cu$ decays with a half-life of 12.7 hours by three different pathways: (1) electron capture (41%) emitting Auger and conversion electrons, (2) positron ($\beta^+$) emission (19%), and (3) $\beta^-$ emission (40%). The $\beta^-$ emissions make the radionuclide useful not only for PET but also for radiotherapy of tumors. Indeed, in one study, $^{64}Cu$-labeled octreotide inhibited the growth of CA20948 rat pancreatic tumors in Lewis rats [217]. In another study, $^{64}Cu$-labeled 1A3 mAbs were used to treat hamsters implanted s.c.

with GW39 human colon carcinoma xenografts [218]. $^{64}$Cu is most commonly produced in a biomedical cyclotron by the $^{64}$Ni(p,n)$^{64}$Cu reaction using a $^{64}$Ni-enriched target [219, 220]. The positron energy of $^{64}$Cu is almost identical to that of $^{18}$F (0.657 MeV vs. 0.635 MeV, respectively), and the intrinsic spatial resolution is indistinguishable (0.73 vs. 0.70 mm, respectively) [13, 221]. However, due to the lower abundance of positron emission with $^{64}$Cu compared with $^{18}$F (19% vs. 97%), the sensitivity for PET with $^{64}$Cu is about fivefold lower than that with $^{18}$F [221]. The short-lived positron-emitter, $^{61}$Cu (half-life 3.32 h), has a greater abundance of positron emission than $^{64}$Cu (60% vs. 19%) and has been used for labeling octreotide for PET of CA20948 tumors in rats [222]. $^{67}$Cu (Table 3) can be used for labeling mAbs and peptides for targeted *in situ* radiotherapy. $^{67}$Cu decays with a half-life of 2.6 days emitting $\beta^-$ particles with energies of 0.395 (51%), 0.484 (28%), and 0.577 (20%) MeV as well as several $\gamma$-photons with energies ranging from 91 to 300 keV that are imageable [223]. $^{67}$Cu is produced in a biomedical cyclotron using the $^{nat}$Zn(p,2p)$^{67}$Cu or $^{68}$Zn(p,2p)$^{67}$Cu reactions or in a nuclear reactor using the $^{nat}$Zn(n,p)$^{67}$Cu reaction [223].

DTPA (Figure 8) and benzyl-EDTA were initially examined for chelating $^{67}$Cu, but it was soon found that these complexes formed were unstable, releasing 70–95% of their radiolabel *in vitro* in serum over 5 days [224]. In contrast, the macrocyclic chelator, 1,4,8,11-tetraazacyclotetradecane-N′,N″,N‴-tetraacetic acid (TETA; Figure 9), provided a more stable $^{67}$Cu complex that lost only 2–6% of the radiolabel in serum when conjugated to Lym-1 antibodies [224]. This greater stability of $^{67}$Cu–TETA vs. $^{67}$Cu–DTPA or Bz–EDTA complexes is not due to a higher thermodynamic stability, but it is believed to be due to the structural rigidity of the $^{67}$Cu–TETA complex, which shields the radionuclide from attack by endogenous copper binding proteins. Copper is bound in plasma by albumin and transcuprein, which transport the metal to hepatocytes [216]. Copper internalized by hepatocytes is used to synthesize metalloenzymes such as superoxide dismutase (SOD), is stored bound to metallothionein (MT), or is secreted back into the plasma complexed to ceruloplasmin. DOTA (Figure 9) and its analogs form stable complexes with copper radionuclides, but TETA is the most widely used chelator [216].

BAT (Figure 9) is a modified form of TETA, which contains a bromoacetamidobenzyl side chain that is reactive with antibodies modified with 2-minothiolane (2-IT) to present a free thiol. The inclusion of the 2-IT spacer improves the labeling efficiency of the antibodies with $^{67}$Cu and $^{64}$Cu [225]. DOTA and TETA chelators can be introduced into peptides during solid phase synthesis for labeling with $^{64}$Cu/$^{67}$Cu [226, 227]. Labeling of TETA-conjugated mAbs or peptides is achieved by incubation with $^{64}$Cu/$^{67}$Cu in a 100-mM ammonium citrate buffer, pH 5.5, at room temperature for 30–60 minutes [226, 228]. Disodium EDTA is added to chelate any unbound radiometal, and the $^{64}$Cu/$^{67}$Cu-labeled biomolecules are purified by SEC. $^{64}$Cu/$^{67}$Cu labeled peptides are usually purified on a C-18 Sep-Pak (Waters Associates Inc., Milford, MA). Using TETA, mAb1A3 IgG and its F(ab′)$_2$ fragments were labeled with $^{64}$Cu for PET in a phase I/II trial of 36 patients with colorectal carcinoma [229] and Lym-1 antibodies were labeled with $^{67}$Cu for RIT of 12 patients with non-Hodgkin's lymphoma [228]. Several different peptides have been labeled with $^{64}$Cu using DOTA or TETA for PET of tumors, including octreotide analogs [226, 227, 230, 231], RGD peptides that recognize $\alpha_v\beta_3$ integrins [232], vasoactive intestinal polypeptide (VIP) analogs [233], and bombesin derivatives [234].

$^{64}Cu/^{67}Cu$ labeled intact IgG antibodies exhibit high accumulation in the liver, whereas $^{64}Cu/^{67}Cu$-labeled $F(ab')_2$ fragments and peptides are retained by the kidneys. In rats administered 1A3 mAbs labeled with $^{67}Cu$ using four different chelators, including 1-[(1,4,8,11-tetraazacyclotetradec-1-yl)methyl]benzoic acid (CPTA), a macrocycle related to TETA, it was found that there was a slow rate of transchelation of $^{67}Cu$ to SOD in the liver [235]. In contrast, $^{67}Cu$-labeled $F(ab')_2$ fragments of mAb 1A3 were rapidly catabolized to $^{67}Cu$–CPTA–ε–lysine. Similarly, transchelation of radiometal to SOD in the liver was observed for $^{64}Cu$–TETA–octreotide in rats [236]. However, in a clinical study, in which 10 patients received $^{67}Cu$–2IT–BAT–Lym-1 antibodies [237], only a small fraction (0.8–7.8%) of the injected dose of $^{67}Cu$ was recycled by the liver and, in this case, mostly to ceruloplasmin. There was no transchelation to SOD detectable. The higher concentration of SOD in the liver of rats compared with humans may account for this species-dependent catabolic route for $^{64}Cu/^{67}Cu$-labeled biomolecules [236].

### 6.7 Lutetium Radionuclides

Lutetium-177 ($^{177}Lu$; Table 3) is a β-emitting radionuclide (Eβ = 0.495 MeV) with a half-life of 6.7 days that is useful for labeling mAbs and peptides for targeted *in situ* radiotherapy of tumors. In addition, $^{177}Lu$ emits γ-photons of energies 113 and 208 keV that can be imaged. $^{177}Lu$ is produced in a nuclear reactor by neutron irradiation of $^{176}Lu_2O_3$. Several chelators have been employed for labeling mAbs and peptides with $^{177}Lu$ including derivatives of DOTA (e.g., PA–DOTA and CA–DOTA) and analogs of DTPA (e.g., CHX–A–DTPA and SCN–Bz–DTPA) [238]. The difference between the two DOTA chelators is that in PA–DOTA, the benzylisothiocyanate side chain used for conjugation is attached to one nitrogen in the macrocycle, whereas in CA–DOTA, it is attached to a methylene carbon. PA–DOTA forms a less stable complex with $^{177}Lu$ than CA–DOTA or CHX–DTPA [238]. The instability of the $^{177}Lu$–PA–DOTA complex may account for the prolonged retention of radioactivity in the reticuloendothelial system (RES), including the bone marrow of patients administered $^{177}Lu$–PA–DOTA–CC49 mAbs for RIT of TAG-72 positive malignancies and, consequently, the observed dose-limiting hematopoietic toxicity [239]. $^{177}Lu$ released from biomolecules is expected to be sequestered in the skeleton because lutetium competes with calcium for bone deposition [240]. It has been shown that by adding DTPA to $^{177}Lu$–DOTA$^0$–Tyr$^3$]octreotate used for PDRT of SMSR-positive tumors, the small amount (<2–5%) of free $^{177}Lu$ impurity in the radiopharmaceutical can be rapidly eliminated from the body by renal excretion as $^{177}Lu$–DTPA, thus minimizing bone uptake of radioactivity [241]. In addition to labeling intact IgG CC49 mAbs with $^{177}Lu$ [239, 242], dimeric scFv fragments of this antibody have been labeled with $^{177}Lu$ for radiotherapeutic purposes [243].

### 6.8 Lead, Bismuth, and Actinium Radionuclides

DOTA (Figure 9) is a also useful bifunctional chelator for labeling antibodies with radionuclides of lead, bismuth, and actinium for targeted *in situ* radiotherapy [2]. Lead-212 ($^{212}Pb$) decays to the α-emitter, bismuth-212 ($^{212}Bi$). The α-emitter,

actinium-225 ($^{225}$Ac), decays to a series of daughter radionuclides that are α-emitters ($^{221}$Fr, $^{217}$At, $^{213}$Bi) or β-emitters ($^{213}$Po, $^{209}$Tl, $^{209}$Pb, $^{209}$Bi). The use of $^{225}$Ac as a radiolabel for mAbs has been termed an "atomic nanogenerator" because the decay of $^{225}$Ac generates the daughter radionuclides locally in tumors that then emit several different forms of radiation that kill cancer cells [244]. The main concern for employing radionuclides such as $^{212}$Pb or $^{225}$Ac that do not decay directly to stable elements is that the decay process may decrease the stability of the metal chelate complex due to the transformation of one element into another (i.e., $^{212}$Pb is converted into $^{212}$Bi). In particular, bismuth has an affinity for the kidneys, and it was shown that even low doses of $^{225}$Ac-labeled HuM195 antibodies caused severe renal toxicity and anemia in monkeys, likely due to release of $^{213}$Bi from the antibodies and redistribution of the radionuclide to the kidneys [245]. Bismuth radionuclides are also stably bound by CHX–A–DTPA and CHX–B–DTPA, two analogs of DTPA [246, 247]. HER2/neu mAbs AE1 and trastuzumab (Herceptin) have been labeled with $^{212}$Pb or $^{225}$Ac, respectively, using DOTA as the metal chelator [248, 249].

## 7 CHARACTERIZATION OF RADIOLABELED MABS AND PEPTIDES

It is important to fully characterize the properties of radiolabeled mAbs and peptides intended for tumor imaging or targeted *in situ* radiotherapy of malignancies. Characterization includes (1) analytical tests that evaluate homogeneity and radiochemical purity; (2) radioligand binding assays that assess the ability to specifically bind target antigens/receptors *in vitro*; (3) stability studies that assess the loss of radiolabel *in vitro* in biologically relevant media; and (4) biodistribution, pharmacokinetic, dosimetry, and imaging studies that reveal tumor and normal tissue uptake *in vivo* in an animal model and predict radiation absorbed doses in humans.

### 7.1 Evaluation of the Homogeneity of Radiopharmaceutical Bioconjugates

Evaluation of homogeneity is especially important for mAbs or peptides conjugated to chelators for labeling with radiometals (Section 6). Substitution of these biomolecules with too many chelators may significantly decrease their immunoreactivity or receptor-binding affinity [159]. The substitution level (moles chelators/mole of biomolecule) may be measured by spectrophotometric, fluorescence, or radiochemical assays. For example, HYNIC substitution in biomolecules intended to be labeled with $^{99m}$Tc or $^{186}$Re/$^{188}$Re (see sections 1 and 2) can be measured by a colorimetric assay that relies on reaction of the HYNIC groups with p-nitrobenzaldehyde to form a complex that absorbs at 385 nm [162]. DTPA substitution for labeling biomolecules with $^{111}$In (see Section 6.3) can be measured by several different techniques: (1) radiochemical assays that rely on the binding of $^{111}$In or $^{57}$Co by DTPA [141, 250]; (2) spectrophotometric assays that generate a colored arsenazo III DTPA complex [251]; (3) fluorescence assays that measure the binding of europium(III) by DTPA [252]; or (4) immunoelectrofocusing (IEF) that reveals changes in pI associated with DTPA substitution [253]. The most

common method is simply to radiolabel an aliquot of the impure DTPA conjugation mixture with a trace amount of $^{111}$In and separate the resulting $^{111}$In–DTPA–biomolecule from free $^{111}$In–DTPA by silica gel instant thin-layer chromatography (ITLC–SG) developed in 100-mM sodium citrate, pH 5.0. The conjugation efficiency is then calculated from the ITLC–SG results and multiplied by the molar ratio of cDTPAA : antibody/peptide used in the reaction to estimate the average number of moles of DTPA per mole of biomolecule. Typically, a substitution level of one to two moles DTPA per mole of biomolecule is desirable to minimize polymerization and/or interference in immunoreactivity or receptor-binding affinity [159]. The number of DOTA chelators introduced into antibodies or peptides for labeling with $^{90}$Y, $^{68}$Ga, $^{177}$Lu, $^{212}$Bi, or $^{225}$Ac (see Sections 4, 5, 7, and 8) can be measured by a spectrophotometric assay, which measures the decrease in absorbance at 656 nm for a Pb(II)–arsenazo III complex upon transchelation of $Pb^{2+}$ to DOTA [254].

The polymerization of biomolecules as a consequence of metal chelator substitution may be assessed by size-exclusion HPLC with UV detection (Figure 10) or by sodium dodecylsulfonate–polyacrylamide gel electrophoresis (SDS–PAGE; Figure 10). Polymerization is most commonly due to cross-linking of mAbs or peptides by bifunctional chelators such as cDTPAA that contain two reactive groups that

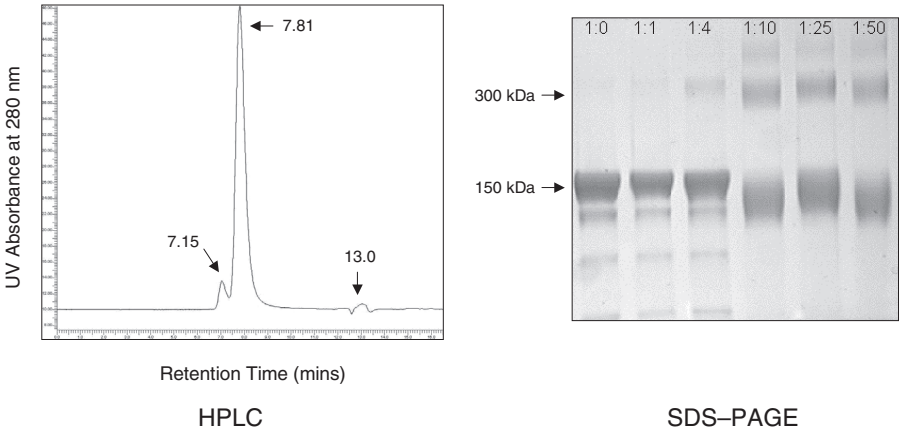


**Figure 10.** Two analytical methods used to assess the purity and homogeneity of metal-chelated monoclonal antibodies or peptides. *Left Panel*: Size-exclusion HPLC analysis of trastuzumab (Herceptin) IgG conjugated with the bicyclic anhydride of DTPA (cDTPAA) on a BioSep-SEC-S2000 column eluted with 150-mM sodium chloride/10-mM sodium phosphate buffer, pH 6.8, at a flow rate of 0.6 mL/min with UV detection at 280 nm. The molar ratio of cDTPAA : IgG was 4 : 1. Peak with a retention time ($t_R$) of 7.81 minutes represents a DTPA-conjugated trastuzumab monomer. Peak with $t_R$ of 7.15 minutes represents a DTPA–trastuzumab dimer caused by cross-linking of the antibody molecules through the two chemically reactive groups on cDTPAA. Peak with $t_R$ of 13.0 minutes represents a small amount of unconjugated DTPA (detected through a change in refractive index). Right Panel: SDS–PAGE analysis of trastuzumab (Herceptin) IgG conjugated with cDTPAA on a 4–10% Tris HCl gradient mini-gel stained with Coomassie Brilliant Blue. There is an increase in the proportion of dimerized (300 kDa) and polymerized (>300 kDa) IgG species as the molar ratio of cDTPAA : IgG is increased from 1 : 1 to 50 : 1. (Data generously provided by Kristin McLarty.)

interact with ε-amino groups on lysines on two separate biomolecules (intermolecular linkage) or within one biomolecule (intramolecular linkage) [141, 151]. Such polymerization can be eliminated by using a bifunctional chelator that contains only a single reactive group (e.g., SCN–Bz–DTPA). The level of polymerization that is tolerable depends on the biomolecule, but less than 10% of higher molecular weight species is desirable to avoid diminishing immunoreactivity or receptor-bindng affinity, as well as to minimize sequestration by the reticuloendothelial system (RES) [158].

### 7.2 Measurement of Radiochemical Purity (RCP)

In our laboratory, we determine the radiochemical purity (RCP) of radiolabeled mAbs or peptides by size-exclusion HPLC on a BioSep SEC-S 2000 column (Phenomenex, Canada) eluted with 100-mM sodium phosphate buffer, pH 7.4 at a flow rate of 1.0 mL/minute, and interfaced with a PerkinElmer diode array detector set at 280 nm and an FSA radioactivity detector (Figure 11). The RCP of radiolabeled peptides could also be measured by reversed-phase HPLC on a C-18 column eluted with trifluoroacetic acid/methanol/water combinations. Through overlaying the UV and radioactivity traces, these HPLC chromatograms confirm that the biomolecule is radiolabeled. Furthermore, they identify and quantify any radiochemical impurities, such as free radiolabeled chelators or radionuclides. However, for rapid measurement of RCP, paper chromatography or ITLC systems are more commonly used. The level of RCP purity required for radiolabeled mAbs and peptides depends on their intended clinical application and the toxicity associated with the impurities. An RCP > 90% would be acceptable for mAbs or peptides labeled with a single γ-photon-emitter (e.g., $^{99m}$Tc or $^{111}$In) or a short-lived positron-emitter (e.g., $^{18}$F or $^{68}$Ga) and intended for tumor imaging. On the other hand, biomolecules labeled with α-emitters (e.g., $^{211}$At) or β-emitters (e.g., $^{131}$I or $^{90}$Y) for targeted *in situ* radiotherapy of malignancies require a higher RCP (>95%) because of the larger amounts of radioactivity administered and the inherent toxicity of the radiochemical impurities, particularly if they concentrate in a sensitive organ (e.g., bone or thyroid). For example, free $^{90}$Y is sequestered *in vivo* by bone, increasing the risk for bone marrow toxicity [255].

### 7.3 Measurement of Immunoreactivity or Receptor-Binding Properties

The immunoreactivity or receptor-binding characteristics of mAbs or peptides conjugated to chelators for radiometal labeling must be assessed because these effect tumor uptake *in vivo* [159]. Similarly, these properties should be evaluated for the final radiolabeled biomolecules. Two parameters are measured: (1) the antigen/receptor-binding affinity and (2) the immunoreactive fraction (IRF) or receptor-binding fraction (RBF). The affinity constant ($K_a$) or dissociation constant ($K_d$) are measured in direct or indirect competition antigen/receptor-binding assays and compared with those of the unmodified mAbs or peptides. Direct binding assays (previously known as Scatchard assays) are performed by incubating increasing concentrations of radiolabeled mAbs or peptides with tumor cells that express the target epitopes/receptors. These assays may also be performed by incubating the radiolabeled biomolecules with purified antigens/receptors coated onto wells in a

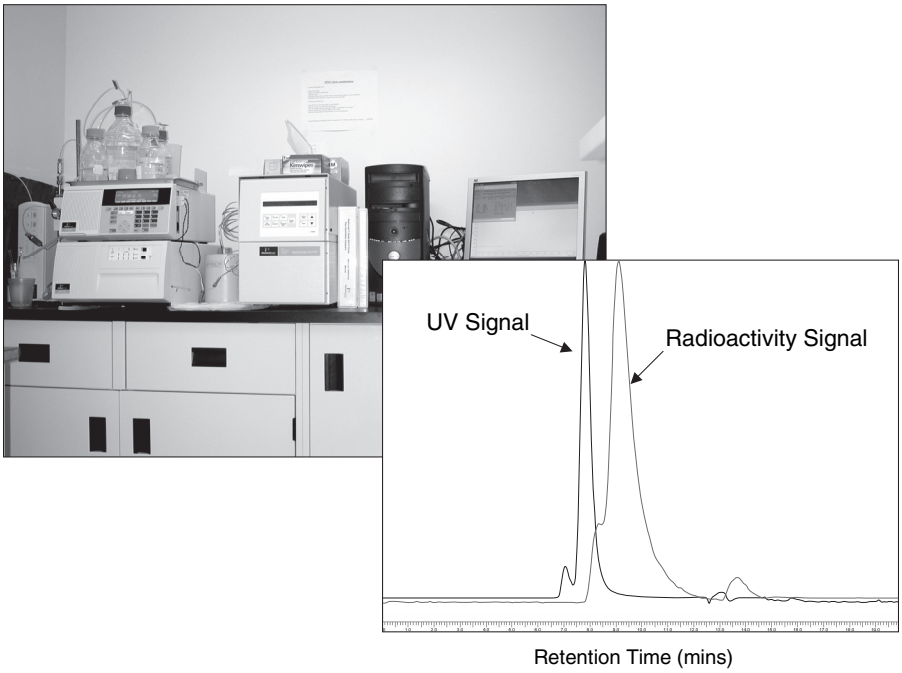


**Figure 11.** *Left Panel*: A size-exclusion HPLC system with a diode array UV detector and FSA flow-through radioactivity detector for evaluating the radiochemical purity and homogeneity of radiolabeled monoclonal antibodies and peptides. *Right Panel*: Typical HPLC chromatograms obtained for analysis of $^{111}$In–DTPA-conjugated trastuzumab (Herceptin). The radioactivity signal is offset from the UV signal because of the distance of tubing between the two detectors, which are in sequence. The radioactivity signal has less resolution than the UV signal due to its larger flow cell. $^{111}$In–DTPA–trastuzumab IgG in this example has less than 5% free $^{111}$In–DTPA impurity (indicated by small peak on the radioactivity trace at $t_R$ = 13.8 minutes). The major peak with $t_R$ of 7.7 minutes in the UV trace and 9.5 minutes in the radioactivity trace represents a $^{111}$In–DTPA–trastuzumab IgG monomer. The smaller peak with $t_R$ of 7.1 minutes in the UV trace and 8.25 minutes in the radioactivity trace represents a $^{111}$In–DTPA–trastuzumab IgG dimer. (Data generously provided by Kristin McLarty.) (This figure is available in full color at ftp://ftp.wiley.com/public/sci_ tech_med/pharmaceutical_biotech/.)

microELISA plate. The binding of the radiolabeled biomolecules to the cells is measured in the absence [total binding (TB)] or presence [nonspecific binding (NSB)] of an excess (e.g., 100 nM) of non-radiolabeled biomolecules to saturate the epitopes/receptors on the cells. Specific binding (SB) is obtained by subtracting NSB from TB and is plotted versus the concentration of radiolabeled mAbs or peptides (Figure 12). The curve is fitted to a direct antigen/receptor-binding model using nonlinear fitting software (e.g., Prism; Graphpad Software, San Diego, CA) and the $K_a$ and maximum number of binding sites/cell ($B_{max}$) estimated. Competition radioligand binding assays can be performed by measuring the binding of radiolabeled mAbs or peptides to cells displaying the target epitopes or receptors in the presence of increasing concentrations of unmodified biomolecules. The $K_d$ is then estimated from the concentration of unmodified antibodies or peptides required

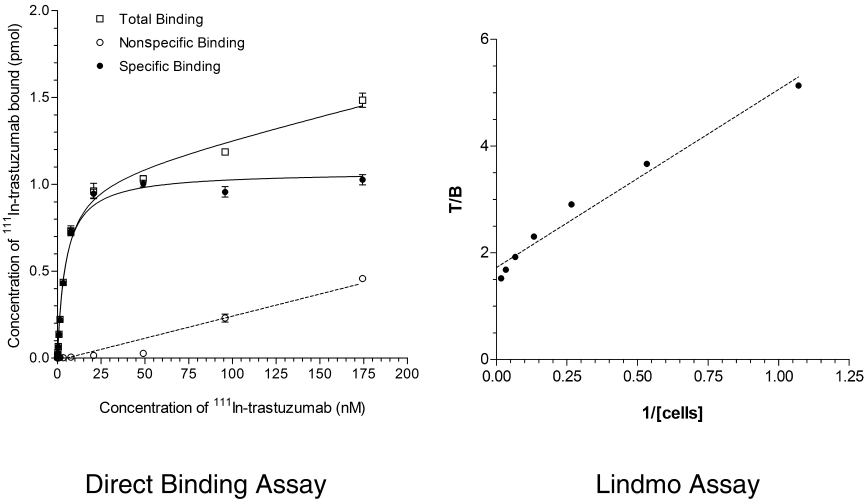


**Figure 12.** *Left Panel*: Results for a direct binding assay for a $^{111}$In–DTPA–trastuzumab (Herceptin) IgG incubated with SK–BR-3 human breast cancer cells overexpressing HER2/neu. The $K_a$ and $B_{max}$ in this example were $1.2 \times 10^8$ L/mole and $1.0 \times 10^6$ receptors/cell, respectively. *Right Panel*: Results for measurement of the immunoreactive fraction (IRF) of a $^{111}$In–DTPA–trastuzumab IgG incubated with SK–BR-3 cells using the Lindmo method [256]. In this example, the intercept on the ordinate (1/IRF) was 1.7 and the IRF was therefore 0.59. (Data generously provided by Kristin McLarty.)

to displace 50% of the initial binding of the radiolabeled biomolecules to the cells. Finally, the IRF or RBF of radiolabeled mAbs or peptides may be measured by the Lindmo assay [256]. In this assay, a small amount of radiolabeled biomolecules (e.g., 5–10 ng) is incubated with increasing concentrations of cancer cells that display the target epitopes/receptors. The total radioactivity counts added divided by the bound counts (T/B) is plotted versus the inverse of the cell concentration (1/[cells]). The intercept on the ordinate of this plot is 1/IRF or 1/RBF (Figure 12, right panel). The IRF or RBF provides information that is different than that of $K_a$ or $K_d$, in that these parameters describe the fraction of radiolabeled biomolecules that can bind their target epitopes/receptors at "infinite antigen excess," irrespective of their binding affinity. Ideally, measurement of $K_a$ or $K_d$ as well as determination of IRF or RBF should be performed to assess immunoreactivity or receptor-binding characteristics.

### 7.4 Evaluation of *In Vitro* and *In Vivo* Stability

Studies to evaluate the *in vitro* and *in vivo* stability of radiolabeled mAbs and peptides are an integral part of their development. Kits used to prepare radiolabeled biomolecules (Section 7.7) must meet specifications established for labeling efficiency, immunoreactivity/receptor binding, homogeneity, sterility and apyrogenicity, and other quality control tests over the expected storage period [145]. Similarly, the radiolabeled biomolecule must maintain the specified level of RCP over the interval from the time of labeling until it is administered. The results of these stability studies will define the expiry times for the kits and/or radiolabeled mAbs

or peptides [145]. The stability of the radiolabeled biomolecules *in vitro* in biologically relevant media such as plasma or serum must also be determined. The anticipated mechanism of loss of radiolabel *in vivo* should be taken into account in designing these studies. For example, studies examining the transchelation of radiometal to transferrin in serum/plasma are necessary for $^{111}$In- or $^{67}$Ga-labeled biomolecules [151], whereas studies that measure transchelation to albumin or other copper-binding proteins are required for $^{64}$Cu/$^{67}$Cu-labeled biomolecules [224]. Cysteine-challenge studies are required for $^{99m}$Tc-labeled biomolecules, especially those labeled using the Schwartz technique [94, 257]. Frequently, stability *in vitro* does not predict stability *in vivo*, because the radiolabeled biomolecules may be catabolized in tissues, resulting in release and redistribution of the radionuclide. For example, radioiodinated proteins are stable *in vitro* in serum/plasma but *in vivo* may be internalized by tumor and normal cells, in which they are rapidly catabolized by lysosomal proteases and deiodinases, resulting in release and redistribution of free radioiodine. This instability *in vivo* can be addressed by employing residualizing radio-iodination techniques (Section 5.1). Nevertheless, due to the limitations of *in vitro* stability studies, it is important to evaluate the stability of radiolabeled mAbs and peptides *in vivo* by chromatographic analysis of plasma/serum samples to identify any circulating radioactive catabolites, as well as by monitoring changes in tissue distribution of radioactivity over time by imaging or biodistribution studies (see Section 7.5). In these studies, it is critical to sample tissues that are known to sequester the radionuclide, should it be released from the biomolecule, i.e., bone for $^{90}$Y or $^{177}$Lu and thyroid and stomach for iodine radionuclides. Noninvasive SPECT or PET imaging studies in humans allow these stability studies to be continued into clinical trials of the radiolabeled biomolecules.

### 7.5 Preclinical Biodistribution, Tumor Imaging, and Dosimetry Studies

It is important to evaluate the tumor and normal tissue distribution of radiolabeled biomolecules in an animal model. For radiolabeled mAbs and peptides intended to target solid tumors, these studies are performed in athymic mice implanted subcutaneously (s.c.) with human tumor xenografts [18, 108, 120, 132, 258]. In the case of hematological malignancies (e.g., B-cell lymphoma or leukemias), athymic or non-obese diabetic severe combined immunodeficient (NOD-scid) mice may be engrafted s.c. or inoculated intravenously (i.v.) with malignant cells to establish a tumor model [259–261]. Tumor-bearing mice are injected i.v. with 25–100 μCi (0.9–3.7 MBq) of radiolabeled mAbs or peptides. Control groups may consist of (1) mice bearing the same tumors but injected with radiolabeled nonspecific biomolecules that do not recognize the target epitopes/receptors, (2) mice bearing these tumors but injected with the radiolabeled mAbs or peptides mixed with a large excess of unlabeled biomolecules to saturate the epitopes/receptors, or (3) mice implanted with a tumor that does not express the target epitopes/receptors and receiving the radiolabeled mAbs or peptides. Goups of five to six mice are sacrificed at selected times postinjection (p.i.) depending on the anticipated pharmacokinetics. The tumor and samples of blood and normal tissues are obtained and weighed, and the radioactivity in each is measured by γ-scintillation counting. For pure β-emitters (e.g., $^{90}$Y), liquid scintillation counting may be used to measure the radioactivity in tissues, or alternatively, the Bremstrahlung radiation caused by interaction of

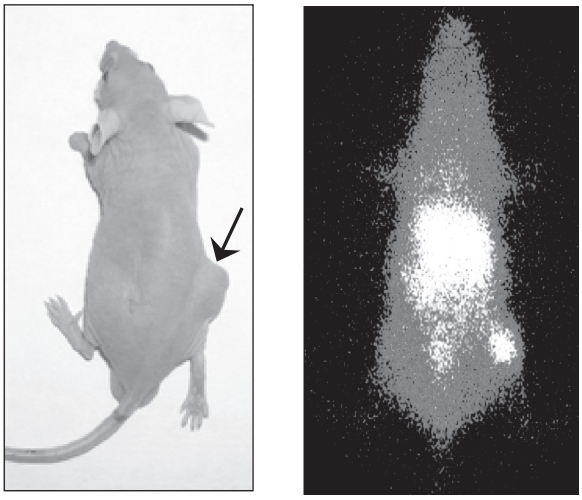

**Figure 13.** Imaging of an athymic mouse bearing a subcutaneous HER2/neu-positive BT-474 human breast cancer xenograft (arrow) at 72 hours postintravenous (tail vein) injection of $^{111}$In-DTPA-trastuzumab (Herceptin). The tumor is clearly visualized using the radiopharmaceutical. (Image generously provided by Kristin McLarty and Deborah Scollard.) (This figure is available in full color at ftp://ftp.wiley.com/public/sci_ tech_ med/pharmaceutical_biotech/.)

the β-particles with the tissues can be measured in a γ-counter. The tumor and normal tissue uptake is expressed as the percent injected dose per gram of tissue (% i.d./g). The tumor/blood (T/B) and T/NT ratios are calculated. Tumor uptake varies from 1% to 2% i.d./g for radiolabeled peptides to as high as 10% to 20% i.d./g for intact IgG mAbs [18]. Tumor uptake should be significantly greater than that in control groups of mice. The T/B ratio should be >2:1 and, ideally, as high as 5:1 to 10:1. In addition to biodistribution studies, imaging studies of γ-emitting or positron-emitting biomolecules may be performed using dedicated high-resolution (1–2 mm) microSPECT or microPET small animal imaging devices (Figure 13) [262, 263]. These images can be quantified by region-of-interest (ROI) analysis and are useful to follow any changes in the biodistribution of radioactivity in an individual animal over time. In addition, preclinical small animal imaging studies provide "proof-of-principle" for radiolabeled biomolecules intended for tumor imaging applications in humans. The results of biodistribution and imaging studies can be used to predict the radiation absorbed doses to organs in humans for subsequent clinical trials using computer software such as OLINDA [264, 265].

### 7.6 Preclinical Studies to Evaluate Anti-Tumor Effects and Normal Tissue Toxicity

Radiolabeled mAbs and peptides intended for RIT or PDRT of malignancies need to be first evaluated preclinically in mouse tumor xenograft models. These studies involve administration of increasing doses of the radiolabeled biomolecules to mice and monitoring the tumor growth or survival of the animals over time [258].

Control groups of mice receive either the formulation vehicle (e.g., saline) or the nonspecific radiolabeled mAbs of the same class and isotype or peptides that do not recognize the target epitopes/receptors. Generalized and gastrointestinal toxicity is monitored by weighing the animals every few days to identify any weight loss (>10%), whereas bone marrow toxicity is assessed by determining leukocyte, red blood cell, and platelet counts in blood samples. Liver toxicity is evaluated by following serum transaminases, whereas kidney toxicity is determined by measuring serum creatinine. Hematopoietic toxicity, if present, is usually observed within 2–3 weeks after injection of the radiolabeled biomolecules and normalizes within 6–8 weeks, whereas liver and kidney toxicity may require an extended observation period (4–8 weeks). An ideal radiotherapeutic agent will exhibit clearly evident and specific antitumor effects with minimal–moderate and manageable normal tissue toxicity [258].

### 7.7 Kit Formulation and Associated Pharmaceutical Testing

Labeling methods for mAbs and peptides that involve the chelation of radiometals (e.g., $^{99m}$Tc, $^{111}$In, $^{68}$Ga, $^{90}$Y, $^{177}$Lu, and $^{64}$Cu) allow formulation of radiopharmaceutical kits [99, 145]. These kits consist of a unit dose of a solution of the metal chelator–biomolecule conjugates dispensed into a glass vial sealed with a rubber septum and aluminum crimp. The kits may be stored at 2–8°C, kept frozen at –10°C, or lyophilized, depending on the particular mAb or peptide. Radiolabeling is performed simply by adding a buffered solution of the radionuclide to the vial and incubating for a predetermined period of time and temperature. The kits are designed such that the labeling efficiency is >90–95%, thus requiring no post-labeling purification and only minimal quality control testing (e.g., assay for total radioactivity, pH measurement, and RCP testing). Other quality control testing procedures are performed on the kit formulation before its use for labeling biomolecules. These tests include protein/peptide purity and homogeneity, evaluation of immunoreactivity or receptor-binding properties, estimation of the level of chelator substitution, measurement of protein concentration and pH, and stability studies. Key pharmaceutical tests such as the USP Sterility Test and USP Bacterial Endotoxins Test also need to be performed on the kits. The radiolabeled biomolecules may be tested retrospectively for sterility and apyrogenicity, provided that the kits have been validated using trial batches to produce a sterile, apyrogenic product.

## 8 SUMMARY

Radiolabeled mAbs and peptides offer the opportunity for molecular imaging of tumors in order to probe their phenotypic properties. Various strategies have been developed for labeling these biomolecules with single γ-photon-emitters or positron-emitters for SPECT and PET imaging, respectively. These approaches can be extended for labeling biomolecules with α- or β-emitters or radionuclides that emit Auger and conversion electrons for targeted *in situ* radiotherapy of malignancies. Establishment of specifications for radiolabeled mAbs and peptides supported by comprehensive characterization testing procedures assures the quality of these novel biopharmaceuticals for human evaluation in clinical trials.

## ACKNOWLEDGMENT

The author would like to acknowledge the financial support for research originating in his laboratory at the University of Toronto described in this chapter. Funding was received from the U.S. Department of Defense Breast Cancer Research Program, the Canadian Institutes of Health Research, the Canadian Breast Cancer Research Alliance, the Canadian Breast Cancer Foundation, the Cancer Research Society, Inc., the Susan G. Komen Breast Cancer Foundation, the Natural Sciences and Engineering Research Council of Canada, the James Birrell Neuroblastoma Research Fund, and the Ontario Cancer Research Network through funding provided by the Province of Ontario. The author would also like to acknowledge the outstanding contributions by the graduate students, post-doctoral fellows, and research technicians who have been members of his research group over the years.

## REFERENCES

1. Ferlay J, Bray F, Pisani P, et al. (2000). Cancer incidence, prevalence and mortality worldwide. *IARC CancerBase, 5th ed*: IARC Press, Lyon, France.
2. Reilly R M (2005). Biomolecules as targeting vehicles for in situ radiotherapy of malignancies. In J Knaeblein, R Mueller, (eds.), *Modern Biopharmaceuticals: Design, Development and Optimization.* Wiley-VCH, Weinheim, Germany, pp. 497–526.
3. Goldenberg D M (1989). Future role of radiolabeled monoclonal antibodies in oncological diagnosis and therapy. *Semin. Nucl. Med.* 19(4):332–339.
4. Baselga J (2001). Clinical trials of Herceptin (trastuzumab). *Eur. J. Cancer.* 37: S18–S24.
5. Behr T M, Béhé M, Wörmann B (2001). Trastuzumab and breast cancer. *N. Engl. J. Med.* 345:995–996.
6. Krenning E P, Kwekkeboom D J, Bakker W H, et al. (1993). Somatostatin receptor scintigraphy with [$^{111}$In-DTPA-D-Phe$^{1}$]- and [$^{123}$I-Tyr$^{3}$]-octreotide: The Rotterdam experience with more than 1000 patients. *Eur. J. Nucl. Med.* 20:716–731.
7. Reilly R M, Sandhu J, varez-Diez T M, et al. (1995). Problems of delivery of monoclonal antibodies. Pharmaceutical and pharmacokinetic solutions. *Clin. Pharmacokinet.* 28(2):126–142.
8. Yokota T, Milenic D E, Whitlow M, et al. (1992). Rapid tumor penetration of a single-chain Fv and comparison with other immunoglobulin forms. *Cancer Res.* 52(12): 3402–3408.
9. Saga T, Neumann R D, Heya T, et al. (1995). Targeting cancer micrometastases with monoclonal antibodies: a binding-site barrier. *Proc. Natl. Acad. Sci. USA.* 92(19):8999–9003.
10. van Osdol W, Fujimori K, Weinstein J N (1991). An analysis of monoclonal antibody distribution in microscopic tumor nodules: consequences of a "binding-site barrier". *Cancer Res.* 51:4776–4784.
11. Behr T M, Goldenberg D M, Becker W (1998). Reducing the renal uptake of radiolabeled antibody fragments and peptides for diagnosis and therapy: Present status, future prospects and limitations. *Eur. J. Nucl. Med.* 25:201–212.
12. Jamar F, Barone R, Mathieu I, et al. (2003). $^{86}$Y-DOTA$^{0}$-D-Phe$^{1}$-Tyr3-octreotide (SMT487)- a phase 1 clinical study: pharmacokinetics, biodistribution and renal pro-

tective effect of different regimens of amino acid co-infusion. *Eur. J. Nucl. Med.* 30:510–518.

13. Pagani M, Stone-Elander S, Larsson S A (1997). Alternative positron emission tomography with non-conventional positron emitters: effects of their physical properties on image quality and potential clinical applications. *Eur. J. Nucl. Med.* 24:1301–1327.
14. Xue L Y, Butler N J, Makrigiorgos G M, Adelstein S J, Kassis A I (2002). Bystander effect produced by radiolabeled tumor cells in vivo. *J. Nucl. Med.* 43(Suppl): 276P–277P.
15. Finn R, Cheung N-KV, Divgi C, et al. (1991). Technical challenges associated with the radiolabeling of monoclonal antibodies utilizing short-lived, positron emitting radionuclides. *Nucl. Med. Biol.* 18:9–13.
16. Hunter W M, Greenwood F C (1962). Preparation of $^{131}$I labeled human growth hormone of high specific activity. *Nature.* 194:495–496.
17. Fraker P J, Speck J C, Jr (1978). Protein and cell membrane iodinations with a sparingly soluble chloroamide, 1,3,4,6-tetrachloro-3a,6a-diphrenylglycoluril. *Biochem. Biophys. Res. Commun.* 80(4):849–857.
18. Reilly R M, Kiarash R, Sandhu J, et al. (2000). A comparison of EGF and MAb 528 labeled with 111In for imaging human breast cancer. *J. Nucl. Med.* 41(5):903–911.
19. Zalutsky M R, Narula A S (1987). A method for the radiohalogenation of proteins resulting in decreased thyroid uptake of radioiodine. *Appl. Radiat. Isot.* 38:1051–1055.
20. Zalutsky M R, Narula A S (1988). Radiohalogenation of a monoclonal antibody using an N-succinimidyl 3-(tri-n-butylstannyl) benzoate intermediate. *Cancer Res.* 48:1446–1450.
21. Zalutsky M R, Noska M A, Colapinto E V, et al. (1989). Enhanced tumor localization and in vivo stability of a monoclonal antibody radioiodinated using N-succinimidyl 3-(tri-n-butylstannyl) benzoate. *Cancer Res.* 49:5543–5549.
22. Garg S, Garg P K, Zalutsky M R (1991). N-succinimidyl 5-(trialkylstannyl)-3-pyridinecarboxylates: A new class of reagents for protein radioiodination. *Bioconjug. Chem.* 2:50–58.
23. Reist C J, Garg P K, Alston K L, et al. (1996). Radioiodination of internalizing monoclonal antibodies using N-succinimidyl 5-iodo-3-pyridinecarboxylate. *Cancer Res.* 56:4970–4977.
24. Reist C J, Archer G E, Wikstrand C J, et al. (1997). Improved targeting of an anti-epidermal growth factor receptor variant III monoclonal antibody in tumor xenografts after labeling using N-succinimidyl 5-iodo-3-pyridinecarboxylate. *Cancer Res.* 57: 1510–1515.
25. Garg P K, Alston K L, Welsh P C, et al. (1996). Enhanced binding and inertness to dehalogenation of α-melanotropic peptides labeled using N-succinimidyl 3-iodobenzoate. *Bioconjug. Chem.* 7:233-239.
26. Vaidyanathan G, Affleck D J, Li J, et al. (2001). A polar substituent-containing acylation agent for the radioiodination of internalizing monoclonal antibodies: N-succinimidyl 4-guanidinomethyl-3-[$^{131}$I]iodobenzoate ([$^{131}$I]SGMIB). *Biocon-jug. Chem.* 12:428–438.
27. Vaidyanathan G, Affleck D J, Bigner D D, et al. (2002). Improved xenograft targeting of tumor-specific anti-epidermal growth factor receptor variant III antibody labeled using N-succinimidyl 4-guanidinomethyl-3-iodobenzoate. *Nucl. Med. Biol.* 29:1–11.
28. Shankar S, Vaidyanathan G, Affleck D, et al. (2003). N-succinimidyl 3-[$^{131}$I]iodo-4-phosphonomethylbenzoate ([$^{131}$I]SIPMB), a negatively charged substituent-bearing

acylation agent for the radioiodination of peptides and mAbs. *Bioconjug. Chem.* 14:331–341.

29. Govindan S V, Mattes M J, Stein R, et al. (1999). Labeling of monoclonal antibodies with diethylenetriaminepentaacetic acid-appended radioiodinated peptides containing D-amino acids. *Bioconjug. Chem.* 10:231–240.
30. Franano F N, Edwards W B, Welch M J, et al. (1994). Metabolism of receptor targeted $^{111}$In-DTPA-glycoproteins: Identification of $^{111}$In-DTPA-ε-lysine as the primary metabolic and excretory product. *Nucl. Med. Biol.* 21:1023–1034.
31. Dumas P, Maziere B, Autissier N, et al. (1973). Specificite de L'iodotyrosine desiodase des microsomes thyroidiens et hepatiques. *Biochim. Biophys. Acta.* 293:36–47.
32. Stein R, Govindan S V, Mattes M J, et al. (2003). Improved iodine radiolabels for monoclonal antibody therapy. *Cancer Res.* 63:111–118.
33. Govindan S V, Griffiths G L, Stein R, et al. (2005). Clinical-scale radiolabeling of a humanized anticarcinoembryonic antigen monoclonal antibody, hMN-14, with residualizing $^{131}$I for use in radioimmunotherapy. *J. Nucl. Med.* 46:153–159.
34. Foulon C F, Reist C J, Bigner D D, et al. (2000). Radioiodinatin via D-amino acid peptide enhances cellular retention and tumor xenograft targeting of an internalizing anti-epidermal growth factor receptor variant III monoclonal antibody. *Cancer Res.* 60:4453–4460.
35. Thorpe S E, Baynes J W, Chroneos Z C (1993). The design and application of residualizing labels for studies of protein catabolism. *FASEB.* 7:399–405.
36. Stein R, Goldenberg D M, Thorpe S R, et al. (1995). Effects of radiolabeling monoclonal antibodies with a residualizing iodine radiolabel in the accretion of radioisotope in tumors. *Cancer Res.* 55:3132–3139.
37. Reist C J, Archer G E, Kurpad S N, et al. (1995). Tumor-specific anti-epidermal growth factor receptor variant III monoclonal antibodies: use of the tyramine-cellobiose radioiodination method enhances cellular retention and uptake in tumor xenografts. *Cancer Res.* 55:4375–4382.
38. Zalutsky M R, Xu F J, Yu Y, et al. (1999). Radioiodinated antibody targeting of the HER-2/neu oncoprotein: Effects of labeling method on cellular processing and tissue distribution. *Nucl. Med. Biol.* 26:781–790.
39. Sundberg Ä L, Blomquist E, Carlsson J, et al. (2003). Cellular retention of radioactivity and increased radiation dose. Model experiments with EGF-dextran. *Nucl. Med. Biol.* 30:303–315.
40. Townsend D W, Carney J P J, Yap J T, et al. (2004). PET/CT today and tomorrow. *J. Nucl. Med.* 45(suppl.):4S–14S.
41. Chatziionnou A F (2002). Molecular imaging of small animals with dedicated PET tomographs. *Eur. J. Nucl. Med.* 29:98–114.
42. Gambhir S S, Czernin J, et al. (2001). A tabulated summary of the FDG PET literature. *J. Nucl. Med.* 42(suppl.):1S–93S.
43. Bakir M A, Eccles S A, Babich J W, et al. (1992). c-erbB2 Protein overexpression in breast cancer as a target for PET using iodine-124-labeled monoclonal antibodies. *J. Nucl. Med.* 33:2154–2160.
44. Larson S M, Pentlow K S, Volkow N D, et al. (1992). PET scanning of iodine-124-3F9 as an approach to tumor dosimetry during treatment planning for radioimmunotherapy in a child with neuroblastoma. *J. Nucl. Med.* 33:2020–2023.
45. Lee F T, Hall C, Rigopoulos A, et al. (2001). Immuno-PET of human colon xenograft-bearing BALB/c nude mice using $^{124}$I-CDR-grafted humanized A33 monoclonal antibody. *J. Nucl. Med.* 42:764–769.

46. Verel I, Visser G W M, Vosjan M J W D, et al. (2004). High-quality $^{124}$I-labelled monoclonal antibodies for use as PET scouting agents prior to $^{131}$I-radioimmunotherapy. *Eur. J. Nucl. Med. Mol. Imaging.* 31:1645–1652.
47. Robinson M K, Doss M, Shaller C, et al. (2005). Quantitative immuno-positron emission tomography imaging of HER2-positive tumor xenografts with an iodine-124 labeled anti-HER2 diabody. *Cancer Res.* 65:1471–1478.
48. Sundaresan G, Yazaki P J, Shively J E, et al. (2003). $^{124}$I-labeled engineered anti-CEA minibodies and diabodies allow high-contrast, antigen-specific small-animal PET imaging of xenografts in athymic mice. *J. Nucl. Med.* 44:1962–1969.
49. Iozzo P, Osman S, Glaser M, et al. (2002). *In vivo* imaging of insulin receptors by PET: preclinical evaluation of iodine-125 and iodine-124 labelled human insulin. *Nucl. Med. Biol.* 29:73–82.
50. Eary J F (2001). PET imaging for planning cancer therapy. *J. Nucl. Med.* 42:770–771.
51. Tolmachev V, Lovqvist A, Einarsson L, et al. (1998). Production of $^{76}$Br by a low-energy cyclotron. *Appl. Radiat. Isot.* 49:1537–1540.
52. Lovqvist A, Sundin A, Ahlstrom H, et al. (1995). $^{76}$Br-labeled monoclonal anti-CEA antibodies for radioimmuno positron emission tomography. *Nucl. Med. Biol.* 22:125–131.
53. Sundin J, Tolmachev V, Koziorowski J, et al. (1999). High yield direct $^{76}$Br-bromination of monoclonal antibodies using chloramine-T. *Nucl. Med. Biol.* 26:923–929.
54. Lovqvist A, Sundin A, Roberto A, et al. (1997). Comparative PET imaging of experimental tumors with bromine-76-labeled antibodies, fluorine-18-fluorodeoxyglucose and carbon-11-methionine. *J. Nucl. Med.* 38:1029–1035.
55. Lovqvist A, Sundin A, Ahlstrom H, et al. (1997). Pharmacokinetics and experimental PET imaging of a bromine-76-labeled monoclonal anti-CEA antibody. *J. Nucl. Med.* 38:395–401.
56. Hoglund J, Tolmachev V, Orlova A, et al. (2000). Optimized indirect $^{76}$Br-bromination of antibodies using N-succinimidyl para-[$^{76}$Br]bromobenzoate for radioimmuno PET. *Nucl. Med. Biol.* 27:837–843.
57. Wilbur D S, Hylarides M D (1991). Radiolabeling of a monoclonal antibody with N-succinimidyl para-[$^{77}$Br]bromobenzoate. *Nucl. Med. Biol.* 18:363–365.
58. Mume E, Orlova A, Malmström P-U, et al. (2005). Radiobromination of humanized anti-HER2 monoclonal antibody trastuzumab using N-succinimidyl 5-bromo-3-pyridinecarboxylate, a potential label for immunoPET. *Nucl. Med. Biol.* 32:613–622.
59. Mume E, Orlova A, Larsson B, et al. (2005). Evaluation of ((4-hydroxyphenyl)ethyl) maleimide for site-specific radiobromination of anti-HER2 affibody. *Bioconjug. Chem.* 16:1547–1555.
60. Steffen A-C, Wikman M, Tolmachev V, et al. (2005). In vitro characterization of a bivalent anti-HER-2 affibody with potential for radionuclide based diagnostics. *Cancer Biother. Radiopharm.* 20:239–248.
61. Bruskin A, Sivaev I, Persson M, et al. (2004). Radiobromination of monoclonal antibody using potassium [$^{76}$Br] (4 isothiocyanatobenzyl-ammonio)-bromo-decahydro-closo-dodecaborate (bromo-DABI). *Nucl. Med. Biol.* 31:205–211.
62. Winberg K J, Persson M, Malmström P-U, et al. (2004). Radiobromination of anti-HER2/neu/ErbB-2 monoclonal antibody using the p-isothiocyanatobenzene derivative of the [$^{76}$Br]undecahydro-bromo-7,8-dicarba-nido-undecaborate(1-) ion. *Nucl. Med. Biol.* 31:425–433.
63. Okarvi S M (2001). Recent progress in fluorine-18 labelled peptide radiopharmaceuticals. *Eur. J. Nucl. Med.* 28:929–938.

64. Garg P K, Garg S, Zalutsky M R (1991). Fluorine-18 labeling of monoclonal antibodies and fragments with preservation of immunoreactivity. *Bioconjug. Chem.* 2:44–49.
65. Garg P K, Garg S, Bigner D D, et al. (1992). Localization of fluorine-18-labeled Mel-14 monoclonal antibody $F(ab')_2$ fragment in a subcutaneous xenograft model. *Cancer Res.* 52:5054–5060.
66. Page R L, Garg P K, Garg S, et al. (1994). PET imaging of osteosarcoma in dogs using a fluorine-18-labeled monoclonal antibody Fab fragment. *J. Nucl. Med.* 35:1506–1513.
67. Lang L, Eckelman W C (1994). One-step synthesis of $^{18}$F labeled [$^{18}$F]-N-succinimidyl 4-(fluoromethyl) benzoate for protein labeling. *Appl. Radiat. Isot.* 45:1155–1163.
68. Vaidyanathan G, Zalutsky M R (1994). Improved synthesis of N-succinimidyl 4-[$^{18}$F]fluorobenzoate and its application to the labeling of a monoclonal antibody fragment. *Bioconjug. Chem.* 5:352–356.
69. Page R L, Garg P K, Vaidyanathan G, et al. (1994). Preclinical evaluation and PET imaging of 18F-labeled Mel-14 $F(ab')_2$ fragment in normal dogs. *Nucl. Med. Biol.* 21:911–919.
70. Vaidyanathan G, Zalutsky M R (1997). Fluorine-18-labeled [Nle$^4$,D-Phe$^7$]-α-MSH, an a-melanocyte stimulating hormone analogue. *Nucl. Med. Biol.* 24:171–178.
71. Guhlke S, Coenen H H, Stocklin G (1994). Fluoroacylation agents based on small n.c.a. [$^{18}$F]fluorocarboxylic acids. *Appl. Radiat. Isot.* 45:715–727.
72. Guhlke S, Wester H-J, Bruns C, et al. (1994). (2-[$^{18}$F]fluoropropionyl-(D)phe$^1$)-octreotide, a potential radiopharmaceutical for quantitative somatostatin receptor imaging with PET: synthesis, radiolabeling, *in vitro* validation and biodistribution in mice. *Nucl. Med. Biol.* 21:819–825.
73. Poethko T, Schottelius M, Thumshirn G, et al. (2004). Two-step methodology for high-yield routine radiohalogenation of peptides: $^{18}$F-labeled RGD and octreotide analogs. *J. Nucl. Med.* 45:892–902.
74. Wester H-J, Brockman J, Rosch F, et al. (1997). PET-pharmacokinetics of $^{18}$F-octreotide: A comparison with $^{67}$Ga-DFO- and $^{86}$Y-DTPA-octreotide. *Nucl. Med. Biol.* 24:275–286.
75. Chen X, Park R, Shahinian A H, et al. (2004). $^{18}$F-labeled RGD peptide: initial evaluation for imaging brain tumor angiogenesis. *Nucl. Med. Biol.* 31:179–189.
76. Haubner R, Kuhnast B, Mang C, et al. (2004). [$^{18}$F]galacto-RGD: Synthesis, radiolabeling, metabolic stability, and radiation dose estimates. *Bioconjug. Chem.* 15:61–69.
77. Wester H-J, Hamacher K, Stocklin G (1996). A comparative study of n.c.a. fluorine-18 labeling of proteins via acylation and photochemical conjugation. *Nucl. Med. Biol.* 23:365–372.
78. Chang Y S, Jeong J M, Lee Y-S, et al. (2005). Preparation of $^{18}$F-human serum albumin: A simple and efficient protein labeling method with $^{18}$F using a hydrazone-formation method. *Bioconjug. Chem.* 16:1329–1333.
79. Couturier O, Supiot S, Degraef-Mougin M, et al. (2005). Cancer radioimmunotherapy with alpha-emitting nuclides. *Eur. J. Nucl. Med. Mol. Imaging.* 32:601–614.
80. Visser G W M, Diemer E L, Kaspersen F M (1981). The nature of the astatine-protein bond. *Int. J. Appl. Radiat. Isot.* 32:905–912.
81. Aaij C, Tschroots W R J M, Lindner L, et al. (1975). The preparation of astatine labelled proteins. *Int. J. Appl. Radiat. Isot.* 26:25–30.
82. Vaughan A T M, Fremlin J H (1978). The preparation of astatine labelled proteins using an electrophilic reaction. *Int. J. Nucl. Med. Biol.* 5:229–230.

83. Zalutsky M R, Narula A S (1988). Astatination of proteins using an N-succinimidyl tri-n-butylstannyl benzoate intermediate. *Appl. Radiat. Isot.* 39:227–232.
84. Zalutsky M R, Garg P K, Friedman H S, et al. (1989). Labeling monoclonal antibodies and $F(ab')_2$ fragments with the α-particle-emitting nuclide astatine-211: Preservation of immunoreactivity and *in vivo* localizing capacity. *Proc. Natl. Acad. Sci. USA.* 86:7149–7153.
85. Garg P K, Harrison C L, Zalutsky M R (1990). Comparative tissue distribution in mice of the alpha-emitter $^{211}$At and $^{131}$I as labels of a monoclonal antibody and $F(ab')_2$ fragment. *Cancer Res.* 50:3514–3520.
86. Wilbur D S, Vessella R L, Stray J E, et al. (1993). Preparation and evaluation of *para*-[$^{211}$At]astatobenzoyl labeled anti-renal cell carcinoma antibody A6H $F(ab')_2$. *In vivo* distribution comparison with *para*-[$^{125}$I]iodobenzoyl labeled A6H $F(ab')_2$. *Nucl. Med. Biol.* 20:917–927.
87. Zalutsky M R, Zhai X-C, Alston K L, et al. (2001). High-level production of α-particle-emitting $^{211}$At and preparation of $^{211}$At-labeled antibodies for clinical use. *J. Nucl. Med.* 42:1508–1515.
88. Pozzi O R, Zalutsky M R (2005). Radiopharmaceutical chemistry of targeted radiotherapeutics, Part I: Effects of solvent on the degradation of radiohalogenation precursors by $^{211}$At α-particles. *J. Nucl. Med.* 46:700–706.
89. Pozzi O R, Zalutsky M R (2005). Radiopharmaceutical chemistry of targeted radiotherapeutics, Part 2: Radiolytic effects of $^{211}$At α-particles influence N-succinimidyl 3-$^{211}$At-astatobenzoate synthesis. *J. Nucl. Med.* 46:1393–1400.
90. Reist C J, Foulon C F, Alston K, et al. (1999). Astatine-211 labeling of internalizing anti-EGFRvIII monoclonal antibody using N-succinimidyl 5-[$^{211}$At]astato-3pyridinecarboxylate. *Nucl. Med. Biol.* 26:405–411.
91. Vaidyanathan G, Affleck D, Welsh P, et al. (2000). Radioiodination and astatination of octreotide by conjugation labeling. *Nucl. Med. Biol.* 27:329–337.
92. Aurlien E, Larsen R H, Bruland O S (2000). Demonstration of highly specific toxicity of the α-emitting radioimmunoconjugate $^{211}$At-rituximab against non-Hodgkin's lymphoma cells. *Br. J. Cancer.* 83:1375–1379.
93. Liu S, Edwards D S, Barrett J A (1997). $^{99m}$Tc labeling of highly potent small peptides. *Bioconjug. Chem.* 8:621–636.
94. Reilly R M (1993). Immunoscintigraphy of tumors using $^{99}$Tc$^{m}$-labelled monoclonal antibodies: A review. *Nucl. Med. Commun.* 14(5):347–359.
95. Schwartz A, Steinstrasser A (1987). A novel approach to Tc-99m labeled monoclonal antibodies. *J. Nucl. Med.* 28:721.
96. Mather S J, Ellison D (1990). Reduction-mediated technetium-99m labeling of monoclonal antibodies. *J. Nucl. Med.* 31:692–697.
97. Stopar T G, Milinaric-Rascan I, Fettich J, et al. (2006). $^{99m}$Tc-rituximab radiolabelled by photoactivation: a new non-Hodgkin's lymphoma imaging agent. *Eur. J. Nucl. Med. Mol. Imaging.* 33:53–59.
98. Stalteri M A, Mather S J (1996). Technetium-99m labelling of the anti-tumor antibody PR1A3 by photoactivation. *Eur. J. Nucl. Med.* 23:178–187.
99. Morales A A, Nunez-Gandolff G, Perez N P, et al. (1999). Freeze-dried formulation for direct $^{99m}$Tc-labeling ior-egr/r3 mab: Additives, biodistribution, and stability. *Nucl. Med. Biol.* 26:717–723.
100. Ellman G L (1958). A colorimetric method for determining low concentrations of mercaptans. *Arch. Biochem. Biophys.* 74:443–450.

101. Iznaga-Escobar N, Morales A, Nunez G (1996). Micromethod for quantification of SH groups generated after reduction of monoclonal antibodies. *Nucl. Med. Biol.* 23:641–644.

102. Marks A, Ballinger J R, Reilly R M, et al. (1995). A novel anti-seminoma monoclonal antibody (M2A) labelled with technetium-99m: Potential application for radioimmunoscintigraphy. *Br. J. Urol.* 75(2):225–229.

103. Steffens M G, Oosterwijk E, Kranenborg M H G C, et al. (1999). In vivo and in vitro characterizations of three $^{99m}$Tc-labeled monoclonal antibody G250 preparations. *J. Nucl. Med.* 40:829–836.

104. Vallis K A, Reilly R M, Chen P, et al. (2002). A phase I study of $^{99m}$Tc-hR3 (DiaCIM), a humanized immunoconjugate directed towards the epidermal growth factor receptor. *Nucl. Med. Commun.* 23(12):1155–1164.

105. Torres L A, Perera A, Batista J F, et al. (2005). Phase I/II clinical trial of the humanized anti-EGF-r monoclonal antibody h-R3 labelled with $^{99m}$Tc in patients with tumor of epithelial origin. *Nucl. Med. Commun.* 26:1049–1057.

106. Fichna J, Janecka A (2003). Synthesis of target-specific radiolabeled peptides for diagnostic imaging. *Bioconjug. Chem.* 14:3–17.

107. Abrams M J, Juweid M, tenKate C I, et al. (1990). Technetium-99m-human polyclonal IgG radiolabeled via the hydrazino nicotinamide derivative for imaging focal sites of infection in rats. *J. Nucl. Med.* 31(12):2022–2028.

108. Tang Y, Scollard D, Chen P, et al. (2005). Imaging of HER2/neu expression in BT-474 human breast cancer xenografts in athymic mice using [$^{99m}$Tc]-HYNIC-trastuzumab (Herceptin) Fab fragments. *Nucl. Med. Commun.* 26(5):427–432.

109. Ono M, Arano Y, Mukai T, et al. (2001). Plasma protein binding of $^{99m}$Tc-labeled hydrazino nicotinamide derivatized polypeptides and peptides. *Nucl. Med. Biol.* 28(2):155–164.

110. Ultee M E, Bridger G J, Abrams M J, et al. (1997). Tumor imaging with technetium-99m-labeled hydrazinenicotinamide-Fab′ conjugates. *J. Nucl. Med.* 38: 133–138.

111. Bangard M, Behe M, Guhike S, et al. (2000). Detection of somatostatin receptor-positive tumors using the new $^{99m}$Tc-tricine-HYNIC-D-Phe$^{1}$-Tyr$^{3}$-octrotide: First results in patients and comparison with $^{111}$In-DTPA-D-Phe$^{1}$-octreotide. *Eur. J. Nucl. Med.* 27:628–637.

112. Decristoforo C, Mather S J, Cholewinski W, et al. (2000). $^{99m}$Tc-EDDA/HYNIC-TOC: A new $^{99m}$Tc-labelled radiopharmaceutical for imaging somatostatin receptor-positive tumors: First clinical results and intra-patient comparison with $^{111}$In-labelled octreotide derivatives. *Eur. J. Nucl. Med.* 27:1318–1325.

113. Decristoforo C, Melendez-Alafort L, Sosabowski J K, et al. (2000). $^{99m}$Tc-HYNIC-[Tyr$^{3}$]-octrotide for imaging somatostatin-receptor-positive tumors: Preclinical evaluation and comparison with $^{111}$In-octreotide. *J. Nucl. Med.* 41:1114–1119.

114. Cornelissen B, Kersemans V, Burvenich I, et al. (2005). Synthesis, biodistribution and effects of farnesyltransferase inhibitor therapy on tumur uptake in mice of $^{99m}$Tc labelled epidermal growth factor. *Nucl. Med. Commun.* 26:147–153.

115. Babich J W, Fischman A J (1995). Effect of "co-ligand" on the biodistribution of $^{99m}$Tc-labeled hydrazino nicotinic acid derivatized chemotactic peptides. *Nucl. Med. Biol.* 22:25–30.

116. Decristoforo C, Mather S J (1999). Technetium-99m somatostatin analogues: Effect of labelling methods and peptide sequence. *Eur. J. Nucl. Med.* 26:869–876.

117. Decristoforo C, Mather S J (1999). 99m-Technetium-labelled peptide-HYNIC conjugates: Effects of lipophilicity and stability on biodistribution. *Nucl. Med. Biol.* 26:389–396.

118. Decristoforo C, Cholewinski W, Donnemiller E, et al. (2000). Detection of somatostatin receptor-positive tumors using the new $^{99m}$Tc-tricine-HYNIC-D-Phe$^{1}$-Tyr$^{3}$-octreotide: first results in patients and comparison with $^{111}$In-DTPA-D-Phe1-octreotide. *Eur. J. Nucl. Med.* 27:1580.

119. Ono M, Arano Y, Mukai T, et al. (2001). $^{99m}$Tc-HYNIC-derivatized ternary ligand complexes for $^{99m}$Tc-labeled polypeptides with low in vivo protein binding. *Nucl. Med. Biol.* 28(3):215–224.

120. Tang Y, Wang J, Scollard D A, et al. (2005). Imaging of HER2/neu-positive BT-474 human breast cancer xenografts in athymic mice using $^{111}$In-trastuzumab (Herceptin) Fab fragments. *Nucl. Med. Biol.* 32(1):51–58.

121. Sattelberger A P, Atcher R W (1999). Nuclear medicine finds the right chemistry. *Nat. Biotechnol.* 17:849–850.

122. Waibel R, Alberto R, Willuda J, et al. (1999). Stable one-step technetium-99m labeling of His-tagged recombinant proteins with a novel Tc(I)-carbonyl complex. *Nat. Biotechnol.* 17:897–901.

123. Schibi R, Schwarzbach R, Alberto R, et al. (2002). Steps toward high specific activity labeling of biomolecules for therapeutic application: Preparation of precursor [Re-188($H_2O$)(3)(CO)(3)](+) and synthesis of tailor-made bifunctional ligand systems. *Bioconjug. Chem.* 13:750–756.

124. Blok D, Feitsman H I J, Kooy Y M C, et al. (2004). New chelation strategy allows for quick and clean $^{99m}$Tc-labeling of synthetic peptides. *Nucl. Med. Biol.* 31:815–820.

125. Fritzberg A R, Abrams P G, Beamier P L, et al. (1988). Specific and stable labeling of antibodies with technetium-99m with a diamide diothiolate chelating agent. *Proc. Natl. Acad. Sci. USA.* 85:4025–4029.

126. Rogers B E, Parry J J, Andrews R, et al. (2005). MicroPET imaging of gene transfer with a somatostatin receptor-based reporter gene and $^{94m}$Tc-demotate. *J. Nucl. Med.* 46:1889–1897.

127. Griffiths G L, Goldenberg D M, Jones A L, et al. (1992). Radiolabeling of monoclonal antibodies and fragments with technetium and rhenium. *Bioconjug. Chem.* 3(91):99.

128. Griffiths G L, Goldenberg D M, Knapp Jr. F F, et al. (1991). Direct radiolabeling of monoclonal antibodies with generator-produced rhenium-188 for radioimmunotherapy: Labeling and animal biodistribution studies. *Cancer Res.* 51:4594–4602.

129. Iznaga-Escobar N (2001). Direct radiolabeling of monoclonal antibodies with rhenium-188 for radioimmunotherapy of solid tumors—a review of radiolabeling characteristics, quality control and in vitro stability studies. *Appl. Radiat. Isot.* 54:399–406.

130. John E, Thakur M L, DeFulvio J, et al. (1993). Rhenium-186-labeled monoclonal antibodies for radioimmunotherapy: preparation and evaluation. *J. Nucl. Med.* 34: 260–267.

131. Rhodes B A, Lambert C R, Marek M J, et al. (1996). Re-188 labelled antibodies. *Appl. Radiat. Isot.* 47:7–14.

132. Reilly R M, Ng K, Polihronis J, et al. (1994). Immunoscintigraphy of human colon cancer xenografts in nude mice using a second-generation TAG-72 monoclonal antibody labelled with $^{99}$Tcm. *Nucl. Med. Commun.* 15(5):379–387.

133. Ferro Flores G, Torres-Garcia E, Garcia-Pedroza L, et al. (2005). An efficient, reproducible and fast preparation of $^{188}$Re-anti-CD20 for the treatment of non-Hodgkin's lymphoma. *Nucl. Med. Commun.* 26:793–799.

134. Goldrosen M H, Biddle W C, Pancock J, et al. (1990). Biodistribution, pharmacokinetic, and imaging studies with $^{186}$Re-labeled NR-LU-10 whole antibody in LS174T colonic tumor-bearing mice. *Cancer Res.* 24:7973–7978.

135. Visser G W, Gerretsen M, Herscheid J D, et al. (1993). Labeling of monoclonal antibodies with rhenium-186 using the MAG3 chelate for radioimmunotherapy of cancer: A technical protocol. *J. Nucl. Med.* 34:1953–1963.

136. Zamora P O, Gulhke S, Bender H, et al. (1996). Experimental radiotherapy of receptor-positive human prostate adenocarcinoma with 188Re-RC-160, a directly-radiolabeled somatostatin analogue. *Int. J. Cancer.* 65:214–220.

137. Safavy A, Khazaeli M B, Safavy K, et al. (1999). Biodistribution study of $^{188}$Re-labeled trisuccin-HuCC49 and trisuccin-HuCC49$\Delta CH_2$ conjugates in athymic nude mice bearing intraperitoneal colon cancer xenografts. *Clin. Cancer Res.* 5(suppl.): 2994S–3000S.

138. Smith C J, Sieckman G L, Owen N K, et al. (2003). Radiochemical investigations of [$^{188}$Re$(H_2O)(CO)_3$-diaminopropionic acid-SSS-bombesin(7-14)$NH_2$]: Synthesis, radiolabeling and in vitro/in/vivo GRP receptor targeting studies. *Anticancer Res.* 23:63–70.

139. Lubberink M, Tolmachev V, Widström C, et al. (2002). $^{110m}$In-DTPA-D-Phe$^1$-octreotide for imaging of neuoendocrine tumors using PET. *J. Nucl. Med.* 43:1391–1397.

140. Tolmachev V, Bernhardt P, Forssell-Aronsson E, et al. (2000). $^{114m}$In, a candidate for radionuclide therapy: low-energy cyclotron production and labeling of DTPA-D-Phe$^1$-octreotide. *Nucl. Med. Biol.* 27:183–188.

141. Hnatowich D J, Childs R L, Lanteigne D, et al. (1983). The preparation of DTPA-coupled antibodies radiolabeled with metallic radionuclides: An improved method. *J. Immunol. Meth.* 65:147–157.

142. Krejcarek G E, Tucker K L (1977). Covalent attachment of chelating groups to macromolecules. *Biochem. Biophys. Res. Commun.* 77:581–585.

143. Brechbiel M W, Gansow O A, Atcher R W, et al. (1986). Synthesis of 1-(p-isothiocyanatobenzyl) derivatives of DTPA and EDTA. Antibody labeling and tumor-imaging studies. *Inorg. Chem.* 25:2772–2781.

144. Westerberg D A, Carney P L, Rogers P E, et al. (1988). Synthesis of novel bifunctional chelators and their use in preparing monoclonal antibody conjugates for tumor targeting. *J. Med. Chem.* 32:236–243.

145. Reilly R M, Scollard D A, Wang J, et al. (2004). A kit formulated under good manufacturing practices for labeling human epidermal growth factor with $^{111}$In for radiotherapeutic applications. *J. Nucl. Med.* 45(4):701–708.

146. Zoghbi S S, Neumann R D, Gottschalk A (1986). The ultrapurification of indium-111 for radiotracer studies. *Invest. Radiol.* 21:710–713.

147. Reilly R, Sheldon K, Marks A (1989). Labelling of monoclonal antibodies 10B, 8C, and M2A with indium-111. *Int. J. Appl. Radiat. Isot.* 40(4):279–283.

148. Carney P L, Rogers P E, Johnson D K (1989). Dual isotope study of iodine-125 and indium-111-labeled antibody in athymic mice. *J. Nucl. Med.* 30:374–384.

149. McLarty K, Cornelissen B, Scollard D, et al. (2006). Imaging of HER-2/neu positive human breast cancer xenografts in CD1$^{nu/nu}$ mice using $^{111}$In-labeled trastuzumab (Herceptin). *Canadian Breast Cancer Research Alliance: Reasons for Hope Conf.* Montreal, Quebec, Canada.

150. Hnatowich D J, Griffin T W, Kosciuczyk C, et al. (1985). Pharmacokinetics of an indium-111-labeled monoclonal antibody in cancer patients. *J. Nucl. Med.* 26:849–858.

151. Reilly R M, Marks A, Law J, et al. (1992). *In-vitro* stability of EDTA and DTPA immunoconjugates of monoclonal antibody 2G3 labelled with In-111. *Appl. Radiat. Isot.* 43:961–967.

152. Goodwin D A, Meares C F, McTigue M, et al. (1986). Metal decomposition rates of $^{111}$In-DTPA and EDTA conjugates of monoclonal antibodies *in vivo*. *Nucl. Med. Commun.* 7:831–838.

153. Reilly R M, Maiti P K, Kiarash R, et al. (2001). Rapid imaging of human melanoma xenografts using an scFv fragment of the human monoclonal antibody H11 labelled with $^{111}$In. *Nucl. Med. Commun.* 22(5):587–595.

154. Orlova A, Briskin A, Sjöström A, et al. (2000). Cellular processing of $^{125}$I- and $^{111}$In-labeled epidermal growth factor (EGF) bound to cultured A431 tumor cells. *Nucl. Med. Biol.* 27:827–835.

155. Bakker W H, Albert R, Bruns C, et al. (1991). [$^{111}$In-DTPA-D-Phe$^{1}$]-octreotide, a potential radiopharmaceutical for imaging of somatostatin receptor-positive tumors: Synthesis, radiolabeling and in-vitro validation. *Life. Sci.* 49:1583–1591.

156. Cole W C, DeNardo S J, Meares C F, et al. (1987). Comparative serum stability of radiochelates for antibody radiopharmaceuticals. *J. Nucl. Med.* 28:83–90.

157. Deshpande S V, Subramanian R, McCall M J, et al. (1990). Metabolism of indium chelates attached to monoclonal antibody: Minimal transchelation of indium from benzyl-EDTA chelate in vivo. *J. Nucl. Med.* 31:218–224.

158. Blend M J, Greager J A, Atcher R W, et al. (1988). Improved sarcoma imaging and reduced hepatic activity with indium-111-SCN-Bz-DTPA linked to MoAb 19–24. *J. Nucl. Med.* 29:1810–1816.

159. Reilly R (2005). The immunoreactivity of radiolabeled antibodies—its impact on tumor targeting and strategies for preservation. *Cancer Biother. Radiopharm.* 19:669–672.

160. Rodwell J D, Alvarez V L, Lee C, et al. (1986). Site-specific covalent modification of monoclonal antibodies: In vitro and in vivo evaluation. *Proc. Natl. Acad. Sci. USA.* 83:2632–2636.

161. Abraham R, Moller D, Gabel D, et al. (1991). The influence of periodate oxidation on monoclonal antibody avidity and immunoreactivity. *J. Immunol. Meth.* 144:77–86.

162. King T P, Zhao S W, Lam T (1986). Preparation of protein conjugates via intermolecular hydrazone linkage. *Biochem.* 25(19):5774–5779.

163. Chan C, Sandhu J, Guha A, et al. (2005). A human transferrin-vascular endothelial growth factor (hnTf-VEGF) fusion protein containing an intergrated binding site for $^{111}$In for imaging tumor angiogenesis. *J. Nucl. Med.* 46:1745–1752.

164. Remy S, Reilly R M, Sheldon K, et al. (1995). A new radioligand for the epidermal growth-factor receptor—In-111 labeled human epidermal growth-factor derivatized with a bifunctional metal-chelating peptide. *Bioconj. Chem.* 6(6):683–690.

165. Wang J, Chen P, Su Z F, et al. (2001). Amplified delivery of indium-111 to EGFR-positive human breast cancer cells. *Nucl. Med. Biol.* 28(8):895–902.

166. Manabe Y, Longley C, Furnanski P (1986). High-level conjugation of chelating agents onto immunoglobulins: Use of an intermediary poly(L-lysine)-diethylenetriamine-pentaacetic acid carrier. *Biochim. Biophys. Acta.* 883:460.

167. Kobayashi H, Sato N, Saga T, et al. (2000). Monoclonal antibody-dendrimer conjugates enable radiolabeling of antibody with markedly high specific activity with minimal loss of immunoreactivity. *Eur. J. Nucl. Med.* 27:1334–1339.

168. Wen X, Wu Q-P, Lu Y, et al. (2001). Poly(ethylene glycol)-conjugated anti-EGF receptor antibody C225 with radiometal chelator attached to the termini of polymer chains. *Bioconjug. Chem.* 12:545–553.

169. Wen X, Wu Q-P, Ke S, et al. (2001). Conjugation with $^{111}$In-DTPA-poly(ethylene glycol) improves imaging of anti-EGF receptor antibody C225. *J. Nucl. Med.* 42:1530–1537.

170. Duncan J R, Welch M J (1993). Intracellular metabolism of indium-111-DTPA labeled receptor targeted proteins. *J. Nucl. Med.* 34:1728–1738.

171. Rogers B E, Franano F N, Duncan J R, et al. (1995). Identification of metabolites of $^{111}$In-diethylenetriaminepentaacetic acid-monoclonal antibodies and antibody fragments in vivo. *Cancer Res.* 55:5714–5720S.

172. Akizawa H, Arano Y, Uezono T, et al. (1998). Renal metabolism of $^{111}$In-DTPA-D-Phe$^{1}$-octreotide in vivo. *Bioconjug. Chem.* 9:662–670.

173. Bass L A, Lanahan M V, Duncan J R, et al. (1998). Identification of the soluble in vivo metabolites of indium-111-diethylenetriaminepentaacetic acid-D-Phe$^{1}$-octreotide. *Bioconjug. Chem.* 9:192–200.

174. Lewis M R, Raubitschek A, Shively J E (1994). A facile, water-soluble method for modification of proteins with DOTA. Use of elevated temperature and optimized pH to achieve high specific activity and high chelate stability in radiolabeled immunoconjugates. *Bioconjug. Chem.* 5:565–576.

175. de Jong M, Bakker W H, Krenning E P, et al. (1997). Yttrium-90 and indium-111 labelling, receptor binding and biodistribution of [DOTA$^{0}$,D-Phe$^{1}$,Tyr$^{3}$]octreotide, a promising somatostatin analogue for radionuclide therapy. *Eur. J. Nucl. Med.* 24: 368–371.

176. Williams L E, Lewis M R, Bebb G C, et al. (1998). Biodistribution of $^{111}$In- and $^{90}$Y-labeled DOTA and maleimidocysteinamido-DOTA ocnjugated to chimeric anticarcinoembryonic antigen antibody in xenograft-bearing nude mice: comparison of stable and chemically labile linker systems. *Bioconjug. Chem.* 9:87–93.

177. Tsai S W, Li L, Williams L E, Anderson A L, et al. (2001). Metabolism and renal clearance of $^{111}$In-labeled DOTA-conjugated antibody fragments. *Bioconjug. Chem.* 12:264–270.

178. Li L, Olafsen T, Anderson A L, et al. (2002). Reduction of kidney uptake in radiometal labeled peptide linkers conjugated to recombinant antibody fragments. Site-specific conjugation of DOTA-peptides to a Cys-diabody. *Bioconjug. Chem.* 13:985–995.

179. Cremonesi M, Ferrari M, Zoboli S, et al. (1999). Biokinetics and dosimetry in patients administered with $^{111}$In-DOTA-Tyr$^{3}$-octreotide: Implications for internal radiotherapy with $^{90}$Y-DOTATOC. *Eur. J. Nucl. Med.* 26:877–886.

180. Onthank D C, Liu S, Silva P J, et al. (2004). $^{90}$Y and $^{111}$In complexes of a DOTA-conjugated integrin $\alpha_v\beta_3$ receptor antagonist: different but biologically equivalent. *Bioconjug. Chem.* 15:235–241.

181. Yoo J, Tang L, Perkins T A, et al. (2005). Preparation of high specific activity $^{86}$Y using a small biomedical cyclotron. *Nucl. Med. Biol.* 32:891–897.

182. Chinol M, Hnatowich D J (1987). Generator-produced yttrium-90 for radioimmunotherapy. *J. Nucl. Med.* 28:1465–1470.

183. Harwood S J, Gibbons L K, Goldner P J, et al. (2002). Outpatient radioimmunotherapy with Bexxar. *Cancer.* 94:1358–1362.

184. Garmestani K, Milenic D E, Plascjak P S, et al. (2002). A new and convenient method for purification of $^{86}$Y using a Sr(II) selective resin and comparison of biodistribution of $^{86}$Y and $^{111}$In labeled Herceptin$^{TM}$. *Nucl. Med. Biol.* 29:599–606.
185. Liu S, Edwards D S (2001). Bifunctional chelators for therapeutic lanthanide radiopharmaceuticals. *Bioconjug. Chem.* 12:7–34.
186. Hnatowich D J, Virzi F, Doherty P W (1985). DTPA-coupled antibodies labeled with yttrium-90. *J. Nucl. Med.* 26:503–509.
187. Brouwers A H, van Eerd J E M, Frielink C, et al. (2004). Optimization of radioimmunotherapy of renal cell carcinoma: labeling of monoclonal antibody cG250 with $^{131}$I, $^{90}$Y, $^{177}$Lu or $^{186}$Re. *J. Nucl. Med.* 45:327–337.
188. Griffiths G L, Govindan S V, Sharkey R M, et al. (2003). $^{90}$Y-DOTA-hLL2: An agent for radioimmunotherapy of non-Hodgkin's lymphoma. *J. Nucl. Med.* 44:77–84.
189. Deshpande S V, DeNardo S J, Kukis D L, et al. (1990). Yttrium-90-labeled monoclonal antibody for therapy: Labeling by a new macrocyclic bifunctional chelating agent. *J. Nucl. Med.* 31:473–479.
190. McCall M J, Diril H, Meares C F (1990). Simplified method for conjugating macrocyclic bifunctional chelating agents to antibodies via 2-iminothiolane. *Bioconjug. Chem.* 1:222–226.
191. Lewis M R, Kao J Y, Anderson A-L J, et al. (2001). An improved method for conjugating monoclonal antibodies with N-hydroxysulfosuccinimidyl DOTA. *Bioconjug. Chem.* 12:320–324.
192. Kukis D L, DeNardo S J, DeNardo G L, et al. (1998). Optimized conditions for chelation of yttrium-90-DOTA immunoconjugates. *J. Nucl. Med.* 39:2105–2110.
193. Breeman W A P, de Jong M, Visser T J, et al. (2003). Optimising conditions for radiolabelling of DOTA-peptides with $^{90}$Y, $^{111}$In and $^{177}$Lu at high specific activities. *Eur. J. Nucl. Med. Mol. Imaging.* 30:917–920.
194. Hainsworth J E S, Mather S J (2005). Regressive DOTA labelling performance with indium-111 and yttrium-90 over a week of use. *Eur. J. Nucl. Med. Mol. Imaging.* 32:1348.
195. Salako Q A, O'Donnell R T, DeNardo S J (1998). Effects of radiolysis on yttrium-90-labeled Lym-1 antibody preparations. *J. Nucl. Med.* 39:667–670.
196. Liu S, Edwards D S (2001). Stabilization of $^{90}$Y-labeled DOTA-biomolecule conjugates using gentisic acid and ascorbic acid. *Bioconjug. Chem.* 12:554–558.
197. Liu S, Ellars C E, Edwards D S (2003). Ascorbic acid: Useful as a buffer agent and radiolytic stabilizer for metalloradiopharmaceuticals. *Bioconjug. Chem.* 14:1052–1056.
198. Kosmas C, Snook D, Gooden C S, et al. (1992). Development of humoral immune responses against a macrocyclic chelating agent (DOTA) in cancer patients receiving radioimmunoconjugates for imaging and therapy. *Cancer Res.* 52:904–911.
199. Watanabe N, Goodwin D A, Meares C F, et al. (1994). Immmunogenicity in rabbits and mice of an antibody-chelate conjugate: Comparison of (S) and (R) macrocyclic enantiomers and an acyclic chelating agent. *Cancer Res.* 54:1049–1054.
200. DeNardo G L, Mirick G R, Kroger L A, et al. (1996). Antibody responses to macrocycles in lymphoma. *J. Nucl. Med.* 37:451–456.
201. Perico M E, Chinol M, Nacca A, et al. (2001). The humoral immune response to macrocyclic chelating agent DOTA depends on the carrier molecule. *J. Nucl. Med.* 42:1697–1703.
202. Saha G B (2004). Diagnostic uses of radiopharmaceuticals in nuclear medicine. In *Fundamentals of Nuclear Pharmacy*, 5th ed. Springer, New York, 247–329.

203. Michel R B, Brechbiel M W, Mattes M J (2003). A comparison of 4 radionuclides conjugated to antibodies for single-cell kill. *J. Nucl. Med.* 44:632–640.
204. Maecke H R, Hofmann M, Haberkorn U (2005). $^{68}$Ga-labeled peptides in tumor imaging. *J. Nucl. Med.* 46:172S–178S.
205. Schumacher J, Maier-Borst W (1981). A new $^{68}$Ge/$^{68}$Ga radioisotope generator system for production of 68Ga in dilute HCl. *Int. J. Appl. Radiat. Isot.* 32:31–36.
206. Ugur O, Kothari P J, Finn R D, et al. (2002). Ga-66 labeled somatostatin analogue DOTA-D-Phe1-Tyr3-octreotide as a potential agent for positron emission tomography imaging and receptor mediated internal radiotherapy of somatostatin receptor positive tumors. *Nucl. Med. Biol.* 29:147–157.
207. Smith-Jones P M, Stolz B, Bruns C, et al. (1994). Gallium-67/gallium-68-[DFO]-octreotide-a potential radiopharmaceutical for PET imaging of somatostatin receptor-positive tumors: Synthesis and radiolabeling in vitro and preliminary in vivo studies. *J. Nucl. Med.* 35:317–325.
208. Breeman W A P, de Jong M, de Blois E, et al. (2005). Radiolabelling DOTA-peptides with $^{68}$Ga. *Eur. J. Nucl. Med. Mol. Imaging.* 32:478–485.
209. Henze M, Schuhmacher J, Hipp P, et al. (2001). PET imaging of somatostatin receptors using [$^{68}$Ga]DOTA-D-Phe$^{1}$-Tyr$^{3}$-octreotide: First results in patients with meningiomas. *J. Nucl. Med.* 42:1053–1056.
210. Velikyan I, Beyer G L, Langström B (2004). Microwave-supported preparation of $^{68}$Ga bioconjugates with high specific activity. *Bioconjug. Chem.* 15:554–560.
211. Reubi J C, Schar J-C, Waser B, et al. (2000). Affinity profiles for human somatostatin receptor subtypes SST1-SST5 of somatostatin radiotracers selected for scintigraphic and radiotherapeutic use. *Eur. J. Nucl. Med.* 27:273–282.
212. Hofmann M, Maecke H, Börner A R, et al. (2001). Biokinetics and imaging with the somatostatin receptor PET radioligand $^{68}$Ga-DOTATOC: preliminary data. *Eur. J. Nucl. Med.* 28:1751–1757.
213. Wild D, Macke H R, Waser B, et al. (2005). $^{68}$Ga-DOTANOC: A first compound for PET imaging with high affinity for somatostatin receptor subtypes 2 and 5. *Eur. J. Nucl. Med. Mol. Imaging.* 32:724.
214. Froideveaux S, Calame-Christe M, Schumacher J, et al. (2004). A gallium-labeled DOTA-α-melanocyte-stimulating hormone analog for PET imaging of melanoma metastases. *J. Nucl. Med.* 45:116–123.
215. Velikyan I, Sundberg ÄL, Lindhe Ö, et al. (2005). Preparation and evaluation of $^{68}$Ga-DOTA-hEGF for visualization of EGFR expression in malignant tumors. *J. Nucl. Med.* 46:1881–1888.
216. Blower P J, Lewis J S, Zweit J (1996). Copper radionuclides and radiopharmaceuticals in nuclear medicine. *Nucl. Med. Biol.* 23:957–980.
217. Anderson C J, Jones A L, Bass L A, et al. (1998). Radiotherapy, toxicity and dosimetry of copper-64-TETA-octreotide in tumor-bearing rats. *J. Nucl. Med.* 39:1944–1951.
218. Connett J M, Anderson C J, Guo L Y, et al. (1996). Radioimmunotherapy with a $^{64}$Cu-labeled monoclonal antibody: a comparison with $^{67}$Cu. *Proc. Natl. Acad. Sci. USA.* 93:6814–6818.
219. McCarthy D W, Shefer R E, et al. (1997). Efficient production of high specific activity $^{64}$Cu using a biomedical cyclotron. *Nucl. Med. Biol.* 24:35–43.
220. Obata A, Kasamatsu S, McCarthy D W, et al. (2003). Production of therapeutic quantities of $^{64}$Cu using a 12 MeV cyclotron. *Nucl. Med. Biol.* 30:535–539.
221. Williams H A, Robinson S, Julyan P, et al. (2005). A comparison of PET imaging characteristics of various copper radioisotopes. *Eur. J. Nucl. Med. Mol. Imaging.* 32:1473–1480.

222. McCarthy D W, Bass L A, Cutler P D, et al. (1999). High purity production and potential applications of copper-60 and copper-61. *Nucl. Med. Biol.* 26:351–358.
223. Novak-Hofer I, Schubiger P A (2002). Copper-67 as a therapeutic nuclide for radioimmunotherapy. *Eur. J. Nucl. Med.* 29:821–830.
224. Cole W C, DeNardo S J, Meares C F, et al. (1986). Serum stability of $^{67}$Cu chelates: comparison with $^{111}$In and $^{57}$Co. *Nucl. Med. Biol.* 13:363–368.
225. Anderson C J, Schwarz S W, Connett J M, et al. (1995). Preparation, biodistribution and dosimetry of copper-64-labeled anti-colorectal carcinoma monoclonal antibody fragments 1A3-F(ab′)$_2$. *J. Nucl. Med.* 36:850–858.
226. Lewis J S, Srinivasan A, Schmidt M A, et al. (1999). In vitro and in vivo evaluation of $^{64}$Cu-TETA-Tyr$^3$-octreotate. A new somatostatin analog with improved target tissue uptake. *Nucl. Med. Biol.* 26:267–273.
227. Li W P, Lewis J S, Kim J, et al. (2002). DOTA-D-Tyr$^1$-octreotate: A somatostatin analogue for labeling with metal and halogen radionuclides for cancer imaging and therapy. *Bioconjug. Chem.* 13:721–728.
228. O'Donnell R T, DeNardo G L, Kukis D L, et al. (1999). A clinical trial of radioimmunotherapy with $^{67}$Cu-2-IT-BAT-Lym-1 for non-Hodgkin's lymphoma. *J. Nucl. Med.* 40:2014–2020.
229. Philpott G W, Schwarz S W, Anderson C J, et al. (1995). RadioimmunoPET: detection of colorectal carcinoma with positron-emitting copper-64-labeled monoclonal antibody. *J. Nucl. Med.* 36:1818–1824.
230. Anderson C J, Dehdashti F, Cutler P D, et al. (2001). $^{64}$Cu-TETA-octreotide as a PET imaging agent for patients with neuroendocrine tumors. *J. Nucl. Med.* 42:213–221.
231. Anderson C J, Pajeau T S, Edwards W B, et al. (1995). In vitro and in vivo evaluation of copper-64-octreotide conjugates. *J. Nucl. Med.* 36:2315–2325.
232. Chen X, Hou Y, Tohme M, et al. (2004). Pegylated Arg-Gly-Asp peptide: $^{64}$Cu labeling and PET imaging of brain tumor $\alpha_v\beta_3$-integrin expression. *J. Nucl. Med.* 45:1776–1783.
233. Thakur M L, Aruva M R, Gariepy J, et al. (2004). PET imaging of oncogene overexpression using $^{64}$Cu-vasoactve intestinal peptide (VIP) analog: comparison with $^{99m}$Tc-VIP analog. *J. Nucl. Med.* 45:1381–1389.
234. Rogers B E, Bigott H M, McCarthy D W, et al. (2003). MicroPET imaging of a gastrin-releasing peptide receptor-positive tumor in a mouse model of human prostate cancer using a $^{64}$Cu-labeled bombesin analogue. *Bioconjug. Chem.* 14:756–763.
235. Rogers B E, Anderson C J, Connett J M, et al. (1996). Comparison of four bifunctional chelates for radiolabeling monoclonal antibodies with copper radioisotopes: biodistribution and metabolism. *Bioconjug. Chem.* 7:511–522.
236. Bass L A, Wang M, Welch M J, et al. (2000). In vivo transchelation of copper-64 from TETA-octreotide to superoxide dismutase in rat liver. *Bioconjug. Chem.* 11:527–532.
237. Mirick G R, O'Donnell R T, DeNardo S J, et al. (1999). Transfer of copper from a chelated $^{67}$Cu-antibody conjugate to ceruloplasmin in lymphoma patients. *Nucl. Med. Biol.* 26:841–845.
238. Milenic D E, Garmestani K, Chappell L L, et al. (2002). In vivo comparison of macrocyclic and acyclic ligands for radiolabeling of monoclonal antibodies with $^{177}$Lu for radioimmunotherapeutic applications. *Nucl. Med. Biol.* 29:431–442.
239. Mulligan T, Carrasquillo J A, Chung Y, et al. (1995). Phase I study of intravenous $^{177}$Lu-labeled CC49 murine monoclonal antibody in patients with advanced adenocarcinoma. *Clin. Cancer Res.* 1:1447–1454.

240. Müller W A, Linzner U, Schäffer E H (1978). Organ distribution studies of lutetium-177 in mouse. *Int. J. Nucl. Med. Biol.* 5:29–31.
241. Breeman W A P, van der Wansem K, Bernard B F, et al. (2003). The addition of DTPA to [$^{177}$Lu-DOTA$^{0}$,Tyr$^{3}$]octreotate prior to administration reduces rat skeleton uptake of radioactivity. *Eur. J. Nucl. Med.* 30(312):315.
242. Alvarez R D, Partridge E E, Khazaeli M B, et al. (1997). Intraperitoneal radioimmunotherapy of ovarian cancer with $^{177}$Lu-CC49: A phase I/II study. *Gynecol. Oncol.* 65:94–101.
243. Chauhan S C, Jain M, Moore E D, et al. (2005). Pharmacokinetics and biodistributon of $^{177}$Lu-labeled multivalent single-chain Fv construct of the pancarcinoma monoclonal antibody CC49. *Eur. J. Nucl. Med. Mol. Imaging.* 32:264–273.
244. McDevitt M R, Dangshe M, Lai L, et al. (2001). Tumor therapy with targeted atomic nanogenerators. *Science.* 294:1537–1540.
245. Miederer M, McDevitt M R, Sgouros G, et al. (2004). Pharmacokinetics, dosimetry, and toxicity of the targetable atomic generator, $^{225}$Ac-HuM195, in nonhuman primates. *J. Nucl. Med.* 45:129–137.
246. Milenic D E, Roselli M, Mirzadeh S, et al. (2001). In vivo evaluation of bismuth-labeled monoclonal antibody comparing DTPA-derived bifunctional chelates. *Cancer Biother. Radiopharm.* 16:133–146.
247. Yao Z, Garmestani K, Wong K J, et al. (2001). Comparative cellular catabolism and retention of astatine-, bismuth-, and lead-radiolabeled internalizing monoclonal antibody. *J. Nucl. Med.* 42:1538–1544.
248. Ballangrud A M, Yang W-H, Palm S, et al. (2004). Alpha-particle emitting atomic generator (actinium-225)-labeled trastuzumab (Herceptin) targeting of breast cancer spheroids: Efficacy versus HER2/neu expression. *Clin. Cancer Res.* 10:4489–4497.
249. Horak E, Hartmann F, Garmestani K, et al. (1997). Radioimmunotherapy targeting of HER2/neu oncoprotein on ovarian tumor using lead-212-DOTA-AE1. *J. Nucl. Med.* 38:1944–1950.
250. Meares C F, McCall M J, Rearden D T, et al. (1984). Conjugation of antibodies with bifunctional chelating agents: isothiocyanate and bromoacetamide reagents: Methods of analysis and subsequent addition of metal ions. *Anal. Biochem.* 142:68.
251. Pippin C J, Parker T A, McMurry T J, et al. (1992). Spectrophotometric method for the determination of a bifunctional DTPA ligand in DTPA-monoclonal antibody conjugates. *Bioconjug. Chem.* 3:342–345.
252. Hartikka M, Vihko P, Södervall M, et al. (1989). Radiolabelling of monoclonal antibodies: Optimization of conjugation of DTPA to $F(ab')_2$-fragments and a novel measurement of the degree of conjugation using Eu(III) labelling. *Eur. J. Nucl. Med.* 15:157–161.
253. Pham D T, Kaspersen F M, Bos E S (1995). Electrophoretic method for the quantitative determination of a benzyl-DTPA ligand in DTPA monoclonal antibody conjugates. *Bioconjug. Chem.* 6:313–315.
254. Dadachova E, Chappell L L, Brechbiel M W (1999). Spectrophotometric method for determination of bifunctional macrocyclic ligands in macrocyclic ligand-protein conjugates. *Nucl. Med. Biol.* 26:977–982.
255. Breeman W A P, de jong M T h M, de Jong M, et al. (2004). Reduction of skeletal accumulation of radioactivity by co-injection of DTPA in [$^{90}$Y-DOTA$^{0}$,Tyr$^{3}$]octreotide solutions containing free $^{90}Y^{3+}$. *Nucl. Med. Biol.* 31:821–824.
256. Lindmo T, Boven E, Cuttitta F, et al. (1984). Determination of the immunoreactive fraction of radiolabelled monoclonal antibodies by linear extrapolation of binding to infinite antigen excess. *J. Immunol. Meth.* 72:77–89.

257. Mardirossian G, Wu C, Rusckowski M, et al. (1992). The stability of $^{99}Tc^{m}$ directly labelled to an Fab′ antibody via stannous ion and mercaptoethanol reduction. *Nucl. Med. Commun.* 13:503–512.

258. Chen P, Cameron R, Wang J, et al. (2003). Antitumor effects and normal tissue toxicity of 111In-labeled epidermal growth factor administered to athymic mice bearing epidermal growth factor receptor-positive human breast cancer xenografts. *J. Nucl. Med.* 44(9):1469–1478.

259. Sharkey R M, Karacay H, Chang C H, et al. (2005). Improved therapy of non-Hodgkin's lymphoma xenografts using radionuclides pretargeted with a new anti-CD20 bispecific antibody. *Leukemia.* 19:1064–1069.

260. Michel R B, Rosario A V, Brechbiel M W, et al. (2003). Experimental therapy of disseminated B-cell lymphoma xenografts with $^{213}$Bi-labeled anti-CD74. *Nucl. Med. Biol.* 30:715–723.

261. Bonnet D, Bhatia M, Wang J Y C, et al. (1999). Cytokine treatment or accessory cells are required to initiate engraftment of purified primitive human hematopoietic cells transplanted at limiting doses into NOD/SCID mice. *Bone. Marrow. Transplant.* 23:203–209.

262. Iwata K, MacDonald L R, Hwang A B, et al. (2002). CT-SPECT for small animal imaging with A-SPECT. *J. Nucl. Med.* 43(suppl.):10P–Abst. No. 35.

263. Lewis J S, Achilefu S, Garbow J R, et al. (2002). Small animal imaging: Current technology and perspectives for oncological imaging. *Eur. J. Cancer.* 38:2173–2188.

264. Stabin M G, Sparks R B, Crowe E (2005). OLINDA/EXM: the second-generation personal computer software for internal dose assessment in nuclear medicine. *J. Nucl. Med.* 46:1023–1027.

265. Reilly R M, Chen P, Wang J, et al. (2006). Preclinical pharmacokinetic, biodistribution, toxicology and dosimetry studies of $^{111}$In-DTPA-hEGF—an Auger electron-emitting radiotherapeutic agent for EGFR-positive breast cancer. *J. Nucl. Med.* In press.

# GENE THERAPY—BASIC PRINCIPLES AND THE ROAD FROM BENCH TO BEDSIDE

Gabor M. Rubanyi
*Cardium Therapeutics, Inc., San Diego, California*

**Chapter Contents**

1 Introduction
2 Basic Principles
  2.1 From Nucleic Acid Delivery to Therapeutic Effects
  2.2 Essential Components and Selected Technical Features of a Gene Therapy Product
  2.3 Disease Targets for Gene Therapy
3 From Bench to Bedside: The Development of Ad5FGF-4 to Treat Patients with Recurrent Angina
  3.1 Preclinical Efficacy
  3.2 Toxicology and Biodistribution
  3.3 cGMP Manufacturing and Characterization of the Gene Therapy Vector
  3.4 Clinical Studies
  3.5 Lessons Learned from Past Failures
4 Conclusions
References

## 1 INTRODUCTION

Human gene therapy (HGT) is defined as the transfer of nucleic acids (DNA or RNA) to somatic cells of a patient, which results in a therapeutic effect, by either (1) correcting genetic defects, (2) overexpressing proteins that are therapeutically useful, or (3) inhibiting the production of "harmful" proteins (e.g., by shRNA).

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

Although it has several theoretical advantages, so far HGT has not delivered the promised results: Convincing clinical efficacy in pivotal phase III clinical trials has not yet been demonstrated in the United States or Europe (although two gene therapy products for cancer treatment have recently been approved and marketed in China). HGT is a complex process, involving multiple steps in the human body (delivery to organs, tissue targeting, cellular trafficking, regulation of gene expression level and duration, biological activity of therapeutic protein, safety of the vector and gene product, to name just a few), most of which are not completely understood.

In most applications, gene therapy represents a new, innovative drug delivery system making use of the technical and scientific advances of the last two decades in microbiology, virology, organic chemistry, molecular biology, biochemistry, cell biology, genetics, genomics, and genetic engineering. It is more than "gene transfer," which is only a part of the complex, multiphase process of identification, manufacturing, preclinical testing, and clinical development of gene therapy products (see Section 3).

The prerequisites of successful HGT include therapeutically suitable genes (with a proven role in pathophysiology of the disease either by its lack, mutation, or overexpression), appropriate gene delivery systems (e.g., viral and nonviral vectors), proof of principle of efficacy and safety in appropriate preclinical models, and suitable manufacturing and analytical processes to provide well-defined HGT products for clinical investigations.

The principle of gene therapy has specific therapeutic advantages over existing therapeutic modalities (such as small molecules or biologics) in certain disease indications. These include (1) correction of the genetic cause of a disease, (2) selective treatment of affected (diseased) cells and tissues (the cells and tissues produce their own "remedy"), and (3) long-term treatment after a single application. It may be best suited for so-called "nondrugable" targets (i.e., the modulation of target molecule is not feasible with small-molecule or protein drugs). Based on these theoretical principles, at the time of its first introduction more than 17 years ago, gene therapy promised to be an effective and safe treatment modality, which will soon cure diseases and replace classic therapies.

Over the past few years, significant progress has been made in various enabling technologies and in the molecular understanding of diseases and the manufacturing of vectors. These advances, combined with the growing experience with gene transfer in humans, promises to contribute to the ultimate success of this new class of therapeutics.

## 2 BASIC PRINCIPLES

### 2.1 From Nucleic Acid Delivery to Therapeutic Effects

Gene therapy consists of multiple biological processes in the body, the exact nature of which is in most cases still unknown (Figure 1). Gene therapy is initiated with the introduction of an appropriate vector (viral or nonviral) either into the body locally (direct tissue injection), into body cavities (e.g., peritoneum or cerebrospinal fluid), or into the bloodstream (systemic delivery). The vector needs to "find"

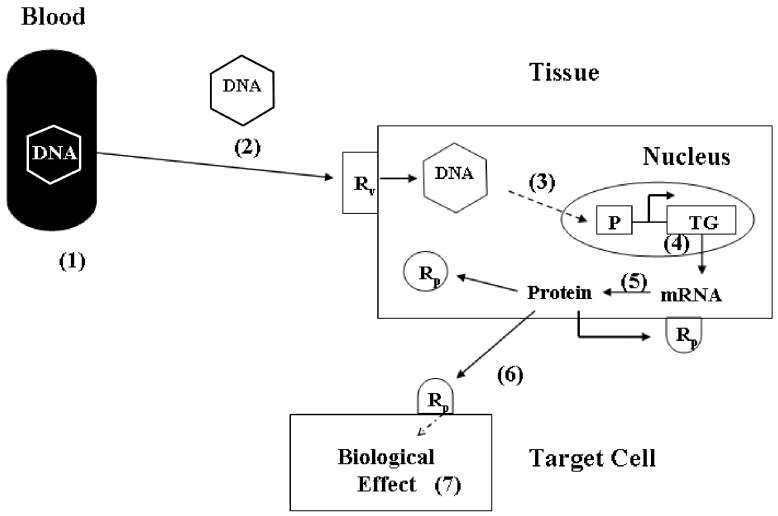


**Figure 1. Multiple biological processes of in vivo gene transfer.** The following main steps are required for successful gene transfer in the human body: (1) vector (viral, nonviral, cell-based) delivery (localized tissue delivery or systemic delivery via blood circulation) of a gene (DNA); (2) vector "recognition" by specific receptors (Rv) on cells in target tissue; (3) uptake of the vector by cells, trafficking to the nucleus and delivery of vector DNA in to the nucleus; (4) transcription (expression) of therapeutic (trans) gene in the nucleus; (5) translation of mRNA into therapeutic protein in the cytoplasm; (6) interaction of therapeutic protein with its receptors (Rp) within the "producing" cell (intracrine mechanism), on the surface of "producing" cell (autocrine mechanism) or on neighboring "target" cell (paracrine mechanism). For some applications, the therapeutic protein enters the circulation and acts distant from the target tissue (endocrine mechanism) (e.g., erythropoietin, coagulation factors VIII and IX, and growth hormone); and (7) after interaction with its receptor, the protein induces a biological effect, which results in therapeutic benefits. (Reproduced with permission from Pergamon Press.)

its target tissue, after which it enters the target cell membrane and traffics through the cytoplasm to reach and enter the nucleus. Once there, the therapeutic (trans)gene needs to be transcribed and the formed mRNA needs to be appropriately translated into the therapeutic protein. The protein then acts on its receptor(s) either on the cell that produced it (intracrine or autocrine mechanism), on neighboring cells (paracrine mechanism), or at distant sites after entering the blood circulation (endocrine mechanism, e.g., erythropoietin, coagulation factors, and growth hormone). Finally, after interacting with its receptor, the protein needs to induce an appropriate biological effect that results in therapeutic benefits. For gene correction or gene knock-down approaches, the steps are similar to those outlined above, except that the last step is modification of the genome ("gene correction") or blockade of mRNA transcription (siRNA/shRNA) of endogenous genes, respectively.

Although the factors needed for successful gene therapy are not different from any new therapeutic modality (which include technical [gene delivery and expression], clinical [therapeutic efficacy and safety], and socioeconomic factors), the specific technical success factors are unique for gene therapy approaches. They

include the choice of appropriate therapeutic gene(s) (with a proven role in the pathomechanism of the disease, specifically targeted and of sufficient potency), gene delivery systems (of sufficient targeting ability, transfection efficiency, and safety), and gene expression regulation systems to control the level and timing of transgene expression.

The therapeutic and socioeconomic success of gene therapy products includes the requirement that the benefits of gene therapy should outweigh the risks and should offer advantages over conventional (usually less expensive) treatments, before this new approach will become accepted in the general medical practice.

## 2.2 Essential Components and Selected Technical Features of a Gene Therapy Product

***Gene Delivery Vectors.*** The goal in the discovery and development of most gene therapy vectors is to deliver and express genes at the appropriate site and at therapeutically meaningful levels in a controlled manner. The first-generation gene therapy vectors have used the ability of viral systems to deliver genetic information to human cells. Attempts are also made to develop nonviral synthetic vectors and hybrid synthetic-viral systems that are potentially safer alternatives for gene delivery. A third approach uses isolated human cells (stem cells, progenitor cells, or somatic cells) as a means to introduce the therapeutic genes into specific human cell populations where the therapeutic product is required.

*Viral Vectors.* Viruses "acquired" numerous biological properties over millions of years of evolution that allow them to effectively recognize and enter cells, traffic within the cytosol to the nucleus, translocate into the nucleus, and express their genes in the host cell (Figure 2).

The most frequently used viral vectors in clinical trials so far are retroviruses and adenoviruses (Table 1). Several other viral vectors are in preclinical development or are under clinical evaluation, including adeno-associated virus (AAV), lentivirus, herpes simplex virus (HSV), and others.

*Retroviruses* can lead to a stable integration of the transfected gene into the host genome and therefore produce long-lasting gene transfer. Replication-deficient retroviruses are produced *in vitro* in specific packaging cells transfected previously with retroviral genes (G, P, E) that have been deleted from the genome of the therapeutic retroviruses. Major limitations of the retroviruses are their low titers, their inability to infect nondividing cells, and the potential risk of insertional mutagenesis. The development of new pseudotyped retroviruses has increased virus titers that will permit more efficient gene transfer.

There are more than 50 known serotypes of *adenoviruses*, but until recently, serotype 5 (Ad5) and to a lesser extent serotype 2 (Ad2) (both class C adenoviruses) have been used as recombinant, nonreplicating gene delivery vectors. *Recombinant Ad5* can be produced in high titer. Ad5 vectors do not lead to stable integration of the transgene into the host genome (only at very low frequency in cell culture [2]), and they usually remain extrachromosomal and cause only a transient transgene expression. Replication-deficient adenoviruses are produced *in vitro* in specific packaging cells that complement gene products (e.g., *E1* and *E3*) deleted from the genome of the therapeutic adenoviruses. They give effective tran-

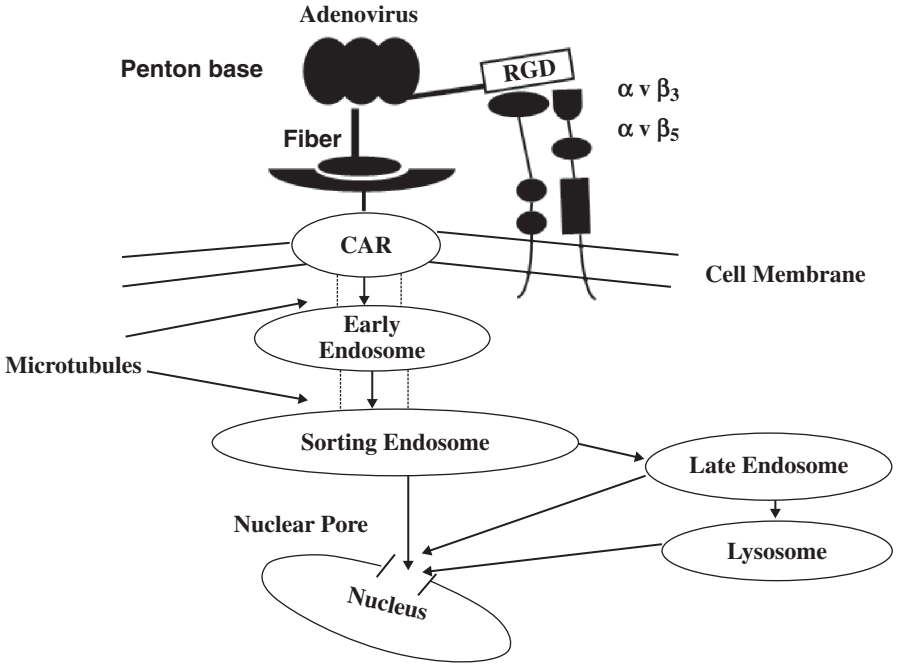


**Figure 2. Cellular recognition, uptake, and trafficking of adenoviral gene delivery vectors.** Receptor-mediated entry of adenovirus into cells depends on interaction of two of its coat proteins with two different cell-surface receptors. The viral fiber knob protein mediates attachment to the cell via the CAR. After attachment, an RGD tripeptide motif in the penton base protein binds to intergrins (αvβ3 or αvβ5), which mediates internalization. Knowledge of these domains and processes helped to design novel strategies to change the natural tropism of adenoviral vector ("re-targeting"; see text for details.) Once inside the cell, the adenovirus effectively and quickly (within less than 60 minutes) reaches the nucleus (via the endosome, microtubules, and the lysosome) and translocates into the nucleus via nuclear pores with the help of the nuclear localization signal (NLS) peptide in the viral coat protein. (Reproduced with permission from Pergamon Press.)

**TABLE 1. Gene Delivery Systems (Vectors) Used in Gene Therapy Clinical Trials (1989–Present)**[1]

| Delivery System | Number of Trials |
|---|---|
| Retrovirus | 254 |
| Adenovirus | 240 |
| Naked DNA | 132 |
| Lipofection | 85 |
| Poxvirus | 52 |
| Vaccinia virus | 30 |
| Herpes simplex virus | 26 |
| Adeno-associated virus | 19 |
| RNA transfer | 10 |

[1] From Ref. 1.

sient gene expression in proliferating and nonproliferating cells, but first-generation adenoviruses have the disadvantage of producing immunological and inflammatory reactions. These complications should be lessened with second-generation adenoviral vectors [3].

Some adenoviruses are human pathogens (wild-type Ad5, for example, causes the "common cold" syndrome), and most patients have already been exposed to them during their lifetime, resulting in the presence of various levels of circulating neutralizing anti-viral antibodies (NABs). This exposure may hinder the effectiveness of their systemic application. Use of viral vectors that are not human pathogens (e.g., various serotypes of adenoviruses, AAV, or nonhuman adenoviruses) can avoid this problem. Immune response evoked by the first application of the viral vector (human or nonhuman) may interfere with their repeated application (although the immune response may depend on the route of delivery and dose used). First- and second-generation adenoviral vectors have limited cloning size (<10 kb), which prevents the use of large therapeutic genes, multiple genes, or complex gene regulatory elements. The use of "gutless" adenovirus avoids this problem, allowing a cloning capacity of up to ~30 kb, but filling the vector with "useless" DNA is a challenge, and so is the proper manufacturing of these modified vectors, which require the presence of helper viruses.

*AAVs* have been used for effective gene transfer to muscle [4]. Numerous AAV serotypes have been identified (AAV1–9) recently, with a different tissue tropism, which gives us hope for using them in a tissue-specific manner [5].

In addition to pseudotyping (retroviruses) or deletion of nonessential viral genes (adenoviruses), several other strategies are pursued to improve viral vectors for human therapeutic use. This is a very active field, the overview of which is beyond the scope of this chapter. Some notable activities include changing viral coat proteins (fibers, pentons, etc.) to redirect the tropism of the virus and incorporation of specific gene regulatory systems (gene-switches, tissue-specific promoters, etc.) to allow proper timing, duration, extent, and localization of therapeutic gene expression.

*Nonviral Vectors.* The existing synthetic vectors (naked DNA, cationic liposomes, etc.) are far from being perfect delivery systems. Although they are less pathogenic and may have reduced toxicity compared with some existing viral vectors, depending on the dose injected, liposomes may aggregate in the blood and can cause severe toxic reactions [6].

Plasmid and liposome complexes are easy to produce and are safer, but they have low gene-transfer efficiency. However, novel lipid formulations and synthetic cationic polymer carriers have clearly improved the effectiveness of plasmid-mediated gene transfer [7–9].

In the future, the "perfect" gene delivery vector may be synthetic, incorporating many advantages of viruses (which over their evolution acquired the perfection of gene delivery to host organisms, such as dense DNA "packaging," cell recognition, cellular uptake, cytosolic trafficking, efficient nuclear uptake, and gene expression in the host cell nucleus, Figure 3), but avoiding the unwanted properties of viruses (e.g., off-target tropism, pathogenicity, cell toxicity, and immune and inflammatory reactions).

Dense DNA packaging (a prerequisite for efficient gene transfer) can be achieved by using, for example, protamine sulfate (a "trick" borrowed from sperm cells,

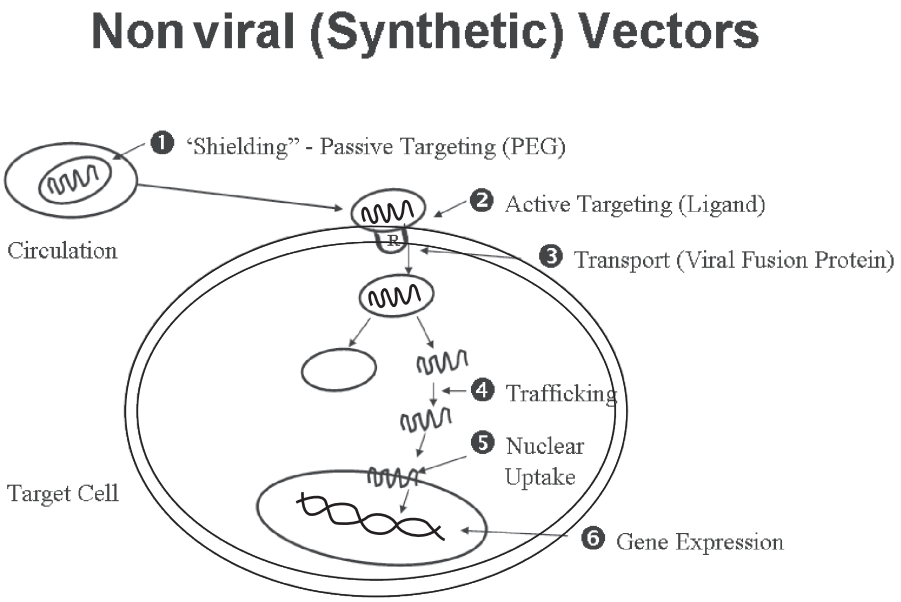


**Figure 3. The "ideal" synthetic (nonviral) gene delivery vector.** After dense DNA packaging is accomplished (e.g., by protamine sulfate), the surface of synthetic particles (which is usually positively charged) needs to be "shielded" (e.g., by poly (ethylene-glycol) [PEG]) so that they do not attach to blood elements or to each other and, therefore, have an extended circulating plasma half-life (1) (passive targeting to "leaky" vessels"). The surface of the particles will contain specific ligands for active targeting to selected cells/tissues (2). By engineering viral fusion proteins to the particle coat, cell entry is facilitated (3). Cellular trafficking will be enhanced and intracellular degradation of DNA prevented (4). Nuclear uptake will be facilitated by viral nuclear localization signal (NLS) peptides (5). Chromosomal localization will be augmented (e.g., by the *rep* gene, which allows targeting to chromosome 19) and gene expression regulated by specific transcriptional control elements (6). (Reproduced with permission from Pergamon Press.)

which similar to viruses package DNA efficiently). The particle should be "shielded" from binding to plasma proteins, blood cells, or to each other to allow longer circulating plasma half-life and more efficient uptake in target tissues (e.g., at the site of leaky vessels in tumors) ("passive" targeting). They will be engineered to contain cell recognition ligands (e.g., transferring for proper "active" targeting of cancer cells), cell membrane fusion proteins, and nuclear localization signals (all "borrowed" from viruses) for efficient cellular trafficking. Chromosomal localization and insertion for long-term expression of the therapeutic gene will be achieved by adding, for example, the *rep* gene, which allows targeting and insertion of the DNA to chromosome 19 [10]. These examples are just a few that are being tested today. However, even more sophisticated systems can be expected in the future, along with appropriate manufacturing and analytical processes, which will allow their introduction to human subjects.

*Cell-Based Delivery of Therapeutic Genes.* Although cell therapies have been used in medicine for several decades (e.g., blood transfusion), the use of cells

manipulated *ex vivo* with therapeutic genes and then reintroduced into patients offers a new strategy by which to deliver therapeutic genes.

Human hematopoietic stem cells, mesenchymal stem cells, neuronal stem cells, embryonic stem cells, and peripheral T cells are the focus of ongoing research efforts [11, 12]. The panel of stem cells available for gene therapy purposes will increase as isolation and culturing procedures improve and appropriate factors are identified that can be used to drive their differentiation along distinct cell lineage pathways. Issues currently being addressed include the development of vectors (e.g., lentiviruses) for efficient stem cell gene transduction, expression and regulation of therapeutic genes during lineage progression from stem cells to differentiated cells, control of stem cell growth, expansion *ex vivo*, and engraftment and differentiation *in vivo*. Genetically engineered stem cells or less pluripotent progenitor cells are currently being tested for therapeutic angiogenesis (endothelial cell progenitors), Parkinson's disease (neuronal stem cells), bone marrow transplantation (hematopoiectic stem cells), and AIDS (e.g., hematopoietic stem cells transfected with the *RevM10* gene) [11–13].

*"Customized" Gene Delivery Vectors.* It is important to emphasize that there will not be a "universal" vector that is optimally useful for all indications. On the contrary, each disease target will have a specific set of technical requirements, and the "perfect" vector for a specific disease should be optimized according to these specific criteria. For example, some diseases will require local delivery (e.g., ischemia, retinitis pigmentosa, and Parkinson's disease), whereas others necessitate systemic delivery (e.g., cancer). For certain diseases, the gene of a secreted protein (e.g., coagulation factors VIII and IX for hemophilia A and B, respectively; growth hormone; and erythropoietin) can be expressed in almost any tissue of the body. Sometimes only transient, short-lived gene expression will be needed (e.g., therapeutic angiogenesis and cancer), whereas in other cases long-term (sometime lifelong) gene expression duration will be necessary (e.g., most monogenic diseases, such as familial hypercholesterolemia, hemophilia, and SCID). For certain disease targets, most, if not all, target cells need to be transfected (cancer), whereas in other cases, this will not be necessary (e.g., with most secreted therapeutic proteins).

In certain diseases, tight control of gene expression will not be important (e.g., coagulation factors), whereas in others very tight regulation of the gene expression will be essential (e.g., insulin). Some diseases will require specific targeting of the vector for efficient and safe delivery after systemic application (e.g., cancer). Other disease targets will require tissue-or disease-specific promoter elements (e.g., arteriosclerosis and cancer). In some instances, conditionally inducible gene expression regulation (gene-switch) will allow precise dosing and timing of gene expression, which will also be an essential element for safety reasons (i.e., ability to "turn off" the gene if serious side effects occur).

Most vectors optimized for a certain disease target will consist of multiple "components," each fulfilling some necessary technical, biological, or therapeutic need. As the different elements will probably be perfected (and patented) by different companies, the introduction of the optimal multicomponent vector may be difficult because of intellectual property rights and commercial obstacles.

***Gene Delivery Targeting.*** The effectiveness of gene therapy is determined by a combination of the effects of gene delivery into the target tissue, the entry of the new genetic material into cells, and the expression of the transfected gene in the target tissue (Figure 1). When specific physical or biological targeting methods are available, they generally improve the expression of the transfected gene in the target organ and reduce gene expression in off-target tissues, which may lead to unwanted side effects.

*Physical Targeting.* A variety of physical gene delivery methods has been introduced to achieve better local tissue targeting of vectors. An example of the effective physical targeting is catheter-mediated gene transfer to various regions of the body (e.g., feed arteries of organs, such as leg muscles, heart, and liver, or retrograde injection via veins) [14–16]. Intramuscular injection of plasmid DNA or viral vectors encoding angiogenic growth factors has been used in ischaemic myocardium [17] and peripheral vascular disease [18, 19]. Another approach to local delivery to small arterioles and capillaries is injection of biodegradable microspheres coated with recombinant growth factors or plasmid DNA [20]. Ultrasonography [21], alone or in combination with microbubbles [22], can also potentially be used to improve the efficiency of gene transfer. For facilitation of intramuscular, intratumoral, or intradermal delivery of naked, plasmid, or cationic liposome-carried DNA, gene gun technology [23] and electroporation [24] can be used. Although these delivery methods offer certain advantages in specific disease targets, they are being progressively replaced by more specific biological targeting methods.

*Biological Targeting.* Biological vector targeting uses modification of viral coat proteins (for viral vectors) or surface properties of synthetic vectors (e.g., liposomes).

*Passive targeting* makes use of alteration of the pharmacokinetics of vectors by "shielding" them from binding to blood cells, plasma proteins, immunoglobulins, unwanted tissues, or to each other, allowing them to circulate for longer periods of time in the blood and accumulate in specific tissues with "leaky" blood vessels (such as tumors).

*Active (vector) targeting* aims at directing vector binding and uptake to specific cells in the body by selective "targeting" molecules attached to the surface of the vectors by a variety of technologies (e.g., chemical and genetic). Recent advances in the biological understanding of adenovirus structure and adenovirus receptor interactions have lead to the development of targeted adenovirus vectors. Receptor-mediated entry of adenovirus serotype 5 (Ad5) into cells has been found to depend on two of its coat proteins. The fiber knob protein mediates primary attachment to the cell via the coxsackie-adenovirus receptor (CAR) protein [25]. After cell attachment via fiber–CAR interaction, an RGD tripeptide motif in the penton base protein binds to integrins ($\alpha_v\beta_3$ and $\alpha_v\beta_5$) that mediate cellular internalization [26, 27] (Figure 2). Other adenovirus serotypes use different cell-surface receptors (e.g., the group B adenoviruses bind to CD46 [28]), most of which have not been identified yet.

Two basic requirements are necessary to create a (re)targeted Ad5 vector: Interaction of adenovirus with its native receptors (e.g., CAR) must be first removed,

and novel, tissue-specific ligands must be added to the virus [29–31]. Three general approaches have been used to achieve these basic requirements. In the "two-component" approach, a bispecific molecule is complexed with the adenovirus [29, 32]. The bispecific component simultaneously blocks native receptor (CAR) binding and redirects virus binding to a tissue-specific receptor (e.g., to integrins using the RGD motif). In the "one-step" approach, the adenovirus is genetically modified in the fiber protein domains to remove native receptor interactions and a novel ligand is genetically incorporated into one adenovirus coat protein [31–33]. The third approach is to use specific polymers to cover the entire virus and "shield" it from unwanted interactions. This method creates a truely "stealth" virus, preventing its interaction, not only with its primary (CAR, via fiber) and secondary cell-surface receptors (integrins, via penton RGD), but also with plasma proteins (including neutralizing antibodies), blood cells, and most tissue sites, providing an ideal situation for extended persistence in the circulation and *passive targeting* [34]. The polymer coat allows covalent attachment of specific "ligands" of choice to its surface, effectively retargeting the vector to a desired tissue (*active targeting*) [35].

*Transcriptional Targeting.* Receptor targeting technology can be combined with "transcriptional targeting" approaches (by the use of cell-, tissue- or disease-specific promoters).

A tissue- or disease-specific promoter will allow therapeutic gene expression only in cells that express transcription factor proteins binding to these specific promoter sites. Some tissue- or disease-specific promoters, identified and tested in animal models to date, include the promoter of the prostate-specific antigen (PSA) gene [36], osteocalcin gene [37], glucose "sensing" molecules for insulin gene delivery, and hypoxia response element (HRE) activated by hypoxia inducible factors (HIFs) in hypoxic/ischemic tissues (e.g., tumors) [38].

***Gene Expression Control Systems.*** Regulating therapeutic gene expression will be necessary for both the clinical efficacy and the safety of most HGT applications. Gene therapy offers the promise of replacing frequent injections of an expensive therapeutic protein or antibody with an infrequent or even one-time administration of a gene delivery vector, which would then provide continuous therapeutic protein production at the desired site. An appropriate system of gene expression will control and allow titration of protein levels, dosing to be adjusted as the disease evolves, and therapy to be initiated repeatedly or terminated at will. Regulation can be achieved by a physiological/pathological signal (e.g., glucose and hypoxia) [39, 40] or via dose-dependent ligand binding and activation of chimeric transcription factor proteins, which then interact with DNA elements incorporated into the vector and regulate therapeutic gene expression [41].

Much progress has been made on control systems, where gene expression is regulated pharmacologically by a small-molecule drug. Four major systems are known to date that have already been tested in animals: those regulated by the antiobiotic tetracycline (Tet) [42], the insect steroid ecdysone or its analogs [43], the anti-progestin mifepristone (RU486) [44], and chemical dimerizers such as the immunosuppressant rapamycin and its analogs [45–47]. They all involve small-molecule-dependent recruitment of a transcriptional activation domain to a basal

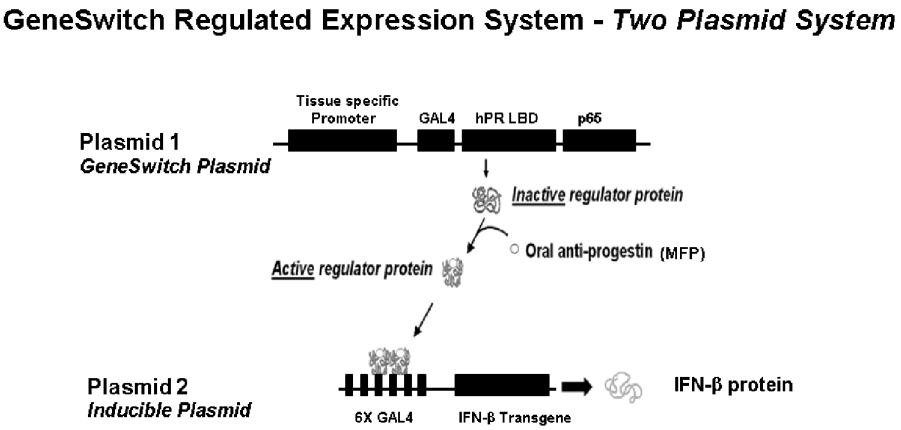


**Figure 4. The two-plasmid GeneSwitch gene expression control system.** Plasmid 1 contains the DNA of the mutated ligand binding domain (LBD) of the human progesterone receptor (hPR) driven by a tissue specific promoter (e.g., actin promoter for muscle applications). It also contains the GAL4 DNA-binding domain and the p65 activator domain (all combined called the "regulator protein"). After expression, the regulator protein remains inactive until an anti-progestin small molecule compound (e.g., mifepriston, MFP) binds to it, and initiates conformation change and dimerization. The active regulator protein binds to the promoter region of Plasmid 2, which induces the expression of the therapeutic gene of interest (in this example interferon β, IFNβ).

promoter driving the gene of interest, but they differ in the mechanism of recruitment. Such a pharmacological gene expression regulation system should meet the following criteria. Basal expression should be very low and inducible to high levels over a wide dose range. Induction should be a positive effect (adding rather than removing a drug) and use an orally active small molecule that has no untoward pharmacological effects in humans. The regulatory protein(s) should have no effects on endogenous gene expression and should be of human origin to minimize immunogenicity. An example of one of the most advanced systems (GeneSwitch), developed by Valentis, Inc. (Burlingame, CA) is shown in Figure 4. It has been successfully used for the timing (Figure 5) and dosing (Figure 6) of circulating plasma levels of therapeutic proteins (such as IFNβ) expressed in the hind limb muscle of mice after i.m. delivery of their DNA in the form of AAV-1 (Figure 5) or naked plasmid combined with electroporation (Figure 6).

## 2.3 Disease Targets for Gene Therapy

Ever since the first clinical attempt of gene transfer more than 17 years ago, numerous monogenic (Table 2) and complex (multigenic) diseases (Table 3) were targeted with a large variety of gene therapy strategies.

In 1989, Rosenberg et al. performed the first human gene therapy trial [49] when they used a retrovirus to introduce the gene coding for resistance to neomycin into human tumor-infiltrating lymphocytes before infusing them into five patients with advanced melanoma.

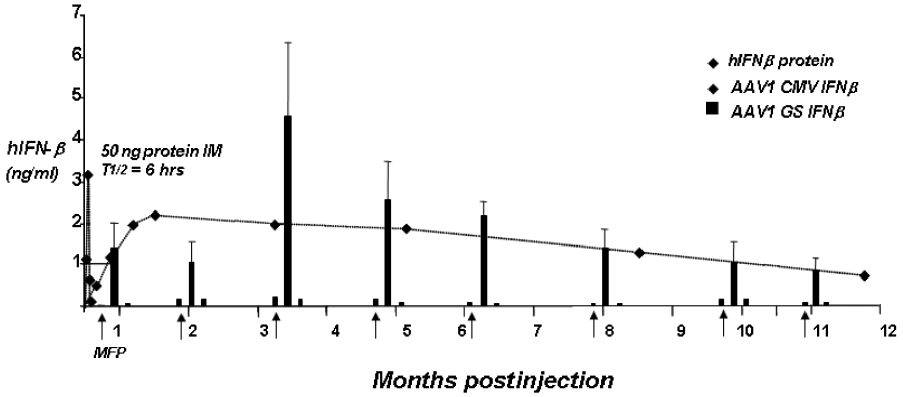


**Figure 5. Long-term regulated expression of human Interferon β (hIFNβ) in mice.** Three groups of male C57B16/J mice (n = 10 each) were injected with hIFNβ protein (50 ng, intramuscular, i.m.; diamonds), adeno-associated virus serotype 1 (AAV1) carrying the hIFNβ gene driven by the constitutive CMV promoter ($10^{10}$ vp/animal; i.m.; dotted line), or AAV1 carrying the hIFNβ gene regulated by the GeneSwitch (GS) inducible promoter (columns). Circulating plasma hIFNβ level (ng/mL) was measured by ELISA. I.m. injection of IFNβ protein resulted in transient increase in plasma IFNβ levels with very rapid clearance (T1/2: 6 hours). In contrast, single i.m. injection of AAV1 CMV IFNβ resulted in long-term (at least up to 12 months) elevation of plasma hIFNβ protein levels. IFNβ protein expression could be repeatedly induced by intraperitoneal (i.p.) administration of MFP (once every 6 weeks up to 12 months; arrows) after single i.m. administration of AAV1GS IFNβ. The induced plasma hIFNβ levels from the GS construct were similar to the sustained plasma levels produced by the constitutive (CMV) promoter.

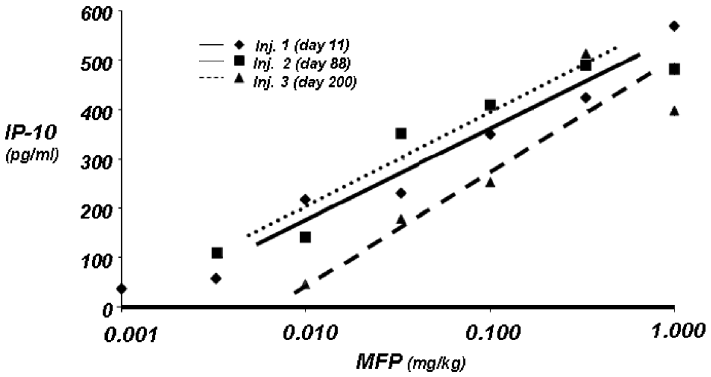


**Figure 6. Reproducible dose-dependent induction of IFNβ by Mifepristone (MFP).** C57B16/J mice (n = 35) were injected intramuscularly (followed by electroporation) with a plasmid (250 μg) containing the murine (m) IFNβ gene regulated by GeneSwitch upon induction by MFP. Groups of mice (n = 5 each) were given different doses (0 to 1 mg/kg) of MFP intraperitonally (i.p.). Plasmid injection and MFP dose-response were repeated three times, every other month, for 6 months. The level of mIFNβ protein in plasma was estimated by the IFNβ-induced marker protein, IP-10. Note the linear dose-response to MFP over a two log dose-range (0.01–1.0 mg/kg), which was reproducible upon readministration of the plasmid.

**TABLE 2. Disease Targets for Gene Therapy: Monogenic Diseases**

| Disease | Gene(s) | Number of Clinical Protocols/Trials (1990–1999)[1] |
|---|---|---|
| Cystic fibrosis | CFTR, α-1-anti-trypsin | 24 |
| Severe combined immunodeficiency (SCID) | ADA | 3 |
| Gaucher disease | Glucocerebrosidase | 3 |
| Canavan disease | Aspartoacylase | 2 |
| Hemophilia A | Factor VIII | 2 |
| Hemophilia B | Factor IX | 2 |
| Chronic granulomatous disease | Gp91 phox | 2 |
| ALS | CNTF | 1 |
| Familial hypercholesterolemia | LDL-R | 1 |
| Hunter disease | ldurinate-2-sulfatase | 1 |
| Leukocyte adherence deficiency | CD 18 | 1 |
| Muscular dystrophy | Sarcoglycan, dystrophin, utrophin | 1 |
| Fanconi anemia | Group A gene | 1 |
| Purine nucleoside phosphorylase deficiency | PNP | 1 |
| Ornithin transcarbamylase deficiency | OTC | 1 |

[1] From Ref. 48.

**TABLE 3. Disease Targets for Gene Therapy: Multigenic Diseases**[1]

**Cancer**
Gynecological: breast, ovary, cervix
Nervous system: glioblastoma, leptomeningeal carcinomatosis, glioma, astrocytoma, neuroblastoma
Gastrointestinal: colon, colorectal, liver mestastases, post-hepatitis liver cancer
Genito urinary: prostate, renal
Skin: melanoma
Head and neck
Lung: adenocarcinoma, small cell, non-small cell
Mesothelioma
Hematological: leukemia, lymphoma, multiple
Myeloma
Sarcoma
Germ cell tumors

**Infectious**
HIV/AIDS
Tetanus
CMV infection
Adenovirus infection

**Vascular disease**
Peripheral arterial disease
Coronary heart disease
Venous ulcers
Vascular complications of diabetes

**Other diseases**
Inflammatory bowel disease
Rheumatoid arthritis
Chronic renal disease
Carpal tunnel syndrome
Alzheimer's disease
Fractures
Diabetic neuropathy
Parkinson's disease
Erectile dysfunction
Superficial corneal opacity
Retinitis pigmentosa
Glaucoma

[1] From Ref. 1.

The first trials have been designed to establish feasibility and safety, to demonstrate the expression of therapeutic protein(s) *in vivo* by the genes transferred, and in some cases, to show therapeutic benefit.

As of January 31, 2004, there were 918 trials in 24 countries on record [1]. The U.S. accounts for two thirds of these trials. Cancer is by far the most common disease indication, followed by inherited monogenic diseases and cardiovascular diseases. Viral vectors have been the most frequently used vehicles for transferring genes into human cells, with retroviruses and adenoviruses representing the vast majority. Over 100 distinct genes have been transferred.

## 3 FROM BENCH TO BEDSIDE: THE DEVELOPMENT OF AD5FGF-4 TO TREAT PATIENTS WITH RECURRENT ANGINA

A brief summary of the preclinical and early clinical development and manufacturing of the Ad5FGF-4 gene therapy product for the facilitation of angiogenesis in patients with chronic myocardial ischemia (recurrent angina) is provided below for the purpose of illustrating the complex processes involved in the many stages of preclinical and clinical development of an HGT product.

### 3.1 Preclinical Efficacy

Specific criteria need to be fulfilled for angiogenic gene therapy to be successful. The gene selected should code for a protein with proven angiogenic activity, the vector should provide high gene-transfer efficiency, the delivery technique should target the ischemic tissue, and the procedure should be safe, in both the short and the long term [50]. In addition, appropriate current Good Manufacturing Practices (cGMP) need to be established (for details, see Section 3.3). Preclinical and initial clinical studies indicate that angiogenic gene transfer with Ad5FGF-4, delivered by a single intracoronary injection, may meet these criteria [16, 50–55].

***Proof-of-Concept Study.*** Preclinical investigation of Ad5FGF-4 involved demonstration of myocardial angiogenesis in a porcine model of stress-induced myocardial ischemia [51]. In this experimental model, an ameroid constrictor is placed around the proximal left circumflex coronary artery, leading to gradual closure of the artery over the following weeks. The subsequent development of collateral vessels allows normal myocardial function and blood perfusion at rest, but blood flow is insufficient to prevent ischemia during periods of increased oxygen demand. In the proof-of-concept studies, stress-induced (atrial pacing) left ventricular function and blood flow changes were evaluated by two-dimensional echocardiography approximately 35 days after ameroid placement. Ad5FGF-4 ($10^{11}$ viral particle [v.p.] per animal) (n = 6) was then administered by intracoronary injection. Ventricular function and blood flow were reassessed 14 days later, and the animals were killed to quantify the presence of adenoviral DNA and FGF-4 mRNA and protein expression in the myocardium.

Polymerase chain reaction (PCR) and reverse-transcription–PCR (RT–PCR) analysis demonstrated the presence of Ad5 DNA and FGF-4 mRNA, respectively, in the injected pig hearts [51]. Immunoblotting showed that pigs treated with

Ad5FGF-4 expressed FGF-4 protein in heart tissue but not in other organs, such as the liver and eye. Contractile function, assessed as the degree of wall thickening of the ischemic region during atrial pacing, had improved significantly 2 weeks after FGF-4 gene transfer compared with preinjection (Figure 7A). This functional

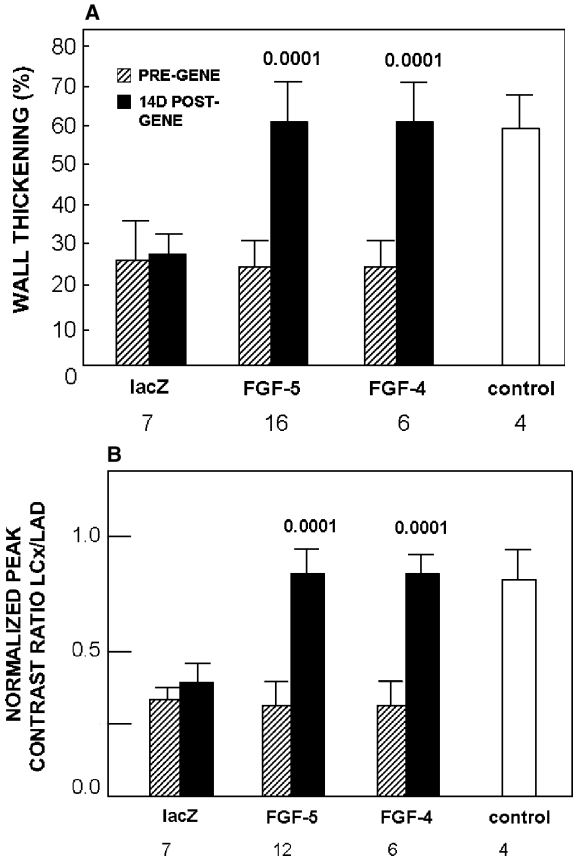


**Figure 7. Proof-of-principle pig studies for angiogenic gene therapy after intracoronary delivery of Ad5FGF-4.** (A) Regional contractile function. Basal wall thickening in the ischemic region of the ameroid-equipped pig heart was normal, but atrial pacing (200 bpm) was associated with reduced wall thickening. At 2 weeks after intracoronary Ad5FGF-4 gene transfer ($10^{12}$ v.p.), there was a 2.3-fold increase in wall thickening in the ischemic region during pacing ($p < 0.0001$); similar to results observed after Ad5FGF-5 gene transfer. Pigs receiving Ad5LacZ showed a similar degree of pacing-induced deficit in the ischemic region before and 2 weeks after gene transfer. (B) Regional myocardial blood flow. Before gene transfer, animals showed a deficit in blood flow in the ischemic region during pacing. At 2 weeks after Ad5FGF-4 gene transfer, the animals showed improved flow in the ischemic region ($p < 0.0001$); this was similar to results observed after Ad5FGF-5 gene transfer. Pigs receiving Ad5LacZ showed a similar degree of pacing-induced deficit in the ischemic region before and 2 weeks after gene transfer. Columns represent mean values for number of experiments shown below the columns; error bars denote 1 SD. (Reproduced with permission from Mary Ann Liebert, Inc.)

improvement was associated with a normalization of regional blood flow during atrial pacing (Figure 7B) [51]. Thus, this study provided experimental evidence that a single intracoronary injection of Ad5FGF-4 ameliorates deficits in myocardial blood flow and function in the setting of chronic myocardial ischemia in pigs.

Follow-up studies in the same pig model showed that the effect of single Ad5FGF-4 injection lasted at least for 3 months, and that the effect was dose dependent [51].

### 3.2 Toxicology and Biodistribution

A Good Laboratory Practices (GLP) toxicology and biodistribution study was performed to determine the potential adverse effects of Ad5FGF-4 in healthy pigs after either intracoronary or left ventricular (systemic) administration ($10^{12}$ v.p. left ventricular, $10^{10}$–$10^{12}$ v.p. intracoronary). Systemic biodistribution of the product was also assessed.

This study found no significant test article-related toxicologic effects [51]. PCR analysis for adenoviral DNA and RT–PCR analysis for the transgene (FGF-4) mRNA were performed to assess vector biodistribution. Adenoviral DNA was detected in 27 of 110 organs examined from animals injected intracoronary with $10^{12}$ v.p. Ad5FGF-4. Adenoviral DNA was detected in the lung, liver, spleen, and testis. Detectable adenoviral DNA typically decreased over time in most organs. The presence of viral DNA in the testis was observed at 5 days but not at 28 or 84 days. In no case in which adenoviral DNA was present in extracardiac sites was transgene expression detectable at the mRNA level after RT–PCR analysis [51]. These results (along with cGMP; see Section 3.3) supported the submission of an investigational new drug (IND) application in 1998.

### 3.3 cGMP Manufacturing and Characterization of the Gene Therapy Vector

This section summarizes the various activities required in the course of cGMP manufacturing and testing of the Ad5FGF-4 product for the initiation and advancement of its clinical development and reaching a "well characterized product" status with the Regulatory Authorities.

***Vector Construction.*** Construction of the recombinant replication incompetent (E1/2 deleted) adenovirus serotype 5 (Ad5) expressing the human FGF-4 cDNA required three components: the FGF-4 transgene, a plasmid shuttle vector to carry the transgene as well as 5′ Ad5 sequences, and a second plasmid carrying the bulk of the Ad5 genome.

The full-length cDNA for human FGF-4 (the "transgene") was isolated from a cDNA library, which was constructed from mRNA of Kaposi's sarcoma DNA transformed NIH3T3 cells [56]. The cDNA that encodes the FGF-4 peptide is approximately 1.2 kB in size, and the FGF-4 protein has 206 amino acids, including a 33-amino-acid signal peptide at the N-terminus.

Plasmid pACCMVplpASR(-) (the "shuttle vector") contains the heterologous CMV promoter, a polylinker, and SV40 polyadenylation sequences flanked by

partial human Ad5 sequences [57]. The FGF-4 cDNA was subcloned, as an Eco R1 fragment, into this adenovirus shuttle vector at its single Eco R1 site.

Plasmid pJM17 (the "adenovirus plasmid") contains the required Ad5 sequences except that the E1 region is disrupted by the insertion of pBR322 sequences [58]. A unique feature of this plasmid is the presence of some nonviral sequences in the E3 region [59] that become relevant for the viral vector identity testing.

To generate the Ad5FGF-4 viral vector, plasmids pACSR/FGF-4 and pJM17 were co-transfected into HEK 293 cells using a calcium phosphate method. The cells were overlaid with nutrient agarose. Homologous recombination between the vectors created E1-deleted, FGF-4 gene-containing vector genomes capable of replication in HEK 293 cells.

***Master Virus Bank.*** Six clones were isolated from one round of virus plaque purification and were screened for protein expression. One clone was selected, amplified on HEK 293 cells, purified by cesium chloride (CsCl) ultracentrifugation, and used for preparation of a Master Virus Bank (MVB). A purified Ad5FGF-4 virus seed was propagated through three consecutive rounds of plaque purification. A final plaque was then expanded and purified by anion exchange chromatography, sterilized by filtration, and stored at –70°C. This virus stock was checked for sterility and absence of measurable replication competent adenovirus (RCA). An MVB was then created by one additional cycle of propagation in serum-free suspension culture of HEK 293 cells, and this virus was purified, aliquoted, and frozen. The MVB was tested and confirmed to be free of RCA as well as adventitious agents. The MVB was stored at –70°C.

***Master Cell Bank and Manufacturer's Working Cell Bank.*** A scalable manufacturing process to meet clinical, as well as commercial, needs for recombinant adenovirus vector would only be attractive if it could be performed without using bovine serum in the culture medium. For these reasons, attachment and serum-dependent HEK 293 cells were first adapted to suspension culture by direct transfer into shake flasks in a modified William's essential medium containing 2% fetal bovine serum (Hyclone, Gamma irradiated). The cultures were passaged continually for over 6 weeks in the same environment until the cells exhibited a consistent growth pattern of less than 48 hours of doubling time and greater than 90% viability. These suspension-adapted cells were subsequently adapted to serum-free medium (modified IS293, Irvine Scientific) by gradual weaning. Cells from this bank were thawed and expanded and then used to generate a Master Cell Bank (MCB). Culture identification as human cells, growth on soft agarose to assess potential tumorigenicity, and screening for adventitious agents were also performed (according to current cell-line testing guidelines). Starting from the MCB, a Manufacturer's Working Cell Bank (MWCB) was prepared and then similarly tested before routine use.

***Manufacture of the Bulk Drug Substance.*** The Ad5FGF-4 vector was manufactured under cGMP in a state-of-the-art, validated facility, following Biosafety Level 2 (BL2) practices. For the preparation of the Ad5FGF-4 drug substance, cells from the MWCB and virus from the MVB were employed. The virus was propagated in suspension cultures of HEK 293 cells using serum-free medium and the virus

progeny was purified using a combination of anion exchange chromatography (AIEX) and ultra-filtration (UF). The purified bulk drug substance was stored at –70°C.

***Drug Substance (Virus) Purification.*** Medium components and extracellular (nonviral) contaminants were largely removed from adenoviral infected HEK 293 cells using three successive physiological buffer solution (PBS) washes and centrifugation of the cells at the time of harvest. The harvested cells were then frozen in a solution of PBS with 2% sucrose and stored at –70°C, until purification. The virus purification process was initiated with two additional cycles of freeze and thaw steps. After the final thaw, the ruptured cell suspension was centrifuged to remove cell debris. Virus purification was established with a protocol based on traditional protein purification techniques: column chromatography and ultra-filtration. Separation from nucleic acid contaminants was achieved by combining strategic peak collection and utilization of a tangential flow ultra-filtration step with a 100-kDa molecular-weight cutoff pore size membrane. Both of these purification procedures are readily and linearily scaled up and thus can support the large-scale production of Ad5FGF-4 for both clinical testings and commercial use.

***Manufacture Scale-Up.*** The adenovirus harvest process consisted of three general steps: concentration, lysis, and clarification. In the small-scale process, the intact adenovirus-infected cells in culture medium were aliquoted and batch centrifuged to achieve a 30-fold concentration, and the supernatant was manually removed and replaced with a smaller volume of the freeze buffer. Although sufficing for harvest volumes of no more than 10 L, such a manual process would prove to be inadequate on a larger manufacturing scale.

A successful harvest method was identified using continuous centrifugation in a Westfalia CSC-6 disk-stack device with a hydroheratic feed system (HHFS). Scale-up of the adenovirus harvest process using a continuous centrifuge with HHFS has been shown to efficiently separate and concentrate an infected cell pellet while allowing the relatively easy exchange of growth medium for freezing buffer without open, manual manipulation. This patented process step combined separation of the cell pellet from culture medium, medium/buffer exchange, cell concentration, and then cell lysis in a single-unit operation step. The resulting lysate could be clarified and the virus purified.

***Analytical Techniques and Release Specifications of Bulk Virus Lots.*** The crude viral harvest was tested for adventitious agents as needed to comply with current regulatory guidelines. Purified virus was tested to confirm the absence of RCAs, residual host cell proteins and DNA, and endotoxin, and to quantify total and infectious vector particles and, thus, the infectivity ratio. Bulk virus lots were released after meeting specifications, which included <1 endotoxin unit/mL and infectivity ratios of >2% (infectious titer/total viral particles).

*Chromatographic Assays for the Quantitation of Intact Viral Particles.* One main challenge in the development of the adenovirus purification process is the quantitation of intact viral particles in crude samples. Nonaggregated, purified viral samples were easily quantified by the reverse phase (RP)–high-performance liquid chro-

matography (HPLC) method [60]. The RP–HPLC assay quantitates the individual structural proteins with UV absorbance at 214 nm as the basis for the measurement of the adenovirus concentration. Unfortunately this assay could not be used to calculate the yields of the purification steps because it could not reliably quantitate crude samples. To solve this problem, an anion exchange (AIEX)–HPLC procedure was developed using a trimethyl anion exchange (TMAE) resin. The TMAE–HPLC was suitable for the analysis of all in-process samples ranging from clarified cell harvest to the final formulated, purified recombinant adenovirus product.

*Endpoint Dilution (EPD) Assay for Infectivity.* To determine infectivity, an endpoint dilution assay was developed [61]. The precision of this assay has a standard deviation of 0.2 when viral titers are expressed in $\log_{10}$.

*Replication Competent Adenoviruses (RCA).* Recombination between the viral E1 gene carried by HEK 293 cells and the E1 deleted recombinant adenovirus can and does occur at low frequency to generate RCAs, essentially wild-type Ad5 lacking the FGF-4 transgene. A highly sensitive assay involving viral amplification and cytopathic effect (CPE) readout has been described [62]. The limit of detection of this assay was one RCA in $3.2 \times 10^{12}$ virus particles. Virus banks and bulk Ad5FGF-4 virus batches testing positive for RCA were not used in clinical studies.

*Host Cell Protein Detection by ELISA.* Typically biotechnology products have been produced in cell lines not of human origin. In such cases, residual host cell proteins (HCPs) are perceived as a safety concern to the patient. In the case of Ad5FGF-4 produced in HEK 293 cells, it is expected that the patient may respond to viral proteins with an immune response, but it is less clear that human HCPs will be immunogenic or pose a serious safety concern. One approach to detection of such HCP is by enzyme-linked immunosorbent assay (ELISA) using a polyclonal antibody reagent such as the HEK 293 HCP detection kit offered by Cygnus Technologies (Renchesler, MD). This ELISA was used to show a consistent, high HCP titer in crude cell lysates and to show that the diethyl-aminoethyl (DEAE) chromatography step is sufficient and reproducible to reduce HCPs to below the level of quantitation (20 ng/mL).

*Residual DNA.* Two methods were used to quantitate residual host cell (HEK 293) DNA in Ad5FGF-4 preparations. The first, using membrane hybridization of an oligonucleotide probe followed by chemiluminescent detection, was based on a procedure described by Walsh et al. [63]. Extracted samples were blotted onto a positively charged nylon membrane using a slot blot apparatus, the blot was then hybridized with a biotinylated oligonucleotide probe specific to the primate alpha-satellite sequence D17Z1, and the bound probe was detected with a streptavidin–alkaline phosphatase conjugate, followed by a chemiluminescent alkaline phosphatase substrate. The second method used TaqMan PCR quantitation of *Alu* repeat sequences in virus samples extracted using the Qiagen QIAAmp Viral RNA Mini Kit. In-process sample testing showed that the two methods gave similar results, both demonstrating a greater than 3 log reduction in the level of host cell DNA by the purification process, with final purified material containing less than 0.5-ng/mL DNA.

***Product Identity Tests by PCR.*** The adenoviral backbone of the Ad5FGF-4 genome is derived from *dl*309, an Ad5 mutant, commonly used for the construction of E1-deleted adenoviral vectors. This mutant has a short stretch of foreign DNA inserted in place of a portion of the E3 region [59]. PCR confirmation of the type of adenovirus backbone (*dl*309 or wild-type Ad5 E3 region) served as a test for product identity and purity. It was possible to confirm transgene orientation as well as differentiate the product digest pattern from those obtained from either wt Ad5, RCA, or vectors containing different transgenes.

***Product Potency Tests.*** Two assays were developed that measure the potency of the FGF-4 protein. In the first case, a "one-step" growth promotion assay is conducted on normal, human retinal pigment epithelial cells (ARPE-19). The assay measures metabolic activity (Alamar blue dye metabolism) after infection of ARPE-19 cells with a serial dilution of the drug substance (Ad5FGF-4). The increase in metabolic activity correlated with FGF-4 protein production (determined by an FGF-4 ELISA), increased *de novo* DNA synthesis (measured by BrdU incorporation), and an increase in cell number.

The second assay measured the production of FGF-4 protein produced in and secreted by A549 cells after infection with different doses of Ad5FGF-4. An FGF-4 specific ELISA (R&D Systems) allowed quantitation of FGF-4 protein present in the cell culture medium. The FGF-4 produced was confirmed as biologically active by stimulating the growth of ARPE-19 cells (see above).

***DNA Sequencing of the Entire Viral Vector.*** The complete DNA sequence of the Ad5FGF-4 genome was determined using double-stranded DNA sequencing of Ad5FGF-4 DNA extracted from an MVB expansion. A total of 166 oligonucleotide primers were used to obtain overlapping readings from both strands of DNA. The observed sequence was compared with that predicted for Ad5FGF-4 by insertion of the FGF-4 transgene sequence into the wild-type Ad5 sequence [64], (Genbank Accession Number M73260, originally deposited as NC001406), into which the *dl*309 specific mutations had been incorporated [59]. The observed sequence matched the predicted 5′ 4.9 kB of vector sequence including the FGF-4 transgene. The sequence data distal to the transgene (3' end) confirmed that the Ad5FGF-4 backbone was derived from the Ad5 mutant *d*/309.

***Genome Analysis.*** As a further safety precaution, and as a specific regulatory request, the DNA sequence of the FGF-4 transgene was analyzed for other possible open reading frames (ORFs) that might encode unexpected foreign proteins upon infection of target cells with the recombinant adenovirus. This analysis also included any reading frames that crossed the transgene/virus junctions and might therefore involve viral sequences. The genome analysis confirmed that only FGF-4 protein is made from the transgene carried in Ad5FGF-4.

The cGMP manufacturing and analytical process (outlined above) and the facility where it took place (Berlex Biosciences, Richmond, CA) have been approved by both the U.S. Food and Drug Administration (FDA) and the European Regulatory Agencies. Based on the test data, the FDA qualified Ad5FGF-4 as a "well characterized product."

### 3.4 Clinical Studies

There are several goals of Ad5FGF-4 angiogenic gene therapy for patients with chronic myocardial ischemia. It should promote new collateral vessel formation in the heart. In turn, this should increase perfusion of ischemic regions, leading to improved myocardial oxygen delivery and left ventricular function. From the patient's perspective, the ultimate goal of treatment should be to reduce or ameliorate symptoms of angina, increase exercise capacity, improve quality of life, and decrease the long-term risk of acute coronary events (such as acute myocardial infarction).

The clinical development program was initiated in 1998 to determine whether the improvement in cardiac perfusion and function (without any product-related adverse effects) detected in pigs translates into a clinical therapeutic benefit in patients with chronic stable angina. Four clinical studies [Angiogenic Gene Therapy (AGENT) phase 1/2 trial, AGENT 2 phase 2 trial, and AGENT 3 and 4 phase 2B/3 trials] have been initiated, and some (AGENT and AGENT-2) have been completed to date.

***The Phase 1/2 AGENT Trial.*** The Angiogenic Gene Therapy Trial (AGENT) was the first ever randomized, double-blind, placebo-controlled (12 sites) U.S. clinical trial of angiogenic gene therapy for myocardial ischemia. The objectives of the AGENT trial were to evaluate the safety and anti-ischemic effects of five ascending doses of Ad5FGF-4 ($3.2 \times 10^8$ to $3.2 \times 10^{10}$ v.p.), randomizing in a ratio of 1:2 (placebo:active) in patients with chronic stable exertional angina, who had recurring angina despite optimal medical treatment and were not in need of immediate surgical revascularization.

A total of 79 patients with chronic stable angina were enrolled. Patients could exercise for ≥3 minutes in an exercise treadmill test (ETT) using the modified Balke protocol. The adenovirus vector was infused over 90 seconds through subselective catheters into all major patent coronary arteries and grafts, 40% into the right coronary distribution and 60% into the left coronary distribution. Repeat treadmill exercise tests were performed at 4 and 12 weeks posttreatment.

The increase from baseline in treadmill exercise duration at weeks 4 and 12 was greater among patients receiving $10^{10}$ v.p. Ad5FGF-4 than among those receiving placebo (the difference at week 4 was statistically significant) (Figure 8) [52]. Analysis of the subgroup of patients with a baseline ETT time of <10 minutes showed the gene therapy produced a statistically significant increase in exercise capacity compared with placebo at 12 weeks (27% vs. 12%; $p = 0.01$).

***The Phase 2 AGENT-2 Trial.*** AGENT 2 was designed to assess whether Ad5FGF-4 improved myocardial perfusion in patients with stable angina. It was also designed to further evaluate safety. Based on the results of the AGENT trial, a dose of $10^{10}$ v.p. was selected. The primary endpoint was the change in stress-related (adenosine-induced) reversible perfusion defect size (RPDS) as assessed by single-photon emission computed tomography (SPECT), 8 weeks after treatment [52].

A total of 52 patients underwent double-blind randomization (35 to Ad5FGF-4, 17 to placebo). Total perfusion defect size (PDS) at baseline was 32% and RPDS was 20%.

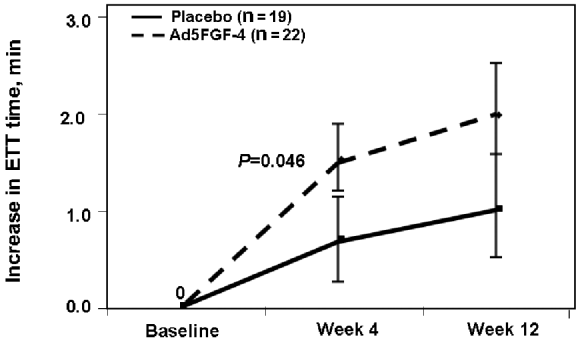


**Figure 8. AGENT Trial ETT Results.** Mean (± SEM) change in total exercise treadmill test (ETT) time for patients at 4 and 12 weeks after intracoronary infusion of $10^{10}$ v.p. of Ad5FGF4 (n = 22) or placebo (n = 19).

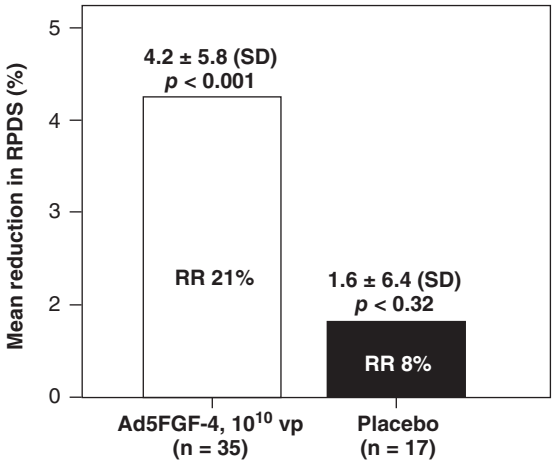


**Figure 9. AGENT-2 Trial Perfusion Results.** Mean change in reversible perfusion defect size (RPDS) compared with baseline for all active (n = 35) and placebo (n = 17) patients at 4 and 8 weeks after intracoronary infusion of Ad5FGF-4 ($10^{10}$ v.p.) or placebo. RR = relative reduction in RPDS compared with baseline. (Reproduced with permission from Elsevier.)

The mean reduction in RPDS from baseline at 8 weeks posttreatment in the Ad5FGF-4 group was 4.2 ± 5.6% ($p < 0.001$), a 21% decrease from baseline, versus a reduction in the placebo group of only 1.6 ± 5.4% ($p = 0.32$), corresponding to a decrease of 8% from baseline (Figure 9). Similar results were observed for the change in total PDS from baseline at 8 weeks posttreatment: Ad5FGF-4 decreased total PDS by a mean of 4.6 ± 5.6% ($p < 0.001$) compared with a reduction of 2.4 ± 6.5% with placebo ($p$ = NS).

More of the patients who received the active product than those who received placebo reported complete resolution of angina (30% vs. 13%) and no nitroglycerin use (43% vs. 17%) at 8 weeks. In addition, the incidence of worsening/unstable angina and revascularation by coronary artery bypass grafting or angioplasty at 12

months was considerably lower in the Ad5FGF-4 group (6% and 6%, respectively) compared with those in the placebo group (24% and 16%, respectively).

***Safety in AGENT and AGENT 2 Trials.*** In the AGENT and AGENT-2 trials combined, a total of 131 patients (95 on Ad5FGF-4 and 36 on placebo) were treated and followed for 12 months. Overall, Ad5FGF-4 was well tolerated. There was no rise in cardiac enzymes, electrocardiographic change, or clinical evidence of myocarditis associated with the treatment [52, 53].

Vector distribution after intracoronary administration was examined in the AGENT trial [52]. Adenovirus could be detected by endpoint dilution infectivity assay in the pulmonary artery during intracoronary infusion and in peripheral venous blood 1 hour later. The frequency of positive samples increased with increasing doses of Ad5FGF-4 (Figure 10). Virus was not detected in the urine collected over 6 hours after treatment. Neutralizing antibody titer to Ad5 increased in most patients, but no FGF-4 protein was detected in the circulation at any time. Semen samples (n = 8) were tested by PCR (8 weeks posttreatment) and found to be negative for Ad5FGF-4 DNA. Ad5FGF-4-related adverse effects included dose-related transient fever (n = 8[8%]) and transient increase in liver enzymes (n = 3 [3%]).

Overall, the safety profile was reassuring and consistent with safety data in other cardiovascular adenoviral gene therapy trials [65, 66]. There was no evidence of myocarditis, retinal neovascularization, or angioma formation.

The positive trend toward efficacy and the safety observed in these two trials were in good agreement with the preclinical safety and efficacy data obtained in pigs, and they provided the basis to progress the clinical development to larger, multinational phase 2b/3 trials (AGENT-3 and AGENT-4), the results of which will be reported in the future.

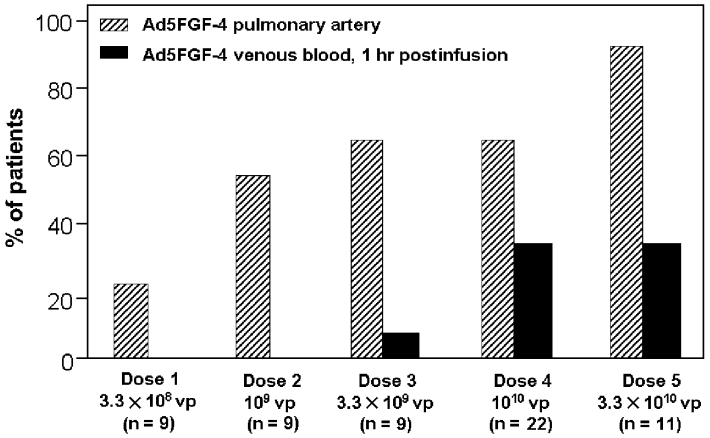


**Figure 10. AGENT Trial Virus Distribution.** Detection of adenovirus in the pulmonary artery during intrcoronary infusion and in venous blood at 1 hour after infusion in patients treated with increasing doses of Ad5FGF-4 in the AGENT trial (n = 60). (Reproduced with permission from Lippincott, Williams & Wilkins.)

### 3.5 Lessons Learned from Past Failures

The remarkable clinical safety observed in most cardiovascular gene therapy trials to date [52, 53, 65, 66] is of no coincidence. Lessons learned from past failures at the manufacturing, preclinical efficacy and safety, and clinical development levels were all taken into account in the development of Ad5FGF-4.

Failures were due primarily to two factors: safety issues and lack of clinical efficacy.

The tragic death of Jesse Gelsinger at the University of Pennsylvania treated with a very high dose (>10 e13 vp) of a first-generation adenoviral vector raised the issue of dose and innate immune response. We learned that side effects of Ad5 are dose dependent, which follow a so-called "elbow" shape (i.e., no effect up to a certain dose and severe effects at the next dose). Therefore we have tested the dose-response relationship of efficacy, biodistribution, side effects, and toxic effects very thoroughly in our preclinical models, and we chose the doses of10e9 and 10e10 vp (more than three orders of magnitude lower than the dose used in the "Gelsinger" case).

Another gene therapy clinical trial that led to serious side effects was the SCID trial in France, where several of the children treated developed leukemia. These later were attributed to retrovirus integration at a site in the genome, which led to significant clonal expansion of the treated lymphocytes. Therefore, the choice of vector could be of great importance. In our product candidate, we intentionally avoided the use of integrating vectors (e.g., retrovirus, lentivirus, and AAV). As relatively short-term and transient gene expression is sufficient for therapeutic angiogenesis, we did not need to use these integrating vectors.

Both the "Gelsinger" case and the "SCID Trial" intensified regulatory scrutiny, which served the field very well. Small (mostly physician sponsored) trials using poorly characterized material are not allowed anymore. Both the FDA and the European Regulatory Agencies require well-controlled cGMP Manufacturing to be in place before clinical testing can start.

The recent failure of Hemofilia B trials due to lack of efficacy raised other important issues. All animal tests using the AAV.fIX product showed positive results from mouse to subhuman primates. The lack of efficacy in humans is now attributed (at least in part) to a memory T-cell mediated immune response against the AAV vector, which is absent in animals. However, no human-specific immune response to Ad5 vectors has been identified to date.

## 4 CONCLUSIONS

The progress of gene therapy in the past 17 years has been slower than was expected, which is due to several factors. Gene therapy is a pioneering new therapeutic modality based on complex biological systems occurring at the leading edge of biomedical knowledge. Incomplete knowledge of the genes involved in the pathomechanism of diseases constitutes a limit to generate clinically effective gene therapies, especially in complex diseases with multiple interacting genetic and environmental factors. Stringent and time-consuming safety studies are needed as is the establishment of new regulatory frameworks essential to control the applications of gene therapy and ensure safety to the patient and the population at large.

High costs are involved in the R&D of gene therapy and complex issues of intellectual property and commercial rights need to be resolved.

Despite the early high exceptions and the subsequent setbacks, one has to recognize that this new therapeutic modality is still in its infancy. It will neither deliver medical "miracles" (as its early proponents predicted) nor will it "disappear" because of a few disappointing setbacks (as some of its recent antagonists predict). As with all new technologies, gene therapy has to run its course in its current "development phase" before it reaches "maturation," when its full potential will be exploited.

## REFERENCES


1. Edelstein M L, Abedi M R, Wixon J, Edelstien R M (2004). Gene therapy clinical trials worldwide: 1989–2004—an overview. *J. Gene Med*. 6:597–602.
2. Harui A, Suzuki S, Kochanek S, et al. (1999). Frequency and stability of chromosomal integration of adenovirus vectors. *J. Virol*. 73:6141–6146.
3. Wilson J M (1996). Adenoviruses as gene delivery vehicles. *New Eng. J. Med*. 334: 1185–1187.
4. Svensan E C, Marshall D J, Woodard K, et al. (1999). Efficient and stable transduction of cardiomyocytes after intramocardial injection or intracoronary perfusion with recombinant adeno-associated virus vector. *Circulation*. 99:201–205.
5. Gao G P, Alvira M R, Wang L, et al. (2002). Novel adeno-associated viruses from rhesus monkeys as vectors for human gene therapy. *Proc. Nat. Acad. Sci. USA*. 99(18):11854–11859.
6. Li S, Huang L. (2000). Nonviral gene therapy: Promises and challenges. *Gene Ther*. 7:31–34.
7. Plank C, Mechtler K, Szoka Jr F C, et al. (1996). Activation of the complement system by synthetic DNA complexes: A potential barrier for intravenous gene delivery. *Human Gene Ther*. 7:1437–1446.
8. Stephan D J, Yang Z-Y, San H, et al. (1996). A new cationic liposome DNA complex enhances the efficiency of arterial gene transfer in vivo. *Human Gene Ther*. 7:1803–1812.
9. Turunun M P, Hiltunen M O, Ruponen M, et al. (1999). Efficient adventitial gene delivery to rabbit carotid artery with cationic polymer-plasmid complexes. *Gene Ther*. 6:6–11.
10. Young Jr S M, McCarty D M, Degtyareva N, et al. (2000). Roles of adeno-associated virus Rep protein and human chromosome 19 in site-specific recombination. *J. Virol*. 74:3953–3966.
11. Asahara T, Kalka C, Isner J M (2000). Stem cell therapy and gene transfer for regeneration. *Gene Ther*. 7:451–457.
12. Gage F H (1998). Cell therapy. *Nature*. 392(suppl.):18–24.
13. Su L, Lee R, Bonyhadi M, et al. (1997). Hematopoietic stem-cell based gene therapy for acquired immuno-deficiency syndrome: Efficient transduction and expression of RevM10 in myeloid cells in vivo and in vitro. *Blood*. 89:2283–2290.
14. Boekstegers P, von Degenfeld G, Giehrl W, et al. (2000). Myocardial gene transfer by selective pressure regulated retroinfusion of coronary veins. *Gene Ther*. 7:232–240.
15. Takeshita S, Zheng L P, Brogi E, et al. (1994). Therapeutic angiogenesis: A single intraarterial bolus of vascular endothelial growth factor augments revascularization in a rabbit ischemic hind limb model. *J. Clini. Invest*. 93:662–670.

16. Giordano F J, Ping P, McKirnan M D, et al. (1996). Intracoronary gene-transfer of fibroblast growth factor-5 increases blood flow and contractile function in an ischemic region of the heart. *Nature Med*. 2:534–539.
17. Mack C A, Aptel S R, Schwarz E A, et al. (1998). Biologic bypass with the use of adenovirus-mediated gene transfer of the complementary deoxyribonucleic acid for vascular endothelial growth factor 121 improves myocardial perfusion and function in the ischemic porcine heart. *J. Thorac. Cardiovasc. Surg*. 115:168–177.
18. Shyu K G, Manor O, Magner M, Yancopoulos G D, Isner J M (1998). Direct intramuscular injection of plasmid DNA encoding angiopoietin-1 but not angiopoietin-2 augments revascularization in the rabbit ischemic hindlimbs. *Circulation*. 98:2081–2087.
19. Tsurumi Y, Takeshita S, Chen D, et al. (1996). Direct intramuscular gene transfer of naked DNA encoding vascular endothelial growth factor augments collateral development and tissue perfusion. *Circulation*. 94:3281–3290.
20. Banai S, Jaklitsch M T, Shou M, et al. (1994). Angiogenic-induced enhancement of collateral blood flow to ischemic myocardium by vascular endothelial growth factor in dogs. *Circulation*. 89:2183–2189.
21. Lawrie A, Brisken A F, Francis S E, et al. (1999). Ultrasound enhances reporter gene expression after transfection of vascular cells in vitro. *Circulation*. 99:2617–2620.
22. Villanueva F S, Jankowski R J, Klibanov S, et al. (1998). Microbubbles targeted to intercellular adhesion molecule-1 bind to activated coronary artery endothelial cells. *Circulation*. 98:1–5.
23. Yang N-S, Sun W H (1995). Gene gun and other non-viral approaches for cancer gene therapy. *Nature Med*. 1:481–483.
24. Mir L M, Burean M F, Gehl J, et al. (1999). High-efficiency gene transfer into skeletal muscle mediated by electric pulses. *Proc. Nat. Acad. Sci. USA*. 96:4262–4267.
25. Bergelson J M, Cunningham G D, Kurt-Jones E A, et al. (1997). Isolation of a common receptor for coxsacki B viruses and adenoviruses 2 and 5. *Science*. 275:1320–1323.
26. Nemerow G R, Stewart P L (1999). Role of αv integrins in adenovirus cell entry and gene delivery. *Microbiol. Molec. Biol. Rev*. 63:725–734.
27. Wickham T J, Mathias P, Cheresh D A, et al. (1993). Integrins αvβ3 and αvβ5 promote adenovirus internalization but not virus attachment. *Cell*. 73:309–319.
28. Gaggar A, Shayakhmetov D M, Lieber A (2003). CD46 is a cellular receptor for group B adenoviruses. *Nature Med*. 9(11):1408–1412.
29. Douglas J T, Rogers B E, Rosenfeld M E, et al. (1996). Targeted gene delivery by tropism-modified adenoviral vectors. *Nature Biotechnol*. 14:1574–1578.
30. Wickham T J, Roelvink P W, Brough D E, et al. (1996). Adenovirus targeted to heparin-containing receptors increases its gene delivery efficiency to multiple cell types. *Nature Biotechnol*. 14:1570–1573.
31. Wickham T J, Tzeng E, Shears L L, et al. (1997). Increased in vitro and *in* vivo gene transfer by adenovirus vectors containing chimeric fiber proteins. *J. Virol*. 71:8221–8229.
32. Krasnykh V N, Douglas J T, van Beusechem V W (2000). Genetic targeting of adenoviral vectors. *Molec. Ther*. 1:391–405.
33. Wickham T J (2000). Targeting adenovirus. Gene *Ther*. 7:110–114.
34. Green N K, Herbert C W, Hale S J, et al. (2004). Extended plasma circulation time and decreased toxicity of polymer-coated adenovirus. *Gene Ther*. 11:1256–1263.
35. Fisher K D, Stallwood Y, Green N K, et al. (2001). Polymer-coated adenovirus permits efficient retargeting and evades neutralizing antibodies. *Gene Ther*. 8(5):341–348.

36. Pang S, Dannull J, Kaboo R, et al. (1997). Identification of a positive regulatory element responsible for tissue-specific expression of prostate-specific antigen. *Cancer Res.* 57:495–499.
37. Ko S C, Cheon J, Kao C, et al. (1996). Osteocalcin promoter-based toxic gene therapy for the treatment of osteosarcoma in experimental models. *Cancer Res.* 56:4614–4619.
38. Shibata T, Giaccia A J, Brown J M (2000). Development of a hypoxia-responsive vector for tumor-specific gene therapy. *Gene Ther.* 7:493–498.
39. Thule P M, Liu J, Phillips L S (2000). Glucose regulated production of human insulin in rat hepatocytes. *Gene Ther.* 7:205–214.
40. Varley A W, Munford R S (1998). Physiologically responsive gene therapy. *Molecular Med. Today.* 4:445–451.
41. Clackson T (2000). Regulated gene expression systems. *Gene Ther.* 7:120–125.
42. Gossen M, Bujard H (1992). Tight control of gene expression in mammalian cells by tetracycline-responsive promoters. *Proc. Nat. Acad. Sci. USA*. 89:5547–5551.
43. No D, Yao T P, Evans R M (1996). Ecdysone-inducible gene expression in mammalian cells and transgenic mice. *Proc. Nat. Acad. Sci. USA*. 93:3346–3351.
44. Wang Y, O'Malley Jr B W, Tsai S Y, et al. (1994). A regulatory system for use in gene transfer. *Proc. Nat. Acad. Sci. USA*. 91:8180–8184.
45. Ho S N, Biggar S R, Spencer D M, et al. (1996). Dimeric ligands define a role for transcriptional activation domains in reinitiation. *Nature*. 382:822–826.
46. Magari S R, Rivera V M, Iuliucci J D, et al. (1997). Pharmacologic control of humanized gene therapy system implanted into nude mice. *J. Clin. Invest*. 100:2865–2872.
47. Rivera V M, Clackson T, Natesan S, et al. (1996). A humanized system for pharmacologic control of gene expression. *Nature Med*. 2:1028–1032.
48. Rosenberg S A, Blaese R M, Brenner M K, et al. (2000). Human gene marker/therapy clinical protocols. *Hum Gene Ther.* 11:919–979.
49. Rosenberg S A, Aebersold P, Cornetta K, et al. (1990). Gene transfer into humans—immunotherapy of patients with advanced melanoma, using tumor-infiltrating lymphocytes modified by retroviral gene transduction. *New Eng. J. Med*. 323:570–578.
50. Rubanyi G M (2001). The future of human gene therapy. *Molec. Aspects Med*. 22: 113–142.
51. Gao M H, Lai N C, McKirnan M D, et al. (2004). Increased regional function and perfusion after intracoronary delivery of adenovirus encoding fibroblast growth factor-4: Report of preclinical data. *Human Gene Ther.* 15:574–587.
52. Grines C L, Watkins M, Mahmarian J, et al. (2003). A randomized double blind placebo-controlled trial of Ad5FGF-4 gene therapy and its effect on myocardial perfusion in patients with stable angina. *J. Amer. Coll. Cardiol*. 42:1339–1347.
53. Grines C L, Watkins M W, Helmer G, et al. (2002). Angiogenic Gene Therapy (AGENT) trial in patients with stable angina pectoris. *Circulation*. 105(11):1291–1297.
54. Rubanyi G M (2004). The design and preclinical testing of Ad5FGF-4 to treat chronic myocardial ischemia. *Euro. Heart J.* 6(suppl. E):E12–E17.
55. Watkins M W, Rubanyi G M (2003). Gene therapy for coronary artery disease: Preclinical and initial clinical results with intracoronary administration of Ad5FGF-4. In G. M., Rubanyi, S., Ylä-Herttuala, (eds.), *Human Gene Therapy: Current Opportunities and Future Trends*, Springer, Berlin, Germany, pp. 61–80.
56. Delli Bovi P, Curatola A M, Kern F G, et al. (1987). An oncogene isolated by transfection of Kaposi's sarcoma DNA encodes a growth factor that is a member of the FGF family. *Cell*. 50:729–737.

57. Gomez-Foix A M, Coats W S, Baque S, et al. (1992). Adenovirus-mediated transfer of the muscle glycogen phosphorylase gene into hepatocytes confers altered regulation of glycogen metabolism. *J. Biol. Chem*. 267:25129–25134.
58. McGrory W J, Bautista D S, Graham F L (1988). A simple technique for the rescue of early region 1 mutations into infectious human Adenovirus Type 5. *Virol*. 163:614–617.
59. Bett A J, Krougliak V, Graham F L (1995). DNA sequence of the deletion/insertion in early region 3 of Ad5 dl309. *Virus Res*. 39:75–82.
60. Lehmberg E, Traina J A, Chakel J A, et al. (1999). Reversed-phase high performance liquid chromatographic assay for the adenovirus type 5 proteome. *J. Chromatog.* B. 732:411–423.
61. Lehmberg E, McCaman M, Traina J, et al. (2004). Analytical assays to characterize adenoviral vectors and theoretic applications. In G. Subramanian, (ed.), *Manufacturing of Gene Therapeutics*, Kluwer Academic/Plenum Press, Dordrecht, the Netherlands, pp. 210–225.
62. Murakami P, Pungor E, Files J, et al. (2002). A single short stretch of homology between adenoviral vector and packaging cell line can give rise to cytopathic effect-inducing, helper-dependent E1-positive particles. *Human Gene Ther*. 13:909–920.
63. Walsh P S, Varlaro J, Reynolds R (1992). A rapid chemiluminescent method for quantitation of human DNA. *Nucleic Acids Res*. 20:5061–5065.
64. Chroboczek J, Bieber F, Jacrot B (1992). The sequence of the genome of adenovirus type 5 and its comparison with the genome of adenovirus type 2. *Virology*. 186(1):280–285.
65. Isner J M, Vale P R, Symes J F, et al. (2001). Assessment of risks associated with cardiovascular gene therapy in human subjects. *Circulation Res*. 89:389–400.
66. Ylä-Herttuala S, Martin J F (2000). Cardiovascular gene therapy. *Lancet*. 355:213–222.

## FURTHER READING

Isner J M (1997). Angiogenesis and collateral formation. In K.L. March, (ed.), *Gene Transfer in the Cardiovascular System: Experimental Approaches and Therapeutic Implications*. Kluwer Academic Publishers, Boston, M.A., pp. 307–330.

# PLASMID DNA AND MESSENGER RNA FOR THERAPY

STEVE PASCOLO
*University of Tübingen, Institute for Cell Biology, Tübingen, Germany*

**Chapter Contents**

1 Nonviral Coding Nucleic Acids: Definition and Characteristics in Relation to Potential Therapeutic Use
  1.1 Historical Overview
  1.2 DNA: The Stable Genetic Information Store
  1.3 RNA: The MultiFunction Molecule
  1.4 Recognition of Nucleic Acids by Immune Cells
  1.5 Production of Plasmid DNA and Messenger RNA
  1.6 Packaging of Nucleic Acids for Therapy
2 Plasmid DNA and Messenger RNA for Therapy
  2.1 Design and Optimization of pDNA and mRNA Vectors
  2.2 Vaccination
  2.3 Gene Therapy
  2.4 Production and Regulations
3 Conclusions and Perspectives
  References

## 1 NONVIRAL CODING NUCLEIC ACIDS: DEFINITION AND CHARACTERISTICS IN RELATION TO POTENTIAL THERAPEUTIC USE

### 1.1 Historical Overview

Extracted from the nucleus of human cells in the kitchen of Tübingen's castle (Germany), Nuclein is an heterogeneous product that was introduced in 1868 by Friedrich Miescher (1844–1895). Miescher's pupil, Richard Altmann (1852–1900),

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

further fractionated Nuclein and isolated in 1889 the substance that he named "nucleic acids." These molecules were characterized by Albrecht Kossel (1853–1927, winner of the Nobel Prize for Medicine in 1910). Kossel used chemical hydrolysis of nucleic acids to identify their basic components: the nucleotides. Five nucleotides are found in the Nuclein: two purines [adenylic acid (A) and guanylic acid (G)] and three pyrimidines [cytidylic acid (C), uridylic acid (U), and thymidylic acid (T)]. In 1910, Kossel could also distinguish between two families of nucleic acids present in Nuclein: one is degraded at high pH and contains A, C, G, and U; one is resistant and contains A, C, G, and T (Figure 1). The precise chemical structure of the nucleotides was deciphered by a student of Kossel: Phoebus Levene (1869–1940) together with Walter Abraham Jacob (1883–1967) in 1929. These researchers found that nucleotides consist of a nitrogen base (Adenine, Guanine, Cytosine, Thymine, or Uracile), a sugar (pentose), and a phosphate group. They also found that nucleotides are linked linearly in the order of phosphate-sugar thanks to phosphodiester bonds. This way, they generate a polymer with a phosphodiester backbone. The pentose in the high-pH-sensitive nucleic acid is a ribose, whereas in the high-pH-resistant nucleic acid, it is a 2′-deoxyribose. Based on this chemical difference are the names of the two nucleic acids: RNA for

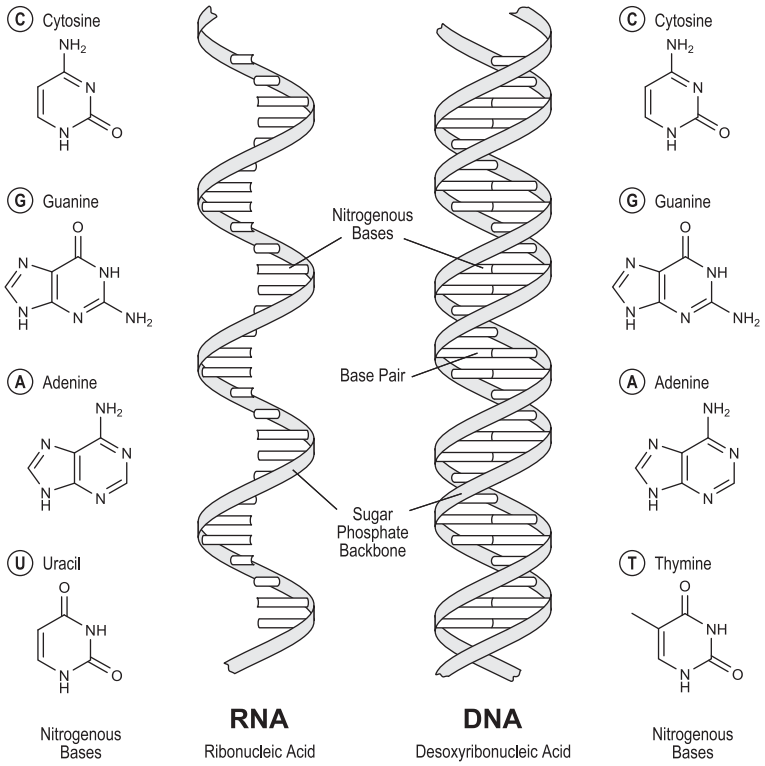


**Figure 1.** Structure of DNA and RNA molecules. The structure of the four nucleotides that compose RNA (on the left) or DNA (on the right) is shown. The representation of linear RNA and DNA molecules is shown in the middle. Actually, RNA makes a secondary structure through intra-strand base pairing (not shown).

the ribo-nucleic acid and DNA for the deoxyribo-nucleic acid. Although the Nuclein contains RNA and DNA, the cytosole of cells was shown by Tornbjörn Caspersson (1911–1998) and Jean Brachet (1909–1988) to contain only RNA. In 1953, Jim Watson (1928–), Francis Crick (1916–2004), and Maurice Wilkins (1917–), who received the Nobel Prize in 1962 for their resolution of the structure of DNA, demonstrated that, thanks to specific hydrogen bonds between their base moieties, nucleotides are complementary: A pairs with T or U using two hydrogen bonds and G pairs with C using three hydrogen bonds. Helped by the x-ray photograph of DNA (called photograph 51) obtained by Rosalind Franklin (1920–1958), they proposed that DNA is a molecule with two antiparallel strands shaped in a regular helix of a 2 nm diameter and 3.4 nm step (10 nucleotides for one turn) with its phosphodiester backbone on the outside and the bases on the inside of the helix (Figure 1). One year earlier, in 1952, Alfred Hershey (1908–1997, Nobel Prize winner in 1969) and Martha Chase (1927–2003) demonstrated using T2 phages that DNA is the genetic material in bacteria. There results confirmed the conclusions of Oswald Theodore Avery (1877–1955) who showed in 1943 that the "transforming principle" that can modify *Pneumococcus*'s phenotype from "smooth" to "rough" is DNA. By deduction and because of the location and behavior of DNA in eukaryotic cells, it was admitted, before being experimentally demonstrated, that DNA is the genetic material in higher organisms as well.

Meanwhile, the characterization of the cytosolic RNA-rich granules called ribosomes and identified by Albert Claude (1898–1983, Nobel Prize in 1974) indicated that RNA is involved in protein synthesis [1, 2]. Such a function of RNA was first hypothesized by Torbjoern Caspersson (1910–1997) and Jean Brachet (1909–1998) at the end of the 1930s and was then based on the relation between the RNA content of cells and their activity (proliferation). Ribosomes contain ribosomic RNA (rRNA) and are the site of protein production. The key intermediate between nucleic acids and proteins is the hybrid aminoacyl-RNA molecule called transfer RNA (tRNA) identified in 1958 [3]. It consists in a family of at least 20 members in any organism. A tRNA is a short (between 75 and 85 nucleotides) RNA molecule linked at its 3′ end to one amino-acid [4, 5]. Finally, the identification of an unstable RNA called messenger (mRNA) [6, 7] which is a copy of genes to be translated into proteins completed the basic description of the fundamental protein machinery. It is summarized in Crick's central dogma of molecular biology (Figure 2): DNA contained in the nucleus is transcribed into (messenger) RNA, which is translated into proteins [8]. Since the formulation of this central dogma, the molecular mechanisms involved in these processes have been characterized in detail. Of interest, we now know that the transient nature of all RNA molecules compared with DNA is not due to the chemical structure difference (in cells, pH are neutral or acidic: DNA and RNA are practically equally stable at the chemical level) but due to the abundance inside and outside the cells of potent and stable ribonucleases (RNases). We have also learned that RNA has more functions than initially postulated: It can replace DNA as genome in viruses and eventually (i.e., in retroviruses) be reverse transcribed into DNA. Moreover, several types of RNA molecules with various biological activities (structure, enzyme activities, containment of gene information) are involved in each step of gene expression (see Section 1.3). Nowadays, thanks to the molecular biology methods, we can also introduce designed foreign DNA and RNA molecules into cells (Figure 2). As proposed by Crick

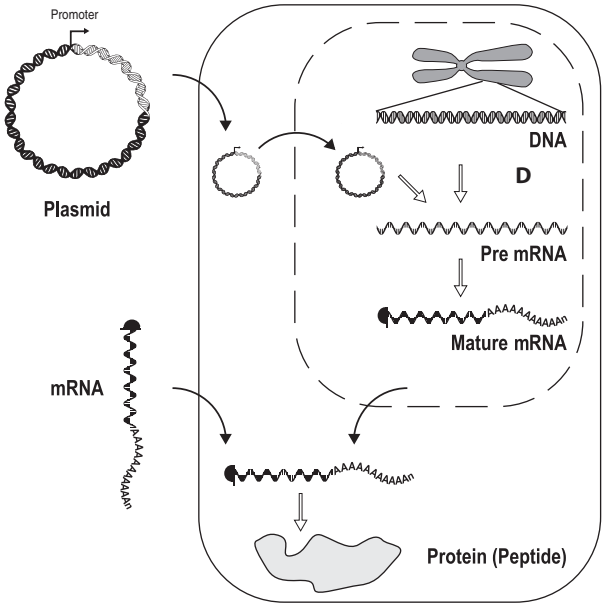


**Figure 2.** Central Dogma. The genes located on the chromosomes are transcribed into mRNA, which is exported from the nucleus to the cytosole where it is translated. Thanks to molecular biology techniques, it is possible to bring exogenous genetic information. Foreign DNA molecules must cross two membranes: the cytoplasmic and the nuclear envelopes, whereas mRNA must cross only the cytoplamic membrane to be active.

et al. in 1961, through a genetic code based on three letters (three consecutive bases: a codon), the nucleic acid information copied from the DNA genome into mRNA is translated in proteins. Each possible codon ($4^3 = 64$ possibilities) corresponds to one precise amino acid. This genetic code is universal: It is conserved between all living organisms. As there are only 20 amino acids, most of them are encoded by more than one codon. Some codons are favored (more frequently found in highly expressed genes) in a species-specific manner. For example, alanine, which is encoded by four codons would be preferably encoded by GCA in *Bacillus subtilis*, by GCC in *Homo sapiens*, by GCG in *Escherichia coli*, and GCT in *Saccharomyces cerevisiae*. The correspondence between codons and amino acids, i.e., the fundamental translation process, is performed by specialized enzymes called aminoacyl tRNA synthetases. There are at least 20 different aminoacyl tRNA synthetases in any organism. Each is responsible for loading one amino acid on the adequate tRNA(s). By transferring a specific amino acid on a specific tRNA, which has the adequate anticodon (complementary to the codon in the mRNA), aminoacyl tRNA synthetase convert genetic information (nucleotides) into protein information (amino acids).

## 1.2 DNA: The Stable Genetic Information Store

In eukaryotic cells, DNA molecules are located in the nucleus. During the interphase (between two cell divisions) of the cell cycle, DNA molecules are packed by histones into an heterogenic structure called the chromatin. Through intense con-

densation of the chromatin, each DNA molecule can be visualised as a chromosome during mitosis or meiosis. A human cell contains 23 pairs of DNA molecules that consist altogether ca. $3.15 \times 10^9$ bases. These molecules stretched would represent altogether a ca. 2-m-long chain. They have a stable nucleotide sequence that is eventually modified only at the ends (the telomeres). At this location, an inducible ribo–nucleo–protein complex that contains a reverse transcriptase (telomerase) uses a matrix RNA molecule to elongate the DNA molecules with a repetitive sequence motif. In proliferating cells not expressing the telomerase, the telomeres shorten at each division due to incomplete replication of DNA molecules at their extremities. Below a certain size of the telomeres (the Hayflick limit), the cell undergoes senescence and starts a self-destruction program (apoptosis). Thus the shortening of telomeres at each replication cycle is a method to "date" a cell: to know how many times it divided. For cells that need to proliferate (germ and stem cells, for example), the telomerase complex is induced and maintains or extends the telomeres' length, keeping the cell "young." Aside from the telomeres, the sequence of the chromosomic DNA molecules is unmodified in healthy cells. However, genomic DNA molecules have a dynamic structure in constant reorganization and modification: Methylations of residues, condensation into a closed chromatin configuration, opening into an accessible chromatin configuration, or changes of the double-helix structure (A, B or Z) can affect the availability of the DNA sequences for gene expression. Several important conserved sequences define a functional DNA stretch. Many of them have a role that is not yet characterized. Only ca. 2% of the human genome is functionally identified as being genes. These entities are characterized by a promoter that is the starting point of mRNA transcription. Promoters are recognized by the RNA polymerase II (the RNA polymerase responsible for the production of mRNA in eukaryotes). They have a minimal core sequence of a few nucleotides [usually a "TATA(A/T)A (A/T)" sequence called a TATA box or Hogness box], 19 to 27 bases upstream of the initiation site of the mRNA. The TATA box is usually surrounded by G- and C-rich sequences. Its efficacy largely depends on upstream proximal control sequences. The so-called "CAAT Box" (consensus GG(C/T)CAATCT) and "GC box" (consensus GGGCGG) enhance the activity of the promoter, i.e., its utilization by the RNA polymerase II. Distal upstream and downstream sequences called "Enhancers" also impact the activity of the promoter. An efficacious promoter in eukaryotic cells requires ca. 100 bases upstream from the TATA box. Highly specialized promoters such as "Locus Control Regions" that also affect the local chromatin structure and keep it in an opened configuration can increase the efficacy of transcription. Downstream of the promoter is the transcribed sequence that is segmented between exons (sequences that will be present in the cytosolic mRNA) and introns (sequences that will be spliced-out from the neo-transcribed mRNA in the nucleus). Transcription of mRNA stops at a transcription stop signal, which is usually a U-repeat. Before it leaves the nucleus, the immature mRNA is further processed as described bellow.

### 1.3 RNA: The MultiFunction Molecule

As opposed to DNA that has a well-determined unique function that is to stably store the genetic information, RNA is a pluri- (or even omni-)potent molecule.

RNA molecules can have all functions of leaving: storage of genetic information, enzymatic activity, and construction of defined three-dimensional structures. RNA is the only biomolecule that has all of these activities. Thus, it was proposed as the primordial macromolecule that, on its own, started the concept of life 4.2 billion years ago [9].

***Structure-related Functions (rRNA, tRNA, aptamers).*** Because of its single-stranded nature, RNA refolds according to the most thermodynamically favored base-pairing. This way, it generates stem loops which position each other into complex three-dimensional structures. It is still hard to predict the dominant structures that will adopt a precise RNA sequence, but several algorithms can help in predicting them (see www.bioinfo.rpi.edu/applications/mfold/rna/form1.cgi).

The sequence-dictated three-dimensional conformation of an RNA molecule is essential in the functionality of several RNA types: The ribosomic RNAs fold into structures that scaffold the ribosomes, and the tRNAs fold into defined structures that allow their recognition by both amino-acyl tRNA synthetase and ribosomes. In both cases, some nucleotide sequences that remain single stranded are involved in the enzymatic processes performed by the molecules. Thus, the sequence of an RNA dictates its structure and allows it to perform its function. Both are strictly connected. One single mutation in a rRNA or a tRNA may completely change the favored three-dimensional structure of the molecule and totally impair the RNA's function(s).

The scaffolding characteristic of RNA molecules is used in drugs under the generic name of aptamers. Selected for their capacity to interact precisely with a target structure similarly to the antibody–antigen interaction, aptamers can specifically bind to any molecule or complex of interest. The first pharmaceutical aptamer is Macugen™ (Eyetech Pharmaceuticals, Inc., and Pfizer, Inc.). It recognizes and neutralizes specifically the vascular endothelial growth factor, thus inhibiting pathologic blood vessel growth. It is used as treatment against age-related acute macular degeneration.

***Regulation of Protein Expression-related Functions (antisense, siRNA, miRNA).*** Small RNA molecules of ca. 20 nucleotides in length can remain linear with all their bases available to pair with a complementary sequence present on another RNA molecule. Such "antisense" RNA oligonucleotides can this way impair the function (block translation) of mRNA. Another type of short RNA molecules of ca. 20 nucleotides in length is double stranded and has two bases overhangs. It is called small inhibitory RNA (siRNA) when naturally produced in the cytosole out of foreign double-stranded RNA molecules or micro-RNA (miRNA) when encoded in the DNA genome. Through its recognition by a proteic complex called RISK, one strand is discarded (the sense strand) and the other strand (the antisense) is used as a guide to direct RISK to a complementary sequence in a mRNA and induce a cleavage on this target molecule. Thus, siRNA or miRNA are a catalytic version of antisense RNA: one molecule can inhibit (induce the degradation of) many mRNA. Some miRNAs can also modify the DNA sequence in the promoter region by inducing methylations. This way, they dictate the level of packaging into the chromatin and the expression of a gene. For more information on small RNA molecules as regulators of gene expression and their use as therapeutics, refer to Chapter 7.6.

***Enzymatic Functions (ribozymes).*** Through a fixed three-dimensional structure and the availability of certain nonpaired residues, RNA molecules can fold into enzymes called ribozymes. The evidence of such an enzymatic capacity of RNA was discovered by Altmann [10] and Cech [11]. It was acknowledged with a Nobel Prize (Chemistry, 1989). Initially, this enzymatic activity was restricted to the capacity of ribozymes to recognize and cleave or religate in an autonomous way a target RNA (mRNA) sequence. Later, other enzymatic capacities of ribozymes were evidenced: Selected ribozymes can catalyse the synthesis of nucleotides [12] or the creation of amide bonds [13].

***Coding Functions (mRNA).*** The pre-messenger RNA produced by transcription of the genomic DNA is processed before leaving the nucleus: Splicing of introns is made by spliceosomes that are ribo–proteic complexes. Within these structures, the RNA provides the sequence specificity of the spliceosome's activity. An exon does not have any sequence recognized by the spliceosome. It is the intron, and its borders that are recognized and excised. Evolutionarily, introns may derive from movable genetic elements called transposons. Thus, theoretically, introns can be introduced naturally (through evolution) or experimentally (thanks to molecular biology methods) within any coding sequence. They present the advantage that they enhance the efficacy of gene expression, probably through the facilitation of mRNA export to the cytosole. Moreover, thanks to regulated alternative splicing mechanisms, some specific exons may be spliced out with flanking introns. Such alternatively spliced mRNAs encode proteins with missing domain(s) and thus with different function(s). Meanwhile, the deregulation of the splicing mechanisms in dysfunctional cells, such as tumor cells, may allow introns to remain in the mature mRNA. Then, a protein with an extra domain and eventually truncated (if the intron contains a stop codon in the frame of translation) is generated. Aside from being spliced, the immature mRNA is also capped at its 5′ end (the addition of a methylated guanine followed by three phosphates bonds) and poly-adenylated at its 3′ end (the addition of several hundreds of adenine residues). The spliced, capped, poly-adenylated mRNA is exported to the cytoplasm through nuclear pores. In the cytosole, several elongation initiation factors recognize the Cap structure, whereas the poly-A binding protein binds to the 3′ poly A tail. These proteins allow the ribosomes to bind to the mRNA and start to scan from the 5′ end the nucleotide sequence until it finds a start codon (ATG) in a "Kozak" surrounding (A/GNNNAUGG). At this location, the first amino acid, always a Methionine, is brought by a tRNA that has the 5′ CAU 3′ anticodon. Then, the ribosome translocates toward the 3′ end to the next codon where an adequate tRNA will bring the second amino acid encoded by the gene. The process continues toward the 3′ end of the mRNA and stops when the ribosome meets a UGA, UAA, or UAG "termination" codon. The neosynthesized protein and the ribosomes are then released. Several ribosomes are sitting on one mRNA. According to its stability and efficacy of translation, one mRNA can produce many proteins. Ultimately, the mRNA is degraded by intracellular RNases. This catabolic process is highly regulated: Several sequences in the mRNA, and especially at its ends, before and after the coding sequence (UnTranslated Regions: UTR), have primary and secondary structures that are recognized by specific proteins in the cytosole. Those proteins may induce or, on the contrary, prevent the detection and degradation of the mRNA by the intracellular RNases, which are in general exonucleases [10].

This means that they degrade the mRNA starting by the 5′ or 3′ ends. Both ends are protected in the mature mRNA: the Cap on the 5′ and the poly-A tail covered by the poly-A binding proteins on the 3′. Specific and regulated mechanisms of decapping and deadenylation will render the mRNA accessible to the exonucleases. Accelerated degradation (shortening mRNA half-life) can be induced by the presence of AU-rich sequences called AUREs in the 3′ UTR, or by the presence of a specific restriction site. This latter possibility is illustrated by the regulation of iron metabolism: An iron-induced endoribonuclease recognizes a sequence present in the mRNA coding for the transferrin receptor (TfR) that is responsible for iron-loaded transferin uptake. The presence of high iron stocks in the cell induces the endonuclease, which cleaves the TfR-encoding mRNA. Thus, it prevents TfR expression and further iron uptake. Inhibition of mRNA degradation (increasing mRNA half-life) involves other untranslated sequences as illustrated with the very stable globin mRNAs: They contain in their 3′ UTR, pyrimidine-rich sequences that are recognized by a ubiquitous protein complex called the alpha-complex. It stabilizes the poly-A binding proteins associated with the poly-A tail and secures the integrity of the 3′ end of the mRNA, thus increasing the half-life of the whole RNA molecule. Depending on the presence of stabilization and destabilization sequences in its UTRs, its length, structure, and efficacy of translation, the intracellular mRNA can have a half-life varying from minutes to weeks [11]. Outside the cells, naked mRNA are degraded within seconds because of the abundant, ubiquitous, stable, and processive extracellular RNases [14]. The presence of such an efficacious machinery to degrade extracellular RNA may have been developed as a protection against viruses, especially those with RNA genomes. It may also control the secretion and recapture of RNA by neighboring cells, which is hypothesized to represent a way for cells to communicate [15].

### 1.4 Recognition of Nucleic Acids by Immune Cells

In eukaryotic cells, DNA is strictly in the nucleus and in energy-producing sub-compartments (mitochondria and chloroplast), whereas RNA is strictly in those same compartments plus the cytosole. In healthy situations there is no nucleic acid in the endoplasmic reticulum, golgi apparatus, endosomes or lysosomes, and associated vesicles. The immune system developed the tools to scrutinize for the presence of nucleic acids in those compartments because it would be associated with a pathogenic situation such as the presence of an infectious agent. A family of receptors similar to immunity-related receptors from flies called Tolls receptors is dedicated to the detection of such "danger signals." This family of Toll-like receptors (TLRs) contains 11 known members. Four of them were shown to recognize nucleic acids: TLR3 is specific for double-stranded RNA (dsRNA), TLR7 and TLR8 recognize single-stranded RNA (ssRNA), and TLR9 recognizes an unmethylated DNA pattern ("CpG motif") that is found specifically in the genome of microorganisms such as bacteria (Figure 3). These four receptors are located in the lumen of cytoplasmic vesicles (mainly endosomes) in immune cells like B cells, NK cells, granulocytes, macrophages, or dendritic cells (DCs). TLR3, 7, and 8 detect the infection by viruses such as influenza (dsRNA genome) or HIV (ssRNA genome), whereas TLR9 signals the presence of bacteria in the endosomes. Upon engagement of TLRs, immune cells are activated. They trigger an adaptive (involving T

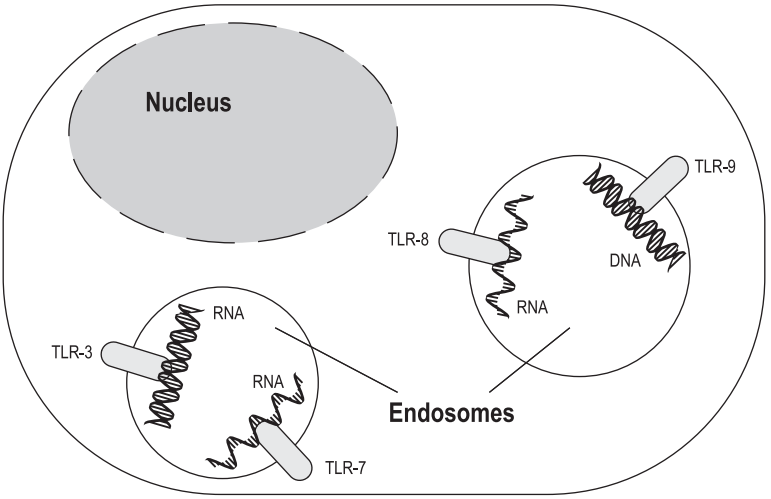


**Figure 3.** Recognition of nucleic acids by TLR. TLR 3, –7, –8, and –9 are located in the endosomes and recognize dsRNA (TLR3), stabilized ssRNA (TLR7 and TLR8), and bacteria DNA (TLR9).

and B lymhocytes) immune response specific for the antigens associated with or encoded by the foreign nucleic acid(s). This natural immune detection of exogenous nucleic acids is an advantage for the utilization of these molecules in vaccines but a disadvantage for the utilization of nucleic acids in gene therapy.

## 1.5 Production of Plasmid DNA and Messenger RNA

***Plasmid DNA.*** Plasmid DNA (pDNA) are small (from 2 to 15 kb) circular DNA molecules that contain a bacterial origin of replication. They can proliferate within a bacteria and represent up to 1% of the microorganims's mass. They are manipulated to code for a protein that neutralizes an antibiotic. Moreover, they can contain any desired sequence, such as for example a promoter, a gene from bacterial or non-bacterial origin, and a poly-adenylation signal. Plasmids are purified from bacteria grown in the presence of the antibiotic that they can neutralize. This guarantees that the bacteria will conserve and amplify their plasmids during fermentation. Frequently, the antibiotic is ampicillin, which is deactivated by the protein β-lactamase encoded by many standard plasmid vectors. The bacteria are pelleted, usually by centrifugation, and cells are lysed, usually thanks to sodium dodecyl sulfate (SDS) and sodium hydroxide. After neutralization with potassium acetate, most proteins and the genomic DNA precipitate while the plasmids remain in the aqueous phase. Several methods exist to purify pDNA from these solutions, but the most commonly used is anionic exchange chromatography (see, for example, the protocole by Qiagen at www.qiagen.com). Protocols are available to eliminate most lipopolysaccharides coming from the bacteria and that may affect the functionality of the pDNA for therapeutic utilizations. However, standard purification methods will provide the pDNA with three topologies: Most is supercoiled circular DNA, but some is in the form of a relaxed circle (one strand has at least one nick) or linear (both strands were broken at the same place). The

supercoiled plasmid is the form that should be used for research and therapy because, based on *in vivo* transcription studies, it is the most active form [16, 17]. Although the nicked pDNA molecules are most often not strongly affecting the functionality of the plasmid *in vitro* or *in vivo*, it is custom to use pDNA with a low amount of these topologies for preclinical and clinical utilizations. At laboratory scale, the purification of the supercoiled pDNA can be obtained using caesium chloride gradients. For larger scale and clinical trials, specialized companies offer the production of high-quality pDNA for research and therapy [Good Manufacturing Practice (GMP) certification; see, for example, plasmid factory at www.plasmidfactory.com]. One liter of bacteria culture can yield up to 30 mg of pDNA.

***Messenger RNA.*** Messenger RNA are produced from purified pDNA using bacterial RNA polymerases in a reaction called "*in vitro* transcription." The pDNA used for *in vitro* transcription must contain a specific promoter recognized by the bacteriophage RNA polymerases T3, T7, or SP6 in front of the coding sequence of interest. It is linearized using a unique restriction site that must be at the end of the sequence to be transcribed. Then, the linearized plasmid is mixed with the RNA polymerase and the four nucleotides plus, in standard reactions, an excess of Cap analog (m7G(5′)ppp(5′)G) compared with GTP. Usually the Cap analog is at ca. 6 mM in the reaction; ATP, UTP, and CTP are at 4 mM; and GTP is at 1.5 mM in the final reaction. The RNA polymerase synthesizes more than 100 mRNA molecules from one pDNA template. After the synthesis of the mRNA, 1 hour at 37°C, the DNA is destroyed by a DNase and the mRNA is recovered by precipitation, usually with lithium chloride. Afterward, the mRNA is resuspended in the adequate solution.

## 1.6 Packaging of Nucleic Acids for Therapy

Although some cells in skin, muscle, brain, and so on can spontaneously take up exogenous injected naked pDNA or mRNA [18], this process is relatively inefficient and can consequently be used only when a local expression of a protein is desired, i.e., for vaccination. Even if physical methods exist to enhance the uptake of naked nuclei acids (electroporation or hydrodynamic delivery; see Section 2.2), increasing the efficacy of transfection *in vivo* is best reached using en-capsulation of the nucleic acids. Two methods are used: cationic polymers or liposomes.

***Polymeric Particles.*** Cationic polymers can spontaneously associate with the anionic nucleic acids. Such complexes can keep a positively charged surface that allows an interaction with the negatively charged cell surfaces. Commonly used cationic polymers for the production of micro- or nanoparticles include chitosan, protamine, poly-L-lysine (PLL), poly-ethylenimine (PEI), and polyamidoamine dendrimers [19]. Usually, a simple mixing of the cationic polymers and the nucleic acid, eventually with high-speed homogenization, is enough to obtain particles. Thus, production is very easy and upscalable. However, some of these cationic polymers, such as PEI or PLL, are toxic or have secondary effects (for example, chitosan induces hypocholesterolemia). Moreover, particles generated by simple

mixing of nucleic acids and cationic polymers may be very heterogenous in size and physical properties, making them hard to qualify for pharmaceutical use. A more controlled method to produce particles is based on poly(lactide co-glycolide) (PLG). It is a biodegradable and biocompatible molecule used for several biomedical products including sutures. As it is a negatively charged molecule, it needs to be mixed with cationic polymers such as those listed above or chemical products [cetyltrimethylammonium bromide (CTAB), dimethyl dioctadecyl ammonium bromide (DDA), or 1,2-dioleoyl-1,3-trimethylammoniopropane (DOTAP), for example]. Optimally, a mixture of several reagents allows the production of homogenous particles with a defined size (from ca. 100 nm up to μm) and surface charge (zeta potential). When using PLG, protocols based on the solvent evaporation process are used [20]. Cationic particles can be manufactured and then coated with the nucleic acid or produced in the presence of the nucleic acid. In the first case, the genetic material is on the outside of the particles, and in the second in case, it is inside. Because the production of the PLG particles involves high shear, organic/aqueous interface, high temperature, and freeze-drying, the nucleic acid may be denatured or degraded during the process. Thus, for a production that is reliable, reproducible, and easy to upscale in GMP conditions, the coating of nucleic acids on the surface of premade cationic PLG particles is preferred. In both cases, the nucleic acid associated to the particles is protected from the activity of nucleases and can be applied by injection (the particles are phagocytosed by cells) or, for vaccination, by gene-gun [21]. The release of the nucleic acid is largely determined by the nature of the polymers used for the preparation of the particles. *In vitro*, the particles may readily release quickly a significant amount of nucleic acid in the injection solution. For example, PLG/CTAB microparticles released ca. 35% of their loaded pDNA within 1 day [20]. For this reason, particles should be stored lyophilized and injected quickly after reconstitution in the injection's buffer. *In vivo*, the biodegradability of the polymers and the site/method of delivery will dictate the release of the genetic information. For some particles, the release may be performed slowly, over several days [20]. Polymers with specific degradation patterns (for example, in endosomes: at low pH) can be engineered to increase the efficacy of delivery and regulate the time frame in which nucleic acids are available for expression within the cell. Some layer-per-layer production processes allow the production of particles with several layers of nucleic acids separated by layers of polymers (see www.capsulution.de). The choice of the polymer(s) will dictate the degradation profile of the particles and thus the release of the several layers of nucleic acids. Theoretically, such a strategy allows the slow release of nucleic acids within cells over a long time.

***Liposomes.*** Liposomes are vesicles that consist of an aqueous compartment enclosed in a phospholipids bilayer. As they lack proteinaceous components, liposomes are not immunogenic. A broad variety of cationic liposomes containing eventually modified lipids have been used to deliver pDNA and mRNA for therapeutic utilizations. These formulations are usually based on a cationic lipid such as 1,2-dioleol-3-trimethylammonium propane (DOTAP) and a fusogenic lipid such as 1,2-dioleoyl-sn-glycero-3-phosphoethanolamine (DOPE) [22, 23]. Despite their ease of production and efficacy *in vitro* for transfecting cultured cells, cationic liposomes face toxicity issues *in vivo* [24, 25]. Meanwhile, their heterogeneous size

and instability are drawbacks for a pharmaceutical production and qualification. Moreover, cationic liposomes are often inactivated in the presence of serum. As possible improvements of cationic liposomes, immunoliposomes, which are liposomes coated with an antibody, are developed. Thanks to their targeting, they are more efficacious than classic liposomes and can consequently be administered at a lower dosage. This way, the immunoliposomes are less toxic. Another variation of liposomes are called stealth liposomes. They consist in poly(ethylene glycol) (PEG)-coated liposomes. PEG prevents adsorption of the liposomes by endothelial cells after intravenous delivery. Thus, stealth liposomes are remaining in the circulation and are distributed more homogenously through the body than classic liposomes.

## 2 PLASMID DNA AND MESSENGER RNA FOR THERAPY

Plasmid DNA (pDNA) and messenger RNA (mRNA) are attractive tools for therapy because they are simple molecules that are easy to produce, store, detect, and manipulate using molecular biology techniques. They can be used to quickly turn into a drug any genetic information. Alternatives to pDNA and mRNA are virus- or bacteria-derived vehicles that can contain the genetic information to be expressed as a drug. The drawbacks of these methods are as follows:

1. Safety issues: The modified infectious agents may revert to a wild-type phenotype or may recombine with wild-type viruses to produce new sorts of pathogens.
2. Specificity: Not only the gene of interest is injected into a patient but also the proteins and eventually parts of the genetic information of the pathogen. The function of these proteins may interfere with the function of the transgene (in vaccination, for example, some pathogen-specific mechanisms may interfere with the priming of the immune response against the transgene), and moreover, an immune response against the vehicle (virus or bacteria) may be dominant. Not only could this hide the response against the relevant antigen for vaccination, but it could also prevent the repeated or chronic delivery of the drug for gene therapy.

Plasmid DNA and mRNA, which are both nonviral coding nucleic acid vehicles, are minimal genetic vectors: in their human host, they express only the therapeutic protein of interest. The two main utilizations of pDNA and mRNA for therapy are as follows:

1. Vaccination using simply the gene sequences of pathogens (viruses, bacteria, parasites, or tumors).
2. Genetic complementation after the gene responsible for a genetic disease is characterized.

Vaccination is by far the most exploited feature of nonviral coding nucleic acids. This is due to

1. The fact that a local, transient expression of a foreign protein is enough to trigger an immune response.
2. The natural capacity of pDNA and mRNA to stimulate the immune system through TLR.

As the production of nucleic acids is barely affected by their sequence, generic production facilities and methods were developed to produce GMP-certified pDNA or mRNA. Several specialized bioindustries throughout the world can supply within a few weeks any nucleic acid to be used in therapy. This feature is unique because all other polymeric biomolecules used in therapy (synthetic peptides, recombinant proteins, or viruses) face more or less severe production and storage issues that are changing depending on their sequence. Thus, pDNA and mRNA seem to be the most versatile and simplest biomolecules to be used as active pharmaceutical ingredients in new drugs. Both nucleic acids were used in clinical trials and are known to be safe. Efficacy was reported, but pDNA- and mRNA-based drugs should be optimized before such treatments are available on the market.

## 2.1 Design and Optimization of pDNA and mRNA Vectors

***Coding Sequence.*** As mentioned in Section 1, most amino acids are encoded by several codons, some of them being more favorable because they correspond to more abundant tRNA. Unfavorable codons or "rare" codons (as opposed to frequent, i.e., favorable codons) force the ribosome to pause during translation, and this may result in the premature ending of translation that generates abortive proteins. In any case, unfavorable codons decrease the efficacy of translation of full-length proteins. This phenomenon is particularly pronounced when the vaccine nucleic acid is a copy of a gene from a non-mammalian organism: Genes from bacteria or parasites have a codon usage very different from mammals. Moreover, some human virus genes and some endogenous human genes have an unfavorable codon usage that may be part of the mechanisms of regulation of their specific expression. For example, the HIV GAG gene has a "bad" codon usage that will limit the efficacy of gag production. This is actually important in the regulation of the virus replication's cycle. Thus, for human genes and more importantly for nonhuman genes, a codon optimization should preferably be performed in the coding part of therapeutic nucleic acids. The design of the gene is performed starting from the protein sequence to be expressed. Through algorithms that are based on the precise human codon usage a coding sequence will be proposed that should be ideally translated by the human cell machinery. The corresponding gene is synthesized by assembly of synthetic oligonucleotides. This process has another advantage, which is to generate a totally synthetic gene that can be very well documented and consequently easily introduced in a GMP manufacturing process (see Section 2.4).

***Noncoding Sequences.*** The nucleic acid vector can be optimized in the noncoding (untranslated) parts of the mRNA. Mainly, as mentioned in Section 1, the AURES destabilizing sequences should be deleted because they would destabilize the mRNA and limit the expression of the antigen. In general, only the coding sequence of the gene of interest should be used for constructing the therapeutic vector; any

sequence upstream of the ATG and downstream of the stop codon should be avoided. Meanwhile, stabilization thanks to 3′ UTR [26] in the form, for example, of the ca. 180 bases or ca. 80 bases UTR from the β- or α-globin genes, respectively, can be added after the coding sequence of interest (after the stop codon). Both of these sequences are recognized by the ubiquitous α-complex, which stabilizes the mRNA and enhances its translation rate. They are routinely used in the pDNA and mRNA vectors used for therapy [27–30].

***Specific Optimization of pDNA Vectors.*** The promoter used for the expression (transcription) of the transgene offers the possibility to control the level, site, and eventually time of expression of the transgenic protein. Concerning the level of expression, strong ubiquitous promoters helped by enhancers will be adequate when a high and broad expression of the transgene is needed. However, if an expression of the transgene is expected in specific cell types, investigators can use the large panel of restricted promoters that are active thanks to cell-type-specific transcription factors. This way, the expression of the transgene can be restricted to the liver (albumin promoter, for example), lymphocytes (immunoglobulin promoter or CD2 locus control region, for example), antigen presenting cells (fascin promoter for example), and so on. The drawback of these promoters is that they may lead to a relatively weak transcription activity compared with ubiquitous promoters. The control of time of expression is a critical parameter for all pDNA-based approaches because a long-term expression may have serious negative effects on the efficacy of the treatment (immune tolerance instead of immune activation, for example) or unaccept-able side effects (overexpression of growth factors or other proteins with physiological activity that are used in gene therapy). A way to finely tune the expression of a transgene is to use an inducible promoter such as the Tet-On or Tet-Off promoters [30a]. Their activity depends on the Tet-On or Tet-Off transcription factors that must be constitutively encoded by the therapeutic pDNA construct. The Tet-On and Tet-Off proteins should be expressed thanks to ubiquitous promoters. The presence of tetracyclin or its analog Doxicyclin, which can just be added in the drinking water, will activate (Tet-On) or deactivate (Tet-Off) the transcription capacity of the Tet protein. Accordingly, it will turn ON (Tet-On) or OFF (Tet-Off) the transcription of the transgene of interest. The drawback of these systems is that they are "leaky," thus always giving a basal expression of the transgene even in conditions where it should be silenced (in the presence of Doxicyclin for Tet-off and in its absence for Tet-On).

The other functional parts of the plasmid such as the transcription termination signal, the bacterial origin of replication, and the antibiotic used for production of the plasmid do not impact the efficacy of the pDNA as a therapeutic vector. However, the antibiotic-resistance gene expressed by the plasmid when it is in a bacteria (the gene is expressed thanks to bacteria-specific controlling sequences) is relevant for the regulatory aspects (see Section 2.4). Moreover, the overall sequence of the plasmid is active at the level of immune stimulation: The more the plasmid contains unmethylated CpG immnostimulating motifs, the more it may induce signaling through TLR9 in human immune cells, and consequently, the stronger may be the priming of the immune response. Thus, for vaccination approaches, pDNA vectors should be enriched in CpG motifs, whereas for gene therapy approaches, pDNA should be depleted of such sequences.

***Specific Optimization of mRNA Vectors.*** As mentioned, a 3′ globin (α or β) UTR at the end of the mRNA provides higher intracellular stability to the mRNA, which is of advantage whether the mRNA is the vaccine vector or a product of the vaccine vector (with pDNA vaccines as mentioned in the above paragraph). Aside from this, specific modifications of the mRNA vector are the poly-A tail and the 5′ Cap structure. Concerning the poly A tail, it should be of a minimum of 30 residues and preferably exceed 100 residues [31]. To generate the poly-A tail, there are two methods: Either it is encoded in the plasmid DNA that is used as a matrix for *in vitro* transcription or it is added by a specific polymerase (Poly(A) polymerase) at the end of the mRNA transcripts. In the first case, a A-tail can be of a maximum of ca. 120 As because more residues would titrate the adenosine triphosphate in the transcription reaction and limit its output. The other method is to submit the *in vitro* transcribed mRNA to the activity of the Poly(A) polymerase, which is a commercially available enzyme that polymerizes A residues at the 3′ end of a ribonucleic acid, in a matrix-independent manner. The enzyme can add several hundred A residues at the 3′ end of the mRNA molecules. This method is preferred to produce efficacious and stable mRNA for research purposes. However, the production of such poly-adenylated mRNA in GMP conditions is limited by the fact that the poly-adenylation process is not totally controlled and may generate heterogeneous molecules with more or less long A tails. This could make endproduct controlling, characterization of a batch, and batch-to-batch reproducibility more difficult than when the poly-A tail is encoded in the plasmid and is consequently clearly defined in length.

As mentioned in Section 1, at the 5′ end of mRNA molecules to be translated by eukaryotic cells, there is the characteristic Cap structure. It consists of a 7-methylated guanine residue linked through three phosphates to the first residue of the mRNA. A noncapped mRNA is not translated in eukaryotic cells. Not only is this structure essential for recognition of the mRNA by the translation machinery (the cap is recognized by the initiation factor 4: eIF4), but it also protects the 5′ end of the mRNA against exonucleases. Similarly to what was described for the poly-A tail, either the Cap structure is incorporated during the transcription process or it is added afterward on the *in vitro* transcribed mRNA.

In the first case, a fourfold excess of synthetic Cap analog (m7G(5′)ppp(5′)G) versus GTP in the transcription mixture guarantees that ca. 80% of the produced molecules will start with a Cap instead of a canonical G residue (the T7, T3, or SP6 promoters are made in a way that they will start transcription with a G residue). The drawback of this method is that the GTP content in the transcription reaction must be lowered compared with the other residues and thus becomes the limiting factor for the transcription. It significantly reduces the amount of recovered mRNA compared with an *in vitro* transcription reaction without Cap and with an identical concentration of all four residues. However, this method to produce capped mRNA is efficacious, fast, and reliable. Some modifications of the Cap analog can improve the efficacy of the mRNA vector: The standard synthetic Cap is a di-guanine molecule with one of its G methylated and three phosphates in-between the nucleotides. It can be incorporated at the start of the mRNA in two orientations: the methylated G being at the 5′ end or the non-methylted G at the 5′ end. Only the former situation will give a functional mRNA. Thus, statistically half of the capped mRNA molecules are nonfunctional. Strategies to avoid this phenomenon include

the anti-reverse Cap structure (ARCA) [32]. It consists in a Cap analog in which the sugar carrying the methylated guanine is modified in its 3′ position ($OCH_3$). Thus, the dinucleotide can be used as the start nucleotide by the RNA polymerase only in the functional orientation where the 3′OH of the non-methylated based is used to extend the mRNA. ARCA Cap enhances the functionality of *in vitro* transcribed mRNA vectors. However, ARCA Cap is relatively difficult to synthesize and consequently more costly than the standard Cap analog.

The second method for capping mRNA takes advantage of the activity of the vaccinia virus capping enzyme, also known as guanylyltransferase. This enzyme is commercially available. In the presence of GTP and S-adenosyl methionine (SAM), it can add a natural Cap structure (7-methylguanosine) to the 5′ triphosphate of a RNA molecule. As it is an enzymatic reaction, it can bring a correct Cap to all mRNA molecules, and thus, it is optimal compared with *in vitro* transcription in the presence of standard Cap but similar theoretically to *in vitro* transcription in the presence of ARCA Cap.

Because of the lack of a reliable quantitative assay to control for the presence of the Cap structure, aside from a functional assay of the mRNA expression (transfection of the mRNA in cells and detection of the translated protein), the enzymatic capping of mRNA is rarely used to produce mRNA for research or therapy. The utilization of the synthetic Cap (nonmodified analog or ARCA) in the *in vitro* transcription reaction is the standard method to produce capped mRNA.

Aside from the 5′ and 3′ ends, some internal modifications of the mRNA such as a phosphorothiate backbone or 2′ modifications of residues can be brought into the mRNA as long as the modified residues are substrate for the RNA polymerases routinely used for *in vitro* transcription (T7, T3, or SP6 RNA polymerases) and do not interfere with the translation process. It seems that only a few modifications such as phosphorothioate nucleotides and 2′ amino residues can be used [14]. However, they lower the efficacy of transcription and translation while they do not noticeably increase the stability of the mRNA toward RNases. Some more work may allow the identifications of modifications that may allow the efficacious production of mRNA, which resists extracellular RNases and remains well translatable by the ribosome. Such mRNA would be optimized compared with native mRNA for therapeutic uses.

## 2.2 Vaccination

***Introduction.*** Upon entry of a pathogen (virus, bacteria, or parasite) or upon pathologic genetic modifications of a cell of the body (tumors), pathogen associated molecular patterns (PAMPs) or "danger signals" such as dsRNA or stabilized ssRNA (virus), CpG DNA or flagelin (bacteria), uric acid, or heat shock proteins (tumors), for example, are recognized by immune receptors such as TLR receptors or scavenger receptors. These activating receptors are expressed by cells of the innate immunity such as macrophages and DCs that are professional antigen presenting cells (APCs). Some PAMPs also stimulate B cells, NK cells, or NKT cells. Professional APCs loaded with a pathogen's derived material and stimulated by the associated PAMP migrate from the site of detection of the danger signal to the draining secondary lymphoid organs such as lymph nodes and activate antigen-specific T and B lymphocytes through their clonotypic receptors: T-cell receptor

(TCR) for the former and surface immunoglobulin for the latter. There are approximately $5 \times 10^9$ T cells (ca. 70% are CD4 positive helper cells and ca. 30% are CD8 positive cytotoxic cells called CTLs) and $2 \times 10^9$ B cells in an adult human body. Generally, only a few hundred naïve lymphocytes specific for a precise antigen are available (frequency ca. $10^{-7}$) [33]. Upon an encounter with the antigen-loaded activated APC, specific lymphocytes are triggered and proliferate. A lymphocyte clone may expand so much that it reaches a frequency of $10^{-2}$ [34]. Activated lymphocytes become armed effector cells. Armed B lymphocytes secrete the clonotypic immunoglobulin that sprays through the body, recognize its antigen, and inactivate it or target it to destruction by effector mechanisms of the innate immunity. Armed CD4-positive T lymphocytes secrete cytokines (mainly interferon-γ for Th1 cells and Il-4 for Th2 cells) that are necessary for immune cells' activity and they stimulate APCs. Armed CD8-positive T lymphocytes migrate to the site of antigen expression and kill cells expressing it. Upon disappearance of the antigen (complete elimination of the pathogen), armed effector cells disappear (contraction of the immune response) and some memory cells remain for the individual's whole life. Not only does this memory cell allow a higher number of specific lymphocytes to be present in the body at an eventual second appearance of the antigen than at the time of the primary encounter, but also memory lymphocytes have the peculiarity to react very quickly and efficiently (higher affinity of the antibodies for example). Thus, in the case of reappearance of the antigen, the neutralization of the pathogen (infectious agent or tumor cells) will be much faster than at the time of first priming. Prophylactic vaccination mimics a first contact between the immune system and a pathogen. It guaranties that, should the pathogen be present in the body, its recognition and elimination will be efficient and fast, preventing acute or chronic (persistent) disease. Therapeutic vaccination or immunotherapy aims at boosting the more or less preexisting immune response (triggered naturally by the already present pathogen), sustaining its strength (stimulating the proliferation of specific lymphocytes) and broadness (stimulating a larger spectrum of lymphocytes specific for the pathogen) to enhance its efficacy. As shown for several pathogens such as for example, Hepatitis B Virus (HBV), Human Immunodeficiency Virus (HIV), or tumor, a strong immune response against the pathogen is associated with a better control of the disease. High morbidity and mortality is often due to a weak immune response against those pathogens. Thus, immunotherapy aims at enhancing the natural capacity of the immune system to help control the pathogens. Another type of vaccination is used to shift or tolerize a preexisting pathologic immune response: in allergies or autoimmune diseases. Those immunotherapeutic regimen allow the manipulation of a preexisting immune response to render it armless.

***Nucleic Acids for Vaccination.*** In all cases where the antigen of interest is a protein, nucleic acids can be used as a method to vaccinate. An adequate pDNA or mRNA vector can be produced *in vitro* (see above Section 2.1), formulated for the delivery (see Section 1.6) and eventually co-injected with an adjuvant in the form of a danger signal or cytokine. The local expression of the foreign nucleic acid-encoded protein associated with the local activation of innate immunity will trigger a T- and B-lymphocyte-mediated antigen-specific immune response that can protect against a challenge with the pathogen (infectious agent or tumor cell)

that expresses this protein. It can also modify an existing pathologic immune response (autoimmunity, allergy) to render it armless. Both pDNA- and mRNA-based vaccines were reported to be capable of inducing humoral and cellular immunity (review by Liu and Ulmer for pDNA [35] and Pascolo for mRNA [36]). The particularly advantageous feature of this vaccination technique is that because antigens are made by the cells of the body, their epitopes are presented on major histocompatibility complex (MHC) I molecules and a CTL response is generated. Other vaccine methods based on proteins or inactivated infectious agents do not usually efficiently induce CTL. These cells are particularly relevant for antivirus, anti-intracellular pathogen and antitumor immunity. Thus, nucleic-acid based vaccines are theoretically very attractive modalities in the context of these pathogens.

It is usually accepted, although not firmly demonstrated, that the stronger is the transgene's expression (the highest is the amount of produced antigen), the better is the triggering of the immune response. Thus, enhancing the injected nucleic acid's translation (see 2.1) is a common method for improving the efficacy of nucleic acid-based vaccination. Improving the efficacy of the processing and presentation of the coded antigen was also reported to strongly impact the vaccine's potential (see below). However, of foremost importance is probably the site of delivery, the formulation, and the used adjuvant, which have a very strong and documented impact on the efficacy of the pDNA- or mRNA-based vaccines.

*Optimization of the Antigen Processing and T-cell Priming.* When the antigen encoded by the nucleic acid is to be recognized by T lymphocytes, its derived peptide epitopes must be efficaciously presented by MHC class I or class II molecules. Although any cytosolic, membrane or secreted protein can be used to produce MHC I and MHC II associated epitopes, several methods to enhance this presentation were described.

For augmenting the presentation by MHC I molecules, a sequence coding for ubiquitin can be inserted at the beginning of the gene. The produced protein will be cleaved after the ubiquitin, thus exposing a non-methionine N-terminal residue [37]. Should this residue be a "destabilizing" one, the protein is quickly degraded by the proteasome and its derived epitope efficiently presented on MHC I proteins. Another method to enhance MHC I epitope presentation is to insert a sequence coding for Herpes Simplex Virus (HSV) VP22 protein or heat schock protein 60 or calreticulin at the beginning of the gene [38–40]. HSV VP22, HSP60, and calreticulin moieties have the particularity to enhance MHCI-antigen presentation of the conjugated protein.

For augmenting the presentation by MHC II molecules, the engineering of genes encoding Invariant chain (Ii) [41] or lysosome-associated membrane protein-1 (LAMP-1) [42] fused to the antigen is recommended. Both Ii and LAMP-1 can direct the foreign nucleic acid-encoded protein to the cellular compartments where the degradation of antigens and the loading of MHC II molecules take place.

Shipping the antigen to the extracelullar compartment and at the same time targeting it to APC is another method to enhance T-cell priming: Whatever cell expresses the foreign nucleic acid, the antigen will be released and captured by APCs. To this end, the antigen can be fused for example to CD40L, Flt-3L, or CTLA4 [43–45] because these molecules bind to APC-specific surface proteins.

These strategies enhanced the efficacy of pDNA-based vaccines but may have the theoretical drawback of triggering an immune response against the self-protein used as fusion. Such an immune response would lead to autoimmune disorders targeting the immune system.

Finally, for enhancing the priming of the T cell, especially when the immune system may be tolerant (self-proteins used in anticancer vaccination), the addition of a "foreign" sequence to the antigen such as for example the Pan DR Epitope (PADRE: AKFVAAWTLKAAA) [46] or the utilization of a xenogenic antigen (the sequence of the antigen in species other than human) that has a ca. 90% homology to the nominal antigen are shown to increase the vaccine potential of nucleic acids [47, 48].

*Specific Functional Optimization of pDNA Vectors for Vaccination.* For a good expression of the vaccine transgene, a strong promoter must be used. Investigators can chose between (1) ubiquitous promoters such as the commonly used promoter of the early genes of CMV, eventually enhanced using additional regulatory elements (Operators [49]); and (2) cell-type specific promoters such as those that can promote transcription only in APCs (promoter exclusively used by dermal dendritic cells such as the fascin gene [50], for example) or locus control regions [12]. In the first case, independently of the site of injection and of the cell type that take up the pDNA, there will be a local expression of the protein, and through direct presentation (when the plasmid was taken by APC) or cross-presentation (when the plasmid was taken up by a somatic cell and the produced protein phagocytosed by neighboring APCs), specific lymphocytes are activated. In the latter case, expression of the antigen is limited to APCs. This has several advantages: As APCs are terminally differentiated cells, they cannot persist or proliferate; thus, long-term uncontrolled expression of the transgene and risks of transformation (when the transgene is an oncogene to be used in antitumor vaccination) are excluded. The induction of immune tolerance due to long-term expression of the antigen by somatic cells is also excluded. Moreover, the delivery of vaccine plasmids that have an expression restricted to APCs was reported to enhance the efficacy of vaccination compared with ubiquitously expressed plasmids [13].

Should the plasmid contain no CpG motifs, the introduction of some of these sequences outside of the functional sequences (i.e., origin of replication, antibiotic resistance, and relevant gene expression cassette) would be needed to render the plasmid functional for vaccination purposes.

### *Delivery of Nonviral Coding Nucleic Acids for Vaccination*

*Injection of Naked Nucleic Acids.* Although *in vitro* cultured cells do not usually take up naked nucleic acids (pDNA or mRNA) spontaneously, skin cells and muscle cells do in vivo. As published in the milestone article by Wolff et al. [18] and presented in Figure 4, the direct injection of pDNA or mRNA resuspended in an isotonic buffer results in the local uptake of the nucleic acid that can be visualized by the expression of the encoded protein (luciferase, EGFP, CAT, or β-galactosidase are easily detectable markers) or by the induction of an immune response directed against the encoded protein. This is true whether the injection is made intramuscular, intradermal, or subcutaneous (review by Wolff and Budker

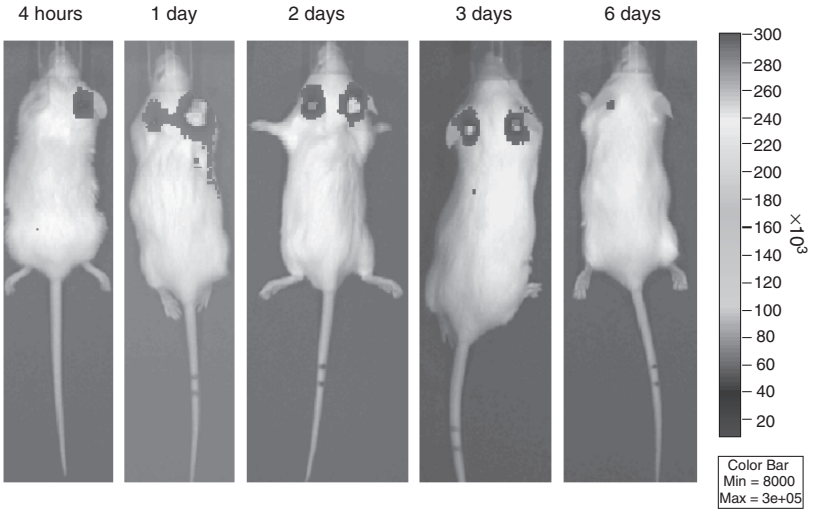


**Figure 4.** Expression of luciferase *in vivo* after injection of nonviral coding nucleic acids. Mice were injected in the ear pina with 10 μg of naked mRNA (right ear "R") or 10 μg of naked pDNA (left ear "D") coding luciferase. At the indicated time points, mice received a systemic injection of luciferin and were monitored by an *in vivo* imaging system for luciferase-mediated light emission. The injection of mRNA quickly results in strong luciferase expression at the site of application and totally vanishes after a few days. Plasmid DNA gives an expression at the site of injection that peaks later than the one obtained with mRNA injections, and that remains for a long time (up to weeks, depending on the injections).

[50]). To get an immune response, standard injections in mice consist of 50 to 200 μg of nucleic acids. A great optimization is obtained by direct intra-lymph-node injections, which proved in preclinical models to be much more efficacious than other routes, i.e., giving similar immunity than a 100-fold higher dose injected intramuscularly or intradermally [51]. Such a method was evaluated in clinical trials and found to be a feasible and safe approach [52].

The mechanisms and molecules necessary for the spontaneous uptake of exogenous nucleic acids are not fully characterized. However, it is probable that both pDNA and mRNA are phagocytosed, eventually in a receptor-mediated way (the reporter gene expression, after pDNA or mRNA injection, is a saturating process that can be competed with irrelevant nucleic acids), and travel through endosomes before they can reach, in an unknown way, the cytosole and, for pDNA, subsquently the nucleus (Figure 2). As pDNA and mRNA vectors contain intrinsic danger signals (nonmethylated CpG sequences recognized by TLR9 for the former and any nonmodified ribonucleic acids recognized by TLR7 and TLR8 for the latter) that are detected in the endosomes of cells that have phagocytosed the exogenous nucleic acid (Figure 3), the expression of the encoded protein parallels a local activation of the immune system. This results in the development of a B-cell- and T-cell-mediated immune response specific for the protein encoded by the injected nucleic acid. After the antigen is not expressed any longer, the triggered immune response becomes a memory response that provides an advantage to the treated

individual, should he be re-exposed to the nominal antigen expressed by a pathogen.

An *in vivo* electroporation method was developed and tested in order to enhance the penetration of injected naked nuclei acids in cells at the site of delivery (for review, see Prud'homme et al. [53]). After the injection of the nucleic acid, a short electric pulse is given through the injected tissue thanks to electrodes connected to a programmed generator. This method allows a strong enhancement of the uptake of the injected pDNA. So far, however, it has not been successful in enhancing the uptake of injected mRNA (unpublished personal observation). For pDNA vaccination, *in vivo* electroporation is becoming a standard method to enhance the efficacy of vaccination in mice and it is being tested in a phase I trial in humans (see www.clinicaltrials.com).

*Delivery of Vaccine Nucleic Acids Using Particles or Cells.* One problem when using injections of naked nucleic acids for vaccination is that a big amount of vector is needed to vaccinate mice and *a fortiori* humans. This is partly due to the quick degradation of the extracellular vector (half-life within seconds for the mRNA [14] and eventually minutes for pDNA [54]) and its relatively poor uptake by cells. Only a few thousand plasmids are expressed at the site of injection [55]. For vaccination, three strategies were developed that can protect the nucleic acid and/or deliver it efficiently to its functional site (cytosol for mRNA and nucleus for pDNA): Gene Gun, *in vitro* transfection of professional APCs before adoptive transfer, and finally encapsulation in particles or liposomes.

GENE-GUN. The gene-gun technology is based on a ballistic delivery. The nucleic acid is precipitated on micrometric gold particles that are resuspended in a buffer and quickly injected into an horizontal tubing of 4 mm in diameter and 75 cm in length. The particles are left there to sediment. Thereafter, the buffer is slowly discarded and the tube is turned 180° along its axis in a way that the initial bead's bed is on the "roof" and disperses slowly through gravity on the inside of the tubing, covering it homogeneously. After it is dried in the inside, the tube is chopped into the size of cartridge (ca. 1.25-cm-long tubes). The cartridges are placed in the cartridge holder of a gun that is connected to pressured helium. Upon firing, the gas goes through the cartridge and propels the beads with high speed. Should the gun be a few centimeters from the skin, the beads penetrate the *stratum corneum* and stop in the dermis. They release the nucleic acid in the hit cells or cells' nuclei. Within these cells are dermis DCs and Langerhans cells. They are activated probably through signaling by TLR9 (pDNA vaccines [56]) or TLR7 and TLR8 (mRNA vaccines [57]) and migrate to the draining lymph node where they can activate antigen-specific lymphocytes.

ADOPTIVE TRANSFER OF *IN VITRO* TRANSFECTED APCS. As the key event in vaccination is the presentation of the antigen and its derived epitopes by professional APCs, researchers investigated the possibility of generating a pure APC population out of blood cells and of modifying these cells with exogenous nucleic acids before re-implanting them in the body. Because the transfection of pDNA in APCs such as DCs is inefficient and, when successful, induces apoptosis, the methods of genetic modification of APCs were principally developed using mRNA vectors. As APC,

the monocyte-derived dendritic cells are most often used. They are easily produced, even in GMP conditions, starting from blood monocytes that are extracted from total blood either using their characteristic to stick quickly on plastic surfaces or by sorting (magnetic beads or FACS) with monoclonal antibodies specific for monocytes such as CD14. The monocytes are cultivated 1 week with GM–CSF and interleukine-4 to generate immature DCs. These cells can be easily transfected with mRNA using simple co-incubation or, for a more efficient, more controlled uptake, electroporation. Eventually, the cells are rendered mature by the addition of one or several danger signals and further cultivated 1 or 2 days. Then the cells are re-injected into the patient. Although the method was shown to work in animals and in humans (see Section 2.1), the optimal danger signal(s) and site of injection are still to be precisely defined. For the maturation, a cocktail containing IL-1β, IL-6, TNF-α, and PGE2 may be optimal to generate mature DCs that can migrate to the lymph nodes draining the site of adoptive transfer without dying too early from overstimulation [58]. Using this type of cocktail, DCs injected subcutaneous or intradermally are expected to go to the secondary lymphoid organs and prime antigen-specific lymphocytes. Alternatively, the direct intra-lymph-node injection of mature transfected DCs may be the best method to guarantee that most *in vitro* generated APCs are at their site of activity after adoptive re-implantation. The utilization of this method in many human clinical trials is ongoing throughout the world, and optimized protocols will soon be deducted from these studies.

PARTICLES. Both pDNA and mRNA vectors entrapped in particles or coated on particles were shown to be potent vaccines. In all cases, the particle strategy allowed a drastic reduction of the amount of the nucleic acid needed to get a significant immunity. For mRNA vectors, Martinon et al. [59] were the first to report that influenza nucleoprotein coding mRNA entrapped in liposomes can induce an immune response against influenza virus in mice. It was then shown that minimal amounts of mRNA (down to 1 μg of encapsulated mRNA injected intradermal or intravenous) are enough to prime an antigen-specific immune response [60]. This technology has not yet been transferred to human clinical trial. Encapsulated pDNA, however, were shown earlier to be potent vaccine formulations. Plasmid DNA entrapped [61] or coated [20] on cationic microparticles or enclosed in liposomes [62] are more potent than naked pDNA as vaccines. Using either intramuscular or intravenous routes of delivery, doses as low as 1 μg of encapsulated pDNA were shown to prime humoral and cellular immune responses in mice. Using these formulations, oral [61, 63, 64], intranasal [65], or other mucosal delivery routes (intrarectal [66], for example) can be used to stimulate mucosal and systemic immunity in animals (mice or macaques).

***Adjuvants.*** Although nucleic acids have the intrinsic capacity to signal "danger" through TLR7, 8, or 9, extra-immunostimulation can enhance or polarize the immune response. It can be achieved by injection of cytokines or nucleic-acid-coding cytokines together with the vaccine. The cytokine GM–CSF is one of the most studied adjuvants in the context of nucleic acid vaccinations. It was shown to enhance and to polarize toward Th1 immune responses after pDNA [67] or mRNA [68] applications. The optimal timing of application of such an adjuvant is not clearly established. Theoretically, the local delivery of GM–CSF should attract

immature DCs that would render the site more reactive to a consecutive nucleic acid injection. However, practically, in mice it looks like the injection of GM–CSF should at best follow the injection of pDNA or mRNA. The optimal timing of GM–CSF injections in humans would need to be deducted from clinical trials especially designed to decipher this matter. Inflammatory cytokines such as IL-12 are also foreseen as promising adjuvants to be used in combination with nucleic acid vaccines.

***Preclinical Results and Clinical Applications.*** Both pDNA- and mRNA-based vaccinations were demonstrated to be efficacious in animal models as prophylactic or therapeutic immunotherapies against tumors, infectious diseases, and allergy. Two pDNA-based vaccines are commercialized for veterinary use: an anti-equine fever and an anti-infectious hematopoietic necrosis virus (IHNV) for farm-raised salmons. In humans, several formulations of nucleic acid vaccines are tested in clinical trials (see the actualized list of trials at www.clinicaltrials.gov). Although pDNA-based vaccine trials were reported in the context of antitumor, antivirus (HIV, influenza virus, HBV) and antiparasite (*Plasmodium falciparum*) approaches, mRNA-based vaccines were up to now tested only as immunotherapies against cancer (review by Liu and Ulmer for pDNA [35] and Pascolo for mRNA [36]).

*Anti-infectious Diseases Vaccines (HIV, Plasmodium falciparum, influenza virus, HBV, Mycobacterium tuberculosis).* Theoretically, any infectious agent can be recognized by the immune system, and thus, vaccination is of interest to prevent acute infection or reduce chronic replication. Preclinical results have shown at least in mice that for all economically or socially relevant pathogens, pDNA-based vaccination is a robust method to prevent or treat the infections. Several trials in the human population are ongoing (see www.clinicaltrials.com) and address the safety of pDNA vaccines designed to trigger immunity against HIV, *Plasmodium falciparum*, influenza virus, HBV, *Mycobacterium tuberculosis*, severe acute respiratory syndrome (SARS) virus, West Nile Virus, or Ebola virus. Most are phase I trials and thus address only the safety and dose. The following paragraphs describe the results obtained for several of these test vaccines.

AIDS. The Acquired Immunodeficiency Syndrome (AIDS) is the fourth cause of death in the world (an estimated 3.5 million people die from AIDS every year). An estimated 40.3 million persons are living with HIV, and 5 million people per year become infected. An affordable and efficacious anti-AIDS vaccine is urgently needed (review by Girard et al. [69]). Due to the potency of pDNA vaccines to induce CTL in mice and the known efficacy of these cells to control HIV replication in humans, the first trials using direct injection of nucleic acids were performed in HIV-infected individuals in the mid-1990s. Follow-up trials were performed in uninfected volunteers. These studies showed that anti-HIV pDNA vaccines are feasible and safe [70]. The induced immune responses were less strong than expected based on preclinical studies. However, CTL responses could be recorded after injection of naked pDNA [70]. Interestingly, in infected patients, the pDNA vaccines could induce new anti-HIV immune specificities, not triggered by the infection itself. Thus, in the context of immunotherapy, pDNA vaccination has indeed the potential to boost the preexisting immunity but also to broaden it by

inducing new virus-specific CTL and antibodies. Up to now, no clinical efficacy of pDNA vaccines against HIV could be demonstrated. However, strategies using pDNA as a prime followed by recombinant viruses or proteins as boost are being evaluated as vaccine protocols in high-risk populations [71, 72].

MALARIA. *Plasmodium falciparum* (*P. falciparum*) is the most pathogenic parasite causative of malaria. Yearly, 300–400 millions clinical cases of *P. falciparum* infections are reported, among which 2–3 million are deadly. Moreover, an increase of the incidence of the disease has been observed due to the emergence of drug-resistant parasites and of insecticide-resistant mosquitos. For these reasons, a protective and cost-effective vaccine against *P. falciparum* is needed. *P. falciparum*, being an intracellular parasite, its control requests the recognition by CTL. Thus again, nucleic acid-based immunotherapies are theoretically adapted for the development of a preventive or therapeutic vaccine against malaria. The same year as the report of anti-HIV pDNA-based vaccines was published so was the result of the first anti-malaria pDNA-based vaccine [73]. In this and follow-up studies, the induction by the pDNA vaccine of CTL against *P. falciparum* epitopes was documented. However, no antibody response was induced. The codon usage of *P. falciparum* is very inadequate for mammal cells. Thus, the pDNA vaccines based on native *P. falciparum* genes may have been relatively inefficiently translated *in vivo*. This may account for the generation of CTL, which are recognizing fragments of the protein but fail to generate antibodies that require the production of the full-length foreign protein. To conclude, although the documented induction of anti-malaria T cells by pDNA applications [74, 75] is an encouraging progress toward the development of an antimalaria vaccine, a randomized placebo-controlled phase III trial is required to see whether pDNA-induced anti-*P. falciparum* immunity results in the reduction of the burden of infection, morbidity, and mortality within populations living in endemic areas.

HEPATITIS B. The hepatitis B virus (HBV), which may have infected more than one third of the population, is responsible for approximately 1 to 2 million deaths every year. After infection, most people clear the virus through an efficient immune response. However, the virus persists in ca. 10% of the people and eventually leads to liver diseases such as hepatocellular cancer. Although the anti-HBV vaccine based on the recombinant surface antigen is efficient and safe, 5% to 10% of healthy immunocompetent subjects do not respond to it [75]. For those people and the ca. 350 million carriers, an alternative vaccine that would trigger a B-cell but also a T-cell response is of great interest. Indeed, a potent T-cell response of the Th1 type is believed to lead to the control of HBV. Several prophylactic and therapeutic trials are ongoing. One of the most successful methods to trigger immunity against HBV using pDNA is based on the utilization of the gene-gun: A trial showed that ballistic delivery of pDNA coding HBS could induce CD4 and CD8 T cells as well as protective levels of antibodies in all treated hepatitis-naïve volunteers [76]. Moreover, no or low responders to the standard vaccine were found to respond to the pDNA applied with the gene-gun (12 out of 16 patients responded [77]). Each dose of the vaccine contained no more than 4 μg of plasmid. Subjects received between one and three doses, with a two-month interval in-between applications. Thus pDNA-based vaccination is a safe, feasible, cost-effective, and pro-

mising prophylactic and therapeutic approach against HBV. The results of phase II and III trials are needed to prove clinical efficacy and allow commercialization of this vaccine.

TUBERCULOSIS. *Mycobacterium tuberculosis* is the etiologic agent of tuberculosis. The disease is due to the immune response against the pathogen more than to the pathogen itself. Immunotherapy using the whole inactivated pathogen or BCG may enhance rather than decrease the disease. Nucleic acid vaccination offers the possibility to perform immunotherapy with the immunogenic parts of the pathogen that are associated with a protective immune response rather than with a pathologic immune response. Heat Shock Protein (HSP) 65 is a dominant mycobacterial antigen. Vaccination with HSP65-encoding pDNA in mice was superior to infection in priming protective T cells [78]. In this model, the same method was found to be efficacious in a therapeutic setting [79]. However, these results were not confirmed by another laboratory, where vaccination of mice with pDNA encoding mycobacterium HSP65 enhanced the pathology. These contradictive results may be due to the ratio of Th1 versus Th2 cells primed by the vaccine, which may be linked with the purity of the nucleic acid (the presence of endotoxin, which orientate the response toward Th2), the site of injection (intradermal delivery orientates the response toward Th2), or the level of immunological naivety of mice that depends on the clean conditions in which they are kept. As an improvement of the approach, a combination between efficacious anti-mycobacterium chemotherapy and pDNA vaccination may allow a sustained immune control of the pathogen [80]. This regimen may be used in infected patients suffering from tuberculosis. However, the risk of increasing rather than decreasing the pathologies associated with *Mycobacterium tuberculosis* using pDNA vaccination is a drawback that limits the testing of this immunotherapy.

*Antitumor Vaccines.* Many, if not all, tumor cells express characterized specific antigens (self-proteins or virus-encoded proteins) that make them recognizable and controllable by the immune system, especially by T cells. The capacity of immune-stimulation approaches, among them, nucleic acid-based vaccines, to control tumors was clearly proven in many animal models. For this reason, nucleic acid vaccines coding tumor antigens such as virus-encoded oncogenes (for example, HPV 16 E7), self-proteins (prostate antigens for prostate cancer, clonotypic antibodies for B-cell leukemia, melanocyte-specific proteins for melanoma, or shared tumor antigens such as testis-antigens for many tumor types, for example) or mutated oncogenes (p53 or cRas, for example) were tested as immunotherapy in cancer patients [81]. Prophylactic vaccination against cancer is validated in mice [82] but not yet transferred to healthy, tumor-prone individuals identified thanks to family history and detection through genotyping using known tumor-predisposition markers.

PLASMID DNA-BASED ANTI-TUMOR VACCINES. One of the first reports of an antitumor pDNA-vaccine phase I trial used a dual expression plasmid coding HBS and carcino embryonic antigen (CEA), a protein overexpressed on many epithelial cancers. The plasmid was repetitively injected intramuscularly. Per injection, up to 2 mg was applied. Some patients developed a good immune response toward HBS, but there was only a very limited response toward CEA. No objective clinical

response was recorded. Similar results were obtained in another phase I trial using a pDNA coding gp100 (an enzyme involved in melanine synthesis and expressed in most melanoma [83]) whether the DNA was injected intramuscularly or intradermally. As a superior vaccination strategy that may allow breaking tolerance, the intranodal delivery of pDNA coding tyrosinase (like gp100, it is a protein expressed in melanocytes and melanoma) was tested in melanoma patients [52]. Patients received up to 800 μg of pDNA every 2 weeks for four cycles. The treatment was feasible and safe. Nearly half of the patients (11 of 26) showed the triggering of an immune response against tyrosinase. Overall survival in this small cohort was found to be higher than expected. Meanwhile the utilization of an adjuvant in the form of recombinant cytokines (IL-2 and GM-CSF) applied subctaneously was tested in a phase I trial involving prostate tumor patients [84]. The pDNA (up to 900 μg per injection, five cycles) was given intramuscularly and intradermally. Cellular and humoral immune responses targeted to PSA could be detected in two of three patients injected with the highest dose. Biochemical regression was recorded in these two immunologically responding patients. Finally, as an immunotherapy strategy targeting the idiotype of B-cell lymphoma, pDNA expressing patient-specific clonotype were delivered with a needleless device (Biojector). Patients received monthly injections of up to 1800 μg of pDNA [85]. Here again, an immune response (B or T cell) could be triggered against the tumor-specific antigen (the antibody's idiotype) in most patients (7 out of 12).

Viral oncogenes are a privileged target for antitumor immunotherapies: Being foreign molecules, they are usually efficaciously recognized by the immune system and can be used as prophylactic antitumor vaccination targets. Human Papilloma Virus-16 (HPV-16) encodes two oncogens: E6 and E7. PLG-encapsulated pDNA coding for a few HLA-A2 epitopes of HPV-16 E7 administered intramuscularly could induce an immune response against HPV-16 in most treated patients suffering from HPV-16 associated anal dysplasia [86] or cervical intraepithelial neoplasia [87]. Clinical responses (histological responses) were observed in some patients (3 out of 12 in the study by Klencke et al.; 5 out of 15 in the study by Sheets et al.). Thus, pDNA-based vaccinations against HPV can be envisioned as a possible prophylactic and therapeutic antitumor immunotherapy regimen. This technology may become a standard anti-HPV therapy in the relatively near future.

MESSENGER RNA-BASED ANTITUMOR VACCINES. The only published results where mRNA vectors were used to vaccinate cancer patients are based on the utilization of mRNA-transfected DCs as described above. This popular method first described in 1996 [88] was validated in several preclinical mouse models and used to vaccinate prostate or renal cancer patients with different mRNA preparations (review by Gilboa [89]). In most patients, an increased T-cell immune response against the antigen(s) encoded by the mRNA(s) could be observed. Thus, mRNA-based vaccines using *in vitro* transfection of autologous DCs is a feasible and efficacious approach to stimulate a T-cell immune response against tumor antigen. Phase II and III trials are needed to show the clinical efficacy of mRNA-transfected DCs as anticancer immunotherapy. However, the need for a large amount of blood to prepare a few vaccine doses, the costly and variegating production of DC in GMP-conditions, as well as the logistic issues restrict the utilization of this technology. As an alternative, direct injections of naked or protamine-encapsulated mRNA

was shown to be efficacious for priming cellular and humoral immune responses in mice [27, 90]. It is being evaluated in human as anticancer immunotherapy (ongoing phase I/II studies).

CONCLUSION ON ANTITUMOR VACCINES BASED ON PDNA OR MRNA. Most published results show limited efficacy of nucleic acid-based vaccines as anticancer immunotherapies, highlighting the need for improving these strategies. However, the possibility of triggering an immune response against self-tumor antigens with pDNA or mRNA vaccines sustains the hope that once the correct antigens and optimized immunization methods will be identified, antitumor vaccinations would be a way to get regression or at least stabilization of tumor diseases in many patients. The fundamental knowledge of immune-tolerance and control of autoimmunity will help to find the protocols that will allow nucleic-acid based vaccines to become potent antitumor therapies. Of particular and new interest is the role of regulatory T cells (Treg) that can suppress immunity and are usually increased in number in the blood of tumor patients. Antitumor vaccination may enhance the activity of suppressive cells rather than the one of effector antitumor cells, thus leading to a failure of the therapeutic approach [91]. Controlling, i.e., reducing, the activity of Treg and of other immunomodulatory cells or molecules may be crucial for the success of anticancer immunotherapies. Methods such as chemotherapies, which are lymphoablative or specific inhibition of gene expression using siRNA (anti-TGF-β or IL-10, for example), may allow a transient reduction of natural immunosuppression that could be used for intensive vaccination protocols. Such combinations are being evaluated in clinical trials.

*Anti-allergy Vaccines.* Allergic diseases affect approximately one third of the population, and for unclear reasons, prevalence of these diseases is increasing. Through the triggering of Th2 type of T helper cells, environmental antigens induce the release of Il-4, Il-5, and Il-13 activate basophil, eosinophil, mast cells, and IgE, which eventually induce the development of inflammatory allergic diseases such as asthma, eczema, hay fever, and rhinitis. The desensitization of a patient can be obtained through repeated injections of increasing doses (starting at very low doses) of the nominal allergen over the course of several years. However, this protocol has been reported to have inconsistent efficacy and in rare cases very strong toxicity [92]. As vaccination with pDNA was reported to induce a Th1 response and inhibit the Th2 response, it was anticipated that pDNA vaccines could have a potential for the prevention and cure of allergic diseases. Indeed, it was demonstrated in animal models that the injection of pDNA coding for an allergen (β-galactsidase in experimental models or relevant allergen such as house dust mite Derp5, latex Hevb5, peanut Arah2, honeybee venom phospholipase A2, pollen allergen Cryj1, or Betv) can be a prophylactic or therapeutic treatment against allergies [93]. However, a clear control of the Th2 response, eventually using the co-injection of the vaccine with nucleic acids coding Il-12 or IL-15 as enhancers of the Th1 response, is needed before pDNA desensitization protocols can be planned for testing in the human population.

***Conclusion on Vaccination Using Nonviral Coding Nucleic Acids.*** Both pDNA and mRNA vectors can induce antigen-specific humoral and cellular immunity of

the Th1 and/or Th2 types. When using direct intramuscular injections of naked plasmid, a minimum of 800 μg of pDNA seems to be necessary to get a detectable immune response. Optimization of the vectors as described in Section 2.1 is a standard way to increase the efficacy of the nucleic acid-based vaccines and should be used for any clinical trials. The choice of the relevant pathogen-derived antigen, especially in the case of antitumor immunotherapies, is critical: This antigen should be well immunogenic and recognized by T and/or antibodies on pathogenic cells (infected and malignant) or by antibodies on infectious agents. The delivery method should be cost-efficient, praticable, and result in the induction of a strong immune response. Although the efficient intra-lymph-node delivery method may be adequate for therapy of a limited number of patients with advanced nontreatable diseases (tumor patients, for example), the gene-gun delivery sounds like the most favorable delivery method for easy, painless, efficacious, and cost-effective nucleic-acid-based vaccination of large populations. The utilization of a nontoxic adjuvant is always relevant, and at the moment, GM–CSF, which is the most frequently used cytokine in the context of nucleic acid vaccination trials, looks like a fairly good candidate. Meanwhile the control of Tregs is fundamental to trigger efficient immunity, especially in anticancer immunotherapies. The choice of the vehicle (pDNA or mRNA) needs to be addressed in clinical trials where both vectors are compared. Although mRNA looks safer due to its transient nature, pDNA may be more efficacious and require less injections for reaching protective immunity (high antibody titers and high T-cell frequencies). On another hand, long-term persistence of antigen expression when using pDNA may interfere with the development of memory T cells [94] and thus render a prophylactic vaccine inefficient. The induction of memory cells by mRNA- and pDNA-based vaccines needs to be evaluated and compared in human populations, especially when prophylactic vaccination is envisioned. Thus, it could be stated that the prevention of diseases or the treatment of chronic non-life-threatening diseases (i.e, allergy) may be most adequately addressed by mRNA vectors, whereas the treatment of late-stage patients (AIDS or tumor patients, for example) may be at the moment best addressed with pDNA vectors.

## 2.3 Gene Therapy

***Introduction.*** Known mutated genes that are the cause of recessive genetic diseases are, for example, adenosine deaminase in severe combined immunodeficiency disease (SCID), cystic fibrosis transmembrane conductance regulator gene (CFTR) in cystic fibrosis, hypoxanthine-guanine phosphoribosyl transferase in Lesch-Nyhan disease, glucocerebrosidase enzyme in Gacher disease, dystrophin in Duchenne muscular dystrophy, and factors XIII and XI for some hemophilia. In these cases, using foreign recombinant nucleic acids, it is possible to bring the protein that is missing or deficient. This technology is called gene complementation. The challenge of nucleic-acid-based gene therapies is to bring in all defined target cells an amount of nucleic acid that would provide a functional quantity of the encoded protein without over- or under-expression by some cells and without stimulating immunity or inducing apoptosis. General optimizations of gene expression as described for vaccination with pDNA and mRNA vectors in Section 2.1 should be performed to obtain the highest possible expression, thus efficacy, from the nuclei-acid-based

drugs. At the same time, the vectors should be designed to have low immunogenicity, for example, by eliminating CpG sequences in pDNA vehicles. Using pDNA vectors, the utilization of promoters that specifically drive gene expression in the adequate tissue or cell type is a method to lower side effects such as expression of the exogenous vector in irrelevant organs or cells. Although delivery of naked nucleic acid is evidenced *in vivo* and can be used for some approaches, the efficacious systemic delivery of nucleic acids needed for gene therapy is best obtained thanks to encapsulation [19]. Both methods are presented below. None of them has been used in human clinical trials. Up to now, gene therapy approaches tested in humans are based on virus-derived constructs that can efficiently penetrate cells and integrate in the nucleus. However, because of safety issues, pDNA and eventually mRNA (not used in preclinical models up to now) may appear again as favorable vehicles for gene complementation.

***Naked pDNA for Gene Therapy.*** As published by Wolff et al. in 1990, the injection of naked pDNA in the mouse skeletal muscle results in local transgene expression [18]. Since then, other species such as rats, cats, or monkeys and other sites such as skin, liver, brain, urological organs (assisted by *in vivo* electroporation), thyroid, and tumors were shown to be permissive to the local uptake of injected naked pDNA [95]. However, the local transgene expression obtained by direct injection of naked pDNA into an organ is usually too low to be used in the context of genetic disease, even when the missing function is contained in a soluble protein secreted in the bloodstream (for example, coagulation factors in hemophilia). It was estimated that after injection of naked pDNA, only a few thousand plasmids are functionally retained within cells of the injected tissue [56]. To treat a genetic disease, a physiological expression of the therapeutic molecule by all cells that need the missing function or a functional level of the protein in the body fluids is required. For this reason, gene therapy approaches based on naked pDNA use a systemic delivery method such as intravenous or, at best when organs other than lungs are targeted, intra-arterial injections. Indeed, the intravascular injection of naked pDNA surprisingly results in the expression of the encoded protein, especially in the liver [50]. As it is the case for intramuscular and intraskin injections, the intravascular injection of naked pDNA was initially a (negative) control of experiments designed to test the potency of transfection reagents. To reach a good expression level in mouse hepatocytes, the pDNA should be injected in a relatively large volume and short delivery time ("hydrodynamic" delivery) through the tail vein, the portal vein, the hepatic vein, or the bile duct. One injection can transfect about 1% of the hepatocytes throughout the entire mouse liver. This method can be used to produce *in situ* liver-specific proteins and soluble proteins such as coagulation factors XIII and XI that are missing in hemophilic patients. The production of critical proteins such as growth hormones with this technology is theoretically possible but would be hazardous because the amount of protein produced and the duration of the expression are not controlled.

Although naked siRNA can be delivered using the same procedure, naked mRNA would not be suitable because it is degraded within seconds in the serum [14].

Using hydrodynamic injection in a muscle's artery or vein, the expression of naked pDNA can also be obtained in the muscle. However, for a high expression,

the blood flow through the targeted vessel must be stopped surgically during the injection [96, 97]. Expression of the naked pDNA can also be obtained in the cardiac myocytes after injection in the left ventricular wall.

These preclinical observations made in mice, rats, and monkeys allow us to foresee the utilization of naked pDNA as a treatment of muscle diseases such as Duchenne muscular dystrophy, peripheral limb ischemia, and cardiac ischemia. However, the methods described above for animals (hydrodynamic injection, transient stop of blood flow through a muscle) are difficult to transpose to the human situation. Instead of an hydrodynamic delivery, human may benefit from the utilization of a vasodilator before pDNA injection. This treatment could increase the permeability of discontinuous endothelium such as those in liver, spleen, and bone marrow and induce permeability through continuous endothelium such as those in the brain. This way, plasmid DNA may go through the vascular wall and reach the target cells of the organs. As mentioned, the utilization of pDNA for gene therapy in humans has not yet been evaluated in clinical trials. Virus-based gene complementation is still observed as the most feasible method although safety issues are raised. Encapsulated pDNA is possibly an improved version of naked pDNA for gene therapy and has better safety features than virus-based delivery systems.

***Encapsulated pDNA for Gene Therapy.*** Both cationic liposomes and polymeric particles were used to deliver efficiently pDNA in animal models for gene therapy approaches. Although in these preclinical assays the expression of the transgene vanished after few days, the uncontrolled long-term expression as well as a possible overexpression after repeated delivery of encapsulated pDNA are safety issues that must be addressed before these methods are evaluated in clinical trials. However, thanks to the recent knowledge on gene regulation (utilization of inducible promoters such as Tet On and Tet Off promoters), immune activation (CpG motifs), immune modulation (Tregs for immunosuppression, for example), we can anticipate new generations of safe (controllable and nonimmunogenic) encapsulated plasmid constructs that will be the basis of future drugs designed to treat genetic diseases.

*Preclinical Results Obtained with Liposome Formulations.* One of the most frequent genetic diseases with high morbidity and mortality for which the deficient gene has been identified is cystic fibrosis. It is due to mutations in the CFTR, which is a transporter of chloride anions. This deficiency is recessive: Bringing the correct gene in the cells can correct it. As cystic fibrosis affects the lung, local gene therapy approaches where the genetic vehicle is delivered through airways is foreseen as an efficacious and safe method. As early as 1992, liposomes encapsulating pDNA coding for CFTR were found to be capable of bringing this protein function to airway epithelium of the lung after intratrachea instillation in mice [98]. Optimization of the delivery can be obtained by modifying the liposome formulation [99]. This promising method was not evaluated in patients.

Meanwhile, the intravenous delivery of lipid-entrapped pDNA was shown to efficaciously lead to gene expression, especially in liver, lung, and kidneys [100]. The addition of targeting molecules can efficiently target the vesicles to specific cells. For example, the utilization of galactosyl-cholesterol or mannosylated-

cholesterol in the liposome formulation allows efficient delivery of the pDNA to the liver in mice [101, 102].

To conclude, cationic liposome formulations were shown in animal models to be efficient, safe, and versatile delivery vehicles for pDNA molecules. The engineering of liposomes coated with targeting molecules is one method to enhance the efficacy and specificity of these vehicles. Clinical trials in humans are needed to demonstrate the feasibility, safety, and efficacy of these approaches.

*Preclinical Results Obtained with Particles.* Poly-L-lysine and poly-L-Ornithine were used to generate particles that can entrap pDNA. The cationic polymers can be coupled to a targeting molecule such as galactose or mannose for enhanced delivery to the liver after intravenous or intraportal injections in mice [102–104]. Expression in other organs than the liver (for example, kidneys) is observed; however, expression in the liver using these particles is dominant. Such particles are found to be more efficient and to persist for a longer time in the circulation than liposomes. Hepatic expression of the pDNA delivered by poly-L-Lysine particles can be found several months after delivery [105]. These promising formulations were not evaluated in humans.

***Conclusion on Gene Therapy Using Nonviral Coding Nucleic Acids.*** As opposed to vaccination strategies, the local delivery of naked pDNA in an organ is not foreseen as a possible gene therapy approach because the protein expression is too low. However, the systemic delivery through the blood circulation of naked or entrapped pDNA was found to result in strong expression of the transgene, mainly in the liver. The utilization of hepatocyte-specific promoters in the pDNA would allow us to turn these methods into liver-specific expression systems. They may be used to produce soluble proteins such as coagulation factors XIII and XI or erythropoietin in patients suffering from genetic deficiencies for these molecules. Encapsulated pDNA expressed by hepatocytes could also theoretically be used for the sustained production of therapeutic monoclonal antibodies (the anti-Her-2/neu antibody: Herceptin, for example) by the body's own cells in tumor patients. Aside from that, some methods (hydrodynamic delivery, occlusion of veins, particles) can allow delivery of pDNA in the muscle as a possible treatment for myopathies. Finally, inoculation through the airways can result in gene expression in the lung and could be used as a treatment against cystic fibrosis. Thus, although they are not yet tested in humans, the gene therapy approaches based on simple pDNA coding for a therapeutic protein are very exciting and promising approaches that offer a broad range of applications. However, several issues have to be addressed:

1. The natural immunogenicity of the pDNA. Even without any known CpG sequence in the plasmid, repeated applications may still induce the activation of the immune system. This immune response could neutralize the produced transgenic protein and result in the specific killing of cells expressing it. As a remedy, the utilization of immunosuppressive drugs concomitantly with the application of pDNA-based gene therapies should probably be envisioned.
2. The uncontrolled persistence and expression of the transgene prevents its utilization for the production of active molecules such as growth factors. Repeated applications may result in some patients in the accumulation of

pDNA and overexpression of the transgene. This would have negative effects on the producing organs (necrosis) and pathogenic systemic effect due to the protein's own activity. The utilization of inducible promoters and a careful monitoring of the patients before each application of the pDNA would be needed to perform safe clinical trials.

Cationic liposomes that were frequently used and validated in several laboratories and animal models may be the first approaches used for gene therapy using pDNA in humans. Meanwhile, the utilization of encapsulated mRNA appears theoretically as a method to circumvent all limitations of pDNA-based approaches for gene therapy. Encapsulated mRNA were not yet reported as gene therapy tools but because they are transient molecules, readily expressed in the cytosole, and if nonstabilized, barely activate immunity (TLR7 and TLR8), they can be foreseen as safe (transient), efficient (expressed in the cytosole), and nonimmunogenic (not stabilized) genetic information vehicles to be used in gene therapy.

## 2.4 Pharmaceutical Production and Regulations

***2.4.1 Plasmid DNA.*** Several companies in the United States and Europe offer the contract manufacture of documented pDNA for preclinical use and of GMP-certified pDNA for clinical utilisations (see, for example, the Plasmid Factory at www.plasmidfactory.com). It is regulatory practice that ampicilin is not used in GMP production in order to avoid cross-contaminations and concerns with penicillin-allergies. Instead, plasmids containing a kanamycin-resistance gene are frequently used. At best, the starting material is a pDNA containing a synthetic gene constructed by assembly of oligonucleotides. This way, the history of the pDNA can be clearly reported in the production's file. The whole manufacturing process is free of bovine-derived products. The plasmid of interest is tested in different *E. coli* strains and several culture media to find out the best conditions for producing the highest amount and best quality plasmid. The final product should meet regulatory guidelines such as those of the FDA: http://www.fda.gov/cberg/gdlns/plasdnavac.htm. It should appear as a clear and colorless solution by visual inspection. The final controls must meet the following criteria:

1. Identity: Sequencing should show 100% identity with the expected sequence. Electrophoresis or chromatography studies should be used to show that the pDNA is predominantly supercoiled. The restriction enzyme digest pattern analyzed by gel electrophoresis can be included. Susceptibility to DNase can also be used as a proof of molecular identity.
2. Content: Quantification should be performed by absorbance at 260 nm. Osmolarity and pH should also be measured using standard methods.
3. Purity: Residual proteins, chromosomal bacteria DNA, RNA, and endotoxin must be below specified limits. Sterility must be controlled by standard microbiological assays.

Moreover, some tests can be performed to check counter-ions (should be NaCl, thanks to precipitation of the pDNA with alcohol plus NaCl), residual solvents (if they are used during the production), and potency (functional assay using, for

example, transfection of cells and verification of the expression of the protein of interest or testing of the immune response after injection in an animal model).

Production of pDNA using up to 100-$m^3$ scale fermentation is performed by specialized industries [106]. Batches of more than 1 kg of pDNA can be produced. As functional doses for vaccination in humans were of ca. 800 μg (intramuscular injection) or 4 μg (gene-gun), 1 kg of pDNA is enough for 1,250,000 or 250,000,000 doses, respectively. This is largely enough for pDNA-based therapies to be commercialized worldwide as treatments against infectious diseases, cancer, or genetic diseases.

Preclinical toxicology studies should address the standard systemic and local reactogenicity, histopathology, and toxicity (acute and chronic dosage) in rodent and nonrodent animals. Also, as pDNA-based therapies are classified as gene therapy approaches, additional strict safety issues must be addressed before starting a clinical trial (for a review, see Ref. 35). The potency of the pDNA to integrate in the genome of the injected animal must be tested using assays that can detect one integrated pDNA in more than 150,000 nuclei. The potential of the pDNA preparation to induce anti-DNA antibodies (the etiologic agents of systemic lupus erythematosus) must be evaluated. The biodistribution and persistence of the pDNA must also be measured: Tissues and body fluids are analyzed at different time points for the presence of the pDNA using PCR methods. Of particular relevance is the study of pDNA in germ cells because such an event could theoretically generate transgenic descendants.

These toxicology studies were performed many times using many different pDNA batches. Close to no integration of pDNA in genomes, no induction of pathogenic anti-DNA antibodies, and a clearance of injected pDNA within days were reported. However, long-term (months) antigen expression in mice using intramuscular injections of a pDNA that may replicate in mammalian cells was found [107]. In this case, the cellular response triggered by the vaccine was weak and may in part account for the persistence of the foreign nucleic acid. The antigen coded by the vaccine in this preclinical study (Hepatitis Surface Antigen) associated with specific antibodies formed circulating immune complexes that caused pathologic lesions in liver and kidney. Thus, special care should be devoted to the design of the plasmid to guarantee that it cannot replicate in mammalian cells. Meanwhile, the regulatory authorities acknowledged the fact that since the first report of trials in humans using injections of up to 800 μg of nonviral coding pDNA (1998), no adverse events have ever been registered. Thus, especially for phase I trials, the toxicology studies can now be facilitated in agreement with the relevant authorities. For vaccination, gene-gun offers the extra-safety advantages that very low doses are used and that the pDNA is delivered exclusively in the dermis, avoiding systemic distribution and potential side effects while guaranteeing the elimination of most of the injected product through natural skin regeneration.

***2.4.2 Messenger RNA.*** Two companies offer contract manufacture of mRNA at laboratory and clinical grade (GMP-certified): Asuragen, a spin-off of Ambion, in the United States and CureVac in Europe. Here again, the whole manufacturing process is free of bovine-derived products. There are no official guidelines for the production of mRNA; however, it can be proposed that the final product should be checked by the standard quality controls. It should appear as a clear and

colorless solution by visual inspection. The following criteria would have to be met by the results of the final controls:

1. Identity: Sequencing of the plasmid used for *in vitro* transcription should show 100% identity with the expected sequence. At best, reverse transcription and sequencing of the final mRNA could be performed. It should also show 100% identity with the expected sequence. Susceptibility to RNase could additionally be used as a proof of molecular identity.
2. Content: Quantification should be performed by absorbance at 260 nm. Osmolarity and pH should also be measured using standard methods.
3. Purity: Residual proteins, chromosomal bacteria DNA, plasmid DNA, bacteria RNA, aberrant mRNA transcripts (smaller or larger by-products of the transcription) and endotoxin shoud be below specified limits. Sterility must be controlled by standard microbiological assays.

As is the case for pDNA, some physical tests can be performed to check counterions, residual solvents (if they are used during the production), and potency (functional assay using, for example, transfection of cells and verification of the expression of the protein of interest or testing of the immune response after injection in an animal model).

Preclinical toxicology studies should address the standard systemic and local reactogenicity, histopathology, and toxicity (acute and chronic dosage) in rodent and nonrodent animals. Since the FDA and some European authorities decided to classify mRNA-based therapies as no-gene therapy (for nonreplicative mRNA as depicted in this chapter), the implementation of clinical trials does not require additional specific toxicology testing.

## 3 CONCLUSIONS AND PERSPECTIVES

Since the initial publication by Wolff et al. showing that the injection of pDNA or mRNA simply coding for a protein can result in gene expression *in vivo*, many studies have been performed in animal models and humans showing that this method is safe and can elicit the expected results, which is the expression of the protein. Because pDNA and stabilized mRNA have the natural capacity to activate the immune system through the triggering of TLR, pDNA- and mRNA-based therapeutic utilizations are mostly tested in the field of vaccination. However, since gene therapies based on viruses have shown severe toxicity features in recent human trials, some pDNA or mRNA formulations may now be intensely developed as safer alternatives. Of special interest is the yet unexplored use of (encapsulated) mRNA for gene therapy. As this transient nucleic acid cannot persist or accumulate and would not generate unwanted long-term side effects even after repeated injections, it may seem as the optimal active pharmaceutical ingredient for gene therapies. Meanwhile, vaccination strategies based on nonviral coding nucleic acids are very advanced in terms of time to commercialization. The injection of naked pDNA and the adoptive transfert of APCs transfected with mRNA were demonstrated in many human clinical trials to be feasible and safe and to result in the development of the expected antigen-specific immunity. Optimal formulation, dosage, frequency of application, and adjuvant still need to be clearly defined.

Improvements of the methods include the optimization of the gene sequence, the optimization of the antigen processing, the enhancement of uptake using for example electrical methods, or the use of particles for formulating the vaccines. Concerning the last point, three types of particles were evaluated: liposomes, cationic degradable polymers, and coated gold particles delivered by gene-gun. This latter technology seems to be at the moment the most efficacious, versatile, cost-efficient method for vaccination with nucleic acids. Moreover, it is painless for the patient and very reproducible because the operator cannot influence the delivery (as opposed to injections, especially intradermal injections). The company Powdermed (www.powdermed.com) is developing and implementing the utilization of gene-gun-based therapies to deliver pDNA as vaccines against infectious diseases and cancer. Meanwhile, for autologuous vaccination, which may be needed for example with the mutating and patient-specific HIV virus, vaccination with adoptively transferred mRNA-transfected DCs as developed by Argos pharmaceuticals (www.argospharmaceuticals.com) may be of high potency. The choice between pDNA and mRNA for clinical utilizations is a matter of cost and efficacy. Although there was an intense development and optimization of pDNA-based vaccines, accompanied by the implementation of industrial production facilities, mRNA was until recently, relatively unexplored. However, as summarized in Table 1, mRNA offers many advantages over pDNA in the context of therapeutic utilizations. Being active in the cytosole, mRNA does not need to cross the very selective nuclear envelope to be expressed. Being transient, mRNA can allow the controlled production in terms of quantity and time frame of the protein of interest. Moreover, as it is a single-stranded molecule, mRNA does not have the discrete topologies that have pDNA and that could affect its clinical efficacy. This feature allows mRNA to have theoretically a higher functional batch-to-batch reproducibility. At the level of efficacy for vaccination, pDNA is assumed currently to be superior to mRNA; however, the direct comparison of optimal formulations of pDNA and mRNA using gene-gun delivery, for example, as a consistent and reliable method, still needs to be performed. Meanwhile the unique possibility that offers mRNA to transfect, without inducing death, DCs *in vitro*, is a guarantee of vaccine efficacy using a minimum amount of nucleic acid. Thus, although pDNA-based vaccination approaches are very advanced and may soon be commercialized, it can be

**TABLE 1.**

| Parameter | pDNA | mRNA |
|---|---|---|
| Delivery to the functional site | Hard (Nucleus) | Easy (Cytosole) |
| Control of level of expression | Hard | Easy |
| Control of duration of expression | Hard (Days to months) | Easy (Hours to few days) |
| Large-scale GMP production | Easy | Easy |
| Conservation | Easy | Easy |
| Batch-to-batch reproducibility | Low (Three topologies) | High |
| Regulatory issues | Severe (Gene therapy) | Easy |
| Immunogenicity (vaccines) | High | Low |
| Usage for gene therapy | Efficacious | Not tested |

envisioned that due to their safety features, mRNA-based approaches will be the second generation of nucleic-acid-based vaccination strategies.

Several companies active in the field of nucleic-acid-based vaccines will probably manage to get their product(s) and methods brought to the market for disease-specific utilizations. Prevention and treatments of hepatitis B, AIDS, influenza, and some cancer are the most advanced. Thus, in the near future, drugs containing pDNA or mRNA as an active pharmaceutical ingredient should be available in pharmacies and help in treating or preventing pandemic infections (HIV, influenza virus, *plasmodium falciparum*, etc.) as well diseases with unmet medical needs such as certain types of cancers. The continuous optimization of these methods ensures that the future pDNA- or mRNA-based drugs will get more and more efficacious and will be used for a large area of therapeutic applications.

## REFERENCES


1. Littlefield J W, Keller E B, Gross J, et al. (1955). Studies on cytoplasmic ribonucleoprotein particles from the liver of the rat. *J. Biol. Chem.* 217:111–123.
2. Palade G E (1955). A small particulate component of the cytoplasm. *J. Biophys. Biochem. Cytol.* 1:59–68.
3. Hoagland M B, Stephenson M L, Scott J F, et al. (1958). A soluble ribonucleic acid intermediate in protein synthesis. *J. Biol. Chem.* 231:241–257.
4. Chapeville F, Lipmann F, Von Ehrenstein G, et al. (1962). On the role of soluble ribonucleic acid in coding for amino acids. *Proc. Natl. Acad. Sci. U.S.A.* 48:1086–1092.
5. Feldmann H, Zachau H G (1964). Chemical evidence for the 3′-linkage of amino acids to s-RNA+. *Biochem. Biophys. Res. Commun.* 15:13–17.
6. Brenner S J F M M (1961). An unstable intermediiate carrying information from genes to ribosome for protein synthesis. *Nature.* 190:576–581.
7. Gros F, Gilbert W, Hiatt H H, et al. (1961). Molecular and biological characterization of messenger RNA. *Cold Spring Harb. Symp. Quant. Biol.* 26:111–132.
8. Crick F (1970). Central dogma of molecular biology. *Nature.* 227:561–563.
9. Joyce G F (2002). The antiquity of RNA-based evolution. *Nature.* 418:214–221.
10. Tourriere H, Chebli K, Tazi J (2002). mRNA degradation machines in eukaryotic cells. *Biochimie.* 84:821–837.
11. Ross J (1995). mRNA stability in mammalian cells. *Microbiol. Rev.* 59:423–450.
12. Festenstein R, Kioussis D (2000). Locus control regions and epigenetic chromatin modifiers. *Curr. Opin. Genet. Dev.* 10:199–203.
13. Sudowe S, Ludwig-Portugall I, Montermann E, et al. (2006). Prophylactic and therapeutic intervention in IgE responses by biolistic DNA vaccination primarily targeting dendritic cells. *J. Allergy Clin. Immunol.* 117:196–203.
14. Probst J, Brechtel S, Scheel B, et al. (2006). Characterization of the ribonuclease activity on the skin surface. *Genet. Vaccines. Ther.* 4:4.
15. Benner S A (1988). Extracellular ‘communicator RNA’. *FEBS Lett.* 233:225–228.
16. Kano Y, Miyashita T, Nakamura H, et al. (1981). In vivo correlation between DNA supercoiling and transcription. *Gene.* 13:173–184.
17. Sekiguchi J M, Swank R A, Kmiec E B (1989). Changes in DNA topology can modulate in vitro transcription of certain RNA polymerase III genes. *Mol. Cell Biochem.* 85:123–133.

18. Wolff J A, Malone R W, Williams P, et al. (1990). Direct gene transfer into mouse muscle in vivo. *Science*. 247:1465–1468.
19. Patil S D, Rhodes D G, Burgess D J (2005). DNA-based therapeutics and DNA delivery systems: A comprehensive review. *AAPS. J.* 7:E61–E77.
20. Singh M, Briones M, Ott G, et al. (2000). Cationic microparticles: A potent delivery system for DNA vaccines. *Proc. Natl. Acad. Sci. U.S.A.* 97:811–816.
21. Uchida M, Natsume H, Kishino T, et al. (2004). Immunization by particle bombardment of antigen-loaded poly-(DL-lactide-co-glycolide) microspheres in mice. *Vaccine*. 24:2120–2130.
22. Felgner J H, Kumar R, Sridhar C N, et al. (1994). Enhanced gene delivery and mechanism studies with a novel series of cationic lipid formulations. *J. Biol. Chem.* 269:2550–2561.
23. Hofland H E, Shephard L, Sullivan S M (1996). Formation of stable cationic lipid/DNA complexes for gene transfer. *Proc. Natl. Acad. Sci. U.S.A.* 93:7305–7309.
24. Dokka S, Toledo D, Shi X, et al. (2000). Oxygen radical-mediated pulmonary toxicity induced by some cationic liposomes. *Pharm. Res.* 17:521–525.
25. Filion M C, Phillips N C (1997). Toxicity and immunomodulatory activity of liposomal vectors formulated with cationic lipids toward immune effector cells. *Biochim. Biophys. Acta.* 1329:345–356.
26. Holcik M, Liebhaber S A (1997). Four highly stable eukaryotic mRNAs assemble 3′ untranslated region RNA-protein complexes sharing cis and trans components. *Proc. Natl. Acad. Sci. U.S.A.* 94:2410–2414.
27. Hoerr I, Obst R, Rammensee H G, et al. (2000). In vivo application of RNA leads to induction of specific cytotoxic T lymphocytes and antibodies. *Eur. J. Immunol.* 30:1–7.
28. Klencke B, Matijevic M, Urban R G, et al. (2002). Encapsulated plasmid DNA treatment for human papillomavirus 16-associated anal dysplasia: a Phase I study of ZYC101. *Clin. Cancer Res.* 8:1028–1037.
29. Conry R M, LoBuglio A F, Wright M, et al. (1995). Characterization of a messenger RNA polynucleotide vaccine vector. *Cancer Res.* 55:1397–1400.
30. Malone R W, Felgner P L, Verma I M (1989). Cationic liposome-mediated RNA transfection. *Proc. Natl. Acad. Sci. U.S.A.* 86:6077–6081.

30a. Gossen M, Bonin AL, Freundlieb S, Bujard H (1994). Inducible gene expression for higher eukaryotic cells. *Curr. Opin. Biotechnol. Oct.* 5(5):516–520 (Review).

31. Mockey M, Goncalves C, Dupuy F P, et al. (2006). mRNA transfection of dendritic cells: synergistic effect of ARCA mRNA capping with Poly(A) chains in cis and in trans for a high protein expression level. *Biochem. Biophys. Res. Commun.* 340: 1062–1068.
32. Stepinski J, Waddell C, Stolarski R, et al. (2001). Synthesis and properties of mRNAs containing the novel "anti-reverse" cap analogs 7-methyl(3′-O-methyl)GpppG and 7-methyl (3′-deoxy)GpppG. *RNA*. 7:1486–1495.
33. Boon T, Coulie P G, Van Den Eynde B J, et al. (2006). Human T cell responses against melanoma. *Annu. Rev. Immunol.* 24:175–208.
34. Selin L K, Vergilis K, Welsh R M, et al. (1996). Reduction of otherwise remarkably stable virus-specific cytotoxic T lymphocyte memory by heterologous viral infections. *J. Exp. Med.* 183:2489–2499.
35. Liu M A, Ulmer J B (2005). Human clinical trials of plasmid DNA vaccines. *Adv. Genet.* 55:25–40.

36. Pascolo S (2004). Messenger RNA-based vaccines. *Expert. Opin. Biol. Ther.* 4: 1285–1294.
37. Chau V, Tobias J W, Bachmair A, et al. (1989). A multiubiquitin chain is confined to specific lysine in a targeted short-lived protein. *Science.* 243:1576–1583.
38. Anthony L S, Wu H, Sweet H, et al. (1999). Priming of CD8+ CTL effector cells in mice by immunization with a stress protein-influenza virus nucleoprotein fusion molecule. *Vaccine.* 17:373–383.
39. Chhabra A, Mehrotra S, Chakraborty N G, et al. (2004). Cross-presentation of a human tumor antigen delivered to dendritic cells by HSV VP22-mediated protein translocation. *Eur. J. Immunol.* 34:2824–2833.
40. Peng S, Trimble C, He L, et al. (2006). Characterization of HLA-A2-restricted HPV-16 E7-specific CD8(+) T-cell immune responses induced by DNA vaccines in HLA-A2 transgenic mice. *Gene Ther.* 13:67–77.
41. Momburg F, Fuchs S, Drexler J, et al. (1993). Epitope-specific enhancement of antigen presentation by invariant chain. *J. Exp. Med.* 178:1453–1458.
42. Ruff A L, Guarnieri F G, Staveley-O'Carroll K, et al. (1997). The enhanced immune response to the HIV gp160/LAMP chimeric gene product targeted to the lysosome membrane protein trafficking pathway. *J. Biol. Chem.* 272:8671–8678.
43. Boyle J S, Brady J L, Lew A M (1998). Enhanced responses to a DNA vaccine encoding a fusion antigen that is directed to sites of immune induction. *Nature.* 392:408–411.
44. Hung C F, Hsu K F, Cheng W F, et al. (2001). Enhancement of DNA vaccine potency by linkage of antigen gene to a gene encoding the extracellular domain of Fms-like tyrosine kinase 3-ligand. *Cancer Res.* 61:1080–1088.
45. Xiang R, Primus F J, Ruehlmann J M, et al. (2001). A dual-function DNA vaccine encoding carcinoembryonic antigen and CD40 ligand trimer induces T cell-mediated protective immunity against colon cancer in carcinoembryonic antigen-transgenic mice. *J. Immunol.* 167:4560–4565.
46. Alexander J, Sidney J, Southwood S, et al. (1994). Development of high potency universal DR-restricted helper epitopes by modification of high affinity DR-blocking peptides. *Immunity.* 1:751–761.
47. Gregor P D, Wolchok J D, Turaga V, et al. (2005). Induction of autoantibodies to syngeneic prostate-specific membrane antigen by xenogeneic vaccination. *Int. J. Cancer.* 116:415–421.
48. Pupa S M, Iezzi M, Di Carlo E, et al. (2005). Inhibition of mammary carcinoma development in HER-2/neu transgenic mice through induction of autoimmunity by xenogeneic DNA vaccination. *Cancer Res.* 65:1071–1078.
49. Barouch D H, Yang Z Y, Kong W P, et al. (2005). A human T-cell leukemia virus type 1 regulatory element enhances the immunogenicity of human immunodeficiency virus type 1 DNA vaccines in mice and nonhuman primates. *J. Virol.* 79:8828–8834.
50. Wolff J A, Budker V (2005). The mechanism of naked DNA uptake and expression. *Adv. Genet.* 54:3–20.
51. Maloy K J, Erdmann I, Basch V, et al. (2001). Intralymphatic immunization enhances DNA vaccination. *Proc. Natl. Acad. Sci. U.S.A.* 98:3299–3303.
52. Tagawa S T, Lee P, Snively J, et al. (2003). Phase I study of intranodal delivery of a plasmid DNA vaccine for patients with Stage IV melanoma. *Cancer.* 98:144–154.
53. Prud'homme G J, Glinka Y, Khan A S, et al. (2006). Electroporation-enhanced nonviral gene transfer for the prevention or treatment of immunological, endocrine and neoplastic diseases. *Curr. Gene. Ther.* 6:243–273.

54. Robertson J S, Griffiths E (2001). Assuring the quality, safety, and efficacy of DNA vaccines. *Mol. Biotechnol.* 17:143–149.
55. Ledwith B J, Manam S, Troilo P J, et al. (2000). Plasmid DNA Vaccines: Investigation of Integration into host cellular DNA. *Intervirology.* 43:258–272.
56. Tang D C, DeVit M, Johnston S A (1992). Genetic immunization is a simple method for eliciting an immune response. *Nature.* 356:152–154.
57. Qiu P, Ziegelhoffer P, Sun J, et al. (1996). Gene gun delivery of mRNA in situ results in efficient transgene expression and genetic immunization. *Gene Ther.* 3:262–268.
58. Jonuleit H, Kuhn U, Muller G, et al. (1997). Pro-inflammatory cytokines and prostaglandins induce maturation of potent immunostimulatory dendritic cells under fetal calf serum-free conditions. *Eur. J. Immunol.* 27:3135–3142.
59. Martinon F, Krishnan S, Lenzen G, et al. (1993). Induction of virus-specific cytotoxic T lymphocytes in vivo by liposome-entrapped mRNA. *Eur. J. Immunol.* 23:1719–1722.
60. Hess P R, Boczkowski D, Nair S K, et al. (2006). Vaccination with mRNAs encoding tumor-associated antigens and granulocyte-macrophage colony-stimulating factor efficiently primes CTL responses, but is insufficient to overcome tolerance to a model tumor/self antigen. *Cancer Immunol. Immunother.* 55:672–683.
61. Jones D H, Corris S, McDonald S, et al. (1997). Poly(DL-lactide-co-glycolide)-encapsulated plasmid DNA elicits systemic and mucosal antibody responses to encoded protein after oral administration. *Vaccine.* 15:814–817.
62. Perrie Y, Frederik P M, Gregoriadis G (2001). Liposome-mediated DNA vaccination: the effect of vesicle composition. *Vaccine.* 19:3301–3310.
63. Chen S C, Jones D H, Fynan E F, et al. (1998). Protective immunity induced by oral immunization with a rotavirus DNA vaccine encapsulated in microparticles. *J. Virol.* 72:5757–5761.
64. Herrmann J E, Chen S C, Jones D H, et al. (1999). Immune responses and protection obtained by oral immunization with rotavirus VP4 and VP7 DNA vaccines encapsulated in microparticles. *Virol.* 259:148–153.
65. Singh M, Vajdy M, Gardner J, et al. (2001). Mucosal immunization with HIV-1 gag DNA on cationic microparticles prolongs gene expression and enhances local and systemic immunity. *Vaccine.* 20:594–602.
66. Sharpe S, Hanke T, Tinsley-Bown A, et al. (2003). Mucosal immunization with PLGA-microencapsulated DNA primes a SIV-specific CTL response revealed by boosting with cognate recombinant modified vaccinia virus Ankara. *Virol.* 313:13–21.
67. Kusakabe K-I, Xin K-Q, Katoh H, et al. (2000). The timing of GM-CSF expression plasmid administration influences the Th1/Th2 response induced by an HIV-1-specific DNA vaccine. *J. Immunol.* 164:3102–3111.
68. Carralot J-P, Probst J, Hoerr I, et al. (2004). Polarization of immunity induced by direct injection of naked sequence-stabilized mRNA vaccines. *Cell. Molec. Life Sci.* 61(18):2418–2424.
69. Girard M P, Osmanov S K, Kieny M P (2006). A review of vaccine research and development: the human immunodeficiency virus (HIV). *Vaccine.* 24:4062–4081.
70. Calarota S, Bratt G, Nordlund S, et al. (1998). Cellular cytotoxic response induced by DNA vaccination in HIV-1-infected patients. *Lancet.* 351:1320–1325.
71. Calarota S A, Weiner D B (2004). Approaches for the design and evaluation of HIV-1 DNA vaccines. *Expert. Rev. Vacc.* 3:S135–S149.
72. Giri M, Ugen K E, Weiner D B (2004). DNA vaccines against human immunodeficiency virus type 1 in the past decade. *Clin. Microbiol. Rev.* 17:370–389.

73. Wang R, Doolan D L, Le T P, et al. (1998). Induction of antigen-specific cytotoxic T lymphocytes in humans by a malaria DNA vaccine. *Science.* 16:476–480.
74. Wang R, Richie T L, Baraceros M F, et al. (2005). Boosting of DNA vaccine-elicited gamma interferon responses in humans by exposure to malaria parasites. *Infect. Immun.* 73:2863–2872.
75. Zuckerman A J (2006). HIV variants and hepatitis B surface antigen mutants. *J. Med. Virol.* 78(suppl 1):S1–S2.
76. Roy M J, Wu M S, Barr L J, et al. (2000). Induction of antigen-specific CD8+ T cells, T helper cells, and protective levels of antibody in humans by particle-mediated administration of a hepatitis B virus DNA vaccine. *Vaccine.* 19:764–768.
77. Rottinhaus S T, Poland G A, Jacobson R M, et al. (2003). Hepatitis B DNA vaccine induces protective antibody responses in human non-responders to conventional vaccination. *Vaccine.* 21:4604–4608.
78. Lowrie D B, Silva C L, Colston M J, et al. (1997). Protection against tuberculosis by a plasmid DNA vaccine. *Vaccine.* 15:834–838.
79. Lowrie D B, Tascon R E, Bonato V L, et al. (1999). Therapy of tuberculosis in mice by DNA vaccination. *Nature.* 400:269–271.
80. Ha S-J, Jeon B-Y, Kim S-C, et al. (2003). Therapeutic effect of DNA vaccines combined with chemotherapy in a latent infection model after aerosol infection of mice with Mycobacterium tuberculosis. *Gene Ther.* 10(18):1592–1599.
81. Shaw D R, Strong T V (2006). DNA vaccines for cancer. *Frontiers Biosci.* 11: 1189–1198.
82. Lollini P L, De Giovanni C, Pannellini T, et al. (2005). Cancer immunoprevention. *Future Oncol.* 1:57–66.
83. Rosenberg S A, Yang J C, Sherry R M, et al. (2003). Inability to immunize patients with metastatic melanoma using plasmid DNA encoding the gp 100 melanoma-melanocyte antigen. *Hum. Gene Ther.* 14(8):709–714.
84. Pavlenko M, Roos A K, Lundqvist A, et al. (2004). A phase I trial of DNA vaccination with a plasmid expressing prostate-specific antigen in patients with hormone-refractory prostate cancer. *Br. J. Cancer.* 91(4):688–694.
85. Timmerman J M, Singh G, Hermanson G, et al. (2002). Immunogenicity of a plasmid DNA vaccine encoding chimeric idiotype in patients with B-Cell lymphoma. *Cancer Res.* 62:5845–5852.
86. Klencke B, Matijevic M, Urban R G, et al. (2002). Encapsulated plasmid DNA treatment for human papillomavirus 16-associated anal dysplasia: a Phase I study of ZYC101. *Clin. Cancer Res.* 8:1028–1037.
87. Sheets E E, Urban R G, Crum C P, et al. (2003). Immunotherapy of human cervical high-grade cervical intraepithelial neoplasia with microparticle-delivered human papillomavirus 16 E7 plasmid DNA. *Am. J. Obstet. Gynecol.* 188:916–926.
88. Boczkowski D, Nair S K, Snyder D, et al. (1996). Dendritic cells pulsed with RNA are potent antigen-presenting cells in vitro and in vivo. *J. Exp. Med.* 184:465–472.
89. Gilboa E, Vieweg J (2004). Cancer immunotherapy with mRNA-transfected dendritic cells. *Immunol. Rev.* 199:251–263.
90. Granstein R D, Ding W, Ozawa H (2000). Induction of anti-tumor immunity with epidermal cells pulsed with tumor-derived RNA or intradermal administration of RNA. *J. Invest. Dermatol.* 114:632–636.
91. Nishikawa H, Kato T, Tanida K, et al. (2003). $CD4^+$ $CD25^+$ T cells responding to serologically defined autoantigens suppress antitumor immune responses. *Proc. Natl. Acad. Sci. U.S.A.* 100(19):10902–10906.

92. Wolf B L, Hamilton R G (2002). Near-fatal anaphylaxis after Hymenoptera venom immunotherapy. *J. Aller. Clin. Immunol.* 102(3):527–528.
93. Chua K Y, Huangfu T, Liew L N (2006). DNA vaccines and allergic diseases. *Clin. Exper. Pharmacol. Physiol.* 33(5–6):546–550.
94. Wherry E J, Barber D L, Kaech S M, et al. (2004). Antigen-independent memory CD8 T cells do not develop during chronic viral infection. *Proc. Natl. Acad. Sci. U.S.A.* 101:16004–16009.
95. Nishikawa M, Hashida M (2002). Nonviral approaches satisfying various requirements for effective in vivo gene therapy. *Biol. Pharm. Bull.* 25:275–283.
96. Budker V, Zhang G, Danko I, et al. (1998). The efficient expression of intravascularly delivered DNA in rat muscle. *Gene Ther.* 5(2):272–276.
97. Zhang G, Budker V, Williams P, et al. (2001). Efficient expression of naked dna delivered intraarterially to limb muscles of nonhuman primates. *Hum. Gene Ther.* 12(4): 427–438.
98. Yoshimura K, Rosenfeld M A, Nakamura H, et al. (1992). Expression of the human cystic fibrosis transmembrane conductance regulator gene in the mouse lung after *in vivo* intratracheal plasmid-mediated gene transfer. *Nucleic Acids Res.* 20:3233–3240.
99. McCluskie et al. (1998). *Antisense Nucl. Acid Drug Devel.* 8:401–414.
100. Liu Y, Mounkes L C, Liggitt H D, et al. (1997). Factors influencing the efficiency of cationic liposome-mediated intravenous gene delivery. *Nature Biotechnol.* 15:167–173.
101. Kawakami S, Fumoto S, Nishikawa M (2000). In vivo gene delivery to the liver using novel galactosylated cationic liposomes. *Pharm. Res.* 17(3):306–313.
102. Kawakami S, Sato A, Nishikawa M, et al. (2000). Mannose receptor-mediated gene transfer into macrophages using novel mannosylated cationic liposomes. *Gene Ther.* 7:292–299.
103. Wu G Y, Wu C H (1988). Receptor-mediated gene delivery and expression in-vivo. *J. Biol. Chem.* 263:14621–14624.
104. Nishikawa M, Takemura S, Yamashita F (2000). Pharmacokinetics and in vivo gene transfer of plasmid DNA complexed with mannosylated poly(L-lysine) in mice. *J. Drug Target.* 8:29–38.
105. Perales J C, Ferkol T, Beegen H, et al. (1994). Gene Transfer in vivo: Sustained expression and regulation of genes introduced into the liver by receptor-targeted uptake. *Proc. Natl. Acad. Sci. U.S.A.* 91:4086–4090.
106. Hoare M, Levy M, Bracewell D G, et al. (2005). Biopocess engineering issues that would be faced in producing a DNA vaccine at up to 100 $m^3$ fermentation scale for an influenza pandemic. *Biotechnol. Prog.* 21:1577–1592.
107. Zi, Yao, Zhu, et al. (2006). *Euro. J. Immuno.* 36(4):875–886.

# FORMULATIONS AND DELIVERY LIMITATIONS OF NUCLEIC-ACID-BASED THERAPIES

TATIANA SEGURA
*University of California, Los Angeles, Los Angeles, California*

**Chapter Contents**

1 Introduction
2 Systemic and Direct Injection Delivery
  2.1 Formulations
3 Matrix-Based Delivery
  3.1 Direct Encapsulation and Release
  3.2 Matrix-Tethered Delivery
4 Delivery Limitations and Current Solutions
  4.1 Extracellular Limitations
  4.2 Intracellular Limitations
5 Outlook
  References

## 1 INTRODUCTION

The ability to control the gene expression of a desired cellular population via the delivery of nucleic acids has inspired applications in clinical and basic science. In the past, nonviral gene delivery referred to the delivery of plasmid DNA (pDNA) whereby the expression of a target protein, which could restore normal biological function or aid during wound healing, would be increased. More recently, however, the term *gene delivery* has expanded to include the delivery of small interfering RNA (siRNA) and oligonucleotides (ON), which can be used to decrease or

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

downregulate the expression of a target protein. Ongoing clinical trials using gene delivery as a therapeutic agent include cancer [1], cystic fibrosis [2–4], adenosine deaminase deficiency [5], and infectious diseases such as acquired immunodeficiency syndrome (AIDS) [6].

Gene delivery employs both viral and nonviral vectors to deliver nucleic acids. Although viral delivery strategies are in general more efficient than their nonviral counterparts, an increasing concern about immune responses to viral vectors *in vivo*, which could pose additional risks for patients undergoing gene therapy [7], has motivated the development of nonviral delivery vectors that aim to synthetically mimic biologically produced viruses. Nonviral delivery vectors have to be engineered to be able to package plasmid DNA to form particles that can (1) travel through biological fluids without losing activity, (2) target appropriate cells, (3) be internalized by the targeted cell population, and (4) finally trafficked to the intracellular site of action (e.g., nucleus and/or cytosol; Figure 1). Furthermore, the ideal delivery system would achieve the above-mentioned points while being nonimmunogenic and nontoxic. Although there have been numerous approaches to the engineering of delivery vectors, this chapter will focus mainly on nonviral delivery vectors that show promising approaches to overcome the delivery limitations mentioned above or that have been tested *in vivo*.

Strategies to administer nonviral delivery vectors *in vivo* can be divided into two main approaches: systemic delivery or direct injection of DNA nanoparticles and the implantation of a matrix that releases "naked" DNA or DNA nanoparticles. Systemic delivery aims at engineering delivery vectors that can be delivered noninvasively through a bodily fluid such as intravenously, transdermally, and orally, while targeting an appropriate tissue (e.g., tumor) or organ (e.g., lung). Local delivery, on the other hand, aims at bypassing the tissue-targeting step by delivering the nonviral vectors directly (or nearby) to the site of action, typically with the use of a polymeric matrix. The polymeric matrix can be designed to slowly release the

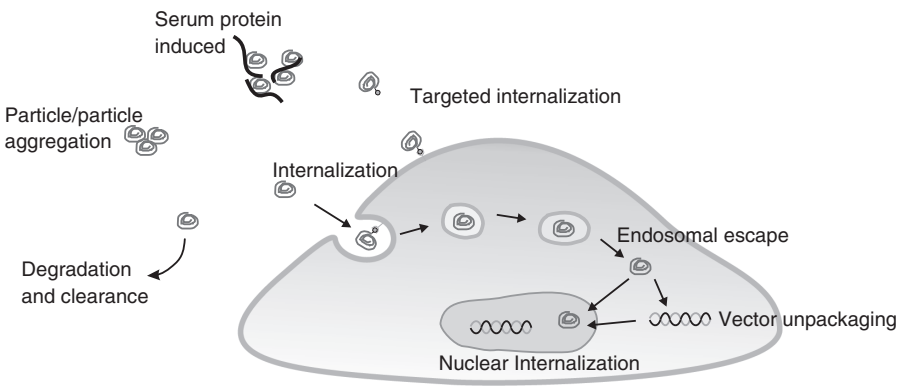


**Figure 1.** Limitations of nonviral gene delivery formulations are shown. Limitations can be divided into extracellular limitations, such as (1) interaction with serum components, (2) particle–particle interaction, (3) clearance from target site, and (4) targeting, and intracellular limitations, such as (1) internalization, (2) endosomal escape, (3) vector unpackaging, and (4) nuclear internalization.

DNA nanoparticles to the local environment where cells can internalize them or can be designed so that it can be infiltrated with local cells that can take up the DNA nanoparticles residing inside the matrix.

Design parameters for targeted cellular internalization and intracellular trafficking are similar for systemic and local delivery approaches; however, the design of the vector physical properties such as surface coating and particle size must be tailored to the method of application. This chapter will be divided into sections with Section 2 discussing systemic and direct injection formulations (Tables 1–4), Section 3 discussing matrix-based delivery formulations (Tables 2–5), and Section 4 discussing limitations with nonviral nucleic acid delivery and current solutions (Tables 4 and 7).

## 2 SYSTEMIC AND DIRECT INJECTION DELIVERY

*In vivo* production and secretion of therapeutic proteins by DNA delivery can be achieved through either systemic or local administration, each providing a unique opportunity for gene therapy. Systemic delivery allows noninvasive access to many target cells and tissue that are not accessible by direct administration. The most common approach involves the complexation of naked DNA with cationic lipids (lipoplexes) or cationic polymers (polyplexes) that package the DNA into nano-sized aggregates. The major factors limiting the *in vivo* effectiveness for systemically delivered lipoplexes and polyplexes is colloidal instability and ineffective targeting. For intravenous administration, complexes are rapidly eliminated from the blood stream (<30 minutes) and found mostly in the liver and to a lesser extent in organs with fine capillary beds like skin, muscle, and intestine [8]. The interaction of biomolecules from the physiological fluid with the complexes induces aggregation and dissociation of the complexes, which limits bioavailability and internalization [9, 10]. The association of serum components (albumin, heparin, lipoprotein, or specific opsonins) with polyplexes and lipoplexes, which can target the complex for clearance by macrophages and is affected by the surface charge density and surface morphology of complexes [11]. Cationic formulations (i.e., lipoplexes and polyplexes) can provide an enhanced stability against degradation and may allow for a more efficient interaction with the negatively charged endothelial cell; however, noncationic or less cationic formulations may be required to avoid accumulation in the reticuloendothelial system. Variations in the charge ratio may be used as part of a passive targeting strategy. For lipoplexes delivered intravenously, different formulations (e.g., charge ratio) can be used to target the lung, liver, blood cells, or tumors [12, 13].

To avoid the difficulties of crossing the endothelium and nonspecific uptake by the liver, DNA can be delivered directly to the target tissue by injection of polyplexes, lipoplexes, or microspheres with encapsulated DNA. Injection of lipoplexes and polyplexes into the desired tissue can be used to achieve *in vivo* gene transfer. However, transfection depends on the physico-chemical properties of the complexes and the rate of clearance from the tissue, which varies from tissue to tissue and depends on factors such as the lymph supply. Nevertheless, polyplexes have been found to be sufficiently small and stable so as to diffuse throughout the brain ventricular spaces after a local injection injection [14].

## 2.1 Formulations

Unprotected plasmid DNA (naked DNA) is unstable under *in vivo* conditions, due to rapid degradation by serum nucleases [10, 15, 16]. Therefore, carriers or vectors are necessary to protect DNA or RNA from degradation to facilitate uptake into specific cells, and to transfer the DNA or RNA their target location (e.g., nucleus or cytosol). Delivery vectors for systemic delivery are designed to increase gene transfer function by enhancing the stability of DNA, the efficiency of cellular uptake and intracellular trafficking, and the biodistribution of DNA. In this section, the formulations that have been developed to deliver nucleic acids such as plasmid DNA (pDNA), oligonucleotide (ON), or small interfering RNA (siRNA) will be reviewed. Tables 1 and 2 summarize the most widely used cationic polymer and cationic lipid formulations.

***Cationic Polymers.*** Cationic polymers have been used since the late 1980s [20] as formulations for nucleic acid delivery. Cationic polymers contain high densities of primary, secondary, or tertiary amines, some of which are protonated at neutral pH. This high density of positive charges allows the cationic polymers to form stable condensed structures with pDNA, termed polyplexes, which are capable of entering the cell. Furthermore, polyplexes protect the DNA from nucleases found in serum and other extracellular environments. Table 1 shows the structure of frequently used cationic polymers, which have been used *in vivo* to deliver therapeutic genes, although many novel polymers are being developed, including polyallylamine [29], peptoids [30], poly(dimethyl aminoethyl methacrylate) [31, 32], poly(trime-thyl aminoethyl methacrylate) [33], poly(β-amino ester) [34, 35], and poly(phosphoester) [36–38]. Cationic polymers vary widely in molecular architecture, ranging from linear to highly branched molecules, which influences their complexation with nucleic acids as well as their transfection efficiency. In addition to providing positive charges for DNA complexation, the primary amines also serve as functional groups to chemically modify the polymers with ligands and peptides that can enhance one or more of the steps in the transfection process (Section 4, Tables 3, 4). Furthermore, tertiary and some secondary amines are typically neutral at phy-siological pH and have been found to aid in intracellular trafficking by facilitating the release of the polyplexes from the endosome though endosomal buffering (Section 4.2). The use of cationic polymers and lipids to deliver therapeutically relevant genes and their therapeutic outcome is summarized in Tables 5, 6, and 7 for formulations delivered systemically or via a matrix-based delivery approach.

*Interaction of Cationic Polymers with DNA.* Cationic polymers form aggregates with DNA, termed polyplexes, via electrostatic interactions, which protect the nucleic acid from degradation [121]. During the complexation of DNA with cationic polymers, the extended structure of DNA is changed to a more condensed configuration forming aggregates in the nanometer size range. Polyplexes typically give rise to particles with spherical, globular, or rod-like structures, which are 20 to 200 nm in diameter [122, 123] but can reach up to 1000 nm. Evidence that electrostatic interactions mediate the complexation of cationic polymers include Fourier transform infrared resonance (FTIR) data showing a reduction of the asymmetric

**TABLE 1. Structure of Commonly Used Cationic Polymer Formulations for Nonviral Gene Delivery**

| PEI, Poly(ethylene imine) [17] | PAMAM, Poly(amidoamine) dendrimer [18] | Cyclodextrin containing polymers [19] |
|---|---|---|
| 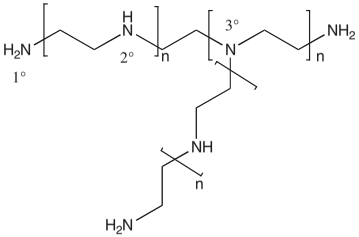 | 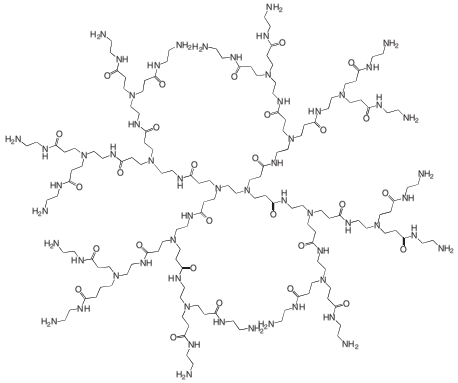 | 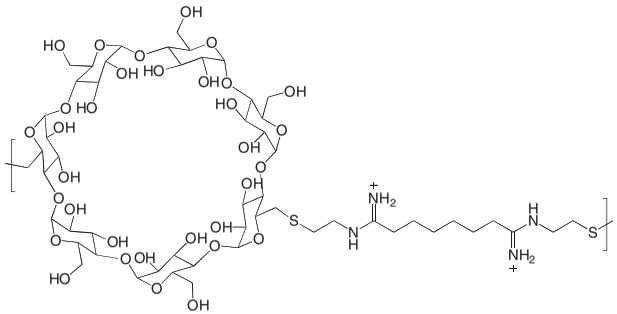 |
| **Poly(amino acids) [20]** | **Chitosan [21]** | **Dextran Spermine [22]** |
| 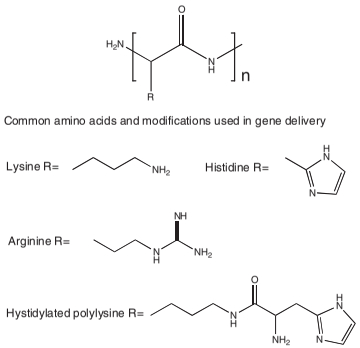  | 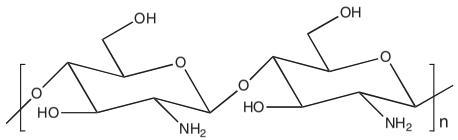 | 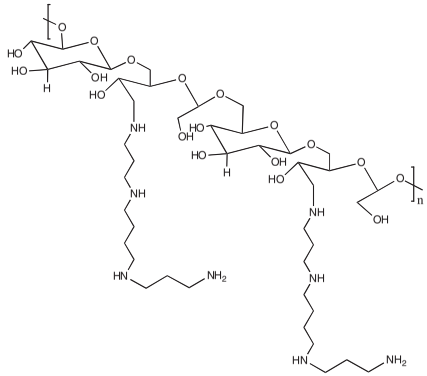 |

**Table 2. Structure of Commonly Used Lipid Formulations for Nonviral Gene Delivery**

| Cationic lipids | Neutral lipids |
|---|---|
| DOTMA [2,3-bis(oleoyl)propyl]trimethyl ammonium chloride [23] | DOPE dioleoylphosphatidyl ethanolamine [24] |
| | |
| DMRIE 1,2-dimyristyloxypropyl-3-dimethyl-hydroxyethyl ammonium bromide [25] | DOPC dioleoylphosphatidyl choline, Chol: cholesterol [26] |
| | |
| DODAB dioctadecyldimethyl ammonium bromide [27] | Chol cholesterol [28] |
| | |
| DOTAP 1,2-diacyl-3-trimethylammonium propane [29a] | DC-Chol 3[N-(N′,N′-dimethylaminoethane)-carbamoyl] cholesterol [28] |
| | |

**TABLE 3. Extracellular limitations to nonviral gene delivery**

| Vector Stabilization | Mechanism | Refs. |
|---|---|---|
| PEG | Provides stability against serum components | [39] |
| Cyclodextrin | Reduces toxicity and inflammation, provides stability against serum components | [40] (Rev. [41]) |
| Chitosan | Provides stability against serum components | [42] |
| Poly(glucaramidoamine) | Provides stability against serum components | [43] |
| CHKKKKKKHC | Covalent stabilization | [44] |

| Targeting | Cell Surface Target | Target Cell/Tissue | Refs. |
|---|---|---|---|
| Folate | Folate receptor | Cancer cells/ tumor | [45] (Rev. [46]) |
| Transferrin | Transferring receptor | Cancer cells/ tumor | [47] (Rev. [48]) |
| Mannose | Mannose receptor | Dendritic cells | [49, 50] |
| Galactose | Gal/GalNAc receptor | Hepatocytes/dendritic cells | [51] |
| Anti-OV-TL16 antibody fragment | OA3 receptor | Ovarian carcinoma cells | [52] |
| Anti-HER2 antibody | Human epidermal growth factor receptor-2 | Breast, ovarian cancer cells | [53] |
| Anti-PECAM antibody | Platelet endothelial cell adhesion molecule | Lung endothelia | [54] |
| RGD peptide | Integrin | Broad range of cells | [55] |
| Anti-CD3 | CD3 receptor | Lymphocytes | [56] |
| EGF (epidermal growth factor) | EGF receptor | Cancer cells | [57] |

phosphate stretching vibration of plasmid DNA after complexation with polyethyleneimine (PEI) [124]. Furthermore, microcalorimetric measurements have shown that complex formation between the graft copolymer poly(ethylene oxide)-PEI and a poly(dTA) result from the formation of ion pairs between ionized amino groups of PEI segments of the copolymer and the phosphate groups of DNA [125]. An increase in salt concentration generally results in a decrease binding affinity between the cationic polymer and the DNA probably due to a charge shielding effect at the higher salt concentrations [125–127].

The complexation and condensation behavior depend on the polymer's physical properties, including molecular weight, density of charges, whether it is linear or branched, and the ratio of polymer to DNA [10, 15, 127]. The molecular weight of the cationic polymer influences both the condensation behavior as well as the complex size. In general, for most cationic polymers, an increase in molecular weight results in a decrease of complex size until a condensation limit is achieved. For example, for PEI, it has been found that molecular weights higher than 25 kDa showed no further increase of complex size, whereas molecular weights of 2 kDa or lower result in a decreased ability to complex DNA to form small complexes

**TABLE 4. Intracellular Limitations to Nonviral Gene Delivery**

| Toxicity | Mechanism | Refs. |
|---|---|---|
| PEG | Charge shielding | [47, 58, 59] |
| Cyclodextrin | Charge shielding | [40] (Rev. [41]) |
| Poly(glucaramidoamine) | Charge shielding | [43] |
| P(EI-co-LSA) | Degradability | [60] |
| PEI-PEG-PEL | Degradability | [61] |

| Endosomal Escape | Mechanism | Sequence | Refs. |
|---|---|---|---|
| Secondary and tertiary amines | Endosomal buffering | Not applicable | [17, 18] |
| INF (influenza virus-derived sequence) | Endosomolytic at pH 5.0 by adopting an amphipathic α-helical conformation | GLFEAIEGFIENGWEGMIDGWYG | [62] |
| KALA | pH-dependent endosomolytic activity by adopting an α-helical conformation | WEAKLAKALAKALAKHLAKALA KALKACEA | [63, 64] |
| H5WYG | pH-dependent endosomolytic between pH 7 and 6. | GLFHAIAHFIHGGWHGLIHGWYG | [65] |
| GALA | pH-dependent membrane fusion | WEAALAEALAEALAEHLAEALA EALEALAA | [66, 67] |
| poly(α-alkyl acrylic acids) | pH-dependent hydrophilic-to-hydrophobic conversion causes membrane disruption | Poly(propyl acrylic acid) | [68, 69] |

| Vector Unpackaging | Mechanism | Refs. |
|---|---|---|
| Polylysine | Partial carbohydrate substitutions to reduce electrostatic interactions | [70] |
| Polylysine | Partial polyhydroxyalcanoyl (gluconoyl) substitution to reduce electrostatic interactions | [71] |
| Polylysine | Short polylysine facilitates unpacking | [72] |
| PEI | Partial acetylation PEI decreases the physiological buffering capacity | [70, 71, 73] |
| Chitosan | The chitosan is degraded by the predelivered chitosanase | [74] |
| Chitosan-NIPAAm | Thermoresponsive enhances decomplexation below critical solubility temperature | [75] |
| CHKKKKKKHC | Environmentally sensitive disulfide bonds | [44] |

| Nuclear Localization | Mechanism | Sequence | Refs. |
|---|---|---|---|
| SV40 | Importin-α | PKKKRKVEDPY | [76] |
| M9 | Transportin | GNQSSNFGPMKGGNFGGRSSGP YGGGGQYFAKPRNQGGY | [77] |
| Importin-β | NPC | Importin-β (1-643aa) | [78] |
| Mu | Unknown | MRRAHHRRRRASHRRMRGG | [79] |

**TABLE 5. *In vivo* Systemic and Direct Injection Delivery Studies Using Polyplexes Vectors to Deliver Therapeutically Relevant Genes**

| Formulation | Therapeutic Target | Nucleic Acid | Gene | Species/Delivery | Ref. |
|---|---|---|---|---|---|
| PEI | Viral infection | siRNA | Influenza virus | Mice/retro-orbitally | [80] |
| PEI | Cancer | siRNA | HER-2 | Mice/i.p. injection | [81] |
| PEI | Cancer | DNA | sst2 receptor | Mice, hamster/local | [82] |
| PEI | Cancer | DNA | IL-12 | Mice/inhalation | [83] |
| PEI | Cancer | DNA | p53 | Mice/inhalation | [83] |
| PEG-PEI-Chol | Cancer | DNA | pmIL-12 | Mice/Local | [84] |
| PEI/Pluronic F127 | Vascular diseases | DNA | Hu-uPA | Rats/Local | [85] |
| Chitosan | Viral infection | DNA | RSV antigens | Mouse/Intranasal | [86] |
| Cyclodextrin | Cancer | siRNA | EWS-FLI1 | Mice/Tail vein injection | [87] |
| Cyclodextrin-PEG | Cancer | DNA | DNAzyme | Mice/Intraveneous and intraperitoneal | [88] |

**TABLE 6.** ***In vivo*** **Systemic and Direct Injection Studies Using Lipoplex Formulations to Deliver Therapeutically Relevant Nucleic Acids**

| Formulation | Therapeutic Target | Nucleic Acid | Gene | Species/Delivery | Ref. |
|---|---|---|---|---|---|
| DOTAP:Chol | Cancer | DNA | Mda-7/IL-24 | Mice/intravenous, local | [89] |
| DC-Chol:DOPE | Cancer | DNA | HLA-A2 | Human/direct injection | [90] |
| DC-Chol:DOPE | Cancer | DNA | HLA-B13 | Human/direct injection | [90] |
| DC-Chol:DOPE | Cancer | DNA | H-2K | Human/direct injection | [90] |
| DC-Chol:DOPE | Cancer | DNA | HLA-B7 | Human/intratumorally and intraveneous | [91] |
| DC-Chol | Cancer | DNA | HER-2/neu | Mice/intraperitoneal injection | [92] |
| DC-Chol | Cancer | DNA | HER-2/neu | Human/intracavitary injection | [93] |
| LSFV | Cancer | DNA | IL-12 | Human/intraveneously | [94] |
| DMRIE:DOPE | Cancer | DNA | HLA-B7 | Human/intratumorally | [95] |
| TMAG:DLPC:DOPE | Cancer | DNA | IFN-β | Human/direct injection | [96] |
| DMRIE:DOPE | Cancer | DNA | HLA-B7 and β2-microglobulin | Human/intratumorally | [97] |
| DC-Chol:DOPE | Cystic fibrosis | DNA | CFTR | Human/direct instillation | [98] |
| DC-Chol:DOPE | Cystic fibrosis | DNA | CFTR | Human/topical application | [99] |
| 67A (GL-67:DOPE:DMPE:PEG) | Cystic fibrosis | DNA | CFTR | Human/aerosolization | [100] |
| DC-Chol:DOPE | Cystic fibrosis | DNA | CFTR | Human/direct instillation | [101] |
| DOTAP | Cystic fibrosis | DNA | CFTR | Human/aerosolization | [102] |
| DOTMA:DOPE | Anti-inflammatory | siRNA | $\alpha_1$-Antitrypsin | Human/direct application | [103] |
| Oligofectamine | Viral infection | siRNA | Nucleoprotein acidic polymerase | Mice/intranasal | [104] |

**TABLE 7. Matrix Delivery *in vivo* Studies Using Nonviral Vectors to Deliver Therapeutically Relevant Genes**

| Material/Matrix | Nucleic Acid | Result | Gene | Species/Delivery | Ref. |
|---|---|---|---|---|---|
| Collagen | DNA | Bone formation | BMP-4/hPTH1-34 | Rat/femur | [105] |
| Collagen | DNA | Bone formation | hPTH1-34 | Canine/bone | [106] |
| Atelocollagen | DNA | Increase platelet number | FGF-4 | Mice/inttramuscular | [107] |
| Atelocollagen | siRNA | Tumor growth suppression | FGF-4 | Mice/intratumorally | [108] |
| Atelocollagen | siRNA | Tumor growth suppression | VEGF | Mice/intratumorally | [109] |
| Collagen | DNA | Granulation tissue and ephithelization | PDGF | Rabbit ear | [110] |
| Collagen | DNA/PLL | Survival of axotomized RGCs | FGF2, BDNF, NT-3 | Rat optic nerve | [111] |
| PLG | DNA | Small blood vessel formation | Endothelial locus-1 | Mouse subdermal | [112] |
| PLG | DNA | Enhanced matrix deposition and vascularization | PDGF | Rat/subdermal | [113] |
| PLGA | DNA | Bone formation | BMP-4 | Rat cranium | [114] |
| PLG | DNA | Increase blood vessel density | VEGF | Mice/subcutaneous | [115] |
| Gelatin | DNA | Decreased blood urea nitrogen | MMP-1 | Mouse/renal subcapsule | [116] |
| Gelatin | DNA | Angiogenesis | FGF-4 | Rabbit/hindlimb | [117] |
| EVAc | DNA | Induction of specific IgA | LDH-C4 | Mice/intravaginal | [118] |
| Fibrin | DNA/PLL | Mature blood vessel formation | pSG5-$VEGF_{165}$ | Rabbit/hindlimb schemia model | [119] |
| Fibrin | DNA/PLL | Mature blood vessel formation | HIF-1$\alpha\Delta$ODD | Mice/back wound | [120] |

[128]. The stability and size of the polyplexes formed between cationic polymers and DNA have also been correlated to primary amine content [10, 129]. Low branched PEI (low density of primary amines) requires higher N/P rations to complete condensation compared with highly branched derivatives (high density of primary amines) [130]. Last, complex size tends to decrease with an increasing polymer-to-DNA ratio [10, 121, 131].

*Polyplexes as Cellular Transfection Agents.* In addition to reducing the size of DNA and protecting the DNA from degradation by serum nucleases, binding of polycations to DNA typically results in polyplexes having a net positive charge. This positive charge has been found to be essential for efficient transfection [14, 17, 130]. It is thought that the positively charged polyplexes electrostatically interact with the negatively charged proteoglycans of the cell membrane [132], resulting in nonspecific adsorptive endocytosis (Figure 1). The polyplexes, which are now trapped inside the endosomes, need to escape before endosome–lysosome fusion occurs, which would result in polyplex degradation inside the lysosomes (Section 4.2). Cationic polymers such as PEI have the advantage over other cationic polymers [e.g., poly(amino acids)] in that they do not require an endosomal lysotropic agent to be added to achieve efficient transfection [17, 18]. Once released from the endosome the polyplexes need to transport the DNA to the nucleus where it can be transcribed or siRNA or ON to the cytosol, where they can mediate gene downregulation. The transition from the cytosol to the nucleus is not well understood, and it depends on the physical properties of the polymer and the cell cycle and degree of mitotic activity of the cells. To aid the transition from the cytosol to the nucleus cationic polymers have been modified with nuclear localization sequences typically derived from viruses (Section 4.2). Full unpackaging or decomplexation of the siRNA and ON in the cytosol is critical for their action. Strategies to enhance decomplexation involve the introduction of environmentally sensitive bonds [44], reducing the affinity of the cationic polymer for the DNA [70–73] and the delivery of enzymes that can degrade the cationic polymer [47] (Table 4 in Section 7.4.3).

Although positively charged polyplexes promote efficient transfection *in vitro*, they lead to aggregation and low transfection *in vivo* with most polyplexes accumulated in the liver after systemic administration [133]. The aggregation of polyplexes has been attributed to the presence of negatively charged proteins in serum, which can mediate bridging of multiple polyplexes together [8]. Strategies to generate neutral complexes, which do not aggregate or interact with serum components, have been developed and are discussed in Section 4.1.

*Poly(ethylenimine).* In recent years, poly(ethylenimine) (Table 1, PEI) has emerged as a widely used cationic polymer to mediate gene transfer *in vivo* and *in vitro* (for review, see Ref. 10), resulting in two commercially available transfection products ExGene (Fermentas, Inc., Hanover, MD) and jetPEI (Polyplus-transfection, San Marcos, CA). PEI has a protonatable hydrogen every two carbons in its structure, making it the polycation with the highest density of protonable amines [17]. Branched PEI is synthesized using the acid catalyzed ring opening polymerization of aziridine, resulting in a theoretical 1:2:1 ratio of primary to secondary to tertiary amines. However, C-13 nuclear resonance spectroscopy

measurements have shown that the degree of branching is closer to 1:1:1 for most commercially available PEIs [134], which indicates a higher degree of branching. Linear PEI, on the other hand, is synthesized by ring opening polymerization of 2-ethyl-2-oxazoline followed by acid hydrolysis [135]. Linear and branched PEI can be purchased in a wide variety of molecular weights ranging from less than 1000 Da to $1.6 \times 10^3$ kDa; however, the range most typically used for gene delivery is 5 to 25 kDa, mostly due to toxicity caused by higher molecular weight PEIs (Section 4.2). PEI forms complexes with DNA ranging from 77 to 300 nm in size at N/P ratios ranging from 6 to 10 [10]. High-molecular-weight PEI forms smaller, more stable particles and achieves higher transfection efficiencies than low-molecular-weight PEI; however, it is more toxic [128]. Similarly, highly branched PEI forms smaller particles that are more stable than low-branched PEI and achieves higher transfection efficiencies, but again, it is more toxic than low-branched PEI.

PEI has be used extensively *in vivo* to deliver DNA and siRNA locally to the brain [14, 136] cornea [137], tumors [82, 84], and vasculature [85] and systemically to the lung [80, 83] and tumors [81]. The charge of the DNA/PEI complexes delivered *in vivo* have a profound effect on their organ targeting. Delivery of positively charged complexes in the tail vein in mice resulted in gene transfer primarily in the lungs, whereas surface shielding transferring-PEI–DNA complexes resulted in preferential gene delivery to distantly growing tumors [138]. Tables 5, 6 (systemic), and 7 (matrix) summarize *in vivo* studies involving nonviral vectors and their therapeutic outcome.

*Poly(amino acids).* Cationic poly(amino acids) (peptide, Table 1) are some of the most commonly used cationic polymers for gene delivery. Cationic peptides such as polylysine (Table 1, PLL) are commercially available in a variety of molecular weights. However, commercially available PLLs are highly polydisperse. To synthesize monodisperse peptides, solid phase synthesis is used that involves the immobilization of the growing peptide on a solid support and a series of protecting/deprotecting synthetic steps (e.g., Fmoc chemistry). The sequential addition of each amino acid provides peptides an absolute level of control over the sequence of the growing peptide, which allows for the specific attachment of targeting ligands or other desired modifications anywhere along the molecule.

PLL and its derivatives are the most commonly used cationic peptides for gene delivery, typically used at charge ratios (+/−) ranging from 3:1 to 6:1. As increasing amounts of PLL are added to DNA, the structure changes from circular to thick, flattened to compact, and finally to toroids and rods at a charge ratio of 6:1 [122]. The diameter and cross section of the toroids are approximately 140 nm and 44 nm, respectively [122]. As with other cationic polymers, the ideal length of the PLL represents a balance between two competing effects: effective condensation and cytotoxicity. Relative to low molecular weight, the high-molecular-weight PLL forms tighter, smaller condensates that are more resistant to the effects of salt concentration and sonication [139]. Cell transfection by PLL is typically lower than other cationic polymers that possess a buffering capacity such as poly(ethylene imine) and to cationic lipids. Extensive modifications aiming at improving the transfection efficiency of PLL have been performed, including the partial modification of the PLL side chains with histidine and imidazole groups, to enhance endosomal buffering and escape (Section 4.2) [65, 140], the addition of PEG to enhance

polyplex stability, the addition of targeting ligands to enhance internalization and targeting, and the addition of covalent bonds within the polyplex to enhance its stability and polyplex unpackaging (Section 4). Furthermore, peptides containing multiple lysines have been synthesized such as Cys–His–$(Lys)_6$–His–Cys, which enhances endosomal escape, while providing covalent stabilization of the polyplexes via disulfide bonds [44].

PLL and its derivatives have not been as extensively used *in vivo* as other cationic polymers to deliver therapeutically relevant genes; however, they have found some success in systemic delivery to the liver [141] and matrix-based delivery to enhance angiogenesis [120]. Tables 5, 6 (systemic), and 7 (matrix) summarize the *in vivo* studies involving nonviral vectors and their therapeutic outcome.

*Cationic Dendrimers.* Dendrimers are polymers that branch out from a multifunctional core in a symmetric fashion. After each subsequent monomer addition (a generation), the number of attachment points increases symmetrically generating more branches. Cationic dendrimers have an attractive architecture for gene transfer because their well-defined structure and robust chemistry enables the synthesis of many generations of protonatable amines (for review of dendrimers as gene delivery vectors, see 142). Dendrimers such as poly(amidoamine) dendrimers (PAMAM, Table 1) and poly(propylenimine) (PPI, Table 1) have been shown to be capable of mediating gene delivery *in vitro* and *in vivo* [18, 131, 143, 144]. PAMAM is commercially available and is synthesized from either an ammonia or ethylenediamine core by successive addition of methyl acrylate and ethylenediamine [142]. PPI is also commercially available and is synthesized from a butylenediamine core by successive addition of acrylonitrile to a primary amino group followed by hydrogenation of nitrile groups to primary amino groups. The surface charge and diameter of the dendrimers is determined by the number of synthetic steps (i.e., number of generations) [142]. The generation of PAMAM used to complex DNA determines particle size and transfection efficiency. Complexation of DNA with fifth-generation (G5) PAMAM dendrimers produces monodisperse condensates with a radius below 200 nm [18]. At charge ratios (+/−) above or equal to one, no free DNA is observed [18]. Although PAMAM dendrimers of generations 3 (G3) to 10 (G10) can form stable complexes with DNA, the higher generations of dendrimers (G5 to G10) can transfect cells at higher efficiencies [18, 145]. Transfection efficiency of PAMAM is enhanced by partial degradation of the dendrimer structure, forming a less homogeneous structure with a higher density of primary amines [144]. The partially degraded PAMAM dendrimer has resulted in the transfection reagent Superfect. Cationic dendrimers have been used *in vivo* to deliver nucleic acids locally to the cornea [146], tumors [147], heart [148], and lung [145] and systemically to the lung [149], liver [16, 131], spleen [149], and tumors [16]. Tables 5, 6 (systemic), and 7 (matrix) summarize the *in vivo* studies involving nonviral vectors and their therapeutic outcome.

***Polysaccharide Containing Polymers.*** Chitosan (Table 1) is composed of 2-amino-2-deoxy β-D-glucan and is prepared from naturally occurring chitin via alkaline deacetylation. Unlike other cationic polymers chitosan is nontoxic and biodegradable, making it an ideal candidate for therapeutic applications. Although the density of positive charges of chitosan is lower than for other cationic polymers,

it has been shown to form particles with DNA, which protect the DNA from degradation [21, 150, 151]. Chitosan/DNA complexes have been found to be in the order of 100 nm with a spherical shape [21, 151]. Like other cationic polymers, chitosan has been chemically modified with a variety of ligands or other functionalities to enhance its transfection efficiency [21]. Chitosan has been recently used to deliver DNA mainly intranasally to the lungs for genetic immunization applications. After intranasal administration, DNA/chitosan complexes encoding for RSV antigen could attenuate pulmonary inflammation and reduce viral titer and viral antigen load [86].

Recently dextran grafted with spermine (Table 1, D-SPM) has been investigated to mediate DNA delivery. D-SPM are synthesized via reductive amination and can form aggregates with plasmid DNA and mediate efficient gene transfer *in vitro* and *in vivo* [22, 152]. Interestingly, the addition of hydrophobic domains, N-oleyl (ODS), to D-SPM (D-SPM–ODS) enhanced the stability of complexes in serum containing media. Furthermore, the addition of ODS groups achieved high transfection in the presence of serum similar to that achieved with linear PEI [153]. These results suggest that the hydrophobic nature of the complexes may be another strategy to improve transfection in the presence of serum. However, the addition of the ODS groups also increased the toxicity of the D-SPM–ODS when compared with D-SPM, but the toxicity was similar to linear PEI and branched PEI [153].

Cyclodextrins (Table 1, CDs) are cyclic, cup-shaped molecules with a hydrophilic exterior and a hydrophobic interior, which allows them to form inclusion complexes with hydrophobic molecules. Inclusion complexes are bimolecular complexes in which the "host" forms a cavity into which the "guest" molecule binds through noncovalent interactions. CDs are composed of six, seven, or eight glucose units, termed α, β, and γ, respectively; are water soluble; and are U.S. Food and Drug Administration (FDA) approved as solublizing agents for hydrophobic drugs. Davis et al. have investigated cationic molecules containing CDs as gene delivery vectors since 1999 [19]. Cationic CDs could form stable complexes with DNA with sizes in the 100-nm range. More recently CD–PAMAM [149, 154, 155] and CD–PEI [156, 157] have been investigated and were found to efficiently transfect cells and be less toxic than their parent molecules. Furthermore, both CD–PAMAM and CD–PEI have been successfully used *in vivo* [155, 156]. Furthermore, cyclodextrin containing polymers have been recently used to deliver siRNAs to inhibit tumor growth [87]. As mentioned for other cationic polymers, modifications to include targeting functionality are often desirable of a gene delivery formulation. A key advantage of cyclodextrin containing gene delivery formulations is that functionality can be added through inclusion complexes without covalent modifications. Therefore, targeting molecules, which contain a hydrophobic molecule bound to them, can be "tethered" to the polyplex via cyclodextrin. This approach has been used to introduced PEG [158], galactose [158], transferring [88, 159], and insulin [157] to cationic cyclodextrin formulations.

***Cationic Lipids.*** One of the most investigated approaches for condensing nonviral DNA for efficient gene transfer is the use of cationic lipids. Felgner et al. [23] and Bennett et al. for ON [160] used cationic lipids for the first time to deliver DNA and ONs. More recently, siRNA has been delivered to mammalian cells using commercially available lipids [161–163]. The structures of several frequently used

cationic lipids are shown in Table 2. Although some cationic lipids are used individually to deliver nonviral DNA (e.g., DOTAP), many formulations of cationic lipids also contain a zwitterionic or neutral colipid, such as DOPE (Table 2) or cholesterol, to enhance transfection. Formulations of cationic lipids have been widely applied for *in vitro* nucleic acid transfection, and more than 30 products are commercially available for this purpose, including Lipofectin (a 1:1 mixture of DOTMA and DOPE), Transfectam, Lipofectase, Lipofect-AMINE, and LipoTaxi [164].

The main components of a cationic lipid are a hydrophilic lipid anchor, a linker group, and a positively charged head group. The lipid anchor is typically either a fatty chain (e.g., derived from oleic or myristic acid) or a cholesterol group, which determines the physical properties of the lipid bilayer, such as flexibility and the rate of lipid exchange [24]. The linker group is an important determinant of the chemical stability, biodegradability, and transfection efficiency of the cationic lipid. Biodegradable lipids are being developed, which can be metabolized by various enzymes (e.g., esterases and peptidases) to minimize toxicity [165, 166]. The linker can also provide sites for the introduction of novel side chains to enhance targeting, uptake, and trafficking. The positively charged head group on the cationic lipid self-assembles with the negatively charged DNA and is a critical determinant of the transfection and cytotoxic properties of liposome formulations. The head groups differ markedly in structure and may be single- or multiple-charged as primary, secondary, tertiary, and/or quaternary amines. The synthesis of novel cationic lipids from libraries of building blocks has provided insight into some structure-activity relationships [167–169]. The hydrophobicity of the lipid moiety has a crucial effect on *in vitro* gene transfer. Multivalent head groups, such as spermine, in a "T-shape" configuration tend to be more effective than their monovalent counterparts at facilitating gene transfer [24, 168, 170]. Generally, increases in the linker length correspond to increases in the gene delivery activity [169]. Continued progress toward a comprehensive relationship among lipid structure, complexation with DNA, and subsequent interaction with the biological environment (e.g., cell membrane and extracellular membrane components) is necessary to facilitate the design of cationic lipids with optimal properties.

Mixing of DNA and cationic lipid results in the collapse of DNA to form a condensed structure (lipoplex), in which nucleic acids are buried within the lipid. The thermodynamic driving force for association of the DNA and lipid is the entropy increase from the release of counter-ions and bound water associated with DNA and the lipid surface [29, 171]. Liposome association with DNA has resulted in tube-like bilayers [28, 172], multilameller complexes [27, 29, 173, 174], as well as structures containing non-bilayer elements [26]. Multilamellar organization, in which DNA is intercalated between cationic lipid bilayers, has been reported for lipolexes resulting from the interaction of DNA with unilameller vesicles of a cationic lipid, such as DOTAP [29] and DODAB [27], respectively, combined with DOPE and cholesterol. Both of these studies used x-ray scattering to show that a periodicity of 6.5 nm occurs in the lamellar structure. The cationic lipid bilayer thickness was 3.9 nm, and the thickness of the water layer was 2.6 nm, which is sufficient to include a hydrated DNA double helix with a total diameter of 2.5 nm. A second periodicity was observed, which was attributed to a DNA–DNA correlation, which ranged from 2.45 nm to 5.71 nm as the concentration of helper lipid was

increased. These lamellar complexes can be converted into hexagonal complexes with DNA confined in inverted lipid micelles by changing the membrane spontaneous curvature or the membrane flexibility [175]. This hexagonal arrangement has demonstrated increased transfection, which is attributed to the relative instability of the complexes due to rapid fusion with anionic vesicles and subsequent DNA release. X-ray diffraction studies and microscopy have illustrated the high degree of variability in lipoplex structure and the need to understand and control the parameters that govern the organization of these structures.

The colloidal properties (e.g., size and stability) of the lipoplexes are principally determined by the cationic lipid/DNA charge ratio and not the composition of the lipid or the helper lipid [176]. The charge ratio (+/−) is typically defined as the number of amines on the cationic lipid relative to the number of phosphate groups on the DNA. A neutral charge ratio (1 : 1 charge ratio for lipid/DNA) is typically avoided because it results in the formation of large aggregates (>1 μm) [176]. Lipoplexes prepared at a positive charge ratio and a negative charge ratio likely represent structures with different lipid and DNA packaging [176]. At a positive charge ratio, large multilamellar vesicles (LMVs, diameter 300–700 nm) transfected cells more efficiently than the small unilameller vesicles (SUVs, diameter 50–200 nm) [24, 177, 178]. These observations were consistent whether the structures were formed as LMVs or as SUVs that aggregated to form LMVs [177]. The order in which DNA and lipid are mixed is critical and significantly affects the lipid and DNA packing [171, 176]. For the addition of DNA to lipid, a gradual increase in size was observed. When adding lipid to DNA, the particle size remains roughly constant until the amount of lipid positive charge exceeds the nucleic acid negative charge, whereupon the particles grow rapidly in size [171].

The net charge on the lipoplex affects its interactions with other components present *in vivo* and *in vitro* (e.g., media, serum, extracellular matrix glycoproteins, and mucosal secretions), which can limit the transfection efficiency. A positive charge ratio, which facilitates interactions with the cell membrane, is frequently used for *in vitro* studies (3 : 1), whereas *in vivo* studies may require the charge ratio to be altered because of interactions with components of the physiological environment [11]. The charge ratio of the complex determines the zeta potential, which ranges from −55 mV to +55 mV as the charge ratio is increased [179]. Multivalent anions present in the serum or media can facilitate fusion of the lipids causing an increase in the size of the particle. Polyanions with adequate anionic charge density (e.g., heparin) release DNA from the complex by binding the cationic lipid [179]. Serum can be a complicating factor for positively charged complexes, possibly causing premature release of the DNA from the complex and enhancing degradation by nucleases. For ON : lipid complexes, the various components of serum (e.g., BSA, lipoproteins, macroglobulin) interact with the complexes and alter the complex diameter, zeta potential, and interfere with cellular uptake and nuclear trafficking [180].

Aggregation of lipoplexes in polyelectrolyte solutions occurs rapidly, which can result in loss of activity in less than 24 hours; thus, strategies are being developed to stabilize the particles and prolong their shelf life. To improve lipoplex stability, PEG–PE has been incorporated into the cationic liposome. PEG containing liposomes are prevented from aggregating and interacting with serum components, which increases their stability [181–183]. Alternatively, new preparation methods

are being developed in which a detergent is present in solution with the cationic lipid and DNA [184]. Removal of the detergent by dialysis allows the formation of uniform complexes. This process yielded a lipid/DNA suspension that was able to transfect tissue culture cells up to 90 days after formation with no loss in activity. Lyophilization, which is a common approach used for many pharmaceuticals, is also being applied to lipoplexes to increase their shelf life. Cryoprotectants (e.g., sucrose and trehalose) are typically added to prevent aggregation and fusion of plasmid/lipid complexes during lyophilization. Complexes lyophilized in the presence of 0.5-M sucrose or trehalose maintained transfection rates and the sizes of rehydrated complexes as compared with nonlyophhilized controls [185].

Cationic polymers such as poly(amino acids) have also been used in combination with liposomes. DNA is initially complexed with polylysine (PLL) at low charge ratios and cationic lipids are subsequently added to completely condense the DNA. Alternatively, the PLL condensed DNA containing a net positive charge can subsequently be complexed with an anionic lipid [168]. Precondensation with polylysine has been shown to reduce serum inhibition and to enhance the transfection efficiency [186, 187].

Catinic polymer formulations have been extensively used to deliver nucleic acids *in vivo* for the treatment of cancer [89–97] and cystic fibrosis [98–102] as an anti-inflammatory agent [103] and to prevent or reduce viral infections [104]. Table 6 summarizes the use of cationic polymers to deliver nucleic acids *in vivo* and for a therapeutic purpose.

## 3 MATRIX-BASED DELIVERY

Matrix-based delivery has been proposed to enhance gene transfer *in vivo* by delaying clearance from the desired tissue, protecting the nucleic acid from degradation, and extending opportunities for internalization. Furthermore, matrix-based delivery can provide sustained delivery to maintain the vector at effective levels within the target tissue. Many of these properties have been observed for controlled release systems that deliver proteins [188]. Gene transfer from tissue engineering matrices may increase the number of cells expressing the transgene along with the extent of transgene expression, while minimizing the quantity of vector used. Additionally, matrix-based delivery may reduce the number of dosages or the required cumulative dose [188]. Matrix-based delivery of nucleic acids can be divided into two delivery approaches: (1) direct encapsulation and release, where naked DNA, polyplexes, or lipoplexes encapsulated into the matrix for later release; and (2) matrix-tethered delivery (also called substrate-mediated or solid phase delivery), where polyplexes or lipoplexes are immobilized to the matrix for release after matrix or tether degradation (Figure 2). To date most matrix-based delivery approaches involve the encapsulation of naked DNA and its subsequent release to transfect surrounding cells. Polymer encapsulation strategies can shield the vector against degradation, clearance, and an immune response. Drug release from the matrix into the tissue can be designed to occur rapidly, as in a bolus delivery, or over an extended period of time, which may affect the local concentration and cellular internalization. For rapid release, levels would be expected to quickly rise and decline as the DNA is cleared or degraded. For sustained delivery, the

**Figure 2.** Matrix-based DNA delivery can be divided into encapsulation and release approaches, where the nucleic acid is encapsulated for later release, and matrix-tethered delivery, where nucleic acid polyplexes or lipoplexes are immobilized directly to a matrix that also supports cell adhesion. These approaches are typically used for applications in tissue engineering where the delivery of nucleic acids is used to augment tissue formation.

concentration may be maintained within an appropriate range by adjusting the release rate (e.g., through the polymer choice). Release of DNA from tissue engineering scaffolds has been observed for times ranging from hours to several weeks [113, 115, 189–191]. Variations in the polymer composition and physical form, which includes properties such as porosity, mass, and size, affects the diffusion of DNA from the vehicle and hence the release rate.

Tethered delivery approaches aim to significantly slow down release kinetics by immobilizing polyplexes or lipoplexes directly to the matrix. Because the polyplexes are immobilized to the matrix, the matrix must also support cell adhesion and infiltration by the surrounding cells rather than function as a nucleic acid reservoir. Internalization of the immobilized polyplexes results after the tether between the polyplex and the matrix is degraded, the matrix is degraded or the affinity of the polyplex for the matrix is reduced due to environmental changes. The release kinetics of the polyplexes in this case depends on the chemistry used for polyplex immobilization and the chemistry of the matrix. Biotynilated polyplexes immobilized to a NeutrAvidin coated substrate was the first example of matrix-tethered delivery [192]. This matrix-tethered delivery strategies resulted in a 100-fold increase of transgene expression compared with bolus delivery, likely due to an increase of the polyplex concentration at the cellular microenvironment. Furthermore, immobilization of polyplexes to the material surface has shown to enhace the stability of the polyplexes in polyelectrolyte solutions, preventing their aggregation [193]. The strength of the tether bond or the number of tethers affects both the density of immobilized polyplexes as well as the transfection efficiency. A high density of tethers results in high surface densities of DNA with low ability to transfect, probably due to the polyplex being too tightly bound to the matrix [192]. More recently, matrix-tethered strategies rely on electrostatic interactions between poly or lipoplexes and the matrix to immobilize the complexes [194, 195].

Matrix-based delivery has been investigated mostly for regenerative medicine applications. Regenerative medicine aims to regenerate tissue by implanting biocompatible and biodegradable scaffolds at sites of injury or disease. The implanted scaffold must provide the mechanical support for the growing tissue and the

bioactive signals needed for proper tissue formation while allowing for cell attachment and growth (for reviews on tissue engineering, see Refs. 196 and 197. Delivery of nucleic acids from tissue engineering scaffolds could be used as the bioactive signals needed to recreate the environments needed for tissue formation allowing for direct manipulation of cellular gene expression patterns. In the following sections, matrix-based nucleic acid delivery will be overviewed with Section 3.1 focusing on encapsulation and release approaches and Section 3.2 on matrix-tethered delivery approaches. Table 7 summarizes *in vivo* studies using matrix-based nucleic acid delivery.

### 3.1 Direct Encapsulation and Release

Direct encapsulation and release strategies have been proposed to enhance gene transfer *in vivo* by delaying clearance from the desired tissue, protecting the DNA from degradation and extending the opportunities for internalization (Figure 2). Furthermore, matrix-based delivery can provide sustained delivery to maintain the vector at effective levels within the target tissue.

***Synthetic Polymers.*** Polymers composed of lactide and glycolide (PLGA) are perhaps the most widely used and recognized biodegradable synthetic polymers because these polymers are FDA approved and are generally considered to be biocompatible. Hydrolytic degradation of the polymer produces lactic acid and glycolic acid, two substances naturally involved in metabolic pathways of the body. The degradation rate can be controlled through the composition of the polymer and molecular weight of the chains. These polymers exhibit relatively even polymer chain scission throughout their bulk after being placed *in vivo*. These properties have led to the use of these materials for drug delivery and as scaffolds for tissue engineering [198]. Drug release from these systems typically occurs by a combination of polymer degradation and drug diffusion from the polymer. Gene delivery from polymeric matrices has been applied to tissue repair and wound healing [106, 113]. These matrices are typically implanted at a specific anatomic location where they serve multiple roles as described previously [139]. Initially, the matrix functions to create and maintain a space *in vivo*. However, the matrix also acts as a scaffold to support cell migration, proliferation, and differentiation of healthy cells from the surrounding tissue. As cells invade the matrix, they encounter DNA that is either released from or entrapped within the matrix.

A gas foaming/particulate leaching process can be employed to fabricate interconnected open pore structures of PLGA for controlled release of DNA [113, 199]. This process employs carbon dioxide to process a mixture of polymer and porogen, in order to fuse adjacent polymer particles into an interconnected structure. The DNA can be lyophilized with the microspheres [113] or encapsulated within the microspheres [200, 201]. Lyophilization of DNA with the microspheres can provide large quantities of incorporated DNA, with relatively rapid release kinetics. Incorporation of DNA into the microspheres provides for a more sustained release relative to the lyophilization method [200], with the release kinetics dependent on the polymer molecular weight and microsphere size [201]. DNA can be incorporated into polymer microspheres using several approaches [201–205]. Subcutaneous implantation of scaffolds results in transfected cells observed within the scaffold

and the tissue immediately adjacent to the scaffold, with protein production sufficient to promote physiological responses [113, 115]. An alternative approach to fabrication of PLG scaffolds for DNA delivery is electrospinning [206]. Electrospinning creates nonwoven, nanofibered membranous structures that release DNA, with maximal release occurring at approximately 2 hours.

Microspheres can also be employed to deliver DNA to tissues. Microspheres loaded with nonviral DNA can be fabricated from nondegradable and degradable polymers in sizes ranging from 0.1 to 100 μm and have been used for applications such as DNA vaccines and systemic protein delivery [198]. One of the main advantages for delivery vehicles of this size is that they can be administered in a minimally invasive manner (e.g., direct injection and oral delivery). The use of biodegradable polymers provides the additional advantage of not having to retrieve the implant after DNA release. Unfortunately, these two qualities (size and degradability) make removal of the devices difficult, should the therapy need to be terminated prematurely. The loading of DNA into the polymer ranges from 0.1- to 10-μg DNA per milligram of polymer using various techniques based on either an emulsion or phase inversion process [204, 207, 208]. A double emulsion process has been used to incorporate aqueous solutions of nonviral DNA, both supercoiled and complexed with poly-L-lysine, into polymer. Incorporation efficiencies ranging from 20% to 80% have been obtained, although the typical incorporation efficiency is approximately 30% [106]. When supercoiled DNA is incorporated using this procedure, the incorporated DNA is structurally intact with 39% of the DNA in the supercoiled conformation. Protecting the DNA by either condensation [207] or using a cryogenic approach [204] can be used to increase the DNA present in the supercoiled form to more than 80%.

DNA polyplexes have been encapsulated and released from polymer microspheres, which may enable these microspheres to be fabricated into matrices using an approach such as the gas-foaming procedure. PLL/DNA complexes have been incorporated into PLG microspheres using a double emulsion process. DNA is incorporated with efficiencies ranging from 30% to 45%, is released over approximately 35 days, and retains its integrity [207, 208]. Alternatively ONs complexed with PEI have been incorporated and released from PLG microspheres. The release profile of the ON/PEI complexes depended on the size, loading, and pore structure of the microspheres. The sustained release of ON/PEI complexes resulted in improved intracellular penetration of the delivered vector as compared with uncomplexed DNA [209, 210]. PLG/DNA scaffolds have successfully been used *in vivo* to enhance matrix deposition [113], angiogenesis [112, 113, 115], and bone formation [114].

Synthetic PEG hydrogels have been extensively investigated as cell transplantation vehicles and as tissue engineering scaffolds. Synthetic hydrogels allow for the complete engineering of the extracellular environment given that they lack direct cellular interactions and have limited protein absorption. Thus, every aspect of the extracellular matrix environment must be engineered, including integrin binding sites, growth factors, and other bioactive signals. DNA delivery from synthetic hydrogels based on PEG have shown that the release rate can be modulated depending on the degree of cross-linking and degradation rate of the hydrogel [211, 212]. Naked DNA encapsulated in photo-cross-linked PEG hydrogels could modulate the release profile of the DNA ranging from linear to a delayed release profile [211].

Naked DNA encoding for transforming growth factor beta-1 encapsulated in thermosensitive hydrogel scaffolds resulted in reepitelization of wounds in diabetic mice if applied at early stages during wound healing.

***Natural Polymers.*** Collagens are the major component of mammalian connective tissue and thus have been widely investigated as a biomaterial for cell growth and drug delivery. At least 14 types of collagen are distributed throughout the body. The most abundant is type I collagen. Type I collagen is found in high concentrations in tendon, skin, bone, and fascia, and thus, these tissues are sources for its isolation. After isolation, typically from bovine or porcine sources, collagen is preserved using techniques such as fixation with aldehydes or other chemical preservatives, gamma-irradiation, or lyophilization [213, 214]. Collagen can be lyophilized to form a spongy product or can be formed into a hydrogel by using its natural ability to self-assemble at neutral pH and body temperature (37°C) or by covalently cross-linking it using aldehydes or carbodiimides [105–107, 215].

Collagen has been used as a matrix for numerous tissue engineering applications, including bone, skin, nerve, and cartilage [216–218]. The collagen serves as a scaffold for the migration of repair cells into the matrix and serves to either retain the DNA within the scaffold [106] or provide gradual release [191]. Collagen-based release may limit vector degradation and can induce transgene expression for up to 40 days [191]. Naked DNA delivery from collagen matrices has been employed to promote tissue formation by transfecting invading fibroblasts. Transfected fibroblasts within DNA-loaded collagen scaffolds, also termed gene-activated matrices (GAMs), subsequently act as bioreactors for localized production of tissue inductive factors. Matrices were prepared by lyophilization of type I collagen and subsequent immersion in a DNA solution. For applications in tissue engineering, collagen/DNA constructs have been implanted into an adult rat femur [105] and a canine bone defect model [106]. Matrices loaded with 1 mg of DNA were capable of transfecting cells *in vivo*, which resulted in protein production for up to 3 weeks postimplant. For the canine model, however, regeneration required 100 mg of plasmid delivered from the matrix for regeneration [106]. Atelocollagen/DNA constructs have also been used for therapeutic applications, including implanted constructs in the muscle to increase platelet number [191], and in the rabbit ear to aid in would healing [110]. Furthermore, atelocollagen/siRNA constructs have been delivered intratumorally to suppress tumor growth [108, 109].

DNA complexed with cationic lipids or cationic polymers can also be incorporated and released from collagen-based matrices, while maintaining their activity. Naked DNA delivery produces substantial transgene expres-sion *in vivo*, but it results in low levels of transgene expression *in vitro*. DNA complexes can transfect cells *in vitro*; thus, the ability to release DNA complexes can extend the applicability of nonviral DNA release matrices to the engineering of tissues *in vitro*. Note, however, that release of DNA complexes may differ significantly from that for naked DNA, due to the different physical properties of complexes relative to naked DNA. Collagen matrices loaded with DNA by pipetting solutions of naked DNA, PEI/DNA complexes, and lipoplexes onto collagen showed different release kinetics [189]. *In vitro* release studies demonstrated that naked DNA was rapidly released, PEI/DNA and lipid complexes were slowly released, and that PEI/DNA complexes with a protective copolymer had intermediate release kinetics. The

PEI/DNA complexes with the protective copolymer gave the highest transfection *in vitro* and *in vivo*, with the highest *in vivo* expression occurring at 4 days and measurable quantities observed at 7 days [189]. PLL/DNA complexes encapsulated in a collagen sponge have been implanted into severed rat optic nerves as a means to promote neuron survival and promote regeneration [111]. The PLL was modified with bFGF to facilitate the internalization and intracellular trafficking. DNA was detected in the retina for up to 3 months. Nerve terminals were observed extending into the collagen, and they seemed appeared capable of transporting the DNA by retrograde transport.

Hydrogels based on agarose, fibrin, hyaluronic acid (HA), and chitosan have been employed independently as biomaterials for fabrication of tissue engineering matrices [219–221] or as materials to regulate DNA delivery. Fibrin matrixes have been extensively used as tissue engineering scaffolds, sealants, and drug delivery matrices. Fibrin sealants have been employed for the delivery of plasmids to promote angiogenesis, with fibrin-based delivery providing similar responses to delivery in PBS solution [119].

Agarose gels have been used to encapsulate PLL/DNA polyplexes and mediate sustained release to transfect smooth muscle cells *in vitro* with an efficiency less than that obtained by freshly formed complexes, but greater than that obtained with naked DNA [222]. Hydrogels based on fibrin and agarose primarily function to limit the release of DNA; however, hydrogels employing chitosan offer the potential to condense the DNA. Chitosan is a positively charged, naturally occurring polysaccharide that can form complexes with DNA. Chitosan/DNA complexes exhibit minimal cytotoxicity, destabilize the lipid bilayer to facilitate internalization [223], and produce high levels of transfection *in vitro* and *in vivo* [42, 224]. Tissue engineering matrices based on these hydrogels, or combinations of these materials, provide a variety of approaches to regulate DNA delivery *in vitro* and *in vivo*. HA-based matrices have not been used extensively for gene delivery, although their potential as tissue engineering scaffold has long been recognized. DNA loaded, cross-linked HA-based delivery systems have been formed in the form of matrices and microspheres. A solution of HA and plasmid DNA was lyophilized to form a spongy material before cross-linking with adipic acid dihydrazide to form a stable three-dimensional matrix. This HA matrix demonstrated the capacity for sustained release of naked DNA, with release likely occurring after degradation of the matrix [225]. The release rate of naked DNA from the matrices, some of which may be associated with HA fragments, could be modulated by the extent of cross-linking in the hydrogel. DNA-loaded hyaluronic acid microspheres have been formed and found to be able to deliver genes both *in vivo* and *in vitro* [226]. The microspheres were formed using a water-in-oil emulsion and covalent cross-linking with adipic acid dihydrazide to stabilize the microspheres. These microspheres could then be implanted to mediate *in vivo* gene delivery or delivered to plated cells. In both cases, the released DNA mediates transgene expression.

HA has also been cross-linked with collagen and used as gene delivery matrices. The procedure used can be similar to that used for collagen gene delivery systems or that used to form HA hydrogels. HA is mixed with collagen and lyophilized to form a spongy material. This material is then cross-linked using carbodiimide chemistry, dehydrothermal treatment, or exposure to ultraviolet light. After cross-linking, the matrices are soaked in a DNA solution [227] to introduce the DNA to

the porous scaffolds. These matrices have been shown to mediate transgene expression *in vivo* and *in vitro* [227].

### 3.2 Matrix-Tethered Delivery

Material-tethered delivery combines cationic polymer and cationic lipid formulations with matrix delivery to generate a more efficient delivery strategy by immobilizing the polyplexes or lipoplexes directly to a matrix that also supports cell adhesion and cellular infiltration (Figure 2). This approach places the nucleic acid directly at the cellular microenvironment, removing the need for diffusion of the complexes to the cell membrane Matrix tethered has not been as extensively used *in vivo* as the direct encapsulation methods described above; however, it has found success mediating DNA delivery from PLGA scaffolds and hyaluronic acid and fibrin hydrogels.

***Synthetic Polymers.*** Matrix-tethered delivery approaches in synthetic polymeric matrices have focused on nonspecific interactions (adsorption) to immobilize the complexes to the matrix. PLGA scaffolds have been coated with a variety of polyplexes such as PAMAM/DNA and PEI/DNA. PAMAM/DNA complexes were dried on a porous PLG scaffold and shown to retain their activity, being able to mediate gene transfer both *in vitro* and *in vivo*, and were found to be a function of DNA concentration and charge ratio [228]. PEI/DNA complexes were absorbed to PLG scaffolds without allowing them to dry on the surface, in contrast to the method described above. It was found that delivery of DNA via matrix-based delivery resulted in similar levels of expression compared with bolus delivery; however, a lower dose of DNA was used. The amount of DNA immobilized was dependent on the time of incubation of the polyplexes with the scaffold, the N/P ratio used, and the initial amount of DNA used [195]. Interestingly, higher amounts of immobilized DNA did not result in higher transgene expression, with 45 ng resulting in more efficient transfection than 600 ng [195].

***Natural Polymers.*** Hyaluronic acid and fibrin were employed independently as biomaterials for fabrication of tissue engineering matrices [221, 229–231] or as materials to mediate gene delivery [232, 233]. HA has been cross-linked with a wide variety of chemistries to form stable hydrogels. DNA/PEI complexes were immobilized to hyaluronic acid hydrogels via biotin–avidin bonds. PEI was modified with biotin groups and subsequently used to complex DNA to form biotynilated polyplexes. The polyplexes were then immobilized to a HA hydrogel, which had been chemically modified to introduce neutravidin, a biotin binding protein. Fibroblasts, which were directly in contact with the polyplex/HA hydrogel, were transfected; however, those directly adjacent to the hydrogel were not, indicating that direct contact of the cell with the tethered polyplexes is essential for transfection [233]. The size of the tethered polyplexes resulted in different transgene expression levels and percent of cells expressing the transgene. Small complexes (~150 nm) resulted in lower levels of transgene expression compared with large complexes (~1000 nm); however, small complexes resulted in ~50% of the cells transfected, which is double that of large complexes. Futhermore, the HA hydrogels used in these studies were topographically patterned with groves and ridges, which

resulted in spatially controlled gene expression along the ridges of the hydrogel surface [232].

Unlike HA, fibrinogen forms a stable hydrogel matrix upon thrombin activation to form a fibrin matrix. Fibrin matrices can be further stabilized by the addition of factor XIIIa, which forms covalent bonds between glutamine and lysine residues. Schense and Hubbell have used factor XIIIa chemistry extensively to bind biologically active peptides [234] and growth factors [235] to fibrin matrices. More recently, Hubbell et al. used the same strategy to immobilize polyplexes to fibrin hydrogels [120, 232]. Two different 21 amino acid peptides were designed to have a DNA binding sequences (Cys–His–(Lys)6–His–Cys) in combination with a transglutaminase substrate site (Asn–Gln–Glu–Gln–Val–Ser–Pro–Leu) or a nuclear localization site (from SV40; see Table 4). The polyplexes were formed with mixtures of the two poly(amino acids) and mixed with fibrinogen, thrombin, and factor XIIIa, resulting in the covalent immobilization of the polyplexes within the fibrin matrix. Transfection of COS-7 cells in a two-dimensional (2D) sandwich assay, in which cells are plated on top of a polyplex/fibrin hydrogel and then a second gel is cast on top, resulted in ~25% of the cells transfected [232]. These hydrogels have been used *in vivo* to deliver a mutant HIF-1α plasmid, which has the transcription factor constantly active, and resulted in enhanced wound healing in a skin wound model [120]. Furthermore, the delivery of a HIF-1α encoding plasmid was able to result in the formation of more mature blood vessels when compared with the delivery of a vascular endothelial growth factor (VEGF) encoding plasmid.

## 4 DELIVERY LIMITATIONS AND CURRENT SOLUTIONS

The ultimate applicability of the various biomaterials for gene delivery rests on their ability to effectively deliver the nucleic acids to their target intracellular location (Figure 1). An effective gene delivery system will protect the nucleic acid from degradation, target the appropriate cell population, be efficiently internalized by the cell, avoid degradative pathways, and ultimately localize to the nucleus (DNA) or cytosol (siRNA or ON, Figure 1). The materials under development can specifically enhance one or more of these steps in the transfection process. The following sections describe how the materials interact with biological systems and how they can be designed to overcome the obstacles to effective gene transfer. Table 3 also summarizes this data.

### 4.1 Extracellular Limitations

***Vector Stabilization and Fast Clearance from Body.*** Interaction of nonviral formulations with serum components results in the deactivation of the lipoplexes or polyplexes due to aggregation [9, 10]. Furthermore, aggregated lipoplexes and polyplexes can lead to rapid clearance of the polyplexes by phagocytic cells and the reticuloendotheial sytem [8]. Hydrophilic polymers such as poly(ethylene glycol) (PEG) [39, 47, 236], N-(2-hydroxypropyl)methacrylamide) (HPMA) [236–239], and oligosaccharides [40, 41, 43, 236, 240] have been found to stabilize complexes against salt and protein aggregation, which results in longer circulation times. The increased stability probably results from steric effects that (1) prevent interaction

with serum proteins, cells, and tissue; (2) increase solubility of the complexes in aqueous milieu; and (3) prevent particle–particle interactions [241]. Traditional cationic polymers for gene delivery such as PEI, PAMAM, and PLL have been made more biocompatible using this approach [18, 241–245].

***Targeted Internalization.*** The plasma membrane of the cell provides a protective shell that serves to limit the transport of undesired molecules, such as DNA. However, cells must communicate with their environment and have thus developed mechanisms to transport high- and low-molecular-weight macromolecules across the membrane. Nonclathrin-coated pit internalization can occur through smooth invaginations of 150–300 nm or via potocytosis, which involves the invagination of caveolae-rich 50–100-nm-diameter vesicles from the cell surface. DNA complexes are thought to enter cells primarily through clathrin-coated pits. These pits, with diameters of approximately 150 nm, are internalized from the plasma membrane to form coated vesicles. Interestingly, recent studies using a series of inhibitors for different endosomal pathways have elucidated different mechanisms of internalization for lipoplexes and polyplexes [246]. Lipoplex internalization and lipoplex-mediated nucleic acid delivery was strongly inhibited by potassium depletion or using chlorpromazine, indicating that clathrin endocytosis is the primary internalization pathway of lipoplexes [246]. On the contrary, polyplex internalization was inhibited by 25% when the caveolae pathway was inhibited using filipin or genistein and by 20% when clathrin-mediated endocytosis was inhibited [246], indicating that polyplexes use two internalization pathways. Gene transfer mediated by polyplexes was completely abolished when the caveolae pathway was inhibited but not when the clathrin-mediated pathway was inhibited [246], suggesting that internalization via caveolae-mediated endosomes leads to efficient transfection. Particle size also affects the endosomal internalization pathway used. Particles with a diameter <200 nm were found to be internalized via clathrin-mediated endocytosis and were ultimately delivered to the lysosome. In contrast, particles that were >500 nm entered the cells via caveolae and never reached the lysosomal compartment [247]. Furthermore, multiple studies have shown that large PEI/DNA complexes >500 nm were more efficient at mediating high transgene expression than small complexes <200 nm [233, 248], suggesting that it is due to mode of internalization and intracellular trafficking.

Receptor-mediated gene delivery offers a promising approach to create a specific interaction with the cell surface and to target a particular internalization pathway. Synthetic materials termed molecular conjugates are being developed and are composed of two domains: a DNA binding domain and a receptor binding domain. The DNA binding domain is in charge of linking the DNA to the rest of the molecular conjugate molecule. The DNA binding domain is typically composed of a cationic polymer or a cationic lipid, which self-assembles with DNA. The receptor binding domain is frequently attached to functional groups on the cationic polymer or lipid before complexation with DNA [169]; however, complexes can be initially formed and subsequently coupled to a ligand [182, 249]. The function of the receptor binding domain is to direct the DNA/molecular conjugate complex to a receptor and guide the complex through the internalization pathway that leads to endosomal entrapment. The intracellular fate of DNA complexes can depend on the type of endocytic process involved in its internalization [246]. The diversity and abundance

of membrane bound receptors and ligands and the high efficiency of internalization and turnover ratio makes receptor-mediated endocytosis a powerful tool for controlling uptake. Targets for which the attached ligand binds include the receptors for asialoglycoprotein [20, 250, 251], transferrin [47, 252–254] (review [48]), folate [45, 255, 256] (review [46]), manose [49, 50], galactose [51], lectins [257–259], integrins [55, 179] antibodies [52–54, 56] and EGF [57, 260]. The attachment of a targeting ligand to the lipoplex has been shown to increase transfection by 1000-fold as compared with lipid with no targeting ligand [261]. In principle, any monoclonal antibody, Fab fragment of a monoclonal antibody, peptide, peptide mimetic, proteins, or peptide fragment can be added to the molecular conjugate using either a covalent or an ionic attachment.

Maximal gene delivery through the receptor-mediated endocytosis pathway requires that the design parameters be optimized, which includes properties such as the ligand binding affinity, the length of the linker between the ligand and the complex, the complex charge ratio and structure, and the number of ligands per complex [260, 262]. Ligand–receptor interactions can be very specific and may be negatively affected when covalently bound to a molecular conjugate. Knowing whether the receptor binding domain is interacting to its receptor with the same efficiency as the free ligand is important to achieve high gene delivery levels. In a model using an EGF–polylysine conjugate, it was found that a longer spacer arm, between the two domains, resulted in more native-like EGF–receptor binding and higher transfection efficiency than shorter spacer arms [260]. The presence of the ligand can alter the structure of the nucleic acid complex. Incorporation of asialoorosomucoid into PLL/DNA complexes was more efficient at condensing DNA than PLL alone, probably due to an aggregation of many PLLs around the negatively charged protein [122]. At low charge ratios, specific (ligand mediated) and nonspecific endocytosis had insignificant levels of transfection. As the charge ratio was increased to 4, the transfection efficiency by specific binding enhanced transfection by a factor of 3 over nonspecific endocytosis [260]. Finally, the number of ligands per complex can affect the binding of the complexes to cells. An excessive number of ligands may inhibit the binding of complexes [260]. Optimization of the parameters that affect the surface binding of the molecular conjugates improves the specificity and gene transfer efficiency for gene delivery.

### 4.2 Intracellular Limitations

The steps after internalization of naked-DNA, polyplexes or lipoplexes, endosomal escape, and nuclear localization are thought to be rate limiting for the transfection of many cell types. Internalization of the plasmid does not necessarily correlate to transfection [263]. Much of the DNA that is internalized into the endosome is either retained or degraded within the endosome. The DNA that does escape the endosome and enters the cytoplasm must subsequently avoid degradation and be transported to the nucleus for successful gene transfer [264]. *In vitro* studies have demonstrated that although greater than 95% of the cells were positive for plasmid (>100,000 copies per cells), less than 50% of the cells were expressing the transgene [25, 265]. The following sections describe how the materials interact with intracellular biological systems and how they can be designed to overcome the obstacles for effective gene transfer. Table 4 also summarizes this data.

***Toxicity.*** One common drawback of efficient nonviral gene delivery strategies is that they are generally toxic. This finding is not surprising given that efficient nucleic acid delivery formulations are also efficient at getting into the cell and moving through the intracellular space most likely in an unspecific manner. For example, PEI has been found to colocalize with the nucleous and with the delivered DNA, whereas PLL has not [266]. As reviewed, PEI is more efficient at gene transfer than PLL and PEI is more toxic than PLL. Nuclear PEI may nonspecifically interact with genomic DNA, preventing its normal transcription, thus causing toxicity. Furthermore, the efficient endosomal buffering of PEI and PAMAM dendrimers, while improving their effectiveness to deliver DNA, may also contribute toxicity by preventing the natural acidification of endosomal vesicles and degradation of unwanted endosomal cargo. Thus, can we have an efficient nucleic acid delivery strategy without some toxicity? Is it possible to have an effective therapeutic approach with absolutely no side effects? The answer is that it needs to be attempted although the blockbuster of nonviral gene delivery formulations may not be completely nontoxic. Some general characteristics of cationic polymers and lipids make them toxic such as molecular weight, density of positive charges, and lack of degradability. This section briefly outlines the most common features associated with toxicity and their current solutions.

Nonviral formulations are generally more efficient and more toxic as the polymer-to-nucleic acid ratio is increased (higher N/P ratios), suggesting a role of free uncomplexed polymer in the solution in efficient delivery and toxicity. A recent report has shown that polyplexes of DNA and PEI contain an average of 3.5 plasmids (5800 base pairs) and 30 PEI (25 kDa) molecules when prepared at N/P ratios of 6 and 10, assuming that the DNA is completely complexed [267]. Based on these calculations, there is 86% of free PEI in the complex mixture [267, 268]. Purification of the PEI/DNA complexes by dialysis has shown a reduction of toxcicity; however, it also reduced transfection efficiency [269]. Efficient gene transfer was restored when free PEI was added to the mixture [269].

High-molecular-weight cationic polymers are generally more toxic than low-molecular-weight polymers [124], most likely due to aggregation of free polymer with the cell membrane [130] or other cellular components. Together these findings suggest that free high-molecular-weight polymer leads to efficient but toxic gene transfer. Strategies to prevent such toxicity have focused on reducing the density of positive charges (number of primary amines) without affecting the buffering capacity of the polymer and, thus, reducing its interaction with the cell membrane and other cellular components.

Another structural aspect of cationic polymers that contributes to their efficiency as gene transfer agents and their toxicity is the degree of branching and backbone flexibility. Highly branched cationic polymers form smaller particles and mediate more efficient gene transfer than low-branched polymers but are more toxic compared with low-branched or linear polymers [130]. The high toxicity of branched polymers has been associated with polymer backbone stiffness; more flexible, hyperbranched PEI derivatives with additional secondary and tertiary groups show lower toxicity *in vitro* than commercially available branched PEI [270].

One approach that could reduce the toxicity associated with a high molecular weight and a high degree of branching is making the delivery formulation

degradable. Degradable PEIs are of particular interest because the basic structure of PEI is not biologically or chemically degradable. Efforts to make PEI degradable typically involve the cross-linking of low-molecular-weight PEI (600–1800 kDa) with disulfide linkages [61] or oligo(L-lactic acid cosuccinic acid) [60] to form the higher molecular weight PEIs required to achieve efficient transfection. Furthermore, low-molecular-weight PEI has been cross-linked with PEI through a degradable ester bond [242]. Cytotoxicity experiments showed decreased toxicity or complexes formed with degradable PEG–PEI/DNA complexes when compared with complexes formed with 25-kDa PEI [242]. However, transfection studies showed that the degradable PEI was less efficient at mediating gene transfer than 25-kDa PEI [242]. Although PAMAM dendrimers have been found to be three orders of magnitude less toxic than PEI (600–1000 kDa) and PLL (36.6 kDa), they still suffer from toxic effects [124]. The toxicity of PAMAM dendrimers seems to have a different mechanism from that of toxicity caused by linear cationic polymers in that it does not cause membrane hemolysis at a similar concentration of PEI and PLL [271]. It has been reported that primary amines are more toxic than secondary or tertiary amines for PEI-type polymers [10]. However, for PAMAM dendrimers, this relationship was not true. Although cytotoxicity increased with generation (higher molecular weight), it was surprisingly independent of surface charge [272, 273]. Strategies to reduce toxicity of PAMAM dendrimers include quaternization [131], reducing or reversal of surface charge [274], and sterically shielding the surface groups by binding C12 lauroyl groups or PEG2000 [272] Cytoxicity of PLL, although low relative to many nondegradable delivery agents, is not negligible. Dendritic PLL has been found to have cytoxicity that is much lower than that of PLL and is nearly the same as naked DNA, although the PLL itself leads to cell viability of about 80% of naked DNA as determined by 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide (MTT) assay [275].

***Endosomal Escape.*** After internalization, the coated vesicle transforms into an early endosome, which is accompanied by acidification of the vesicular lumen that continues into the late endosomal and lysosomal compartments, reaching a final pH in the perinuclear lysosome of approximately 4.5. For lipoplex-mediated delivery, the interaction of the lipids with the endosomal membrane is thought to facilitate escape of the DNA to the cytoplasm before its degradation in the lysosome. Although it is not well understood, some lipoplexes preferentially fuse with the early endosome, whereas others fuse with the late endosome [24]. pH-sensitive liposomes can take advantage of the acidification process to facilitate the release of plasmids into the cytoplasm before lysosomal degradation [276]. Release of the DNA from the complex may occur at the wall of the endosome [179, 277]. Xu and Szoka proposed a model in which destabilization of the endosomal membrane causes a flip-flop of anionic lipids from the cytoplasmic facing monolayer, which diffuse laterally into the complex and form a charge neutral ion pair with the cationic lipid [179]. This pairing results in displacement of the DNA from the cationic lipid and release of the DNA into the cytoplasm. Displacement of the cationic lipid from the DNA before it enters the nucleus is critical for the ultimate expression of the gene.

Materials containing secondary and tertiary amines that are protonated at acidic pH, and not at neutral pH like primary amines, can act as proton sponges that

buffer the decrease in pH and ultimately cause DNA release into the cytoplasm. For example, PEI shows a level of protonation of 20% at pH = 7.4 compared with about 45% at pH = 5 [278]. This proton sponge effect is thought to result from the protonation of amino moieties on the polymer as the pH decreases inside the endocytic vesicle. The influx of counter $Cl^-$ ions, which occurs to maintain electroneutrality, induces osmotic swelling and rupture of the vesicle membrane [29]. Evidence for the proton sponge hypothesis include decelerated acidification of endosomal vesicles, as well as elevated chloride accumulation and a 140% increase in the relative volume in PEI-containing endosomes [279]. Modifications investigated for PLL to introduce endosomal buffering include the partial modification of the PLL side chains with imidazoles [140, 280] or histidine [65]. Both approaches have been found to enhance the release of DNA complexed with PLL from the endosome.

Endosomal disrupting peptides, also termed fusogenic peptides, can be covalently incorporated into DNA/lipid and DNA/cationic polymer complexes to enhance escape from the endosome [281, 282]. Viruses and bacteria have evolved sophisticated endosomal release pathways [283]. Based on the viral and bacterial pathways, peptides for gene delivery have been identified from the viral fusion proteins and have been used successfully to enhance endosomal escape. An extensive list of fusogenic peptides can be found in Ref. 284. The peptides range in length from 15 to 30 amino acids and form stable amphipathic α-helices primarily due to alternating hydrophilic and hydrophobic amino acids. Most peptides are amphipathic pH sensitive being only active at low pH [283]. There are two mechanisms by which fusion peptides disrupt endosomal membranes. First, the peptide causes the rearrangement of the lipid packaging, thereby changing the membrane integrity and causing release of the endosomal contents. Second, the peptide causes pore formation within the endosomal membrane, without affecting the membrane integrity [283]. Most peptides induce endosomal release by membrane disruption rather than by membrane fusion [283]. For example, GALA, which was one of the first synthetic peptides to be designed and synthesized, undergoes a conformational change to form an alpha-helix at acidic pHs [66, 67, 284]. The α-helix formed induces endosomal leakage by disrupting the endosomal membrane. A similar peptide, KALA, also functions to disrupt endosomal membranes but is positively charged and is capable of condensing DNA and mediating gene delivery [63, 64].

More recently, a family of acid responsive polymers based on α-alkyl acrylic acids such as methacrylic acid (MAA), ethylacrylic acid (EAA), propylacrylic acid, and butylacrylic acid (BAA) and their copolymers with alkyl acrylates or methacrylates have been explored for their ability to disrupt membranes and their potential as endosomal disruption polymers. The key feature of these poly(α-alkyl acrylic acids) is that they switch from a hydrophilic to a hydrophobic character as they become protonated (pH sensitive), and the switch to a hydrophobic character has been shown to disrupt membranes [68, 69]. The polymer, PPAA, has been shown to be 15 times more effective than PEAA at membrane disruption and to have maximum hemolytic activity at pH ≤ 6, which is in the range of endosomal pH [68, 285]. PPAA has been successfully been used to enhance lipolyplex-mediated gene transfer [286] *in vitro* and wound healing by altering extracellular matrix organization and greater vascularization *in vivo* [286].

***Nuclear Localization.*** The transport of DNA from the cytoplasm to the nucleus may be the most significant limitation to successful gene transfer. Plasmids injected far from the nuclei (60–90 μm) had less protein expression than plasmids injected near the nuclei [287]. In addition to cytoplasmic transport limitations, the size of DNA is problematic for crossing into the nucleus. The nuclear pores allow free diffusion entry of only small particles (less than approximately 70 kDa) [288]. Entry of the DNA can be facilitated by the breakdown of the nuclear membrane, which occurs during cell divi-sion [289]. However, for delivery to nondividing cells, nuclear localization sequences (NLSs) can be incorporated into the DNA complexes to direct the plasmid into the nucleus.

NLSs are short peptide sequences (5–25 amino acids) that are necessary and sufficient for nuclear localization of their respective proteins [288]. These sequences can be incorporated into complexes with cationic lipids [290] and cationic polymers [291, 292], or they can be directly linked to the plasmid [293, 294]. Some nuclear localization signals have stretches of positive charge, which has led to speculation that cationic polymers function as nuclear localization signals [186]. The cationic polymer PEI rapidly accumulates in the nucleus of cells and to a greater extent than PLL [266]. Nevertheless, the attachment of NLS does significantly enhance nuclear accumulation [291]. For the transfection of nondividing endothelial cells, lipoplex transfection resulted in 5% of the cells testing positive for transfection, whereas the incorporation of an NLS resulted in more than 80% of the cells testing positive [290]. Examples of nuclear localization sequences include PKKKRKVEDPY (SV40) [76], GNQSSNFGPMKGGNFGGRSSGPYGGGGQYFAKPRNQGGY (M9) [77], importin-β (1–643aa) [78], and MRRAHHRRRRASHRRMRGG (mu) [79].

***Vector Decomplexation.*** Recently, vector decomplexation has been viewed as another limiting step to efficient gene transfer [72]. Vector decomplexation involves the events that need to take place for the dissociation of the delivery formulation from the nucleic acid. This step is critical for efficient gene transfer to take place given that without it DNA cannot be transcribed and siRNA/ON cannot efficiently mediate gene downregulation. Strategies to enhance decomplexation involve the introduction of environmentally sensitive bonds (mostly disulfide bonds) [244, 295] and reducing the affinity of the cationic polymer for the DNA [70, 71, 73]. More recently, a thermosensitive copolymer, N-isopropylacrylamide-co-vinyl laurate, was covalently coupled to chitosan and employed to enhance unpacking. At temperatures below the polymer's critical solution temperature, the polymer extended and was soluble, causing the enhanced unpackaging of the chitosan/DNA polyplexes [75]. Furthermore, enzymatically degradable formulations (chitosan) have been used to enhance unpackaging and gene transfer by encouraging intracellular enzymatic degradation (via chitosinase) of the polyplexes [74].

## 5 OUTLOOK

The design and construction of ideal formulations for nonviral gene delivery will continue to focus on generating materials that are "smarter," taking advantage of the cell's natural environments and processes. One approach to the design and

synthesis of smarter nonviral gene delivery formulations is to better understand the intracellular and extracellular limitations at the molecular level and to understand the role (positive or negative) of the intracellular and extracellular molecules and environments that the lipoplexes and polyplexes come in contact with during the transfection process. The cell is a complex structure that contains a myriad of molecules and events occurring simultaneously, which result in specific cell phenotypes. How the addition of lipoplexes and polyplexes affect this harmonious environment is essential for the design and generation of nonviral formulations that are less toxic and more effective, and this design is likely to be different for different cell types. For example, recent studies have shown that the rigidity of the matrix where the cells are attached affects polyplex uptake and transgene expression, with stiffer materials resulting in enhanced uptake, unpackaging, and gene expression [296]. Although the molecular mechanism for this enhancement was not elucidated in this study, it is clear that the cellular environment (extracellular and intracellular) and cell "state" affects the effectiveness of nonviral formulations. During the past decade, nonviral gene delivery has identified key limiting steps to polyplex and lipoplex formulations and has addressed these limitations with clever solutions such as the attachment of specific cell binding ligands, fusogenic peptides, nuclear localization sequences, and degradability. The next decade will undoubtedly come with more advances for nonviral formulations that can more elegantly bypass the different barriers encountered during gene transfer.

## REFERENCES


1. Merdan T, Kopecek J, Kissel T (2002). Prospects for cationic polymers in gene and oligonucleotide therapy against cancer. *Adv. Drug Deliv. Rev.* 54:715–758.
2. Ziady A G, et al. (2003). Transfection of airway epithelium by stable PEGylated poly-L-lysine DNA nanoparticles in vivo. *Mol. Ther.* 8:936–947.
3. Ferrari S, Geddes D M, Alton E W (2002). Barriers to and new approaches for gene therapy and gene delivery in cystic fibrosis. *Adv. Drug Deliv. Rev.* 54:1373–1393.
4. Griesenbach U, et al. (2002). Gene therapy progress and prospects: Cystic fibrosis. *Gene Ther.* 9:1344–1350.
5. Onodera M, Sakiyama Y (2000). Adenosine deaminase deficiency as the first target disorder in gene therapy. *Expert Opin. Investig. Drugs.* 9:543–549.
6. Buchschacher G L, Jr, Wong-Staal F (2001). Approaches to gene therapy for human immunodeficiency virus infection. *Hum. Gene Ther.* 12:1013–1019.
7. Gansbacher B (2003). Report of a second serious adverse event in a clinical trial of gene therapy for X-linked severe combined immune deficiency (X-SCID). Position of the European Society of Gene Therapy (ESGT). *J. Gene Med.* 5:261–262.
8. Dash P R, et al. (1999). Factors affecting blood clearance and in vivo distribution of polyelectrolyte complexes for gene delivery. *Gene Ther.* 6:643–650.
9. Ledley F D (1996). Pharmaceutical approach to somatic gene therapy. *Pharm. Res.* 13:1595–1614.
10. Neu M, Fischer D, Kissel T (2005). Recent advances in rational gene transfer vector design based on poly(ethylene imine) and its derivatives. *J. Gene Med.* 7:992–1009.
11. Mahato R I, Smith L C, Rolland A (1999). Pharmaceutical perspectives of nonviral gene therapy. *Adv. Genet.* 41:95–156.

12. Cohen-Sacks H, et al. (2002). Novel PDGFbetaR antisense encapsulated in polymeric nanospheres for the treatment of restenosis. *Gene Ther.* 9:1607–1616.
13. Takakura Y, Mahato R I, Hashida M (1998). Extravasation of macromolecules. *Adv. Drug Deliv. Rev.* 34:93–108.
14. Goula D, et al. (1998). Size, diffusibility and transfection performance of linear PEI/DNA complexes in the mouse central nervous system. *Gene Ther.* 5:712–717.
15. Mahato R I (1999). Non-viral peptide-based approaches to gene delivery. *J. Drug Target.* 7:249–268.
16. Dufes C, et al. (2005). Synthetic anticancer gene medicine exploits intrinsic antitumor activity of cationic vector to cure established tumors. *Cancer Res.* 65:8079–8084.
17. Boussif O, et al. (1995). A versatile vector for gene and oligonucleotide transfer into cells in culture and in vivo: Polyethylenimine. *Proc. Natl. Acad. Sci. USA.* 92:7297–7301.
18. Haensler J, Szoka F C, Jr. (1993). Polyamidoamine cascade polymers mediate efficient transfection of cells in culture. *Bioconjug. Chem.* 4:372–379.
19. Gonzalez H, Hwang S J, Davis M E (1999). New class of polymers for the delivery of macromolecular therapeutics. *Bioconjug. Chem.* 10:1068–1074.
20. Wu G Y, Wu C H (1987). Receptor-mediated in vitro gene transformation by a soluble DNA carrier system. *J. Biol. Chem.* 262:4429–4432.
21. Mansouri S, et al. (2004). Chitosan-DNA nanoparticles as non-viral vectors in gene therapy: Strategies to improve transfection efficacy. *Eur. J. Pharm. Biopharm.* 57:1–8.
22. Hosseinkhani H, et al. (2004). Dextran-spermine polycation: An efficient nonviral vector for in vitro and in vivo gene transfection. *Gene Ther.* 11:194–203.
23. Felgner P L, et al. (1987). Lipofection: A highly efficient, lipid-mediated DNA-transfection procedure. *Proc. Natl. Acad. Sci. USA.* 84:7413–7417.
24. Felgner J H, et al. (1994). Enhanced gene delivery and mechanism studies with a novel series of cationic lipid formulations. *J. Biol. Chem.* 269:2550–2561.
25. Zabner J, et al. (1995). Cellular and molecular barriers to gene transfer by a cationic lipid. *J. Biol. Chem.* 270:18997–19007.
26. Mok K W, Cullis P R (1997). Structural and fusogenic properties of cationic liposomes in the presence of plasmid DNA. *Biophys. J.* 73:2534–2545.
27. Lasic D D, et al. (1997). The structure of DNA-liposome complexes. *J. Amer. Chem. Soc.* 119:832–833.
28. Sternberg B, Sorgi F L, Huang L (1994). New structures in complex formation between DNA and cationic liposomes visualized by freeze-fracture electron microscopy. *FEBS Lett.* 356:361–366.
29. Boussif O, et al. (1999). Synthesis of polyallylamine derivatives and their use as gene transfer vectors in vitro. *Bioconjug. Chem.* 10:877–883.
29a. Radler J O, et al. (1997). Structure of DNA-cationic liposome complexes: DNA intercalation in multilamellar membranes in distinct interhelical packing regimes. *Science.* 275:810–814.
30. Murphy J E, et al. (1998). A combinatorial approach to the discovery of efficient cationic peptoid reagents for gene delivery. *Proc. Natl. Acad. Sci. USA.* 95:1517–1522.
31. van de Wetering P, et al. (1999). Structure-activity relationships of water-soluble cationic methacrylate/methacrylamide polymers for nonviral gene delivery. *Bioconjug. Chem.* 10:589–597.
32. van de Wetering P, et al. (1998). 2-(Dimethylamino)ethyl methacrylate based (co)polymers as gene transfer agents. *J. Control. Rel.* 53:145–153.

33. Wolfert M A, et al. (1996). Characterization of vectors for gene therapy formed by self-assembly of DNA with synthetic block co-polymers. *Hum. Gene Ther.* 7:2123–2133.
34. Jon S, Anderson D G, Langer R (2003). Degradable poly(amino alcohol esters) as potential DNA vectors with low cytotoxicity. *Biomacromol.* 4:1759–1762.
35. Akinc A, et al. (2003). Synthesis of poly(beta-amino ester)s optimized for highly effective gene delivery. *Bioconjug. Chem.* 14:979–988.
36. Wen J, et al. (2004). Biodegradable polyphosphoester micelles for gene delivery. *J. Pharm. Sci.* 93:2142–2157.
37. Zhao Z, et al. (2003). Polyphosphoesters in drug and gene delivery. *Adv. Drug Deliv. Rev.* 55:483–499.
38. Wang J, et al. (2002). Enhanced gene expression in mouse muscle by sustained release of plasmid DNA using PPE-EA as a carrier. *Gene Ther.* 9:1254–1261.
39. Kim J K, et al. (2003). Enhancement of polyethylene glycol (PEG)-modified cationic liposome-mediated gene deliveries: Effects on serum stability and transfection efficiency. *J. Pharm. Pharmacol.* 55:453–460.
40. Popielarski S R, Mishra S, Davis M E (2003). Structural effects of carbohydrate-containing polycations on gene delivery. 3. Cyclodextrin type and functionalization. *Bioconjug. Chem.* 14:672–678.
41. Davis M E, et al. (2004). Self-assembling nucleic acid delivery vehicles via linear, water-soluble, cyclodextrin-containing polymers. *Curr. Med. Chem.* 11:179–197.
42. Leong K W, et al. (1998). DNA-polycation nanospheres as non-viral gene delivery vehicles. *J. Control. Rel.* 53:183–193.
43. Liu Y, et al. (2004). New poly(d-glucaramidoamine)s induce DNA nanoparticle formation and efficient gene delivery into mammalian cells. *J. Am. Chem. Soc.* 126:7422–7423.
44. McKenzie D L, et al. (2000). Low molecular weight disulfide cross-linking peptides as nonviral gene delivery carriers. *Bioconjug. Chem.* 11:901–909.
45. Gottschalk S, et al. (1994). Folate receptor mediated DNA delivery into tumor cells: potosomal disruption results in enhanced gene expression. *Gene Ther.* 1:185–191.
46. Ward C M (2000). Folate-targeted non-viral DNA vectors for cancer gene therapy. *Curr. Opin. Mol. Ther.* 2:182–187.
47. Ogris M, et al. (1999). PEGylated DNA/transferrin-PEI complexes: Reduced interaction with blood components, extended circulation in blood and potential for systemic gene delivery. *Gene Ther.* 6:595–605.
48. Li H, Qian Z M (2002). Transferrin/transferrin receptor-mediated drug delivery. *Med. Res. Rev.* 22:225–250.
49. Diebold S S, et al. (1999). Mannose polyethylenimine conjugates for targeted DNA delivery into dendritic cells. *J. Biol. Chem.* 274:19087–19094.
50. Sato A, et al. (2001). Enhanced gene transfection in macrophages using mannosylated cationic liposome-polyethylenimine-plasmid DNA complexes. *J. Drug Target.* 9:201–207.
51. Sagara K, Kim S W (2002). A new synthesis of galactose-poly(ethylene glycol)-polyethylenimine for gene delivery to hepatocytes. *J. Control. Rel.* 79:271–281.
52. Merdan T, et al. (2003). Pegylated polyethylenimine-Fab' antibody fragment conjugates for targeted gene delivery to human ovarian carcinoma cells. *Bioconjug. Chem.* 14:989–996.
53. Chiu S J, Ueno N T, Lee R J (2004). Tumor-targeted gene delivery via anti-HER2 antibody (trastuzumab, Herceptin) conjugated polyethylenimine. *J. Control. Rel.* 97:357–369.

54. Li S, et al. (2000). Targeted gene delivery to pulmonary endothelium by anti-PECAM antibody. *Am. J. Physiol. Lung Cell Mol. Physiol.* 278:L504–511.
55. Kunath K, et al. (2003). Integrin targeting using RGD-PEI conjugates for in vitro gene transfer. *J. Gene Med.* 5:588–599.
56. Buschle M, et al. (1995). Receptor-mediated gene transfer into human T lymphocytes via binding of DNA/CD3 antibody particles to the CD3 T cell receptor complex. *Hum. Gene Ther.* 6:753–761.
57. Blessing T, et al. (2001). Different strategies for formation of pegylated EGF-conjugated PEI/DNA complexes for targeted gene delivery. *Bioconjug. Chem.* 12:529–537.
58. Erbacher P, et al. (1999). Transfection and physical properties of various saccharide, poly(ethylene glycol), and antibody-derivatized polyethylenimines (PEI). *J. Gene Med.* 1:210–222.
59. Kwoh D Y, et al. (1999). Stabilization of poly-L-lysine/DNA polyplexes for in vivo gene delivery to the liver. *Biochim. Biophys. Acta.* 1444:171–190.
60. Petersen H, et al. (2002). Poly(ethylenimine-co-L-lactamide-co-succinamide): A biodegradable polyethylenimine derivative with an advantageous pH-dependent hydrolytic degradation for gene delivery. *Bioconjug. Chem.* 13:812–821.
61. Gosselin M A, Guo W, Lee R J (2001). Efficient gene transfer using reversibly cross-linked low molecular weight polyethylenimine. *Bioconjug. Chem.* 12:989–994.
62. Kichler A, et al. (1997). Influence of membrane-active peptides on lipospermine/DNA complex mediated gene transfer. *Bioconjug. Chem.* 8:213–221.
63. Wyman T B, et al. (1997). Design, synthesis, and characterization of a cationic peptide that binds to nucleic acids and permeabilizes bilayers. *Biochem.* 36:3008–3017.
64. Baru M, et al. (1998). Lysosome-disrupting peptide increases the efficiency of in-vivo gene transfer by liposome-encapsulated DNA. *J. Drug Target.* 6:191–199.
65. Pichon C, Goncalves C, Midoux P (2001). Histidine-rich peptides and polymers for nucleic acids delivery. *Adv. Drug Deliv. Rev.* 53:75–94.
66. Li W, Nicol F, Szoka F C, Jr. (2004). GALA: A designed synthetic pH-responsive amphipathic peptide with applications in drug and gene delivery. *Adv. Drug Deliv. Rev.* 56:967–985.
67. Subbarao N K, et al. (1987). pH-dependent bilayer destabilization by an amphipathic peptide. *Biochem.* 26:2964–2972.
68. Murthy N, et al. (1999). The design and synthesis of polymers for eukaryotic membrane disruption. *J. Control. Rel.* 61:137–143.
69. Thomas J L, Barton S W, Tirrell D A (1994). Membrane solubilization by a hydrophobic polyelectrolyte: surface activity and membrane binding. *Biophys. J.* 67:1101–1106.
70. Erbacher P, et al. (1995). Glycosylated polylysine/DNA complexes: gene transfer efficiency in relation with the size and the sugar substitution level of glycosylated polylysines and with the plasmid size. *Bioconjug. Chem.* 6:401–410.
71. Erbacher P, et al. (1997). The reduction of the positive charges of polylysine by partial gluconoylation increases the transfection efficiency of polylysine/DNA complexes. *Biochim. Biophys. Acta.* 1324:27–36.
72. Schaffer D V, et al. (2000). Vector unpacking as a potential barrier for receptor-mediated polyplex gene delivery. *Biotechnol. Bioeng.* 67:598–606.
73. Forrest M L, et al. (2004). Partial acetylation of polyethylenimine enhances in vitro gene delivery. *Pharm. Res.* 21:365–371.

74. Liang D C, et al. (2006). Pre-deliver chitosanase to cells: A novel strategy to improve gene expression by endocellular degradation-induced vector unpacking. *Int. J. Pharm.* 314:63–71.
75. Sun S, et al. (2005). A thermoresponsive chitosan-NIPAAm/vinyl laurate copolymer vector for gene transfection. *Bioconjug. Chem.* 16:972–980.
76. Ritter W, et al. (2003). A novel transfecting peptide comprising a tetrameric nuclear localization sequence. *J. Mol. Med.* 81:708–717.
77. Bremner K H, et al. (2004). Factors influencing the ability of nuclear localization sequence peptides to enhance nonviral gene delivery. *Bioconjug. Chem.* 15:152–161.
78. Nagasaki T, et al. (2005). Enhanced nuclear import and transfection efficiency of plasmid DNA using streptavidin-fused importin-beta. *J. Control. Rel.* 103:199–207.
79. Akita H, et al. (2006). Evaluation of the nuclear delivery and intra-nuclear transcription of plasmid DNA condensed with micro (mu) and NLS-micro by cytoplasmic and nuclear microinjection: A comparative study with poly-L-lysine. *J. Gene Med.* 8:198–206.
80. Thomas M, et al. (2005). Full deacylation of polyethylenimine dramatically boosts its gene delivery efficiency and specificity to mouse lung. *Proc. Natl. Acad. Sci. USA.* 102:5679–5684.
81. Urban-Klein B, et al. (2005). RNAi-mediated gene-targeting through systemic application of polyethylenimine (PEI)-complexed siRNA in vivo. *Gene Ther.* 12:461–466.
82. Carrere N, et al. (2005). Characterization of the bystander effect of somatostatin receptor sst2 after in vivo gene transfer into human pancreatic cancer cells. *Hum. Gene Ther.* 16:1175–1193.
83. Densmore C L (2003). Polyethyleneimine-based gene therapy by inhalation. *Expert Opin. Biol. Ther.* 3:1083–1092.
84. Fewell J G, et al. (2005). Synthesis and application of a non-viral gene delivery system for immunogene therapy of cancer. *J. Control. Rel.* 109:288–298.
85. Zaitsev S, et al. (2004). Polyelectrolyte nanoparticles mediate vascular gene delivery. *Pharm. Res.* 21:1656–1661.
86. Mohapatra S S (2003). Mucosal gene expression vaccine: a novel vaccine strategy for respiratory syncytial virus. *Pediatr. Infect. Dis. J.* 22:S100–103; discussion S103–104.
87. Hu-Lieskovan S, et al. (2005). Sequence-specific knockdown of EWS-FLI1 by targeted, nonviral delivery of small interfering RNA inhibits tumor growth in a murine model of metastatic Ewing's sarcoma. *Cancer Res.* 65:8984–8992.
88. Pun S H, et al. (2004). Targeted delivery of RNA-cleaving DNA enzyme (DNAzyme) to tumor tissue by transferrin-modified, cyclodextrin-based particles. *Cancer Biol. Ther.* 3:641–650.
89. Ramesh R, et al. (2004). Local and systemic inhibition of lung tumor growth after nanoparticle-mediated mda-7/IL-24 gene delivery. *DNA Cell Biol.* 23:850–857.
90. Hui K M, et al. (1997). Phase I study of immunotherapy of cutaneous metastases of human carcinoma using allogeneic and xenogeneic MHC DNA-liposome complexes. *Gene Ther.* 4:783–790.
91. Nabel G J, Felgner P L (1993). Direct gene transfer for immunotherapy and immunization. *Trends Biotechnol.* 11:211–215.
92. Xing X, Yujiao Chang J, Hung M (1998). Preclinical and clinical study of HER-2/neu-targeting cancer gene therapy. *Adv. Drug Deliv. Rev.* 30:219–227.
93. Hortobagyi G N, Hung M C, Lopez-Berestein G (1998). A Phase I multicenter study of E1A gene therapy for patients with metastatic breast cancer and epithelial ovarian

cancer that overexpresses HER-2/neu or epithelial ovarian cancer. *Hum. Gene Ther.* 9:1775–1798.

94. Ren H, et al. (2003). Immunogene therapy of recurrent glioblastoma multiforme with a liposomally encapsulated replication-incompetent Semliki forest virus vector carrying the human interleukin-12 gene—a phase I/II clinical protocol. *J. Neurooncol.* 64:147–154.
95. Nabel G J, et al. (1996). Immune response in human melanoma after transfer of an allogeneic class I major histocompatibility complex gene with DNA-liposome complexes. *Proc. Natl. Acad. Sci. USA*. 93:15388–15393.
96. Yoshida J, et al. (2004). Human gene therapy for malignant gliomas (glioblastoma multiforme and anaplastic astrocytoma) by in vivo transduction with human interferon beta gene using cationic liposomes. *Hum. Gene Ther.* 15:77–86.
97. Stopeck A T, et al. (2001). Phase II study of direct intralesional gene transfer of allovectin-7, an HLA-B7/beta2-microglobulin DNA-liposome complex, in patients with metastatic melanoma. *Clin. Cancer Res.* 7:2285–2291.
98. Gill D R, et al. (1997). A placebo-controlled study of liposome-mediated gene transfer to the nasal epithelium of patients with cystic fibrosis. *Gene Ther.* 4:199–209.
99. Caplen N J, et al. (1995). Liposome-mediated CFTR gene transfer to the nasal epithelium of patients with cystic fibrosis. *Nat. Med.* 1:39–46.
100. Ruiz F E, et al. (2001). A clinical inflammatory syndrome attributable to aerosolized lipid-DNA administration in cystic fibrosis. *Hum. Gene Ther.* 12:751–761.
101. Hyde S C, et al. (2000). Repeat administration of DNA/liposomes to the nasal epithelium of patients with cystic fibrosis. *Gene Ther.* 7:1156–1165.
102. Porteous D J, et al. (1997). Evidence for safety and efficacy of DOTAP cationic liposome mediated CFTR gene transfer to the nasal epithelium of patients with cystic fibrosis. *Gene Ther.* 4:210–218.
103. Brigham K L, et al. (2000). Transfection of nasal mucosa with a normal alpha1-antitrypsin gene in alpha1-antitrypsin-deficient subjects: comparison with protein therapy. *Hum. Gene Ther.* 11:1023–1032.
104. Tompkins S M, et al. (2004). Protection against lethal influenza virus challenge by RNA interference in vivo. *Proc. Natl. Acad. Sci. USA*. 101:8682–8686.
105. Fang J, et al. (1996). Stimulation of new bone formation by direct transfer of osteogenic plasmid genes. *Proc. Natl. Acad. Sci. USA*. 93:5753–5758.
106. Bonadio J, et al. (1999). Localized, direct plasmid gene delivery in vivo: prolonged therapy results in reproducible tissue regeneration. *Nat. Med.* 5:753–759.
107. Ochiya T, et al. (1999). New delivery system for plasmid DNA in vivo using atelocollagen as a carrier material: The Minipellet. *Nat. Med.* 5:707–710.
108. Minakuchi Y, et al. (2004). Atelocollagen-mediated synthetic small interfering RNA delivery for effective gene silencing in vitro and in vivo. *Nucleic Acids Res.* 32:e109.
109. Takei Y, et al. (2004). A small interfering RNA targeting vascular endothelial growth factor as cancer therapeutics. *Cancer Res.* 64:3365–3370.
110. Tyrone J W, et al. (2000). Collagen-embedded platelet-derived growth factor DNA plasmid promotes wound healing in a dermal ulcer model. *J. Surg. Res.* 93:230–236.
111. Berry M, et al. (2001). Sustained effects of gene-activated matrices after CNS injury. *Mol. Cell Neurosci.* 17:706–716.
112. Eliaz R E, Szoka F C, Jr. (2002). Robust and prolonged gene expression from injectable polymeric implants. *Gene Ther.* 9:1230–1237.
113. Shea L D, et al. (1999). DNA delivery from polymer matrices for tissue engineering. *Nat. Biotechnol.* 17:551–554.

114. Huang Y C, et al. (2005). Bone regeneration in a rat cranial defect with delivery of PEI-condensed plasmid DNA encoding for bone morphogenetic protein-4 (BMP-4). *Gene Ther.* 12:418–426.
115. Jang J H, Rives C B, Shea L D (2005). Plasmid delivery in vivo from porous tissue-engineering scaffolds: Transgene expression and cellular transfection. *Mol. Ther.* 12:475–483.
116. Aoyama T, et al. (2003). Local delivery of matrix metalloproteinase gene prevents the onset of renal sclerosis in streptozotocin-induced diabetic mice. *Tissue Eng.* 9:1289–1299.
117. Kasahara H, et al. (2003). Biodegradable gelatin hydrogel potentiates the angiogenic effect of fibroblast growth factor 4 plasmid in rabbit hindlimb ischemia. *J. Am. Coll. Cardiol.* 41:1056–1062.
118. Shen H, Goldberg E, Saltzman W M (2003). Gene expression and mucosal immune responses after vaginal DNA immunization in mice using a controlled delivery matrix. *J. Control. Rel.* 86:339–348.
119. Jozkowicz A, et al. (2003). Delivery of high dose VEGF plasmid using fibrin carrier does not influence its angiogenic potency. *Int. J. Artif. Organs.* 26:161–169.
120. Trentin D, et al. (2006). Peptide-matrix-mediated gene transfer of an oxygen-insensitive hypoxia-inducible factor-1{alpha} variant for local induction of angiogenesis. *Proc. Natl. Acad. Sci. USA.* 103:2506–2511.
121. Abdelhady H G, et al. (2003). Direct real-time molecular scale visualisation of the degradation of condensed DNA complexes exposed to DNase I. *Nucleic Acids Res.* 31:4001–4005.
122. Golan R, et al. (1999). DNA toroids: stages in condensation. *Biochem.* 38:14069–14076.
123. Dunlap D D, et al. (1997). Nanoscopic structure of DNA condensed for gene delivery. *Nucleic Acids Res.* 25:3095–3101.
124. Fischer D, et al. (2003). In vitro cytotoxicity testing of polycations: influence of polymer structure on cell viability and hemolysis. *Biomat.* 24:1121–1131.
125. Bronich T, Kabanov A V, Marky L A (2001). A thermodynamic characterization of the interaction of a cationic copolymer with DNA. *J. Phys. Chem. B.* 105:6042–6050.
126. Kabanov V A, Kabanov A V (1998). Interpolyelectrolyte and block ionomer complexes for gene delivery: physico-chemical aspects. *Adv. Drug Deliv. Rev.* 30:49–60.
127. Chen W, Turro N J, Tomalia D A (2000). Using ethidium bromide to probe the interactions between DNA and dendrimers. *Langmuir.* 16:15–19.
128. Kunath K, et al. (2003). Low-molecular-weight polyethylenimine as a non-viral vector for DNA delivery: Comparison of physicochemical properties, transfection efficiency and in vivo distribution with high-molecular-weight polyethylenimine. *J. Control. Rel.* 89:113–125.
129. Tang M X, Szoka F C (1997). The influence of polymer structure on the interactions of cationic polymers with DNA and morphology of the resulting complexes. *Gene Ther.* 4:823–832.
130. Fischer D, et al. (1999). A novel non-viral vector for DNA delivery based on low molecular weight, branched polyethylenimine: Effect of molecular weight on transfection efficiency and cytotoxicity. *Pharm. Res.* 16:1273–1279.
131. Schatzlein A G, et al. (2005). Preferential liver gene expression with polypropylenimine dendrimers. *J. Control. Rel.* 101:247–258.
132. Erbacher P, Remy J S, Behr J P (1999). Gene transfer with synthetic virus-like particles via the integrin-mediated endocytosis pathway. *Gene Ther.* 6:138–145.

133. Thomas M, Klibanov A M (2003). Non-viral gene therapy: polycation-mediated DNA delivery. *Appl. Microbiol. Biotechnol*. 62:27–34.
134. von Harpe A, et al. (2000). Characterization of commercially available and synthesized polyethylenimines for gene delivery. *J. Control. Rel.* 69:309–322.
135. Brissault B, et al. (2003). Synthesis of linear polyethylenimine derivatives for DNA transfection. *Bioconjug. Chem*. 14:581–587.
136. Abdallah B, et al. (1996). A powerful nonviral vector for in vivo gene transfer into the adult mammalian brain: polyethylenimine. *Hum. Gene Ther.* 7:1947–1954.
137. Kuo C N, et al. (2005). Dehydrated form of plasmid expressing basic fibroblast growth factor-polyethylenimine complex is a novel and accurate method for gene transfer to the cornea. *Curr. Eye Res*. 30:1015–1024.
138. Kircheis R, et al. (2001). Tumor targeting with surface-shielded ligand–polycation DNA complexes. *J. Control. Rel.* 72:165–170.
139. Adami R C, et al. (1998). Stability of peptide-condensed plasmid DNA formulations. *J. Pharm. Sci*. 87:678–683.
140. Pack D W, Putnam D, Langer R (2000). Design of imidazole-containing endosomolytic biopolymers for gene delivery. *Biotechnol. Bioeng*. 67:217–223.
141. Harada-Shiba M, et al. (2002). Polyion complex micelles as vectors in gene therapy–pharmacokinetics and in vivo gene transfer. *Gene Ther.* 9:407–414.
142. Dufes C, Uchegbu I F, Schatzlein A G (2005). Dendrimers in gene delivery. *Adv. Drug Deliv. Rev*. 57:2177–2202.
143. Qin L, et al. (1998). Efficient transfer of genes into murine cardiac grafts by Starburst polyamidoamine dendrimers. *Hum. Gene Ther.* 9:553–560.
144. Bielinska A U, Kukowska-Latallo J F, Baker J R, Jr. (1997). The interaction of plasmid DNA with polyamidoamine dendrimers: mechanism of complex formation and analysis of alterations induced in nuclease sensitivity and transcriptional activity of the complexed DNA. *Biochim. Biophys. Acta*. 1353:180–190.
145. Kukowska-Latallo J F, et al. (1996). Efficient transfer of genetic material into mammalian cells using Starburst polyamidoamine dendrimers. *Proc. Natl. Acad. Sci. USA*. 93:4897–4902.
146. Hudde T, et al. (1999). Activated polyamidoamine dendrimers, a non-viral vector for gene transfer to the corneal endothelium. *Gene Ther.* 6:939–943.
147. Maruyama-Tabata H, et al. (2000). Effective suicide gene therapy in vivo by EBV-based plasmid vector coupled with polyamidoamine dendrimer. *Gene Ther.* 7:53–60.
148. Wang Y, et al. (2001). Combination of electroporation and DNA/dendrimer complexes enhances gene transfer into murine cardiac transplants. *Am. J. Transplant*. 1:334–338.
149. Kihara F, et al. (2003). In vitro and in vivo gene transfer by an optimized alpha-cyclodextrin conjugate with polyamidoamine dendrimer. *Bioconjug. Chem*. 14:342–350.
150. Cui Z, Mumper R J (2001). Chitosan-based nanoparticles for topical genetic immunization. *J. Control. Rel.* 75:409–419.
151. Illum L, et al. (2001). Chitosan as a novel nasal delivery system for vaccines. *Adv. Drug Deliv. Rev*. 51:81–96.
152. Eliyahu H, et al. (2006). Dextran-spermine-based polyplexes–evaluation of transgene expression and of local and systemic toxicity in mice. *Biomat.* 27:1636–1645.
153. Eliyahu H, et al. (2005). Novel dextran-spermine conjugates as transfecting agents: comparing water-soluble and micellar polymers. *Gene Ther.* 12:494–503.
154. Arima H, et al. (2001). Enhancement of gene expression by polyamidoamine dendrimer conjugates with alpha-, beta-, and gamma-cyclodextrins. *Bioconjug. Chem*. 12:476–484.

155. Kihara F, et al. (2002). Effects of structure of polyamidoamine dendrimer on gene transfer efficiency of the dendrimer conjugate with alpha-cyclodextrin. *Bioconjug. Chem*. 13:1211–1219.
156. Pun S H, et al. (2004). Cyclodextrin-modified polyethylenimine polymers for gene delivery. *Bioconjug. Chem*. 15:831–840.
157. Forrest M L, Gabrielson N, Pack D W (2005). Cyclodextrin-polyethylenimine conjugates for targeted in vitro gene delivery. *Biotechnol. Bioeng*. 89:416–423.
158. Pun S H, Davis M E (2002). Development of a nonviral gene delivery vehicle for systemic application. *Bioconjug. Chem*. 13:630–639.
159. Bellocq N C, et al. (2003). Transferrin-containing, cyclodextrin polymer-based particles for tumor-targeted gene delivery. *Bioconjug. Chem*. 14:1122–1132.
160. Bennett C F, et al. (1998). Structural requirements for cationic lipid mediated phosphorothioate oligonucleotides delivery to cells in culture. *J. Drug Target*. 5:149–162.
161. Xia H, et al. (2002). siRNA-mediated gene silencing in vitro and in vivo. *Nat. Biotechnol*. 20:1006–1010.
162. Bertrand J R, et al. (2002). Comparison of antisense oligonucleotides and siRNAs in cell culture and in vivo. *Biochem. Biophys. Res. Commun*. 296:1000–1004.
163. Perelson A S, et al. (1984). Mechanism of cell-mediated cytotoxicity at the single cell level. VIII. Kinetics of lysis of target cells bound by more than one cytotoxic T lymphocyte. *J. Immunol*. 132:2190–2198.
164. Ferrari M E, et al. (1998). Analytical methods for the characterization of cationic lipid-nucleic acid complexes. *Hum. Gene Ther*. 9:341–351.
165. MacDonald R C, et al. (1999). O-ethylphosphatidylcholine: A metabolizable cationic phospholipid which is a serum-compatible DNA transfection agent. *J. Pharm. Sci*. 88:896–904.
166. Wang J, et al. (1998). Synthesis and characterization of long chain alkyl acyl carnitine esters. Potentially biodegradable cationic lipids for use in gene delivery. *J. Med. Chem*. 41:2207–2215.
167. Byk G, et al. (1998). Novel non-viral vectors for gene delivery: synthesis of a second-generation library of mono-functionalized poly-(guanidinium)amines and their introduction into cationic lipids. *Biotechnol. Bioeng*. 61:81–87.
168. Lee E R, et al. (1996). Detailed analysis of structures and formulations of cationic lipids for efficient gene transfer to the lung. *Hum. Gene Ther*. 7:1701–1717.
169. Byk G, et al. (1998). Synthesis, activity, and structure–activity relationship studies of novel cationic lipids for DNA transfer. *J. Med. Chem*. 41:229–235.
170. Falk K, et al. (2001). Reduction of experimental adhesion formation by inhibition of plasminogen activator inhibitor type 1. *Br. J. Surg*. 88:286–289.
171. Kennedy M T, et al. (2000). Factors governing the assembly of cationic phospholipid-DNA complexes. *Biophys. J*. 78:1620–1633.
172. Gershon H, et al. (1993). Mode of formation and structural features of DNA-cationic liposome complexes used for transfection. *Biochem*. 32:7143–7151.
173. Templeton N S, et al. (1997). Improved DNA: liposome complexes for increased systemic delivery and gene expression. *Nat. Biotechnol*. 15:647–652.
174. Gustafsson J, et al. (1995). Complexes between cationic liposomes and DNA visualized by cryo-TEM. *Biochim. Biophys. Acta*. 1235:305–312.
175. Koltover I, et al. (1998). An inverted hexagonal phase of cationic liposome-DNA complexes related to DNA release and delivery. *Science*. 281:78–81.
176. Xu Y, et al. (1999). Physicochemical characterization and purification of cationic lipoplexes. *Biophys. J*. 77:341–353.

177. Turek J, et al. (2000). Formulations which increase the size of lipoplexes prevent serum-associated inhibition of transfection. *J. Gene Med*. 2:32–40.
178. Ross P C, Hui S W (1999). Lipoplex size is a major determinant of in vitro lipofection efficiency. *Gene Ther*. 6:651–659.
179. Xu Y, Szoka F C, Jr. (1996). Mechanism of DNA release from cationic liposome/DNA complexes used in cell transfection. *Biochem*. 35:5616–5623.
180. Zelphati O, et al. (1998). Effect of serum components on the physico-chemical properties of cationic lipid/oligonucleotide complexes and on their interactions with cells. *Biochim. Biophys. Acta*. 1390:119–133.
181. Mok K W, Lam A M, Cullis P R (1999). Stabilized plasmid-lipid particles: Factors influencing plasmid entrapment and transfection properties. *Biochim. Biophys. Acta*. 1419:137–150.
182. Meyer O, et al. (1998). Cationic liposomes coated with polyethylene glycol as carriers for oligonucleotides. *J. Biol. Chem*. 273:15621–15627.
183. Hong K, et al. (1997). Stabilization of cationic liposome-plasmid DNA complexes by polyamines and poly(ethylene glycol)-phospholipid conjugates for efficient in vivo gene delivery. *FEBS Lett*. 400:233–237.
184. Hofland H E, Shephard L, Sullivan S M (1996). Formation of stable cationic lipid/DNA complexes for gene transfer. *Proc. Natl. Acad. Sci. USA*. 93:7305–7309.
185. Anchordoquy T J, Carpenter J F, Kroll D J (1997). Maintenance of transfection rates and physical characterization of lipid/DNA complexes after freeze-drying and rehydration. *Arch. Biochem. Biophys*. 348:199–206.
186. Vitiello L, et al. (1996). Condensation of plasmid DNA with polylysine improves liposome-mediated gene transfer into established and primary muscle cells. *Gene Ther*. 3:396–404.
187. Gao X, Huang L (1996). Potentiation of cationic liposome-mediated gene delivery by polycations. *Biochem*. 35:1027–1036.
188. Langer R (1998). Drug delivery and targeting. *Nature*. 392:5–10.
189. Scherer F, et al. (2002). Nonviral vector loaded collagen sponges for sustained gene delivery in vitro and in vivo. *J. Gene Med*. 4:634–643.
190. Cleek R L, et al. (1997). Inhibition of smooth muscle cell growth in vitro by an antisense oligodeoxynucleotide released from poly(DL-lactic-co-glycolic acid) microparticles. *J. Biomed. Mater. Res*. 35:525–530.
191. Ochiya T, et al. (2001). Biomaterials for gene delivery: Atelocollagen-mediated controlled release of molecular medicines. *Curr. Gene Ther*. 1:31–52.
192. Segura T, Shea L D (2002). Surface-tethered DNA complexes for enhanced gene delivery. *Bioconjug. Chem*. 13:621–629.
193. Segura T, Volk M J, Shea L D (2003). Substrate-mediated DNA delivery: Role of the cationic polymer structure and extent of modification. *J. Control. Rel*. 93:69–84.
194. Bengali Z, et al. (2005). Gene delivery through cell culture substrate adsorbed DNA complexes. *Biotechnol. Bioeng*. 90:290–302.
195. Jang J H, et al. (2006). Surface adsorption of DNA to tissue engineering scaffolds for efficient gene delivery. *J. Biomed. Mater. Res. A*. 77:50–58.
196. Langer R, Vacanti J P (1993). Tissue engineering. *Science*. 260:920–926.
197. Lutolf M P, Hubbell J A (2005). Synthetic biomaterials as instructive extracellular microenvironments for morphogenesis in tissue engineering. *Nat. Biotechnol*. 23:47–55.
198. Saltzman M W, Baldwin S P (1998). Materials for protein delivery in tissue engineering. *Adv. Drug Deliv. Rev*. 33:71–86.

199. Harris L D, Kim B S, Mooney D J (1998). Open pore biodegradable matrices formed with gas foaming. *J. Biomed. Mater. Res.* 42:396–402.

200. Nof M, Shea L D (2002). Drug-releasing scaffolds fabricated from drug-loaded microspheres. *J. Biomed. Mater. Res.* 59:349–356.

201. Jang J H, Shea L D (2003). Controllable delivery of non-viral DNA from porous scaffolds. *J. Control. Rel.* 86:157–168.

202. Luo D, et al. (1999). Controlled DNA delivery systems. *Pharm. Res.* 16:1300–1308.

203. Hsu Y Y, Hao T, Hedley M L (1999). Comparison of process parameters for microencapsulation of plasmid DNA in poly(D,L-lactic-co-glycolic) acid microspheres. *J. Drug Target.* 7:313–323.

204. Ando S, et al. (1999). PLGA microspheres containing plasmid DNA: Preservation of supercoiled DNA via cryopreparation and carbohydrate stabilization. *J. Pharm. Sci.* 88:126–130.

205. Hedley M L, Curley J, Urban R (1998). Microspheres containing plasmid-encoded antigens elicit cytotoxic T-cell responses. *Nat. Med.* 4:365–368.

206. Luu Y K, et al. (2003). Development of a nanostructured DNA delivery scaffold via electrospinning of PLGA and PLA-PEG block copolymers. *J. Control. Rel.* 89:341–353.

207. Capan Y, et al. (1999). Influence of formulation parameters on the characteristics of poly(D, L-lactide-co-glycolide) microspheres containing poly(L-lysine) complexed plasmid DNA. *J. Control. Rel.* 60:279–286.

208. Capan Y, et al. (1999). Preparation and characterization of poly (D,L-lactide-co-glycolide) microspheres for controlled release of poly(L-lysine) complexed plasmid DNA. *Pharm. Res.* 16:509–513.

209. De Rosa G, et al. (2002). Biodegradable microparticles for the controlled delivery of oligonucleotides. *Int. J. Pharm.* 242:225–228.

210. De Rosa G, et al. (2003). Long-term release and improved intracellular penetration of oligonucleotide-polyethylenimine complexes entrapped in biodegradable microspheres. *Biomacromol.* 4:529–536.

211. Quick D J, Anseth K S (2004). DNA delivery from photocrosslinked PEG hydrogels: Encapsulation efficiency, release profiles, and DNA quality. *J. Control. Rel.* 96:341–351.

212. Lee P Y, Li Z, Huang L (2003). Thermosensitive hydrogel as a Tgf-beta1 gene delivery vehicle enhances diabetic wound healing. *Pharm. Res.* 20:1995–2000.

213. Pachence J M (1996). Collagen-based devices for soft tissue repair. *J. Biomed. Mater. Res.* 33:35–40.

214. Schoen F J, Levy R J (1999). Tissue heart valves: Current challenges and future research perspectives. *J. Biomed. Mater. Res.* 47:439–465.

215. Pieper J S, et al. (1999). Preparation and characterization of porous crosslinked collagenous matrices containing bioavailable chondroitin sulphate. *Biomat.* 20:847–858.

216. Pieper J S, et al. (2002). Crosslinked type II collagen matrices: Preparation, characterization, and potential for cartilage engineering. *Biomat.* 23:3183–3192.

217. Verdu E, et al. (2002). Alignment of collagen and laminin-containing gels improve nerve regeneration within silicone tubes. *Restor. Neurol. Neurosci.* 20:169–179.

218. Wisser D, Steffes J (2003). Skin replacement with a collagen based dermal substitute, autologous keratinocytes and fibroblasts in burn trauma. *Burns.* 29:375–380.

219. Bellamkonda R, Ranieri J P, Aebischer P (1995). Laminin oligopeptide derivatized agarose gels allow three-dimensional neurite extension in vitro. *J. Neurosci Res.* 41:501–509.

220. Madihally S V, Matthew H W (1999). Porous chitosan scaffolds for tissue engineering. *Biomat.* 20:1133–1142.

221. Sakiyama-Elbert S E, Panitch A, Hubbell J A (2001). Development of growth factor fusion proteins for cell-triggered drug delivery. *Faseb. J.* 15:1300–1302.

222. Meilander N J, et al. (2003). Sustained release of plasmid DNA using lipid microtubules and agarose hydrogel. *J. Control. Rel.* 88:321–331.

223. Fang N, et al. (2001). Interactions of phospholipid bilayer with chitosan: effect of molecular weight and pH. *Biomacromol.* 2:1161–1168.

224. Mao H Q, et al. (2001). Chitosan-DNA nanoparticles as gene carriers: Synthesis, characterization and transfection efficiency. *J. Control. Rel.* 70:399–421.

225. Kim A, et al. (2003). Characterization of DNA-hyaluronan matrix for sustained gene transfer. *J. Control. Rel.* 90:81–95.

226. Yun Y H, et al. (2004). Hyaluronan microspheres for sustained gene delivery and site-specific targeting. *Biomat.* 25:147–157.

227. Samuel R E, et al. (2002). Delivery of plasmid DNA to articular chondrocytes via novel collagen-glycosaminoglycan matrices. *Hum. Gene Ther.* 13:791–802.

228. Bielinska A U, et al. (2000). Application of membrane-based dendrimer/DNA complexes for solid phase transfection in vitro and in vivo. *Biomat.* 21:877–887.

229. Campoccia D, et al. (1998). Semisynthetic resorbable materials from hyaluronan esterification. *Biomat.* 19:2101–2127.

230. Bulpitt P, Aeschlimann D (1999). New strategy for chemical modification of hyaluronic acid: preparation of functionalized derivatives and their use in the formation of novel biocompatible hydrogels. *J. Biomed. Mater. Res.* 47:152–169.

231. Duranti F, et al. (1998). Injectable hyaluronic acid gel for soft tissue augmentation. A clinical and histological study. *Dermatol. Surg.* 24:1317–1325.

232. Trentin D, Hubbell J, Hall H (2005). Non-viral gene delivery for local and controlled DNA release. *J. Control. Rel.* 102:263–275.

233. Segura T, Chung P H, Shea L D (2005). DNA delivery from hyaluronic acid-collagen hydrogels via a substrate-mediated approach. *Biomat.* 26:1575–1584.

234. Schense J C, Hubbell J A (1999). Cross-linking exogenous bifunctional peptides into fibrin gels with factor XIIIa. *Bioconjug. Chem.* 10:75–81.

235. Zisch A H, et al. (2001). Covalently conjugated VEGF—fibrin matrices for endothelialization. *J. Control. Rel.* 72:101–113.

236. Toncheva V, et al. (1998). Novel vectors for gene delivery formed by self-assembly of DNA with poly(L-lysine) grafted with hydrophilic polymers. *Biochim. Biophys. Acta.* 1380:354–368.

237. Howard K A, et al. (2000). Influence of hydrophilicity of cationic polymers on the biophysical properties of polyelectrolyte complexes formed by self-assembly with DNA. *Biochim. Biophys. Acta.* 1475:245–255.

238. Dash P R, et al. (2000). Decreased binding to proteins and cells of polymeric gene delivery vectors surface modified with a multivalent hydrophilic polymer and retargeting through attachment of transferrin. *J. Biol. Chem.* 275:3793–3802.

239. Oupicky D, et al. (2000). Steric stabilization of poly-L-Lysine/DNA complexes by the covalent attachment of semitelechelic poly[N-(2-hydroxypropyl)methacrylamide]. *Bioconjug. Chem.* 11:492–501.

240. Wang W, Tetley L, Uchegbu I F (2001). The level of hydrophobic substitution and the molecular weight of amphiphilic poly-L-lysine-based polymers strongly affects their assembly into polymeric bilayer vesicles. *J. Colloid Interface Sci.* 237:200–207.
241. Katayose S, Kataoka K (1997). Water-soluble polyion complex associates of DNA and poly(ethylene glycol)-poly(L-lysine) block copolymer. *Bioconjug. Chem.* 8:702–707.
242. Ahn C H, et al. (2002). Biodegradable poly(ethylenimine) for plasmid DNA delivery. *J. Control. Rel.* 80:273–282.
243. Choi Y H, et al. (1999). Characterization of a targeted gene carrier, lactose-polyethylene glycol-grafted poly-L-lysine and its complex with plasmid DNA. *Hum. Gene Ther.* 10:2657–2665.
244. Carlisle R C, et al. (2004). Polymer-coated polyethylenimine/DNA complexes designed for triggered activation by intracellular reduction. *J. Gene Med.* 6:337–344.
245. Shuai X T, et al. (2003). Novel biodegradable ternary copolymers hy-PEI-g-PCL-b-PEG: Synthesis, characterization, and potential as efficient nonviral gene delivery vectors. *Macromol.* 36:5751–5759.
246. Rejman J, Bragonzi A, Conese M (2005). Role of clathrin- and caveolae-mediated endocytosis in gene transfer mediated by lipo- and polyplexes. *Mol. Ther.* 12:468–474.
247. Rejman J, et al. (2004). Size-dependent internalization of particles via the pathways of clathrin- and caveolae-mediated endocytosis. *Biochem. J.* 377:159–169.
248. Ogris M, et al. (2001). DNA/polyethylenimine transfection particles: influence of ligands, polymer size, and PEGylation on internalization and gene expression. *AAPS PharmSci.* 3:e21.
249. Templeton N S, Lasic D D (1999). New directions in liposome gene delivery. *Mol. Biotechnol.* 11:175–180.
250. Baatz J E, et al. (1994). Utilization of modified surfactant-associated protein B for delivery of DNA to airway cells in culture. *Proc. Natl. Acad. Sci. USA.* 91:2547–2551.
251. Plank C, et al. (1992). Gene transfer into hepatocytes using asialoglycoprotein receptor mediated endocytosis of DNA complexed with an artificial tetraantennary galactose ligand. *Bioconjug. Chem.* 3:533–539.
252. Zenke M, et al. (1990). Receptor-mediated endocytosis of transferrin-polycation conjugates: an efficient way to introduce DNA into hematopoietic cells. *Proc. Natl. Acad. Sci. USA.* 87:3655–3659.
253. Wagner E, et al. (1990). Transferrin-polycation conjugates as carriers for DNA uptake into cells. *Proc. Natl. Acad. Sci. USA.* 87:3410–3414.
254. Cotten M, et al. (1990). Transferrin-polycation-mediated introduction of DNA into human leukemic cells: stimulation by agents that affect the survival of transfected DNA or modulate transferrin receptor levels. *Proc. Natl. Acad. Sci. USA.* 87:4033–4037.
255. Weitman S D, et al. (1992). Cellular localization of the folate receptor: potential role in drug toxicity and folate homeostasis. *Cancer Res.* 52:6708–6711.
256. Leamon C P, Low P S (1991). Delivery of macromolecules into living cells: a method that exploits folate receptor endocytosis. *Proc. Natl. Acad. Sci. USA.* 88:5572–5576.
257. Stewart A J, et al. (1996). Enhanced biological activity of antisense oligonucleotides complexed with glycosylated poly-L-lysine. *Mol. Pharmacol.* 50:1487–1494.
258. Erbacher P, et al. (1996). Gene transfer by DNA/glycosylated polylysine complexes into human blood monocyte-derived macrophages. *Hum. Gene Ther.* 7:721–729.

259. Midoux P, et al. (1993). Specific gene transfer mediated by lactosylated poly-L-lysine into hepatoma cells. *Nucleic Acids Res*. 21:871–878.
260. Schaffer D V, Lauffenburger D A (1998). Optimization of cell surface binding enhances efficiency and specificity of molecular conjugate gene delivery. *J. Biol. Chem*. 273:28004–28009.
261. Remy J S, et al. (1995). Targeted gene transfer into hepatoma cells with lipopolyamine-condensed DNA particles presenting galactose ligands: A stage toward artificial viruses. *Proc. Natl. Acad. Sci. USA*. 92:1744–1748.
262. Ziady A G, et al. (1999). Chain length of the polylysine in receptor-targeted gene transfer complexes affects duration of reporter gene expression both in vitro and in vivo. *J. Biol. Chem*. 274:4908–4916.
263. Reimer D L, Kong S, Bally M B (1997). Analysis of cationic liposome-mediated interactions of plasmid DNA with murine and human melanoma cells in vitro. *J. Biol. Chem*. 272:19480–19487.
264. Lechardeur D, et al. (1999). Metabolic instability of plasmid DNA in the cytosol: a potential barrier to gene transfer. *Gene Ther*. 6:482–497.
265. Tseng W C, Haselton F R, Giorgio T D (1997). Transfection by cationic liposomes using simultaneous single cell measurements of plasmid delivery and transgene expression. *J. Biol. Chem*. 272:25641–25647.
266. Godbey W T, Wu K K, Mikos A G (1999). Tracking the intracellular path of poly(ethylenimine)/DNA complexes for gene delivery. *Proc. Natl. Acad. Sci. USA*. 96:5177–5181.
267. Clamme J P, Krishnamoorthy G, Mely Y (2003). Intracellular dynamics of the gene delivery vehicle polyethylenimine during transfection: Investigation by two-photon fluorescence correlation spectroscopy. *Biochim. Biophys. Acta*. 1617:52–61.
268. Clamme J P, Azoulay J, Mely Y (2003). Monitoring of the formation and dissociation of polyethylenimine/DNA complexes by two photon fluorescence correlation spectroscopy. *Biophys. J*. 84:1960–1968.
269. Boeckle S, et al. (2004). Purification of polyethylenimine polyplexes highlights the role of free polycations in gene transfer. *J. Gene Med*. 6:1102–1111.
270. Banerjee P, et al. (2004). Novel hyperbranched dendron for gene transfer in vitro and in vivo. *Bioconjug. Chem*. 15:960–968.
271. Ruponen M, Yla-Herttuala S, Urtti A (1999). Interactions of polymeric and liposomal gene delivery systems with extracellular glycosaminoglycans: physicochemical and transfection studies. *Biochim. Biophys. Acta*. 1415:331–341.
272. Malik N, et al. (2000). Dendrimers: relationship between structure and biocompatibility in vitro, and preliminary studies on the biodistribution of 125I-labelled polyamidoamine dendrimers in vivo. *J. Control. Rel*. 65:133–148.
273. Zinselmeyer B H, et al. (2002). The lower-generation polypropylenimine dendrimers are effective gene-transfer agents. *Pharm. Res*. 19:960–967.
274. Jevprasesphant R, et al. (2003). The influence of surface modification on the cytotoxicity of PAMAM dendrimers. *Int. J. Pharm*. 252:263–266.
275. Ohsaki M, et al. (2002). In vitro gene transfection using dendritic poly(L-lysine). *Bioconjug. Chem*. 13:510–517.
276. Budker V, et al. (1996). pH-sensitive, cationic liposomes: A new synthetic virus-like vector. *Nat. Biotechnol*. 14:760–764.
277. Zelphati O, Szoka F C, Jr. (1996). Mechanism of oligonucleotide release from cationic liposomes. *Proc. Natl. Acad. Sci. USA*. 93:11493–11498.

278. SUH J, Paik H J, Hwang B K (1994). Ionization of Poly(Ethylenimine) and Poly(Allylamine) at Various Phs. *Bioorg. Chem.* 22:318–327.
279. Sonawane N D, Szoka F C, Jr, Verkman A S (2003). Chloride accumulation and swelling in endosomes enhances DNA transfer by polyamine-DNA polyplexes. *J. Biol. Chem.* 278:44826–44831.
280. Midoux P, Monsigny M (1999). Efficient gene transfer by histidylated polylysine/pDNA complexes. *Bioconjug. Chem.* 10:406–411.
281. Cristiano R J, et al. (1993). Hepatic gene therapy: efficient gene delivery and expression in primary hepatocytes utilizing a conjugated adenovirus-DNA complex. *Proc. Natl. Acad. Sci. USA.* 90:11548–11552.
282. Curiel D T, et al. (1991). Adenovirus enhancement of transferrin-polylysine-mediated gene delivery. *Proc. Natl. Acad. Sci. USA.* 88:8850–8854.
283. Plank C, Zauner W, Wagner E (1998). Application of membrane-active peptides for drug and gene delivery across cellular membranes. *Adv. Drug Deliv. Rev.* 34:21–35.
284. Kakudo T, et al. (2004). Transferrin-modified liposomes equipped with a pH-sensitive fusogenic peptide: An artificial viral-like delivery system. *Biochem.* 43:5618–5628.
285. Cheung C Y, et al. (2001). A pH-sensitive polymer that enhances cationic lipid-mediated gene transfer. *Bioconjug. Chem.* 12:906–910.
286. Kyriakides T R, et al. (2002). pH-sensitive polymers that enhance intracellular drug delivery in vivo. *J. Control. Rel.* 78:295–303.
287. Dowty M E, et al. (1995). Plasmid DNA entry into postmitotic nuclei of primary rat myotubes. *Proc. Natl. Acad. Sci. USA.* 92:4572–4576.
288. Jans D A, Chan C K, Huebner S (1998). Signals mediating nuclear targeting and their regulation: application in drug delivery. *Med. Res. Rev.* 18:189–223.
289. Tseng W C, Haselton F R, Giorgio T D (1999). Mitosis enhances transgene expression of plasmid delivered by cationic liposomes. *Biochim. Biophys. Acta.* 1445:53–64.
290. Subramanian A, Ranganathan P, Diamond S L (1999). Nuclear targeting peptide scaffolds for lipofection of nondividing mammalian cells. *Nat. Biotechnol.* 17:873–877.
291. Chan C K, Jans D A (1999). Enhancement of polylysine-mediated transferrinfection by nuclear localization sequences: Polylysine does not function as a nuclear localization sequence. *Hum. Gene Ther.* 10:1695–1702.
292. Singh D, et al. (1999). Peptide-based intracellular shuttle able to facilitate gene transfer in mammalian cells. *Bioconjug. Chem.* 10:745–754.
293. Zanta M A, Belguise-Valladier P, Behr J P (1999). Gene delivery: A single nuclear localization signal peptide is sufficient to carry DNA to the cell nucleus. *Proc. Natl. Acad. Sci. USA.* 96:91–96.
294. Sebestyen M G, et al. (1998). DNA vector chemistry: The covalent attachment of signal peptides to plasmid DNA. *Nat. Biotechnol.* 16:80–85.
295. Wetzer B, et al. (2001). Reducible cationic lipids for gene transfer. *Biochem. J.* 356:747–756.
296. Kong H J, et al. (2005). Non-viral gene delivery regulated by stiffness of cell adhesion substrates. *Nat. Mater.* 4:460–464.

# PHARMACOKINETICS OF NUCLEIC-ACID-BASED THERAPEUTICS

John C. Schmitz,[1] Aleksandra Pandyra,[2,3] James Koropatnick,[2,3,4] and Randal W. Berg[2,3,4]

[1] *VACT Healthcare System, VA Cancer Center, and Yale Cancer Center, Yale University School of Medicine, West Haven, Connecticut*
[2] *Cancer Research Laboratory Program, London Regional Cancer Program, London, Ontario, Canada*
[3] *The University of Western Ontario, London, Ontario, Canada*
[4] *Lawson Health Research Institute, London Health Sciences Centre, London, Ontario, Canada*

**Chapter Contents**

1 Introduction: Nucleic-Acid-Based Therapeutics
2 Antisense Chemistries—Modifications to DNA/RNA Structure
3 Additional Synthetic Chemistries
4 Methods for Detection of ASOs in Tissues and Fluids
  4.1 Labeled ASOs
  4.2 Nondenaturing Polyacrylamide and Capillary Gel Electrophoresis
  4.3 Hybridization-Based Approaches
5 Routes of Administration
6 PK (ADME) of ASOs Administered by Injection
  6.1 Absorption and Distribution
  6.2 Metabolism and Elimination
  6.3 PK Properties of 2′-MOE, PS, and PD ASOs
  6.4 PK Analyses from Selected Clinical Trials
  6.5 Target Downregulation in Tissues of Interest
7 RNA Interference
  7.1 siRNA Specificity
  7.2 Cell Culture
  7.3 *In vivo* Studies with "Naked" siRNAs
  7.4 Hydrodynamic High-Pressure "Naked" siRNA Injection
  7.5 Chemically Modified siRNAs

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

7.6 Ligand-Mediated siRNA Delivery
7.7 Intranasal siRNA Injection
7.8 *In vivo* Studies with Encapsulated siRNAs
7.9 Linear Polyethylenimine
7.10 Cationic Lipids
7.11 siRNAs in Clinical Trials
8 Conclusions
8.1 Nucleic-Acid-Based Therapeutics
8.2 PK (ADME) and Toxicology Differences
8.3 Relative Clinical Utility of the Different Approaches
References

## 1 INTRODUCTION: NUCLEIC-ACID-BASED THERAPEUTICS

Nucleic-acid-based therapies include various forms of gene and antisense therapeutics. Gene therapy encompasses the use of plasmid vectors, retroviral and lentiviral vectors, adeno-associated virus (AAV) and other methods of introducing a gene expression or gene silencing nucleic acid sequence into a target tissue or cell. Gene silencing has been achieved with ribozymes, small interfering double-stranded RNAs (siRNAs), and synthetic antisense oligonucleotides (ASOs), whether these are unmodified or chemically modified DNA or RNA structures (Figure 1). Additional chemically synthesized, but not natural, oligonucleotides include DNA-like peptide nucleic acids (PNAs) and phosphorodiamidate morpholino oligonucleotides (PMOs) (Figure 1). Some nucleic-acid-based therapeutics are designed to influence the splicing of pre-mRNA to modify or alter the protein produced, rather than to silence the gene expression entirely [1, 2, 3]. Many gene therapy and antisense approaches have unique applications based on local delivery, which limits the distribution of the therapeutic in the body and precludes considerations of their pharmacokinetics (PK). We have included key citations describing PK for some gene-based and oligonucleotide-based approaches in Table 1, but the remainder of the chapter is limited to ASOs and siRNAs for brevity and clarity.

Most PK studies of nucleic-acid-based therapies are, in fact, devoted to description of the PK properties of ASOs. Many academic and industry researchers have chosen to focus on the use and development of synthetic ASOs as therapeutics. RNA- and DNA-based ASOs designed against a large number of targets have progressed through preclinical stages and into clinical trials. As a consequence, there is a large body of literature describing the PK properties of these ASOs in animals and humans. We will first discuss the various chemical structures of ASOs being developed as therapeutics, and then we will focus on the published animal and human PK data regarding ASOs of two chemistries, exemplified by Genasense (Oblimersen, G3139, targeting the mRNA for the anti-apoptotic protein Bcl-2) and OGX-011 (targeting the mRNA encoding the chaperone protein clusterin).

The second major (and more recent) focus in the literature is on the PK properties of various embodiments of siRNA therapeutics. Excitement has grown recently with our increased understanding of the mechanisms responsible for the endogenous pathways of posttranscriptional gene silencing, also known as RNA inter-

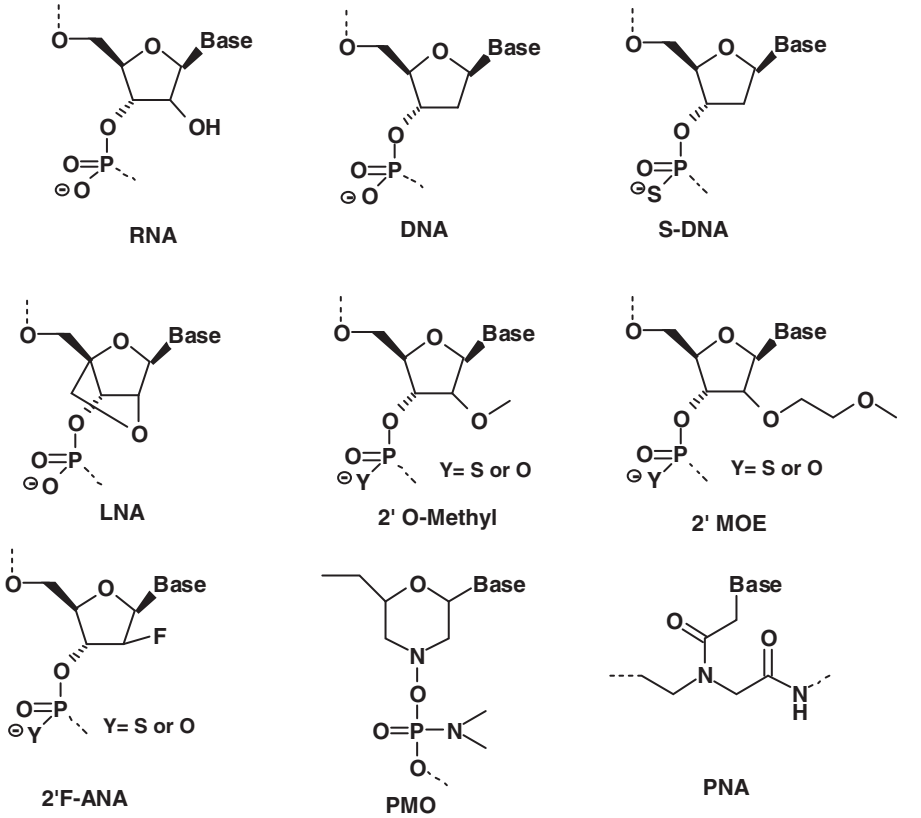


**Figure 1.** Nucleic acid analogs and artificial backbones commonly used in ASOs and siRNAs.

**TABLE 1. Selected PK Studies of Nucleic-Acid-Based Therapies**

| Category | Comments | Citations |
|---|---|---|
| Plasmid and viral | Reviews of PK and clinical trials | [4–6] |
| Ribozymes | Targeting HEP C virus in mice | [7] |
| | Review of anti-VEGF-R1 | [8] |
| | PK in monkeys | [9] |
| | Trial in healthy volunteers | [10] |
| | Trials in cancer patients | [11, 12] |
| PMO | AVI-4126, targeting c-myc | [13–15] |
| LNA | | [16–18] |
| PNA | PK in rats | [19] |
| | PK in mice | [20] |

ference (RNAi). This understanding has led to design and synthesis of expression vectors, short hairpin RNA (shRNA), and small interfering RNA (siRNA) molecules to reduce or silence expression of target genes. Although use of RNAi technology is widespread in laboratories, therapeutic applications are not yet as far advanced as for ASOs. Studies describing PK properties of siRNA in animal

models are limited, and only three siRNA molecules are currently in human clinical trials. The relevant preclinical PK studies and the ongoing clinical trials are discussed in the latter portion of the chapter.

## 2 ANTISENSE CHEMISTRIES—MODIFICATIONS TO DNA/RNA STRUCTURE

ASOs are designed to reduce expression of a target protein based on double-stranded Watson–Crick base pairing with the mRNA coding for the protein. RNA-based ASOs act by inhibiting translation of the target mRNA, and subsequent normal protein turnover (degradation) results in reduced target protein levels in cells [21, 22]. DNA-based ASOs, also known as oligodeoxynucleotides (ODNs), mediate mRNA degradation by activation of endogenous ribonuclease H and subsequent exonuclease cleavage of the target mRNA [23–26]. Again, routine protein turnover reduces target protein levels in cells in the absence of the mRNA to provide template for new protein synthesis. Advantages of ASOs over other nucleic-acid-based strategies include ease of delivery into cells and tissues, improved specificity for individual mRNA target molecules, flexibility in target choice and reagent design, and relative ease of chemical synthesis that facilitates large-scale production of pharmaceutical-grade material [22, 27–29].

Early studies with unmodified phosphodiester (PD) DNA ASOs revealed various problems, including delivery into target cells and nuclease sensitivity in body fluids and in cells. Since then, numerous chemical modifications to the DNA structure have been incorporated into ASOs (see Figure 1). One of the first improvements was inclusion of phosphorothioate internucleotide linkages (PS DNA), where the nonbridging oxygen is replaced with a sulfur atom. This improved stability against endo- and exonucleases inside cells and in body fluids [22, 30]. "Second generation" ASOs were designed to include 2′-methoxyethoxylation (2′-MOE), 2′-O-methyl, or other modifications of the ribose sugar group [31–33]. These modifications at the 2′ position enhance uptake and stability of antisense ODNs [22, 32], but they preclude RNase H activity. ASOs containing 2′ modifications at the 5 or 6 nucleotides on each end, with DNA bases in the central region (called "gapmers"), retain the ability to provide a suitable target for ribonuclease H [25, 32, 34, 35]. Several PS DNA ASOs and gapmer ASOs have progressed through preclinical development stages, including PK and toxicological studies, and many are currently in human clinical trials (see Table 2).

In terms of PK properties, ASOs of the same chemistry but with different nucleotide base sequences seem to behave identically. In contrast, differences in PK profiles depend critically on the ASO chemistry and modifications. That is, ASOs with the same base sequence but having PS versus PD, or DNA versus RNA versus 2′-MOE modifications will have slightly or vastly different PK profiles [32, 34]. Furthermore, separately from PK properties, differences in the nucleotide sequence incorporated into ASOs with identical chemistry do affect physiological responses, including toxicities and off-target effects. For example, ASOs containing CpG dinucleotide motifs often activate particular toxicities and immune responses [62–64]. Toxicological changes are also dependent on backbone and sugar modifications, such as activation of the complement cascade by ASOs with PS inter-

**TABLE 2. Selected Clinical Trials of DNA-Based ASOs**

| ASO | Target | Chemistry | Citations |
|---|---|---|---|
| G3139 | Bcl-2 | PS DNA | [36–40] |
| GTI-2040 | RNR | PS DNA | [41] |
| OL(1)p53 | p53 | PS DNA | [42, 43] |
| MG98 | DNMT1 | PS DNA | [44] |
| GEM91 | HIV gag | PS DNA | [45, 46] |
| ISIS 5132 | c-raf-1 | PS DNA | [47–49] |
| ISIS 2503 | H-ras | PS DNA | [50, 51] |
| ISIS 2302 | ICAM-1 | PS DNA | [52–54] |
| ISIS 3521 | PKC-α | PS DNA | [47, 55, 56] |
| LY900003 | PKC-α | (= ISIS 3521) | [57] |
| GEM231 | PKA | 2′-Me Gapmer | [58, 59] |
| ISIS 104838 | TNF-α | 2′-MOE Gapmer | [60] |
| OGX-011 | Clusterin | 2′-MOE Gapmer | [61] |

nucleotide linkages but not by those with PD backbones [45, 65, 66]. These toxicological issues are not discussed in detail here, but they have been reviewed recently elsewhere [62, 65, 67].

## 3 ADDITIONAL SYNTHETIC CHEMISTRIES

Several groups of researchers have been pursuing alternative synthetic chemistries to achieve the desirable properties of an ideal ASO. Some of these designs that are in more-advanced stages of development as potential drug candidates are shown in Figure 1. These include locked nucleic acids (LNAs) [16, 17, 68–71], 2′-fluoroarabinonucleic acids (FANAs) [72–74], PNAs [75–79], and PMOs [13, 14, 80–83]. Only a limited number of ASOs with these chemistries have progressed to PK studies in rodent models, primate models, or human clinical trials. Relevant PK studies regarding this group of nucleic-acid-based ASOs are given in Table 1. For brevity and due to the lack of human PK data, we will limit our discussion of these other chemistries.

## 4 METHODS FOR DETECTION OF ASOs IN TISSUES AND FLUIDS

### 4.1 Labeled ASOs

In our work [84], we have used T4 polynucleotide kinase to end-label ASOs using radioactive [$^{32}$P]-γ-adenosine triphosphate (ATP). Methods to end-label or synthesize ASOs with biotin, fluorescein, digoxigenin, and so on, are also available. ASOs are amenable to internal labeling with radioactive halogens [85]. Radioactively labeled ASOs are tracked using phospho-imaging, gel electrophoresis, and other techniques. Fluorescence microscopy is used to monitor fluorescently labeled ASOs, whereas detection of biotin-labeled ASOs is done with anti-biotin antibodies or avidin-conjugated enzymes and appropriate substrates.

### 4.2 Nondenaturing Polyacrylamide and Capillary Gel Electrophoresis

Acrylamide gels (15% to 20%) can be used effectively to separate ASOs from longer intact nucleic acids and shorter degradation products [72]. Radioactively labeled ASOs are detected using standard autoradiography or phospho-imaging techniques [35]. Using acrylamide gels cast in capillary tubes, ASOs can also be separated, identified, and quantified after electrophoresis. Absorbance at 260 nm, mass spectrometry, and so on, can be used to detect ASOs as they leave the capillary gels [35, 86–88].

### 4.3 Hybridization-Based Approaches

Methods have been developed to detect ASOs in plasma and other biological samples, based on hybridization with labeled complementary probes [89]. The probes are tagged at one end with biotin and the other end with digoxigenin. After hybridization and binding to neutravidin-coated 96-well plates, nuclease S1 is added to degrade unhybridized probe. Anti-digoxigenin antibodies and enzyme-linked secondary antibodies are then added sequentially, followed by enzyme substrate (for example, AttoPhos, which fluoresces after enzymatic cleavage) for detection and quantitation [60, 61].

We have developed a hybridization approach to monitor the intracellular uptake and distribution of ASOs in cell culture experiments [R. Berg et al., unpublished]. Cells treated with ASOs were fixed in paraformaldehyde and then biotin-labeled complementary oligonucleotides were added in hybridization buffer. Detection was with an avidin–horseradish peroxidase conjugate, which cleaves diaminobenzidine to produce a colored precipitate within the cells. An alternative detection approach uses fluorescein-labeled streptavidin, which is visualized by fluorescence and/or confocal microscopy [S. Fard et al., manuscript submitted]. We are currently adapting this protocol for use on tissue and tumor sections taken from mice treated systemically with ASOs.

A hybridization-based approach has also been developed to quantitate levels of ribozymes in serum using paired complementary oligonucleotide probes, one labeled with biotin and the other with digoxigenin [90]. The annealed "triplex" is first collected on streptavidin-coated 96-well plates, then an anti-digoxigenin antibody conjugated with alkaline phosphatase is added, followed by addition of the enzyme substrate p-nitrophenyl phosphate, which is cleaved into a soluble colored product that is quantitated by absorbance at 405 nm.

## 5 ROUTES OF ADMINISTRATION

As with any drug, the PK profile of nucleic-acid-based therapeutics varies according to the route of administration. Before focusing on the PK properties of ASOs administered intravenously, we briefly discuss the published PK analysis of topical, inhaled, and oral formulations of ASOs.

Topical administration of ASOs has been investigated for a small number of indications, and it has been the subject of several recent reviews [91–93]. To date, topical administration has not progressed past preclinical animal models, although these include human skin transplanted (grafted) onto immunocompromised mice [81, 94]. The distribution of ASOs applied to the skin is generally limited to epiderimis and

dermis, and it may be limited in intact skin but more advantageous in conditions that include inflammation or other damage, especially psoriasis [93, 95]. Distinct chemistries of ASOs, such as second-generation versus PMOs, may behave differently when applied topically. Furthermore, the formulation of the application compound [96–98] and the use of delivery methods such as sonophoresis (ultrasound) [99, 100] may also affect the delivery and PK properties of ASOs applied topically.

Respirable, or inhaled, ASOs have been developed as potential therapeutics for treatment of asthma, allergic rhinitis, and other pulmonary diseases [101–104]. In animal models including rodents and rabbits, inhaled ASOs are detectable in all cell types in the lungs but only in small amounts (<5–10% of the administered dose) in extra-pulmonary organs, and they exhibit minimal systemic bioavailability of intact, full-length ASO [103, 105, 106]. Less than 5% of the delivered dose was detected in the lung 72 hours after treatment (the elimination half-life is approximately 30 hours). At 72 hours, nearly 70% of the administered dose was excreted in the urine, the major pathway of elimination. ASOs designed and formulated for delivery by inhalation may have particular strengths as potential therapeutics for pulmonary diseases. One ASO, EPI-2010 (targeting the adenosine signaling pathway in asthma), completed phase I and II human trials, but further development was not pursued due to reduced efficacy in important patient markets [103].

Stability of unmodified DNA ASOs is limited after oral administration, due to degradation in the stomach and small intestine. Although backbone and ribose modifications substantially improved the stability of ASOs administered orally [107], intestinal permeability of ASOs is limited due to hydrophilicity, high negative charge, and high molecular mass [108, 109]. Formulation of ASOs in sodium caprate further improves stability and intestinal permeability of orally administered ASOs [108, 110]. Oral formulations have progressed to animal studies in rodents, pigs, and dogs [107–110], but not yet to human trials.

Lipid delivery vehicles such as stealth liposomes [111] and cationic liposome formulations [112–114], and additional vehicles such as polymers of chitosan [115], are being pursued as potential means to improve uptake and distribution of ASOs. Although these novel formulations and alternative routes of administration might ultimately prove to be more successful, the intravenous route has been the most successful thus far in the development of therapeutic ASOs.

## 6 PK (ADME) OF ASOs ADMINISTERED BY INJECTION

The PK properties of PS ASOs in animal models have been extensively studied and reviewed [33, 62, 116–118]. Several routes of administration, including intraperitoneal (i.p.), intravenous (i.v.), intradermal (i.d.), or subcutaneous (s.c.) injection, have been compared in animal models. PK profiles of ASOs in animals correlate well with those observed in human clinical trials, where ASOs are usually administered by i.v. or s.c. injection.

### 6.1 Absorption and Distribution

ASOs are highly bound to albumin and α2-macroglobulin in the blood, with binding affinities in the micro- to millimolar range [32, 65, 119]. In mice, rats,

monkeys, and humans, more than 96% of administered PS ASO was bound to plasma proteins [32, 119]. This high protein binding affinity limits renal clearance, but PS ASOs are rapidly cleared from plasma to tissues. Plasma clearance is dose-dependent, with plasma half-life ranging from 30 to 60 minutes [32, 119]. Widespread systemic distribution of ASOs is observed after i.d. and s.c. injections [65], but s.c. injection resulted in increased bioavailability and steady-state ASO levels [120]. Minimal differences in distribution were observed when comparing i.p and i.v. injection [116, 118]. In animal models, ASOs accumulated to highest levels in the liver and kidney, followed by spleen and lung, with only minor distribution to the brain and testes [65, 120, 121]. Distribution was independent of ASO length and nucleotide sequence [65]. Subcutaneous tumor xenografts accumulated levels of ASOs similar to spleen and lung [84]. Recent clinical trials of G3139 also measured similar disposition of intact and partially shortened ASO in humans, as discussed below [122, 363].

### 6.2 Metabolism and Elimination

Metabolism of ASOs occurs primarily through degradation by exonucleases and endonucleases. Degradation products have lower protein binding affinity and higher solubility, facilitating their excretion in urine and exhaled air [65, 121, 123–125].

### 6.3 PK Properties of 2′-MOE, PS, and PD ASOs

In a recent study the PK profiles of three distinctly modified ASOs were compared after i.v. injection into monkeys [34]. They compared a PD ASO with 2′-MOE modifications on all nucleotides, a PS ASO with 2′-MOE modifications only on the eight 3′-end residues, and a PS ASO with 2′-MOE modifications on the six nucleotides on each end (i.e., a gapmer). The plasma half-lives of the two PS ASOs were similar (~45 minutes), whereas that of the PD ASO was markedly shorter (~14 minutes). The two PS ASOs also had very similar tissue distribution, with the highest concentrations in kidney, liver, lymph nodes, and spleen. However, the gapmer exhibited a significantly longer tissue half-life, compared with the ASO with only the 3′ MOE modifications (~22 days vs. ~5 days). This suggests that the presence of both 5′-end and 3′-end MOE modifications on the gapmer protects the ASO from degradation in tissues and that the position of the modifications could significantly affect biological lifetime and efficacy in a clinical setting. This study confirmed that the chemistry of the ASO contributes to PK behavior and nuclease sensitivity, as suggested by earlier studies [119]. Although the extent of the 2′-MOE modification did not alter the distribution properties, the presence of this modification is important in the tissue distribution patterns and in determining nuclease resistance within tissues.

### 6.4 PK Analyses from Selected Clinical Trials

Several clinical trials using Genasense (G3139, a PS DNA ASO targeting the anti-apoptotic protein Bcl-2) have included detailed PK analyses. One of the earliest trials of G3139 used a 14-day continuous subcutaneous infusion at doses ranging

from 4.6 to 195.8 mg/m$^2$/day [126]. The mean plasma half-life of G3139 was 7.46 hours. More recent trials measured PK profiles of G3139 administered at 7 mg/kd/day by continuous i.v. infusion for 8 days [36], 10 days [37], or 5 days [122]. All showed similar PK results for plasma clearance, metabolism, and urinary excretion.

The first report of administration into humans of a 2′-MOE gapmer (ISIS 104838, targeting tumor necrosis factor α) included extensive PK analysis of blood and urine samples [60]. This trial also compared i.v. and s.c. dosing regimens. The plasma half-life of the ASO was 1 to 1.8 hours at doses greater than 1 mg/kg. Urinary excretion of full-length ASO during the initial 24 hours postadministration was less than 10%, whereas excretion of 8 to 12 nucleotide metabolites (i.e., degradation products) increased over time thereafter.

Similarly, a recent clinical trial of the 2′-MOE gapmer OGX-011 (targeting the chaperone protein clusterin) in prostate cancer patients indicated dose-dependent PK and pharmacodynamic properties [61]. Plasma distribution half-lives ranged from 0.476 to 3.83 hours, and other plasma PK parameters were similar to the PS ASO G3139 [126]. Estimates of ASO concentrations in the target tissue ranged from 1.67 to 4.82 μg/g of prostate tissue, depending on the dose of ASO delivered. Metabolism and elimination were not measured in the study.

### 6.5 Target Downregulation in Tissues of Interest

Drug efficacy for ASOs can be measured in terms of reductions in target mRNA and protein. The outcomes of drug treatment are not formally PK issues, but they are within the realms of pharmacodynamics and toxicology. It is important to consider, however, how the PK profile of a systemically administered ASO affects the pharmacodynamic and toxicological properties of the drug. Many clinical trials of ASOs include some measures of target mRNA and protein in target or surrogate marker tissues. In one of the early trials of the PS DNA ASO G3139 in patients with non-Hodgkin's lymphoma, Bcl-2 mRNA and protein were demonstrably reduced in isolated peripheral blood lymphocytes, but only in 7 of 16 patients [126]. Recent clinical trials have shown similar interpatient variability in target downregulation [36, 37]. In the case of ASOs targeting clusterin, *in vitro* experiments revealed the improved efficacy of 2′-MOE modified gapmers compared with PS DNA ASOs [35]. Superior target downregulation and increased antitumor activity were also demonstrated in animal models of prostate cancer tumor growth [35]. Dose-dependent reductions in clusterin mRNA and protein in prostate tumor samples were also shown in the phase I clinical trial of OGX-011 [61], indicating that the intended biological activity was achieved in the intended target tissue. This represents a tremendous success for antisense therapeutics and should provide encouragement for others developing ASOs as potential clinically useful drugs.

## 7 RNA INTERFERENCE

RNA interference (RNAi) is a cellular regulatory mechanism in which double-stranded RNA (dsRNA) induces the specific degradation of its homologous target mRNA. This process was discovered by Guo et al. who found that injection of both

antisense and sense RNAs into *Caenorhabditis elegans* resulted in a decrease of their target mRNA [127]. Several years later, Fire et al. demonstrated that this observation was due to the single-stranded RNA populations containing small amounts of double-stranded RNAs. Using highly purified RNA preparations, they showed that dsRNA injected into *C. elegans* resulted in specific and potent interference of cellular gene expression [128]. This inhibition was much greater than that observed with injection of either the antisense or the sense strands by themselves. The molecular mechanism of RNAi involves cleavage of dsRNAs by an RNase III-like endonuclease termed Dicer into 21–25 nucleotide (nt) dsRNAs. These small interfering dsRNAs (siRNAs) are bound in a multiprotein RNA-induced silencing complex (RISC), which unwinds the siRNA, and one strand is then used as a guide to target complementary mRNAs for degradation [129–131].

RNAi gained rapid acceptance as a tool for functional genomics in *C. elegans* and *Drosophila* model systems. It was initially unclear as to whether RNAi existed in mammalian cells, because treatment of mammalian cells with dsRNA results in global degradation of RNAs and repression of protein synthesis. As researchers toiled diligently to understand the detailed molecular mechanisms involved in the RNAi pathway, Elbashir et al. identified 21- to 22-nt dsRNAs as the effector molecules that induce RNAi in *Drosophila* embryos [132]. Additionally, they demonstrated that chemically synthesized siRNAs could be transfected into mammalian cells and induce RNAi without inducing global RNA degradation [133]. Since this discovery, the field of siRNA and related methodologies has grown at an exponential rate. Although siRNAs have competed with ASOs as tools for functional genomic analysis and target validation, one primary interest is to develop siRNAs as therapeutic agents.

### 7.1 siRNA Specificity

One of the most important issues regarding development of new pharmaceuticals is their specificity of effect. The overall intent is to design molecules that inhibit their intended target without undesirable side effects. As with ASOs and other nucleic-acid-based therapies, demonstrating that the observed cellular affects are due to inhibition of the intended mRNA target remains a challenge. Initial studies suggested that one or two mismatches could alleviate the silencing ability of an siRNA [132]. However, recent studies have identified siRNAs with effects on unintended targets with as little as seven complementary nucleotides [134]. Moreover, in addition to the G:U wobble base pair, it has been shown that mismatches formed between adenine (A) and cytosine (C) are well tolerated [135]. Thus, the ability of siRNAs to specifically downregulate predicted targets is being called into question. Many investigators have turned to microarray analysis to validate the specificity of their siRNAs. This approach may nonetheless provide researchers with a false sense of specificity. Saxena et al. showed that siRNAs with multiple mismatches located in the center of the molecule can function as micro-RNAs that inhibit expression of the target protein through translational arrest and not mRNA degradation [136]. Thus, the microarray patterns from two different siRNAs targeting the same mRNA may be similar, but the siRNAs could still inhibit protein expression through translational inhibition. It remains unclear to what extent each of these potential nonspecific effects may impact their potential therapeutic use. With

a more thorough understanding of the RNAi mechanism, new selection guidelines will identify optimal target sites that account for these sequence-based nonspecific effects. Identification of the most active siRNA with limited nonspecific effects will be critical for the success of siRNA molecules as therapeutic agents.

### 7.2 Cell Culture

Since their discovery only a few years ago, many researchers in both academic and industrial settings have used siRNAs to knock down mRNA targets within cultured cells. Unlike ASOs, which can enter cells by themselves when used at high concentrations, siRNAs cannot cross the cellular membrane by themselves due to their two negatively charged phosphate backbones. Cationic lipids such as Lipofectamine 2000 (L2K) and Oligofectamine are commonly used to complex the siRNA and allow uptake into cultured cells. However, a recent publication has demonstrated that cationic lipids can have unintended effects. Fedorov et al. transfected siRNAs into HeLa cells using either L2K or electroporation and performed microarray studies [137]. Although both transfection methods resulted in a similar knockdown of the target mRNA, the number of genes affected by each method differed significantly. Lipid transfection resulted in increased expression of 65 genes, whereas electroporation increased only 11 genes. Over 50% of the upregulated genes could be directly attributed to L2K. Because L2K is also commonly used for transfecting DNA plasmids and ASOs, investigators should thoroughly examine the effect that the transfection reagent has on gene expression patterns. In addition to these lipids, other cationic molecules such as cell-penetrating peptides and polyethyleneimine have also been used to transport siRNAs across the cell membrane [138, 139]. Further studies are needed to determine what affects these transport molecules have on global gene expression.

### 7.3 *In vivo* Studies with "Naked" siRNAs

Although most siRNA researchers will readily acknowledge that the siRNAs are more potent than the ASOs and result in a more potent and reproducible target suppression, others have shown siRNAs to be equivalent to ASOs [140]. The real difference between these two approaches becomes clear once the nucleic acids are injected into animals. ASOs have therapeutic activity even when they are injected intravenously without any cationic carrier molecule, as discussed in the preceding section. Although most studies show that naked siRNA injections have no effect, some have demonstrated that uncomplexed, unmodified siRNAs can be injected into animals and inhibit the target protein in tissues. Verma et al. intravenously injected approximately 3 μg of a β-catenin-targeted siRNA into the tail vein of tumor-bearing nude mice. After 72 hours, tumors were harvested and significant decreases in β-catenin protein levels could be observed by Western blot analysis [141]. Similarly, Filleur et al. injected intraperitonealy 3 μg of siRNAs targeting either luciferase or vascular endothelcal growth factor (VEGF) and could demonstrate a 50% reduction in luciferase activity and a 70% reduction in VEGF expression. VEGF siRNA injections also delayed the onset of tumors [142]. Furthermore, Duxbury et al. delivered similar amounts of an siRNA targeting ribonucleotide reductase mRNA. After 6 weeks of biweekly tail vein siRNA injections, the levels

of target protein in the tumor were suppressed and a small decrease in the tumor growth rate was observed. More importantly, the combination of the siRNA and gemcitabine resulted in a synergistic suppression of tumor growth. The control siRNA did not enhance the antitumor affect of gemcitabine [143]. All of these results using naked siRNAs are surprising because siRNAs are rapidly degraded in plasma with a half-life of about 3 minutes [144]. Given that many investigators have shown that unmodified, uncomplexed siRNAs have little to no activity (as discussed below), it remains unclear as to whether systemic injection of unprotected siRNAs is a viable therapeutic option.

### 7.4 Hydrodynamic High-Pressure "Naked" siRNA Injection

To overcome the problem of siRNA delivery into tissues, several investigators have employed the high-pressure hydrodynamic method. This method involves rapid injection of siRNA in a large volume (1–2 mL) of saline or phosphate-buffered solution (PBS). Surprisingly, the mice can tolerate this harsh administration quite well [145]. Song et al. was able to reduce Fas mRNA and protein levels in mouse liver for up to 10 days after hydrodynamic injection (50 μg siRNA in 1-mL PBS). They also demonstrated that Fas siRNA in-jection could protect mice from fulminant hepatitis [146]. Another group targeted caspase 8 with siRNAs and showed decreased liver expression of caspase 8 and prevented Fas (CD95)-mediated apoptosis after hydrodynamic injection (120-μg siRNA in 2-mL PBS). As with the previous Fas-targeted approach, caspase 8 siRNAs improved the survival of two animal models of acute liver failure [147]. Layzer et al. used quantitative whole-body imaging to demonstrate that hydrodynamic coinjections of luciferase reporter plasmids and luciferase-targeted siRNAs could reduce luciferase expression by 85%. Injection of a nuclease-resistant 2′-F siRNA resulted in a similar reduction in luciferase expression. However, if siRNAs were injected under low pressure, no decrease in luciferase activity was observed [148]. Although the use of hydrodynamic injection in animal models is relatively simple and effective, it is not feasible in the clinical setting. An additional limitation of this methodology is that the liver is the primary organ for siRNA uptake with only limited uptake in other tissues.

### 7.5 Chemically Modified siRNAs

Contrary to the studies demonstrating activity with unmodified siRNAs, the major limitation for the use of nucleic acids as therapeutic agents is their susceptibility to attack from serum and intracellular nucleases. The antisense field has identified a host of ribose ring and phosphate backbone modifications that enhance the stability of ASOs [Figure 1]. Extensive analysis has been performed to demonstrate which chemical modifications enhance siRNA stability and still allow for efficient silencing [149–151]. Altering the phosphate backbone of an siRNA to phosphorothioate reduces its ability to induce RNAi by 50% and increases nonspecific cellular toxicity. Designing siRNAs entirely with 2′-O-methyl moieties improves their nuclease stability dramatically and eliminates their silencing ability. One nucleotide modification that seems to be the most promising is the 2′-fluorine modification on the cytidines and uridine residues. 2′-F-modified siRNAs have plasma half-lives of days instead of minutes and can inhibit their targets with similar potency as the

unmodified siRNAs. Researchers at Sirna Therapeutics have developed siRNAs in which all 2′-OH groups have been substituted with deoxy-, 2′-O-methyl-, or 2′-fluoro-nucleotides, which are termed "short interfering nucleic acid" (siNA). In human serum, siNAs have a half-life on the order of 3 days [144]. Using a hepatitis B viral mouse model, they injected HBV-targeted siNAs by standard low-pressure tail vein injections (60–600-μg siRNA) thrice daily for 2 days. After 24 hours, the levels of serum HBV DNA were significantly reduced in a dose-dependent manner. The inverted control siNA had little to no effect on serum levels of HBV DNA. No significant antiviral activity was observed with unmodified siRNAs at any dose. Although this study demonstrated *in vivo* activity of a fully modified siRNA, it is uncertain whether these siNAs are effectively taken up into tissues by standard intravenous injections. One assumes that lower serum titers are the result of lower viral production from the tissues, but no data were presented regarding decreased HBV tissue expression. Given the high doses required for activity, more studies are needed to validate the therapeutic effectiveness of siNAs.

### 7.6 Ligand-Mediated siRNA Delivery

Another approach to improve tissue delivery of siRNAs is to attach small molecules on the end of the siRNA. Soutschek et al. conjugated a cholesterol molecule on the 3′-end of the sense strand of a partial chemically modified siRNA to generate "chol-siRNA" [152]. A chol-siRNA targeting luciferase could reduce luciferase expression in HeLa cells with an IC50 of about 200 nM without any transfection reagents. Upon intravenous injection into animal models, the addition of cholesterol improved the siRNAs pharmacokinetic properties by increasing the elimination half-life and reducing plasma clearance in rodents. Chol-siRNAs, as detected by RNase protection assay, were broadly distributed throughout the tissues of the mouse confirming the improved pharmacological properties. No detectable amounts of unconjugated siRNAs were observed in any tissues. Intravenous injections of a chol-siRNA targeting apolipoptrotein B (1 mg) daily for 3 days resulted in inhibition of apoB mRNA in liver and intestine, decreased serum levels of apoB protein, and a reduction of total plasma cholesterol. Hence, this study demonstrated the importance of conjugating a small molecule to the siRNA that will enhance tissue uptake. However, further optimization is required to identify modifications that will permit more clinically acceptable doses and schedules.

### 7.7 Intranasal siRNA Injection

As an alternative approach to systemic siRNA injections, several groups have demonstrated that intranasal administration of siRNA can be effective for certain disease applications [153–155]. X. Zhang et al. [154, 155] were the first to demonstrate that unmodified siRNA could be administered intranasally and silence the intended target mRNA without the need for transfection agents. Bitko et al. [153] designed siRNAs targeting the respiratory syncytial virus and parainfluenza virus. A single intranasal siRNA administration (70 μg) provided significant protection against respiratory infection from viral challenge. Two days after intranasal injection, siRNAs could still be detected in the mouse lungs. There are several advantages for this route of administration. It is relatively painless and noninvasive. The

risk of potential side effects is also reduced because no cationic carrier molecules are required for cellular uptake. It remains unclear as to the mechanism of cellular uptake after intranasal siRNA administration.

### 7.8 *In vivo* Studies with Encapsulated siRNAs

An alternative approach for *in vivo* delivery of siRNAs is to encapsulate or complex the siRNA with cationic molecules. These positively charged molecules serve several functions. After interacting with the negatively charged phosphate backbone of the nucleic acid, they condense the nucleic acid into a more compact structure, protect against degradation by plasma nucleases, and enhance tissue uptake. The use of these cationic synthetic molecules is not limited to siRNA delivery. ASOs and DNA plasmid vectors can also be delivered with these molecules. As well, it remains unclear as to whether chemical modifications to the siRNA are necessary if one is using cationic molecules for delivery. In cell culture, lipid-complex siRNAs are protected from degradation by serum nucleases. One benefit of using natural, unmodified siRNAs is that when they actually do degrade, the nucleotides should not have any nonspecific effects of cellular metabolism.

### 7.9 Linear Polyethylenimine

The cationic molecule polyethylenimine (PEI) has been shown to effectively transport nucleic acids into cells in both the *in vitro* and the *in vivo* setting. PEI condenses nucleic acids into 0.2–0.4-nm positively charged particles, is taken up into cells by endocytosis, and then releases the nucleic acid contents into the cytoplasm efficiently via a process called the "proton sponge" [156]. Urban-Klein et al. demonstrated that PEI complexation of an unmodified siRNA protects the siRNA from nuclease degradation. In cultured ovarian carcinoma cells, they showed that linear low-molecular-weight PEI could transport their HER2-targeted siRNA into cells, inhibit HER2 protein expression by 70%, and reduce colony formation by 50% with a single PEI/siRNA transfection. In a tumor-bearing nude mouse model, intraperitoneal injection of PEI/siRNA resulted in a rapid uptake of the P32-labeled siRNA into the tumor, muscle, liver, and kidney with little uptake into the lungs. No siRNA was observed in the tissues after injection of uncomplexed siRNAs. As they observed no histological changes in the lungs and the absence of visible side effects, intraperitoneal injections of PEI/siRNA seem to avoid systemic PEI-associated toxicity previously reported [157]. Furthermore, they showed that intraperitoneal injections of PEI/HER-2 siRNA (10 μg) two to three times a week significantly reduced growth of tumor xenografts. HER2 mRNA levels were inhibited by 50% by this treatment regimen. As important controls for specificity, injection of either a nonspecific siRNA complexed with PEI or uncomplexed HER-2 siRNA had no effect on tumor growth. However, a recent report demonstrated that tail vein injections of PEI along with other cationic molecules can have antitumor activity by themselves [158]. Although this effect may be cell-line dependent, caution should be used when interpreting the therapeutic effects of siRNAs when cationic molecules are used for delivery.

One main concern with using PEI as a nucleic acid transporter is the lack of tissue specificity. To improve the ability of PEI to target tumors, Schiffelers et al.

designed ligand-targeted, sterically stabilized nanoparticles (RPPs) for complexation with siRNAs [159]. These particles comprise branched PEI that is PEGylated with an RGD peptide on the end of the polyethylene glycol (PEG). Once the siRNA is complexed in the nanoparticles, the RGD peptide should preferentially target the tumor neovasculature. One hour after intravenous injection of a fluorescently labeled siRNA complexed with RPP, significant siRNA uptake was observed in the tumor with poor liver and lung accumulation. Conversely, siRNA complexed with only branched PEI did not accumulate in the tumor to any appreciable extent but did accumulate in the liver and lung. To verify whether the siRNA distributed into the tumor was active, they evaluated the expression of luciferase in the tumor and its knockdown by a siRNA specific for luciferase. Thus, luciferase was complexed with either PEI or RPP particles and injected into tumor-bearing mice. Although luciferase levels in the tumor were equivalent for both PEI and RPP particles, only RPP could selectively target the tumor. Two hours after plasmid injections, luciferase siRNA–RPP nanoparticles were injected into the same mice. After 24 hours, the luciferase activity had decreased by 90% demonstrating sequence-specific inhibition. To evaluate whether these nanoparticles have therapeutic potential, siRNAs targeting the VEGF receptor were complexed with RPP and injected into tumor-bearing mice every 3 days at a dose of 40 μg administered in the tail vein. They observed a strong inhibition of tumor growth, a reduction in blood vessel formation around the tumor, and a decrease in the expression of tumor VEGF R2 protein levels. The control siRNA had no effect, demonstrating specificity.

### 7.10 Cationic Lipids

In addition to PEI-related molecules, various cationic lipids have been used for nucleic acid transport *in vivo*. Lipids such as the commercially available DOTAP have been effective in delivering siRNAs into tissues such as kidney and spleen [160]. Others have developed their own delivery lipids. Yano et al. designed a less-toxic cationic liposome (LIC-101) and demonstrated its *in vitro* ability to deliver a Bcl-2-targeted siRNA into four different cell lines and inhibit Bcl-2 protein expression [161]. Furthermore, tail vein injection of LIC-101/Bcl-2 siRNA inhibited tumor growth in two different mouse models. Chien et al. developed a novel cationic liposome based on a synthetic cationic cardiolipin analog (CCLA) [162]. They showed their lipid had significantly less-toxic side effects than DOTAP and could suppress tumor growth by >70% in a severe combined immunodeficiency (SCID) mouse xenograft model. Injection of a mismatched siRNA complexed with CCLA had no effect on tumor growth. Morrissey et al., in an attempt to improve delivery of their HBV-targeted siRNA, incorporated their siRNA into a specialized liposome to form a stable nucleic-acid–lipid particle (SNALP) [163]. In a mouse model, an unmodified siRNA had a plasma elimination half-life of approximately 2 minutes. Injection of a nuclease-stabilized siRNA improved substantially the elimination half-life to 49 minutes. However, siRNA encapsulation in SNALP improved the plasma pharmacokinetics most dramatically with an elimination half-life of 6.5 hours. They demonstrated that SNALP–siRNA could significantly suppress serum levels of HBV in HBV-challenged mice with much lower doses and reduced dosing frequency than previously shown with uncomplexed siRNA [144].

### 7.11 siRNAs in Clinical Trials

Although some uncertainly remains for the feasibility and effectiveness of siRNA therapeutics, several companies are rapidly moving ahead with siRNA molecules in clinical trials. In the Fall of 2004, Acuity Pharmaceuticals initiated the first human siRNA clinical trial for treatment of wet age-related macular degeneration (AMD) with their lead compound Cand5, which silences VEGF. Phase I studies showed that intravitreal injections of an unmodified siRNA, Cand5, was safe and well tolerated [164]. Pharmacokinetic analysis revealed that the siRNA was not detected in the plasma of any patient. Lack of systemic exposure due to localized siRNA injection likely prevented the adverse side effects of anti-VEGF therapy. However, given that unmodified siRNAs are very unstable in plasma, it is unclear whether the siRNA could not escape the vitreal cavity or whether it was degraded so rapidly after entry into the bloodstream that it could not be detected in plasma. Acuity Pharmaceuticals has already initiated phase II studies with its lead siRNA candidate. The second siRNA to enter clinical trials was Sirna-027 from Sirna Therapeutics. Although both Cand5 and Sirna-027 are being investigated for treatment of AMD, Sirna-027 targets the VEGF receptor as opposed to VEGF, and is chemically modified to enhance stability. Interim phase I results were recently reported and demonstrated that Sirna-027 seems to be safe and well tolerated [165]. All patients experienced improvement or stabilization of visual acuity, and 23% of the patients showed a clinically significant improvement in visual acuity. A third company (Alnylam Pharmaceuticals) has recently initiated a phase I study with its siRNA (ALN-RSV01), which targets respiratory syncytial virus. Using their Direct RNAi approach, the siRNA will be administered directly to the lungs and silence a key RSV gene, which hopefully will prevent spreading of the viral infection.

Other pharmaceutical companies are also seeking to initiate their own siRNA clinical trials. An Australian company (Benitec) in collaboration with the City of Hope is planning to enter their RNA-based HIV drug into clinical trials sometime in 2006. Unlike Cand5 and Sirna-027, which are siRNA molecules, Benitec specializes in RNAi plasmid and vector expression systems. Patients who have become resistant to HAART therapy would receive a transplant of their own stem cells containing an RNAi expression system targeting HIV to generate HIV resistance. Nucleonics, Inc. also uses RNAi expression systems for gene silencing and expects to file its first IND for hepatitis B and hepatitis C in 2006. CytRx Corporation is rapidly advancing their preclinical siRNA studies for the treatment of obesity, type 2 diabetes, and Lou Gehrig's disease. Thus, many pharmaceutical companies have decided to develop siRNAs as therapeutic agents. However, as with ASOs, the fate of therapeutic siRNAs will be determined from the initial siRNA clinical trials.

## 8 CONCLUSIONS

### 8.1 Nucleic Acid-Based Therapeutics

Plasmids and viral vectors have been developed to deliver gene therapy in animal models and human clinical trials. These approaches aim to express a gene of interest in target tissue. Many of these approaches use local delivery, and thus, PK analysis is limited to the short-range disposition, metabolism, and excretion of the

nucleic acid therapeutic. ASOs and siRNA molecules are being developed as gene therapeutics to reduce the expression of target genes, through double-stranded nucleic acid interactions with target mRNA. Development of orally bioavailable formulations and lipid-encapsulated antisense molecules is being pursued, but most preclinical studies and clinical trials employ intravenous injection to deliver these therapeutics.

### 8.2 PK (ADME) and Toxicology Differences

For ASOs, the differences in PK and toxicology properties depend largely on the chemistry of the backbone and the modifications to the sugars. Thus, the ADME characteristics are independent of nucleotide base sequence and, therefore, do not change significantly according to the gene or protein being targeted. The toxicological properties do depend to a certain extent on the nucleotide sequence, for example, the effects of CpG dinucleotides and G-quartets. Whether these principles will hold for siRNA therapeutics will likely been seen in the next few years, as the development of siRNA with novel chemical modifications progresses.

### 8.3 Relative Clinical Utility of the Different Approaches

The relative effectiveness of ASOs and siRNAs, in terms of target downregulation, have been studied in cell culture and animal experiments [29, 140, 166, 167]. Although it seems that siRNA may be more effective than ASOs in most experimental circumstances, the development of modified ASOs with improved PK properties, including stability to nucleases, uptake into target tissues, and tissue half-life, has progressed much further than for siRNA. Indeed, although many ASOs of various chemistries have progressed into phase II and III clinical trials and thus the PK studies are widely reported, siRNAs are just now entering human clinical trials and their PK properties are not yet fully described or understood. It seems likely that both types of antisense therapeutics will demonstrate clinical utility in a variety of indications in the coming years.

## REFERENCES

1. Wilton S D, Fletcher S (2005). RNA splicing manipulation: strategies to modify gene expression for a variety of therapeutic outcomes. *Curr. Gene Ther.* 5:467–483.
2. Kole R, Vacek M, Williams T (2004). Modification of alternative splicing by antisense therapeutics. *Oligonucleotides.* 14:65–74.
3. Sazani P, Kole R (2003). Therapeutic potential of antisense oligonucleotides as modulators of alternative splicing. *J. Clin. Invest.* 112:481–486.
4. Edelstein M L, Abedi M R, Wixon J, Edelstein R M (2004). Gene therapy clinical trials worldwide 1989–2004-an overview. *J. Gene Med.* 6:597–602.
5. Shah P B, Losordo D W (2005). Non-viral vectors for gene therapy: Clinical trials in cardiovascular disease. *Adv. Genet.* 54:339–361.
6. Nishikawa M, Takakura Y, Hashida M (2005). Pharmacokinetics of plasmid DNA-based non-viral gene medicine. *Adv. Genet.* 53PA:47–68.

7. Lee P A, Blatt L M, et al. (2000). Pharmacokinetics and tissue distribution of a ribozyme directed against hepatitis C virus RNA following subcutaneous or intravenous administration in mice. *Hepatol.* 32:640–646.
8. Schubert S, Kurreck J (2004). Ribozyme- and deoxyribozyme-strategies for medical applications. *Curr. Drug Targets.* 5:667–681.
9. Parry T J, Bouhana K S, Blanchard K S, et al. (2000). Ribozyme pharmacokinetic screening for predicting pharmacodynamic dosing regimens. *Curr. Issues Mol. Biol.* 2:113–118.
10. Sandberg J A, Parker V P, et al. (2000). Pharmacokinetics and tolerability of an anti-angiogenic ribozyme (ANGIOZYME) in healthy volunteers. *J. Clin. Pharmacol.* 40:1462–1469.
11. Kobayashi H, Eckhardt S G, et al. (2005). Safety and pharmacokinetic study of RPI.4610 (ANGIOZYME), an anti-VEGFR-1 ribozyme, in combination with carboplatin and paclitaxel in patients with advanced solid tumors. *Cancer Chemother. Pharmacol.* 56:329–336.
12. Weng D E, Masci P A, et al. (2005). A phase I clinical trial of a ribozyme-based angiogenesis inhibitor targeting vascular endothelial growth factor receptor-1 for patients with refractory solid tumors. *Mol. Cancer Ther.* 4:948–955.
13. Amantana A, Iversen P L (2005). Pharmacokinetics and biodistribution of phosphorodiamidate morpholino antisense oligomers. *Curr. Opin. Pharmacol.* 5:550–555.
14. Devi G R, Beer T M, et al. (2005). In vivo bioavailability and pharmacokinetics of a c-MYC antisense phosphorodiamidate morpholino oligomer, AVI-4126, in solid tumors. *Clin. Cancer Res.* 11:3930–3938.
15. Arora V, Devi G R, Iversen P L (2004). Neutrally charged phosphorodiamidate morpholino antisense oligomers: uptake, efficacy and pharmacokinetics. *Curr. Pharm. Biotechnol.* 5:431–439.
16. Jepsen J S, Sorensen M D, Wengel J (2004). Locked nucleic acid: a potent nucleic acid analog in therapeutics and biotechnology. *Oligonucleotides.* 14:130–146.
17. Fluiter K, ten Asbroek A L, et al. (2003). In vivo tumor growth inhibition and biodistribution studies of locked nucleic acid (LNA) antisense oligonucleotides. *Nucleic Acids Res.* 31:953–962.
18. Wahlestedt C, Salmi P, et al. (2000). Potent and nontoxic antisense oligonucleotides containing locked nucleic acids. *Proc. Natl. Acad. Sci. U.S.A.* 97:5633–5638.
19. McMahon B M, Mays D, et al. (2002). Pharmacokinetics and tissue distribution of a peptide nucleic acid after intravenous administration. *Antisense Nucleic Acid Drug Dev.* 12:65–70.
20. Kristensen E (2002). In vitro and in vivo studies on the pharmacokinetics and metabolism of PNA constructs in rodents. *Methods Mol. Biol.* 208:259–269.
21. Crooke S T (2004). Antisense strategies. *Curr. Mol. Med.* 4:465–487.
22. Kurreck J (2003). Antisense technologies. Improvement through novel chemical modifications. *Eur. J. Biochem.* 270:1628–1644.
23. Giles R V, Spiller D G, Tidd D M (1993). Chimeric oligodeoxynucleotide analogues: enhanced cell uptake of structures which direct ribonuclease H with high specificity. *Anticancer Drug Des.* 8:33–51.
24. ten Asbroek A L, van Groenigen M, Nooij M, Baas F (2002). The involvement of human ribonucleases H1 and H2 in the variation of response of cells to antisense phosphorothioate oligonucleotides. *Eur. J. Biochem.* 269:583–592.
25. Zamaratski E, Pradeepkumar P I, Chattopadhyaya J (2001). A critical survey of the structure-function of the antisense oligo/RNA heteroduplex as substrate for RNase H. *J. Biochem. Biophys. Meth.* 48:189–208.

26. Stein C A, Cohen J S (1988). Oligodeoxynucleotides as inhibitors of gene expression: A review. *Cancer Res.* 48:2659–2668.
27. Achenbach T V, Brunner B, Heermeier K (2003). Oligonucleotide-based knockdown technologies: Antisense versus RNA interference. *Chembiochem.* 4:928–935.
28. Scanlon K J (2004). Anti-genes: siRNA, ribozymes and antisense. *Curr. Pharm. Biotechnol.* 5:415–420.
29. Grunweller A, Wyszko E, et al. (2003). Comparison of different antisense strategies in mammalian cells using locked nucleic acids, 2′-O-methyl RNA, phosphorothioates and small interfering RNA. *Nucleic Acids Res.* 31:3185–3193.
30. Micklefield J (2001). Backbone modification of nucleic acids: synthesis, structure and therapeutic applications. *Curr. Med. Chem.* 8:1157–1179.
31. Geary R S, Watanabe T A, et al. (2001). Pharmacokinetic properties of 2′-O-(2-methoxyethyl)-modified oligonucleotide analogs in rats. *J. Pharmacol. Exp. Ther.* 296:890–897.
32. Geary R S, Yu R Z, Levin A A (2001). Pharmacokinetics of phosphorothioate antisense oligodeoxynucleotides. *Curr. Opin. Investig. Drugs.* 2:562–573.
33. Agrawal S (1999). Importance of nucleotide sequence and chemical modifications of antisense oligonucleotides. *Biochim. Biophys. Acta.* 1489:53–68.
34. Yu R Z, Geary R S, et al. (2004). Tissue disposition of 2′-O-(2-methoxy) ethyl modified antisense oligonucleotides in monkeys. *J. Pharm. Sci.* 93:48–59.
35. Zellweger T, Miyake H, et al. (2001). Antitumor activity of antisense clusterin oligonucleotides is improved in vitro and in vivo by incorporation of 2′-O-(2-methoxy)ethyl chemistry. *J. Pharmacol. Exp. Ther.* 298:934–940.
36. Tolcher A W, Chi K, et al. (2005). A phase II, pharmacokinetic, and biological correlative study of oblimersen sodium and docetaxel in patients with hormone-refractory prostate cancer. *Clin. Cancer Res.* 11:3854–3861.
37. Marcucci G, Stock W, et al. (2005). Phase I study of oblimersen sodium, an antisense to Bcl-2, in untreated older patients with acute myeloid leukemia: Pharmacokinetics, pharmacodynamics, and clinical activity. *J. Clin. Oncol.* 23:3404–3411.
38. Tolcher A W, Kuhn J, et al. (2004). A Phase I pharmacokinetic and biological correlative study of oblimersen sodium (genasense, g3139), an antisense oligonucleotide to the bcl-2 mRNA, and of docetaxel in patients with hormone-refractory prostate cancer. *Clin. Cancer Res.* 10:5048–5057.
39. Rudin C M, Kozloff M, et al. (2004). Phase I study of G3139, a bcl-2 antisense oligonucleotide, combined with carboplatin and etoposide in patients with small-cell lung cancer. *J. Clin. Oncol.* 22:1110–1117.
40. Morris M J, Tong W P, et al. (2002). Phase I trial of BCL-2 antisense oligonucleotide (G3139) administered by continuous intravenous infusion in patients with advanced cancer. *Clin. Cancer Res.* 8:679–683.
41. Desai A A, Schilsky R L, et al. (2005). A phase I study of antisense oligonucleotide GTI-2040 given by continuous intravenous infusion in patients with advanced solid tumors. *Ann. Oncol.* 16:958–965.
42. Bishop M R, Iversen P L, et al. (1996). Phase I trial of an antisense oligonucleotide OL(1)p53 in hematologic malignancies. *J. Clin. Oncol.* 14:1320–1326.
43. Bayever E, Iversen P L, et al. (1993). Systemic administration of a phosphorothioate oligonucleotide with a sequence complementary to p53 for acute myelogenous leukemia and myelodysplastic syndrome: initial results of a phase I trial. *Antisense Res. Dev.* 3:383–390.

44. Stewart D J, Donehower R C, et al. (2003). A phase I pharmacokinetic and pharmacodynamic study of the DNA methyltransferase 1 inhibitor MG98 administered twice weekly. *Ann. Oncol.* 14:766–774.
45. Sereni D, Tubiana R, et al. (1999). Pharmacokinetics and tolerability of intravenous trecovirsen (GEM 91), an antisense phosphorothioate oligonucleotide, in HIV-positive subjects. *J. Clin. Pharmacol.* 39:47–54.
46. Zhang R, Yan J, et al. (1995). Pharmacokinetics of an anti-human immunodeficiency virus antisense oligodeoxynucleotide phosphorothioate (GEM 91) in HIV-infected subjects. *Clin. Pharmacol. Ther.* 58:44–53.
47. Tolcher A W, Reyno L, et al. (2002). A randomized phase II and pharmacokinetic study of the antisense oligonucleotides ISIS 3521 and ISIS 5132 in patients with hormone-refractory prostate cancer. *Clin. Cancer Res.* 8:2530–2535.
48. Rudin C M, Holmlund J, et al. (2001). Phase I trial of ISIS 5132, an antisense oligonucleotide inhibitor of c-raf-1, administered by 24-hour weekly infusion to patients with advanced cancer. *Clin. Cancer Res.* 7:1214–1220.
49. Stevenson J P, Yao K S, et al. (1999). Phase I clinical/pharmacokinetic and pharmacodynamic trial of the c-raf-1 antisense oligonucleotide ISIS 5132 (CGP 69846A). *J. Clin. Oncol.* 17:2227–2236.
50. Adjei A A, Dy G K, et al. (2003). A phase I trial of ISIS 2503, an antisense inhibitor of H-ras, in combination with gemcitabine in patients with advanced cancer. *Clin. Cancer Res.* 9:115–123.
51. Cunningham C C, Holmlund J T, et al. (2001). A Phase I trial of H-ras antisense oligonucleotide ISIS 2503 administered as a continuous intravenous infusion in patients with advanced carcinoma. *Cancer.* 92:1265–1271.
52. Kahan B D, Stepkowski S, et al. (2004). Phase I and phase II safety and efficacy trial of intercellular adhesion molecule-1 antisense oligodeoxynucleotide (ISIS 2302) for the prevention of acute allograft rejection. *Transplan.* 78:858–863.
53. Yacyshyn B R, Barish C, et al. (2002). Dose ranging pharmacokinetic trial of high-dose alicaforsen (intercellular adhesion molecule-1 antisense oligodeoxynucleotide) (ISIS 2302) in active Crohn's disease. *Aliment. Pharmacol. Ther.* 16:1761–1770.
54. Glover J M, Leeds J M, et al. (1997). Phase I safety and pharmacokinetic profile of an intercellular adhesion molecule-1 antisense oligodeoxynucleotide (ISIS 2302). *J. Pharmacol. Exp. Ther.* 282:1173–1180.
55. Marshall J L, Eisenberg S G, et al. (2004). A phase II trial of ISIS 3521 in patients with metastatic colorectal cancer. *Clin. Colorectal Cancer.* 4:268–274.
56. Mani S, Rudin C M, et al. (2002). Phase I clinical and pharmacokinetic study of protein kinase C-alpha antisense oligonucleotide ISIS 3521 administered in combination with 5-fluorouracil and leucovorin in patients with advanced cancer. *Clin. Cancer Res.* 8:1042–1048.
57. Villalona-Calero M A, Ritch P, et al. (2004). A phase I/II study of LY900003, an antisense inhibitor of protein kinase C-alpha, in combination with cisplatin and gemcitabine in patients with advanced non-small cell lung cancer. *Clin. Cancer Res.* 10:6086–6093.
58. Goel S, Desai K, et al. (2003). A safety study of a mixed-backbone oligonucleotide (GEM231) targeting the type I regulatory subunit alpha of protein kinase A using a continuous infusion schedule in patients with refractory solid tumors. *Clin. Cancer Res.* 9:4069–4076.
59. Chen H X, Marshall J L, et al. (2000). A safety and pharmacokinetic study of a mixed-backbone oligonucleotide (GEM231) targeting the type I protein kinase A by two-

hour infusions in patients with refractory solid tumors. *Clin. Cancer Res.* 6:1259–1266.

60. Sewell K L, Geary R S, et al. (2002). Phase I trial of ISIS 104838, a 2′-methoxyethyl modified antisense oligonucleotide targeting tumor necrosis factor-alpha. *J. Pharmacol. Exp. Ther.* 303:1334–1343.
61. Chi K N, Eisenhauer E, et al. (2005). A phase I pharmacokinetic and pharmacodynamic study of OGX-011, a 2′-methoxyethyl antisense oligonucleotide to clusterin, in patients with localized prostate cancer. *J. Natl. Cancer Inst.* 97:1287–1296.
62. Jason T L, Koropatnick J, Berg R W (2004). Toxicology of antisense therapeutics. *Toxicol. Appl. Pharmacol.* 201:66–83.
63. Krieg A M (1999). Mechanisms and applications of immune stimulatory CpG oligodeoxynucleotides. *Biochim. Biophys. Acta.* 1489:107–116.
64. Ioannou X P, Gomis S M, et al. (2002). CpG-containing oligodeoxynucleotides, in combination with conventional adjuvants, enhance the magnitude and change the bias of the immune responses to a herpesvirus glycoprotein. *Vaccine.* 21:127–137.
65. Levin A A (1999). A review of the issues in the pharmacokinetics and toxicology of phosphorothioate antisense oligonucleotides. *Biochim. Biophys. Acta.* 1489:69–84.
66. Shaw D R, Rustagi P K, et al. (1997). Effects of synthetic oligonucleotides on human complement and coagulation. *Biochem. Pharmacol.* 53:1123–1132.
67. Gleave M E, Monia B P (2005). Antisense therapy for cancer. *Nat. Rev. Cancer.* 5:468–479.
68. Fluiter K, Frieden M, et al. (2005). On the in vitro and in vivo properties of four locked nucleic acid nucleotides incorporated into an anti-H-Ras antisense oligonucleotide. *Chembiochem.* 6:1104–1109.
69. Jepsen J S, Wengel J (2004). LNA-antisense rivals siRNA for gene silencing. *Curr. Opin. Drug Discov. Devel.* 7:188–194.
70. Vester B, Wengel J (2004). LNA (locked nucleic acid): high-affinity targeting of complementary RNA and DNA. *Biochem.* 43:13233–13241.
71. Frieden M, Christensen S M, et al. (2003). Expanding the design horizon of antisense oligonucleotides with alpha-L-LNA. *Nucleic Acids Res.* 31:6365–6372.
72. Lok C N, Viazovkina E, et al. (2002). Potent gene-specific inhibitory properties of mixed-backbone antisense oligonucleotides comprised of 2′-deoxy-2′-fluoro-D-arabinose and 2′-deoxyribose nucleotides. *Biochem.* 41:3457–3467.
73. Mangos M M, Min K L, et al. (2003). Efficient RNase H-directed cleavage of RNA promoted by antisense DNA or 2′F-ANA constructs containing acyclic nucleotide inserts. *J. Am. Chem. Soc.* 125:654–661.
74. Wilds C J, Damha M J (2000). 2′-Deoxy-2′-fluoro-beta-D-arabinonucleosides and oligonucleotides (2′F-ANA): synthesis and physicochemical studies. *Nucleic Acids Res.* 28:3625–3635.
75. Kaihatsu K, Huffman K E, Corey D R (2004). Intracellular uptake and inhibition of gene expression by PNAs and PNA-peptide conjugates. *Biochem.* 43:14340–14347.
76. Nielsen P E (2005). Gene targeting using peptide nucleic acid. *Methods Mol. Biol.* 288:343–358.
77. Marin V L, Roy S, Armitage B A (2004). Recent advances in the development of peptide nucleic acid as a gene-targeted drug. *Expert. Opin. Biol. Ther.* 4:337–348.
78. Foubister V (2003). New PNA tool makes great antisense. *Drug Discov. Today.* 8:520–521.
79. Koppelhus U, Nielsen P E (2003). Cellular delivery of peptide nucleic acid (PNA). *Adv. Drug Deliv. Rev.* 55:267–280.

80. Stephens A C (2004). Technology evaluation: AVI-4126, AVI BioPharma. *Curr. Opin. Mol. Ther.* 6:551–558.
81. Pannier A K, Arora V, Iversen P L, et al. (2004). Transdermal delivery of phosphorodiamidate Morpholino oligomers across hairless mouse skin. *Int. J. Pharm.* 275:217–226.
82. Iversen P L (2001). Phosphorodiamidate morpholino oligomers: favorable properties for sequence-specific gene inactivation. *Curr. Opin. Mol. Ther.* 3:235–238.
83. Iversen P L, Arora V, Acker A J, et al. (2003). Efficacy of antisense morpholino oligomer targeted to c-myc in prostate cancer xenograft murine model and a Phase I safety study in humans. *Clin. Cancer Res.* 9:2510–2519.
84. Berg R W, Werner M, et al. (2001). Tumor growth inhibition in vivo and G2/M cell cycle arrest induced by antisense oligodeoxynucleotide targeting thymidylate synthase. *J. Pharmacol. Exp. Ther.* 298:477–484.
85. Kuhnast B, Dolle F, et al. (2000). General method to label antisense oligonu-cleotides with radioactive halogens for pharmacological and imaging studies. *Bioconjug. Chem.* 11:627–636.
86. Yu R Z, Geary R S, Levin A A (2004). Application of novel quantitative bioanalytical methods for pharmacokinetic and pharmacokinetic/pharmacodynamic assessments of antisense oligonucleotides. *Curr. Opin. Drug Discov. Devel.* 7:195–203.
87. Gilar M, Fountain K J, et al. (2003). Characterization of therapeutic oligonucleotides using liquid chromatography with on-line mass spectrometry detection. *Oligonucleotides.* 13:229–243.
88. Freudemann T, von Brocke A, Bayer E (2001). On-line coupling of capillary gel electrophoresis with electrospray mass spectrometry for oligonucleotide analysis. *Anal. Chem.* 73:2587–2593.
89. de Serres M, McNulty M J, Christensen L, et al. (1996). Development of a novel scintillation proximity competitive hybridization assay for the determination of phosphorothioate antisense oligonucleotide plasma concentrations in a toxicokinetic study. *Anal. Biochem.* 233:228–233.
90. Brown-Augsburger P, Yue X M, et al. (2004). Development and validation of a sensitive, specific, and rapid hybridization-ELISA assay for determination of concentrations of a ribozyme in biological matrices. *J. Pharm. Biomed. Anal.* 34:129–139.
91. Brand R M (2001). Topical and transdermal delivery of antisense oligonucleotides. *Curr. Opin. Mol. Ther.* 3:244–248.
92. Brand R M, Iversen P L (2005). Transdermal delivery of antisense oligonucleotides. *Methods Mol. Med.* 106:255–269.
93. White P J, Atley L M, Wraight C J (2004). Antisense oligonucleotide treatments for psoriasis. *Expert. Opin. Biol. Ther.* 4:75–81.
94. Dokka S, Cooper S R, Kelly S, et al. (2005). Dermal delivery of topically applied oligonucleotides via follicular transport in mouse skin. *J. Invest Dermatol.* 124:971–975.
95. White P J, Gray A C, et al. (2002). C-5 propyne-modified oligonucleotides penetrate the epidermis in psoriatic and not normal human skin after topical application. *J. Invest Dermatol.* 118:1003–1007.
96. Ramakumar S, Phull H, et al. (2005). Novel delivery of oligonucleotides using a topical hydrogel tissue sealant in a murine partial nephrectomy model. *J. Urol.* 174:1133–1136.
97. Bochot A, Couvreur P, Fattal E (2000). Intravitreal administration of antisense oligonucleotides: potential of liposomal delivery. *Prog. Retin. Eye Res.* 19:131–147.

98. Klimuk S K, Semple S C, et al. (2000). Enhanced anti-inflammatory activity of a liposomal intercellular adhesion molecule-1 antisense oligodeoxynucleotide in an acute model of contact hypersensitivity. *J. Pharmacol. Exp. Ther.* 292:480–488.
99. Tezel A, Dokka S, Kelly S, Hardee G E, Mitragotri S (2004). Topical delivery of anti-sense oligonucleotides using low-frequency sonophoresis. *Pharm. Res.* 21:2219–2225.
100. Regnier V, Preat V (1998). Localization of a FITC-labeled phosphorothioate oligodeoxynucleotide in the skin after topical delivery by iontophoresis and electroporation. *Pharm. Res.* 15:1596–1602.
101. Sandrasagra A, Leonard S A, et al. (2002). Discovery and development of respirable antisense therapeutics for asthma. *Antisense Nucleic Acid Drug Dev.* 12:177–181.
102. Nyce J (2002). Respirable antisense oligonucleotides: a new, third drug class targeting respiratory disease. *Curr. Opin. Allergy Clin. Immunol.* 2:533–536.
103. Ball H A, Van Scott M R, Robinson C B (2004). Sense and antisense: therapeutic potential of oligonucleotides and interference RNA in asthma and allergic disorders. *Clin. Rev. Allergy Immunol.* 27:207–217.
104. Ball H A, Sandrasagra A, et al. (2003). Clinical potential of respirable antisense oligonucleotides (RASONs) in asthma. *Am. J. Pharmacogenom.* 3:97–106.
105. Ali S, Leonard S A, et al. (2001). Absorption, distribution, metabolism, and excretion of a respirable antisense oligonucleotide for asthma. *Am. J. Respir. Crit. Care Med.* 163:989–993.
106. Templin M V, Levin A A, et al. (2000). Pharmacokinetic and toxicity profile of a phosphorothioate oligonucleotide following inhalation delivery to lung in mice. *Antisense Nucleic Acid Drug Dev.* 10:359–368.
107. Geary R S, Khatsenko O, et al. (2001). Absolute bioavailability of 2′-O-(2-methoxyethyl)-modified antisense oligonucleotides following intraduodenal instillation in rats. *J. Pharmacol. Exp. Ther.* 296:898–904.
108. Raoof A A, Chiu P, et al. (2004). Oral bioavailability and multiple dose tolerability of an antisense oligonucleotide tablet formulated with sodium caprate. *J. Pharm. Sci.* 93:1431–1439.
109. Agrawal S, Zhang X, et al. (1995). Absorption, tissue distribution and in vivo stability in rats of a hybrid antisense oligonucleotide following oral administration. *Biochem. Pharmacol.* 50:571–576.
110. Raoof A A, Ramtoola Z, et al. (2002). Effect of sodium caprate on the intestinal absorption of two modified antisense oligonucleotides in pigs. *Eur. J. Pharm. Sci.* 17:131–138.
111. Yu R Z, Geary R S, et al. (1999). Pharmacokinetics and tissue disposition in monkeys of an antisense oligonucleotide inhibitor of Ha-ras encapsulated in stealth liposomes. *Pharm. Res.* 16:1309–1315.
112. Miyano-Kurosaki N, Barnor J S, et al. (2004). In vitro and in vivo transport and delivery of phosphorothioate oligonucleotides with cationic liposomes. *Antivir. Chem. Chemother.* 15:93–100.
113. Yoshida J, Mizuno M (2003). Clinical gene therapy for brain tumors. Liposomal delivery of anticancer molecule to glioma. *J. Neurooncol.* 65:261–267.
114. Rudin C M, Marshall J L, et al. (2004). Delivery of a liposomal c-raf-1 antisense oligonucleotide by weekly bolus dosing in patients with advanced solid tumors: A phase I study. *Clin. Cancer Res.* 10:7244–7251.
115. Springate C M, Jackson J K, Gleave M E, et al. (2005). Efficacy of an intratumoral controlled release formulation of clusterin antisense oligonucleotide complexed with

chitosan containing paclitaxel or docetaxel in prostate cancer xenograft models. *Cancer Chemother. Pharmacol.* 56:239–247.

116. Agrawal S, Temsamani J, Tang J Y (1991). Pharmacokinetics, biodistribution, and stability of oligodeoxynucleotide phosphorothioates in mice. *Proc. Natl. Acad. Sci. U.S.A.* 88:7595–7599.
117. Dvorchik B H (2000). The disposition (ADME) of antisense oligonucleotides. *Curr. Opin. Mol. Ther.* 2:253–257.
118. Saijo Y, Perlaky L, Wang H, et al. (1994). Pharmacokinetics, tissue distribution, and stability of antisense oligodeoxynucleotide phosphorothioate ISIS 3466 in mice. *Oncol. Res.* 6:243–249.
119. Geary R S, Henry S P, Grillone L R (2002). Fomivirsen: clinical pharmacology and potential drug interactions. *Clin. Pharmacokinet.* 41:255–260.
120. Cotter F E, Waters J, Cunningham D (1999). Human Bcl-2 antisense therapy for lymphomas. *Biochim. Biophys. Acta.* 1489:97–106.
121. Raynaud F I, Orr R M, et al. (1997). Pharmacokinetics of G3139, a phosphorothioate oligodeoxynucleotide antisense to bcl-2, after intravenous adminis-tration or continuous subcutaneous infusion to mice. *J. Pharmacol. Exp. Ther.* 281:420–427.
122. O'Brien S M, Cunningham C C, et al. (2005). Phase I to II multicenter study of oblimersen sodium, a Bcl-2 antisense oligonucleotide, in patients with advanced chronic lymphocytic leukemia. *J. Clin. Oncol.* 23:7697–7702.
123. Lopes de Menezes D E, Mayer L D (2002). Pharmacokinetics of Bcl-2 antisense oligonucleotide (G3139) combined with doxorubicin in SCID mice bearing human breast cancer solid tumor xenografts. *Cancer Chemother. Pharmacol.* 49:57–68.
124. Cossum P A, Truong L, et al. (1994). Pharmacokinetics of a 14C-labeled phosphorothioate oligonucleotide, ISIS 2105, after intradermal administration to rats. *J. Pharmacol. Exp. Ther.* 269:89–94.
125. Cossum P A, Sasmor H, et al. (1993). Disposition of the 14C-labeled phosphorothioate oligonucleotide ISIS 2105 after intravenous administration to rats. *J. Pharmacol. Exp. Ther.* 267:1181–1190.
126. Waters J S, Webb A, et al. (2000). Phase I clinical and pharmacokinetic study of bcl-2 antisense oligonucleotide therapy in patients with non-Hodgkin's lymphoma. *J. Clin. Oncol.* 18:1812–1823.
127. Guo S, Kemphues K J (1995). par-1, a gene required for establishing polarity in C. elegans embryos, encodes a putative Ser/Thr kinase that is asymmetrically distributed. *Cell.* 81:611–620.
128. Fire A, Xu S, et al. (1998). Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. *Nature.* 391:806–811.
129. Zamore P D (2001). RNA interference: Listening to the sound of silence. *Nat. Struct. Biol.* 8:746–750.
130. Sharp P A (2001). RNA interference–2001. *Genes Dev.* 15:485–490.
131. Bernstein E, Denli A M, Hannon G J (2001). The rest is silence. *RNA.* 7:1509–1521.
132. Elbashir S M, Martinez J, Patkaniowska A, et al. (2001). Functional anatomy of siRNAs for mediating efficient RNAi in Drosophila melanogaster embryo lysate. *EMBO J.* 20:6877–6888.
133. Elbashir S M, Harborth J, et al. (2001). Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. *Nature.* 411:494–498.
134. Lin X, Ruan X, et al. (2005). siRNA-mediated off-target gene silencing triggered by a 7 nt complementation. *Nucleic Acids Res.* 33:4527–4535.

135. Du Q, Thonberg H, Wang J, et al. (2005). A systematic analysis of the silencing effects of an active siRNA at all single-nucleotide mismatched target sites. *Nucleic Acids Res.* 33:1671–1677.

136. Saxena S, Jonsson Z O, Dutta A (2003). Small RNAs with imperfect match to endogenous mRNA repress translation. Implications for off-target activity of small inhibitory RNA in mammalian cells. *J. Biol. Chem.* 278:44312–44319.

137. Fedorov Y, King A, et al. (2005). Different delivery methods-different expression profiles. *Nat. Methods.* 2:241.

138. Muratovska A, Eccles M R (2004). Conjugate for efficient delivery of short interfering RNA (siRNA) into mammalian cells. *FEBS Lett.* 558:63–68.

139. Urban-Klein B, Werth S, Abuharbeid S, et al. (2005). RNAi-mediated gene-targeting through systemic application of polyethylenimine (PEI)-complexed siRNA in vivo. *Gene Ther.* 12:461–466.

140. Vickers T A, Koo S, et al. (2003). Efficient reduction of target RNAs by small interfering RNA and RNase H-dependent antisense agents. A comparative analysis. *J. Biol. Chem.* 278:7108–7118.

141. Verma U N, Surabhi R M, Schmaltieg A, Becerra C, Gaynor R B (2003). Small interfering RNAs directed against beta-catenin inhibit the in vitro and in vivo growth of colon cancer cells. *Clin. Cancer Res.* 9:1291–1300.

142. Filleur S, Courtin A, et al. (2003). SiRNA-mediated inhibition of vascular endothelial growth factor severely limits tumor resistance to antiangiogenic thrombospondin-1 and slows tumor vascularization and growth. *Cancer Res.* 63:3919–3922.

143. Duxbury M S, Ito H, Zinner M J, et al. (2004). RNA interference targeting the M2 subunit of ribonucleotide reductase enhances pancreatic adenocarcinoma chemosensitivity to gemcitabine. *Oncogene.* 23:1539–1548.

144. Morrissey D V, Blanchard K, et al. (2005). Activity of stabilized short inter-fering RNA in a mouse model of hepatitis B virus replication. *Hepatol.* 41:1349–1356.

145. Lewis D L, Hagstrom J E, Loomis A G, et al. (2002). Efficient delivery of siRNA for inhibition of gene expression in postnatal mice. *Nat. Genet.* 32:107–108.

146. Song E, Lee S K, et al. (2003). RNA interference targeting Fas protects mice from fulminant hepatitis. *Nat. Med.* 9:347–351.

147. Zender L, Hutker S, et al. (2003). Caspase 8 small interfering RNA prevents acute liver failure in mice. *Proc. Natl. Acad. Sci. U.S.A.* 100:7797–7802.

148. Layzer J M, McCaffrey A P, et al. (2004). In vivo activity of nuclease-resistant siRNAs. *RNA.* 10:766–771.

149. Chiu Y L, Rana T M (2003). siRNA function in RNAi: A chemical modification analysis. *RNA.* 9:1034–1048.

150. Czauderna F, Fechtner M, et al. (2003). Structural variations and stabilising modifications of synthetic siRNAs in mammalian cells. *Nucleic Acids Res.* 31:2705–2716.

151. Harborth J, Elbashir S M, et al. (2003). Sequence, chemical, and structural variation of small interfering RNAs and short hairpin RNAs and the effect on mammalian gene silencing. *Antisense Nucleic Acid Drug Dev.* 13:83–105.

152. Soutschek J, Akinc A, et al. (2004). Therapeutic silencing of an endogenous gene by systemic administration of modified siRNAs. *Nature.* 432:173–178.

153. Bitko V, Musiyenko A, Shulyayeva O, Barik S (2005). Inhibition of respiratory viruses by nasally administered siRNA. *Nat. Med.* 11:50–55.

154. Zhang W, Yang H, et al. (2005). Inhibition of respiratory syncytial virus infection with intranasal siRNA nanoparticles targeting the viral NS1 gene. *Nat. Med.* 11:56–62.

155. Zhang X, Shan P, et al. (2004). Small interfering RNA targeting heme oxygenase-1 enhances ischemia-reperfusion-induced lung apoptosis. *J. Biol. Chem.* 279:10677–10684.

156. Boussif O, Lezoualc'h F, et al. (1995). A versatile vector for gene and oligonucleotide transfer into cells in culture and in vivo: polyethylenimine. *Proc. Natl. Acad. Sci. U.S.A.* 92:7297–7301.

157. Chollet P, Favrot M C, Hurbin A, et al. (2002). Side-effects of a systemic injection of linear polyethylenimine-DNA complexes. *J. Gene Med.* 4:84–91.

158. Dufes C, Keith W N, et al. (2005). Synthetic anticancer gene medicine exploits intrinsic antitumor activity of cationic vector to cure established tumors. *Cancer Res.* 65:8079–8084.

159. Schiffelers R M, Ansari A, et al. (2004). Cancer siRNA therapy by tumor selective delivery with ligand-targeted sterically stabilized nanoparticle. *Nucleic Acids Res.* 32: e149.

160. Sioud M, Sorensen D R (2003). Cationic liposome-mediated delivery of siRNAs in adult mice. *Biochem. Biophys. Res. Commun.* 312:1220–1225.

161. Yano J, Hirabayashi K, et al. (2004). Antitumor activity of small interfering RNA/cationic liposome complex in mouse models of cancer. *Clin. Cancer Res.* 10:7721–7726.

162. Chien P Y, Wang J, et al. (2005). Novel cationic cardiolipin analogue-based liposome for efficient DNA and small interfering RNA delivery in vitro and in vivo. *Cancer Gene Ther.* 12:321–328.

163. Morrissey D V, Lockridge J A, et al. (2005). Potent and persistent in vivo anti-HBV activity of chemically modified siRNAs. *Nat. Biotechnol.* 23:1002–1007.

164. Prenner J L, Thompson J T, Miller D G, et al. (2005). An Open Label Study for the Evaluation of Tolerability of Five Dose Levels of Intravitreous VEGF siRNA (Cand5) in Patients With Wet Age-Related Macular Degeneration. *American Academy of Ophthalmology Annual Meeting,* poster 463.

165. Quinlan E, Nguyen Q, Kaiser P, et al. (2005). A Phase I, open-label, dose escalation trial of intravitreal injection of small interfering RNA molecule in subjects with neovascular AMD. *American Academy of Ophthalmology Annual Meeting.* Available: http://www.sirna.com/media/pdfs/AAOSIRNA.pdf.

166. Miyagishi M, Hayashi M, Taira K (2003). Comparison of the suppressive effects of antisense oligonucleotides and siRNAs directed against the same targets in mammalian cells. *Antisense Nucleic Acid Drug Dev.* 13:1–7.

167. Bertrand J R, Pottier M, et al. (2002). Comparison of antisense oligonucleotides and siRNAs in cell culture and in vivo. *Biochem. Biophys. Res. Commun.* 296:1000–1004.

# CASE STUDIES—DEVELOPMENT OF OLIGONUCLEOTIDES

Ezharul Hoque Chowdhury and Toshihiro Akaike
*Tokyo Institute of Technology, Yokohama, Japan*

**Chapter Contents**

1 Introduction
2 Oligonucleotides in Functional Genomics for Drug Target Discovery
  2.1 Northern Blotting
  2.2 Polymerase Chain Reaction (PCR)
  2.3 High-Density Oligonucleotide Arrays
3 Oligonucleotides in Drug Target Validation and Potential Therapeutics
  3.1 Diversity in Antisense Technology
4 Oligonucleotides in Genotyping Polymorphisms
  4.1 Gel-Based Genotyping for Detecting Known Polymorphisms
  4.2 Non-Gel-Based High-Throughput Genotyping Technologies
5 Conclusion
  References

## 1 INTRODUCTION

The human genome is estimated to consist of 30,000 genes, up to 20% of which are considered to be expressed in a cell at a certain time. The human genome has now been completely sequenced, and huge amounts of data are being presented in public databases [1, 2]. However, these data do not provide information about the function of the genes or the significance of their expression in the specific state of

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

a disease. Thus, to fully exploit the massive DNA sequence information, smart technologies are necessary for rigorous analysis of gene expression. The collection of genes that are transcribed from genomic DNA, referred to as the "transcriptome," is a major determinant of cellular phenotype and function. The transcription of genomic DNA to produce mRNA is the first step for protein synthesis, and differences in the process are responsible for both morphological and phenotypic differences as well as indicative of cellular responses to environmental stimuli and perturbations. Unlike the genome, the transcriptome is highly dynamic and changes rapidly and dramatically in response to perturbations or during normal cellular events [3, 4]. Understanding the function of genes with respect to the time, location, and degree of expression is central to elucidating the activity and biological roles of its encoded protein. In addition, changes in the multigene expression patterns can provide clues about regulatory mechanisms and broader cellular functions and biochemical pathways. Thus, the knowledge gained from these types of measurements through implementation of proper and efficient technology can help determine the causes and consequences of disease, how drugs and drug candidates work in cells and organisms, and what gene products might have therapeutic uses or may be appropriate targets for therapeutic intervention.

Oligonucleotides have emerged as essential tools in pharmaceutical biotechnology by virtue of their small size, low-cost chemical synthesis, and precise base pairing capability with other complementary sequences either in the transcriptome or in the genome, according to the Watson–Crick model. Functional genomics-based screening, identification, and validation of drug targets, development of target protein-specific new generation therapeutics, and genotyping of many polymorphisms to develop tailor-made drugs for individuals are all being accomplished by many smart technological appro-aches using synthetic oligonucleotides of variable sizes and sequences.

## 2 OLIGONUCLEOTIDES IN FUNCTIONAL GENOMICS FOR DRUG TARGET DISCOVERY

The classic sciences of chemistry, pharmacology, microbiology, and biochemistry have shaped the course of drug discovery. Additionally, molecular biology has exerted a profound influence on drug discovery, allowing the concept of genetic information to be dealt with in very concrete chemical and biochemical terms. Although the earlier influence of molecular biology was restricted to cloning and expressing genes that encode therapeutically useful protein drugs (mostly recombinant proteins and monoclonal antibodies), recently the main promise of molecular biology for drug discovery lies in the potential to understand disease processes at the molecular genetic level and determine the optimal targets of drug intervention. The next step is to design novel pharmaceutical compounds either by high-throughput screening (HTS), a process in which batches of compounds are tested for binding activity or biological activity against target molecules, or simulation of drug–target interactions in order to design a drug to fit a target structure precisely.

However, the molecular biology techniques that are most commonly used for drug target discovery are fully or partially based on synthetically designed oligo-

nucleotides for detection and quantitation of specific mRNAs as representatives of target proteins.

### 2.1 Northern Blotting

Northern analysis remains a standard method for detection and quantitation of mRNA levels despite the advent of powerful techniques, such as reverse transcriptase–polymerase chain reaction (RT–PCR), real-time PCR, and microarray. It is used to quantitate a specific mRNA, determine transcript size, detect alternative splice variants of a gene, and identify closely related species. RNA samples are first separated by size via electrophoresis in an agarose gel under denaturing conditions, transferred to a membrane, cross-linked, and hybridized with a labeled probe, such as oligonucleotide (cDNA or RNA can also be applicable). Sensitivity can be improved with oligo deoxythymidine (dT) selection for enrichment of mRNAs because the dT oligonucleotide anchor prime [oligo(dT)] in a column hybridizes to the 3′-poly(A) tail of the mRNA. The radiolabeled or nonisotopically labeled probe has a sequence complementary to that of target mRNA facilitating efficient hybridization between them. Nonisotopic labeling of oligonucleotides can be broadly classified into two categories: indirect and direct. The indirect method involved the incorporation of a hapten into the probe molecule that is subsequently detected by affinity recognition with an antibody or a binding moiety, such as streptavidin. The binding moiety is, in turn, coupled to a signaling enzyme, such as horseradish peroxidase. Direct methods, on the other hand, covalently attach the signaling enzymes to the probe molecules. Factors that influence the hybridization efficiency include temperature, ionic strength, destabilizing agents, mismatched base pairs, duplex length, viscosity, and base composition. High salt favors hybridization reactions (i.e., less specificity and greater background), whereas decreased salt and/or increased detergent or temperature increases hybridization specificity and reduces background.

### 2.2 Polymerase Chain Reaction (PCR)

The PCR is a test-tube system for DNA replication that allows a "target" DNA sequence to be selectively amplified several million-fold in just a few hours. During PCR, high temperature is used to separate the DNA molecules into single strands, and synthetic sequences of single-stranded DNA (20–30 nucleotides) serve as primers to enable DNA polymerase to synthesize the complementary strand. Two different primer sequences are used to bracket the target region to be amplified. One primer is complementary to one DNA strand at the beginning of the target region; a second primer is complementary to a sequence on the opposite DNA strand at the end of the target region. Thus, PCR technology consists of three steps: denaturation of double-stranded target DNA at 94–96°C, annealing (hybridization) of primers to their complementary sequences on either side of the target sequence at 50–65°C and finally extending a complementary DNA strand from each primer with the help of a heat-stable DNA polymerase at 72°C. As amplification proceeds, the DNA sequence between the primers doubles after each cycle. After 30 such cycles, a theoretical amplification factor of one billion is attained.

For the discovery of drug target proteins based on functional genomics, some advanced PCR techniques are now playing the vital roles.

***Reverse Transcriptase (RT)–PCR.*** RT–PCR is the most sensitive technique currently available for mRNA detection and quantitation. Compared with the two other commonly used techniques (Northern blot analysis and RNase protection assay) for quantifying mRNA levels, RT–PCR can be used to quantify mRNA levels from much smaller samples. In fact, this technique is sensitive enough to enable quantitation of RNA from a single cell. RT–PCR is a technique in which an RNA strand is "reverse" transcribed into its DNA complement, followed by amplification of the resulting DNA using PCR. Transcribing an RNA strand into its DNA complement is termed reverse transcription (RT), and it is accomplished through the use of an RNA-dependent DNA polymerase (reverse transcriptase). Afterward, a second strand of DNA is synthesized through the use of a deoxy-oligonucleotide primer and a DNA-dependent DNA polymerase. The complementary DNA and its antisense counterpart are then exponentially amplified via PCR. The original RNA template is degraded by RNase H leaving behind DNA only.

***Real-Time PCR.*** Over the last several years, the development of novel chemistries and instrumentation platforms enabling detection of PCR products on a real-time basis has led to widespread adoption of real-time RT–PCR as the method of choice for quantitating changes in gene expression. It enables quantitation of the initial amount of the template most specifically, sensitively, and reproducibly by monitoring the fluorescence emitted during the reaction as an indicator of amplicon production during each PCR cycle (i.e., in real time), and it is a preferable alternative to other forms of quantitative RT–PCR that detect the amount of final amplified product at the endpoint [5, 6]. Furthermore, real-time RT–PCR has become the preferred method for validating results obtained from array analyses and other techniques that evaluate gene expression changes on a global scale.

The real-time PCR system is based on the detection and quantitation of a fluorescent reporter [7, 8]. This signal increases in direct proportion to the amount of PCR product in a reaction. By recording the amount of fluorescence emission at each cycle, it is possible to monitor the PCR reaction during the exponential phase where the first significant increase in the amount of PCR product correlates to the initial amount of target template. The higher the starting copy number of the nucleic acid target, the sooner a significant increase in fluorescence is observed. A significant increase in fluorescence above the baseline value measured during the 3–15 cycles indicates the detection of accumulated PCR product.

Real-time quantitative PCR analysis can be performed with oligonucleotide-based hydrolysis probes or hybridization probes. Hydrolysis probes include TaqMan probes [9], molecular beacons [10, 11], and scorpions [12]. They use the fluorogenic 5′ exonuclease activity of Taq polymerase to measure the amount of target sequences in cDNA samples (Figure 1). TaqMan probes, Molecular Beacons, and Scorpions depend on fluorescence resonance energy transfer (FRET) to generate the fluorescence signal via the coupling of a fluorogenic dye molecule and a quencher moiety to the same or different oligonucleotide substrates.

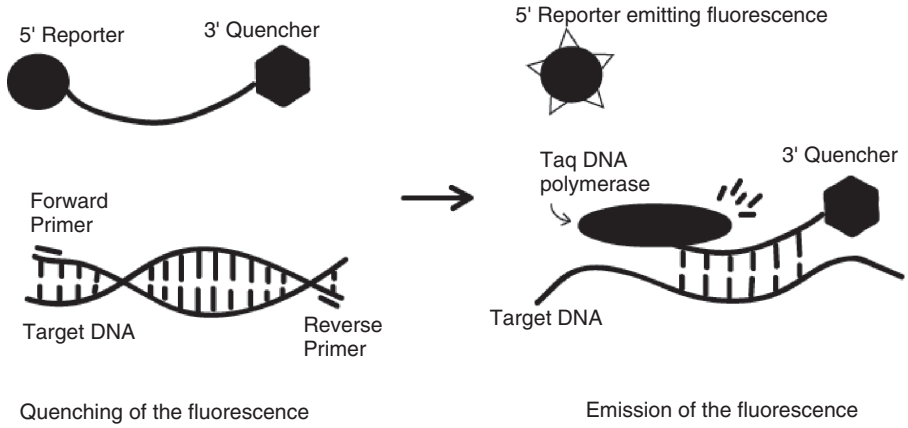


**Figure 1.** A model for dual-labeled fluorogenic probes for real-time PCR. The probe is labeled with two different dyes: A reporter dye is located on the 5′ end, and a quencher dye is located on the 3′ end. The quencher dye inhibits the natural fluorescence emission of the reporter dye by FRET. During the elongation step, Taq DNA polymerase 5′–3′ exonuclease activity hydrolyzes the probe bound to the specific DNA template, releasing the reporter dye from the target/oligonucleotide-quencher hybrid and causing an increase of fluorescence in proportion to the amount of target DNA.

### *Hydrolysis Probes*

TAQMAN PROBES. These are oligonucleotide probes longer than the primers (20–30 bases long with a Tm value of 10°C higher) that contain a fluorescent dye usually on the 5′ base and a quenching dye (usually TAMRA) typically on the 3′ base. When irradiated, the excited fluorescent dye transfers energy to the nearby quenching dye molecule in the FRET mechanism [13]. Thus, the close proximity of the reporter and quencher prevents emission of any fluorescence while the probe is intact. TaqMan probes are designed to anneal to an internal region of a PCR product. When the polymerase replicates a template on which a TaqMan probe is bound, its 5′ exonuclease activity cleaves the probe [14]. This ends the activity of quencher (no FRET), and the reporter dye starts to emit fluorescence, which increases in each cycle proportional to the rate of probe cleavage. Accumulation of PCR products is detected by monitoring the increase in fluorescence of the reporter dye. Because the cleavage occurs only if the probe hybridizes to the target, the origin of the detected fluorescence is specific amplification.

MOLECULAR BEACONS. Like TaqMan probes, molecular beacons also use FRET to detect and quantitate the synthesized PCR product via a fluorescent dye coupled to the 5′ end and a quencher attached to the 3′ end of an oligonucleotide substrate. Unlike TaqMan probes, molecular beacons are designed to remain intact during the amplification reaction, and they must rebind to the target in every cycle for signal measurement. Molecular beacons form a stem-loop structure when free in solution. Thus, the close proximity of the fluorescent and quencher dyes prevents the probe from fluorescing. When a molecular beacon hybridizes to a target, the fluorescent and quencher dyes are separated, FRET does not occur, and the fluorescent dye emits light upon irradiation.

SCORPIONS. With scorpion probes, sequence-specific priming and PCR product detection is achieved using a single oligonucleotide. The scorpion probe maintains a stem-loop configuration in the unhybridized state. The fluorophore is attached to the 5′ end and is quenched by a moiety coupled to the 3′ end. The 3′ portion of the stem also contains a sequence that is complementary to the extension product of the primer. This sequence is linked to the 5′ end of a specific primer via a nonamplifiable monomer. After extension of the Scorpion primer, the specific probe sequence can bind to its complement within the extended amplicon, thus opening the hairpin loop. This prevents the fluorescence from being quenched and a signal is observed.

*Hybridization Probes*

LIGHTCYCLER. A LightCycler probe (Roche Applied Science, Indianapolis, IN) is a pair of single-stranded fluorescent-labeled oligonucleotides. Oligo Probe I is labeled at its 3′ end with a donor fluorophore dye, and Oligo Probe 2 is labeled at its 5′ end with one of two available acceptor fluorophore dyes. The free 3′ hydroxyl group of Probe 2 must be blocked with a phosphate group to prevent Taq DNA polymerase extension. To avoid any steric problems between the donor and the acceptor fluorophores on both sides, there should be a spacer of one to five nucleotides to separate the two probes from each other. During the annealing step, the PCR primers and the LightCycler probes hybridize to their specific target regions, which causes the donor dye to come into close proximity to the acceptor dye. When the donor dye is excited by light from the LightCycler instrument, energy is transferred by FRET from the donor to the acceptor dye, which causes the acceptor dye to emit light at a longer wavelength than the light emitted by the instrument. The acceptor fluorophore's emission wavelength is detected by the LightCycler instrument's optical unit. The significant increase of measured fluorescence signal is directly proportional to the amount of target DNA.

***Differential Display (DD)–PCR.*** DD–PCR is a PCR-based technique that can detect differentially expressed genes. It provides a picture of the transcript pool of cells or tissues by displaying subsets of mRNAs. Subsets obtained from different cell types or tissues can be compared and used for isolation of the genes of interest. The general strategy for DD–PCR consists of the combination of (1) reverse transcription using a dT oligonucleotide anchor primer (oligo$(dT)_n$) to produce cDNA; (2) performing PCR using the cDNA as a template with the anchor primer and an arbitrary primer; and (3) separation of the PCR product by electrophoresis and visualization. The differential display method is thus far unique in its potential to visualize all expressed genes in a eukaryotic cell in a systematic and sequence-dependent manner by using multiple primer combinations.

***Serial Analysis of Gene Expression (SAGE).*** SAGE analysis is a method derived to provide a readout, via sequencing, of the spectrum of genes being expressed in a cell or tissue in both a qualitative and a quantitative manner [15, 16]. Three principles underlie the SAGE methodology:

1. A short sequence tags (10–15 bp) contains sufficient information to uniquely identify a transcript provided that the tag is obtained from a unique position within each transcript.

2. Sequence tags can be linked together to form long serial molecules that can be cloned and sequenced.
3. Quantification of the number of times a particular tag is observed provides the expression level of the corresponding transcript.

A detailed description of SAGE is provided in Figure 2.

Important advantages of SAGE over DD–PCR are that SAGE data are quantitative and cumulative. Accurate, quantitative transcript profiles describing the abundance of all genes expressed in a cell or tissue are generated by SAGE, provided sufficient sequencing is completed. The resulting data may then be compared with all existing and future SAGE databases constructed in a similar manner.

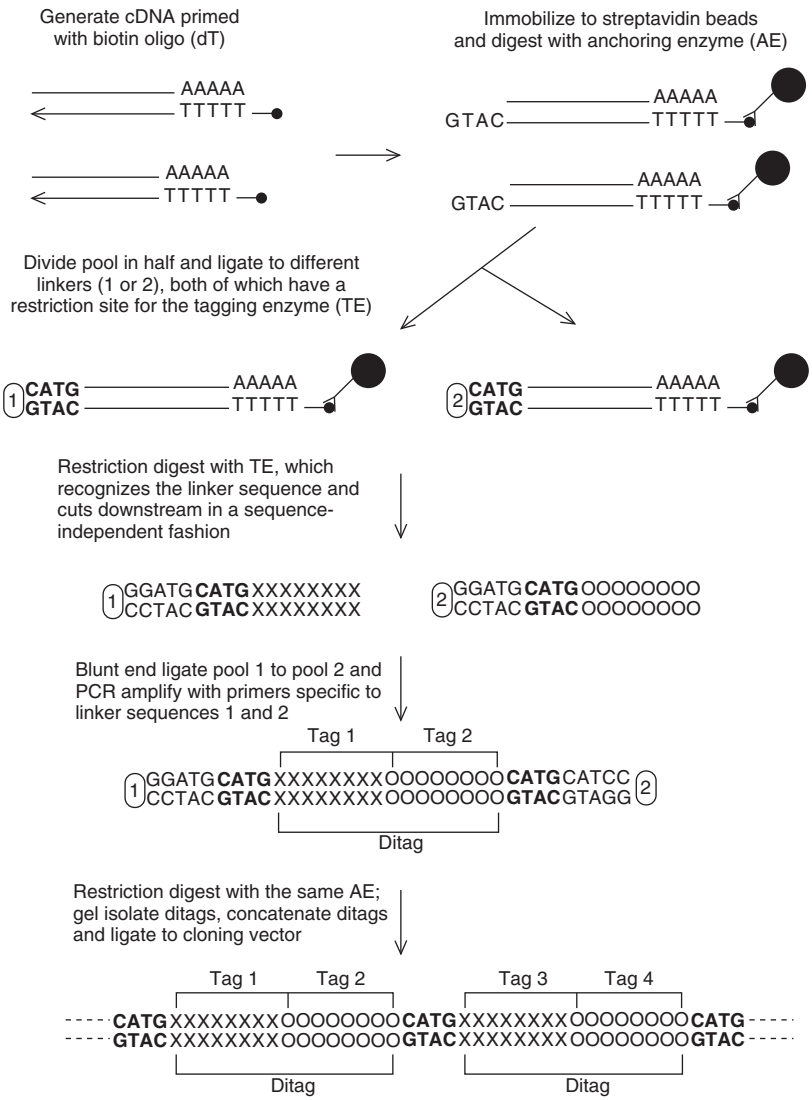


**Figure 2.** The steps involved in SAGE.

## 2.3 High-Density Oligonucleotide Arrays

High-density arrays of oligonucleotides are the most powerful and versatile tools for functional genomics. They work by hybridization of labeled RNA or DNA in solution to DNA molecules attached at specific locations on a surface. So DNA microarrays may be defined as miniaturized, ordered arrangements of nucleic acid fragments derived from individual genes located at defined positions on a solid support, which enables the analysis of thousands of genes in parallel by specific hybridization. DNA microarrays provide information on how several genes are abnormally regulated in a disease. For example, a microarray of 100 genes that have a role in inflammation was used to examine rheumatoid tissue. The analysis revealed the upregulation of genes encoding interleukin 6 and several matrix metalloproteinases [17]. The microarray technology will continue to contribute to the understanding of responses to drug treatments [18]. Based on the fabrication technique used, planar microarrays can be broadly classified into microarrays prepared by *in situ* synthesis of oligonucleotides and direct spotting of oligonucleotides on glass, membrane, or other derivatized surfaces by microspotting or inkjet printing methods.

***Microarrays Fabrication by in-situ Synthesis.*** A combination of photoli-thography and combinatorial chemistry is used to develop very high-density DNA chips by Affymetrix (http://www.affymetrix.com, Santa Clara, CA). Fused silica wafers are hydroxylated and then silanized. The silanes contain a linker molecule and a protective group that can be activated by light. The silane film provides uniform hydroxyl density to initiate the microarray fabrication process. Photolithographic masks are then used to illuminate specific locations on the silanized wafer. Ultraviolet (UV) exposure causes the deprotection of groups on the silane, after which the activated groups couple nucleotides onto the silica wafer. Additional capping steps prevent unattached molecules from becoming probes. The side chains on the nucleotides are protected to prevent branched-chain formation. These steps of deprotection, coupling, and capping are continued until probes of full length are fabricated (Figure 3).

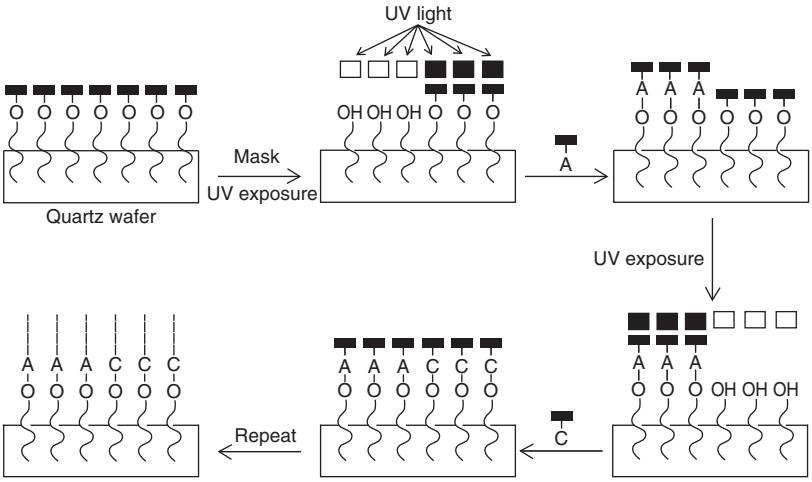


**Figure 3.** Fabrication of high-density oligonucleotide array by photoactivation and deprotection of nucleic acids. Photomasks are used to pattern UV light at localized regions to selectively synthesize patterned array.

In an alternative approach developed by NimbleGen (http://www.nimblegen.com/, Madison, WI), arrays are built using photodeposition chemistry based on a maskless array synthesizer (MAS). The MAS system is a high-density DNA fabrication instrument that uses a maskless light projector as a "virtual mask" instead of the physical chromium masks used by Affymetrix. The virtual mask is an array of hundreds of thousands of individually addressable aluminum mirrors on a computer chip. These mirrors function as virtual masks that reflect the desired pattern of UV light and are controlled by the computer.

*In situ* synthesized oligonucleotide probes generally do not exceed 25 bases in length because of reaction efficiency limits. As a result, mismatches and spurious target-probe binding can take place due to the limited specificity and binding affinity for a 25-residue oligonucleotide. To overcome this problem, a series of oligonucleotides that differ by a one-base mismatch from the gene-specific probe is also included on the array and can be used to determine the amount of mismatch hybridization, which can then be substracted from the signal [19, 20]. The Affymetrix arrays contain between 40,000 and 60,000 probes (including multiple mismatch controls for each gene) and provide the highest density of probes of any array.

***Microarrays Fabrication by Direct Spotting.*** Microarrays can be fabricated by directly spotting oligonucleotides on microscope glass slides, membranes, or other surfaces. Various types of spotters and derivatized substrate surfaces have been developed to fabricate microarrays [21–23]. The surface groups on the derivatized surface determine how the oligonucleotides will be immobilized on the surface. DNA molecules contain phosphate groups, which are commonly used for immobilization on positively charged surfaces. Amine surfaces carry a positive charge at neutral pH. Therefore, negatively charged DNA molecules form ionic bonds with the positively charged amine-derivatized surface. Often, UV or thermal treatment is used to covalently link the DNA molecules to amine-derivatized slides that are most suitable for immobilizing long oligonucleotides. Sometimes, to improve the binding efficiency, a longer chain spacer is attached to the aminated surface to provide sufficient separation between the amino groups and substrate surface. Poly-1-lysine-coated slides containing a dense layer of amino groups are similarly used to attach DNA molecules via ionic interactions. Additionally, amino-modified can be covalently immobilized on aldehyde- and epoxy-derivatized glass slides.

Microarray is now an essential tool in functional genomics for drug target discovery. However, in addition to the obvious use of functional genomics in basic research and target discovery, there are many other specific uses, including biomarker determination to find genes that correlate with and presage disease progression; pharmacology to determine differences in gene expression in tissues exposed to various doses of compounds; toxico-genomics to find gene expression patterns in a model tissue or organism exposed to a compound and their use as early predictors of adverse events in humans; target selectivity to define a compound by the gene expression patterns; prognostic tests to find a set of genes that accurately distinguishes one disease from another; and disease subclass determination to find multiple subcategories of tumors in a single clinical diagnosis [24].

## 3 OLIGONUCLEOTIDES IN DRUG TARGET VALIDATION AND POTENTIAL THERAPEUTICS

Because their sequence is complementary to that of mRNA, antisense oligonucleotides (ON) form a very powerful weapon for studying gene function (functional genomics) and for discovering new and more specific treatments of human diseases (antisense therapeutics). The binding or hybridization of antisense nucleic acid sequences to a specific mRNA target will interrupt normal cellular processing of the genetic message of a gene through at least three mechanisms: (1) the oligonucleotide: RNA duplex may form substrate for endogenous RNase H, leading to mRNA cleavage; (2) the duplex may prevent the productive assembly of the ribosomal complex preventing translation; and (3) the duplex may arrest a ribosomal complex already engaged in translation leading to a truncated protein [25]. However, the successful therapeutic application of antisense ONs has been delayed because of delivery problems. Natural ONs are not membrane-permeable, and even after delivery into cells with the help of amphipathic cations or liposomes, antisense ONs can still be hydrolyzed by cellular nucleases before reaching their intended target mRNAs. To improve their bioavailability, various types of chemically modified antisense ONs have been produced [26–29]. However, chemical modification has also led to undesirable complications such as decreased sequence-specificity and general toxicity. On the other hand, the discovery of RNA interference (RNAi) started a new era in antisense technology [30–32]. Double-stranded short interfering RNAs (siRNA) could silence their target mRNA almost 100 times more effectively than single-stranded ONs [33]. However, delivery of siRNA is still a problem as successful administration often requires the use of plasmids or viral vectors, leading to the risk of random integration into chromosomal DNA, which is a serious threat to therapeutic safety [34].

### 3.1 Diversity in Antisense Technology

Many types of chemically modified antisense ONs have been manufactured with the changes in the overall electronic charge as well as the incorporation of non-phosphate oligonucleotide backbones. They are commonly grouped into three "generations" based on the type of modifications (Figure 4).

***First-Generation Antisense ONs.*** These ONs were manufactured through the substitution of one nonbridging oxygen atom in the phosphate group of ONs with either a sulphur or a methyl group. Those DNA analoges with a sulfur group are known as phosphorothioate ONs, which is the most widely used antisense ON to date. The main advantages of phosphorothioate ONs are resistance against nucleases, ability to recruit RNase H to cleave target mRNA, ease of synthesis, and attractive pharmacokinetic properties. The first U.S. Food and Drug Administration (FDA)-approved antisense drug, Vitravene from Isis (Carlsbad, CA), which treats a condition called cytomegalovirus (CMV) retinitis in people with AIDS, and the majority of antisense compounds in clinical trials to date are based on this chemical design.

***Second Generation Antisense ONs.*** These contain nucleotides with alkyl modification at the 2′ position of the ribose. 2′-O-methyl and 2′-O-methoxyethyl

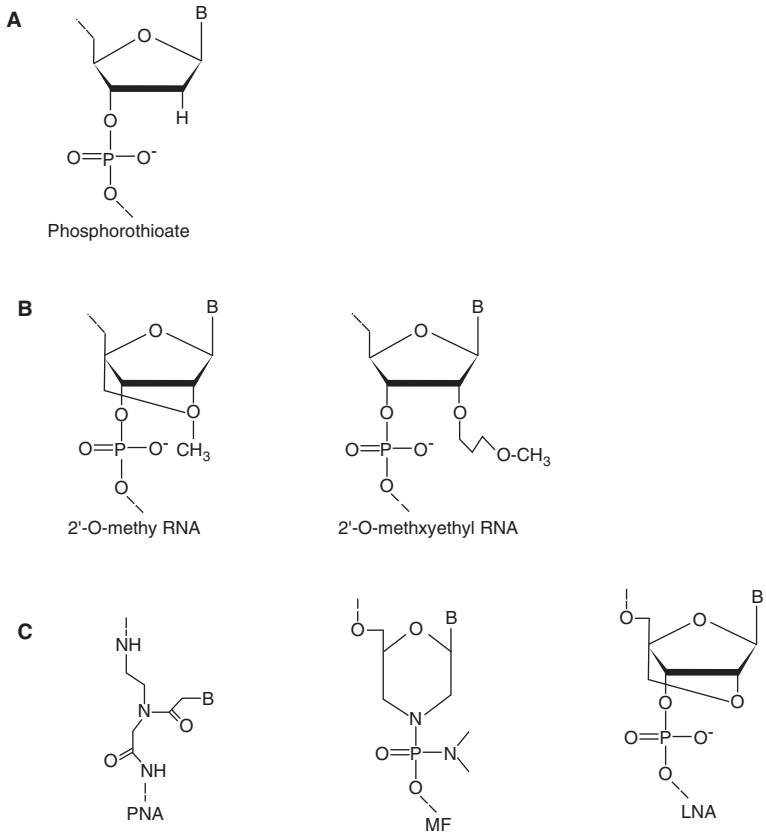


**Figure 4.** Nucleotide modifications for use in antisense technology. (A) The first-generation phosphorothioate backbone modification. (B) The second ribonucleotides modified at the 2′ hydroxyl by adding a methyl or methoxy–ethyl group. (C) The third-generation modifications involving the complete replacement of phosphodiester linkages as in PNA and MF or the conformational lock as in LNA.

RNAs are the most important representatives. The significant improvements are a reduction in general toxicity, increased hybrid stability, and increased nuclease resistance. However, the only disadvantage is that 2′-O-methyl RNA cannot induce the RNase H cleavage of the target mRNA. To induce RNase H cleavage, mixed backbone ONs (MBOs) were developed by surrounding a phosphorothioate-modified deoxyribose core that retains RNase H activity with nuclease-resistant arms, such as 2′-O-methyl ribonucleosides [35, 36]. The antisense activity, pharmacokinetics, *in vivo* degradation, and safety profile of an MBO can be modulated by combining appropriate oligonucleotide segments and backbone modifications at defined sites. GEM231, which targets the RI(alpha) subunit of protein kinase A (PKA), is built on an MBO platform and is being developed by Hybridon (Cambridge, MA) as a cancer therapeutic [37]. It is an 18-mer oligonucleotide with four 2′-O-methyl ribonucleosides at both the 3′- and the 5′-ends surrounding the remaining deoxynucleosides, all with phosphorothioate internucleotide linkages.

***Third-Generation Antisense ONs.*** These are DNA and RNA analoges with modified phosphate linkages or ribose as well as nucleotides with a completely different chemical moiety replacing furanose ring. Peptide nucleic acids (PNAs), morpholino phosphoroamidates (MFs), and locked nucleic acids (LNAs) are three interesting RNA/DNA analoges in this class. These compounds are essentially nuclease resistant while maintaining good hybridization affinity with their complementary mRNA.

PNA and MF possess the structures where the phosphodiester linkages are completely replaced, respectively, with a polyamide (peptide) and phosphoroamidate backbone (Figure 4) [38, 39]. They both form tight bonds with their RNA targets and probably exert their effects by blocking translation, as neither molecule effectively activates RNase H. Noncharged backbones prevent PNA and MF from binding to proteins that normally recognize polyanions in a nonspecific manner. However, due to the electrostatically neutral property, solubility and cellular uptake are serious problems for both PNA and MF. The ability of PNA to recognize duplex DNA makes PNA more promising candidates for modulating gene expression or inducing mutations by strand invasion of chromosomal duplex DNA [40] because the PNA:DNA duplex is more stable than a DNA:DNA duplex [41]. Effective gene knockdown has also been shown by antisense MF, and one such MF targets the *c-myc* oncogene [42] and is being developed by AVI BioPharma (Portland, OR).

LNAs [also known as bridged nucleic acids (BNAs)] are another type of novel high-affinity molecules that provide major improvements in several key properties. They contain a methylene bridge connecting the 2′-oxygen of the ribose with the 4′-carbon (Figure 4), leading to improved binding to complementary DNA and RNA sequences [43]. Although the 2′–4′ linkage reduces or eliminates activation of mRNA cleavage by RNase H, it is straightforward to synthesize chimeric "gapmers," in which a central DNA portion is flanked by LNA to enhance the stability of binding and enable recruitment of RNase H. LNAs have been successfully used to suppress the expression of an RNA polymerase, resulting in the inhibition of tumor growth in a xenograft model [44].

***siRNAs and Traditional Antisense Technology.*** siRNA are the effector molecules of the RNAi pathway in which endogenously produced double-stranded RNA is cleaved by the Dicer enzyme within the cells into 21- to 28-nucleotide siRNA for targeted degradation of complementary mRNA through the formation of RNA-induced silencing complexes (RISCs). siRNA can be synthetically produced or expressed from vectors transcribing short double-stranded hairpin-like RNAs that are processed into siRNAs inside the cell. Unlike ONs, siRNA cannot effectively target pre-mRNA for degradation in mammalian cells [45]. However, siRNA is now a reliable gene-silencing tool in functional genomics for the validation of potential drug targets identified by the techniques described above. The low concentration of siRNA required to elicit effective gene silencing and the fact that siRNAs are specifically and rapidly incorporated into RISC diminish the potential for the nonspecific binding with proteins as happened in phosphorothioate-modified ONs, which makes them toxic at a high concentration. However, it is important to identify effective siRNAs so that the lowest possible concentration of siRNA can be allowed for gene silencing [46–48]. If siRNAs are to be used for therapeutic

purposes, methods must be developed that protect the possible side effects of siRNAs by either modification of the siRNA backbone [49–51] or avoidance of immunostimulatory sequence motifs [51, 52].

## 4 OLIGONUCLEOTIDES IN GENOTYPING POLYMORPHISMS

Single nucleotide polymorphisms (SNPs), the most frequent DNA sequence variations found in the human genome (about 1 per 350-bp frequency), are being uncovered and assembled into large SNP databases that promise to enable the dissection of the genetic essence of disease and drug response [53–57]. The identification of a complex set of genes that cause a disease will require both linkage and association analyses of thousands of polymorphisms across the human genome in thousands of individuals. In addition, certain genetic polymorphisms cause significantly different responses among individuals on exposure to a particular drug [58]. Understanding the role of genetic polymorphisms in drug responses will help to increase drug efficacy and decrease adverse effects by tailoring medications to patients' genetic makeup. Advances in the field, which is known as "pharmacogenetics," have important clinical implications and practical value for the design of dosing regimens. There are two broad areas in pharmacogenetics, specifically pharmacokinetics (PK) describing drug adsorption, distribution, metabolism and elimination, and pharmacodynamics (PD) elucidating the pharmacological effects of a drug on the body (either desired or undesired) and differences in either PK or PD can lead to variable drug efficacy or toxicity risk [59]. The Human Genome Project has provided a windfall of sequence polymorphism data, and because of the collaborative SNP discovery initiatives such as the SNP Consortium (TSC), millions of human SNPs have been catalogued, many of which are publicly available in the TSC and NCBI dbSNP repositories (http://snp.cshl.org; http://www.ncbi.nlm.nih.gov/SNP/). The high-density SNP map will allow researchers to expand their capabilities for identification of critical diseases and drug-response genes. Genotyping SNPs in large-scale pharmacogenetic studies is an integrated part of the drug discovery and development process. Advanced technologies to identify genetic polymorphisms rapidly, accurately, and economically are becoming a priority in the implementation of pharmacogenetics to drug development, clinical trials, and clinical monitoring for drug efficacy and toxicity [60]. SNP genotyping is now performed by several advanced technologies, most of which are based on synthetic oligonucleotide primers or probes.

### 4.1 Gel-based Genotyping for Detecting Known Polymorphisms

***PCR–Restriction Fragment Length Polymorphism (RFLP) Analysis.*** Commonly used methods include gel-electrophoresis-based techniques such as PCR coupled with RFLP analysis. Specific regions of DNA sequences can be PCR amplified by using specific primers. The PCR products are then digested with appropriate restriction enzymes and visualized by staining the gel after electrophoresis. If the genetic polymorphism produces a gain or loss of the restriction site, a different restriction digestion pattern will be obtained [60]. A major limitation is the requirement that the polymorphisms alter a restriction enzyme cutting site [61].

***Oligonucleotide Ligation Assay (OLA) Genotyping.*** OLA relies on hybridization with specific oligonucleotide probes that can effectively dis-criminate between the wild-type and variant sequences. Three oligonucleotide probes are used in OLA: two allele-specific probes (one specific for the wild-type allele and the other specific for the mutant allele) and a fluorescent common probe. The 3′ ends of the allele-specific probes are immediately adjacent to the 5′ end of the common probe. The gene fragment containing the polymorphic site is amplified by PCR and incubated with the probes. In the presence of thermally stable DNA ligase, ligation of the fluorescent-labeled probe to the allele-specific probe(s) occurs only when there is a perfect match between the variant or the wild-type probe and the PCR product template. These ligation products are then separated by electrophoresis, thus enabling the identification of the wild-type genotypes, the variants, the heterozygotes, and the variants (Figure 1). By varying the combinations of color dyes and probe lengths, multiple mutations can be detected in a single reaction [62].

### 4.2 Non-Gel-Based High-Throughput Genotyping Technologies

***TaqMan Genotyping.*** The assay is based on TaqMan real-time PCR technology as described above. However, SNP genotyping requires two TaqMan probes that differ at the polymorphic site, with one probe comple-mentary to the wild-type allele and the other to the variant allele. A 5′ reporter dye and a 3′ quencher dye are covalently linked to both wild-type and variant allele probes. During the PCR annealing step, the TaqMan probes hybridize to the targeted polymorphic site. During the PCR extension phase, only the perfectly hybridized probe will be cleaved by the 5′ nuclease activity of the *Taq* polymerase, leading to an increase in the characteristic fluorescence of the reporter dye. A mismatched probe will not be recognized by the *Taq* polymerase, resulting in the quenching of fluorescence due to the physical proximity of the reporter and quencher dyes.

***Molecular Beacons.*** As described, the ends of molecular beacons are complementary to each other, forming a stem structure, whereas the intervening loop is complementary to a sequence within the PCR amplified product. When hybridized to the right target sequence, the two fluorescent dyes (donor and acceptor) at opposite ends of the probes are separated and the fluorescence is increased dramatically [63]. Mismatched probetarget hybrids dissociate at a substantially lower temperature than exactly complementary hybrids. This thermal instability of mismatched hybrids increases the specificity of molecular beacons. For SNP genotyping, two molecular beacons with exact sequence matches to the wild-type and variant alleles are used in the same PCR reaction. These two molecular beacons are labeled with different fluorophores that emit fluorescent light at distinct optical wavelengths. The use of two differentially labeled molecular beacons in the same PCR reaction allows simultaneous determination of three possible allelic combinations.

***Invader Assay.*** The Invader assay developed by Third Wave Technologies, Inc. (Madison, WI) is an attractive FRET-based technique to genotype SNPs without PCR amplification (Figure 5). Two oligonucleotides, a wild-type or variant signal probe plus an upstream Invader probe, are used in each reaction. These probes

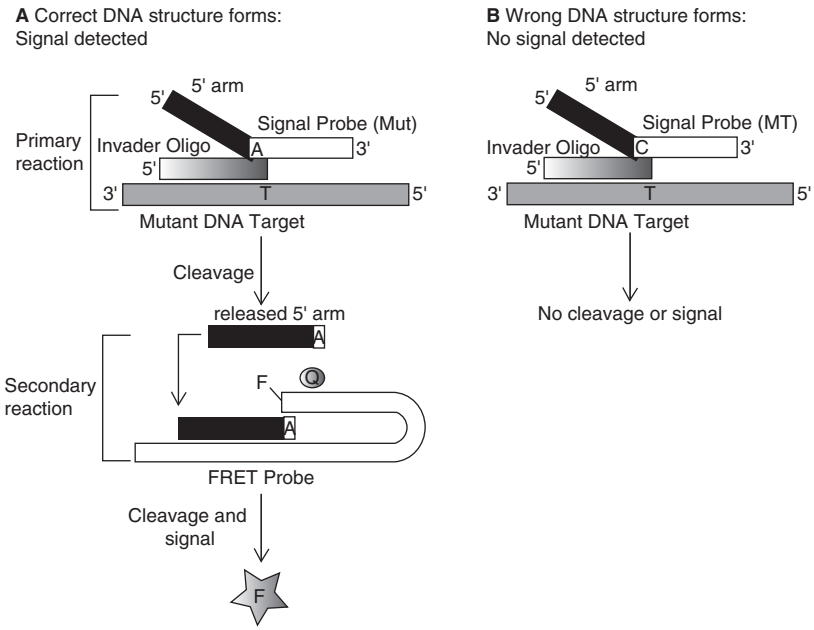


**Figure 5.** Invader SNP assay.

hybridize in tandem to a specific region of genomic DNA. When the 3′ end of an upstream oligonucleotide (Invader probe) overlaps the hybridization site of the 5′ end of a downstream oligonucleotide (signal probe) by at least one base pair, the structure will be recognized and cleaved by Cleavase [64, 65]. A single-nucleotide mismatch immediately upstream of the cleavage site renders the conformation unrecognized by Cleavage. Each cleavage product then serves as an invader oligonucleotide in a secondary reaction where it directs the cleavage of a combined labeled FRET probe-template construct. The secondary oligonucleotide probe is 5′ end-labeled with the donor flurophore, which is quenched by an internal acceptor dye. When the DNA is cleaved, the donor and acceptor dyes are no longer in close proximity, the quenching is abolished and fluorescence is generated. The whole process repeats multiple times, producing linear amplification of the fluorescent signal. Assays are read with a fluorescence plate reader and genotypes are assigned after determination of the net wild-type/variant signal ratio for each sample [64, 66].

***FRET-based PCR–OLA.*** The three dye-labeled ligation probes for each SNP are designed to have low melting temperatures. A 5′ donor dye-labeled common probe terminates one base immediately upstream from the polymorphic site. Two allele-specific 5′-phosphorylated, 3′-acceptor dye-labeled probes have polymorphic nucleotides at the 5′ end. A thermostable DNA polymerase with no 5′ nuclease activity and a thermostable DNA ligase are used. The first stage of PCR reaction is kept at high temperature, and the ligation probes cannot to anneal. After sufficient PCR products are generated, the second stage of the reaction, with a low annealing temperature, allows ligation to occur. By analyzing the fluorescence signals of all dyes, individual genotypes can be determined directly after one reaction using

real-time PCR or by endpoint signal analysis using a fluorescent plate reader [67].

***Rolling Circle Amplification (RCA).*** Like the Invader assay, RCA is also sensitive enough to work directly from genomic DNA [68, 69]. For each SNP, two linear, allele-specific probes are designed that circularize when they anneal to the target sequence. Each probe consists of a single oligonucleotide, 80–90 bases in length. The 5′-end of the probe is phosphorylated and bears a sequence of 20 nucleotides that will hybridize to the region immediately 5′ of the SNP. The 3′-end of the probe contains 10–20 nucleotides complementary to the region immediately 3′ of the SNP. Both allele-specific probes are identical with the exception of the 3′-base, which is varied to complement the polymorphic site. Sandwiched between the allele-specific probe arms is a generic backbone sequence of 40–50 nucleotides that provides binding sites for two RCA amplification primers to be used in the second stage of the assay.

The first stage of the assay is the allele discrimination step. Target genomic DNA is denatured to make it single stranded and then hybridized with a pair of single-allele-specific, open-circle oligonucleotide probes resulting in circularization of the probe. Allele discrimination is performed by a thermophilic DNA ligase. Ligation of the probe will preferentially occur when the 3′-base exactly complements the target DNA. If there is a mismatch between the 3′-allele-specific base and the target, the ligation product will be produced at a much-reduced rate. In a homozygous sample, the ligation of one allele-specific probe predominates over the other, and in a heterozygous sample, both probes are ligated with equal efficiency.

Subsequent detection of the circularized oligonucleotide probes is by RCA, using the two non-cross-reacting primers. The first primer is complementary to 20 bases of the circularized oligonucleotide probe in the backbone region and is extended by an exonuclease-deficient DNA polymerase. The extending probe eventually displaces itself at its 5′-end once one revolution of the circularized probe is completed. Continued polymerization and displacement result in the generation of a single-stranded product of complementary concatamers. The second primer is the same sense as the probe. It binds to each tandem repeat of the first strand product at sites corresponding to a second region of the backbone, where it primes DNA synthesis at each site. As these multiple binding events elongate, strand displacing activity of the DNA polymerase causes these nascent strands to be displaced and a multitude of new recognition sites for the first primer are exposed in a process known as branching. Eventually, strand displacement results in the release of double-stranded DNA fragments from the parent molecule, thus generating up to $10^9$ copies of each circular oligonucleotide probe under isothermal conditions. A FRET-based RCA assay called SNIPER has been developed by Amersham Pharmacia Biotech. [70], in which the second amplification primer is an Amplifluor labeled oliogonucleotide with a 5′-hairpin loop containing FRET dyes. As the complementary strands to the hairpin primers are synthesized, the polymerase reads through the stem of the hairpin loop and opens out the 5′-end. In this linearized conformation, the fluorophore is no longer quenched and fluorescence can be detected. Probes that fail to hybridize do not give fluorescence signal in solution, eliminating the need for purification or separation steps, which makes automation of the RCA assay easier.

***DNA Microarray Genotyping.*** A high-density oligonucleotide microarray offers simultaneous analysis of many polymorphisms. The DNA sample of interest is PCR amplified to incorporate fluorescently labeled nucleotides and then hybridized to hundreds of thousands of oligonucleotides attached to a solid silicon surface in an ordered array. Each oligonucleotide in the high-density array acts as an allele-specific probe. Perfectly matched sequences hybridize more efficiently to their corresponding oligomers on the array and, therefore, give stronger fluorescent signals over mismatched probe-target combinations [71, 72]. The hybridization signals are quantified by high-resolution fluorescent scanning and analyzed by computer software.

Sequencing by hybridization can be an efficient method to monitor many SNPs. It is possible to array a set of short oligonucleotides covering the entire DNA fragment on a DNA chip. Because the precise sequence of the oligonucleotide at each location on the chip is known, the pattern of hybridization can be determined using fluorescently labeled DNA probes. The advantage of this method is that a large DNA fragment or a large collection of small PCR products can be scanned in one hybridization. More than 100 such arrays have been used for high-throughput screening of SNPs covering 2.3 Mb of genomic DNA [73] and thousands of SNPs have been screened rapidly by use of chip-based resequencing [74].

As mentioned, almost all currently available SNP genotyping starts with a locus-specific amplification step, followed by an allele discrimination step [59, 73, 75]. At capacities of 1000 SNPs or less, locus-specific amplification is economically feasible in terms of oligonucleotide synthesis and other reagent costs. At capacities of 10,000 SNPs and greater, the costs of designing, synthesizing, and managing such an enormous number of oligonucleotides become prohibitive. Additionally, large amounts of starting sample DNA are required to genotype tens of thousands of SNPs in a locus-specific manner. Thus, considering the high complexity of human genome, the most common strategy is now to reduce the complexity by amplifying portions of the genomic DNA by PCR. This can be done in a random or semi-random fashion, for example, by using restriction enzyme-based adapter ligation PCR [76]. This representational approach, which is used to detect changes in genomic copy number, has also been applied to genotyping of $10^4$–$10^5$ or more SNPs from a single sample preparation [77, 78]. Genomic complexity reduction assay involves five primary steps, starting with restriction digestion, ligation of adaptor, amplification, fragmentation, and labeling, before hybridization to the oligonucleotde array [78]. Highly multiplexed PCR-based approaches have recently been developed to genotype specific loci of interest, enabling targeted genotyping of $10^3$–$10^4$ or more SNPs from a single sample [79]. These targeted genotyping approaches include a pre-PCR enzymatic SNP scoring step, the addition of universal priming sites by ligation, and subsequent universal primer PCR. Today, almost all high-throughput genotyping is done using one of these methods [80].

## 5 CONCLUSION

The automated oligonucleotide synthesizers now enable large-quantity, low-cost synthesis of precisely designed different sequences of oligonucleotide based in a parallel fashion. On the other hand, depending on the specificity of sequences for

precise base pairing to each other, oligonucleotides have the enormous potential to facilitate manufacturing of innovative numerous molecular machinery with the help of molecular biology enzymes, for being used as genomics tools as well as therapeutic tools. It should be emphasized that the new high-throughput approaches, such as microarrays, do not replace the conventional standard methods. The standard methods, such as northern blots or RT–PCR, are simply used in a more targeted fashion to complement the broader measurements and to follow up on the genes, pathways, and mechanisms implicated by array results.

## REFERENCES


1. Lander E S, et al. (2001). Initial sequencing and analysis of the human genome. *Nature*. 409:860–921.
2. Venter J C, et al. (2001). The sequence of the human genome. *Science*. 291:1304–1351.
3. Spellman P T, et al. (1998). Comprehensive identification of cell cycle-regulated genes of the yeast *Saccharomyces cerevisiae* by microarray hybridization. *Mol. Biol. Cell*. 9:3273–3297.
4. Lockhart D J, Winzeler E A (2000). Genomics, gene expression and DNA arrays. *Nature*. 405:827–836.
5. Freeman W M, Walker S J, Vrana K E (1999). Quantitative RT-PCR: Pitfalls and potential. *Biotechniques*. 26:112–122, 124–125.
6. Raeymaekers L (2000). Basic principles of quantitative PCR. *Mol. Biotechnol*. 15:115–122.
7. Lee L G, Connell C R, Bloch W (1993). Allelic discrimination by nick-translation PCR with fluorogenic probes. *Nucleic Acids Res*. 21:3761–3766.
8. Livak K J, Flood S J, Marmaro J, et al. (1995). Oligonucleotides with fluorescent dyes at opposite ends provide a quenched probe system useful for detecting PCR product and nucleic acid hybridization. *PCR Methods Appl*. 4:357–362.
9. Heid C A, Stevens J, Livak K J, et al. (1996). Real time quantitative PCR. *Genome Res*. 6:986–994.
10. Tan W, Wang K, Drake T J (2004). Molecular beacons. *Curr. Opin. Chem. Biol*. 8:547–553.
11. Vet J A M, Marras S A E (2004). Design and optimization of molecular beacon real-time polymerase chain reaction assays. *Oligonucleotide Syn. Meth. Appl*. 288: 273–290.
12. Solinas A, et al. (2001). Duplex Scorpion primers in SNP analysis and FRET applications. *Nucleic Acids Res*. 29:e96.
13. Chen X, Zehnbauer B, Gnirke A, et al. (1997). Fluorescence energy transfer detection as a homogeneous DNA diagnostic method. *Proc. Natl. Acad. Sci. U.S.A*. 94:10756–10761.
14. Holland P M, Abramson R D, Watson R, et al. (1991). Detection of specific polymerase chain reaction product by utilizing the 5′–3′ exonuclease activity of Thermus aquaticus DNA polymerase. *Proc. Natl. Acad. Sci. U.S.A*. 88:7276–7280.
15. Velculescu V E, Zhang L, Vogelstein B, et al. (1995). Serial analysis of gene expression. *Science*. 270:484–487.
16. Tuteja R, Tuteja N (2004). Serial analysis of gene expression: Applications in Human studies. *J. Biomed. Biotechnol*. 2:113–120.

17. Heller R A, et al. (1997). Discovery and analysis of inflammatory disease-related genes using cDNA microarrays. *Proc. Natl. Acad. Sci. U. S. A*. 94:2150–2155.
18. Lennon G L (2000). High-throughput gene expression analysis for drug discovery. *Drug Discov. Today*. 5:59–66.
19. Lockhart D J, et al. (1996). Expression monitoring by hybridization to high-density oligonucleotide arrays. *Nat. Biotechnol*. 14:1675–1680.
20. Wodicka L, Dong M, Mittmann M H, et al. (1997). Genome-wide expression monitoring in Saccharomyces cerevisiae. *Nat. Biotechnol*. 15:1359–1366.
21. Bowtell D D (1999). Options available–from start to finish–for obtaining expression data by microarray. *Nat. Genet*. 21(suppl 1):25–32.
22. Jain K K (2000). Applications of biochip and microarray systems in pharmacogenomics. *Pharmacogenom*. 1:289–307.
23. Okamoto T, Suzuki T, Yamamoto N (2000). Microarray fabrication with covalent attachment of DNA using bubble jet technology. *Nat. Biotechnol*. 18:384–385.
24. Butte A (2002). The use and analysis of microarray data. *Nat. Rev. Drug. Discov*. 1:951–960.
25. Dagle J M, Weeks D L (2001). Oligonucleotide-based strategies to reduce gene expression. *Differentiation*. 69:75–82.
26. Kurreck J (2003). Antisense technologies. Improvement through novel chemical modifications. *Eur. J. Biochem*. 270:1628–1644.
27. Cho-Chung Y S, et al. (1997). Antisense-protein kinase A: A single-gene-based therapeutic approach. Antisense *Nucleic Acid Drug Dev*. 7:217–223.
28. Srivastava R K, et al. (1999). Growth arrest and induction of apoptosis in breast cancer cells by antisense depletion of protein kinase A-R1 alpha subunit: p53-independent mechanism of action. *Mol. Cell. Biochem*. 195:25–36.
29. Wang H, et al. (1999). Antitumor activity and pharmacokinetics of a mixed-backbone antisense oligonucleotide targeted to the R1alpha subunit of protein kinase A after oral administration. *Proc. Natl. Acad. Sci. U.S.A*. 96:13989–13994.
30. Elbashir S M, et al. (2001). Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. *Nature*. 411:494–498.
31. Fire A, et al. (1998). Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. *Nature*. 391:806–811.
32. Hamilton A J, Baulcombe D C (1999). A species of small antisense RNA in posttranscriptional gene silencing in plants. *Science*. 286:950–952.
33. Miyagishi M, et al. (2003). Comparison of the suppressive effects of antisense oligonucleotides and siRNAs directed against the same targets in mammalian cells. Antisense *Nucleic Acid Drug Dev*. 13:1–7.
34. Stevenson M (2004). Therapeutic potential of RNA interference. *N. Engl. J. Med*. 351:1772–1777.
35. Agrawal S, Zhao Q (1998). Mixed backbone oligonucleotides: Improvement in oligonucleitide-induced toxicity in vivo. Antisense *Nucleic Acid Drug Dev*. 8:135–139.
36. Shen L X, et al. (1998). Impact of mixed-backbone oligonucleotides on target binding affinity and target cleaving speficity and selectivity by *Escherichia coli* RNase H. *Bioorg. Med. Chem*. 6:1695–1705.
37. Goel S, et al. (2003). A safety study of a mixed-backbone oligonucleotide (GEM231) targeting the type 1 regulatory subunit alpha of protein kinase A using a continuous schedule in patients with refractory solid tumors. *Clin. Cancer Res*. 9:4069–4076.
38. Summerton J (1999). Morpholino antisense oligomers: the case for an RNase H-dependent structure type. *Biochim. Biophys. Acta*. 1489:141–154.

39. Nielsen P E, Egholm M (1999). An introduction to peptide nucleic acid. *Curr. Issues Mol. Biol*. 1:89–104.
40. Braasch D A, Corey D R (2002). Novel antisense and peptide nucleic acid strategies for controlling gene expression. *Biochem*. 41:4503–4510.
41. Egholm M, et al. (1993). PNA hybridizes to complementary oligonucleotides obeying the Watson-Crick hydrogen-bonding rules. *Nature*. 365:566–568.
42. Kurreck J (2003). Antisense technologies. Improvement through novel chemical modifications. *Eur. J. Biochem*. 270:1628–1644.
43. Braasch D A, Corey D R (2001). Locked nucleic acid (LNA): Fine-tuning the recognition of DNA and RNA. *Chem. Biol*. 8:1–7.
44. Fluiter K, et al. (2003). *In vivo* tumor growth inhibition and biodistribution studies of locked nucleic acid (LNA) antisense oligonucleotides. *Nucleic Acids Res*. 31:953–962.
45. Zeng Y, Cullen B R (2002). RNA interference in human cells is restricted to the cytoplasm. *RNA*. 8:855–860.
46. Persengiev S P, Zhu X, Green M R (2004). Nonspecific, concentration-dependent stimulation and repression of mammalian gene expression by small interfering RNAs (siRNAs). *RNA*. 10:12–18.
47. Sledz C A, Holko M, de Veer M J, et al. (2003). Activation of the interferon system by short-interfering RNAs. *Nature Cell Biol*. 5:834–839.
48. Bridge A J, Pebernard S, Ducraux A, et al. (2003). Induction of an interferon response by RNAi vectors in mammalian cells. *Nature Genet*. 34:263–264.
49. Dorsett Y, Tuschl T (2004). siRNAs: Application in functional genomics and potential as therapeutics. *Nature Rev. Drug Disc*. 3:318–329.
50. Hornung V, et al. (2005). Sequence-specific potent induction of IFN-alpha by short interfering RNA in plasmacytoid dendritic cells through TLR7. *Nat. Med*. 11: 263–270.
51. Marques J, Williams R G (2005). Activation of the mammalian immune system by siRNAs. *Nat. Biotechnol*. 23:1399–1405.
52. Judge A D, et al. Sequence-dependent stimulation of the mammalian innate immune response by synthetic siRNA. *Nat. Biotechnol*. 23:457–462.
53. Collins F S, Guyer M S, Chakravarti A (1997). Variations on a theme: Cataloging human DNA sequence variation. *Science*. 278:1580–1581.
54. Wang D G, et al. (1998). Large-scale identification, mapping and genotyping of single-nucleotide polymorphisms in the human genome. *Science*. 280:1077–1082.
55. Collins F S, Patrinos A, Jordan E, et al. (1998). New goals for the U.S. Human Genome Project 1998–2003. *Science*. 282:682–689.
56. McCarthy J J, Hilfiker R (2000). The use of single-nucleotide polymorphism maps in pharmacogenomics. *Nat. Biotechnol*. 18:505–508.
57. Marshall E (2000). Drug firms to create public database of genetic mutations. *Genomics*. 284:406–407.
58. Kleyn P W, Vesell E S (1998). Genetic variation as a guide to drug development. *Science*. 281:1820–1821.
59. Johnson J A (2003). Pharmacogenetics: Potential for individualized drug therapy through genetics. *Trends Gene*. 19:660–666.
60. Shi M M (2001). Enabling large-scale pharmacogenic studies by high-throughput mutation detection and genotyping technologies. *Clin. Chem*. 47:164–172.
61. Shi M M, Bleavins M R, de la Iglesia F A (1999). Technologies for detecting genetic polymorphisms in pharmacogenomics. *Mol. Diagn*. 4:343–351.

62. Baron H, Fung S, Aydin A, et al. (1996). Oligonucleotide ligation assay (OLA) for diagnosis of familial hypercholesterolemia. *Nat. Biotechnol.* 14:1279–1282.
63. Tyagi S, Bratu D P, Kramer F R (1998). Multicolor molecular beacons for allele discrimination. *Nat. Biotechnol.* 16:49–53.
64. Lyamichev V, et al. (1999). Polymorphism identification and quantitative detection of genomic DNA by invasive cleavage of oligonucleotide probes. *Nat. Biotechnol.* 17:292–296.
65. Kaiser M W, et al. (1999). A comparison of eubacterial and archaeal structure-specific 5′-exonucleases. *J. Biol. Chem.* 274:21387–21394.
66. Kwiatkowski R W, Lyamichev V, de Arruda M, et al. (1999). Clinical, genetic and pharmacogenic applications of the Invader assay. *Mol. Diagn.* 4:353–364.
67. Chen X, Livak K J, Kwok P-Y (1998). A homogeneous, ligase-mediated DNA diagnostic test. *Genome Res.* 8:549–556.
68. Lizardi P, Huang X, Zhu Z, et al. (1998). Mutation detection and single-molecule counting using isothermal rolling-circle amplification. *Nat. Genet.* 19:225–232.
69. Baner J, Nilsson M, Mendel-Hartvig M, et al. (1998). Signal amplification of padlock probes by rolling circle replication. *Nucleic Acids Res.* 26:5073–5078.
70. Faruqi F A, et al. (2001). High-throughput genotyping of single nucleotide polymorphisms with rolling circle amplification. *BMC Genom.* 2:1–10.
71. Chee M, et al. (1996). Accessing genetic information with high-density DNA arrays. *Science.* 274:610–614.
72. Lipshutz R J, Fodor S P A, Gingeras T R, et al. (1999). High density synthetic oligonucleotide arrays. *Nat. Genet.* 21(suppl 1):20–24.
73. Wang D G, et al. (1998). Large-scale identification, mapping and genotyping of single nucleotide polymorphisms in the human genome. *Science.* 280:1077–1082.
74. Hacia J G, et al. (1999). Determination of ancestral alleles for human single-nucleotide polymorphisms using high-density oligonucleotide arrays. *Nat. Genet.* 22:164–167.
75. Kwok P Y (2001). Methods for genotyping single nucleotide polymorphisms. *Annu. Tev. Genomics Hum. Genet.* 2:235–258.
76. Lucito R, et al. (1998). Genetic analysis using genomic representations. *Proc. Natl. Acad. Sci. U.S.A.* 95:4487–4492.
77. Kennedy G C, et al. (2003). Large-scale genotyping of complex DNA. *Nat. Biotechnol.* 21:1233–1237.
78. Matsuzaki H, et al. (2004). Parallel genotyping over 10,000 SNPs using a one-primer assay on a high-density oligonucleotide array. *Genome Res.* 14:414–425.
79. Hardenbol P, et al. (2003). Multiplexed genotyping with sequence tagged molecular inversion probes. *Nat. Biotechnol.* 21:673–678.
80. Gunderson K L, Steemers F J, Lee G, et al. (2005). A genome-wide scalable SNP genotyping assay using microarray technology. *Nat. Genetics.* 37:549–554.

# RNA INTERFERENCE: THE NEXT GENE-TARGETED MEDICINE

Andrew V. Oleinikov[1] and Matthew D. Gray[2]

[1] *Seattle Biomedical Research Group, Seattle, Washington*
[2] *MDG Associates, Inc., Seattle, Washington*

**Chapter Contents**

1 Introduction
2 Basic Principles and Molecular Mechanisms of RNAi
3 RNAi as a Tool for Identifying and Validating Drug Targets
  3.1 Low-Throughput Screening for Drug Targets
  3.2 High-Throughput Screening for Drug Targets
4 RNAi as an Instrument for Therapeutic Use
  4.1 Cancer
  4.2 AIDS and Other Viral Infections
  4.3 Parasite Infections
  4.4 Neurological Disorders
5 Basic Theoretical and Practical Concerns for Application of RNAi in Therapy
  5.1 Variable Efficiency of Different siRNAs
  5.2 Nonspecific Systemic and Off-Target Effects, and the Potential for Pleiotropic Effects
6 Technical Challenges of siRNA Therapy: *In vivo* Delivery and Stability
  6.1 Delivery of RNAi Compounds
  6.2 Stability of RNAi Compounds
7 RNAi Drugs under Development
  7.1 Cancer
  7.2 Viral and Respiratory Diseases
  7.3 Arthritis
  7.4 Autoimmune Diseases
  7.5 Neurodegenerative Disorders
  7.6 Ocular Disorders

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

8 Comparison of siRNAs with other Pharmaceuticals (Targeted and Nontargeted Type), Including Closely Related Gene Specific Antisense-Based Drugs
9 Biotech and Drug Companies Currently Invested in RNAi Drug Discovery and Development
10 Concluding Remarks
Acknowledgment
References

## 1 INTRODUCTION

Advances in medical technology have depended on cellular and molecular research to gain an in-depth understanding of the basic mechanisms of homeostasis. Medical investigators have used their knowledge of cellular processes to search for small chemical compounds that can alter specific biochemical reactions in cells and tissues. Many of these compounds function in a general fashion through an alteration in biochemical pathways or via stimulation of the immune system. Drugs that can modulate gene expression in a targeted fashion, however, offer the potential of a much higher level of specificity and control over cellular processes. The newest laboratory method for modulating gene expression, RNA interference (RNAi), has gained rapid interest throughout the world of molecular research, in general, and is being actively pursued as the next "gene-targeted medicine."

RNAi is a process in which double-stranded RNA (dsRNA) directs the specific degradation of a corresponding target mRNA. The mediators of this process are small dsRNAs, of 21 bp in length, called short interfering RNAs (siRNAs). The phenomenon of RNAi was first reported in 1998, by Fire et al. [1], who observed that dsRNA could inhibit gene expression in *Caenorhabditis elegans*. Although the underlying mechanism of this effect was poorly understood at the time, the generality of the phenomenon and its utility in the laboratory as a method of posttranscriptional gene silencing were quickly confirmed by the work of several groups [2–8]. The process whereby dsRNA is cleaved into siRNAs was initially described in plants [9] and then in *Drosophila* [10, 11]. The cellular protein factors associated with this conversion in the cell were identified soon afterward ([12] and recently reviewed in Refs. 13 and 14).

The rapid growth and popularity of RNAi as a research topic of interest and as a useful tool in the laboratory is evidenced by the tremendous yearly increase in publications on the subject since its initial description less than a decade ago. A current search of the scientific literature, using the terms "RNAi" and "siRNA" as keywords, reveals that over 10,000 articles have been published on the subject within the last 5 years, since the beginning of this century (Figure 1). This type of response to a technical discovery is wholly unprecedented and could only perhaps be reasonably compared with the "explosion" of significant and widespread advances that the invention of the polymerase chain reaction (PCR) [15–17] contributed to biological research and current biotechnology toward the end of the last century.

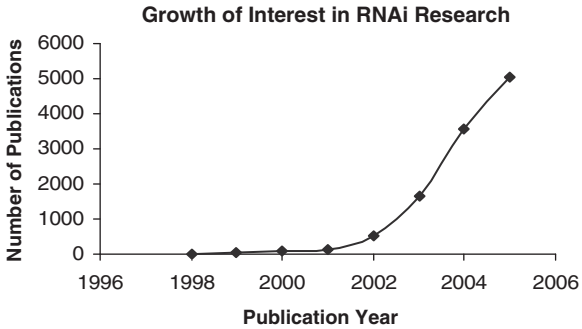


**Figure 1.** Growth of interest in RNAi research.

Our aim in this chapter is not to give an extensive review of the rapidly expanding literature in this field but, rather, to describe the main ideas, approaches, and technological directions that may guide the medical research-er interested in using RNAi methodologies to identify and/or develop new drug candidates. We have included several examples and discussions of the concepts involved to provide a clear picture of the potential advantages and limitations of specific RNAi-based gene-targeted medicine.

## 2 BASIC PRINCIPLES AND MOLECULAR MECHANISMS OF RNAi

For many years scientists involved in biomedical research have searched for simple methods for targeting the regulation of a gene's activity. These methods would be used for characterizing gene functions as well as for developing targeted medicines to treat and correct a variety of diseases and disorders. In 1990, when Napoli et al. reported on the specific repression of the chalcone synthase (CS) gene in petunia flowers by introduction of a chimeric CS gene and demonstrated that it was the result of a reduction in the level of CS mRNA [18], it would have been almost impossible to suggest that in just a few years time there would be a revolution in our ability to specifically regulate the expression of virtually any gene in any cell and in many organisms.

Work starting from that of Andrew Fire et al. [1] describing the phenomenon of RNAi in *C. elegans* in 1998, up until now, has shown that RNAi is a conserved and universal mechanism of defense [19] (against viruses and mobile genetic elements, retrotransposons) and gene regulation that is widely spread among different biological kingdoms [20]. For the efficient silencing of any target gene, a corresponding homologous dsRNA can be delivered into the cells and tissues of organisms by several different methods: direct injection, feeding, soaking, electroporation, transfection, and so on. RNAi occurs by two parallel and, in many respects, similar pathways involving many common cellular proteins: the so-called short interfering RNA (siRNA) and microRNA (miRNA) pathways (Figures 2 and 3). Both pathways lead to the specific silencing of genes, either through degradation or a translational block of the corresponding mRNA.

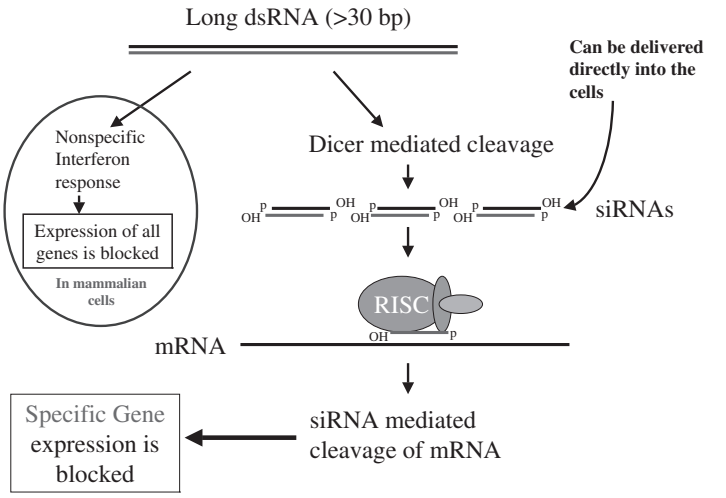


**Figure 2.** Mechanism for inhibition of gene expression via the siRNA branch of the RNAi pathway. Short interfering RNAs can be delivered directly into the target cells to avoid nonspecific interferon (IFN) response.

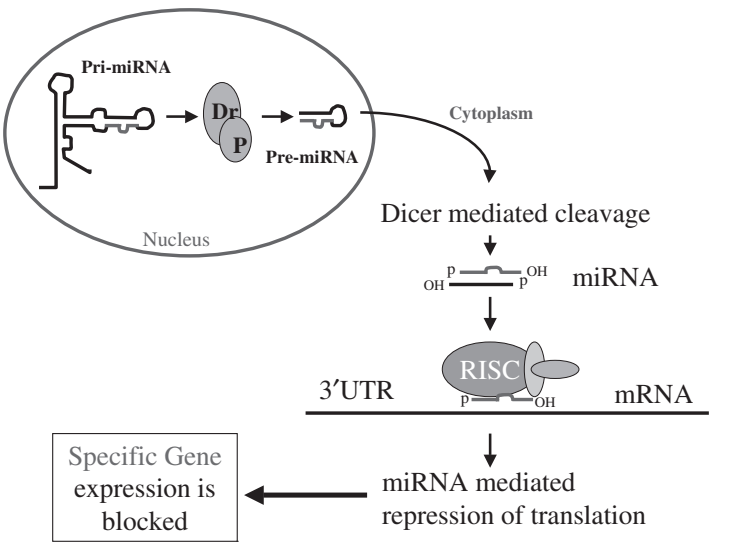


**Figure 3.** Mechanism for inhibition of gene expression via the miRNA branch of the RNAi Pathway. Proteins involved in Microprocessor: Dr—Drosha, P—Pasha, UTR—untranslated region of mRNA.

The first step in the siRNA pathway is degradation of the long dsRNA into short double-stranded fragments, later termed siRNAs (Figure 2). siRNAs, which have two overhanging 3′-nucleotides and 5′-phosphate groups [9–11, 21], are converted from longer dsRNAs by a cellular enzyme with RNase III activity called dicer ([22] and reviewed in Ref. 23). At the next step, siRNAs bind to the cellular complex called the RNA-induced silencing complex (RISC) [22, 24], of which the main

component is a protein called argonaute [19, 24]. Only one strand of the siRNA molecule specifically interacts with, and hence "enters," RISC [25, 26] and determines the specificity of target RNA cleavage.

With the success of RNAi-dependent targeted inhibition of gene expression in different lower organisms, immediate efforts were made to apply this technology in mammalian cells and organisms by the intracellular delivery of long dsRNAs cognate to the gene assigned to downregulation. However, in mammalian cells, there is a well-known nonspecific antiviral defense mechanism based on activation of the interferon (IFN) pathway in response to intracellular invasion by viral dsRNA [27]. In this pathway, several IFN-induced genes (including the RNA-dependent protein kinase (PKR), the 2′,5′-oligoadenylate synthetase (OAS) and RNase L, and the Mx protein GTPase) orchestrate a complex cascade of events leading to the degradation of the foreign dsRNA, a general arrest of the translational machinery of the host cell, and other downstream effects important for host cell defenses over viral infections. Moreover, PKR can be differentially modulated by different species of dsRNA. This effect may depend on the sequence and secondary structure of a dsRNA, and PKR can be stimulated with as few as 11 bp of dsRNA [28]. In any case, this complex mixture of different nonspecific responses can completely abolish the sequence-specific gene-targeted effects of dsRNAs.

Nonetheless, an understanding of the molecular mechanisms of RNAi proceeding through the endogenous generation of siRNAs [10, 21] allowed Elbashir et al. [29] to design a smart way to avoid most of these nonspecific IFN-related effects. They designed and delivered into the cells gene-specific dsRNA as short dsRNAs (siRNAs, 21–22 bp), as they occur in their mature form during the natural RNAi mechanism. Using this modification, most of the IFN response, which is induced by longer dsRNAs (>30 bp), is avoided. Their experiments demonstrated the use of siRNA compounds as efficient and specific downregulators of gene expression. This breakthrough began a new era for the use of RNAi technology in biomedical research for pathway profiling, drug candidate identification, and targeted therapy using mammalian cells and organisms. Nevertheless, some types of nonspecific effects still present a concern for the use of RNAi technology for therapeutic applications, as will be discussed later on in this chapter.

A detailed understanding of the molecular mechanism underlying RISC activity has led to improvements in the design of siRNA compounds. The current principles for siRNA design are (1) the strand with the more loosely base-paired 5′ end in the siRNA duplex enters RISC, (2) the 5′ end of this strand has to be phosphorylated, and (3) the siRNA duplex needs to have 3′ overhanging ends of 2-nucleotides in length. Despite this knowledge, only about 20% of siRNAs designed are active in RNAi, and hence, all require experimental validation of their activity. Furthermore, the efficiency of repression of the target gene by siRNA may not be absolute, as the degree of inhibition observed depends on several factors. These factors include the chemical stability and specific activity of the siRNA compound; the "effective" concentration delivered to the cells, which depends on the number and duration of treatments; the physiological state of the target cell; the level of target mRNA; and, especially, the level and turnover rate of the target protein in the cell or tissue. A number of these factors will be discussed in detail below.

With intense research efforts in the field of RNAi, it became clear that another cellular pathway of gene regulation, discovered earlier in *C. elegans* and based on

the transcription of small untranslated RNAs [30, 31], is closely related to siRNA-dependent gene regulation [32]. This second branch of RNAi, which has also remained extremely conserved during evolution, was termed the microRNA, or miRNA, pathway (Figure 3). MicroRNAs have been identified across many different biological kingdoms (from plants to nematodes to humans), and they function to regulate the expression of different genes. Both pathways, siRNA and miRNA, share common elements of their core machinery [33], such as dicer and RISC; however, there are some obvious differences and specialized components. The miRNA process begins in the nucleus, where long (up to several thousand bp [34]) capped pri-miRNA molecules are transcribed by RNA-polymerase II from corresponding miRNA genes. These complex RNA molecules are processed into hairpin-like pre-miRNA molecules (about 65–70 bp) by a multiprotein complex, called microprocessor, containing at least an RNaseIII enzyme Drosha and its partner Pasha [35]. Pre-miRNA molecules are then transported to the cytoplasm where they are converted to miRNAs by dicer (Figure 3) [23, 32]. The main difference between the two pathways is that miRNA molecules are not 100% homologous with the target mRNA and may contain up to several mismatches. In addition, miRNA-dependent inhibition of expression, which most likely takes place in specialized subcellular compartments called P-bodies [36], occurs through the interaction of RISC-bound miRNAs with the 3′-untranslated region of the target mRNA and results in an arrest of translation rather than degradation of the message, as in the siRNA pathway.

Currently, it is not clearly understood what targets a short dsRNA down the miRNA pathway. It is possible that the presence of some nonhomologous regions between an siRNA molecule and some unintended target RNA may direct this complex into the miRNA pathway and produce some off-target effects. In any case, a more detailed understanding of the miRNA pathway will clearly provide additional insights for the further development of powerful RNAi-based tools for gene regulation and therapy, and for the evasion of nonspecific and off-target effects.

Since its discovery less than a decade ago, the RNAi field has progressed at an incredible pace, permitting manipulations of gene expression that were previously impossible. As a result, this technology immediately acquired an important role in pharmaceutical science for the screening, identification, and validation of drug targets, a process that, before the discovery of RNAi, usually expended several years of intense effort. Moreover, the specific features of RNAi technology, such as high target specificity, efficiency, and to some extent, the ease of compound design, have implicated this technology as a potential instrument for therapeutic use. In the following sections we discuss a variety of biological and technical aspects of applying RNAi technologies to the discovery and development of customized, state-of-the-art, gene-targeted medicines of the future.

## 3 RNAi AS A TOOL FOR IDENTIFYING AND VALIDATING DRUG TARGETS

The traditional drug discovery process implemented by many pharmaceutical companies includes drug target identification, characterization of the target, high-throughput screening (HTS) of small-molecule libraries, and medicinal chemistry

optimization. The bottleneck of this scheme has traditionally been the identification of targets due to the absence of appropriate high-throughput screening approaches. The recent discovery and development of RNAi technology, however, has provided a new approach that will accelerate the rate and efficiency of the initial and overall screening process.

With the emergence of siRNA technology, its first useful application was obvious: to switch off (or downregulate) a certain gene of interest to determine what phenotypic changes might occur within the cell (or organism). These phenotypic changes can be monitored at different levels by measuring (1) the relative expression of selected genes of interest or the global changes in expression of all (or most) genes within the cell (organism), and (2) the inhibition or stimulation of other physiological parameters of the cell or organism, such as its growth rate, shape, biochemical markers, and specific functions. If downregulation/switching off a particular gene leads to the desirable phenotypic effect, this gene or gene product is then selected as a target for the development of a drug, which will have a similar effect on the inhibition of activity of the corresponding protein. By screening multiple genes (one by one, or in combinations) for downregulation of their expression and estimating the presence and magnitude of the desirable phenotypic outcome, novel genes can be identified as targets for the future development of therapeutic compounds. In this respect, systematic genome-wide screens using RNAi technology and libraries of siRNA compounds seem to offer the most powerful approach for the identification of multiple drug target genes [37].

At the same time, RNAi technology may also make it possible to identify any undesirable effects due to the downregulation of certain genes of interest, which could have detrimental effects upon the pathway, cell, or organism. Therefore, this approach may provide an effective means for the early elimination of such genes from the list of drug target candidates; which is likely to save time and money in the subsequent drug development process. Below, we will review the methods currently being employed using siRNA technology for drug target identification and validation, as well as the specific target genes and respective diseases that are currently under investigation using this approach.

The most common way of exploiting the siRNA approach for target validation is to perform experiments in cultured cells. There are several methods of introducing siRNAs into mammalian cells. Chemically synthesized siRNAs or siRNAs prepared by *in vitro* transcription can be delivered into cells through a process similar to cell transfection with DNA vectors, using several different agents to facilitate such delivery [38]. Recently, several companies have developed transfection reagents specifically designed to improve the delivery of siRNA oligonucleotides in cell culture (Ambion, Austin, TX; Invitrogen, Carlsbad, CA; QIAGEN, The Netherlands).

The second popular method for induction of RNAi is to transfect cells with vectors containing siRNA expression constructs. This is similar to the expression of exogenous genes using vector-mediated transfection except that the result of this process is not the translation of the transcribed mRNA but the assembly of a double-stranded siRNA or hairpin loop siRNA, called shRNA, within the cell. Both forms can efficiently inhibit the expression of target genes [39–43]. A major advantage of "DNA-directed" RNAi is its potential for a more sustained effect, because the siRNA-expression vector may persist in the cells for several

generations providing a constant supply of siRNA molecules. Also, the recombinant expression of the siRNA can be controlled by specifically selected regulatory elements incorporated into the vector, providing more precise control over the timing, dose, and cell specificity of target gene inhibition. It is also possible to introduce siRNA vectors into cells using DNA electroporation [44]. This method is not as convenient as transfection when working with cell cultures *in vitro*, but it may be especially well suited for the localized delivery of siRNAs when working *in vivo*.

### 3.1 Low-Throughput Screening for Drug Targets

The first direct approach to identifying plausible drug targets would be the application of siRNA technology to well-characterized biochemical pathways involved in some disease of interest. When proteins involved in a particular pathway are known as well as their functions, it is straightforward to select a few target mRNAs against which siRNAs can be prepared and optimized for their effect. Also, with fewer initial targets, the outcome of RNAi screening can be studied in greater detail. Below, we describe several examples where siRNA technology has been used to identify critical checkpoints in a known pathway and to test whether desirable phenotypic effects could be obtained.

***Cancer.*** Grillo et al. [45] used siRNA to validate the cyclin D1 and CDK4 protein complex as a drug target for cancer drug discovery. The inhibition of endogenous cyclin D1 or CDK4 expression by RNAi resulted in hypophosphorylation of the retinoblastoma gene product and accumulation of MCF-7 breast cancer cells in G1. These results support the prevailing view that pharmacological inhibition of cyclin D1/CDK4 complexes is an effective strategy for inhibiting the growth of tumors.

Another recent example demonstrates that inhibition of an enzyme in the same pathway in cells of different species may have different, and even opposite, effects on phenotypic outcome. This emphasizes that careful selection of model cells and characterization of molecular detail of the biochemical pathway, including tissue- or cell-specific expression of different proteins, is critical to selecting a suitable drug target. Hill et al. [46] demonstrated that rat (NMU and 13762), but not human (MCF-7 and T47D), breast cancer cell lines express Heme oxygenase-1 (HO-1), whereas indoleamine 2,3 dioxygenase (IDO) is expressed by both. IDO inhibition with siRNA resulted in diminished proliferation of rat cells, whereas HO-1-negative human cell lines increased proliferation upon IDO inhibition. As IDO inhibits the antiproliferative effects of HO-1, these authors concluded that IDO has the opposite effect on proliferation, dependent upon the co-expression of HO-1. This example clearly shows that results obtained on cells from other organisms are not automatically transferable to human cells, even if they seem to have highly homologous proteins and very conserved pathways. It also shows that a drug specifically targeting the expression or activity of a particular gene might have the opposite effect in cell types that differ in their expression profiles. These potential problems need to be considered when screening for therapies directed at a particular target protein.

***Respiratory Diseases.*** An interesting example of hypothesis-driven identification of a potential drug target *in vivo* using siRNA is described in the work of

Lomas-Neira et al. [47]. In their study of acute lung injury (ALI), they hypothesized that local silencing of keratinocyte-derived chemokine (KC) or macrophage-inflammatory protein-2 (MIP-2), via the local administration of siRNA against KC or MIP-2, after traumatic shock/hemorrhage would suppress signaling for the influx of polymorphonuclear neutrophils (PMNs) to the lung, thereby reducing ALI associated with a secondary septic challenge (cecal ligation and puncture). This hypothesis was tested using intratracheal instillation in an *in vivo* model (mice). They found that 24 h after this treatment intratracheal MIP-2 siRNA significantly reduced tissue and plasma interleukin-6, tissue MIP-2, as well as neutrophil influx. In contrast, KC siRNA treatment reduced plasma KC, tissue KC, and IL-6 but produced no significant reduction in plasma IL-6 or neutrophil influx, thus identifying MIP-2, but not KC, as a potent drug target. In addition, this study demonstrates the potential use of siRNA as a therapeutic.

***Inflammation.*** The exact roles of many cytokine genes that influence rheumatoid arthritis are still unknown. Inhibition of some of these genes can strongly reduce joint inflammation [48]. In an *in vivo* arthritic mouse model, Schiffelers et al. [49] used electroporation [44] to deliver a tumor-necrosis factor alpha (TNFα) siRNA-coding vector into joint tissue. They demonstrated not only that TNFα is an effective target for reducing joint inflammation, but also a new strategy for therapeutic intervention for rheumatoid arthritis. This approach, hence, may clearly serve as a model for the study of arthritis disease pathways through gene-specific loss-of-function phenotypes.

***Parkinson's Disease.*** Animal disease models play an extremely important role in identification of drug lead candidates. Some animal models can be created simply by making gene knockouts. However, if the targeted gene is an essential gene for survival of the organism, these animals cannot be easily created. For example, mice in which the gene encoding tyrosine hydroxylase (TH) has been knocked out die shortly after birth, making it impossible to study the neurological effects of this alteration in the fully developed animal. RNAi technology can help to overcome this obstacle by creating an animal disease model through the local and/or temporal knocking out of essential genes. Hommel et al. [50] demonstrated that injection of an adeno-associated virus encoding a TH-specific shRNA into the midbrain of 9-week-old mice generated an attenuated locomotor response to amphetamine, an effect that is also characteristic for toxin-induced rat models of Parkinson's disease. These authors, hence, suggest that similar knockout strategies may be effectively used in other disease model systems. In addition, as we mentioned above, because the RNAi effect obtained by treatment with siRNA compounds may not be absolute, there may also be a certain advantage in a knock-down vs. knockout phenotype; in that, a knock down might be more appropriate for modeling some complex human diseases [51].

### 3.2 High-Throughput Screening for Drug Targets

The second valuable approach to using RNAi for drug target identification is a high-throughput strategy (HTS), where a library of siRNAs directed against many genes is screened simultaneously or sequentially to identify target genes and their products for further evaluation as drug leads. As described above, this approach might

be used against a restricted set of genes that are members of a certain pathway, for example, a "pathway approach," or against most of the genes in the genome, a "global approach." For these approaches, traditional HTS equipment, such as robotic systems, can be implemented, facilitating the processing of thousands of cell-based screening protocols a day [52]. Both chemically synthesized siRNAs as well as vectors coding for functional si/shRNAs can be delivered into the cells using this high-throughput format. Retroviral [53] and lentiviral [54] vectors have been shown to be particularly useful vehicles for high-throughput screening of siRNA and shRNA libraries due to their robust infection and expression in a wide range of mammalian cell types. To simplify the transfection procedure on multiwell plates or slides, Chang et al. [55] suggested the pretreatment of surfaces with cationic polymers. This provides for the simple addition of naked DNA vectors or siRNAs directly to the cells in these pretreated wells in order to achieve efficient transfection. Another method of high-throughput screening using vector-encoded or chemically synthesized siRNAs is the so-called "reverse-transfection," or solid phase, method [56–58], originally designed for the expression of cDNA libraries [59]. In this method, a number (up to several thousands) of individual siRNAs or siRNA-encoding vectors are spotted in a predetermined array onto a microscope glass slide, which is then treated with transfection reagent and covered with recipient cells. Individual cells seeded directly over each spot of material take up only the siRNAs or vectors localized at that spot. That, hence, results in a virtual array of siRNA-treated cell colonies. After 24–48 hours, the array of treated cells is evaluated for the desired phenotype to identify effective siRNAs and, more importantly, the corresponding target gene. It has been estimated that, by this method, between 100 and 500 reverse-transfections could be performed using the same amount of reagents required for a single transfection in a well of a 96-well plate [58].

***Cancer.*** Using a human cell line and an shRNA library (containing 23,742 distinct shRNAs, 3 shRNAs per gene) cloned into a retroviral vector and directed against of 7914 genes, Berns et al. [60] identified genes involved in the p53 pathway. This protein is a known tumor-supressor and induces cell cycle arrest in cells that have undergone DNA damage. By inhibiting an array of genes using their shRNA library, the authors identified five new genes, and p53 itself, whose expression are required for proliferation arrest of cells. After infection of cells with the pooled shRNA library, the relevant genes associated with this specifically desired phenotype were rapidly identified by unique "molecular bar-codes," which were introduced into each siRNA construct.

The HTS–RNAi approach is extremely promising in providing rapid identification of proteins, which control or are associated with disease-relevant phenotypes. A lentiviral library of shRNA targeting 9610 human and 5563 mouse genes (human homologues) was efficiently used to identify proteins that compromise the 26S proteasome function [61]. The 26S proteasome, a well-studied biochemical machinery involved in the critical degradation of many unwanted proteins including some oncoproteins (e.g., c-Myc), has been implicated in certain diseases. About 30 proteins, mostly proteasome subunits, including five out of five proteasome ATPases and the two largest non-ATPases, were identified in their screen. This work also exploited a "mole-cular bar-coding" system, which facilitated the identification of effective shRNAs.

A single, well-constructed library can be used to screen for gene products involved in completely different disease-relevant pathways. The most important step in applying this approach is the design of a functional assay that will clearly discriminate between the genes whose inhibition results in the desired phenotypes to be obtained by screen. In each case, the specialized development of an appropriate cell-based assay is required. Principles and strategies for developing such assays for the identification of drug-targets for different diseases, including rheumatoid arthritis, osteoarthritis, asthma, osteoporosis, Alzheimer's disease, and cancer, are discussed in the recent review by van Es and Arts [62].

The combination of a high-throughput microarray analysis of differential gene expression followed by RNAi inhibition of selected genes also represents a powerful methodology for the identification of new drug targets [63]. Using this approach, Williams et al. [64] analyzed the differential expression of genes in normal tissue versus colon tumors and polyps isolated from 20 patients. In this study, they identified 574 genes that were upregulated in tumor patients. Out of this large set, they selected a single gene, survivin, a potent inhibitor of apoptosis, and downregulated its expression by RNAi. Inhibition of survivin expression via RNAi resulted in severely reduced tumor growth *in vitro* as well as in an *in vivo* mouse xenograft model. Future analyses using RNAi inhibition of other upregulated genes will undoubtedly make it possible to identify additional precise targets for drug development.

Using this combined approach plus the construction of gene knockout mice, Gunton et al. [65] identified the transcription factor ARNT as an important factor involved in impaired glucose-stimulated insulin release and in causing changes in gene expression similar to those in human type 2 islets. Hence, RNAi was used here to identify and validate a new mouse model for type 2 diabetes.

The reverse application of these two technologies, where cells are first subjected to RNAi treatment and the resulting changes in global gene expression are then assessed by microarray analysis, allows for the dissection of cellular pathways of interest [66]. Such analyses are extremely important: (1) for understanding the global changes that occur in the cellular response to the downregulation of the selected target, (2) for elucidating the role of the target gene in intracellular pathways, (3) for identification of unwanted or detrimental effects, and (4) for selecting the best checkpoints for therapeutic intervention.

These examples demonstrate that RNAi technology offers scientific promise for the identification of effectual drug targets and an excellent perspective for pursuing the development of corresponding small molecules to modulate these targets. Nevertheless, some caution should be reserved in its application because, although RNAi technology may provide a clear understanding of a particular protein's role in a pathway and the phenotypic effects of its downregulation, a small drug directed at the same cellular checkpoint or mechanism might produce different effects [67].

## 4 RNAi AS AN INSTRUMENT FOR THERAPEUTIC USE

With the apparent great success of using RNAi technology for the inhibition of specific genes in cultured cells, and in some animal models, there has been a

corresponding great expectation of efforts to directly transfer this technology for therapeutic use in humans. Several companies are currently developing RNAi-based drugs to target different diseases (see Table 1, and discussion below). A variety of serious challenges for the development of such drugs will be discussed in the next two sections. Herein we will review the basic principles and methods for therapeutic treatment, as well as a few prospective directions for the application of RNAi in medicinal therapies.

The most important aspect of using siRNAs as therapeutic agents is their specificity to the target RNA. An optimally designed siRNA for therapeutic use should cause the destruction or repression of only its specific target RNA. This principle puts particular restrictions upon the choice of potential target sequences. For example, if several variants of a protein exist, among which only some forms are disease-relevant and, hence, represent a subset of targets for repression, the drug must be designed based on sequential differences between the intended target and the nontarget forms of the protein. Such variants can be the products of different alleles of the same gene, the results of alternative splicing of a single gene, or the products of a family of related genes. In this respect, dominant mutations that define specific disorders are particularly good candidates for RNAi-based treatment due to the inherent potential ability of an siRNA to discriminate between the normal gene product and the "mutant target" RNAs containing just a few, or even a single, nucleotide mutation(s) [68–70].

As RNAi acts at the RNA level, either through the degradation of the target mRNA by an siRNA-directed mechanism or through the inhibition of the target mRNA's translation by the miRNA pathway, its final phenotypic effect will depend significantly on the amounts, stabilities, and turnover rates of the target RNA and its corresponding protein. In the general case, when the ultimate goal of treatment

**TABLE 1. Companies Developing RNAi Drugs**

| Company | Targeted Disease or Agent and Stage |
|---|---|
| Sirna | AMD—phase I |
| | HC, Asthma, DB, Cancer—preclinical |
| | HD—proof of principle |
| Alnylam Pharmaceuticals | RSV—phase I |
| | AMD, DR, DME, SCI—preclinical |
| | INFL, CF, PD—research |
| Acuity Pharmaceuticals | AMD, DR—phase I |
| Benitech | HC, HIV—preclinical (scheduled for phase I in 2006) |
| Genesis R&D Corp. Ltd. | Allergy—preclinical |
| ToleroTech Inc. | Transplant rejection—preclinical |
| Intradigm Corp. | Cancer (anti-angiogenesis)—preclinical, IND in 12 months |
| CytRx | Obesity, Type 2 DB, ALS, CMV—research |

Data obtained from company websites.

*Abbreviations:* AMD, age-related macular degeneration; HC, Hepatitis C; DB, Diabetes; HD, Huntington's Disease; RSV, Respiratory Syncytial Virus; DR, diabetic retinopathy; DME, diabetic macular edema; SCI, Spinal Cord Injury, INFL, Influenza; CF, Cystic Fibrosis; PD, Parkinson's Disease; HIV, Human Immunodeficiency Virus; ALS, amyotrophic lateral sclerosis; CMV, Cytomegalovirus.

is to completely abolish a target protein, a successful siRNA compound should be targeted to a gene encoding a relatively short-lived protein with a high turnover rate, preferably one with a low total amount in the cell or organism. Following these criteria, the inhibition of the production of the targeted gene product will most effectively result in a significant decrease in the concentration of this protein in the cell and, hence, will optimally affect the downstream biochemical events in order to treat the disease condition.

A promising class of gene target candidates for siRNA intervention is genes with temporal profiles of expression. If an siRNA drug is delivered to cells before or at the beginning of expression of these genes, there are better chances that the production of their products will be inhibited sufficiently to produce the desirable therapeutic effect; even better if these proteins are normally produced in small quantities. A good example of such genes is viral genes that start to express early after viral infection. Certain specific viral transcription factors, generally expressed in small quantities, should be particularly well suited as targets for siRNA treatment. Several examples have demonstrated that siRNA directed against viral transcription factors can efficiently inhibit viral replication. Lee et al. successfully inhibited the replication of human immunodeficiency virus type 1 (HIV-1) by Tat-specific siRNA [71]. Simultaneous inhibition of viral transcription factor and surface antigen by a vector expressing two siRNAs specific for HBx and HBs significantly reduced hepatitis B virus replication in cell culture [72].

Another suitable group of targets for siRNA treatment in the case of viral infection is host proteins involved in infection and propagation of the virus. The key consideration here is that downregulation of these molecules must not be detrimental to the host. In the study mentioned above, by Lee et al. [71], inhibition of HIV-1 was achieved not only by Tat-specific siRNA but also by siRNAs specific for the CCR-5 cellular coreceptor. Using RNAi, Chiu et al. [73] inhibited expression of human positive transcription elongation factor P-TEFb. P-TEFb is involved in transcriptional regulation of cellular genes as well as HIV-1 genes. Targeting of endogenous P-TEFb resulted in the inhibition of Tat transactivation and HIV-1 replication in host cells. Interestingly, P-TEFb knockdown was not lethal to the host cells, which showed normal P-TEFb kinase activity. These authors suggested that there may be a different critical threshold of P-TEFb activity required for cell viability than for HIV replication. This example shows that RNAi-mediated gene inhibition may provide an opportunity to specifically inhibit a disease-causing agent while minimizing the effect on normal cellular function. Nevertheless, the inhibition of any cellular gene for the purpose of disease treatment or prevention should be considered very carefully, not only at cellular level but at the level of the whole organism as well.

As mentioned in Section 2, using siRNA it is rare to achieve an effect of 100%, where the target RNA is completely abolished or "knocked out." More often the result is a "knock-down" effect. In some cases, such a partial reduction of gene expression is sufficient for obtaining the desired phenotypic or therapeutic effect. It is also prudent to be able to switch "on" and "off" the expression of siRNA at certain times or in specific tissues. Fritsch et al. designed a CRE-lox-based strategy that allows one to repress gene activity through shRNA expression in a time-dependent manner in cells, and in a time- or tissue-dependent manner in animals [74, 75]. Other vectors with chemically inducible promoters have been designed for the

controlled expression of shRNAs using tetracycline [76, 77], ecdisone [78], macrolide [77], and butyrolactone [79]. Still, in some cases, an incomplete shut-off of the target gene may not produce the desired therapeutic effect. In these cases, RNAi therapy could be used effectively in combination with other drugs directed to the same, or a different, checkpoint in order to achieve a more satisfactory therapeutic effect. Such parallel treatments, with drugs having different modes of action, also make it possible to overcome potential antagonistic factors due to mutations or alternative pathways leading to a disease.

A prolonged treatment with any drug may encounter the emergence of mutations causing resistance to the treatment. RNAi technology is inherently susceptible to this problem. However, expression vectors able to express two shRNAs simultaneously have been designed and used to help overcome potential loss of RNAi-drug efficacy [80]. Using two siRNAs directed against the 3D-RNA-dependent RNA polymerase of coxsackievirus B3, Shubert et al. obtained efficient inhibition of virus propagation in HeLa cells and reduced virus titers by up to 90%. Introducing point mutations into either target site can abrogate the inhibitory effect of that siRNA, when used separately. These authors constructed an siRNA double-expression vector (SiDEx) to achieve simultaneous expression of both siRNAs from a single plasmid. Their "double-drug" construct successfully downregulated the point-mutated target gene. In addition, these vectors could be used for the simultaneous knockdown of two different targeted genes for functional studies or therapeutic use.

In Section 3, we described the application of RNAi for drug target identification and validation in different diseases. In examples where inhibition of expression of specific genes led to the desirable phenotypic effects, those targets may consequently be considered for RNAi therapeutic use. Below we provide additional examples of how the principles and methods described in this section can be applied to the development of therapies for different diseases.

### 4.1 Cancer

A significant proportion of all cancers result from mutations in cellular proto-oncogenes. Nearly half of these cancers contain point mutations in the tumor-suppressor gene, p53. The ability of an siRNA compound to discriminate between wild-type (WT) and mutated forms of p53 mRNA allowed for the specific knockdown of a dominant-negative mutation, resulting in an inhibition of tumor growth and restoration of WT protein function [68].

Fusion proteins, which result from chromosomal translocations, are commonly associated with the growth transformation that occurs in some cancers. Leukemic fusion gene AML1/MTG8 is associated with up to 15% of all *de novo* cases of acute myeloid leukemia. Heidenreich et al. [81] used a human leukemic cell line, Kasumi-1, to demonstrate that efficient suppression of the disease-associated fusion protein AML1/MTG8, but not WT AML1, can be achieved by siRNAs directed against the fusion site of the AML1/MTG8 mRNA. This treatment, in combination with TGF beta(1)/vitamin D(3), significantly reduced the clonogenicity of Kasumi-1 cells.

One subtype of acute lymphoblastoid leukemia is characterized by a t(1;19) chromosomal translocation resulting in the expression of a novel fusion protein

E2A-Pbx1. This mutant protein activates the expression of a secreted Wnt16 glycoprotein [82], which is widely involved in cell proliferation, differentiation, and oncogenesis, and promotes transformation and leukemogenesis. Mazieres et al. [83] demonstrated that only one of two Wnt isoforms (Wnt16b) is overexpressed in t(1;19)-containing cell lines. Using isoform-specific siRNAs and an anti-Wnt16 antibody, they showed that targeted inhibition of the overexpressed Wnt16b leads to apoptotic cell death.

A common problem in modern cancer therapy is the development of cellular resistance to conventional chemotherapeutic compounds. In many cases, this resistance is due to the overexpression of the multidrug resistance gene, MDR1. Targeting of this gene using RNAi technology may significantly decrease or even completely reverse the multidrug-resistance phenotype [84–86]. Thus, again, suggesting the potential utility of using RNAi in combination with traditional strategies.

Inhibition of oncogene expression, or of other components of molecular pathways involved in cell proliferation, may slow the growth of cancer cells in a tumor without leading to apoptotic cell death; and hence, it may not always eliminate these cells from the organism. In addition, because RNAi rarely leads to the complete knockout of target gene expression, it might be advantageous to use a combined regimen of RNAi and a small drug and/or radiation treatments for more efficient combat against cancer. RNAi compounds can be directed against genes responsible for anti-apoptotic effects [87–89], against tumor-suppressor genes [90], against dsDNA-brake repair proteins [91, 92], or against other genes in cellular survival pathways [93]. In this scheme, RNAi-induced suppression of normally protective, repair-activated, or damage-induced genes in tumor cells may result in even greater increases in proliferation of tumor cells and, hence, an increased sensitivity to the combined chemotherapeutics or radiation.

### 4.2 AIDS and Other Viral Infections

Interestingly, to date, almost all HIV-1 genes have been targeted by RNAi, at least in cell culture experiments. It has proven possible to block the replication of HIV-1 in cultured cells by siRNAs targeted to viral sequences [94, 95]. The main obstacle to therapy, though, is the specific delivery of siRNAs to cells infected or susceptible to HIV-1 infection *in vivo*. We shall discuss this problem of cell-specific targeting in detail later. However, it has been shown recently that HIV-1 can suppress the function of the host cellular RNAi machinery [96]. This suppression of RNA silencing may potentially affect the efficiency of RNAi drugs directed against HIV-1 gene targets, and against other target viruses that may have evolved similar anti-host mechanisms. In addition, HIV infections are particularly difficult to treat, in part, due to the viruses' extreme ability to transform as a result of high mutation rates, especially of surface receptors, which gives rise to the high sequence diversity of HIV genomes among infected subjects. One measure against this tendency for viral sequence diversity might be the simultaneous use of several siRNAs directed against several viral genes. On the other hand, nonvariable human cell receptors for HIV infection represent a great target for siRNA treatment. Martinez et al. [97] demonstrated that RNAi directed against CD4 coreceptors CXCR4 and CCR5 efficiently blocked acute infection by X4 (NL4-3) and R5 (BaL) HIV-1 strains but

had no effect on CD4 cells or control green fluorescence protein expression. In another example, Nair et al. [98] presented evidence that siRNA directed against the CD4 independent attachment receptor (DC–SIGN) significantly inhibits HIV infection of dendritic cells. Nevertheless, as mentioned above, any targeting of host genes should be evaluated for unwanted or detrimental effects, especially at the organism level. Some negative effects could significantly restrict the usage of endogenous gene targets.

Many other viral infections, including the ones most threatening to humans, like hepatitis C, B, and A viruses (HCV, HBV, HAV), human papillomavirus viruses (HPVs), West Nile virus, and human herpes viruses (i.e., HSV, CMV, EBV, and HHV-8) are also the targets of intense research for the development of effective RNAi treatments. It has been demonstrated that RNAi can disrupt the replication of many viruses, including HCV, HBV, and HAV in cell culture [99, 100] and, as later demonstrated in mouse models, *in vivo* [101, 102]. The major limitation, again, is the delivery and stability of RNAi compounds for application as human therapies in clinically relevant systems, which will be discussed further below. These examples demonstrate, however, that RNAi technology has great potential for the treatment of viral disease, yet clearly requires further development for clinical applications.

### 4.3 Parasite Infections

Despite the biotechnological and drug development booms of the last several decades, parasite infections (malaria, African sleeping sickness, leishmaniasis, Chagas disease, etc.) remain the foremost health burden and source of mortality and morbidity on our planet, especially in the tropics. Malaria, alone, contributes to about 500 million cases of disease per year that result in more than 2–2.5 million deaths, mostly in sub-Saharan Africa [103]. Not only are newer drugs still expensive and unavailable to most of the population in these regions, but parasites also develop drug resistance rapidly, especially to the drugs extensively used, like chloroquine [104]. Due to many reasons, vaccine development has not yet been successful against these diseases. Therefore, the appearance of RNAi technology was perceived with great interest and was immediately explored in studies aimed at defining treatments for parasite infections.

It was quickly discovered that RNAi works in *Trypanosome brucei* as well as it does in *C. elegans* [105]. Since that time, RNAi mechanisms have been extensively studied in *T. brucei* [106]. The technology is now being used extensively to dissect different biochemical pathways in *T. brucei* [107] as well as to study potential therapeutic applications [109–110].

Unfortunately, it is still not clear whether RNAi functions in malaria parasites. Very few publications have described RNAi in malaria [111–114]. Genomic data mining does not clearly identify any members of the RNAi machinery in these parasites by homology searches. Given the enormous increase in the number of publications on RNAi in other systems, and the likely numerous efforts made in malaria research to use RNAi, there seems to be a surprising lack of published evidence demonstrating RNAi in malaria. It is possible that in this parasite the RNAi pathway was lost or has diverged during evolution to perform other functions in the organism, which is incompatible with the classic utilization of the RNAi mechanisms being explored in other organisms [106]. It is also possible that the

effects described in the papers published on malaria actually reflect other processes and pathways, such as the abundant presence of antisense RNAs [106] as recently described in malaria [115, 116].

### 4.4 Neurological Disorders

Many neurological disorders, including Alzheimer's disease, Parkinson's disease, fragile-X syndrome, and amyotrophic lateral sclerosis (ALS), result from dominant mutations in a single allele. Therefore, as discussed above, this class of "disease genes" is perfectly suited as candidates for treatment by sequence-specific RNAi therapy. In the example of ALS, given below, stepwise improvements led to progress in the development and use of RNAi technology for therapeutic treatment of the disease in an animal model.

ALS is a familial neurodegenerative disease characterized by motor neuron degeneration, paralysis, and death. One cause of this disease (2–3% of cases) is mutations in the Cu,Zn superoxide dismutase (SOD1) gene. Ding et al. [117] identified a specific shRNA that could specifically degrade mutant SOD1 mRNA. Later, Maxwell et al. [118] investigated the functional effects of RNAi-mediated silencing of mutant SOD1 in cultured murine neuroblastoma cells and found that silencing of mutant SOD1 protects these cells against cyclosporin A-induced cell death. Recently, Raoul et al. [119] demonstrated that in SOD1 (G93A) transgenic mice, a model for familial ALS, intraspinal injection of a lentiviral vector that produces RNAi-mediated silencing of SOD1 substantially retards both the onset and the progression rate of the disease. Another group [120] injected lentiviral vector expressing SOD1 RNAi compound into various muscle groups of SOD1 (G93A) mice. This treatment also resulted in an efficient and specific reduction of SOD1 expression and improved survival of vulnerable motor neurons in the brain stem and spinal cord. They observed a considerable delay in the onset of ALS symptoms by more than 100% and an extension in survival by nearly 80% of their normal life span. Despite this success in mice, there is an additional complexity with this disease in humans. There are more than 100 different known mutations in the SOD1 allele. Whereas it is difficult even to propose to design efficient RNAi compounds to address each of these mutations in individual patients, Xia et al. [121] suggested another elegant approach based on RNAi technology to overcome this problem. They designed a replacement RNAi strategy, where all mutant and wild-type forms of the gene are inhibited by RNAi. The wild-type SOD1 function is then replaced by a custom designed wild-type SOD1 gene that is resistant to the RNAi. These ALS studies clearly demonstrate how therapeutic strategies based on RNAi technology can be improved over a short period of time with the accumulated incorporation of different molecular aspects of a certain disease into improvements in the application of the technology itself.

## 5 BASIC THEORETICAL AND PRACTICAL CONCERNS FOR APPLICATION OF RNAi IN THERAPY

Despite the great potential of RNAi technology for therapeutic applications, there are a number of intrinsic features of this process, which present a considerable challenge to its practical use.

### 5.1 Variable Efficiency of Different siRNAs

As discussed in the second section, different dsRNA fragments cognate to the target RNA have different efficiencies in inducing its degradation. Even more uncertainties exist about the miRNA-dependent mechanism of inhibiting mRNA translation. Because the relative efficiencies of different siRNAs in RNAi may differ significantly, currently there is no way to predict their individual potentials. Screening for the most efficient siRNAs using cell cultures can be a laborious and expensive procedure. In addition, natural nucleotide polymorphism and mutations in a gene target site may affect the efficacy of RNAi [122–124]. Therefore, efforts have been made to overcome this problem of variable efficiency of different siRNAs.

Several methods based on the use of siRNA mixtures, which may contain a particular efficient siRNA (or several), have been developed. These include the preparation of siRNA mixtures using RNaseIII [125] or dicer [126] enzymes to digest longer dsRNAs. The short RNAs produced as a result of these digestions have been found to be efficient in RNAi. The disadvantage of these digestive approaches, however, is in their potential for nonspecific effects. If a randomly chopped long dsRNA contains areas of homology to other genes, or splice variants of the same gene, the expression of these genes might also be inhibited by cognate siRNAs (see below). In addition, obtaining long dsRNAs also adds to the potential total cost of, and time required for, these approaches.

It would be more advantageous to use defined mixtures of siRNAs, complex enough to inhibit the desired gene with high probability, but devoid of any siRNAs with unwanted homologies. Besides creating higher target efficiency, the use of such mixtures would also alleviate the need to screen for individually active compounds. We developed a simple, fast, and inexpensive method for the preparation of defined-siRNA mixtures using custom DNA oligonucleotide microarrays as a starting material [127]. In this study, we demonstrated that siRNA pools, prepared using array-derived DNA oligonucleotide mixtures, are an efficient tool for the inhibition of exogenous as well as endogenous genes in cell cultures [128]. In this approach, many (up to 30) different siRNAs are designed against a single gene. As the sequence of the human genome is known, any unwanted homologous siRNA sequences could be excluded from the designed pool. Many different siRNA pools can be synthesized using a single DNA oligonucleotide microarray, which makes this approach extremely affordable relative to the construction of defined complex mixtures via the synthesis and/or cloning of many different individual siRNA species.

Although the approaches using siRNA mixtures prepared as described above are efficient for use in cultured cells for pathway validation and drug target identification experiments, their use as therapeutic agents might, nevertheless, be problematic due to manufacturing issues as well as issues related to the specific delivery of siRNAs to the target cell or tissue (see below). In any case, the identification of an efficient single or a mixture of RNAi compounds requires a careful selection process and experimental testing to maximize gene silencing and minimize potential off-target and nonspecific effects.

### 5.2 Nonspecific Systemic and Off-Target Effects, and the Potential for Pleiotropic Effects

We described, in the second section, how long dsRNAs may induce nonspecific effects in mammalian cells due to the activation of interferon and the induction of

the cellular dsRNA-dependent protein kinase response. Therefore, to avoid these nonspecific effects, the use of siRNA (<~30 bp) was proposed [29] and proved to be an efficient technology for posttranscriptional gene silencing. Nevertheless, it was later noted that even the short dsRNAs used to generate RNAi effects, and designed in the form of either siRNA or shRNA, can still induce some mammalian genes involved in the interferon response [129], although results from different laboratories were conflicting [130]. To circumvent these effects, different measures can be implemented, including different administration routes, dose of vectors, selection of promoters, and use of additional immunosuppression [131].

Interferon-independent nonspecific effects also exist. Scacheri et al. found that siRNAs could induce dramatic and significant changes in the levels of p53 and p21 proteins that were unrelated to silencing of the target gene [132]. These authors speculated that partial complementary sequence matches to off-target genes may have resulted in a microRNA-like inhibition of translation. In this respect, it is, actually, difficult to distinguish between nonspecific and sequence-specific off-target effects (see below). It has also been reported that si/shRNAs induce immune activation in macrophages and dendritic cells through toll-like receptor 3 [133, 134]. Moreover, Persengiev et al. [135] found that 21-bp siRNAs nonspecifically stimulated or repressed more than 1000 genes involved in diverse cellular functions. The effects on gene expression were dependent on siRNA concentration and were stable throughout the course of siRNA treatment. Again, here too, it is difficult to distinguish the nonspecific systemic effects from off-target effects. In addition, there is a potential for pleiotropic effects due to a genuine physiological response of downstream genes due to the downregulation of the target genes. If off-target genes become downregulated, these may also produce a cascade of unintended physiological responses. Such a cascade of multiple events could result in the global dysregulation of many unrelated genes. Therefore, potential widespread, nonspecific effects on mammalian gene expression must be carefully considered in the design of therapeutic RNAi applications.

Besides nonspecific cellular effects due to the RNAi compounds themselves, there are additional sources for potential nonspecific effects; these include the vehicles for, and methods of, delivery of RNAi compounds to cells and tissues. For example, viral vectors, as well as other components such as transfection reagents [136], chemical modifications to RNAi compounds, or chemical substances used to facilitate the delivery of RNAi compounds, might produce unwanted immune and other cellular responses [137]. These responses should be considered as additional possible sources for nonspecific effects in therapeutic applications of RNAi.

In addition to the aforementioned nonspecific effects, one additional problem has recently become evident: that the cellular RNAi machinery, normally involved in the regulation of some endogenous genes' expression and in maintaining a stable chromosome structure, might be indirectly affected by large amounts of exogenous siRNAs introduced by transfection or vector-directed transcription, for example. This may result in an imbalance or saturation of the cellular resources of endogenous molecules in the pathway.

The possibility of saturating the RNAi machinery was raised in early studies [122], and it has recently been supported by several observations [138–140]. Moreover, it has been suggested that perturbation of the cellular RNAi machinery might

be involved in different disorders and diseases [141]. Therefore, exploiting the RNAi machinery for therapeutic interventions could present additional nonspecific complications for cellular and organismal homeostases that should be carefully evaluated in the design and use of RNAi compounds, including the consequences of the dosage and timing of treatment.

Off-target effects are another serious concern in the application of RNAi for therapy. Such effects might be due to the partial homology of an RNAi compound to an irrelevant, or related but different, gene, which is not a target for downregulation. These effects will clearly be different for different sequences, and the parameters and degree of homology, which confer the extent of off-target effects caused by a particular RNAi compound, are complex and not fully understood. It has been shown that, depending on the sequence, most RNAi compounds can tolerate one [142], or even several [143], mismatches, though, typically with reduced efficiency. Moreover, it was shown recently that in some cases mismatches can even enhance the activity of RNAi compounds, possibly through the miRNA-regulatory pathway [144]. In fact, as few as 11 base pairs of homology with single-stranded RNA can be sufficient to produce an inhibitory effect [145]. Therefore, careful elimination of possible antagonistic homologies is an important consideration in the design of RNAi compounds. This is feasible with the great wealth of sequence data and bioinformatics currently available from human and pathogen genomes. In this respect, treatment with complex mixtures of defined siRNA molecules [128] may also offer a unique advantage over current methods. Because each efficient siRNA has an optimum concentration when its effect is maximal (the practical effective concentration range in cell culture is 10 to 100 nM), if such an siRNA molecule, homologous to an off-target gene, were present in a defined mixture, it is likely that its individual concentration would be well below the effective range. Thus, because a complex mixture targeted at a single gene may include 30–50 different siRNAs at an active total concentration of 10–100 nM, this may, theoretically, lessen the off-target effects of each siRNA molecule contained within the mixture. Nevertheless, despite the incorporation of even the most conscientious theoretical design elements to minimize nonspecific and off-target effects, the empirical testing of each compound is requisite for its evaluation and development for therapeutic application.

## 6 TECHNICAL CHALLENGES OF SIRNA THERAPY: *IN VIVO* DELIVERY AND STABILITY

The use of any drug for clinical application involves two important issues: the delivery of the drug to the affected cells or tissues, and the stability of the drug while in the organism. These factors directly influence the determination and choice of effective dosage, delivery time, and routes. RNAi compounds are not unique in this respect, and their physical–chemical and biological characteristics determine the limits and specifics of RNAi therapy. The physical stability of any compound might depend significantly on its particular route of delivery, and therefore, these parameters are potentially restrictive to each other and may, hence, limit different options for use.

### 6.1 Delivery of RNAi Compounds

RNAi compounds for therapeutic treatment can be generated in two ways: chemically synthesized si/shRNA molecules and vector-encoded molecules, which are synthesized by the cells after intracellular delivery of the corresponding expression vectors. The choice between these two approaches will determine the specifics of treatment: different targets and routes of delivery, different stability of compounds and efficiency of treatment, different possible side effects, etc. In many respects, vector-encoded delivery seems to offer advantages over the more traditional "pharmaceutical" approach of drug delivery. For example, the duration of effect delivered by chemically synthesized si/shRNA molecules might be shorter than the duration expected from vector-encoded molecules due to their continuous supply by intracellular transcription. In addition, these "internally synthesized" RNAi compounds are not subject to the potentially destructive environments of the organism, such as chemically synthesized and systemically or locally delivered siRNAs might be, for example, through gastrointestinal (GI), intravenous, or intramuscular delivery routes. It is usually sufficient to deliver only a small number of siRNA-expressing vector molecules per cell to obtain efficient silencing. Therefore, the delivery of RNAi compounds by these two approaches should have completely different dosage and timing strategies. In turn, the vectors for delivering RNAi compounds can be constructed using viral or DNA plasmid systems, which also might provide for or necessitate different delivery routes. It is also probable that siRNA-expression vectors will have to be customized for the delivery of targeted compounds to specific tissues (i.e., employing tissue-specific promoters, or viruses with tissue-type limited host ranges). There are currently three kinds of viral vectors in use for RNAi compound delivery: adenovirus, adeno-associated virus, and lentivirus [146]. In general, plasmid and viral RNAi-vectors, as methods of gene therapy, could have some advantages over siRNA oligonucleotides in delivery, stability, and efficacy. However, vector-associated gene therapies have their own long history of safety concerns and potential for nonspecific effects [147].

There are two fundamental means of drug delivery: local delivery and systemic delivery; each with its inherent advantages and disadvantages. Local delivery is not always practical or feasible, although it might be conceptually preferred for its potential efficiency in targeting the specific tissue or even a pool of target cells. Systemic delivery suffers from low efficiency of the relevant cells being targeted. Morrissey et al. [148] calculated that only 1% of intravenously administered (30 mg/kg) siRNAs reached the targeted organ (liver). Combined with the aforementioned differences in the forms of RNAi compounds, all these issues make the delivery problem quite complex. In these respects, nasal delivery represents one of the most convenient and powerful methods of delivery for both local and systemic delivery [149]. The substantial previous experience gained by advances made in the delivery of antisense drugs (one of which, Vitravene, was approved by the FDA in 1998) may contribute significantly to the related RNAi-based therapies [150].

Recently, substantial success in the targeted delivery of chemically synthesized siRNAs was achieved using siRNAs in the form of a supramolecular complex with a fusion protein consisting of a protamine-antibody directed at a specific cell-surface receptor [151, 152]. These complexes were found to be efficient in silencing target genes when delivered either locally (by intratumoral injection) or

systemically (by intravenous injection). The use of viral vectors might also prove efficient for systemic delivery and, as discussed above, are likely to provide long-term efficacy. Moreover, specially designed viral vectors, targeting specific tissues or even specific cells within the target tissue, might significantly improve delivery of RNAi compounds to the intended target cells. Gou et al. [153] recently demonstrated, both *in vitro* and *in vivo*, that the use of a cell-specific promoter introduced into an adenoviral vector could direct the specific expression of shRNA in one type of cell without interfering with other cell systems present. This approach seems extremely useful for the delivery of RNAi drugs. For example, as these authors suggested, to target the expression of an oncogene in cancer cells, but not in normal cells, a telomerase reverse transcriptase promoter (active in cancer cells, but inactive in normal somatic cells) [154] could be used to drive shRNA expression specifically in cancer cells.

### 6.2 Stability of RNAi Compounds

As with any drug delivered into an organism, an RNAi compound is subject to inactivation, degradation, and elimination. The good news, though, is that because RNAi drugs share the same basic chemical composition independent of the targeted disease, any findings on stabilization of compound(s) in one model should be applicable to other models and compounds, if the compounds are delivered through a similar route.

si/shRNAs are subject to enzymatic degradation in the body, as are any RNA molecules. Due to their small size and double-stranded nature, though, these structures seem to be more stable than many types of RNA and even more stable than the single-stranded DNA oligonucleotides used as antisense drugs [155]. Nevertheless, because degradation is a common problem, several different approaches have been undertaken to create compounds that may resist destruction in the cell. One way to stabilize a nucleic acid compound in the biological environment is to chemically modify it so that it will become resistant to enzymatic degradation while maintaining its intended biological function. Several such modifications to siRNAs, and the influence of these modifications on their pharmacokinetic properties, have been described [148, 156–158]. Another way of stabilizing RNAi compounds is the formation of supramolecular complexes, such as nanoparticles, which have been extensively explored for the delivery of RNAi compounds [159–163], or of liposome-based complexes, which also can be used for plasmid vector-based siRNA delivery [164–166]. Despite the method(s) chosen for delivery of chemically prepared si/shRNAs, the achievement of intracellular delivery at therapeutically effective concentrations in the target cells still represents a major challenge to the practicable use of RNAi technology for therapy, especially via systemic delivery routes.

## 7 RNAi DRUGS UNDER DEVELOPMENT

No RNAi drugs are currently on the market for use in humans. Only a few compounds have entered phase I clinical trials, and some will be starting soon (see below). In the last few years, a tremendous investment has been made in the development and application of RNAi compounds as drugs. Several recent reviews

describe these extensive efforts and the progress made in several small animal models [51, 167–169]. The effective use of RNAi–*in vivo* animal models may not only lead to considerable savings for the drug industry in the expensive development of small drugs, but it will also likely indicate whether particular RNAi compounds have significant potential for further development into drugs. Our aim in this section is to describe several interesting recent examples of RNAi drugs under development for different diseases in animal models. These studies illustrate the application of many principles and concerns discussed above.

### 7.1 Cancer

Hu-Lieskovan et al. [170] described systemic RNAi therapy for metastatic cancer in a murine model of metastatic Ewing's sarcoma. They prepared siRNA against the EWS-FLI1 gene product and delivered it using a targeted, nonviral delivery system. SiRNA was packaged into cyclodextrin-containing polycation complexes, which contained not only siRNA but also transferrin as a targeting ligand for specific delivery to transferrin receptor-expressing tumor cells. This delivery system transiently reduced expression of the target gene in metastatic cells and slowed tumor growth. Using long-term twice-weekly treatments, they demonstrated absence of any tumor growth in 80% of mice. Control experiments revealed that only the transferrin-targeted system was efficient in antitumor effects. Moreover, they found that low-pressure, low-volume tail-vein administrations were efficient in the delivery of RNAi compounds and resulted in significant improvements over the high-pressure, high-volume intravenous injections previously used in many studies of RNAi therapy in mouse models [171–173], but that are untenable and unacceptable in humans in routine clinical settings. It was also very important to the success of this study that no abnormalities in interleukin-12 and IFN-alpha, liver and kidney function tests, complete blood counts, or pathology of major organs, i.e., no off-target side effects, were observed from the treatments described. This delivery system was found to be ineffective for brain tumors, however, because the siRNA complexes used did not cross the blood-brain barrier. Consequently, different RNAi approaches will have to be developed to specifically address the treatment of brain tumors.

Yano et al. [174] used mouse models for liver metastasis and prostate cancer to test a human bcl-2 oncogene-specific siRNA complexed with a novel cationic liposome, LIC-101. These complexes were delivered by bolus intravenous injections into the livers of disease-model mice and completely suppressed the growth of liver tumors. The injection volume used was 0.1 mL per 10 g of body weight, which is normally acceptable for human therapy. In addition, in their prostate cancer model, siRNA complexes were administered subcutaneously (0.1 mg in 0.1 mL per mouse) near the tumor five times a week for 2 weeks in mice bearing PC-3 xenografts. The average tumor volume on day 36 of the treatment was 1300 $mm^3$ in the control group, compared with 487 $mm^3$ in the siRNA compound-treated group. The authors of this study also indicated that other cationic liposomes caused much greater toxicity than LIC-101, as judged by hemolysis activity on human erythrocytes, although these data were not presented.

KITENIN, a member of the Tetraspanin family, is expressed by many different tumors. Its expression is correlated with the acquisition of metastatic phenotypes

[175]. Lee et al. [176] treated mice with established colon tumors with KITENIN-targeted siRNAs cloned into pSUPER vectors. Weekly, or semiweekly, injections of these siRNA-encoding vectors, delivered in suspension with FuGENE 6 transfection reagent (Roche) for 1 month into tail veins, resulted in a marked regression of tumor size and inhibited metastases. Delivery of this siRNA also significantly prolonged the survival of mice compared with the vector alone ($77 \pm 12.2$ versus $25 \pm 9.9$ days, $p < 0.01$). In this study, it was demonstrated that the antitumor effects of KITENIN siRNA derived from both the generation of a tumor-specific immune response *in vivo* and the suppression of tumor invasion.

### 7.2 Viral and Respiratory Diseases

Hepatitis B virus (HBV) infections cause acute and chronic hepatitis and hepatocellular carcinomas. Uprichard et al. [177] demonstrated, in a mouse model of established Hepatitis B infection, that infection with an adenovirus vector carrying shRNA directed at HBV-specific viral RNA could suppress preexisting HBV gene expression and replication to almost undetectable levels for at least 26 days. Morrissey et al. [148] obtained similar results in a study using hydrodynamic tail vein injections of HBV RNA-specific siRNA stabilized by means of chemical modifications.

Influenza A virus infects about 15–20% of the population each year and is a major source of morbidity and mortality worldwide. Antiviral vaccines only afford protection against a limited range of strains, due to substantial mutation rates in these viruses. Ge et al. [178] demonstrated that siRNAs specific for conserved regions of influenza virus genes can be used to treat and prevent influenza virus infections in mice. In this study, slow i.v. administration of small volumes containing siRNAs complexed with a polycationic carrier was used. Virus titers in lungs were reduced from 10- to 1000-fold by single doses of 60–120 μg of siRNA. Similar effects were also observed when mice were given DNA vectors by i.v. or intranasally, from which siRNA precursors were transcribed. In another study using the same siRNAs, but different carriers, doses, strains of mice, and virus strains, Tompkins et al. [179] observed a significant increase in the survival of mice after lethal challenge with viruses. In both studies, the effects were siRNA-specific and not due to nonspecific antiviral IFN responses. These experiments provide a good basis for studies of prophylaxis and therapy of influenza virus infections in human populations using siRNA compounds.

Respiratory syncytial virus (RSV) and parainfluenza virus (PIV) are leading causes of respiratory disease in infants, young children, immuno-compromized patients, and the elderly. Bitko et al. [140] used viral-specific siRNAs for the treatment of these respiratory syndromes in an animal model. In this study, either, or both, viral infections could be specifically prevented and inhibited when corresponding siRNAs were infused intranasally in mice, with or without transfection reagents. These investigators noted an excellent correlation between the activities observed in their siRNAs in *in vitro* cell cultures and those observed in their animal model. This type of approach presents a fast screen for efficient siRNA sequences before more expensive and demanding applications are pursued in animals. In addition, nasal administration of siRNA compounds seems to provide a convenient

and potent route for delivery of these drugs directed against, at least, respiratory viral diseases in humans, and perhaps against many others.

### 7.3 Arthritis

We described, in Section 3, the use of electroporation for the local delivery of αTNF-specific siRNAs to inhibit joint inflammation in a mouse model of collagen-induced arthritis [50]. In another example, Inoue et al. [180] used electro-transfer of an αTNF-specific siRNA-polyamine complex to significantly ameliorate collagen-induced arthritis in rats. As αTNF plays an important role in many processes in different organs, local delivery was used, in this case, to restrict the effects to the target organ, the joint, and minimize the potential effects on other αTNF-dependent processes elsewhere in the organism.

### 7.4 Autoimmune Diseases

Autoreactive T cells play an important role in many autoimmune diseases, such as multiple sclerosis (MS), type 1 diabetes, systemic lupus erythematosus (SLE), psoriasis, asthma, and experimental autoimmune encephalomyelitis (EAE). Lovett-Racke et al. [181] studied inhibition of T-bet, a member of the T-box family of transcription factors, which is expressed in Th1 cells, but not Th2 cells, in EAE, an inflammatory and demyelinating disease of the central nervous system and model for the human disease MS. A single injection (20 or 50 μg) into the tail vein of T-bet-specific naked siRNA at the time of immunization (to induce EAE) reduced the incidence of EAE disease by as much as 75% depending on the siRNA dose. It is possible that modification of the delivery method to one more appropriate for human routine clinical settings, such as those that employ cell-specific targeting as described above in the subsection on *Cancer*, might provide a better basis for clinical trials of RNAi compounds in autoimmune disorders. In addition, the use of a targeted cell delivery system, for example, using an epitope from the T-cell receptor, similar to the transferrin-directed complexes used in [170], might provide more specific elimination of disease-relevant T cells.

### 7.5 Neurodegenerative Disorders

Alzheimer disease is a neurodegenerative disorder characterized by an accumulation of β-amyloid (Aβ), a proteolytic product of the amyloid precursor protein (APP) [182]. Increase in β-secretase (BACE1) activity has been directly associated with the production and accumulation of Aβ [183–185]. Singer et al. [186] used lentiviral vectors expressing BACE1-targeted siRNAs delivered locally by intracranial injection into the hippocampi of APP transgenic mice (a model of Alzheimer disease). This treatment specifically reduced both the cleavage of APP into Aβ and neurodegeneration in this *in vivo* model. RNAi-mediated reduction of Aβ-reactive plaques was observed only in the hippocampus and not in other parts of the brain. Behavioral deficits were, nevertheless, also improved by this treatment. Another important issue in this study was that siRNA-targeted inhibition downregulated, rather than completely knocked out, the target gene; this may have been advantageous, in this case, because BACE1 has several potential substrates and the complete switching off of its activity could produce undesired or deleterious effects. In

the effective case of RNAi-mediated downregulation of a target gene, therefore, disease improvements might be observed by lowering the level of production of a relevant protein below some critical threshold level, without significant effect on normal functions.

Huntington's disease (HD) is a dominantly inherited neurodegenerative disorder associated with polyglutamine expansion in the huntingtin (htt) protein [187], for which, currently, there is no effective treatment. Wang et al. [188] used an siRNA compound directed against human mutant-htt, containing expanded polyglutamine tracts, delivered locally into the brains of newborn HD transgenic model mice by intraventricular injections. This treatment significantly increased the longevity of HD mice, resulted in improved motor function, and slowed HD-associated loss of body weight. Despite these successes, this method is currently problematic for human therapeutic applications; not only because of the delivery route used, but also due to the challenge of designing an siRNA sequence that is highly selective for the mutant form of htt, because reduced expression of the wild-type human gene could be deleterious.

### 7.6 Ocular Disorders

In age-related macular degeneration, vascular endothelial growth factor (VEGF) plays a crucial role in destructive vascularization. Downregulation of VEGF by antibody [189], or by siRNA [190], has been shown to reduce vascular invasion. Recently, Tolentino et al. [191] used intravitreal injection of VEGF-specific siRNA in a non-human primate model of laser-induced choroidal neovascularization (CNV). They demonstrated that VEGF-specific siRNA treatment is capable of inhibiting vascular growth and permeability in a dose-dependent manner. This study served as a preclinical "proof-of-principle," and phase I clinical studies of VEGF siRNA in patients with exudative age-related macular degeneration are currently in progress [192].

## 8 COMPARISON OF SIRNAS WITH OTHER PHARMACEUTICALS (TARGETED AND NONTARGETED TYPE), INCLUDING CLOSELY RELATED GENE SPECIFIC ANTISENSE-BASED DRUGS

Most drugs on the market today are small-molecule compounds, which can be systemically delivered in the form of pills. This is certainly the most convenient method. However, it might not be the most efficient method of drug delivery, due to the different specific physical-chemical nature of the drugs. Nevertheless, the consumer market is heavily affected by a demand for consumer convenience, and pharmacological companies have generally followed this trend with much success. There is a substantial number of drugs, however, which are not pills and could not be pills due to their physical-chemical, biological, and pharmacological properties. RNAi compounds will most likely belong to this class of drugs, at least in the near future, unless substantial efforts will be invested in the specific and robust protection of the extremely sensitive and reactive RNA molecules from the enzymes of the digestive tract. It is impractical, perhaps, to expect such efforts until the efficacy and safety of these drugs are proven in clinical trials, which will most likely be decided over the next 5–7 years.

Efficient delivery of a drug to its target organ(s), tissue(s), cell(s), and molecule(s) is an important requirement for the success of any therapeutic treatment. Any single improvement in targeted delivery, even of existing compounds, may significantly reduce the exposure of irrelevant cells and organs, which would result in less potential side effects and improved drug efficacy.

Many common drugs target proteins. The interaction of small molecules with proteins often treats or lessens the disease by correcting some biochemical aspect of a pathological pathway. RNAi drugs are completely different from these common drugs in that their immediate target is an RNA molecule, and not a protein. The effect on the "target protein" is a secondary one, by downregulation of the abundance or translation of the mRNA coding for this protein. Therefore, there may be a substantial delay in the desired effects of RNAi if the turnover rate of the target protein is low, or if the protein is localized in a different tissue than where it is expressed so that this protein and its mRNA are spatially separated. For example, a small drug directed at a protein present in the blood can be delivered directly into the bloodstream by a variety of methods. To achieve a similar effect, an RNAi compound must be delivered specifically to the organs and cells that produce this protein. If a particular serum protein has a very slow turnover rate, it might not even be feasible to obtain the desired effect in a reasonable period of time using RNAi. Another example of a potentially deceptive RNAi target is anucleated cells, such as erythrocytes, which are deficient in active mRNA and protein syntheses. In this case, to make a correction in these cells, RNAi compounds should be directed at their progenitors. Yet another difference is that RNAi drugs act by reducing the *amount* of the target protein, and not by inhibiting its activity as many small-molecule drugs do. Reducing the amount of a target protein might produce effects different from the inhibition of its activity, because the mere concentration of a particular protein in a cell or tissue could determine the regulation of other processes or biochemical pathways in the organism. All of these potentially problematic aspects that derive from differences between protein and RNA targeting should, therefore, be carefully considered when conceiving the targeted design of an RNAi drug against a specific protein.

A clear advantage of RNAi drugs is their common physical-chemical nature, unlike the disparate nature of the kingdom of small molecules. Therefore, advances made in any aspect of development of a specific RNAi compound (e.g., delivery, stability, and modification) will likely translate to improvements of other RNAi compounds directed against different targets. Other obvious advantages to RNAi drugs are ease of design and synthesis, high specificity, ease of target validation, and fast initial development and redevelopment [193]. It is much easier do design and test RNAi compounds against several targets or against several sites of the same target, a task not feasible for the common small drug. This property of RNAi drugs should significantly improve efficacy and resistance against mutations and natural diversity or polymorphism.

Most similar to RNAi compounds in their application and chemical nature are antisense compounds; one of which, Vitravene, reached the market as an FDA-approved drug in 1998 [194, 195]. Antisense oligonucleotides have been studied for their ability to knock down gene expression since 1978 [196, 197]. Later studies support that the mechanism of antisense inhibition acts either through the inhibition of translation or through RNase H-dependent degradation of mRNA [198].

Both technologies, RNAi and antisense, can successfully inhibit target gene expression; and both share many practical problems, such as delivery, stability, and potential off-target effects. Nevertheless, a significant difference between these two technologies is their relative efficiency of compound design. An efficient antisense oligonucleotide must be able to bind its target mRNA under the physiological conditions that exist inside a cell. However, mRNAs in cells have folded structures, with more than 90% of their sequence unavailable for hybridization with short oligonucleotides [199]. Therefore, not all possible antisense oligonucleotides would be able to hybridize to a target mRNA. Although the RNAi mechanism involves a specialized apparatus for RNA targeting and efficient hybridization of siRNA, antisense oligonucleotides do not have this help. Hence, the ability to efficiently hybridize to a specific mRNA directly determines the potential success or failure of an antisense experiment. As a result, where one out of every five randomly chosen double-stranded short RNA molecules is likely to be effective in RNAi, only about one out of every hundred antisense oligonucleotides would be expected to have some activity. As physical forces playing a role in the hybridization process are complex, a more reliable means to screen for efficient antisense oligonucleotides would be an empirical approach [200], which can be an expensive, labor, and time-consuming process. In addition, the concentrations of RNAi compounds required for inhibition are lower than those required for antisense oligonucleotides, which might result in less nonspecific toxicity. For an extensive description and detailed comparison of both technologies, see the comprehensive review by Achenbach et al. [201]. In general, and by many criteria, RNAi technology seems to provide a more convenient and powerful technology for therapeutic treatment. Nevertheless, therapeutic application of RNAi technology is still in its infancy; although therapeutic applications of antisense technology have been under development for decades and have achieved some substantial success, with one drug on the market and several promising drug candidates in advanced stages of clinical trials [150, 194]. In fact, many advances in antisense technology (e.g., modification and stabilization of compounds, delivery routes, aversion of side effects) have been, and will continue to be, applicable to the related field of RNAi technology. Hence, it is, perhaps, currently difficult to judge which technology may be superior, but the potential of both for delivering effective therapeutic applications will definitely continue to be explored in the future.

## 9 BIOTECH AND DRUG COMPANIES CURRENTLY INVESTED IN RNAI DRUG DISCOVERY AND DEVELOPMENT

Naturally, recognizing its great utility and success in multiple research applications and its high potential for therapy, numerous industrial companies and research institutes have taken an interest and joined the RNAi-therapy game. Over a dozen biotech companies are currently competing in the development of RNAi pharmaceuticals. There seem to be even more businesses exploiting the RNAi boom by providing different kinds of RNAi-related products and services to desirous research teams, such as the synthesis of RNAi compounds and libraries; the design, construction, and distribution of RNAi vectors; and customized transfection

reagents. (e.g., Invitrogen, Ambion, QIAGEN, Dharmacon). Changes are occurring fast on this playing field, as some companies have quickly been "taken out of the game" by others (e.g., Atugen was recently acquired by SR Pharma plc), and additional companies join in by changing their focus to RNAi technology (e.g., Genta). Several biotech companies have filed for FDA approval of clinical trials with RNAi compounds (Table 1).

Collectively, the companies invested in the RNAi drug business have over 40 RNAi compounds engaged in therapeutic projects, which cover a wide range of diseases, including cancers (10 compounds), infectious diseases (mostly viral, 11 compounds), neurological disorders (5 compounds), allergy and inflammation (2 compounds), diabetes, obesity, and others [202]. Some major players have been Sirna Therapeutics, Alnylam Pharmaceuticals, and Acuity Pharmaceuticals, which all compete in the development of RNAi drugs directed against age-related macular degeneration (AMD). This common choice of target disease does not seem to be coincidental, as (1) no cure is currently available for AMD, (2) the target gene product VEGF has been clearly implicated in the development of the disease, and (3) VEGF-specific siRNA compounds were found in prior studies to inhibit its expression and the growth and vascular permeability associated with the disease [189–191] (also, as discussed in Section 7). Most importantly, local delivery through topical application is feasible in this disease of the eye, assuring appropriate therapeutic concentrations of the compound and minimizing potential side effects. If there are no complications due to safety issues, in a few years we may see who will win the race to establish this new class of gene targeted medicines [203]. In the near future, the market for RNAi technology and related products will surely continue its rapid expansion [204], and more exciting news from biotech companies on advances in RNAi drug development will come into play.

## 10 CONCLUDING REMARKS

Extensive research and development efforts in RNAi technology during the last few years have resulted in enormous success in our ability to downregulate the expression of almost any gene in cell culture and in many different animal models. This success has created great enthusiasm and hope that the exploitation of this powerful technology will soon result in highly effective and safe new drugs, able to cure diseases, which were incurable or difficult to cure before. A significant concentration of research efforts in four key areas: target selection, targeted delivery, compound stability, and minimization of off-target effects should result in superior RNAi drugs. We shall see, probably within the next 5–7 years, whether expectations can be realized and deliver the first RNAi-based drug for human use on the market.

## ACKNOWLEDGMENT

We thank Susan Francis and Vlad Malkov for their critical review of the manuscript and valuable discussion.

## REFERENCES


1. Fire A, et al. (1998). Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. *Nature.* 391(6669):806–811.
2. Powers J, et al. (1998). A nematode kinesin required for cleavage furrow advancement. *Curr. Biol.* 8(20):1133–1136.
3. Kennerdell J R, Carthew R W (1998). Use of dsRNA-mediated genetic interference to demonstrate that frizzled and frizzled 2 act in the wingless pathway. *Cell.* 95(7):1017–1026.
4. Kennerdell J R, Carthew R W (2000). Heritable gene silencing in Drosophila using double-stranded RNA. *Nat. Biotechnol.* 18(8):896–898.
5. Li Y X, et al. (2000). Double-stranded RNA injection produces null phenotypes in zebrafish. *Dev. Biol.* 217(2):394–405.
6. Zhou Y, et al. (2002). Post-transcriptional suppression of gene expression in Xenopus embryos by small interfering RNA. *Nucleic Acids Res.* 30(7):1664–1669.
7. Svoboda P, et al. (2000). Selective reduction of dormant maternal mRNAs in mouse oocytes by RNA interference. *Development.* 127(19):4147–4156.
8. Svoboda P, Stein P, Schultz R M (2001). RNAi in mouse oocytes and preimplantation embryos: Effectiveness of hairpin dsRNA. *Biochem. Biophys. Res. Commun.* 287(5):1099–1104.
9. Hamilton A J, Baulcombe D C (1999). A species of small antisense RNA in post-transcriptional gene silencing in plants. *Science.* 286(5441):950–952.
10. Zamore P D, et al. (2000). RNAi: Double-stranded RNA directs the ATP-dependent cleavage of mRNA at 21 to 23 nucleotide intervals. *Cell.* 101(1):25–33.
11. Hammond S M, et al. (2000). An RNA-directed nuclease mediates post-transcriptional gene silencing in Drosophila cells. *Nature.* 404(6775):293–296.
12. Moss E G (2001). RNA interference: It's a small RNA world. *Curr. Biol.* 11(19): R772–R775.
13. Filipowicz W, et al. (2005). Post-transcriptional gene silencing by siRNAs and miRNAs. *Curr. Opin. Struct. Biol.* 15(3):331–341.
14. Tang G (2005). siRNA and miRNA: An insight into RISCs. *Trends Biochem. Sci.* 30(2):106–114.
15. Mullis K, et al. (1986). Specific enzymatic amplification of DNA in vitro: The polymerase chain reaction. *Cold Spring Harb. Symp. Quant. Biol.* 51(Pt 1):263–273.
16. Scharf S J, Horn G T, Erlich H A (1986). Direct cloning and sequence analysis of enzymatically amplified genomic sequences. *Science.* 233(4768):1076–1078.
17. Saiki R K, et al. (1986). Analysis of enzymatically amplified beta-globin and HLA-DQ alpha DNA with allele-specific oligonucleotide probes. *Nature.* 324(6093):163–166.
18. Napoli C, Lemieux C, Jorgensen R (1990). Introduction of a chimeric chalcone synthase gene into petunia results in reversible co-suppression of homologous genes in trans. *Plant Cell.* 2(4):279–289.
19. Tabara H, et al. (1999). The rde-1 gene, RNA interference, and transposon silencing in C. elegans. *Cell.* 99(2):123–132.
20. Hammond S M, Caudy A A, Hannon G J (2001). Post-transcriptional gene silencing by double-stranded RNA. *Nat. Rev. Genet.* 2(2):110–119.
21. Elbashir S M, Lendeckel W, Tuschl T (2001). RNA interference is mediated by 21- and 22-nucleotide RNAs. *Genes Dev.* 15(2):188–200.
22. Bernstein E, et al. (2001). Role for a bidentate ribonuclease in the initiation step of RNA interference. *Nature.* 409(6818):363–366.

23. Hammond S M (2005). Dicing and slicing: The core machinery of the RNA interference pathway. *FEBS Lett.* 579(26):5822–5829.
24. Hammond S M, et al. (2001). Argonaute2, a link between genetic and biochemical analyses of RNAi. *Science.* 293(5532):1146–1150.
25. Schwarz D S, et al. (2003). Asymmetry in the assembly of the RNAi enzyme complex. *Cell.* 115(2):199–208.
26. Khvorova A, Reynolds A, Jayasena S D (2003). Functional siRNAs and miRNAs exhibit strand bias. *Cell.* 115(2):209–216.
27. Samuel C E (2001). Antiviral actions of interferons. *Clin. Microbiol. Rev.* 14(4):778–809.
28. Karpala A J, Doran T J, Bean A G (2005). Immune responses to dsRNA: Implications for gene silencing technologies. *Immunol. Cell Biol.* 83(3):211–216.
29. Elbashir S M, et al. (2001). Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. *Nature.* 411(6836):494–498.
30. Lee R C, Feinbaum R L, Ambros V (1993). The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. *Cell.* 75(5):843–854.
31. Wightman B, Ha I, Ruvkun G (1993). Posttranscriptional regulation of the heterochronic gene lin-14 by lin-4 mediates temporal pattern formation in C. elegans. *Cell.* 75(5):855–862.
32. Ruvkun G, Wightman B, Ha I (2004). The 20 years it took to recognize the importance of tiny RNAs. *Cell.* 116(suppl 2):S93–96, 2 p following S96.
33. Tijsterman M, Plasterk R H (2004). Dicers at RISC; the mechanism of RNAi. *Cell.* 117(1):1–3.
34. Cai X, Hagedorn C H, Cullen B R (2004). Human microRNAs are processed from capped, polyadenylated transcripts that can also function as mRNAs. *Rna.* 10(12): 1957–1966.
35. Denli A M, et al. (2004). Processing of primary microRNAs by the Microprocessor complex. *Nature.* 432(7014):231–235.
36. Liu J, et al. (2005). A role for the P-body component GW182 in microRNA function. *Nat. Cell Biol.* 7(12):1161–1166.
37. Carpenter A E, Sabatini D M (2004). Systematic genome-wide screens of gene function. *Nat. Rev. Genet.* 5(1):11–22.
38. Keown W A, Campbell C R, Kucherlapati R S (1990). Methods for introducing DNA into mammalian cells. *Methods Enzymol.* 185:527–537.
39. Brummelkamp T R, Bernards R, Agami R (2002). A system for stable expression of short interfering RNAs in mammalian cells. *Science.* 296(5567):550–553.
40. Sui G, et al. (2002). A DNA vector-based RNAi technology to suppress gene expression in mammalian cells. *Proc. Natl. Acad. Sci. USA.* 99(8):5515–5520.
41. Yu J Y, DeRuiter S L, Turner D L (2002). RNA interference by expression of short-interfering RNAs and hairpin RNAs in mammalian cells. *Proc. Natl. Acad. Sci. USA.* 99(9):6047–6052.
42. Miyagishi M, Taira K (2002). U6 promoter-driven siRNAs with four uridine 3′ overhangs efficiently suppress targeted gene expression in mammalian cells. *Nat. Biotechnol.* 20(5):497–500.
43. Paddison P J, et al. (2002). Short hairpin RNAs (shRNAs) induce sequence-specific silencing in mammalian cells. *Genes Dev.* 16(8):948–958.
44. Katahira T, Nakamura H (2003). Gene silencing in chick embryos with a vector-based small interfering RNA system. *Dev. Growth Differ.* 45(4):361–367.

45. Grillo M, et al. (2005). Validation of cyclin D1/CDK4 as an anticancer drug target in MCF-7 breast cancer cells: Effect of regulated overexpression of cyclin D1 and siRNA-mediated inhibition of endogenous cyclin D1 and CDK4 expression. *Breast Cancer Res. Treat.* 95:185–194.
46. Hill M, et al. (2005). Heme oxygenase-1 inhibits rat and human breast cancer cell proliferation: Mutual cross inhibition with indoleamine 2,3-dioxygenase. *Faseb J.* 19(14):1957–1968.
47. Lomas-Neira J L, et al. (2005). In vivo gene silencing (with siRNA) of pulmonary expression of MIP-2 versus KC results in divergent effects on hemorrhage-induced, neutrophil-mediated septic acute lung injury. *J. Leukoc. Biol.* 77(6):846–853.
48. Arend W P (2001). Physiology of cytokine pathways in rheumatoid arthritis. *Arthritis Rheum.* 45(1):101–106.
49. Schiffelers R M, et al. (2005). Effects of treatment with small interfering RNA on joint inflammation in mice with collagen-induced arthritis. *Arthritis Rheum.* 52(4): 1314–1318.
50. Hommel J D, et al. (2003). Local gene knockdown in the brain using viralmediated RNA interference. *Nat. Med.* 9(12):1539–1544.
51. Leung R K, Whittaker P A (2005). RNA interference: From gene silencing to gene-specific therapeutics. *Pharmacol. Ther.* 107(2):222–239.
52. Xin H, et al. (2004). High-throughput siRNA-based functional target validation. *J. Biomol. Screen.* 9(4):286–293.
53. Devroe E, Silver P A (2002). Retrovirus-delivered siRNA. *BMC Biotechnol.* 2:15.
54. Abbas-Terki T, et al. (2002). Lentiviral-mediated RNA interference. *Hum. Gene Ther.* 13(18):2197–2201.
55. Chang F H, et al. (2004). Surfection: A new platform for transfected cell arrays. *Nucleic Acids Res.* 32(3):e33.
56. Kumar R, Conklin D S, Mittal V (2003). High-throughput selection of effective RNAi probes for gene silencing. *Genome Res.* 13(10):2333–2340.
57. Mousses S, et al. (2003). RNAi microarray analysis in cultured mammalian cells. *Genome Res.* 13(10):2341–2347.
58. Silva J M, et al. (2004). RNA interference microarrays: High-throughput loss-of-function genetics in mammalian cells. *Proc. Natl. Acad. Sci. USA.* 101(17):6548–6552.
59. Ziauddin J, Sabatini D M (2001). Microarrays of cells expressing defined cDNAs. *Nature.* 411(6833):107–110.
60. Berns K, et al. (2004). A large-scale RNAi screen in human cells identifies new components of the p53 pathway. *Nature.* 428(6981):431–437.
61. Paddison P J, et al. (2004). A resource for large-scale RNA-interference-based screens in mammals. *Nature.* 428(6981):427–431.
62. van Es H H, Arts G J (2005). Biology calls the targets: Combining RNAi and disease biology. *Drug Discov. Today.* 10(20):1385–1391.
63. Zhou A, et al. (2003). Identification of NF-kappa B-regulated genes induced by TNFalpha utilizing expression profiling and RNA interference. *Oncogene.* 22(13): 2054–2064.
64. Williams N S, et al. (2003). Identification and validation of genes involved in the pathogenesis of colorectal cancer using cDNA microarrays and RNA interference. *Clin. Cancer Res.* 9(3):931–946.
65. Gunton J E, et al. (2005). Loss of ARNT/HIF1beta mediates altered gene expression and pancreatic-islet dysfunction in human type 2 diabetes. *Cell.* 122(3):337–349.

66. Semizarov D, Kroeger P, Fesik S (2004). siRNA-mediated gene silencing: A global genome view. *Nucleic Acids Res.* 32(13):3836–3845.
67. Fitzgerald K (2005). RNAi versus small molecules: Different mechanisms and specificities can lead to different outcomes. *Curr. Opin. Drug Discov. Devel.* 8(5):557–566.
68. Martinez L A, et al. (2002). Synthetic small inhibiting RNAs: Efficient tools to inactivate oncogenic mutations and restore p53 pathways. *Proc. Natl. Acad. Sci. USA.* 99(23):14849–14854.
69. Fluiter K, et al. (2003). Killing cancer by targeting genes that cancer cells have lost: Allele-specific inhibition, a novel approach to the treatment of genetic disorders. *Cell Mol. Life Sci.* 60(5):834–843.
70. Miller V M, et al. (2003). Allele-specific silencing of dominant disease genes. *Proc. Natl. Acad. Sci. USA.* 100(12):7195–7200.
71. Lee M T, et al. (2003). Inhibition of human immunodeficiency virus type 1 replication in primary macrophages by using Tat- or CCR5-specific small interfering RNAs expressed from a lentivirus vector. *J. Virol.* 77(22):11964–11972.
72. Wu K L, et al. (2005). Inhibition of Hepatitis B virus gene expression by single and dual small interfering RNA treatment. *Virus Res.* 112(1–2):100–107.
73. Chiu Y L, et al. (2004). Inhibition of human immunodeficiency virus type 1 replication by RNA interference directed against human transcription elongation factor P-TEFb (CDK9/CyclinT1). *J. Virol.* 78(5):2517–2529.
74. Fritsch L, et al. (2004). Conditional gene knock-down by CRE-dependent short interfering RNAs. *EMBO Rep.* 5(2):178–182.
75. Ventura A, et al. (2004). Cre-lox-regulated conditional RNA interference from transgenes. *Proc. Natl. Acad. Sci. USA.* 101(28):10380–10385.
76. Matsukura S, Jones P A, Takai D (2003). Establishment of conditional vectors for hairpin siRNA knockdowns. *Nucleic Acids Res.* 31(15):e77.
77. Malphettes L, Fussenegger M (2004). Macrolide- and tetracycline-adjustable siRNA-mediated gene silencing in mammalian cells using polymerase II-dependent promoter derivatives. *Biotechnol. Bioeng.* 88(4):417–425.
78. Gupta S, et al. (2004). Inducible, reversible, and stable RNA interference in mammalian cells. *Proc. Natl. Acad. Sci. USA.* 101(7):1927–1932.
79. Weber W, et al. (2005). Engineered Streptomyces quorum-sensing components enable inducible siRNA-mediated translation control in mammalian cells and adjustable transcription control in mice. *J. Gene Med.* 7(4):518–525.
80. Schubert S, et al. (2005). Maintaining inhibition: siRNA double expression vectors against coxsackieviral RNAs. *J. Mol. Biol.* 346(2):457–465.
81. Heidenreich O, et al. (2003). AML1/MTG8 oncogene suppression by small interfering RNAs supports myeloid differentiation of t(8;21)-positive leukemic cells. *Blood.* 101(8):3157–3163.
82. McWhirter J R, et al. (1999). Oncogenic homeodomain transcription factor E2A-Pbx1 activates a novel WNT gene in pre-B acute lymphoblastoid leukemia. *Proc. Natl. Acad. Sci. USA.* 96(20):11464–11469.
83. Mazieres J, et al. (2005). Inhibition of Wnt16 in human acute lymphoblastoid leukemia cells containing the t(1;19) translocation induces apoptosis. *Oncogene.* 24(34):5396–5400.
84. Nieth C, et al. (2003). Modulation of the classical multidrug resistance (MDR) phenotype by RNA interference (RNAi). *FEBS Lett.* 545(2–3):144–150.

85. Peng Z, et al. (2004). Reversal of P-glycoprotein-mediated multidrug resistance with small interference RNA (siRNA) in leukemia cells. *Cancer Gene Ther.* 11(11):707–712.
86. Stege A, et al. (2004). Stable and complete overcoming of MDR1/P-glycoprotein-mediated multidrug resistance in human gastric carcinoma cells by RNA interference. *Cancer Gene Ther.* 11(11):699–706.
87. Futami T, et al. (2002). Induction of apoptosis in HeLa cells with siRNA expression vector targeted against bcl-2. *Nucleic Acids Symposium Ser.* 2(1):251–252.
88. Monks N R, Biswas D K, Pardee A B (2004). Blocking anti-apoptosis as a strategy for cancer chemotherapy: NF-kappaB as a target. *J. Cell Biochem.* 92(4):646–650.
89. Tarnawski A, et al. (2005). Rebamipide inhibits gastric cancer growth by targeting survivin and Aurora-B. *Biochem. Biophys. Res. Commun.* 334(1):207–212.
90. Mack P C, et al. (2004). Enhancement of radiation cytotoxicity by UCN-01 in non-small cell lung carcinoma cells. *Radiat. Res.* 162(6):623–634.
91. Peng Y, et al. (2002). Silencing expression of the catalytic subunit of DNA-dependent protein kinase by small interfering RNA sensitizes human cells for radiation-induced chromosome damage, cell killing, and mutation. *Cancer Res.* 62(22):6400–6404.
92. Collis S J, et al. (2005). The life and death of DNA-PK. *Oncogene.* 24(6):949–961.
93. Tagliaferri P, et al. (2005). Antitumor therapeutic strategies based on the targeting of epidermal growth factor-induced survival pathways. *Curr. Drug Targets.* 6(3):289–300.
94. Bagasra O (2005). RNAi as an antiviral therapy. *Expert Opin. Biol. Ther.* 5(11):1463–1474.
95. Cullen B R (2005). Does RNA interference have a future as a treatment for HIV-1 induced disease? *AIDS Rev.* 7(1):22–25.
96. Bennasser Y, et al. (2005). Evidence that HIV-1 encodes an siRNA and a suppressor of RNA silencing. *Immunity.* 22(5):607–619.
97. Martinez M A, et al. (2002). Suppression of chemokine receptor expression by RNA interference allows for inhibition of HIV-1 replication. *Aids.* 16(18):2385–2390.
98. Nair M P, et al. (2005). RNAi-directed inhibition of DC-SIGN by dendritic cells: Prospects for HIV-1 therapy. *Aaps J.* 7(3):E572–578.
99. Gitlin L, Andino R (2003). Nucleic acid-based immune system: The antiviral potential of mammalian RNA silencing. *J. Virol.* 77(13):7159–7165.
100. Randall G, Rice C M (2004). Interfering with hepatitis C virus RNA replication. *Virus Res.* 102(1):19–25.
101. Giladi H, et al. (2003). Small interfering RNA inhibits hepatitis B virus replication in mice. *Mol. Ther.* 8(5):769–776.
102. Klein C, et al. (2003). Inhibition of hepatitis B virus replication in vivo by nucleoside analogues and siRNA. *Gastroenterol.* 125(1):9–18.
103. Nahlen B L, et al. (2005). Malaria risk: Estimating clinical episodes of malaria. *Nature.* 437(7056):E3; discussion E4–5.
104. Plowe C V (2005). Antimalarial drug resistance in Africa: Strategies for monitoring and deterrence. *Curr. Top. Microbiol. Immunol.* 295:55–79.
105. Ngo H, et al. (1998). Double-stranded RNA induces mRNA degradation in Trypanosoma brucei. *Proc. Natl. Acad. Sci. USA.* 95(25):14687–14692.
106. Ullu E, Tschudi C, Chakraborty T (2004). RNA interference in protozoan parasites. *Cell Microbiol.* 6(6):509–519.
107. Djikeng A, et al. (2004). Analysis of gene function in Trypanosoma brucei using RNA interference. *Meth. Mol. Biol.* 265:73–83.

108. Montalvetti A, et al. (2003). Farnesyl pyrophosphate synthase is an essential enzyme in Trypanosoma brucei. In vitro RNA interference and in vivo inhibition studies. *J. Biol. Chem.* 278(19):17075–17083.

109. Witola W H, et al. (2004). RNA-interference silencing of the adenosine transporter-1 gene in Trypanosoma evansi confers resistance to diminazene aceturate. *Exp. Parasitol.* 107(1–2):47–57.

110. Sheader K, et al. (2005). Variant surface glycoprotein RNA interference triggers a precytokinesis cell cycle arrest in African trypanosomes. *Proc. Natl. Acad. Sci. USA.* 102(24):8716–8721.

111. McRobert L, McConkey G A (2002). RNA interference (RNAi) inhibits growth of Plasmodium falciparum. *Mol. Biochem. Parasitol.* 119(2):273–278.

112. Kumar R, et al. (2002). Characterisation and expression of a PP1 serine/threonine protein phosphatase (PfPP1) from the malaria parasite, Plasmodium falciparum: Demonstration of its essential role using RNA interference. *Malar. J.* 1:5.

113. Malhotra P, et al. (2002). Double-stranded RNA-mediated gene silencing of cysteine proteases (falcipain-1 and -2) of Plasmodium falciparum. *Mol. Microbiol.* 45(5):1245–1254.

114. Mohmmed A, et al. (2003). In vivo gene silencing in Plasmodium berghei–a mouse malaria model. *Biochem. Biophys. Res. Commun.* 309(3):506–511.

115. Patankar S, et al. (2001). Serial analysis of gene expression in Plasmodium falciparum reveals the global expression profile of erythrocytic stages and the presence of antisense transcripts in the malarial parasite. *Mol. Biol. Cell.* 12(10):3114–3125.

116. Militello K T, et al. (2005). RNA polymerase II synthesizes antisense RNA in Plasmodium falciparum. *RNA.* 11(4):365–370.

117. Ding H, et al. (2003). Selective silencing by RNAi of a dominant allele that causes amyotrophic lateral sclerosis. *Aging Cell.* 2(4):209–217.

118. Maxwell M M, et al. (2004). RNA interference-mediated silencing of mutant superoxide dismutase rescues cyclosporin A-induced death in cultured neuroblastoma cells. *Proc. Natl. Acad. Sci. USA.* 101(9):3178–3183.

119. Raoul C, et al. (2005). Lentiviral-mediated silencing of SOD1 through RNA interference retards disease onset and progression in a mouse model of ALS. *Nat. Med.* 11(4):423–428.

120. Ralph G S, et al. (2005). Silencing mutant SOD1 using RNAi protects against neurodegeneration and extends survival in an ALS model. *Nat. Med.* 11(4):429–433.

121. Xia X G, et al. (2005). An RNAi strategy for treatment of amyotrophic lateral sclerosis caused by mutant Cu, Zn superoxide dismutase. *J. Neurochem.* 92(2):362–367.

122. Parrish S, et al. (2000). Functional anatomy of a dsRNA trigger: Differential requirement for the two trigger strands in RNA interference. *Mol. Cell.* 6(5):1077–1087.

123. Sharp P A (2001). RNA interference–2001. *Genes Dev.* 15(5):485–490.

124. Gitlin L, Karelsky S, Andino R (2002). Short interfering RNA confers intracellular antiviral immunity in human cells. *Nature.* 418(6896):430–434.

125. Yang D, et al. (2002). Short RNA duplexes produced by hydrolysis with Escherichia coli RNase III mediate effective RNA interference in mammalian cells. *Proc. Natl. Acad. Sci. USA.* 99(15):9942–9947.

126. Myers J W, et al. (2003). Recombinant Dicer efficiently converts large dsRNAs into siRNAs suitable for gene silencing. *Nat. Biotechnol.* 21(3):324–328.

127. Oleinikov A V, et al. (2003). Self-assembling protein arrays using electronic semiconductor microchips and in vitro translation. *J. Proteome Res.* 2:313–319.

128. Oleinikov A V, Zhao J, Gray M D (2005). RNA interference by mixtures of siRNAs prepared using custom oligonucleotide arrays. *Nucl. Acids Res.* 33(10):e92.
129. Sledz C A, et al. (2003). Activation of the interferon system by short-interfering RNAs. *Nat. Cell Biol.* 5(9):834–839.
130. Robbins M A, Rossi J J (2005). Sensing the danger in RNA. *Nat. Med.* 11(3):250–251.
131. Zhou H S, Liu D P, Liang C C (2004). Challenges and strategies: The immune responses in gene therapy. *Med. Res. Rev.* 24(6):748–761.
132. Scacheri P C, et al. (2004). Short interfering RNAs can induce unexpected and divergent changes in the levels of untargeted proteins in mammalian cells. *Proc. Natl. Acad. Sci. USA.* 101(7):1892–1897.
133. Kariko K, et al. (2004). Small interfering RNAs mediate sequence-independent gene suppression and induce immune activation by signaling through toll-like receptor 3. *J. Immunol.* 172(11):6545–6549.
134. Kariko K, et al. (2004). Exogenous siRNA mediates sequence-independent gene suppression by signaling through toll-like receptor 3. *Cells Tissues Organs.* 177(3):132–138.
135. Persengiev S P, Zhu X, Green M R (2004). Nonspecific, concentration-dependent stimulation and repression of mammalian gene expression by small interfering RNAs (siRNAs). *RNA.* 10(1):12–18.
136. Ma Z, et al. (2005). Cationic lipids enhance siRNA-mediated interferon response in mice. *Biochem. Biophys. Res. Commun.* 330(3):755–759.
137. Tomanin R, Scarpa M (2004). Why do we need new gene therapy viral vectors? Characteristics, limitations and future perspectives of viral vector transduction. *Curr. Gene Ther.* 4(4):357–372.
138. McManus M T, et al. (2002). Small interfering RNA-mediated gene silencing in T lymphocytes. *J. Immunol.* 169(10):5754–5760.
139. Holen T, et al. (2002). Positional effects of short interfering RNAs targeting the human coagulation trigger Tissue Factor. *Nucl. Acids Res.* 30(8):1757–1766.
140. Bitko V, et al. (2005). Inhibition of respiratory viruses by nasally administered siRNA. *Nat. Med.* 11(1):50–55.
141. Gong H, et al. (2005). The role of small RNAs in human diseases: Potential troublemaker and therapeutic tools. *Med. Res. Rev.* 25(3):361–381.
142. Pusch O, et al. (2003). Nucleotide sequence homology requirements of HIV-1-specific short hairpin RNA. *Nucl. Acids Res.* 31(22):6444–6449.
143. Saxena S, Jonsson Z O, Dutta A (2003). Small RNAs with imperfect match to endogenous mRNA repress translation. Implications for off-target activity of small inhibitory RNA in mammalian cells. *J. Biol. Chem.* 278(45):44312–44319.
144. Holen T, et al. (2005). Tolerated wobble mutations in siRNAs decrease specificity, but can enhance activity in vivo. *Nucl. Acids Res.* 33(15):4704–4710.
145. Jackson A L, et al. (2003). Expression profiling reveals off-target gene regulation by RNAi. *Nat. Biotechnol.* 21(6):635–637.
146. Devroe E, Silver P A (2004). Therapeutic potential of retroviral RNAi vectors. *Expert Opin. Biol. Ther.* 4(3):319–327.
147. Yi Y, Hahm S H, Lee K H (2005). Retroviral gene therapy: Safety issues and possible solutions. *Curr. Gene Ther.* 5(1):25–35.
148. Morrissey D V, et al. (2005). Activity of stabilized short interfering RNA in a mouse model of hepatitis B virus replication. *Hepatol.* 41(6):1349–1356.

149. Johnson P H, Quay S C (2005). Advances in nasal drug delivery through tight junction technology. *Expert Opin. Drug. Deliv.* 2(2):281–298.
150. Patil S D, Rhodes D G, Burgess D J (2005). DNA-based therapeutics and DNA delivery systems: A comprehensive review. *Aaps J.* 7(1):E61–77.
151. Song E, et al. (2005). Antibody mediated in vivo delivery of small interfering RNAs via cell-surface receptors. *Nat. Biotechnol.* 23(6):709–717.
152. Williams B R (2005). Targeting specific cell types with silencing RNA. *N. Engl. J. Med.* 353(13):1410–1411.
153. Gou D, et al. (2004). Gene silencing in alveolar type II cells using cell-specific promoter in vitro and in vivo. *Nucl. Acids Res.* 32(17):e134.
154. Gu J, et al. (2002). A novel single tetracycline-regulative adenoviral vector for tumor-specific Bax gene expression and cell killing in vitro and in vivo. *Oncogene.* 21(31): 4757–4764.
155. Bertrand J R, et al. (2002). Comparison of antisense oligonucleotides and siRNAs in cell culture and in vivo. *Biochem. Biophys. Res. Commun.* 296(4):1000–1004.
156. Soutschek J, et al. (2004). Therapeutic silencing of an endogenous gene by systemic administration of modified siRNAs. *Nature.* 432(7014):173–178.
157. Chen X, et al. (2005). Chemical modification of gene silencing oligonucleotides for drug discovery and development. *Drug Discov. Today.* 10(8):587–593.
158. de Fougerolles A, et al. (2005). RNA interference in vivo: Toward synthetic small inhibitory RNA-based therapeutics. *Meth. Enzymol.* 392:278–296.
159. Guo S, et al. (2005). Specific delivery of therapeutic RNAs to cancer cells via the dimerization mechanism of phi29 motor pRNA. *Hum. Gene Ther.* 16(9):1097–1109.
160. Kakizawa Y, Furukawa S, Kataoka K (2004). Block copolymer-coated calcium phosphate nanoparticles sensing intracellular environment for oligodeoxynucleotide and siRNA delivery. *J. Control. Rel.* 97(2):345–356.
161. Schiffelers R M, et al. (2004). Cancer siRNA therapy by tumor selective delivery with ligand-targeted sterically stabilized nanoparticle. *Nucl. Acids Res.* 32(19):e149.
162. Lochmann D, et al. (2005). Albumin-protamine-oligonucleotide nanoparticles as a new antisense delivery system. Part 1: Physicochemical characterization. *Eur. J. Pharm. Biopharm.* 59(3):419–429.
163. Khaled A, et al. (2005). Controllable self-assembly of nanoparticles for specific delivery of multiple therapeutic molecules to cancer cells using RNA nanotechnology. *Nano. Lett.* 5(9):1797–1808.
164. Safinya C R (2001). Structures of lipid-DNA complexes: Supramolecular assembly and gene delivery. *Curr. Opin. Struct. Biol.* 11(4):440–448.
165. Paul C P, et al. (2002). Effective expression of small interfering RNA in human cells. *Nat. Biotechnol.* 20(5):505–508.
166. Sioud M (2005). siRNA delivery in vivo. *Meth. Mol. Biol.* 309:237–249.
167. Uprichard S L (2005). The therapeutic potential of RNA interference. *FEBS Lett.* 579(26):5996–6007.
168. Pai S I, et al. (2005). Prospects of RNA interference therapy for cancer. *Gene Ther.* 8:8.
169. Spankuch B, Strebhardt K (2005). RNA interference-based gene silencing in mice: The development of a novel therapeutical strategy. *Curr. Pharm. Des.* 11(26):3405–3419.
170. Hu-Lieskovan S, et al. (2005). Sequence-specific knockdown of EWS-FLI1 by targeted, nonviral delivery of small interfering RNA inhibits tumor growth in a murine model of metastatic Ewing's sarcoma. *Cancer Res.* 65(19):8984–8992.

171. Lewis D L, et al. (2002). Efficient delivery of siRNA for inhibition of gene expression in postnatal mice. *Nat. Genet.* 32(1):107–108.
172. McCaffrey A P, et al. (2002). RNA interference in adult mice. *Nature.* 418(6893): 38–39.
173. Song E, et al. (2003). RNA interference targeting Fas protects mice from fulminant hepatitis. *Nat. Med.* 9(3):347–351.
174. Yano J, et al. (2004). Antitumor activity of small interfering RNA/cationic liposome complex in mouse models of cancer. *Clin. Cancer Res.* 10(22):7721–7726.
175. Lee J H, et al. (2004). KAI1 COOH-terminal interacting tetraspanin (KITENIN), a member of the tetraspanin family, interacts with KAI1, a tumor metastasis suppressor, and enhances metastasis of cancer. *Cancer Res.* 64(12):4235–4243.
176. Lee J H, et al. (2005). Suppression of progression and metastasis of established colon tumors in mice by intravenous delivery of short interfering RNA targeting KITENIN, a metastasis-enhancing protein. *Cancer Res.* 65(19):8993–9003.
177. Uprichard S L, et al. (2005). Clearance of hepatitis B virus from the liver of transgenic mice by short hairpin RNAs. *Proc. Natl. Acad. Sci. USA.* 102(3):773–778.
178. Ge Q, et al. (2004). Inhibition of influenza virus production in virus-infected mice by RNA interference. *Proc. Natl. Acad. Sci. USA.* 101(23):8676–8681.
179. Tompkins S M, et al. (2004). Protection against lethal influenza virus challenge by RNA interference in vivo. *Proc. Natl. Acad. Sci. USA.* 101(23):8682–8686.
180. Inoue A, et al. (2005). Electro-transfer of small interfering RNA ameliorated arthritis in rats. *Biochem. Biophys. Res. Commun.* 336(3):903–908.
181. Lovett-Racke A E, et al. (2004). Silencing T-bet defines a critical role in the differentiation of autoreactive T lymphocytes. *Immunity.* 21(5):719–731.
182. Selkoe D J, (2004). Alzheimer disease: Mechanistic understanding predicts novel therapies. *Ann. Intern. Med.* 140(8):627–638.
183. Holsinger R M, et al. (2002). Increased expression of the amyloid precursor beta-secretase in Alzheimer's disease. *Ann. Neurol.* 51(6):783–786.
184. Fukumoto H, et al. (2002). Beta-secretase protein and activity are increased in the neocortex in Alzheimer disease. *Arch. Neurol.* 59(9):1381–1389.
185. Li R, et al. (2004). Amyloid beta peptide load is correlated with increased beta-secretase activity in sporadic Alzheimer's disease patients. *Proc. Natl. Acad. Sci. USA.* 101(10):3632–3637.
186. Singer O, et al. (2005). Targeting BACE1 with siRNAs ameliorates Alzheimer disease neuropathology in a transgenic model. *Nat. Neurosci.* 8(10):1343–1349.
187. Gusella J F, MacDonald M E (2000). Molecular genetics: Unmasking polyglutamine triggers in neurodegenerative disease. *Nat. Rev. Neurosci.* 1(2):109–115.
188. Wang Y L, et al. (2005). Clinico-pathological rescue of a model mouse of Huntington's disease by siRNA. *Neurosci. Res.* 53(3):241–249.
189. Krzystolik M G, et al. (2002). Prevention of experimental choroidal neovascularization with intravitreal anti-vascular endothelial growth factor antibody fragment. *Arch. Ophthalmol.* 120(3):338–346.
190. Reich S J, et al. (2003). Small interfering RNA (siRNA) targeting VEGF effectively inhibits ocular neovascularization in a mouse model. *Mol. Vis.* 9:210–216.
191. Tolentino M J, et al. (2004). Intravitreal injection of vascular endothelial growth factor small interfering RNA inhibits growth and leakage in a nonhuman primate, laser-induced model of choroidal neovascularization. *Retina.* 24(1):132–138.
192. Check E (2005). A crucial test. *Nat. Med.* 11(3):243–244.

193. Barik S (2004). Development of gene-specific double-stranded RNA drugs. *Ann. Med.* 36(7):540–551.
194. Graham M J (1999). Oligonucleotide Therapeutics–IBC Sixth International Conference. *La Jolla, CA. IDrugs.* 2(7):653–655.
195. Orr R M (2001). Technology evaluation: Fomivirsen, Isis Pharmaceuticals Inc/CIBA vision. *Curr. Opin. Mol. Ther.* 3(3):288–294.
196. Zamecnik P C, Stephenson M L (1978). Inhibition of Rous sarcoma virus replication and cell transformation by a specific oligodeoxynucleotide. *Proc. Natl. Acad. Sci. USA.* 75(1):280–284.
197. Stephenson M L, Zamecnik P C (1978). Inhibition of Rous sarcoma viral RNA translation by a specific oligodeoxyribonucleotide. *Proc. Natl. Acad. Sci. USA.* 75(1): 285–288.
198. Wagner R W (1994). Gene inhibition using antisense oligodeoxynucleotides. *Nature.* 372(6504):333–335.
199. Sohail M (2001). Antisense technology: Inaccessibility and non-specificity. *Drug. Discov. Today.* 6(24):1260–1261.
200. Sohail M, Southern E M (2002). Oligonucleotide scanning arrays: Application to high-throughput screening for effective antisense reagents and the study of nucleic acid interactions. *Adv. Biochem. Engineer./Biotechno.* 77:43–56.
201. Achenbach T V, Brunner B, Heermeier K (2003). Oligonucleotide-based knockdown technologies: antisense versus RNA interference. *Chembiochem.* 4(10):928–935.
202. Beal J (2005). Silence is golden: Can RNA interference therapeutics deliver? *Drug Discov. Today.* 10(3):169–172.
203. Fishman M C, Porter J A (2005). Pharmaceuticals: a new grammar for drug discovery. *Nature.* 437(7058):491–493.
204. Howard K (2003). Unlocking the money-making potential of RNAi. *Nat. Biotechnol.* 21(12):1441–1446.

# DELIVERY SYSTEMS FOR PEPTIDES/ OLIGONUCLEOTIDES AND LIPOPHILIC NUCLEOSIDE ANALOGS

R.A. Schwendener[1] and Herbert Schott[2]

[1] *Institute of Molecular Cancer Research, University of Zurich, Zurich, Switzerland*
[2] *Institute of Organic Chemistry, University of Tübingen, Tübingen, Germany*

**Chapter Contents**

1 State of the Art of Nanosized Delivery Systems
2 Comparison of Viral vs. NonViral Gene Delivery Systems
3 Viral Systems
  3.1 Adenoviruses
  3.2 Poxviruses
  3.3 Herpex Simplex Virus
  3.4 Lentiviruses
  3.5 Adeno-Associated Virus
4 NonViral Systems
  4.1 Cationic Polymers, Lipoplexes
  4.2 NonLipidic Polycation Gene Delivery Systems
  4.3 Other Methods
  4.4 Virosomes
5 Liposomes and Cationic Liposome–DNA Complexes (Lipoplexes)
6 Peptide and Oligonucleotide Liposome Vaccine Formulations
7 Liposomes as Carriers of Lipophilic and Amphiphilic Nucleoside Analogs
8 Outlook and Future Directions
  References

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

## 1 STATE OF THE ART OF NANOSIZED DELIVERY SYSTEMS

The earliest developments of drug delivery systems (DDS) date back to the 1950s where the first microencapsulated drugs were introduced. In the 1960s polymer-based slow release systems appeared, and as the first nanosized DDS with spherical shape properties, the liposomes were recognized as models to study membranes and as carriers of both hydrophilic and lipophilic molecules. Liposomes, small unilamellar phospholipid bilayer vesicles, undoubtedly represent the most extensively and advanced drug delivery vehicles. After a long period of research and development efforts, liposome-formulated drugs have now entered the clinics to treat cancer and systemic fungal infections, mainly because they are biologically inert and biocompatible and do not cause unwanted toxic or antigenic reactions [1–3]. For the delivery of genetic material (DNA, ribozymes, DNAzymes, aptamers, (antisense-) oligonucleotides, small interfering RNAs), the liposomes, in particular lipid–DNA complexes termed lipoplexes, compete with viral gene transfection systems, as will be outlined in the following sections. Nanoparticles, nanospheres, polymersomes, nanogels, micelles, dendrimers, and virosomes are other main types of nanocarrier systems used for drug delivery [4–7]. As shown in Figure 1, these drug delivery systems vary in their compositions, shapes, sizes, drug loading capacity, as well as their pharmacokinetic and organ or tissue targeting properties [8].

DDS are developed for drugs with nonideal properties that include (1) poor solubility, where a conventional pharmaceutic formulation is difficult to prepare as poorly water soluble drugs may precipitate in aqueous media. (2) Tissue damage caused by inadvertent extravasation of drugs, e.g., tissue necrosis caused by cytotoxic drugs. (3) Loss of drug activity after administration, e.g., enzymatic and fast metabolic degradation. (4) Unfavorable pharmacokinetic properties and poor biodistribution. (5) Lack of selectivity for target organs or tissues. Systemic drug distribution may cause toxic side effects, and low concentrations in target tissues may cause suboptimal therapeutic effects.

The formulation of pharmacologically active drug molecules in DDS can improve or abolish these unfavorable properties. However, there are also drawbacks in DDS development, such as system complexity, unwanted biologic effects, stability, costs of development and scale-up, as well as intellectual property issues. In the limited format of this review, it is not possible to cover all methods and references in the field. Hence, we concentrate this review on DDS for the delivery of peptides, DNA, plasmids, oligodeoxynucleotides, siRNA, and lipophilic nucleoside derivatives. Vaccine delivery systems will also be mentioned, and examples will be provided to demonstrate the general development trends.

## 2 COMPARISON OF VIRAL VS. NONVIRAL GENE DELIVERY SYSTEMS

Viral gene delivery has become an important therapeutic strategy for the development of novel treatment approaches. A deeper understanding of vector biology and the molecular mechanisms of disease together with remarkable advances in molecular biology and vector technology have considerably advanced the field of human

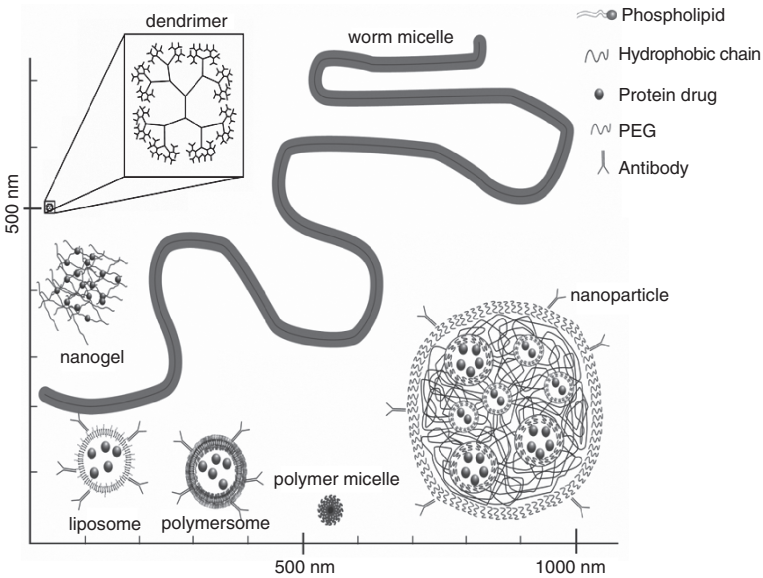


**Figure 1.** Nanocarriers for vascular drug delivery. Schematic representation of several main types of nanocarriers depicted in relative size scale with limited and oversimplified structural features. All nanocarriers can be surface-conjugated with targeting antibodies (or alternative affinity moieties) and PEG polymer providing stealth features. Dendrimers are the smallest of nanocarriers, in the maximum range of tens of nanometers. They possess multiple end groups suitable for a high extent of coupling targeting or active agents. Liposomes, composed of biologically derived phospholipids, are the most common form of nanocarriers with large aqueous loading potential; yet they are also the least stable of carriers. Polymersomes, one of three self-assembled polymeric nanocarriers, are the synthetic polymer analog of liposomes, possessing enhanced stability, dense PEG coating, and therefore prolonged circulation *in vivo*. Worm micelles are long, flexible cylindrical polymer micelles that possess one of the longest circulation times recorded *in vivo* with a yet unknown potential for therapeutic applications. Polymeric micelles are the smallest of the self-assembling polymer aggregate carriers and are the least stable self-assembler. Nanogels are composed of cross-linked polymers with drug entrapped into the ensuing matrix. Their circulation and potential therapeutic uses remain to be studied. Nanoparticles are solid polymer structures formed through processing rather than through self-assembly methods. They represent the largest of the carriers, have the greatest active protein loading capacity measured to date, and can protect encapsulated therapeutic enzymes from external proteolysis. (With permission from Ref. 8.)

gene therapy development. However, most viral gene delivery systems used to date have demonstrated limitations in practicality and safety, mainly due to low levels and short duration of recombinant transgene expression, induction of host immunogenicity to vector constituents, and suboptimal transgene expression to tissues or cells. A recent, additional cause for concern over using viral vectors is the phenomenon known as insertional mutagenesis, in which the chromosomal integration of viral gene material either interrupts the expression of a tumor suppressor gene or activates an oncogene, leading to the malignant transformation of cells.

Thus, safer, nonviral delivery approaches are needed and the latest advances indicate that efficient, long-term gene expression can be achieved by nonviral

means. In particular, integration of DNA can be targeted to specific genomic sites without harmful consequences, and it is possible to maintain transgenes as small episomal plasmids or artificial chromosomes. The application of these approaches to human gene therapy is progressively becoming a reality. Here, we briefly compare use, properties, advantages, and disadvantages of viral and nonviral gene delivery, the two main types of DDS that are used in gene therapy and vaccine approaches.

## 3 VIRAL SYSTEMS

The lack of efficient nontoxic gene delivery systems is still the major impediment to the successful application of gene therapy. Having evolved to deliver their genes to target cells, viruses are effective means of gene delivery and they can be manipulated to express therapeutic genes or to replicate specifically in certain cells.

The first viral vector systems were developed more than 25 years ago [9], and since then viral gene therapy strategies has been progressively developed [10]. A variety of virus vectors has been employed and modified to deliver genes to cells to provide either transient, such as adenovirus [11–13], poxviruses (vaccinia) [14], or herpes virus [15], or permanent, such as retroviruses (lentivirus) [16–18] and adeno-associated virus [19], transgene expression; each approach has its characteristic advantages and disadvantages.

### 3.1 Adenoviruses

Replication-defective adenoviruses are vectors of choice for delivering corrective genes into human cells. Major efforts are directed to design new generations of adenoviral vectors that feature reduced immunogenicity and improved targeting ability. Various targeting strategies have been attempted aiming at obtaining optimized and specific cellular transduction, including that of genetic manipulation of the viral capsid. Modification of the tropism-determining fiber protein and other capsid proteins has yielded vectors that are superior to the first-generation adenoviruses employed for gene therapy [11]. Adenoviral-based vectors are susceptible both to cytotoxic T-lymphocyte and humoral immune responses. In addition, leaky adenoviral genes also render transduced cells susceptible to host immune responses. These are the main reasons why adenoviral-based vectors are not suitable to correct genetic disorders, which require long-term expression of the transgene [20]. The production of adenoviral vectors for gene therapy applications still faces several challenges that limit the availability of high-quality material for clinical applications [21, 22].

### 3.2 Poxviruses

Poxviruses represent a heterogenous group of DNA viruses that have been used to express a multitude of foreign genes. Vaccinia virus is the prototypical recombinant poxvirus that can generate potent antibody and T-cell responses. Recombinant vaccinia viruses (rVVs) as nonreplicating viral vectors have been demonstrated for their great potential as vaccines, diminished cytopathic effects, high levels of protein expression, and strong immunogenicity, and they are relatively safe in

animals and in human patients. These properties have led to the use of rVVs as vaccines against HIV and cancer [14, 23].

### 3.3 Herpes Simplex Virus

Attenuated genetically engineered herpes simplex virus (HSV) vectors are potential vectors for several human therapy applications. These include delivery and expression of human genes to cells of the nervous systems, selective destruction of cancer cells, prophylaxis against infection with HSV or other infectious diseases, and targeted infection to specific tissues or organs [15].

### 3.4 Lentiviruses

Lentiviruses, members of the retroviral family, have the ability to infect cells at both mitotic and postmitotic stages of the cell cycle, thus opening the possibility to target nondividing cells and tissues. Human Immunodeficiency Virus (HIV)-based vectors have been used *in vitro* and *in vivo* in several situations; however, safety concerns still exist. Therefore, the development of vector systems based on primate as well as nonprimate lentiviruses is ongoing. Recent developments in the modification of the virus coat allow more targeted approaches and open new possibilities for the systemic delivery of therapeutic viruses [24]. However, the specific mechanisms used by different retroviruses to efficiently deliver their genes into cell nuclei remain largely unclear. Understanding these molecular mechanisms may reveal features to improve the efficacy of current retroviral vectors [25].

### 3.5 Adeno-Associated Virus

Vectors based on the adeno-associated virus (AAV) have attracted much attention as potent gene-delivery vehicles, mainly because of the persistence of this nonpathogenic virus in the host cell and its sustainable therapeutic gene expression. The principal historical limitation of this vector system, efficiency of recombinant AAV-mediated (rAAV) transduction, has recently observed a dramatic increase as the titer, purity, and production capacity of rAAV preparations have improved [26]. AAV vectors have been used in phase I clinical trials for the treatment of neurological disorders, such as Parkinson's and Canavan's diseases. Indeed, AAV-mediated gene transfer is a promising tool for the delivery of the therapeutic gene into the central and peripheral nervous systems. AAV-mediated gene transfer was also applied in phase I and II clinical trials for the treatment of cystic fibrosis and in phase I trials for the treatment of hemophilia B. In the context of cancer, the ability of attenuated viruses to replicate specifically in tumor cells has already yielded some impressive results in clinical trials, allowing the design of new therapeutic approaches, particularly when combined with other approved anticancer therapies. Despite the remark-able progress that has been reported, the design of further optimized vectors is still required. As it stands, AAV-mediated gene transfer has a limited capacity in accommodating foreign genes. In addition, some preclinical studies have shown that AAV-derived vectors can cause tumors in animals due to mutagenic random vector integration into the genome [27]. To circumvent this problem, a novel approach to AAV-mediated gene therapy based on gene targeting through homologous recombination has been introduced that allows efficient, high-fidelity, nonmutagenic gene repair in a host cell [28].

## 4 NONVIRAL SYSTEMS

Currently, the most severe limitations of the nonviral gene therapy systems are low transfection efficiency of gene material into the target cells, physico-chemical instability, and cytotoxicity. The major obstacles encountered in the transfer of foreign genetic material into the body include interactions with blood components and vascular endothelial cells and uptake by the mononuclear phagocyte system (MPS). The degradation of DNA by serum nucleases is another major obstruction for functional delivery to the target. In addition to targeting a specific cell type, an ideal nonviral vector (liposomes, lipoplexes, virosomes), once taken up by a target cell, must manifest an efficient endosomal escape, provide sufficient protection of DNA in the cytosol and help facilitate an easy passage of cytosolic DNA to the nucleus (Figure 2).

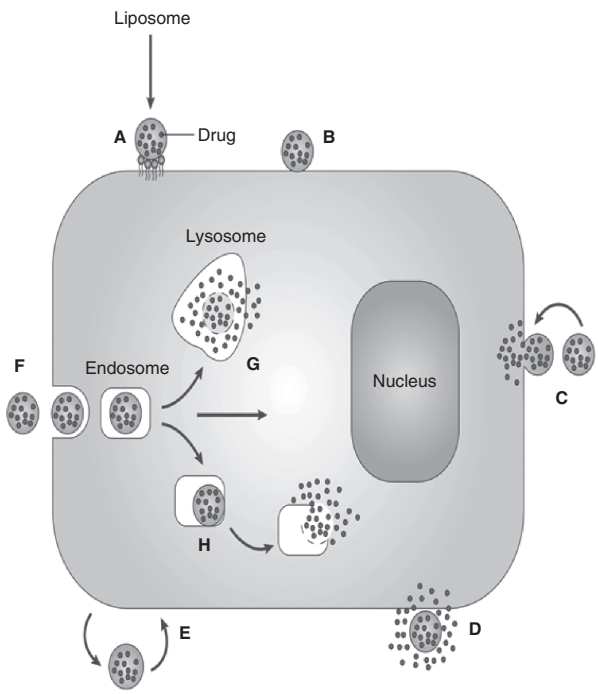


**Figure 2.** Liposome–cell interactions. Drug-loaded liposomes can specifically (**A**) or nonspecifically (**B**) adsorb onto the cell surface. Liposomes can also fuse with the cell membrane (**C**) and release their contents into the cell cytoplasm, or they can be destabilized by certain cell membrane components when adsorbed on the surface (**D**) so that the released drug can enter the cell via micro-pinocytosis. Liposome can undergo the direct or transfer-protein mediated exchange of lipid components with the cell membrane (**E**) or be taken up by specific or nonspecific endocytosis (**F**). In the case of endocytosis, a liposome can be delivered by the endosome into the lysosome (**G**), or en route to the lysosome, the liposome can provoke endosome destabilization (**H**), which results in drug liberation into the cytoplasm. (With permission from Ref. 29.)

### 4.1 Cationic Polymers, Lipoplexes

Cationic polymers have a great potential for DNA complexation and have shown to be useful as nonviral vectors for gene therapy applications. For the past 15–20 years liposomes composed of cationic lipids, termed lipoplexes, have routinely been used for the delivery of nucleic acids such as plasmids, oligodeoxynucleotides, and siRNA to cells in culture and *in vivo*. Many of these reagents are commercially available or can be formulated in the laboratory [30–32]. Most cationic lipid–DNA complexes form a multilayered structure with DNA sandwiched between the cationic lipids. Much more rarely, an inverted hexagonal structure with single DNA strands encapsulated in lipid tubules is observed [33]. Among other advantages, lipoplexes have the ability to transfer very large genes into cells. However, as the understanding of their mechanisms of action is still incomplete, their transfection efficiencies remain low compared with those of viruses. Particularly in cultured cells, toxicity remains a significant problem. In addition, these complexes are immunostimulatory, a fact that may either be harmful or beneficial. The development of cationic lipids that are safe to use, especially for *in vivo* applications, and possess enhanced transfection capabilities is an ongoing process. The lipoplexes are described in detail in Section 5.

### 4.2 NonLipidic Polycation Gene Delivery Systems

Of the many nonlipidic polycation gene delivery systems developed in the past decades, poly(L-lysine) (PLL) was the first polycation used for nonviral gene delivery [34]. Among a vast number of other positively charged polymers, polyethylenimine (PEI) has been widely used for nonviral transfection *in vitro* and *in vivo* and has an advantage over other polycations in that it combines strong DNA condensation capacity with an intrinsic endosomolytic activity [35–38].

Other synthetic and natural polycations developed as nonviral vectors are polyamidoamine dendrimers, the synthetic cationic polymer poly(2-dimethylamino)ethyl methacrylate (PDMAEMA) [39], and chitosan. Polyamidoamine (PAMAM) dendrimers represent a novel class of polycationic synthetic polymers that can be used for gene transfer [40, 41]. The three-dimensional spherical structure of dendrimers offers synthesis control of the molecule in terms of degree and generation of branching. The control of branching of the dendrimers during synthesis allows the production of polymer particles with a very high degree of monodispersity, which is a significant advantage over other polymers such as polylysine that generate highly polydisperse particles (see Figure 1). Low polydispersity can lead to reproducible gene delivery and a clinically reliable formulation. The cationic amino acid residues in the polymeric structure of PAMAM dendrimers can help in DNA condensation and endosome release. Dendrimers that protect oligonucleotides from serum nucleases have been used to enhance oligonucleotide delivery [42].

Chitosan, a natural-based polymer obtained by alkaline deacetylation of chitin, is nontoxic, biocompatible, and biodegradable. These properties make chitosan a promising candidate for conventional and novel drug delivery systems. Because of the high affinity of chitosan for cell membranes, it has been used as a coating agent for liposome formulations [43–45].

### 4.3 Other Methods

Various other methods, such as DNA electrotransfer, electroporation-based gene transfer, calcium phosphate nanoparticles, peptide nucleic acids, and cell penetrating peptides, round off the currently used methods in the field of nonviral gene delivery techniques.

DNA electrotransfer, the use of electric pulses to transfect various types of cells, is well known and regularly used *in vitro* for bacteria and eukaryotic cell transformation. Electric pulses can also be delivered *in vivo* either through the skin or with electrodes in direct contact with the target tissues. After injection of naked DNA in a tissue, appropriate local electric pulses can result in a very high expression of the transferred genes [46].

Electroporation has been applied in preclinical autoimmune and/or inflammatory diseases to deliver either cytokines and anti-inflammatory agents or immunoregulatory molecules. The method is also effective for the intratumoral delivery of therapeutic vectors, and it strongly boosts DNA vaccination against infectious agents or tumor antigens. Electroporation gene therapy has become a widely used method for nonviral gene delivery, including applications for intramuscular and intratumoral electro-gene transfer and for the transfection of dendritic and stem cells [47, 48]. The current challenges faced by both *in vitro* and *in vivo* applications comprise the enhancement of transfection efficiency, extention of the duration of gene expression, and increase of the survival rate for *in vitro* cell transfections.

Virus- or liposome-like-sized calcium phosphate nanoparticles of 20–200 nm mean diameter have been found to overcome many of the known limitations in delivering genes to the nucleus of specific cells. It has been demonstrated that calcium ions play an important role in endosomal escape, cytosolic stability, and enhanced nuclear uptake of DNA through nuclear pore complexes. The role of exogenous calcium ions to overcome the major obstacles encountered in the practical accomplishment of gene delivery suggests that calcium phosphate nanoparticles can be designated as a new generation of nonviral vectors [49].

Peptide nucleic acid (PNA) is a powerful new biomolecular tool with a wide range of important applications. PNA mimics the behavior of DNA and binds complementary nucleic acid strands and RNA sequences with high affinity and selectivity. The unique chemical, physical, and biological properties of PNA are exploited to produce powerful biomolecular tools, antisense and antigene agents, molecular probes, and biosensors [50–53].

Cell penetrating peptides (CPPs) have proven to be an efficient intracellular delivery system overcoming the lipophilic barrier of cell membranes [54]. CPPs can deliver a wide range of large cargo molecules such as proteins, peptides, oligonucleotides, and even small particles as liposomes to a variety of cell types and to different cellular compartments. The CPPs are basic, lysine-, or arginine- rich amphipathic peptides. The peptides originate from different sources, either as naturally occurring peptide sequences, virally derived (TAT, VP22), from transcription factors (pAntp), as chimeric (transportan) or as synthetic peptides (polyarginines) and others [55].

As depicted in Figure 3, CPPs can either form complexes with peptides, proteins, plasmids, oligonucleotides, or siRNA molecules or they can be covalently linked to these cargo molecules [54]. Larger structures such as liposomes have also been

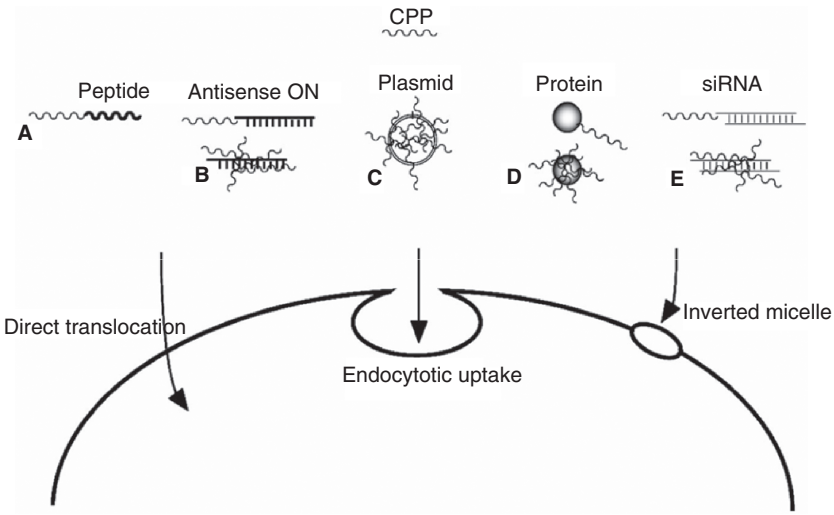


**Figure 3.** Cell penetrating peptides, CPPs. Suggested uptake mechanisms for CPPs and examples of delivered cargoes. (A) CPP and peptide in single amino acid chain. (B) Oligodeoxynucleotides either in complex or covalently linked. (C) Plasmid in complex by electrostatic interaction. (D) Protein either as fusion protein or in complex with CPP. (E) siRNA, covalently linked or as complex. (With permission from Ref. 54.)

decorated with the TAT [56] or pAntp CPPs [57], demonstrating higher cell uptake rates *in vitro*.

Unfortunately, their usefulness as drug delivery systems is hampered by their ability to penetrate virtually any cell type both *in vitro* and *in vivo*. This feature makes CPP applications as target specific drug delivery systems complicated and their application as therapeutic drug carrier systems seems unlikely, unless their target cell specificity can be significantly improved.

### 4.4 Virosomes

Virosomes were developed from liposomes by combining liposomes with fusogenic viral envelope proteins. Almeida et al. [58] were the first to report on the generation of lipid vesicles containing viral spike proteins derived from influenza virus. Using preformed liposomes and hemagglutinin (HA) and neuraminidase (NA), purified from influenza virus, they succeeded to generate membrane vesicles with spike proteins protruding from the vesicle surface. Visualization of these vesicles by electron microscopy revealed that they very much resembled native influenza virus. Consequently, they were named virosomes. Reconstituted viral envelopes (virosomes, artificial viral envelopes) appear to be ideally suited as vaccine formulations for the delivery of protein antigens to the cytosol of antigen presenting cells (APCs), and thus for the introduction of antigenic peptides into the MHC class I presentation pathway. Cytotoxic T-cell activity can be induced by immunization of mice with an antigenic peptide or an entire protein encapsulated in virosomes. As shown schematically in Figure 4, the action of influenza virosomes is likely to involve both delivery of the enclosed antigen to the cytosol of antigen presenting

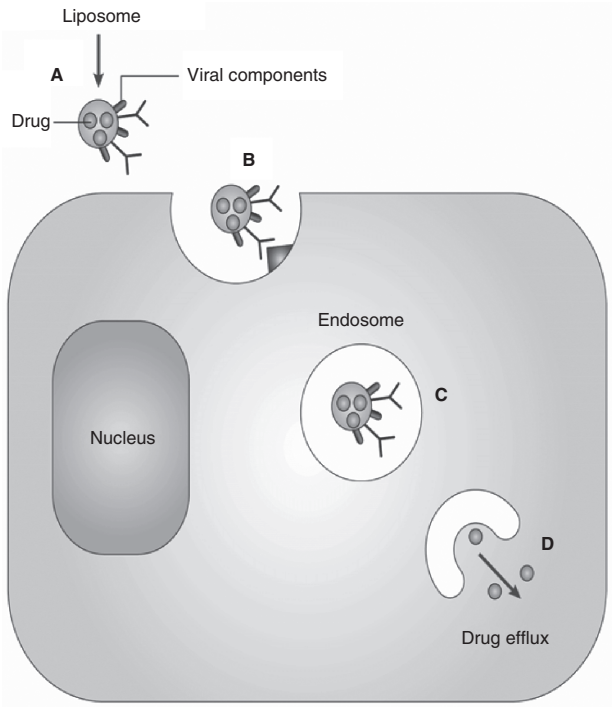


**Figure 4.** Virosome–cell interactions. Liposome modified with specific viral components (**A**) and loaded with a drug can specifically interact with cells (**B**), provoke endocytosis, and via the interaction of viral components with the inner membrane of the endosome (**C**), allow for drug efflux into the cell cytoplasm (**D**). (With permission from Ref. 29.)

cells and the powerful helper activity of the virosomal hemagglutinin. Although the immune responses elicited by DNA-virosomes are moderate, they are promising and warrant further research to ultimately develop effective DNA-based virosomal vaccines [59–61].

## 5 LIPOSOMES AND CATIONIC LIPOSOME-DNA COMPLEXES (LIPOPLEXES)

Liposomes have become known as one of the most versatile tools for the delivery of DNA, DNA-related, and many other therapeutic molecules [6, 7, 29]. Liposomes are spherical vesicles that consist of an aqueous compartment enclosed in a phospholipid bilayer. If multiple bilayers of lipids are formed around the primary core, the structures that are generated are known as multilamellar vesicles (MLVs). MLVs are formed spontaneously by reconstitution of lipid films in aqueous media. Small unilamellar vesicles (SUVs) of specific size (100–500 nm) are produced by high-pressure extrusion of MLVs through polycarbonate membranes. SUVs (25–90 nm) are also obtained by sonication of MLVs or larger SUVs, by detergent dialysis [62] and by many other, less important methods. Both hydrophilic and

hydrophobic drugs can be entrapped in the liposomes, and the choice of the lipid composition as well as the surface modification of the liposomes provides them with a high versatility such as long circulation half-life and sustained and targeted delivery (Figure 5) [29, 62].

Liposomes can be used as DNA drug delivery systems either by entrapping the DNA-based therapeutics inside the aqueous core or by complexing them to positively charged lipids (lipoplexes, see below). Liposomes offer significant advantages over viral delivery systems; for example, liposomes are generally nonimmunogenic because of the absence of proteinaceous components. As the phospholipid composition in the liposome bilayers can be varied, liposomal delivery systems are of high versatility and customized formulations can be easily engineered to obtain desired sizes, surface charge, composition, and morphology. Liposome encapsulated DNA molecules are protected from nuclease activity and thus improve their biological stability.

Long circulating ("stealth") liposomes are sterically stabilized liposomal formulations that include poly(ethylene glycol) (PEG)-conjugated lipids or other hydrophilic coating molecules. PEGylation prevents the opsonization and recognition of the liposomal vesicles by the MPS. PEGlyation has also been used in conjunction with other polymeric delivery systems such as poly(L-lysine) to achieve longer circulation half-lives.

Immunoliposomes are complex drug or gene delivery systems that can be used for cell targeting by the incorporation of functionalized antibodies attached to lipid bilayers. The "state-of-the-art" immunoliposomes are long circulating PEG-liposomes to which receptor specific molecules are attached, preferably at the distal tips of the PEG chains [62–65]. Immunoliposomes target specific receptors and facilitate receptor-mediated endocytosis for cell uptake (see Fig. 2). Immunoliposomes decorated with single-chain antibody fragments against the ED-B isoform of fibronectin were successfully used in targeted delivery of cytotoxic drugs into tumors *in vivo* [66]. Tissue-specific gene delivery using immunoliposomes has also been achieved by antitransferrin receptor immunoliposomes [67].

To release encapsulated material into the cytoplasm, pH-sensitive liposomes can be generated by the inclusion of dioleyl-phosphatidylethanolamine (DOPE) into liposomes composed of acidic lipids such as cholesterylhemisuccinate or oleic acid. At the neutral cellular pH 7, these lipids have the typical bilayer structure; however, upon endosomal compartmentalization, they undergo protonation and collapse into nonbilayer structures, thereby leading to the disruption and destabilization of the endosomal bilayer, which in turn helps in the rapid release of encapsulated molecules into the cytoplasm [7].

There is an important difference between liposome vesicles and cationic lipid–DNA complexes, the lipoplexes. Lipoplexes are cationic lipid–DNA complexes that are formed spontaneously in aqueous media upon mixing DNA and preformed liposome vesicles composed of cationic and neutral lipids [68]. From a physical point of view, lipoplexes are ordered, self-assembled, composite aggregates whose spatial geometry and phase behavior are controlled by the electrostatic interactions between the positively charged lipids and the negatively charged DNA molecules. Numerous studies have demonstrated the use of cationic liposome formulations for the delivery of different plasmid constructs in a wide range of cells, both *in vivo* and *in vitro*. Despite their cytotoxicity, their nonimmunogenic nature and the

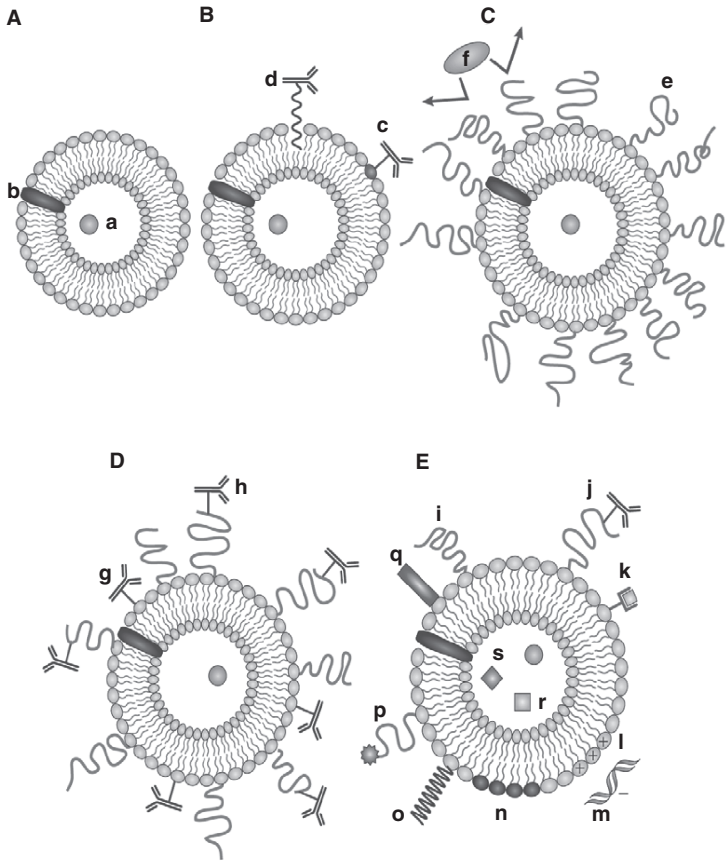


**Figure 5.** Evolution of liposomes. (**A**) Early traditional "plain" liposomes with water-soluble drug (**a**) entrapped into the aqueous liposome interior, and lipophilic drug (**b**) incorporated into the liposomal membrane. (**B**) Antibody-targeted immunoliposome with antibody covalently coupled (**c**) to the reactive phospholipids in the membrane, or hydrophobically anchored (**d**) into the liposomal membrane after preliminary modification with a hydrophobic moiety. (**C**) Long circulating liposome grafted with a protective polymer (**e**) such as PEG, which shields the liposome surface from the interaction with opsonizing proteins (**f**). (**D**) Long circulating immunoliposome simultaneously bearing both protective polymer and antibody, which can be attached to the liposome surface (**g**) or, preferably, to the distal end of the grafted polymeric chain (**h**). (**E**) New-generation liposome, the surface of which can be modified (separately or simultaneously) by different ways. Among these modifications are the attachment of protective polymer (**i**) or protective polymer and targeting ligand, such as antibody (**j**); the attachment/incorporation of a diagnostic label (**k**); the incorporation of positively charged lipids (**l**) allowing for the complexation with DNA yielding lipoplex structures (**m**); the incorporation of stimuli-sensitive lipids (**n**); the attachment of a stimuli-sensitive polymer (**o**); the attachment of a cell-penetrating peptide (**p**); and the incorporation of viral components (**q**). In addition to a drug, liposomes can be loaded with magnetic particles (**r**) for magnetic targeting and/or with colloidal gold, silver particles, or fluorescent molecules (**s**) for microscopic analysis. (Adapted with permission from Ref. 29.)

simplicity of production of these systems make them attractive tools for gene transfer. Currently, many gene therapy trials in progress employ nonviral liposomal vectors for transgene delivery.

Lipoplex formulations generally consist of mixtures of cationic and neutral (zwitterionic) lipids. Commonly used cationic lipids are 1,2-dioleoyl-3-trimethylammonium propane (DOTAP), N-[1-(2,3-dioleyloxy)propyl]-N,N,N-trimethylammonium chloride (DOTMA), 2,3-dioleoyloxy-N-[2-(sperminecarboxamido)ethyl]-N,N-dimethyl-1-propanaminium (DOSPA), dioctadecyl amido glycil spermine (DOGS), and 3,[N-(N1,N-dimethylethylenediamine)-carbamoyl] cholesterol (DC-chol) [69–71]. Commonly used neutral molecules, also known as helper or colipids, are DOPE and cholesterol. The cationic lipids in the liposomal formulation serve as a DNA complexation and DNA condensation agent during the formation of the lipoplex. The positive charge also helps in cellular association. The colipids facilitate membrane perturbation and fusion. Many proprietary reagents of cationic lipids such as Lipofectamine (Invitrogen, Carlsbad, CA), Effectene (Qiagen, Valencia, CA), and Tranfectam (Promega, Madison, WI) are commercially available. However, most of these transfection reagents can only be used for *in vitro* gene transfection applications. Despite the appreciable success of cationic lipids in gene transfer, toxicity is a main issue for both *in vitro* and *in vivo* applications. Inflammatory toxicity represents a typical toxicity associated with systemic administration of lipoplexes. Results obtained from *in vivo* studies indicate that lipoplex gene delivery systems mediate uptake of plasmid DNA by the liver, mainly by Kupffer cells, in which a large amount of cytokines is produced [72, 73]. Upon administration via the airways, cytokine-mediated pulmonary toxicity and TNF-α induction by cationic lipids in lung tissue have been reported [74]. Reduction in toxicity was observed by a modification of DOPE polyplexes with cetylated PEI, resulting in remarkable transgene efficiency with low cytotoxicity [72]. In another study, a sterically stabilized immunolipoplex composed of a p53 DNA–lipid complex to which PEG molecules and an antitransferrin receptor single-chain antibody fragment were attached resulted in improved delivery of the complex to tumor cells *in vivo* [75]. The negative factors of lipoplex-mediated gene transfer are low transfection efficiencies, which have been attributed to the heterogeneity and instability of the lipoplex formulations. Lipoplex size heterogeneity also adversely affects their quality control, scale-up, and long-term shelf stability, which are important issues for their pharmaceutical development. Compared with viral vectors, the transfection efficiencies of cationic liposomal vectors are significantly lower. Another drawback in the use of cationic lipids is the rapid inactivation of their cargo in the presence of serum [76].

As an alternative to cationic lipids, the potential of anionic lipids for DNA delivery has been investigated. However, in recent years, only a few studies using anionic liposomal DNA delivery vectors have been reported [77, 78]. These vectors have limited applications, mainly because of inefficient entrapment of DNA molecules within anionic liposomes and lack of toxicity data. Lacking progress of these systems may be attributed to the poor association between DNA molecules and anionic lipids, which is caused by electrostatic repulsion between these negatively charged species.

## 6 PEPTIDE AND OLIGONUCLEOTIDE LIPOSOME VACCINE FORMULATIONS

With the increasing availability of defined antigens such as highly purified proteins or synthetic peptides, more effective and safer vaccines are developed. As many antigens are often poorly immunogenic when administered alone, the development of suitable adjuvants, possessing the ability to potentiate the immunogenicity of a given antigen, preferably with little or no side effects, is required. Based on their principal mechanisms of action, adjuvants can be divided into two groups: (1) vaccine delivery systems and (2) immunostimulatory adjuvants. As described in this chapter, vaccine delivery systems are generally particulate delivery systems of size dimensions comparable with pathogens as bacteria and viruses (e.g., liposomes, microemulsions, immunostimulatory complexes, and other nano- or microparticle systems in the size range of 20–500 nm). Their function is mainly to target antigens to antigen-presenting cells (APCs; dendritic cells, macrophages) and to act as adjuvants. The rationale for the development of an optimal vaccine is to ensure that both antigen and adjuvant are delivered into the same population of APCs, thereby reducing systemic distribution and minimizing the potential to induce adverse reactions. Small unilamellar liposomes have an important potential as delivery systems for the coadministration of antigens and of immunostimulatory adjuvants, including synthetic oligodeoxynucleotides containing immunostimulatory deoxycytidylyl-deoxyguanosine dinucleotides (CpGs) or DNA encoding antigens [79–81]. Additionally, the efficiency of the liposomes can be improved by targeting them more effectively and specifically to the APCs by exploiting various scavengers and other receptors as their targets or by enhancing their cell uptake by modification with cell penetrating peptides as recently shown by us [57].

Using the lymphocytic choriomeningitis virus (LCMV) model system, we successfully prepared efficient peptide vaccines with liposomes as the carrier [82]. Liposome-encapsulated antigenic peptides were highly immunogenic when administered intradermally, and they elicited protective antiviral immunity. An optimized formulation contained immunostimulatory oligonucleotides leading to activation of dendritic cells and antitumor immunity in LCMV peptide transfected EL4 thymomas. In a follow-up study we could confirm the efficacy of such peptide–liposome vaccines in a hepatitis C virus model in HLA-A2.1 transgenic mice [83]. These findings clearly indicate that liposomal antigen delivery *in vivo* is a promising approach to induce efficient antiviral and antitumor immune responses with relevance for human applications.

A cautionary note to the potential dangers of all viral gene products, transgenes, viral proteins and peptides, and CpG DNA sequences in siRNA or plasmids formulated in DDS has to be given. Immune responses induced by these molecules may lead to problems such as transient gene expression, nonefficient readministration of the same vectors and to severe side effects in clinical trials [84]. Due to their particulate nature, the DDS are recognized as foreign in an organism that reacts with an immune response. However, the immunomodulating activities of the DDS depend largely on their composition, size, and homogeneity. Synthetic polymers can exhibit significant immunomodulating activity, whereas liposomes prepared with natural phospholipids and cholesterol are known to be less immunogenic [85].

## 7 LIPOSOMES AS CARRIERS OF LIPOPHILIC AND AMPHIPHILIC NUCLEOSIDE ANALOGS

Most applications of liposomes as therapeutic drug carrier systems are based on the encapsulation of water-soluble molecules within the trapped aqueous volume of the liposomes. Long circulating PEG-modified liposomes with cytotoxic antitumor drugs doxorubicine, paclitaxel, vincristine, and methotrexate are examples of clinically applied chemotherapeutic liposome formulations [86–88]. In contrast to the extensive exploitation of the trapped aqueous volume of the liposomes that serves as a nanocontainer for the water-soluble molecules, the phospholipid bilayer has not been given the same attention for its use as the carrier matrix for lipophilic drugs. Hence, the development of liposomal drug formulations with lipophilic drugs is less popular. This difference may have several reasons, the main probably consisting in the chemistry required to transform water-soluble molecules into lipophilic compounds that allow incorporation into the lipid bilayer core. The most favorable chemical modifications to obtain a molecule that intercalates in a stable fashion into the lipophilic moiety of a lipid bilayer consist in the attachment of long-chain fatty acyl or alkyl residues, for example saturated or unsaturated fatty acids, preferably palmitic or stearic acid and alkylamines, preferably hexadecyl- or octadecylamine to a suitable functional group of the hydrophilic part of the molecule. Some recent examples of lipophilic modifications of antitumor drugs and their formulation in liposomes are gemcitabine, 5-iodo-2′-deoxyuridine, methotrexate, paclitaxel, cytosine arabinoside, and a lipophilic topoisomerase inhibitor, DB67 [89–96].

Drugs that are highly lipophilic by their own nature, e.g., taxanes, epothilones, and cyclosporins, can only be used therapeutically by the addition of possibly toxic solubilizing agents (e.g., Cremophor EL) in complex pharmaceutical formulations [97–99]. One of several feasible means of obtaining nontoxic parenterally applicable formulations of such drugs is their incorporation into the bilayer matrix of phospholipid liposomes.

Nucleoside analogs are a major class of chemotherapeutic agents for the treatment of cancer and viral diseases. Natural endogenous nucleosides must be phosphorylated to corresponding 5′-triphosphates to be incorporated into the DNA or RNA synthesized within the cell. The first phosphorylation step, leading to the formation of nucleoside-5′-monophosphate, is commonly performed by a nucleoside kinase encoded by either the host cell or the virus infecting the host cell. Hence, cellular and virally encoded kinases play a vital role in the metabolism and replication of cells and viruses. Nucleoside analogs used for chemotherapy of cancer and viral infections are in essence prodrugs because they must be phosphorylated in the cytoplasm like the natural nucleosides to triphosphates before they can exert their activities. Thus, administration of phosphorylated nucleoside analogs would circumvent the enzymatic phosphorylation step, which transforms inactive nucleosides into active drugs. Phosphate groups possess an anionic charge at nearly all physiological pH values, making them very polar. This high polarity can be responsible for many deficiencies in terms of efficient drug delivery. Nucleotides are too hydrophilic to penetrate the lipid-rich cell membrane. In addition, blood and cell-surface phosphohydrolases rapidly metabolize the nucleotides to the corresponding nucleosides. Due to their polarity, nucleotides often exhibit a low

volume of distribution and are therefore efficiently cleared by renal elimination. In an attempt to overcome these shortcomings, various prodrugs or "pronucleotide" approaches have been devised and investigated for the *in vivo* delivery of pharmacologically active nucleotides. The general aim of these approaches has been to promote passive diffusion through cell membranes and to increase the bioavailability of phosphorylated nucleoside analogs [100, 101]. A simple solution for the delivery of nucleoside monophosphate analogs into cells consists in the chemical neutralization of the ionizable phosphate group via chemical derivatization for example by esterification, resulting in nucleotide phosphodi- and phosphotriester derivatives or by the synthesis of nucleotide phophorodiamidates, cyclic phosphoramidates, and phosphoramidate mono- and diesters. These compounds have a neutral charge and should be capable of entering cells via passive diffusion. To retain the advantage of nucleoside phosphotriesters with regard to their improved cellular uptake, a variety of biolabile moieties have been evaluated as potential protecting groups for nucleoside phosphotriesters.

We chose the approach of the chemical transformation of water-soluble nucleosides of known cytotoxic and antiviral properties into lipophilic drugs or prodrugs. Thus, we reversed the paradigm of transforming lipophilic molecules into hydrophilic derivatives [102]. The first cytotoxic nucleoside that we transformed into lipophilic derivatives was 1-β-D-arabino-furanosyl cytosine (ara-C) because its major clinical disadvantages are a very short plasma half-life and rapid degradation by deamination to the inactive metabolite 1-β-D-arabino-furanosyluracil (ara-U), a shortcoming that also impedes the oral application of ara-C. To reduce these limitations, a large number of 5′- and $N^4$-substituted ara-C derivatives have been synthesized and characterized in the past (reviewed in Ref. 103). We synthesized a new class of $N^4$-alkyl-ara-C derivatives with alkyl chain lengths ranging between 6 and 22 C-atoms, demonstrating a typical structure-activity correlation between the length of the alkyl side chains and their antitumor activity profile [104]. The most effective derivative, $N^4$-octadecyl-ara-C (NOAC), which is highly lipophilic and extremely resistant toward deamination exerted excellent antitumor activity after oral and parenteral therapy in several mouse tumor models and showed to have distinct pharmacological properties compared with ara-C [105].

Consequently, we further modified NOAC by the synthesis of a new generation of lipophilic/amphiphilic heterodinucleoside phosphate derivatives, termed "duplex drugs" that combine the clinically used antitumor drugs ara-C and 5-fluorodeoxyuridine (5-FdU) with NOAC yielding the heterodinucleoside phosphates arabinocytidylyl-$N^4$-octadecyl-1-β-D-arabino-furanosylcytosine (ara-C-NOAC) and 2′-deoxy-5-fluorouridylyl-$N^4$-octadecyl-1-β-D-arabinofuranosy-lcytosine (5-FdU-NOAC) (Figure 6) [106, 107]. Ethynylcytidine (1-(3-C-ethynyl-β-D-ribopentafuranosyl)-cytosine, ETC) is a novel nucleoside that was found to be highly cytotoxic [108–110]. Thus, by combination of ETC with NOAC, we obtained the lipophilic duplex drug ETC-NOAC (3′-C-ethynylcytidylyl-(5′ → 5′)-$N^4$-octadecyl-1-β-D-arabinofuranosylcytosine).

Due to the combination of the effects of both active molecules that can be released into the cytoplasm as monomers or as the corresponding monophosphates (MPs), the cytotoxic activity of the duplex drugs is expected to be more pronounced as compared with the monomeric drugs. Furthermore, it can be anticipated that the monophosphorylated nucleosides ara-CMP, 5-FdU-MP, and ETC-

MP, respectively, are directly released into the cell after enzymatic cleavage of the duplex drugs. Thus, monophosphorylated molecules would not have to pass the first phosphorylation step, which is known to be rate limiting [100, 101].

The lipophilic side chains allow a stable incorporation of these duplex drugs into liposomes, allowing the exploitation of the advantages liposome formulations are offering. Due to their high polarity, the nonderivatized heterodinucleoside phosphodiesters are less suited for liposomal formulations. In comparison with the nonpolar heterodinucleoside phosphotriesters that have good properties to be taken up by cells, but whose capacity to be cleaved enzymatically is limited, the cleavage of the natural phosphodiester bond of the duplex drugs is not constrained. A delayed intracellular release of nucleoside and nucleotide analogs from the duplex molecules provides a depot effect that may be therapeutically of advantage.

A structure-related disadvantage of the duplex drugs is that upon enzymatic cleavage of the phosphodiester bonds, a 1-to-1 ratio of nucleoside to nucleotide is obtained (see Figure 6). Thus, the desired 5′-phosphorylated nucleotide is only formed at maximally 50%. Additionally, one of the two nucleosides has to be transformed into a lipophilic derivative without loss of antitumor or antiviral activity. On the other hand, it was shown that the lipophilic derivatization of nucleosides can result in enhanced activity and modulation of cell specificity [106].

The disadvantage of obtaining only a 50% yield of 5′-phosphorylated nucleoside analogs, which results from the enzymatic cleavage of the duplex drugs, can be avoided by the synthesis of glycerol–lipid–heteronucleotides, the so-called "multiplex drugs." Figure 7 shows an example of such a compound where the two cytotoxic nucleosides ara-C and 5-FdU are linked at their 5′-hydroxyl groups via a phosphodiester to the 1,3-hydroxy groups of glycerol. To introduce amphiphilic/

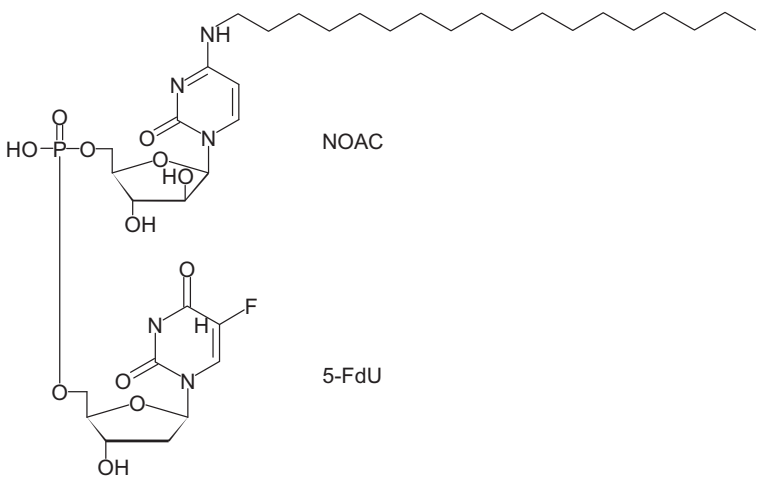


**Figure 6.** Duplex drugs. Structure of the anticancer amphiphilic heterodinucleosidyl phosphodiester (Duplex-drug) combining the hydrophilic 5-fluoro-deoxyuridine (5-FdU) with the lipophilic $N^4$-octadecylarabino-furanosylcytosine (NOAC). Upon enzymatic hydrolysis of the duplex drugs, different anticancer nucleotides and nucleosides are formed with additive or synergistic activities.

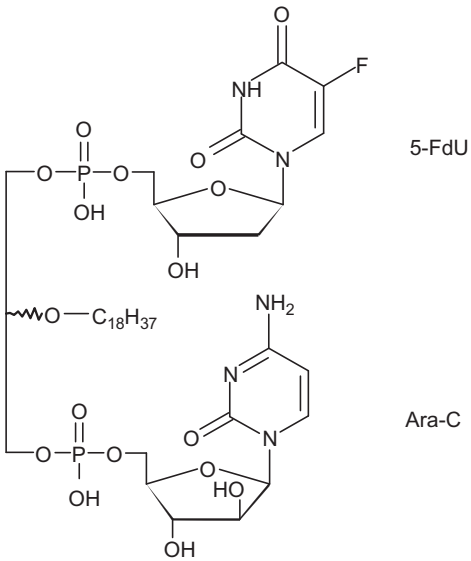


**Figure 7.** Multiplex drugs. Example of an amphiphilic glyceryllipid-heterodinucleotide (Multiplex drug). The two hydrophilic nucleoside-5′-monophosphates p5-FdU and paraC are esterified with the terminal hydroxyl residues of the lipophilic 2-octadecylglycerol. By the enzymatic cleavage, several different anticancer metabolites with additive or synergistic activities are obtained.

lipophilic properties, an octadecyl chain is coupled to the 2-hydroxy position of glycerol.

After metabolic degradation by phosphodiesterase, both nucleosides are released as 5′-nucleotides. In case of different hydrolysis kinetics, other active intermediate products may be formed. The structure of the glycerol–lipid–heteronucleotides provides a programmed release of differently active drugs that can penetrate a cell membrane when the cleavage takes place outside of a cell. These compounds may also be formulated in DDS, such as liposomes or micellar systems.

*In vitro* tests of the glycerol–lipid–heteronucleotides shown in Figure 7 revealed that these multiplex drugs inhibited colony formation of 5-FU sensitive and resistant human colon tumor cell lines and induced dose-dependent apoptosis in colon tumor cells as well as in mouse leukemia cells. No significant difference in the cytotoxicity could be observed between 5-FU sensitive and resistant cells, indicating that the multiplex drugs might be useful for the treatment of 5-FU resistant tumors [111, 112]. The effectiveness of the postulated mechanisms and advantages of the multiplex drugs will have to be elucidated in *in vivo* experiments.

We conclude that the chemical modification of water-soluble molecules by attachment of long alkyl chains and their stable incorporation into the bilayer membranes of small unilamellar liposomes represent a very promising example of taking advantage of the high loading capacity lipid bilayers offer for lipophilic drugs. The combination of chemical modifications of water-soluble drugs with their pharmaceutical formulation in liposomes is a valuable method for the development of novel pharmaceutical preparations not only for the treatment of tumors or infectious diseases, but also for many other disorders.

## 8 OUTLOOK AND FUTURE DIRECTIONS

The development of DDS is a challenging venture that combines research efforts of experts in various areas, including bioengineering, nanotechnology, biomaterials, pharmaceutics, biochemistry, and cell and molecular biology. Specific characteristics of pathological processes and cell or tissue types that are the subject of therapeutic interventions govern the path from target selection to the development of specific DDS formulations. The identification of novel cellular targets, for example, easily accessible vascular endothelial cells, in contrast to tumor cells or other less reachable tissues, will lead to optimized pharmaceutical drug delivery formulations and preparation technologies. Refinement of DDS to overcome unwanted properties such as toxicity, unspecific tissue distribution, and uncontrolled release of entrapped active molecules will be the major challenges in the field. Future DDS will mostly be based on DNA therapeutics such as DNA, ribozymes, DNAzymes, aptamers, (antisense-) oligonucleotides, small interfering RNAs, and their next-generation analogs and derivatives. The use of high-throughput systems for lead identification will yield many new therapeutic targets. Validation and optimization of these disease targets will provide a tremendous stimulus in developing newer potent molecules and their corresponding delivery systems. Further advances in the study of gene function and identification of single-nucleotide polymorphisms will not only help in fine-tuning DNA-based therapeutics but will also fulfill the ultimate goal of providing individualized therapies and medicines.

## REFERENCES


1. Bangham A (1992). Liposomes: Realizing their promise. *Hospital Prac.* 27:51–62.
2. Kim S (1993). Liposomes as carriers of cancer chemotherapy. *Drugs.* 46:618–638.
3. Drummond D C, Meyer O, Hong K, et al. (1999). Optimizing liposomes for delivery of chemotherapeutic agents to solid tumors. *Pharmacol. Rev.* 51:691–743.
4. Haag R, Kratz F (2006). Polymer therapeutics: Concepts and applications. *Angewandte Chemie Internat. Edi. Eng.* 45:1198–1215.
5. Tiera M J, Winnik F O, Fernandes J C (2006). Synthetic and natural polycations for gene therapy: state of the art and new perspectives. *Curr. Gene Ther.* 6:59–71.
6. Patil S D, Rhodes D G, Burgess D J (2005). DNA-based therapeutics and DNA delivery systems: A comprehensive review. *AAPS J.* 7:E61–E77.
7. Torchilin V P (2006). Recent approaches to intracellular delivery of drugs and DNA and organelle targeting. *Annual Rev. Biomedi. Engin.* 8:1.1–1.31.
8. Ding B-S, Dziubla T, Shuvaev V V, et al. (2006). Advanced drug delivery systems that target the vascular endothelium. *Molec. Intervent.* 6:98–112.
9. Wei C M, Gibson M, Spear P G, et al. (1981). Construction and isolation of a transmissible retrovirus containing the src gene of Harvey murine sarcoma virus and the thymidine kinase gene of herpes simplex virus type 1. *J. Virol.* 39:935–944.
10. Young L S, Searle P F, Onion D, et al. (2006). Viral gene therapy strategies: from basic science to clinical application. *J. Pathol.* 208:299–318.
11. Noureddini S C, Curiel D T (2005). Genetic targeting strategies for adenovirus. *Molec. Pharmaceuti.* 2:341–347.

12. Barouch D H (2006). Rational design of gene-based vaccines. *J. Pathol.* 208:283–289.
13. Nadeau I, Kamen A (2003). Production of adenovirus vector for gene therapy. *Biotechnol. Adv.* 20:475–489.
14. Moroziewicz D, Kaufman H L (2005). Gene therapy with poxvirus vectors. *Curr. Opin. Molec. Therapeuti.* 7:317–325.
15. Argnani R, Lufino M, Manservigi M, et al. (2005). Replication-competent herpes simplex vectors: design and applications. *Gene Ther.* 12:*S1*, S170–S177.
16. Wiznerowicz M, Trono D (2005). Harnessing HIV for therapy, basic research and biotechnology. *TRENDS Biotechnol.* 23:42–47.
17. Sinn P L, Sauter S L, McCray Jr P B (2005). Gene therapy progress and prospects: Development of improved lentiviral and retroviral vectors—design, biosafety, and production. *Gene Ther.* 12:1089–1098.
18. Trono D (2000). Lentiviral vectors: turning a deadly foe into a therapeutic agent. *Gene Ther.* 7:20–23.
19. Vasileva A, Jessberger R (2005) Precise hit: adeno-associated virus in gene targeting. *Nat. Rev. Microbiol.* 3:837–847.
20. Romano G (2006). The controversial role of adenoviral-derived vectors in gene therapy programs: Where do we stand? *Drug News Perspecti.* 19:99–106.
21. Kaplan J M (2005). Adenovirus-based cancer gene therapy. *Curr. Gene Ther.* 5:595–605.
22. Kanerva A, Hemminki A (2005). Adenoviruses for treatment of cancer. *Ann. Med.* 37:33–43.
23. Guo Z S, Bartlett D L (2004). Vaccinia as a vector for gene delivery. *Expert Opin. Biolog. Ther.* 4:901–917.
24. Bartosch B, Cosset F L (2004). Strategies for retargeted gene delivery using vectors derived from lentiviruses. *Curr. Gene Ther.* 4:427–443.
25. Anderson J L, Hope T J (2005). Intracellular trafficking of retroviral vectors: Obstacles and advances. *Gene Ther.* 12:1667–1678.
26. Kapturczak M H, Flotte T, Atkinson M A (2001). Adeno-associated virus (AAV) as a vehicle for therapeutic gene delivery: improvements in vector design and viral production enhance potential to prolong graft survival in pancreatic islet cell transplantation for the reversal of type 1 diabetes. *Curr. Molec. Med.* 1:245–258.
27. Romano G (2005). Current development of adeno-associated viral vectors. *Drug News Perspecti.* 18:311–316.
28. Vasileva A, Jessberger R (2005). Precise hit: Adeno-associated virus in gene targeting. *Nat. Rev. Microbiol.* 3:837–847.
29. Torchilin V P (2005). Recent advances with liposomes as pharmaceutical carriers. *Nat. Rev. Drug Disc.* 4:145–160.
30. Ewert K, Evans H M, Ahmad A, et al. (2005). Lipoplex structures and their distinct cellular pathways. *Adv. Genet.* 53:119–155.
31. Spagnou S, Miller A D, Keller M (2004). Lipidic carriers of siRNA: differences in the formulation, cellular uptake, and delivery with plasmid DNA. *Biochem.* 43:13348–13356.
32. Shuey D J, McCallus D E, Giordano T (2002). RNAi: gene-silencing in therapeutic intervention. *Drug Disc. Today.* 7:1040–1046.
33. Safinya C R (2001). Structures of lipid-DNA complexes: supramolecular assembly and gene delivery. *Curr. Opin. Struct. Biol.* 11:440–448.

34. Wu G Y, Wu C H (1988). Evidence for targeted gene delivery to Hep G2 hepatoma cells *in vitro*. *Biochem*. 27:887–892.
35. Lungwitz U, Breunig M, Blunk T, et al. (2005). Polyethylenimine-based non-viral gene delivery systems. *Euro. J. Pharm. Biopharm*. 60:247–266.
36. Neu M, Fischer D, Kissel T (2005). Recent advances in rational gene transfer vector design based on poly(ethylene imine) and its derivatives. *J. Gene Med*. 7: 992–1009.
37. Cho Y W, Kim J D, Park K (2003). Polycation gene delivery systems: escape from endosomes to cytosol. *J. Pharm. Pharmacol*. 55:721–734.
38. Demeneix B, Behr J P (2005). Polyethylenimine (PEI). *Adv. Genet*. 53:217–230.
39. Haensler J, Szoka F C Jr (1993). Polyamidoamine cascade polymers mediate efficient transfection of cells in culture. *Bioconj. Chem*. 4:372–379.
40. Eichman J D, Bielinska A U, Kukowska-Latallo J F, et al. (2000). The use of PAMAM dendrimers in the efficient transfer of genetic material into cells. *Pharmaceutical Sci. Technol. Today*. 3:232–245.
41. Zinselmeyer B H, Mackay S P, Schatzlein A G, et al. (2002). The lower-generation polypropylenimine dendrimers are effective gene-transfer agents. *Pharmaceut. Res*. 19:960–967.
42. Yoo H, Juliano R L (2000). Enhanced delivery of antisense oligonucleotides with fluorophore-conjugated PAMAM dendrimers. *Nucl. Acids Res*. 28:4225–4231.
43. Lee K Y, Kwon I C, Kim Y H, et al. (1998). Preparation of chitosan self-aggregates as a gene delivery system. *J. Control. Rel*. 51:213–220.
44. Prabaharan M, Mano J F (2005). Chitosan-based particles as controlled drug delivery systems. *Drug Del*. 2:41–57.
45. Shi C, Zhu Y, Ran X, et al. (2006). Therapeutic potential of chitosan and its derivatives in regenerative medicine. *J. Surg. Res*. 133:185–192.
46. Andre F, Mir L M (2004). DNA electrotransfer: Its principles and an updated review of its therapeutic applications. *Gene Ther. S1*, S33–42.
47. Li S (2004). Electroporation gene therapy: New developments *in vivo* and *in vitro*. *Curr. Gene Ther*. 4:309–316.
48. Prud'homme G J, Glinka Y, Khan A S, et al. (2006). Electroporation-enhanced non-viral gene transfer for the prevention or treatment of immunological, endocrine and neoplastic diseases. *Curr. Gene Ther*. 6:243–273.
49. Maitra A (2005). Calcium phosphate nanoparticles: Second-generation nonviral vectors in gene therapy. *Exp. Rev. Molec. Diagnos*. 5:893–905.
50. Nielsen P E, Egholm M (1999). An introduction to peptide nucleic acid. *Curr. Iss. Molec. Biol*. 1:89–104.
51. Marin V L, Roy S, Armitage B A (2004). Recent advances in the development of peptide nucleic acid as a gene-targeted drug. *Exp. Opin. Biolog. Ther*. 4:337–348.
52. Simonson O E, Svahn M G, Tornquist E, et al. (2005). Bioplex technology: Novel synthetic gene delivery pharmaceutical based on peptides anchored to nucleic acids. *Curr. Pharmac. Design*. 11:3671–3680.
53. Nielsen P E, Egholm M, Buchardt O (1994). Peptide nucleic acid (PNA). A DNA mimic with a peptide backbone. *Bioconj. Chem*. 5:3–7.
54. Järver P, Langel U (2006). Cell-penerating peptides—A brief introduction. *Biochim. Biophys. Acta*. 1758:260–263.
55. Wagstaff K M, Jans D A (2006). Protein transduction: cell penetrating peptides and their therapeutic applications. *Curr. Med. Chem*. 13:1371–1387.

56. Torchilin V P, Rammohan R, Weissig V, et al. (2001). TAT peptide on the surface of liposomes affords their efficient intracellular delivery even at low temperature and in the presence of metabolic inhibitors. *Proc. Nati. Acad. Sci. USA*. 98:8786–8791.
57. Marty C, Meylan C, Schott H, et al. (2004). Enhanced heparan sulfate proteoglycan-mediated uptake of cell-penetrating peptide-modified liposomes. *Cell. Molec. Life Sci*. 61:1785–1794.
58. Almeida J D, Edwards D C, Brand C M, et al. (1975). Formation of virosomes from influenza subunits and liposomes. *Lancet*. 2:899–901.
59. Kaneda Y (2000). Virosomes: Evolution of the liposome as a targeted drug delivery system. *Adv. Drug Del. Rev.* 43:197–205.
60. Felnerova D, Viret J F, Gluck R, et al. (2004). Liposomes and virosomes as delivery systems for antigens, nucleic acids and drugs. *Curr. Opin. Biotechnol.* 15:518–529.
61. Daemen T, de Mare A, Bungener L, et al. (2005). Virosomes for antigen and DNA delivery. *Adv. Drug Del. Rev.* 57:451–463.
62. Allen T M (2002). Ligand-targeted therapeutics in anticancer therapy. *Nat. Rev. Cancer.* 2:750–763.
63. Fenske D B, Cullis P R (2005). Entrapment of small molecules and nucleic acid-based drugs in liposomes. *Meth. Enzymol.* 391:7–40.
64. Sapra P, Tyagi P, Allen T M (2005). Ligand-targeted liposomes for cancer treatment. *Curr. Drug Del.* 2:369–381.
65. Park J W, Benz C C, Martin F J (2004). Future directions of liposome- and immunoliposome-based cancer therapeutics. *Seminars Oncol.* 6(suppl 13):196–205.
66. Marty C, Odermatt B, Schott H, et al. (2002). Cytotoxic targeting of F9 teratocarcinoma tumours with anti-ED-B fibronectin scFv antibody modified liposomes. *British J. Cancer.* 87:106–112.
67. Xu L, Huang C C, Huang W, et al. (2002). Systemic tumor-targeted gene delivery by anti-transferrin receptor scFv-immunoliposomes. *Molec. Cancer Ther.* 1:337–346.
68. May S, Ben-Shaul A (2004). Modeling of cationic lipid-DNA complexes. *Curr. Med. Chem.* 11:151–167.
69. Ferrari M E, Rusalov D, Enas J, et al. (2002). Synergy between cationic lipid and co-lipid determines the macroscopic structure and transfection activity of lipoplexes. *Nucleic Acids Res.* 30:1808–1816.
70. Tranchant I, Thompson B, Nicolazzi C, et al. (2004). Physicochemical optimisation of plasmid delivery by cationic lipids. *J. Gene Med.* 6:S24–S35.
71. Zhdanov R I, Podobed O V, Vlassov V V (2002). Cationic lipid-DNA complexes-lipoplexes-for gene transfer and therapy. *Bioelectrochem.* 58:53–64.
72. Matsuura M, Yamazaki Y, Sugiyama M, et al. (2003). Polycation liposome-mediated gene transfer *in vivo*. *Biochim. Biophys. Acta.* 1612:136–143.
73. Zhang J S, Liu F, Huang L (2005). Implications of pharmacokinetic behavior of lipoplex for its inflammatory toxicity. *Adv. Drug Del. Rev.* 57:689–698.
74. Freimark B D, Blezinger H P, Florack V J, et al. (1998) Cationic lipids enhance cytokine and cell influx levels in the lung following administration of plasmid: cationic lipid complexes. *J. Immunol.* 160:4580–4586.
75. Yu W, Pirollo K F, Rait A, et al. (2004). A sterically stabilized immunolipoplex for systemic administration of a therapeutic gene. *Gene Ther.* 11:1434–1440.
76. Audouy S, Molema G, De Leij L, et al. (2000). Serum as a modulator of lipoplex-mediated gene transfection: dependence of amphiphile, cell type and complex stability. *J. Gene Med.* 2:465–476.

77. Son K K, Tkach D, Hall K J (2000). Efficient *in vivo* gene delivery by the negatively charged complexes of cationic liposomes and plasmid DNA. *Biochim. Biophys. Acta.* 1468:6–10.
78. Lakkaraju A, Dubinsky J M, Low W C, et al. (2001). Neurons are protected from excitotoxic death by p53 antisense oligonucleotides delivered in anionic liposomes. *J. Biol. Chem.* 276:32000–32007.
79. Rothenfusser S, Tuma E, Wagner M, et al. (2003). Recent advances in immunostimulatory CpG oligonucleotides. *Curr. Opin. Molec. Ther.* 5:98–106.
80. Wang H, Rayburn E, Zhang R (2005). Synthetic oligodeoxynucleotides containing deoxycytidyl-deoxyguanosine dinucleotides (CpG ODNs) and modified analogs as novel anticancer therapeutics. *Curr. Pharmaceu. Des.* 11:2889–2907.
81. Chen W C, Huang L (2005). Non-viral vector as vaccine carrier. *Adv. Gene.* 54: 315–337.
82. Ludewig B, Barchiesi F, Pericin M, et al. (2001). *In vivo* antigen loading and activation of dendritic cells via a liposomal peptide vaccine mediates protective antiviral and anti-tumour immunity. *Vaccine.* 19:23–32.
83. Engler O B, Schwendener R A, Dai Wen J, et al. (2004). A liposomal peptide vaccine inducing CD8+ T cells in HLA-A2.1 transgenic mice, which recognise human cells encoding hepatitis C virus (HCV) proteins. *Vaccine.* 23:58–68.
84. Zhou H, Liu D, Liang C (2004). Challenges and strategies: The immune responses in gene therapy. *Med. Res. Rev.* 24:748–761.
85. Storni T, Kündig T M, Senti G, et al. (2005). Immunity in response to particulate antigen-delivery systems. *Adv. Drug Del. Rev.* 57:333–355.
86. Rose P G (2005). Pegylated liposomal doxorubicin: optimizing the dosing schedule in ovarian cancer. *Oncol.* 10:205–214.
87. Cattel L, Ceruti M, Dosio F (2004). From conventional to stealth liposomes: A new Frontier in cancer chemotherapy. *J. Chemother.* S4:94–97.
88. Straubinger R M, Arnold R D, Zhou R, et al. (2004). Antivascular and antitumor activities of liposome-associated drugs. *Anticancer Res.* 24(2A):397–404.R
89. Immordino M L, Brusa P, Rocco F, et al. (2004). Preparation, characterization, cytotoxicity and pharmacokinetics of liposomes containing lipophilic gemcitabine prodrugs. *J. Control. Rel.* 100:331–346.
90. Bergman A M, Kuiper C M, Noordhuis P, et al. (2004). Antiproliferative activity and mechanism of action of fatty acid derivatives of gemcitabine in leukemia and solid tumor cell lines and in human xenografts. *Nucleosides Nucleot. Nucleic Acids.* 23:1329–1333.
91. Harrington K J, Syrigos K N, Uster P S, et al. (2004). Targeted radiosensiti-sation by pegylated liposome-encapsulated 3′, 5′-O-dipalmitoyl 5-iodo-2′-deoxyuridine in a head and neck cancer xenograft model. *British J. Cancer.* 91:366–373.
92. Pignatello R, Puleo A, Puglisi G, et al. (2003). Effect of liposomal delivery on in vitro antitumor activity of lipophilic conjugates of methotrexate with lipoamino acids. *Drug Del.* 10:95–100.
93. Stevens P J, Sekido M, Lee R J (2004). A folate receptor-targeted lipid nanoparticle formulation for a lipophilic paclitaxel prodrug. *Pharmaceu. Res.* 21:2153–2157.
94. Lundberg B B, Risovic V, Ramaswamy M, et al. (2003). A lipophilic paclitaxel derivative incorporated in a lipid emulsion for parenteral administration. *J. Control. Rel.* 86:93–100.
95. Bergman A M, Kuiper C M, Myhren F, et al. (2004). Antiproliferative activity and mechanism of action of fatty acid derivatives of arabinosylcytosine (ara-C) in leukemia and solid tumor cell lines. *Nucleosides Nucleot. Nucleic Acids.* 23:1523–1526.

96. Lopez-Barcons L A, Zhang J, Siriwitayawan G, et al. (2004). The novel highly lipophilic topoisomerase I inhibitor DB67 is effective in the treatment of liver metastases of murine CT-26 colon carcinoma. *Neoplasia*. 6:457–467.
97. Fahr A, van Hoogevest P, May S, et al. (2005). Transfer of lipophilic drugs between liposomal membranes and biological interfaces: consequences for drug delivery. *Euro. J. Pharmaceu. Sci.* 26:251–265.
98. Strickley R G (2004). Solubilizing excipients in oral and injectable formulations. *Pharmaceu. Res.* 21:201–230.
99. ten Tije A J, Verweij J, Loos W J, et al. (2003). Pharmacological effects of formulation vehicles: Implications for cancer chemotherapy. *Clin. Pharmacokin.* 42:665–685.
100. Krise J P, Stella V J (1996) Prodrugs of phosphates, phosphonates, and phosphinates. *Adv. Drug Del. Rev.* 19:287–310.
101. Wagner C R, Iyer V V, McIntee E J (2000). Pronucleotides: Toward the in vivo delivery of antiviral and anticancer nucleotides. *Med. Res. Rev.* 20:417–451.
102. Schwendener R A, Schott H (2005). Lipophilic arabinofuranosyl cytosine derivatives in liposomes. *Meth. Enzymol.* 391:58–70.
103. Hamada A, Kawaguchi T, Nakano M (2002). Clinical pharmacokinetics of cytarabine formulations. *Clinical Pharmakinet.* 41:705–718.
104. Schwendener R A, Schott H (1996). Lipophilic 1-β-D-arabino-furanosylcytosine derivatives in liposomal formulations for oral and parenteral antileukemic therapy in the murine L1210 leukemia model. *J. Cancer Res. Clini. Oncol.* 122:723–726.
105. Schwendener R A, Friedl K, Depenbrock H, et al. (2001). In vitro activity of liposomal $N^4$octadecyl-1-β-D-arabino-furanosylcytosine (NOAC), a new lipophilic derivative of 1-β-D-arabino-furanocylcytosine on biopsized clonogenic human tumor cells and hematopoietic precursor cells. *Investigat. New Drugs.* 19:203–210.
106. Cattaneo-Pangrazzi R M C, Schott H, Wunderli-Allenspach H, et al. (2000). The novel heterodinucleoside dimer 5-FdU-NOAC is a potent cytotoxic drug and a p53-independent inducer of apoptosis in the androgen-independent human prostate cancer cell lines PC-3 and DU-145. *Prostate.* 45:8–18.
107. Cattaneo-Pangrazzi R M C, Schott H, Wunderli-Allenspach H, et al. (2000). New amphiphilic heterodinucleoside phosphate dimers of 5-fluorodeoxyuridine (5FdUrd): Cell cycle dependent cytotoxicity and induction of apoptosis in PC-3 prostate tumor cells. *Biochem. Pharmacol.* 60:1887–1896.
108. Takatori S, Kanda H, Takenaka K, et al. (1999). Antitumor mechanisms and metabolism of the novel antitumor nucleoside analogues, 1-(3-C-ethynyl-β-D-ribopento-furanosyl)cytosine and 1-(3-C-ethynyl-β-D-ribopento-furanosyl)-uracil. *Cancer Chemother. Pharmacol.* 44:97–104.
109. Tabata S, Tanaka M, Endo Y, et al. (1997). Anti-tumor mechanisms of 3′-ethynyluridine and 3′-ethynylcytidine as RNA synthesis inhibitors: Development and characterization of 3′-etyhynyluridine-resistant cells. *Cancer Lett.* 116:225–231.
110. Yokogawa T, Naito T, Kanda H, et al. (2005). Inhibitory mechanisms of 1-(3-C-ethynyl-β-D-ribopentofuranosyl)uracil (EUrd) on RNA synthesis. *Nucleosides Nucleot. Nucleic Acids.* 24:227–232.
111. Ludwig P S, Schwendener R A, Schott H (2005). Synthesis and anticancer activities of amphiphilic 5-fluoro-2′-deoxyuridylic acid prodrugs. *Euro. J. Med. Chem.* 40:494–504.
112. Saiko P, Horvarth Z, Bauer W, et al. (2004). *In vitro* and *in vivo* antitumor activity of novel amphiphilic dimers consiting 5-fluorodeoxyuridine and arabinofuranosylcytosine. *Internat. J. Oncol.* 25:357–364.

# GROWTH FACTORS AND CYTOKINES

Mandeep K. Mann and Brian E. Cairns
*University of British Columbia, Vancouver, British Columbia, Canada*

**Chapter Contents**

1 Background
- 1.1 Neuropathic and Nociceptive Pain Mechanisms
- 1.2 Introduction to Nerve Growth Factor
- 1.3 Introduction to Tumor Necrosis Factor α
- 1.4 Clinical Significance
- 1.5 NGF and Experimental Muscle Pain
- 1.6 TNFα and Experimental Joint Pain

2 Case Studies
- 2.1 Effect of IM injection of NGF in humans
- 2.2 NGF effects on masticatory muscle nociceptors

References

## 1 BACKGROUND

### 1.1 Neuropathic and Nociceptive Pain Mechanisms

***Role of Growth Factors and Cytokines in These Pain Mechanisms (Peripheral).*** Pain is generally classified into two main categories: neuropathic and nociceptive (inflammatory) pain. Nociceptive pain is caused by the activation of primary afferent fibers known as nociceptors [1]. Primary nociceptive afferents (Aδ and C fibers) can become sensitized as a result of various factors such as tissue trauma,

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

inflammation, and ischemia [2]. Peripheral injury (tissue damage) results in the synthesis and release of various inflammatory mediators that induce inflammation and edema as part of the healing process. These mediators include bradykinin, substance P, histamine, serotonin (5-HT), glutamate, acetylcholine (Ach), ATP, ions ($H^+$, $K^+$), cholecystokinin, nerve growth factor (NGF) [3], cytokines such as tumor necrosis factor α (TNFα), interleukin-1β (IL-1β) and interleukin-6 (IL-6), and eicosanoids (such as prostaglandin $E_2$, prostaglandin $I_2$, and leukotriene $B_4$), which excite and sensitize nociceptors as well as recruit additional nociceptors to enhance pain perception [1, 4]. Peptide neurotransmitters such as substance P (SP) and calcitonin gene-related peptide (CGRP) can be released from the nerve terminals of nociceptors into damaged tissues, a process that serves to augment the effects of inflammatory agents by causing vasodilation and increased capillary permeability. Bradykinin (released from blood vessels) and other chemical agents such as 5-HT released from platelets, histamine from mast cells, and eicosanoids from various cellular elements further contribute to sensitization of nociceptors, by either opening ion channels or activating second messenger systems [5].

The process of peripheral sensitization is manifested by a decrease in the activation threshold of nociceptors (allodynia) and an increase in the response of nociceptors to noxious stimulation (hyperalgesia) [6, 7]. In addition, peripheral sensitization may be enhanced due to the recruitment of "silent" nociceptors [6, 8], which are normally unresponsive to thermal or mechanical stimuli in healthy tissue. The release of various chemical mediators in the periphery also causes repetitive activity in primary afferent fibers. This barrage of input results in the central release of neurotransmitters and neuropeptides that act on brain cells or neurons in the central nervous system to increase their response to tissue stimulation. This phenomenon is known as central sensitization and results in the amplification and persistence of nociceptive signals that are eventually perceived as pain. It is theorized that central sensitization is the mechanism underlying pain referral and clinical signs of secondary hyperalgesia, which refers to the sensitization of noninflamed tissue surrounding the site of inflammation in the periphery [9–12].

Normally, as the injured tissue heals and inflammation subsides, the phenomena of peripheral and central sensitization dissipate [1]. However, under certain conditions, pain does not resolve with tissue healing and becomes chronic [1, 4]. It is thought that the transition from acute to chronic pain under these conditions may reflect damage or dysfunction of the peripheral and/or central nervous system, which results in abnormal processing of sensory input, and thus, this type of pain is called neuropathic pain [1, 13, 14]. The pathophysiology of neuropathic pain syndromes (such as postherpetic neuralgia, diabetic neuropathy, posttraumatic neuralgia etc.) is not completely understood, but it is considered to be complex and multifactorial [1, 13–15]. Neuropathic pain syndromes and peripheral nerve injury are generally accompanied by a cascade of inflammatory events ("neuroinflammatory cascade") mediated by a group of cytokines and other molecules, as mentioned above [4]. Cytokines such as TNFα and IL-1β, released at the site of inflammation, bind to their receptors on primary afferent fibers and activate and/or sensitize nociceptors, thus inducing and maintaining pain as well as a state of hyperalgesia [4]. Studies have shown that the expression of the receptors of these cytokines increases after peripheral nerve injury [16–18] and likely contributes to increased

neuronal excitability. Also, use of cytokine antagonists has been shown to reduce nociceptive behavior in rodents [19–21], suggesting a role for TNFα and IL-1β in mediating neuropathic-like pain. Similarly, neurotrophic factors such as NGF have also been shown to be involved in nociceptive and neuropathic pain conditions such as diabetic peripheral neuropathies.

This chapter will focus on NGF and TNFα and their role in the mechanisms of neuropathic and nociceptive pain.

## 1.2 Introduction to Nerve Growth Factor

***Discovery of NGF.*** Rita Levi-Montalcini, an Italian scientist, discovered NGF in the early 1950s. Levi-Montalcini's research on the central nervous system and its peripheral targets led her to hypothesize that "the failure of neurons to thrive in the absence of peripheral target tissue was because of a degenerative process" [22]. Through her research collaboration with Victor Hamburger, a neuroembryologist, Levi-Montalcini gained further evidence that confirmed her hypothesis. Their work on chick embryos led them to hypothesize that "developing nerve cells depend on feedback signals that are in limited supply, and the neuronal targets provide a specific signal that is required for neuronal survival" [22, 23]. This "specific signal" was later identified to be "a diffusible agent that stimulated the growth and differentiation of nerve cells" and thus named the "nerve growth factor." The biological characterization (isolation and purification) of NGF was achieved through experiments performed by Levi-Montalcini in collaboration with Stanley Cohen, a biochemist who discovered the epidermal growth factor [22]. In 1986, both Rita Levi-Montalcini and Stanley Cohen received a Nobel Prize in medicine for their discovery of NGF [23].

NGF was the first neurotrophic factor to be characterized [24]. The structure of NGF comprises 2α, 2β, and 2γ subunits and one or two zinc ions [25]. The β subunit of the complex is released as a result of autocatalytic cleavage initiated by the loss of zinc ions. The biologically active form of NGF is NGF-β, which is a dimeric molecule (2β subunits) weighing 26.8 kDa, with each monomer consisting of 118 amino acids and three disulfide bridges [25, 26].

***Role/Function of NGF.*** NGF is a neurotrophic protein molecule essential for the growth and survival of sympathetic and small-diameter afferent neurons (nociceptors) [27, 28] and is involved in neuronal function and plasticity [24]. NGF belongs to the neurotrophin family, which also includes the brain-derived neurotrophic factor (BDNF), glial-derived neurotrophic factor (GDNF), neurotrophin-3 (NT-3), and neurotrophin 4/5 (NT-4/5) [29, 30].

***Sources of NGF.*** Neurons as well as non-neuronal cells such as mast cells, fibroblasts, eosinophils, T and B lymphocytes, and epithelial cells synthesize NGF [31, 32]. NGF and other neurotrophins can also be synthesized and secreted by sympathetic and sensory target organs [24]. High levels of NGF are present in the central nervous system (CNS) where it is known to play a crucial role in growth and plasticity [29]. During development, NGF is secreted by target tissues of sensory axons/neurons [24]. Also, during peripheral nerve injury, NGF synthesis is initiated in Schwann cells and fibroblasts within the injured nerve by cytokines

released in response to the injury (for survival and regeneration of the injured nerve) [24].

***NGF Receptors.*** NGF exerts its biological effects by acting on two different receptors found on target cells: tyrosine kinase receptor A (TrkA; tropomyosin related kinase A) and p75 receptor [29]. The TrkA receptor is a tyrosine kinase consisting of intracellular, transmembrane, and extracellular domains, and NGF binding results in phosphorylation of key transduction-related proteins or ion channels causing downstream receptor modulation [29, 32]. The TrkA receptor has a very slow dissociation rate for NGF and is therefore called a high-affinity NGF receptor [33]. In the periphery, TrkA receptors are expressed on small-diameter sensory neurons, inflammatory cells, and sympathetic neurons [30, 34]. The p75 receptor is a member of the superfamily of TNF (tumor necrosis factor) receptor-related molecules that has extracellular and transmembrane domains. The p75 receptor can bind all neurotrophins [29]. The p75 receptor is known as the low-affinity NGF receptor because it has fast association and dissociation rates for NGF binding [33]. The binding of NGF to TrkA and p75 receptors results in the activation of downstream signal transduction pathways that mediate the physiological effects of NGF [35].

***Downstream Signaling Pathways of NGF.*** NGF binding to the TrkA receptors causes receptor activation and dimerization, which results in transautophosphorylation and activation of intracellular signaling cascades like extracellular signal-regulated kinase (ERK), phosphatidylinositol 3-kinase (PI3K), and phospholipase Cγ (PLC-γ) pathways [24, 35]. The Src homologous and collagen-like (Shc) adaptor protein binds to the activated Trk receptor, and its phosphorylation results in an increase in the activities of PI3K and Akt (protein kinase B), which are involved in neuronal survival [35]. The Shc binding to TrkA receptors also increases the activities of Ras and ERK, which can influence transcriptional events such as induction of cyclic AMP-response element binding (CREB) transcription factor that is involved in neurite outgrowth, cell cycle, and synaptic plasticity [35]. It is thought that sustained activation of the ERK pathway requires the internalization of the NGF–TrkA receptor complex into membrane vesicles [24, 36]. The PLC-γ signaling cascade is initiated by direct PLC-γ binding to the activated Trk receptor and results in protein kinase C activation and the release of inositol phosphates and calcium [24, 35]. It has been suggested that activation of these signaling pathways may result in the modulation of ion channels such as the capsaicin receptor channel known as the TRPV1 or VR1 channel [35, 37]. The p75 receptor also activates signaling components such as Jun N-terminal kinase (JNK), nuclear factor-κB (NF-κB), and ceramide to mediate processes such as apoptosis or cell death [35].

***NGF Tissue Levels.*** As compared with neurotransmitters, neurotrophic factors such as NGF exist in very low concentrations; therefore, their detection is very difficult [38]. The maximum peak level of NGF in the interstitial fluid of the brain of patients with severe head injury was found to be 1100 pg/mL [39]. A study using an enzyme-linked immunosorbent assay (ELISA) (measuring unbound β-NGF concentration) has shown that men have significantly higher serum NGF

concentration (40.8 ± 10.8 pg/mL) than women (8.2 ± 1.4 pg/mL in the follicular phase and 14.4 ± 2.9 pg/mL in the luteal phase) [40]. As differences in the NGF concentrations in both the follicular and the luteal phase of the menstrual cycle were reported to be statistically significant, it has been suggested that female sex hormones (and androgens) influence circulating NGF levels [40].

### 1.3 Introduction to Tumor Necrosis Factor α

***Discovery of TNFα.*** In 1975, Carswell et al. published a study that showed that endotoxin treated serum of bacillus Calmette-Guerin (BCG)-infected mice induces the release of a substance that is selectively toxic for tumor/malignant cells [41]. They reported that this substance was as effective at causing necrosis of tumors as endotoxin itself and named it "the tumor necrosis factor" (TNF) [41]. TNF is now considered a part of a large and diverse group of polypeptides called cytokines, whose major role is to mediate interactions between the inflammatory and the immune system [42, 43]. There are two forms of TNF: TNFα and TNFβ. TNFα is synthesized as a prohormone, and its soluble form (17 kD, nonglycosylated protein) is released from the 26-kD transmembrane protein via specific proteolytic cleavage [44, 45]. TNFβ or lymphotoxin shares 28% sequence homology with TNFα [46].

***Role/Function of TNFα.*** TNFα belongs to a family of cytokines that also includes interleukins, interferons, and transforming growth factors [42]. TNFα is involved in various processes such as inflammation, cytotoxicity, (apoptotic) cell death, metabolism (insulin resistance), thrombosis, and fibrinolysis [44, 47]. TNFα is also involved in exerting endocrine, paracrine, and autocrine control of inflammatory responses. It can activate macrophages [46], initiate a cascade of other cytokines (IL-1β, IL-6, and IL-8) and growth factors (NGF), and can recruit circulating inflammatory cells to the local site of inflammation to induce edema [47, 48]. TNFα-induced inflammatory response also involves the release of arachidonic acid, which results in the production of pro-inflammatory mediators such as prostaglandins and leukotrienes [49]. TNFα plays a major role in the pathogenesis of septic shock syndrome that results from the body's reaction to a bacterial, viral, or parasitic infection [49]. In the CNS, TNFα has both neurotoxic and neuroprotective effects [50, 51]. Recent research has suggested that TNFα (and other cytokines) link the immune and nervous system and may be involved in the generation of pain and hyperalgesia [52]. It has also been suggested that cytokines like TNFα can induce the production of NGF in both the peripheral and the central nervous system [53].

***Sources of TNFα.*** Various stimuli such as endotoxins, superantigens, osmotic stress, injury, inflammation, and radiation can cause the release of TNFα [48]. TNFα is produced mainly by monocytes and/or macrophages and T cells [45, 48]. However, small amounts of TNFα can be released by various other cell types [44, 52]. For instance, Schwann cells can produce TNFα after injury. Like several other cytokines, TNFα can induce its own production as well [52].

***TNFα Receptors.*** TNFα exerts its effect through two high-affinity, cell-surface receptors: p55 and p75 receptors [42, 52]. Both TNF receptors are expressed in the

CNS [42]. Both receptors are glycoproteins and have a single-membrane spanning hydrophobic segment [49]. At least one of these two TNF receptors is present on virtually all cells, explaining the pleiotropic nature of TNFα [44, 45]. It has been shown that tissues including kidney, liver, adipose, and muscle tissue express the p55 receptor [46]. Most of the effects of TNFα are mediating via its binding and activation of the p55 receptor. Some processes such as cytotoxicity and thymocyte proliferation are also associated with the p75 receptor [49]. The binding of TNFα to its receptors results in the activation of a multitude of complex signaling pathways. TNFβ also binds to the same receptors as TNFα and mediates similar biological actions [49].

***Downstream Signaling Pathways of TNFα.*** The binding of TNFα to its receptors results in the activation of multiple signal transduction pathways, kinases, and transcription factors that activate several cellular genes [45]. These signaling pathways include those that activate transcription factors (such as NF-κB and AP1), protein kinases (such as mitogen activated protein kinase or MAPK, JNK, p38), and proteases [44]. The binding of TNFα to the p55 receptor results in the activation of phospholipase C (PLC) and production of diacylglycerol (DAG), which then activates a calcium-independent protein kinase C (PKC) isotype [49]. Through an unknown pathway, PKC then activates Jun and Fos proteins that are components of AP-1 transcription factor [49, 54] and therefore results in the induction of AP-1 responsive genes [54]. DAG can also activate a C type phospholipase known as SMase, which hydrolyzes sphingomyelin (SM) to ceramide, a second messenger for processes such as apoptosis. In addition, p55 receptor activation can result in the phosphorylation of phospholipase $A_2$ ($PLA_2$), which triggers the release of arachidonic acid (AA) and production of eicosanoids involved in the process of inflammation [49]. $PLA_2$ activation is also thought to be involved in TNFα-induced cytotoxicity. TNFα can also activate transcription factors such as NF-κB and JNK. Several adaptor proteins that initiate these pathways and link TNF receptors to their downstream targets have been identified. One such protein is called the death domain protein, which is involved in the process of apoptosis [44, 51]. Another adaptor protein is part of the TNF receptor associated factors (TRAFs) family, which consists of six distinct proteins, but only TRAF2, TRAF5, and TRAF6 are thought to mediate NF-κB and JNK activation [44]. Another molecule required for TNFα mediated NF-κB activation is known as the receptor interacting protein (RIP) [44]. NF-κB activation results in the synthesis of new proteins, some of which are involved in cell death, whereas others result in cell proliferation. Therefore, depending on the cell cycle, TNFα can either initiate apoptosis (programmed cell death) or be protective [51].

***TNFα Tissue Levels.*** Cytokines are produced on demand, and they only travel over short distances. It is thought that *in vivo* concentrations of cytokines are in the range of a few picograms to nanograms per milliliter [52]. In a study conducted on baboons, infusion of a lethal dose of *Escherichia coli* increased plasma TNFα concentration to a peak level of 20,500 ± 9890 pg/mL within 90 minutes [55]. Similarly, it has been reported that in human volunteers, a bolus injection of endotoxin results in a burst increase in TNFα levels (peak level: 358 ± 166 pg/mL) followed by a rapid decline to undetectable levels [45, 56].

## 1.4 Clinical Significance

Currently available surgical techniques and powerful analgesics such as opioids are relatively ineffective for the treatment of painful neuropathies or, in the case of opioid analgesics, have other limitations such as adverse side effects [4]. Therefore, researchers have targeted other cellular and molecular mechanisms and mediators that are thought to be involved in neuropathic and inflammatory pain conditions in an effort to develop new drugs and therapeutic strategies. One such molecule is the nerve growth factor.

***NGF.*** NGF levels change during some neurodegenerative disease states and other pathological conditions [57, 58]. NGF levels have been increased exogenously in experiments investigating diabetic peripheral neuropathies. There is a loss of dorsal root ganglion neurons and degeneration of Schwann cells and small-diameter sensory neuronal fibers during diabetic peripheral neuropathies [59]. Experimentally, it has been shown that induction of diabetes with streptozotocin causes a decrease in NGF levels in the sensory neurons and target tissues of rats [60, 61] and that human diabetic patients have low NGF levels in the skin [62]. Also, studies have shown that NGF uptake and retrograde transport mechanisms are impaired in animal models of diabetes [63, 64]. Therefore, it was thought that NGF replacement therapy might have a beneficial effect on the treatment of this condition due to its involvement in the promotion of neuronal growth and survival as well as its ability to prevent neuronal damage to small-diameter sensory and autonomic neurons [65].

Initial animal studies showed that increasing NGF levels in experimental models of peripheral neuropathy had beneficial effects (prevented the decrease in SP and CGRP levels in dorsal root ganglion neurons and prevented the loss of sensitivity to thermal noxious stimuli, i.e., prevented the onset of symptoms associated with polyneuropathy) [65–67]. This led to phase I (0.03–1-μg/kg human NGF) and II (0.1- or 0.3-μg/kg human NGF three times a week for 6 months) clinical trials that involved subcutaneous administration of human NGF into healthy human volunteers and diabetic patients suffering from polyneuropathy [68, 69]. Although symptoms of polyneuropathy were decreased by NGF treatment, these trials also revealed that human NGF administration results in serious dose-limiting, pain-related systemic side effects such as hyperalgesia at the site of injection, diffuse myalgias or arthralgias, and leg cramps (more frequent and severe with higher doses) [68, 69]. The positive results from phase II clinical trials resulted in a phase III clinical trial (0.1-μg/kg human NGF three times a week for 48 weeks) that also reported pain-related side effects in response to human NGF administration but failed to show the efficacy of human NGF in treating diabetic polyneuropathy [70]. Although initially NGF seemed to be an ideal therapeutic agent for the treatment of diabetic neuropathy, the failure of these subsequent clinical trials has served to highlight the potentially important role altered NGF levels may play in the development of chronic joint and muscle pain.

NGF is also considered to be a novel therapeutic target for the treatment of other types of neuropathic pain. Studies conducted on animal models of neuropathic pain suggest that there might be a relationship between neuropathic pain and NGF [71–74]. Procedures used to induce neuropathic pain such as sciatic nerve

constriction [73] or spinal nerve ligation [75] result in pain behaviors such as allodynia and mechanical and thermal hyperalgesia [71, 73, 74] and lead to increases in the levels of NGF [75, 76]. Administration of NGF antiserum has been shown to inhibit mechanical and thermal hyperalgesia caused by elevated endogenous NGF levels [72, 77], which suggests that NGF is a mediator of mechanical sensitivity. Also, neuropathic patients with chronic hyperalgesia and allodynia have increased levels of NGF in the area of peripheral neuropathy [78]. However, some studies conducted on animal models of nerve constriction injury have reported contradictory results. For example, exogenous NGF administration (chronic infusion of 10 μL of 0.5-mg/mL NGF) has been shown to abolish behavioral hyperalgesia (decrease in mechanical threshold) caused by chronic constriction injury of the rat sciatic nerve [79]. Also, a study has shown that intrathecal rat NGF infusion restores the antiallodynic and antihyperalgesic effects of morphine in a rat model of sciatic nerve constriction injury [80]. Overall, these studies suggest that there might be a correlation between NGF levels in the periphery and pain behaviors such as mechanical allodynia and thermal hyperalgesia that are hallmarks of neuropathic pain.

Tissue inflammation is known to result in an increase in the production and release of NGF [32, 81–83]. Studies have shown that inflammatory mediators such as interleukin-1 (IL-1), IL-4, IL-5, tumor necrosis factor-α, and interferon-γ induce NGF release. In turn, NGF can augment neurogenic inflammation by promoting the release of inflammatory mediators from basophils, mast cells, macrophages, and T and B lymphocytes [32, 34]. Various inflammatory diseases such as rheumatoid arthritis [84], multiple sclerosis [31], and systemic lupus erythematosus [31, 81] lead to upregulation of NGF synthesis. Anti-NGF antibodies have been shown to significantly reduce inflammatory hypersensitivity caused by high levels of NGF [85]. Therefore, NGF is considered to be a key mediator in the production of inflammatory pain. It has been suggested that NGF influences inflammatory responses by affecting immune cell function, altering neuropeptide (SP and CGRP) levels in sensory fibers [1, 31, 32, 86], or sensitizing nerve terminals through receptor phosphorylation [34]. Therefore, elevated endogenous levels of NGF seem to be associated with neuropathic and inflammatory pain. However, the use of anti-NGF antibodies for the treatment of these conditions is still under research, as some studies have produced contradictory results.

***TNFα.*** Another novel therapeutic target that has been identified is TNFα. TNFα (and IL-1) plays a major role in the pathogenesis of inflammatory diseases such as arthritis [47]. It has been shown that there is an increased expression of TNFα producing cells in the synovial tissue samples of osteoarthritis patients [87] and that TNFα levels have also been detected in the synovial fluid and synovium of rheumatoid arthritis patients [47, 88]. It is thought that TNFα (acting synergistically with IL-1) increases the synthesis of matrix degrading proteins like interstitial collagenase and stromelysin [49, 89] and causes tissue (cartilage) destruction observed in rheumatoid arthritis, osteoarthritis, and other joint diseases [46, 51, 88]. Injections of TNFα have been shown to accelerate the onset or increase the severity of collagen-induced arthritis in rats and mice [90, 91]. The use of anti-TNFα antibodies or other TNFα inhibitors has been shown to treat collagen-induced arthritis in mice [92, 93]. Also, a study has shown that anti-TNFα antibodies

can reduce inflammation (foot swelling) and joint destruction even when used after disease onset [94].

The positive results of these animal studies led to clinical trials with monoclonal antibodies of TNFα [95–97]. An open label trial was conducted with chimeric (75% human and 25% murine) monoclonal anti-TNFα antibody cA2 (infliximab) to test its safety and efficacy in the treatment of rheumatoid arthritis [97, 98]. Infliximab was infused two to four times over 2 weeks at a total dose of 20-mg/kg body weight in 20 patients with rheumatoid arthritis unresponsive to disease-modifying anti-rheumatic drugs (DMARD) therapy. It was found that this treatment led to a rapid improvement in patient symptoms; i.e., patients had decreased number of swollen/tender joints, decreased degree of pain, and stiffness in the joints. Also, there was a decrease in the laboratory measurements of inflammatory activity [erythrocyte sedimentation rate and C-reactive proteins (CRP)]. The maximum clinical response (change from baseline) was achieved within the first month of treatment and lasted for 8–22 weeks. The infusions were reported to be safe as they were well tolerated and did not cause any adverse effects. It was reported that TNFα monoclonal antibody treatment might be useful for long-term therapy (for, e.g., for controlling chronic diseases) as seven patients that were retreated (given repeated infusions) after a relapse also showed significant improvement [96].

Another TNFα monoclonal antibody known as etanercept (a fusion protein of human IgG and two p75 TNF receptors) [98] that is a competitive inhibitor of TNF's cell-surface receptor has been developed. Clinical studies have shown that etanercept therapy results in significant clinical benefits and is well tolerated and safe for long-term use [99–101]. Etanercept is given subcutaneously (25 mg) twice weekly for the treatment of rheumatoid arthritis [98]. It is also used in the treatment of psoriatic arthritis, and a recent clinical trial has reported that in addition to decreasing joint pain and fatigue, etanercept also decreases symptoms of depression associated with this chronic disease [102].

Recently, a fully human monoclonal TNFα antibody called adalimumab (a humanized $IgG_1$) [98] has also been developed for the treatment of rheumatoid arthritis. Double-blind, placebo-controlled clinical trials have shown that the addition of adalimumab to methotrexate therapy resulted in significant, long-term improvement (slowed progression of structural joint damage and a reduction in signs and symptoms) in rheumatoid arthritis patients [103, 104]. Therefore, adalimumab has been approved for treating rheumatoid arthritis as a monotherapy and in combination with methotrexate [98]. Adalimumab is administered subcutaneously every 2 weeks (45 mg) [98]. Currently, all three of these TNFα inhibitors are being used in the treatment of various inflammatory conditions, such as rheumatoid arthritis, ankylosing spondylitis, psoriatic arthritis, and Crohn's disease.

### 1.5 NGF and Experimental Muscle Pain

***Animal Research.*** Studies in adult rats [83, 105] have shown that NGF administration leads to hyperalgesia and sensitization of cutaneous nociceptive neurons [6, 82, 106–108]. Local or systemically administered mouse or human NGF resulted in two phases of hyperalgesia in response to thermal and mechanical stimuli in adult male and female rats: an early phase starting 30 minutes after NGF administration and a late, but longer lasting phase starting several hours after NGF

administration and lasting for several days [30, 105–107, 109]. Thermal hyperalgesia occurred during both phases but mechanical hyperalgesia only during the late phase. Local, but not systemic, administration of NGF resulted in hyperalgesia that was limited to the site of injection. Interestingly, injection of human NGF has not been reported to result in any overt pain behavior in these animals [30]. It was subsequently demonstrated that human NGF applied directly to the receptive fields of cutaneous afferent fibers decreased the thermal or heat threshold without changing the mechanical or cold threshold [30, 107]. These results suggest that the first phase of thermal hyperalgesia is due to the peripheral effects of NGF, whereas the second phase of thermal and mechanical hyperalgesia are likely a result of central mechanisms.

***Human Research.*** In healthy men and women, single intravenous or subcutaneous doses of human NGF (0.03 to 1 μg/kg) produced mild-to-moderate muscle pain, especially pain in the masseter muscle that increased with chewing, beginning 1–1.5 hours postinjection [69]. Beginning at 60–90 minutes after human NGF administration, subjects reported muscle pain that increased at 4 to 6 hours and then slowly decreased over 2 to 8 days. Female subjects reported greater muscle pain as compared with male subjects, and the effect was found to be dose-dependent, which suggests that human NGF may have sex-dependent effects [69]. Subcutaneous injection of human NGF resulted in hyperalgesia to touch and heat at the site of the injection (in addition to mild, diffuse myalgias), which lasted for 7 weeks and was found to be dose-dependent [69]. This study demonstrated that human NGF administration results in diffuse muscle pain (including pain in the orofacial region) that peaks hours after injection and increases with muscle use. Another study conducted by Dyck et al. [110] had reported that healthy human subjects injected with intradermal injection of human NGF in the forearm (1 or 3 μg) developed pressure allodynia and heat-pain hyperalgesia within 3 hours on the NGF-injected side. The authors suggested that local tissue mechanisms might be responsible for the rapid onset of these processes [110].

### 1.6 TNFα and Experimental Joint Pain

***Animal Research.*** In 1992, Cunha et al. conducted a study in rats to determine the hyperalgesic effects of various cytokines including TNFα [111]. They found that TNFα played a crucial role in the development of inflammatory hyperalgesia and blockade of TNFα activity completely abolished the inflammatory response. They reported that TNFα-induced hyperalgesia occurs through both prostaglandin and sympathetically mediated pathways [111]. Another study conducted on male rats has shown that repeated intraplantar injections of TNFα into the rat paw (2.5 pg daily for 30 days) caused persistent mechanical hypersensitivity in nociceptors that lasted for at least 30 days after the treatment was stopped [112]. The authors reported that hypersensitivity was not induced when indomethacin and atenolol were injected with TNFα, suggesting that TNFα-induced mechanical hypersensitivity was mediated via the release of eicosanoids and sympathomimetic mediators, as suggested by the Cunha et al. study [112]. Various other studies have also shown that intraplantar injections of TNFα can induce mechanical allodynia and hyperalgesic effects in rats [52, 111, 113, 114].

Researchers have also used an *in vitro* rat skin model to determine the effects of TNFα on heat-induced release of CGRP (neuropeptide) from nociceptors [115]. They found that, in contrast to the excitatory effects of TNFα *in vivo*, it did not induce any CGRP release at 32°C; however, TNFα produced a transient (short-term) and dose-dependent heat sensitization. Thus, the authors speculated that, in addition to their role in mediating long-term sensitization in various chronic inflammatory conditions [52, 116], cytokines such as TNFα might also play a role in the acute phase of inflammation, i.e., cause heat sensitization through receptor-associated mechanisms (phosphorylation of ion channels) [115] rather than through signaling pathways. A recent study has also suggested that TNFα-induced rapid thermal and mechanical sensitization processes (after peripheral administration) are probably due to a direct effect on afferent fibers [117].

TNFα is also thought to be one of the mediators involved in formalin-induced orofacial nociception in behavioral rat models [118] and carrageenan (CG)-induced inflammation in mice [119, 120]. A study has shown that intraarticular injections of TNFα could cause an intense incapacitation response in carrageenan-primed rat knee joints that lasts for more than 8 hours [121]. The treatment of knee joints with anti-TNFα antibodies (intraarticular injection) significantly inhibited CG-induced incapacitation response, which suggests that TNFα is a mediator of CG-induced inflammatory incapacitation and is involved in maintaining nociceptive responses over the long term [121]. Another study has shown that induction of arthritis in mice (streptococcal cell wall arthritis model) caused the release of TNFα, which resulted in significant knee joint swelling [122]. The use of anti-TNFα treatment markedly reduced the joint swelling, which suggests a role for TNFα in this arthritis model [122]. The contribution of TNFα to inflammatory pain has been assessed in another experimentally induced model of arthritis (complete Freund's adjuvant or CFA) [123]. It was found that CFA-induced joint inflammation in rats was associated with increased expression of both TNFα receptors in the dorsal root ganglion cells. Also, the number of macrophages was correlated with the development of hyperalgesia, which suggests that TNFα induces inflammatory hyperalgesia through its direct actions on neuronal receptors as well as by increasing the expression of macrophages [123].

The effects of subcutaneously injected TNFα on the mechanical sensitivity and activity of C fibers have also been tested in anesthetized rats [124]. It was found that the 5-ng dose of TNFα decreased the mechanical thresholds of ~67% of the C fibers, and this sensitization began within 30 minutes of the injection and lasted for at least 2 hours. Also, the authors reported that TNFα caused significant plasma protein extravasation that was dose-dependent [124]. Therefore, these studies suggest that TNFα released after tissue injury or disease is involved in the generation of increased pain sensitivity and inflammation, which is observed in patients suffering from inflammatory diseases. Further evidence for TNFα's role in pain conditions comes from studies that have shown strong correlations among tissue levels of TNFα, pain, and hyperalgesia [52, 118, 125–127].

***Human Research.*** Animal research has provided strong evidence for the role of TNFα in mediating inflammatory pain sensitivity and tissue destruction associated with chronic inflammatory conditions, and the successful use of anti-TNFα antibodies in animals has led to their use in human subjects. Various clinical trials

have shown positive results (as discussed) and led to the approval of these agents for use in the treatment of conditions like rheumatoid arthritis.

TNFα has also been thought to be involved in the pathogenesis of other conditions such as migraine headaches, multiple sclerosis, and depression. Recently, a study was conducted to determine the role of TNFα in a migraine. It was found that migraine patients had significantly higher plasma TNFα levels during an attack as compared with headache-free periods, which suggests a potential role for TNFα [178]. Similarly, serum TNFα levels have been reported to be significantly higher in patients suffering from depression and multiple sclerosis than in healthy controls [129]. Also, it has been reported that TNFα may be important in the pathogenesis of temporomandibular disorders (TMDs) and the measurement of TNFα levels in the synovial aspirate may be useful for the diagnosis of temporomandibular joint (TMJ)-related pain conditions [130–133].

## 2 CASE STUDIES

### 2.1 Case Study I: Effect of Intramuscular Injection of NGF in Humans

To investigate the effects of intramuscular injection of NGF on muscle pain, a study was conducted on 12 male volunteers with a mean age of 24.8 years. A baseline measurement of pressure pain threshold (PPT) and pressure tolerance threshold (PTOL) was conducted with a pressure algometer for both the masseter and the temporalis muscles. PPT is a measure of mechanical allodynia and refers to the lowest pressure that the subjects perceived to be painful. PTOL is considered to be a measure of mechanical hyperalgesia and refers to the maximum pressure that the subjects could tolerate. Each subject received two injections (buffered isotonic saline control and ~0.1-μg/kg recombinant human NGF; injection volume = 0.2 mL), one in each masseter muscle. The side and sequence of injections was randomized and blinded. After the first injection, the subjects were asked to rate the pain experienced on an electronic visual analog scale (VAS) of 1 (no pain) to 10 (worst imaginable pain). The pain scores were used to determine the pain intensity felt by the subjects. Fifteen minutes after the first injection, the second injection was made in the contralateral masseter muscle and the pain score measurement was repeated. Subjects were also asked to fill out a pain questionnaire and rate their pain intensity on a 0–10 numerical rating scale (NRS) during activities involving jaw movements such as chewing, yawning, and talking. Both PPT and PTOL measurements were repeated 1, 7, 14, 21, and 28 days after the injections [134].

The results showed that intramuscular injection of NGF did not evoke any significant pain in the masseter muscle, as the VAS scores after the NGF injection were not significantly different from the low VAS scores reported after the control saline injection. There was no significant difference in the baseline PPT and PTOL values in either temporalis or masseter muscles. The NGF injection caused a decrease in the PPT values (as compared with the baseline) in the masseter muscle, with significantly lower PPTs at 1 and 7 days after injection (Figure 1). The saline injection also caused a significant decrease in PPTs 1 day after injection; however, the PPT values increased significantly above baseline after 14 and 21 days. At 1 day after injection, the mean PPT value in the NGF-injected muscles was significantly lower than that in the saline-injected muscles (Figure 1). The PPT

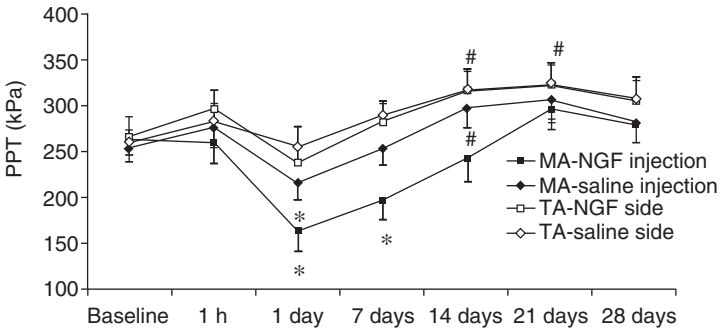


**Figure 1.** PPT measured in four test muscles for 28 days. NGF or isotonic saline was injected into the masseter (MA) muscle. Temporalis (TA) muscle PPTs were also assessed bilaterally on the side of NGF and saline injection. Each point indicates mean ± SEM ($n$ = 12). The symbol (*) indicates significantly lower and (#) significantly higher compared with baseline (Dunnett: $p < 0.05$). (Reproduced from Ref. 134, with the permission of the International Association for the Study of Pain.)

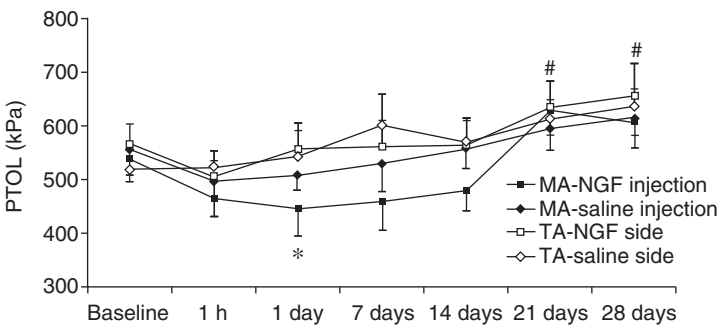


**Figure 2.** Pressure pain tolerances (PTOL) measured on four test muscles over 28 days. NGF or isotonic saline was injected into the masseter (MA) muscle. Temporalis (TA) muscle PPTs were also assessed bilaterally on the side of NGF and saline injection. Each point indicates mean ± SEM ($n$ = 12). The symbol (*) indicates significantly lower and (#) significantly higher compared with baseline ($p < 0.05$). (Reproduced from Ref. 134, with the permission of the International Association for the Study of Pain.)

values in both temporalis muscles increased after (control saline or NGF) injections in the masseter muscles. The NGF injection also caused a decrease in the PTOL values, with significantly lower PTOL values at 1 day (compared with baseline) (Figure 2). However, at 21 days, the mean PTOL value in NGF-injected muscles increased significantly above the baseline value. The control saline injection did not cause any significant change in the PTOL values. The PPT values of the temporalis muscle ipsilateral to saline-injected masseter muscle were significantly higher at 21 and 28 days after injection. However, there was no significant change in the PPTs of the temporalis muscle ipsilateral to the NGF-injected masseter muscle. Subjects reported pain upon chewing and yawning in the NGF-injected masseter muscle, whereas activities such as talking, swallowing, and smiling did not evoke pain in the NGF-injected masseter muscle (Figure 3) [134].

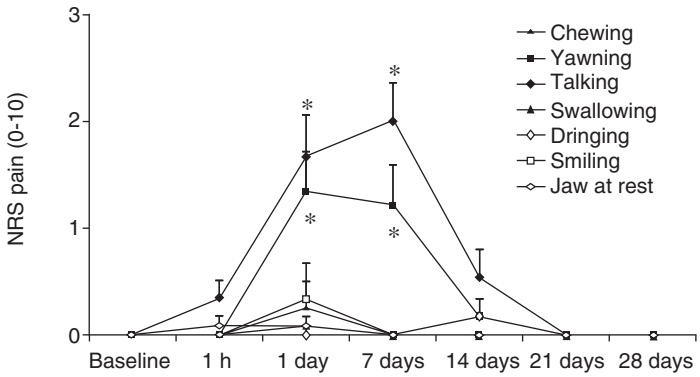


**Figure 3.** Perceived pain intensity measured on a 0–10 NRS in response to various jaw functions before and after injection of NGF into the masseter muscle. Each point indicates mean ± SEM ($n = 12$). The symbol (*) indicates significantly higher NRS pain scores compared with baseline (Dunnett: $p < 0.05$). (Reproduced from Ref. 134, with the permission of the International Association for the Study of Pain.)

The results of this study suggested that intramuscular injection of NGF did not evoke any significant pain but caused mechanical sensitization at the site of injection in the masseter muscle that lasted for at least 7 days and resulted in TMD-like symptoms. A significant reduction in the PPT and PTOL values in the NGF-injected muscles indicated that NGF can cause mechanical allodynia and hyperalgesia. Also, a significant number of subjects reported pain during oral functions such as chewing and yawning, which is also a characteristic symptom of TMD. The authors suggested that intramuscular administration of NGF into the masseter muscle could be used to model TMD-like pain and sensitization and to further study the pathophysiology of TMD. In addition, the authors commented that it was not clear whether NGF was causing mechanical sensitization through a peripheral or central mechanism [134]. Therefore, the mechanisms responsible for NGF-induced sensitization need to be further investigated.

## 2.2 Case Study II: NGF Effects on Masticatory Muscle Nociceptors

An animal study was undertaken to determine whether NGF significantly alters the masseter muscle afferent fiber mechanical threshold as part of the mechanism whereby it decreases PPTs in human subjects. In this study, *in vivo* single afferent neuron recordings were conducted on anesthetized, adult female Sprague–Dawley rats ($n = 17$). A recording electrode was lowered into the trigeminal ganglion to record orthodromic action potentials of afferent fibers activated by applying mechanical force to the masseter muscle. A blunt mechanical search stimulus was used to find afferent fibers in the rat masseter muscle. An antidromic collision technique [135] was used to confirm the projection of a fiber to the central nervous system (Figure 4). The mechanical threshold (MT) of each masseter muscle afferent fiber was assessed with an electronic von Frey (VF) hair (Figure 5). For each fiber, a baseline MT recording was conducted every minute for 10 minutes. Each primary afferent fiber was assigned to one of the following two groups: vehicle

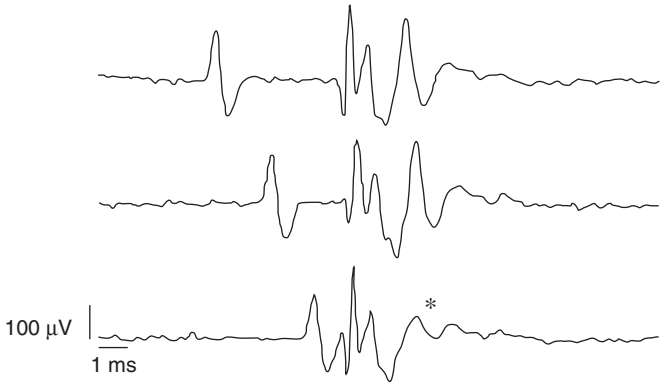


**Figure 4.** An example of a collision between an orthodromic (activated by mechanical stimulation of the masseter muscle) and an antidromic (activated by electrical stimulation of the caudal brain stem) is illustrated. *Collision resulted in the disappearance of the antidromic spike. This method was used to confirm the projection of the masseter afferent fiber from the masseter muscle to the caudal brain stem, a region believed to be responsible for the integration of muscle pain input.

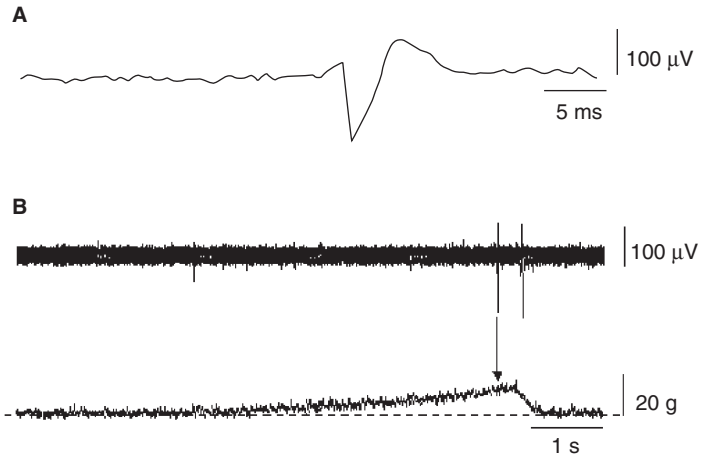


**Figure 5.** (**A**) The shape of a single action potential is shown. (**B**) The top trace shows the action potentials recorded from the trigeminal ganglion. The bottom trace shows an increase in the force applied with an electronic von Frey (VF) hair on the masseter muscle until the fiber started firing action potentials. The threshold of the fiber (7.71 g) was determined by subtracting the baseline from the minimum force required to activate the fiber.

control (10 μL of phosphate-buffered isotonic saline and albumin, $n$ = 10) or 25-μg/mL human NGF (1 μg/kg, respectively, 10 μL, $n$ = 7) dissolved in phosphate-buffered isotonic saline and albumin. After the baseline recording, an injection was made into the masseter muscle and the evoked response was recorded for 10 minutes. Sixty minutes after the injection, 10 consecutive MT recordings were conducted at 1-minute intervals. These recordings were repeated every hour for a total of 3 hours. Also, the effect of NGF on plasma protein extravasation in the masseter muscle was determined.

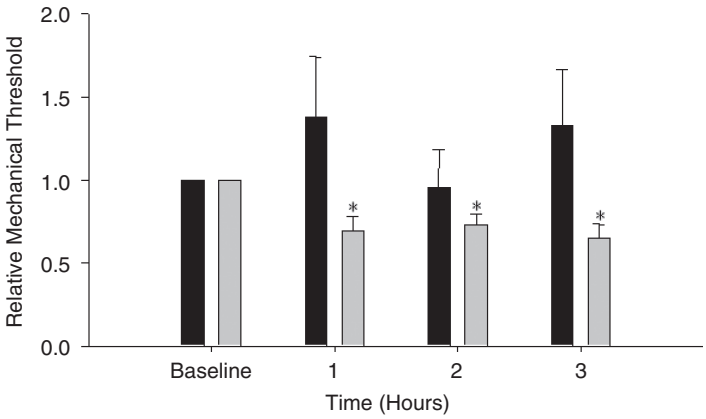


**Figure 6.** The vertical bar graph illustrates the mean relative mechanical threshold of afferent fibers at various time points after a 10-μL injection of phosphate-buffered saline control (black bars, $n = 10$) or 25-μg/mL human NGF (gray bars, $n = 7$) into the masseter muscle. The error bars represent the standard error of the mean. $*p < 0.05$: ANOVA and post hoc Holm–Sidak method compared with the baseline relative mechanical threshold.

Intramuscular injection of human NGF into the rat masseter muscle failed to evoke afferent discharge; however, it decreased the MT of masseter afferent fibers compared with the baseline (Figure 6). There was a significant decrease in the MT 1 hour postinjection (as compared with the baseline), which lasted for 3 hours. The postinjection MT of saline-injected fibers was not significantly different from the baseline. There was no significant difference between the level of plasma protein extravasation between control saline and human NGF groups.

The injection of human NGF caused mechanical sensitization that lasted for at least 3 hours; however, there was no indication that human NGF produced these effects indirectly through local muscle inflammation. The peripheral effects of NGF are likely mediated through a receptor mechanism, for example, TrkA receptor mediated pathways [136]. It is possible that nonspecific mechanisms, such as foreign protein reactions, decreased degradation of human NGF by rat proteases, slowed clearance from the muscle, modulation of intracellular calcium concentration (calcium uptake) [137, 138], actions on bradykinin receptors [139], capsaicin receptors [37, 140], or sodium channels [141, 142] might contribute to the effect of human NGF in rats.

The results of this study suggested that intramuscular injection of human NGF at a concentration equivalent to that which resulted in mechanical sensitization in humans [134] caused a significant decrease in the mechanical threshold of rat masseter muscle afferent fibers. Therefore, the method of injecting human NGF into the rat masseter muscle could be used to study peripheral mechanisms associated with the development of NGF-induced masseter muscle sensitization in humans.

## REFERENCES

1. Pasero C (2004). Pathophysiology of neuropathic pain. *Pain Manage. Nursing.* 5:3–8.

2. Mense S (1993). Nociception from skeletal muscle in relation to clinical muscle pain. *Pain*. 54:241–289.
3. DeLeo J A, Winkelstein B A (2002). Physiology of chronic spinal pain syndromes: From animal models to biomechanics. *Spine*. 27:2526–2537.
4. White F A, Bhangoo S K, Miller R J (2005). Chemokines: integrators of pain and inflammation. *Nat. Rev. Drug Disc*. 4:834–844.
5. Willis W D (1998). The somatosensory system. In R M Berne, M N Levy, B M Koeppen, et al. (eds.), *Physiology*: Mosby, Inc., St. Louis, MD, pp. 109–128.
6. Dmitrieva N, McMahon S B (1996). Sensitisation of visceral afferents by nerve growth factor in the adult rat. *Pain*. 66:87–97.
7. Meyer R A, Campbell J N, Raja S N (1994). Peripheral neural mechanisms of nociception. In P D Wall, R Melzack (eds.), *Textbook of Pain*: Churchill Livingstone, Edinburgh, Scotland, pp. 13–44.
8. Almeida T F, Roizenblatt S, Tufik S (2004). Afferent pain pathways: A neuroanatomical review. *Brain Res*. 1000:40–56.
9. McMahon S B, Lewin G R, Wall P D (1993). Central hyperexcitability triggered by noxious inputs. *Curr. Opin. Neurobiol*. 3:602–610.
10. Ren K, Dubner R (1999). Central nervous system plasticity and persistent pain. *J. Orofacial Pain*. 13:155–163.
11. Sessle B J (2000). Acute and chronic craniofacial pain: Brainstem mechanisms of nociceptive transmission and neuroplasticity, and their clinical correlates. *Crit. Rev. Oral Biol. Med*. 11:57–91.
12. Woolf C J (1991). Generation of acute pain: central mechanisms. *British Med. Bull*. 47:523–533.
13. Dworkin R H, Backonja M, Rowbotham M C, et al. (2003). Advances in neuro-pathic pain: diagnosis, mechanisms, and treatment recommendations. *Archives Neurol*. 60:1524–1534.
14. Harden R N (2005). Chronic neuropathic pain. Mechanisms, diagnosis, and treatment. *Neurolog*. 11:111–122.
15. Hansson P T, Dickenson A H (2005). Pharmacological treatment of peripheral neuropathic pain conditions based on shared commonalities despite multiple etiologies. *Pain*. 113:251–254.
16. Ohtori S, Takahashi K, Moriya H, et al. (2004). TNF-alpha and TNF-alpha receptor type 1 upregulation in glia and neurons after peripheral nerve injury: studies in murine DRG and spinal cord. *Spine*. 29:1082–1088.
17. Schafers M, Geis C, Svensson C I, et al. (2003). Selective increase of tumour necrosis factor-alpha in injured and spared myelinated primary afferents after chronic constrictive injury of rat sciatic nerve. *Euro. J. Neurosci*. 17:791–804.
18. Shubayev V I, Myers R R (2000). Upregulation and interaction of TNFalpha and gelatinases A and B in painful peripheral nerve injury. *Brain Res*. 855:83–89.
19. Lindenlaub T, Teuteberg P, Hartung T, et al. (2000). Effects of neutralizing antibodies to TNF-alpha on pain-related behavior and nerve regeneration in mice with chronic constriction injury. *Brain Res*. 866:15–22.
20. Raghavendra V, Tanga F, DeLeo J A (2003). Inhibition of microglial activation attenuates the development but not existing hypersensitivity in a rat model of neuropathy. *J. Pharmacol. Exper. Therapeut*. 306:624–630.
21. Sommer C, Lindenlaub T, Teuteberg P, et al. (2001). Anti-TNF-neutralizing antibodies reduce pain-related behavior in two different mouse models of painful mononeuropathy. *Brain Res*. 913:86–89.

22. Aloe L (2004). Rita Levi-Montalcini: the discovery of nerve growth factor and modern neurobiology. *Trends Cell Biol.* 14:395–399.
23. Levi-Montalcini R (1987). The nerve growth factor 35 years later. *Sci.* 237:1154–1162.
24. Huang E J, Reichardt L F (2001). Neurotrophins: roles in neuronal development and function. *Ann. Rev. Neurosci.* 24:677–736.
25. Holland D R, Cousens L S, Meng W, et al. (1994). Nerve growth factor in different crystal forms displays structural flexibility and reveals zinc binding sites. *J. Molec. Biol.* 239:385–400.
26. McDonald N Q, Lapatto R, Murray-Rust J, et al. (1991). New protein fold revealed by a 2.3-A resolution crystal structure of nerve growth factor. *Nature.* 354:411–414.
27. Aloe L, Bracci-Laudiero L, Bonini S, et al. (1997). The expanding role of nerve growth factor: From neurotrophic activity to immunologic diseases. *Allergy.* 52:883–894.
28. Owolabi J B, Rizkalla G, Tehim A, et al. (1999). Characterization of Antiallodynic Actions of ALE-0540, a Novel Nerve Growth Factor Receptor Antagonist, in the Rat. *J. Pharmacol. Exper. Therapeut.* 289:1271–1276.
29. Aloe L, Allevab E, Fiore M (2002). Stress and nerve growth factor: Findings in animal models and humans. *Pharmacol. Biochem. Behavior.* 73:159–166.
30. Andreev N Y, Dimitrieva N, Koltzenburg M, et al. (1995). Peripheral administration of nerve growth factor in the adult rat produces a thermal hyperalgesia that requires the presence of sympathetic post-ganglionic neurons. *Pain.* 63:109–115.
31. Bracci-Laudiero L, Aloe L, Buanne P, et al. (2002). NGF modulates CGRP synthesis in human B-lymphocytes: a possible anti-inflammatory action of NGF? *J. Neuroimmunol.* 123:58–65.
32. de VRIES A, Dessing M C, Engels F, et al. (1999). Nerve Growth Factor Induces a Neurokinin-1 Receptor-Mediated Airway Hyperresponsiveness in Guinea Pigs. *Amer. J. Respir. Criti. Care Med.* 159:1541–1544.
33. Chao M V, Hempstead B L (1995). p75 and Trk: A two-receptor system. *Trends Neurosci.* 18:321–326.
34. Woolf C J, Ma Q P, Allchorne A, et al. (1996). Peripheral cell types contributing to the hyperalgesic action of nerve growth factor in inflammation. *J. Neurosci.* 16:2716–2723.
35. Chao M V (2003). Neurotrophins and their receptors: A convergence point for many signaling pathways. *Neurosci.* 4:299–309.
36. York R D, Molliver D C, Grewal S S, et al. (2000). Role of phosphoinositide 3-kinase and endocytosis in nerve growth factor-induced extracellular signal-regulated kinase activation via Ras and Rap1. *Molec. Cell. Biol.* 20:8069–8083.
37. Chuang H, Prescott E D, Kong, H, et al. (2001). Bradykinin and nerve growth factor release the capsaicin receptor from PtdIns(4,5)P2-mediated inhibition. *Nat. Rev.* 411:957–962.
38. Humpel C, Ebendal T, Olson L (1996). Microdialysis: a way to study in vivo release of neurotrophic bioactivity: a critical summary. *J. Molec. Med.* 74:523–526.
39. Winter C D, Iannotti F, Pringle A K, et al. (2002). A microdialysis method for the recovery of IL-1beta, IL-6 and nerve growth factor from human brain in vivo. *J. Neurosci. Meth.* 119:45–50.
40. Martocchia A, Sigala S, Proietti A, et al. (2002). Sex-related variations in serum nerve growth factor concentration in humans. *Neuropept.* 36:391–395.
41. Carswell E A, Old L J, Kassel R L, et al. (1975). An endotoxin-induced serum factor that causes necrosis of tumors. *Proc. Nat. Acad. Sci. USA.* 72:3666–3670.

42. Allan S M, Rothwell N J (2001). Cytokines and acute neurodegeneration. *Nat. Rev. Neurosci.* 2:734–744.
43. Rothwell N J, Luheshi G, Toulmond S (1996). Cytokines and their receptors in the central nervous system: physiology, pharmacology, and pathology. *Pharmacol. Therapeut.* 69:85–95.
44. Darnay B G, Aggarwal B B (1999). Signal transduction by tumour necrosis factor and tumour necrosis factor related ligands and their receptors. *Annals Rheumat. Dis.* 58(suppl 1):I2–I13.
45. Vilcek J, Lee T H (1991). Tumor necrosis factor. New insights into the molecular mechanisms of its multiple actions. *J. Biol. Chem.* 266:7313–7316.
46. Sherry B, Cerami A (1988). Cachectin/tumor necrosis factor exerts endocrine, paracrine, and autocrine control of inflammatory responses. *J. Cell Biol.* 107:1269–1277.
47. Burger D, Dayer J M (2002). Cytokines, acute-phase proteins, and hormones: IL-1 and TNF-alpha production in contact-mediated activation of monocytes by T lymphocytes. *Annals NY Acad. Sci.* 966:464–473.
48. Van Deventer S J (1997). Tumour necrosis factor and Crohn's disease. *Gut.* 40:443–448.
49. Heller R A, Kronke M (1994). Tumor necrosis factor receptor-mediated signaling pathways. *J. Cell Bio.* 126:5–9.
50. Allan S M (2000). The role of pro- and antiinflammatory cytokines in neurodegeneration. *Annals NY Acad. Sci.* 917:84–93.
51. Dinarello C A (2000). Proinflammatory cytokines. *Chest.* 118:503–508.
52. Sommer C, Kress M (2004). Recent findings on how proinflammatory cytokines cause pain: peripheral mechanisms in inflammatory and neuropathic hyperalgesia. *Neurosci. Lett.*, 361:184–187.
53. Steiner P, Pfeilschifter J, Boeckh C, et al. (1991). Interleukin-1 beta and tumor necrosis factor-alpha synergistically stimulate nerve growth factor synthesis in rat mesangial cells. *Amer. J. Physiol.* 261:F792–798.
54. Brenner D A, O'Hara M, Angel P, et al. (1989). Prolonged activation of jun and collagenase genes by tumour necrosis factor-alpha. *Nature.* 337:661–663.
55. Tracey K J, Fong Y, Hesse D G, et al. (1987). Anti-cachectin/TNF monoclonal antibodies prevent septic shock during lethal bacteraemia. *Nature.* 330:662–664.
56. Hesse D G, Tracey K J, Fong Y, et al. (1988). Cytokine appearance in human endotoxemia and primate bacteremia. *Surgery Gynecol. Obstet.* 166:147–153.
57. Chaldakov G N, Fiore M, Hristova M G, et al. (2003). Metabotrophic potential of neurotrophins: implication in obesity and related diseases? *Med. Sci. Monitor.* 9: HY19–21.
58. Hellweg R, Hartung H D (1990). Endogenous levels of nerve growth factor (NGF) are altered in experimental diabetes mellitus: A possible role for NGF in the pathogenesis of diabetic neuropathy. *J. Neurosci. Res.* 26:258–267.
59. Leinninger G M, Vincent A M, Feldman E L (2004). The role of growth factors in diabetic peripheral neuropathy. *J. Periph. Nerv. Syst.* 9:26–53.
60. Brewster W J, Fernyhough P, Diemel L T, et al. (1994). Diabetic neuropathy, nerve growth factor and other neurotrophic factors. *Trends Neurosci.* 17:321–325.
61. Fernyhough P, Diemel L T, Brewster W J, et al. (1994). Deficits in sciatic nerve neuropeptide content coincide with a reduction in target tissue nerve growth factor messenger RNA in streptozotocin-diabetic rats: effects of insulin treatment. *Neurosci.* 62:337–344.

62. Anand P, Terenghi G, Warner G, et al. (1996). The role of endogenous nerve growth factor in human diabetic neuropathy. *Nature Med.* 2:703–707.
63. Hellweg R, Raivich G, Hartung H D, et al. (1994). Axonal transport of endogenous nerve growth factor (NGF) and NGF receptor in experimental diabetic neuropathy. *Exper. Neurol.* 130:24–30.
64. Jakobsen J, Brimijoin S, Skau K, et al. (1981). Retrograde axonal transport of transmitter enzymes, fucose-labeled protein, and nerve growth factor in streptozotocin-diabetic rats. *Diabetes.* 30:797–803.
65. Apfel S C (2002). Nerve growth factor for the treatment of diabetic neuropathy: what went wrong, what went right, and what does the future hold? *Internat. Rev. Neurobio.* 50:393–413.
66. Apfel S C, Arezzo J C, Brownlee M, et al. (1994). Nerve growth factor administration protects against experimental diabetic sensory neuropathy. *Brain Res.* 634:7–12.
67. Fernyhough P, Diemel L T, Hardy J, et al. (1995). Human recombinant nerve growth factor replaces deficient neurotrophic support in the diabetic rat. *Euro. J. Neurosci.* 7:1107–1110.
68. Apfel S C, Kessler J A, Adornato B T, et al. (1998). Recombinant human nerve growth factor in the treatment of diabetic polyneuropathy. NGF Study Group. *Neurol.* 51:695–702.
69. Petty B G, Cornblath D R, Adornato B T, et al. (1994). The effect of systemically administered recombinant human nerve growth factor in healthy human subjects. *Ann. Neurol.* 36:244–246.
70. Apfel S C, Schwartz S, Adornato B T, et al. (2000). Efficacy and safety of recombinant human nerve growth factor in patients with diabetic polyneuropathy: A randomized controlled trial. rhNGF Clinical Investigator Group. *J. Amer. Med. Associ.* 284:2215–2221.
71. Deng Y S, Zhong J H, Zhou X F (2000). Effects of endogenous neurotrophins on sympathetic sprouting in the dorsal root ganglia and allodynia following spinal nerve injury. *Exper. Neurol.* 164:344–350.
72. Li L, Xian C J, Zhong J H, et al. (2003). Lumbar 5 ventral root transection-induced upregulation of nerve growth factor in sensory neurons and their target tissues: a mechanism in neuropathic pain. *Molec. Cell. Neurosci.*, 23:232–250.
73. Ramer M S, French G D, Bisby M A (1997). Wallerian degeneration is required for both neuropathic pain and sympathetic sprouting into the DRG. *Pain.*, 72:71–78.
74. Ruiz G, Ceballosc D, Banos J-E (2004). Behavioral and histological effects of endoneurial administration of nerve growth factor: Possible implications in neuropathic pain. *Brain Res.* 1011:1–6.
75. Oh E J, Yoon Y W, Lee S E, et al. (2000). Changes in nerve growth factor levels in dorsal root ganglia and spinal nerves in a rat neuropathic pain model. *Exper. Brain Res.* 130:93–99.
76. Heumann R, Korsching S, Bandtlow C, et al. (1987). Changes of nerve growth factor synthesis in nonneuronal cells in response to sciatic nerve transection. *J. Cell Biol.* 104:1623–1631.
77. Theodosiou M, Rush R A, Zhou X F, et al. (1999). Hyperalgesia due to nerve damage: role of nerve growth factor. *Pain.* 81:245–255.
78. Anand P (2004). Neurotrophic factors and their receptors in human sensory neuropathies. *Prog. Brain Res.* 146:477–492.
79. Ren K, Thomas D A, Dubner R (1995). Nerve growth factor alleviates a painful peripheral neuropathy in rats. *Brain Res.* 699:286–292.

80. Cahill C M, Dray A, Coderre T J (2003). Intrathecal nerve growth factor restores opioid effectiveness in an animal model of neuropathic pain. *Clin. Neuropharmacol.* 45:543–552.
81. Bielefeldt K, Ozaki N, Gebhart G F (2003). Role of nerve growth factor in modulation of gastric afferent neurons in the rat. *Amer. J. Phys. Gastrointest. Liver Physio.* 284: G499–G507.
82. Djouhri L, Dawbarn D, Robertson A, et al. (2001). Time course and nerve growth factor dependence of inflammation-induced alterations in electrophysiological membrane properties in nociceptive primary afferent neurons. *J. Neurosci.* 21:8722–8733.
83. Woolf C J, Safieh-Garabedian B, Ma Q P, et al. (1994). Nerve growth factor contributes to the generation of inflammatory sensory hypersensitivity. *Neurosci.* 62:327–331.
84. Aloe L, Tuveri M A, Carcassi U, et al. (1992). Nerve growth factor in the synovial fluid of patients with chronic arthritis. *Arthritis Rheumat.* 35:351–355.
85. Woolf C J (1996). Phenotypic modification of primary sensory neurons: the role of nerve growth factor in the production of persistent pain. *Philosoph. Trans. Royal Soc. London. Series B, Biolog. Sci.* 351:441–448.
86. Lindsay R M, Harmar A J (1989). Nerve growth factor regulates expression of neuropeptide genes in adult sensory neurons. *Nature Rev.* 337:362–364.
87. Benito M J, Veale D J, FitzGerald O, et al. (2005). Synovial tissue inflammation in early and late osteoarthritis. *Annals Rheumatic Dis.* 64:1263–1267.
88. Feldmann M (1996). The cytokine network in rheumatoid arthritis: definition of TNF alpha as a therapeutic target. *J. Royal Coll. Physicians of London.* 30:560–570.
89. Fiers W (1991). Tumor necrosis factor. Characterization at the molecular, cellular and in vivo level. *FEBS Lett.* 285:199–212.
90. Brahn E, Peacock D J, Banquerigo M L, et al. (1992). Effects of tumor necrosis factor alpha (TNF-alpha) on collagen arthritis. *Lymphokine Cytokine Res.* 11:253–256.
91. Cooper W O, Fava R A, Gates C A, et al. (1992). Acceleration of onset of collagen-induced arthritis by intra-articular injection of tumour necrosis factor or transforming growth factor-beta. *Clin. Experim. Immunol.* 89:244–250.
92. Piguet P F, Grau G E, Vesin C, et al. (1992). Evolution of collagen arthritis in mice is arrested by treatment with anti-tumour necrosis factor (TNF) antibody or a recombinant soluble TNF receptor. *Immunol.* 77:510–514.
93. Thorbecke G J, Shah R, Leu C H, et al. (1992). Involvement of endogenous tumor necrosis factor alpha and transforming growth factor beta during induction of collagen type II arthritis in mice. *Proc. Nat. Acad. Sci. USA.* 89:7375–7379.
94. Williams R O, Feldmann M, Maini R N (1992). Anti-tumor necrosis factor ameliorates joint disease in murine collagen-induced arthritis. *Proc. Nat. Acad. Sci. USA.* 89:9784–9788.
95. Elliott M J, Maini R N, Feldmann M, et al. (1994). Randomised double-blind comparison of chimeric monoclonal antibody to tumour necrosis factor alpha (cA2) versus placebo in rheumatoid arthritis. *Lancet.* 344:1105–1110.
96. Elliott M J, Maini R N, Feldmann M, et al. (1994). Repeated therapy with monoclonal antibody to tumour necrosis factor alpha (cA2) in patients with rheumatoid arthritis. *Lancet.* 344:1125–1127.
97. Elliott M J, Maini R N, Feldmann M, et al. (1993). Treatment of rheumatoid arthritis with chimeric monoclonal antibodies to tumor necrosis factor alpha. *Arthrit. Rheumat.* 36:1681–1690.

98. Nash P T, Florin T H (2005). Tumour necrosis factor inhibitors. *Med. J. Australia.* 183:205–208.
99. Flagg S D, Meador R, Hsia E, et al. (2005). Decreased pain and synovial inflammation after etanercept therapy in patients with reactive and undifferentiated arthritis: An open-label trial. *Arthrit. Rheumatism.* 53:613–617.
100. Moreland L W, Schiff M H, Baumgartner S W, et al. (1999). Etanercept therapy in rheumatoid arthritis. A randomized, controlled trial. *Ann. Intern. Med.* 130:478–486.
101. Weinblatt M E, Kremer J M, Bankhurst A D, et al. (1999). A trial of etanercept, a recombinant tumor necrosis factor receptor:Fc fusion protein, in patients with rheumatoid arthritis receiving methotrexate. *New Eng. J. Med.* 340:253–259.
102. Tyring S, Gottlieb A, Papp K, et al. (2006). Etanercept and clinical outcomes, fatigue, and depression in psoriasis: double-blind placebo-controlled randomised phase III trial. *Lancet.* 367:29–35.
103. Keystone E C, Kavanaugh A F, Sharp J T, et al. (2004). Radiographic, clinical, and functional outcomes of treatment with adalimumab (a human anti-tumor necrosis factor monoclonal antibody) in patients with active rheumatoid arthritis receiving concomitant methotrexate therapy: a randomized, placebo-controlled, 52-week trial. *Arthrit. Rheumatism.* 50:1400–1411.
104. Weisman M H, Moreland L W, Furst D E, et al. (2003). Efficacy, pharmacokinetic, and safety assessment of adalimumab, a fully human anti-tumor necrosis factor-alpha monoclonal antibody, in adults with rheumatoid arthritis receiving concomitant methotrexate: A pilot study. *Clin. Therapeut.* 25:1700–1721.
105. Lewin G R, Ritter A M, Mendell L M (1993). Nerve growth factor-induced hyperalgesia in the neonatal and adult rat. *J. Neurosci.* 13:2136–2148.
106. Rueff A, Dawson A J, Mendell L M (1996). Characteristics of nerve growth factor induced hyperalgesia in adult rats: Dependence on enhanced bradykinin-1 receptor activity but not neurokinin-1 receptor activation. *Pain.* 66:359–372.
107. Rueff A, Mendell L M (1996). Nerve Growth Factor and NT-5 induce increased thermal sensitivity of cutaneous nociceptors in vitro. *J. Neurophysiol.* 76:3596–3596.
108. Thompson S W N, Dray A, McCarson K E, et al. (1995). Nerve growth factor induces mechanical allodynia associated with novel A fibre-evoked spinal reflex activity and enhanced neurokinin-1 receptor activation in the rat. *Pain.* 62:219–231.
109. Lewin G R, Rueff A, Mendell L M (1994). Peripheral and central mechanisms of NGF-induced hyperalgesia. *Euro. J. Neurosci.* 6:1903–1912.
110. Dyck P J, Peroutka S, Rask C, et al. (1997). Intradermal recombinant human nerve growth factor induces pressure allodynia and lowered heat-pain threshold in humans. *Neurol.* 48:501–505.
111. Cunha F Q, Poole S, Lorenzetti B B, et al. (1992). The pivotal role of tumour necrosis factor alpha in the development of inflammatory hyperalgesia. *British J. Pharmacol.* 107:660–664.
112. Sachs D, Cunha F Q, Poole S, et al. (2002). Tumour necrosis factor-alpha, interleukin-1beta and interleukin-8 induce persistent mechanical nociceptor hypersensitivity. *Pain.* 96:89–97.
113. Perkins M N, Kelly D, Davis A J (1995). Bradykinin B1 and B2 receptor mechanisms and cytokine-induced hyperalgesia in the rat. *Canadian J. Physiol. Pharmacol.* 73:832–836.
114. Safieh-Garabedian B, Poole S, Allchorne A, et al. (1995). Contribution of interleukin-1 beta to the inflammation-induced increase in nerve growth factor levels and inflammatory hyperalgesia. *British J. Pharmacol.* 115:1265–1275.

115. Opree A, Kress M (2000). Involvement of the proinflammatory cytokines tumor necrosis factor-alpha, IL-1 beta, and IL-6 but not IL-8 in the development of heat hyperalgesia: effects on heat-evoked calcitonin gene-related peptide release from rat skin. *J. Neurosci.* 20:6289–6293.
116. Sommer C, Schmidt C, George A (1998). Hyperalgesia in experimental neuropathy is dependent on the TNF receptor 1. *Exper. Neurol.* 151:138–142.
117. Jin X, Gereau R W T (2006). Acute p38-mediated modulation of tetrodotoxin-resistant sodium channels in mouse sensory neurons by tumor necrosis factor-alpha. *J. Neurosc.* 26:246–255.
118. Tak P P, Smeets T J, Daha M R, et al. (1997). Analysis of the synovial cell infiltrate in early rheumatoid synovial tissue in relation to local disease activity. *Arthrit. Rheumatism.* 40:217–225.
119. Cunha T M, Verri W A, Jr., Silva J S, et al. (2005). A cascade of cytokines mediates mechanical inflammatory hypernociception in mice. *Proc. Nat. Acad. Sci. USA.* 102:1755–1760.
120. Frode T S, Souza G E, Calixto J B (2001). The modulatory role played by TNF-alpha and IL-1 beta in the inflammatory responses induced by carrageenan in the mouse model of pleurisy. *Cytokine.* 13:162–168.
121. Tonussi C R, Ferreira S H (1999). Tumour necrosis factor-alpha mediates carrageenin-induced knee-joint incapacitation and also triggers overt nociception in previously inflamed rat knee-joints. *Pain.* 82:81–87.
122. Kuiper S, Joosten L A, Bendele A M, et al. (1998). Different roles of tumour necrosis factor alpha and interleukin 1 in murine streptococcal cell wall arthritis. *Cytokine.* 10:690–702.
123. Inglis J J, Nissim A, Lees D M, et al. (2005). The differential contribution of tumour necrosis factor to thermal and mechanical hyperalgesia during chronic inflammation. *Arthrit. Res. Ther.* 7:R807–816.
124. Junger H, Sorkin L S (2000). Nociceptive and inflammatory effects of subcutaneous TNFalpha. *Pain.* 85:145–151.
125. Barnes P F, Chatterjee D, Brennan P J, et al. (1992). Tumor necrosis factor production in patients with leprosy. *Infec. Immunity.* 60:1441–1446.
126. Lindenlaub T, Sommer C (2003). Cytokines in sural nerve biopsies from inflammatory and non-inflammatory neuropathies. *Acta Neuropathol.* 105:593–602.
127. Shafer D M, Assael L, White L B, et al. (1994). Tumor necrosis factor-alpha as a biochemical marker of pain and outcome in temporomandibular joints with internal derangements. *J. Oral Maxillofac Surg.* 52:786–791.
128. Perini F, D'Andrea G, Galloni E, et al. (2005). Plasma cytokine levels in migraineurs and controls. *Headache.* 45:926–931.
129. Mikova O, Yakimova R, Bosmans E, et al. (2001). Increased serum tumor necrosis factor alpha concentrations in major depression and multiple sclerosis. *Euro. Neuropsychopharmacol.* 11:203–208.
130. Emshoff R, Puffer P, Rudisch A, et al. (2000). Temporomandibular joint pain: relationship to internal derangement type, osteoarthrosis, and synovial fluid mediator level of tumor necrosis factor-alpha. *Oral Surg., Oral Med., Oral Pathol., Oral Radiol., Endodon.* 90:442–449.
131. Emshoff R, Puffer P, Strobl H, et al. (2000). Effect of temporomandibular joint arthrocentesis on synovial fluid mediator level of tumor necrosis factor-alpha: implications for treatment outcome. *Internat. J. Oral Maxillofacial Surg.* 29:176–182.
132. Kaneyama K, Segami N, Sun W, et al. (2005). Analysis of tumor necrosis factor-alpha, interleukin-6, interleukin-1beta, soluble tumor necrosis factor receptors I and II,

interleukin-6 soluble receptor, interleukin-1 soluble receptor type II, interleukin-1 receptor antagonist, and protein in the synovial fluid of patients with temporomandibular joint disorders. *Oral Surg., Oral Med., Oral Pathol., Oral Radiol., Endodon.*, 99:276–284.

133. Kaneyama K, Segami N, Sun W, et al. (2005). Levels of soluble cytokine factors in temporomandibular joint effusions seen on magnetic resonance images. *Oral Surg., Oral Med., Oral Pathol., Oral Radiol., Endodont.* 99:411–418.
134. Svensson P, Cairns B E, Wang K, et al. (2003). Injection of nerve growth factor into human masseter muscle evokes long lasting mechanical allodynia and hyperalgesia. *Pain.* 104:241–247.
135. Cairns B E, Gambarota G, Svensson P, et al. (2002). Glutamate-induced sensitization of rat masseter muscle fibers. *Neurosci.* 109:389–399.
136. Malik-Hall M, Dina O A, Levine J D (2005). Primary afferent nociceptor mechanisms mediating NGF-induced mechanical hyperalgesia. *Euro. J. Neurosci.* 21:3387–3394.
137. Jiang H, Guroff G (1997). Actions of the neurotrophins on calcium uptake. *J. Neurosc. Res.* 50:355–360.
138. Jiang H, Takeda K, Lazarovici P, et al. (1999). Nerve growth factor (NGF)-induced calcium influx and intracellular calcium mobilization in 3T3 cells expressing NGF receptors. *J. Biolog. Chem.* 274:26209–26216.
139. Dray A, Perkins M (1993). Bradykinin and inflammatory pain. *Trends Neurosci.* 16:99–104.
140. Zhu W, Galoyan S M, Petruska J C, et al. (2004). A developmental switch in acute sensitization of small dorsal root ganglion (DRG) neurons to capsaicin or noxious heating by NGF. *J. Neurophysiol.* 92:3148–3152.
141. Fjell J, Cummins T R, Dib-Hajj S D, et al. (1999). Differential role of GDNF and NGF in the maintenance of two TTX-resistant sodium channels in adult DRG neurons. *Brain Res. Molec. Brain Res.* 67:267–282.
142. Fjell J, Cummins T R, Fried K, et al. (1999). In vivo NGF deprivation reduces SNS expression and TTX-R sodium currents in IB4-negative DRG neurons. *J. Neurophysiol.* 81:803–810.

# GROWTH FACTORS, CYTOKINES, AND CHEMOKINES: FORMULATION, DELIVERY, AND PHARMACOKINETICS

Heping Cao and Rui Lin
*Peace Technology Development, North Potomac, Maryland*

**Chapter Contents**

1 Types of Growth Factors, Cytokines, and Chemokines
2 Functions of Growth Factors, Cytokines, and Chemokines
3 Some Drugs Involving Growth Factors, Cytokines, and Chemokines
4 Formulations of Growth Factors, Cytokines, and Chemokines
5 Delivery of Growth Factors, Cytokines, and Chemokines
6 Pharmacokinetics of Growth Factors, Cytokines, and Chemokines
7 Potential Applications of Tristetraprolin Family Proteins in Inflammation-Related Diseases by Causing Cytokine mRNA Instability
References

## 1 TYPES OF GROWTH FACTORS, CYTOKINES, AND CHEMOKINES

Growth factors refer to a large class of proteins capable of stimulating cellular proliferation and differentiation. Examples of growth factors include cytokines and hormones that bind to specific receptors on the surface of their target cells.

Cytokines are a subclass of growth factors. Cytokines are primarily secreted from leukocytes. Cytokines that are secreted from lymphocytes are called lymphokines and those secreted by monocytes or macrophages are called monokines. Growth factors are often used interchangeably with cytokines. Traditionally,

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

cytokines are associated with hematopoietic cells and the immune system, including lymphocytes, spleen, thymus, and lymph nodes. However, some of the same signaling proteins are also being used by other cells and tissues. Growth factors imply positive effects on cell division, but cytokines are neutral in terms of their functions. Some cytokines promote cell growth or proliferation, such as granulocyte colony stimulating factor (GCSF) and granulocyte macrophage colony stimulating factor (GM-CSF). Other cytokines exhibit inhibitory effects on these processes, such as Fas ligand as "death" signal causing target cells to undergo programmed cell death or apoptosis.

Chemokines are a family of pro-inflammatory activation-inducible cytokines or small protein signals secreted by cells. Chemokines induce directed chemotaxis in nearby responsive cells and therefore the name *chemo*tactic cyto*kines*. Four types of chemokines exist: (1) C chemokine: the first and the third conserved cysteine residues in the mature protein are missing; (2) CC chemokine: the first two conserved cysteine residues are adjacent in the mature protein; (3) CXC chemokine: one amino acid residue separates the first two conserved cysteine residues in the mature protein; and (4) CX3C chemokine: three amino acid residues separate the first two conserved cysteine residues in the mature protein.

Table 1 lists examples of the major types of growth factors, cytokines, and chemokines. This list is not intended to be complete. New members of growth factors, cytokines, and chemokines are discovered continuously.

## 2 FUNCTIONS OF GROWTH FACTORS, CYTOKINES, AND CHEMOKINES

Growth factors bind to receptors on the cell surface and subsequently activate cellular proliferation or differentiation. Many growth factors stimulate cellular division in numerous cell type, whereas others are specific to particular cell types. They often promote cell differentiation and maturation, whereas varies between growth factors. For example, bone morphogenic proteins stimulate bone cell differentiation, whereas vascular endothelial growth factors stimulate blood vessel differentiation.

Cytokines stimulate the humoral and cellular immune responses and activate phagocytic cells. Each cytokine binds to a specific cell-surface receptor, initiates intracellular signaling cascades, and alters cellular functions. The functional consequences may include upregulation or downregulation of gene expression, resulting in the production of other cytokines; increases in the number of surface receptors for other molecules; or the suppression of their own effects by feedback inhibition. The major functions of cytokines could be classified into three categories: (1) autocrine: cytokine acts on the cells that secrete it; (2) paracrine: cytokine action is restricted to the immediate vicinities of cytokine's secretion; and (3) endocrine: cytokine diffuses to distant regions of the body to affect different tissues. However, cytokines exhibit considerable "redundancy" because many cytokines share similar functions. Cytokines are also pleiotropic factors because they act on many different cell types. The reasons for these effects of cytokines are that a certain type of cell may express receptors for more than one cytokine and different cells may express receptors for the same cytokine.

**TABLE 1. Types of Growth Factors, Cytokines, and Chemokines**

| Type | Representative growth factors, cytokines, and chemokines |
|---|---|
| Adipose-derived cytokine | Leptin |
| Angiogenic cytokine | VEGFa, VEGFb |
| C chemokine | XCL1 (Lymphotactin), XCL2 (SCM-1β) |
| CC chemokine | CCL1 (I-309), CCL2 (MCP-1), CCL3 (MIP-1α), CCL4 (MIP-1β), CCL5 (RANTES), CCL6 (C10), CCL7 (MIP-3), CCL8 (MCP-2), CCL9 (MIP-1γ), CCL10 (MRP-2), CCL11 (Eotaxin), CCL12 (MCP-5), CCL13 (MCP-4), CCL14 (HCC-1), CCL15 (MIP-5), CCL16 (HCC-4), CCL17 (TARC), CCL18 (MIP-4), CCL19 (MIP-3β), CCL20 (MIP-3α), CCL21 (Exodus-2), CCL22 (MDC), CCL23 (MIP-3), CCL24 (Eotaxin-2), CCL25 (TECK), CCL26 (Eotaxin-3), CCL27 (CTACK), CCL28 (MEC) |
| CXC chemokine | CXCL1 (GRO-α), CXCL2 (GRO-β/MIP-2), CXCL3 (GRO-γ), CXCL4 (PF-4), CXCL5 (ENA-78/LIX), CXCL6 (GCP-2), CXCL7 (NAP-2), CXCL8 (IL-8), CXCL9 (MIG), CXCL10 (IP-10), CXCL11 (I-TAC), CXCL12 (SDF-1α/1β), CXCL13 (BCA-1), CXCL14 (BRAK), CXCL15 (Lungkine), CXCL16 (small inducible cytokine B6) |
| CX3C chemokine | CX3CL1 (Fractalkine) |
| Hematopoietic growth factor | IL-20, M-CSF, SCF, SCGF-α, SCGF-β |
| Immunoregulatory lymphokine | IL-2, IL-9, IL-10, IL-13, IL-15 |
| Interferon | INF-α, INF-β, INF-γ, INF-λ |
| Lectin | GALECTIN-1, GLECTIN-3 |
| Neurotrophic factor | ART, BDNF, CNTF, GDNF, MK, NTN/NRTN, β-NGF, NNT-1, NT-3, NT-4, PSP, PTN |
| Osteoinductive cytokine | BMP-2 |
| Other growth factor | EGF, EG-VEGF, Flt3 ligand, G-CSF, GM-CSF, HGF, IGF-1, IGF-II, IL-3, IL-5, IL-7, IL-25 (SF20), KGF/FGF-7, OSM, PDGF, TGFA, TPO |
| Pleiotropic cytokine | CT-1, IL-4, IL-6 |
| Pro-inflammatory cytokine | IL-1α, IL-1β, IL-8, IL-17, IL-19, TNF-α, TNF-β, MIP-1α, MIP-1β, MIP-1γ, MIP-2, MIP-3, MIP-3α, MIP-3β, MIP-4, MIP-5 |
| TGF superfamily | TGF-α, TGF-β1 to TGF-β4, GDF-1 to GDF-15, BMP-1 to BMP-15 |
| TNF superfamily | TNFSF1 (TNF-β), TNFSF2 (TNF-α), TNFSF3 (TNFC/p33), TNFSF4 (TXGP1/gp34), TNFSF5 (CD40L/gp39), TNFSF6 (FasL), TNFSF7 (CD70), TNFSF8 (CD30L), TNFSF9 (CD137L), TNFSF10 (TRAIL/Apo2L), TNFSF11 (RANKL), TNFSF12 (TWEAK/Apo3L), TNFSF13 (APRIL/TALL2), TNFSF13B (THANK/TNFSF20), TNFSF14 (LIGHT), TNFSF15 (TL1/VEG1), TNFSF18 (AITRL/TL6) |

Chemokines induce directed chemotaxis in nearby responsive cells. They are released from various cells in response to bacteria and viruses infection and in response to agents that cause physical damage such as silica or urate crystals. The main functions of chemokines are chemoattractants for leukocytes. They help to recruit monocytes, neutrophils, and other effector cells from the blood to the sites of infection or damage. They serve to guide cells involved in innate immunity and in the adaptive immune system. Some chemokines have other roles in the development of lymphocytes, migration, and the growth of new blood vessels.

Table 2 lists examples of the major functions of growth factors, cytokines, and chemokines. This list is not intended to be complete. New functions of growth factors, cytokines, and chemokines are discovered rapidly.

**TABLE 2. Major Functions of Growth Factors, Cytokines, and Chemokines**

| No. | Abbreviation | Full Name | Major Function |
|---|---|---|---|
| 1 | AITRL | Activation-inducible TNF receptor ligand | Regulates T-cell proliferation and survival; promotes the interaction between T lymphocytes and endothelial cells |
| 2 | Apo-SAA | Apo-serum amyloid A protein | Circulates with high-density lipoproteins; increased by inflammatory stimulus |
| 3 | APRIL | A proliferating-inducing ligand | Stimulates proliferation of various tumor cell lines and B and T cells |
| 4 | ART | Artemin | Supports the survival of peripheral ganglia |
| 5 | 4-1BBL | | Provides a co-stimulatory signal for T cell activation and expansion |
| 6 | BAFF | B cell activating factor | Stimulates B and T cell function |
| 7 | BCA-1 | B cell attracting chemokine | Behaves as a potent chemoattractant for B lymphocytes; induces a weak chemotactic response in T cells and macrophages |
| 8 | BD-1, -2, -3 | β-Defensins | Provides antimicrobial activity; behaves as chemoattractant for immature dendritic cells and memory cells |
| 9 | BDNF | Brain-derived neurotrophic factor | Supports neuron proliferation and survival |
| 10 | BMP | Bone morphogenetic protein | Modulates cell proliferation, differentiation, matrix synthesis, apoptosis; initiates, promotes, and regulates the development, growth, and remodeling of bone and cartilage |
| 11 | BMP-2 | Bone morphogenetic protein-2 | Induces bone and cartilage formation; plays an important role in cardiac morphogenesis |

**TABLE 2.** *Continued*

| No. | Abbreviation | Full Name | Major Function |
|---|---|---|---|
| 12 | BMP-13 | Bone morphogenetic protein-13 | Regulates articular cartilage; repairs oteochondral defects |
| 13 | BMP-14 | Bone morphogenetic protein- 14 | Plays a role in long bones during embryonic development and postnatally in articular cartilage |
| 14 | BRAK | Breast and kidney-expressed chemokine | Behaves as a highly selective monocyte chemoattractant |
| 15 | C10 | CCL6 | Behaves as a chemoattractant B cells, CD4+ T cells, monocytes, and NK cells; exhibits suppressive activity on colony formation |
| 16 | CT-1 | Cardiotrophin-1 | Induces myocyte hypertrophy; enhances the survival of cardiomyocyte and different neuronal populations |
| 17 | CD40L | CD40-Ligand | Plays an important role in B cell proliferation and differentiation, and immunoglobulin class switching |
| 18 | CNTF | Ciliary neurotrophic factor | Provides a vital role in the survival of neural cells |
| 19 | CTACK | Cutaneous T cell-attracting chemokine | Attracts CLA+ T cells and directs them into the skin |
| 20 | CXCL16 | | Attracts lymphocyte subsets during inflammation; facilitates certain immune response |
| 21 | EGF | Epidermal growth factor | Stimulates proliferation of epidermal and epithelial cells; inhibits gastric secretion; involves in wound healing |
| 22 | EG-VEGF | Endocrine-gland-derived vascular endothelial growth factor | Induces proliferation and migration fenestration in capillary endothelial cells |
| 23 | EMAP-II | Endothelial-monocyte activating polypeptide II | Inhibits endothelial cell proliferation, vasculogenesis, neovessel formation; inhibits angiogenesis of vascular beds; suppresses tumor growth; induces apoptosis and myeloperoxidase activity from neutrophils |
| 24 | ENA-78 | Epithelial neutrophil activating peptide 78 | Binds to CXCR2 receptor |
| 25 | Eotaxin | CCL11 | Binds to CCR3 receptor; chemoattracts eosinophils; plays a key role in the regulation of eosinophil recruitment in asthmatic lung and allergic reactions |
| 26 | Eotaxin-2 | CCL24 | Binds to CCR3 receptor; chemoattracts cells expressing CCR3 including eosinophils, |

**TABLE 2.** *Continued*

| No. | Abbreviation | Full Name | Major Function |
|---|---|---|---|
| | | | basophils, Th2 T cells, mast cells; inhibits the proliferation of multipotential hematopoietic progenitor cells |
| 27 | Eotaxin-3 | CCL26 | Similar to Eotaxin and Eotaxin-2 |
| 28 | Exodus-2 | CCL21 | Binds to CCR7 receptor; chemoattracts T and B lymphocytes; inhibits hematopoiesis |
| 29 | FasL | Fas ligand | Binds to Fas receptor; induces apoptosis of Fas-containing cells; kills T cells; activates B cells leading to downregulation of immune response |
| 30 | FGF family | Fibroblast growth factor family | Promotes cell proliferation and differentiation |
| 31 | Flt3L | Flt3-ligand | Binds to cells expressing tyrosine kinase receptor Flt3; regulates proliferation of early hematopoietic cells |
| 32 | FS | Follistatin | Binds to TGF-β ligands; inhibits their access to their signaling receptors |
| 33 | Fractalkine | CX3CL1 | chemoattracts monocytes, microglia cells, and NK cells |
| 34 | Galectin-1 | S-LAC lectin-1 | Binds to β-galactoside moieties; interacts with CD3, CD4, and CD45; induces apoptosis of activated T cells and T leukemia cells; inhibits the protein phosphatase activity of CD45 |
| 35 | Galectin-3 | Galactose-specific lectin-3 | Binds to β-galactoside moieties; regulates embryogenesis, inflammatory responses, cell progression, and metastasis |
| 36 | gAcrp30 | Globular domain of adipocyte complement-related protein of 30 kDa | Binds to AdipoR1 and AdipoR2 receptors; decreases hyperglycemia and reverses insulin resistance; promotes fat loss |
| 37 | GCP-2 | Granulocyte chemotactic protein-2 | Attracts neutrophils; exhibits anti-angiogenic activity |
| 38 | G-CSF | Granulocyte colony-stimulating factor | Stimulates the development of committed progenitor cells to neutrophils; enhances the functional activities of the mature end-cell; facilitates hematopoietic recovery after bone marrow transplantation or chemotherapy |
| 39 | GDF-11 | Growth/differentiation factor-11 | Suppresses neurogenesis through a myostatin-like pathway |

**TABLE 2.** *Continued*

| No. | Abbreviation | Full Name | Major Function |
|---|---|---|---|
| 40 | GDNF | Glial cell line-derived neurotrophic factor | Promotes dopamine uptake, survival and differentiation of midbrain neurons; improves bradykinesia, rigidity, and postural instability |
| 41 | GM-CSF | Granulocyte/macrophage colony-stimulating factor | Stimulates the development of neutrophils and macrophages; promotes proliferation and development of early erythroid megakaryocytic and eosinophilic progenitor cells; inhibits neutrophil migration; enhances the functional activities of the mature end-cells; increases the recovery rate of hematopoietic cells after bone marrow transplantation |
| 42 | GRO | Growth regulated protein/ melanoma growth stimulatory activity family (MGSA) | Chemoattracts and activates neutrophils and basophils |
| 43 | HCC-1 | Hemafiltrate CC chemokine-1 | Chemoattracts blood monocytes |
| 44 | HGF | Hepatocyte growth factor | Induces cell proliferation, motility, morphogenesis; inhibits cell growth; enhances neuron survival; helps liver regeneration |
| 45 | I-309 | CCL1 | Chemoattracts monocytes and Th2-differentiated T cells |
| 46 | IFN-α | Interferon alpha | Induces nonspecific resistance against viral infections; regulates expression of MHC class I antigens; affects cell proliferation; modulates immune responses; treats viral infection and neoplasia |
| 47 | IFN-γ | Interferon gamma | Increases the surface expression of class I MHC antigens to T-helper (CD4+) cells; regulates the antigen-specific phases of the immune responses; stimulates lymphoid cell functions |
| 48 | IFN-λs | Interferon-lambdas | Induces antiviral defense |
| 49 | IGF-I/II | Insulin-like growth factor I and II | Stimulates the proliferation and survival of various cell types similar to insulin but with higher growth-promoting activity |
| 50 | IL-1α | Interleukin-1 alpha | Stimulates thymocyte proliferation by inducing IL-2 release; promotes B cell maturation and proliferation; exhibits mitogenic FGF-like activity; stimulates the release of prostaglandin and collagenase from synovial calls |

**TABLE 2.** *Continued*

| No. | Abbreviation | Full Name | Major Function |
|---|---|---|---|
| 51 | IL-1β | Interleukin-1 beta | Similar to IL-1α. IL-α is a mainly cell-associated cytokine, but IL-β is a secreted cytokine |
| 52 | IL-2 | Interleukin-2 | Stimulates growth and differentiation of B cells, NK cells, lymphokine-activated killer cells, monocytes, macrophages, and oligodendrocytes; enhances the immune system of patients for the treatment of cancer and infectious diseases |
| 53 | IL-3 | Interleukin-3 | Promotes the survival, differentiation, and proliferation of committed progenitor cells of the megakaryocyte, granulocyte-macrophage, erythroid, eosinophil, basophil, and mast cell lineages; enhances thrombopoiesis, phagocytes, and antibody-mediated cellular cytotoxicity; activates monocytes |
| 54 | IL-4 | Interleukin-4 | Regulates diverse T and B cell responses, including cell proliferation, survival, and gene expression; regulates the differentiation of native CD4+ T cells into helper Th2 cells; regulates immunoglobulin class switching to the IgG1 and IgE isotypes |
| 55 | IL-5 | Interleukin-5 | Stimulates the proliferation and activation of eosinophils; induces cell-mediated immunity against parasitic infections and some tumors |
| 56 | IL-6 | Interleukin-6 | Regulates immunes and inflammatory responses; stimulates B cell differentiation and antibody production; induces expression of hepatic acute-phase proteins; regulates bone metabolism |
| 57 | IL-7 | Interleukin-7 | Affects early B and T cells mainly; costimulates mature T cells with other factors |
| 58 | IL-8 | Interleukin-8 | Chemoattracts and activates neutrophils |
| 59 | IL-9 | Interleukin-9 | Enhances the proliferation of T lymphocytes, mass cells, erythroid precursor cells, and megakaryoblastic leukemia cells |

**TABLE 2.** *Continued*

| No. | Abbreviation | Full Name | Major Function |
|---|---|---|---|
| 60 | IL-10 | Interleukin-10 | Inhibits the expression of pro-inflammatory cytokines such as IL-1 and TNF-α; enhances humoral immune responses and attenuates cell-mediated immune responses |
| 61 | IL-11 | Interleukin-11 | Regulates hematopoiesis by stimulating growth of myeloid, erythroid, and megakaryocyte progenitor cells; regulates bone metabolism; inhibits production of pro-inflammatory cytokines; protects against gastromucosal injury |
| 62 | IL-12 | Interleukin-12 | Induces IFN-γ expression by NK and T cells; promotes the growth and activity of activated NK, CD4+, and CD8+ cells; induces the development of INF-γ-producing Th1 cells |
| 63 | IL-13 | Interleukin-13 | Plays a role in the expulsion of gastrointestinal parasites; exhibits anti-inflammatory effects on monocytes and macrophages; inhibits pro-inflammatory cytokine expression; enhances B cell proliferation; increases IgE production by inducing isotype switching |
| 64 | IL-15 | Interleukin-15 | Stimulates the proliferation of T lymphocytes; relates to rheumatoid arthritis, inflammatory bowel diseases |
| 65 | IL-16 | Interleukin-16 | Induces chemotaxis of CD4+ T cells and monocytes and eosinophils; induces expression of IL-2R and MHC class II proteins on CD4+ T cells |
| 66 | IL-17A | Interleukin-17A | Induces production of pro-inflammatory and hematopoietic molecules |
| 67 | IL-17B | Interleukin-17B | Stimulates the release of TNF-α and IL-1β from cells of monocyte lineage |
| 68 | IL-17D | Interleukin-17D | Stimulates the production of IL-6, IL-8, and GM-CSF; inhibits hemopoiesis of myeloid progenitor cells |
| 69 | IL-17E | Interleukin-17E | Stimulates the secretion of IL-8; induces activation of NFκB in cells expressing IL-17BR receptor |

**TABLE 2.** *Continued*

| No. | Abbreviation | Full Name | Major Function |
|---|---|---|---|
| 70 | IL-17F | Interleukin-17F | Stimulates proliferation and activation of T cells and PBMCs; regulates cartilage matrix turnover; inhibits angiogenesis |
| 71 | IL-19 | Interleukin-19 | Upregulates IL-6 and TNF-α; induces apoptosis through TNF-α |
| 72 | IL-20 | Interleukin-20 | Stimulates colony formation by CD34+ multipotential progenitors; induces inflammatory skin diseases |
| 73 | IL-22 | Interleukin-22 | Inhibits IL-4 production by Th2 cells; induces acute-phase reactants in the liver and pancreas |
| 74 | IP-10 | γ-interferon inducible protein 10 | Chemoattracts Th1 lymphocytes and monocytes; inhibits cytokine-stimulated hematopoietic progenitor cell proliferation; behaves angiostatic and mitogenic for vascular smooth muscle cells |
| 75 | I-TAC | Interferon inducible T-cell alpha chemokine | Chemoattracts IL-2-activated T cells |
| 76 | JF | CCL2/MCP-1 | Chemoattracts and activates monocytes, activated T cells, basophils, NK cells, and immune dendritic cells |
| 77 | KC | CXCL1/GROα | Chemoattracts and activates neutrophils and basophils |
| 78 | KGF | Keratinocyte growth factor | Regulates keratinocyte growth |
| 79 | LAG-1 | Lymphocyte activation gene-1 protein | Chemoattracts monocytes; exhibits activity in HIV suppressive factor |
| 80 | LD78β | CCL3L1 | Exhibits activity in HIV suppression assays |
| 81 | LEC | Liver-expressed chemokine (CCL16) | Chemoattracts monocytes and lymphocytes |
| 82 | Leptin | Obesity protein | Suppresses appetite; increases thermogenesis; reduces body weight, food consumption, and plasma glucose levels |
| 83 | LIGHT | TNFSF14 | Activates MFlB; costimulates the activation of lymphocytes; induces apoptosis in some tumors |
| 84 | Lymphotactin | XCL-1 | Chemoattracts lymphocytes |
| 85 | M-CSF | Macrophage colony-stimulating factor | Regulates cell proliferation, differentiation, and survival of blood monocytes, tissue macrophages, and their |

**TABLE 2.** *Continued*

| No. | Abbreviation | Full Name | Major Function |
|---|---|---|---|
| | | | progenitor cells; modulates dermal thickness and male and female fertility; treats infection, malignancies, and atherosclerosis; facilitates hematopoietic recovery after bone marrow transplantation |
| 86 | MCPs | Monocyte chemoattractant proteins | Chemoattracts and activates monocytes, activated T cells, basophils, NK cells, and immune dendritic cells |
| 87 | MDC | Macrophage-derived chemokine (CCL22) | Chemoattracts monocytes, NK cells, and dendritic cells; exhibits HIV suppressive activity |
| 88 | MEC | Mucosae-associated epithelial chemokine (CCL28) | Chemoattracts resting CD4 and CD8 T cells and eosinophils |
| 89 | MK | Midkine | Chemoattracts embryonic neurons, neutrophils, and macrophages; exhibits angiogenic, growth, and survival activities during tumorgenesis |
| 90 | MIG | Monokine-induced by interferon-γ (CXCL9) | Chemoattracts Th1 lymphocytes; inhibits tumor growth, angiogenesis, and colony formation of hematopoietic progenitors |
| 91 | MIP-1α/1β | Macrophage inflammatory protein-1 alpha/-1 beta | Regulates the trafficking and activation state of inflammatory cells such as macrophages, lymphocytes, and NK cells; chemoattracts B cells, eosinophils, and dendritic cells |
| 92 | MIP-1γ | Macrophage inflammatory protein-1 gamma | Chemoattracts neutrophils; inhibits colony formation of bone marrow myeloid immature progenitors |
| 93 | MIP-2 | Macrophage inflammatory protein-2 (CXCL2/GROβ) | Chemoattracts and activates neutrophils and basophils |
| 94 | vMIP-2 | Viral macrophage inflammatory protein-2 | Chemokine analog of human herpes virus; exhibits antagonistic activity toward several chemokine receptors |
| 95 | MIP-3 | Macrophage inflammatory protein-3 (CCL23) | Chemoattracts monocytes and resting T lymphocytes; inhibits colony formation of bone marrow myeloid immature progenitors |

**TABLE 2.** *Continued*

| No. | Abbreviation | Full Name | Major Function |
|---|---|---|---|
| 96 | MIP-3α | Macrophage inflammatory protein-3 alpha (CCL20) | Chemoattracts lymphocytes and dendritic cells; promotes the adhesion of memory CCD4+ T cells; inhibits colony formation of bone marrow myeloid immature progenitors |
| 97 | MIP-3β | Macrophage inflammatory protein-3 beta (CCL19) | Chemoattracts T and B lymphocytes and myeloid progenitor cells |
| 98 | MIP-4 | Macrophage inflammatory protein-4 (CCL18) | Chemoattracts lymphocytes; exhibits activity on CD4+ and CD8+ T cells |
| 99 | MIP-5 | Macrophage inflammatory protein-5 (CCL15) | Chemoattracts T cells and monocytes |
| 100 | Myostatin | GDF-8 | Inhibits skeletal muscle growth; interacts with Activin type I and II receptors; suppresses myoblast proliferation by arresting cell cycle in G1 phase |
| 101 | Myostatin propeptide | | Binds to and inhibits myostatin; increases skeletal muscle growth |
| 102 | NAP-2 | Neutrophil activating protein-2 (CXCL7) | Chemoattracts and activates neutrophils |
| 103 | NTN | Neurturin | Promotes neuron development and survival |
| 104 | β-NGF | Beta-nerve growth factor | Promotes neuron development and preservation |
| 105 | NNT-1 | Novel neurotrophin-1/ B-cell stimulating factor-3 (BSF-3) | Binds to and activates glycoprotein 130 and leukemia inhibitory factor receptor member β; induces tyrosine phosphorylation of these receptors |
| 106 | Noggin | | Binds to ligands of TGF-β family and regulates their activity by interfering their access to signaling receptors |
| 107 | NP-1 | Neutrophil peptide-1 | Exhibits antimicrobial activity and chemotactic activity on dendritic cells |
| 108 | NT-3 | Neurotrophin-3 | Promotes the growth and survival of nerve and glial cells |
| 109 | NT-4 | Neurotrophin-4 | Promotes the survival of peripheral sensory sympathetic neurons |
| 110 | OSM | Oncostatin M | Regulates neurogenesis, osteogenesis, and hematopoiesis; stimulates the proliferation of fibroblast, smooth muscle cells, and Kaposi's sarcoma cells; inhibits some tumor cells; promotes cytokine release from |

**TABLE 2.** *Continued*

| No. | Abbreviation | Full Name | Major Function |
|---|---|---|---|
| | | | endothelial cells; enhances the expression of low-density lipoprotein receptor in hematoma cells |
| 111 | OPG | Osteoprotegerin | Binds to sRANKL; inhibits osteoclastogenesis by interrupting the signaling between stromal cells and osteoclastic progenitor cells and leads to excess accumulation of bone and cartilage |
| 112 | PDGF | Platelet-derived growth factor (AA, BB, AB forms) | Acts as mitogens for cells like smooth muscle cells, connective tissue cells, bone and cartilage cells, and blood cells; relates to hyperplasia, chemotaxis, embryonic neuron development, and respiratory tubule epithelial cell development |
| 113 | PSP | Persephin | Promotes the survival of ventral mid-brain dopaminergic neurons and motor neurons after sciatic nerve oxotomy; promotes ureteric bud branching |
| 114 | PF-4 | Platelet factor 4 (CXCL4) | Chemoattracts neutrophils and monocytes; inhibits angiogenesis |
| 115 | PTN | Pleiotrophin | Functions in neurogenesis, cell migration, secondary organogenetic induction, and mesoderm-epithelial interaction |
| 116 | PTHrP | Parathyroid hormone-related protein | Regulates extracellular concentrations of calcium and phosphorous |
| 117 | RANKL | Receptor activator of NFκB ligand | Regulates specific immunity and bone turnover; binds to RANK receptor; stimulates naïve T cell proliferation; promotes survival of RANK +T cells; regulates T cell-dependent immune response; binds to OPG; inhibits osteoclastogenesis and leads to excess accumulation of bone and cartilage |
| 118 | RANK | Receptor-activator of NF- kappa B | Binds to RANKL |
| 119 | RANTES | Regulated upon activation normal T cell express sequence | Chemoattracts monocytes, memory T cells (CD4+/CD45R+), basophils, and eosinophils; inhibits some strains of HIV-1, HIV-2, and simian immunodeficiency virus (SIV) |

**TABLE 2.** *Continued*

| No. | Abbreviation | Full Name | Major Function |
|---|---|---|---|
| 120 | RELMα | Resistin-like molecule alpha | Tissue-specific cytokine; physiological roles unknown |
| 121 | RELMβ | Resistin-like molecule beta | Tissue-specific cytokine; physiological roles unknown |
| 122 | Resistin | Adipose tissue-specific secretory factor | Tissue-specific cytokine; physiological roles unknown |
| 123 | SCF | Stem cell factor | Essential for the survival, proliferation, and differentiation of hematopoietic cells committed to the melanocyte and germ cell lineages |
| 124 | SCGFα/β | Stem cell growth factor alpha/beta | Supports the growth of primitive hematopoietic cells; promotes proliferation of erythroid or myeloid progenitor cells |
| 125 | SDF-1α/β | Stromal cell derived factor-1 alpha/beta (CXCL12) | Chemoattracts T and B cells; induces migration of CD34+ stem cells; exhibits HIV suppressive activity in cells expressing CXCR4 receptor |
| 126 | SF20 | IL-25 | Binds to the surface of cells expressing TSA-1 (Thymic shared Ag-1) receptor; stimulates the proliferation of FDCP2 cells and lymphoid cells |
| 127 | Shh | Sonic Hedgehog | Regulates developmental processes |
| 128 | TACI | Transmembrane activator and CAML interactor (TNFRSF13B) | Binds to APRIL and BAFF; stimulates the activation of NFκB and AP-1; mediates calcineurin-dependent activation of NF-AT; stimulates B and T cell function |
| 129 | TARC | Thymus and activation regulated chemokine (CCL17) | Chemoattracts T cells |
| 130 | TECK | Thymus expressed-chemokine (CCL25) | Chemoattracts activated macrophages, thymocytes, and dendritic cells |
| 131 | TGF-α | Transforming growth factor-alpha | Stimulates the proliferation of epidermal and endothelial cells; induces transformation and anchorage independence in cultured cells |
| 132 | TGF-β1 | Differentiation inhibiting factor and cartilage-inducing factor | Regulates cell proliferation, growth, differentiation, motility, synthesis and deposition of extracellular matrix, embryogenesis, tissue remodeling, and wound healing |
| 133 | TGF-β2 | Glioblastoma-derived T cell suppressor factor | Regulates cell proliferation, growth, differentiation, motility, synthesis and deposition of extracellular matrix, embryogenesis, tissue remodeling, and wound healing |

**TABLE 2.** *Continued*

| No. | Abbreviation | Full Name | Major Function |
|---|---|---|---|
| 134 | TGF-β3 | | Regulates cell proliferation, growth, differentiation, motility, synthesis and deposition of extracellular matrix, embryogenesis, tissue remodeling, and wound healing |
| 135 | TNF-α | Tumor necrosis factor alpha | Inhibits tumor cells; mediates immune response against bacterial infection; induces septic shock, autoimmune diseases, rheumatoid arthritis, inflammation, and diabetes |
| 136 | TNF-β | Tumor necrosis factor beta | Mediator of inflammatory and immune response; similar to TNF-α |
| 137 | TPO | Thrombopoietin | Stimulates megakaryocyte proliferation and maturation |
| 138 | TRAIL | TNF-related apoptosis-inducing ligand (Apo2L) | Activates rapid apoptosis in tumor cells |
| 139 | TWEAK | TNF-related weak inducer of apoptosis | Induces NFκB activation and chemokine secretion; promotes endothelial cell proliferation and migration |
| 140 | VEGF | Vascular endothelial growth factor | Stimulates endothelial cell proliferation and migration; promotes angiogenesis and vascular permeability |

## 3 SOME DRUGS INVOLVING GROWTH FACTORS, CYTOKINES, AND CHEMOKINES

It was estimated that the revenue potential involving growth factors, cytokines, and chemokines could reach at least tens of billions of U.S. dollars. The research based on these proteins is changing the treatment of various diseases such as inflammation and infectious and neoplastic diseases. Current product development is focused on autoimmunity, allergy, cancer, infection, transplantation, and wound healing. A number of the first-generation cytokine-based drugs, including IFN-1β and TNF-α, are being replaced by next-generation drugs developed from studies using more refined approaches, causing the knowledge of growth factors/cytokines/chemokine networks to become more clarified. The increased market potential of approved drugs targeting growth factors, cytokines, chemokines, and their receptors has encouraged major biotechnology and pharmaceutical companies to devote their pipelines to the development of new therapeutics.

Table 3 lists a few examples of the therapeutic targets currently under development targeting growth factors, cytokines, and chemokines.

**TABLE 3. Representative Drugs Involving Growth Factors, Cytokines, and Chemokines**

| Type | Representative Drug | Company |
|---|---|---|
| TNF superfamily | HGS-ETR1 (agonistic human monoclonal antibody to TRAIL receptor 1) | Human Genome Sciences |
| | HGS-ETR2 (agonistic human monoclonal antibody to TRAIL receptor 2) | Human Genome Sciences |
| | HGS-TR2J (agonistic human monoclonal antibody to TRAIL receptor 2) | Human Genome Sciences |
| Growth factor | Neupogen (Filgrastim, r-metHuG-CSF) | Amgen |
| | Neulasta (Pegfilgrastim, PEG-r-metHuG-CSF) | Amgen |
| | Nutropin [somatropin (rDNA origin) for injection] | Amgen |
| | Growth hormone | GENENTECH |
| | Kepivance (a recombinant form of human keratinocyte growth factor) | Amgen |
| Pro-inflammatory cytokine | Enbrel (etanercept) (TNF blocker) | Amgen |
| | Kineret (a recombinant form of a naturally occurring protein that regulates interleukin-1) | Amgen |
| C chemokine | CCR5 mAb | Human Genome Sciences |
| Angiogenic cytokine | Avastin (bevacizumab) (anti-VEGF antibody) | GENENTECH |
| Interferon | INTRON A (recombinant interferon α-2b) | Schering-Plough Corp. |
| | Albuferon (albumin-interferon α) | Human Genome Sciences |

## 4 FORMULATIONS OF GROWTH FACTORS, CYTOKINES, AND CHEMOKINES

The formulations for growth factors, cytokines, and chemokines include low concentrations of nontoxic detergents, such as Tween 80, and osmotic agents, such as sorbitol. Additional agents are also included for providing optimum pH for their activity. Some salts are usually included in the formulations. A number of formulations for growth factors, cytokines, and chemokines exist. These formulations are based on saline, liposome, conjugation, and coadministration. The goals of developing optimum formulations are designed to protect the activities of these proteins and to effectively deliver the biologics to the targets.

Table 4 lists a few examples of the formulations for growth factors, cytokines, and chemokines (please refer to Refs. 1–12 for more details).

**TABLE 4. Formulation of Growth Factors, Cytokines, and Chemokines**

| Formulation | Growth factor and cytokine | Reference |
|---|---|---|
| Saline | G-CSF | [*1*] |
| | IL-2 | [*2*] |
| Liposome | G-CSF/ProGelz interleukin-2 | [*1*] |
| | GM-CSF and TNF-alpha | [*2*] |
| | VEGF/PLGA | [*3*] |
| Conjugation | PEGfilgrastim, MonoPEGylated INF-alpha2b | [*4*] |
| | bio-bFGF/OX26-SA | [*5*] |
| | Fc-EPO (NDS) fusion protein | [*6*] |
| | (99m)Tc-IL-8 | [*7*] |
| | PEG-rHuMGDF | [*8*] |
| | PEGlylated INF | [*9*] |
| Coadministration | IL-2/INF-alpha/histamine | [*10*] |
| | TNF-alpha/Doxil | [*11*] |
| | TNF-alpha/MTX | [*12*] |

## 5 DELIVERY OF GROWTH FACTORS, CYTOKINES, AND CHEMOKINES

A number of routes for the delivery of growth factors, cytokines, and chemokines have been studied in detail. Those routes include: (1) oral administration; (2) intravenous injection; (3) intraperitoneal injection; (4) intramuscular injection; and (5) subcutaneous injection. Most of the delivery methods for these compounds are accomplished through injection because of their protein nature.

The U.S. FDA recently approved the first inhaled version of insulin (Exubera) for adults with type 1 or type 2 diabetes. Pfizer and Aventis developed Exubera, in collaboration with Nektar Therapeutics (formerly Inhale Therapeutics, a company that specializes in finding delivery solutions for oral, injectable, and pulmonary drug administration). It weighs about 4 ounces and is about the size of an eyeglass case when closed. It is portable, but not necessarily discrete. Exubera is a short-acting powder form of insulin that can be taken before meals. Inhaled insulin enters the bloodstream more rapidly than that delivered by subcutaneous injection. However, most people with diabetes still need to get long-acting insulin via an injection. Some side effects are associated with the inhaled insulin, including coughing, shortness of breath, sore throat, and dry mouth. It is also possible that overdose for smokers may be developed because smokers get more of the inhaled insulin entering their bloodstream than nonsmokers. Much still needs to be done to effectively deliver growth factors, cytokines, and chemokines to the patients.

Table 5 lists a few examples of the delivery for growth factors, cytokines, and chemokines (please refer to Refs 2, 3, 5, 9, 10, 13–17 for more details).

## 6 PHARMACOKINETICS OF GROWTH FACTORS, CYTOKINES, AND CHEMOKINES

Numerous published papers exist dealing with the pharmacokinetics of growth factors, cytokines, and chemokines. Those studies were performed using cells,

**TABLE 5. Routes of Administration of Growth Factors, Cytokines, and Chemokines**

| Route of delivery | Growth factor and cytokine | Reference |
|---|---|---|
| Oral administration | TLF | [*13*] |
| Intravenous injection | VEGF | [*3*] |
| | BFGF | [*5*] |
| | IL-2 | [*2*] |
| | Apo2L/TRAIL | [*14*] |
| | IL-1 | [*15*] |
| Intraperitoneal injection | INF-gamma | [*16*] |
| Intramuscular injection | INFs | [*9*] |
| Subcutaneous injection | VEGF | [*3*] |
| | IL-2/IFN-alpha | [*10*] |
| | G-CSF analog | [*17*] |
| | INFs | [*9*] |

tissues, animals, and patients. The experiments tested the various formulations and delivery methods mentioned above.

Table 6 is a summary of the selected studies on the formulation, delivery, and pharmacokinetics for growth factors, cytokines, and chemokines (Please refer to Refs. 1–18 for more details).

## 7 POTENTIAL APPLICATIONS OF TRISTETRAPROLIN FAMILY PROTEINS IN INFLAMMATION-RELATED DISEASES BY CAUSING CYTOKINE mRNA INSTABILITY

Tristetraprolin (TTP) and related proteins are well-established factors that destabilize cytokine mRNAs [*19*, *20*]. TTP is the best-understood member of a small family of tandem CCCH zinc finger proteins. Similar zinc finger sequences are found in numerous species in the GenBank database ranging from human to yeast and plants (including human, cow, mouse, rat, sheep, chimpanzee, dog, horse, pig, *Xenopus*, carp, zebrafish, *Drosophila*, *C. elegans*, baker's yeast, fission yeast, oyster, rice, and *Arabidopsis*.) In humans, three members of this family have been characterized: TTP (ZFP36, TIS11, G0S24, and NUP475), ZFP36L1 (TIS11B, cMG1, ERF1, BRF1, and Berg36), and ZFP36L2 (TIS11D, ERF2, and BRF2) [*19*]. They are encoded by different genes and their patterns of cell- and tissue-specific expression and agonist-stimulated expression are quite different. However, they share certain properties: All have highly conserved tandem zinc finger domains, in which each C8xC5xC3xH zinc finger is preceded by the sequence R/KYKTEL, and the two fingers are separated by 18 amino acids [*19*]; all are nuclear-cytoplasmic shuttling proteins [*21*, *22*]; and all are capable of binding AREs within single-stranded RNA [*23–29*] and promoting the deadenylation and subsequent destruction of those transcripts, both in transfection studies and in cell-free experiments [*26–28*, *30*].

TTP is an extraordinarily low-abundance, inducible, stable cytosolic, and hyperphosphorylated mRNA binding protein [*24*, *25*, *31–33*]. TTP mRNA and protein

**TABLE 6. Summary of the Formulation, Delivery, and Pharmacokinetics for Growth Factors, Cytokines, and Chemokines**

| Drug | Experimental system | Formulation | Delivery | Pharmacokinetics | Reference |
|---|---|---|---|---|---|
| Apo2L/ TRAIL | Mice model of human colon carcinoma | Apo2L/ TRAIL | Intravenous injection | Intact [125I]-Apo2L/TRAIL was detectable in the solid tumor at all time points and was the only tissue in which radioactivity transiently increased over time. Apo2L/ TRAIL was stable in the circulation, localized to human solid xenograft tumors, and was primarily eliminated through the kidney | [*14*] |
| Epo | Mice myeloma cells | Fc-Epo(NDS) fusion protein | | Fc-Epo(NDS) was secreted almost exclusively as a dimer, was relatively stable to the removal of N-linked oligosaccharides, had much improved pharmacokinetic properties, and had a significantly improved effect on RBC production | [*6*] |
| bFGF | Rat | bFGF, bio-bFGF, and bio-bFGF/OX26-SA conjugation to a blood-brain barrier peptide drug delivery vector | Intravenous injection | The brain uptake of the [$^{125}$I]-bio-bFGF/OX26-SA was increased five-fold, although the uptake of the conjugate by peripheral tissues was decreased relative to the unconjugated bio-bFGF. Conjugation of bio-bFGF to a BBB drug delivery vector (1) caused only a minor decrease in affinity for the bFGF receptor, | [*5*] |

**TABLE 6.** *Continued*

| Drug | Experimental system | Formulation | Delivery | Pharmacokinetics | Reference |
|---|---|---|---|---|---|
| | | | | (2) decreased the peripheral organ uptake of the bFGF, and (3) increased the brain uptake of the neurotrophin | |
| G-CSF | Mice | ProGelz (PG) (17% poloxamer-407 and 5% hydroxypropyl methylcellulose) vs. saline | One injection in PG vs. multiple injections in saline | PG increased bioavailability (>1.5-fold), the time required to achieve the maximum concentration (Tmax) (six fold), and prolonged elimination (Tbeta) half-life (>three fold). Mobilized HPC was transported more rapidly to the spleen than that to the blood (289-fold vs. eight-fold increase) | [*1*] |
| GM-CSF & TNF-α | Normal and tumor-bearing mice | Liposomal cytokine vs. soluble cytokines | | Liposome-entrapped GM-CSF and TNF-α was more efficacious immunomodulators than the soluble cytokines: (1) the entrapped cytokine displayed high stability; (2) release of TNF-α, but not of GM-CSF, from the liposomes was required for their biological activity in vitro; (3) plasma half-lives and the area under the curve of the entrapped cytokines were 10–20 times greater than those of the soluble cytokines; (4) the toxicity of liposomal TNF-α was 1/3–1/7 that of soluble TNF-α; (5) liposomal GM-CSF led to a 2–4-fold increase | [*18*] |

| | | | | | |
|---|---|---|---|---|---|
| | | | | vs. soluble GM-CSF in the number of peritoneal and spleen leukocytes and of GM colony-forming cells in the spleen; and (6) liposomal GM-CSF markedly enhanced the level of blood granulocytes | |
| G-CSF analog | Monkeys | | Subcutaneous injection | The analog enhanced stability and resulted in greater systemic exposure, probably because of improved absorption from the subcutaneous compartment | [*17*] |
| IFN-α2b | Rat | MonoPEGylated IFN-α2b vs. native IFN-α | | PEGylation markedly enhanced both the resistance to tryptic degradation and the thermal stability of IFN-alpha2b. The serum half-life of 40K PEG-IFN was 330-fold longer, while plasma residence time was increased 708 times compared with native IFN. The PEG2,40K conjugate of IFN-α2b has higher stability than the native cytokine | [*4*] |
| IFN-γ | Mice | | Intraperitoneal and intravenous injection | IFN-γ pretreatment for 24 h resulted in retardation of plasma elimination of the drug with a concomitant increase of its tissue levels in liver, kidney, and intestine. A more pronounced effect might be elicited by IFN-γ for common substrates of P-gp and CYP3A | [*16*] |
| IFNs | | Pegylated interferon | Intramuscular, muscular, intravenous, and subcutaneous injection | The decline in serum concentrations of IFNs was rapid after intravenous administration. The volume of distribution was approximately 20–60% of body weight. Terminal elimination half-lives ranged from | [*9*] |

**TABLE 6.** *Continued*

| Drug | Experimental system | Formulation | Delivery | Pharmacokinetics | Reference |
|---|---|---|---|---|---|
| | | | | 4 to 16 hrs, 1 to 2 hrs, and 25 to 35 min for alpha, beta, and gamma, respectively. Intramuscular and subcutaneous administration of interferons alpha and beta resulted in protracted but fairly good absorption >80% for interferon alpha to 70% for interferon gamma. Pegylated interferon safety and pharmacodynamic profiles were comparable. Pegylated interferon demonstrated delayed clearance compared with non-pegylated interferon. The severity of adverse effects was dose-dependent. Most patients treated had influenza-like syndrome within 2 to 8 hrs of drug administration. Other effects such as fatigue, lethargy, and anorexia were usually dose-limiting. Neuropsychiatric reactions might also become dose-limiting. Interferon induced the formation of serum neutralizing antibodies in approximately 10–20% of treated patients | |
| IL-1α | Phase I trial | | Intravenous injection | IL-1 had measurable effects on tumor blood flow and caused a significant decrease in blood flow. This decrease was temporally associated | [*15*] |

| | | | | | |
|---|---|---|---|---|---|
| | | | | with a significant leukopenia and an increased expression of the adhesion integrin CD11b on the circulating cell surface. IL-1 caused decreased tumor blood flow in vivo in human cancer patients, an effect that was temporally related to cytokine-induced peripheral blood cellular changes | |
| IL-2, IFN-α | Patients with stage IV melanoma | IL-2 2.4 MIU/m2, IFN-alpha 3 MIU, histamine as an adjunct | Subcutanous injections | Histamine, IL-2, and IFN-α treatment was safe, well-tolerated, and tumor responses were observed | [*10*] |
| IL-2 | Mice | Liposomal IL-2 vs. soluble IL-2 | Intravenous injection | Liposomal IL-2 exhibited a superior immunomodulatory activity and more efficiently stimulated spleen cell proliferation and lymphokine-activated killer (LAK) cell activation in vitro, fastened the circulation time, and increased in the leukocyte levels in blood, spleen, and peritoneal exudates | [*2*] |
| IL-8 | Neutropenic rabbits | IL-8 was coupled with (99 m)Tc | | (99 m)Tc-IL-8 was localized in the abscess, mainly bound to peripheral neutrophils. Accumulation in the abscess was a highly specific, neutrophil-driven process. Total fraction that accumulated in the inflamed tissue was extremely high | [*7*] |
| MGDF | Monkey bone marrow-derived mononuclear cell | Pegylated MGDF (PEG-rHuMGDF) | Infusion | The study confirmed predictions from PK/PD modeling of PEG-rHuMGDF (pegylated recombinant human megakaryocyte growth and development factor) that | [*8*] |

**TABLE 6.** *Continued*

| Drug | Experimental system | Formulation | Delivery | Pharmacokinetics | Reference |
|---|---|---|---|---|---|
| | | | | thrombocytopenia is preventable after AuBMT | |
| TLF | Patients with advanced solid tumors | | Oral administration | Significant levels of talactoferrin were undetectable in circulation, but a statistically significant increase in circulating IL-18 occured, a pharmacodynamic indicator of talactoferrin activity | [*13*] |
| TNF-α | Melanoma-bearing mice | Co-administered Doxil | Coinjection | Systemic application of clinically tolerable doses of TNF might improve drug distribution and tumor response | [*11*] |
| TNF-α | Patients with RA | Adalimumab (a human anti-TNF α monoclonal antibody) added to continuing MTX therapy | Coadministration of adalimumab and MTX | Adalimumab exhibited linear pharmacokinetics. Among patients with active RA who had not had an adequate response to MTX, addition of adalimumab to MTX achieved long-term improvement compared with placebo plus MTX | [*12*] |
| VEGF | Rat | Poly(lactic-co-glycolic) acid microspheres vs. protein solutions | Subcutaneous vinjection of microspheres vs. intravenous injection of solutions | VEGF in microspheres administered subcutaneously was detected with low plasma concentrations and high subcutaneous concentrations over a period of 7 weeks. VEGF in solution administered intravenously was rapidly cleared with high plasma concentrations as expressed by rapid absorption and elimination | [*3*] |

are detected in a number of tissues including spleen, thymus, lymph node, lung, liver, and intestine, consistent with its proposed antiinflammatory role [*31*, *34*]. The expression levels of TTP mRNA and protein in mammalian cells are increased in response to several kinds of stimuli, including insulin and other growth factors, and stimulators of innate immunity, such as the endotoxin lipopolysaccharide (LPS) [*31*, *34*]. TTP is an mRNA-binding protein with high binding specificity for the so-called class II AREs within the 3′-untranslated mRNAs [*23–27*, *29*, *30*, *35*, *36*]. TTP and its related proteins can bind to the 3′-untranslated AREs of certain clinically important mRNAs, such as those coding for TNFα [*24–27*], GM-CSF [*30*], IL-3 [*36*], cyclooxygenase-2 [*37*, *38*], and plasminogen activator inhibitor type 2 [*39*]. This specific binding of TTP to the AREs results in the destabilization of the mRNAs in transfected human embryonic kidney (HEK) 293 cells [*26*, *27*]. In animals, TTP deficiency causes a profound inflammatory syndrome with cachexia, dermatitis, erosive arthritis, autoimmunity, and myeloid hyperplasia, apparently because of excessive production of TNFα and GM-CSF, both of whose mRNAs are direct targets of TTP and are stabilized in cells from the knockout (KO) mice [*27*, *30*, *40*]. For these reasons, TTP can be thought of as an anti-inflammatory or arthritis-suppressor protein [*25*, *40*, *41*]. Therefore, TTP and related proteins are potential therapeutic agents for inflammation-related diseases.

## REFERENCES


1. Robinson S N, Chavez J M, Blonder J M, et al. (2005). Hematopoietic progenitor cell mobilization in mice by sustained delivery of granulocyte colony-stimulating factor. *J. Interferon Cytokine Res*. 25:490–500.
2. Kedar E, Gur H, Babai I, et al. (2000). Delivery of cytokines by liposomes: Hematopoietic and immunomodulatory activity of interleukin-2 encapsulated in conventional liposomes and in long-circulating liposomes. *J. Immunother*. 23:131–145.
3. Kim T K, Burgess D J (2002). Pharmacokinetic characterization of 14C-vascular endothelial growth factor controlled release microspheres using a rat model. *J. Pharm. Pharmacol*. 54:897–905.
4. Ramon J, Saez V, Baez R, et al. (2005). PEGylated interferon-alpha2b: A branched 40K polyethylene glycol derivative, *Pharm. Res*. 22:1374–1386.
5. Wu D, Song B W, Vinters H V, et al. (2002). Pharmacokinetics and brain uptake of biotinylated basic fibroblast growth factor conjugated to a blood-brain barrier drug delivery system. *J. Drug Target* 10:239–245.
6. Way J C, Lauder S, Brunkhorst B, et al. (2005). Improvement of Fc-erythropoietin structure and pharmacokinetics by modification at a disulfide bond. *Protein Eng. Des. Sel*. 18:111–118.
7. Rennen H J, Boerman O C, Oyen W J, et al. (2003). Kinetics of 99mTc-labeled interleukin-8 in experimental inflammation and infection. *J. Nucl. Med*. 44:1502–1509.
8. Farese A M, MacVittie T J, Roskos L, et al. (2003). Hematopoietic recovery following autologous bone marrow transplantation in a nonhuman primate: effect of variation in treatment schedule with PEG-rHuMGDF. *Stem Cells*. 21:79–89.
9. Arnaud P (2002). The interferons: Pharmacology, mechanism of action, tolerance and side effects, *Rev. Med. Inter*. 23 (suppl 4):449s–458s.

10. Lindner P, Rizell M, Mattsson J, et al. (2004). Combined treatment with histamine dihydrochloride, interleukin-2 and interferon-alpha in patients with metastatic melanoma. *Anticancer Res*. 24:1837–1842.
11. Brouckaert P, Takahashi N, van Tiel S T, et al. (2004). Tumor necrosis factor-alpha augmented tumor response in B16BL6 melanoma-bearing mice treated with stealth liposomal doxorubicin (Doxil) correlates with altered Doxil pharmacokinetics. *Int. J. Cancer*. 109:442–448.
12. Weisman M H, Moreland L W, Furst D E, et al. (2003). Efficacy, pharmacokinetic, and safety assessment of adalimumab, a fully human anti-tumor necrosis factor-alpha monoclonal antibody, in adults with rheumatoid arthritis receiving concomitant methotrexate: a pilot study. *Clin. Ther*. 25:1700–1721.
13. Hayes T G, Falchook G F, Varadhachary G R, et al. (2006). Phase I trial of oral talactoferrin alfa in refractory solid tumors. *Invest. New Drugs*. 24:233–240.
14. Xiang H, Nguyen C B, Kelley S K, et al. (2004). Tissue distribution, stability, and pharmacokinetics of Apo2 ligand/tumor necrosis factor-related apoptosis-inducing ligand in human colon carcinoma COLO205 tumor-bearing nude mice. *Drug Metab. Dispos*. 32:1230–1238.
15. Logan T F, Jadali F, Egorin M J, et al. (2002). Decreased tumor blood flow as measured by positron emission tomography in cancer patients treated with interleukin-1 and carboplatin on a phase I trial. *Cancer Chemother. Pharmacol*. 50:433–444.
16. Kawaguchi H, Matsui Y, Watanabe Y, et al. (2004). Effect of interferon-gamma on the pharmacokinetics of digoxin, a P-glycoprotein substrate, intravenously injected into the mouse. *J. Pharmacol. Exp. Ther*. 308:91–96.
17. Luo P, Hayes R J, Chan C, et al. (2002). Development of a cytokine analog with enhanced stability using computational ultrahigh throughput screening. *Protein Sci*. 11:1218–1226.
18. Kedar E, Palgi O, Golod G, et al. (1997). Delivery of cytokines by liposomes. III. Liposome-encapsulated GM-CSF and TNF-alpha show improved pharmacokinetics and biological activity and reduced toxicity in mice. *J. Immunother*. 20:180–193.
19. Blackshear P J (2002). Tristetraprolin and other CCCH tandem zinc-finger proteins in the regulation of mRNA turnover. *Biochem. Soc. Trans*. 30:945–952.
20. Carrick D M, Lai W S, Blackshear P J (2004). The tandem CCCH zinc finger protein tristetraprolin and its relevance to cytokine mRNA turnover and arthritis. *Arthritis Res. Ther*. 6:248–264.
21. Murata T, Yoshino Y, Morita N, et al. (2002). Identification of nuclear import and export signals within the structure of the zinc finger protein TIS11. *Biochem. Biophys. Res. Commun*. 293:1242–1247.
22. Phillips R S, Ramos S B, Blackshear P J (2002). Members of the tristetraprolin family of tandem CCCH zinc finger proteins exhibit CRM1-dependent nucleocytoplasmic shuttling. *J. Biol. Chem*. 277:11606–11613.
23. Blackshear P J, Lai W S, Kennington E A, et al. (2003). Characteristics of the interaction of a synthetic human tristetraprolin tandem zinc finger peptide with AU-rich element-containing RNA substrates. *J. Biol. Chem*. 278:19947–19955.
24. Cao H, Dzineku F, Blackshear P J (2003). Expression and purification of recombinant tristetraprolin that can bind to tumor necrosis factor-alpha mRNA and serve as a substrate for mitogen-activated protein kinases. *Arch. Biochem. Biophys*. 412:106–120.
25. Cao H (2004). Expression, purification, and biochemical characterization of the anti-inflammatory tristetraprolin: a zinc-dependent mRNA binding protein affected by posttranslational modifications. *Biochem*. 43:13724–13738.

26. Carballo E, Lai W S, Blackshear P J (1998). Feedback inhibition of macrophage tumor necrosis factor-alpha production by tristetraprolin. *Sci.* 281:1001–1005.
27. Lai W S, Carballo E, Strum J R, et al. (1999). Evidence that tristetraprolin binds to AU-rich elements and promotes the deadenylation and destabilization of tumor necrosis factor alpha mRNA. *Mol. Cell. Biol.* 19:4311–4323.
28. Lai W S, Carballo E, Thorn J M, et al. (2000). Interactions of CCCH zinc finger proteins with mRNA. Binding of tristetraprolin-related zinc finger proteins to Au-rich elements and destabilization of mRNA. *J. Biol. Chem.* 275:17827–17837.
29. Worthington M T, Pelo J W, Sachedina M A, et al. (2002). RNA binding properties of the AU-rich element-binding recombinant Nup475/TIS11/tristetraprolin protein. *J. Biol. Chem.* 277:48558–48564.
30. Carballo E, Lai W S, Blackshear P J (2000). Evidence that tristetraprolin is a physiological regulator of granulocyte-macrophage colony-stimulating factor messenger RNA deadenylation and stability. *Blood.* 95:1891–1899.
31. Cao H, Tuttle J S, Blackshear P J (2004). Immunological characterization of tristetraprolin as a low abundance, inducible, stable cytosolic protein. *J. Biol. Chem.* 279:21489–21499.
32. Cao H, Deterding L J, Venable J D, et al. (2006). Identification of the antiinflammatory protein tristetraprolin as a hyperphosphorylated protein by mass spectrometry and site-directed mutagenesis. *Biochem. J.* 394:285–297.
33. Carballo E, Cao H, Lai W S, et al. (2001). Decreased sensitivity of tristetraprolin-deficient cells to p38 inhibitors suggests the involvement of tristetraprolin in the p38 signaling pathway. *J. Biol. Chem.* 276:42580–42587.
34. Lai W S, Stumpo D J, Blackshear P J (1990). Rapid insulin-stimulated accumulation of an mRNA encoding a proline-rich protein. *J. Biol. Chem.* 265:16556–16563.
35. Bakheet T, Williams B R G, Khabar K S A (2003). ARED 2.0: An update of AU-rich element mRNA database. *Nucl. Acids Res.* 31:421–423.
36. Lai W S, Blackshear P J (2001). Interactions of CCCH zinc finger proteins with mRNA: Tristetraprolin- mediated AU-rich element-dependent mRNA degradation can occur in the absence of a poly(A) tail. *J. Biol. Chem.* 276:23144–23154.
37. Sawaoka H, Dixon D A, Oates J A, et al. (2003). Tristetraprolin binds to the 3′-untranslated region of cyclooxygenase-2 mRNA. a polyadenylation variant in a cancer cell line lacks the binding site. *J. Biol. Chem.* 278:13928–13935.
38. Sully G, Dean J L, Wait R, et al. (2004). Structural and functional dissection of a conserved destabilizing element of cyclo-oxygenase-2 mRNA: evidence against the involvement of AUF-1 [AU-rich element/poly(U)-binding/degradation factor-1], AUF-2, tristetraprolin, HuR (Hu antigen R) or FBP1 (far-upstream-sequence-element-binding protein 1). *Biochem. J.* 377:629–639.
39. Yu H, Stasinopoulos S, Leedman P, et al. (2003). Inherent instability of plasminogen activator inhibitor type 2 mRNA is regulated by tristetraprolin. *J. Biol. Chem.* 278:13912–13918.
40. Taylor G A, Carballo E, Lee D M, et al. (1996). A pathogenetic role for TNF alpha in the syndrome of cachexia, arthritis, and autoimmunity resulting from tristetraprolin (TTP) deficiency. *Immunity.* 4:445–454.
41. Phillips K, Kedersha N, Shen L, et al. (2004). Arthritis suppressor genes TIA-1 and TTP dampen the expression of tumor necrosis factor alpha, cyclooxygenase 2, and inflammatory arthritis. *Proc. Natl. Acad. Sci. U. S. A.* 101:2011–2016.

# PROTEIN ENGINEERING WITH NONCODED AMINO ACIDS: APPLICATIONS TO HIRUDIN

VINCENZO DE FILIPPIS
*University of Padova, Padova, Italy*

**Chapter Contents**

1 Introduction
2 Hirudin–Thrombin Interaction
3 Selection of Amino Acid Replacements
  3.1 Val1
  3.2 Ser2
  3.3 Tyr3
4 Structure-Activity Relationships
  4.1 Val1→X
  4.2 Ser2→X
  4.3 Tyr3→X: Effect of Side-Chain Volume and Hydrophobicity
  4.4 Tyr3→X: Orientation Effects
  4.5 Tyr3→X: Electronic Effects
5 Cumulative Amino Acid Substitutions in Hirudin Yield a Highly Potent Thrombin Inhibitor
6 Structural Mapping of Thrombin Recognition Sites in the $Na^+$-Bound and $Na^+$-Free Form
7 Incorporation of Noncoded Amino Acids as Spectroscopic Probes in the Study of Protein Folding and Binding
  7.1 7-Azatryptophan
  7.2 3-Nitrotyrosine
8 Concluding Remarks
  Acknowledgment
  References

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

## 1 INTRODUCTION

During the past decade, the incorporation of noncoded amino acids into proteins has emerged as a novel and promising approach in protein science. In its infancy, the approach of protein engineering was essentially restricted to the possibility to chemically modify, in a rather unspecific fashion, particular amino acid side chains in proteins [1]. More recently, the advent of recombinant DNA technology allowed the site-specific alteration of a given polypeptide chain at a glance, thus, greatly expanding the tools available for studying the molecular mechanisms of protein folding, stability, and function [2]. Nevertheless, a quantitative description of the physical and chemical basis that makes a polypeptide chain efficiently fold into a stable and functionally active conformation is still elusive [3]. This mainly originates from the fact that nature combined, in a yet unknown manner, different properties (i.e., hydrophobicity, conformational propensity, polarizability, and hydrogen bonding capability) into the 20 standard protein amino acids, thus making it difficult, if not impossible, to univocally relate the change in protein stability or function to the variation of physico-chemical properties caused by amino acid exchange(s). In this view, incorporation of noncoded amino acids with tailored side chains, allowing investigators to finely tune the structure at a protein site, would facilitate to dissect the effects of a given mutation in terms of one or a few physico-chemical properties, thus, greatly expanding the scope of physical-organic chemistry in the study of proteins [4].

Incorporation of noncoded amino acids has been widely exploited in the study of bioactive peptides [5]. However, the results of these studies provide only a qualitative picture of the mechanism of the ligand–receptor interaction, and thus, they are of limited predictive power, as also documented by the explosion in the last few years of serendipity-based approaches in peptide design and drug discovery [6]. These difficulties stem primarily from the fact that introduction of single or multiple amino acid exchanges into a short peptide is expected not only to alter the interaction energy at the binding site(s) of the receptor, but also to affect (in a yet unpredictable way) fundamental properties of the free ligand, including electrostatic potential, hydration energy, and conformational entropy. Contrary to short peptides (5–20 amino acids), longer polypeptide chains (>50 amino acids) are generally characterized by a well-defined and stable three-dimensional (3D) structure, thus serving as macromolecular scaffolds, by keeping the global properties of the molecule rather constant and allowing the changes in binding free energy to be related in a more predictable way to the local variations of the physico-chemical properties at the mutation site(s). Hence, the incorporation of noncoded amino acids into proteins represents a further, (almost) obligatory extension of the studies aimed at elucidating the relationships existing among the structure, stability, and function in proteins (i.e., protein engineering).

In this view, several different strategies have been pursued to incorporate noncoded amino acids, including peptide synthesis [7] native chemical ligation [8, 9], enzyme-catalyzed semisynthesis [10, 11], biosynthetic incorporation *via* auxotrophic bacterial strain expression [12, 13], and nonsense suppression methodologies in cell-free [4] or whole-cell [14] expression systems. Both native chemical ligation and enzymatic semi-synthesis require careful tailoring of the experimental procedures, whereas biosynthetic incorporation methods are not

site-specific and are restricted to those cases when the structure of the noncoded amino acid (e.g., p-fluorephenylalanine, selenomethionine, and 7-zatryptophan) is similar to that of the natural counterpart to be replaced. On the other hand, despite significant advancements during the last decade [15–17], genetic methods are still limited by exceedingly low amounts of the resulting mutant protein, usually micrograms, that impair a thorough structural and functional characterization. However, stepwise solid-phase chemical synthesis remains the easiest and fastest approach to site-specifically incorporate in high yields any noncoded amino acid into even long (50–80 amino acids) polypeptide chains approaching the size of real proteins [18].

Relevant applications from our laboratory, regarding protein engineering with noncoded amino acids as a tool in the study of protein folding and function, will be presented. In particular, we will discuss the use of noncoded amino acids in structure-activity relationship (SAR) studies of hirudin binding to thrombin, as well as the incorporation of noncoded analogs of tryptophan and tyrosine (i.e., 7-azatryptophan and 3-nitrotyrosine) as spectroscopic probes for studying the hirudin–thrombin interaction.

## 2 HIRUDIN–THROMBIN INTERACTION

Thrombin is a serine protease of the chymotrypsin family that plays a key role at the interface among coagulation, infammation, and cell growth [19], and it exerts either procoagulant or anticoagulant functions in hemostasis [20]. The procoagulant role entails conversion of fibrinogen into fibrin and platelets activation, whereas the anticoagulant role regards the activation of protein C [21]. The most effective modulator of thrombin function in solution is $Na^+$, which triggers the transition of the enzyme from an anticoagulant (*slow*) form to a procoagulant (*fast*) form [22–24]. The $Na^+$-bound (*fast*) form displays procoagulant properties, because it cleaves more specifically fibrinogen and the protease-activated receptor (PAR), whereas the $Na^+$-free (*slow*) form is anticoagulant because it retains the normal activity toward protein C but cannot promote acceptable hydrolysis of procoagulant substrates [23, 24]. The importance of $Na^+$ on thrombin function is outlined by the fact that under physiological conditions the two forms are almost equally populated and that natural thrombin variants with compromised $Na^+$ binding result in bleeding phenotypes [24]. In this view, molecules capable of affecting the equilibrium between these two forms can have great potential therapeutic impact on the treatment of coagulative disorders [25].

Among natural anticoagulants, hirudin is the most potent and specific inhibitor of thrombin, with a dissociation constant in the 20–200-fM range [26, 27]. Due to its important pharmacological implications, the hirudin–thrombin pair has been the object of thorough structural and biochemical investigations. With respect to this, structural studies conducted on hirudin in the free [28–30] and thrombin-bound state [31] indicate that this inhibitor is composed of a compact N-terminal region, encompassing residues 1–47 and cross-linked by three disulfide bridges, and a flexible negatively charged C-terminal tail that binds to the fibrinogen-recognition site (exosite I) on thrombin. The N-terminal domain covers the active site of thrombin and through its first three amino acids extensively penetrates into the specificity

pockets of the enzyme (see Figure 1). Notably, the N-terminal tripeptide makes about half of the total contacts observed for the binding of the core domain 1–47 to thrombin (PDB code 4htc.pdb) [31] and accounts for ~30% of the total free energy of binding [33]. By limited proteolysis of full-length hirudin, we could produce the peptide fragment corresponding to the N-terminal domain 1–47 of hirudin HM2 [27]. Although far less active (at least $2 \times 10^5$-fold) than intact hirudin, this fragment (like the parent hirudin molecule) binds ~30-fold more tightly to the *fast* form of thrombin than to the *slow* form [34], which suggests that the structural determinants for this behavior are stored in the N-terminal domain. Hence, mutational studies on hirudin fragment 1–47 can be useful not only to investigate the physico-chemical determinants responsible for the extraordinary

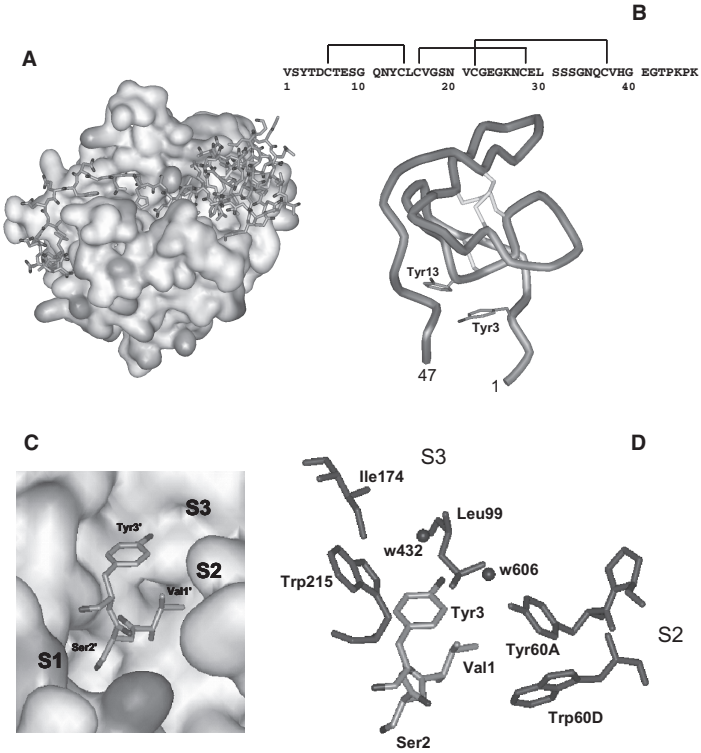


**Figure 1.** (**A**) Schematic representation of the interaction of full-length hirudin (stick) with thrombin (van der Waals surface), based on the crystal structure of the hirudin–thrombin complex [31]. (**B**) Amino acid sequence [32] and schematic representation of the solution structure [30] of the N-terminal fragment 1–47 of hirudin HM2 from *Hirudinaria manillensis*. Disulfide bonds are indicated by plain lines. (**C–D**) Structural details of the interaction of the N-terminal tripeptide of hirudin with thrombin. Val1′ of hirudin contacts the S2 site of thrombin, shaped by Tyr60a and Trp60d; Ser2′ covers, but does not penetrate, the S1 site, containing Asp189 at the bottom; Tyr3′ fills the apolar S3 site, formed by Trp215, Leu99, and Ile174. Structural water molecules at the hirudin–thrombin interface in the S2–S3 sites are also indicated. The ribbon drawing was generated using the program WebLab Viewer-Pro vs. 4.0 (Molecular Simulations Inc., 2000).

affinity and specificity of hirudin for thrombin, but also to probe the structural properties of the enzyme in the *slow* or *fast* forms.

## 3 SELECTION OF AMINO ACID REPLACEMENTS

### 3.1 Val1

Nuclear magnetic resonance (NMR) studies indicate that Val1 is almost fully exposed to solvent and highly flexible in the free hirudin [29, 30, 35]. Conversely, in the thrombin-bound state, Val1 is completely buried into the active site of the enzyme and fixed into a single side-chain conformation [31]. In particular, the α-$NH_2$ group of Val1 forms a hydrogen bond with the Oγ of the catalytic Ser195, whereas its side chain makes numerous hydrophobic contacts with Tyr60A and Trp60D, shaping the S2 of the enzyme (Figure 1). This loop, absent in other homologous trypsin-like proteases, defines the S2 specificity site on thrombin and narrows the access to the active site such that only small-sized apolar residues are allowed.

Considering the structural properties of hirudin in the free and bound state and the steric requirements at the S2 site of thrombin, we shaved the Val1 side chain with Ala or replaced it with *tert*-butylglycine (*t*Bug) (Figure 2). In fact, both Val and *t*Bug have comparable side-chain volume [36] and strong β-forming propensities [37], with minimum energy points at $\phi/\psi = -90°, 100°$ for Val and $\phi/\psi = -130°$,

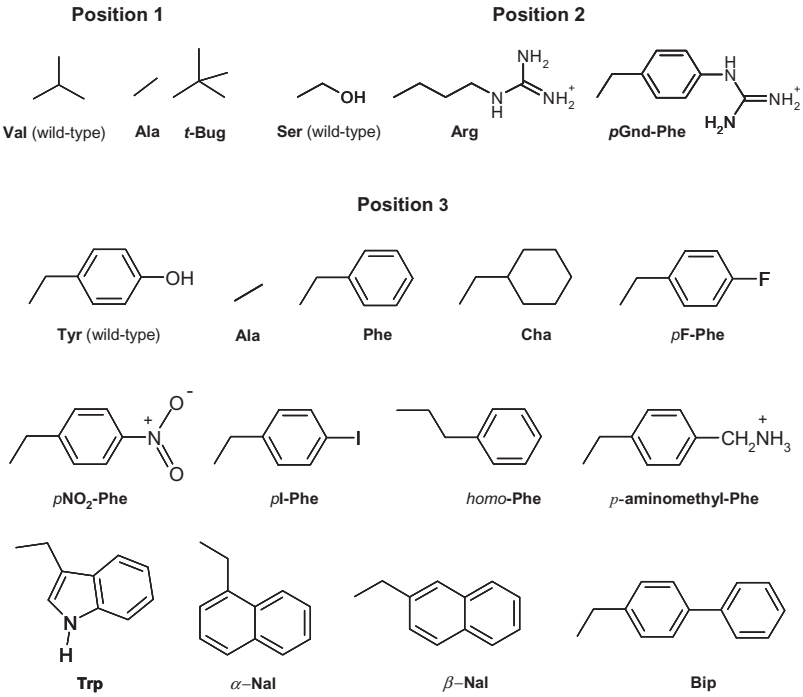


**Figure 2.** Amino acid replacements introduced into the hirudin fragment 1–47.

140° for *t*Bug [37]. On the other hand, the addition of a methyl group to the Cβ of valine is expected to restrict the backbone conformations available to *t*Bug to about one third of those allowed to Val [37].

### 3.2 Ser2

Analysis of the 3D structure of hirudin thrombin complex reveals that position 2 of hirudin is located at the entrance to, but does not enter, the primary specificity (S1) site [31]. Hence, the occupancy of the S1 site is not strictly required for binding. To evaluate the effects of perturbation of Asp189, located at the bottom of the S1 site, on the affinity of hirduin fragment 1–47 to thrombin, we have replaced Ser2 with Arg and its more rigid analog, *p*-guanidophenylalanine (*p*Gnd-Phe).

### 3.3 Tyr3

Contrary to the high conformational flexibility observed for the first two residues of hirudin, Tyr3 has a well-defined conformation in both the free and the bound state [29–31]. In the hirudin–thrombin complex, the side chain of Tyr3 projects into the apolar binding site of thrombin (the S3 site), which is formed by a large hydrophobic cavity comprising residues Trp215, Leu99, and Ile174 [31]. The importance of position 3 is confirmed by the fact that Tyr3 is highly conserved through the hirudin family [38]. To probe the S3 site of thrombin, we introduced at position 3 of hirudin 1–47 relatively large structural and chemical diversity by systematically replacing Tyr3 with natural and non-natural amino acids having different side-chain volume, hydrophobicity, electronic, and conformational properties (Figure 2).

## 4 STRUCTURE-ACTIVITY RELATIONSHIPS

Chemical synthesis of peptide analogs of the N-terminal domain 1–47 of hirudin was carried out by a combination of manual and automated solid-phase synthesis using the synthetic strategy previously described [34, 39]. The crude peptides with the Cys-residues in the reduced state were allowed to fold (2 mg/mL) under air-oxidizing conditions in bicarbonate buffer, pH 8.3, in the presence of 100-μM β-mercaptoethanol [39]. As an example, the reversed phase (RP)–high-performance liquid chromatography (HPLC) analyses of the crude synthetic analog Tyr3*homo*-Phe in the reduced and disulfide-oxidized state are reported in Figure 3. The chemical identity of the disulfide folded peptides was established by N-terminal sequence analysis and electrospray ionization (ESI)–time-of-flight (TOF) mass spectrometry, and for some synthetic peptides, the exact topology of disulfide bonds was established by enzymatic fingerprint analysis [39]. All peptides were purified by preparative RP–HPLC, lyophilized, and used for subsequent conformational and functional characterization. The results of conformational characterization, conducted by far- and near-Ultraviolet (UV) circular dichroism (CD) indicated that amino acid exchanges do not appreciably affect the conformation of hirudin fragment 1–47, allowing us to interpret the differences in affinity to thrombin exclusively on the basis of the variation of the physico-chemical properties at

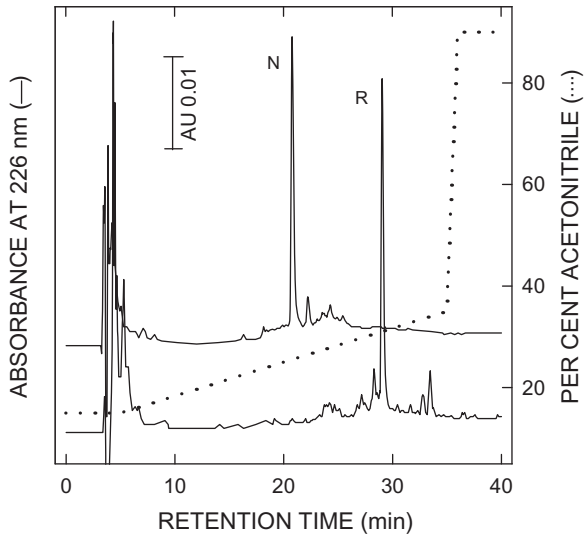


**Figure 3.** RP–HPLC analysis of the crude synthetic analog Tyr3*homo*-Phe in the reduced (*lower trace*) and disulfide oxidized state (*upper trace*). The chromatographic separations were conducted on a Vydac C18 analytical column (4.6 × 250 mm, 5 μm) eluted with a linear acetonitrile-0,05% TFA (. . . .) gradient, at a flow rate of 0.8 mL/min.

the mutation site. The inhibitory potency of the synthetic analogs toward the procoagulant *(fast)* and anticoagulant *(slow)* form of thrombin was determined by measuring at 405 nm the release of *p*-nitroaniline from the synthetic substrate D-Phe-Pro-Arg-*p*-NA, under temperature (25°C) and salt conditions in which the enzyme predominantly (>90%) exists in the *fast* (0.2-M NaCl) or in the *slow* (0.2-M choline chloride ChCl) form [22, 23] (Table 1).

### 4.1 Val1 → X

Shaving of Val1 with Ala (Val1Ala) reduces affinity for the *fast* and *slow* form of thrombin by 15- and 3-fold, respectively. On the other hand, replacement of Val1 with *tert*-butylglycine (*t*Bug) enhances the affinity of fragment 1–47 for the fast form of thrombin by about 3-fold, with a gain in the free energy of binding ($\Delta G_b$) of 0.65 kcal/mol. This result is unprecedented, considering that almost all mutations at this position reported so far dramatically decrease binding [33, 41].

Model building studies indicate that *t*Bug fits snugly into the S2 subsite of the enzyme without steric hindrance and buries approximately the same amount of apolar surface as Val1 upon binding to thrombin. Likely, the enhanced binding of V1tBug is due to the higher symmetry of the *tert*-butyl group of *t*Bug over the isopropyl side chain of Val. In the case of *t*Bug, three energetically equivalent side-chain rotamers of *t*Bug can bind thrombin into the functionally active conformation, leading to a reduction of the entropy change of binding compared with Val ($\Delta S_{b,\mathrm{Val}\to t\mathrm{Bug}}$), which binds thrombin in only one totamer (i.e., the *trans* rotamer). An estimate of $\Delta S_{b,\mathrm{Val}\to t\mathrm{Bug}}$ can be obtained from the equation $\Delta S_{b,\mathrm{Val}\to t\mathrm{Bug}} = R\cdot\ln(\gamma_{b,t\mathrm{Bug}}/\gamma_{b,\mathrm{Val}}) = R\cdot\ln(3/1)$ [42], where $\gamma_{b,t\mathrm{Bug}}$ and $\gamma_{b,\mathrm{Val}}$ are the number of functionally active conformations allowed to *t*Bug and Val in the hirudin–thrombin

complex, respectively. Thus, a relative stabilization of $\Delta G_{\mathrm{Val}\rightarrow t\mathrm{Bug}} = -T \cdot \Delta S_{\mathrm{Val}\rightarrow t\mathrm{Bug}} = -0.65$ kcal/mol at 298K is expected for the Val1*t*Bug analog over the natural species, in excellent agreement with the experimental value ($\Delta G_{\mathrm{Val}\rightarrow t\mathrm{Bug}} = -0.63$ kcal/mol) reported in Table 1. Our results suggest that proper introduction of symmetric groups into amino acid side chains can significantly improve binding by increasing the entropy of the ligand in the bound state, thus reducing the overall change in $\Delta S_b$. With this respect, amino acid substitutions like valine → *tert*-butylglycine or leucine → *tert*-butylalanine represent safe mutations, enabling us to improve binding with minimal steric requirements.

### 4.2 Ser2 → X

Replacement of Ser2 with Arg or its more rigid analog *p*Gnd-Phe induces a strong enhancement of the affinity of fragment 1–47 for thrombin (see Table 1), which is partly anticipated by modeling studies showing that the Arg side chain can be electrostatically coupled to Asp189, positioned at the bottom of the S1 site. Notably, perturbation of Asp189 enhances the affinity of hirudin preferentially for the *slow* form of the thrombin (see below).

### 4.3 Tyr3 → X: Effect of Side-Chain Volume and Hydrophobicity

The effects of coarse variations in the side-chain volume at position 3 on the binding to the *fast* or *slow* form were investigated by replacing Tyr3 with a smaller amino acid, like Ala, and with much larger residues, like Trp, α- and β-naphthylalanine (αNal and βNal), and biphenylalanine (Bip) (see Figure 2) [34, 43]. Thrombin dinding data, reported in Table 1, indicate that shaving of Tyr3 at Cβ reduces affinity almost exclusively for the *fast* form of thrombin, whereas enlargement of the side-chain volume at position 3 enhances binding in all cases, but preferentially to the procoagulant *fast* form of the enzyme. The lower affinity of Tyr3Ala can be reasonably explained either by the lower hydrophobicity of Ala or by the loss of numerous van der Waals contacts with the S3 site of the enzyme. Modeling of the Tyr3Ala–thrombin complex reveals that shaving of the Tyr3 side chain creates a cavity of about 150 $Å^3$ at the inhibitor–enzyme interface within the S3 region, which allows penetration of water at the binding interface, with a resulting destabilization of the hirudin–thrombin complex. Increasing the side-chain volume at position 3 also increases the hydrophobic effect, due to the burial of larger apolar surface upon binding. Strikingly, replacement of Ala with Bip enhances the affinity of hirudin 1–47 for thrombin by more than $1.3 \times 10^4$ times.

Clearly, thrombin binding data indicate that hydrophobicity is a major driving force for interaction. However, in the following we show that, beyond hydrophobicity, both side-chain orientation and electronic effects also play an important role in binding.

### 4.4 Tyr3 → X: Orientation Effects

As shown in Table 2, Trp is less hydrophobic than αNal. Nevertheless, the corresponding 1–47 analogs Tyr3Trp and Tyr3αNal display very similar inhibitory activity. Conversely, although αNal and βNal have the same hydrophobicity value,

**TABLE 1. Thermodynamic Data for the Binding of Synthetic Analogs of Fragment 1–47 to the Procoagulant (*fast*) and Anticoagulant (*slow*) Form of Thrombin**[1]

| Side Chain | *Fast form* | | *Slow form* | | |
|---|---|---|---|---|---|
| | $K_d$ (nM) | $\Delta\Delta G_b$ (kcal/mol) | $K_d$ (nM) | $\Delta\Delta G_b$ (kcal/mol)[2] | $\Delta G_c$ (kcal/mol)[3] |
| | | | Position 1 | | |
| Val (wild-type) | 42 ± 0.5 | — | 1460 ± 20 | — | –2.12 |
| Ala | 630 ± 7 | 1.61 | 4080 ± 40 | 0.61 | –1.10 |
| *t*Bug | 14 ± 0.4 | –0.63 | 1000 ± 80 | –0.24 | –2.5 |
| | | | Position 2 | | |
| Ser (wild-type) | 42 ± 0.5 | — | 1460 ± 20 | — | –2.12 |
| Ser2Arg | 1.7 ± 0.02 | –1.90 | 12 ± 2 | –2.84 | –1.16 |
| Ser2*p*GdnPhe | 2.5 ± 0.1 | –1.67 | 16 ± 2 | –2.67 | –1.10 |
| | | | Position 3 | | |
| Tyr (wild-type) | 42 ± 0.5 | — | 1460 ± 20 | — | –2.12 |
| Ala | 2750 ± 23 | 2.47 | 2600 ± 30 | 0.32 | 0.03 |
| Phe | 4 ± 0.1 | –1.37 | 203 ± 3 | –1.18 | –2.31 |
| Cha | 49 ± 0.6 | 0.09 | 1740 ± 23 | –0.09 | –2.12 |
| *p*F-Phe | 18 ± 0.3 | –0.48 | 583 ± 6 | 0.56 | –2.04 |
| *p*$NO_2$-Phe | 185 ± 3.0 | 0.88 | 2035 ± 20 | 0.22 | –1.46 |
| *p*I-Phe | 24 ± 0.2 | –0.33 | 1020 ± 12 | –0.23 | –2.22 |
| *homo*-Phe | 157 ± 2.5 | 0.79 | 3290 ± 38 | 0.46 | –1.79 |
| *p*AM-Phe | 140 ± 2.0 | 0.72 | 1590 ± 22 | 0.03 | –1.44 |
| Trp | 6.6 ± 0.7 | –1.08 | 550 ± 40 | –0.60 | –2.6 |
| αNal | 6.4 ± 0.4 | –1.10 | 530 ± 30 | –0.62 | –2.6 |
| βNal | 1.1 ± 0.1 | –2.14 | 94 ± 4 | –1.64 | –2.64 |
| Bip | 0.2 ± 0.01 | –3.17 | 12 ± 1 | –2.84 | 2.43 |
| BugArgNal | 15.0 ± 3 pM | –4.67 | 220 ± 50 pM | –5.23 | –1.6 |
| *r*-Hirudin HM2 | 0.20 ± 0.03 pM | –7.25 | 5.5 ± 0.6 pM | –7.36 | –1.9 |

[1] All measurements were carried out at 25°C in 5-mM Tris, pH 8.0, containing 0.1% PEG in the presence of 200-mM NaCl for the *fast* form or 200-mM ChCl when the *slow* form was being studied.
[2] $\Delta\Delta G_b$ is the difference in the free energy change of binding to thrombin between the synthetic analog ($\Delta G_b^*$) and the natural fragment ($\Delta G_b^{wt}$): $\Delta\Delta G_b = \Delta G_b^* - \Delta G_b^{wt}$. A negative value of $\Delta\Delta G_b$ indicates that the mutated species binds more tightly to thrombin than the natural fragment. Errors are ±0.1 kcal/mol or less.
[3] $\Delta G_c$ is the free energy of coupling to thrombin measured as $\Delta G_c = \Delta G_{b,\text{fast}} - \Delta G_{b,\text{slow}}$ [40]. The value of $\Delta G_c$ is negative if the inhibitor binds to the *fast* form with higher affinity than to the *slow* form. Binding data for the triple mutated species BugArgNal and recombinant hirudin HM2 variant from *Hirudinaria manillensis* are also reported.

Tyr3βNal is sixfold more potent than Tyr3αNal. Model building studies and accessible surface area (ASA) calculations provide reasonable explanation for these results. In fact, both Trp and αNal have similar side-chain orientation and bury approximately the same amount of apolar surface area upon complex formation. In addition, the side-chain of the residue in position 3 points toward the S2 site of thrombin in both Tyr3Trp and Tyr3αNal. Conversely, the βNal isomer in the analog Tyr3βNal buries a larger amount of apolar surface area upon binding, and

**TABLE 2. Physico-Chemical Properties of Amino Acid Side Chains at Position 3 of Hirudin Fragment 1–47**

| Side Chain at Position 3 | Volume (Å$^3$)[1] | logP (π)[2] | F[3] | R[3] | μ[4] (Debye) |
|---|---|---|---|---|---|
| Tyr (wild-type) | 138 | 1.97 | 0.29 | –0.64 | –1.57 |
| | | (0.96) | | | 2.02[5] |
| Ala | 38 | 0.60 | | | 0.0 |
| | | (0.31) | | | |
| Phe | 127 | 2.73 | 0.00 | 0.00 | 0.36 |
| | | (1.79) | | | |
| Cha | 146 | 3.88 | | | 0.0 |
| | | (2.72) | | | |
| *p*F-Phe | 133 | 2.87 | 0.43 | 0.34 | 1.79 |
| *p*$NO_2$-Phe | 157 | 2.42 | 0.67 | 0.16 | 4.38 |
| | | (1.96) | | | |
| *p*I-Phe | 160 | 3.85 | 0.40 | –0.19 | 1.72 |
| *homo*-Phe | 147 | 3.15 | | | 0.39 |
| | | (2.10) | | | |
| *p*AM-Phe | 165 | –0.80 | | | n.a. |
| Trp | 170 | 2.60 | | | n.a. |
| | | (2.25) | | | |
| αNal | 180 | 3.87 | | | n.a. |
| | | (3.08) | | | |
| βNal | 180 | 4.00 | | | 0.44 |
| | | (3.15) | | | |
| Bip | 215 | 4.63 | | | n.a. |

[1] Side-chain molecular volumes are referred to the volume within van der Waals surface and are calculated for the corresponding organic compounds taken as suitable models for the amino acid side chains (e.g., *p*-cresol for Tyr, toluene for Phe).
[2] LogP values are those of the corresponding model compounds determined experimentally at 25°C and reported by Sangster [44]. When available, the hydrophobic substituent constants, π, of amino acid side chains are given in parenthesis [36].
[3] *F* and *R* are the field (*F*) and resonance (*R*) contribution to the Hammett electronic substituent constant, $\sigma_p$, according to Hansch et al. [45]. Electron-withdrawing groups possess positive F and *R* values, whereas electron-releasing groups have negative *R* values.
[4] μ is the electric dipole moment of organic model compounds of the amino acid side chains determined at 25°C in benzene [46].
[5] The μ value of Tyr was corrected for the absence of resonance effects. A positive sign of μ stands for a negative endpointing away from the phenyl ring, toward the *para*-X substituent.
*Abbreviation*: n.a., not available.
*Note*: A detailed description of the procedures employed to get estimates of side-chain physicochemical properties is reported in Ref 43.

it favorably interacts with in the S3 site (see Figure 1D). Further support to the importance of orientation effects to the binding of hirudin to thrombin comes from the low affinity of the analog of hirudin 1–47 in which Tyr3 was replaced by *homo*-Phe. Although *homo*-Phe is much more hydrophobic than Tyr and Phe (Table 2), the analog Tyr3*homo*Phe is less potent than the wild-type species and Tyr3Phe analogs by 4- and 40-fold, respectively (Table 1). Due to steric clashes, the side chain of *homo*-Phe cannot favorably interact with the apolar S3 site, but

more likely it points toward Tyr60A and Trp60D in the S2 site, leaving an uncompensated cavity at the S3 site.

An important aspect emerging from our work is that the intrinsic aversion of nonpolar groups for water is the dominant driving force for ligand binding only when the removal of these groups from the aqueous solvent leads to favorable specific interactions with the receptor binding-site(s), which suggest that ligand–receptor association is strongly influenced by both hydrophobic and shape-dependent packing effects.

### 4.5 Tyr3 → X: Electronic Effects

Aromatic–aromatic interactions have been identified as important factors for protein stability and binding [47]. Indeed, edge-to-face interaction between two aromatic side chains allows the $\delta^+$ hydrogen atoms of the edge of one aromatic ring to approach the $\delta^-$ π-electron cloud of the other ring, favorably contributing to protein stability and binding by 0.6–1.3 kcal/mol [48]. Analysis of the 3D structure of the hirudin–thrombin complex reveals that the $\delta^+$ hydrogens on the edge of the Tyr3 ring can favorably interact in a T-shaped conformation with the $\delta^-$ π-electron cloud of the aromatic ring of Trp215 at the S3 site on thrombin (see Figure 1D). To probe the importance of aromatic–aromatic interactions in the hirudin–thrombin system, we replaced Tyr3 with Phe and its saturated analog, cyclohexylalanine (Cha). Compared with Tyr or Phe, Cha is similar in size but significantly more hydrophobic (Table 2). Hence, if hydrophobicity were the dominant driving force for binding, Tyr3Cha would be expected to be a more potent inhibitor than the wild-type and Tyr3Phe derivatives. Conversely, the Tyr3 → Phe exchange improves affinity for thrombin by 10-fold, whereas saturation of the aromatic ring of Phe (by Phe → Cha substitution) reduces binding by a similar amount. These findings clearly indicate that the enhanced affinity of Tyr3Phe and wild-type species over Tyr3Cha is primarily due to the stabilizing interaction with Trp215, which is possible only for the aromatic side chain of Phe or Tyr and not for its saturated analog, Cha, thus demonstrating that, beyond hydrophobicity, electronic effects, and specifically aromatic–aromatic interactions, can markedly modulate binding.

With the aim to strengthen aromatic–aromatic interactions, we systematically replaced Tyr3 with noncoded amino acids retaining the aromatic nucleus of Tyr, as well as similar size and hydrophobicity, but possessing electron-withdrawing substituents on the aromatic ring (i.e., *p*-fluoro-, *p*-iodo-, *p*-nitro-Phe, *p*-aminomethyl-Phe; see Table 2). The presence of electron-attracting groups in the *para*-position of Phe3 was expected to enhance binding by making the hydrogens of the aromatic nucleus at position 3 more electron-deficient, thus reinforcing the aromatic electrostatic interaction with the π-electron cloud of Trp215. Contrary to expectations, electron-withdrawing substituents failed to improve thrombin binding. Conversely, their presence resulted in a significant reduction in the affinity by 4 (i.e., Tyr3*p*F-Phe) to 46-fold (i.e., Tyr3*p*$NO_2$-Phe) in respect to that of Tyr3Phe (Table 1). Rather surprisingly, the presence of *p*-aminomethyl-Phe (*p*AM-Phe), carrying a net positive charge at position 3 ($pK_a$ 9.36), reduced affinity for the *fast* form of thrombin by only threefold and did not affect the binding strength to the *slow* form. Notably, the charged amino-group of *p*AM-Phe is shielded from water and buried into the apolar S3 site. Both theoretical and experimental work estimated an energy cost

of 9–10 kcal/mol to bury a charge in the protein interion [49, 50]. In our case, the difference of hydrophobicity (see Table 2) between *p*-aminomethyl-Phe and Tyr or Phe would yield an unfavourable increase of $\Delta G_b$ to thrombin by 3.8 and 4.8 kcal/mol, respectively. Therefore, the low energetic penalty of 0.72 and 0.03 kcal/mol experimentally derived for the binding of Tyr3*p*AM-Phe to the *fast* or *slow* form of the enzyme would lead us to reconsider the apolar properties of the S3 site or, more reasonably, to invoke specific interactions of the charged amino group within the S3 site (see below) that would compensate for its unfavorable desolvation properties.

Given the apolar character of the S3 site, traditionally referred to as the "aryl binding site" of thrombin [51], we first attempted to relate the affinity of the synthetic analogs of hirudin fragment 1–47 for the fast form of thrombin to the hydrophobicity value of the amino acid side chain at position 3. The contribution of hydrophobicity to the free energy change of binding ($\Delta G_b$) was estimated using the approach of Eisenberg and McLachlan [52], which assumes that desolvation free energy change of binding, $\Delta G_{desolv} = G_{desolv}(\text{complex}) - [G_{desolv}(\text{thrombin}) + G_{desolv}(\text{hirudin})] = -\Sigma_i \Delta\sigma_i \times \Delta\text{ASA}_i$, is related to the amount of polar and apolar surface area of both ligand and receptor that becomes buried upon complex formation [53]. $\Delta\text{ASA}_i$ is the change in the ASA for the atom-type *i* upon binding of hirudin to thrombin, and $\Delta\sigma_i$ (i.e., the atomic solvation parameter) is the solvation free energy change per unit area of atom type *i* that becomes buried upon binding. Desolvation free energy change of binding for a mutated hirudin analog (M) was calculated relatively to that of the wild-type fragment 1–47 (WT) as $\Delta\Delta G_{desolv} = \Delta G_{desolv}(\text{M}) - \Delta G_{desolv}(\text{WT})$.

Data reported in Figure 4A clearly indicate that the experimental free energy change of $\Delta G_b$ is linearly related to the variation of desolvation free energy ($\Delta G_{desolv}$), which is calculated on theoretical grounds. Only two data points strongly deviate from the regression line, namely Tyr3*p*I-Phe and Tyr3*p*AM-Phe. In particular, these analogs were found to bind thrombin much more tightly than predicted by desolvation free energy calculations. Structural analysis of the corresponding complexes with thrombin reveals that both the iodine atom of *p*I-Phe and the charged nitrogen of *p*AM-Phe can productively interact with the aromatic nucleus of Trp215 at the S3 site, thus compensating (or overwhelming) their unfavorable desolvation free energy of binding. In the case of the Tyr3*p*I-Phe analog, the favorable interaction of iodine with Trp215 is predominantly driven by dispersive forces related to the high polarizability of iodine. In this regard, $I_2$ forms stable complexes with benzene [54]. On the other hand, the surprisingly high affinity of the Tyr3*p*AM-Phe analog can be rationalized by taking into account a specific charge–π interaction between the protonated amino group of *p*-aminomethyl-Phe and the π-electrons of Trp215. Since the seminal work of Burley and Petsko [47] on "weakly polar interactions," a great deal of experimental evidences have been accumulated over the years, leading to the conclusion that "nonconventional" hydrogen bonds of the type X–H . . . π and cation–π interactions can play a key role in protein structure and stability, as well as in ligand recognition [55]. With respect to this, theoretical calculations indicate that, among many other possible combinations of cations and aromatics, the interaction of $NH_4^+$ (or $K^+$) with indole provides the most favorable interaction energy [56]. These predictions have been confirmed by a comprehensive structural analysis of X–H . . . π hydrogen bonding in proteins, in which the indolyl

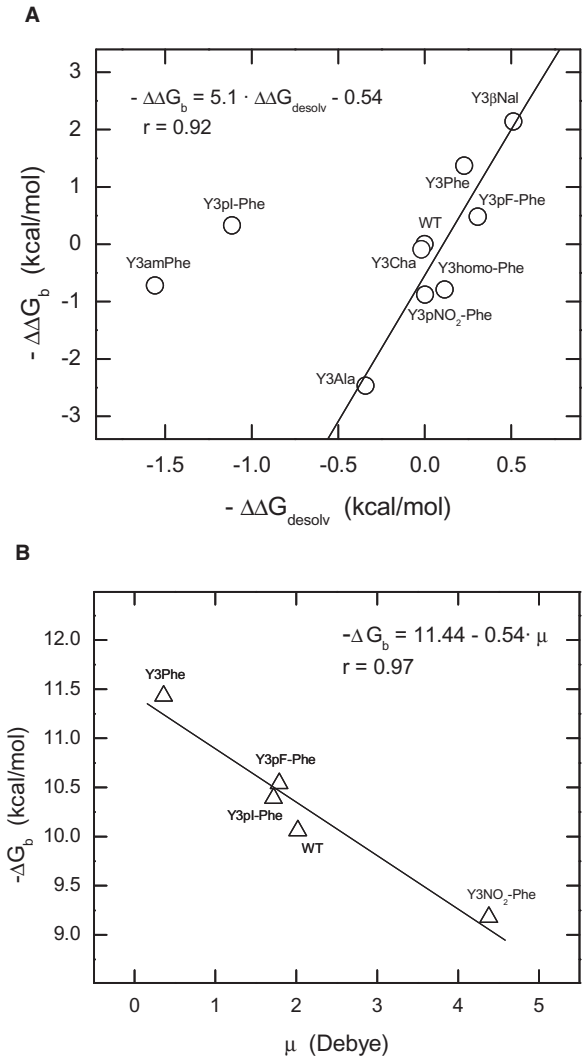


**Figure 4.** (**A**) Plot of the experimental free energy change of binding to the *fast* form of thrombin ($-\Delta\Delta G_b$) for the synthetic analogs of fragment 1–47 mutated at position 3 as a function of free energy change of desolvation ($\Delta\Delta G_{desolv}$) upon hirudin–thrombin interaction. (**B**) Plot of $\Delta G_b$ *versus* the electric dipole moment μ of the amino acid side chain at position 3. A positive sign of μ stands for a negative endpoint away from the phenyl ring toward the *para*-X substituent. For details, see Ref 43. **r** is the correlation coefficient of the linear fitting.

nucleus of Trp was found by far to be the most frequent π-acceptor of hydrogen and Lys-Trp to be the most represented donor–acceptor pair [57].

To better understand the physical nature of these forces driving the hirudin–thrombin interaction, we plotted the experimental $\Delta G_b$ of hirudin analogs *versus* the electric dipole moment, μ, of the *para*-X substituted (X = –H, –F, –I, –OH and $-NO_2$) amino acid side chain at position 3 (see Table 2). The data shown in Figure 4B indicate that the affinity of hirudin fragment 1–47 for thrombin is inversely

related to the value of μ, which suggests that hirudin–thrombin interaction is destabilized by the presence of a partial negative charge at the *para*-position of the phenyl ring, as in the case of *p*NO2-Phe, where the two oxygen atoms are strongly electron-dense. A reasonable explanation for these results may be found in the specific electrostatic interactions with the electronic π-system of Trp215, which for *p*-aminomethyl-Phe are attractive, whereas for *p*NO2-Phe are repulsive.

Taken together, our results indicate that the relative change in the affinity of hirudin 1–47 analogs can be accounted for by the apolar character of the ligand quite well, but on the other hand, they also emphasize the importance of "hidden interactions," mainly of electrostatic nature, in strengthening hirudin–thrombin binding. If a reasonable estimate of the hydrophobic effect can be obtained by calculating the free energy change due to desolvation of apolar surfaces that become buried upon ligand–receptor association, electrostatic effects are much more difficult to evaluate and predict, because they are mediated by weakly polar interactions, whose strength is strongly dependent on the electron density and polarizability, as well as on the orientation and distance of the interacting groups, such that even subtle perturbations in the ligand or receptor structure may dramatically alter binding. In this perspective, the possibility to introduce at a given protein site noncoded amino acids with larger structural and chemical diversity will improve our understanding of the mechanisms dictating molecular recognitions in proteins.

## 5 CUMULATIVE AMINO ACID SUBSTITUTIONS IN HIRUDIN YIELD A HIGHLY POTENT THROMBIN INHIBITOR

Hirudin offers numerous advantages over the existing anticoagulants heparins and cumarins [58, 59], and its use in clinical practice has been recently introduced. However, hirudin administration necessitates careful dose-titration, and bleeding effects are not rare [60]. These problems stem primarily from the intrinsic instability of the highly flexible C-terminal tail of full-length hirudin to degradation by endogenous proteases, generating truncated N-terminal fragments that, however, are dramatically less potent as thrombin inhibitors than the intact molecule [39, 61].

To minimize the hirudin sequence binding to thrombin and to improve its therapeutic profile, several N-terminal fragments of hirudin have been prepared as potential anticoagulants. Their use would provide a more predictable effect *in vivo* compared with intact hirudin, due to their stability to denaturants and proteolytic attack and lower immunogenicity [61]. Moreover, N-terminal core fragments targeting solely the active site of thrombin are expected to have a safer therapeutic profile, in keeping with the notion that active-site reversible inhibitors of thrombin display a better antithrombotic/hemorrhagic balance than bivalent inhibitors [62–64]. However, a major limitation of these fragments resides in their poor affinity for thrombin ($K_d$ = 30–400 nM), compared with hirudin ($K_d$ = 0.2–1.0 pM) [27, 33, 61].

To possibly obtain a mini-hirudin retaining the highly potent antithrombin activity of full length hirudin, but lacking the susceptibility to proteolysis, we selected the best-performing amino acid exchanges tested in our previous work (i.e., Val1*t*Bug,

Ser2Arg, and Tyr3βNal) [34] and combined them in the same analog, denoted as BugArgNal [65], which was synthesized in high yields by standard Fmoc chemistry. The introduction of only three mutations at the N-terminal end of hirudin fragment 1–47 yields a molecule that inhibits the *fast* or *slow* form by 2670- and 6820-fold more effectively than the natural counterpart and that binds exclusively at the active site of thrombin with an affinity ($K_{d,fast}$ = 15 pM, $K_{d,slow}$ = 220 pM) comparable with that of full-length hirudin ($K_{d,fast}$ = 0.2 pM, $K_{d,slow}$ = 5.5 pM). Strikingly, BugArgNal induced a change in the coagulative parameters (i.e., thrombin time, prothrombin time, and activated partial thromboplastin time) comparable with that evoked by intact hirudin (unpublished data). BugArgNal is also highly stable to low pH and resistant to the action of numerous proteases, including trypsin, chymotrypsin, thermolysin, and pepsin, and like full-lemgth hirudin, it displays almost absolute selectivity for thrombin over other closely related, physiologically important serine proteases, including plasmin, factor Xa, and tissue plasminogen activator, up to the highest concentration of inhibitor tested (10 μM). Only a slight inhibition was observed for factor Xa, with an estimated $K_d$ value higher than 8 μM.

The issue of protease selectivity is crucial in the design of novel thrombin inhibitors, because inhibition of other physiologically relevant serine-proteases (e.g., tissue plasminogen activator and plasmin) can impair their clinical use [66–68]. The results reported in this study demonstrate that the presence of an Arg-residue at the N-terminal end of hirudin fragment 1–47 strongly improves binding, while retaining the extraordinary selectivity of the natural product. This is in contrast with the results obtained for low-molecular-weight inhibitors, where it is paradigmatic that the introduction of a positive charge at P1-position improves binding, but strongly reduces selectivity [66–68], because the positive charge of the inhibitor interacts in the S1 site of thrombin with Asp189, which is highly conserved among the endogenous enzymes prevalent in the vascular system. Likely, in the case of BugArgNal, many weak favorable contacts operate at the inhibitor–thrombin interface to cooperatively encode protease specificity.

As shown in Table 1, the effects of the amino acid replacements are additive in both the *fast* and the *slow* forms of thrombin, indicating that S1, S2, and S3 sites behave independently; that is, perturbation of a given site of the enzyme does not affect the binding properties of the other two sites [69]. This result is of particular relevance, because it would allow the properties of a multiple mutant to be inferred directly from those of the singly mutated species and to engineer incremental increases in binding strength and selectivity for either allosteric form of thrombin. For instance, substitution of Val1 with *t*Bug, Ser2 with Arg, and Tyr3 with Bip would yield a synthetic analog about 15,750 more potent than the wild-type species, with a predicted $K_d$ value of about 2.5 pM.

## 6 STRUCTURAL MAPPING OF THROMBIN RECOGNITION SITES IN THE $NA^+$-BOUND AND $NA^+$-FREE FORM

The effect of $Na^+$ binding on thrombin function is allosteric in nature [23, 70], and several crystal structures of the enzyme with and without $Na^+$ bound have been recently reported [71–74]. The structures of the pseudo-wild-type thrombin mutant (Arg77aAla), reported by Di Cera et al. [71, 72], in the presence or absence of $Na^+$,

display only small changes in the side-chain orientation of Ser195 in the active site, Asp189 in the S1 site, Glu192 and Asp222 on the protein surface, and some rearrangement of the water molecules filling the S1 site. Conversely, the structures of the *fast* and *slow* form reported by Huntington et al. [73, 74] show significant differences at the level of the S2 and S3 sites, which in the *slow* form protrude onto the protein surface and limit the access to the catalytic pocket. In particular, the apolar cavity of the S3 site is restricted by protrusion of Trp215, which is possibly caused by reorientation of the underlying Phe227 and 168–182 disulfide bond. Partial unfolding of the $Na^+$ site is observed, with a collapse of the 148-loop onto the groove leading to the catalytic pocket. Furthermore, all putative structures of the *fast* and slow forms reported so far display numerous contacts between thrombin monomers in the crystal lattice [71–74], and therefore, the packing effects can also influence the thrombin structure in the crystal.

Hence, we decided to use the N-terminal hirudin domain 1–47 as a molecular probe of the solution conformation of thrombin recognition sites in the *fast* and *slow* form [75]. The guiding idea is that structural information on thrombin sites in the two allosteric forms can be gained from the physico-chemical properties of the mutated residue at position 1–3 of hirudin and from the effects of the perturbations introduced on the value of coupling free energy $\Delta G_c$, which is the difference in standard free energy of binding of the inhibitor to the *fast* and *slow* form of thrombin, $\Delta G_c = \Delta G^f - \Delta G^s$. Thrombin binding data reported in Table 1 can be summarized as follows: First, Ala-shaving at either position 1 or position 3 reduces affinity almost exclusively for the fast form; second, side-chain enlargement at position 3 with bulky and hydrophobic amino acids (i.e., βNal and Bip) strongly enhances affinity for both forms, but preferentially for the fast form; third, electrostatic perturbation of the primary specificity site S1, by Ser2 → Arg or Ser2 → *p*Gnd-Phe exchange, enhances affinity preferentially for the slow form. These data were analyzed within the theoretical framework of site-specific thermodynamics [75–77] and used to extract structural information on the specificity sites of thrombin in the two allosteric forms:

$$\begin{array}{ccccc} & & \Delta G^s & & \\ & S & \Leftrightarrow & SI & \\ \Delta G^0 & \Updownarrow & & \Updownarrow & \Delta G^1 \\ & F & \Leftrightarrow & FI & \\ & & \Delta G^f & & \end{array}$$

The *slow* (*S*) and *fast* (*F*) forms bind the inhibitor (*I*) with a standard free energy change $\Delta G^s$ and $\Delta G^f$, whereas $\Delta G^0$ and $\Delta G^1$ represent the free energy changes for switching from the *slow* to the *fast* form in the absence or presence of the inhibitor. The coupling free energy ($\Delta G_c$) for the cycle is given by the equation $\Delta G_c = \Delta G^f - \Delta G^s = \Delta G^1 - \Delta G^0$, where $\Delta G^f$ and $\Delta G^s$ can be determined experimentally. The preferential loss (or gain) in affinity of a mutated inhibitor for one of the two allosteric forms of the enzyme is a measure of the energetic contribution of the interactions being lost (or gained) upon mutation and provide strong, albeit indirect, means for identifying those regions on thrombin that have different structural features in the *slow* or *fast* form. On the other hand, mutations that affect both forms to the same extent reveal that the perturbations introduced in the inhibitor are important for binding to thrombin and that the site is not involved in the *slow* → *fast* transition

or, otherwise, that the entity of the perturbation introduced is too small to elicit a different behavior at that site in the two allosteric forms. In addition, because $Na^+$ binds the hirudin–thrombin complex with ~20-fold higher affinity than the free enzyme ($\Delta G^\circ \neq \Delta G^1$) [40], then from the linkage principles mentioned above, it follows that the inhibitor must bind with different affinity to the *slow* and *fast* forms ($\Delta G^f \neq \Delta G^s$), ruling out other possibilities.

In the case of Tyr3 → Ala and Val1 → Ala exchanges, elimination of the interactions of Tyr and Val side chains beyond the Cβ strongly reduces affinity for the *fast* form, by 65- and 15-fold, respectively, whereas it is practically ineffective on the *slow* form, which suggests that the presence of Tyr and Val is crucial to enhance binding exclusively to the procoagulant *fast* form. On the other hand, the presence of the larger side chain of βNal at position 3 enhances affinity for the *fast* form by about 40-fold and by only 15-fold for the *slow* form. Taken together these results suggest that the S3 site of thrombin in the procoagulant (*fast*) form is in a more open and accessible conformation in respect to the less forgiving structure it acquires in the anticoagulant *slow* form. Also consistent with our model are the effects of the replacement of Ser2 with Arg or *p*Gnd-Phe, whose long, charged side chain is expected to facilitate penetration of the inhibitor into thrombin recognition sites, to a greater extent in the case of the more closed *slow* form than in the case of the *fast* form of the enzyme, which is already accessible for binding. From the effects of amino acid substitutions on the affinity of fragment 1–47 for the enzyme allosteric forms, we conclude that the specificity sites of thrombin in the $Na^+$-bound form are in a more open and permissible conformation, compared with the more closed structure they assume in the $Na^+$-free form [75].

The structural picture of thrombin allosteric forms proposed above is consistent with detailed molecular dynamics (MD) simulations carried out for 18 ns in full explicit water [43], showing two well-defined conformational minima on the energy landscape. After about 5 ns, a concerted conformational transition, involving the S2/S3 sites, the 148-loop, and the fibrinogen binding site, leading the thrombin molecule from a more compact and closed form, which we propose can be related to the anticoagulant (*slow*) form in the $Na^+$-free state, to a more open and accessible conformation, which we propose can be related to the procoagulant (*fast*) form in the $Na^+$-bound state (see Figure 5). Moreover, the results of

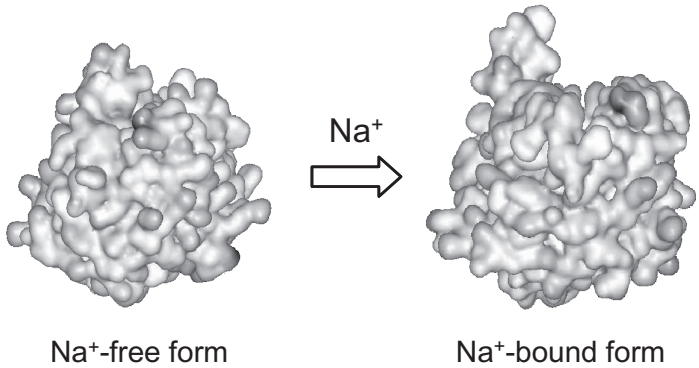


**Figure 5.** Surface representation of thrombin in the putative *slow* and *fast* form. The model was built on the basis of the coordinates derived form 18-ns MD simulations in explicit water solvent [43].

MD analysis outline the high degree of correlation existing between the motions of all these regions of thrombin, all occurring after about 5 ns, and suggest that a structural network is present, capable of communicating the conformational changes, induced by $Na^+$ binding, between different structural domains of the enzyme.

Our model also provides reasonable explanation for the fact that those substrates that are related to the procoagulant activities of thrombin (e.g., fibrinogen, PAR-1, and factor XIII) orient a bulky side chain deep into the S3 site of the enzyme and, as expected, are cleaved by the *fast* form of thrombin 20–40-fold more efficiently than by the *slow* form [23]. In particular, fibrinogen interacts at the S3 site of the enzyme through the bulky Phe8 having a side-chain volume of 127 $Å^3$ [1bbr.pdb; 78], PAR-1 through Leu38 (Leu = 100 $Å^3$) [1nrs.pdb; 79], and factor XIII through two Val-residues (Val29 and Val34; Val = 79 $Å^3$) [1de7.pdb; 80]. On the other hand, protein C, which is related to the anticoagulant function of thrombin, does not seem to extensively interact with the S3 site of the enzyme [43], and as expected, it is cleaved with similar specificity by either the *slow* or the *fast* form ($\Delta G_c = 0.2$ kcal/mol) [23].

## 7 INCORPORATION OF NONCODED AMINO ACIDS AS SPECTROSCOPIC PROBES IN THE STUDY OF PROTEIN FOLDING AND BINDING

A major application of protein engineering with noncoded amino acids regards the introduction into proteins of biophysical probes possessing physico-chemical properties (e.g., side-chain volume, hydrophobicity) similar to those of the corresponding natural amino acids, but spectral features distinct from those of the natural counterparts and highly sensitive to the chemical environment in which the probe is located [81]. Hence, by the use of the so-called "spectrally enhanced proteins" [82], it should be possible to effectively monitor the local structure and dynamics of the mutated protein during key events, such as protein folding and denaturation or ligand binding, without significantly perturbing the kinetics and equilibrium properties of the process under investigation. With respect to this, in a recent study, Cohen et al. could site-specifically introduce 6-dimethylamino-2-acyl-naphthylalanine (Aladan) into the B1 domain of staphylococcal protein G to obtain estimates of the local dielectric constant of the protein at different sites [83].

More specifically, the development of new spectroscopic tools for studying protein–protein interactions is central to many disciplines, including structural biology, biotechnology, and drug discovery [84]. Traditionally, the change in tryptophan (Trp) fluorescence has been exploited to study ligand–protein interactions [85]. However, the fluorescence signal of many proteins is insensitive to ligand binding [86], because fluorescence changes are mostly restricted to those cases where Trp-residues are embedded in the ligand–protein interface or when the ligand binding induces conformational changes in the protein, remote from the binding region and involving one or more Trp-residues. Furthermore, the presence of multiple tryptophans in proteins may lead to compensating effects that often complicate interpretation of the fluorescence data [85].

To overcome these problems, several extrinsic spectroscopic probes, characterized by well-defined spectral properties, have been covalently bound to protein

functional groups (i.e., Cys and Lys), to act as energy donors or acceptors in fluorescence resonance energy transfer (FRET) studies [84, 87–89]. This approach, however, is limited by possible labeling heterogeneity, nonquantitative modification, structural alteration of the proteins resulting from the labeling *per se*, and perturbation of the binding process, due to the large size of the fluorescent labels used [87, 88].

In the following, we show the utility of two noncoded analogs of tyrosine and tryptophan, namely 3-nitrotyrosine (NT) and 7-azatryptophan (AW), in the study of hirudin folding and binding to thrombin [90, 91].

### 7.1 7-Azatryptophan

Among the noncoded tryptophan analogs studied so far (i.e., 5-hydroxy- and 5-metoxy-Trp, benzo[b]thiophenylalanine and the spectrally silent fluorotryptophans), 7-azatryptophan (AW), an isostere of tryptophan (W), displays interesting absorption and fluorescence properties [92]. Unlike tryptophan, free AW displays single exponential fluorescence decay and the presence of a nitrogen-atom at position 7 in the indolyl-nucleus results in a red shift of 10 nm in the absorption and 46 nm in the emission of AW compared with Trp. Furthermore, the fluorescence $\lambda_{max}$ and quantum yield of 7-azaindole (7AI) are strongly influenced by the polarity of the chemical environment. In particular, on going from cyclohexane to water the emission fluorescence of 7AI is shifted from 325 to 400 nm and the quantum yield is decreased by 10-fold. The quantum yield of AW increases from 0.01 in aqueous solution, pH 7, to 0.25 in acetonitrile [93]. Hence, it should be possible to selectively excite the fluorescence of AW at the red edge of its absorption (between 310 and 320 nm), where Tyr does not absorb and the contribution of Trp is negligible and, thus, investigate protein folding and binding processes through variation of the AW fluorescence signal.

Recently, we exploited the unique spectroscopic properties of AW to probe the disulfide-coupled folding of the hirudin N-terminal domain 1–47 and the binding to its the target enzyme, thrombin [90]. Before chemical synthesis, the resolution of the commercially available enantiomeric mixture of AW was carried out by treating the racemic mixture with acetic anhydride and subsequent enantioselective deacylation with immobilized *Asperigillus oryzae* acylase-I to yield L-AW (Scheme 1). Purified L-AW was then reacted with 9-fluorenylmethoxycarbonyl chloride

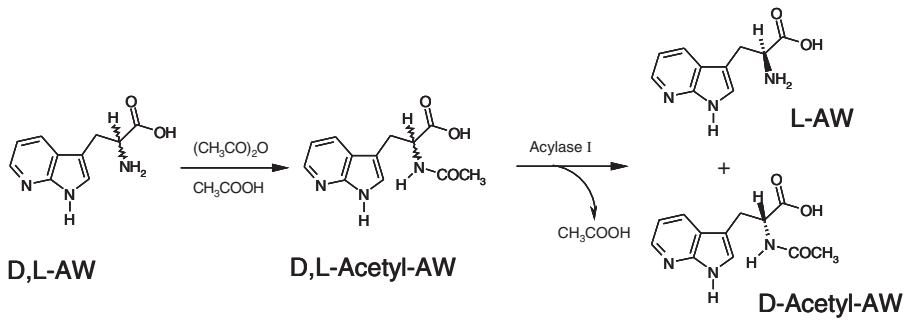


**Scheme 1.**

(Fmoc-Cl), to quantitatively obtain the Fmoc-derivative, which was subsequently used in the solid-phase synthesis of the analog of hirudin 1–47 in which Tyr3 was replaced by AW.

The replacement of tryptophan with the isosteric 7-azatryptophan leads to a significant reduction (~10-fold) in the affinity of fragment 1–47 for the *fast* form of thrombin, indicating that exchange of even a single atom (C → N) can substantially affect binding to the enzyme. These results can be explained by the lower hydrophobicity of 7-azatryptophan compared with that of tryptophan, as given by the values of octanol→water partition coefficient (logP) of indole ($\log P_{indole} = 2.33$) and 7-azaindole ($\log P_{azaindole} = 1.72$), determined experimentally [90].

The results of spectroscopic characterization indicate that the $\lambda_{max}$ values in the absorption and fluorescence spectra of Y3AW are red-shifted by ~10 and 40 nm, respectively, compared with those of Y3W (Figure 6). The fluorescence spectra of Y3AW with the six Cys-residues in the fully reduced or correctly folded state are shown in Figure 7 and compared with those of the corresponding Y3W analog. The emission spectrum of the reduced, unfolded Y3AW reveals the presence of two distinct, well-resolved bands at 305 and 397 nm, assigned to the contribution of Tyr13 and AW at position 3, respectively. In the folded state the Tyr-band disappears, whereas the fluorescence of AW is blue-shifted to 390 nm and enhanced by about 20%. Our results can be rationalized by considering that in the reduced state, hirudin fragment 1–47 is in a random coil conformation (unpublished results), with the donor (Tyr13) and acceptor (AW3) amino acids far apart in space. In the native state, the chain folding brings Tyr13 in close proximity to AW3 and allows the energy absorbed by Tyr to be efficiently transferred to AW. In the case of Y3W, a single band at 350 nm is observed in the native state, whereas in the reduced state the contribution of Tyr13 appears as a very weak shoulder at 303 nm, overwhelmed by the stronger emission of Trp3 at 355 nm (Figure 7C). This

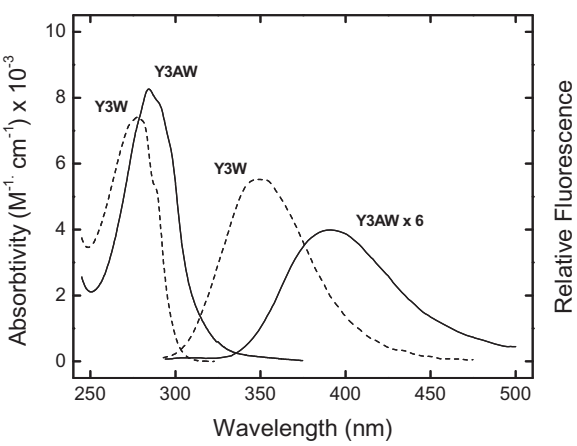


**Figure 6.** UV absorption and emission fluorescence spectra of Y3AW (—) and Y3W (--) analogs of hirudin fragment 1–47. The fluorescence spectrum of free Y3AW analog was multiplied by sixfold. All measurements were carried out at 25°C in 5-mM Tris-HCl buffer, pH 8.0, containing 0.2-M NaCl and 0.1% PEG-8000, at a protein concentration of 2 μM. Sample excitation was at 280 nm.

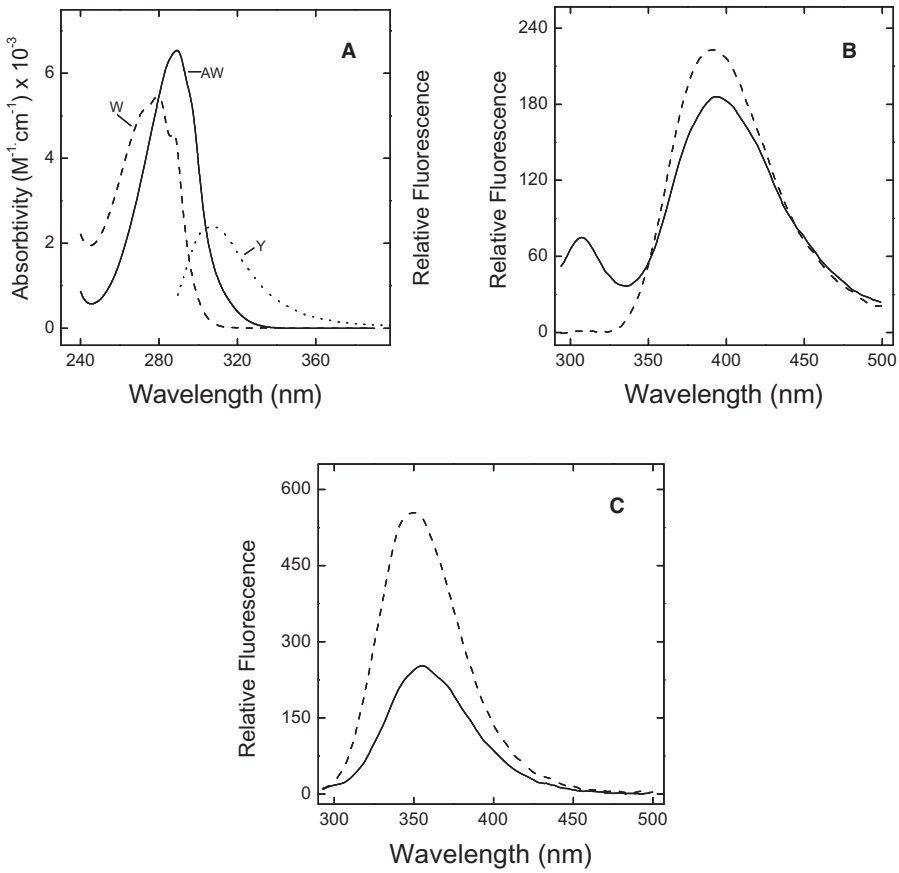


**Figure 7.** Disulfide oxidative folding of Y3AW analog monitored by fluorescence spectroscopy. (**A**) UV-absorption spectra of free tryptophan (W, --) and 7-azatryptophan (AW, —), acting as energy acceptors, and fluorescence spectrum of tyrosine (Y, · · ·), acting as an energy donor. Emission fluorescence spectra of Y3AW (**B**) and Y3W (**C**) in the reduced (—) and disulfide oxidized, native state (--). All spectra were taken at 25°C by exciting the samples (5 μM) at 280 nm in 0.1-M $NaHCO_3$, pH 8.3, except for that of the fully reduced form, which was recorded in 0.1-M morpholinoethane sulfonic acid buffer, pH 6.0. This pH value is sufficiently low to impair disulfide formation for at least 4 hours (not shown) and, concomitantly, avoid protonation of $N^7$-atom of the azaindole nucleus, which has a $pK_a$ value of 4.5 [94].

actually makes it difficult (if not impossible) to follow the folding process of hirudin by Tyr → Trp energy transfer measurements.

The strong dependence of the fluorescence properties of AW on solvent polarity [95] was exploited to investigate the binding of Y3AW analog to thrombin (Figure 8). To minimize the contribution of the nine Trp-residues present in the thrombin sequence, we excited AW at the red edge of its absorption range (320 nm), where the contribution of Trp is expected to be negligible [93]. The spectra reported in Figure 8 clearly indicate that the fluorescence of AW is strongly quenched upon binding to thrombin. Given the known solvent-dependent emission of AW and the apolar character of the S2/S3 binding sites of thrombin, this result is surprising. In

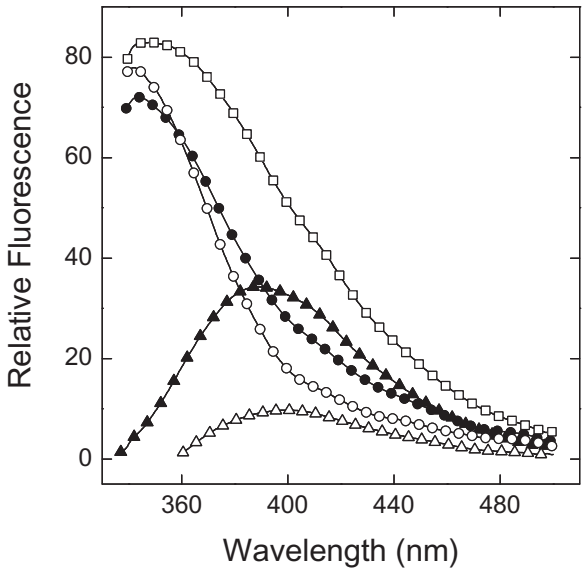


**Figure 8.** Hirudin–thrombin interaction probed by fluorescence spectroscopy. Fluorescence emission spectra of isolated thrombin (○–○) and Y3AW analog (▲–▲) and equimolar mixture of the enzyme and inhibitor (●–●). All spectra were recorded at 25°C in 5-mM Tris-HCl buffer, pH 8.0, containing 0.2-M NaCl and 0.1% PEG-8000, at a protein concentration of 2 μM, and by exciting the samples at 320 nm. Theoretical sum spectrum (□–□), obtained by adding the spectrum of free thrombin to that of the Y3AW analog. The contribution of Y3AW in the thrombin-bound form is estimated by the difference spectrum (△–△), obtained by subtracting the spectrum of isolated thrombin from that of enzyme-inhibitor complex. Given a $K_d$ value of 60 nM for the thrombin–Y3AW complex and a concentration of enzyme and inhibitor of 2 μM, it is expected that about 20% of Y3AW remains in the free form. Notably, the intensity of the difference spectrum accounts for 25% of that of the free inhibitor, which suggests that the fluorescence of Y3AW in the complex with thrombin is almost totally quenched.

fact, we would have expected that binding of Y3AW to thrombin is accompanied by a blue-shifted emission of AW and a substantial increase in its fluorescence intensity. The fluorescence spectra of the model compound 7-azaindole (7AI) in different solvents (Figure 9) well document the extraordinary dependence of the fluorescence signal of 7AI on the solvent polarity, shifting from 325 nm in cyclohexane (not shown) to 345 nm in diethylether and 362 nm in acetonitrile. In particular, in water-restricted environments (e.g., water-saturated diethyl ether), the fluorescence signal is dramatically reduced, due to the formation of a 1 : 1 7AI-$H_2O$ cyclic adduct that promotes formation in the excited state of a "tautomer" species that is poorly fluorescent [96–98].

In the light of these considerations, the strong quenching effect on the fluorescence of 7AW, observed upon binding of Y3AW to thrombin, can be explained on the basis of the model structure of Y3AW bound to thrombin. As shown in Figure 10, the –NH group of AW interacts with water molecule w432 (B-factor: 36 Å$^2$; occupancy: 1.0), whereas $N^7$ may be linked to Tyr60a through a water bridge involving the structural water molecule w606 (B-factor: 22 Å$^2$; occupancy: 0.52) [31].

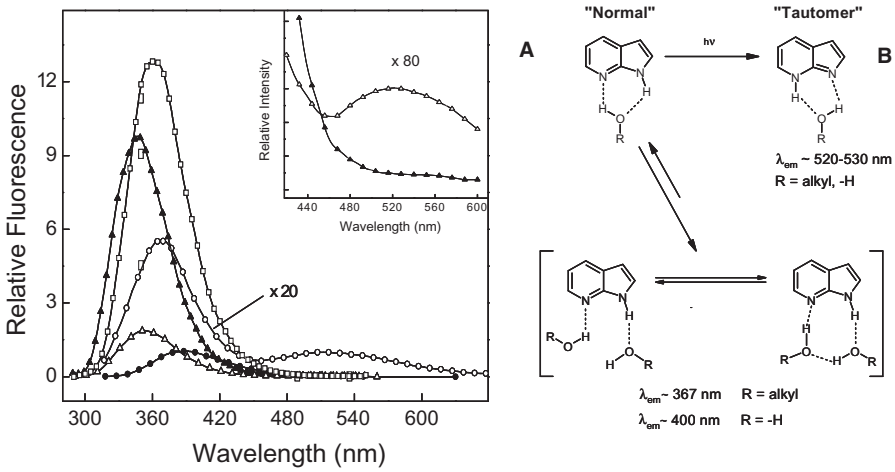


**Figure 9.** (**A**) Fluorescence emission spectra of 7-azaindole (7AI) in different solvents: acetonitrile (□–□), diethyl ether (▲–▲), *n*-propanol (○–○), water-saturated diethyl ether (△–△), and deionized water (●–●). The emission of 7AI in *n*-propanol is multiplied by 20-fold. *Inset*, fluorescence spectra of 7AI in diethyl ether (▲–▲) and water-saturated diethyl ether (△–△) expanded by a factor of 80 to show the growth of the second band at ~520 nm in organic-water solution. All spectra were recorded at 25°C and 2-μM 7AI, by exciting the samples at 280 nm. (**B**) Consensus scheme explaining the emission properties of 7AI in alcohol (R = alkyl) or water solution (R = –H). The polyhydrated species in water are more populated and emit radiation at 400 nm. In water-restricted environments (e.g., water-saturated diethyl ether), the concentration of the "normal" monohydrated species increases, and upon excitation, it is converted into the "tautomer" form, which is responsible for the longer wavelength emission at ~520 nm (*Inset* to panel **A**) [90].

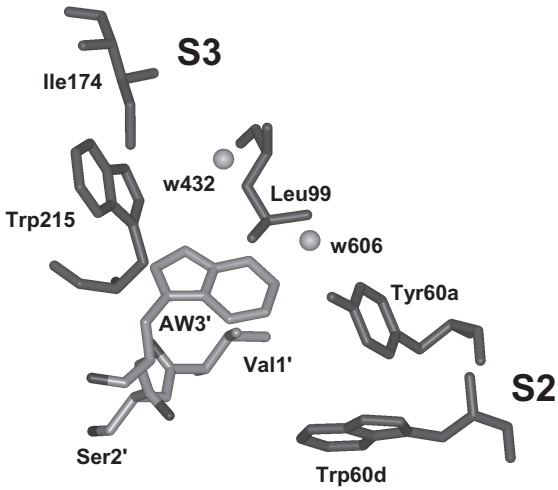


**Figure 10.** Schematic representation of the interaction of the N-terminal tripeptide of hirudin analog Y3AW with thrombin. The model structure of Y3AW was obtained by keeping the position of all atoms unchanged and using the same dihedral angles that Tyr3 has in the x-ray structure of wild-type hirudin complexed to thrombin ($\chi^1 = -61°$, $\chi^2 = -56°$, [4htc.pdb; 31]. Water molecule w432 (B-factor 36 Å$^2$, occupancy 1.0) is in the same plane of the azaindole ring of AW at position 3, whereas $N^7$ occupies the same position of the –OH group in Tyr3, which is linked to Tyr60a through a water bridge involving w606 (B-factor 22 Å$^2$, occupancy 0.52).

Hence, the rigid, structural water molecules at the hirudin–thrombin interface can have a crucial role in quenching the fluorescence of 7AW, because they promote the nonradiative decay of AW in the excited state more effectively than the labile water molecules solvating the 7AW in the free Y3AW analog.

In conclusion, our data demonstrate that the incorporation of 7-azatryptophan into proteins can be of broad applicability in structure-activity relationship studies, where a Trp-isostere is required, or as a spectroscopic probe in the study of protein folding and binding [90].

### 7.2 3-Nitrotyrosine

3-Nitrotyrosine (NT) is produced *in vivo* by reaction of protein tyrosines with peroxynitrite [99]. The NT side chain is only 30 $Å^3$ larger than the unmodified Tyr, and the presence of the electron-withdrawing nitro-group makes the phenolic hydrogen of free NT about $10^3$-fold more acidic ($pK_a$ 6.8) [100]. At pH < $pK_a$, where the neutral form is predominant, NT is more hydrophobic than Tyr, whereas at higher pH, where NT exists in the ionized form, it is much more polar [101, 102]. NT can form an internal hydrogen bond, and its absorption properties are strongly pH-dependent (see Figure 11). In particular, at basic pH, the UV/Vis spectrum of

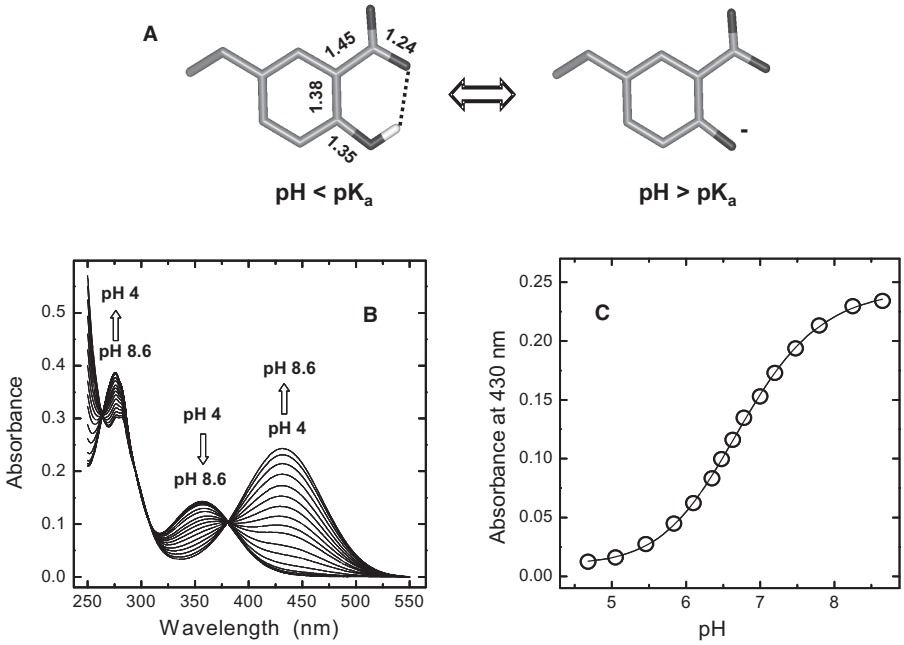


**Figure 11.** (**A**) Structure of 3-nitrotyrosine. Bond distances (Å) are taken from the crystallographic structure of free NT [103]. (**B–C**) Spectrophotometric titration of Y3NT by UV/Vis absorption spectroscopy. Absorption spectra (**B**) and plot of the absorbance values at 430 nm of Y3NT (2 mL, 48 μM) as a function of pH (**C**). Measurements were carried out in 2-mM citrate–borate–phosphate buffer, at the indicated pH. Fitting data points yields a $pK_a$ value of 6.74 ± 0.02 for NT.

free NT displays a major band at 422 nm, characteristic of the ionized form, whereas at acidic pH a prominent band appears at 355 nm, assigned to the contribution of the neutral form [100]. NT is essentially nonfluorescent and absorbs radiation in the wavelength range where both Tyr and Trp emit fluorescence, with a Trp-to-NT Förster's distance (i.e., the donor-acceptor distance at which the FRET efficiency is 50%) as large as 26 Å [104]. For these reasons, NT has great potential as an energy acceptor in FRET studies, and indeed, direct chemical nitration of Tyr was used to investigate the structural and folding properties of calmodulin [104] and apomyoglobin [105–107]. However, very little is known about the possibility of exploiting the unique spectral properties of NT to study molecular recognition [100, 108, 109].

Hence, we chose the hirudin–thrombin system as a suitable model for evaluating the potentialities of NT as a spectroscopic probe in the study of protein–protein interactions [91]. To this aim, we synthesized two analogs of the N-terminal domain (residues 1–47) of hirudin: Y3NT, in which Tyr3 was replaced by NT, and S2R/Y3NT, containing the cumulative substitutions Ser2→Arg and Tyr3→NT. In the presence of saturating concentrations of Y3NT or S2R/Y3NT, the fluorescence of thrombin is strongly quenched and approaches a similar value, under either *fast* (0.2-M NaCl) and *slow* (0.2-M ChCl) conditions (Figure 12A, B). We have demonstrated that quenching of fluorescence is mainly caused by FRET, occurring between (some of) the Trp-residues of thrombin (i.e., the donors) and the single 3-nitrotyrosine of the inhibitors (i.e., the acceptor).

FRET is a nonradiative decay process occurring between a donor and an acceptor, which interact *via* electromagnetic dipoles transferring the excitation energy of the donor to the acceptor. For a one-donor–one-acceptor system, the efficiency of energy transfer depends on the extent of spectral overlap of the emission spectrum of the donor with the absorption spectrum of the acceptor, on the donor quantum yield, on the inverse sixth power of the distance separating the donor and acceptor, and on their orientation [88]. In the case of hirudin–thrombin interaction, there is an extensive overlap of the emission spectrum of the enzyme (i.e., the donor) with the absorption spectrum of the inhibitor (i.e., the acceptor) (see Figure 12C). In addition, Trp-to-NT energy transfer is also favored by the relatively short distances separating Trp-residues and NT in the enzyme–inhibitor complex [31, 91]. To estimate the possible contribution of spectroscopic effects other than FRET (e.g, unspecific binding, dynamic or static quenching, inner filter effect), the fluorescence of thrombin was measured in the presence of increasing concentrations of free NT. The data reported in Figure 12D indicate that NT slightly (~14%) reduces the fluorescence of the enzyme, in keeping with the notion that nitrocompounds (e.g., nitromethane and nitrobenzene) quench the emission of polycyclic aromatic hydrocarbons mainly by a dynamic mechanism [110]. Moreover, we found that for concentrations of NT-containing analogs lower than 20 μM, inner filter effect can be neglected. These considerations allow us to conclude that quenching of thrombin fluorescence by Y3NT (or S2R/Y3NT) is mainly caused by Trp-to-NT energy transfer.

The quenching data reported above were used to obtain quantitative estimates (i.e., $K_d$ values) of the binding of hirudin analogs to thrombin allosteric forms (see Figure 13 and Table 3). For both analogs, the excellent fit of the experimental data to the curve describing one-site binding mechanism is a stringent, albeit indirect,

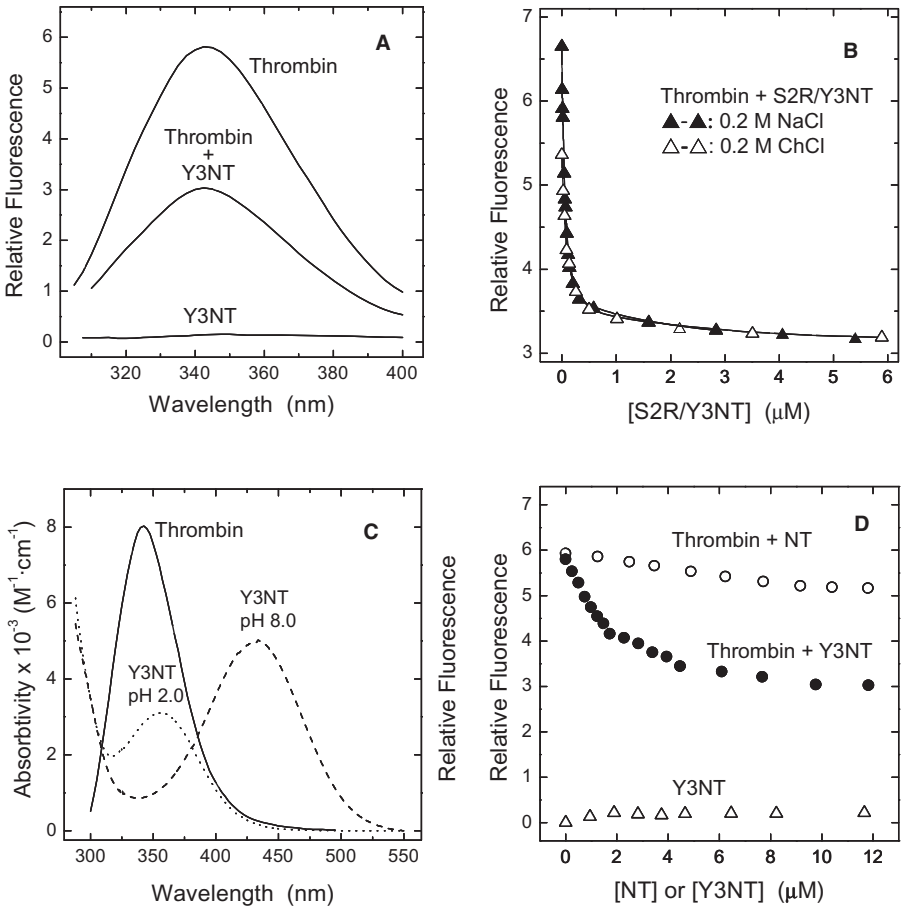


**Figure 12.** Binding of Y3NT and S2R/Y3NT to thrombin, monitored by Trp-to-NT fluorescence energy transfer. (**A**) Fluorescence spectra of thrombin alone (50 nM) and in the presence of Y3NT (10 μM). For comparison, the spectrum of the free inhibitor Y3NT (10 μM) is also reported. (**B**) Change in the fluorescence of thrombin as a function of S2R/Y3NT concentration, under fast (▲–▲, 0.2-M NaCl) and slow (△–△, 0.2-M ChCl) conditions. (**C**) Superimposition of the fluorescence spectrum of thrombin (continuous line) with the absorption spectra of Y3NT at pH 2 and 8.0 (dashed/dotted lines). (**D**) Change in thrombin fluorescence as a function of Y3NT concentration (●–●). As a control, the fluorescence intensity of thrombin in the presence of free NT (○–○) is reported. The signal of Y3NT alone (△–△) is also included. All measurements were carried out at 25°C by exciting the protein samples at 295 nm in 5-mM Tris-HCl buffer, pH 8.0, containing 0.1% (w/v) PEG 8000 and 0.2-M salt, as indicated, and recording the fluorescence signal at 342 nm.

proof of 1:1 binding stoichiometry. The replacement of Tyr3 with NT resulted in a drop in the affinity of Y3NT for thrombin, which was restored in the doubly substituted analog S2R/Y3NT by replacing Ser2 with Arg. The structural model of Y3NT bound to thrombin (see Figure 14), based on the crystallographic structure of the hirudin–thrombin complex [31], reveals that the $-NO_2$ group of NT might be easily accommodated into the S2 specificity site of the enzyme without

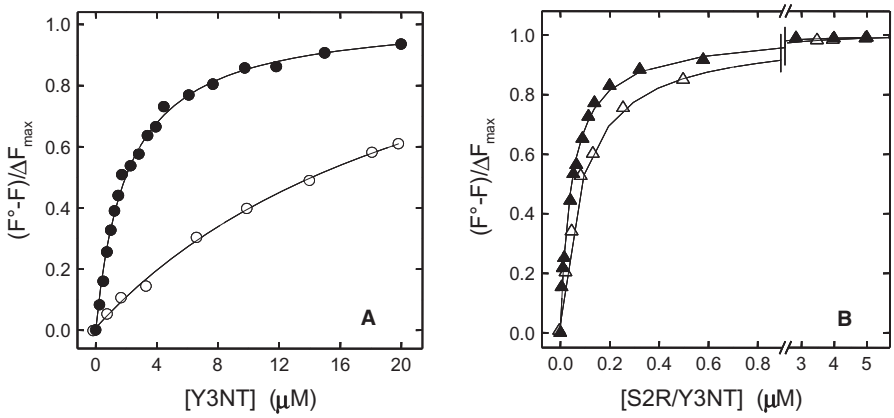


**Figure 13.** Determination of the dissociation constant ($K_d$) of the complexes formed by the synthetic analogs Y3NT (**A**) (●, ○) and S2R/Y3NT (**B**) (▲, △) with thrombin, under fast (filled symbols) and slow (empty symbols) conditions. Continuous lines represents the best fit of the data points to the equation describing tight binding [91], which allowed us to obtain the $K_d$ values reported in Table 3.

**TABLE 3. Thrombin Binding Data of the Synthetic Hirudin Analogs, as Obtained by Fluorescence Energy Transfer[1] and Enzyme Inhibition Assays[2]**

| 1–47 Analogs | Fast Form | | Slow Form | | |
|---|---|---|---|---|---|
| | $K_d$ (nM) | $\Delta\Delta G_b$ (kcal/mol) | $K_d$ (nM) | $\Delta\Delta G_b$ (kcal/mol) | $\Delta G_c$ (kcal/mol) |
| WT[2] (Tyr3) | 42 ± 0.5 | — | 1,460 ± 20 | — | –2.10 |
| Y3NT[1] | 1,325 ± 80 | 2.06 | 17,000 ± 500 | 1.45 | –1.51 |
| S2R[2] | 1.7 ± 0.02 | –1.90 | 12 ± 2 | –2.84 | –1.16 |
| S2R/Y3NT[1] | 45 ± 2.2 | 0.06 | 91 ± 4.2 | –1.64 | –0.41 |

[1] The $K_d$ values of Y3NT and S2R/Y3NT were obtained by fitting FRET data to the equation describing tight binding [91].

[2] The inhibitory potency of the wild-type (WT) and S2R analog was determined at 25°C by measuring at 405 nm the release of *p*-nitroanilide, *p*NA, from the synthetic substrate D-Phe-Pro-Arg-*p*NA. Additivity of mutational effects was calculated by the equation $\Delta G_I = \Delta\Delta G_b(\text{S2R/Y3NT}) - [\Delta\Delta G_b(\text{Y3NT}) + \Delta\Delta G_b(\text{S2R})]$, where $\Delta G_I$ is the free energy term that accounts for the energetic interaction between the mutated sites [9].

requiring steric distortion. Likely, the lower affinity of Y3NT reflects the lower hydrophobicity of NT at pH 8.0, where it exists by ~95% in the ionized form. With respect to this, the logP value of 2-nitrophenol, taken as a suitable model of the NT side chain, is –1.47 at pH 8.0 [101], whereas that of phenol, taken as a model of Tyr, is +1.50 [102]. Besides hydrophobicity, the presence of the nitro-group introduces a net (i.e., at pH 8.0) negative charge at position 3 of hirudin, which can oppose binding through unfavorable electrostatic interaction with the strong negative potential of the thrombin active site [111], in agreement with our previous structure-activity relationship studies [43]. The reliability of these data was verified by comparing the $K_d$ values of Y3NT and S2R/Y3NT, obtained by FRET

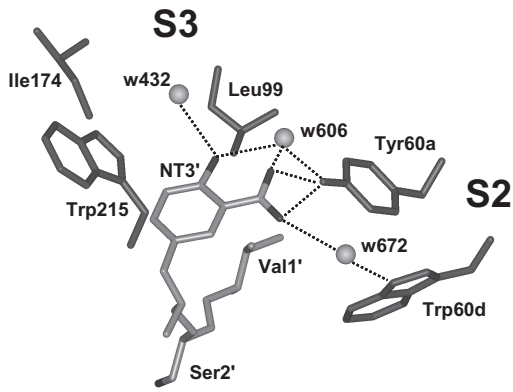


**Figure 14.** Schematic representation of the interaction of the N-terminal tripeptide of Y3NT with thrombin. The structure of Y3NT was modeled on the structure of the hirudin–thrombin complex [4htc.pdb; 31]. Relevant NT-thrombin distances, in the 2.5–3.5-Å range, are indicated by dashed lines. Of note, w606, which in the wild-type hirudin–thrombin structure connects Tyr3′ of the inhibitor to Tyr60a of the enzyme, is well suited as a hydrogen bond donor to stabilize the NT ring system [91].

measurements, with those determined by classic enzyme inhibition experiments, in which the rate of thrombin-mediated substrate hydrolysis was measured as a function of inhibitor concentration [91]. Strikingly, the $K_d$ values for the binding of Y3NT and S2R/Y3NT to thrombin *fast* form were determined as 1.4 ± 0.1 μM and 41 ± 2 nM, respectively, in agreement (8–10%) with those obtained by FRET (Table 3).

The absorption spectra of S2R/Y3NT, recorded in the absence and presence of thrombin, are reported in Figure 15. Under *fast* conditions, the 430-nm band of S2R/Y3NT, assigned to the contribution of the ionized form of NT at pH 8.0, is reduced by 54%, whereas an additional band of similar intensity appears at about 362 nm, characteristic of NT in the neutral form. Of note, the binding of Y3NT to thrombin *fast* form yields very similar results [91]. These observations can be explained on the basis of the modeled structure of Y3NT bound to thrombin and assuming that NT interacts with the enzyme in the neutral form.

As shown in Figure 14, three structural water molecules at the enzyme–inhibitor interface (i.e., w432, w606, and w672), characterized by low thermal factors and high occupancy values, can variably interact with NT. In particular, w606, which in the structure of the wild-type hirudin–thrombin complex connects Tyr3′ of the inhibitor to Tyr60a of the enzyme [31], is suitably positioned as a hydrogen bond donor to stabilize the six-membered ring system of NT (see Figure 11A and Figure 14). As a result, the contribution of the protonated NT in the bound form appears as a distinct band at about 362 nm in the absorption spectrum of the thrombin–S2R/Y3NT complex (Figure 15). The residual intensity of the 430-nm band is contributed by the ionized form of NT in the free inhibitor that exists in equilibrium with the thrombin-bound form. Hence, we conclude that the phenate moiety of NT in the free state becomes protonated to phenol upon binding to thrombin and that a water molecule at the hirudin–thrombin interface, likely w606, functions as a hydrogen

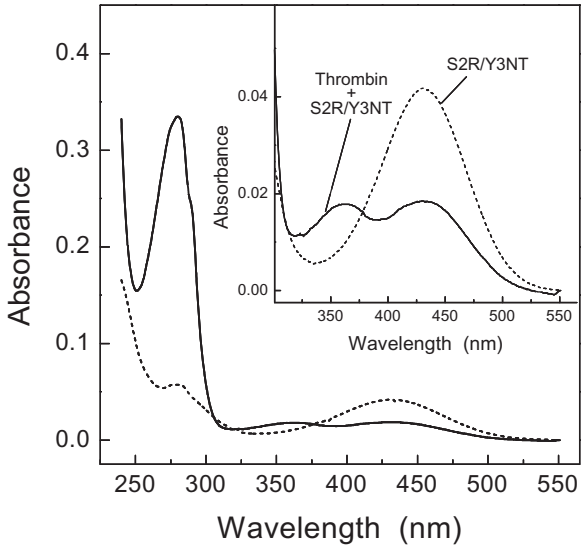


**Figure 15.** Binding of the synthetic analog S2R/Y3NT to thrombin, monitored by UV/Vis absorption spectroscopy. Spectra of the inhibitor (8.6 μM) were taken at 25° in 5-mM Tris-HCl buffer, pH 8.0, containing 0.1% PEG 8000 and 0.2 M NaCl, in the absence (--) and presence (--) of thrombin (4.1 μM). For clarity, the *Inset* shows the spectra in the wavelength range 300–550 nm.

donor. Notably, w432 and w606 are conserved in the structure of thrombin bound to hirugen (i.e., the 53–64 peptide of hirudin) [1hah.pdb; 112], where the specificity sites of the enzyme are unoccupied, thus suggesting that these water molecules represent constant spots in the solvation shell of thrombin and, perhaps, key elements for molecular recognition, in keeping with the key role that protein–water interactions play in ligand binding [113].

Our results demonstrate that NT is a suitable spectroscopic probe for investigating ligand–protein interactions and suggest that its incorporation into proteins may have vast applications for identifying ligand–protein interaction.

## 8 CONCLUDING REMARKS

It is widely accepted that even the knowledge of the structure of the ligand–receptor complex provides only partial information for predicting ligand binding energetics [114]. In this view, we have demonstrated here that it is possible to transfer the quantitative structure-activity relationship (QSAR) approach, traditionally applied to low-molecular-weight bioactive compounds, to the study of recognition in macromolecular systems, such as the hirudin–thrombin complex. Recent advances in chemical and genetic methodologies will, hopefully, allow the researchers to extend this approach to other protein systems of relevant pharmacological application. Moreover, we have shown that incorporation of noncoded amino acids, possessing peculiar spectroscopic properties, can have great potentialities in biotechnology and pharmacological screening.

## ACKNOWLEDGMENT

This work was supported by grants PRIN-2003 and PRIN-2005 from the Italian Ministry of University and Scientific Research. I would like to thank Prof. Angelo Fontana for stimulating discussions. The support of Dr. Olmetta Iadicicco is also gratefully acknowledged.

## REFERENCES

1. Freedman R B (1971). Applications of the chemical reactions of proteins in studies of their structure and function. *Q. Rev. Chem. Soc.* 25:431–462.
2. Fersht A, Winter G (1992). Protein engineering. *Trends Biochem. Sci.* 17:292–294.
3. Smock R G, Gierasch L M (2005). Finding the fittest fold: using the evolutionary record to design new proteins. *Cell.* 122:832–834.
4. Cornish V W, Mendel, D, Schultz P G (1995). Probing protein structure and function with an expanded genetic code. *Angew. Chem.* 34:621–633.
5. Giannis A, Rübsam F (1997). Peptidomimetics in drug design. *Adv. Drug Res.* 29:1–78.
6. Hogan J G Jr (1997). Combinatorial chemistry in drug discovery. *Nat. Biotech.* 15:328–330.
7. Kent S B (1988). Chemical synthesis of peptides and proteins. *Annu. Rev. Biochem.* 57:957–989.
8. Dawson P, Kent S B (2000). Synthesis of native proteins by chemical ligation. *Annu. Rev. Biochem.* 69:923–960.
9. Nilsson B L, Soellner M B, Raines R T (2005). Chemical synthesis of proteins. *Annu. Rev. Biophys. Biomol. Struct.* 34:91–118.
10. Wallace C J (1993). Understanding cytochrome c function: Engineering protein structure by semisynthesis. *FASEB J.* 7:505–515.
11. De Filippis V, De Antoni F, Frigo M, et al. (1998). Protein stabilization by Ala→Aib replacements. *Biochem.* 37:1686–1696.
12. Wong C-Y, Eftink M R (1997). Biosynthetic incorporation of tryptophan analogues into staphylococcal nuclease: Effect of 5-hydroxytryptophan and 7-azatryptophan on structure and stability. *Protein Sci.* 6:689–697.
13. Budisa N, Minks C, Mediano F, et al. (1998). Residue-specific bioincorporation of non-natural, biologically active amino acid into proteins as possible drug carriers: Structure and stability of the *per*-thiaproline mutant of annexin V. *Proc. Natl. Acad. Sci.* 95:455–459.
14. Dougherty D A (2000). Unnatural amino acids as probes of protein structure and function. *Curr. Opin. Chem. Biol.* 4:645–652.
15. Cropp A T, Schultz P G (2004). An expanding genetic code. *Trends Genet.* 20:625–630.
16. England P M (2004). Unnatural amino acid mutagenesis: A precise tool for probing protein structure and function. *Biochem.* 43:11623–11629.
17. Hendrickson T L, de Crécy-Lagard V, Schimmel P (2004). Incorporation of nonnatural amino acids into proteins. *Annu. Rev. Biochem.* 73:147–176.
18. Albericio F (2004). Developments in peptides and amide synthesis. *Curr. Opin. Chem. Biol.* 8:211–221.

19. Minami T, Sugiyama A, Wu S O, et al. (2004). Thrombin and phenotypic modulation of the endothelium. *Arterioscler. Thromb. Vasc. Biol.* 24:41–53.
20. Griffin J H (1995). The thrombin paradox. *Nature*. 378:337–338.
21. Esmon C T (2003). The Protein C pathway. *Chest*. 124:26S–32S.
22. Dang Q D, Vindigni A, Di Cera E (1995). An allosteric switch controls the procoagulant and anticoagulant activities of thrombin. *Proc. Natl. Acad. Sci. USA* 92:5977–5981.
23. Di Cera E, Dang Q D, Ayala Y M (1997). Molecular mechanisms of thrombin function. *Cell Mol. Life Sci*. 53:701–730.
24. Di Cera E (2003). Thrombin interactions. *Chest*. 124:11S–17S.
25. Berg D T, Wiley M R, Grinnell B W (1996). Enhanced Protein C activation and ihibition of fibrinogen cleavage by a thrombin modulator. *Science*. 273:1389–1391.
26. Braun P J, Dennis S, Hofsteenge J, et al. (1988). Use of site-directed mutagenesis to investigate the basis for the specificity of hirudin. *Biochem*. 27:6517–6522.
27. Vindigni A, De Filippis V, Zanotti G, et al. (1994). Probing the structure of hirudin from *Hirudinaria manillensis* by limited proteolysis: Isolation, characterization and thrombin-inhibitory properties of N-terminal fragments. *Eur. J. Biochem*. 226:323–333.
28. Folkers P J M, Clore G M, Driscoll P C, et al. (1989). Solution structure of recombinant hirudin and the Lys-47→Glu mutant: A nuclear magnetic resonance and hybrid distance geometry-dynamical simulated annealing study. *Biochem*. 28:2601–2617.
29. Szyperski T, Güntert P, Stone S R, et al. (1992). Nuclear magnetic resonance solution structure of hirudin (1–51) and comparison with corresponding three-dimensional structures determined using the complete 65-residue hirudin polypeptide chain. *J. Mol. Biol*. 228:1193–1205.
30. Nicastro G, Baumer L, Bolis G, et al. (1997). NMR solution structure of a novel hirudin variant HM2, N-terminal 1–47 and N64 → V+G mutant. *Biopolym*. 41:731–749.
31. Rydel T J, Tulinsky A, Bode W, et al. (1991). Refined structure of the hirudin-thrombin complex. *J. Mol. Biol*. 221:583–601.
32. Scacheri E, Nitti G, Valsasina B, et al. (1993). Novel hirudin variants from the leech *Hirudinaria manillensis*. Amino acid sequence, cDNA cloning and genomic organization. *Eur. J. Biochem*. 214:295–304.
33. Betz A, Hofsteenge J, Stone S R (1992). Interaction of the N-terminal region of hirudin with the active-site cleft of thrombin. *Biochem*. 31:4557–4562.
34. De Filippis V, Quarzago D, Vindigni A, et al. (1998). Synthesis and characterization of more potent analogues of hirudin fragment 1–47 containing non-natural amino acids. *Biochem*. 37:13507–13515.
35. Szyperski T, Güntert P, Stone S R, et al. (1992). Impact of protein-protein contacts on the conformation of thrombin-bound hirudin studied by comparison with the nuclear magnetic resonance solution structure of hirudin (1–51). *J. Mol. Biol*. 228:1206–1211.
36. Fauchère J-L, Charton M, Kier L B, et al. (1988). Amino acid side-chain parameters for correlation studies in biology and pharmacology. *Int. J. Pept. Prot. Res*. 32:269–278.
37. Paterson Y, Leach S J (1978). The effect of side-chain branching on the theoretically predicted conformational space available to amino acid residues. *Macromol*. 11:409–415.

38. Steiner V, Knecht R, Börsen K O, et al. (1992). Primary structure and function of novel *O*-glycosylated hirudins from the leech *Hirudinaria manillensis*. *Biochem*. 31:2294–2298.
39. De Filippis V, Vindigni A, Altichieri L, et al. (1995). Core domain of hirudin from leech *Hirudinaria manillensis*: Chemical synthesis, purification and characterization of a Trp3 analog of fragment 1–47. *Biochem*. 34:9552–9564.
40. Ayala Y, Di Cera E (1994). Molecular Recognition by thrombin. Role of the slow → fast transition, site-specific ion binding energetics and thermodynamic mapping of structural components. *J. Mol. Biol*. 235:733–746.
41. Wallace A, Dennis S, Hofsteenge J, et al. (1989). Contribution of the N-terminal region of hirudin to its interaction with thrombin. *Biochem*. 28:10079–10084.
42. Wang J, Szewczuk Z, Yue S-Y, et al. (1995). Calculation of relative binding energies and configurational entropies: A structural and thermodynamic analysis of the nature of non-polar binding of thrombin inhibitors based on hirudin. *J. Mol. Biol*. 253:473–492.
43. De Filippis V, Colombo G, Russo I, et al. (2002). Probing hirudin-thrombin interaction by incorporation of noncoded amino acids and molecular dynamics simulation. *Biochem*. 43:1537–1550.
44. Sangster J (1989). Octanol-water partition coefficients of simple organic compounds. *J. Phys. Chem*. 18:1111–1229.
45. Hansch C, Leo A, Unger S H, et al. (1973). "Aromatic" substituent constants for structure-activity correlations. *J. Med. Chem*. 16:1207.
46. Lien E J, Guo Z-R, Li R-L, et al. (1982). Use of dipole moment as a parameter in drug-receptor interaction and quantitative structure-activity relationship studies. *J. Pharm. Sci*. 71:641–655.
47. Burley S K, Petsko G A (1988). Weakly polar interactions in proteins. *Adv. Protein Chem*. 39:125–189.
48. Serrano L, Bycroft M, Fersht A R (1991). Aromatic-aromatic interactions and protein stability. *J. Mol. Biol*. 218:465–475.
49. Gilson M K, Rashin A, Fine R, et al. (1985). On the calculation of electrostatic interactions in proteins. *J. Mol. Biol*. 184:503–516.
50. Dao-Pin S, Anderson D E, Baase W A, et al. (1991). Structural and thermodynamic consequences of burying a charged residue within the hydrophobic core of T4 lysozyme. *Biochem*. 30:115121–11529.
51. Berliner L J, Shen Y Y L (1977). Physical evidence for an apolar binding site near the catalytic centre of human α-thrombin. *Biochem*. 16:4622–4626.
52. Eisenberg D, McLachlan A D (1986). Solvation energy in protein folding and binding. *Nature*. 319:199–203.
53. Horton N, Lewis M (1992). Calculation of the free energy of association for protein complexes. *Protein Sci*. 1:169–181.
54. Legon A C (1999). Prereactive complexes of dihalogens XY with Lewis bases B in the gas phase: A systematic case for the halogen analogue B–XY of the hydrogen bond B–HX. *Angew. Chem., Int. Ed*. 38:2686–2698.
55. Kim K S, Tarakeshwar P, Lee J Y (2000). Molecular clusters of pi-systems: Theoretical studies of structures, spectra, and origin of interaction energies. *Chem. Rev*. 100: 4145–4185.
56. Mecozzi S, West A P, Dougherty D A (1996). Cation-pi interactions in aromatics of biological and medicinal interest: Electrostatic potential surfaces as a useful qualitative guide. *Proc. Natl. Acad. Sci. USA*. 93:10566–10571.

57. Steiner T, Koellner G (2001). Hydrogen bonds with pi-acceptors in proteins: Frequencies and role in stabilizing local 3D structures. *J. Mol. Biol.* 305:535–557.
58. Markwardt F (1994). The development of hirudin as an antithrombotic drug. *Thromb. Res.* 74:1–23.
59. Pineo G F, Hull R D (1995). Hirudin and hirudin analogues as new anticoagulant agents. *Curr. Opin. Hematol.* 2:380–385.
60. Fenton J W Jr, Ofosu F A, Brezniak D V, et al. (1998). Thrombin and antithrombotics. *Semin. Thromb. Hemostasis.* 24:87–91.
61. Chang J-Y (1990). Production, properties, and thrombin inhibitory mechanism of hirudin amino-terminal core fragments. *J. Biol. Chem.* 265:22159–22166.
62. Callas D, Fareed J (1995). Comparative pharmacology of site directed antithrombin agents. Implication in drug development. *Thromb. Haemostasis.* 74:473–781.
63. Hursting M J, Alford K L, Becker J P, et al. (1997). Novastan®: A small-molecule, direct thrombin inhibitor. *Sem. Thromb. Hemostasis.* 23:503–516.
64. Verstraete M (1997). Direct thrombin inhibitors: Appraisal of the antithrombotic/hemorrhagic balance. *Thromb. Haemostasis.* 78:357–363.
65. De Filippis V, Russo I, Vindigni A, et al. (1999). Incorporation of noncoded amino acids into the N-terminal domain 1–47 of hirudin yields a highly potent and selective thrombin inhibitor. *Protein Sci.* 8:2213–2217.
66. Tapparelli C, Metternich R, Ehrhardt C, et al. (1993). Synthetic low-molecular weight inhibitors: Molecular design and pharmacological profile. *Trends Pharmacol. Sci.* 14:366–376.
67. Das J, Kimball D S (1995). Thrombin active site inhibitors. *Bioorg. Med. Chem.* 3:999–1007.
68. Sanderson P E J, Naylor-Olsen A M (1998). Thrombin inhibitor design. *Curr. Med. Chem.* 5:289–304.
69. Wells J A (1990). Additivity of mutational effects in proteins. *Biochem.* 29: 8509–8517.
70. Di Cera E, Guinto E R, Vindigni A, et al. (1995). The $Na^+$ binding site of thrombin. *J. Biol. Chem.* 270:22089–22092.
71. Pineda A O, Savvides S N, Waksman G, et al. (2002). Crystal structure of the anticoagulant *slow* form. *J. Biol. Chem.* 277:40177–40180.
72. Pineda A O, Carrell C J, Bush L A, et al. (2004). Molecular dissection of $Na^+$ binding to thrombin. *J. Biol. Chem.* 279:31842–31853.
73. Huntington J A, Esmon C (2003). The molecular basis of thrombin allostery revealed by a 1.8-Å structure of the "slow" form. *Structure.* 11:469–479.
74. Carter W J, Myles T, Gibbs C S, et al. (2004). Crystal structure of anticoagulant thrombin variant E217K provides insights into thrombin allostery. *J. Biol. Chem.* 279:26387–26394.
75. De Filippis V, De Dea E, Lucatello F, et al. (2005). Effect of $Na^+$ binding on the conformation, stability, and molecular recognition properties of thrombin. *Biochem. J.* 390:485–492.
76. Wyman J (1968). Regulation in macromolecules as illustrated by haemoglobin. *Q. Rev. Biophys.* 1:35–80.
77. Di Cera E (1998). Site-specific analysis of mutational effects in proteins. *Adv. Protein Chem.* 51:59–119.
78. Stubbs M T, Oschkinat H, Mayr I, et al. (1992). The interaction of thrombin with fibrinogen: A structural basis for its specificity. *Eur. J. Biochem.* 206:187–195.

79. Mathews I I, Padmanabhan K P, Ganesh V, et al. (1994). Crystallographic structures of thrombin complexed with thrombin receptor peptides: Existence of expected and novel binding modes. *Biochem*. 33:3266–3279.
80. Sadasivan C, Yee V C (2000). Interaction of the factor XIII activation peptide with α-thrombin: Crystal structure of its enzyme-substrate analog complex. *J. Biol. Chem*. 275:36942–36948.
81. Cornish V W, Benson D R, Altenbach C A, et al. (1994). Site-specific incorporation of biophysical probes into proteins. *Proc. Natl. Acad. Sci*. 91:2910–2924.
82. Wong C-Y, Eftink M R (1998). Incorporation of tryptophan analogues into staphylococcal nuclease: Stability toward thermal and guanidine-HCl induced unfolding. *Biochem*. 37:8947–8953.
83. Cohen B E, McAnaney T B, Park E S, et al. (2002). Probing protein electrostatics with a synthetic fluorescent amino acid. *Science*. 296:1700–1703.
84. Hovius R, Vallotton P, Wohland T, et al. (2000). Fluorescence techniques: Shedding light on ligand-receptor interactions. *Trends Pharmacol. Sci*. 21:266–273.
85. Eftink M R (1997). Fluorescence methods for studying equilibrium macromolecule-ligand interactions. *Methods Enzymol*. 278:221–257.
86. Jameson D M, Croney J C, Moens P D (2003). Fluorescence: Basic concepts, practical aspects, and some anecdotes. *Methods Enzymol*. 360:1–43.
87. Wu P, Brand L (1994). Resonance energy transfer: Methods and applications. *Anal. Biochem*. 18:1–13.
88. Selvin P R (1995). Fluorescence resonance energy transfer. *Methods Enzymol*. 246:300–334.
89. Yan Y, Marriott G (2003). Analysis of protein interactions using fluorescence technologies. *Curr. Opin. Chem. Biol*. 7:635–640.
90. De Filippis V, De Boni S, De Dea E, et al. (2004). Incorporation of the fluorescent amino acid 7-azatryptophan into the core domain 1–47 of hirudin as a probe of hirudin folding and thrombin recognition. *Protein Sci*. 13:1489–1502.
91. De Filippis V, Frasson R, Fontana A (2006). 3-Nitrotyrosine as a spectroscopic probe for investigating protein-protein interactions. *Protein Sci*. 15:976–986.
92. Twine S M, Szabo A G (2003). Fluorescent amino acid analogs. *Methods Enzymol*. 360:104–127.
93. Ross A, Szabo A G, Hogue C W V (1997). Enhancement of protein spectra with tryptophan analogs: Fluorescence spectroscopy of protein-protein and protein-nucleic acid interactions. *Methods Enzymol*. 278:151–190.
94. Negrerie M, Gai F, Bellefeuille S M, et al. (1991). Photophysics of a novel optical probe: 7-Azaindole. *J. Phys. Chem*. 95:8663–8670.
95. Chapman C F, Maroncelli M (1992). Excited-state tautomerization of 7-azaindole in water. *J. Phys. Chem*. 96:8430–8441.
96. Chou P-T, Martinez M L, Cooper W C, et al. (1992) Monohydrate catalysis of excited-state double-proton transfer in 7-azaindole. *J. Phys. Chem*. 96:5203–5205.
97. Smirnov A V, English D S, Rich R L, et al. (1997). Photophysics and biological applications of 7-azaindole and its analogs. *J. Phys. Chem. B*. 101:2758–2769.
98. Mente S, Maroncelli M (1998). Solvation and the excited-state tautomerization of 7-azaindole and 1-azacarbazole: Computer simulations in water and alcohol solvents. *J. Phys. Chem*. 102:3860–3876.
99. Halliwell B (1997). What nitrates tyrosine? Is nitrotyrosine specific as a biomarker of peroxynitrite formation *in vivo*? *FEBS Lett*. 411:157–160.

100. Riordan J F, Sokolovsky M, Vallee B L (1967). Environmentally sensitive tyrosyl residues. Nitration with tetranitromethane. *Biochem*. 6:358–361.
101. Csizmadia F, Tsantili-Kkoulidou A, Panderi I, et al. (1997). Prediction of distribution coefficient from structure. 1. Estimation method. *J. Pharm. Sci*. 7:865–871.
102. Abraham M H, Du C M, Platts J A (2000). Lipophilicity of the nitrophenols. *J. Org. Chem*. 65:7114–7118.
103. Mostad A, Natarajan S (1990). Crystal and molecular structure of 3-nitro-4-hydroxyphenylalanine nitrate. *Z. Kristall*. 193:127–136.
104. Steiner R F, Albaugh S, Kilhoffer M-C (1991). Distribution separations between groups in an engineered calmodulin. *J. Fluoresc*. 1:15–22.
105. Rischel C, Poulsen F M (1995). Modification of a specific tyrosine enables tracing of the end-to-end distance during apomyoglobin folding. *FEBS Lett*. 374:105–109.
106. Rischel C, Thyberg P, Rigler R, et al. (1996). Time-resolved fluorescence studies of the molten globule state of apomyoglobin. *J. Mol. Biol*. 257:877–885.
107. Tcherkasskaya O, Ptitsyn O B (1999). Direct energy transfer to study the 3D structure of non-native proteins: AGH complex in the molten globule state of apomyoglobin. *Protein Eng*. 12:485–490.
108. Juminaga D, Albaugh S A, Steiner R F (1994). The interaction of calmodulin with regulatory peptides of phosphorylase kinase. *J. Biol. Chem*. 269:1660–1667.
109. Mezo A R, Cheng R P, Imperiali B (2001). Oligomerization of uniquely folded miniprotein motifs: Development of a homotrimeric ββα peptide. *J. Am. Chem. Soc*. 123:3885–3891.
110. Dreeskamp H, Koch E, Zander M (1975). On the fluorescence quenching of polycyclic aromatic hydrocarbons by nitromethane. *Z. Naturforsch*. 30a:1311–1314.
111. Karshikov A, Bode W, Tulinsky A, et al. (1992). Electrostatic interactions in the association of proteins: An analysis of the thrombin-hirudin complex. *Protein Sci*. 1:727–735.
112. Vijayalakshmi J, Padmanabhan K P, Mann K G, et al. (1994). The isomorphous structures of prethrombin-2, hirugen-, and PPACK-thrombin: Changes accompanying activation and exosite binding to thrombin. *Protein Sci*. 3:2254–2271.
113. Mattos C (2002). Protein-water interactions in a dynamic world. *Trends Biochem. Sci*. 27:203–208.
114. Engh R A, Brandstetter H, Sucher G, et al. (1996). Enzyme flexibility, solvent and weak interactions characterize thrombin-ligand interactions: Implications for drug design. *Structure*. 4:1353–1362.

# PRODUCTION AND PURIFICATION OF ADENOVIRUS VECTORS FOR GENE THERAPY

D. M. F. Prazeres and J. A. L. Santos
*Instituto Superior Técnico, Lisbon, Portugal*

**Chapter Contents**

1 Gene Therapy
  1.1 Milestones
  1.2 Vectors
2 Adenovirus in Gene Therapy
  2.1 Introduction
  2.2 Biology and Properties
  2.3 Recombinant Adenovirus Vectors
  2.4 Targeting Strategies
  2.5 Applications
  2.6 Safety Aspects
3 Adenovirus Manufacturing
  3.1 GMPs and Validation
  3.2 Product Specifications and Quality Control
  3.3 Environmental and Safety Issues
  3.4 Process Considerations
  3.5 Production
  3.6 Downstream Processing
4 Analysis and Evaluation of an Adenovirus Production Process
  4.1 Process Description
  4.2 Inventory Analysis
  4.3 Cost Analysis and Economic Assessment
5 Concluding Remarks and Outlook
  References

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

# 1 GENE THERAPY

## 1.1 Milestones

Gene therapy is a conceptually simple and attractive process, which consists of the introduction of one or more functional genes in a human/non-human receptor (*in vivo*, *in situ*, or *ex vivo*). It constitutes a promising alternative for the treatment, diagnosis, or cure of genetic defects such as cystic fibrosis [1] or acquired diseases like cancer [2] and AIDS [3]. DNA vaccines can also be developed on the basis of genes [4] to provide immunity against infectious agents (e.g., malaria [5]) or treat noninfectious diseases such as tumors and allergies [6].

Although gene therapy is a recent endeavor of the human mind, it has nevertheless come a long way since the early, nonauthorized administration of Shope papilloma virus to argininemia-suffering patients by Stanfield Rogers et al. [7, 8] (Table 1). Despite the flawed design and consequent failure of this clinical trial, Rogers was one of the first scientists to anticipate the therapeutic potential of viruses as carriers of genetic information [8]. The first federally approved gene therapy clinical trials took place in 1990 when an adenosine deaminase (ADA)-deficient patient was given her own T cells engineered with a retroviral vector carrying a normal ADA gene [9]. This experiment paved the way for further clinical trials. In 1993 an adenovirus vector (AdV) was first used in a clinical trial designed to evaluate the potential of direct transfer of cystic fibrosis transmembrane conductance regulator (CFTR) cDNA in the treatment of cystic fibrosis (CF) [10].

**TABLE 1. Gene Therapy Milestones**

| Year | Vector | Target Disease | Comments | Reference |
|---|---|---|---|---|
| 1970 | Shope virus | Argininemia | First human trial, unauthorized | [7, 8] |
| 1990 | Retrovirus | ADA-SCID | First federally approved human trial | [9] |
| 1993 | Adenovirus | Cystic fibrosis | First clinical trial with an adenovirus vector | [10] |
| 1999 | Adenovirus | OTC deficiency | Death of patient due to an adverse reaction to the adenovirus vector | [11, 12] |
| 2000 | Retrovirus | SCID-X1 syndrome | Apparent clinical cure of two recipients | [13] |
| 2003 | Retrovirus | SCID-X1 syndrome | Development of leukemia-like syndrome in recipients due to retrovirus integration | [14] |
| 2004 | Adenovirus | Head and neck carcinoma | The first human GT product, Gendicine, received approval from the Chinese FDA | [15] |

*Abbreviations*: ADA = adenosine deaminase; OTC = ornythine transcarbamylase; SCID = severe combined immunodeficiency.

The first setback faced by gene therapy came in 1999 when a patient suffering from ornithine transcarbamylase (OTC) deficiency died after administration of an adenovirus vector encoding OTC [11, 12]. The year 2000 saw gene therapy's first major success: A gene therapy protocol could correct the phenotype of an X-linked severe combined immunodeficiency (SCID-X1) syndrome in two patients who had been re-infused with autologous CD34 bone marrow cells transduced *ex vivo* with a retrovirus vector encoding the γC receptor gene [13]. These successes were later shadowed by the development of leukemia-like syndrome in recipients of the treatment as a consequence of retrovirus integration in proximity to the LMO2 proto-oncogene promoter [14]. In October 16, 2003, a recombinant adenovirus vector expressing the tumor-suppressor gene p53 was approved by the State Food and Drug Administration (SFDA) of China for the treatment of head and neck squamous cell carcinoma [15–17]. Developed and manufactured by Shenzen SiBiono GeneTech (China) and trademarked under the name Gendicine, it became the first human gene therapy product to reach the market in April 2004 [15]. As of September 2005, 2600 patients had been treated with Gendicine, with projections estimating 50,000 patients to receive the product by 2006 [18].

### 1.2 Vectors

The transport of the therapeutic transgenes toward the nuclei of the target cells can be carried out both by viral [19] and nonviral vectors such as plasmid DNA [20]. Plasmid DNA molecules are extra-chromosomal carriers of genetic information that have the ability to replicate autonomously. These vectors constitute an attractive gene transfer system because they are safer and easier to produce when compared with viral vectors [20–22]. However, plasmid DNA vectors are less effective in transfecting cells when compared with viral vectors, which have a natural ability to deliver and express their genes in a wide variety of cell types and tissues [11, 23].

The major drawback of viral vectors is undoubtedly related to safety aspects. When confronted with viral vector particles, the human immune system, which has evolved to tackle wild-type infections, will likewise generate an immune response [19]. The potency and severity of this response can ultimately lead to death, as occurred in the 1999 OTC gene therapy clinical trial (Table 1) [12]. Thus, a safe and effective use of viral vectors in gene therapy requires the modification of natural viruses to impair replication and expression of viral proteins and thus minimize toxicity and immunogenicity [19]. The infection pathway of these recombinant viral vectors should of course remain unaltered. Further goals of virus modification are to provide room in the viral genome for transgenes, to modulate transgene expression, and to improve the selectivity of infection.

Most recombinant viral vectors used in gene therapy clinical trials belong to one of the following categories (Figure 1): (1) adenovirus, (2) retrovirus/lentivirus, (3) pox/vaccinia virus, (4) adeno-associated virus (AAV), and (5) herpes simplex-1 virus (HSV-1). According to the data provided in the *Journal of Gene Medicine Database* (Wiley Database, http://www.wiley.co.uk/genmed/clinical/) and updated in July 2005, 25% of the total 1076 gene therapy clinical trials were using AdVs (Figure 1). The popularity and high expectations generated toward AdVs is well expressed by the fact that both the first gene therapy tragedy and success used adenovirus to deliver transgenes (Table 1).

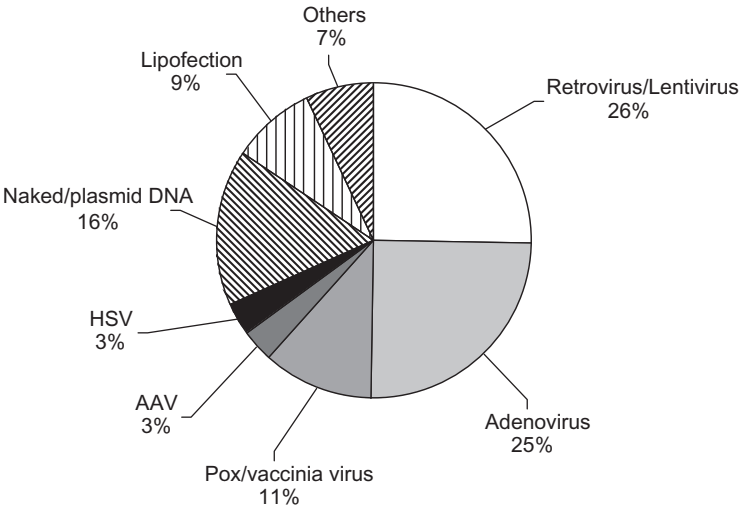


**Figure 1.** Breakdown of gene therapy clinical trials in terms of vector type ($N$ = 1076) (source http://www.wiley.co.uk/genmed/clinical/, July 2005). Abbreviations: AAV—adeno-associated virus; HSV—herpes simplex virus.

**TABLE 2. Characteristics of Adenoviruses Relevant for Human Gene Therapy Applications**

| |
|---|
| Efficient transduction of a large variety of quiescent and actively dividing cells |
| Relatively high capacity for transgene insertion |
| Convenient and simple methods of vector construction |
| Ability to replicate at high titers in complementing cell lines |
| High stability, which allows efficient purification and long-term storage |
| Biology has been extensively studied and is well understood |
| Transient transgene expression due to episomal location of viral genome |
| Immunogenicity |
| Insert-size limit of 7.5 kb |

## 2 ADENOVIRUS IN GENE THERAPY

### 2.1 Introduction

Adenoviruses were discovered in cultures of human adenoids in 1953 by Rowe et al. [24] and have since then been implicated in several respiratory, ocular, and gastrointestinal human diseases [25, 26]. Nonhuman adenoviruses have also been found in many other mammalian (dogs, horses, sheeps, chimpanzees, etc.) and nonmammalian (ducks, fowl, geese, etc.) species [26]. Overall, more than 100 members have been included in the *Adenoviridae* family. This section will only give a brief description of the major characteristics of the adenovirus and its vectors because this topic has been covered in depth in several recent reviews [26–31].

The extensive studies that followed the discovery of adenoviruses unveiled many features that make them a popular gene delivery vector (Table 2) [11, 26, 32]. For instance, adenoviruses are extremely efficient in the transduction of both

quiescent and actively dividing cells in most tissues [19, 26] and their genome can be manipulated easily to generate recombinant adenoviruses with improved properties. It is also possible to propagate them to high titers, and thus, it becomes relatively easy to generate sufficient amounts for research purposes and small-scale clinical trials [26]. On the down side, adenoviruses can generate potent immunogenic reactions in human recipients [28], although in some instances this immunogenicity may enhance antitumor effects [19] and vaccination efficiency [33]. Furthermore, and because adenoviruses persist in the nucleus of transduced cells as extra-chromosomal episomes, the expression of transgenes is transient [19, 26].

## 2.2 Biology and Properties

The 51 distinct serotypes of human adenovirus (Ad1 to Ad51) are classified into six groups (A to F) on the basis of sequence homology and hemagglutination properties [25]. Most AdVs used in gene therapy are derived from serotype 2 (Ad2) and 5 (Ad5) of the subgroup C [26]. The efficacy of Ad2 and Ad5 derived vectors, however, may be limited by the preexistence of humoral and/or cellular immunity to these serotypes in most human populations. Thus, other serotypes such as Ad11 and Ad35 to which most humans do not have neutralizing antibodies may be clinically useful [34, 35]. The antivector immunity may also be circumvented by the physical shielding of the adenovirus coat and the use of nonhuman adenovirus serotypes [36].

Adenoviruses are nonenveloped viruses with 26- to 45-kbp-long linear genomes of double-stranded DNA (the genome of human Ad2 comprises 35,937 base pairs) [37]. The genome is encapsidated in an icosahedral protein coat (12 vertices, 20 surfaces) made essentially of 240 nonvertex hexons and 12 vertex pentons, each with one or two protruding fibers (Figures 2 and 3). Each hexon protein is a trimer of the identical polypeptide II (pII), and each penton protein is formed by the interaction of five polypeptides (pIII). These pentons are tightly associated with one or two fibers made of three polypeptides (pIVs) each [37]. Overall, the fiber protein is composed of an N-terminal tail for penton binding, a rigid shaft, and a distal globular knob domain responsible for interaction with host cell receptors [38, 39]. The net charge of each hexon monomer in the Ad5 serotype is –23.8. This makes adenoviral capsids highly negative with an overall surface charge of over –17,000 [40, 41]. Other minor protein components can be found in the capsid, including protein IIIa, VI, VIII, and IX [37]. The size of the adenovirus particle is around 70–110 nm, and its molecular weight falls within the range $150–180 \times 10^6$ [39]. The buoyant density of an adenoviral particle in CsCl is 1.32–1.35 $g\,cm^{-3}$ [39]. Assuming a $170 \times 10^6$ MW and a 35,937-bp-long genome [42], the protein and DNA content of one Ad2 viral particle (VP) can be estimated to be $24 \times 10^{-5}$ (86% of dry weight) and $4 \times 10^{-5}$ (14% of dry weight) pg/VP, respectively. The corresponding water content, if a 100-nm viral particle is assumed, is approximately 60%.

The core of an adenoviral particle consists of the DNA genome complexed with four polypeptides (pV, pVII, mu, TP). Figure 4 schematically shows the structure of an adenovirus genome. The extremities contain inverted terminal repeat (ITR) sequences (100–140 bp), covalently linked to terminal proteins (TPs), which function as replication origins [23]. The nearby ψ sequence at the left end of the genome

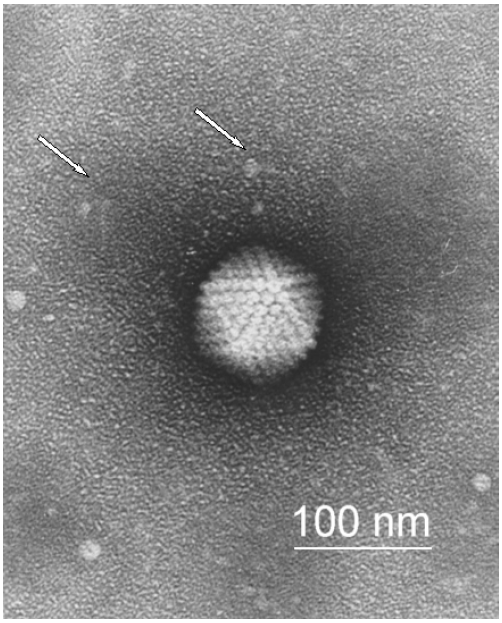


**Figure 2.** Human adenovirus C. The bar represents 100 nm. The arrows, upper left, point to fiber knobs. (Micrograph reprinted with authorization from Dr. H.-W. Ackermann, Department of Microbiology Medical Faculty Laval University, Quebec, Canada.)

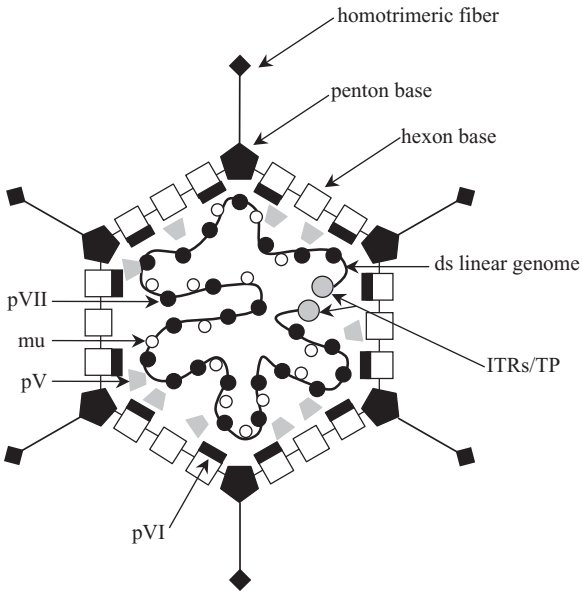


**Figure 3.** Schematic structure of an adenovirus particle. Major capsid (hexon, penton, fiber) and core (pV, pVII, Mu) proteins are shown. Abbreviation: ITR—inverted terminal repeat; TP—terminal protein. (Adapted from Ref. 27.)

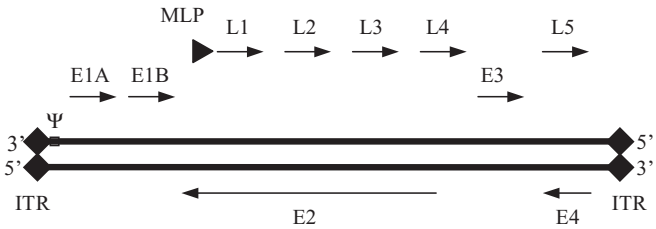


**Figure 4.** The adenovirus genome. The early (E1–E4) and late (L1–L5) transcription units are identified by the arrows. The inverted terminal repeats (ITRs), packaging sequence (ψ), and major late promoter (MLP) are also shown. (Adapted from Ref. 26.)

consists of a series of seven repeats and is required for efficient packaging [26]. The genes in both strands are grouped into early and late transcriptional units. The basis of this classification is the two-phase infectious cycle (30–40 h), which is characteristic of adenoviruses. In the "early" phase, the virus particles enter the host cell through binding of the homotrimeric protruding fibers to the coxsackievirus B and adenovirus receptor (CAR) [26]. After endosomal uptake, release, and capsid dismantling, the viral genome is transported and delivered to the nucleus [38]. The early genes E1 to E5 are then selectively transcribed and translated to modulate functions of the host cell and thus create an optimal environment for virus replication [26]. Specifically, this involves driving the host cell into S-phase (E1A gene), suppressing the host's cell apoptotic machinery (E1B gene), modifying the host's immunological environment (E3 gene), encoding proteins for viral DNA replication (E2 gene), and blocking cellular protein synthesis [43].

The "late" phase of the infectious cycle involves the synthesis of structural proteins of the virus, capsid assembly, genome encapsidation, and maturation of fully infectious viral particles. These events are associated with the transcription and translation of the "late" genes L1 to L5 [26, 27]. A key player in the transcription of the late genes is the major late promoter (MLP), which is activated early on via the E2 gene [27, 29].

### 2.3 Recombinant Adenovirus Vectors

A safe and effective use of AdVs requires the engineering of natural viruses into useful recombinant adenovirus vectors by using molecular biology tools [26]. These modifications are essentially designed to: (1) impair replication and expression of viral proteins and thus minimize toxicity and immunogenicity [19], (2) provide room in the viral genome for the therapeutic transgenes, and (3) modulate transgene expression. Once these recombinant vectors are available, transgenes can be inserted in their genome. In many cases, tissue- and cell-specific heterologous promoters are inserted along side to provide better expression [27, 29].

***First-Generation Vectors.*** Most first-generation AdVs have been constructed by deleting the viral early gene E1 (Figure 5). The goal underlying this deletion is to impair viral replication and production of capsid proteins and thus prevent the vectors from causing disease in the gene therapy recipient [27]. Simultaneously, room is made for transgenes up to 6.5 kb, which are usually expressed under the

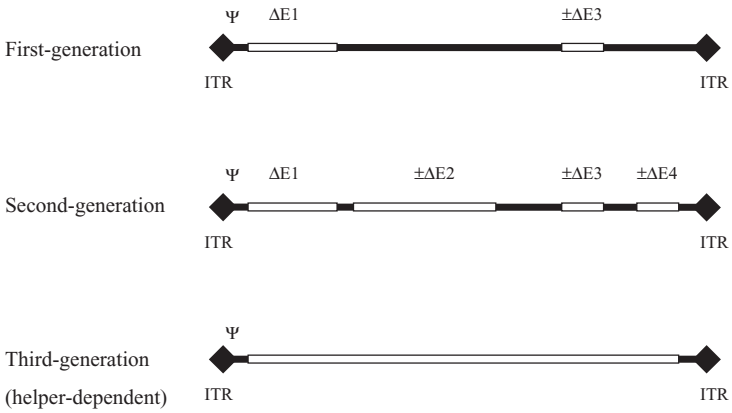


**Figure 5.** The three generations of adenovirus vectors. (A) First-generation vectors lack the E1 and sometimes the nonessential E3 region. (B) Second-generation vectors contain additional deletions of the E2 or E4 region. (C) Third-generation or "gutless" vectors are constructed by removing all or most viral genes and maintaining the *cis*-acting elements ITR and ψ. (Adapted from Ref. 26.)

control of an heterologous promoter [27]. The propagation of these E1 replication defective vectors can only be accomplished in complementing cells lines such as the human embryonic kidney 293 (HEK-293) [44] or the human retinoblast (PER. C6) [45]. These cells are transformed with the E1 gene and therefore transcomplement its deficiency in the recombinant vectors.

First-generation AdVs are characterized by several problems. For instance, recombination between the replication-defective virus and the E1 sequences in the complementing HEK-293 cell line can occur during propagation, giving rise to replication-competent adenoviruses (RCA) that will contaminate the recombinant virus being produced [26]. This problem has been circumvented in the PER.C6 cell line in which a rational design of the E1 transgene prevents generation of RCAs [45]. Another problem with first-generation vectors is that despite the deletion of the E1 gene, viral protein expression is not completely shut off. The concomitant low level of viral replication taking place in the transduced host cells, thus, can generate a cytotoxic-T lymphocyte (CTL) immune response that in the end will destroy the transgene-expressing cells and contribute to gene silencing [26]. Despite this residual immunogenicity, first-generation vectors may find application in cases such as cancer therapy and vaccination where a short-term transgene expression is desirable [26]. It is worth mentioning that the transgene product itself, depending on its own intrinsic immunogenicity, can also trigger a cellular immune response capable of destroying the transduced cells [29]. However, the magnitude and nature of this immune response seems to vary considerably depending on the specific transgene product being expressed [46].

***Second-Generation Vectors.*** Second-generation AdVs have been constructed by additionally deleting the viral regulatory E2 and/or E4 coding sequences (Figure 5), on the expectation that progressive deletions should reduce viral antigen expression, increase *in vivo* persistence, and reduce antiviral immune response

[47]. The E2 gene encodes proteins that are required for initiation and elongation of viral DNA synthesis and the activation of the MLP, whereas E4 encodes proteins involved in the accumulation, splicing, and transport of early and late viral mRNA and in DNA replication and virus particle assembly [47]. As in the first generation, complementing cell lines able to express the missing functions must be generated and isolated to ensure propagation of the recombinant viruses. The HEK-293 cell line can be conveniently adapted to this end [48]. A first advantage of these doubly deleted vectors is the reduced probability of the emergence of RCAs, because this would require simultaneous reversions in the E1 and E2 or E4 regions [47]. Several studies with E1/E2 and E1/E4 deleted vectors have shown reduced synthesis of viral proteins, extended transgene expression, and reduced toxicity [49–51]. Other studies, however, have challenged these results [47].

***Third-Generation Vectors.*** A third generation of AdVs has been further developed by deleting most or virtually all viral genes in the genome (Figure 5) [52–55]. These gutless or helper-dependent vectors can accommodate inserts up to 30 kbp and have shown reduced immunogenicity and prolonged transgene expression in mice [55]. The reduced immunogenicity still observed should be attributed either to the injected viral proteins or to the transgene product [29]. Growth of helper-dependent vectors depends on coinfection of an E1-complementing cell line such as HEK-293 by an E1-deleted helper virus (hence, the name helper-dependent) that provides missing replication and assembly functions *in trans* [26, 52]. The simultaneous propagation of the helper virus can be limited by resorting to a Cre-lox mechanism that allows the packaging sequence to be deleted [53]. In short, the helper genome is provided with loxP-sites that flank the ψ signal, whereas complementing 293 cells are engineered to stably express the enzyme Cre-recombinase. During infection, Cre-recombinase excises the loxP-flanked signal rendering the helper genome unpackable [29]. Nevertheless, the two types of viral particles (helper-dependent and helper) must be separated after propagation, a task that is difficult to accomplish [52].

***Oncolytic Adenoviruses.*** A class of AdVs has been developed to specifically kill cancer cells, an approach that has been termed virotherapy [56]. Contrary to the approach pursued with first-, second-, and third-generation adenovirus vectors, these oncolytic or conditionally replicating adenoviruses (CRAds) are engineered in such a way that they retain the ability to replicate. This replication however, is tumor selective—the virus will only infect and proli-ferate in malignant cells with mutations in specific tumor suppressor genes [2, 19, 56]. Thus, a transgene is actually not required for the therapeutic effect to take place.

## 2.4 Targeting Strategies

Several therapeutically relevant human cells (e.g., skeletal and smooth muscle cells, endothelial cells, hematopoietic cells, and some tumor cells) are not easily transduced by adenovirus vectors due to low or inexistent levels of expression of the CAR receptor. On the other hand, the broad tropism of adenoviruses may lead to the undesirable expression of the transgene in nontarget cells. Thus, an improvement in the transduction selectivity of existing adenovirus vectors is crucial to

circumvent these limitations and improve efficacy in many clinical applications. The following strategies have been devised to meet this end: (1) structural retargeting, (2) tropism ablation, and (3) transcriptional targeting [38].

Structural targeting relies on the structural modification of the adenoviral capsid by genetic incorporation of peptides [57], IgG-binding domains [58], and fiber proteins from other serotypes [59], metabolic biotinylation [60], and PEGylation [61]. These ligands are selected or designed to direct vector attachment to alternative cell receptors in CAR-deficient cells [38]. The goal of tropism ablation, on the other hand, is to reduce transduction of nontarget, CAR-expressing cells. This "de-targeting" strategy can be accomplished by constructing "knobless" vectors or by introducing point mutations in the fiber protein knob [30, 38]. The capsid of these mutant vectors is then modified with ligands (e.g., peptides) adequate for transduction of the target cells (retargeting) [62–64].

In transcriptional targeting, cell-type-specific promoters are used to restrict transgene expression to specific tissues. The adenoviral particles may infect different cells, but the transgene is expressed only in those cells that actively express transcription factors required to drive expression from the cell-specific promoter [19, 33]. For instance, this approach has been used in the context of adenoviral-mediated treatment of gastrointestinal cancer [65]. This application has been limited by the undesirable expression of the transgene (which codes for HSVtk-herpes simplex virus thymidine kinase) in the liver due to the vector hepatotropism. Thus, the transgene was placed under the control of the cyclooxygenase-2 (cox-2) promoter, which is inactive in liver cells but active in many gastrointestinal cancers. Experiments showed that the cox-2 promoter could confine the cytocidal effect of HSVtk specifically to cyclooxygenase-2-positive gastrointestinal cancer, while mitigating the otherwise fatal hepatotoxicity [65].

## 2.5 Applications

Most adenovirus vector applications and clinical trials described and reported in the literature have targeted genetic diseases and cancer. This section will briefly mention the most significant applications. The reader is directed to several reviews for a more comprehensive description of applications [2, 19, 27, 33, 56, 66].

***Genetic Diseases.*** In the case of genetic diseases, the transgene encodes a protein that is missing or is defective in the host organism. A typical and well-studied example is cystic fibrosis, the first human disease targeted in an adenovirus gene therapy clinical trial [10]. Cystic fibrosis is an inherited, recessive disease caused by a variety of mutations in the gene encoding the CFTR protein [1, 67, 68]. Although multiple organs are affected, the lung is the life-threatening organ [25]. The virus-vector mediated shuttling of a normal copy of the CFTR gene toward the affected lung cells could potentially prevent the onset, or halt the progression, of the disease. The choice of adenoviruses as CFTR gene delivery vectors is logical given their natural ability to infect lung cells. Nevertheless, barriers such as the lack of CARs in airway epithelial cells and alveolar macrophages, and the specific pulmonary-associated T-helper cell response, have prevented a successful transgene expression [27]. Significantly, all clinical trials currently under way involving AdVs and directed toward genetic diseases target cystic fibrosis (*Journal of Gene Medicine Database*, http://www.wiley.co.uk/genmed/clinical/).

Adenoviral gene therapy has also targeted muscular dystrophy, a genetic disease characterized by progressive muscle weakness [69]. Muscular dystrophy is caused by mutations in the X-linked dystrophin gene (DMD), which lead to prematurely aborted dystrophin synthesis. The lack of this important structural protein in muscle cells causes fiber damage and membrane leakage. In this context, gene therapy attempts to deliver a dystrophin expression vector to the nuclei of striated muscle cells. Given the huge size of the DMD gene (2.6 Mb), mini-gene cassettes (14 kb) have been generated that are capable of expressing therapeutic levels of a functional dystrophin protein [70]. Nevertheless, vectors with a capacity large enough to accommodate the 14-kb dystrophin cDNA are still required. Unlike first- and second-generation AdVs that are limited to 7–8-kb transgenes, third-generation adenovirus vectors adequately meet this requirement. The use of such vectors has resulted in a prolonged expression of the transgene in mice muscle cells [55, 69, 70]. Despite these promising results, barriers such as the lack of adenovirus receptors in the target muscle cells and potent immune response have to be overcome before a therapeutic use is developed [27, 69]. Further examples of the application of AdVs to treat genetic diseases include OTC deficiency, factor VIII deficiency, Tay-Sachs disease, and glycogen storage disease II [27].

***Cancer.*** Different strategies have been pursued in an attempt to treat cancer via gene therapy, by taking advantage of molecular differences between normal and tumor cells [2, 27, 33]. Once transferred to the target cells, expression of the transgene delivers the antitumor effect. Clinical data suggest excellent safety when AdVs are injected locally. Synergistic effects with treatment options such as radiotherapy and chemotherapy have further improved the efficacy of cancer gene therapy [33]. Most efforts have relied on one of the following approaches: (1) tumor suppression, (2) suicide therapy, (3) cancer vaccination, and (4) virotherapy.

The tumor suppression approach attempts to induce apoptosis by delivering a tumor suppressor gene that is missing or defective in the tumor cells [2]. On the contrary, normal cells infected by the tumor suppressor delivery vector will not be detrimentally affected. The approach is perfectly illustrated with p53, a gene whose mutations have been associated with several tumors. The delivery of wild-type p53 gene efficiently induces apoptosis in cells of different tumors, as demonstrated in several phase I and II clinical trials [43, 71]. Additionally, this toxic effect can extend to neighbor, uninfected tumor cells, a phenomenon known as the "bystander effect". This effect has been attributed to the ability of p53 to block angiogenesis [2]. Combination of p53 with immunomodulatory genes, cytotoxic drugs, or radiotherapy may further improve efficacy [2, 27]. Recombinant adenoviruses encoding the p53 gene have been used to treat more than 20 kinds of cancer indications, including head and neck squamous cell carcinoma, lung cancer, breast cancer, and liver cancer [15]. Studies have focused both on the effect of adenoviral gene therapy alone or in combination with conventional therapies such as chemotherapy, radiotherapy, and surgery [15]. The efficiency of the adenovirus/p53 strategy is well illustrated with the results of a phase II/III clinical trial using the commercial product Gendicine. In this study, 135 patients with head and neck squamous cell carcinoma were divided into two groups. The first group (63) received Gendicine in combination with radiotherapy, whereas the second group (72) received

radiotherapy alone. Significant difference in terms of complete or partial tumor regression was shown between the two groups, with 93% of the patients responding in the Gendicine/radiotherapy group *versus* 79% in the radiotherapy group [15].

The second approach used to kill tumor cells is suicide gene therapy, which combines the delivery of suicide transgenes with the separate administration of a harmless prodrug. Once reaching the target cancer cells, the suicide gene expresses an enzyme that metabolizes the prodrug into a cytotoxic agent that kills cells. The diffusion of the cytotoxic agent into neighbor cells further generates a "bystander" effect that increases the efficacy of the strategy [2, 27]. An example of such a transgene/prodrug combination that has been tested in the clinic is herpes simplex virus thymidine kinase (HSVtk)/ganciclovir [2, 27, 33]. The suicide HSVtk transgene is first delivered to cancer cells, for example, via AdVs. The administered ganciclovir is then metabolized by the expressed HSVtk into ganciclovir triphosphate, a nucleotide analoge that blocks DNA synthesis [72–75]. When a DNA strand incorporates this analoge, chain termination results and cells die upon induction of apoptosis [76]. This effect is, of course, more pronounced in tumor cells, which divide much more actively when compared with native cells. Another feature that contributes to the attractiveness of the approach is the use of a drug such as Ganciclovir that is widely used clinically [76]. Another example of suicide cancer therapy is the combined use of *Escherichia coli* cytosine deaminase (CD) with the prodrug 5-fluorocytosine. In this case, 5-fluorocytosine is metabolized by the expressed CD into 5-fluorouracil, a pyrimidine antagonist that blocks DNA and RNA synthesis [2]. The subsequent direct and bystander inhibitory effect can be enhanced further by combination with radiotherapy [77, 78].

Cancer can also be treated by the adenovirus-mediated delivery of transgenes to tumor cells with the goal of boosting antitumor immunity [27]. These adenoviral "cancer vaccines" can harbor either immunomodulatory genes (e.g., IL2 or IL12) and/or tumor antigens (e.g., MART 1 or gp 100 melanoma antigens), which once expressed should induce tumor regression. Human dendritic cells isolated from patients have also been transduced with AdVs designed to express tumor antigens. The therapeutic effect is achieved after reinfusion of the transformed cells [27].

Virotherapy constitutes a fourth approach that has been attempted clinically to treat cancer. The oncolytic vector Onyx-015 (previously d-1042) is a characteristic example of virotherapy. The mode of action can be explained as follows. Both adenoviruses and tumor cells need to block the p53 function to replicate. In normal cells the intact p53 function blocks replication of the Onyx-015 vector [2, 19, 56]. In tumor cells, however, and because the p53 gene is mutated, infection with the adenovirus vector Onyx-015 is followed by replication and cell destruction. Furthermore, because after cell destruction an increased number of viral particles is released, the infection/propagation/cell death process can continue in neighbor cells [43]. This approach has been tested in several phase I and phase II clinical trials as described in Ref. 43. In one phase II study, the intratumoural injection of Onyx-015 combined with chemotherapy resulted in an 83% tumor response in head and neck cancer patients [79]. The efficacy of oncolytic vectors can be further improved by resorting to transductional targeting, such has the genetic insertion of an integrin binding RGD-4C motif in the fiber proteins of the vector. This modification resulted in enhanced infectivity in ovarian cancer cells, which suggests improvements in clinical efficacy [33].

### 2.6 Safety Aspects

Adenovirus infection triggers both cellular and humoral responses. Viral proteins (either synthesized *de novo* or not) are processed by antigen-presenting cells and presented to CD8+ T cells (by means of MHC class I molecules) and CD4+ T-helper cells (by means of MHC class II molecules). This induces proliferation of cytotoxic T lymphocytes that specifically destroy the infected cells. CD4+ T-helper cells are also involved in the production of adenovirus-specific neutralizing antibodies directed toward the virus capsid [29]. The risks associated with this immunogenicity have been vividly demonstrated by the death of a patient in the 1999 OTC clinical trial. After the direct administration of a high dose of the adenoviral vector ($3.8 \times 10^{13}$ viral particles) to the liver, wide dissemination into the circulation triggered a massive activation of innate immunity followed by systemic inflammation that led to fever, intravascular coagulation, and multiorgan failure [19]. Responsibility for eliciting the immune response was attributed to viral capsid proteins rather than the transgene [19]. In subsequent years, a wealth of evidence accumulated confirming that AdVs may induce harmful immune and inflammatory responses, especially when large doses are administered systemically [19]. Adverse effects to adenovirus administration, however, may vary from individual to individual and depend on predisposing and underlying conditions [19].

The improvement of safety through minimization of toxicity and immunogenicity has been one of the major drivers in the development of recombinant AdVs. Nevertheless, even when third-generation "gutless" vectors are used, reduced immunogenicity due to the injected viral capsid proteins or to the transgene product is still observed [29, 55]. Thus, a need exists for the development of safe and effective methods capable of downregulating the host immune response against both adenoviral capsid proteins and transgene products. For instance, conventional immunosuppressive agents such as cyclosporine and FK506 may partially reduce immune responses and thus increase transduction and long-term transgene expression [46]. However, low doses of these agents must be used to prevent substantial organ toxicity. Improved results have been obtained by combining the use of immunomodulatory immunoglobulins and immunosuppressive agents. With this strategy, the immune response against adenovirus proteins and the transgene product dystrophin has been abrogated to a degree not achievable with the use of either agent alone [46].

Another problem associated with immunity to AdVs is the generation of memory cells. Upon subsequent administration of the same vector, the immune response is boosted by these cells, effectively reducing the efficacy of the repeated dosing [11].

## 3 ADENOVIRUS MANUFACTURING

The increasing number of adenoviral gene therapy applications that are moving from the laboratory to the clinic is creating a need for large amounts of highly purified recombinant AdVs. This demand is expected to increase as the first products reach the market. Thus, there is a clear need for a parallel development of efficient, scalable, and reproducible adenovirus manufacturing processes capable of delivering high amounts of infective AdV particles [80].

### 3.1 GMPs and Validation

Recombinant AdVs, such as all products that are to be administered to humans or animals, must be manufactured in accordance with a set of regulations issued by regulatory authorities such as the U.S. Food and Drug Administration (FDA) or the European Medicines Agency (EMEA) in the European Union. These regulations are known as current Good Manufacturing Practices (cGMPs) and cover all aspects of the production, from choosing and testing raw materials, to utilities, packaging, shipping, and transferring of final products to the clinic [32, 81]. If these items are not in conformity with GMPs, the product is deemed to be adulterated and cannot be legally approved [82]. In the specific case of AdVs, the facilities and processes used in manufacturing should be designed carefully to guarantee maximum protection to the product, to personnel, and to the environment [32, 83]. For instance, dedicated facilities should be used to prevent cross-contamination from previous or parallel batches of other bioproducts [32]. Confinement of production operations to class C clean rooms is also advisable to avoid dissemination of recombinant adenoviruses into the air [32].

One cornerstone provision of cGMPs is validation, a concept introduced to assure product consistency [84]. The validation of downstream processing operations aims to prove that they are capable of consistently removing impurities (e.g., host cell components, process-related materials, adventitious agents) to acceptable levels. Additionally, acceptance limits and operating ranges for each step must be determined [84]. Validation studies usually lead to optimized processes with reduced variability and, as a consequence, to a decrease in the number of failed batches. Thus, the development of adenovirus manufacturing processes (and associated facility) should be undertaken with validation in mind, not only to improve quality assurance and accelerate approval, but also to reduce costs [85, 86].

### 3.2 Product Specifications and Quality Control

One core concept hovering GMPs is quality assurance (QA). Among other attributes, QA is a means of guaranteeing the excellence, security, and dependability of the manufacturing process or its product [81]. QA is a key issue in process development, validation, and product approval, as well as on the assessment of the endproduct quality in comparison with product specifications. The characterization of recombinant AdVs thus constitutes a crucial aspect in all steps of product development, from basic research to clinical trials [87]. As it is the case for other biologics, adenovirus products have an inherent variability in their composition, stability, and potency. They are also subject to the variability inherent to the biological nature of some of the methods used to test them. An AdV should be well characterized to demonstrate that it is consistent in composition, exhibits long-term physico-chemical and biological stability, and is free of adventitious agents (microorganisms, adeno-associated viruses), contaminants, and impurities [88]. This requires the development and setup of a range of "validatable" analytical methodologies capable of fully characterizing the product during processing and in its final formulation, while ensuring the production of a consistent product [87]. The molecular structure of AdVs, hence their biological activity (potency), may be sensitive to factors such as temperature (labile for $T > 40°C$), pH (labile for $pH < 6$), and freeze/thawing cycles [89, 90]. Thus, a correct evaluation of product stability

requires the use of convenient analytical techniques (e.g., circular dichroism and light scattering [89]). Analytical techniques also play an important role in the analysis of source materials, in the assessment of the impact of changes in manufacturing processes, and in the validation of processes and cleaning.

The exact final specifications (identity, efficacy, safety, potency, and purity) for an adenovirus vector product will usually depend on the intended therapeutic use, and thus, they are defined during clinical trials. However, regulatory agencies such as the FDA, the EMEA, or the SFDA provide guidelines and quality standards that are helpful during product and process development [91–93]. Table 3 exemplifies some of the assays, methods, and specifications used in the quality control of Gendicine (adapted from Refs. 15 and 93).

### 3.3 Environmental and Safety Issues

An assessment of the environmental impact of a process is an essential part of the design [94, 95]. The costs associated with the treatment and disposal of the waste generated by a particular process solution should be estimated beforehand. This information is one key element in the final decision on which process to select. Environmentally unfriendly operations such as those that generate large amounts

**TABLE 3. Specifications and Recommended Assays for Assessing AdV Purity, Safety, and Potency (adapted and abbreviated from Refs. 15 and 93).**

| Parameter | Analytical Method | Specification |
|---|---|---|
| | *Adenovirus-transgene* | |
| Appearance | Opalescence | Visual inspection |
| pH | pH meter | 8.0–8.5 |
| Identity | | |
| adenovirus genome | Restriction mapping | According to sequence |
| p53 gene | PCR | 396 bp |
| Viral titer | | |
| viral particles (VP) | Absorbance at 260 nm | $\geq 6.7 \times 10^{11}$ VP/mL |
| infectious units (IU) | $TCID_{50}$ | $2–4 \times 10^{10}$ IU/mL |
| IU/VP | — | ≥3.3% |
| Purity | A260/A280 ratio | 1.2–1.3 |
| Potency | | |
| expression | Western blot | Positive |
| bioactivity | Saos-2 cell bioassay | Positive |
| | *Impurities* | |
| RCA | A549 cell bioassay | $\leq 1$ RCA/$3 \times 10^{10}$ VP/mL |
| AAV | PCR | Negative |
| Proteins | ELISA | ≤66 ng/mL |
| Genomic DNA | Southern blot | ≤6.6 ng/mL |
| Endotoxins | *Limulus* ameobocyte lysate (LAL) | ≤6.6 EU/mL |

*Abbreviations*: $TCID_{50}$ = tissue culture infective dose; RCA = replication-competent adenovirus; AAV = adeno-associated virus.

of hazardous solvents and materials, which are usually costly to dispose of, should definitely be avoided. Additionally, adequate systems should be in place to provide an efficient decontamination of solid (e.g., autoclaving and incineration) and liquid wastes (e.g., heat or caustic inactivation) before release from the manufacturing area [32, 96, 97]. Once inactivated, these wastes should pose limited risk to the environment. Nevertheless, process optimization should always be geared toward minimization of waste generation [94, 95]. The formation of aerosols during processing should be minimized because adenovirus can spread via airborne transmission and contribute to dissemination within the manufacturing rooms and facility. Unit operations that need special safety precautions (e.g., explosion-proof tanks, blow-out walls, emission containment, personnel protection, etc.) to operate should also be carefully considered, because these requirements can dramatically increase the cost of equipment, building design, and construction. Overall, environmental and safety issues require an intimate knowledge of the process technology solutions available.

### 3.4 Process Considerations

A process for the manufacture of an AdV consists of several activities aimed at the production of a certain amount (measured as number of viral particle [VP] units or infectious viral particle [IVP] units) of the target product at an acceptable cost and quality (Figure 6). The generation of fully qualified cell banks and virus seeds is at the forefront of these activities [32, 83]. These banks should be extensively tested for contaminants, including RCAs [83]. A rigorous screening of raw materials, especially those of biological origin, is also mandatory to guarantee that extraneous organisms and contaminants are not introduced into the process [83]. Upstream and downstream processing unit operations are selected, arranged, designed, and operated to manufacture a bulk product. After filling and finishing, the product can be distributed and shipped. These activities should be GMP-compliant and thoroughly scrutinized under a quality control program, as discussed in the previous sections (Figure 6).

Flowsheets for the recovery and purification of biologicals are usually established with several rules of thumb and on the basis of accumulated experience with the target product [95]. Simulation tools can also be used to rapidly evaluate bioprocess alternatives and speed up development [95, 98–100]. The specifics of adenovirus purification were first addressed by researchers that have developed a range of

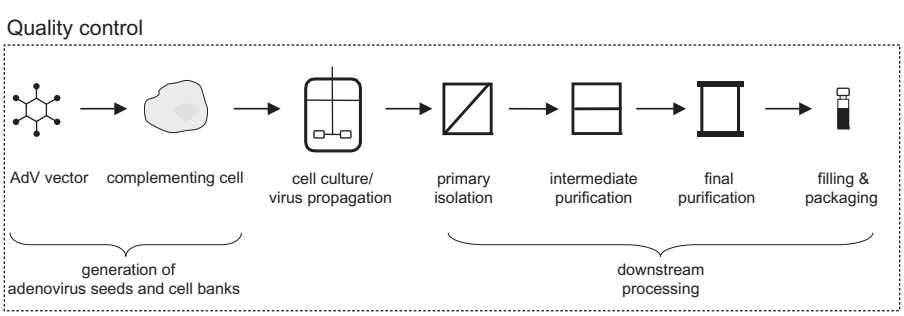


**Figure 6.** Outline of the activities involved in a typical AdV production process.

efficient lab-scale protocols, most of them based on CsCl density-gradient ultra-centrifugation [101]. Unfortunately, many of these protocols use reagents (e.g., CsCl is toxic) that are not acceptable for the manufacturing of a biological pharmaceutical. Furthermore, ultra-centrifugation is not amenable to scale-up due to the limited capacity of commercial ultra-centrifuges [41]. Nevertheless, many published adenovirus production processes include modifications or adaptations of specific steps used in these laboratory procedures [40, 101–107].

Cell culture is usually optimized to obtain high cell densities (cells/mL) and specific titers of infectious viral particles (IVPs/cell). The final goal is of course to maximize overall mass production (specific titer × cell density = IVP/mL) [80]. Next, a sequence of unit operations must be set up to recover the viral particles and eliminate host cell impurities (genomic DNA, RNA, proteins, RCAs, etc.) until the desired level of purity is met (Table 3). For a product with an intended use in humans, this removal of impurities is mandatory to avoid side effects upon administration to patients. The downstream processing unit operations can be grouped into three different stages: primary isolation, intermediate purification, and final purification (Figures 6 and 7). Ideally, the overall process should have a limited number of high-recovery steps, so that processing costs are reduced and acceptable yields are obtained [108]. The process should also use Generally Regarded As Safe (GRAS) reagents. Furthermore, lengthy operations and processes should be avoided to cut costs from overhead, amortization of equipment, and direct labor charges [109]. Processes with overall adenoviral particle yields of 32% [105], 50% [107], 60% [106], and 71% [104] have been reported in the literature.

### 3.5 Production

Most E1-deficient recombinant adenoviruses are propagated in the HEK-293 and PER.C6 cell lines (and derivatives thereof) as discussed in Section 2.3. The generation of RCAs is minimal in PER.C6 but remains a concern for HEK-293. Both cell lines have been documented for GMP manufacturing [110]. The production of these cells can be accomplished by a variety of methods, which depend on whether adherent or suspension cell lines are being used.

An adenoviral production process with HEK-293 and PER.C6 is a two-phase process. In the first phase, cultures are started at an appropriate seeding density (e.g., $0.3 \times 10^6$ cells/mL) and allowed to grow until a cell density around $0.5–1 \times 10^6$ cells/mL is reached. The second phase starts at this point with the infection of the culture at a ratio of virus titer to cell density (known as multiplicity of infection—MOI) adequately chosen (e.g., MOI = 10). An exchange of medium is usual, but not mandatory, before infection [111–113]. The cell density at the time of infection is a crucial parameter, with specific viral particle productivity (IVP/cell) usually dropping for cell densities higher than $0.5 \times 10^6$ cells/mL [80]. After infection, viral particles propagate and accumulate in the cells with maximum titers ($\sim 10^{10}–10^{11}$ VP/mL) typically obtained 48 hours postinfection (hpi). Maximum cell densities of $1 \times 10^7$ cells/mL have been reported [111]. Loss of viability and cell death ensue thereafter [80].

***Adherent Cell Culture.*** Adherent or anchorage-dependent cell lines require attachment to a surface for their survival and replication. These cultures are labor

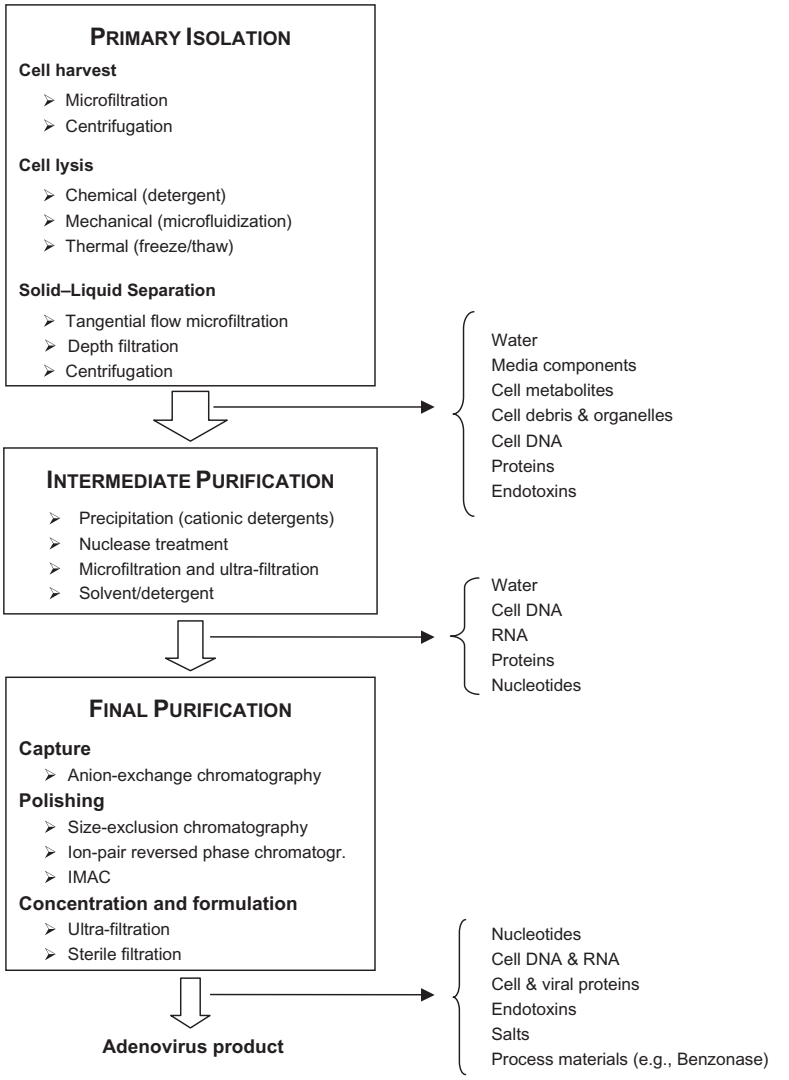


**Figure 7.** Generic block of downstream processing of AdV vectors showing unit operation options in each section.

intensive, require both a large amount of space and specific equipment handling, and have a limited potential for scale-up. Thus, they are more adequate during the early stages of product and process development [110, 114, 115].

Adherent cells can be grown in stationary or microcarrier cultures [116]. Stationary cultures use devices such as flasks, roller bottles, and cell factories that provide wall surface area for cell attachment [117]. Microcarrier cultures, on the other hand, can be implemented in agitated bioreactors [114, 116]. These microcarriers are small porous particles (0.2 mm) that provide large surface areas for cell attachment and can be suspended in the culture medium by gentle agitation. The culture environment is thus easily controlled and the scale-up is done by increasing the

bioreactor volume. In comparative studies, stationary cultures usually yield higher adenovirus titers (up to $1.1 \times 10^4$ VP/cell) than microcarrier cultures (up to $8.5 \times 10^3$ VP/cell) [114, 116].

***Suspension Cell Culture.*** Suspension cell lines are more convenient for large-scale manufacturing [111, 112]. A further advantage is that they can be grown in serum-free media, without the addition of bovine-derived components [110]. This advantage reduces costs and batch-to-batch variability and facilitates downstream processing and validation [111]. Conveniently, both the HEK-293 [80, 111–114] and the PER.C6 [118, 119] cell lines can be adapted to grow on suspension cultures. When grown in serum-free media, these cell lines allow high growth rates and cell densities as well as high yields of adenoviral particles [111]. The tendency of cell derivatives to aggregate, however, is usually a recurrent concern with suspension cultures, especially at high cell densities [111, 120].

The cultivation of cells and adenovirus propagation in suspension bioreactors has been performed in batch, fed-batch, and perfusion modes. In batch operation, no nutrients are added during the course of growth or infection. Although batch cultures are simple to perform and contamination risks are low, essential nutrients are rapidly depleted from the medium and cell viability decreases [110, 121].

The fed-batch mode is used to maintain cell viability after infection by supplementing limiting nutrients (e.g., glucose, glutamine, and amino acids) or reducing the accumulation of toxic metabolites (e.g., ammonia and lactate) [48, 80, 112, 113]. It also improves viral titers by increasing the cell density at which cells can be infected up to $2 \times 10^6$ cells/mL, without reducing the per-cell yield of product [121]. Both the medium composition as well as the feeding strategy are important parameters [112]. Nutrient depletion and the buildup of toxic metabolites cannot be completely alleviated by fed-batch cultures [121]. Operation under perfusion mode can overcome this limitation by ensuring a constant renewal of medium in the bioreactor. Briefly, cells are retained at high concentrations inside the bioreactor by using devices such as hollow fibers and membrane units, whereas fresh medium is continuously supplemented [121]. Consequently, the cell density at which cells can be infected can be increased. With this strategy, HEK-293 cell densities of $8 \times 10^6$ cells/mL could be achieved at the time of infection, and viral titers of $7.8 \times 10^9$ IVP/mL were obtained at 60 hpi [121].

### 3.6 Downstream Processing

***Primary Isolation and Intermediate Purification.*** The suspension generated during cell culture constitutes the starting point for downstream processing. In the primary isolation stage, cells are harvested and viral particles are released from the cells together with other impurities. A significant reduction in volume occurs, and the most plentiful impurities such as extracellular liquid, proteins, genomic DNA, and cell debris are removed. Cells are typically harvested by centrifugation or microfiltration. The resulting cell slurry is then resuspended in an adequate volume (1/10 to 1/20 of the culture volume) of a suitable buffer or of the culture medium itself. The subsequent cell lysis operation is typically performed using freeze/thaw cycles [40, 106, 118, 119], osmotic shock [80], sonication [101, 122], microfluidization [102], or by the addition of detergents such as Triton X-100 or

Tween-20 [104, 107]. Unlike lysis by freeze/thawing and sonication, cell shearing by microfluidization is more rapid, reproducible, and easy to scale-up [102].

After lysis, cleared lysates are obtained by removing cell debris and larger unruptured organelles with centrifugation [40, 118, 119], depth filtration [107], or tangential flow microfiltration [102, 106] operations. Treatment with nucleases (e.g., Benzonase [Merck, Gibbstown, NJ], Pulmozyme [Roche, Basel, Switzerland], DNase, RNase T1, and RNase I) is usually performed before or after clarification, to reduce the cellular DNA and RNA load [40, 80, 104]. This step not only improves purity, hence the safety of the viral product [40], but also it reduces agglomeration of viral particles, which is usually induced by adhesion of nucleic acids [104]. In some cases, nuclease treatment can be carried out at the end of the process, after chromatographic purification [106].

As an alternative to nuclease digestion, cationic detergents can be added during lysis to selectively precipitate cellular DNA. Recent experiments with the detergent domiphen bromide have shown that it is possible to obtain three logs of DNA clearance without losses in the infectivity of purified viral particles [107]. Thus, the use of DNA removal operations such as nuclease treatment and anion exchange chromatography may be eliminated or reduced.

The viral particle-containing solutions resulting from nuclease treatment and cationic detergent precipitation are typically filtered, concentrated, and conditioned before chromatographic purification [40, 80, 101, 106, 107]. The inclusion of a solvent/detergent step at this stage may be included to inactivate potential enveloped viruses that could have been coamplified [32].

***Final Purification.*** A pharmaceutical product such as an AdV requires a high degree of purity. This level of purity is usually achieved with a combination of chromatography and filtration operations. The goal is to separate the viral particles from the most recalcitrant impurities that persist in the streams. The chromatographic operations described in the literature for virus purification explore properties such as size, charge, hydrophobicity, and metal affinity [41, 80, 102–107].

Anion-exchange (AEX) chromatography is widely used as a viral particle capture step. It explores the interaction between the negatively charged hexons in viral capsids and the stationary phases bearing positively charged ligands such as quaternary amines [41]. Impurities (media components, low-molecular-weight DNA, penton, hexon, and fiber proteins are eluted at low salt (<0.25-M NaCl), whereas bound capsids are displaced with a salt gradient [41, 102, 103, 106]. A recent study shows that the NaCl concentration required to elute viral capsids from the anion exchange column is a function of the serotype being purified (ranging from 0.27 to 0.45 M), and it correlates very well to the electrostatic properties of the hexon protein in each serotype [41]. Tentacular (e.g., DEAE-Fractogel; Merck, Gibbstown, NJ), perfusion (e.g., POROS.RTM.50D; PerSeptive Biosystems, Framingham, MA), and soft gel in rigid shell (e.g., Ceramic HyperD.TM.F; BioSepra, Villeneuve-La-Garenne, France) materials have all been used for AEX purification of adenovirus particles [102, 106]. The reported AEX viral particle yields range from 63% [104] to 80% [80].

AEX can also be performed in expanded-bed adsorption (EBA) mode [102]. Under this mode of operation, cell lysates can be applied directly to the column from below. Large debris and unlysed cells that can move freely around the anion-exchanger beads eventually leave through the top of the column, whereas

adenoviral particles bind to the anion-exchange matrix. Extensive washing from below limits nonspecific interactions between the particulates and the resin. Finally, the flow is reversed, the anion-exchanger beads are allowed to pack, and the viral particles are eluted under a salt gradient [102]. Due to its early use in the downstream processing, EBA can be considered as an intermediate purification unit operation.

Adenovirus particles have also been purified by affinity chromatography by taking advantage of an ingenious targeting approach. Briefly, the fiber capsid protein of adenovirus 5 vectors was genetically fused to a biotin acceptor peptide (BAP). These BAP-modified fibers were then metabolically biotinylated during virus propagation in the HEK-293 cell line. The resulting covalently biotinylated viral particles could then be purified from crude lysates by avidin-affinity chromatography [122].

AEX alone is unlikely to be sufficient to produce an adenovirus product with the required purity [106]. The inclusion of a chromatographic polishing step may thus be necessary as a means to remove traces of DNA and protein impurities. Size-exclusion [80, 103], ion-pair reversed phase [106], and metal affinity [105] chromatography have all been used toward this end. In the case of size-exclusion, if an adequate matrix is selected (e.g., Superdex 200; GE Healthcare, Piscataway, NJ), virus elute in the flowthrough because of its large size, whereas low-molecular-weight impurities are retarded. The column loading volume can be increased up to 20% of the bed volume, because a group separation is achieved [102]. An extra advantage of size exclusion as a polishing step is that it can also be used to exchange buffer [103]. This enables formulation to be carried out simultaneously with polishing. In the case of ion-pair reverse phase chromatography, elution conditions are carefully selected to retain impurities in a PolyFlo matrix (Puresyn, Inc., Malvern, PA) and to allow the viral particles not to bind. Under this negative chromatography operation mode, viral particles are recovered in the flowthrough with an 87% recovery yield [106]. This leads to some dilution of the adenovirus stream, as is always the case when products are recovered in the flowthrough.

Processes for adenovirus purification typically end with concentration, formulation, and sterile filtration operations [40, 80, 106]. Concentration and formulation are usually carried out in ultra-filtration units equipped with 100–300-kDA membranes [40, 106]. The exact composition of the formulation buffer will depend on the intended application, mode of administration (injectable, aerosol), and required short-term and shelf stability [104, 123]. A typical liquid formulation may include an aqueous buffer supplemented with cryoprotectants (e.g., sucrose) and stabilizers such as the nonionic-surfactant polysorbate-80, the chelating agent EDTA, and the oxidation inhibitors ethanol and histidine [123]. Filtration under sterile conditions is typically performed with 0.22-μm membranes [103, 106].

## 4 ANALYSIS AND EVALUATION OF AN ADENOVIRUS PRODUCTION PROCESS

In this section, a pilot-scale process that has been developed specifically for the production and purification of AdVs for gene therapy [80] is analyzed and evaluated for large-scale manufacturing with the use of the process simulator software SuperPro Designer (Intelligen, Inc., Scotch Plains, NJ). The major objective of the

analysis presented here is to estimate the cost of the production of an adenovirus therapeutic product. Results further provide insights into process weaknesses and strengths, effectively directing bioprocess engineers toward better processes.

## 4.1 Process Description

As a design basis we have assumed a plant capacity of around $1 \times 10^{18}$ viral particles of purified recombinant adenovirus produced per year. This amount of product is sufficient to treat 125,000 patients per year on the basis of $1 \times 10^{12}$ VP doses given weekly for a total of 8 weeks, as described in a Gendicine clinical trial [15]. The plant is designed to operate 330 days a year, with a new batch initiated every 11 days—this corresponds to 30 batches at $33.3 \times 10^{15}$ VP/batch. The total batch time is around 17 days. Guidelines and quality standards issued by regulatory agencies have been used to set up product specifications in terms of final purity (Table 3). Finally, we have assumed that, at the end of the process, the bulk adenoviral product will be distributed in vials, each containing a $1 \times 10^{12}$ VP dose of 0.6 mL of sterile formulation buffer. This corresponds to a viral titer of $167 \times 10^{11}$ VP/mL, which is higher that the specified value of $6.7 \times 10^{11}$ VP/mL (Table 3).

The entire flowsheet for the production of adenovirus is divided into cell culture and virus propagation (Figure 8) and downstream processing (Figure 9) sections.

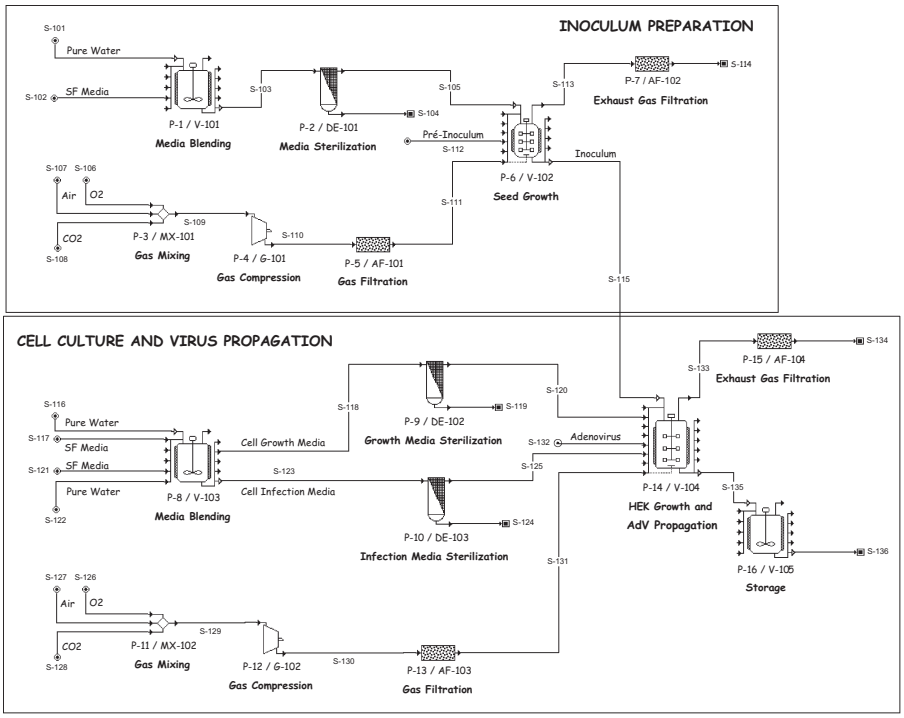


**Figure 8.** Adenoviral vector production flowsheet showing the cell culture and virus propagation section (implemented with the process simulator software SuperPro Designer [Intelligen, Inc., Scotch Plains, NJ]).

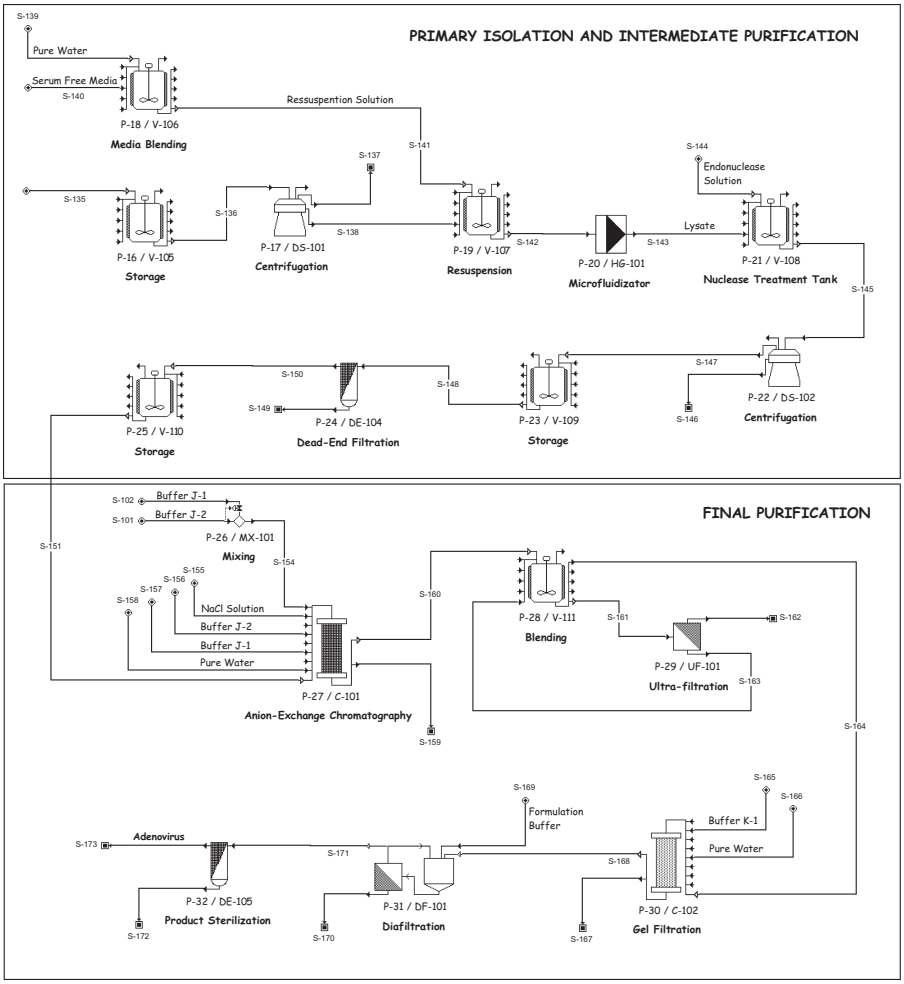


**Figure 9.** Adenoviral vector production flowsheet showing downstream processing section (implemented with the process simulator software SuperPro Designer [Intelligen, Inc., Scotch Plains, NJ]).

The overall adenovirus particle recovery yield per batch is around 67% ($33.3 \times 10^{15}$ VP are recovered out of the $5 \times 10^{16}$ VP that are present in the whole cell lysate). This figure is within yields reported in the literature for similar processes [104]. Input data used in the simulation software SuperPro Designer were taken from the reference publication [80] and supplemented with information from other published processes [40, 106, 107]. Educated guesses were made to provide for missing data. Table 4 summarizes the most important data—process volumes, step yields, and viral particle titers.

***Cell Culture Section.*** The activities in this section include media sterilization, inoculum preparation, cell growth, and virus propagation (Figure 8). Serum-free (SF) medium is sterilized by 0.2-μm filtration and used for both inoculum preparation and cell growth/virus propagation. The culture is assumed to take

**TABLE 4. Summary of Input Data Used in the Bioprocess Simulator SuperPro Designer**

| Step | Volume (L) | AdV Titer ($\times 10^{10}$ VP/mL) | Step Yield (%) |
|---|---|---|---|
| Cell culture | 1000 | 5.0 | — |
| Harvest + resuspension | 109 | 46 | 100 |
| Cell Lysis | 109 | 45 | 98 |
| Benzonase digestion | 109 | 45 | 100 |
| Centrifugation + dead-end filtration | 106 | 46 | 99 |
| Anion exchange chromatography | 118 | 36.5 | 89 |
| Ultrafiltration | 11.8 | 350.0 | 96 |
| Size exclusion chromatography | 59.0 | 66.5 | 95 |
| Ultrafiltration | 19.7 | 169.3 | 85 |
| Sterile filtration | 19.7 | 169.3 | 100 |

*Notes*: An overall yield of 67% is obtained. The final viral titer is higher than the specified value ($67 \times 10^{10}$ VP/mL).

place in fed-batch mode (with a single nutrient addition), in a bioreactor with a working volume of 1000 L. This size is a reasonable assumption, given that cultures with volumes as high as 10,000 L are currently being developed for production of HIV adenoviral vaccines [118]. The inoculum is added to 500 L of culture medium at a seeding density of $3 \times 10^5$ cell/mL, and cells are allowed to grow for 6 days. Then, 500 L of fresh SF medium are added, and cells at a concentration of $0.5 \times 10^6$ cell/mL are infected at a MOI of 10. Cell growth and virus propagation are conducted at 37°C and pH 7.2, with constant addition of a gaseous mixture (80% air, 10% $CO_2$, 10% $O_2$) at 0.05 vvm [80]. Harvest is performed 48 hpi, with typical values assumed for the final cell concentration ($1 \times 10^6$ cells/mL), adenovirus titer ($5 \times 10^4$ VP/cell), and ratio of infectious viral particle to total viral particles (1 : 10) [80].

***Downstream Processing Section.*** The activities in this section include cell harvest and cell lysis, nuclease digestion, and chromatographic purifications (Figure 9). Cells from the suspension culture (1 $m^3$) are harvested by continuous centrifugation, resuspended in SF medium (1/10 of the original volume), and lysed by microfluidization (e.g., 2000 psi and 1 pass) [80]. Cell DNA is then drastically reduced (≈4.5 logs) by adding Benzonase up to a final concentration of 150 IU/mL [107]. After removal of debris by centrifugation, the supernatant is clarified in a 0.2-μm dead-end filtration unit [80]. Adenovirus capture and purification is performed by AEX chromatography column (Fractogel EMD-DEAE, 0.4-m bed height, 0.6-m bed diameter) according to the instructions presented by Tang et al. [103]. Briefly, each cycle comprises six distinct operations: (1) equilibration with 22 bed volumes (BVs) of buffer J1 (50-mM sodium phosphate pH 7.5, 265-mM NaCl, 2-mM $MgCl_2$, 2% (w/v) sucrose) at 4 cm/min, (2) loading of 106 L of feed at 1 cm/min, (3) washing with 4 BVs of buffer J1 at 2 cm/min, (4) washing with 8 BVs of 94% buffer J1 and 6% buffer J2 (50-mM sodium phosphate pH 7.5, 600-mM NaCl, 2-mM $MgCl_2$, 2% (w/v) sucrose), (5) elution of bound adenovirus particles with 10 BVs of linear gradient from 6% to 100% buffer J2 at 2 cm/min, and (6)

column cleaning with 4 BVs of 1-M NaCl at 4 cm/min. The total cycle time is 12.26 h. The bound viral particles are eluted with an overall yield of 89% [103]. The adenovirus pool is then concentrated 10 times by ultra-filtration using a 100-kDa membrane [80]. A step yield of 96% is assumed based on data published for ultra-filtration of different viral particles [124]. A size exclusion chromatography column (Superdex 200, 0.8-m bed height, 0.4-m bed diameter) is included as a polishing step (95% step yield) [103]. The column is loaded with the totality of the viral solution (12% of the bed volume), and upon elution with an adequate buffer (e.g., 20-mM sodium phosphate pH 8.0, 100-mM NaCl, 2-mM $MgCl_2$, 2% (w/v) at 0.43 cm/min), a pool of adenoviral particles is obtained that is five times diluted relatively to the feed [103]. The total cycle time is 18.3 h. Concentration and formulation into an adequate buffer (2.5% glycerol, 25-mM NaCl, 20-mM Tris, pH 8 [80]) is subsequently carried out in an ultra-filtration unit equipped with a 100-kDA membrane. Finally, the adenovirus product is sterile filtered into glass vials ($1 \times 10^{12}$ VP in 0.6 mL).

***Process Scheduling.*** The scheduling and equipment utilization for one production batch is shown in Figure 10. The plant batch time is approximately 405 hours, with a new batch (i.e., inoculum preparation) initiated every 264 hours. This batch start time roughly corresponds to the fourth day of the cell culture in the previous batch. The inoculum preparation, cell culture, and virus propagation procedures, with a duration of approximately 367 hours, are clearly identified in the chart as the time bottleneck. Comparatively, the downstream processing is complete within a mere 38 hours. The time bottleneck in this section is the size-exclusion chromatography polishing step, which takes about 18 hours.

### 4.2 Inventory Analysis

The overall material balance per year is summarized in Table 5. Remarkably, the annual mass production of purified adenovirus vector constitutes a minor fraction (~0.3 ppm) of the total amount of materials required for its production. Apart from the adenovirus vector, cell debris, and gases, all output materials end up in liquid waste streams, which are disposed of after adequate treatment (e.g., heat or caustic inactivation) to minimize environmental impacts. Water is the major raw material used (~97%), most of it for equipment cleaning. This is typical in the production of biopharmaceuticals as can be seen by checking IgG [95], recombinant β-glucuronidase [125], and plasmid DNA [100] production examples. Serum-free medium (~0.2%) and gases (~1%) are used in the cell culture and virus propagation step. Large amounts of sodium chloride, sodium phosphate, and sucrose (~0.8%) are required as buffer components in the chromatographic operations. A substantial amount of sodium hydroxide (~0.5%) is also used for equipment cleaning.

### 4.3 Cost Analysis and Economic Assessment

Table 6 shows the key economic evaluation results for this project. Economic evaluations were based on the following assumptions: (1) the entire direct fixed capital is depreciated linearly over a period of 10 years assuming a 10% salvage

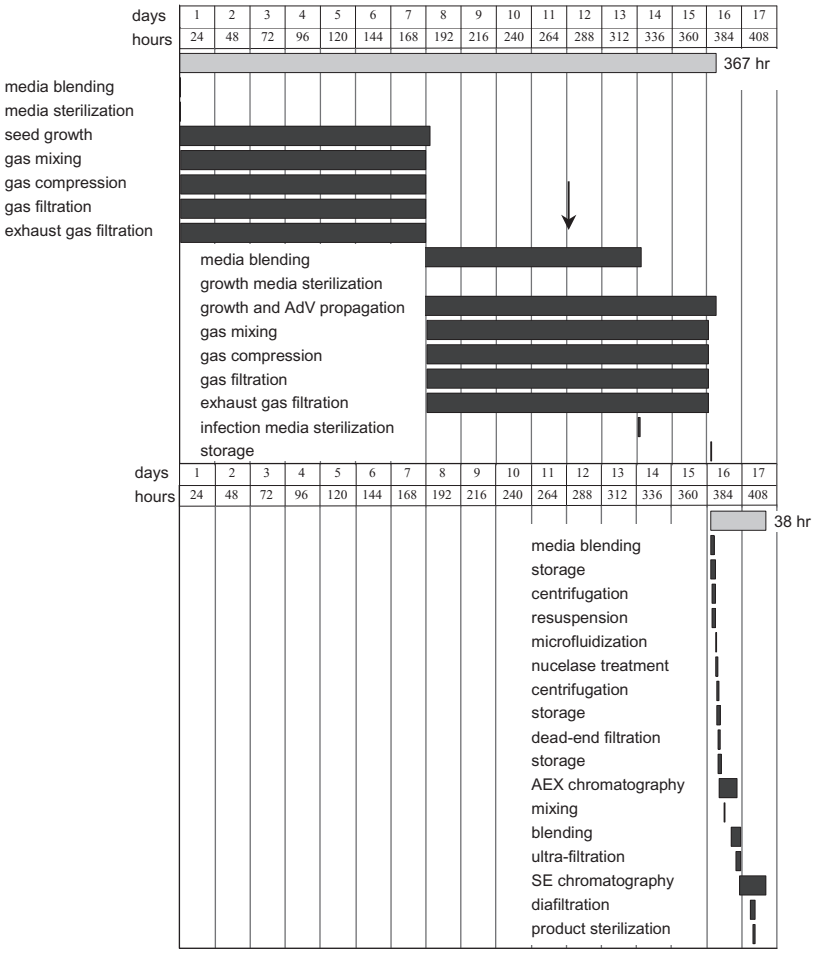


**Figure 10.** Gantt chart for production scheduling of one batch of adenoviral vector. The top and bottom charts show the duration of each operation in the cell culture and downstream processing sections, respectively. The total process time is 405 hours. The arrow indicates the start of a new batch at 264 hours.

value for the entire plant, (2) the project lifetime is 15 years, and (3) $1 \times 10^{18}$ VP of adenoviral product will be produced per year. For a plant of this capacity, the total capital investment is around \$17.8 million. The unit production cost is ~\$7.2/dose ($10^{12}$ VP) or \$24,715/g. This figure is considerably higher when compared with production costs of other biopharmaceuticals such as β-glucuronidase (\$43.0/g [125]), insulin (\$42.2/g [95]), IgG (\$908/g [95]), and plasmid DNA (\$375/g, [100]).

The total equipment purchase cost was estimated to be around \$2.8 million. The most expensive piece of equipment is the bioreactor used for cell culture and virus propagation, priced at \$506,000. The cost of unlisted equipment (including the equipment used in the inoculum preparation section) was assumed to represent 20% of the total equipment cost.

**TABLE 5. Overall Material Balances for the Annual Production of $1 \times 10^{18}$ VP (~0.285 kg) of AdV (kg/year). This Amount Corresponds to 67% of the Total AdV Propagated in the Culture Section.**

| Component | Total Inlet | Total Outlet |
|---|---|---|
| Adenovirus | 0 | 0.425 |
| Carbon dioxide | 1,050 | 1,077 |
| Debris | 0 | 17 |
| DNase | 0.095 | 0.095 |
| Glycerol | 91 | 91 |
| Magnesium chloride | 30 | 30 |
| Nitrogen | 6,444 | 6,444 |
| Nucleic acids | 0 | 0 |
| Nucleotides | 0 | 9 |
| Oxygen | 3,006 | 2,995 |
| Proteins | 0 | 81 |
| SF media | 1,628 | 1,498 |
| Sodium chloride | 3,582 | 3,582 |
| Sodium hydroxide | 5,013 | 5,013 |
| Sodium phosphate | 1,066 | 1,066 |
| Sucrose | 3,171 | 3,171 |
| Tris base | 11 | 11 |
| Process water | 734,446 | 734,446 |
| Water for injection | 210,860 | 210,860 |
| Total | 970,398 | 970,392 |

**TABLE 6. Key Economic Evaluation Results for Adenovirus Production**

| | |
|---|---|
| Direct fixed capital | \$16,593,000 |
| Total capital investment | \$17,820,000 |
| Plant throughput | 0.29 kg ($10^{18}$ VP) AdV/year |
| Operating cost | \$7,149,000/year |
| Unit production cost | \$7.17/dose |

The breakdown of the annual operating cost (AOC) is shown in Figure 11. Facility-dependent cost (44% of the AOC) is the principal operating cost in this process, as is typical for high-value products that are produced in small quantities [95]. Labor-related costs come next, accounting for 34% of the AOC. The annual cost of raw materials is around \$1.0 million, 92% of which are associated with the culture SF medium (priced at \$28/L). The cost of consumables is around \$193,000, representing approximately 3% of the AOC. The chromatographic resins in the AEX and size-exclusion chromatographic steps represent 84% of this value. The cost of utilities (electricity, steam, and cooling agents) is minimal, representing less than 0.1% of the AOC.

Figure 12 shows the return on investment (ROI) and payback time for different product selling prices (\$/dose). The ROI increases around 3.5% for every \$1 increase in the selling price of each AdV dose, whereas the payback time declines for every \$1 increase. If one assumes that a dose of product is sold at \$30, the annual

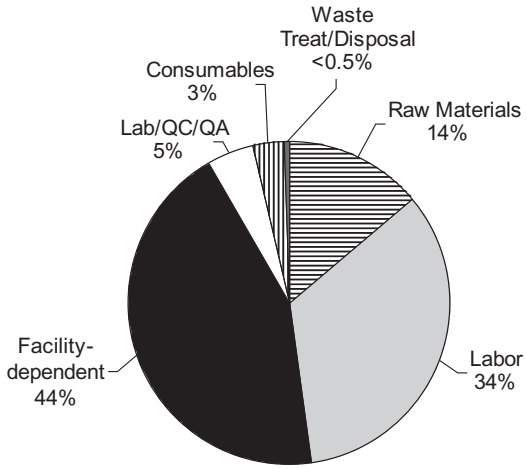


**Figure 11.** Breakdown of the annual operating cost.

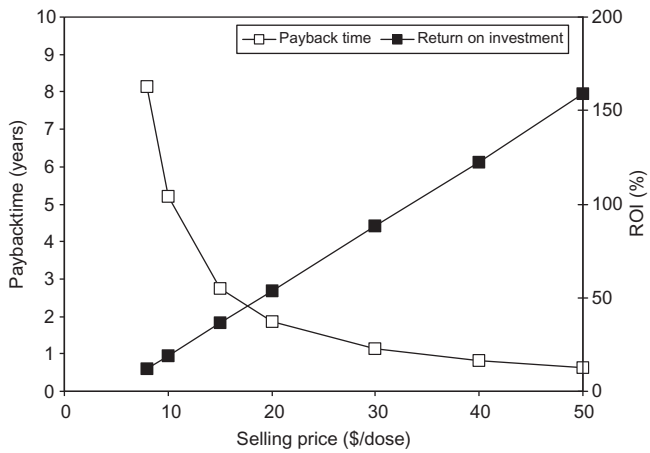


**Figure 12.** Return on investment and payback time at different AdV selling prices.

revenue will amount to $30,662,000 for an annual production of $1 \times 10^{18}$ VP. This corresponds to a gross profit (i.e., revenue-operating cost) of $23,513,000. At this selling price, the payback time is 1.14 years with an ROI of 88%. This economic picture could change slightly by the inclusion of costs that were not accounted for in this case study (filling & packaging section, R&D, process validation).

## 5 CONCLUDING REMARKS AND OUTLOOK

Adenoviral gene therapy has matured to the point where several products should be hitting the market in the wake of Gendicine. Many of the initial drawbacks and barriers (e.g., immunogenicity, lack of specificity, and transient expression of

transgenes) have been tackled with innovative solutions and strategies (targeting, gutless vectors) that have proved valuable in some cases. To provide sufficient material for the ongoing gene therapy clinical trials, production and purification processes must be developed and optimized in parallel to deliver the required selectivity, yield, efficiency, and productivity. The design of these processes under cGMPs should be based on a profound knowledge of the target molecule and associated impurities and on the performance of the available unit operations. Although several well-established processes have been documented in the literature, new developments are likely to surface within the coming years because of increased investments from academia and industry in the area. The analysis presented in this chapter indicates that the production and purification of a therapeutic adenovirus product is economically viable, even with a suboptimized process. Process improvements will certainly reduce costs. As an educated guess, we may estimate selling prices to be in the range \$20–\$50/$10^{12}$ VP.

## REFERENCES


1. Davies J C, Geddes D M, Alton E W (2001). Gene therapy for cystic fibrosis. *J. Gene Med.* 3:409–417.
2. McCormick F (2001). Cancer gene therapy: Fringe or cutting edge? *Nat. Rev. Cancer.* 1:130–141.
3. Strayer D S, Akkina R, Bunnell B A, et al. (2005). Current status of gene therapy strategies to treat HIV/AIDS. *Mol. Ther.* 11:823–842.
4. Donnelly J J, Wahren B, Liu M A (2005). DNA vaccines: Progress and challenges. *J. Immunol.* 175:633–639.
5. Tuteja R (2002). DNA vaccine against malaria: A long way to go. *Crit. Rev. Biochem. Molec. Biol.* 37:29–54.
6. Leitner W W, Thalhamer J (2003). DNA vaccines for non-infectious diseases: New treatments for tumour and allergy. *Expert Opin. Biol. Ther.* 3:627–638.
7. Anderson W (1997). Human gene therapy: The initial concepts. In K Brigham (ed.), *Gene Therapy for Diseases of the Lung*: Marcel Dekker Inc, New York, pp. 3–16.
8. Friedmann T (2001). Stanfield Rogers: Insights into virus vectors and failure of an early gene therapy model. *Molec. Ther.* 4:285–288.
9. Blaese R M, Culver K W, Miller A D, et al. (1995). T lymphocyte-directed gene therapy for ADA-SCID: Initial trial results after 4 years. *Science.* 270:475–480.
10. Zabner J, Couture L A, Gregory R J, et al. (1993). Adenovirus-mediated gene transfer transiently corrects the chloride transport defect in nasal epithelia of patients with cystic fibrosis. *Cell.* 75:207–216.
11. Somia N, Verma I M (2000). Gene therapy: Trials and tribulations. *Nature Rev. Genet.* 1:91–99.
12. Teichler Zallen D (2000). US gene therapy in crisis. *Trends. Genet.* 16:272–275.
13. Cavazzana-Calvo M, Hacein-Bey S, de Saint Basile G, et al. (2000). Gene therapy of human severe combined immunodeficiency (SCID)-X1 disease. *Science.* 288:669–672.
14. Hacein-Bey-Abina S, Von Kalle C, Schmidt M, et al. (2003). LMO2-associated clonal T cell proliferation in two patients after gene therapy for SCID-X1. *Science.* 302:415–419.

15. Peng Z (2005). Current status of gendicine in China: Recombinant human Ad-p53 agent for treatment of cancers. *Human Gene Ther.* 16:1016–1027.
16. Surendran A (2004). News in brief: China approves world's first gene therapy drug. *Nat. Med.* 10:9.
17. Rolland A (2005). Gene medicines: The end of the beginning? *Adv. Drug Del. Rev.* 57:669–673.
18. Wilson J M (2005). Gendicine: The first commercial gene therapy product. *Human Gene Ther.* 16:1014–1015.
19. Thomas C E, Ehrhardt A, Kay M A (2003). Progress and problems with the use of viral vectors for gene therapy. *Nat. Rev. Genet.* 4:346–358.
20. Li S, Huang L (2000). Nonviral gene therapy: Promises and challenges. *Gene Ther.* 7:31–34.
21. Ledley F D (1995). Nonviral gene therapy: The promise of genes as pharmaceutical products. *Human Gene Ther.* 6:1129–1144.
22. Prazeres, D M F, Ferreira, G N M, Monteiro G A, et al. (1999). Large-scale production of pharmaceutical-grade plasmid DNA for gene therapy: Problems and bottlenecks. *Trends Biotechnol.* 17:169–174.
23. Robbins P D, Tahara H, Ghivizzani S C (1998). Viral vectors for gene therapy. *Trends Biotechnol.* 16:35–40.
24. Rowe W P, Huebner R J, Gilmore L K, et al. (1953). Isolation of a cytopathogenic agent from human adenoids undergoing spontaneous degeneration in tissue culture. *Proc. Soc. Exper. Biol. Med.* 84:570–573.
25. Wolff G (2004). Adenovirus for Gene Therapy. In T Blankenstein (ed.), *Gene Therapy: Principles and Applications*: Birkhäuser Verlag, Basel, Switzerland, pp. 29–45.
26. McConnell M J, Imperiale M J (2004). Biology of adenovirus and its use as a vector for gene therapy. *Human Gene Ther.* 15:1022–1033.
27. Russell W C (2000). Update on adenovirus and its vectors. *J. Gen. Virol.* 81:2573–2604.
28. Muruve D A (2004). The innate immune response to adenovirus vectors. *Human Gene Ther.* 15:1157–1166.
29. Schagen F H, Ossevoort M, Toes R E, et al. (2004). Immune responses against adenoviral vectors and their transgene products: A review of strategies for evasion. *Crit. Rev. Oncol. Hematol.* 50:51–70.
30. Nicklin S A, Wu E, Nemerow G R, et al. (2005). The influence of adenovirus fiber structure and function on vector development for gene therapy. *Molec. Ther.* 12:384–393.
31. Fabry C M S, Rosa-Calatrava M, Conway J F, et al. (2005). A quasi-atomic model of human adenovirus type 5 capsid. *EMBO J.* 24:1645–1654.
32. Lusky M (2005). Good manufacturing practice production of adenoviral vectors for clinical trials. *Human Gene Ther.* 16:281–291.
33. Bauerschmitz G J, Barker S D, Hemminki A (2002). Adenoviral gene therapy for cancer: from vectors to targeted and replication competent agents (review). *Internat. J. Oncol.* 21:1161–1174.
34. Seshidhar Reddy P, Ganesh S, Limbach M P, et al. (2003). Development of adenovirus serotype 35 as a gene transfer vector. *Virol.* 311:384–393.
35. Lemckert A A, Sumida S M, Holterman L, et al. (2005). Immunogenicity of heterologous prime-boost regimens involving recombinant adenovirus serotype 11 (Ad11) and Ad35 vaccine vectors in the presence of anti-ad5 immunity. *J. Virol.* 79:9694–9701.

36. Holterman L, Vogels R, van der Vlugt R, et al. (2004). Novel replication-incompetent vector derived from adenovirus type 11 (Ad11) for vaccination and gene therapy: Low seroprevalence and non-cross-reactivity with Ad5. *J. Virol.* 78:13207–13215.
37. Regenmortel M H V V, Fauquet C M, Bishop D H L (eds.) (2001). *Virus Taxonomy: Seventh Report of the International Committee on Taxonomy of Viruses*, Academic Press Inc, London.
38. Volpers C, Kochanek S (2004). Adenoviral vectors for gene transfer and therapy. *J. Gene Med.* 6(suppl 1):S164–171.
39. ICTVdB Management C (2006). 00.001 Adenoviridae. In C Büchen-Osmond (ed.), *ICTVdB—The Universal Virus Database*, version 3, Columbia University, New York.
40. Konz J O, Lee A L, Lewis J A, et al. (2005). Development of a purification process for adenovirus: controlling virus aggregation to improve the clearance of host cell DNA. *Biotechnol. Progress.* 21:466–472.
41. Konz J O, Livingood R C, Bett A J, et al. (2005). Serotype specificity of adenovirus purification using anion-exchange chromatography. *Human Gene Ther.* 16:1–8.
42. van Oostrum J, Burnett R M (1985). Molecular composition of the adenovirus type 2 virion. *J. Virol.* 56:439–448.
43. Vecil G G, Lang F F (2003). Clinical trials of adenoviruses in brain tumors: A review of Ad-p53 and oncolytic adenoviruses. *J. Neurooncol.* 65:237–246.
44. Graham F L, Smiley J, Russell W C, et al. (1977). Characteristics of a human cell line transformed with by DNA from human adenovirus type 5. *J. Gen. Virol.* 36:59–74.
45. Fallaux F J, Bout A, van der Velde I, et al. (1998). New helper cells and matched early region 1-deleted adenovirus vectors prevent generation of replication-competent adenoviruses. *Human Gene Ther.* 9:1909–1917.
46. Guibinga G H, Lochmuller H, Massie B, et al. (1998). Combinatorial blockade of calcineurin and CD28 signaling facilitates primary and secondary therapeutic gene transfer by adenovirus vectors in dystrophic (mdx) mouse muscles. *J. Virol.* 72: 4601–4609.
47. Lusky M, Christ M, Rittner K, et al. (1998). In vitro and in vivo biology of recombinant adenovirus vectors with E1, E1/E2A, or E1/E4 deleted. *J. Virol.* 72:2022–2032.
48. Nadeau I, Sabatie J, Koehl M, et al. (2000). Human 293 cell metabolism in low glutamine-supplied culture: Interpretation of metabolic changes through metabolic flux analysis. *Metabol. Engineer.* 2:277–292.
49. Dedieu J F, Vigne E, Torrent C, et al. (1997). Long-term gene delivery into the livers of immunocompetent mice with E1/E4-defective adenoviruses. *J. Virol.* 71:4626–4637.
50. Engelhardt J F, Ye X, Doranz B, et al. (1994). Ablation of E2A in recombinant adenoviruses improves transgene persistence and decreases inflammatory response in mouse liver. *Proc. Nat. Acad. Sci. U S A.* 91:6196–6200.
51. Gao G P, Yang Y, Wilson J M (1996). Biology of adenovirus vectors with E1 and E4 deletions for liver-directed gene therapy. *J. Virol.* 70:8934–8943.
52. Sandig V, Youil R, Bett A J, et al. (2000). Optimization of the helper-dependent adenovirus system for production and potency in vivo. *Proc. Nat. Acad. Sci. U S A.* 97:1002–1007.
53. Parks R J, Chen L, Anton M, et al. (1996). A helper-dependent adenovirus vector system: Removal of helper virus by Cre-mediated excision of the viral packaging signal. *Proc. Nat. Acad. Sci. U S A.* 93:13565–13570.
54. Parks R J, Graham F L (1997). A helper-dependent system for adenovirus vector production helps define a lower limit for efficient DNA packaging. *J. Virol.* 71:3293–3298.

55. Chen H H, Mack L M, Kelly R, et al. (1997). Persistence in muscle of an adenoviral vector that lacks all viral genes. *Proc. Nat. Acad. Sci. U S A*. 94:1645–1650.
56. Kirn D, Martuza R L, Zwiebel J (2001). Replication-selective virotherapy for cancer: Biological principles, risk management and future directions. *Nat. Med*. 7:781–787.
57. Ghosh D, Barry M A (2005). Selection of muscle-binding peptides from context-specific peptide-presenting phage libraries for adenoviral vector targeting. *J. Virol*. 79:13667–13672.
58. Volpers C, Thirion C, Biermann V, et al. (2003). Antibody-mediated targeting of an adenovirus vector modified to contain a synthetic immunoglobulin g-binding domain in the capsid. *J. Virol*. 77:2093–2104.
59. Stevenson S C, Rollence M, Marshall-Neff J, et al. (1997). Selective targeting of human cells by a chimeric adenovirus vector containing a modified fiber protein. *J. Virol*. 71:4782–4790.
60. Campos S K, Parrott M B, Barry M A (2004). Avidin-based targeting and purification of a protein IX-modified, metabolically biotinylated adenoviral vector. *Molec. Ther*. 9:942–954.
61. Eto Y, Gao J Q, Sekiguchi F, et al. (2005). PEGylated adenovirus vectors containing RGD peptides on the tip of PEG show high transduction efficiency and antibody evasion ability. *J. Gene Med*. 7:604–612.
62. Henning P, Andersson K M, Frykholm K, et al. (2005). Tumor cell targeted gene delivery by adenovirus 5 vectors carrying knobless fibers with antibody-binding domains. *Gene Ther*. 12:211–224.
63. Nicklin S A, Von Seggern D J, Work L M, et al. (2001). Ablating adenovirus type 5 fiber-CAR binding and HI loop insertion of the SIGYPLP peptide generate an endothelial cell-selective adenovirus. *Molec. Ther*. 4:534–542.
64. van Beusechem V W, van Rijswijk A L, van Es H H, et al. (2000). Recombinant adenovirus vectors with knobless fibers for targeted gene transfer. *Gene Ther*. 7:1940–1946.
65. Yamamoto M, Alemany R, Adachi Y, et al. (2001). Characterization of the cyclooxygenase-2 promoter in an adenoviral vector and its application for the mitigation of toxicity in suicide gene therapy of gastrointestinal cancers. *Molec. Ther*. 3:385–394.
66. St George J A (2003). Gene therapy progress and prospects: Adenoviral vectors. *Gene Ther*. 10:1135–1141.
67. Matsuse T, Teramoto S (2000). Progress in adenovirus-mediated gene therapy for cystic fibrosis lung disease. *Curr. Therapeu. Res*. 61:422–434.
68. Harvey B G, Leopold P L, Hackett N R, et al. (1999). Airway epithelial CFTR mRNA expression in cystic fibrosis patients after repetitive administration of a recombinant adenovirus. *J. Clin. Invest*. 104:1245–1255.
69. van Deutekom J C, van Ommen G J (2003). Advances in Duchenne muscular dystrophy gene therapy. *Nat. Rev. Genet*. 4:774–783.
70. Chamberlain J S (2002). Gene therapy of muscular dystrophy. *Human Molec. Genet*. 11:2355–2362.
71. Lang F F, Bruner J M, Fuller G N, et al. (2003). Phase I trial of adenovirus-mediated p53 gene therapy for recurrent glioma: biological and clinical results. *J. Clin. Oncol*. 21:2508–2518.
72. Moolten F L (1986). Tumor chemosensitivity conferred by inserted herpes thymidine kinase genes: Paradigm for a prospective cancer control strategy. *Cancer Res*. 46:5276–5281.
73. Tsuchiyama T, Kaneko S, Nakamoto Y, et al. (2003). Enhanced antitumor effects of a bicistronic adenovirus vector expressing both herpes simplex virus thymidine kinase

and monocyte chemoattractant protein-1 against hepatocellular carcinoma. *Cancer Gene Ther.* 10:260–269.

74. Lanuti M, Gao G P, Force S D, et al. (1999). Evaluation of an E1E4-deleted adenovirus expressing the herpes simplex thymidine kinase suicide gene in cancer gene therapy. *Human Gene Ther.* 10:463–475.
75. Mizuguchi H, Hayakawa T (2002). Enhanced antitumor effect and reduced vector dissemination with fiber-modified adenovirus vectors expressing herpes simplex virus thymidine kinase. *Cancer Gene Ther.* 9:236–242.
76. Vile R G (2004). Thymidne kinases. In T Blankenstein (ed.), *Gene Therapy: Principles and Applications*. Birkhäuser Verlag, Basel, Switzerland, pp. 247–266.
77. Ueno M, Koyama F, Yamada Y, et al. (2001). Tumor-specific chemo-radio-gene therapy for colorectal cancer cells using adenovirus vector expressing the cytosine deaminase gene. *Anticancer Res.* 21:2601–2608.
78. Anello R, Cohen S, Atkinson G, et al. (2000). Adenovirus mediated cytosine deaminase gene transduction and 5-fluorocytosine therapy sensitizes mouse prostate cancer cells to irradiation. *J. Urol.* 164:2173–2177.
79. Khuri F R, Nemunaitis J, Ganly I, et al. (2000). A controlled trial of intratumoral ONYX-015, a selectively-replicating adenovirus, in combination with cisplatin and 5-fluorouracil in patients with recurrent head and neck cancer. *Nat. Med.* 6:879–885.
80. Kamen A, Henry O (2004). Development and optimization of an adenovirus production process. *J. Gene Med.* 6(suppl 1):S184–192.
81. Kanarek A D (2001). A guide to Good Manufacturing Practice. In S C DiClemente (ed.), *D&MD How-to Guides*: D&MD, Westborough, MA.
82. Doblhoff-Dier O, Bliem R (1999). Quality control and assurance from the development to the production of biopharmaceuticals. *Trends Biotechnol.* 17:266–270.
83. Boyd J E (1999). Facilities for large-scale production of vectors under GMP conditions. In A Meager (ed.), *Gene Therapy Technologies, Applications and Regulations—from Laboratory to Clinic*: John Wiley & Sons LTD, Chichester, UK, pp. 383–400.
84. Sofer G, Hagel L (1997). Handbook of process chromatography: A guide to optimization, scale-up and validation, 1st ed. In G Sofer, L Hagel (eds.), Academic Press; San Diego, CA.
85. Akers J, McEntire J, Sofer G (1994). Biotechnology product validation, part 2: A logical plan. *Pharmaceut. Technol. Euro.* 6:230–234.
86. Tolbert W R, Merchant B, Taylor J A, et al. (1996). Designing an initial gene therapy manufacturing facility. *BioPharm Internat.* Nov.: 32–40.
87. Roitsch C, Achstetter T, Benchaibi M, et al. (2001). Characterization and quality control of recombinant adenovirus vectors for gene therapy. *J. Chromatog. B.* 752:263–280.
88. Meager A, Vocke T, Zimmermann, G. (1999). The development of the regulatory process in Europe for biological medicines: How it affects gene therapy products. In A Meager (ed.), *Gene Therapy Technologies, Applications and Regulations—from Laboratory to Clinic*. John Wiley & Sons LTD, Chichester, UK, pp. 319–346.
89. Rexroad J, Wiethoff C M, Green A P, et al. (2003). Structural stability of adenovirus type 5. *J. Pharmaceut. Sci.* 92:665–678.
90. McGann L E, Yang H Y, Walterson M (1988). Manifestations of cell damage after freezing and thawing. *Cryobiol.* 25:178–185.
91. CBER (1996). Addendum to the points to consider in Human somatic cell and gene therapy (1991). US FDA, Center for Biologics Evaluation and Research, Rockville, MD.

92. EMEA (1999). Note for guidance on the quality, preclinical and clinical aspects of gene transfer medicinal products. The European Agency for the Evaluation of Medicinal Products, London.
93. Peng Z (2004). Points to consider for human gene therapy and product quality control State Food and Drug Administration of China. *BioPharm Internat.* May: 6–9.
94. Prazeres, D M F, Ferreira G N M (2004). Design of flowsheets for the recovery and purification of plasmids for gene therapy and DNA vaccination. *Chem. Engin. Proc.* 43:615–630.
95. Petrides D (2003). Bioprocess Design. In R G Harrison, P W Todd, S R Rudge, et al. (eds.), *Bioseparations Science and Engineering*. Oxford University Press, Oxford, UK, pp. 319–372.
96. Gregoriades N, Luzardo M, Lucquet B, et al. (2003). Heat inactivation of mammalian cell cultures for biowaste kill system design. *Biotechnol. Prog.* 19:14–20.
97. Jannat R, Hsu D, Maheshwari G (2005). Inactivation of adenovirus type 5 by caustics. *Biotechnol. Prog.* 21:446–450.
98. Petrides D P, Koulouris A, Lagonikos P T (2002). The role of process simulation in pharmaceutical processes development and product commercialization. *Pharmaceut. Engin.* 22:1–8.
99. Steffens M A, Fraga E S, Bogle I D (2000). Synthesis of bioprocesses using physical properties data. *Biotechnol. Bioengin.* 68:218–230.
100. Freitas S S, Santos J A L, Prazeres D M F (2006) Plasmid DNA production. In E Heinzle, A Biwer, C Cooney (eds.), *Development of Sustainable Bioprocesses: Modeling and Assessment*. Wiley, New York, pp. 271–285.
101. Ugai H, Yamasaki T, Hirose M, et al. (2005). Purification of infectious adenovirus in two hours by ultra-centrifugation and tangential flow filtration. *Biochem. Biophys. Res. Comm.* 331:1053–1060.
102. Carrion M E, Menger M, Kovesdi I (2003). Efficient purification of adenovirus. US Patent appl. 6586226.
103. Tang J C-T, Vellekamp G, Bondoc L L (2001). Methods for purifiying viruses. US Patent appl. 20010036657.
104. Zhang S, Thwin C, Wu Z, Cho T (2001). Method for the production and purification of adenoviral vectors. US Patent appl. 6194191 B1.
105. Huyghe B G, Liu X, Sutjipto S, et al. (1995). Purification of a type 5 recombinant adenovirus encoding human p53 by column chromatography. *Human Gene Ther.* 6:1403–1416.
106. Green A P, Huang J J, Scott M O, et al. (2002). A new scalable method for the purification of recombinant adenovirus vectors. *Human Gene Ther.* 13:1921–1934.
107. Goerke A R, To B C, Lee A L, et al. (2005). Development of a novel adenovirus purification process utilizing selective precipitation of cellular DNA. *Biotechnol. Bioengin.* 91:12–21.
108. Knight P (1989). Downstream processing. *Bio/technol.* 7:777–782.
109. Wheelwright S M (1987). Designing downstream processes for large-scale protein purification. *Bio/technol.* 5:789–793.
110. Nadeau I, Kamen A (2003). Production of adenovirus vector for gene therapy. *Biotechnol. Adv.* 20:475–489.
111. Cote J, Garnier A, Massie B, et al. (1998). Serum-free production of recombinant proteins and adenoviral vectors by 293SF-3F6 cells. *Biotechnol. Bioengin.* 59:567–575.

112. Nadeau I, Gilbert P A, Jacob D, et al. (2002). Low-protein medium affects the 293SF central metabolism during growth and infection with adenovirus. *Biotechnol. Bioengin.* 77:91–104.
113. Ferreira T B, Ferreira A L, Carrondo, M J T, et al. (2005). Effect of refeed strategies and non-ammoniagenic medium on adenovirus production at high cell densities. *J. Biotechnol.* 119:272–280.
114. Iyer P, Ostrove J M, Vacante D (1999). Comparison of manufacturing techniques for adenovirus poduction. *Cytotechnol.* 30:169–172.
115. Braas G, Searle P F, Slater N K, et al. (1996). Strategies for the isolation and purification of retroviral vectors for gene therapy. *Biosepar.* 6:211–228.
116. Wu S C, Huang G Y, Liu J H (2002). Production of retrovirus and adenovirus vectors for gene therapy: A comparative study using microcarrier and stationary cell culture. *Biotechnol. Prog.* 18:617–622.
117. Okada T, Nomoto T, Yoshioka T, et al. (2005). Large-scale production of recombinant viruses by use of a large culture vessel with active gassing. *Human Gene Ther.* 16:1212–1218.
118. Xie L, Metallo C, Warren J, et al. (2003). Large-scale propagation of a replication-defective adenovirus vector in stirred-tank bioreactor PER.C6 cell culture under sparging conditions. *Biotechnol. Bioengin.* 83:45–52.
119. Xie L, Pilbrough W, Metallo C, et al. (2002). Serum-free suspension cultivation of PER.C6(R) cells and recombinant adenovirus production under different pH conditions. *Biotechnol. Bioengin.* 80:569–579.
120. Tsao, Y-S, Condon R, Schaefer E, et al. (2001). Development and improvement of a serum-free suspension process for the production of recombinant adenoviral vectors using HEK293 cells. *Cytotechnol.* 37:189–198.
121. Cortin V, Thibault J, Jacob D, et al. (2004). High-titer adenovirus vector production in 293S cell perfusion culture. *Biotechnol. Prog.* 20:858–863.
122. Parrott M B, Adams K E, Mercier G T, et al. (2003). Metabolically biotinylated adenovirus for cell targeting, ligand screening, and vector purification. *Molec. Ther.* 8:688–700.
123. Evans R K, Nawrocki D K, Isopi L A, et al. (2004). Development of stable liquid formulations for adenovirus-based vaccines. *J. Pharmaceu. Sci.* 93:2458–2475.
124. Azari M, Boose J A, Burhop K E, et al. (2000). Evaluation and validation of virus removal by ultrafiltration during the production of diaspirin crosslinked haemoglobin (DCLHb). *Biologicals.* 28:81–94.
125. Evangelista R L, Kusnadi A R, Howard J A, et al. (1998). Process and economic evaluation of the extraction and purification of recombinant beta-glucuronidase from transgenic corn. *Biotechnol. Prog.* 14:607–614.

# ASSESSING GENE THERAPY BY MOLECULAR IMAGING

Pascal Delépine and Claude Férec
*EFS Bretagne—Site de Brest, INSERM U613, Brest, France*

**Chapter Contents**

1 Technological Advances Applied to Imaging of Gene Transfer
2 *In Vitro* Gene Imaging Studies
  2.1 Characterization of Gene Transfer Agents
  2.2 Gene Transfer Agents Pathway: Localisation and Persistence of the Transgene at its Target
  2.3 Evaluation of Gene Transfer Efficiency
3 *In Vivo* Imaging Experiments
  3.1 Imaging Biodistribution of Gene Transfer Agents
  3.2 Imaging Gene Transfer Efficacy
  3.3 Advantages and Drawbacks of Imaging Approaches
4 Conclusion
  References

## 1 TECHNOLOGICAL ADVANCES APPLIED TO IMAGING OF GENE TRANSFER

At the dawn of the twenty-first century, cancer and cardiovascular disease are the first causes of mortality and the focus of considerable attention and mobilization of financial and human resources. Increased understanding of pathogenesis and its underlying molecular processes has led to proven or investigational therapies, such as gene therapy. The concept of gene therapy (and gene transfer) first appeared in the 1960s when the structure of DNA was defined, and much effort has since been devoted to developing suitable strategies. Although gene therapy was initially considered primarily in terms of treatment of genetic disease, the concept has evolved,

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

and now the two main fields of gene therapy are cancer and cardiovascular disease. Gene therapy consists of overexpressing a functional gene, replacing a mutated gene, or expressing an exogenous gene with the aim of correcting a malfunction, inducing an immune response, or stimulating angiogenesis. Currently, stem cell research is generating new hope in the field of cell therapy, with which gene transfer is increasingly associated.

The first challenge was to transfer DNA into cells, and *in vitro* and *in vivo* gene transfer techniques were developed: mechanical (gene gun), physical (calcium phosphate precipitation), chemical (synthetic gene transfer agents), and biological (viruses). Some of these techniques are commonly used, at least *in vitro*, and have been evaluated in clinical trials, whereas others are currently under investigation. Clinical testing is contingent on technical quality and reproducibility and finding answers to a variety of questions:

- Does the vector or gene transfer procedure work as expected? Optimally?
- Is the preparation optimal?
- What about the route of transfer?
- What is the fate of the nucleic acid and of the vector (if any)?
- When do the vector and nucleic acid dissociate?
- Does the nucleic acid reach its target (and only its target)?
- Does the procedure have any deleterious effect?
- How long is it before expression occurs?
- How long does the transgene persist in the targeted cell or organ?
- How long is transgene expression effective?

Before addressing these questions, it is first essential to ensure that the nucleic acid is transferred into the cell (or organ) and is active. Transfer can be observed *in vitro* by means of microscopy or electron microscopy. Using nucleic acid probes tagged with gold particles, it is possible to observe the presence of a DNA sequence [1]. Direct labeling is also possible using radionuclides, such as phosphorus 32, and is mostly applied *in vivo* in animal models because of its low resolution. Direct labeling is sometimes impractical, so indirect labeling has been developed. Staining techniques have been improved in parallel.

Although imaging approaches can be used to observe, understand, and document all phenomena involved in gene transfer, many are destructive or "invasive" and often require sacrifice of the animal, as in all histopathological, immunohistological, and immunofluorescent techniques. These techniques yield information that is valuable but that only offers a snapshot of a particular moment in time: Researchers are usually interested in a longer-term image of induced phenomena. The use of several groups of the same animal species to obtain this longer-term image is not really satisfying, and so noninvasive techniques tend to replace invasive or destructive ones. These noninvasive techniques can track processes in a single group of animals and have the ethical advantage that they reduce the number of animals used. Also, invasive approaches can be applied to *in vitro* and *in vivo* research experiments, but not to clinical studies, which may explain why we lack information on the fate of gene transfer agents and transgenes in humans. This technological drawback may be partially responsible for the slowness with which gene therapy is being transferred from the lab bench to the bedside.

The most studied imaging methods are positron emission tomography (PET), single-photon emission (computed) tomography (SPE(C)T), bioluminescence, magnetic resonance imaging (MRI), and gamma imaging. Whereas TEP, SPECT, MRI, and gamma imaging are easily performed on animals, bioluminescence is largely limited because of tissue absorption, with a 90% decrease in signal for each centimeter of tissue. Given this limitation, it is clear that better reporter proteins and reporter genes are needed. Optical imaging is usually performed with proteins that emit green light [luciferase or green fluorescent protein (GFP)], and a shift toward red or near-infrared emission is considered to restrict this absorption and allow the signal to leave the animal's body. Bioluminescence usually yields just planar information, but never devices give 3D information. One technique combines laser scanning, which defines the animal's topography, and filters to select the emitted photons (different wavelengths). Software is used to reconstruct a 3D image of the animal. In another, tomographic, approach, a charge-coupled device (CCD) camera rotates around the animal giving whole-body information on photon emission.

Improvments in molecular biology also exists in the MRI field. The ferritin reporter gene was developed for use with MRI. Overexpression of the transgene in the modified cells (various methods of gene transfer are possible) leads to intracellular ferritin excess and iron entrapment. Ferritin associates with iron, which accumulates (aggregates) in the cell as nanomagnets that can be imaged by MRI [2]. Initially, the very common reporter gene beta-galactosidase was used for MRI. The paramagnetic ion is "blocked" using a galactopyrannoside screen and so is not accessible to water. When the reporter gene beta-galactosidase is expressed, its enzymatic activity alters the galactopyrannoside structure and frees the paramagnetic ion, which can then be imaged by MRI [3]. Another approach for MRI uses a reporter gene that encodes a membrane transferrin receptor. The transgene receptor is modified to be resistant to downregulation, so the receptor is largely overexpressed in comparison with untransfected cells. Internalization of superparamagnetic structures (such as MION) conjugated to transferrin is increased in modified cells [4].

As noted above, gene transfer is increasingly associated with cell therapy. Imaging can provide information on the fate of the reinfused modified cells and also on the expression of the transgene or the corrected gene. By loading the cells of interest with contrast agents, it is possible to use MRI to track their movements through the animal's body [5]. The exitation devoted to quantum dots also spreads to cell therapy. Indeed, recently the fate of cells reinjected to mice was studied after labeling those cells by means of self-illuminating quantum dots. It was then possible to image the cells migration by fluorescence imaging *in vivo* [6].

## 2 *IN VITRO* IMAGING STUDIES

### 2.1 Characterization of Gene Transfer Agents

***Structure.*** Unless DNA is transferred by a physical method, it is necessary to generate a structure able to carry the nucleic acid. When virus capsid or synthetic molecules are involved, they must be characterized physically to allow rational studies on how they ensure vectorization. The structures of gene transfer agents have to be defined to understand how they interact or associate with other

components. Visualization of the structures alone is currently possible only by microscopy. Classically, transmission electron microscopy (TEM) is performed after negative staining with uranyl acetate [7] (Figure 1). Vector characterization is frequently combined with characterization of the "nucleic acid-vector" association. TEM is also used, for example, to confirm that the DNA has intertwined with the polymer assembly [8]. Atomic force microscopy is a relatively new tool that enables the visualization of nanostructures. It will probably give much information on the

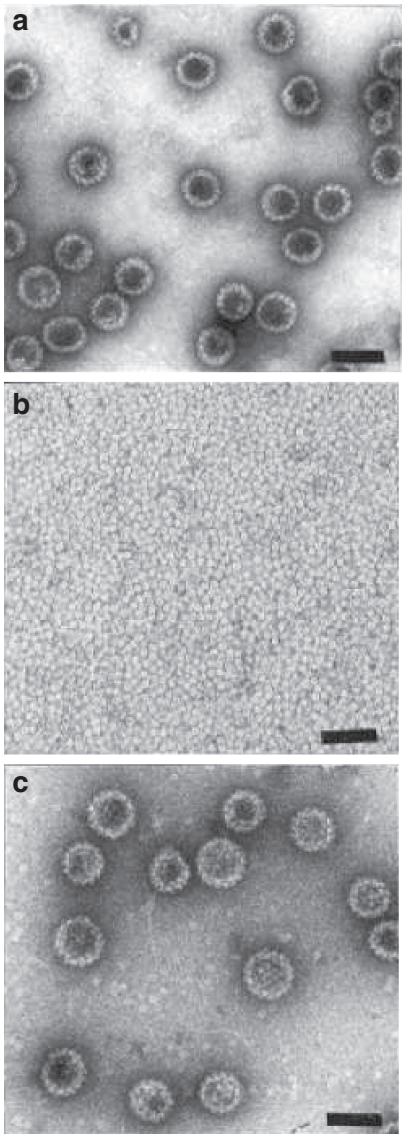


**Figure 1.** Electron microscopy to image gene transfer structures. In this example, the authors have imaged the gene transfer structure they developed (human papillomavirus-like particles) after uranyl acetate negative staining. Reprinted from Ref. 7, with permission from Oxford Journals.

structures obtained according to the preparation way and also on the association of the nucleic acid with its vector.

***Nucleic Acid—Vector Interaction and Kinetics.*** One way to confirm that the vector interacts with or encapsulates the nucleic acid is to visualize the interaction. Scanning force microscopy has shown that polymers and cationic lipids interact with DNA in a similar way [9] (Figure 2). TEM has been used to observed this kind of interaction between DNA and synthetic gene transfer agents [8]. Although fluorescence resonance energy transfer (FRET) data are usually represented graphically [10], it is theoretically possible to obtain images of the fusion events using fluorescence microscopy, which shows how lipids used to prepare lipoplexes interact and mix and also gives information on the capacity of a structure to interact with a membrane model representing the cells.

## 2.2 Gene Transfer Agent Pathways: Localization and Persistence of the Transgene at its Target

To improve gene transfer agents, we need to know which hurdles they will have to overcome. Early studies of intracellular trafficking of the vectors were performed by means of electron microscopy.

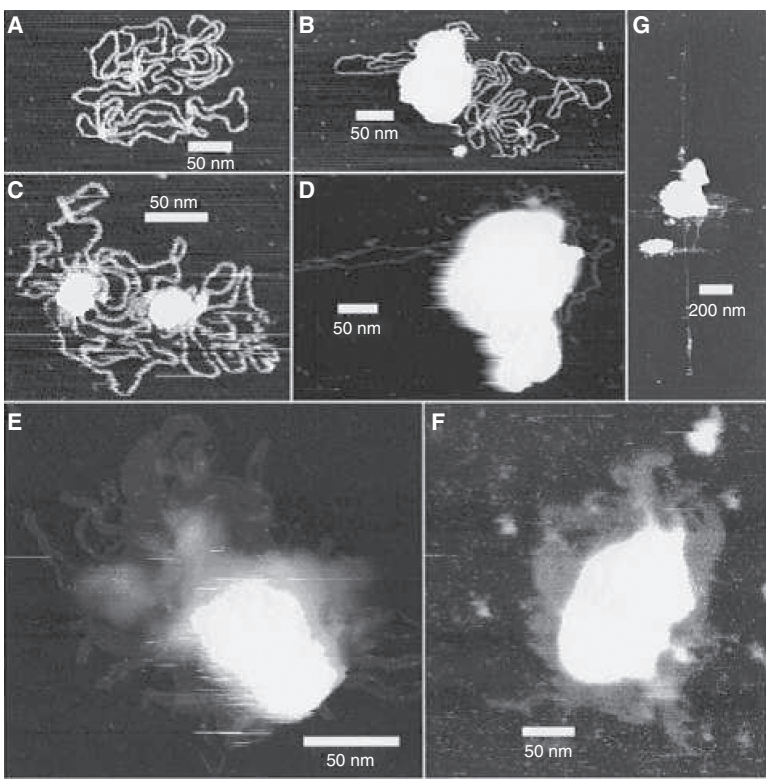


**Figure 2.** Scanning force microscopy to image the association between the nucleic acid and the vector. The authors have imaged DNA and its condensation by synthetic gene transfer agents. Reprinted from Ref. 9, with permission from Oxford Journals.

***Transmission Electron Microscopy.*** To observe polyplexes, thin cell sections are negatively stained using uranyl acetate and observed by TEM [8], which can track gene transfer agents. Existing structures can be observed, but when the polymer dissociates, nothing is visible any more. To observe the nucleic acid either free or associated with its vectors, a labeling using gold-labeled probes can be used. The nucleic acid sequence is recognized by a probe that has been tagged with gold particles. The intracellular progression of the nucleic acid can thus be visualized [1] (Figure 3).

***Fluorescence Microscopy.*** This microscopic approach is used to study the trafficking of either the vector or the nucleic acid. By a fluorescent labeling of the structure of interest (with fluorochrome such as FITC or TRITC), it is possible to visualize the internalization and trafficking processes using fluorescence microscopy or confocal microscopy. Using this approach, Midoux's group studied the internalization of polyplexes in HepG2 cells [11]. An interest of this last technique is that both the vector and the nucleic acid can be labeled. The complex evolution (and its disruption) can be observed.

***Magnetic Resonance Imaging.*** MRI can be used to track complexes during transfection even at the cellular level [12]. MRI microscopy achieves a resolution of about 1 μm. It is based on the co-complexation of the DNA and the contrast agents with a cationic polymer (theoretically, it is also possible with all cationic gene transfer agents). By adding selective molecules, it is possible to target specific cells. A gadolinium derivative is used to follow the fate of the complex (DNA

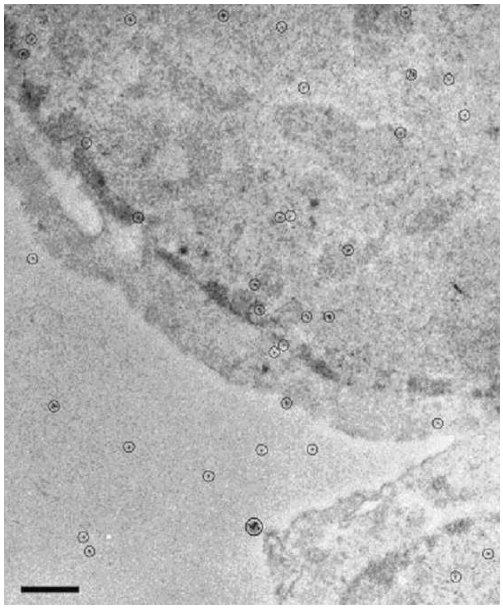

**Figure 3.** Study of the intracellular trafficking of a plasmid complexed with a synthetic phosphonolipid using electonic microscopy. Cell lines were transfected with lipoplexes. Then, DNA probes specific to the plasmid used and tagged with gold particles was incubated with the cells pellet sections. The DNA migration was also observed. Reprinted from Ref. 1, with permission from Elsevier.

polymer) within the cell. This real-time visualization reveals only the contrast agent aggregates. No information is given on the organization of the visualized structure, and it is not possible to say whether or when the complexes have been disrupted.

### 2.3 Evaluation of Gene Transfer Efficiency

Given how hard it is to demonstrate the functional activity of genes, a nucleic acid delivery system is usually developed using reporter genes encoding proteins that are easily detected through enzymatic reaction or their intrinsic fluorescence.

***Fluorescence Microscopy.*** To ensure that a delivery system or a vector efficiently transfers a nucleic acid into a cell, the easiest approach is to use a reporter gene encoding a protein that can be visualized directly. Reporter genes such as GFP or its new derivatives RFP (red) or BFP (blue) are satisfying because gene expression can be easily visualized by fluorescence microscopy. Moreover, by using a fusion protein, this approach not only reveals expres-sion but also defines cellular localization, particularly when using confocal microscopy. Indirect visualization can also be performed using a nucleic acid that encodes a tagged protein. By incubating the cells with a fluorescent ligand that specifically forms a covalent bond with the tagged protein, the transfected cells appears fluorescent.

***Bioluminescence.*** The bioluminescence approach is not frequently used for *in vitro* studies because of the scarcity of the equipment required. The classic reporter gene luciferase can be used to visualize changes in cells, which is of particular interest when working with a nucleic acid that encodes both luciferase and its substrate, luciferin.

***Magnetic Resonance Imaging.*** When used with MRI, classic contrast agents can track complexes but give no information on gene expression, which is why new reporter genes were developed to modify and hence activate contrast agents, including the gene-encoding ferritin, transfer agents, and enzymes that enable "smart" MRI contrast agent activation [13]. These smart contrast agents are prepared in a weak relaxivity state. When they enter a cell genetically modified by gene transfer, they are transformed into a strong relaxivity state, which is the case of EGad, a gadolinium complex. When Egad interacts with β-galactosidase, a widely used reporter protein, the Egad chelate is disrupted and the relaxation properties of gadolinium can be imaged.

All these MRI approaches applied to cells and to animal models are potentially applicable to humans. MRI is particularly interesting because it gives real-time information potentially at high resolution.

## 3 *IN VIVO* IMAGING EXPERIMENTS

### 3.1 Imaging Biodistribution of Gene Transfer Agents

It is essential to study the biodistribution of components involved in gene transfer (virus, synthetic vector, or nucleic acid) to identify all *in vivo* barriers. This study indicates what proportion of the administered dose actually reaches the target, as well as any unexpected localizations (according to the quantity at each site) or blocking sites, which is of particular interest because it allows prediction of the

dose that will produce a significant therapeutic effect. Biodistribution studies indicate the time when the nucleic acid is released from its vector. Using markers specific to each component (by labeling or conjugating), it is possible to follow both of them separately and also to observe their separation and fate (i.e., their elimination route and metabolism).

The most frequently used approach is direct, in which a radionuclide (usually a β-emitter) is used to label the different components. Indirect studies are also possible, such as the use of markers that are supposed to mimic the behavior of the studied molecules. Invasive imaging methods are used less frequently. In some, the compounds involved are radiolabeled and administered before sacrifice of the animal. The animal's body is thinly sectioned using a cryomicrotome and visualization is done either by classic autoradiography or phosphorus imaging. Gene vectors may also be labeled with fluorescent markers and visualized on microsections. As a result of whole-body dilution, this approach is essentially used to pinpoint a component in the cells of a defined organ, after localization by scintigraphy or autoradiography. Although invasive approaches have been widely used, live imaging is now largely preferred, mainly using the methods described below.

***Radioisotopic Imaging.*** Radioisotopic imaging (using γ-emitters) efficiently tracks the distribution of viral [14] and nonviral vectors [15, 16]. In these two cases, the adenovirus and the GLB43 lipid vector (Figure 4) were labeled with $^{99m}$Tc pertechnetate using a stannous tin procedure. Although most available devices give planar imaging, this approach is quantitative. Scintigraphic imaging can be applied to all kinds of animals and also to humans (depending on the radioactive dose). The method used to label the vector is not always suitable for the nucleic acid, as nucleic acid chelation may inhibit its expression. It is, therefore, necessary to choose between information on expression and on location.

***Magnetic Resonance Imaging.*** MRI is a multivalent approach that also predicts gene transfer efficiency (i.e., whether the transgene actually reached the targeted site). The transgene pathway can be mimicked using a contrast agent, which is administered as the nucleic acid would have been (i.e., associated with a gene vector or directly injected into the targeted tissue). Thus, a gene transfer approach can also be validated by MRI. Moreover, as MRI visualizes tissues and organs, any deleterious effects after gene transfer can be monitored [17].

### 3.2 Imaging Gene Transfer Efficacy

The sites of gene transfer clearly must be defined whatever the animal model used or clinical application. For most applications, a specific target is considered. Different administration routes will result in particular expression levels and accuracy of targeting. Direct visualization on thin sections is a common approach in which immunofluorescent probes of the expressed protein are used to study transgene expression (Figure 5). For example, Bartoli et al. visualized the transgenic protein (α-sarcoglycan) by immunohistochemistry using a secondary antibody conjugated to the fluorochrome Alexa488 in animal models of muscular dystrophies [18].

Animal care and ethical issues are of increasing importance even in research, and so when suitable technology is available, it is preferable to use noninvasive methods

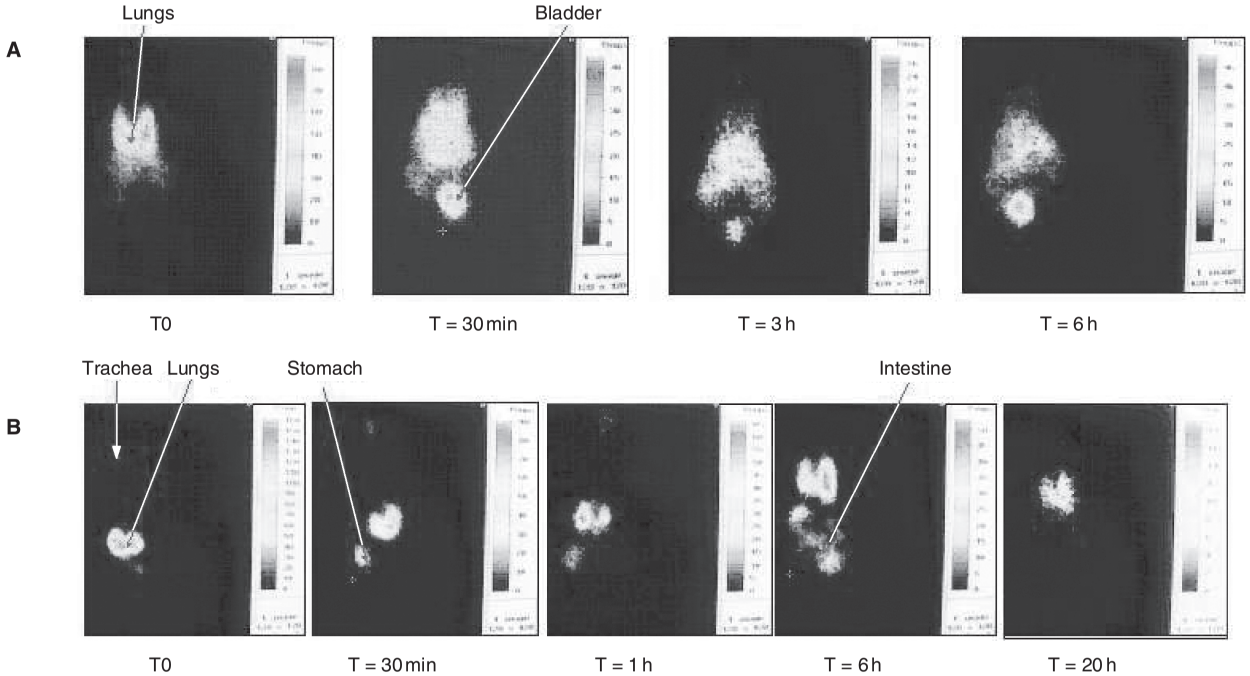


**Figure 4.** Scintigraphic biodistribution study. A plasmid encoding the reporter protein luciferase was labeled with $^{99m}$Tc. The lipoplexes, resulting from the association of the phosphonolipid GLB43 with the radiolabeled plasmid, was administrated to mice either by intravenous injection or endotracheal nebulization. The animals were imaged using a gamma camera dedicated to small animals. Reprinted from Ref. 15, with permission from Nature Publication Group. (This figure is available in full color at ftp://ftp.wiley.com/public/sci_ tech_med/pharmaceutical_biotech/.)

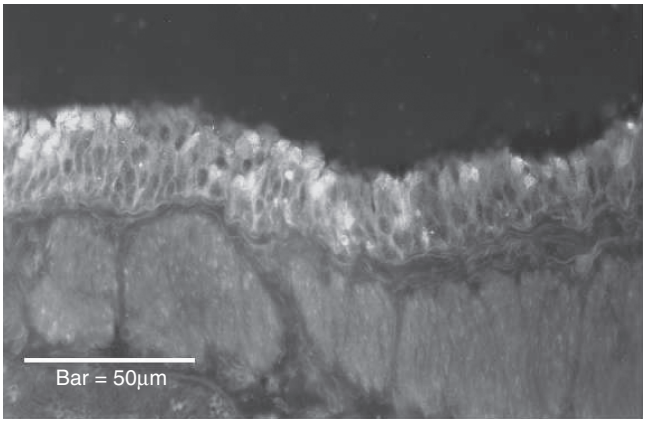


**Figure 5.** Immunohistofluorescence imaging to visualize the cells that express the transgene. After mice transfection with a reporter gene (luciferase) associated with the synthetic amphiphiphilic vector (KLN47), the transgene expression was visualized using a primary antibody directed against the luciferase protein. The secondary antibodies were FITC labeled. The expressing cells are thus identified in the targeted tissue (from Delépine et al., unpublished data). (This figure is available in full color at ftp://ftp.wiley.com/public/sci_tech_med/pharmaceutical_biotech/.)

**Figure 6.** *In vivo* bioluminescence imaging. The luciferase transgene expression was observed using a CCD camera 24 hours after an intravenous injection of a lipoplexe (KLN20 phosphonolipid + pCMVLuc) [19]. (This figure is available in full color at ftp://ftp.wiley.com/public/sci_tech_med/pharmaceutical_biotech/.)

to study gene transfer. Tomography is useful in this regard as it gives information on gene expression in terms of time lag, duration, magnitude, and location.

***Bioluminescence [19]*** *(Figure 6)*. Since a pioneering 1998 study of bioluminescence in living mammals [20], bioluminescence technology has advanced greatly and is

probably now the most accessible and common approach to *in vivo* imaging. It is, however, limited by the data acquisition, as the most frequently used devices equipped with a CCD camera do not give tomographic information (because of the cost of the tomographic apparatus) and the main part of the emitted light is absorbed by the tissues. Moreover, this absorption depends on tissue type and depth, so it is difficult to extrapolate the results from one animal to another. Another limitation is the diffusion of the substrate. In most cases, the luciferase substrate is injected intraperitoneally. If the transfected site is less accessible because of disease or another cause, the substrate concentration is reduced locally, as is the luminescent signal. This bias can be prevented by using a genetic construction that simultaneously expresses both luciferase and its substrate.

New tomographic devices based on bioluminescence are able to generate a 3D reconstruction, using software to calculate photon absorption for the various tissues. Two advantages accrue from this ability: First, this approach is quantitative and, second, it more precisely localizes the source(s) of luminescence induced by the gene transfer procedure. Although bioluminescence techniques are evolving, they are still only applicable to small animals. The hurdles inherent to this imaging approach and to the use of such reporter genes prevent its application to humans.

***Fluorescence.*** Most bioluminescence imaging devices are also equipped for fluorescence imaging. Fluorescent markers are less sensitive than bioluminescent ones and so are used less. Most of the frequently used reporters emit at a wavelength similar to that of natural molecules, particularly GFP and DsRed. The background is also high with these probes, so few markers are wholly satisfying. However, whole-body fluorescence imaging has been used to visualize gene expression [21].

***Positron Emission Tomography.*** PET can be used to study gene transfer expression in living animals. It has a satisfactory resolution (around 1 mm) for definition of organ targeting and, above all, gives a quantitative signal. The most commonly used positron-emitting radioisotopes are $^{18}F$, $^{15}O$, $^{13}N$, and $^{11}C$. PET imaging is used diagnostically in humans to monitor function or localize tumors. Its potential has been progressively enhanced and the increased resolution of microPET methods enables imaging in small laboratory animals [22] (Figure 7). The three main PET approaches to *in vivo* imaging involve an intracellular enzyme-encoding reporter gene, a membrane receptor-encoding reporter gene, or a membrane transporter-encoding reporter gene [23, 24].

The enzyme approach consists of a genetic modification using an enzyme-encoding reporter gene, usually encoding an HSV-tk mutant (HSV1-sr39tk) [25]. When the genetic modification is performed, the neo-synthesized enzyme is able to phosphorylate uracil derivatives such as [$^{124}$I]FIAU and acycloguanosine derivatives such as [$^{18}$F]FHBG. The phosphorylated substrate is unable to leave the cell and the intracellular accumulation of radiolabeled molecules generates a detectable signal. Using the reporter gene HSV-tk, a significant correlation has been demonstrated in rats (after adenovirus gene transfer) between the transgene expression and tissue accumulation of the reporter probe [$^{18}$F]FGCV [26].

The receptor approach consists of a genetic modification using a reporter gene encoding a receptor, usually the dopamine D2 receptor [27] or the SSTr2 receptor [28]. Using positron-emitting labeled probes that specifically target the receptor, it is possible to identify the sites where the genetic modification occurred.

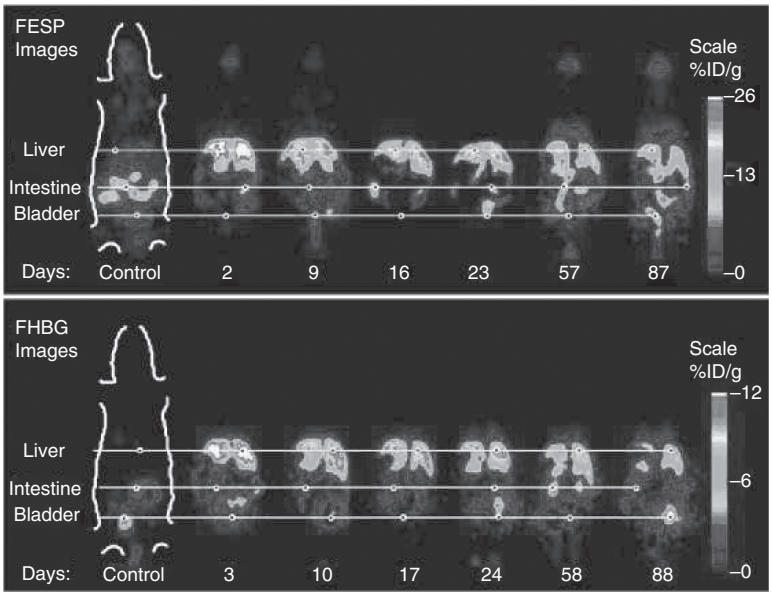


**Figure 7.** MicroPET imaging to visualize the kinetics of gene expression in mice. Mice were transduced with either a D2R PET or a HSV1-sr39TK PET reporter gene by intravenous injection of adDTm. Reprinted from Ref. 22, with permission from Elsevier. (This figure is available in full color at ftp://ftp.wiley.com/public/sci_tech_med/pharmaceutical_biotech/.)

The transporter PET strategy consists of modifying the cells of interest with a gene that encodes a transporter able to generate intracellular accumulation of a radiolabeled reporter probe. The sodium iodide symporter (NIS) is one of the most used transporters in this application [29].

Using sequences that enable simultaneous expression of two genes (such as IRES), it is theoretically possible to define expression of the gene of interest in terms of location, magnitude, and duration by studying reporter gene expression [30].

***Magnetic Resonance Imaging.*** MRI can be used to assess gene transfer expression, which is visualized and localized in the animal's body through planar or spatial (2D or 3D) images. *In vivo* transgene expression can be observed in real time, thereby giving information on the time lag between gene transfer and gene expression. MRI has the great advantage of visualizing surrounding tissues. It is therefore possible not only to assess gene transfer efficiency but also any deleterious (inflammation, necrosis, and so on) or beneficial (such as tumor regression) effects [31]. Any effects of gene transfer are defined early and at high resolution and sensitivity, whatever the gene transfer process adopted. MRI can be used for whole-body imaging of gene expression. In mouse brain, gene expression can be monitored after stereotaxic injection of recombinant adenovirus that encodes ferritin [2] (Figure 8). The relevance of gene transfer to rat muscle by electrotransfer has also been documented by MRI. Expression of the reporter gene luciferase was well correlated with contrast agent entrapment (Gd-DOTA) when administered by the same procedure [17]. Coadministration of the contrast agent and the reporter gene

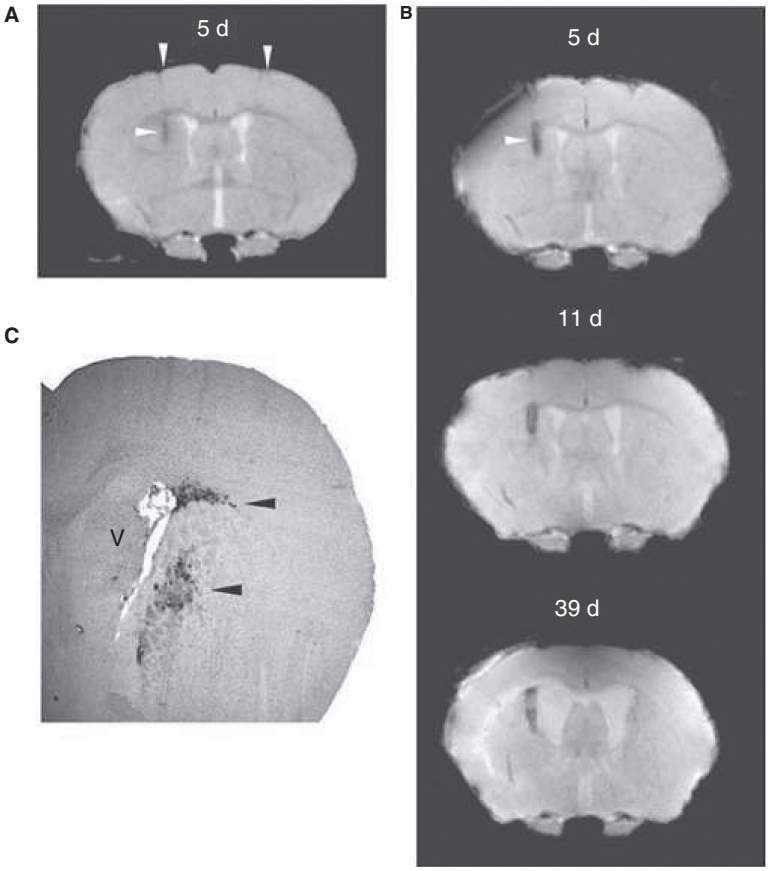


**Figure 8.** MRI imaging for gene expression study. In this example, Genove et al. [2] have documented the expression of a ferritin-encoding gene after stereotaxic injection of a recombinant adenovirus. In this study, the MRI imaging approach was validated by an X-Gal staining after injection of an AdV-LacZ. MRI allows a kinetic study of the gene expression. Reprinted from Ref. 2, with permission from Nature Publication Group. (This figure is available in full color at ftp://ftp.wiley.com/public/sci_tech_med/pharmaceutical_biotech/.)

does not modify reporter gene expression. So, once the experimental parameters have been optimized, MRI can be used to indirectly evaluate gene transfer expression and location.

### 3.3 Advantages and Drawbacks of Imaging Approaches

Gene transfer approaches are not only used to study gene transfer agents. Some reporter genes introduce tags into *ex vivo* modified cells, thereby allowing imaging of their fate (migration, homing, and so on).

To complete the gene transfer studies, safety can be studied in animal models by using ultrasound or X-ray computed tomography to document any tissue or organ alteration. Planar imaging is valuable and easy to perform, but does not really

reflect the true situation, as photon emission depends on tissue type and depth. Moreover, emission is multidirectional, whereas collection is planar, which leads to a large loss of information. PET is currently only used to characterize gene expression (time lag, level, and duration), which may seem quite limiting with regard to the questions initially posed.

Biodistribution studies using PET can be imagined in which a positron emitter is used to label vectors (either viral or nonviral) or the nucleic acid. Such an approach will be contingent on efficient labeling that does not alter the various components.

MRI of cells or small biological samples requires high magnetic fields (often up to 11.7 T), and the requisite equipment is limited in availability because of its prohibitive cost. Experimentally, MRI is very satisfying because the signal obtained not only yields an image but also provides information on sample content (MR spectroscopy). This twofold analysis will become increasingly powerful as higher magnetic field strengths are used. MRI will soon be able to give information simultaneously on changes in structure (by imaging) and in chemical composition (by spectroscopy) of the targeted site.

Bioluminescence devices are also able to give 3D images, but quantification is indirect and based on mathematical treatment that involves considerable extrapolation.

## 4 CONCLUSION

Advances in gene transfer have gone hand in hand with developments in the imaging technologies used to assess transfer efficiency. In gene expression studies, the most frequently used imaging methods are optical (particularly bioluminescence) and nuclear (PET). Whereas optical approaches are limited by the tissue absorption of the signal, the nuclear approach suffers from insufficient resolution. Magnetic resonance offers high resolution, but higher magnetic fields and more relevant and sensitive probes are needed.

No single imaging method can fulfill all research requirements. For large-scale screening or preliminary validation of a new approach, the least expensive method is to be preferred. More costly methods can be applied once the gene transfer approach has been shown to be feasible for gene therapy. Some molecular systems enable two imaging strategies. Using a fusion protein encoding both luciferase and an HSVtk enzyme, gene transfer has been documented using optical bioluminescence and PET [32]. Such fusion proteins are very promising and current improvements might be of great value in gene transfer imaging. An alternative to these fusion proteins is to use two reporter genes (delivered by the same route) to enable a combination of methods, such as PET and fluorescence imaging. Reporter genes may also be coupled with a therapeutic gene through an IRES sequence to monitor modified cells after subcutaneous injection in the mouse [33]. This approach visualizes where expression occurs and its therapeutic effects. But what of functional changes induced by the reporter gene? In the case of PET, three approaches can be envisioned, involving enzymes, receptors, or transporters. How might these proteins impact on cell function? Is therapeutic gene expression really representative of the effects in normal conditions? And how reliable a picture of reality is

given by imaging techniques? Proposed solutions to such outstanding questions may be interesting in research terms but are not always appropriate in a clinical setting. It seems best to tackle these problems by comparing and combining data from various imaging approaches.

## REFERENCES


1. Montier T, Cavalier A, Delépine P, et al. (2003). The use of in situ hybridization to study the transgene pathway following cellular transfection with cationic phosphonolipids. *Blood Cells Molec. Dis.* 30:112–123.
2. Genove G, DeMarco U, Xu H, et al. (2005). A new transgene reporter for in vivo magnetic resonance imaging. *Nat. Med.* 11:450–454.
3. Louie A Y, Huber M M, Ahrens E T, et al. (2000). In vivo visualization of gene expression using magnetic resonance imaging. *Nat. Biotechnol.* 18:321–325.
4. Weissleder R, Moore A, Mahmood U, et al. (2000). In vivo magnetic resonance imaging of transgene expression. *Nat. Med.* 6:351–355.
5. Dodd S J, Williams M, Suhan J P, et al. (1999). Detection of single mammalian cells by high-resolution magnetic resonance imaging. *Biophys. J.* 76:103–109.
6. So M K, Xu C, Loening A M, et al. (2006). Self-illuminating quantum dot conjugates for in vivo imaging. *Nat. Biotechnol.* 24:339–343.
7. Touze A, Courasaget P (1998). *In vitro* gene transfer using human paillomavirus-like particles. *Nucleic Acids Res.* 26:1317–1323.
8. Labat-Moleur F, Steffan A M, Brisson C, et al. (1996). An electron microscopy study into the mechanism of gene transfer with lipopolyamines. *Gene Ther.* 3:1010–1017.
9. Dunlap D D, Maggi A, Soria M R, et al. (1997). Nanoscopic structure of DNA condensed for gene delivery. *Nucleic Acids Res.* 25:3095–3101.
10. Kumar V V, Pichon C, Refregiers M, et al. (2003). Single histidine residue in headgroup region is sufficient to impart remarkable gene transfection properties to cationic lipids: evidence for histidine-mediated membrane fusion at acidic pH. *Gene Ther.* 10:1206–1215.
11. Goncalves C, Mennesson E, Fuchs R, et al. (2004). Macropinocytosis of polyplexes and recycling of plasmid via the clathrin-dependent pathway impair the transfection efficiency of human hepatocarcinoma cells. *Molec. Ther.* 10:373–385.
12. Kayyem J F, Kumar R M, Fraser S E, et al. (1995). Receptor-targeted co-transport of DNA and magnetic resonance contrast agents. *Chem. Biol.* 2:615–620.
13. Bell J D, Taylor-Robinson S D (2000). Assessing gene expression in vivo: Magnetic resonance imaging and spectroscopy. *Gene Ther.* 7:1259–1264.
14. Lerondel S, Le Pape A, Sene C, et al. (2001). Radioisotopic imaging allows optimization of adenovirus lung deposition for cystic fibrosis gene therapy. *Human Gene Ther.* 12:1–11.
15. Delépine P, Montier T, Guillaume C, et al. (2002). Visualization of the transgene distribution according to the administration route allows prediction of the transfection efficacy and validation of the results obtained. *Gene Ther.* 9:736–739.
16. Delépine P, Guillaume C, Montier T, et al. (2003). Biodistribution study of phosphonolipids: A class of non-viral vectors efficient in mice lung-directed gene transfer. *J. Gene Med.* 5:600–608.
17. Leroy-Willig A, Bureau M F, Scherman D, et al. (2005). In vivo NMR imaging evaluation of efficiency and toxicity of gene electrotransfer in rat muscle. *Gene Ther.* 12:1434–1443.

18. Bartoli M, Poupiot J, Goyenvalle A, et al. (2006). Noninvasive monitoring of therapeutic gene transfer in animal models of muscular dystrophies. *Gene Ther.* 13:20–28.
19. Delépine P, Montier T, Lerondel S, et al. (2002). Imaging of gene distribution and gene expression: A new approach for gene therapy studies on living little rodents. *10th meeting of the ESGT. Antibes-Juan les Pins, France.*
20. Contag P R, Olomu I N, Stevenson D K, et al. (1998). Bioluminescent indicators in living mammals. *Nature Med.* 4:245–247.
21. Yang M, Baranov E, Moossa A R, et al. (2000). Visualizing gene expression by whole-body fluorescence imaging. *Proc. Nat. Acad. Sci. USA.* 97:12278–12282.
22. Liang Q, Gotts J, Satyamurthy N, et al. (2002). Noninvasive, repetitive, quantitative measurement of gene expression from a bicistronic message by positron emission tomography, following gene transfer with adenovirus. *Molec. Ther.* 6(1):73–82.
23. Penuelas I, Boan J, Marti-Climent J M, et al. (2004). Positron emission tomography and gene therapy: Basic concepts and experimental approaches for *in vivo* gene expression imaging. *Molec. Imag. Biol.* 6:225–238.
24. Dharmarajan S, Schuster D P. (2005). Molecular imaging of pulmonary gene expression with positron emission tomography. *Proc. Amer. Thoracic Soc.* 2:549–552, 514–516.
25. Gambhir S S, Bauer E, Black M E, et al. (2000). A mutant herpes simplex virus type 1 thymidine kinase reporter gene shows improved sensitivity for imaging reporter gene expression with positron emission tomography. *Proc. Nat. Acad. Sci. U S A.* 97:2785–2790.
26. Gambhir S S, Barrio J R, Phelps M E, et al. (1999). Imaging adenoviral-directed reporter gene expression in living animals with positron emission tomography. *Proc. Nat. Acad. Sci. U S A.* 96:2333–2338.
27. MacLaren D C, Gambhir S S, Satyamurthy N, et al. (1999). Repetitive, non-invasive imaging of the dopamine D2 receptor as a reporter gene in living animals. *Gene Ther.* 6:785–791.
28. Zinn K R, Buchsbaum D J, Chaudhuri T R, et al. (2000). Noninvasive monitoring of gene transfer using a reporter receptor imaged with a high-affinity peptide radiolabeled with 99mTc or 188Re. *J. Nucl. Med.* 41:887–895.
29. Groot-Wassink T, Aboagye E O, Glaser M, et al. (2002). Adenovirus biodistribution and noninvasive imaging of gene expression in vivo by positron emission tomography using human sodium/iodide symporter as reporter gene. *Human Gene Ther.* 13:1723–1735.
30. Inubushi M, Tamaki N (2005). Positron emission tomography reporter gene imaging in the myocardium: for monitoring of angiogenic gene therapy in ischemic heart disease. *J. Cardiac Surg.* 20:S20–24.
31. Kuehn B M (2005). MRI reveals gene activity in vivo. *JAMA* 293:2584.
32. Ray P, Wu A M, Gambhir S S (2003). Optical bioluminescence and positron emission tomography imaging of a novel fusion reporter gene in tumor xenografts of living mice. *Cancer Res.* 63:1160–1165.
33. Luker G D, Sharma V, Pica C M, et al. (2002). Noninvasive imaging of protein-protein interactions in living animals. *Proc. Nat. Acad. Sci. USA.* 99:6961–6966.

# OVERVIEW OF STEM AND ARTIFICIAL CELLS

Alejandro Soto-Gutierrez, Nalu Navarro-Alvarez, Jorge David Rivas-Carrillo, and Naoya Kobayashi

*Okayama University Graduate School of Medicine and Dentistry, Okayama, Japan*

**Chapter Contents**

1 Introductory Remarks
2 Embryonic Stem Cells
  2.1 Introduction
    2.1.1 Properties of Embryonic Stem Cells
    2.1.2 Use and Applications of ES Cells
  2.2 Conclusions
    References
3 Adult Somatic Stem Cells or Postnatal Stem Cells
  3.1 Introduction
    3.1.1 General Concepts of Stem Cells
    3.1.2 General Classification of Stem Cells
    3.1.3 General Characteristics of Stem Cells
    3.1.4 Hematopoietic Stem Cells (HSCs)
    3.1.5 Mesenchymal Stem Cells (MSCs)
    3.1.6 Multipotent Adult Progenitor Cells (MAPCs)
    3.1.7 Side-Population Phenotype Cells (SPs)
    3.1.8 Tissue-Specific Cell Progenitors (TSCPs)
    3.1.9 Umbilical Cord Blood-Derived Stem Cells (UCBSCs)
    3.1.10 Heterogenous Populations of Stem Cells in Bone Marrow
    3.1.11 Uses and Applications of Stem Cells in Toxicology
    References
4 Artificial Cells
  4.1 Introduction
    4.1.1 Immortalization of Human Cells
    4.1.2 Tightly Immortalized Human Cell Lines
    4.1.3 Conditionally Immortalized Cell Lines

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

4.1.4 Ultra-Transform Immortalized Transgenic Cell Lines
4.1.5 Reversibly Immortalized Cell Lines
4.1.6 Applications
4.2 Conclusions
References

## 1 INTRODUCTORY REMARKS

Research on stem cells allows us to get the information about how an organism grows and develops from a single cell and how healthy normal cells replace damaged cells in adult organisms. This promising area of stem cell studies is fascinating scientists to investigate the possibility of cell-based therapies to treat a wide range of diseases, which is referred to as regenerative medicine. Stem cells have two important characteristics that distinguish them from other types of cells. First, they are capable of renewing themselves for long periods through cell division. Second, under some experimental conditions, the cells can be induced to be the functional cells, such as the beating cells of the heart muscle, the albumin-producing cells of the liver, or the insulin-secreting cells of the pancreas. Thus, it is now hypothesized by researchers that stem cells may, in the near future, play a basic role in treating diseases such as heart disease, liver disease, and diabetes. Toward that goal, it is important to understand stem cell biology and its therapeutics. Eventually, control of the growth and differentiation of stem cells will be a big tool in the fields of regenerative medicine, tissue engineering, drug discovery, and toxicity testing.

Although normal human cells are ideal to develop cell therapy, it is unlikely that human cells can be isolated on a scale sufficient to treat many patients. The use of animal cells results in the concerns related to the transmission of infectious pathogens and immunologic and physiologic incompatibilities between the donor and humans. Human embryonic stem cells and bone marrow multipotent adult progenitor cells have receive much attention as a possible source for such cell therapy and drug discovery. It is unlikely that perfect control of differentiation of these multipotent cells will be achieved in very near future. Another attractive cell source is human-derived cell lines. In particular, the use of tightly regulated clonal human cell lines are of value. Such cell lines grow economically in tissue culture and provide the advantage of uniformity, sterility, and freedom of pathogens. Reversible immortalization mediated by Cre/loxP site recombination seems to be the most reliable approach to construct human cells for the clinical setting. In Section 2, the authors review the nature of embryonic stem cells and their potential. In Section 3, somatic stem cells are discussed. In Section 4, the authors describe the general concepts of senescence, crisis, telomeres, telomerase, and immortalization of the cells. Finally, the authors introduce the construction of artificial human liver cells and pancreatic beta cells by the use of Cre/loxP-based reversible immortalization.

## 2 EMBRYONIC STEM CELLS

### 2.1 Introduction

Embryonic stem cells have become a very important source for basic research and possible clinical applications since more than 20 years ago when the establishment of mouse embryonic stem cells (ES cells) was achieved by Evans and Kaufman [1] and Martin [2]. They established the pluripotent cells from the inner cell mass of the blastocyst, mouse ES cells, and developed the culture conditions for the cells *in vitro*. The achievement evolved the studies on teratocarcinomas, tumors that arise in the gonads of several inbred strains and consist of an array of somatic tissues juxtaposed together in a disorganized fashion, and gave the origins of the concept of embryonal carcinoma (EC). As teratocarcinomas could also induced by grafting the blastocyst to ectopic sites, it was likely that pluripotent cell lines could be derived directly from the blastocyst. As expected, a stable diploid cell line that could differentiate into all embryonic cell types was established and formed functional germ cells after transplantation into chimeric mice [1–3]. Testicular teratocarcinomas occur spontaneously in humans, and pluripotent cell lines were derived from the teratocarcinomas [4]. In 1998, human ES cell lines were established from pre-implanted embryos by Thomson et al. [5]. Frozen human embryos that were produced by *in vitro* fertilization at an early stage were thawed and cultured to the blastocyst stage. Fourteen inner cell masses were isolated, and five ES cell lines were established. During the first 8 months of culture, no period of replicative crisis was observed in any cell lines. The principal characteristics of the established cells were that the cells expressed high levels of telomerase activity, which suggests that their life span exceeds that of somatic cells. These cells also expressed surface markers that are typical of human embryonic carcinoma cells, such stage-specific embryonic antigen (SSEA)-3, SSEA-4, TRA-1-60, TRA-1-81, and alkaline phosphatase. The cells also expressed the Transcription Factor Octamer binding protein 4 (Oct-4) to undergo somatic differentiation [6]. ES cells could be used for many different purposes, including early development research [7], toxicology and drug screening [8], and gene and protein screening. One of the most important uses is for regenerative medicine and cell therapy, which involves the transplantation of healthy, functional, and propagating cells to restore the viability or function of deficient tissues and organs [9]. The availability of reliable cells is essential for the toxicology and drug discovery process. The primary cells, immortalized cells, and genetically modified cells have been used for drug discovery; however, the inconsistent availability of the primary cells and the genetic abnormalities of transformed cells are the current problems with such application. ES cells could offer considerable advantages due to their plasticity, proliferative capacity, and the ability to undergo homologous recombination at relatively high frequency. In conclusion, ES cells may offer several important advantages in the field of basic biology, drug discovery, and the future cell therapies in various human diseases.

#### *2.1.1 Properties of Embryonic Stem Cells*

*Pluripotency*

MOUSE ES CELLS. Culturing mouse ES cells from the inner cell mass of the preimplanted blastocyst were first reported more than 20 years ago [1, 2]. To date only

three species of mammals have yielded long-term cultures of self-renewing ES cells: mice, monkeys, and humans [1, 5, 10]. Mouse ES cells lines have shown an unlimited capacity of proliferation and ability to contribute to all cell lineages, which are defined as pluripotency to all definitive tissues: ectoderm, mesoderm, and endoderm. Leukemia inhibitory factor (LIF), a cytokine belonging to the IL-6 family, has been identified as an exogenous signal to maintain ES self-renewal [11]. LIF was initially identified by its activity to induce differentiation of M1 leukemia cells [12], whereas it mediates the opposite cellular responses in ES cells. The signal transduction of LIF consists in the LIF-specific receptor subunit LIFRβ and the common signal transducer gp130, which is shared between the members of the IL-6 cytokine family; the gp130 signaling regulates several cell functions through signal transduction and activation of transcription factor STAT3 that may interact and affect the function of common target genes. Cytokines, including IL-6, IL-11, oncostatin M, ciliary neurotrophic factor, and cardiotrophin-1, show similar properties maintaining the pluripotency of mES cells [12]. A coculture of mouse ES cells on an inactivated mouse embryonic fibroblast layer is also required to maintain the undifferentiated stage. Thus, production of some critical factors of the fibroblast layer is required either to promote self-renewal of mouse ES cells or to suppress their differentiation. The activation of STAT3 is essential to the LIF signaling pathway, but it plays an accessory role to maintain ES cell identity [13]. Moreover, there are two major pathways of intracellular signal transduction downstream of gp130; the Jak-Stat pathway and the Shp2-Erk pathway. As Jak and Shp2 interact with separate subdomains of the intracellular domain of gp-130, it has been demonstrated that the activation of Jak but not Shp2 is sufficient to preserve an undifferentiated status of mouse ES cells. The finding means that the LIF signal is mainly transmitted to the nuclei by the Jak-STAT signal pathway [14] and the Shp2-Erk pathway does not contribute directly to stem cell renewal as demonstrated when adding Erk kinase inhibitor PD98059 in the medium resulting in self-renewal [14]. Although STAT3 acts as a transcription factor to activate target genes, there is only one gene whose specific function in pluripotent cell population is confirmed, that is, the POU-family transcription factor octamer-3/4 (Oct-3/4) encoded by Pou5f1 [12]. The essential role of Oct-3/4 in mouse development has been revealed by targeting gene deletion [15]. Oct-3/4-deficient embryos fail to initiate fetal development because the prospective founder cells of the ICM do not require pluripotency and become diverted into the trophoectoderm lineage, which indicates that Oct-3/4 is essential to establish a pluripotent cell population in preimplantation development [15]. ES cells require a critical level of Oct-3/4 to maintain stem cell renewal, and at least a twofold increase in the expression of Oct-3/4 causes differentiation of mouse ES cells into the endoderm and mesoderm, whereas reduction to less than 50% of the normal expression level of Oct-3/4 triggers de-differentiation of mouse ES cells into the trophectoderm [16]. Recently identification of the homeodomain protein Nanog as another key regulator of pluripotentiality has opened another door of the complicated system of pluripotency; the dosage of Nanog is a critical determinant of cytokine-independent colony formation, and the forced expression of this protein confers constitutive self-renewal in ES cells without gp-130 stimulation; Nanog may act to restrict the differentiation-inducing potential of Oct-3/4 [17]. Other investigators have reported that the induction of the expression of Inhibitor of differentiation (Id) by addition of TGF-β1/bone morphogenetic protein

(BMP) in combination with LIF sustains self-renewal via the Smad pathway [18]. Recent studies have implicated the importance of Wnt-signaling pathways in the maintenance of ES cell pluripotency [19]. Components of the β catenin Wnt-signaling pathway are expressed in ES cells, and it is likely that activation of the Wnt-signaling pathway using a glycogen synthase kinase (GSK)-3-specific inhibitor (BIO) that maintains the pluripotent state of human ES and mouse ES cells [19] (see Figure 2-1).

The pluripotency of ES cells depends on the balance of various signaling molecules, and thus, such imbalance causes differentiation of ES cells. Many other molecules, such as Genesis [20], Rex-1 [21], Sox2 [22], GBX2 [23], and UTF1 [24], have been identified with a potential role in defining pluripotency.

Human ES Cells. The murine models of isolation, derivation, culture, and characterization in ES cells have provided valuable information to the generation of human ES (hES) cell lines derived from the embryos at the preimplantation stage, which involves culturing embryos to the morula or blastocyst stage (see Figure 2-2). These embryos are donated for research to establish hES cell lines. Thomson et al. isolated from the ICM of human blastocysts, placed on inactivated murine feeder

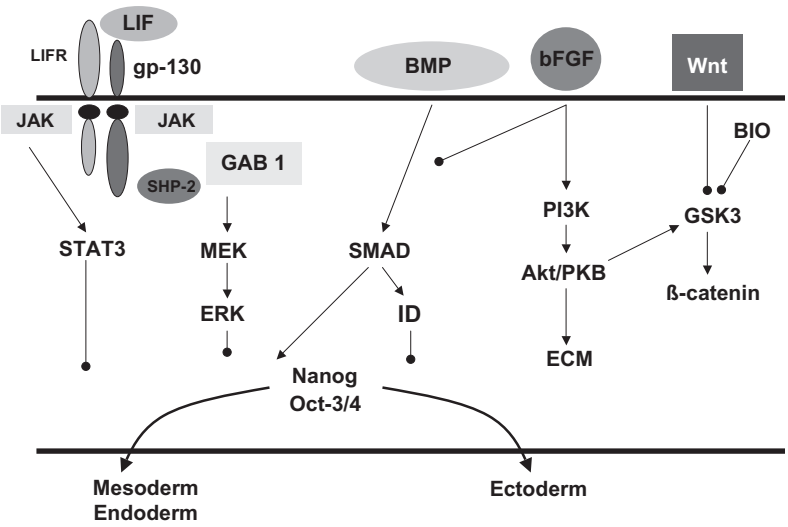


**Figure 2-1.** ***Regulation pathways of self-renewal in the undifferentiated stage of ES cells.*** The undifferentiation of ES cells is regulated by Nanog, Oct-3/4, and interactions between LIF-dependent JAK/STAT3 pathways principally in mouse ES cells. In human ES cells, the mechanism involved is the BMP-dependent activation of Id target genes; the role of bFGF is to activate the (PI3K)/Akt/PKB pathway, which subsequently down-regulates the expression of ECM molecules; and finally the Wnt pathway activation by specific pharmacological inhibitor BIO of GSK-3 maintains the undifferentiated phenotype in both mouse and human ES cells. The balanced expression level of Oct-3/4 determines the fate of ES cells. Adapted from Niwa.[30] Abbreviations: LIF (Leukemia Inhibitor factor), BMP (bone morphogenetic protein), bFGF (basic fibroblast growth factor), ECM (extracellular matrix), BIO (6-bromoinduribin-3′-oxime), GSK3 (glycogen synthase kinase-3). (This figure is available in full color at ftp://ftp.wiley.com/public/sci_tech_med/pharmaceutical_biotech/.)

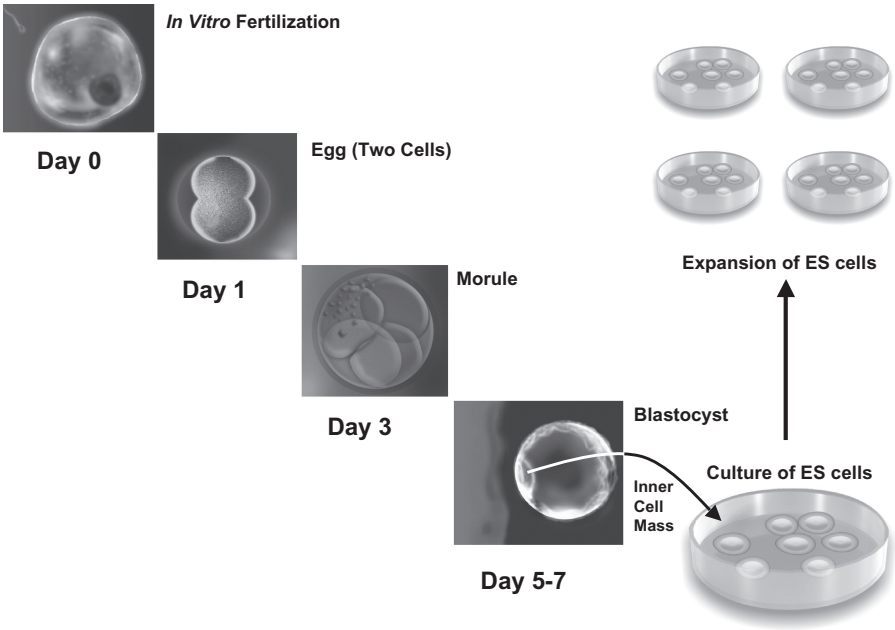


**Figure 2-2.** ***Developmental origin of embryonic stem cell lines.*** Illustration demonstrates the derivation of embryonic stem cells and their culture.

cells, and successfully performed initial derivations of hES cell lines [5]. Since that time, many laboratories have applied the same techniques to derivate hES cell lines (see Figure 2-2). hES cells have been currently characterized by a set of markers and their differentiation capacity. These criteria include the expression of several surface markers and transcription factors that associated with an undifferentiated state. In addition, maintenance of extended proliferating capacity, pluripotency, and normal euploid karyotype without marked change in the epigenetic status of hES cells is a crucial issue for the future use of the cells in clinical trials. Several surface markers have been identified to characterize hES cells, in which glycolipids and glycoproteins, such as SSEA-4, TRA-1-60, and TRA-1-81, that are expressed in human embryocarcinoma cells are present in hES cells. Some surface antigens initially described in other stem cells are AC133, CD-9, CD-117, and CD-135, which are also expressed in hES cells [25]. The stability of the expression of these surface markers in hES cells after culture for prolonged periods of time has been maintained.

Several critical transcription factors that play a critical role in maintaining self-renewal of stem cells have now been identified, and these analyses are also useful for characterization of hES cells. One of these transcriptional factors is Oct-3/4, and several Oct-3/4-targeting genes have been identified in hES cells, which are Utf-1, Rex-1, PDGFαR, Otx-2, Lefty-1, and Nanog. However, their specific roles have not been identified, and their expression has been retained in hES cells for over a year in culture [15–19, 25]. Additionally, all hES cell lines express high levels of telomerase, an enzyme that helps to maintain telomeres that protect the end of the chromosomes. Telomerase activity and long telomeres are characteristics of proliferating

cells in embryonic tissue and germ cells. As cells divide and differentiate throughout the life span of an organism or cell line, the telomeres become progressively shortened and lose the ability to maintain their length. The functions of telomeres and telomerase seem to be important in the cell division, normal development, and aging [26]. Mouse ES cells have shown to require activation of the gp-130/STAT3 pathway to maintain the cells in the undifferentiated stage. This activation is generally achieved by the addition of LIF to the culture medium. In contrast, it has been demonstrated that the addition of exogenous LIF to hES cell culture does not maintain the pluripotent capacity of hES cells [27]. However, common intracellular signaling pathways exist between mES cells and hES cells to regulate self-renewal and to maintain an undifferentiated state. These signaling pathways have not yet been clarified, although transcriptional profiling or gene expression technology has identified several genes, transcription factors, ligand/receptor pairs, and secreted inhibitors of signaling pathways. hES cells can be maintained in several different conditions by the use of growth factors belonging to the TGF-β1/BMP superfamily, fibroblast growth factor (FGF) family, and Wnt family.

Previous studies have demonstrated that prolonged propagation of undifferentiated hES cells requires culture of the cells on embryonic fibroblast feeder layers. The use of bFGF has become a common thread in the culture medium formulas, which have been recently developed by using feeder-free conditions [28]. hES cells can be maintained in serum replacement-containing medium that has been supplemented with 36–40 ng/mL of bFGF, in which culture condition the cells can retain the expression of surface markers, transcription factors, normal karyotype, telomerase activity, and pluripotency of hES cells conventionally cultured on embryonic fibroblast feeder layers. The mechanism of bFGF activity in hES cell culture remains unclear, but it has been demonstrated that treatment with 40–100 ng/mL of bFGF inhibits BMP signaling in hES cells. A further role of bFGF in hES cells to maintain pluripotentiality is activation of the phosphatidylinositol 3-kinase (P13K/Akt, PKB) pathway, which subsequently enhances the expression of extracellular matrix molecules (ECMs). Removal of bFGF in hES cell culture or the treatment of the cells with chemical inhibitors of the PKB pathway results in downregulation of the expression of hES cell markers as well as a decrease in ECM components [29] (see Figure 2-1).

*Differentiation.* Pluripotency is one of the defining features of ES cells. The most definitive test of pluripotency of the cells is the formation of chimeras in mice in which the mES cells are injected into the blastocyst. The approach cannot be applied to assess pluripotency of hES cells, therefore, and teratoma formation after injection of embryonic bodies (EBs) of hES cells *in vitro* into immunocompromised mice is currently used to validate the pluripotency of the established hES cell lines in culture. In the case of mES cells, once differentiation of ES cells has started, the cells representing the primary germ layers spontaneously develop *in vitro* in the absence of LIF. The culture conditions to form EBs include hanging drops [30], suspension mass culture [31], or the use of methylcellulose [32]. Initially, an outer layer of endoderm-like cells forms within ESs, followed by the development of an ectodermal layer and subsequent specification of mesodermal cells over a period of a few days [33]. The generation of specific functional cell types from hES cells has been demonstrated both *in vitro* and *in vivo*. In fact, with the rapid interest in

gene targeting for the development of genetically modified mice, most of the efforts directed toward ES cells have been made in the maintenance of ES cells in an undifferentiated state. Although research on *in vitro* differentiation of ES cells has been limited, investigators have demonstrated differentiation protocols, which assess the differentiated cells via expression analysis of cell-specific markers. However, very few markers are specific for one cell type, and for that reason, panels of markers must be used in these experiments. Understanding of cellular differentiation during embryogenesis has led to methods for enriching populations of specific cell types: First, genetic manipulation of ES cells facilitates differentiation and serves as a framework for genetic engineering of ES cells by key transcription factors to regulate their cell fate [34]. Second, culture conditions supplemented with established growth factors can be used. Several samples of clinically and pharmacologically relevancy have been proposed, including almost any kind of cells, such as chondrocytes [35], atrial and ventricular cardiomyocytes [30], hepatocytes [36], pancreatic islet cells [37], and motor neurons [38]. Third, coculture conditions have been reported to facilitate and produce differentiated cells [39] (see Figure 2-3). Fourth, ectopic implantation of ES cells into syngenic or immune-compromised mice produces specific cell types [40]. Although such transplantation is a powerful approach, this is not a practical method for deriving cells to be evaluated in clinical experiment. Currently the desired cells cannot be developed on a stable large scale. Inductive differentiation protocols have generated many cell types to further enrich *in vitro* differentiated populations by the use of selective methods that are based on the expression of specific marker proteins. Such selective

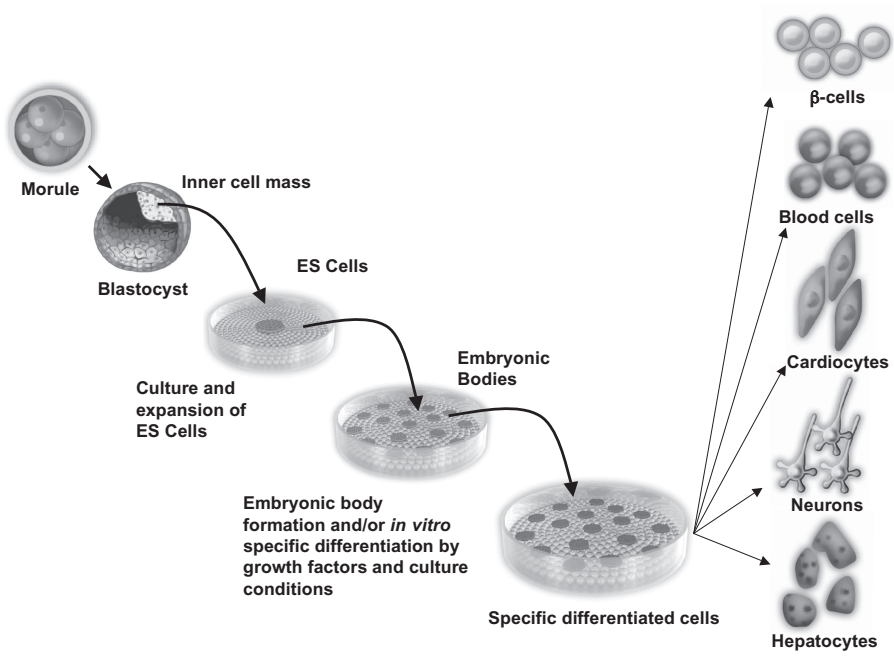


**Figure 2-3.** ***Differentiation potential of embryonic stem cells.*** Illustration shows the possible application of ES cells for cell therapy.

protocols can include promoter-based recovering strategies, based on GFP-dependent cell sorting [41], or antibiotic selection [42], and have allowed the enrichment of increasingly defined phenotypes from ES cells. Cell sorters have been applied to enrich the populations by using specific cell surface markers of ES cell derivatives.

Investigators have used the forced gene expression to influence *in vitro* differentiation of ES cells. For example, constitutive overexpression of murine *Pax 4* in ES cells combined with an inductive protocol has resulted in an enrichment of nestin-positive progenitors and insulin-producing cells among other cells found in pancreatic islets [43]. Finally, the next step in developing clinical trials and drug discovery is the use of internal bioimplants using bioactive materials that provide biological signals at the site of damaged tissues *in vivo*. This next generation of implantable bioartificial devices can permit the evaluation of the potential of ES cell derivatives *in vivo* without having direct contact with the blood circulation. Such an approach allows researchers to avoid a risk of tumor formation of ES cell derivatives and to examine the fate of such cells in an isolation manner from the body.

ENDODERMAL DIFFENTIATION. The pancreas and liver are the derivatives of the definitive endoderm; hepatic and pancreatic cells are of special therapeutic interest for the treatment of diabetes mellitus and hepatic failure. Theoretically, both cells can be generated from ES cells. Researchers have demonstrated different strategies *in vitro* to differentiated mES cells in hepatocyte-like cells; these cells have shown specific transcription factors and proteins of normal hepatocytes; they also have been shown the presence of the two lineages of the liver cells, bile duct epithelial cells and oval cells. Once the cells are transplanted, these cells can integrate into the hepatic parenchyma and function in the host liver. It has been previously demonstrated that ES cells can differentiate into hepatocyte-like cells [36, 40, 44] through *in vivo* differentiation of mES cells in the damaged host liver and subsequent recovery of mES cell-derived hepatocytes by the use of an albumin-promoter-derived GFP expression system [40]. Briefly, mES cells in which an albumin-promoter-derived GFP expression cassette was externally introduced were injected into the spleen of the hepatic-injured mice and tumors were developed after 3 weeks. Examination of the tumors showed the presence of hepatocytes derived from mES cells. These hepatocytes were positive for hepatic markers for at least 3 weeks and were capable of proliferating. The cells were highly characterized by revealing correspondence in 98% of gene expression profiles compared with primary mouse hepatocytes [45]. It is known that the formation of EBs itself can raise the cell population expressing hepatocyte phenotype [46], but such a population in EBs is estimated to be considerably small. Differentiation and isolation of hepatocyte-like cells from hES cells has been demonstrated by using hES cells that are stably transfected with the reporter gene of GFP fused to an albumin-promoter [47]. The generation of ES-derived insulin-producing cells may represent a critical cell source for the treatment of diabetes. In this field, ES cells hold a great hope as a source of β-cells, but unfortunately this has proven to be more complicated than expected. Both mouse and human ES cells can be manipulated to contain insulin and even to regulate insulin secretion [37, 48, 49]. Several methods have been used to obtain enriched populations by the stable transfection with a cDNA construct

containing a neomycin-resistance gene under the control of the insulin or Nkx6.1 promoter [49], by the selection with nestin-positive cells [37, 50], or induction of critical transcription factors, such as pax4 and pdx-1 [43, 51]. Although these reports have encouraged that ES cells can generate cells containing insulin and pancreatic β-cell markers, a recent report has demonstrated that insulin-positivity of ES cell derivatives was considerable due to insulin uptake from culture media by apoptotic cells [52]. Measurement of C-peptide, which is excised from proinsulin in the process of maturation of insulin, has been strongly recommended when demonstrating differentiation of ES cells to pancreatic beta cells. Differentiation protocols have to clarify that ES cells are directed to definitive endoderm, not to visceral endoderm, which has similar markers of gene expression but different pathways from pancreatic beta cell lineage. According to the latest research, the differentiation ratio of insulin-positive cells from ES cell populations is about 2.7% compared with less than 1% in undifferentiated controls [53].

ECTODERMAL DIFFERENTIATION. Epithelial cell differentiation from ES cells has been identified by the presence of cytokeratines and specific keratinocyte markers [54]. Enrichment of keratinocytes *in vitro* from ES cells has been achieved by seeding the cells on various extracellular matrices in the presence of bone morphoprotein (BMP-4) and/or ascorbate; such protocols promote the formation of epidermal equivalents. It has been reported that the resulting tissue displays patterns similar to the embryonic skin. The cells express the late differentiation markers of fibroblasts, which suggests that ES cells have the capacity to reconstitute fully differentiated skin *in vitro* [55]. Researchers have reported a differentiation capacity of ES cells into neurons and glial cells [56, 57]. The neural differentiation ratios have significantly improved by the introduction of numerous strategies, including lineage selections (dopaminergic, serotonergic [57–63], or γ-aminobutyric acid (GABA)-ergic neurons [57]), astrocytes, oligodendrocytes [64], glutamatergic and cholinergic neurons [58], and growth factors. The possibility of producing dopaminergic neurons from ES cells has been demonstrated by the use of both fibroblast growth factor 8 (FGF8) and sonic hedgehog, which were implicated as tandem initiators of dopaminergic neurogenesis [58]. Later, the enrichment of dopaminergic neurons has been acheived by mimicking the oxygen tension of the developing midbrain [59]. When such cells were implanted into 6-OHDA-lesioned rats, where nigrostrial dopaminergics afferents are largely lost, the brain function restored normalcy to the dopamine-depleted animals [60]. Similar results were observed in animal experiments by the transplantation of monkey ES cell-derived dopaminergic neurons into 1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine (MPTP)-lesioned adult cymolgus monkeys, in which treatment-associated behavioral improvement was noted by 10 weeks after transplantation [61]. In general, proliferation of neural precursor cells is induced by the addition of FGF and epidermal growth factor (EGF) into the culture medium. Thereafter, neural differentiation can be facilitated by the addition of neural differentiation factors, such as glial cell line-derived neurotrophic factors (GDNFs), neurturin (NT), TGF-β3, and IL-1β [62]. Although the ability of hES cells to generate derivatives of the neural ephitelium has been demonstrated, the selective derivation of neuron subtypes has been difficult to achieve until now. Therefore, long-term survival of the grafted cells has not been

successful in experimental models, and a potential risk of teratoma formation by undifferentiated cell populations after implantation should be carefully studied before any therapeutic clinical trials can be considered.

MESODERMAL DIFFERENTIATION. The mesodermal germ layer has the capacity to differentiate into muscle, bone, cartilage, blood, and connective tissues. ES cells also have been successfully used to reconstitute mesodermal developmetal processes *in vitro* by generating several mesodermal cell types, such as adipogenic cells [63], chondrogenic cells [65], osteoblasts [66], and myogenic cells [67]. One of the mostly studied cell lineages in ES cells is cardiomyocytes. Similar to other differentiation protocols, the generation of cardiomyocytes from ES cells requires an initial aggregation step to form EBs. Within EBs, cardiomyocytes are located between an epithelial layer and a basal layer of mesenchymal cells [30]. Cardiomyocytes are easy to identify *in vitro*, because they spontaneously contract after 3–4 days of culture. The number of spontaneously beating cells can be increased by adding differentiation-inducing factors, such as dimethyl sulfoxide (DMSO), retinoic acid, Dynorphin B, and cardiogenol derivatives [30, 68–73]. Developmental changes of cardiomyocytes can be correlated with the length of time in culture and divided into three stages: early (pacemaker-like cells), intermediate, and terminal (atrail-, ventricular-, nodal-, His-, and purkinje-like cells) [30]. hES cell-derived cardiomyocytes have showed the expected molecular, structural, electrophysiologic, and contractile properties of nascent embryonic myocardium [68, 69]. hES cell-derived cardiomyocytes differ from those of mouse in an important and potencially exploitable capacity, that is, proliferation. In contrast to the limited proliferation of mouse ES cell-derived cardiomyocytes [70], human ES cell-derived cardiomyocytes show sustained cell cycle activity both *in vitro* [71] and *in vivo* (in the heart of the nude rats) [72]. *In vivo* studies with hES cell-derived cardiomyocytes have increased lately [72, 73] and suggested that hES cell-resulting cardiomyocytes engraft and integrate into the host myocardium in experimental immunosuppressed animals [73]. These data represent an exciting proof-of-concept evidence for the potential application of hES cell-derived cardiomyocytes in the formation of biological pacemakers. Several important challenges include long-term follow-up of the grafts to confirm whether such cells maintain their pace-making ability over time and development of efficient protocols for large-scale production of highly purified cells without any risks of teratoma formation of the cells *in vivo*.

Hematopoietic cells and blood vessels are believed to arise from common progenitor cells, hemangioblasts. Cystic EBs increase the generation of blood islands containing erythrocytes and macrophages [31], and differentiation of the cells on semi-solid medium is efficient for the formation of neutrophils, mast cells, macrophages, and erythrocytes [32]. Application of fetal calf serum (FCS), cytokines such as IL-1 and IL-3, or granulocyte–macrophage colony stimulating factor (GM–CSF) in ES cell culture generates early hematopoietic precursor cells expressing both embryonic z globin (βH1) and adult β major globin RNAs. Experiments to identify potential inducers of the hematopoietic lineage indicate that Wnt3 is one of the most important signaling molecules that play a significant role in enhancing hematopoietic commitment during *in vitro* differentiation of ES cells [74].

***2.1.2 Use and Applications of ES Cells*** One of the uses of mES cells is the development of genetically modified mice. Researchers have conducted genetic modifications in mES cells to generate genetically modified mice to evaluate specific functions of the targeted genes. These engineered mice have selectively inactivated specific genes to evaluate their target functions and toxicity [75]. This technology can be applied to *in vivo* assays to test for clinical conditions of the central nervous system, cardiovascular diseases, tumors, and metabolic changes in muscle, fat, and bone. The approach has become a crucial method for the utility of different targets in toxicology. Another significant use of ES cells is achieved by a homologous recombination that allows the preplanned replacement of mutations in endogenous murine alleles. Base changes allow an inactivation of specific domains or an alteration of binding sites of a specific protein. For example, single-base changes allow the expression of conserved human mutations. An allelic variant in presenilin 1 is associated with early onset of Alzheimer's disease [76].

*Undifferentiated ES Cells.* The use of undifferentiated ES cells has been rapidly facilitated to gain functional information and to generate and evaluate knockout mice. An example was the development and evaluation of a nuclear factor-κβ precursor p105 in ES cells [77]. The genetically modified ES cells had specific deletions in the C-terminal region of the p105 gene. These cells demonstrated that p105 was important in the control of NF-κβ-binding activity, detailing the role of NF-κβ in inflammatory responses.

Researchers have also used ES cells deficient in specific genes in the undifferentiated state to evaluate the gene function in the cell development *in vitro* [78].

*Differentiated ES Cells.* Researchers have demonstrated that cells derived from ES cells *in vitro* reflect the cellular physiology of relevant to primary cells. Gene inactivation has been applied to study events in signaling pathways that could be relevant in specific cell types. For example, there is the role of mitogen-activated protein kinase kinase-1 (MEKK1) in a cellular model of postischemic reperfusion injury using ES cell-derived cardiomyocytes [79]. This approach clearly allows time- and cost-saving evaluation of target genes. In addition, results from knockout cells derived from ES cells can provide clues as to the developmental cause of embryonic lethality.

Teratoma formation has been used to evaluate the role of target gene expression on the cell proliferation and differentiation of the resulting tumor formation. This kind of strategy was applied to demonstrate a reduced proliferative capacity of ES cells null for cyclooxygenase 2 compared with cyclooxygenase 1-deficient or wild-type ES cells [80]. The strategy can be used to evaluate the role of potential cancer-inducing targets on cell proliferation. Reproductive toxicity is a target where ES cells could be applied to *in vitro* teratology assays [81]. Such a procedure could reduce the use of animal procedures. It has been reported that 78% of correlation exists between the data of *in vivo* and *in vitro* experiments using ES cells as an alternative to animal testing in teratology and embryo-toxicity testing [82]. The aim is to overcome the limitation of availability of primary cells or the genetic abnormality of immortalized cell lines. Generation of specific cell and tissue types is necessary rather than heterogeneous populations that can result from spontaneous differentiation of ES cells. For this reason, ES cell lines expressing GFP

selectively in the specific differentiation to cardiomyocytes [83] or hepatocytes [40] have been designed [84]. Finally, therapeutic concepts of ES cells must be well argued. Some difficulties can be identified in the application of adult stem cells or embryonic stem cells, in order to develop stem cell-replacement therapy. We need to meticulously select the best candidates of cells in terms of the facilities to propagate, manipulate, and select the most purified population of the desired cell type. We need to produce immune tolerance when allogenic cells are used for stem cell-based cell therapy. In addition, ES cells bring certain advantages against the rest of stem cells. Firstly, adult stem cells are difficult to isolate and hard to propagate in culture; in contrast, ES cells are derived easily once an embryo has been obtained and they can grow indefinitely in culture. ES cells can be manipulated genetically by homologous recombination to correct a genetic defect [85]. On the other hand, adult stem cells can be genetically manipulated only by the introduction of retroviral gene delivery, which overexpress the transduced genes and could happen in an insertional mutagenesis [86]. ES cells have the capacity to differentiate into any kind of tissue of the body, that is, pluripotent, but adult stem cells are multipotent and their differentiation capacity seems to be restricted. Protocols must be well developed in ES cells to generate an enrichment population of cells of a specific lineage or cell type. Furthermore, if therapeutic application is the final goal of ES cells, culturing techniques need to be scaled up for mass production of clinically relevant quantities of the specified cells. Finally, immunological barriers must be overcome. Regarding this issue some promising alternatives have been suggested lately. Nuclear transplantation denotes the introduction of a nucleus from an adult donor cell into an enucleated oocyte to generate a cloned embryo. When transferred to the uterus of a female, this embryo has the potential to become an infant that is a clone of the adult donor cell; this process is referred as to "reproductive cloning" (see Figure 2-4). When explanted in culture, this embryo can give rise to ES cells that have the potential to become almost any type of cells present in the adult body. The resulting ES cells by nuclear transfer are genetically identical to the donor and thus potentially useful for therapeutic applications; this process is called "therapeutic cloning" (see Figure 2-4). Therapeutic cloning may substantially improve the treatment for neurodegenerative diseases, blood disorders, or diabetes, because the therapy for such diseases is currently limited by the availability or immunocompatibility of tissue transplants. Indeed, experiments in animals have shown that nuclear cloning combined with gene and cell therapy represents a valid strategy for treating genetic disorders [85].

## 2.2 Conclusions

Considerable progress has been made toward the generation of more defined culture conditions of ES cells since the initial isolation and growth conditions were described. There have been many modifications of the procedures, including growth of ES cells, differentiation potential, and cell lines establishment. The availability of hES cells represents an extraordinary opportunity for cell transplantation that may be applicable to humans in the future. Despite exciting advantages and advances in the field of hES cell research, many challenges have to be addressed in the near future. The culture conditions have to improved and be humanized. The cells cultured in xenogeneic conditions are likely to be considered and

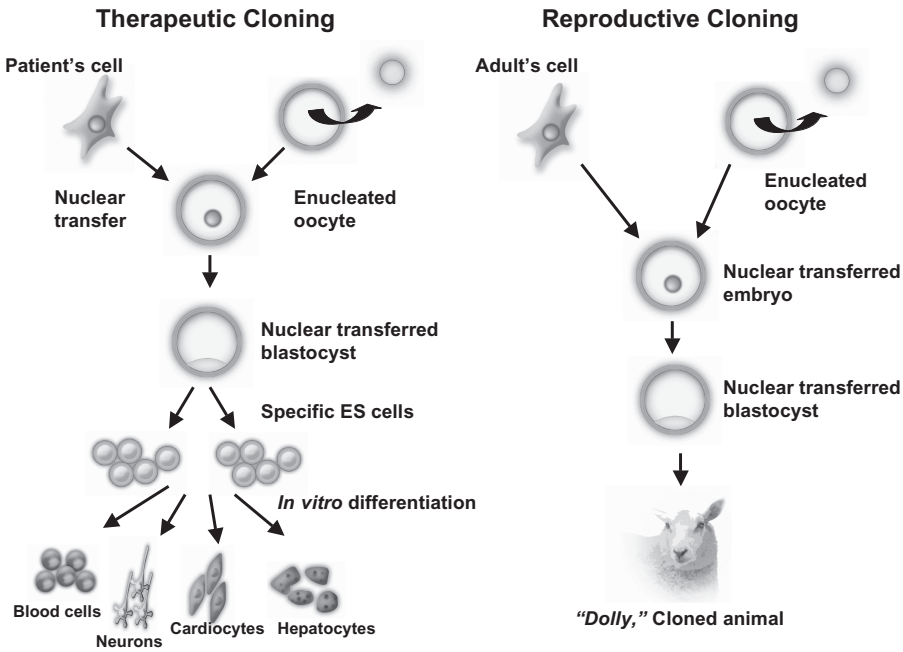


**Figure 2-4.** ***Comparison of therapeutic cloning and reproductive cloning.*** In therapeutic cloning (left panel), the diploid nucleus of an adult donor cell is introduced into an enucleated oocyte and it divides into cloned blastocysts that will provide lines of embryonic stem cells that can be differentiated *in vitro* into any types of cell for therapeutic purposes. In contrast, reproductive cloning (right panel) requires the transfer of the cloned blastocysts into surrogate mothers and a few of the cloned blastocysts will give rise to the cloned animals.

regulated. Concerns about the infection from nonhuman pathogens are the most discussed issues for the clinical application. Although ES cells offer great promise for regenerative medicine, the near-term applications could be in drug discovery. Researchers might have a store of stem cells lines derived from many human populations from different ethnicities and gene variations. These cells can be of extreme importance in the development of effective and safe therapeutics.

## References

1. Evans M J, Kaufman M H (1981). Establishment in culture of pluripotential cells from mouse embryos. *Nature.* 292:154–156.
2. Martin G R (1981). Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. *Proc. Natl. Acad. Sci. USA*. 78:7634–7638.
3. Bradley A, Evans M, Kaufman M H, et al. (1984). Formation of germ-line chimaeras from embryo-derived teratocarcinoma cell lines. *Nature.* 309:255–256.
4. Andrews P W, Damjanov I, Simon D, et al. (1984). Pluripotent embryonal carcinoma clones derived from the human teratocarcinoma cell line Tera-2. Differentiation in vivo and in vitro. *Lab. Invest.* 50:147–162.

5. Thomson J A, Itskovitz-Eldor J, Shapiro S S, et al. (1998). Embryonic stem cell lines derived from human blastocysts. *Science.* 282:1145–1147.
6. Reubinoff B E, Pera M F, Fong C Y, et al. (2000). Embryonic stem cell lines from human blastocysts: Somatic differentiation in vitro. *Nat. Biotechnol.* 18:399–404.
7. Pera M F, Trounson A O (2004). Human embryonic stem cells: Prospects for development. *Develop.* 131:5515–5525.
8. McNeish J (2004). Embryonic stem cells in drug discovery. *Nat. Rev. Drug Discov.* 3:70–80.
9. Strom T B, Field L J, Ruediger M (2002). Allogeneic stem cells, clinical transplantation and the origins of regenerative medicine. *Curr. Opin. Immunol.* 14:601–605.
10. Thomson J A, Kalishman J, Golos T G, et al. (1995). Isolation of a primate embryonic stem cell line. *Proc. Natl. Acad. Sci. USA.* 92:7844–7848.
11. Williams R L, Hilton D J, Pease S, et al. (1988). Myeloid leukaemia inhibitory factor maintains the developmental potential of embryonic stem cells. *Nature.* 336:684–687.
12. Niwa H (2001). Molecular mechanism to maintain stem cell renewal of ES cells. *Cell Struct. Funct.* 26:137–148.
13. Niwa H, Burdon T, Chambers I, et al. (1998). Self-renewal of pluripotent embryonic stem cells is mediated via activation of STAT3. *Genes Dev.* 12:2048–2060.
14. Burdon T, Stracey C, Chambers I, et al. (1999). Suppression of SHP-2 and ERK signalling promotes self-renewal of mouse embryonic stem cells. *Dev. Biol.* 210:30–43.
15. Nichols J, Zevnik B, Anastassiadis K, et al. (1998). Formation of pluripotent stem cells in the mammalian embryo depends on the POU transcription factor Oct4. *Cell.* 95:379–391.
16. Niwa H, Miyazaki J, Smith A G (2000). Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells. *Nat. Genet.* 24:372–376.
17. Chambers I, Colby D, Robertson M, et al. (2003). Functional expression cloning of Nanog, a pluripotency sustaining factor in embryonic stem cells. *Cell.* 113:643–655.
18. Ying Q L, Nichols J, Chambers I, et al. (2003). BMP induction of Id proteins suppresses differentiation and sustains embryonic stem cell self-renewal in collaboration with STAT3. *Cell.* 115:281–292.
19. Paling N R, Wheadon H, Bone H K, et al. (2004). Regulation of embryonic stem cell self-renewal by phosphoinositide 3-kinase-dependent signaling. *J. Biol. Chem.* 279:48063–48070.
20. Sutton J, Costa R, Klug M, et al. (1996). Genesis, a winged helix transcriptional repressor with expression restricted to embryonic stem cells. *J. Biol. Chem.* 271:23126–23133.
21. Hosler B A, Rogers M B, Kozak C A, et al. (1993). An octamer motif contributes to the expression of the retinoic acid-regulated zinc finger gene Rex-1 (Zfp-42) in F9 teratocarcinoma cells. *Mol. Cell Biol.* 13:2919–2928.
22. Avilion A A, Nicolis S K, Pevny L H, et al. (2003). Multipotent cell lineages in early mouse development depend on SOX2 function. *Genes. Dev.* 17:126–140.
23. Chapman G, Remiszewski J L, Webb G C, et al. (1997). The mouse homeobox gene, Gbx2: Genomic organization and expression in pluripotent cells in vitro and in vivo. *Genomics.* 46:223–233.
24. Okuda A, Fukushima A, Nishimoto M, et al. (1998). UTF1, a novel transcriptional coactivator expressed in pluripotent embryonic stem cells and extraembryonic cells. *Embo. J.* 17:2019–2032.
25. Carpenter M K, Rosler E S, Fisk G J, et al. (2004). Properties of four human embryonic stem cell lines maintained in a feeder-free culture system. *Dev. Dyn.* 229:243–258.

26. Amit M, Carpenter M K, Inokuma M S, et al. (2000). Clonally derived human embryonic stem cell lines maintain pluripotency and proliferative potential for prolonged periods of culture. *Dev. Biol.* 227:271–278.
27. Sato N, Sanjuan I M, Heke M, et al. (2003). Molecular signature of human embryonic stem cells and its comparison with the mouse. *Dev. Biol.* 260:404–413.
28. Xu C, Rosler E, Jiang J, et al. (2005). Basic fibroblast growth factor supports undifferentiated human embryonic stem cell growth without conditioned medium. *Stem Cells.* 23:315–323.
29. Kim S J, Cheon S H, Yoo S J, et al. (2005). Contribution of the PI3K/Akt/PKB signal pathway to maintenance of self-renewal in human embryonic stem cells. *FEBS Lett.* 579:534–540.
30. Boheler K R, Czyz J, Tweedie D, et al. (2002). Differentiation of pluripotent embryonic stem cells into cardiomyocytes. *Circ. Res.* 91:189–201.
31. Doetschman T C, Eistetter H, Katz M, et al. (1985). The in vitro development of blastocyst-derived embryonic stem cell lines: Formation of visceral yolk sac, blood islands and myocardium. *J. Embryol. Exp. Morphol.* 87:27–45.
32. Wiles M V, Keller G (1991). Multiple hematopoietic lineages develop from embryonic stem (ES) cells in culture. *Develop.* 111:259–267.
33. Wobus A M, Boheler K R (2005). Embryonic stem cells: prospects for developmental biology and cell therapy. *Physiol. Rev.* 85:635–678.
34. Chung S, Sonntag K C, Andersson T, et al. (2002). Genetic engineering of mouse embryonic stem cells by Nurr1 enhances differentiation and maturation into dopaminergic neurons. *Eur. J. Neurosci.* 16:1829–1838.
35. Hegert C, Kramer J, Hargus G, et al. (2002). Differentiation plasticity of chondrocytes derived from mouse embryonic stem cells. *J. Cell Sci.* 115:4617–4628.
36. Lavon N, Yanuka O, Benvenisty N (2004). Differentiation and isolation of hepatic-like cells from human embryonic stem cells. *Different.* 72:230–238.
37. Lumelsky N, Blondel O, Laeng P, et al. (2001). Differentiation of embryonic stem cells to insulin-secreting structures similar to pancreatic islets. *Science.* 292:1389–1394.
38. Wichterle H, Lieberam I, Porter J A, et al. (2002). Directed differentiation of embryonic stem cells into motor neurons. *Cell.* 110:385–397.
39. Yamane T, Kunisada T, Yamazaki H, et al. (1997). Development of osteoclasts from embryonic stem cells through a pathway that is c-fms but not c-kit dependent. *Blood.* 90:3516–3523.
40. Yamamoto H, Quinn G, Asari A, et al. (2003). Differentiation of embryonic stem cells into hepatocytes: Biological functions and therapeutic application. *Hepatol.* 37:983–993.
41. Kolossov E, Fleischmann B K, Liu Q, et al. (1998). Functional characteristics of ES cell-derived cardiac precursor cells identified by tissue-specific expression of the green fluorescent protein. *J. Cell Biol.* 143:2045–2056.
42. Li M, Pevny L, Lovell-Badge R, et al. (1998). Generation of purified neural precursors from embryonic stem cells by lineage selection. *Curr. Biol.* 8:971–974.
43. Blyszczuk P, Czyz J, Kania G, et al. (2003). Expression of Pax4 in embryonic stem cells promotes differentiation of nestin-positive progenitor and insulin-producing cells. *Proc. Natl. Acad. Sci. USA.* 100:998–1003.
44. Yamada T, Yoshikawa M, Kanda S, et al. (2002). In vitro differentiation of embryonic stem cells into hepatocyte-like cells identified by cellular uptake of indocyanine green. *Stem Cells.* 20:146–154.

45. Yamamoto Y, Teratani T, Yamamoto H, et al. (2005). Recapitulation of in vivo gene expression during hepatic differentiation from murine embryonic stem cells. *Hepatol.* 42:558–567.
46. Chinzei R, Tanaka Y, Shimizu-Saito K, et al. (2002). Embryoid-body cells derived from a mouse embryonic stem cell line show differentiation into functional hepatocytes. *Hepatol.* 36:22–29.
47. Heins N, Englund M C, Sjoblom C, et al. (2004). Derivation, characterization, and differentiation of human embryonic stem cells. *Stem Cells.* 22:367–376.
48. Soria B, Roche E, Berna G, et al. (2000). Insulin-secreting cells derived from embryonic stem cells normalize glycemia in streptozotocin-induced diabetic mice. *Diabetes.* 49:157–162.
49. Assady S, Maor G, Amit M, et al. (2001). Insulin production by human embryonic stem cells. *Diabetes.* 50:1691–1697.
50. Kahan B W, Jacobson L M, Hullett D A, et al. (2003). Pancreatic precursors and differentiated islet cell types from murine embryonic stem cells: an in vitro model to study islet differentiation. *Diabetes.* 52:2016–2024.
51. Miyazaki S, Yamato E, Miyazaki J (2004). Regulated expression of pdx-1 promotes in vitro differentiation of insulin-producing cells from embryonic stem cells. *Diabetes.* 53:1030–1037.
52. Rajagopal J, Anderson W J, Kume S, et al. (2003). Insulin staining of ES cell progeny from insulin uptake. *Science.* 299:363.
53. Ku H T, Zhang N, Kubo A, et al. (2004). Committing embryonic stem cells to early endocrine pancreas in vitro. *Stem Cells.* 22:1205–1217.
54. Bagutti C, Wobus A M, Fassler R, et al. (1996). Differentiation of embryonal stem cells into keratinocytes: Comparison of wild-type and beta 1 integrin-deficient cells. *Dev. Biol.* 179:184–196.
55. Coraux C, Hilmi C, Rouleau M, et al. (2003). Reconstituted skin from murine embryonic stem cells. *Curr. Biol.* 13:849–853.
56. Fraichard A, Chassande O, Bilbaut G, et al. (1995). In vitro differentiation of embryonic stem cells into glial cells and functional neurons. *J. Cell Sci.* 108(Pt 10):3181–3188.
57. Goridis C, Rohrer H (2002). Specification of catecholaminergic and serotonergic neurons. *Nat. Rev. Neurosci.* 3:531–541.
58. Ye W, Shimamura K, Rubenstein J L, et al. (1998). FGF and Shh signals control dopaminergic and serotonergic cell fate in the anterior neural plate. *Cell.* 93:755–766.
59. Studer L, Csete M, Lee S H, et al. (2000). Enhanced proliferation, survival, and dopaminergic differentiation of CNS precursors in lowered oxygen. *J. Neurosci.* 20:7377–7383.
60. Kim J H, Auerbach J M, Rodriguez-Gomez J A, et al. (2002). Dopamine neurons derived from embryonic stem cells function in an animal model of Parkinson's disease. *Nature.* 418:50–56.
61. Takagi Y, Takahashi J, Saiki H, et al. (2005). Dopaminergic neurons generated from monkey embryonic stem cells function in a Parkinson primate model. *J. Clin. Invest.* 115:102–109.
62. Rolletschek A, Chang H, Guan K, et al. (2001). Differentiation of embryonic stem cell-derived dopaminergic neurons is enhanced by survival-promoting factors. *Mech. Dev.* 105:93–104.
63. Bost F, Caron L, Marchetti I, et al. (2002). Retinoic acid activation of the ERK pathway is required for embryonic stem cell commitment into the adipocyte lineage. *Biochem. J.* 361:621–627.

64. Angelov D N, Arnhold S, Andressen C, et al. (1998). Temporospatial relationships between macroglia and microglia during in vitro differentiation of murine stem cells. *Dev. Neurosci.* 20:42–51.
65. Kramer J, Hegert C, Hargus G, et al. (2005). Mouse ES cell lines show a variable degree of chondrogenic differentiation in vitro. *Cell Biol. Int.* 29:139–146.
66. Bourne S, Polak J M, Hughes S P, et al. (2004). Osteogenic differentiation of mouse embryonic stem cells: differential gene expression analysis by cDNA microarray and purification of osteoblasts by cadherin-11 magnetically activated cell sorting. *Tissue Eng.* 10:796–806.
67. Rohwedel J, Maltsev V, Bober E, et al. (1994). Muscle cell differentiation of embryonic stem cells reflects myogenesis in vivo: developmentally regulated expression of myogenic determination genes and functional expression of ionic currents. *Dev. Biol.* 164:87–101.
68. Kehat I, Kenyagin-Karsenti D, Snir M, et al. (2001). Human embryonic stem cells can differentiate into myocytes with structural and functional properties of cardiomyocytes. *J. Clin. Invest.* 108:407–414.
69. He J Q, Ma Y, Lee Y, et al. (2003). Human embryonic stem cells develop into multiple types of cardiac myocytes: Action potential characterization. *Circ. Res.* 93:32–39.
70. Klug M G, Soonpaa M H, Field L J (1995). DNA synthesis and multinucleation in embryonic stem cell-derived cardiomyocytes. *Am. J. Physiol.* 269:H1913–1921.
71. Snir M, Kehat I, Gepstein A, et al. (2003). Assessment of the ultrastructural and proliferative properties of human embryonic stem cell-derived cardiomyocytes. *Am. J. Physiol. Heart Circ. Physiol.* 285:H2355–2363.
72. Kehat I, Khimovich L, Caspi O, et al. (2004). Electromechanical integration of cardiomyocytes derived from human embryonic stem cells. *Nat. Biotechnol.* 22:1282–1289.
73. Xue T, Cho H C, Akar F G, et al. (2005). Functional integration of electrically active cardiac derivatives from genetically engineered human embryonic stem cells with quiescent recipient ventricular cardiomyocytes: insights into the development of cell-based pacemakers. *Circul.* 111:11–20.
74. Lako M, Lindsay S, Lincoln J, et al. (2001). Characterisation of Wnt gene expression during the differentiation of murine embryonic stem cells in vitro: Role of Wnt3 in enhancing haematopoietic differentiation. *Mech. Dev.* 103:49–59.
75. Abuin A, Holt K H, Platt K A, et al. (2002). Full-speed mammalian genetics: In vivo target validation in the drug discovery process. *Trends Biotechnol.* 20:36–42.
76. Grilli M, Diodato E, Lozza G, et al. (2000). Presenilin-1 regulates the neuronal threshold to excitotoxicity both physiologically and pathologically. *Proc. Natl. Acad. Sci. USA.* 97:12822–12827.
77. Ishikawa H, Ryseck R P, Bravo R (1996). Characterization of ES cells deficient for the p105 precursor (NF-kappa B1): Role of p50 NLS. *Oncogene.* 13:255–263.
78. Gowen L C, Johnson B L, Latour A M, et al. (1996). Brca1 deficiency results in early embryonic lethality characterized by neuroepithelial abnormalities. *Nat. Genet.* 12:191–194.
79. Minamino T, Yujiri T, Papst P J, et al. (1999). MEKK1 suppresses oxidative stress-induced apoptosis of embryonic stem cell-derived cardiac myocytes. *Proc. Natl. Acad. Sci. USA.* 96:15127–15132.
80. Zhang X, Morham S G, Langenbach R, et al. (2000). Lack of cyclooxygenase-2 inhibits growth of teratocarcinomas in mice. *Exp. Cell Res.* 254:232–240.
81. Klemm M, Genschow E, Pohl I, et al. (2001). Permanent embryonic germ cell lines of BALB/cJ mice—an in vitro alternative for in vivo germ cell mutagenicity tests. *Toxicol. In Vitro.* 15:447–453.

82. Seiler A, Visan A, Buesen R, et al. (2004). Improvement of an in vitro stem cell assay for developmental toxicity: The use of molecular endpoints in the embryonic stem cell test. *Reprod. Toxicol.* 18:231–240.
83. Bremer S, Worth A P, Paparella M, et al. (2001). Establishment of an in vitro reporter gene assay for developmental cardiac toxicity. *Toxicol. In Vitro.* 15:215–223.
84. Paparella M, Kolossov E, Fleischmann B K, et al. (2002). The use of quantitative image analysis in the assessment of in vitro embryotoxicity endpoints based on a novel embryonic stem cell clone with endoderm-related GFP expression. *Toxicol. In Vitro.* 16:589–597.
85. Rideout W M, 3rd, Hochedlinger K, Kyba M, et al. (2002). Correction of a genetic defect by nuclear transplantation and combined cell and gene therapy. *Cell.* 109:17–27.
86. Check E (2003). Second cancer case halts gene-therapy trials. *Nature.* 421:305.

## 3 ADULT SOMATIC STEM CELLS OR POSTNATAL STEM CELLS

### 3.1 Introduction

In the past few decades, a true revolution has been occurring in a spectacular way in the field of medicine and biology. Regenerative medicine with the intention of tissue regeneration, and thus the curative treatment of diseases, has awoken maximum interest from scientific communities all over the world [1].

The likelihood that the human body contains the cells capable of regenerating and repairing the damaged or disease tissues has turned from an implausible subject to a virtual belief. It has been well known that stem cells with differentiating potential to replenish progeny are present in postnatal tissues of mammals. Recently many studies have demonstrated the abilities of stem cells to form multiple types of cells and the presence of such cells in an increasing number of tissues. We attempt here to provide an overview of adult stem cells in terms of possible mechanisms of their differentiation and their potential in therapeutic use.

The presence of adult somatic stem cell or cell progenitor has been clearly identified in several tissues. It has primarily been easily identified in tissues with high cell replication. These cells contribute to repair and regeneration throughout the life span in adult tissues, which also have the capacity of cell renewal [2, 3]. Recently, the concept of stem cell niches was proposed, in which stem cells exist in small numbers in silence under normal conditions but become activated after tissue injury or other pathological conditions [4]. Some stem cells possess enormous plasticity to differentiate into any cell types derived from the germinal layers (mesoderm, ectoderm, and endoderm). In contrast, other stem cells can only terminally differentiate into the same germinal layer from which they originated. Several factors, which are not clearly understood, are involved in the regulation of stem cells in terms of their potential of differentiation, tissue repair, and regeneration. On the other hand, the tragedy includes the development and progression of malignancy due to their longevity with respect to the body life span and their ability to give rise to multiple cell phenotypes [5]. Based on the nature of stem

cells, there is a worldwide interest in stem cell research to understand the natural growth and senescence of the cells and to balance tissue repair and regeneration through the life span. Studies on the mechanism that controls the maintenance of stem cells and differentiation signals is useful to understand how external (toxins, radiation, etc.) and internal injuries (cytogenetic alterations, mutation, etc.) potentially modify the fate of stem cells and eventually lead them to malignant transformation [5–7]. Stem cells can withstand over stressful environmental events associated with tissue damage after surgical procedures [8], exposure to toxics agents [8, 9], and extreme cold [10] and then repopulate and repair adult tissues. Although such *in vivo* environments seem to be transient and hostile against host cells and mature cells are easily destroyed, stem cells can survive, differentiate, and regenerate the tissue [8]. Their differentiation capacities investigated under a variable controlled environment may allow us a window into the regenerative process of adult tissues. To facilitate stem cell research, *in vitro* culture should mimic native niches of stem cells *in vivo* by application of matrices, growth factors, and drug delivery systems [11–15].

***3.1.1 General Concepts of Stem Cells*** The name *stem cells* is commonly used to refer the cells that are relatively undifferentiated while retaining the ability to divide and proliferate throughout postnatal life, providing progenitor cells that can differentiate into tissue-specific cells. The ability of stem or progenitor cells to give rise to different populations of terminally differentiated cells is referred to as plasticity. Thus, their potential is called totipotent, pluripotent, or multipotent. The term *totipotent* should be strictly reserved for the unique stem cells that can form embryonic and extra-embryonic membranes [6, 16], which are the most primitive about 4 days after fecundation [17]. The term *pluripotent* is used to refer to the stem cells' ability to form all cell types of the proper embryo, except for the extra-embryonic membranes and their tissues. In contrast, the term *multipotent* is used to name those that give rise to a subset of cell lineages. During the development process, the totipotent stem cells originate and pluripotent stem cells come out of the germ layers. Afterward multipotent stem cells are generated [16]. These cells, in turn, form the oligopotent progenitor cells in the developing organs. Moreover, intestinal progenitor cells are considered to be quadripotent; they can form progeny that become mucous, absorptive, neuroendocrine, or Paneth cells. In the case of bronchial lining cells, the progenitors are tripotent: Progeny turn into neuroendocrine, mucous, or ciliated cells. The oval cells of the liver are bipotent; they can derivate into duct cells and hepatocytes. Other type of cells, such as epidermal progenitors, are unipotent to produce only a single progeny [6, 16, 17].

*Tissue Homeostasis and Stem Cell Renewal.* Physiological tissue renewal is accomplished by a delicate balance among specialized cells, young progeny, and their tissue-specific stem cells. They perpetually renew, which is evident in many tissues, such as, blood, skin, gastrointestinal tract, respiratory tract, and testis. In contrast, other tissues, including cardiac and neural tissues, seem to have a limited response under regenerative circumstances. Thus, it was considered that cardiac and neural tissues had no regenerative capacities [16, 18]. Asynchronous division is a typical *in vivo* pattern of stem cells; it is defined by the division of one stem

cell to give rise to one daughter cell that remains as a stem cell while another undergoes the process of differentiation. The transiently amplifying cells provide an expanded population that differentiates into more mature cells, which can no longer proliferate and eventually die [3, 6, 17]. Such phenomenon is observed in most tissues, however, others, such as the liver [19] and pancreas [20], are mainly driven by replication of mature cells rather than stem cell division and differentiation under normal conditions. Occasionally, in those tissues under toxic DNA-damage conditions when replication of mature cells is compromised, facultative stem cells accomplish the tissue regeneration [19].

*Stem Cells and Their Niches.* The microenvironment or the stem cell niche tightly regulates the behavior of the cells. A combination of cells and extracellular matrix components, soluble factors delivered to the tissue from the vasculature, and the growth factors produced by the cells govern all aspects of the behavior of stem cells. For example, in the intestinal mucosa, the pericryptal myofibroblasts that surround the crypts may serve as niche cells, whereas in the hair follicles, the region just below the sebaceous glands seems to be a stem cell niche [17, 18]. The niches themselves control many dynamic facets of the stem cells, intrinsically regulating the internal signaling, synthesis of structural and metabolic proteins, their mitotic activities, axes, and growth pattern. These events certainly play a pivotal role in leading intrinsic and extrinsic factors that will define the physiological function of stem cells or the first step in the pathological transformation of stem cells in carcinogenesis [17, 18].

***3.1.2 General Classification of Stem Cells*** Because adult somatic stem cells consist of a different population of the cells that share some common characteristics, it is difficult to rigorously divide the cells into some classifications. A new theory has currently proposed that stem cells are generated from a single cell source [21–24]. Thus, we review here adult stem cells according to the commonest classification as follows:

1. Hematopoietic stem cells (HSCs)
2. Mesenchymal stem cells (MSCs)
3. Multipotent adult progenitor cells (MAPCs) isolated by fluorescent-activated cell sorting (FACS) from the bone marrow
4. Side-population phenotype cells (SPs)
5. Tissue-specific cell progenitors (TSCPs)
6. Umbilical cord blood-derived stem cells (UCBDSs)

*Bone Marrow Stem Cells.* The bone marrow has been considered to provide an adequate microenvironment for stem cells to survive forever. Stem cells in the bone marrow can migrate into blood vessels, circulate around the body, and return home when needed [25]. A long time ago, only a few stem cells were recognized in humans and they were thought to be a restricted population with a limited potential to differentiate in a single organ system. Such reasoning was examined through the study of bone marrow, which contains a wide range of cell populations. The complex constitution of bone marrow also attracted the interest of scientists. Such investigations changed the original thought to a new one, which implies that stem

cells also have the ability to give rise to the cells of other tissue types as well as to the cells of the original, referred to as the "plasticity of stem cells." Bone marrow has been described to contain at least two [26, 27] different types of stem cells. One type includes HSCs, which produce the entire progeny of blood cells of the body, and the other includes MSCs, which are a promising source for tissue repair [28]. It is also believed that another population exists, called MAPCs [27], which could be generated from the adult bone marrow of several species, and can differentiate into multiple cell types *in vitro* [29, 30] (Figure 3-1).

***3.1.3 General Characteristics of Stem Cells*** Stem cells possess a wide range of different characteristics, which have made them an attractive cell source for cell biology, ontogeny, toxicological studies, and cell therapy [12–16, 18, 24, 31]. Several types of stem cells have been reported, and each possesses particular characteristics that determine their potential (Table 3-1).

### *3.1.4 Hematopoietic Stem Cells (HSCs)*

*Characteristics.* For the past three decades, HSCs have been considered important cells. In 1963, the origin of hematopoietic cells was first reported [33]. One of the main characteristics of HSCs is the capacity to give rise to intermediate precursor

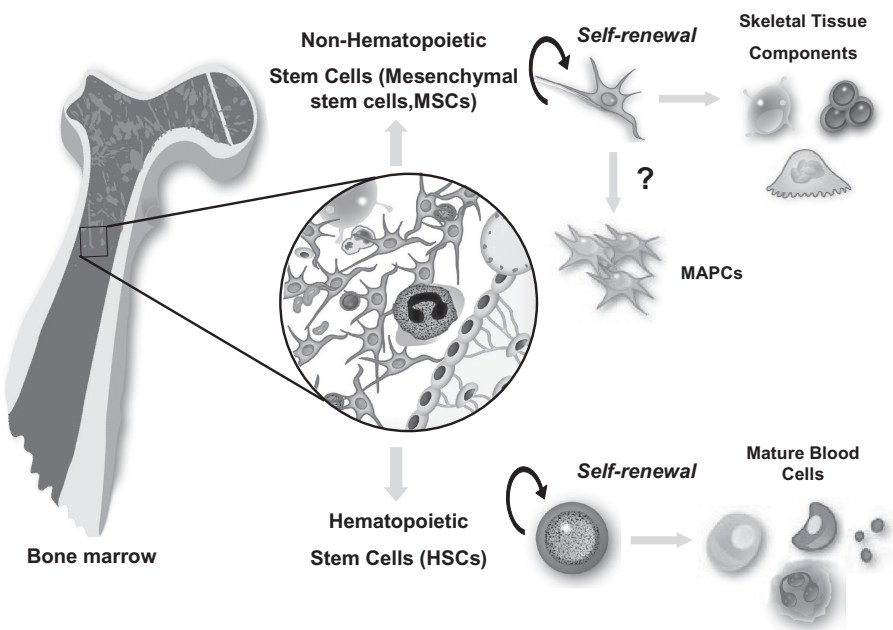


**Figure 3-1.** ***Schematic model of bone marrow niche.*** Bone marrow (BM) harbors a heterogeneous population of progenitor cells, which can generate a diversity of different cells. BM is composed of the osteoblastic zone and the vascular zone. In the vascular zone, HSCs exist that are capable of giving rise to all hematopoietic progeny, in direct contact with MSCs, which are known as progenitors for skeletal tissue components such as bone, cartilage, hematopoietic-suporting stroma, and adipose tissue. (This figure is available in full color at ftp://ftp.wiley.com/public/sci_tech_med/pharmaceutical_biotech/.)

**TABLE 3-1. Stem Cell Types and Their General Characteristics**

| | Tissue-specific cell progenitors[1] | Pluripotent and multipotent stem cells[2] | Embryonic stem cells[3] |
|---|---|---|---|
| Cell commitment | Lineage-commitment* | Uncommitment | Uncommitment |
| Cell quiescence under neither serum supplemented medium nor analogous knockout replacement | Quiescence | Quiescence | Non-quiescence |
| Telomerase activity | Absent | Most present but variable activity | High activity |
| Cell life span | Hayflick's limited | Extended | Unlimited |
| Cell growth at confluence | Inhibited by cell contact | Inhibited by cell contact | Non-contact inhibition |
| Inducible plasticity | Limited* | Usually unlimited | Unlimited |
| *In vitro* preservation of undifferentiated stage | Possible but limited | Possible but limited | Certainly possible |
| Cell membrane markers or antigens | Often lineage-commitment* | Various according to their cell phenotypes | Alkaline phosphatase SSEA-1,-3,-4 Oct-3/4 |

[1] Tissue-specific cell progenitors are commonly quadri-, tri-, or bipotent; therefore, even though they give rise to progeny with different phenotypes, the fate of the progenitor cells is lineage-specific. Each progenitor cell has a unique profile on the cell surface markers, but the profile is similar to that of both the tissue and the progeny that they will originate [6, 16, 32].

[2] Pluripotent and multipotent stem cells are heterogeneous population. These cells are practically the entire populations of postnatal stem cells [16]. They include HSCs, MSCs, SPs, MAPCs, and UCBDSs. Although they have different phenotypes, some of their characteristics are frequently shared.

[3] Embryonic stem cells are the prototype of all stem cells. Characteristics are well-defined because ES cell lines were established.

* Even though these populations are localized in adult tissues having lineage-commitment and exhibiting limited plasticity and local cell markers, some of these cells are capable of de-differentiating and then becoming multipotent stem cells.

or progenitor cell populations that partially differentiate and commit to various types of blood cell lineages [34, 35]. HSCs need to possess the hallmark properties to equilibrate cell self-renewal, whereas the cells quickly generate progenitors as a workforce as well as additional stem cells without depleting the reserves. Therefore, HSCs need to be multipotent; that is, a single HSC can produce several different lineages of mature blood cells and proliferate to yield a broad number of mature progeny. HSCs in the bone marrow of the mice are a rare population with a frequency of 1 in 10,000 to 100,000 of total blood cells, and the cells may be even less in humans [36–38]. HSCs are thought to be relatively in-active in the adult

hematopoietic system, with 1–3% in the progression of the cell cycle and approximately 90% in the cell cycle G0. The cells divide only a few or not at all until they are required to differentiate [39–42]. Abnormal clonal expansion of HSCs may result in chronic myelogenous leukemia, polycythemia vera, and myelodysplastic syndromes. HSCs are believed to have the ability to live for a long period of time, probably a lifetime in the recipient after bone marrow transplantation. In fact, HSCs require the bone marrow microenvironment, which regulates their migration, proliferation, and differentiation, to maintain active hematopoiesis throughout their lifetime [43, 44].

*Isolation and Phenotypic Characteristics of HSCs.* Recent experiments have demonstrated that HSCs in bone marrow from many different species can be purified by FACS as SP cells [40, 45, 46]. In 1994, multipotent cells committed to the hematopoietic lineage in mouse according to their different array of cell-surface markers were isolated [39]. This cells were described as KTLS c-kit$^{+}$ (K), Thy-1.1$^{low}$ (a marker on stem cells) (T), Lin$^{-/low}$ (Lineage-marker) (L), and Sca-1$^{+}$ (Stem cell antigen-1) (S) [1]. The cells showed >80% for hematopoietic multilineage differentiation and represented only 1/2000 cells in the bone marrow. HSCs were proposed to be divided into three compartments based on both the expression of the surface markers and their self-renewal ability [39, 47].

1) Long-term HSCs (LT–HSCs): LT–HSCs habitually reside in the bone marrow and have for all intents and purposes six developmental alternatives: remain quiescent, differentiate, self-renew, migrate, enter senescence, or undergo programmed cell death. The cells represent only 0.007% of cells in the bone marrow. In young adult mice, approximately 8% of LT–HSCs arbitrarily enters into cell division per day [42], and half of their progeny are LT–HSCs to maintain the level of steady state [48]. LT–HSCs perform self-renewal perpetually without depleting the pool of stem cells, and the cells are in charge of producing proliferative short-term HSCs. LT–HSCs express the surface markers of Thy1.1$^{low}$Flk-2$^{-}$. The expression of Flk-2 is upregulated and the expression of Thy-1.1 is downregulated as self-renewal capacity of the cells diminishes [49].
2) Short-term HSCs (ST–HSCs): ST–HSCs engender the lineage-committed progenitors to produce the billions of differentiated hematopoietic cells in the peripheral blood daily. The self-renewal life span of ST–HSCs is 6–8 weeks, and afterward, the cells fade away from the bone marrow. ST–HSCs represent 0.01% of the cells in the bone marrow of young adult C57BL mice and have the phenotype of thy1.1$^{low}$ Flk-2$^{+}$.
3) Multipotent progenitor cells (MPs): MPs have restricted self-renewal potential for less than 2 weeks. The expression of thy1.1$^{-}$Flk-2$^{+}$ is also identified in MPs [49]. The offspring of HSCs have been characterized and lineage restricted oligopotent progenitor cells for *lymphoid*, common lymphoid progenitor (CLP), and *myeloid*, common myeloid progenitor (CMP), granulocyte-monocyte progenitor (GMP), and megakaryocyte-erythrocyte progenitor (MEP) lineages (Figure 3-2).

*Comparable Phenotypic Markers of HSCs in Human and Mouse.* CD34 was the first marker found in human hematopoietic progenitors [50]. Most human HSCs express CD 34, which is also expressed in the committed progenitors [51] and

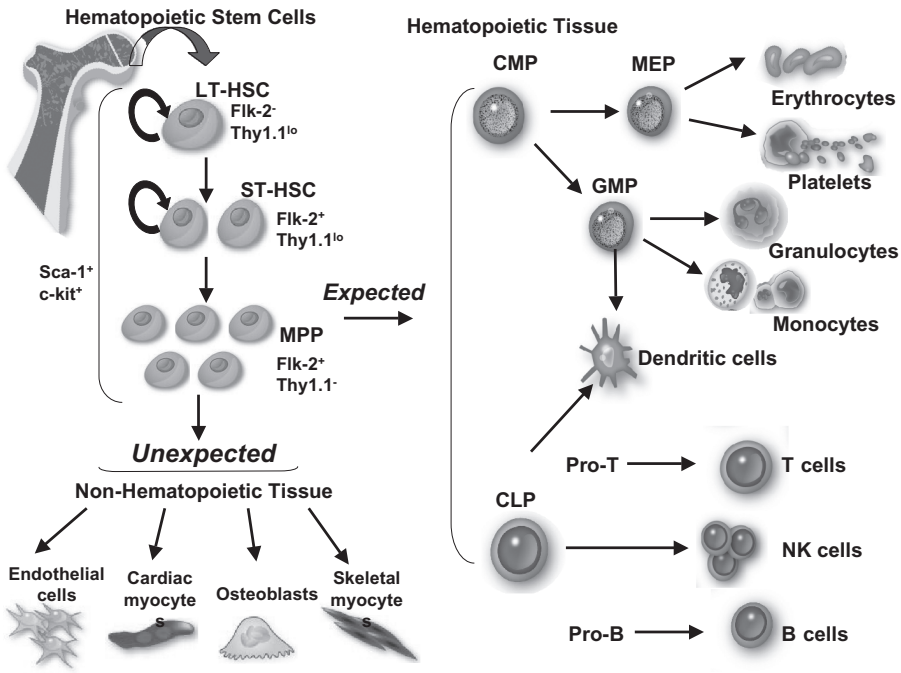


**Figure 3-2.** ***Model of hematopoietic stem cell lineage.*** Long-term hematopoietic stem cells (LT–HSCs) have the ability of unlimited self-renewal and of generating other precursors that give rise to population of the hematopoietic tissue cells (expected), including mature blood cells, CMPs (common myeloid progenitors), CLPs (common lymphoid progenitors), MEPs (megakaryocytes/erythrocyte progenitors), and GMPs (granulocytes/monocyte progenitosr). HSCs also generate a population of the non-hematopoietic tissue cells (unexpected), including ostoeblasts, skeletal myocytes, cardiac myocytes, and endothelial cells. (This figure is available in full color at ftp://ftp.wiley.com/public/sci_tech_med/pharmaceutical_biotech/.)

nonhematopoietic progenitors [52, 53]. The characteristic phenotype of human HSCs includes the lack of expression of lineage markers (Lin-) and expression of Thy-1, c-Kit, and Sca-1, CD45 [54] without CD38 expression [51, 55, 56]. Nonetheless, some human HSCs can be found in the fraction of CD34-negative population [57, 58]. CD34- negative HSCs are a precursor fraction of CD34-positive HSCs. Human HSCs also express Bcrp, known also as ABCG2 transporter, which outflows particular molecules as Hoechst-33342 staining [40, 59]. Detection of CD34 expression in HSC could be performed by the use of HCC-1 antibody in CD59 family members [60] (the sca-1 [61] antibody for mouse detects CD59 family members). Recently CD133 has been identified and could be used as another marker for human HSCs instead of CD34 [62]. It is important to remember that the $CD34^+$ bone marrow population denotes 1–6% of the cells in bone marrow, whereas the HSC compartment corresponds to only 0.05% [38]. Hence the $CD34^+$ population includes HSCs and a small fraction of the non-HSC cell population. Most quiescent LT–HSCs are $CD34^+Thy^+Lin^-CD38$ [63] (Figure 3-3).

*Mechanism of Activation Cycle of HSCs.* During the late phase of embryonic development, HSCs are located in the fetal liver, where they undergo massive

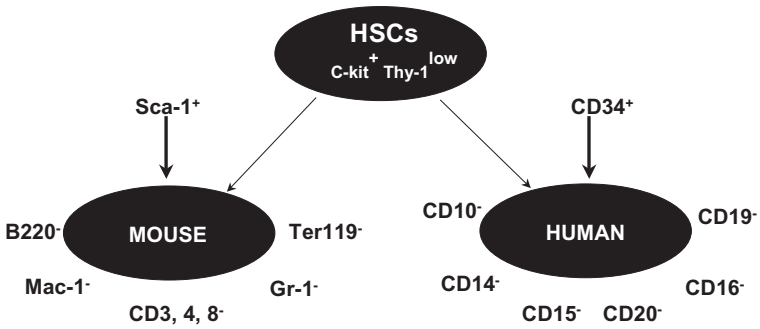


**Figure 3-3.** ***Marker profiles of HSCs in human and mouse.*** Human HSCs are positive for c-kit$^+$, Thy-1$^{low}$, and CD34$^+$, and mouse HSCs are positive for c-kit$^+$, Thy-1$^{low}$, and Sca-1$^+$. (This figure is available in full color at ftp://ftp.wiley.com/public/sci_tech_med/pharmaceutical_biotech/.)

expansion before they enter into the bone marrow and express AA4.1 and Mac 1, which are usually absent in HSCs in a quiescent state [64]. It has been generally believed that the microenvironment affects the fate of the cells. When HSCs are unperturbed in their quiescence niche, the cells express receptors associated with metabolism and aging (IGF1R) and show the activity of tyrosine kinase Tie1, which allows the cells to respond to multiple mitogenic signals. HSCs also express high levels of transcription factors, such as c-fos and GATA-2, which enable quick activation of the cells. HSCs are ready to act in response to changes in their environment "state of readiness." Immediately encountering the stress, HSCs pause totally by remaining quiescent in their niche, while they prepare to proliferate. This phase is mediated by upregulation of TIMP and serine proteinase inhibitor A-3g and by antiproliferative genes, such as Tob1, p21, and Btg3. Interferon-induced genes are also upregulated, indicating that HSCs are responding to pro-inflammatory signals caused by the stress. The signals induce a proliferative status, which is divided into early and late proliferation phases. In the early phase, the expression of genes involved in the regulation of replication and repair of DNA is enhanced in HSCs [65]. In the late phase when most of HSCs are in cycle, other genes related to energy production are expressed, indicating an increase in the metabolic activity of HSCs. In this context, two proposed approaches are associated with the proliferation state, mobilization [66] and migration [67]. HSCs need to move out of their quiescence niches and enter into a proliferative zone. Downregulation of $\alpha$-4 integrin is necessary for proliferation of HSCs, and downregulation of c-kit has been linked to mobilization of HSCs [68]. At the final phase, HSCs are required to return to their niches in order to stay at the initial state of bone marrow. This step usually starts the day when the damage is ameliorated, when several cells in cycle decrease, and is related with the expression of antiproliferative genes [69, 70].

*Plasticity of HSCs.* The hematopoietic tissues such as the bone marrow and peripheral blood have heterogeneous stem cell populations, including hematopoietic stem cells, mesenchymal stem cells, multipotent adult progenitor cells, and endothelial precursor cells. HSCs have been widely used to study the plasticity of adult stem cells [23]. Stem cell plasticity can be defined as a unique property of tissue-specific adult stem cells. Approximately more than 80% of studies reporting

plasticity of adult stem cells have been performed using the cells derived from the bone marrow or mobilized peripheral stem cells [23, 71, 72]. Although a previous concept was that differentiation of HSCs was restricted only to a hematopoietic lineage, current studies have reported that HSCs derived from bone marrow can give rise to hematopoetic precursors and multiple "unexpected" cell types, such as neural cells [73–75], hepatic cells [54, 76–78], cardiac muscle cells [79–81], and skeletal muscle cells [82, 83]. Many attempts have been proposed to define the mechanism for how stem cell plasticity occurs. Despite these efforts, there is still no clear evidence regarding whether stem cell plasticity really exists. Some investigators have suggested "trans-differentiation" as a possible cause of stem cell plasticity [84, 85], and others have proposed the cell fusion effect [86]. Even though many theories about plasticity have been proposed, nobody knows for certain whether stem cell plasticity is common or infrequent, whether it might be a vestige of the potential expression during embryonic development, or whether it has a physiologic role in the repair and homeostasis of the tissues. The plasticity of adult HSCs, even though it is rare, has suggested that the environment can reprogram the fate of the cells.

*Potential Implications of HSCs.* Studies on bone marrow cells differentiating into nonhematopoietic lineages have proposed that the cells are a representa-tive source of cell "replacement" in the treatment of numerous diseases. To amplify and generalize this issue for clinical translation, we should accumulate the useful processes of regeneration of blood formation by HSCs [36], skin replacement by putative epidermal stem cells [87], and the benefits reported in treating patients with myocardial infarction with bone marrow cells, mobilized peripheral HSCs, or hematopoietic progenitors [88, 89].

### *3.1.5 Mesenchymal Stem Cells (MSCs)*

*Characteristics.* MSC is a wide population that exists in the niche of the bone marrow and has attracted the attention of researchers by their potential for self-renewal and differentiation into functional cell types in the intrinsic tissue where they reside. These cells have also been isolated from the fat tissues [90]. MSCs can be expanded in culture for several passages without losing their differentiation potential and have been widely accepted as stem cells due to their usefulness in the clinical treatment of osteogenesis imperfecta [91], bone tissue regeneration [92], and hematopoietic recovery [93].

*Self-Renewal and Multilineage Differentiation Potential of MSCs.* Characterization of MSCs is subject of active investigation, because diversities of techniques to isolate the cells and different ways to analyze the self-renewal ability of the cells have been documented. Considering the lack of reliable specific markers and locating site of MSCs, further experiments are needed to address these issues. To identify MSCs, a combination of several monoclonal antibodies has been tested [94]. As of now, markers for MSCs include SH-2 [95], SH-3, SH-4 [96], SB-10 [97], CD29, CD44, CD90, and STRO-1 [98]. In addition, MSCs are negative to hematopoietic markers CD34 and CD45 and the cells are human leukocyte antigen (HLA) class I-positive and HLA class II-negative [99]. The first experiment

supporting the presence of MSCs was conducted in the early 1960s [100, 101]. Aspirated bone marrow cells were cultured at low density. The resultant cells formed colonies of fibroblasts, and they were responsible for osteogenesis [102, 103]. Murine MSCs showed self-renewal activity, and the cells differentiated into many types of cell lineages, such as osteoblasts and chondroblasts [104–109], adipocytes [110], neuronal progenitors [111], and myocytes [83] with the stimulation of cytokines, growth factors, and chemicals *in vitro*. However, it is controversial whether such plasticity of MSCs is induced *in vivo* even after minimal injury. To address this issue, researchers have used animal models of injury [112, 113]. Such experiments have shown that MSCs contribute to the repair or regeneration of damaged bone, cartilage, and infracted myocardial tissues [114–116].

*Isolation of MSCs.* Since 1980 when the characterization of human bone marrow fibroblast colony-forming cells (CFU-Fs) was performed, which was the gold standard to identify MSCs, many researchers have attempted to develop different methods for isolation of MSCs [117]. Now MSCs can be identified by their ability to adhere to a static surface and their proliferat-ing potential. Approximately 30% of human bone marrow-aspirated cells adhering to plastic culture dishes are considered to be MSCs [110]. MSCs in the bone marrow are a heterogeneous population that contains not only putative cells but also tripotent (ability to differentiate into osteocyte, chondrocyte, and adipose lineage-osteo/chondro/adipo), bipotent (osteo/chondro), or unipotent cells (osteo) [118–121] (Figure 3-4).

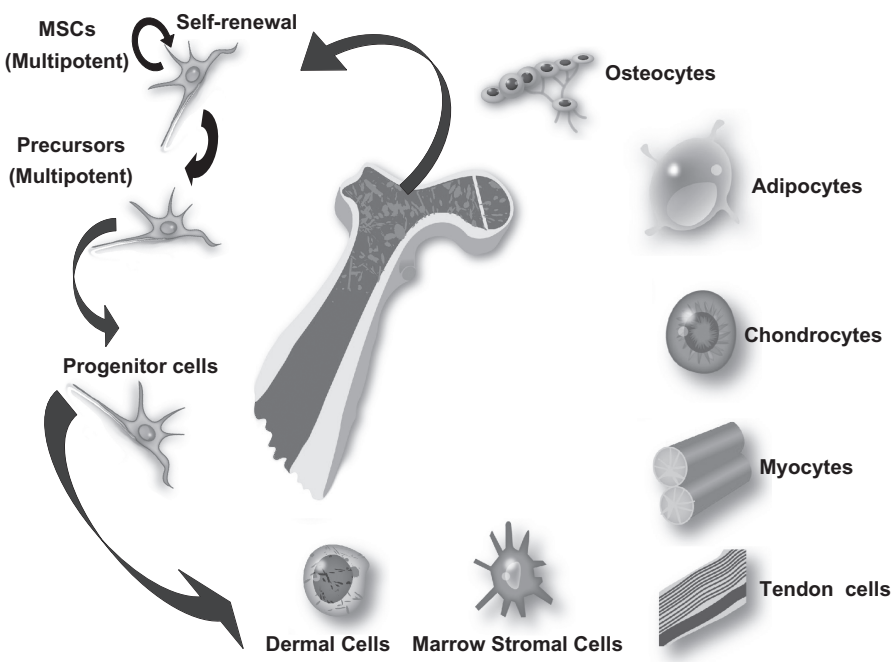


**Figure 3-4.** ***Multilineage differentiation of mesenchymal stem cells.*** Mesenchymal stem cells (MSCs) have the ability to differentiate into all types of cells of connective tissue. (This figure is available in full color at ftp://ftp.wiley.com/public/sci_tech_med/pharmaceutical_biotech/.)

*Future Prospectives of MSCs.* Adult MSCs have shown great promise in cell therapy in humans due to their multipotentiality, capacity for extensive self-renewal, and lack of induction of immune response. MSCs may be a valuable source to introduce foreign genes and could be a useful tool for gene therapy for bone regeneration and tissue engineering [122]. Researchers have also proposed MSCs as a useful means to protect the brain tissue from ischemic damage [123, 124]. MSCs ameliorate functional deficits after stroke in rats probably by releasing protective cytokines by the cells. Another potential clinical application is the use of MSCs as a drug delivery vehicle for the treatment of invasive malignant tumors [125]. The behavior of MSCs has been exploited as a tumor-targeting gene therapy in gliomas. Many studies regarding genetically modified MSCs have revealed extraordinary antitumor effects of the cells in experimental models of gliomas [126–127]. MSCs have the advantage of easy propagation *in vitro*. Moreover, implantation of autologous MSCs into patients with malignant gliomas is ethically nonproblematic.

***3.1.6 Multipotent Adult Progenitor Cells (MAPCs)*** MAPCs are bone marrow-derived stem cells with an extensive *in vitro* expansion ability, more than 80 population doublings, as well as a capacity to differentiate *in vivo* and *in vitro* into tissue cells of all three germinal layers; ectoderm, mesoderm, and endoderm. These cells have been isolated from three different species, including mouse, rat, and human [27].

*Isolation and Culture of MAPCs.* MAPCs of different species require different isolation and culture procedures. Human MAPCs are isolated after seeding bone marrow mononuclear cells (BMMNCs) of $CD45^-$ $GlyA^-$ at low densities ($1–3 \times 10^3/cm^2$) onto fibronectin-coated plastic dishes, with <2% FCS (fetal calf serum), EGF (epidermal growth factor), and PDGF-BB (platelet-derived growth factor). During the culture, depletion of the mono-nuclear cells is observed. The remnant clones that emerge over time are then harvested, and approximately 20% of the total cells seeded. The cells should be plated again at a density of $0.5–1 \times 10^3$ cells/$cm^2$. Maintenance of the stem cell phenotype is critically dependent on such low-density inoculation. The culture conditions of mouse MAPCs are similar to those described in human MAPC culture, except for the addition of the leukemia inhibitor factor (LIF).

*Phenotypic Characteristics and Proliferation Capacity of MAPCs.* Human MAPCs are characterized by the lack of hematopoietic markers such as CD34 and CD45 and the presence of the low expression of VCAM, CD44, MHC class I, and endoglin [29, 128]. The cells can be cultured *in vitro* for more than 50–80 population doublings, while preserving their normal karyotype. Mouse MAPCs are positive for CD34, CD45, CD44, c-Kit, CD3, Gr-1, Mac-1, and CD19 and negative for major histocompatibility complex (MHC) class I and class II. The cells express a lower level of Flk1 and Sca1 and a higher level of CD13 and SSEA-1 [129]. Mouse MAPCs express pluripotent markers of Oct-4, rex-1, and nanog, which were observed only in ES cells [129]. The cells can be cultured for more than 80–150 population doublings, but the karyotype of the cells becomes unstable.

*In Vitro Differentiation of MAPCs.* The differentiation of MAPCs to several lineages is achieved when the cells are re-plated at high density ($1–2 \times 10^4$ cells/$cm^2$). The cells were cultured in the absence of the serum, but with lineage-specific cytokines required for the differentiation of the desired tissues.

1) ***Mesoderm.*** By the use of VEFG, the differentiation of MAPCs has been demonstrated into functional mature endothelial cells expressing endothelial markers such as CD31, flk-1, and vWF. Such MAPC-derived endothelial cells are capable of contributing to neoangionesis and wound healing *in vivo* [29, 130].
2) ***Ectoderm.*** Verfaille et al. have shown the potential characterization of MAPCs to differentiate into functional neuroectodermal cells, including astrocytes, oligodendrocytes, and neurons [27, 131, 132].
3) ***Endoderm.*** MAPCs can differentiate into hepatocyte-like cells by using FGF4 and HGF. Such cells expressed hepatic markers of CK19, AFP, CK18, albumin, HepPar-1, and CD26 and produced albumin, urea, and glycogen [30].

*In vivo Differentiation of MAPCs.* The potential of multipotency of MAPCs has also been examined *in vivo* [27] by intravenous transplantation of the cells into postnatal murine recipients treated either without radiation or with sublethal irradiation. The observation of chimaerism in many somatic tissues of the mice derived from blastocysts demonstrated the engraftment of MAPCs in various tissues, including hematopoietic, lung, gut, and liver. In these tissues, MAPCs have acquired phenotypic characteristics of the respective cells, indicating their multipotenty.

*Future Prospectives of MAPCs.* MAPCs have been compared with ES cells, for their similar *in vitro* potential to give rise to tissue cells of all three germ layers. MAPCs have shown the same ability as ES cells *in vivo* experiments, although MAPCs do not develop tumors. Due to the nontumorigenecity, MAPCs can serve as a source for the production of a wide spectrum of transplantable cells with an enormous promise for the treatment of many degenerative or inherited diseases. Taking it into account that MAPCs can be recovered from the patients themselves, the clinical use of the cells would be advantageous, avoiding the need for immunosuppressive therapy. MAPCs provide a prospectively beneficial tool to study developmental biology of the stem cells as well as drug discovery.

***3.1.7 Side-Population Phenotype Cells (SPs)*** Purification of stem cells was reported based on an efflux of fluorescent dyes, such as Rhodamine 123 and Hoechst 33342 [133]. The purified cell population, referred to as SP, has been identified as a small fraction in the bone marrow in all species examined [45]. SP cells can yield about 0.05–0.1% from total bone marrow cells by the use of a flow cytometry cell sorting [40, 45, 59]. Despite being a small population, it possesses interesting capacities, including extensive proliferation and multipotential to generate the entire adult hematopoietic cell lineages [40]. Indeed, only a single transplantation of 200 SP cells can fully repopulate the bone marrow of lethally

irradiated mice [40, 134]. It has been reported that a successful expression of dystrophin is achieved in a mouse model of Duchenne's muscular dystrophy after transplantation of SP cells [82]. SP cells have contributed to regenerate the infarcted myocardium by developing cardiocytes and vascular endothelium cells [81]. SP cells have been identified in others several tissues, such as Posterior, by identifying the expression of the ABC transporter Bcrp1/ABCG2 [59]. Because of Hoechst 33342-related toxicity [46], other strategies have been proposed to recover SP cells by using monoclonal antibodies. Even though SP cells seem to be a small source for basic research and possible cell therapy, the cells have shown greater potential. Since SP cells were identified in the bone marrow, the cells have held great promise due to their plasticity and replicating capacity *in vivo* without any tumorigenic transformation [40, 45, 46, 81, 82]. *In vivo* differentiation of SP cells has succeeded in only mesoderm tissues (skeletal muscle, cardiocytes, and vascular cells) [81, 82]. Studies have shown that hematopoietic stem cells in the bone marrow are capable of trans-differentiating and giving rise to neuronal progenitor cells [73, 75, 135], but that such trans-differentiation has not occurred in SP cells [136].

*Tissue-Specific SP Phenotype Cells from Different Compartments.* For a period of time, the SP phenotype cells were identified in many other different tissues, including the brain [137, 138], pituitary gland [139], skin [140–143], corneal [144] and limbal epithelial ocular tissue [145], mammary glands [146], skeletal muscle [138, 143, 147, 148], lung [134, 138, 149–151], heart [138, 152, 153], liver [138, 154], spleen [138], pancreas [155], small intestine [138], kidney [138, 156–158], testis [159, 160], peripheral blood [138, 161], and in umbilical cord blood [45].

*Main Characteristics of SP Cells from Diverse Compartments.* A variable number of SP cells have been distributed in the different tissues of the body. The SP cells are a heterogeneous population that possesses the ability to give rise to their lineage-specific progenitor cells [138]. Some of these cells have been analyzed and found to be multipotent [139]. *In vitro* cultures of SP cells have shown differences in their propagation capacities and quiescent state [140, 145, 161].

### *3.1.8 Tissue-Specific Cell Progenitors (TSCPs)*

*Neural Cell Progenitor.* The cells involving the development of mature central nervous system (CNS) are precisely regulated by both temporal and local patterns of differentiation, which determine the appropriate cell function. The role of niches shows plasticity coordinated by both the intrinsic factors and the extrinsic soluble signals. Therefore, the specific neural differentiation is affected by this regional patterning. Neurogenesis involves the opposing soluble signals that determine dorsal and ventral patterning: Sonic hedgehog (Shh), bone morphogenetic protein (BMP) antagonists, chordin, and noggin are secreted from the floor plate, whereas other signals originating from the roof plate result in the creation of the gradient of signal concentrations [162, 163]. Precise concentration and ratio of each independent signal derive the development of the specific neuronal phenotypes at different points in accordance with these gradients. Such phenomena cause the different expression patterns of the cells and create the groups of neurons with different patterns of cell division and differentiation [163]. At the beginning,

the CNS stem cells divide symmetrically to enrich their population pool, or asymmetrically to generate more differentiated progeny, which become mature cells of neuronal and glial lineages. The spinal cord can produce both oligodendrocytes and neurons from common precursors, and the final decision of the cell fate depends on extrinsic signals and specific patterns of transcriptional activation [164].

*General Characteristics of Neural Progenitor Cells.* In rodent models, neural progenitor cells have been identified in specific regions of adult CNS. The cells isolated from different sites of CNS are not identical and thus show different growth characteristics, trophic factor requirements, and specific patterns of differentiation [36, 165]. Growth factor requirements can define at least two major types of neural progenitor cells isolated from CNS. One cell type requires a high dose of EGF and can be expanded as floating neurospheres, while preserving their phenotype after many passages, and eventually they come to be FGF-responsive [166, 167]. In contrast, another cell type isolated from CNS is exclusively FGF-dependent. Such cells can be propagated, as adherent cultures with FGF-dependent and the cells do not respond to EGF. The cells do not express EGF receptor either *in vivo* or *in vitro*. As these cells have been isolated from multiple brain areas, it is likely that both of neural precursor cells coexist. Both cells can self-renew and differentiate into the three neural cell types *in vitro*. Differentiation of the cells in tissue culture can be induced by withdrawal of growth factors or by induction of specific signals [166–168]. It seems that distinct stem cell populations are programmed to differentiate into specific cell types during the CNS development based on their response to growth factors. Moreover, some of these phenotypes expressed in progenitor cells in CNS have been observed in progenitors isolated from the pancreas of adult mice, which can generate neural and pancreatic lineages [169].

*Epidermal Progenitor Cells.* Keratinocytic populations can be grown *in vitro*, while displaying clonal growth. Keratinocytes can generate normal epidermis when they are appropriately grafted [170, 171]. Furthermore, the proliferative heterogeneity of the keratinocytes reveals the presence of stem cells, that is, epidermal progenitor cells. β1 integrin seems to play a central role in the regulation of the progenitor cells. When β1 integrin binds its ligand of type IV collagen, the cells exhibit high proliferative capacity, while maintaining their undifferentiated state. Therefore, the number and patterning of progenitor cells *in vivo* are autoregulated by the stem cell niche and turnover of the cells provides regulatory feedback signals [172]. The differentiating potential of epidermal progenitor cells has totally denied the previous speculation, implying that the cells were not multipotent [173–175].

*Progenitor Cells in the Respiratory Tract.* The respiratory tract seems to have a local progenitor population, which gives rise to mature respiratory cells and preserves the native architecture along the tract. Consequently, tissue homeostasis is maintained by these cells, in which the trachea and bronchus segments, bronchiole, and alveolus are maintained by mucous secreting basal cells, by Clara cells, and by type II pneumocytes, respectively. These cells serve as structural and functional parenchyma and work for quick regeneration if any injury occurs in the respiratory tract [176]. In addition, the recent reports have demonstrated that progenitors

existing in the populations of basal cells, Clara cells, and type II pneumocytes are not only able to provide tissue function, but also possess renewal capacities [177, 178]. These progenitor cell populations contain both SP cells and hematopoietic markers-expressing cells and are capable of repopulating the bone marrow as HSCs do. These findings have sustained the theory that bone marrow stem cells serve as the whole postnatal stem cells and their own progenies to provide a common interchangeable cell source [23, 177, 178].

*Skeletal Muscle Progenitor Cell or Satellite Cells.* The embryonic mesoderm generates skeletal musculature. Satellite cells are skeletal muscle progenitor cells that are responsible for postnatal growth and repair [179, 180]. Injury in skeletal musculature results in detriment in skeletal muscle mass, but injury itself triggers muscle regeneration of muscle tissue, which is accomplished by differentiation of native progenitor cells [180]. The tissue-specific cell progenitors are so-called satellite cells [179–181]. Despite the replenishment of the satellite cell pool during muscle growth, the number of satellite cells, which is highest in postnatal muscle, declines with aging [181]. This unique population of the cells exists between the muscle fiber sarcolemma and its covering basement membrane. Anatomically, the satellite cells can be clearly defined. The cells are quiescent lying outside the myofibers but beneath the basement membranes [179, 180]. Quiescent muscle satellite cells are characterized by the expression of surface markers such as M-cadherin, syndecan 3 and 4, and CD34 [180]. Even using a standard dissociation technique of the muscle tissue, an efficient isolation of myogenic progentior cells has not been achieved [179, 180]. Many growth factors are implicated in the regulation, chemotaxis, proliferation, and differentiation of satellite cells [179, 181]. Basic fibroblast growth factor (bFGF), platelet-derived growth factor (PDGF), transferring, and hepatocyte growth factor (HGF) have been identified as potent mitogens for satellite cells. HGF, bFGF, and IGF-1 and TGF-β can also promote chemotaxic activity of satellite cells in tissue culture [179, 182]. Based on their SP phenotype, Hoechst 33342 dye efflux-used FACS has become the easiest way to isolate skeletal muscle progenitor cells. This SP population can generate terminally differentiated skeletal muscle cells and completely repopulate the bone marrow after transplantation [179].

*Hepatic Cell Progenitors.* Hepatocytes possess the ability to rapidly respond to the parenchymal loss with turning on active mitosis. After massive loss of hepatocytes induced by two-thirds partial hepatectomy, the remaining hepatocytes make a cell-cycle progression at two to three times to restore preoperative cell volume. Such an enormous repopulating capacity is con-ducted by unknown potential and properties of hepatocytes [183]. These crucial properties that defines a stem cell are capable of generating a large progeny. This phenomenon can be observed at least in some hepatocytes [184–187]. Putative hepatic stem cells have been identified by the cell membrane markers, such as the lowered expression of the asialoglyco-protein receptor [188, 189]. Induction of a massive liver damage compromises the regeneration of hepatocytes, and then facultative hepatic progenitor cells distributed along the small branches of the intrahepatic biliary trees are activated to divide. Oval cells that express the same markers as HSCs (c-kit, flt-3, Thy-1, and CD34) transiently amplify the biliary population and then differentiate into hepatocytes

[19, 190, 191]. This particular phenomenon may provide information as to how the liver provides niches for facultative stem cells or offers an opportunity for cell fusion [19].

*Progenitor Cells in the Gastrointestinal Tract.* The presence of progenitor cells in the intestinal crypts has been observed [17]. Intestinal crypts and gastric glands are considered as specific niches that contain noncommitment and probably multipotent stem cells [17, 192–195]. The epithelial intestinal cells form the lining invaginates with numerous crypts and finger-shaped projections along the gastrointestinal tract and possess a rich population of neuroenteroendocrine cells disseminated throughout their epithelium. The plasticity of progenitor cells in the intestinal epithelium has been shown by differentiation of the cells to insulin-secreting cells using GLP-1 [196]. Intestinal crypts and gastric glands are enclosed by protective fenestrated intestinal subepithelial myofibroblasts and disperse in the lamina propria with merging blood vessels. The myofibroblasts secrete HGF, TGF-beta, and KGF to regulate the differentiation of the intestinal epithelial progenitor cells. Thus, myofibroblast cells play a central role in regulating the differentiation of the progenitor cells located in Lieberkühn crypts [197, 198].

*Adult Pancreatic Progenitor Cells.* The adult pancreatic tissue is composed of the exocrine and endocrine cells. During embryological development, the pancreatic tissue is generated from differentiation of the ductal epithelium. The exocrine tissue is first differentiated and then endocrine tissue is differentiated [199]. Islets of Langerhans under physiological conditions turn over continuously, but slowly. Terminally differentiated pancreatic beta cells have a life span of approximate 50 days [200, 201]. The cells have a balance of apoptosis and replacement, which is conducted by replication of differentiated beta cells [20]. The replicating capacity of the beta cells is limited [202]. The presence of pancreatic progenitor cells that can give rise to pancreatic and neuronal cell lineages has been demonstrated [169]. Ductal and acinar cells have been proven to have plasticity to differentiate into insulin-secreting cells or to trans-differentiate into hepatocytes [203–209]. The beta cells certainly replicate; however, the senescence of the cells still determines their potential [200, 202]. Otherwise, beta cells may be reprogrammed by cell fusion. *In vitro* culture human islets gradually lose their insulin expression, but they preserve the expression of PDX-1 [208, 210, 211]. Recent research has suggested that adult pancreatic islets contains progenitor cells [206]. Pancreatic progenitor cells may also exist in the non-endocrine tissue [208]. These progenitor cells are capable of generating terminally differentiated endocrine and exocrine progeny [155, 203–206, 212]. They are identified by the positive expression of nestin, referred as to nestin-positive progenitor cells (NIPs) [155, 206, 212], and are dispersed in the whole pancreatic tissue [213]. In addition, SP phenotype cells were observed in NIPs [155] and the cells can give rise to vascular and neuronal progenies [169, 214].

***3.1.9 Umbilical Cord Blood-Derived Stem Cells (UCBSCs)*** Stem cells were isolated from umbilical cord blood (UCB) and cultured *in vitro* [215]. Umbilical cord blood-derived stem cells (UCBSCs) had been proposed as an alternative source of regeneration of hemopoietic tissue by allogeneic transplantation [216].

The successful hematopoietic reconstitution in a patient with Fanconi's anemia has proved the potential of UCBSCs [217, 218]. Usually small cell fractions are recovered from the cord blood, but the primitive hematopoietic population possesses a high proliferative capacity. UCB cells contain a larger hematopoietic population of $CD34^-$ as well as a subset of $CD34^+$ and $CD38^-$, of which population ratio is about fourfold higher than that in bone marrow cells [219]. Moreover, the most primitive phenotype ($CD34^+CD38^-CD45RA^{low}CD71^{low}Thy\text{-}1^+c\text{-}kit^{low}Rh^{low}$)-expressing cells present in UCB at 0.003% of the entire population of nucleated cells [220]. Their proliferative capacity reach to a 50-fold expansion of colony forming cells [221, 222]. Successful *in vitro* expansion of the cells depends on the cytokines contained in the culture. Such cytokines include SCF, IL-1, IL-3, IL-6, GM-CSF, G-CSF, M-CSF, erythropoietin, and thrombopoietin [223]. The cells also possess multipotent characteristics and outstanding plasticity [220]. Therefore, there have been several approaches of differentiation of non-hematopoietic tissues ongoing in the clinical tissue regeneration settings [224–236] and *in vitro* toxicological studies [237].

***3.1.10 Heterogeneous Populations of Stem Cells in Bone Marrow*** Bone marrow contains two main well-defined populations of stem cells, HSCs and MSCs. They renew by themselves and give rise to all of the terminally differentiated blood lineages. In addition, bone marrow possesses a small subset population of non-hematopoietic stem cells (NHSCs). This population of NHSCs is identified by the expression of $CXCR4^+$, $CD34^+$, $AC133^+$, $lin^-$, and $CD45^-$ [22, 23]. Moreover, there is evidence that tissue damage in the body itself is capable of triggering the bone marrow stem cells to get into the circulation [238–242]. Subsequently, the stem cells modify their phenotype in accordance with the tissue where they migrate. Specific activation of membrane surface receptors and subsequent activation of transcription factors determine a modulation of the phenotypes and migratory properties of the stem cells. Chemotactic or homing signaling is involved in the regulation of the migration. The bone marrow stem cells have an exceptional ability to adapt and survive in different microenvironments. The cells can respond to several cytokines and growth factor, such as HGF, VEGF, LIF, and bFGF. The axis SDF-1-$CXCR4^+$ seems to be the most crucial regulatory factor, which is involved in honing of the stem cells. Therefore, it may be considered that bone marrow stem cells can serve as the source of all adult somatic stem cells [23, 238–242]. UCBSCs are considered as HSCs because of their strong proliferative capacity. UCBSCs certainly display higher potentials compared with adult stem cells [217, 220] and have been proposed as the best candidate for regenerating the damaged tissues in the clinical setting among the stem cells [231, 232].

***3.1.11 Uses and Applications of Stem Cells in Toxicology*** Besides the further clinical application of stem cells, utilization of the cells in the pharmacological field is also of great interest. As mentioned, stem cells are specialized cells found within many tissues of the body, in which they have the role to maintain homeostasis and repair of the damage tissue. The advantages of unlimited proliferation and subsequent differentiation of the stem cells can be of great use for *in vitro* screening of the newly developed drugs, allowing us to predict possible adverse effects without

the need to test them in animals. For example, hepatocytes derived from stem cells can be routinely screened for new drugs and chemicals to evaluate their liver toxicity.

## References


1. Lagasse E, Shizuru J A, Uchida N, et al. (2001). Toward regenerative medicine. *Immunity.* 14:425–436.
2. Punzel M, Ho A D (2001). Divisional history and pluripotency of human hematopoietic stem cells. *Ann. N Y Acad. Sci.* 938:72–81; discussion 81–82.
3. Punzel M, Liu D, Zhang T, et al. (2003). The symmetry of initial divisions of human hematopoietic progenitors is altered only by the cellular microenvironment. *Exp. Hematol.* 31:339–347.
4. Nagy P (1995). The facultative stem cell: A new star in liver pathology. *Pathol. Oncol. Res.* 1:23–26.
5. Almeida-Porada G, Porada C, Zanjani E D (2001). Adult stem cell plasticity and methods of detection. *Rev. Clin. Exp. Hematol.* 5:26–41.
6. Sell S (2004). Stem cell origin of cancer and differentiation therapy. *Crit. Rev. Oncol. Hematol.* 51:1–28.
7. Huntly B J, Gilliland D G (2005). Leukaemia stem cells and the evolution of cancer-stem-cell research. *Nat. Rev. Cancer.* 5:311–321.
8. Fausto N, Campbell J S (2003). The role of hepatocytes and oval cells in liver regeneration and repopulation. *Mech. Dev.* 120:117–130.
9. Allard E K, Boekelheide K (1996). Fate of Germ Cells in 2,5-Hexanedione-Induced Testicular Injury. *Toxicol. Appl. Pharmacol.* 137:149–156.
10. Miyagi K, Yamazaki T, Tsujino I (2001). Application of hypothermia to autologous stem cell purging. *Cryobiol.* 42:190–195.
11. Czyz J, Wobus A (2001). Embryonic stem cell differentiation: The role of extracellular factors. *Different.* 68:167–174.
12. Rohwedel J, Guan K, Hegert C, et al. (2001). Embryonic stem cells as an in vitro model for mutagenicity, cytotoxicity and embryotoxicity studies: present state and future prospects. *Toxicol. In Vitro.* 15:741–753.
13. Wobus A M, Guan K, Pich U (2001). In vitro differentiation of embryonic stem cells and analysis of cellular phenotypes. *Methods Mol. Biol.* 158:263–286.
14. Davila J C, Cezar G G, Thiede M, et al. (2004). Use and application of stem cells in toxicology. *Toxicol. Sci.* 79:214–223.
15. Wobus A M, Boheler K R (2005). Embryonic stem cells: Prospects for developmental biology and cell therapy. *Physiol. Rev.* 85:635–678.
16. Wagers A J, Weissman I L (2004). Plasticity of adult stem cells. *Cell.* 116:639–648.
17. Alison M R, Poulsom R, Forbes S, et al. (2002). An introduction to stem cells. *J. Pathol.* 197:419–423.
18. Blau H M, Brazelton T R, Weimann J M (2001). The evolving concept of a stem cell: entity or function? *Cell.* 105:829–841.
19. Grompe M (2005). The origin of hepatocytes. *Gastroenterol.* 128:2158–2160.
20. Dor Y, Brown J, Martinez O I, et al. (2004). Adult pancreatic beta-cells are formed by self-duplication rather than stem-cell differentiation. *Nature.* 429:41–46.
21. Kucia M, Ratajczak J, Ratajczak M Z (2005). Are bone marrow stem cells plastic or heterogenous—that is the question. *Exp. Hematol.* 33:613–623.

22. Kucia M, Ratajczak J, Ratajczak M Z (2005). Bone marrow as a source of circulating CXCR4+ tissue-committed stem cells. *Biol. Cell.* 97:133–146.
23. Kucia M, Reca R, Jala V R, et al. (2005). Bone marrow as a home of heterogenous populations of nonhematopoietic stem cells. *Leukemia.* 19:1118–1127.
24. Majka M, Kucia M, Ratajczak M Z (2005). Stem cell biology—a never ending quest for understanding. *Acta Biochim. Pol.* 52:353–358.
25. Lapidot T, Dar A, Kollet O (2005). How do stem cells find their way home? *Blood.* 106:1901–1910.
26. Mazurier F, Doedens M, Gan O I, et al. (2003). Rapid myeloerythroid repopulation after intrafemoral transplantation of NOD-SCID mice reveals a new class of human stem cells. *Nat. Med.* 9:959–963.
27. Jiang Y, Jahagirdar B N, Reinhardt R L, et al. (2002). Pluripotency of mesenchymal stem cells derived from adult marrow. *Nature.* 418:41–49.
28. Bianco P, Riminucci M, Gronthos S, et al. (2001). Bone marrow stromal stem cells: nature, biology, and potential applications. *Stem Cells.* 19:180–192.
29. Reyes M, Lund T, Lenvik T, et al. (2001). Purification and ex vivo expansion of postnatal human marrow mesodermal progenitor cells. *Blood.* 98:2615–2625.
30. Schwartz R E, Reyes M, Koodie L, et al. (2002). Multipotent adult progenitor cells from bone marrow differentiate into functional hepatocyte-like cells. *J. Clin. Invest.* 109:1291–1302.
31. Vats A, Bielby R C, Tolley N S, et al. (2005). Stem cells. *Lancet.* 366:592–602.
32. Young H E, Black A C, Jr (2004). Adult stem cells. *Anat. Rec. A Discov. Mol. Cell. Evol. Biol.* 276:75–102.
33. Becker A J, Mc C E, Till J E (1963). Cytological demonstration of the clonal nature of spleen colonies derived from transplanted mouse marrow cells. *Nature.* 197:452–454.
34. Weissman I L, Anderson D J, Gage F (2001). Stem and progenitor cells: origins, phenotypes, lineage commitments, and transdifferentiations. *Annu. Rev. Cell Dev. Biol.* 17:387–403.
35. Kondo M, Scherer D C, King A G, et al. (2001). Lymphocyte development from hematopoietic stem cells. *Curr. Opin. Genet. Dev.* 11:520–526.
36. Weissman I L (2000). Stem cells: Units of development, units of regeneration, and units in evolution. *Cell.* 100:157–168.
37. Bonnet D (2002). Haematopoietic stem cells. *J. Pathol.* 197:430–440.
38. Morrison S J, Uchida N, Weissman I L (1995). The biology of hematopoietic stem cells. *Annu. Rev. Cell Dev. Biol.* 11:35–71.
39. Morrison S J, Weissman I L (1994). The long-term repopulating subset of hematopoietic stem cells is deterministic and isolatable by phenotype. *Immunity.* 1:661–673.
40. Goodell M A, Brose K, Paradis G, et al. (1996). Isolation and functional properties of murine hematopoietic stem cells that are replicating in vivo. *J. Exp. Med.* 183:1797–1806.
41. Bradford G B, Williams B, Rossi R, et al. (1997). Quiescence, cycling, and turnover in the primitive hematopoietic stem cell compartment. *Exp. Hematol.* 25:445–453.
42. Cheshier S H, Morrison S J, Liao X, et al. (1999). In vivo proliferation and cell cycle kinetics of long-term self-renewing hematopoietic stem cells. *Proc. Natl. Acad. Sci. USA.* 96:3120–3125.
43. Nibley W E, Spangrude G J (1998). Primitive stem cells alone mediate rapid marrow recovery and multilineage engraftment after transplantation. *Bone Marrow Trans.* 21:345–354.

44. Lanzkron S M, Collector M I, Sharkis S J (1999). Hematopoietic stem cell tracking in vivo: A comparison of short-term and long-term repopulating cells. *Blood*. 93:1916–1921.
45. Goodell M A, Rosenzweig M, Kim H, et al. (1997). Dye efflux studies suggest that hematopoietic stem cells expressing low or undetectable levels of CD34 antigen exist in multiple species. *Nat. Med*. 3:1337–1345.
46. Uchida N, Fujisaki T, Eaves A C, et al. (2001). Transplantable hematopoietic stem cells in human fetal liver have a CD34(+) side population (SP)phenotype. *J. Clin. Invest*. 108:1071–1077.
47. Morrison S J, Wandycz A M, Hemmati H D, et al. (1997). Identification of a lineage of multipotent hematopoietic progenitors. *Develop*. 124:1929–1939.
48. Morrison S J, Wandycz A M, Akashi K, et al. (1996). The aging of hematopoietic stem cells. *Nat. Med*. 2:1011–1016.
49. Christensen J L, Weissman I L (2001). Flk-2 is a marker in hematopoietic stem cell differentiation: A simple method to isolate long-term stem cells. *Proc. Natl. Acad. Sci. USA*. 98:14541–14546.
50. Civin C I, Strauss L C, Brovall C (1984). Antigenic analysis of hematopoiesis. III. A hematopoietic progenitor cell surface antigen defined by a monoclonal antibody raised against KG-1a cells. *J. Immunol*. 133:157–165.
51. Larochelle A, Vormoor J, Hanenberg H (1996). Identification of primitive human hematopoietic cells capable of repopulating NOD/SCID mouse bone marrow: Implications for gene therapy. *Nat. Med*. 2:1329–1337.
52. Asahara T, Murohara T, Sullivan A, et al. (1997). Isolation of putative progenitor endothelial cells for angiogenesis. *Science*. 275:964–967.
53. Peichev M, Naiyer A J, Pereira D, et al. (2000). Expression of VEGFR-2 and AC133 by circulating human CD34(+) cells identifies a population of functional endothelial precursors. *Blood*. 95:952–958.
54. Lagasse E, Connors H, Al-Dhalimy M, et al. (2000). Purified hematopoietic stem cells can differentiate into hepatocytes in vivo. *Nat. Med*. 6:1229–1234.
55. Sutherland H J, Eaves C J, Eaves A C, et al. (1989). Characterization and partial purification of human marrow cells capable of initiating long-term hematopoiesis in vitro. *Blood*. 74:1563–1570.
56. Kawashima I, Zanjani E D, Almaida-Porada G, et al. (1996). CD34+ human marrow cells that express low levels of Kit protein are enriched for long-term marrow-engrafting cells. *Blood*. 87:4136–4142.
57. Nakamura Y, Ando K, Chargui J, et al. (1999). Ex vivo generation of CD34(+) cells from CD34(–) hematopoietic cells. *Blood*. 94:4053–4059.
58. Verfaillie C M, Almeida-Porada G, Wissink S, et al. (2000). Kinetics of engraftment of CD34(–) and CD34(+) cells from mobilized blood differs from that of CD34(–) and CD34(+) cells from bone marrow. *Exp. Hematol*. 28:1071–1079.
59. Zhou S, Schuetz J D, Bunting K D, et al. (2001). The ABC transporter Bcrp1/ABCG2 is expressed in a wide variety of stem cells and is a molecular determinant of the side-population phenotype. *Nat. Med*. 7:1028–1034.
60. Hill B, Rozler E, Travis M, et al. (1996). High-level expression of a novel epitope of CD59 identifies a subset of CD34+ bone marrow cells highly enriched for pluripotent stem cells. *Exp. Hematol*. 24:936–943.
61. Bradfute S B, Graubert T A, Goodell M A (2005). Roles of Sca-1 in hematopoietic stem/progenitor cell function. *Exp. Hematol*. 33:836–843.
62. Yin A H, Miraglia S, Zanjani E D, et al. (1997). AC133, a novel marker for human hematopoietic stem and progenitor cells. *Blood*. 90:5002–5012.

63. Uchida N, Tsukamoto A, He D, et al. (1998). High doses of purified stem cells cause early hematopoietic recovery in syngeneic and allogeneic hosts. *J. Clin. Invest.* 101:961–966.
64. Jordan C T, McKearn J P, Lemischka I R (1990). Cellular and developmental properties of fetal hematopoietic stem cells. *Cell.* 61:953–963.
65. Randall T D, Weissman I L (1997). Phenotypic and functional changes induced at the clonal level in hematopoietic stem cells after 5-fluorouracil treatment. *Blood.* 89:3596–3606.
66. Wright D E, Cheshier S H, Wagers A J, et al. (2001). Cyclophosphamide/granulocyte colony-stimulating factor causes selective mobilization of bone marrow hematopoietic stem cells into the blood after M phase of the cell cycle. *Blood.* 97:2278–2285.
67. Venezia T A, Merchant A A, Ramos C A, et al. (2004). Molecular signatures of proliferation and quiescence in hematopoietic stem cells. *PLoS. Biol.* 2:e301.
68. Heissig B, Hattori K, Dias S, et al. (2002). Recruitment of stem and progenitor cells from the bone marrow niche requires MMP-9 mediated release of kit-ligand. *Cell.* 109:625–637.
69. Levy D E, Lee C K (2002). What does Stat3 do? *J. Clin. Invest.* 109:1143–1148.
70. Raz R, Lee C K, Cannizzaro L A, et al. (1999). Essential role of STAT3 for embryonic stem cell pluripotency. *Proc. Natl. Acad. Sci. USA.* 96:2846–2851.
71. Krause D, Cantley L G (2005). Bone marrow plasticity revisited: Protection or differentiation in the kidney tubule? *J. Clin. Invest.* 115:1705–1708.
72. Kashofer K, Bonnet D (2005). Gene therapy progress and prospects: Stem cell plasticity. *Gene. Ther.* 12:1229–1234.
73. Brazelton T R, Rossi F M, Keshet G I, et al. (2000). From marrow to brain: Expression of neuronal phenotypes in adult mice. *Science.* 290:1775–1779.
74. Eglitis M A, Mezey E (1997). Hematopoietic cells differentiate into both microglia and macroglia in the brains of adult mice. *Proc. Natl. Acad. Sci. USA.* 94:4080–4085.
75. Mezey E, Chandross K J, Harta G, et al. (2000). Turning blood into brain: Cells bearing neuronal antigens generated in vivo from bone marrow. *Science.* 290:1779–1782.
76. Petersen B E, Bowen W C, Patrene K D, et al. (1999). Bone marrow as a potential source of hepatic oval cells. *Science.* 284:1168–1170.
77. Theise N D, Badve S, Saxena R, et al. (2000). Derivation of hepatocytes from bone marrow cells in mice after radiation-induced myeloablation. *Hepatol.* 31:235–240.
78. Alison M R, Poulsom R, Jeffery R, et al. (2000). Hepatocytes from non-hepatic adult stem cells. *Nature.* 406:257.
79. Orlic D, Kajstura J, Chimenti S, et al. (2001). Bone marrow cells regenerate infarcted myocardium. *Nature.* 410:701–705.
80. Orlic D, Kajstura J, Chimenti S, et al. (2001). Mobilized bone marrow cells repair the infarcted heart, improving function and survival. *Proc. Natl. Acad. Sci. USA.* 98:10344–10349.
81. Jackson K A, Majka S M, Wang H, et al. (2001). Regeneration of ischemic cardiac muscle and vascular endothelium by adult stem cells. *J. Clin. Invest.* 107:1395–1402.
82. Gussoni E, Soneoka Y, Strickland C D, et al. (1999). Dystrophin expression in the mdx mouse restored by stem cell transplantation. *Nature.* 401:390–394.
83. Ferrari G, Cusella-De Angelis G, Coletta M, et al. (1998). Muscle regeneration by bone marrow-derived myogenic progenitors. *Science.* 279:1528–1530.

84. Krause D S (2005). Engraftment of bone marrow-derived epithelial cells. *Ann. NY Acad. Sci.* 1044:117–124.
85. Orlic D (2005). BM stem cells and cardiac repair: where do we stand in 2004? *Cytother.* 7:3–15.
86. Terada N, Hamazaki T, Oka M, et al. (2002). Bone marrow cells adopt the phenotype of other cells by spontaneous cell fusion. *Nature.* 416:542–545.
87. Pellegrini G, Ranno R, Stracuzzi G, et al. (1999). The control of epidermal stem cells (holoclones) in the treatment of massive full-thickness burns with autologous keratinocytes cultured on fibrin. *Transplant.* 68:868–879.
88. Stamm C, Westphal B, Kleine H D, et al. (2003). Autologous bone-marrow stem-cell transplantation for myocardial regeneration. *Lancet.* 361:45–46.
89. Strauer B E, Brehm M, Zeus T, et al. (2002). Repair of infarcted myocardium by autologous intracoronary mononuclear bone marrow cell transplantation in humans. *Circulat.* 106:1913–1918.
90. Zuk P A, Zhu M, Mizuno H, et al. (2001). Multilineage cells from human adipose tissue: implications for cell-based therapies. *Tissue Eng.* 7:211–228.
91. Horwitz E M, Gordon P L, Koo W K, et al. (2002). Isolated allogeneic bone marrow-derived mesenchymal cells engraft and stimulate growth in children with osteogenesis imperfecta: Implications for cell therapy of bone. *Proc. Natl. Acad. Sci. USA.* 99:8932–8939.
92. Petite H, Viateau V, Bensaid W, et al. (2000). Tissue-engineered bone regeneration. *Nat. Biotechnol.* 18:959–963.
93. Koc O N, Gerson S L, Cooper B W, et al. (2000). Rapid hematopoietic recovery after coinfusion of autologous-blood stem cells and culture-expanded marrow mesenchymal stem cells in advanced breast cancer patients receiving high-dose chemotherapy. *J. Clin. Oncol.* 18:307–316.
94. Haynesworth S E, Baber M A, Caplan A I (1992). Cell surface antigens on human marrow-derived mesenchymal cells are detected by monoclonal antibodies. *Bone.* 13:69–80.
95. Barry F P, Boynton R E, Haynesworth S, et al. (1999). The monoclonal antibody SH-2, raised against human mesenchymal stem cells, recognizes an epitope on endoglin (CD105). *Biochem. Biophys. Res. Commun.* 265:134–139.
96. Barry F, Boynton R, Murphy M, et al. (2001). The SH-3 and SH-4 antibodies recognize distinct epitopes on CD73 from human mesenchymal stem cells. *Biochem. Biophys. Res. Commun.* 289:519–524.
97. Bruder S P, Ricalton N S, Boynton R E, et al. (1998). Mesenchymal stem cell surface antigen SB-10 corresponds to activated leukocyte cell adhesion molecule and is involved in osteogenic differentiation. *J. Bone Miner. Res.* 13:655–663.
98. Simmons P J, Torok-Storb B (1991). Identification of stromal cell precursors in human bone marrow by a novel monoclonal antibody, STRO-1. *Blood.* 78:55–62.
99. Le Blanc K, Tammik C, Rosendahl K, et al. (2003). HLA expression and immunologic properties of differentiated and undifferentiated mesenchymal stem cells. *Exp. Hematol.* 31:890–896.
100. Friedenstein A J, Piatetzky S, II, Petrakova K V (1966). Osteogenesis in transplants of bone marrow cells. *J. Embryol. Exp. Morphol.* 16:381–390.
101. Friedenstein A J, Gorskaja J F, Kulagina N N (1976). Fibroblast precursors in normal and irradiated mouse hematopoietic organs. *Exp. Hematol.* 4:267–274.
102. Keating A, Singer J W, Killen P D, et al. (1982). Donor origin of the in vitro haematopoietic microenvironment after marrow transplantation in man. *Nature.* 298:280–283.

103. Owen M, Friedenstein A J (1988). Stromal stem cells: Marrow-derived osteogenic precursors. *Ciba Found Symp.* 136:42–60.

104. Beresford J N (1989). Osteogenic stem cells and the stromal system of bone and marrow. *Clin. Orthop. Relat. Res.* 240:270–280.

105. Caplan A I (1991). Mesenchymal stem cells. *J. Orthop. Res.* 9:641–650.

106. Bruder S P, Jaiswal N, Haynesworth S E (1997). Growth kinetics, self-renewal, and the osteogenic potential of purified human mesenchymal stem cells during extensive subcultivation and following cryopreservation. *J. Cell. Biochem.* 64:278–294.

107. Jaiswal N, Haynesworth S E, Caplan A I, et al. (1997). Osteogenic differentiation of purified, culture-expanded human mesenchymal stem cells in vitro. *J. Cell. Biochem.* 64:295–312.

108. Bruder S P, Kurth A A, Shea M, et al. (1998). Bone regeneration by implantation of purified, culture-expanded human mesenchymal stem cells. *J. Orthop. Res.* 16:155–162.

109. Kadiyala S, Young R G, Thiede M A, et al. (1997). Culture expanded canine mesenchymal stem cells possess osteochondrogenic potential in vivo and in vitro. *Cell Transplant.* 6:125–134.

110. Prockop D J (1997). Marrow stromal cells as stem cells for nonhematopoietic tissues. *Science.* 276:71–74.

111. Woodbury D, Schwarz E J, Prockop D J, et al. (2000). Adult rat and human bone marrow stromal cells differentiate into neurons. *J. Neurosci. Res.* 61:364–370.

112. Hofstetter C P, Schwarz E J, Hess D, et al. (2002). Marrow stromal cells form guiding strands in the injured spinal cord and promote recovery. *Proc. Natl. Acad. Sci. USA.* 99:2199–2204.

113. Kopen G C, Prockop D J, Phinney D G (1999). Marrow stromal cells migrate throughout forebrain and cerebellum, and they differentiate into astrocytes after injection into neonatal mouse brains. *Proc. Natl. Acad. Sci. USA.* 96:10711–10716.

114. Bianco P, Gehron Robey P (2000). Marrow stromal stem cells. *J. Clin. Invest.* 105:1663–1668.

115. Pereira R F, O'Hara M D, Laptev A V, et al. (1998). Marrow stromal cells as a source of progenitor cells for nonhematopoietic tissues in transgenic mice with a phenotype of osteogenesis imperfecta. *Proc. Natl. Acad. Sci. USA.* 95:1142–1147.

116. Toma C, Pittenger M F, Cahill K S, et al. (2002). Human mesenchymal stem cells differentiate to a cardiomyocyte phenotype in the adult murine heart. *Circulat.* 105: 93–98.

117. Castro-Malaspina H, Gay R E, Resnick G (1980). Characterization of human bone marrow fibroblast colony-forming cells (CFU-F) and their progeny. *Blood.* 56:289–301.

118. Pittenger M F, Mackay A M, Beck S C, et al. (1999). Multilineage potential of adult human mesenchymal stem cells. *Science.* 284:143–147.

119. Muraglia A, Cancedda R, Quarto R (2000). Clonal mesenchymal progenitors from human bone marrow differentiate in vitro according to a hierarchical model. *J. Cell. Sci.* 113 (Pt 7):1161–1166.

120. Bianco P, Cossu G (1999). Uno, nessuno e centomila: Searching for the identity of mesodermal progenitors. *Exp. Cell. Res.* 251:257–263.

121. Bianco P, Riminucci M, Kuznetsov S, et al. (1999). Multipotential cells in the bone marrow stroma: Regulation in the context of organ physiology. *Crit. Rev. Eukaryot. Gene Expr.* 9:159–173.

122. Tsuda H, Wada T, Ito Y, et al. (2003). Efficient BMP2 gene transfer and bone formation of mesenchymal stem cells by a fiber-mutant adenoviral vector. *Mol. Ther.* 7:354–365.
123. Kurozumi K, Nakamura K, Tamiya T, et al. (2004). BDNF gene-modified mesenchymal stem cells promote functional recovery and reduce infarct size in the rat middle cerebral artery occlusion model. *Mol. Ther.* 9:189–197.
124. Kurozumi K, Nakamura K, Tamiya T, et al. (2005). Mesenchymal stem cells that produce neurotrophic factors reduce ischemic damage in the rat middle cerebral artery occlusion model. *Mol. Ther.* 11:96–104.
125. Nakamura K, Ito Y, Kawano Y, et al. (2004). Antitumor effect of genetically engineered mesenchymal stem cells in a rat glioma model. *Gene Ther.* 11:1155–1164.
126. Benedetti S, Pirola B, Pollo B (2000). Gene therapy of experimental brain tumors using neural progenitor cells. *Nat. Med.* 6:447–450.
127. Ehtesham M, Kabos P, Kabosova A (2002). The use of interleukin 12-secreting neural stem cells for the treatment of intracranial glioma. *Cancer Res.* 62:5657–5663.
128. Reyes M, Verfaillie C M (2001). Characterization of multipotent adult progenitor cells, a subpopulation of mesenchymal stem cells. *Ann. NY Acad. Sci.* 938:231–233, discussion 233–235.
129. Thomson J A, Itskovitz-Eldor J, Shapiro S S, et al. (1998). Embryonic stem cell lines derived from human blastocysts. *Science.* 282:1145–1147.
130. Reyes M, Dudek A, Jahagirdar B, et al. (2002). Origin of endothelial progenitors in human postnatal bone marrow. *J. Clin. Invest.* 109:337–346.
131. Keene C D, Ortiz-Gonzalez X R, Jiang Y (2003). Neural differentiation and incorporation of bone marrow-derived multipotent adult progenitor cells after single cell transplantation into blastocyst stage mouse embryos. *Cell Transplant.* 12:201–213.
132. Jiang Y, Henderson D, Blackstad M, et al. (2003). Neuroectodermal differentiation from mouse multipotent adult progenitor cells. *Proc. Natl. Acad. Sci. USA.* 100(suppl 1):11854–11860.
133. Wolf N S, Kone A, Priestley G V, et al. (1993). In vivo and in vitro characterization of long-term repopulating primitive hematopoietic cells isolated by sequential Hoechst 33342-rhodamine 123 FACS selection. *Exp. Hematol.* 21:614–622.
134. Summer R, Kotton D N, Sun X, et al. (2004). Translational physiology: Origin and phenotype of lung side population cells. *Am. J. Physiol. Lung. Cell Mol. Physiol.* 287: L477–483.
135. Mezey E, Chandross K J (2000). Bone marrow: A possible alternative source of cells in the adult nervous system. *Eur. J. Pharmacol.* 405:297–302.
136. Castro R F, Jackson K A, Goodell M A, et al. (2002). Failure of bone marrow cells to transdifferentiate into neural cells in vivo. *Science.* 297:1299.
137. Murayama A, Matsuzaki Y, Kawaguchi A, et al. (2002). Flow cytometric analysis of neural stem cells in the developing and adult mouse brain. *J. Neurosci. Res.* 69:837–847.
138. Asakura A, Rudnicki M A (2002). Side population cells from diverse adult tissues are capable of in vitro hematopoietic differentiation. *Exp. Hematol.* 30:1339–1345.
139. Chen J, Hersmus N, Van Duppen V, et al. (2005). The adult pituitary contains a cell population displaying stem/progenitor cell and early embryonic characteristics. *Endocrinol.* 146:3985–3998.
140. Terunuma A, Jackson K L, Kapoor V, et al. (2003). Side population keratinocytes resembling bone marrow side population stem cells are distinct from label-retaining keratinocyte stem cells. *J. Invest. Dermatol.* 121:1095–1103.

141. Yano S, Ito Y, Fujimoto M, et al. (2005). Characterization and localization of side population cells in mouse skin. *Stem Cells*. 23:834–841.

142. Montanaro F, Liadaki K, Volinski J, et al. (2003). Skeletal muscle engraftment potential of adult mouse skin side population cells. *Proc. Natl. Acad. Sci. USA*. 100:9336–9341.

143. Montanaro F, Liadaki K, Schienda J, et al. (2004). Demystifying SP cell purification: Viability, yield, and phenotype are defined by isolation parameters. *Exp. Cell Res.* 298:144–154.

144. Budak M T, Alpdogan O S, Zhou M (2005). Ocular surface epithelia contain ABCG2-dependent side population cells exhibiting features associated with stem cells. *J. Cell Sci*. 118:1715–1724.

145. Umemoto T, Yamato M, Nishida K, et al. (2006). Limbal epithelial side-population cells have stem cell-like properties, including quiescent state. *Stem Cells*. 24:86–94.

146. Alvi A J, Clayton H, Joshi C, et al. (2003). Functional and molecular characterisation of mammary side population cells. *Breast Cancer Res.* 5:R1–8.

147. Liadaki K, Kho A T, Sanoudou D, et al. (2005). Side population cells isolated from different tissues share transcriptome signatures and express tissue-specific markers. *Exp. Cell Res.* 303:360–374.

148. Asakura A, Seale P, Girgis-Gabardo A, et al. (2002). Myogenic specification of side population cells in skeletal muscle. *J. Cell. Biol.* 159:123–134.

149. Giangreco A, Shen H, Reynolds S D, et al. (2004). Molecular phenotype of airway side population cells. *Am. J. Physiol. Lung Cell. Mol. Physiol.* 286:L624–630.

150. Majka S M, Beutz M A, Hagen M, et al. (2005). Identification of novel resident pulmonary stem cells: Form and function of the lung side population. *Stem Cells.* 23:1073–1081.

151. Summer R, Kotton D N, Liang S, et al. (2005). Embryonic lung side population cells are hematopoietic and vascular precursors. *Am. J. Respir. Cell. Mol. Biol.* 33:32–40.

152. Pfister O, Mouquet F, Jain M, et al. (2005). CD31– but Not CD31+ cardiac side population cells exhibit functional cardiomyogenic differentiation. *Circ. Res.* 97:52–61.

153. Meissner K, Heydrich B, Jedlitschky G, et al. (2005). The ATP-binding cassette transporter ABCG2 (BCRP), a marker for side population stem cells, is expressed in human heart. *J. Histochem. Cytochem.* 54:215–221.

154. Shimano K, Satake M, Okaya A, et al. (2003). Hepatic oval cells have the side population phenotype defined by expression of ATP-binding cassette transporter ABCG2/BCRP1. *Am. J. Pathol.* 163:3–9.

155. Lechner A, Leech C A, Abraham E J, et al. (2002). Nestin-positive progenitor cells derived from adult human pancreatic islets of Langerhans contain side population (SP) cells defined by expression of the ABCG2 (BCRP1) ATP-binding cassette transporter. *Biochem. Biophys. Res. Commun.* 293:670–674.

156. Hishikawa K, Marumo T, Miura S, et al. (2005). Musculin/MyoR is expressed in kidney side population cells and can regulate their function. *J. Cell. Biol.* 169:921–928.

157. Hishikawa K, Marumo T, Miura S, et al. (2005). Leukemia inhibitory factor induces multi-lineage differentiation of adult stem-like cells in kidney via kidney-specific cadherin 16. *Biochem. Biophys. Res. Commun.* 328:288–291.

158. Iwatani H, Ito T, Imai E, et al. (2004). Hematopoietic and nonhematopoietic potentials of Hoechst(low)/side population cells isolated from adult rat kidney. *Kidney Int.* 65:1604–1614.

159. Falciatori I, Borsellino G, Haliassos N, et al. (2004). Identification and enrichment of spermatogonial stem cells displaying side-population phenotype in immature mouse testis. *Faseb. J.* 18:376–378.
160. Lassalle B, Bastos H, Louis J P (2004). "Side Population" cells in adult mouse testis express Bcrp1 gene and are enriched in spermatogonia and germinal stem cells. *Develop.* 131:479–487.
161. Preffer F I, Dombkowski D, Sykes M, et al. (2002). Lineage-negative side-population (SP) cells with restricted hematopoietic capacity circulate in normal human adult blood: Immunophenotypic and functional characterization. *Stem Cells.* 20:417–427.
162. Briscoe J, Chen Y, Jessell T M, et al. (2001). A hedgehog-insensitive form of patched provides evidence for direct long-range morphogen activity of sonic hedgehog in the neural tube. *Mol. Cell* 7:1279–1291.
163. Briscoe J, Ericson J (2001). Specification of neuronal fates in the ventral neural tube. *Curr. Opin. Neurobiol.* 11:43–49.
164. Jessell T M (2000). Neuronal specification in the spinal cord: Inductive signals and transcriptional codes. *Nat. Rev. Genet.* 1:20–29.
165. Morrison S J (2001). Neuronal potential and lineage determination by neural stem cells. *Curr. Opin. Cell Biol.* 13:666–672.
166. Ciccolini F, Svendsen C N (1998). Fibroblast growth factor 2 (FGF-2) promotes acquisition of epidermal growth factor (EGF) responsiveness in mouse striatal precursor cells: Identification of neural precursors responding to both EGF and FGF-2. *J. Neurosci.* 18:7869–7880.
167. Weiss S, Dunne C, Hewson J, et al. (1996). Multipotent CNS stem cells are present in the adult mammalian spinal cord and ventricular neuroaxis. *J. Neurosci.* 16:7599–7609.
168. Palmer T D, Takahashi J, Gage F H (1997). The adult rat hippocampus contains primordial neural stem cells. *Mol. Cell Neurosci.* 8:389–404.
169. Seaberg R M, Smukler S R, Kieffer T J (2004). Clonal identification of multipotent precursors from adult mouse pancreas that generate neural and pancreatic lineages. *Nat. Biotechnol.* 22:1115–1124.
170. Compton C C, Nadire K B, Regauer S, et al. (1998). Cultured human sole-derived keratinocyte grafts re-express site-specific differentiation after transplantation. *Different.* 64:45–53.
171. Jones P H, Watt F M (1993). Separation of human epidermal stem cells from transit amplifying cells on the basis of differences in integrin function and expression. *Cell.* 73:713–724.
172. Levy L, Broad S, Diekmann D (2000). beta1 integrins regulate keratinocyte adhesion and differentiation by distinct mechanisms. *Mol. Biol. Cell.* 11:453–466.
173. Fernandes K J, McKenzie I A, Mill P, et al. (2004). A dermal niche for multipotent adult skin-derived precursor cells. *Nat. Cell. Biol.* 6:1082–1093.
174. Liang L, Bickenbach J R (2002). Somatic epidermal stem cells can produce multiple cell lineages during development. *Stem Cells.* 20:21–31.
175. Toma J G, Akhavan M, Fernandes K J, et al. (2001). Isolation of multipotent adult stem cells from the dermis of mammalian skin. *Nat. Cell. Biol.* 3:778–784.
176. Borthwick D W, Shahbazian M, Krantz Q T, et al. (2001). Evidence for stem-cell niches in the tracheal epithelium. *Am. J. Respir. Cell. Mol. Biol.* 24:662–670.
177. Griffiths M J, Bonnet D, Janes S M (2005). Stem cells of the alveolar epithelium. *Lancet.* 366:249–260.

178. Kotton D N, Summer R, Fine A (2004). Lung stem cells: New paradigms. *Exp. Hematol.* 32:340–343.
179. Charge S B, Rudnicki M A (2004). Cellular and molecular regulation of muscle regeneration. *Physiol. Rev.* 84:209–238.
180. Montarras D, Morgan J, Collins C, et al. (2005). Direct isolation of satellite cells for skeletal muscle regeneration. *Science.* 309:2064–2067.
181. Morgan J E, Partridge T A (2003). Muscle satellite cells. *Int. J. Biochem. Cell. Biol.* 35:1151–1156.
182. Bischoff R (1997). Chemotaxis of skeletal muscle satellite cells. *Dev. Dyn.* 208: 505–515.
183. Forbes S, Vig P, Poulsom R, et al. (2002). Hepatic stem cells. *J. Pathol.* 197:510–518.
184. Overturf K, Al-Dhalimy M, Finegold M, et al. (1999). The repopulation potential of hepatocyte populations differing in size and prior mitotic expansion. *Am. J. Pathol.* 155:2135–2143.
185. Overturf K, al-Dhalimy M, Ou C N, et al. (1997). Serial transplantation reveals the stem-cell-like regenerative potential of adult mouse hepatocytes. *Am. J. Pathol.* 151:1273–1280.
186. Tateno C, Yoshizato K (1996). Growth and differentiation in culture of clonogenic hepatocytes that express both phenotypes of hepatocytes and biliary epithelial cells. *Am. J. Pathol.* 149:1593–1605.
187. Tateno C, Yoshizato K (1996). Long-term cultivation of adult rat hepatocytes that undergo multiple cell divisions and express normal parenchymal phenotypes. *Am. J. Pathol.* 148:383–392.
188. Hirose S, Ise H, Uchiyama M, et al. (2001). Regulation of asialoglycoprotein receptor expression in the proliferative state of hepatocytes. *Biochem. Biophys. Res. Commun.* 287:675–681.
189. Ise H, Sugihara N, Negishi N, et al. (2001). Low asialoglycoprotein receptor expression as markers for highly proliferative potential hepatocytes. *Biochem. Biophys. Res. Commun.* 285:172–182.
190. Paku S, Schnur J, Nagy P, et al. (2001). Origin and structural evolution of the early proliferating oval cells in rat liver. *Am. J. Pathol.* 158:1313–1323.
191. Petersen B E, Goff J P, Greenberger J S, et al. (1998). Hepatic oval cells express the hematopoietic stem cell marker Thy-1 in the rat. *Hepatol.* 27:433–445.
192. Cohn S M, Roth K A, Birkenmeier E H, et al. (1991). Temporal and spatial patterns of transgene expression in aging adult mice provide insights about the origins, organization, and differentiation of the intestinal epithelium. *Proc. Natl. Acad. Sci. USA.* 88:1034–1038.
193. Gordon J I, Schmidt G H, Roth K A (1992). Studies of intestinal stem cells using normal, chimeric, and transgenic mice. *Faseb. J.* 6:3039–3050.
194. Karam S M (1999). Lineage commitment and maturation of epithelial cells in the gut. *Front. Biosci.* 4:D286–298.
195. Roth K A, Hermiston M L, Gordon J I (1991). Use of transgenic mice to infer the biological properties of small intestinal stem cells and to examine the lineage relationships of their descendants. *Proc. Natl. Acad. Sci. USA.* 88:9407–9411.
196. Suzuki A, Nakauchi H, Taniguchi H (2003). Glucagon-like peptide 1 (1–37) converts intestinal epithelial cells into insulin-producing cells. *Proc. Natl. Acad. Sci. USA.* 100:5034–5039.

197. Powell D W, Mifflin R C, Valentich J D (1999). Myofibroblasts. II. Intestinal subepithelial myofibroblasts. *Am. J. Physiol.* 277:C183–201.
198. Powell D W, Mifflin R C, Valentich J D, et al. (1999). Myofibroblasts. I. Paracrine cells important in health and disease. *Am. J. Physiol.* 277:C1–9.
199. Gittes G K, Galante P E, Hanahan D, et al. (1996). Lineage-specific morphogenesis in the developing pancreas: Role of mesenchymal factors. *Develop.* 122:439–447.
200. Finegood D T, Scaglia L, Bonner-Weir S (1995). Dynamics of beta-cell mass in the growing rat pancreas. Estimation with a simple mathematical model. *Diabetes.* 44:249–256.
201. Hunziker E, Stein M (2000). Nestin-expressing cells in the pancreatic islets of Langerhans. *Biochem. Biophys. Res. Commun.* 271:116–119.
202. Bonner-Weir S, Toschi E, Inada A, et al. (2004). The pancreatic ductal epithelium serves as a potential pool of progenitor cells. *Pediatr. Diabetes.* 5(suppl 2):16–22.
203. Lechner A, Nolan A L, Blacken R A, et al. (2005). Redifferentiation of insulin-secreting cells after in vitro expansion of adult human pancreatic islet tissue. *Biochem. Biophys. Res. Commun.* 327:581–588.
204. Song K H, Ko S H, Ahn Y B, et al. (2004). In vitro transdifferentiation of adult pancreatic acinar cells into insulin-expressing cells. *Biochem. Biophys. Res. Commun.* 316:1094–1100.
205. Lardon J, De Breuck S, Rooman I, et al. (2004). Plasticity in the adult rat pancreas: transdifferentiation of exocrine to hepatocyte-like cells in primary culture. *Hepatol.* 39:1499–1507.
206. Zulewski H, Abraham E J, Gerlach M J, et al. (2001). Multipotential nestin-positive stem cells isolated from adult pancreatic islets differentiate ex vivo into pancreatic endocrine, exocrine, and hepatic phenotypes. *Diabetes.* 50:521–533.
207. Heller R S, Stoffers D A, Bock T, et al. (2001). Improved glucose tolerance and acinar dysmorphogenesis by targeted expression of transcription factor PDX-1 to the exocrine pancreas. *Diabetes.* 50:1553–1561.
208. Bonner-Weir S, Taneja M, Weir G C, et al. (2000). In vitro cultivation of human islets from expanded ductal tissue. *Proc. Natl. Acad. Sci. USA.* 97:7999–8004.
209. Bogdani M, Lefebvre V, Buelens N, et al. (2003). Formation of insulin-positive cells in implants of human pancreatic duct cell preparations from young donors. *Diabetol.* 46:830–838.
210. Beattie G M, Montgomery A M, Lopez A D, et al. (2002). A novel approach to increase human islet cell mass while preserving beta-cell function. *Diabetes.* 51:3435–3439.
211. Beattie G M, Itkin-Ansari P, Cirulli V, et al. (1999). Sustained proliferation of PDX-1+ cells derived from human islets. *Diabetes.* 48:1013–1019.
212. Kim S Y, Lee S H, Kim B M, et al. (2004). Activation of nestin-positive duct stem (NPDS) cells in pancreas upon neogenic motivation and possible cytodifferentiation into insulin-secreting cells from NPDS cells. *Dev. Dyn.* 230:1–11.
213. Ueno H, Yamada Y, Watanabe R, et al. (2005). Nestin-positive cells in adult pancreas express amylase and endocrine precursor Cells. *Pancreas.* 31:126–131.
214. Treutelaar M K, Skidmore J M, Dias-Leme C L, et al. (2003). Nestin-lineage cells contribute to the microvasculature but not endocrine cells of the islet. *Diabetes.* 52:2503–2512.
215. Knudtzon S (1974). In vitro growth of granulocytic colonies from circulating cells in human cord blood. *Blood.* 43:357–361.

216. Nakahata T, Ogawa M (1982). Hemopoietic colony-forming cells in umbilical cord blood with extensive capability to generate mono- and multipotential hemopoietic progenitors. *J. Clin. Invest.* 70:1324–1328.
217. Gluckman E, Broxmeyer H A, Auerbach A D, et al. (1989). Hematopoietic reconstitution in a patient with Fanconi's anemia by means of umbilical-cord blood from an HLA-identical sibling. *N. Engl. J. Med.* 321:1174–1178.
218. Cohen Y, Nagler A (2004). Umbilical cord blood transplantation–how, when and for whom? *Blood Rev.* 18:167–179.
219. de Wynter E A, Testa N G (2001). Interest of cord blood stem cells. *Biomed. Pharmacother.* 55:195–200.
220. Mayani H, Lansdorp P M (1998). Biology of human umbilical cord blood-derived hematopoietic stem/progenitor cells. *Stem Cells.* 16:153–165.
221. Donaldson C, Denning-Kendall P, Bradley B, et al. (2001). The CD34(+)CD38(neg) population is significantly increased in haemopoietic cell expansion cultures in serum-free compared to serum-replete conditions: dissociation of phenotype and function. *Bone Marrow Trans.* 27:365–371.
222. Hows J M (2001). Status of umbilical cord blood transplantation in the year 2001. *J. Clin. Pathol.* 54:428–434.
223. Moore M A, Hoskins I (1994). Ex vivo expansion of cord blood-derived stem cells and progenitors. *Blood Cells.* 20:468–479; discussion 479–481.
224. Buzanska L, Habich A, Jurga M, et al. (2005). Human cord blood-derived neural stem cell line-Possible implementation in studying neurotoxicity. *Toxicol In Vitro.* 19(7): 991–999.
225. Bieback K, Kern S, Kluter H, et al. (2004). Critical parameters for the isolation of mesenchymal stem cells from umbilical cord blood. *Stem Cells.* 22:625–634.
226. Sanberg P R, Willing A E, Garbuzova-Davis S, et al. (2005). Umbilical cord blood-derived stem cells and brain repair. *Ann. N.Y. Acad. Sci.* 1049:67–83.
227. Chen N, Hudson J E, Walczak P, et al. (2005). Human umbilical cord blood progenitors: The potential of these hematopoietic cells to become neural. *Stem Cells.* 23:1560–1570.
228. Watt S M, Contreras M (2005). Stem cell medicine: Umbilical cord blood and its stem cell potential. *Semin. Fetal Neonat. Med.* 10:209–220.
229. Schmidt D, Breymann C, Weber A, et al. (2004). Umbilical cord blood derived endothelial progenitor cells for tissue engineering of vascular grafts. *Ann. Thorac. Surg.* 78:2094–2098.
230. Ma N, Stamm C, Kaminski A, et al. (2005). Human cord blood cells induce angiogenesis following myocardial infarction in NOD/scid-mice. *Cardiovasc. Res.* 66:45–54.
231. Leor J, Guetta E, Feinberg M S (2006). Human umbilical cord blood-derived CD133+ cells enhance function and repair of the infarcted myocardium. *Stem Cells.* 24:772–780.
232. Korbling M, Robinson S, Estrov Z, et al. (2005). Umbilical cord blood-derived cells for tissue repair. *Cytother.* 7:258–261.
233. Ishikawa F, Yasukawa M, Yoshida S, et al. (2004). Human cord blood- and bone marrow-derived CD34+ cells regenerate gastrointestinal epithelial cells. *Faseb J.* 18:1958–1960.
234. Pessina A, Eletti B, Croera C, et al. (2004). Pancreas developing markers expressed on human mononucleated umbilical cord blood cells. *Biochem. Biophys. Res. Commun.* 323:315–322.

235. Gang E J, Jeong J A, Hong S H, et al. (2004). Skeletal myogenic differentiation of mesenchymal stem cells isolated from human umbilical cord blood. *Stem Cells.* 22:617–624.
236. Gang E J, Hong S H, Jeong J A, et al. (2004). In vitro mesengenic potential of human umbilical cord blood-derived mesenchymal stem cells. *Biochem. Biophys. Res. Commun.* 321:102–108.
237. Jeong J A, Gang E J, Hong S H, et al. (2004). Rapid neural differentiation of human cord blood-derived mesenchymal stem cells. *Neuroreport.* 15:1731–1734.
238. Kucia M, Dawn B, Hunt G, et al. (2004). Cells expressing early cardiac markers reside in the bone marrow and are mobilized into the peripheral blood after myocardial infarction. *Circ. Res.* 95:1191–1199.
239. Kucia M, Jankowski K, Reca R, et al. (2004). CXCR4-SDF-1 signalling, locomotion, chemotaxis and adhesion. *J. Mol. Histol.* 35:233–245.
240. Kucia M, Ratajczak J, Reca R, et al. (2004). Tissue-specific muscle, neural and liver stem/progenitor cells reside in the bone marrow, respond to an SDF-1 gradient and are mobilized into peripheral blood during stress and tissue injury. *Blood Cells Mol. Dis.* 32:52–57.
241. Ratajczak M Z, Kucia M, Reca R, et al. (2004). Stem cell plasticity revisited: CXCR4-positive cells expressing mRNA for early muscle, liver and neural cells "hide out" in the bone marrow. *Leukemia.* 18:29–40.
242. Wojakowski W, Tendera M, Michalowska A, et al. (2004). Mobilization of CD34/CXCR4+, CD34/CD117+, c-met+ stem cells, and mononuclear cells expressing early cardiac, muscle, and endothelial markers into peripheral blood in patients with acute myocardial infarction. *Circulat.* 110:3213–3220.

## 4 ARTIFICIAL CELLS

### 4.1 Introduction

A major limitation to the clinical application of cell therapy and drug discovery is the current inability to isolate an adequate number of functional and transplantable cells [1]. Unfortunately, the number of human organs available for transplantation or cell isolation is severely limited. Considering the cost and difficulty in some cases of cell isolation and the need for immediate availability of consistent and functionally uniform cell preparations, human cells cannot be isolated on a scale sufficient to treat more than a fraction of the patients who need organ transplantation. The use of animal cells would result in additional concerns related to the transmission of infectious pathogens and immunologic and physiologic incompatibilities between donors and humans [2, 3]. Thus, other alternative cell sources, such as stem cells, have been explored. In particular, embryonic stem cells have unlimited proliferative capacity and theoretically can differentiate all kinds of cell types. In contrast, somatic stem cells have finite proliferation ability in the currently available culture system [4]. The cultivation of mammalian cells is an important experimental procedure that is a fundamental tool in biological studies [5]. Establishment of methods to control differentiation and proliferation in stem cells is not yet achieved. Thus, possible tumor development will be occurred after de-differentiated or trans-differentiated stem cells. To overcome this issue, great efforts have been made to construct immortalized human cell lines with an unlimited replicative potential.

However, despite the widespread use of the culture system, we have recently begun to understand the essential mechanisms that control cell growth and division. This section discusses recent advances in our understanding of the molecular mechanisms that regulate the life span of cell lineages *in vitro* and an approach to construct human cell lines with currently available genetic manipulation for cell therapy and the study of pharmacology and toxicology. The definition of cellular immortality is infinite survival with an unlimited proliferative potential [6]. In fact, primary human cells in tissue culture rarely undergo spontaneous immortalization process [7]. Thus, researchers tried to immortalize human cells by using x-ray [8] and chemical carcinogens [7], but it turned out to be very inefficient. Currently, human cells have been immortalized by an introduction of the transforming genes of DNA tumor viruses, such simian virus 40 large T antigen (SV40LT) [9], papillomaviruses [10], and the catalytic unit of the human telomerase reverse transcriptase [11].

### *4.1.1 Immortalization of Human Cells*

*Limited Proliferative Life Span or Senescence of Human Cells.* Senescence is a terminally arrested growth state of the cells. It differs from the nonproliferative state of terminally differentiated cells. As senescent cells remain viable, senescence can be distinguished with cell death processes such as apoptosis or necrosis. Senescent cells are arrested at the G1/S phase of the cell cycle and are thus distinct from nondividing G0-arrested quiescent cells [5, 12]. Human cells are defined as senescent when they fail to respond to mitogens and the population of the cells does not divide in a certain period of time, for example, 30 days. Additional characteristics of senescent cells include (1) expression of β-galactosidase, (2) increased lysossomal biogenesis, (3) decreased rates of protein synthesis and degradation that are distinct from their pre-senescent ancestors, (4) increased cell size, (5) multinucleation, (6) cytoplasmic vacuolation, and (7) decreased membrane fluidity. The rate of protein, DNA, and RNA syntheses is reduced in senescent cells. Senescent cells accumulate altered macromolecules and exhibit DNA alterations such as shortened telomeres, decreased chromatin condensation, increased karyotypic abnormalities, and decreased methylation. Senescence does not lead directly to cell death, and senescent cells can persist in culture for up to 2 years if fed regularly [13].

Senescent cells may accumulate in aged tissue [14] and could compromise tissue function by both their altered pattern of gene expression and their nonproliferative state, resulting in aged phenotypes of the tissues [15]. A link between *in vitro* senescence and *in vivo* aging is suggested by the observation that the *in vitro* life span of cells from various mammalian species correlates with their *in vivo* life span and that the cells from individuals with premature ages suffering from disorders such as Werner's syndrome, Down's syndrome, and progeria have a shorter *in vitro* life span than that of the cells from normal individuals. However, evidence that the *in vitro* life span of normal human cells correlates with the donor age remains controversial [16].

*Immortal or Extended Life Span of Human Cells.* More than 30 years ago, it was demonstrated for the first time that SV40 [17], a DNA tumor virus of the papova

virus family isolated by Sweet and Hillman [18], can morphologically transform human fetal and adult skin fibroblast. The transformation of mammalian cells by SV40 is known to require expression of only the early region of the viral genomes, which encodes two proteins of large T-antigen (94 kd) and small t-antigen (17 kd) [18]. Transfer and expression of specific oncogenes, such as simian virus 40 large T antigen (SV40 LT), in primary human cells can generate cell populations that propagate for extended periods of time *in vitro*, presumably because they bypass the first senescence crisis through the inactivation of p53 and Rb, and thus extend their life span by about 20 population doubling (PD) [19]. This extended life span ends in M2 crisis stage (Figures 4-1 and 4-2), which typically coincides with

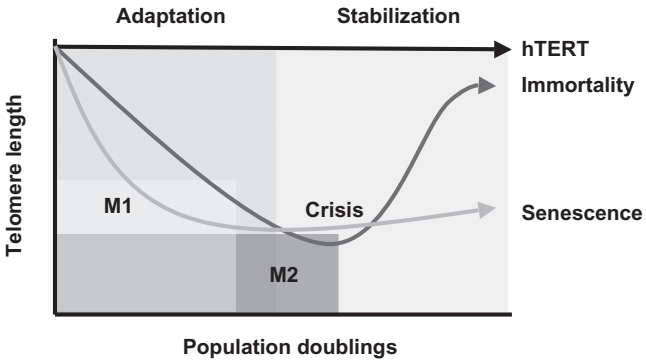


**Figure 4-1.** ***Cellular senescence and immortalization.*** Telomere length is maintained by telomerase, and most human somatic cells have lower levels of telomerase. The cells are telomerase-negative and experience telomere shortening with each cell division. Shortened telomeres may start the cells to enter senescence at the Hayflick limit, or M1. This proliferative checkpoint can be overcome by an inactivation of pRB/p16 or p53, for example, by the use of SV40 or human papilloma virus oncoproteins. Such cells continue to suffer telomere erosion and ultimately enter crisis, or M2, characterized by cell death. Quite a few surviving cells acquire stabilization of telomere length and unlimited proliferative potential, mostly due to activation of telomerase. (This figure is available in full color at ftp://ftp.wiley.com/public/sci_tech_med/pharmaceutical_biotech/.)

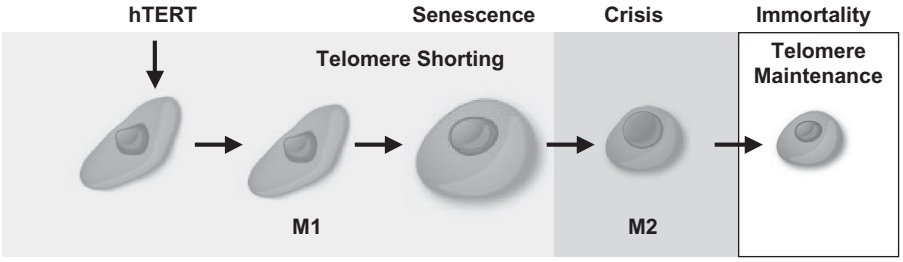


**Figure 4-2.** ***Immortalization of cells with hTERT.*** Ectopic expression of hTERT allows cells to bypass proliferation barriers and become immortal. Telomeres are very important in booth senescence (M1) and crisis (M2) as hTERT introduction either before or after M1 results in cell immortalization. If introduction of hTERT does not result in immortalization in the cells, another type of growth arrest called "premature senescence" or "cell culture shock" will be occurred, which is telomere-independent. (This figure is available in full color at ftp://ftp.wiley.com/public/sci_tech_med/pharmaceutical_biotech/.)

dangerously shortened telomeres, and every mitosis is accompanied by enormous chromosomal instability in the cells. Most of the cells fail to survive crisis, and the frequency rate of the cells that overcome the crisis is very low in humans. The *in vitro* spontaneous immortalization of human cells using SV40T is approximately $10^6$–$10^7$ and is believed to result from activation of the endogenous telomerase [20]. Such cells are infrequently tumorogenic, and tumorigenicity occurs only after long periods of continuous culture of the cells [10]. These phenomena have suggested that establishment of permanent cell lines in culture involves two distinct processes: (1) an initial adaptation of cells to grow in the unnatural environment of culture dish, and (2) an acquired ability of these adapted cells to proliferate indefinitely in tissue culture (Figure 4-1). The life span of cells is measured in cell divisions [5]; thus, cells must possess a molecular "clock" that counts the number of times they have divided. In a telomere model of senescence, telomere length acts as this counting mechanism. Nucleoprotein structures that constitute the ends of linear chromosomes became the primary candidates to fulfill this role in human cells. It was for the first time described molecularly in 1981 [21]; telomeres serve to protect the ends of chromosomes from illegitimate fusions and other damage and shield the ends of chromosomes from recognition by the cellular DNA repair machinery [22]. Telomeric DNA is maintained at a constant length, and telomeres shorten at a constant rate with progressive cell divisions in human cells (Figure 4-3). Telomere shortening occurs for at least two reasons: (1) Telomeres shorten with successive replication because the polymerases involved in conventional DNA replication cannot fully copy telomeric DNA [23] and (2) telomerase, an enzyme responsible for telomere maintenance, is tightly repressed in most human somatic cells [24]. Most immortalized human cells express telomerase [24], and telomere length is stable. These observations have suggested that some elements of telomere structures monitor cell proliferation and order a signal of the onset of replicative senescence. The ectopic expression of telomerase in pre-senescent cells halts telomere shortening and permits cells to avoid replicative senescence. Telomerase is a ribonucleoprotein composed of an RNA subunit, *hTERC*, that is ubiquitously expressed [25] and a reverse transcriptase protein catalytic subunit, *hTERT*, whose expression correlates with telomerase activity in immortal cells [26].

*hTERT* is indeed undetectable in most types of normal human cells, but it is readily detectable in most immortal cell lines, cells derived from human cancers, and human tumor samples. Expression of *hTERT* in pre-senescent human cells confers telomerase activity [10], resulting in telomere lengthening or stabilization, and it allows the cells to bypass replicative senescence. Cells expressing *hTERT* maintain the normal karyotype and continue to respond in the same manner as pre-senescent, nonimmortalized cells [27]. Thus, activation of telomerase by expression of *hTERT* permits the cells to bypass senescence and become immortalized in a single step. Although maintenance of telomere plays an important role in replicative senescence, it has been proved that both pRB and p53 tumor suppressor pathways also play a prominent role in regulating the onset of replicative senescence (Figure 4-4). When cells expressing SV40LT continue to repress hTERT expression and lack a detectable telomerase activity, the telomere length of the cells gradually shortens with continued cell proliferation. Eventually such cells enter the second period of diminished growth, which referred is to as crisis or M2 [28]. Telomere shortening may be one signal that activates the pRB or p53 pathway and initiates replicative senescence. These observations have suggested that both pRB

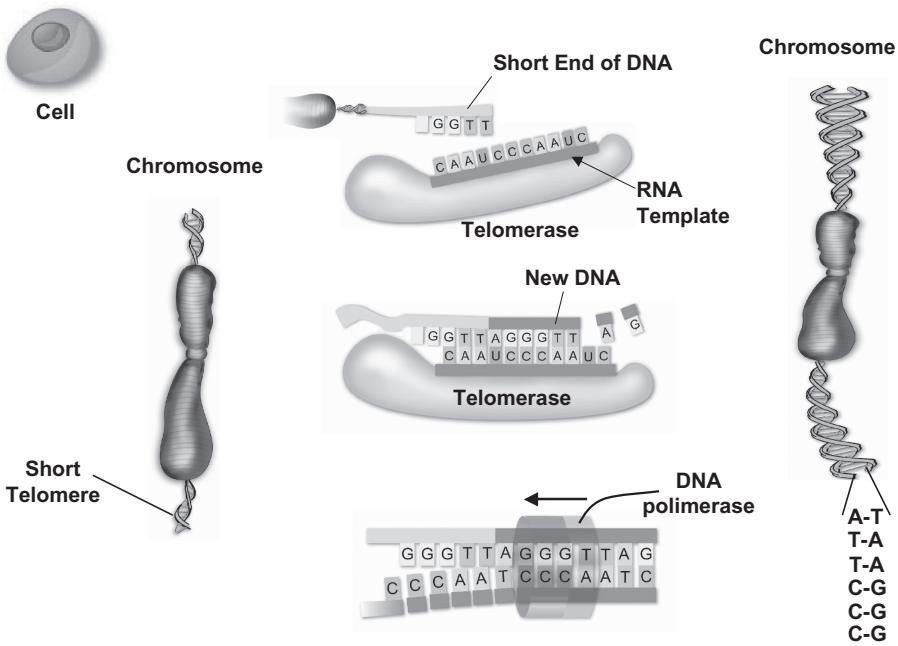


**Figure 4-3.** ***Telomeres and telomerase.*** Telomere is a repeating sequence of double-stranded DNA located at the ends of chromosomes. Long telomere length is associated with immortalized cell lines, cancer cells, and embryonic stem cells. Cells experience telomere shortening with each cell division. Telomerase is an enzyme that lengthens telomeres by adding on repeating sequences of DNA. Telomerase binds the ends of the telomere via an RNA template that is used for the attachment of a new strand of DNA. Telomerase adds several repeated DNA sequences, then releases to a second enzyme, DNA polymerase, and attaches the opposite or complementary strand of DNA, completing the double-stranded extension of the chromosome ends. The function of telomere and telomerase seem to be important in cell division, normal development, and cancer research. (This figure is available in full color at ftp://ftp.wiley.com/public/sci_tech_med/pharmaceutical_biotech/.)

and p53 pathways in conjunction with telo-mere shortening play significant roles in governing replicative senescence (Figure 4-1).

*Crisis stage.* Crisis (M2) is distinguished from senescence by cell death in the presence of ongoing cell division [29]. Although most cells that enter crisis die by apoptosis, a few cells overcome crisis and become immortal [30]. Immortal cells that have survived crisis typically exhibit aneuploidy and extensive nonreciprocal chromosomal translocations, which suggests that substantial genomic rearrangements accompany the selection of these rare surviving cells. These observations indicate that crisis is precipitated by critically shortened telomeres that have lost their ability to protect chromosomes in the setting of p53 loss. Immortal cells that emerge from crisis show the stabilized telomere lengths with extended passage, which suggests that only those cells that acquire the ability to maintain stable telomere lengths survive. It is important to note that, in general, immortal cells fail to form tumors in immunodeficient animals [31]. Experimentally, the introduction of *hTERT* in post-senescent, pre-crisis cells confers telomerase activity, stabilizes telomere

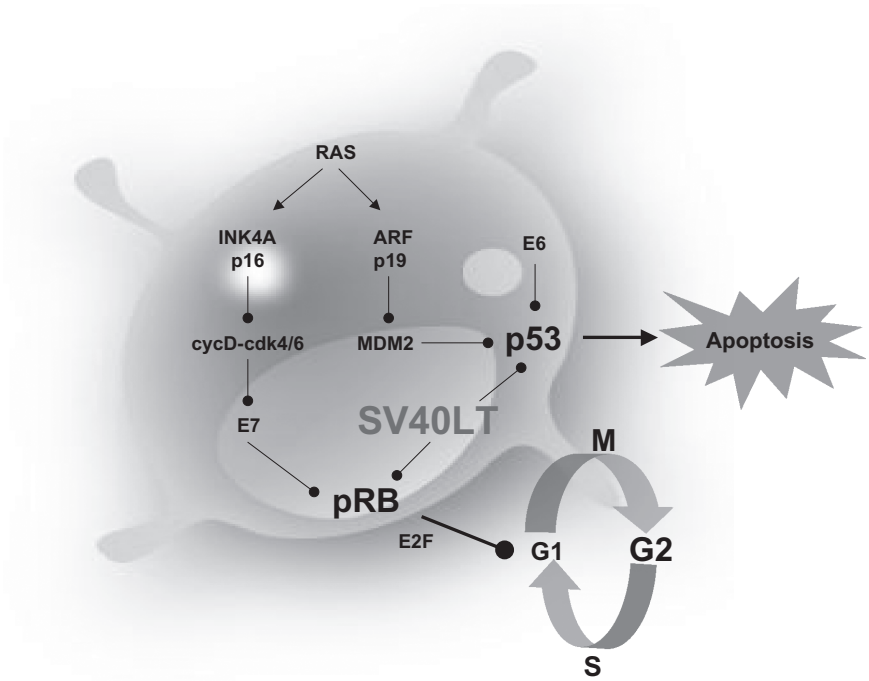


**Figure 4-4.** ***pRB and p53 tumor suppressors pathways.*** These pathways play critical roles in the cell physiology and mutations. Their affectation is found in most human cancers. The viral oncoprotein SV40 LT targets pRB and p53 pathways. (This figure is available in full color at ftp://ftp.wiley.com/public/sci_tech_med/pharmaceutical_biotech/.)

length, and permits these telomerase-expressing cells to become immortal [10, 27, 32–35]. Human cells immortalized in this manner often exhibit near-diploid karyotypes, which strongly suggests that critical telomere shortening in the absence of pRB and p53 functions initiates crisis. Two barriers limit proliferative lifespan in human cells: (1) replicative senescence and (2) crisis. Furthermore, it is obvious that telomerase, pRB, and p53 play critical roles in regulating an entry of human cells into these two states.

*Immortality.* Several groups have demonstrated that the introduction of pairs of cooperating oncogenes, such as myc and ras [36], or the adenoviral E1A protein and ras [37] into the primary rodent cells leads to direct transformation of the cells. However, several laboratories have now shown that telomere biology differs in important ways between human and murine cells [38]. Unlike most human cells in which telomerase is tightly repressed, most murine cells express detectable telomerase activity. In addition, telomeres are maintained at much longer lengths in cells derived from inbred mice than are observed in human cells [38]. These observations have suggested that the replicative life span of murine cells is not limited by telomere length. With extended cell division, telomere shortening

activates either replicative senescence or crisis and telomeres play in protecting chromosomal integrity, which suggests that telomeres play an important role in repressing tumor formation. In contrast, the maintenance of telomere length confers cell immortality. These observations also indicate that telomeres serve as a factor that restricts and/or promotes malignant transformation. Under most circumstances, short telomeres limit the life span of the cells, depending on the status of the pRB and p53 pathways, to greatly reduce the pool of premalignant cells (Figure 4-2).

To address cellular immortality, researchers have made great efforts to establish an immortalized human cell lines. Such cell lines provide the advantage of uniformity and sterility, grow in unlimited quantity, and are far less costly than isolating the primary cells. One approach to construct clonal cell lines is transduction of the primary cells with genes from DNA tumor viruses, such as SV40, human papillomavirus, and Epstein–Barr virus [39]. Transformation of normal cells with early region genes of SV40, typically by transfection with expression plasmids, remains a common immortalization technique. For example, human fetal hepatopcytes were successfully transduced with pSV3neo DNA containing both large and small T antigens of the early region of SV40 and bacterial neomycin phosphotransferase gene (Figure 4-3). After G418 treatment, one line of the surviving clones, OUMS-29, grew well in the chemically defined serum-free medium without any crisis and showed liver-specific functions [40]. When transplanted into immunodeficient mice, OUMS-29 cells were not tumorigenic. The potential risk of malignant formation of OUMS-29 cells cannot be precluded in humans. Safeguards, including the introduction of suicide genes, should be considered in immortalized cells for human application.

In parallel, cells are recovered from surgically resected tumor specimens and cultured. Cell lines have been established for the study of different pathophysiologies. These cell lines grow indefinitely in tissue culture and have been widely used all over the world. However, genetic constitution and alterations of such cells are often observed, and several important specific functions, including several receptors and transporters, have been unfortunately deregulated [41].

***4.1.2 Tightly Immortalized Human Cell Lines*** Transduction of immortalized cell lines with these suicide genes would provide a way to eliminate the cells after transplantation. Cells modified to express a herpes simplex virus–thymidine kinase (HSV–TK) gene become sensitive to ganciclovir (GCV) [42]. Thus, researchers have introduced the HSV–TK gene into cell lines. For example, OUMS-29/TK cells, immortalized human fetal hepatocytes expressing HSV–TK, were more than 100 times sensitive to GCV treatment than unmodified parental OUMS-29 cells. OUMS-29/TK cells stopped proliferation in the presence of 5-μM GCV [40].

***4.1.3 Conditionally Immortalized Cell Lines*** An approach to create a conditionally immortalized cell line is the use of a transforming gene containing a temperature-sensitive mutation. Primary rat hepatocytes were successfully immortalized with a thermolabile mutant SV40T (encoded by the early region mutant tsA58) and functioned as well as primary hepatocytes after transplantation [43–45]. However, the continued presence of SV40T in the transplanted cells may increase the risk of malignant transformation of the cells after transplantation in the recipients.

***4.1.4 Ultra-Transform Immortalized Transgenic Cell Lines*** This approach has proven to be successful because the resulting cell lines showed stability in culture and sensitivity to chemical exposure. The strategy implicates the development of a bi-transgenic hepatocyte cell line to evaluate the ability of various organic and inorganic chemicals to induce the expression of the HSP70-driven reporter gene. Development of two types of transgenic mice is necessary in advance. One is a transgenic mice (Hsp70/hGH) [46] secreting high levels of human growth hormone (hGH), and another is transgenic model (AT/cytoMet) [46] allowing the reproducible immortalization of untransformed hepatocytes retaining liver functions. Both strains are crossed, and the resulting transgenic strain permits a reproducible immortalization of untransformed hepatocytes. Several stable hepatic cell lines (MMH–GH) showing highly differentiated phenotypes have been generated from the double transgenic animals. This strategy is valuable in the field of toxicology and in the development of chemical and physical xenobiotics. The technology provides a simple biological system that reduces the need for animal experimentation and/or continuously isolating fresh hepatocytes [46].

***4.1.5 Reversibly Immortalized Cell Lines*** To generate an immortalized hepatocyte cell line more suitable for clinical use, a more tightly regulated system for cell growth should be considered. An attractive system using site-specific recombination has been documented [47]. DNA sequences intervened by loxP recombination targets can be excised after expression of Cre recombinase [47]. A Cre/loxP system allows the construction of reversibly immortalized cell lines. To provide more stringent control over the expression of transforming genes, human hepatocytes were transduced with a retro-viral vector SSR#69 expressing SV40T and selectable positive (hygromycin resistance gene) and negative (HSV-TK) markers that were intervened by a pair of loxP recombination targets and subsequently excised by Cre/loxP recombination [31, 48]. One emerging clone after SSR#69-transduction, NKNT-3, was a highly differentiated hepatocyte cell line. NKNT-3 cells were sensitive to 5-μM GCV. Adenovirus-mediated Cre recombinase expression was efficiently performed to remove SV40T from NKNT-3 cells. Intrasplenic transplantation of such reverted NKNT-3 cells significantly improved the survival of 90% hepatectomized rats [31]. Cre/loxP-based reversible immortalization has been achieved in several types of human cells [31, 48, 49]. In our studies, SV40T-transduced normal human endothelial cells (ECs) acquired an extensive proliferation capacity to population doubling level (PDL) 65 to 80, but complete immortalization was not achieved [49]. This was explained by the absence of spontaneous activation of endogenous telomerase, which is known as one of the essential participants in cellular immortalization processes in SV40T-transduced ECs [50–52]. To achieve immortalization of normal human ECs, a retroviral vector SSR#197 expressing hTERT and GFP cDNAs flanked by a pair of loxB target sequences was constructed. Cotransduction of human ECs with retroviral vectors SSR#69 and SSR#197 facilitated establishment of a completely immortalized cell line TMNK-1 that expresses a differentiated endothelial phenotype. Feasibility of reversible immortalization in TMNK-1 cells was demonstrated by using TAT-derived HIV-mediated Cre/loxP recombination followed by GFP-negative cell sorting and drug selection [27]. Lately, the system has been applied to establish hepatic stellate cell lines [33], cholangiocyte cell lines [10], hepatic progenitor cell lines [53], bone marrow-derived human cells [54], and human pancreatic beta cell lines [32]. To

establish immortalized human pancreatic beta cell lines, freshly isolated human pancreatic islet cells were transduced with SSR#69, followed by hygromycin selection. Then, the resultant cells were super-infected with SSR#197 for immortalization (Figure 4-5). At the first screening, tumorigenic assay of the resulting cell lines was performed in immunodeficient mice. Then, gene expression analysis was conducted in nontumorigenic clones. Based on these findings, NAKT-15 turned out to be highly differentiated. To obtain the reverted form of NAKT-15 cells, the cells were infected with a recombinant adevovirus virus vector (AxCANCre) expressing Cre recombinase tagged with a nuclear localization signal (NLS) expressing the recombinant adenovirus vector. After AxCANCre infection, GFP-

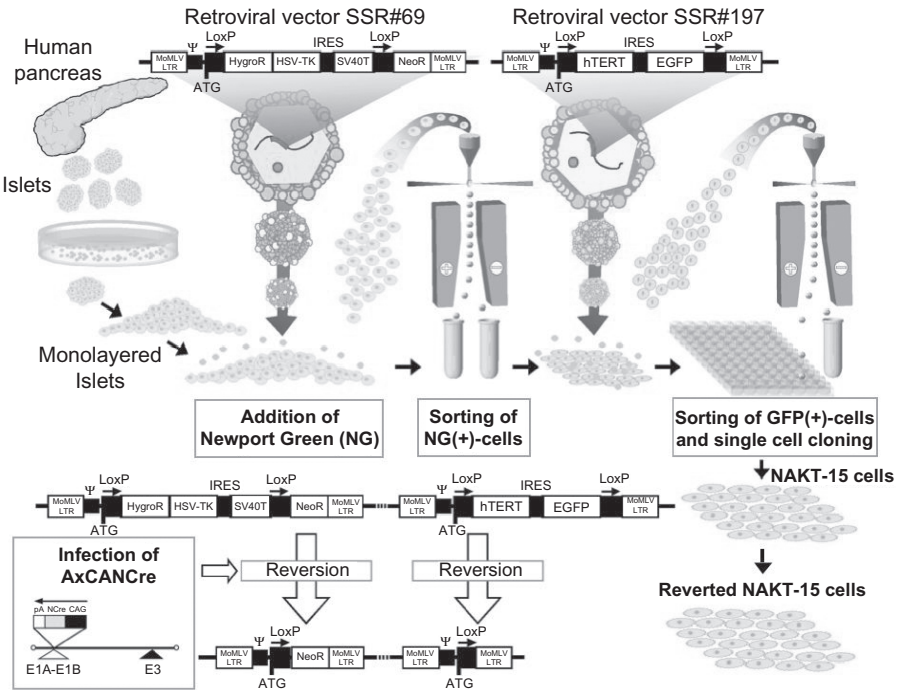


**Figure 4-5.** ***Scheme for the establishment of reversibly immortalized human β-cell lines.*** Freshly isolated human islets were cultured in monolayer and transduced with the retroviral vector SSR#69. After hygromycin selection of SSR#69-transduced cells, Newport Green dye was added and Newport Green-positive cells were sorted. After transduction with SSR#197, EGFP-positive cells were sorted and subjected to single-cell cloning, yielding the clone NAKT-15. Transient expression of Cre recombinase was induced by infecting cells with a recombinant adenovirus expressing Cre recombinase (AxCANCre), which removed the SV40T and TERT genes that were flanked by loxP sites. Neomycin (G418) selection and EGFP-negative cell sorting resulted in the collection of reverted NAKT-15 cells. CAG, promoter consisting of cytomegalovirus IE enhancer, chicken b-actin promoter, and rabbit b-globin polyadenylation signal; HSV-TK, herpes simplex virus thymidine kinase; HygroR, hygromycin-resistance gene; IRES, internal ribosomal entry site; MoMLV LTR, long terminal repeat of Moloney murine leukemia virus leader sequence; NCre, nuclear localizing signal-tagged Cre recombinase; pA, polyadenylation signal; j, packaging signal. (This figure is available in full color at ftp://ftp.wiley.com/public/sci_tech_med/pharmaceutical_biotech/.)

negative cell populations were recovered by a MoFlo cell sorter and then cultured in the presence of neomycin analog G418 [32]. These procedures are well illustrated in Figure 4-5. This strategy can be applicable to various types of human cells.

***4.1.6 Applications*** Numerous genomics-based technologies are now routinely applied to drug discovery. The central role for these technologies is validating the next generation of therapeutics from novel targets identified through genomics. The aim of such technologies is to accurately identify the next generation of targets that demonstrate therapeutic efficiency and safety. Immortal human cell lines will have unique attributes that can be used for drug discovery, and if cell therapy is the goal of such cells, the addition of redundant safeguards into the cells and accumulation of experimental data will be required before clinical trials. The ultimate goal of cell transplantation is an autologous setting in which the patient's own cells are genetically modified *ex vivo* with a reversible immortalization method, and then a reverted form of the cells can be transplanted back to the patient. Immortalized or reversibly immortalized cells offer continuous availability, uniformity, and sterility, and the cells can be functionally cryopreserved for the future use. Production of such cell lines may overcome the shortage of donors, and thus, the established human cell clones may be considered as a potential cell source for the treatment of diabetic patients or liver failure patients with transplantation. We anticipate that the production of artificial cells will be useful not only to provide carcinogenesis models, but also to develop drug discovery and possible cell therapy in the future. Reversible immortalization of human cells has important applications in biological research, biotechnology, and medicine.

## 4.2 Conclusions

The ethical issue of transplanting immortalized cells is still controversial, but clinical approval of a protocol involving transplantation of these cells depends on the balance of risks and benefits. If patients with high risk were to receive such transplants with considerable benefits, approval would be obtained after reliable accumulation of data regarding the safety and efficacy of immortalized cells. We here represent an important step in the development of a useful strategy for resolving the organ shortage that now limits the use of normal human cells for cell therapies. Such technology can be applicable to a variety of somatic cells and would potentially be used to treat a large number of patients with clinically significant pathologic conditions.

## References

1. McNeish J (2004). Embryonic stem cells in drug discovery. *Nat. Rev. Drug Discov.* 3:70–80.
2. van der Laan L J, Lockey C, Griffeth B C, et al. (2000). Infection by porcine endogenous retrovirus after islet xenotransplantation in SCID mice. *Nature.* 407:90–94.
3. Butler D (2002). Xenotransplant experts express caution over knockout piglets. *Nature.* 415:103–104.

4. Hayflick L, Moorhead P S (1961). The serial cultivation of human diploid cell strains. *Exp. Cell Res.* 25:585–621.
5. Masters J R (2000). Human cancer cell lines: Fact and fantasy. *Nat. Rev. Mol. Cell Biol.* 1:233–236.
6. Cooper C S, Park M, Blair D G, et al. (1984). Molecular cloning of a new transforming gene from a chemically transformed human cell line. *Nature.* 311:29–33.
7. Rhim J S, Park J B, Jay G (1989). Neoplastic transformation of human keratinocytes by polybrene-induced DNA-mediated transfer of an activated oncogene. *Oncogene.* 4:1403–1409.
8. Borek C (1980). X-ray induced in vitro neoplastic transformation of human diploid cells. *Nature.* 283:776–778.
9. Rhim J S, Jay G, Arnstein P, et al. (1985). Neoplastic transformation of human epidermal keratinocytes by AD12-SV40 and Kirsten sarcoma viruses. *Science.* 227:1250–1252.
10. Munger K, Howley P M (2002). Human papillomavirus immortalization and transformation functions. *Virus Res.* 89:213–228.
11. Maruyama M, Kobayashi N, Westerman K A, et al. (2004). Establishment of a highly differentiated immortalized human cholangiocyte cell line with SV40T and hTERT. *Transplant.* 77:446–451.
12. Sherwood S W, Rush D, Ellsworth J L, et al. (1988). Defining cellular senescence in IMR-90 cells: a flow cytometric analysis. *Proc. Natl. Acad. Sci. USA.* 85:9086–9090.
13. Goldstein S (1990). Replicative senescence: the human fibroblast comes of age. *Science.* 249:1129–1133.
14. Dimri G P, Lee X, Basile G, et al. (1995). A biomarker that identifies senescent human cells in culture and in aging skin in vivo. *Proc. Natl. Acad. Sci. USA.* 92:9363–9367.
15. Kipling D, Cooke H J (1990). Hypervariable ultra-long telomeres in mice. *Nature.* 347:400–402.
16. Cristofalo V J, Allen R G, Pignolo R J, et al. (1998). Relationship between donor age and the replicative lifespan of human cells in culture: a reevaluation. *Proc. Natl. Acad. Sci. USA.* 95:10614–10619.
17. Shein H M, Enders J F (1962). Transformation induced by simian virus 40 in human renal cell cultures. I. Morphology and growth characteristics. *Proc. Natl. Acad. Sci. USA.* 48:1164–1172.
18. Sweet B H, Hilleman M R (1960). The vacuolating virus, S.V. 40. *Proc. Soc. Exp. Biol. Med.* 105:420–427.
19. Fanning E (1992). Simian virus 40 large T antigen: the puzzle, the pieces, and the emerging picture. *J. Virol.* 66:1289–1293.
20. Ray F A, Kraemer P M (1993). Iterative chromosome mutation and selection as a mechanism of complete transformation of human diploid fibroblasts by SV40 T antigen. *Carcinogen.* 14:1511–1156.
21. Blackburn E H, Chiou S S (1981). Non-nucleosomal packaging of a tandemly repeated DNA sequence at termini of extrachromosomal DNA coding for rRNA in Tetrahymena. *Proc. Natl. Acad. Sci. USA.* 78:2263–2267.
22. Blackburn E H (2001). Switching and signaling at the telomere. *Cell.* 106:661–673.
23. Meyerson M (2000). Role of telomerase in normal and cancer cells. *J. Clin. Oncol.* 18:2626–2634.
24. Kim N W, Piatyszek M A, Prowse K R, et al. (1994). Specific association of human telomerase activity with immortal cells and cancer. *Science.* 266:2011–2015.

25. Feng J, Funk W D, Wang S S, et al. (1995). The RNA component of human telomerase. *Science.* 269:1236–1241.
26. Nakamura T M, Morin G B, Chapman K B, et al. (1997). Telomerase catalytic subunit homologs from fission yeast and human. *Science.* 277:955–959.
27. Matsumura T, Takesue M, Westerman K A, et al. (2004). Establishment of an immortalized human-liver endothelial cell line with SV40T and hTERT. *Transplant.* 77:1357–1365.
28. Counter C M, Avilion A A, LeFeuvre C E, et al. (1992). Telomere shortening associated with chromosome instability is arrested in immortal cells which express telomerase activity. *Embo. J.* 11:1921–1929.
29. Wei W, Sedivy J M (1999). Differentiation between senescence (M1) and crisis (M2) in human fibroblast cultures. *Exp. Cell. Res.* 253:519–522.
30. Macera-Bloch L, Houghton J, Lenahan M, et al. (2002). Termination of lifespan of SV40-transformed human fibroblasts in crisis is due to apoptosis. *J. Cell Physiol.* 190:332–344.
31. Kobayashi N, Fujiwara T, Westerman K A, et al. (2000). Prevention of acute liver failure in rats with reversibly immortalized human hepatocytes. *Science.* 287:1258–1262.
32. Narushima M, Kobayashi N, Okitsu T, et al. (2005). A human beta-cell line for transplantation therapy to control type 1 diabetes. *Nat. Biotechnol.* 10:1274–1282.
33. Watanabe T, Shibata N, Westerman K A, et al. (2003). Establishment of immortalized human hepatic stellate scavenger cells to develop bioartificial livers. *Transplant.* 75:1873–1880.
34. Shibata N, Watanabe T, Okitsu T, et al. (2003). Establishment of an immortalized human hepatic stellate cell line to develop antifibrotic therapies. *Cell Transplant.* 12:499–507.
35. Okitsu T, Kobayashi N, Jun H S, et al. (2004). Transplantation of reversibly immortalized insulin-secreting human hepatocytes controls diabetes in pancreatectomized pigs. *Diabetes.* 53:105–112.
36. Land H, Parada L F, Weinberg R A (1983). Tumorigenic conversion of primary embryo fibroblasts requires at least two cooperating oncogenes. *Nature.* 304:596–602.
37. Ruley H E (1983). Adenovirus early region 1A enables viral and cellular transforming genes to transform primary cells in culture. *Nature.* 304:602–606.
38. Wright W E, Shay J W (2000). Telomere dynamics in cancer progression and prevention: fundamental differences in human and mouse telomere biology. *Nat. Med.* 6:849–851.
39. Katakura Y, Alam S, Shirahata S (1998). Immortalization by gene transfection. *Methods Cell Biol.* 57:69–91.
40. Kobayashi N, Miyazaki M, Fukaya K, et al. (2000). Transplantation of highly differentiated immortalized human hepatocytes to treat acute liver failure. *Transplant.* 69: 202–207.
41. Blumrich M, Zeyen-Blumrich U, Pagels P, et al. (1994). Immortalization of rat hepatocytes by fusion with hepatoma cells. II. Studies on the transport and synthesis of bile acids in hepatocytoma (HPCT) cells. *Eur. J. Cell Biol.* 64:339–347.
42. Culver K W, Ram Z, Wallbridge S, et al. (1992). In vivo gene transfer with retroviral vector-producer cells for treatment of experimental brain tumors. *Science.* 256: 1550–1552.
43. Fox I J, Chowdhury N R, Gupta S, et al. (1995). Conditional immortalization of Gunn rat hepatocytes: an ex vivo model for evaluating methods for bilirubin-UDP-glucuronosyltransferase gene transfer. *Hepatol.* 21:837–846.

44. Schumacher I K, Okamoto T, Kim B H, et al. (1996). Transplantation of conditionally immortalized hepatocytes to treat hepatic encephalopathy. *Hepatol.* 24:337–743.
45. Nakamura J, Okamoto T, Schumacher I K, et al. (1997). Treatment of surgically induced acute liver failure by transplantation of conditionally immortalized hepatocytes. *Transplant.* 63:1541–1547.
46. Sacco M G, Amicone L, Cato E M, et al. (2004). Cell-based assay for the detection of chemically induced cellular stress by immortalized untransformed transgenic hepatocytes. *BMC Biotechnol.* 4:5.
47. Sternberg N, Hamilton D, Austin S, et al. (1981). Site-specific recombination and its role in the life cycle of bacteriophage P1. *Cold Spring Harb. Symp. Quant. Biol.* 45(Pt 1):297–309.
48. Westerman K A, Leboulch P (1996). Reversible immortalization of mammalian cells mediated by retroviral transfer and site-specific recombination. *Proc. Natl. Acad. Sci. USA.* 93:8971–8976.
49. Noguchi H, Kobayashi N, Westerman K A, et al. (2002). Controlled expansion of human endothelial cell populations by Cre-loxP-based reversible immortalization. *Hum. Gene. Ther.* 13:321–334.
50. Bodnar A G, Ouellette M, Frolkis M, et al. (1998). Extension of life-span by introduction of telomerase into normal human cells. *Science.* 279:349–352.
51. Halvorsen T L, Leibowitz G, Levine F (1999). Telomerase activity is sufficient to allow transformed cells to escape from crisis. *Mol. Cell Biol.* 19:1864–1870.
52. Zhu J, Wang H, Bishop J M, et al. (1999). Telomerase extends the lifespan of virus-transformed human cells without net telomere lengthening. *Proc. Natl. Acad. Sci. USA.* 96:3723–3728.
53. Delgado J P, Parouchev A, Allain J E, et al. (2005). Long-term controlled immortalization of a primate hepatic progenitor cell line after Simian virus 40 T-Antigen gene transfer. *Oncogene.* 24:541–551.
54. Nishioka K, Fujimori Y, Hashimoto-Tamaoki T, et al. (2003). Immortalization of bone marrow-derived human mesenchymal stem cells by removable simian virus 40T antigen gene: analysis of the ability to support expansion of cord blood hematopoietic progenitor cells. *Int. J. Oncol.* 23:925–932.

# REGULATION OF SMALL-MOLECULE DRUGS VERSUS BIOLOGICALS VERSUS BIOTECH PRODUCTS

María de los Angeles Cortés Castillo and José Luis Di Fabio
*Pan American Health Organization, Washington, DC*

**Chapter Contents**

1 Historical Perspective on Regulation
2 Issues Related to Regulation for Biologicals (Vaccines) Compared with Regulation for Other Medicinal Products
  2.1 Regulation of Small Molecules
  2.2 Regulation of Classic Vaccines and Biologicals
  2.3 Regulation of Biotechnology Products
3 Main Regulatory Agencies
  3.1 Food and Drug Administration (FDA)
  3.2 European Medicines Agency (EMEA)
  3.3 International Committee for Harmonization (ICH)
4 Global Role of WHO in Drugs and Biologicals Regulation
  4.1 WHO Procedure for Assessing Drug Regulatory Authorities
5 Other Initiatives for Harmonized Regulation
6 Future Perspective of Regulation
  References

## 1 HISTORICAL PERSPECTIVE ON REGULATION

Regulation is the logical consequence of the search from health institutions and governments to ensure public protection from unsafe products.

The regulation of drugs emerged first in developed countries. The first initiatives of regulation appeared in 1862 with the Bureau of Chemistry in the Department of Agriculture of the United States, lead by Harvey Wiley, trying to regulate food,

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

drug, and biological products. At that time, adulterated food and drugs as The Great Blood Purifier System (Peter's specific) and Mrs. Winslow's Soothing Syrup for teething and colicky babies (which was related with morphine, without a label indication) were the main targets of regulation [1].

The Bureau of Chemistry enforced a 1906 law until 1927, when it was reorganized to form the Food, Drug and Insecticide Administration, later renamed, in 1931, as the Food and Drug Administration [1]. The 1906 Act prohibited the sale of adulterated or misbranded drugs and prohibited interstate transport of disallowed food and drugs. It also mandated that a variety of information be stated on the drug package. Changes to labeling requirements and a provision to label certain dangerous ingredients led many patent medicine makers to reformulate their products because many of the products contained harmful chemicals [1].

After an event in 1937, when a marketed anti-infective called Elixir Sulfanimide caused over 100 deaths regulators were forced to change the law to require that drugs meet certain safety requirements before being marketed. A new 1938 Act called for evidence of drug safety, prohibited false therapeutic drug claims, and also included cosmetics and therapeutic devices under FDA regulation. The 1938 Act established a listing of active ingredients on the drug label and required that drugs be labeled with adequate directions for safe use [1].

Even after all these efforts to ensure the safety of drugs were implemented, many incidents (included the thalidomide tragedy) occurred and the U.S. Congress had to pass major amendments to the Act in 1962. In addition to safety issues, for the first time drug manufacturers were required to prove the effectiveness of their drug products before marketing.

In parallel with the development of drug regulations, biological regulations were historically developed in connection to safety concerns. The 1902 Biologics Control Act enacted by the U.S. Congress established the initial concepts used for the regulation of biologicals. These provisions were revised in 1944 and expanded in the 1950s. Attention to regulation of vaccines and toxins was enforced until a major tragedy occurred in the United States, and only then were actions taken by the federal government to ensure public protection from unsafe products. The incident occurred in 1901, in St. Louis, Missouri, when 20 children became ill and 14 died after administration of an equine-derived diphtheria antitoxin (prepared from horse serum) contaminated with tetanus toxin. The serum had been manufactured in local establishments designed without any of the current knowledge of Good Manufacturing Practices, particularly without procedures that guarantee the product's safety and potency. This event forced legislation to regulate the sale of biologicals. In July 1902, Congress passed the Biologics Control Act also known as the "Virus-Toxin" law [2], which is the first time that the United States' government took control over biologicals' production, consequently with the responsibility to ensure their safety for the public. In this act, the need to include the control of the manufacturing establishments and facilities was recognized, as assurance of purity could not be demonstrated if control was limited to inspection and testing of a final product. This act settled the essential concepts for ensuring the safety of biologicals. These ideas are used as the basis for ensuring safety and effectiveness of biologicals throughout the world [3].

The history of vaccine control and regulation in developed nations has been one of increasing complexity. By 1955, many biologicals, including blood products and

vaccines, had been licensed and the need for appropriate regulation was evident. In the United States, the use of inactivated polio vaccines from different manufacturers, licensed in the country, provoked a number of children to develop poliomyelitis, and some of the children died. The reason was incompletely inactivated vaccines [3]. This incident led to the expansion of the biologicals control function to the establishment of the Division of Biological Standards (DBS) within the National Institutes of Health. Later, in 1972, the DBS was transferred to FDA and became the Bureau of Biologics. In 1982, the Bureau of Biologics was renamed the Office of Biologics Research and Review (OBRR) and combined with the Office of Drugs Research and Review (ODRR) formed the Center for Drugs and Biologics. In 1987, the OBRR was separated and renamed the Center for Biologics Evaluation and Research (CBER) [4].

The need to develop and institute an independent evaluation of medicinal products before they are allowed into the market for public consumption was reached at different times in different regions, generally associated with the level of technological development. In Japan, government regulations requiring all medicinal products to be registered for sale started in the 1950s. In many countries in Europe, the thalidomide tragedy of the 1960s was the trigger. This tragedy revealed that the new generation of synthetic drugs, which were revolutionizing medicine at the time, had the potential to harm as well as to cure [5].

Between 1960 and 1970, important developments transpired regarding laws, regulations, and guidelines for reporting and evaluation of data on safety, quality, and efficacy of new medicinal products. At the same time, the industry was becoming more international and seeking new global markets, but still the registration of medicines remained a national responsibility.

The urgent need to rationalize and harmonize regulation has been forced by the need of countries to have access to drugs and biologicals to cure, treat, and prevent diseases according to the development of important technological advances in science and responding to the effort of manufacturers to gain access to more and new markets.

For the purpose of this review, drugs will be classified as small molecules, biologicals, and biotech products:

- Small molecules are most medicines or drugs chemically synthesized with molecular weights of typically less than 500 Da applicable to prevention, treatment, or cure of diseases or injuries in man.
- Biologicals are any virus, therapeutic serum, toxin, antitoxin, vaccine, blood component or derivative, allergenic extract, or analogous product applicable to the prevention, treatment, or cure of diseases or injuries in man, generally with molecular weights larger than thousands of Daltons.
- Biotechnological products are any biological or pharmaceutical product made using recombinant DNA techniques and intended to prevent, treat, or cure diseases or injuries in man. This group includes mainly proteins of high molecular weight.

Drugs or small molecules of natural origin that are not produced through chemical synthesis are not considered in this review.

## 2 ISSUES RELATED TO REGULATION FOR BIOLOGICALS (VACCINES) COMPARED WITH REGULATION FOR OTHER MEDICINAL PRODUCTS

WHO has established that effective drug regulation promotes and protects public health by ensuring that [6]:

- medicines are of the required quality, safety, and efficacy;
- health professionals and patients have the necessary information to enable them to use the medicines rationally;
- all premises, persons, and practices engaged in the development, manufacture, importation, exportation, wholesale, supply, dispensing, and promotion of drugs comply with approved standards, norms, procedures, and requirements;
- product information is unbiased, accurate, and appropriate;
- illegal manufacturing and trade are detected and adequately sanctioned;
- promotion and advertising are fair, balanced, and aimed at rational drug use; and
- access to medicines is not hindered by unjustified regulatory work.

National drug budgets, as a proportion of total health budgets, currently range from 7% to 66% worldwide. The proportion is higher in developing countries (24–66%) than in developed countries (7–30%) [7]. People and governments willingly spend money on drugs because of the role they can play in saving lives, restoring health, preventing diseases, and stopping epidemics. However, to accomplish these goals, drugs must be safe, effective, and of good quality and used appropriately, which means, in turn, that their development, production, importation, exportation, and subsequent distribution must be regulated to ensure that they meet prescribed standards.

In the search of implementing regulations for drugs, biologicals, and all medicines to guarantee the protection and benefit of the users, governments have established the national drug regulatory authorities (DRAs). These institutions must be developed with a clear mission, solid legal basis, realistic objectives, appropriate organizational structure, adequate number of qualified staff, sustainable financing, access to technical literature and information, and with the capacity to exert effective market control. DRAs must be accountable to both the government and the public and their decision-making processes should be transparent. Monitoring and evaluating mechanisms should be built into the regulatory system to assess attainment of established objectives [6].

Regulatory authority is generally based on laws, which represent policy choices. This authority is assigned to designated organizations, generally part of the bureaucratic institutions, with the mission to carry out drug regulation. Several factors regarding the authority and the capacity for exercising such authority affect the operation and drug regulatory activities [7]. Among the most important are:

- availability of standards, procedures, and guidelines to be used as guidance in performing the authorized functions;

- consideration of strategies to overcome structural and resources constraints, including financial adequacy and sustainability;
- establishment of clear strategies for planning, monitoring, and evaluation;
- availability of appropriate human resources including number, qualifications, remuneration, and human resources development; and
- definition of the scope of regulatory authority, in relation to type and extent of the activities carried out to implement legal provisions, including sanctions for noncompliance.

The value of drug regulatory activities depends on whether they produce the intended outcomes in terms of the following [7]:

- the quality of pharmaceutical products marketed;
- the proportion of licensed pharmaceutical facilities meeting international standards in Good Manufacturing Practices, Good Laboratory Practices, Good Clinical Practices, and so on;
- the number of illegal products on the market; and
- the number of illegal facilities marketing pharmaceutical products.

In addition to effectiveness, policy makers must also address regulatory efficiency, transparency, and accountability when evaluating regulatory policies. The cost effectiveness of drug regulation; the cost for pharmaceutical businesses and consumers of regulatory delays; political and commercial influence over regulatory decisions, public access to regulatory procedures and decision-making criteria; communication among the regulatory authority, its clients, and the consumer; and the accountability for the results of regulatory actions are performance indicators that must be assessed in the evaluation of regulatory agencies.

### 2.1 Regulation of Small Molecules

Since the mid-1930s, many new pharmaceutical products have emerged, and trade in the pharmaceutical industry has taken on international dimensions. At the same time, however, the circulation of toxic, substandard, and counterfeit drugs on the markets has increased, mainly because of ineffective regulation of production and trade of pharmaceutical products in exporting and importing countries. Problems relating to drug safety and efficacy are generally caused by the use of drugs containing toxic substances or impurities, drugs whose claims have not been verified or that have unknown severe adverse reactions, as well as substandard preparations or counterfeits drugs. All these problems can be tackled effectively only by establishing an effective drug regulatory system.

The development process stages for a small molecule drug that are relevant to the regulatory scrutiny are as follows:

- drug discovery,
- drug purification,
- preclinical research, and
- clinical trials (phase I, II, and III).

When the dossier of a drug is submitted to the DRA, an exhaustive review of the medical/clinical, chemistry/manufacture, pharmacology/toxicology aspects, and their statistical validity is conducted. The main element of the evaluation is to certify its safety (clinical trials phase I or II) and, during this stage of the approval, the manufacturer can be required to submit additional information to ensure the safety of the product. When safety is acceptable, the DRA proceeds with complete reviews to ensure efficacy, stability, and all other specifications that guarantee the product complies with the required norms. At any stage of the trial, the DRA of the country where the trial is being conducted should have the authority to put it on hold if deficiencies or safety issues appear and until they are resolved. If results of clinical trials demonstrate that the drug is safe and efficacious over existing treatment drugs, an application is made in most of the regulatory agencies to seek approval for marketing the drug.

As part of the new drug approval process, the DRA conducts inspection for Good Manufacturing Practices (GMP inspection) of the facilities where the drug is being manufactured. GMP is a quality concept and consists of a set of policies and procedures for manufacturing processes to guarantee that drugs are pure, consistent, safe, and effective. These policies and procedures describe the facilities, equipment, methods, and controls for producing drugs with the intended quality. The guiding principle for GMP is that quality cannot be tested in a product, but must be designed and built into each batch of the drug product throughout all aspects of its manufacturing processes [8].

## 2.2 Regulation of Classic Vaccines and Biologicals

Well-managed programs of vaccination have brought about profound reductions in the impact of diseases in terms of morbidity and mortality in the majority of countries of the world. One major disease, smallpox, has been totally eradicated largely as a result of vaccination and another, poliomyelitis, is absent from large areas of the world and on target for eradication during the next few years, and the incidence of measles has been greatly reduced. Vaccines provide one of the most cost-effective public health interventions and are among the safest medicinal products [9] (Table 1).

These enormous achievements have been possible in part because internationally agreed principles and procedures are in place to secure high levels of safety, efficacy, and quality of vaccines. Vaccines differ from therapeutic medicines first because of the biological and thus inherently variable nature of the products themselves, the raw material used in their production, and the biological methods used to test them. Thus, special expertise and procedures are needed for their manufacture, control, and regulation. The use of appropriate standard materials and reference preparations, whenever they exist, is fundamental to the standardization and control of vaccines [9].

Vaccines are unique in the fact that they are usually administered to very large numbers of healthy people, mostly infants, in national immunization programs. Thus, safety and quality are of high relevance. Although vaccines have a key role in preventive medicine, recent history of their use has shown a general high level of safety compared with their benefit. In most cases, minor adverse reactions may occur, but these reactions do not challenge the risk-benefit advantage of vaccina-

**TABLE 1. Classification and Characteristics of Vaccines**

- **Classic vaccines:** products of a whole or part (subunit) of a microorganism. They can be attenuated or inactivated vaccines, subunits, or toxoids products of the microorganism's metabolism, for instance, Diphtheria-Tetanus-Pertussis (DTP), Measles–Mumps–Rubella (MMR), oral polio vaccine (OPV), inactivated polio vaccine (IPV), smallpox, rotavirus, and so on.
- **Recombinant vaccines:** obtained as products of gene modification of different organisms (yeast or bacteria) with posterior purification to get the immunogenic protein (example Hepatitis B vaccine).
- **Synthetic vaccines:** Part or the whole antigen is a product of chemical synthesis, proteic nature (peptides and oligopeptides), or carbohydrates (saccharides and oligosaccharides) (polyribosylribitol phosphate (PRP) synthetic antigen conjugated to tetanus toxoid).
- **Combined vaccines**: including different combination of classic, recombinant, or synthetic vaccines (DTP-HepB-Hib).
- New approach of vaccines: DNA vaccines and plant-derived vaccines, under development but still not licensed.

tion. A number of potential and theoretical risks are implicit in their use. They include, in particular, the presence of adventitious agents derived from the source materials or introduced during manufacture or, as is the case of live vaccines, the presence of virulent organisms caused by reversions of the vaccine virus or bacteria to its virulent form. In these cases, a possible risk to the community at large may exist in addition to the vaccinees [9].

A remarkable difference between vaccines and biologicals and other synthetic pharmaceutical drugs consists of the difficulty of evaluating the activity for those biological products, and thus the differences among them in their control and regulation. In general, for synthetic drugs and medicinal products, the active ingredient is totally defined and the activity can be evaluated by quantitative methods. They are synthesized in the laboratory in a precisely controlled way by a defined sequence of chemical reactions, and the nature of these molecules is such that, once they are made, their exact structure and composition can be determined by laboratory tests. Today, technology allows the modeling of small drug-receptor interactions that can be studied using computational chemistry (*in silico*) methods and extraction of valuable data and information from databases (bioinformatics). The modeled drug is then synthesized using combinatorial chemistry. The application of technologies such as X-ray crystallography and Nuclear Magnetic Resonance [8] are standard procedures for the 3D structural determinations and drug-receptor interactions.

For biologicals, the biological activity and its consequences must be evaluated in an integrated way before starting clinical trials. The complexity and challenge that regulators face with vaccines can be exemplified with the case of combined vaccines, where different epitopes are being introduced in the human body, inducing different responses to the regulatory mechanisms of the immune system, while not causing interference in the responses between them. New vaccines challenge regulators with the development of novel procedures for the proper evaluation of their efficacy and safety.

It is important that the standardization and control of vaccines by the national regulatory authority and the manufacturer are continually reviewed and modified

so as to reflect the current state of science and technology incorporating an improved understanding of quality and safety issues. Regulatory authorities must thus be proactive and maintain an acute awareness of scientific developments in the vaccine field [9].

### 2.3 Regulation of Biotechnology Products

Recombinant insulin was the first biotechnology product, which emerged in 1982. Today, biopharmaceutical products have diverged to encompass not only recombinant forms of natural proteins and biologicals derived from natural sources, including recombinant plant-derived pharmaceutical proteins, but also monoclonal antibody (mAb)-based therapeutics. All these products require special consideration from a regulatory point of view.

In the past, microbial and animal cell culture was the most common method of producing recombinant proteins. Currently, the interest is focusing increasingly on transgenic animals because of their apparent ability to produce complex proteins at a high volume and a low cost [6]. Many of these products are now in different phases of clinical trials, although some of them without encouraging results [10].

During 2002 and 2003, regulators in North America and Europe approved a total of 64 biopharmaceuticals for human use. These approved drugs included hormones, blood factors, thrombolytics, vaccines, interferons, monoclonal antibodies, and therapeutic enzymes (not including new indications for products already approved) [11]. Data published by Walsh [11] suggest that during the past three years, over a quarter of all new drug approvals were biopharmaceuticals, taking into account that these numbers do not reflect duplication for products with different formulations and different commercial names for the same active ingredient. All of the new biopharmacuticals approved in United States in 2003 are protein-based drugs (rather than nucleic acid-based drugs). According to estimates at that time, around 500 biopharmaceuticals were undergoing clinical trials [11].

To date, no gene therapy drug has gained marketing approval as a direct consequence of the technical, manufacturing, and regulatory difficulty that still beset this class of therapeutic compounds. Besides, it is important to consider that a bad reputation on these products exists as a consequence of reported incidents of adverse reactions after their use [11]. Monoclonal antibody therapeutics have been hit particularly hard regarding warnings mainly related to hypersensitivity or anaphylactic reactions or cytokine storms [12].

The major target indications of biopharmaceuticals currently undergoing clinical trials include cancer, cardiovascular, and infectious diseases, the major killers within the developed world. Although in excess of 30 nucleic acid-based drugs are currently being evaluated for gene therapy, vaccines, and other applications, the majority remain in the early stages of clinical development (phase I/II) [11].

Another kind of biotechnological product that is acquiring interest as a therapeutic agent is the biopharmaceuticals derived from plants. As a result of the limited production capacity of many pharmaceutical products, the production in plants represents an innovative tool that can expand the yields of active principles because the large volume of biomass that can be developed at the fields. An extra advantage is that these products are free of human diseases and mammalian viral vectors. In products derived from bacteria, the bacterial endotoxin content of the

final product has always been of concern, which is not an issue in products derived from plants.

To date, only proteins have been developed through genetically modified plant, including mammalian antibodies, blood components, coagulation factors, vaccines (edible vaccines), hormones (insulin, somatotropin, and erythropoietin), various interferons, and other therapeutics agents such as enzymes and interleukins [13].

Monoclonal antibodies for therapeutic and diagnostic use seem to be the best potential successful products produced in plants, but they currently have not reached the marketplace. Tobacco plants are the most commonly used plant in these developments. Hepatitis B surface protein and rabies virus glycoprotein are examples of advances in edible vaccines.

In the production of biopharmaceuticals derived from genetically modified plants, several steps are part of the process, and they commonly include:

- selection of host plant;
- selection of the gene expression system;
- location of gene expression within the plant (green matter, the seed, or other tissues);
- production of biomass; and
- implementation of purification systems for the protein.

Several of these steps will require new and special regulatory developments. The products obtained from genetically modified plants must be carefully studied by regulators focusing on some main issues such as the development of Good Manufacturing Practices appropriate for the use in the fields that can guarantee the consistency of production during the manufacturing process. Environmental issues also exist that must be taken into account. Potential effects on nontarget species such as butterflies, honeybees, and other wildlife at or near the growing sites are important to be considered [13]. The possible transgene escape through pollen or seed dispersal, the potential for recombinant molecules to enter the food chain [14], and the proper evaluation of the impact that the introduction of virus from plants into the human body can cause are issues that have to be resolved for the success of this modern biotechnology tool.

Table 2 summarizes some of the main differential issues regarding small molecules, biologicals, and biotechnological products.

## 3 MAIN REGULATORY AGENCIES

### 3.1 Food and Drug Administration (FDA)

FDA is recognized as the strongest regulatory agency in the world. Although the agency has limited its regulatory functions to products to be used in the United States, this agency has been the pioneer in the development of the regulation for drugs, including vaccines, blood products, and a lot of new biotechnology products emerging in the field that are used worldwide by other regulatory agencies.

Since 1962, the key elements of product safety and efficacy have not changed, although the agency's level of participation, collaboration, and transparency has

**TABLE 2. Major Differences that Impact the Regulation of Small Products versus Biologicals versus Biotech Products**

| Activity or issue | Small molecules | Biologicals | Biotech products |
|---|---|---|---|
| Control of starting materials | Generally purified and well-characterized chemicals | Most of culture media for fermentation processes are not totally synthetic.<br>Use of animal-derived additives (fetal calf serum) is requested | Requirements for: Baseline data on the host and the gene (vector) used for production; stability of the host-vector expression system under storage and recovery conditions |
| Control of the manufacturing process | In general, well-established procedures are available for synthesis and purification of drugs and medicines | Great diversity of production and purification procedures, depending on strain of origin.<br>Procedures developed case by case | Great diversity of production processes depending on the host and gene used.<br>Requirements for control at genetic, post-transcriptional, post-translational level or during production and purification |
| Quality control tests | Well-established qualitative and quantitative control tests are available to measure the activity of medicines and drugs.<br>Mainly required for final product | Difficulties in the evaluation of biological activity.<br>In many cases, "in process" quality control tests are requested to ensure the safety and effectiveness of the final product.<br>It is very important that the correlation of biological activity with efficacy is shown in clinical trials.<br>Need to show absence of potentially hazardous contaminants from animal origin, cell substrates, or additives | Similar to biologicals, "in process" quality control tests are requested to ensure the safety and effectiveness of the product.<br>Need to show absence of potentially hazardous contaminants mainly oncogenic DNA |

increased as demonstrated with review goal dates, meetings with industry, guidance documents, and public and private partnerships. Today, according to FDA estimates, 10,510 approved drugs and more than 100,000 over-the-counter drugs are on the market [15].

As discussed by Ray and Stein [16], "the current regulatory process does not have systematic provisions for obtaining important data needed to guide clinical practice. Planned data collection happens almost exclusively during pre-marketing testing. The FDA approves medications on the basis of studies of limited duration that include relatively small numbers of patients who are often healthier than the target populations for the new drug. Although many important effects of a new medication almost certainly will be unknown at the time of licensing, there are no systematic provisions for post-marketing studies."

This problem is common for regulatory agencies around the world. The lack of information presented by manufacturers regarding the safety during the use by populations at large, the identification of the best indicators for efficacy, and the differentiation that manufacturers try to obtain between products with the same active ingredient, frequently make the regulation function difficult to perform. In addition, in developing countries, the lack of expertise in the different areas of regulation, particularly on clinical trial evaluation and interpretation of statistical results, increases the high uncertainty of the regulatory agencies' results.

### 3.2 European Medicines Agency (EMEA)

The European Medicines Agency (EMEA) is a decentralized body of the European Union created with the objective to harmonize regulatory activities in Europe. EMEA began its activities in 1995, when the European system for authorizing medicinal products was introduced, providing for a centralized and a mutual recognition procedure. Its main responsibility is the protection and promotion of public and animal health through the evaluation and supervision of medicines for human and veterinary use.

Before 1995, the legal framework in Europe for drugs was 25 independent national regulatory authorities and thus 25 parallel national review processes, with the same number of independent marketing authorizations. The consequence was poor resource utilization, divergent scientific opinions, and confusion by physicians, nurses, and health workers in general when decisions had to be made. At the present time, EMEA coordinates the evaluation and supervision of medicinal products throughout the European Union. The Agency brings together the scientific expertise of the 25 European Union Member States in a network of 42 national competent authorities. It cooperates closely with the international partners, reinforcing the European Union contribution to global harmonization [17].

At the beginning, the EMEA set up two procedures for product authorization in Europe, the centralized and decentralized procedures.

The centralized procedure is as follows [17]:

- submission of the marketing authorization application to EMEA;
- a unique evaluation is carried out through the Committee for Medicinal Products for Human Use (CHMP);

- if the committee concludes that the product has proven its quality, safety, and efficacy, it adopts a positive opinion; and
- the positive opinion is sent to the commission to be transformed into a single market authorization valid for the whole of the European Union.

The decentralized procedure establishes a mutual recognition between Member States. Briefly, this procedure is based on:

- a marketing license application is submitted to one Member State;
- if the national regulatory authority concludes that the product has proven its quality, safety, and efficacy, it grants a marketing authorization to the product; and
- the marketing authorization is extended to one or more additional Member States.

Today, EMEA uses both procedures, but is primarily involved in the centralized procedure [17]. The regulation 2309/93 has defined a mandatory centralized procedure for medicinal products developed by means of one of the following biotechnological processes: recombinant DNA, controlled expression of genes coding, hybridoma, and monoclonal antibody methods. In addition, after 2005, the regulation establishes the mandatory centralized procedure for medicinal products containing new active substances for AIDS, cancer, diabetes, and neurodegenerative diseases, including orphan designated medicinal products. The regulation still allows the mutual recognition procedure for generic centralized products or for medicinal products with a significant therapeutic benefit, products of scientific or technical innovation, or products answering the interest of patients or animal health at the community level.

The fact that more and more products are being included in the mandatory centralized procedure is indicative of a general reliance on this procedure more than on the mutual recognition procedure; it can also reflect that not all national regulatory authorities have developed the same regulatory capabilities to authorize the marketing of new products for the whole European Community.

Although European countries have faced many challenges in the establishment of regulatory activities of the agency, mainly because the diverse conflicting priorities between the Member States, the centralized procedure has generated harmonization in the scientific opinion in Europe, getting one marketing authorization valid in the European Community, one common name, and, most important for the impact on health professionals and users (potentially 370 million), one common product information.

After 10 years of been established, EMEA is a good example of harmonization, with both centralized and mutual recognition procedures, that can be extrapolated to other regions of the world, mainly in the developing world where the few countries that have developed strong regulatory systems can share technical knowledge and expertise in the different areas of regulations with those countries in the corresponding regions with limited resources to the benefit of the whole region.

### 3.3 International Committee for Harmonization (ICH)

The ICH is not a regulatory agency, it is a unique project that brings together the regulatory authorities of Europe, Japan, and the United States and experts from the pharmaceutical industry in the three regions to discuss scientific and technical aspects of product registration.

The purpose is to make recommendations on ways to achieve greater harmonization in the interpretation and application of technical guidelines and requirements for product registration to reduce or obviate the need to duplicate the testing carried out during the research and development of new medicines. The objective of such harmonization is a more economical use of human, animal, and material resources and the elimination of unnecessary delay in the global development and availability of new medicines while maintaining safeguards on quality, safety, and efficacy and meeting regulatory obligations to protect public health. The guidelines established by ICH are followed by many countries around the world, sometimes as the only guide available to establish harmonized standards on quality, safety, and efficacy for their products. However, its impact in classic vaccine regulation is limited, being more developed regarding biotechnological products.

Six founding members of ICH are directly involved in the decision-making process. They represent the regulatory bodies and the industry in the European Union, Japan, and the United States. They are the European Commission and the European Federation of Pharmaceutical Industries and Associations (EFPIA) in Europe; the Ministry of Health Labor and Welfare and the Japan Pharmaceutical Manufacturers Association (JPMA) in Japan; the FDA and the Pharmaceutical Research and Manufacturers of America (PhRMA); and a group of observers including WHO, The European Free Trade Area (EFTA represented at ICH by Swissmedic Switzerland), and Canada represented by Health Canada.

## 4 GLOBAL ROLE OF WHO IN DRUGS AND BIOLOGICALS REGULATION

Although WHO is not a regulatory agency, it provides guidelines for acceptability of products taking in consideration the specifications needed in regions where the regulatory institutions have not fully developed. With the responsibility to guarantee the quality, safety, and efficacy of products all around the world, WHO develops guidelines joining efforts from experts in the most diverse regions, getting a generalized opinion about the best way to establish technical guidelines. WHO provides relevant expertise ant technical assistance through such activities as guidelines development; workshops and training courses; coordination and promotion of anti-counterfeiting measures; prequalification of medicines for priority diseases; pharmacovigilance for global medicine safety, regulatory, and other information exchange; and review of narcotic and psychotropic substances for scheduling within the 1961 Convention on Narcotic Drugs and the 1971 Convention on Psychotropic Substances.

### 4.1 WHO Procedure for Assessing Drug Regulatory Authorities

As WHO establishes [9], the overall objective of a DRA is to ensure that medicinal products (pharmaceuticals, biologicals including vaccines, blood products, and other biologicals) are of acceptable quality, safety, and efficacy and that they are manufactured and distributed in ways that ensure their quality until they reach the patient/consumer and their commercial promotion is accurate.

According to WHO [9], the main functions of a DRA are as follows:

- registration (licensing) of products,
- inspection and licensing of manufacturers,
- inspection and licensing of distributors,
- post-marketing surveillance, and
- regulation of claims that can be made for commercial promotion of products as well as authorization of clinical trials.

A DRA can be effective only if it has the following:

- a legal basis for all its functions in legislation and regulations,
- sufficient human and financial resources,
- access to appropriate scientific expertise, and
- access to a quality control laboratory.

Particularly for vaccines and biologicals, in comparison with medicines, because the inherent variability of these products caused by the biological nature of their starting materials, their manufacturing processes, and their test methods, WHO has identified six essential regulatory functions to evaluate a regulatory system as effective [9]. These functions are as follows:

- a published set of clear requirements for licensing,
- surveillance of vaccine field performance (safety and efficacy),
- system of lot release,
- use of laboratory when needed,
- regular inspections of manufacturer for Good Manufacturing Practices compliance, and
- evaluation of clinical performance in clinical trials.

An extra function not included in the six established by WHO must be considered during the evaluation of regulatory authority for vaccines. This function must include the distribution and vaccine's shipment mechanisms; it must be able to guarantee the delivery of vaccines to all points of use with an acceptable level of quality.

Finally, research in the regulatory agencies, looking to have experts in each function updated in all relevant aspects involved in the regulation activities, is an aspect that must be considered during the implementation of regulatory systems in the countries.

The difference of the regulatory activities that WHO performs in comparison with the activities of other regulatory institutions is the impact that WHO guidelines, norms, and procedures have all over the world. The documents are developed looking for consensus with globally recognized experts. They are a reflection of the activities that are being developed in the world, and it facilitates their implementation.

## 5 OTHER INITIATIVES FOR HARMONIZED REGULATION

The Pan American Network for Drug Regulatory Harmonization (PANDRHA) is an initiative to support the process of drug regulatory harmonization throughout the Americas, with the mission to promote drug regulatory harmonization for all aspects of quality, safety, and efficacy of pharmaceutical products as a contribution to the quality of life and health care of the citizens of the member countries of the Americas. It includes the participation of all regulatory authorities in the region along with representatives of academia, industries, and other experts. The main forum for developing the PANDRHA activities is the Pan-American Conference on Drug Regulatory Harmonization, where there is also participation from other organizations such as the Caribbean Common Market (CARICOM), the Southern Countries Common Market (Mercado Comun del Sur, MERCOSUR), the North American Free Trade Agreement (NAFTA), and the Central American Integration System (Sistema de Integracion Centro Americano, SICA) and constitutes a way to disseminate the decision on drug regulatory harmonization of global initiatives such as the ICH. The network functions through working groups assigned to different areas including: good manufacturing practices, bioequivalence, good clinical practices, combat to drug counterfeiting, pharmacopoeia, drug classification, medicinal plants, pharmacovigilance, good laboratory practices, vaccines, and drug registration.

Other initiatives around the world exist that are incorporating, among many objectives, the harmonization of drug regulation. Although not all of these organizations have been developed exclusively for regulation, such initiatives can be found in the Association of South-East Asian Nations (ASEAN) [18], Andean Community of Nations (CAN), the Collaboration Agreement of Drug Regulatory Authorities in European Union Associated Countries (CADREAC), the European Union, the Gulf Central Committee for Drug Registration (GCC-DR), the Southern Common Market (MERCOSUR), and the Southern African Development Community (SADC).

## 6 FUTURE PERSPECTIVE OF REGULATION

Regulators must be prepared to face challenges resulting from the fast scientific, technological, and methodological development of drugs, biologicals, and biotechnological products. This effort will require that DRAs establish stronger alliances with the research community, universities, and academia in general so as to have available experts in all fields. Still, much remains to be done in the development of clinical trials with products derived from new technologies, mainly focusing on

those that target the immune system's natural control mechanisms, including therapeutic monoclonal antibodies, interleukins, and some vaccine's adjuvant, to avoid the recent disastrous example in Europe with the use of a monoclonal antibody [10].

Regulatory authorities have been receiving important pressure from industry to decrease approval times. Some agencies are trying to reduce approval times by making changes to the review system such as: improving communication and consultation by applicants [19], which leads to better preparation of dossiers; eliminating unnecessary administrative time loss; interacting with similar regulatory agencies; and increasing the number of employed reviewers with appropriate expertise. The latter has impacted through the accumulation of experience, know-how, and proficiency within the regulatory agency. However, it must be clear that from the point of view of society and improving public health, approval times alone are not the only goal. It is more critical to make the best decision to secure safety, efficacy, and effectiveness of the new drugs.

At the same time, for product developers, the recommendation to integrate regulation in the very early stages of their research and development work exist. This strategy can contribute as a success factor instead of a bottleneck at the later stages of licensing and market authorization. Incorporating a regulatory perspective will have a positive effect on the cost and time frame of product development. Furthermore, having a good understanding of regulatory requirements will also assist the developer in identifying nonviable projects at an early stage, thus reducing the risk of the loss of investment.

Regulation of drugs must balance the extreme efforts to regulate and control the presence of adverse effects while trying to address health care needs of the population in general. New products are costly to develop and produce. To guarantee their safety, specifically for possible rare events associated to its use, the need of designing and conducting large clinical trials to show statistical significance exist. This cost will be included in the product's price. As a recent example, in the vaccine area, large clinical trials had to be conducted to demonstrate safety for two new rotavirus vaccines. Over 60,000 children were involved in trials for each vaccine. These regulatory requirements add not only onto the cost of development but also onto the timing for the products to access the markets.

Post-marketing surveillance activities follow programs based in passive collection of spontaneous reports for ensuring safety using voluntary reporting of adverse reactions from physicians and other health workers is a common practice in developed countries, in many cases without or with poor participation from the regulatory agencies. This procedure shows deficiencies such as underreporting, difficulty in calculating rates because incomplete numerator data along with unreliable denominators, and limited ability to establish cause and effect [20, 21].

Recent withdrawals from the market of high-profile drugs are more often being carried out by the manufacturer who proceeds to withdraw the product before the regulatory agency does so, in principle to protect the public from further harm, but, at the same time, the manufacturer avoids the drastic consequences of penalties that the law establishes if the recall is done by the regulatory agency. Although much has been done to strengthen quality and safety during the development and production processes, post-marketing surveillance involvement by the regulatory authorities is still a pending issue.

The strongest regulatory agencies are still looking for reforms on drug regulation [16, 22] focused to improve the current regulatory process. Modifications proposed for the regulation in developed countries generally are focused on solving the particular problems in terms of the country or region that is involved, and, in general, are not applicable to regulation in developing countries. To be able to address the enormous challenges faced by regulation of new products, regulatory agencies in developing countries will have to urgently strengthen their networks and establish new harmonized regulatory and methodological procedures.

## REFERENCES


1. Chabra R, Kremzner M E, Killany B J (2005). FDA policy on unappproved drug products: Past, present and future. *Ann. Pharmacother.* 39:1260–1264.
2. Bren L (2006). The road to the biotech revolution: Highligths of 100 years of biologics regulation. *FDA Consumer Mag.* Centen. Edn. 1–7.
3. Offit PA (2005). The Cutter Incident. Yale University Press, New Haven, CT, pp. 83–131.
4. Baylor N W, Midthun K (2004). Regulation and testing of vaccines. In B A Plotkin, W A Orenstein (eds.), *Vaccines*, 4th ed.: Saunders, Philadelphia, PA, pp. 1539–1555.
5. Structure of ICH. Available: www.ich.org/cache/html/355-272-1.html.
6. WHO/Drug regulation. Available: www.who.int/medicines/areas/quality_safety/.
7. Ratanawijitrasin S, Wondemagegnehu E (2002). Effective drug regulation. A multicountry study. *WHO.* pp. 7–10.
8. Ng R (2004). *Drugs. From Discovery to Approval.* John Wiley and Sons, Inc., New York, pp. 51, 176–178, 212.
9. World Health Organization, Geneva (1999). Regulation of vaccines: Building on existing drug regulatory authorities. WHO/V&B/99. pp. 1–28.
10. Wadman M (2006). London's disastrous drug trial has serious side effects for research. *Nature.* 440:388–389.
11. Walsh G (2003). Biopharmaceutical benchmarks. *Nat. Biotechnol.* 21:865–870.
12. Hopkin M (2006). Can super-antibody drugs be tamed? *Nature.* 440:855–859.
13. Goldstein D A, Thomas J A (2004). Biopharmaceuticals derived from genetically modified plants. *QJ Med.* 97:705–716.
14. Ma J K, Barros E, Bock R, et al. (2005). Molecular farming for new drugs and vaccines. Current perspectives on the production of pharmaceuticals in transgenic plants. The European Union Framework 6 Pharma-Planta Consortium. *EMBO Reports.* 6(7):593–599.
15. Walsh G (2000). Biopharmaceuticals benchmarcks. *Nat. Biotechnol.* 18:831–833.
16. Wayne A, Ray W A, Stein M (2006). Reform on drug regulation—beyond and independiente drug safety board. *N. Engl. J. Med.* 354:194–200.
17. About the EMEA—Structure. Available: www.emea.eu.int.
18. Awang D C M Z B C (2003). ASEAN Initiatives toward pharmaceutical regulatory harmonization. *Drug Info. J.* 37:55–58.
19. Heidenreich K (2006). Dialogue between companies and the EMEA regarding the centralized procedure from an industry's point of view. *Drug Info. J.* 40:15–21.
20. Fontanarosa P B, Rennie D, De Angelis C D (2004). Post marketing surveillance—lack of vigilance, lack of trust. *JAMA.* 292(21):2647–2650.

21. Lexchin J (2006). Is there still a role for spontanous reporting of adverse drug reactions? *CMAJ. J.* 174:191–192.
22. Nutt D J (2006). Informed consent—a new approach to drug regulation? *J. Psychopharmacol.* 20:3–4.

# INTELLECTUAL PROPERTY AND BIOTECHNOLOGY

Tania Bubela[1] and Karen Lynne Durell[2]

[1] *University of Alberta, Edmonton, Alberta, Canada*
[2] *McGill University, Montreal, Quebec, Canada*

**Chapter Contents**

1 Introduction
2 Overview of Intellectual Property Rights for Biotechnology
  2.1 Patents
  2.2 Other Intellectual Property Rights
3 Biotechnology Patents, Public Policy, Public Opinion, and Morality
4 International Initiatives
  4.1 The World Trade Organization (WTO)
  4.2 World Intellectual Property Office (WIPO)
  4.3 The World Health Organization (WHO)
  4.4 Union for the Protection of New Varieties of Plants (UPOV)
  4.5 Convention on Biological Diversity (CBD)
  4.6 Food and Agricultural Organization (FAO)
5 National Patent Laws
  5.1 Substantive Patent Law
  5.2 Procedural Patent Law
  5.3 Legal Challenges: Infringement, Exemptions to Infringement, and Validity
6 The Effect of the Regulatory Environment
  6.1 Biosimilars
7 Reform Proposals for Innovation Governance in Biotechnology
  7.1 Strengthening Research Exemptions
  7.2 Patent Pooling
  7.3 Open Source Patents
  7.4 Compulsory Licensing
References

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

## 1 INTRODUCTION

Intellectual property laws and regulations govern the creation, use, and exploitation of intellectual activity in the industrial, scientific, literary, or artistic fields within national boundaries [1, 2]. They establish property protection over intangible items such as inventions, signs, and information. In theory, intellectual property protection provides incentives for the production and distribution of new knowledge that may benefit society by establishing a mechanism for the creator of that knowledge to obtain compensation for the use of protected works.

Intellectual property laws and regulations are designed to manage the competing interests of those who produce new knowledge, those who want to make use of new knowledge, and those who want simply to buy the products and services made through the use of new knowledge. It is becoming increasingly difficult to balance these competing desires in light of rapid advances in fields such as biotechnology and information technology. More broadly, designing sets of laws and regulations, along with the governmental institutions and practices that administer them in keeping with business practices and public interests, is becoming increasingly difficult, which is especially so given the lack of empirical data on how the entire intellectual property system works for innovation in emerging technological fields [3, 4]. The central issue is the appropriate balance to be struck between setting intellectual property rights that encourage and reward innovators while still permitting affordable and ready access to the public and other users. The latter concern is paramount in the context of health biotechnology.

This chapter provides an overview of intellectual property rights available for biotechnological innovation and the national and international laws and regulations governing those rights. The focus is on patent rights, the most widely sought intellectual property right in the field of biotechnology. We compare the national patent laws of the United States, Europe, Japan, and Canada.

Unfortunately, most patent laws and regulations were designed for inanimate matter. Significant difficulties develop in the application of arcane laws to technological innovation involving biological material, especially self-replicating life forms. Patent rights have become a flash point for many of the social and ethical concerns that derive in the context of biotechnology. Although the controversy centers on patenting of biotechnological innovation, the underlying concern is the increasing commercialization of biotechnological research. We conclude with widely raised policy reforms designed to address issues of balance, access, and public support for research and innovation in health biotechnology.

## 2 OVERVIEW OF INTELLECTUAL PROPERTY RIGHTS FOR BIOTECHNOLOGY

### 2.1 Patents

Patents, the focus of this chapter, are the most prevalent form of intellectual property right (IPR) sought for biotechnological innovation [5]. Patent rights are nega-

tive rights because they confer upon the holder the right to prevent others from selling, using, importing, and making the invention for a set period of time [6–8]. In effect, patent rights grant a holder exclusive control over an invention—a monopoly right. This broad exclusionary right is an important and valuable tool for industry as it can translate into the ability to limit entry by others into a market sector and to set the market price for an invention.

Patent rights may be granted over several aspects of an invention including the product itself and the process to manufacture the product. Several formal steps exist that must be completed to obtain a patent. The two most important are: (1) submission of a patent application, drafted to meet legislated format and content requirements; and (2) a successful review of the patent application by a national patent office based on the patent criteria. The most crucial aspect of the patent application is the claims; only those elements of the invention included in the claims will be granted patent protection. Nonclaimed elements will not be granted patent protection [9, 10]. The patent criteria, with some variation between national patent systems, are that the invention be new, useful, and involve an inventive step. An invention that fails to meet any of the criteria, or any other requirements set out in national patent laws, will not be granted a patent.

As patent rights are national in scope, inventors who want to secure patent rights in more than one nation must submit a patent application and comply with the patent legislation in each country individually. Patent rights granted by one country will not be respected in any other country. For example, if an inventor is granted a patent in the United States but fails to seek patent protection in Canada, the inventor's patent rights will end at the U.S./Canadian border and Canadians will be free to make use of the invention. Moreover, if an inventor is granted a first patent in the United States and a second patent in Canada for the same invention, the U.S. legislation will govern the use of the patent rights granted in the United States and the Canadian legislation will govern the use of the patent rights granted in Canada. Generally, no overlap of patent legislation exists between nations.

The modern version of patent rights has evolved and changed significantly from its historical role. For example, in Europe, patents were originally granted as letters patent, which were intended to reward inventors for their industrial activities. The original letters patent were different from today's patents in that they granted privileges rather than property rights. The sovereign had the right to grant a privilege to an inventor, and in so doing, encourage the development of new industries within a nation [11]. Letters patent were not necessarily granted for inventions that were absolutely new but for those that were newly introduced to the nation [11]. Copying a foreign invention and introducing it to a new country was an entirely permissible and desired outcome. For example, a patent was granted in France by the French crown "for glassware according to the manner of Venice" [11], although it was common knowledge that the glassware technique was not novel but was previously practiced in Venice. The fact that it was new to France was sufficient to garner a patent right.

The modern patent system is largely a creation of the nineteenth century and is popularly understood to respond to the argument that inventors have a natural, inherent property right in their inventions that the state has an obligation to protect [11]. This argument is directly related to the realization that innovation can repre-

sent significant commercial value. Accordingly, the state agrees to protect the inventor's natural property right in an innovation, but considers public benefit by requiring the inventor to disclose publicly the details of the invention. Disclosure allows for the generation and circulation of technical information, which is known as the "information function" of the patent system [2].

The second main justification for patents is that they provide an incentive for the production of new innovation and improvements on old innovations. The key incentive is the 20-year patent term. The establishment of 20 years as a firm and unbending time period during which inventions may be exploited solely by the inventor has many benefits. In particular, it offers certainty in the length of the right, a major factor in attracting investment for research and development.

Modern patent legislation, which emerged in the United States and the United Kingdom in the 1970s, formalized the information function of patent law. It requires an inventor to describe the invention in such a way that it can be put into practice by a person skilled in the art. Modern-day patent law also rationalizes the procedural component of patent examination by establishing set criteria that must be met for an invention to merit a grant of patent rights [11]. The outcome of these changes is that skilled examiners are now required to evaluate whether a patent application is complete and ultimately whether an invention is patentable.

## 2.2 Other Intellectual Property Rights

A number of other forms of IPRs, besides patent rights, may also apply to protect elements of biotechnological innovations. Here, we present a brief overview of these IPRs, which include trade secret, plant breeders' rights, copyright, database rights, and trade mark.

***Trade Secret.*** Trade secret grants a right over information that is treated as confidential by a corporation or other organization. The right translates into the ability to sue for a breach of confidentiality if someone, for example, a key employee, discloses confidential information to a competitor. The wronged party is entitled to a remedy usually in the form of monetary compensation for the harm caused by the wrongful disclosure of the secret information.

Trade secret provides a necessary supplement to other IPRs because it can be used to protect technical information and know how during research, development, and testing stages of a biotechnology project. Unlike other IPRs, trade secret is rarely rooted in statute, but is usually established by the collective history of judgments from litigated cases known as jurisprudence [12, 13].

Trade secret protection is acknowledged when information is treated as confidential, which usually means those to whom it is properly disclosed realize that it is secret and agree to protect the confidentiality of the information. In many cases, companies establish secrecy through the use of confidentiality agreements with employees and other customers or partners. As the right relies on absolute secrecy, trade secret protection is fragile; the right can be lost immediately upon nonconfidential disclosure of the secret information. Companies therefore usually opt to seek patent protection once a product is introduced to the market and only rely on trade secret during the research and development stages. It is particularly true for

innovations that can be easily reversed-engineered. Such inventions cannot be adequately protected by way of trade secret because they are subject to copying.

However, in some cases, trade secret protection may constitute a competitive advantage by sustaining marketplace confidentiality [14]. If the technical details of an innovation remain secret, it is often difficult for competitors to recreate a product, invent around it, or improve on it, especially if the innovation is complex. These latter types of activities are commonly known as competitor piggy-backing. If adequate confidentiality measures are put in place and enforced, trade secret can be an effective method of protecting information from the public and securing a market sector monopoly [15].

***Plant Breeders' Rights.*** Generally, plant breeders' rights are granted to a person who breeds a new plant variety and for this reason are helpful tools for many biotechnology projects. Like patents, plant breeders' rights require a formal application and review process, which is undertaken nationally. The strength of plant breeders' rights varies greatly between nations. The rights granted may extend to the plant, as in the United States where plant rights are a subset of the patent system, or may only include the propagating material and not the plant, as set out by Canadian legislation. It is argued that strong plant breeders' rights negatively affect farmers because plant breeders may exercise their right to limit the subsequent uses of their plants. For example, a farmer who grows a crop from protected seeds may be prevented from saving the seeds from that crop for the purpose of cultivating a subsequent crop [16, 17]. Concerns over farmers' rights vary globally resulting in an international collection of plant breeders' rights statutes that are inconsistent in strength and scope.

***Copyright.*** Copyright regulates the creation and use of a range of original works such as books, songs, films, and computer programs. Copyright protects the expression of an idea, not the idea itself. Thus, it protects the work that is produced but, unlike a patent, cannot prevent the making, use, or sale of an identical work by an independent originator [18]. Copyright differs from patent and trade mark in another important way—it arises automatically and confers upon the owner the right to prohibit copying or performance of a work. For example, a researcher has a copyright in the scientific papers they seek to publish, although he or she generally transfers or assigns those rights to the journal before publication. Copyright is granted by national statutes for the lifetime of the author plus anywhere from 50 to 70 years after the death of the author, depending on the country. However, because of international treaties, the national foundation of copyright law is less oppressive than for patent rights. Copyright can cross borders—a work created in the United States will be acknowledged as a copyright protected work in Canada as well [19]. Copyright is of limited use for the products and processes of biotechnology, even in the area of collections of biological data, gene sequences, and plants.

***Database & Material Deposit Rights.*** One emerging area of biotechnology that is testing the limits of intellectual property regimes is the production of academic data collections, especially biological and genetic data linked to long-term health of the individuals who provided the data. Such databases require substantial

intellectual investment. Traditional IPRs, however, such as patents and copyright, do not protect mere collections of data. Indeed, the prevailing thought in legal discourse over IPRs and data is that data *per se* should not be protected based on the public nature of many large data collection efforts [20]. However, there may be need for some form of property right to prevent misappropriation and misuse of data. One such model exists, namely the European Database Directive, which was introduced in 1996 by the European Union. The directive vests an exclusive right in the producer of a database to grant permission to extract and reuse the contents of the database [11]. The debate about the appropriateness of database rights in the context of biological data is ongoing [20].

It should be noted that some national patent laws require the deposit of biological and genetic materials to supplement the disclosure of an invention. Materials are submitted to authorized depositaries that eventually find themselves in possession of a collection of biological samples. Biological material depositaries are governed at the international level by the Budapest Treaty on the International Recognition of the Deposit of Microorganisms for the Purpose of Patent [21], as established in 1977 (discussed in more detail below). The treaty and national patent rules cooperatively establish authorized depositories and sets rules for the deposit as well as the availability to the public of the material.

***Trade Mark.*** National trade mark legislation grants protection for certain words or symbols, usually trade names such as Microsoft or pictures such as the Nike swoosh, when the words or symbols are used in connection with specific marketplace goods or services [2]. Trade mark holders are granted the right to prevent others from using the protected words or symbols to represent the same or similar goods or services. Generally, trade marks must be registered at a national office to be effective, although some jurisdictions acknowledge that a right can be established through consistent long-term use.

## 3 BIOTECHNOLOGY PATENTS, PUBLIC POLICY, PUBLIC OPINION, AND MORALITY

The patent system is traditionally viewed as a tool for promoting economic goals [22]. It is understood to work by encouraging the advancement of new industries, research and development, or innovation. In contrast, noneconomic factors such as health, human rights, environment, or ethics have historically been viewed as extrinsic to the activities regulated by the patent system or as undesirable side effects that require mitigation. Accordingly, patent law has been largely unconcerned with moral arguments. A few notable exceptions exist including medical treatments and inventions designed for purely criminal applications.

With the advent of the patenting of the products and processes of biotechnology, moral arguments have entered the patent arena with more force than ever, especially in Europe [23]. Indeed, the commercialization and patenting of biotechnological innovation has become a focal point for opposition to biotechnology *per se*. The core objection is that some applications of biotechnology, such as the genetic manipulation of higher organisms or embryonic stem cell research, are in them-

selves wrong because they are against God, nature, or human dignity. The patenting of human genes also draws special criticism as establishing property rights over the common heritage of humankind.

Intuitively, biotechnological inventions do not fit neatly into the customary model of patentable subject matter because the original law was developed for mechanical inventions. Two reasons why the patent system is not necessarily a neat fit with biological matter are the ability of living organisms to self-replicate, which has significant implications for the exclusive rights patents grant over the use and manufacture of a product, and because many biological products, such as genes, may be seen more as discoveries than as new inventions. But, despite the ill fit, patents have been issued over life forms since the mid-1800s and biological inventions have been patented since the 1970s [5]. That said, some forms of biotechnology remain unpatentable; for example, genes and gene fragments are not patentable in their natural state, but only in isolated form.

In the field of biotechnology, policy makers must achieve a delicate balance between the advancement of research and biomedical innovation and the many social, legal, and ethical challenges raised by this field of science. The latter, coupled with the fast pace of advances in biomedical research, makes it essential that there be an ongoing public review accompanied by appropriate legislative and regulatory responses [23]. Unfortunately, legislative and regulatory responses generally lag behind research and corresponding moral and ethical dilemmas. It is therefore necessary to keep law making and regulatory processes flexible and responsive to scientific advances, including those in the realm of IPRs.

There is considerable debate about the appropriateness of using the patent system to regulate innovation on the basis of public morality [3]. Those opposed state that the patent system is ethically neutral, concerning itself only with the technical requirements for patentability spelled out in national laws and influenced by international standards. Historically, patents are concerned primarily with economics and innovation while other regulatory mechanisms are both more flexible and better at optimizing health, safety, and environmental outcomes. However, others argue that the balance has shifted too far toward granting property rights over biotechnological innovation at the expense of the beneficial social outcomes of information diffusion and access to new innovation [3]. Globally, opposition to the patenting of living organisms, genes, and gene fragments may be viewed as broad opposition to the commercialization or commodification of biological matter [24]. Public unease may be augmented by distorted media reporting of sensitive issues, or these reports may merely reflect public perceptions. Either way, such reports may lead to a general public mistrust of government regulators and politicians [25].

Although such concerns may seem unfounded, it is not wise simply to discard public sentiment that opposes biotechnology. It is overly simplistic to state that public opinion, particularly public concern, is based on a combination of ignorance and misinformation. Proponents of genetic technologies and products use this argument to delegitimize public opposition. Media sensationalism is "blamed for amplifying and exploiting that public ignorance" [26]. Indeed, public opinion does not necessarily show a correlation between a level of education and support for biotechnology but depends, instead, on the segment of the public and on the specific genetic technology or product [25].

On the whole, polling studies have shown that the public is largely supportive of biomedical research [27], especially when that research is perceived to be independent of industry and conducted at publicly funded research institutions [24, 28]. Public support for agricultural biotechnology, on the other hand, is lower, especially in Europe, partly because the public has not been convinced of the utility of the applications of agricultural biotechnology, which translates into a higher level of risk tolerance by the public for health biotechnology. That said, the scientific community must be willing to accept that some research avenues are simply not acceptable to the majority of the public, and should potentially be reconsidered. Rather than blaming the media and other information conduits that are generally presenting a pro-science position, especially in the biomedical research, scientists need to demonstrate the utility of their research along with a fair assessment of the risk [29, 30].

In a democratic society, it is important to maintain public trust and confidence because the lay public can exert substantial influence on their public representatives who in turn fund or regulate scientific research and the use of science-based technologies [27]. One substantial set of studies estimates that public opinion, as measured by polling, becomes functioning law about two-thirds of the time [31]. Public representatives may also ban or criminalize scientific research that is perceived as immoral, dangerous, or unjust. Indeed, such steps have been taken in Canada with the recent enactment of the *Assisted Human Reproduction Act*, which has criminalized a range of research endeavors such as human cloning, somatic cell nuclear transfer, and germ-line alteration [32].

One factor that may significantly influence public opinion and public support for health biotechnology is the increasing trend toward commercialization of publicly funded research. Many major public funding agencies and research institutions, including publicly funded universities, now set commercialization of research activities as an explicit goal. Indeed, the rise of modern biotechnology is often explained by two events: First, the United States Supreme Court allowed the first patent of a genetically modified bacterium in 1980 [33], giving industry an incentive to invest and engage in biotechnological research; and second, the United States Congress passed the Bayh–Dole Act in 1980, providing universities with the right to obtain patents in the results of federally funded research and encouraging universities to transfer technology to industry. The success of technology transfer between universities and industry is measured by the annual survey of the Association of University Technology Managers [34].

However, evidence exists that the public is becoming increasingly suspicious of the ties between industry and researchers, at a time when academic biomedical research receives unprecedented levels of funding from industry and research institutions are increasingly focused on policies aimed at commercialization [35–38]. Survey data show that a perceived connection with commercial forces has an adverse impact on the perceived credibility of researchers. If there is a perception that funding comes from less independent sources, credibility deteriorates. For example, one recent survey of the Canadian and U.S. public on biotechnology issues found that publicly funded university researchers are highly trusted and credible. Only the World Health Organization (WHO) and peer-reviewed scientific journals were rated higher. In contrast, scientists working for biotechnology com-

panies and university researchers funded by industry were not considered credible [39]. A 2000 focus group study done on behalf of the Canadian government came to a similar conclusion, finding that "many people say university scientists are much more credible than other scientists because it is assumed they are free from funding pressures and therefore more 'independent'" [40].

Some of the most highly publicized controversies in science policy have been a direct result of the growing ties between biomedical science and commerce. For example, in the United States, the death of Jesse Gelsinger raised concerns about whether financial considerations had had an inappropriate impact on the running of the gene therapy research trial [41]. Other concerns about commercialization are directly related to the patenting process, the Myriad Genetics controversy being the most obvious. As described by Professor Caulfield, "the decision by Myriad, a Utah based company, to enforce its patents over the BRCA1/2 genes was, arguably, the most significant recent catalyst of national policy-making activity, media coverage, and public outrage on the topic of gene patents. In 2001, the company sent a 'cease-and-desist' letter to most Canadian provinces that used the test as part of the publicly funded health care system. The case was viewed, rightly or not, as a harbinger of the policy challenges created by gene patents" [42].

Other high-profile biotechnology patent stories exist, such as Harvard College's attempts to patent a mouse, genetically modified to be susceptible to cancer, known as the "onco-mouse." A patent was granted both for the method of making the mouse and the mouse itself by the United States Patent and Trademark Office (USPTO) in 1989, but the application for the mouse patent was initially rejected by the European Patent Office (EPO). Unlike the United States, Europe allows for the rejection of patents considered to be contrary to public morality or the *ordre public*. The patent was eventually granted, but the EPO developed a statement on the patenting of transgenic animals. The analysis developed by the EPO requires an evaluation of the risks, including moral risks, and benefits of a new technology, with the recognition that every new technology is accompanied by some risk. In the case of the onco-mouse, the benefits to humankind of the mouse model for cancer research outweighed the public morality concerns about protecting the environment and the suffering of the experimental animals. The risks were further allayed by the lack of research alternatives to mouse models and the additional efficiency of using the onco-mouse in reducing the number of mice required for experimentation. The primary importance of this analysis is that biotechnology is to be treated the same as any other new technology in evaluating risks and benefits, and genetically modified organisms are not excluded out of hand from patentability on the basis of immorality.

The Canadian courts have taken a different stance on the patentability of higher life forms in a number of high-profile cases. In *Harvard College v. Canada (Commissioner of Patents)*, the Supreme Court of Canada (SCC), in a slim majority, concluded that the onco-mouse was not patentable subject matter under Canadian patent law. The SCC held that the drafters of the Patent Act had not intended higher life forms to fit within the definition of patentable subject matter [43]. To confuse the issue further, the patentability of higher life forms was addressed again only two years later in the 2004 case of *Monsanto Canada Inc. v. Schmeiser* [44]. In that case, a farmer had grown a crop of Monsanto's genetically modified canola

without authorization from Monsanto. In its patent, Monsanto had claimed all products and processes related to the genetically modified canola except for the canola plant itself, because higher life forms are not patentable in Canada. The SCC did not explicitly overturn its decision that higher life forms are not patentable, but ruled that if a nonclaimed element, in this case the canola plant, is determined to have an "important role in production," then patent protection will extend to that element as well as the claimed elements [44]. In other words, having a patent over a gene inserted into a higher life form grants patent protection over the higher life form as a whole, even though technically the higher life form cannot be claimed as an invention in Canada.

The effect of the decision was to grant patent protection to a nonpatentable plant (a higher life form) because its bioengineered cells are patentable [44]. In application, this doctrine means that patent scope can be understood to cover both the claimed invention as well as any broader structure that encapsulates or incorporates the invention [44]. These Canadian decisions have been referred to as "the most significant decisions in the area of biotech patent law anywhere in the world," and the esteem attributed to them makes their inconsistencies that much more disconcerting [45].

These examples in Canada and Europe have generated a high degree of public and policy-making interest. Indeed, in these regions, a high level of public discomfort exists with the patenting of higher life forms and human genes, which is augmented by the morally contentious nature of much biotechnology research, for example, the ongoing debates about embryonic stem cell research. In controversial areas, public trust is fragile. Professor Caulfield summarizes the dilemma as follows: "The public may be setting aside moral concerns based on the belief that the research is being done in the public's interest. If the public comes to believe that commercial interests are dominating the research process, trust could easily be lost. And, it should not be forgotten that many in the public are already suspicious of the motivations behind biotechnology innovation" [42].

In addition, numerous other policy challenges exist that stem back to the social benefit side of the patent bargain. Significant concerns exist that a commercialization agenda for biomedical research will skew the research agenda away from basic research and that new biotechnological innovations will drive up the cost of health care in publicly funded systems, such as Europe and Canada, and increase the disparity between patients or regions that can afford new technologies and those who cannot. These concerns may partially be addressed by a well-formulated and well-balanced patent system and government regulation.

## 4 INTERNATIONAL INITIATIVES

As a result of the national scope of many intellectual property laws, problems sometimes develop where inventions, such as biopharmaceuticals, are traded internationally. International trade can bring divergent national legislation into conflict. Countries that are net exporters of innovations protected by IPRs, such as the United States, the European Union, and Japan, advocate strongly in the international arena for the strengthening of national intellectual property laws and international harmonization of these laws [46]. Although international harmonization

has not been achieved, intercountry trade has lead to a number of international initiatives that significantly impact the intersection between IPRs and biotechnological innovation.

Here, we review the main international treaties relating to either IPRs or biotechnology and the institutions that administer them. The treaties range from attempts to coordinate the filing of patents at an international level to guidelines for the sharing of benefits derived from biologic resources. Depending on the form of a biotechnological innovation, several of these treaties may be applicable to a single invention. Thus, simultaneous compliance with several treaties may be required. As the treaties have not been drafted to reflect and consider each other, but do in some instances overlap, compliance can be difficult in some areas of biotechnological research and commercialization.

### 4.1 The World Trade Organization (WTO)

The function of the WTO, to establish trade standards at an international level, is particularly important in today's era of globalization. The WTO engages in "analysis and debate about the relationship between international trade and investment, and its implications for economic growth and development" [47]. It promotes trade and eliminates inefficiencies [47] by imposing obligations on members to accept common rules and arbitration methods. The WTO addresses trade issues in a number of different categories, one of which is the protection of IPRs and trade.

The WTO membership includes countries at varying levels of development, and from divergent political and legal systems, which are differentially impacted by globalization. This breadth of input lends credibility to both the organization and the standards created by WTO membership [48]. Despite criticism, the membership of the WTO causes its principles to be adopted widely around the world.

***Trade-Related Aspects of Intellectual Property Rights (TRIPs).*** TRIPs, the most influential multilateral agreement on IPRs, was created by the WTO at its 1986–1994 Uruguay Round of negotiations [49]. TRIPs regulates trade exchanges through uniform transnational rules and encompasses issues related to biotechnology within its framework. The TRIPs negotiations were instigated by the realization in the 1970s of the negative effect of counterfeiting of trade-marked products on international trade [50]. Although there was opposition to the WTO assuming a role in global intellectual property issues, a move that was thought to usurp the authority of the World Intellectual Property Organization (WIPO), a General Agreement on Tariffs and Trade decision of January 28, 1987 established guidelines for the negotiation process that eventually produced TRIPs [50].

The official goal of TRIPs is to "reduce distortions and impediments to international trade" posed by divergent national IPRs [49]. It presents a common minimum standard of IPRs for the adoption within national intellectual property laws of each WTO member country. The common standards are enforced by the WTO, which adjudicates disagreements between nations [50]. Each WTO member country retains the right to legislate more stringent IPR regimes [50].

TRIPs sets standards for many forms of IPRs—copyright, plant breeders' rights, and trade marks—but the provisions relevant to patent rights are particularly significant for biotechnological innovations. To be eligible for patent protection,

TRIPs requires an invention to be new, to have an industrial application, and to have an inventive step [49]. Inventions that meet these criteria are protected against use, offer for sale, and sale and importation by anyone who is not the rights-holder for a minimum term of 20 years [49]. Moreover, TRIPs requires nations to grant patent rights for all fields of technology without discrimination and for both inventive products and processes [49].

Despite the harmonizing nature of TRIPs, a number of controversies have developed over the implementation of provisions related to biotechnological innovation and higher life forms. Article 27.3(b) states that "plants and animals other than microorganisms, and essentially biological processes for the production of plants or animals other than non-biological and microbiological processes" may be excluded from patent protection. However, plant varieties must be protected within national legislation through patents, *sui generis* protection, or a combination of both. Patents are well understood and *sui generis* protection is commonly provided as plant variety protection or plant breeders' rights legislation, compliant with UPOV, discussed below. The correct blend of these methods of protection has been subject to varying interpretations, particularly in the field of biotechnological invention. The issue of integrating the protection of traditional knowledge about genetic resources and benefit sharing with traditional knowledge holders and plant variety protection is particularly controversial.

Another controversial aspect of TRIPs is the compulsory licensing provisions in Article 31, which allow a government to override the rights to exclusive use and manufacture of a product or process granted by a patent. National IPR regimes may allow for compulsory licensing under limited circumstances. Conditions for granting a compulsory license include a national emergency and public noncommercial use for public health, national defense, or environmental protection. The provisions of a compulsory license may likewise be limited by national IPR laws. For example, negotiations with the holder of patent rights must start before the granting of a license, the scope and duration of the license must be limited, the rights holder must be compensated, and the license must be nonexclusive. However, these strictures have not quelled the debate, which is fueled by the characterization of the compulsory license issue as a struggle between developed nations, who support very limited compulsory licensing, and developing nations, who are interested in the extension of broad compulsory licensing rights to combat public health crises such as the HIV/AIDS pandemic [50].

TRIPs does address some concerns specific to developing countries. Article 66 exempts developing countries from compliance with most provisions of TRIPs for a term of 10 years because of "their economic, financial, and administrative constraints, and their need for flexibility to create a viable technological base" [49]. A further extension may be granted on request. Moreover, TRIPs admonishes developed nations to engage in technology transfer initiatives with less-developed countries to facilitate the establishment of a sound and viable technological base in those countries [49].

TRIPs also establishes minimum standards and definitions for undisclosed information or trade secrets.

***Sanitary and Phytosanitary Agreement (SPS).*** The SPS is a multilateral mechanism formulated to protect human, animal, and plant health in WTO member

countries. The SPS details how governments can apply food safety and animal and plant health measures. These measures are based on standards articulated by a number of international organizations including the food safety standards formulated by Codex Alimentarius Commission, the animal health and disease control measures drafted by the International Office of Epizootics, and the steps for plant quarantine and phytosanitary certificate regulation articulated by the International Plant Protection Convention. SPS measures are based on a risk assessment—risks should be evaluated and the measures adopted should be commensurate with the risk. The SPS recognizes the sovereignty of nations and acknowledges that a country may maintain measures that are stricter than the SPS measures. Each nation has a right to determine its appropriate level of risk. Set standards will only be questioned if they result in discrimination or a disguised restriction on trade [51].

### 4.2 World Intellectual Property Office (WIPO)

WIPO is a specialized agency of the United Nations (UN) system of organizations, with a mandate to administer intellectual property matters recognized by the member states of the UN [52]. It administers 23 treaties and various programs that aim to harmonize national intellectual property legislation and procedures, provide services for international applications for industrial property rights, exchange intellectual property information, provide legal and technical assistance to developing and other countries, facilitate the resolution of private intellectual property disputes, and marshal information technology as a tool for storing, accessing, and using valuable intellectual property information.

***Patent Cooperation Treaty (PCT).*** The Patent Cooperation Treaty (PCT) makes it possible to reserve the right to seek patent protection for an invention simultaneously in each of a large number of countries by filing an "international" patent application [53]. Such an application may be filed by anyone who is a national or resident of a contracting state. It may be filed in the applicant's national or regional patent office. Copies of a filed application are then forwarded to the International Bureau and an International Search Authority for processing and review.

The aim of the PCT system is to consolidate and streamline patenting procedures. A PCT application reserves the right to file patent applications in all of the member states at a future date. It gives the applicant more time to submit patent applications to national offices than is allotted under national patent systems. This additional period of time (up to 42 months depending on the residence of the applicant and whether a priority filing has been made) gives an applicant more time to decide where to continue with national patent applications. The PCT process also provides the applicant with valuable information about the potential patentability of the invention through an international search report and the optional international preliminary examination report.

The PCT system is expanding rapidly: the number of member states, including many European countries, Japan, and Canada, has more than doubled in the last eight years, to 125, and the number of international applications has grown from 2600 in 1979 to about 110,065 in 2003.

***Budapest Treaty on the International Recognition of the Deposit of Microorganisms for the Purposes of Patent Procedure (Budapest Treaty) [21].*** Patents require the disclosure of the invention, usually in writing, with sufficient detail that a person skilled in the art can reproduce the invention, which raises difficulties when the invention involves a microorganism or the use of a microorganism. The Budapest Treaty allows for disclosure to be made by depositing a sample of the microorganism with a specialized institution [54]. In practice, the term "microorganism" is interpreted in a broad sense, covering a range of biological material the deposit of which is necessary for the purposes of disclosure, in particular regarding inventions relating to the food and pharmaceutical fields.

An "international depositary authority" (IDA) is a scientific institution—typically a "culture collection"—that is capable of storing microorganisms. Such institutions have been formally nominated by the contracting state within which they reside and have agreed to abide by the treaty provisions. Under the treaty, the deposit with any IDA satisfies disclosure requirements of national patent offices of contracting states if those states allow or require deposits of microorganisms in their national patent legislation. The regional offices of The African Regional Intellectual Property Organization (ARIPO), the Eurasian Patent Organization (EAPO), and the European Patent Office (EPO) have recognized this method of disclosure, as have the United States, the United Kingdom, Japan, and Canada.

### 4.3 The World Health Organization (WHO)

The WHO is the UN specialized agency for health. Although it does not administer any intellectual property treaties, it is involved in intellectual property issues in the field of health. Indeed, the independent Commission on Intellectual Property Rights, Innovation and Public Health, established by the World Health Assembly in 2003 has recently completed a report entitled, *Public Health, Innovation and Intellectual Property Rights* [55]. This report considers "intellectual property rights, innovation, and public health, including the question of appropriate funding and incentive mechanisms for the creation of new medicines and other products against diseases that disproportionately affect developing countries . . ." [56].

### 4.4 Union for the Protection of New Varieties of Plants (UPOV)

UPOV is an intergovernmental organization with headquarters in Geneva, Switzerland [57]. It administers the UPOV System of Protection of Plant Varieties, established in Paris in 1961, with the adoption of the International Convention for the Protection of New Varieties of Plants [58]. The goal of the UPOV Act is to encourage the development of new plant varieties and to benefit society through the grant of a *sui generis* IPR to plant breeders.

Members of UPOV are required to grant and protect plant breeders' rights, including exclusive rights for the production, reproduction, and conditioning, for the purpose of propagation, offering for sale, selling or marketing, importing, exporting, and stocking, of plant varieties. The UPOV Act also sets criteria for plant variety protection, namely novelty, distinctness, uniformity, and stability [59].

The impetus behind plant breeders' rights varies globally, and, thus, the rights granted vary from nation to nation, despite the harmonizing role of UPOV. For example, countries that allow for plant patents offer broader, stronger enforcement options than countries with *sui generis* plant breeders' rights. The extent of rights granted to plant breeders reflects issues of national concern, such as the rights of farmers to save and reuse seeds, among others. The stronger rights for plants provided by patents are favored by multi-national agribusiness and countries such as the United States.

Recent changes to the UPOV Act have resulted in significant controversy about the appropriate scope of plant variety protection. Specifically, the act was extended to provide protection to an "essentially derived variety" of plant, which has left confusion about exactly which plants are now covered by UPOV. Moreover, debate rages about provisions that allow a breeder to impose royalty payments on farmers who save seeds for the purpose of reusing them in the future [17]. Royalty payments may be a significant hurdle for some farmers, particularly those in developing countries, and the provisions may augment food insecurity issues [60].

Another controversial modification to the act allows for *ordre public* considerations to supersede infringement if equitable remuneration is offered, which opens the door to a system in which *ordre public* may be invoked to avoid seeking the consent of a rights-holder to use a protected plant variety. Together, these changes to the UPOV Act are worrisome to those who perceive it to be unjust and those who consider the act to offer insufficient solutions for the problems facing farmers in developing countries [61].

### 4.5 Convention on Biological Diversity (CBD)

The CBD, established in 1992, provides a coordinated international framework for the commercialization and sustainable use of genetic resources [62]. It forms the basis for a fair and equitable means of sharing benefits derived from the commercial use of genetic resources. The CBD tries to balance the sovereign rights of states to draft legislation regarding the natural environment within their borders with obligations states owe to other nations. For example, Article 3 of the CBD acknowledges a state's right to exploit its own resources, but simultaneously asserts the responsibility of a state to refrain from harming areas lying outside its borders. Article 15 addresses genetic resources specifically. It permits states to regulate access to genetic resources. However, at the same time, that access must be subject to prior informed consent obtained from the party providing the resources. This provision is intended to ensure benefits flow to indigenous or other marginalized communities, which are often the custodians of traditional knowledge and rich biodiversity within their territories. The CBD is binding for its signatories, which presently include 188 parties. However, it does not include the United States.

The CBD is not formally recognized by the WTO and, consequently, is considered to lack significant legal weight. Developing countries in particular continue to advocate for the adoption of the CBD by the WTO. They argue that the CBD promotes fair and equitable use of genetic resources and, therefore, its integration with the WTO is necessary to formulate balanced international standards for genetic resources. The CBD standards may also complement and enhance the

TRIPs provisions, which promote technology transfer between developed and developing countries.

Parties attending the conference that adopted the CBD have also accepted a supplementary agreement known as the Cartagena Protocol on Biosafety as of January 29, 2000 [63]. The Protocol protects biological diversity from potential risks posed by genetically modified organisms (GMOs). One initiative is for an advance informed agreement procedure that will ensure all relevant information is provided to a country before the importation of GMOs. The protocol also establishes a biosafety clearinghouse, intended to facilitate an exchange of information on GMOs and aid the implementation of the protocol.

### 4.6 Food and Agricultural Organization (FAO)

The FAO administers the International Treaty on Plant Genetic Resources for Food and Agriculture (FAO Treaty), in force as of June 29, 2004, which encompasses all plant genetic resources that fall within the categories of food or agriculture [64]. It sets out access and benefit sharing rules for these resources. The provisions of the FAO Treaty accord with the CBD, although its subject matter is limited in comparison. The FAO Treaty sets up a multilateral system to facilitate access to plant genetic resources for food and agriculture, allowing countries that have ratified the treaty to work together to create equitable material transfer agreements, in compliance with the treaty. It also coordinates a pool of crop resources, collected from the members, which facilitates sharing of information among farmers and breeders. In general, the FAO Treaty aims to protect genetic resources relating to specific plants and is one of the first treaties to propose an operational system of plant genetic resource distribution at an international level.

## 5 NATIONAL PATENT LAWS

In this section, we compare patent rules and patent criteria for biotechnological innovation in the United States, Europe, Japan, and Canada. Although attempts have been made to harmonize patent laws globally, so far it has been impossible to achieve international consensus. TRIPs sets minimum standards that must be included in national patent legislation enforced by all WTO members, but it has not had the effect of creating consistent national patent laws.

Some attempts at regional harmonization of patent rules have been implemented. The European Patent Convention (EPC) guides the granting of patents in European countries by offering an alternative means of filing a patent in several European jurisdictions simultaneously [65]. A party seeking a patent in Europe has two options: file an application in national offices or file with the European Patent Office (EPO). An application filed with the EPO, upon grant of patent, is equivalent to having a patent in each member state of the EPC [not necessarily the same as the member states of the European Union (EU)]. All patent applications filed under the EPO system are required to submit to EPC rules. However, the jurisdiction and activities of the EPC are limited to the granting of patents. The EPC does not deal with issues developing after the grant of a patent such as infringement. Thus, the EPC governs the adjudication of the granting of patens, but national laws govern post-grant challenges.

Regional filing initiatives aside, patent laws are generally national in scope and apply only to the use of patented inventions within national borders. A patent application must, therefore, be filed in each country where patent protection is sought. The problem is that the patent law of each country is slightly different. TRIPs has created consistency in some aspects of patenting such as the patent term, which must be at least 20 years from the filing date. However, national variation may affect the subject matter that can be patented, the criteria used to judge whether a patent may be awarded, and the scope of the right granted.

It is tempting to formulate judgments about the relative strength of national patent laws based solely on patentable subject matter and the scope of the right to exclusivity. However, it is important to note that patent laws have many elements. Subject matter and scope are only two aspects of any bundle of patent rights. Patent subject matter, criteria, permissible invalidity and infringement challenges, claims construction tests, and drafting rules all play important roles in patent law as well.

### 5.1 Substantive Patent Law

***Patentable Subject Matter.*** A threshold issue for patentability is whether the subject matter of the patent falls within the categories of patentable subject matter set by national patent law. In some jurisdictions, such as the United States, what is considered patentable subject matter is broadly interpreted. In the 1980 United States Supreme Court decision of *Diamond v. Chakrobarty*, which revolved around the patentability of a genetically modified bacterium capable of metabolizing oil, the Supreme Court noted that U.S. patent law embodies Thomas Jefferson's philosophy that "ingenuity should receive a liberal encouragement." [33]. The court made the famous observation that an inventor should be able to patent "anything under the sun that is made by man." Based on this justification, the United States has gone on to grant patent rights over innovations that are excluded from patentability in other countries. One such category is plant patents. The USPTO also grants patents over culture methods, differentiated cells derived from human embryonic stem cells, and the human embryonic stem cells themselves [66], subject matter that is considered not to be patentable in other jurisdiction because of moral concerns.

Adopting a narrower definition of patentable subject matter, other countries have drawn lists, of varying lengths, of inventions explicitly excluded from patentability. In Europe, the EU adopted Directive 98/44/EC of the European Parliament on the legal protection of biotechnological inventions as a standard to be adopted by all its member states and complied with by the EPO [67]. The provisions of the directive outline a restrictive approach to patenting biotechnological inventions, stating in Article 6 that inventions, including "(1) processes for cloning human beings; (2) processes for modifying the germ line genetic identity of human beings; (3) uses of embryos for industrial or commercial purposes; processes for modifying the genetic identity of animals which are likely to cause them suffering without any substantial medical benefit to man or animal, and also (4) animals resulting from such processes," and "where their commercial exploitation would be contrary to *ordre public* or morality" are not patentable. Recital 38 of the

directive confirms that this list is not exhaustive and that any process whose application offends against human dignity is also excluded from patentability.

The EPC specifically limits patentable subject matter for patent applications filed with the EPO, excluding inventions that are contrary to *ordre public* or morality, as do many national European patent systems [65, 67]. The EPC further identifies discoveries, scientific theories, mathematical methods, aesthetic creations, games, business methods, computer programs, presentations of information, and methods for treatment of the human or animal body by surgery or therapy and diagnostic methods as unpatentable [65]. Furthermore, the human body, at the various stages of its formation and development, and the simple discovery of one of its elements, including the sequence or partial sequence of a gene, cannot constitute patentable inventions. However, an element isolated from the human body or otherwise produced by means of a technical process may constitute a patentable invention, even if the structure of that element is identical to a natural element. Plant and animal varieties and essential biological processes for production of plants and animals are also explicitly excluded from protection [65]. In keeping with the requirements of TRIPs, however, microbiological processes and their resultant products are patentable [65]. In compliance with Article 6 of the directive, the EPO has amended the EPC to include Rule 23d, which states that the list of exclusions included in Article 6 of the directive will likewise be unpatentable under the EPC.

In practice, genetically modified mice have been found to be patentable because they provide research platforms with significant benefits for research into human diseases [43]. Human embryonic stem cells, on the other hand, are not patentable on moral grounds because of the constraints imposed by Rule 23d(c) of the EPC, which prohibits the patenting of the "human embryo"[66]. The EPO, in two leading rulings now under appeal, has interpreted Rule 23(d) to exclude not only process patents for creating stem cell lines, thereby destroying the embryo, but also patents derived from already existing stem cell lines. Adult stem cells are not caught within the moral exclusion of subject matter from patentability.

The national patent laws of European nations have not been as well tested as the EPC provisions and the application of the directive within jurisdictions is open to interpretation and therefore can be adjudicated in variant manners. For example, the United Kingdom has interpreted the legal prohibition of patenting human embryos more narrowly than the EPO [66]. The United Kingdom Patent Office has issued a Practice Notice stating that although patents will not be granted for processes of obtaining stem cells from human embryos, nor for human totipotent cells because these have the potential to develop into a human being, patents will be granted for pluripotent stem cells, which develop from further division of totipotent cells and no longer have the ability to develop into a human being. Both the United Kingdom and Sweden have issued stem cell patents filed by the Wisconsin Alumni Research Foundation. The U.K. practice has not yet been challenged on moral grounds in the national courts.

Japanese patent laws are different still from those of either the United States, EPO, or individual European nations. Japan identifies a broad range of inventions as patentable. It does, however, impose two limits on patentability; those innovations that contravene public order or morality are deemed unpatentable [68].

The Canadian Patent Act identifies several innovations as nonpatentable subject matter, including games, medical treatments, scientific principles, and abstract

theorems. As has been discussed earlier, officially higher life forms are not patentable in Canada [43], although recent case law suggests that the rights associated with patent protection are extended to genetically modified higher life forms, such as plants, that are made up of patented genes [44].

There is a lack of clarity on the patentability of biotechnological innovation, which can have a number of unwanted outcomes. The scope of the right over some biotechnology inventions, such as genetically modified plants, can be difficult to determine. Such uncertainty may lead to increased litigation for lucrative inventions. The ethical and moral boundaries of patent rights are unknown and are differentially applied, which can create problems for international technology transfer. The same invention can receive very different treatment as patentable subject matter depending on the region or nation where an application is filed.

***Patent Criteria.*** Patent criteria are applied to determine whether a product or process described in a patent application is an invention. If the product or process fails to meet the criteria, the innovation is not patentable. The deliberation on whether an invention meets the patent criteria is distinct from whether the invention falls within the category of patentable subject matter. An invention may meet all the patent criteria, but its subject matter may fall into a category that *a priori* cannot receive patent protection.

Most countries have three criteria that must be met for patent protection to be granted. Among countries, the criteria are often similar in nature and intent, but not identical. These differences can have a significant effect for biotechnological innovations. In the United States and Canada, the criteria are utility, novelty, and non-obviousness. The EPC defines the criteria as industrial application, novelty, and inventive step. In Japan, the criteria are industrial application, novelty, and inventive step.

*Utility or Industrial Application.* Utility and industrial application are similar, but not identical, criteria. In the United States, the definition of utility is not mere functionality but, rather, whether the invention is useful for its described purposes. For inventions involving DNA sequences, utility is considered to be established if a substantial, credible, and specific use is demonstrated in the patent application. In Canada, utility is understood to mean whether the invention works [69]. It does not have to work well or for a prolonged period of time, nor does a model or prototype have to be presented. Simple evidence that the invention, as it is claimed in the patent application, will work is sufficient.

Industrial application, the standard applied by the EPO and several European countries, is a narrower criterion than utility because it requires an invention to be applicable for use in industry. However, the interpretation of what constitutes an industrial application varies among national patent laws. For example, in the United Kingdom, an invention must be capable of being made or used in some kind of industry, which is broadly defined as anything that is distinct from a purely intellectual or aesthetic activity. Industry does not mean commercial use *per se*, and it includes agriculture [70]. In France, industrial application is more narrowly defined than in the United Kingdom, as industry is interpreted to mean an endeavor that turns a profit. Sweden applies an interpretation similar to the United Kingdom: An invention having industrial application must be of a technical nature, have technical effect, and be reproducible [71]. In Germany, an invention has industrial

application if it can be made or used in any kind of industry, including agriculture. The German definition is similarly worded to the provisions of the United Kingdom and France, but it is much more narrowly interpreted so that manufacture or use must occur in an industrial enterprise.

The European Biotechnology Directive, which must be adhered to by all EU member states and the EPO, sets out specific requirements for the industrial application of DNA patents. The directive states that an invention involving DNA must indicate the function of the DNA and "in order to comply with the industrial application criterion, it is necessary in cases where a sequence or partial sequence of a gene is used to produce a protein or part of a protein, to specify which protein or part of a protein is produced or what function it performs" [67].

The definition of industrial application included in the Japanese patent law is similar to that imposed by the EPO in the EPC [72]. However, the fact that Japanese patent applications are rarely refused for lack of industrial application suggests that Japan's application of this criterion may be less stringent than that of the EPO.

*Novelty.* The novelty criterion simply identifies whether the innovation is truly new, meaning it is not previously known or existing. To determine novelty, patent offices conduct a search for prior disclosure, including prior art materials that describe the invention and were available publicly before the date the patent application was filed. Patent offices in different countries conduct searches with varying degrees of diligence.

One factor that determines the breadth of the search undertaken is the definition of what constitutes disclosure as spelled out in national patent law. The United States considers use of the invention, printed publications, or patents from any country that pre-date the day when the applicant created the invention [73]. The EPO criterion includes use as well as both oral descriptions and written descriptions of an invention pre-dating the filing of the patent application [65]. Oral disclosure is a hot issue in the debate over the patentability of traditional knowledge, as traditional knowledge is often passed down orally rather than in written form, and, accordingly, patent laws that only recognize written disclosure do not consider oral information to represent a barrier to patentability. Japan encompasses public knowledge or public working of inventions and printed publications that precede the filing date of the patent application [68]. In Canada, written descriptions of the invention as well as public disclosure, which may include oral disclosure, pre-dating the filing of the patent application all constitute disclosure [74].

Some prior disclosure of an invention is permissible. In the United States and Canada, the inventor is granted a one-year grace period during which he can disclose the invention before filing a patent application without the disclosure barring the right to a patent [73, 74]. Without this grace period, as soon as an invention is disclosed, it can no longer be considered novel, and therefore is unpatentable. Japan and the EPO allow only a six-month grace period [65]. The conditions for the grace period under the EPC are very stringent; only disclosures because of inclusion in an officially recognized international exhibition or that is caused by an "evident abuse" (e.g., breach of a confidentiality agreement) are permissible. Some European national patent laws, such as those in force in the United Kingdom, do not permit a patent application to be filed after any public disclosure of an inven-

tion [75]. A prudent inventor will delay disclosure of an invention until he is certain of rules in the jurisdictions where patent rights will be sought or until after an application is filed.

Another factor affecting the scope of prior art searches is the amount of time that examiners have to devote to a search, which is directly related to the staffing levels and expertise of examiners working in national patent offices. For some examiners, searching time is short and prior art searches are consequently limited [76, 77], which has implications if a patent is later challenged on the ground of novelty. A search for prior art conducted by an examiner that is limited in scope because of time constraints may fail to discover prior art, which may later be discovered by persons wishing to oppose the patent. Any prior art can be used to challenge a patent and ultimately render the patent invalid for lack of novelty.

The novelty criterion further identifies a distinction that is particularly relevant to biotechnological innovation—the difference between an invention and a discovery. An invention is a product or process that did not exist previously, whereas a discovery refers to something that already exists and is a product of nature [78]. In practice, albeit somewhat counterintuitively, the discovery/invention distinction means that a gene sequence that exists *in situ* is not patentable, whereas an isolated gene sequence, with further restrictions in some countries, is patentable [79]. When a patent is sought for a gene, it is, in fact, sought for the complementary DNA (cDNA) sequence. That molecule is a manmade product derived from nature. Thus, although patentability of human body parts is controversial, isolated human genes, proteins, and cell lines are patentable.

*Non-Obviousness or Inventive Step.* Non-obviousness or inventive step is the third criterion applied to establish whether an innovation meets the definition of an invention set by patent legislation. Non-obviousness invokes the expertise of a person skilled in the art and asks whether such a person would have considered the invention to be obvious in light of the ordinary knowledge of the art [80]. The date of the relevant knowledge of the art differs between countries. The United States requires a consideration of knowledge existing when the invention was made, whereas Canada requires it as of the claim date of the patent application. Canadian jurisprudence distinguishes obviousness and novelty in the following manner: saying that an invention is obvious if a person skilled in the art sees it and says, "any fool could have done that," whereas novelty invokes a reaction of "your invention, though clever, was already known" [81].

Inventive step is a similar criterion to non-obviousness. The EPO actually defines inventive step as existing if an invention is non-obviousness in regard to the state of the art. Japan's test states that there is no inventive step if the invention could have been easily made before the filing date by a person skilled in the art on the basis of pre-existing inventions [68].

As is true of utility/industrial application and novelty as well, each of the provisions for non-obviousness and inventive step are similar. It is, therefore, the wording and interpretation of each nation's criterion that is critical because it ultimately determines if an invention is patentable within that jurisdiction. These variations in national patent criteria may cause an invention to be denied patent protection in one country while being granted such protection in another.

## 5.2 Procedural Patent Law

***The Patent Application.*** A patent is a legal and technical document generally drafted by a patent agent. It at once discloses information about the invention and establishes the boundary around the rights given to the patent holder and is intended to give fair warning about infringing activities. To this end, the drafting of patents is a highly technical art, controlled by strict rules and procedural requirements [82]. For example, patents must be drafted clearly and completely so that the invention can be directly performed by a person skilled in the art. However, a patent rarely results in such clear and compre-hensible disclosure. Key details of the optimum manufacture or working of an invention are often withheld to preserve marketplace advantage because patent documents are available for public perusal.

Patents are generally made up of four parts: an abstract, a description of the invention, one or more claims, and any drawing or technical illustrations referred to in the description or claims. The exact nature of these components may vary from nation to nation. The abstract is a concise summary of the invention and generally identifies the field of art and the technical problem the invention addresses. In some jurisdictions, a word limit is imposed on the abstract [83]. The drawings provide a representation of embodiments of the invention. Generally, elements of the invention that are depicted are numbered for the ease of reference. The description, also known as the specification, provides background information about the field of art of the invention and prior art, addresses why the invention overcomes a technical problem not addressed by the prior art, and offers a complete description of the invention in its various embodiments. The description may reference the drawings for the purpose of better describing the invention. As has been discussed, if an invention involves biological material that cannot be adequately described, the applicant may deposit a sample of this biological material at a recognized institution or depository, depending on the rules of the jurisdiction where the patent is filed [21]. Special rules also exist in some jurisdictions for applications relating to nucleotide and amino acid sequences [65]. Above all else, the description is intended to support the claims. To fulfill this role, the description may be relied on to offer insight into the intention of the inventor for the purpose of any interpretation of the claims.

The claims are the most important element of a patent because only the elements of an invention that are claimed are granted patent protection. Claims are highly technical and challenging to draft because of the strict rules governing their formulation and organization.

Claims can describe various types of inventions or embodiments of the same invention. The most common types of claims are product claims and process claims. Product claims confer protection over any aspects of a product specified, regardless of how that product was derived. Process claims describe a particular means of making a product, and it is the described process that is granted patent protection. Also relevant for biotechnology patents are Markush claims, which are used where embodiments of the invention can include a class of compounds (e.g., chemicals or DNA sequences). By opting to use a Markush claim, the applicant need not spell out every possible combination individually, instead he can claim a range of products that may be used in an embodiment of the invention. This range is compiled on the basis of a limited number of representative samples.

Every patent application must include all of the sections prescribed by national patent law. A patent right may be refused on the basis of an application that is deficient. However, a well-drafted patent application can result in a grant of significant patent rights and great economic value for an owner.

### *Patent Filing*

*First-to-File versus First-to-Invent.* The United States applies a "first-to-invent" patent system, which means that if two different patent applications describe the same invention, the patent right will be issued to the invention that can be proven to have been invented first [84]. By contrast, the patent law of Canada, the EPO, and Japan, impose a "first-to-file" system, meaning that if two patent applications for the same invention are filed, the patent right will be granted to the applicant who filed an application at the earliest date [85], which is true even if the two identical inventions were arrived at independently. For this reason, first-to-file systems are often referred to as creating a "high stakes race" [86].

*When to File.* The first step toward obtaining patent protection for an invention is to file a patent application. The date the patent application is filed is a key reference point for the patent examination and establishes the beginning of the patent term. In accordance with TRIPs, most countries now grant patent rights for 20 years beginning on the earliest filing date [49]. Considerations of the desired patent term and the possibility of being usurped by competitors should dictate when a patent application is filed. Patent term is important because an invention that is in early-stage development and will not reach the market for many years may benefit from a decision to delay filing. A later filing date will ensure that a greater period of the patent term will occur while the invention is marketable (e.g., if the invention takes 15 years post-filing to complete research and development, it will only be on the market during 5 years of the patent term, so waiting to file can extend the period of patent protection while the product is on the market). The position of competitors is important because a delay in filing a patent means that a competitor working on an identical or similar invention may file an application first, and thereby extinguish patent rights for others in the future.

The decision of when to file a patent application can be tricky for complex biotechnological inventions that involve a long research and development period. In the case of the NF-κB patent application, it took 16 years to complete the review at the USPTO before the patent was issued [87]. This significant review period was, at least in part, caused by the large number of claims that were required because of the complexity of the invention.

*How to File.* There are three possible ways to file a patent application, as a national patent, a regional patent, or a Patent Cooperation Treaty (PCT) filing. The decision of which type of patent filing to pursue will depend on commercial, strategic, and cost considerations. For example, the cost of filing a PCT patent is greater than filing in a national patent office. Strategic considerations may relate to the speed of the grant of a patent under each type of application, which is often related to the type of patent review employed. The PCT includes an extended review period of 30 months pre-national filing, the EPO requires a full examination of the patent, whereas some national patent offices require no examination (e.g., Belgium, the Netherlands, Switzerland, and Ireland).

National patent offices act within the boundaries of the country they exist within and apply that country's patent law. Patent applicants may file applications directly with national patent offices in each country where patent protection is required. The decision of where to file an application is generally based on the predicted potential markets for the invention. For example, a crop plant genetically modified to be resistant to frost may only be marketable in countries with colder climates and with markets open to genetically modified foods. Another consideration is the patentable subject matter allowed in each jurisdiction. For this product, an inventor can seek patent protection for a plant in the United States, but not Canada or Europe. The most cost-effective strategy for an invention that is only marketable in a small number of countries is to file patent applications directly with individual national patent offices.

A further consideration is the differences in disclosure rules set by different countries, as discussed above under "novelty." As has been discussed, if an invention is publicly disclosed before an application is filed, it may limit the countries in which patent protection may be sought. For example, once disclosure is made, an inventor may be barred from obtaining a patent in the United Kingdom. For this reason, disclosure should never occur before the decision about where patent rights will be sought, as patent applications themselves are a form of disclosure.

In some circumstances, patent filings can rely on each other for the purpose of establishing an early filing date. The Paris Convention allows for priority filings among signatory nations. A priority filing is one that relies on the filing date of a previously filed patent application. For example, if on April 19, 2006 an inventor files a patent application for protein A in the United States, a Paris Convention signatory, a second patent application for protein A can be filed in Canada, also a Paris Convention signatory, up to 12 months later. If the Canadian application is identified as a priority filing based on the U.S. application, then the priority date granted to the Canadian application will be the filing date of the U.S. filing—April 19, 2006—which takes off some of the pressure of deciding exactly what type of patent to file immediately while holding an early filing date for subsequent filings to rely on, which is crucial in a first-to-file patent system.

Inventors who wish to obtain patent rights in multiple countries represented by a region may file applications with a regional office, if such an office exists. The EPO is a regional patent office serving several European countries, and ARIPO accepts applications for multiple African nations. Regional patent offices allow for a filing process that reserves rights in all nations within the region simultaneously. For example, once an inventor files a patent application with the EPO, the right to national patent protection in each of the participating European nations is reserved, so that after a successful examination, the inventor is granted a patent right in each of the individual nations represented by the EPO.

Regional patent offices are distinct from national patent offices. National patent offices not only undertake the filing of patents, they also administer the national patent laws. Regional patent offices are limited to undertaking the grant of patents. Post-grant patent rights will be dealt with in accordance with the national patent laws of individual nations. For example, an infringement dispute erupting in France over a patent application that was filed with the EPO will be resolved in accordance with the national patent laws of France.

Regional offices, in developing nations in particular, play another key role besides granting patent applications. They allow for the integration of services and free flow of information. ARIPO was mainly established to pool the resources of its member countries in industrial property matters to avoid duplication of financial and human resources [88].

The PCT system facilitates the filing of a patent application on a even broader scale, globally [53]. Upon filing, a PCT patent application reserves for an inventor the right to seek patent protection in all countries that are signatories to the PCT at the time of filing. This option is generally chosen by an inventor for two reasons, first, because the inventor is unsure of the scope of the market for the invention, but believes it could be global, and second, because the inventor needs time to raise investor funds. The time advantage is achieved because the process of PCT examination of a patent application can add several months (at least 30 months) to the filing process, a significant period of time to be able to put off a final determination about where patent rights are required internationally for an invention, especially considering the pressure to file patent applications as soon as possible imposed by investors and competitors, and the expense that is incurred to seek patent protection in multiple countries. A PCT filing offers some breathing space to an inventor. Moreover, an additional time advantage can be achieved if a PCT filing is made as a priority filing based on a previous foreign filing. For example, a filing made in the United States on April 19, 2006 for a widget can be identified as the priority filing for a subsequent PCT application for the widget, so that the PCT filing would be granted the priority date of April 19, 2006. Through this process, the 30-month lapse during which an inventor can decide which countries to pursue patent rights in is increased to 42 months.

It is important to remember that the PCT is not a patent-granting system but only a patent-filing system [89]. Once a PCT patent application has been examined and deemed patentable, the inventor is then required to file national applications in each of the countries where patent rights are desired. The initial PCT filing merely reserves the right to file in each of these nations (so that applications will not be refused because they missed disclosure deadlines, etc.), it does not take the place of national filings—the PCT is a preliminary step to national filings. However, as the patent application has already been examined by PCT examiners, the expectation is that the national patent filing will be a simple and straightforward process, devoid of extensive discussion between the inventor and the examiner. The drawback of a PCT filing is that it is slightly more expensive than normal national filings; a fee is charged for the PCT filing in addition to the filing fees that must be paid in each country where patent applications are subsequently sought. However, for inventions with a truly global appeal, such as with some types of biotechnology, PCT filing may be worth the added cost.

The determination of how a patent application will be filed—nationally, regionally, or PCT—can greatly affect the ultimate scope of the freedom to operate (FTO) achieved for a commercial patent owner. Evidence exists that for biotech start-up companies, FTO considerations are surpassing the previous focus on exclusion of competition [90]. Issues such as freedom from third-party contracts and properly timed disclosure can also affect FTO; but, ultimately, securing the right to seek patent rights in key marketplaces will ensure significant FTO success.

### 5.3 Legal Challenges: Infringement, Exemptions to Infringement, and Validity

***The Scope of Patent Protection.*** The scope of patent protection granted to an invention defines the breadth of the patent right, sometimes referred to as the fence around the invention [80, 91]. Only when armed with an understanding of patent scope can a party know which activities will constitute infringement of the patent. The interpretation of patent protection differs by jurisdiction. In some countries, patent protection is limited to commercial activities involving the invention whereas in other jurisdictions the range of activity is broader. For example, in the United States and Canada, patent legislation does not limit patent protection to the commercial working of the invention. In contrast, in Japan, the patent holder's right is limited to the commercial application of the patented invention. For patents granted by the EPO, the applicable patent protection applied in an infringement proceeding is prescribed by the national patent law in the jurisdiction in which the alleged infringement occurred; but, generally, most European states impose a level of industrial application as a boundary for patent protection. The infringement test and specific details of infringement may also differ from one jurisdiction to the next.

***Infringement.*** The test for infringement varies by jurisdiction. A U.S. patent is infringed when an unauthorized person makes, uses, offers to sell, or sells any patented invention within the United States or imports a patented invention into the United States [73]. Those who actively induce infringement or who sell or import critical components of inventions, knowing the sale will aid others to infringe the patent, may also be found liable for contributory infringement.

Infringement proceedings occur before a national court in the United States. The test for patent infringement asks whether the patent has been either literally or substantively infringed. The test determines the scope of the invention claimed in the patent and decides whether the patent has been infringed as a related consideration. Literal infringement means that the product or process of the invention has been duplicated as it was claimed. Substantive infringement, on the other hand, also known as the doctrine of equivalents, considers the "spirit of the invention" [82]. The question here is whether the alleged infringing product or process copies aspects of the invention that, although not being expressly included in the text of the patent, should nevertheless be considered part of the invention as it was conceived of or intended by the inventor. Thus, the substantive test is not limited to the text of the patent, but can be read into the document.

A further aspect of the U.S. infringement test that is unique is the doctrine of prosecution history estoppel. The doctrine prevents a patent holder from alleging substantive infringement based on a claim, or portion of a claim, that formed part of the original patent application but was removed during the office action stage [92]. The office action stage is the examination process when the patent applicant and patent examiner enter into discussion about a filed patent application. Generally, during the office action stage, the patent examiner will refuse claims included in the patent application on the basis of prior art; the patent applicant will either agree or disagree with the challenge. As a result of this process, the claims of the patent application will be modified at the behest of the patent examiner. The result

is that the claims included in the filed patent application are often entirely different from those included in the granted patent, as claims are often reduced in scope or removed altogether during this process. Prosecution history estoppel holds that any element of an invention removed during the office action stage cannot later be considered protected by the patent as a result of any interpretation of the claims.

The remedies for infringement available in the United States include an injunction (court order to stop the infringing activity or to destroy the infringing product) or damages (monetary or other compensation calculated to reimburse the patent holder for the losses suffered) [73]. Attorney fees may also be claimed.

As has been stated, infringement of patents granted by the EPO will be evaluated in accordance with the national patent law of the jurisdiction where infringement is alleged to occur. As an example, infringement occurs in the United Kingdom if there is disposal of, offer to dispose of, use, or importation of a patented product or a product created using a patented process or unauthorized use of a patented process in the United Kingdom [93]. Both substantive and contributory infringement are prohibited. Proceedings may occur before a court or comptroller, and possible infringement claims include an injunction, an order to deliver or destroy the patented product, damages, an accounting of profits (a calculation of the profits of the infringer derived from the infringement), and a declaration that a patent is valid and has been infringed. The test applied to infringement is the Catnic test [94], which first establishes patent scope by reducing the invention into essential and nonessential elements. Only the essential elements are then considered in the subsequent evaluation as to whether infringement occurred. The considerations of patent scope and infringement are distinct steps in the test, and scope is determined in a manner that takes no account of the infringement allegation.

Japan applies a literal/substantive test, similar to that used in the United States and has adopted the doctrine of equivalents [95]. The Japanese version of the test, however, also takes into account the weight or importance of the claimed elements of the invention in a manner similar to the United Kingdom [96, 97]. Infringement proceedings occur within the courts but rarely rely on oral arguments as the majority of the proceedings are presented in the form of written briefs. Damages may be claimed as a remedy for infringement, as can lost profits. Infringement awards were low for a long period in Japan, but evidence exists that these have been rising steadily since 2002 [98]. The burden of proof is considered to be higher in Japanese infringement proceedings than in U.S. cases (i.e., it is more difficult to prove infringement in Japan than in the United States).

In Canada, anyone who makes, constructs, uses, or sells a patented invention without authorization to do so can be found guilty of infringement. An infringement action may proceed in the higher courts in each province or in the more specialized federal court [74]. Remedies include damages, injunction, and accounting of profits. In Canada, as is true in most jurisdictions, an infringement allegation can only be raised after a patent is granted; however, Canada is unique in that the damages claimed can include a period of time between the date when the patent application was disclosed to the public and the patent was granted, in addition to the post-grant period [74].

The test applied to determine infringement in Canada is the purposive construction test [99, 100]. This test is based on the U.K. Catnic test, although the tests are not absolutely identical. The first step of the test is a determination of the scope of

a patent, which involves the break down of the claims into essential and nonessential elements. As is true of the U.K. test, this step is undertaken without any consideration of the infringement allegation. An important distinction between the Canadian test and that of the United States is that it is limited to the "four corners" of the patent document, whereas the U.S. test allows for consideration of extrinsic documents, such as the prosecution history, otherwise known as the file wrapper information. Once patent scope is determined, the Canadian test compares the alleged infringing innovation to the essential elements of the patented invention. No infringement occurred if an essential element is different or omitted in the alleged infringing device; but infringement may have occurred if nonessential elements are substituted or omitted.

Infringement is a common challenge raised in the biotechnology patent world against competitors. Myriad Genetics, as described above, applied the threat of an infringement suit to great advantage. Ariad Pharmaceuticals, owner of the NF-κB patent, likewise has "gone to extraordinary lengths to ensure that anything that comes within so much as a whiff of NF-κB will be drawn into the '516' patent's black hole" [87].

***Exemptions to Infringement: Exemptions for Research and Generic Drug Manufacturing.*** Certain activities are exempt from patent infringement challenges, and, again, these vary according to national patent laws. Exemptions may be raised to defend against allegations of infringement. The most pertinent exemption for biopharmaceuticals and biotechnology is the research exemption. The research or experimental use exemption curtails a patent holder's rights by permitting researchers and research institutions to make certain uses of a patented invention without compensating the patent holder. Many policy documents argue that the research exemption for free access to basic research materials should be strengthened (see discussion below) [101]. However, it is becoming increasingly difficult to justify a research exemption as the lines between private-sector and public-sector research blur and the commercialization activities of public-sector research institutions increase.

In the United States, no statutory research exemption exists; instead, a very narrow experimental use exemption has been carved out by judges and exists in the common law [102, 103]. Previously, the exemption did not apply if the research has "the slightest commercial implication" [104]. However, the decision of the United States Court of Appeals for the Federal Circuit in *Madey v. Duke University* [105] has further narrowed the exemption. Now, any conduct by a research institution that is in keeping with its legitimate business, regardless of the commercial implications of the research, cannot rely on the research exemption. In the case of *Madey*, Duke University was using patented equipment for teaching and research. The court characterized both of these activities as part of the legitimate business activities of the university and therefore decided that Duke University could not be exempted from patent infringement. The result is that the research exemption has been restricted only to the most exceptional circumstances [102].

Most importantly, for biopharmaceuticals, however, and the eventual entry of generic products of biotechnological research into the marketplace (see discussion below) is the Restoration Act or Hatch–Waxman Amendments to the Drug Price Competition and Patent Term Restoration Act of 1984 [106]. This act creates an

exemption to the normal rules governing infringement for parties seeking Food and Drug Administration (FDA) approval for generic drugs. Section 271(e) of the U.S. Patent Act now states that "use of a patented invention solely for purposes reasonably related to gathering data to support an FDA application for generic versions of [patented] drugs" does not constitute infringement. However, the generic drug manufacturer, although having the ability to do research to show that their drug is equivalent to the patented original, must still wait until the patent expires before it can seek FDA approval.

In Europe, the issue of infringement is still largely determined by national patent law. Article 31 of the Community Patent Convention creates an exemption for acts "done for experimental purposes relating to the subject matter of the patented invention" [107]. All members of the EU except Austria have introduced a similar provision into their national patent laws [102]. However, the interpretation of the scope of the exemption still varies between countries. For example, in Germany, the scope of the experimental use exemption is broad and applies to activities that test the viability and potential for development of patented inventions. In the United Kingdom, however, the exemption is narrower. It does not apply for experimentation in the context of regulatory approval in contrast to Canada, the United States and other countries of the EU [102].

Infringement exceptions in Japan's patent law include some research uses [68]. Further exemptions include, in the United States patented inventions in use upon vessels, aircraft, or vehicles that are in the Unites States on a temporary basis (and will be used solely aboard the vessel, aircraft, or vehicle); and if a person other than the patent holder reduced the invention to practice at least one year before the filing date of the patent, that person can continue to use the invention [73].

The United Kingdom also exempts patented inventions used on an aircraft that is temporarily landing in or crossing over the United Kingdom [93]. Canada offers a similar exemption for ships, vessels, aircraft, or land vehicles within Canada, allowing temporary use of patented inventions solely for their needs and upon the craft; and they also acknowledge that any prior holder (meaning a person who had the patented product in their possession before the claim date of the patent) has a right to continue to use the product [74].

***Challenging the Validity of a Patent.*** The validity of the claims made in a patent is likely to be challenged in one of two situations: directly in a post-grant opposition or validity challenge proceeding, depending on the jurisdiction; or alternatively, as a defense to an action for infringement. The latter use of a validity challenge may be formulated in a variety of ways, but will most likely argue that the invention is not patentable subject matter, that the patent criteria is not met (e.g., prior art exists that establishes that the invention is not new), or that claims are invalid (e.g., claims are not supported by the description or are affected by a technical deficiency). In some countries, pre-grant validity challenges may be raised as well; for example, in Canada, a prior art filing process can be invoked [74]. This challenge involves the delivery of prior art that proves that the invention was not novel at the date of filing to the patent office by a person wishing to thwart the grant of a patent. In Canada, such challenges are unpopular because the person filing the prior art has no right to involvement in the proceedings after delivery to the patent office and is bound by the decision. Canadian validity challenges are normally only raised

after a patent has been granted, at which point the challenger can have full involvement in the proceedings and a right of appeal.

An opposition proceeding is a form of post-grant validity challenge available for EPO patents [65]. As the EPO is solely a patent-granting organization, the effect of the issuance of an EPO patent is effectively to create multiple patents from a single application, one patent in each of the European countries designated by the applicant during the patent prosecution stage. However, if an EPO opposition proceeding is launched within nine months of the granting of the EPO patent, it has the effect of pulling back each of the national patents, so that a single EPO patent exists [108]. It is the validity of this single EPO patent that is challenged. The benefit of the opposition proceeding is that only one challenge needs to be made; this option is more cost-effective than pursuing multiple proceedings in each individual European nation where the patent exists post-grant. An opposition proceeding may be brought by any person and is based on three grounds.

(1) that the subject matter is not patentable under European law;
(2) that the European patent does not disclose the invention so that it could be carried out by a person skilled in the art; and
(3) that the subject matter of the European patent extends beyond the content.

In the United States, post-grant oppositions are under discussion as a major feature of proposed patent reform [109–111]. Advocates of opposition proceedings hope that allowing an opposition to be filed in the USPTO within a limited time window, immediately after the patent issues, would help weed out invalid patents in a procedurally simple, relatively inexpensive forum. Presently, a re-examination process is set out in the U.S. patent law, which is a challenge to the validity of a patent based on the presentation of prior art [73]. However, the outcome of the re-examination is binding and no right to appeal is granted, and, like the Canadian filing of prior art process, the re-examination process is rarely used [108].

In Japan, the opposition procedure after an examiner's decision to grant a patent was abandoned in 2005. The alternative is a trial for the invalidation of the patent [112].

Canada does not offer an opposition procedure, but includes a reexamination procedure in its patent law. Re-examination may be requested for any claim or claims at any time during the term of the patent, but the evaluation is undertaken solely on the basis of prior art [74]. The Commissioner of Patents appoints a three-person re-examination board to undertake the review. Unlike the re-examination process in the United States, the board's decision is not the end, and an appeal can be launched within three months.

## 6 THE EFFECT OF THE REGULATORY ENVIRONMENT

Biotechnological innovations that are eligible for patent or other IPR protection can simultaneously be regulated by relevant IPR laws and other legislation or regulation. For example, an overlap between patent law and legislation directed specifi-

cally at biologic products may affect biopharmaceuticals. Biologics, as defined by the FDA, "in contrast to drugs that are chemically synthesized, are derived from living sources (such as humans, animals, and microorganisms). Most biologics are complex mixtures that are not easily identified or characterized, and many biologics are manufactured using biotechnology. Biological products often represent the cutting-edge of biomedical research and, in time, may offer the most effective means to treat a variety of medical illnesses and conditions that presently have no other treatments available" [113]. A biotechnological innovation that meets the definition of biologic and is granted patent rights will need to comply with food and drug legislation as well as patent law.

Simultaneous compliance with two sources of legislation has the potential to affect the scope of patent rights. For example, the parameters for the introduction of generic versions of the patented drug into the marketplace are set not only by patent law but also by food and drug legislation. In Canada, the patent regime has been tied to the food and drug legislation to address concerns that patent rights over drugs may negatively affect public health by limiting access to drugs. Cooperation between the food and drug regime and the patent regime also ensures that the term of a patented drug is respected. Specifically, legislation prohibits generic drug manufacturers from introducing an equivalent drug into the marketplace during the patent term, thereby introducing patent considerations into the regulatory approval system.

### 6.1 Biosimilars

An emerging issue in the field of biopharmaceuticals is the issue of biogenerics and how these may be introduced into the marketplace. Biogenerics are the generic version of the original, usually patented, biologic product; the same relationship is found between the original pharmaceutical and its generic counterpart. The primary distinguishing feature between biologics and traditional pharmaceuticals is the former uses recombinant DNA technology to produce macromolecules whereas the latter conforms to a "chemical synthesis/small molecule paradigm" [114]. Diverse biotechnological products include hormones, cellular growth factors, enzymes, clotting and anticlotting factors, monoclonal antibodies, and fusion proteins. The very diversity and complexity of biologics leads to problems in setting the regulatory environment.

To date, no specific regulations or universally applicable guidelines exist governing biogenerics. Each product is reviewed on a case-by-case basis according to existing regulations governing generic pharmaceuticals. Such regulations exist in most jurisdictions and provide a simplified approval process based on the proof of "essential similarity" or product equivalence of the generic pharmaceutical compared with the patented pharmaceutical already available in the marketplace. Equivalency may be in terms of chemical composition, pharmaceutical equivalence or dosage forms, and thera-peutic equivalence. Generic pharmaceuticals are almost identical to the original product or even contain an active ingredient for which the patent has expired.

The problem with biopharmaceuticals is that these are generally characterized by properties that may complicate the demonstration of equivalence because of the difficulty of characterizing the physico-chemical makeup [115]. The product

properties, such as the impurity profile, glycosylation, and protein folding, may be highly dependent on the process by which they are manufactured, adding to the complexity of a determination of equivalency [115].

A further problem is the issue of the patent thicket that surrounds much biotechnological innovation. There may be multiple patents held by multiple parties for constituent ingredients, processes required for the development of biopharmaceuticals, and the products involved in those processes (e.g., gene and protein patents). Navigating the thicket to ensure that a biogeneric violates no extant patents may be quite challenging. Thickets greatly impede FTO.

These problems will become increasingly significant as the patents for top-selling biopharmaceuticals expire. Several of the top-15 selling biopharmaceuticals come off patent in the next five years and a number of large generic manufacturers have begun development of biogenerics [115]. Indeed, if a similar market penetration is assumed for biogenerics as for generic pharmaceuticals (10–15%), the worldwide market for biogenerics will be $2–3B U.S. [114, 115].

Some hope exists that the regulatory framework will become responsive to the emergence of biogenerics. The shift will require a move from essential similarity (or demonstration that the generic is almost identical to the origi-nal pharmaceutical) to product comparability [115] and clearer working definitions for different classes of biologics [114]. These issues are under consideration by the FDA, and the European Commission has passed legislation for the approval of "similar biological medicinal products" [116].

## 7 REFORM PROPOSALS FOR INNOVATION GOVERNANCE IN BIOTECHNOLOGY

The rapidly changing area of biotechnology innovation does not necessarily fit comfortably within the confines of traditional intellectual property laws. Some reforms have been made at national and international levels to adapt intellectual property law, particularly patent laws, to biotechnology. These reforms generally take the form of specific provisions and adjunct statutes, regulations, and practice guidelines within patent offices designed to address patent rights in biotechnology. Many of the reforms are driven by moral and ethical concerns that develop from the research *per se* and discomfort with the commodification of products and processes that are perceived to belong to nature or life. However, some more pragmatic concerns exist, developing from considerations of the appropriate balance to be struck between the rights of innovators in biotechnology and access to the innovation by the public, by government, and by researchers, among others. In this final section, we explore some of the reforms that have been proposed as solutions or compromises to moral and ethical dilemmas and to the issue of striking the right balance for patent protection of biotechnology innovation.

### 7.1 Strengthening Research Exemptions

Many jurisdictions are currently considering reform proposals for a broad research exemption from patent infringement [117, 118]. The concerns prompting the call for policy reform are set against a backdrop of rapid growth of DNA-related

patents, particularly in the United States with growth of about 5% per annum [119]. It is estimated that over 3000 new DNA-related patents per year have been issued in the United States since 1998 and more than 40,000 such patents have been granted [120]. On the human front, at least 20% of the human genes are associated with at least one U.S. patent, and many have multiple patents [121].

The first concern, raised by Heller and Eisenberg, is the plethora of overlapping patents that may result, in the research context, in a "tragedy of the anti-commons" [122]. That is, the large number and diversity of ownership of gene patents may make it difficult to acquire all of the rights necessary for products and processes required for research (e.g., a patent thicket), resulting in the underuse of valuable technologies and the abandonment of promising avenues of research. Second, concern exists that patent holders will exercise their right of exclusivity in an overly restrictive manner that prevents others from developing or accessing a particular technology, for example, by refusing to license the technology for use by others [123]. The case of Myriad Genetics' zealous enforcement of its BRCA1/2 gene patents to prevent government-run diagnostic laboratories in Ontario from testing for genetic predisposition to breast cancer was such an example. Denial of access to patented genes and other biological products raises particular concerns because of the likely inability to invent around product patents because of the unique role these molecules play in biological processes. Third, the increasing focus on commercialization of public-sector research institutions will undermine norms of open scientific collaboration, especially the sharing of research results and materials [124, 125]. Increasing levels of industry funding will augment this effect and contribute to delays in the publication and dissemination of scientific research.

To date, the empirical evidence indicates that most of these concerns have not materialized [126]. However, real worries exist about the sharing of research materials, resulting in an increasingly competitive environment, which is likely tied to the commercialization ethos, industry funding, and the practical difficulties, such as time limitations, of transferring research materials [126].

### 7.2 Patent Pooling

A combination of the controversial nature of some areas of biotechnological innovation and a general distrust of commercially driven research are two factors that will influence public support and policy setting for both the private research sector and the public research sector. Patent pools may be a potential solution for maintaining public confidence in the research enterprise [42].

A patent pool is an arrangement in which "two or more patent owners agree to license certain of their patents to one another and/or third parties" [127, 128]. Patent pools bring together patent holders in a specific area of innovation, such as a viral genome, to facilitate the efficient use and development of a technology. The patents are "pooled" in the sense that the arrangement allows inventors in the pool to use all patented invention under favorable licensing terms. Any benefits that may materialize are then shared among the group. Patent pools have been used successfully in the motion picture industry, in aeronautics, and in the development of video and DVD technology [129].

As commercial interests are part of the problem, patent pools could be set up as independent, nonprofit corporations with an independent governance structure

in which researchers are removed from direct involvement with industry and commercial forces [42, 130]. The governing body of the patent pool would make decisions about commercialization, access, and licensing. These decisions would be open to public scrutiny to ensure that research is in the public interest and not solely concerned with the profit motive. Such a structure is being considered for the Canadian Stem Cell Network and research related to the SARS genome, for example.

Of course, many challenges face the setting up of a patent pool. Patent pools may trigger anticompetition laws [127]. Researchers may not join in the patent pool because, although patent pools lower research transaction costs and spread risk, they also decrease the potential for the attainment of large profits from an invention. In the end, "the major incentive for all parties is economic benefit. In order for a patent pool to be an effective solution, the right balance has to be achieved between the cost of creating a pool and the prospect of adequate revenue generated by royalties on the endproduct" [131].

### 7.3 Open Source Patents

Another governance model designed to improve access to information and biotechnological research is open source patents. Open source patent systems share the goal of promoting free dissemination of biotechnological research. The aim is to foster an environment of sharing between inventors and the public rather than marketplace monopolies. Open source systems can be directed at endproducts or research tools used to develop products. Some consider open source to represent a grassroots movement to return to precommercialization sensibilities about scientific research and development, where there was a greater ethos of freely sharing scientific information among members of the scientific community [132].

The concept of open source rights as an alternative to IPRs originated with copyright. The rights granted by copyright arise automatically with the creation of a work. Copyright holders are then free to decide who may use the work. Open source systems, often found to protect software available for download on the World Wide Web, are created when copyright holders agree to license their works in a nonexclusive manner to anyone who agrees to abide by certain terms. The terms usually ensure that the copyrighted works themselves and any derivative works will be made available to future users in the same nonexclusive manner. As patent rights do not arise automatically, but must be actively applied for, open source patent systems have more options available for their formulation. Open source patent systems operate in a diverse manner, and some include inventions that are protected by patent rights whereas others apply to innovations that are not patent-protected. The similarity with all open source copyright is that open source patent initiatives build a free flow of information upon the foundation of innovation [133].

Several functioning examples of open source patent systems exist. The Public Patent Foundation (PPF) is one such initiative that facilitates the creation of a commons wherein patents may be pooled and made freely available to other participants [134]. The specific method applied by the PPF is to grant a nonexclusive and payment-free license to participants. The Biological Innovation of Open Society (BIOS) also involves technologies for which patent rights have been granted.

Its operation is focused on research tools rather than final products, and its goal is to "assemble groups of enabling technologies that together provide the pieces necessary for a particular form of research investigation" [134]. Both PPF and BIOS set licensing terms that achieve their specific goals [134].

The Tropical Disease Initiative (TDI) is an example of an open source patent system in which inventions need not be protected by patent rights [135]. Its aim is to create an accessible Web database to facilitate "searching for new targets, finding chemicals to attack known targets, and posting data from related chemistry and biological experiments" [135], which makes information readily accessible by researchers in developing and developed nations. Eventually, TDI hopes its open source system will aid virtual pharmas—organizations that are not engaged in in-house development, but that use a grouping of commercial and academic partnerships to create a portfolio of promising drug innovations [135].

### 7.4 Compulsory Licensing

The debate about compulsory licenses for patented products occurred in the context of access to HIV/AIDs drugs in Africa. At issue was the effect of patent rights on the accessibility and affordability of pharmaceutical products and vaccines in developing nations. Similar concerns around access and cost are likely to develop with biopharmaceuticals and other innovations resulting from research in health biotechnology.

One policy option for ensuring the accessibility and affordability of biotechnological innovation is for the state to implement compulsory licenses. Compulsory licenses override the exclusive patent right granted to a patent holder. A number of reasons why a state may take such a step exist, including public emergency and public health. Many national patent laws include provisions that permit the application of compulsory licenses. TRIPs also permits WTO members to include compulsory licensing within national patent legislation, in the instances of either "national emergency or extreme urgency," and sets out that the invention is to be applied to a "public noncommercial use" [49].

At its Cancun Ministerial Meeting in 2003, the WTO decided to implement changes to enable the least-developed countries to import cheaper generic drugs manufactured under compulsory licensing in other countries in the event that they were themselves unable to manufacture the pharmaceuticals. Calling it an "historic agreement for the WTO," Director General Supachai Panitchpakdi recognized that it was the first time that the WTO both validated and facilitated the full use of TRIPs' flexibility by developing countries [136].

Canada was the first country to introduce legislation to provide access to affordable medicines under the Cancun agreement. Canada's Patent Act has been amended to permit the award of compulsory licenses to export medications to developing countries in need. The amendments stipulate the conditions under which the government can issue a compulsory licence to a pharmaceutical manufacturer for the manufacture and export of a patented drug to the world's developing and least-developed countries. However, it is too early to tell whether the compulsory licensing regime will actually result in increased access to medicines [137, 138].

## REFERENCES


1. World Intellectual Property Office (WIPO), World Intellectual Property Organization Convention (1967). Available: http://www.wipo.int/treaties/en/convention/trtdocs_wo029.html.
2. Bently L, Sherman B (2004). *Intellectual Property 2d*, Oxford University Press, New York. 1131p.
3. Gold R, et al. (2004). The unexamined assumptions of intellectual property: Adopting an evaluative approach to patenting biotechnological innovation. *Public Aff. Quart.* 18:299–344.
4. Centre for Intellectual Property Policy (CIPP). Available: http://www.cipp.mcgill.ca/en. Accessed 21 July 2006.
5. Westerlund L (2002). *Biotech Patents: Equivalency and Exclusions under European and US Patent Law*, Kluwer Law International, New York, pp. 351.
6. Carrier M (2004). Cabining intellectual property through a property paradigm. *Duke L.* 54:1–124.
7. Seymore S (2006). My patent, your patent, or our patent? Inventorship disputes within academicresearch groups. *Alb. L.J. Sci. & Tech.* 16:125–167.
8. Tomlinson N (2006). Tax abuse halting progress? An inside look at patent donations and their tax deductibility. *SW. U. L. Rev.* 35:183–206.
9. *Whirlpool Corp. v. Camco Inc.* (2000). 2. S.C.R. 1067.
10. *Mallinckrodt, Inc. v. Medipart, Inc.* (1992). 976 Fed. Cir. 700.
11. Walterscheid W (1994). The early evolution of the United States patent law: Antecedents. *J. Pat. & Trademark Off. Soc.* 76:697–715.
12. *Lac Minerals Ltd. v. International Corona Resources Ltd* (Canada). (1989). 2 S.C.R. 74.
13. *Coco v. Clark* (UK). (1969). *RPC* 41.
14. Isaacs R (2004). How not to tell all: Find out how to preserve the company's competitive edge by preventing proprietary information from being given away. *Security Mgmt.* 48(5):101–106.
15. The Week's Business: Coca Cola Caper, *AFX International Focus.* 7 July 2006.
16. Shand H (1999). Legal and technological measures to prevent farmers from saving seed and breeding their own plant varieties. In J Alexandria (ed.), *Perspectives on New Crops and New Uses*: ASHS Press, Alexandria, VA, pp. 124–126.
17. National Farmers Union (2004). Undermining farmers' rights to their seed. Available: http://nfu.ca/seedsavercampaign/NFU_Seeds_Fact_Sheet_2.pdf. Accessed 20 July 2006.
18. Vaver D (1997). *Intellectual Property Law.* Irwin Law, Concord, Ontario, p. 22.
19. *Berne Convention for the Protection of Literary and Artistic Works* (1971). Paris, France.
20. Bovenberg J A (2004). Inalienably yours? The new case for an inalienable property right in human biological material: Empowerment of sample donors or a recipe for a tragic Anti-Commons? *J. Law Technol.* 1(4):591–616.
21. World Intellectual Property Organization (1997). *Budapest Treaty on the International Recognition of the Deposit of Microorganisms for the Purpose of Patent Procedure.* Geneva, Switzerland, 1989.
22. Arai H (1999). *Intellectual Property Policies for the Twenty-First Century: The Japanese Experience in Wealth Creation. WIPO*, pp. 1–107.

23. Mills O (2005). *Biotechnology Inventions: Moral Restraints and Patent Law,* Ashgate Publishing, Ltd. Aldershot, UK, p. 195.
24. Einsiedel E F (2001). *Biotechnology and the Canadian Public: 1997 and 2000*. University of Calgary, Alberta, Canada.
25. Einsiedel E F (2000). Cloning and its discontents- a Canadian perspective. *Nature Biotech*. 18:943–944.
26. Priest S, Gillespie A (2000). Seeds of discontent: Expert opinion, mass media messages, and the public image of agricultural biotechnology. *Sci. Engin. Ethics*. 6:529–539.
27. Condit C (2001). What is "public opinion" about genetics? *Nature Rev. Genet.* 2:811–815.
28. Gaskell G, Bauer M, Durant J, et al. (1999). Worlds apart? The reception of genetically modified foods in europe and the U.S. *Science*. 285:384–387.
29. Bubela T, Caulfield T (2005). Media Representations of Genetic Research. In E F Einsiedel, F Timmermans (eds.), *Crossing Over. Genomics in the Public Arena*, University of Calgary Press, Alberta, Canada.
30. Holtzman N, Bernhardt B, Mountcastle-Shah E, et al. (2005). The quality of media reports on discoveries related to human genetic diseases. *Community Genet.* 8:133–144.
31. Page B, Shapiro R (1992). *The Rational Public: Fifty Years of Trends in Americans' Policy Preferences,* University of Chicago Press, Chicago, IL, pp. 67–116.
32. Parliament of Canada (2003). *Assisted Human Reproduction Act*. Bill C-13: LS-434E.
33. *Diamond, Commissioner of Patents & Trademarks v. Chakrabarty.* 1980. 447 U.S. 303.
34. Association of University Technology Managers (AUTM). Available: http://www.autm.net/index.cfm.
35. Caulfield T (2005). ASHG: Human stem cell biology: Principles, applications and ethical considerations—patent law, *American Society of Human Genetics, 55th Annual Meeting*, Salt Lake City, Utah.
36. Caulfield T (2006). Stem cell patents and social controversy: A speculative view from Canada. *Med. Law Internat.* 7:219–232.
37. Morin K, et al. (2002). Managing conflicts of interest in the conduct of clinical trials. *JAMA*. 287:78–84.
38. Lemmens T (2004). Leopards in the temple: Restoring scientific integrity to the commercialized research scene. *J. Law, Med. Ethics*. 32:641–657.
39. Canadian Biotechnology Secretariat Industry Canada (2005). International public opinion research on emerging technologies: Canada-U.S. survey results. Available: http://www.biostrategy.gc.ca/CMFiles/E-POR-ET_200549QZS-5202005-3081.pdf. Accessed 22 July 2006.
40. Pollara and Earnscliffe Research and Communications (2000). Third wave: Executive summary (Prepared for and presented to the Biotechnology Assistant Deputy Minister Coordinating Committee (BACC), Government of Canada). Available: http://www.biostrategy.gc.ca/english/view.asp?x=551&all=true. Accessed 22 July 2006.
41. Nelson D, Weiss R (2000). Penn Researchers Sued in Gene Therapy Death. *The Washington Post.* September 19, A:03.
42. Caulfield T (2004). Biotechnology Patents, Public Trust and Patent Pools: The Need for Governance? Health Biotechnology and Intellectual Property: A New Framework. Hosted at the European University Institute, Florence, Italy.

43. *Harvard College v. Canada (Commissioner of Patents)*. (2002). 4 S.C.R. 45.
44. *Monsanto Canada Inc. v. Schmeiser.* (2004). 1 S.C.R. 902.
45. Morrow D, Ingram C (2005). Of transgenic mice and round-up ready canola: The decisions of the supreme court of canada in Harvard College v. Canada and Monsanto v. Schmeiser. *U.B.C. Law Rev.* 38:189–222.
46. McCalman P (1999). Reaping what You Sow: An Empirical Analysis of International Patent Harmonization. Available: http://www2.cid.harvard.edu/cidtrade/Issues/mccalman.pdf. Accessed 20 July 2006.
47. Hamilton C, Rochwerger P (2005). Trade and Investment: Foreign Direct Investment Through Bilateral and Multilateral Treaties. *N.Y. Internat. Law Rev.* 18:1–59.
48. Wolfe R (2005). Decision-making and Transparency in the 'Medieval' WTO: Does the Sutherland report have the right prescription? *J. Int. Econ. L.* 8:632–633.
49. World Trade Organization (1995). Trade-Related Aspects of Intellectual Property Rights (TRIPs).
50. Blakeney M (1996). *Trade Related Aspects of Intellectual Property Rights: A Concise Guide to the TRIPS Agreement*, Sweet & Maxwell, London, pp. 1–158.
51. World Trade Organization (1998). Agreement on the Application of Sanitary and Phytosanitary Measures. Available: http://www.wto.org/English/tratop_e/ sps_e/sps_e.htm.
52. World Intellectual Property Organization (WIPO). Available: http://www.wipo.int/portal/index.html.en.
53. World Intellectual Property Organization (WIPO) (1970). Patent Cooperation Treaty (PCT). Available: http://www.wipo.int/pct/en/texts/articles/atoc.htm.
54. World Intellectual Property Organization (WIPO) (1970). Summary of the Patent Cooperation Treaty (PCT). Available: http://www.wipo.int/treaties/en/registration/pct/summary_pct.html. Accessed 20 July 2006.
55. Commission on Intellectual Property Rights, Innovation and Public Health. April 2006. Public health, Innovation and Intellectual Property Rights.
56. World Health Organization (2003). *Public Health, Innovation and Intellectual Property Rights*. Available: http://www.who.int/intellectualproperty/documents/thereport/en/index.html. Accessed 20 July 2006.
57. Union for the Protection of New Varieties of Plants (UPOV). Available: http://www.upov.int.
58. World Intellectual Property Organization (1961). Contracting parties and signatories to the International Convention for the Protection of New Varieties of Plants (UPOV). Available: http://www.upov.int/en/publications/conventions/1991/pdf/act1991.pdf.
59. Halewood M (1999). Indigenous and local knowledge in international law: A preface to Sui Generis intellectual property protection. *McGill Law J.* 44:953–996.
60. Shiva V (2005). The Indian seed act and patent act: sowing the seeds of dictatorship. *ZNet*, Available: http://www.zmag.org/content/print_article.cfm. Accessed 20 July 2006.
61. Wendt J, Izquierdo J (2001). Biotechnology and development: A balance between IPR and protection and benefit-sharing. *Electr. J. Biotechnol.* 3, Available: http://www.agbios.com/docroot/articles/01-361-002pdf. Accessed 21 July 2006.
62. Convention on Biological Diversity (CBD). Available: http://www.biodiv.org/default.shtml.
63. Secretariat of the Convention on Biological Diversity (2000). Cartagena Protocol on Biosafety, Geneva.

64. Food and Agriculture Organization of the United Nations (2001). International Treaty on Plant Genetic Resources for Food and Agriculture. Available: http://www.fao.org/ag/cgrfa/itpgr.htm.
65. European Patent Convention (EPC) (1973). (rev. 2002). European Patent Office. Available: http://www.european-patent-office.org/legal/epc/e/ma1.html.
66. Porter G, Denning C, Plomer A, et al. (2006). The patentability of human embryonic stem cells in Europe. *Nat. Biotechnol.* 24:653–655.
67. Directive 98/44/EC of the European Parliament and of the Council of 6 July 1998 on the legal protection of biotechnological inventions. Available: http://europa.eu.int/smartapi/cgi/sga_doc?smartapi!celexapi!prod!CELEXnumdoc&lg=EN&numdoc=31998L0044&model=guichett.
68. *Japan Patent Law*, as amended (2003).
69. Canadian Intellectual Property Office (CIPO) (2006). A Guide to Patents: Utility. Available: http://strategis.ic.gc.ca/sc_mrksv/cipo/patents/pat_gd_protect-e.html#sec2. Accessed 20 July 2006.
70. UK Patent Office. What is a Patent? (2006). Available: http://www.patent.gov.uk/patent/ definition.htm. Accessed 20 July 2006.
71. Swedish Patent and Registration Office (2006). Which Inventions are Patent-able? Available: http://www.prv.se/english/patents/which.html. Accessed 20 July 2006.
72. Japan Patent Office (2006). Examination Guidelines for Patent and Utility Model in Japan: Part II Requirements for Patentability. Available: http://www.jpo.go.jp/tetuzuki_e/index.htm. Accessed 20 July 2006.
73. US Patent Act (2004). 35 U.S.C.
74. *Patent Act*, R. S. C. 1985, C. p-4, as amended.
75. UK Patent Office (2005). Confidentiality and Confidentiality Agreements, DDU85/01-05. Available: http://www.patent.gov.uk/patent/info/cda.pdf. Accessed July 21 2006.
76. Elder G (1985). A practical guide for proving fraud on the patent and trademark office: J.P. Stevens & Co. v. Lex Tex Ltd. *Am. U. L. Rev.* 34:729–746.
77. *Control Data Canada Ltd. v. Senstar Corp.* (1989). 23 C.P.R. 3d.
78. Patent Appeal Board (1982). Application of Abitibi Co. 62 C.P.R. 2d.
79. UK Patent Office (2005). Examination Guidelines for Patent Applications relating to Biotechnological Inventions in the UK Patent Office.
80. Gold R, Durell K (2005). Innovating the skilled reader: Tailoring patents to new technologies. *Intellectual Prop. J.* 19:189–226.
81. *Beloit v. Valmet* (1986). 8 C.P.R. 3d.
82. Hitchman C, MacOdrum D (1990). Don't fence me in: Infringement in substance in patent actions. *C.I.P.R.* 7:167–225.
83. Canada, Patent Rules (SOR/96-423), R. 79(6).
84. Mossinghoff G (2005). Small Entities and the "First to Invent" Patent System: An Empirical Analysis. *Washington Legal Foundation*. Available: http://www.wlf.org/upload/MossinghoffWP.pdf. Accessed 20 July 2006.
85. Canadian Intellectual Property Office (2006). A Guide to Patents: Glossary. Available: http://strategis.ic.gc.ca/sc_mrksv/cipo/patents/pat_gd_gloss-e.html. Accessed 20 July 2006.
86. Cameron D, Davis M (2001). Patents on Bay Street? *The Globe and Mail*. 24 August.
87. Marshall A, et al. (2006). A license to print money? *Nat. Biotechnol.* 24:593.

88. African Regional Industrial Property Organization (ARIPO) (2006). Background on the ARIPO. Available: http://www.aripo.wipo.net/background.html. Accessed 20 July 2006.
89. World Intellectual Property Organization (2006). Protecting your Inventions Abroad: Frequently Asked Questions about the Patent Cooperation Treaty. Available: http://www.wipo.int/pct/en/basic_facts/faqs_about_the_pct. Accessed 20 July 2006.
90. Anderegg M, et al. (2006). Trendspotting: A shift in intellectual property focus. *Nat. Biotechnol.* 24:609–611.
91. *Minerals Separation North American Corp. v. Noranda Mines, Ltd.* (1947). Ex. C.R. 306.
92. *Festo Corp. v. Shoketsu Kinzoku Kogyo Kabushiki Co.* (2002). 304 Fed. Cir. 3d 1289.
93. *Patent Act*, 1977 (UK), c. 37, as amended.
94. *Catnic Components Ltd. v. Hill & Smith Ltd.* (1982). R.P.C. 183 H.L.
95. Takenaka T (2000). Patent infringement damages in Japan and the United States: Will increased patent infringement damage awards revive the Japanese economy? *Re-Engin. Patent Law.* 2:301–309.
96. *THK Co Ltd v. Tsubakimoto Precision Products Co Ltd*. (1994). Tokyo High court.
97. *Genentech Inc. v. Sumitomo Pharmaceuticals Company* Ltd. (1996). Osaka High Court.
98. Ladas & Parry LLP (2002). Japan—Damage Awards in Patent Infringement Actions. August Newsletter.
99. *Free World Trust v. Électro Santé Inc.* (2000). 2 S.C.R. 1024.
100. *Whirlpool Corp. v. Camco Inc.* (2000). 2. S.C.R. 1067.
101. Kleiman M, Sontag L (2006). Patenting Pathways. *Nat. Biotechnol.* 24:616–617.
102. Center for Intellectual Property Policy and Health Law Institute (2004). The Research or Experimental Use Exemption: A Comparative Analysis. Report prepared for Health Canada. Available: http://www.cipp.mcgill.ca/data/publications/ Accessed 23 July 2006.
103. Gold E, Joly Y, Caulfield T (2005). Genetic research tools, the research exception and open science. *GenEdit.* 3:1–11.
104. *Embrex v. Service Engineering* (2000). 216 F.3d 1343.
105. *Madey v. Duke University* (2002). No. 01-1567 CAFC.
106. Drug Price Competition and Patent Term Restoration Act (1984). 35 U.S.C. §271(e).
107. Convention for the European Patent for the Common Market *Community Patent Convention*. 1975, 1976 O.J. L 17.
108. Grubert A (2006). Understanding European Patent Opposition In Light Of U.S. Patent Practice. *Intellect. Prop. Report*. 4:39. Available: http://www.imakenews.com/bakerbotts/e_article000274940.cfm. Accessed 21 July 2006.
109. Cochran D (2006). United States: Toward a Post-Grant Opposition Proceeding in the United States. *Jones Day*. 21 April 2006. Available: http://www.mondaq.com/article.asp?articleid=39280. Accessed 23 July 2006.
110. Kirk M (2004). Patent quality improvement: Post-grant opposition. *Statement before the Subcommittee on Courts, the Internet, and Intellectual Property Committee on the Judiciary U.S. House of Representatives.* Available: http://www.aipla.org/Content/ContentGroups/Legislative_Action/108th_Congress1/Testimony2/Testimony_on_Patent_Quality_Improvement_Post-Grant_Opposition.htm. Accessed 24 July 2006.

111. United States Congress (2005). Bill 2795 H.R.
112. Trilateral Statistical Report (2004). *Patent Activity at Trilateral Offices*. Trilateral Cooperation, Munich, Germany. Available: <http://www.trilateral.net/tsr/tsr_2004/tsr2004.pdf. Accessed 24 July 2006.
113. U.S. Food and Drug Administration, Centre for Biologics Evaluation and Research (2006). About Us. Available: http://www.fda.gov/cber/about.htm. Accessed 20 July 2006.
114. Dudzinski D (2005). Reflections on historical, scientific, and legal issues relevant to designing approval pathways for generic versions of recombinant protein-based therapeutics and monoclonal antibodies. *Food Drug Law J.* 60:143–259.
115. Polastro E (2001). Biogenerics: Myth or reality? *Innov. Pharm. Technol.* Dec:63–65.
116. Class S. Biogenerics: Waiting for the Green Light. *IMS Company Profiles*. Available: http://www.imshealth.com/web/content/0,3148,64576068_63872702_70261006_71226827,00.html. Accessed 22 July 2006.
117. Ontario Ministry of Health (2002). Genetics, Testing and Gene Patenting: Charting New Territory in Healthcare. Government of Ontario, Toronto, Canada.
118. National Academy of Sciences (2004). *A Patent System for the 21st Century*, The National Academies Press, Washington, D.C.
119. Williams-Jones B (2002). History of a gene patent: Tracing the development and application of commercial BRCA testing. *Health Law J.* 10:123–146.
120. National Academy of Sciences (2005). *Reaping the benefits of Genomic and Proteomic Research: Intellectual Property Rights, Innovation, and Public Health*; National Academies Press, Washington, D.C.
121. Jensen K, Murray F (2005). Enhanced: Intellectual property landscape of the human genome. *Science*. 310:239–240.
122. Heller M, Eisenberg R (1998). Can patents deter innovation? The anticommons in biomedical research. *Science*. 280:698–701.
123. Caulfield T (2005). Policy conflict: Gene patents and health care in Canada. *Community Genet.* 8:223–227.
124. Nelson R R (2004). The market economy and the scientific commons. *Res. Policy.* 33:455–471.
125. David P A (2004). Can 'Open Science' be protected form the evolving regime of IPR protection? *J. Theoret. Institut. Econ.* 160:1–26.
126. Walsh J P, Cho C, Cohen W M (2005). View from the bench: Patents and material transfers. *Science*. 309:2002–2003.
127. US Department of Justice and the Federal Trade Commission (1995). Antitrust Guidelines for the Licensing of Intellectual Property. 4 Trade Reg. Rep. CCH.
128. Goldstein J, et al. (2005). Patent pools as a solution to the licensing problem of diagnostic genetics. *Drug Dis. World*. Spring, 86–90.
129. Clarke J, et al. (2000). Patent Pools: A Solution to the Problem of Access in Biotechnology Patents?" United States Patent and Trademark Office. Available: http://www.uspto.gov/web/offices/pac/dapp/opla/patentpool.pdf.
130. Resnik D B (2002). The commercialization of human stem cells: Ethical and policy issues. *Health Care Anal.* 10:127–136.
131. Verbeure B, Van Zimmeren E, Matthijs G, et al. (2006). Patent pools and diagnostic testing. *Trends Biotechnol.* 24:114–120.
132. Opderbeck D (2004). The penguin's genome, or coase and open source biotechnology. *Harv. J. Law & Tec.* 18:167–227.

133. Burke D (2001). Open Source Genomics. *B.U. J. Sci. & Tech. Law*. 8:255–256.
134. Feldman R (2004). The open source biotechnology movement: Is it patent misuse? *Minn. J.L. Sci. & Tech*. 6:117–167.
135. Maurer S, et al. (2004). Finding cures for tropical diseases: Is open source and answer? *Minn. J.L. Sci. & Tech*. 6:169–175.
136. World Trade Organization (2003). Decision Removes Final Patent Obstacle to Cheap Drug Imports. Press/350/Rev. Available: http://www.wto.org/english/news_e/pres03_e/pr350_e.htm. Accessed August 2005.
137. Canadian HIV/AIDS Legal Network (2004). Global Access to Medicines: Will Canada Meet the Challenge? *Standing Committee on Industry, Science and Technology regarding Bill C-9, An Act to amend the Patent Act and the Food and Drugs Act*. Available: http://www.aidslaw.ca/Maincontent/issues/cts/patent-amend/SCIST%20Submission_Feb2604.PDF. Accessed August 2005.
138. Canadian HIV/AIDS Legal Network (2003). Amendment to Canada's Patent Act to Authorize Export of Generic Pharmaceuticals—Update. Available: http://www.aidslaw.ca/Maincontent/issues/cts/patent-amend/PatentActAmendment_Update.pdf. Accessed August 2005.

# COMPARABILITY STUDIES FOR LATER-GENERATION PRODUCTS—PLANT-MADE PHARMACEUTICALS

PATRICK A. STEWART
*Arkansas State University, State University, Arkansas*

**Chapter Contents**

1 Introduction
2 Benefits Offered by Plant-Made Pharmaceuticals (PMPs)
3 Threats Posed by Plant-Made Pharmaceuticals (PMPs)
  3.1 Starlink Bt Corn
  3.2 Prodigene PMP Corn
  3.3 Syngenta Bt Corn
  3.4 Conclusions
4 Public Response to Plant-Made Pharmaceuticals (PMPs)
  4.1 Interested Public/Pressure Group Comments
  4.2 Public Opinion Concerning PMPs
  4.3 Conclusions
5 Regulation of Plant-Made Pharmaceuticals (PMPs) in the Fields
6 Conclusions
  References

## 1 INTRODUCTION

The future of pharmaceutical drugs in the United States is likely that of continuing economic growth, as there is increasing use of pharmaceutical drugs in the general population, with an estimated 43% of Americans using at least one pharmaceutical drug between 1999 and 2000 [1]. This fact, in conjunction with an aging population needing and demanding more and cheaper pharmaceutical drugs, with an average of 30 filled prescriptions for non-institutionalized Medicare enrollees aged 65 and

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

older [2], suggests a need for alternative approaches to traditional "vat-based" mammalian cell-based production. Although such production approaches as animal-based "pharming" are promising, among the most promising of approaches is that of Plant-Made Pharmaceuticals (PMPs). These PMPs use genetically engineered plants, a technology used for over 20 years, to express substances that may be used to treat and cure such diseases as cancer, Alzheimer's, diabetes, HIV, heart disease, and others. By using naturally occurring processes in the plant, along with the ability to scale up production quickly, PMPs are a promising technology. However, as with any new technology, the potential pitfalls are plentiful, chief among which are the perceptions of the public, which is largely unaware of this technology, and whose perceptions place it on a threshold of acceptance or rejection, making it a "boundary technology".

PMPs may be considered a boundary technology for a variety of reasons. First, it is a subset of a relatively new technology, that of genetically engineered plants. As a result of its newness, risks from genetic engineering of plants are not well characterized. Although recent events suggest the potential risks, and the current inadequacy of the regulatory system to either prevent or respond to crises (see Section 3), from genetically engineered plants have not been tangibly experienced by the average American. Instead of actual events, characterized risks tend to exist as either extrapolations from laboratory experiments or as hypothetical events. Although laboratory experiments do provide a good deal of insight into the mechanisms by which events may occur, they do not replicate the reality of the fields in which PMPs may be grown with the field's dynamic, nonlinear, complex systems affected by multiple interacting factors such as weather, soil, cropping systems, chemical usage, and so on. Hypotheticals about PMPs tend to be extreme in nature and politically charged, with Cassandras of doom preaching worst-case scenarios and Candides of optimism trumpeting the best possible future. Neither perspective, however, appreciates the probabilistic nature of risk. Here, risk is defined as hazard multiplied by exposure. And, in the words of Howard and Donnelly, "(I)n order for the risk to be zero, either the exposure or the hazard must be zero" [3], which, of course, is unlikely if not impossible, especially considering increased sensitivity of detection methods and understanding of the effect of different substances on different individuals.

A second reason to consider PMPs as a boundary technology is that they promise to radically change public perceptions by drawing attention to the changed character of the plants themselves. It is here that hypothetical scenarios concerning PMPs make their mark, leading to acceptance or rejection of this technology. Although genetically engineered plants are currently a pervasive part of the agricultural landscape and consumed extensively, with the staple crops of corn and soybean grown in genetically engineered variants in the United States and comprising 45% and 85% of total acreage in 2004 [4], respectively, there is little public awareness of their presence in the food supply [5–7], which is mainly a result of them being genetically engineered to express agronomic qualities, such as herbicide tolerance and pest resistance, which are typically not brought to the attention of the public. Although plants have long been used for medicinal purposes and food can have healing properties, genetically engineered plants, especially dealing with food crops, will likely draw public attention by underlining the transgenic nature of the products that span species and may even incorporate human DNA. Some precedent in public response to genetic engineering may be seen in "second-

generation" food products with enhanced product quality, which, although offering useful characteristics, have yet to be economically successful.

The final reason PMPs may be considered a boundary technology derives from the previous reason. PMPs may be recognized as both a food and a medicinal application of genetic engineering. Although it is highly unlikely that the same plant will be processed for both food and pharmaceutical purposes, it is quite likely the public will perceive them as a unified technology. In terms of public response to date, there have been quite different paths in the reception of genetically engineered pharmaceuticals and medical products and that of genetically engineered plants. There has been little controversy over genetically engineered pharmaceutical products, likely because people are willing to accept risk when it comes to recovering the health they have lost, whereas people are likely to be risk averse when it comes to changing their diet, especially when they perceive nothing wrong with it. By exhibiting both characteristics, PMPs are likely to arouse ambiguity in response, depending on the personal salience of particular pharmaceutical products being developed.

Considering these reasons, this chapter will explore PMPs and their potential with an eye toward public response and policy actions. In this chapter, we first consider potential benefits of new products moving toward the market, such as enhanced access and decreased price. We next look at the risks of this boundary technology, especially adventitious presence in which PMPs might accidentally enter the food supply, which is illustrated by recounting three recent events in which the food chain was either breached or threatened by unapproved food products. We then consider public opinion toward PMPs, both in response to these incidents, as seen in comments to Federal Register changes, and through a regional survey of public opinion. Finally, we consider the federal regulations currently in place and recent political developments before drawing conclusions.

## 2 BENEFITS OFFERED BY PLANT-MADE PHARMACEUTICALS (PMPs)

The benefits of PMPs have been well characterized and documented by proponents of this technology. Indeed, much of the research and development of this technology is driven by optimism that PMPs will have tremendous health benefits for an aging and growing population and the potential for economic gain, especially when compared with current production approaches. This outlook is bolstered by awareness of the sheer size of the pharmaceutical drug market. Of this market, an increasing proportion builds on genetic engineering.

According to Elbehri [8], three factors commend PMPs over current approaches to express pharmaceuticals, such as bacteria, yeast, and mammalian cells, and more recent technologies, such as insect cells and transgenic animals. These three factors are cost advantage, production scalability, and safety. The cost advantage of PMPs over the current use of mammalian cell cultures is great, with mammalian cell bioreactors ranging from $106–650 per gram of antibody, whereas PMPs would likely cost four to five times less for the same pharmaceutical.

The second factor is the capacity to more easily scale-up production than is currently the case with mammalian cell-based systems, which is based on the ability to increase production by growing more plants on added acreage in a short period of

time, avoiding the need to build more production facilities with buildings, vats, and other associated capital costs [8]. However, this benefit is offset by production being driven by growing seasons and the ability of plants to reproduce seeds enough to scale-up production [9]. In other words, pharmaceuticals will be produced on the basis of an agricultural cycle, with the inherent difficulties of predicting weather, pest outbreaks, and other factors that influence natural systems. Although storage may not prove problematic, especially if the pharmaceutical is expressed in the seed (which might not be the case), predicting market demand is yet another difficult factor. Another difficulty with scaling up lies with enforcing regulatory controls. Expanding production acreage implies increased potential for inadvertent movement of the PMPs' genes, whether by cross-breeding with other plants or their being blown into neighboring fields, bringing with it legal liabilities for environmental and health damage if they occur. Therefore, expanded acreage implies expanded implementation of safeguards and increased surveillance.

The third and final factor is the belief that PMPs are safer than using proteins from recombinant microorganisms or cells, which is because "PMPs do not carry potentially harmful human or animal viruses into the drug" [8], as plants do not serve as hosts for such human pathogens as HIV, prions, hepatitis, and so on [10]. Furthermore, plants are typically constructed to express a portion of a pathogen or toxin, reducing possible infection or innate toxicity [11]. This, however, does not alleviate concerns over potential allergenicity, which will be present in most pharmaceuticals.

In spite of the benefits offered, PMP development is still in its infancy, with retardation of its progress caused by reasons discussed in the next section. However, analysis of U.S. Department of Agriculture (USDA) data concerning release permits for "pharmaceuticals, industrial, and value-added proteins" over the last two years, during which time stricter regulatory standards were enforced, suggests a still-fertile field of research. Overall, 35 permits were applied for during 2004 and 2005, with permit applications increasing from 15 in 2004 to 20 in 2005, of which nearly 70% were granted and 60% were released into the fields.

Institutions doing research in PMPs tend to be dominated by small research companies (see Figure 1). Although research universities, mainly Iowa State

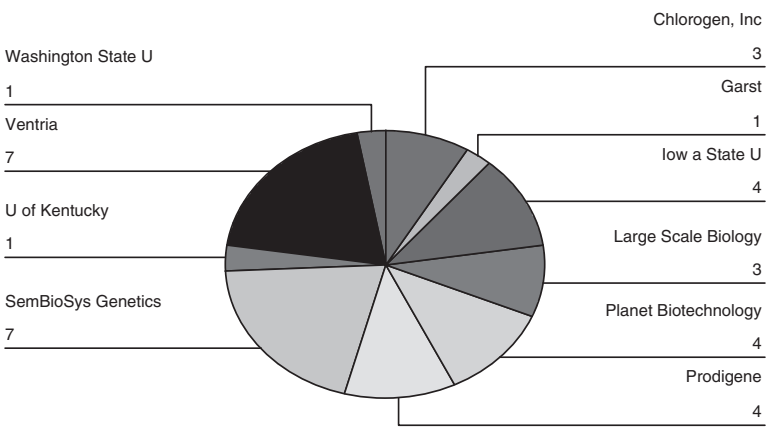


**Figure 1.** Institution.

University (four applications) and the University of Kentucky and Washington State University (one application apiece), are involved in research, companies such as SemBioSys Genetics and Ventria (seven applications apiece), Planet Biotechnology and Prodigene (four applications), Large Scale Biology and Chlorogen, Inc. (three applications), and Garst (one application) are the mainstays of research currently being sent to the fields.

The five plants serving as hosts for the experimental activities reflect research realities (see Figure 2). Tobacco is the preferred host, having nearly one third of the permit applications, reflecting its long-time use as a research plant in genetic engineering. Although corn has long history of experimentation and being used extensively in the field, and, as a result, is the second most popular plant for experimentation, the change in regulations in response to critiques of the regulatory system, and fears of it entering the food supply, has led to its reduced relevance in 2005, where only two university-based field trials were proposed, compared with its more extensive use (seven applications) in 2004. The remaining plants, safflower (seven applications), rice (six applications), and barley (two applications), might prove to be the preferred future host plants for PMPs and industrial products.

Examples of products being developed include Ventria BioScience's work on human recombinant lactoferrin and lysozyme, which may be used as an oral rehydration solution [12]; SemBioSys has been developing PMP-based human insulin (to meet the needs of an increasing population of diabetics) and human APO IP, which reduces and stabilizes arterial plaque associated with "acute coronary syndrome" such as heart attacks and angina, as well as strokes [13]; Planet Biotechnology's treatments of tooth decay by bacterial infection, hair loss through specific forms of chemotherapy, and treatment for the common cold by blocking infection by rhinovirus are likewise promising [14]. Prodigene has brought to market recombinant trypsin, normally produced by the pancreas, which may be used in the production of insulin, human and veterinary vaccines, cell cultures, and wound care [15]. It has also developed a vaccine targeting swine transmissible gastroenteritis [16]. Finally, Chlorogen is developing a "chloroplast transformation technology" that would introduce genes into plants, allowing large-scale production without potential genetic transfer through pollen [17]. These products and

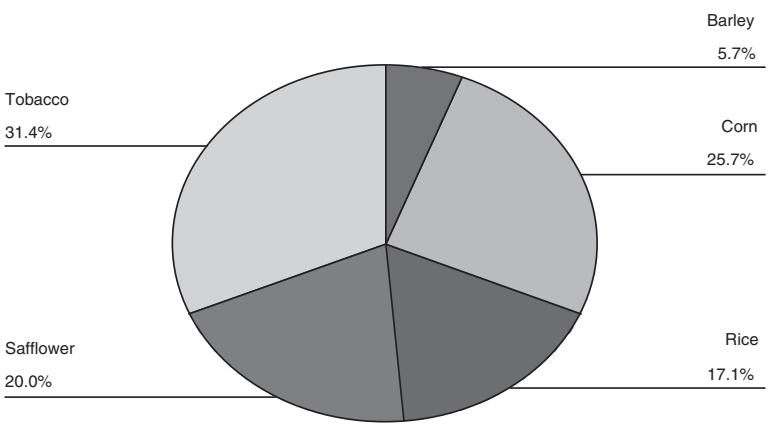


**Figure 2.** Host plant.

others promise increased product availability and a range of new products, with health benefits and economic growth implied.

Although the future of PMPs may, like any technology, be difficult to predict, the experimental activity concerning these plants appears to be relatively robust considering recent regulatory changes and increased public skepticism. Although the amount of permit applications are relatively small, especially compared with the sheer numbers of experimentation involving genetically engineered plants, PMPs and associated technologies appear to be growing in importance and in the fields. And given the benefits they offer in the production of new pharmaceutical drugs, this growth can be expected.

## 3 THREATS POSED BY PLANT-MADE PHARMACEUTICALS (PMPs)

In spite of the benefits specifically offered by PMPs, and by genetically engineered plants generally, apprehension exists over their release into the environment. According to the National Research Council report "Bioconfinement of Genetically Engineered Organisms" [18], three concerns exist over the unintentional movement of genes placed into target plants. The first is that the inserted genes will lead to the target plant or relatives cross-breeding and becoming weeds by obtaining benefits that enable it to overrun or disrupt ecosystems. The second concern is that wild relatives, especially those with economic, social, or moral values, will be at risk of extinction by either being outcompeted or through hybridization with genetically engineered plants. The third concern, that the genetic material will spread to other domesticated varieties [18], is perhaps one of the highest profile concerns for PMPs, especially when considering that many PMPs are currently being expressed in food plants. Although these concerns are currently being researched and will likely continue to pose a threat as genetically engineered plants dominate the commodity agriculture landscape, no evidence of these concerns coming to fruition have come to light to date.

Another major concern, if not the major concern, is that PMPs will somehow find their way into the food supply. As a result, it is feared the public will respond with decreased institutional trust and will make market decisions that negatively affect food producers. As a result, recent changes to the regulatory structure concerning agricultural biotechnology are occurring because of concerns "that the expansion in agricultural biotechnology increasingly will put pressure on seed production and commodity handling systems" [19] to segregate and control its products. Specifically, concerns exist that the current regulatory scheme is not able to adequately address the amount and variety of new genetically engineered plants, with PMPs chief among them. Doubts over the agricultural biotechnology regulatory system have been raised by three recent events, calling into question the scientific basis for regulation, the effectiveness of regulatory enforcement, and the integrity of the food system.

### 3.1 StarLink Bt Corn

The first major event to focus national attention on the potential violability of the U.S. food chain occurred when StarLink corn made its way into the human food

supply in 1999 and 2000. StarLink corn used the Cry9c variant of *Bacillus thuringiensis* (Bt), a protein that kills targeted insects by eating through their guts, leading to sepsis and the inability to digest food. StarLink corn was not deemed fit for either human consumption or export because of presumed possible human allergic reactions. As a result, it was to be used only for animal feed or nonfood industrial purposes.

In September and October of 2000, attention was drawn to the presence of StarLink corn in the food chain when the consumer group Genetically Engineered Foods Alert performed tests on taco shells and other corn-based products being sold in grocery stores and in fast food restaurants [20]. Estimates by the USDA's Economic Research Service suggest 123.8 million bushels were co-mingled in 2000, whereas Aventis', the company responsible for StarLink, estimates that over 430 million bushels were co-mingled in 1999 and 2000 [21] with an estimated 340,908 acres in 28 states planted with StarLink Bt corn [20]. Overall, a total of nearly 300 products were recalled, including 70 types of corn chips, 80 types of taco shells, and almost 100 restaurant food products.

In an attempt to minimize economic loss and public concern, Aventis and USDA bought back existing grain supplies and recalled food containing StarLink corn. However, significant damage was done. In addition to illustrating the failure of the U.S. regulatory system concerning genetically engineered plants, there was disruption of the downstream food manufacturing marketplace with corn exports being affected. Indeed, within a single year, of 110,000 grain tests by federal inspectors, StarLink corn showed up in one-tenth [22]. Although the complexity of market forces allows only estimation, a significant decrease in trade with Japan and South Korea occurred immediately after StarLink corn was found [21]. Since then, the European Union has required labeling and tracing requirements for genetically engineered goods that are seen as greatly hindering U.S. farm product competitiveness, especially with respect to corn and soybean products. Overall, estimates suggest the StarLink event and the recall in its aftermath have cost the food industry $1 billion [23].

### 3.2 Prodigene PMP Corn

The second event, that of Prodigene's PMP corn, is one immediately relevant for PMP production. In September and October of 2002, in Iowa and Nebraska respectively, APHIS found "volunteer" corn plants genetically engineered to produce a pharmaceutical to prevent "traveler's diarrhea" growing in soybean fields in violation of permit conditions. Specifically, Prodigene did not abide by the conditions of their field release permit to eradicate all traces of the experimental crop from the fields, as small quantities of this corn ended up in soybean that was to be processed and sold for human consumption.

Although none of this corn made it into the food supply, this near miss drew attention to the regulatory system's lack of readiness in dealing with the next generation of genetically engineered crops, those producing pharmaceutical and industrial products. Prodigene had to pay a civil penalty of $250,000; destroy 500,000 bushels, or $2.7 million dollars worth, of soybean in Nebraska; and incinerate 155 acres of corn in Iowa because of concern that cross-pollination occurred, as well

as posting a $1 million bond and accede to higher compliance standards for future field tests [24]. Perhaps more important in terms of long-term political implications, the Grocery Manufacturers of America (GMA) and other food processing interest groups expressed concern over PMP field test regulations, with John R. Cady, CEO of the National Food Processors Association commenting "(it is) nothing short of alarming to know that at the earliest stages of development of crops for plant-made pharmaceuticals, the most basic preventative measures were not faithfully observed. This apparent violation of rules . . . very nearly placed the integrity of the food supply in jeopardy" [25].

### 3.3 Syngenta Bt Corn

The third and most recent event came to light in March 2005 when the seed company Syngenta revealed that it had accidentally sold approximately 146,000 tons of experimental Bt-10 corn that had been approved for experimental use only by USDA but was grown on 37,000 acres over a four-year period (2001–2004). The antibiotic marker gene is often used in the experimental phase of many genetically engineered plant studies to test for the presence or absence of altered genetics.

Although the fact that the Syngenta Bt-10 corn inadvertently entered the food chain raises questions about the effectiveness of the USDA regulatory system, especially after regulatory changes were put in place in response to the StarLink and Prodigene situations, the antibiotic marker gene entering the food supply has raised concern because of perceived problems with increased antibiotic resistance with human infections. Although many other factors contribute to antibiotic resistance, including their extensive use in farm animals, concern over the use of antibiotics in genetically engineered crops date back to the earliest days of genetically engineering when, in 1990, Calgene asked for an opinion concerning their use from FDA, and it was found that there was reasonable certainty no harm would result from its use. However, since that time, concerns are still being raised concerning their near ubiquitous use, whether in genetically engineered plants or in farm animals [26].

Although Syngenta officials argue that the release was relatively small, affecting 0.01% of all corn grown from 2001 to 2004 [23], and has not appeared to affect U.S. consumers, response from abroad suggests that U.S. producers will be affected in the export marketplace. Specifically, on April 15, 2006, the European Union placed a temporary ban on U.S. corn imports unless they are proven to be Syngenta Bt-10-free; on May 25, 2006, they enforced this ban by impounding a U.S. shipment of corn gluten upon its arrival in Ireland when it was found to be tainted with Bt-10. Likewise, in South Korea, imports of U.S. corn are being investigated with certification of future shipments as Bt-10-free required for entry, while most recently (June 1, 2006), a shipment of corn to Japan found to be Bt-10-contaminated was found and held [27].

As a result, Syngenta has been fined $375,000 and ordered to sponsor a compliance training conference by USDA [28–30]. However, the regulatory response appears to be subdued at best, especially in light of previous failures by biotechnology companies to implement federal regulations.

### 3.4 Conclusions

According to the 1984 Coordinated Framework for the Regulation of Biotechnology, USDA, EPA, and FDA are responsible for regulating agricultural biotechnology. Although this framework has proved to be serviceable for most of its existence, recent events suggest it is not meeting the challenges posed by the increased volume and variety of genetically engineered crops. The instances discussed above illustrate some of the problems facing U.S. regulators, especially the reduced trust in institutions. However, what is most important for producers is that an inefficient and ineffective regulatory system might push consumers elsewhere.

Although critics argue that the above cases show the system is working by identifying flaws in the regulatory system and changing practices of both government agencies and industry, in the end, consumer trust is what matters most. Although lost consumer trust in domestic markets is a threat, with StarLink-tainted products being recalled from grocer's shelves in 2000, long-lasting effects in the U.S. market have not materialized, especially when the recent Syngenta Bt-10 corn case is considered. International trade disruption, however, has occurred in both StarLink and Syngenta cases and has been costly in terms of trade with the European Union and with Asian trade partners South Korea and Japan. Perhaps most important, trust in the U.S. regulatory system by our trade partners is waning, with a recent editorial title concerning the Syngenta case in the respected science journal *Nature* underscoring diminishing respect: "Don't rely on Uncle Sam" [30].

## 4 PUBLIC RESPONSE TO PLANT-MADE PHARMACEUTICALS (PMPs)

Public response to PMPs suggests they are of two minds concerning PMPs. On the one hand, PMPs promise to offer tangible benefits that directly enhance the quality of life by producing less expensive pharmaceutical drugs that may treat a broader range of ailments. This perspective is bolstered in a recent PEW report that showed a strong relationship between perception of PMPs as being a good use of genetic engineering and the likelihood of it having a positive impact on the respondent's family [6].

On the other hand, the risks of PMPs entering the food supply or remaining in the environment as weeds gives pause to critics of the technology. The StarLink and Syngenta Bt-10 cases stand out as examples of the regulatory and food systems' failure to prevent the unwanted presence of genetically engineered crops in the food supply. Although the Prodigene case was a near miss, it also focused attention on the flaws in the regulatory system and may have had a chilling effect on the development of both PMPs and PMIPs. Although these three cases reflect "growing pains" in regulating the technology as more plant products with a greater variety of uses enter the fields, the question remains as to whether this problem will be endemic or a "mere blip" that has been dealt with and will not recur.

### 4.1 Interested Public/Pressure Group Comments

The amount of comments in response to the 2003 Federal Register notices dealing with changes to the regulation of PMPs and PMIPs, which in turn are arguably in

response to the Starlink and Prodigene incidents discussed in the previous section, reflect the new-found importance of the field release of genetically engineered plants to the public. The March 10, 2003 Federal Register notice concerning PMP and PMIP field testing requirements attracted at least 847 comments (of which 77 were late). In comparison, changes to the APHIS regulations in 1993 garnered 84 comments, whereas even more wide-ranging regulatory changes in 1997 attracted only 50 comments [31].

Although the numbers of comments received in response to PMP regulatory changes are exponentially higher than those received in response to prior Federal Register notices, multiple factors are responsible. E-government with electronic dockets easing comments on Federal rule-making has expanded public participation, especially with salient issues. For example, proposed organic standards posted in the Federal Register received an unprecedented 275,000 comments. Here, although the usual policy actors interested in genetically engineered plants are still involved, a high percentage of comments were sent by individuals and organizations not normally associated with this debate [32].

As expected, support for the new technology came from biotechnology and biotechnology-related companies, agricultural organizations, state agricultural departments, universities, and agricultural sector pressure groups. Critiques were also received from those who appeared to have ties with the organic movement or environmental groups, as evidenced by the roughly 600 comments e-mailed as cut-and-paste forwards.

Unexpectedly, especially in comparison with previous regulatory changes, concerns were raised by other powerful groups. Namely, the GMA and affiliated groups expressed concern over uncontained field release of PMPs and PMIPs, especially in food and feed plants. Interestingly enough, concern by consumer groups and traditional biotechnology opponents was tempered, likely mitigated by the potential for medical benefits from this new technology.

Comments by the interested public in response to the changes to PMP regulations give insight into concerns the general public may have once their awareness of these new technologies are raised, especially if PMPs and PMIPs enter into widespread production. A few themes stand out when comments of those critiquing the proposed USDA regulations are analyzed. The first theme concerns the science used and its effectiveness in regulating the threats posed by PMPs and PMIPs to the food supply and the environment. The next two themes expand on institutional distrust suggesting that governmental institutions, namely USDA, are either negligent, showing disregard for American farmers and the public by allowing these "dangerous" plants to be grown, or have been captured by corporate interests and are doing their bidding despite the threat posed. The risk perceptions of those commenting on the regulations suggest this technology is unknown, unnatural, uncontrolled, and likely to end in disasters disproportionately affecting future generations [32]. As can be expected by the foregoing responses, PMPs and PMIPs engender a good deal of fear and anger, emotional responses we deal with more systematically in the next section.

### 4.2 Public Opinion Concerning PMPs

Although the general public's opinion on PMPs is best analyzed through surveys, little research on their opinion has been carried out in the United States. In

Canada, Einsiedel and Medlock [33] carried out a series of focus groups analyzing attitudes toward PMPs and industrial products after providing information on plant molecular farming generally and five specific examples of this technology. Findings by them suggest greater support for novel medical products benefiting human health or the environment. Risk perceptions were driven by concerns over contamination of the food supply or the environment with higher risk seen as deriving from these products being grown outdoors in food plants that are able to go to seed or that flower, thus dispersing their genetic material to food plants that might be consumed by humans. When risk-benefit analyses were made, Einsiedel and Medlock found "acceptability is predicated on the idea that there are stringent approval processes and long-term measures in place to ensure safety" [33].

One exception to the lack of systematic public opinion research on PMPs is a study carried out by Stewart and McLean [7] using a telephone survey examining consumers living in Arkansas, Louisiana, Texas, Oklahoma, and New Mexico. This study was conducted between May 9 and June 10, 2004 by the Arkansas State University Center for Social Research using computer-assisted telephone interviewing (CATI) technology and had a very good response rate (a total of 680 interviews completed for a response rate of 61% and a margin of error of +/–3.5%). Exact wording of the questions are contained in Appendix A.

Five questions were asked of participants and ascertained personal benefits, likelihood of PMPs entering the U.S. food supply, the likelihood of respondents eating food from PMPs, how worried the respondents would be if they ate food with PMPs in it, and how angry they would be if they ate it without their knowledge. PMPs were characterized as "plants genetically modified to produce pharmaceutical drugs. These are plants that are modified to produce compounds used in manufacturing vaccines for diarrhea, antibodies to fight cancer, and drugs to treat such illnesses as cystic fibrosis."

Findings suggest the respondents have mixed feelings about PMPs. Generally speaking, PMPs are seen as presenting personal benefits to respondents, with two-thirds seeing either "a great deal" or "some benefit" (Figure 3). Of these, a plurality (36.5%) see a great deal of benefit for themselves, whereas only 12% see themselves as receiving no benefit at all from PMP-based products. Finally, only 12% had no opinion about PMPs.

Although respondents see benefits from PMPs, a majority see adventitious presence as likely to occur, with nearly 60% perceiving PMPs as accidentally entering the U.S. food supply. Only one fourth of respondents see this event as either "somewhat" or "very unlikely" to occur (Figure 4). Whether this pattern of response is because of the Starlink or Prodigene controversies discussed earlier in this chapter, because of a lack of faith in governmental and food production institutions to prevent this event from occurring, or the result of some other unspecified factor is not known; it may be assumed, however, that a relatively high level of public skepticism exists.

Although respondents see themselves as less likely to eat food with PMP components than see the potential for adventitious presence, those seeing it as either "very" or "somewhat likely" make up a substantial proportion of respondents, with a combined 45% (Figure 5). At the same time, 38% perceive themselves as able to avoid such foods, responding to the question that they are either "somewhat" or "very unlikely" to eat food coming from PMP plants.

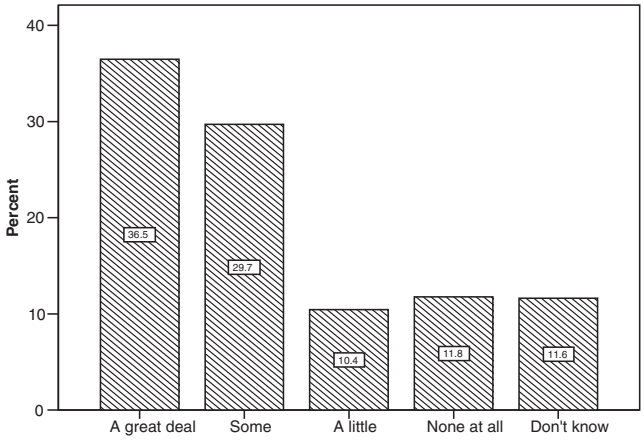


**Figure 3.** How much benefit would you get from PMP GM plants?

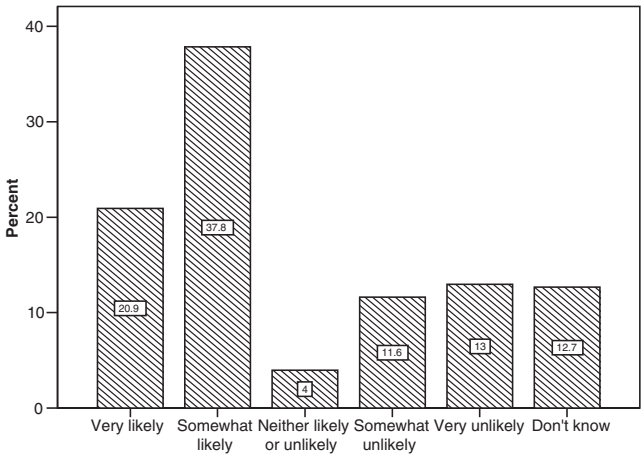


**Figure 4.** How likely is it that PMPs might accidentally enter the U.S. food supply?

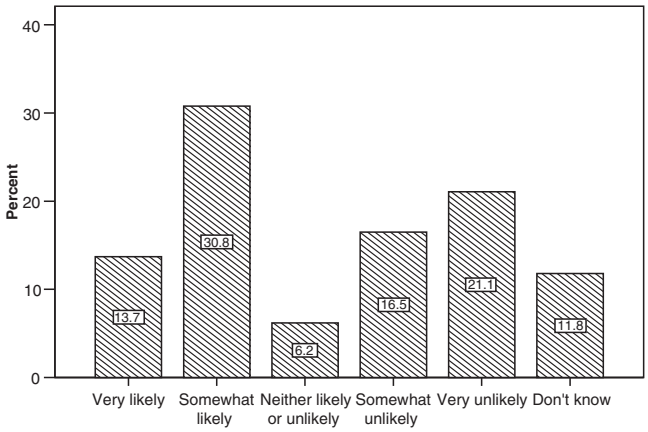


**Figure 5.** How likely is it that you would eat food coming from PMP GM plants?

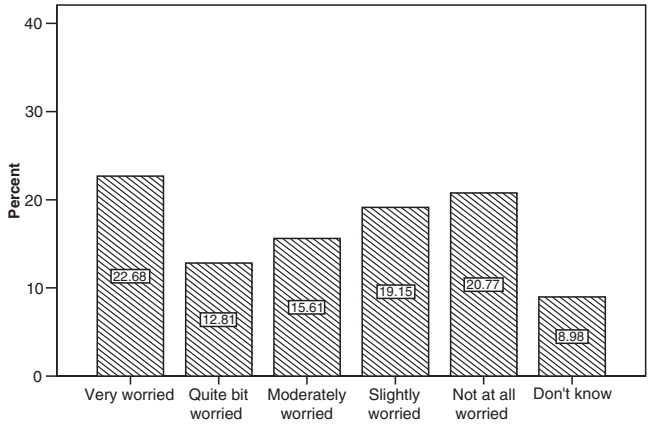


**Figure 6.** How worried would you be if you ate food from PMP GM plants?

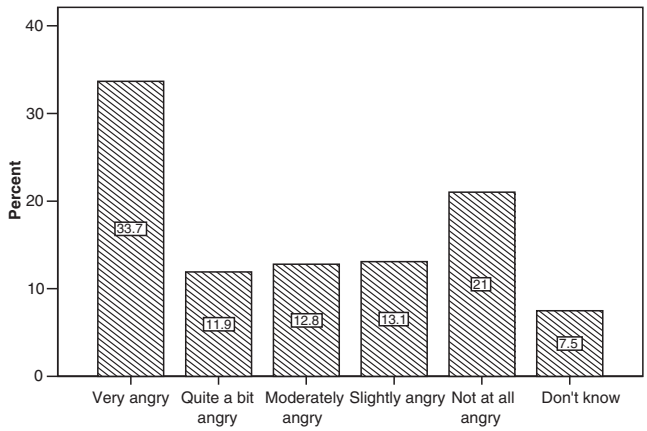


**Figure 7.** How angry would you be if you ate food from PMP GM plants?

Emotional response to eating food with PMPs in it suggests survey respondents are at least somewhat worried about the potential for it to occur (Figure 6). Overall, 23% are "very worried" and 13% are "quite a bit worried" about this event occurring. However, an increasing proportion of respondents show lesser levels of concern, with 16% "moderately worried," 19% "slightly worried," and 21% "not at all worried."

Although levels of concern are not as high as can be expected, great potential exists for upsetting the public. The pattern of response that can be seen in Figure 7 suggests polarized opinions. Specifically, one-third of respondents state they would be "very angry" if they ate food with PMPs in it without their knowledge, whereas only one-fifth of respondents would be "not at all angry."

## 4.3 Conclusions

Research available concerning public attitudes toward PMPs emphasizes it's "boundary technology" nature. Great promise and great risk are inherent in PMPs

because of the personal benefits of increased access to a wider array of pharmaceutical drugs balanced by their potential for accidental release into the food supply or the environment. These ambivalent feelings are likely to persist until either the benefits of PMPs are felt by their products entering the marketplace or an environmental or health catastrophe involving these products occurs. In either case, the absence of awareness of PMPs (and even genetically engineered plants) is a driving issue.

## 5 REGULATION OF PLANT-MADE PHARMACEUTICALS (PMPs) IN THE FIELDS

The regulation of PMPs will establish a different, potentially more circuitous, path to the marketplace because of its "boundary technology" nature. Although, as a pharmaceutical, it will be regulated in a traditional manner by FDA, as a plant, it must adhere to standards set forth by USDA. As a genetically engineered plant, it triggers a set of regulations specific to this technology as a whole, and as a PMP specifically, as discussed below.

On the whole, the regulation of agricultural biotechnology has attracted relatively little public involvement until recently, which is most likely because of its highly technical nature and the downstream nature of most of its products, which have predominantly been concerned with replacing or reducing production inputs such as herbicides and pesticides. More recently, the promise of new products and dearth of experienced difficulties in the field experiments have led to relaxed regulations. However, recent events detailed above have publicized and politicized the regulation of agricultural biotechnology with stricter regulations and greater scrutiny by an expanding range of parties with economic and political clout [34].

To address the decreasing trust in the regulatory structure, the Office of Science and Technology Policy (OSTP) published "Proposed Federal Actions To Update Field Test Requirements for Biotechnology Derived Plants and To Establish Early Food Safety Assessments for New Proteins Produced by Such Plants" in August 2002. Specifically, the notice was published to provide guidance to USDA, EPA, and FDA to update field testing requirements for food and feed crop plants and establish early food safety assessments for new plant proteins, most specifically PMPs and PMIPs, in line with the 1986 Coordinated Framework. Three principles in the document are relied on in updating the Coordinated Framework: (1) The level of field test confinement should be consistent with the level of environmental, human, and animal health risk associated with the introduced proteins and trait(s); (2) if a trait or protein presents an unacceptable or undetermined risk, field test confinement requirements would be rigorous enough to restrict outcrossing or comingling of seed and, further, the occurrence of these gene or gene products from these field tests would be prohibited in commercial seed, commodities, and processed food and feed; and (3) even if these traits or proteins do not present a health or environmental risk, field test requirements should still minimize the occurrence of outcrossing and comingling of seed, although low levels of genes and gene products could be found acceptable based on meeting applicable regulatory standards [19].

In light of concerns raised by increased experimentation with PMPs and PMIPs, and addressed by OSTP in their notice [19], USDA's Animal and Plant Health Inspection Service (APHIS) changed rules concerning their field testing of PMPs in March 2003 [35]. Although the resulting regulations are expected to be modified further over the coming years, they currently incorporate significant changes in how PMPs and PMIPs are regulated [36]. Specifically, for all plants genetically engineered to produce pharmaceutical or industrial compounds and field tested under permit, APHIS established seven conditions that can be grouped into three categories. The first considers field test siting, the second the dedication of equipment and facilities to their production, and the third considers procedural matters.

Field test siting regulations proposed by APHIS provide two conditions to be met, with special consideration for pharmaceutical corn. First, the perimeter fallow zone will be increased from 25 feet to 50 feet to prevent inadvertent comingling with plants to be used for food or feed. Second, production of food and feed plants at the field test site and perimeter fallow zone will be restricted for the following season to prevent inadvertent harvesting. Furthermore, specific permit conditions for pharmaceutical corn have been instituted, likely a result of corn being the organism of choice until the Prodigene event drew public concern and regulatory changes were put in place in its wake.

A second theme concerns the dedication of farm equipment and facilities to the production of such crops. First, APHIS requires planters and harvesters to be dedicated to the test site for the duration of the tests, and although tractors and tillage attachments do not have to be dedicated, they have to be cleaned in accordance with APHIS protocols. The equipment and regulated articles must be stored in dedicated facilities for the field test's duration.

The final three requirements from the proposed rules concern procedural aspects of dealing with field tests of PMPs and plants producing industrial compounds. First, APHIS requires submission of cleaning procedures to minimize risk of seed movement. Second, procedures for seed cleaning and drying are required to be submitted and approved to confine plant material and minimize risk of seed loss or spillage. Finally, permittees will be required to implement an APHIS-approved training program to successfully comply with the stated permit conditions [35].

A key factor in any regulatory arrangement is the ability to ensure that those regulated are complying with the requirements set forth. As a result of the potentially contentious nature of PMPs and plant-produced industrial compounds, APHIS plans to increase the number of field site inspections "to correspond with critical times relevant to the confinement measures" [35]. Therefore, in addition to maintaining records of activities related to meeting permitting conditions, and increasing the likelihood of audit-ing them to verify required permit conditions were met, APHIS might inspect permitted field tests up to five times during the growing season: once at pre-planting to evaluate the site location, once at the planting stage to verify site coordinates and adequate cleaning of planting equipment, at midseason to verify reproduction isolation protocols and distances, at harvest to verify cleaning of equipment and their appropriate storage, and again at post-harvest to verify cleanup of the field site. In addition, two post-harvest inspections may occur to verify that the regulated articles do not persist in the environment. Finally, APHIS may inspect more frequently if deemed necessary [35].

As a result of the potential for both PMPs and PMIPs entering into the food supply, and using the PMP regulatory changes as a starting point, APHIS took immediate action to remove the notification track option, requiring complete permit track review in their recent (August 6, 2003) interim rule and request for comments [36].

The rationale given in the interim rule and request for comments was that although 14 field releases (9 notifications, 5 permits) had been carried out to date, the type of genetic engineering being carried out was to enhance such nutritional components as oil content. However, recent genetic modifications have been for "non-food traits with which APHIS has little regulatory experience or scientific familiarity" [36]. As such, the definition of PMIPs has three criteria: (1) the plants produce compounds new to the plant; (2) this compound has not normally been used in food or feed; and (3) the compound is being expressed for non-food/feed purposes [36].

An administrative reorganization of how USDA-APHIS regulates biotechnology created the Biotechnology Regulatory Services (BRS) to address concerns raised by PMPs and PMIPs specifically and genetically engineered organisms generally. According to USDA-APHIS, "Given the growing scope and complexity of biotechnology, now more than ever, APHIS recognizes the need for more safeguards and greater transparency of the regulatory process to ensure that all those involved in the field testing of GE crops understand and adhere to the regulations set forth by BRS."

Changes instituted by BRS include new training for APHIS inspectors in auditing and inspections of field trials, the use of new technologies such as global positioning systems, and analysis of historical trends to inform monitoring and inspection. According to APHIS, there are nine overarching goals that the changes will serve with key components being: (1) enhanced and increased inspections in which risk-based criteria, along with other factors, will be used to assess field test sites with higher-risk sites being inspected at least once a year, with other sites being randomly selected for yearly inspections; (2) auditing and verification of records of businesses and organizations to verify accuracy and implementation; (3) remedial measures to protect "agriculture, the food supply, and the environment in the event of compliance infraction" with the establishment of a "first-responder" group to deal with serious infractions; (4) standardized infraction resolution in which criteria will be established to determine the extent of an infraction and the response, whether further investigation, the issuance of a guidance letter, the issuance of a written warning, or referral to APHIS' Investigative and Enforcement Services (IES) unit for further action; (5) documentation, in which a database will be set up to track field test inspections and resulting compliance infractions, and transparency to keep stakeholders and the public informed on the regulatory decision-making process; (6) continuous process improvements where, as the science of biotechnology advances, so too will regulations and permit conditions to allow safe field testing; (7) an emergency response protocol, being developed with input from EPA and FDA, in which a quick response plan will be put in place "to counteract potential impacts on agriculture, the food supply, and the environment"; (8) training for field test inspectors in their dealings with PMP and PMIP field test sites, as well as the latest in auditing; and (9) certification concerning compliance with the highest level of auditing standards [37].

Although the reorganization can be seen as streamlining and focusing enforcement efforts, the potential for unduly high levels of workload stresses placed on this 26-member unit can be foreseen. First, BRS draws on APHIS inspectors to inspect field tests; however, more than 2600 of these agriculture quarantine inspectors have been transferred to the Department of Homeland Security [36]. The current agreement between USDA-APHIS and DHS allows for continued access by APHIS and BRS, although it can be expected that problems might occur as a result of split responsibilities and duties.

Implementation considerations are still an issue. Although thousands of field tests have been carried out in a variety of sites through the USDA-APHIS field permitting, notification, and petitioning program, these tend to be short-term tests, typically lasting for one growing season and collecting limited data that considers the ecological effect. Little work has been done in the way of long-term effects on the ecology of the regions in which they are sited.

Also, although genetically engineered plants have been in the fields for at least a decade, in the case of herbicide-tolerant and pest-resistant plants, their management has left something to be desired with large numbers of farmers not following federal regulatory standards [38, 39]. Although USDA states that of the 7402 field tests carried out between 1990 and 2001 and regulated by APHIS, only 115 resulted in compliance infractions and, for the most part, these were relatively minor infractions that did not raise public concern [40], concerns still exist. However, regarding PMPs, because of the regulations concerning them and the potential for controversy, it is likely that the regulations will be more stringently followed.

## 6 CONCLUSIONS

The "boundary technology" nature of PMPs places it firmly into limbo in terms of social acceptability and, ultimately, economic viability. In terms of benefits, the opportunities afforded by this subset of genetic engineering of plants are in terms of increased diversity of products, reduced production costs, scalability on relatively short notice, and reduced potential for health risks of toxicity and pathogenicity because of less-easily crossed species barriers. The risks of unintentional movement of genes into other plants or the entry of PMP products into the food supply, however unrealized to date, loom large. These threats are especially salient in light of likely scaling up of plantings of PMPs and greater diversity of PMP products. And when current regulatory difficulties are taken into account, it is an especially daunting challenge.

Although the potential benefits of PMPs are large and the risks may be perceived as negligible, even if such products enter the food supply as feared, public opinion looms large in whether this boundary technology is accepted. Given that most Americans are not aware of their current consumption of genetically engineered foods and the extent of their market presence, PMPs entering the food supply would likely be their first "real" exposure to the risks of genetically engineered plants. The resulting response—apathy, hysteria, or somewhere in between—would have an extremely large impact on the future of PMPs and their growth, which is not to discount the effect of other factors that may hasten or retard the development of PMPs. Economic climate, government policies and regulations, international

trade, and a slew of other factors will impact the future of PMPs. What remains is the question of how far and how fast this technology will develop.

## REFERENCES


1. NCHS (National Center for Health Statistics) (2005). FASTATS—Drug Use Health Facts. Available: http://www.cdc.gov/nchs/drugs/htm. Accessed 7 November 2005.
2. Federal Interagency Forum on Aging Related Statistics (2005). Federal Forum Reports Americans Aging Well, But Gaps Remain. Press release. Available: http://www.agingstats.gov. Accessed 7 November 2005.
3. Howard J A, Donnelly K C (2004). Quantitative safety assessment model for transgenic protein products produced in agricultural crops. *J. Agricul. Environ. Ethics.* 17:545–558.
4. PEW (Pew Initiative on Food and Biotechnology) (2004). Factsheet: Genetically Modified Crops in the United States. Available: http://pewagbiotech.org/resources/factsheets/. Accessed 23 September 2005.
5. Hallman W K, Hebden W C, Cuite C L, et al. (2004). Americans and GM Food: Knowledge, Opinion and Interest in 2004. (Publication number RR-1104-007). Food Policy Institute, New Brunswick, NJ.
6. PEW (Pew Initiative on Food and Biotechnology) (2004). *Public Sentiment About Genetically Modified Food.* Mellman Group, Inc./Public Opinion Strategies for the Pew Initiative. Pew Initiative on Food and Biotechnology: Washington, D.C. Available: http://pewagbiotech.org/research/2004update/overview.pdf. Accessed 27 May 2005.
7. Stewart P A, McLean W P (2005). Public Opinion Toward the First, Second and Third Generations of Plant Biotechnology. *In Vitro Cell Developmen. Biol.—Plant* 41(6): 718–724.
8. Elbehri A (2005). Biopharming and the food system: Examining the potential benefits and risks. *AgBioForum.* 8(1):18–25.
9. Delaney D E (2002). Choice of crop species and development of transgenic product lines. In E E Hood, J A Howard (eds.), *Plants As Factories for Protein Production*: Kluwer Academic Publishers: Boston, MA. pp. 139–158.
10. Larrick J W, Yu L, Naftzger C, et al. (2002). Human pharmaceutical products in plants. In E E Hood, J A Howard (eds.), *Plants As Factories for Protein Production.* Kluwer Academic Publishers: Boston, MA. pp. 79–102.
11. Jilka J M, Streatfield S J (2002). Animal health. In E E Hood, J A Howard (eds.), *Plants As Factories for Protein Production*. Kluwer Academic Publishers: Boston, MA. pp. 103–118.
12. BIO PMP&AMP Update: Quarterly News (2005). Lactoferrin and Lysozyme may address health concerns in the developing world. *Biotechnology Industry Organization.* 1–2.
13. SemBioSys Website (2005). Pipeline: Pharmaceutical Products. Available: http://www.goodmedia.com/equicom/sembiosys/Main.aspx. Accessed 2 November 2005.
14. Planet Biotechnology—Products. Available: http://www.planetbiotechnology.com/products.html. Accessed 7 November 2005.
15. Prodigene. Prodigene Brings Plant-Produced Recombinant Trypsin to Market. Available: http://www.prodigene.com/news_releases/04-04-28_TrypZean(tm).html. Accessed 7 November 2005.

16. Prodigene. Prodigene Shows Lactogenic Immunity With an Oral Swine Vaccine. Available: http://www.prodigene.com/news_releases/04-06-23_TGEV.html. Accessed 7 November 2005.
17. Chlorogen News. Chlorogen, Sigma-Aldrich Sign Joint Development Agreement to Produce First Ever Products from Chloroplasts. Available: http://www.chlorogen.com/news.htm. Accessed 7 November 2005.
18. NRC (National Research Council) (2004). *Biological Confinement of Genetically Engineered Organisms*. The National Academies Press: Washington, D.C.
19. OSTP (Office of Science and Technology Policy) (2002). Proposed Federal Actions to Update Field Test Requirements for Biotechnology Derived Plants and To Establish Early Food Safety Assessments for New Proteins Produced by Such Plants; Notice." *67 Fed. Reg.* 149:50577–50580.
20. Harl N, Moline S (2003). The StarLink situation. Ames: Iowa State University. White Paper. Available: http://www.exnet.iastate.edu/Pages/grain/publications/starlink.html. Accessed 5 June 2005.
21. Lin W, Allen E (2001). Starlink: Impacts on the U.S. corn market and world trade. Feed Yearbook/April 2001. Economic Research Service, USA, pp. 40–48. Available: http://www.ers.usda.gov/Briefing/Biotechnology/starlinkarticle.pdf. Accessed 20 August 2004.
22. Ellstrand N C (2003). Going to "great lengths" to prevent the escape of genes that produce specialty chemicals. *Plant Physiol.* 132:1770–1774.
23. Macilwain C (2005). U.S. launches probe into sales of unapproved transgenic corn. *Nat.* 434:807.
24. USDA (U.S. Department of Agriculture) (2003). USDA Announces Actions Regarding Plant Protection Act Violations Involving ProdiGene, Inc. Press Release. December 6. Available: http://www.usda.gov/news/releases/2002/12/0498.htm. Accessed 7 November 2005.
25. Fox J L (2003). Puzzling industry response to prodigene fiasco. *Nat. Biotechnol.* 21(1):3–4.
26. Nestle M (2003). *Safe food: Bacteria, biotechnology, and bioterrorism*. University of California Press: Berkeley, CA.
27. Reuters News Service (2005). Chronology-recent events in Bt-10 corn controversy. Available: http://www.reuters.com/newsArticle.jhtml? type=topNews&storyID=8669951. Accessed 4 June 2005.
28. Herrera S (2005). Syngenta's gaffembarrasses industry and White House. *Nat. Biotechnol.* 23:514.
29. USDA (U.S. Department of Agriculture) (2005). What are the result of BRS' 2005 compliance investigations? Available: http://www.aphis.usda.gov/brs/compliance12.html. Accessed 5 December 2005.
30. Nature (2005). Don't rely on Uncle Sam. *Nat.* 434:807.
31. NRC (National Research Council) (2002). *Environmental Effects of Transgenic Plants: The Scope and Adequacy of Regulation*. National Academies Press: Washington, D.C.
32. Stewart P A, McLean W P (2005). Fear and hope over the third generation of agricultural biotechnology. *AgBioForum: J. Agrobiotechnol. Manag. Econom.* 7(3):133–141.
33. Einsiedel E F, Medlock J (2005). A public consultation on plant molecular farming. *AgBioForum.* 8(1):26–32.
34. Stewart P A, Knight A J (2005). Trends affecting the next generation of U.S. agricultural biotechnology: Politics, policy and plant made pharmaceuticals. *Technolog. Forecas. Soc. Change.* 72(5):521–534.

35. USDA (U.S. Department of Agriculture) (2003). USDA Announces Actions Regarding Plant Protection Act Violations Involving ProdiGene, Inc. Press Release. December 6. Available: http://www.usda.gov/news/releases/2002/12/0498.htm. Accessed 5 December 2005.
36. USDA (U.S. Department of Agriculture) (2003). Field testing of plants engineered to produce pharmaceutical and industrial compounds. *68 Fed. Reg.* 46:11337–11340.
37. USDA (U.S. Department of Agriculture) (2003). United States Department of Agriculture Pre-Briefing for reporters on USDA's Federal Register Notice on Field Testing of Pharmaceutical-Producing Plants. March 6. Available: www.usda.gov/news/releases/2003/03/084.htm. Accessed 4 June 2005.
38. Jaffe G (2003). Planting Trouble: Are Farmers Squandering Bt. Corn Technology? An Analysis of USDA Data Showing Significant Non-Compliance with EPA's Refuge Requirements. Center for Science in the Public Interest. Available: www.cspinet.org. Accessed 27 May 2005.
39. Stewart P A, Harding D, Day E (2002). Regulating the New Agricultural Biotechnology by Managing Innovation Diffusion. *Amer. Rev. Public Admin.* 32(1):78–99.
40. USDA (U.S. Department of Agriculture) (2003). Biotechnology Regulatory Services: Compliance and Enforcement. Agricultural Biotechnology Website. Available: www.aphis.usda.gov/brs/compliance. Accessed 27 May 2005.

## APPENDIX I: PLANT-MADE PHARMACEUTICALS QUESTIONS

I would next like to ask you a few questions about plants genetically modified to produce pharmaceutical drugs. These plants are modified to produce compounds used in manufacturing vaccines for diarrhea, antibodies to fight cancer, and drugs to treat such illnesses as cystic fibrosis.

How much benefit do you believe you would personally get from this type of plant? (1, a great deal; 2, some; 3, a little; 4, none at all; 8, DK; 9, refused)

How likely is it that these types of plants might accidentally enter the U.S. food supply? (1, very likely; 2, somewhat likely; 3, neither likely nor unlikely; 4, somewhat unlikely; 5, very unlikely; 8, DK; 9, refused)

How likely would you be to eat food coming from this type of plant? (1, very likely; 2, somewhat likely; 2, neither likely nor unlikely; 4, somewhat unlikely; 5, very unlikely; 8, DK; 9, refused)

How worried would you be if you ate food coming from this type of plant? (1, very worried; 2, quite a bit worried; 3, moderately worried; 4, slightly worried; 5, not at all worried; 8, DK; 9, refused)

How angry would you be if you ate food coming from this type of plant without your knowledge? (1, very angry; 2, quite a bit angry; 3, moderately angry; 4, slightly angry; 5, not at all angry; 8, DK; 9, refused)

# BIOSIMILARS

H. Schellekens, W. Jiskoot, and D.J.A. Crommelin
*Utrecht Institute for Pharmaceutical Sciences, Utrecht University, Utrecht, The Netherlands*

When a patent of a drug expires, competitors are allowed to introduce copies of the drug to the market. In the case of classic drugs, which are in general produced by chemical synthesis, a submission for marketing approval needs to contain data showing the generic to be chemically identical to the innovative drug, e.g., with spectroscopic and chromatographic separation techniques, and pharmaceutically acceptable [1]. In addition, in the case of oral products, the bioavailability of the generic product must be shown to be equivalent to that of the reference product as assessed by the results of pharmacokinetic studies demonstrating an equivalent rate and extent of absorption. Bioavailability studies are typically conducted in healthy volunteers in whom the geometric means for pharmacokinetic parameters such as area under the drug plasma concentration-time curve (AUC) and maximum observed plasma concentration ($C_{max}$) fall within an acceptable range (usually no more than ±20% with a 90% degree of confidence). These bioequivalence requirements make the assumption that two chemically identical products that demonstrate an identical pharmacokinetic profile will also have identical clinical effects. Thus, the randomized clinical trials conducted to demonstrate safety and efficacy that are required for the approval of innovator products are not necessary for the approval of conventional generics as the data generated with the original product can be extrapolated to the generic. Long-term experience with chemical drugs and the availability of sophisticated methods for analysis ensure that the generic products are safe and effective. Because the company marketing a generic product does not need to make the large investments necessary to develop and launch an innovative drug and can take advantage of the extensive experience and possible extensions of the indications gained during the period the drug was patent protected, the price of a generic is relatively low. The introduction of a generic also leads to

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

a price drop of the original product. Because reducing health costs is a political priority, most countries promote the introduction and use of generics. The competition with generics is also an important incentive for the innovator to improve the original poduct. So generics are an essential part of the innovation cycle of drugs.

For several reasons the classic generic paradigm can only partly be extrapolated to therapeutic proteins [2–5]. The major reasons are listed in Box 1. Protein pharmaceuticals are mostly large complex molecules showing heterogeneity. This heterogeneity can be due to natural processes in the host cells such as variations in glycosylation or protein clipping. Modifications of the product can also be introduced during production, purification, formulation, and storage. Besides these product-related impurities, host cells and (biological) materials used during production and purification may introduce process-related impurities.

Therapeutic proteins are manufactured by complicated processes, and all steps in production, purification, and formulation may influence the biological and clinicial properties of the final product. The different steps are, therefore, carefully monitored by analytical methods using in-house standards. These analytical methods have, in part, been developed (or at least refined) for each specific product. For many products and especially the first for which the patents will expire shortly, the production process has undergone continuous improvement based on manufacturing and clinical experience.

The features of a particular biopharmaceutical are the result of the basic characteristics of the molecule such as amino acid sequence and three-dimensional structure as well as the specific production, purification, formulation, and storage conditions (Box 2). To produce a biopharmaceutical of constant required quality, a company also needs the experience and the in-house standards to apply the methods used to analyze the structure of a given product. There are various guidelines of the European Medicine Evaluation Agency, the Food and Drug Administration, the Japanese Ministry of Health and Welfare, and the ICH, which require manufacturers to show that they control the production process and are capable of reproducibly manufacturing batches that not only meet product specifications, but also conform to the definition of the product as established through full characterization. Modifications of the established process are only accepted if the manufacturer can show that the product of the new process is comparable with the initially manufactured product. Comparability studies include revalidation of the process,

**BOX 1. CHARACTERISTICS OF THERAPEUTIC PROTEINS**

- Size
  - —100–500 times larger than classic drugs
  - —Cannot be completely characterized by physico-chemical methods
- Immunogenicity
- Structural heterogeneity
- Relatively high biological activity
- Relatively unstable

**BOX 2. FACTORS INFLUENCING THE ACTIVITY OF THERAPEUTIC PROTEINS**

- Gene and promotor
- Host cell
- Culture conditions
- Purification
- Formulation
- Storage and handling
- Unknown factors

**BOX 3. WHAT IS IN A NAME?**

- Biogenerics
- Second entry biologicals
- Subsequent entry biologicals
- Off-patent biotech products
- Multisource products
- Follow-up biologics
- Biosimilars
- Similar biological medicinal products

reevaluation of process and product-related impurities, recharacterization of the product in side-by-side analyses by all available state-of-the-art methods, and stability studies with the product from the new process in addition to release testing. When considered relevant, these comparability studies also may include clinical studies examining the pharmacokinetic, pharmacodynamic, and immunogenic properties of the new product and some efficacy and safety studies as well.

Little, if any, of the expertise, analytical methods and in-house standards, specifics of the production process, historical process, and validation data or full characterization data required for comparability assessment of therapeutic proteins are available in the public domain. As a rule, they are proprietary knowledge. It is inconceivable, in most cases, that another manufacturer, on the basis of the patent or published data, is able to manufacture a protein pharmaceutical that can be assumed similar enough to the original innovative product that only a limited documentation of physico-chemical characteristics would be sufficient to show equivalence. In most cases only limited data are available in pharmacopoeial monographs and scientific papers. Moreover, even the most sophisticated analytical tools are not sensitive enough to fully predict the biological and clinical characteristics of the product.

Because the generic approach is not applicable to protein drugs, the term *biogenerics* is considered misleading. Other terms have been used over the years, which are listed in Box 3. Similar biological medicinal products is the official

**TABLE 1. Expiration of Biologics in the United States (US) and European Union (EU)**

| Pioneer Company | Product | Indication(s) | US Patent/ Market Exclusivity Expires | EU Patent/ Market Exclusivity Expires |
|---|---|---|---|---|
| Genentech | Nutropin™ (somatropin) | Growth disorders | Expired | Expired |
| Abbott | Abbokinase™ (eudurase urokinase) | Ischemic events | Expired | Expired |
| Eli Lilly | Humulin™ (recombinant insulin) | Diabetes | Expired | Expired |
| Genzyme | Ceredase™ (alglucerase); Cerezyme™ (imiglucerase) | Gaucher disease | Expired | Expired |
| AstraZeneca | Streptase™ (streptokinase) | Ischemic events | Expired | Expired |
| Biogen/Roche | Intron A™ (IFN-alfa-2b) | Hepatitis B and C | 2002 | 2003 (France) 2007 (Italy) |
| Serono | Serostim™ (somatropin) | AIDS wasting | 2003 | NA |
| Eli Lilly | Humatrope™ (somatropin) | Growth disorders | ODE*** 2003 | NA |
| Amgen | Epogen™, Procrit™, Eprex™ (erythropoietin) | Anemia | 2013 | 2004 |
| Roche | NeoRecormon™ (erythropoietin) | Anemia | NA | 2005 |
| Genentech | TNKase™ (tenecteplase TNK-tPA) | Acute myocardial infarction | 2005 | 2005 |
| InterMune | Actimmune™ (IFN-gamma-Ib) | Chronic granulomatous disease (CGD), malignant osteopetrosis | 2005, 2006, 2012 | 2002, 2004 |
| Genentech | Activase™, Alteplase™ (tPA) | Acute myocardial infarction | 2005, 2010 | 2005 |
| Chiron | Proleukin™ (IL-2) | HIV | 2006, 2012 | 2005 |
| Amgen | Neupogen™ (filgrastim G-CSF) | Anemia, leukemia, neutropenia | 2015 | 2006 |

terminology in the European Union (EU), but biosimilars has become the preferred terminology both in scientific and regulatory discussions and publications and will also be used in this chapter.

Table 1 shows the expiration dates of the patents of several recombinant DNA-derived therapeutic proteins. The patents of products as interferon alpha and epoetin alpha have already expired, and others will follow shortly. Table 1 also shows the patent and data protection to last longer in the United States than in Europe. This partly explains why the regulations regarding the conditions for marketing authorization in the EU are more developed than in the United States.

In the EU legislation, which became effective in November 2005, biosimilars have been set apart from the generic drugs. Clinical data have legally become mandatory for a marketing authorization submission for biosimilars to the EMEA, which like innovative biotechnology products will need to be evaluated through a centralized procedure and not by national regulatory approval. However, the European regulators do not require a full dossier; the idea of extrapolation of data as with classic generics is maintained, as it is considered unethical to ask for a complete set of clinical data. The CHMP, the scientific committee of the EMEA, has issued several guidelines concerning the data required for marketing authorization. The main issues of these guidelines are listed in Box 4 [6–8].

Only biosimilars of therapeutic proteins marketed within the EU are allowed. A biosimilar marketing authorization submission should contain the same

**BOX 4. MAIN ELEMENTS CHMP GUIDELINES CONCERNING BIOSIMILARS**

- The concept of similar biological product is applicable to any biological medicinal product, but it is more likely applied to highly purified products, which can be thoroughly characterized
- In order to support pharmacovigilance monitoring, the specific product given to the patient should be clearly identified
- The active substance of the biosimilar product must be similar in molecular and biological terms to the active substance of the reference medicinal product e. IFN alpha 2a is not similar to IFN alpha 2b
- The same reference product throughout the comparability program
- The pharmaceutical form, dose, and route of administration of the biosimilar and the reference product should be the same
- If the reference product has more than one indication, the safety and efficacy for all indications have to be justified or demonstrated for each indication separately
- The clinical safety must be monitored on an ungoing basis after marketing approval
- The issue of immunogenicity should always be addressed, and its long-term monitoring is necessary

extensive data on quality and safety as an innovative protein drug. In addition, the submission needs a supplement showing similarity in quality safety and efficacy between the biosimilar and the same reference product.

A biosimilar is defined as any therapeutic protein by its own manufacturing process that may influence the characteristics of the product itself but also introduce specific process-related impurities. The producers of biosimilars therefore need to demonstrate consistency and robustness of their production process. They need to do formulation studies showing stability and compatibility, even if the formulation is identical to the reference product.

The generation of comparative data concerning physico-chemical characteristics and preclinical testing will be complicated because unformulated bulk material of the innovator protein is unavailable to the biosimilar manufacturer. So the competitor can only use formulated marketed material. For some analytical techniques, isolation of the protein from the formulation is necessary. This may induce modifications that can hamper the comparisons. Comparing with the biosimilar isolated from the same formulation is being used to compensate for these possible modifications.

The biosimilar producer needs to do at least one comparative trial, usually a clinical trial with a sufficient number of patients to give the statistical power to find differences in efficacy and safety. The results may be extrapolated to other indications if the mode of action in other disease conditions is identical and the toxicity profile comparable with the population suffering from the other conditions. An example could be interferon alpha 2. If the biosimilar is proven to be similar in efficacy with the innovator protein in chronic hepatitis C virus infection, it will also be active in hepatitis B. However, an extension to cancer indications such as hairy cell leukemia would be blocked by a different mode of action in cancer and a different toxicity profile in tumor patients. For example, the immunogenicity of interferon alpha 2 is significantly different in cancer patients compared with patients with viral infections.

An important issue in the regulatory guidelines is immunogenicity, which is considered a major difference between protein drugs and low-molecular-weight drugs [9–11]. According to the CHMP guidelines on biosimilars, the issue of immunogenicity always needs to be addressed. There are only limited possibilities for the prediction of immunogenicity *in vitro* or in animal studies. This implies that the immunogenicity always needs to be monitored in clinical trials. The assay strategy and the assays employed need to be validated. Moreover, the assays need to meet specifications that allow the detection of immunogenicity differences between the biosimilar and the innovative protein therapeutic.

Sometimes the incidence of the induction of antibody formation in patients is too low to be evaluated in the clinical trials of the biosimilar, but the consequences are so severe that this side effect needs to be excluded for a biosimilar as is the case with epoetin-associated pure red cell aplasia. This can only be accomplished by strict postmarketing surveillance protocols.

The companies marketing biosimilars will need to show their capacity to manage postmarketing protocols, not only to monitor rare side effects but also because the biological and clinical characteristics of therapeutic proteins cannot be completely predicted by physico-chemical analyses, and possible batch-associated side effects need to be identified.

Another specific problem of protein therapeutics is their relative instability, which calls for strict regimens in storage and handling. In principle, protein drugs need a cold chain that is maintained from the manufacturing plant until administration to the patient. The CHMP has indicated that biosimilar producers need to show that they control the distribution of their products.

Nearly all innovative proteins are being administered parentally. They are mainly prescribed by hospital medical specialists for sometimes relatively rare, knowledge intensive conditions needing individualized treatment and specific diagnostic facilities. The marketing of biosimilars needs therefore to be accompanied by specific information to ensure their proper and safe use.

Considering the large investments manufacturers have to do to develop, produce, distribute, and market biosimilars, a large price drop by their introduction will be unlikely. Unlike with classic generics, the market will not be conquered by price alone, but the biosimilar manufacturers and distributors need to show quality and added value of their products.

Another consequence of the relative modest price drop by the introduction of biosimilars on the expensive biotechnology-derived protein drug treatments will be the attractiveness for racketeers to counterfeit protein drugs. Several biosimilars have been produced for many years in countries where patent protection or enforcement is lacking. The quality of these products varies (Figure 1). Proteins as epoetin, growth hormone, and insulin are also popular as performance enhancers in sports. This is an important part of the illegal demand. Both unbranded products as well as counterfeits, which look like authorized products, have been discovered. Counterfeits of therapeutic proteins have also been detected in the regular medical supply chains of these products. And products as growth hormone are also popular among the illegal drugs offered by the Internet (Figure 2).

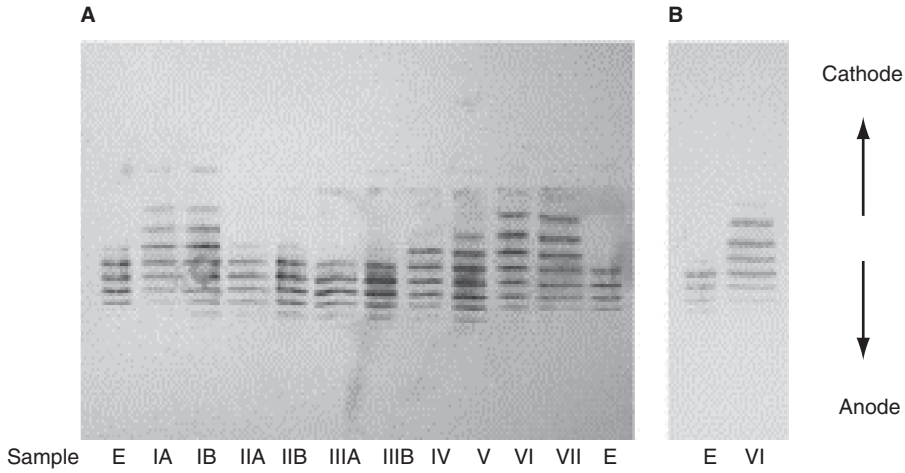


**Figure 1.** An example of biosimilar epoetins. This figure shows epoetin alpha's marketed in Asia and South America analyzed by iso-electrical focusing. The different products show large differences in isotype content. Difference can also be seen between different batches of the same manufacturer (IA and IB; IIA and IIB; IIIA and IIIB). E is the epoetin alpha marketed in Europe.

**Figure 2.** An example of a counterfeited growth hormone.

In Europe the first patents of protein drugs have expired and the legal conditions for launching biosimilars have been met, the general guidelines describing the regulatory demands have been implemented, and product-specific draft guidelines are under evaluation. The first marketing authorization requests have been submitted, and the marketing of these products will follow in 2006.

The biggest uncertainty concerning biosimilars is the level of similarity the regulatory bodies will find acceptable [12–13]. The only definition of a biosimilar that can be given is a protein that is similar in quality, safety, and efficacy to a reference product. Still, similarity is a matter of taste. Differences between products will always be found, and the question the regulatory agencies need to answer is how relevant these differences are. This is not an easy question because structurally comparable products can show significant differences in biological and clinical properties [14]. There are also examples of products with comparable clinical effects but considerably different structures. With room for interpretation about what a biosimilar is, legal procedures by the innovator or the generic industry challenging the decisions of the EU about marketing authorizations are to be expected. So, in the end the European Court of Justice will define what is a biosimilar.

## REFERENCES

1. Crommelin D J, Storm G, Verrijk R, et al. (2003). Shifting paradigms: Biopharmaceuticals versus low molecular weight drugs. *Int. J. Pharm.* 266(1–2):3–16.
2. Crommelin D, Bermejo T, Bissig M, et al, (2005). Biosimilars, generic versions of the first generation of therapeutic proteins: Do they exist? *Contrib. Nephrol.* 149:287–294.
3. Schellekens H, Ryff J C (2002). "Biogenerics": The off-patent biotech products. *Trends Pharmacol. Sci.* 23(3):119–121.
4. Schellekens H (2005). Follow-on biologics: Challenges of the "next generation". *Nephrol. Dial. Transplant.* 20(suppl 4):iv31–36.
5. Schellekens H (2004). When biotech proteins go off-patent. *Trends Biotechnol.* 22(8):406–410.

6. Guidelines on comparability of medicinal products containing biotechnology-derived proteins as active substance: Quality issues. EMEA/CPMP/BWP/3207/00/Rev 1.
7. Guidelines on comparability of medicinal products containing biotechnology-derived proteins as active substance: Non-clinical and clinical issues. EMEA/CPMP/3097/02/Final.
8. Chamberlain P (2004). Biogenerics: Europe takes another step forward while the FDA dives for cover. *Drug Discov. Today.* 9(19):817–820.
9. Schellekens H (2002). Immunogenicity of therapeutic proteins: Clinical implications and future prospects. *Clin. Ther.* 24(11):1720–1740, discussion 1719.
10. Schellekens H (2002). Bioequivalence and the immunogenicity of biopharmaceuticals. *Nat. Rev. Drug Discov.* 1(6):457–462.
11. Schellekens H (2004). Biosimilar therapeutic agents: Issues with bioequivalence and immunogenicity. *Eur. J. Clin. Invest.* 34(12):797–799.
12. Schellekens H (2004). How similar do "biosimilars" need to be? *Nat. Biotechnol.* 22(11):1357–1359.
13. Combe C, Tredree R L, Schellekens H (2005). Biosimilar epoetins: An analysis based on recently implemented European medicines evaluation agency guidelines on comparability of biopharmaceutical proteins. *Pharmacother.* 25(7):954–962.
14. Nieminen O, Nordstrom K (2004). Regulation of biogenerics: A survey of viewpoints. *BioDrugs.* 18(6):399–406.

# THE PROMISE OF INDIVIDUALIZED THERAPY

Michael Oettel
*Jena, Germany*

**Chapter Contents**

1 Background
2 Drug Metabolizing Enzymes
3 Drug Transporters
4 Ion Channels
5 Drug Receptors
6 Cancer Drugs
7 Cardiovascular Drugs
8 Drugs Acting on the CNS
9 Endocrinology
10 Environmental Factors
11 Ethnicity
12 Technological Aspects
13 Limitations
14 Educational Aspects
15 Concluding Remarks
References

## 1 BACKGROUND

The rapid progress in molecular medicine has sought to understand the molecular basis of human disease with an ultimate goal of developing rationally designed therapies. The *gene discovery phase* has been largely driven by key technological advances including polymerase chain reaction (PCR) and other strategies and

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

methods, high-throughput sequencing, and bioinformatics. Now that the genome is sequenced, there are ongoing efforts to identify genetic polymorphisms (e.g., single nucleotide polymorphisms, SNPs) that may point to disease predisposition or unique responses to therapy such as untoward drug side effects.

Presently, physicians have to optimize a dosage regimen for an individual patient by a trail-and-error method. This kind of blind approach may cause adverse drug reactions (ADRs) in some patients. In fact, ADRs are found to occur in more than 2 million hospitalizations including approximate 100,000 deaths per year in the United States [1]. Similarly, according to a German study, about 6% of ADRs are attributed to new hospital admissions, and these ADRs were found to be preventable [2].

This interindividual variability in drug response could be caused by multiple factors such as disease determinants, genetic and environmental factors, and variability in drug target response (pharmacodynamic response) or idiosyncratic response. These factors affect drug absorption, distribution, metabolism, and excretion [3]. Drug concentrations in plasma can vary more than 600-fold between two individuals of the same weight on the same drug dosage [4]. An understanding of the variability in efficacy and toxicity of the same doses of medications in the human population, therefore, may provide safer and efficient drug therapy.

In general, genetic factors are estimated to account for 15–30% of interindividual differences in drug metabolism and response; but for certain drugs or classes of drugs, genetic factors are of the utmost importance and can account for up to 95% of interindividual variability in drug disposition and effects [4].

The idea that drug response is determined by genetic factors that alter pharmacokinetics and pharmacodynamics of compounds evolved in the late 1950s, when an inherited deficiency of glucose-6-phosphate dehydrogenase was shown to cause the severe hemolysis observed in some patients exposed to the antimalarial primaquine. This discovery explained why hemolysis was observed mainly in African-Americans, in whom the deficiency is common, and rarely in Caucasians of northern, western, and eastern descent [5]. In 1959, Vogel coined the term pharmacogenetics to describe inherited differences in drug responses [6].

Later in 1962, W. Kalow defined **pharmacogenetics** as ". . . the study of heredity and the response to drugs" [7]. It is a well-recognized fact that individuals respond differently to drug therapy; some drugs that are effective or well-tolerated by some people may be ineffective or toxic to others. This variability can often be traced by SNPs in genes encoding drug-metabolizing enzymes, transporters, ion channels, and drug receptors, all of which have been known to be associated with interindividual variation in drug response and have aroused considerable interest in recent years [8].

Taken together, pharmacogenetics is the study of genetic polymorphisms in drug-metabolizing enzymes and the translation of inherited differences to the variability in drug effects. Genes are described as "polymorphic" when allelic variants exist in the population, one or more of which alters the activity of the encoded protein compared with the wild-type sequence. Typically, the polymorphism leads to reduced activity of the encoded protein. Although the focus of pharmacogenetics is the study of drug-metabolizing enzymes like the CYP family, polymorphisms in drug transporters (such as P-glycoprotein) band drug targets (such as receptors) have also received attention.

The original task of pharmacogenetic research is to aid physicians in the prescription of the appropriate medicine in the right dosage in an attempt to attain maximum efficacy and minimum toxicity based on a genetic test, performed before the initiation of the therapy. The new paradigm moves toward the approach of screening for polymorphisms associated with a drug response and tailoring clinical and therapeutic decisions for an individual patient. This strategy of targeting drugs according to the patient's genetic constitution is termed "personalized medicine" [8, 9].

Toward the goal of personalized medicine, after completion of the human genome project, a haplotype map (HapMap) has been developed by the International HapMap Consortium with an intention of profiling DNA sequence variations across the human genome, which should provide a powerful tool to understand the genetic variants and drug responses (biomarkers). This knowledge may ultimately allow the development of personalized medications based on the genotype of each patient [10].

The term "**pharmacogenomics**" was introduced to reflect the transition from genetics to genomics and the use of genome-wide approaches to identify genes that contribute to a specific disease or drug response. A pharmacogenomics approach could allow a specific drug therapy to be targeted to genetically defined subsets of patients and could lead to a new disease and treatment classification on the molecular level.

Although the "blueprints" of human disease and for the drug action may be genetically encoded, the execution of the disease process and the pharmacodynamics of a given drug occurs through altered protein function. Researchers in molecular medicine are currently going from genomics to proteomics. One goal for clinical proteomics will be to characterize the information flow within single cells, tissues, or entire organisms under normal or disease conditions [11]. Therefore, identifying the genetic or epigenetic events leading to wanted or unwanted effects of drugs requires subsequent understanding of the proteomic consequences of these events. Therefore, **pharmacoproteomics** are focused on the influence of drugs on protein–protein interactions, their localization, or whether the encoded proteins are stable expressed, phosphorylated, cleaved, acetylated, glycosylated, or functionally "active." Mounting evidence confirms that the low-molecular-weight (LMW) range of circulatory proteome contains a rich source of information that may be able to detect or better characterize drug–protein interactions and stratify toxicological risk. Current mass spectrometry (MS) platforms can generate a rapid and high-resolution portrait of the LMW proteome. Emerging novel nanotechnology strategies to amplify and harvest these LMW biomarkers *in vivo* or *ex vivo* will greatly enhance our knowledge about pharmacoproteomics [12]. As an example for an pharmacoproteomic approach, it was found that in comparison with tamoxifen-sensitive breast tumor cells, in an tamoxifen-resistant line, 12 proteins were found upregulated, whereas nine were downregulated. Three of the identified proteins (AGL-2 interacting protein and two GDP-dissociation inhibitors) could be directly involved in the resistance phenomenon [13]. Curiously enough, the erythrocyte sedimentation reaction (ESR), which could be defined as a primitive forerunner of functional proteomics, has been used for a long time for clinical diagnostics and for monitoring of pharmacological therapy.

Although genetic variation is clearly important, it seems unlikely that personalized drug therapy will be enabled for a wide range of major diseases using genomic knowledge alone. Therefore, a third modern possibility for enabling personalized pharmacological therapy is the phenotyping by **pharmacometabonomics**. A major factor underlying interindividual variation in drug effects is variation in metabolic phenotype, which is influenced not only by genotype but also by different factors such as *in utero* effects, lifestyle, nutritional status, the gut microbiota, age, disease, environment, and the co- or pre-administration of other drugs. A new approach to personalized drug treatment is the examination of the metabolic profile. The profile, which is a measurement of small molecules such as sugars and amino acids, could be used to predict the pharmacodynamic or toxic response to drugs [14].

Pharmacometabonomics use a combination of pre-dose metabolite profiling and chemometrics to model and predict the responses of individual subjects. $^{1}$H nuclear magnetic resonance (NMR) spectroscopy has been applied as a metabolite profiling tool for metabonomic studies, as it enables many endogenous metabolites to be quantified rapidly and reproducibly in biological fluids without derivatization or separation. Pharmacometabonomics has an important theoretical advantage over pharmacogenomics in that it can potentially take account of both genomic and environmental factors affecting drug-induced responses [15].

However, because the time of practical use of pharmacoproteomics and pharmacometabonomics is obviously more distant from today as such for the genetic approaches, we will focus this chapter more on pharmacogenetics and pharmacogenomics. The present knowledge certainly does not allow or recommend individualized therapy on the basis of pharmacoproteomics as well as pharmacometabonomics. They are not ready for profound optimization of clinical drug application or rationale drug development to become a reality in the near future. Intensification of research capacities is assumed, providing optimism for personalized medicine.

In the following sections, depending on the target respective, selected examples for the influence of genetic variability on the drug response are discussed. In most of the cases, we will cite publications from the first half of the year 2006, indicating the enormous progress in this field in the last months. For excellent reviews on individualized therapy solely from 2006, see Refs. 4, 8, 14, 16–25.

## 2 DRUG METABOLIZING ENZYMES

A considerable body of evidence suggests that SNP in genes encoding drug-metabolizing enzymes might determine drug efficacy and toxicity. The cytochrome P450 (CYP) enzymes consist of a superfamily of haem-containing monooxygenases, and multiple forms of CYP exist in all mammals. CYPs are responsible for the oxidation of many drugs, environmental chemicals, and endogenous substrates. They exist in the liver and in extrahepatic tissues such as the intestine, lung, and kidney. In humans, xenobiotics are metabolized primarily by three CYP subfamilies: CYP1, CYP2, and CYP3 [26, 27] (Table 1). Genetic polymorphisms of the genes for CYP2C9, CYP2C19, and CYP2D6 affect the metabolism of 20–30% of clinically used drugs. The frequency of variant alleles of CYP families varies among populations according to the race and ethnic background [28].

**TABLE 1. Selected Drugs Metabolized by Various Cytochrome P450 (CYP) and Other Enzymes**

| Enzyme | Drug |
|---|---|
| CYP1A2 | Imipramine, Tacrine, Propranolol |
| CYP1B1 | 17β-Estradiol, Estrone, Testosterone, Tamoxifen |
| CYP2B6 | Carbamazepine, Efavirenz |
| CYP2C8 | Paclitaxel, Carbamazepine |
| CYP2C9 | Cyclosporine, Ifosfamide, Etoposide Nefazodone, Losartan, Warfarin, Tamoxifen, Phenytoin |
| CYP2C19 | Omeprazole, Lansoprazole, Citalopram, Thalidomide |
| CYP2D6 | Fluoxetine, Paroxetine, Desipramine, Amitryptiline, Imipramine, Nortryptilin, Maprotilin, Metoprolol, Propranolol, Tamoxifen |
| CYP 3A (responsible for the oxidative metabolism of more than 50% of clinically used drugs) | |
| CYP3A4 | Amitryptiline, Imipramin, Cyclosporine, Docetaxel, Epipodophyllotoxin, Erythromycin, Losartan, Diazepam, Carbamazepine, Nefadozone, Nifedipine, Omeprazole, Terfenadine, Chlorpyrifos (Insecticide), Testosterone, Estradiol, Progesterone, and synthetic sexual steroids |
| CYP 3A5 | Etoposide, Vincristine, Midazolam |
| CYP7A1 | 17α-Ethinylestradiol, Statins |
| CYP7B1 | 17α-Ethinylestradiol |
| CYP8B1 | 17α-Ethinylestradiol |
| Dihydropyrimidine Dehydrogenase (DPD) | Fluorouracil, Capecitabine, Nifedipine |
| Glucuronosyl transferase | Labetalol, Morphine, Naloxone |
| Glutathione—Transferase (GST) | |
| GSTA1 | Cyclophosphamid |
| GSTM1 | Anthracyclines (Doxorubicin, Idarubicin, Mitoxantrone) |
| GSTM3 | Cisplatin |
| GSTP1 | Etoposide, Doxorubicin, Cisplatin, Oxaliplatin, Carboplatin |
| GSTT1 | Prednisolon |
| N-Acetyltransferase | Amonafide |
| S-Methyltransferase | Captopril |
| Sulfotransferase1A1 (SULT1A1) | Tamoxifen |
| Thiopurine Methyltransferase (TPMT) | 6-Mercaptopurine, Azathioprine, Methotrexate |
| Uridine Diphosphate Glucuronosyltransferase (UGT) | Irinocetan, Etoposide, Epirubicine, Tipifarnib |

For instance, there are 78 reported variants of CYP 2D6 that associated with adverse drug reactions. Many of these polymorphic genes encode inactive enzymes. However, these inactive enzymes may produce adverse drug reactions among patients because of their poor metabolic activity (e.g., the adverse effects of the neuroleptic risperidone) [29].

Similarly, several inactivating genetic polymorphisms have been reported in another member of the CYP family, namely CYP2C19 (CYP2C19*2 and CYP2C19*3), which is also associated with adverse drug reactions. This enzyme is responsible for the metabolism of proton pump inhibitors (e.g., omeprazole and lansoprazole) used for the treatment of gastric acid-related disorders such as peptic ulcer and gastroesophageal reflux disease. Approximately 2–4% of Caucasians and 4% of African-Americans have poor metabolism of these drugs [26]. Poor metabolizing patients of proton pump inhibitors carry two nonfunctional alleles, heterozygous extensive metabolizers have one nonfunctional and one wild-type allele, and extensive metabolizers are homozygous for the wild-type allele [4].

Another example is the coumarin warfarin, which is widely used for oral anticoagulation. Major side effects include bleeding complications. It was found that CYP2C9*2 and CYP2C9*3 alleles reduce the clearance of warfarin and increase the risk of bleeding [30]. The data indicate that patients carrying at least one variant CYP2C9 allele require lower maintenance doses and have a significantly higher risk of bleeding. However, SNPs in vitamin K epoxide reductase (VKORC1) may be more important. Recent studies have identified haplotype-dependent predictions for warfarin dosing. VKORC1 haplotype A predicted 21–25% of the required warfarin dose, and the inclusion of CYP2C9 genotypes reduced the required warfarin dose up to 31% in Caucasians. Combining nongenetic factors such as age, sex, body surface area, and drug interactions with the genotype information predicts up to 60% of warfarin dose. The remaining 40% of warfarin dosing variability remains unexplained [31].

CYP2C9*13 allele is associated with reduced metabolism of lornoxicam [32]. Similarly, CYP2C8 plays a role in the disposition of some therapeutic drugs [33].

The intestinal epithelium and liver contain the most abundant member of the CYP family, namely CAP3A, and these enzymes are responsible for the metabolism of more than half of the therapeutic drugs. Its activity also varies among members of a given population. In addition, this enzyme may undergo induction (rifamycins) and inhibition (calcium channel blockers) depending on the drug administration, which may account for its poor or higher metabolic activity. The interindividual variation in the immunosuppressive drugs cyclosporine and tacrolimus could be caused by interindividual differences in the expression of CYP 3A4 and 3A5 and the drug transporter P-glycoprotein. The most frequent allelic variant of CYP3A5 is CYP3A5*3, with a frequency of 87% of all alleles in a French population [34, 35].

An important P450 database in terms of genetic polymorphisms is the CYP alleles database (http://www.imm.ki.se/CYPalleles). However, genetic variants identified in the CYP3A4 and CYP3A5 genes have only a limited impact on the CYP3A-mediated drug metabolism [36], and hence the identification of the genotype for the ABCB1 transporter gene may provide further clues for the individualization of therapy with certain drugs (see below).

Apart from CYPs, some other enzymes involved in the biotransformation of drugs include *N*-acetyltransferase (NAT), glutathione-*S*-transferase, and uridine

diphosphate-glucuronosyl transferase 1A1 (UGT1A1). One of the earlier discoveries of pharmacogenetics is the attribution of the neurological side effects of the antituberculosis drug isoniazid to genetic variability of NAT2. A 98.1% correlation between genotyping of NAT2 and acetylation phenotype has been demonstrated using an allele-specific PCR [37].

Gastric cancer patients treated with 5-FU and cisplatin and possessing the glutathione S-tranferase PI-105 Valine/Valine (GSTPI-105VV) genotype showed a response rate of 67% and medial survival time of 15 months compared with 21% and 6 months, respectively, in patients harboring one GSTPI-105 Isoleucine (GSTPI-1051) allele [38].

Pharmacogenetics of the UGT1A1 gene is known to influence irinotecan-induced diarrhea mediated via the glucuronidation of the active metabolite SN-38 [39]. Irinocetan is an inhibitor of topoisomerase used for the treatment of lung and colon cancer in adults and pediatric solid tumors such as rhabdomyosarcoma and neuroblastoma. The presence of seven repeats, instead of the wild-type number of six (UGT1A1*28) is associated with reduced UGT1A1 expression, leading to reduced SN-38 glucuronidation. It has been shown that the UGT1A1*28 allele leads to significantly increased amounts of SN-38 and a heightened risk of irinocetan-caused diarrhea and leukopenia. Patients homozygous or heterozygous for seven TA repeats have a sevenfold higher likelihood of diarrhea or leukopenia with irinocetan therapy than do patients with the wild-type genotype (six repeats). In a study conducted on Asians, UGT1A1*28 was found to be a common allele in Indians [40]. Thus, it is possible that UGT1A1 genotyping can be used to predict toxicity to irinocetan therapy. The UGT database contains data on genetic polymorphisms of the various UGT alleles (http://som.flinders.edu.au/FUSA/ClinPharm/UGT).

A good example for clinical reality of pharmacogenetic methods is the very frequently performed test for pseudocholinesterase. Inherited deficiency in pseudocholinesterase activity results in prolonged respiratory paralysis when deficient patients receive standard doses of the neuromuscular blockers suxamethonium or mivacurium [41].

## 3 DRUG TRANSPORTERS

Genetic variability in drug transporters also plays an important role for individual drug response (see Table 2). For instance, polymorphism in the ABC-binding cassette (ABC) gene may affect the function and expression of proteins [42] by inducing tumor cell resistance to anticancer therapy, altered disposition of chemotherapeutic agents, and associated chemotherapy toxicity, which may cause certain drug-induced side effects and uncertainty in treatment efficacy. ABCB1, also known as P-glycoprotein (Pgp) alias MDR1 for multidrug resistance, is a member of the ABC family that participates in the energy-dependent efflux of various substrates. A synonymous polymorphism in exon 26 (C345T) was found to influence drug response and exhibit interethnic variability [43]. The importance of haplotype analysis was shown by a meta-analysis of the influence of this SNP on digoxin pharmacokinetics and Pgp gene expression [44].

The ABCB1 genotype of the donor, but not of the recipient, may be a major risk factor for cyclosporine-related chronic nephrotoxicity in recipients of renal transplants. The ABCB1 3435TT genotype, which is associated with lower Pgp

**TABLE 2. Selected Drug Transporters Associated with Individual Variations in Drug Response**

| Protein | Drug |
|---|---|
| P-Glycoprotein; Synonyms: Multidrug Resistance 1 (MDR1, ABCB1) | Digoxin, Vinblastine, Vincristine, Irinocetan |
| Multidrug Resistance-Associated Protein 2 (Canalicular-Multispecific Organic Anion Transporter, MRP2; ABCC2) | Doxorubicin, Vinblastine, Sulfinpyrazone, Irinocetan, SN38, Methotrexate |
| Breast Cancer Resistance Protein (BCRP, ABCBG2, ABCP) | Mitoxantrone, Methotrexate, Doxorubicin, Camptothecin-based drugs like Topocetan and SN-38 |
| ABCG5/8 | 17α-Ethinylestradiol |
| $Na^+$/taurocholate cotransporting polypeptide (NTCP) | 17α-Ethinylestradiol |
| Organic Anion Transporter Polypeptide-1B1 (OATP1B1; *SLCO1B1*) | Irinocetan |
| OATP1/2 | 17α-Ethinylestradiol |
| OATP1A2 (OATP-A) | Ajmalin, DHEA-Sulfate, 17β-Estradiol, Estrone, Microcystin, N-Methyl-quinidine, Ouabain, Prostaglandin $E_2$, Rocuronium, Thyroxin (T4), Triiodothyronine (T3) |
| OATP1B1 (OATP-C/OATP2) | Atorvastatin, Cerivastatin, Fluvastatin, Benzylpenicillin, Rifampicin |
| OATP1B3 (OATP8) | CCK-8, Digoxin, Fexofenadine, Paclitaxel |

expression in renal parenchymal cells, is strongly associated with cyclosporine nephrotoxicity (odds ratio 13.4) [45]. On the other hand, the cyclosporin disposition in heart transplant patients may be influenced by Pgp haplotypes rather than genotypes [46].

Genetic polymorphisms in ABCB1 and ABCg2 may be important also in influencing the pharmacokinetics of irinotecan and its metabolites (see above).

Another notable example is that in certain patients the reduced rate of methotrexate metabolism produced a severe methotrexate overdosing and nephrotoxicity. This defect is attributed to the heterozygous mutation (R412G) in the highly conserved amino acid arginine [47] of the ABCC2 gene, which encodes the human multidrug resistant protein-2 (MRP-2). Interestingly, this mutated region is associated with substrate affinity and hence the mutant protein has a reduced rate of methotrexate elimination. In some other cases, a long-term use of methotrexate induces pancytopenia, which is determined by white blood cells and platelet counts [48]. However, it is also known that polymorphisms always need not have to produce functionally defective proteins. For example, in the multidrug resistant gene (MDR1), certain polymorphisms may not have any effect on the drug response [49]. However, this could be caused by nonsignificant statistical power.

Gwee et al. [50] have devised a rapid and robust assay to simultaneously screen SNPs of the MDR1 gene using a single-tube multiplex minisequencing strategy. Finally, there are several web-based transporter databases (http://www.tp-search.jp; http://www.gene.ucl.ac.uk/nomenclature/genefamily/abc.html and http://lab.digibench.net/transporter).

## 4 ION CHANNELS

Organic anion transporting polypeptides (OATPs) mediate the uptake of a broad range of compounds into cells. Substrates for members of the OATP family include bile salts, hormones, and steroid conjugates as well as drugs like the HMG-CoA-reductase inhibitors (statins), cardiac glycosides, anticancer agents like methotrexate, and antibiotics like rifampicin. The identification and functional characterization of naturally occurring variations in genes encoding human OATP family members is in the focus of transporter research [51]. There is a high degree of functional heterogeneity among OAT3 variants, with three variants (p.Arg149Ser, o. Gln239Stop, and p.Ile260 Arg) that results in complete loss of transporter function, and several other variants with significantly reduced function [52].

Common variation in the gene SCN1A affects the maximum dose of phenytoin and carbamazepine, which act on the sodium channel subunit encoded by this gene [53].

The etiology of drug-induced long QT syndrome (LQTS) could also be based on gene variability. A life-threatening form of cardiac arrhythmia has been associated with mutations in the ion channel genes [54]. Therefore, screening of LQTS-associated genes before the initiation of therapy with known QT-prolonging drugs could serve as a precautionary measure against life-threatening adverse effects by avoiding such drugs.

## 5 DRUG RECEPTORS

Genetic polymorphisms in drug receptors may alter pharmacological response. For example, variations in β-adrenoceptors, angiotensin-converting enzyme (ACE), and 5-hydroxytryptamine (5-HT) receptors alter drug response to β-adrenoceptor agonists and blockers, ACE inhibitors, and antipsychotic agents, respectively (see the sections cardiovascular dugs and drugs acting on the CNS in this chapter).

Genetic polymorphisms affecting amino acids at positions 16 and 27 within the $\beta_2$-adrenoceptor gene have been implicated in the asthma phenotypes and influence on the variability observed in response to use bronchodilator agents. It was found that Arg16 allele was slightly more frequent within the group with the unwanted tachyphylaxis phenomenon, whereas Gly16 allele carriers were overrepresented within the group of good responders (59.7%, $P = 0.0028$). On the other hand, the allele frequency of Gln27 and the proportion of Gln27 carriers was higher within the group with tachyphylaxis and Glu27 allele carriers were overrepresented within the group of good responders ($P = 0.026$) [55]. For asthma, see the section "Cardiovascular Drugs" in this chapter.

Concerning pharmacogenetic aspects of hormone receptors, see the "Endocrinology" section in this chapter.

## 6 CANCER DRUGS

Cancer chemotherapy is an area that requires continual monitoring and adjustment of antineoplastic agents to achieve optimal therapeutic outcome, which makes pharmacogenetic approaches attractive in this field.

Drugs such as azathioprines, mercaptopurines, and thioguanine have been used extensively to treat childhood acute lymphoblastic leukemia, rheumatic disease, inflammatory bowel disease, and used for solid organ transplantation. Thiopurine S-methyltransferase (TPMT) is a cytosolic enzyme that is involved in the metabolism of thiopurines (Table 1). It has been shown that polymorphisms in TPMT result in severe toxicity for patients prescribed with normal doses of the cytotoxic agents mercaptopurine or azathioprine. The variant enzyme was shown to misfold and subsequently form aggresome [56]. It has been reported that the TPMT genotype has a substantial impact on the mercaptopurine treatment response [57]. Previous studies also have shown that patients with homozygous mutant TPMT alleles exhibit very low enzyme activity and develop a severe hematopoietic toxicity after treatment with standard doses of thiopurines [58–60]. TPMT-deficient patients tend to accumulate excessive thioguanine nucleotide concentrations and are, therefore, at higher risk for hematological toxicity. TPMT deficiency can be largely attributed to three mutant alleles (TPMT*2, TPMT*3A, and TPMT*3C). Allele-specific PCR or PCR-restriction fragment length polymorphism strategies have been used to detect the three signature mutations, hence offering a rapid and affordable assay for identifying >90% of all mutant alleles [61]. Today, one of the most frequently performed pharmacogenetic tests are for TPMT. Patients who inherit two nonfunctional variant alleles should be given 6–10% of the standard dose of thiopurines. TPMT deficiency has also been associated with a high risk of irradiation-induced brain tumors in patients given thiopurines concomitantly with radiation therapy [62].

Similarly, the response rate of 5-fluorouracil (5-FU)-based treatment of advanced colorectal cancer is significantly linked to 677 C → T polymorphism in the methylenetetrahydrofolate reductase gene [63]. Additionally, polymorphisms in the thymidylate synthase gene promoter (TYMS enhancer region, TSER) has been linked to tumor downstaging in patients with rectal cancer who were treated preoperatively with 5-FU-based chemoradiation. In a study of 65 patients with stage T2–T4 rectal cancer, patients with at least one TSER*2 allele had a 38% increased frequency of tumor downstaging at the time of surgical resection compared with TSER*3/TSER*3 patients [64].

The first genotype-guided clinical trial in North America is based on TYMS TSER genotype. Rectal cancer patients (stage T3 and T4) with the "good risk" TSER*2 allele are treated in a phase II study consisting of standard therapy (radiation and 5FU). The sample size was calculated to detect a downstaging rate of 60%, compared with the historical downstaging rate of 45%. Patients homozygous for TSER*3 ("bad risk" genotype) are also enrolled in a phase II study, in which they receive the standard radiation and 5-FU along with additional irinocetan. The sample size was calculated to detect an improvement from the previously reported TSER*3/TSER*3 downstaging rate of 22–45%. Preliminary data implied an improved response rate in both treatment groups, suggesting an enrichment for positive response [22].

Genetic polymorphisms in the epidermal growth factor receptors (EGFR) impact on pharmacological response to gefitinib and erlotinib, tyrosine kinase inhibitors used as monotherapy in the treatment of metastatic non-small cell lung cancer. The drugs are effective in only 10–15% of patients. Responders were found to harbor activating mutations in the gene coding for EGFR [65]. The EGFR assay for tumor mutation analysis provides a platform in personalized therapy with the EGFR inhibitors gefitinib or erlotinib (http://www.dxsgenotyping.com).

The use of the other tyrosine kinase inhibitor imatinib, which blocks the enzymatic action of the BCR-ABL fusion protein, has represented a critical advance in chronic myeloid leukaemia (CML) treatment. However, a subset of patients initially fails to respond to this treatment. Use of complementary DNA (cDNA) microarray expression profiling, a set of 46 genes was differentially expressed in imatinib responders and non-responders. A six-gene prediction model was constructed, which was capable of distinguishing cytogenetic response with an accuracy of 80% [66].

Another case of already practical use of pharmacogenetic methods is the test for mutations in tumors that overexpress the human EGFR, HER2. Trastuzumab, a humanized monoclonal antibody, is effective in only 10–15% of breast cancer patients whose tumors overexpress HER2 [67]. Therefore, the pretreatment detection of HER2 is essential for the trastuzumab therapy.

## 7 CARDIOVASCULAR DRUGS

Variation in two genes encoding angiotensin-converting enzyme and endothelial nitric oxide synthase influence the effects of standard therapies [68]. In addition, polymorphism in the sodium channel gamma-subunit promoter region is significantly associated with blood pressure response to hydrochlorothiazide [69]. Similarly, SNPs in angiotensinogen (T1198C), apolipoprotein B (G10108A), and adrenoreceptor alpha 2A (A1817G) significantly predict the change in left ventricular mass during antihypertensive treatment [70]. Although common variants may influence the blood pressure response to a given class of antihypertensive medication, studies of polymorphisms have generally provided conflicting results [71]. For instance, polymorphisms in the alpha 2B adrenergic receptor does not show any association with azepexole hypertensive response [72]. However, patients with Gly 389 variant and Ser 49 homozygous of the beta-adrenergic receptor require increases in heart failure medication [73, 74].

One study has reported that the effect of statins in lowering low-density lipoprotein (LDL)-cholesterol levels was slightly greater in -204AA homozygotes of CYP7A1 [75].

Most of the studies to date have failed to demonstrate any link between polymorphism in tumor necrosis factor alpha and both cardiomyopathy and coronary artery disease [76].

In the case of asthma that causes substantial economic burden, morbidity, and mortality, patients exhibit an extensive interindividual variation in the response to beta-agonists acting at beta 2 adrenergic receptors, which could be caused by one nonsynonymous polymorphism [1772M9 of adenylyl cyclase type 9 (AC 9)] gene. This variation results in decreased catalytic activity (M772) and, therefore, alters

albuterol (bronchodilator) responsiveness in the presence of a corticosteroid [77]. Additionally, in an Indian population, response to salbutamol treatment of asthmatic patients depends on polymorphism of the beta 2 adrenergic receptor [78].

## 8 DRUGS ACTING ON THE CNS

A meta-analysis of the quantitative contribution of CYP2D6 polymorphism to the interindividual variation in dosage of antidepressants has shown that the metabolism and dosage of imipramine, doxepin, maprotiline, trimipramine, desipramine, nortryptiline, clomipramine, and, partially, paroxetine depend on the CYP2D6 genotype and phenotype. Pharmacokinetic data suggest dose adjustments for these drugs that range from 28% to 60% of the normal dose for poor metabolizers and from 180% to 14% of the normal dose for ultrarapid metabolizers. In general, based on the impact of CYP2D6 on dosage adaption of antidepressants and antipsychotics, 40–50% of drugs may be subject to important pharmacokinetic alterations owing to CYP2D6 polymorphism [79].

A considerable variability also exists in efficiency and toxicity of other antipsychotic drugs. For instance, in the case of mood disorder, approximately 30–40% of patients do not completely respond to pharmacological treatment [80, 81]. However, serotonin transporter gene promoter (5HTTLPR) length polymorphisms has been implicated in the pathogenesis of mood disorders as well as in the therapeutic response to serotonergic drugs [82]. Reduction in the Liebowitz social anxiety scale and in the brief social phobia score during treatment with serotonin reuptake inhibitors (SSSRIs) was significantly associated with 5HTTLPR genotype [83]. When patients were treated with serotonin-blocking antidepressants, a significantly higher occurrence of side effects was found in patients with the HTTVNTR2.10/2.10 genotype (52.6%) than in patients with the 2.10/2.12 (12.5%) and 2.12/2.12 (0%) genotypes [84].

In patients with schizophrenia, Taq I polymorphism in the dopamine D2 receptor is associated with greater improvement of symptoms after treatment. Similarly, Gly 9 allele (Ser 9 Gly) of the dopamine D3 receptor and His 452 Tyr polymorphism in the 5-hydroxytryptamine 2A receptor (5-HT2A) are associated with response to clozapine. The side effects (weight gain) induced by antipsychotics seems to be associated with the -759C allele of the 5-HT2C receptor. Additionally, Gly 9-variant of dopamine D3, the 102C-variant of the 5-HT2A, and the Ser 23-variant of the 5-HT2C receptors (in females) seem to increase the susceptibility to tardive dyskinesia [85, 86].

Epilepsy is a difficult disease to treat because different patients require different ranges of doses, and some patients may even experience side effects such as increase in seizures, depression, and double vision. To control epilepsy, drugs such as phenytoin and carbamazepine have been extensively prescribed throughout the world. At present, evaluation of the allelic variation between individuals relies on the prior identification of candidate genes and their therapeutic effects of antiepileptic drugs [87]. Variants in the CYP2C9 and SCN1A (encodes a brain protein) genes are often found in patients treated with the highest doses of both phenytoin and carbamazepine [53]. Additionally, in Han Chinese, the carbamazepine side effects like Stevens–Johnson syndrome and toxic epidermal necrolysis are strongly associated

with the HLA-B*1502 gene, which also means that genetic susceptibility to carbamazepine-induced cutaneous adverse drug reactions is phenotype-specific [88]. Pharmacoresistant epilepsy is still a major clinical problem in epilepsy, and it could be caused by multiple factors, but also multidrug transporters may play a key role in resistance phenotypes. However, studies on one variant in the ABCB1 gene provided inconclusive evidence so far [89].

Long-term treatment of Parkinson patients with L-Dopa exhibits L-Dopa-induced dyskinesis in some patients, which could be caused by genetic polymorphisms among patients. Therefore, pharmacogenetic studies may provide an explanation of neuronal plasticity among Parkinson patients [90].

Furthermore, drug addictions are major social and medical problems and therefore impose a significant burden on society. Epidemiological, linkage, and association studies have shown a significant contribution of genetic factors to the addictive diseases. Studies of polymorphisms in the mu-opioid receptors and transporter genes have contributed significantly to the knowledge of genetic influence on opioid and cocaine addiction and the efficacy of opioid therapy in pain management [91–94].

## 9 ENDOCRINOLOGY

At present, a lot of scientific efforts are being employed to use pharmacogenetics/pharmacogenomics for protein/peptidergic hormone therapy as well as for treatment with steroids. The main topics are polymorphisms in the membranous and nuclear receptors, including their subtypes and isoforms, and the genetics of steroid transforming enzymes (aromatases, 5α-reductases, sulfotransferases). It seems that, in comparison with other fields of pharmacological approaches, the pharmacogenetics in endocrinology are relatively advanced and, in certain parts, ready for clinical use.

Growth hormone receptor (GHR) transcripts exist in several isoforms in humans, among which is the retention (GHRfl) or exclusion (GHRd3) of exon 3, which encodes a 22-residue sequence in the extracellular domain of the membrane-located receptor. In Western Europe an populations, it has been estimated that 68–75% of alleles are GHRfl, whereas 25–32% are GHRd3. In short children, the homozygous or heterozygous presence of GHRd3 resulted in a significantly greater growth response in both year 1 and year 2 of GH therapy [95, 96]. Logically, patients who are homozygous for GHRd3 were less responsive to short-term and long-term hGH therapy [97].

Exogenous sexual hormones are used worldwide by women as oral contraceptives and hormonal replacement therapy. Some epidemiological studies have shown an increased risk of venous thromboembolism (VTE). It was found, that the risk/benefit ratio could be, in part, mediated by the genetic predisposition of women. Genetic thrombophilia might be implicated in the risk of VTE patients who use exogenous hormones.The most common causes of genetic hypercoagulability known today are factor V Leiden, G20210A prothrombin polymorphisms, and the genetic variant C677T of the methylenetetrahydrofolate reductase (MTHFR) gene. Therefore, an increasing number of kits for these two thrombophilic mutations are becoming commercially available, and screening for inherited thrombotic risk

before giving the "pill" is among the most requested genetic tests in molecular diagnostic laboratories [98].

It seems that use of oral contraceptives or postmenopausal hormone replacement in women with the germ line mutations in the two genes BRCA1 and BRCA2 are more at risk for breast cancer than women also carrying these mutations but without hormonal interventions [99].

CYP1A2, CYP2C19, and CYP3A5 are responsible for estrone oxidation. These enzymes are all known to be genetically variant in the human population, and studies to asses the role of these CYP P450 enzymes in breast cancer risk are indicated [100].

The estrogen receptor-subtype ERα mediates the hepatotoxicity of 17α-ethinylestradiol (EE2). Upon EE2 treatment, ERα represses the expression of bile acid and cholesterol transporters (bile salt export pump, BSEP), $Na^+$/taurocholate cotransporting polypeptide (NTCP), OATP1, OATP2, ABCG5, and ABCG8 in the liver [101]. The genetic variability of some of these transporters is well known and could explain, at least in part, the interindividual differences for the tolerability of oral contraceptives.

Estrogen receptor α variations increase or decrease the action of estrogens; but, at present, no clinical studies are available that show that pharmacogenomic checking of ERα before estrogen-treatment (OCs or hormone replacement) can reduce the risk of adverse drug reactions.

The progesterone receptor 660L allele (PGR-12(rs1042638)V660L) may be associated with a moderately increased risk of breast cancer [102].

The polymorphism of UDP-glucuronosyl transferase (UGT2B17) is strongly associated with the bimodal distribution of the testosterone excretion [103].

Interestingly enough, besides the encouraging findings of genetics in the field of clinical endocrinology, proteomic approaches on the effects of estrogens, progestins, and androgens on the mammary gland are advancing step-by-step [13, 104, 105].

And what happens with diabetes mellitus? In 525 Caucasian type 2 diabetic patients, the common E23K variant of KCNJ11 encoding the pancreatic β-cell adenosine 5′-triphosphate-sensitive potassium channel subunit Kir6.2 was associated with increased risk for secondary failure to treatment with sulfonylurea-like glibenclamide [106].

## 10 ENVIRONMENTAL FACTORS

Needless to say, the genetic background and the gene variability is only one aspect of pharmacodynamics. Apart from potential gene–gene interactions, drug actions are also deeply affected by gene-environment interactions such that a particular genetic marker may present a variable pharmacological response in individuals with different nutritional states, lifestyle habits, and general well-being. Thus, the genetic information is not a reliable predictor of drug response, and therapeutic drug monitoring, with several notable exceptions, remains generally empirical. The sum and substance is as follows: Prescription genotyping serves more a predictive rather than a diagnostic role [8].

Here, we give only two examples for the role of the environment. Interindividual variability has been seen in liver UDP-glucuronosyltransferase 1A6 (UGT1A6)

enzyme activity that glucuronates various drugs and toxins (Table 1). Its expression is associated with polymorphisms in the 5′-regulatory and exon 1 regions. The three most common non-synonymous polymorphisms are S7A, T181A, and R184S. However, it did not explain the interindividual variability in glucuronidation and alcohol consumption, which suggests that environmental factors may have a significant role in alcohol consumption [107, 108]. Similarly, alcohol dependence is not associated with single-nucleotide polymorphisms in the corticotrophin releasing hormone receptor 1 (CHRH 1) gene [109].

## 11 ETHNICITY

To use genomic knowledge to develop drugs and to improve health, we need to consider ethnical differences in different populations [110, 111]. There exists inter-ethnical differences in polymorphisms of genes encoding drug metabolizing enzymes, transporters, and disease-associated proteins [112, 113]. Meanwhile, a population genetics-based method to calculate the probability value for a variation in the gene is proposed [114]. Genetic differences are greater within socially defined racial groups than between other groups [115]. Additionally, it has been found that genetic diversity decreases in noncoding regions, whereas diversity of coding non-synomous SNPs is lower in regions containing a known protein sequence motif in individuals of European origin [116].

Drug treatment may be tailored for greater effect if important genetic variation exists between racial and ethnic groups. By knowing these variants, patients can be classified into low-, intermediate-, and high-dose groups [31, 117]. For instance, coumarins are characterized by a narrow therapeutic index and a wide interindividual variability in dose response; daily maintenance doses of warfarin range from less than 1 mg to over 20 mg [30] and, additionally, warfarin therapy shows a wide variation among patients of different ancestries. This variation could be caused by polymorphisms in the gene encoding vitamin K epoxide reductase complex 1. Accordingly, Chinese patients require lower dosages of heparin and warfarin than those usually recommended for white patients [118, 119]. Additionally, the combination of isosorbide dinitrate and hydralazine for treatment of heart failure in African-American heart patients reduced mortality by 43%, claiming that African-Americans and Caucasians differentially respond to the treatment, which is claimed to be because of genetic differences in the pathophysiology of heart failure between the two groups [120]. In other words, biological differences exist between the two racial groups. However, in this study, there is no comparison population and hence results should be interpreted cautiously. The distribution of haplotype profile of MDR1 (Pgp, ABCB1) has also been shown to exhibit inter-ethnic variabililty [121].

The ethical and moral concerns that develop in the midst of genetic testing are factors that hamper the development of personalized medicine. The deciphering of the genetic code may pose a threat to the protection of one's privacy. Moreover, some variants that predict drug response are also markers for disease predisposition. For example, the apolipoprotein E4 allele known to influence response to cholesterol-lowering (statin) is also associated with an increased risk of Alzheimer's disease, which may subsequently lead to medico-legal implications, such as the issue of data confidentiality and the possibility of stigmatization: whether

employers and insurance companies should be given rights to asses the genetic data and the chance of the information falling into the hands of unauthorized parties. In addition, a positive result for a genetic determinant underlying therapeutic failure for a critical illness may inflict additional emotional trauma and dampen the willpower of the patient to combat the disease [8].

Nevertheless, the debate on the biological basis of race and ethnicity and pharmacogenetics may provide a useful understanding of ethnic and racial differences. Even in this case, however, we should not ignore several important parameters such as diet, economic, environmental, and psychosocial factors. However, pharmacogenetic studies on race and ethnicity are worthwhile because they are useful indicators of genetic variation. However, this kind of race and ethnicity classification for medical treatment could lead to discrimination [122].

## 12 TECHNOLOGICAL ASPECTS

Mutation screening technologies can generally be categorized into mass screening for novel variants and specific genotyping approaches. Although the former approach is more often adopted in academic research to uncover novel mutations and unravel functional consequences, the latter is more suited for practical diagnostic purposes. Rapid, precise, and cost-effective high-throughput technological platforms are essential for performing large-scale mutational analysis of genetic markers. However, genotyping techniques have generally been laborious in nature, rendering large-scale analysis time-consuming and inefficient from a cost perspective. However, SNP detection technologies have recently evolved to some of the most highly automated, robust, and affordable methods in biomedical research. Genotyping is often performed in conjunction with phenotyping (e.g., pharmacometabonomics) [8].

It is not the task of this chapter to discuss the advantages or disadvantages of the different technological platforms and bioinformatics tools in detail. Only a short presentation is possible (for review, see Ref. 8). Commercially provided services, such as Signature Genetics (http://www.signaturegenetics.com) are based on the analysis of integrated results from a detailed genetic test and comprehensive questionnaire. The report addresses the efficacy and toxicity of medications, potential drug interactions, and customized information on nutrition and recommended lifestyle modifications. Another example of a personalized medicine company (DxS, Manchester, U.K.) is focused on SNP testing, haplotyping, and clinical genotyping. DxS has applied the Amplification Refractory Mutation System and Scorpions (a homogenous fluorescent PCR detection system) technologies to the development of a highly sensitive oncology test panel [123].

The Roche AmpliChip CYP Genotyping test is FDA-approved and combines Roche's PCR amplification technology and Affymetrix high-density microarray technology to allow rapid, simultaneous analysis of multiple SNPs within the CYP family. This genotyping strategy relies on the hybridization of complementary fluorescent-tagged DNA sequences to an array of sequence-specific oligonucleotide probes (http://www.roche-diagnostics.com/press_releases/archive/2003_06_25.html). Drug MEt is another microarray-based pharmacogenetic test used for simultaneous detection of 29 SNPs of CYP and phase II enzymes involved in drug

metabolism (http://arrayit.com/Products/Microarray/DrugMEt/drugmet.html). Invader UGT1A1 Molecular Assay is an *in vitro* diagnostic test for genotyping UGT1A1 alleles and is FDA-approved. This assay appears to be an accurate method for the rapid detection of UGT1A1 polymorphisms (http://www.ons.org/fda/documents/FDA93995insert.pdf).

The TRUGENE Human Immunodeficiency Virus (HIV-1) Genotyping Kit and OpenGene DNA Sequencing System is yet another example illustrating the use of genetic testing in personalized medicine. TRUGENE is a sequence-based assay designed for detecting HIV genomic mutations (in the protease and part of the reverse transcriptase regions of HIV) that confer resistance to certain antiretroviral drugs. The assay has been shown to be robust, reproducible, and accurate and is considered a significant advance in the treatment of HIV infection [124].

New technologies like matrix-assisted laser desorption ionization-time-of-flight (MALDI-TOF) mass spectrometry (MS) and GOOD assay (requires no purification steps) have to bring in wider use. The next challenge is the demand for more accurate, economical, and large-scale technologies for SNP association studies. There are several review papers describing the technological platforms in pharmacogenetic research [8, 125–127].

The technological challenges for pharmacoproteomics are exceptional (see Table 3). Protein microarrays are an emerging class of nanotechnology for tracking many different proteins simultaneously. However, translation into the medical practice is very slow. On the other hand, proteomic changes in cultured cell lines might not fully reflect pharmacodynamic interactions because of the lack of the tissue microenvironment [11].

Although the molecular genotyping and phenotyping techniques are well established in major research institutes, the facilities for genetic testing and measurement of parent and metabolic concentrations are not always accessible in the diagnostic laboratory. Furthermore, mutational screening using the current state-of-the-art technology is still laborious and time-consuming. The hassle of having to courier samples to an external laboratory and the turnaround time for sample processing diminish the feasibility of adopting the approach in the fast-paced healthcare setting.

**TABLE 3. The Hurdles for Protein Compared with Nucleic Acid Analysis in Human Tissue [11]**

- More than 1 million proteins estimated (compared to 22.000 human genes)
- More than 300 posttranslational modifications known (phosphorylation, glycosylation)
- Wide dynamic range for protein abundances: $10^{10}$ (for mRNA:$10^4$)
- No protein amplification method available (comparable to PCR in genomics)
- Specific detection methods (only a few very good antibodies available)
- Senstivity (fM to aM needed, like ELISA)

PCR, polymerase chain reaction
ELISA, enzyme linked immunosorbent assay
fM, femtomolar ($10^{-15}$ molar)
aM, attomolar ($10^{-18}$ molar)

## 13 LIMITATIONS

A big problem limiting the progress of the pharmacogenetic approach is the fidelity of genotyping results and the confidence in associating SNPs with altered drug response. The ambiguity that sometimes develops in classifying an individual's genotype based on the laboratory results is another important contributory factor. In addition, the phenotyping method may give rise to false-positive results because the metabolic ratio can be influenced by other factors, such as epigenomic signaling, concomitantly administered drugs, nutritional state, and general health, hence affecting the validity of genotype-phenotype correlations [8].

The exact association between many SNPs of the drug target genes with therapeutic outcome is still unclear. However, few mutations have been characterized to ascertain their potential functional severity and limited definitive functional correlates established. As such, the functionality of these SNPs and their causative role remain largely speculative. In cases of functionally characterized SNPs, care should be exercised when translating research findings into an investigative tool, particularly in the extrapolation of observations from *in vitro* studies to a physiological effect. Even if the pharmacokinetic parameters are altered by genetic variants, the impact on pharmacodynamic or therapeutic effects may not be apparent. No standard guidelines exist on how the dosages of a drug should be adjusted, as a specific drug target may affect its panel of substrates to different extents, which increases the complexity of drug prescription because the dose adjustments differ among the various substrates in individuals carrying the same gene mutation [8].

Clinical use of pharmacogenetic testing has been severely limited by a lack of prospective clinical trials. Such trials are required to establish that pharmacogenetic testing benefits the selection of the appropriate drug and dose for the individual patient, thereby improving therapeutic responses or reducing ADRs. One key point that will affect the integration of pharmacogenetics into clinical practice will be the cost-effectiveness of these approaches, which may be influenced by several factors. Drugs with a narrow therapeutic index with more severer and expensive side effects are ideal candidates for pharmacogenetic testing. Drugs for which there are no established methods for monitoring adverse events (e.g., methotrexate) are also best-suited for pharmacogenomic analyses. However, for such approaches to be cost-effective, a well-established association should exist between genotype and clinical phenotype, and the frequency of the variant gene should be high. For example, if the frequency of a vriant allele is only 0.5%, then ~200 patients will have to be tested to identify one patient with the variant allele. Similarly, the strength of association between genotype and clinical phenotype will be important [128].

## 14 EDUCATIONAL ASPECTS

The resistance in the medical community to switching from the "trial-and-error" treatment approach to the gene-based approach is still prominently evident. Physicians in clinical practice, trainee physicians, and medical undergraduates have not been adequately educated in the field; the concept of pharmacogenetics has not been incorporated in the curriculum of medical courses worldwide. They are, thus,

neither well-versed in the selection of target genes for ordering a genetic test nor equipped with the knowledge to interpret and analyze the report. Thus, the bridging of the gap between basic science and medicine requires the collaborative efforts of researchers and clinicians.

Professionals in medicine and the life sciences must be prepared to adapt to this new approach. However, systems-based pharmacogenomics is unlikely to be ready for clinical application in the near future. To benefit patients today, the already available options of pharmacogenetics should be carefully implemented in clinical practice as soon as possible. Teaching the current, continuously updated knowledge of pharmacogenomics should not be postponed until the new paradigm arrives [19].

## 15 CONCLUDING REMARKS

The well-known interindividual differences in drug response could be caused by genetic and environmental factors and by the dose-response curve of a given drug. Knowledge of the individual genetic variability in drug response is, therefore, clinically and economically very important. Pharmacogenetics (focus is on single genes) and pharmacogenomics (focus is on many genes) are the two recent developments to investigate interindividual variations of drug response. This type of genetic profiling of the population doubtless provides benefits for future medical care by predicting the individual drug response.

The field of pharmacogenetics/pharmacogenomics has seen exciting advances in the recent past. The Human Genome Project and International HapMap projects have uncovered a wealth of information for researchers. The discovery of clinically predictive genotypes (e.g., UGT1A1*28 for irinocetan therapy; TPMT alleles for avoiding severe ADRs if patients receive standard doses of mercaptopurine and azathiopurine; TYMS TSER for treatment of rectal cancer; HER2 for optimizing the trastuzumab treatment in mammary Ca patients; BCA1, BRCA2, and factor V Leiden for improvement of oral contraception as well as hormone replacement), haplotypes (e.g., VKORC1 haplotype A for individualization of warfarin therapy), and somatic mutations (e.g., epidermal growth factor receptor for tailoring of tyrosine kinase inhibitors), along with the introduction of FDA-approved pharmacogenetic tests (UGT1A1*28) and the initiation of a genotype-guided clinical trial for cancer therapy (TYMS TSER in rectal cancer) have provided the first steps toward the integration of pharmacogenomics into clinical practice [22].

The translation from population-based (one dose fits all) to personalized medicine in the clinical setting is progressing at an incredibly slow pace. But why?

Several issues and problems need to be considered and solved before pharmacogenetics can be fully integrated into clinical practice (and also into drug development in the pharmaceutical industry). The ideal pharmacogenetic assay would quickly, accurately, and inexpensively provide composite genotypes for an individual patient to allow selection of the most suitable drug for the patient. Today, some approaches including suitable assays are very successful and have reached the level of clinical routine methods. Most other approaches (some are presently under investigation) are not mature. However, ongoing research is sure to bring one of the promises of the human genome project to fruition soon, that being

individualized drug therapy. We should always keep in mind that, although in some cases polymorphism in a gene is associated with poor efficacy and adverse drug reactions, in many cases the clinical relevance remains to be understood or is irrelevant. Therefore, pharmacogenetics/pharmacogenomics may not be applicable to all diseases and all treatments [129]. However, the many pharmacogenomic complexities, and particularly time-dependent changes of gene expression, will never allow personalized medicine to become an error-free entity.

The suggestion that individuals will be genotyped at birth and their "HapMap genotype" carried lifelong as an implantable identity may be too Orwellian for some. However, it may be close to the reality of clinical pharmacology practice in decades to come. By the way, we will have the data needed to extract the genetic risk from a genome and effectively model dosages and the risks of adverse reactions based on an individual's genotype, which is further than most imagined electronic prescription would go, but it is an attractive prospect. Pharmacogenomics has a long way to go to achieve this goal, but its potential is clear enough [130].

However, at present, pharmacogenomics' practical impact on medicine is more or less minimal, and the greatest challenge is to understand the genotype-environmental factor interactions, extensive geographic variations in genes (ethnicity) and to optimize study design for the accuracy, high level of quality, and consistency of technologies [131].

In comparison with pharmacogenomics, the successful transition of proteomic technologies (including novel nanotechnology strategies) from research tools to integrated diagnostic platforms will require much more effort and much more time.

After the great enthusiasm about mastering of the Human Genome Project, at present, we are in the post-genetics skepticism, which should not block our efforts for bringing pharmacogenomics and pharmacoproteomics into the clinical practice, which, however, is a long and interesting road. Like other methods in medicine, pharmacogenomics as well as pharmacoproteomics will optimize only certain parts of pharmacological therapy, not the whole field of clinical pharmacology. Taken together, prescription genotyping is only intended to aid the doctor in making individualized therapeutic decisions and is not a substitute for a physician's judgment and clinical experience.

## REFERENCES

1. Lazarou J, Pomeranz B H, Corey P N (1998). Incidence of adverse drug reactions in hospitalized patients: a meta-analysis of prospective studies. *JAMA*. 279:1200–1205.
2. Dorman H, Neubert A, Criegee-Rieck M, et al. (2004). Readmissions and adverse drug reactions in internal medicine: the economic impact. *J. Int. Med.* 255:653–663.
3. Zheng C J, Sun L Z, Han L Y, et al. (2004). Drug ADME associated protein data base as a resource for facilitating pharmacogenomics research. *Drug. Dev. Res.* 62:134–142.
4. Eichelbaum M, Ingelman-Sundberg M, Evans W E (2006). Pharmacogenomics and individualized drug therapy. *Annu. Rev. Med.* 57:119–137.
5. Beutler E (1969). Drug induced haemolytic anaemia. *Pharmacol. Rev.* 21:73–103.

6. Vogel F (1959). Moderne Probleme der Humangenetik. *Ergebn Inn Med Kinderheilk*. 12:52–125.
7. Kalow W (1962). *Pharmacogenetics, Heredity and the Response to Drugs*. WB Saunders, Philadelphia, PA.
8. Koo S H, Lee E D (2006). Pharmacogenetics approach to therapeutics. *Clin. Exp. Pharmacol. Physiol.* 33:525–532.
9. Webster A, Martin P, Lewis G, et al. (2004). Integrating pharmacogenetics into society: In search of a model. *Nat. Rev. Genet.* 5:663–669.
10. Lin M, Aquilante C, Johnson J A, et al. (2005). Sequencing drug response with HapMap. *Pharmacogenom. J.* 5:149–156.
11. Becker K-F, Metzger V, Hipp S, et al. (2006). Clinical proteomics: New trends for protein microarrays. *Curr. Med. Chem.* 13:1831–1837.
12. Calvo K R, Liotta L A, Petricoin E F (2005). Clinical proteomics: From biomarker and cell signalling profiles to individualized personal therapy. *Biosci. Reports*. 25:107–125.
13. Besada V, Diaz M, Becker M, et al. (2006). Proteomics of xenografted human breast cancer indicates novel targets related to tamoxifen resistance. *Proteom.* 6:1038–1048.
14. Haselden J N, Nicholls A W (2006). Personalized medicine in progress. *Nat. Med.* 12:510–511.
15. Clayton T A, Lindon J C, Cloarec O, et al. (2006). Pharmaco-metabonomic phenotyping and personalized drug treatment. *Nature.* 440:1073–1077.
16. Arnold H P, McHale D (2006). Pharmacogenetics: Developmental issues and solutions for safe and effective medicines. *Pharmacogen.* 7:149–155.
17. Bosch T M, Meijerman I, Beijnen J H, et al. (2006). Genetic polymorphisms of drug-metabolising enzymes and drug transporters in the chemotherapeutic treatment of cancer. *Clin. Pharmacokinet.* 45:253–285.
18. Kalow W (2006). Pharmacogenetics and pharmacogenomics: Origin, status, and the hope for personalized medicine. *Pharmacogen. J.* 6:162–165.
19. Lunshof J (2006). Teaching and practicing pharmacogenomics: A complex matter. *Pharmacogen.* 7:243–246.
20. Lunshof E J, Pirmohamed M, Gurwitz D (2006). Personalized medicine: Decades away? *Pharmacogen.* 7:237–241.
21. Mahlknecht U (2006). Pharmakogenomik: Prinzip and Perspektive. *Dtsch Med Wschr.* 131:310–313.
22. Marsh S, McLeod H L (2006). Pharmacogenomics: From bediside to clinical practice. *Human Molec. Genet.* 15:R89–R93.
23. Mendrick D L, Brazell C, Mansfield E A, et al. (2006). Pharmacogenomics and regulatory decision making: an international perspective. *Pharmacogen. J.* 6:154–157.
24. Shabo A (2006). Clinical genomics data standards for pharmacogenetics and pharmacogenomics. *Pharmacogen.* 7:247–253.
25. Shastry B S (2006). Pharmacogenetics and the concept of individualized medicine. *Pharmacogen. J.* 6:16–21.
26. Wilkinson G R (2005). Drug metabolism and variability among patients in drug response. *N. Engl. J. Med.* 352:2211–2221.
27. Emoto C, Iwasaki K I (2006). Enzymatic characteristics of CYP3A5 and CYP3A4: A comparison of *in vitro* kinetic and drug-drug interaction patterns. *Xenobio.* 36:219–233.

28. Xie H-G, Kim R B, Wood A J J, et al. (2001). Molecular basis of ethnic differences in drug disposition and response. *Ann. Rev. Pharmacol. Toxicol.* 41:815–850.
29. De Leon J, Susce M T, Pan R M, et al. (2005). The CYP2D6 poor metabolizer phenotype may be associated with risperidone adverse drug reactions and discontinuation. *J. Clin. Psychiatry.* 66:15–27.
30. Voora D, Eby C, Linder M W, et al. (2005). Prospective dosing of warfarin based on cytochrome P450 2C9 genotype. *Thromb. Haemost.* 93:7000–7005.
31. Rieder M J, Reiner A P, Gage B F, et al. (2005). Effect of VKORC1 haplotypes on transcriptional regulation and warfarin dose. *N. Engl. J. Med.* 352:1185–2293.
32. Guo Y, Zhang Y, Wang Y, et al. (2005). Role of CYP2C9 and its variants (CYP2C9*3) and CYP2C9*13) in the metabolism of lornoxicam in humans. *Drug. Metab.* 33:749–753.
33. Totah R A, Rettie A E (2005). Cytochrome P450 2C8: substrates, inhibitors, pharmacogenetics and clinical relevance. *Clin. Pharmacol. Ther.* 77:341–352.
34. Hesselink D A, van Gelder T, van Schaik R H (2005). The pharmacogenetics of calcineurin inhibitors: One step closer toward individualized immunosuppression? *Pharmacogen.* 6:323–337.
35. Rogausch A, Brockmoller J, Himmel W (2005). Pharmacogenetics in future medical care—implications for patients and physicians. *Gesundheitswesen.* 67:257–263.
36. He P, Court M H, Greenblatt D J, et al. (2005). Genotype-phenotype associations of cytochrome P450 3A4 and 3A5 polymorphism with midazolam clearance *in vivo*. *Clin. Pharmacol. Ther.* 77:373–378.
37. Zhao B, Seow A, Lee E J, et al. (2000). Correlation between acetylation phenotype and genotype in Chinese women. *Eur. J. Clin. Pharmacol.* 56:689–692.
38. Goekkurt E, Hoehn S, Wolschke C, et al. (2006). Polymorphisms of glutathione S-transferase (GST) and thymidylate synthase (TS)—novel predictors for response and survival in gastric cancer patients. *Br. J. Canc.* 94:281–286.
39. Iyer L, Das S, Janisch L, et al. (2002). UGT1A1*28 polymorphism as a determinant of irinocetan disposition and toxicity. *Pharmacogen. J.* 2:43–47.
40. Balram C, Sabapathy K, Frei G, et al. (2002). Genetic olymorphisms of UDP-glucuronosyltransferase in Asians: UGT1A1*28 is a common allele in Indians. *Pharmacogen.* 12:81–83.
41. Kalow W, Grant D M. Pharmacogenetics. In C R Scriver, A L Beaudet, W S Sly, (eds.), *The Metabolic and Molecular Basis of Inherited Disease*. McGraw-Hill, New York, pp. 293–326.
42. Lepper E R, Nooter K, Verweij J, et al. (2005). Mechanism of resistance to anticancer drugs: The role of the polymorphic ABS transporters ABCB1 and ABCG2. *Pharmacogenom.* 6:115–138.
43. Kafka A, Sauer G, Jaeger C, et al. (2003). Polymorphism C3435T of the MDR-1 gene predicts response to properative chemotherapy in locally advanced breast cancer. *Int. J. Oncol.* 22:1117–1121.
44. Chowbay B, Li H, David M, et al. (2005). Meta-analysis of the influence of MDR1 C3435T polymorphism on digoxin pharmacokinetics and MDR1 gene expression. *Br. J. Clin. Pharmacol.* 60:159–171.
45. Hauser I A, Schaeffeler E, Gauer S, et al. (2005). ABCB1 genotype of the donor but not of the recipient is a major risk factor for cyclosporine-related nephrotoxicity after renal transplantation. *J. Am. Soc. Nephrol.* 16:1501–1511.
46. Chowbay B, Cumaraswamy S, Cheung Y B, et al. (2003). Genetic polymorphisms in MDR1 and CYP3A4 genes in Asians and the influence of MDR1 haplotypes on cyclosporine disposition in heart transplant recipients. *Pharmacogen.* 13:89–95.

47. Hulot J S, Villard E, Maguy A, et al. (2005). A mutation in the drug transporter gene ABCC2 associated with impaired methotrexate elimination. *Pharmacogenet. Genomics.* 15:277–285.
48. Lim A Y, Gaffney K, Scott D G (2005). Methotrexate-induced pancytopenia: serious and under reported? Our experience of 25 cases in 5 years. *Rheumatol.* 44:1051–1055.
49. Sills G J, Mohanraj R, Butler E, et al. (2005). Lack of association between the C3435T polymorphism in the human multidrug resistance (MDR1) gene and response to antiepileptic drug treatment. *Epilepsia.* 46:643–647.
50. Gwee P C, Tang K, Chua J M, et al. (2003). Simultaneous genotyping of seven single-nucleotide polymorphisms in the MDR1 gene by single-tube multiplex minisequencing. *Clin. Chem.* 49:672–676.
51. König J, Seithel A, Gradhand U, et al. (2006). Pharmacogenomics of human OATP transporters. *Naunyn-Schmiedeberg's Arch. Pharmacol.* 372:432–443.
52. Erdman A R, Mangravite L M, Urban T J, et al. (2006). The human organic anion transporter 3 (OAT3; SLC22A8): Genetic variation and functional genomics. *A. J. Physiol. Renal. Physiol.* 290:F905–F912.
53. Tate S K, Depondt C, Sisodiya S M, et al. (2005). Genetic predictors of the maximum doses patients receive during clinical use of the anti-epileptic drugs carbamazepine and phenytoin. *Proc. Natl. Acad. Sci. USA.* 102:5507–5512.
54. Paulussen A D, Gilissen R A, Armstrong M, et al. (2004). Genetic variations of KCNQ1, KCNH2, SCN5A, KCNE1, and KCNE2 in drug-induced long QT syndrome patients. *J. Mol. Med.* 82:182–188.
55. Tellería J J, Blanco-Quirós A, Muntión S, et al. (2006). Tachyphylaxis to $\beta_2$-agonists in Spanish asthmatic patients could be modulated by $\beta_2$-adrenoceptor gene polymorphisms. *Reg. Med.* 100:1072–1078.
56. Wang L, Nguyen T V, McLaughlin R W, et al. (2005). Human thiopurine S-methyltransferase pharmacogenetics: Variant allozyme misfolding and aggresome formation. *Proc. Natl. Acad. Sci. USA.* 102:9394–9399.
57. Stanulla M, Schaeffeler E, Flohr T, et al. (2005). Thiopurine methyltransferase (TPMT) genotype and early treatment response to mercaptopurine in childhood acute lymphoblastic leukaemia. *JAMA.* 293:1485–1489.
58. McLeod H L, Krynetski E Y, Relling M V, et al. (2000). Genetic polymorphism of thiopurine methyltransferase and its clinical relevance for childhood acute lymphoblastic leukaemia. *Leukemia.* 14:567–572.
59. Schaeffeler E, Fischer C, Brockmeier D, et al. (2004). Comprehensive analysis of thiopurine S-methyltransferase (TPMT) phenotype-genotype correlation in a large population of German-Caucasians and identification of novel TPMT variants. *Pharmacogen.* 14:407–417.
60. Gearry R B, Barclay M L (2005). Azathioprine and 6-mercaptopurine pharmacogenetics and metabolite monitoring in inflammatory bowel disease. *Gastroenterol. Hepatol.* 20:1149–1157.
61. Yates C R, Krynetski E Y, Loennechen T, et al. (1997). Molecular diagnosis of thiopurine *S*-methyltransferase deficiency: Genetic basis for azathioprine and mercaptopurine intolerance. *Ann. Intern. Med.* 126:608–614.
62. Relling M V, Rubnitz J E, Rivera G K, et al. (1999). High incidence of secondary brain tumors related to radiation and antimetabolite therapy. *Lancet.* 354:34–39.
63. Etienne M C, Formento J L, Chazal M, et al. (2004). Methylenetetrahydrofolate reductase gene polymorphisms and response to fluorouracil-based treatment in advanced colorectal cancer patients. *Pharmacogen.* 14:785–792.

64. Villafranca E, Okruzhnov Y, Dominguez M A, et al. (2001). Polymorphisms of the repeated sequences in the anhancer region of the thymidilate synthase gene promoter may predict downstaging after properative chemoradiation in rectal cancer. *J. Clin. Oncol.* 19:1779–1786.
65. Lynch T J, Bell D W, Sordella R, et al. (2004). Activating mutations in the epidermal growth factor receptor underlying responsiveness of non-small-cell lung cancer to gefitinib. *N. Engl. J. Med.* 350:2129–2139.
66. Villuendes R, Steegmann J L, Pollán M, et al. (2006). Identification of genes involved in imatinib resistance in CML: a gene-expression profiling approach. *Leukemia.* 20:1047–1054.
67. Couzin J (2004). Pharmacogenomics: Cancer sharpshooters rely on DNA tests for a better aim. *Science.* 305:1222–1223.
68. McNamara D M (2004). Pharmacogenetics in heart failure: Genomic markers of endothelial and neurohumoral function. *Congest. Heart Failure.* 10:302–308.
69. Maitland-van der Zee A H, Turner S T, Schwartz G L, et al. (2005). A multilocus approach to the antihypertensive pharmacogenetics of hydrochlorothiazide. *Pharmacogenet. Genomics.* 15:287–293.
70. Liljedahl U, Kahan T, Malmqvist K, et al. (2005). Single nucleotide polymorphisms predict the change in left ventricular mass in response to antihypertensive treatment. *J. Hyperten.* 22:2273–2275.
71. Mellen P B, Herrington D M (2005). Pharmacogenomics of blood pressure response to antihypertensive treatment. *J. Hyperten.* 23:1311–1325.
72. King D, Etzel J P, Chopra S, et al. (2005). Human response to alpha 2-adrenergic agonist stimulation studied in an isolated vascular bed *in vivo*: Biphasic influence of dose, age, gender and receptor genotype. *Clin. Pharmacol.* 77:388–403.
73. Terra S G, Pauly D F, Lee C R, et al. (2005). Beta-adrenergic receptor polymorphisms and responses during titration of metoprolol controlled release/extended release in heart failure. *Clin. Pharmacol. Ther.* 77:127–137.
74. Taylor M R, Bristow M R (2004). The emerging pharmacogenomics of the beta-adrenergic receptors. *Congest. Heart Failure.* 10:281–288.
75. Hubacek J A, Bobkova D (2006). Role of cholesterol 7α-hydroxylase (CYP7A1) in nutrigenetics and pharmacogenetics of cholesterol lowering. *Mol. Diag. Ther.* 10:93–100.
76. Vadlamani L, Lyenger S (2004). Tumor necrosis factor alpha polymorphism in heart failure/cardiomyopathy. *Congest. Heart Failure.* 10:289–292.
77. Tantisira K G, Small K M, Litonjua A A, et al. (2005). Molecular properties and pharmacogenetics of a polymorphism of adenylyl cyclase type 9 in asthma: interaction between beta-agonist and corticosteroid pathways. *Hum. Mol. Genet.* 14:1671–1677.
78. Kukreti R, Bhatnagar P, B-Rao C, et al. (2005). Beta (2)-adrenergic receptor polymorphisms and response to salbutamol among Indian asthmatics. *Pharmacogen.* 6:399–410.
79. Kirchheiner J, Nickchen K, Bauer M, et al. (2004). Pharmacogenetics of antidepressants and antipsychotics: The contribution of allelic variations to the phenotype of drug response. *Mol. Psychiat.* 9:442–473.
80. Serrette A, Artioli P, Quartesan R (2005). Pharmacogenetics in the treatment of depression: pharmacodynamic studies. *Pharmacogenet. Genomics.* 15:61–67.
81. Sink K M, Holden K F, Yaffe K (2005). Pharmacological treatment of neuropsychiatric symptoms of dementia. *JAMA.* 293:596–608.

82. Rybakowski J K, Suwalska A, Czerski P M, et al. (2005). Prophylactic effect of lithium in bipolar affective illness may be related to serotonin transporter genotype. *Pharmacol. Rep.* 57:124–127.
83. Stein M S, Seedat S, Gelernter J (2006). Serotonin transporter gene promotor polymorphism predicts SSRI response in generalized social anxiety disorder. *Psychopharmacol.* 187:68–72.
84. Popp J, Leucht S, Heres S, Steimer W (2006). Serotonin transporter polymorphisms and side effects in antidepressant therapy—a pilot study. *Pharmacogen.* 7:159–166.
85. Reynolds G P, Yao Z, Zhang X, et al. (2005). Pharmacogenetics of treatment in first-episode schizophrenia: D3 and 5-HT2C receptor polymorphisms separately associate with positive and negative symptom response. *Eur. Neuropsychopharmacol.* 15: 143–151.
86. Wilffert B, Zaal R, Brouwers J R (2005). Pharmacogenetics as a tool in the therapy of schizophrenia. *Pharm. World Sci.* 27:20–30.
87. Ferraro T N, Buono R J (2005). The relationship between pharmacology of antiepileptic drugs and human gene variation: An overview. *Epilepsy Behav.* 7:18–36.
88. Hung S L, Chung W-H, Jee S-H, et al. (2006). Genetic susceptibilty to carbamazepine-induced cutaneous adverse drug rections. *Pharmacogen. Genomics.* 16:297–306.
89. Soranzo N, Goldstein D B, Sisodiya S M (2005). The role of common variation in drug transporter gene in refractory epilepsy. *Expert. Opin. Pharmacother.* 6:1305–1312.
90. Linazasoro G (2005). New ideas on the origin of l-Dopa-induced dyskinesias: Age, genes and neural plasticity. *Trends Pharmacol. Sci.* 26:391–397.
91. Stamer U M, Bayerer B, Stuber F (2005). Genetics and variability in opioid response. *Eur. J. Pain.* 9:101–104.
92. Kreek M J, Bart G, Lilly C, et al. (2005). Pharmacogenetics and human molecular genetics of opiate and cocaine addictions and their treatments. *Pharmacol. Rev.* 57:1–26.
93. Gourlay G K (2005). Advances in opioid pharmacology. *Support Care Cancer.* 13:153–159.
94. Mogil J S, Ritchie J, Smith S B, et al. (2005). Melanocortin-1 receptor gene variants affect pain and (micro)-opioid analgesin in mice and humans. *J. Med. Genet.* 42:583–587.
95. Rosenfeld R G (2006). Editorial: The Phasrmacogenomics of Human Growth. *J. Clin. Endocrinol. Metab.* 91:795–796.
96. Dos Santos C, Essioux L, Teinturier C, et al. (2004). A common polymorphism of the growth hormone receptor is associated with increased responsiveness to growth hormone. *Nat. Genet.* 36:720–724.
97. Jorge A A L, Marchisotti F G, Montenegro L R, et al. (2006). Growth hormone (GH) pharmacogenetics: Influence of GH receptor exon 3 retention or deletion on first-year growth response and final height in patients with severe GH deficiency. *J. Clin. Endocrinol. Metab.* 91:1076–1080.
98. Andreassi M G, Botto N, Maffei S (2006). Factor V Leiden, prothrombin G20210A substitution and hormone therapy: Indications for molecular screening. *Clin. Chem. Lab. Med.* 44:514–521.
99. Sade R B B, Chetrit A, Figer A, et al. (2006). Hormone replacement therapy is more prevalent among Jewish BRCA1/2 mutation carriers. *Eur. J. Cancer.* 42:650–655.
100. Cribb A E, Knight M J, Dryer D, et al. (2006). Role of polymorphic human cytochrome P450 enzymes in estrone oxidation. *Cancer Epidemiol. Biomarkers Prev.* 15:551–558.

101. Yamamoto Y, Moore R, Hess H A, et al. (2006). Estrogen receptor α mediates 17α-ethynylestradiol causing hepatotoxicity. *J. Biol. Chem.* 281:16625–16631.
102. Pooley K A, Healey C S, Smith P L, et al. (2006). Association of the progesterone receptor gene with breast cancer risk: A single-nucleotide polymorphism tagging approach. *Cancer Epidemiol. Biomarkers Prev.* 15:675–682.
103. Jakobsson J, Ekström L, Inotsume N, et al. (2006). Large differences in testosterone excretion in Korean and Swedish men are strongly associated with a UDP-glucuronosyl transferase 2B17 polymorphism. *J. Clin. Endocrinol. Metab.* 91:687–693.
104. Traub F, Hess R, Schorn K, et al. (2006). Peptidomic analysis of breast cancer reveals a putative surrogate marker for estrogen receptor-negative carcinomas. *Lab. Invest.* 86:246–253.
105. Zhang L, Liu X, Zhang J, et al. (2006). Proteome analysis of combined effects of androgen and estrogen on the mouse mammary gland. *Proteomics.* 6:487–497.
106. Sesti G, Laratta E, Cardellini M, et al. (2006). The E23K variant of KCNJ11 encoding the pancreatic β-cell adenosine 5′-triphosphate-sensitive potassiuj channel subunit Kir&.2 is associated with an increased risk of secondary failure to sulfonylurea in patients with type 2 diabetes. *J. Clin. Endocrinol. Metab.* 91:2334–2339.
107. Krishnaswamy S, Hao Q, Al-Rohaimi A, et al. (2005). UDP Glucuronosyltransferase (UGT) 1A6 pharmacogenetics: identification of polymorphisms in the 5′-regulatory and exon 1 regions, and association with human liver UGT1A6 gene expression and glucuronidation. *Pharmacol. Exp. Ther.* 313:1331–1339.
108. Krishnaswamy S, Hao Q, Al-Rohaimi A, et al. (2005). UDP glucuronosyltransferase (UGT) 1A6 pharmacogenetics: functional impact of the three most common non-synonymous UGT1A6 polymorphisms (S7A, T181A, and R184S) *Pharmacol. Exp. Ther.* 313:1340–1346.
109. Dahl J P, Doyle G A, Oslin D W, et al. (2005). Lack of association between single nucleotide polymorphisms in the corticotrophin releasing hormone receptor 1 (CRHR 1) gene and alcohol dependence. *J. Psychiatr. Res.* 39:475–479.
110. Daar A S, Singer P A (2005). Pharmacogenetics and geographical ancestry: Implications for drug development and global health. *Nat. Rev. Genet.* 6:241–246.
111. Rahemtulla T, Bhopal R (2005). Pharmacogenetics and ethnically targeted therapies. *BMJ.* 330:1036–1037.
112. Chowbay B, Zhou S, Lee E J (2005). An interethnic comparison of polymorphisms of the genes encoding drug-metabolizing enzymes and drug transporters: Experience in Singapore. *Drug Metab. Rev.* 37:327–378.
113. Mori M, Yamada R, KLobayashi K, et al. (2005). Ethnic differences in allele frequency of autoimmune disease-associated SNPs. *J. Hum. Genet.* 50:264–266.
114. Mitchell A A, Chakravarti A, Cutler D J (2005). On the probability that a novel variant is a disease-causing mutation. *Genome Res.* 15:960–966.
115. Kaplan J B, Bennett T (2003). Use of race and ethnicity in biomedical publication. *JAMA.* 289:2709–2716.
116. Freudenberg-Hua Y, Freudenberg J, Winantea J, et al. (2005). Systematic investigation of genetic variability in 111 human genes—implication for studying variable drug response. *Pharmacogen. J.* 5:183–192.
117. Hall A M, Wilkins M R (2005). Warfarin: A case history in pharmacogenetics. *Heart.* 9:563–564.
118. Yu C M, Chan T Y K, Critchley J A J H, et al. (1996). Factors determining the maintenance dose of warfarin in Chinese patients. *Q. J. Med.* 89:127–135.
119. Yu C M, Chan T Y, Tsoi W C, et al. (1997). Heparin therapy in the Chinese—lower doses are required. *Q. J. Med.* 90:535–543.

120. Taylor A L, Zieshe S, Yancy C, et al. (2004). Combination of isosorbide dinitrate and hydralazine in blacks with heart failure. *N. Engl. J. Med.* 351:2049–2057.
121. Tang K, Ngoi S M, Gwee P C, et al. (2002). Distinct haplotype profiles and strong linkage disequilibrium at the MDR1 multidrug transporter gene locus in three ethnic Asian populations. *Pharmacogen.* 12:437–450.
122. Harty L, Johnson K, Power A (2006). Race and ethnicity in the era of emerging pharmacogenomics. *J. Clin. Pharmacol.* 46:405–407.
123. Little S (2003). DxS Ltd. *Pharmacogen.* 4:97–101.
124. Jagodzinski L L, Colley J D, Weber M, et al. (2003). Performance characteristics of human immunodeficiency virus type 1 (HIV-1) genotyping systems in sequence-based analysis of subtypes other than HIV-1 subtype B. *J. Clin. Microbiol.* 41:998–1003.
125. Shi M M (2001). Enabling large-scale pharmacogenetic studies by high-throughput mutation detection and genotyping technologies. *Clin. Chem.* 47:164–172.
126. Jannetto P J, Laleli-Sahin E, Wong S H (2004). Pharmacogenomic genotyping methodologies. *Clin. Chem. Lab.* 42:1256–1264.
127. Koch W H (2004). Technology platforms for pharmacogenomic diagnostic assays. *Nat. Rev. Drug Discov.* 3:749–761.
128. Ranganathan P, McLeod H L (2006). *Arthritis Rheumat.* 54:1366–1377.
129. Lindpaintner K (2005). Pharmacogenetics and pharmacogenomics. *Methods Mol. Med.* 108:235–260.
130. O'Shaughnessy (2006). HapMap, pharmacogenomics, and the goal of personalized prescribing. *Br. J. Clin. Pharmacol.* 61:783–786.
131. Stoughton R B, Friend S H (2005). How molecular profiling could revolutionize drug discovery. *Nat. Rev. Drug Discov.* 4:345–350.

# ENHANCED PROTEOMIC ANALYSIS BY HPLC PREFRACTIONATION

Pierre C. Havugimana, Peter Wong, and Andrew Emili
*University of Toronto, Toronto, Ontario, Canada*

**Chapter Contents**

1 Introduction
2 Materials
  2.1 Chemicals
  2.2 Yeast Protein Extraction
  2.3 IEX–HPLC Analysis
  2.4 Precipitations and Proteolysis of Proteins in IEX–HPLC Fractions
  2.5 1D-LC–MS Shotgun Tandem Mass Spectrometry Proteomic Analysis
3 Method
  3.1 HPLC Sample Preparation
  3.2 HPLC Sample Prefractionation
  3.3 LC–MS Sample and Data Analysis
References

## 1 INTRODUCTION

Proteomics is often cited as the global (genome-wide) study of protein expression in a given living cell, tissue, or organism. Over the past decade, tandem mass spectrometry (MS/MS) has emerged as an essential analytical tool for large-scale

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

proteomic analysis [1, 2]. The critical challenge is to accurately and reproducibly identify and quantify large numbers of proteins in a given sample with good overall sensitivity and dynamic range. Currently, two shotgun (peptide level) tandem mass spectrometry-based proteomic approaches are widely employed to survey protein levels in complex biological samples. The first, known as MudPIT (for Multidimensional Protein Identification Technology), was developed by Yates and colleagues as an effective brute-force method for the comprehensive shotgun sequencing of large collections of peptide products of proteolytic digests of entire proteomes [3–5], whereas the second approach, prefractionation, requires sample separations to alleviate protein complexity before MS/MS analysis. In the latter approach, two avenues have generally been explored: gel electrophoresis (GE) and high-performance liquid chromatography (HPLC) pre-fractionations [6–8].

GE is a well-documented method [9] and is systematically used around the globe even though it is tedious, time-consuming, and often difficult to achieve reproducible results. On the other hand, HPLC prefractionation techniques [10, 11] such as size exclusion chromatography, reverse phase chromatography, affinity chromatography, and anion exchange chromatography represent more robust, simple-to-execute, rapid, and ultimately more reproducible sample prefractionation platforms that are also amenable to automation. These different HPLC methods can be characterized in terms of resolution as well as whether the method is denaturing. Choice of HPLC method will depend on the separation goals, with the primary goal usually being of obtaining the maximum nonoverlapping polypeptide fractions. Reverse phase HPLC, a popular and particularly effective separation method that fractionates proteins according to hydrophobicity, is typically used if sample denaturation (and the subsequent loss in protein activity) is not a consideration [12]. Alternatively, size exclusion chromatography, which separates proteins according to molecular size, is an excellent method to recover the proteins in their native conformation, albeit with relatively low resolution. Affinity chromatography is a powerful way to isolate proteins for which a biospecific interaction to a suitable selective adsorbent is available, but it is not suitable for the study of a whole proteome. The most flexible, convenient, and "widely used" method, however, for protein fractionation in biochemical studies is ion-exchange chromatography, or IEX–HPLC. Similar to size exclusion and affinity chromatography, this method also allows for the collection of proteins in their native conformations, without significant sample loss, but maintains the maximum possible resolution suitable for effective separations of complex biological samples. Surprisingly, despite its exciting potential, only limited use of IEX–HPLC in combination with proteomic analysis has been reported to date.

In IEX–HPLC, the resolution of a mixture of proteins is performed according to the differential retention due to inherent variations in the surface charge properties of the globular proteins in solution [13]. It involves the selective binding of target proteins to a suitably charged solid phase material or adsorbent, followed by sequential elution by increasing the mobile phase salt concentration (or, alternatively, under denaturing conditions, by altering the pH) to disrupt ionic interactions. Many such resins or prepacked columns are commercially available and can be used to efficiently resolve protein mixtures in a reasonable time frame. Anion exchange columns, which consist of positively charged functional groups, can be used to separate negatively charged proteins (which represent most typical cellular soluble proteins under near physiological pH), whereas cation exchange

columns resolve positively (e.g., histidine, lysine, and arginine rich) proteins. In principle, the development of a method that combines the two types of columns should provide a more effective analytical separation tool for protein fractionations before proteomic analysis using MS/MS that could be applied to different types of biological samples.

In this chapter, we described an easily implemented, effective, and highly reproducible dual-column HPLC prefractionation method that we have developed for improving routine proteomic analyses. The approach results in multifold increases in the numbers of proteins that can be confidently identified by LC–MS of whole cell lysates without fractionation. We outline the key steps in the overall procedure, from sample preparation through to MS/MS and attendant data analysis, using yeast soluble protein extract as a test mixture.

## 2 MATERIALS

The chemicals, biologicals, equipment, and instruments listed are those routinely used in our laboratory for proteomics analyses. Substitutions could be made with equivalent items as available.

### 2.1 Chemicals

1. Tris-(hydroxymethyl)-aminomethane (TRIS base), Sodium chloride (NaCl), Trichloroacetic acid (TCA), Sodium azide ($NaN_3$), Ammonium bicarbonate ($NH_4HCO_3$), HPLC-grade water, Calcium chloride ($CaCl_2$), HPLC grade acetonitrile (ACN), and HPLC-grade glacial acetic acid (AA) were all purchased from Fisher Scientific (Whitby, ON, Canada).
2. Hydrochloric acid (6N HCl) was from VWR International (Mississauga, ON, Canada).
3. HPLC grade acetone was obtained from Sigma-Aldrich (Oakville, ON, Canada).
4. Bio-Rad dye reagent (Bio-Rad, Mississauga, ON, Canada).
5. Heptafluorobutyric acid (HFBA) was purchased from Pierce (Rockford, IL).

### 2.2 Yeast Protein Extraction

1. Standard laboratory *Saccharomyces cerevisiae* yeast strain W303 (Mata ade2-1, ura3-1, his3-11,15, trp1-1, leu2-3,112, can1-100) (Open Biosystems, Huntsville, AL).
2. Lysis Solution: 10-mM Tris-HCl, 30-mM NaCl, and 3-mM $NaN_3$ (pH 7.8) to minimize microbial growth in the buffer.
3. Protease cocktail inhibitor (Roche Diagnostic, Laval, QC, Canada).
4. Bovine serum albumin (BSA) (Sigma-Aldrich, Oakville, ON, Canada).
5. Silica beads (Fisher scientific, Whitby, ON, Canada).
6. Ice bucket.

7. Microtube vortex adaptor (placed in cold room).
8. Eppendorf centrifuge (cooled at 4°C) or placed in cold room.
9. 1.5-mL Eppendorf microcentrifuge tubes cooled at 4°C.
10. 1.0-mL plastic cuvettes (Bio-Rad, Mississauga, ON, Canada).
11. Ultrospec1100 pro UV/VIS Spectrophotometer (Biochrom, Cambridge, U.K.) or an equivalent instrument.

### 2.3 IEX-HPLC Analysis

1. 80–100-μg total soluble protein from yeast (W303) crude soluble cell lysate as measured by Bradford (approximate concentration: [$10\,\mu g \cdot \mu L^{-1}$]).
2. Buffer A (binding buffer): 10-mM Tris-HCl, pH 7.5 with 3-mM $NaN_3$.
3. Buffer B (eluting buffer): Buffer A with 600-mM NaCl.
4. 0.22-μm membrane vacuum-driven disposable bottle top filter (Millipore, Cambridge, ON, Canada).
5. Anion exchange column (PolyWAX LP, 50 × 2.1 mm i.d., 5 μm, 1000 Å), cation exchange column (PolyCAT A, 50 × 2.1 mm i.d., 5 μm, 1000 Å), anion exchange guard cartridge (PolyWaX LP, 10 × 2.1 mm i.d., 5 μm, 1000 Å), and pre-guard column sieves with high flow, 2 μm (IT9085-20-10) were from Canadian Life Science (Peterborough, ON, Canada).
6. Agilent 1100 quaternary HPLC pump system (Agilent Technologies, Mississauga, ON, Canada).
7. Microcentrifuge tubes (precooled at 4°C).

### 2.4 Precipitations and Proteolysis of Proteins in IEX–HPLC Fractions

1. 100% TCA in HPLC-grade water: dissolve 20 g of solid TCA in 9.0 mL of water (prepare fresh solution).
2. Applied Biosystems immobilized trypsin beads (Applied Biosystems, Foster City, CA).
3. Pierce immobilized TPCK-trypsin beads (Pierce, Rockford, IL).
4. Digestion buffer: 50-mM $NH_4HCO_3$, 1-mM $CaCl_2$, pH 8.0.
5. 1-M $CaCl_2$ in HPLC-grade water.
6. Ice cold HPLC-grade acetone (keep at –20°C until use).
7. Eppendorf centrifuge 5417C (14,000 rpm capacity) or its equivalent.

### 2.5 1D-LC-MS Shotgun Tandem Mass Spectrometry Proteomic Analysis

1. An LTQ linear ion trap tandem mass spectrometer (Thermo Finnigan Corp, San Jose, CA) is routinely used in our laboratory. XCalibur Software is used to automate the acquisition of tandem mass spectra over a 400–1600-m/z range and to control precursor ion selection by the instrument in a data-dependent, data-acquisition mode, with dynamic exclusion activated.
2. 100-μm capillary microcolumn silica tubing (Polymicro Technologies, Phoenix, AZ).
3. 5-μm pore size C18 reverse-phase packing material (Zorbax eclipse XDB-$C_{18}$ resin) (Agilent Technologies, Mississauga, ON, Canada).
4. Solvent A: 5% ACN, 0.5% AA, and 0.02% HFBA.

5. Solvent B: 100% ACN.
6. HPLC quaternary gradient pump, with flow splitter.

## 3 METHOD

The following method is designed to allow the routine prefractionation, leading to enhanced identification of proteins from complex sample matrices, such as a representative crude soluble cell extract prepared from a standard laboratory *S. cerevisiae* yeast strain, by IEX–HPLC and LC–MS techniques. This method can be readily customized to investigate other biological samples. Specific details regarding sample preparation, chromatography prefractionation, precipitation, and tryptic proteolysis of the protein samples followed by proteomic LC–MS analysis are included. The major steps of the entire procedure, from protein sample mixture prefractionation to subsequent analysis by LC–MS, are summarized schematically in Figure 1.

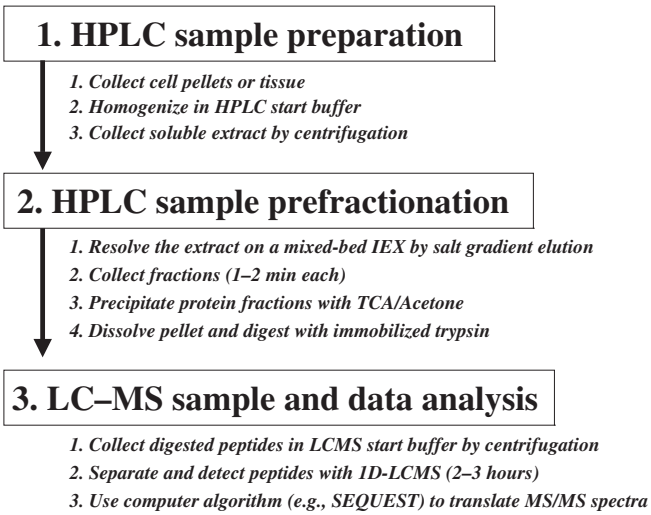


**Figure 1. Representative workflow for IEX–HPLC prefractionation of a proteome sample (e.g., the yeast *S. cerevisiae*) before proteomic analysis by LC–MS.** (1) Sample preparation is performed in IEX–HPLC starting buffer (pH 7.8) containing protease inhibitors. The soluble extract is recovered and clarified by centrifugation. (2) The sample is prefractionated using IEX–HPLC salt gradient, and a total of 27 fractions (1.3 minutes each) is collected using a dual-column system (tandem anion and cation-exchange columns placed in series) and a standard HPLC binary pump system. The collected protein fractions and a crude soluble sample lysate, as a reference control, are subjected to precipitation and tryptic digestion. (3) The peptide mixtures are separated and detected by single-dimension liquid chromatography coupled to automated tandem mass spectrometry. The resulting tandem mass spectra are computationally translated to a short amino acid sequence, and their corresponding cognate protein identity is determined using a suitable database algorithm searching against a relevant protein sequence database. High confidence protein identifications (≥90% probability) are parsed into an in-house database, and diverse data clustering and mining strategies used to find interesting patterns of protein expression for biological validation and detailed analysis. (This figure is available in full color at ftp://ftp.wiley.com/public/sci_tech_med/pharmaceutical_biotech/.)

### 3.1 HPLC Sample Preparation

To facilitate the process and to maintain reproducible results, minimal sample handling is highly recommended for success of an IEX–HPLC experiment. To meet this requirement, a one-step sample pretreatment approach is employed. This approach allows recovering the sample in a low salt buffer compatible with HPLC while preserving the native conformation of the proteins. Here, a simple Tris-HCl-based buffer system at or near the physiological pH is used to solubilize the yeast cell pellet and to collect a soluble fraction by high-speed centrifugation. The yield of this crude extract is approximately 10-mg protein/mL. The protocol below allows for collection of a protein sample mixture ready for the direct injection onto a standard analytical grade HPLC column without further treatment.

1. To approximately 500 μL of *S. cerevisiae* cell pellets on ice, add an equal volume of the fresh Lysis Buffer Solution (described in Section 2.2, prefiltered through a 0.22-μm membrane) containing protease inhibitors (follow the vendor's instructions for preparation) and vortex extensively to fully resuspend the pellet.
2. Allow the mixture to stand for 1 minute on ice and add a half volume of glass beads (stored at 4°C) for cell disruption.
3. Insert tube into a vortex adaptor placed in the cold room and vortex vigorously at maximum speed for a period of 1 minute; repeat 10 times, equilibrating the tubes on ice for 2 minutes between bursts.
4. Allow to stand for 10 minutes and sediment the cell debris and the beads by microcentrifugation at 14,000 rpm for 10 minutes.
5. Collect the supernatant in a new prechilled microtube with a micropipette[1] and clarify the sample by centrifugation 10 minutes at 14,000 rpm in an eppendorf centrifuge.[2]
6. Measure protein concentration before injection onto the column.[3] Take 1- to 10-μL aliquots to measure the protein concentration, following the standard procedure Bio-Rad Assay for microtiter plate instructions, using BSA as a standard reference protein.
7. Dispense sample by dividing into 20-μL aliquots in the 1.5-mL microtubes for injection and immediately freeze the remainder at −80°C (if possible, freeze the sample in nitrogen liquid first).

[1] The tube contains two layers; take only the top layer that contains soluble protein, and estimate the concentration of protein for the subsequent step.

[2] The sample to be loaded onto the HPLC system must be free of the particulates susceptible to clog its valves, lines, and columns. The particulates may be removed either by filtration with 0.22–0.45-μm membrane or by centrifugation. Yeast crude cell lysate solution can be difficult to filter due to the presence of glass beads so centrifugation is the method of choice.

[3] Failure to measure the amount of the protein in the sample may lead to the injection of an excessively concentrated sample, which may cause the protein to precipitate out in the HPLC tubing or within the column itself, causing a pressure rise. The shelf lifetime of the column is also shortened by loading the protein in excess of the maximum capacity of the column.

### 3.2 HPLC Sample Prefractionation

To achieve the high degree of resolution while maintaining results reproducibility and getting fractionation of the protein in a reasonable time frame, a binary or quaternary eluent strategy is required. Our basic binary strategy involves the separation of protein mixtures using a binary eluent IEX–HPLC method. This approach uses a combination of two mobile phase buffers differing only in salt concentration, either 10-mM Tris-HCl, pH 7.5, no salt (*buffer A*), or 10-mM Tris-HCl + 600-mM NaCl, pH 7.5. The fractionation is accomplished by applying the gradient of salt during a 30–90-minute chromatographic run to resolve and separate the proteins.

The protocol described in this section is an example of the one currently performed in our laboratory for routine proteomic analyses. All of our fractionation experiments are carried out on the Agilent 1100 HPLC system equipped with a vacuum degasser, a binary pump, a refrigerated autosampler with a 100-μL injector loop, 2-D column compartments with a thermostat, a multiwavelength detector (MWD), a cooled autocollection fraction module, and an Agilent ChemStation for chromatograms acquisition and instrument control. As a test case, soluble yeast proteins were fractionated using two combined column chromatography, an anion-exchange column (PolyWAX LP 50 × 2.1 mm i.d., 5 μm, 1000 Å) connected upstream of the cation-exchange column (PolyCAT A, 50 × 2.1 mm i.d., 5 μm, 1000 Å) and protected with a precolumn guard cartridge.[4]

Using the above-described HPLC system and columns together with the salt gradient and the appropriate instrument preparation, 27 HPLC fractions may be collected in less than 35 minutes. Figure 2 depicts the typical chromatograms obtained when the following protocol was applied to the yeast soluble cell lysate:

1. Flush and equilibrate the entire HPLC system with 40 mL of the binding buffer (Buffer A) at 2 mL. min$^{-1}$ using an empty union in place of the column.
2. Stop the flow, remove the union, and install the guard cartridge and the columns (make sure your columns are compatible with the buffer systems); use a 10-cm-long Peek tubing (0.13 mm i.d.) to connect the anion exchanger to a cation-exchange column.
3. Set the column temperature at 17°C to enhance protein integrity during the run, and turn on the HPLC system, flush columns with 80% of Buffer B at a flow rate of 250 μL. min$^{-1}$ for 20 minutes. Follow the manufacturer's instructions when using new columns, and carefully monitor the column pressure for any backpressure indicating possible clogging.
4. Condition columns with pure Buffer A (10-mM Tris-HCl, 3-mM $NaN_3$, pH 7.5) at the above flow rate and time or until you get a flat Ultraviolet baseline at 280 nm.

[4] As discussed, crude cell lysate can contaminate the anion-exchange column resulting in increased backpressure and decreased efficiency. Poly WAX LP (10 × 2.1 mm, 5 μm, 1000 Å) precolumn is used before the PolyWAX LP column to act as a prefilter because the material is the same as the anion exchanger but is inexpensive. The use of a preguard column extends the lifetime of the anion-exchange column and so can save both time and money.

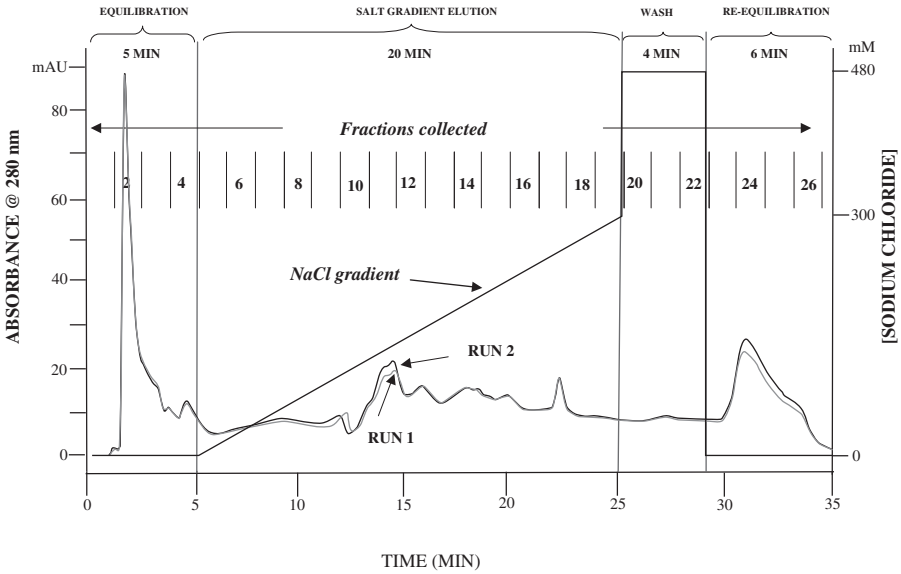


**Figure 2. Repeat ultraviolet-traces (280 nm) recorded for a representative yeast whole cell lysate separated by IEX–HPLC showing the reproducibility of the procedure**. Chromatography was performed at 17°C using a mixed bed of ion-exchange columns, PolyWAX LP (50 × 2.1 mm i.d., 5 µm, 1000 Å) and PolyCAT A (50 × 2.1 mm i.d., 5 µm, 1000 Å). The injection of 80-µg total protein onto the column was followed by 5 minutes of column equilibration with 100% of buffer A, and then 20 minutes of protein fractionation with a linear NaCl gradient from 0% to 50% of buffer B. After protein separation, the column was flushed with 80% buffer B for 4 minutes and then re-equilibrated for 6 minutes with the starting buffer A. (This figure is available in full color at ftp://ftp.wiley.com/public/sci_tech_med/pharmaceutical_biotech/.)

5. Apply 5–10 µL of the centrifuged crude soluble cell extracts (80–100-µg total protein) onto the column using automated (autosampler) injector that has been kept at 4°C.
6. Refer to Figure 2 for protein fractionation with the given salt gradient condition.
7. Monitor protein separation with a MWD set at 280 nm.
8. Collect HPLC fractions based on time (e.g., 1–2 minutes per fraction) by use of an automated sample collector, kept at 4°C, with 1.5-mL Agilent Well Plate fraction collector (325-µL fractions are typically collected per 1.5-mL microtubes).

PROTEOLYSIS AND LC–MS SAMPLE PREPARATION. The following procedure allows for digestion of the proteins in collected HPLC fractions or the control unfractionated crude cell lysate to enable their analysis by LC–MS. After fractionation with HPLC, fractions are precipitated and digested as follows:

1. To each HPLC fraction, add one-tenth cold TCA (100%, fresh solution); vortex the tubes and incubate 30 minutes on ice or overnight at 4°C (cold room).

2. Spin the tubes for 30 minutes at maximum speed (14,000 rpm) in a cooled or refrigerated microcentrifuge; aspirate half of the supernatant (be careful not to disturb the pellet).
3. To each tube, add 600-μL ice-cold HPLC-grade acetone and incubate at –20°C for 10–20 min.
4. Spin tubes as in step 2 and aspirate the total supernatant out of the tube without breaking the pellet.
5. Leave the cap open for air drying (in the fume-hood) for 20 minutes with a protective tissue (e.g., Kimwipes) cover held loosely over it (check the sample frequently to prevent overdrying).
6. Dissolve the pellet in 30 μL of the immobilized trypsin solution (600 μL of the digestion buffer, 60 μL of Pierce immobilized trypsin beads, 30 μL of the Applied Biosciences immobilized trypsin beads, and 2 μL of 1-M $CaCl_2$); check the pH (should be ~8.0) with the pH strip; and neutralize with 5 μL of 1-M TRIS (pH 8.0) if necessary.
7. Incubate the sample for two days at 30°C with rotation or agitation.
8. Stop digestion by adding 30 μL of the LC–MS solvent A, and spin the peptides mixtures for 5 minutes at maximum speed (room temperature) in a microcentrifuge to sediment the trypsin beads.
9. Carefully, recover 20 μL of the peptide mixtures into small disposable (e.g., PCR) tubes, and either immediately analyze by LC–MS or store at –20°C.

### 3.3 LC-MS Sample and Data Analysis

A single capillary scale reverse phase high-performance liquid chromatography system coupled online to automated electrospray ion trap tandem mass spectrometry (1D–LCMS) is usually performed to characterize the protein fractions. The following protocol is an example of the implementation of the method to identify the peptide components of tryptic digests of the yeast HPLC fractions analysis using the ion trap tandem mass spectrometer instrument described in Section 2.5, and typical results are shown in Figure 3.

1. Pack a silica capillary-scale microcolumn (150 μm i.d.) 7.5 cm of C18 reverse phase packing material (Zorbax eclipse XDB-$C_{18}$ resin).
2. Place the packed column in line with the LC-MS buffer pump and ion source power supply.
3. Load the peptide mixture onto the column using an autosampler.
4. The digested proteome is chromatographically resolved and the peptides subject to automated precursor ion selection. The following gradient may be increased or decreased according to sample complexity: 100% of solvent A for 1 minute, 0% to 5% of solvent B in 1 minute, 5% to 30% of solvent B in 58 minutes, 30% to 80% of solvent B in 10 minutes, 80% of solvent B for 5 minutes, 80% to 30% of solvent B during 5 minutes, and then step the gradient back at 100% of solvent A. The flow rate at the tip of the needle is set to 150 μL. $min^{-1}$ for 2 minutes, rise to 175 μL. $min^{-1}$ in 58 minutes, and maintain the flow constant for 25 minutes and then reduce to 5 μL. $min^{-1}$ in 5 minutes.

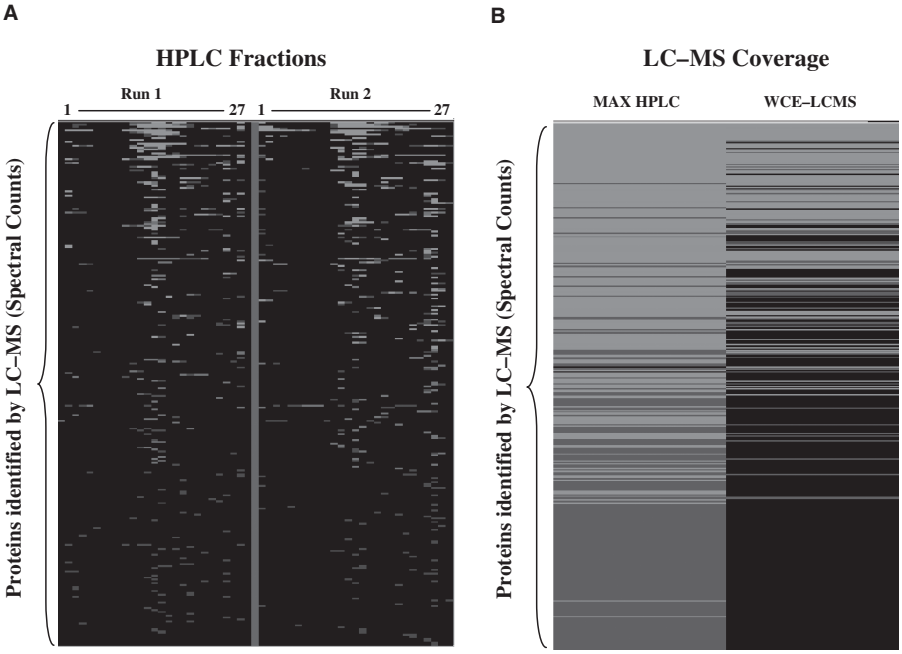


**Figure 3. Yeast cell extract HPLC fractionation and LC–MS analysis**. (*A*) Comparison of representative proteomic patterns recorded for two consecutive HPLC prefractionations. (*B*) Graphical comparison of HPLC prefractionation and straight analysis of the whole cell extract (WCE) by LC–MS (a maximum protein spectral count was recorded for two successive HPLC runs and compared with the corresponding value for the WCE). Black-to-red color shading indicates increasing spectral counts. (This figure is available in full color at ftp://ftp.wiley.com/public/sci_tech_med/pharmaceutical_biotech/.)

5. The mass spectrometer cycles through a full mass precursor ion scan, followed by 2–10 successive tandem mass scans using data-dependent ion isolation and fragmentation of the 10 most intense ions as the gradient progresses. Dynamic exclusion is implemented to prevent repeated fragmentations of the same peptide.
6. Candidate peptide sequences from the mixture are later identified using a computerized database search algorithm (e.g., SEQUEST) [14] and validated using the STATQUEST probabilistic scoring program [15].
7. Comparison of relative protein expression levels between chromatographic runs can be performed using the clustering algorithm Cluster 3.0 and the patterns visualized in a "heat map" format using Java Treeview freeware software.

## REFERENCES

1. Domon B, Aebersold R (2006). Mass spectrometry and protein analysis. *Science*. 312:212–217.
2. Meyers R A (2005). *Encyclopedia of Molecular Cell Biology and Molecular Medicine*. Wiley-VCH Verlag, Weinheim, Germany, pp. 1–43.

3. Kislinger T, Emili A (2003). Going global: Protein expression profiling using shotgun mass spectrometry. *Curr. Opin. Mol. Ther.* 5:285–293.
4. Link A J, Eng J, Schieltz D M, et al. (1999). Direct analysis of protein complexes using mass spectrometry. *Nat. Biotechnol.* 17:676–682.
5. Washburn M P, Wolters D, Yates J R, 3rd. (2001). Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat. Biotechnol.* 19:242–247.
6. Badock V, Steinhusen U, Bommert K, et al. (2001). Prefractionation of protein samples for proteome analysis using reversed-phase high-performance liquid chromatography. *Electrophor.* 22:2856–2864.
7. Gao M-X, Hong J, Yang P-Y, et al. (2005). Chromatographic prefractionation prior to two-dimensional electrophoresis and mass spectrometry identifies: Application to the complex proteome analysis in rat liver. *Anal. Chim. Acta.* 553:83–92.
8. Issaq H J, Conrads T P, Janini G M, et al. (2002). Methods for fractionation, separation and profiling of proteins and peptides. *Electrophor.* 23:3048–3061.
9. Righetti P G, Castagna A, Antonioli P, et al. (2005). Prefractionation techniques in proteome analysis: the mining tools of the third millennium. *Electrophor.* 26:297–319.
10. Barnea E, Sorkin R, Ziv T, et al. (2005). Evaluation of prefractionation methods as a preparatory step for multidimensional based chromatography of serum proteins. *Proteom.* 5:3367–3375.
11. Zhang Z, Smith D L, Smith J B (2001). Multiple separations facilitate identification of protein variants by mass spectrometry. *Proteom.* 1:1001–1009.
12. Snyder L R, Glajch J L, Kirkland J J (1997). Practical HPLC method development. John Wiley & Sons, New York.
13. Lescuyer P, Hochstrasser D F, Sanchez J C (2004). Comprehensive proteome analysis by chromatographic protein prefractionation. *Electrophor.* 25:1125–1135.
14. Eng J K, McCormack A L, Yates I I I J R (1994). An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J. Amer. Soc. Mass Spectrom.* 5:976–989.
15. Kislinger T, Rahman K, Radulovic D, et al. (2003). PRISM, a generic large scale proteomic investigation strategy for mammals. *Mol. Cell Proteom.* 2:96–106.

# AN OVERVIEW OF METABONOMICS TECHNIQUES AND APPLICATIONS

John C. Lindon
*Imperial College London, South Kensington, London, United Kingdom*

**Chapter Contents**

1 Introduction
2 Metabonomics Analytical Technologies
  2.1 NMR Spectroscopy
  2.2 Mass Spectrometry
  2.3 Other Technologies
3 Data Analysis using Chemometrics
  3.1 Chemometrics Methods
  3.2 Biomarker Identification using Chemometrics
4 Selected Applications of Metabonomics
  4.1 Phenotypic and Physiological Effects
  4.2 PreClinical Drug Candidate Safety Assessment
  4.3 Disease Diagnosis and Therapeutic Efficacy
5 Integration of -Omics Results
6 Future Prospects
  References

## 1 INTRODUCTION

Metabonomics is the comprehensive and simultaneous systematic profiling of metabolite levels and their temporal changes in whole organisms through the study of biofluids, tissues, and tissue extracts [1, 2]. A parallel approach mainly from plant science and from the study of *in vitro* systems has led to the term *metabolomics* also

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

being coined [3], and the methods and approaches used in the two disciplines are highly convergent. For the pharmaceutical and medical communities, this multivariate approach holds out the promise of a means by which real disease and drug effect endpoints can be obtained. In this monograph, the main technologies used in metabonomics are summarized, brief details of the types of samples used are given, and the current applications of metabonomics are described. Some prospects for the future are then discussed.

One main problem with integrating information at the three main levels of biomolecular organization and control, transcriptomic, proteomic, and metabonomic, is because these are highly interdependent and can have very different time scales of change. As a result, difficulties can occur in correlation of effects seen by the different omics approaches because some time courses can be very rapid (gene switching), some require much longer time scales (protein synthesis), or some encompass enormous ranges of time scales (metabolite levels). Additionally, biochemical changes do not always occur in the order that would intuitively be expected, i.e., transcriptomic, proteomic, and metabolic, because, for example, pharmacological or toxicological effects at the metabolic level can induce subsequent adaptation effects at the proteomic or transcriptomic levels. One important potential role for metabonomics, therefore, could be to direct the timing of proteomic and genomic analyses in order to maximize the probability of observing "omic" biological changes that are relevant to functional outcomes.

In addition, overlaid with this complexity, is the fact that environmental and lifestyle effects have a large effect on gene and protein expression and on metabolite levels and these have to be considered as part of intersample and interindividual variation. Interpretation of genomic data, in terms of real biological endpoints, is therefore a major challenge because of the modification of gene expression levels by environmental factors. The modeling of such diverse information sets poses significant challenges in terms of bioinformatics. Highly complex animals such as man can be considered as "superorganisms" with an internal ecosystem of diverse symbiotic gut microbiota and parasites that have interactive metabolic processes and for which, in many cases, the genome is not known. The many levels of complexity of the mammalian system and the diverse features that need to be measured to allow "omic" data to be fully used have been reviewed recently [4]. In addition, novel approaches will continue to be required to measure and model metabolic processes in various compartments in different interacting cell types, with genomes that are connected by cometabolic processes in such a global mammalian system [5].

Typically, mammalian metabonomics studies of relevance to the pharmaceutical industry are carried out on biofluids because these are often easy to obtain and can provide an integrated view of the whole system's biology. The biochemical profiles of the main diagnostic fluids, blood plasma, cerebrospinal fluid (CSF), and urine, can reflect both normal variation and the impact of disease and drug toxicity or efficacy on single or multiple organ systems. Urine and plasma are obtained essentially noninvasively, and hence, are most appropriate for clinical trials monitoring and disease diagnosis. However, there is a wide range of fluids that can be, and have been, studied, including seminal fluids, amniotic fluid, synovial fluid, digestive fluids, blister and cyst fluids, lung aspirates, and dialysis fluids. In addition, several metabonomics studies have used analysis of tissue biopsy samples and their

lipid and aqueous extracts. This is particularly true for studies in fields other than mammalian systems, such as for plants, or for model organisms such as yeast and cell culture studies. Additionally, the approach can be used to characterize *in vitro* cell systems such as Caco-2 cells commonly used for cell uptake studies [6] or tissue spheroids, which can be used, for example, as model systems for liver or tumor investigations [7].

As described in more detail in Section 2, the main analytical techniques that are employed for metabonomic studies are based on nuclear magnetic resonance (NMR) spectroscopy and mass spectrometry (MS). The latter technique requires a preseparation of the metabolic components using either gas chromatography (GC) after chemical derivatization or liquid chromatography (LC), with the newer method of ultra-high-pressure LC (UPLC) being used increasingly. The use of capillary electrophoresis (CE) coupled to MS has also shown promise. Other more specialized techniques such as Fourier transform infrared spectroscopy and arrayed electrochemical detection have been used in some cases.

All metabonomics studies result in complex multivariate datasets that require a variety of chemometric, bioinformatic, and visualization tools for effective interpretation. The aim of these procedures is to produce biochemically based fingerprints that are of diagnostic or other classification value. A second stage, crucial in such studies, is to identify the substances causing the diagnosis or classification, and these become the combination of biomarkers that reflects actual biological events. Thus, metabonomics studies allow real-world or biomedical endpoint observations to be obtained.

There have been several reviews of metabonomics and metabolomics recently that describe the various techniques used and that summarize the main areas of application. These provide more detail than can be given here and serve to act as pointers to the original literature studies [8–11].

## 2 METABONOMICS ANALYTICAL TECHNOLOGIES

### 2.1 NMR Spectroscopy

NMR spectroscopy is a nondestructive technique, widely used in chemistry, that provides detailed information on molecular structure, both for pure compounds and in complex mixtures [12]. NMR spectroscopic methods can also be used to probe metabolite molecular dynamics and mobility as well as substance concentrations through the interpretation of NMR spin relaxation times and by the determination of molecular diffusion coefficients [13].

Standard pulse-Fourier transform NMR spectra typically take only a few minutes to acquire, often using robotic flow-injection methods, with automatic sample preparation involving buffering and addition of $D_2O$ as a magnetic field lock signal for the spectrometer. For large-scale studies, bar-coded vials containing the biofluid are used and the contents of these can be transferred and prepared for analysis using robotic liquid handling technology into 96-well plates with the whole process under LIMS system control. Using NMR flow probes, the capacity for NMR analysis has increased enormously recently, and around 200 samples per day can be measured on one spectrometer. Alternatively, for more precious samples or for

those of limited volume, conventional 5-mm NMR tubes are usually used, either individually or using a commercial sample tube changer and automatic data acquisition. Typical NMR spectra of human biofluids are shown in Figure 1.

A $^{1}$H NMR spectrum of urine acquired as seen in Figure 1, typically contains thousands of sharp lines from predominantly low-molecular-weight metabolites. This spectrum is the result of 64 co-added scans each requiring about 5 s, thus yielding a total data acquisition time of around 5 minutes. The large interfering NMR signal arising from water in all biofluids is easily eliminated by use of appropriate standard NMR solvent suppression methods, either by secondary radio-frequency (RF) irradiation at the water peak chemical shift or by use of a specialized NMR pulse sequence that does not excite the water resonance. The position of each spectral band (known as its chemical shift and measured in frequency terms, in ppm, from that of an added standard reference substance) gives information on molecular group identity and its molecular environment, e.g., methyl group on an aromatic ring, olefinic hydrogen, pyridyl ring proton, or aldehyde. The reference compound used in aqueous media is usually the sodium salt of 3-trimethylsilylpropionic acid (TSP) with the methylene groups deuterated to avoid giving rise to peaks in the $^{1}$H NMR spectrum. The multiplicity of the splitting pattern on each band, and the magnitudes of the splittings (caused by a nuclear spin–spin interaction mediated through the electrons of the chemical bonds, and known as

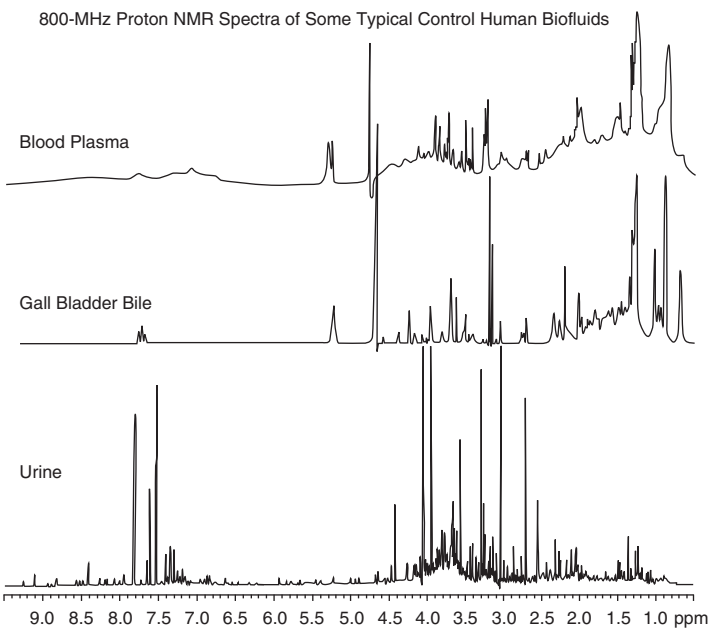


**Figure 1.** Shown are 800-MHz $^{1}$H NMR spectra of control human biofluids. The peaks arise from different chemical types of hydrogen in the biochemicals present. The peak areas are related to molar concentrations, and the peak positions and splittings allow information to be obtained, after expert interpretation, of the molecules responsible for the peaks. The signal from water has been suppressed by an NMR procedure to avoid problems of dynamic range in the detection process.

J-coupling), provide knowledge about nearby protons, their through-bond connectivities, the relative orientation of nearby C–H bonds, and hence molecular conformations. The band areas relate directly to the number of protons, giving rise to the peak and hence to the relative concentrations of the substances in the sample. Absolute concentrations can be obtained if the sample contains an added internal standard of known concentration, if a standard addition of the analyte of interest is added to the sample, or if the concentration of a substance is known by independent means (e.g., glucose in plasma can be quantified by a conventional biochemical assay).

Blood plasma and serum contain both low- and high-molecular-weight components, and these give a wide range of signal line widths. Broad bands from protein and lipoprotein signals contribute strongly to the $^1$H NMR spectra, with sharp peaks from small molecules superimposed on them. Standard NMR pulse sequences, where the observed peak intensities are edited on the basis of molecular diffusion coefficients or on NMR relaxation times (known as $T_1$, $T_{1\rho}$, and $T_2$), can be used to select only the contributions from proteins, and other macromolecules and micelles, or alternatively to select only the signals from the small-molecule metabolites, respectively [12]. A typical $^1$H NMR spectrum from human blood serum is shown in Figure 1, and a series of edited NMR spectra, from rat blood serum, based on the approaches described above are given in Figure 2.

The spin-echo spectrum has the broader peaks from nuclei with shorter spin relaxation times attenuated (these are macromolecules and substances involved in

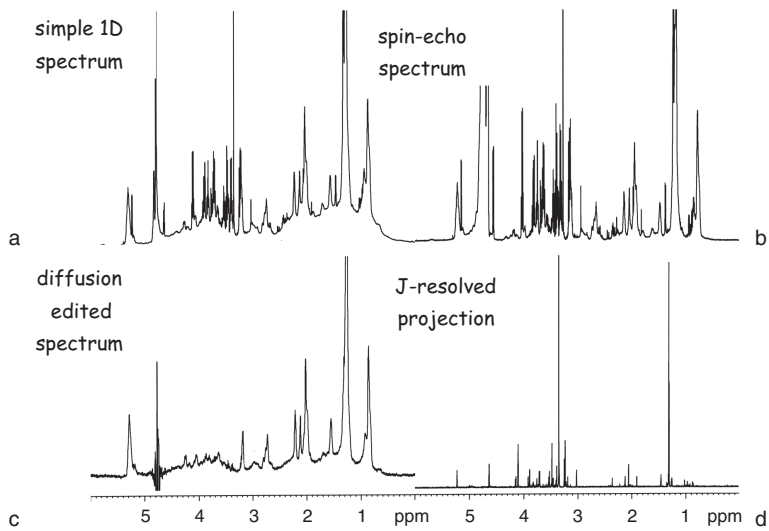


**Figure 2.** $^1$H NMR spectra of rat serum illustrating the various NMR responses that are possible through the use of different pulse sequences, which edit the spectral intensities: (a) standard water suppressed spectrum, showing all metabolites; (b) CPMG spin-echo spectrum, with attenuation of peaks from fast relaxing components such as macromolecules and lipoproteins; (c) diffusion-edited spectrum, with attenuation of peaks from fast diffusing components such as small molecules; and (d) a projection of a 2D J-resolved spectrum on to the chemical shift axis, showing removal of all spin–spin coupling and peaks from fast relaxing species.

chemical exchange). The diffusion-edited spectrum, on the other hand, has peaks only from macromolecules and other slow-moving species present, with the peaks from highly mobile small molecules attenuated. Thus, it is also possible to use these approaches to investigate molecular mobility and flexibility, and to study intermolecular interactions such as the reversible binding between small molecules and proteins [14].

Two recent improvements in NMR detector technology have resulted in major improvements in sample quantity requirements for NMR spectroscopy. The first is the commercialization of miniaturized detectors. Now it is possible to study metabolic profiles by NMR using as little as 2–20 μL of sample, and examples have been published using CSF and blood plasma [15]. Secondly, the availability of cryogenic NMR probe technology where the detector coil and preamplifier are cooled to around 20K has provided an improvement in spectral signal-to-noise ratios of up to a factor of 5 by reducing the thermal noise in the electronics of the spectrometer. This allows use of smaller samples, and when this approach is combined with the use of miniaturized probes as described above, optimum sensitivity and sample requirements are achieved. Conversely, because the NMR signal-to-noise ratio is proportional to the square root of the number of co-added scans, shorter data acquisition times by up to a factor of 20–25 become possible for the same amount of sample. NMR spectroscopy of biofluids detecting the much less sensitive $^{13}C$ nuclei, which also only have a natural abundance (1.1%), also becomes possible because of the increase in signal-to-noise ratio [16].

Within the last few years, the development of a technique called high-resolution $^{1}H$ magic angle spinning (MAS) NMR spectroscopy has made feasible the acquisition of high-resolution NMR data on small pieces of intact tissues with no pretreatment [17–21]. Rapid spinning of the sample (typically at ~4–6 kHz) at an angle of 54.7° relative to the applied magnetic field serves to reduce the loss of information caused by line broadening effects observed in nonliquid samples such as tissues. These broadenings are caused by sample heterogeneity and residual anisotropic NMR parameters that are normally averaged out in free solution where molecules can tumble isotropically and rapidly. NMR spectroscopy on a tissue matrix in an MAS experiment is the same as solution-state NMR, and all common pulse techniques can be employed to study metabolic changes and to perform molecular structure elucidation. Typical $^{1}H$ NMR spectra from a range of tissue types are shown in Figure 3. In most cases, a standard set of one-dimensional (1D) sequences is used to describe the biochemical changes in the pool of low-molecular-weight metabolites and lipids. The different spin properties of macromolecules and small molecules can also be exploited using spectral editing techniques to filter out certain subgroups of peaks, as in solution-state NMR spectroscopy.

MAS NMR spectroscopy has straightforward sample preparation, although this still has to be carried out manually. Snap-frozen tissue samples (as little as 10 mg), which have been stored at −80°C, are defrosted and cut to select the region of interest. The original samples should be frozen rapidly using small specimens in liquid nitrogen to avoid microcrystallization of the water in the cells and consequent cell damage. Then, the tissue is rinsed with 0.9% $D_2O$/saline to wash off remaining blood and the specimen is transferred into a 4-mm-diameter zirconia rotor and a Teflon spacer is used to restrict the sample volume, to eliminate trapped air bubbles, and to increase sample homogeneity. As soon as the sample has been

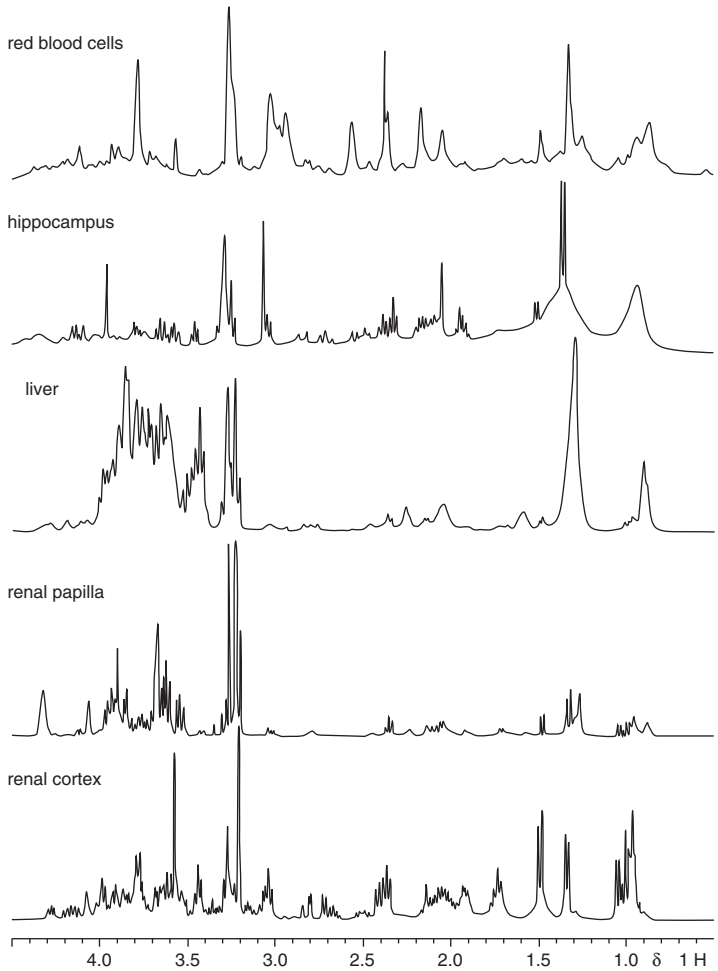


**Figure 3.** High-resolution 400-MHz $^1$H MAS NMR spectra of various tissues, with sample spinning at 4.2 kHz.

transferred into the rotor, a trace volume of 0.9% $D_2O$/saline is added to provide a field-frequency lock for the $^1$H MAS NMR data acquisitions. In cases where the chemical shifts of signals have not been identified previously, it is necessary to add a reference standard such as TSP in the saline solution, although sometimes it is possible to use a well-characterized peak of an easily assigned compound, such as the anomeric proton resonance of α-glucose, as a secondary chemical shift standard. Fast rotation of the sample can cause frictional heating, and this is prevented by a controlled cooling of the input gas to compensate.

Identification of biomarkers can involve the application of a range of techniques, including two-dimensional (2D) NMR experiments [12]. Although, all of the armory of the usual analytical physical chemistry can and should be used, including mass spectrometry, $^1$H NMR spectra of urine and other biofluids, even though they are very complex, allow many resonances to be assigned directly based on their chemical shifts, signal multiplicities, and by adding authentic material, and further

information can be obtained by using spectral editing techniques, as described above.

Two-dimensional NMR spectroscopy can be useful for increasing signal dispersion and for elucidating the connectivities between signals, thereby enhancing the information content and helping to identify biochemical substances. These include the $^{1}H$-$^{1}H$ 2-D J-resolved experiment, which attenuates the peaks from macromolecules and yields information on the multiplicity and coupling patterns of resonances, which is a good aid to molecule identification. The projection of such a spectrum on to the chemical shift axis yields a fingerprint of peaks from only the most highly mobile small molecules, with the added benefit that all spin-coupling peak multiplicities have been removed (see Figure 2). Other 2D experiments known as COSY and TOCSY provide $^{1}H$–$^{1}H$ spin–spin coupling connectivities, thus giving information on which hydrogens in a molecule are close in chemical bond terms. Use of other types of nuclei, such as naturally abundant $^{13}C$ or $^{15}N$ or where present $^{31}P$, can be important to help assign NMR peaks, and here heteronuclear correlation NMR experiments can be obtained by use of appropriate NMR pulse sequences. These benefit by the use of so-called inverse detection, where the lower sensitivity or less abundant nucleus NMR spectrum (such as $^{13}C$) is detected indirectly using the more sensitive/abundant nucleus ($^{1}H$) by making use of spin–spin interactions such as the one-bond $^{13}C$–$^{1}H$ spin–spin coupling between the nuclei to effect the connection. These yield both $^{1}H$ and $^{13}C$ NMR chemical shifts of CH, $CH_2$, and $CH_3$ groups, which are useful again for identification purposes. There is also a sequence that allows correlation of protons to quaternary carbons based on long-range $^{13}C$–$^{1}H$ spin–spin coupling between the nuclei.

## 2.2 Mass Spectrometry

Mass spectrometry has also been widely used in metabolic fingerprinting and metabolite identification, with most studies to date on plant extracts and model cell system extracts, although its application to mammalian studies is increasing. In general, with the exception of some studies using Fourier transform, or ion-cyclotron-MS, a prior separation of the complex mixture sample using chromatography is required. MS is inherently considerably more sensitive than NMR spectroscopy, but in complex mixtures of very variable composition such as biofluids, it is necessary generally to employ different separation techniques (e.g., different LC column packings) for different classes of substances. MS is also a mainstay technique for molecular identification purposes, especially through the use of $MS^n$ methods for fragment ion studies. Analyte quantitation by MS in complex mixtures of highly variable composition can also be impaired by variable ionization and ion suppression effects. Chemical derivatization might be necessary to ensure volatility and analytical reproducibility, in complex mixtures such as biofluids.

For both profiling and metabolite identification, most published metabonomics studies on mammalian biological systems have used NMR spectroscopy, but HPLC–MS techniques are increasing in usage, particularly using electrospray ionization. Thus, for metabonomics applications on biofluids such as urine, an HPLC chromatogram is generated with MS detection, and usually both positive and negative ion chromatograms are measured. At each sampling point in the chromatgram, there is a full mass spectrum and so the data are three-dimensional in nature, i.e.,

retention time, mass-to-charge ratio, and intensity. Given this very high resolution, it is easy to cut out any mass peaks from interfering substances such as drug metabolites, without compromising the dataset.

For plant metabonomics studies, the principal approach has been to make extracts of the samples followed by chemical derivatization, and then to employ separation using GC with molecular identification using MS by comparing spectra against publically available and single-laboratory databases [10, 11].

UPLC is a new combination of a 1.7-μm reversed phase packing material and a chromatographic system, operating at around 12,000 psi. This has enabled a marked improvement in chromatographic performance to be obtained for complex mixture separation, with better peak resolution and increased speed and sensitivity. UPLC gives more than doubling of peak capacity compared with HPLC, an almost 10-fold increase in speed, and a 3- to 5-fold increase in sensitivity compared with that generated with a conventional stationary phase. Because of the much improved chromatographic resolution of UPLC, the problem of ion suppression from co-eluting peaks is greatly reduced. UPLC–MS has been used for metabolic profiling of urines from males and females of two groups of phenotypically normal mouse strains and a nude mouse strain [22].

Recently, CE coupled to mass spectrometry has also been explored as a suitable technology for metabonomics studies [23]. Charged metabolites are first separated by CE based on charge and size and then selectively detected using MS by monitoring over a large range of $m/z$ values. This method was used to measure 352 metabolic standards and then employed for the analysis of 1692 metabolites from *Bacillus subtilis* extracts, revealing significant changes in metabolites during the bacterial sporulation.

For biomarker identification, it is also possible to separate out substances of interest from a complex biofluid sample using techniques such as solid phase extraction or HPLC. For metabolite identification, directly coupled chromatography–NMR spectroscopy methods can be used. The most powerful of these "hyphenated" approaches is HPLC–NMR–MS [24] in which the eluting HPLC peak is split with parallel analysis by directly coupled NMR and MS techniques. This can be operated in on-flow, stopped-flow, and loop-storage modes and thus can provide the full array of NMR and MS-based molecular identification tools. These include MS–MS for identification of fragment ions and FT–MS or TOF–MS for accurate mass measurement and hence derivation of molecular empirical formulae.

In summary NMR and MS approaches are highly complementary, and use of both is often necessary for full molecular characterization. MS can be more sensitive with lower detection limits provided the substance of interest can be ionized, but NMR spectroscopy is particularly useful for distinguishing isomers, for obtaining molecular conformation information, and for studies of molecular dynamics and compartmentation.

### 2.3 Other Technologies

Although most metabonomics applications have used either NMR spectroscopy or chromatography coupled to mass spectrometry, other techniques have been explored. Fourier transform infrared spectroscopy has been applied to a few metabolomics investigations including the evaluation of different *E. coli* mutants by

analysis of their secreted metabolites [25]. Although in principal, different responses can be achieved from different sample classes, and analyzed using pattern recognition methods, the main limitation of this technique is the low level of detailed molecular identification that can be achieved, and indeed in the case quoted above, MS was also employed for metabolite identification.

An alternative approach has been pioneered by Lewitt et al. [26]. This uses an array of coulometric detectors following HPLC separation to detect redox-active compounds in a complex mixture such as in CSF or tissue extracts. Although this approach does not identify compounds directly, the combination of retention time and redox properties can serve as a basis for database searching of libraries of standard compounds. The separation output can also be directed to a mass spectrometer for additional identification experiments.

## 3 DATA ANALYSIS USING CHEMOMETRICS

### 3.1 Chemometrics Methods

The complex data that arise from, for example, an NMR spectrum of a sample can be thought of as an object with a multidimensional set of metabolic coordinates, the values of which are the spectral intensities at each data point. Thus, each spectrum becomes a point in a multidimensional metabolic hyperspace. In chemistry, the term *chemometrics* is generally applied to describe the use of both parametric and nonparametric multivariate statistical approaches to chemical numerical data. The general aim is to classify an object based on identification of inherent patterns in a set of experimental measurements or descriptors and to identify those descriptors responsible for the classification. The approach can also be used for reducing the dimensionality of complex datasets, for example, by 2D or 3D mapping procedures, to enable easy visualization of any clustering or similarity of the various samples. Alternatively, in what are known as "supervised" methods, multiparametric datasets can be modeled so that the class of separate samples (a "validation set") can be predicted based on a series of mathematical models derived from the original data or "training set" [27].

Principal components analysis (PCA) is one of the simplest techniques that has been used extensively in metabonomics, and this expresses most of the variance within a dataset using a smaller number of factors or principal components. Each PC is a linear combination of the original data parameters whereby each successive PC explains the maximum amount of variance possible, not accounted for by the previous PCs. Each PC is orthogonal and therefore independent of the other PCs. Thus, the variation in the spectral set is usually described by many fewer PCs than the original data point values because the less important PCs describe the noise variation in the spectra. Conversion of the data matrix to PCs results in two matrices known as scores and loadings. Scores, the linear combinations of the original variables, are the coordinates for the samples in the established model and may be regarded as the new variables. In a scores plot, each point represents a single sample spectrum. The PC loadings define the orientation of the computed PC with respect to the original variables and thus indicate which variables carry the greatest weight in transforming the position of the original samples from the data matrix into their new position in the scores matrix. In the loadings plot, each point represents a dif-

ferent spectral intensity. Thus, the cause of any spectral clustering observed in a PC scores plot is interpreted by examination of the loadings that cause any cluster separation. In addition, there are many other visualization methods such as non-linear mapping and hierarchical cluster analysis.

The upper part of Figure 4 shows a PC scores plot where each point is based on the $^{1}$H NMR spectrum of a rat urine sample. In this case, the open and filled symbols represent samples of rat urine received from two different sites, from animals dosed with a model liver toxin, hydrazine, and sampled at various time points in hours after dosing as shown [28].

One of many widely used supervised methods (i.e., using a training set of data with known outcomes) is partial least squares (PLS). This method relates a data matrix containing independent variables from samples, such as spectral intensity values (an X matrix), to a matrix containing dependent variables (or measurements of response) for those samples (a Y matrix). PLS can also be used to examine the

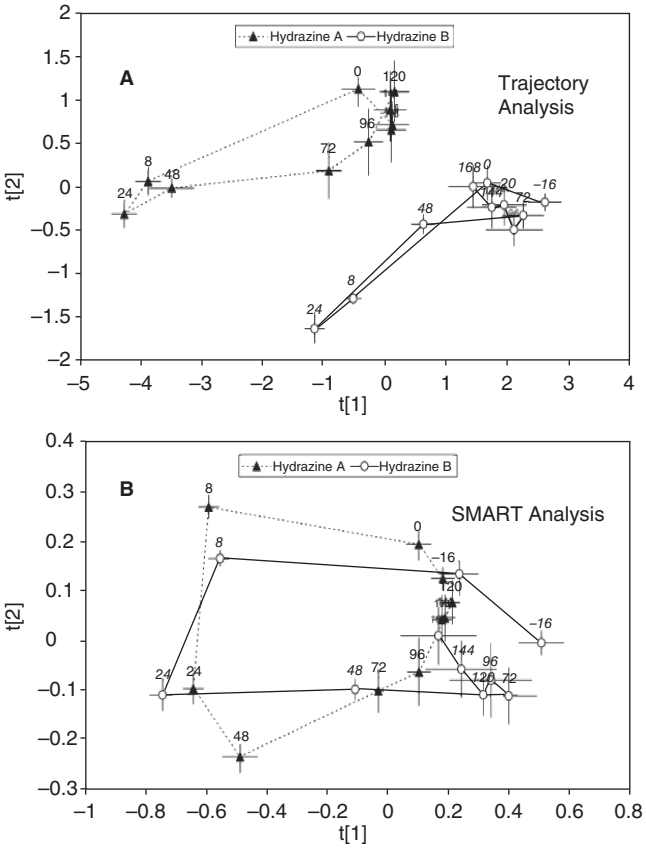


**Figure 4.** Principal components scores plots (PC1 vs. PC2) based on 600-MHz $^{1}$H NMR spectra of rat urine. (a) Standard trajectory analysis of urinary NMR data obtained after treatment with hydrazine from identical studies repeated at different sites, open and closed symbols. (b) The same data after application of the trajectory matching technique, SMART. Error bars represent the standard error for each time point average [28].

influence of time on a dataset, which is particularly useful for biofluid NMR data collected from samples taken over a time course of the progression of a pathological effect. PLS can also be combined with discriminant analysis (DA) to establish the optimal position to place a discriminant surface that best separates classes. It is possible to use such supervised models to provide classification probabilities or even quantitative response factors for a wide range of sample types, but given the strong possibility of chance correlations when the number of descriptors is large, it is important to build and test such chemometric models using independent training data and validation datasets. Extensions of this approach allow the evaluation of those descriptors that are completely independent (orthogonal) to the Y matrix of endpoint data. This orthogonal signal correction (OSC) can be used to remove irrelevant and confusing parameters.

Apart from the methods described above that use linear combinations of parameters for dimension reduction, classification, or regression, other methods exist that are not limited in this way. For example, neural networks comprise a widely used nonlinear approach for modeling data. A training set of data is used to develop algorithms, which "learn" the structure of the data and can cope with complex functions. The basic software network consists of three or more layers, including an input level of neurons (spectral descriptors or other variables), one or more hidden layers of neurons that adjust the weighting functions for each variable, and an output layer that designates the class of the object or sample. Recently, probabilistic neural networks, which represent an extension to the approach, have shown promise for metabonomics applications in toxicity [29]. Other approaches that are currently being tested include genetic algorithms, machine learning, and Bayesian modeling [30].

### 3.2 Biomarker Identification using Chemometrics

Recently, a new method for identifying multiple NMR peaks from the same molecule in a complex mixture, hence providing a new approach to molecular identification, has been introduced. This is based on the concept of statistical total correlation spectroscopy and has been termed *STOCSY* [31]. This takes advantage of the multicollinearity of the intensity variables in a set of spectra (e.g., $^{1}$H NMR spectra) to generate a pseudo-two-dimensional NMR spectrum that displays the correlation among the intensities of the various peaks across the whole sample. This method is not limited to the usual connectivities that are deducible from more standard two-dimensional NMR spectroscopic methods, such as TOCSY. Added information is available by examining lower correlation coefficients or even negative correlations, because this leads to a connection between two or more molecules involved in the same biochemical pathway. In an extension of the method, the combination of STOCSY with supervised chemometrics methods offers a new framework for analysis of metabonomic data. In a first step, a supervised multivariate discriminant analysis can be used to extract the parts of NMR spectra related to discrimination between two sample classes. This information is then cross-combined with the STOCSY results to help identify the molecules responsible for the metabolic variation. To illustrate the applicability of the method, it has been applied to 114 $^{1}$H NMR spectra of urine from a metabonomic study of a model of insulin resistance based on the administration of a carbohydrate diet to three different mice strains in

which a series of metabolites of biological importance could be conclusively assigned and identified by use of the STOCSY approach [31].

## 4 SELECTED APPLICATIONS OF METABONOMICS

### 4.1 Phenotypic and Physiological Effects

To determine therapeutic or toxic effects or to understand the biochemical alterations caused by disease, it is necessary first to understand any underlying physiological sources of variation. To this end, metabonomics has been used to separate classes of experimental animals such as mice and rats according to several inherent and external factors based on the endogenous metabolite patterns in their biofluids [32]. Such differences may help explain differential toxicity of drugs between strains and interanimal variation within a study. Many other effects can be distinguished using metabonomics, including male/female differences, age-related changes, estrus cycle effects in females, diet, diurnal effects, and interspecies differences and similarities [32].

Metabonomics has also been used for the phenotyping of mutant or transgenic animals and the investigation of the consequences of transgenesis such as the transfection process itself [33]. Genetic modifications used in the development of genetically engineered animal models of disease are often made using such transfection procedures, and it is important to differentiate often-seen unintended consequences of this process from the intended result. Metabonomic approaches can give insight into the metabolic similarities or differences between mutant or transgenic animals and the human disease processes that they are intended to simulate and hence their appropriateness for monitoring the efficacy of novel therapeutic agents. This suggests the method may be appropriate for following treatment regimes such as gene therapy.

The importance of gut microfloral populations on urine composition has been highlighted by a study in which axenic (germ free) rats were allowed to acclimatize in normal laboratory conditions and their urine biochemical makeup was monitored for 21 days [34]. The combined influence of gut microflora and parasitic infections on urinary metabolite profiles has also been elucidated [35].

### 4.2 PreClinical Drug Candidate Safety Assessment

Despite the regulatory requirements and the huge investment by pharmaceutical companies in safety testing in animals, unexpected results are sometimes observed in the clinic and drugs occasionally still have to be withdrawn from the marketplace. The selection of robust candidate drugs for development based on mimimization of the occurrence of drug adverse effects is therefore one of the most important aims of pharmaceutical R&D, and the pharmaceutical industry is now embracing metabonomics for evaluating the adverse effects of candidate drugs. The National Center for Toxicological Research, a part of the U.S. Food and Drug Administration, is also investigating the usefulness of the approach.

In this application, NMR-based metabonomics can be used for (1) definition of the metabolic hyperspace occupied by normal samples; (2) the consequential rapid classification of a biofluid sample as normal or abnormal; (3) if abnormal,

classification of the target organ or region of toxicity; (4) biochemical mechanism of that toxin; (5) identification of combination biomarkers of toxic effect; and (6) evaluation of the time course of the effect, e.g., the onset, evolution, and regression of toxicity. An example of a metabolic trajectory is shown in Figure 4(a), where each point represents an NMR spectrum of a rat urine sample at various time points after dosing with the model liver toxin, hydrazine [28]. There have been many studies using $^{1}$H NMR spectroscopy of biofluids to characterize drug toxicity going back to the 1980s [36], and the role of metabonomics in general, and magnetic resonance in particular, in toxicological evaluation of drugs has been comprehensively reviewed recently [37].

Metabonomics has already been applied in fields outside human and other mammalian systems. For example, studies in the environmental pollution field have highlighted the potential benefits of this approach by studies of caterpillar hemolymph [38] and earthworm biochemical changes as a result of soil pollution by model toxic substances [39]. In addition, a study of heavy metal toxicity ($As^{3+}$ and $Cd^{2+}$) in wild rodents living on polluted sites has been concluded successfully [40]. In terms of monitoring water quality, one study has evaluated adverse effects in abalone using NMR-based metabonomics [41].

The usefulness of metabonomics for the evaluation of xenobiotic toxicity effects has recently been comprehensively and successfully explored by the Consortium for Metabonomic Toxicology (COMET). This was formed by five pharmaceutical companies and Imperial College, London, United Kingdom [42], with the aim of developing methodologies for the acquisition and evaluation of metabonomic data generated using $^{1}$H NMR spectroscopy of urine and blood serum from rats and mice for preclinical toxicological screening of candidate drugs. The successful outcome is evidenced by the generated databases of spectral and conventional results for a wide range of model toxins (147 in total) that served as the basis for computer-based expert systems for toxicity prediction. The project goals of the generation of comprehensive metabonomic databases (now around 35,000 NMR spectra) and multivariate statistical models (expert systems) for prediction of toxicity, initially for liver and kidney toxicity in the rat and mouse, have now been achieved, and the predictive systems and databases have been transferred to the sponsoring companies.

A feasibility study was carried out at the start of the project, using the same detailed protocol and using the same model toxin, over seven sites in the companies and their appointed contract research organizations. This was used to evaluate the levels of analytical and biological variation that could both arise through the use of metabonomics on a multisite basis. The intersite NMR analytical reproducibility revealed the high degree of robustness expected for this technique when the same samples were analyzed both at Imperial College and at various company sites. This gave a multivariate coefficient of regression between paired samples of only about 1.6% [43].

Additionally, the biological variability was evaluated by a detailed comparison of the ability of the companies to provide consistent urine and serum samples for an in-life study of the same toxin, with all samples measured at Imperial College. There was a high degree of consistency between samples from the various companies, and dose-related effects could be distinguished from intersite variation [43].

As a precursor to developing the final predictive expert systems, metabonomic models were constructed for urine from control rats and mice, enabling identifica-

tion of outlier samples and the metabolic reasons for the deviation. To achieve the project goals, new methodologies for analyzing and classifying the complex datasets were developed. For example, as the expert system takes into account the metabolic trajectory over time, a new way of comparing and scaling these multivariate trajectories was developed (called SMART), and this is illustrated in Figure 4 [28]. Additionally, a novel classification method for identifying the class of toxicity based on all NMR data for a given study has been generated. This has been termed "Classification Of Unknowns by Density Superposition (CLOUDS)" and is a novel non-neural implementation of a classification technique developed from probabilistic neural networks [44]. Modeling the urinary NMR data according to organ of effect (control, liver, kidney, or other organ), using a model training set of 50% of the samples and predicting the other 50%, over 90% of the test samples were classified as belonging to the correct group with only a 2% misclassification rate between these classes. This work showed that it is possible to construct predictive and informative models of metabonomic data, delineating the whole time course of toxicity, the ultimate goal of the COMET project.

### 4.3 Disease Diagnosis and Therapeutic Efficacy

Many examples exist in the literature on the use of NMR-based metabolic profiling to aid human disease diagnosis, including the use of plasma to study diabetes, CSF for investigating Alzheimer's disease, synovial fluid for osteoarthritis, seminal fluid for male infertility, and urine in the investigation of drug overdose, renal transplantation, and various renal diseases. Most of the earlier studies using NMR spectroscopy have been reviewed [45]. For example, a promising use of NMR spectroscopy of urine and plasma, as evidenced by the number of publications on the subject, is in the diagnosis of inborn errors of metabolism in children [46].

Some studies have been undertaken in the area of cancer diagnosis using perchloric acid extracts of various types of human brain tumor tissue [47], and the spectra could be classified using neural network software giving ~85% correct classification. Tissues can be studied by metabonomics through the MAS NMR technique, and published examples include prostate cancer [16, 48], renal cell carcinoma [18], breast cancer [19, 49], and various brain tumors [50]. Other recent studies include an NMR-based urinary metabonomic study of multiple sclerosis in humans and non-human primates [51].

Recently metabonomics has been applied to provide a method for diagnosis of coronary artery disease noninvasively through analysis of a blood serum sample using NMR spectroscopy [52]. Patients were classified, based on angiography, into two groups, those with normal coronary arteries and those with triple coronary vessel disease. Around 80% of the NMR spectra were used as a training set to provide a two-class model after appropriate data filtering techniques had been applied, and the samples from the two classes were easily distinguished. The remaining 20% of the samples were used as a test set, and their class was then predicted based on the derived model with a sensitivity of 92% and a specificity of 93% based on a 99% confidence limit for class membership.

It was also possible to diagnose the severity of the disease that was present by employing serum samples from patients with stenosis of one, two, or three of the coronary arteries. Although this is a simplistic indicator of disease severity, separation of the three sample classes was evident even though none of the wide range

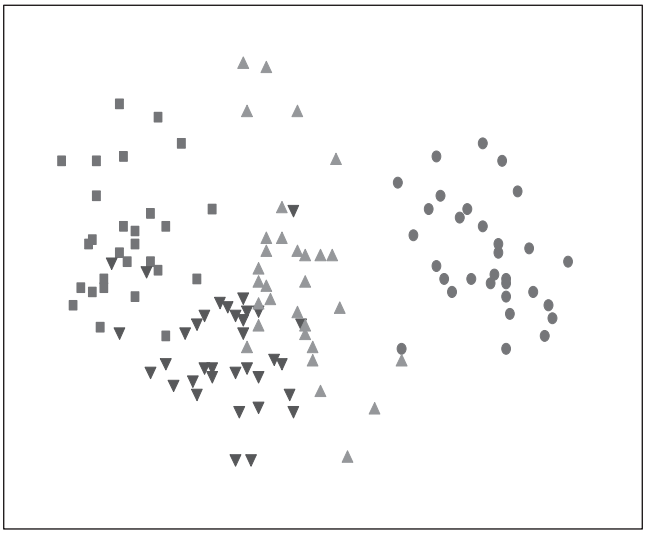

**Figure 5.** Partial least-squares discriminant analysis model for the classification of blood plasma samples in terms of coronary artery disease, based on their $^{1}$H NMR spectra with visualization of the degree of coronary artery occlusion. Each point is based on data from a $^{1}$H NMR spectrum of human blood plasma from subjects with different degrees of coronary artery occlusion. Circles—no stenosis, triangles—stenosis of one artery, inverted triangles—stenosis of two arteries, and squares—stenosis of three arteries. (This figure is available in full color at ftp://ftp.wiley.com/public/sci_tech_med/pharmaceutical_biotech/.)

of conventional clinical risk factors that had been measured was significantly different between the classes. The visualization of the separation of patients with zero, one, two, or three coronary arteries occluded is shown in Figure 5.

## 5 INTEGRATION OF -OMICS RESULTS

The value of obtaining multiple datasets from various biofluid samples and tissues of the same animals collected at different time points has been demonstrated. This procedure has been termed "integrated metabonomics" [2] and can be used to describe the changes in metabolic chemistry in different body compartments affected by exposure to toxic drugs. An illustration of the types of information that can be obtained is shown in Figure 6, from an NMR spectroscopic study of the acute toxicity of α-naphthylisothiocyanate, a model liver toxin [53]. Such timed profiles in multiple compartments are characteristic of particular types and mechanisms of pathology and can be used to give a more complete description of the biochemical consequences than can be obtained from one fluid or tissue alone [54].

Integration of metabonomics data with that from other multivariate techniques in molecular biology such as from gene array experiments or proteomics is also feasible. Thus, it has also been possible to integrate data from transcriptomics and metabonomics to find, after acetaminophen administration to mice, common metabolic pathways implicated by both gene expression changes and changes in metabolism [55]. In a similar fashion, changes in gene expression detected in microarray

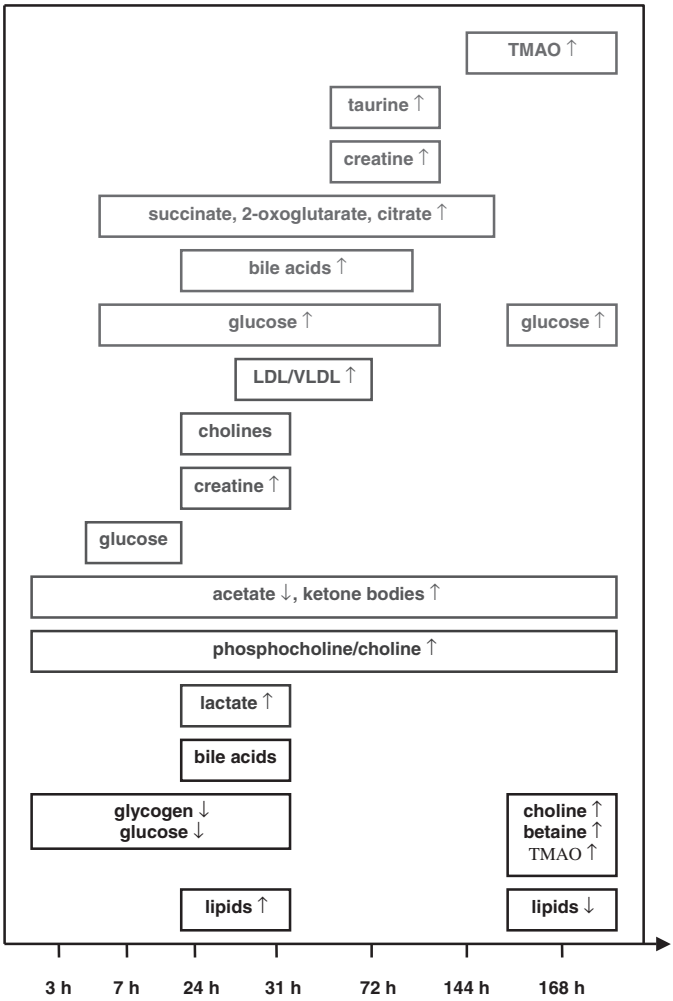


**Figure 6.** Representation of the information that can be obtained from an integrated metabonomics study. This shows the metabolites that are changed in a range of biofluids and tissues as a function of time after a single acute administration of the liver toxin α-naphthylisothiocyanate [53]. (This figure is available in full color at ftp://ftp.wiley.com/public/sci_tech_med/pharmaceutical_biotech/.)

experiments can lead to the identification of changed enzyme activity, and this can also be achieved by analysis of metabolic perturbations [56].

## 6 FUTURE PROSPECTS

It has become accepted that the main pharmaceutical areas where metabonomics is impacting include validation of animal models of disease, including genetically modified animals; preclinical evaluation of drug safety studies; allowing ranking of candidate compounds; assessment of safety in humans in clinical trials; after product launch, quantitation, or ranking of the beneficial effects of pharmaceuticals,

both in development and clinically; improved understanding of the causes of highly sporadic idiosyncratic toxicity of marketed drugs; and patient stratification for clinical trials and drug treatment (pharmaco-metabonomics).

In addition, in terms of disease studies, metabonomics is playing a role in improved, differential diagnosis and prognosis of human diseases, particularly for chronic and degenerative diseases, and for diseases caused by genetic effects. A better understanding of large-scale human population differences through epidemiological studies is also being achieved.

Other applications include nutritional studies, sports medicine, and lifestyle studies, including the effects of diet, exercise, and stress, and evaluation of the effects of interactions between drugs, and between drugs and diet. Additionally, the metabolic profiles of plants, including genetically modified species, and quality control studies (essential equivalence) are now being achieved together with applications in microbiological characterization; forensic science, including the biochemical effects of drug overdose; and environmental effects of pollutants monitored using marker species, both terrestrial and aquatic.

One long-term goal of using pharmacogenomic approaches is to understand the genetic makeup of different individuals (their genetic polymorphism) and their varying abilities to handle pharmaceuticals both for their beneficial effects and for identifying adverse effects. If personalized health care is to become a reality, an individual's drug treatments must be balanced so as to achieve maximal efficacy and avoid adverse drug reactions. Very recently, for the first time, an alternative approach has been proposed for understanding intersubject variability in response to drug treatment using a combination of multivariate metabolic profiling and chemometrics to predict the metabolism and toxicity of a dosed substance, based solely on the analysis and modeling of a predose metabolic profile [57]. Unlike pharmacogenomics, this approach, which has been termed *pharmaco-metabonomics*, is sensitive to both the genetic and the modifying environmental influences that determine the basal metabolic fingerprint of an individual, because these will also influence the outcome of a chemical intervention. This new approach has been illustrated with studies of the toxicity and metabolism of acetaminophen (paracetamol) administered to rats.

A major initiative has been under way to investigate the reporting needs and to consider recommendations for standardizing reporting arrangements for metabonomics studies, and to this end, a Standard Metabolic Reporting Structures (SMRS) group was formed (see www.smrsgroup.org). This group has produced a draft policy document that covers all of those aspects of a metabolic study that are recommended for recording, from the origin of a biological sample, the analysis of material from that sample, and chemometric and statistical approaches to retrieve information from the sample data, and a summary publication has ensued [58]. The various levels and consequent detail for reporting needs, including journal submissions, public databases, and regulatory submissions, have also been addressed. This has been followed up with a workshop and discussion meeting sponsored by the U.S. National Institutes of Health, from which firm plans are being developed to define standards in a number of areas relevant to metabonomics, including characterizarition of sample-related meta-data, technical standards, and related data; meta-data and QC matters for the analytical instrumentation; data transfer methodologies and schema for implementation of such activities; and development of standard vocabularies to enable transparent exchange of data.

NMR- and MS-based metabonomics are now recognized as independent and widely used techniques for evaluating the toxicity of drug candidate compounds, and they have been adopted by several pharmaceutical companies into their drug development protocols. For drug safety studies, it is possible to identify the target organ of toxicity, derive the biochemical mechanism of the toxicity, and determine the combination of biochemical biomarkers for the onset, progression, and regression of the lesion. Additionally, the technique has been shown to be able to provide a metabolic fingerprint of an organism ("metabotyping") as an adjunct to functional genomics and hence has applications in design of drug clinical trials and for evaluation of genetically modified animals as disease models.

Using metabonomics, it has proved possible to derive new biochemically based assays for disease diagnosis and to identify combination biomarkers for disease, which can then be used to monitor the efficacy of drugs in clinical trials. Thus, based on differences observed in metabonomic databases from control animals and from animal models of disease, diagnostic methods and biomarker combinations might be derivable in a preclinical setting. Similarly, the use of databases to derive predictive expert systems for human disease diagnosis and the effects of therapy require compilations from both normal human populations and patients before, during, and after therapy.

It is expected that further integration of the techniques into pharmaceutical R&D activities will continue. One potential advantage of this approach is the transferability in principle of metabolic biomarkers between species and ultimately into humans.

The ultimate goal of systems biology must be the integration of data acquired from living organisms at the genomic, protein, and metabolite levels. In this respect, transcriptomics, proteomics, and metabonomics will all play an important role. Through the combination of these, and related approaches, will come an improved understanding of an organism's total biology, and with this, better understanding of the causes and progression of human diseases and, given the twenty-first century goal of personalized health care, the improved design and development of new and better targeted pharmaceuticals.

## REFERENCES


1. Nicholson J K, Lindon J C, Holmes E (1999). "Metabonomics": Understanding the metabolic responses of living systems to pathophysiological stimuli via multivariate statistical analysis of biological NMR spectroscopic data. *Xenobiot.* 29:1181–1189.
2. Nicholson J K, Connelly J, Lindon J C, et al. (2002). Metabonomics: A platform for studying drug toxicity and gene function. *Nat. Rev. Drug Disc.* 1:153–162.
3. Fiehn O (2002). Metabolomics—the link between genotypes and phenotypes. *Plant Molec. Biol.* 48:155–171.
4. Nicholson J K, Wilson I D (2003). Understanding "global" systems biology: Metabonomics and the continuum of metabolism. *Nat. Rev. Drug Disc.* 2:668–676.
5. Nicholson J K, Holmes E, Lindon J C, et al. (2004). The challenges of modeling mammalian biocomplexity. *Nat. Biotechnol.* 22:1268–1274.
6. Lamers R J A N, Wessels E C H H, van der Sandt J J M, et al. (2003). A pilot study to investigate effects of inulin on Caco-2 cells through *in vitro* metabolic fingerprinting. *J. Nutr.* 133:3080–3084.

7. Bollard M E, Xu J S, Purcell W, Griffin J L, et al. (2002). Metabolic profiling of the effects of D-galactosamine in liver spheroids using $^{1}$H NMR and MAS NMR spectroscopy. *Chem. Res. Toxicol.* 15:1351–1359.
8. Lindon J C, Nicholson J K, Holmes E, et al. (2000). Metabonomics: Metabolic processes studied by NMR spectroscopy of biofluids. *Concepts Magn. Reson.* 12: 289–320.
9. Lindon J C, Holmes E, Bollard M E, et al. (2004). Metabonomics technologies and their applications in physiological monitoring, drug safety assessment and disease diagnosis. *Biomarkers.* 9:1–31.
10. Bino R J, Hall R D, Fiehn O, et al. (2004). Potential of metabolomics as a functional genomics tool. *Trends Plant Sci.* 9:418–425.
11. Sumner L W, Mendes P, Dixon R A (2003). Plant metabolomics: Large-scale phytochemistry in the functional genomics era. *Phytochem.* 62:817–836.
12. Claridge, T D W (1999). *High-Resolution NMR Techniques in Organic Chemistry.* Elsevier Science, Oxford, UK, pp. 384.
13. Liu M, Nicholson J K, Lindon J C (1996). High resolution diffusion and relaxation edited one- and two-dimensional $^{1}$H NMR spectroscopy of biological fluids. *Analyt. Chem.* 68:3370–3376.
14. Nicholson J K, Foxall P J D, Spraul M, et al. (1995). 750 MHz $^{1}$H and $^{1}$H-$^{13}$C NMR spectroscopy of human blood plasma. *Analyt. Chem.* 67:793–811.
15. Khandelwal P, Beyer C E, Lin Q, et al. (2004). Nanoprobe NMR spectroscopy and *in vivo* microdialysis: New analytical methods to study brain neurochemistry. *J. Neurosci. Meth.* 133:181–189.
16. Keun H C, Beckonert O, Griffin J L, et al. (2002). Cryogenic probe $^{13}$C NMR spectroscopy of urine for metabonomic studies. *Analyt. Chem.* 74:4588–4593.
17. Tomlins A, Foxall P J D, Lindon J C, et al. (1998). High resolution magic angle spinning $^{1}$H nuclear magnetic resonance analysis of intact prostatic hyperplastic and tumour tissues. *Analyt. Communicat.* 35:113–115.
18. Garrod S L, Humpfer E, Spraul M, et al. (1999). High-resolution magic angle spinning $^{1}$H NMR spectroscopic studies on intact rat renal cortex and medulla. *Magn. Reson. Med.* 41:1108–1118.
19. Moka D, Vorreuther R, Schicha H, et al. (1998). Biochemical classification of kidney carcinoma biopsy samples using magic-angle-spinning $^{1}$H nuclear magnetic resonance spectroscopy. *J. Pharmaceut. Biomed. Anal.* 17:125–132.
20. Cheng L L, Chang I W, Smith B L, et al. (1998). Evaluating human breast ductal carcinomas with high-resolution magic-angle spinning proton magnetic resonance spectroscopy. *J. Magn. Reson.* 135:194–202.
21. Cheng L L, Chang I W, Louis D N, et al. (1998). Correlation of high-resolution magic angle spinning proton magnetic resonance spectroscopy with histopathology of intact human brain tumor specimens. *Cancer Res.* 58:1825–1832.
22. Wilson I D, Nicholson J K, Castro-Perez J, et al. (2005). High resolution "ultra performance" liquid chromatography coupled to oa-TOF mass spectrometry as a tool for differential metabolic pathway profiling in functional genomic studies. *J. Proteome Res.* 4:591–598.
23. Soga T, Ohashi Y, Ueno Y, et al. (2003). Quantitative metabolome analysis using capillary electrophoresis mass spectrometry. *J. Proteome Res.* 2:488–494.
24. Lindon J C, Nicholson J K, Wilson I D (2000). Directly-coupled HPLC-NMR and HPLC-NMR-MS in pharmaceutical research and development. *J. Chromatogr.* B. 748:233–258.

25. Kaderbhai N N, Broadhurst D I, Ellis D I, et al. (2003). Functional genomics via metabolic footprinting: Monitoring metabolite secretion by *Escherichia coli* tryptophan metabolism mutants using FT-IR and direct injection electrospray mass spectrometry. *Compar. Funct. Genom.* 4:376–391.
26. Lewitt P A, Galloway M P, Matson W, et al. (1992). Markers of dopamine metabolism in Parkinsons disease. *Neurol.* 42:2111–2117.
27. Lindon J C, Holmes E, Nicholson J K (2001). Pattern recognition methods and applications in biomedical magnetic resonance. *Progr. NMR Spectros.* 39:1–40.
28. Keun H C, Ebbels T M D, Bollard M E, et al. (2004). Geometric trajectory analysis of metabolic responses to toxicity can define treatment specific profiles. *Chem. Res. Toxicol.* 17:579–587.
29. Holmes E, Nicholson J K, Tranter G (2001). Metabonomic characterization of genetic variations in toxicological and metabolic responses using probabilistic neural networks. *Chem. Res. Toxicol.* 14:182–191.
30. Kell D B (2002). Metabolomics and machine learning: Explanatory analysis of complex metabolome data using genetic programming to produce simple, robust rules. *Molec. Biol. Rep.* 29:237–241.
31. Cloarec O, Dumas M E, Craig A, et al. (2005). Statistical total correlation spectroscopy: An exploratory approach for latent biomarker identification from metabolic $^{1}$H NMR data sets. *Analyt. Chem.* 77:1282–1289.
32. Bollard M E, Stanley E G, Lindon J C, et al. (2005). NMR-based metabonomics approaches for evaluating physiological influences on biofluid composition. *NMR Biomed.* 18:143–162.
33. Lehtimaki K K, Valonen P K, Griffin J L, et al. (2003). Metabolite changes in BT4C rat gliomas undergoing ganciclovir-thymidine kinase gene therapy-induced programmed cell death as studied by H-1 NMR spectroscopy in vivo, ex vivo, and in vitro. *J. Biolog. Chem.* 278:45915–45923.
34. Nicholls A W, Mortishire-Smith R J, Nicholson J K (2003). NMR spectroscopic based metabonomic studies of urinary metabolite variation in acclimatizing germ-free rats. *Chem. Res. Toxicol.* 16:1395–1404.
35. Wang Y, Holmes E, Nicholson J K, et al. (2004). Metabonomic investigations in mice infected with *Schistosoma mansoni*: An approach for biomarker identification. *Proc. Nat. Acad. Sci. USA.* 101:12676–12681.
36. Nicholson J K, Wilson I D (1989). High-resolution proton magnetic resonance spectroscopy of biological fluids. *Prog. Nucl. Magn. Reson. Spectros.* 21:449–501.
37. Lindon J C, Holmes E, Nicholson J K (2004). Toxicological applications of magnetic resonance. *Prog. Nucl. Mag. Reson. Spectros.* 45:109–143.
38. Phalaraksh C, Lenz E M, Nicholson J K, et al. (1999). NMR spectroscopic studies on the haemolymph of the tobacco hornworm, *Manduca sexta*: Assignment of $^{1}$H and $^{13}$C NMR spectra. *Insect Biochem. Molec. Biol.* 29:795–805.
39. Bundy J G, Lenz E M, Bailey N J, et al. (2002). Metabonomic investigation into the toxicity of 4-fluoroaniline, 3,5-difluoroaniline and 2-fluoro-4-methylaniline to the earthworm *Eisenia veneta (Rosa)*: Identification of novel endogenous biomarkers. *Environ. Toxicol. Chem.* 21:1966–1972.
40. Griffin J L, Walker L A, Shore R F, et al. (2001). High-resolution magic angle spinning $^{1}$H NMR spectroscopy studies on the renal biochemistry in the bank vole (*Clethrionomys glareolus*) and the effects of arsenic ($As^{3+}$) toxicity. *Xenobiot.* 31:377–385.
41. Viant M R, Rosenblum E S, Tjeerdema R S (2003). NMR based metabolomics: a powerful approach for characterizing the effects of environmental stressors on organism health. *Environ. Sci. Technol.* 37:4982–4989.

42. Lindon J C, Nicholson J K, Holmes E, Breau A P, Cantor G H, Bible R H, et al. (2003). Contemporary issues in toxicology: the role of metabonomics in toxicology and its evaluation by the COMET project. *Toxicol. Appl. Pharmacol.* 187:137–146.
43. Keun H C, Ebbels T M D, Antti H, et al. (2002). Analytical reproducibility in $^{1}$H NMR-based metabonomic urinalysis. *Chem. Res. Toxicol.* 15:1380–1386.
44. Ebbels T, Keun H, Beckonert O, et al. (2003). Toxicity classification from metabonomic data using a density superposition approach: "CLOUDS". *Analyt. Chim. Acta.* 490: 109–122.
45. Lindon J C, Nicholson J K, Everett J R (1999). NMR spectroscopy of biofluids. In G A Webb (ed.), *Annual Reports on NMR Spectroscopy*, Vol. 38. Academic Press, Oxford, UK, pp. 1–88.
46. Moolenaar S H, Engelke U F H, Wevers R A (2003). Proton nuclear magnetic resonance spectroscopy of body fluids in the field of inborn errors of metabolism. *Annals Clin. Biochem.* 40:16–24.
47. Maxwell R J, Martinez-Perez I, Cerdan S, et al. (1998). Pattern recognition analysis of $^{1}$H NMR spectra from perchloric acid extracts of human brain tumor biopsies. *Magn. Reson. Med.* 39:869–877.
48. Swanson M G, Vigneron D B, Tabatabai Z L, et al. (2003). Proton HR-MAS spectroscopy and quantitative pathologic analysis of MRI/3D-MRSI-targeted postsurgical prostate tissues. *Magn. Reson. Med.* 50:944–954.
49. Sitter B, Sonnewald U, Spraul M, et al. (2002). High-resolution magic angle spinning MRS of breast cancer tissue. *NMR in Biomed.* 15:327–337.
50. Barton S J, Howe F A, Tomlins A M, et al. (1999). Comparison of in vivo $^{1}$H MRS of human brain tumors with $^{1}$H HR-MAS spectroscopy of intact biopsy samples in vitro. *Magma.* 8:121–128.
51. 't Hart B A, Vogels J T W E, Spijksma G, et al. (2003). NMR spectroscopy combined with pattern recognition analysis reveals characteristic chemical patterns in urines of MS patients and non-human primates with MS-like disease. *J. Neurolog. Sci.* 212: 21–30.
52. Brindle J T, Antti H, Holmes E, et al. (2002). Rapid and noninvasive diagnosis of the presence and severity of coronary heart disease using H-1 NMR-based metabonomics. *Nat. Med.* 8:1439–1445.
53. Waters N J, Holmes E, Williams A, et al. (2001). NMR and pattern recognition studies on the time-related metabolic effects of α-naphthylisothiocyanate on liver, urine, and plasma in the rat: An integrative metabonomic approach. *Chem. Res. Toxicol.* 14:1401–1412.
54. Coen M, Lenz E M, Nicholson J K, et al. (2003). An integrated metabonomic investigation of acetaminophen toxicity in the mouse using NMR spectroscopy. *Chem. Res. Toxicol.* 16:295–303.
55. Coen M, Ruepp S U, Lindon J C, et al. (2004). Integrated application of transcriptomics and metabonomics yields new insight into the toxicity due to paracetamol in the mouse. *J. Pharmaceut. Biomed. Anal.* 35:93–105.
56. Griffin J L, Bonney S A, Mann C, et al. (2004). An integrated reverse functional genomic and metabolic approach to understanding orotic acid-induced fatty liver. *Physiolog. Genom.* 17:140–149.
57. Clayton T A, Lindon J C, Antti H, et al. (2005). Pharmaco-metabonomic phenotyping: A new paradigm for personalized healthcare? *Nat. Biotechnol.* submitted.
58. Lindon J C,, Nicholson J K, Holmes E, et al. (2005). Summary recommendations for standardization and reporting of metabolic analyses. *Nat. Biotechnol.* 23:833–838.

# BIOTERRORISM

DANY SHOHAM
*Begin-Sadat Center for Strategic Studies, Bar Ilan University, Ramat-Gan, Israel*

**Chapter Contents**

1 Preface
2 Definition and Essence of Bioterrorism
3 Concept and Incentives
4 Bioterrorism-Patterned Natural Epidemics
  4.1 Ebola Hemorrhagic Fever
  4.2 AIDS
  4.3 SARS
  4.4 West-Nile Encephalitis
  4.5 Avian Influenza
  4.6 Cholera
5 Significance of Bioterrorism-Patterned Natural Epidemics
6 Spectrum of Applicable Pathogens and Toxins
7 Bacterial Agents
  7.1 Plague
  7.2 Legionnaires Disease (Legionellosis)
  7.3 Brucellosis
  7.4 Tuberculosis (TB)
8 Viral Agents
  8.1 Rift Valley Fever (RVF)
  8.2 Yellow Fever
  8.3 Norwalk
  8.4 Hepatitis A
9 Toxins
  9.1 Botulism Toxin
  9.2 Staphylococcal Enterotoxin B (SEB)
  9.3 T-2 Toxin
  9.4 Aflatoxin
  9.5 Aconitine

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

10 Availability of Infective Agents and Toxins
11 Technical Feasibility of Biosabotage Acts
12 Operational Modes
13 Spatially and Temporally Varying Impacts
14 History and Evolution
15 State-Sponsored Bioterrorism
15.1 Germany
15.2 Japan
15.3 United States
15.4 South Africa
15.5 USSR and Russia
15.6 Iraq
15.7 Iran
16 Nonstate-Sponsored Bioterrorism
16.1 al-Qaeda
16.2 Jemaah Islamiah
16.3 Hezbollah
16.4 Radical Palestinian Terror Organizations
16.5 Aum Shinrikyo Cult
16.6 Bhagwan Shri Rajneesh Cult—Typhoid Epidemic
16.7 Shigella Dysenteriae Type 2 Outbreak
16.8 Hepatitis A Outbreak
16.9 Ascaris Suum Outbreak
17 The Anthrax Letters
18 The Ricin Course
19 The Smallpox Complex
20 Agroterrorism
21 Narcoterrorism
22 Novel Agents
22.1 Anti-Organism Power Multipliers
22.2 Supremely Engineered Bacteria and Viruses
22.3 Anti-Inanimate Object Power Multipliers
22.4 Bioengineering Availability
23 Exploitability of Biotechnology and Biomedicine
24 Preventative and Countermeasures
25 Projections and Prognosis
References

## 1 PREFACE

*The very fundamental characteristics of bioterrorism are presented, alongside with the distinct multidisciplinary configuration of the term and phenomenon. Some unique attributes are outlined that make bioterrorism an entity nourished by different essentials and bearing various, in part far-reaching, implications.*

Infectious diseases—the most effective extraneous regulator controlling the size of many animal populations in nature—are largely restrained in relation to mankind, thanks to the advancement of preventive and therapeutic medicine. The balance thus created is at any rate fragile, especially as human population density is steadily increasing, overall, whereas pathogens persist in acquiring drug-resistance. Pharmaceutical biotechnology is one major contributing arena to medicine, in that context; bioterrorism constitutes a counter-arena that unfortunately may amplify or substitute for the natural control mechanisms signified for by infectious diseases. Paradoxically, yet, pharmaceutical biotechnology in itself represents a potential key resource for bioterrorism. This is the case because pharmaceutical biotechnology is becoming an unparalleled front area, bearing a melting pot of remarkably sophisticated and multidisciplinary know-how, substances, and appliances. Large portions of which are dual use and may readily be exploited decently or viciously.

This built-in, intriguing duality is an essential feature, in that bioterrorism may imbibe extra potency from the very capacity destined to prevent and cope with bioterrorism. Although a very old way of sabotage, bioterrorism is presently, or soon to be, immeasurably more powerful, potentially, than ever, due to marked breakthroughs achieved in biopharmacology and biomedicine. Scientifically, this evolving, dichotomic course is inevitable, albeit significantly augmenting bioterrorism resources. The comprehension and practice allowing for combating pathogens and toxins may serve then, as well, to upgrade bioterrorism. It is becoming but a matter of their degree of availability of those bioagents. The same principle applies for pathogens of farm animals and plants, which form an additional dimension of bioterrorism—agricultural bioterrorism (agroterrorism). Also, the increasing availability of recreational drugs—mostly a sort of phytotoxins, in their essence—propels another terrorism variant—narcoterrorism.

In a sense, then, bioterrorism may embody the golem that rose against its creator. It can trigger a confined outbreak, an extensive epidemic, or a colossal pandemic, intentionally engined by man himself. The perpetrators may not even figure the anticipated impact—medically, psychologically, and logistically—which may immensely exceed what they want to achieve. Inversely, the in-effect impact may be but marginal, much less than that quested for. This uncertainty stems from various unknowns still marking the behavioral profile of infectious diseases. The interface between pathogens and man retains its enigmatic identity, although largely deciphered. At any rate, as James Woolsey, former CIA director has put it: "Germ terrorism is the single most dangerous threat to our national security in the foreseeable future." The magnitude of bioterrorism, yet, extends far beyond the boundaries of the United States.

Man only can eradicate, either locally or globally, long-lasting formidable pathogens—as was the case with smallpox—by means of extraneous, far-reaching intervention, namely vaccination; antibiotics are not less important for wiping out certain bacterial pathogens. Yet man is, at the same time, the prime messenger of their possible reemergence, particularly through bioterrorism. Conceivably, his intelligence and impulse both shape, thus, a delicate balance, one that might branch unthinkable ramifications. It is, then, his own and absolute responsibility to pursue a steady state that would eliminate or minimize the risk of epidemic self-destruction. In general, calculated self-destruction is unsupprisingly, antagonistic to the very nature of most species.

The menace of terrorism at large, in its broad sense, is an unparalleled phenomenon that may be considered, basically, to be a sort of a limited self-destruction mechanism for mankind. In a way, it is even worse then wars, because of its inherent unpredictability, irrationalism, and wide scope, as well as the built-in, unapparent, vitality, and durability marking it. When the dimension of biological warfare is added to this featuring, it turns out to pose uncontrollable potentiality.

Addressing the complicated domain of bioterrorism has to necessarily be conducted through a multidisciplinary approach. Therefore, the following chapter covers varitype aspects altogether, namely, microbiological, epidemiological, medical, technological, demographical, political, and strategic; the integration of which is essential, so as to comprehend the substance, complexity, and implications of bioterrorism. The entire issue of bioterrorism may thus gain a typical multidimensional structure and meaning.

## 2 DEFINITION AND ESSENCE OF BIOTERRORISM

*Integrating two spheres that are innately unconnected to each other—terrorism and pathogens—the entity of bioterrorism reflects a complex, although in a sense fairly old, phenomenon of mankind. Given this starting point, the definition and essence of bioterrorism, together with its meaningfulness, are complicated. At the same time, the protean nature of bioterrorism endows it with uncommon attractiveness and advantageousness that may meet various courses.*

The two elements comprising bioterrorism, namely terrorizing by means of biological agents, represent two distinct spheres, each bearing multiple contents. Those two spheres form, thereby, a singular, mighty conjunction between strategic studies and life sciences.

Terrorism at large, can be variedly featured, according to the following parameters:

- Conducted by a country, organization, group, or individual.
- Against a country, organization, group, or individual.
- By threatening and/or attacking.
- Covertly or overtly.
- Terror may be the goal or a by-product of an in-effect damage.
- The direct impact may be the ultimate objective or propel the occurrence of the ultimate objective.
- The impact is intended to form in the short, medium, or long run.
- With or without taking responsibility by the perpetrator.

Biological agents include, basically, pathogens and toxins, which may be classified in the following ways:

- Live—hence reproducing—agents (pathogens) or nonviable agents (toxins). Prions, as reproducing molecules, constitute a notable, exceptional intermediate.

- Lethal or sublethal (not a clear-cut division).
- Transmissible—hence epidemic—pathogens or nontransmissible pathogens.
- Natural or modified/engineered pathogens and toxins.
- Abruptly or gradually affecting the target.
- Affecting humans, husbandry, crops, or materials.
- The vehicle may be natural (infected arthropod, animal, or human being) or artificial (man-made disseminator, i.e., sprayers and envelopes).
- The route of penetrating the body is the respiratory system, alimentary tract, or skin.

Remarkably, almost any combination of the various mentioned parameters featuring terrorism and biological agents is feasible. That is one main attribute underlying the attractiveness and potential might of bioterrorism. Another one is the outstanding ratio between the amount of biological agent to be used and the resulting impact, especially with reference being made to epidemic pathogens. At the extreme, which is certainly feasible, an individual saboteur can disperse a scarcely existent, unapparent amount of a fully epidemic, virulent pathogen, be it pestilence, smallpox, or influenza, and give rise to a transgressing, questionably controllable, lethal plague.

Biocrimes, biosabotage, and biowarfare are terms appreciably dovetailing with bioterrorism, in different manners. Although biological crimes accentuate the illegal dimension of bioterrorism, they pertain, as well, to a variety of acts, such as the very holding or transferring of a certain microorganism, which are in violation of national or international rules and conventions. Biological sabotage is the in-effect employment of biological agents for whatever operational sabotage purpose, whether or not terrorism-oriented, usually by means of guerilla warfare. Biological warfare basically reflects a military confrontation in which biological weaponry is used; yet broadly, it may be waged against civilian target populations, thus explicitly having the quality of bioterrorism.

Bioterrorism may be perceived, overall, as the calculated, deliberate use of pathogens or toxins (or threat of using them) against civilian populations or economic/logistical infrastructures, in order to attain goals that are political, social, religious, financial, ideological, or personal in nature; this is done through intimidation or coercion or instilling fear. It may still be subclassified, then, as follows:

- The unlawful release of biological agents or toxins with the intent to intimidate or coerce a government or civilian population to further political or social objectives. Humans, farm animals, and cultivars are often targets.
- Use of microorganisms or toxins to kill, sicken, or cause other malfunction in people, animals, plants, or useful materials.

In its broad scope, thus, bioterrorism includes, beyond its classic targeting against humans:

- Agroterrorism, which may involve anti-plant pests, in addition to anti-plant and anti-animal pathogens.

- Narcoterrorism, namely the constant on-purpose input of recreational drugs onto a target population.
- Any other type of sabotage achievable by means of microbes, such as fuel-eating bacteria, asphalt eating-bacteria, and alike.

In a way, the core of bioterrorism, both objectively and subjectively, is horrifying, whereas the involved weapon in itself consists in, absurdly, some of the most powerful natural foes of humanity, and can in practice bring about a very significant direct impact, far beyond horror. Hence, it may regretfully be regarded, in a sense, as the potential promoter of various infectious diseases that are currently being suppressed, pharmacologically and medically, just waiting for their opportunity to come.

One chief character of bioterrorism is, nonetheless, the conceivably immeasurable disproportion between effort and outcome, particularly when the latter is at its maximum. It thus poses a strong temptation. Any other weapons, terror means—and apparently any other context—are not competent in sustaining such disproportion. Technically, then, this attribute mostly represents the essence of bioterrorism, such that nears a legendary identity. The distance from reality is at any rate secondary, especially as it can readily be bridged over, in affect, quite soon. In that concern, recent breakthroughs made in the field of genetic engineering amplify at the same time, paradoxically, both imagination and practicality.

All in all, within the sphere of terrorism at large, three distinctions typify the essence of bioterrorism:

1. The distinction between bioterrorism and conventional terrorism.
2. Between bioterrorism and other unconventional terrorism (chemical, radiological, nuclear).
3. Between bioterrorism and biological warfare.

The projection of those three distinctions altogether reflects the whole substance of bioterrorism, as aimed, among other things, to be shown in this chapter.

## 3 CONCEPT AND INCENTIVES

*Several common denominators are outlined, which mark both terrorism-oriented states and terror organizations, on their way to resorting to bioterrorism. Other elements of their concept and incentives regarding bioterrorism comparatively differ, naturally, owing to their dissimilar essences. Various motives and objectives may then underlie their encouraged attitude toward bioterrorism. The roots are at any rate historical, appreciably shaping during time.*

Intuitively, not having any concrete knowledge in terms of toxinology and bacteriology until the nineteenth century, man adopted poisonous and infective substances for attacking rivals. It was mainly practical common sense that guided him. Much later, paving its way toward scientific horizons, man gradually mastered the related technical fundamentals. The recognition of infective microorganisms and naturally produced toxic molecules accumulated and ripened. It reached the level of having the full capacity to operate them as warfare agents.

But, anthropologically, the concept underlying bioterrorism extends far beyond technical and operational features, whether state or non-state sponsored. The very notion of afflicting passive civilian populations with infectious diseases seems to be, objectively, profoundly evil, even if accompanied by different considerations and calculations. In its simplest form, then, bioterrorism reflects but an extreme way of malevolence. Variably, the dimensions of governance and insidiousness may certainly play a roll, in conjunction. *A priori*, the power to generate epidemics was possessed only by God and nature. Thus, biblically, God waged, initially, the epidemics included within the Ten Plagues of Egypt. Later on, comprehending the principal mechanism and impact of epidemics, man tried to emulate and immobilize that route against adversaries. As a matter of fact, it was around 1500 BC, when the Hittites sent plague victims into the lands of their enemies [1].

The fundamental incentive and concept of bioterrorism did not substantially change, up to these days. What did alter, obviously, are the know-how, the technical tools, and the strategic approach. Strategically, there are three basic alternatives shaping the concept of bioterrorism acts:

- The perpetrator (either state or non-state sponsored) would seek to remain unidentified.
- The perpetrator would be indifferent as to its identity.
- The perpetrator would seek to be identified.

The first and third alternatives have each two variants, respectively:

- Anonymity would rely on objective inability to epidemiologically determine whether the event was an act of bioterrorism.
- Anonymity would rely on objective inability to trace the perpetrator by means of investigation or intelligence (even if epidemiologically affirmed to be an act of bioterrorism).
- Identification would rely on direct announcement, either before or subsequent to the act of bioterrorism.
- Identification would rely on investigative and intelligence findings—either evidential or, deliberately, merely circumstantial (hence less valid)—without taking responsibility.

Several technological, political, and psychological explanations underly the quest of terrorists and terrorism-oriented states to opt for acquiring weapons of mass destruction (WMD), particularly biological agents, among them are the cheap cost of such weapons and the sense of prestige and security they grant to the owner. Parallel to offensive considerations, the inverse dimension of deterring capacity also plays a roll. All in all, terrorists or terrorism-oriented states may choose bioterrorism among many options, both conventional and unconventional. Preference may be given to bioterrorism due to various reasons. In terms of a pragmatic, clinical feasibility study, bioterrorism is, at least potentially, the most efficient form of sabotage, particularly when conducted indistinguishably from a natural event. Moreover, it may have a fully strategic impact; hence, it is often reckoned to be the ultimate mode of terrorism. Notably, most biological agents are not contagious; yet

they might generate extensive outbreaks, if not epidemics, due to their transmissibility in other ways.

A terrorist group or terrorism-oriented state, intending to employ biological agents, will likely be attracted to their mass killing potential [2]. In addition, their delayed impact may also enable the saboteur to escape detection. Also, an attacker might reckon the resemblance between a natural and an unnatural epidemic to be close enough to divert suspicion, or at least make it extremely difficult to trace [3]. That would make it difficult to retaliate with a massive punitive strike, the expectation of which ordinarily serves as deterrent to an unconventional assault. Thus, the subsequent intimidation and chaos resulting from the inherently unknowable may make biological agents the terrorist weapon of choice [4]. On the whole, the related line of terms accentuating the significance and potential impact of bioterrorism at large include, then: high consequence terrorism; catastrophic terrorism; total terrorism; asymmetric terrorism; superterrorism; ultraterrorism, and so on.

Beyond, terrorists and terrorism-oriented states are certain to decide for themselves whether the conduction of bioterrorism serves their cause. Moreover, it is believed that biological weapons are preferable for terrorist groups, and that the allure for terrorists of biological weapons is currently intensifying due to the increase in availability of biotechnologists and sophistication of manufacturing methods [5].

Overall, one fundamental motive of bioterrorism (within the context of WMD at large), then, is the will to generate what is currently called high consequence terrorism, total terrorism, or catastrophic terrorism. Under certain circumstances it may be regarded as pure asymmetric terrorism, the purpose of which is not to take lives or destroy property. It is rather the mechanism by which the attacker seeks to achieve weakening of the sense of cohesion that binds communities together, to reduce its social capital, and to sow distrust, fear, and insecurity. Within this narrow course, it is but a method of waging a sort of social and psychological warfare.

The availability of pathogenic agents, alongside with technical feasibility and operational modality of bioterrorism at large, are basically rather in favor of the saboteur or terrorism-oriented state. Still, the bioterrorism events that already took place overall, seemingly exhibit but a portion of its potentiality, whereas the essence of the related threshold of proliferation is vague. Resultantly, built-in limitations interfere with thoroughly comprehending and foreseeing the course of bioterrorism.

## 4 BIOTERRORISM-PATTERNED NATURAL EPIDEMICS

*The apparent resemblance between natural epidemics and induced ones mostly leads to recognizing the relevance of bioterrorism-patterned natural epidemics as prime demonstrative occurrences. Let alone, that various infectious diseases, chiefly viral, are emerging or reemerging in an unpredictable, often overwhelming, manner. It would therefore be advisable to diligently observe some prominent natural epidemics that took place during the recent decade.*

The most distinct bioterrorism-patterned natural epidemics are typically those concerned with water- and food-borne infectious gastroenteritis, stemming from

the contamination of a current collective source. Such epidemics would include, principally, two phases:

- The first wave of infections is generated directly by the consumption of the collective source.
- An additional wave is brought about, dependent on the degree of ongoing communicability or contagiousness of the causative agent toward potential secondary victims that are prone to be exposed, irrespective of the initial collective source.

This type of epidemic locally occurs very often, worldwide, in the form of relatively restricted innocent outbreaks. They bear the potency to expand, however. And they are quite easy to be produced deliberately, in practice. But, at any rate, they represent the simplest bioterrorism-patterned natural epidemics.

Bioterrorism-patterned natural epidemics might be by far more complex and hazardous, nonetheless. It so happened that the past 10 years were marked by some outstanding epidemic events perfectly illustrating the tremendous potentiality of biological agents at large, as well as the candidature of certain pathogens for acts of bioterrorism, in particular. Those events are presented here in detail, then, so as to demonstrate the general effectuality of various bioterrorism-patterned scenarios. They include six virus- and germ-propelled catastrophic epidemics.

### 4.1 Ebola Hemorrhagic Fever

The 14th of April, 1995, Zaire (now the Democratic Republic of Congo). A 36-year-old laboratory technician checked into the medical clinic in Kikwit, complaining of a severe headache, stomach pains, fever, dizziness, weakness, and exhaustion [6].

Surgeons did an exploratory operation to try to find the cause of his illness. To their horror, they found his entire gastrointestinal tract was necrotic and putrefying. He bled uncontrollably and within hours was dead. By the next day, the five medical workers who cared for him, including an Italian nun who assisted in the operation began to show similar symptoms, including high fevers, fatigue, bloody diarrhea, rashes, red and itchy eyes, vomiting, and bleeding from every body orifice. Less than 48 hours later, they, too, were dead, and the disease turned to form a deadly, uncontrollable epidemic.

As panicked residents fled into the bush, government officials responded to calls for help by closing off all travel, including humanitarian aid into or out of Kikwit, about 400 km from Kinshasa, the national capital. Fearful neighboring villages felled trees across the roads to seal off the pestilent city. No one dared enter houses where dead corpses rotted in the intense tropical heat. Boats plying the adjacent Kwilu River refused to stop to take on or discharge passengers or cargo. Food and clean water became scarce. Hos-pitals could hardly function as medications and medical personnel became scarce.

Deadly tropical fevers are an unfortunate fact of life in Central Africa but rarely are they this contagious or lethal. The plague that afflicted Kikwit was Ebola hemorrhagic fever, a viral disease for which there is no known treatment. Within a few weeks, about 400 people in Kikwit had contracted the virus and 350 were dead.

Where a 10% mortality rate is considered high for most infectious diseases, Ebola kills up to 90% of its victims, usually within only a few days after exposure. Still, contagiousness seems to be low.

The Kikwit Ebola outbreak was neither the first nor the last appearance of this dread disease. The first recognized Ebola epidemic occurred in 1976 in Yambuku, Zaire (near the Ebola River, after which the virus was named), where at least 280 people died. Three years later, 22 patients died in Sudan from a slightly different and less virulent form of Ebola. In 1996, about 100 people were killed by Ebola in two separate episodes in Gabon, and in 1999, an outbreak in the gold mining town of Durba in the Democratic Republic of Congo killed at least 63 people.

Ebola is one of two members of a family of RNA viruses called the Filoviridae (the other one being Marburg fever). No host or vector is known for Ebola, but it has been observed that monkeys and other primates can contract related diseases. Why viruses remain peacefully in their hosts for many years without causing much more trouble than a common cold, but then erupt sporadically and unpredictably into terrible human epidemics, is a new and growing question in environmental health and risk assessment. Although certainly a formidable pathogen, this virus exhibits changing rates of virulence (which is at times extremely high, more than any other pathogen), communicability, and, hence, epidimicity, the latter being rather restrained, in effect (400 infected and diseased people, out of 600,000 residents in Kikwit). Therefore, a shift in the transmissibility level of Ebola virus may be crucial and is immensely feared.

## 4.2 AIDS

AIDS, or acquired immunodeficiency syndrome, is a fatal disease caused by a rapidly mutating retrovirus that attacks the immune system and leaves the victim vulnerable to infections, malignancies, and neurological disorders. It was first recognized as a disease in 1981. The virus was isolated for the first time not long ago, in 1983. It is steadily spreading. During 2004, as a representation, around five million adults and children became infected with HIV (human immunodeficiency virus), the virus that causes AIDS. By the end of the year, an estimated 39.4 million people worldwide were living with HIV/AIDS. The year also saw more than three million deaths from AIDS, despite the availability of HIV antiretroviral therapy, which reduced the number of deaths in high-income countries. African and South-American tropical regions are heavily afflicted in particular for long periods of time.

But in certain regions the disease has risen remarkably swiftly during recent years. One of those arenas is Leningrad, were a notably exponential AIDS epidemic started in 1997. Beginning in that year and proceeding up to 2001, all patients in the Narcology Hospital of Leningrad Regional Center of Addictions (LRCA), were tested for HIV antibody [7]. These clinical records (i.e., serostatus, gender, age, and addiction) and data from the HIV/AIDS Center in the Leningrad Region were reviewed.

It has thus been shown that HIV prevalence at the LRCA increased from 0% to 12.7% overall, 33.4% among drug-dependent patients, and 1.2% among alcohol-dependent patients. During the same 5-year period (1997–2001), 2826 persons were registered at the HIV/AIDS Center: 6, 6, 51, 780, and 1983 persons in 1997, 1998, 1999, 2000, and 2001, respectively.

So, HIV infection is exploding in the Leningrad Region, currently in injection drug users, but potentially more broadly. The known high-per-capita alcohol intake in Russia heightens concern regarding the sexual transmission of HIV. Overall, the number of HIV-infected people in the Leningrad Region has risen to 2929 by the beginning of 2001, which is 2.3 times more than in 2000. Young people aged from 14 to 30, mostly drug addicts, account for 85% of the overall number of HIV-infected people. The number of teenagers sick with AIDS has risen by 2.5 times since last year. Particular concern was caused by a 10-fold increase in the number of AIDS victims among pregnant women. A total of 350 births given by HIV-infected mothers has been registered in the past year, and in 10 cases, parents gave up their babies [8].

However, the AIDS pandemic, in its broad essence, is actually a global long-term eroding factor toward mankind at large. Caused by a type of attritional bioterrorism agent, analogicolly, its ultimate real impact is expected sometime in the future, unforeseeably. For now, and in relation to the magnitude of humanity, it is just crawling, slowly yet steadily. It may be regarded as an inherent, low-rate—although undefeatable, for the time being—cryptic, ostensibly overt pandemic, accumulating its critical mass.

### 4.3 SARS

The SARS (sever acute respiratory syndrome) virus emerged—and possibly rather formed, somehow, little earlier, in reality—on the 16 of November 2002. In Foshan, southern China, on the Pearl River Delta, the first case of SARS involved a man, a "super-infector" who subsequently infected four others [9].

It soon exploded. By November 31, 2002, a deadly outbreak of SARS carried on in China and has killed at least 34 people in the south and three in Beijing. Hundreds have been infected. Clearly an infectious disease, the illness was named and registered only clinically, owing to the unidentified nature of its casual agent. It was certainly recognized to be a virulent virus, at any rate. At that stage, the new virus seemed to be fairly confined. But during December, the disease "hitch-hiked" 200 km to Heyuan city, the provincial capital of Guangzhou. Virus spread to five hospital workers. Panic ensued in Heyuan, and the local authorities tried to calm the people with news that there was not epidemic. Within days, the alarm subsided. Somewhere during this time, provincial party leaders curbed the media from publishing details of the disease, hindering further efforts to contain the virus. By the end of December 2002, Guangdong had reported at least 300 cases.

Although showing to be a new virus, a comprehensive genomic analysis has established that the relationship between SARS Coronavirus and group 2 CoVs is monophyletic [10]. Apparently, then, the SARS virus evolved as a strange zoonotic offshoot of group 2 CoVs, bearing greatly enhanced infectivity toward man.

Actually, SARS is the first major novel infectious disease to hit the international community in the twenty-first century. It originated in southern China in November 2002 and spread rapidly thereafter to 29 countries/regions on five continents. At the end of the epidemic, the global cumulative total was 8098 with 774 deaths. Seven Asian countries/regions were among the top 10 on the list. It has alarmed the world with its infectivity and significant morbidity and mortality, its lack of a rapid, reliable diagnostic test, and lack of effective specific treatment and

vaccination. The adverse impact on travel and business around the world, particularly in Asia, has been enormous [11].

The consequences of the SARS epidemic were remarkable. In Taiwan, as an example, the impact of the SARS epidemic on the utilization of medical services was studied [12]. Using interrupted time-series analysis and National Health Insurance data between January 2000 and August 2003, this study assessed the impact of SARS epidemic on medical service utilization in Taiwan. At the peak of the SARS epidemic, significant reductions in ambulatory care (23.9%), inpatient care (35.2%), and dental care (16.7%) were observed. People's fears of SARS appear to have had a strong impact on access to care. Adverse health outcomes resulting from accessibility barriers posed by the fear of SARS are significant.

### 4.4 West-Nile Encephalitis

In late August of 1999, an unusual cluster of severe cases of encephalitis was noted in an area around Queens in New York City [13]. An epidemiologic investigation by the New York City Department of Health identified eight such cases and revealed that all of the patients had been previously healthy, had resided within the same 16 square mile area, and had recently engaged in outdoor activities. All but one had developed severe acute flaccid paralysis in the setting of encephalitis. At that point, the ethological agent had not yet been deciphered, though it seemed to be a basic virulent virus.

Both before and during the human encephalitis investigation, an epizootic among birds associated with a high fatality rate had been noted in and around New York City [14]. Pathologic assessment of the dead birds displayed involvement of multiple organs, including evidence of encephalitis; however, common avian pathogens were not detected [15]. The diseased birds were initially felt to be unrelated to the human epidemic. But genomic analyses using polymerase chain reaction and genome sequencing with specimens from New York City birds, infected mosquitoes collected in Connecticut, and human brain tissue from a fatal case of encephalitis, as well as expanded serological testing of specimens from suspected human cases identified WNE virus, a mosquito-borne pathogen, as the etiologic agent of this outbreak—4 weeks after the outbreak in humans was first reported to New York City public health officials [16–19]! By the end of the summer of 1999, 62 patients with serologic evidence of acute WNE virus infection, including 59 hospitalized patients, had been identified. Seven died. Similar to more recent outbreaks of the virus, the epidemic seemed to be associated with a high rate of CNS involvement and a preponderance of cases in patients older than 60 years [20].

The introduction of WNE virus in North America was followed by progressive spread throughout the United States. During the summer of 2000, 21 cases of human WNE virus illnes, were identified, 2 died occurred among 10 counties in northeastern states [21]. The following year, 66 cases were detected among a much more widespread geographic area, involving 38 counties in 10 states. The spread of human cases seemed to follow avian deaths; thus, avian death surveillance and, to a lesser extent, mosquito pool surveillance became important parts of public health efforts to track the virus and predict potential human cases [21]. In addition to avian and human illness, a substantial number of equine cases were documented throughout the United States, a pattern also being observed in other parts of the world.

An eco-epidemiological critical mass has been neared in 2002. During the summer of 2002 the number of WNV cases in North America ascended, and was the largest outbreak of arboviral meningoencephalitis ever documented in the Western Hemisphere. The WNE virus expanded its geographic range from the Mississippi River area at the conclusion of the 2001 season to the Pacific Coast by the end of 2002. Virus activity considerably increased and then found expression in the following figures: 2002—1460 cases; 66 deaths; 2003—9862 cases; 264 deaths; 2004—2470 cases; 88 deaths.

### 4.5 Avian Influenza

During 1997, an epidemic in Hong Kong of avian influenza in chickens and ducks occurred and was—merely marginally, then—transmitted to humans. Fortunately, the virus did not move from person to person and seemingly died out after the reservoir of domestic birds was completely depleted by the killing of approximately 1.6 million chickens, ducks, and geese by Hong Kong authorities. A variant of the type A H5N1 flu virus has been identified to be the cause of that event [22].

It was but an ostensible defeat of that epidemic H5N1 variant. In actuality it silently proliferated spatially, and then catastrophic outbreaks of influenza H5N1 took place among poultry in many countries in Asia during late 2003 and early 2004. At that time, more than 100 million birds in the affected countries either died from the disease or were killed to try to control the outbreak. (Thailand 36 million, Vietnam 36 million, Indonesia 15 million, China 5 million, Pakistan 4 million) (Cambodia, Laos South Korea, and Japan have been less affected, so far.)

By March 2004, the outbreak was reported to be under control. Beginning in late June 2004, however, new deadly outbreaks of influenza H5N1 among poultry were reported by several countries in Asia (Cambodia, China, Indonesia, Malaysia [first-time reports], Thailand, and Vietnam). These outbreaks—since extended—have been ongoing, and in 2005 reached Europe. They reached Africa in 2006.

The H5N1 virus does not usually infect humans. In 1997, however, the first case of spread from a bird to a human being was observed during the outbreak of bird flu in poultry in Hong Kong. The virus caused severe respiratory illness in 18 people, 6 of whom died. Since that time, there have been other cases of H5N1 infection among humans. Most recently, human cases of H5N1 infection have occurred in Thailand, Vietnam, and Cambodia during large H5N1 outbreaks in poultry. The death rate for these reported cases has been about 50%. Most of these cases occurred from contact with infected poultry or contaminated surfaces. Remarkably, however, it is thought that a few cases of human-to-human spread of H5N1 have occurred.

So far, spread of H5N1 virus from person to person has been rare, and proliferation has not continued beyond one person. Yet, because all influenza viruses have the ability to change, there is a great concern that the H5N1 virus could one day be able to become contagious and spread easily from one person to another. Because these viruses do not commonly infect humans, there is little or no immune protection against them in the human population. If the H5N1 virus were able to infect people and spread easily from person to person, a new flu pandemic could probably begin. It is impossible to foresee when a pandemic might occur. However, experts from around the world are watching the H5N1 situation in Asia very closely

and are preparing for the possibility that the virus may begin to spread more easily and widely from person to person. Still, the H5N1 strain is but one variant within an immeasurable natural gallery of avian flu viruses. A few of them have already infected man, whereas the majority are taking their time, for now. A highly communicable derivative of them or of the H5N1 type will most probably generate the next colossal pandemic.

Until now, the most demanding flu pandemic that has been documented was the "Spanish Flu" (also known as "Swine Flu"), which killed 40–90 million people during just two years (1918–1919). The virus was of the H1N1 subtype. Later on, in 1957 a new serotype, H2N2, emerged and generated another pandemic, substituting the H1N1 virus. The next pandemic was given rise to in 1968 by the serotype H3N2, which replaced the H2N2. In 1977, a mild H1N1 virus resurfaced, bringing about another pandemic, but it did not substituted the H3N2 virus, and they are both prevailing currently. As a matter of fact, all pandemic and any other influenza virus strains are derived from avian viral strains. Moreover, at least some of them are likely preserved in perennial lake ice, mostly in the Northern Hemisphere, and may reintroduce—subsequent to ice thawing—genes or entire genomes onto the current genetic pool of that protean virus [23]. In regard to the current H5N1 strain, which is considerably virulent but noncontagious, for now, it may attain transmissibility by even a minor genetic alteration, the full essence of which is not yet clear.

In case this virus would retain its ongoing high communicability and virulence toward farm birds (currently chickens and ducks mainly) concomitantly with such genetic alteration, then an unparalleled pathogen, both pandemic and epizootic, may rise. Theoretically, it could disease pigs and horses as well, as the influenza A virus is in general infective and at times virulent toward those domestic mammals. But apparently this cannot occur in actuality, because there is, in all likelihood, some inherent viral incompetence to have such amplified affinity toward more then one host simultaneously. The remarkable potentiality of influenza A virus to attack various farm animals signifies, at any rate, its distinct candidature in the context of agroterrorism (see below). Yet even in terms of merely an anti-human bioterrorism weapon, this varitype virus has been reckoned to be an ultimate pathogen, in some respects even exceeding the formidability of the notorious smallpox virus [24].

Pandemic spread would presumably bring about 60–80 million deaths worldwide. According to US Homeland Security Council's estimates, 90 million people could be infected in the US alone in case a bird flu pandemic breaks out, of which 45 million will require medical care, 10 million will require hospital admission and 2 million people could die [25].

A British government report warns that The health service will be plunged into chaos if Britain is struck by a bird flu pandemic. Faced with a possible 4.5 million victims, demand for hospital beds would outstrip supply and doctors might have to deny treatment to the sick and elderly to save younger, fitter patients [26].

Not casually, all of the above-detailed illustrative epidemics represent, basically, bioterrorism-patterned scenarios specifically propelled by viruses. Antibiotics, the main combat-tool against germs, are of no relevance against viruses. Antiviral drugs are very limited in their scope. Vaccines are efficient against most viruses, but they cannot cope with viruses that are currently dynamically altering. Antisera against viruses are of marginal pharmacological contribution, for now. Viruses pose, then, in a sense, the ultimate biological weapon, and can be used for bio-

terrorism purposes. Still, the cultivation and handling of viruses is complicated in comparison with bacteria, due to the necessity of host tissue, whereby viruses obligatorily propagate.

### 4.6 Cholera

But, beyond viruses, other classes of pathogens and toxins are not at all to be depreciated, certainly. The germ causing cholera—mostly a water-borne disease—is nowadays the most epidemic and demanding bacterial pathogen, with reference being made to acute illness (parallel to the inversely lingering course of illness of tuberculosis, another remarkably demanding bacterial disease). In contrast to viruses, the cardinal characteristic marking pathogenic bacteria is their interface with various antibiotics, which is a very tangled and dynamic one. Thus, during March and April 2002, a resurgence of the causative agent *Vibrio cholerae* O139 occurred in Dhaka and adjoining areas of Bangladesh with an estimated 30,000 cases of cholera [27]. The reemerged O139 strains belong to a single ribotype corresponding to one of two ribotypes that caused the initial O139 outbreak in 1993. Unlike the strains of 1993, however, the recent strains are susceptible to trimethoprim, streptomycin, and sulphamethoxazole, but resistant to nalidixic acid. This alternating profile exemplifies a predominant intensifying property of various bacterial pathogens resisting antibacterial drugs in an unpredictable fashion.

In general, cholera is an acute intestinal infection caused by ingestion of contaminated water or food. It has a short incubation period, from less than 1 day to 5 days, and produces an enterotoxin that causes copious, painless, watery diarrhea that can quickly lead to severe dehydration and death if treatment is not promptly given. Vomiting also occurs in most patients. Man-made and natural disasters can intensify the risk of epidemics considerably, as can conditions in crowded refugee camps. Explosive outbreaks with high case-fatality rates are often the result. For example, in the aftermath of the Rwanda crisis in 1994, outbreaks of cholera caused at least 48,000 cases and 23,800 deaths within one month in the refugee camps in Goma, the Congo [28]. Although rarely so deadly, cholera outbreaks continue to be a major public health concern, causing considerable socioeconomic disruption as well as loss of life. In 2001 alone, the WHO and its partners in the Global Outbreak Alert and Response Network participated in the verification of 41 cholera outbreaks in 28 countries.

Throughout history, populations all over the world have sporadically been affected by devastating outbreaks of cholera. Records from Hippocrates (460–377 BC) and Galen (129–216 AD) already described an illness that might well have been cholera, and numerous hints indicate that a cholera-like malady has also been known in the plains of the Ganges River since antiquity.

Modern knowledge about cholera, however, dates only from the beginning of the nineteenth century when researchers began to make progress toward a better understanding of the causes of the disease and its appropriate treatment. The first cholera pandemic, or global epidemic, started in 1817 from its endemic area in Southeast Asia and subsequently spread to other parts of the world. The first and subsequent pandemics inflicted a heavy toll, spreading all over the world before receding.

In 1961, the seventh cholera pandemic wave began in Indonesia and spread rapidly to other countries in Asia, Europe, Africa, and finally in 1991 to Latin America, which had been free of cholera for more than one century. The disease spread rapidly in Latin America, causing nearly 400,000 reported cases and over 4000 deaths in 16 countries of the Americas that year [29]. In 1992, a new serogroup—a genetic derivative of the EI-Tor biotype—emerged in Bangladesh and caused an extensive epidemic. Designated *V. cholerae* 0139 Bengal (as mentioned above), the new serogroup has now been detected in 11 countries and likewise warrants close surveillance. Although no evidence is available to gauge the significance of these developments, the possibility of a new pandemic cannot be excluded. EI-Tor, for example, was originally isolated as an avirulent strain in 1905 and subsequently acquired sufficient virulence to cause the current pandemic.

The economic and social impacts of cholera are immense. In addition to human suffering caused by cholera, cholera outbreaks generate panic, disrupt the social and economic structure, and can impede development in the affected communities. Unjustified panic-induced reactions by other countries include curtailing or restricting travel from countries where a cholera outbreak is occurring or imposing import restrictions on certain foods. For example, the cholera outbreak in Peru in 1991 cost the country US$ 770 million due to food trade embargos and adverse effects on tourism.

The germ *Vibrio cholerae* is indeed a current unrestrained foe. Prominently transmissible among humans, this pathogen may readily be used for bioterrorism purposes.

## 5 SIGNIFICANCE OF BIOTERRORISM-PATTERNED NATURAL EPIDEMICS

*The principles and practicality of differentiating between epidemics stemming from bioterrorism acts and natural epidemics are outlined, to the extent attainable. The comprehension of bioterrorism-patterned natural epidemics thus turns out to be essential, and the related factors are presented in details, so as to facilitate the analysis of abrupt outbreaks.*

The occurrence of epidemics, or outbreaks, at large, is the outcome of three alternative courses:

- A natural biological process or move, such as the emergence of a new pathogenic variant, the spread of an infected vector onto a new biotope, and so on.
- An innocent, either passive or active, artificial intervention, such as the arrival of an infected person at an epidemiologically virgin area/place, with relation to the pathogen harbored by him.
- An act of bioterrorism (or a military biological war fare operation).

The significance of bioterrorism-patterned natural epidemics appears to be contributive in two major senses. It helps to appraise the profile and impact of bioterrorism scenarios, and it may facilitate the differentiation of an epidemic (or

an outbreak), being a natural or man-made—either intentional or unintentional—occurrence. Hence, the recognition of a bioterrorism act is thus facilitated as well. This paradigmatic axis is vital for the very basic classification and handling of any epidemic or outbreak of an infectious disease or poisoning. In certain cases of diseased individuals, it ought to be applied as well, so as to arrive at the respecting differentiation. Thereby, the field of forensic microbiology plays an important roll.

In practice, classification and handling of various occurrences and events of concern ought to rely on the following elements:

- Biotlogical and chemical analysis of specific markers related to the pathogen/toxin involved, so as to trace its provenance.
- Epidemiological analysis through which gauging of the event being a natural, innocently man-made, or deliberately man-made one, is allowed for.
- Information obtained or assessment made by intelligence sources regarding identified or possible perpetrators, if any.

One essential trait of bioterrorism, then, is the degree of resemblance of a natural event. The perpetrator might be trying to emulate a natural occurrence—so as to mask his act—or might be indifferent in that respect. Being an artificial intervention, bioterrorism can rarely or scarcely initiate an infectious course that would otherwise take place naturally in the same way of appearance. Yet this incriminating disadvantage would in most cases remain uncertain owing to ambiguity, let alone events of innocent human intervention. To minimize that gray area, the mechanisms underlying natural epidemiological phenomena ought to be meticulously and thoroughly recognized. In general, they are rather gradually evolving, mostly unapparent, although their outcome may be abrupt and drastic, eventually. Acts of bioterrorism, even if looking similar in their resultant impact, lack the preceding evolving links that mark the natural apparatus. Of paramount importance, then, is the formation of a formulated, multifactorial epidemiological profile with respect to every infectious disease or toxin of special concern. This is particularly important regarding the deciphering of unexplained biological invasions of new pathogens, toxins or infected vectors/carriers.

A proper case study is the totally unexpected appearance of West Nile encephalitis (WNE) in New York City, which reflected, actually, a much broader and fundamental phenomenon, namely the introduction of a pathogen prevailing for long in the Eastern Hemisphere onto the Western Hemisphere (or vice versa). In that case, inherent difficulty emerged in both identifying the etiological agent and deciphering, later on, its provenance. Although a well-known disease in the Eastern Hemisphere, it was impossible, subjectively, to clinically identify, or even suspect, WNE. This means, instantly, that the inventory of diseases being looked into ought to be, *a priori*, perfectly global, even in terms of initial symptomatology. The term "exotic" (disease or pathogen)—hence less likely, ostensibly—largely loses its essence, therefore. As the primary suspicion was that of an arthropod-borne virus (arbovirus) encephalitis, early testing was directed at common eastern North-American arboviruses. The synchronic mortality of collocated birds could have been instrumental for postulating that the concerned virus is basically an avian pathogen. Early serologic testing displayed IgM antibodies against the St. Louis

encephalitis virus by enzyme-linked immunosorbent assay, a basically correct finding that later turned to be misleading, showing a built-in limitation in identifying a pathogen new to a given locality.

Misdiagnosis is at times worse than lack of diagnosis, particularly with respect to infectious diseases, let alone in the context of epidemiological tracing intended to gauge bioterrorism-patterned events. Deliberate introduction of an encephalitic virus has been considered, among other possibilities, and has emerged to be more specific when it became clear that the etiological agent is WNE virus. It was a conjunction of three elements that propelled such assessment:

- Information indicating that Iraq cultivated WNE virus as a biological agents [30].
- An Iraqi intention to strike the United States by whatever means [30].
- Ostensible infeasibility of WNE virus natural transfer from the Western to the Eastern Hemisphere.

Sensibly, however, the possibility of bioterrorism was not regarded to be a prime one, in this case. The probability of a natural or accidental event still appeared to be higher. Although the mechanism of the introduction of the WNE virus into North America remains unknown, it seems clear that the source of the WNE virus strain detected in New York City originated in the Middle East. A similar avian epizootic among domestic geese in Israel during 1997 and 1998 had been attributed to WNE virus [16, 31]. Human cases of WNE virus occurred simultaneously in Israel and New York in August 1999, and when the genomic sequences of WNE virus isolates or infected human brain tissue from the New York City outbreak were compared with various non-U.S. strains, the greatest homology was found with a WNE virus strain isolated from a goose from the Israeli 1998 epizootic and subsequently with a strain detected in the brain tissue of an Israeli patient who died of WNE in 1999 [16]. In addition, both the pattern of high avian mortality previously not associated with WNE virus outbreaks as well as the severity of human CNS disease seen in New York City and Israel were similar during the 1999 outbreaks. In between 1998 and 1999, this specific strain of WNE virus could have probably been harbored by certain migrating avian host species (transfected by mosquitoes) along an axis extending from the Western Hemisphere, through Greenland, and, later, onto New York [32].

Deliberate introduction or release of infected birds or mosquitos compatible with the above-described natural apparatus could be equally feasible, conceivably; yet certainly extremely sophisticated, somewhat beyond the expected capacity of a terror organization or even a terrorism-sponsoring state. On the other hand, the employment of a current epidemic strain of cholera, for instance, through introducing it into food supplies or water systems, even if conducted far away from an ongoing epidemic, could generate a wide outbreak that by all means may seem natural, resembling an infected traveler that innocently took a transcontinental flight. The chain of transmission in such cases is much simpler; hence, they are liable to be followed, untraceably, in the form of an act of bioterrorism. Furthermore, concentrating on AIDS and Ebola, an elaborated analysis inquired into those aspects, and aroused certain viewpoints that illustrate the significance of bioterrorism-patterned natural epidemics [33].

Notably, a recent outbreak of mumps (June 2006, Kansas)—a viral, often epidemic disease—is expected to help health officials prepare for an act of bioterrorism or a flu pandemic [34]. The Kansas Health and Environment Department had found 761 confirmed or suspected cases of mumps around the state. In dealing with the outbreak, the department for the first time used its incident command system to spread information to agencies around the state. The same system would be used in the event of a natural or intentional disease outbreak.

Once the bioterrorism-employed pathogen has been introduced into the targeted site, the subsequent epidemiological featuring of the resultant events may expectedly be entirely futile, in terms of pointing to an act of bioterrorism, because those events would be shaped by whatever environmental, medical, and demographic circumstances prevailing at any rate in the targeted site. In the absence of direct evidence witnessing an act of bioterrorism, the only potentially incriminating indications can stem from the very specificity of the pathogenic strain involved, the particular manner of introduction, and the existence or lack of collateral factors such as natural vectors, carriers, or reservoirs, including natural abiotic vehicles, like rivers, of the pathogen in concern. Currently, the only fully reliable way to establish specificity is sequence analysis of both genomic and extragenomic DNA/RNA. This would be a necessary, yet usually insufficient, precondition/prerequisite for deciphering the provenance of a pathogen used in an act of bioterrorism.

Therefore, diligent modeling is vital. Evaluating surveillance systems for the early detection of bioterrorism is particularly challenging when systems are designed to detect events for which there are few or no historical examples [35]. One approach to benchmarking outbreak detection performance is to create semisynthetic datasets containing authentic baseline patient data (noise) and injected artificial patient clusters, as signal. This useful pattern has very recently been developed.

It has thereby been pointed out, indeed, that doctors in training had difficulties diagnosing diseases associated with bioterrorism. The study involved 631 doctors, mostly medical residents—doctors still in training—in 30 internal medicine residency programs in 16 states and Washington, D.C.; it was found that half of the doctors surveyed misdiagnosed botulism, 84% misdiagnosed plague, and chickenpox was misdiagnosed as smallpox 42% of the time [36].

Yet beyond the clinical domain, various, at times crucial, gaps still prevailing within our overall knowledge about the biology of pathogens may play a significantly negative roll in that they can possibly bring about false incrimination or, inversely, false discrimination, referring to bioterrorism-resembling events. That is but one reason, even if a major one, for deepening our research and understanding of the essence and pronunciations of the complex mechanisms underlying the emergence and prognosis of infectious diseases. Mankind, at large, would expectedly endeavor to be most familiar with its prime extraneous foes, especially as their might is currently and steadily increasing.

## 6 SPECTRUM OF APPLICABLE PATHOGENS AND TOXINS

*A general featuring of the spectrum of pathogens and toxins applicable for bioterrorism is given. Basic distinctions among various bioterrorism warfare agents, together with the implications they bear, are thereby presented.*

Basically, then, any epidemic pathogen may be regarded, potentially, as a candidate for carrying out a bioterrorism act, the germ causing plague, for example. To those agents should at any rate be added many nonepidemic—namely, incommunicable—pathogens, such as anthrax, and a large variety of toxins. The spectrum thus formed is in a way misleading in its amplitude.

In actuality, a much narrower range of pathogens and toxins that may be used, practically, as weapons is notable, although but a marginal fraction, quantitatively, of the entire class of bacteria, viruses, fungi, and other infectious agents plus numerous different toxins found in nature. This being the case, their characteristics should meet, one way or another, at least part of the following fundamental criteria:

- Infective/toxic and virulent.
- Procurable or reproducible.
- Conveyable and stable.
- Introducible into the targeted area and people.

Still, the spectrum of agents applicable for bioterrorism acts is broader than those applicable for armed conflicts, due to the greater vulnerability and wider variety of civilian infrastructures that may be targeted. Qualitatively, however, this specific offense-oriented range of pathogens and toxins is marked by very potent and aggressive biological agents, some of which are extremely contagious and medically untreatable.

Preferring, mostly justifiably, a comprehensive approach, the United Kingdom recently increased the number of biological agents that must be secured to ensure they are not obtained and used for acts of terrorism. The government boosted, thus, the number of restricted agents listed in the 2001 Antiterrorism, Crime and Security Act from 47 to 103. The list of controlled bioagents includes 45 viruses, 21 bacteria, 2 fungi, 13 toxins and 18 animal pathogens [37].

It seems, then, as if the most significant distinctions can be made between epidemic and nonepidemic agents, on the one hand, and, independently, between treatable and untreatable agents, on the other hand. Although the former distinction relates equally to bacterial and viral pathogens, the latter reflects a fundamental difference between those two major classes. Regardless of antisera, antiviral preparations are of limited efficacy, although they are expectedly being upgraded. Connectedly, within the sphere of toxins, the related spectrum may basically be divided into specifically treatable—protein toxins, hence may be coped by antitoxins—and specifically untreatable ones, which are otherwise chemically structured, like alkaloids. Vaccines, as prophylactic measures, are in principle efficient against viruses, bacteria, and protein toxins.

At the same time, it would be insensible to refer to those distinctions as absolute categorizations, because some pathogens represent intermediaries, to a certain degree. Other distinctions such as between lethal and sublethal pathogens are rather artificial, although of course not meaningless. Influenza, as one example, is generally regarded as an incapacitating disease, while potentially much more virulent, in reality. A genuine distinction, obviously, is the one categorizing self-reproducing agents—hence a multiplying mass of weapon—as against unviable

agents, meaning toxins. The former, if effectively applied, constitute an infectious apparatus that intensify itself; the latter rather resemble, in that sense, chemical warfare agents, bearing their own immense capacity.

## 7 BACTERIAL AGENTS

*Varitype pathogenic germs may be used for bioterrorism purposes through different modes, and they are here represented by four entities: plague, brucellosis, Legionnaires disease, and tuberculosis. Their heterogeneity with respect to source of infection, impact, as well as short-, medium-, and long-term endurance is thereby accentuated.*

Within the context of bioterrorism, bacterial pathogens are signified for, foremost, by the germs responsible for the diseases of plague, anthrax, and alimentary tract infections (the latter group already typified above in the form of cholera). Other bacterial pathogens, such as those causing tularemia, brucellosis, glanders, Legionnaires disease, Klebsiella-associated pneumonia, and Q-fever, are of secondary—certainly not negligible—importance. A third category of potential microbial agents may be added, in case the perpetrator chooses to employ a slow-acting, scarcely traceable type of germ. Those three categories are here visited, then, through four representatives, namely, the causative agents of plague, Legionnaires disease, brucellosis, and tuberculosis. Anthrax is separately covered.

### 7.1 Plague

The plague bacterium is an old, salient foe to mankind. Posing three significantly different forms of an aggressive disease—bubonic, pneumonic, and septicemic plague—this germ is chiefly marked by two hosts, prior to man infection, namely rat and flee; hence, it can in principal be employed as a bioterrorism agent through each of those three modes, namely, inhalable air-borne bacteria, infected rats, or infected flees. Naturally, the humane plague cycle is usually initiated by the bite of a rodent flea or occasionally by the human flea.

The bubonic plague is the common form, so called because of the buboes or swellings, which are the obvious visible symptom, communicated by flea bites and attacking the human lymphatic system [38]. Simple bubonic plague may kill within 5 days, if antibiotic is not applied. It is regarded as the "true" or classic form of the disease. Pneumonic plague occurs when a victim develops pneumonia as a consequence of first getting the bubonic plague. The pneumonic variant occurs when the bacilli colonize the alveoli and cause latent suppurative pneumonia. This results in hemorrhagic sputum that can spread the disease via aerosolized droplets. In that case, the infection is then communicated by exhalation and affects the bronchial system of the next host directly. The pneumonic variant kills within 1–3 days. The third variant, the septicemic plague, is extremely rare and occurs where the bloodstream is directly affected raising the concentration of bacilli in the bloodstream to a level at which it can be transmitted directly by the human flea without passing through the intermediary host of the rat. This plague variant can kill within hours and explains contemporary accounts of people who went to bed apparently well and when they awoke in the morning they were, like the biblical Assyrians (also victims of a god-sent plague), all dead men.

The bubonic form of the plague is the one most widely described by chroniclers and, contrary to popular belief, can be survived, especially if the buboes burst or are lanced, although this was not understood at the time. Both the pneumonic and the septicemic varieties were fulminant and 100% lethal. The pneumonic variety, however, could be transmitted only by close human contact, in the same way as influenza today, but it accounted for the high mortality among monks, priests, and others tending to the dying, putting their faces close to the patient. As those infected in this way were unlikely to get far in the period between infection and death, the pneumonic plague remained localized within families and communities, for example, monasteries where it could spread like wildfire. Henry of Knighton, a regular Canon in the Abbey at Leicester and opponent of the Franciscan Order, reported the death of the entire population of 150 Friars in a Franciscan monastery in Marseilles, adding as a malicious afterthought, a good thing too. The bubonic version also kills its hosts, but more slowly, especially in the case of the flea that can survive dormant for weeks, even months, in the goods of a traveling merchant. The causative agent, *Yersinia pestis*, will, however, die off when it has itself killed off all accessible and susceptible hosts, whether human, rat, or flea. To sustain itself the bacillus must find a host that is not affected by it or that reproduces faster than the bacillus can kill the host.

### 7.2 Legionnaires Disease (Legionellosis)

The particular significance of *Legionalla* sp. as a human pathogen was not recognized until 1976, when a mysterious epidemic of pneumonia struck members of the Pennsylvania American Legion. Indeed, the most common presentation of *Legionella pneumophila* is acute pneumonia (legionellosis); potentially any species (about 40, allover) of *Legionella* may cause the disease. Extrapulmonary disease (e.g., pericarditis and endocarditis) is rare. Less often, the disease presents as a nonpneumonic epidemic, influenza-like illness called Pontiac fever [39].

Remarkably, *Legionella* bacilli are primarily waterborne. They reside in surface and drinking water and are usually transmitted to humans in watery aerosols. The pathogenesis of *Legionella* infections begins with a supply of water containing virulent bacteria and with a means for dissemination to humans. Person-to-person transmission has never been observed, and *Legionella* is not a member of the bacterial flora of humans. Infection starts in the lower respiratory tract. Alveolar macrophages, which are the primary defense against bacterial infection of the lungs, engulf the bacteria; however, *Legionella* is, unlike most bacterial pathogens, a facultative intracellular parasite and multiplies freely in alveolar macrophages. Recruited neutrophils and monocytes, as well as bacterial enzymes, produce destructive alveolar inflammation. Direct inoculation of surgical wounds by contaminated tap water has been described.

In most cases, the initial infective link is an aerosol of water contaminated with the organisms. Evaporative condensers and cooling towers are proven sources of outdoor infection. Indoors, nebulizers, and humidifiers filled with contaminated drinking water have disseminated *Legionella* to susceptible patients. The automatic misting devices that keep supermarket produce fresh have even been fingered as culprits in outbreaks of pneumonia. Notably, aerosols are produced in numerous ways in our environment, from taking a shower to flushing the toilet. An epidemic

of Pontiac fever caused by *Legionella anisa* was associated with an ornamental fountain in a public place. Clusters of *Legionella* pneumonia have occurred after exposure to whirlpool spas in hotels or cruise ships. In most cases, the source of the infection remains unknown. Direct infection of surgical wounds has been linked to washing of patients with tap water that harbored pathogenic *Legionella* organisms.

*Legionella* infections may be sporadic or epidemic, community acquired, or nosocomial. There is great geographic variation in the frequency of infection even within communities, presumably reflecting the presence of suitable aquatic environments and susceptible subjects. Both sporadic and epidemic cases are more common during summer than winter months, apparently because of increased use of air-cooling equipment that generates aerosols. As a bioterrorism agent, this prominent bacterium is prone to be applied, then, in both water sources and air circulating systems.

### 7.3 Brucellosis

Brucellosis represents a narrow, uniquely significant category of diseases combining proximate and lingering effects altogether. This clinical profile may meet, then, certain needs of both short- and long-term warfare. In the short run, it is featured by an acute phase. After a relatively long and variable incubation period (1–8 weeks), it most often manifests as an acute bacteremic disease, and it may be complicated by secondary hematogenous localization in almost every organ [40, 41]. Inversely, chronic brucellosis is defined as a disease evolving for more than a year, without evident secondary localization. During asymptomatic infection, the pathogen is sequestered in lymphoid tissues, bone, and liver. It would tend to persist for many years, at times in a steadily debilitating incurable state. Four bacterial species are responsible for human brucellosis: *Brucella suis* (from pigs; high pathogenicity), *Brucella melitensis* (from sheep; highest pathogenicity), *Brucella abortus* (from cattle; moderate pathogenicity), and *Brucella canis* (from dogs; moderate pathogenicity). Many animals are potential reservoirs for various *Brucella* species [40, 41].

Epidemiologically, brucellosis is a zoonotic infection transmitted from animals to humans by ingestion of infected food products (milk and dairy products are the main sources for human contamination), direct contact with an infected animal, or inhalation of aerosols. This last mode of transmission is remarkably efficient given the relatively low concentration of organisms (as few as 10–100 bacteria) needed to establish infection in humans and has brought renewed attention to this old disease. Descriptions of the disease date back to the days of Hippocrates, although the organism was not isolated until 1887, when British Army physician David Bruce isolated the organism that bears his name from the spleens of five patients with fatal cases on Malta. The disease gets its additional names from both its course (undulant fever) and location (Malta fever, Crimean fever).

Given the ease of effectiveness of aerosol transmission of *Brucella* species, researchers attempted to develop it into a biological weapon beginning in 1942. In 1954, it became the first agent weaponized by the old U.S. offensive biological weapons program. Field testing on animals soon followed. By 1955, the United States was producing *B suis*-filled cluster bombs for the U.S. Air Force at the Pine

Bluff Arsenal in Arkansas. Not much later, it was weaponized and standardizes by the USSR. The pathogen may meet both large-scale and guerilla warfare purposes.

### 7.4 Tuberculosis (TB)

Although AIDS represents an attritional viral disease, as described, tuberculosis is, in parallel, a bacterial disease of a long-lasting debilitating course, both individually and demographically. Migration of contagion within community and the course of the disease in itself are very clumsy, yet consistent, aided by fortified stability toward antibacterial drugs. A staggering 1.9 million around the globe die of tuberculosis each year—another 1.9 billion are infected with the causative agent, *Mycobacterium tuberculosis*, and are at risk for active disease [42]. It's sobering to realize that this ancient scourge, which has been found in 2000-year-old mummies (but much earlier evolved), remains a severe global health threat despite modern medicine. In fact, TB is a leading infectious disease killer in the world, alongside AIDS and malaria.

Millions of people—fully one third of the world's population—are infected with TB. Five to 10% of those will develop the active disease, for reasons only partially understood. This is, certainly, a key trait. The bacteria can lie dormant for many years and become active when the host is weakened by factors such as stress, poor nutrition, diabetes, or HIV infection. Once active, TB destroys tissue in the lungs and becomes contagious. Because it is airborne, there is little that can be done to protect against infection.

Antibiotics and public health measures failed to wipe out TB, although until the mid-1980s its rates had been brought very low, even in the developed world. Then the TB bacteria acquired immunity to many drugs, spreading rapidly among homeless people, AIDS patients, and other vulnerable groups worldwide. Several drugs can cure TB, but only if they are taken for many months. With abbreviated treatment, bacteria that have evolved ways of evading the drugs can escape eradication and proliferate as a resistant strain. Then, even when treated with the strongest drugs for 2 years, the resistant TB is fatal about 60% of the time.

Russian prisons are a major source of multidrug-resistant TB. As tens of thousands of infected inmates reenter the civilian world every year, TB spreads when they cough and sneeze. Some experts fear that with the speed and frequency of modern travel, the resistant epidemic could spread to Europe and the United States. To counter the threat of a worldwide epidemic, the costly second-line drugs are supplied, and prison health workers are being helped to ensure that infected inmates take their entire course of drugs—even if it means the prisoners have to remain in confinement for up to 2 years past the start of treatment.

In 2005, in Japan, the Tuberculosis Prevention Law was set to be abolished so that bacteria that cause the disease can be included in counterterrorism regulations in an updated Infectious Disease [43]. The Tuberculosis Prevention Law (passed in Japan in 1951) does not contain counterterrorism provisions or regulations requiring the quick investigation of an epidemic. The change is aimed at restricting transfer of the disease (alongside with other ones), allowing inspections, and requiring registration of ownership of samples. The bacterium causing TB is regarded, in a sense, as an attritional bioterrorism agent.

## 8 VIRAL AGENTS

*In addition to the viruses mentioned above, some other viral pathogens are noticeable, of which several are profiled herewith as varitype representatives: Rift Valley fever, yellow fever, Norwalk, and hepatitis A. The smallpox virus is presented later. Notably, the handling of viruses is much more complicated then that of bacteria; hence, it is less likely to be mastered by terrorist organizations, individuals, or perpetrators, with reference being made to the viruses mentioned below as well as others.*

### 8.1 Rift Valley Fever (RVF)

The RVF virus typically represents a zoonotic hemorrhagic pathogen equally potent against man as well as various mammalian farm animals (such as cattle, buffalo, sheep, goats, and camels). For viruses, this wide spectrum featuring is rare, making this pathogen a biological agent of appreciable uniqueness. This trait characteristically found expression in the late 1970s, in Egypt. After an initial epizootic of RVF in Egypt in 1971, the first recorded epidemic of RVF in Egypt, in 1977–1980, inflicted an estimated 200,000 human cases, with some 600 reported deaths, plus enormous losses of farm animals [44]. An epidemic of the disease occurred again in Egypt in 1993. RVF virus is naturally a mosquito-borne pathogen, but it is adequately infective in the form of an air-borne aerosol, making it a biological agent applicable through those two modes. It is found in the Old World only, principally in Africa. In September 2000, an RVF outbreak was reported in Saudi Arabia and subsequently Yemen. These cases represent the first RVF cases identified outside Africa. Fortunately, it is not carried by migrating birds, as is the case with many other arboviruses (including the parallel mosquito-borne West Nile virus) and with influenza A virus. RVF virus, then, is basically a geographically slowly crawling virus of notable potentiality, particularly with respect to tropical and subtropical regions. It may be employed as a bioterrorism agent in the context of both antihuman bioterrorism as well as agroterrorism. Basically a mosquito-dependent tropical virus, it could doubtfully colonize temperate regions, however.

RVF is generally observed during years in which unusually heavy rainfall and localized flooding occur [45]. The excessive rainfall allows mosquito eggs, usually of the genus *Aedes*, to hatch. The mosquito eggs are naturally infected with the RVF virus, and the resulting mosquitoes transfer the virus to the livestock on which they feed. Once the livestock is infected, other species of mosquitoes can become infected from the animals and can spread the disease among other hosts, including man.

### 8.2 Yellow Fever

The Yellow Fever virus is another mosquito-borne pathogen, in that case affecting both man and primates. In the past, it was mass-accumulated as a biological weapon, in the form of infected mosquitoes, by the U.S. Army. Although somewhat complex, this is an effective way of employing this virus as a weapon either through guerilla or apparent warfare. It so happened that this standardized weapon-form

resembled and reestablished the first human virus ever isolated. It was in 1901 that the pathogen causing yellow fever was proved to be a filterable virus carried by mosquitoes.

Yellow fever is a viral hemorrhagic fever. There are three types of transmission cycle: sylvatic (or jungle), intermediate, and urban. All three cycles exist in Africa, but in South America, only sylvatic and urban yellow fever occur [46]. Sylvatic yellow fever occurs in tropical rainforests where monkeys, infected by sylvatic mosquitoes, pass the virus onto other mosquitoes that feed on them; these mosquitoes, in turn, bite and infect humans entering the forest. This produces sporadic cases, the majority of which are often young men working in the forest, e.g., logging.

On occasion, the virus spreads beyond the affected individual. The intermediate cycle of yellow fever transmission occurs in humid or semi-humid savannahs of Africa and can produce small-scale epidemics in rural villages. Semi-domestic mosquitoes infect both monkey and human hosts and increased contact between man and infected mosquito leads to disease. This is the most common type of outbreak observed in recent decades in Africa. Urban yellow fever results in large explosive epidemics when travelers from rural areas introduce the virus into areas with high human population density. Domestic mosquitoes, most notably *Aedes aegypti*, carry the virus from person to person. These outbreaks tend to spread outward from one source to cover a wide area.

### 8.3 Norwalk

The Norwalk virus is an enterovirus with pronounced eco-epidemiological traits. It causes acute gastroenteritis with nausea, vomiting, fever, and myalgia that lasts 24–48 hours. The virus is transmitted through fecal-oral contact. The Norwalk virus is well established as the chief cause of viral gastroenteritis epidemics. The disease occurs throughout the year without a seasonal predominance.

Norwalk virus was first associated with gastroenteritis in 1972. For a long time, it was an unnoticed, although extremely important, pathogen. It was identified by electron microscopy of stool samples that had been saved from a 1968 gastroenteritis epidemic that occurred in Norwalk, Ohio [47]. In a 2-day period, acute gastroenteritis developed in 50% of 232 students or teachers in an elementary school. The virus initially was labeled as a small, round, structured virus, and it was named after the city in which the outbreak occurred. The virus is transmitted through contaminated food, water, or infected contacts. After ingestion, the virus infects the mucosa of the proximal small intestine, damages microvilli, and causes malabsorption. Although no histopathological lesions can be found in the stomach mucosa, the virus causes abnormal gastric motility and delayed gastric emptying.

In the United States, the Norwalk virus is estimated to cause approximately 40% of cases of nonbacterial gastroenteritis. The Norwalk virus is, then, the leading cause of viral gastroenteritis in the United States. January 1, 2002, to December 2, 2002: Norwalk virus was attributed to 9 of the 21 outbreaks of acute gastroenteritis on cruise ships reported to the U.S. Centers for Disease Control and Pre-

vention's Vessel Sanitation Program in this time period. The Norwalk virus brings about approximately 23 million cases of acute gastroenteritis each year and is the major cause of outbreaks of gastroenteritis. An extensive outbreak has been recorded in Europe (Finland). As a bioterrorism weapon, this enterovirus is a typical incapacitating agent.

### 8.4 Hepatitis A

Inversely a water/food-borne virus of a relatively long incubation period (several weeks) and a causative agent of a lingering incapacitating infectious disease, hepatitis A virus (HAV), has been intended mainly for attrition warfare purposes and has apparently been employed by the Soviet army in Afghanistan, through contaminating water or food sources. Also, hepatitis A virus is reported to have been studied as a possible bioterrorism agent by South Africa.

Infectious hepatitis caused by HAV is common worldwide and is usually spread as a result of poor personal hygiene and can be a problem in day care centers, refugee camps, and among troops in the field. The virus is hardy and can withstand relatively harsh chemical treatments and can survive outside the body for days. It attacks humans only, invading the liver and generating, hence, a systemic debilitating illness.

The disease has an incubation period that is relatively long for a biological weapon at 3–4 weeks, but the onset of the disease is abrupt and victims become exhausted and develop jaundice as their livers lose function [48]. Victims typically take about 10 weeks to recover. The disease is rarely fatal, and victims have lifelong immunity. Overall, hepatitis A virus may be regarded, then, as a representative of a class of water/food borne antihuman viruses, in that case, a pathogen of a rather attritional nature. Lasting for more than 3 months, the entire course of the virus since contracted by a victim produces a fairly chronic infection.

## 9 TOXINS

*Overall, about 400 toxins have been identified, which are certainly but a small proportion of what is actually found in nature. Out of this remarkable inventory, several are reckoned as typical bioterrorism agents. Presented here are few prominent toxins: two protein toxins (botulinum and SEB)—hence detectible and treatable by anti-toxins—and three non-protein toxins (T-2 toxin, aflatoxin and aconitine), hence hardly detectable or treatable by antidotes. Etiological diagnosis of toxins is often extremely complicated.*

The realm of toxins is a very wide one. Those microbial, fungal, plant, and animal biomolecules exhibit an unthinkably diversified inventory of natural substances, all poisonous toward man, by definition. They are extremely variable chemically and physically. But dosage plays a cardinal role, both in terms of its great differential among toxins, as well as the delicate transition onto the tangential domain of toxin-based pharmacological applications. Moreover, most usable for man—as proteins, chiefly—in preparing toxoids and antisera, toxins are inwardly

inert—sometimes rather immensely beneficial—within their own toxinogenic biosystem, or extremely poisonous, outwardly, as natural toxicants. Fundamentally, they constitute the pristine form of biochemical warfare in nature, much earlier to the emergence of mankind. Thus, for instance, Hydrogen cyanide, reckoned as a classical chemical warfare agent, is found in various toxic plants, thereby equipping them with a powerful poison.

Overall, about 400 toxins have been identified, which are certainly but a small proportion of what is actually found in nature. They include, however, bacterial toxins—49; plant—36; fungal—26; algal—22; (other) marine organism—111; snake—124; insect—22; amphibian—5; and in total—395. Out of this remarkable inventory, several are reckoned as typical bioterrorism warfare agents. Presented here are few prominent toxins: two protein toxins (botulinum and SEB)—hence detectible and treatable by antitoxins—and three nonprotein toxins (T-2 toxin, aflatoxin, and aconitine), which are hardly detectable and treatable by antidotes. Ricin, an outstanding bioterrorism toxin on its own, is covered separately. Other notable toxins, such as tetradotoxin, saxitoxin, mamba toxin, and many others are not discussed, although certainly potential candidates for bioterrorism tasks, particularly state-sponsored ones. Mamba toxin, as one example, has been developed as a weapon in South Africa (see below). Also, various plant and fungal psychotoxins—part of which constitute important psychopharmacological agens—have in some countries been exploited experimentally, to the least, as a major means to attain compelled mind control over normal human beings— an act of governance regardable as bioterrorism.

### 9.1 Botulinum Toxin

Botulism toxin, a bacterial protein, is reputed as the most known toxic molecule. It is produced by *Clostridium botulinum*, a gram-positive spore-forming anaerobe germ that dwells in the soil. There are 7 different antigenic types of this toxin: A through G. Disease in humans is caused by types A, B, E, or F. Aerosol exposure studies in rhesus monkeys indicate that type F is the most toxic and that it is 60× more toxic than type B, the least toxic. Extrapolations from primate studies predict that the fatal dose of Botulinum toxin in humans is on the order of:

- 70 μg by mouth
- 0.7–0.9 μg via the respiratory route
- 0.09 μg by intravenous or intramuscular routes

The toxin is a 150-kDa zinc-dependent metalloproteinase that cleaves proteins involved in the docking and fusion of synaptic vesicles to the membrane at the neuromuscular junction. The toxin consists of two chains: a heavy chain (100 kDa) and a light chain (50 kDa) linked by a disulfide bond. The crystal structure of the molecule has been solved to 3.3-A resolution [49]. The protein structure revealed a 50-residue belt that partially obscures the active site access channel; the authors note that this unusual feature makes rational inhibitor design more difficult.

The deadly toxin irreversibly blocks acetylcholine release from peripheral nerves resulting in muscle paralysis. It has been developed as a biological weapon by many government research programs, including the United States, Russia, and Iraq. Currently, public health services classifies botulinum toxin as a Category A Bioterrorism Agent, indicating that it is a high-priority risk to public health. As a bioterrorism weapon, it is a potent substance that is easy to produce and transport. A large population of victims requiring intensive care could easily overwhelm health-care systems. It can be introduced into the food and water supply or be aerosolized. The toxin is unstable after a few days in surface water, and it cannot withstand chlorine treatment. Although it is more stable in packaged foods, it is inactivated by normal cooking practices (heat > 85°). Aerosolization is technically difficult but would bring about the poisoning of a large number of people.

At the conclusion of the Gulf War, the Iraqis reported that they produced 19,000 L of pure botulinum toxin to a UN team. A total of 19,000 L of botulinum toxin is estimated, just arithmetically, to be sufficient to kill three times the current human population by inhalation. Overall, 10,000 L of the toxin had been loaded into weapons. They loaded botulinum toxin into thirteen 600-km missiles and one hundred 400-lb bombs. Some of the toxin has not been accounted for. The JAMA consensus statement points out that it is "noteworthy that Iraq chose to weaponize more botulinum toxin than any other of its known biological agents."

Incubation times vary according to the amount of toxin ingested and the toxin type. The range reported from natural infections is between 6 hours and 10 days. Symptoms most commonly present between 12 and 36 hours of eating contaminated food. Inhalational exposures are predicted to progress more rapidly. Regardless of the route of intoxication, the neurological symptoms are similar. Gastrointestinal cases may also display nausea and vomiting. Botulinum toxicity presents as a febrile, descending paralysis. In cases of severe exposure, it is possible that respiratory failure will occur suddenly. Without treatment, the cause of death is respiratory failure due to obstruction from the pharyngeal muscles and failure of the diaphragm to move adequate volumes of air.

The standardized antitoxin is equine in origin, and of high efficacy, provided that it is administrated quite early. When treatment is administered promptly, the risk of death is significantly decreased. In the past 50 years, the mortality has decreased from 50% to 8%. It is important to note that the antitoxin prevents worsening of the condition, but recovery may not occur until after several weeks to months of intensive care. Complications such as long-lasting weakness, aspiration pneumonia, and nervous system dysfunction may occur. This would at any rate make botulinum toxin a deadly and/or medium-term debilitating agent. The toxoid is highly efficient as prophylaxis.

Most experts predict that a terrorist attack with botulinum toxin would attempt to distribute the agent via aerosolization to intoxicate the largest number of people. The toxin does not survive long in the water supply. In fact, there are no reported cases of water-transmitted botulism. The food supply is at some risk, but the agent can be inactivated by food preparation. Weapons programs, such as the one in Iraq, have demonstrated that it is not difficult to produce and store large amounts of concentrated toxin.

### 9.2 Staphylococcal Enterotoxin B (SEB)

One of seven staphylococcal enterotoxins, this bacterial protein is the best studied potential bioterrorism weapon of a group of molecules described as superantigens; others include two streptococcal pyrogenic toxins, and the toxic shock syndrome toxin [50]. Superantigens create stable bonds across class II molecules of major histocompatibility complex (MHC) and specific variable (Vß) T-cell receptors on $T_H$ cells, causing a release of disease-producing cytokines. Naturally, SEB intoxication occurs due to consumption of contaminated foods or drinks. SEB may be aerosolized or used to sabotage food supplies. As a weapon it would induce substantial, prolonged morbidity (up to 2 weeks) with little mortality.

From 1 to 6 hours after aerosol exposure to SEB, a distinct syndrome of high fever, chills, myalgias, nonproductive cough, dyspnea, and severe substernal chest pain may appear. Headache is also common, and nausea, vomiting, and anorexia may occur. Illness usually lasts for a few days, but it is rarely fatal. Physical examination may be unremarkable or reveal inspiratory and/or expiratory rales. After SEB ingestion, symptoms are primarily vomiting and diarrhea; fever only occurs in about one quarter of ill persons and respiratory involvement is absent. Although symptoms may be quite severe and lead to dehydration or even shock, illness usually lasts less than 12 hours. Physical examination may be unremarkable or reveal inspiratory and/or expiratory rales. Were Toxic shock syndrome toxin-1 (TSST-1) or streptococcal pyrogenic exotins to be used, they would likely cause acute erythroderma followed by desquamation and multiorgan failure. Differential diagnosis is complicated by many other diseases. Yet although toxin is transient in serum, it accumulates in the urine and is detectable for several hours postexposure. Antibodies generally develop within 6 days of exposure.

SEB, a very potent and enduring protein toxin, was included in the past U.S. BW program and standardized as a powerful biological warfare agent. Although reckoned to be an effectual incapacitating agent, it has not been connected, so far, with terror organizations.

### 9.3 T-2 Toxin

The Soviet military discovered—and later on materialiged—the potential use of the fungal trichothecene toxins—chiefly T-2 toxin—as aggressive bioterrorism agents shortly after World War II, when many Russian civilians ate bread baked from flour, naturally contaminated with extremely toxinogenic species of *Fusarium* mold. During this vast epidemic, numerous victims developed a protracted lethal illness characterized by initial symptoms of abdominal pain, diarrhea, vomiting, and exhaustion followed within days by fever, chills, muscular pain, and an imbalance of the red and white blood cells accompanied by pus-forming or other disease-causing organisms or their toxins in the blood or tissues [51]. It was then named Alimentary Toxic Aleucia.

The trichothecene mycotoxins are nonvolatile compounds produced by molds. They are very stable and resist heat- and ultraviolet light-induced inactivation. Only after heating at 500°F for 30 minutes will the toxins inactivate.

This discussion focuses on the hemorrhagic T-2 mycotoxin, a highly toxic agent that causes several illnesses in humans and animals, as described. From the 1970s and 1980s, trichothecene mycotoxins surfaced in the press as bioterrorism warfare

agents in incidents labeled "yellow rain" attacks against civilians in Southeast Asia [52]. Although such attacks have not been verified officially, it has been observed that trichothecene mycotoxins, T-2 in particular, have been produced in the USSR, weaponized, transferred to North Vietnam, and employed by North Vietnamese airplanes. The impact was vast and terrorizing, owing to many civilians that were vulnerably exposed, and then suffered a prolonged course of conspicuous pathological effects, with high mortality rate. Somehow, the evident employment of trichothecene mycotoxins in Southeast Asia, destined for bioterrorism—as well as live human field experimentation—purposes, turned into a saga involving alternative, natural explanations for the "yellow rain" phenomenon, which are certainly legitimate but do not refute the coexistence of the two—vicious and naïve—occurrences.

Reportedly, the Egyptian Air Force employed trichothecene mycotoxins during the Yemen War (in the 1960s) against Yemenis civilians—the main impact being terrorizing [53]. Later on, the Iraqi Air Force allegedly employed trichothecene mycotoxins against Iranian soldiers during the Iraq–Iran War in 1984 [54, 55]. According to UNSCOM, the Iraqis researched trichothecene mycotoxins, including T-2 [56]; they probably weaponized the T-2 toxin.

Unlike most biotoxins and microorganisms that do not affect the skin, T-2 toxin is an aggresive dermal irritant and can severely harm an unprotected person's skin and eyes. The pain associated with the exposure occurs within seconds to minutes. Larger doses produce incapacitation and death within minutes to hours. A larger amount of T-2 toxin is required for a lethal dose than of the chemical warfare agents VX, soman, or sarin. Comparisons with blister agents such as sulfur mustard show the T-2 toxin is about 400 times more efficient in producing blisters: It takes approximately 50 ng of T-2 toxin to produce the same injury to the skin as 20 μg (20,000 ng) of mustard. The T-2 toxin has a diverse effect depending on the manner and amount of exposure with vomiting and diarrhea noted at exposure doses one fifth to one tenth the lethal dose [52].

Exposure causes skin pain, itching, redness, blisters, and sloughing (shedding) of dead skin. Effects on the airway include nose and throat pain, nasal discharge, itching and sneezing, cough, shortness of breath, wheezing, and chest pain; the victim spits blood as a result of pulmonary or bronchial hemorrhage. The T-2 toxin also produces effects after ingestion or eye contact. Severe poisoning results in prostration, weakness, jerky movement, collapse, shock, and death [57]. The only protection against T-2 toxin effects is the individual protective mask and chemical protective overgarment. No chemotherapy, vaccine, or specific antidote is available [57].

Altogether, the trichothecene mycotoxins are reckoned as primarily blister agents that, at lower exposure concentrations, would cause extreme skin and eye irritation, and at larger doses would produce considerable incapacitation and death within minutes to hours [52].

### 9.4 Aflatoxin

As a public health hazard, the major routes for aflatoxin exposure are inhalation, pulmonary mycotoxicosis, especially of grain dusts, and ingestion as a result of eating food made with contaminated grains. The aflatoxin problem was first

recognized in 1960, when there was a severe outbreak of a disease referred to as "Turkey 'X' Disease" in the United Kingdom, in which over 100,000 turkey poults died. The cause of the disease was due to toxins in peanut meal infected with *Aspergillus flavus*, and the toxins were called aflatoxins. The potentialities of aflatoxins as toxic carcinogens, mutagens, teratogens, and immunosuppressive agents are well documented [58]. They are produced as secondary metabolites by the fungus *Aspergillus flavus* and *A. parasiticus* on a variety of food products. Chemically, aflatoxins normally refer to the group of difuranocoumarins.

Two forms of illness have been observed:

1. (Primary) Acute aflatoxicosis is produced when moderate-to-high levels of aflatoxins are consumed. Specific, acute episodes of disease may include hemorrhage, acute liver damage, edema, alteration in digestion, absorption and/or metabolism of nutrients, and possibly death.
2. (Primary) Chronic aflatoxicosis results from ingestion of low-to-moderate levels of aflatoxins. The effects are usually subclinical and difficult to recognize. Some common symptoms are impaired food conversion and slower rates of growth with or without the production of an overt aflatoxin syndrome. In the long term, exposure to aflatoxin is thought to explain the high rates of primary liver cancer in Africa and parts of Asia. Pooled data from Kenya, Mozambique, Swaziland, and Thailand show a positive correlation between daily dietary aflatoxin intake (in the range of 3.5- to 222.4-ng/kg body weight per day) and the crude incidence rate of primary liver cancer (ranging from 1.2 to 13.0 cases per 100,000 people per year).

Iraq developed and weaponized aflatoxins, chiefly as a bioterrorism agent. It has then been noted that "The discovery (by UNSCOM, during the 90s) that Iraq was researching aflatoxin, not a traditional BW candidate, was a cause for some surprise. It is a carcinogen, the effects of which manifest themselves only after many years, and several Western experts have rationalized this Iraqi program only in terms of genocidal goals. If aflatoxin were used against the Kurds, for instance, it would be impossible definitively to prove the use of BW once the symptoms emerged. Another possible explanation is its potential use as an immune suppressant, making victims more susceptible to other agents. However, the aflatoxin declaration may also hide other aspects of Iraq's BW program: according to Iraq's depositions, the production program never encountered any mishap (as other parts of the BW program had) and, to judge from the declared time-frame for the total amount produced, production could never have stopped, even for cleaning of the equipment. This raises the suspicion that Iraq declared an excessive amount of aflatoxin in order to disguise the fact that other, more destructive agents had been produced in greater quantities [59]."

Iraq developed and produced aflatoxin as a long-term debilitating bioterrorism agent, intended to be used against the Kurds. There is documentary evidence and statements obtained by UNSCOM that Iraq was mixing aflatoxin with riot-control gas [60]. Although seemingly peculiar, some rational can be figured out regarding this type of weaponization (much more detail is given in Section 15.6).

### 9.5 Aconitine

This relatively unnoticed plant toxin is an intensely poisonous alkaloid obtained from aconite root. It occurs in colorless crystals. Aconite tubers are most toxic wild plants distributed from Asia to western Europe [61–64]. It achieved some notoriety in the nineteenth century as an agent for homicides and suicides. In modern times, aconite has been used as Chinese herbal medicine, which is freely purchased from herb shops and consumed as a doctation by a herbal practitioner for pain control in the Northern Hemisphere [65]. In Japan, some cases of aconite poisoning appeared as a result of committing suicide or accidental ingestion, which were mistaken for edible grass. However, aconite alkaloids have the potential for serious and even fatal cardiotoxity, which management has continued to fail save patients with therapeutic resistant fatal arrhythmia [66, 67].

Serious aconitine poisoning is characterized by hypotension, palpitations, shock, delay in myocardial conduction, and dysrhythmia beginning within 6 hours of ingestion. Respiration is progressively depressed by the effect of aconitine on the bulbar respiratory center. Tingling of the tongue, mouth, and skin, followed by numbness and anesthesia, are characteristic signs of aconitine poisoning. Excessive salivation is a characteristic sign of poisoning. The skin becomes cold, clammy, sweaty, and pale. Initial bradycardia is due to vagal stimulation. Soon after ingestion, aconite causes a tingling, burning sensation on the lips, tongue, mouth, and throat, which is followed by numbness and constriction of the throat. When applied to the skin, aconite causes a tingling sensation and then numbness. Blurring of vision occurs. Initial miosis is followed by mydriasis. There is a feeling of constriction in the throat [68].

According to some reports, Egyptian Armed Forces Commander-in-Chief Field Marshall Mohammed Abd el Hakim Amer, who headed the defeated Egyptian Army in the 1967 Six Days War, was later poisoned, institutionally, by this substance. Allegedly, he committed suicide by means of aconitine that had been introduced to him while arrested [69]. An unidentified toxicant, possibly the same plant toxin, was used to replace the antidote with which Egyptian autoinjectors were filled, for personal assassinations [70].

Another trivial, yet peculiar episode, involved an English doctor who used aconitine to murder his crippled 18-year-old nephew for his inheritance. Percy Malcolm John was a resident at Blenheim School in Wimbledon and owned a small property that would go to his uncle on his death. Lamson visited the boy, bringing a cake and a capsule that he said was medicine. A few hours later he died in agony from aconitine poisoning [71]. Reportedly, aconitine toxin has been adopted by some states for assassination purposes.

## 10 AVAILABILITY OF INFECTIVE AGENTS AND TOXINS

*A broad spectrum of availability of pathogens and toxins is discussed, especially pertaining to potential resources worldwide. Still, the spectrum is much narrower regarding weaponized agents, making the course of the attacker complicated, unless "raw" infective agents or toxins are to be employed. Different episodes illustrating that variability are presented.*

The common nonstate-sponsored scenario of a just-graduated microbiologist cultivating a pathogenic germ in a rudimentary household laboratory, so as to produce it as a weapon for bioterrorism acts, is partially feasible. It is much less likely with regard to weaponizing or engineering bioterrorism agents. A significant advantage emerges concerning state-sponsored bioterrorism, because the relevant pathogens and toxins are appreciably more procurable, reproducible, and conveyable for a state, generally speaking. Therefore, the inventory of agents, their handling and the feasibility of conducting bioterrorism are scaled up when a state is involved, either intending to carry out an act of bioterrorism on its own or to assist a terror organization. Any step made by a country may basically be institutionalized, hence be more legitimate compared with an unaffiliated or noninstitutionalized body, element, or person.

Overall, the wide range of bioterrorism agents is reflected by considerably varying degrees of availability. Ricin toxin can readily be produced, roughly, by unqualified saboteurs. Conversely, an expedient way to meet the need for high-leveled know-how often faced by terror organizations would be in the form of "scientific mercenaries." At large, varitype availability of pathogens and toxins is demonstrated though the following episodes (and further ones mentioned in Sections 14, 15 and 16).

In Japan, in 1965, outbreaks of typhoid and dysentery were intentionally induced by a local bacteriologist, with professional direct access to those germs [72].

In 1972 two men affiliated with the U.S. group "Order of the Rising Sun," who eventually fled to Cuba, had conspired to contaminate the water supplies of some large U.S. Midwestern cities with stocks of typhoid fever germs cultivated by one of them. Up to 40 kg of bacteria cultures were found in a college laboratory [73].

In 1973, the American leftist terrorist group, the Weather Underground, reportedly attempted to blackmail a homosexual officer at USAMRIID into supplying organisms that would be used to contaminate municipal water supplies in the United States. The plot was discovered when the officer requested several items "unrelated to his work."

In 1975, the Symbionese Liberation Army was found in possession of technical manuals on how to produce bioterrorism weapons.

Moving to Europe, in 1980, police raided a German Red Army Faction apartment in Paris and found a miniature laboratory containing a culture medium of the germ that produces botulinum toxin. Notes about bacteria-induced diseases were found in the apartment as well [74]. Once again in Europe, several Muslim terrorists affiliated with the Algerian "Armed Muslim Group" were arrested in Belgium, holding a lot of information about biological (and chemical) weapons and about the World Cup football games [75].

Outstandingly, protesters claimed to have taken infected soil from the Scottish island of Gruinard and placed it at the Microbiological Defence Establishment at Porton Down, Britain. The island has been closed to the public since germ warfare experiments on sheep were conducted there in 1941. The anthrax spores used in the experiments can remain dangerous for decades.

The American Type Culture Collection (ATCC) has often been approached. Two Canadians attempted to procure botulinum and tetanus cultures from the ATCC. Reportedly the first phone order, of less deadly cultures, was fulfilled, and

it was not until the second order that ATCC employees became sufficiently suspicious to notify authorities [76].

A member of the white supremacist Aryan Nation acquired freeze-dried bubonic plague bacteria from the ATCC [77]. In May 1995, in the United States, Larry Wayne Harris was arrested for illegally obtaining the plague bacteria *Yersinia pestis*. Using his previous employer's certification, Harris obtained the samples through the mail from the ATCC. He was sentenced to 18 months probation and 200 hours of community service. Harris was again arrested in 1998 when he and another individual were found allegedly in possession of anthrax cultures, which were later determined to be anthrax vaccine. Due to the ease of obtaining dangerous pathogens, the CDC established rigorous guidelines for shipment of specific pathogens that may be used as bioterrorism agents.

Yet, as a matter of fact, a wide variety of pathogens and toxins serving the Iraqi BW program during the 1980s has been freely procured from different Western and Eastern sources, mostly the ATCC. The Iraqi case study provides remarkable demonstration, indeed, of repeated procurements of highly dangerous pathogens and toxins that readily took place albeit an integral part of the past Iraqi BW program [78]. And beyond, paradoxically, anti-bioterrorism programs worldwide bring about academic, commercial, and industrial institutionalization of numerous new pathogens-holdings labs which constitute potential resources for pathogens seekers. It so happened, that 245 facilities are now authorized to work with live anthrax in the US, and about 100 actually do so , compared to roughly 12 prior to the 2001 anthrax mailings [79].

Inadequately secured state-held stockpiles of BW constitute a potential resource. Reportedly, bin Laden's associates managed to receive anthrax and plague cultures or weaponized agents, from former Soviet facilities in Kazakhstan, or elsewhere. This has not been evidenced, although attempts to obtain those and other germs have certainly been made by al-Qaeda, perhaps fruitlessly. Still, assuming that al-Qaeda is responsible (together with Iraq) for the Sept. 2001 anthrax letter attacks, its members were at least involved in the final step of the envelope posting, thus possessing, temporarily, this pathogen [80]. Also, aerosol-released bioterrorism weapons were often in the possession of Spetsnaz operatives (Soviet Special Forces) who were believed to be involved with terrorists, especially those associated with bin Laden.

A typical example demonstrating an endeavor to exploit an institutionalized researcher as a 'scientific mercenary' was the case in which al-Qaeda had direct connections with a senior Pakistani microbiologist, Dr. Abdur Rauf, who attempted to support al-Qaeda's pursue of BW [81].

Adriana Stujit, an acknowledged Dutch journalist, contended that radical Muslims have gotten control over South Africa's biological weapons stockpiles. Quantities of anthrax, Ebola, Congo fever, and other agents are missing from South Africa, with the only explanation being theft. There has been no explanation or accurate documentation of the amounts, locations, and contents of the stockpiles [82]. This is not at all surprising, considering the picture described below regarding South Africa.

Obtaining pathogens and toxins from their natural environment is another sensible, though skill-demanding mode. In 1993, the Japanese cult Aum Shinrikyo sent a group of 16 cult doctors and nurses to Zaire, on a supposed medical mission. The

actual purpose of the trip to Central Africa was to learn as much as possible about and, ideally, to bring back samples of Ebola virus. In early 1994, cult doctors were quoted on Russian radio as discussing the possibility of using Ebola as a bioterrorism weapon. The attempt was fruitless. Also, they attempted to purchase a Q-fever culture from a Japanese academic researcher but were rebuffed. But they did obtain botulinum germs from earth in Japan, anthrax germs, and cholera germs. The anthrax strain was consistent with strain Sterne 34F2, which is used in Japan for animal prophylaxis against anthrax [83].

Domestic naive facilities may readily be contributive. In Israel, much concern was raised that regular microbiological laboratories—whether private or institutionalized, and particularly those located in the Palestinian-occupied territories—may be used as minifactories for producing biological agents in quantities sufficient for bioterrorism purposes [84].

A broad spectrum of resources of pathogens and toxins is notable, then, especially pertaining to potential resources worldwide. Still, it is much narrower regarding weaponized agents, making the course of the attacker complicated, unless "raw" infective agents or toxins are to be employed in an act of bioterrorism. But, on the other hand, dozens of biotech firms now offer to synthesize complete genes from the chemical components of DNA, making it feasible, seemingly, that a bioterrorist, armed with only a fake e-mail address, could probably order such deadly biological components online and receive them by mail [85]. He would doubtfully be able to transform, however, those components into a weapon, if not substantially assisted.

## 11 TECHNICAL FEASIBILITY OF BIOSABOTAGE ACTS

*Pragmatically, biosabotage acts are featured by high technical feasibility, making them attractive, in that sense. Various operational options allowing for such feasibility are presented and discussed. Still, an appreciable level of know-how is needed for all preparatory stages preceding the in-effect act of bioterrorism.*

The most plain, nearly self-evident, act of bioterrorism is demonstrated by one person emptying a tube into a water source so as to contaminate it with a waterborne pathogen. Against such an act stands mainly the level of water chlorination, which does not interfere, practically, with the distinct technical feasibility of this deed. It is, indeed, a simple thing to do.

But the technical feasibility of bioterrorism may extend much further. A fairly sophisticated practice involves the preparation and dispersing of a toxin or pathogen in an inhalable aerosol form, either dry or wet. Various advanced aerosol technology-based devices structured to dispersing inhalable aerosols of pharmacological quality are available and may likewise release a pathogen or toxin. Alternatively, dissemination may be achieved by infected insects. During an infamous biowarfare attack in 1941, the Japanese Military released an estimated 150-million plague-infected fleas from airplanes over villages in China and Manchuria, resulting in several plague outbreaks in those villages [86]. The victims were, deliberately, merely civilians. The same principle may be applied for disseminating plague by the release of infected rats in an enemy's settlements. Another example, similarly complex, yet plainly feasible act of bioterrorism, rabid, pre-symptomatic dogs or

other canines infiltrated covertly, steadily, and untraceably in the wild through borderlines, may then bring about the spread of rabies onto local dogs and farm animals plus humans, gradually elevating the overall level of rabies endemicity.

Currently, effective worldwide dissemination is achievable through postal systems. The following recent incident, although certainly a naïve one, illustrates the technical feasibility of biosabotage acts at large, as one, outstanding example:

Thousands of scientists were scrambling at the urging of global health authorities to destroy vials of a pandemic flu strain sent to laboratories in 18 countries as part of routine testing. The rush, urged by the World Health Organization, was sparked by a slim, but real, risk that the samples could spark a global flu epidemic. The vials of virus sent by a U.S. company went to nearly 5000 laboratories, mostly in the United States [87].

Referring to the incident, WHO's influenza chief, Klaus Stohr contended that "The risk is relatively low that a lab worker will get sick, but a large number of labs got it and if someone does get infected, the risk of severe illness is high and this virus has shown to be fully transmissible. The risk is low but things can go wrong as long as these samples are out there and there are some still out there."

The 1957 pandemic strain, which killed between 1 million and 4 million people, was in the proficiency test kits routinely sent to laboratories. This strain has not been included in the flu vaccine since 1968, and anyone born after that date has no immunity to it.

Most samples were sent at the request of the College of American Pathologists, which helps laboratories do proficiency testing. A private company, Meridian Bioscience Inc. of Cincinnati, Ohio, is paid to prepare the samples. The firm was told to pick an influenza type A virus sample and chose from its stockpile the deadly 1957 H2N2 strain. The reason for choosing this past, mighty strain, and not a current strain, remains unclear [88]. Still, some other test kit providers besides the college also used the 1957 pandemic strain in samples sent to laboratories in the United States.

The majority of the laboratories that got the test kits are in the United States. Fourteen were in Canada, and 61 samples went to laboratories in 16 other countries in Europe, Asia, the Middle East, and South America, according to the WHO. The test kits are used for internal quality control checks to demonstrate that a laboratory is able to correctly identify viruses or as a way for laboratories to get certified by the College of American Pathologists. The kits involve blind samples. The laboratory then has to correctly identify the pathogen in the vial in order to pass the test. Usually, the influenza virus included in these kits is one that is currently circulating, or at least one that has recently been in circulation. The WHO then notified the health authorities in all countries that received the kits and recommended that all samples be destroyed immediately. That same day, the College of American Pathologists faxed the laboratories asking them to immediately incinerate the samples and to confirm in writing that the operation had been completed.

Unintentionally, this demonstration of the technical feasibility of biosabotage operations marks both simultaneous, wide-range distributable shipment of a virulent—in that case, rather recoverably pandemic—pathogen in a stable infective form, as well as specifically, on-spot targeted sites intended to be attacked biologically. The two options are equally feasible.

Actually, two events of unintentional postal distribution of smallpox—yet in that case generating epidemics in effect—happened to occur already in 1901. Smallpox had developed in a woman in Saginaw, Michigan, after she received a letter form her sweetheart, a soldier in Alaska.

He had written it while recovering from this disease. The infection subsequently spread to 33 other persons in Saginaw [89]. Also in 1901, an outbreak of 5 cases of smallpox was recorded at the Mormon headquarters in Nottingham, England, apparently after receipt of "letters or other fomites" from Salt Lake City, Utah, where smallpox was widespread [90].

Heavily crowded facilities single out the technical feasibility of bioterrorism. Thus, the British army tested model bacilli spore powder released (1963) from a window of a tube train traveling in the London Underground [91]. The trial concluded that the spores can be carried for several miles on the tube system, and locally can persist as an aerosol of high concentration for a considerable period. Also, widespread dispersal of bacteria was found in a May 1965 secret release of innocent bacilli at Washington's National Airport and its Greyhound bus terminal, according to declassified military reports. More than 130 passengers who had been exposed to the bacteria traveled to 39 cities in seven states in the two weeks following the mock attack [92]. By 1966, in a similar trail in New-York Subway, bacilli powder-carrying light bulbs were dropped [93].

Water is an easy target, basically. The feasibility of water contamination has been pointed out to be concrete in Israel: Deliberate influx of sewage from Judea and Samaria hills onto the Israel Coast Plain was been observed during the 1990s, according to the Israel Minister for Environment Quality, causing the contamination of Israeli water systems [94, 95]. Objectively, this may become much more feasible as a result of the 2005 Israeli disengagement. It should be mentioned that, in the past, outbreaks of cholera occurred in Judea and Samaria as a result of sewage being routinely and innocently used for the irrigation of vegetables. Connectedly, the factual situation is that most streams found in Israel stem from the mountainous aquifers, which are mostly located in the Judean and Samarian Hills, parts of which are now outside of Israeli territorial control. The other Israeli water sources stem from the Golan Heights and Southern Lebanon.

Furthermore, rather unusual modes illustrating the technical feasibility of biosabotage acts have been noted in Israel:

In the body of one person who survived suicide bomber sabotage, bone fragments were discovered that are believed to come from a suicide bomber. The bone fragments tested positive for hepatitis B. Medical experts believe "This is possibly the first report of human bone fragments acting as foreign bodies in a blast injury, and consequently all survivors of these attacks in Israel are now vaccinated for hepatitis B." The experts suggested that bone fragments embedded in attack victims should be routinely tested [96]. Theoretically, bone fragments might also spread other diseases including HIV, dengue fever, syphilis, and Creutz–Jakob disease [97].

In August 2004, an Iranian-made unmanned drone launched from southern Lebanon by the Iranian-supported Shi'lte militia, the Hezbollah flew for about 15 minutes along Israel's northern Mediterranean coast until it reached the coastal resort of Nahariyah. "Iran has not only supplied Hezbollah with these UAVs but

has also trained 30 of the group's members to operate them," an official (a senior commander in the Iranian Revolutionary Guards) told the London-based Arab daily *al-Sharq al-Awsat* 11 Nov 2004 [98]. The efficiency of UAVs as aerosol disseminators of biological agents is well known.

Drifted saboteurs may even be full residents of the country to be attacked. A terrorist state or organization might use such residents, particularly extremists, to carry out bioterrorism acts, thus considerably elevating the technical feasibility of those operations. The attacked might catch the extremists, but not understand they were minor figures, and present them instead as major figures. All in all, the practicability of bioterrorism is fairly plain, although a degree of professional knowledge is needed for planning and preparations.

It is also possible to disseminate infected insects or infected animals (like rats, for instance). Also, deliberate infection of contacts by an infected individual harboring a contagious disease (AIDS, for example—a plot already carried out not infrequently) is, in a sense, an act of bioterrorism. Water and ventilation systems may be targeted. During the First Gulf War, there were serious concerns in the United States, that the ventilation systems of buildings might be attacked by terrorists using BW agents [99].

According to Norqvist, terrorists might effectively and easily disseminate biological agents through water systems, resulting in a high number of casualties, the terrorism focusing either on large cities or military facilities. It has thereupon been pointed out that chlorination can be neutralized by using naturally chlorine-resistant microorganisms or constructing bacteria resistant to chlorine concentrations normally used in the municipal drinking water systems [100]. Simple devices might be used to deliver biological agents into the ventilation systems of buildings. Open-air experiments carried out in the 1960s demonstrated that throwing a light bulb filled with biological agent before an incoming subway train during rush hour is sufficient to infect tens of thousands of people [101]. Plausibly, state-supported terrorist groups would be the most likely of all terrorist groups to get a hold of biological weapons by the supporting government transferring entire systems from their national programs [5].

## 12 OPERATIONAL MODES

*A wide variety of modes that may serve for operating bioterrorism is described. Essentially aimed at attaining advantage through asymmetric warfare, they may lead, at most, to an impact equaling catastrophic terrorism. They might be limited, inversely, to merely assassination. Hoax acts are visited as well. One-man operational mode can bring about any of the mentioned outcomes.*

The operational modes marking bioterrorism are closely related to asymmetric and catastrophic warfare. Asymmetric warfare is a military term to describe warfare in which the two belligerents are mismatched in their total capabilities or accustomed methods of engagement such that the inferior side must press its special advantages or effectively exploit its enemy's particular weaknesses if they are to have any hope of *prevailing* [102–104]. Thus, asymmetric warfare is a recipe for engaging an opponent who has superior military power at his disposal and where

the main target is not the armed forces but the fabric of society itself. Catastrophic terrorism differs from the more traditional forms of terrorism because it has no closely defined political aim. It is based on the belief that the current world order is dominated by certain "universal values" that the terrorists despise and that this order must be destroyed.

Consequence management roles vary depending on whether the attack is a terrorist event or an act of war. Yet, the distinction between terrorist and military use of WMDs is increasingly problematic. State adversaries, perhaps acting through terrorist surrogates, may be inclined to use WMDs early, unconventionally, and if possible, anonymously. They may use WMDs against such military targets as ports, airfields, staging areas, and overseas bases to prepare the military battlefield by slowing logistics and power projection. Similarly, they may psychologically undermine public support or politically divide a coalition for an operation by attacking domestic or allied civilian targets. Eager to deter regional involvement while avoiding an overwhelming retaliatory response, perpetrators may try to obscure their identity. Under these circumstances, the line between terrorism and symmetric warfare may then vanish.

The stealthy qualities of BW further complicate the distinction between terrorism and war. An adversary with effective agent dissemination capabilities could employ BW as part of a covert attack nearly impossible to detect until casualties appear. Depending on the agent used, the attacked might not know whether an outbreak was natural, a terrorist attack, or the opening assault of a war. For all the legitimate concern about bioterrorism, systemized weapons are mostly still difficult to employ effectively without access to state-developed technology. Rogue states and well-financed terrorists with access to state-developed technology remain the most serious danger from the effective, large-scale use of biological agents through guerilla warfare.

The repertoire of bioterrorism modes is potentially an incredible one, extending from ricin-based assassination of one person to a worldwide smallpox pandemic, even if not aimed, a priori, to reach such magnitude. Those two edges may each take shape in actuality due to a seemingly slight act carried out by a single perpetrator. The former mode has practically been materialized through the murder of Markov, the Bulgarian journalist in London, whereas instead of ricin the lethal toxin used could variedly be of another type, as detailed above. The latter mode is a continuously floating menace, the possible realization of which is attentively being faced, in terms of both preventive measures and preparedness for handling such occurrence. The one-man mode was exemplified in 1998, when the British government issued a warning to all ports about an Iraqi attempt to bring large quantities of the deadly germs of anthrax into Britain (and other countries) inside cosmetics bottles, cigarette lighters, and perfume sprays, disguised as duty-free goods [104].

A bioterrrorism attack could consist, then, of nothing more than a person deliberately coughing on people and surfaces after being infected with a pathogen, as correctly pointed out in a study conducted by the Australian Strategic Policy Institute. This could happen by a terrorist knowingly infected and incubating an infection flying to Australia and then walking around a crowded shopping center for some hours, coughing near people and over surfaces. The study found that even without their knowledge, infected individuals could likewise be used as carriers for bioterrorism acts [105].

One major operational mode of massive bioterrorism may focus on endpoints of collective source reservoirs right before their physical division into much smaller to individually consumed portions. Such reservoirs may include blood, sera, or other preparations intended to be introduced into the body as well as foods, drinks, and water. Thus, certain collective infusion fluids may be contaminated before being packed: AIDS, hepatitis B, hepatitis C, arboviruses, and other blood-adapted pathogens.

Milk supplies, as another example, meet that very same biosabotage principle and have aroused much worry. About a third of an ounce of botulism toxin poured by bioterrorists into a milk truck en route from a dairy farm to a processing plant could cause hundreds of thousands of deaths and billions of dollars in economic losses, according to a scientific analysis [106].

The analysis considered what might happen if terrorists poured into a milk tanker truck a couple of gallons of concentrated sludge containing as much as 10 g of botulinum toxin. Because milk from many sources is combined in huge tanks holding hundreds of thousands of gallons, the toxin would get widely distributed in low, but potentially lethal, concentrations and within days be consumed by about 568,000 people, the report concludes. The researchers acknowledge that their numbers are very rough. But depending on how thoroughly the milk was pasteurized (which partially inactivates toxins) and how promptly the outbreak was detected and supplies recalled, about 400,000 people would be likely to fall ill, they conclude. Although only 6% of victims would generally be expected to die, the death rate could easily hit 60%, they conclude, because there would not be nearly enough mechanical ventilators or doses of antitoxin to treat so many victims.

A different mode of bioterrorism may aim to bring about the outward leakage of infectious microorganisms by physically damaging facilities containing them. A variety of installations may serve for such a mode, including P-3 and P-4 laboratories, vaccine factories, and institutionalized depositories. Such sabotage may questionably be regarded as bioterrorism, particularly if directed toward a facility engaged in biological warfare, even if but defensive. Yet the consequences and ultimate impact may be very severe. Hardly accusable may as well be bioterrorists choosing a mode of amplifying an ongoing outbreak or epidemic, especially if the very same causative pathogenic strain is concurrently employed by them for that purpose.

The various options of bioterrorism conducted as a sole operative mode or in conjunction with other sabotage acts add a further dimension. Bioterrorism acts may on the spot be accompanied by simultaneous conventional and/or unconventional—mostly chemical or radiological—sabotage. Such synchronism may form camouflage and be misleading, making the bioterrorism component scarcely thinkable or detectable. Spatial diversity may concomitantly increase the blurring effect. Combined employment of WMDs indeed drew the Pentagon's attention. The Pentagon has implemented, thus, a plan to expand its support of civilian authorities in the event of multiple domestic attacks involving chemical, biological, radiological, or nuclear weapons, The Virginia-based Joint Task Force-Civil Support was previously the Pentagon's only force dedicated to such a mission [107].

"We have identified capabilities within our force structure—beyond Joint Task Force-Civil Support—in order to ensure that we could respond not simply to a domestic attack involving a WMD, but to multiple attacks at diverse locations,

several cities perhaps at once where terrorists might have employed weapons of mass destruction," Paul McHale, deputy assistant secretary of defense for homeland defense, said. "It is now the established policy of the Department of Defense that we will train and equip for the mission requirement of multiple WMD response," said McHale [107].

Two overt modes are notable:

1. Suicide by a person carrying a detectable contagion.
2. Bioterrorism combined with conventional terrorism, simultaneously serving for two advantages:
   a. Temporary misleading (until infected cases present).
   b. Increasing vulnerability to bioterrorism agents by conventional wounding.

Opposite in polarity to the extreme mode involving in-effect bioterrorism in conjunction with multi-WMDs, are bioterrorism hoaxes. Bioterrorism hoaxes—whether consisting of unviable or viable substances that are benign—constitute an increasing mode of bioterrorism. The anonymously angry of the world are creating serious trouble today with just a stamp, an envelope, and a household product. It takes only moments for someone to send a package containing a suspicious but ultimately harmless powder. Moreover, hoaxers often see the powder-filled letters as a way to send a message, without considering that they are committing a crime. The act, however, can force entire buildings to be locked down and tie up emergency personnel for hours. Each incident must be taken seriously in case the contents turn out to be anthrax rather than crushed aspirin.

It so happened that the uncertainty prevailing around a hoax caused Australian Prime Minister John Howard to say in the early hours of a crisis thereupon created due to an powder containing envelope received in the Indonesian Embassy (June 2005): "This is a very serious development for our country, and I can't overstate the sense of concern I feel that such a recklessly criminal act should have been committed." He added he believed the threat was meant as retaliation for the 20-year prison sentence Australian national Schapelle Corby received in Indonesia for smuggling drugs. Such acts could damage relations between the two nations and invite retaliations against Australians in Indonesia [108].

Actually, however, people were putting powder in the mail to scare others years before someone sent envelopes laced with anthrax that killed five American people in the fall of 2001. There were 22 incidents reported worldwide in 2000 involving faked biological agents, according to the Monterey Institute's Center for Nonproliferation Studies. The count jumped to about 730 in 2001, due largely to Clayton Lee Waagner's campaign against abortion providers [109].

Waagner was arrested late that year after sending roughly 550 powder-filled letters, and the number of bioterrorism hoaxes dropped to 70 in 2002. However, excluding his count, both 2001 and 2002 saw significant increases in such incidents from 2000, spurred by the anthrax mailings and the resulting media coverage, according to Sundara Vadlamudi, research associate for the center's WMD Terrorism Research Program.

Although certainly the most relentless, Waagner was not the first anti-abortion bioterrorism hoaxer or the last. More than 20 letters containing fake anthrax were

mailed to abortion providers and abortion rights organizations in January 2002, several weeks after his arrest. Currently, however, the trend is increasing; notable receiving powder-filled envelopes since April 2005: the Israeli Embassy in Washington, D.C., the Danish embassies in Stockholm and Vienna, the New Mexico state capitol, a Slovenian government office, the office of Quebec Premier Jean Charest, a Vermont multimedia company, Israel Army Radio, the Bank of Israel, and NATD's Joint Warfare Center, Norway.

## 13 SPATIALLY AND TEMPORALLY VARYING IMPACTS

*The expected impact of bioterrorism acts, together with the featuring of their spatial and temporal variance, are configured. A typical multifactorial system is thus formed, the handling of which is remarkably complicated, demanding, and vital. It includes a far-reaching range of bioterrorism scenarios, some evident, some tentative, beginning with hoaxes scenarios, and ending in the form of a pandemic.*

The impact of bioterrorism acts is appreciably varied, both spatially and temporally. Its variability is shaped by the following factors:

- Initial area coverage (via air, water, food, or animal vectors/carriers)
- Duration of pathogenetic course
- Curability
- Environmental stability of the pathogen/toxin
- Contagiousness
- Demographic conditions
- Climatic conditions
- Conduction and effectiveness of preventing measures (before and after the act of bioterrorism)

The multifactorial integral thus formed is complex. For automating the testing and validation of spatial and temporal cluster detection algorithms, and enabling the ready creation of datasets for benchmarking outbreak detection systems, a specific software tool has been developed [35]. It allows for the creation of simulated clusters with controlled feature sets, varying the desired cluster radius, density, distance, relative location from a reference point, and temporal epidemiological growth pattern. This tool does not require the use of an external geographical information system program for cluster creation. Based on user-specified parameters describing the location, properties, and temporal pattern of simulated clusters, it creates clusters accurately and uniformly.

The impact is pronounced within several domains:

- Personal illness that may lead to death
- Paralyzed manpower
- Logistical efforts needed to medically support and isolate the infected/sickened victims
- Meticulous, extremely demanding managing of the apparently uninfected population

- Demoralization that may ascend to total panic
- Overall instability

The remarkable psychological responses subsequent to a bioterrorism attack were summarized as follows: horror, wrath, panic, paranoia, demoralization, magical thinking about germs and viruses, fear of invisible pathogenic agents, fright of contagion, anger at terrorists, government, or both, attribution of arousal symptoms to infection, scapegoating, social isolation, and loss of faith in social institutions [110]. Panic is a power multiplier in itself. Some plead that the use of biological agents can cause severe panic and hysteria among the civic population. For example, during the Sept. 2001 anthrax letter attack, thousands of civilians came to hospitals, although only few of them were found to be suffering from exposure to the anthrax powder. When BW agents are used, there is a great fear that the number of civilians turning to get medical help will overload the health service systems and those who really need treatment will not be able to get it. This overload, besides keeping treatment from those who need it, will also cause mass hysteria. Mass hysteria is very significant. In case of contagious diseases, there is a grave concern that the mass hysteria and the urge to flee the scene will only cause the disease to spread. That is contrary to the proper way of fighting it, namely, closing and isolating the contaminated area.

A typically acute bioterrorism episode would be featured by a toxin or pathogen of a short incubation period, pronounced illness, recovery or death, absence of secondary epidemic waves, and the creation of long-lasting immunity of the surviving population, which would inhibit relapsing occurrences. Such a prototypic episode may at times not exceed a week, just as it had been designed by the perpetrator. But, contrastingly, attritional or demographically debilitating impact may result from a delayed, relapsing, or slowly progressing pathogenetic course that may prevail for many years in a given population.

Spatially, however, any combination with the above-mentioned temporal variations is possible, basically. This means that in a relatively confined area a several-weeks transient bioterrorism episode or a lingering impact stemming from a bioterrorism act lasting for years, may take place. Equally, a vast territory may be afflicted by a swiftly moving epidemic or, alternatively, by a clumsily evolving pathogen, in terms of epidemic rate and virulence, whether it is an incapacitating or deadly agent. The dynamics underlying this complex variability is foremost shaped by the nature of the toxin or pathogen employed in-effect, and the countermeasures to be specifically taken against it. Fundamentally, this fluctuating dynamics is well illustrated, both spatially and temporally, through the varitype bacteria, viruses, and toxins presented within the above corresponding sections. It prevails, thereby, for both the dimension of the individual victim as well as the demographic dimension at large.

On the other hand are the bioterrorism hoax incidents, of which the spatial and temporal impacts, though, are still significant. Even an absolutely empty post envelope bearing but one word, like "biohazard," would bring about those undesirable impacts. If not empty, it may contain but paper and/or innocent chemical, toxin, non-infective microorganism, infective yet attenuated microorganism, or pathogenic microorganism. In this order of possibilities, there is an ascending line of spatial and temporal impacts. (The envelope content may as well comprise insects, infected or naïve.)

Three recent examples are herewith mentioned, then, resembling the spatial and temporal variance of impacts:

After a harmless bacillus powder was found in an post envelope sent to the Indonesian Embassy in Australia, the Australian capital of Canberra sustained a barrage of bioterrorism hoax mailings beginning 1 June 2005 to its Parliament building, Prime Minister and Cabinet Department, and the embassies of Indonesia, the United States, United Kingdom, Japan, Italy, and South Korea [111]. Emergency personnel in Canberra had to decontaminate 46 staffers at the Indonesian Embassy when it appeared some form of powered toxin had been sent to the building, according to Australian media reports. Tests later indicated the substance was not dangerous [112].

Some days later, continuation took place far away. On 10 June 2005, an employee in the collections department of Imperial Parking in Vancouver, British Columbia, opened an envelope containing a similarly suspicious white powder. Vancouver emergency services were quickly alerted [113]. The affected section of the building was locked down—the air-conditioning system and elevators were shut off, and no one was allowed to enter or leave. Twenty-five workers were quarantined for several hours until tests determined the substance was not dangerous. Forty-eight emergency workers and 18 pieces of equipment were sent to the powder scare at the Imperial Parking office. Firefighters suited up in protective gear to retrieve a sample, whereas others managed the quarantine and organized decontamination for the hazardous materials workers. Emergency medical personnel examined firefighters before they entered the building and after they exited, and waited to see whether they would have to take anyone to the hospital. Police officers handled crowd control and traffic and provided an escort when the sample was taken for further testing at a laboratory. This occurred, unfortunately, while the city's emergency services were also handling a hotel fire, an injured person, and the nearby appearance of the Aga Khan, spiritual leader of Shia Imami Ismaili Muslims.

A third episode followed soon. An Israeli embassy employee in the United States checking mail at 5 p.m. 16 June 2005 opened one envelope that contained a small amount of a white powder. A section of the building was closed, and some employees remained inside until tests indicated at 11 p.m. that the substance was benign. The letter, which reportedly contained anti-Semitic language, was traced to a man jailed in North Carolina, according to the FBI [113]. The logistical, mental (and medical) impacts would be much worse, certainly, both temporally and spatially, in case a mailed envelope contains a living pathogen, as was the 2001 anthrax letter attack (presented below).

## 14 HISTORY AND EVOLUTION

*An old tool of mankind, the chronology of bioterrorism from 1500 BC to now is presented. It demonstrates the gradual, consistent transition from intuitive bioterrorism to knowledge-based bioterrorism. It shows, as well, the past, recent, and present trends of bioterrorism as an evolving course, the extrapolative featuring of which is troublesome.*

The evolutionary course of biological warfare at large is fascinatingly long. Almost as soon as humans figured out how to make arrows, they were dipping them in animal feces or sick persons so as to contaminate them. Practically, bioterrorism

agents were the earliest weapons of mass destruction ever used by mankind. Long before having any idea regarding the in-effect nature of poisonous and infectious substances, man used toxins, feces, carcasses, and the like as effective weapons, in order to disease, kill, and terrorize opponents and enemies, thus following, from the very beginning, the three very basic fundamentals: incapacitating, killing, and intimidating. Rather a sort of intuitive bioterrorism, primarily.

Still, most of the ancient world liked to believe it fought with a code of honor. According to Adrienne Mayor, "Archers were disdained because they shot safely from afar: long range missiles implied unwillingness to face the enemy at close range. And long range missiles daubed with poisons seemed even more cowardly." Yet Odysseus, hardly a coward, returns home to kill his wife's suitors with poison arrows [1].

The roots of repugnance toward contaminated weapons are deep. Such weapons were singled out for disdain more than 2000 years ago in Greek and Roman codes of conduct, as well as in Hindu writings. But the course of using pathogens and toxins for terrorism continued. The rulers of ancient India were no less conflicted than the Greeks about arrows "barbed, poisoned or blazing with flame." These instruments violated the "traditional Hindu laws of conduct for Brahmans and high castes, the Laws of Manu." But in the Arthashastra, the Brahman military strategist Kautilya advised his king to use whatever means necessary to attain his military goals, including poisons.

In the Near East, it was after the mythic Greek hero Hercules slew the multi-headed Hydra that he developed a technology at the heart of today's most pressing international issue. In her illuminating history of warfare, the acknowledged folklorist Adrienne Mayor argues, in a sense, that "by steeping his arrows in the monster's venom, Hercules created the first biological weapon [1]."

Amazingly, then, the ancient world already made an important distinction between using disease weapons for purely defensive purposes as opposed to "first strikes." Although this constraint was rooted partially in ethics, Mayor contends it also reflected a shrewd understanding of epidemiology: "The principle of summoning plague for self-defense may be related to the reality that invaders are 'immunologically naïve' and therefore more vulnerable to epidemic diseases in foreign lands than the local population." Mayor is intrigued with the often fantastic devices the ancients used in war, but "Greek Fire" also excavates ancient attitudes toward biological arms and terrorism tools that are startlingly relevant today. Poisonous arrows were the Bronze Age's terror weapons. "Almost as soon as they were created," Mayor writes, "poison weapons set in motion a relentless train of tragedies for Hercules and the Greeks—not to mention the Greeks' enemies, the Trojans."

Similar, unrecorded biotterorism-like events followed, most probably. Yet, beyond Greek mythology, even earlier, the ancient world also contained examples of bioterrorism warfare against settled populations stretching back to 1500 BC, when the Hittites sent plague victims into the lands of their enemies. Later, it was in about 600 BC that Solon of Athens put hellebore roots in the drinking water of Kirrha to kill the inhabitants. At about the same time, the Assyrians used to poison enemy wells with a fungus that would make the enemy delusional. Around 500 BC, Socrates was poisoned by the juice of Conium, which was the state poison of the Athenians. Necrotizing germs were intuitively employed in 400 BC, whereas Scythian archers systematically used arrows dipped in blood and manure or decom-

posing bodies to prevent wounds from healing. There are accounts from 300 BC of parts of the dead bodies of humans or animals being used by the Romans to poison water supplies. There have also been instances of the dead bodies of those who died from plague being catapulted into besieged cities. Later on, by 200 BC, Carthaginians used Mandrake root left in wine to sedate the enemy. Furthermore, in 184 BC, Carthaginian leader Hannibal is credited with an interesting use of bioterrorism weapons. In anticipation of a naval battle with the Pergamenes at Eurymedon, he ordered his troops to fill clay pots with snakes. During the battle, Hannibal sent the pots crashing down on the deck on the Pergamene ship. The confused Pergamenes lost the battle, having to fight both Hannibal's forces and a ship full of snakes. The act brought about panic and injured enemy sailors. Pergamenes, a king in Asia Minor, remarked that "he did not think any general would want to obtain a victory by the use of means which might in turn be directed against himself [1]." Honey bees are known to have been used as weapons since Roman times, as was then the case of Lucullus, who defeated Mithridates (74 BC) via bees.

As man expended onto the Far East and, later, the Americas, tropical plants became a common source for toxins. Thus, poisoned arrows are used widely throughout the jungles of Burma, Malaysia, and Assam. The principal sources of arrow poison are varieties of *Antiaris, Strychnos*, and *Strophanthus* [114]. *Antiaris toxicaria*, for example, is a tree of the mulberry and breadfruit family, common in Java and the neighboring islands. The active agents in all of these are either contained within the milky sap or the juice of crushed seeds. This is smeared behind the arrow point on its own, or mixed with another plant latex. When introduced into the bloodstream, the active ingredient (either antiarin, strychnine, or strophanthin depending on the species) acts quickly, attacking the central nervous system causing paralysis, convulsions, and cardiac failure.

The story of the *Strychnos*-derived toxin Curare is remarkable, in that it perfectly embodies a complete bivalent evolutionary line, starting with a very old arrow-poison—still in its very same traditional use presently—and ending, for the time being, as an important muscle relaxant. Contemporarily, curare is usable for selected assassinations just like "modern" toxins. In the sixteenth century, a group of Spanish explorers traveled the Amazon River. During the voyage, one explorer was hit in the hand by an arrow and died soon after. The culprit was curare, used widely as an arrow poison by many Amazon Indian groups (as it is still used by a few today). The complex processes used to make curare were a guarded secret. Often 30 or more ingredients could be found in one recipe. Indigenous Amazonians often mixed plants of different genera to concoct their potent toxins; their skill and knowledge in safely preparing these poisons is a testimony to their incredible ingenuity. Amazonian curares are divided into two groups based on the container the plant is stored in: pots or tubes. Pot curare in the East Amazon is predominately from the species *Strychnos guianensis*. Tube curare in the West Amazon is from *Chrondrodendron tomentosum*. (The curare in modern medicine is made from the latter species, therefore, its name: tubocurarine.)

For many centuries, the exact content of curare remained a mystery to Western observers; not until 1800 did Alexander Von Humboldt witness and document the preparation of curare by the Indians from the Orinco River. In 1814, an explorer named Charles Waterton injected a donkey with curare. Within 10 minutes, the donkey appeared dead. Waterton cut a small hole in her throat and inserted a pair

of bellows, and then pumped to inflate the lungs. The donkey held her head up and looked around. Waterton continued artificial respiration for 2 hours until the effects of curare had worn off. Curare was found to block the transmission of nerve impulses to muscle, including the diaphragm muscle, which controls breathing.

Back to Europe, the Middle Ages, the use of infectious diseases to break sieges of castles and fortified towns is widespread. The most common method is to use catapults to hurl dead human or animal bodies over walls to spread disease. This same method is used to poison water sources. Plagued rats and infected flies were employed as well for introducing and disseminating the contagion. This was the case in 1155, when Barbarossa uses dead bodies to spread pathogens among the enemy during the battle of Tortona [115]. About 200 years later, in 1340 attackers hurled dead horses and other animals by catapult at the castle of Thun L'Eveque in Hainault, in what is now northern France. The defenders reported that "the stink and the air were so abominable . . . they could not long endure" and negotiated a truce.

In 1346, Tartar forces led by Khan Janibeg attacked the city of Kaffa (now Feodossia, Ukraine), catapulting the plague-infected bodies of their own men over the city's walls, and forcing the defending Genoese to abandon it when plague spread. Connectedly, ships carrying plague infected refugees (and possibly rats) sailed to Constantinople, Genoa, Venice, and other Mediterranean ports and are thought to have contributed to the second plague pandemic. (The first plague pandemic in 541 AD spread from Egypt to other parts of the world and killed 50–60% of the world population). It was perhaps the trigger of a subsequent outbreak of Bubonic plague that swept medieval Europe, causing 25 million deaths. Using dead bodies and excrement as weapons continued in Europe during the Black Plague of the fourteenth and fifteenth centuries. Even as late as 1710, Russian troops fighting Sweden resorted to catapulting plagued bodies over the city walls of Reval.

During that era, three peculiar events took place, as well:

In 1422, at Karlstein in Bohemia, attacking forces launched the decaying cadavers of men killed in battle over the castle walls. They also stockpiled animal manure in the hope of spreading illness. Yet the defense held fast, and the siege was abandoned after 5 months. In 1495, the Spanish tried wine infected with leprosy patients' blood against the French near Naples, and, in 1650, Polish artillery General put saliva from rabid dogs into hollow spheres for firing against his enemies.

In between, a sort of bioterrorism came to the New World in the fifteenth century, aimed to defeat the Indians. Spanish conquistador Pizarro gave clothing contaminated with the smallpox virus to natives in South America. During the French and Indian War (1754–1767) Sir Jeffrey Amherst, commander of British forces in North America, suggested the deliberate use of smallpox to "reduce" Native American tribes hostile to the British [116]. An outbreak of smallpox at Fort Pitt results in the opportunity to execute Amherst's plan. On June 24, 1763, Captain Ecuyer, Amherst's subordinate, gives blankets and a handkerchief from the smallpox hospital to the Native Americans and records in his journal, "I hope it will have the desired effect." This was followed by an epidemic of smallpox among Native American tribes in the Ohio River valley, which may also have been spread by contact with settlers. Transmission of smallpox by fomites (on blankets) is inefficient compared with respiratory droplet transmission.

Still, although the ancient world's arsenal of primitive bioterrorism weapons was trivial compared with the horrors of the modern world, those weapons raised the same terrifying moral and political dilemmas then as now. Thus, the absence of brought out events whereby BW were employed throughout the nineteenth century was possibly the latency preceding proliferation. Let alone that the nineteenth century marked the first isolation of germs in science.

During the twentieth century, then, an ascending course took place. In modern times, BW was used first for sabotage by Germany during WWI. German forces reportedly spread glanders and anthrax to debilitate enemy cavalries (detailed above).

In 1918, the Japanese formed a biological weapons section in the Japanese Army (Unit 731). Later on, in 1931, Japan expanded its territory into Manchuria and made available "an endless supply of human experiment materials" (prisoners of war, mostly civilians) for Unit 731. Biological weapons experiments in Harbin, Manchuria, continued until 1945. A post-World War II autopsy investigation of 1000 victims revealed that most were exposed to aerosolized anthrax. It is estimated that up to 3000 more prisoners and Chinese nationals may have died in this facility. During an infamous biowarfare attack in 1941, the Japanese Military released an estimated 150-million plague-infected fleas from airplanes over villages in China and Manchuria, resulting in several plague outbreaks in those villages. Overall, the Japanese Imperial Army experimented with and operated about 16 biological agents as tools of warfare and terrorism between 1932 and 1945. This took place in numerous locations in Asia, and it has been estimated that a total of 10,000 Chinese prisoners, U.S. prisoners of war, and British detainees were killed by some of the most gruesome human experimentation in history. The Japanese used BW agents such as anthrax, plague, tularemia, and smallpox [117, 118].

During World War II, another effort, taken by Britain, was the production of over 5 million anthrax infected cattle cakes, which would have been dropped over Germany in an attempt to decrease meat stocks by some 30%. Events overtook plans to put these into operation [119].

Ken Alibek, formerly a chief scientist of the Soviet offensive biological warfare program, has alleged that the Soviets employed the germs causing tularemia during World War II. In his book *Biohazard*, he states that there is evidence tularemia was used by the Soviet troops to help stop the German panzer troops in the Battle of Stalingrad. The resulting tularemia outbreak may have halted the Nazi advance, but the Soviet troops also developed the disease because of what Alibek suspects was a sudden change in wind direction. Over 100,000 cases of tularemia were reported in the Soviet Union in 1942, a 10-fold increase in incidence experienced in 1941 and 1943. Seventy percent of the cases were the respiratory form of the disease, which is the form that would have been expected from a BW rather than a natural outbreak of the disease [120]. Nonetheless, this episode may constitute a good example of biological warfare, rather than bioterrorism.

Avner Cohen, an expert on unconventional weapons proliferation, has catalogued reported uses of bioterrorism weapons by Jewish forces during the 1948 War of Independence in Palestine. Connectedly, the Israeli historian Uri Milstein alleged that "in many conquered Arab villages, the water supply was poisoned to prevent the inhabitants from coming back." Milstein states that one of the largest of such covert operations caused the typhoid outbreak in Acre in May 1948 [121]. Within

that context, the Palestinian Arab Higher Committee reported in July 1948 that there was some evidence that Jewish forces were responsible for a cholera outbreak in Egypt in November 1947 and in Syrian villages near the Palestinian–Syrian border in February 1948. Furthermore, in May 1948, the Egyptian ministry of defense stated that four "Zionists" had been captured while trying to contaminate artesian wells in Gaza with "a liquid which was discovered to contain germs of dysentery and typhoid."

A pause took place for almost two decades (apart from minor or marginal incidents), but since 1965, the trend significantly changed. Multiple bioterrorism occurrences have been observed, as follows.

1965—In Japan, outbreaks of typhoid and dysentery were deliberately induced by a local bacteriologist [72].

1970—In Canada, several students became badly ill after eating food deliberately contaminated with the eggs of parasitic ringworm [122].

During the 1970s, the Soviets used mycotoxins in Laos and Cambodia and hepatitis A virus in Afganistan.

1972—Two men affiliated with the U.S. group "Order of the Rising Sun," who eventually fled to Cuba, had conspired to contaminate the water supplies of some large Midwestern cities with stocks of typhoid fever germs cultivated by one of them. Up to 40 kg of bacteria cultures were found in a college laboratory [73].

1975—The Symbionese Liberation Army was found in possession of technical manuals on how to produce bioweapons.

1980—Assassination of CIA agent Boris Korczak in McLean, Virginia, Tyson's Corner, using a ricin weapon, possibly in umbrella configuration [75].

Police raided a German Red Army Faction apartment in Paris and found a miniature laboratory containing a culture medium of the germ that produces botulinum toxin. Notes about bacteria-induced diseases were found in the apartment as well [74].

Protesters claimed to have taken infected soil from the Hebridean island of Gruinard and placed it at the chemical defense establishment at Porton Down. The island has been closed to the public since germ warfare experiments on sheep were conducted there in 1941. The anthrax spores used in the experiments can remain dangerous for decades [75].

1983—The FBI obtained one ounce of ricin in a 35-mm film canister from an individual in Springfield, Massachusetts, who had manufactured it himself. This is believed to be one of several confiscations of ricin [75].

Some 750 people were sickened due to typhoid fever, consequent to bacterial contamination of restaurant salad bars in Oregon, conducted by a local cult attempting to affect the outcome of a local election [73].

1984—Australian authorities received an anonymous threat warning that foot-and-mouth disease virus would be released among livestock if reforms in Queensland Prison were not implemented [123].

A Cuban expert defected and testified that one third of the United States could have been contaminated if a stockpile of toxins held by Cuba were to be "strategically placed in the Mississippi River [75]".

Two Canadians attempt to procure tetanus and botulism cultures from ATCC. Reportedly the first phone order, of less deadly cultures, was fulfilled, and it was not until the second order that ATCC employees become sufficiently suspicious to notify authorities [71].

1990–1995—Japanese cult Aum Shinriky attempted at least three times to disperse aerosolized botulinum toxin and anthrax in downtown Tokyo and at U.S. military installations in Japan. All attempts failed.

1990—In Scotland, a limited outbreak of giardiasis occurred as result of deliberate water contamination [124].

1991—During the Gulf War, there were serious concerns in the United States, that the ventilation systems of buildings might be attacked by terrorists using BW agents [98].

1993—An Arkansas man with survivalist group connections attempted to smuggle 130 g of ricin from Alaska into Canada to use as a weapon [125].

1994—An Iraqi scientist specializing in genetic engineering and implanted in New York by Saddam Hussein's regime intended to conduct an act of bioterrorism, due to having access to various local laboratories [126].

1995—Two members of the Minnesota Patriots Council were convicted of conspiracy to assassinate a deputy U.S. Marshal and International Revenue Service agents by ricin [127].

A member of the white supremacist Aryan Nation acquired freeze-dried bubonic plague bacteria from the ATCC [76].

1996—In Texas, 12 laboratory workers at a medical center became ill as a result of eating muffins and doughnuts intentionally contaminated by dysentery germs type 2 [185].

1998—An Iraqi terrorist network was maintained in the United States, intending to conduct acts of bioterrorism and reportedly furnished with BW agents by Iraqi women that smuggle agents filled vials into the United States within their bodies [128].

The British government has issued a warning to all ports about an Iraqi attempt to bring large quantities of the deadly germs of anthrax into Britain (and other countries) inside cosmetics bottles, cigarette lighters, and perfume sprays, disguised as duty-free goods [104].

Several Muslim terrorists affiliated with the Algerian "Armed Muslim Group" were arrested in Belgium, holding a lot of information about biological and chemical weapons and about the World Cup football games [129].

1999—A group of medical workers, mostly Bulgarian, employed in the Benghazi children's hospital, Libya, were detained, subsequent to an explosive AIDS epidemic that apparently started in 1998 in that hospital, and involved, thereupon, hundreds of children. By the year 2000, the medics were charged with deliberately infecting 393 Libyan children in their care with HIV, by injecting them with infected products. The HIV type was the same in each case—a rare and previously unrecorded strain that originated in West Africa when two existing viruses combined [130]. An

analysis conducted by European experts suggested that the outbreak has the hallmarks of accidental cross-contamination, where poor hygiene and ineffective sterilization procedures allowed contaminated blood to be spread between patients from a single infected child. Highly involved in the affair, the charged medics were then sentenced to death, however. Libyan leader Muammar Gadhafi said he believes the medics are guilty.

2001—The anthrax letters were sent in the USA, containing highly sophisticated militory-grade spore powder (details are given bellow).

2003—Former Texas Tech professor Thomas Butler said that 30 vials of plague bacteria were missing from the university. In fact, he stole, smuggled, and illicitly transported the bacteria to Tanzania [131].

Further episodes have been discussed above. All in all, the various above-described episodes are at any rate but a small segment of the entire picture. By 1998, Carus undertook a comprehensive inventory and assessment of bioterrorism (together with biocrimes) in the twentieth century. He has documented 222 cases, categorizing the cases and number of reported cases [132]:

- Confirmed use of bioterrorism agents—24
- Probable or possible use—28
- Threatened use (probable or confirmed possession)—11
- Threatened use (no confirmed possession)—121
- Confirmed possession (no known attempts or threats to use)—5
- Probable or possible possession—6
- Possible interest in acquisition (no known possession)—13
- False cases and hoaxes—14

Surprisingly, there have been only 222 bioterrorism-related incidents in a 100-year period and in only 24 cases have there been confirmed attacks—an average of 1 every 4 years worldwide. Most were abortive. Fourteen of the 24 confirmed cases of bioterrorism or biocrimes are food or agriculture-related; of these cases, 11 involved food poisoning and only 3 targeted commercial animals or plants. Of the 222 documented incidents, only 6 appear to be clearly linked to attacks on commercial plants and animals.

Significantly, yet, the survey made by Carus points at 144 incidents that occurred in the 1990s, meaning nearly two thirds of the total. This may reflect better incident tracking and record keeping in recent years, or it may indicate a dramatic increase in the propensity of terrorists or criminals to employ bioterrorism agents. Available evidence supports the latter premise. For example, FBI statistics indicate that U. S. incidents involving weapons of mass destruction using chemical, biological, radiological, or nuclear materials have soared from 37 in 1996 to over 200 in 1999, with three fourths of the cases involving bioterrorism agents—usually the threatened release of anthrax. Notably, the vast majority of incidents have been directed against individuals or small groups, not mass populations. On the whole, if any sort of extrapolation can be made, then the next decade is supposed to be a disturbing one, at the least.

## 15 STATE-SPONSORED BIOTERRORISM

*State-sponsored bioterrorism may be carried out by saboteurs either affiliated with the concerned state or acting on their own but institutionally assisted by some country specifically aware of the ultimate outcome. Biosabotage programs and projects, in part realized, have been identified mainly in Germany, Japan, USSR, the United States, South Africa, and Iraq. They are here presented in detail, illustrating the conceptual and practical paradigm of state-sponsored bioterrorism.*

States running a methodical biological weapons program would usually adopt a collateral subsystem dealing with bioterrorism means and operations. But the opposite equation remains open, in that the absence of a methodical program may not at all impair an effectual framework in charge of developing, manufacturing, and deploying of terrorism-oriented biological agents. The perpetrators tasked for carrying out the biosabotage acts may equally be affiliated with the involved state, or with another state, be it a friendly or a hostile state, including the target country itself. State-sponsored bioterrorism may as well be directed against domestic opponents and carried out internally. Another variance lies between bioterrorism initiated by a state and such that is initiated by a terror organization (or a second state) but is crucially and knowingly assisted by the concerned state. All those modes may be categorized as state-sponsored bioterrorism and are looked into through the following case studies.

### 15.1 Germany

From 1915 to 1918, Germany waged an ambitious campaign of covert biosabotage on animals being shipped from neutral countries to the Allies. The program used the germs causing glanders and anthrax, and employed secret agents to administer the bacterial cultures to animals penned for shipment. The cultures were sometimes injected using needles dipped into the cultures, sometimes poured onto feed, or (later in the war) contained in capillary tubes embedded in lumps of sugar that were fed to the animals. Horses and mules were the main targets, but in some cases sheep and cattle appear to have been targeted as well.

The programs were initiated nearly simultaneously in Romania and the United States. The Romanian campaign was administered by Major Nodolny of the German General Staff, through his Military Attaché in Bucharest. The agents disseminating the cultures were Bulgarian, run by the Bulgarian embassy. Cultures were shipped from Berlin. The program lasted until the August 1916 Romanian declaration of war against Austro-Hungary and the expulsion of German diplomats.

The campaign in the United States was operated by a U.S.-born, German-raised, physician, Anton Dilger. Dr. Dilger brought seed cultures with him to the United States in 1915 and set up a culture facility in the suburban Washington, D.C. home that he rented. He supplied cultures to the German merchant-ship captain Hinsch, stranded in the United States by the British naval blockade, who ran the agents, largely stevedores. Dr. Dilger returned to Germany in early 1916, and the campaign came to a halt a few months later [133, 134]. Similar campaigns were conducted concomitantly in Argentina, Mesopotamia, and Norway. During WWII, the only known conduction of biosabotage by Germany was the contamination of a

large reservoir in Bohemia with sewage, in 1945. Throughout that war, however, the Nazis routinely used non-German-captured civilians for induced infection experimentations.

## 15.2 Japan

Civilians were the main object of mass terror attacks launched by the Japanese bacteriological Unit 731 on multiple Chinese targets during the 1930s and 1940s, various bombs and other disseminating devices being thereupon applied. Furthermore, if terrorism at large is a term of relevance with respect to enemy civilian prisoners, then the vast biological experimentations conducted by Unit 731 on thousands of arrested civilian Chinese during the 1930s and 1940s are certainly a sort of an extremely brutal bioterrorism. First-hand accounts testify the Japanese infected civilians through the distribution of contaminated foodstuffs, such as dumplings and vegetables. There are also reports of contaminated water supplies. Such estimates report over 580,000 victims, largely due to plague, anthrax, tularemia, smallpox, and cholera outbreaks [135]. Plague was often induced through infected flees, as well. In addition, repeated seasonal outbreaks after the conclusion of the war brought the death toll to much higher. On one occasion at least (1939), the Japanese military contaminated Soviet water sources with typhoid bacteria at the former Mongolian border.

## 15.3 United States

Declassified documents reveal the past existence of at least two bioterrorism operations worked out, if not conducted, by the CIA against Cuba:

Operation FULL-UP, the objective of which was to destroy confidence in fuel supplied by the Soviet Bloc by indicating it is contaminated. The operation was to be accomplished by introducing a known biological agent into jet fuel storage facilities. This agent flourishes in jet fuel and grows until it consumes all the space inside the tank [136].

Operation MONGOOSE: A document entitled "Project Cuba", dated 18 January 1962, sets forth the aims and the 32 original tasks of what subsequently became known as "Operation MONGOOSE [137]." For task 21, it states that, on 15 February 1962, the CIA would submit a plan to disrupt the harvest of food crops in Cuba. The following two sections of the declassified text of this document, which might be expected to clarify the method to be used in pursuit of this objective, appear to be censored: Clearly their content was so repugnant that even the officials responsible for declassifying the document saw fit to keep that part of it secret. However, in materials provided to the U.S. Senate Select Committee on Intelligence Activities by the CIA during the mid-1970s, the CIA acknowledged that it had developed methods and systems for carrying out a covert attack against crops and causing severe crop loss. The CIA denied that it had ever employed such systems [138].

The United States has repeatedly been accused by Cuba of biosabotage acts that allegedly occurred. Reportedly, in 1977, a U.S. intelligence source admitted that the United States used the swine-fever virus as biological warfare against Cuba. The agent told U.S. media he was ordered to transport the virus from a U.S. Army Base and CIA training center in the Panama Canal Zone to a group of right-wing

Cuban exiles who in turn delivered it to operatives inside Cuba in March 1971 [139]. This may seemingly be connected to the outbreak of swine-fever virus in Cuba on 6 May, 1971—its first appearance in the Western Hemisphere. The highly contagious virus is lethal to pigs. Six weeks into the epidemic, the Cubans were forced to slaughter a half-million pigs to stem the spread of the epidemic. Also, the *New York Times* reported in 1983 how the head of a Miami-based anti-Cuban terrorist group admitted in a U.S. court that he had taken germs to Cuba in 1980.

Furthermore, Project MKNAOMI is notable as well. In the 1970s, the CIA publicly revealed that a U.S. Army team called the Special Operations Division (SOD) at Fort Detrick, Maryland, developed biological and chemical weapons for the CIA under a Top Secret project that would last almost 20 years [140]. This project, MKNAOMI, was practically unknown at the CIA due to the extreme sensitivity of its mission. Few written records were kept. CIA personnel working at Fort Detrick used the cover of Special Support Staff of the Department of Defense. Indeed, on 23 October 1962, the U.S. Patent Office granted patent 3,060,165 to four persons "as represented by the Secretary of the Army." The patent was first filed 3 July 1952, Serial Number 297,142 for the use of ricin as a biological weapon. The strikingly honest descriptive language used to apply for this U.S. patent, in 1952, is very revealing: "Ricin is a protoplasmic poison prepared from castor beans after the extraction of castor oil therefrom. It is most effective as a poison when injected intravenously or inhaled. A very fine particle size was necessary so that the product might be used as a toxic weapon." Also, in the early 50s the SOD conducted—as revealed in a document partially declassified by the FBI—a biosabotage experiment, using the Pentagon building as a model target, so as to prove that all persons present therein would be unknowingly exposed within a relatively short period of time [141]. Additional model targets where selected to simulate biosabotage operations, for example the Washington's National Airport and the New-York Subway during the 60s (details given above). The CIA and possibly other elements within the U.S. intelligence community apparently persisted in holding bioterrorism agents after the United States destroyed its biological weapons arsenal. Besides, an increasing debate has aroused as to the shear innocence of the US outstandingly growing research of virulent pathogens, referring to the fragile borderline separating between defensive- and offensive-oriented implications. However, the very same debate may equally pertain, in principle, to other states.

## 15.4 South Africa

Rather unexpectedly, an extremely meticulous biosabotage program has been uncovered in South Africa. Apparently, it constitutes an outstanding case of a state-sponsored bioterrorism paradigm that has been fairly brought out in details [142].

The most characteristic feature of the South African biological (and chemical) terrorism program was undoubtedly the development, testing, and utilization of a wide array of hard-to-trace agents to assassinate "enemies of the state [143]." As insider testimony and the notorious "sales list" of 1989 (Telemedicine Research Center document 52) indicate, several of the highly poisonous substances produced at both Delta G and Roodeplaat Research Laboratories (ad hoc front companies)

were actually deployed by clandestine units ADF and SAP "death squads," above all the Security Branch's C[ounterinsurgency]1 section (later renamed C10), housed at the Vlakplaas base, in covert assassination operations [144].

There is no doubt whatsoever that high-ranking officers within the SADF and SAP and other "securocrats" within the government were generally aware of these activities, many of which they in fact authorized. Some civilian Afrikaner paramilitary groups, whose pro-apartheid members remain violently opposed to black majority rule, have even publicly threatened to attack their enemies with biological and chemical agents [145, 146].

In the mid-1980s, a higher level and more formalized assassination program formed when the Teen-Rewolusionêre Inligting Taakspan (TREWITS: Counter-Revolutionary Intelligence Task Force) was created. Then, on the verbal instructions of a senior cardiologist, Colonel Dr. Wouter Basson—head of Project Coast: Apartheid's Chemical and Biological Warfare Program—a host of freeze-dried pathogens and highly toxic substances that had been produced either at Delta G or RRL—was secretly transferred and thereafter stored in a refrigerator inside a fireproof and bombproof walk-in safe in his own office to military and police personnel through various channels. The deadly agents were passed on, either to the aforementioned persons in innocuous public places like restaurants or to Basson himself in the latter's office at South African Medical Services (SAMS) headquarters in Centurion. The specific recipients of these lethal substances and contaminated items were operatives of the "death squads," officers who either deployed some of them personally or later distributed them to the so-called "hit team" members; an ex-psychologist who in 1988 assumed control over Systems Research and Development; a bioengineering company set up in part to manufacture special "applicators," i.e., arcane assassination devices such as rings, screwdrivers, walking sticks, and umbrellas that had been transformed into weapons by means of the addition of poison compartments and injectors or firing mechanisms for poisoned pellets; and Basson himself. Furthermore, holes were drilled in cans of Game orange soda, into which some substance was injected, and then closed by means of soldering so that they were no longer visible.

The actual substances included the viruses of Ebola, Congo, and Marburg; lethal toxins such as mamba toxin, botulinum, and ricin; bacterial agents such as plague, anthrax, brucellosis, salmonellosis, and bottles of cholera bacteria; and a wide variety of foodstuffs, beverages, household items, and cigarettes that had been contaminated with these biological agents [147]. There can be little doubt that several of these materials, items, or devices were subsequently often used to murder or sicken opponents of the apartheid regime. If one excludes the hundreds of drugged and secretly disposed guerrilla opponents, the total number of victims appears to have been in the dozens.

International ties formed as well within that dark context. Dr. Wouter Basson, or other Coast personnel, may have transferred dangerous biological warfare materials and know-how to elements of a loose international network of right-wing extremists. Fears have been expressed that Basson and other Coast scientists were associated with an even broader international right-wing network, purportedly known as Die Organisasie (The Organization), among whose members are said to be expatriate Rhodesians and South Africans who emigrated to other countries both during the apartheid era and as the apartheid system was collapsing [148]. Basson also had strong bonds with Arabic elements, particularly Iraqi and Libyan.

Connectedly, according to a pair of Federal Bureau of Investigation (FBI) informants, in the mid-1980s, the American doctor Larry Ford transferred a suitcase full of dangerous "kaffir-killing" pathogens to Surgeon-General Knobel at the Los Angeles residence of the South African trade attaché, Gideon Bouwer [149]. It has also emerged that at Knobel's request Ford lectured Coast scientists about the contamination of household items with biological agents [150–153]. The whole apparatus was dismantled during the 1990s. Still, such a complex state-sponsored bioterrorism program, operating both inside South Africa and beyond, all over, has not elsewhere been uncovered publicly, yet certainly not an isolated one, worldwide; the Soviet, and later Russian parallel system did not—and apparently does not—lag far behind.

## 15.5 USSR and Russia

Two bioterrorism affairs, one momentary—the planned assassination of an opponent by ricin toxin in London—and the other one lasting for years—the relapsing employment of fungal toxins against civilians in South East Asia—were sponsored, in effect, by USSR, whereas Bulgaria and North Vietnam were involved as partners, respectively. Not singular (referring to the Soviet apparatus at large), those two affairs attracted much attention. Although the latter has been surreptitiously—yet massively—conducted by the Soviet MOD (details given in Section 9.3), the former was a fine product of the KGB.

Various modes of biosabotage have been specifically encountered with regard to the USSR and Russia, in the book titled *Biological Espionage*, written by former KGB officer Alexander Kouzminov who worked for the so-called "Department 12, Directorate S (the KGB Operational Technical Support Directorate) [154]." This directorate oversees Moscow's "illegals" (i.e., Russian agents posing as Westerners, operating under deep cover). Department 12 (and Department 8) of the KGB were tasked with preparing "clandestine acts of biological sabotage against 'potential strike targets' on the enemy's territory." These potential targets include military research laboratories, combat units, weapon stockpiles, public drinking water, food stores, vaccine repositories, pharmaceutical plants, and the overall economy of the target country. These two departments would also assassinate, incapacitate, or kidnap foreign officials, political enemies, and "important persons" (where "important" is determined by the exigencies of war).

Department 12 was referred to in intelligence circles as the "Chamber" or "Kamera." KGB General Viktor Chebrikov, Andropov's closest subordinate in the KGB, was then the Director of this department. The "Chamber" developed, among many technical devices and substances various toxins, such as ricin, which is but one example. The work of Department 12 "has grown" since the collapse of the Soviet Union. Genetic engineering has brought forth new bioterrorism horrors. And these are to be unleashed in the event of something called "Day X," which signifies, conceptually, the beginning of the next world war—a large-scale war. Soviet military thinkers believe that such a war will tentatively involve the mass use of nuclear as well as biological weapons, concomitant with multiple bioterrorism operations. The KGB thus had its own R&D centers for special toxic matters including bioterrorism toxins and pathogens for use in espionage aims and sabotage (so-called "Fleita" program). In parallel, aerosol-released bioterrorism weapons were often in the possession of Spetsnaz operatives (Soviet military Special Forces).

Connectedly, the manners of the KGB are well illustrated in detail through the Soviet–Bulgarian ricin plot. Georgi Ivanov Markov was born 01.03.1929 in Sofia, Bulgaria and died in London due to induced ricin intoxication. He was executed by DS (Durzhavna Sigornost—Bulgarian State Security) in London on 11.09.1978, thanks to vital assistance knowingly afforded by the KGB. A totally independent journalist, Markov was Bulgaria's most revered dissident and Bulgarian Communism's arch enemy. Bulgarian Communist dictator, Todor Zhivkov, was very well informed about Markov's activities by the DS [155]. In 1977, Zhivkov asked the KGB to help him silence Markov. The Russians did not hesitate. Both President Yuri Andropov and Vladimir Krutchkov, head of KGB, personally approved and ordered General Sergei Golubev (Chief of the Security Service and specialist on "murder") to cooperate with DS. The KGB granted DS access to the resources in the "Chamber."

Golubev received instructions in the KGB headquarter relevant "Chamber," and the next week he flew to Sofia with Ivan Surov. Surov's job was to transmit the Bulgarian Intelligence Service practical know-how in the use of special poisons, which could not be traced after the victim's death. Golubev and Surov discussed with the Bulgarian's intelligence officers the various options of killing Markov. They worked out one plan to use a poison that could be surreptitiously dissolved in tea, coffee, or any liquid that Markov might drink. In 1978, Gen. Golubev traveled to Sofia three or four times to help DS with planning of the secret operation. Three attempts to assassinate Markov followed. The first attempt was made in Munich in the spring of 1978 when Markov was visiting friends and colleagues at Radio Free Europe. Someone put a toxin into Markov's drink at a dinner in his honor. The attempt to kill him failed. The second assassination effort occurred on the Italian island of Sardinia, where Markov was on vacation with his family. The plan also failed for reasons unknown.

Golubev returned to Sofia to work out a new plan to kill Markov. The KGB decided to use a camouflaged weapon. A folding umbrella was adapted with a firing mechanism and silencer to shoot a small pellet at close range, one and a half to two meters. Golubev requested that the KGB Residency in Washington purchase several U.S.-manufactured umbrellas and send them to the Center. The KGB head resident in Washington bought several umbrellas and sent them to the KGB, and an OTU operational technical unit rebuilt the umbrellas. The "Chamber" then adapted the umbrella tip to enable it to shoot the victim with a tiny metal pellet containing ricin. Golubev then took the converted umbrellas to Sofia to instruct the assassin on how to use this weapon. The pellet was supposed to penetrate the clothing and be lodged in the upper skin layer. Consequently, the final, and successful, attempt was staged in London 07.09.1978, on Bulgarian's Communist dictator Todor Zhivkov's sixty-seventh birthday. Before his assassination, Markov received a threatening anonymous phone call: "Not this time," said anonymous caller. "This time you will not become a martyr. You will simply die of natural causes. You will be killed by a poison that the West cannot detect nor treat." The tasked ricin toxin did the job, indeed.

On the day of his assassination, Markov worked a double shift at the BBC. After finishing the early morning shift, he went home for rest and lunch. Returning to work by car, he drove to a parking lot on the south side of Waterloo Bridge. It was his habit to take a bus across the half-mile Waterloo Bridge to the BBC head-

quarters in the Bush House. After parking his car in a parking lot near the Waterloo Bridge, Markov climbed the stairs to the bus stop. As he neared the queue of people waiting for the bus, he experienced a sudden stinging pain in the back of his right thigh. He turned and saw a man bending to pick up a dropped umbrella. The man who was facing away from Markov, apologized. The assassin then hailed a taxi and departed. Although in pain, Markov boarded the bus to work. But the pain continued.

Markov noticed a small blood spot on his jeans. He told colleagues at the BBC what happened and showed one friend a pimple-like red swelling on his thigh. By evening, he had developed a high fever. He was hospitalized and treated for an undetermined form of blood poisoning. His condition fast worsened. He was not responding to doctor's efforts. The next day he went into shock, and after three days of agony and delirium, he died on 11.09.1978.

An autopsy was performed at Wandsworth Public Mortuary. The doctors found a tiny metal sphere the size of a pinhead in the wound. When they attempted to extract the "pin," a tiny pellet fell on the table. The police took the pellet to the Chemical and Microbiological Warfare Establishment at Porton Down, commonly called the "Germ Warfare Center." There, a team of the England's foremost specialists in forensic medicine, and, reportedly, Dr. Christopher Green of the CIA examined the pellet. The pellet was 1.52 mm in diameter, embedded in his calf, and composed of 90% platinum and 10% iridium [156]. They found that two 0.34-mm holes had been drilled in the pellet, possibly using a high-technology laser at right angles to each other, producing an X-shaped cavity. The holes were empty.

This prevented investigators from establishing the type of substance that had been used, but it was sufficient to determine that Markov had "not died of natural causes." BATS (British Anti-Terrorist Squad), detectives then joined the Scotland Yard investigating team. After weeks of research and experimentation, in January 1979, a Coroner's Inquest in London Gavin Thurston ruled that Georgi Markov had been killed via ricin toxin. Traces of ricin were possibly found later thereupon.

But several years later, two former top KGB officers, Oleg Gordievsky and Oleg Kalugin, publicly admitted Soviet complicity in Markov's murder by means of ricin toxin. The case was dormant until after the arresting of Communist dictator Todor Zhivkov and the fall of the Communist government in Bulgaria in 1989. Although the DS act about the case Markov is destroyed or has been sent to Moscow, there is no doubt that some DS and KGB officers know the true about his death.

In 1991, former chief of Bulgarian Foreign Intelligence Vasil Kotsev, who was identified as the person in charge of the Markov operation, died in a questionable and mysterious automobile accident. A second suspect, General Stoyan Savov, preferred to commit suicide on 09.01.1992 rather than face trial also for destroying the documents. In 1994 the British Parliament asked Russia to help them find 15 past KGB agents who might have been involved in or had knowledge of the murder. The request remains unanswered. The Markov murder case remains officially unsolved. No one has been brought to justice for the murder of Markov, although *prima facie* evidence points fairly clearly at the involvement of Soviet and Bulgarian elements, chiefly the KGB.

The Russian KGB officially ceased to exist in November 1991, but its successor organization, the FSB, is functionally extremely similar to the KGB. Even after

many details about the related bioterrorism mechanisms were revealed, during the 1990s, and although many alterations took place in Russia, this system did not substantially change.

## 15.6 Iraq

In the past, Iraq conducted bioterrorism in the following cases:

In 1988, the Iraqi army deliberately introduced typhoid bacteria into the water supply of the Kurdish city of Sulaimaniyah, bringing about an outbreak [157].

In 1989, outbreaks of cholera were generated as a result of BWA being experimented with on Kurdish populations by Iraq [158].

In 1990, an apparent, severe malaria outbreak that exceptionally occurred in the Kurdish Biharka concentration camp in Iraq has been attributed to Iraqi experimental employment of a BWA [159].

Also, Iraq developed and produced aflatoxin (at least 2200 L) as a long-term debilitating bioterrorism agent intended to be used against the Kurds. It has thereupon been noted that:

> The discovery (by UNSCOM, during the 1990s) that Iraq was researching aflatoxin, not a traditional BW candidate, was a cause for some surprise. It is a carcinogen, the effects of which manifest themselves only after many years, and several Western experts have rationalized this Iraqi program only in terms of genocidal goals. If aflatoxin were used against the Kurds, for instance, it would be impossible definitively to prove the use of BW once the symptoms emerged. Another possible explanation is its potential use as an immune suppressant, making victims more susceptible to other agents. However, the aflatoxin declaration may also hide other aspects of Iraq's BW program: according to Iraq's depositions, the production program never encountered any mishap (as other parts of the BW program had) and, to judge from the declared time-frame for the total amount produced, production could never have stopped, even for cleaning of the equipment. This raises the suspicion that Iraq declared an excessive amount of aflatoxin in order to disguise the fact that other, more destructive agents had been produced in greater quantities [59].

Being a very potent, although typically slow-acting hepatotoxin and nephrotoxin, aflatoxin may have been developed as a BWA by the Iraqis for purposes of long-term terrorism, or some short-term acute impact of yet unknown nature [5]. Iraq was unable to justify the weaponization of aflatoxin from the research data obtained from its own experimentation. One bomb that by Iraq's account should have contained aflatoxin instead tested positive for botulinum toxin and negative for aflatoxin [160]; this indicated another mode of weaponization. And indeed, there is documentary evidence and statements obtained by UNSCOM that Iraq was mixing aflatoxin with riot-control gas. They note that it would not be unthinkable for a leader who has used chemical weapons on part of his population. "It is a great way to keep colonels from becoming generals," said an UNSCOM inspector familiar with the searches. "Saddam Hussein hasn't made any weapon that he hasn't used on his own population, with the exception of Scud missiles [60]." Doubtfully, the entire, tainted past Iraqi–Kurdish interface has been revealed by him since being in prison.

Notably, the Iraqi Intelligence Service provided the BW program with security and participated in bioterrorism research, probably for its own purposes, from the beginning of Iraq's BW effort in the early 1970s until the final days of Saddam Hussein's regime. The Iraqi Intelligence Service had a series of laboratories that conducted bioterrorism work including research into BW agents for assassination purposes until the mid-1990s. The Iraqi Survey Group found that the Iraqi Intelligence service produced ricin and tested it on political prisoners. ISG has not been able to establish the scope and nature of the work at these laboratories or determine whether any of the work was related to military development of a BW agent [161]. But they could certainly suffice to sponsor and support the 2001 anthrax letters' sabotage, together with al-Qaeda, as has previously been suggested [80]. All in all, the option for bioterrorism operations was persistently maintained by Iraq until its collapse in 2003 [162].

### 15.7 Iran

According to Eisenstadt, Iran has probably deployed BWs, which it could deliver via terrorist saboteurs, aerosol tanks mounted on aircraft or ships, or via missiles [163]. Taking care, thus, of the entire spectrum of BWs, Tehran did not ignore their importance in the context of terrorist actions; it equipped itself with micro-warfare means destined to employ bioterrorism agents by on-spot spraying and by contaminating water systems. Iran, as well as Muslim BW possessors like Sudan and Syria, can readily be assisted by some Muslim terror organization, in case they want to carry out bioterrorism acts through nonresidential saboteurs. More concretely, such cooperation between Iran and its terror organization ally—the Hezbollah—has been feared.

Other states runing BW programs that may potentially support of sponsor bioterrorism acts include, mainly, Syria, North Korea and Cuba.

## 16 NONSTATE-SPONSORED BIOTERRORISM

*Persisting organizations, temporary communes, or sporadic individuals may initiate and conduct bioterrorism acts, without there being any institutionalized assistance from any country. The following section deals with those three types of elements, practically or potentially involved in bioterrorism-related activities. Also, it presents in detail prominent episodes of in-effect bioterrorism (further episodes are included in Section 14, 17 and 18).*

### 16.1 al-Qaeda

This terror organization has been, and still is, the most significant one worldwide, in terms of bioterrorism, especially that it was involved, most probably, in the Sept. 2001 anthrax letter attack. In the mid-1990s, while being sheltered in Sudan, its head, Osama bin Laden, financed, in part, the construction of BW (and CW) facilities in Sudan. Later on, members of al-Qaeda were apparently trained in Iraq (by its intelligence apparatus) for BW (and other nonconventional weapons) employment. In 1998, relationship with Iraqi intelligence was established, so as to obtain

poisons—appatently toxins—and gases training. After the USS Cole bombing in 2000, two al-Qaeda operatives were sent to Iraq for CBW-related training beginning in Dec. 2000. Iraqi intelligence was "encouraged" after the embassy and USS Cole bombings to provide this training [164].

Moreover, in August 2002, reports have emerged that Ansar al-Islam, an al-Qaeda affiliate active in Iraqi Kurdistan since September 2001 as a militant group, has been involved in testing various poisons including ricin [165]. According to an ABC News report, the group tested ricin powder as an aerosol on animals such as donkeys and chickens, and perhaps even an unwitting human subject. The experiments were ordered and financed by a senior al-Qaeda official, who was providing money and guidance from elsewhere in the region.

Also, selected al-Qaeda terrorists were guided as to the methods of cultivating germs and preparing toxins to be used as bioterrorism agents and how to convert them into weapons, exploiting easily available equipment and materials. Those who plotted in the caves of Afghanistan have left behind diagrams of American cities and landmarks. Also, manufacturing instructions have been posted on the Internet. One of the simulated operations in Kandahar, Afghanistan, was contaminating a water main of a European city (apparently London) using equipment that could fit inside a backpack. Several al-Qaeda cells have been trained in Afghanistan, where they have learned to use bioterrorism agents, including anthrax, ricin, and botulism toxins. Later, after the fall of the Taliban regime, those groups continued their experiments in the Pankisi Gorge, on the territory of Georgia, bordering Chechnya [166, 167]. Not too far away, in Fallugah, Iraq, an improvised laboratory belonging to Iraqi al-Qaeda-directed rebels was revealed in 2004. It was found to contain materials for making chemical blood agents, as well as a "cookbook" on how to produce a deadly form of anthrax [168]. Two months earlier, notably, the very same Abu Musan al-Zarkawi's derivative group of al-Qaedat ("Tahwid and Jihad"—"Oneness of God and Holy War") tried hard to take care of two women known as the Iraqi anthrax masterminds. It kidnapped two Americans and a British man, threatening to kill them if Iraqi women prisoners were not released. The two Iraqi masterly women were not set free, and the three kidnapped men were then killed [169].

Another training base was located at Zenica, Bosnia. It is not clear, however, whether any experimentation practically included, beyond toxins, in-effect living germs. Reportedly, bin Laden's associates managed to receive anthrax and plague biological weapons from former Soviet facilities in Kazakhstan. Records and operations manuals captured in Afghanistan and elsewhere disclose that bin Laden has devoted money and personnel to pursue smallpox, among other biological weapons. Furthermore, uncorroborated testimony in a high-profile Egyptian trial in 1999 indicated that al-Qaeda had equired Ebola virus and Salmonella germs.

al-Qaeda, and particularly its second in command, Ayman al-Zawahiri, a physician, had long been eager to acquire biological agents, particularly anthrax, according to Khalid Shaikh Mohammed, one of Osama bin Laden's top lieutenants. Also, al-Qaeda agents had inquired about renting crop-dusters to spread pathogens, especially anthrax. During interrogations with terror suspects in custody, it has come to light that the al-Qaeda terror group has been actively seeking weapons-grade anthrax [170]. According to an article by Milton Leitenberg, computer hard drives and handwritten notes seized at the home where Mohammed was arrested included

an order to buy anthrax, along with other evidence of an interest in acquiring anthrax and other dangerous germs [171]. Until the American invasion of Afghanistan, launched after the Sept. 11, 2001 attacks, al-Qaeda's anthrax program was based in Kandahar, Afghanistan, and was led by two men: Riduan Isamuddin, known as Hambali, and Yazid Sufaat, a Malaysian member of Jemaah Islamiyah, an al-Qaeda-affiliated group. Although Sufaat tried to acquire anthrax, there is no evidence that he was able to procure the appropriate strain used for the 2001 attacks (Ames strain). This does not mean, yet, that al-Qaeda did not obtain that deadly strain through other channels. Hambali had been trying to open a new BW project for al-Qaeda in the Far-East, when he was arrested. Reportedly, anthrax powder was found, eventually, in March 2006 in a house occupied by the Taliban in Afghanistan [172].

A London-based offshoot was engaged in preparing ricin and botulinum toxins. In January 2003, British authorities arrested six Arab men that intended to produce ricin in their north London apartment, led by Kamel Bourgass, who had been trained in al-Qaeda terror camps in Afghanistan and was specially selected for instruction in making poisons [173]. Accurate recipes and ingredients for poisons including botulinum and ricin (plus cyanide), and the blueprint for a bomb were found in the apartment, in addition to castor beans.

Nonetheless, after the 2001 anthrax letter attack (covered bellow in special subchapter) and the U.S. military operations against al-Qaeda in Afghanistan, on many occasions repeated bioterrorism warnings emerged, stemming from intelligence information and from various announcements made by al-Qaeda. Just as one example, six flights bound for America, including four from Britain, were cancelled after intelligence suggested that al-Qaeda was planning a spectacular attack using a hijacked airliner and weapons of mass destruction, particularly BWs [174]. Alongside with cases of in-effect bioterrorism-oriented activities, like the ricin affair in London, various bioterrorism hoaxes that could not be identified may probably be attributed to al-Qaeda. At any rate, al-Qaeda certainly wants to use biological weapons and widen the scope. The British M15 recently reported, thereby, that al-Qaeda operatives are training in germ warfare, and trying to recruit university students with access to microbiological laboratories, so as to steal virulent pathogens [175].

## 16.2 Jemaah Islamiah

A Southeast Asian Islamic militant group affiliated with al-Qaeda, this organization does not lag far behind the latter. Connectedly, Dr Rohan Gunaratna, who heads the terror unit at Singapore's Institute of Defence and Strategic Studies, has warned Australia of a new, global generation of Islamist terrorists "armed, trained financed and ideologicised" to use biological weapons. In declaring that part of this new global terror wave would come out of Southeast Asia, he added he found the Jemaah Islamiah group had come extremely close to developing bioterrorism-chemical weapons, recounting an alarming analysis of a Jemaah Islamiah biological-terror rudimentary training manual that was taken from a Philippines safe house in late 2003. His further generalization is notable, saying "It is only a question of time that a group that has these intentions will have the capability to develop them. We are seeing a new generation of terrorists being trained in the use of biotlogical

weapons. We rarely saw this in the 1990s but today we are seeing increasingly this kind of training. Groups are being trained to use these agents, though the probability of attack is still low [176]."

## 16.3 Hezbollah

Former CIA director James Wolsey described the Hezbollah as a potential tool for bioterrorism. That threat has been accentuated and is particularly realistic, because the Hezbollah is a radical terror organization, directly supported and encouraged by Iran, which possesses biological weapons. The evident acquisition and successful employment of unmanned drones by the Hezbollah, as well as its access to the sources of the Jordan River, may increase a potential bioterrorism threat posed by this dangerous organization. In 1998 the Hezbollah attempted to obtain biological and chemical weapons through two businessmen situated in Switzerland [177]. Elements of Hezbollah and al-Qaeda joined forces to plan a series of attacks using toxic substances, apparently chemicals and toxins. Groups of al-Qaeda and Hezbollah activists from the Sidon area of Lebanon met in Africa in December 2002, to set up the needed preparations there or in Sidon [178]. The 2006 confrontation between Hezbollah and Israel, demonstrated effective operational capacities of the Hezbollah to employ a variety of Iranian rockets and missiles, some of which are tipped—in Iran—with biological warheads.

## 16.4 Radical Palestinian Terror Organizations

An article appeared on the 13 August 2001 edition of the Lebanon-based Palestinian weekly, *Al-Manar*, stating that there is "serious thinking" among the Palestinians about obtaining BWs. It says regarding BWs that "obtaining its primary components is possible without too much effort, let alone the fact that there are hundreds of experts who are capable of handling them and use them as weapons of deterrence, thus creating a balance of horror. Anyone who is capable, with complete self-control, of turning his body into shrapnel and scattered organs, is also capable of carrying a small device that cannot be traced and throw it in the targeted location [179]." It has been pointed out that Palestinian organizations are inclined to resort to the acquisition and introduction of nonconventional capabilities—in particular biological and chemical—mainly for deterrence purposes. According to that source, many Palestinians believe that biological and chemical weapons are a legitimate and desirable means in the struggle against Israel [180]. Basically, a radical terror organization, like the Hamas, is apt to acquire BWs, although so far it is not known to have obtained them, in that case.

Notably, a novel mode of infection with hepatitis B virus has been suggested, following penetrating human bone fragments due to the explosion of a Palestinian suicide bomber [96]. Connectedly, perhaps, the Tanzim planned to use a suicide-bomber arrested with explosives bearing AIDS-infected blood, the thinking being that anyone who survived would get AIDS. The operation was foiled [181]. Another Palestinian body, the al-Aqsa Martyrs' Brigades (affiliated with the Palestinian Fatah political party) sent an announcement in 2006 to the Ramattan News Agency, stating: "we are pleased to say that we succeeded in developing some 20 different

types of biological and chemical weapons, this after a three-year effort", adding that those agents might be employed under certain circumstances [182].

## 16.5 Aum Shinrikyo Cult

The Aum Shinrikyo Cult ("religion of supreme truth"), a Japan-based apocalyptic religious sect, produced biological agents and tried to use them. The Japanese police discovered that the Aum included among its members skilled scientists and technicians, including some with training in microbiology, who attempted to generate weapons using anthrax, cholera, botulinum toxin, Q-fever, and even Ebola virus [183]. These accounts also suggest that there were four separate attempts to use biological agents, including anthrax once and botulinum toxin three times:

- In April 1990, the Aum Shinrikyo outfitted an automobile to disseminate botulinum toxin through the engine's exhaust. The car was then driven around Japan's parliament building.
- In early June 1993, the cult attempted to disrupt the planned wedding of Prince Naruhito, Japan' s Crown Prince, by spreading botulinum toxin in downtown Tokyo using a specially equipped automobile.
- In late June 1993, the cult attempted to spread anthrax in Tokyo using a sprayer system on the roof of an Aum-owned building in eastern Tokyo. The anthrax was disseminated for four days.
- On March 15, 1995, the Aum planted in the Tokyo subway three briefcases designed to release botulinum toxin. Apparently, the individual responsible for filling the botulinum toxin had qualms about the planned attack and substituted a nontoxic substance. The failure of this attack led the cult to use sarin in its March 20, 1995 subway attack. A helicopter and two UAVs were purchased by the cult and intended to be equiped with sprayers.

Fortunately, the Aum scientists apparently made mistakes in either the way they produced or disseminated the agents, and no one became ill or died from the attacks.

## 16.6 Bhagwan Shri Rajneesh Cult—Typhoid Epidemic

As the summer of 1984 waned, a cult led by the Bhagwan Shri Rajneesh near the town of The Dalles, Oregon, used a biological agent to sicken hundreds in an apparent dress rehearsal to sway the outcome of a local election. Three years earlier, the Bhagwan brought followers with him from India when he immigrated in 1981. The hard-working Rajneeshees were isolationist, with some 150 armed people to keep outsiders away from their ranch. When the Bhagwan decided to enlarge his ranch and his flock, he took over the small town of Antelope, christening the new town Rajneesheepuram.

Oregon's attorney general stated, however, that the municipality was unconstitutional because it did not separate church and state. To outmaneuver the attorney

general, the cult's hierarchy hatched a plan to make the Wasco county residents too sick to vote in November 1984, enabling the Rajneeshees to seat their favored candidate on the county court. The cult's nurse was the scientific brain behind attempts to put a bioterrorism plan into action [184].

Although the cultists considered additional organisms (AIDS and *Salmonella typhi*), the Rajneeshees decided upon *Salmonella typhimurium*, so as to bring about a massive food contamination. The cult bought bactrol disks from a Seattle medical supply company under false pretenses. A trio of cult members worked in a laboratory equipped with an incubator and freeze dryer to brew what they called a "salsa." Several more Rajneeshees were involved in distributing the agent on various occasions. Starting on 29 August 1984, the Rajneeshees began sprinkling their *S. typhimurium* in personal drinking glasses, on doorknobs, and urinal handles; on produce at the local supermarket; and on salad bars in 11 restaurants.

Soon, a steady stream of patients was reporting to local physicians and hospitals, with symptoms ranging from nausea and diarrhea to headache and fever. In total, 751 fell ill. Wasco County commissioners and ordinary citizens were among the victims. Within 4 days, local health-care providers were able to identify the *S. typhimurium* as the source, but over a year passed before there was confirmation that a single strain caused all of the illnesses and the Centers for Disease Control filed its report. No one died in this test to see whether a ballot box could be fixed, but the Bhagwan reportedly observed that one should not worry if a few perished. Law enforcement authorities thought that the Rajneeshees were practicing to poison the water system of The Dalles. Cult members had already put dead rodents and perhaps raw sewage and salmonella salsa into The Dalles' water supply.

Several cult members were involved in the planning and execution of the salmonella attacks, but only two of the Baghwan's chief lieutenants were prosecuted. This pair received multiple concurrent 20 year sentences, among other penalties. According to one analyst who studied the Rajneeshees carefully, the cult did not work its way up the ladder of violence to bioterrorism. Rather, the Rajneeshees appear to have abruptly embarked on their salmonella spree as a means to a specific end.

An epidemiologic analysis was then done. Cohort and case-control investigations were conducted among groups of restaurant patrons and employees to identify exposures associated with illness. All 751 persons with Salmonella gastroenteritis were associated with eating or working at area restaurants. Most patients were identified through passive surveillance; active surveillance was conducted for selected groups. A case was defined either by clinical criteria or by a stool culture yielding *S. typhimurium*. It has been found that the outbreak occurred in two waves, September 9 through 18 and September 19 through October 10. Most cases were associated with 10 restaurants, and epidemiologic studies of customers at 4 restaurants and of employees at all 10 restaurants implicated eating from salad bars as the major risk factor for infection. Eight (80%) of 10 affected restaurants compared with only 3 (11%) of the 28 other restaurants in The Dalles operated salad bars (relative risk, 7.5; 95% confidence interval, 2.4–22.7; $P < 0.001$). The implicated food items on the salad bars differed from one restaurant to another. The investigation did not identify any water supply, food item, supplier, or distributor common to all affected restaurants, nor were employees exposed to any single common source. In some instances, infected employees may have contributed to the spread of illness by inadvertently contaminating foods. However, no evidence

was found linking ill employees to initiation of the outbreak. Errors in food rotation and inadequate refrigeration on ice-chilled salad bars may have facilitated growth of the *S. typhimurium* but could not have caused the outbreak. A subsequent criminal investigation revealed that members of a religious commune had deliberately contaminated the salad bars. An *S. typhimurium* strain found in a laboratory at the commune was indistinguishable from the outbreak strain [184].

Although the above information pertains to organizational or cultist bioterrorism, the following episodes represent individual bioterrorism.

## 16.7 Shigella Dysenteriae Type 2 Outbreak

Another case of intentional contamination of food with a bacterial pathogen occurred in 1996, involving clinical laboratory workers infected with *S. dysenteriae* type 2 [185]. *S. dysenteriae* type 2 is a rare organism, and outbreaks are seldom seen in the general population. After consumption of muffins and donuts anonymously placed in the break room of a Texas laboratory, 12 of 45 laboratory workers developed severe, acute diarrheal illness. Rapid onset of symptoms allowed investigators to retrieve a muffin sample, later cultured and found to contain *S. dysenteriae* type 2, matching those cultured from infected workers. Laboratory reference cultures were found in disarray, and cultures were missing, indicating laboratory cultures had been taken, grown, and used to contaminate foodstuffs. All case patients reported having eaten muffins or doughnuts placed in the staff break room on October 29. Isolates from nine case patients were indistinguishable from *S. dysenteriae* type 2 recovered from an uneaten muffin and from the laboratory's stock strain, a portion of which was missing.

## 16.8 Hepatitis A Outbreak

Another instance of apparent intentional contamination of food with hepatitis A virus was documented in 1965. Twenty-three cases of hepatitis were reported among personnel at a Naval Air Station in 1961 [186]. The course of disease among those infected was moderately severe with no deaths. At the time, hepatitis A (infectious hepatitis) and hepatitis B (serum hepatitis) diagnoses were made on clinical and epidemiological grounds. To determine the cause of the infections, an extensive investigation of risk factors was conducted. Dining hall records and food questionnaires identified potato salad as the most likely contaminated food source. Questioning of the sole salad cook revealed that he had experienced symptoms resembling hepatitis, and a background investigation revealed aberrant social behavior. Connectedly, two incidents of inappropriate urination gave investigators a possible route of contamination of the potato salad when coupled with the individuals' symptomology.

## 16.9 Ascaris Suum Outbreak

The last instance of intentional contamination involved four college students who were peculiarly exposed to a massive dose of embryonated *Ascaris suum* eggs (a large ringworm infecting pigs) while attending a Winter Carnival in Canada during February 1970. The most likely source of infection was a meal served to the students during the Winter Carnival [187].

More episodes, mostly of secondary importance, are mentioned below in Section 13.14.4.

## 17 THE ANTHRAX LETTERS

*A colossal event generated by means of a minute amount of anthrax powder, the September 2001 anthrax letter attack, is discussed in detail. The remarkable potentiality of anthrax as a bioterrorism agent in general, and within the context of the letter attack in particular, is demonstrated. Resultant short- and long-term impacts are accentuated, alongside with the severe inability to point at the provenance of the anthrax powder contained in the letters.*

The germ *Bacillus anthracis* is the etiological agent of anthrax, a disease primarily of herbivores, but one that humans occasionally acquire through contact with infected animals or with contaminated animal products. The mode of infection may be cutaneous, intestinal, or pulmonary. The latter represents, as well, a distinct form of employing anthrax germs as an inhalable bioterrorism agent, either sprayed (wet aerosol) or powdered (dry aerosol). The efficacy of the aerosol thus generated is shaped by a variety of inherent attributes largely depending on the way the aerosol material has thereby been structured. In general, the aerosol material contains anthrax germs spores—a resilient form of life typifying all species of the genus *Bacillus* sp.—plus various, most vital, additives.

Under optimal meteorological conditions, producing 50% fatalities over a one-square-mile area would require about just 10 g of anthrax spores, in case there is no medical intervention, whatsoever [188]. Additional data show the impact of anthrax in terms of some 20,000 to 80,000 casualties afflicted by 30 kg of this biological agent, as compared with 400 to 6000 casualties afflicted by 300 kg of sarin nerve gas, and 80,000 casualties afflicted by 20 kiloton of a nuclear weapon [189]. Moreover, official U.S. estimates for casualties produced by an airplane, flying upwind of a city, releasing successfully a cloud of anthrax germs, range from 100,000 to three million dead. An individual driving a car around a medium-sized city spewing anthrax out the tailpipe would cause some 70,000 fatalities, and two individuals in two cars in two cities, 140,000 (referring to an unprotected population). These are, of course, traditional theoretical estimates; yet they illustrate the proportional impact of a biological agent, in this case, anthrax—a noncontagious bacterial pathogen. Still, even by 2003, it was held that a little more than two pounds of anthrax spores efficiently spilled into the air over a city the size of New York could be expected to kill more than 120,000 people unless state and federal officials respond much more aggressively than they currently plan to, according to the first comprehensive computer model of such a terrorist act [190].

The September 2001 anthrax letter attack marked an outstanding, most significant milestone in the course of bioterrorism. It was a quantum leap, scarcely expected in practical terms, albeit with a lot of indicative intelligence preceding the event. Overall, this bioterrorism campaign has been, operationally, complicated, assuming that the spore powder was produced outside the United States. A most comprehensive picture has been presented by the School of Public Health, Department of Epidemiology, University of California. It meticulously covers the technical, epidemiological, medical, and logistical aspects altogether. Seven letters

containing anthrax spores were probably mailed, with similar written messages from Trenton, NJ. Five letters were sent on September 18 (postal facilities [191, 192] [probable cross-contamination]), one going to American Media in Boca Raton, Florida [193, 194] (not recovered); a second to the *New York Post* (recovered) [195–197]; a third to Tom Brokaw of NBC News (recovered) [198, 199]; a fourth to ABC News (not recovered) [200]; and a fifth to Dan Rather of CBS News (not recovered) [201]. On October 9, two more letters were sent from Trenton, NJ. (probably cross-contamination) [202–206] via Brentwood mail processing facility [207–210], one to Senator Tom Daschle (recovered), and the other to Senator Patrick Leahy (recovered) [211]. Letter(s) cross-contaminated with the Daschle and Leahy letters were sent from Trenton, NJ to Wallingford, CT, with at least one letter probably going to Oxford, CT [212]. The final anthrax case in the outbreak remains a mystery, but possibly arose from contact with the September 18 letters or cross-contamination with the October 9 Leahy letter in Trenton, NJ. To do so, spores from the Leahy letter would need to have adhered to an envelope of another letter destined for the Bronx, New York City [213]. In summary, the 22 cases (one was removed by the CDC) that comprised the American Anthrax Outbreak of 2001 likely had contact with one or more of seven spore-laden envelopes.

Four letters laden with anthrax spores were discovered, all dated by an unknown author as "09-11-01," and all sent from Trenton, NJ. Two letters were postmarked September 18, 2001 in Trenton, one of which was sent to the *New York Post* where it was handled by several staff members, and the other to Tom Brokaw of NBC, opened September 19–25 but not found until October 12, 2001 in case 2's file drawer [198]. The *New York Post* letter, handled but not opened, was found on October 19, 2001. It was dampened before being discovered, turning the spore contents into a granular or clumped state.

The second two letters were postmarked on October 9, 2001 and mailed to the Washington, D.C. offices of Senator Tom Daschle of North Carolina, Majority Leader, and Senator Patrick Leahy of Vermont, Chair of the Judiciary Committee. Both letters went though Washington, D.C.'s Brentwood mail processing facility, which handles all incoming federal government mail. Both letters contained the same anthrax strain and were of the same potency.

The Daschle letter was opened in the sixth floor office at 9:45 a.m. by an aide in the Senator's Hart Senate Office Building suite on October 15, 2001. It was believed to contain about 2 g of powder comprising 200 billion to 2 trillion spores. Based on nasal swabs, all 18 persons who were in the area of Daschle's sixth floor office tested positive for anthrax exposure, as did 7 of 25 (i.e, 28%) in the area of the Senator's fifth floor office (an open staircase connected the two offices).

The Leahy letter never arrived at his office. Instead an optical reader misread the handwritten 20510 ZIP code for the Capitol as 20520, which serves the State Department. As a result, the letter was routed to the State Department, where it arrived on October 15, infecting a State Department postal worker (case 20) [211]. Shortly thereafter, all mail was isolated and sealed in plastic bags for a later search.

On November 16, 2001, the Leahy letter was found; then after special preparation, it was opened on December 6 in a laboratory setting. It contained about 1 g of anthrax, made fresh no more than 2 years before it was sent. The contents of the enclosed letter were identical to the wording of the Daschle letter. The anthrax

spores in the Daschle and Leahy envelopes were uniformly between 1 and 3 μm in size, and were coated with fine particles of frothy silica glass. More investigation is underway.

Also processed at the Trenton, NJ postal facility was a small number of letters sent to the Southern Connecticut Processing and Distribution Center in Wallingford, CT. Here letters arrived on October 11 that had been cross-contaminated with anthrax spores from the October 9 Daschle or Leahy envelops. Anthrax spores were found on mail-sorting equipment in Wallingford. One letter that went through the Wallingford distribution center was found in Seymour, nearby to Oxford where case 23 resided. Likely case 23 was infected via a similar cross-contaminated letter that came in contact with mail in the distribution center in Wallingford (not discovered) [212].

Most uncertain is the origin of case 22, although there is a connection between case 22's neighborhood and the Trenton, NJ post office [213]. A printout from the post office showed that an unrelated letter went to a shop around the corner from case 22's home. This unrelated letter was processed 2 minutes after the Leahy letter and 18 minutes before the Daschle letter. Thus the mail sent to case 22 might also have been cross-contaminated with spores from the Leahy or Daschle envelopes. Alternatively, case 22 might have had contact with one or more of the unrecovered September 18th letters following their disposal in Manhattan, similar to case 19 and case 21 [196].

But the anthrax letters affair was not limited to the US. The American embassy in Vilnius, Lithuania, was likewise concurrently targeted. Also, Three businesses in Karachi, Pakistan, were targeted, and received anthrax letters. Anthrax spores were found in a letter sent from Europe to a doctor's surgery in Chile, according to Chilean officials [214].

For the time being, the culmination of bioterrorism worldwide has been this act of distributing mail envelopes containing anthrax spore powder. It reflected noticeable supremacy of a simple act of bioterrorism (irrespective of preparing the anthrax powder in itself, which was very sophisticated), in several senses:

a. Uncontrollable preparing of the postal envelopes containing the anthrax powder
b. Uncontrollable, repeated mailings
c. Undetectable conveying of the mailed envelopes until reaching their various destinies
d. Untraceable footprints of the perpetrators
e. Inability to even postulate whether the sabotage was state or nonstate sponsored

Despite or, rather, because of those points of supremacy, the leading agencies responsible for foreseeing or, at the least, deciphering in retrospect such strikes—FBI, CIA, DIA, and the like—were after nearly 2 years still reluctant to grant public access to an unclassified report on lessons learned from the anthrax letter attacks of 2001; the Pentagon has finally released a redacted version of the document [215]. "The anthrax attacks revealed weaknesses in almost every aspect of U.S. bioterrorism-preparedness and response," the report said. "As simple as these

attacks were, their impact was far-reaching." The report, authored by David Heyman of the Center for Strategic and International Studies, is based on a day-long forum convened by CSIS under contract to the Defense Threat Reduction Agency in December 2001.

It provides a detailed and informative but hardly unsuspected inventory of shortcomings in emergency preparedness and response. Ironically, the Defense Department's two-year denial of repeated requests for release of the document exemplifies one of the central problems identified in the report: "The failure to communicate a clear message to the public was one of the greatest problems observed during the anthrax attacks . . . [including] failure to provide timely and accurate information." The Department's inability to efficiently process requests for this document suggests that it still has a long way to go to remedy this particular failure. The redacted version of the report was finally released in response to a Freedom of Information Act (FOIA) appeal [215].

Also, the Pentagon did not contend that the withheld portions of the anthrax report are classified (exemption 1), even though they are said to concern "vulnerabilities and capabilities of the US Government to respond to another . . . attack." Rather, in an expansive interpretation of the law, it said that release of the redacted portions would make it possible to "circumvent Department of Defense rules and practices. . . ." (exemption 2). See the March 15 transmittal letter from H.J. McIntyre of the DoD FOIA Policy Office here: Mike Denny Senior Force Protection Consultant Guardian Group International (GGI) [216].

In the long run, the impact of the anthrax letter attack still prevails in several respects:

- Individually, those who survived the infection have not been fully cured.
- Globally, the short-run impact encouraged terror-oriented entities to persistently explore or increase bioterrorism options.
- Practically, numerous consequent hoaxes of bioterrorism threats through powder containing post envelopes extremely proved burdensome to wide logistical systems.
- Environmentally, even as of 2005, rooms contaminated in 2001 are repeatedly disinfected.
- Publicly, the interface with institutional classified information holders sharpened.
- Strategically, the provenance of the anthrax powder has not been deciphered, although somewhat traced [217].

Thus, since the anthrax attacks in 2001, work at the nation's post offices has been disrupted by more than 20,000 incidents of suspicious powder leaking from envelopes and packages. All but a few have turned out to be nothing more than soap, dust, talc, or other nonlethal substances. Sand was the culprit in one incident included in wedding invitations for a beach ceremony. Still, the scares have taken a toll in nerves, lost time, and money. Postal workers have been instructed, if they see something like that to consider it dangerous. The area then is sealed off, and local hazardous materials teams and the Postal Inspection Service are called in. Leaking suspicious powders that turned out to be harmless include powdered Alfredo sauce, ground lentils, pudding mix, and coffee creamer, officials said.

Other cases have included leaking samples of detergent, sugar, or baking soda [218].

In November 2003, a secret cabinet-level "tabletop" exercise, designed to be very stressful to the system was thus conducted, which simulated the simultaneous release of anthrax in different types of aerosols in several American cities. The drill, code named Scarlet Cloud, found that the country was better able to detect an anthrax attack than it was 2 years ago, said officials knowledgeable about the exercise. But they said the exercise also showed that antibiotics in some cities could not be distributed and administered quickly enough and that a widespread attack could kill thousands; particularly, enormous difficulties stopping the spread of contamination through the country and into Canada were observed [219].

The drill was notable for the top-level attention it drew and the gaps it showed in the effort to protect against bioterrorism. About three dozen senior officials involved in domestic defense, including two cabinet officers—the secretary of homeland security, and the secretary of transportation—as well as the head of the White House's Homeland Security Council, participated in the exercise at the Pentagon's National Defense University.

Moreover, the contents of the former American Media Inc. building in Boca Raton, FL—namely, thousands of boxes—exposed to anthrax in 2001, were set to be decontaminated for a second time to ensure all of the pathogen is eliminated, given the approval of the Environmental Protection Agency and the Palm Beach Country Health Department [220].

Overall, the impact of the anthrax letter attack was colossal. In a sense, the very fact that the anthrax letter attack has not been deciphered since 2001 is no less meaningful and important than this unprecedented act of bioterrorism itself. In contrast to the wealth of empirical data collected and published with respect to the specific Ames strain of the 2001 anthrax letter attack and the resultant medical cases, a relative paucity of information has been brought out regarding the chemical and physical properties of the substance contained in the mail envelopes, namely the anthrax spore powder. Primarily its provenance has not yet been deciphered. This peculiar powder certainly reflects an outstanding assortment of different disciplines: microbiological, technological, medical, political, and strategic. Connectedly, it still poses, in parallel, an enigmatic, extremely complex intelligence issue that ought to be elucidated, for numerous reasons. Although the anthrax germs used for this act of bioterrorism have been fully identified to be indistinguishable from the strain Ames—a remarkably virulent strain initially isolated in 1981 in Texas but later shared with laboratories outside the United States—it is still not even known whether the anthrax powder contained in the envelopes has been structured in the United States or elsewhere [221]. Various indications point at Al-Qaeda together with Iraq as being the perpetrators [80].

Overall, dozens of buildings were contaminated with anthrax as a result of the five mailings, which contained, altogether, about 18 gr. of the sabotage spore powder. The decontamination of the Brentwood postal facility took 26 months (and cost US$130 million). The Hamilton, NJ postal facility remained closed for 41 months (its cleanup cost US$65 million). The Environmental Protection Agency spent US$41.7 million to clean up government buildings in Washington, D.C. [222]. One FBI document said the total damage exceeded US$1 billion [223].

## 18 THE RICIN COURSE

*Various traits making ricin the toxin of choice for bioterrorism purposes are discussed. The actual outcome of that appeal, as shown, has for long been relapsing episodes involving ricin, worldwide. Clear tendency emerges of al-Qaeda and affiliated groups, at the least, to persist in deploying ricin, and is doubtfully controllable.*

Ricin molecules contain a heterodimeric type 2 ribosome-inactivating enzyme (32 kDa), also known as the A chain. It is a very potent cytotoxic RNA N-glycosidase that inactivates eukaryotic ribosomes. It is linked by a disulfide bond to the galactose/N-acetylgalactosamine-binding lectin (34 kDa), also called the B chain. Weakness, fever, cough and pulmonary edema occur 18–24 hours after inhalation exposure, followed by severe respiratory distress and death from hypoxemia in 36–72 hours. Direct administration into the blood system brings about a similar picture, whereas food poisoning is basically slower, certainly dose-dependent [224].

This protein plant toxin may be regarded, foremost, as the prime toxin for bioterrorism, owing to the integral formed by several different characteristics, altogether:

a. High toxicity
b. Atypical symptomatology
c. Effective contraction through respiratory and alimentary tract
d. Effectual administration through blood stream
e. Availability of the raw material (castor beans)
f. Typical dual usability of the raw material
g. Easiness of manufacturing

Ricin's properties have been known since ancient times, and it has a long history of accidental and intentional intoxication [225]. Ricin was studied as a possible weapon in World War I, and the United States, Canada, and the United Kingdom developed it as a field weapon during World War II, also known as Agent W [226]. The World War II effort led to refined ricin in a crystalline form, although it was the amorphous form that was used in high-explosive bombs and shells and in more specialized delivery systems, such as plastic containers and cluster bombs. Attaining effective particle sizes with ricin powder is difficult, and use of the material in water or suspended in glycerol or carbon tetrachloride is more effective [227]. The agent is difficult to detect. It is fairly stable in clear, dry weather, persisting in the soil or environment for up to 3 days [228].

But the perfect usability of ricin-oriented for sabotage acts, in parallel, has not at all been ignored. The course of ricin oriented sabotage began long before 1978; yet its in-effect emergence took place in that year. It was the successful ricin-based assassination of the Bulgarian anticommunist reporter Georgi Markov that then occurred in London (full details given above). Two years later, an assassination of CIA agent Boris Korczak occurred in McLean, Virginia, Tyson's Corner, ricin being used as weapon, possibly repeatedly in umbrella configuration [76].

The attractiveness of ricin did not diminish because it found new grounds in the United States. In 1983, the FBI obtained one ounce of ricin oriented in a 35-mm film canister from an individual in Springfield, Massachusetts, who had manufactured it himself. This is believed to be one of several confiscations of ricin [76]. In 1993, an Arkansas man with survivalist group connections attempted to smuggle 130 g of ricin from Alaska into Canada to use as a weapon [164]. Later, in 1995 two members of the Minnesota Patriots Council were convicted of conspiracy to assassinate a deputy U.S. Marshal and International Revenue Service agents by ricin [127].

Syria and Iran acquired ricin as a weapon during the 1990s. Also, the Iraqi Survey Group concluded that the Iraqi Intelligence Service produced ricin contemporarily, and tested it on political prisoners. The toxin subsequently became an agent of increasing interest within some Muslim terrorist organizations. Thus, in August 2001, the FSB (Russian Federal Security Service) claimed it had intercepted a recorded conversation between two Chechen field commanders about instructions on the "homemade production of poison" for use against Russian soldiers. Russian authorities reportedly then seized materials, including confiscated papers also containing instructions on how to produce ricin from castor beans. The Russian response did not take place too late, employing the lethal castor molecule. In March 2002 the FSB apparently continued its long tradition of poisoning, while killing Chechen rebel Amir Khattab, by sending him a letter impregnated with ricin [229].

al-Qaeda got into action. Traces of ricin and instructions on its use were discovered in 2002 in an al-Qaeda house in Afghanistan [230]. The head of an al-Qaeda affiliate network, Abu Massab al-Zarqawi, was trained in the use of ricin in Afghanistan before he relocated to Iraq in 2002. It was regarded to be an ideal bioterrorism weapon for terrorists. An al-Qaeda suspect arrested in Italy before the 2003 Iraq War told his interrogators that members of the al-Zarqawi network had purchased toxins from Iraq [231]. Ricin, botulinum, and gas-gangrene toxins were the main toxins possessed by Iraq.

And, indeed, in August 2002 reports have emerged that Ansar al-Islam, an al-Qaeda affiliate active in Iraqi, has been involved in testing poisons and chemicals including ricin. The toxin powder has been tested as an aerosol on animals such as donkeys and chickens, perhaps even on an unwitting human subject.

The outcome has first surfaced in London, marking the ricin route from Asia onto Europe: London, 5 January 2003. British authorities arrested six men suspected of producing ricin in their north London apartment. All six, including two teenaged asylum seekers and four individuals in their twenties and thirties, are believed to be Arabs from Algeria or other North African countries. On 8 January, a seventh man, age 33, was also arrested in connection with the case [232]. The next day sources in Whitehall indicated that at least one of these seven men had attended an al-Qaeda training camp in Afghanistan, whereas others appear to have received terrorist training in Chechnya and the Pankisi Gorge region of the Newly Independent State of Georgia [233–235].

In the apartment where the six original arrestees resided, the authorities found several castor oil beans and equipment that could be used to process those beans. The toxin has not been detected. Five other locales were subsequently searched in conjunction with this incident, and on 13 January, Scotland Yard officials arrested

five more men and a woman in Bournemouth. One day later, another Islamist who was being arrested by Manchester police attacked them with a knife, killing one officer and wounding four others [236]. British antiterrorist investigators soon identified this same 27-year-old Algerian, "Kamel Bourgass" (also known as Nadir Habra), as "a very senior player" in the network thought to be behind the ricin plot [237].

These events occurred after a series of arrests of other Islamist radicals in Rome, Paris, and London. Because some of these individuals, particularly those in the Rome case, were allegedly planning to carry out terrorist attacks using poisons, the most recent incident in London raised several potentially worrisome questions [238]. One is whether the arrestees actually intended to employ the deadly toxin as an assassination or mass casualty weapon. A second is whether they were linked to components of the al-Qaeda network, such as the Groupe Salafiste pour la Predication et le Combat (GSPC: Salafist Group for Preaching and Fighting—see footnote) in Algeria, or were instead affiliated with some other Islamist terrorist organization such as the Algerian Groupe Islamique Armée (GIA: Armed Islamic Group). A third is whether certain states that have previously produced and tested ricin as a potential weapon played any role at all in transmitting their technical expertise or unused stocks of ricin to violent anti-Western Islamist groups. Iran, Syria, and Iraq (until occupied, if not later) are known to possess ricin.

Still in Europe, ricin was then discovered in Paris, two months later. Traces of the toxin have been found at a railway station in Paris, according to the French interior ministry. Two vials of the potentially deadly substance were found inside a locker at the Gare de Lyon, according to ministry officials. The locker contained two vials with a powder, a bottle filled with a liquid and two smaller bottles also containing a liquid. The two smaller bottles contained traces of ricin in a mixture that turned out to be a highly toxic [218].

The United States was repeatedly the next station of the ricin course, chronologically, although apparently involving but domestic American perpetrators questing to achieve influence. In October 2003, a metallic container was discovered at a Greenville, SC postal facility with ricin in it. The small container was in an envelope along with a threatening note. Authorities did not believe this was a terrorism-related incident. The note expressed anger against regulations overseeing the trucking industry. But one month later, the target was the White House.

The U.S. Secret Service intercepted a letter addressed to the White House in November 2003 that contained sparkled ricin powder, but it never revealed the incident publicly and delayed telling the FBI and other agencies; the affair was disclosed in February 2004. It was addressed to the U.S. Department of Transportation. The letter—signed by "Fallen Angel" and containing complaints about trucking regulations—was nearly identical to one discovered in October 2003 at a Greenville, SC, mail-sorting facility. But the existence of a similar letter sent to the White House was not disclosed until yesterday, and then only by law enforcement

[1] The name *Salafi*, which derives from the Arabic verb *salafa* (to precede), refers to the original companions of Muhammad, who are collectively known as al-Salaf al-Salih (the virtuous forefathers of the faith). In this context, a Salafi is a traditionalist who demands that all Muslims follow the exemplary, pious, and uncorrupted behavior of Muhammad and his trusted original companions. Note also that the GSPC is a breakaway faction of the GIA that later allied with al-Qaeda.

officials who asked not to be identified by name. Until then, the whole thing was absolutely kept under wraps on a national security basis.

Although seemingly—at the least—purely domestic, Secret Service spokeswoman declined to comment on details of the case or why it was kept secret, citing the ongoing investigation. He also declined to comment as to whether workers at the mail facility were tested or underwent decontamination procedures, and said the facility's location was kept secret for security reasons. Officials have previously noted that one mail facility used by the White House is located at Bolling Air Force Base. The letter is believed to have been sent from the Chattanooga area as it passed through a mail facility there [240]. The subsequent and last link in the United States, for now, has been a ricin-containing postal enveloped mailed in Feb. 2004 to the U.S. Senate office of the Majority Leader. Federal investigators have examined about 20,000 pieces of mail in hopes of finding the source of the ricin discovered on Capitol Hill, but so far they have turned up nothing to lead them to a suspect in the case [241].

One month later, possibly by sheer coincidence, FBI agents searching Miron Tereshchuk's home in March 2004 found firecrackers, chemicals, recipes for ricin, and a postcard with the message "Here is your poison—enjoy," according to documents filed in the Virginia court. Back to Iraq, the same month, the U.S. Army staged raids in Baghdad, uncovering evidence of an attempt to produce ricin, so as to put some in mortar rounds [242]. And again, in the United States, in November 2005, plans for producing ricin, along with bomb materials and diagrams were found in the home of Sergio Maldonado, a 20-year-old man known as a criminal [243].

In a markedly similar event (June 2006), a jar containing ricin derived form castor beans, along with ricin residue in a bowl, gun silencers, pipe bombs and bomb-making materials, were found in a shed in Nashville, Tennessee [244].

Most probably, the ricin course did not reach an endpoint.

## 19 THE SMALLPOX COMPLEX

*The unique properties and particular importance of smallpox virus are discussed, both as a bioterrorism threat and as a major biohazard in general. It appears to pose, potentially, a colossal menace, which ought to be intensely coped with. The issues of renewed vaccination and ongoing research are then addressed, showing the complexity related to that singular pathogen.*

The virus causing smallpox appears to be, objectively, the most devastating pathogen ever faced—as far as known—by humanity. Presumably, many billions of people lost their life due to smallpox during human history (an estimated 300-million people worldwide in the twentieth century alone). The virus has been effectively used for bioterrorism purposes (details given above). It has successfully been eradicated worldwide, primarily thanks to the fact that it is infective toward man only, hence not having the capacity to endure within any other creature. This plainly means that the absence of this pathogen from human populations signifies for its final elimination, except for persisting in the frozen state in laboratories or, perhaps, in nature. Moreover, the lacking of alternative hosts, apart from man, considerably restricts—both quantitatively and qualitatively—the overall gene pool

of this virus and, therefore, its ability to markedly evolve and give rise to novel variants capable of evading human herd immunity, which has extremely been fortified in the past by mass inoculations. Resultantly, the magnitude of susceptible unvaccinated segments of human populations was, at the time, too small to sustain the virus. Consequently, the virus vanished.

Furthermore, the genetic distance between *Variola major*, the virus that generates smallpox, and the monkey-pox virus, which is sometimes regarded as a possible future reproducer of a new variant of human pox, is still appreciable. But the genetic distance from certain viral strains held in some laboratories over the world is zero. And, at the same time, human herd immunity against smallpox is steadily diminishing, since inoculations were stopped (around 1980). The temporal equation thus formed seems to be formidable. Moreover, the sound, natural equilibrium between the spatial distribution of immunized and nonimmunized segments of human populations worldwide has substantially been impaired as a result of vaccinations, particularly since the elimination of the virus (as a natural selection factor) from the wild. Objectively, favoring for a moment an authentic natural perspective, this seems to be, then, a sort of anomaly. Such circumstances may immensely increase the vulnerability of mankind to the impact of an act of bioterrorism involving the smallpox virus.

If not held in some laboratories until now, it would apparently be, then, a nonrevivable creature. But in actuality, things are different. The story begins in that there are no clear data as to the overall local attempts to isolate—and, hence, preserve—the virus during the occurrences of natural smallpox epidemics over many countries worldwide throughout the twentieth century, until the last 1972 epidemic. Still, even if assuming the total and complete destruction of this virus worldwide, it is apt to be recreated, based on the knowledge attained on its full genome.

Smallpox (also known by the Latin names *Variola* or *Variola vera*) is a highly contagious disease unique to humans. The causative virus has two variants called *Variola major* and *Variola minor. V. major* is the more deadly form, with a typical mortality of 20–40% of those infected. The other type, *V. minor*, only kills 1% of its victims. Many survivors are left blind in one or both eyes from corneal ulcerations, and persistent skin scarring—pockmarks—is nearly universal. Smallpox was responsible for an estimated 300–500 million deaths in the twentieth century. In 1967, for instance, still prevailing worldwide in spite of massive inoculations, the WHO estimated that 15 million people contracted the disease and that 2 million died in that year [245].

Transmission is by droplets, and infection in the natural disease will be via the lungs. The incubation period for obvious disease is around 12 days. In the initial growth phase, the virus seems to move from cell to cell, but around the twelfth day, lysis of many infected cells occurs and the virus will be found in the bloodstream in large numbers. The initial or prodromal symptoms are essentially similar to other viral diseases such as influenza and the common cold—fevers, muscle pain, stomach aches, etc. The digestive tract is commonly involved, leading to vomiting. Most cases will be prostrated.

The smallpox virus preferentially attacks skin cells, and by days 14–15, smallpox infection becomes obvious. The attack on skin cells causes the characteristic pimples associated with the disease. The pimples tend to erupt first in the mouth, then the arms and the hands, and later the rest of the body. At that point, the

pimples, called macules, should still be fairly small. At this stage, the victim is most contagious.

By days 15–16 the condition worsens; at this point, the disease can take two vastly different courses. The first form is of classic ordinary smallpox, in which the pimples grow into pauples, and then fill up with pus (turning them into pustules). Ordinary smallpox generally takes one of two basic courses. In *discrete* ordinary smallpox, the pustules stand out on the skin separately—there is a greater chance of surviving this form. In *confluent* ordinary smallpox, the blisters merge together into sheets that begin to detach the outer layers of skin from the underlying flesh—this form is usually fatal. If a victim of ordinary smallpox survives for the course of the disease, the pustules will deflate in time (the duration is variable), and will start to dry up, usually beginning on day 28. Eventually the pustules will completely dry and start to flake off. Once all of the pustules flake off, the patient is considered cured.

In the other form of smallpox, known as hemorrhagic smallpox, an entirely different set of symptoms starts to develop. The skin does not blister, but it remains smooth. Instead, bleeding will occur under the skin, making the skin look charred and black (this is known as black pox). The eyes will also hemorrhage, making the whites of the eyes turn deep red (and, if the victim lives long enough, eventually black). At the same time, bleeding begins in the organs. Death may occur from bleeding (fatal loss of blood or by other causes such as brain hemorrhage), or from loss of fluid. The entry of other infectious organisms, because the skin and the intestine are no longer a barrier, can also lead to multiorgan failure. This form of smallpox occurs in anywhere from 3% to 25% of fatal cases (depending on the virulence of the smallpox strain).

Edward Jenner developed a smallpox vaccine by using cowpox fluid (hence the name vaccination, from the Latin *vacca*, cow); his first inoculation occurred on May 14, 1796. After independent confirmation, this practice of vaccination against smallpox spread quickly in Europe. The first smallpox vaccination in North America occurred on June 2, 1800. National laws requiring vaccination began appearing as early as 1805 [245]. The cowpox virus is still the immunogenic tool, dismissing, fortunately, the need to use the smallpox virus itself for vaccine preparation, as is the case with many pathogens. Compulsory vaccinations were increased, and the virus has been wiped out globally.

The last case of wild smallpox occurred on September 11, 1977. One last victim was claimed by the disease in the United Kingdom in September 1978, when Janet Parker, a medical photographer in the University of Birmingham Medical School, contracted the disease and died, and the Professor responsible for the unit killed himself. A research project on smallpox was being conducted in the building at the time, although the exact route by which Ms. Parker became infected has never been fully elucidated.

After successful vaccination campaigns, the WHO in 1979 declared the eradication of smallpox, although cultures of the virus were kept by the Centers for Disease Control and Prevention (CDC) in the United States and at the Vector Institute in Koltsovo, Novosibirsk in Siberia, Russia, where a regiment of troops guard it. Under such tight control, smallpox would, it was thought, never be let out again. All other known stocks of smallpox were ostensibly destroyed. Smallpox vaccination was discontinued in most countries in the late 1970s; the risks of vaccination include death (~1 per million), among other serious side effects.

Nonetheless, after the 2001 anthrax attacks took place in the United States, concerns about smallpox have resurfaced as a possible agent for bioterrorism. As a result, there has been increased concern about the availability of vaccine stocks. Moreover, President George W. Bush has ordered all American military personnel to be vaccinated against smallpox and has implemented a voluntary program for vaccinating emergency medical personnel. Vaccination is seriously being reconsidered over many countries. It is also feared that additional stocks of the virus may exist in research collections, the product of the accumulatory nature of microbiologists. Additional collections of the virus almost certainly exist as the result of certain military and biologicol warfare programs, particularly in Russia. Iran, Syria, Israel, China, North Korea, and Cuba have been mentioned as current possessors, as well. The possibility that the virus is currently being held by a terrorist organization has not been excluded.

In light of those circumstances, and considering the extremely undesirable possession of this virus by various countries—hence its expected migration to terror elements, as well—the final and complete destruction of virus stocks held anywhere, and particularly in the CDC and Vector, was ordered in 1993, 1994, 1995, and 1996, but they have not yet been destroyed in those two facilities. Yet the continuation of this actuality has been approved. The World Health Assembly has passed resolutions (WHA52.10) (WHA 55.15) authorizing temporary retention of the existing stocks of variola virus for the purpose of further essential research [246].

The program of research is overseen by the WHO Advisory Committee on Variola Virus Research, composed of members from all WHO regions and advised by some 10 scientific academic experts from such areas as public health, fundamental applied research, and regulatory agencies. Current reports are submitted to the World Health Assembly and are worth detailing due to the uniqueness of the topic at large, as follows.

**Articles summarizing recent research overseen by the WHO Advisory Committee on Variola Virus Research**

- Exploring the potential of variola virus infection of cynomolgus macaques as a model for human smallpox [pdf 564 kb] [247]
- The host response to smallpox: Analysis of the gene expression program in peripheral blood cells in a nonhuman primate model [pdf 897 kb] [248]

**Abstracts summarizing recent research overseen by the WHO Advisory Committee on Variola Virus Research—2004**

- Diagnostic Development at the CDC: Update 2004
- Inferring the Phylogeny of Variola from Genomic SNPs Analysis
- Variola Morphogenesis effected by an erb-B Tyrosine Kinase Inhibitor: Possible therapeutic functions
- The WHO Collaborating Center for Smallpox and other Poxviruses at the Centers for Disease Control and Prevention Atlanta: 2004 report on the variola repository [249]

**Abstracts summarizing recent research overseen by the WHO Advisory Committee on Variola Virus Research—2003**

- Analysis of Nucleotide Sequences of Individual Orthopoxvirus Genes
- Comparative restriction analysis of genomic DNAs of the variola virus strains from the Russian collection
- Human Combinatorial Antibodies against Orthopoxviruses
- Update on variola and orthopoxvirus diagnostic development and use
- Update on search for antivirals against human-pathogenic orthopoxviruses
- Viability estimation of variola virus isolates from the Russian collection [250]

**Abstracts summarizing recent research overseen by the WHO Advisory Committee on Variola Virus Research—2002**

- Capillary-electrophoresis restriction fragement length polymorphism. A new method for poxvirus fingerprinting [251]
- Cidofovir Treatment of Variola (Smallpox) in the Hemorrhagic Smallpox Primate Model and the IV Monkeypox Primate Model [252]
- Genome-wide analysis of the host response to variola infection [253, 254]
- Lethal infection of primates with variola virus as a model for human smallpox [255]
- New PCR assays for identification of smallpox virus [232]
- Variola virus genomics [256]

So far, so reasonable, perhaps. But a dangerous shift might have taken place. American scientists are now awaiting World Health Assembly approval to begin experiments to genetically modify the smallpox virus. Researchers have already been given the go-ahead by a technical committee (WHO Advisory Committee on Variola Virus Research), which accepts the argument that the research could bring new vaccines and treatments for smallpox closer. Still, in the debate during the full assembly of the WHO, whose representatives from 192 member states met from 16 to 25 May 2005, WHO said it will ensure that any research will only be conducted after detailed proposals have been thoroughly examined on a case-by-case basis by the Advisory Committee, paying particular attention to biosafety and biosecurity issues [257].

Some 200 tons of smallpox virus have been produced by the USSR as a weapon and inherited by Russia. Their fate is unclear. However, details of the development of smallpox as a weapon by the Soviets became available. A report was elicited from General Prof. Peter Burgasov, former Chief Sanitary Physician of the Soviet Army and a senior researcher within the BWP. Admitting that development of BW by the Soviets did take place, in the form of live field tests, he described a "smallpox incident" that happened in the 1970s, and was then hashed up: "On Vozrazhdenie Island in the Aral Sea, the strongest recipes of smallpox were tested. Suddenly I was informed that there were mysterious cases of mortalities in Aralsk. A research ship of the Aral fleet came 15 km away from the island (it was forbidden to come

any closer than 40 km). The lab technician of this ship took samples of plankton twice a day from the top deck. The smallpox formulation—400 gr. of which was exploded on the island—'got her' and she became infected. After returning home to Aralsk, she infected several people including children. All of them died. I suspected the reason for this and called the Chief of General Staff of Ministry of Defense and requested to forbid the stop of the Alma-Ata—Moscow train in Aralsk. As a result, the epidemic around the country was prevented. I called Andropov, who at that time was Chief of KGB, and informed him of the exclusive recipe of smallpox obtained on Vozrazhdenie Island" [258].

Actually, it was a new lethal strain of smallpox that traveled far and away from the island-based BW testing facility in the Aral Sea, to infect people downwind on a ship. Most of the adults exposed to the strain contracted smallpox despite being immunized. Promptly, remarkably extensive countermeasures, including vaccination, disinfection, and quarantining were conducted [259].

In March 2003, smallpox scabs were found tucked inside an envelope in a book on Civil War medicine in Santa Fe, New Mexico [260]. The envelope was labeled as containing the scabs and listed the names of the patients that were vaccinated with them. Assuming the contents could be dangerous, the librarian who found them did not open the envelope. This was fortunate, as unlike bacteria (with the exception of those that produce spores), viruses can theoretically survive for many years. The scabs ended up with employees from the National Center for Disease Control, who responded quickly once in-formed of the discovery. The discovery raised concerns that smallpox DNA could be extracted from these and other scabs and used for a bioterrorism attack.

Nevertheless, the chances of successfully doing that seem to be slim, for now. Notably, Ramses V, who lived some 3000 years ago, is the earliest known victim of smallpox, based on an analysis of his mummy [261]. Many virus-like particles were revealed by electron microscopy and identified serologically as smallpox virus in a 400-year-old mummy from Italy. The antigenic structure of the particles was well preserved [262]. The virions in the mummy's skin had lost their viability. Viral antigen was not be detected by EIA or RPHA, and its DNA was not detected by molecular hybridization [263]. Attempts to recover the virus from the frozen bodies of persons who died of smallpox continue in Arctic sites in Siberia [264], and in Canada [265].

The *Variola major* virus is very stable and survives in exudates from patients for many months. The virus is unlikely to survive in dried crusts for more than a year [266]. Yet, it can be preserved in sealed ampules at 4°C for many years, and indefinitely by freeze-drying. This leads to the presumption that the virus still may be preserved alive within bodies of victims buried in permafrost. Much more feasible, however, the practical employment of smallpox virus for bioterrorism purposes is remarkably worrisome and reckoned to be a prime threat.

The current, perpetual phase along the smallpox virus course is facilitated, in effect, by an informal Russian–American status quo, one that has virtually been established lately, acknowledging the nontermination of BWs, for the time being. In a sense, the cardinal, unsolved issue of the smallpox virus constitutes an illustrative reflection of that much broader, not as yet untied, tangle. Sheltered by seemingly acceptable justifications of essential ongoing research into the most dreadful pathogen of mankind is thus still being kept in Russia, the United States, and most

likely, some other places. It perfectly symbolizes the bivalent potential of further, extremely ominous pathogens held, militarily, by Russia and additional states. In Russia, it is retained, probably, as a stockpiled weapon, alongside with an arsenal of various pathogens and toxins. The argument that full acquaintances with pathogens reckoned as BWA—and, likewise, with their aggressive attributes, in terms of bioterrorism threats—are necessary for achieving effective protection, is a very reasonable one, and may hardly be debated. It is therefore, objectively, a salient dual-purpose factor; particularly as engineered pathogens and bioterrorism menaces are becoming common. Concurrently, and perhaps genuinely, this attitude is applied within the U.S. Army as defensively oriented "bioprofiling" [202]. At any rate, the U.S. Army Medical Research Institute of Infectious Diseases, at Fort Detrick, has accumulated "credible evidence" that several terrorist groups and nations have obtained, or are trying to obtain, clandestine stocks of smallpox and are actively trying to produce weaponized armaments based on the virus [267].

Connectedly, an international exercise in dealing with a series of international terrorist attacks involving smallpox virus has shown that the impact would be catastrophic for almost the entire world, and that the lives of millions of people could be lost. This outcome of the "Atlantic Storm" session seems all the more alarming given all the measures taken to combat bioterrorism since the 2001 anthrax letters attacks [268]. The fictitious scenario of "Atlantic Storm" centered on a wave of attacks with the smallpox virus. The exercise was held at a hotel in Washington, D.C., with some 150 observers present to see how 11 former government officials and leaders of inter-governmental organizations, all with the necessary experience, would deal with such a calamity.

## 20 AGROTERRORISM

*The essence and characteristics of agricultural terrorism (agroterrorism) are presented. Reference is here made to infective agents only, meaning pathogens of farm animals and cultivars. Mention is not made here of bioterrorism warfare agents common to man and animals, such as anthrax (those are discussed above). Accordingly, various episodes of agroterrorism that took place since WWII are reveiwed.*

Agroterrorism is the deliberate introduction of pathogens or chemicals (toxic/radioactive), either against livestock or crops, for the purpose of causing economic losses, undermining stability and/or generating fear. The direct impacts are economical, logistical, and demographical (up to hunger). Agroterrorism may be conducted in the form of state-sponsored or nonstate-sponsored acts. The outcomes show in the form of epizootics (animal epidemics) or epiphytotics (plant epidemics).

The United States, the USSR, Germany, Iraq, and Iran. foremost, are known to have had programs that included potential agroterrorism pathogens (in Russia and Iran, at least, this inheritance presently persists), as follows:

**United States**: The viruses causing fowl plague (epizootic avian influenza), rinderpest (cattle plague), hog cholera (swine contagion), and Newcastle disease (poultry contagion), plus late blight of potato fungus, were explored by the United States as candidates for agroterrorism. Weaponized have been the

fungal causative agents of wheat rust and rice blast [269], making the former, explored ones, of no less importance. A wealth of British scientific experience has thereupon been exploited by the US. One salient finding has been in that 3 g of the rice blast fungi per hectare could infect between 50% and 90% of the crops exposed [270]. Other fungal agents evaluated—in that case against drug crops, mainly poppy and coca fields—were *Fusarium oxysporum* and *Pleospora papaveracea*.

**Germany:** Explored: foot and mouth disease, potato beetle, potato stalk rot, potato tuber decay, wheat fungus, turnip weevils and antler moths.

**USSR**: Explored: African swine fever virus, rinderpest virus, and the fungi casing wheat stem rust and rice blast [271]. An uncertain part of those pathogens have been weaponized, plus, apparently, cow pox and sheep pox viruses [272].

**Iraq**: Explored: sheep pox, goat pox, camel pox, and foot-and-mouth disease viruses; wheat smut and cereal rust fungi. Weaponization of the fungus causing weat smut has been evidenced [30].

**Iran**: Rinderpest and foot-and-mouth disease viruses, as well as crop fungal pathogens apparently have been explored and possibly weaponized [273].

Those lists exhibit, basically, the severity of agroterrorism, although principally the spectrum of potential agents is much broader. Thus, mention should be made, in addition, of some further examples. For instance, a 2005 GAO report accentuates the special significance of Asian soybean rust [274]. In an experts meeting "Biosystems and Bioterrorism: Can Arthropods Be Used as Agents of Destruction," the immense importance of arthropods as disease disseminators has been pointed at, with reference being made, among other, to the glassy-winged sharpshooter scare, the Mediterranean fruit fly and the potential for Tsetse and Trypanosomosis reinfestation or establishment in free areas [275].

Illustratively, the devastating New World screwworm fly, regarded to be the most serious insect pest of cattle in the New World, appeared for the first time in the Old World in Libya, during 1988. It aroused, justifiably, great fear, and the countermeasure conducted was an act of biological warfare in itself, namely the sterile male technique: from December 1990 and until October 1991, sterile male flies at the rates of 3.5 building up to 40 million per week were released so as to suppress the fertile male invaders [276]. The release of sterile flies was terminated 6 months after the last detected case of screwworm myiasis in Libya. Notably, Libya accused the United States of intentionally delivering the initial, reproductive pest, although it was the Mexican-American Commission for Eradication of Screwworms that thereafter carried out the eradication campaign.

Peculiarly, an outstanding agroterrorism-related interface evolved between Cuba and the United States. Cuba has blamed the United States for attacking Cuban crops and livestock on as many as 21 different occasions. According to Raymond Zilinskas, of the few incidents for which information is available, agents include Newcastle Disease (1962), African wine sever (1971, 1979–1980), Tobacco Blue Mold Disease (1979–1980), Sugarcane Rust Disease (1978), and Thrips insect infestation (1997). Only in the case of the Thrips did Cuba make a formal complaint. According to the Cuban account, the United States flew a crop-duster

operated by the State Department over Cuba and released the insects. The United States has denied the allegation, and while U.S. agriculture experts discount Cuban claims, the United Nations has undertaken an investigation. Considering the above-mentioned information on the U.S. programs as presented in this section and the Section 15.3, there is support for some Cuban allegations. Yet, according to Zilinskas, the most likely explanation for all these incidents was nature or accidental human transmittal through commerce [5]. Connectedly, a Florida university professor informed the CIA that a Florida citrus canker outbreak was the result of a Cuban bioterrorism weapons program. Although the CIA could not substantiate the claim, it did investigate the case. During that same time period, Cuba claimed that an outbreak of Thrips Palmi disease on the island was bioterrorism warfare introduced by the United States [277]. Apparently, no final conclusion can be reached concerning those issues.

At any rate, the Cuban file was not the only one within the context of agroterrorism affairs. Since 1915, a variety of agroterrorism-related incidents have taken place, worldwide [278], and included, mostly, the following events involving animal and plant pathogens/toxins:

Richard Ford, a prominent British naturalist, has made the accusation that Germany dropped Colorado Potato Beetles on the United Kingdom during WWII, accounting for their unusual appearance in parts of the United Kingdom. According to Ford the bombs were made of cardboard and contained 50 to 100 beetles. The allegation has not been verified or refuted.

In a Ministry of Forestry report dated June 15, 1950, the East German government accused the United States of scattering Colorado Potato beetles over potato crops in May and June of 1950 [279, 280].

The Mau Mau, a nationalist liberation movement, poisoned some 33 steers at a British mission station, using what is believed to be a local toxic plant known as "African milk bush" [281–285].

Ken Alibek, First Deputy Chief of the Russian Biopreparat system, alleged he was informed by a senior Soviet military officer that the Soviet Union attacked the Afghan Mujaheddin with glanders on at least one occasion. According to Alibek, this would have the dual effect of sickening the Mujaheddin and killing their horses, their main mode of transportation [120, 280, 286].

Sometime from 1983 to 1987, a Tamil militant group threatened to use bioterrorism agents against Sinhalese and crops in Sri Lanka. The communiqu' threatened to introduce foreign diseases into the local tea crop and to use Leaf Curl to infect rubber trees [287–290].

Queensland's State Premier received a letter threatening to infect wild pigs with foot-and-mouth disease (FMD), which was feared might spread to cattle and sheep, unless prison reforms were implemented within 12 weeks. Ultimately, this incident proved to be a hoax as the perpetrator turned out to be a 37-year-old murderer serving a life sentence in a local jail. In December 1984, Queensland's premier received a similar letter from an unidentified individual [120, 280, 286].

An outstanding pattern emerged regarding the spread of the Mediterranean fruit fly, a major threat to agriculture in California, in 1989. Despite heroic attempts to eradicate this insect, new infestations repeatedly appeared in odd and unexpected places. Concomitantly, the Mayor of Los Angeles received several letters from a group calling itself The Breeders, which claimed to be spreading the insect to protest California's agricultural practices [287–290].

In 2003, Israel has been blamed for spreading foot-and-mouth disease in the Palestinian Authority territory [291, 292].

It so happened that the majority of antihuman bioterrorism agents are basically and currently pathogens of animals, making them zoonotic. The sense lying in that phenomenon is in that, naturally, those pathogens rarely infect humans; hence human herd immunity against them is marginal, bringing about considerable vulnerability. Some of those zoonotic pathogens are important as both antihuman and anti-animal agents, like Rift Valley fever, epiornithic avian influenza ("fowl-plague"), glanders, anthrax, and brucellosis. They are potentially usable for both agro-and bioterrorism. But another class of pathogens includes specifically livestock attackers. The FMD virus is seemingly their best representative. It is, indeed a formidable agroterrorism agent.

It was again, thus, a viral epidemic—rather an epizootic—this time belonging to the category of livestock attacking agents, which exhibited the aggressiveness of an agroterrorism-patterned course, in that case—the FMD virus. Although very rarely infective toward humans, it is extremely contagious and virulent for mammalian farm animals. It is therefore a prime candidate for agricultural bioterrorism and a very effective one.

Great Britain was the arena, 2001. Despite a single outbreak of FMD in cattle on the Isle of Wight, Hampshire in 1981, there had been no outbreaks of FMD in Great Britain since 1968, until 2001. On 20 February 2001 an outbreak of FMD caused by the O1 Pan Asia strain of virus was confirmed in pigs in an abattoir in Essex. The source of the infection was traced to a pig unit in northeast England where disease was believed to have been introduced at the beginning of February. The provenance of the causative agent could not be identified. Deliberate introduction has not been ruled out. Sheep on a neighboring holding became infected by airborne spread from the pig unit. These sheep were subsequently moved through markets in Northumberland and Cumbria between 13 and 22 February 2001. Disease was then disseminated to many other parts of Great Britain and Northern Ireland as a result of movements through markets and dealers before the existence of disease in the country was recognized. Notably, between 20 February and 30 September 2001, a total of 2025 apparently resultant FMD outbreaks were confirmed in Great Britain; a further 4 were confirmed in Northern Ireland. Over 4 million animals were killed as part of the FMD control program. The epidemic was essentially sheep based, other classes of livestock becoming infected through direct contact with infected sheep or through the movements of people, vehicles, and fomites [293].

The impact of the epidemic was colossal. The overall economic costs in the food and farming sectors of the U.K. economy totaled an estimated £5 billion, roughly $10 billion [9]. This figure compares with an annual gross output of the entire U.K. agricultural sector of £25 billion [10]. The U.K. economy suffered other economic costs associated with the outbreak of FMD. Outside the agricultural sector costs, the United Kingdom endured additional losses to the leisure and tourism sector of the economy. The U.K. Department of Environmental Food and Rural Affairs estimated that the leisure and tourism sector of the economy lost £5–6 billion pounds as a result of the outbreak of FMD in 2001 [11].

Finally, an image of a list of livestock diseases found in a cave in Afghanistan was used as evidence that terrorists are considering attacks on agriculture. It has been accentuated, in connection, that there are many operatives that were planning such acts [180].

## 21 NARCOTERRORISM

*The nature and specific features of narcoterrorism are discussed. Various connotations of narcoterrorism, which are compatible or concerned with biosabotage-related activities, are presented. The main terrorist organizations involved in narcoterrorism worldwide are outlined.*

Primarily, narcoterrorism is a term coined by former President Belaunde Terry of Peru in 1983, when describing terrorist-type attacks against his nation's antinarcotics police. In the original context, narcoterrorism is understood to mean the attempts of narcotics—actually, recreational drugs—traffickers to influence the policies of government, the enforcement of the law, and the administration of justice by the systematic threat or use of violence. Pablo Escobar's ruthless dealings with the Colombian government is probably the best known and best documented example of narcoterrorism [296].

The term has become a subject of controversy, and it is being increasingly used for known terrorist organizations that engage in drug trafficking activity to fund their operations and gain recruits and expertise. Such organizations include, mainly, FARC, ELN, and AUC in Colombia, Hezbollah in Lebanon, and al-Qaeda throughout the Middle East, Europe, and Central Asia. Although al-Qaeda is often said to finance its activities through drug trafficking, the 9/11 Commission Report notes that "While the drug trade was a source of income for the Taliban, it did not serve the same purpose for al Qaeda, and there is no reliable evidence that Bin Laden was involved in or made his money through drug trafficking."

Although financial considerations seem to be the leading ones, narcoterrorism is often marked, substantially, by a constant attempt to sustain and enlarge the proportion of drug consumers in a given target population, hence weakening the latter physically, mentally, and economically. In that sense, it is indeed a sort of terrorism. But beyond the direct effects of narcoterrorism, it is very meaningful, although secondarily, in that it greatly facilitates the ongoing spread of blood-transmitted pathogens, such as AIDS and hepatitis, having their own significant and demographic attritional impact.

Moreover, usually the drugs consumed by narcomans are in fact psychotropic plant toxins: heroin, cocaine, opium, hashish, marijuana, and so on. Narcoterrorism is often based on massive cultivars of the related plants that are currently being overseen by the concerned terror organization, namely self-production of the involved hallucinogenic phytotoxins thereafter consumed by narcomans. In addition, the terrorists engaged in producing those phytotoxins accumulate the know-how and experience needed, basically, for the manufacturing of typical plant toxins, such as ricin, aconitine, and other potent ones. Chemical refinement is often essential. Also, the financial, technical, and logistical infrastructures involved in drug trading may be appreciably supportive of narcoterrorism, bioterrorism, and agroterrorism altogether. Like agroterrorism, narcoterrorism may be regarded, thus, as an additional variant of bioterrorism, although rather of a different sphere.

Most (more than half, certainly) of drug consumption worldwide is fueled by directed narcoterrorism. Some terrorist groups, like Colombia's FARC, collect taxes from people who cultivate or process illicit drugs on lands that it controls; others, including Hezbollah and Colombia's AUC, traffic in drugs themselves.

Moreover, some terrorist groups are supported by states funded by the drug trade; Afghanistan's former Taliban rulers, for instance, earned an estimated $40 million to $50 million per year from taxes related to opium. The drug trade is also a significant part of the economies of Syria—which has funded terrorist organizations such as Hezbollah, the Popular Front for the Liberation of Palestine-General Command, and Palestinian Islamic Jihad—and Lebanon, a haven for numerous terrorist groups including Hezbollah and Hamas [297].

The drug trade is extremely lucrative. Heroin, cocaine, and marijuana are uncomplicated and cheap to produce, but because they are illegal and therefore risky to supply, they can earn more than their weight in gold on the vast international black market. The United Nations estimated that the illicit drug business generates about $400 billion per year. Also, because the drug trade is secretive, terrorists can amass large sums of cash without being detected by authorities.

The terror organizations involved in narcoterrorism at large are mostly the following ones [298]:

- The Revolutionary Armed Forces of Colombia (FARC), a Colombian leftist group, which raises funds by taxing coca farmers in the Switzerland-sized zone of the country it controls. Experts say FARC may force peasant farmers to grow the coca used to make cocaine. It also makes money by protecting cocaine laboratories and clandestine airstrips and by trafficking in drugs locally.
- The National Liberation Army (ELN), another Colombian leftist group, taxes growers of marijuana and opium poppies and protects drug-lab operations. But it generates far less of its funding from drugs than does FARC.
- The United Self-Defense Forces of Colombia (AUC), which includes several right-wing paramilitary groups, says it gets 70% of its income from processing and exporting cocaine. It claims to be leaving the drug business, but experts doubt that all of its members will comply.
- Remnants of Shining Path, a Peruvian leftist group, finance some operations by "protecting" cocaine smugglers in jungle areas under its control and by taxing the coca trade.
- Some members of the Liberation Tigers of Tamil Eelam, a Sri Lankan separatist group, traffic in heroin, and the group reportedly has close ties to drug-trafficking networks in nearby Burma.
- Hezbollah smuggles Latin American cocaine to Europe and the Middle East and has smuggled opiates out of Lebanon's Bekaa Valley, although poppy cultivation there is declining.
- The Kurdistan Workers' Party (PKK), a Marxist separatist group based in Turkey, taxes ethnic Kurdish drug traffickers, and individual PKK cells traffic in heroin.
- The Real IRA, an Irish Republican Army (IRA) splinter group that opposes the peace process in Northern Ireland, is suspected of trafficking drugs, although the extent of its involvement is unclear.
- Basque Fatherland and Liberty (ETA), a separatist group in Spain, is reportedly involved in drug trafficking.

- al-Qaeda does not appear to have direct links to the drug trade. But its former protector in Afghanistan, the Taliban, supported itself in part through opium poppy production and trafficking.

Chaotic countries constitute warm nests for narcoterrorism. Drug traffickers and terrorists tend to thrive in failed states with ineffective governments that have been destabilized by war and internal conflict, experts say. For example, Colombia—a large, fragmented country in the throes of a decades-long conflict over power and resources—produces 80% of the world's cocaine and 70% of the U.S. heroin supply. Lebanon has been plagued by drug traffickers and terrorist groups since its own harrowing 15-year civil war began in 1975. In Afghanistan, the 1990 retreat by occupying Soviet troops left the economically devastated country vulnerable to control by warlords and Islamist extremists. Furthermore, by promoting violence, tax evasion, and lawlessness, terrorists and drug traffickers make it harder for a weakened state to form a stable central government. All in all, the magnitude of narcoterrorism is a remarkable one, and coping with it is difficult and complicated. Narcoterrorism is, in a sense, a long-term attritional mode of terrorism. It is unlikely to diminish, and the main way to effectively struggle against it is expediently to remove the drug producing cultivars themselves. Significant achievements were thus reached by mass employment of fungi that specifically attack those cultivars: cocaine plantations in Columbia and opium plantations in Uzbekistan—a sensible mode of biological warfare by itself.

Paradoxically—yet sensibly, by all means—vast narcoterrorism cultivars were eliminated by agroterrorism-like warfare launched in the last decade, including the effective use of specific fungi capable of destroying poppy fields in Uzbekistan and coca fields in Columbia.

## 22 NOVEL AGENTS

*Novel biological agents constitute a category of enormous interest, in many senses. It is apt to chiefly be pronounced in terms of paradigmatic combat power multipliers on the one hand, and plain bioterrorism parameters, on the other hand. Reference is made, however, to the inability of terrorist organizations to develop or acquire novel bioterrorism agents, unless extraneously assisted.*

A combat power multiplier is defined, primarily, as a warfare element sufficient enough to balance a quantitative inferiority of troops. This concept bears, however, a broader meaning in the context of strategy and military affaires. In these areas, the power multipliers have greater and more complex implications, starting with terror attacks and ending in an overall geopolitical power balance at its highest level, such as international pacts as we saw in the rivalry between NATO and the Warsaw pact.

The future battle field is absorbing, tentatively, a variety of multifunctional technologies. So, the sheer importance of these power multipliers is less in balancing the quantity disadvantage and more as a pragmatic method to decrease the number of the fighting corps (order of battle). Connectedly, a unique range of action is dedicated to biological, chemical, and radiological guerilla warfare tech-

nologies, both terrorism-oriented and military-oriented. This section will discuss these technologies and their ramifications at the macrolevel, within the biological sphere, synthetic biology being, thereby, the chief platform.

## 22.1 Anti-Organism Power Multipliers

Anti-organism power multipliers are based on technologies designed to harm humans, farm animals, or cultivars, whether during offensive, defensive, or calm situations. Although the new generation of these power multipliers is characterized by nonlethal (or sublethal) weaponry, some of them are distinctively lethal.

This new generation include:

- Novel toxins
- Bacteria and viruses in a supremely engineered condition

Toxins are being upgraded as power multipliers. This being the case, through the following biotechnological featuring:

- Molecular design, avoiding or decreasing the body ability to activate an immunological response
- Molecular design that reinforces their toxicity
- Embedding coded genes (regarding protein toxins) in noncontagious bacteria and turning these into "production lines"
- Embedding coded genes in contagious bacteria and viruses and making these even more harmful

The range of toxins designed to perform as power multipliers is vast. Here are two examples, representing both ends of the spectrum: The Pepper spray and the group of agents called "Prions."

The Oleoresin Capsicum (OC) called "the Pepper spray" can be naturally found in hot chilies. The OC has been recently found as more efficient than the common police incapacitating agents, because it causes a faster and longer lasting reaction. The spray can be used for on-spot or wider purposes. Simultaneously, the "Pepper gel" was also developed. This gel is launched from its container by air-pressure; when contact occurs, the gel clings to any surface (if it touches the face, it may cause temporary blindness). The gel formula contains 10% of OC mixed with the gel. The gel is not flammable [299].

Prions, on the other hand, are extremely destructive protein molecules. They are the cause of "Mad Cow Disease" (Creutzfeldt–Jacob Disease). The prions are infectious molecules, which means their penetration to the body brings about normally functioning proteins that turn into abnormal proteins like themselves. The infection-like condition they cause is developing slowly but deadly. They are host-specific. The research of prions, chemically and biologically, is in progressive process and some see them as one of the future measures designed to target a specific civil population for long-term influence. Some prions harm a variety of farm animals [300].

## 22.2 Supremely Engineered Bacteria and Viruses

Genetic engineering is intruding this field, in two main ways:

- The probability of finding gene fragments of highly virulent extinct pathogens is more likely than finding the complete genome and succeeding to revive it. At any rate, an intervention of genetic engineering is necessary to reconstruct and assemble the genome so it can imitate the natural pathogen as a living-reproducing system. In that manner, the polio virus has been recently produced, out of the blue, in laboratories. Similarly, the devastating 1918 "Swine Flu" virus has lately been reconstructed and resurrected.
- The ability of the defender to immunize its population even against the most virulent variants forces the offender to engineer them so immunity will turn out as ineffective.

Genetic engineering is covering far more than those two just mentioned, by creating different and sophisticated power multipliers. The current front in genetic research is promising great achievements for the medical treatments of humans, animals, and plants disease by what is referred to as gene therapy. But, a military aspect of this research also exists. This aspect permits a new form of bioterrorism warfare to take place—genetic warfare. An unavoidable outcome is the discussion on how to use these methods of genetic treatment as a weapon: how to deploy these weapons of genetic warfare and how it is possible to expose the deployment of these weapons [301].

In his futuristic–realistic article, Mark Willis predicts the appearance of the biological genetic-engineered warfare agents according to the following criteria [302]:

- Synthetic viruses
- Cellular pathogens bearing unusual virulence
- Cell-like synthetic entities that cannot reproduce, as vectors of biochemical warfare agents
- Dormant and secret pathogens
- Pathogens specific to farm animals and agricultural vegetation
- Pathogens specific to ethnic groups
- Long-lasting or delayed-effect biological warfare agents
- Ethnic pathogens that can cause autoimmune diseases such as infertility

In practice, at the last decade, we saw the use of specific fungus able to destroy poppy fields in Uzbekistan and coca fields in Columbia. In future perspective, there is no doubt that this is just the tip of the iceberg.

## 22.3 Anti-Inanimate Object Power Multipliers

Anti-inanimate object power multipliers are derived from technologies aimed to harm, disrupt, or jam weapon systems or other valuable assets. They can be implemented during times of offense, defense, and calm. In this context, power multi-

pliers designed to contaminate water or food sources will not be regarded as an anti-inanimate object power multiplier because water and food are used, in that case, to attack humans, serving as vehicles. To our discussion, mention is made, then, of bacteria that can be fed from key materials, naturally or genetically engineered. That includes [303]:

- Plastic, like polyurethane (for example, digesting and removing the special cover enables specific aircrafts to reduce their radar signatures, and by that allowing their detection)
- Rubber, steel, and paint used in strategic and logistic facilities
- Asphalt used in runways
- Crude oil and petrol

## 22.4 Bioengineering Availability

This ethical–technological entanglement has begat a discussion, both locally and globally, concerning increasing or reducing bioengineering availability of the power multipliers being discussed. That, while the freedom of information principle, universal communication interfaces, and the natural curiosity of different scientific circles are taking their place and diverting the reality toward a constant increasing technology availability at large. The fact that a lot of these novel agents are defined as nonlethal (ethical, so to speak) and antiriot countermeasures is contributing to this trend. This trend is also encouraged by the involvement of a university's laboratories in the military R&D process. Those universities, especially in Western countries, mostly get the publication they want.

The latest developments in the fields of political and strategic thinking regarding the use of existing or developing weapon systems (including nonlethal biological weapons) and the ramifications over the operational doctrine in the context of the contemporary trends in the international relations are very significant [304]. This last publication maintains that the alleged revolution of the military-technology affairs has to be considered in light of discussions on military use of nonlethal weapons in existing and future conflicts. The issue is presented from a political-strategic point of view, criticizing the revolution in military affairs as a starting point in the discussion on the use of nonlethal weapons during conflicts. These issues are presented to politicians and strategic analysts asked to formulate a technology-based policy that may be used in future political/social conflicts. The danger of neglecting other important dimensions in politics and strategy, emphasizing the "force to force" aspect, are considered.

Most of the scientific and technological developments mentioned here are openly reported and documented in the United States, which is probably the world leader on these subjects. Russia is, probably, not far behind, but seems to lessen the publishing of this domain. The Russian BWs are very advanced [305], and Russia also puts a lot of effort in biological espionage, trying to get its hands on relevant knowledge and technological means [306].

At large, the global research for upgrading biological power multipliers is striving to develop:

- New measures: both new categories and new means included in already known categories.
- Measures that although they are known, the enemy will not be able to detect, identify, and handle. That may happen due to objective-inherent technological limitations (in such case, then, measures that have to be activated from a far, with no offender presence) or by using a known technological weakness the enemy suffers from.
- Measures that will complicate the enemy's ability—military demogrophical, or economical—significantly and even decisive, although known by the enemy.
- Specific measures designed against a given rival, affecting his military forces, population, or assets uniquely.

The range covered by these new agents is amazing by both its extent and its intensity, starting with incapacitating agents like Pepper spray and ending, hypothetically for now, with ethnical weapons, targeting specific races only. Not in vain, these two examples represent, even if superficially, derivatives of two powerful, inexhaustible technologies—molecular biology and genetic engineering. So, the Pepper spray is becoming a more favorable substitute to the regular tear-gases, being a molecular copy of the biological extract in the hot Mexican chile. So, the biological-ethnic weapons—whose mechanism is using specific genetic elements in the target race or nationality—are conceivably prone to be the unavoidable outcome of highly complex genetic mechanisms, being deciphered by man.

Toxins can be produced in three alternative methods:

- Extracting it from a natural source (the classic mode)
- Chemically synthesizing the biological molecule, after its full and accurate identification
- Producing it by a gene able to code for it (when it is a protein molecule)

The range of scenarios, where the novel agents discussed can be used is vast, in any parameter:

- Activation by state or terror organization, if only by an individual (that can also be sent by state)
- Activation in a concealed or open manner
- Activation in order to achieve an affect in the short, medium, or long range
- Activation during a military conflict or during an international nonmilitary conflict
- Activation against humans, animals, plants, or still objects
- Activation at times of offense, defense, or at calm

Any combination of the parameters or the subparameters is possible.

Today and in the near future it is not expected that the availability of sophisticated biologically engineered power multipliers will proliferate. It will probably be an asset of few supremely qualified laboratories. However, with time they will

become more and more accessible, presumably. With no connection to the BW agents that the Iraqi government has produced in its time (undoubtedly, industrial amounts), it is appropriate to mention two agents reflecting an unusual way of thinking suiting aspects aforementioned, although on an embryonic level [30]:

- Camel poxvirus—a pathogen that can attack humans, but not those who live next to camels and so are naturally immune. The Iraqis have also used this virus as a model to the small pox virus.
- The fungal toxin aflatoxin, which generates terrible diseases, but only on the long term (was apparently destined to be used against the Kurds in Iraq).

In Russia, or elsewhere, chimeras of VEE-smallpox virus, Ebola-smallpox virus, and snake-toxin-gene-bearing influenza virus have possibly been developed successfully, but this has not been verified or refuted.

Furthermore, it was recently argued that advances in nanotechnology could be used in the development of new types of biological (and chemical) weapons, based on the dual-use application of nanotechnology in the fields of biotechnology and medicine, which can be used for offensive purposes. Such a potential military application of nanotechnolgoy could undermine existing international laws that ban biological weapons [307].

To sum, the technological absorbability (of a country) is indeed the cardinal factor, but it is not sufficient for these frontline technologies to be adopted (even more so—for their upgrading) by one. But mention is made of the option that hostile countries or organizations can get their hands on standardized, instantly usable, weapons. The discussed novel agents are very sophisticated, and so transferring, maintaining, and operating them may be very complicated. However, it is possible that hostile elements will get extraneous assistance at these aspects as well. Particularly, a great deal of these weapons includes, so-called, nonlethal and anti-inanimate agents, hence more obtainable.

Referring to the increasing threat of novel bioagents, Steven Block, former president of the Biophysical Society, was quite pessimistic: "The biological weapons threat is multiplying and will do so regardless of the countermeasures we try to take. You can't stop it, any more than you can stop the progress of mankind. You just have to hope that your collective brainpower can muster more resources than your adversaries [308]."

## 23 EXPLOITABILITY OF BIOTECHNOLOGY AND BIOMEDICINE

*The vast space inevitably formed by biotechnology and biomedicine, in terms of dual usability, is briefly presented. Its affinity to bioterrorism-oriented activities is shown, with regard to both pathogens and toxins. Two prominent toxins—botulinum and ricin—are thereby visited, to accentuate the scarcely observable borderline lying in between legitimacy and illegitimacy.*

Apparently, the dual usability marking the fields of biotechnology and biomedicine, namely their capacity to serve for desirable and undesirable purposes at the same time, is the most extreme one, comparing with almost any other scientific

domain. A relatively simple instance of this: the undisputable need to thoroughly understand the mechanisms underlying the above-described alternating profile of drug resistance of *Vibrio cholerae* potentially enables, in parallel, the creation of a multistable pathogen. The shift is fairly slight, if any.

Fundamentally, all categories of biomedicine targeted at infectious diseases—antibacterial and antiviral drugs, vaccines, and antisera—as well as gene therapy are apt to be exploited for improper development, parallel to their impressive, invaluable advancements. Moreover, regretfully, a principle of "amplified reversibility" underlies the undesired, yet inevitable, usability of biomedicine and bioengineering resources for the purpose of bioterrorism. That is to say, although most pathogens and toxins have found a way to become harmless and beneficial, thanks to the scientific comprehension gained from their usefulness in various respects, that very paved course may readily be reciprocal; in particular their road back to pathogenicity and virulence—often appreciably amplified, artificially—may take place in the form of accidental leakage or bioterrorism acts.

In pragmatic terms, two dimensions of exploitability can be observed: One is the general modularity of equipment and materials that can equally be employed for innocent or bioterrorism-oriented activities. The other one is the various untainted activities currently engaging specific pathogens and toxins that are reckoned, at the same time, as potential agents for bioterrorism; hence, they are prone to undergo, in principle, a turnover at the level of intentions and practice of their possessors. Thus, the harvest of a seed virus propagated for subsequent production of a vaccine by attenuation may equally be kept as is, before attenuation takes place, serving as a weapon. In actuality, the two courses are not in contrast with each other; hence, they may expediently be applied in parallel, in whatever proportions the possessor chooses, so as to attain effective camouflage.

The principle is fairly plain, even an old one. In 1939, the active ingredient of curare—an ancient plant toxin weapon still in use by Indians—was isolated for the first time. In 1943, it was introduced successfully into anesthesiology. Curare provided adequate muscle relaxation without the depressant effect of deep anesthesia induced by ether or chloroform. Over the last 20 years, physicians have used curare to ease the stiffened muscles caused by polio and to treat such diverse conditions as lockjaw, epilepsy, and cholea (a nervous disorder characterized by uncontrollable muscle movements). Eventually, more effective treatments were found for these illnesses, but the active ingredient of curare, d-tubocurarine, led to the skeletal muscle relaxant Intocostrin, which has been used in surgery ever since. Synthetic analogs of d-tubocurarine are used tens of thousands of times per day in the operating room [309].

Tubocurarines reflect, yet, but one, relatively marginal example of the dual exploitability of biotechnology and biomedicine. Ricin toxin and botulinum toxin are more actual agents, as shown. Global castor seed production for civilian purposes is around 1 million tons per year. The beans are widely used for the production of castor oil, making their control impossible, practically. The castor oil manufacturer may readily use the remains of the beans for obtaining ricin. Moreover, ricin in itself is typically a dual-use substance.

Because of its cytotoxic potency, modified ricin is being used for the selective killing of unwanted cells and for the toxigenic ablation of cell lineages in transgenic organisms. Ricin is the most commonly used toxin in conjugates for selective cell

killing, so-called targeted toxins. Beneficial uses of ricin include, potentially the treatment of cancer and AIDS. Recently, there is, indeed, increasing interest in the medicinal applications of ricin as immunotoxin, which have been successfully applied in several human diseases. In addition, the mole-cule of ricin in itself is the precursor for the preparation of toxoid and anti-serum.

Ricin can be targeted to specific cells, such as cancer cells, by conjugating the RTA subunit to antibodies or growth factors that preferentially bind the unwanted cells. These immunotoxins have worked very well for *in vitro* applications, e.g., bone marrow transplants. Although they have not worked very well in many *in vivo* situations, progress in this area of research shows promise for the future. In bone marrow transplant procedures, RTA-immunotoxins have been used successfully to destroy T lymphocytes in bone marrow taken from histocompatible donors. This reduces rejection of the donor bone marrow, a problem called "graft-vs-host disease" (GVHD). In steroid-resistant, acute GVDH situations, RTA-immunotoxins helped alleviate the condition. Also, in autologous bone marrow transplantation, a sample of the patients own bone marrow is treated with anti-T-cell immunotoxins to destroy malignant T cells in T-cell leukemias and lymphomas [310]:

For the *in vivo* treatment of solid tumors, considerable problems can arise due to poor access of the immunotoxin to the tumor mass, lack of immunotoxin specificity, and further disadvantages. Still, research efforts to expand and develop immunotoxins and therapies for clinical use in cancer and AIDS are continuing with strategies using recombinant DNA technology [311]. On the whole, the following examples of recent studies may be noted: Oncologic applications: Immunotoxin treatment of brain tumors; cytotoxins directed at Interleukin-4 receptors as therapy for human brain tumors; bispecific monoclonal antibodies for the targeting of type I ribosome-inactivating proteins against hematological malignancies; tyrosine kinase inhibitors against EGF receptor-positive malignancies; targeting tumor vasculature using VEGF-toxin conjugates; and gene therapy with immunotoxins. Alternative applications: effects of selective immunotoxic lesions on learning and memory; targeting toxins to neural antigens and receptors; *in vivo* testing of anti-HIV immunotoxins. All in all promising, apparently, but yet legitimating the access to purified ricin.

Botulinum toxin constitutes another notable case. The most poisonous molecule ever identified, compared with any other natural or synthetic molecule, could not gain, apparently, a title better then "A Bug with Beauty and Weapon [312]."

The germ *Clostridium botulinum*, a gram-positive, anaerobic spore-forming bacterium, is distinguished by its significant clinical applications as well as its potential to be used as a producer of a singular bioterror agent. Growing cells secrete botulinum neurotoxin, a multitype protein bearing exceptional toxicity. Although botulinum toxin is the causative agent of deadly neuroparalytic botulism, it also permits a remarkably effective treatment for involuntary muscle disorders such as torticollus, dystonia blepharospasm, strabismus, hemifacial spasm, certain types of spasticity in children, and other ailments. It is also used for "off label" indications such as migraine headaches. Furthermore, this extraordinarily potent toxin is also used in cosmetology for the treatment of glabellar lines and is well-known as the active component of the anti-aging medications Botox and Dysport. In addition, recent reports show that botulinum neurotoxin can be used as a tool

for pharmaceutical drug delivery. However, botulinum toxin remains the deadliest of all toxins and is a potent agent of bioterrorism. Among seven serotypes, *C. botulinum* type A is responsible for the highest mortality rate in botulism, and thus it has the greatest potential to act as a bioterrorism weapon. Genome sequencing of *C. botulinum* type A Hall strain (ATCC 3502) has been completed. It may readily serve to dually engineer the related molecule in terms of toxicity, antigenicity, and stability.

The availability of medicinal Botox is not likely to contribute to a terrorist attack as the vials contain dilute toxin: approximately 0.3% of the lethal inhalational dose estimate and 0.005% of the lethal oral dose estimate. Yet if available right before being diluted, it would certainly be an attractive object. Obviously, the toxin is needed for producing a toxoid and an antiserum.

Finally, the next two examples may tentatively illustrate an entire scope of the exploitability of biomedicine and biotechnology through its two edges, namely the rudimentary one as against the sophisticated one. The former can be demonstrated, then, by crude fermentation of various fungi for the manufacturing of antibiotics or mycotoxins (a dual-use biotechnology largely adopted by the then USSR), whereas the latter one is the usage—for now experimental—recently made of viruses as drug delivery systems, owing to their unique capacity to inwardly parasitize host cells. This biomedical horizon appears to be fascinating, and a far reaching one, in both a desirable and an undesirable fashion.

## 24 PREVENTIVE AND COUNTERMEASURE

*Having been recognized as a prime threat with global potentiality, different approaches and schools configured to fight bioterrorism are here discussed, rather exhibiting the absence of adequate preparedness. Moreover, various factors, some inherent, may hinder the effectual establishment of future preparedness and are considered. Much progress has though been achieved and is reviewed, particularly in the United States. An endeavor is taking place in Europe and Russia. All over, the efforts are doctrinal, scientific, and logistical, both preventive and reactive.*

Aiming to defy any sort of assistance lent by (or from) states to terrorists within the entire context of biological, chemical, or nuclear weapons, the UN Security Council formulated in 2004 a resolution saying that "... All States shall refrain from providing any form of support to non-state actors that attempt to develop, acquire, manufacture, possess, transport, transfer or use nuclear, chemical or biological weapons and their means of delivery. All States, in accordance with their national procedures, shall adopt and enforce appropriate effective laws which prohibit any non-state actor to manufacture, acquire, possess, develop, transport, transfer or use nuclear, chemical or biological weapons and their means of delivery, in particular for terrorist purposes, as well as attempts to engage in any of the foregoing activities, participate in them as an accomplice, assist or finance them ..." [313].

Globally, apart from many national frameworks, various international initiatives have been founded, aimed at the prevention of, and coping with, possible bioterrorism scenarios. In 2006, the U.N. General Assembly released its new counterterrorism strategy. It recommends development of a "biological incidents" database

by the United Nations and its member countries, in order to fight the threat of bioterrorism [314].

In Europe, much progress was achieved during 2004–2006, owing to the New Dedence Agenda (NDA) (later on renamed as Security and Defence Angenda—SDA)—a regular professional discussion forum involving NATO, the EU, and the World Health Organization plus national ministries, industry figures, and journalists, to debate safety and defense issues. It aims to raise awareness in Europe of the bioterrorism threat and to define a set of recommendations for EU policy makers to prevent and protect against attacks.

The 2004 NDA conference outlined that security awareness at epidemiology research laboratories across the industrialized world is lagging behind the growing threat of bioterrorism. Tighter cooperation is needed among laboratories, international customs, and transport authorities, and NATO could provide important logistical and communications support in the event of a bioterror attack in Europe. Notably, Roger Roffey, research director of the Swedish Defense Ministry's Department of International and Security Affairs, accentuated that "The risk of bioterrorism is on the rise, and one of the deficiencies we all face concerns the control of bioterrorism agents within laboratories and their transfer beyond, especially at labs across the former Soviet Union. We should expect bioterrorists to go beyond an attack by aerosol delivery into product-tampering of crops and animals, into our very food and water supplies. We need new multilateral initiatives to improve biosecurity facilities [315]."

The 2005 NDA conference was an example of excellent EU–U.S. collaboration in the context of bioterrorism. If there was one conclusion that could be drawn from the conference, it was that such teamwork had to be duplicated in the actual fight against bioterrorism. Not that the event lacked ideas, as these were ever present. But there were few signs of real cooperation, and the specter of different "threat perceptions" was forever hovering in the background [316]. This line strengthened in the 2006 SDA conference.

The Interpol and the European Homeland Security Association pay much attention to bioterrorism threats. Interpol President Jackie Selebi recently observed that "Major panic, temporary paralysis of government functions and private businesses and even civil disorder are all likely outcomes of a bioterrorism attack. In fact, bioterrorism appears particularly well suited to the small, well-informed groups. A bioterrorist's lab could well be the size of a household kitchen and the weapon built there could be smaller than a toaster, and the range of options available to terrorists will continue to grow." Furthermore, he warned of bioterror attacks on livestock or the food chain [317].

Interpol Secretary General Ronald Noble added that "There is no criminal threat with greater potential danger to all countries, regions and people in the world than the threat of bioterrorism. There is no crime area where police have as little training than in preventing—or responding to—bioterrorist attacks. Terrorists do want to use biological weapons. The threat is worthy of immediate preparation" [318].

French Interior Minister Dominique de Villepin called for creation of a UN-affiliated organization to track potential biowarfare agents and keep them away from terrorists. He did not propose giving inspection powers to such an agency, but said that biotechnology companies, laboratories, hospitals, and universities need to

better monitor themselves on issues of hiring, pathogen work, and access to sensitive areas [317]. Connectedly, in Britain, MI5 officials recently warned British laboratories that Islamist terrorists—chiefly al-Qaeda, through recruited university students—may try to steal deadly pathogens. Scientists and lab staff handing biological agents such as samples of avian flu, tuberculosis, rabies and polio, have been told their security measures will be vetted by police [319]. In consequence, police is supposed to conduct background checks on scientists and others who work with restricted materials at universities, hospitals and pharmaceutical firms [320]. Government officials are intended to inspect such laboratories, and regular audits are planned of agent inventories. The idea is that it will be up to scientists to prove they have good reason for using the materials in their work, according to a senior of the British National Institute for Biological Standards and Controls.

During 2006, cooperation between prominent European states and Russia considerably strengthened. Thus, in Kyiv, Russia, the highly qualified Interpol Workshop on Preventing Bioterrorism took place. Also, a fairly thorough "Strategic Study on Bioterrorism" produced by 20 high level bio-experts from the Russian Federation and other European countries, was released by the Center for Strategic and International Studies.

In Israel, much concern was raised that regular microbiological laboratories—particularly those located in the Palestinian-occupied territories—may be used as minifactories for producing bioterrorism agents. It has therefore been decided that the National Security Council—in coordination with the General Security Service, Ministry of Health, academic institutions, and industrial bodies—will oversee such laboratories [84]. The Israeli Parliament building and facilities are expected to be reviewed for potential vulnerabilities to biological (and chemical) terrorism. Air conditioning and other ducts are expected to be examined to determine any entry point for airborne agents [321].

In some Muslim states, professional forums intended to fight bioterrorism have been formed. In Iran, parallel to its ongoing, progressing BW program—which includes, among other things, guerilla means for applying bioterrorism warfare—The First Conference on [the] Campaign Against Bioterrorism was held in Tehran. The conference, which was a joint project of the Iranian Red Crescent Society (IRCS) and the infectious diseases department of the Tehran Medical Science University examined the health risks posed by bioterrorism and attempts to develop an action plan for use by relief workers in the event of a bioterror event. Over 300 IRCS relief instructors attended the event along with IRCS managers and students from the University [322].

In Saudi Arabia, where there are still elements in support of al-Qaeda, the Center of Studies and Research at Naif Arab University for Security Sciences in Saudi Arabia held a seminar on bioterrorism. Experts participating in the seminar were from Saudi Arabia, Jordan, Bahrain, Comoros, Sudan, Syria, Palestine, Qatar, Iraq, Kuwait, Lebanon, and Egypt [323].

Beyond comparison, however, is the effort against bioterrorism made by the United States. The preventive and countermeasures taken by the United States against bioterrorism are indeed the most impressive ones, in terms of concept, thoroughness, and scope. They are nourished by fundamentals that preceded the 2001 anthrax attack, which means, dually, that on the one hand the need for preemptive preparedness was soberly recognized in advance, and on the other hand,

that the saboteurs succeeded to carry out their attack plan despite the defensive preparedness already established, then, by the United States.

In principle, a variety of ways in which the U.S. agencies were implementing President George W. Bush's April 2004 Biodefense for the 21st Century initiative, organized around the four "pillars" of awareness, prevention, detection, and response [324]. Project Bioshield Presidential law reflects the genuine national American incentive to get prepared toward the bioterrorism threat. In this fashion, lately (2004), the President of the United States signed a new law to implement a $7.8-billion (estimated) project aimed to develop advanced vaccines. This law permits the government to use, in case of national emergency, medications and treatment not yet approved by the FDA. President Bush said this step will help the country to be better prepared in case of a terror attack. Also, said Bush: "It sends a message about our direction in the war on terror. We refuse to remain idle while modern technology might be turned against us." The President also added: "We will rally the great promise of American science and innovation to confront the greatest danger of our time [325]."

The U.S. Homeland Security Subcommittee on Prevention of Nuclear and Bioterrorism Attack handles the bioterrorism threat. It pointed out that the United States must seek the proper balance between agility of response and countermeasure stockpiling in defending against bioterrorism. Moreover, the United States is pursuing multiple avenues of defense against a possible terrorist attack using bioengineered pathogens [326]. Top leaders such as the Homeland Security Secretary have stepped up calls for greater use of terrorist threat information in setting a hierarchy of planning and spending priorities. Witnesses warned the Government Reform National Security, Emerging Threats and International Relations Subcommittee, though, about limits to what can be known about potential engineered threats. They stressed the need for maintaining a broad, flexible array of countermeasures [327]. Homeland Security and the Health and Human Services Department have developed a strategy to address the potential for a bioengineered attack, a document that highlights monitoring of scientific research around the world, as well as broad countermeasure development to give the United States flexibility when confronting an unknown pathogen. Complementarily, two annexes of the new U.S. National Response Plan are particularly relevant to bioterrorism: Emergency Support Function 8 (Public Health and Medical Services) and the Biological Incident Annex. These annexes describe specialized application of the NRP to the delivery of public health and medical services and biological incidents [328].

Within that colossal program, the Bush administration's fiscal 2006 budget plans for civilian bioterrorism defense measures total at least $5.1 billion, according to a new nongovernmental analysis. The amount brings the total requested since 2001 to at least $27.7 billion. The budget request includes increases for protecting national food and water supplies and a significant decrease for funding state and local public health departments, compared with requested spending last year [329].

Further budgetary and organizational details are worth mentioning, then, as presented in that publication [330]. The fiscal 2006 budgeting, an aggregate of spending across numerous agencies, reflects as much as a $2.5 billion drop from what was sought for fiscal 2005, but the decrease results mostly from the absence of a one-time appropriation last year for drug and vaccine purchases through 2008 as part of the Project Bioshield law. "Civilian biodefense spending, not including

the Bioshield bill, has reached a consistent level of about $5 billion from fiscal 2003 to fiscal 2006," the article says. The fiscal 2006 total in the study does not include Defense Department budgeting for civilian bioterrorism defenses, as do previous years' figures, because the Pentagon "was unable to furnish numbers for the requisite programs," the study says.

Military bioterrorism defense spending for civilians in the previous two fiscal years averaged about $200 million. The analysis says, though, that those figures do not truly account for all Pentagon funding for civilian bioterrorism defense measures. "Some DOD research has direct civilian benefit, but because the majority of these funds are primarily military in application, these lines were excluded from calculation of total DOD expenditures," it says.

As in previous years, the largest amount of money—$4.1 billion according to the article—was budgeted for the Health and Human Services Department, which funds the National Institutes of Health, the Food and Drug Administration, and the Centers for Disease Control and Prevention. The article describes a planned $130 million cut to CDC funding for state and local public health departments, bringing the total CDC budget request down to $797 million. Another substantial cut is a $119 million reduction from the National Institute of Allergy and Infectious Diseases' budget for research facility construction, down to $30 million, and intended to "offset the increase in research funding," the analysis says. The Homeland Security Department's $362 million budget is roughly the same as for fiscal 2005—except for the $2.5 billion drop reflecting the advanced purchase for fiscal 2005 under the Bioshield law.

The Agriculture Department was budgeted $354 million, a 26% increase, for bioterrorism defense activities, the State Department $71.8 million, and the National Science Foundation $31.3 million, particularly, in accordance with Homeland Security Presidential Directive 9 (HSPD-9), Defense of United States agriculture and food, January 30, 2004. Finally, the Environmental Protection Agency received an 87% increase to $185 million, primarily for decontamination capabilities, to protect water and food supplies, and for training, the study says.

Apart from doctrinal and laboratory studies conducted by those agencies, field studies are being carried out as well. Scientists from several agencies released two tracer gases in a 2-$km^2$ area of midtown Manhattan. In a series of experiments, they are tracking how harmful particles might disperse through street canyons, subway tunnels, and buildings. Dubbed the "Urban Dispersion Project," this Department of Homeland Security-sponsored effort aims to produce a computer model of airflow patterns that could help state and local officials better respond to an emergency or to a bioterrorism act [331]. The studies are intended to conclude in 2007.

Overall, a possible counterproductive outcome has been assessed, naturally, as some scientists are concerned that the increase in the number of researchers working in the United States to counter bioterrorism elevates, potentially, the risk of an attack and the accidental release of biological agents [332]. More than 300 institutions and 12,000 individuals have access to weaponizable biological agents, said Richard Ebright, a Rutgers University molecular biologist and critic of the expansion in biodefense since the anthrax attacks of 2001. Biodefense watchdog, the Sunshine Project, claims that 97% of principal investigators who received grants from the U.S. National Institute for Allergy and Infectious Diseases from

2001 to 2005 to study six biological agents had not previously conducted similar work. The explosion of "NIAID newbies" increases the likelihood of accidents, said Edward Hammond, U.S. director for the Sunshine Project. Jeanne Guillemin, a senior fellow at the Security Studies Program at the Massachusetts Institute of Technology, has the view that increasing access to pathogens heightens the chances that rogue scientists could use them in an attack. "What [NIAID Director Anthony Fauci] and others haven't thought through is the particular kind of expertise, from basic bench work to aerosolization, that comes with defensive biological weapons programs" [332].

Furthermore, the extreme, objective complexity of implementing bioterrorism-defense measures aroused, unavoidably, critics saying that despite substantial funding on bioterrorism defenses since the Sept. 11 terror attacks, the country remains substantially unprepared for a mass-casualty bioterrorism attack [330]. A variety of concrete steps and moves has been conducted, however, as follows.

Nonetheless, a large bioterrorism research network has been established in the United States. The U.S. National Institute of Allergy and Infectious Diseases supports 11 academic institutions throughout the United States, which operate as Regional Centers of Excellence for Biodefense and Emerging Infectious Diseases Research. The new network is working diligently to uncover new knowledge and create preventive, therapeutic, and diagnostic tools that will leave us far less vulnerable to bioterrorism, and is regarded as a key element of a U.S. strategic plan to counter bioterrorism and emerging infectious diseases. Each Center for Excellence leads a group of local universities to conduct research on next-generation treatments for anthrax, smallpox, plague, and other diseases. The consortiums encourage bioterrorism research, train personnel, maintain support resources, push for research on the development of new countermeasures, open facilities to researchers from academia and the business world, and provide support for first responders [333]. Within that framework, an 83,154-square foot Biosafety Level 3 laboratory is to be constructed as part of George Mason University National Center for Biodefense and Infectious Diseases. Personnel at the facility will conduct research on development of techniques and products to combat bioterrorism and treat natural infectious disease outbreaks. Researchers will focus on bioterror threats identified by the U.S. government, such as anthrax, tularemia, and plague, along with diseases such as SARS, West Nile virus, and influenza [334]. In addition, Army Medical Research and Materiel Commander Eric Schoomaker focused on the coming benefits of interagency cooperation at the planned National Interagency Biodefense Campus at Fort Detrick, MD. The Defense, Health and Human Services and Homeland Security departments are participating in the campus project.

At the same time, the multiplicity of pathogen holding research facilities is disadvantageous in that they are vulnerable to technical intelligence assaults, thefts and sabotage acts which may serve bioterrorism. This applies, in principle, to a variety of biotechnological resources at large, and thereby propelled, for example, the Congressional Seminar on Preventing Terrorist Exploitation of the Biotechnology Revolution, (June 5, 2006). Connectedly, perhaps, an FBI secrecy initiative formed with key bio-facilities such as the Public Health Research Institute (PHRI) in Newark, New Jersey, so as to refrain from "publicly disclose which specific select agent pathogens and/or strains are stored at PHRI [335]."

In practical terms, early detection systems of various pathogens gained high priority. BioWatch—a Department of Homeland Security's effort to collect air samples form dozens of cities around the country to detect biological agents—biosensors have been deployed in more than 30 urban areas (at a cost of $79 million for 2006). Air is thus monitored air 24 hours a day and samples are collected daily and taken to labs that are part of the Centers for Disease Control and Prevention Laboratory Response Network. The results are provided within 12 to 36 hours [336].

Connectedly, the U.S. Postal Service has installed more than 1000 biological agents—anthrax plus two undisclosed agents—detectors at 271 mail processing facilities since the 2001 anthrax mailings. Moreover, President Bush has staked out the right of the U.S. government to open citizens' mail without first obtaining a warrant [337].

A further methodology, mass spectrometry based "finger print" data base, has been developed by the Food and Drug Administration (FDA) National Center of Toxicological Research (NCTR), for immediate differentiation between biological contaminants, toxins and other, nontoxic substances. Although the concerned device cannot distinguish between living and dead cells, it is highly advantageous in terms of instant findings [338].

Also, nanotechnology could significantly increase the speed of detection of an act of bioterrorism or a natural outbreak. Using a diagnostic test, researchers can measure the frequency change of a nearinfrared laser while it scatters a virus' DNA or RNA. The change in frequency is distinct [339].

Beyond, the most advanced biodetectors are the Triangulation Identification for Genetic Evaluation of Risks—the first multi-purpose diagnostic system that can work with samples of various materials. It can simultaneously identify all pathogens in a soil, water, air or blood samples within four hours, and roughly five such devices are now being used in the US by the US Army and Agriculture Department [340]. In connection, a comprehensive national surveillance network—Project Tripwire—aimed at detecting diseases in wildlife that may be linked to bioterrorism, is being developed by the Wildlife Center of Virginia [341].

Coping with specific pathogens is complicated, naturally, though the very fact that a certain pathogen is being referred to removes a lot of vagueness. Illustratively, the optional adoption of a pre-attack line of defense, as against a post-attack one, is indeed an intriguing one. Researchers from United States and Canada said post-attack anthrax immunization and antibiotic therapy is more effective than pre-attack vaccination [342]. Still, this is doubtfully the correct strategy concerning smallpox, for instance, particularly considering the irrelevance of antibiotic drugs.

Currently, cities and counties throughout the United States are creating stockpiles of anthrax countermeasures that mirror the drugs available in the Strategic National Stockpile [343]. The distributed stockpiles are necessary because it would take too long to distribute countermeasures from the National Stockpile. The drugs, given to cities to prepare for a terrorist attack, are meant to immediately treat first responders in the event of an anthrax attack without having to wait for drugs from the National Stockpile.

The 2001 anthrax letters sabotage was at any rate a cardinal turning point. Using those attacks as a case study, a group of U.S. researchers have developed a risk-

based decision-making system that they believe will allow governments to make better judgments following a bioterrorism attack [344]. In "Bayes, Bugs and Bioterrorists: Lesson Learned from the Anthrax Attacks," Kimberly Thompson, Robert Armstrong, and Donald Thompson argue that governments must develop "decision trees," or methods for evaluating several courses of action, following a bioterrorism attack. They claim this would improve the evaluation of costs, risks, and benefits and would create more effective policy development [344].

"Using this type of approach, the government can better characterize the costs, risks and benefits of different policy options and ensure the integration of policy development," the report states. "Additionally, confirmed use and refinement of decision trees during exercises will provide analysis of the long-term consequences of decisions made during an event and give policymakers insights to improve initial decisions." The study, dated April 2005 and published by the National Defense University's Center for Technology and National Security Policy, says poor planning prior to the fall 2001 bioterrorism attacks in which anthrax was sent to U.S. Senate office buildings through the mail led to improper allocation of resources after the attack [345]. The report estimates the direct costs to the Postal Service could exceed $3 billion, with additional expenditures of over $1 billion for unnecessary countermeasures.

Despite the attack, coordination between U.S. agencies has made development of a comprehensive response plan difficult. To remedy the problem, the report urges establishing a decision tree so that multiple paths of action can be evaluated at once, making individual agency's responsibility more clear. "With this approach, analysts can quickly communicate with decision makers about the implications of combinations of options," the report says. "We emphasize that this approach of focusing on decisions provides a means to cross interdisciplinary and other boundaries . . . and consequently it provides a useful organization and communication tool to promote effective management." The report adds that decisions should be separated into different categories, including who should be immunized, what response should be, how to allocate Strategic National Stockpile resources, how to contain biological agents, and how much information should be made public. The Health and Human Services Department should head the effort to form the decision tree, drawing on the expertise of other agencies when necessary.

The 2001 anthrax attacks clearly demonstrated the need for a better decision-making process, the report says. Lack of investment in the public health infrastructure, a poor understanding of the disease caused by anthrax, inadequate training of first responders, and poor communication are just a few of the problems that plagued the response to the attack. The report argues that a decision tree would have vastly improved response, allowing for a better understanding of who should be vaccinated, how relevant agencies should have responded, a better plan for containment, and improved information management. The report is careful to say that it is not critical of the 2001 response. However, researchers believe a decision tree would have prevented panic and yielded a more appropriate response. Researchers hope decision trees can guide policy discussions as a comprehensive response plan is formulated. "Given our recent experience with anthrax, the specific decision trees for anthrax are offered, as an analytical tool to aid future policy decisions," the report says. "Indeed, the lessons learned from the 2001 attack should facilitate the use of these trees."

Plague has been visited as well, when a wide-scale drill took place in the United States in November 2003. The drill was an effort to follow up on weaknesses in federal emergency response plans identified in a simulated pneumonic plague bioterrorism attack. That exercise, called Top Off 2, was organized by the Department of Homeland Security and involved 8000 local, state, and federal officials. It simulated a pneumonic plague attack on Chicago (and a radiological attack on Seattle) [219].

As for smallpox, the United States intends to acquire up to 80 million doses of the modified virus Ankara smallpox vaccine, as part of its defense against the threat of smallpox virus being used for bioterrorism purposes. It would be useful for people who could suffer reactions to the existing vaccine. That includes patients with suppressed immune systems or those with the skin condition eczema. The vaccine contract could be worth more than $1 billion [346]. President George W. Bush has ordered all American military personnel to be vaccinated against smallpox and has implemented a voluntary program for vaccinating emergency medical personnel. Stockpiling and vaccination is seriously being reconsidered over many countries. Japan is a recent example; after announcing in 2001 that it would begin stockpiling enough vaccine for 3 million people, a governmental panel of experts recommended—in July 2005—the stockpiling of smallpox vaccine for 56 million people in preparation for a possible bioterrorism attack [347]. Russia, with its efficient oral—plus aerosol—vaccines, does not lag far behind. Lately, an international exercise—the "Atlantic Storm Scenario" in dealing with a series of international terrorist attacks involving smallpox virus has shown that the impact would be catastrophic for almost the entire world, and that the lives of millions of people could be lost. This outcome of the "Atlantic Storm" session seems all the more alarming given all the measures taken to combat bioterrorism since 11 September 2001 [268]. The Atlantic Storm Scenario has been preceded by "Exercise Global Mercury," which was the first worldwide simulation project modeling a smallpox bioterrorism attack [348].

Considering influenza virus to be the most powerful potential bioterrorism weapon, Madjid et al. proposed several steps to address the threat of influenza as a terrorist weapon, including the following [349]:

- The Centers for Disease Control and Prevention should classify influenza virus as a "critical agent" for bioterrorism.
- Immunization should be expanded, possibly by requiring it for all medical personnel.
- Laboratories that work with influenza virus should strengthen their security.
- Antiviral drugs should be stockpiled, and vaccine-making capacity should be increased.
- The government should consider a gene-sequencing and vaccine development program.
- Surveillance efforts should be increased and should include incentives for reporting of clinical cases.
- The fitting of ventilation systems with virus detection and inactivation.

A super-flu virus, particularly a pandemic derivative of the current H5N1 avian flu, would presumably kill 200,000 and up to 1.9 million Americans, plus many millions

sickened, according to recent estimates. A government's plan to fight such contagion designates not just who cares for the diseased, but who will keep the country running amid the chaos. Specific countermeasures include stockpiling of both the anti-flu drug Tamiflu as well as avian flu vaccine [350]. Still, the protean nature of influenza type A virus at large may be defying. Experts in Hong Kong observed that the human H5N1 strain that surfaced in northern Vietnam during 2005 had proved to be resistant to Tamiflu [351]. Also, the present antigenic quality of the H5N1 may undesirably alter. Also, the present intact strain of avian flu, as isolated from its human victims, and although not yet a contagious pathogen, is a putative bioterrorism agent itself, due to its respiratory infectivity, lethality, and panic-causing nature.

In general, then, ideally—although seemingly impractically—a complete genetic inventory of all pathogenic strains found worldwide could provide the proper basis for initial, rather essential, inquiry into bioterrorism acts. It may facilitate, as well, controlling the registration of culture collection so as to hamper theft of dangerous pathogens. Thus oriented, the United States is making an effort to fully catalog and selectively to obtain prioritized pathogenic strains such as those then included in the Soviet BW program. Within that context, Kazakhstan, Azerbaijan, Ukraine, and Georgia are appreciably cooperative.

Consequently, more than 60 pathogens—including anthrax and plague strains—from the former Soviet Union's Azerbaijan-based biological weapons program plus epidemiological system arrived at Dover Air Force Base in Delaware [352]. In Azerbaijan—largely a Muslim country—several facilities—including field-test site—were engaged in the then USSR BW development program plus epidemic control. The samples were transferred as part of the Cooperative Threat Reduction program with Azerbaijan, aimed at getting the U.S. help to improve security for pathogens.

Moving to Russia, a different picture emerges with regard to a country fully acknowledging the essentiality of anti-bioterrorism moves, while still possessing an inventory of BWs and an active BW program [305]. The anti-bioterrorism outline and legislative base that were established in Russia—so as to deny possible access by terrorist states and terrorist groups to dual-use biotechnologies, plus materials—have been described by Vorobiev. He thereupon pointed at potential international cooperative efforts to be implemented, accordingly [353]. Eventually, in 2001, a commitment was achieved between Russian and U.S. Presidents, Vladimir V. Putin and George W. Bush, to pursue cooperation to counter the threat of bioterrorism, including a focus on health-related measures.

American assistance to Russia within the "Reduce the Common (WMD) Threats" program fared much. In the bioterrorism sphere, a chief course of that program for conversion has been the anti-bioterrorism one. The Novosibirsk-based Biopreparat-affiliated Vector microbiological center has been a pioneering facility, within that context. A long time before others, Vector's Director, Academician Sandakhchiev, started to cooperate with international institutions, and already in the early 1990s focused on issues related to monitoring and prevention of bioterrorism [354]. In addition, there are plenty of microbiological laboratories all over Russia eager—for financial, professional, or even ideological reasons—to save forbidden strains, toxins, equipment, and information. On the other hand, with time the Americans became more attentive to Russian arguments about development of anti-terror measures.

Consequently, The Russian-U.S. BioIndustry Initiative (BII) began in 2002, constituting the newest proliferation threat reduction program. It aims to counter bioterrorism through targeted reconfiguration of large-scale, formerly Soviet BW research, development, and production facilities for civilian purposes, by creating Russian-U.S. research partnerships [355]. Collaboration has then been established in 2003 between the International Science and Technology Center (ISTC), Moscow, and the Boston-based Center for Integration of Medicine and Innovative Technology (CIMIT), to implement the BII. Conversion is thus intended to take place through formation of systems in Russia to link scientists, physicians, and engineers to solve medical and scientific problems, and to identify innovative technologies and commercialization opportunities [356]. Also, the U.S. Defense Department has been increasingly engaged in efforts to secure from proliferation dozens of former Soviet pathogen collection and research facilities included within the "anti-plague system," widespread across 11 former Soviet states (excluding Russia). Apparently, many of those facilities lack sufficient safety and security and their scientists on average are poorly paid [357]. The concerned, so called Anti-Plague Institutes, are thus being turned into controllable "central reference laboratories" [358].

## 25 PROJECTIONS AND PROGNOSIS

*Built-in limitations, which interfere with comprehending and foreseeing the course of bioterrorism, are outlined and discussed, together with objective constrains that restrict the scope of countering it. Nonetheless, the conjunction of methodological intelligence monitoring, systematic biosecurity globalization, and creative scientific research may furnish a crucial tool for effectively fighting bioterrorism.*

The very supremacy of mankind, in terms of know-how and technology, may possibly make it prone, paradoxically, to self-destruction, in various manners. Bioterrorism is one main way to materialize that undesirable predisposition. Apparently, an outstanding conjunction marked the recent decade, bringing about the prominence of bioterrorism as a colossal issue, including:

- Internet-contained unclassified and declassified information.
- The outcomes of the birth of the formerly Soviet republics—particularly the Muslim and semi-Muslim ones—with their Soviet BW inheritance, namely highly qualified, equipped facilities, culture collections, and many unemployed experts, plus partial Islamic orientation.
- The accelerated augmentation of international terrorism, at large.
- The rise and persistence of al-Qaeda with its radical philosophy and unlimited financial resources.
- The global strengthening of Islam and its fundamentalistic inclination.
- Extremely meaningful breakthroughs made in life sciences and apparatus engineering.
- The increasing emergence and reemergence of infectious diseases worldwide.

Altogether, the ongoing integration of those various, interacting factors will probably shape, both conceptually and practically, the future of bioterrorism, one way or another. If a new breed of anarchistic multinational terrorists—not necessarily connected with al-Qaeda—is currently sprouting up, as at times claimed, the outcome may be unforeseeably catastrophic.

The outlining—let alone structuring—of a defensive alignment that would fully address the bioterrorism threat is, objectively, an impossible mission. Even while referring, primarily, to natural (nonengineered) pathogens and toxins, and assuming a consensus prioritizing anthrax, smallpox, plague, influenza, ricin, and botulinum, one can hardly face the two main resultant questions:

a. How to be best prepared toward each of those agents?
b. Should other potential bioterrorism agents be totally ignored, and if not—what ought to be the respecting mode of preparedness?

And, beyond, of course—what about potential engineered pathogens and toxins? Is it unfeasible, for instance, to form an anthrax germ resistant to the antibiotic Ciproflaxin? Overall, the immeasurable complexity thus formed seems to be insoluble even if budgetary aspects are ignored. Although frustrating, those issues can and should in part be practically administrated by bioterrorism-strategic planners, provided they are fully aware of and acknowledging the concurrent constraints and limitations.

Increasing success of acts of so-called conventional terrorism is not regarded to decrease the potentiality and potency of bioterrorism (or other WMD-terrorism). Recurrent operations of conventional sabotage—even if fully effective—may intensify a perpetrator's daring quest for escalation and resonance through varitype means and modes of terrorism. Perhaps the most significant feature of bioterrorism, the ratio between the low probability of threat realization on the one hand, as against the high impact generated in case of threat realization on the other hand, is an extremely problematic factor for the defender. Let alone, that bioterrorism act that took place—the enthrax letters—in effect looks, in part, successful. In a sense, the very fact that the anthrax letter attack—apparently the most significant bioterrorism act ever conducted—has not been deciphered since 2001 is no less meaningful and important than this unprecedented act of bioterrorism itself.

Apparently, the prime and ultimate apparatus to cope with the bioterrorism threat is effective monitoring by intelligence sources, at least in that if adequately structured it would expediently shape the real and concrete needs and deeds of defense, in whatever sense. Therefore, intelligence ought to be supremely prioritized. All defensive measures have to actually form as derivatives of the intelligence picture, which may certainly be dynamic, though, in itself. The defense planner would not like that dynamics, in case the latter is not a plainly evolving one, but this is an inherent disadvantage. Conversely, yet, or rather complementarily, the defense planner may rely on his own considerations, as long as they can objectively be figured out. The significance of intelligence has been accentuated by the U.S. House of Representatives Homeland Security Subcommittee on Prevention of Nuclear and Biological Attack Chairman John Linder, who called for better coordination between intelligence and disease-fighting agencies. "Science, tools, reagents and technology may be ubiquitous. Scientists, however, are not," Linder

said. "We have to do a better job of keeping track of those individuals with skill sets that are attractive to potential terrorists [359]."

In conjunction, it has been accentuated that a harmonized international regime that enhances biosecurity is essential for reducing the risk of bioterrorism [360]. Presumably, like other security regimes, this will entail mutually reinforcing strands, which need to include enactment of legally binding control of access to dangerous pathogens, transparency for sanctioned biodefense programs, technology transfer, and assistance to developing countries to jointly advance biosafety and biosecurity, global awareness of the dual-use dilemma and the potential misuse of science by terrorists, and development of a global ethic of compliance. To work, this effort must be undertaken collectively, using the international and regional institutions that already have a role to play in providing safety and security.

To those two essential elements—methodological intelligence monitoring and systematic international endeavor—should be added the dimension of creative scientific research. Apparently, the prospects for attaining a breakthrough in upgrading preparedness lie in the integral formed, then, by those three different spheres (beyond the self-evident readiness achieved by means of doctrine-structuring and medical stockpiling). The scientific sphere seems to bear enormous potential. Specific immunoglobulins may constitute a good example of a not yet adequately developed defensive tool, which potentially can be very efficient against pathogens and protein toxins, both for instant prophylaxis and treatment for as well as detection. This tool may be valid against both recognized and future engineered agents, provided that biotechnological processes will allow for at once production of fully specific immunoglobulins against a given agent.

Seemingly, to the least, this direction is worth elaborating on. Unlike vaccines and antibiotics, which are largely prioritized, antisera are less noticed, apparently unjustifiably. Antisera could constitute an optimal means for treatment, provided that instant and specific ethiological diagnosis is available. Studies and case reports evaluating convalescent plasma as therapy (or prophylaxis) of infectious diseases showed to be fairly promising. In the event of a bioterrorism attack, the usefulness of active immunization may be limited. Vaccine efficacy often requires time, multiple doses, and a competent immune system. Prophylactic immunization might provide an effective defense against bioterrorism agents but has the disadvantage that many individuals would have to be vaccinated to protect against an attack that might never occur. In this situation, even a small number of vaccine-related side effects would be unacceptable. In addition, vaccines do not induce protective immunity in all recipients, especially immunocompromised individuals [361]. Such inadequacies should carefully be considered. Yet in contrast, passive immunization involves the administration of preformed antibody to provide a state of immediate immunity. The two modes of immunization are used together in certain circumstances, such as rabies prophylaxis following possible exposure, where a passive antibody provides immediate protection and a vaccine elicits a protective immune response.

A protective antibody suitable for passive immunization could be used in concert with vaccines and drugs to provide a multilayered defense against attacks with biological agents. It should be possible to create a strategic reserve of immunoglobulins against the major biological warfare agents that can be rapidly administered to exposed individuals in the event of an attack. The availability of a strategic

reserve of specific antibody preparations would have a significant deterrent value, as aggressors would be aware that the lethality of their weapon could potentially be counteracted by prompt administration of antibody to susceptible individuals. As antibody can be administered intramuscularly, immunoglobulin preparations could be packaged in self-injectable, disposable, single-use containers. Self-administration of antidote would avoid taxing the health-care system with the need for intravenous administration. Given the stability of immunoglobulin preparations, it should be possible to store antibody for many years. Developing, producing, and stockpiling antibody reagents for defense against bioterrorism agents is a sensible strategy that needs to be considered as steps are taken to prepare against the threat of bioterrorism [336]. On the whole, so it seems, the practicality of antisera should rather equalize that of vaccines and drugs, if not beyond. Connectedly, it has been highlighted that, in the U.S. National Institute of Allergy and Infectious Diseases work on boosting the human innate immune system, a strategy is formed that could lead to countermeasures that would be useful against a wide variety of different agents [359].

Such strategy may become vital, owing to the considerably diminishing usefulness of antibiotics against bacterial pathogens, as predicted by Prof George Poste, Director of the Biodesign Institute at Arizona State University and an advisor to the U.S. President: "Frankly, most governments are asleep at the switch"; he predicts that from 2010 to 2015 will be a "window of vulnerability" when the toll of the super-bug will reach its peak as a result of naturally augmenting antibiotic resistance [362]. Antibiotics may therefore be appreciably replaced by other protective agents, such as antibodles, receptor decoys, dominant-negative inhibitors of translocation, small-molecule inhibitors, and substrate analogues [363].

A new horizon may be gained through the Neugene antisense technology. Neugene antisense "rapid response therapeutics" are synthetic compounds that mirror a critical portion of a disease-causing organism's genetic code and bind to specific portions of the target genetic sequence. Like a key in a lock, Neugene antisense compounds are designed to match up perfectly with a specific gene sequence, blocking the function of the target gene [364]. Resultant antisense preparations have been able to block the cellular mechanisms used by the causative agents of anthrax, Ebola and Marburg, during pathogenesis. Another variant has been reported to block ricin toxin as well [365].

Relying on a similar principle, a novel drug has lately been developed that could expectedly be used to treat people exposed to anthrax bacteria specifically engineered to overcome antibiotics, due to an increasing concern that therapeutics developed for bioterrorism agents may be rendered ineffective if the microbial target is altered intentionally. This problem could be overcome, however, by designing inhibitors that block host proteins used by pathogens or their toxins to cause disease [366].

Thus, although the bioterrorism threat stemming from nonengineered pathogens and toxins is more or less characterized, in terms of expected impacts (although not in terms of likelihood, timing, and locality), the vagueness marking the threat posed by engineered pathogens and toxins is intriguing. Synthetic biology may be a key arena. Security experts fear that scientists building synthetic biochemical compounds could create a pathogen with no natural countermeasures, hence the possibility of making an unprecedented deadly bioterrorism weapon [367].

Synthetic biologists are combining existing chemical components of DNA or RNA taken from cells and viruses to form compounds that do not occur naturally. Scientists hope to use this technology to produce computers, medicines, and energy sources. The field is attracting attention from investors and prominent biologists; yet some biologists and security experts warn that these new compounds could be used to develop a biological weapon. "There are certainly a lot of national security implications with synthetic biology."

Concerns were played down that a new agent could exterminate the human race but warned that the threat of new, engineered pathogens remains serious [368]. Scientists are already exploring ways to self-police synthetic biological research, including requiring reports of sales of materials that could be used to create a tailored biological-weapon. "There are now tens of thousands of people, worldwide, who could engineer drug-resistant anthrax," said Professor Kenneth Alibek, who as a consultant to the U.S. government has received numerous briefings on U.S. and Soviet biological weapon programs.

A testimony prepared to the House of Representatives Homeland Security Subcommittee on Prevention of Nuclear and Biological Attack may well illustrate the issue of bioengineered agents. Addressing it, National Institute of Allergy and Infectious Diseases Director Anthony Fauci contented that agents could be made more virulent through "resistance to one or more antibiotic or antiviral drugs, increased infectiousness or pathogenicity or, in the somewhat longer term, a new virulent pathogen made by combining genes from more than one organism. As the power of biological science and technology continues to grow, it will become increasingly possible that we will face an attack with a pathogen that has been deliberately engineered for increased virulence." Fauci added, however, that creating an agent whose transmissibility could be sustained on such a scale, even as authorities worked to counter it, would be a daunting task. "Would you end up with a microbe that functionally will . . . essentially wipe out everyone from the face of the Earth? . . . It would be very, very difficult to do that," he said.

Centers for Disease Control and Prevention Director Julie Gerberding said a deadly agent could be engineered with relative ease that could spread throughout the world if left unchecked, but that the outbreak would be unlikely to defeat countries' detection and response systems. "The technical obstacles are really trivial," Gerberding said. "What's difficult is the distribution of agents in ways that would bypass our capacity to recognize and intervene effectively."

Today, the CDC's list of pathogens that must be reviewed comprises more than 60 agents. Each of them may completely be explored, so as to become fully manageable. But the catch is in that the existent pathogens are recognized—hence controllable—yet currently available for the potential attacker, whereas the unknown future engineered pathogens will scarcely be available, yet hardly controllable. The dichotomic menace thus emerging is an extremely challenging one. Possibly, as mentioned, immunoglobulins, polyvalent inhibitors, and other opproaches, may furnish a solution.

All in all, the threat of bioterrorism is indeed a challenging one and bears, potentially, overwhelming, somewhat enigmatic impacts. Henry Crumpton, the US State Department Coordinator for Counter-Terrorism, described a biological attack on West "simply a matter of time", adding that such an attack could pose an even greater menace to security than a nuclear strike [369].

U.N. Secretary General Kofi Annan, in the General Assembly, laid emphasis on the hazardous conjunction of bioterrorism and biotechnology:

"The international community must give a higher priority to developing new strategies against bioterrorism. Biotechnology has value in the fight against disease, but scientific advances could also "bring incalculable harm if put to destructive use by those who seek to develop designer drugs and pathogens. Soon, tens of thousands of laboratories worldwide will be operating in a multibillion-dollar industry," Annan said, according to the Associated Press. "Even students working in small laboratories will be able to carry out gene manipulation. U.N. nations should consider organizing of forum of governmental officials and science and public health experts to develop a strategy "to ensure that biotechnology's advances are used for the public good and that the benefits are shared equitably around the world [370]."

Finally, perhaps most plainly forwarded by the head of the Interpol, Ronald Noble, the bioterrorism treat has been featured as a global, severe, unprecedented menace: "The world is ill prepared for the looming threat of a bioterrorism attack. The danger of an al-Qaeda attack has not diminished since the 9/11 strikes on the US. The potential cost of a bioterrorism attack left no room for complacency. When you talk about bioterrorism, that's one crime we can't try to solve after it happens because the harm will be too great. How could we ever forgive ourselves if millions or hundreds . . . or tens of thousands of people were killed simply because our priorities did not include bioterrorism?" Noble acknowledged that governments and security agencies were better organized against the threat than ever before, but "none of us can let our guards down and assume that the problem has been addressed". Were al-Qaeda to launch a "spectacular bioterrorism attack which could cause contagious disease to be spread, no entity in the world is prepared for it," he said. "Not the US, not Europe, not Asia, not Africa [371]." He may possibly be right.

## REFERENCES


1. Currie J. The earliest weapons of mass destruction. *Christ. Sci. Mon.* Oct. 9, 2003 Available: http://www.csmonitor.com/2003/1009/p17s01-bogn.html.
2. Simon J D (1997). Biological terrorism. Preparing to meet the threat. *J. Amer. Med. Assoc.* 278(5):428–430.
3. SIPRI (1993). The Problem of Chemical and Bioterrorism Warfare, Vol. 2. Humanities Press, New York, p. 143.
4. Zelicoff A P (1996). Preparing for bioterrorism terrorism: First, do not harm. *Politics Life Sci.* 15(2):235–236.
5. Zilinskas R (1999). Cuban allegations of biological warfare by the United States: Assessing the evidence. *Crit. Rev. Microbiol.* 25(3):173–227.
6. Cunningham W, Cunningham M (1998). Principles of Environmental Science. Available: http://highered.mcgraw-hill.com/sites/0072452706/student_view0/chapter8/additional_case_studies.html#ddt.
7. Krupitsky E, et al. (2004). The onset of HIV infection in the Leningrad region of Russia: A focus on drug and alcohol dependence. *HIV Med.* 5(1):30.
8. TASS, Online, Moscow, 2 April 2002.

9. SARS- Origins of the Virus. Available: http://www.bikesutra.com/sars/origins_old.html.
10. Zhu G, Chen H-W (2004). Monophyletic relationship between Severe Acute Respiratory Syndrome Coronavirus and Group 2 Coronaviruses. *J. Infec. Dis.* 189:1676.
11. Lam W K, et al. (2003). Overview on SARS in Asia and the world. *Respirol.* 8(s1): S2.
12. Chang H-J, et al. (2004). The impact of the SARS epidemic on the utilization of medical services: SARS and the fear of SARS. *Amer. J. Pub. Health.* 94(4):562–564.
13. Nash D, Mostashari F, Fine A, et al. (2001). The Outbreak of West Nile virus infection in the New York City area in 1999. *N. Engl. J. Med.* 344:1807–1814.
14. Komar N, Panella N A, Burns J E, et al. (2001). Serologic evidence for West Nile virus infection in birds in the New York City vicinity during an Outbreak in 1999. *Emerg. Infect. Dis.* 7:621–625.
15. Steele K E, Linn M J, Schoepp R J, et al. (2000). Pathology of fatal West Nile virus infections in native and exotic birds during the 1999 Outbreak in New York City, New York, *Vet. Pathol.* 37:208–224.
16. Giladi M, Metzkor-Cotter E, Martin D A, et al. (2001). West Nile encephalitis in Isael, 1999: The New York connection. *Emerg. Infect. Dis.* 7:659–661.
17. Briese T, Jia X Y, Huang C, et al. (1999). Identification of a Kunjin/West Nile-like flavivirus in brains of patients with New York encephalitis. *Lancet.* 354: 1261–1262.
18. Jia X Y, Briese T, Jordan I, et al. (1999). Genetic analysis of West Nile New York 1999 encephalitis virus. *Lancet.* 354:1971–1972.
19. Lanciotti R S, Roehrig J T, Deubel V, et al. (1999). Origin of the West Nile virus responsible for an Outbreak of encephalitis in the northeastern United States. *Science.* 286:2333–2337.
20. Weiss D, Carr D, Kellachan J, et al. (2000). Clinical findings of West Nile virus infection in hospitalized patients, New York and New Jersey, 2000. *Emerg. Infect. Dis.* 7:654–658.
21. Marfin A A, Petersen L R, Eidson M, et al. (2000). Widespread West Nile virus activity, eastern United States, 2000. *Emerg. Infec. Dis.* 7:730–735.
22. WHO. Avian Influenza. Available: www.who.int/csr/disease/avian_influenza/ en/.
23. Shoham D (1993). Biotic-abiotic mechanisms for long-term preservation and reemergence of influenza type A virus genes. *Prog. Med. Virol.* 40:178–192.
24. Madjid M, Lillibridge S, Mirhaji P, et al. (2003). Influenza as a bioweapon. *J. R. Soc. Med.* 96:345–346.
25. US ready with guide on pandemics, 8 Jan. 2007, Times of India, New Delhi.
26 NHS "meltdown" predicted by Government bird flu report, UK Telegraph, Aug. 28, 2006.
27. Faruque S M, Chowdhury N, Kamruzzaman M, et al. (2003). Reemergence of epidemic Vibrio cholerae O139, Bangladesh. *Emerg. Infect. Dis.* 9:1116–1122.
28. Global Epidemics and Impact of Cholera. Available: http://www.who.int/topics/cholera/impact/en/.
29. Cloete T E, Nevondo T S (2001). The Global Cholera Pandemic. Available: http://www.scienceinafrica.co.za/2001/September/cholera.htm.
30. Shoham D (2000). Iraq's biological warfare agents: A comprehensive analysis. *Crit. Rev. Microbiol.* 26:179–204.
31. Malkinson M, Banet C, Weisman Y, et al. (2002). Introduction of West Nile virus in the Middle East by migrating white storks. *Emerg. Infect. Dis.* 8:392–397.

32. Shoham D (2005). *Viral Pathogenes of Humans Likely to be Preserved in Natural Ice*. Princeton University Press, Princeton, NJ, p. 221.
33. Horowitz G L (1996). *Emerging Viruses: AIDS and Ebola-Nature, Accident or Intentional?* Tetrahedron Inc., Sandpoint, ID.
34. Mumps Outbreak Aids Kansas Biological Preparedness, Associated Press/The Wichita Eagle, June 16, 2006.
35. Cassa C A, Iancu K, Olson K L, et al. (2005). A software tool for creating simulated Outbreaks to benchmark surveillance systems. *BMC Med. Informat. Decis. Making.* 5:22.
36. Cosgrove S E, Perl T M, Song X, et al. (2005). Ability of Physicians to diagnose and manage illness due to Category A bioterrorism agents. *Arch. Intern. Med.* 165:2002–2006.
37. U.K. Increases Biological Security, Tendler/McGrory, The Times, Jan. 25, 2007.
38. The Black Death Available: http://www.studentcentral.co.uk/black_death_9886.
39. Winn Jr. W C (1996). Medical microbiology. In B Samuel (ed.), *Legionella*. The University of Texas Medical Branch at Galveston, Galveston, TX.
40. Corbel M J (1997). Brucellosis: An overview. *Emerg. Infect. Dis.* 3:213–221.
41. Young E J (1995). An overview of human brucellosis. *Clin. Infect. Dis.* 21:283– 290.
42. Dye C. Scheele S, Dolin P, et al. (1999). Consensus statement. Global burden of tuberculosis: Estimated incidence, prevalence, and mortality by country. WHO Global Surveillance and Monitoring Project. *JAMA*. 282:677–686.
43. Anonymous (2005). TB to be Subject to Parameters of Bioterrorism Law. *The Japan Times.*, Sept. 20, 2005 Available: http://www.japantimes.co.jp/cgi-bin/makeprfy.pl5?nn20050920b1.htm.
44. Meegan J M, Hoogstraal H, Moussa M I (1979). An epizootic of Rift Valley fever in Egypt in 1977. *Vet. Rec.* 105:124–125.
45. CDC. Rift Valley Fever Fact Sheet. Available: http://www.cdc.gov/ncidod/dvrd/spd/mnpages/dispages/Fact_Sheets/Rift_Valley_Fever_Fact_Sheet.pdf.
46. WHO (2001). Yellow Fever Fact Sheet. Available: http://www.cdc.gov/ncidod/dvrd/spd/mnpages/dispages/Fact_Sheets/Rift_Valley_Fever_Fact_Sheet.pdf.
47. Jaworski M (2003). Norwalk Virus. Available: http://www.emedicine.com/med/topic1648.htm#section~author_information.
48. Rubins K H, Hensley L E, Jahrling P B, et al. (2004). The host response to smallpox: Analysis of the gene expression program in peripheral blood cells in nonhuman primate model. *Proc. Natl. Acad. Sci. USA*. 101:15190– 15195.
49. Lacy D B, Tepp W, Cohen A C, et al. (1998). Crystal structure of botulinum neurotoxin type A and implications for toxicity. *Nat. Struct. Biol.* 5:898–902.
50. McCormick J B, Steele J H, Hendricks K (2005). Staphylococcal Enterotoxin B-Factsheet. Available: http://www.dshs.state.tx.us/idcu/disease/staphylococcal/seb/factsheet.
51. *US Army Medical Research Insitute of Infectious Diseases*, (1998) Medical Management of Biological Causalties Handbook 3rd ed.: Frederick, Fort Detrick, MD, p. 108, 109.
52. Wannemacher Jr. R W, Wiener S L (1997). Trichothecene Mycotoxins, Medical aspects of chemical and biological warfare, F R Sidell, COL E T Takafuji, COL D R Franz (eds.). In: B G R Zajtchuk, COL R F Bellamy, *Part I, Warfare, Weapons, and the Casualty, Textbook of Military Medicine: Medical Aspects of Chemical and Biological Warfare*, Office of the Surgeon General, Walter Reed Army Medical Center, Washington, D.C., p. 656.

53. Seagrave, S (1981). *Yellow Rain*. Evans & Co., New York, P. 134.
54. Kadivar H, Adams S (1991). Treatment of chemical and bioterrorism warfare injuries: insights derived from the 1984 Iraqi attack on Majnoon Island. *Military Med*. 156: 170.
55. Heyndrickx A, et al. (1989). Detection of trichothecene mycotoxins in Iranian soldiers treated as victims of a gas attack. *Rivista Toxicol. Speriment*. 19:7.
56. United Nations, Department of Public Information, JY (1996). The United Nations and the Iraq-Kuwait Conflict, 1990–1996, In: *The United Nations Blue Book Series*, p. 784.
57. US Army Medical Research Institute of Infectious Diseases (1998). *Medical Management of Biological Casualties Handbook*, 3rd ed.: Frederick, Fort Detrick, MD, p. 107.
58. Yourtee D M, Raj H G, Prasanna H R, et al. (eds) (1989). Proceedings of the international symposium on agricultural and biological aspects of aflatoxin related health hazards. *J. Toxicol. Toxin Rev*. 8:3–375.
59. SIPRI Fact Sheet (1998). Iraq: The UNSCOM Experience. www.sipri.org/contents/webmaster/publications.
60. UNSCOM Periodical Report (1999). Current e-mail notifications on Iraq; by Lautie Mylroie sam11@erols.com, January 25.
61. French G (1958). Aconitine-induced cardiac arrhythmia. *Br. Heart J*. 20:140–142.
62. Tai Y T, But P P, Young K, et al. (1992). Cardiotoxity after accidental herb-induced Aconite poisoning. *Lancet*. 340:1254–1256.
63. Chan T Y (1994). Aconitine poisoning: A global perspective. *Vet. Hum. Toxicolo*. 36:326–328.
64. Chan T Y, Tomlinson B, Tse L K, et al. (1994). Aconitine poisoning due to Chinese herbal medicines: A review. *Vet. Hum. Toxicolo*. 36:452–455.
65. Fatovich D M (1992). Aconite: A lethal Chinese herb. *Ann. Emerg. Med*. 21:309–311.
66. Kolev S T, Leman P, Kite G C, et al. (1996). Toxicity following accidental ingestion of Acotinum containing Chinese remedy. *Hum. Exp. Toxicol*. 15:839–842.
67. Honerjager P, Meissner A (1983). The positive inotropic effect of Aconitine. *Naunyn Schmiedebergs Arch. Pharmacol*. 322:49–58.
68. H'ctor A, de Landoni H J (1990). Aconitum Napellus spp. Available: http://www.inchem.org/documents/pims/plant/aconitum.htm#SectionTitle: 12.3%20Other.
69. El–Siad (Lebanon), The Death of Marsall Amer, Editorial 14 August 1975.
70. Anonymous (1976). Latest dirty trick is fake nerve as antidote. *New Scientist*. 20 May, p. 423.
71. Adam H L (1953). Trial of Dr. Lamson (The Blenheim School Murder). *J. Crim. Law, Criminol. Police Sci*. 44:361–362.
72. Anonymous (1966). Deliberated spreading of Typhoid in Japan. *Sci. J*. 2:11–12.
73. U.S. Congress Office of Technology Assessment (1992). *Technology against terrorism: structuring security*, OTA-ISC-511: US Government Printing Office, Washington, D.C.
74. Bishop S (1980). GRA Seeks Botulin *International Herald Tribune*, November 8–9.
75. Shalev R, Bioterrorists Arrested in Belgium. Ma'ariv, March 23, 1998, p. 3.
76. Douglass J D (1990). America the Vulnerable, Lexingtom Books, p. 196.
77. CNN, Plague Plot, Feb 19, 1998.
78. Shoham D (2000). *Crit. Rev. Microb*. 26:(3):179–204.

79. Dave Altimari, Hartford Courant, Oct. 9, 2006.
80. Shoham D (2003). The anthrax evidence points to Iraq. *Intern. J. Intell. CounterIntell.* 16(1):39–68.
81. Joby Warrick, Suspect and A Setback In Al-Qaeda Anthrax Case, Washington Post, October 31, 2006; A01.
82. September 19, 2001. Available: Worldnetdaily.com. Biological Weapons in South Africa.
83. Keim P, Smith K L, Keys C, et al. (2001). Molecular investigation of the Aum Shinrikyo anthrax release in Kameido, Japan. *J. Clin. Microbiol.* 39:4566–4567.
84. Alon B 19 April 2005. Ma'ariv (Israel). Palestinian Labs Worry, p. 4.
85. Anonymous (2005). Your bioweapon is in the mail. UPI, Washington, Nov 9.
86. Bureau of Emergency Preparedness and Response: Arizona Department of Health Service. History of biowarfare and bioterrorism. Available: http://www.azdhs.gov/phs/edc/edrp/es/bthistor2.htm.
87. Ross E (2005). Labs urged to destroy pandemic flu strain. *Associated Press*, London, U.K., April 12.
88. WHO Memorandum (2005). Missing Samples of Deadly Flu Strain, April 19, 2005.
89. Anonymous, A letter blamed for an epidemic of smallpox, *N Y Med. J.* 1901;73:600.
90. Boobbyer P (1901). Smallpox in Nottingham, *Br. Med. J.* 1901;1:1054.
91. Audrey Woods (2002). 1960s germ tests carried out in London Underground, Associated Press February 26, 2002, Available from: URL http://www.nti.org/d_newswire/issues/2002/2/27/13s.html.
92. Carlton, Jim, Of Microbes and Mock Attacks: Years Ago, The Military Sprayed Germs on U.S. Cities, Wall Street Journal, October 22, 2001.
93. BBC Radio 4 Documentary, Hotel Anthrax, February 13, 2006.
94. Alon B Maariv, (Israel). 3 November, 1996. Sewage Influx Causes Concerns, p. 6.
95. Alon B Maariv, (Israel). 1 December, 1996. Sewage Contamination Increases, p. 5.
96. Braveman I, Wexler D, Oren M (2002). A novel mode of infection with hepatitis B: Penetrating bone fragments due to the explosion of a suicide bomber. *Isr. Med. Assoc. J.* 4:528–529.
97. MacKenzie D (2002). Suicide Bombers May Spread Disease. Available: http://www.newscientist.com/article.ns?id=dn2588.
98. Abdel-Aziz G Arab daily "al-Sharq al-Awsat" 11 Nov 2004. Iranian UAVs Operated by Hezbollah, p. 2.
99. Anonymous Janes Defence Weekly, 13 August, 1997. Ventilation Systems Vulnerable to Bioterrorism, p. 14.
100. Norqvist A (1995). Covert use of bioterrorism and toxin weapons—For what purpose? *Proc. 5th International Symposium Protection Against CB Warfare Agents*, Stockholm, Sweden, pp. 367–372.
101. Geissler E (1996). Joint international action is necessary to counter the threat of C/B terrorism. *Polit. Life Sci.* 15(2):205–207.
102. Grange D L (2001). Asymmetric warfare. *The Officer.* Available: http://www.findarticles.com/p/articles/mi_m0IBY/is_2_77/ai_73326816#continue.
103. McKenzie K (2001). The Rise of Asymmetric Threats: Priorities for Defense Planning, Institute for National Strategic Studies. Available: http://www.ndu.edu/inss/press/QDR_2001/sdcasch03.html.
104. (1998). BBC, March 23.
105. Blenkin, Max, Australian Associated Press/The Age, Dec. 1, 2005.

106. Faruque S, Chowdhury N, Kamruzzaman M, et al. Reemergence of epidemic-Vibrio cholerae O139, Bangladesh. Emerging Infections Diseases, 9(9):1116–1122, 2003.
107. McHale P (2005). Pentagon preparing for multiple WMD attacks. *Inside Missile Defense*, May 25.
108. Brissenden M (2005). Biological agent sent to Indonesian Embassy. *The 7:30 Report* ABC Australia.
109. Associated Press, July 8, 2005. Man Sentenced to 19 years for Anthrax Hoaxes.
110. Holloway H C, et al. (1997). The threat of bioterrorism weapons. *JAMA*. 278(5): 425–427.
111. McPhedran I (2005). Suspicious Powder letter, Daily Telegraph June, 4.
112. Australian AP, June 2, 2005. Osborne P Powder Letter Found to be Harmless.
113. Schneidmiller C. Years after Anthrax Attacks, Hoaxes Persist. Global Security Newswire. Available: http://www.nti.org/d_newswire/issues/2005/7/19/9c68a408-b230-431e-9769-6805f6855b99.html.
114. Poisoned Arrows, V&A. Available: http://www.vam.ac.uk/collections/asia/object_stories/arrows/.
115. Above Top Secret—War On Terrorism. Available: http://www.abovetopsecret.com/forum/viewthread.php?tid=18930&s.
116. B-WAR. Available: www.junglelogic.tv/artf/b_war.html.
117. McGovern T W, Christopher G W (2000). Biological warfare and its cutaneous manifestations. In: Dr. R Drugge, H A Dunn (eds.), *The Electronic Textbook of Dermatology*. Available: http://www.telemedicine.org/BioterrorismWar/bioterrorismlogic.htm.
118. Harris S H (1994). *Factories of Death: Japanese Biological Warfare*, *1932–1945, and the American Cover-Up*. Routledge, New York, pp. 1–147.
119. Spencer, R C, Bacillus anthracis, Journal of Clinical Pathology, 2003, 56:182–187.
120. Alibek K, Handelman S (1999). *Biohazard*. Random House, New York, pp. 29–31.
121. US Politics Online: A Political Discussion Forum—what about Israel?! Available: www.uspoliticsonline.com/forums/archive/index.php/t-18141.htm.
122. Phills J A, et al. (1972). Pulmonary ringworm infestation in man. *N. Eng. J. Med*. 2:11–12.
123. Simon J D (1997). Biological terrorism. Preparing to meet the threat. *J. Amer. Med. Assoc*. 278(5):8.
124. Robert S (1991). Root-Bernstein, Infectious Terrorism, The Atlantic Monthly, 267: 44–50.
125. Kifner, J (1995), Man Arrested in a Case Involving Ricin Ny Times, Dec. 23.
126. Shemer P (1994), Yediot Ahronot, 10 October, 1994. Iragi Scientist Implanted in NY, p. 8.
127. Herbert B (1995). Militia Madness. *The New York Times*, June 7.
128. Rapaport A (1998), Ma'ariv, 22 February, 1998. Iragi Women Intend to Smuggle Bioagents, p. 3.
129. Alon B (1998). Muslim Bioterrorists in Belgium, p. 2. Ma'ariv, 23 March, 1998.
130. Sarrar S (2005). Libya Acts in HIV Row with Bulgaria. *The Guardian*, April 13.
131. Blaney B (2005) Everything is Under Consideration. *Associated Press/Star-telegram*, October 25.
132. Carus W S (1998). *Bioterrorism and Biocrimes*: *The Illicit Use of Bioterrorism Agents in the 20 Century*. Center for Counterproliferation Research, National Defense University, Washington, D.C.

133. Wheelis M (2004). *A short History of Biological Warfare and Weapons*. ISO Press, Amsterdam, The Netherlands.
134. Wheelis M (1999). Bioterrorism sabotage in World War I. In: E Geissler, J Ellis, van Courtland Moon (eds.), *Biological and Toxin Weapons: Research, Development, and Use from the Middle Ages to 1945*: Oxford University Press, Oxford, UK, pp. 35–62.
135. Biological Warfare. Available: http://www.nutrition-report.com/articles/Biological_warfare.
136. Nelson A K (1998). JFK assassination review board releases top secret documents. *Organization of American Historians Newsletter.* Available: http://www.indiana.edu/%7Eoah/nl/98feb/jfk.html.
137. Ad Hoc Group of the States Parties to the BW Convention (2000). University of Bradford. Available: http://www.brad.ac.uk/acad/sbtwc/ahg52wp/wp417.pdf.
138. Livingstone N C, Douglass J D (1984). CBW: The poor man's atomic bomb. *National Security Papers*. Vol 1. Institute for Foreign Policy Analysis, Inc., Cambridge, MA.
139. Anonymous Book review: Swine fever warfare (1997). *Hartford Web Publishing*. Available: http://www.hartford-hwp.com/archives/43b/069.html.
140. Cummings R H (1996). Code Name Piccadilly—The Murder of Georgi Markov. *Balkan Info*. Available: http://www.b-info.com/places/Bulgaria/news/96–02/feb09.scb.
141. FBI File: Bacteriological Warfare in the Unted States (released May 2006) www.thememoryhole.org/fbi/biowar.htm.
142. South Africa Profile Available: http://www.nti.org/e_research/profiles/SAfrica/Biological/index.html#fnB2.
143. Truth and Reconciliation Commission, South Africa (1998). Chemical and Biological Warfare Hearings. Available: http://www.totse.com/en/politics/the_world_beyond_the_usa/camera.html.
144. Burger M, Gould C (2002). *Secrets and Lies: Wouter Basson and South Africa's Chemical and Biological Warfare Program*. Zebra, Cape Town, South Africa, pp. 34–35.
145. Kemp A (1990). *Victory or Violence: The Story of the AWB*. Forma, Pretoria, South Africa, p. 47.
146. Van Rooyen J (1994). *Hard Right: The New White Power in South Africa*. I.B.Tauris, New York, pp. 96, 97, 199.
147. NTI: Country Overviews: South Africa: Chemical Overview. Available: www.nti.org/e_research/profiles/SAfrica/Chemical.
148. Mangold T, Goldberg J (1999). *Plague Wars: The Terrifying Reality of Biological Warfare*. St. Martin's, New York, pp. 250, 254, 277–279.
149. Humes E (2001). The Medicine Man. *Los Angeles Magazine*, July:95–96.
150. CBS News (2002). "Dr. Death" and his Accomplice. *60 Minutes*, November 7.
151. Collins J (2000). Ford advised S. Africa on Warfare Devices. *Orange County Register*, 15 March.
152. Allen A (2002). Mad Scientist. *Salon Magazine*, June 26. Available: http://archive.salon.com/health/feature/2000/06/26/biofem.
153. Bell R (2004). Court TV's Crime Library. Available: www.crimelibrary.com/terrorists_spies/larry_ford/1.html.
154. Kouzminov A (2005). *Biological Espionage: Special Operations of the Soviet and Russian Foreign Intelligence Services in the West*. Greenhill, London.
155. Markov G. Available: http://www.videofact.com/english/defectors_33_en.html.

156. Ridley Y (2005). Don't Let Facts Spoil a Good Story. Available: http://www.ukwatch.net/article/450.
157. Miller J (1993). Evidence Grows on Bioterrorism Weapons. *New York Times Magazine*, January 3, p. 33.
158. Anonymous (1990). A Case of Finding Evidence, Jane's Defence Weekly, July 14, p. 54.
159. Anonymous (1990). On the Street. *The ASA Newsletter*, August issue.
160. UNSCOM Periodical Report (1998) Current e-mail notifications on Iraq; by Laurie Mylroie sam11@erols.com, June 3.
161. Duelfer C (2004). Comprehensive Report of the Special Advisor to the DCI on Iraq's WMD. Available: http: //www.cia.gov/cia/reports/iraq_wmd_2004/index.html.
162. Shoham D (2006). An antithesis on the fate of Iraq's chemical and biological weapons. *Internat J. Intell. Counterintell*. 19(1):59–83.
163. Eisenstadt M (1996). Iranian *Military Power: Capabilities and Intentions*. Washington Inst. for Near East Policy, Washington, D.C., p. 26.
164. Gaffney Jr. F J (2004). Terror-tied by Memo. *The Washington Times*, May 9.
165. Schanzer J (2003). *A. Al-Islam: Iraq's Al-Qaeda Connection*. The Washington Institute for Near East Policy, Washington, D.C. Available: http://www.freerepublic.com/focus/news/824263/posts.
166. Warrick J, al-Qaeda's Bio-lab, Washington Post May 5, 2004.
167. Parry D. Interfax/Mosnews.com, March 1, 2005. al-Qaeda Explores Bioagents.
168. Gertz B (2004). Iraqi Bomb Labs Signal Attacks in the Works. *The Washington Times*, November 30.
169. Anonymous (2004). Iraqi Women Not Being Released. *BBC News World Edition*, September 22.
170. McMahon M. NewsMax.com Monday, December 29, 2003. al-Qaeda Seeks Anthrax; no page.
171. Leitenberg M (2002). Biological weapons and bioterrorism in the first years of the twenty-first century. *Politics Life Sci*. 21(2):3–27.
172. Taliban Official Said Found With Anthrax, Agence France-Presse/The Nation, Jan. 17, 2007.
173. Associated Press (2005). 8 of 9 Acquitted in U.K Ricin Trial. April 13.
174. Uhlig R, Fleming N (2004). Flights Axed in New Terror Alert. Available: http://www.telegraph.co.uk/news/main.jhtml?xml=/news/2004/02/02/nterr02.xml.
175. Anonymous, MI5 warns UK labs of threat from al Qaeda, London Daily Mail, January 25, 2007.
176. Bockmann M W (2004). Biochemical Threat Next Terror Wave. *The Australian*.
177. Anonymous *The Times*, 3 March, 1998. Hezbollah Trying to Obtain Biochemicals page unknown.
178. *Reuters*, 9 Feb. 2003.
179. Al-Manar (2001). Will We Reach the Option of Bioterrorism Deterrence? August 13.
180. Luft, G (2003). The Palestinian security forces and the second intifada. *Institute for the Analysis of Global Security, Presentation given at the National Press Club*. Washington, D.C.
181. Rapaport A, Ma'ariv, 13 April 2004. Tanzim Plants to Use Aids Virus, p. 7.
182. Palestinians Claim Biological and Chemical Weapons Capability, Israel Today, June 26, 2006.

183. Kaplan D E, Marshall A (1996). *The Cult at the End of the World*. Crown Publishers, Inc., New York.
184. Torok T J, Tauxe R V, Wise R P, et al. (1997). A large community outbreak of salmonellosis caused by intentional contamination of restaurant salad bars. *JAMA*. 278: 389–395.
185. Kolavic S A, Kimura A, Simons S L, et al. (1997). An outbreak of Shigella dysenteriae type 2 among laboratory workers due to intentional food contamination. *JAMA*. 278:396–398.
186. Joseph P R, Millar J D, Henderson D A (1965). An outbreak of hepatitis traced to food contamination. *N. Engl. J. Med*. 273:188–194.
187. Phills J A, Harrold A J, Whiteman G V, et al. (1972). Pulmonary infiltrates, asthma and eosinophilia due to Ascaris suum infestation in man. *N. Engl. J. Med*. 286:965–970.
188. Chester C V, Zimmerman G P (1994). Civil Defense Implications of Bioterrorism Weapons. *J. Civil Def.* 16(6):6–12.
189. Fetter S (1991). International Security, Summer 1991, p. 27. Ballistic Missiles and WMD.
190. MSNBC, 13 March 2003. The Impact of Anthrax Available: http://www.msnbc.com/news/886761.asp?0cv=CB20&cp1=1.
191. American Anthrax Outbreak of 2001, Case 3. UCLA School of Public Health, Los Angeles, CA. Available: http://www.ph.ucla.edu/epi/bioter/detect/antdetect_case3.html.
192. American Anthrax Outbreak of 2001, Case 4. UCLA School of Public Health, Los Angeles, CA. Available: http://www.ph.ucla.edu/epi/bioter/detect/antdetect_case4.html.
193. American Anthrax Outbreak of 2001, Case 5. UCLA School of Public Health, Los Angeles, CA. Available: http://www.ph.ucla.edu/epi/bioter/detect/antdetect_case5.html.
194. American Anthrax Outbreak of 2001, Case 7. UCLA School of Public Health. Available: http://www.ph.ucla.edu/epi/bioter/detect/antdetect_case7.html.
195. American Anthrax Outbreak of 2001, Case 1. UCLA School of Public Health, Los Angeles, CA. Available: http://www.ph.ucla.edu/epi/bioter/detect/antdetect_case1.html.
196. American Anthrax Outbreak of 2001, Case 19. UCLA School of Public Health, Los Angeles, CA. Available: http://www.ph.ucla.edu/epi/bioter/detect/antdetect_case19.html.
197. American Anthrax Outbreak of 2001, Case 21. UCLA School of Public Health, Los Angeles, CA. Available: http://www.ph.ucla.edu/epi/bioter/detect/antdetect_case21.html.
198. American Anthrax Outbreak of 2001, Case 2. UCLA School of Public Health, Los Angeles, CA. Available: http://www.ph.ucla.edu/epi/bioter/detect/antdetect_case2.html.
199. American Anthrax outbreak of 2001, Case 6. UCLA School of Public Health, Los Angeles, CA. Available: http://www.ph.ucla.edu/epi/bioter/detect/antdetect_case6.html.
200. American Anthrax Outbreak of 2001, Case 8. UCLA School of Public Health, Los Angeles, CA. Available: http://www.ph.ucla.edu/epi/bioter/detect/antdetect_case8.html.
201. American Anthrax outbreak of 2001, Case 9. UCLA School of Public Health, Los Angeles, CA. Available: http://www.ph.ucla.edu/epi/bioter/detect/antdetect_case9.html.

202. American Anthrax Outbreak of 2001, Case 10. UCLA School of Public Health, Los Angeles. Available: http://www.ph.ucla.edu/epi/bioter/detect/antdetect_case10.html.
203. American Anthrax Outbreak of 2001, Case 11. UCLA School of Public Health, Los Angeles, CA. Available: http://www.ph.ucla.edu/epi/bioter/detect/antdetect_case11.html.
204. American Anthrax Outbreak of 2001, Case 12. UCLA School of Public Health, Los Angeles, CA. Available: http://www.ph.ucla.edu/epi/bioter/detect/antdetect_case12.html.
205. American Anthrax Outbreak of 2001, Case 13. UCLA School of Public Health, Los Angeles, CA. Available: http://www.ph.ucla.edu/epi/bioter/detect/antdetect_case13.html.
206. American Anthrax Outbreak of 2001, Case 18. UCLA School of Public Health, Los Angeles, CA. Available: http://www.ph.ucla.edu/epi/bioter/detect/antdetect_case18.html.
207. American Anthrax Outbreak of 2001, Case 14. UCLA School of Public Health, Los Angeles, CA. Available: http://www.ph.ucla.edu/epi/bioter/detect/antdetect_case14.html.
208. American Anthrax Outbreak of 2001, Case 15. UCLA School of Public Health, Los Angeles, CA. Available: http://www.ph.ucla.edu/epi/bioter/detect/antdetect_case15.html.
209. American Anthrax Outbreak of 2001, Case 16. UCLA School of Public Health, Los Angeles, CA. Available: http://www.ph.ucla.edu/epi/bioter/detect/antdetect_case16.html.
210. American Anthrax Outbreak of 2001, Case 17. UCLA School of Public Health, Los Angeles, CA. Available: http://www.ph.ucla.edu/epi/bioter/detect/antdetect_case17.html.
211. American Anthrax outbreak of 2001, Case 20. UCLA School of Public Health, Los Angeles, CA. Available: http://www.ph.ucla.edu/epi/bioter/detect/antdetect_case20.html.
212. American Anthrax Outbreak of 2001, Case 23. UCLA School of Public Health, Los Angeles, CA. Available: http://www.ph.ucla.edu/epi/bioter/detect/antdetect_case23.html.
213. American Anthrax Outbreak of 2001, Case 22. UCLA School of Public Health, Los Angeles, CA. Available: http://www.ph.ucla.edu/epi/bioter/detect/antdetect_case22.html.
214. BBC News, November 22, 2001.
215. Heyman D (2002). Lessons from the anthrax attacks. Intelligence Resource Program. Available: http://www.fas.org/irp/threat/cbw/dtra02.pdf.
216. Mcintyre H J (2004). Project on Government Secrecy. Ref: 04-F-0521(A), DoD FOIA Pollicy Office.
217. Shoham D, and Jacobsen S T (2007). Technical intelligence in retrospect: the 2001 anthrax letters powder, *International J. of Intelligence and Counterintelligence*, 20(1):79–105, 2007.
218. Associated Press, Washington D.C. 11 March 2004, Powder Letters Hoaxes.
219. Miller J (2003). U.S. Has New Concerns about Anthrax Readiness. December 28.
220. Valdemoro T (2005). Cleaning of Anthrax Contents Continues. *Cox News Service-Herald Today*, July 7. Available: http://www.bradenton.com/mld/bradenton/news/local/12072828.htm.

221. Matsumoto G (2003). Anthrax powder: State of the art? *Science*. 302:1492–1497.
222. 2001 anthrax attacks—Wikipedia, the free encyclopedia, en.wikipedia.org/wiki/Website.
223. Allan Lengel, Little Progress In FBI Probe of Anthrax Attacks, Washington Post, September 16, 2005; A01.
224. Ricin Toxin from Castor Bean Plant. Cornell University, Ithaca, NY. Poisonous Plants Informational Database Available: http://www.ansci.cornell.edu/plants/toxicagents/ricin/ricin.html.
225. Klain G J, Jaeger J J (1990). *Castor Seed Poisoning In Humans*: *A Review*. Letterman Army Inst. of Research, San Francisco, CA.
226. Historical records from Army Office of Research and Development (OSRD), 1946, Agent W-Ricin.
227. Franke S (1976). *Bacterial, Animal and Plant Toxins as Combat Agents, Manual of Military Chemistry*, Vol. 2. Militaerverlag der DDR., Berlin, Germany, pp. 484, 485, 488–496.
228. Sanches M L, Russell C R, Randolph C L (1993). *Chemical Weapons Convention (CWC) Signatures Analysis: Technical Analysis*. Defense Nuclear Agency, Alexandria, VA, pp. B171–788.
229. Usama bin Laden. Available: www.nmhschool.org/tthornton/mehistorydatabase/afterseptembereleven.htm.
230. Frankel G (2003). *The Washington Post*. Jan. 8, Ricin Found in Afghomistan.
231. Gold D (2004). *Tower of Babble*. Crown Forum, New York, p. 129.
232. Seventh Arrest in Ricin Case. *BBC News Online*, 8 January 2003.
233. Norton-Taylor R, Hopkins N, Henley J (2003). Poison Suspect Trained at al-Qaida Camp. *The Guardian* (London), January 10.
234. Sergeyev V (2003). London Poisoners Came from Chechnya. *Gazeta* (Moscow), January 10.
235. Ivanov S (2003) France, UK Security Follow Trail of New Terrorist Structures with Chechen cell. *ITAR-TASS News Agency* (Moscow), January 13.
236. Gray C, Herbert I, Bennetto J (2003). Policeman Stabbed to Death in Terror Raid. *The Independent* (London), January 15.
237. Carter H, Ward D, Hopkins N (2003). Murder Suspect "is Senior Player" in Ricin Plot Network. *The Guardian* (London), January 16.
238. Chemical Terrorist Plot in Rome? Center for Nonproliferation Studies. Available: http://cns.miis.edu/pubs/week/020311.htm.
239. Anonymous (2003). Ricin found in Paris. *BBC News*, March 21.
240. Eggen D (2004). Letter with Ricin Vial Sent To White House, November Discovery Was Kept Quiet. *The Washington Post*, February 4, A:07.
241. Lengel A (2004). Ricin Investigation Still Wide Open. *The Washington Post*, April 6, B:01.
242. Barry R (2004). Rebels Tried to Make Ricin. *Associated Press*, Washington, D.C., June 10.
243. Woodward C (2005). Ricin Intended to be Produced *Associated Press/RedOrbit.com*, November 21.
244. Jar of Ricin Found in Tennessee Shed, Associated Press/San Diego Union-Tribune, June 2, 2006.
245. What Is Smallpox? Available: http://skin-care.health-cares.net/smallpox.php.

246. Smallpox: Destruction of variola virus stocks. World Health Organization. Available: http://www.who.int/gb/ebwha/pdf_files/WHA58/A58_10-en.pdf.
247. Jahrling P B, Hensley L E, Martinez M J, et al. (2004). Exploring the potential of variola virus infection of cynomolgus macaques as a model for human smallpox. *Proc. Natl. Acad. Sci. USA*. 101:15196–15200.337.
248. Rubins K H, Hensley L E, Jahrling P B, et al. (2004). The host response to smallpox: analysis of the gene expression program in peripheral blood cells in a nonhuman primate model. *Proc. Natl. Acad. Sci. USA*. 101:15190–15195.
249. WHO (2004). Abstracts summarizing recent research overseen by the WHO Advisory Committee on Variola Virus Research. Available: http://www.who.int/csr/disease/smallpox/abstracts2004/en/index1.html.
250. WHO (2003). Abstracts summarizing recent research over seen by the WHO Advisory Committee on Variola Virus Research. Available: http://www.who.int/csr/disease/smallpox/abstracts2003/en/index1.html.
251. Li Y, Kline R, Regnery R. (2002). Capillary-Electrophoresis Restriction Fragement Length Polymorphism: A New Method for Poxvirus Fingerprinting. Available: http://www.who.int/csr/disease/smallpox/capillary/en/index.html.
252. Huggins J W, Zwiers S H, Baker R O, et al. Cidofovir Treatment of Variola (Smallpox) in the Hemorrhagic Smallpox Primate Model and the IV Monkeypox Primate Model. Available: http://www.who.int/csr/disease/smallpox/cidofovirtreatment/en/index.html.
253. Rubins K, Whitney A, Jahrling P, et al. Genome-Wide Analysis of the Host Response to Variola Infection. Available: http://www.who.int/csr/disease/smallpox/genomewide/en/index.html.
254. Pulford D J, Damon I, Ulaeto D. New PCR Assays for Identification of Smallpox Virus. Available: http://www.who.int/csr/disease/smallpox/newpcr/en/index.html.
255. Jahrling P B, Hensley L, Huggins J, et al. Lethal Infection of Primates with Variola Virus as a Model for Human Smallpox. Available: http://www.who.int/csr/disease/smallpox/lethalinfection/en/index.html.
256. Esposito J J, Sammons S, Frace M, et al. Variola Virus Genomics. Available: http://www.who.int/csr/disease/smallpox/virusgenomics/en/index.html.
257. Boseley S, Borger J (2005). US Scientists Push for Go-Ahead to Genetically Modify Smallpox Virus. *The Guardian*, May 16.
258. Burgasov P (2001). Smallpox Field Tested By USSR *Courier*, November 24.
259. Anonynmous (2003). Soviet Smallpox Most Lethal *Vedomosty* (Moscow), March 27.
260. Smallpox: Facts and Information. Available: http://www.absoluteastronomy.com/encyclopedia/s/sm/smallpox.htm.
261. Hopkins D R, Ramses V (1980). Earliest Known Victim? *World Health*, p. 220.
262. Fornaciari G, Marcetti A (1986). Intact smallpox virus particles in an Italian mummy of 16th century. *The Lancet*. 2(8507):625.
263. Marennikova S S, Shelukhina E M, Zhukova O A, et al. (1990). Smallpox diagnosed 400 years later: Results of skin lesions examination of 16th century Italian mummy. *J. Hyg. Epidemiol. Microbiol. Immunol*. 34:227–231.
264. Orent W (1998). Escape from Moscow. *The Sciences*, pp. 29–31, May 22.
265. Lewin P K (1985). Mummy riddles unraveled. *Microsc. Soc. Can. Bull*. 12:4–8.
266. Arita I (1980). Can we stop smallpox vaccination? *World Health*. May:27–29.
267. Thomas, Gordon (2003). Ethnic Bomb to Kill Whites Only Tied to North Korea. *American Free Press*, October, 17.

268. Sombre Conclusions follow "Atlantic Storm". Radio Netherlands. Available: www2.rnw.nl/rnw/en/currentaffairs/region/internationalorganisations/ter050117.
269. National Security Archive (1970). The September 11th Source Books: National Security Archive Online Readers on Terrorism, Intelligence and the Next War. Volume III: BIOWAR: The Nixon Administration's Decision to End U.S. Biological Warfare Programs. Available: http://www.gwu.edu/~nsarchiv/NSAEBB/NSAEBB58/RNCBW22.pdf.
270. Beckett B (1983). *Weapons of Tomorrow*. Plenum Press, New York.
271. Leitenberg M (2001). Biological Weapons in the Twentieth Century: A Review and Analysis. Available: http://www.fas.org/bwc/papers/bw20th.htm.
272. Anonymous Times, Sept. 14, 1999, Soviet Bioagents Against Livestock.
273. Shoham D (2005). Image versus reality of Iranian chemical and biological weapons. *Internat. J. Intell. Counterintell*. 18(1):89–141.
274. Robinson R A. USDA's Preparation for Asian Soybean Rust. Available: http://www.gao.gov/new.items/d05668r.pdf.
275. Vargas R I (2002). Biosystems and Bioterrorism: Can Arthropods Be Used as Agents of Destruction. Entomological Society of America Meeting, Fort Lauderolale, FL.
276. Lindquist D A, Abusowa M, Hall MJ (1992). The New World screwworm fly in Libya: A review of its introduction and eradication. *Med. Vet. Entomol.* 6:2–8.
277. Windrem R (1999). U.S. to launch war on "agro-terror." *MSNBC.com*, September 22. Available: http://www.msnbc.com/news/314627.asp.
278. Treble A (2005). Agro-Terrorism. Available: http://cns.miis.edu/research/cbw/agromain.htm.
279. Stockholm International Peace Research Institute (1971). *The Problem of Chemical and Biological Warfare*, Vol. I. Humanities Press, New York, p. 224.
280. Treble A (2002). Chemical and Biological Weapons: Possession and Programs Past and Present. Available: http://cns.miis.edu/research/cbw/possess.htm#US.
281. Thorold P W (1953). Suspected malicious poisoning. *J. South African Vet. Med. Assoc.* 24:215–217.
282. Morton J F (1977). Poisonous and injurious higher plants and fungi. In: C C Tedeschi, W G Eckert, L G Tedeschi (eds.), *Forensic Medicine: A study in trauma and environment, Volume III: Environmental Hazards*. W.B. Saunders, Philadelphia, PA, p. 1504.
283. Verdcourt B, Trump E C (1969). *Common Poisonous Plants of East Africa*. Collins, London, p. 62.
284. Maloba W O (1993). Mau *Mau and Kenya: An Analysis of a Peasant Revolt*. Indiana University Press, Bloomington, IN.
285. Edgerton R B (1989). *Mau Mau: An African Crucible*. Free Press, New York.
286. Alibek K (1999). The Soviet Union's anti-agricultural biological weapons. *Annals N. Y. Acad. Sci*. 894:18–19.
287. Narayan Swamy M R (1994). *Tigers of Lanka, From Boys to Guerrillas*. Konark Publishers, Delhi.
288. Hellmann-Rajanayagam D (1994). *The Tamil Tigers: Armed Struggle for Identity*. Franz Steiner Verlag, Stuttgart.
289. O'Balance E (1989). *The Cyanide War: Tamil Insurrection in Sri Lanka 1973–1988*. Brassey's, Washington, D.C.
290. Gunaratna R (1987). *War and Peace in Sri Lanka*. Institute of Fundamental Studies, Sri Lanka, pp. 51, 52.

291. Duboudin T (1984). Australian Livestock Threatened. *The Times* (London), January 21, p. 5.
292. Duboudin T (1984). Murderer in Court Over Virus Threat. *The Times* (London), February 22, p. 5.
293. Root R S (1991). Infectious terrorism. *Atlantic Monthly*. May, p. 4450.
294. Office International des Eepizootics, Headquarter, Paris (2002). Foot and mouth disease in the United Kingdom. 2001 Report to the OIE Part I: Great Britain Department for Environment, Food & Rural Affairs 14 January, 2002.
295. Anonymous (2005). Putative Agroterrorism Bioagents in Afghanistan Associated Press, 17 July, 2005.
296. Narcoterrorism. Available: http://www.sciencedaily.com/encyclopedia/narcoterrorism.
297. What is Narcoterrorism? Terrorism: Questions & Answers. Available: http://cfrterrorism.org/terrorism/narcoterrorism_print.html.
298. Narco-Terrorismus. Zentraler Informatikdienst- Universität Wien. Available: http://gerda.univie.ac.at/ie/ws03/fischer/wiki/index.php/Narco-Terrorismus.
299. Department of the Navy (1998). Legal Review of Oleoresin Capsicum (OC) Pepper Spray, (Ser 103/353). Office of the Judge Advocate General, Alexandria, Virginia.
300. Common and New Types of Biological Weapons. Available: whalonlab.msu.edu/2003_EC_Projects/Biological_Weapons/ISB%20Extra%20Credit%20Project/Common%20Types.html.
301. Black J L (2003). Genome projects and gene therapy: Gateways to next generation biological weapons. *Military Med*. 168(11):864–871.
302. Wheelis A M (2004). Will the new biology lead to new weapons? *Arms Control Today*. 34(6):6–13.
303. van Aken J, Hammond E (2003). Genetic engineering and biological weapons: New technologies, desires and threats from biological research. *Euro. Molec. Biol. Org. Rep*. 4 (suppl 1):S57–S60.
304. Quille G (2001). The revolution in military affairs debate and non-lethal weapons, medicine, conflicts and survival. *Internat. Security Informat. Serv. London, UK*. 17(3):207–220.
305. Shoham D, Wolfson Z (2004). The Russian biological weapons program: Vanished or disappeared? *Crit. Rev. Microbiol*. 30(4):241–261.
306. Former KGB spy, living in New Zealand, soon to publish a book on Russian biological warfare. Available: http://www.freerepublic.com/focus/f-news/1347344/posts.
307. Oppenheimer A (2005). Nanotechnology paves way for new weapons. Jane's Chem-Bio Web. www.janes.com/defence/news/jcbw.
308. Joby Warrick, Engineered Microbes Pose New Bioterror Threat, Washington Post, July 31, 2006.
309. King S R (1992). Medicines that changed the world. Available: http://www.accessexcellence.org/RC/Ethnobotany/page4.html. Ethnobotany Index.
310. Ricin toxin from castor bean plant. Available: http://7plus7.com/MedicalNews/News/Ricin.htm.
311. Lord J M, Roberts L M, Robertus J D (1994). Ricin: Structure, mode of action and some current applications. *FASEB J*. 8(2):201–208.
312. Shukla H, Sharma S (2005). Clostridium botulinum: A bug with beauty and weapon. *Crit. Rev. Microbiol*. 31(1):11–18.
313. UN Security Council Resolution 1540 (2004), United Nations, New York City, April 28.

314. U.N. Releases Counterterrorism Strategy, Associated Press, Sep. 4, 2006.
315. Tigner B (2004). European group urges bioterror defense. *Defense News*, June 29.
316. Bioterrorism Reporting Group Report (2005). Countering bioterrorism: How can the EU and US work together? New Defence Agenda *NDA* July, 18.
317. Trevelyan M (2005). *Reuters*, March 1. Outcomes of a Biotterorism Attack.
318. Ceran F (2005). *Associated Press*, March 1, 2005. The Threat of Bioterrorism.
319. MI5 warns UK labs of threat from al-Qaeda, London Daily Mail, January 25, 2007.
320. U.K. Increases Biological Security, Tendler/McGrory, The Times, January 25, 2007.
321. Rozen K (2005). NIS 500,000 to Find Knesset Vulnerabilities to Non-Conventional Weapons. *Israel National News*, September 22.
322. Anonymous (2003). Iran Red Crescent holds first conference on campaign against bioterrorism, May 29. Available: http://www.fbis.gov. FBIS doc IAP 20030529000082.
323. Anonymous (2005). Seminar on Bioterrorism Opens in Saudi Capital. *Financial Times Information*, March 7.
324. Biodefense for the 21st century. Available: http://www.whitehouse.gov/homeland/20040430.html.
325. Remarks by the President at the Signing of S.15-Project BioShield Act of 2004, The White House, July 21, 2004.
326. Fiorill J (2005). Experts paint dire picture of bioterrorism threat. *Global Security Newswire*, July 13.
327. Fiorill J (2005). U.S. plans to defend against engineered bioattack. *Global Security Newswire*, June 15.
328. Bioterrorism and the National Response Plan 2005. Available: apha.confex.com/apha/133am/techprogram/paper_110332.htm.
329. Schuler A (2005). Billions for Biodefense: Federal Agency Biodefense Budgeting, FY2005-FY2006. *Biosecur. Bioterror.* 3(2):91–101.
330. Ruppe D (2005). *Global Security Newswire*, July 14, 2005. Bush Budget Holds Steady for Civiliam Biodefences.
331. Ember L (2005). Manhattan experiments track gases for terror response biomonitoring. *Chem. Engin. News.* 83(33):10.
332. Schwellenbach N (2005). Biodefense: A Plague of Researchers Bulletin of the Atomic Scientists, May–June. pp. 8–10.
333. National Institute of Allergy and Infectious Diseases release, June 1, 2005. US to Complete Bioterrorism Research Network.
334. Anonymous. *Associated Press*, 8 September 2005. New Biocontainment Lab.
335. FBI Wants Biodefense Work Secret says Newark, New Jersey Lab. The Sunshine Project—Biosafety Bites (v.2) p. 17 September 20, 2006.
336. Jennifer Lebovich, BioWatch Air Samplers Used to Detect Deadly Germs, jlebovich@MiamiHerald.com Nov. 01, 2006.
337. Hall/Jackson, Bush Asserts Mail-Opening Authority, USA Today, Jan. 5, 2007.
338. FDA's National Center for Toxicological Research Publishes Study on Distinguishing Potential Hoax Materials from Bioterror Agents, FDA New, PO6-109, August 3, 2006.
339. Nanotechnology Offers Fast Virus Detection, United Press International, RedOrbit.com, Nov. 15, 2006.
340. Alison Walker-Baird, U.S. Researchers Develop Tool Against Agroterrorism, The Frederick News-Post, October 31, 2006.

341. Network Would Help Detect Bioterrorism, The Richmond Times Dispatch, May 23, 2006.
342. Anonymous *Biotech Week*, June 8, 2005. Post-Attack Anthrax Medical Measures.
343. Barrett G (2005). *Baltimore Sun*, June 10, 2005. Stockpiling Bioterror Countermeasures.
344. Francis D. Researchers suggest risk-based decision making model to combat bioterrorism attacks. *Global Security Newswire*. June 3, 2005.
345. Thompson K M, Armstrong R E, Thompson D F (2005). Bayes, Bugs, and Bioterrorists. National Defense University, Washington, D.C. Available: http://www.ndu.edu/ctnsp/Def_Tech/DTP14%20Bayes%20Bugs%20Bioterrorists.pdf.
346. Harrison P (2005). Smallpox Vaccine to be Acquired *Reuters*, August 16, 2005.
347. *Daily Yomiuri*, July 12, 2005. Japan Intends to Stockpile Smallpox Vaccine.
348. Exercise Global Mercury: Post Exercise Report. Available: www.dh.gov.uk . . . /ChiefMedicalOfficer/Features/FeaturesArticle/fs/en.
349. Madjid M, Lillibridge S, Parsa M, et al. (2003). Influenza as a bioweapon. *J. R. Soc. Med*. 96:345, 346.
350. Neergaard L (2005). U.S. working on plan for flu pandemic. *Associated Press*, Washington D.C., October 8.
351. Anonymous (2005). Avian Flu Resistant to Tamiflu. CNN (*Reuters*), September 30.
352. Zeleny J (2005). Pathogens Transferred from Formen USSR to USA. *Chicago Tribune*, September 3.
353. Vorobiev A (1996). Countering chemical/biological terrorism in the former Soviet Union: The need for cooperative efforts. *Politics Life Sci*. 15:233–235.
354. Vector Outline (2003). Available: http://www.vector.nsc.ru/art1-r.htm.
355. Available: www.state.gov/documents/organization/28781.
356. Gulobev E (2003). New Russian-American Collaboration Against Bioterrorism: Moscow's International Science and Technology Center (ISTC) participating, Pravda, 17 September 2003. Pravda Online English Version.
357. Ruppe D (2005). U.S. helps secure former soviet antiplague sites. *Global Security Newswire*, November 23.
358. David Francis, U.S. Program to Secure Former Soviet Pathogens Grows, Global Security Newswire, March 29, 2006.
359. Fiorill J (2005). Top U.S. disease fighters warn of new engineered pathogens but call bioweapons doomsday unlikely. *Global Security Newswire*, July 29.
360. Atlas R M, Reppy J (2005). Globalizing biosecurity, *Biosecurity and Bioterrorism*. 3(1):51–60.
361. Casadevall A (2002). Antibodies for defense against biological attack. *Nat. Biotechnol*. 20:114.
362. Anonymous *Associated Press* (2005). Antibiotics Less Useful March 3.
363. Rainey G J, Young J A (2004). Antitoxins: Novel strategies to target agents of bioterrorism. *Nat. Rev. Microbiol*. 2(9):721–726.
364. AVI BioPharma Receives Notice of Allowance for Key Patent Application Covering NEUGENE Antisense Technology for RNA Viruses, BioExchange Newsroom. Available: http://www.bioexchange.com/news/news_page.cfm?id=20709.
365. AVI BioPharma Info (2005). Promising new technology may block biological agent effects, *GSN*, October 18. Available: http://www.mmrs.fema.gov/news/publichealth/2005/oct/nph2005–10–18a.aspx.

366. Basha, Saleem et al, Polyvalent inhibitors of anthrax toxin that target host receptors, PNAS, September 5, 2006 vol. 103, no. 36, 13509–13513.
367. Elias P (2005). *Associated Press/Journal-Gazette*, August 19. Super Pathogens.
368. Fiorill J (2005). *GSN*, July 13. Experts Paint Dire Picture of Bioterrorism Threat.
369. Biological attack on West "simply a question of time": US official (AFP) Jan 17, 2006.
370. Edith Lederer, Annan Calls for Bioterror Defense Strategies, Associated Press/ phillyBrubs.com, May 3, 2006.
371. Anonymous (2005). Interpol sounds bioterrorism alarm. *BBC*, February 23. Available: http://news.bbc.co.uk/go/pr/fr/-/1/hi/world/europe/4289485.stm.

# BIOTECHNOLOGY-DERIVED DRUG PRODUCT DEVELOPMENT

Stephen M. Carl,[1] David J. Lindley,[1] Gregory T. Knipp,[1] Kenneth R. Morris,[1] Erin Oliver,[2] Gerald W. Becker,[3] and Robert D. Arnold[4]

[1]*Purdue University, West Lafayette, Indiana*
[2]*Rutgers, The State University of New Jersey, Piscataway, New Jersey*
[3]*SSCI, West Lafayette, Indiana*
[4]*The University of Georgia, Athens, Georgia*

**Contents**

1 Introduction
2 Formulation Assessment
  2.1 Route of Administration and Dosage
  2.2 Pharmacokinetic Implications to Dosage Form Design
  2.3 Controlled-Release Delivery Systems
3 Analytical Method Development
  3.1 Traditional and Biophysical Analytical Methodologies
  3.2 Stability-Indicating Methodologies
  3.3 Method Validation and Transfer
4 Formulation Development
  4.1 Processing Materials and Equipment
  4.2 Container Closure Systems
  4.3 Sterility Assurance
  4.4 Excipient Selection
5 Drug Product Stability
  5.1 Defining Drug Product Storage Conditions
  5.2 Mechanisms of Protein and Peptide Degradation
  5.3 Photostability
  5.4 Mechanical Stress
  5.5 Freeze–Thaw Considerations and Cryopreservation
  5.6 Use Studies
  5.7 Container Closure Integrity and Microbiological Assessment
  5.8 Data Interpretation and Assessment

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

6 Quality by Design and Scale-Up
6.1 Unit Operations
6.2 Bioburden Considerations
6.3 Scale-Up and Process Changes
7 Concluding Remarks
References

## 1 INTRODUCTION

Although the origins of the first biological and/or protein therapeutics can be traced to insulin in 1922, the first biotechnology-derived pharmaceutical drug product approved in the United States was Humulin in 1982. In the early stages of pharmaceutical biotechnology, companies that specialized primarily in the development of biologicals were the greatest source of research and development in this area. Recent advances in molecular and cellular biological techniques and the potential clinical benefits of biotechnology drug products have led to a substantial increase in their development by biotechnology and traditional pharmaceutical companies. In terms of pharmaceuticals, the International Conference on Harmonization (ICH) loosely defines biotechnology-derived products with biological origin products as those that are "well-characterized proteins and polypeptides, their derivatives and products of which they are components, and which are isolated from tissues, body fluids, cell cultures, or produced using rDNA technology" [1]. In practical terms, biological and biotechnology-derived pharmaceutical agents encompass a number of therapeutic classes, including cytokines, erythropoietins, plasminogen activators, blood plasma factors, growth hormones and growth factors, insulins, monoclonal antibodies, and vaccines [1]. Additionally, short interfering and short hairpin ribonucleic acids (siRNA, shRNA) and antisense oligonucleotide therapies are generally characterized as biotechnology-derived products.

According to the biotechnology advocacy group, The Biotechnology Industry Organization (BIO), pharmaceutical-based biotechnology represents over a $30 billion dollar a year industry and is directly responsible for the production of greater than 160 drug therapeutics and vaccines [2]. Furthermore, there are more than 370 biotechnology-derived drug products and vaccines currently in clinical trials around the world, targeting more than 200 diseases, including various cancers, Alzheimer's disease, heart disease, diabetes, multiple sclerosis, acquired immunodeficiency syndrome (AIDS), and arthritis. While the clinical value of these products is well recognized, a far greater number of biotechnology-derived drug products with therapeutic potential for life-altering diseases have failed in development.

As the appreciation of the clinical importance and commercial potential for biological products grows, new challenges are arising based on the many technological limitations related to the development and marketing of these complex agents. Additionally, the intellectual property protection of an associated agent might not

provide a sufficient window to market and regain the costs associated with the discovery, research, development, and scale-up of these products. Therefore, to properly estimate the potential return on investment, a clear assessment of potential therapeutic advantages and disadvantages, such as the technological limitations in the rigorous characterization required of these complex therapeutic agents to gain Food and Drug Administration (FDA) approval, is needed prior to initiating research. Clearly, research focused on developing methodologies to minimize these technological limitations is needed. In doing so we hypothesize the attrition rate can be reduced and the number of companies engaged in the development of biotechnology-derived products and diversity of products will continue to expand.

Technological limitations have limited the development of follow-on, or generic biopharmaceutical products that have lost patent protection. In fact, the potential pitfalls associated with developing these compounds are so diverse that regulatory guidance concerning follow-on biologics is relatively obscure and essentially notes that products will be assessed on a case-by-case basis. The reader is encouraged to see Chapter Regulatory Considerations in Approval of Follow-On Protein Drug Products for a more detailed discussion concerning regulatory perspectives pertaining to follow-on biologics.

Many of the greatest challenges in producing biotechnology-derived pharmaceuticals are encountered in evaluating and validating the chemical and physical nature of the host expression system and the subsequent active pharmaceutical ingredient (API) as they are transferred from discovery through to the development and marketing stages. Although this area is currently a hotbed of research and is progressing steadily, limitations in analytical technologies are responsible for a high degree of attrition of these compounds. The problem is primarily associated with limited resolution of the analytical technologies utilized for product characterization. For example, without the ability to resolve small differences in secondary or tertiary structure, linking changes to product performance or clinical response is impossible. The biological activity of traditional small molecules is related directly to their structure and can be determined readily by nuclear magnetic resonance (NMR), X-Ray crystallography (X-ray), mass spectrometry (MS), and/or a combination of other spectroscopic techniques. However, methodologies utilized for characterizing biological agents are limited by resolution and reproducibility. For instance, circular dichroism (CD) is generally considered a good method to determine secondary structural elements and provides some information on the folding patterns (tertiary structure) of proteins. However, CD suffers from several limitations, including a lower resolution that is due in part to the sequence libraries used to deconvolute the spectra. To improve the reliability of determining the secondary and tertiary structural elements, these databases need to be developed further. An additional example is the utility of two-dimensional NMR (2D-NMR) for structural determination. While combining homonuclear and heteronuclear experimental techniques can prove useful in structural determination, there are challenges in that 2D-NMR for a protein could potentially generate thousands of signals. The ability to assign specific signals to each atom and their respective interactions is a daunting task. Resolution between the different amino acids in the primary sequence and their positioning in the covalent and folded structures become limited with increasing molecular weight. Higher dimensional techniques can be used to improve resolution; however, the resolution of these methods remains limited as the number of amino acids is increased.

Understanding the limitations of the analytical methodologies utilized for product characterization has led to the development of new experimental techniques as well as the refined application of well-established techniques to this emerging field. Only through application of a number of complementary techniques will development scientists be able to accurately characterize and develop clinically useful products. Unfortunately, much of the technology is still in its infancy and does not allow for a more in-depth understanding of the subtleties of peptide and protein processing and manufacturing. For instance, many of the analytical techniques utilized for characterization will evaluate changes to product conformation on the macroscopic level, such as potential denaturation or changes in folding, as observed with CD. However, these techniques do not afford the resolution to identify subtle changes in conformation that may either induce chemical or physical instabilities or unmask antigenic epitopes.

Further limiting successful product development is a lack of basic understanding as to critical manufacturing processes that have the potential to affect the structural integrity and activity of biopharmaceuticals. As with traditional small molecules, stresses associated with the different unit operations may affect biopharmaceutical products differently. In contrast to traditional small molecules, there is considerable difficulty in identifying potentially adverse affects, if any, that a particular unit operation may have on the clinically critical structural elements of a drug. Considering that many proteins exhibit a greater potential for degradation from shear stress, it is particularly important to assess any negative effects of mixing, transport through tubing, filtration, and filling operations. Essentially all unit operations for a given manufacturing process could create enough shear stress to induce minor structural changes that could lead to product failure. The difficulty is establishing what degree of change will have an impact on the stability, bioactivity, or immunogenic potential of the compound. Unfortunately, unless exhaustive formulation development studies are conducted, coupled with a comprehensive spectrum of analytical methodologies, these effects may not be readily evident until after scale-up of the manufacturing process or, worse yet, in the clinical setting. Moreover, modeling shear and stress using fluid dynamic structurally diverse molecules is a foreboding task. Extending these models to validate process analytical technologies (PAT) and incorporate critical quality by design (QbD) elements in the development process for a collection of biopharmaceuticals would be largely hindered by the daunting nature of the task at hand.

The use of biological systems to produce these agents results in additional variability. Slight changes in nutrient profile could affect growth patterns and protein expression of cultured cells. Furthermore, microbial contamination in the form of viruses, bacteria, fungi, and mycoplasma can be introduced during establishment of cell lines, cell culture/fermentation, capture and downstream processing steps, formulation and filling operations, or drug delivery [3]. Therefore, establishing the useful life span of purification media and separation columns remains a critical issue for consistently producing intermediates and final products that meet the defined quality and safety attributes of the product [4]. In short, understanding the proper processability and manufacturing controls needed has been a major hurdle that has kept broader development of biopharmaceutical products relatively limited.

Notwithstanding the many technological hurdles to successfully develop a pharmaceutically active biotechnology product, they offer many advantages in terms of

therapeutic potency, specificity, and target design (not generally limited to a particular class or series of compounds). This is an iterative approach, whereby every new approved compound, new lessons, and applications to ensure successful product development are realized, thereby adding to our knowledge base and facilitating the development of future products. This chapter will discuss some of the fundamental issues associated with successful biopharmaceutical drug product development and aims to provide an understanding of the subtleties associated with their characterization, processing, and manufacturing.

## 2 FORMULATION ASSESSMENT

In order to select the most appropriate formulation and route of administration for a drug product, one must first assess the properties of the API, the proposed therapeutic indication, and the requirements/limitations of the drug and the target patient population. Development teams are interdisciplinary comprised of individuals with broad expertise, for example, chemistry, biochemistry, bioengineering, and pharmaceutics, that can provide insight into the challenges facing successful product development. As such, knowledge gained through refinement of the API manufacturing process through to lead optimization is vital to providing an initial starting point for success. Information acquired, for example, in the way of analytical development and API characterization, during drug discovery or early preclinical development that can be applied to final drug product development may contribute to shorter development times of successful products.

The host system utilized for API production is critical to the production of the final product and will determine the basic and higher order physicochemical characteristics of the drug. Typically biopharmaceuticals are manufactured in *Escherichia coli* as prokaryotic and yeast and Chinese hamster ovary (CHO) cells as eukaryotic expression systems [5]. While general procedures for growth condition optimization and processing and purification paradigms have emerged, differences in posttranslational modifications and host–system related impurities can exist even with relatively minor processing changes within a single production cell line [5]. Such changes have the potential to alter the biopharmaceutical properties of the active compound, its bioactivity, and its potential to elicit adverse events such as immunogenic reactions. These properties will be a common theme as they could potentially play a major role in both analytical and formulation development activities.

During the process of lead optimization, characterization work is performed that would include a number of parameters that are critical to formulation and analytical development scientists. The following information is a minimalist look at what information should be available to support product development scientists:

- Color
- Particle size and morphology (for solid isolates)
- Thermoanalytical profile (e.g., $T_g$ for lyophiles)
- Hygroscopicity
- Solubility with respect to pH
- Apparent solution pH

- Number and p$K_a$ of ionizable groups
- Amino acid sequence
- Secondary and tertiary structural characteristics
- Some stability parameters with respect to
  - pH
  - Temperature
  - Humidity
  - Light
  - Mechanical stress
  - Oxygen sensitivity
- Impurity profile
  - Misfolded/misaligned active
  - Potential isoforms
  - Expression system impurities
- Potency [median inhibitory concentration ($IC_{50}$)]
- Animal Pharmacokinetic/Pharmacodynamic (PK/PD) and $T_{ox}$ profiles

All of the above information will prove invaluable in determining the potential methods for rational drug delivery. Particular attention should be paid to the relative hygroscopicity of the API, of course, any stability information, as well as the impurity profile and ADMET (absorption, distribution, metabolism, excretion, and toxicity) information. In short, the more information that is available when development activities are initiated, the easier it is to avoid common pitfalls and make development decisions more rationally.

### 2.1 Route of Administration and Dosage

Biologics are traditionally very potent molecules that may require only picomolar blood concentrations to elicit a therapeutic effect. Given that the amount of drug required per dosage will be commensurate with the relative potency of the molecule, small concentrations are generally required for any unit dose. Biopharmaceuticals typically have large molecular weights relative to conventional pharmaceutical agents, which may be increased further by posttranslational modifications. The pharmacokinetics (ADMET) of biotechnology products have been reviewed elsewhere [6], but generally they have short circulating half-lives [7]. As such, biological products are most often delivered parenterally and formulated as solutions, suspensions, or lyophilized products for reconstitution [8, 9]. However, one must first ascertain the potential physiological barriers to drug delivery and efficacy before assessing potential routes of administration. These barriers may include actual physical barriers, such as a cell membrane, that could restrict the drug from reaching its site of action or chemical barriers, including pH or enzymatic degradation. Based on current drug delivery approaches, the proteinaceous nature of biological products limits their peroral delivery due to their susceptibility to proteases and peptidases present in the gastrointestinal tract as well as size limitations for permeating through absorptive enterocytes [10].

Difficulties in peroral delivery have stimulated researchers to explore alternate delivery mechanisms for biologics, such as through the lungs or nasal mucosa [11, 12]. Further, advances in technology and our understanding of the mechanisms limiting oral delivery of biotechnology products have led to innovative drug delivery approaches to achieve sufficient oral bioavailability. However, no viable products have successfully reached the market [13]. As a result of the technological limitations inherent in biopharmaceutical delivery, these compounds are largely delivered parenterally through an injection or implant.

When assessing the potential routes of administration, one must consider the physicochemical properties of the drug, its ADMET properties, the therapeutic indication, and the patient population, some of which are discussed below. Table 1 provides a list of some of those factors that must be addressed when determining the most favorable route of administration and the subsequent formulation for delivery. Ideally the route of administration and subsequent formulation will be optimized after identifying critical design parameters to satisfy the needs of patients and health care professionals alike while maintaining the safety and efficacy of the product.

Parenteral administration is the primary route of delivering biopharmaceutical agents (e.g., insulin); however, issues associated with patient compliance with administration of short-acting molecules are a challenge. Yet, the risk-to-benefit ratio must be weighed when determining such fundamental characteristics of the final dosage form. For instance, a number of biopharmaceutical compounds are administered subcutaneously, but this route of parenteral administration exhibits the highest potential for immunogenic adverse events due to the presence of Langerhans cells [14]. A compound's immunogenic potential is related to a host of factors, both

**TABLE 1 Factors That Determine Route of Administration**

| |
|---|
| Site of action |
| Therapeutic indication |
| Dosage |
| Potency/biological activity |
| Pharmacokinetic profile |
| Absorption time from tissue vs. IV |
| Circulating half-life |
| Distribution and elimination kinetics |
| Toxicological profile |
| Immunogenic potential |
| Patient population characteristics |
| Disease state |
| Pathophysiology |
| Age |
| Pharmacodynamic profile |
| Onset and duration of action |
| Required clinical effect |
| Formulation considerations |
| Stability |
| Impurity profile |

patient and treatment related; however, if an alternate, potentially safer route of administration is available, it may be prudent to consider it. Other factors, such as the frequency of dosing (especially into an immune organ such as the skin) and the duration of treatment, can also dramatically increase the potential for immunogenic reactions [14]. Many of the factors that contribute to the immunogenic potential of biopharmaceuticals, such as impurities, degradation products, and native antigenic epitopes, can be mitigated through altering the physicochemical properties of the drug (e.g., pegylation [15, 16], acylation [17, 18], increased glycosylation to mask epitopes [19]) or changing the characteristics of the formulation [20, 21]. In reality, the pharmaceutical industry has done a good job of recognizing the potential implications of immunogenic reactions and readily embraced technologies that can either mask or eliminate potential antigenic epitopes. However, additional research is needed to further identify and remove immunogenic epitopes.

### 2.2 Pharmacokinetic Implications to Dosage Form Design

Biological agents are generally eliminated by metabolism into di- and tripeptides, amino acids, and smaller components for subsequent absorption as nutrients or clearance by the kidney, liver, or other routes. Renal elimination of peptides and proteins occur primarily via three distinct mechanisms. The first involves the glomerular filtration of low-molecular-weight proteins followed by reabsorption into endocytic vesicles in the proximal tubule and subsequent hydroysis into small peptide fragments and amino acids [22]. Interleukin 11 (IL-11) [23], IL-2 [24], insulin [25], and growth hormone [26] have been shown to be eliminated by this method. The second involves hydrolysis of the compound at the brush border of the lumen and subsequent reabsorption of the resulting metabolites [6]. This route of elimination applies to small linear peptides such as angiotensin I and II, bradykinin, glucagons, and leutinizing hormone releasing hormone (LHRH) [6, 27, 28]. The third route of renal elimination involves peritubular extraction from postglomerular capillaries and intracellular metabolism [6]. Hepatic elimination may also play a major role in the metabolism of peptides and proteins; however, reticuloendothelial elimination is by far the primary elimination route for large macromolecular compounds [29].

Biopharmaceutical drug products are subject to the same principles of pharmacokinetics and exposure/response correlations as conventional small molecules [6]. However, these products are subject to numerous pitfalls due to their similarity to nutrients and endogenous proteins and the evolutionary mechanisms to break them down or prevent absorption. The types of pharmacokinetic-related problems that a biotechnology drug development team may encounter range from lack of specificity and sensitivity of bioanalytical assays to low bioavailability and rapid drug elimination from the system [6]. For example, most peptides have hormone activity and usually short elimination half-lives which can be desirable for close regulation of their endogenous levels and function. On the other hand, some proteins such as albumin or antibodies have half-lives of several days and formulation strategies must be designed to account for these extended elimination times [6]. For example, the reported terminal half-life for SB209763, a humanized monoclonal antibody against respiratory syncytial virus, was reported as 22–50 days [30]. Furthermore, some peptide and protein products that persist in the bloodstream exhibit the potential for idiosyncratic adverse affects as well as increased immunogenic poten-

tial. Therefore, the indication and formulation strategy can prove crucial design parameters simply based on clearance mechanisms.

### 2.3 Controlled-Release Delivery Systems

Given that the majority of biopharmaceutical products are indicated for chronic conditions and may require repeated administrations, products may be amenable to controlled-release drug delivery systems. Examples include Lupron Depot (leuprolide acetate), which is delivered subcutaneously in microspheres [31], and Viadur, which is implanted subcutaneously [32]. Various peptide/protein controlled delivery systems have been reviewed recently by Degim and Celebi and include biodegradable and nondegradable microspheres, microcapsules, nanocapsules, injectable implants, diffusion-controlled hydrogels and other hydrophilic systems, microemulsions and multiple emulsions, and the use of iontophoresis or electroporation [33]. These systems offer specific advantages over traditional delivery mechanisms when the drug is highly potent and if prolonged administration greater than one week is required [5, 33]. However, each of these systems has its own unique processing and manufacturing hurdles that must be addressed on a case-by-case basis. These factors, coupled with the difficulties of maintaining product stability, limit the widespread application of these technologies. However, the introduction of postapproval extended-release formulations may also provide the innovator company extended patent/commercial utility life and, as such, remains a viable option for postmarketing development. A current example of this is observed in the development of a long-acting release formulation of Amylin and Eli Lilly's co-marketed Byetta product.

## 3 ANALYTICAL METHOD DEVELOPMENT

The physical and chemical characterization of any pharmaceutical product is only as reliable as the quality of the analytical methodologies utilized to assess it. Without question, the role of analytical services to the overall drug product development process is invaluable. Good analytical testing with proper controls could mean the difference between a marketable product and one that is eliminated from development. Analytical methodologies intended for characterization and/or assessment of marketed pharmaceutical products must be relevant, validatable, and transferable to manufacturing/quality assurance laboratories.

### 3.1 Traditional and Biophysical Analytical Methodologies

Typically, there are a handful of traditional analytical methodologies that are utilized to assess the physical, chemical, and microbiological attributes of small-molecule pharmaceutical products. While many of these testing paradigms can still be utilized to assess biopharmaceuticals, these molecules require additional biophysical, microbiological, and immunogenic characterization as well. In brief, analytical methodologies should evaluate the purity and bioactivity of the product and must also be suitable to assess potential contaminants from expression systems as well as different isoforms and degradation products of the active. Biophysical

methodologies allow for assessment of the structural elements of the product with respect to its activity. Such assessments include structural elements, such as the folding of the molecule, and also encompass potential posttranslational modifications and their impact on structure. A list of typical analytical parameters and methodologies utilized to assess those parameters can be found in Table 2.

The impact of a molecule's biophysical characteristics on its clinical efficacy should be readily quantifiable. With respect to rational drug design, it is also extremely important to minimize external factors that may influence the formation of any adverse response. One such factor is the presence of degradation products and drug-related impurities that may be responsible for an immune response. One such industrial example is granulocyte-macrophage colony-stimulating factor [GM-CSF, or Leukine (sargramostim), by Berlex Co.], which is produced as a recombinant protein synthesized and purified from a yeast culture, *Saccharomyces cerevisiae*. As expected, the expression system has an impact on the final product: sargramostim, manufactured from *S. cerevisiae*, yields an O-glycosylated protein, while molgramostim (Leucomax), synthesized using an *E. coli* expression system, is nonglycosylated [34]. The *E. coli*–derived product exhibited a higher incidence of adverse reactions in clinical trials and never made it to the market. With respect to the drug product, the immunogenic reactions included [34, 35]:

**TABLE 2 Analytical Methodologies and Their Utility for API and Drug Product Characterization**

| Parameter Assessed | Methodologies | Utility |
|---|---|---|
| Appearance | Visual appearance, colorimetric assays, turbidity | Simple determination of physical stability, i.e., are there particles in solution, is the solution the correct color/turbidity? Is the container closure system seemingly intact? |
| Purity, degradation products and related substances | GPC/SEC-HPLC, RP-HPLC, gel electrophoresis, immunoassays, IEF, MS, CD, CE | Gives a general idea of the relative purity of the API and the drug product. Are there impurities related to the expression system? Are there alternate API isoforms present? Can degradation products be distinguished from the active component(s)? |
| Molecular weight determination | GPC/SEC-HPLC, gel electrophoresis, multiangle laser light scattering (MALLS), laser diffraction | Is the product a single molecular weight or polydisperse? Is the molecular weight dependent on posttranslational modifications? |
| Potency | Biological activity (direct or indirect) | Does the compound have reproducible in vitro activity and can this be correlated to in vivo? |
| pH | Potentiometric assays | Is the product pH labile or do pH changes affect potency is such ways that are not evident in other assays, i.e., minimal degradation and/or unfolding? |

**TABLE 2** ***Continued***

| Parameter Assessed | Methodologies | Utility |
|---|---|---|
| Primary structural elements | Protein sequencing, *N*-term degradation (Edman degradation), peptide mapping, amino acid composition, 2D-NMR | Verifies primary amino acid sequence and gives preliminary insight into activity. |
| Secondary structural elements | CD, 2D-NMR, in silico modeling from AA sequence | Secondary structural elements result from the primary sequence and help define the overall conformation (3D folding) of the compound. |
| Tertiary structural elements | Disulfide content/position, CD | Determines correct folding and overall integrity of the 3D product. Qualitative determination for denaturation potential. Also correlates to immunogenic potential. |
| Agglomeration/ aggregation | Subvisual and visual Particle size analysis, immunogenicity | Indicator of physical instability. Also gives an indication of immunogenic potential. |
| Carbohydrate analysis | RP-HPLC, gel electrophoresis, AE-HPLC, CE, MALDI-MS, ES-MS, enzyme arrays | Ensures proper posttranslational modifications and carbohydrate content. |
| Water content (lyophilized products) | Karl Fischer, TGA, NIR | Indicator of hydrolytic potential and process efficiency. |
| Immunogenic potential | Surface plasmon resonance, ELISA, immunoprecipitation | Methodologies generally only give positive/negative indicators of immunogenic potential. In vitro methodologies do not always correlate to in vivo. |
| Sterility | Membrane filtration | Indicator of microbial contaminants from manufacturing operations. |
| Bacterial endotoxins | Limulus amebocyte lysate (LAL) | Gives an idea of processing contaminants and potentially host organism contaminants. |
| Container closure integrity | Dye immersion, NIR, microbial ingress/sterility | Demonstrates viability of container closure system over the life of the product. |

*Abbreviations*: gel permeation chromatography (GPC), size exclusion chromatography (SEC), high-performance, or high-pressure, liquid chromatography (HPLC), reverse phase (RP), isoelectric focusing (IEF), mass spectrometry (MS), circular dichroism (CD), capillary electrophoresis (CE), nuclear magnetic resonance (NMR), anion exchange (AE), matrix-assisted laser desorption ionization (MALDI), electrospray ionization (ES), thermogravimetric analysis (TGA), near infrared (NIR), enzyme-linked immunosorbent assay (ELISA)

1. Formation of antibodies which bind and neutralize the GM-CSF
2. Formation of antibodies which bind but do not affect the efficacy of GM-CSF
3. Antibody formation against proteins not related to GM-CSF, but to proteins from the expression system (*E. coli*)
4. Antibodies formed against both product- and non-product-related proteins
5. No antibody formation

This example clearly illustrates not only the range of clinical manifestations with respect to antibody formation to drug therapy but also how the choice of an expression system can affect the final product. In this example, the expression system was responsible for the adverse events reported. This finding is certainly clinically relevant considering the homologous product, sargramostim, has been on the U.S. market for quite some time.

The above example also gives an indication of the relative importance of carbohydrate analysis. Without question, protein glycosylation is the most complex of all posttranslational modifications made in eukaryotic cells, the importance of which cannot be underestimated. For many compounds, glycosylation can readily affect protein solubility (as influenced by folding), protease resistance, immunogenicity, and pharmacokinetic/pharmacodynamic profiles (i.e., clearance and efficacy) [36]. Typical analytical methodologies used to assess carbohydrate content are also listed in Table 2.

### 3.2 Stability-Indicating Methodologies

Analytical methodologies that are specific to the major analyte that are also capable of separating and quantifying potential degradation products and impurities, while simultaneously maintaining specificity and accuracy, are deemed stability indicating. Traditional stability-indicating high-performance liquid chromatography (HPLC) methodologies for small molecules are developed and validated with relative ease. Typically, the stability-indicating nature of an analytical method can be demonstrated by subjecting the product to forced degradation in the presence of heat, acid, alkali, light, or peroxide [37]. If degradation products are sufficiently well resolved from the active while maintaining specificity and accuracy, the method is suitable. In contrast to small molecules, there is no one "gold standard" analytical methodology that can be utilized to determine the potential degradation products and impurities in the milieu that may constitute a biopharmaceutical drug product. Furthermore, a one-dimensional structure assessment (e.g., in terms of an absorption spectrum) does not give any indication of the overall activity of the product, as is the case with traditional small molecules. Thus, the stability assessment of biopharmaceuticals will typically comprise a multitude of methodologies that when taken together give an indication of the stability of the product. The overall goal is to assess the structural elements of the compound as well as attempt to determine the relative quantities of potential degradation products, as well as product isoforms and impurities, that are inherent to the expression systems utilized for API manufacture. However, it is still advised that bioactivity determinations are made at appropriate intervals throughout the stability program, as discussed below. Furthermore, any biopharma-

ceutical stability program should also minimally include an evaluation of the in vitro immunogenicity profile of the product with respect to time, temperature, and other potential degradative conditions.

### 3.3 Method Validation and Transfer

Analytical method validation is the process by which scientists prove that the analytical method is suitable for its intended use. Guidances available on validation procedures for some traditional analytical methodologies [38] can be adapted to nontraditional methodologies. The United States Pharmacopeia (USP) and National Formulary (NF) do provide some guidance on designing and assessing biological assays [39], as does the U.S. FDA [40]. Essentially, validation determines the acceptable working ranges of a method and the limitations of that method. At a minimum the robustness, precision, and accuracy of quantitative methodologies should be determined during support of API iteration and refinement, while at the very least the robustness of qualitative methodologies should be assessed. Of particular importance for successful analytical method validation is ensuring that the proper standards and system suitability compounds have been chosen and are representative or analogous to the compound to be analyzed and traceable to a known origin standard, such as the National Institute of Standards and Technology (NIST) or USP/NF. If a reference standard from an "official" source is not available, in-house standards may be used provided they are of the highest purity that can be reasonably obtained and are thoroughly characterized to ensure its identity, strength, quality, purity, and potency.

Methods developed and validated during the product development phase are routinely transferred to quality control or contract laboratories to facilitate release and in-process testing of production batches. Ensuring that method transfer is executed properly, with well-defined and reproducible system suitability and acceptance criteria, is the responsibility of both laboratories. Experiments should consist of all those parameters assessed during method validation and should include an evaluation of laboratory-to-laboratory variation. This information will give an idea of the reliability of the methodology and equipment used under the rigors of large-scale manufacturing.

## 4 FORMULATION DEVELOPMENT

The previous sections have highlighted some of the limitations and difficulties in developing biotechnology-derived pharmaceuticals. Although there are major technological limitations in working with these products, their synthesis and manufacturing are significantly more reproducible compared to naturally derived biologics. Determining the most appropriate route of administration and subsequent formulation is dependent on a number of factors, including the product's indication, duration of action, pharmacokinetic parameters, stability profile, and toxicity. As mentioned previously, biopharmaceuticals are typically delivered parenterally, and thus we will focus on those studies required to successfully develop a parenteral formulation of a biopharmaceutical agent. The goal of formulation development is to design a dosage form that ensures the safety and efficacy of the product through-

out its shelf life while simultaneously addressing the clinical needs of both the patient and caregivers to ensure compliance. Formulation development is truly a balancing act, attempting to emphasize the benefits of the therapy and patient compliance while maximizing drug efficacy and minimizing toxicity. As such, a number of studies are required to properly design and develop a formulation, many of which are discussed below.

### 4.1 Processing Materials and Equipment

An important factor in the quality and reproducibility of any formulation development activity is the materials utilized for formulating and processing studies. In addition, the choice of container closure systems for the API and the formulation needs to be considered carefully to provide maximum product protection and optimal stability. Variability between small- and larger scale development stages may also be significant depending on the API and materials involved during process scale-up. It is important to conduct process development studies utilizing equipment representative of what will be used for large-scale production, if possible. Implementing this design approach will enable at least some limited dimensional analysis, allowing for early identification of critical design parameters, thereby facilitating scale-up or permitting earlier attrition decisions and cost savings. Regardless, it is important to consider the chemical composition and material properties of every manufacturing component that may contact the drug product. For instance, processing vessels may be made of glass, glass-lined steel, or bare steel, while stir paddles used to ensure homogeneity made be manufactured of a number of different materials. In short, any manufacturing unit that could potentially come into intimate contact with either the formulation or the API should be demonstrated to be compatible with the product, including sampling instruments, sample vials, analytical and processing tubing, and so forth. Material incompatibility could result in something as simple as unexplained analytical variability due to a loss of drug through adsorptive mechanisms to something as serious as a loss of bioactivity or an increase in immunogenic potential. Therefore, equipment design and materials would ideally be consistent from formulation development through to scale-up and process validation; however, this may not be readily feasible. As such, determining the chemical and physical compatibility of each piece of processing equipment with the API is critical to maintaining the physical and chemical attributes of the product. Furthermore, such studies help eliminate potential sources of experimental variability and give a better indicator as to the relative technological hurdles to successful product development.

Material compatibility protocols must be clearly defined and require that analytical methodologies be suitable for their intended use. Typically, product purity methods and cleaning methodologies utilized to determine residual contaminating product on processing equipment are used for compatibility studies as they are sufficiently sensitive and rugged to accurately determine product content in the presence of a multitude of potential confounding factors. This is particularly important when assessing potential metal, glass, and tubing compatibilities. Compatibility is a function not only of the product's intimate contact with surrounding materials but also of the contact time and surface area with these equipment. As such, protocols should be designed to incorporate expected real-world conditions the product will

see when in contact with the material. For instance, temperature, light, and mechanical stimulation should mimic usage conditions, although study duration should include time intervals that surpass expectations to estimate a potential worst case. These factors should all be considered when examining potential process-related stability.

### 4.2 Container Closure Systems

The ICH guideline for pharmaceutical development outlines requirements for container closure systems for drugs and biologics [41]. The concept paper prepared for this guidance specifically states that "the choice of materials for primary packaging should be justified. The discussion should describe studies performed to demonstrate the integrity of the container and closure. A possible interaction between product and container or label should be considered" [42]. In essence, this indicates that the container closure system should maintain the integrity of the formulation throughout the shelf life of the product. In order to maintain integrity, the container closure system should be chosen to afford protection from degradation induced by external sources, such as light and oxygen. In addition to the primary container, the stability of the product should also be examined in the presence of IV administration components if the product could be exposed to these conditions (see Section 5.6). Understanding the potential impact of product-to-container interactions is integral to maintaining stability and ensuring a uniform dosage. For example, adsorption of insulin and some small molecules has been demonstrated to readily occur in polyvinyl chloride (PVC) bags and tubing when these drugs were present as additives in intravenous (IV) admixtures [43].

In addition to their use in large-volume parenterals and IV sets, thermoplastic polymers have also recently found utility as packaging materials for ophthalmic solutions and some small-volume parenterals [43]. However, there are many potential issues with using these polymers as primary packaging components that are not major concerns with traditional glass container closure systems, including [44]:

1. Permeation of vapors and other molecules in either direction through the wall of the plastic container
2. Leaching of constituents from the plastic into the product
3. Sorption (absorption and/or adsorption) or drug molecules or ions on the plastic material

These concerns largely preclude the utility of thermoplastic polymers as the primary choice of container closure system for protein and peptide therapeutics, although the formulation scientist should be aware of the potential advantages of these systems, such as the ease of manufacturability and their cost. These systems are also finding greater utility in intranasal and pulmonary delivery systems.

Parenterally formulated biopharmaceuticals are typically packaged in glass containers with rubber/synthetic elastomeric closures. Pharmaceutical glass is composed primarily of silicon dioxide tetrahedron which is modified with oxides such as sodium, potassium, calcium, magnesium, aluminum, boron, and iron [45]. The USP classifies glass formulations as follows:

Type I, a borosilicate glass
Type II, a soda–lime treated glass
Type III, a soda–lime glass
NP, a soda–lime glass not suitable for containers for parenterals

The tendency of peptides to adsorb onto glass surfaces is well known and a major concern in the pharmaceutical industry. This is especially important when the dose of the active ingredient is relatively small and a significant amount of drug is adsorbed to these surfaces. In addition, the leaching of atoms or elements in the glass's silicate network into solution is also a potential issue. This is especially important for terminally heat sterilized products where oxide additives included in the silicate network are relatively free to migrate/leach, resulting in increased solution pH, reaction catalysis, and so on [45]. As such, only type 1 treated glass is traditionally used for parenterally administered formulations, where these alkaline-rich phases in the glass have been eliminated, thus decreasing the potential for container closure system interactions. Additional approaches, including surface treatment with silicone (siliconization), have also been developed to minimize the interaction of biotechnology products with free silanols (Si–OH) [46].

Elastomeric closures are typically used for syringe and vial plungers and closures. For vials, elastomers provide a soft and elastic material that can permit the entry of a hypodermic needle without loss of the integrity [45]. For syringes, the closures not only provide a permeation barrier but also allow for a soft gliding surface facilitating plunger movement and drug delivery. Elastomeric polymers, however, are very complex materials composed of multiple ingredients in addition to the basic polymers, such as vulcanizing agents, accelerators, activators, antioxidants, fillers, lubricating agents, and pigments [45]. As leaching of these components into solution is a potential issue, the compatibility of the drug formulation with the closures must be studied early during the formulation development process. The choice and type of elastomeric closure depends on the pH and buffer, if any preservatives are present, the sterilization method, moisture vapor/gas protection, and active compatibility [47]. In addition, the problem of the additives in rubber leaching into the product can be reduced by the coating with specific polymers such as Teflon [48].

Container closure systems required for implantable devices are further restricted by the fact that they are required to be compatible with the formulation over the intended shelf life and therapeutic application time as well as being biocompatible. This means that the system not only must afford protection to and contain the formulation but also cannot cause any potential adverse effects, such as allergy. Typically, implantable systems are composed of biocompatible metals, such as titanium or polymers such as polyethylene glycol or polylactic-co-glycolic acid.

### 4.3 Sterility Assurance

Maintaining the sterility of biopharmaceutical products is especially important due to the relative potency and their innate potential for immunogenic reactions. Further, the biochemical nature of these compounds enables them to serve as potential nutrients for invading organisms. Methods for sterilizing small molecules include

heat terminal sterilization, terminal filtration coupled with aseptic processing techniques, ultraviolet (UV) and gamma irradiation, ethylene oxide exposure (for containers and packaging only), and electron beam irradiation. While terminal heat sterilization is by far the most common sterilization technique, it normally cannot readily be utilized for peptide or protein formulations due to the potential effects of heat and pressure on the compound's structure [48]. Furthermore, irradiation can affect protein stability by cross-linking the sulfur-containing and aromatic residues, resulting in protein aggregation [49].

To overcome these issues, sterile filtration coupled with aseptic processing and filling is the preferred manufacturing procedure for biopharmaceuticals. Garfinkle et al. refer to aseptic processing as "those operations performed between the sterilization of an object or preparation and the final sealing of its package. These operations are, by definition, carried out in the complete absence of microorganisms" [50]. This highlights the importance of manufacturing controls and bioburden monitoring during aseptic processes. Newer technologies such as isolator technology have been developed to reduce human intervention, thereby increasing the sterility assurance. These technologies have the added benefit of facilitating aseptic processing without construction of large processing areas, sterile suites, or gowning areas [50].

Even the most robust monitoring programs do not ensure the sterility of the final formulation. As such, aseptically processed formulations are traditionally filtered through a retentive final filter, which ensures sterility. Coupled with proper component sterilization, traditionally by autoclaving, these processes ensure product sterility. However, filtration is a complex unit operation that can adversely affect the drug product through increased pressure, shear, or material incompatibility. Therefore, filtration compatibility must be assessed thoroughly to demonstrate both product compatibility, and sufficient contaminant retention [51]. Parenteral Drug Association (PDA) technical report 26 provides a thorough systematic approach to selecting and validating the most appropriate filter for a sterilizing filtration application [51].

### 4.4 Excipient Selection

Pharmaceutical products are typically formulated to contain selected nonactive ingredients (excipients) whose function is to promote product stability and enable delivery of the active pharmaceutical ingredient(s) to the target site. These substances include but are not limited to solubilizers, antioxidants, chelating agents, buffers, tonicity contributors, antibacterial agents, antifungal agents, hydrolysis inhibitors, bulking agents, and antifoaming agents [45]. The ICH states that "the excipients chosen, their concentration, and the characteristics that can influence the drug product performance (e.g. stability, bioavailability) or manufacturability should be discussed relative to the respective function of each excipient" [42]. Excipients must be nontoxic and compatible with the formulation while remaining stable throughout the life of the product. Excipients require thorough evaluation and optimization studies for compatibility with the other formulation constituents as well as the container/closure system [52]. Furthermore, excipient purity may be required to be greater than that listed in the pharmacopeial monograph if a specific impurity is implicated in potential degradation reactions (e.g., presence of trace metals) [48].

One of the critical factors in excipient selection and concentration is the effect on preferential hydration of the biopharmaceutical product [53, 54]. Preferential hydration refers to the hydration layers on the outer surface of the protein and can be utilized to thermodynamically explain both stability enhancement and denaturation. Typical excipients used in protein formulations include albumin, amino acids, carbohydrates, chelating and reducing agents, cyclodextrins, polyhydric alcohols, polyethylene glycol, salts, and surfactants. Several of these excipients increase the preferential hydration of the protein and thus enhance its stability. Cosolvents need to be added in a concentration that will ensure their exclusion from the protein surface and enhance stability [54]. A more comprehensive review of excipients utilized for biopharmaceutical drug products is available elsewhere [48].

***Buffer Selection*** In addition to maintaining solution pH, buffers serve a multitude of functions in pharmaceutical formulations, such as contributing toward overall isotonicity, preferential hydration of proteins and peptides, and serving as bulking agents in lyophilized formulations. The buffer system chosen is especially important for peptide and proteins that have sensitive secondary, tertiary, and quaternary structures, as the overall mechanisms contributing to conformational stabilization are extremely complex [48]. Furthermore, a protein's propensity for deamidation at a particular pH can be significant, as illustrated by Wakankar and Borchardt [55]. This study illustrated stability concerns with peptides and proteins at physiological pH in terms of asparagine (Asn) deamidation and aspartate (Asp) isomerization, which can be a major issue with respect to circulating half-life and potential in vivo degradation. This study and others also provide insight into predicting potential degradative mechanisms based on primary and secondary structural elements allowing for formulation design with these pathways in mind.

Selecting the appropriate buffer primarily depends on the desired pH range and buffer capacity required for the individual formulation; however, other factors, including concentration, effective range, chemical compatibility, and isotonicity contribution, should be considered [56]. Some acceptable buffers include phosphate (pH 6.2–8.2), acetate (pH 3.8–5.8), citrate (pH 2.1–6.2, p$K$ 3.15, 4.8, and 6.4), succinate (pH 3.2–6.6, p$K$ 4.2 and 5.6), histidine (p$K$ 1.8, 6.0, and 9.0), glycine (p$K$ 2.35 and 9.8), arginine (p$K$ 2.18 and 9.1), triethanolamine (pH 7.0–9.0), tris-hydroxymethylaminomethane (THAM, p$K$ 8.1), and maleate buffer [48]. Additionally, excipients utilized solely for tonicity adjustment, such as sodium chloride and glycerin, may not only differ in ionic strength but also could afford some buffering effects that should be considered [52].

***Preservatives*** In addition to those processing controls mentioned above (Section 4.3 of From Pilot Plant to Manufacturing: Effect of Scale-Up on Operation of Jacketed Reactors), the sterility of a product may be maintained through the addition of antimicrobial preservatives. Preservation against microbial growth is an important aspect of multidose parenteral preparations as well as other formulations that require preservatives to minimize the risk of patient infection upon administration, such as infusion products [52]. Aqueous liquid products are prone to microbial contamination because water in combination with excipients derived from natural sources (e.g., polypeptides, carbohydrates) and proteinaceous active ingredients may serve as excellent media for the growth [57]. The major criteria for the selection of an appropriate

preservative include efficiency against a wide spectrum of microorganisms, stability (shelf life), toxicity, sensitizing effects, and compatibility with other ingredients in the dosage form [57]. Typical antimicrobial preservatives include *m*-cresol, phenol, parabens, thimerosal, sorbic acid, potassium sorbate, benzoic acid, chlorocresol, and benzalkonium chloride. Cationic agents such as benzalkonium chloride are typically not utilized for peptide and protein formulations because they may be inactivated by other formulation components and their respective charges may induce conformational changes and lead to physical instability of the API. Further, excipients intended for other applications, such as chelating agents, may exhibit some antimicrobial activity. For instance, the chelating agent ethylenediaminetetraacetic acid (EDTA) may exhibit antimicrobial activity, as calcium is required for bacterial growth.

Identifying an optimal antimicrobial preservative is based largely on the effectiveness of that preservative at the concentration chosen. In short, it is not enough to assess the compatibility of the preservative of choice with the API and formulation and processing components. There also needs to be a determination of whether the preservative concentration is sufficient to kill certain standard test organisms. The USP presents standard protocols for assessing the relative efficacy of a preservative in a formulation using the antimicrobial effectiveness test (AET) [58]. Briefly, by comparing the relative kill efficiency of the formulation containing varying concentrations of the preservative, the formulator can determine the minimal concentration required for preservative efficacy and design the formulation accordingly.

## 5 DRUG PRODUCT STABILITY

### 5.1 Defining Drug Product Storage Conditions

From a regulatory standpoint, the primary objective of formulation development is to enable the delivery of a safe and efficacious drug product to treat and/or mitigate a disease state throughout its proposed shelf life. The efficacy and in many cases the safety of a product are directly related to the stability of the API, both neat and in the proposed formulation under processing, storage, and shipping conditions as well as during administration. As such, the concept of drug stability for biotechnology-derived products does not change substantially from that of small molecules, although the level of complexity increases commensurate with the increased complexity of the APIs in question and the formulation systems utilized for their delivery.

Stability study conditions for biotechnology-derived APIs and their respective drug products are not substantially different from those studies conducted for small molecules. Temperature and humidity conditions under which to conduct said studies are outlined in ICH Q1A(R2), which incorporates ICH Q1F, stability study conditions for zones III and IV climactic conditions [59]. Additional guidance specific to conducting stability studies on biopharmaceutical drug products is given in ICH Q5C [1]. However, the intention of ICH Q5C is not to outline alternate temperature and humidity conditions to conduct primary stability studies; rather it provides guidance with respect to the fact that the recommended storage conditions and expiration dating for biopharmaceutical products will be different from product to product and provides the necessary flexibility in letting the applicant determine

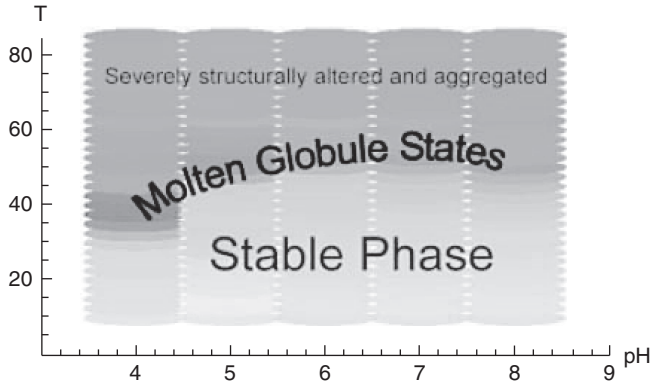


**FIGURE 1** Empirical phase diagram for ricin toxin A-chain generated using CD molar ellipticity at 208 nm, ANS fluorescence, and intrinsic Trp fluorescence intensity data. Labels indicate the state of the protein within the same region of color based on evaluation of a compilation of data sets. (Reproduced with permission from ref. 62.)

the proper storage conditions for their respective product. Furthermore, this document provides general guidance in directing applicants in the types of analytical methodologies that may be used and direction on how to properly assess the stability of these complex molecules [1]:

> Assays for biological activity, where applicable, should be part of the pivotal stability studies. Appropriate physicochemical, biochemical and immunochemical methods for the analysis of the molecular entity and the quantitative detection of degradation products should also be part of the stability program whenever purity and molecular characteristics of the product permit use of these methodologies.

One recent approach to aid in defining the design space for protein and peptide therapeutics has been to create empirical phase diagrams indicating the relative stability of compounds based on altering conditions and assessing conformational changes via a compilation of analytical techniques (Figure 1) [60–62]. These empirical phase diagrams can be generated based on pH, temperature, salt concentration, and so on, and, although seemingly laborious at first glance, could provide invaluable information in defining the extremes to which a compound may be subjected without altering its conformation. For instance, if an empirical phase diagram determines the safe temperature range for a compound is up to 35 °C and an excursion occurs to 33 °C, this information would give the stability scientist a guideline as to the appropriate course of action. Under the traditional testing paradigm of ICH Q1A, where stability testing is limited to 25, 30, and 40 °C, one may not know the compound's upper transition temperature to induce conformational changes. If the information is not already available, then additional excursion studies may need to be conducted to assimilate this information and take the appropriate course of action.

### 5.2 Mechanisms of Protein and Peptide Degradation

The inherent heterogeneity of peptide and protein drug substances results in their relative sensitivity to processing, storage, and handling conditions as well as a mul-

**TABLE 3 Potential Degradative Mechanisms of Peptides and Proteins**

| | Degradative Mechanism | Site of Occurrence |
|---|---|---|
| **Chemical degradative mechanisms** | Oxidation | Intrachain disulfide linkages Met, Trp, Tyr |
| | Peptide bond hydrolysis | AA backbone |
| | N-to-O migration | Ser and Thr |
| | α- to β-Carboxy migration | Asp and Asn |
| | Deamidation | Asn and Gln |
| | Acylation | α-Amino and ε-amino group |
| | Esterification/carboxylation | Glu, Asp, and C-term |
| **Physical degradative mechanisms** | Unfolding | Partial unfolding of tertiary structure |
| | Aggregation | Aggregation of subunits could result in precipitation |
| | Adsorption | Adsorption to processing equipment and container closure systems |

*Source*: Modified from Crommelin et al. [5].

titude of other factors. Most importantly, this heterogeneity results in a whole host of potential degradative mechanisms, some of which are compiled in Table 3 and include chemical instability pathways such as oxidation, hydrolysis of side chains and potentially the peptide backbone, and deamidation of Asn and Gln side chains. Also, physical instability manifesting in the form of protein unfolding, formation of intermediate structures, aggregation, and adsorption to the surfaces of containers and other equipment can be a major technical hurdle in developing any biopharmaceutical and may or may not be related to chemical instability [63]. Further complicating matters is that instability can potentially manifest in various ways and may or may not be detectable by any one method. Taken together, however, the compilation of methodologies utilized for stability assessment should give a good approximation as to the degradative mechanisms of the compound in its respective formulation. Further, bioactivity and immunogenicity assays should play integral roles in assessing the relative stability of any biopharmaceutical compound. Briefly stated, the chemical and physical stability of products is extraordinarily difficult to assess and will not be belabored here as good reviews on this topic are readily available in the literature [63, 64].

### 5.3 Photostability

In certain cases, exposure of pharmaceutical compounds to UV and visible light could result in electronic excitation, termed vertical transition, that could ultimately result in light-induced degradation. The ICH guideline Q1B [65] is a reference on how to conduct photostability stress testing for pharmaceutical compounds. In brief, compounds are exposed to an overall illumination of not less than 1.2 million lux hours and an integrated near-UV energy of not less than 200 $Wh/m^2$ [65]. These requirements are in addition to normal stability stress testing and require the additional caveat that analytical methodologies are suitable to also detect photolytic degradation products, as discussed above. A comprehensive discussion of small-molecule photolytic degradative mechanisms is available for further review [66].

### 5.4 Mechanical Stress

Regulatory guidance on appropriate methods to evaluate the effect of shear stress and process-handling stability studies is not available. However, these studies are integral in determining the relative stability of the product with respect to mechanical stresses introduced during development and manufacturing. Although not typically recognized as a major degradative pathway for most small-molecule dosage forms, the introduction of mechanical stress is recognized as a major challenge in the formulation of semisolids and can potentially induce physical instability of biopharmaceuticals, although the extent of this effect is currently unknown. For example, processing shear may influence the protein's outer hydration shell, altering the stabilizing energy provided from preferential hydration and resulting in the exposure of internal, nonpolar residues. This may facilitate aggregation if enough shear force is provided. Alternately, the shear energy required to force unfolding has been studied but has not been related to the fluid dynamic shear experienced during processing. Therefore, stress studies should include meticulous controls in the form of temperature, light and humidity, and fluid dynamic shear as a function of time. Data from these studies could be incorporated into empirical phase diagrams, and/or response surfaces, to help further define the design space for the active and finished drug product. Understanding the effects of stress introduced during the manufacturing processing of biopharmaceutical products could facilitate the selection of appropriate PAT tools and QbD incorporation in the development of these products. Clearly, there is a considerable need for research in this area, and until the extent of the possible effects are understood, this lack of knowledge poses an unknown risk and prevents adequate risk assessment for biopharmaceutical development activities consistent with ICH Q9.

### 5.5 Freeze–Thaw Considerations and Cryopreservation

The rapid or continuous freezing and thawing of protein products could contribute significantly to instability of the API. Such studies are typically designed to assess the implications of potential transport and handling conditions. These conditions include not only the manufacturing processing, storage, and shipment to warehouses and pharmacies but also subsequent pharmacy storage and patient handling [52]. Unpredictable and somewhat modest temperature fluctuations could easily induce degradation or conformational changes that may reduce bioactivity or expose antigenic epitopes [5]. These effects could also be a result of altered preferential hydration at the surface of the peptide or protein through salting-out effects induced by rapid freezing, which could easily denature the product [67].

### 5.6 Use Studies

Stability of biopharmaceutical compounds should also be determined under conditions that mimic their normal usage. For instance, the stability of reconstituted lyophilized products should be assessed with respect to time and temperature and, if applicable, light and mechanical stimuli. Likewise, the stability of a compound included in implantable devices and controlled-release microsphere formulations should be determined over the course of its required use, under conditions which mimic the heat, moisture, light, and enzymatic physiological conditions to which it

will be implanted. Such studies should also determine the release profile of the compound over these specified conditions.

Drug products intended for IV administration are generally dosed as an initial bolus followed by a slow infusion. Consequently, admixture studies of the compound in potential IV fluids, such as 0.9% (w/v) saline, 5% (w/v) dextrose, and Ringer's solution, should also be assessed to determine the relative stability of the compound in this new environment. These studies are critical as the formulation dynamic that protected and stabilized the compound has now been altered dramatically with dilution. This environmental change could potentially impact the preferential hydration of the compound as well as directly induce conformational changes based on the diluent chosen and the compound's potential degradative mechanism(s). Additional contributing factors to instability in admixture solutions could be due to changes in pH, mechanical mixing of the compound in the IV bag, adsorption of the compound to the bag itself (which is typically polymeric), or IV sets used for administration, as well as an increased potential for oxidative degradation. The suitability of analytical methodologies should also be determined in the presence of these additional analytes.

### 5.7 Container Closure Integrity and Microbiological Assessment

Ensuring that parenteral pharmaceuticals maintain their sterility over the course of their shelf life is an integral part of any stability assessment [68]. Parenteral dosage forms must be free from microbiological contamination, bacterial endotoxins, and foreign particulate matter. Selection of the adequate sterile manufacturing process has been briefly discussed above. Determining the microbiological integrity of the product over its shelf life also gives an indication of the relative quality of the container closure system chosen for the formulation. Compendial sterility and endotoxin testing are often used for this purpose; however, sampling is dependent on a statistical evaluation of the batch size, unit fill volume, and method of product sterilization [68]. Additionally, since these tests are destructive, it would be impossible to test an entire stability batch to ensure viability of a container closure system. Other nondestructive tests have been developed to determine the integrity of a container's closure system [69]. These tests could also serve as a surrogate indicator of product manufacturing quality over time.

### 5.8 Data Interpretation and Assessment

Interpretation of primary stability data for determining expiration dating and primary storage conditions has been outlined by ICH Q1E [70]. This guidance document delineates broad methodologies for interpreting primary and accelerated stability data and extrapolation of said data for determining expiry dating. Of course, expiry dating cannot be made without reference to specifications for those primary stability-indicating parameters assessed, which is discussed below. Traditionally, stability assessments performed during preformulation will give an indication of the potential storage conditions as well as allow for extrapolation of accelerated stability studies to kinetic degradation rates. Typically this is done through Arrhenius manipulations. However, as one would expect, these analyses are not readily useful for biopharmaceutical products, as there is rarely a linear correlation between

temperature and the compound's degradative rate. This is primarily due to the complex and often competing degradative mechanisms as well as the potential for so-called molten globule intermediate phases. In spite of these limitations, ICH Q5C does provide relevant guidance in illustrating the flexibility required for determining storage conditions, as these products usually require a very narrow temperature condition to maintain optimal stability. Further guidances may be needed to enhance uniformity in testing methodology and enable the utilization of validated PAT methodologies.

## 6 QUALITY BY DESIGN AND SCALE-UP

### 6.1 Unit Operations

Unit operations are defined as the individual basic steps in a process that when linked together define the process train and result in the final product. In practical terms, a unit operation is often defined as an individual step that is carried out on one piece of equipment. Typical biopharmaceutical API unit operations may include fermentation or bioreactor processes, cell separation through centrifugation or microfiltration, virus removal or inactivation, cell lysis and inclusion body precipitation, product refolding, and purification steps [71]. Conversely, those unit operations for drug product manufacturing procedures would be similar to those seen in the manufacture of a small molecule of comparable dosage form, namely mixing, fluid transfer, sterile filtration, dose filling, lyophilization, and so on. Of course, unit operations will be dependent on the manufacturing process for the specific dosage form, but careful preformulation and characterization studies will enable relatively straightforward process design and ease subsequent scale-up activities. Modeling of unit operations for both small and large molecules is a recognized gap in our ability to achieve QbD. The application of accepted engineering methods to the problem is the subject of active research.

### 6.2 Bioburden Considerations

Bioburden refers to the amount of microbial flora that can be detected on an item, on a surface, or in a solution [68]. As mentioned previously, microbial contamination and bioburden are especially important for biotechnology-derived parenteral products since these products are typically capable of supporting microbial growth. Special care should be taken to ensure not only that the final packaged product does not contain microbial contamination but also that manufacturing equipment is also free from contamination. Monitoring bioburden and determining potential levels of microbial contamination on equipment surfaces are particularly important with respect to the material being evaluated.

In general, bioburden counts in parenteral solutions are obtained by conducting the total aerobic counts and total yeast and mold counts as specified in the USP microbial limits test (61) or an equivalent test [72]. In addition, membrane filtration of larger than specified volumes may also be used to detect any microbial contamination when sample results are expected to contain a negligible number of microbial flora or in the presence of potential confounding factors, such as antimicrobial

preservatives [68, 72]. It is important to note that the presence of a high bioburden count can present an endotoxin contamination problem, as whole microbial cells and spores can be removed by sterilizing grade filtration (0.2 μm), while endotoxins are not [68]. These issues also underscore the importance of cleaning methods and their respective validation as well as assessing relevant product contamination on manufacturing equipment.

### 6.3 Scale-Up and Process Changes

The FDA defines process validation as "establishing documented evidence that provides a high degree of assurance that a specific process will consistently produce a product meeting its predetermined quality attributes" [73]. While validation studies are typically performed at full scale, in most cases scale-down or laboratory-scale models were used to initially develop the manufacturing process. Consequently, scale-down process precharacterization and characterization studies are considered crucial to successful process validation for both API and drug product manufacturing schemes [74]. Although they do require qualification work and a significant commitment of time and resources, characterization studies provide significant insight into the critical process and control parameters for each unit operation as well as improved success rates for process validation due to a better, more complete understanding of the process [74]. In engineering terms, characterization studies identify the critical parameters useful for dimensional analysis that enable successful process scale-up.

While the above explanation attempts to simplify the scale-up process, it is not meant to trivialize it. In fact, scale-up is probably the most difficult manufacturing challenge for traditional small molecules, let alone biopharmaceuticals. Issues such as homogeneous mixing, bulk product holding and transfer, and sterile filtration could all be potentially compounded due to the increased scale and introduced stress. However, a QbD approach to rational drug design should enable simplified process scale-up and validation. This is only true if experimental design approaches have been utilized to identify the design space for the processes involved in the production of the molecule. This is also where the greatest benefit of developing empirical phase diagrams early in development could materialize. Essentially, the QbD approach identifies the quality attributes of the product based on scientific rationale as opposed to attempting to fit the proverbial square peg into a round hole through a trial-and-error approach. This rational design approach goes further to identify the limiting factors of each unit operation and provides the means of attempting to correlate how each unit operation affects the final product quality attributes.

In order to initiate a successful QbD program, the first step is to identify those process parameters that are essential to product quality and develop well-validated analytical methodologies to monitor those parameters. In short, the process involves identification of the potential design space for production of the molecule and confirmation that design space through rational, deliberate experimentation. Ideally, process monitoring should be done in real time to minimize production time and if possible online; however, this may not always be the case or even necessary depending upon the relative duration of the process to the test. Recognizing potential quality metrics earlier in the development process could also potentially facilitate

greater flexibility during product development and subsequent process characterization [74]. Certainly, manufacturing site-specific differences could also potentially introduce variability into processes. It is for this reason that site-specific personnel training, process/technology transfer and validation, and stability assessments are required to ensure product quality.

By definition, a process designed under the auspices of QbD should enable a degree of process knowledge that allows for controlled process changes without affecting the final product or requiring regulatory approval. For immediate- and controlled-release solid dosage products, SUPAC guidelines provide direction on the studies to conduct to determine the impact of a process change. Although there is some regulatory guidance available for biological products (e.g., "Changes to an Approved Application for Specified Biotechnology and Specified Synthetic Biological Products" or "FDA Guidance Concerning Demonstration of Comparability of Human Biological Products, Including Therapeutic Biotechnology-Derived Products"), process changes need to be evaluated on a case-by-case basis. The comparative analysis of process changes should also be evaluated with respect to defined product specifications. PAT will be invaluable in determining the potential impact of process changes. While stability is often the main metric for small-molecule drug product, bioactivity and immunogenicity will need to be added metrics for biopharmaceuticals. Therefore, any process change should be approached subjectively and care should be taken to validate the relative impact on the safety and efficacy of the product.

## 7 CONCLUDING REMARKS

Although the goals are the same, developing biotechnology molecules presents challenges that are unique compared to the development of conventional small molecules. The innate complexity of the molecular and macromolecular structures requires three dimensionally viable stability assays and understanding. The complexity of possible physiological responses and interactions requires an enhanced understanding of the formulation and processing stresses to identify the minor but critical changes that result in product unacceptability. A key to addressing these challenges is the development of analytical techniques with the sensitivity and reliability to detect and monitor such changes and to provide data to another gap-closing activity—modeling unit operations. Also the need to develop meaningful kinetic models is obvious to everyone involved in the development of both large and small molecules. Linking this type of information to the major efforts in the discovery arena is a necessary step to bringing the products of the future to market.

The use of biotechnology products is increasing exponentially and many opportunities exist to improve their development. The first step may be defining rational biotechnology-derived drug "developability" standards that can be assessed during preclinical/early development testing. Such a tiered approach based upon the potential risk, the confidence in methodology, and benefit has of course been a proven strategy for small molecules, and a preliminary version applicable to biotechnology drug products is likely possible today given the topics discussed in this chapter.

## ACKNOWLEDGMENTS

The authors would like to thank The School of Pharmacy and Pharmaceutical Sciences, the Department of Industrial and Physical Pharmaceutics of Purdue University and the National Institutes of General Medical Sciences (R01-GM65448) for their financial support.

## REFERENCES

1. International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use, Q5C: Quality of biotechnological products: Stability testing of biotechnological/biological products, Nov. 30, 1995.
2. Biotechnology Industry Organization. Available: http://www.bio.org/, accessed Jan. 10, 2007.
3. Nims, R., Presene, E., Sofer, G., Phillips, C., and Chang, A. (2005), Adventitious agents: Concerns and testing for biopharmaceuticals, in Rathore, A. S., and Sofer, G., Eds., *Process Validation in Manufacturing of Biopharmaceuticals: Guidelines, Current Practices, and Industrial Case Studies*, Taylor and Francis, Boca Raton, FL.
4. Rathore, A. S., and Sofer, G. (2005), Life span studies for chromatography and filtration media, in Rathore, A. S., and Sofer, G., Eds., *Process Validation in Manufacturing of Biopharmaceuticals: Guidelines, Current Practices, and Industrial Case Studies*, Taylor and Francis, Boca Raton, FL.
5. Crommelin, D. J. A., Storm, G., Verrijk, R., de Leede, L., Jiskoot, W., and Hennink, W. E. (2003), Shifting paradigms: Biopharmaceuticals vs. low molecular weight drugs, *Int. J. Pharm.*, 266, 3–16.
6. Tang, L., Persky, A. M., Hochhaus, G., and Meibohm, B. (2004), Pharmacokinetic aspects of biotechnology products, *J. Pharm. Sci.*, 93(9), 2184–2204.
7. Roberts, M. J., Bentley, M. D., and Harris, J. M. (2002), Chemistry for peptide and protein pegylation, *Adv. Drug Deliv. Rev.*, 54, 459–476.
8. Frokjaer, S., and Otzen, D. (2005), Protein drug stability: A formulation challenge, *Nat. Rev.*, 4, 298–306.
9. Niu, C., and Chiu, Y. (1998), FDA perspective on peptide formulation and stability issues, *J. Pharm. Sci.*, 87, 1331–1334.
10. Washington, N., Washington, C., and Wilson, C. G. (2001), *Physiological Pharmaceutics: Barriers to Drug Absorption*, 2nd ed., Taylor and Francis, New York.
11. Hussain, A., Arnold, J. J., Khan, M. A., and Ashan, F. (2004), Absorption enhancers in pulmonary protein delivery, *J. Controlled Release*, 94, 15–24.
12. Alpar, H. O., Somavarapu, S., Atuah, K. N., and Bramwell, V. W. (2005), Biodegradable mucoadhesive particulates for nasal and pulmonary antigen and DNA delivery, *Adv. Drug Deliv. Rev.*, 57, 411–430.
13. Thanou, M., Verhoef, J. C., and Junginger, H. E. (2001), Chitosan and its derivatives as intestinal absorption enhancers, *Adv. Drug Deliv. Rev.*, 50, 91–101.
14. Schellekens, H. (2002), Immunogenicity of therapeutic proteins: Clinical implications and future prospects, *Clin. Ther.*, 24, 1720–1740.
15. Bhadra, D., Bhadra, S., Jain, P., and Jain, N. K. (2002), Pegnology: A review of PEG-ylated systems, *Pharmazie*, 57, 5–29.
16. Matthews, S. J., and McCoy, C. (2004), Peginteferon α2a: A review of approved and investigational uses, *Clin. Ther.*, 26, 991–1025.

17. Knudsen, L. B., Nielsen, P. F., Huusfeldt, P. O., Johansen, N. L., Madsen, K., Pedersen, F. Z., Thogersen, H., Wilken, M., and Agerso, H. (2000), Potent derivatives of glucagon-like peptide-1 with pharmacokinetic properties suitable for once daily administration, *J. Med. Chem.*, 43, 1664–1669.
18. Foldvari, M., Attah-Poku, S., Hu, J., Li, Q., Hughes, H., Babiuk, L. A., and Kruger, S. (1998), Palmitoyl derivatives of interferon α: Potential for cutaneous delivery, *J. Pharm. Sci.*, 87, 1203–1208.
19. Egrie, J.C., and Browne, J. K. (2001), Development and characterization of novel erythropoiesis stimulating protein (NESP), *Nephrol. Dial. Transplant.*, 16(Suppl 3), 3–13.
20. Haselbeck, A. (2003), Epoetins: Differences and their relevance to immunogenicity, *Curr. Med. Res. Opin.*, 19, 430–432.
21. Hochuli, E. (1997), Interferon immunogenicity: Technical evaluation of interferon-α2A, *J. Int. Cytokine Res.*, 17, S15–S21.
22. Maack, T., Johnson, V., Kau, S. T., Figueiredo, J., and Sigulem, D. (1979), Renal filtration, transport, and metabolism of low-molecular-weight proteins: A review, *Kidney Int.*, 16, 251–270.
23. Takagi, A., Masuda, H., Takakura, Y., and Hashida, M. (1995), Disposition characteristics of recombinant human interleukin-11 after a bolus intravenous administration in mice, *J. Pharmacol. Exp. Ther.*, 275, 537–543.
24. Anderson, P. M., and Sorenson, M. A. (1994), Effects of route and formulation on clinical pharmacokinetics of interleukin-2, *Clin. Pharmacokinet.*, 27, 19–31.
25. Rabkin, R., Ryan, M. P., and Duckworth, W. C. (1984), The renal metabolism of insulin, *Diabetologia*, 27, 351–357.
26. Johnson, V., and Maack, T. (1977), Renal extraction, filtration, absorption, and catabolism of growth hormone, *Am. J. Phsiol.*, 233, F185–F196.
27. Carone, F. A., and Peterson, D. R. (1980), Hydrolysis and transport of small peptides by the proximal tubule, *Am. J. Physiol.*, 238, F151–F158.
28. Carone, F. A., Peterson, D. R., and Flouret, G. (1982), Renal tubular processing of small peptide hormones, *J. Lab. Clin. Med.*, 100, 1–14.
29. Braeckman, R. (2000), Pharmacokinetics and pharmacodynamics of protein therapeutics, in Reid, R., Ed., *Peptide and Protein Drug Analysis*, Marcel Dekker, New York.
30. Meissner, H. C., Groothuis, J. R., Rodriguez, W. J., Welliver, R. C., Hogg, G., Gray, P. H., Loh, R., Simoes, E. A., Sly, P., Miller, A. K., Nichols, A. I., Jorkasky, D. K., Everitt, D. E., and Thompson, K. A. (1999), Safety and pharmacokinetics of an intramuscular monoclonal antibody (SB 209763) against respiratory syncytial virus (RSV) in infants and young children at risk for severe RSV disease, *Antimicrob. Agent Chemother.*, 43, 1183–1188.
31. Prescribing information for Lupron Depot®. Manufactured for TAP Pharmaceuticals, Inc., Lake Forest, IL 60045.
32. Prescribing information for Viadur®. Manufactured by Alza Corporation, Mountain View, CA 94043.
33. Degim, I. T., and Celebi, N. (2007), Controlled delivery of peptides and proteins, *Curr. Pharm. Des.*, 13, 99–117.
34. Sylvester, R. K. (2002), Clinical applications of colony-stimulating factors: A historical perspective, *Am. J. Health-Syst. Pharm.*, 59, s6–s12.
35. Dorr, R. T. (1993), Clinical properties of yeast-derived versus *Escherichia coli*-derived granulocyte-macrophage colony-stimulating factor, *Clin. Ther.*, 15, 19–29.
36. Jenkins, N., Shah, P. M., and Buckberry, L. D. (2000), Carbohydrate analysis of glycoproteins and glycopeptides, in Reid, R., Ed., *Protein and Peptide Drug Analysis*, Marcel Dekker, New York.

37. Berglund, M., Bystroem, K., and Persson, B. (1990), Screening chemical and physical stability of drug substances, *J. Pharm. Biomed. Anal.*, 8, 639–643.
38. International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use, Q2(R1): Validation of analytical procedures: Text and methodology, Nov. 2005.
39. U.S. Pharmacopeia (USP), Chapter ⟨111⟩ Design and analysis of biological assays, USP 26, 2003, Rockville, MD.
40. Center for Drug Evaluation and Research, FDA (2001, May). Guidance for industry, bioanalytical method validation, FDA, Washington, DC.
41. International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use, Q8: Pharmaceutical development, Nov. 10, 2005.
42. International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use, Q8: Concept paper, available: http://www.ich.org/LOB/media/MEDIA3096.pdf, accessed Sept. 19, 2003.
43. Avis, K. E., and Levchuk, J. W. (2000), *Remington 20th Edition: Parenteral Preparations*, Lippincott, Williams & Wilkins, Philadelphia.
44. Autian, J. (1968), Interrelationship of the properties and uses of plastics for parenterals, *Bull. Parenteral Drug Assoc.*, 22, 276–288.
45. Avis, K. E. (1986), Sterile Products, in Lachman, L., Lieberman, H. A., and Kanig, J. L., Eds., *The Theory and Practice of Industrial Pharmacy,* 3rd ed., *Sterile Products*, 3rd Ed., Lippincott, Williams and Wilkins, reprint with permission by Stipes Publishing, Champaign, IL.
46. Gombotz, W., Pankey, S., Bouchard, L., Phan, D., and MacKenzie, A. (2002), Stability, characterization, formulation, and delivery system development for transforming growth factor-beta1, in Pearlman, R., and Yang, Y., Eds., *Pharmaceutical Biotechnology*, Vol. 9, *Formulation, Characterization, and Stability of Protein Drugs: Case Histories*, Springer, New York.
47. Bontempo, J. A. (1997), Considerations for elastomeric closures for parenteral biopharmaceutical drugs, in Bontempo, J. A. Ed., *Development of Biopharmaceutical Parenteral Dosage Forms*, Marcel Dekker, New York.
48. Banga, A. K. (2006), *Therapeutic Peptides and Proteins: Formulation, Processing and Delivery Systems*, 2nd ed., Taylor and Francis, Boca Raton, FL.
49. Yamamoto, O. (1992), Effect of radiation on protein stability, in Ahern, T. J., and Manning, M. C. Eds., *Pharmaceutical Biotechnology*, Vol. 2, *Stability of Protein Pharmaceuticals. Part A: Chemical and Physical Pathways of Protein Degradation*, Plenum, New York.
50. Garfinkle, B. D., and Henley, M. W. (2000), *Remington 20th Edition: Sterilization*, Lippincott, Williams & Wilkins, Philadelphia.
51. PDA Technical Report No. 26 (1998), Sterilizing filtration of liquids, *PDA J. Pharm. Sci. Tech.*, 52(3, Suppl).
52. Defelippis, M. R., and Akers, M. J. (2000), *Pharmaceutical Formulation Development of Peptides and Proteins: Peptides and Proteins as Parenteral Suspensions: An Overview of Design, Development, and Manufacturing Considerations*, Taylor and Francis, Philadelphia.
53. Timasheff, S. N. (1998), Control of protein stability and reactions by weakly interacting cosolvents: The simplicity of the complicated, *Adv. Protein Chem.*, 51, 355–432.
54. Shimizua, S., and Smith, D. J. (2004), Preferential hydration and the exclusion of cosolvents from protein surfaces, *J. Chem. Phys.*, 121, 1148–1154.
55. Wakankar, A. A., and Borchardt, R. T. (2006), Formulation considerations for proteins susceptible to asparagine deamidation and aspartate isomerization, *J. Pharm. Sci.*, 95, 2321–2336.

56. Windheuser, J. J. (1963), The effect of buffers on parenteral solutions, *Bull. Parenteral Drug Assoc.*, 17, 1–8.
57. Im-Emsap, W., Paeratakul, O., and Siepmann, J. (2002), Disperse Systems, in Banker, G. S., and Rhodes, C. T., Eds., *Modern Pharmaceutics*, 4th Ed., Marcel Dekker, New York.
58. U.S. Pharmacopeia (USP 26 2003), Chapter ⟨51⟩ Antimicrobial effectiveness testing, Rockville, MD.
59. International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use, Q1A(R2): Stability testing of new drug substances and products, Feb. 6, 2003.
60. Fan, H., Li, H., Zhang, M., and Middaugh, C. R. (2007), Effects of solutes on empirical phase diagrams of human fibroblast growth factor 1, *J. Pharm. Sci.*, 96, 1490–1503.
61. Harn, N., Allan, C., Oliver, C., and Middaugh, C. R. (2006), Highly concentrated monoclonal antibody solutions: Direct analysis of physical structure and thermal stability, *J. Pharm. Sci.*, 96, 532–546.
62. Peek, L. J., Brey, R. N., and Middaugh, C. R. (2007), A rapid, three-step process for the preformulation of a recombinant ricin toxin A-chain vaccine, *J. Pharm. Sci.*, 96, 44–60.
63. Violand, B. N., and Siegel, N. R. (2000), Protein and peptide chemical and physical stability, in Reid, R., Ed., *Protein and Peptide Drug Analysis*, Marcel Dekker, New York.
64. Manning, M. C., Patel, K., and Borchardt, R. T. (1989), Stability of protein pharmaceuticals, *Pharm. Res.*, 6, 903–918.
65. International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use, Q1B: Stability testing: Photostability testing of new drug substances and products, Nov. 6, 1996.
66. Fasani, E., and Albini, A. (2005), Photostability stress testing, in Baertschi, S. W., Ed., *Pharmaceutical Stress Testing: Predicting Drug Degradation*, Taylor and Francis, Boca Raton, FL.
67. Ghosh, T., Kalra, A., and Garde, S. (2005), On the salt-induced stabilization of pair and many-body hydrophobic interactions, *J. Phys. Chem. B. Condens. Matter Mater. Surf. Interfaces Biophys.*, 109, 642–651.
68. Birrer, G. A., Murthy, S. S., Liu, J., and Estrada, J. (2001), Parenteral dosage forms, in Ahuja, S., Ed., *Handbook of Modern Pharmaceutical Analysis*, Academic, New York.
69. Birrer, G. A., Liu, J., Halas, J. M., and Nucera, G. G. (2000), Evaluation of a container closure integrity test model using visual inspection with confirmation by near infra-red spectroscopic analysis, *PDA J. Pharm. Sci. Technol.*, 54, 373–382.
70. International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use, Q1E: Evaluation for stability data, Feb. 6, 2003.
71. Godavarti, R., Petrone, J., Robinson, J., Wright, R., and Kelley, B. D. (2005), Scale-down models for purification processes: Approaches and applications, in Rathore, A. S., and Sofer, G. Eds., *Process Validation in Manufacturing of Biopharmaceuticals: Guidelines, Current Practices, and Industrial Case Studies*, Taylor and Francis, Boca Raton, FL.
72. U.S. Pharmacopeia (USP 26 2003), Chapter ⟨61⟩ Microbial limits test, USP, Rockville, MD.
73. Center for Drug Evaluation and Research, FDA (1987, May), Guideline on general principles of process validation.
74. Seely, J. E. (2005), Process characterization, in Rathore, A. S., and Sofer, G. Eds., *Process Validation in Manufacturing of Biopharmaceuticals: Guidelines, Current Practices, and Industrial Case Studies*, Taylor and Francis, Boca Raton, FL.

## Regulatory Considerations in Approval of Follow-On Protein Drug Products

# REGULATORY CONSIDERATIONS IN APPROVAL OF FOLLOW-ON PROTEIN DRUG PRODUCTS

Erin Oliver,[1] Stephen M. Carl,[2] Kenneth R. Morris,[2] Gerald W. Becker,[3] and Gregory T. Knipp[1]

[1]*Rutgers, The State University of New Jersey, Piscataway, New Jersey*
[2]*Purdue University, West Lafayette, Indiana*
[3]*SSCI, West Lafayette, Indiana*

**Contents**

1 Introduction
  1.1 Emergence of Biotechnology Industry
  1.2 Challenges Facing "Biogenerics"
2 History of Biologics Regulation in United States
  2.1 Early Biologics Regulation (1800s–1990s)
  2.2 Modern Biologics Regulation (1990s–Today)
3 Regulatory Classification of Proteins
  3.1 Definitions and Key Terminology
  3.2 Application of Definitions to Proteins: Is It a Drug or a Biologic?
  3.3 Regulatory Approval Path for Proteins
4 Regulation of Generic Drugs
  4.1 History of Generic Drug Legislation in United States
  4.2 Approval Process for Generic Drugs
  4.3 Application of Generic Regulations to Biologics
5 Legal Arguments Related to Follow-On Proteins
  5.1 Constitutionality of 505(b)(2) Process for Drugs
  5.2 Constitutionality of 505(b)(2) Process for Follow-On Proteins
  5.3 Applicability of 505(j)(1) or ANDA Process to Biogenerics
  5.4 Current Rules Relating to Bioequivalence of Generic Drugs
  5.5 Statutory Authority
6 Scientific Issues Related to Follow-On Proteins (Data Requirements)
  6.1 "Sameness" as per Orphan Drug Regulations
  6.2 "Sameness" as per Postapproval Change Guidances

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

7 Proposed Regulatory Paradigm: Case Studies
7.1 Case Study 1: Fortical [Calcitonin-Salmon (rDNA Origin)]
7.2 Case Study 2: Omnitrope [Somatropin (rDNA Origin)]
7.3 Case Study 3: Generic Salmon Calcitonin
8 Summary and Conclusions
References

## 1 INTRODUCTION

The ongoing need to provide the U.S. population with cost-effective pharmacological therapies has led to an emergent public health initiative in this country, namely for generic versions of therapeutic proteins. Greater access to generic drugs was made possible by the passage of the 1984 Drug Price Competition & Patent Term Restoration Act, commonly referred to as Hatch–Waxman. Generics have historically afforded considerable savings to the American consumer in need of prescription medication. Ten years after the Hatch–Waxman amendments, the Congressional Budget Office estimated that purchasers saved a total of $8–$10 billion on prescriptions at retail pharmacies by substituting generic drugs for their brand-name counterparts in 1994 [1]. To put those numbers in the context of today's pharmaceutical landscape, a recent report issued by the U.S. Department of Health and Human Services estimates that generic drugs constitute 63% of the total prescription medicines sold in the United States [2]. This same report suggests that generic drugs cost approximately 11% of the total cost of branded pharmaceuticals (on a per-dose basis).

At the same time, the development and use of therapeutic proteins have increased dramatically, with more than 850 biotechnology drug products and vaccines currently in trials [3]. Further, it is estimated that by the year 2010 nearly one-half of all newly approved medicines will be of biological origin [4]. The industrial financial incentives for the pursuit of follow-on biologics (heretofore termed biogenerics) are substantial with sales of biotechnology medicines in the United States rising 17% to approximately $30 billion in 2005 and growing at an annual rate of about 20% thereafter [3].

Not unexpectedly, the U.S. Food and Drug Administration (FDA) is experiencing mounting pressure to progress the cause of biogenerics. In a letter dated February 10, 2006, Senators Henry Waxman and Orrin Hatch (authors of the original "generic" legislation) urged the FDA to develop and implement clear guidelines for the approval of follow-on biological products for certain well-characterized proteins like insulin and human growth hormone (HGH) [5]. Additionally, recent litigation has compelled the FDA to take action on a pending drug application for a follow-on protein (FOP) drug product [Omnitrope, somatropin (recombinant DNA, rDNA origin)] [6].

A significant barrier to the emergence of "biogenerics" is the absence of a clear, efficient abbreviated pathway for approval. This hurdle is linked to significant scientific and legal issues in the United States in terms of how proteins are classified (drug vs. biologic) and subsequently regulated as well as how "generics" are tradi-

tionally defined in terms of equivalence and substitutability. However, an examination of the vast array of biologicals on the market today reveals that not all proteins are created equal. This range of complexity may provide an opportunity for stepwise progress on the regulatory front. This chapter presents the background to this multifaceted issue and examines the key regulatory challenges facing biogenerics today. An appropriate regulatory paradigm for the approval of FOPs is proposed and supported though a discussion of recent case studies.

### 1.1 Emergence of Biotechnology Industry

The explosion of scientific advances over the last quarter century has spawned the biotechnology industry and whole new classes of therapeutic agents for the treatment and prevention of disease. In October of 1982, the FDA approved the first protein-based therapeutic created by DNA technology in the form of Humulin (recombinant insulin). Developed by Eli Lilly & Co., with technical assistance from Genentech, Humulin is indicated for the treatment of diabetes. At the time, the use of recombinant technology was somewhat limited to the production of smaller, nonglycosylated proteins such as insulin (51 amino acids) and HGH (191 amino acids) using bacterial hosts. The seminal discovery by Columbia's Richard Axel of the process of cotransformation enabled complex protein production and glycosylation and thus spurred the emergence of the modern biotechnology industry [7].

The phenomenal growth observed in the biotechnology sector is notable in terms of the extraordinary number and diversity of therapeutic peptides and proteins that have been developed within a period of only about 20 years. Examples of therapeutic proteins in current use include cytokines, clotting factors, vaccines, and monoclonal antibodies, as illustrated in Table 1 [8].

As presented in Table 2, many of these "early" biotechnology products have reached the end of their period of patent exclusivity [4–9]. Thus, it is appropriate to now consider the next steps in the "life cycle" of these products as potential generic drugs.

### 1.2 Challenges Facing "Biogenerics"

The diversity and complexity of biologic molecules that drive their utility as therapeutic agents also contribute to the difficulty in classifying them as pharmacological entities, namely, whether they are drugs or biologics. This difficulty in classification is of profound importance since there are fundamental differences in how the FDA regulates drugs and biologics.

To appreciate the current challenges facing the pharmaceutical and biotechnology industry, it is informative to review the historical background associated with the classification and regulation of biologics in the United States, particularly in the context of the nation's evolving drug regulation system.

## 2 HISTORY OF BIOLOGICS REGULATION IN UNITED STATES

Due to the scientific limitations of the early to mid-1900s, significant differences existed between the approaches taken to manufacture and analyze biologics and

**TABLE 1 Examples of Therapeutic Peptide and Protein Molecules Currently Marketed in United States**

| | |
|---|---|
| Peptides | ***Antibiotics:*** bacitracin, bleomycin, gramicidine, capreomycin |
| | ***Hormones:*** corticotropin, glucagon, gonadrolein HCl, leuprolide acetate, histrelin acetate, oxytocin, secretin, goserelin acetate, vassopressin |
| | ***Others:*** polymixin B, eptifibatide, cyclosporine |
| Nonglycosylated proteins | ***Interleukins:*** andresleukin (IL-1), denileukin diftitox (fusion, protein-IL-2+ DT), anakinra (IL-2) |
| | ***Interferons:*** interferon alpha-n1, interferon alpha-n3, interferon alpha-2a, peg interferon alfa-2b, interferon alfacon-1, Interferon alpha-2b, interferon beta-1b, interferon gamma-1b, |
| | ***Enzymes/inhibitors:*** anistreplase, asparaginase, lactase, trypsin, alpha-1 proteinase inhibitor, urokinase, deoxyribonuclease, fibrinolysin, chymotrypsin, pancreatin, papain, urokinase |
| | ***Growth factors/hormones:*** Filigrastim pegfilgrastim, somatropin, becaplermin, somatrem, menotropins |
| | ***Antithrombotic agents:*** thrombin, fibrinogen, hirudin, hirulog, fibrin |
| | ***Others:*** insulin, gelatin, prolactin, albumin (human), hemoglobin, collagen |
| Glycosylated proteins | Interferon beta-1a |
| | ***Antithrombotic agents:*** alteplase, drotrecogin alfa, antithrombin III |
| | ***Antianemic:*** darbopoetin alfa, erythropoietin |
| | ***Growth hormones:*** follitropin alpha, follitropin beta, chorionic gonadotropin (Human) |
| | ***Immuno globulins (IG):*** pertusssis IG, rabies IG, tetanus IG, hepatitis B IG, varicella zoster IG, rho(D) IG, normal immune globulin, lymphocyte anti-thymocyte, IB (equine) |
| | ***Coagulation factors:*** factor VII antihemophilic factor, factor IX (human, recombinant) |
| | ***Factor VIII (others):*** etanercept (CSF), sargramostim (TNF) |
| Monoclonal antiobodies | avciximab, alemtuzamub, basiliximab, gentuzumab, satumomab, inflixibam, palivizumab |

drugs. This reality led to the creation of separate and distinct regulatory pathways for drugs and biologics. As noted earlier, the developments in analytical chemistry and improvements in process technologies have, in recent times, blurred the lines between drug and biologic drug development. In the current era of pharmaceutical development and standards harmonization, one might question the continued need for two distinct pathways. Recognizing the shifting paradigm of drug development, the history of biologics regulation is discussed below in two parts: early history and present day.

### 2.1 Early Biologics Regulation (1800s–1990s)

This country's earliest experience with biologics dates back to the infectious scourges of the late 1800s and early 1900s when epidemics of typhoid, yellow fever, smallpox, diphtheria, and tuberculosis were being battled by new advances in immunology. The discovery and development of vaccines and antitoxins led to the creation of a

**TABLE 2 Patent Expiration Dates for U.S. Marketed Biologics**

| Brand Name | Generic Name | Indication | Company | Patent Expiry |
|---|---|---|---|---|
| Humulin | Recombinant insulin | Diabetes | Eli Lilly | Expired |
| Nutropin | Somatropin | Growth disorders | Genentech | Expired |
| Abbokinase | Eudurase urokinase | Ischaemic events | Abbott | Expired |
| Ceredase | Alglucerase | Gaucher disease | Genzyme | Expired |
| Cerezyme | Imiglucerase | Gaucher disease | Genzyme | Expired |
| Streptase | Streptokinase | Ischaemic events | AstraZeneca | Expired |
| Intron A | IFN-α-2b | Hepatitis B and C | Biogen/Roche | Expired |
| Serostim | Somatropin | AIDS wasting | Serono | Expired |
| Humatrope | Somatropin | Growth disorders | Eli Lilly | Expired |
| Geref | Sermorelin | Growth hormone deficiency | Serono | Expired (2004) |
| Synagis | Palivizumab | Respiratory syncytial virus | Abbott | Expired (2004) |
| Novolin | Human insulin | Diabetes | Novo Nordisk | 2005 |
| Protropin | Somatrem | Growth hormone deficiency | Genentech | 2005 |
| TNKase | Tenecteplase TNK-tPA | Acute myocardial infarction | Genentech | 2005 |
| Actimmmune | IFN-γ-1b | Chronic granulomatous disease; malignant osteoporosis | InterMune | 2005, 2006, 2012 |
| Activase, Alteplase | tPA | Acute myocardial infarction | Genentech | 2005, 2010 |
| Proleukin | IL-2 | HIV | Chiron | 2006, 2012 |
| Epogen, Procrit, Eprex | Erythropoietin | Anemia | Amgen | 2013 |
| Neupogen | Filgrastim (G-CSF) | Anemia, leukemia, neutropenia | Amgen | 2015 |

*Note:* Based on our search of available patent sites for only the reference product.
IFN-Interferon; tPA-Tissue Plasminogen Activator, IL-interleukin; HIV-Human Immunodeficiency Virus; G-CSF- Granulocyte-Colony Stimulating Factor; TNKase- Tenecteplase.

whole new "biopharmaceutical" industry. As demand increased, the pharmaceutical manufacturers responded and in turn supplanted the government's role in the public supply of vaccines (per Vaccine Act of 1813) [10]. Unfortunately, the commercialization of vaccines by smaller, less experienced, and likely less scrupulous manufacturers led to problems. Similar to the history of drug regulation, early advances in biologics regulation could be characterized as responsive rather than proactive. Change often occurred following tragedy and the result of government's attempt to respond. Some of the key milestones of early biologics regulation are summarized in Table 3. The following years saw many administrative changes in terms of the specific governmental agency responsible for regulating biologics, but with few substantive changes to the regulations themselves.

**TABLE 3 Key Milestones in Early Biologics Regulation**

| | |
|---|---|
| 1901 | Ten children died in St. Louis from administration of tetanus-contaminated diphtheria antitoxin. In this case, no safety testing had been performed prior to use. |
| 1902 | Biologics Control Act (BCA) signed into law:<br>• Authorizing the regulation of commercial viruses, serums, toxins, and analogous products<br>• Requiring the licensure of biologics manufacturers and establishments<br>• Providing governmental inspectional authority<br>• Making it illegal for the commercial distribution of product not manufactured and labeled in accordance with the act |
| 1906 | Pure Food and Drug Act enacted (precursor of modern-day drug regulation). Lack of mention of biologics as a class effectively represents first distinction between drug and biologic regulation. |
| 1919 | BCA amended:<br>• Required reporting of changes in equipment, manufacturing processes, personnel; establishment of formal quality control procedures; and submission of samples for regulatory inspection and approval for release<br>• Recognized potential that slight changes to manufacturing conditions (raw materials, process, personnel, etc.) could have significant and adverse effect on product quality<br>• Required strict control of input (environment and manufacturing conditions) rather than end-stage testing of quality attributes due to limitations in analytical methodology to detect these effects |
| 1937 | Elixir sulfanilamide, containing the poisonous solvent diethylene glycol, kills 107, many of whom are children. |
| 1938 | Food, Drug and Cosmetic Act (FDCA) enacted:<br>• Established concept of "new drugs" requiring proof of safety prior to marketing<br>• Required submission of an investigational new Drug (IND) application prior to clinical use of an experimental drug in humans<br>• Required approval of a new drug application (NDA) prior to commercial sale of drugs<br>• Granted federal government power of seizure of misbranded or adulterated drugs<br>• Defined "drugs" comprehensively; not excluding potential of "biologics" to function as drugs |
| 1941 | • Approximately 300 deaths and injuries result from distribution of sulfathiazole tablets tainted with the sedative phenobarbital.<br>• Insulin Amendment passed to require FDA testing/certification of purity and potency. |
| 1944 | Public Health Service (PHS) Act enacted to consolidate and codify previous biologics laws:<br>• Outlined licensing requirements for biologics—for both product (product licensing application, or PLA) and establishment where the product was manufactured (establishment licensing application, or ELA)<br>• Required submission of samples of each manufactured lot of all biologicals for government testing and certification prior to commercial release<br>• Required sponsors to own all of manufacturing facilities, effectively eliminating multiparty or contract manufacturing |

### 2.2 Modern Biologics Regulation (1990s–Today)

Whereas early biologics regulation was grounded by technical limitations, modern biologics regulation is driven by tremendous advances in scientific knowledge. Development of analytical tools and techniques has dramatically increased the ability to characterize proteins and substantiate the structure, composition, and function of the therapeutic molecule. These advances enable the detection of small differences in molecular weight; elucidation of primary, secondary, and tertiary protein structures; detection and quantification of posttranslational modifications (i.e., patterns of glycosylation); and improved understanding of structure–function relationships and potential immunogenic responses. Simultaneously, developments in the fields of pharmaceutical and biotechnological manufacturing have greatly improved process efficiency and control. This recent technological evolution has had a direct impact on biologics regulation as reflected below in several key events:

- In 1995, the FDA agreed to eliminate lot testing requirements for certain highly characterized products once the company's ability to consistently manufacture product of acceptable quality was established.
- In 1996, the FDA and Congress dismantled the dual-licensing process, requiring the submission of a single BLA (biologics license application), making the content and format of a biologics application similar to that required for new drug applications (NDAs).
- In 1996, the Center for Biologics Evaluation and Research (CBER) liberalized its definition of "legal manufacturer" and eliminated many of the barriers to cooperative, multiparty manufacturing arrangements.
- In 1997, Congress passed a noteworthy piece of legislation affecting modern pharmaceutical regulation in the Food and Drug Modernization Act (FDAMA). Among the many goals of the act was to harmonize the drug and biologic approval processes.

In fact, current pharmaceutical/regulatory initiatives appear to extract the best practices from biologic and drug approaches which can apply equally to both classes of products:

- The Quality Systems Approach and GMPs for the 21st Century, two initiatives being pursued by the FDA for drugs and devices, emphasize the utility of building quality into the process, consistent with the strict control of "input factors" seen in early biologic regulation.
- Initiatives such as Process Analytical Technologies build on the concept of conventional drug product testing using increasingly sophisticated analytical techniques to provide continuous process monitoring and finished-product quality assurance of multiple pharmaceutical dosage forms.
- The current global initiative to harmonize electronic submission format and content requirements effectively creates one standard data package for drugs or biologics. Thus, the eNDA (electronic new drug application) or eBLA (electronic biologics license application) will eventually be replaced by the eCTD (electronic common technical document).

## 3 REGULATORY CLASSIFICATION OF PROTEINS

Despite the blurring of lines between drugs and biologics, there remain two different mechanisms to bring protein drug products to the U.S. marketplace. The choice of approval framework is dependent on the protein's classification as a drug or biologic. The history of this regulatory distinction is rooted in the technical differences between small-molecule drugs and macromolecular biologics. Traditionally, drugs were characterized as having well-defined chemistry. Conversely, biologics were large, complex macromolecules whose active moiety defied characterization and quantitation. By necessity, different means of assuring the safety and efficacy of these therapeutic products were required at the time. The modern-day consequence is a legal system that distinguishes between proteins as drugs and proteins as biologics. The distinction is based on statutory definitions as well as historical precedent and has implications in terms of the approval pathways for original and follow-on products.

### 3.1 Definitions and Key Terminology

*Drugs* are defined by the U.S. Food and Drug Act [FD&C Act, 21 U.S.C. 321(g)(1)]*by function* as any article Federal Food, Drug and Cosmetic Act (a) intended for use in the diagnosis, cure, mitigation, treatment, or prevention of disease in man or animals and (b) intended to affect the structure or function of the body [11].

*Biologics* as a class may be regulated as drugs but are defined within the Public Health Service Act [PHSA, 42 U.S.C. 262(a)] *by category* as "a virus, therapeutic serum, toxin, antitoxin, vaccine, blood, blood component or derivative, allergenic product, or analogous product, or arsphenamine (or any other trivalent organic arsenic compound), applicable to the prevention, treatment, or cure of diseases or injuries of humans" [12].

A cursory examination of these definitions reveals that they are not mutually exclusive, leading to confusion about how to appropriately and consistently apply them. This point is illustrated when one reviews the history of how the FDA has categorized and subsequently regulated these drugs and biologics as shown below.

### 3.2 Application of Definitions to Proteins: Is It a Drug or a Biologic?

The answer to this fundamental question is not straightforward and has evolved over time. Historically, some natural-source-derived proteins such as insulin, hyaluronidase, menotropins, and Human Growth Hormone (HGH) have been regulated as drugs. While other natural-source-proteins such as blood factors were regulated as biologics. When recombinant proteins and monoclonal antibodies began development in the 1970s–1980s, these were regulated as follows:

1. By the Center for Drug Evaluation and Research (CDER) under the Food, Drug and Cosmetic Act (FDCA) as drugs when they were hormones such as insulin, HGH, and parathyroid hormone (PTH) derivatives
2. By the CBER under the PHSA as biologics when they were cytokines or blood factors such as factor VIII for hemophilia

**TABLE 4 FDA Center Regulatory Responsibility for Therapeutic Biological Products**

| CDER | CBER |
|---|---|
| Monoclonal antibodies (in vivo use) | Cellular product, including products composed of human, bacterial, or animal cells |
| Proteins intended for therapeutic use: | Vaccines |
| Cytokines (e.g., interferons) | Allergenic extracts |
| Enzymes (e.g., thrombolytics) | Antitoxins, antivenoms, venoms |
| other novel proteins except those assigned to CBER | Blood, blood components, plasma-derived products (e.g., albumin, immunoglobulins, clotting factors, fibrin sealants, proteinase inhibitors), recombinant and transgenic versions of plasma derivatives |
| Immunomodulators (nonvaccine, nonallergenic) | |
| Growth factors, cytokines, some hormones and monoclonal antibodies intended to mobilize, stimulate, decrease, or otherwise alter the production of hematopoietic cells in vivo | |

As other recombinant proteins and monoclonal antibodies came under development, the CBER held primary responsibility for this review, with the CDER retaining responsibility for hormones such as insulin and HGH. However, in 2003 all therapeutic proteins were transferred from the CBER to the CDER. This reassignment of review responsibility did not impact the legal classification of these protein products, such that the Center for *Drug* Evaluation and Research assumed responsibility for the review and approval of *biologics* approved under Section 351 of the PHSA.

The basic distribution of these therapeutic biologics to the respective FDA center is reflected in Table 4; however, many of the current complex biotechnology-derived products do not fit neatly into accepted definitions and require case-by-case classification [13].

### 3.3 Regulatory Approval Path for Proteins

The relevance of the preceding discussion becomes important with the understanding that therapeutic proteins classified as drugs are governed under a different set of laws than those classified as biologics. Drugs are approved via submission of NDAs under Section 505 of the FD&C Act, while biologics are supported by BLAs under the PHSA. These two approval paths are similar in terms of application content, that is, requirement of complete reports of clinical safety and efficacy data to support approval. However, only the drug regulation, that is, Section 505 of the FD&C Act, has been amended to outline an abbreviated approval mechanism for generic products.

## 4 REGULATION OF GENERIC DRUGS

### 4.1 History of Generic Drug Legislation in United States

In 1984, Congress responded to America's need for safe, affordable medicines by passing a pivotal piece of legislation, The Drug Price Competition and Patent Term

Restoration Act (Hatch–Waxman amendments). The intent of this act was to effectively balance the need to encourage pharmaceutical innovation with the desire to accelerate the availability of lower cost alternatives to approved drugs. The act also sought to eliminate unnecessary or redundant clinical testing to protect patients (reduce the number of patients in need receiving placebo in controlled clinical trials) and conserve industry and agency resources. To accomplish the goal of faster to market, cheaper alternatives, the amendments stipulated the following [14]:

- *For Innovator Companies* The act encouraged continued innovation, research, and development activities by providing manufacturers with meaningful incentives in the form of patent protection/restoration and marketing exclusivity, thus allowing them to recoup some of their investments.
- *For Generic Companies* The act provided access to certain innovator information without the threat of legal action via patent infringement suits (safe harbor provisions), allowing generics the opportunity to prepare for market introduction prior to the expiration of patent/exclusivity terms. This effectively limited the period of innovator exclusivity to the statutory timelines.

### 4.2 Approval Process for Generic Drugs

The act served as a boon to the generic industry by paving the path to abbreviated and accelerated drug approvals. From a legal perspective, the Hatch–Waxman amendments modified Section 505 of the FD&C Act to create two new abbreviated approval pathways (see Table 5) [14].

In essence, the abbreviated NDA (ANDA) and 505(b)(2) processes allow generic manufacturers the ability to rely on what is already known about the drug and refer to the agency's finding of safety and efficacy for the innovator. For an ANDA, the generic product must meet certain criteria related to bioequivalence and product sameness. However, a 505(b)(2) application often describes a drug with substantial differences to the innovator (which would seem more closely related to FOPs).

### 4.3 Application of Generic Regulations to Biologics

A central question is "Do biologics fall under the provisions of the Hatch–Waxman Act?" Since the Hatch–Waxman Act specifically amended the FD&C Act, biologics

**TABLE 5 Description of NDA Approval Mechanisms**

| | |
|---|---|
| Traditional path | 1. 505(b)(1)—Application that contains full reports of investigations of safety and effectiveness to which sponsor has right of reference (stand-alone NDA) |
| Abbreviated path | 2. 505(b)(2)—Application that contains full reports of investigations of safety and effectiveness, where the sponsor relies on studies conducted by someone else to which the sponsor does not have right of reference |
| Abbreviated path | 3. 505(j)(1)—Abbreviated new drug application (ANDA) containing information to show the product is a duplicate of an already approved drug product |

approved via a BLA under the PHSA are not covered by this legislation nor does the PHSA have similar provisions for biogenerics. However, those few therapeutic proteins approved via Section 505 of the FDCA as NDAs are covered by the Hatch–Waxman amendments and thus are *legally* considered appropriate for filing a 505(b)(2) or 505(j)(1) application. For simple, well-characterized peptides and proteins regulated under Section 505 of the FD&C Act, mechanisms are already in place to bring FOPs to the market. In fact, several FOPs have already been approved by the FDA, including GlucaGen (glucagon recombinant for injection), Hylenex (hyaluronidase recombinant human), Hydase and Amphadase (hyaluronidase), Fortical (calcitonin salmon recombinant) Nasal Spray, and Omnitrope [somatropin (rDNA origin)] [15]. Further details related to the latter two are presented in the discussion of actual case studies.

## 5 LEGAL ARGUMENTS RELATED TO FOLLOW-ON PROTEINS

The legal arguments regarding the approval of biogenerics relate to several different aspects of drug/biologics law.

### 5.1 Constitutionality of 505(b)(2) Process for Drugs

The agency's authority to grant approval of drugs via the 505(b)(2) process has previously been challenged by several companies. The nature of these challenges has questioned the FDA's right to use proprietary information of the innovator in support of another company's drug approval. Recall that the 505(b)(2) process allows a company to use data for which it does not have right of reference (i.e., another company's safety and efficacy data) in support of its own application. The FDA's long-standing interpretation of the statute seems firm and well founded in precedent since over 80 applications for drugs have been approved via the 505(b)(2) route since its inception with indications ranging from cancer pain to Attention Deficit Disorder (ADD) [16].

### 5.2 Constitutionality of 505(b)(2) Process for Follow-On Proteins

The constitutionality issues related to FOPs are similar to those mentioned above for drugs, namely protection of proprietary information and intellectual property rights. Some critics opine that issues unique to FOPs create additional legal hurdles. For example, the rules pertaining to the disclosure of safety and effectiveness information are different for biologics licensed under the PHSA and drugs approved under the FDCA. When the rules were originally written (1974), it was thought that safety and effectiveness for one biologic would not support the licensure of another. So these data were deemed not to be protected trade secrets and could be publicly disclosed immediately after issuance of the biologic's license [see 21 CFR 601.51(e), 1974]. However, since this language applies strictly to the PHSA, it has no bearing on discussions related to the 505(b)(2) process.

In other public challenges, opponents argue that the unique and complex nature of biologics and the close relationship between their method of preparation and clinical attributes require that the FDA use and disclose the manufacturing methods

and process information contained in an innovator's application. Further, this use and disclosure would violate Trade Secret and Constitutional Law (Fifth Amendment "taking clause") [17, 18].

The concept of "the product is the process" may have been applicable to early biologics, but current capabilities allow the chemical, biologic, and functional comparison of well-characterized protein drugs. The follow-on manufacturer need not necessarily utilize the identical method of manufacture or proprietary technology to reproduce a follow-on biologic with similar clinical safety and efficacy. Additionally, it is important to distinguish between the regulatory requirements for approval of an actual generic protein (duplicate of innovator; see discussion below) and those associated with a 505(b)(2), which requires a showing of similarity between two products. Any differences between the two would need to be adequately supported by bridging studies and appropriate clinical and/or nonclinical data.

The FDA has confirmed this interpretation in its response to petitions filed regarding FOPs (both in general and targeted to specific applications). The FDA has clearly said, "the use of the 505(b)(2) pathway does not entail disclosure of trade secret or confidential commercial information, nor does it involve unauthorized reliance on such data" [18].

### 5.3 Applicability of 505(j)(1) or ANDA Process to Biogenerics

Biogenerics per se, that is, protein drug products approved via 505(j)(1), would need to demonstrate their bioequivalence to the innovator protein. However, due to their complexity and heterogeneity, the classical biopharmaceutical principles upon which the current ratings of therapeutic equivalence are based do not apply in their current language to complex macromolecules. For example, due to the nature and complexity of an immunogenic response, one concern would be if traditional bioequivalence appropriately addresses the complex safety issues associated with biologics.

### 5.4 Current Rules Relating to Bioequivalence of Generic Drugs

The list of approved drug products with therapeutic equivalence (Orange Book) was originally intended as an information source to states seeking formulary guidance [19]. The list provides the FDA's recommendations as to which generic prescription drug products are acceptable substitutes for innovator drugs. The term *innovator* is used to describe the reference listed drug, or RLD [21 CFR 314.94(a)(3)], upon which an applicant (generic) relies in seeking approval of its ANDA. In layman's terms the RLD describes the original NDA-approved drug and is often referred to as the "pioneer" drug.

Under the Drug Price Competition and Patent Term Restoration Act of 1984, manufacturers seeking approval to market a generic drug need to submit data to the FDA demonstrating that their proposed drug product is bioequivalent to the pioneer (innovator) drug product. A major premise underlying the 1984 law is that bioequivalent drug products are therapeutically equivalent, will produce the same clinical effect and safety profile as the innovator product, and are therefore, interchangeable [19].

So how would FOPs be classified using conventional definitions of bioequivalence? To answer this question, it is necessary to review current legal definitions of bioequivalence terms [19]:

- Two products are bioequivalent in " the absence of a significant difference in the rate and extent to which the active ingredient or active moiety in *pharmaceutical equivalents* or *pharmaceutical alternatives* becomes available at the site of drug action when administered at the same molar dose under similar conditions in an appropriately designed study" [21 CFR 320.1(e)]. An appropriately designed comparison could include (1) pharmacokinetic (PK) studies, (2) pharmacodynamic (PD) studies, (3) comparative clinical trials, and/or (4) in vitro studies.
- *Pharmaceutical equivalents* are those drug products which are formulated to contain the same amount of active ingredient in the same dosage form to meet the same (compendial or other applicable) standards of quality.
- *Pharmaceutical alternatives* are drug products that contain the same therapeutic moiety, or its precursor, but not necessarily in the same amount or dosage form. Drug products are considered to be *therapeutic equivalents* only if they are pharmaceutical equivalents and if they can be expected to have the same clinical effect and safety profile when administered to patients under the conditions specified in the labeling. Although pharmaceutical alternatives may ultimately be proven bioequivalent, given their differences they are not automatically presumed to be.

Given these definitions, FOPs would likely be considered pharmaceutical alternatives if one presumes that pioneer and follow-on proteins are identical at a precursor stage, prior to potential post-translational modification. This presumption may also be consistent with the similarity standard the agency applies to ascertain orphan drug status (see discussion in Section 6). Follow-on proteins cannot be considered to be therapeutic equivalents since they are not pharmaceutical equivalents and cannot be expected to have the same clinical effect and safety profile in the absence of testing. This assertion is supported by the following:

- The potential impact of how posttranslational modifications, such as glycosylation, can directly impact protein conformation and subsequently affect biological activity, including the overall safety and efficacy of the drug product.
- An underlying premise of bioequivalence assessments is a clearly defined pharmacokinetic/pharmacodynamic relationship; however, the relation between blood levels and effect is less clearly established for proteins [20].

Consequently, within the current regulatory framework, FOPs are unique products that may be "similar" but are not the same as innovator proteins, consistent with their approval via a 505(b)(2) pathway. This interpretation is supported by the FDA's designation of Omnitrope as having a BX rating in the Orange Book. The code BX in the Orange Book refers to drug products for which the data are insufficient to determine therapeutic equivalence as compared to a therapeutic rating of A indicative of interchangeability. This concept of similarity is also consistent with the definitions proposed by the European Agency for the Evaluation of Medicinal Products (EMEA) for generic versions of proteins [21]:

> Bio-similar products: second and subsequent versions of biologics that are independently developed and approved after a pioneer has developed an original version. Bio-similar products may or may not be intended to be molecular copies of the innovator's product; however, they rely on the same mechanism of action and therapeutic indication.

### 5.5 Statutory Authority

Unlike the FDCA, which affords therapeutic protein drugs the legal pathway of abbreviated drug approval for a FOP, the PHSA currently has no similar provisions. Such a pathway for approval or licensure of FOP products under the PHSA would require new legislation and recent congressional developments suggest that work is underway to create this statutory pathway.

Legislation proposed on September 29, 2006, by U.S. Representative Henry Waxman (D-CA) and Senator Charles Schumer (D-NY) seeks to amend the PHSA to authorize the FDA to approve abbreviated applications for biologic products that are "comparable" to previously approved (brand name) biologic products. Entitled The Access to Life-Saving Medicine Act, this bill outlines a process by which the FDA could determine, on a product-by-product basis, the studies necessary to demonstrate comparability of a FOP product to a brand name product and assure its safety and effectiveness. The act allows for an applicant to seek interchangeability with a brand name product, recognizing that the extent of data to support such a designation must be discussed with the FDA. To encourage the development of interchangeable products, the bill would authorize tax incentives and periods of marketing exclusivity. The bill would also seek to create an improved process to facilitate early resolution of patent disputes which might otherwise delay competition [22].

## 6 SCIENTIFIC ISSUES RELATED TO FOLLOW-ON PROTEINS (DATA REQUIREMENTS)

The challenge of FOPs demonstrating similar quality, safety, and efficacy to the innovator product relates to the poor predictability of physicochemical characteristics and biologic activity. For example, there are several different interferon-α and erythropoietin α and β products currently on the market. These variants are characterized by differences in sequence, glycosylation pattern, and in vitro measures of specific activity; however, their clinical safety and efficacy profiles are considered similar [20].

In contrast, different formulations of insulin and growth hormone containing the same active ingredient exhibit significant differences in bioavailability [20]. Additionally, the inability to adequately predict immunogenic responses from in vitro data or animal studies remains a concern.

The answer to the challenge is that generic manufacturers must go through a similar process of in-depth characterization, including identification of critical structural elements of the product (structure/function) when developing a FOP. Although the regulatory standards for demonstrating similarity are currently undefined, some insight can be gleaned from consideration of FDA expectations in terms of granting orphan drug status to similar proteins and assessing postapproval Chemistry, Manufacturing and Controls (CMC) changes for innovator proteins.

### 6.1 "Sameness" as per Orphan Drugs Regulations

The Orphan Drug Act of 1983 was implemented in response to the government's concern that viable treatments for rare diseases were not being explored due to excessive costs of drug development in comparison to the relatively small popula-

tion of potential users (and sales). Orphan drugs are (a) those used to treat rare diseases, defined by the act as affecting <200,000 persons in the United States, or (b) those drugs whose development costs would not be recovered through sales of the drug. To encourage development, the government authorized incentives in the form of marketing exclusivity (seven years), tax credits, protocol assistance, and grants/contracts, with the first being of primary importance to most drug sponsors.

Since exclusivity is awarded only to the first designated product to obtain approval for a given drug/indication, competition is fierce. No approval would be given to a subsequent sponsor's application for the same product/indication unless it was shown to be clinically superior (i.e., not the same). Thus, the agency needed to develop criteria upon which it would make these determinations.

In 1992, the FDA's orphan drug regulations first established the conditions under which the agency could determine product "sameness" of protein drugs and therefore take action to block the approval of a second orphan drug product: "two protein drugs would be considered the same if the only differences in structure between them were due to post-translational events, or infidelity of translation or transcription, or were minor differences in amino acid sequence; other potentially important differences, such as different glycosylation patterns or different tertiary structures, would not cause the drugs to be considered different unless the differences were shown to be clinically superior" [23]. It should be noted that there may exist exceptions to this rule that depend on the interpretation of each individual case. For example, Eli Lilly & Co. successfully received orphan drug status in the late 1980s for the naturally occurring HGH to compete with the previously marketed Met-HGH, which only differed in the N-terminal methionine.

The support for clinical superiority could be based on evidence of greater effectiveness and increased safety or represent a "major contribution to patient care." In short, orphan drug regulations utilize clinical data to demonstrate product differences. Examples include [23]:

- 1996: Biogen's Avonex (interferon β) was considered to be clinically superior to Berlex's Betaseron based on improved safety (fewer site injection reactions).
- 1999: In a law suit involving generic paclitaxel, Baker Norton, challenged the FDA's sameness determinations based on active moiety alone, arguing that factors such as formulation and labeling should be considered. The challenge was unsuccessful.
- 2002: Serono's Rebif (interferon $\beta_{1a}$) was awarded exclusivity based on the clinical demonstration of improved efficacy (reduced Multiple Sclerosis (MS) exacerbations).

Therefore, it would appear that the orphan drug regulations provide some flexibility to the sponsor (generic) in establishing product sameness but also reaffirm the important role of clinical data in supporting product safety and efficacy.

### 6.2 "Sameness" as per Postapproval Change Guidances

Guidelines for supporting postapproval changes to the chemistry, manufacturing, and controls of approved products (SUPAC guidances) take a somewhat different approach to establishing sameness. In essence, the SUPAC guidelines reflect risk

management practices in evaluating the potential of certain CMC changes to impact the identity, strength, quality, purity, and potency of the product as they may relate to overall safety and efficacy.

A long-held contention within the biologics industry is that the product is the process and, by extension, change is strongly discouraged. Without qualification, this rather dated thinking is inconsistent with the flexibility required in managing change throughout the life cycle of a product. Further, this thinking may serve to discourage the implementation of advanced technologies designed to improve not only efficiency but also product quality. Even current biologics regulations recognize the need to accommodate change; 21 CFR 601.12 (for biologics) states that for changes in the product, production process, quality controls, equipment, facilities, and so on, an applicant must assess the effects of the change and demonstrate through appropriate validation and/or other clinical and/or nonclinical laboratory studies the lack of adverse effect of the change on the identity, strength, quality, purity, or potency of the product as they may relate to the safety or effectiveness of the product.

In fact, many of the challenges that generic manufacturers face in demonstrating sameness of FOPs to reference listed drugs are similar to those encountered by innovators in managing the dynamic CMC life cycle of a product. One of the tools available to assess the potential impact of product differences is a comparability protocol. The FDA described its expectations of the data requirements necessary to support postapproval CMC changes to protein drug product and biologic products in a Guidance to Industry on the use of comparability protocols for such products issued in 2003 [24]. Underpinning the successful application of a comparability protocol are extensive product development and characterization.

***Initial Product Development*** Prior to undertaking any comparative analysis, a manufacturer must perform two critical steps. First, the manufacturer needs to conduct thorough process development and optimization of the therapeutic protein product. Second, the sponsor (generic or innovator) needs to prospectively examine the impact of changes to all critical processing parameters during the development phase and determine the minimum data requirements necessary to assure the absence of adverse impact to product quality, safety, or efficacy. The current state of technology provides us with better tools to more fully characterize the protein drug substance and drug product at all stages of production.

*Physicochemical Characterization and Process Development* Some of the key steps to process development and product characterization include:

- Production of a cell line/clone
- Identification and characterization of critical raw materials (media, resins, formulation excipients)
- Development of internal standards, in-process controls, product specifications
- Conduct of extensive pilot-scale manufacturing development: fermentation and downstream processing (separation and purification)
- Performance of process scale-up and optimization studies

**TABLE 6 Analytical Techniques for Physicochemical Characterization of Proteins**

| Parameter | Test |
|---|---|
| Primary structure | Amino acid sequencing, N-terminal Edman sequencing, peptide mapping |
| Higher order structure | CD, NMR, FTIR, Raman |
| Mass | LC-ESI-MS, MALDI-TOF-MS |
| Size | SDS-PAGE, DLS, SEC-MALLS |
| Hydrophobicity | RP-HPLC |
| Binding | Immunological binding |
| Sulfhydryl groups/disulfide bridges | Peptide mapping (under reducing and nonreducing conditions) |
| Glycan analysis: | |
| Monosaccharide analysis | HPLC, MS |
| Sialic acid content | HPLC |
| Molecular weight | MALDI-MS, ESI-MS |
| Impurity profile | |
| Process-related impurities | Immunoassay, HPLC, SDS-PAGE, MS, CD, capillary gel electrophoresis, size exclusion chromatography |
| • Cell substrate derived | |
| • Cell culture derived | |
| • Downstream derived | |
| Product-related impurities | |
| • Truncated forms | |
| • Other modified forms (i.e., deamidated, isomerized) | |
| • Aggregates | |
| Evaluation of stability | HPLC |

CD, Circular Dichroism; NMR, Nuclear Magnetic Resonance; FTIR, Fourier transform infrared spectroscopy; LC-ESI-MS, Liquid chromatography electrospray ionisation mass spectrometry; MALDI-TOF-MS, Matrix-assisted laser desorption ionization-time of flight-mass spectrometry; SDS-PAGE, Sodium dodecyl sulfate (SDS) polyacrylamide gel electrophoresis; DLS. Dynamic light scattering; SEC-MALLS, Size exclusion chromatography-multi-angle laser light scattering; RP-HPLC, Reversed phase-high performance liquid chromatography; HPLC, High performance liquid chromatography; MALDI-MS, Matrix-assisted laser desorption ionization mass spectrometry; ESI-MS, Electrospray ionisation mass spectrometry; MS, Mass spectrometry.

- Application of a comprehensive array of analytical techniques to fully characterize the drug product at each stage of development. Table 6 provides examples of methods to probe virtually every property of the protein and develop a fingerprint of the molecule.

*Other Testing Requirements* The need for additional supportive studies beyond physicochemical characterization will increase proportionately with the complexity of the protein drug. The entire battery of tests may not be required for each FOP but may include the following data, bioassay, preclinical (pharmacology/toxicology/pharmacokinetic/pharmacodynamic), clinical safety and efficacy, and immunogenicity. The nature, number, and size of the trials should relate directly to the particular drug/indication/patient population.

*Bioassay* A biological assay, or "bioassay," is an analytical procedure capable of measuring the biologic activity of a substance based on a specific functional, biologic

response of the test system. Bioassays should be predictive of clinical effect and are therefore used as a means of quantifying activity (in nonclinical manner) and ensuring efficacy throughout development. They are informative in equivalence studies to the extent that a change affects a part of the molecule, which in turn impacts the molecule's biologic activity. Bioassays may be based on animal models, in vitro cell lines, cell-based biochemical assays (i.e., kinase receptor activity), receptor binding assays, or enzyme assays. The selection of an appropriate bioassay is driven in part by the ability to demonstrate a correlation to clinical effect. An example of a predictive bioassay is the measurement of the antiviral activity of interferon as a function of its cytopathic effect on host cells [25].

*Nonclinical (Pharmtox, PK, PD)* In the context of FOPs, the original sponsor will have demonstrated what the molecule per se does to the body; however, since the formulation is likely different, nonclinical studies are useful in demonstrating a lack of adverse impact due to dosage form, route of administration, excipient changes, manufacturing contaminants, and supporting sameness of the active moiety.

Appropriate toxicology studies would include acute or subchronic testing in at least one relevant small animal species. Pharmacokinetic studies are highly useful in assessing the impact of changes in the manufacture of natural-source- and recombinant-derived proteins. Standard approaches used in bioequivalence studies [measurement of the area under the curve (AUC), $C_{max}$, $t_{max}$] can be used to make direct comparisons of innovator and follow-on profiles. Pharmacodynamic studies are similarly very informative. Direct comparison of innovator and follow-on products can be made by evaluating appropriate surrogate markers of efficacy (i.e., platelet aggregation following anticoagulation therapy).

*Clinical (PK, PD, Safety and Efficacy)* Human clinical studies can range in complexity from standard-design PK studies to complicated, long-term efficacy trials evaluating one or more indications in multiple populations. Human PK studies are used as the benchmark for establishing bioequivalence of conventional dosage forms. For traditional pharmaceuticals for which reliance on systemic exposure may not be suitable, PD or clinical safety and efficacy may be performed to show equivalence.

The appropriate clinical program is influenced by many factors, including the degree of molecular complexity of the particular protein and the extent of physicochemical characterization; the mode of action, indication(s), and use population(s); the presence of established structure–activity relationships and validated bioassays; and the results of preclinical testing.

As such, the nature and scope of each clinical support program need to be determined on a case-by-case basis in consultation between the sponsor and the regulatory agency.

*Immunogenicity* The observation of serious adverse events with the use of some recombinant and natural source proteins [i.e., pure red cell aplasia (PRCA) detected with erythropoietin use] has highlighted immunogenicity as a major issue for consideration when assessing within and between manufacturer changes [8]. Although the exact immunological mechanism responsible for the increased number of PRCA cases is unknown, it appears to be linked to a formulation change associated with Eprex, a European epoetin-α product. Replacement of the stabilizer human serum

albumin with polysorbate 80 and glycine correlated with a surge in PRCA reported cases [4]. An immunogenic effect may have no clinical impact or it could have serious clinical consequences as seen above. The immune response of the therapeutic protein should be fully characterized using both immunoassays which detect antibodies that bind to the drug as well as bioassays which detect neutralizing antibodies that might block the protein's desired biological effect. Ultimately, this testing needs to be performed in humans, as animal testing is not truly predictive of human immune response. Antibody detection techniques include enzyme-linked immunosorbent assay radioimmunoassay, (ELISA), and surface plasmon resonance [8].

***Comparability Testing to Demonstrate "Sameness"*** Following the developmental studies described above, comparative studies to directly evaluate pre- and post-change materials to one another and assess the impact of any process changes may be conducted. In a similar manner, comparative studies between pioneer drug and the follow-on can be used to systematically evaluate the impact of any differences between reference listed drug and proposed generic protein drug. When compiling information into an analytical characterization database, the data should be directly compared to the reference product and variation observed in multiple batches of test product (generic) should be similar to that of the reference innovator product.

The FDA's expectations in this regard are apparent in their description of the CMC data package supporting the comparability of Omnitrope to the innovator protein Genotropin. The FDA asserted [18]:

> Each biotechnology manufacturer, whether producing a new molecular entity or a follow-on product must independently develop its own cell expression, fermentation, isolation and purification systems for the active ingredient in its product. Thus, the manufacturing process for each active ingredient is unique to each manufacturer. Nevertheless, as Sandoz has demonstrated in its Omnitrope application, for this relatively simple recombinant protein, it is possible to determine that the end products of different manufacturing processes are highly similar, without having to compare or otherwise refer to the [proprietary] processes.

## 7 PROPOSED REGULATORY PARADIGM: CASE STUDIES

Based on the nature and complexity of therapeutic protein products, an approval pathway for follow-ons may require moving away from the traditional generic paradigm in place for small molecules and creating a biosimilar paradigm for complex molecules. The proposed regulatory paradigm for the approval of FOP products could be similar for protein drugs approved under Section 505 of the FDCA or licensed as biologics under the PHSA and mirror the current 505(b)(2) process. This pathway permits the sponsor and agency to determine exactly what studies are necessary to support the proposed differences (see 21 CFR 314.54(a) ["a 505(b)(2) application need contain only that information needed to support the modification(s) of the listed drug"]. Application of a 505(b)(2) paradigm removes the need to demonstrate bioequivalence per se and potentially reduces innovator intellectual property concerns that arise if a generic must "duplicate" the innovator. Guidance as to how similar a "biosimilar" needs to be exists in the form of current regulations related to orphan drugs and postapproval manufacturing changes.

Several recent drug approvals illustrate how this regulatory framework may be applied and are described in the sections to follow.

### 7.1 Case Study 1: Fortical [Calcitonin-Salmon (rDNA origin)]

On August 17, 1995, the FDA approved Novartis's NDA for Miacalcin (calcitonin-salmon) Nasal Spray (Miacalcin NS) for the treatment of postmenopausal osteoporosis in females greater than five years postmenopause with low bone mass relative to healthy premenopausal females. The active ingredient in Miacalcin NS is synthetic salmon calcitonin. On March 6, 2003, Unigene submitted a new drug application under Section 505(b)(2) for Fortical [calcitonin-salmon (rDNA origin)] Nasal Spray which relied in part on data submitted in the Miacalcin NS NDA.

***Comparability Program*** Fortical and Miacalcin NS differed in certain aspects, such as the use of recombinant versus synthetic salmon calcitonin and the use of different types and amounts of excipients. Given these differences, Unigene was required to submit data to establish that the findings of safety and efficacy for Miacalcin were relevant to Fortical (i.e., contain the same active ingredient and have comparable bioavailability) and that the formulation differences did not impact previous clinical profile [26, 27].

***Comparability Results***

*Physicochemical Analysis* Salmon calcitonin is a 32-amino-acid, nonglycosylated peptide hormone. It is structurally simple, possessing limited secondary structure and a single disulfide bond. The physicochemical characterization studies demonstrated that the primary and secondary structure of Fortical's recombinant salmon calcitonin (sc) was identical to that of Miacalcin's synthetic sc or naturally occurring sc. Further, the tertiary structures of the three were indistinguishable.

*Nonclinical* PK/$T_{ox}$ The pharmacokinetic profile of Fortical by different routes of administration was compared to Miacalcin, demonstrating similarity in PK profiles between the synthetic and recombinant peptides and toxicity results (28-day rat intranasal toxicity study) were acceptable, particularly in light of clinical safety data.

*Clinical* PK/PD Calcitonin has a well-established mechanism of action; published literature supports that salmon calcitonin, mediated through calcitonin receptors located on osteoclasts, inhibits bone resorption, thereby increasing bone mineral density. Since serum beta-CTx (C-telopeptides of type 1 collagen, corrected for creatinine) is a recognized marker of bone resorption, the effect of administered salmon calcitonin on serum beta-CTx is considered to be an adequate surrogate for pharmacodynamic comparisons.

Fortical's PD equivalence was shown in a double-blind, active-controlled, 24-week study in 134 postmenopausal women randomized to Fortical (200 IU per day) or Miacalcin (200 IU per day). The primary outcome measure was change in serum beta-CTx from baseline. The results fell within prespecified PD equivalence limits (–0.08 to 0.06 ng/mL; equivalence margin of ±0.2 ng.mL) and indicated Fortical was not inferior to Miacalcin.

Fortical's PK equivalence was assessed by comparing the relative bioavailability of Fortical to Miacalcin in a multidose, crossover study of 47 healthy female volunteers. Results indicated that Fortical was slightly more bioavailable than Miacalcin, but given the demonstration of similar PD activity, this difference were not considered to be clinically significant.

*Immunogenicity* Archived samples from the 24-week PD study were used to compare the immunogenicity potential of both products. The results indicated there was no difference in terms of total immune response and the response of neutralizing antibodies between the two drugs.

***Conclusion to Case Study 1*** On August 12, 2005, the FDA approved Unigene's 505(b)(2) application for Fortical for the same indication as Miacalcin NS [26, 27]. In the FDA's analysis no statistically and/or clinically significant differences were noted in any aspect of the comparability profile, including clinical performance, and Fortical was approved.

The basis of this comparison was strongly challenged in a citizen petition claiming that (1) recombinant salmon calcitonin is not the same as the synthetic version which could potentially cause differences in product efficacy, safety, or both and (2) only a long-term clinical study (actual bone fracture data) would provide adequate support of sameness [28]. The FDA responded to this citizen petition by asserting its decision that the comparability data presented above collectively constituted sufficient demonstration of sameness [27, 29].

### 7.2 Case Study 2: Omnitrope [Somatropin (rDNA origin)]

On August 24, 1995, the FDA approved NDA20-280 filed by the Pharmacia & Upjohn Company for Genotropin (somatropin) (rDNA origin) for injection. Since that time, Genotropin has been marketed as a safe and effective therapy for growth hormone deficiency (GHD) in children and adults.

On July 30, 2003, Sandoz submitted a 505(b)(2) application for the approval of its recombinant HGH product (recombinant somatropin) indicated for long-term treatment of pediatric patients who have growth failure due to an inadequate secretion of endogenous growth hormone and for long-term replacement therapy in adults with GHD of either childhood or adult onset. This application relied in part on data submitted in the Genotropin NDA.

***Comparability Program*** As with the Fortical case study, Omnitrope and Genotropin differed in certain aspects. As such, Sandoz was required to submit substantial data to establish that Omnitrope was sufficiently similar to Genotropin to warrant reliance on FDA's finding of safety and effectiveness for Genotropin to support the approval of Omnitrope [18].

#### *Comparability Results*

*Physicochemical Analysis* In terms of complexity, HGH is fairly simple and well-characterized. Human growth hormone is a single-chain, 191-amino-acid, nonglycosylated protein with two intramolecular disulfide bonds. Sandoz used a variety of physicochemical tests and analytical methods to confirm the primary, secondary, and

tertiary structures, molecular weight, and impurity profile. Characterization studies performed to verify somatropin as the active ingredient in Omnitrope included reverse-phase liquid chromatography/mass spetrometry (RP-HPLC/MS), DNA sequencing, N-terminal and C-terminal sequencing, peptide mapping, circular dichroism (CD) analysis, UV spectroscopy, one-dimensional nuclear magnetic resonance spectroscopy (1D NMR), two-dimensional (2D) NMR, size exclusion chromatography (SEC), isoelectric focusing (IEF), sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS–PAGE), and capillary zone electrophoresis.

*Nonclinical* PK/$T_{ox}$ Minimal toxicity data were needed on recombinant HGH (rHGH) itself, since the clinical effects of HGH excess are well established and understood and are extensively documented in published literature. Sandoz performed toxicity studies to appropriately qualify impurities specific to Omnitrope, that is, a subacute 14-day rat study and a local (skin) tolerance study in rabbits. Further, the bioactivity of Omnitrope was assessed using a validated weight gain bioassay using a hypophysectomized (growth-hormone-deficient) rats.

*Clinical PK/PD* HGH has a well-established mechanism of action. Omnitrope was demonstrated to be pharmacokinetically and pharmacodynamically "highly similar" to Genotropin. The dataset comprised a total of three PK/PD studies, including a double-blind, randomized, two-way crossover study comparing Omnitrope and Genotropin. Additionally, Sandoz conducted three sequential, multicenter phase 3 pivotal trials in 89 pediatric patients with GHD providing data in some patients for up to 30 months. A fourth phase 3 trial ($n$ = 51, 24 months) was submitted as part of its safety update. Collectively, these data in conjunction with the demonstrated comparability to the reference listed product provide substantial evidence of Omnitrope's safety and effectiveness.

*Immunogenicity* A significant number of patients who were administered an earlier version Omnitrope developed anti–growth hormone antibodies during the first and second phase 3 clinical trials. In response, Sandoz implemented changes to the drug product to address this immunogenicity and evaluated the impact of these changes clinically. Data from the 24-month clinical study demonstrated that Omnitrope has a low and acceptable level of immunogenicity (comparable to other rHGH products) as none of the patients developed anti–growth hormone antibodies during the duration of the study and only one patient developed anti–host cell protein antibodies, which were of no detectable clinical consequence.

***Conclusion to Case Study 2*** This case provoked significant challenges from interested parties voiced via several citizen petitions [18]. Furthermore, the FDA's delay in approval prompted Sandoz to file suit to compel the FDA to rule on its application. On April 10, 2006, the Washington, D.C., District Court ruled that the FDA must meet its statutory obligations and take action on Sandoz's outstanding NDA[6]. On May 30, 2006, the FDA approved Omnitrope [somatropin (rDNA origin)] as a "follow-on protein product" for use in the treatment of pediatric GHD. At the same time, the FDA responded to the related citizen petitions and defended its position that the data were adequate to demonstrate that Omnitorpe was sufficiently similar to Genotropin to enable reliance on the agency's previous findings of safety and

efficacy for Genotropin. These data, in conjunction with the independent evidence of safety and efficacy provided by Sandoz, supported Omnitrope's approval.

### 7.3 Case Study 3: Generic Salmon Calcitonin

On February 17, 2004, Nastech Pharmaceutical Company announced its filing of an ANDA for a salmon calcitonin nasal spray drug product for the treatment of post-menopausal osteoporosis. As with Fortical, Novartis's Miacalcin was cited as the reference listed drug; however, Nastech chose to submit an ANDA via the 505(j)(1) route, rather than a 505(b)(2) application.

The distinction between the two regulatory routes has significant implications for FOPs. Whereas 505(b)(2) allows products to be "sufficiently similar," an ANDA requires the applicant establish "sameness" of the active ingredients. The scope of data necessary to demonstrate that the actives are the same is unclear. Additionally, use of the ANDA route is appropriate for circumstances in which "clinical studies are not necessary to show safety and effectiveness." If clinical data are required as proof of sameness, as in the previous example where clinical data were used to demonstrate comparable immunogenicity, then the ANDA route may not represent a viable regulatory path.

On July 10, 2006, Nastech was notified by the FDA that its ANDA for intranasal calcitonin salmon was not approvable at present based on concerns relating to the potential for immunogenicity that might result from a possible interaction between calcitonin salmon and chlorobutanol, the preservative in the formulation.

Nastech has indicated it will continue to work with the agency to understand the data requirements and regulatory options, but the final resolution remains presently unknown. This case study highlights the fact that demonstration of *sameness* of therapeutic proteins is more complex than for other drugs and that true "biogenerics" may be hard to come by due to the complexity in establishing sameness versus similarity.

## 8 SUMMARY AND CONCLUSIONS

This chapter provides an overview of the complex scientific, legal, and policy issues facing the development of biogenerics today. Given the rising cost of health care and prescription medications in this country and the pivotal and expanding role of biologically derived products within the pharmaceutical landscape, these issues present a challenge to industry, regulators, and legislators alike. Substantial progress has already been made and the regulatory climate continues to evolve in response to advancing science and technology. Recent FDA approvals provide insight into the technical requirements for approval of well-characterized FOP products. They also demonstrate the appropriate use of an abbreviated approval pathway, that is, the 505(b)(2) pathway in place for drugs approved under the FDCA. Importantly, recent legislative proposals seek to amend the PHSA to eliminate the current legal barriers which prohibit abbreviated approval of protein biologics. This legislation reaffirms the need for the FDA to determine on a case-by-case basis the nature and extent of supporting data required for a given product.

## REFERENCES


1. Congressional Budget Office (1998), *How Increased Competition from Generic Drugs Has affected Prices and Returns in the Pharmaceutical Industry*, Congressional Budge Office, Washington, DC.
2. Crawford, L. M., Acting Commissioner of the Food and Drug Administration, in a Speech to the Generic Pharmaceutical Association on February 26, 2005, available: http://www.fda.gov/oc/speeches/2005/GPhA0301.html, accessed Apr. 23, 2005.
3. Comments of the Generic Pharmaceutical Association (GPhA) (Sept. 29, 2006), available: http://www.gphaonline.org/AM/Template.cfm?Section=Media&Template=/CM/HTMLDisplay.cfm&ContentID=2849, accessed Jan. 23, 2007.
4. Schellekens, H. (2005), Follow-on biologics: Challenges of the "next generation", *Nephrol. Dial. Transplant*. 20(Suppl. 4), iv31–iv36.
5. Congressional letter from Senators O. Hatch and H. Waxman to Andrew von Eschenbach, Acting Commissioner of the Food and Drug Administration (Feb. 10, 2006), available: http://www.henrywaxman.house.gov/news_letters_2006.htm, accessed Dec. 21, 2006.
6. Messplay, G. C., and Heisey, C. (2006), Follow-on biologics: The evolving regulatory landscape, *Bioexec Int.*, May, 42–45.
7. *Biotechnol. Law Rept.*, 2003, 22(5), 485–508.
8. Comments from R. Williams, U.S. Pharmacopoeia (USP), to FDA Docket No. 2004N-0355, Mar. 15, 2005.
9. Herrera, S. (2004), Biogenerics standoff, *Nat. Biotechnol.*, 22(11), 1343–1346.
10. Scott, S. R. (2004) What is a biologic?, Chapter 1 in Mathieu, M., Ed., *Biologics Development: A Regulatory Overview*, 3rd ed., Paraxel Intl., Waltham, MA, pp. 1–16.
11. Federal Food Drug and Cosmetic Act, available: http://www.fda.gov/opacom/laws/fdcact/fdctoc.htm, accessed Apr. 21, 2005.
12. Public Health Service Act, available: http://www.fda.gov/opacom/laws/phsvcact/phsvcact.htm, accessed Apr. 21, 2005.
13. U.S. Department of Health and Human Services, Food and Drug Administration Transfer of Therapeutic Products to the Center for Drug Evaluation and Research, available: http://www.fda.gov/cber/transfer/transfer.htm, accessed Apr. 23, 2005.
14. U.S. Department of Health and Human Services (DHHS) (1999, Oct.), Food and Drug Administration, Center for Drug Evaluation and Research, Guidance for Industry: Applications covered by Section 505(b)(2), DHHS, Washington, DC.
15. U.S. Department of Health and Human Services, Food and Drug Administration, Omnitrope (somatropin [rDNA origin]) questions and answers, available: http://www.fda.gov/cder/drug/infopage/somatropin/qa.htm, accessed Dec. 21, 2006.
16. Letter of J. Woodcock. (CDER, FDA) to Docket Nos. 2001P-0323/CP1, 2002P-0447/CP1, and 2003P-0408/CP1 (Oct. 14, 2003).
17. Glidden, S. (2001), The generic industry going biologic, *Biotechnol. Law Rept*, 20(2), 172–181.
18. Letter from S. Galson (CDER, FDA) in response to Docket Nos. 2004P-023 11CP1 and SUP 1,2003P-0 1 76lCP 1 and EMC 1, 2004P-0171lCP1 and 2004N-0355 (May 30, 2006).
19. U.S. Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research, Office of Pharmaceutical Science, Office of Generic Drugs, Electronic orange book: Approved drug products with therapeutic equivalence evaluations, available: http://www.fda.gov/cder/ob/default.htm, accessed Apr. 21, 2005.
20. Schellekens, H. (2004), How similar do "biosimilars" need to be? *Nat. Biotechnol*. 22(11), 1357–1359.

21. Webber, K. (2005), Relevant terminology. A presentation conducted at the Public Workshop on the Development of Follow-On Protein Products, Sept. 14, 2004, available: http://www.fda.gov/cder/meeting/followOn/followOnPresentations.htm, accessed Feb. 2, 2005.
22. Waxman, H., Schumer, C. E., and Clinton, H. R. (2006), Congress of the United States, H.R. 6257, Access to Life-Saving Medicine Act," available: http://www.waxman.house.gov/pdfs/bill_generic_biologics_9.29.06.pdf, accessed Sept. 29, 2006.
23. Mathieu, M., and Evans, A. G. (2005), The FDA's Orphan Drug Development Program, in Ed., *New Drug Development: A Regulatory Overview*, 7th ed., Paraxel Intl., Waltham, MA, pp. 307–317.
24. U.S. Department of Health and Human Services (DHHS) (2003, Sept.), Food and Drug Administration, Center for Drug Evaluation and Research, Guidance for industry: Comparability protocols—Protein drug products and biological products—Chemistry, manufacturing and controls information, DHHS, Washington, DC.
25. Beatrice, M. (2002), Regulatory considerations in the development of protein Pharmaceuticals, in Nail, S., and Akers, M., Eds., *Development and Manufacture of Protein Pharmaceuticals, Pharmaceutical Biotechnology*, Vol. 14, Kluwer Academic/Plenum, New York, pp. 405–457.
26. Letter from R. Levy (Unigene) to FDA Docket No. 2004P-0015 (Apr. 11, 2005).
27. Letter from S. Galson (CDER, FDA) in response to Docket No. 2004P-0015/CP1 (Aug. 12, 2005).
28. Letter from N. Buc to FDA Docket No. 2004P-0115/CP1 (Jan. 9, 2004).
29. *FDA Week*, 11(34), Aug. 26, 2005.

# RADIOPHARMACEUTICAL MANUFACTURING

Brit S. Farstad[1] and Iván Peñuelas[2]

[1]*Institute for Energy Technology, Isotope Laboratories, Kjeller, Norway*
[2]*University of Navarra, Pamplona, Spain*

## Contents

1 Introduction
- 1.1 Radiopharmacy
- 1.2 Characteristics of Radiopharmaceuticals
- 1.3 Ideal Characteristics of Radiopharmaceuticals
- 1.4 Radioactive Decay
- 1.5 Principles of Radiation Protection
- 1.6 Detection Devices for Clinical Nuclear Imaging

2 Product Development
- 2.1 Radionuclides
- 2.2 Carrier Molecules/Active Ingredients
- 2.3 Radiolabeling Techniques
- 2.4 Manufacturing Scale-Up
- 2.5 Automation

3 Manufacturing Aspects
- 3.1 Design of Manufacturing Sites
- 3.2 Design of Production Processes
- 3.3 Design of Production Equipment
- 3.4 Cleaning and Sanitation of Production Equipment
- 3.5 Environmental Control
- 3.6 Sterilization of Radiopharmaceuticals
- 3.7 Starting Materials
- 3.8 Labeling and Packaging

4 Product Manufacturing
- 4.1 Production of Radionuclides
- 4.2 Production of Radiopharmaceuticals

5 Quality Considerations

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

5.1 Documentation
5.2 Qualification of Personnel
5.3 Quality Control
5.4 Validation and Control of Equipment and Procedures
5.5 Stability Aspects of Radiopharmaceuticals
6 Extemporaneous Preparation of Radiopharmaceuticals
References
Further Readings

## 1 INTRODUCTION

### 1.1 Radiopharmacy

Radiopharmacy is a patient-oriented science that includes the scientific knowledge and professional judgment required to improve and promote health through assurance of the safe and efficacious use of radiopharmaceuticals. Radiopharmacy encompasses studies related to the pharmaceutical, chemical, physical, biochemical, and biological aspects of radiopharmaceuticals.

Radiopharmacy comprises a rational understanding of the design, preparation, and quality control of radiopharmaceuticals, the relationship between the physicochemical and biological properties of radiopharmaceuticals and their clinical applications, as well as radiopharmaceutical chemistry and issues related to the management, selection, storage, dispensing, and proper use of radiopharmaceuticals.

### 1.2 Characteristics of Radiopharmaceuticals

A radiopharmaceutical is any medicinal product which, when ready for use, contains one or more radionuclides (radioactive isotopes) included for a medicinal purpose. This generic definition of radiopharmaceutical thus includes both diagnostic and therapeutic radiopharmaceuticals.

A radiopharmaceutical can be as simple as a radioactive element such as $^{133}$Xe, a simple salt such as $^{131}$INa, a small labeled molecule such as L-(*S*-[$^{11}$C]methyl)methionine, or a protein labeled with a radionuclide such as $^{99m}$Tc-labeled albumin or $^{90}$Y-labeled monoclonal antibodies.

In clinical nuclear medicine, roughly 95% of radiopharmaceuticals are used with diagnostic purposes. Radiopharmaceuticals are administered to the patients only once, or a few times at most, in their lifetime. They contain minute amounts of active ingredients, with a radionuclide somehow linked to or being the active ingredient itself, with the main purpose of obtaining an image or a measure of their biodistribution. Radiopharmaceuticals do not usually show any measurable pharmacodynamic activity, as they are used in tracer quantities. Hence, there is no dose–response relationship in this case and thus differs significantly from conventional drugs.

Radiation is of course an inherent characteristic of all radiopharmaceuticals. Hence, patients always receive an unavoidable radiation dose. In the case of therapeutic radiopharmaceuticals, radiation is what produces the therapeutic effect.

The terms *tracer, radiotracer*, and *radiodiagnostic agent*, although long used as equivalent to radiopharmaceutical, should be avoided. The preferred and correct term is *radiopharmaceutical*, as the other names can be confusing or do not clearly show the nature of these compounds as pharmaceuticals.

The composition of radiopharmaceuticals is not constant as it varies with time as the radionuclide disintegrates. Very often, the half-life of the labeled molecule is so short that it must be readily prepared just before its administration to the patient. This implies in many cases the use of "semimanufactured", such as radionuclide generators, precursors, and cold kits that are also considered a medicinal product according to directive 2001/83/EC.

### 1.3 Ideal Characteristics of Radiopharmaceuticals

Radiopharmaceuticals should have several specific characteristics that are a combination of the properties of the radionuclide used as the label and of the final radiopharmaceutical molecule itself. The radiopharmaceutical should ideally be easily produced (both the radionuclide and the unlabeled molecule) and readily available. The half-life of the radionuclide should be adequate to the diagnostic or therapeutic purpose for which it is designed. It has to be considered that radiopharmaceuticals disappear from the organism by a combination of two different processes. The biological half-life (showing the disappearance of a radiopharmaceutical from the body due to biological processes such as metabolization, excretion, etc.) and the physical half-life (due to the radioactive decay of the radionuclide). The combination of both parameters gives the *effective half-life*:

$$T_e = \frac{T_p T_b}{T_p + T_b}$$

where $T_e$ is the effective half-life, $T_p$ the physical half-life, and $T_b$ the biological half-life. Radiopharmaceuticals should have an effective half-life adequate to the use for which they are intended. It should be short (hours) for diagnostic radiopharmaceuticals (not longer than the time necessary to complete the study in question) and longer for therapeutic radiopharmaceuticals (most often days) as the intended effect should have a sufficient duration.

The type of decay of the radiopharmaceutical should also be adequate for its intended use. Diagnostic radiopharmaceuticals should decay by $\gamma$ emission, electron capture, or positron emission, and never emit $\alpha$ or even $\beta$ particles. On the contrary, therapeutic radiopharmaceuticals should decay by $\alpha$ or $\beta$ emission because the intended effect is in fact radiation damage to specific cells.

Regarding the energy emission of diagnostic radiopharmaceuticals, the finally produced $\gamma$ rays should be powerful enough to be detected from outside of the body of the patient. The ideal energy for nuclear medicine equipment is around 150 keV. $\gamma$ rays should be monochromatic and photon abundance should be high to decrease the imaging time.

### 1.4 Radioactive Decay

Radionuclides are unstable nuclei that are stabilized upon radioactive decay. More than 2000 unstable nuclides have been described so far, most of them radioactive.

The stabilization process can proceed by several different processes, such as spontaneous fission, α-particle emission, β-particle emission, positron emission, γ-ray emission, or electron capture. In all decay processes the mass, energy, and charge of radionuclides must be conserved, and many nuclides can decay by a combination of any of the above-mentioned processes.

Fission is the process in which a nucleus breaks down into two fragments (thus leading to two different new nuclides) with an emission of two or three neutrons and a lot of energy. Spontaneous fission is a rare process that can only occur in heavy nuclei. Fission can also be produced by bombardment of certain nuclides with high-energy particles (such as neutrons) and is in fact the nuclear process used for the production of energy in nuclear energy plants by bombardment of highly enriched uranium with neutrons.

The usual decay process of heavy nuclei is α-particle emission. An α particle is a helium ion containing two protons and two neutrons. Alpha particles are heavy particles that have a very short range in matter due to their mass, and radiopharmaceuticals labeled with α emitters are used only with therapeutic purposes. Their clinical use is very limited, and they are mainly used for research purposes or in early phase clinical studies.

Radioactive nuclides that are neutron rich disintegrate by β decay. A $\beta^-$ particle is originated by the conversion of a neutron into a proton, along with the emission of an antineutrino to conserve energy in the decay process. Beta-emitting radionuclides are also used in radiopharmaceuticals for therapeutic purposes.

Positron decay occurs in proton-rich nuclei. In this case, the positron (or $\beta^+$ particle) is originated by conversion of a proton into a neutron, along with the emission of a neutrino to conserve the energy. Positrons are the antiparticle of electrons. In a very fast process ($10^{-12}$ s), emitted positrons collide with an electron of a nearby atom and both particles disappear in a process called annihilation. The necessary conservation of mass and energy accounts for the transformation of the mass of both particles into energy, which is characteristically emitted in the form of two 511-keV photons almost in opposite directions. Consequently, positron emitters are used to label radiopharmaceuticals produced with diagnostic purposes by imaging.

Proton-rich nuclei can also decay by electron capture. In this process, an electron from the innermost electron shell orbitals is captured into the nucleus and transforms a proton into a neutron (and a neutrino is emitted for conservation of energy). The vacancy created by the lost electron is filled by the transition of an electron from a higher level orbital, and the energy difference between the intervening orbitals is emitted as energy in the form of an X ray.

For any particular nucleus, several different energy states can be defined by quantum mechanics. All the excited energy states above the ground state are referred to as isomeric states and decay to the ground state by the so-called isomeric transition. In β, positron, or electron-capture decay processes, the parent nucleus may reach any of these isomeric states of the daughter nucleus. The energy difference between the nuclear energy states can be emitted as γ rays. A particular situation for isomeric transition is that in which the excited state is long lived and is then called the metastable state.

***Radioactive Decay Equations, Magnitudes, and Units*** Radioactive decay is a random process, being impossible to tell which particular atom from a group of atoms will decay at a specific moment. It is then only possible to talk about the average

number of atoms that disintegrate during a certain period of time, giving the disintegration rate ($-dN/dt$) of a particular radionuclide that is proportional to the total number of radioactive atoms present at that time. This magnitude is usually called the radioactivity (or mainly simply the activity) of a radionuclide and denoted by

$$A = -\frac{dN}{dt} = \lambda N$$

where $\lambda$ is the decay constant and $N$ the number of radioactive atoms. The previous differential equation is mathematically solved leading to the exponential equation

$$N_t = N_0 e^{-\lambda t}$$

where $N_t$ and $N$ are the number of radioactive atoms present at time $t = 0$ and $t = t$, respectively.

Radioactivity is expressed in becquerels (Bq), the Internationale System (SI) unit for the magnitude $A$. One Becquerel is defined as one disintegration per second (dps). Usual activities used in radiopharmacy are in the range of megabecquerels or gigabequerels. There is (as usual) a non-SI unit called the curie (Ci). It was initially defined in a trivial way as the disintegration rate of one gram of radium, which was considered to be $3.7 \times 10^{10}$ dps. Thus the equivalence between the becquerel and the curie is as follows:

$$1\,\text{Bq} = 2.7 \times 10^{-11}\,\text{Ci} \qquad 1\,\text{Ci} = 37\,\text{GBq}$$

The decay constant $\lambda$ is a specific characteristic of any single radionuclide, but being related to probability, it is difficult to understand its meaning. Thus, a new magnitude is defined: the half-life ($t_{1/2}$), which is the time required to reduce the initial activity of a radionuclide to one-half. In consequence, after one half-life the activity of a radionuclide would be $A/2$, after two half-lives $A/4$, after three half lives $A/8$, and so on.

The relationship between the decay constant $\lambda$ and the half-life $t_{1/2}$ can be derived from the general radioactive decay equation

$$t_{1/2} = \frac{\ln 2}{\lambda}$$

An additional (and commonly misunderstood concept) is the mean life $\tau$,which is the average life of a certain group of radioactive atoms that is mathematically also derived from the decay constant $\lambda$ as

$$\tau = \frac{1}{\lambda} = \frac{t_{1/2}}{\ln 2} = 1.44 t_{1/2}$$

## 1.5 Principles of Radiation Protection

Production, transportation, and use of radiopharmaceuticals, as radioactive products, is governed by regulatory agencies dealing with radiation protection and nuclear safety.

In any case, and albeit the different regulation in different countries, as a general principle only licensed personnel working in an authorized facility are authorized to handle and use radiopharmaceuticals. Facilities and procedures are subject to periodic inspection by official radiation safety officers that control production and handling of radioactive material, its transportation, proper use, as well as personnel dosimetry and radioactive waste disposal.

The general principles of radiation protection are very simple:

*Justification.* All procedures involving radioactive material must be justified.

*Optimization.* The radiation exposure to any individual should be as low as reasonable achievable. This principle is the widely known ALARA concept, an acronym derived from as low as reasonable achievable.

*Limitation.* The radiation dose received by the personnel handling radioactive material will never exceed the legally established dose limits. It has to be taken into account that such limitations do not apply to patients receiving radiopharmaceuticals as either diagnostic or therapeutic agents. But nuclear medicine physicians, nuclear physicists, and radiopharmacists must ensure that the amount of radiopharmaceutical administered to a patient is adapted to his or her disease and optimized to obtained the intended result.

***Operational Radiation Protection*** The fundamentals of operational radiation protection (i.e., how to proceed when working with radioactive products) are based on three factors: distance, time, and shielding. In any case, it is obvious that the radiation hazard is increased with the activity of the radiation source, as can be derived from the mathematical equation to calculate the exposure rate $X$ given by

$$X = \frac{A\Gamma}{d^2}$$

where $A$ is the activity of the radiation source, $\Gamma$ a constant that is characteristic of every radionuclide, and $d$ the distance to the source.

Distance should be increased as much as possible to decrease exposition and exposure time should be reduced to a minimum. Adequate shielding (depending on the radionuclide and its emission characteristics) should be used whenever possible and handling of high activities should only be carried out by either automated systems or proper manipulators.

## 1.6 Detection Devices for Clinical Nuclear Imaging

Diagnostic radiopharmaceuticals are mostly used for in vivo imaging of the biodistribution of the radiopharmaceutical. Depending on whether $\gamma$ or positron emitters are used, different devices are employed for clinical imaging. In any case, imaging devices are based on detection of the high-energy photons coming from the body of the patient upon administration and specific uptake of a radiopharmaceutical. Advances in nuclear medicine imaging devices now permit in vivo noninvasive imaging of such biodistribution and to obtain tomographic (i.e., three-dimensional)

images that can also give quantitative or semiquantitative information about the amount of radiopharmaceutical and even its kinetics.

## 2 PRODUCT DEVELOPMENT

### 2.1 Radionuclides

When designing a radiopharmaceutical one should have in mind the potential hazard the product may have to the patient. The goal must be to have maximum amounts of photons with a minimum radiation exposure of the patient. For use in therapy, β emitters and α emitters are particularly useful. For diagnostic purposes, γ emitters are most widely used. In general, those γ emitters with a short physical half-life and with a γ energy between 100 and 300 keV are most widely used in medical application, since these can easily be detected by standard γ cameras.

However, positron emission tomography (PET) radiopharmaceuticals involve short-lived radionuclides (positron emitters) giving a double set of photons at 511 keV each.

### 2.2 Carrier Molecules/Active Ingredients

The function of the carrier molecule is to carry the radioactivity to the target organ and to make sure the radioactivity stays there. The uptake of radioactivity should be as specific as possible in order to minimize irradiation of other organs and parts of the body. This is particularly important when using radiopharmaceuticals for therapy. But, also, for use in diagnostics, it is desirable that the radiopharmaceutical is localized preferentially in the organ under study since the activity from nontarget areas can obscure the structural details of the pictures of the target organ. It is therefore important to know the specific uptake in an organ for a potential chemical carrier and also the rate of leaking out of the organ/organ system. Thus, the target-to-background activity ratio should be large. There are several approaches to develop targeting radiopharmaceuticals. Radioimmunotargeting is one approach frequently used for radiopharmaceuticals, where monoclonal antibodies (MAbs) or fractions of MAbs are the carrier molecules for the radioactivity. These are binding specifically to receptors on cell surfaces in the target organs.

The target-binding surface of the cell has been well explored with a range of tumor-associated and other antigens, identified, and used for pathological tissue characterizations.

The active analog approach in general, whereby a set of compounds is synthesized so as to mimic features of a chosen natural compound, has been successful [1]. The active analog approach includes the pharmacophore. The concept of the pharmacophore is to look at features common to a set of drugs or compounds binding to and acting on the same receptors.

### 2.3 Radiolabeling Techniques

When a labeled compound is to be prepared, the first criterion to consider is whether the label can be incorporated into the molecule to be labeled [2]. This may be

assessed from knowledge of the chemical properties of the two partners. Furthermore, one needs to know the amount of each component to be added. This is particularly important in tracer level chemistry and in $^{99m}$Tc chemistry.

In a radiolabeled compound, atoms, or groups of atoms of a molecule, are substituted by similar or different radioactive atoms or groups of atoms. Saha [2] lists six major methods employed in the preparation of labeled compounds for clinical use: isotope exchange reactions, introduction of a foreign label, labeling with bifunctional chelating agents, biosynthesis, recoil labeling, and excitation labeling. Among these, three frequently used methods in radiopharmaceutical synthesis are briefly described below.

***Isotope Exchange Reactions*** In isotope exchange reactions, isotopes of the same elements having different mass numbers replace one or more atoms in a molecule. Examples are labelling of iodide-containing material with iodine radioisotopes. Since the radiolabeled and parent molecules are identical except for the isotope effect, they are expected to have the same biological and chemical properties.

***Introduction of a Foreign Label*** In this type of labelling, a radionuclide is incorporated into a molecule primary by the formation of covalent or coordinated covalent bonds. The tagging radionuclide is foreign to the molecule and does not label it by exchange of one of its isotopes. Examples are $^{99m}$Tc–DTPA (Diethylenetriamine-pentacetic acid), $^{51}$Cr-labeled red blood cells, and many iodinated proteins and enzymes. In many compounds of this category, the chemical bond is formed by chelation. In chelation, one atom donates a pair of electrons to the foreign acceptor atom, which is usually a transition metal. Most of the $^{99m}$Tc-labeled compounds used in nuclear medicine are formed by chelation.

***Labeling by Bifunctional Chelating Agents*** In this approach, a bifunctional chelating agent is conjugated to a macromolecule (e.g., protein) on one side and to a metal ion by chelation on the other side. Examples of bifunctional chelating agents are DTPA, metallothionein, diamide dimercaptide ($N_2S_2$), and dithiosemicarbazone [2].

There are two methods: the preformed radiometal–chelate method and the indirect chelator—antibody method. Various antibodies are labelled by the latter, where the bifunctional chelating agent is initially conjugated to a macromolecule, which is then allowed to react with a metal ion, to form a metal–chelate–macromolecule complex. Due to the presence of the chelating agent, the biological properties of the labeled protein may be altered and must be assessed before clinical use.

## 2.4 Manufacturing Scale-Up

As the radiolabeled substances emerge from the laboratory to the clinics, there will be a need for scaling up the batch size of the product. This can be done by increasing either the total volume of the produced batches or the specific activity of the product or both. When doing this, the following aspects should be considered:

The influence on the stability of the product itself due to possible radiolysis

The need for additional operator protection due to handling of increased amounts of radioactivity

***Product Stability*** The stability of a labeled compound is one of the major problems in labeling chemistry. It must be stable both in vitro and in vivo. Many labeled compounds are decomposed by radiation emitted by the radionuclides in them. This kind of decomposition is called radiolysis. Radiation may also decompose the solvent, producing free radicals that can break down the chemical bonds of the labeled compounds (indirect radiolysis). In general, the risk of radiolysis increases with higher specific activity of the product. In addition, the more energetic the radiation, the greater is the radiolysis. Alpha emitters, leaving most of its energy close by the molecules, and thus a high potential risk of radiolysis, give rise to major challenges when scaling up is necessary.

***Operator Radiation Protection*** Even for the largest commercial manufacturer of radiopharmaceuticals, the batch volumes are small compared to nonradioactive pharmaceuticals. So even a scaled-up production batch can be contained within a limited space. When scaling up a radiopharmaceutical production, one always has to assure that the radiation outside the contained work unit is acceptable for the operator.

The production of a radiopharmaceutical will normally take place within a contained box unit. Depending on the kind of radionuclides used and the amount of radioactivity handled in the production process, the box units are shielded by lead walls, typically 5–15 cm in thickness. When the box is used for production of radiopharmaceuticals incorporating α- or β-emitting radionuclides, closed box units without any lead coating are sufficient. Working with these types of radionuclides or with smaller amounts of γ-emitting radioactivity, as in research scale, suitable glove boxes can be used. When working with larger quantities of γ-emitting radionuclides, the material must be handled by either remote control equipment or manipulator tongs incorporated in the wall.

## 2.5 Automation

Because of the unique operational and safety requirements of radiopharmaceutical synthesis, the motivation for the development of automated systems is clear. These unique constraints include short synthesis times and control from behind bulky shielding structures that make both access to and visibility of radiochemical processes and equipment difficult. The need for automated systems is particularly expressed for PET radiopharmaceutical synthesis, with the short-lived radionuclides emitting high-energy γ photons at 511 keV. Automated synthesis systems require no direct human participation. The short half-lives of the PET radionuclides may require repeated synthesis during the day, thus being a potential radiation burden for the operator when not using automated systems.

Furthermore, radiopharmaceutical synthesis must be reliable and efficient and result in pharmaceutical-quality products. In addition, the processes must be well documented and controlled. Automated systems may support all these challenges and requirements.

One must keep in mind, though, that success in synthesis automation requires first and foremost innovative chemistry. PET radiosynthesis draws from a broad chemistry knowledgebase rooted in synthetic organic chemistry [3].

## 3 MANUFACTURING ASPECTS

### 3.1 Design of Manufacturing Sites

The manufacturing of radiopharmaceuticals is potentially hazardous. Both small- and large-scale production must take place on premises designed, constructed, and maintained to suit the operations to be carried out. Radiation protection regulations stipulate that radionuclides must only be used in specially designed and approved "radioisotope laboratories." National regulations with regard to the design and classification of radioisotope laboratories must be fulfilled. Such laboratories are normally classified according to the amount of the various radionuclides to be handled at any time and the radiotoxicity grading given to each radionuclide. When planning the layout of the laboratory, it is recommended to allocate separate working areas or contained units for the various procedures to avoid possible cross-contamination of radionuclides [4, 5]. Premises must be designed with two important aspects in mind:

The product should not be contaminated by the operator.

The operator and the environment should be protected from contamination by the radioactive product.

This is the basic principle of good radiopharmaceutical practice (GRPP).

One of the most important factors in planning a radioisotope laboratory is the design of the ventilation system. Laboratories with medium and high grading must be designed with the purpose of protecting the personnel from inhaling radioactive gases or particles. The system should be designed to provide lower pressure at the actual working area compared to the surrounding environment. Furthermore, the system should have an appropriate number of air changes per hour and the replacement air should be filtered. Air extracted from the area where radioactive products are handled, though, should not be recirculated. Exhaust air to the environment should be monitored for radioactivity, and it may be necessary to install active charcoal filters to absorb radioactive gases and small particles [4].

Aseptic production of radiopharmaceuticals, that is, when the products cannot be terminally sterilized, will increase the requirements for the design and construction of the premises. Contained workstations and clean-room technology will be applied to a much higher degree. The general requirements for the design of such premises are the same as for nonradioactive pharmaceuticals, including entry of staff and the introduction of materials through air locks. The main difference is found in the planning and design of the ventilation system. Laboratories for aseptic work normally have a positive pressure relative to the surrounding areas. On the other hand, in laboratories for work with radioactivity, it is good practice to have a negative pressure to avoid the spread of radioactive material. In order to meet both pharmaceutical and radiation protection requirements, it is necessary to balance carefully the air pressures in the clean rooms, the air locks, and the surrounding areas. From a pharmaceutical point of view a negative pressure in the area designated for aseptic work can only be accepted in special cases. There are various ways to meet the required balance between these apparently contradictory principles. A frequently chosen solution is to use sealed production units or contained work-

stations supplied with unidirectional airflow (UDAF) and with a lower pressure compared to the aseptic laboratory. The laboratory itself may then have positive pressure in relation to the surrounding premises.

Waste management is an important aspect when planning a radiopharmaceutical manufacturing site. The key factor is to reduce the amount of radioactive waste to a minimum. There should be a system for dividing the waste according to physical half-life and radiotoxicity, both for solid and liquid waste. As an example, waste containing α emitters is normally kept separately, when possible. National legislation will vary considerably and influence the requirement that must be set for handling of radioactive waste material.

### 3.2 Design of Production Processes

The design of a radiopharmaceutical production process depends very much upon the kind of radiopharmaceutical to be made. Although most radiopharmaceuticals are intended for parenteral use, also oral radiopharmaceuticals in different forms are widely used. One must emphasize different factors when planning for production of parenteral radiopharmaceuticals compared to oral radiopharmaceuticals. Still, a common factor is the involvement of radioactive materials, and the radiation protection of the personnel must always be an integral part of the design.

The production of a radiopharmaceutical will normally take place within a contained box unit, consisting of either plastic walls or a combination of plastic and stainless steel. The latter is more optimal for clean-room work. The box units may be shielded by lead, either as large lead panels or as lead brick walls (see Figure 1). Depending on the kind of radionuclides used and the amount of radioactivity handled in the box, the walls are typically 5–15 cm in thickness. Shielded production units like these are often called "hotcells."

When the box is used for productions of radiopharmaceuticals incorporating α- or β-emitting radionuclides, closed box units without any lead coating may be sufficient. When handling radionuclides with mixed emitting properties, a possibility is to concentrate the shielding to critical parts of the process. This can be done by use of local shielding inside the production unit. However, for aseptic production, one must keep in mind a potential disturbance of the airflow inside the box.

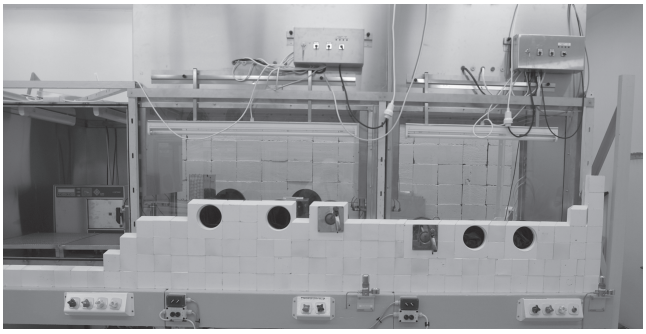

**FIGURE 1** Shielding of box units (hot cells) with lead bricks. (Photo courtesy of Institute for Energy Technology.)

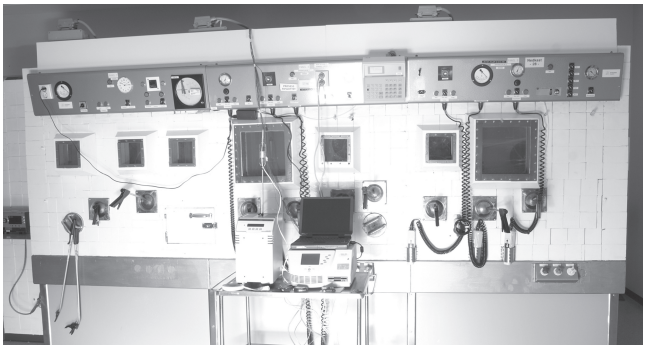

**FIGURE 2** Hot cells for manufacturing of larger quantities of γ-emitting radionuclides. (Photo courtesy of Institute for Energy Technology.)

**FIGURE 3** Lead bricks are interlocked when they are stacked on top of each other. (Photo courtesy of Institute for Energy Technology.)

Working with α and β radionuclides and also limited amounts of γ-emitting radioactivity, the boxes may be mounted with special protection gloves. When working with larger quantities of γ-emitting radionuclides, the material must be handled by either remote control equipment or manipulator tongs incorporated in the wall (see Figure 2).

The design of the elements and their assembly on the production unit should be such that there are no radiation leaks at the interface.

When using lead bricks to construct the wall of the production unit, they should have a special design. When they are stacked on top of each other, they should interlock (see Figure 3). This is important to avoid cracks in the wall through which radiation can escape.

Manipulator tongs are fitted into the wall as part of a large tungsten sphere which acts as a ball bearing and thereby allows more flexibility for the movement of the tong inside the box (see Figure 2). Lead glass windows, with good shielding properties, are fitted in the lead bricks to allow the operator to overlook the process.

When large lead panels are used, they should be reinforced with suitable steel structures [International Organization for Standardization (ISO) 10648-1: 1997].

The surfaces of the lead shielding must be smooth and easy to clean. This can be achieved by painting the surface of the wall.

In general, the manufacturing of most radiopharmaceuticals consists of the following:

Nuclear synthesis, synthesis of the radionuclide
Synthesis of the radiolabeled compound
Pharmaceutical formulation of the radiopharmaceutical

***Nuclear Synthesis*** Except for radionuclides with ultrashort half-lives, like most PET radionuclides, the production of these is normally performed well in advance (see Section 4.1). Thus, the radionuclide is considered as a starting material and must undergo controls as a starting material.

***Synthesis of Radiolabeled Compound*** The complexity of a radiopharmaceutical may differ greatly, with the radioactive element itself or simple salts as the less complex. Very often, though, the radiolabel is part of a larger molecule, and thus a radiolabeling procedure is required. This is part of the synthesis of the radiopharmaceuticals, which also may involve chemical alteration of a precursor of the active ingredient. Both labeling methods and synthesis may involve steps at elevated temperatures or even cooling steps. Thus, equipment for heating or cooling must be part of the production line. Furthermore, an important part of a synthesis is often the purification step, and equipment for this must be available. Typically, this is simple chromatographic or ion exchange columns.

Planning of the process very much depends on the complexity of the process. In general, keeping in mind the limited possibilities of direct handling of the materials, it is important to keep the processes as simple as possible. For more complex processes, automation may be the best solution, if available.

***Pharmaceutical Formulation*** Even when the radiochemical part of a product is simple, the radiopharmaceutical may be a complex solution. A pharmaceutical formulation often contains additives in the form of buffers and preservatives: buffers to keep the solution at a pH suitable for injection and preservatives to preserve the integrity and efficiency of the radiopharmaceutical.

Ideally, a solution for injection should be an isotonic solution with a neutral (physiological) pH. However, the pH of a radiopharmaceutical is very important for its stability, and for labelled compounds, the pH for optimal stability is not always equivalent to physiological pH. For iodide solutions, the pH should be alkaline to prevent loss of radioiodine. Reducing agents, such as thiosulfate, are often added to radioiodide solutions to help this situation. A preservative can act as a stabilizer, an antioxidant, or a bactericidal agent.

Some additives, like benzyl alcohol, are added for a double action. Benzyl alcohol 0.9% is widely used as a bactericide. In addition, benzyl alcohol reduces radiolysis in radiopharmaceuticals and thus acts as a stabilizer.

### 3.3 Design of Production Equipment

The equipment used for manufacturing operation should be reserved exclusively for radiopharmaceuticals [6]. Furthermore, two principles are of utmost importance in the design of production equipment [4]:

The equipment must be easy to repair after it has been installed in the production unit.

The equipment must have a simple construction and be easy to assemble, so a substitution can be done quickly when total renovation of the equipment is necessary.

Glass is an important material in the construction of production equipment for radiopharmaceuticals. This material will become discolored and brittle when affected by radiation, and thus repair and/or change of parts of the equipment may be necessary. Due to radioactive contamination of the equipment, repair and maintenance can often be complicated, and time for decay must be included in the maintenance period. To secure the continuous supply of products, it may be necessary to construct two production lines in separate production units, where one is kept as a backup facility.

Sometimes it will be necessary to substitute not only parts of a production line but also the assembly of equipment as a whole. To facilitate this operation and thereby reduce time and radiation exposure, it can be advantageous to build the whole production line on a stainless steel support frame fitted with simple connections to electricity, water, and air supplies [4]. The complete withdrawal of a production line from a box and the introduction of a new one can then be performed in a very short time.

It is also important to keep in mind, when designing production equipment, that all sense of touch is lost when fingers are replaced by remote handling tongs.

The design of the equipment must therefore be as simple as possible. On the other hand, when using hot-cell units mounted with handling tongs, it may be favorable to use more automated systems in the production line. Systems like these can be run and controlled from steering panels outside the box unit.

Finally, equipment should be constructed so that surfaces that come in contact with the product are not reactive, additive, or absorptive so as to alter the quality of the radiopharmaceutical.

### 3.4 Cleaning and Sanitation of Production Equipment

Preparation equipment should be designed so it can be easily and thoroughly cleaned. Procedures for cleaning, sanitation, and storage of production equipment used in radiopharmaceutical production must be established. Special training is necessary for personnel involved in this kind of work with regard to both clean-room aspects and radiation protection aspects.

Before any equipment or materials used during production are removed from the production unit, a check for radioactive contamination must be performed. After removal, the equipment should be allowed to decay further in a special storage area before it is cleaned and made ready for assembly again.

Glass equipment will normally be sterilized by dry-heat sterilization. Smaller equipment, like plastic tubes and rubber stoppers, can be sterilized by autoclaving.

If available locally, also $\gamma$ irradiation may be a suitable method for sterilization of equipment. One must keep in mind, though, that sterilization by irradiation may change the composition of plastic and rubber materials. In addition, glass materials may be discolored by $\gamma$ irradiation.

Production equipment that cannot be sterilized must be sanitized and disinfected by an appropriate method. This can be done by use of biocides like alcohols (70%), hydrogen peroxide, or formaldehyde-based chemicals or a combination of these. These can either be used for surface disinfections by wiping or spraying or even better by use of gas or dry fog systems for application of the disinfectants. The effect of cleaning and sanitation should be monitored. Microbiological media contact plates can be used to test critical surfaces, as inside the hot cells or glove boxes. The test samples must then be handled and monitored as radioactive contaminated units.

A system must be established for sanitation of all equipment before these are transferred into clean areas.

### 3.5 Environmental Control

Workstations and their environment should be monitored with respect to radioactivity, particulate, and microbiological quality. Active air sampling from production units for radioactive products (hot cells or glove boxes) is subject to a safety consideration. There is always a risk of bringing radioactive contaminated air outside the workstation. To avoid the spread of radioactivity during the test, all possible exhaust from the test equipment must be sampled and/or controlled.

A possible approach for testing of particulate and microbiological quality of air inside the hot cells or glove boxes is to gain information about airborne particles during simulated operations (without radioactivity).

The use of settle plates is common practice for monitoring of the microbiological quality of air inside production units. These must then be placed as close a possible to critical parts of the production process in order to show the real microbiological burden to the product.

Warning systems must be installed to indicate failure in the filtered air supply to the laboratory. Recording instruments should monitor the pressure difference between areas where this difference is of importance.

### 3.6 Sterilization of Radiopharmaceuticals

Sterile radiopharmaceuticals may be divided into those which are manufactured aseptically and those which are terminally sterilized. In general, it is advisable to use a terminal sterilization whenever this is possible. Terminal sterilization is defined as a process that subjects the combined product/container/closure system to a sterilization process that results in a specified assurance of sterility [7]. Since sterilization of solutions normally means autoclaving (steam sterilization), one must assure that the radiopharmaceutical product does not decompose when it is heated to temperatures above 120°C. Many radiolabeled compounds are susceptible to decomposition at higher temperatures. Proteins, such as albumin, are good examples of this. Others, such as $^{18}$F-fluodeoxyglucose (FDG), can be autoclaved in some formulation but not in others.

Furthermore, these processes take time, typically 20–30 min in total when heating up to 121°C. For very short-lived product with a half-life of only a few minutes, this is not an adequate method. On the other hand, these short-lived products are not subject to any storage, and thus the risk of microbiological growth is more limited.

Alternatively, a shorter cycle at a higher temperature might be used, assuming that the temperature does not decompose the radiopharmaceutical.

If terminal sterilization is not possible, aseptic processing must be performed. Aseptic processing is a process that combines presterilized materials and presterilized equipments in a clean area.

Heating of radioactive solutions, particularly under elevated pressure (e.g., steam sterilization), is also a matter of safety. In order to avoid any contaminated air to escape if a container or a seal is broken, autoclaves used for radioactive solutions should be placed inside negative-pressure sealed units. Autoclaves used for sterilizing high-energy γ-emitting radiopharmaceuticals should in addition be supplied with proper lead shielding.

### 3.7 Starting Materials

As for manufacturing of other pharmaceuticals, a system should be established to verify the quality of the starting materials used in manufacturing radiopharmaceuticals. This system must assure that no material is used for production until it has been released by a competent person [qualified person (QP) or others given this responsibility].

The starting materials as well as the packaging materials should be purchased from qualified vendors. It is recommended to use materials described in a pharmacopoeia, whenever this is available. Supplier approval should include an evaluation that provides adequate assurance that the material consistently meets specifications.

Radionuclides involved in manufacturing radiopharmaceuticals must be considered as starting materials. For very short-lived radionuclides, where batch analysis is not possible, the validation of the production process of the radionuclide is of utmost importance.

### 3.8 Labeling and Packaging

Packaging material should be purchased from qualified vendors. Primary containers and closures must be tested to verify that there are no interactions between the radiopharmaceutical and packaging material during storage of the product.

Due to the risk of radiation exposure, it is accepted that most of the labeling of the primary (direct) container is done prior to manufacturing. The empty vial can be prelabeled with partial information prior to filtration and filling [6]. This procedure should be designed so as to not compromise sterility or prevent visual inspection of the filled vial. After filling of radioactive products, the primary containers (vials) must be placed within a shielded container. These containers, which can be made of lead or tungsten, vary in size and thickness depending on the amount of radioactivity in the vial as well as the radiation properties of the radionuclide. Radiopharmaceuticals containing α or β emitters may be placed in thin lead pots, typically 2–4 mm in wall thickness. On the other hand, for vials containing regular doses of high-energy γ emitters, such as PET radionuclides, shielding with 3–5 cm lead/tungsten may be needed.

Necessary information about the product must be given on the label of the lead or tungsten container. Hence, there is no need to study the label on the direct con-

tainer. The name of the radiopharmaceutical, including the radionuclide, together with the amount of radioactivity in the vial at a stated calibration time is part of the necessary information. So is the expiry date of the product. Furthermore, the symbol for radioactivity, designed as a black propeller, is obligatory on labels for radioactive solutions.

When the products are intended for distribution and transport, the packaging and labeling of the outer packages must be done according to the national regulation of the country from which the shipments will depart, transfer, and arrive. The outer packaging material must be properly tested in accordance with the type of shipment, most frequently type A packages for radiopharmaceuticals. Furthermore, the packages must be labeled with radionuclide data, such as type and amount of radioactivity, along with the transport index (TI), which indicates the radiation from the package at 1 m distance. While the information on the product itself (outside the lead pot) is intended for the physicians, the information outside the package is intended for the transport personnel.

## 4 PRODUCT MANUFACTURING

### 4.1 Production of Radionuclides

Radiopharmaceuticals are labeled with artificial radionuclides that are obtained by bombardment of stable nuclei with subatomic particles or photons. Nuclear reactions produced in such a way convert stable in unstable (radioactive nuclei). Several kind of devices are used for such purposes, including nuclear reactors, particle accelerators, and generators.

Various types of targets have been designed and used for both reactor and cyclotron irradiation. In the design of targets, primary consideration is given to heat deposition in the target by irradiation with neutrons in the reactors or charged particles in the cyclotrons [2]. As the temperature can rise to 1000°C during irradiation in both reactors and cyclotrons, the target needs proper cooling to avoid burning. Most often, the targets are designed in the form of a foil to maximize the heat dissipation. The target element should ideally be monoisotopic or an enriched isotope to avoid extraneous nuclear reactions.

***Nuclear Reactors*** Nuclear reactors are highly complex systems in which two kinds of nuclear reactions are useful for the production of clinically useful radionuclides: Neutrons produced by the fission of heavy nuclides (such as $^{235}U$ or $^{239}Pu$) are used in a neutron capture (n, γ) reaction to produce an isotope of the same element that is bombarded by the neutrons. Such reactions can be produced almost in all elements with different probability. Examples of useful nuclear reactions are $^{130}Te(n, \gamma)^{131}Te$ (which produces $^{131}I$ after emission of β particles with a half-life of 25 min), $^{50}Cr(n, \gamma)^{51}Cr$, $^{58}Fe(n, \gamma)^{59}Fe$, and $^{98}Mo(n, \gamma)\ ^{99}Mo$. The second possibility for the use of nuclear reactors is to use fission reactions (n,f) in which a heavy nuclide is broken down into two fragments. Many clinically relevant radionuclides can be produced from thermal fission of $^{235}U$, such as $^{131}I$, $^{117}Pd$, $^{133}Xe$, and $^{137}Cs$. The isotopes produced by this kind of fission reaction must be separated and purified by appropriate chemical procedures, but since the chemical behavior of many different heavy

elements is similar, contamination can often become a problem in the isolation of the radionuclide of interest.

As an example, and due to the particular interest of $^{99}$Mo in radiopharmacy (as it is the parent nuclide of $^{99m}$Tc in the $^{99}$Mo–$^{99}$mTc generator), the complex process used to produce and purify $^{99}$Mo is described below.

Molybdenum-99 is produced by fission of $^{236}$U as follows:

$$^{235}\mathrm{U} + {}^{1}\mathrm{n} \rightarrow {}^{236}\mathrm{U} \rightarrow {}^{99}\mathrm{Mo} + {}^{135}\mathrm{Sn} + 2{}^{1}\mathrm{n}$$

After irradiation of the uranium target, it is dissolved in nitric acid and the final solution adsorbed on an alumina column that is washed with nitric acid to remove uranium (and other fission products). Molybdenum is finally eluted with ammonium hydroxide and further purified by absorption on an anion exchange column from which ammonium molibdate is eluted with dilute hydrochloric acid after washing the resin with concentrated HCl. The $^{99}$Mo is obtained in no-carrier-added conditions, and the most common contaminants can be $^{131}$I and $^{103}$Ru.

***Particle Accelerators: Cyclotrons*** Both linear and circular particle accelerators (cyclotrons) can be used, but the latter have many advantages and are mainly used for the production of clinically relevant radionuclides.

A cyclotron is basically a cylinder-shaped high-vacuum chamber in which by means of a magnetic field and a radio-frequency system used to generate an alternating electric field, elemental particles can be accelerated to very high energies and used as projectiles. The bombardment of stable elements loaded in a properly designed target (either solid or filled with a liquid or a gas) induces different types of nuclear reactions that finally lead to the production of radioactive elements.

Most cyclotrons accelerate negative particles (such as $^{2}$H, $^{1}$H, or even heavier particles such as helium cations) that are stripped off the electrons in the stripping foils that are used also to focus the beam on the target. As the energy of the incident particle is increased, a much greater variety of nuclides can be produced.

When the nuclides produced have atomic numbers different from those of the target elements, such preparations have no stable isotope of the intended element and can be considered to be produced in no-carrier-added conditions.

The target material should ideally be monoisotopic to avoid the production of extraneous radionuclides. However, in many cases this is not possible and only isotopically enriched targets can be used, thus leading to the production of different radionuclides. In this case appropriate methods must be used to separate the different elements produced in the target.

An interesting concept that must always be taken into account in cyclotron-produced radionuclides is the saturation activity characteristic of each target and each nuclear reaction. The saturation activity is the activity of the radionuclide in which the secular equilibrium is obtained between the activity produced in the target and the disintegration of the radioisotope. The activity produced at a target can be calculated by the equation

$$A = A_S(1 - e^{-(\ln 2)t/T})(\mu\mathrm{A})$$

where $A$ is the activity obtained for a radionuclide with a half-life of $T$ after irradiation of the target during a time of $t$ at a current of μA microamperes. From the

practical point of view, almost 97% of the saturation activity value is reached after irradiation of the target for five half-lives of the radionuclide. Longer irradiation times do not produce significant increases in the activity obtained. Methods to obtain several cyclotron-produced radionuclides are described below.

Iodine-123 can be produced either directly or indirectly in a cyclotron. Direct reactions usually lead to $^{123}$I contaminated with other iodine radioisotopes, such as $^{124}$I or $^{125}$I, due to side nuclear reactions. Using nuclear reactions such as $^{123}$Te(p,n)$^{123}$I, $^{122}$Te(d,n)$^{123}$I, or $^{124}$(p,2n)$^{123}$I produces $^{123}$I that is obtained after dissolving the target in hydrochloric acid by distillation into dilute NaOH.

In the indirect methods the radionuclide produced after bombardment of the target is not $^{123}$I, but a radionuclide that decays to $^{123}$I with a short half-life. The most widely used nuclear reactions produce $^{133}$Xe (which decays to $^{123}$I with a half-life of 2.1 h) by bombardment with high-energy $^{3}$He or $^{4}$He particles or $^{123}$Cs (which decays to $^{123}$Xe with a half-life of 5.9 min, and then $^{123}$Xe decays to $^{123}$I) after irradiation of $^{124}$Xe with high-energy protons. Complex processing and purification processes must be used to obtain $^{123}$I in any of these cases, and adequate design and composition of the target are critical to facilitate the process.

Thallium-201 is obtained using an indirect reaction such as $^{203}$Tl(p,3n)$^{201}$Pb in which $^{201}$Pb decays to $^{201}$Tl with a half-life of 9.4 h. Thallium-201 can in this way be obtained pure and free from other contaminants after several purification steps and letting the target product decay for 35 h.

Indium-111 is produced by a direct nuclear reaction by irradiation of an $^{111}$Cd target with 15-MeV protons. After irradiation the target is dissolved in HCl and purified in an anion exchange column.

Positron emission tomography has become a widely used diagnostic technique in nuclear medicine. Ultrashort half-live radionuclides are used in these cases, and such radionuclides are mostly obtained in small cyclotrons with high yields and short irradiation times. The overall process will be described further in this chapter when PET radiopharmaceuticals are described.

***Generators*** A generator is constructed on the principle of the decay–growth relationship between a parent radionuclide with longer half-life that produces by disintegration a daughter radionuclide with shorter half-life. The parent and the daughter radionuclide must have sufficiently different chemical properties in order to be separated. The daughter radionuclide is then used either directly or to label different molecules to produce radiopharmaceutical molecules.

A typical radionuclide generator consists of a column filled with adsorbent material in which the parent radionuclide is fixed. The daughter radionuclide is eluted from the column once it has grown as a result of the decay of the parent radionuclide. The elution process consists of passing through the column a solvent that specifically dissolves the daughter radionuclide leaving the parent radionuclide adsorbed to the column matrix.

The main advantage of the generators is that they can serve as top-of-the-bench sources of short-lived radionuclides in places located far from the site of a cyclotron or nuclear reactor facilities.

A generator should ideally be simple to build, the parent radionuclide should have a relatively long half-life, and the daughter radionuclide should be obtained by a simple elution process with high yield and chemical and radiochemical purity. The generator must be properly shielded to allow its transport and manipulation.

Several different generators are used in radiopharmaceutical procedures, but the $^{99}$Mo/$^{99m}$Tc is with great difference the most important generator of all of them and will be described in detail later on in this chapter.

## 4.2 Production of Radiopharmaceuticals

More than 90% of the radiopharmaceuticals used in nuclear medicine are for diagnostic use. PET radiopharmaceuticals, with their ultrashort half-lives, have become a significant part of this group of products. Hence PET investigation has been the fastest growing imaging modality worldwide the last few years [8].

Also for conventional radiopharmaceuticals used in diagnostic, it is favorable to use products with short half-lives. Radionuclide generator systems are widely used for supply of short-lived radionuclides/radiopharmaceuticals. Several generator systems are available and routinely in use within nuclear medicine. Some of these are listed in Table 1.

Because of the short half-life, the coupling of the radionuclide to the carrier molecule must be done immediately before the administration. Hence, there is a need to have a constant supply of carrier molecules that can be labeled efficiently on site. For this purpose, several preparation kits have been developed.

Ready-for-use diagnostic radiopharmaceuticals which are intended for transport over some distance typically include radionuclides with half-lives from 13 h and up.

Among these, products involving the radionuclide $^{131}$I are used for both diagnostic and therapeutic indications. This is based upon the mixed emitting properties of the radionuclide, giving both β and γ emission. The availability, price, and half-life (8 days) of this radionuclide, together with the physical properties, have probably made it the most commonly used radionuclide in radiotherapy. Although $^{131}$I also is frequently used for diagnostic purposes, the radiation characteristics of this radionuclide are not really ideal for use in conventional scintigraphy (SPECT) due to the high γ energies. In addition, the β emission from this radionuclide gives the patients an unnecessary radiation burden. Hence, other radionuclides are preferred for use in diagnostic nuclear medicine.

The radioiodine $^{123}$I, on the other hand, is very useful in nuclear medicine because it has good radiation characteristics for scintigraphy, such as decay by electron capture, a half-life of 13 h, and γ emmision of 159 keV. However, the much shorter half-life, together with the more complex radionuclide production, makes this radionuclide less available and more expensive compared to $^{131}$I.

There are several $^{131}$I and $^{123}$I radiopharmaceuticals on the market, for both oral and parenteral administration. Ready-for-use radiopharmaceuticals that contain

**TABLE 1 Several Radionuclide Generator Systems Useful in Nuclear Medicine**

| Parent Nuclide | $t_{1/2}$ | | Daughter Nuclide | $t_{1/2}$ |
|---|---|---|---|---|
| $^{68}$Ge | 280 days | → | $^{68}$Ga | 68 min |
| $^{81}$Rb | 4.7 h | → | $^{81m}$Kr | 13 s |
| $^{99}$Mo | 66 h | → | $^{99m}$Tc | 6 h |
| $^{113}$Sn | 117 days | → | $^{113m}$In | 100 min |
| $^{188}$W | 69.4 days | → | $^{188}$Re | 17 h |

these radionuclides will normally be manufactured by radiopharmaceutical companies and distributed to the marked according to a marketing authorization (MA).

Although therapeutic application represents less than 10% of the nuclear medicine investigations, therapeutic radiopharmaceuticals are a very important group of radiopharmaceuticals. Hence, a brief description is outlined for production of therapeutic radiopharmaceuticals following some other selected groups of radiopharmaceuticals.

***$^{99}$Mo/$^{99m}$Tc Generators*** The essential part of the most commonly available generator system is a simple chromatography column to which the mother radionuclide is absorbed on a suitable support material. The daughter radionuclide is a decay product of the mother nuclide. Since it is the daughter nuclide that is used to label the carrier molecules, it must be possible to separate this from the parent nuclide by a chemical separation.

In a $^{99}$Mo/$^{99m}$Tc generator, the $^{99}$Mo (molybdenum) is fixed as molybdate to aluminum oxide in the column. The daughter nuclide, $^{99m}$Tc (technetium), is eluted from the column as pertechnetate when using saline solution. Molybdenum-99 has a half-life of 66 h, while $^{99m}$Tc has a half-life of 6 h. This is an ideal combination of half-lives, giving a system where the daily supply of $^{99m}$Tc can easily be calculated from the known amount of $^{99}$Mo on the column. The half-life of $^{99m}$Tc, along with the radiation characteristics of the nuclide, makes it excellent for use in nuclear medicine imaging. After reconstitution of kits and formation of various radiopharmaceuticals, this radionuclide is used in a major part of all nuclear medicine procedures.

Although the principle for the generators is similar, the design of $^{99}$Mo/$^{99m}$Tc generators from different manufacturers can differ a lot. A drawing of a $^{99}$Mo/$^{99m}$Tc generator is shown in Figure 4. In general, the generator consists of a column with adsorbent material where the radionuclide $^{99}$Mo is applied. The column is combined with a needle system necessary for the elution process. A sterile filter is fitted on the air inlet side of the needles to keep an aseptic system during elution. The saline solution for elution may be supplied as a bulk solution sufficient for several elutions

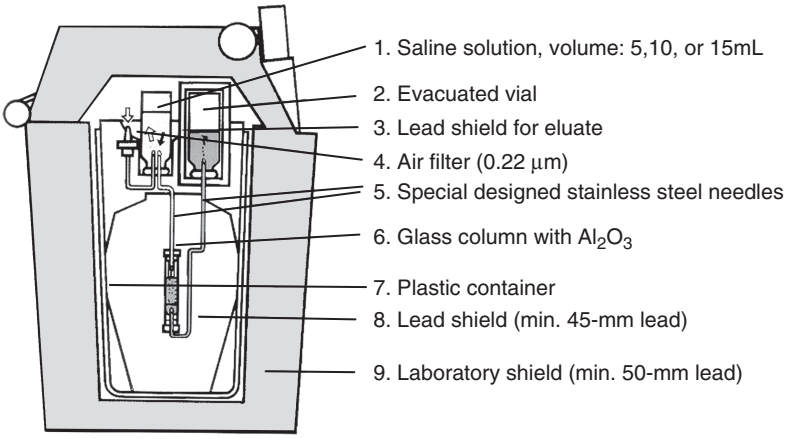


**FIGURE 4** Typical radionuclide generator system (ISOTEC, GE Healthcare, AS).

or dispensed volumes sufficient for a single elution. For both, vacuum is normally used to run the elution of the column using sterile evacuated vials.

Finally, due to the relatively high radiation from $^{99}$Mo, the system must be properly shielded by either lead or a combination of lead and tungsten.

Whether the column is designed to contain liquid after and between elutions, determine if this is a wet-column generator or a dry-column generator. When liquid is retained at the column (wet generator), radiolysis of water on the column may occur as a result β irradiation from $^{99}$Mo. This may change the chemistry on the column and thus reduce the yield when eluting the generator. Most commonly, when manufacturing wet-column generators, oxidizing agents are added either to the saline or to the column itself to avoid reduction of pertechnetate on the column.

A radionuclide generator must be sterile and pyrogen free. Most commonly, the generator is sterilized by autoclaving the entire column after the molybdate has been bound to the aluminum oxide. Other critical procedures during the production and the assembly of the generator must be performed under aseptic conditions. Elution of the generator must also be carried out under aseptic conditions while using only sterile accessories.

***Other Generators*** Of the generators listed in Table 1, two systems are of particular interest in nuclear medicine today along with the $^{99}$Mo/$^{99m}$Tc generator, namely the $^{68}$Ge/$^{68}$Ga generator and the $^{81}$Rb/$^{81m}$Kr generator.

*$^{68}$Ge/$^{68}$Ga Generator* Germanium-68 has a half-life of 271 days, and $^{68}$Ga (gallium) a half-life of 68 min. Gallium-68 is a PET emitter, and this generator system is a valuable source of a short-lived radionuclide in a radiopharmacy or nuclear medicine department. However, the system is not as easy or efficient as the $^{99}$Mo/$^{99m}$Tc generator. On the other hand, the longer half-life of the mother nuclide allows use of the system for several months.

This generator can be made up of aluminum loaded on a plastic or glass column. Carrier-free $^{68}$Ge in concentrated HCl is neutralized in ethylenediaminetetraacetic acid (EDTA) solution and adsorbed to the column. Then $^{68}$Ga is eluted from the column with 0.005 *M* EDTA solution. Alternatively, $^{68}$Ge is adsorbed on a stannous dioxide column and $^{68}$Ga is eluted with 1 *N* HCl [2].

*$^{81}$Rb/$^{81m}$Kr Generator* Rubidium-81 has a half-life of 4.6 h and decays to $^{81m}$Kr (krypton) by electron capture. Krypton-81m has a half-life of 13 s and decays by isomeric transition emitting γ rays of 190 keV. Being an inert gas $^{81m}$Kr is used for lung ventilation study.

The parent $^{81}$Rb is adsorbed on an ion exchange resin, and the daughter $^{81m}$Kr is eluted with air. Because of the very short half-life of 13 s, the studies can be repeated every few minutes, and no radiation safety precaution for trapping $^{81m}$Kr is needed [2].

***Radiopharmaceutical Kits*** Radiopharmaceutical kits are nonradioactive ("cold") products containing the sterile ingredients needed to prepare the final radiopharmaceutical. Immediately before administration to the patient, the radionuclide is added. From the point of licensing, these semimanufactured products are defined as radiopharmaceuticals, as they have no other application in medicine [2].

Most of these preparation kits have been developed for labeling of various substances with $^{99m}$Tc. Labeling is normally a single- or two-step procedure consisting of adding a solution of $^{99m}$Tc-pertechnetate to the preparation kit. The preparation kit contains the ingredient necessary for labeling, such as the substance or ligand to be labeled, a reducing agent, buffers for pH adjustments, and various stabilizers. The reducing agent, very often a stannous salt, is added to bring the radionuclide into a valence state with high reactivity.

Most preparation kits are lyophilized, and the reason for this is to extend the shelf life of the products. Some preparation kits can in fact be stored for more than one year. Since these products are not radioactive, conventional clean rooms and clean-room technology can be applied for production of preparation kits. Most of these products have to be produced aseptically, as they cannot be sterilized with other methods. During lyophilization of the preparation kits used for $^{99m}$Tc labeling, it is very important to remove all the oxygen from the kit vial. This is to ensure the right valence of the tin salt. Normally, the vials are filled with an inert gas, such as nitrogen, before the vials are closed completely. It is important, though, that the gas is dried. Some manufacturer chose to not completely replace the removed oxygen, giving a slightly negative pressure inside the kit vial. This may be favorable for the kit-labeling procedure.

***Therapeutic Radiopharmaceuticals*** Radiopharmaceuticals used for therapy (radiotherapy) are designed such that, after administration, they act locally at a target by either damaging or killing cells by irradiation. One of the attractions of radionuclide therapy is the existence of radiation with quite different dimensions of effectiveness, ranging from subcellular (Auger electrons) to hundreds of cell diameters (β particles). In between, α emitters have a tissue range equivalent to only a few cell diameters [9]. Alpha emitters have a very high linear energy transfer (LET), being very potent at short distances.

Table 2 lists a selection of radionuclides and radiopharmaceuticals used in radiotheraphy.

**TABLE 2 Selected Radionuclides and Radiopharmaceuticals Used for Radiotherapy in Routine Use or as Part of Clinical Investigations**

| Radionuclide | Mode of decay | $t_{1/2}$ | Radiopharmaceuticals |
|---|---|---|---|
| $^{131}$I | $\beta^-/\gamma$ | 8.04 days | $^{131}$I-NaI, $^{131}$I-MIBG, $^{131}$I-mAbs |
| $^{90}$Y | $\beta^-$ | 2.7 days | $^{90}$Y-colloid, $^{90}$Y-DOTATOC, $^{90}$Y-mAbs |
| $^{186}$Re | $\beta^-/\gamma$ | 3.8 days | $^{186}$Re-sulfide, $^{186}$Re-HEDP |
| $^{188}$Re | $\beta^-/\gamma$ | 17 h | $^{188}$Re-HEDP |
| $^{177}$Lu | $\beta^-/\gamma$ | 6.6 days | $^{177}$Lu-DOTA-Tyr3-octreotide |
| $^{153}$Sm | $\beta^-/\gamma$ | 1.9 days | $^{153}$Sm-EDTMP |
| $^{89}$Sr | $\beta^-$ | 50.6 days | $^{89}$Sr-chloride |
| $^{223}$Ra | $\alpha/\gamma$ | 11.4 days | $^{223}$Ra-chloride |
| $^{211}$At | α | 7.2 h | $^{211}$At-mAbs |
| $^{213}$Bi | α | 46 min | $^{213}$Bi-mAbs |
| $^{166}$Ho | $\beta^-/\gamma$ | 26.8 days | $^{166}$Ho-colloid |
| $^{169}$Er | $\beta^-$ | 9.4 days | $^{169}$Er-citrate colloid |
| $^{165}$Dy | $\beta^-/\gamma$ | 2.3 h | $^{165}$Dy-ferric hydroxide macroaggregate |
| $^{32}$P | β– | 14.3 days | $^{32}$P-*ortho*-phosphate |

Pure α and β emitters are easy to shield, and thus production involving these can be performed in sealed production units with no lead protection. One must keep in mind, though, the potential hazard when inhaling some of these materials. Moreover, many radionuclides used for radiotherapy have an additional γ component. Hence, local lead shielding may be necessary. If the γ component is larger or represents very high energy emission, a total lead shielded unit may be necessary. The latter will be the case when manufacturing $^{131}$I radiopharmaceuticals for therapy, since $^{131}$I is a radionuclide consisting of a high-energy γ photon together with the β component.

Radiopharmaceuticals for therapeutic use must have a high target-to-background ratio. Targeted radiotherapy involves the use of molecular carrier such as a receptor-avid compound or an antibody to deliver a radionuclide to cell populations.

A challenge when performing radiolabeling of carrier molecules for targeted radiotherapy is the potential risk of radiolysis due to the radiation characteristics of the radionuclides involved. When increasing the specific activity, as part of the scaling up, the risk of radiolytic decomposition of the labeled compound also increases. This is particularly pronounced when using α emitters. The addition of stabilizers in the form of scavengers can reduce this risk. Benzyl alcohol is an example of a compound that acts as a scavenger by catching up with free radicals in the solution.

Another approach is to use kit formulations also for this kind of product. Therapeutic radiopharmaceuticals have been developed where the carrier molecule is formulated in a lyophilized kit and supplied together with the radionuclide. An example of this is the MAb ibritumomab tiuxetan formulated for labeling with the β-emitting radionuclide $^{90}$Y. Yttrium-90 ibritumomab tiuxetan (Zevalin) is used in the treatment of non-Hodgkin's lymphoma (NHL).

The labeling is performed in a centralized radiopharmacy, hospital radiopharmacy, or nuclear medicine department immediately before use.

***Radioactive Sanitary Products*** Radioactive sanitary products could be considered as radiopharmaceuticals according to the definition given in directive 2001/83/EC, although there are significant differences between radioactive sanitary products and classical radiopharmaceuticals. The former can in fact be considered as encapsulated radioactive sources, although with the use of microencapsulated sanitary products (such as micrometer-sized glass or polymer beads loaded with a radionuclide), the difference between both types is becoming more difficult to establish.

In any case, radioactive sanitary products are delivered locally (and not systemically or orally) for the local treatment of a disease. The idea is to give a high dose of radiation to a specific part of the body by the implantation of the corresponding sanitary product in the desired zone. The sanitary product must not be metabolized, destroyed, or removed from the place it has been located during a sufficiently long time as to give the desired high radiation dose.

The most commonly used radioactive sanitary products are millimeter-sized seeds or needles loaded with $^{103}$P, $^{192}$Ir, $^{90}$Sr, or $^{125}$I. Currently micrometer-sized or even nanometer-sized beads loaded with $^{90}$Y are being used for the treatment of specific diseases.

***PET Radiopharmaceuticals*** PET radiopharmaceuticals are labeled with short-lived positron-emitting radionuclides. Such radionuclides can either be produced in a cyclotron or obtained from an appropriate radionuclide generator.

*General Considerations* The synthesis of PET radiopharmaceuticals has several peculiarities substantially different from the procedures followed to prepare conventional γ-emitting radiopharmaceuticals. A very important issue that must be considered is the specific activity. For all radiopharmaceuticals it is usually very high and can be calculated from the formula

$$A_e = \frac{k}{AT_{1/2}}$$

where $A_e$ is the specific activity, $A$ the mass number of the radionuclide, and $T_{1/2}$ its half-life. It is then clear that the achievable specific activity is higher for radionuclides with shorter half-lives, as is the case for the most relevant PET radionuclides ($^{18}F$ and $^{11}C$). As an example, $^{18}F$ produced in no-carrier-added conditions can be obtained with specific activities of almost $10^{10}$ Ci/mmol, resulting in PET radiopharmaceuticals with extremely high specific activities.

For PET radiopharmaceuticals we must always consider that synthesis processes must be extremely fast. Consequently, synthesis schemes with as few steps as possible must be used, and each of the steps must proceed with high efficiency. The incorporation of the radionuclide to the molecule should ideally be done in the final steps of the synthesis. In this way two objectives can be achieved: reduce the overall synthesis time (thus increasing the yield) and reduce the number of side reactions and secondary undesired products obtained during the synthesis.

The synthesis of PET radiopharmaceuticals is always carried out at very small scale (only a few dozen micrograms of the radiopharmaceutical are obtained) and each batch can sometimes only be used for a single patient or a few patients at most. Consequently, there is always a big excess of the precursor in the reaction medium, and proper purification systems must be used to get rid of all the possible contaminants. Such systems must also be very efficient and fast, and the most usual is to apply either semipreparative high-performance liquid chromatography (HPLC) or solid-phase extraction-based procedures.

The position of the radionuclide in the molecule of interest is also critical as it will affect the biological behavior of the radiopharmaceutical. Chemical reactions must be designed to be stereospecific in many cases, as the production of a mixture of different stereoisomers complicates the purification of the final radiopharmaceutical. Synthesis procedures must also be easy to automate, as very elevated activities are used for the synthesis of PET radiopharmaceuticals (several curies usually) and appropriate radiation protection systems must be used.

*PET Generators* Table 3 summarizes the characteristics of some PET generators. So far, the most widely used system has been the $^{82}Sr/^{82}Rb$ generator, although due to the specific physical and chemical characteristics of the daughter radionuclide and the half-life of the parent radionuclide, the $^{68}Ge/^{68}Ga$ generator is probably one of the most interesting systems. Recent advances in gallium chemistry have permitted the development of $^{68}Ga$ radiopharmaceuticals of clinical interest making

**TABLE 3 Selected of PET Generators**

| Generator | Parent | $T_{1/2}$ | Daughter | $T_{1/2}$ |
|---|---|---|---|---|
| Fe/Mn | $^{52}$Fe | 8.27 h | $^{52m}$Mn | 21.1 min |
| Zn/Cu | $^{62}$Zn | 9.13 h | $^{62}$Cu | 9.73 min |
| Ge/Ga | $^{68}$Ge | 270.8 days | $^{68}$Ga | 68.3 min |
| Sr/Rb | $^{82}$Sr | 25.6 days | $^{82}$Rb | 76.4 s |

**TABLE 4 Physical Characteristics of Some Positron Emitters of Clinical Use**

| Isotope | $T_{1/2}$ (min) | % | $E_{\beta+}$ (keV) |
|---|---|---|---|
| $^{11}$C | 20.4 | 99.7 | 960 |
| $^{13}$N | 9.9 | 99.8 | 1198 |
| $^{15}$O | 2.0 | 99.9 | 1732 |
| $^{18}$F | 109.6 | 96.7 | 634 |

**TABLE 5 Nuclear Reactions for Production of Most Widely Used Positron Emitters**

| $^{11}$C | $^{13}$N | $^{15}$O | $^{18}$F |
|---|---|---|---|
| **$^{14}$N(p,α)$^{11}$C** | **$^{16}$O(p,α)$^{13}$N** | **$^{14}$N(d,n)$^{15}$O** | **$^{18}$O(p,n)$^{18}$F** |
| $^{10}$B(d,n)$^{11}$C | $^{13}$C(p,n)$^{13}$N | $^{15}$N(p,n)$^{15}$O | $^{20}$Ne(d,α)$^{18}$F |
| $^{11}$B(d,2n)$^{11}$C | $^{12}$C(d,n)$^{13}$N | | $^{16}$O(α,pn)$^{18}$F |
| $^{11}$B(p,n)$^{11}$C | | | $^{19}$F(p,pn)$^{18}$F |
| $^{12}$C(p,pn)$^{11}$C | | | |

*Note:* Most common reactions used in small cyclotrons are bolded. Different energies of the incident particle are needed for the different nuclear reactions

available PET studies at stand-alone PET centers without a cyclotron with other compounds different from the classical $^{18}$FDG.

*PET Cyclotrons* Cyclotrons used to produce positron emitters of clinical interest (see Tables 4 and 5), mainly $^{18}$F and $^{11}$C, do not need to be very big. In fact, small devices installed in hospital or academic institutions have long been used for such purposes (see Figure 5). These devices are easy to operate and maintain, and even with single-particle low-energy cyclotrons, it is possible to produce multicurie amounts of $^{18}$F and $^{11}$C.

*Some Positron Emitters of Clinical Interest* Fluorine-18 is undoubtedly the most widely used positron-emitting radionuclide. This is mainly due to the wide use of $^{18}$FDG, the PET radiopharmaceutical that has permitted PET to become an everyday clinical tool. With the exception of $^{18}$FDG and probably $^{18}$FDOPA, the use of other $^{18}$F-labeled radiopharmaceuticals is very limited. However, the chemical and physical characteristics of $^{18}$F are excellent:

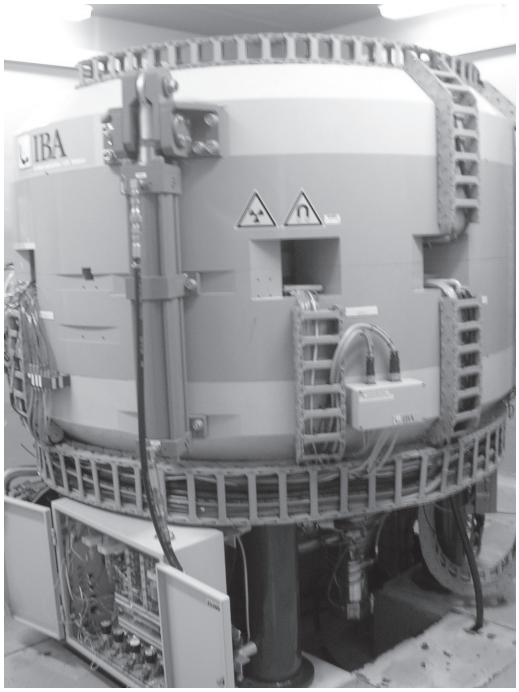


**FIGURE 5** Small (less than 2 m in diameter) dual-beam negative ion cyclotron capable of easily producing multicurie amounts of $^{18}F$ and $^{11}C$. (Photo courtesy of PET-CUN Center, University of Navarra.)

It can easily be produced in very high quantities (up to 7–9 Ci per batch) even in small cyclotrons with just a few hours irradiation time.

The mean positron emission energy of $^{18}F$ is just 0.64 MeV (the lowest of all positron emitters with clinical use) and this has several important consequences: The dose of radiation received by the patient will be lower and the distance between disintegration of the radionuclide and the annihilation site (after collision of the positron with an electron) is reduced, thus making PET images with higher resolution possible.

The half-life of $^{18}F$ (109 min) is sufficiently long to carry out complex synthesis procedures, apply long PET imaging protocols, and carry out metabolite analysis. Furthermore, it is possible to produce the radiopharmaceutical in a laboratory and transport it to a distant site only equipped with an imaging device. These kinds of "satellite PET centers" have boomed all around the world and permitted the fast expansion of PET as an everyday clinical tool in certain pathologies (mainly in oncological diseases).

Fluorine is not common in biological molecules, but many drugs contain this atom. Fluorine and hydrogen have quite similar radii, and changing a hydrogen to a fluorine atom in a molecule does not usually generate substantial steric differences between both molecules. Nonetheless, the electronegativity of fluorine is usually

going to change substantially the physicochemical properties of the molecule (reactivity, hydrogen bonding, interactions with cognate receptors, metabolization, etc.). It is not possible to assume that the biological behavior of a molecule and its fluorinated analog is going to be similar. On the contrary, it is advisable to find substantial differences in lipophilicity, biodistribution, protein binding, affinity for receptors, and so on. However, such modifications are in many cases very useful to permit the use of a $^{18}$F-fluorinated analog as a PET radiopharmaceutical. In fact, that is the case for the most widely use one: FDG. This compound, which accounts for probably more than 90% of the PET studies performed in the world every day, is a glucose analog that is taken up by the cells by GLUT transporters and metabolized just as glucose at the very first steps of glicolysis. But as a consequence of the change of the $C_2$ OH group in natural glucose by a $^{18}$F atom in FDG, the latter cannot be isomerized (once phosphorilated) and suffers metabolic trapping being specifically accumulated in tumoral cells.

Carbon-11 has a very short half-life (just 20.4 min) but the chance to substitute a carbon atom in any biological molecule by a positron-emitting $^{11}$C is a very interesting possibility. This has led to a substantial development of $^{11}$C-labeled tracers. The short half-life conditions everything and only PET centers equipped with a cyclotron can have a clinical program with $^{11}$C tracers. The production of the radiopharmaceutical must in these cases be performed just before the imaging study and is usually not started until the patient is already on the PET scanner.

The $^{12}$C–$^{11}$C substitution will produce chemically identical molecules and give the chance to study many biological processes by this noninvasive methodology and can also be used in new-drug research and development (R&D).

*Synthesis of PET Radiopharmaceuticals* Albeit the requirements for the synthesis of PET radiopharmaceuticals previously described, the synthesis process could conceptually be reduced to a very simple scheme, as shown in Figure 6.

The concept is really simple, but there are considerable difficulties in each of the steps. In many cases it is difficult to synthesize a properly designed cold precursor that will permit a simple direct reaction with few secondary products. No modifica-

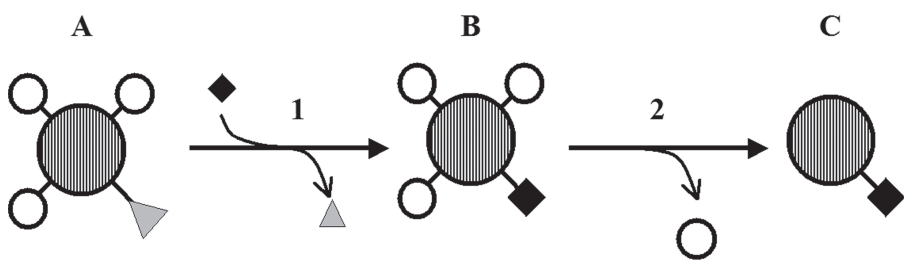


**FIGURE 6** General reaction scheme for synthesis of PET radiopharmaceuticals. The precursor molecule (A) is designed with the adequate protecting groups (○) and a reactive leaving group (Δ). A reactive form of the radionuclide (■) is covalently joined to the precursor at the reaction site, while the leaving group is eliminated. An intermediate radioactive product (B) is obtained that is hence deprotected (2) to produce the final radiopharmaceutical (C). A fast and efficient purification process of C is needed to get read of unreacted cold precursor, radionuclide, and intermediate products.

tions in the configuration of the chiral centers should be produced during the overall process and a simple purification system able to purify the final product in a very short time should be found. Additionally, all the reactions should be very fast (just several minutes at most) and be easy to automate to be performed in a computer-controlled device placed in a shielded hot cell.

*Production Process and Quality Control* The production process includes the following:

- Production of the radionuclide in the cyclotron and sending it to the PET radiopharmaceutical laboratory
- Reaction of the radionuclide with an appropriate cold precursor, either in solution or in solid phase
- Purification of the radiopharmaceutical, usually by semipreparative radio HPLC or solid-phase extraction
- Formulation of the final product as an injectable solution (frequently including phase change in a rotary evaporator) and the adjustment of tonicity and pH
- Sterile filtration or autoclaving

The quality control of the final product must be carried out before release of the batch (except for the sterility and the endotoxin tests for extremely short-lived radionuclides). Consequently, all procedures must not only be very fast but also very accurate, and in all cases it is very important to have a properly established quality assurance system that might permit parametric release of the produced batches. The quality control assays that must be carried out in the radiopharmaceutical includ the following:

- Radionuclidic purity
- Radionuclidic identity
- Chemical purity
- Radiochemical purity
- Specific activity
- Residual solvents
- Visual inspection
- Tonicity
- pH
- Sterility
- Endotoxin

A PET radiopharmaceutical laboratory must include the cyclotron bunker (where positron-emitting radionuclides are produced), the production laboratory, the quality control laboratory, and several different ancillary areas.

In the production laboratory all synthesis and purification processes are carried out in remote-operated fully automated computer-controlled systems (synthesis modules, see Figure 7) located in heavily shielded hot cells (see Figure 8). Dispensing of individual doses is in many cases also carried out by automated systems.

**FIGURE 7** Automated synthesis module for PET radiopharmaceutical synthesis located in a shielded hot cell. (Photo courtesy of PET-CUN Center, University of Navarra.)

**FIGURE 8** Production laboratory for PET radiopharmaceuticals. The 10-cm lead shielded hot cells contain computer-controlled automated synthesis modules. (Photo courtesy of PET-CUN Center, University of Navarra.)

## 5 QUALITY CONSIDERATIONS

### 5.1 Documentation

Good documentation constitutes an essential part of the quality assurance system. As claimed in the European Community (EC) Guide to Good Manufacturing Practice (GMP), Chapter 4: "Clearly written documentation prevents errors from spoken communications and permits tracing of batch history." In general, the requirements for documentation related to manufacturing of pharmaceuticals, as set in the GMP

regulations, are also valid for manufacturing of radiopharmaceuticals. A recent draft proposal of EC GMP Annex 3, "Manufacture of Radiopharmaceuticals," outlines the following regarding this issue:

All documents related to the manufacture of radiopharmaceuticals should be prepared, reviewed, approved, and distributed according to written procedures.

Specifications should be established and documented for raw materials, labeling and packaging materials, critical intermediates, the finished radiopharmaceutical, and any other critical material.

Acceptance criteria should be established for the radiopharmaceutical, including criteria for release and shelf life specifications.

Records of major equipment use, cleaning, sanitization or sterilization, and maintenance should show the product, batch number, date and time, and signatures of the persons involved.

Records should be retained for at least three years unless another time frame is specified in national requirements.

It is of utmost importance to have a system for implementing such documents. Any new master document or a new version of such a document must be followed by a training process for relevant operators. This training must be recorded as well.

The recording of production data will make it necessary to bring batch documentation into the radioisotope laboratory. Hence, it is important to have routines that minimize the risk for radioactive contamination of the documents and to ensure that any contaminated documents will not leave the controlled area. Today, the use of computers instead of paper documents in the laboratory leaves most of the paperwork outside the controlled area.

### 5.2 Qualification of Personnel

As a general principle in GMP, there should be sufficient qualified personnel to carry out all the tasks that are the responsibility of the manufacturer. Furthermore, individual responsibilities should be clearly understood by the individuals and recorded.

For personnel working with radiopharmaceuticals, training and qualification should cover general principles of GMP and radiation protection. This includes also personnel in charge of cleaning premises and equipment used for this type of production. All manufacturing operations should be carried out under the responsibility of a QP with additional competence in radiation protection.

### 5.3 Quality Control

All quality control procedures that are applied to nonradioactive pharmaceuticals are in principle applicable to radiopharmaceuticals. In addition, tests for radionuclidic and radiochemical purity must be carried out. Furthermore, since radiopharmaceuticals are short-lived products, methods used for quality control should

be fast and effective. Still, some radiopharmaceuticals with very short half-lives may have to be distributed and used after assessment of batch documentation even though all quality control tests have not been completed. It is acceptable, though, for these products to be released in a two-stage process, before and after full analytical testing. In this case there should be a written procedure detailing all production and quality control data that should be considered before the batch is dispatched. A procedure should also describe the measures to be taken by the QP if unsatisfactory test results are obtained after dispatch (GMP, Annex 3).

The quality control tests fall in two categories: biological tests and physiochemical tests. The biological tests establish the sterility and apyrogenicity, while the physiochemical tests include radionuclidic, chemical, and radiochemical purity tests along with determination of pH, osmotic pressure, and physical state of the sample (for colloids).

For lyophilized preparation kits containing reducing agents, such as $^{99m}$Tc kits, a test for moisture content can be necessary. Residual water in the freeze-dried pellet may lead to oxidation of the reducing agent.

***Radionuclidic Purity*** Radionuclidic purity is defined as the fraction of the total radioactivity in the form of the desired radionuclide present in a radiopharmaceutical. Radionuclide impurities may arise from impurities in the target material or from fission of heavy elements in the reactor [2]. In radionuclide generator systems, the appearance of the parent nuclide in the daughter nuclide product is a radionuclidic impurity. In a $^{99}$Mo/$^{99m}$Tc generator, $^{99}$Mo may be found in the $^{99m}$Tc eluate due to breakthrough of $^{99}$Mo on the aluminum column. The presence of these extraneous radionuclides increases the radiation dose to the patient and may also obscure the scintigraphic image.

Radionuclidic purity is determined by measuring the characteristic radiations emitted by individual radionuclides. Gamma emitters are distinguished from another by identification of their $\gamma$ energies on the spectra obtained from a NaI crystal or a Ge (germanium) detector. This method is called $\gamma$ spectroscopy.

Pure $\beta$ emitters are not as easy to check as the $\gamma$ emitters. However, they may be checked for purity with a $\beta$ spectrometer or a liquid scintillation counter.

***Radiochemical Purity*** The radiochemical purity (RCP) of a radiopharmaceutical is the fraction of the total radioactivity in the desired chemical form in the radiopharmaceutical. Radiochemical impurities arise from decomposition due to the action of solvent, change in temperature or pH, light, presence of oxidizing or reducing agents, and radiolysis [2]. Examples of radiochemical purity are free $^{99m}$Tc-pertechenetate and hydrolyzed $^{99m}$Tc in labeled $^{99m}$Tc radiopharmaceuticals. The presence of radiochemical impurities in a radiopharmaceutical results in poor-quality images due to the high background from the surrounding tissues and blood. It also gives the patient unnecessary radiation doses.

A number of analytical methods are used to detect and determine the radiochemical impurities in a given radiopharmaceutical. Most commonly used are methods like paper (PC), thin-layer (TLC), and gel chromatography, paper and gel electrophoresis, HPLC, and precipitation. A common principle for the different methods is that they can chemically separate the different radiolabeled components in the radiopharmaceutical. It may sometimes be necessary to perform more than

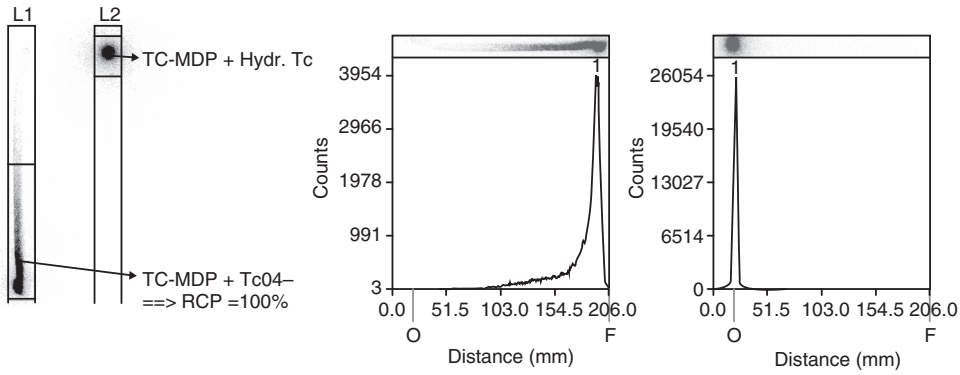


**FIGURE 9 Typical chromatograms for $^{99m}$Tc-MDP.** The left strip and chromatogram are obtained with ITLC-SG in sodium acetate. The right strip and chromatogram are obtained in methyl ethyl ketone (MEK). When combining these, any free pertechnetate ($^{99m}TcO^{4-}$) and/or hydrolyzed $^{99m}$Tc can be detected. Thus the fraction representing $^{99m}$Tc-MDP (RCP) can be calculated.

one test method, for instance, TLC and HPLC, to get a complete picture of the different radiochemical impurities. Alternatively, one can use one chromatographic method consisting of a constant stationary phase but varying the mobile phase (solvent). An example is the radiochemical purity test of $^{99m}$Tc-methylenediphosphate (MDP), a radiolabeled phosphate used in bone scintigraphy. When using two TLC systems, one with sodium acetate as a solvent and one with methyl ethyl ketone (MEK) as a solvent, the different $^{99m}$Tc compunds in the product can be determined. A small aliquot of the radiopharmaceutical preparation is spotted on an instant thin-layer chromatography (ITLC) strip. The strip is dipped into the chromatography flask while keeping the spot above the solvent. During the chromatography process, the different components of the sample distribute differently in the ITLC strip, depending on the solubility and polarity of the components. In systems like this, each component is characterized by an $R_f$ value, defined as the ratio of the distance traveled by the component to the distance the solvent front has advanced from the original point of application of the test material. The distribution of the radioactive components on the strips can be monitored by use of an appropriate device for measuring radioactivity and printed in a chromatogram. Figure 9 shows typical chromatograms for $^{99m}$Tc-MDP in the TLC systems described above.

***Chemical Purity*** The chemical purity of a radiopharmaceutical is the fraction of the material in the desired chemical form. Chemical impurities may arise from the breakdown of the material either before or after labeling. Chemical impurities may also arise from the manufacturing process, such as aluminum in a $^{99m}$Tc eluate, coming from the aluminum column on the generator. Residuals of solvent from the radiopharmaceutical synthesis are also considered as chemical impurities. If the chemical impurity is present before labeling, the result may be undesirable labeled molecules. Furthermore, chemical impurities may cause a toxic effect. High-performance liquid chromatography and gas chromatography (GC) are important methods for determination of chemical impurities in a radiopharmaceutical.

***Sterility and Pyrogen Testing*** Sterility indicates the absence of any viable bacteria or microorganisms in a radiopharmaceutical preparation. Hence, sterility testing is performed to prove that radiopharmaceuticals are essentially free of viable microorganism. The test for microbial contamination of these products is best carried out with filter methods. It is a great advantage to incubate only the filters instead of the radioactive solutions.

The test is performed according to the Ph.Eur/USP monograph on Sterility tests [13, 14], but with an important modification. Small batch sizes, typical for radiopharmaceuticals, make it necessary to use smaller test volumes than required in the monographs. Also the risk for radiation exposure supports this modification.

All radiopharmaceuticals for human administration are required to be pyrogen free. Also the tests for apyrogenicity must be modified when applied for these products. The classical rabbit test for pyrogens was never a convenient test for parenteral radiopharmaceuticals. Practical problems due to radioactive rabbits and the need for larger test volumes made this a difficult task. Today, the Limulus amebocyte test (LAL) is the method of choice and has been accepted by the Ph. monographs for many years. This test is normally done within an hour, compared to several days for the rabbit test.

However, even the LAL test may be too time consuming for the very short lived PET radiopharmaceuticals. Hence, less time consuming methods are in progress and will probably improve this situation. Meanwhile, it is accepted that the test for apyrogenicity, like the sterility test is for most radiopharmaceuticals, is finished after release of the most short lived radiopharmaceuticals.

***Bubble Point Testing of Filters*** Parenteral radiopharmaceuticals that are not terminally sterilized must undergo a sterile filtration process as part of the aseptic production procedure. Although the supplier certifies the filters used, they must be checked for integrity after use to assure that there has been no leakage during the filtration. The integrity of the filter may be demonstrated by *bubble point testing*. In this test, the filter is placed and monitored under controlled pressure. When the test is done on wet filters, the pressure needed to push gas through the filter is defined as the bubble point. A filter with given pore width has a corresponding bubble point value. Most frequently, sterile filtration is performed by 0.22-μm filters; hence the bubble point is about 3–4 bars. However, the filter supplier should specify the bubble point valid for a specific filter.

Since this is an in-process test, special caution must be given to radiation protection. The test equipment should be placed within a closed and shielded unit and a system should be in place to collect any radioactive spill from the test.

When the filter integrity test fails, the sterile filtration process must be rejected.

***Visual Inspection of Finished Product*** As part of the quality control, all parenterals will be subject to an inspection for the possible content of particles. Visual inspection of radiopharmaceuticals is more complicated than for other pharmaceuticals, as radiation protection guidelines strongly discourage any direct eye contact with radioactive sources. Normally, the visual inspection of a radiopharmaceutical is performed by placing the vial on a rotating station connected to a camera. The station is properly shielded, and the operators can study the solution on a distant screen.

### 5.4 Validation and Control of Equipment and Procedures

Preventive maintenance, calibration, and qualification programs should be operated to ensure that all facilities and equipment used in the manufacture of radiopharmaceuticals are suitable and qualified (GMP, Annex 3). Special emphasis should be put on critical equipment for handling of radiopharmaceuticals, such as dose calibrators that are used to check the accuracy of the dispensing of patient doses. Particular programs are outlined for checking the dose calibrator, including constancy, accuracy, linearity, and geometry. The general principles of validation outlined in the GMP regulations are valid for radiopharmaceuticals as well as for other pharmaceuticals. All validation activities should be planned and clearly defined and documented in a validation master plan (VMP). Special emphasis should be given on the validation of aseptic processes in the production of radiopharmaceuticals. Studies, including media fill tests, must be performed and recorded to demonstrate maintenance of sterility throughout the production process. This is particularly important since most radiopharmaceuticals are dispatched and used before the sterility test is finished.

### 5.5 Stability Aspects of Radiopharmaceuticals

As discussed already, radiopharmaceuticals are exposed to stability problems, particularly when radiolabeled compounds are involved. Decomposition of labeled compounds by radiolysis depends on the specific activity of the radioactive material, the energy of the emitted radiation, and the half-life of the radionuclide. Particles, such as $\alpha$ and $\beta$ radiation, are more damaging than $\gamma$ rays, due to their short range and local absorption in matter. The stability of a compound is time dependent on exposure to light, change in temperature, and radiolysis. The longer a compound is exposed to these conditions, the more it will tend to break down.

Stabilizers such as ascorbic acid and benzyl alcohol may be added to inhibit or delay the decompostion. Many preparations are stored in the dark under refrigeration to slow down the degradation of the material [2]. The expiry date of a radiopharmaceutical is based upon data from stability studies designed to demonstrate the described effects on the product after storage.

Hence, for most stability studies on radiolabeled compounds, the radiochemical purity and pH are the most important physiochemical parameters to study. Moreover, for parenteral radiopharmaceuticals, a stability study also has to demonstrate the maintenance of sterility and apyrogenicity after storage.

## 6 EXTEMPORANEOUS PREPARATION OF RADIOPHARMACEUTICALS

An extemporaneous preparation is defined as a product which is dispensed immediately after preparation and not kept in stock [10]. Hence, many radiopharmaceuticals could fall into this category due to their limited shelf life.

The use of extemporaneous preparation should be limited to situations where there is no product with marketing authorization (MA) available. This could be prepared based upon a prescription for a named patient (magistral preparation) or a production based upon a formula and prepared on a regular basis. The latter is a

common situation for many radiopharmaceuticals. For radiopharmaceuticals with short half-lives or rare indications, no sizable commercial market exists. Consequently, no pharmaceutical company will be prepared to obtain a MA for a product that will not yield a profit due to these limitations. Still, there is a need from a medical point of view to have such products available. For radiopharmaceuticals incorporating radionuclides with a physical half-life of only a few minutes, only local production is feasible. They are therefore prepared in hospital pharmacies or laboratories and supplied for individual or small numbers of patients on a daily basis.

The extemporaneous preparation of radiopharmaceuticals is regulated on a national level, and hence this regulation may differ from country to country. The Pharmaceutical Inspection Convention (PIC/S) has drafted a guide to good practices for preparations of medicinal products in pharmacies [10], valid for medicinal products that do not have a MA, prepared extemporaneously or for stock. For medicinal products prepared to a larger extent or for use in clinical trials, industrial GMPs are applicable. Although the suggested guide outlines a general principle according to GMP, different requirements are particularly evident when it comes to documentation and quality control testing. There is also a discussion about the grades of background environment needed for production, with a differentiation between products with shelf lives less than or longer than 24 h [10]. While aseptic manufacturing according to industrial GMP has to be performed in grade A with a grade B background, this proposed guide opens for a relaxation to this. For an aseptic preparation of a product with a shelf life of less than 24 h, using a biohazard safety cabinet (BSC), the background environment may be grade D. Even for products with a shelf life longer than 24 h, an extensively documented procedure may allow grade C in background, as long as grade B clothing is worn. In general, the referred draft guide is based much upon a risk related approach and is graduated, depending on the size and type of prepared medicinal products.

As to documentation for extemporaneous prepared products, the proposed guide set as a minimum requirement to specify the name, strength, and expiry date of the product. If a product is prepared for a single patient (magistral production), it is assumed that no end product testing will be required. For radiopharmaceuticals, though, the activity in each dose must be measured before administration. Chemical and microbiological quality control is not required for products that have a shelf life of 24 h or less, provided that frequent process validation is performed. In addition, chemical and microbiological information must be available to justify the shelf life for the product.

For products that are prepared extemporaneously at a regular basis or even for a limited stock, a product specific documentation (product file) is needed. This will include specifications, instructions, and records but also a pharmaceutical assessment of safety data, toxicity, biopharmaceutical aspects, stability, and product design. The product file should also include a product review as soon as a product is used repeatedly or over longer periods.

Furthermore, the drafted guide suggests that the level of end-product testing for those products will depend on the associated risk connected to the scale of operation, shelf life of the product, frequency of preparation, as well as type of product (parenterals, orals) and type of facility where the product has been prepared.

Independent of which regulation applies at a national level to extemporaneous or magistral preparation of radiopharmaceuticals, the patients should be entitled to expect that these products are prepared accurately, are suitable for use, and will meet the expected standards for quality assurance. Pharmacists involved in this kind of production must ensure that they and any other staff involved are competent to undertake the tasks to be performed and that the requisite facilities and equipment are available [11]. As for other radiopharmaceutical production, systems must be in place to ensure the operator safety due to handling of radioactive materials. All involved staff must have sufficient training in radiation safety issues, in addition to training in GMP.

## REFERENCES


1. Britton, K. (1996), Radiopharmaceuticals for the future, *Curr. Dir. Radiopharma. Res. Dev*. (Ed. by Stephen Mather), viii. Developments in Nuclear medicine, Vol XXX. London, UK.
2. Saha, G. B. (1998), *Fundamentals of Nuclear Pharmacy*, 4th ed., Springer, Heidelberg, Germany.
3. Alexoff, D. L. Automation for the synthesis and application of PET radiopharmaceuticals, BNL-68614 Officinal File Copy.
4. Bremer, P. O. (1995), Aseptic production of radiopharmaceuticals, in *Aseptic Pharmaceutical Manufacturing*, Vol. II, *Application for the 1990s*, Interpharm, Michael J. Groves and Ram Murty, pp. 153–180.
5. Nordic Council on Medicines. (1989), *Radiopharmacy: Preparation and Control of Radiopharmaceuticals in Hospitals*, NLN Publications No. 26, Uppsala, Sweden.
6. European Commision (2003), *EU Guide to Good Manufacturing Practice*, Annex 1 and 3, Brussels, Belgium, October 8.
7. Dabbah, R. (1995), Controlled environments in the pharmaceutical and medical products industry: A global review from regulatory, compendial, and industrial perspectives, in *Aseptic Pharmaceutical Manufacturing,* Vol. II, *Application for the 1990s*, Interpharm, Michael J. Groves and Ram Murty, pp. 11–40.
8. Lee, M. C., PET and PET/CT are the fastest growing imaging modalities worldwide, paper presented at the 5th International Conference on Isotopes (5ICI), Brussels, Belgium, Apr. 25–29, 2005.
9. Zalutsky, M. R., Pozzi, O., and Vaidyanatha, G., Targeted radiotherapy with alpha particle emitting radionuclides, paper presented at the International Symposium on Trends in Radiopharmaceuticals (ISTR-2005), Vienna, Austria, Nov. 14–18, 2005.
10. Pharmaceutical Inspection Convention (2006, Aug.), PIC/S guide to good practices for preparation of medicinal products in pharmacies, PE 010-1 (Draft 2), Geneva, Switzerland.
11. Standards for good professional practice (2000), *Pharm. J*. 265(7109), 233.
12. Kowalsky, R. J., and Falen, S. W. (2004), *Radiopharmaceuticals in Nuclear Pharmacy and Nuclear Medicine*, American Pharmacists Association, Forrester Center, WV.


## FURTHER READINGS


European Commision. (2006), *EU Guide to Good Manufacturing Practice*, Annex 3; draft proposal, Brussels, Belgium, Apr. 12.

Rootwelt, K. (2005), *Nukleærmedisin*, 2nd ed. Gyldendal Norsk Forlag AS, Oslo, Norway.

Schwochau, K. (2000), *Technetium: Chemistry and Radiopharmaceutical Applications*, VCH Verlagsgesellschaft Mbh, Weinheim, Germany.

Welch, M. J., and Redvanly, C. S., Eds. (2002), *Handbook of Radiopharmaceuticals*, Wiley, Hoboken, NJ.

# STERILE PRODUCT MANUFACTURING

James Agalloco[1] and James Akers[2]

[1]*Agalloco & Associates, Belle Mead, New Jersey*
[2]*Akers Kennedy & Associates, Kansas City, Missouri*

## Contents

1 Introduction
2 Process Selection and Control
  2.1 Formulation and Compounding
  2.2 Primary Packaging
  2.3 Process Objectives
3 Facility Design
  3.1 Warehousing
  3.2 Preparation Area
  3.3 Compounding Area
  3.4 Aseptic Compound Area (If Present)
  3.5 Aseptic Filling Rooms and Aseptic Processing Area
  3.6 Capping and Crimping Sealing Areas
  3.7 Sterilizer Unload (Cooldown) Rooms
  3.8 Corridors
  3.9 Aseptic Storage Rooms
  3.10 Lyophilizer Loading and Unloading Rooms
  3.11 Air Locks and Pass-Throughs
  3.12 Gowning Rooms
  3.13 Terminal Sterilization Area
  3.14 Inspection, Labeling, and Packaging
4 Aseptic Processing Facility Alternatives
  4.1 Expandability
5 Utility Requirements
  5.1 Water for Injection
  5.2 Clean (Pure) Steam
  5.3 Process Gases
  5.4 Other Utilities

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

6 Sterilization and Depyrogenation
6.1 Steam Sterilization
6.2 Dry-Heat Sterilization and Depyrogenation
6.3 Gas and Vapor Sterilization
6.4 Radiation Sterilization
6.5 Sterilization by Filtration
7 Facility and System: Qualification and Validation
8 Environmental Control and Monitoring
8.1 Sanitization and Disinfection
8.2 Monitoring
9 Production Activities
9.1 Material and Component Entry
9.2 Cleaning and Preparation
9.3 Compounding
9.4 Filling
9.5 Stoppering and Crimping
9.6 Lyophilization
10 Personnel
11 Aseptic Processing Control and Evaluation
11.1 In-Process Testing
11.2 End-Product Testing
11.3 Process Simulations
12 Terminal Sterilization
13 Conclusion
Appendix
References
Additional Readings

## 1 INTRODUCTION

The manufacture of sterile products is universally acknowledged to be the most difficult of all pharmaceutical production activities to execute. When these products are manufactured using aseptic processing, poorly controlled processes can expose the patient to an unacceptable level of contamination. In rare instances contaminated products can lead to microbial infection resulting from products intended to hasten the patient's recovery. The production of sterile products requires fastidious design, operation, and maintenance of facilities and equipment. It also requires attention to detail in process development and validation to ensure success. This chapter will review the salient elements of sterile manufacturing necessary to provide acceptable levels of risk regarding sterility assurance.

Commensurate with the criticality associated with sterile products, the global regulatory community has established a substantial number of the basic requirements that firms are expected to adhere to in the manufacture of sterile products. The most extensive of these are those defined by the Food and Drug Administration

(FDA) in its 2004 Guideline on Sterile Drug Products Produced by Aseptic Processing and the European Agency for the Evaluation of Medicinal Products (EMEA) Annex 1 on Sterile Medicinal Products [1, 2]. Substantial additional information is available from the International Organization for Standardization (ISO), the Parenteral Drug Association (PDA), and the International Society for Pharmaceutical Engineering (ISPE) (see Appendix) [3]. The organizations have provided a level of practical, experience-based detail not found in the regulatory documents, thereby better defining practices that are both compliant with regulatory expectations and based upon rational, evidence-based science and engineering.

Consideration of patient risk associated with pharmaceutical production emerged largely from regulatory impetus, by which the regulatory community stated its intended goal to structure its inspectional process using patient safety as a major focus in determining where to allocate their inspectional and review resources. Emanating from the International Conference on Harmonization (ICH) efforts to produce a harmonized approach to pharmaceutical regulation, risk-based compliance has been adopted in Europe, Japan, and the United States [4, 5]. Sterile products, especially those made by aseptic processing, have been properly identified as a high priority by the global regulatory community. Several risk analysis approaches have been developed that can help the practitioner review practices with the goal of minimizing risk to the patient [6–8].

## 2 PROCESS SELECTION AND DESIGN

The production of sterile products is profoundly impacted both by formulation and the selection of primary packaging components. Design parameters for a facility and selection of appropriate manufacturing technologies for the product require that the formulation process and packaging components be chosen and evaluated in advance.

### 2.1 Formulation and Compounding

The vast majority of parenteral formulations are solutions requiring a variety of tankage, piping, and ancillary equipment for liquid mixing (or powder blending), filtration, transfer, and related activities. Suspensions, ointments, and other similar products, including the preparation of the solutions for lyophilized products, can be manufactured in the same or very similar equipment. The scale of manufacturing can vary substantially, with the largest batches being well in excess of 5000 L (typically for large-volume parenteral production), down to less than 50 mL for radiopharmaceuticals or biologicals customized for a particular patient.

The majority of this equipment is composed of 300 series austenitic stainless steel, with tantalum or glass-lined vessels employed for preparation of formulations sensitive to iron and other metal ions. The vessels can be equipped with external jackets for heating and/or cooling and various types of agitators, depending upon the mixing requirements of the individual formulation. In many facilities, a variety of tank sizes are available for use. Larger facilities may have the high-capacity tanks permanently installed and permanently connected to process utilities. Smaller vessels are generally mobile and positioned in individual processing booths or rooms as needed.

After sterilizing filtration (or sterilization by heat or other means), comparably sized vessels are sometimes utilized to contain the product prior to and during the filling process. These holding vessels are often steam sterilized along with the connecting piping prior to use. There are a number of firms that fill directly from the compounding vessel using in-line filtration eliminating the intermediate vessel. When this approach is used, a small moist-heat-sterilized surge tank or reservoir tank may be required, particularly with modern time–pressure filling systems. This practice may reduce initial facility and equipment cost but places additional constraints on operational flexibility. The use of disposable equipment for compounding and holding of sterile formulations is coming into greater use. This eliminates the cleaning of vessels prior to reuse, but confirmation of material compatibility is required. Disposable equipment is often used with products manufactured in small to moderate volumes, and while reducing initial equipment expenses disposable equipment also results in contaminated waste, which cannot be recycled or reused and must be treated appropriately.

Aseptic compounding as required for suspensions and other formulations in which open-vessel processes are required mandate an ISO 5 environment providing ideally >400 air changes/hour in which these steps can be performed with minimal opportunity for adventitious contamination. This could be accomplished using a protective curtain and a unidirectional flow hood (UFH) or other more evolved designs such as a restricted access barrier (RABs) system or an isolator (technologies that provide a higher level of employee separation from the area in which materials are handled can get by with lower air exchange rates). All activities requiring opening of processing lines such as sampling or filter integrity testing should be performed using similar protective measures. The preparation of sterile suspensions requires a facility/equipment design capable of safe addition of sterile solids to a liquid vehicle and is conventionally performed using a specifically designed processing area to minimize contamination potential. Comparable and greater complexity is generally required for creams, ointments, emulsions, and the increasingly common liposome formulations.

Some sterile powder formulations (these are predominantly, but not exclusively, antibiotics) may require sampling, mixing, milling, and subdivision activities similar to those found in oral powder manufacturing. The facilities and equipment utilized for these products is substantially different from that used for liquids, and the production area bears little resemblance to that utilized for liquids. These materials are received sterile and must be processed through sterilized equipment specifically intended for powder handling in a fully aseptic environment with ISO 5 protection over all open container activities.

## 2.2 Primary Packaging

The primary package for parenteral formulations provides protection to the sterile materials throughout the shelf life. The components of the primary package are every bit as important to contamination control and hence safety of the finished product as the formulation itself, and their preparation must be given a comparable level of consideration. The most commonly used container is glass; vials are still the most common, although increasingly prefilled syringes are chosen. Glass ampoules are still seen. However, although convenient from a manufacturing perspective, the

difficulty involved in opening ampoules while at the same time avoiding problems with glass particulate or microbial contamination has reduced their popularity. The use of plastic containers (as vials, ampoules, or syringes) is increasingly common given their reduced weight and resistance to breakage. Blow-fill seal (BFS) and form-fill seal (FFS) are utilized for the filling of numerous ophthalmic and other noninjectable formulations in predominantly low-density polyethylene (LDPE) containers. With the exception of ampoules and BFS/FFS, an elastomeric closure system is also necessary to seal the containers. Some delivery systems (i.e., prefilled syringes, multichamber vials, and others may require more than one elastomeric component to operate properly. In the case of vials, an aluminum crimp is applied to secure the closure to the vial. Prefilled syringes may require the preparation and assembly of additional components such as needles, needle guards, stoppers, diaphragms, or plungers, depending on the specifics of the design. Lyophilization is required to ensure the stability of some formulations and requires the use of closures that allow venting of the container during the freeze-drying process. Full seating of the closure is accomplished within the lyophilizer using moving shelves to seat the closure.

Glass is ordinarily washed prior to sterilization/depyrogenation to reduce contamination with foreign material prior to filling. In aseptic fill processes, the glass is then depyrogenated using dry heat. This can be accomplished using either a continuous tunnel (common for larger volumes and high-speed lines) or a dry heat oven (predominantly for small batches). The depyrogenation process serves to sterilize the glass at the same time, and thus the glass components must be protected post-processing. This is generally accomplished by short-term storage in an ISO 5 environment often accompanied by covering within a lidded tray. There are suppliers that offer depyrogenated glass vials and partially assembled syringes in sealed packages for filling at a customer's site. In this instance, the supplier assumes responsibility for the preparation, depyrogenation, and aseptic packaging. Glass ampoules are available presealed and depyrogenated; the end user has merely to open, fill, and reseal the syringe under appropriate conditions.

Plastic components (whether container or closure) can be sterilized using steam, ethylene oxide, hydrogen peroxide, or ionizing radiation. The $\gamma$ irradiation is accomplished off-site by a subcontractor with appropriate expertise as these methods are considered the province of specialists because of the extreme health hazards directly related to the sterilization method. Electron beam sterilization may also be done by a contractor, although compact lower energy electron beam systems have been introduced that allow sterilization in-house. Steam sterilization is ordinarily performed in house, though many common components are becoming available presterilized by the supplier. Preparation steps prior to sterilization vary with the component and the methods used to produce the component. Rubber components are washed to reduce particles, while this is less common with plastic materials.

Syringes vary substantially in design details and can be aseptically assembled from individual components. However, increasingly, these are supplied as presterilized partial assemblies in sealed containers.

The BFS and FFS are unique systems in that the final container is formed as a sterile container just prior to the aseptic filling step. The BFS requires careful control over the endotoxin content of the LDPE (and other polymeric materials) beads used to create the containers as well as the melting conditions utilized to form them.

The FFS utilizes in-line sterilization/drying of the film prior to shaping of the containers.

### 2.3 Process Objectives

The production of parenteral products requires near absolute control over microorganisms. Endotoxin contamination is a serious health concern, particularly among neonates and infants and also requires a high level of control and validation. Additionally, the control of foreign matter, including particles and fibers of various types, is also vitally important to end-user safety. Assuring appropriate control over these potential contaminants requires careful attention to several factors: facility design, equipment selection, sterilization procedures, cleaning regimens, management of personnel, and the process details associated with compounding, filling, and sealing of product containers. Each of these will be discussed in detail.

## 3 FACILITY DESIGN

To provide control of microbial, pyrogen, and particles controls over the production environment are essential. The facility concerns encompass the entire building, but the most relevant components are those in which production materials are exposed to the environment.

### 3.1 Warehousing

Environmental protection of materials commences upon receipt where samples for release are taken from the bulk containers. Protection of the bulk materials is accomplished by the use of ISO 7 classified environments for sampling. All samples should be taken aseptically, which mandates unidirectional airflow and full operator gowning. This practice is mandated by current good manufacturing practice (CGMP) and assures that sampling does not introduce contaminants to the materials that will be used in the production. Where central weighing/subdivision of active ingredients and excipients are performed, similar protection is provided for identical reasons. The expectation is that these measures reduce the potential for contamination ingress into materials that have yet to receive any processing at the site. Materials and components that are supplied sterile are received in this area, but samples are often packaged separately by the supplier to eliminate the need for potentially invasive sampling of the bulk containers. Where so-called delivery samples are used, it is critical that these samples are known to be fully representative of the production process. Additionally, where sterility or bioburden control of sampled materials is critical, thought must be given to the methods used to reseal the containers to ensure that moisture levels, bioburden levels, or in the case of sterile products sterility assurance are not compromised.

### 3.2 Preparation Area

The materials utilized for production of sterile processes move toward the filling area through a series of progressively cleaner environments. Typically, the first step

is transfer into an ISO 8 [Class 100,000, European Union (EU) Grade D] environment in which the presterilization preparation steps are performed. Wooden pallets and corrugated materials should always be excluded from this zone (and any classified environment), and transfers of materials are performed in air locks designed to reduce the potential for particle ingress and to a lesser extent microbial ingress. Preparation areas provide protection to materials and components for a variety of activities: component washing (glass, rubber, and other package components), cleaning of equipment (product contact fill parts, process tools, etc.), and preassembly/wrapping for sterilization. In some facilities, this area is also utilized to support compounding operations in which case process utensils, small containers, and even portable equipment will be cleaned and prepared for sterilization.

Careful attention must be given to material flow patterns for clean and dirty equipment to prevent cross contamination. In larger facilities, the equipment wash room may be a separate room proximate to the preparations area with defined flows for materials and personnel. Ideally, materials should move through the facility in a unidirectional fashion, with no cross over of any kind.

The preparations area typically includes storage areas where clean and wrapped change parts, components, and vessels can be held until required for use in the fill or compounding areas. (Just-in-time practices are desirable for all parenteral operations to avoid extensive and extended storage of materials in the higher classified fill or compounding areas.) The preparations area is ordinarily located between the warehouse and the filling/compounding areas and connected to each of those by material/equipment air locks.

Preparation areas are supplied with high-efficiency particulate air (HEPA) filters (remote-mounted HEPAs are commonplace). The common design requirement is more than 20 air changes per hour, turbulent airflow (see below), and temperature and relative humidity controlled for personnel comfort. As in any clean room area designed for total particulate control, the air returns should be low mounted. Wall and ceiling surfaces should be smooth, easily cleaned, and tolerant of localized high humidity. Floors should be typically monolithic with integral drains to prevent standing water. Common utilities are water for injection, deionized water, compressed air, and clean/plant steam. Clean-in-place (CIP) and sterilize-in-place (SIP) connections may be present if the prep area supports compounding as well.

Ordinarily, present within the preparation area are localized areas of ISO 5 unidirectional airflow (Class 100) utilized to protect washed components prior to sterilization and/or depyrogenation. These areas are not aseptic and should not be subjected to the more rigorous microbial expectations of aseptic processing. They are designed to reduce/eliminate the potential for particle contamination of unwrapped washed materials. Operators accessing these protective zones wear gloves at all times when handling materials.

Operators in the preparations area are typically garbed in low particle uniforms (or suits) with shoe, hair, and beard covers. The use of latex or other gloves is required when contacting washed components. Sterilized gowns and three-stage gowning facilities are not required to enter or work in this ISO 8 environment. Gowns are generally donned within a single-stage airlock, which is maintained at a pressure slightly negative to the ISO 8 working environment. Separate personnel entry/exit are not typically necessary for this lower classified environment.

Equipment within the preparations area varies with the practices of the firm and can include manual or ultrasonic wash/rinse sinks; single or double door automated parts washers; batch or continuous glass washers; stopper washers for closure components; CIP/SIP stations; equipment wrap areas (as described above); and staging areas for incoming (prewash) components, dirty equipment, and cleaned components/equipment. An adjacent classified storage area(s) may be present in larger facilities to accommodate the full variety of change parts and equipment that is not in immediate use. Where the preparations area also supports compounding, it may include additional equipment such as pH meters, filter integrity apparatus, and the like in support of those operations. (*Note:* Where compounding requires aseptic conditions for rigorous control of bioburden, as is the case for unpreserved biologics and other contamination-sensitive products, it is best to provide separate entry for compounding. The moisture level and hence contamination potential in a typical preparation area is unsuitable for entry into an aseptic compounding area).

Depending on the scale of the operation, the preparations area may include the loading areas for both sterilizers and ovens. In high-throughput operations where the use of tunnels for glass depyrogenation is more prevalent, glass washers and tunnels for each filling line may be in separate ISO 8 rooms accessed from the preparations area.

### 3.3 Compounding Area

The manufacture of parenteral solutions is ordinarily performed in ISO 7 (Class 10,000, EU Grade C) controlled environments in which localized ISO 5 unidirectional flow hoods are utilized to provide greater environmental control during material addition. These areas are designed to minimize the microbial, pyrogen, and particle contributions to the formulation prior to sterilization. Depending upon the scale of manufacture, this can range from small containers (up to 200 L) (disposable containers are coming into use for these applications), to portable tanks (up to 600 L) to large fixed vessel (10,000 L or more have been used) in which the ingredients are formulated using mixing, heating, cooling, or other unit operations. Smaller vessels are placed or rolled onto scales, while fixed vessels are ordinarily mounted on weigh cells. The vessels may be equipped for temperature and pressure measurement instruments, as mandated by process requirements. Compounding areas often include equipment for measuring mass and volume of liquid and solid materials including, for example, graduated cylinders, and scales of various ranges, transfer and metering pumps, homogenizers, prefilters, and a variety of other liquid/powder handling equipment. Liquid handling may be accomplished by single-use flexible hose, assemblies of sanitary fittings, or some combination thereof. A range of smaller vessels to be used for the addition of formulation subcomponents or excipients to the primary compounding tank may be required as well. Because parenteral formulations can include aqueous and nonaqueous vehicles, suspensions, emulsions, and other liquids, the capabilities of the compounding area may vary. Agitators can be propeller, turbine, high shear, or anchor designs depending upon the requirements of the products being manufactured, and it is not uncommon to find examples of each in larger facilities. It is preferable to perform as much of the process as possible while the formulated liquid is nonsterile to ease sterilization requirements, although precautions to prevent microbial and endotoxin contamination are important risk abatement features.

The formulation area is customarily a combination of open floor space, adjacent to three-sided booths and individual processing rooms in which the ingredients are handled and individual batches are produced. Walls and ceiling materials are selected to be impervious to liquids and chemical spills and are easy to clean. Floors in these areas are monolithic and should be sloped (at 1–3:100) to drains with appropriate design elements and control procedures to eliminate backflow potential (regulatory bans on drains in classified areas are focused on protecting aseptic environments and are inappropriate for nonsterile compounding areas). Pit scales should be avoided in new installations; floor-mounted scales intended for cleaning underneath the base are preferable.

Compounding areas are supplied with HEPA filters (ceiling-mounted terminal HEPAs are more common, though central supply is possible in areas of low contamination risk). The common design requirement is more than 50–60 air changes per hour, turbulent airflow (see below), with temperature and relative humidity for personnel comfort. Air returns may be at or near floor level, with localized extraction provided as necessary to minimize dusting of powder materials. Where substantial heat is generated from processing or sterilization, a ceiling or high wall return may be more appropriate. Wall and ceiling surfaces are smooth, easy to clean, and tolerant of localized high humidity. Floors are typically monolithic with integral drains to prevent standing water. Common utilities are water for injection, deionized water, nitrogen, compressed air, clean/plant steam, and heating and cooling media for the fixed and portable tanks. Water for injection use points are often equipped with sanitizable heat exchangers for operator safety.

Cleaning of the fixed vessels and portable tanks is accomplished using either manual sequenced cleaning procedures or more commonly with a CIP system. Cleaning of other items can be accomplished in a wash area accessed from the compounding area or in a common wash room incorporating both filling and compounding equipment. Sterilization of the nonsterile processing equipment and vessel is often provided for as an option, even where it is not routinely required to control product bioburden. Where production volumes or physical location dictate, the compounding area may have a separate preparations area from that utilized to support filling operations.

Personnel working in the compounding area typically wear a coverall (which may be sterilized for contamination control as required), with head/beard covers, as well as dust masks and sterile gloves. Additional personnel protective equipment may be necessary for some of the materials being processed. A fresh gown should be donned upon each entry into the compounding area. Separate gowning/degowning rooms should be provided to minimize cross-contamination potential for personnel working with different materials. As nonsterile compounding areas are often ISO 6–7 environments but are not aseptic, the more rigorous contamination controlling designs required of aseptic gowning areas (see below) are somewhat reduced.

### 3.4 Aseptic Compounding Area (If Present)

Where products are filled using in-line filtration direct to the filling machine, an aseptic compounding area may not be present. In those instances the final sterilizing filter will be located in the fill room.

Products that are held/processed in sterilized vessels prior to filling require an aseptic compounding area. This is typically an ISO 7 in environment with localized

ISO 5 unidirectional flow present where open-product containers or aseptic operations are conducted. Some products may require larger ISO 5 suites with full HEPA coverage rather than the more common ISO 5 clean booth design. Fixed vessels in this area are cleaned and sterilized in situ, while portable vessels are typically relocated to the wash area for cleaning. Sterilization of portable vessels may be accomplished at an SIP station in the aseptic core, compounding, or preparations areas. When accomplished outside the aseptic processing area, resterilization of the connecting lines may be appropriate. Filters for sterilization of solutions from compounding to holding vessels are typically located in this environment as well, with sterilization by either SIP or sterilization in an autoclave. The use of integrated, programmable logic controlled (PLC) filter skids with automatic CIP/SIP and filter integrity testing is frequently seen for contamination sensitive products.

Depending upon the formulations being produced, additional sterilized processing equipment may be present in this area for use in the process. This can include in-line homogenizers, static mixers, and colloid mills. Where sterile powders are produced, the aseptic compounding processes can include blending, milling, and subdivision equipment.

Aseptic compounding areas typically require a means to introduce sterile equipment, tubing, and other items, so access to a sterilizer is desirable. The aseptic compounding area may be contiguous to the aseptic filling suites. If it is not, separate gowning areas must be provided for personnel as well as separate air locks/pass-throughs (see below).

Personnel working in aseptic compounding wear full aseptic garb: sterile gown, hood, face mask, goggles, foot covers, and gloves. Adaptations may be necessary for potent/toxic compounds to assure operators are properly protected from hazardous materials. Gowning areas are ordinarily shared with aseptic filling, but where they are not shared a comparable design, albeit on a smaller scale, is appropriate.

The facility design features match that of the aseptic filling room/aseptic processing areas described in greater detail below. Utility services would mimic those utilized in the nonsterile compounding area that is usually adjacent (next to or above) to the aseptic compounding area. Temperature and humidity should be controlled to similar levels as those required for aseptic filling. Since CIP/SIP systems tend to generate heat and humidity, sufficient capacity must be available to control temperatures to approximately 18–20°C and <50% relative humidity (RH).

### 3.5 Aseptic Filling Rooms and Aseptic Processing Area*

The filling of aseptic formulations (and many terminally sterilized products as well, by reason of their lesser number) is performed in an ISO 5 (Class 100) environment, which is accessed from an ISO 6/7 background environment in which personnel are present. Some measure of physical separation is provided between the ISO 5 and ISO 6/7 environments as a means of environmental protection as well as a reminder to personnel to restrict their exposure to ISO 5.

*This section describes the conventional manned clean room; a later section in this chapter will address alternative aseptic processing environmental control designs with somewhat different features and control measures.

In large operations an aseptic filling room is generally one of a multiple suite of aseptic rooms which allow simultaneous production of multiple products. The filling rooms are independent of each other; however, sharing the supporting rooms is common. Sterilizer unload rooms, corridors, air locks, storage rooms, lyophilizer loading rooms, and gowning rooms (each will be briefly described as well) may all be present, and their arrangement must suit production volumes. Where shared common areas are required, the design should feature unidirectional materials flow to prevent cross-contamination and to minimize the potential for mix-ups. In the smallest facilities, only the gowning area might be separate from the fill room, and all of the supportive activities could be inclusive in a single room (however, unloading activities should not occur during filling operations). All of these aseptic processing areas (APAs) are built to the same design standards: smooth, impervious ceilings, walls and floors, flush-mounted windows, clean room door designs, coved corners, finishes capable of withstanding the aggressive chemicals utilized for cleaning and sanitization. Air returns throughout the APA are located at or near floor level. Unidirectional airflow is provided over all exposed sterile materials, that is, fill zone, sterilizer/oven/tunnel unload areas, and anywhere else sterile materials are exposed to the environment. Air changes in these ISO 5 environments can approach 600 per hour, though lesser values have proven successful. Air changes in the background environment vary from 60 to 120 per hour.

The glass container fill rooms filling machines are connected to depyrogenating tunnels and exit ports leading to capping stations. Batch handling of glass is discouraged unless isolator systems are employed. In some operations, the in-feed and discharge of containers/components may utilize trays, tubs, or bag systems for material feed/discharge. Wherever possible, automation of component feeding should be considered to reduce contamination risk. Supportive equipment present might include carts, weigh stations, stoppering, crimping, sealing, and other fill system related machinery depending upon requirements.

The product contact surfaces in this environment are typically removed for cleaning; however, in some installations, the sterilization, transfer, and reinstallation of the component feed hoppers present such difficulty that these systems are decontaminated in situ with a sporicidal agent, rather than removed after each use. These units should still be removed for cleaning and sterilization on a validated periodic basis to prevent the buildup of residues that might impact their in-situ decontamination or create particle control problems. All other product contact surfaces should be sterilized prior to each use. Nonsterilized items should not be allowed to enter the ISO 5 portion of the fill zone, and sanitization is essential for all nonproduct surfaces in the fill zone, as well as the surrounding background environment.

Discharge of sealed containers can be accomplished via a exit port or "mouse hole" that allows for the passage of the containers from the APA to the surrounding environment. Proper design of the mouse hole system ensures protection of the classified fill area from contamination flowing against the flow of the containers. In many instances the discharge is into a nonclassified inspection area that may lead directly to the secondary labeling/packaging area.

Personnel working in aseptic compounding wear full aseptic garb: sterile gown, hood, face mask, goggles, foot covers, and gloves. Adaptations may be necessary for potent/toxic compounds to assure operators are properly protected from hazardous materials.

### 3.6 Capping and Crimp Sealing Areas

The application of aluminum seals over rubber stoppers is essential to secure them properly. In many older facilities this was accomplished outside the aseptic processing area in an unclassified environment. Current practice requires that air supplied to this activity meet ISO 5 under static conditions. The protection of crimping has resulted in a variety of designs to meet the requirement: Sterile crimps can be applied with the aseptic core on the filling line; sterile crimps can be applied in a separate crimping room accessible from the filling room. If the crimpling operation is located within the APA, it should be in a separate room maintained at a negative pressure differential relative to the filling environment. Crimping may alternatively be performed in a classified room accessed from a controlled but unclassified environment. In this case it is imperative to verify that the environmental controls satisfy regulatory expectations for all relevant markets.

### 3.7 Sterilizer Unload (Cooldown) Rooms

Sterilizers/ovens are unloaded and items staged prior to transfer to the individual fill rooms. ISO 5 air is provided over the discharge area of ovens (and autoclaves if items are sterilized unwrapped) to provide protection until the items are ready for transfer. The heat loads in this room may be such that special high-temperature sprinkler heads may be necessary to avoid unintentional discharge when unloading hot materials. This room may not be separate from the corridor used to connect the fill rooms. It is ordinarily adjacent to any aseptic storage area.

### 3.8 Corridors

Corridors serve to interconnect the various rooms that comprise the APA. Fill rooms, air locks, and gowning rooms are accessed from the corridor. They can also be utilized for modest storage as well.

### 3.9 Aseptic Storage Rooms

In general, extensive use of in-process storage areas should be avoided. It is best to operate the aseptic facility in a just-in-time mode in which components and equipment are sterilized shortly before they are required for use in the filling or compounding areas. Some limited storage is necessary for nonproduct contact materials such as sanitizing agents, environmental supplies and equipment, and other items.

### 3.10 Lyophilizer Loading and Unloading Rooms

The loading of lyophilizers is accomplished under ISO 5 environmental conditions within the aseptic processing area. Several possible locations are possible: within the aseptic fill room itself, in a separate room adjacent to the fill room, or in a separate room remote from the fill room. There are pros and cons with each of these selections which should be carefully considered in the facility design. There is a

universal expectation that filled containers of product should be maintained under ISO 5 conditions during transfer and lyophilizer loading. Many modern facilities incorporate automatic lyophilizer loading and unloading. Automation of loading, unloading, and in the case of vials transfer to the crimping station greatly reduces contamination risk and is highly recommended.

If manual transfer is unavoidable, location of the lyophilizer relatively close to the filling line enables protected transfer to be accomplished rather easily. Remote locations may require transfer of product in carts capable of providing ISO 5 quality air. These carts will generally require battery power in order to run the necessary air blowers and control systems. Alternatively, product trays could be placed in airtight carriers; this activity and the sealing of the carriers would have to be accomplished under ISO 5 conditions. Locating the lyophilizer in the fill room may restrict the ability to unload the dryer while the filling line is in use, particularly if the lyophilizer is loaded and unloaded manually, which would increase the clean room personnel load and potentially increase contamination risk.

The use of trays during lyophilization is less common, nevertheless, ring trays with removable bottoms are sometimes used to transfer vials to/from the lyophilizer. Where trays are used, they must be cleaned and sterilized prior to each batch. Large lyophilization facilities will sometimes use an automated loading/unloading system in which all shelves or a shelf at a time are processed. Regardless of the practice, ISO 5 conditions are required for all areas of the facility in which partially stoppered containers are transferred or handled. As previously mentioned, it may be possible in some operations to transfer containers in a manner that they are not exposed to the environment during transfer.

Upon completion of the drying process, the containers will ordinarily have their stoppers fully seated on the container within the freeze dryer. The stoppered containers are then passed through a sealing station in which aluminum crimps are applied. This may be accomplished on the fill line, or using a separate crimping machine. Precautions will need to be taken to ensure that only fully stoppered vials are transferred to the crimping station. This can be accomplished by automatic inspection systems of various designs. It is increasingly common for product transfer to crimping and crimping itself to be done under unidirectional airflow. It should be noted that a crimpling station will generally not meet ISO 5 particulate air quality requirements when the crimper is operating since the generation of relatively high levels of particulate is an inherent feature of this process.

### 3.11 Air Locks and Pass-Throughs

Air locks serve as transition points between one environment and another. Ordinarily, they are designed to separate environments of different classification: that is, ISO 6 from ISO 7. When this is the case, they are designed to achieve the higher of the two air quality levels in operation. If they are utilized for decontamination purposes for materials/equipment that cannot be sterilized, but must be introduced into the higher air quality environment, they may be fitted with ultraviolet (UV) lights, spray systems, vapor phase hydrogen peroxide generators, or other devices that may be effectively utilized for decontamination of materials. Regardless of the design or the decontamination method employed, the process should be validated to ensure

consistent efficacy. The doors at each end can be automatically interlocked or managed by standard operating procedure. In some instances a demarcation line is used to delineate the extent to which individuals from one side should access the air lock. It is good practice to carefully control and to minimize the time that any operator spends accessing an air lock, therefore transfer of materials should be carefully planned to minimize frequent and spontaneous access. Additionally, the capacity of the air lock should be carefully considered relative to the actual production requirements. Air locks that lack sufficient capacity and that cannot provide sufficient air exchange will be less suited to the control of contamination into more critical areas of the aseptic processing environment.

A smaller scale system with comparable capabilities is the pass-through. This differs from the air lock primarily in dimension, as items are typically placed into the pass-through by personnel, whereas the air lock is customary for pallet, portable tanks, and larger items that are either rolled or mechanically lifted into position. The operation of the pass-through can be either manual or automatic with similar capabilities to that of the air lock described above. In general pass-throughs should be supplied with HEPA filters and should be designed to meet the air quality level of the higher air quality classification room served. Pass-throughs should also be interlocked and provide adequate facilities for decontamination of materials being transferred.

Air locks and pass-throughs are bidirectional and can be used for movement in either direction. When used as an exit route, the decontamination procedure can be omitted. Where production volumes warrant separate entry and exit, air locks may be necessary to maintain both adequate capacity and separation between clean and used items. In an emergency, airlocks can serve as emergency exits for personnel, in which case the interlocks can be overridden.

### 3.12 Gowning Rooms

The gowning area used for personnel entry/exit presents some unique problems. Gowning facilities must be designed to the standards of the aseptic processing area, yet personnel upon entry are certainly not gowned. Because ungowned staff will release higher concentrations of contaminants into the environment, gown rooms must be designed with sufficient air exchange so that this contamination is effectively and promptly removed. In general, the contamination load within a gowning environment will require air exchange rates at the high end of recommended levels for a given ISO 14644 air quality classification. Gowning areas are separated into well-defined zones where personnel can progress through the various stages of the gowning process.

The most common approach in industry is a three-stage gowning area design in which three linked rooms with increasing air quality levels are utilized to efficiently and safely affect clothing change. Staff should enter the first state of the gowning room wearing plant uniforms. No articles of outerwear worn outside the facility should be worn to the gowning area. Therefore, a pregowning room equipped with lockers is required so that operators can change into dedicated plant clothing prior to moving to the gowning area. Generally, the pregowning locker area is not classified, although entry is controlled and temperature and humidity are maintained at 20–24°C and 50% ± 10%. The pregown area should have extensive hand-washing

facilities equipped with antibacterial soap, warm water, and brushes for cleaning finger nails. Soap and water dispensing should be automatic and hands should be air rather than towel dried. The pregown area should have typical clean room wall and floor finishes along for frequent and rigorous cleaning and sanitization. The pregown area is bidirectional as it is used as both an entry and exit point. Separate pregown areas are required for female and male personnel. A typical complement of garments for exit of the pregown area includes surgical scrubs or other nonparticulate shedding plant uniform. Ideally, the uniform should have a high neck and sleeves which extend to the lower wrist. Hair covers and beard covers are donned in the pregown area.

Upon entry into the first-stage gowning room, which is generally designed to an ISO 7 air quality level, the operators often don a second hair cover, sterilized gloves, and a sterilized surgical mask. In the second and third stages of the gowning area room classification is typically ISO 6 or ISO 6 followed by ISO 5 at the exit point. Different firms have different gowning sequences. However, in every case the flow of personnel and arrangement of gowning materials should be such that personnel flow is in one direction. In the last of the three gowning stages, secondary protective equipment can be donned, including sleeve covers and a second set of gloves. Some firms will use tape to secure the gloves to the sleeves to prevent separation. A dry glove decontamination point utilizing disinfectant foam is generally provided prior to exiting the gowning area; this should be a hands-free operation. In some facilities air showers, which provide a high-intensity blast of HEPA air for a predetermined length of time, are employed after gowning is completed. Side-by-side gowning of personnel should be avoided to preclude adventitious contamination. Similarly, personnel exiting the aseptic area should use a separate degowning area. These design practices are appropriate in all but the very smallest facilities where only a single aseptic operator is present.

### 3.13 Terminal Sterilization Area

The terminal sterilization of finished product containers may be performed in the same sterilizers utilized to supply the aseptic processing operations. The differing process needs of terminal sterilization will sometimes dictate the use of sterilizers specifically designed for terminal sterilization incorporating air-over pressure systems, internal fans, and spray cooling. Where this is the case, the terminal sterilizer is located proximate to the crimping/sealing areas. A double-door sterilizer design is preferred with staging areas for filled containers to be sterilized and a separate area for containers that have completed the process. Classification of these areas is not required as the containers are closed throughout the sterilization process. The flooring materials in this area should be monolithic to allow for easy cleanup in the event of container breakage.

### 3.14 Inspection, Labeling, and Packaging

These activities are performed on finished product containers in unclassified environments. The primary design requirements are straightforward: separation of products to prevent mix-up, adequate lighting for the processes, and control over labeling materials.

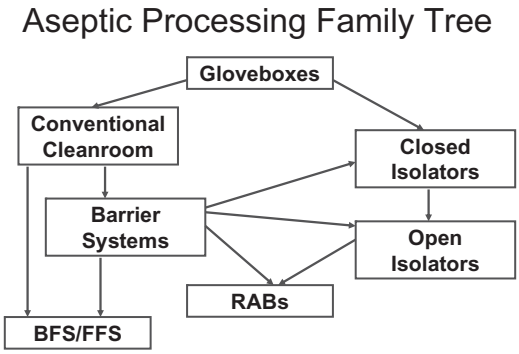


**FIGURE 1** Aseptic processing family tree.

## 4 ASEPTIC PROCESSING FACILITY ALTERNATIVES

The successful production of parenteral drugs by aseptic processing requires an environment in which microorganisms and particles are very well controlled. The means to accomplish this has undergone substantial change over the last 50 years (see Figure 1) with continuing refinement. The earliest aseptic processing systems used glove boxes with minimal (if any) airflow and manual disinfection in which manual processes were performed. The availability of HEPA filters in the late 1950s led to human-scale clean rooms in which processing equipment could be installed. Aseptic processing changed radically once entire clean rooms became feasible.

As it had always been recognized that personnel were the dominant source of contamination, the majority of designs utilized some measure of physical separation between the operator and the critical zone (sterile field) in which the aseptic processing activities were performed. Separative devices (a term that is now embodied in ISO 14644-7 Separative Enclosures) of different design and varying capability have been successfully employed including flexible curtains and fixed plastic shields with or without integrated gloves/sleeves [9]. In the most evolved designs operation of the equipment is interlocked with the surrounding enclosure, such that equipment stops running when the doors are opened. These latter designs represented the pinnacle of clean room-based aseptic processing into the early 1990s.

Isolators represent a return to operator separation principles utilized during the glove box era, albeit with substantial improvements in the form of rapid transfer ports for material transfer, air-handling systems utilizing modern HEPA filters, and reliable decontamination systems. The salient element of all isolator designs is the completeness of separation between the internal and external environments. This single feature affords vastly superior performance relative to manned clean rooms in excluding personnel-derived contamination and has comparable advantages for the containment of potent compounds. While initial adoption of the technology was slowed by the novelty that isolators presented to users, much of the initial reluctance has been overcome [10, 11]. Isolators for aseptic processing vary in complexity, size, and amount of processing equipment. They can be utilized for processing ranging

from manual compounding of small batches to high-speed filling of final product containers. Depending upon the process requirements, isolators can be utilized for containment of potent compounds (under negative pressure while still nonsterile) during the compounding, aseptic operation (under positive pressure) for preparation and transfer of components and aseptic containment (also under positive pressure) for aseptic filling of the potent drug solution.

Firms that were intimidated by or unconvinced of the superiority of isolators developed the restricted-access barrier (RAB) system as a potentially less complex and less costly alternative [12]. The real-world utility of RABs systems is unknown; there are still relatively few installations; thus, the experience base is still emerging. Also unconfirmed at this point are the actual validation and ongoing process control requirements which make direct comparison of project time lines and overall costs with isolators somewhat speculative.

There are specialized technologies such as BFS and FFS that are appropriate for aseptic processing, but these are restricted to filling processes only. A number of other new technologies are being developed for use in aseptic processing, including vial isolators and closed vial filling [13–15]. All of these have the objective of reducing contamination through reduction in human involvement or increased protection of the container. Further advances in processing including gloveless isolator designs, robotics, and others are already under active development to further improve the safety of parenteral products.

### 4.1 Expandability

Large facilities often include design elements that facilitate later expansion of the facility to add additional capacity. The most common of these is extension of an aseptic corridor to additional filling suites; reservation of space for additional sterilizers; and allocating space for additional or oversizing initial utility systems. Obviously, these types of changes require careful design and must be properly managed during execution to avoid impact on existing operations.

Isolation technology changes this dynamic significantly by eliminating most of the disruption on current activities, as fabrication of the isolator occurs off-site, and installation can be minimally disruptive compared to what is required with a cleanroom design. Isolators are generally installed in ISO 8 space; therefore, it is possible to build a rather large ISO 8 facility in which equipment can be moved, replaced, or reconfigured quite easily compared to conventional human-scale zoned aseptic processing areas.

## 5 UTILITY REQUIREMENTS

Any utility in direct product contact is subject to formal qualification through confirmation of the quality of the delivered material at each use point. Water-for-injection (WFI) systems are considered the most critical of all, and the qualification period for WFI is the longest and may be as long as 3 months. The remaining product contact utilities can be qualified more rapidly. Nonproduct utilities requirements can be satisfied by commissioning.

### 5.1 Water for Injection

The most important utility in sterile manufacturing is WFI. Not only is it a major component in many formulations, it is also utilized as a final rinse of process equipment, product contact parts, utensils, and components. In some facilities it may be the only grade of water available and is used for initial cleaning of items as well. The WFI may be produced by either distillation (multiple effect or vapor compression) or reverse osmosis (generally in conjunction with deionization) and is ordinarily stored and recirculated at an elevated temperature greater than 70°C to prevent microbial growth [16, 17]. Where cold water is required, it may be supplied by use point heat exchangers or using a separate cold loop (usually without a storage capability). Point-of-use cool water drops and reduced temperature circulation loops are generally sterilized or high-temperature sanitized at defined and validated intervals. The design details of the WFI system varies with the incoming water quality, local utility costs, and operational demands. Very small operations may not have a WFI system and will utilize larger (5 L or larger) packages of WFI for formulation and cleaning.

Other grades of water may be present in parenteral facilities for use as initial rinses and detergent cleaning. The water utilized for these purposes is generally of relatively low bioburden and is often deionized, softened, ultra-filtered, or in some instances prepared by distillation or reverse osmosis, resulting in chemical purity similar to, if not identical to, WFI. Systems for the preparation of this water are subject to qualification, validation, and routine analysis to assure consistent quality.

### 5.2 Clean (Pure) Steam

Sterilizers and SIP systems in the facility are supplied with steam which upon condensation meets WFI quality requirements (testing steam condensate for microbial content is not fruitful). The steam can be produced directly from the water of sufficient purity to meet the input requirements of the steam generator. Steam generators are phase transition technologies that operate like a still, so it is no more necessary to provide these devices with WFI feed water than it would be to double distill WFI. (Production from WFI is certainly possible, but that is both expensive and an unnecessary precaution.) Modest quantities of steam can be produced from the first effect of a multiple effect WFI still, however, with a resultant loss of WFI output [18].

### 5.3 Process Gases

Air or nitrogen used in product contact is often supplied in stainless steel piping and ordinarily equipped with point-of-use filters; quite often an additional filter is placed within the distribution loop or at the entry point into a room resulting in a form of redundant filtration. Compressed air is typically provided by oil-free compressors to minimize potential contaminants and is often treated with a drier to obviate the possibility of condensation within the lines which could be a source of contamination. Nitrogen is supplied as a bulk cryogenic liquid. Argon and carbon dioxide have also been utilized as inerting gases, while propane or natural gas may be needed for sealing of ampoules.

### 5.4 Other Utilities

The operation of a parenteral facility often entails other utilities for the operation of the equipment. These include plant steam, jacket cooling water, and instrument air.

## 6 STERILIZATION AND DEPYROGENATION

The preparation of the drug formulation, components, and equipment entails the use of various sterilization/depyrogenation treatments to control bioburden, avoid excessive pyrogens, and to sterilize. The selection of the specific process must always fully consider the impact of the treatment on the items being sterilized/depyrogenated. Sterilization and heat depyrogenation processes must balance the effect of the treatment on the microorganism with the effect of that same treatment on the materials being processed. The choice of one method over another is often based upon achieving the desired sterilization/depyrogenation effect with minimal impact on the items critical quality attributes.

### 6.1 Steam Sterilization

The method of choice in nearly every instance is moist heat due to its lethality, simplicity, speed, and general ease of process development and validation. For the majority of items, this is accomplished in a double-door steam sterilizer, which is conventionally located between the preparations and aseptic processing (filling or compounding) areas. Steam sterilizers are routinely utilized for items such as elastomeric closures, process and vent filters, product contact parts, heat stabile environmental monitoring equipment, tools and utensils, hoses, sample containers, and other items unaffected by contact with saturated steam at commonly used sterilizing temperature and pressure [19]. Similar items utilized in the nonsterile compounding area would be processed in a similar manner. Regardless of their final destination or usage, items for steam sterilization should be protected from post-sterilization contamination by materials that are permeable to steam, air, or condensate but impenetrable by microorganisms. The wrapping materials would be maintained on the sterilized items until just prior to use. There are numerous publications that provide additional details on steam sterilization procedures [19–21].

Sealed containers of aqueous solutions, suspensions, and other liquids can be processed through steam sterilizers as well. These liquids might be used in formulation or cleaning procedures, and sterilization in this manner may be more efficient and more reliable than sterilizing filtration. Larger volumes of aqueous liquids are often sterilized in bulk using a jacketed and agitated pressure vessel (the vessel is usually rated for full vacuum as well).

Steam SIP is a widely used practice for the sterilization of equipment prior to the introduction of process materials and is the method of choice for holding tanks, process transfer lines, lyophilizers, and other large items. Conceptually, it has many similarities to sterilization in autoclaves but differs markedly due to the often custom designs of process equipment requiring SIP. Systems must be designed with careful consideration given to air removal and condensate draining, process sequenc-

ing, and poststerilization integrity to assure success [22]. Terminal sterilization of finished product containers is addressed later in this chapter.

### 6.2 Dry-Heat Sterilization and Depyrogenation

The use of dry heat for depyrogenation (and sterilization) is almost universal for glass containers. Temperatures of 250°C or higher are utilized to render the glass endotoxin free. The depyrogenation is necessary because the washing of glass to reduce particles can introduce unacceptable levels of gram-negative microorganisms whose presence could result in pyrogen formation. The depyrogenation process can assist in component surface treatment (siliconization is required for some formulations) and will also render the glass sterile as well (depyrogenation temperature conditions far exceed those needed for sterilization [23]).

Sterilization by dry heat is only infrequently used, preference being given to the use of steam (due to its higher speed) or dry-heat depyrogenation (affording an added measure of safety using the same equipment). Where it is employed temperatures in the range of 170–180°C are employed, and a batch oven is customarily used.

Dry-heat processes are conducted in either batch ovens or continuous tunnels, which are also installed between preparations and aseptic processing areas. Ovens have lower capacity and are typically found in smaller facilities. They offer the ability to handle items other than final product containers and thus can replace autoclaves in facilities where filling parts, feed hoppers, tools, and other items that must be extremely dry. Ovens should be equipped with internal HEPA filters, recirculating fans, heating/cooling coils, and a sophisticated control system [24]. Items prepared for dry-heat treatment in ovens are inverted or covered to protect them after exiting from the oven as there are no sealed protective systems suitable for the higher temperatures necessary for dry-heat depyrogenation or sterilization. Oven discharge is typically into a cool-down area (usually the same as that used for the sterilizer), though in small facilities it might discharge directly into the fill room. Unless ovens are used in conjunction with isolators, they require direct operator intervention to transfer containers to the filling line and to charge the line with depyrogenated glass. This constitutes a risky intervention which should be avoided. For this reason, batch glass processing is rare in all but the lowest throughput facilities.

Dry-heat tunnels are typically utilized where the production volumes are higher and allow for continuous supply of depyrogenated glass to the aseptic fill room. Tunnels are operated at high temperatures (>300°C) to increase processing speed and include a cooling zone that facilities discharge at or near room temperature. Typically, heating of the glass to 300°C or more for 3 or more minutes will result in much greater than the three-log endotoxin reduction required in current industry standards. The air inside the tunnel is HEPA filtered, and newer designs allow for dry-heat sterilization of the cooling zone as an added protective measure. Tunnels must be positioned with some care as they ordinarily will terminate into a fill room. A pressure differential between the cooling zone of the tunnel and the fill room is critical for proper operation of the tunnel. The pressure differential must conform to the requirements stipulated by the tunnel manufacturer. It is not necessary to have a >12.5 PA (particulate air) differential between the in-feed side of the heating zone of the tunnel and the exit side of the cooling zone. It has been suggested by

some that, since the in-feed side of the tunnel is typically in ISO 7 or 8 space, a greater differential is required; however, this is not true since the cooling zone is ISO 5, and the heating zone is certain to be sterile and is also ISO 5 in terms of particulate air quality. Their in-feed is often direct from a glass washer, which may be remote from the main preparations area utilized for washing, wrapping, and sterilizer loading.

### 6.3 Gas and Vapor Sterilization

The sterilization of materials using noncondensing gases (ethylene oxide, chlorine dioxide, or ozone) or condensing vapors such as hydrogen peroxide is a supplementary process intended for items that cannot be exposed to heat. The utilization of gas/vapor designs is coming into increased use as a supportive technology for isolation technology for presterilized items such as syringes and stoppers that must be introduced into the isolators aseptic zone. Air locks using these agents can be utilized in similar fashion for the supply of materials to manned clean rooms. Control over agent concentration or injection mass, relative humidity, and temperature may be required for these systems. There are different types of vapor processes available, and users should generally follow the cycle development strategy suggested by the manufacturer of the equipment they have chosen. Specific temperature and humidity ranges may be required for some vapor processes to assure appropriate efficacy [25, 26].

### 6.4 Radiation Sterilization

The use of radiation within a parenteral facility would have been considered unthinkable prior to the start of the twenty-first century. While $\gamma$ irradiation is typically a contracted service provided off-site, electron beam sterilization advances can make the installation of an in-house (and generally an in-line) system a real possibility. An in-line system would be utilized similarly to the gas/vapor systems described above for treatment of external surfaces for entry into either a clean room or isolator-based aseptic processing facility. The use of this same technology for terminal sterilization is also possible [1]. Association for the Advancement of Medical Instrumentation (AAMI)/ISO 11137 provides widely accepted guidance on the development and validation of radiation sterilization processes.

### 6.5 Sterilization by Filtration

Filters are utilized to sterilize liquids and gases by passage through membranes that retain microorganisms by a combination of sieve retention, impaction, and attractive mechanisms [27]. In contrast with the other forms of sterilization that are destructive of the microorganisms, filters rely on separation of the undesirable items (microorganisms as well as nonviable particles) from the fluid. Because filtration requires passage of the fluid from the "dirty" (upstream) side of filter to the clean (downstream) side of the filter, the downstream piping and equipment must be both "clean" and sterile prior to the start of the filtration process. This will ordinarily require the use of SIP procedures or sterilization followed by aseptic assembly.

Sterilizing filtration of parenterals is a complex and often inadequately considered subject, and numerous controls are required on the filter, fluid, and sterilizing/

operating practices employed. PDA Technical Reports 26 and 40 can be instructive in understanding the relevant concerns [28, 29].

## 7 FACILITY AND SYSTEM: QUALIFICATION AND VALIDATION

Facilities for the manufacture of sterile products require the qualification/validation of the systems/equipment and procedures utilized for that production. Each system described above and others with a direct/indirect impact on the quality of the products being produced should be placed into operation using a defined set of practices. The general approach is described below, and best practices include the development of traceable documentation from project onset. The preferred approach begins during a project's conceptual design phase where provisions for meeting the CGMP expectations and user requirement specifications establishing the technical basis for the processes are first defined. This is commonly followed by the validation master planning exercise in which the user requirement specifications are used as a basis for the development of acceptance criteria for process control studies. This effort should be accompanied by an analysis of risk that considers product attributes, target patient population, as well as technical and compliance requirements. Detailed design follows in which the specifics of the various systems are refined. Construction of the facility and fabrication of the process equipment follows and a variety of controls are necessary during these activities to satisfy user requirements for compliance of the various elements of the facility. Typically, factory acceptance testing (FAT) will be done on all key process equipment, usually at the manufacturer's plant site; much of the information gathered during FAT can be referenced in the qualification activities to follow. Physical completion is followed by a well-defined step termed commissioning in which construction and fabrication errors and omissions are addressed. Site acceptance testing of installed process equipment may be done in parallel with facility commissioning. Formal qualification of the facility ensues in which the installed systems and equipment are evaluated for their conformance to the design expectations. The very last steps in this process are variously termed performance qualification. Detailed discussion of these subjects is not possible within the constraints of this chapter, however the qualification/validation of equipment, systems, and processes has been extensively addressed in the literature [30].

## 8 ENVIRONMENTAL CONTROL AND MONITORING

Confirmation of appropriate conditions for aseptic processing and its supportive activities is required by regulation. In the highest air quality environment utilized for aseptic processing, ISO 5, there is a general expectation that the air and surfaces be largely free of microbial contamination and the number of particles be within defined limits (less than 3500 particles greater than 0.5 μm/m$^3$). Proving the complete absence of something is an impossible requirement, so the usual expectation is that 99+% of all samples taken from this most critical environment be free of detectable microorganisms. The minimum monitoring expectations for these environments as defined by the regulators are consistently attainable in nearly all instances,

especially those with lesser expectations. This is accomplished by proper design, periodic facility disinfection, and measures to control the ingress of microorganisms and particles for materials entering each environment from adjacent less clean areas [31].

### 8.1 Sanitization and Disinfection

Disinfection is customarily performed by gowned personnel during nonoperating periods using such agents as phenolics, quaternary ammonium compounds, aldehydes, and other nonsporicidal agents. The frequency of treatment varies with the ability of the facility to maintain the desired conditions between disinfection. Sporicidal agents such as dilute hydrogen peroxide or bleach are reserved for those occasional periods when control over the spore population warrants and is often employed after lengthy maintenance shutdowns or at the end of construction. Isolation technology replaces the manual disinfection with reproducible decontamination with a sporicidal agent and thus assures a superior level of environmental control as compared to manned environments. The manual treatments fall short of this level of control due to the uncertainties of the manual procedure and recontamination of the environment as a consequence of the very personnel and activities utilized to disinfect it. To mitigate these weaknesses, automatic sporicidal disinfection of manned clean spaces has been developed by multiple vendors. Disinfection of the less critical environments is accomplished in the same manner albeit on a less frequent interval befitting their higher allowable levels of microorganisms.

### 8.2 Monitoring

Aseptic environments are subject to a variety of monitoring systems including air, surface, and personnel monitoring for viable microorganisms and for nonviable particles. Environmental monitoring programs are often developed during the qualification of a new facility using a multiphase approach. Methods for the monitoring and expectations for performance have been extensively discussed in the literature and will only be addressed briefly in the context of this chapter [1, 2, 31, 32]. In general, the frequency and intensity of monitoring and concern for cleanliness increases as the product progresses from preparation steps (typically in ISO 7/8 environments) to more important activities (nonsterile compounding in ISO 6) and ultimately into the aseptic core (aseptic compounding and filling in ISO 5). Sampling site and time selection should be a balance between the need to collect meaningful data and avoidance of sampling interventions that could adversely (and inadvertently) impact product quality. Microbiological sampling must always be done by well-trained staff utilizing careful aseptic technique. This will both minimize risk to the product and also improve the reliability of the data by reducing the likelihood of false-positive results.

***Air Sampling*** The relative cleanliness of air in the most critical environment is assessed using passive sampling systems such as settle plates or estimated volumetrically using active air samplers. Active air samplers should be designed to be isokinetic in operation to avoid disruptions to unidirectional airflow. Considerable variability has been reported among the several sampling methods employed for

active air sampling, and there are also reports that active air sampling may have advantages in terms of sensitivity. Passive sampling using settle plates can be a useful adjunct in critical areas with limited access and where an active sampler might interfere with airflow or entail a worrisome intervention risk. It must be recognized that attempts to support the "sterility" of the cleanest aseptic environments (those in ISO 5) by aggressive sampling may have exactly the opposite effect. Sampling too frequently will increase process contamination risk by causing critical interventions that are best avoided within these very clean environments. As personnel are the greatest single source of microbial contamination and conduct the sampling, sampling intensity should be carefully considered. There is no value to taking air samples beyond those required to assess the relative cleanliness level within the environment.

***Surface Sampling*** Surfaces in the classified environments are monitored using a variety of methods but most commonly with contact plates (on smooth surfaces) or swabs (for irregular surfaces). Surface sampling in aseptic environments (ISO 5/6) is typically performed after the completion of the process to avoid the potential for adventitious contamination of the production materials as a consequence of sampling activities during the process. Fortunately, studies indicate that contamination does not build up during typical processing operations in modern clean rooms. Sampling with these materials may leave a trace of media or water on the sampled surface, and cleaning of the surface immediately after sampling is commonplace. Sampling of product contact surfaces (i.e., fill needles, feeder bowls, etc.) should only be performed after completion of the process, and the results of this testing should **not** be considered as an additional sterility test on the products. As in any form of manual environmental sampling, the risk of contamination by samplers during the processing of a sample makes the data less than completely reliable. Sampling of surfaces such as walls and floors should not be overdone because with good attention to aseptic technique they should be of little concern relative to actual process risk. Sampling on these surfaces is probably most useful in assessing ongoing changes in microflora and to confirm the adequacy of the disinfection program.

***Personnel Sampling*** The monitoring of personnel gown surfaces is an adaptation of surface sampling in which samples are taken from surfaces on the operator. In ISO 5 environments, this ordinarily entails the gloved hands and perhaps forearms. As with any other sampling of a critical surface (the gloved hand is often in closest proximity to sterile product contact surfaces and sterilized components), the sampling should be performed at the conclusion of the aseptic activity. Sampling during the midst of the process risks contamination of the product and should be avoided. Sampling of other aseptic gown surfaces is ordinarily restricted to gowning certification or postmedia fill testing, where more aggressive sampling can sometimes be informative. Whenever a gowned individual is sampled, the sample should be taken in the background environment (not ISO 5), and the individual should immediately exit and regown before continuing any further activity in the aseptic core area. Sampling of personnel in less critical environments can be useful; however, meeting regulatory expectations in these areas is ordinarily straightforward. Recommended contamination levels often distinguish among the different room classification levels found within clean rooms. While this may seem reasonable, it is not completely

logical since operators often move frequently between these different levels of classification during the conduct of their work.

***Total Particulate Monitoring*** Confirming the ability of the facility's heating, ventilation, and air-conditioning (HVAC) system to maintain the appropriate conditions throughout (to the extent practical) the classified environments is most easily accomplished using electronic total particle counters that can provide near immediate feedback on conditions during production operations. Total particle samples can be taken automatically, using permanently installed probes oriented into the unidirectional airflow. As such, they can be positioned proximate to critical activities to reaffirm the continued quality of the air in the vicinity of the sterile materials and surfaces. Manual total particulate air sampling can be a dangerous intervention and therefore if required should be timed so as to minimize risk to product. Attempts to correlate total particle counts with microbial counts have proven difficult. Correlations are only meaningful when the source of foreign material is personnel since people are the only source of airborne contamination within an aseptic processing area. When personnel are the only source of particulate, the ratio between viable and nonviable particles have been consistently found to be >1000:1, which means that in ISO 5 environments even relatively large total particulate count excursions would typically contribute microbial contamination that fell far below the limit of detection. Process equipment can and often does contribute airborne particulate matter but not detectable levels of microbial contamination. Also, microbial sampling is highly variable with respect to sensitivity, accuracy, precision, and limit of detection making correlations, particularly in rooms of highest air quality. So, it might seem logical to think that particle excursions are indicative of coincident microbial excursions especially in the cleaner environments (ISO 5) where the aseptic process takes place.

It is common practice for firms to interrupt their aseptic processes when atypical total particulate excursions are observed so that the scientists and engineers can determine the source of the foreign material. Monitoring frequency and expectations in the less critical environments is always reduced relative to the critical aseptic environments.

Where firms have introduced unidirectional air systems in preparations and compounding areas for particle control, there is often the temptation to expect these areas to meet the same microbial limits that these locations might attain in the aseptic core. This temptation should be resisted to avoid unnecessary sampling and deviations associated with expecting these environs to meet the conditions of aseptic areas where sanitization frequency, background environment, and most importantly personnel gowning are far superior to that found in the less clean locales [33].

***Housekeeping*** An important component of environmental control are the housekeeping activities utilized to clean the facility external to the controlled environments. Aseptic operations utilize a series of protective environments to protect the sterile field. Controls on the surrounding unclassified areas are an important part of the overall control scheme for sterile manufacturing. These unclassified areas support sterile operations in a variety of ways, and it is important to conduct activities therein that assist in the environmental control. Routine housekeeping, periodic sanitization, and even occasional environmental monitoring may be appropriate to

assure that microbial and particle loads on items, equipment, and personnel entering the classified environments is appropriately controlled.

## 9 PRODUCTION ACTIVITIES

The preparation of sterile materials requires execution of a number of supportive processes that together constitute the manufacturing process. They are intended to control bioburden, reduce particle levels, remove contaminants, sterilize, and/or depyrogenate. Nearly all of these activities occur within the controlled environments and are subject to qualification/validation.

### 9.1 Material and Component Entry

Prior to the start of any production activity, materials and components must be transferred from a warehouse environment into a classified environment. For most items this will necessitate removal from boxes or cartons, transfer to a nonwooden pallet, and passage through an air lock which serves as the transfer system between the controlled and uncontrolled environments. Often components are contained within plastic bags within a box or carton, and in some cases there are multiple bag layers to facilitate disinfection and passage through air locks into different zones of operation within the aseptic area. The firm may utilize an external disinfection of the materials in conjunction with this transfer. The concern is for minimization of particles and bioburden on these as yet unprocessed items in order to protect the controlled environment.

Raw materials may be weighed in a weigh area in which they are transferred to plastic bags and/or noncorrugate containers prior to the transfer. The weighing area provides ISO 7 or better conditions, and may be a dedicated portion of the warehouse proper; in a central weighing/dispensing area; or in a location contiguous to the compounding area. Sterile ingredients are never opened anywhere other than an aseptic environment and must be handled aseptically at all times including sampling and processing of samples.

### 9.2 Cleaning and Preparation

Once the container component items have been introduced into the preparations area, they must be readied for sterilization/depyrogenation. For many items this consists of washing/rinsing processes designed to remove particles and reduce bioburden and endotoxin levels. The application of silicone suspensions for glass or closure materials is sometimes employed to provide lubrication allowing smoother feeding of components or dispensing (elimination of product accumulation on vial). Following the cleaning, items for sterilization are dried, wrapped, and staged/stored for steam sterilization. Washed containers are either placed in trays or boxes for depyrogenation in ovens or are directly loaded into dry-heat tunnels. It is common practice to protect all washed items with ISO 5 air from the completion of washing, through either wrapping or placement into a sterilizer or oven for passage into the aseptic area. The intention is to avoid foreign matter that could result in contamination of product.

It is increasingly common for components to be supplied by the vendor in a ready-to-sterilize condition (washed and pretreated as necessary). Some items are available in a ready-to-use configuration with the supplier providing sterile and pyrogen-free components. The use of supplier-prepared items eliminates the need for preparation activities at the fill site and requires modification of material in-feed practices relative to on-site prepared items.

The process equipment (portable tanks, valves, fill needles, etc.) and consumable materials (filters, hoses, gaskets, etc.) are prepared using a variety of methods. Portable tanks are subjected to CIP (and perhaps SIP as well) in the preparation area. Smaller items are disassembled (if necessary) and cleaned either manually or in a cabinet washer. After cleaning they are wrapped and staged/stored prior to sterilization. Tubing should not be reused; its preparation typically consists of flushing with WFI followed by cutting to the required length. It is best to preassemble fill sets with tubing, filters, and fill needles/pumps and then wrap them in preparation for sterilization. This process obviates poststerilization assembly steps and therefore mitigates contamination risk. These steps may be performed in ISO 5 environments to reduce total particulate contamination on the items.

There are items that must be transferred into the aseptic processing area that cannot be treated within a sterilizer/oven. These include portable tanks, electronic equipment, and containers of sterile materials (ready-to-use items, sterile powders, environmental monitoring media, etc.). Air locks, pass-throughs, and similar designs are employed in which the exterior surfaces of the items are disinfected. The disinfection process may be completed by personnel outside and/or inside the aseptic area depending upon the specifics of the design.

At the completion of the cleaning process, the items should be free of contaminating residues including traces of prior products, free of endotoxin, and well-controlled in terms of total particulate and microbial levels. This level of control would be appropriate regardless of whether the items, equipment, or components are to be sterilized or not. Sterilization, other than by relatively high temperature dry heat, has only a modest impact on endotoxin levels; cleaning provides the only means to control endotoxin for materials and equipment that is sterilized by other means.

### 9.3 Compounding

Fixed equipment in the compounding area (nonaseptic or aseptic) is cleaned in place. This eliminates traces of prior products, particles, and pyrogens. Sterilization in place is required for the aseptic fixed equipment and is sometimes employed for the nonaseptic equipment as well as a bioburden control measure. Fixed transfer lines must be cleaned and sterilized as well, and this is accomplished independently or in conjunction with the vessels. The reuse of hoses and tubing is discouraged as cleaning and extractables cannot be confirmed beyond a single use.

The preparation of the product is performed within a classified environment with careful attention to the batch record, especially for time limits and appropriate protection of materials during handling to guard against all forms of contamination. This is proper for nonsterile compounding to minimize contamination prior to filtration/sterilization and is required for aseptic compounding activities. Barrier designs and other means of physically separating the worker from the product are recommended as a minimum even in nonaseptic compounding. As compounding may

expose the worker to a variety of potent/toxic materials, the use of personnel protective equipment may be required. In extreme cases, the use of containment system may be required to protect the compounding operator.

Where the compounding is nonaseptic, careful control over the environment, materials, and equipment is still appropriate to reduce viable/nonviable levels and to reduce the potential for endotoxin. Time limits should be imposed on manufacturing operations for additional control over microorganisms and thus microbial toxins.

Once the materials have been sterilized, interventions near either the formulation or product contact surfaces/parts should be minimized. Direct handling of these materials should only be done with sterilized tools or implements; nonsterile objects, such as operator gloves, should never directly contact a sterilized surface. Sampling, filter integrity testing, process connection, and other activities should all be designed to eliminate the need for personnel exposure to sterile items.

Aseptic compounding is often a required activity for sterile products that cannot be filter sterilized. The preparation of the sterile solids for use in these formulations is outside the scope of this chapter, but it is often acknowledged as the most difficult of all pharmaceutical processes to properly execute. Handling these materials at the fill site is performed using ISO 5 environments, and the use of closed systems is preferred [34].

## 9.4 Filling

Aseptic filling is performed in ISO 5 environments, and a variety of approaches are utilized with the technology choice largely dependent upon the facility design, batch size, and package design. Older plants utilize manned clean rooms in which aseptically gowned personnel operate the filling equipment: performing the setup, supplying components, making any required adjustments, and conducting the environmental monitoring. As human operators are directly or indirectly responsible for essentially all microbial contamination, aseptic filling operations are increasingly designed to minimize the potential for operator contamination to enter the critical environment. Barriers of various sophistication and effectiveness are employed to increase the protection afforded to sterile materials. The most evolved of the cleanroom designs are RAB systems in which personnel interventions are restricted to defined locations. Many newer facilities utilize isolation technology in which the filling environment is fully enclosed and personnel contamination is completely avoided.

Filling designs for syringes and ampoules differ only with respect to the details of component handling and closure design. However, it is wise not to underestimate the influence of both component quality and component handling reliability on contamination control in aseptic processing. Components that minimize the need for intervention and equipment that is rather tolerant of component variability will result in better contamination control performance. Aside from these distinctions, the range of filling technologies previously described is also possible.

The filling of plastic containers is accomplished using two very different approaches. Pre-formed containers can be sterilized in bulk, introduced into the aseptic suite via air locks, oriented (unscrambled), and filled. Blow-fill-seal prepares sterile bottles (most often LDPE) on line just prior to filling and sealing.

Filling of suspensions, emulsions, and other liquids may require slightly different filling designs to assure uniformity of dose in each container. Ointments and creams are sometimes filled at elevated temperatures to improve their flow properties through the delivery and filling equipment. These are ordinarily filled into presterilized plastic tubes that have largely replaced aluminum tubes for these formulations. Powders are typically filled in vials using equipment specifically engineered for that purpose.

An inerting gas (typically nitrogen, but other gases can be utilized) may be added to the headspace of the container to protect formulations that are oxygen sensitive. If the product is particularly sensitive to oxygen, purging may be done in the empty container prior to filling and again immediately after filling. Products may also be filled in an isolator under a nitrogen atmosphere if required. Products that require inert gas purging will also generally require inert gas for pressurization of tanks to provide motive force to drive the product through the filter(s) and into the filling reservoir.

## 9.5 Stoppering and Crimping

If the product is not freeze dried, the primary closure or "stopper" is applied shortly after completion of the filling process to better assure the sterility of the contents. When the product is to be lyophilized, the stopper may be partially inserted after filling and be fully seated after completion of the lyophilization cycle. Alternatively, the container could be left open and a stopper applied after completion of the drying.

Crimping is the act of securing the closure to the vial. It must be performed with sufficient uniform downward force to assure the container is properly secured. Too little downward force results in inadequately secured closures, while excessive force can result in container breakage. The force contributed by the crimp roller may be controllable as well.

Applying the closure to syringes, ampoules, and other containers usually differs in methodology from the approaches used for vials, but the objective is identical to secure the container's contents fully assuring the product's critical quality attributes (especially sterility) are maintained throughout its shelf life.

## 9.6 Lyophilization

Lyophilization (or freeze-drying) is a process utilized to convert a water-soluble material filled into a container to a solid state by removal of the liquid while frozen. The process requires the use of deep vacuums and careful control of temperatures. By conducting the process under reduced pressure, the water in the container converts from ice directly to vapor as heat is applied and is removed from the container by the vacuum. The dissolved solids in the formulation cannot undergo this phase change and remain in the container. At the completion of the cycle, the container will be returned to near atmospheric pressure; stoppers are applied or fully seated and crimped as described above. Lyophilization is particularly common with biological materials whose stability in aqueous solution may be relatively poor. The time period in solution and the temperature of the solution are kept at a specified low temperature to prevent product degradation [35].

As partially stoppered but unsealed containers must be transferred to the lyophilizer from the fill line, various designs have been utilized to protect the containers during this transit. Among the common alternatives utilized are the following:

- Placement of the lyophilization in the wall of the fill room to allow for direct loading
- Battery-operated unidirectional airflow carts to a remote lyophilizer
- ISO 5–protected conveyors with single shelf loading
- Transfer utilizing isolator technology

The use of trays for supporting the containers during the transfer, loading, lyophilization, and unloading steps was at one time common. The major problem with the use of trays for this purpose was the heat/handling-related distortion of the tray bottom that impacted the uniformity of the heating process in the freeze dryer. This was overcome by the use of trays with bottoms that were removed after loading and reinserted after completion of the drying. The current preference is for the placement of the containers directly on the shelf eliminating the trays entirely. This is accomplished by single height loading/unloading of the individual shelves with various pusher designs.

The use of thermocouples to monitor product temperature inside selected vials with the lyophilizer is still the prevalent practice. The utility of this data is questionable and the current trend is to eliminate this "requirement" as soon as possible to better assure sterility of the unsealed vials by eliminating placement of the thermocouples.

The lyophilizer chamber and condenser should be cleaned with a CIP system after each batch to prevent cross-contamination and, after cleaning, both should be sterilized. If a slot door loading system is utilized, periodic opening of a full door in the lyophilizer may be required to remove stoppers and glass that may have fallen.

## 10 PERSONNEL

Aseptic processing in the pharmaceutical industry is almost entirely dependent upon the proficiency of the personnel assigned to this most critical of all activities. The operators must be able to consistently aseptically transfer sterile equipment and materials in a manner that avoids contamination of those materials [1]. This is no mean feat given the contamination continuously released by personnel and the prevailing need for personnel for execution of the process activities.

Personnel proficiency in aseptic operations must be firmly established before they are allowed to conduct critical aseptic process steps. Operators must master a number of relevant skills in order to be declared competent. The usual progression is from classroom training (CGMP, microbiology, sterilization, etc.) to relevant practical exercises (aseptic media transfers, aseptic gowning rehearsals) and ultimately to the core aseptic skills required (aseptic gowning certification, aseptic

assembly/technique) using a growth medium. Through this approach the operator gradually acquires the necessary skills to be a fully qualified member of the production staff. Training/qualification of personnel is an ongoing requirement and must be repeated periodically to assure the skills are maintained. Continuing evaluation of operator qualification is accomplished using written examinations, practical challenges, documented observation, and participation in process simulation trials.

There is general acknowledgment of the risk associated with heavy reliance on personnel for aseptic processing. This has fostered much of the innovative designs for aseptic filling such as RABS and isolators where personnel are largely removed from the critical environment. The future will undoubtedly witness aseptic technologies where human interaction with sterile materials has been eliminated.

## 11 ASEPTIC PROCESSING CONTROL AND EVALUATION

The preparation of any pharmaceutical product requires controls over the production operations to assure the end result is a product that meets the required quality attributes. The methods utilized for this control are supported by formalized validation studies in which proof of consistency is demonstrated by appropriately designed experiments. The definition of appropriate operating parameters is the primary objective of the development activities and is further confirmed during scale-up to commercial operations. The validation supports that the routine controls applied to the process are appropriate to assure product quality [36]. This is typically accomplished in formalized validation activities in which expanded sampling/testing of the product materials is performed to substantiate their uniformity and suitability for use [30].

### 11.1 In-Process Testing

The sampling and testing of in-process materials during the course of the manufacturing process can confirm that essential conditions have been provided. This is appropriate in preparation, compounding, and filling activities. Sampling in preparation processes can confirm the absence of particles, proper siliconization levels, and cleanliness of equipment to assure that production items and equipment are suitable for use. Samples for microbiological quality, must, as previously mentioned, always be done by fully gowned staff under ISO 5 conditions using excellent aseptic techniques. During compounding, in-process testing can confirm proper pH, dissolution of materials, bioburden, and potency prior to filling. Filling operations can be monitored for fill volume (weight), headspace oxygen, and particles. These activities can all be automated to reduce interventions. These are typical examples of in-process controls utilized to assure acceptability of the process while it is underway. In the event of an abnormal result, corrective measures could be applied before further processing. The validation effort supports that these control measures are sufficient to assure product quality, when met during production operations. The sample intervals, sizes, and locations for in-process testing are chosen to enhance the validation. The tolerance limits are usually tightened relative to the release requirements to further assure that no out-of-tolerance materials are produced.

## 11.2 End-Product Testing

Upon completion of the process, samples are taken to establish that the batch meets the final product specifications defined for release. Predefined sampling plans are utilized to obtain representative samples of the entire batch, the prior validation effort having assured through an expanded sampling effort that the process provides a uniform product. End-product sampling often suffers from the inability to link an anomalous result with a specific portion/segment of the batch. If the validation is insufficiently rigorous, an out-of-specification result will ordinarily result in rejection of the batch and little opportunity to take effective corrective action.

The FDA has been supportive of the use of process analytical technologies (PATs) as an improvement on end-product testing [37]. These are intended to act as on-line indicators of critical product attributes enabling immediate corrective action and preventing the production of off-specification materials. This approach is common in the continuous process industries where feedforward controls are often employed. Their application to the more batch-oriented pharmaceutical/biotechnology industry is an acknowledgment that this approach can assure product quality more fully than a sampling-based approach. The PAT applications are still relatively few in number, but their utility in lieu of traditional quality methods is certainly promising.

The preceding relates solely to product quality attributes, based upon chemical or physical requirements. Assurance of sterility, the most critical of all the quality components for an aseptically filled sterile product relies on the following:

- The validation of the various sterilization processes for preparation of materials, equipment, and formulations
- The design of the aseptic manufacturing process and facility
- The establishment and maintenance of a proper processing environment
- Most importantly, the proficiency of the operating personnel directly involved with the aseptic process

There is no direct means to evaluate the cumulative capability of these measures. We infer success in aseptic processing through the evaluation of indirect measures of performance: air pressure differentials, total particle counts, viable monitoring results, and end-product sterility testing. The enormous challenge of aseptic processing is that none of the in-process or end-product testing results can prove that the attribute of sterility is attained with a high degree of certainty. Therefore, we rely on validation and the demonstration of a validated state of control to infer the adequacy of our contamination control efforts.

## 11.3 Process Simulations

An indirect means of assessing a facility's aseptic processing performance is the process simulation (or media fill) test [38]. This test substitutes a growth medium for the product in the process from the point of sterilization through to closure of the product container. The expectation is that successful handling of the growth media through the operating steps provides assurance that product formulations handled in a similar fashion would also be successful [39]. Process simulations

culminate in the incubation of the media-filled containers with success defined as a limited number of contaminated units in a larger number of filled units. The result is a contamination rate for the media fill, and not a direct indication of the level of sterility assurance afforded to aseptically processed materials using the same procedures. At the present time, the level of sterility provided to aseptically processed materials cannot be measured. The FDA and EMEA have harmonized their expectations relative to process simulation performance, but they have also asserted that the goal in every process simulation is zero contamination [1, 2]. This formalized expectation and recognition that patient safety should always be preeminent have resulted in substantial improvements in aseptic processing technology over the last 20 years.

## 12 TERMINAL STERILIZATION

Terminal sterilization is a process by which product is sterilized in its final container. Terminal sterilization is the method of choice for products that are sufficiently stabile when subjected to a compatible lethal treatment. Because the process utilized is expected to be lethal to the microorganisms present, is highly reproducible, and generally readily validated, there is a clear preference for its use [1, 40, 41].

The predominant method for terminal sterilization is moist heat, and a substantial percentage of sterile products are processed in this manner. (Estimates range from 5 to 15% of all sterile products are terminally sterilized.) The sterilization often requires the attainment of a balance between sterility assurance and degradation of the material's essential properties [42]. The overkill sterilization method is preferred for heat-resistant materials, and may be usable for terminal sterilization where the formulation can tolerate substantial heat input. The bioburden/biological indicator approach uses less heat input but requires increased control over the titer and resistance of the bioburden organisms present.

The large-volume parenteral (LVP) industry sometimes uses dedicated nonaseptic filling systems for its containers prior to subjecting them to terminal treatments. These LVP systems may approach the aseptic designs described earlier, but they are not supported by the same levels of environmental monitoring nor process simulation. Application of terminal sterilization at small volume parenteral producers may be done after the product is aseptically filled, although this practice is usual only where the firm produces predominantly aseptically filled products and would not have a filling system dedicated to terminally sterilized formulations. Product that will be subject to terminal sterilization may be filled under clean conditions with reduced environmental monitoring and control. However, control of total particulate levels requires unidirectional airflow for critical filling or assembly processes.

Terminal sterilization is most commonly accomplished by moist heat. Terminal sterilization by other means is certainly possible, and a very limited number of parenteral drugs are treated with dry heat or radiation after filling. There is growing interest in the use of radiation, including low-energy E-beam, as a terminal treatment suggesting more products will be processed in this manner.

Although there are numerous advantages to terminal sterilization, there can be very good reasons for aseptically filling products that are stabile enough to be com-

patible with a sterilization process. For example, multichamber containers that cannot withstand terminal sterilization may provide a very important safety benefit to the patient by reducing aseptic admixture or reconstitution in the clinic. These aseptic activities when conducted in clinics are generally not able to be done within anything like the controls required in industrial aseptic processing. It is often beneficial to discuss processing technology choices with regulatory authorities early in the development of a new product.

## 13 CONCLUSION

The manufacture of parenteral drugs by aseptic processing has long been considered a difficult technical challenge. These products require careful control and stringent attention to detail to assure their safety. Aseptic processing done with discipline and taking advantage of the numerous technical developments that have occurred over the years results in sterile products that can be administered with complete confidence. The wider adaptation of advanced aseptic processing will result in further evolutionary improvements in aseptic processing. The industry is at the beginning of an era in which human-scale aseptic processing will be completely replaced by separative technologies and process automation. Additionally, improved in-process controls are likely to be implemented making validation easier and easing the compliance burden.

## APPENDIX

**Parenteral Drug Association, Bethesda, Maryland**

TM 1: Validation of Steam Sterilization Cycles, 1978

TR 3: Validation of Dry Heat Processes used for Sterilization & Depyrogenation, 1981

TR 7: Depyrogenation, 1985

TR 11: Sterilization of Parenterals by Gamma Irradiation, 1988

TR 13: Fundamentals of an Environmental Monitoring Program, 2001

TR 22: Process Simulation Testing for Aseptically Filled Products, 1996

TR 26: Sterilizing Filtration of Liquids, 1998

TR 28: Process Simulation Testing for Sterile Bulk Pharmaceutical Chemicals, 2006

TR 34: Design & Validation of Isolator Systems for the Manufacture & Testing of Health Care Products, 2001

TR 36: Current Practices in the validation of Aseptic Processing, 2002

TR 40: Sterilizing Filtration of Gases, 2005

**International Society For Pharmaceutical Engineering, Tampa, Florida**

Baseline Guide, Vol. 3: Sterile Manufacturing Facilities, 1999

Baseline Guide, Vol. 4: Water and Steam Systems, 2001

Baseline Guide, Vol. 5: Commissioning and Qualification, 2001

## REFERENCES


1. U.S. Food and Drug Administration (FDA) (2004), Guideline on sterile drug products produced by aseptic processing, FDA, Washington, DC.
2. European Union (EU) (2006), Annex 1—Sterile medicinal products—draft revision.
3. International Organization for Standardization (ISO), international standard 14644 1-3.
4. U.S. Food and Drug Administration (FDA) (2004), Pharmaceutical CGMPs for the twenty-first century—A risk-based approach, FDA, Washington, DC.
5. International Conference on Organization (ICH) (2005), Draft consensus guideline quality risk management Q9, draft.
6. Whyte, W., and Eaton, T. (2004), Microbiological contamination models for use in risk assessment during pharmaceutical production, *Eur J Parenteral Pharm Sci*, 9(1).
7. Whyte, W., and Eaton, T. (2004), Microbial risk assessment in pharmaceutical cleanrooms, *Eur J Parenteral Pharm Sci*, 9(1).
8. Agalloco, J., and Akers, J. (2006), Simplified risk analysis for aseptic processing: The Akers-Agalloco method, *Pharm Technol*, 30(7), 60–76.
9. International Organization for Standardization (ISO) (2004), Cleanrooms and associated controlled environments—Part 7: Separative devices (clean air hoods, gloveboxes, isolators and mini-environments), ISO 14644-7.
10. Agalloco, J. (2006), Thinking inside the box: The application of isolation technology for aseptic processing, *Pharm Technol*., p. S8–11.
11. Lysford, J., and Porter, M. (2003), Barrier isolators history and trends, *Pharm Eng*, 23(2), 58–64.
12. ISPE (2005), Restricted access barrier systems (RABS) for aseptic processing, ISPE definition, Aug. 16.
13. Wikol, M. (2004), GoreTM vial isolator, ISPE presentation, Feb. 12.
14. Py, D. (2004), Development challenges for intact sterile filling, PDA presentation, Mar. 9.
15. Thilly, J. (2004), CVFL technology from lab scale to industry, PDA presentation, Mar. 8.
16. ISPE (2001), Water and Steam Systems Baseline® guide.
17. ISPE (1999), Sterile Manufacturing Facilities Baseline® guide.
18. ISPE (2001), Water and Steam Systems Baseline® guide.
19. PDA (2006), Technical Monograph 1, Industrial moist heat sterilization in autoclaves, draft 17.
20. Perkins, J. (1969), *Principles and Methods of Sterilization in Health Sciences*, Charles Thomas, Springfield, IL.
21. Phillips, G. B., and Morrissey, R. F. (1993), *Sterilization Technology: A Practical Guide for Manufacturers and Users of Health Care Products*, Van Nostrand Reinhold, New York.
22. Agalloco, J. (1998), Sterilization in place technology and validation, in Agalloco, J., and Carleton, F. J., Eds., *Validation of Pharmaceutical Processes: Sterile Products*, Marcel Dekker, New York.
23. PDA (1981), Technical Report 3, Validation of dry heat processes used for sterilization and depyrogenation.
24. Case, L., and Heffernan, G. (1998), Dry heat sterilization and depyrogenation: Validation and monitoring, in Agalloco, J., and Carleton, F. J., Eds., *Validation of Pharmaceutical Processes: Sterile Products*, Marcel Dekker, New York.
25. Burgess, D., and Reich, R. (1993), Industrial ethylene oxide sterilization, in Phillips, G. B., and Morrissey, R. F. Eds., *Sterilization Technology: A Practical Guide for Manufacturers and Users of Health Care Products*, Van Nostrand Reinhold, New York.

26. Sintim-Damao, K. (1993), Other gaseous sterilization methods, in Phillips, G. B., and Morrissey, R. F. Eds., *Sterilization Technology: A Practical Guide for Manufacturers and Users of Health Care Products*, Van Nostrand Reinhold, New Youk.
27. Meltzer, T., Agalloco, J., et al. (2001), Filter integrity testing in liquid applications; Revisited, Part 1, *Pharm Technol*, 25(10), and Part 2, *Pharm Technol*, 25(11).
28. PDA (1998), Technical Report 26, Sterilizing filtration of liquids.
29. PDA (2005), Technical Report 40, Sterilizing filtration of gases.
30. Agalloco, J., and Carleton, F. J., Eds. (1998), *Validation of Pharmaceutical Processes: Sterile Products*, Marcel Dekker, New York.
31. PDA (2001), Technical Report 13, Fundamentals of an environmental control program.
32. USP ⟨1116⟩ (2005), Microbiological control and monitoring environments used for the manufacture of healthcare products, *Pharm Forum*, 31(2), Mar.–Apr.
33. Agalloco, J. (1996), Qualification and validation of environmental control systems, *PDA J Pharm Sci Technol*, 50(5), 280–289.
34. PDA (2006), Technical Report 28, Process simulation testing for sterile bulk pharmaceutical chemicals.
35. Trappler, E. (1998), Validation of lyophilization, in Agalloco, J., and Carleton, F. J., Eds., *Validation of Pharmaceutical Processes: Sterile Products*, Marcel Dekker, New York.
36. Chapman, K. G. (1984), The PAR approach to process validation, *Pharm Technol*, 8(12), 22–36.
37. Food and Drug Administration (FDA) (2004), PAT guidance for industry—A framework for innovative pharmaceutical development, manufacturing, and quality assurance, FDA, Washington, DC.
38. PDA (1998), Technical Report 22, Process simulation testing for aseptically filled products.
39. Agalloco, J., and Akers, J. (2006), Aseptic processing for dosage form manufacture: Organization & validation, in Carleton, F. J., and Agalloco, J. P., Eds., *Validation of Pharmaceutical Processes: Sterile Products*, Marcel Dekker, New York.
40. Food and Drug Administration (FDA) (1991), Use of aseptic processing and terminal sterilization in the preparation of sterile pharmaceuticals, FR 56, 354–358.
41. PIC/S41. (1999), Decision trees for the selection of sterilisation methods (CPMP/QWP/054/98).
42. PDA (2006), Technical Monograph 1, Industrial moist heat sterilization in autoclaves, draft 17.

## ADDITIONAL READINGS

Akers, J. (2001), An overview of facilities for the control of microbial agents, in Block, S. S., Ed., *Disinfection, Sterilization and Preservation*, 5th ed, Lippincott, Williams and Wilkins, Philadelphia, pp. 1123–1138.

Akers, J., and Agalloco, J. (1997), Sterility and sterility assurance, *J Pharm Sci Technol* 51, 72–77.

Cole, J. C. (1990), *Pharmaceutical Production Facilities—Design and Application*, Ellis Norwood, Chicester.

Institute of Environmental Science and Technology (IEST) (1995), Compendium of standards, practices, and similar documents relating to contamination control, CC009/IESCC009.2, IEST, Mt. Prospect, IL.

Ljungvist, B., and Reinmueller, B. (1995), *Ventilation and Airborne Contamination in Clean Rooms*, Pharmacia A/B, Stockholm.

Reinmuller, B. (2000), Microbiological risk assessment of airborne contaminants in clean zones, Bulletin No. 52, Royal Institute of Technology/Building Services and Engineering, Stockholm.

United States Pharmacopoeia/National Formulary (2006), 29, Chapter 1116, Microbial evaluation of clean rooms, Rockville, Maryland, pp. 2969–2976.

# FROM PILOT PLANT TO MANUFACTURING: EFFECT OF SCALE-UP ON OPERATION OF JACKETED REACTORS

B. Wayne Bequette
*Rensselaer Polytechnic Institute, Troy, New York*

**Contents**

1 Motivation
2 Background
  2.1 Pharmaceutical Process Development
  2.2 Batch Reactors
  2.3 Reaction Calorimetry
3 Laboratory Vessels and Reaction Calorimeters
  3.1 Material and Energy Balances
  3.2 Estimating Fluid Properties and Heat Transfer Coefficients from Calorimeter Data
  3.3 Estimating Heat Flows
  3.4 Relating Heat Flows and Conversion
  3.5 Semibatch Reactions
  3.6 Rapid Scale-Up Relationships
  3.7 Strategy under a Cooling System Failure
4 Heat Transfer in Process Vessels
  4.1 Heat Transfer Relationships
  4.2 Effect of Reactor Type, Jacket Heat Transfer Fluid, and Reactor Fluid Viscosity
  4.3 Pilot- and Production-Scale Experiments
5 Dynamic Simulation Studies
6 Summary
  References

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

## 1 MOTIVATION

There are many phases of process development between the discovery of an active pharmaceutical ingredient and the design, construction, and operation of a manufacturing process to produce a drug. A sequence of reactions and separations that is successful at the bench scale may lead to a process that is unsafe, is difficult to operate, or produces unsatisfactory product at the manufacturing scale. A manufacturing process typically has a large sequence of steps, involving several different unit operations (heat exchangers, reactors, separators, etc.), and a complete review of the design and scale-up of these unit operations would constitute a chemical engineering curriculum; thus, the focus of this chapter is the scale-up of jacketed batch chemical reactors from the laboratory to the pilot plant and manufacturing. These reaction vessels often serve many functions, including mixing, heating, cooling, distillation, and crystallization.

Temperature control for laboratory reactors is typically easy because of high heat transfer area–reactor volume ratios, which do not require large driving forces (temperature differences) for heat transfer from the reactor to the jacket. Pilot- and full-scale reactors, however, often have a limited heat transfer capability. A process development engineer will usually have a choice of reactors when moving from the laboratory to the pilot plant. Kinetic and heat of reaction parameters obtained from the laboratory reactor, in conjunction with information on the heat transfer characteristics of each pilot plant vessel, can be used to select the proper pilot plant reactor.

Similarly, when moving from the pilot plant to manufacturing, a process engineer will either choose an existing vessel or specify the design criteria for a new reactor. A necessary condition for operation with a specified reactor temperature profile is that the required jacket temperature is feasible. We have therefore chosen to focus on heat transfer–related issues in scale-up. Clearly there are other scale-up issues, such as mixing sensitive reactions. See Paul [1] for several examples of mixing scale-up in the pharmaceutical industry.

In this chapter we discuss important issues as we move from laboratory to pilot plant and manufacturing. A review of batch process operation and pharmaceutical research is covered in Section 2, followed by laboratory vessels and reaction calorimetry in Section 3. In Section 4 heat transfer in process vessels is presented, including the effect of reactor type and heat transfer fluid on the vessel heat transfer capability. In Section 5 dynamic behavior based on simulation studies is discussed.

## 2 BACKGROUND

### 2.1 Pharmaceutical Process Development

Anderson [2] presents a wide range of topics on pharmaceutical process development, including a number of different problems related to process scale-up, such as solvent and reagent selection, purification, and limitations to various operations. He notes that most reactors used for scale-up operations are selected for flexibility in running many different processes, especially for pilot plants and multiproduct manufacturing plants.

Pisano [3] discusses the management of process development projects in the pharmaceutical industry. Case studies are used to illustrate the effect of resource allocation decisions at different stages of a project. While there has been a focus on product development in the pharmaceutical industry, clearly process development plays an important role in getting a product to market and lowering the long-term product manufacturing costs.

### 2.2 Batch Reactors

Batch processes present challenging control problems due to the time-varying nature of operation. Chylla and Haase [4] present a detailed example of a batch reactor problem in the polymer products industry. This reactor has an overall heat transfer coefficient that decreases from batch to batch due to fouling of the heat transfer surface inside the reactor. Bonvin [5] discusses a number of important topics in batch processing, including safety, product quality, and scale-up. He notes that the frequent repetition of batch runs enables the results from previous runs to be used to optimize the operation of subsequent ones.

LeLann et al. [6] discuss tendency modeling (using approximate stoichiometric and kinetic models for a reaction) and the use of model predictive control (linear and nonlinear) in batch reactor operation. Studies of a hybrid heating–cooling system on a 16-L pilot plant are presented.

Various aspects of the effect of process scale-up on the safety of batch reactors have been discussed by Gygax [7], who presents methods to assess thermal runaway. Shukla and Pushpavanam [8] present parametric sensitivy and safety results for three exothermic systems modeled using pseudohomogenous rate expressions from the literature. Caygill et al. [9] identify the common factors that cause a reduction in performance on scale-up. They present results of a survey of pharmaceutical and fine chemicals companies indicating that problems with mixing and heat transfer are commonly experienced with large-scale reactors.

### 2.3 Reaction Calorimetry

The microanalytical methods of differential thermal analysis, differential scanning calorimetry, accelerating rate calorimetry, and thermomechanical analysis provide important information about chemical kinetics and thermodynamics but do not provide information about large-scale effects. Although a number of techniques are available for kinetics and heat-of-reaction analysis, a major advantage to heat flow calorimetry is that it better simulates the effects of real process conditions, such as degree of mixing or heat transfer coefficients.

Regenass [10] reviews a number of uses for heat flow calorimetry, particularly process development. The hydrolysis of acetic anhydride and the isomerization of trimethyl phosphite are used to illustrate how the technique can be used for process development. Kaarlsen and Villadsen [11, 12] provide reviews of isothermal reaction calorimeters that have a sample volume of at least 0.1 L and are used to measure the rate of evolution of heat at a constant reaction temperature. Bourne et al. [13] show that the plant-scale heat transfer coefficient can be estimated rapidly and accurately from a few runs in a heat flow calorimeter.

Landau et al. [14] use a heat flow calorimeter to investigate feasible pilot plant operating conditions for the production of a pharmaceutical intermediate. They

determine kinetic and heat flow parameters using the calorimeter and further estimate heat transfer parameters for a pilot-scale reactor. Simulation studies are used to find the required jacket temperature for desired batch reactor temperature profiles. Semibatch operation is shown to be safer than normal batch operation. Landau [15] provides a detailed review of reaction calorimetry, including mathematical expressions for energy balances, and a number of application examples.

## 3 LABORATORY VESSELS AND REACTION CALORIMETERS

As reviewed in Section 2.3, reaction calorimeters can be used to better understand and characterize scale-related process phenomena, such as mixing and heat transfer. A heat flow calorimeter, the Mettler RC1e, is shown in Figure 1. A schematic of a similar calorimeter system is shown in Figure 2 [16]. A heat flow calorimeter can be used to estimate:

- Physical parameters (heat capacity)
- Reaction rate constants
- Heat transfer coefficients (overall, $U$ or, or film, $h_i$)

### 3.1 Material and Energy Balances

The overall energy balance for a process with no reaction has the form

Energy accumulation = energy in heat transfer from jacket
+ energy in by calibration probe – energy lost by ambient heat transfer

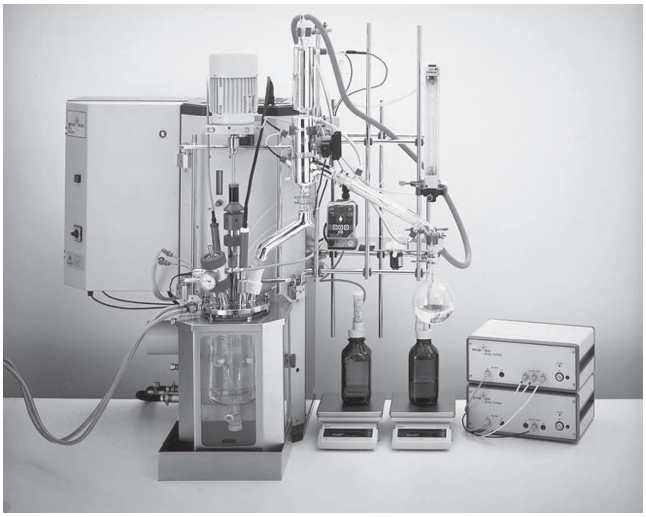

**FIGURE 1** Mettler RC1e heat flow calorimeter system (www.mt.com).

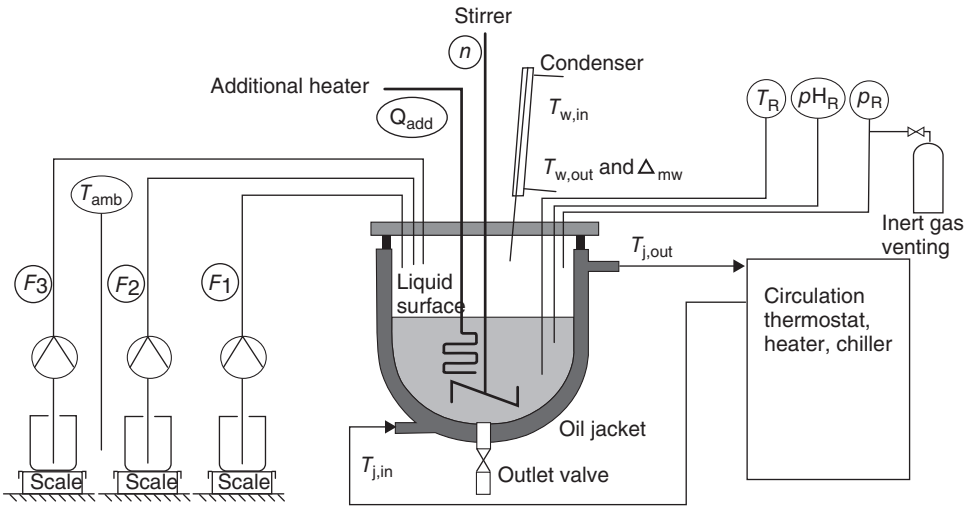


**FIGURE 2** Schematic of HEL SIMULAR reaction calorimeter. From ref. 16.

which is shown mathematically as

$$(mc_p)_r \frac{dT}{dt} = -UA(T - T_j) + q_{\text{cal}} - k_{\text{loss}}(T - T_{\text{amb}}) \tag{1}$$

where $(mc_p)_r$ is the reactor thermal capacitance, $T$ is the reactor temperature, $T_j$ is the jacket temperature, $U$ is the overall heat transfer coefficient, $A$ is the area for heat transfer, $q_{\text{cal}}$ is the heat flow from the calibration probe, and the final term accounts for heat loss from the reactor system. The thermal capacitance is composed of the fluid in the vessel as well as the inert components in contact with the fluid, including the vessel wall, agitator (stirrer), and sensors (e.g., thermocouple), as shown in the equation

$$(mc_p)_r = V\rho c_p + m_v c_{pv} \tag{2}$$

where $V$ is the volume of liquid, $\rho$ is the liquid density, $c_p$ is the liquid heat capacity, $m_v$ is the mass of the vessel wall and other inerts, and $c_{pv}$ is the average heat capacity of the vessel wall and inerts. The inert contributions and heat transfer to the ambient can be found from extensive calibration studies. For small-scale reactors, such as reaction calorimeters, the thermal mass of the inerts can be significant. The thermal capacitance ratio, sometimes called the Lewis number, is given as

$$\phi = \frac{(mc_p)_r}{V\rho c_p} = 1 + \frac{m_v c_{pv}}{V\rho c_p} \tag{3}$$

which can be on the order of 1.5–2 for a small-scale reactors and adiabatic calorimeters but is often 1.05–1.10 for small pilot plant reactors and less than 1.02 for manufacturing-scale reactors.

### 3.2 Estimating Fluid Properties and Heat Transfer Coefficients from Calorimeter Data

In a heat flow calorimeter, a feedback controller is used to maintain a constant desired reactor temperature by adjusting the jacket temperature. From (1), with a constant calibration probe heat flow, at steady state ($dT/dt = 0$), the overall heat transfer coefficient can be found from

$$UA = \frac{q_{\text{cal}} - k_{\text{loss}}(T - T_{\text{amb}})}{T - T_j} \tag{4}$$

Also, the fluid heat capacity can be found by ramping up the reactor temperature and using

$$(mc_p)_r = \frac{-UA(T - T_j) + q_{\text{cal}} - k_{\text{loss}}(T - T_{\text{amb}})}{dT/dt} \tag{5}$$

and solving for $c_p$ from (2), assuming that the reactor inert component contributions are known from previous studies. An example calibration study is shown in Figure 3, where a constant heat flow is applied from 35 to 42 min, enabling the heat transfer coefficient to be estimated from the temperature difference using Equation (4). Then, the heat capacity is estimated from the temperature ramp applied between 5 and 20 min. It should be noted that the heat transfer coefficient and heat capacity of the fluid may vary with concentration and temperature. Typically, calibration experiments are performed before and after the reaction; then the heat transfer coefficient and heat capacity are assumed to vary linearly with conversion or batch time. For polymerization reactions in particular, the viscosity can increase tremendously with conversion, causing a substantial decrease in the heat transfer coeffi-

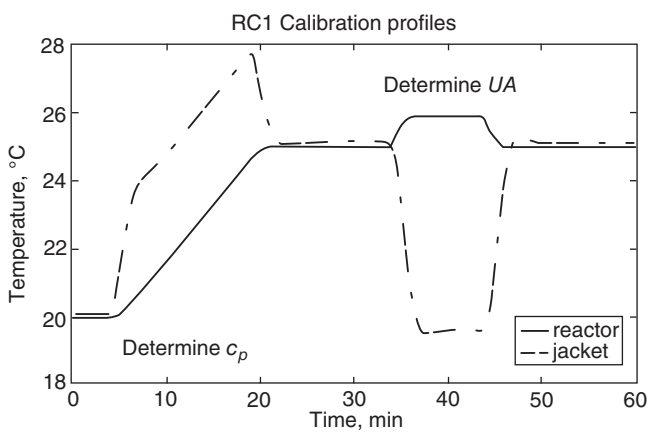


**FIGURE 3** Example reaction calorimetry study without reaction. The overall heat transfer coefficient area can be found during the steady-state temperature difference and known calibration probe heat flow, between 35 and 42 min. The heat capacity can then be found from the temperature ramp between 5 and 20 min.

cient. Reaction experiments can be run at several temperatures to find the functional relationship with temperature.

Since the heat transfer area as a function of liquid volume is known, the overall heat transfer coefficient $U$ can be calculated from (4). The overall heat transfer coefficient is calculated as

$$\frac{1}{U} = \frac{1}{h_i} + \frac{x_g}{k_g} \tag{6}$$

where the jacket side resistance is negligible. The glass vessel heat transfer resistance ($x_g/k_g$, thickness/thermal conductivity) can be used to find the reactor fluid heat transfer coefficient ($h_i$).

### 3.3 Estimating Heat Flows

The reaction heat flow can be found by rearranging (1), with the calibration heat probe replaced by the reaction heat flow, to find

$$q_r = (mc_p)_r \frac{dT}{dt} + UA(T - T_j) + k_{\text{loss}}(T - T_{\text{amb}}) \tag{7}$$

The total heat released during the reaction can be found by integrating (7),

$$Q_{\text{tot}} = \int_0^{t_f} q_r dt \tag{8}$$

or, represented as a scaled (per-unit mass) total heat release,

$$\bar{Q}_{\text{tot}} = \frac{Q_{\text{tot}}}{V\rho} = \frac{Q_{\text{tot}}}{m} \tag{9}$$

The molar heat of reaction can be found from

$$\Delta H_{\text{rxn}} = \frac{-Q_{\text{tot}}}{n_{\text{rxn}}} \tag{10}$$

where $n_{\text{rxn}}$ is the molar amount reacted.

As a "first-pass" calculation, if it is assumed that the dominant heat transfer resistance is on the reactor side, then the overall heat transfer coefficient ($U$) from (4) can be used for scale-up.

### 3.4 Relating Heat Flows and Conversion

The reaction heat flows are directly related to the conversion of reactants [14]. Consider a first-order reaction of a limiting reactant, with the rate expression

$$\frac{dC}{dt} = -kC \tag{11}$$

where $C$ is the molar concentration of the reactant. The heat flow is

$$q_r = \left(\frac{dC}{dt}\right)\Delta H_{\text{rxn}} V = -kC\Delta H_{\text{rxn}} V \tag{12}$$

with an initial heat flow of

$$q_{r0} = -kC_0 \Delta H_{\text{rxn}} V \tag{13}$$

dividing (12) by (13), we find the relationship between concentration and heat flow:

$$\frac{C}{C_0} = \frac{q_r}{q_{r0}} \tag{14}$$

For an isothermal reaction, the solution to (11) is

$$\frac{C}{C_0} = e^{-kt} \tag{15}$$

so, the heat flow for an isothermal reaction is

$$\frac{q_r}{q_{r0}} = e^{-kt} \tag{16}$$

Thus, the reaction rate constant $k$ can be estimated from the reaction heat flow without making any concentration measurements. Assuming an Arrhenius rate expression

$$k = A_0 e^{-E/RT} \tag{17}$$

the rate constant at several temperatures can be used to estimate the frequency factor ($A_0$) and activation energy ($E$). (See ref. 14 for an example application.)

### 3.5 Semibatch Reactions

For extremely exothermic reactions it is necessary to slowly add the feed over time, that is, operate in a semibatch fashion. The heat flow for a semibatch reaction can be found from

$$q_r = UA(T - T_j) + (mc_p)_r \frac{dT}{dt} + \dot{m}_f c_{pf}(T - T_f) \tag{18}$$

where $\dot{m}_f$ is the mass flow rate of the feed stream. If the reactor temperature is maintained constant, this reduces to

$$q_r = UA(T - T_j) + \dot{m}_f c_{pf}(T - T_f) \tag{19}$$

For reactions with essentially instantaneous kinetics, the reaction rate is limited by the feed addition rate. For other reactions, particularly if the reactor is operated at too low of a temperature, a reactant concentration can "build up," eventually reaching an unsafe level that could lead to a rapid temperature rise and explosion. It is important for these reactions to monitor the heat flow to confirm that the reactant concentration is not increasing to unacceptable levels.

### 3.6 Rapid Scale-Up Relationships

Lacking knowledge of the larger scale reactor, it is tempting to simply assume that only the area for heat transfer varies upon scale-up. A natural parameter is the *cooling time*,[1] defined as

$$\tau_{co} = \frac{(mc_p)_r}{UA} = \frac{V\rho c_p \phi}{UA} \tag{20}$$

The heat transfer area varies with the square of the vessel diameter, and the volume varies with the cube of the vessel diameter. Thus the area–volume ratio ($A/V$) varies with volume as

$$\frac{A}{V} \sim \frac{1}{V^{1/3}} \tag{21}$$

The inverse cooling time relationship for scale-up from volume $V_1$ to $V_2$ is

$$\left[\frac{UA}{V\rho c_p \phi}\right]_2 = \left[\frac{UA}{V\rho c_p \phi}\right]_1 \left(\frac{V_1}{V_2}\right)^{1/3} \tag{22}$$

The required reactor-jacket temperature difference on scale-up, with a constant Lewis number, is

$$[T - T_j]_2 = [T - T_j]_1 \left(\frac{V_2}{V_1}\right)^{1/3} \tag{23}$$

so the temperature difference can increase dramatically when a process is scaled up several orders of magnitude. Reactor-jacket temperature difference constraints can be particularly important for glass-lined vessels, where the limit is often 75 °C.

### 3.7 Safety under a Cooling System Failure

In the event of a cooling system failure it can be assumed that the reactor operates adiabatically. The adiabatic temperature rise can be found from

[1]The notion of cooling time can be understood by writing (1) and assuming no calibration energy or heat loss. Then (1) becomes $\tau_{co}(dT/dt) = -(T - T_j)$. If a constant temperature difference $T - T_j$ is applied, it will take $\tau_{co}$ time units for the reactor temperature to change by the temperature difference.

$$\Delta T_{\text{ad}} = \frac{Q_{\text{tot}}}{(mc_p)_r} \tag{24}$$

and the final temperature is

$$T_{\text{final}} = T_{\text{initial}} + \Delta T_{\text{ad}} \tag{25}$$

As long as the final temperature is less than some critical "onset" temperature where a secondary decomposition reaction occurs, then the process can safely handle a cooling system failure. If a batch reactor temperature cannot be assured to remain less than the onset temperature after a cooling system failure, then a semibatch operation should be used. As noted in Section 3.5, it is necessary to assure that reactant concentration is not increasing above an onset concentration where a similar decomposition could occur with a cooling system failure.

## 4 HEAT TRANSFER IN PROCESS VESSELS

Based on initial heat flow calorimetry studies, a process development engineer must choose the appropriate reactor vessels for pilot plant studies. A pilot plant typically has vessels that range from 80 to 5000 L, some constructed of alloy and others that are glass lined. In addition some vessels may have half-pipe coils for heat transfer, while others have jackets with agitation nozzles. A process drawing for a typical glass-lined vessel is shown in Figure 4. In Sections 4.1 and 4.2 we review fundamental heat transfer relationships in order to predict overall heat transfer coefficients. In Section 4.3 we review experimental techniques to estimate heat transfer coefficients in process vessels.

### 4.1 Heat Transfer Relationships

***Reactor-Side Coefficient*** The reactor-side heat transfer coefficient is calculated as

$$h_i = a\frac{k_i}{D_i}\text{Re}_i^{0.67}\,\text{Pr}_i^{0.33} \tag{26}$$

where $a$ is the agitation constant (0.33), $k_i$ is the fluid thermal conductivity, $\text{Re}_i$ is the Reynolds number, and $\text{Pr}_i$ is the Prandtl number,

$$\text{Re}_i = \frac{D_{ag}^2 N \rho_i}{\mu_i} \tag{27}$$

$$\text{Pr}_i = \frac{\mu_i c_{pi}}{k_i} \tag{28}$$

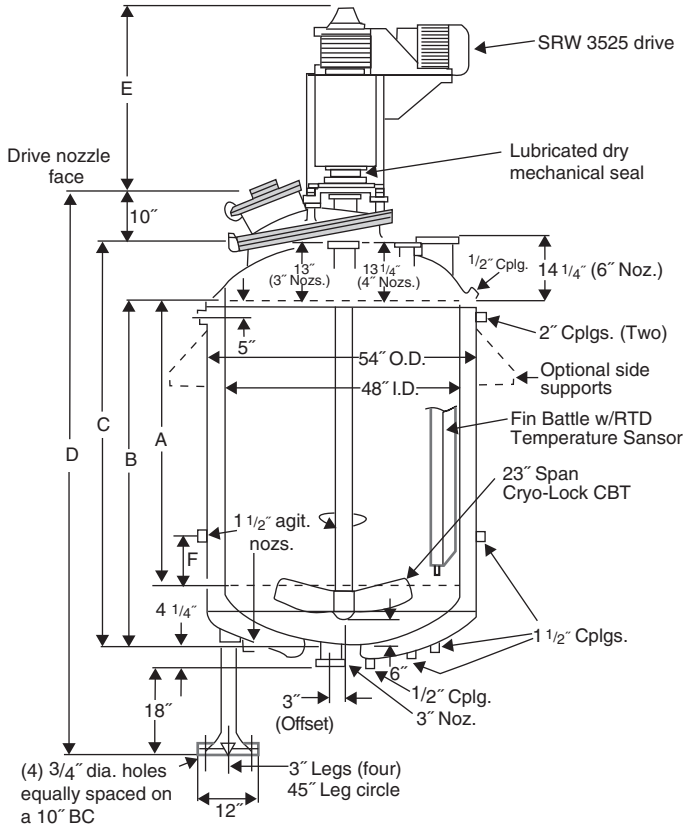


**FIGURE 4** Typical 300- or 500-gal jacketed vessel (www.pfaudler.com).

and $N$ is the agitator rotation rate. It should be noted that the film heat transfer coefficient varies inversely with the viscosity, that is,

$$h_i \sim \frac{1}{\mu_i^{0.33}} \tag{29}$$

Reactions where the viscosity increases substantially with conversion, such as some polymerization reactions, can be particularly difficult to control upon scale-up.

***Jacket-Side Coefficient*** Here the calculations are shown for a jacket equipped with agitation nozzles that greatly increase the jacket fluid velocity. The jacket "swirl velocity" $v_j$ is calculated (iteratively) from the nonlinear algebraic relationship [17]

$$\dot{m}_n(v_n - v_j) = \left(\frac{4fL}{D_e}\right)\left(\frac{v_j^2}{2}\right)\rho A_f \tag{30}$$

where $\dot{m}_n$ the is the nozzle mass flow rate, $v_n$ is the nozzle velocity, the friction factor is

$$f = \frac{2 \times 0.023}{\mathrm{Re}^{0.2}} \tag{31}$$

the jacket-side film coefficient is

$$h_j = \frac{k_j}{D_e} 0.027\,\mathrm{Re}_j^{0.8}\,\mathrm{Pr}_j^{0.33} \tag{32}$$

and the Reynolds and Prandtl numbers are

$$\mathrm{Re}_j = \frac{D_e v_j \rho_j}{\mu_j} \tag{33}$$

$$\mathrm{Pr}_j = \frac{\mu_j c_{pj}}{k_j} \tag{34}$$

***Overall Coefficient*** The overall heat transfer coefficient is found from the sum of the resistances,

$$\frac{1}{U} = \frac{1}{h_i} + \frac{1}{h_j} + \frac{x_m}{k_m} + \frac{x_g}{k_g} + ff_i + ff_j \tag{35}$$

which includes reactor film, jacket film, vessel metal, vessel glass, and fouling factors for both the reactor and jacket sides.

### 4.2 Effect of Reactor Type, Jacket Heat Transfer Fluid, and Reactor Fluid Viscosity

Here we present examples of how the reactor type and heat transfer fluid affect the heat transfer coefficient. When the reactor fluid has a low viscosity, the dominant heat transfer resistance tends to be on the jacket side. When the reactor fluid has a high viscosity, however, the dominant resistance is typically on the reactor side. Parameter values for the studies are presented in Figures 5–7 and are given in the literature [18].

The overall heat transfer coefficient is much higher for an alloy reactor/half-pipe jacket than for a glass-lined carbon steel reactor/agitation nozzle jacket, as shown in Figure 5, where Syltherm is the heat transfer fluid. Syltherm has a significantly lower heat transfer coefficient than an ethylene glycol mixture, as shown in Figure 6, but is capable of operating over a wider range of temperatures. The reactor fluid viscosity has a tremendous effect on the overall heat transfer coefficient, as shown in Figure 7. This can be particularly important in polymerization reactions where viscosity increases with conversion.

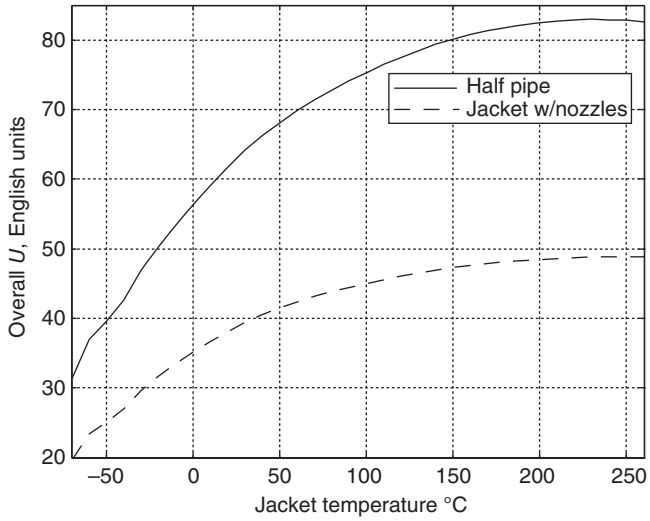


**FIGURE 5** Overall heat transfer coefficient for 500-gal reactors. Comparison of alloy half pipe with glass-lined carbon steel (GLCS). Syltherm is the heat transfer fluid. (From ref. 18, with permission.)

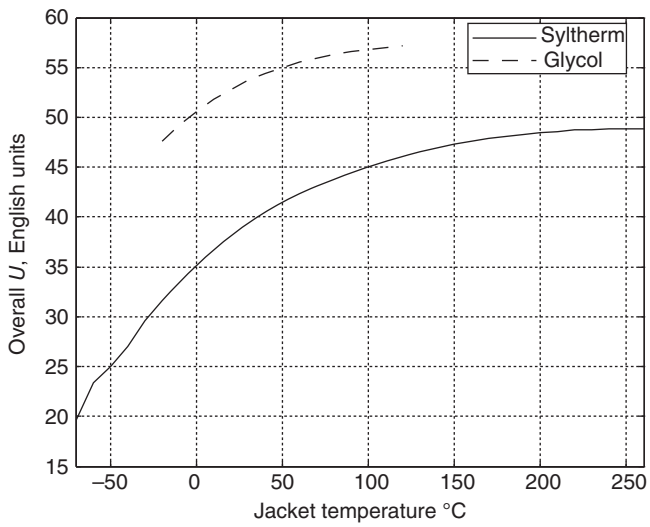


**FIGURE 6** Overall heat transfer coefficient for 500-gal GLCS reactor. Comparison of Syltherm with Glycol. (From ref. 18, with permission.)

### 4.3 Pilot- and Production-Scale Experiments

The relationships shown in Section 3 are also pertinent to large-scale reactors. By using different solvents and volumes of solvent, pilot and production reactor heat transfer characteristics can be determined from a series of experiments. A primary limitation, compared to reaction calorimeter characterization, is that a calibration probe is rarely available. Thus, heat-up and cool-down studies, performed by varying

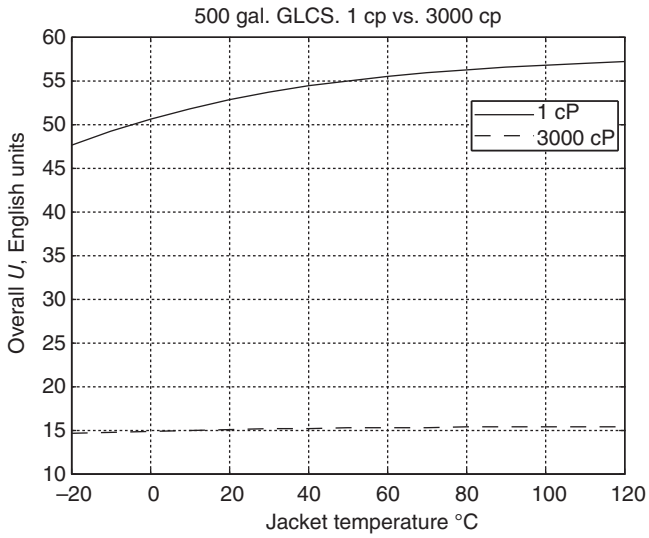


**FIGURE 7** Overall heat transfer coefficient for 500-gal GLCS reactor with glycol heat transfer fluid. Comparison of effect of reactor-side viscosity.

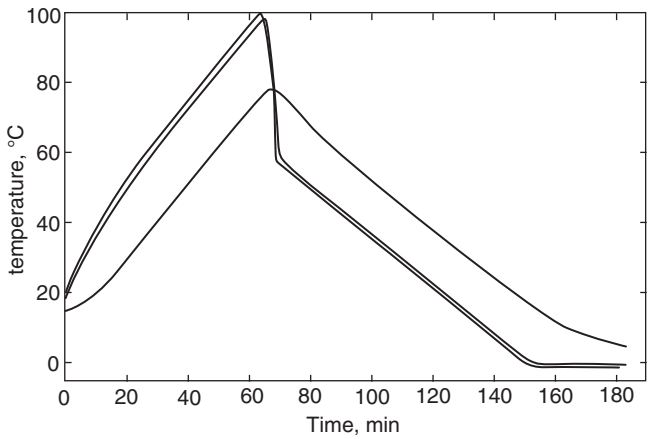


**FIGURE 8** Temperature profiles (jacket inlet, jacket outlet, and reactor) for a pilot plant reactor. (From ref. 19.)

the jacket temperature and observing the changes in the reactor temperature (for solvents with known heat capacity), are used to characterize the reactor. The inverse cooling time,

$$\frac{UA}{(mc_p)_r} = \frac{dT/dt}{T_j - T} \tag{36}$$

can be estimated from the temperature data collected from a heat-up/cool-down study. A characteristic example for a pilot-scale reactor is shown in Figure 8. The

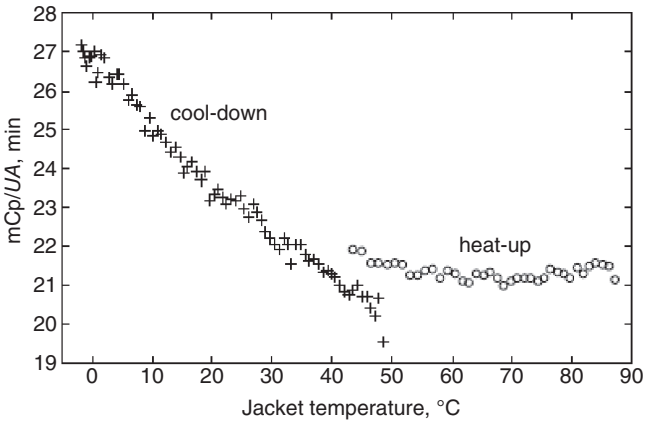


**FIGURE 9** Cooling time estimates based on data presented in Figure 8. (From ref. 19.)

resulting cooling time estimates are shown in Figure 9. Notice that the overall heat transfer coefficient is clearly a function of the jacket temperature. The reduced heat transfer at the lower jacket temperatures is due to the strong relationship between viscosity and temperature for the 40% glycol solution used in the jacket. The discontinuity in the cooling time estimate at around 45 °C may be due to two factors. One factor is the assumption of no heat loss from the vessel, which would tend to lower the *UA* estimates during the heat-up phase. Another factor is the assumption that the metal and glass inerts in the reactor are at the temperature of the reactor; in practice it might be a better assumption that the reactor wall in particular is at a temperature that is intermediate between the jacket and reactor temperatures.

The fluid and inert thermal masses can be independently estimated by conducting experiments with a number of different solvent amounts. From the cooling time expression

$$\frac{(mc_p)_r}{UA} = \frac{m_v c_{pv}}{UA} + \frac{V\rho c_p}{UA} \tag{37}$$

writing this as a function of the reactor fluid volume,

$$\frac{(mc_p)_r}{UA} = \frac{m_v c_{pv}}{UA} + \frac{\rho c_p}{UA} \cdot V \tag{38}$$

and conducting experiments at a number of different fluid volumes or, equivalently, masses ($V\rho$),

$$\frac{(mc_p)_r}{UA} = \frac{m_v c_{pv}}{UA} + \frac{c_p}{UA} \cdot V\rho \tag{39}$$

the linear regression can be used to find the slope and intercept and thus estimate the *UA* and $m_v c_{pv}$ terms [19]. This approach is shown in Figure 10 for a jacket tem-

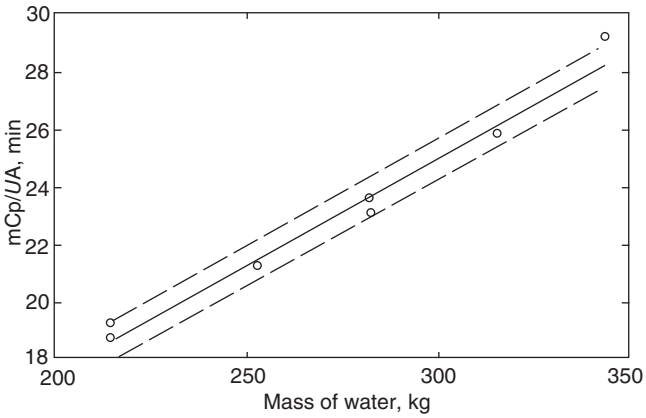


**FIGURE 10** Linear regression to estimate thermal mass and *UA*. (From ref. 19.)

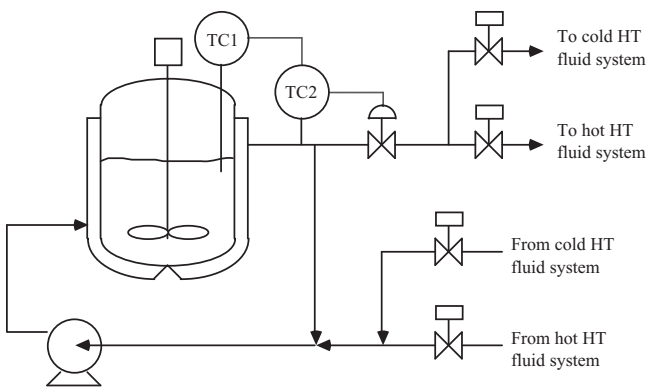


**FIGURE 11** Characteristic pilot plant vessel control strategy. Slave (secondary) controller based on jacket outlet temperature is shown. The control valve is on the outlet stream to minimize temperature gradients (when switching from hot to cold fluids) that would be imposed if the valve was on the inlet. (From ref. 18, with permission.)

perature of 60 °C (based on a total of eight experiments at five different reactor fluid volumes).

## 5 DYNAMIC SIMULATION STUDIES

Older pilot plant and manufacturing processes often used steam for heating and water for cooling, with a switch-and-purge strategy between the two modes. Recent process designs have two heat transfer fluid systems (hot and cold heat transfer fluids) that are used for most of the heating and cooling needs. In addition, some vessels may have nitrogen coolers for cryogenic operation.

A simplified schematic for a jacket heat transfer service is shown in Figure 11 [18]. Here, two separate heat transfer fluid headers are used, and the control valve is on the outlet stream to reduce the temperature shocks that might occur if a single

control valve was on the inlet stream. Depending on the range of temperatures, either ethylene glycol or a proprietary fluid such as Syltherm is used. Depending on whether heating or cooling is needed, either the hot or cold process control valve is open. Similarly, on–off valves return fluid to the appropriate distribution system.

Although the heat transfer fluid can be used over a wide range of temperatures, the film heat transfer coefficient is a strong function of temperature due to viscosity effects. The "cooling time" of a large reactor operating at a low temperature can be substantially longer than that of a small reactor operating at a high temperature due to this strong temperature effect. Simulation studies can be used to:

- Understand the effect of heat transfer fluid
- Understand possible performance limitations due to scale and operating conditions
- Test the effect of specified temperature gradient constraints
- Assist with controller design and selection of tuning parameters for system start-up

Various levels of models can be used to describe the behavior of pilot-scale jacketed batch reactors. For online reaction calorimetry and for rapid scale-up, a simple model characterizing the heat transfer from the reactor to the jacket can be used. Another level of modeling detail includes both the jacket and reactor dynamics. Finally, the complete set of equations simultaneously describing the integrated reactor/jacket and recirculating system dynamics can be used for feedback control system design and simulation. The complete model can more accurately assess the operability and safety of the pilot-scale system and can be used for more accurate process scale-up.

In the simulation studies that follow, it is assumed that the reactor and jacket are well mixed, resulting in differential equations for the material and energy balances [18]. The reactor shell (including a glass lining, if used) and reactor internals (agitator and baffles) are at the same temperature as the reactor, so their "thermal mass" is included in the reactor energy balance. Similarly, the jacket shell is at the jacket temperature, with an associated thermal mass. The heat transfer area $A$ is proportional to the reactor liquid level (between volumes associated with the minimum and maximum heat transfer area); also, the reactor shell thermal mass varies linearly with the liquid level. Heat transfer coefficients are calculated using the relationships presented in Section 4; see Garvin [20] or Dream [21] for detailed examples. Parameters, viscosity in particular, are a function of temperature.

We focus on the effect of reactor size and material of construction on the expected dynamic behavior of the reactors. Details on the model development and simulation environment are presented elsewhere [18]. Figure 12 illustrates that a vessel can have significantly different dynamic behavior depending on whether it is being heated or cooled (for illustrative purposes, the freezing point of water is neglected in this simulation). The increase in reactor temperature results in a much faster response than a decrease for two reasons: (i) the jacket heat transfer fluid has a much higher viscosity (resulting in a lower overall heat transfer coefficient) at low temperatures and (ii) the fluid flow rate/jacket temperature gain is proportional to

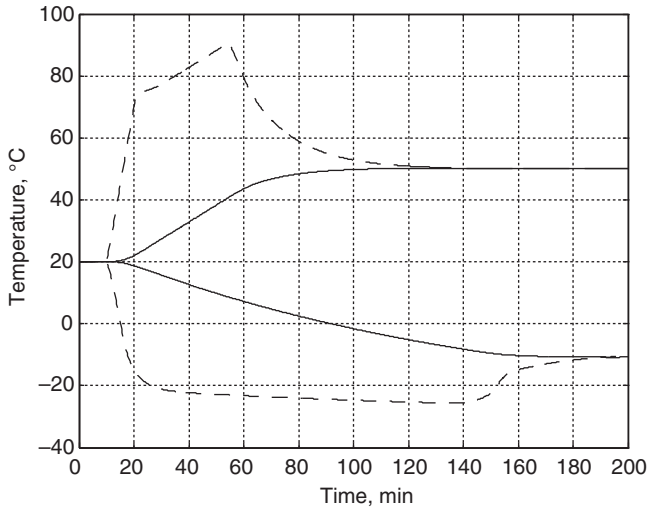


**FIGURE 12** Comparison of responses for ±30 °C reactor temperature setpoint changes at $t$ = 10 min; 500-gal GLCS filled with water (1925 kg).

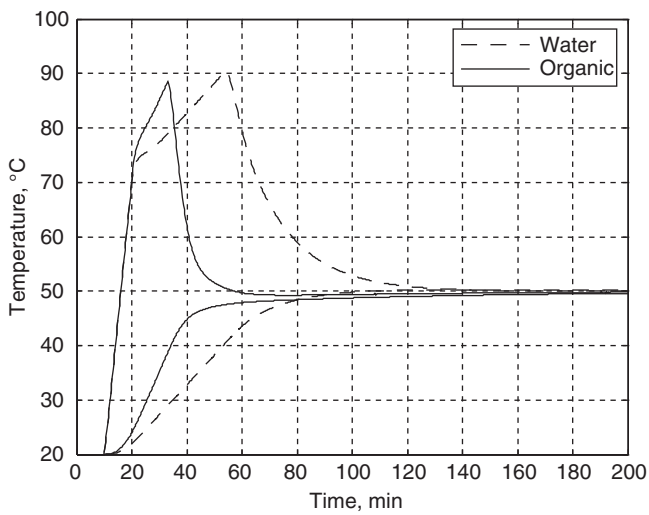


**FIGURE 13** Comparison of temperature responses for 30 °C batch setpoint change; 500-gal GLCS, water (1925 L) vs. organic (1700 L).

the difference between the jacket temperature and make-up fluid temperature, which becomes small at low jacket temperatures. Notice that the initial response for the temperature increase is constrained by the ramp limit of 5 °C/min on the jacket temperature. The temperature response of an organic solvent is much faster than water because of the heat capacity difference, as shown in Figure 13. The previous plots were for simple heating/cooling applications (ref. 18 presents further studies for cryogenic and semibatch systems).

## 6 SUMMARY

In this chapter we have presented an overview of scale-up considerations involved as one moves from bench-scale reaction calorimetry to larger scale pilot plant and production reactors. Our focus has been on heat transfer and single-phase processes, addressing primarily the problem that the heat transfer area per unit reactor volume decreases with scale. Clearly, there are many challenging problems associated with multiphase vessels, with evaporation/distillation and crystallization as obvious examples, but these topics are beyond the scope of this chapter.

## REFERENCES


1. Paul, E. L. (1988), Design of reaction systems for specialty organic chemicals, *Chem. Eng. Sci.*, 43(8), 1773–1782.
2. Anderson, N. G. (2000), *Practical Process Research and Development*, Academic, New York.
3. Pisano, G. P. (1997), *The Development Factory*, Harvard Business School, Boston.
4. Chylla, R. W., and Hasse, D. R. (1993), Temperature control of semi-batch polymerization reactors, *Comp. Chem. Eng.*, 17(3), 257–264.
5. Bonvin, D. (1998), Optimal operation of batch reactors—A personal view, *J. Proc. Cont.*, 8(5–6), 355–368.
6. LeLann, M. V., Cabassud, M., and Casamatta, G. (1999), Modeling, optimization and control of batch chemical reactors in fine chemical production, *Annu. Rev. Control*, 23, 25–34.
7. Gygax, R. W. (1990, Feb.), Scale-up principles for assessing thermal runaway risks, *Chem. Eng. Prog.*, 86(2), 53–60.
8. Shukla, P. K., and Pushpavanam, S. (1994), Parametric sensitivity, runaway, and safety in batch reactors: Experiments and models, *Ind. Eng. Chem. Res.*, 33(12), 3202–3208.
9. Caygill, G., Zanfir, M., and Gavrildis, A. (2006), Scalable reactor design for pharmaceuticals and fine chemicals production. 1: Potential scale-up obstacles, *Org. Proc. Res. Dev.*, 10(3), 539–552.
10. Regenass, W. (1985), Calorimetric monitoring of industrial chemical processes, *Thermochim. Acta*, 95, 351–369.
11. Kaarlsen, L. G., and Villadsen, J. (1987), Isothermal reaction calorimeters—I. A literature review, *Chem. Eng. Sci.*, 42(5), 1153–1164.
12. Kaarlsen, L. G., and Villadsen, J. (1987), Isothermal reaction calorimeters—II. Data treatment, *Chem. Eng. Sci.*, 42(5), 1165–1173.
13. Bourne, J. R., Buerli, M., and Regenass, W. (1981), Heat transfer and power measurements in stirred tanks using heat flow calorimetry, *Chem. Eng. Sci.*, 36, 347–354.
14. Landau, R. N., Blackmond, D. G., and Tung, H.-H. (1994), Calorimetric investigation of an exothermic reaction: Kinetic and heat flow modeling, *Ind. Eng. Chem. Res.*, 33, 814–820.
15. Landau, R. N. (1996), Expanding the role of reaction calorimetry, *Thermochim. Acta*, 289, 101–126.
16. Obenndip, D. A., and Sharratt, P. N. (2006), Towards an information-rich process development. Part I: Interfacing experimentation with qualitatitive/semiquantitative modeling, *Org. Proc. Res. Dev.*, 10(3), 430–440.

17. Bolliger, D. H. (1982), Assessing heat transfer in process-vessel jackets, *Chem. Eng.*, Sept. 20, 95–100.
18. Bequette, B. W., Holihan, S., and Bacher, S. (2004), Automation and control issues in the design of a pharmaceutical pilot plant, *Control Eng. Practice*, 12, 901–908.
19. Zima, A., Spencer, G., and Bequette, B. W. (1996), Model development for batch reactor calorimetry and control, Preprint, presented at the AIChE Annual Meeting, Chicago, IL, Nov. 1996.
20. Garvin, J. (1999), Understand the thermal design of jacketed vessels, *Chem. Eng. Prog.*, 95(6), 61–68.
21. Dream, R. F. (1999), Heat transfer in agitated jacketed vessels, *Chem. Eng.*, Jan., 90–96.

# PACKAGING AND LABELING

Maria Inês Rocha Miritello Santoro and Anil Kumar Singh
*University of São Paulo, São Paulo, Brazil*

**Contents**

1 Introduction
2 Packaging Materials
  2.1 General Considerations
  2.2 Glass as packaging material
  2.3 Plastic as Packaging Material
  2.4 Metal as Packaging Material
  2.5 Applications: Some Examples
3 Quality Control of Packaging Material
  3.1 General Considerations
  3.2 Packaging Components
  3.3 Inhalation Drug Products
  3.4 Drug Products for Injection and Ophthalmic Drug Products
  3.5 Liquid-Based Oral Products, Topical Drug Products, and Topical Delivery Systems
  3.6 Solid Oral Dosage Forms and Powders for Reconstitution
4 Importance of Proper Packaging and Labeling
5 Regulatory Aspects
  5.1 General Considerations
  5.2 Food, Drug and Cosmetic Act
  5.3 New Drugs
  5.4 Labeling Requisites
  5.5 Prescription Drugs
  5.6 Drug Information Leaflet
  5.7 Other Regulatory Federal Laws
  5.8 Fair Packaging and Labeling Act
  5.9 United States Pharmacopeia Center for the Advancement of Patient Safety
  5.10 National Agency of Sanitary Vigilance (ANVISA, Brazil)

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

5.11 International Committee on Harmonization (ICH)
5.12 European Union Regulatory Bodies
References

## 1 INTRODUCTION

The packaging of a pharmaceutical product fulfils a variety of roles, such as product presentation, identification, convenience, and protection until administration or use. Selection of packaging requires a basic knowledge of packaging materials, the environmental conditions to which the product will be exposed, and the characteristics of the formulation. Several types of packaging are used to contain and protect the pharmaceutical preparations, such as the primary packaging around the product and secondary packaging such as a carton and subsequent transit cases [1].

The principal objective of the modern pharmaceutical industry is to manufacture pharmaceutical preparations presenting high quality, identity, purity, effectiveness, and innocuity in order to guarantee the satisfaction and safety of patients. The development of a new drug must involve the synthesis of a molecule, determination of its pharmacological activity, industrial-scale production, and its commercialization to guarantee quality of the final product.

Packaging system development, including primary and secondary packaging components, is of critical importance. The material should be selected based on the characteristics of pharmaceutical product and dosage form. After the production phase, packaging must be planned according to regulatory requirements and its quality should be controlled according to the specifications.

Commercially, the packaging material is used as a barrier to protect the pharmaceutical preparations against external factors that can degrade them and consequently decrease their effectiveness and increase toxic effects.

Once the type of packaging material is decided based on such factors as size, shape, capacity, and physicochemical properties, all these data, including quality control tests, should be included in the specification of the products in order to assure the therapeutic effectiveness during its shelf life.

Several types of materials are in use in the preparation of containers and closure systems: glass, plastics, metals, and combinations of these materials. However, care should be taken in the selection of appropriate material. These materials should not present any physical or chemical reactivity that could modify drug activity, quality, purity, or physical characteristics of the drug and pharmaceutical preparations. Any minor modification in the pharmacopeial specification is acceptable if it does not present a threat to patient's health.

The aim of this chapter is to discuss the importance of the packaging and labeling of pharmaceutical preparations. The role of packaging and labeling in the pharmaceutical industry has grown substantially over the past decade. The total packaging operation is part of any drug development program. Pharmaceutical products generally require a standard of packaging which is superior to that of most other products in order to support and comply with their main requirements, such as efficacy, integrity, purity, safety, and stability.

For these reasons packaging technology should be based on the understanding of pharmaceutical products, characteristics of formulations, and dosage forms, including the physical and chemical properties of the drug substance.

In the past, packaging concerns often arose only during the later steps of product development. Today, packaging is integrated with the development step and is among the earliest considerations of new pharmaceutical preparations being studied. Labels of products can vary from the simple to the extremely complex. But, even at the most basic level, product identification should meet regulatory requirements. More complex are the labels that make use of bar code technologies. New components such as microchips, biosensors, and deoxyribonucleic acid (DNA) arrays are making possible the development of new technologies leading to finished products individually packed that require specialized packaging materials and design expertise. The challenge now is to maintain low packaging cost, that is, always integrated into the cost of the product itself.

Packaging in the post–World War II period benefited immensely from the commercialization of plastics, which were little known or used in prior years. Since then, the packaging industry has openly adopted plastics as a powerful new tool in the development of new packaging forms and functions.

Quality control of a packaging component starts at the design stage. All aspects of package development that may give rise to quality problems must be identified and minimized by good design. Identifying and correcting mistakes in packaging will avoid product recall and rejection of pharmaceutical preparations [2, 3].

## 2 PACKAGING MATERIALS

### 2.1 General Considerations

Packaging refers to all the operations, including filling and labeling, through which a bulk product should pass to become a finished product. Usually, sterile filling is not considered part of the packing process, although the bulk product is contained in a primary container.

A packaging component means any single part of a container closure system. Typical components are containers (e.g., ampules, vials, bottles), container liners (e.g., tube liners), closures (e.g., screw caps, stoppers), closure liners, stopper overseals, container inner seals, administration ports [e.g., on large-volume parenterals (LVPs)], overwraps, administration accessories, and container labels [4].

A primary packaging component is one that is or may be in direct contact with the dosage form. A secondary packaging component is one that is not and will not be in direct contact with the dosage form [4].

A container closure system refers to the sum of packaging components that together contain and protect the dosage form. This includes primary packaging components and secondary packaging components, if the latter are intended to provide additional protection to the drug product. A packaging system is equivalent to a container closure system [4].

The role of packaging material on the overall perceived and actual stability of the dosage form is well established. Packaging plays an important role in quality maintenance, and the resistance of packaging materials to moisture and light can

significantly affect the stability of drugs and their dosage forms. It is crucial that stability testing of dosage forms in their final packaging be performed. The primary role of packaging, other than its esthetic one, is to protect the dosage forms from moisture and oxygen present in the atmosphere, light, and other types of exposure, especially if these factors affect the overall quality of the product on long-term storage [5].

The compliance packaging such as for fixed-dose combination pills and unit dosage form packaging is a therapy-related intervention and is designed to facilitate medication regimens and so potentially improve adherence. Compliance packaging can be defined as a prepackaged unit that provides one treatment cycle of the medication, to both the pharmacist and the patient, in a ready-to-use package. This innovation type of packaging is usually based on blister packaging that contain unit therapeutic dose for one time use. The separate dosage units and separate days are usually indicated on the dosage cards to help remind the patient when and how much of the medication to take, for example, blister packed oral dosage forms with drug information leaflets and contraceptive pills [6, 7].

The selection of packaging material for any pharmaceutical product is as important as proper pharmaceutical dosage form. To guarantee the safe and adequate delivery of drug product to the patient and improve patient compliance, the manufacturer should consider the following factors:

1. Compatibility and safety concerns raised by the route of administration of the drug product and the nature of the dosage form (e.g., solid or liquid based)
2. Kinds of protection the container closure system should provide to the dosage form (e.g., photosensitive, hygroscopic, easily oxidized drug products)
3. Potential effect of any treatment or handling that may be unique to the drug product in the packaging system
4. Patient compliance to the treatment and ease of drug administration
5. Safety, efficacy, and quality of drug product throughout its shelf-life

The acquisition, handling, and quality control of primary and secondary packaging materials and of printed materials should be accomplished in the same way as that for the raw materials. The printed materials should be stocked in a reserved place so the possibility of unauthorized access is avoided. The labels and other rejected printed materials should be stored and transported with proper identification before being destroyed. There should be a destruction record of the printed materials. Each batch of printed material and packaging material should receive a specific reference number for identification.

The identification affixed on the containers, on the equipment, in the facilities, and on the product containers should be clear, without ambiguity, and in a format approved by the company and contain the necessary data. Besides the text, differentiated colors indicating its condition could be used (e.g., in quarantine, approved, rejected, and cleaned).

The packing materials should attend to the specifications, giving emphasis to the compatibility of the same with the pharmaceutical product that it contains. The material should be examined with relation to visible physical and critical defects as well as the required specifications.

## 2.2 Glass as Packaging Material

A packaging system found acceptable for one drug product may not be appropriate for another. Each application should contain enough information to show that each proposed container closure system and its components are suitable for the intended use.

### *Nonsterile Products*

*Solids* Some topical drug products such as powders may be considered for marketing in glass bottles with appropriate dispenser. These topical drug products may be sterile and could be subject to microbial limits.

The most common glass-packed solid oral dosage forms are oral powders and granules for reconstitution. A typical solid oral dosage form container closure system is a glass bottle (although plastic bottles are also used) with a screw-on or snap-off closure. A typical closure consists of a metal cap, often with a liner and frequently with an inner seal.

The dry powders that are reconstituted in their marketed container need not be sterile; however, the possibility of an interaction between the packaging components and the reconstituting fluid can't be discarded. Although the contact time will be relatively short when compared to the component/dosage form contact time for liquid-based oral dosage forms, it should still be taken into consideration when the compatibility and safety of the container closure system are being evaluated.

Powders for oral administration that are reconstituted in their market container, however, have an additional possibility of interaction between the packaging components and the reconstituting fluid. Although the contact time will be relatively short when compared to the component/dosage form contact time for liquid-based oral dosage forms, it should still be taken into consideration when the compatibility and safety of the container closure system are being evaluated.

*Nonsolids* For nonsterile products the preservative provides some protection, but continual microbial challenge will diminish the efficacy of the preservative, and spoilage or disease transmission may occur [8].

Antimicrobial preservatives such as phenylmercuric acetate are known to partition into rubbers during storage, thus reducing the formulation concentration below effective antimicrobial levels [9]. A complication of modern packaging is the need for the application of security seals to protect against deliberate adulteration and maintain consumer confidence.

***Sterile Products*** The sterile dosage forms share the common attributes that they are generally solutions, emulsions, or suspensions and are all required to be sterile. Injectable dosage forms represent one of the highest risk drug products (Table 1). Any contaminants present (as a result of contact with a packaging component or due to the packaging system's failure to provide adequate protection) can be rapidly and completely introduced into the patient's general circulation. Injectable drug products may be liquids in the form of solutions, emulsions, or suspensions or dry solids that are to be combined with an appropriate vehicle to yield a solution or suspension.

Although ophthalmic drug products can be considered topical products, they have been grouped here with injectables because they are required to be sterile and the descriptive, suitability, and quality control information is typically the same as that for an injectable drug product.

The potential effects of packaging component/dosage form interactions are numerous. Hemolytic effects may result from a decrease in tonicity and pyrogenic effects may result from the presence of impurities. The potency of the drug product or concentration of the antimicrobial preservatives may decrease due to adsorption or absorption.

A cosolvent system essential to the solubilization of a poorly soluble drug can also serve as a potent extractant of plastic additives.

A disposable syringe may be made of plastic, glass, rubber, and metal components, and such multicomponent construction provides a potential for interaction that is greater than when a container consists of a single material.

Injectable drug products require protection from microbial contamination (loss of sterility or added bioburden) and may also need to be protected from light or exposure to gases (e.g., oxygen).

Performance of a syringe is usually addressed by establishing the force to initiate and maintain plunger movement down the barrel and the capability of the syringe to deliver the labeled amount of the drug product.

*Solids* For solids that must be dissolved or dispersed in an appropriate diluent before being injected, the diluent may be in the same container closure system (e.g., a two-part vial) or be part of the same market package (e.g., a kit containing a vial of diluent).

Sterile powders or powders for injection may need to be protected from exposure to water vapor. For elastomeric components, data showing that a component meets the requirements of U.S. Pharmacopeia (USP) elastomeric closures for injections will typically be considered sufficient evidence of safety.

*Nonsolids* The package must prevent the entry of organisms; for example, packaging of sterile products must be absolutely microorganism proof—hence the continued use of glass ampules. Liquid injections are classified as small-volume parenterals (SVPs), if they have a solution volume of 100 mL or less, or as LVPs, if the solution volume exceeds 100 mL [10]. Liquid-based injectables may need to be protected from solvent loss.

An SVP may be packaged in a vial or an ampule. An LVP may be packaged in a vial, a glass bottle or, in some cases, as a disposable syringe. Packaging material for vials, and ampules are usually composed of type I or II glass. Stoppers and septa in cartridges, and vials are typically composed of elastomeric materials.

Pharmaceuticals may interact with packaging and containers, resulting in the loss of drug substances by adsorption onto and absorption into container components and the incorporation of container components into pharmaceuticals. Diazepam in intravenous fluid containers and administration sets exhibited a loss during storage due to adsorption onto glass [11, 12].

Glass surfaces are also known to adsorb drug substances. Chloroquine solutions in glass containers decreased in concentration owing to adsorption of the drug onto the glass [13].

Rubber closures are also known to absorb materials, including drugs. Absorption of preservatives such as chlorocresol into the rubber closures of injectable formulations has been studied extensively [13].

The water permeability of rubber closures used in injection vials is considered an important parameter in assessing the closures, but quantitative prediction of water permeability through rubber closures is difficult because the diffusion coefficient of water is dependent on relative humidity [14].

Liquid-based oral drug products are usually dispensed in glass bottles (sometimes in plastic), often with a screw cap with a liner, and possibly with a tamper-resistant seal or an overcap that is welded onto the bottle. The same cap liners and inner seals are sometimes used with solid oral dosage forms. A laminated material can be used to overwrap glass bottles for extra safety.

The USP-grade glass packaging components are chemically resistant and can be considered sufficient evidence of safety and compatibility. In some cases (e.g., for some chelating agents), a glass packaging component may need to meet additional criteria to ensure the absence of significant interactions between the packaging component and the dosage form.

Several ophthalmic preparations are commercialized in glass containers. Although the risk factors associated with ophthalmic preparations are generally considered to be lower than for injectables, any potential for causing harm to the eyes demands caution.

A large-volume intraocular solution (for irrigation) may be packaged in a polyolefin (polyethylene and/or polypropylene) container.

The liquid-based oral dosage forms may be marketed in multiple-unit bottles. The dosage form may be used as is or admixed first with a compatible diluent or dispersant. Liquid-based oral drug products in glass container must meet the requirements for USP containers. Glass containers are accepted as sufficient evidence of safety and compatibility. Performance is typically not a factor for liquid-based oral drug products but should be considered while treating pressurized liquid-based oral drug products (e.g., elixir spray).

Topical dosage forms such as unpressurized sprays, lotions, ointments, solutions, and suspensions may be considered for marketing in glass bottles with appropriate dispenser. Some topical drug products, especially ophthalmic, are sterile or may be subject to microbial limits. In these cases, packaging material and handling should be done as those for injectables.

### 2.3 Plastic as Packaging Material

For plastic components, data from USP biological reactivity tests will typically be considered sufficient evidence of safety. Whenever possible, extraction studies should be performed using the drug product. If the extraction properties of the drug product vehicle may reasonably be expected to differ from that of water (e.g., due to high or low pH or to a solubilizing excipient), then drug product should be used as the extracting medium. If the drug substance significantly affects extraction characteristics, it may be necessary to perform the extractions using the drug product vehicle. If the total extract significantly exceeds the amount obtained from water extraction, then an extraction profile should be obtained. It may be advisable to obtain a quantitative extraction profile of an elastomeric or plastic packaging

component and to compare this periodically to the profile from a new batch of the packaging component. Extractables should be identified whenever possible.

### *Nonsterile Products*

*Solids* The most common solid oral dosage forms are capsules and tablets. A typical solid oral dosage forms container closure system is a plastic, usually high-density polyethylene (HDPE), bottle with a screw-on or snap-off closure and a flexible packaging system such as a pouch or a blister package. A typical closure consists of a cap, often with a liner, frequently with an inner seal. If used, fillers, desiccants, and other absorbent materials are considered primary packaging components.

A change in the selection of packing materials combined with a change in storage conditions or conditions during administration of the drug products may provoke stability problems.

Many studies have been conducted on predicting the role of packaging in moisture adsorption by dosage forms. Adsorption of moisture by tablets contained in polypropylene films was successfully modeled from storage temperature and the difference in water vapor pressure between the inside and outside of the packaging [15].

Chemical and physical degradation of packaged dosage forms caused by moisture adsorption has been predicted from the moisture permeability of the packaging. For example, strength changes of lactose–corn starch tablets in strip packaging [16] and discoloration of sugar-coated tablets of ascorbic acid [17, 18] were predicted using the moisture permeability coefficient of the packaging.

Typical flexible forms of packaging containing solid oral dosage forms are the blister package and the pouch. A blister package usually consists of a lidding material and a forming film. The lidding material is usually a laminate which includes a barrier layer (e.g., aluminum foil) with a print primer on one side and a sealing agent (e.g., a heat-sealing lacquer) on the other side.

The sealing agent contacts the dosage form and the forming film. The forming film may be a single film, a coated film, or a laminate. A pouch typically consists of film or laminate which is sealed at the edges by heat or adhesive.

Solid oral dosage forms generally need to be protected from the potential adverse effects of the following:

1. Water vapor (e.g., moisture may affect the decomposition rate of the active drug substance or the dissolution rate of the dosage form)
2. Incident light (e.g., in case of photosensitive products)
3. Reactive gases (e.g., oxygen could provoke oxidative reactions)

Carefully selected packaging material may help protect drug products. For example, a blister or pouch and use of secondary packing may be used to protect pack photosensitive material, especially when a dark polymeric film with a covering lid made of aluminum is used for blister packing. Blister packaging using multilayer HDPE material and selection of an adequate sealing technique may help prevent moisture in the blister system. However, plastics and glass for packaging of solid oral dosage forms and for powders for reconstitution should meet the requirements

of the USP container test. Incorporating oxygen adsorbents such as iron powder into packaging units can reduce the effect of oxygen. Protection from light can be achieved using primary packaging (packaging that is in direct contact with the dosage forms) and secondary packaging made of light-resistant materials. May be involved in the photolytic degradation kinetics. The velocity of the photochemical reaction may be affected not only by the light source, intensity, and wavelength of the light but also by the size, shape, composition, and color of the container.

Great effort should be taken to stabilize a formulation in such a way that the shelf life becomes independent of the storage conditions. The photostability of drugs and excipients should be evaluated at the formulation development stage in order to assess the effects of packaging on the stability of the final product. Molsidomine tablet preparations in inadequate primary containers (blister) without secondary containers when exposed to irradiation may produce morpholine. These results illustrate the importance of packaging for the stability of molsidomine [19].

Three standard tests for water vapor permeation have been established by the USP for use with solid oral dosage forms.

1. Polyethylene containers (USP ⟨661⟩) [10]
2. Single-unit containers and unit-dose containers for capsules and tablets (USP ⟨671⟩)
3. Multiple-unit containers for capsules and tablets (USP ⟨671⟩) [10]

The cotton and rayon used as fillers in solid oral dosage form containers may not meet pharmacopeial standards, but through appropriate tests and acceptance criteria for identification and moisture content, their adequacy should be shown. For example, rayon has been found to be a potential source of dissolution problems for gelatin capsules and gelatin-coated tablets.

Desiccants are often used to eliminate moisture in packaging when the moisture resistance of the packaging is not sufficient to prevent exposure. The utility of desiccants has been assessed based on a sorption–desorption moisture transfer model [20].

Desiccants or other absorbent materials are primary packaging component. The components should differ in shape and/or size from the tablets or capsules with which they are packaged. Their composition should be provided and their inertness should be proved through appropriate tests, and acceptance criteria should be established.

A topical powder product may be marketed in a sifter-top container made of flexible plastic tubes or as part of a sterile dressing (e.g., antibacterial product). The topical formulations in a collapsible tube can be constructed from low-density polyethylene (LDPE), with or without a laminated material. Normally, there is no product contact with the cap during storage. Thus usually there is no cap liner, especially in collapsible polypropylene screw caps. Normally separate applicator devices are made from LDPE. Product contact is possible if the applicator is part of the closure, and therefore an applicator's compatibility with the drug product should be established, as appropriate (e.g., vaginal applicators).

*Nonsolids* Typical liquid-based oral dosage forms are elixirs, emulsions, extracts, fluid extracts, solutions, gels, syrups, spirits, tinctures, aromatic waters, and suspen-

sions. These products are usually nonsterile but typically need to be protected from solvent loss, microbial contamination, and sometimes exposure to light or reactive gases (e.g., oxygen).

The presence of a liquid phase implies a significant potential for the transfer of materials from a packaging component into the dosage form.

The higher viscosity of semisolid dosage forms and transdermal systems may cause the rate of migration of leachable substances into these dosage forms to be slower than for aqueous solutions. Due to extended contact, the amount of leachables in these drug products may depend more on a leachable material's affinity for the liquid/semisolid phase than on the rate of migration.

In addition to absorption onto and absorption into containers, transfer of container components into pharmaceuticals may affect the perceived stability/quality of drug dosage forms. Adsorption of volatile components from rubber closures onto freeze-dried parenterals during both dosage form processing and storage brought about haze formation upon reconstitution [21–23]. Leaching of dioctyl phthalate, a plasticizer used especially in polyvingl chloride (PVC) plastics, into intravenous solutions containing surfactants was observed [24, 25]. Plastics contain additives to enhance polymer performance. PVC may contain phthalate diester plasticizer, which can leach into infusion fluids from packaging [26].

The liquid-based oral dosage forms may be marketed in multiple-unit bottles or in unit-dose or single-use pouches or cups. The dosage form may be used as is or admixed first with a compatible diluent or dispersant. A liquid-based oral drug pouch may be a single-layer plastic or a laminated material. The pouches may use an overwrap, which is usually a laminated material.

For LDPE components, data from USP container tests are typically considered sufficient evidence of compatibility. The USP general chapters do not specifically address safety for polyethylene (HDPE or LDPE), polypropylene (PP), or laminate components.

In such cases, an appropriate reference to the indirect food additive regulations [27] is typically considered sufficient. This reference is considered valid for liquid-based oral dosage forms which the patient will take only for a relatively short time.

For liquid-based oral drug products which the patient will continue to take for an extended period, that is, months or years, and is expected to extract greater amounts of substances from plastic packaging components than from water (presence of cosolvents), additional extractable information may be needed to address safety issues.

Topical dosage forms such as creams, emulsions, gels, lotions, ointments, pastes, and powders may be marketed in plastic materials. Topical dosage formulations are for local (not systemic) effect and are generally applied to the skin or oral mucosal surfaces. Some vaginal and rectal creams and nasal, otic, and ophthalmic solutions may be considered for topical drug products.

A rigid bottle, a collapsible tube, or a flexible pouch made of plastic material may be used for liquid-based topical product. These preparations are marketed in a single- or multiple-unit container. For example, dissolved drug (or any substance, e.g., benzocaine) may diffuse in the suppository base and can, for instance, partition into polyethylene linings of the suppository wrap.

Topical delivery systems are self-contained, discrete dosage forms that are designed to deliver drug for an extended period via intact skin or body surface, for example, transdermal, ocular, and intrauterine.

These systems may be constructed of a plastic or polymeric material loaded with active ingredients or a coated metal. Each of these systems is generally marketed in a single-unit soft blister pack or a preformed tray with a preformed cover or overwrap. The compatibility and safety for topical delivery systems are addressed in the same manner as for topical drug products. Performance and quality control should be addressed for the rate-controlling membrane.

### *Sterile Products*

*Nonsolids* An SVP may be packaged in a disposable cartridge, a disposable syringe, or a flexible bag made of polymeric plastic. Flexible bags are typically constructed with multilayered plastic (Table 2).

An LVP may be packaged in a vial, a flexible bag, or, in some cases, a disposable syringe. Packaging material for cartridges, syringes, vials, and ampules are usually composed of polypropylene (Table 2).

Stoppers and septa in cartridges and syringes are typically composed of elastomeric materials. An overwrap may be used with flexible bags to retard solvent loss and to protect the flexible packaging system from rough handling.

Diazepam in intravenous fluid containers and administration sets exhibited a loss during storage due to adsorption onto and absorption into plastics [11, 12].

Absorption of clomethiazole edisylate and thiopental sodium into PVC infusion bags was observed [28].

The pH dependence of adsorption/absorption of acidic drug substances such as warfarin and thiopental and basic drug substances such as chlorpromazine and diltiazem indicates that only the un-ionized form of the drug substance is adsorbed onto or absorbed into PVC infusion bags [29].

The absorption was correlated to the octanol–water partition coefficients of the drugs, suggesting that prediction of absorption from partition data is possible [30, 31]. Polymers such as Nylon-6 (polycaprolactam) are known to adsorb drug substances such as benzocaine [32].

The ophthalmic drug products are usually solutions marketed in a LDPE bottle with a dropper built into the neck. A few solution products use a glass container due to stability concerns regarding plastic packaging components.

## 2.4 Metal as Packaging Material

Metal tubes constructed of a single material are the packaging material of choice for topical dosage forms and may be tested readily for stability with a product. Tubes with a coating, however, present additional problems. The inertness of coating material must be established through adequate tests and guarantee that it completely covers underlying material. The coating material must be resistant to creaking, leaking, leaching, and solvent erosion. For example, frequently used aluminum tubes have demonstrated reactivity with fatty alcohol emulsions, mercurial compounds, and preparations with pH below 6.5 and above 8.0. Nonreactive, epoxy linings have been found to make aluminum tubes resistant to attack [6].

**TABLE 1 Plastic Additives**

| Type | Purpose | Examples |
|---|---|---|
| Lubricants | Improve processability | Stearic acid paraffin waxes, polyethylene (PE) waxes |
| Stabilizers | Retard degradation | Epoxy compounds, organotins, mixed metals |
| Plasticizers | Enhance flexibility, resiliency, melt flow | Phthalates |
| Antioxidants | Prevent oxidative degradation | Hindered phenolics (BHT), aromatic amines, thioesters, phosphites |
| Antistatic agents | Minimize surface static charge | Quaternary ammonium compounds |
| Slip agents | Minimize coefficient of friction, especially polyolefins | |
| Dyes, pigments | Color additives | |

*Source*: From ref. 6.

**TABLE 2 Parenteral Drug Administration Devices**

| Sterile Device | Plastic Material |
|---|---|
| Containers for blood products | Polyvinyl chloride |
| Disposable syringes | Polycarbonate, polyethylene, polypropylene |
| Irrigating solution containers | Polyethylene, polypropylene, polyvinyl chloride |
| Intravenous infusion fluid containers | Polyethylene, polypropylene, polyvinyl chloride |
| Administration sets | Nylon (spike), polyvinyl chloride (tubing), polymethylmethacrylate (needle adapter), polypropylene (clamp) |
| Catheter | Teflon, polypropylene, thermoplastic elastomers |

*Source*: From ref. 6.

Some examples of plastic additives and parenteral drug administration devices used as packaging materials for sterile products can be seen in Tables 1 and 2.

Ophthalmic ointments are marketed in a metal tube with an ophthalmic tip. Ophthalmic ointments that are reactive toward metal may be packaged in a tube lined with an epoxy or vinyl plastic coating.

Metal containers, pressurized or not, may also be used for topical drug products. Topical dosage forms include aerosols, emulsions, gels, powders, and solutions and may be marketed in metallic flasks, pressurized or not. Topical dosage formulations are for local (not systemic) effect and are generally applied to the skin or oral mucosal surfaces. Some vaginal and rectal creams and nasal and otic spray drug products may be considered for marketing in metallic containers for topical use.

A number of topical products marketed as a pressurized aerosol may be dispensed in a metallic bottle with a screw cap. Topical dosage forms in aluminum tubes usually include a liner. A tube liner is frequently a lacquer or shellac whose composition should be stated. A metallic pressurized packaging system for a liquid-

based topical product may deter solvent loss and may provide protection from light when appropriate.

The droplet size of topical aerosol sprays does not need to be carefully controlled, and the dose usually is not metered as in inhalers. The spray may be used to apply the drug to the skin (topical aerosol) or mouth (lingual aerosol) and the functionality of the sprayer should be addressed. The drug product has no contact with the cap and short-term contact with the nozzle. A topical aerosol may be sterile or may conform to acceptance criteria for microbial limits. However, the physical stability of aerosols can lead to changes in total drug delivered per dose and total number of doses that may be obtained from the container.

### 2.5 Applications: Some Examples

Many research papers in the scientific literature present studies showing the importance of the effect of packaging materials on the stability of pharmaceutical and cosmetic preparations:

1. Santoro and co-workers [33] presented results of the stability of oral rehydration salts (ORSs) in different types of packaging materials. The objective of the research was to give guidance on the adequate choice of packaging material presenting the indispensable characteristics in order to protect ORS preparation. This pharmaceutical preparation is essential to children living in developing countries with tropical climate and its distribution is one of the programs of the World Health Organization (WHO) [34].

It has been proved in several research papers that water is the most important factor in the component's degradation of ORSs. To proceed with the study, the pharmaceutical formulation was prepared by a pharmaceutical manufacture. The batch was packed in six types of packaging material. After storage of samples for 36 weeks maintained at ambient temperature, at ambient temperature and 76% relative humidity, and at 40°C with 80% relative humidity, analyses of water determination were made at different intervals of time. Water determination was performed by loss on drying at 50°C and Karl Fisher methods.

The studied ORS preparation contained anhydrous glucose (20 g), sodium chloride (3.5 g), trisodium dehydrate citrate (2.9 g), and potassium chloride (1.5 g)

According to the results, the packaging material that better protected the ORS preparation is the one constituted of polyester (18 g), aluminum (35%), and polyethylene (50 g).

2. The effect of packaging materials on the stability of sunscreen emulsions was also studied by Santoro and co-workers [35, 36]. The purpose of the research was to study the stability of an emulsion containing UVA, UVB, and infrared sunscreens after storage in different types of packaging materials (glass and plastic flasks, plastic and metallic tubes). The samples, emulsions containing benzophenone-3 (B-3), octyl methoxycinnamate (OM), and Phycocorail , were stored at 10, 25, 35, and 45°C and representative samples were analyzed after 2, 7, 30, 60, and 90 days. Stability studies were conducted by analyzing samples at predetermined intervals by high-performance liquid chromatography (HPLC) along with periodic rheological measurements.

The proposed HPLC method enabled the separation and quantitative determination of B-3 and OM present in sunscreens. The method was successfully applied in

the stability studies of the emulsions. The method is simple, precise, and accurate; there was no interference from formulation components. The sample emulsions stored at different temperatures presented similar rheological behavior, at least during the period of the study (three months). Most of the samples showed a pseudoplastic non-Newtonian thixotropic profile. There were no significant changes in the physical and chemical stability of emulsions stored in different packaging material. The studied glass and plastic packaging materials were found adequate for storing referred solar protector emulsions.

3. Sarbach and co-workers [48], studied the effect of plastics packaging materials on parenteral pharmaceutical products. Compatibility studies of these containers with different contents are required for drug registration. The authors demonstrated the migration phenomena which occurred between a trilaminated film and a parenteral solution of metronidazole at 0.5%. The main migration products found in the solution were e-caprolactam and a phthalic derivative. The authors also separated several unidentified compounds probably coming from the polyurethane adhesive.

4. Molsidomine is sensitive to light and shows a fast decomposition in solutions and in tablets. Thoma and co-workers [37] showed the importance of light-resistant packaging material for photolabile pharmaceuticals. They irradiated molsidomine preparations over a period of 72 h in a light cabinet according to storage at daylight for about 4–6 weeks. Losses of 23–90% in tablets and 43–60% in solutions were found. The photodegradation could be overcome by selection of suitable packaging materials, colorants or vanillin. The degradation product morpholine after dansylation was determined by HPLC and showed contents of 0.10–0.67 mg in tablets and 0.10–0.38 mg/mL in solution after irradiation.

These examples, among many others described in the scientific literature, illustrate the importance of proper selection of packaging material for the stability and effectiveness of pharmaceutical dosage forms.

## 3 QUALITY CONTROL OF PACKAGING MATERIAL

### 3.1 General Considerations

Several regulatory agencies as well as private agencies [Food and Drug Administration (FDA), British Pharmacopoeia, WHO, USP] [4, 10, 34, 38] have issued guidelines on the safety evaluation of materials and container closure systems. However, the ultimate proof of the safety and suitability of a container closure system and the packaging process is established by full shelf life stability studies. An important step in such evaluations is characterization of the packaging materials and the chemicals that can migrate or extract from container closure system components to the drug product. This extractable material belongs to diverse chemical classes that can migrate from polymeric materials, such as antioxidants, contaminants, lubricants, monomers, plasticizers, and preservatives. Such basic information is critical to understanding the biological safety and suitability of a container.

Establishing the safety of container closure systems is of key importance to the medical and pharmaceutical industries (Table 3). It is no less important than the contents themselves. The FDA's document "Guidance on Container Closure Systems for Packaging Human Drugs and Biologics" makes this point clear [4].

**TABLE 3 Examples of Packaging Concerns for Common Classes of Drug Products**

| Degree of Concern Associated with Route of Administration | Likelihood of Packaging Component–Dosage Form Interaction | | |
|---|---|---|---|
| | High | Medium | Low |
| Highest | Inhalation aerosols and solutions; injection; injectable suspensions | Sterile powders and powders for injections and inhalation powders | |
| High | Ophthalmic solutions and suspensions; transdermal ointments and patches; nasal aerosols and sprays | | |
| Low | Topical solutions and suspensions; topical and lingual aerosols; oral solutions and suspensions | Topical powders; oral powders | Oral tablets and oral (hard and soft; gelatin) capsules |

The FDA's guidance document requires the evaluation of four attributes to establish suitability: protection, compatibility, safety, and performance/drug delivery. The document also provides a structured approach to ranking packaging concerns according to the route of drug administration and likelihood of packaging component–dosage form interaction. A container closure system acceptable for one drug product cannot be assumed to be appropriate for another. Each product should have sufficient information to establish that a container and its components are right for their intended use [4].

To establish suitability, all four attributes must be evaluated and be shown to pose no concern to the drug product or to product performance. Suitability refers to the tests used for the initial qualification of the container closure system with regard to its intended use. The guidance defines what tests must be done to evaluate each of the attributes of suitability.

While the tests and methods described in Table 4 allow one to provide data that the container closure system is suitable for its intended use, an application must also describe the quality control (QC) measures that will be used to ensure consistency in the packaging components. The principal considerations for the QC measures are the physical characteristics and the chemical composition. By choosing two or three of the tests done in the initial suitability study, a QC program can be established that will ensure the consistency of the container closure system (Table 4).

***Protection*** A container closure system should provide the dosage form with adequate protection from factors (e.g., temperature, light) that can cause a degradation in the quality of that dosage form over its shelf life. Common causes of such degradation are exposure to light, loss of solvent, exposure to reactive gases (e.g., oxygen), absorption of water vapor, and microbial contamination.

A container intended to provide protection from light or offered as a light-resistant container must meet the requirements of the USP ⟨661⟩ light transmission test. The procedure requires the use of a spectrophotometer, with the required sensitivity

and accuracy, adapted for measuring the amount of light transmitted by the plastic material used for the container.

The ability of a container closure system to protect against moisture can be ascertained by performing the USP ⟨661⟩ water vapor permeation test. The USP sets limits on the amount of moisture that can penetrate based upon the size and composition of the plastic components [HDPE, LDPE, or polyethylene terephthalate (PET)].

Evaluating the integrity of the container can be done in several ways. Two of the most common tests are dye penetration and microbial ingress. Container closure systems stored in a dye solution and exposed to pressure and vacuum cycles are examined for dye leakage into the container. The microbial ingress is similar in fashion but determines the microbial contamination of the contents when soaked in a media contaminated with bacteria. Other quantitative tests that can be run are vacuum/pressure decay, helium mass spectrometry, and gas detection.

***Compatibility*** Packaging components that are compatible with a dosage form will not interact sufficiently to cause unacceptable changes in the quality of either the dosage form or the packaging component. A leachability study designed to evaluate the amount and/or nature of any chemical migrating from the plastic material to the drug product should be considered. The study should evaluate substances that migrate into the drug product vehicle for the length of shelf life. The drug product should be evaluated at regular intervals, such as at one, three, or six months or one or two years, until the length of the shelf-life claim has been met.

Analytical techniques such as liquid chromatography/mass spectrometry (LC/MS) to evaluate nonvolatile organics, gas chromatography/mass spectrometry (GC/MS) to evaluate semivolatile organics, and inductively coupled plasma (ICP) spectroscopy to detect and quantitate inorganic elements should be a part of this study.

Unknown impurities and degradation products can be identified using liquid or gas chromatography along with MS. Information or substances identified from extractable chemical evaluation can be used to help prepare standards specific for the plastic container being studied during leachability studies. Development and validation of the selective analytical method should be thoroughly studied before its application in the detection of leachable chemicals in active drug substance and drug product.

Organoleptic and chemical changes such as precipitates, discoloration, strange odor, and pH modification are signs of degradation of drug product. Changes in the physical characteristics of the container, such as brittleness, should be evaluated using thermal analysis and hardness testing. An infrared spectroscopic scan can fingerprint the materials and also provide proof of identity. Spectrophotometry and LC with ultraviolet detection can be used for the analysis of drug product stored at different stress conditions. These tests can be used for the quality control of drug product as well as for conducting stability studies on different products stored in the same container material.

***Safety*** Packaging components should be constructed of materials that will not leach harmful or undesirable amounts of substances to which a patient will be exposed when being treated with the drug product. This consideration is especially

**TABLE 4 Properties of Suitability Concerns and Interactions**

| Attributes | Concerns and Interactions | Proposed Methods |
|---|---|---|
| Protection | Exposure to light, moisture, microbial ingress, and oxidation from presence of oxygen | USP⟨661⟩ light transmission and water vapor permeation, container integrity (microbial ingress, dye penetration, helium leak) |
| Compatibility | Leachable induced degradation, absorption or adsorption of drug, precipitation, change in pH, discoloration, brittleness of packaging materials | Leachability study (migration of chemicals into drug product) using LC/MS, GC/MS, ICP/AA, pH, appearance of drug and container, thermal analysis (DSC, TGA), and infrared (IR) |
| Safety | No leached harmful or undesirable amounts of substances to expose patients treated with drug | Extraction study (USP physicochemical tests–plastics), USP elastomeric closures for injections, toxicological evaluation, USP biological reactivity and complies with CFR, additives and purity |
| Performance | Container closure system functionality, drug delivery | Functionality (improved patient compliance or use), delivery (transfer dose in right amount or rate) |

*Abbreviations*: DSC, differential scanning calorimetry; ICP, Inductively coupled plasma spectrometer; AA, Atomic absorption.
*Source*: From ref. 39.

important for those packaging components which may be in direct contact with the dosage form, but it is also applicable to any component from which substances may migrate into the dosage form (e.g., an ink or adhesive). Determining the safety of a packaging component is not a simple process, and a standardized approach has not been established. However, an extraction study should be one of the first considerations.

A good knowledge regarding possible extractable material could help analysts develop specific and selective methods to identify extractables from container closure components under various control extraction study conditions.

Precise information on the synthesis of the polymer and descriptions of the monomers used in the polymerization, the solvents used in the synthesis, and the special additives that have been added during material production as well as knowledge of degradation products that may be released into the drug product are also important.

Some potential extractable chemicals from packaging materials are water soluble, while others are soluble only in nonpolar environments. The USP includes physicochemical tests for plastics based on water extracts, while water, alcohol, and hexane extracts are required for polyethylene containers under controlled temperature and time parameters (70°C for 24 h for water and alcohol and 50°C for 24 h for hexane).

The USP physicochemical tests for extractables should be a part of all suitability programs, regardless of the criticality of the drug dosage form. USP elastomeric closures for injections should also be a part of the extractables study to establish safety. These USP tests, which have evolved over many years, are relevant, sensitive, rapid, and inexpensive. They help establish material safety.

The safety of material can be guaranteed by using appropriate analytical methods and instrumentation to identify and quantitate extracted chemicals. Liquid and gas chromatography and MS are powerful analytical tools that can separate and quantitate volatile and nonvolatile chemicals along with useful structural information. The mass spectrum or fragmentation pattern acquired for each molecule makes these excellent and effective tools for identifying unknown impurities or degradation products.

Toxicological evaluation of identified and unidentified impurities from a container can help improve the safety index of drug products. The toxicological evaluation should take into consideration container closure system properties, drug product formulation, dosage form, route of administration, and dose regimen. A close correlation between chemical and toxicological information can provide better control on safety and compatibility of containers and closures.

***Performance*** The fourth attribute of the suitability of the container closure system, performance and drug delivery, refers to its ability to function in the manner for which it was designed. There are two major considerations when evaluating performance. The first consideration is functionality that may improve patient compliance, [e.g., a two-chamber vial or intravenous (IV) bag], or improve ease of use (e.g., a cap that contains a counter, a prefilled syringe). The second consideration is drug delivery, which is the ability of the packaging system to deliver the right amount or rate (e.g., a prefilled syringe, a transdermal patch, a metered tube, a dropper or spray bottle, a dry-powder inhaler, and a metered-dose inhaler).

### 3.2 Packaging Components

Quality control refers to the tests typically used and accepted to establish that, after the application is approved, the components and the container closure system continue to possess the characteristics established in the suitability studies.

To ensure consistency, protection, compatibility, safety, and performance of the packaging components, it is necessary to define QC measures that will be used to ensure consistency in the packaging components. These controls are intended to limit unintended postapproval variations in the manufacturing procedures or materials of construction for a packaging component and to prevent adverse effects on the quality of a dosage form.

The USP tests and studies for establishing suitability and QC of container closure system and for associated component materials are summarized in Table 5.

Hydrolysis and oxidation are the two main routes of degradation for the majority of drugs. To gain more information, the drug could be subjected to a range of temperature and relative humidity conditions. In addition, photostability studies could be conducted by exposure to artificial or natural light conditions. Elevated temperature, humidity, and light stress the drug and induce rapid degradation. Harmonized guidelines are available for new drug substances and products and may provide useful information to characterize degradation processes and selection of appropriate packaging material.

The primary packaging must physically protect the product from the mechanical stresses of warehousing, handling, and distribution. Mechanical stress may take a

**TABLE 5 U.S. Pharmacopeia General Tests and Assays**

| Chapter | Topic |
|---|---|
| ⟨1⟩ | Injections |
| ⟨51⟩ | Antimicrobial preservatives—effectiveness |
| ⟨61⟩ | Microbial limit tests |
| ⟨71⟩ | Sterility tests |
| ⟨87⟩ | Biological reactivity tests, in vitro |
| ⟨88⟩ | Biological reactivity tests, in vivo |
| ⟨161⟩ | Transfusion and infusion assemblies |
| ⟨381⟩ | Elastomeric closures for injections, biological test procedures, physicochemical test procedures |
| ⟨601⟩ | Aerosols |
| ⟨661⟩ | Containers: light transmission; chemical resistance—glass containers; biological tests—plastics and other polymers; physicochemical tests—plastics; containers for ophthalmics—plastics; polyethylene containers; polyethylene terephthalate bottles and polyethylene; terephthalate G bottles; single-unit containers and unit-dose containers for nonsterile; solid and liquid dosage forms; customized patient medication packages |
| ⟨671⟩ | Containers—permeation: multiple-unit containers for capsules and tablets; single-unit containers and unit-dose containers for capsules and tablets |
| ⟨691⟩ | Cotton (or the monograph for purified rayon USP) |
| ⟨771⟩ | Ophthalmic ointments |
| ⟨1041⟩ | Biologics |
| ⟨1151⟩ | Pharmaceutical dosage forms |

*Source*: From ref. 10.

variety of forms, from impact through vibration in transit and compression forces on stacking.

The demands for mechanical protection will vary with product type: Glass ampules will require greater protection than plastic eye drop bottles, for example. Other protection is required from environmental factors such as moisture, temperature changes, light, gases, and biological agents such as microorganisms and, importantly, humans.

The global market for medicinal products requires that the products are stable over a wide range of temperatures ranging from subzero in the polar region, 15°C in temperate zones, up to 32°C in the tropics. Along with this temperature variation, relative humidity can vary from below 50% to up to 90%, a feature that the packaging should be able to resist if necessary. The majority of packaging materials (including plastics) are to some degree permeable to moisture and the type of closure employed, such as screw fittings, may also permit ingress of moisture. The susceptibility of the product to moisture and its hygroscopicity will have to be considered and may require packaging with a desiccant or the use of specialized strip packs using low-permeability materials such as foil.

Temperature fluctuations can lead to condensation of moisture on the product and, with liquids, formation of a condensate layer on top of the product. This latter problem is well known and can lead to microbiological spoilage as the condensate is preservative free.

If the product is sensitive to photolysis, then opaque materials may be required. Most secondary packaging materials (e.g., cartons) do not transmit light, but in some cases specialized primary packaging designed to limit light transmission is employed.

The package must also prevent the entry of organisms; for example, packaging of sterile products must be microorganism proof—hence the continued use of glass ampules. For nonsterile products the preservative provides some protection, but continual microbial challenge will diminish the efficacy of the preservative, and spoilage or disease transmission may occur.

The packaging material must not interact with the product either to adsorb substances from the product or to leach chemicals into the product. Plastics contain additives to enhance polymer performance. PVC may contain phthalate diester plasticizer, which can leach into infusion fluids from packaging. Antimicrobial preservatives such as phenylmercuric acetate are known to partition into rubbers and plastics during storage, thus reducing the formulation concentration below effective antimicrobial levels.

A complication of modern packaging is the need for the application of security seals to protect against deliberate adulteration and maintain consumer confidence. The active products used must also be stability tested in the proposed packaging material.

The FDA guidance for industry suggests considering consistency in physical and chemical composition. Using a few simple tests, the quality of components and ultimately the container closure system can be monitored.

***Physical Characteristics*** The physical characteristics of interest include dimensional criteria (e.g., shape, neck finish, wall thickness, design tolerances), physical parameters critical to the consistent manufacture of a packaging component (e.g., unit weight), and performance characteristics (e.g., metering valve delivery volume or the ease of movement of syringe plungers). Unintended variations in dimensional parameters, if undetected, may affect package permeability, drug delivery performance, or the adequacy of the seal between the container and the closure. Variation in any physical parameter is considered important if it can affect the quality of a dosage form.

Physical considerations such as water vapor transmission to evaluate seal integrity, thermal analysis such as DSC to monitor melting point and glass transitions of plastics, and IR scanning to prove identity should be part of an ongoing quality control monitoring program.

***Chemical Composition*** The chemical composition of the materials of construction may affect the safety of a packaging component. New materials may result in new substances being extracted into the dosage form or a change in the amount of known extractables. The chemical composition may also affect the compatibility, functional characteristics, or protective properties of packaging components by changing rheological or other physical properties (e.g., elasticity, resistance to solvents, or gas permeability).

The chemical composition should also be evaluated by performing the simple but informative USP physicochemical tests using water, drug product vehicle, and alcohol extractions of plastic components. Specifications should be set for nonvola-

tile residue (total extractables) during the initial suitability tests and then used to monitor the level of polar and nonpolar extractables as part of a quality control plan.

A change in the composition of packaging raw material or a change in formulation is considered a change in the specifications. Due care must be taken to guarantee the safety, compatibility, and performance of a new dosage form in a new packaging system.

The use of stability studies for monitoring the consistency of a container closure system in terms of compatibility with the dosage form and the degree of protection provided to the dosage form is essential. Except for inhalation drug products, for which batch-to-batch monitoring of the extraction profile for the polymeric and elastomeric components is routine, no general policy concerning the monitoring of a packaging system and components with regard to safety is available.

Secondary packaging components are not intended to make contact with the dosage form. Examples are cartons, which are generally constructed of paper or plastic, and overwraps, fabricated from a single layer of plastic or from a laminate made of metal foil, plastic, and/or paper. In special cases, secondary packaging components provide some additional measure of protection to the drug product. In such cases it could be considered a potential source of contamination and the safety of the raw materials should be taken into consideration.

### 3.3 Inhalation Drug Products

Inhalation drug products include inhalation aerosols (metered-dose inhalers); inhalation solutions, suspensions, and sprays (administered via nebulizers); inhalation powders (dry-powder inhalers); and nasal sprays. The carboxymethylcellulose (CMC) and preclinical considerations for inhalation drug products are unique in that these drug products are intended for respiratory tract compromised patients. This is reflected in the level of concern given to the nature of the packaging components that may come in contact with the dosage form or the patient (Table 4).

In October 1998, the FDA issued guidance for industry regarding container closure systems such as metered-dose inhaler (MDI) and dry-powder Inhaler (DPI) drug products.

### 3.4 Drug Products for Injection and Ophthalmic Drug Products

Injectable dosage forms are sterile and represent one of the highest risk drug products. Injectable drug products may be liquids in the form of solutions, emulsions, and suspensions or dry solids that are to be combined with an appropriate vehicle to yield a solution or suspension.

Cartridges, syringes, vials, and ampules are usually composed of type I or II glass or polypropylene frequently used to deliver SVP and LVPs. Flexible bags are typically constructed with multilayered plastic. Stoppers and septa in cartridges, syringes, and vials are typically composed of elastomeric materials. An overwrap may be used with flexible bags to retard solvent loss and to protect the flexible packaging system from rough handling.

Injectable drug products require protection from microbial contamination (loss of sterility or added bioburden) and may also need to be protected from light or

exposure to gases (e.g., oxygen). Liquid-based injectables may need to be protected from solvent loss, while sterile powders or powders for injection may need to be protected from exposure to water vapor.

For elastomeric components, data showing that a component meets the requirements of USP elastomeric closures for injections should typically be performed to assure safety. For plastic components, USP biological reactivity tests are recommended to assure evidence of safety. Whenever possible, the extraction studies described in USP should be performed using the drug product. Extractables should be identified whenever possible. For a glass packaging component, data from USP "Containers: Chemical resistance—Glass containers" will typically be considered sufficient evidence of safety and compatibility. In some cases (e.g., for some chelating agents), a glass packaging component may need to meet additional criteria to ensure the absence of significant interactions between the packaging component and the dosage form.

The performance of a syringe is usually addressed by establishing the force to initiate and maintain plunger movement down the barrel and the capability of the syringe to deliver the labeled amount of the drug product.

Ophthalmic drug products are usually solutions marketed in a LDPE bottle with a dropper built into the neck or ointments marketed in a metal tube lined with an epoxy or vinyl plastic coating with an ophthalmic tip.

Since ophthalmic drug products are applied to the eye, compatibility and safety concerns should also address the container closure system's potential to form substances which irritate the eye or introduce particulate matter into the product (USP ⟨771⟩, ophthalmic ointments).

### 3.5 Liquid-Based Oral Products, Topical Drug Products, and Topical Delivery Systems

The presence of a liquid phase implies a significant potential for the transfer of materials from a packaging component into the dosage form.

***Liquid-Based Oral Drug Products*** Typical liquid-based oral dosage forms are elixirs, emulsions, extracts, fluid extracts, solutions, gels, syrups, spirits, tinctures, aromatic waters, and suspensions. These products are usually nonsterile but may be monitored for changes in bioburden or for the presence of specific microbes.

A liquid-based oral drug product typically needs to be protected from solvent loss, microbial contamination, and sometimes exposure to light or reactive gases (e.g., oxygen). For glass components, data showing that a component meets the requirements of USP "Containers: Glass containers" are accepted as sufficient evidence of safety and compatibility. For LDPE components, data from USP container tests are typically considered sufficient evidence of compatibility.

The USP general chapters do not specifically address safety for polyethylene (HDPE or LDPE), PP, or laminate components. A patient's exposure to substances extracted from a plastic packaging component (e.g., HDPE, LDPE, PP, laminated components) into a liquid-based oral dosage form is expected to be comparable to a patient's exposure to the same substances through the use of the same material when used to package food [27].

***Topical Drug Products*** Topical dosage forms include aerosols, creams, emulsions, gels, lotions, ointments, pastes, powders, solutions, and suspensions. These dosage forms are generally intended for local (not systemic) effect and are generally applied to the skin or oral mucosal surfaces. Topical products also include some nasal and otic preparations as well as some ophthalmic drug products. Some topical drug products are sterile and should be subject to microbial limits.

A rigid bottle or jar is usually made of glass or polypropylene with a screw cap. The same cap liners and inner seals are sometimes used as with solid oral dosage forms. A collapsible tube is usually constructed from metal or is metal lined from LDPE or from a laminated material.

***Topical Delivery Systems*** Topical delivery systems are self-contained, discrete dosage forms that are designed to deliver drug via intact skin or body surface, namely transdermal, ocular, and intrauterine.

Each of these systems is generally marketed in a single-unit soft blister pack or a preformed tray with a preformed cover or overwrap. Compatibility and safety for topical delivery systems are addressed in the same manner as for topical drug products. Performance and quality control should be addressed for the rate-controlling membrane. Appropriate microbial limits should be established and justified for each delivery system.

### 3.6 Solid Oral Dosage Forms and Powders for Reconstitution

The most common solid oral dosage forms are capsules and tablets. A typical container closure system is a plastic (usually HDPE) or a glass bottle with a screw-on or snap-off closure and a flexible packaging system, such as a pouch or a blister package. If used, fillers, desiccants, and other absorbent materials are considered primary packaging components.

Solid oral dosage forms generally need to be protected from the potential adverse effects of water vapor, light, and reactive gases. For example, the presence of moisture may affect the decomposition rate of the active drug substance or the dissolution rate of a dosage form. The potential adverse effects of water vapor can be determined with leak testing on a flexible package system (pouch or blister package). Three standard tests for water vapor permeation have been established by the USP, namely polyethylene containers (USP ⟨661⟩), single-unit containers and unit-dose containers for capsules and tablets (USP ⟨671⟩), and multiple-unit containers for capsules and tablets (USP ⟨671⟩).

## 4 IMPORTANCE OF PROPER PACKAGING AND LABELING

The Poison Prevention Packaging Act (www.cpsc.gov/businfo/pppa.html) requires special packaging of most human oral prescription drugs, oral controlled drugs, certain normal prescription drugs, certain dietary supplements, and many over-the-counter (OTC) drug preparations in order to protect the public from personal injury or illness from misuse of these preparations.

In many countries there are very strict regulations for packaging of many drug substances. Nevertheless, special packaging is not required for drugs dispensed within a hospital setting for inpatient administration. Manufacturers and packagers of bulk-packaged prescription drugs do not have to use special packaging if the drug will be repackaged by the pharmacist.

Various types of child-resistant packages are covered in ASTM International standard D-3475. Medication errors linked to poor labeling and packaging can be controlled through the use of error potential analysis. The recognition that a drug name, label, or package may constitute a hazard to safety typically occurs after the drug has been approved for use and is being marketed. Calls for change almost always result from accumulating reports of serious injuries associated with the use of a drug.

Numerous reports of medication errors are being reported, some of which have resulted in patient injury or death. In a number of these reports, a medication was mistakenly administered either because the drug container (bag, ampule, prefilled syringe and bottle) was similar in appearance to the intended medication's container or because the packages had similar labeling. Obviously, the severity of such errors depends largely on the medication administered.

The problem of medical errors associated with the naming, labeling, and packaging of pharmaceuticals is being very much discussed. Sound-alike and look-alike drug names and packages can lead pharmacists and nurses to unintended interchanges of drugs that can result in patient injury or death. Simplicity, standardization, differentiation, lack of duplication, and unambiguous communication are human factors that are relevant to the medication use process. These factors have often been ignored in drug naming, labeling, and packaging.

The process for naming a marketable drug is always lengthy and complex and involves submission of a new entity and patent application, generic naming, brand naming, FDA—or other corresponding organization all over the world—review, and final approval. Drug companies seek the fastest possible approval and may believe that the incremental benefit of human factor evaluation is small. Very often, the drug companies are resistant to changing, for example, brand names. Although a variety of private-sector organizations in many countries have called for reforms in drug naming, labeling, and packaging standards, the problem remains.

Drug names, labels, and packages are not selected and designed in accordance with human factor principles. FDA standards or other corresponding organizations in other countries do not require application of these principles, the drug industry has struggled with change, and private-sector initiatives have had only limited success.

A number of factors can contribute to the mistaking of one medication for another. Failure to read the package label is one cause. Another if a medication is stored in the wrong location or if clinicians select the medication based solely on the appearance of its package. Also, confusion can occur between medications with names that look alike or sound alike or between premixed medications packaged in similar-looking containers. Another potential source for confusion with premixed medications is the presence of different concentrations of the same medication in a particular location (e.g., a package with 100 mg/mL concentration of a drug could be mistaken for one with 10 mg/mL concentration).

Daily, physicians, nurses, and pharmacists base medical decisions on the information provided by a drug product's labeling and packaging. Unfortunately, poor

labeling and packaging have been linked all too often to medication errors. To help practitioners avoid errors, drug manufacturers should present information in a clear manner that can be grasped quickly and easily.

To determine what presentation is most clear, manufacturers should invite and consider the input of physicians, nurses, and pharmacists, because they work with these products every day and are more likely than label and package designers to discover potential problems. Such input provides the basis for failure and effect analysis (FMEA), also known as error potential analysis or error prevention analysis. FMEA is a systematic process that can predict how and where systems might fail. Using FMEA, health care practitioners examine a product's packaging or labeling in order to identify the ways in which it might fail. A number of steps to reduce confusion and improve the readability of a drug product's label have already been determined through the use of FMEA.

The first step is to reduce label clutter. Only essential information, such as the brand and generic names, strength or concentration, and warnings, should appear prominently on the front label. Numerous deaths have been prevented through the addition of a warning to concentrated vials of injectable potassium chloride, for example. Another step includes the use of typeface to enhance distinctive portions of look-alike drug names on look-alike packaging.

Medication errors are also associated with poor product packaging design. Unfortunately, medication errors linked to poor labeling and packaging are sometimes used in the health care environment to justify the damage. Participation of an expertise from health care practitioners, during labeling and packaging design phase, might have prevented several errors.

Whether for established drugs or new entities going through the approval process, the principles of safe practice in naming, labeling, and packaging are the same and must be very well controlled. Safety experts may differ about specific details, but there is little disagreement about the fundamental principles that should be incorporated into the drug approval process.

Based on reports of errors associated with packaging and labeling, many recommendations have been proposed. Some of them are:

1. Avoid storing medications with similar packaging in the same location or in close proximity.
2. Follow the American Society of Health System Pharmacists (ASHP) guidelines or other legislation of a specific country for preventing hospital medication errors [40, 41]. The ASHP's recommendations include the following:

   Fully document all medication prescription and deliveries and instruct staff that discrepancy or misunderstanding about prescription or patient information should be verified with the prescribing physician. Staff members should be told that all caregivers (regardless of level) have the duty to question the prescribing physician (regardless of the physician's relative position in the hospital hierarchy) if they have concerns about a drug, dose, or patient.

   Periodically train staffs in practices that will help avoid medication errors.

   Ensure that the medication storage and distribution to hospital locations outside the pharmacy are supervised by hospital pharmacy staff only.

Nonpharmacists should not be allowed to enter the pharmacy if it is closed.

3. Perform failure mode and effects analysis. This is a technique used to identify all medication errors that could occur, determine how they occur, and estimate what their consequences would be. Steps then should be taken to prevent errors from occurring, when possible, and to minimize the effects of any errors that do occur.
4. Report any information relating to medication errors to the Medication Errors Reporting Program operated by USP convention [10] and the Institute for Safe Medication Practice (ISMP) or other corresponding institutions in the different countries. The program shares information on medication errors with health care professionals to prevent similar errors from recurring.
5. Hospitals should report incidents in which a device caused or helped cause a medication error.
6. Urge suppliers to provide clear and unique labels and packages for their various individual medications.

Some other considerations relating to standards for drug names, labeling standards, and packaging standards are as follows:

1. *Standard for Drug Names.* The most critical issue in drug name selection is that one name should not be easily confused with another. This applies to both generic and brand names. A name must neither sound like that of another drug nor look like another drug name when it is written out by hand. From the industry's standpoint, the challenge is to find a name that is easy to recollect and appropriate for the connotation desired, do not lead astray (safe), and not already a trade name.

Nowadays, increasing sophisticated and effective methods are available for determining the likelihood of confusion by sound or sight.

2. *Labeling Standards.* To minimize the possibility of error, labels should be easy to read and avoid nonessential material. The name of the drug, and not the name of the manufacturer, should be the most prominent feature and should be in at least 12-point type. The use of color is very controversial; some believe that all colors should be prohibited to force personnel to read the labels.

In the 1990s, a Washington State legislator proposed that every drug product entering the state must have a color-coded label. There was concern on the part of many that the state legislature would turn this idea into law. The prospect of having to color-code all the drugs entering a single state galvanized a response by industry, regulators, practitioners, and safe experts who agreed to revise pharmaceutical labeling. A Committee to Reduce Medication Errors was formed to study the problem. The effort eventually satisfied the color coders and the proposed legislation was dropped.

The committee made several recommendations for standardizing and simplifying labels:

1. Eliminate unnecessary words from the label, such as "sterile," "nonpyrogenic," and "may be habit forming."
2. Allow some abbreviations such as "HCl" and "Inj."
3. Make label information consistent.

4. Require that vials containing medication that must be diluted bear the words "Concentrated, must be diluted" in a box on the label, that the vial have a black flip-top with those words on it, and that the ampules carry a black band.

3. *Packaging Standards.* While there is no evidence that trademark colors and logos on boxes pose a problem, the use of color on bottle tops and labels creates many difficulties. There are dozens of drugs whose names are quite different but whose packages look alike. This creates the potential for error when people "see" what they expect to see on the label.

Standards need to be set for color on both caps and labels. Some believe that prohibiting all color would be safest—in effect, taking away a cue that could divert someone from reading the label.

## 5 REGULATORY ASPECTS

### 5.1 General Considerations

Once the finished dosage form is made, the product should be packed into the primary container and labeled. Additional packaging and labeling are also included. Because of the many products and labeling materials, personnel in this area must be alert to prevent mix-ups. Controls and in-process checks should be carried out throughout the packaging/labeling operation to ensure proper labeling.

Some examples of good manufacturing practices (GMP) requirements specific to packaging and labeling in different countries are as follows:

In the United States the requirements should be written procedures designed to assure that correct labels, labeling, and packaging materials are used for drug products; such written procedures should be followed. These procedures should incorporate features such as prevention of mix-ups and cross-contamination by physical or spatial separation from operations on other drug products.

In Canada, packaging operations are performed according to comprehensive and detailed written operating procedures or specifications, which include identification of equipment and packaging lines used to package the drug, adequate separation, and, if necessary, the dedication of packaging lines packaging different drugs and disposal procedures for unused printed packaging materials. Packaging orders are individually numbered.

In the European Union, the requirements should be formally authorized in the "packaging instructions" for each product containing pack size and type. They are normally included in process controls with instructions for sampling and acceptance limits [42].

### 5.2 Food, Drug and Cosmetic Act

About 100 years after its foundation, the Congress of the United States recognized that subjects related to safety and public health could not exclusively be state dependent and measures should be taken to protect the population in vital areas. Therefore, the federal government became interested in regulating products for consumption.

In 1906, the Congress approved the Wiley Law to avoid the production, sale, or transport of food, medications, and alcoholic beverages that were inadequate or falsified, poisonous, or harmful. It was the first food and medication regulation adopted in interstate commerce. The Congress was given power to regulate commerce between foreign nations and several U.S. states.

In 1912 a civil code law was enacted prohibiting any false affirmation of curing or therapeutic effect on medication labels. The current law was enacted on June 27, 1938, and regulates food, medications, medical devices, and diagnostic and cosmetic products. The law of 1938 stopped regulating the trade of alcoholic beverages. This law stated, among other recommendations, the following:

1. The label of each medication had to give the name of each active component and the quantity of some specific substances, active or not.
2. Cosmetics had to be inoffensive and be properly labeled and packaged.

The 1938 law states that the label of a medication should contain adequate information regarding its use. However, in practice, it became evident that some pharmaceuticals and medications had to be administered by or under the orientation of a medical practitioner, due to the inability of a layman to diagnose a disease, choose an effective treatment, and recognize the cure or the symptoms. Several products were thus classified, but "the prescription concept of a medication" was introduced only after Alteration in the Law of Durham-Humphrey's in 1951. Since then, a label had to carry the warring "Caution, the Federal Law prohibits dispensation without medical prescription." The use of these medications had to be restricted to prescription by a practitioner and the packing or printed material inside had to contain adequate information so that the practitioner could prescribe them safely.

Alterations in 1962 of the 1938 Law constituted an attempt to establish rigid controls on the research, production, divulging, promotion, sale, and use of medications as well as to assure its quality, efficiency, and effectiveness [43].

### 5.3 New Drugs

Before starting clinical trials in humans, an authorization should be obtained from the FDA. This is known as a clinical trial authorization request for a new medication (AEM), on which it is necessary to establish the following:

1. The name that best describes the medication, including the chemical name and the structure of any new molecule
2. A complete list of medication components.
3. A quantitative composition of the medication.
4. The name and address of the vendor and an acquired description of the new drug
5. The methods, facilities, and controls used for the production, processing, and packing of a new medication
6. All available results available from preclinical and clinical trials
7. Copies of medication labels and the informative material that will be supplied to the researchers

8. A description of the scientific training and the appropriate experience considered by the proponent to qualify a researcher as an adequate expert to investigate the medication
9. The names and "curricula vitae" of all researchers
10. An investigation layout planned for test accomplishment in humans

Solicitations for release of new medications are generally very extensive, sometimes thousands of pages. The information has to be enough to justify the affirmations contained in the label of the proposed medication with respect to effectiveness, dosage, and safety. The exact composition of the content on the medication label is usually decided by consensus between the proponent and the FDA.

The requisites for solicitation of new medications, whether by prescription or not, are identical. The instructions contained in the medication labels for use without prescription should demonstrate that the medication can be used safely without medical supervision.

Once the medications are perfected, the publicity related to them has to be routinely presented to the FDA.

The rules of 1985 also changed the requisites regarding addendums that are necessary when alterations are proposed in the medication or in its labeling, for example.

In regulations promulgated by the FDA on February 12, 1972, a clinic should be called upon regarding the effectiveness of a medication. After that the information may be included in the label or in the drug informative leaflet with eligible sentence and defined by dark lines that contour it [43].

Other dispositions contained in the alterations to the 1962 law are as follows:

1. Immediate registration with the FDA before starting the production, repacking, or relabeling of medications and later annual registration, with inspections to be made at least once every two years.
2. Supportive inspections in the factory, particularly where prescription medications are produced.
3. The procedures used by the manufacturers should be in conformity with the good manufacturing practices, which permits the government to better inspect of all the operations.
4. The common name should be presented on the label.
5. The publicity of a prescription medication should present a brief summary mentioning the secondary effects, the contraindications, and the medication effectiveness.
6. All antibiotics are subject for certification procedures.

### 5.4 Labeling Requisites

According to a 1962 law, the main requisites for labeling are as described below. The labeling of over-the-counter medications is regulated by the Food, Drug and Cosmetic Act, which states:

> A medication should be considered falsified unless the label contains: 1. Indications of adequate use and 2. Adequate warnings regarding the pathological indications in those it should not be used or not for children use, when its use can be dangerous for health, of dosages, methods or interval of administration, or unsafe application, of mode and in necessary form for patients' protection.

"Indications of use" were defined in the regulations as information with which even a layman can use the medication safely and for the purpose to which it is designated.

The label of an over-the-counter medication must refer to the active substances, but it is not necessary to indicate its relative quantity, except where the ingredient leads to habituation. In this case the warning "Can lead to habituation" should appear on the label.

A drug can be considered falsified if it does not provide, besides indications of adequate use, warnings against its use in some pathological conditions (or for children) in which the medication can constitute a health risk. Regulations have suggested warnings that can be used for most well-known dangerous substances.

### 5.5 Prescription Drugs

Specific requisites for labeling of ethical medications or of prescription medications are also found in the Food, Drug and Cosmetic Act. These need not to contain "adequate indications of use"; however, they must contain indications for the practitioner, inside or outside the package in which the medication is going to be dispensed, with adequate information for its use. This information may include indications, effects, dosages, route of administration, methods, frequency and duration of administration, important dangers, contraindications, secondary effects, and cautions "according to which the practitioners can prescribe the medication assuredly and for the desirable effects, including those for which it is proclaimed."

Regarding all medications, the act requires that the label present a precise affirmation on the weight of the content, measure or counting, as well as the name and manufacturer's address, packer or distributor.

The label of a prescription medication destined for oral administration has to contain the quantity or proportion of each active substance.

If the medication is for parenteral administration, the quantity or proportion of all the excipients have also to be mentioned on the label, except for those that are added to adjust pH or make it isotonic, in which case only the name and its effect are needed. However, if the vehicle for injection is water, this does not need to be mentioned.

If the medication is not to be administered by any of the routes mentioned above, for example, a pomade or a suppository, all excipients must to be mentioned, except for perfuming agents. Perfumes can be designated as such without the need to mention the specific components.

Coloring agents can be assigned without being specified individually, unless this is required in a separate section for regulation of coloring agents, and inoffensive substances added exclusively for individual identification of each product need not be mentioned.

The only warning that is necessary, "Attention: the Law prohibits the dispensing without prescription," should be on the label of a prescription medication or in its secondary packing if the label is too little to contain it.

### 5.6 Drug Information Leaflet

The inclusion of a drug information leaflet is not compulsory whatever the medication. However, all medications, whether prescription or of over the counter, have to contain a label with adequate indications for use. If the medication label does not have enough space to contain all the information, the drug information leaflet has to be included with necessary information. The drug information leaflet and labels containing indication information must include the date when the text was last revised.

To satisfy the act, the drug information leaflet usually included in the prescription medication packaging should contain "adequate information on usage, including indications, effects, dosages, methods, route, frequency and duration of administration. Any important dangers, contraindications, secondary effects and cautions, based on which the practitioner can prescribe the medication safely and for desirable effects, including those for which a clam is made." To present the information in a uniform manner, the FDA issued labeling policies describing its format and the order and headings for the drug information leaflet description, action, indication, contraindications, alerts, cautions, adverse reactions, dosage and administration, overdose (when applicable), and as it is supplied.

The drug information leaflet can contain the following optional information:

Animal pharmacology and toxicology
Clinical studies
References

Other specific cautions on medication have to appear in a visible manner at the beginning of the drug information leaflet so that practitioners, pharmacists, and patients can easily see them.

According to GMP, an inspector should be cautious with several aspects of drug production, including the following:

1. Product containers and other components have to be tested and be considered adequate for their intended use only if they are not reactive, departure by-products, or even have absorption capacity; so that they do not affect the safety, identity, potency, quality, or purity of the medication or its components.
2. Packing and labeling operations should be adequately controlled to (1) guarantee that only those medications that own quality standards and attain established specifications in their production and control be distributed, (2) avoid mix-ups during the filling operations, packing, and labeling, (3) assure that the labels and labeling used are correct for the medication, and (4) identify the finished product with a batch or a control number that allows determination of the batch production and control history.

Application of the federal law on food, drug, and cosmetics is the FDA's responsibility, which is a subdivision of the Department of Health and Human Services.

The institution is managed by a Commissaries and is subdivided into several departments: Food safety and applied nutrition (CFSAN), Drug evaluation and research (CDER), Biologics evaluation and research (CBER), Devices and radiological health (CDRH), Veterinary medicine (CVM), Toxicological research (NCTR), Regulatory affairs (ORA) and the office of the commissioner (OC) [4, 44].

### 5.7 Other Regulatory Federal Laws

There are other federal laws with which a pharmacist should be familiar. Perhaps the most important are laws on packing and labeling, operations that are regulated by the FDA and the Federal Communications Commission (FCC). The law on packing and labeling is targeted mostly to protecting the consumer. In the case of liquid the ingredients should be on the visible part of the package. The law presents specific requisites concerning the location and size of the type. Violation of this law can lead to apprehension by the FDA or a withdrawal order from the FCC.

Many times a pharmacist involved in developing a product is called upon in the publicity of the medication. For this, he or she must understand the politics of the regulatory agency involved. The FCC, according to the Federal Law of Commerce, has jurisdiction over the announcement and promotion of all consumables, including medications and cosmetics.

This law extends to all publicity and has to do with practices of fraudulent publicity and with promotion that is understood to be false and fraudulent. In general, the FCC controls the publicity of nonprescription drugs and cosmetics with respect to false or fraudulent affirmations, and the FDA is responsible for labeling of medications and for all publicity related to prescription medications. The principal objective of this is to avoid unnecessary duplication of procedures while enforcing the law. The agencies work closely together and the FCC relies strongly on the FDA due to its scientific knowledge. Any government has the right to approve laws for its citizens' protection. This right constitutes the base on which laws regulate the drug substance, the drug product, and its production, distribution, and sale. It is common that these laws exist at a district level, state level, and national level and deal with falsification and adulteration, fraudulent publicity, and maintenance of appropriated sanitary conditions.

Most U.S. states specify the purity requisites, labeling, and applicable packaging of a medication that are generally defined in identical language in federal law. Almost all states, prohibit the commercialization of a new medication until an authorization request for commercialization of a new medication has been submitted to the FDA and has been approved. Medication labeling requisites in each state are established, just as the local laws are defined, taking into consideration arguments and information, such as name and place of activity (production), content quantity, drug name, name of ingredients, quantity or proportion of some ingredients, usage indications, warning regarding dependence, caution against deterioration (degradation), warning about situations in which the use can be dangerous, and special requisites for labeling of official drugs [43].

### 5.8 Fair Packaging and Labeling Act [44]

The FDA through Fair Packaging and Labeling Act regulates the labels on many consumer products, including health products. Title 15: Commerce and Trade

Chapter 39: Fair Packaging and Labeling Program [44]

***Section 1451. Congressional Delegation of Policy*** Informed consumers are essential to the fair and efficient functioning of a free market economy. Packages and their labels should enable consumers to obtain accurate information as to the quantity of the contents and should facilitate value comparisons. Therefore, it is hereby declared to be the policy of the Congress to assist consumers and manufacturers in reaching these goals in the marketing of consumer goods [44].

***Section 1452. Unfair and Deceptive Packaging and Labeling: Scope of Prohibition***

*(a) Nonconforming Labels* It shall be unlawful for any person engaged in the packaging or labeling of any consumer commodity (as defined in this chapter) for distribution in commerce, or for any person (other than a common carrier for hire, a contract carrier for hire, or a freight forwarder for hire) engaged in the distribution in commerce of any packaged or labeled consumer commodity, to distribute or to cause to be distributed in commerce any such commodity if such commodity is contained in a package, or if there is affixed to that commodity a label, which does not conform to the provisions of this chapter and of regulations promulgated under the authority of this chapter.

*(b) Exemptions* The prohibition contained in subsection (a) of this section shall not apply to persons engaged in business as wholesale or retail distributors of consumer commodities except to the extent that such persons (1) are engaged in the packaging or labeling of such commodities, or (2) prescribe or specify by any means the manner in which such commodities are packaged or labeled.

***Section 1453. Requirements of Labeling; Placement, Form, and Contents of Statement of Quantity; Supplemental Statement of Quantity***

*(a) Contents of Label* No person subject to the prohibition contained in section 1452 of this title shall distribute or cause to be distributed in commerce any packaged consumer commodity unless in conformity with regulations which shall be established by the promulgating authority pursuant to section 1455 of this title which shall provide that:

- (1) The commodity shall bear a label specifying the identity of the commodity and the name and place of business of the manufacturer, packer, or distributor;
- (2) The net quantity of contents (in terms of weight or mass, measure, or numerical count) shall be separately and accurately stated in a uniform location upon the principal display panel of that label, using the most appropriate units of both the customary inch/pound system of measure, as provided in paragraph (3) of this subsection, and, except as provided in paragraph (3)(A)(ii) or paragraph (6) of this subsection, the SI metric system;
- (3) The separate label statement of net quantity of contents appearing upon or affixed to any package:
  - (A)
    - (i) if on a package labeled in terms of weight, shall be expressed in pounds, with any remainder in terms of ounces or common or decimal fractions of

the pound; or in the case of liquid measure, in the largest whole unit (quart, quarts and pint, or pints, as appropriate) with any remainder in terms of fluid ounces or common or decimal fractions of the pint or quart;
- (ii) if on a random package, may be expressed in terms of pounds and decimal fractions of the pound carried out to not more than three decimal places and is not required to, but may, include a statement in terms of the SI metric system carried out to not more than three decimal places;
- (iii) if on a package labeled in terms of linear measure, shall be expressed in terms of the largest whole unit (yards, yards and feet, or feet, as appropriate) with any remainder in terms of inches or common or decimal fractions of the foot or yard;
- (iv) if on a package labeled in terms of measure of area, shall be expressed in terms of the largest whole square unit (square yards, square yards and square feet, or square feet, as appropriate) with any remainder in terms of square inches or common or decimal fractions of the square foot or square yard;

- (B) shall appear in conspicuous and easily legible type in distinct contrast (by topography, layout, color, embossing, or molding) with other matter on the package;
- (C) shall contain letters or numerals in a type size which shall be
  - (i) established in relationship to the area of the principal display panel of the package, and
  - (ii) uniform for all packages of substantially the same size; and
- (D) shall be so placed that the lines of printed matter included in that statement are generally parallel to the base on which the package rests as it is designed to be displayed; and

- (4) The label of any package of a consumer commodity which bears a representation as to the number of servings of such commodity contained in such package shall bear a statement of the net quantity (in terms of weight or mass, measure, or numerical count) of each such serving.
- (5) For purposes of paragraph (3)(A)(ii) of this subsection the term "random package" means a package which is one of a lot, shipment, or delivery of packages of the same consumer commodity with varying weights or masses, that is, packages with no fixed weight or mass pattern.
- (6) The requirement of paragraph (2) that the statement of net quantity of contents include a statement in terms of the SI metric system shall not apply to foods that are packaged at the retail store level.

*(b) Supplemental Statements* No person subject to the prohibition contained in section 1452 of this title shall distribute or cause to be distributed in commerce any packaged consumer commodity if any qualifying words or phrases appear in conjunction with the separate statement of the net quantity of contents required by subsection (a) of this section, but nothing in this subsection or in paragraph (2) of subsection (a) of this section shall prohibit supplemental statements, at other places on the package, describing in nondeceptive terms the net quantity of contents: *Provided*, That such supplemental statements of net quantity of contents shall not

include any term qualifying a unit of weight or mass, measure, or count that tends to exaggerate the amount of the commodity contained in the package.

### *Section 1454. Rules and Regulations*

*(a) Promulgating Authority* The authority to promulgate regulations under this chapter is vested in (A) the Secretary of Health and Human Services (referred to hereinafter as the "Secretary") with respect to any consumer commodity which is a food, drug, device, or cosmetic, as each such term is defined by section 321 of title 21; and (B) the Federal Trade Commission (referred to hereinafter as the "Commission") with respect to any other consumer commodity.

*(b) Exemption of Commodities from Regulations* If the promulgating authority specified in this section finds that, because of the nature, form, or quantity of a particular consumer commodity, or for other good and sufficient reasons, full compliance with all the requirements otherwise applicable under section 1453 of this title is impracticable or is not necessary for the adequate protection of consumers, the Secretary or the Commission (whichever the case may be) shall promulgate regulations exempting such commodity from those requirements to the extent and under such conditions as the promulgating authority determines to be consistent with section 1451 of this title:

*(c) Scope of Additional Regulations* Whenever the promulgating authority determines that regulations containing prohibitions or requirements other than those prescribed by section 1453 of this title are necessary to prevent the deception of consumers or to facilitate value comparisons as to any consumer commodity, such authority shall promulgate with respect to that commodity regulations effective to:

- (1) establish and define standards for characterization of the size of a package enclosing any consumer commodity, which may be used to supplement the label statement of net quantity of contents of packages containing such commodity, but this paragraph shall not be construed as authorizing any limitation on the size, shape, weight or mass, dimensions, or number of packages which may be used to enclose any commodity;
- (2) regulate the placement upon any package containing any commodity, or upon any label affixed to such commodity, of any printed matter stating or representing by implication that such commodity is offered for retail sale at a price lower than the ordinary and customary retail sale price or that a retail sale price advantage is accorded to purchasers thereof by reason of the size of that package or the quantity of its contents;
- (3) require that the label on each package of a consumer commodity (other than one which is a food within the meaning of section 321(f) of title 21) bear (A) the common or usual name of such consumer commodity, if any, and (B) in case such consumer commodity consists of two or more ingredients, the common or usual name of each such ingredient listed in order of decreasing predominance, but nothing in this paragraph shall be deemed to require that any trade secret be divulged; or

- (4) prevent the nonfunctional-slack-fill of packages containing consumer commodities. For purposes of paragraph (4) of this subsection, a package shall be deemed to be nonfunctionally slack-filled if it is filled to substantially less than its capacity for reasons other than (A) protection of the contents of such package or (B) the requirements of machines used for enclosing the contents in such package.

*(d) Development by Manufacturers, Packers, and Distributors of Voluntary Product Standards* Whenever the Secretary of Commerce determines that there is undue proliferation of the weights or masses, measures, or quantities in which any consumer commodity or reasonably comparable consumer commodities are being distributed in packages for sale at retail and such undue proliferation impairs the reasonable ability of consumers to make value comparisons with respect to such consumer commodity or commodities, he shall request manufacturers, packers, and distributors of the commodity or commodities to participate in the development of a voluntary product standard for such commodity or commodities under the procedures for the development of voluntary products standards established by the Secretary pursuant to section 272 of this title. Such procedures shall provide adequate manufacturer, packer, distributor, and consumer representation.

*(e) Report and Recommendations to Congress upon Industry Failure to Develop or Abide by Voluntary Product Standards* If (1) after one year after the date on which the Secretary of Commerce first makes the request of manufacturers, packers, and distributors to participate in the development of a voluntary product standard as provided in subsection (d) of this section, he determines that such a standard will not be published pursuant to the provisions of such subsection (d), or (2) if such a standard is published and the Secretary of Commerce determines that it has not been observed, he shall promptly report such determination to the Congress with a statement of the efforts that have been made under the voluntary standards program and his recommendation as to whether Congress should enact legislation providing regulatory authority to deal with the situation in question.

### *Section 1455. Procedures for Promulgation of Regulations*

*(a) Hearings by Secretary of Health and Human Services* Regulations promulgated by the Secretary under section 1453 or 1454 of this title shall be promulgated, and shall be subject to judicial review, pursuant to the provisions of subsections (e), (f), and (g) of section 371 of title 21. Hearings authorized or required for the promulgation of any such regulations by the Secretary shall be conducted by the Secretary or by such officer or employees of the Department of Health and Human Services as he may designate for that purpose.

*(b) Judicial Review; Hearings by Federal Trade Commission* Regulations promulgated by the Commission under section 1453 or 1454 of this title shall be promulgated, and shall be subject to judicial review, by proceedings taken in conformity with the provisions of subsections (e), (f), and (g) of section 371 of title 21 in the same manner, and with the same effect, as if such proceedings were taken by the Secretary pursuant to subsection (a) of this section. Hearings authorized or required

for the promulgation of any such regulations by the Commission shall be conducted by the Commission or by such officer or employee of the Commission as the Commission may designate for that purpose.

*(c) Cooperation with Other Departments and Agencies* In carrying into effect the provisions of this chapter, the Secretary and the Commission are authorized to cooperate with any department or agency of the United States, with any State, Commonwealth, or possession of the United States, and with any department, agency, or political subdivision of any such State, Commonwealth, or possession.

*(d) Returnable or Reusable Glass Containers for Beverages* No regulation adopted under this chapter shall preclude the continued use of returnable or reusable glass containers for beverages in inventory or with the trade as of the effective date of this Act, nor shall any regulation under this chapter preclude the orderly disposal of packages in inventory or with the trade as of the effective date of such regulation.

### 5.9 United States Pharmacopeia Center for the Advancement of Patient Safety [45]

For nearly 33 years, the USP has been reporting programs for health care professionals to share experiences and observations about the quality and safe use of medications. This year, the USP Center for the Advancement of Patient Safety publishes its sixth annual report to the nation on medication errors reported to MEDMARX (Table 6). It was observed that drug product packaging/labeling is one of the main courses of medication errors in hospitals.

### 5.10 National Agency of Sanitary Vigilance (ANVISA, Brazil)

ANVISA is a federal organization linked to Brazil's Health Ministry, which has the incumbency of looking after medication quality and other health products aimed at patients' safety. Several documents regarding GMP and quality control are easily accessed. The agency is also responsible for establishing enforcing the rules and can take corrective measures and punish the offenders [46].

Product stability and compatibility with the conditioning material are distinct, separate, and complementary concepts which should be applied to the pharmaceutical product before being made available for health care.

**TABLE 6 Selected Causes of Error Related to Equipment, Product Packaging/Labeling, and Communication in ICUs**

| Cause of Error | *N* (Nonharmful + Harmful) | Percent Harmful |
|---|---|---|
| Label (the facility's) design | 1,236 | 6,9 |
| Similar packaging/labeling | | |
| Packaging/container design | | |
| Label (manufacturer's) design | | |
| Brand/generic names look-alike | | |

*Source*: MEDMARX Data Report: A Chartbook of 2000–2004 Findings from Intensive Care Units (ICUs) and Radiological Services.

In the compatibility test between formulation and the conditioning material, several options of conditioning materials are evaluated to determine the most adequate for the product.

The environmental conditions and periodicity analyses can be the same as those mentioned for the stability studies for the formulation. In this phase, the possible interactions between the product and the conditioning material which is in direct contact with the medication are verified. Phenomena such as absorption, migration, corrosion, and others that compromise integrity can be observed. Considering that these types of tests are generally destructive, it is necessary to define the number of samples to be tested.

In ANVISA's documents, different types of tests are established that should be carried out with different types of available materials and employed for conditioning medications and cosmetics (cellulose packagings, metallic, plastic, pressurized, etc.) [46].

## 5.11 International Committee on Harmonization (ICH)

In the document "Good Manufacturing Practice Guide for Active Pharmaceutical Ingredients (APIs)" of the ICH Harmonized Tripartite Guideline, the following instructions are given for packaging and identification labeling of APIs and intermediates [47].

### *General*

- There should be written procedures describing the receipt, identification, quarantine, sampling, examination and/or testing and release, and handling of packaging and labeling materials.
- Packaging and labeling materials should conform to established specifications. Those that do not comply with such specifications should be rejected to prevent their use in operations for which they are unsuitable.
- Records should be maintained for each shipment of labels and packaging materials showing receipt, examination, or testing, and whether accepted or reject.

### *Packaging Materials*

- Containers should provide adequate protection against deterioration or contamination of the intermediate or API that may occur during transportation and recommended storage.
- Containers should be clean and, where indicated by the nature of the intermediate or API, sanitized to ensure that they are suitable for their intended use. These containers should not be reactive, addictive, or absorptive so that the quality of the intermediate or API complies with the specifications.
- If containers are reused, they should be cleaned in accordance with documented procedures and all previous labels should be removed or defaced.

### *Label Issuance and Control*

- Access to the label storage areas should be limited to authorized personnel.
- Procedures should be used to reconcile the quantities of labels issued, used, and returned and to evaluate discrepancies found between the number of containers labeled and the number of labels issued. Such discrepancies should be investigated and the investigation should be approved by the quality unit(s).
- All excess labels bearing batch numbers or other batch-related printing should be destroyed. Returned labels should be maintained and stored in a manner that prevents mix-ups and provides proper identification.
- Obsolete and outdated labels should be destroyed.
- Printing devices used to print labels for packaging operations should be controlled to ensure that all imprinting conforms to the print specified in the batch production record.
- Printed labels issued for a batch should be carefully examined for proper identity and conformity to specifications in the master production record. The results of this examination should be documented.
- A printed label representative of those used should be included in the batch production record.

### *Packaging and Labeling Operations*

- There should be documented procedures designed to ensure that correct packaging materials and labels are used.
- Labeling operations should be designed to prevent mix-ups. There should be physical or spatial separation from operations involving other intermediates or APIs.
- Labels used on containers of intermediates or APIs should indicate the name or identifying code, the batch number of the product, and storage conditions, when such information is critical to assure the quality of intermediate API.
- If the intermediate or API is intended to be transferred outside the control of the manufacturer's material management system, the name and address of the manufacturer, quantity of contents and special transport conditions, and any special legal requirements should also be included on the label. For intermediates or APIs with an expiry date, the expiry date should be indicated on the label and certificate of analysis. For intermediates or APIs with a retest date, the retest date should be indicated on the label and/or certificate of analysis.
- Packaging and labeling facilities should be inspected immediately before use to ensure that all materials not needed for the next packaging operation have been removed. This examination should be documented in the batch production records, the facility log, or other documentation system.
- Packaged and labeled intermediates or APIs should be examined to ensure that containers and packages in the batch have the correct label. This examination should be part of the packaging operation. Results of these examinations should be recorded in the batch production or control records.

- Intermediate or API containers that are transported outside of the manufacturer's control should be sealed in a manner such that, if the seal is breached or missing, the recipient will be alerted to the possibility that the contents may have been altered.

### 5.12 European Union Regulatory Bodies

European regulatory requirements say little to date about container closure integrity of parenteral or sterile pharmaceutical products. Regulations provide for package integrity verification of parenteral vials to be supported by the performance of sterility tests as part of the stability program. More specific information is described in the European Union (EU) 1998 "Rules Governing Medical Products in the European Union, Pharmaceutical Legislation." These GMP regulations require that the sealing or closure process be validated. Packages sealed by fusion (e.g., ampules) should be 100% integrity tested. Other packages should be sampled and checked appropriately. Packages sealed under vacuum should be checked for the presence of vacuum. While not as detailed as the FDA guidances, it is evident that the EU rules also require the verification of parenteral product package seal integrity. It is important to note that the EU rules specifically require 100% product testing for fusion-sealed packages, sampling and testing of all other packages, and vacuum verification for packages sealed under partial pressure [42].

The vacuum/pressure decay test is performed by placing the package in a tightly closed test chamber, a pressure or vacuum is applied inside the chamber, and then the rate of pressure/vacuum change in the chamber over time is monitored. The rate or extent of change is compared to that previously exhibited by a control, nonleaking package. Significantly greater change for a test package is indicative of a leak.

## REFERENCES

1. Griffin, J. P. Ed. (2002), *The Textbook of Pharmaceutical Medicine*, 4th ed., BMJ Publishing, London.
2. Harburn, K. (1990), *Quality Control of Packaging Materials in the Pharmaceutical Industry*, Marcel Dekker, New York.
3. O'Brien, J. D. (1990), *Medical Device Packaging Handbook*, Marcel Dekker, New York.
4. U.S. Food and Drug Administration (FDA) (1999, May) Guidance on container closure systems for packaging human drugs and biologics, U.S. Department of Health and Human Services, FDA, Washington, DC.
5. Yoshioka, S. (2000), *Stability of Drugs and Dosage Forms*, Kluwer Academic Publishers: New York, NY, USA, p 272.
6. Banker, G. S., and Rhodes, C. T. (2002), *Modern Pharmaceutics*, 4th ed., rev. and expanded, Marcel Dekker, New York.
7. Connor, J., Rafter, N., and Rodgers, A. (2004), Do fixed-dose combination pills or unit-of-use packaging improve adherence? A systematic review. *Br. World Health Org.*, 82, 935–939.
8. Bloomfield, S. F. (1990), Microbial contamination: Spoilage and hazard, in Denyer, S., and Baird, R., Eds., *Guide to Microbiological Control in Pharmaceuticals*, Ellis Horwood, Chichester, England, pp 29–52.

9. Aspinall, J. A., Duffy, T. D., Saunders, M. B., and Taylor, C. G. (1980), The effect of low density polyethylene containers on some hospital-manufactured eye drop formulations. 1. Sorption of phenyl mercuric acetate, *J. Clin. Hosp. Pharm.*, 5, 21–29.
10. *United States Pharmacopeia* (2006), 29th ed., United States Pharmacopeial Convention, Rockville, MD.
11. Parker, W. A., and MacCara, M. E. (1980), Compatibility of diazepam with intravenous fluid containers and administration sets, *Am. J. Hosp. Pharm.*, 37, 496–500.
12. Mizutani, T., Wagi, K., and Terai, Y. (1981), Estimation of diazepam adsorbed on glass surfaces and silicone-coated surfaces as models of surfaces of containers, *Chem. Pharm. Bull.*, 29, 1182–1183.
13. Yahya, A. M., McElnay, J. C., and D'Arcy, P. F. (1985), Binding of chloroquine to glass, *Int. J. Pharm.*, 25, 217–223.
14. Vromans, H., and Van Laarhoven, J. A. H. (1992), A study on water permeation through rubber closures of injection vials, *Int. J. Pharm.*, 79, 301–308.
15. Matsuura, I., and Kawamata, M. (1978), Studies on the prediction of shelf life. III. Moisture sorption of pharmaceutical preparation under the shelf condition, *Yakugaku Zusshi*, 98, 986–996.
16. Nakabayashi, K., Tuchida, T., and Mima, H. (1980), Stability of packaged solid dosage forms. I. Shelf-life prediction of packaged tablets liable to moisture damage, *Chem. Pharm. Bull.*, 28, 1090–1098.
17. Nakabayashi, K., Shimamoto, T., and Mima, H. (1980), Stability of packaged solid dosage forms. II. Shelf-life prediction for packaged sugar-coated tablets liable to moisture and heat damage, *Chem. Pharm. Bull.*, 28, 1099–1106.
18. Nakabayashi, K., Shimamoto, T., and Mima, H. (1980), Stability of packaged solid dosage forms. III. Kinetic studies by differential analysis on the deterioration of sugar-coated tablets under the influence of moisture and heat, *Chem. Pharm. Bull.*, 28, 1107–1111.
19. Tonnesen, H. H. (1996), *Photostability of Drugs and Drug Formulations*, CRC Press, London.
20. Kontny, M. J., Koppenol, S., and Graham, E. T. (1992), Use of the sorption–desorption moisture transfer model to assess the utility of a desiccant in a solid product, *Int. J. Pharm.*, 84, 261–271.
21. Pikal, M. J., and Lang, J. E. (1978), Rubber closures as a source of haze in freeze dried parenterals: Test methodology for closure evaluation, *J. Parenteral drug Assoc.*, 32, 162–173.
22. Jaehnke, R. W. O., Kreuter, J., and Ross, G. (1990), Interaction of rubber closures with powders for parenteral administration, *J. Parenteral sci. Tech.*, 44, 282–288.
23. Jaehnke, R. W. O., Kreuter, J., and Ross, G. (1991), Content/container interactions: The phenomenon of haze formation on reconstitution of solids for parenteral use, *Int. J. Pharm.*, 77, 4755.
24. Moorhatch, P., and Chiou, W. L. (1974), Interactions between drugs and plastic intravenous fluid bags. II: Leaching of chemicals from bags containing various solvent media, *Am. J. Hosp. Pharm.*, 31, 149–152.
25. Venkataramanan, R., Burckart, G. J., Ptachcinski, R. J., Blaha, R., Logue, L. W., Bahnson, A. C., and Brady, G. J. E. (1986), Leaching of diethylhexyl phthalate from polyvinyl chloride bags into intravenous cyclosporine solution, *Am. J. Hosp. Pharm.*, 43, 2800–2802.
26. Boruchoff, S. A. (1987), Hypotension and cardiac arrest in rats after infusion of mono (2ethylhexyl) phthalate (MEHP), a contaminant of stored blood, *N. Engl. J. Med.*, 316, 1218–1219.

27. U.S. Food and Drug Administration, *Code of Federal Regulations* (CFR)—Title 21, Food and drugs, Chapters 174–186, available: http://www.access.gpo.gov/nara/cfr/index.html, accessed Mar. 11, 2005.
28. Kowaluk, E. A., Roberts, M. S., Blackburn, H. D., and Polack, A. E. (1981), Interactions between drugs and polyvinyl chloride infusion bags, *Am. J. Hosp. Pharm.*, 38, 1308–1314.
29. Illum, L., and Bundgaard, H. (1982), Sorption of drugs by plastic infusion bags, *Int. J. Pharm.*, 10, 339–351.
30. Illum, L., Bundgaard, H., and Davis, S. S. (1983), A constant partition model for examining the sorption of drugs by plastic infusion bags, *Int. J. Pharm.*, 17, 183–192.
31. Atkinson, H. C., and Duffull, S. B. (1990), Prediction of drug loss from PVC infusion bags, *J. Pharm. Pharmacol.*, 43, 374–376.
32. Richardson, N. E., and Meakin, B. J. (1974), The sorption of benzocaine from aqueous solution by nylon 6 powder, *J. Pharm. Phamacol.*, 26, 166–174.
33. Santoro, M. I. R. M., Kedor-Hackmann, E. R. M., and Moudatsos, K. M. (1993), Estabilidade de sais de reidratação oral em diferentes tipos de embalagem. *Bol. Sanit. Panam.*, 115, 310–315.
34. World Health Organization (WHO) (2003), *The International Pharmacopoeia, Tests and General Requirements for Dosage Forms: Quality Specifications for Pharmaceutical Substances and Tablets*, 3rd ed., Vol. 5, WHO, Geneva.
35. Santoro, M. I. R. M., Oliveira, D. A. G. C., Kedor-Hackmann, E. R. M., and Singh, A. K. (2004), Quantifying benzophenone-3 and octyl methoxycinnamate in sunscreen emulsions, *Cosm. & Toil.*, 119, 77–82.
36. Santoro, M. I. R. M., Oliveira, D. A. G. C., Kedor-Hackmann, E. R., and Singh, A. K. (2005), The effect of packaging materials on the stability of sunscreen emulsions, *Int. J. Pharm.*, 13, 197–203.
37. Thoma, K., and Kerker, R. (1992), Photoinstability of drugs. 6. Investigations on the photosansibility of molsidomine, *Pharm. Ind.*, 54, 630–638.
38. *British Pharmacopoeia* (2002), Her Majesty's Stationary Office, London, pp A144, 135–136, 196, 671–673, 778–780, 976–978, 1145–1146.
39. Albert, D. E. (2004), Evaluating pharmaceutical container closure systems, *Pharm. & Med. Packaging News*, 3, 76–78.
40. ASHP Council on Professional Affairs (1993), ASHP Guidelines on preventing medication errors in hospital, *Am. J. Hosp. Pharm.*, 50, 305–314.
41. ASHP Council on Professional Affairs (2001), ASHP guidelines on preventing medication errors in hospital, *Am. J. Hosp. Pharm.*, 58, 3033–3041.
42. *European Pharmacopoeia* (2001), 4th ed., Council of Europe, Strasbourg.
43. Lachman, L., Lieberman, H. A., and Kanig, J. L. (2001), *Teoria e prática na indústria farmacêutica*, Fundação Calouste Gulbenkian, Lisboa.
44. U.S. Food and Drug Administration, Fair Packaging and Labeling Act. Title 15—Commerce and Trade, Chapter 39—Fair Packaging and Labeling Program, available: http://www.fda.gov/opacom/laws/fplact.htm accessed Mar. 11, 2005.
45. Santell, J. P., Hicks, R. W., and Cousins, D. D. (2005), *MEDMARX Data Report: A Chartbook of 2000–2004 Findings from Intensive Care Units and Radiological Services*, USP Center for Advancement of Patient Safety, Rockville, MD.
46. Agência Nacional de Vigilância Sanitária (ANVISA) (2004), *Guia de Estabilidade de Produtos Cosméticos*, ANVISA, Brasília.
47. International Organization on Harmonisation (2000), ICH harmonized tripartite guideline: Good manufacturing practice guide for active pharmaceutical ingredients, available: http://www.ICH.org, accessed June 23, 2005.
48. Sarbach, C., Yagoubi, N., Sauzieres, J., Renaux, C., Ferrier, D., and Postaire, E. (1996), Migration of impurities from a multi-layer plastics container into a parenteral infusion solution, *Int. J. Pharm.*, 140, 169–174.

# CLEAN-FACILITY DESIGN, CONSTRUCTION, AND MAINTENANCE ISSUES

RAYMOND K. SCHNEIDER
*Clemson University, Clemson, South Carolina*

**Contents**

1 Introduction
2 Planning for Project Success
  2.1 Needs Assessment
  2.2 Front-End Planning
  2.3 Preliminary Design
  2.4 Procurement
  2.5 Construction
  2.6 Start-Up and Validation
  2.7 Summary
3 Design Options
  3.1 Clean-Facility Scope
  3.2 Design Parameters
  3.3 Architectural Design Issues
  3.4 Materials of Construction
  3.5 HVAC System
  3.6 Clean-Room Testing
  3.7 Utilities
4 Construction Phase: Clean Build Protocol
  4.1 General
  4.2 Level I Clean Construction
  4.3 Level II Clean Construction
5 Maintenance
Appendix A: Guidelines for Construction Personnel and Work Tools in a Clean Room
Appendix B: Cleaning the Clean Room
Bibliography

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

## 1 INTRODUCTION

While there are discrete steps in the design and construction of a pharmaceutical manufacturing plant project, those projects deemed successful incorporate certain practices that promote flow of the construction process toward completion on time and within budget. Proper front-end planning is not completed until it results in appropriate values for design parameters, "buy-in" at all levels of management, and clear direction for the design phase. Engineering the clean room in accordance with recognized industry practice would produce construction documents that facilitate clear procurement and construction planning as well as a focused, efficient, construction effort. A full return on the energy expended through the construction phase cannot be realized without a well-executed start-up and validation process that provides baseline data for effective ongoing operation and maintenance.

The steps in the clean-room construction project include:

Needs assessment
Front-end planning
Preliminary design
Construction document development
Procurement
Construction
Start-up and validation

One of the truisms of the construction industry is that the greatest impact on the cost of a facility can be made at the earliest stages of the process. The construction process can be likened to a snowball rolling down a snow-covered hill. It grows and gains momentum, seemingly taking on a life of its own, until it can only be brought under control with a major effort. So too with manufacturing plant projects. Careful work during the first three stages will ensure that the project begins on a well-directed course and moves to a successful conclusion.

Sometimes the special nature of pharmaceutical manufacturing plant projects clouds the fact that building such a plant is in fact a construction project. The facility engineering team of a small to medium company may be tempted to turn away from such projects due to the projects' perceived uniqueness and leave the key decision making to others. In fact, it is the construction experience of that team that is most required to keep the project costs under control. The way to accomplish this is for the team to be involved in the process from its earliest stages.

Let us review the steps in such a project and identify what should occur at each step and the potential for trouble.

## 2 PLANNING FOR PROJECT SUCCESS

### 2.1 Needs Assessment

It is during this early stage that a requirement for a clean manufacturing facility is perceived. The need for the facility may be precipitated by a new product, an

improved product, an improved manufacturing methodology, new or more stringent regulation requirements, or perhaps a change in marketing strategy.

At this point a study should be undertaken to determine the benefits to be realized by the new facility as well as the costs to be incurred. Costs arise from not only construction but also ongoing operation and maintenance. These costs are affected by the plant location and the availability of a trained or trainable workforce. Does the day-to-day operation of the facility generally require that special attire be worn? Are special procedures, possibly more time consuming than those presently used, required? It is important that this study is complete and accurate in order to prevent any unrealistic expectations on the part of management and plant operations and to permit advanced planning for revised procedures once the facility is in use.

The study should describe the goals of the project, its impact on present operations, budget restraints, tentative schedule, and path forward. It will serve as the basis for front-end planning and will provide the standard against which the success of the program is measured.

## 2.2 Front-End Planning

While the needs assessment study may be conducted by a limited number of people, the front-end planning process should be open to all. Plant facilities people will be bearing the brunt of the responsibility for bringing the facility online, on schedule, and within budget. Process people are responsible for ensuring that the facility will adequately house the process equipment and that the facility incorporates sufficient space, utilities, process flow considerations, and provision for flow of people and material to support the goals of the building program. Human resources people have to staff the facility, either out of the present employee pool or from the general local labor market. They must know the requirements of potential employees as well as the conditions under which they will be working.

Procurement people will be purchasing furnishings and process equipment for the plant as well as overseeing the contracts let to the design and construction professionals. Operations people should have input regarding design parameters such as temperature, humidity, lighting, vibration, cleanliness class, and energy needs. Materials handling people should participate in order to understand the requirements for storing and transporting raw materials as well as retrieving, storing, and shipping finished goods from the plant.

An integral part of the front-end planning team should be the design professionals charged with developing the plant design based on client input in such a way as to satisfy as many requirements developed in needs assessment as possible. This team may be assembled internally but frequently is drawn from specialty builders, architectural and engineering (A&E) firms, and design/build firms active in the pharmaceutical industry. The team of design professionals should have pharmaceutical experience on facilities comparable in size and complexity to that being planned as well as extensive experience in construction projects of all types. The design team may offer design only, design/build, procurement, construction management, or combinations of these services. This design team should be considered a resource during the front-end planning phase. It is the wise client who takes advantage of the experience of the design team, permitting them a large role as facilitators of the planning sessions.

An appropriate design team will demonstrate expertise in contamination control philosophies, space planning, code compliance, and mechanical and electrical design and will be familiar with materials of construction currently being used in pharmaceutical projects. It is frequently helpful to include a member of the construction team in the front-end planning effort to advise on constructibility of the facility being planned. Unrealistic construction schedules will be avoided and field rework will be minimized if appropriate attention is paid to the construction phase early in the planning process.

## 2.3 Preliminary Design

Front-end planning typically utilizes the expertise of client process people to convey the requirements of the pharmaceutical facility to the design team. With this information in hand the design team begins the facility design incorporating process needs, code requirements, safety issues, material and personnel flow, work-in-process storage, utility needs, and so on, into a first-cut approach.

Client representatives have an opportunity to review the effort and begin fine tuning the design to incorporate late-breaking process changes. The preliminary design is a target that helps both the design team and the client solidify design goals. Change is inexpensive, and therefore encouraged, at this stage and buy-in by all concerned is a major objective of this phase of the design effort.

A budget based on the agreed-upon preliminary design should be developed to make sure that the overall project is on course. This will minimize surprises further along in the design/build process. Ideally the design will be "cast in stone" at the end of the preliminary phase. This permits the production work on the design documents to proceed unhindered. The more unknowns left at the end of the preliminary phase, the more difficult it will be to complete design documents in a timely fashion.

***Construction Document Development*** The construction documents should convey the intent of the design team and client to the construction team. A good set of construction documents should result in a tight spread of construction bids as there should be little room for varying interpretation on the part of the potential construction contractors. The drawings should have sufficient notes to convey the design intent without creating a cluttered appearance. The written specifications should be as brief as possible consistent with clarity.

Complicated documents create the impression that a project may be more involved, and therefore more costly, than it should be. Cautious contractors may unnecessarily inflate their bid to cover perceived contingencies. Specifications that are too wordy may be difficult to follow and similarly result in higher prices as bidders make sure all bases are covered. No one likes surprises.

The development of construction documents should be a straightforward process with little involvement by the client except to monitor the process and ensure that the original design intent is followed. While changes will always occur during this phase ("cast in stone" is a euphemism for "let's keep the changes under control"), they are certainly less costly at this point than during the construction phase. It is desirable to minimize such changes. A continuous sequence of changes suggests that the preliminary design phase was not entered into seriously. It demonstrates a lack

of preparedness on the part of the client and a lack of ability to communicate and draw out the client's needs on the part of the design team. A sense of clarity of purpose slips away with ongoing change and the possibility for errors in construction documents, which eventually surface as costly construction changes, increases.

## 2.4 Procurement

A detailed scope of work describing the materials and services required is a vital part of the procurement process. There is no purpose to keeping the project bidders in the dark regarding what is required of them. The role of the procurement function is to obtain maximum value, that is, the best quality and schedule at the lowest price. The clearer the scope of work and construction documents, the better will be the chance of this happening. A low price is not a good value if the schedule slips by several months as a result. A marginal plant that does not maintain design conditions or meet production goals is a poor value even if it was delivered within schedule.

The procurement process should qualify potential bidders by ensuring that similar pharmaceutical projects have been delivered on time, within budget, and on schedule. References should be checked. It is expected that references offered by a potential bidder would have good things to say about that bidder, but this is not a certainty and pointed questioning about personnel, schedule, quality, change orders, follow-up, and so on, can help develop a warm feeling or an uncertain feeling about potential bidders. If bids are in fact quite close, it is the quality of references that might suggest a particular bidder be given preference.

There are a number of ways in which the project can be procured. Use of in-house engineering and construction expertise may work in special situations or on smaller projects. Typically problems arise when facilities departments, stretched to their limit with ongoing plant requirements, must lower the priority of the new facility to meet other commitments. Schedules may stretch out unacceptably.

A number of specialty contractors have proven over the years to be adept at installing small turnkey facilities of limited complexity in a timely and economical fashion. If extensive engineering is required, if local code compliance becomes an issue, if complex process requirements must be met, or if the client requirements exceed the experience of the supplier there could be cause for concern.

Design/build is a popular approach in that it suggests a single source of responsibility for all phases of the project. Frequently firms billing themselves as "design/build" are strong in either design *or* build, but not both. The strong design firm can put the essentials on paper but the final price and schedule may suffer. The strong construction firm may lack the expertise to create an appropriate manufacturing environment, particularly where clean-room expertise is required. The project may be outstanding in all respects except performance. A good review of references is essential before selecting a design/build firm.

Construction management has been increasingly used on larger projects. A good construction management firm will work closely with the client-selected design company to review constructability and adequacy of construction documents. It will assist to qualify bidders, maintain schedule, track costs, administer and oversee, and generally ensure that a team incorporating the strongest skills is assembled to complete the project. Pharmaceutical experience is essential.

## 2.5 Construction

The construction process should proceed smoothly if the remarks presented above are followed. Cost can increase during this phase if changes must be implemented. While change is inevitable, a construction change procedure negotiated during the bidding phase and in place during construction will keep such change from getting out of control.

The requirement for "building clean" has arisen in recent years as more stringent clean rooms have become more popular. Imposing a clean construction protocol on contractors can lengthen the schedule and increase cost. The protocol should be developed during the construction document phase and be an integral part of the bid documents. Once the decision is made to work clean, protocols developed should be followed by *everyone* on the jobsite associated with the clean areas. A poorly conceived and enforced protocol will be a costly and futile exercise. The tendency to build clean on every new or retrofit project should be carefully evaluated and a practical protocol should be developed consistent with the needs of the project.

Client end users should be encouraged to observe construction as it progresses. They will be more intelligent about how their facility was built and therefore more attuned to maintaining the facility once it is completed and in operation. While suggestions should be welcomed as construction progresses, it is important that a chain of command be enforced. Any questions or suggestions or concerns should not be expressed to workers on the site but rather through project management channels. In this way good ideas can be implemented and bad ideas shelved without impacting the construction effort in a negative manner. Note the one exception to this practice is in regard to safety. Everyone on the site has safety responsibility. Any unsafe acts should be questioned and supervisors consulted immediately.

## 2.6 Start-Up and Validation

Subcontractors on the jobsite should be responsible for start-up as well as installation of equipment. Equipment manufacturers typically have personnel available to ensure appropriate start-up procedures are followed. If several trades are involved in the installation of a particular piece of equipment, then one trade should be assigned, by contract, as having coordinating responsibility for that piece of equipment. This will minimize "finger pointing" when equipment does not start or operate properly. This can be a sensitive issue and a construction manager can set the tone for cooperation in this area.

An independent contractor responsible to the construction manager or owner should do testing and balancing (TAB) of mechanical systems. All start-up should be complete and initial valve or damper settings made (and recorded) by the subcontractor before testing and balancing begins. The TAB contractor should not have to repair equipment or troubleshoot inoperative equipment but rather only adjust and verify performance of equipment.

A separate contractor should certify clean-room areas. This might be the TAB contractor if that firm is suitably qualified. There should be no question of equipment being operative at this stage of the project since start-up and testing and balancing are complete. Certification is the verification of facility compliance

with clean-room specifications. If the facility design is well conceived and the construction team has installed a quality project, any certification test failure will most likely be corrected through fairly minor adjustments. Failure of the clean room to pass certification tests might require redesign but more frequently requires some equipment adjustment or perhaps a filter repair and then a retest. It is important that a clear understanding of responsibility be communicated before problems are encountered. Failure to plan for potential problems could result in extending the schedule and incurring unforeseen costs at a crucial point in the project.

### 2.7 Summary

Recognizing the step-by-step process involved in even the smallest pharmaceutical project can help focus attention in a manner that will result in a successful project. The formal schedule of a well-conceived project will include needs assessment, front-end planning, and preliminary design. It is important that project progress is measured against such a schedule and not just by the visual impact caused by bricks and mortar being installed.

## 3 DESIGN OPTIONS

### 3.1 Clean-Facility Scope

The purpose of this section is to identify design and construction options for those parts of a pharmaceutical facility intended to house process equipment. These suggestions are intended to assure that the facilities, when used as designed, will meet the requirements of current good manufacturing practices (cGMPs). Air cleanliness within the facility may range from International Organization for Standardization (ISO) 5 (Class 100) through ISO 8 (Class 100,000). In addition, areas may be considered clean or labeled as "controlled environment" without having a cleanliness class assigned to the space. Note that throughout this chapter cleanliness class will be described using the designation presented in the new ISO 14644 (e.g., ISO 5, ISO 8) and parenthetically as presented in the currently obsolete (but widely understood and quoted) U.S. Federal Standard 209 (e.g., Class 100, Class 100,000).

A cleanliness classification in accordance with the latest revision of ISO 14644 is generally inadequate by itself to describe a facility used for pharmaceutical processes. The presence of viable particles (living organisms) within the particle count achieved by applying methods described in the standard may affect the product within the facility. A measure of both viable and nonviable particles is required to provide sufficient information upon which to base a decision regarding the suitability of the clean room for its intended purpose.

The options presented herein are intended to provide facilities that will effectively restrict both viable and nonviable particles from entering the clean areas, minimize contamination introduced by the facility itself, and continuously remove contaminants generated during normal operations.

Measurement of total particle count in the clean room is described in ISO 14644. This count may be composed of viable, nonviable, or nonviable host particles with a viable traveler. There is no generally accepted relationship between total particle

count and viable particle count. While maintaining appropriate particle counts is important in clean-room design and operation, a protocol designed to identify viable particles should be inherent in the certification/validation testing of a pharmaceutical clean room.

No facility design can compensate for excessive contamination generated within it. In addition to effective facility design, the user must also institute a routine maintenance program as well as maintain personnel and operational disciplines that limit particles both entering and being generated within the facility.

While this section identifies options for contamination control in facility design, any such options must be implemented in accordance with all appropriate government and regulatory building and safety codes. The design guideline is nonspecific as regards biological or chemical materials that may be used within the facility but generally addresses bulk pharmaceutical chemical plants (BPCs), secondary manufacturing chemical plants, bulk biopharmaceutical plants, and plants used for fill and finish operations. Good practice as well as any regulations governing biological and pharmaceutical processes conducted within the facility must be adhered to as required and could modify some of the suggestions contained herein.

### 3.2 Design Parameters

The design of the facility is based upon specification of certain design parameters. These in turn are used to calculate building system equipment capacities and aid in the selection of the appropriate types of equipment that are required. Design parameters that may be critical are discussed below.

***Cleanliness Classification*** The classification of the clean areas is determined by the using organization consistent with the level of nonviable and viable particulate contamination acceptable to the process conducted within the facility. This may be governed by regulatory agencies, client organizations, or company protocols. Target goals are set for nonviable particle count in accordance with the ISO. Viable particle target goals should be stated in colony-forming units (CFU) per square centimeter. In accordance with ISO 14644, particle goals will typically be identified for "at rest" and "operational" modes.

*In the absence of other guidance* governing the cleanliness classification and acceptable levels of microbial contamination of the clean room, the values presented in Table 1 may be used. The room grades presented are from most critical (A) to least critical (E). The definition of criticality is left to the clean-room user organization.

***Other Design Parameters*** Facility design parameters that support the process within the clean room should be established by the user organization. Parameters such as temperature, humidity, lighting requirements, sound level, and/or vibration may be process driven or comfort driven and therefore are selected to accommodate specific process or comfort requirements as determined by the end user.

***Local Control*** Under some circumstances, cleanliness requirements can be achieved through the use of localized controls such as clean tents, glove boxes, minienvironments, or isolators. These provide unidirectional filtered airflow within

**TABLE 1 An Example of Cleanliness Classification Goals**

| | | Particle Counts[e] | | Microbial Contamination | | | |
|---|---|---|---|---|---|---|---|
| Room Grade | Cleanliness Class[a] | At Rest | Operational | Air Sample (cfu/m$^3$) | Settle Plates (cfu/4 h)[b] | Contact Plates (cfu/plate)[c] | Glove Print (cfu/glove)[d] |
| A[f] | M3.5 (100) | 3,500 | 3,500 | <1 | <1 | <1 | <1 |
| B[g] | M3.5 (100) | 3,500 | 35,000 | 10 | 5 | 5 | 5 |
| C | M5.5 (10000) | 350,000 | 3,500,000 | 100 | 50 | 25 | — |
| D | M6.5 (100000) | 3,500,000 | N/A | 200 | 100 | 50 | — |
| E | Uncontrolled | N/A | N/A | N/A | N/A | N/A | N/A |

[a]In accordance with U.S. Federal Standard 209E.
[b]90-mm-diameter settling plate. These are average values and individual plates may have <4 h of exposure.
[c]55-mm contact plates.
[d]Five-fingered glove.
[e]Maximum particle counts per cubic meter >0.5 μm.
[f]Unidirectional airflow at 90 ft/min.
[g]Non-unidirectional airflow.

a limited area. They may be located within a facility that provides the necessary temperature and humidity conditions or they may be provided with integral environmental control equipment designed to maintain necessary conditions.

***Air Change Rate*** The airflow pattern and air change rate in a clean room largely determines the class of cleanliness that can be maintained during a given operation. Non-unidirectional flow clean rooms rely on air dilution as well as a general ceiling-to-floor airflow pattern to continuously remove contaminants generated within the room. Unidirectional flow is more effective in continuously sweeping particles from the air due to the piston effect created by the uniform air velocity. The desired air change rate is determined based on the cleanliness class of the room and the density of operations expected in the room. An air change rate of 10–25 per hour is common for a large, low-density ISO 8 (Class 100,000) clean room. ISO 7 (Class 10,000) clean rooms typically require 40–60 air changes per hour. In unidirectional flow clean rooms, the air change rate is generally not used as the measure of airflow but rather the average clean-room air velocity is the specified criterion. The average velocity in a typical ISO 5 (Class 100) clean room will be 70–90 ft/min. A tolerance of plus or minus 20% of design airflow is usually acceptable in the clean room. The foregoing values have been found to be appropriate in many facilities. Generally air change rate or air velocity is not a part of regulations. It is left to the user to demonstrate that the selected design parameter is appropriate for the products being manufactured. An exception to this may be in the case of filling operations where a unidirectional flow velocity of 90 ± 20 ft/min may be required.

***Pressurization*** A pressure differential should be maintained between adjacent areas, with the cleaner area having the higher pressure. This will minimize infiltration of external contamination through leaks and during the opening and closing of personnel doors. A minimum overpressure between clean areas of 5 Pa [0.02 in. of

water column (in. WC)] is recommended. The pressure between a clean area and an adjacent unclean area should be 12–14 Pa (0.05 in. WC). Where several clean rooms of varying levels of cleanliness are joined as one complex, a positive-pressure hierarchy of cleanliness levels should be maintained, including air locks and gowning rooms. Note that for certain processes and products it may be desirable to have a negative pressure relative to the surrounding ambient in one or more rooms when containment is a major concern. A "room within a room" may have to be designed to achieve this negative pressure yet still meet the needs of clean operation.

***Temperature Control*** Where occupant comfort is the main concern, a temperature of 68–70°F ± 2°F will usually provide a comfortable environment for people wearing a typical lab coat. Where a full "bunny suit" or protective attire is to be worn, room temperature as low as 66°F may be required. If the temperature is to be controlled in response to process concerns, the value and tolerance should be specified early in the design phase to ensure that system selection is appropriate and that budgeting is accurate. Note that a tight tolerance (e.g., ±1°F or less) will typically be more costly to maintain than a less stringent tolerance.

***Humidity Control*** The humidity requirement for comfort is in the range of 30–60% relative humidity (RH). If process concerns suggest another value, it should be specified as soon as possible in the design process. Biopharmaceutical materials sensitive to humidity variations or excessively high or low values may require stringent controls.

### 3.3 Architectural Design Issues

***Facility Layout*** The facility layout should support the process contained within the clean room. While a rectangular shape is easiest to accommodate, other shapes may be incorporated into the facility as long as appropriate attention is paid to airflow patterns. The facility should be able to accommodate movement of equipment, material, and personnel into and out of the clean room. The layout of the clean suite should facilitate maintaining cleanliness class, pressure differentials, and temperature/humidity conditions by isolating critical spaces and by excluding non-clean operations. See Figure 1. The potential for cross-contamination is addressed as both an architectural and a mechanical issue. Generally, in a facility where multiple products are to be processed, each product has a dedicated space, isolated physically from adjacent spaces, and each has its own air conditioning system, independent of adjacent systems.

***Air Locks or Anteroom*** This is a room between the clean room and an unrated or less clean area surrounding the clean room or between two rooms of differing cleanliness class. The purpose of the room is to maintain pressurization differentials between spaces of different cleanliness class while still permitting movement between the spaces. An air lock can serve as a gowning area. Certain air locks may be designated as an equipment or material air lock and provide a space to remove packaging material and/or to clean equipment or materials before they are introduced into the clean room. Interlocks are recommended for air lock door sets to prevent opening of both doors simultaneously. The air lock is intended to separate the clean from the unclean areas.

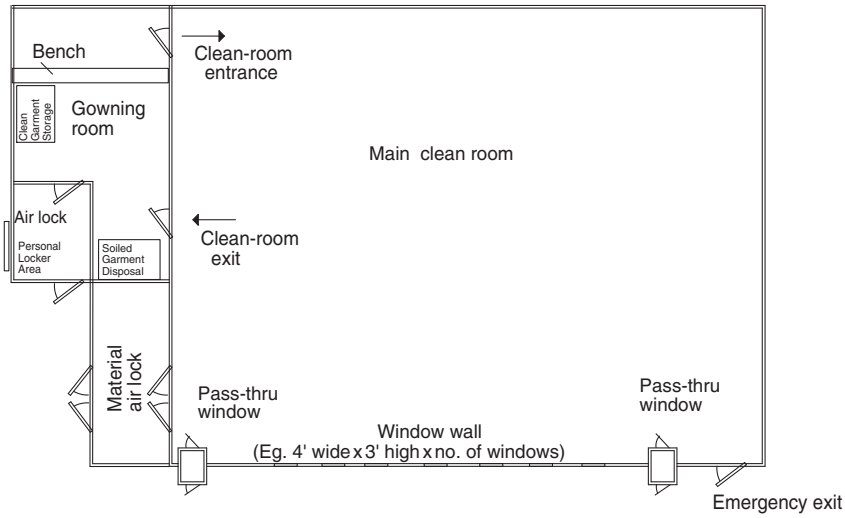


**FIGURE 1** Sample clean-room lay-out.

Prior to equipment or raw materials being introduced into the clean room, they should be prepared. This may mean removing an outer package wrap or perhaps surface cleaning of the object. Material handling equipment used within the clean room should be dedicated to the clean room. Physical barriers may be integrated into the material air lock design to prevent material handling equipment from leaving the clean room or outside equipment from passing into the clean room.

***Windows*** Windows are recommended in interior clean-room walls to facilitate supervision and for safety, unless prohibited by the facility protocol for visual security reasons. Windows in exterior building walls adjacent to a clean space are problematic. Windows can be a source of leakage and can result in contaminants entering the space. Windows should be placed to permit viewing of operations in order to minimize the need for non-clean-room personnel to enter the clean room. Windows should be impact-resistant glass or acrylic, fully glazed, installed in a manner that eliminates or minimizes a ledge within the clean space. Double glazing is frequently used to provide a flush surface on both sides of the wall containing the window. Windows may be included if there is a public relations requirement for visitors to view the operations. Speaking diaphragms or flush, wall-mounted, intercom systems are recommended near all windows to facilitate communication with occupants of the clean room.

***Pass-Through*** A pass-through air lock should be provided for the transfer of product or materials from uncontrolled areas into the clean room or between areas of different cleanliness class (Figure 2). The pass-through may include a speaking diaphragm, intercom, or telephone for communication when items are transferred and interlocks to prevent both doors from being opened at the same time. A cart-size pass-through installed at floor level can be used to simplify the movement of carts between clean areas. Stainless steel is typically the material of choice (Figure 3).

**FIGURE 2** Stainless steel pass-through with interlock designed to permit safe passage of small items between spaces of differing cleanliness. (Courtesy of Terra Universal.)

**FIGURE 3** Cart pass-through enabling larger amounts and sizes of items to be transported. Note that the cart shown is not to be taken from the clean room. Typically a physical barrier is incorporated into the cart pass-through design. (Courtesy of Terra Universal.)

***Gowning Room*** Gowning rooms should be designed to support the garment protocol established for the facility. A typical gowning room may have a wall- or floor-mounted coat rack for clean garment storage (Figure 4); a bench specifically designed for clean-room use (Figure 5); a full-length mirror installed near the door for gowning self-inspection; storage for new packaged garments; and bins for disposal of soiled garments.

Personal lockers and coat racks for the storage of notebooks, coats, and personal items should be located outside the gowning room or in an anteroom separate from

**FIGURE 4** Furnishings in the gowning room are typically of a nonshedding material such as the stainless steel designs shown. The gown rack will generally have a ceiling-mounted HEPA filter above it to continually bathe the garments in clean air. (Courtesy of Terra Universal.)

**FIGURE 5** The stainless steel clean benches have has a perforated seat to permit airflow from ceiling to floor essentially unobstructed. (Courtesy of Terra Universal.)

the clean gowning area. Restroom facilities may also be located outside the gowning room or in an anteroom adjacent to the clean gowning area. A common gowning room design has two areas divided by a bench. The "unclean" area is used to remove and store outer garments. Stepping over the bench as the clean-room footwear is being put on ensures that the "clean" side of the gowning room will remain that way. Final donning of the clean-room garb is then accomplished.

Male and female gowning rooms may be required depending on the make-up of the work force and the type of garments being used.

***Siting*** A clean room that serves as an element of a larger process line should be integrated into the line to permit movement of personnel and materials in and out of the room. A free-standing clean room may be located in any convenient site; however, certain conditions adjacent to the facility may degrade its performance. Vibration sources inside or near a clean room will encourage particle release within the room and under severe conditions may cause leaks in filters and ductwork. Heavy equipment, including the heating, ventilation, and air conditioning (HVAC) system components, pumps, house and vacuum system, ought to be vibration isolated. Location of a clean room directly adjacent to heavy equipment or loading docks that see heavy truck traffic and other sources of vibration, shock, and noise may be problematic. The outdoor air intake for the clean-room makeup air must be carefully located to prevent overloading of filters or entrance of contaminating gases that the filter will not remove. Clean-room air intakes should not be located near loading docks, traffic lanes, or other areas where vehicles may drive through or idle. These intakes should not be located near the exhaust locations of other processing facilities. Use of gas-phase filtration may be required if the quality of make-up air is not acceptable.

### 3.4 Materials of Construction

***Walls*** Generally wall material selection should be based on the operations and material handling equipment to be used within the space. The walls should be strong enough to withstand repeated impact of carts or other equipment without deterioration. The materials should also be selected with the sanitizing protocol in mind. Chemicals, high-pressure wash, and steam can cause reduced wall life if proper materials are not selected. Seamless walls, to the extent possible, are desirable.

Basic steel stud construction with gypsum board paneling can be used in biopharmaceutical clean rooms when appropriately coated with a nonshedding finish. Modular wall systems utilizing coated steel or aluminum panel construction are growing in popularity due to the ability to easily retrofit a lab or production space at a later date with minimal disruption and construction debris. Stainless steel may be appropriate but costly. Modular systems have been developed that address the concerns of the biopharmaceutical clean-room user relative to surface finish integrity and smooth surfaces. The joint between adjacent modular panels is commonly treated with a gunnable sealant to provide a smooth, cleanable joint that will not hold contaminants.

Concrete masonry unit (CMU) construction is widely used (Figure 6). It can prevent buildup of contaminants when finished with an epoxy or other smooth, chemical-resistant coating. Where retrofit is not a regular practice, the strength of concrete block and its long life recommend it.

Rounded, easy-to-clean corners and smooth transitions between architectural features such as windows and walls (Figure 7) should be featured in all wall system designs, whether modular or "stick built."

**FIGURE 6** A CMU wall treated with block filler and epoxy finish to provide a smooth, cleanable wall surface. (Courtesy of Niagara Walls.)

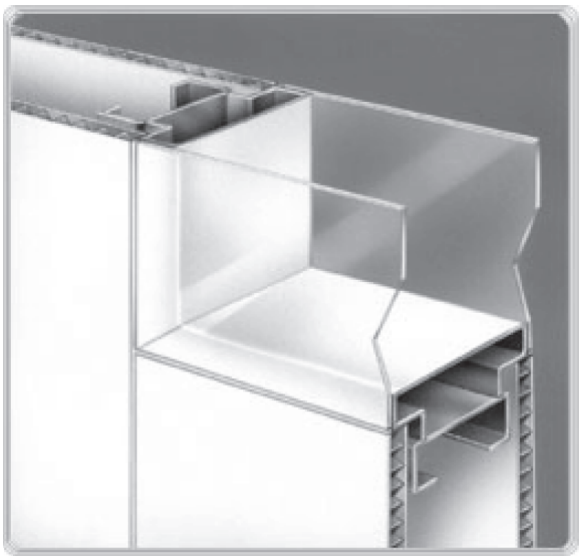

**FIGURE 7** A window detail that provides a smooth, easy-to-clean surface on both wall faces. (Courtesy of Portafab.)

***Wall Finishes*** Inexpensive latex wall paints will deteriorate over time and are unacceptable in clean rooms. Acceptable wall finishes include epoxy paint, polyurethane, or baked enamel of a semigloss or gloss type. These may be applied in the factory to metal wall system panels. Field application of epoxy to gypsum board or CMU should be done to ensure a smooth, nonporous, monolithic surface that will not provide a breeding site for organisms. Exposed outside corners in high traffic areas as well as on lower wall surfaces may have stainless steel facings or guards to prevent impact

**FIGURE 8** A modular wall system has been installed in a manner that provides a smooth surface for cleaning. The fit of the components and the method of sealing are important when a modular wall is selected. (Courtesy of Portafab.)

damage to the wall. This is particularly true when gypsum board construction is used. Corner and wall guards should extend from the floor to at least the 4-ft height. Traditionally the clean room has been white throughout as an indication of the clean nature of the facility and to identify it as a special work space. Other colors may be used in the clean room to provide an interesting environment as long as the materials of construction do not contribute particles to the air stream and will withstand the sanitizing agents and procedures used in the facility (Figure 8).

***Doors*** Entry should be through air locks to maintain clean-room pressure differentials. Emergency exit doors should incorporate a panic-bar mechanism (or a similar emergency opening device) with alarms for exit only. Emergency exit doors must be secured in a manner that prevents entry from the outside yet permits exiting from within. All doors should include essentially air-tight seals. Neoprene seals are generally acceptable. Brush-type door seals are not recommended. Foam rubber door seals are not recommended as these have been found to quickly deteriorate and shed particles. All personnel doors and swinging equipment doors should include self-closing mechanisms. Manual and automatic sliding doors may be useful when space is an issue or to facilitate movement between spaces of similar cleanliness class for personnel whose hands are otherwise engaged. As the mechanism of such doors can generate particles, a design specifically intended for clean-room application should be selected.

***Ceilings*** The ceiling finish should be similar to that used on the walls. The requirements for sanitizing typically address the ceiling as well as the walls and ceiling material and finish selection should reflect this. Suspended ceilings using an inverted-T grid and lay-in panels may have a place in that part of the clean-room suite not subjected to the rigors of frequent sanitizing and where the possibility of trapped

spaces to support organism growth is not considered an issue (Figure 9). When suspended panel ceilings are used, the panels must be securely clipped or sealed in place to prevent movement due to air pressure changes.

Modular wall systems designed for biopharmaceutical applications frequently have a "walk-on" ceiling designed using materials and finish similar to the wall. A rounded, easy-to-clean intersection between ceiling and walls should be a feature of the clean-room ceiling design, whether modular or stick built. Monolithic (seamless) ceilings can be installed using inverted-T grid supports and gypsum panels (Figure 10). This design permits incorporation of filtration and lighting into what is essentially a monolithic ceiling.

**FIGURE 9** A suspended ceiling utilizing lay-in panels and lay-in lighting troffers. A variety of cleanable materials can be used for the panels. The lay-in lights should be of a design that will provide appropriate service based on the cleaning protocol to be used. (Courtesy of CleanTek.)

**FIGURE 10** An area of HEPA filters is installed above a process machining. Tear drop lighting is used to permit maximum filter coverage. A monolithic ceiling construction of gypsum panels suspended from a framework. The panels are finished with an epoxy coating compatible with cleaning/sterilization procedures. (Courtesy of CleanTek.)

**FIGURE 11** The process area is subjected to substantial chemical action due to the sterilizing protocol. It has a troweled epoxy, easy-to-clean finish. (Courtesy of Dex-O-Tex.)

***Floors*** Commonly used floor finishes for biopharmaceutical clean rooms include sheet vinyl installed using heat-welded or chemically fused seams to provide a seamless surface. Troweled epoxy and epoxy paint (Figure 11) have also found wide use. Compatibility of the floor material with solvents, chemicals, and cleaning agents to be used in the room must be considered. A minimum 4-in. cove at the junction of floor and walls is recommended to facilitate cleaning. Some modular wall systems have a recess or offset that permits sheet vinyl to be installed in a manner that creates a seamless junction between floor and wall. When a stick-built approach is used, care should be taken to design cleanable intersections of walls and floors (Figure 12).

### 3.5 HVAC System

***Air Side*** The clean-room HVAC system must be designed to maintain the required particulate cleanliness, temperature, humidity, and positive pressure at the expected outside environmental extremes and during the expected worst-case use operations. Rapid recovery from upset conditions such as door openings and contaminant-generating events is also a consideration. The high cost of conditioning outside air suggests that as much air as possible be recirculated. Recirculated air should be high-efficiency particulate air (HEPA) filtered in those spaces requiring a cleanliness classification in accordance with ISO 14644. Air that may be hazardous to health, even after HEPA filtration, should be exhausted after appropriate treatment. The required quantity of make-up air is calculated based on process exhaust plus air leakage from the clean room. A rate of two air changes per hour for clean room pressurization may be used in the absence of a more detailed calculation of air

**FIGURE 12** The wall system used in the facility incorporates a monolithic sheet vinyl flooring junction between floor and wall face. Note the coving run up the wall around the edges to provide a smooth surface for cleaning. (Courtesy of Portafab.)

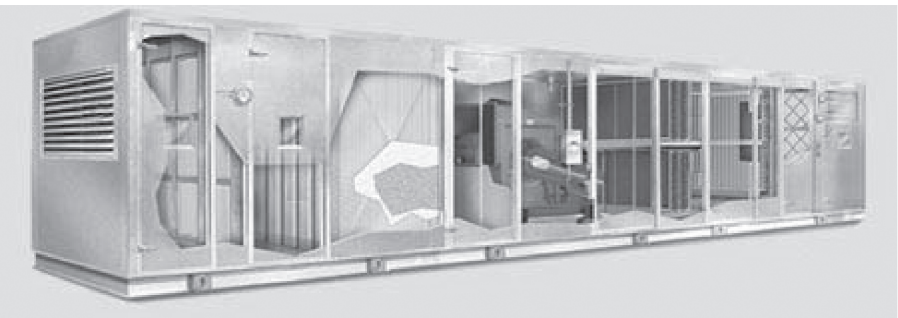

**FIGURE 13** The air handler has several stages of filtration combined with heating, cooling, humidification, and dehumidification capability. (Courtesy of Air Enterprises.)

leakage. Make-up air should be drawn from the outdoors, conditioned, and filtered as necessary before being introduced into the clean-room recirculation air stream. Care should be taken to ensure that make-up air intakes are not drawing in contaminated air.

The potential for cross-contamination is an issue that should be addressed. A flexible manufacturing facility is one in which a variety of products can be manufactured simultaneously. If the facility has a single air-handling system, the likelihood of materials from one space intruding into an adjacent space is high. For this reason each filling or compounding operation, or operation where noncompatible product can be expected to be picked up by the air stream, should be served by its own air-handling system (Figure 13). Isolated systems will minimize the possibility of cross-contamination. This can be a costly option and should not be undertaken

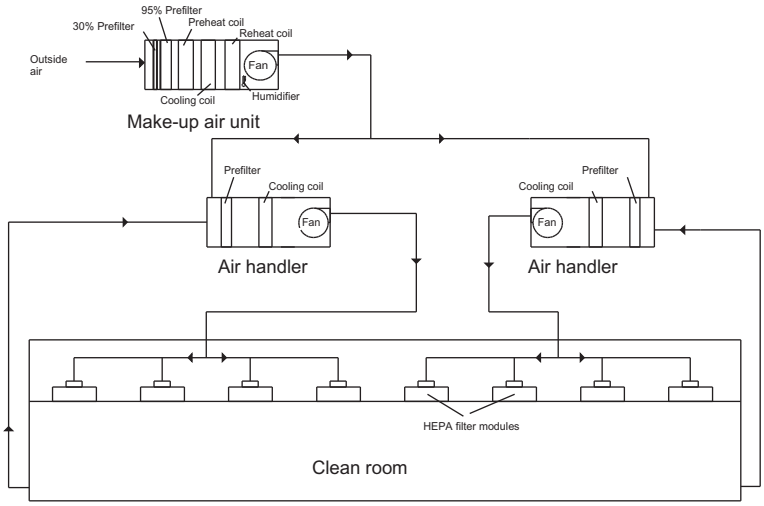


**FIGURE 14** Non-unidirectional clean-room with lay-in HEPA filter modules.

lightly. The current use of the plant and the anticipated future use should be assessed before a blanket decision that may lead to costly duplicated systems is made.

***Filtration*** The filtration system for a biopharmaceutical clean room typically consists of several stages of filters. Prefilters are selected, sized, and installed to maximize the life of the final HEPA filters. With proper selection of prefilters, the final HEPA filters should not require replacement within the life of the filter media and seal materials, a period of several years (perhaps as long as 10–15 years). Make-up air is commonly filtered by a low-efficiency [30% as set by the American Society of Heating, Refrigerating, and Air Conditioning Engineers (ASHRAE)] prefilter followed by an intermediate- (60% ASHRAE) or high-efficiency (95% ASHRAE) final filter (Figure 14). A screen should be included at the make-up air inlet to keep out pests and large debris. The make-up air is then directed to the recirculating air handler which also may have a low-efficiency prefilter, although prefiltration of recirculated clean-room air is often omitted because of its high cleanliness level even after having passed through the clean room. The air is then directed through HEPA filters into the clean room. HEPA filters must be a minimum of 99.97% efficient on 0.3-μm particles in accordance with military standard Mil-F-51068 or the Institute of Environmental Science and Technology IEST-RP-CC001. Note that the filtration system for an unrated "controlled area" is the same, except that the HEPA filter stage may be omitted. Refer to Figure 15.

***Filter Location*** HEPA filters may be installed in a facility either within an air handler or at the inlet to a plenum above the clean room or in the clean room ceiling. High-velocity HEPA filters, that is, filters with a face velocity up to 500 ft/min, are frequently installed in air handlers serving Class 100,000 clean rooms and are also used in make-up air handlers. Where hazardous materials may be trapped by the filters a "bag-in–bag-out" filter arrangement, such as that depicted in Figure 16, may

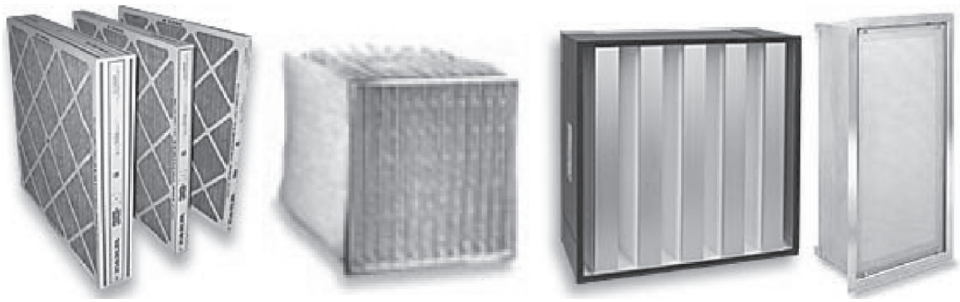

**FIGURE 15** Several panel-type filters commonly used as prefilters in air handlers. Second from left is a high-dust-loading filter available in ASHRAE efficiency as high as 95% frequently used in make-up air handlers. If HEPA filtration of the make-up air is required, the high-velocity duct-mounted HEPA filter third from the left is appropriate. It can tolerate face velocities up to 500 fpm, compared to the standard HEPA, which is usually designed for face velocity on the order of 90–100 fpm. A standard HEPA designed for bio-pharma facility ceiling installation is shown at right. (Courtesy of of CamFil.)

**FIGURE 16** The "bag-in–bag-out" filter unit contains a HEPA filter and permits personnel to change the filter without coming into contact with possibly hazardous materials that may have been filtered from the air. (Courtesy of Flanders Filters Inc.)

be employed. Figure 17 shows a schematic arrangement with HEPA filters installed in the air handler. During the design phase care should be taken to provide access to both the upstream and downstream face of these filters to permit periodic challenging and leak testing.

To provide HEPA filtered air over a limited area within a larger controlled space, a ceiling-mounted pressure plenum may be used. This plenum has an air distribution means at its lower face that permits air to be introduced in a unidirectional manner over the critical process area. Refer to Figure 18.

HEPA filters are installed at the upper face of the pressure plenum and the plenum is pressurized with filtered air. The ceiling-mounted HEPA filters have a face velocity up to 100–120 ft/min. This is somewhat higher than the HEPA filters serving the rest of the clean room. The filters are commonly supplied with air by a

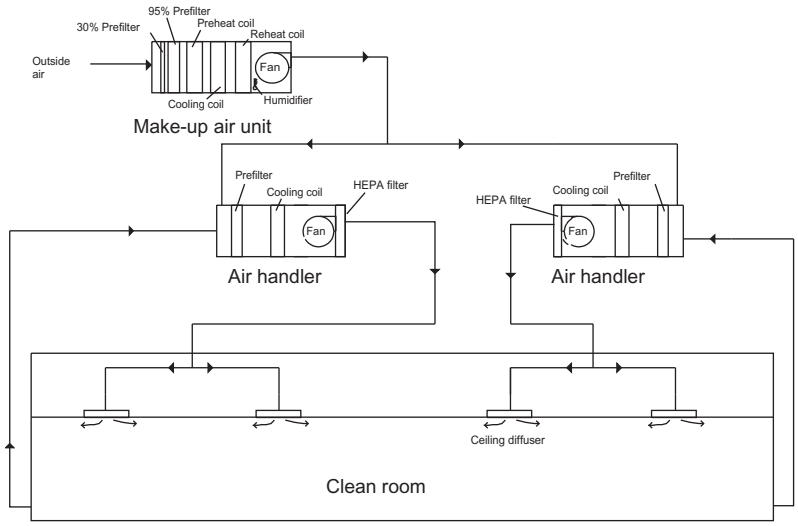


**FIGURE 17** Non-unidirectional clean-room with air handler mounted HEPA filters.

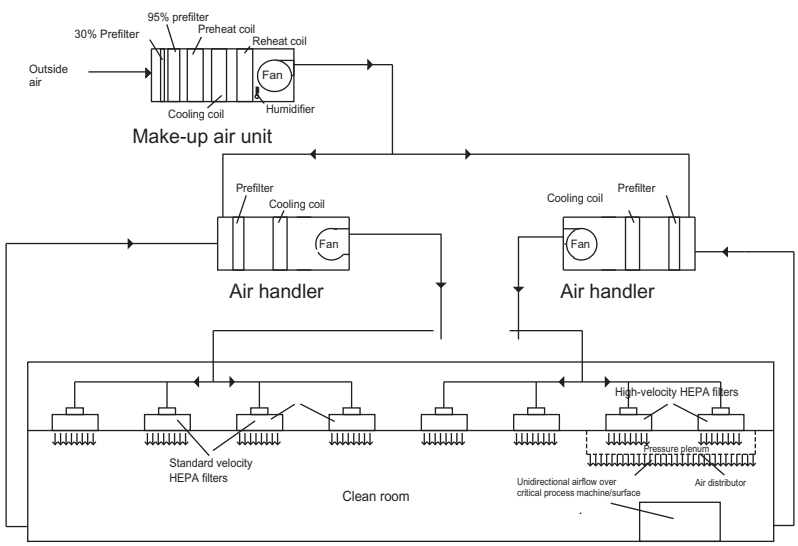


**FIGURE 18** Non-unidirectional clean-room with critical area unidirectional flow plenum.

duct distribution network consisting of rectangular or round trunk ducts and flexible or rigid round branch ducts. Full coverage, typical for ISO 5 (Class 100) clean rooms, or partial coverage, for higher class (less stringent) clean rooms, can be accomplished using 2 × 4-ft lay-in HEPA filter modules installed in the ceiling.

### 3.6 Clean-Room Testing

ISO 14644 describes methodology and instrumentation for particle counting in the clean room. The tests described there are the basis for assigning a cleanliness rating

to the facility. IEST-RP-CC006 similarly provides a procedure for particle counting but goes beyond that to a full series of tests that can be conducted to determine the effectiveness of the clean-room design and operability. The determination of which tests should be run is up to the clean-room end user. As a minimum, particle counting, room pressurization, and filter leakage tests should be run. Other tests dealing with airflow patterns, temperature, humidity, lighting, and sound levels are available. The array of tests selected is determined by the owner based on the effect the various design parameters will have on the product. The data obtained in acceptance tests become baseline data against which future testing is compared to determine if clean-room performance is changing over time. Ongoing periodic monitoring of the facility will ensure that clean-room performance degradation is identified as it occurs. Pass–fail criteria are not part of the ISO standards but are to be developed on a case-by-case basis by the end user of the facility. These standards become part of the operational protocol of the facility.

The clean-room testing described here is part of the commissioning or validation process wherein all equipment in the facility is run, tested, and observed to ensure it is working as designed.

### 3.7 Utilities

Biopharmaceutical clean-rooms typically house process equipment requiring utilities such as pure water, electricity, vacuum, and clean compressed air. The source of these utilities is usually outside the clean room. During the design phase a utility matrix is developed, in conjunction with end users and equipment manufacturers, identifying all equipment and the utilities required. This is the basis for determining the capacity of the utility systems as well as the point-of-use location of specific utilities.

When bringing the utilities to the point of use, care should be taken to ensure that the clean room is not compromised. A clean construction protocol should be implemented and wall, ceiling, and floor penetrations, if needed, should be flashed and sealed in such a manner as to prevent contaminants from entering the clean room. Such entry points should also be smoothly sealed to ensure that there are no crevices to harbor organisms. Drains should be avoided in the clean room wherever possible. When this is not possible, the drains should be covered when not in use with a means specifically designed for biopharmaceutical clean-room application. Such means are tight, smooth, cleanable, and corrosion resistant.

In small facilities an individual pipeline may be run from outside the facility to the point of use. In large facilities a utility chase (Figure 19) that enables major utility lines to be brought to the vicinity of process tools may be provided. Final hook-up between the chase and point of use then becomes a relatively simple, minimally intrusive procedure. The utility chase concept is also beneficial in facilities that undergo frequent retrofit or upgrade.

## 4 CONSTRUCTION PHASE: CLEAN BUILD PROTOCOL

Ongoing experience has demonstrated that an aggressive clean construction protocol program is generally not required for biopharmaceutical facilities that do not

**FIGURE 19** The utility chase is located between two clean rooms. Major utility lines are installed within the chase and hook-up lines for local pieces of process equipment are connected through the clean-room walls. A major benefit of this arrangement is that installers need not be fully garbed in clean-room attire to access the utility lines. (Courtesy of CleanTek.)

carry a cleanliness rating. Where cleanliness classifications less stringent than Class 10,000 are used, standard construction techniques followed by careful cleanup and wipe-down within the clean space have proven quite acceptable. Cleanliness levels of Class 1000 or Class 10,000 are achievable shortly after startup and maintainable thereafter. For cleanliness levels of Class 100 a somewhat more restrictive protocol is required. Once a facility is up and running, any intrusion into clean areas for retrofit work should be done in conjunction with some level of clean build protocol in place, dependent on the rating of the facility and the degree of disruption encountered during the retrofit project.

The levels of clean construction described herein can provide a practical means of meeting operational cleanliness goals in a cost-effective fashion. Each project, whether new construction or retrofit of an existing process, should have as part of it an evaluation of the required elements of the build clean protocol to be employed. The information provided below is broad and can act as a template for the protocol put in place for a specific project.

A key to successful clean construction is the appointment of an individual as a clean-room monitor who is well versed in the clean-room construction protocol. That person is charged with maintaining a clean environment and monitoring the activities of all personnel within the clean area during the construction phase *and* is concerned with maintaining budget and schedule goals. The clean-room monitor should have the confidence to make "real-world" decisions supporting the "spirit" of the protocol as well as the "letter."

### 4.1 General

All clean-facility construction, while employing standard construction techniques, should be accomplished in a manner that does not create excessive particulate contamination. A temporary lay-down area within the building adjacent to the clean area should be set aside for storage of clean construction components. All tools used for clean construction should be in an "as-new" condition and be cleaned and inspected prior to use. The pass–fail criteria for tool and material inspection is "no visible dirt."

Cleanup within the clean area at the end of each shift should consist of broom cleaning *and* vacuum cleaning the floor with a clean vacuum, that is, a vacuum with a HEPA filter (99.97% efficient on 0.3-μm particles).

Clean-facility construction materials should be left in an outer shipping wrap until moved to the temporary lay-down area, where they should then be unwrapped and wiped down before being moved into the clean space.

Adherence to these guidelines will make final clean-up faster and acceptable start-up and certification/validation more certain. While a goal of clean construction is rapid start-up and certification/validation, a long-range goal is the maintenance of the facility cleanliness without intrusion, over an extended period of time, of contaminants deposited during construction due to a poor protocol or improper implementation of the protocol.

Appendix A and Appendix B offer a template for working in a clean environment as well as clean-room cleaning procedures. Procedures should be modified with caution to suit a particular project.

### 4.2 Level I Clean Construction

Level I clean construction is used for all areas with a cleanliness rating of Class 1000 (ISO 6) or higher (less stringent), including those spaces within which clean processes are conducted in minienvironments/isolators and those unrated areas identified as being "controlled environments."

Standard construction techniques are used until the clean-room envelope is completed, HEPA filters with protective film in place are installed and air handlers are ready to start. The clean envelope consists of clean-room walls, ceiling, and floor. Prior to starting the air handlers, a thorough clean-up of the space within the clean envelope is accomplished as described in Appendix B, 2A–2J. Following clean-up and start-up of the clean-space air-handling system, particle counts should quickly drop to well within operational requirements.

Once the clean room is operational, as described above, additional construction related to process equipment installation and facility modification within the clean room can be done in compliance with the Guidelines in Appendix A.

### 4.3 Level II Clean Construction

This is used for construction of those areas rated at Class 100 (ISO 5) employing a 100% HEPA filter ceiling. Generally standard construction techniques should be used. The clean-room envelope includes walls, ceiling, floor, return ductwork, supply fans, and supply ductwork.

All ductwork sections should be cleaned and sealed with plastic wrap at the time of fabrication until just prior to installation or start-up to prevent contaminants from accumulating inside air-handling passageways. The sections of ductwork should be unsealed only as required for installation. Open ends of ducts and fans should remain sealed until connecting duct is about to be installed. A final isopropyl alcohol (IPA) wipe-down of all interior duct sections and fan surfaces should be done immediately prior to installation.

When general construction of the clean room is completed, steps 2A–2J of Appendix B describing coarse cleaning can be implemented. Following coarse cleaning the protective film can be removed from the HEPA filters and the air-handling system started.

Successful completion of the cleaning process described above will indicate that installation of process equipment may begin. Note that procedures described in Appendix A should be followed. After installation of all process equipment or when the clean room is to be prepared for certification, steps 2K–2O of Appendix B for final wipe-down can be followed.

A black-and-white felt rub-down test is performed to demonstrate adequate cleanliness of the interior clean-envelope surfaces. This test consists of both black-and-white felt being wiped over any surface for 1 m linear distance with a firm hand pressure. No residue should be visible on the cloth. Each cloth should be 60 cm square black or white static-free natural fiber felt folded with cut edges inside to a 25-cm square. The cut edges should be sealed with an approved latex sealant.

## 5 MAINTENANCE

To maximize the life and effectiveness of the facility, it must be maintainable. The facility should be designed to permit ongoing day-to-day preventive maintenance of the mechanical systems and, should a failure occur, permit needed repairs to be made in an expeditious manner. Perhaps of equal importance is the janitorial maintenance required to keep the facility suitable for pharmaceutical manufacturing. Proper janitorial maintenance begins with the design of the facility and evolves into an operational protocol, personnel training, and effective implementation.

In the design phase it is important to provide sufficient access to mechanical and process equipment to enable preventive maintenance procedures to be carried out with minimum effort. Typically manufacturer's installation instructions offer guidelines as to how much space should be left open around equipment to permit removal of critical components. One driver of construction cost is floor space. Making a space as small as possible to house an operation presumably will result in first-cost savings. If the space does not provide sufficient access for lubrication, filter changes, belt adjustments, and the like, there is a strong possibility that this preventive maintenance will be ignored. A predictable result is shortened equipment life and the disappearance of any first-cost savings that may have been realized. If there is a major equipment failure that requires replacement of an inaccessible component, the cost associated with knocking down a wall to gain equipment access will very likely negate first-cost savings.

Storage of maintenance items should be identified early in the design process. Spare-parts storage, janitorial supply storage, janitors' sink closets, repair work shops,

and storage space for consumable maintenance items (e.g. air filters) will require floor space in the facility design. Frequently tools are dedicated to the clean facility or are required specifically for unique process equipment and must also have a storage area. Accommodation of these items is an important part of the planning process.

A requirement of a clean facility is that the cleaning materials should be specifically intended for use in a "clean" operation, should be kept in good ("like new") repair, and should not be used in other, nonclean, areas of the facility. Using general cleaning materials manned by the "house" janitorial staff will invariable introduce more contamination into the clean portions of the facility than it removes.

A central housekeeping vacuum is very useful in keeping contamination under control. While "wet-and-dry" versions of the central vacuum are available, the manner in which each is to be used should be carefully reviewed to ensure that it is in keeping with the sanitary requirements of the facility. A common housekeeping procedure addresses spills with local clean-up and uses a dry-type central vacuum for dry particulate contaminants.

## APPENDIX A: GUIDELINES FOR CONSTRUCTION PERSONNEL AND WORK TOOLS IN A CLEAN ROOM

### 1.0 General requirements

A. Makeup will not be allowed inside the clean room.
B. Smoking will not be allowed in or around the clean room.
C. Tobacco chewing will not be allowed in the clean room.
D. Paper or paper by-products will not be allowed in the clean room except clean-room approved paper and pens.
E. Prints or papers will be allowed only if totally laminated in plastic and cleaned with isopropyl alcohol prior to entry.
F. Lead pencils will not be allowed in the clean room. Ball point pens only.
G. Clean-room garments, to include shoe covers, coveralls, and head cover, will be worn within the clean room.
H. Clean-room garments will not be unfastened or unzipped while inside the clean room.
I. No writing will be allowed on the clean-room garments.
J. Food and drink will not be allowed in the clean room.
K. Combing of hair will not be allowed in the clean room or gowning area.
L. Stepping on chairs, work benches, test equipment or process equipment is not allowed.
M. Damaged garments (rips, worn booties, torn gloves) will be replaced immediately. Do not wait for a convenient time. DO IT NOW!
N. Tool pouches are not allowed in a clean room.
O. All work areas and adjacent areas will be vacuumed after completion of work and prior to leaving the clean room.
P. **ALWAYS** wash hands before entering the clean room to remove residues from food, smoke, and/or other sources.

### 2.0 Personnel

A. All personnel working inside a clean room will be required to follow all dress codes associated with the particular clean space.

B. Street clothes or company uniforms will be allowed as standard undergarments provided they are well maintained and clean. No such garments will be allowed that are soiled with grease, dirt, or any detectable stains.
C. Any garments producing excessive fibers (such as fuzzy sweaters) will not be allowed as an undergarment.
D. Standard safety shoes (or other specified footware) will be required. Shoe covers must be worn.
E. Bare feet, socks, and stockings are not allowed inside booties.
F. Coats, lunches, and private items will not be allowed inside the clean room.

**3.0 Gowning procedure**

A. Each individual is responsible for knowing and using the correct method of gowning prior to entering the clean room. (See Figure 20.)
   1. Clean shoes prior to entering the gowning room.
   2. The order of dress should be as follows:
      a. Shoe covers
      b. Hairnet/beard cover (required after final cleaning)
      c. Hood
      d. Coveralls
      e. Face cover (required after final cleaning)
      f. Gloves (required after final cleaning)
B. Ensure that hoods are tucked inside neck opening of coveralls and pants legs are tucked and snapped inside booties. Garments are to be snapped

**FIGURE 20** Clean-room garments are intended to keep contaminants from entering the clean room. In a critical environment the "bunny suit" shown may be required. In a less critical environment a lab coat may suffice. The clean-room construction protocol should identify the type of garment that will be required at various stages of construction and for process equipment installation. (Courtesy of Terra Universal.)

closed at the neck, wrist and ankle opening and sleeves tucked inside gloves.

C. All head hair must be covered at all times.
D. Do not allow garments to touch the floor while dressing or undressing.
E. Avoid leaning on walls, lockers, or other personnel at all times. **DO NOT** place feet on benches.
F. The order of undress should be as follows:
   1. Gloves
   2. Coveralls
   3. Face cover and hood
   4. Shoe covers
G. If you will be reentering the clean room, unsoiled garments may be hung for reuse; gloves are not to be reused.

**4.0 Work tools, parts, and equipment**

A. All tools and equipment used in a clean room should be in like-new condition.
B. All parts will be removed from their shipping container prior to cleaning and introduction into the clean room. **NO PAPER PRODUCTS** will be allowed inside the clean room.
C. All tools, parts, and equipment will be properly cleaned prior to entering the clean room. Minimum cleaning should be a total wipe-down with isopropyl alcohol, using certified clean-room wipes, to assure that the last wipe does not leave visible residue on the wipe. Parts should be blown off *outside* the clean room using filtered nitrogen when available.
D. All parts and equipment should be sent through the equipment wipe-down area (material air lock) and not carried through the gowning area.

**5.0 Working in a Clean Room**

A. A major concern when working in a clean room is the generation of particles of the size that cannot be seen and spreading these particles throughout the clean room. Every possible precaution must be taken to contain these contaminants and protect the clean-room environment. Everything that is done as a standard operation must be analyzed to determine if it will adversely affect the cleanroom. *If you have any concerns, ask the clean-room monitor before you damage the environment and incur unnecessary clean-up cost.*
B. *All procedures must be reviewed with the clean-room monitor to ensure compliance with clean-room operation practices.* All procedures that can generate particles should be done outside the clean-room whenever possible. In the listing below all prohibited procedures are subject to review by the clean-room monitor. The intent is to get the job done; however, some preplanning with the clean-room monitor can result in a positive result and a clean facility.
   1. Drilling:
      a. All power drills will be wrapped to encapsulate any contamination generated during operation.
      b. Drills may be operated in sealed enclosures equipped with an exhaust vacuum.

c. Surface to be drilled will be tented or vacuumed to prevent the spread of contamination.

2. Grinding: NO GRINDING will be allowed in the clean room.
3. Welding: NO WELDING will be allowed inside the clean room.
4. Soldering: May be allowed after total review.
5. Painting: NO PAINTING will be allowed after the start-up of the clean room.
6. Sanding/filing: Will be allowed only within a properly tented space.
7. Cutting: Will be allowed with clean-room approved vacuums removing particles created.

## APPENDIX B: CLEANING THE CLEANROOM

**1.0 Final cleaning:** During this cleaning phase, the clean-room proper should be prepared for certification to the appropriate cleanliness level. Extreme care must be exercised by all those involved in this procedure to minimize the potential for contamination.

**2.0 Procedure**

A. Secure the entrance to the clean space "envelope," the gowning area, and the entrance from the gowning room to the clean room with locks or limited access via card keys. Post a restriction notice: YOU ARE ENTERING A CONTROLLED, PARTICULATE-FREE ENVIRONMENT, CONTACT _____ FOR PERMISSION TO ENTER.

B. Perform two coarse cleanings of the clean room. Each cleaning should include the following:
   1. Wipe-down of the HEPA filter grid, all lights, walls, floors, windows, and all exposed interior surfaces. This will include any outlet boxes or floor/wall recesses.
   2. Wipe-down should be by clean potable water and mild nonphosphate detergent using clean, lint-free cloths approved by the clean-room monitor.
   3. A second wash-down should commence using clean potable water in the same manner.
   4. Floors should be scrubbed and polished using a floor-polishing machine. No wax is to be used.

C. Partitions and floors should be washed to maintain a dust-free condition.

D. Access into the clean room should be restricted to discourage infiltration of outside particulates.

E. Tacky mats 3 ft by 6 ft should be installed inside the entrance of the gowning area as well as at the entrance to the clean room.

F. Foot covers should be worn by all personnel entering the clean room after the second coarse cleaning is complete.

G. Caulking crew should be assigned and commence work after the second coarse cleaning is completed. They should complete all caulking as required by specification.

H. Simultaneously with the caulking procedure, air handlers and associated ducts and plenums should be checked for cleanliness.

I. The HEPA filter protective film should be removed. Air-handling equipment should be activated and the clean space pressurized to maintain a positive static pressure.
J. From this point forward, clean-room garments and head covers should be worn by all personnel entering the clean space.
K. Commence with the first of two final wipe-downs. Nonshedding clean-room wipes (saturated with isopropyl alcohol) or tacky wipes should be used. *All exposed surfaces should be wiped.*
L. A particle counter should be installed in the clean space and samples taken at several control points over the next 48 hours. A steady decrease in the particle count over time should be achieved.
M. If particle counts stabilize at a level above that desired, a search for filter leakage will be required.
N. If the search for filter leakage fails to find a leak, the entire area should be recleaned as described for the final wipe-down.
O. Once the air standards are achieved, final air balance can begin followed by clean-room certification testing.

## BIBLIOGRAPHY

U.S. Food and Drug Administration, Washington, DC

21 CFR Part 210, Current good manufacturing practice in manufacturing, processing, packing, or holding of drugs.

21 CFR Part 211, Current good manufacturing practice for finished pharmaceuticals.

Institute of Environmental Sciences and Technology, Rolling Meadows, IL

IEST-RP-CC001.4: *HEPA and ULPA Filters*, Nov. 7, 2005.

IEST-RP-CC002.2: *Unidirectional Flow Clean-Air Devices*, Jan. 19, 1999.

IEST-RP-CC003.3: *Garments Systems Considerations for Cleanrooms and Other Controlled Environments*, Aug. 11, 2003.

IEST-RP-CC004.3: *Evaluating Wiping Materials Used in Cleanrooms and Other Controlled Environments*, Aug. 23, 2004.

IEST-RP-CC005.3: *Gloves and Finger Cots Used in Cleanrooms and Other Controlled Environments*, May 1, 2003.

IEST-RP-CC006.3: *Testing Cleanrooms*, Aug. 30, 2004.

IEST-RP-CC008-84: *Gas-Phase Adsorber Cells*, Nov. 1, 1984.

IEST-RP-CC012.1: *Considerations in Cleanroom Design*, Mar. 1, 1998.

IEST-RP-CC013-86-T: *Equipment Calibration or Validation Procedures*, Nov. 1, 1986.

IEST-RP-CC016.2: *The Rate of Deposition of Nonvolatile Residue in Cleanrooms*, Nov. 15, 2002.

IEST-RP-CC018.3: *Cleanroom Housekeeping: Operating and Monitoring Procedures*, Jan. 1, 2002.

IEST-RP-CC019.1: *Qualifications for Organizations Engaged in the Testing and Certification of Cleanrooms and Clean-Air Devices*, Jan. 23, 2006.

IEST-RP-CC023.2: *Microorganisms in Cleanrooms*, Jan. 31, 2006.

IEST-RP-CC026.2: *Cleanroom Operations*, July 21, 2004.

IEST-RP-CC027.1: *Personnel Practices and Procedures in Cleanrooms and Controlled Environments*, Apr. 1, 1999.

IEST-RP-CC028.1: *Minienvironments*, Sept. 1, 2002.

IEST-RP-CC034.2: *Hepa and ULPA Filter Leak Tests*, June 23, 2005.

IEST-STD-CC1246D: *Product Cleanliness Levels and Contamination Control Program*, Jan. 1, 2002.

International Organization for Standardization (ISO) Standards

ISO 14644-1: Classification of air cleanliness.

ISO 14644-2: Specifications for testing and monitoring to prove continued compliance.

ISO 14644-3: Test methods.

ISO 14644-4: Design, construction and start-up.

ISO 14644-5: Operations.

ISO 14644-6: Terms and definitions.

ISO 14644-7: Separative devices (clean air hoods, gloveboxes, isolators, and minienvironments).

ISO 14644-8: Classification of airborne molecular contamination.

ISO 14698-1: Bicontamination control—General principles.

ISO 14698-2: Biocontamination control—Evaluation and interpretation of biocontamination data.

ISO 14698-3: Biocontamination control—Methodology for measuring the efficiency of processes of cleaning and/or disinfection of inert surfaces bearing biocontaminated wet soiling or biofilms.

# SOLID DOSAGE FORMS

Barbara R. Conway
*Aston University, Birmingham, United Kingdom*

**Contents**

1 Biopharmaceutics Classification System
2 Systematic Formulation Development
3 Standard and Compressed Tablets
4 Excipients in Solid Does Formulations
  4.1 Diluents
  4.2 Binders
  4.3 Lubricants
  4.4 Glidants and antiadherents
  4.5 Disintegrants
  4.6 Superdisintegrants
  4.7 Added Functionality Excipients
  4.8 Colorants
  4.9 Interactions and Safety of Excipients
5 Coated Tablets
  5.1 Sugar-Coated Tablets
  5.2 Compression Coating and Layered Tablets
  5.3 Film-Coated Tablets
  5.4 Tablet Wrapping or Enrobing
6 Hard and soft gelatin capsules
  6.1 Hard-Shell Gelatin Capsules
  6.2 Manufacture of Hard Gelatin Shells
  6.3 Hard Gelatin Capsule Filling
  6.4 Soft Gelatin Capsules
  6.5 Manufacture of Soft Gelatin Capsules
  6.6 Dissolution Testing of Capsules
7 Effervescent Tablets
  7.1 Manufacture of Effervescent Tablets
8 Lozenges
  8.1 Chewable Lozenges

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

9 Chewable Tablets
9.1 Testing of Chewable Tablets
10 Chewing Gums
10.1 Composition of Chewing Gum
10.2 Manufacture of Chewing Gum
10.3 Drug Release from Chewing Gums
10.4 Applications for Chewing Gums
11 Orally Disintegrating Tablets
11.1 Dissolution Testing of ODTs
12 Solid Dosage Forms for Nonoral Routes
References

Drug substances are most frequently administered as solid dosage formulations, mainly by the oral route. The drug substance's physicochemical characteristics, as well the excipients added to the formulations, all contribute to ensuring the desired therapeutic activity. Tablets and capsules are the most frequently used solid dosage forms, have been in existence since the nineteenth century, and are unit dosage forms, comprising a mixture of ingredients presented in a single rigid entity, usually containing an accurate dose of a drug. There are other types of solid dosage forms designed to fulfill specific delivery requirements, but they are generally intended for oral administration and for systemic delivery. The major solid oral dosage form is the tablet, and these can range from relatively simple, single, immediate-release dosage forms to complex modified-release systems. Tablets offer advantages for both patients and manufacturers (Table 1). Most tablets are intended to be swallowed whole and to rapidly disintegrate and release drug in the gastrointestinal tract. Tablets are classified by their route of administration or their function, form, or manufacturing process. For example, some tablets are designed to be placed in the oral cavity and to dissolve there or to be chewed before swallowing, and there are many kinds of formulation designed for sustained or controlled release (Table 2).

Solid dose formulations, including tablets, must have the desired release properties coupled with manufacturability and aesthetics and must involve rational formulation design. The dose of the drug and its solubility are important considerations

**TABLE 1 Advantages of Tablets as a Dosage Form**

| |
|---|
| Easy to handle |
| Variety of manufacturing methods |
| Can be mass produced at low cost |
| Consistent quality and dosing precision |
| Can be self-administered |
| Enhanced mechanical, chemical, and microbiological stability compared to liquid dosage forms |
| Tamperproof |
| Lend themselves to adaptation for other profiles, e.g., coating for sustained release |

**TABLE 2 Types of Solid Dosage Form**

| Formulation type | Description |
|---|---|
| Immediate-release tablet/capsule | Intended to release the drug immediately after administration |
| Delayed-release tablet/capsule | Drug is not released until a physical event has occurred, e.g., change in pH |
| Sustained-release tablet/capsule | Drug is released slowly over extended time |
| Soluble tablets | Tablet is dissolved in water prior to administration |
| Dispersible tablet | Tablet is added to water to form a suspension prior to administration |
| Effervescent tablet | Tablet is added to water, releasing carbon dioxide to form a effervescent solution |
| Chewable tablet | Tablet is chewed and swallowed |
| Chewable gum | Formulation is chewed and removed from the mouth after a directed time |
| Buccal and sublingual tablets | Tablet is placed in the oral cavity for local or systemic action |
| Orally disintegrating tablet | Tablet dissolves or disintegrates in the mouth without the need for water |
| Lozenge | Slowly dissolving tablet designed to be sucked |
| Pastille | Tablet comprising gelatin and glycerine designed to dissolve slowly in the mouth |
| Hard gelatin capsule | Two-piece capsule shell that can be filled with powder, granulate, semisolid or liquid |
| Soft gelatin capsule (softgel) | One-piece capsule containing a liquid or semisolid fill |

in the design of the formulation as are the type of dosage form and its method of preparation.

## 1 BIOPHARMACEUTICS CLASSIFICATION SYSTEM

Dissolution of the drug must occur before or on reaching the absorption site before absorption can occur, and generally water-soluble drugs do not exhibit formulation difficulties. For poorly water-soluble drugs, the absorption rate may be dictated by the dissolution rate, and, if dissolution is slow, bioavailability may be compromised. The solubility of a drug should, therefore, be considered along with its dose when designing formulations, and unsuitable biopharmaceutical properties is the major reason for the failure of new drugs.

In 1995, the Biopharmaceutics Classification System (BCS) was devised to classify drugs based on their aqueous solubility and intestinal permeability [1]. According to the BCS, drug substances are classified as follows [2]:

Class I: high permeability, high solubility
Class II: high permeability, low solubility
Class III: low permeability, high solubility
Class IV: low permeability, low solubility

A dose solubility volume can be defined for all drugs (i.e., the volume required to dissolve the dose). A drug substance is considered highly soluble when the highest dose strength is soluble in ≤250 mL water over a pH range of 1–7.5. A drug substance is considered highly permeable when the extent of absorption in humans is determined to be ≥90% of an administered dose, based on mass balance or in comparison to an intravenous reference dose. A drug product is considered to be rapidly dissolving when ≥85% of the labeled amount of drug substance dissolves within 30 min using U.S. Pharmacopeia (USP) apparatus I or II in a volume of ≤900 mL buffer solutions.

It was recognized that dissolution rate has a negligible impact on bioavailability of highly soluble and highly permeable (BCS class I) drugs when dissolution of their formulation is sufficiently rapid. As a result, various regulatory agencies including the U.S. Food and Drug Administration (FDA) now allow bioequivalence of formulations of BCS class I drugs to be demonstrated by in vitro dissolution (often called a biowaiver) [3]. Therefore, one of the goals of the BCS is to recommend a class of immediate-release (IR) solid oral dosage forms for which bioequivalence may be assessed based on in vitro dissolution tests.

## 2 SYSTEMATIC FORMULATION DEVELOPMENT

Systematic development approaches are needed to gather a full and detailed understanding of marketable formulations in order to satisfy the requirements of regulatory bodies and to provide a research database. Efficient experimental design using in-house or commercial software packages can ensure quality while avoiding expensive mistakes and lost time. Information from various categories such as the properties of the drug substance and excipients, interactions between materials, unit operations, and equipment are required [4]. Design of experiments (DOE) and statistical analysis have been applied widely to formulation development. Using DOE facilitates evaluation of all formulation factors in a systematic and timely manner to optimize the formulation and manufacturing process. Abbreviated excipient evaluation techniques such as Plackett–Burman design can be applied to minimize the number of experiments and identify critical components or processes. Optimization processes can then be applied. When the formulation and manufacturing processes of a pharmaceutical product are optimized by a systematic approach, the scale-up and processes validation can be very efficient because of the robustness of the formulation and manufacturing process.

Innovations in statistical tools such as multivariate analysis, artificial intelligence, and response surface methodology have enabled rational development of formulations, and such methods allow formulators to identify critical variables without having to test each combination.

## 3 STANDARD AND COMPRESSED TABLETS

The simplest tablet formulations are uncoated products that are made by direct compression or compression following wet or dry granulation. They are a versatile drug delivery system and can be intended for local action in the gastrointestinal

(GI) tract or for systemic effects. General design criteria for tablets are accuracy and uniformity of drug content, stability of the drug candidate and the formulation, optimal dissolution and availability for absorption (whether immediate or extended release), and patient acceptability in terms of organoleptic properties and appearance. Flocculant, low-density drugs can be difficult to compress and formulate into tablets. This is a particular issue with drugs of low potency. Also some poorly water-soluble, poorly permeable drugs or highly metabolized drugs cannot be given this way. Additionally, local irritant effects can be harmful to the mucosa of the GI tract.

Tablets are a popular dosage form due to their simplicity and economy of manufacture, relative stability, and convenience in packaging, shipping, and storage. For the patient, uniformity of dose, blandness of taste, and ease of administration ensure their popularity. Thus, the purpose of the formulation and the identification of suitable excipients are of primary importance in the development of a successful formulation. A well-designed formulation should contain, within limits, the stated quantity of active ingredient, and it should be capable of releasing that amount of drug at the intended rate and site. Tablets need to be strong enough to withstand the rigors of manufacture, transport, and handling, and they need to be of acceptable size, taste, and appearance. A typical manufacturing process for a tablet product includes weighing, milling, granulation and drying, blending and lubrication, compression, and coating. Each processing step involves several process parameters. For a given formulation, all processing steps should be thoroughly evaluated so that a robust manufacturing process can be defined, and DOE can be applied effectively to optimize this process.

Direct compression is a simple process being more economical and less stressful to ingredients in terms of heat and moisture, However, there are limitations governed by the physical properties of the ingredients, and raw materials must be carefully controlled. It is difficult to form directly compressed tablets containing high-dose and poorly compactible drugs. Granulation can be employed to improve the compaction characteristics of the powder. Granulation can also improve flow properties and reduce the tendency for segregation of the mix due to a more even particle size and bulk density. Granules can be produced by either wet or dry methods based on the stability of the drug and excipients.

Although the basic mechanical process of producing tablets by compression has not changed, there has been much work on improving tableting technology [5]. Understanding of the physical and mechanical properties of powders and the compaction process has improved and will continue to improve product design while increases in the speed and uniformity of action of tableting presses improve the process.

## 4 EXCIPIENTS IN SOLID DOSE FORMULATIONS

In addition to the active ingredients, solid oral dosage forms will also contain a range of substances called excipients. The role of excipients is essential in ensuring that the manufacturing process is successful and that the quality of the resultant formulation can be guaranteed. The appropriate selection of excipients and their relative concentrations in the formulation is critical in development of a successful product.

**TABLE 3 Excipients Used in Solid Dose Formulations**

| Excipient Category | Examples |
|---|---|
| Fillers/diluents | Lactose, sucrose, glucose, microcrystalline cellulose |
| Binders | Polyvinyl pyrrolidone, starch, gelatin, cellulose derivatives |
| Lubricants | Magnesium stearate, stearic acid, polyethylene glycol, sodium chloride |
| Glidants | Fine silica, talc, magnesium stearate |
| Antiadherents | Talc, cornstarch, sodium dodecylsulfate |
| Disintegrants and superdisintegrants | Starch, sodium starch glycollate, cross-linked polyvinyl pyrrolidone |
| Colorants | Iron oxide, natural pigments |
| Flavor modifiers | Mannitol, aspartame |

Although they are often categorized as inert, preformulation studies can determine the influence of excipients on stability, bioavailability, and processability. Excipients are categorized into groups according to their main function, although some may be multifunctional, and examples of common excipients used in the manufacture of tablets and capsule are detailed in Table 3.

### 4.1 Diluents

An inert substance is frequently added to increase the bulk of a tablet for processing and handling. The lower weight limit for formulation of a tablet is usually 50 mg. Ideally, diluents should be chemically inert, nonhygroscopic, and hydrophilic. Having an acceptable taste is important for oral formulations, and cost is always a significant factor in excipient selection.

Lactose is a common diluent in both tablets and capsules, and it fulfils most of these criteria but is unsuitable for those who are lactose intolerant. Various lactose grades are commercially available which have different physical properties such as particle size distribution and flow characteristics. This permits the selection of the most suitable material for a particular application. Usually, fine grades of lactose are used for preparation of tablets by wet granulation or when milling during processing is carried out, since the fine size permits better mixing with other formulation ingredients and facilitates more effective action of the binder [6].

Diluents for direct compression formulations are often subject to prior processing to improve flowability and compression, for example, amorphous lactose, but this can contribute to reduced stability especially under high-humidity conditions when reversion to the crystalline form is more likely [6].

Microcrystalline cellulose (Avicel) is purified partially depolymerized cellulose, prepared by treating α-cellulose with mineral acids. In addition to being used as a filler, it is also used as dry binder and disintegrant in tablet formulations. Depending on the preparation conditions, it can be produced with a variety of technical specifications depending on particle size and crystallinity. It is often used as an excipient in direct compression formulations but can also be incorporated as a diluent for tablets prepared by wet granulation, as a filler for capsules and for the production of spheres.

Diluents, although commonly presumed inert, do have the ability to influence the stability or bioavailability of the dosage form. For example, dibasic calcium phosphate (both anhydrous and dihydrate forms) is the most common inorganic salt used as a filler–binder for direct compression. It is particularly useful in vitamin products as a source of both calcium and phosphorous. Milled material is typically used in wet-granulated or roller-compacted formulations. The coarse-grade material is typically used in direct compression formulations. It is insoluble in water, but its surface is alkaline and it is therefore incompatible with drugs sensitive to alkaline pH. Additionally, it may interfere with the absorption of tetracyclines [7].

### 4.2 Binders

Binders (or adhesives) are added to formulations to promote cohesiveness within powders, thereby ensuring that the tablet remains intact after compression as well as improving the flow by forming granules. A binder should impart adequate cohesion without retarding disintegration or dissolution. Binders can be added either as a solution or as a dry powder. Binders added as dry powders are mixed with other powders prior to agglomeration, dissolving in water or solvent added during granulation, or added prior to compaction. Solution binders can be sprayed, poured, or mixed with the powder blend for agglomeration and are generally more effective, but further dry binder can be added prior to tableting. Starch, gelatin, and sugars are used along with gums, such as acacia and sodium alginate, and are used at concentrations between 2 and 10% w/w. Celluloses and polyvinyl pyrrolidone (PVP) are also utilized, often as dry binders.

### 4.3 Lubricants

Lubricants can reduce friction between the tablet and the die wall during compression and ejection by interposing an intermediate film of low shear strength at the interface between the tablet and the die wall. The best lubricants are those with low shear strength but strong cohesive tendencies perpendicular to the line of shear [8]. The hydrophobic stearic acid and stearic acid salts, primarily magnesium stearate, are the most widely used and are included at concentrations less than 1% w/w in order to minimize any deleterious effects on disintegration or dissolution. They should be added after the disintegrant to avoid coating it and preferably at the final stage prior to compression to ensure mixing time is kept to a minimum. Hydrophilic lubricants such as polyethylene glycols (PEGs) and lauryl sulfates can be used to redress the issues with dissolution but may not be as efficient as their hydrophobic counterparts.

### 4.4 Glidants and Antiadherents

Like lubricants, glidants are fine powders and may be required for tablet compression at high production speeds to improve the flow properties of the material into the die or during initial compression stages. They are added in the dry state immediately prior to compression and, by virtue of their low adhesive potential, reduce the friction between particles. Colloidal silica is popular, as are starches and talc.

Antiadherents can also be added to a formulation that is especially prone to sticking to the die surface (or picking). Water-insoluble lubricants such as magnesium stearate can be used as antiadherents, as can talc and starch.

### 4.5 Disintegrants

Disintegrants are added to a formulation to overcome the cohesive strength imparted during compression, thus facilitating break up of the formulation in the body and increasing the surface area for dissolution. They can be either intragranular, extragranular, or both, and there is still a lack of understanding concerning their precise mechanism of action. On contact, disintegrants can draw water into the tablet, swelling and forcing the tablet apart. Starch, a traditional and still widely used disintegrant, will swell when wet, although it has been reported that its disintegrant action could be due to capillary action [6]. Levels can be increased beyond the normal 5% w/w to 15–20% w/w if a rapid disintegration is required. Surfactants can also act as disintegrants promoting wetting of the formulation, and sodium lauryl sulfate can be combined with starch to increase effectiveness.

Tablet disruption following production of carbon dioxide is another mechanism used to enhance disintegration. This uses a mixture of sodium bicarbonate and a weak acid such as citric acid or tartaric acid and is exploited for effervescent formulations.

### 4.6 Superdisintegrants

Compared to the more traditional starch, newer disintegrants are effective at much lower levels and comprise three groups: modified starches, modified cellulose, and cross-linked povidone. Their likely mechanism of action is a combination of proposed theories including water wicking, swelling, deformation recovery, repulsion, and heat of wetting [9]. Superdisintegrants are so called because of the relatively low levels required (2–4% w/w). Sodium starch glycollate (Primojel, Explotab) is made by cross-linking potato starch and can swell up to 12-fold in less than 30 s. Crospovidone is completely insoluble in water, although it rapidly disperses and swells in water, but does not gel even after prolonged exposure. It rapidly exhibits high capillary activity and pronounced hydration capacity with little tendency to form gels and has a greater surface area–volume ratio compared to other disintegrants. Micronized versions are available to improve uniformity of mix. Croscarmellose sodium, a cross-linked polymer of carboxymethyl cellulose sodium is also insoluble in water, although it rapidly swells to 4–8 times its original volume on contact with water [6].

### 4.7 Added Functionality Excipients

Adverse physiochemical and mechanical properties of new chemical entities prove challenging for formulation development. There is an increasing demand for faster and more efficient production processes. Also, biotechnological developments and various emerging protein-based therapies are broadening the definition for excipient products. Although the description of excipients from inactive ingredients is shifting toward functionally active materials and will continue to grow in this area, the intro-

duction of improved versions of long-existing excipients is probably the more successful development. New single-component and coprocessed products have been introduced, for example, filler–binders. In addition, there have been advances in the understanding of how such substances act and hence how they can be optimally designed. Excipients for use in direct compression product forms or physically or chemically modified excipients used in relatively new drug delivery systems, such as patches or inhalation systems, are examples of these developments.

### 4.8 Colorants

Colorants are frequently used in uncoated tablets, coated tablets, and hard and soft gelatin capsules. They can mask color changes in the formulation and are used to provide uniqueness and identity to a commercial product. Concerns over the safety of coloring agents in formulations generally arise from adverse effects in food substances. Colorants are therefore subject to regulations not associated with other pharmaceutical excipients. Legislation specifies which colorants may be used in medicinal products and also provides for purity specifications. The number of permitted colors has decreased in recent years, and a list of approved colorants allowed by regulatory bodies can vary from country to country.

Colorants can be divided into water-soluble dyes and water-insoluble pigments. Some of the insoluble colors or pigments can also provide opacity to tablet coatings or gelatin shells, which can promote stability of light-sensitive active materials. Pigments such as the iron oxides, titanium dioxide, and some of the aluminum lakes are especially useful for this purpose.

Water-soluble dyes are usually incorporated within the granulation process to ensure even distribution throughout the formulation, but there can be an uneven distribution due to migration of the dye during drying processes. Therefore, water-soluble dyes can also be adsorbed into a carrier such as starch or lactose and dry blended prior to the final mix. Water-insoluble pigments are more popular in direct compression and are dry blended with the other ingredients.

Lakes are largely water-insoluble forms of common synthetic water-soluble dyes and are prepared by adsorbing the sodium or potassium salt of a dye onto a very fine substrate of hydrated alumina, followed by treatment with a further soluble aluminum salt. The lake is then purified and dried. Lakes are frequently used in coloring tablet coatings since they are more stable and have greater opacity than a water-soluble dye [6].

### 4.9 Interactions and Safety of Excipients

Because there is such a wide selection available, rational choice of the necessary excipients and their concentration is required. Consideration must also be given to cost, reliability, availability, and international acceptability. Although generally considered inert, formulation incompatibility of excipients is also necessary. Lactose, for example, can react with primary and secondary amines via its aldehyde group by Maillaird condensation reaction [6], and calcium carbonate is incompatible with acids due to acid–base chemical reaction and with tetracyclines due to complexation. Additionally, excipients can contribute to the instability of the active substance through moisture distribution.

Despite the importance of drug–excipient compatibility testing, there is no generally accepted method available for this purpose. After identification of any major known incompatibilities, a compatibility screen needs to be proposed. Issues such as sample preparation, storage conditions, and methods of analysis should be addressed and factorial design applied to reduce the number if tests required. Drug–excipient compatibility studies can be performed with minimal amounts of materials. Usually, small amounts of each material are weighed into a glass vial, in a ratio representative of the expected ratio in the formulation. The vials can be sealed as is or with additional water, either in an air environment or oxygen-free (nitrogen head space) environment, and stored in the presence or absence of ambient light, at various temperatures. Factorial or partial factorial design experiments can be set up to determine important binary and multiple component interaction factors. This information helps determine which excipients should be avoided and whether oxidation or light instability in the formulation is a consideration. Controls consisting of the active pharmaceutical ingredient (API) alone in the various conditions also should be run to determine whether the API is susceptible alone or must have the mediating excipient or water additives for instability.

## 5 COATED TABLETS

Tablets are often coated to protect the drug from the external environment, to mask bitter tastes, add mechanical strength, or to enhance ease of swallowing. A coating can also be used for aesthetic or commercial purposes, improving product appearance and identity.

### 5.1 Sugar-Coated Tablets

Sugar coating can be beneficial in masking taste, odors, and colors. It is useful in protecting against oxidation, and sugar coating was once very common due to its aesthetic results and cheapness of materials. Use has declined in recent years due to the complexity of the process and skills required, but advances in technology have led to a resurgence in popularity. Typical excipients used are sucrose (although this can be substituted with low-calorie alternatives), fillers, flavors, film formers, colorants, and surfactants. It is usually carried out in tumbling coating pans and comprises several stages.

The first sealing stage uses shellac or cellulose acetate phthalate, for example, to prevent moisture from reaching the tablet core. This has to be kept to a minimum to prevent impairment of drug release. The subcoating is an adhesive coat of gum (such as acacia or gelatin) and sucrose used to round off the edges, and the tablets can be dusted with substances such as kaolin or calcium carbonate to harden the coating. A smoothing coat is built up in layers using 70% v/v sucrose syrup and often opacifiers such as titanium dioxide, and the tablets are dried between each application. A colorant is added to the final few layers and followed with a final polishing step which can make further embossing difficult. The coating is relatively brittle, prone to chipping or cracking, and there is a substantial increase in weight, up to 50%, and size of the product.

### 5.2 Compression Coating and Layered Tablets

A coating can be applied by compression using specially designed tablet presses. The same process can be used to produce layered tablets which can comprise two or even three layers if complete separation of the ingredients is required. This process is used when physical separation of ingredients is desired due to incompatibility or to produce a repeat-action product. The formulation can also be designed to provide an immediate and a slow-release component. Release rates can be controlled by modification of the geometry, the composition of the core, and the inclusion of a membrane layer.

The technique involves using a preliminary compression step to produce a relatively soft tablet core which is then placed in a large die containing coating material. Further coating material is added and the content compressed. A similar light compression is used for the production of layers and a final main compression step used to bind the layers together.

### 5.3 Film-Coated Tablets

Film coating, although most often applied to tablets, can also be used to coat other formulations including capsules. Film coating imparts the same general characteristics as sugar coating but weight gain is significantly less (typically up to 5%), it is easier to automate, and it has capacity to include organic solvents if required. The main methods involved are modified conventional coating pans, side-vented pans, and fluid-bed coating. Celluloses are often used as film-forming polymers, as detailed in Table 4, and usually require addition of a compatible plasticizer as glass transition temperatures are higher than the temperatures used in the process. Polyethylene glycol, propylene glycol, and glycerol are commonly used, and colorants and opacifiers can also be added to the coating solution. Specialist coatings such as Opadry fx and Opaglos 2 can be used to give a high gloss finish to improve brand identity and consumer recognition.

### 5.4 Tablet Wrapping or Enrobing[1]

Recent innovations in tablet coating include the use of gelatin and non-animal-derived coatings for tablets that require formulation of a pre-formed film that is then used to encapsulate the product (e.g., Banner's Soflet Gelcaps or Bioprogress' Nrobe technology). The coated formulations are tamper evident and can be designed with different colors for branding purposes. They are reported to be preferred by patients due to their ease of swallowing and superior taste- and odor-masking properties. An alternative is the Press-fit Geltabs system, which uses a high-gloss gelatin capsule shell to encapsulate a denser caplet formulation.

## 6 HARD AND SOFT GELATIN CAPSULES

Capsules are solid oral dosage forms in which the drug is enclosed within a hard or soft shell. The shell is normally made from gelatin and results in a simple, easy-to-swallow formulation with no requirement for a further coating step. They can be

[1]See http://www.banpharm.com/technologiesSofletGelcap.cfm and http://www.fmcmagenta.com/NRobe/tabid/145/Default.aspx.

**TABLE 4 Polymers Commonly Used in Film Coating of Tablets**

| Polymer | Soluble in | Description |
|---|---|---|
| Methylcellulose (MC) | Cold water, GI fluids, and organic solvents | Low-viscosity grades best for aqueous films |
| Ethylcellulose (EC) | Organic solvents and GI fluids (insoluble in water) | Used in combination with water-soluble agents for immediate release |
| Hydroxyethyl cellulose (HEC) | Water and GI fluids | Similar to MC with clear solutions |
| Hydroxypropyl cellulose (HPC) | Cold water, GI fluids, and polar solvents | Results in a tacky coat and used in combination to promote adhesion |
| Hydroxypropylmethyl cellulose (HPMC) | Cold water, GI fluids, and alcohols | Excellent film former, low-viscosity grades best |
| Sodium carboxymethyl cellulose | Water and polar solvents | Cannot be used if presence of moisture is a problem |
| Methylhydroxyethyl cellulose (MHEC) | Water and GI fluids | Similar to HPMC but less soluble in organic solvents |
| Povidone (PVP) | Water, GI fluids, alcohol, and isoproplyl alcohol (IPA) | Can lead to tackiness during drying, often brittle and hygroscopic |
| PEGs | Water, GI fluids, some organic solvents | High molecular weights best for film forming and low molecular weights used as plasticizer; can be waxy |
| Enteric coatings such as poly(methacrylates) or cellulose acetate phthalate | Soluble at elevated pHs | Used for delayed-release formulations |

*Source*: Adapted from refs. 5 and 10.

either hard or soft depending on the nature of the capsule shell, with soft capsules possessing a flexible, plasticized gelatin film. Hard gelatin capsules are usually rigid two-piece capsules that are manufactured in one procedure and packed in another totally separate operation, whereas the formulation of soft gelatin capsules is more complex but all steps are integrated.

There is a growing interest in using non-animal-derived products for formulation of the capsule shells to address cultural, religious, and dietary requirements. HPMC (e.g., V-caps, Quali-VC, Vegicaps) and pullulan shells (NPCaps) and starch are alternatives.

### 6.1 Hard-Shell Gelatin Capsules

Although the challenges of powder blending, homogeneity, and lubcricity exist for capsules as for tablets, they are generally perceived to be a more flexible formulation as there is no requirement for the powders to form a robust compact. This means that they may also be more suitable for delivery of granular and beadlike formulations, fragile formulations that could be crushed by the normal compaction step. They are commonly employed in clinical trials due to the relative ease of blinding and are useful for taste masking.

Capsules are usually more expensive dosage forms than an equivalent tablet formulation due to the increased cost of the shells and the slower production rates. Even with modern filling equipment, the filling speeds of capsule machines are much slower than tablet presses. However, increased costs can be offset by avoiding a granulation step. Capsules, although smoother and easy to swallow, also tend to be larger than corresponding tablet formulations, potentially leading to retention in the esophagus. Humidity needs to be considered during manufacture and storage, with moisture leading to stickiness and desiccation causing brittleness. Cross-linking of gelatin in the formulation can also lead to dissolution and bioavailability concerns.

Capsule excipients are similar to those required for formulation of tablets and include diluents, binders, disintegrants, surfactants, glidants, lubricants, and dyes or colorants. The development of a capsule formulation follows the same principles as tablet development, and consideration should be given to the same BCS issues. The powder for encapsulation can comprise simple blends of excipients or granules prepared by dry granulation or wet granulation. There is a reduced requirement for compressibility, and often the flow properties are not as critical as in an equivalent tablet formulation. The degree of compressibility required is the major difference, and capsules can therefore be employed when the active ingredient does not possess suitable compression characteristics.

### 6.2 Manufacture of Hard Gelatin Shells

Gelatin is a generic term for a mixture of purified protein fractions obtained either by partial acid hydrolysis (type A gelatin) or by partial alkaline hydrolysis (type B gelatin) of animal collagen. Type A normally originates from porcine skin while B is usually derived from animal bones, and they have different isoelectric points (7.0–9.0 and 4.8–5.0, respectively) [6]. The protein fractions consist almost entirely of amino acids joined together by amide linkages to form linear polymers, varying in molecular weight from 15,000 to 250,000. Gelatin can comprise a mixture of both types in order to optimize desired characteristics, with bone gelatin imparting firmness while porcine skin gelatin provides plasticity. Gelatin Bloom strength is measured in a Bloom gelometer, which determines the weight in grams required to depress a standard plunger in a 6.67% w/w gel under standard conditions. Bloom strength and viscosity are the major properties of interest for formulation of capsules, and Bloom strength of 215–280 is used in capsule manufacture.

Gelatin is commonly used in foods and has global regulatory acceptability, is a good film former, is water soluble, and generally dissolves rapidly within the body without imparting any lag effect on dissolution. Gelatin capsules are strong and robust enough to withstand the mechanical stresses involved in the automated filling and packaging procedures.

In addition to gelatin, the shells may contain colorants, opacifiers, and preservatives (often parabens esters). There are eight standard capsule sizes, and the largest capsule size considered suitable for oral use is size 0 (Table 5).

To manufacture the shells, pairs of molds, for the body and the cap, are dipped into an aqueous gelatin solution (25–30% w/w), which is maintained at about 50 °C in a jacketed heating pan. As the pins are withdrawn, they are rotated to distribute the gelatin evenly and blasted with cool air to set the film. Drying is carried out by

**TABLE 5 Capsule Size and Corresponding Volume or Weight of Fill**

| Size | Volume (mL) | Fill weight[a] (g) |
|---|---|---|
| 000 | 1.37 | 1.096 |
| 00 | 0.95 | 0.760 |
| 0 | 0.68 | 0.544 |
| 1 | 0.50 | 0.400 |
| 2 | 0.37 | 0.296 |
| 3 | 0.30 | 0.240 |
| 4 | 0.21 | 0.168 |
| 5 | 0.13 | 0.104 |

*Source*: Adapted from http://capsugel.onlinemore.info/download/BAS192-2002.pdf.

[a]Assumes a powder density of 0.8 g/cm$^3$.

passing dry air over the shell as heating temperatures are limited due to the low melting point of gelatin. The two parts are removed from the pins, trimmed, and joined using a prelock mechanism. The external diameter of the body is usually wider at the open end than the internal diameter of the cap to ensure a tight fit. They can be made self-locking by forming indentations or grooves on the inside of both parts so that when they are engaged, a positive interlock is formed (e.g., Posilok, Conicap, Loxit).

Alternatively, they may be hermetically sealed using a band of gelatin around the seam between the body and the cap (Qualicaps). This can be applied without the application of heat and provide a tamper-evident seal. LEMS (liquid encapsulation microspray sealing) used in Licaps is a more elegant seal in which sealing fluid (water and ethanol) is sprayed onto the joint between the cap and body of the capsule. This lowers the melting point of gelatin in the wetted area. Gentle heat is then applied which fuses the cap to the body of the Licaps capsule. The moisture content of manufactured shells is 15–18% w/w and levels below 13% will result in problems with the capsule filling machinery. Therefore, capsules are stored and filled in areas where relative humidity is controlled to between 30 and 50%.

### 6.3 Hard-Gelatin Capsule Filling

The filling material must be compatible with the gelatin shell and, therefore, deliquescent or hygroscopic materials cannot be used. Conversely, due the moisture content in the capsule shells, they cannot be used for moisture-sensitive drugs. All ingredients need to be free of even trace amounts of formaldehyde to minimize cross-linking of gelatin.

Powders and granules are the most common filling materials for hard-shell gelatin capsules, although pellets, tablets, pastes, oily liquids, and nonaqueous solutions and suspensions have been used. Filling machines are differentiated by the way they measure the dose of material and range in capacity from bench-top to high-output, industrial, fully automated machines. Those that rely on the volume of the shell are known as capsule dependent, whereas capsule-independent forms measure the quantity to be filled in a separate operation. The simplest dependent method of filling is leveling where powder is transferred directly from a hopper to the capsule

**TABLE 6 Liquid Excipients Compatible with Hard Gelatin Capsules**

| Peanut oil | Paraffin oil |
|---|---|
| Hydrogenated peanut oil | Cetyl alchohol |
| Castor oil | Cetostearyl alcohol |
| Hydrogenated castor oil | Stearyl alcohol |
| Fractionated coconut oil | Stearic acid |
| Corn oil | Beeswax |
| Olive oil | Silica dioxide |
| Hydrogenated vegetable oil | Polyethylene glycols |
| Silicone oil | Macrogol glycerides |
| Soya oil | Poloxamers |

*Source*: Adapted from http://www.capsugel.com/products/licaps_oil_chart.php.

body, aided by a revolving auger or vibration. Additional powder can be added to fill the space arising, and the fill weight depends on the bulk density of the powder and the degree of tamping applied.

Most automated machinery is of the independent type and compresses a controlled amount of powder using a low compression force (typically 50–200 *N*) to form a plug. Most are piston-tamp fillers and are dosator or dosing disk machines. The powder is passed over a dosing plate containing cavities slightly smaller than the capsule diameter, and powder that falls into the holes is tamped by a pin to form a plug. This can be repeated until the cavity is full and the plugs (or slugs) are ejected into the capsule shells. The minimum force required to form a plug should be used to reduce slowing of subsequent dissolution.

In the dosator method, the plug is formed within a tube with a movable piston that controls the dosing volume and applies the force to form the plug. The dose is controlled by the dimensions of the dosator, the position of the dosator in the powder bed, and the height of the powder bed. Fundamental powder properties to ensure even filling are good powder flow, lubricity, and compressibility. The auger or screw method, now largely surpassed, uses a revolving archimedian screw to feed powder into the capsule shell.

A liquid fill can be useful when manufacturing small batches if limited quantities of API are available. Liquid fills also offer improved content uniformity for potent, low-dose compounds and can reduce dust-related problems arising with toxic compounds. Two types of liquid can be filled into hard gelatin capsules: nonaqueous solutions and suspensions or formulations that become liquid on application of heat or shear stress. These require hoppers with heating or stirring systems. For those formulations that are liquid at room temperature, the capsule shells need to be sealed after filling to prevent leakage of the contents and sticking of the shells. It is essential to ensure the liquid is compatible with the shell (Table 6).

### 6.4 Soft-Gelatin Capsules

Soft gelatin capsules are hermetically sealed one-piece capsules containing a liquid or a semisolid fill. Like liquid-filled hard capsules, although the drug is presented in a liquid formulation, it is enclosed within a solid, thus combining the attributes

of both. Soft gelatin capsules (softgels) offer a number of advantages including improved bioavailability, as the drug is presented in a solubilized form, and enhanced drug stability. Consumer preference regarding ease of swallowing, convenience, and taste can improve compliance, and they offer opportunities for product differentiation via color, shape, and size and product line extension. The softgels can be enteric coated for delayed release. They are popular for pharmaceuticals, cosmetics, and nutritional products, but highly water-soluble drugs and aldehydes are not suitable for encapsulation in softgels. Formulations are tamper evident and can be used for highly potent or toxic drugs. However, they do require specialist manufacture and incur high production costs.

### 6.5 Manufacture of Soft Gelatin Capsules

The shell is primarily composed of gelatin, plasticizer, and water (30–40% wet gel), and the fill can be a solution or suspension, liquid, or semisolid. The size of a softgel represents its nominal capacity in minims, for example, a 30 oval softgel can accommodate 30 minims (or 1.848 $cm^3$). Glycerol is the major plasticizer used, although sorbitol and propylene glycol can also be used. Other excipients are dyes, pigments, preservatives, and flavors. Up to 5% sugar can be added to give a chewable quality. Most softgels are manufactured by the process developed by Scherer [11]. The glycerol–gelatin solution is heated and pumped onto two chilled drums to form two separate ribbons (usually 0.02–0.04 in. thick) which form each half of the softgel. The ribbons are lubricated and fed into the filling machine, forcing the gelatin to adopt the contours of the die. The fill is manufactured in a separate process and pumped in, and the softgels are sealed by the application of heat and pressure. Once cut from the ribbon, they are tumble-dried and conditioned at 20% relative humidity.

Fill solvents are selected based on a balance between adequate solubility of the drug and physical stability. Water-miscible solvents such as low-molecular-weight PEGs, polysorbates, and small amounts of propylene glycol, ethanol, and glycerin can be used. Water-immiscible solvents include vegetable and aromatic oils, aliphatic, aromatic, and chlorinated hydrocarbons, ethers, esters, and some alcohols. Emulsions, liquids with extremes of pH (<2.5 and >7.5), and volatile components can cause problems with stability, and drugs that do not have adequate stability in the solvents can be formulated as suspensions. In these instances, the particle size needs to be carefully controlled and surfactants can be added to promote wetting.

Vegicaps soft capsules from Cardinal Health are an alternative to traditional softgels, containing carageenan and hydroxyproyl starch. As with traditional soft gelatin capsules, the most important packaging and storage criterion is for adequate protection against extremes of relative humidity. The extent of protection required also depends on the fill formulation and on the anticipated storage conditions.

### 6.6 Dissolution Testing of Capsules

In general, capsule dosage forms tend to float during dissolution testing with the paddle method. In such cases, it is recommended that a few turns of a wire helix around the capsule be used [12]. Inclusion of enzymes in the dissolution media must be considered on case-by-case basis. A Gelatin Capsule Working Group (including

participants from the FDA, industry, and the USP) was formed to assess the noncompliance of certain gelatin capsule products with the required dissolution specifications and the potential implications on bioavailability [13]. The working group recommended the addition of a second tier to the standard USP and new drug and abbreviated new drug applications (NDA/ANDA) dissolution tests: the incorporation of enzyme (pepsin with simulated gastric fluid and pancreatin with simulated intestinal fluid) into the dissolution medium. If the drug product fails the first tier but passes the second tier, the product's performance is acceptable. The two-tier dissolution test is appropriate for all gelatin capsule and gelatin-coated tablets and the phenomenon may have little significance in vivo.

## 7 EFFERVESCENT TABLETS

Effervescence is the reaction in water of acids and bases to produce carbon dioxide, and effervescent tablets are dissolved or dispersed in water before administration. Advantages of effervescent formulations over conventional formulations are that the drug is usually already in solution prior to ingestion and can therefore have a faster onset of action. Although the solution may become diluted in the GI tract, any precipitation should be as fine particles that can be readily redissolved. Variability in absorption can also be reduced. Formulations can be made more palatable and there can be improved tolerance after ingestion. Thus, the types of drugs suited to this formulation method are those that are difficult to digest or are irritant to mucosa. Analgesics such as paracetamol and aspirin and vitamins are common effervescent formulations. The inclusion of buffering agents can aid stability of pH-sensitive drugs. There is also the opportunity to extend market share and to deliver large doses of medication.

Effervescents comprise a soluble organic acid and an alkali metal carbonate salt. Citric acid is most commonly used for its flavor-enhancing properties. Malic acid imparts a smoother after taste and fumaric, ascorbic, adipic, and tartaric acids are less commonly used [14]. Sodium bicarbonate is the most common alkali, but potassium bicarbonate can be used if sodium levels are a potential issue with the formulation. Both sodium and potassium carbonate can also be employed. Other excipients include water-soluble binders such as dextrose or lactose, and binder levels are kept to a minimum to avoid retardation of disintegration. All ingredients must be anhydrous to prevent the components within the formulation reacting with each other during storage.

Lubricants such as magnesium stearate are not used as their aqueous insolubility leads to cloudy solutions and extended disintegration times. Spray-dried leucine and PEG are water-soluble alternatives [15, 16]. Both artificial and natural sweeteners are used and an additional water-soluble flavoring agent may also be required. If a surfactant is added to enhance wetting and dissolution, the addition of an antifoaming agent may also be considered [17].

### 7.1 Manufacture of Effervescent Tablets

Essentially, effervescent formulations are produced in the same way as conventional tablets, although due to the hygroscopicity and potential onset of the effervescence

reaction in the presence of water, environmental control of relative humidity and water levels is of major importance during manufacture. A maximum of 25% relative humidity (RH) at 25°C is required. Closed material-handling systems can be used or open systems with minimum moisture content in the ventilating air.

A dry method of granulation is preferred as no liquid is involved but may not always be possible. Wet granulation can be carried out under carefully controlled conditions using two separate granulators for the alkaline and acid components. Water can be added at 0.1–1.0% w/w, and it initiates a preeffervescent reaction. The cycle is stopped by drying, usually by transfer into a preheated fluidized-bed dryer. Fluid-bed spray granulation is a process wherein granulation and drying are simultaneous and can be useful for effervescent formulations. Water (or a binder solution) is sprayed onto the mixture, which is suspended in a stream of hot, dry air. Organic solvents can also be employed for granulation avoiding the need for water and are useful for heat-labile formulations, although complex handling equipment is required.

Effervescent formulations must contain less than 0.3% w/w water and are often quite large. Sticking due to insufficient lubrication can be overcome by adaptation of punches for external lubrication or using flat-faced punches with disks of poly (tetrafluoroethylene) (PTFE). Poor lubrication can also be the cause of poor flow characteristics, and this can be addressed by using a constant level powder feed system. The tablets should be stored in tightly closed containers or moisture-proof packs. In tube arrangements, dry air is added prior to sealing and desiccants to reduce enclosed moisture levels once the pack has been opened. Foil packaging should be heavy gauge to minimize risk of holes, and the surrounding pocket should be large enough to hold the tablets but minimize inclusion of air.

In-process quality control is of major importance for these formulations as are stability testing and stress testing of packaging. Tablet disintegration and dissolution are of prime importance, and disintegration should be carried out using representative conditions. Hardness and friability are also important as these large tablets tend to chip easily. Common areas for problems are that the packaging permits entry of water, the seal is compromised or that the excipients can react with each other.

## 8 LOZENGES

Lozenges are tablets that dissolve or disintegrate slowly in the mouth to release drug into the saliva. They are easy to administer to pediatric and geriatric patients and are useful for extending drug form retention within the oral cavity. They usually contain one or more ingredient in a sweetened flavored base. Drug delivery can be either for local administration in the mouth, such as anaesthetics, antiseptics, and antimicrobials or for systemic effects if the drug is well absorbed through the buccal lining or is swallowed. More traditional drugs used in this dosage form include phenol, sodium phenolate, benzocaine, and cetylpyridinium chloride. Decongestants and antitussives are in many over-the-counter (OTC) lozenge formulations, and there are also lozenges that contain nicotine (as bitartrate salt or as nicotine polacrilex resin), flurbiprofen (Strefen), or mucin for treatment of dry mouth (A.S Saliva Orthana).

Lozenges can be made by molding or by compression at high pressures, often following wet granulation, resulting in a mechanically strong tablet that can dissolve

in the mouth. Compressed lozenges (or troches) differ from conventional tablets in that they are nonporous and do not contain disintegrant. As the formulation is designed to release drug slowly in the mouth, it must have a pleasant taste, smoothness, and mouth feel. The choice of binder, filler, color, and flavor is therefore most important. The binder is particularly important in ensuring retardation of dissolution and pleasant mouth feel. Suitable binders include gelatin, guar gum, and acacia gum. Sugars such as sucrose, dextrose, and mannitol are preferred to lactose, and xylitol is often included in sugar-free formulations. In order to ensure adequate sweetness and taste masking, artificial sweeteners including aspartame, saccharin, and sucralose are also included subject to regulatory guidelines.

Other variations include hard-candy-type and soft or chewable lozenges. Most hard-candy-type lozenges contain sugar, corn syrup, acidulant, colorant, and flavors. They are made by heating sugars and other ingredients together and then pouring the mixture into a mold. Corn syrup combined with sucrose and dextrose can form an amorphous glass suitable for such formulations [18]. Colorants can be added to enhance product appearance or to mask products of degradation. Stability and compatibility with the drug must be established along with the other excipients. Flavors tend to be complex entities, and stability or compatibility can pose major formulation challenges. Acidulants such as citric and tartaric acids are often added to enhance flavors, thus lowering pH of the formulation as low as 2.5–3.0. Addition of bases such as calcium carbonate, sodium bicarbonate, and magnesium trisilicate is common to increase pH and enhance drug stability. For example, in vivo and in vitro studies confirmed that the pH of the dissolved lozenge solution was the single most influential, readily adjustable formulation parameter influencing the activity of cetylpyridinium chloride activity in candy-based lozenges [19]. The dosage form needs a low moisture content (0.5–1.5% w/w), so water is evaporated off by boiling the sugar mixture during the compounding process, thus limiting the process to nonlabile drugs, and the manufacture requires specialized candy processing facilities. Packaging also needs to protect the formulated product from moisture and ranges from individual bunch wrapping to foil wraps.

### 8.1 Chewable Lozenges

Chewable lozenges are popular with the pediatric population since they are "gummy-type" lozenges. Most formulations are based on a modified suppository formula consisting of glycerin, gelatin, and water. These lozenges are often highly fruit-flavored and may have a slightly acidic taste to cover the acrid taste associated with glycerin. Soft lozenges typically comprise ingredients such as PEG 1000 or 1450, or a sugar–acacia base. Silica gel can be added to prevent sedimentation, and again this dosage form requires flavors and sweeteners to aid compliance. Soft lozenges tend to dissolve faster than gelatin bases and can be used if taste masking is not effective.

## 9 CHEWABLE TABLETS

Chewable tablets are designed to be mechanically disintegrated in the mouth. Potential advantages of chewable tablets are mainly concerning patient convenience and acceptance, although enhanced bioavailability is also claimed. This can be due

to a rapid onset of action as disintegrate is more rapid and complete compared to standard formulations that must disintegrate in the GI tract. The dosage form is an appealing alternative for pediatric and geriatric consumers. Chewable tablets also offer convenience for consumers, avoiding the necessity of coadministration with water, and creation of palatable formulations can increase compliance. Antacids and pediatric vitamins are often formulated as chewable tablets, but other formulations include antihistamines (Zyrtec), antimotility agents (Imodium Plus) and antiepileptic agents (Epanutin Infatabs), antibiotics (Augmentin Chewable), asthma treatments (Singulair), and analgesics (Motrin).

Constraints with these systems are that many pharmaceutical actives have an unpleasant bitter taste that can actually reduce compliance among patients. Iron salts, for example, can impart a rusty taste, and some antihistamines such as promethazine HCl can have a bitter aftertaste. As such, active formulations require very effective taste-masking strategies to provide acceptable patient tolerance and to ensure patient adherence to their pharmaceutical regimen.

Formulation factors governing design are similar to standard formulations (e.g., compactability, flow, etc.), but disintegrants are not included. Organoleptic properties are a major concern, especially in the design of products for children, and usage has been limited as formulators have encountered difficulties in achieving satisfactory sensory characteristics.

Certain diluents are beneficial in the formulation of chewable tablets by compression such as mannitol, lactose, sucrose, and sorbitol. They can aid disintegration upon chewing and can help with acceptable taste and mouth feel. Mannitol, for example, can impart a cooling or soothing sensation. Specialist excipients with improved sensory components such as mouth feel and lack of grittiness have been developed for formulation of chewable tablets. For example, Avicel CE-15 [a combination of microcrystalline cellulose (MCC) and guar gum] can reduce grittiness, leading to a creamier mouth feel and improved overall compatibility.

Citric acid, grape, raspberry, lemon, and cherry flavors are often used in chewable tablets and lozenges (Table 7). Flavoring agents are commonly volatile oils, and they can be dissolved in alcohol and then sprayed onto another excipient or granules. They are usually added immediately prior to compression to avoid loss due to their volatile nature. Dry flavors have advantages in terms of stability and ease of handling and are formed by emulsification of the flavor into an aqueous solution of a carrier followed by drying, encapsulating the flavor within the carrier. This is useful if the agent is prone to oxidation. Common carrier substances are acacia gum, starch, and maltodextrin. Sweeteners such as aspartame can also be added. Low-calorie and non-sugar-based excipients may present a marketing advantage.

Issues of taste masking for chewable formulations may be addressed by coating in wet granulation. The granulating/coating agent should form a flexible rather than

**TABLE 7 Flavor Groups for Taste Types**

| | |
|---|---|
| Sweet | Vanilla, grape, maple, honey |
| Sour | Citrus, raspberry, anise |
| Salty | Mixed fruit, mixed citrus, butterscotch, maple |
| Bitter | Licorice, coffee, mint, cherry, grapefruit |
| Metallic | Grape, lemon, lime |

*Source*: From ref. 18.

brittle film, have no unpleasant taste of its own, not interfere with dissolution, and be insoluble in saliva. Microencapsulation for taste masking can be achieved by phase separation or coacervation and may also impart stability. The same taste-masking technologies may be used to encapsulate drugs for formulation into chewable, softchew, and fast dissolving dosage forms. Coating materials include carboxymethylcellulose, polyvinyl alcohol (PVA), and ethylcellulose. Xylitol is the sweetest sugar alcohol, and it has a high negative heat of solution, making it a good candidate as an excipient for chewable tablets. There are many types of compressible sugars today, and most of them are composed of sucrose granulated with small amounts of modified dextrins in order to make the sucrose more compressible [20]. Modifications to sugar-based excipients such as spray-dried crystalline maltose and directly compressible sucrose (95% sucrose and 5% sorbitol) to facilitate direct compression are also aiding development in this area [21].

### 9.1 Testing of Chewable Tablets

Dissolution testing for chewable tablets should be the same as that used for regular tablets [22]. This is because patients could swallow the dosage form without adequate chewing, in which case the drug would still need to be released to ensure the desired pharmacological action. However, as outlined, chewable tablets will typically have different excipients than standard formulations, including agents to either mask or add flavor, and may undergo a different manufacturing process. Where applicable, test conditions would preferably be the same as used for non-chewable tablets of the same active pharmaceutical ingredient, but because of the nondisintegrating nature of the dosage form, it may be necessary to alter test conditions (e.g., increase the agitation rate) and specifications (e.g., increase the test duration). The reciprocating cylinder (USP apparatus 3) with the addition of glass beads may also provide more intensive agitation for in vitro dissolution testing of chewable tablets. As another option, mechanical breaking of chewable tablets prior to exposing the specimen to dissolution testing could be considered. Chewable tablets should also be evaluated for in vivo bioavailability and/or bioequivalence.

Additional concerns in the testing of chewable tablets are organoleptic, chemical, and physical stability. As it is a critical factor in the design of such formulations, taste masking should be incorporated into excipient testing during preformulation studies. Technologies like the "electronic tongue" can be used to match desirable taste characteristics [23, 24].

## 10 CHEWING GUMS

### 10.1 Composition of Chewing Gum

Medicated chewing gums are gums made with a tasteless masticatory gum base that consists of natural or synthetic elastomers [25]. They include excipients such as fillers, softeners, and sweetening and flavoring agents. Natural gum bases include chicle and smoked natural rubber and are permitted in formulations by the FDA, but modern gum bases are mostly synthetic in origin and approved bases include

**TABLE 8 Typical Formulation of Gum Base**

| Ingredient | Weight (%) | Example |
|---|---|---|
| Elastomer | 10 | Styrene–butadiene rubber |
| Plasticizer | 30 | Rosin esters |
| Texture agent/filler | 35 | Calcium carbonate |
| Wax | 15 | Paraffin wax |
| Lipid | 7 | Soya oil |
| Emulsifier | 3 | Lecithin |
| Miscellaneous | 1 | Colorant, antioxidant |

*Source*: From ref. 26.

**TABLE 9 Example Chewing Gum Formulations**

| Ingredient (%) | Sugar Gum | Sugar-Free Gum |
|---|---|---|
| Gum base | 19.4 | 25.0 |
| Corn syrup | 19.8 | — |
| Sorbitol, 70% | — | 15.0 |
| Sugar | 59.7 | — |
| Glycerin | 0.5 | 6.5 |
| Sorbitol | — | 52.3 |
| Flavor | 0.6 | 1.2 |

*Source*: From ref. 26.

styrene–butadiene rubber, polyethylene, and polyvinylacetate. Gum base usually forms about 40% of the gum, but can comprise up to 65%, and is a complex mixture, insoluble in saliva, comprising mainly of elastomer, plasticizers, waxes, lipids, and emulsifiers (see Table 8). It will also contain an adjuvant such as talc to modify the texture of the gum and low quantities of additional excipients including colorants and antioxidants such as butylated hydroxyanisole. Elastomers control the gummy texture while the plasticizers and texture agents regulate the cohesiveness of the product. The lipid and waxes melt in the mouth to provide a cooling, lubricating feeling while the juicy feel of the gum texture is from the emulsifiers. The choice and formulation of gum base will affect the release of active ingredient, and the texture, stability, and method of manufacture of the product.

The remaining ingredients in the chewing gum itself include drug, sweeteners, softeners, and flavoring and coloring agents. A typical chewing gum formulation is shown in Table 9. The sugar is for sweetening the product while the corn syrup keeps the gum fresh and flexible. Softeners or fillers are included to help blend the ingredients and retain moisture. Sugar-free gum has sorbitol, mannitol, aspartame, or saccharin instead of sugar. Optimized chewing gum formulations will require tailoring for each individual product. For example, nicotine-containing gums are formulated with the nicotine within an ion exchange resin and pH-modifying carbonates and/or bicarbonates to increase the percentage of the drug in its free base form in saliva.

## 10.2 Manufacture of Chewing Gum

The majority of chewing gum delivery systems today are manufactured using conventional gum processes. The gum base is softened or melted and placed in a kettle mixer where sweeteners, syrups, active ingredients, and other excipients are added at a defined time. The gum is then sent to a series of rollers that form it into a thin, wide ribbon. During this process, a light coating of an antisticking agent can be added (e.g., magnesium stearate, calcium carbonate, or finely powdered sugar or sugar substitute). Finally, the gum is cut to the desired size and cooled at a carefully controlled temperature and humidity.

As the heating process involved in conventional methods may limit the applicability of the process for formulation of thermally labile drugs, directly compressible, free-flowing powdered gums such as Pharmagum (SPI Pharma) and MedGumBase (Gumbase Co) have been proposed to simplify the process. These formulations can be compacted into a gum tablet using a conventional tablet press and have the potential to simplify the manufacture, facilitating inclusion of a wider range of drugs.

## 10.3 Drug Release from Chewing Gums

Until recently, the release of substances from chewing gums during mastication was studied using a panel of tasters and chew-out studies. During the mastication process, the medication contained within the gum product should be released into the saliva and is either absorbed through the buccal mucosa or swallowed and absorbed via the GI tract. The need for, and value of, in vitro drug release testing is well established for a range of dosage forms, however, standard dissolution apparatus is not suitable for monitoring release of drug from chewing gums as mastication is essential in order to provide a renewable surface for drug release after chew action. A number of devices to mimic the chewing action have been reported [26–28]. In 2000, the European Pharmacopoeia produced a monograph describing a suitable apparatus for studying the in vitro release of drug substances from chewing gums [25]. The chewing machine consists of a temperature-controlled chewing chamber in which the gum piece is chewed by two electronically controlled horizontal pistons driven by compressed air. The two pistons transmit twisting and pressing forces to the gum while a third vertical piston operates alternately to the two horizontal pistons to ensure that the gum stays in the right place (see Figure 1). The temperature of the chamber can be maintained at 37°C ± 0.5°C and the chew rate varied. Other adjustable settings include the volume of the medium, distance between the jaws and the twisting movement. The European Pharmacopoeia recommends using 20 mL of unspecified buffer in a chewing chamber of 40 mL and a chew rate of 60 strokes per minute. This apparatus has been used to study release of nicotine from commercial gums and directly compressible gums [26].

Factors affecting the release of medicament from chewing gum can be divided into three groups: the physicochemical properties of the drug, the gum properties, and chew-related factors, including rate and frequency. Drugs can be incorporated into gums as solids or liquids. For most pharmaceuticals, aqueous solubility of the drug will be a major factor affecting the release rate. In order for drugs to be

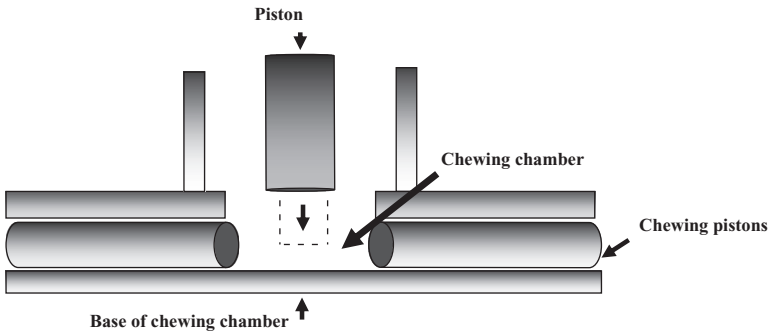


**FIGURE 1** Schematic of chewing chamber of in vitro chewing apparatus [26].

released, the gum would need to become hydrated; the drugs can then dissolve and diffuse through the gum base under the action of chewing.

For treatment of local conditions, a release period less than 1 h may be desirable, but a faster release may be required if a rapid onset of action is required for a systemically absorbed formulation. There are a number of strategies that can be undertaken in order to achieve the desired release rate. Decreasing the amount of the gum base will enhance the release of lipophilic drugs and addition of excipients designed to promote release can also be considered. Release can be sustained using, for example, ion exchange resins as described for nicotine-containing gums. Changes in gum texture as a consequence of changes in excipient levels provide a further challenge to controlling the release of drugs. A quantitative measure of gum texture during the process is possible using texture analysis techniques [26].

## 10.4 Applications for Chewing Gums

The promotion of sugar-free gums to counteract dental caries by stimulation of saliva secretion has led to a more widespread use and acceptance of gums. Medicated gums for delivery of dental products to the oral cavity are marketed in a number of countries, for example, fluoride-containing gums as an alternative to mouthwashes and tablets or chlorhexidine gum for treatment of gingivitis. The potential use of medicated chewing gums in the treatment of oral infections has also been reported. Gums have been prepared containing antifungal agents such as nystatin [29] and miconazole [30] or antibiotics, such as penicillin and metronidazole for the treatment of oral gingivitis [31].

Chewing gum is also useful as a delivery system for agents intended for systemic delivery. Drug that is released from the gum within the oral cavity can act locally, be absorbed via the buccal mucosa, or swallowed with the saliva. The buccal mucosa is well vascularized, and if a drug is absorbed by this route, then first-pass metabolism could be avoided. Associated increases in bioavailability can permit the use of lower dosages. Like orally disintegrating tablets, chewing gum is a convenient dosage form; it can be administered without water and to those who have difficulty swallowing. Although medicated gums are generally intended to be chewed for 10–30 min and can therefore be designed for sustained release, a fast onset of action can result either from buccal absorption or as a consequence of the active being

already dissolved in the saliva prior to swallowing. Guidance can be given regarding chewing conditions (e.g., time, frequency), but factors such as the force of chewing and salivary flow will impact on drug release and the fraction of drug absorbed via the oral mucosa. Released drug can be swallowed with the saliva, therefore leading to multiple absorption sites, which can result in variable pharmacokinetics.

Along with nicotine replacement patches, nicotine chewing gum for smoking cessation therapy has met with major sales success. The principal active ingredient of currently marketed nicotine chewing gums is nicotine polacrilex USP. The nicotine is loaded at approximately 18% w/w on an ion exchange resin (Amberlite IRP64). Recent product variations have been launched with improved flavors such as mint and fruit, rather than the original peppery flavoring, designed to reduce the unpleasant taste and burning sensation arising from nicotine itself and flavored coated gums that are sweeter and easier to chew.

Other applications for chewing gum formulations include delivery of antacids such as calcium carbonate, antiemetics for travel sickness, and vitamins and minerals. However, the potential for a buccal delivery, a fast onset of action, and the opportunity for product line extension makes it an attractive alternative delivery form for other applications.

## 11 ORALLY DISINTEGRATING TABLETS

The demand for fast-dissolving/disintegrating tablets or fast-melting tablets that can dissolve or disintegrate in the mouth has been growing particularly for those with difficulty swallowing tablets such as the elderly and children. They are referred to using a range of terminologies: fast dissolving, orodispersible, and fast melting and the FDA has adopted the term orally disintegrating tablets (ODTs). Patients with persistent nausea or those who have little or no access to water could also benefit from ODTs. Other advantages include product differentiation and market expansion, and applications exist in the veterinary market for oral administration to animals.

Orally disintegrating tablets disintegrate and/or dissolve rapidly in the saliva without the need for water, within seconds to minutes. Some tablets are designed to dissolve rapidly in saliva, within a few seconds, and are true fast-dissolving tablets. Others contain agents to enhance the rate of tablet disintegration in the oral cavity and are more appropriately termed fast-disintegrating tablets, as they may take up to a minute to completely disintegrate. Increased bioavailability using such formulations is sometimes possible if there is sufficient absorption via the oral cavity prior to swallowing [32]. However, if the amount of swallowed drug varies, there is the potential for inconsistent bioavailability. Patented orally disintegrating tablet technologies include OraSolv, DuraSolv, Zydis, FlashTab, WOWTAB, and others. They are generally prepared using freeze drying, compaction, or molding. Examples of marketed products, excipients, and technologies used are given in Table 10.

Platform technologies based on freeze drying include Zydis (Cardinal Health) and Quicksolv (Janssen Pharmaceutica). Zydis was the first ODT to be successfully launched, and it is ideal for poorly soluble drugs. It can incorporate doses up to 400 mg, but high loadings can extend disintegration time. The porous matrix consists of a network of water-soluble carriers and active ingredient. The maximum dose for

**TABLE 10 Examples of Marketed ODT Products and Technologies**

| Name (Company) | Examples | Ingredients[a] | Technology |
|---|---|---|---|
| Zydis (Cardinal Health) | Claritin Reditab | *Micronized loratadine (10 mg)*, citric acid, gelatin, mannitol, mint flavor | Freeze drying |
| Zydis (Cardinal Health) | Zofran ODT | *Ondansetron (4 or 8 mg)*, aspartame, gelatin, mannitol, methylparaben sodium, propylparaben sodium, strawberry flavor | Freeze drying |
| Zydis (Cardinal Health) | Zyprexa Zydis | *Olanzapine (5, 10, 15, or 20 mg)*, gelatin, mannitol, aspartame, methylparaben sodium, propylparaben sodium | Freeze drying |
| Oralsolv (CIMA Labs Inc.) | Remeron Soltab | *Mirtazepine (15, 30, or 45 mg)*, aspartame, citric acid, crospovidone, hydroxypropyl methylcellulose, magnesium stearate, mannitol, microcrystalline cellulose, polymethacrylate, povidone, sodium bicarbonate, starch, sucrose, orange flavor | Compression |
| Durasolv (CIMA Labs Inc.) | Zomig ZMT | *Zolmitriptan (2.5 mg)*, mannitol, microcrystalline cellulose, crospovidone, aspartame, sodium bicarbonate, citric acid, anhydrous, colloidal silicon dioxide, magnesium stearate, orange flavor | Compression |
| WOWTAB (Yamanouchi Pharma Technologies, Inc.) | Benadryl Allergy & Sinus Fastmelt | *Diphenhydramine citrate (19 mg), pseudoephedrine HCl (30 mg),* aspartame, citric acid, D&C red no. 7 calcium lake, ethylcellulose, flavor, lactitol, magnesium stearate, mannitol, and stearic acid | Compression molded tablet |
| Flashtab (Prographarm/ Ethypharm) | Excedrin Quicktabs | *Acetaminophen (500 mg), caffeine (65 mg)*, aminoalkyl methacrylate copolymers, citric acid, colloidal silicon dioxide, crospovidone, distilled acetylated monoglycerides, ethylcellulose, flavors, magnesium stearate, mannitol, methacrylester copolymer, polyvinyl acetate, povidone, propylene glycol, propyl gallate, silica gel, sodium lauryl sulfate, sucralose, talc | Compression |

[a]Active ingredients appear in italics.

water-soluble drugs is 60 mg, and particle sizes of drug and excipients should be below 50 μm.

Excipients used in the formulation usually include a mixture of a water-soluble polymer and a crystalline sugar. Mannitol and natural polysaccharides such as gelatin and alginates are used. Microencapsulation and complexation with ion exchange resins can be combined with additional flavors and sweeteners for taste masking of bitter drugs. The fairly complex nature of manufacture and scale-up contributes to

a relatively high manufacturing cost. Manufacture comprises three stages. The production sequence begins with the bulk preparation of an aqueous drug solution or suspension and subsequent precise dosing into preformed blisters. It is the blister that actually forms the tablet shape and is, therefore, an integral component of the total product package. The second phase of manufacturing entails passing the filled blisters through a cryogenic freezing process to control the ultimate size of the ice crystals. This aids in ensuring porosity and the product is freeze dried. The final phase of production involves sealing the open blisters via a heat-seal process to ensure stability and protect the fragile tablet during removal by the patient.

The manufacture of Flashdose (Fuisz Technologies/Biovail) is patented as Shearform process and utilizes a unique spinning mechanism to produce a flosslike or shear-form crystalline structure, much like cotton candy. The matrix comprises saccharides or polysaccharides which are subjected to simultaneous melting and centrifugal force and then partially recrystallized [33]. High temperatures are involved so the technology is only suitable for thermostable agents. Drug can then be incorporated, either as coated or uncoated microspheres, into the sugar and the formulation is compressed into a tablet. Manufacture of the microspheres is patented as Ceform and will help with taste masking. The final product has a very high surface area for dissolution and it disperses and dissolves quickly once placed onto the tongue. Like freeze-drying processes, the manufacture is expensive and resultant formulations are friable and moisture sensitive, therefore requiring specialized packaging.

Most commercial ODTs have been developed using mannitol as the bulk excipient of choice because of its extremely low hygroscopicity, excellent compatibility, good compressibility, better sweetness, and relatively slower dissolution kinetics. Although lactose also has a relatively low aqueous solubility compared with other excipients that have acceptable palatabilities, the dispersibility of a bulk excipient is more important than its aqueous solubility for a successful ODT formulation. Many of the initially marketed ODTs were prepared by the wet granulation of mannitol followed by direct compression. However, added functionality mannitols are now available to simplify the process of ODT manufacturing by direct compression.

Direct compression is, as for normal tablets, the most straightforward process for manufacturing ODTs. Conventional equipment can be used and high doses can be incorporated. The excipients play a major role in the successful formulation and superdisintegrants, hydrophilic polymers, and effervescent compounds are included. Patented technologies include Orasolv and Durasolv (Cima Labs) and Ziplets (Eurand). The OraSolv technology is best described as a fast disintegrating, slightly effervescing tablet; the tablet matrix dissolves in less than one minute, leaving coated drug powder. Both the coating and the effervescence contribute to taste masking in OraSolv. The tablet is prepared by direct compression but at a low pressure, yielding a weaker and more brittle tablet in comparison with conventional tablets. For that reason, Cima developed a special handling and packaging system for OraSolv called Packsolv. Acidic compounds such as citric or fumaric acid are included in the formulation together with a carbonate or bicarbonate. An advantage that goes along with the low degree of compaction of OraSolv is that the particle coating used for taste masking is not compromised by fracture during processing.

DuraSolv is Cima's second-generation fast-dissolving/disintegrating tablet formulation and is also produced using direct compression but using higher compaction

pressures during tableting, resulting in a stronger product. It is thus produced in a faster and more cost-effective manner and may not require specialized packaging. Large amounts of finely milled conventional fillers are used (mannitol, lactose) while the effervescing agents are reduced. It is best suited to potent drugs, requiring only low doses, and the taste-masking coating can be disturbed following compaction. DuraSolv is currently available in two products: NuLev and Zomig ZMT.

Compression following wet or dry granulation is also employed in the manufacture of ODTs. Patented formulations include WOWTAB and Flashtab. WOWTAB relies on a combination of low moldable sahharides (mannitol, glucose, sucrose) with a highly moldable saccharide (malitol, sorbitol, maltose) using conventional granulation and tableting techniques to form a tablet of suitable mechanical properties with desired disintegration. It is manufactured by compression of molded granules, can accommodate a high level of drug loading (up to 50% in some cases), and can be packed using conventional methodology.

Flashtab (Ethypharm) is the technology behind Exedrin QuickTabs and uses swellable agents and disintegrants along with sugars and polyalcohols to achieve a fast dispersible formulation. The manufacture involves either wet or dry granulation of the excipients, blending with the active followed by direct compression.

### 11.1 Dissolution Testing of ODTs

Taste masking (drug coating) is very often an essential feature of ODTs and thus can also be the rate-determining mechanism for dissolution/release. If taste masking is not an issue, then the development of dissolution methods is comparable to the approach taken for conventional tablets and pharmacopeial conditions should be used [34]. Due to the nature of the product, the dissolution of orally disintegrating tablets is very fast when using USP monograph conditions, and slower paddle speeds can be used to obtain a profile. Other media such as 0.1 N HCl can also be used. USP 2 paddle apparatus is the most suitable and common choice for orally disintegrating tablets, with a paddle speed of 50 rpm commonly used [34]. Faster agitation rates may be necessary in the case of sample mounding. The method can be applied to the ODTs (finished product) as well as to the bulk intermediate (in the case of coated drug powder/granulate). A potential difficulty for in vitro dissolution testing may arise from floating particles [35]. Similarly, difficulties can arise using USP I due to trapping of disintegrated fragments.

A single-point specification is considered appropriate for ODTs with fast dissolution properties. For ODTs that dissolve very quickly, a disintegration test may be used in lieu of a dissolution test if it is shown to be a good discriminating method. If taste masking (using a polymer coating) is a key aspect of the dosage form, a multipoint profile in a neutral pH medium with early points of analysis (e.g., ≤5 min) may be recommended [34].

## 12 SOLID DOSAGE FORMS FOR NONORAL ROUTES

Although the majority of tablets and capsules are intended for oral delivery, there are a number of other delivery routes suitable for drug delivery by these formula-

tions. Some buccal formulations have been discussed above, and tablets can also be administered via the rectal and vaginal routes for local and systemic treatment.

Many types of product have been designed for vaginal administration with creams, gels, and pessaries being most popular, although powders and tablets have also been used. Despite the effectiveness of systemic vaginal absorption, the majority of products administered by this route are for the treatment of localized infections, especially *Candida albicans*, (e.g., Canestan vaginal tablets). Estradiol tablets (Vagifem) were also designed for delivery via vaginal route to address patient preference issues with vaginal creams. The formulations are administered with an applicator and are designed to dissolve or erode slowly in the vaginal secretions [36]. Bioadhesion as a means of retaining the formulation at the site of delivery is widely accepted to retain formulations in the buccal cavity [37] and has also been reported for the vaginal route [38]. An increased residence time may improve drug absorption by these routes.

## REFERENCES


1. Amidon, G. L., Lennernas, H., Shah, V. P., and Crison, J. R. (1995), A theoretical basis for a biopharmaceutic drug classification: The correlation of *in vitro* drug product dissolution and *in vivo* bioavailability, *Pharm. Res.*, 12, 413–420.
2. Waiver of *in-vivo* bioavailability and bioequivalence studies for immediate release solid oral dosage forms based on a biopharmaceutics classification system, available: http://www.fda.gov/cder/OPS/BCS_guidance.html.
3. Waiver of in vivo bioequivalence studies for immediate release solid oral dosage forms based on a biopharmaceutics classification system, available: http://www.fda.gov/cder/guidance/index.html.
4. Zhao, C., Jain, A., Hailemariam, L., Suresh, P., Akkisetty, P., Joglekar, G., Venkatasubramanian, V., Reklaitis, G.V., Morris, K., and Basu, P. (2006), Toward intelligent decision support for pharmaceutical product development, *J. Pharm. Innovation*, 1, 23–35.
5. Banker, G. S., and Rhodes, C. T. (2002), *Modern Pharmaceutics*, 4th ed., Drugs and the Pharmaceutical Sciences 121, Marcel Dekker, New York.
6. Rowe, R. C., Sheskey, P. J., and Weller, P. J. (2001), *Handbook of Pharmaceutical Excipients*, 4th ed., Pharmaceutical Press, London.
7. Weiner, M., and Bernstein I. L. (1989), *Adverse Reactions to Drug Formulation Agents: A Handbook of Excipients*, Marcel Dekker, New York, pp. 93–94.
8. Lachman, L., Lieberman, H. A., and Kanig, J. L. (1986), *The Theory and Practice of Industrial Pharmacy*, 3rd ed., Lea & Febiger, Philadelphia.
9. Augsburger, L. L., Hahm, H. A., Brzeczko, A. W., and Shah, U. (2002), Superdisintegrants: Characterization and function, in Swarbrick, J., and Boylan, J. V., Eds., *Encyclopedia of Pharmaceutical Technology*, Vol. 3, 2nd ed., Marcel Dekker, New York.
10. Gibson, M. (2001), Pharmaceutical preformulation and formulation; a practical guide from candidate drug selection to commercial dosage form, IHS Health Group, Englewood, Colorado.
11. Stanley, J. P. (1986), Soft gelatin capsules, in Lachman, L., Lieberman, H. A., and Kanig, J. L., Eds., *The Theory and Practice of Industrial Pharmacy*, 3rd ed., Lea & Febiger, Philadelphia.

12. *The United States Pharmacopeia*, 26th revision, United States Pharmacopeial Convention, Rockville, MD, 2003.
13. Gelatin Capsule Working Group, Collaborative development of two-tier dissolution testing for gelatin capsules and gelatin-coated tablets using enzyme-containing media, *Pharmacop. Forum*, 24(5), Sept./Oct. 1998.
14. Lee, R. E., Effervescent tablets. Key facts about a unique effective dosage form, *Tablets Capsules*, available: http://www.amerilabtech.com/EffervescentTablets&KeyFacts.pdf, accessed Aug. 6, 2004.
15. Rotthauser, B., Kraus, G., and Schmidt, P. C. (1998), Optimization of an effervescent tablet formulations containing spray-dried L-leucine and polyethylene glycol 6000 as lubricants using a central composite design, *Eur. J. Pharm. Biopharm.*, 46, 85–94.
16. Stahl, H. (2003), Effervescent dosage manufacturing, *Pharm. Technol. Eur.*, 4, 25–28.
17. Lindberg, N.-O., and Hansson, H. (2002), Effervescent pharmaceuticals, in Swarbrick, J., and Boylan, J. V., Eds., *Encyclopedia of Pharmaceutical Technology*, Vol. 2, 2nd ed., Marcel Dekker, New York.
18. Mendes, R. W., and Bhargava, H. (2002), Lozenges, in Swarbrick J., and Boylan, J. V., Eds., *Encyclopedia of Pharmaceutical Technology*, Vol. 2, 2nd ed., Marcel Dekker, New York.
19. Richards, R. M. E., Xing, J. Z., and Weir, L. F. C. (1996), The effect of formulation on the antimicrobial activity of cetylpyridinium chloride in candy based lozenges, *Pharm. Res.*, 13, 583–587.
20. Bolhuis, G., and Armstrong, A. N. (2006), Excipients for direct compaction—An update, *Pharm. Dev. Technol.*, 11, 111–124.
21. Bowe, K. E. (1998), Recent advances in sugar-based excipients, *Pharm. Sci. Technol. Today*, 1(4), 166–173.
22. FDA guidance for industry: Bioavailability and bioequivalence studies for orally administered drug products—General considerations, Oct. 2000.
23. Murray, O. J., Dang, W., and Bergstrom, D. (2004), Using an electronic tongue to optimize taste-masking in a lyophilized orally disintegrating tablet formulation, *Pharm. Technol. Outsourcing Res.*, 42–52.
24. Zheng, J. Y., and Keeney, M. P. (2004), Taste masking analysis in pharmaceutical formulation development using an electronic tongue, *Int. J. Pharm.*, 310, 118–124.
25. *European Pharmacopoeia*, 4th ed., European Directorate for the Quality of Medicines, Strasbourg, 2001.
26. Morjaria, Y., Irwin, W. J., Barnett, P. X., Chan, R. S., and Conway, B. R. (2004), *In vitro* release of nicotine from chewing gum formulations, *Dissolution Technol.*, 11, 12–15.
27. Rider, J. N., Brunson, E. L., Chambliss, W. G., Cleary, R. W., Hikal, A. H., Rider, P. H., Walker, L. A., Wyandt, C. M., and Jones, A. B. (1992), Development and evaluation of a novel dissolution apparatus for medicated chewing gum products, *Pharm. Res.*, 9, 255–260.
28. Kvist, C., Andersson, S.-B., Fors, S., Wennergren, B., and Berglund, J. (1999), Apparatus for studying *in vitro* release from medicated chewing gums, *Int. J. Pharm.*, 89, 57–65.
29. Andersen, T., Gram-Hansen, M., Pedersen, M., and Rassing, M. R. (1990), Chewing gum as a drug delivery system for nystatin. Influence of solubilising agents on the release of water-soluble drugs, *Drug Dev. Ind. Pharm.*, 16, 1985–1994.
30. Pedersen, M., and Rassing, M. R. (1991), Miconazole chewing gum as a drug delivery system test of release promoting additives, *Drug Dev. Ind. Pharm.*, 17, 411–420.
31. Emslie, R. D. (1967), Treatment of acute ulcerative gingivitis. A clinical trial using chewing gums containing metronidazole or penicillin, *Br. Dent. J.*, 122, 307–308.

32. Habib, W., Khankari, R., and Hontz, J. (2002), Fast-dissolve drug delivery system, *Crit. Rev. Therap. Drug Carrier Syst.*, 17, 61–72.
33. Dobetti, L. (2001), Fast-melting tablets: Developments and technologies, *Pharm. Technol. Drug Deliv. Suppl.*, 44–50.
34. Klanke, J. (2003), Dissolution testing of orally disintegrating tablets, *Dissolution Technol.*, 10(2), 6–8.
35. Siewert, M., Dressman, J., Brown, C., and Shah, V. P. (2003), FIP/AAPS. Guidelines for dissolution/*in vitro* release testing of novel/special dosage forms, *Dissolution Technol.*, 10(1), 6–15.
36. Conine, J. W., and Pikal, M. J. (1989), Special tablets, in Lieberman, H. A., Lachman, L., and Schwartz, J. B. Eds., *Pharmaeutical Dosage Forms: Tablets*, Vol. 1, 2nd ed., Marcel Dekker, New York.
37. Sudhakar, Y., Kuotsu, K., and Bandyopadhyay, A. K. (2006), Buccal bioadhesive drug delivery—A promising option for orally less efficient drugs, *J. Controlled Release*, 114, 15–40.
38. Hussain, A., and Ahsan, F. (2005), The vagina as a route for systemic drug delivery, *J. Controlled Release*, 103, 301–313.

# SEMISOLID DOSAGES: OINTMENTS, CREAMS, AND GELS

Ravichandran Mahalingam, Xiaoling Li, and Bhaskara R. Jasti
*University of the Pacific, Stockton, California*

**Contents**

1 Introduction
2 Ointments and Creams
  2.1 Definition
  2.2 Bases
  2.3 Preparation and Packaging
  2.4 Evaluation
  2.5 Typical Pharmacopeial/Commercial Examples
3 Gels
  3.1 Definition
  3.2 Characteristics
  3.3 Classification
  3.4 Stimuli-Responsive Hydrogels
  3.5 Gelling Agents
  3.6 Preparation and Packaging
  3.7 Evaluation
  3.8 Typical Pharmacopeial and Commercial Examples
4 Regulatory Requirements for Semisolids
  References

## 1 INTRODUCTION

Semisolid dosage forms are traditionally used for treating topical ailments. The vast majority of them are meant for skin applications. They are also used for treating ophthalmic, nasal, buccal, rectal, and vaginal ailments. Various categories of drugs

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

such as antibacterials, antifungals, antivirals, antipruritics, local anesthetics, anti-inflammatories, analgesics, keratolytics, astringents, and mydriatic agents are incorporated into these products. Drugs incorporated into semisolids either show their activity on the surface layers of tissues or penetrate into internal layers to reach the site of action. For example, an antiseptic ointment should be able to penetrate the skin layers and reach the deep-seated infections in order to prevent the growth of microbes and heal the wound.

Systemic entry of drugs from these products is limited due to various physicochemical properties of dosage forms and biological factors. The barrier nature of most surface biological layers such as skin, cornea and conjunctiva of the eye, and mucosa of nose, mouth, rectum, and vagina greatly limits their entry into the systemic circulation. Systemic delivery of drugs from topical dosages is however feasible by suitable formulation modifications. Semisolid dosage forms are also used in non-therapeutic conditions for providing protective and lubricating functions. They protect the skin against external environments such as air, moisture, and sun rays and hence their components do not necessarily penetrate the skin layers. Cold creams and vanishing creams are classic examples of such semisolid preparations.

The formulation, evaluation, and regulatory feature of the three most commonly used semisolid dosage forms, ointments, creams, and gels, are described in this chapter.

## 2 OINTMENTS AND CREAMS

### 2.1 Definition

Ointments are semisolid preparations intended for topical application. They are used to provide protective and emollient effects on the skin or carry medicaments for treating certain topical ailments. They are also used to deliver drugs into eye, nose, vagina, and rectum. Ointments intended for ophthalmic purposes are required to be sterile. When applied to the eyes, they reside in the conjunctival sac for prolonged periods compared to solutions and suspensions and improve the fraction of drug absorbed across ocular tissues. Ophthalmic ointments are preferred for night-time applications as they spread over the entire corneal and conjunctival surface and cause blurred vision.

Creams are basically ointments which are made less greasy by incorporation of water. Presence of water in creams makes them act as emulsions and therefore are sometimes referred as semisolid emulsions. Hydrophilic creams contain large amounts of water in their external phase (e.g., vanishing cream) and hydrophobic creams contain water in the internal phase (e.g., cold cream). An emulsifying agent is used to disperse the aqueous phase in the oily phase or vice versa. As with ointments, creams are formulated to provide protective, emollient actions or deliver drugs to surface or interior layers of skin, rectum, and vagina. Creams are softer than ointments and are preferred because of their easy removal from containers and good spreadability over the absorption site.

### 2.2 Bases

Bases are classified based on their composition and physical characteristics. The U.S. Pharmacopeia (USP) classifies ointment bases as hydrocarbon bases (oleaginous

bases), absorption bases, water-removable bases, and water-soluble bases (water-miscible bases) [1].

*Hydrocarbon bases* are made of oleaginous materials. They provide emollient and protective properties and remain in the skin for prolonged periods. It is difficult to incorporate aqueous phases into hydrocarbon bases. However, powders can be incorporated into these bases with the aid of liquid petrolatum. Removal of hydrocarbon bases from the skin is difficult due to their oily nature. Petrolatum USP, white petrolatum USP, yellow ointment USP, and white ointment USP are examples of hydrocarbon bases.

*Absorption bases* contain small amounts of water. They provide relatively less emollient properties than hydrocarbon bases. Similar to hydrocarbon bases, absorption bases are also difficult to remove from the skin due to their hydrophobic nature. Hydrophilic petrolatum USP and lanolin USP are examples of absorption bases.

*Water-removable bases* are basically oil-in-water emulsions. Unlike hydrocarbon and absorption bases, a large proportion of aqueous phase can be incorporated into water-removable bases with the aid of suitable emulsifying agents. It is easy to remove these bases from the skin due to their hydrophilic nature. Hydrophilic ointment USP is an example of a water-removable ointment base.

*Water-soluble bases* do not contain any oily or oleaginous phase. Solids can be easily incorporated into these bases. They may be completely removed from the skin due to their water solubility. Polyethylene glycol (PEG) ointment National Formulary (NF) is an example of a water-soluble base.

Selection of an appropriate base for an ointment or cream formulation depends on the type of activity desired (e.g., topical or percutaneous absorption), compatibility with other components, physicochemical and microbial stability of the product, ease of manufacture, pourability and spreadability of the formulation, duration of contact, chances of hypersensitivity reactions, and ease of washing from the site of application. In addition, bases that are used in ophthalmic preparations should be nonirritating and should soften at body temperatures. White petrolatum and liquid petrolatum are generally used in ophthalmic preparations. Table 1 summarizes

**TABLE 1 Some Compendial Bases Used in Ointments and Creams**

| Name | Synonyms | Official Compendia | Specifications |
|---|---|---|---|
| Carnauba wax | Caranda wax, Brazil wax | BP, JP, PhEur, USPNF | Melting range 80–88 °C[a]; iodine value 5–14[b]; acid value 2–7; saponification value 78–95; total ash ≤0.25% |
| Cetyl alcohol | Cetanol, Avol, Lipocol C | BP, JP, PhEur, USPNF | Melting range 47–53 °C[b]; residue on ignition ≤0.05%[b]; iodine value ≤5.0; acid value ≤2.0; saponification value ≤2.0[a] |
| Cetyl ester wax | Crodamol SS, Ritachol SS, Starfol wax | USPNF | Melting range 43–47 °C; acid value ≤5.0; saponification value 109–120; iodine value ≤1.0 |
| Emulsifying wax | Collone HV, Crodex A, Lipowax PA | BP | Saponification value ≤2.0; iodine value ≤3.0[c] |

**TABLE 1** ***Continued***

| Name | Synonyms | Official Compendia | Specifications |
|---|---|---|---|
| Hydrous lanolin | Hydrous wool fat, Lipolan | BP, JP, PhEur | Melting range 38–44 °C; acid value ≤0.8; saponification value 67–79; nonvolatile matter 72.5–77.5%; iodine value 18–36[b] |
| Lanolin | Wool fat, purified wool fat, Corona | BP, JP, PhEur, USPNF | Melting range 38–44 °C; loss on drying ≤0.25%; residue on ignition ≤0.1%; iodine value 18–36; acid value ≤1.0[b] |
| Lanolin alcohols | Argowax, Ritawax, wool wax alcohol | BP, PhEur, USPNF | Melting range ≥56 °C; loss on drying ≤0.50%; residue on ignition ≤0.15%; acid value ≤2.0; saponification value ≤12 |
| Microcrystalline wax | Petroleum ceresin | USPNF | Melting range 54–102 °C; residue on ignition ≤0.10% |
| Paraffin | Paraffin wax, hard wax, hard paraffin | BP, JP, PhEur, USPNF | Melting range 47–65 °C |
| Petrolatum | Yellow soft paraffin, yellow petroleum jelly | BP, JP, PhEur, USPNF | Melting range 38–60 °C; residue on ignition ≤0.1% |
| Poloxamer | Polyethylene–propylene glycol, Lutrol, Pluronic | BP, PhEur, USPNF | Melting point ≈50 °C |
| Polyethylene glycol (PEG) | Macrogol, Carbowax, PEG, Lutrol | BP, JP, PhEur, USPNF | Melting range of PEG 1000, 37–40 °C; melting range of PEG 8000, 60–63 °C; residue on ignition ≤0.1% |
| Stearic acid | Emersol, Hystrene | BP, JP, PhEur, USPNF | Melting range ≥54 °C; iodine value ≤4.0 |
| Stearyl alcohol | Lipocol S, Cachalot, Rita SA | BP, JP, PhEur, USPNF | Melting range 55–60 °C; residue on ignition 0.05%[b]; iodine value ≤2.0; acid value ≤2.0; saponification value ≤2.0[a] |
| White wax | Bleached wax | BP, JP, PhEur, USPNF | Melting range 62–65 °C; acid value 17–24; saponification value 87–104[a] |
| Yellow wax | Refined wax | BP, JP, PhEur, USPNF | Acid value 17–22[a]; saponification value 87–102[a] |

*Note*: BP, British Pharmacopoeia; JP, Japanese Pharmacopoeia; PhEur, European Pharmacopoeia; USPNF, U.S. Pharmacopeia/National Formulary. All are USPNF specifications, except as indicated below.

[a]European Pharmacopoeia.

[b]Japanese Pharmacopoeia.

[c]British Pharmacopoeia.

compendial status, synonym, and specifications of some of the bases used in ointments and creams.

The following sections describe the source, physicochemical properties, formulation considerations, stability, incompatibility, storage, and hypersensitivity reactions of some of these bases.

***Lanolin*** Lanolin is a refined, decolorized, and deodorized material obtained from sheep wool. It is available as a pale yellow, waxy material with a characteristic odor. It is extensively used in the preparation of hydrophobic ointments and water-in-oil creams. As lanolin is prone to oxidation, antioxidants such as butylated hydroxytoluene are generally included. Although lanolin is insoluble in water, it is miscible with water up to 1 :2 ratio. This property favors in preparing physically stable creams. Addition of soft paraffin or vegetable oil improves the emollient property of lanolin preparations. Exposure of lanolin to higher temperature usually leads to discoloration and rancidlike odor, and hence prolonged heating is avoided during the preparation and preservation of lanolin-containing preparations. Gamma sterilization or filtration sterilization is usually employed for sterilizing ophthalmic ointments containing lanolin. Lanolin and some of its derivatives are reported to cause hypersensitivity reactions and therefore are avoided in patients with known hypersensitivity. One of the reasons for hypersensitivity reactions is free fatty alcohols. Modified lanolins containing reduced levels of free fatty alcohols are commercially available [2, 3].

***Hydrous Lanolin*** Incorporation of about 25–30% of water into lanolin gives hydrous lanolin. Gradual addition of water into molten lanolin with constant stirring helps in water incorporation. It is available as a pale yellow, oily material with a characteristic odor. The water uptake capacity of hydrous lanolin is higher than lanolin, and it is used for preparing topical hydrophobic ointments or water-in-oil creams with larger aqueous phase. Exposure of these preparations to higher temperatures results in separations of oily and aqueous layers. Addition of antioxidants and preservation in well-filled, airtight, light-resistant containers in a cool and dry place improve the stability of lanolin products. Well-preserved preparations can be stored up to two years. Hydrous lanolin that contains free fatty alcohols is avoided in hypersensitive patients [2, 3].

***Lanolin Alcohols*** Lanolin alcohol is prepared from lanolin by the saponification process and is used as a hydrophobic vehicle in pharmaceutical ointments and creams. It is composed of steroidal and triterpene alcohols and is available as a brittle solid material pale yellow in color with a faint characteristic odor. The brittle powder becomes plastic under warm conditions. It is practically insoluble in water and soluble in boiling ethanol. Lanolin alcohol possesses emollient properties, which makes it suitable for preparing dry-skin ointments, eye ointments, and water-in-oil creams. Creams containing lanolin alcohols do not show surface darkening and do not produce objectionable odor compared to lanolin-containing preparations. Inclusion of about 0.1% antioxidant, however, minimizes the oxidation on storage. Preparations containing lanolin alcohols can be stored up to two years if preserved in well-filled, well-closed, light-resistant containers in a cool and dry place. As with

other lanolin bases, hypersensitivity reactions may occur in some individuals while using preparations containing lanolin alcohols [2, 3].

***Petrolatum*** Petrolatum is also known as yellow soft paraffin. It is an inert material obtained from petroleum, which contains branched and unbranched hydrocarbons. It is available as soft oily material and appears pale yellow to yellow in color. Various grades of petrolatum are commercially available with varying physical properties. All these grades are generally insoluble in water and possess emollient properties. Concentrations up to 30% are used in creams. Petrolatum shows phase transitions on heating to about 35 °C. As it possesses a higher coefficient of thermal expansion, prolonged heating is avoided during processing. The presence of minor impurities can oxidize petrolatum and discolor the product. Antioxidants are therefore added to prevent such physical changes in preparations during storage. Butylated hydroxyanisole, butylated hydroxytoluene, or α-tocopherol is generally incorporated as an antioxidant in petrolatum products. In addition, use of well-closed, airtight, light-resistant containers and storage in a cool and dry place improve stability of preparations. Minor quantities of polycyclic aromatic hydrocarbon impurities in petrolatum sometimes cause hypersensitivity reactions. Substituting yellow soft paraffin with white soft paraffin reduces such reactions [4].

***Petrolatum and Lanolin Alcohols*** Various quantities of lanolin alcohols are mixed with petrolatum to form these mixtures. Wool ointment British Pharmacopoeia (BP) 2001 contains 6% lanolin alcohols and 10% petrolatum. These proportions can be varied to alter physical properties such as consistency and melting range. They are available as soft solids pale ivory in color and possess a characteristic odor. These mixtures are insoluble in water, and concentrations ranging 5–50% are used for preparing hydrophobic ointments. They are also used for preparing water-in-oil emollient creams. Preparations containing petrolatum and lanolin alcohols need to be preserved in airtight, well-closed, light-resistant containers in a cool and dry place to avoid oxidation of impurities and discoloration. Antioxidants improve the stability of these products. Although these mixtures are safe for topical applications, hypersensitivity reactions may occur in some individuals due to the presence of lanolin alcohol [5].

***Paraffin*** Paraffin is obtained by distillation of crude petroleum followed by purification processes. The purified fraction contains saturated hydrocarbons. Paraffin is available as a white color solid and does not possess any specific odor or taste. Different purity grades are available. Use of highly purified grades can avoid batch-to-batch variations in formulations, especially the hardness, melting behavior, and malleability. Paraffin is insoluble in water and is generally used to prepare hydrophobic topical ointments and water-in-oil creams. Repeated heating and congealing are avoided during formulation as they change the physical properties of paraffin. These preparations need to be preserved in well-closed container at room temperature. Synthetic paraffins, which melt between 96 and 105 °C, are sometimes used to increase the melting point and stiffness of formulations [6].

***Polyethylene Glycol*** Also known as macrogol, PEG is synthesized by condensation of ethylene oxide and water under suitable reaction conditions. Based on the

**TABLE 2 Properties of Different Grades of PEG**

| | By Grade | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Property | 200 | 400 | 600 | 1000 | 2000 | 3000 | 4000 | 8000 |
| Physical state | Liquid | Liquid | Liquid | Solid | Solid | Solid | Solid | Solid |
| Average molecular weight | 190–210 | 380–420 | 570–613 | 950–1050 | 1800–2200 | 2700–3300 | 3000–4800 | 7000–9000 |
| Melting (°C) | — | — | — | 37–40 | 45–50 | 48–54 | 50–58 | 60–63 |
| Density ($g/cm^3$) | 1.11–1.14 | 1.11–1.14 | 1.11–1.14 | 1.15–1.21 | 1.15–1.21 | 1.15–1.21 | 1.15–1.21 | 1.15–1.21 |
| Kinematic viscosity[a] (cS) | 3.9–4.8 | 6.8–8.0 | 9.9–11.3 | 16.0–19.0 | 38–49 | 67–93 | 110–158 | 470–900 |

[a]At 98.9 °C.

number of oxyethylene groups present, their molecular weights vary from few hundreds to several thousands. Usually the number that follows PEG represents their average molecular weight. They are available as liquids or solids based on molecular weight. PEGs 600 or less are liquids, whereas PEGs above 1000 are solids. PEG liquids are usually clear or pale yellow in color. Their viscosity increases with increase in molecular weight. Solid PEGs are usually white in color and available as pastes, waxy flakes, or free-flowing solids based on their molecular weight. Table 2 shows the physicochemical properties of some PEGs.

PEGs are hydrophilic materials and are extensively used in the preparation of hydrophilic ointments and creams. They are nonirritants and are easily washed from skin surfaces. Products with varying consistency are prepared by mixing different grades of PEGs. Excessive heating is avoided while melting PEGs. This will prevent oxidation and discoloration of products. In addition, use of purified grades that are free from peroxide impurities, inclusion of suitable antioxidants, and heating under nitrogen atmosphere can minimize the oxidation. PEGs are prone to etherification or esterfication reactions due to the presence of two terminal hydroxyl groups. They are incompatible with some antibiotics, antimicrobial preservatives, iron, tannic acid, and salicylic acid and also interact with plastic containers made of polyvinyl chloride and polyethylene. PEG-containing products are usually packed in aluminum, glass, or stainless steel containers to avoid such interactions. Although low-molecular-weight PEGs are hygroscopic, they do not promote microbial growth. PEG-containing products are generally stored in well-closed containers in a cool, dry place. These products can cause stinging sensation on mucus and some hypersensitivity reactions, especially when applied onto open wounds [7, 8].

***Stearic Acid*** Stearic acid is obtained by hydrolysis of fat or hydrogenation of vegetable oils. Compendial stearic acid contains a mixture of stearic acid and palmitic acids. It is available as powder or crystalline solid which is white to yellowish white in color and possesses a characteristic odor. Although stearic acid is insoluble

in water, partially neutralized grades form a cream base when combined with about 10 times its weight of aqueous solvents. The appearance and consistency of these grades are based on the proportion of alkali or triethanolamine used for neutralization. Concentrations up to 20% are used for formulating creams and ointments. Different grades of stearic acids are commercially available with varying stearic acid content, melting temperature, and other physical properties. A suitable antioxidant is included in formulations containing stearic acid. As stearic acid interacts with metals, it is avoided in preparations which contain salts, especially divalent metals such as calcium and zinc. It also reacts with metal hydroxides and some drugs. Compatibility evaluation between stearic acid and other formulation components is therefore essential when formulating newer products with stearic acid [9].

***Carnauba Wax*** Carnauba wax contains a mixture of esters of acids and hydroxyacids isolated from Brazilian carnauba palm. It also contains various resins, hydrocarbons, acids, polyhydric alcohols, and water. It is available as lumps, powder, or flakes which are brown to pale yellow in color and possesses a characteristic odor. Carnauba wax is practically insoluble in water and melts at 80–88 °C. Being a hard material, it improves the stiffness of topical preparations [6].

***Cetyl Alcohol*** Cetyl alcohol is obtained by hydrogenolysis or esterfication of fatty acids and contains not less than 90% cetyl alcohol along with other aliphatic alcohols. It is available as flakes or granules white in color and possesses a characteristic odor. Different grades are commercially available with varying proportions of cetyl alcohol, stearyl alcohol, and related alcohols. Although insoluble in water, cetyl alcohol has good water-absorptive and emulsifying properties. This property makes it suitable for preparing emollient ointments and creams. Its viscosity-enhancing properties reduce coalescence of dispersed phase and improves the physical stability of creams. Concentrations ranging from 2 to 10% are used in topical preparations to impart emollient, emulsifying, water-absorptive, and stiffening properties. Mixtures of petrolatum and cetyl alcohol are sometimes used for preparing creams. Such mixtures minimize the quantity of additional emulsifying agents in preparations. Although cetyl alcohol forms stable preparations, it is incompatible with strong oxidizing materials and some drugs. Compatibility studies are therefore conducted when including cetyl alcohol into formulations. Highly purified grades are free from hypersensitivity reactions [3, 10].

***Emulsifying Wax*** Emulsifying wax, also known as anionic emulsifying wax, is a mixture of cetostearyl alcohol, sodium lauryl sulfate, and purified water. Emulsifying wax BP contains about 90% cetostearyl alcohol, 10% sodium lauryl sulfate, and 4% purified water. Emulsifying wax USP contains nonionic surfactants. It is available as flakes or solids which are white to pale yellow in color and possesses a characteristic odor. Although emulsifying wax is insoluble in water, its emulsifying properties help in preparing hydrophilic oil-in-water emulsions. Ointment bases are prepared by mixing up to 50% emulsifying wax with liquid or soft paraffins. At concentrations up to 10%, it forms creams. Although emulsifying wax is compatible with many acids and alkalis, it is incompatible with many cationic materials and polyvalent metal salts. Stainless steel vessels are preferred for mixing operations.

Preparations containing emulsifying wax are preserved in well-closed container in a cool, dry place [11].

***Cetyl Esters Wax*** Cetyl esters wax is obtained by esterification of some fatty alcohols and fatty acids. It is available as crystalline flakes which are white to off-white in color and possesses a characteristic aromatic odor. It is insoluble in water and has emollient and stiffening properties. About 10% of cetyl ester wax is used for preparing hydrophobic creams and about 20% is used for preparing topical ointments. Various grades of cetyl esters wax are available commercially and vary in their fatty alcohol and fatty acids content and melting range. As this wax is incompatible with strong acids and bases, it should be avoided in certain formulations. Cetyl ester wax–containing products are stored in well-closed containers in a cool, dry place [6].

***Hydrogenated Castor Oil*** It is used as stiffening agent in hydrophobic ointments and creams due to its higher melting point. Hydrogenated castor oil contains triglyceride of hydroxystearic acid and is available as white color flakes or powder. It is insoluble in water and melts at 85–88 °C. Different grades with varying compositions and physical properties are commercially available. Products can be prepared at higher temperatures, as hydrogenated castor oil is stable up to 150 °C. It is compatible with other waxes obtained from vegetable and animal sources. Preparations containing hydrogenated castor oil need to be preserved in well-closed containers in a cool and dry place [12].

***Microcrystalline Wax*** Microcrystalline wax is obtained from petroleum by solvent fractionation and dewaxing procedures. It contains many straight-chain and branched-chain alkanes, with carbon chain lengths ranging from 41 to 57. It is available as fine flakes or crystals which are white or yellow in color. Microcrystalline wax is insoluble in water and possesses a wide melting range (54–102 °C). High-melting and stiffening properties of microcrystalline wax make it suitable for preparing ointments and cream with higher consistency. Acids, alkalis, oxygen, and light do not affect its stability [6].

***Stearyl Alcohol*** Reduction of ethyl stearate in the presence of lithium aluminum hydride yields stearyl alcohol, which contains not less than 90% of 1-octadecanol. It is available as flakes or granules which are white in color and possesses a characteristic odor. It is insoluble in water and melts at 55–60 °C. Stearyl alcohol has stiffening, viscosity-enhancing, and emollient properties and hence is used in the preparation of hydrophobic ointments and creams. Its weak emulsifying properties help in improving the water-holding capacity of ointments. Hypersensitivity reactions are sometimes observed due to the presence of some minor impurities. Stearyl alcohol preparations are compatible with acids and alkalis and are preserved in well-closed containers in a cool and dry place [6].

***White Wax*** White wax is a bleached form of yellow wax which is usually obtained from the honeycomb of bees and hence is known as *bleached wax* or *white bees wax*. It contains about 70% esters of straight-chain monohydric alcohols, 15% free acids,

12% carbohydrates, and 1% free wax alcohols and stearic esters of fatty acids. It is available as granules or sheets which are white in color and possesses a characteristic odor. White wax is insoluble in water and melts between 61 and 65 °C. It has stiffening and viscosity-enhancing properties and therefore is used in hydrophobic ointments and oil-in-water creams. Although it is thermally stable, heating to above 150 °C results in reduction of its acid value. White wax is incompatible with oxidizing agents. The presence of small quantities of impurities results in hypersensitivity reactions in rare occasions. Preparations are stored in well-closed, light-resistant containers in a cool, dry place [13].

***Yellow Wax*** Yellow wax, also known as yellow beeswax, is obtained from honey combs. It contains about 70% esters of straight-chain monohydric alcohols, 15% free acids, 12% carbohydrates, and 1% free wax alcohols and stearic esters of fatty acids. It is available as noncrystalline pieces which are yellow in color and possesses a characteristic odor. It is practically insoluble in water and melts at 61–65 °C. It is used in the preparation of hydrophobic ointments and water-in-oil creams because of its viscosity-enhancing properties. Concentrations up to 20% are used for producing ointments and creams. It is incompatible with oxidizing agents. Esterification occurs while heating to 150 °C and hence should be avoided during preparation. Hypersensitivity reactions sometimes occur on topical application of yellow wax–containing ointments and creams due to the presence of some minor impurities. These products are preserved in well-closed, light-resistant containers [13].

Combinations of bases are sometimes used to acquire better stability. Gelling agents such as carbomers and PEG are also included in some ointment and cream preparations. Table 3 shows examples of cream bases used in some commercial cream preparations.

### 2.3 Preparation and Packaging

In addition to the base and drug, ointments and creams may also contain other components such as stabilizers, preservatives, and levigating agents. Usually levigation and fusion methods are employed for incorporating these components into the base. Levigation involves simple mixing of base and other components over an ointment slab using a stainless steel ointment spatula. A fusion process is employed only when the components are stable at fusion temperatures. Ointments and creams containing white wax, yellow wax, paraffin, stearyl alcohol, and high-molecular-weight PEGs are generally prepared by the fusion process. Selection of levigation or the fusion method depends on the type base, the quantity of other components, and their solubility and stability characteristics.

Oleaginous ointments are prepared by both levigation and fusion processes. Small quantities of powders are incorporated into hydrocarbon bases with the aid of a levigating agent such as liquid petrolatum, which helps in wetting of powders. The powder component is mixed with the levigating agent by trituration and is then incorporated into the base by spatulation. All solid components are milled to finer size and screened before incorporating into the base to avoid gritty sensation of the final product. Roller mills are used for producing large quantities of ointments in pharmaceutical industries. Uniform mixing can be obtained by the geometric dilution procedure, which usually involves stepwise dilution of solids into the ointment

**TABLE 3 Cream Bases Present in Some Commercial Creams**

| Commercial Name | Drug | Cream Base (s) Used |
|---|---|---|
| Dritho-Calp, Psoriatec | Anthralin, 0.5%, 1.0% | White petrolatum, cetostearyl alcohol |
| Temovate E | Clobetasol propionate, 0.05% | Propylene glycol, glyceryl monostearate, cetostearyl alcohol, glyceryl stearate, PEG 100 stearate, white wax |
| Eurax | Crotamiton, 10% | Petrolatum, propylene glycol, cetyl alcohol, carbomer-934 |
| Topicort | Desoximetasone, 0.25% | White petrolatum USP, isopropyl myristate NF, lanolin alcohols NF, mineral oil USP, cetostearyl alcohol NF |
| Apexicon, Maxiflor, Psorcon | Diflorasone diacetate, 0.05% | Hydrophilic vanishing cream base of propylene glycol, stearyl alcohol, cetyl alcohol |
| Lidex Cream, Vanos | Fluocinonide, 0.05%, 0.10% | Polyethylene glycol 8000, propylene glycol, stearyl alcohol |
| Carac | Fluorouracil, 0.5%, 1.0%, 5.0% | Carbomer-940, PEG 400, propylene glycol, stearic acid |
| Halog | Halcinonide, 0.1% | Polyethylene and mineral oil gel base with PEG 400, PEG 6000, PEG 300, PEG 1450 |
| Cortaid, Anusol-Hc, Proctosol HC | Hydrocortisone, 2.5% water washable | Petrolatum, stearyl alcohol, propylene glycol, carbomer-934 |
| Monistat-Derm | Miconazole nitrate, 2% | Water-miscible base consisting of pegoxol 7 stearate, peglicol 5 oleate, mineral oil, butylated hydroxyanisole |

base. The fusion method is followed when the drugs and other solids are soluble in the ointment bases. The base is liquefied, and the soluble components are dissolved in the molten base. The mixture is then allowed to congeal by cooling. Fusion is performed using steam-jacketed vessels or a porcelain dish. The congealed mixture is then spatulated or triturated to obtain a smooth texture. Care is taken to avoid thermal degradation of the base or other components during the fusion process.

Absorption-type ointments and creams are prepared by incorporating large quantities of water into hydrocarbon bases with the aid of a hydrophobic emulsifying agent. Water-insoluble drugs are added by mechanical addition or fusion methods. As with oleaginous ointments, levigating agents are also included to improve wetting of solids. Water-soluble or water-miscible agents such as alcohol, glycerin, or propylene glycol are used if the drug needs to be incorporated into the internal aqueous phase. If the drug needs to be incorporated into the external oily phase, mineral oils are used as the levigating agent. Incorporation of water-soluble components is achieved by slowly adding the aqueous drug solution to the hydrophobic base using pill tile and spatula. If the proportion of aqueous phase is larger, inclusion of additional quantities of emulsifier and application of heat may be needed to achieve uniform dispersion. Care must be taken to avoid excessive heating as it can result in evaporation aqueous phase and precipitation of water-soluble components and formation of stiff and waxy product.

Water-removable ointments and creams are basically hydrophilic-type emulsions. They are prepared by fusion followed by mechanical addition approach. Hydrocarbon components are melted together and added to the aqueous phase that contains water-soluble components with constant stirring until the mixture congeals. A hydrophilic emulsifying agent is included in the aqueous phase in order to obtain stable oil-in-water dispersion. Sodium lauryl sulfate is used in the preparation of hydrophilic ointment USP.

Water-soluble ointments and creams do not contain any oily phase. Both water-soluble and water-insoluble components are incorporated into water-soluble bases by both levigation and fusion methods. If the drug and other components are water soluble, they are dissolved in a small quantity of water and incorporated into the base by simple mixing over an ointment slab. If the components are insoluble in water, aqueous levigating agents such as glycerin, propylene glycol, or a liquid PEG are used. The hydrophobic components are mixed with the levigating agent and then incorporated into the base. Heat aids incorporation of a large quantity of hydrophobic components.

A wide range of machines are available for the large-scale production of ointments and creams. Each of these machines is designed to perform certain unit operations, such as milling, separation, mixing, emulsification, and deaeration. Milling is performed to reduce the size of actives and other additives. Various fluid energy mills, impact mills, cutter mills, compression mills, screening mills, and tumbling mills are used for this purpose. Alpine, Bepex, Fluid Air, and Sturtevant are some of the manufacturers of these mills. Separators are employed for separating materials of different size, shape, and densities. Either centrifugal separators or vibratory shakers are used for separation. Mixing of the actives and other formulation components with the ointment or cream base is performed using various types of low-shear mixers, high-shear mixers, roller mills, and static mixers. Mixers with heating provisions are also used to aid in the melting of bases and mixing of components. Chemineer, Fryma, Gate, IKA, Koruma (Romaco), Moorhouse-Cowles, Ross, and Stokes Merrill are some of the manufacturers of semisolids mixers.

Creams are produced with the help of low-shear and high-shear emulsifiers. These emulsifiers are used to disperse the hydrophilic components in the hydrophobic dispersion phase (e.g., water-in-oil creams) or oleaginous materials in aqueous dispersion medium (oil-in-water creams). Bematek, Fryma, Koruma (Romaco), Lightnin, Moorhouse, and Ross supply various types of emulsifiers. Entrapment of air into the final product due to mixing processes is a common issue in the large-scale manufacturing of semisolid dosage forms. Various offline and in-line deaeration procedures are adopted to minimize this issue. Effective deaeration is generally achieved by using vacuum vessel deaerators. Some of the recent large-scale machines are designed to perform heating, high-shear mixing, scrapping, and deaeration processes in a single vessel. Figure 1 shows the design feature of a semisolid production machine manufactured by Ross.

Various low- and high-shear shifters are used to transfer materials from the production vessel to the packaging machines. In the packaging area, various types of holders (e.g., pneumatic, gravity, and auger holders), fillers (e.g., piston, peristaltic pump, gear pump, orifice, and auger fillers), and sealers (e.g., heat, torque, microwave, indication, and mechanical crimping sealers) are used to complete the unit

**FIGURE 1** Semisolid production machine with heat jacketed vessel, high-shear mixer, scrapper, vacuum attachments, and control station. (Courtesy of Ross, Inc.)

operations. These equipments are supplied by various manufacturers, namely Bosch, Bonafacci, Erweka, Fryma-Maschinenbau, IWKA, Kalish, and Norden.

Sterility of ointments, especially those intended for ophthalmic use, is achieved by aseptic handling and processing. Improper processing, handling, packing, or use of ophthalmic ointments lead to microbial contaminations and eventually result in ocular infections. In general, the empty containers are separately sterilized and filled under aseptic condition. Final product sterilization by moist heat sterilization or gaseous sterilization is ineffective because of product viscosity. Dry-heat sterilization is associated with stability issues. Strict aseptic procedures are therefore practiced when processing ophthalmic preparations. Antimicrobial preservatives such as benzalkonium chloride, phenyl mercuric acetate, chlorobutanol, or a combination of methyl paraben and propyl paraben are included in ophthalmic ointments to retain microbial stability.

***Packaging*** An ideal container should protect the product from the external atmosphere such as heat, humidity, and particulates, be nonreactive with the product components, and be easy to use, light in weight, and economic [14]. As tubes made of aluminum and plastic meet most of these qualities, they are extensively used for packaging semisolids. Aluminum tubes with special internal epoxy coatings are commercially available for improving the compatibility and stability of products. Various modified plastic materials are used for making ointment tubes. Tubes made

**FIGURE 2** Custom-designed LDPE containers made by BFS process for packaging topical products. (Courtesy of Rommelag USA, Inc.)

of low-density polyethylene (LDPE) are generally soft and flexible and offer good moisture protection. Tubes made of high-density polyethylene (HDPE) are relatively harder but offer high moisture protection. Polypropylene containers offer high heat resistance. Plastic containers made of polyethylene terephthalate (PET) are transparent and provide superior chemical compatibility. Ointments meant for ophthalmic, nasal, rectal, and vaginal applications are supplied with special application tips for the ease of product administration.

A recent method known as blow fill sealing (BFS) performs fabrication of container, filling of product, and sealing operations in a single stage and hence is gaining greater attention. The products can be sterile filled, which makes BFS a cost-effective alternative for aseptic filling. All plastic materials are suitable for BFS processing. In most cases, monolayered LDPE materials are used for making small-size containers. If the product is not compatible with the LDPE or sensitive to oxygen, barrier layers are added to the container wall by coextrusion methods. As the container is formed inside the BFS machine, upstream handling problems are avoided. The BFS machine can hand the container off to any secondary packaging operation that needs to be performed. Typically a secondary overwrap is added to the containers prior to cartooning. An additional advantage of BFS containers is the integrated design of the applicator into the product container. Figure 2 shows some of the custom-designed BFS containers for topical products.

### 2.4 Evaluation

Ointments and creams are evaluated for various pharmacopeial and nonpharmacopeial tests to ascertain their physicochemical, microbial, in vitro, and in vivo characteristics. These tests help in retaining their quality and minimizing the batch-to-batch variations. The USP recommends storage and labeling, microbial screening, minimum fill, and assays for most ointments and creams. Tables 4 and 5 summarize the compendial requirements for some pharmacopeial ointments and creams.

**TABLE 4 USP Specifications for Some Official Ointments**

| Drug | Quality Control Tests | Packaging and Storage Requirements |
|---|---|---|
| Acyclovir | *Staphylococcus aureus, Pseudomonas aeruginosa*, minimum fill, limit of guanine, and assay | Tight containers; store between 15 and 25 °C in a dry place |
| Alclometasone dipropionate | *S. aureus, P. aeruginosa*, minimum fill, and assay | Collapsible tubes or tight containers, store at controlled room temperature |
| Amphotericin B | Minimum fill, water, and assay | Collapsible tubes or other well-closed containers |
| Anthralin | Assay | Tight containers; in a cool place; protect from light |
| Bacitracin | Minimum fill, water, and assay | Well-closed containers containing not more than 60 g; controlled temperature |
| Benzocaine | *S. aureus, P. aeruginosa*, minimum fill, and assay | Tight containers; protect from light; avoid prolonged exposure to temperatures exceeding 30 °C |
| Betamethasone valerate | *S. aureus, P. aeruginosa*, minimum fill, and assay | Collapsible tubes or tight containers; avoid exposure to excessive heat. |
| Clioquinol | Assay | Collapsible tubes or tight, light-resistant containers |
| Clobetasol propionate | *S. aureus, P. aeruginosa, Escherichia coli, Salmonella* species, total aerobic microbial count, minimum fill, and assay | Collapsible tubes or in tight containers; store at controlled room temperature; do not refrigerate |
| Erythromycin | Minimum fill, water, and assay | Collapsible tubes or in tight containers at controlled room temperature |
| Fluocinolone acetonide | *S. aureus, P. aeruginosa*, and assay | Tight containers |
| Gentamycin sulfate | Minimum fill, water, and assay | Collapsible tubes or in tight containers; avoid exposure to excessive heat |
| Hydrocortisone valerate | *S. aureus, P. aeruginosa*, total microbial count, minimum fill, and assay | Tight container; store at room temperature |
| Ichthammol | Assay | Collapsible tubes or in tight containers; avoid prolonged exposure to temperatures exceeding 30 °C |
| Lidocaine | *S. aureus, P. aeruginosa*, minimum fill, and assay | Tight containers |
| Mometasone furoate | *S. aureus, P. aeruginosa, E. coli, Salmonella* species, minimum fill, and assay | Well-closed containers |
| Nitrofurazone | Completeness of solution and assay | Tight, light-resistant containers; avoid exposure to direct sunlight, strong fluorescent lighting, and excessive heat |
| Nitroglycerine | Minimum fill, homogeneity, and assay | Tight containers |
| Nystatin | Minimum fill, water, and assay | Well-closed containers at controlled room temperature |
| Tetracycline hydrochloride | Minimum fill, water, and assay | Well-closed containers at controlled room temperature |
| Zinc oxide | Minimum fill, calcium, magnesium, other foreign substances, and assay | Tight containers; avoid prolonged exposure to temperatures exceeding 30 °C |

**TABLE 5 USP Specifications for Some Official Creams**

| Cream | Quality Control Tests | Packaging and Storage Requirements |
|---|---|---|
| Alclometasone dipropionate | Microbial limits, minimum fill, and assay | Collapsible tubes or tight containers; store at controlled room temperature |
| Amphotericin B | Minimum fill and assay | Collapsible tubes or other well-closed containers |
| Benzocaine | Microbial limits, minimum fill, and assay | Tight containers, protected from light, and avoid prolonged exposure to temperatures exceeding 30 °C |
| Betamethasone dipropionate | Minimum fill and assay | Collapsible tubes or tight containers; store at 25 °C; excursions permitted between 15 and 30; protect from freezing |
| Ciclopirox olamine | Minimum fill, pH, content of benzyl alcohol, and assay | Collapsible tubes at controlled room temperature |
| Clobetasol propionate | Microbial limits, minimum fill, pH, and assay | Collapsible tubes or tight containers; store at controlled room temperature; do not refrigerate |
| Clotrimazole | Assay | Collapsible tubes or tight containers at a temperature between 2 and 30 °C |
| Desoximetasone | Minimum fill, pH, and assay | Collapsible tubes at controlled room temperature |
| Dibucaine | Microbial limits, minimum fill, and assay | Collapsible tubes or in tight, light-resistant containers |
| Dienestrol | Minimum fill and assay | Collapsible tubes or in tight containers |
| Diflorasone diacetate | Microbial limits, minimum fill, and assay | Collapsible tubes, preferably at controlled room temperature |
| Fluocinolone acetonide | Microbial limits, minimum fill, and assay | Collapsible tubes or in tight containers |
| Fluorouracil | Microbial limits, minimum fill, and assay | Tight containers and stored at controlled room temperature |
| Gentamycin sulfate | Minimum fill and assay | Collapsible tubes or in other tight containers; avoid exposure to excessive heat |
| Hydrocortisone butyrate | Microbial limits, minimum fill, pH, and assay | Well-closed containers |
| Hydroquinone | Minimum fill and assay | Well-closed, light-resistant containers |
| Lindane | pH and assay | Tight containers |
| Meclocycline sulfosalicylate | Minimum fill and assay | Tight containers, protected from light |
| Miconazole nitrate | Minimum fill and assay | Collapsible tubes or tight containers; store at controlled room temperature |
| Monobenzone | Assay | Tight containers; avoid exposure to temperatures higher than 30 °C |
| Nystatin | Minimum fill and assay | Collapsible tubes or in other tight containers; avoid exposure to excessive heat |
| Prednisolone | Minimum fill and assay | Collapsible tubes or in tight containers |
| Tetracaine hydrochloride | Microbial limits, minimum fill, pH between 3.2 and 3.8, and assay | Collapsible, lined metal tubes |
| Triamcinolone acetonide | Microbial limits, minimum fill, and assay | Tight containers |

***Packaging and Storage*** The USP recommends packaging and storage requirements for each official ointment and cream. Generally collapsible tubes, tight containers, or other well-closed containers are recommended for packing. They are stored in either a cool place or at controlled room temperatures. In some cases, special storage conditions are recommended: for example, protect from light, avoid exposure to excessive heat, avoid exposure to direct sunlight, avoid strong fluorescent lighting, do not refrigerate, and avoid prolonged exposure to temperatures exceeding 30 °C.

***Minimum Fill*** This test is performed to compare the weight or volume of product filled into each container with their labeled weight or volume. It helps in assessing the content uniformity of product. A minimum-fill test is applied only to those containers that contain not more than 150 g or mL of preparation. It is performed in two steps. Initially, labels from the product containers are removed. After washing and drying the surface, their weights are recorded ($W_1$). In the second step, the entire product from each container is removed. After cleaning and drying, the weight of empty containers is recorded ($W_2$). The difference between total weight ($W_1$) and empty-container weight ($W_2$) gives the weight of product. The USP recommends that the average net content of 10 containers should not be less than the labeled amount. If the product weight is less than 60 g or mL, the net content of any single container should not be less than 90% of the labeled amount. If the product weight is between 60 and 150 g or mL, the net content of any single container should not be less than 95% of the labeled amount. If these limits are not met, the test is repeated with an additional 20 containers. All semisolid topical preparations should meet these specifications [15].

***Water Content*** The presence of minor quantities of water may alter the microbial, physical, and chemical stability of ointments and creams. Titrimetric methods (method I) are usually performed for determining the water content in these preparations. These methods are based on the quantitative reaction between water and anhydrous solution of sulfur and iodine in the presence of a buffer that can react with hydrogen ions. Special titration setups and reagents (Karl Fischer, KF) are used in these determinations. In the direct method (method Ia), about 35 mL of methanol is titrated with sufficient quantity of KF reagent to the electrometric or visual endpoint (color change from canary yellow to amber). This blank titration helps to consume any moisture that may be present in the reaction medium. A known quantity of test material (ointment or cream) is added to the reaction medium, mixed, and again titrated with KF reagent to the reaction endpoint. The water content is determined by considering the volume of KF reagent consumed and its water equivalence factor. In the residual titration method (method Ib), a known excess quantity of KF reagent is added to the titration vessel, which is then back titrated with standardized water to the electrometric or visual endpoint. In the coulometric titration method (method Ic), the sample is dissolved in anhydrous methanol and injected into the reaction vessel that contains the anolyte, and the coulometric reaction is performed until the reaction endpoint. In some cases, methanol is replaced with other solvents. The maximum allowable limit of water in ointment preparations varies between 0.5 and 1.0%. The limit of water in bacitracin, chlortetracycline hydrochloride, and nystatin ointments is not more than 0.5%, whereas amphotericin

B, erythromycin, gentamycin sulfate, neomycin sulfate, and tetracycline hydrochloride ointments may contain up to 1% moisture [15].

***Metal Particles*** This test is required only for ophthalmic ointments. The presence of metal particles will irritate the corneal or conjunctival surfaces of the eye. It is performed using 10 ointment tubes. The content from each tube is completely removed onto a clean 60-mm-diameter petridish which possesses a flat bottom. The lid is closed and the product is heated at 85 °C for 2 h. Once the product is melted and distributed uniformly, it is cooled to room temperature. The lid is removed after solidification. The bottom surface is then viewed through an optical microscope at 30× magnification. The viewing surface is illuminated using an external light source positioned at 45° on the top. The entire bottom surface of the ointment is examined, and the number of particles 50 μm or above are counted using a calibrated eyepiece micrometer. The USP recommends that the number of such particles in 10 tubes should not exceed 50, with not more than 8 particles in any individual tube. If these limits are not met, the test is repeated with an additional 20 tubes. In this case, the total number of particles in 30 tubes should not exceed 150, and not more than 3 tubes are allowed to contain more than 8 particles [15].

***Leakage Test*** This test is mandatory for ophthalmic ointments, which evaluates the intactness of the ointment tube and its seal. Ten sealed containers are selected, and their exterior surfaces are cleaned. They are horizontally placed over absorbent blotting paper and maintained at 60 ± 3 °C for 8 h. The test passes if leakage is not observed from any tube. If leakage is observed, the test is repeated with an additional 20 tubes. The test passes if not more than 1 tube shows leakage out of 30 tubes [15].

***Sterility Tests*** Ophthalmic semisolids should be free from anaerobic and aerobic bacteria and fungi. Sterility tests are therefore performed by the membrane filtration technique or direct-inoculation techniques. In the membrane filtration method, a solution of test product (1%) is prepared in isopropyl myristate and allowed to penetrate through cellulose nitrate filter with pore size less than 0.45 μm. If necessary, gradual suction or pressure is applied to aid filtration. The membrane is then washed three times with 100-mL quantities of sterile diluting and rinsing fluid and transferred aseptically into fluid thioglycolate (FTG) and soybean–casein digest (SBCD) medium. The membrane is finally incubated for 14 days. Growth on FTG medium indicates the presence of anaerobic and aerobic bacteria, and SBCD medium indicates fungi and aerobic bacteria. Absence of any growth in both these media establishes the sterility of the product. In the direct-inoculation technique, 1 part of the product is diluted with 10 parts of sterile diluting and rinsing fluid with the help of an emulsifying agent and incubated in FTG and SBCD media for 14 days. In both techniques, the number of test articles is based on the batch size of the product. If the batch size is less than 200 the containers, either 5% of the containers or 2 containers (whichever is greater) are used. If the batch size is more than 200, 10 containers are used for sterility testing [15].

***Microbial Screening*** Semisolid preparations are required to be free from any microbial contamination. Hence, most of the topical ointments are screened for the

presence of *Staphylococcus aureus* and *Pseudomonas aeruginosa*. In some cases, screening for *Escherichia coli, Salmonella* species, and total aerobic microbial counts is recommended by the USP. For instance, clobetasol propionate ointment USP and mometasone furoate ointment USP are screened for all these organisms. In addition, preparations meant for rectal, vaginal, and urethral applications are tested for yeasts and molds [15].

*Test for* S. aureus *and* P. aeruginosa The test sample is mixed with about 100 mL of fluid soybean–casein digest (FSBCD) medium and incubated. If microbial growth is observed, it is inoculated in agar medium by the streaking technique. Vogel–Johnson agar (VJA) medium is used for *S. aureus* screening, and cetrimide agar (CA) medium is used for screening *P. aeruginosa*. The petridishes are then closed, inverted, and incubated under appropriate conditions. The appearance of black colonies surrounded by a yellow zone over VJA medium and greenish colonies in CA medium indicates the presence of *S. aureus* and *P. aeruginosa*, respectively. Various other agar media are also available for screening these organisms. A coagulase test is then performed for confirming the presence of *S. aureus* and oxidase and pigment tests for confirming *P. aeruginosa*.

*Test for* Salmonella *Species and* E. coli The test sample is mixed with about 100 mL of fluid lactose (FL) medium and incubated. If microbial growth is observed, the contents are mixed and 1 mL is transferred to vessels containing 10 mL of fluid selinite cystine (FSC) medium and fluid tetrathionate (FT) medium and incubated for 12–24 h under appropriate conditions. To identify the presence of *Salmonella*, samples from the above two media are streaked over brilliant green agar (BGA) medium, xylose lysine desoxycholate agar (XLDA) medium, and bismuth sulfite agar (BSA) medium and incubated. The appearance of small, transparent or pink-to-white opaque colonies over BGA medium, red colonies with or without black centers over XLDA medium, and black or green colonies over BSA medium indicates the presence of *Salmonella*. It is further confirmed in triple sugar iron agar medium. The presence of *E. coli* is screened by streaking the samples from FL medium over MacConkey agar medium. The appearance of brick red colonies indicates the presence of *E. coli*. It is further confirmed using Levine eosin methylene blue agar medium.

TOTAL AEROBIC MICROBIAL COUNTS The plate method or multiple-tube method is performed to estimate the total count. About 10 g or 10 mL of the test sample is dissolved or suspended in sufficient volume of phosphate buffer (pH 7.2), fluid soybean casein digest (FSBCD) medium, or fluid casein digest–soy lecithin–polysorbate 20 medium to make the final volume 100 mL. In the plate method, about 1 mL of this diluted sample is mixed with molten soybean–casein digest agar (SBCDA) medium and solidified at room temperature. The plates are inverted and incubated for two to three days. The number of colonies that are on the surface of nutrient media are counted. The multiple tube method is performed using sterile fluid SBCD medium. The number of colonies formed should not exceed the limits specified in an individual monograph. For example, clobetasol propionate ointment USP and hydrocortisone valerate ointment USP contains less than 100 colony-forming units (CFU) per gram of sample.

*Test for Yeasts and Molds* The plate method is used for testing molds and yeast in semisolids. The procedure is similar to that of the total count test. Instead of SBCDA medium, Sabouraud dextrose agar (SDA) medium or potato dextrose agar (PDA) medium is used. Samples are incubated for five to seven days at 20–25 °C to identify the presence of yeasts and molds.

***Assay*** The quantity of drug present in a unit weight or volume of ointment or cream is determined by various methods. Spectrophotometric, titrimetric, chromatographic, and in some cases microbial assays are performed. Selection of a particular method is based on the nature of drug, its concentration in the product, interference between the drug and other formulation components, and official requirements. Although spectrophotometric methods are accurate and easy to perform, the complexity of ointment matrix sometimes reduced the specificity of analysis compare to liquid chromatographic methods. The USP prescribes high-performance liquid chromatographic (HPLC) assays for many official ointments due to its specificity, accuracy, and precision. For example, amcinonide, anthralin, betamethasone dipropionate, clobetasol propionate, dibucaine, nitroglycerine, hydrocortisone, and triamcinolone acetonide are assayed by HPLC methods. These methods involve extraction of drug from the formulation matrix using suitable solvents followed by chromatographic separation using suitable reversed-phase columns followed by ultraviolet (UV) detection. Clioquinol preparation is assayed by gas chromatography. The USP also recommends potentiometric titrations (benzocaine, lidocaine, and ichthammol) and complexometric titrations (zinc oxide) for some semisolid preparations.

Microbial assays are recommended for certain preparations containing antibiotics such as amphotericin B, bacitracin, chlortetracycline hydrochloride, gentamycin sulfate, neomycin sulfate, and nystatin. These tests evaluate the potency of an antibiotic by means of its inhibitory effects on specific microorganism. Two types of microbial assays are performed to determine the antibiotic potency. They are known as cylindrical plate or plate assays and turbidimetric or tube assays. The plate method measures the extent of growth inhibition of a particular microorganism in solidified agar medium in the presence of the test antibiotic (commonly known as *zone of inhibition*). The tube method measures the turbidity of a liquid medium that contains a particular organism in the presence and absence of the test antibiotic. These methods involve extracting drug from the formulation matrix, diluting the drug to a known concentration, and measuring the zone of inhibition or turbidity.

***In Vitro Drug Release Studies*** These studies are conducted to ascertain release of drug from the formulation matrix. Open-chamber diffusion cells such as Franz cells are used for performing in vitro studies. These cells consist of a donor side and a receiver side separated by a synthetic membrane such as cellulose acetate/nitrate mixed ester, polysulfone, or polytetrafluoroethylene. The membranes are usually pretreated with the receiver fluid to avoid any lag phase in drug release. The receiver side is filled with a known volume of release medium and is heated to 32 ± 0.5 °C by circulating warm water through an outer jacket. Aqueous buffers are used for water-soluble drugs. Phosphate buffer solution of pH 5.4 is considered most appropriate for dermatological products as it mimics the pH of skin. Hydroalcoholic or other suitable medium may also be used for sparingly water soluble drugs. A known quantity of the test product is applied uniformly over the membrane on the donor

side and samples are withdrawn from the receiver side at different time intervals. After each sampling, an equal volume of fresh medium is replaced to the receiver side. The sampling time points are different for different formulations; however, at least five samples are withdrawn during the study period for determining the release rate. A typical sample time sequence for a 6-h study is 0.5, 1.0, 2.0, 4.0, and 6.0 h. The receiver samples are analyzed by a suitable analytical method to quantify the amount of drug released from the formulation at different time intervals. The slope of the straight line which is obtained from a plot of cumulative amount drug release across 1-$cm^2$ membrane versus the square root of time represents the release rate. Experiments are conducted in hexaplicate to obtain statistically significant results [16].

***In Vivo Bioequivalence Studies*** In vivo studies are conducted to establish the biological availability or activity of the drug from a topically applied semisolid formulation. Dermatopharmacokinetic studies, pharmacodynamic studies, or comparative clinical trials are generally conducted to assess the bioequivalence of topical products [16, 17].

Dermatopharmacokinetic (DPK) studies are applicable for topical semisolid products that contain antifungals, antivirals, corticosteroids, and antibiotics and vaginally applied products. They are not applicable for ophthalmic, otic, and other products that damage stratum corneum. DPK studies involve measurement of drug concentrations in stratum corneum, drug uptake, apparent steady state, and elimination after application of the test product onto skin. These studies are conducted in healthy human subjects adopting crossover design. The test and the reference products are applied onto eight to nine sites in the forearm. The surface area of each site is based on the strength of drug, extent of drug diffusion, exposure time, and sensitivity of the analytical technique. The application site is washed and allowed to normalize for at least 2 h prior to drug application. A known amount of product is applied onto these selected sites. At appropriate time intervals, the excess of drug from each area is removed using cotton swabs or soft tissue papers. Care is taken to avoid stratum corneum damage during sample collection. Stripping of stratum corneum is performed using adhesive tape-strips (e.g., D-Squame, Transpore). In the elimination phase, the excess drug is removed at the steady-state time point, and the stratum corneum is harvested at succeeding times over 24 h. The drug content from strips from each time point are extracted using suitable solvents and quantified by a validated analytical method. A stratum corneum drug levels–time curve is developed, and pharmacokinetic parameters such as maximum concentration at steady state ($C_{max\text{-}ss}$), time to reach $C_{max\text{-}ss}$ ($T_{max\text{-}ss}$), and areas under the curve for the test and standard ($AUC_{test}$ and $AUC_{reference}$) are computed. DPK studies are performed in either one or two occasions. If performed in one occasion, both arms of a single subject are used to compare the test and reference products. If performed in two occasions, a wash-out period of at least 28 days is allowed to rejuvenate the harvested stratum corneum.

Pharmacodynamic (PD) studies are also performed to estimate the bioavailability and bioequivalence of drugs from topically applied semisolids. In this case, known therapeutic responses from drug products such as skin blanching due to vasoconstrictor effects caused by corticosteroids and transepidermal water loss caused by retinoids are measured and compared between the test and reference

products. Comparative clinical studies are rarely conducted due to the difficulties involved in performing the study, variability in study results, and their poor sensitivity.

### 2.5 Typical Pharmacopeial/Commercial Examples

The vast majority of topical ointments and creams are meant for dermatological applications. They are used to treat various skin conditions such as eczema, dermatitis, allergies, inflammatory and pruritic manifestations, minor skin wounds, pain, insect bites, psoriasis, herpes and other infections of the skin (e.g., impetigo), acne, and precancerous and cancerous skin growths. Similarly, ophthalmic conditions such as infections, inflammation, allergy, and dry-eye symptoms are treated with semisolid preparations. Products are also available for certain eye examinations. Vaginal preparations are available for treating genital herpes, yeast infections, and vaginosis caused by bacteria and to reduce menopausal symptoms (e.g., vaginal dryness), and rectal preparations are available for treating minor pain, itching, swelling, and discomfort caused by hemorrhoids and other problems of the anal area. Tables 6 and 7 show some of the commercially available compendial ointment and cream preparations used for treating various topical ailments.

## 3 GELS

### 3.1 Definition

Gels are semisolid preparations that contain small inorganic particles or large organic molecules interpenetrated by a liquid. Gels made of inorganic materials are usually two-phase systems where small discrete particles are dispersed throughout the dispersion medium. When the particle size of the dispersed phase is larger, they are referred to as magmas. Gels made of organic molecules are single-phase systems, where no apparent physical boundary is seen between the dispersed phase and the dispersion medium. In most cases, the dispersion medium is aqueous. Hydroalcoholic or oleaginous dispersion media are also used in some cases. Unlike dispersed systems like suspensions and emulsions, movement of the dispersed phase is restricted in gels because of the solvated organic macromolecules or interconnecting three-dimensional networks of particles.

Gels are attractive delivery systems as they are simple to manufacture and suitable for administering drugs through skin, oral, buccal, ophthalmic, nasal, otic, and vaginal routes. They also provide intimate contact between the drug and the site of action or absorption. With the advancement in polymer science, gel-based systems that respond to specific biological or external stimuli like pH, temperature, ionic strength, enzymes, antigens, light, magnetic field, ultrasound, and electric current are being designed and evaluated as smart delivery systems for various applications.

### 3.2 Characteristics

Gels may appear transparent or turbid based on the type of gelling agent used. They exhibit different physical properties, namely, imbibition, swelling, syneresis, and

**TABLE 6 Examples of Compendial/Commercial Ointments**

| Drug[a] | Category | Indication | Commercial Names | Strength(s) Available (%) |
|---|---|---|---|---|
| Acyclovir | Antiviral | Genital herpes, herpes infections of the skin, and oral herpes | Zovirax | 5 |
| Atropine sulfate (oph) | Mydriatic | Relax muscles in the eye, treat inflammation of certain parts of the eye (uveal tract), and used for certain eye exams | Atropisol, Isopto Atropine | 0.5, 1 |
| Bacitracin | First-aid antibiotic | Treat or prevent skin infections | Baciguent Oint, Bacitracin Top | 500 units/g |
| Bacitracin (oph) | Antibiotic | Treat or prevent eye infections | Ak-Tracin, Bacticin | 500 units/g |
| Benzocaine | Antipruritic and local anesthetic | Itching, minor skin wound pain, and insect bites | Americaine | 20 |
| Ciprofloxacin (oph) | Antibiotic | Eye infections | Ciloxan | 0.3 |
| Clobetasol propionate | Anti-inflammatory agent | Relieve inflammatory and pruritic manifestations of corticosteroid-responsive dermatoses | Temovate Ointment | 0.05 |
| Erythromycin | Antibiotic | Treatment of acne | Akne-Mycin | 2.0 |
| Erythromycin (oph) | Antibiotic | Infections of eye or ear | Erythromycin Ophthalmic | 0.5 |
| Gentamicin sulfate (oph) | Antibacterial | Infections of eye or ear | Gentamicin Sulfate | 0.3 |
| Hydrocortisone | Anti-inflammatory agent | Minor pain, itching, swelling, and discomfort caused by hemorrhoids and other problems of anal area | Cortaid, Anusol-HC, Proctosol HC | 2.5 |
| Mupirocin | Antibiotic | Treat certain skin infections (e.g., impetigo) | Bactroban | 2.0 |
| Sodium chloride (oph) | Miscellaneous | Treat fluid accumulation in cornea of eye causing swelling | Muro-128, Sochlor | 2.0 |

[a]oph: ophthalmic ointment.

**TABLE 7 Examples of Compendial/Commercial Creams**

| Drug[a] | Category | Indication | Commercial Products | Strength(s) Available (%) |
|---|---|---|---|---|
| Alclometasone dipropionate | Anti-inflammatory | Eczema, dermatitis, allergies, and rash | Aclovate | 0.05 |
| Amcinonide | Anti-inflammatory | Lymphomas of the skin, atopic dermatitis, contact dermatitis, and skin rash | Cyclocort | 0.1 |
| Amphotericin B | Antifungal antibiotic | Treat skin infection due to a *Candida* yeast and diaper rash | Fungizone | 3.0 |
| Anthralin | Keratolytic | Long-term psoriasis | Dritho-Calp, Psoriatec | 0.5, 1.0 |
| Betamethasone dipropionate | Anti-inflammatory | Eczema, dermatitis, allergies, and rash | Diprolene AF | 0.05 |
| Butoconazole nitrate (vag) | Antifungal | Vaginal yeast infections | Gynazole-1 | 2.0 |
| Clindamycin phosphate (vag) | Antibiotic | Vaginosis caused by bacteria | Clindesse | 2.0 |
| Clobetasol propionate | Anti-inflammatory | Inflammatory and pruritic manifestations of corticosteroid-responsive dermatoses | Temovate E Cream | 0.05 |
| Clotrimazole | Antifungal | Ringworm of groin area, athlete's foot, ringworm of the body, fungal infection of the skin with yellow patches, skin infection due to a candida yeast, and diaper rash | Lotrimin | 1.0 |
| Crotamiton | Scabicidal and antipruritic | Scabies infection and itching | Eurax | 10.0 |
| Desoximetasone | Anti-inflammatory | Eczema, dermatitis, allergies, and rash | Topicort | 0.25 |
| Dienestrol | Estrogen | Reduce menopause symptoms such as vaginal dryness | Ortho-Dienestrol | 0.01 |
| Diflorasone diacetate | Anti-inflammatory | Eczema, dermatitis, allergies, and rash | Apexicon, Maxiflor | 0.05 |
| Fluocinonide | Anti-inflammatory and antipruritic | Psoriasis, eczema, dermatitis, allergies, and rash | Lidex, Vanos | 0.05, 0.10 |
| Fluorouracil | Anticancer | Precancerous and cancerous skin growths | Fluoroplex, Carac, Efudex | 0.5, 1.0, 5.0 |

**TABLE 7** ***Continued***

| Drug[a] | Category | Indication | Commercial Products | Strength(s) Available (%) |
|---|---|---|---|---|
| Halcinonide | Anti-inflammatory and antipruritic | Eczema, dermatitis, allergies, and rash | Halog | 0.1 |
| Mometasone furoate | Anti-inflammatory | Eczema, dermatitis, allergies, and rash | Elocon | 0.1 |
| Naftifine hydrochloride | Antifungal | Jock itch, athlete's feet, or ringworm | Naftin | 1 |
| Nystatin | Antifungal | Fungal skin infections | Mycostatin | 100,000 units/g |

[a]vag, vaginal cream.

thixotropy. *Imbibition* refers to the uptake of water or other liquids by gels without any considerable increase in its volume. *Swelling* refers to the increase in the volume of gel by uptake of water or other liquids. This property of most gels is influenced by temperature, pH, presence of electrolytes, and other formulation ingredients. *Syneresis* refers to the contraction or shrinkage of gels as a result of squeezing out of dispersion medium from the gel matrix. It is due to the excessive stretching of macromolecules and expansion of elastic forces during swelling. At equilibrium, the system still maintains its physical stability because the osmotic forces of swelling balance the expanded elastic forces of macromolecules. On cooling, the osmotic pressure of the system decreases and therefore the expanded elastic forces return to normal. This results in shrinkage of the stretched molecules and squeezing of dispersion medium from the gel matrix. Addition of osmotic agents such as sucrose, glucose, and other electrolytes helps in retaining higher osmotic pressure even at lower temperatures and avoids syneresis of gels. *Thixotropy* refers to the non-Newtonian flow nature of gels, which is characterized by a reversible gel-to-sol formation with no change in volume or temperature [18].

### 3.3 Classification

Gels are classified as hydrogels and organogels based on the physical state of the gelling agent in the dispersion. Hydrogels are prepared with water-soluble materials or water-dispersible colloids. Organogels are prepared using water-insoluble oleaginous materials.

***Hydrogels*** Natural and synthetic gums such as tragacanth, sodium alginate, and pectin, inorganic materials such as alumina, bentonite, silica, and veegum, and organic materials such as cellulose polymers form hydrogels in water. They may either be dispersed as fine colloidal particles in aqueous phase or completely dissolve in water to gain gel structure. Gums and inorganic gelling agents form gel structure due to their viscosity-increasing nature. Organic gelling agents which are generally high-molecular-weight cellulose polymer derivatives produce gel structure

because of their swelling and chain entanglement properties. The swollen molecular chains are held together by secondary valence forces, which help in retaining their gel structure. The physical strength of the gel structure is based on the quantity of gelling agent, nature and molecular weight of gelling agent, product pH, and gelling temperature. Generally high-molecular-weight polymers at higher concentrations produce thick gels. The gel-forming temperature (*gel point*) varies with different polymers. Generally natural gums form gel at lower temperatures. Gelatin, a natural protein polymer, forms gel at about 30 °C. If the temperature is increased, gel consistency is not obtained even at higher concentrations of gelatin. On the other hand, polymers such as methylcellulose gain gel structure only when the temperature is above 50 °C due to its decreased solubility and precipitation. Knowledge of the gel point for each gelling agent is therefore essential for preparing physically stable hydrogels.

***Organogels*** Organogels are also known as oleaginous gels. They are prepared using water-insoluble lipids such as glycerol esters of fatty acids, which swell in water and form different types of lyotropic liquid crystals. Widely used glycerol esters of fatty acids include glycerol monooleate, glycerol monopalmito stearete, and glycerol monolinoleate. They generally exist as waxes at room temperature and form cubic liquid crystals in water and increase the viscosity of dispersion. Waxes such as carnauba wax, esparto wax, wool wax, and spermaceti are used in cosmetic organogel preparations. A large quantity of water is entrapped between the three-dimensional lipid bilayers. The equilibrium water content in organogels is about 35%. The structural properties of the lipid, quantity of water in the system, solubility of drug incorporated, and external temperature influence the nature of the liquid crystalline phase. The bipolar nature of organogels allows incorporation of both hydrophilic and lipophilic drugs. Release rates can be controlled by altering the hydrophilic and hydrophobic components. Biodegradability of these waxes by the lipase enzyme in the body makes organogels suitable for parenteral administration.

The water present in the gel framework can be completely removed with some gelling agents. Gelatin sheets, acacia tears, and tragacanth ribbons are generally prepared by removal of water from their respective gel matrix. These dehydrated gel frameworks are called as xerogels.

### 3.4 Stimuli-Responsive Hydrogels

The three-dimensional networks of hydrophilic polymers absorb a large quantity of water and form soft structures which resemble biological tissues. Swelling properties of these hydrogels can be altered by various physicochemical parameters. Physical factors such as temperature, pH, and ionic strength of the swelling medium and chemical factors such as the structure of polymer (linear/branched) and chemical modifications (cross-linking) can be altered to tailor their swelling rate. This feature makes them very attractive for drug delivery and biomedical applications [19–23].

***pH-Responsive Hydrogels*** Some polymers show pH-dependent swelling and gelling characteristics in aqueous media. A polymer that exhibits such phase transition properties is very useful from the point of drug delivery. Methacrylic acids

(e.g., carbomers) that contain many carboxylic acid groups exist as solution at lower pH conditions. When the pH is increased, they undergo a sol-to-gel transition. This is because of the increase in the degree of ionization of acidic carboxylic groups at higher pH conditions, which in turn results in electrostatic repulsions between chains and, increased hydrophilicity and swelling. Conversely, polymers that contain amine-pendant groups swell at lower pH environment due to ionization and repulsion between polymer chains. The ionic strength of surrounding fluids significantly influences the equilibrium swelling of these pH-responsive polymers. Higher ionic strength favors gel–counter ionic interactions and reduces the osmotic swelling forces.

***Thermoresponsive Hydrogels*** A dispersion which exists as solution at room temperature and transforms into gel on instillation into a body cavity can improve the administration mode and help in modulating the drug release. Many polymers with thermoresponsive gelling properties are currently being synthesized and evaluated. A triblock copolymer that consists of polyethylene glycol–polylactic acid, glycolic acid–polyethylene glycerol (PEG–PLGA–PEG) is solution at room temperature and gels at body temperature. Poloxamers, which are made of triblock poly(ethylene oxide)–poly(propylene oxide)–poly(ethylene oxide), exhibit gelatin properties at body temperatures. Similarly, xyloglucan and xanthan gum aqueous dispersions are solutions at room temperature and become gel at body temperature. These are considered convenient alternatives for rectal suppository formulations which usually cause mucosal irritations due to their physical state. The physicochemical properties of these chemically modified thermoresponsive hydrogels are altered by changing the ratio of hydrophilic and hydrophobic segments, block length, and polydispersity. ReGel by MacroMed contains a triblock copolymer PLGA–PEG–PLGA, undergoes sol-to-gel transition on intratumoral injection, and releases paclitaxel for six weeks.

***Ionic-Responsive Hydrogels*** Administration of sodium alginate aqueous drops into the eye results in alginate gelation due to its interaction with calcium ions in the tear fluid. Alginate with high guluronic acid and deacetylated gellan gum (Gelrite) show sol-to-gel conversions in the eye due to their interaction with cations in the tear fluid. Timolol maleate sterile ophthalmic gel-forming solution (Timoptic-XE) that contains Gelrite gellan gum is commercially available.

### 3.5 Gelling Agents

A large number of gelling agents are commercially available for the preparation of pharmaceutical gels. In general, these materials are high-molecular-weight compounds obtained from either natural sources or synthetic pathways. They are water dispersible, possess swelling properties, and improve the viscosity of dispersions. An ideal gelling agent should not interact with other formulation components and should be free from microbial attack. Changes in the temperature and pH during preparation and preservation should not alter its rheological properties. In addition, it should be economic, readily available, form colorless gels, provide cooling sensation on the site of application, and possess a pleasant odor. Based on these factors, gelling agents are selected for different formulations. Table 8 summarizes the

**TABLE 8 Some Compendial Gelling Agents Used in Gels**

| Name | Molecular Weight | Gelling Strength (%) | Synonyms | Official Compendia |
|---|---|---|---|---|
| Bentonite | 359.16 | 10–20 | Magnabrite, mineral soap, Polargel, Veegum HS | BP, JP, PhEur, USPNF |
| Carbomer | $7 \times 10^5$–$4 \times 10^9$ | 0.5–2.0 | Acritamer, Carbopol, polyacrylic acid | BP, PhEur, USPNF |
| Carboxymethyl cellulose sodium | 90,000–700,000 | 3.0–6.0 | Akucell, Aquasorb, Sodium CMC, Tylose CB | BP, JP, PhEur, USPNF |
| Carrageenan | ≈1,000,000 | 0.3–2.0 | Gelcarin, Genu, Hygum Marine colloids | USPNF |
| Colloidal silicon dioxide | 60.08 | 2.0–10.0 | Aerosil, colloidal silica, fumed silica | BP, PhEur, USPNF |
| Gelatin | 15,000–25,0000 | 10.0–20.0 | Cryogel, Solugel | BP, JP, PhEur, USPNF |
| Glyceryl behenate | 1059.8 | 1.0–15.0 | Docosanoic acid, glycerol behenate | BP, PhEur, USPNF |
| Guar gum | ≈220,000 | 1.0–5.0 | Galactosal, Guar flour, Jaguar gum, Meyproguar | BP, PhEur, USPNF |
| Hydroxypropyl cellulose | 50,000–1,250,000 | 2.0–5.0 | Hyprolose, klucel, Methocel | BP, JP, PhEur, USPNF |
| Hydroxypropylmethyl cellulose | 10,000–1,500,000 | 1.0–10.0 | Hypromellose | BP, JP, PhEur, USPNF |
| Magnesium aluminum silicate | — | 5.0–15.0 | Veegum, aluminosilicic acid, Carrisorb, Magnabite | BP, PhEur, USPNF |
| Methylcellulose | 10,000–220,000 | 1.0–5.0 | Benecel, Methocel, Metolose | BP, JP, PhEur, USPNF |
| Poloxamer | 2090–17,400 | 15.0–20.0 | Lutrol, Monolan, Pluronic, Supronic | BP, phEur, USPNF |
| Polyvinyl alcohol | ~20,000–200,000 | 2.5–10.0 | Airvol, Elvanol, PVA, vinyl alcohol | USP |
| Povidone | 2500–3,000,000 | 2.0–20.0 | Kollidon, Plasdone, Polyvidone, PVP | BP, JP, PhEur, USPNF |
| Sodium alginate | 20,000–240,000 | 10.0–20.0 | Algin, alginic acid, sodium salt, Protanal | BP, PhEur, USPNF |
| Tragacanth | 840,000 | 1.0–8.0 | Gum Benjamin, Gum dragon, Trag, Tragant | BP, JP, PhEur, USPNF |

*Note*: BP, British Pharmacopoeia; JP, Japanese Pharmacopoeia; PhEur, European Pharmacopoeia; USPNF, U.S. Pharmacopeia/National Formulary

molecular weight, gelling strength, synonyms, and compendial status of some of these agents.

The following sections briefly describe the source, physicochemical properties, formulation, and preservation of some pharmacopeial gelling agents.

***Alginic Acid*** Alginic acid is tasteless and odorless and occurs as a yellowish white fibrous powder. The main source for this naturally occurring hydrophilic colloidal polysaccharide is different species of brown sea weed, known as Phaeophyceae. It consists of a mixture of D-mannuronic acid and L-glucuronic acids. It is used in gels due to its thickening and swelling properties. Alginic acid is insoluble in water; however, it absorbs 200–300 times its own weight of water and swells. The viscosity of alginic acid gels can be altered by changing the molecular weight and concentration. Addition of calcium salts increases the viscosity of alginic acid gels. Its viscosity decreases at higher temperature. Depolymerization due to microbial attack also results in viscosity reduction. Inclusion of an antimicrobial preservative avoids depolymerization and viscosity reduction during storage [6].

***Bentonite*** Bentonite is a naturally occurring colloidal hydrated aluminum silicate and contains traces of calcium, magnesium, and iron. It is odorless, available as fine crystalline powder, and is cream to grayish in color. The particles are negatively charged. Its high water uptake and swelling and thickening properties make it suitable for preparing gels. It swells to about 12-fold when it comes in contact with water. The viscosity of bentonite dispersion increases with increase in concentration. The gel-forming properties increase with addition of alkaline materials such as magnesium oxide and decrease with addition of alcohol or electrolytes. Use of hot water and stirring improve wetting and dispersion of bentonite particles in the preparation of the gel. Mixing with magnesium oxide or zinc oxide prior to addition helps in dispersion of bentonite in water. Prior trituration of bentonite with glycerin also helps in easy dispersibility in water. These dispersions are generally left for about 24 h to complete the swelling process. At lower concentration (10%) bentonite suspension exhibits the properties of shear thinning systems and at high concentrations (about 50–60%) it forms gel with finite yield strength [24].

***Carbomer*** Carbomers are one of the widely used gelling agents in topical preparations due to their extensive swelling properties. They are obtained by cross-linking acrylic acid with allyl sucrose or allyl pentaerythritol. Various grades with varying degree of cross-linking and molecular weight are commercially available. Carbomers are generally available as hygroscopic powders, are white in color, and possess a characteristic odor. Presence of about 60% carboxylic acid in its composition makes them acidic. Carbomer 934P, 971P, 974P, and so on, are used for preparing clear gels. Aqueous dispersions of carbomers are usually low viscous, and on neutralization they form high-viscous gels. Basic materials such as sodium hydroxide, potassium hydroxide, sodium bicarbonate, and borax are being used for neutralizing carbomer dispersions. About 0.4 g of sodium hydroxide is used to neutralize 1 g of carbomer dispersion. The viscosity of gels depends on the molecular weight of carbomer and its degree of cross-linking. Inclusion of antioxidants, protection from light, and preservation at room temperature help in retaining their viscosity for prolonged periods. Microbial stability of carbopol gels can be improved by

adding antimicrobial preservatives. These gels are prone to discoloration in the presence of large amounts of electrolytes, strong acids, and cationic polymers. Glass, plastic, and resin-lined containers which possess good corrosion-resistant properties are used for packing carbomer gels [6].

***Carboxymethylcellulose Calcium (Calcium CMC)*** A calcium salt of polycarboxymethyl ether of cellulose, calcium CMC is obtained by carboxymethylation of cellulose and conversion into calcium salt. Different molecular grades are prepared by changing the degree of carboxymethylation. It is available as a fine powder, white to yellowish white in color, and hygroscopic in nature. Calcium CMC has swelling and viscosity-enhancing properties in water. It can swell twice its volume in water [25].

***Carboxymethylcellulose Sodium (Sodium CMC)*** A sodium salt of polycarboxymethyl ether of cellulose, sodium CMC is obtained by treating alkaline cellulose with sodium monochloroacetate. It is available as white-colored granular powder. Various viscosity grades of sodium CMC commercially available basically differ in their degree of substitution. The degree of substitution represents the average number of hydroxyl groups that are substituted per anhydroglucose unit. It is readily dispersible in water and forms clear gels. The aqueous solubility of CMC sodium is governed by the degree of substitution. Higher concentrations generally yield thicker gels. Although the viscosity of gels is stable over a wide range of pH (4–10), a fall in pH below 2 or a rise to above 10 results in physical instability and viscosity reduction. Higher viscosity is obtained at neutral pH conditions. Exposure of gels to higher temperature also results in viscosity reduction. Preservation at optimum temperature and inclusion of an antimicrobial preservative improve the physical and microbial stability of CMC sodium gels [25].

***Carrageenan*** Extraction of some red seaweed belonging to the Rhodophyceae class with water or aqueous alkali yields carrageenan. It is a hydrocolloid and mainly contains sodium, potassium, calcium, magnesium, and ammonium sulfate esters of galactose and copolymers of 3, 6-anhydrogalactose. They differ in their ester sulfate and anhydrogalactose content. It is available as a coarse to fine powder which is yellow to brown in color. It is odorless and tasteless. Carrageenan is soluble in hot water and forms gels at 0.3–2.0%. ι-Carrageenan and κ-carrageenan show good gelling properties [26].

***Colloidal Silicon Dioxide*** Colloidal silicon dioxide is a fumed silica obtained by vapor hydrolysis of chlorosilanes. It is available as nongritty amorphous powder which is bluish white in color. It is tasteless and odorless and possesses low tapped density. Although insoluble in water, it readily forms a colloidal dispersion due to its fine particle size, higher surface area, and water-adsorbing properties. The bulk density and particle size of colloidal silicon dioxide can be altered by changing the method of manufacture. Transparent gels can be formed by mixing with other materials that possess similar refractive index. Under acidic and neutral pH conditions, it shows viscosity-increasing properties. This property is lost at higher pH conditions because of its dissolution. Viscosity of gels is not generally affected by temperature [27].

***Ethylcellulose*** Ethylcellulose is a synthetic polymer made of β-anhydroglucose units connected by acetyl linkages. It is obtained by ethylating alkaline cellulose solution with chloroethane. Ethylcellulose is available as a free-flowing powder which is tasteless and white in color. Although it is insoluble in water, it is incorporated into topical preparations due to its viscosity-enhancing properties. Ethanol or a mixture of ethanol and toluene (2:8) is used as a solvent. A decrease in the ratio of alcohol increases the viscosity. The viscosity of the dispersion is increased by increasing the concentration of ethylcellulose or by using a high-molecular-weight material. As ethylcellulose is prone to photo-oxidation at higher temperature, and gels are prepared and preserved at room temperature and dispensed in airtight containers [28].

***Gelatin*** Gelatin is a protein obtained by acid or alkali hydrolysis of animal tissues that contain large amounts of collagen. Based on the method of manufacture, it is named type A or type B gelatin. Type A is obtained by partial acid hydrolysis and type B is obtained by partial alkaline hydrolysis. They differ in their pH, density, and isoelectric point. Gelatin is available as yellow-colored powder or granules. It swells in water and improves the viscosity of dispersions. Different molecular weights and particle size grades are commercially available. Gels can be prepared by dissolving gelatin in hot water and cooling to 35 °C. Temperature greatly influences the viscosity and stability of gelatin dispersions. It transforms to a gel at temperatures above 40 °C and undergoes depolymerization above 50 °C. The viscosity of gelatin gel is also affected by microbes [29].

***Guar Gum*** Guar gum is a high-molecular-weight polysaccharide obtained from the endosperms of guar plant. It mainly contains D-galactan and D-mannan. It is available as powder which is odorless and white to yellowish white in color. It readily disperses in water and forms viscous gels. The viscosity of gel is influenced by the particle size of material, pH of the dispersion, rate of agitation, swelling time, and temperature. Viscosity reduces on long-time heating. Maximum viscosity can be achieved within 2–4 h. Gels are stable at pH between 7 and 9 and show liquification below pH 7. Addition of antimicrobial preservatives improves the microbial stability of guar gum gels. Rheological properties of these gels can be modified by incorporating other plant hydrocolloids such as tragacanth and xanthan gum [30].

***Hydroxyethyl Cellulose (HEC)*** HEC is a partially substituted poly(hydroxyethyl) ether of cellulose. It is obtained by treating alkali cellulose with ethylene oxide. HEC is available as a powder and appears light tan to white in color. Different viscosity grades of HEC are commercially available which differ in their molecular weights. Clear gels are prepared by dissolving HEC in hot or cold water. Dispersions can be prepared quickly by altering the stirring rate of dispersion, temperature, and pH. Slow stirring at room temperature during the initial stages favors wetting. Increasing the temperature at this stage increases the rate of dispersion. Although HEC dispersions are stable over a wide pH range, maintaining basic pH improves the dispersion. The preservation temperature, formulation pH, and microbial attack influence the rheological properties of HEC dispersions. Viscosity reduces at higher temperature, but reverts to the original value on returning to room temperature. Lower and higher pH of the vehicle usually results in hydrolysis or oxidation of HEC,

respectively. Some of the enzymes secreted by microbes decrease the viscosity of HEC dispersions. The presence of higher levels of electrolytes may also destabilize HEC dispersions. Inclusion of a suitable antimicrobial preservative is essential to retain the viscosity of HEC gels [31].

***Hydroxyethylmethyl Cellulose (HEMC)*** HEMC is a partially *o*-methylated and *o*-(2-hydroxyethylated) cellulose. It is available as powder or granules which are white, grayish white, or yellowish white in color. Various viscosity grades of HEMC are commercially available, and form viscous colloidal dispersions or gels in cold water which has a pH in the range of 5.5–8 [6].

***Hydroxypropyl Cellulose (HPC)*** HPC is a partially substituted poly(hydroxypropyl) ether of cellulose. It is obtained by treating alkali cellulose with propylene oxide at higher temperatures. It is available as tasteless and odorless powder which is yellowish or white in color. Different viscosity grades are commercially available. Gradual addition of HPC powder into vigorously stirred water yields clear viscous dispersions or gels below 38 °C. Increase in temperature destabilizes the dispersion and leads to precipitation. The viscosity of dispersions can be increased by increasing the concentration of HPC or by using high-molecular-weight grades. Inclusion of a cosolvent such as dichloromethane or methane produces viscous dispersion or gels with modified texture. The viscosity of HPC dispersions can be increased by mixing with an anionic polymer. High concentrations of electrolytes destabilize HPC dispersions. HPC dispersions are neutral in pH (6–8). They undergo acid hydrolysis at lower pH and oxidation at higher pH. Both processes decrease the dispersion viscosity. In addition, certain enzymes produced by microbes degrade HPC and reduce its viscosity. Addition of an antimicrobial preservative is therefore recommended for HPC gels. Preservation of these gels from light can further improve its physical stability [25].

***Hydroxypropylmethyl Cellulose (HPMC)*** HPMC is a partly *o*-methylated and *o*-(2-hydroxypropylated) cellulose obtained by treating alkali cellulose with chloromethane and propylene oxide. It is available as odorless and tasteless granular or fibrous powder which is creamy white or white in color. HPMC is soluble in cold water. Aqueous dispersions are prepared by dispersing material in about 25% hot water (80 °C) under vigorous stirring. On complete hydration of HPMC, a sufficient quantity of cold water is added and mixed. The gel point of HPMC dispersions varies from 50 to 90 °C. Gels are stable over a wide pH range (3–11). The viscosity HPMC dispersions depends on the concentration of material used, its molecular weight, vehicle composition, presence of preservatives, and so on. Viscous gels can be prepared using high concentrations of high-molecular-weight grades. Inclusion of organic solvents such as ethanol or dichloromethane improves the viscosity. Addition of an antimicrobial preservative (e.g., benzalkonium chloride) minimizes microbial spoilage of HPMC gels [25].

***Glyceryl Behenate*** Glyceryl behenate is a mixture of glycerides of fatty acids which is obtained by esterification of glycerin with behenic acid. It may also contain arachidic acid, stearic acid, erucic acid, lignoceric acid, and palmitic acid. It is available as a waxy mass or powder, possesses a faint odor, and is white in color. It is

practically insoluble in water and soluble in dichloromethane and chloroform. It is used as a viscosity-increasing agent in silicon gels [6].

***Glyceryl Monooleate (GMO)*** GMO is a mixture of glycerides of fatty acids obtained by esterification of glycerol with oleic acid. It may also contain linoleic acid, palmitic acid, stearic acid, linolenic acid, arachidic acid, and eicosenoic acid. It is available as a partially solidified or oily liquid. GMO is insoluble in water. Self-emulsifying grades that contain an anionic surfactant disperse easily and swell in water. The nonemulsifying grades are used as emollients in topical preparations and self-emulsifying grades are used as emulsifiers in aqueous emulsions [6].

***Magnesium Aluminum Silicate (MAS)*** MAS is also known as veegum. It is a polymeric complex of magnesium, aluminum, silicon, oxygen, and water and is obtained from silicate ores. Based on the ratio of aluminum and magnesium and viscosity, it is classified as types IA, IB, IC, and IIA. It is available as fine powder that is odorless, tasteless, and off-white to creamy white in color. Although MAS is insoluble in water, it swells to a large extent and produces viscous colloidal dispersions. Use of higher concentration, addition of electrolytes, and heating of dispersion usually improve the viscosity [32].

***Methylcellulose (MC)*** MC is a long-chain cellulose polymer with methoxyl substitutions at positions 2, 3, and 6 of the anhydroglucose ring. It is synthesized by methylating alkali cellulose with methyl chloride. The degree of substitution of methoxy groups influences the molecular weight, viscosity, and solubility characteristics of MC. It is available as powder or granules and is odorless, tasteless, and white to yellowish white in color. MC is insoluble in hot water but slowly swells and forms viscous colloidal dispersions in cold water. Gels can be prepared by initially mixing the methylcellulose with half the volume of hot water ($\approx 70\,^{\circ}C$) followed by addition of the remaining volume of cold water. Viscosity of these dispersions can be increased by using high-concentration or high-molecular-weight grades of methylcellulose. Higher processing or preservation temperatures reduce the viscosity of formulations, which regain their original state on cooling to room temperature. MC aqueous dispersions show pH values of 5–8. Reduction in pH to less than 3 leads to acid-catalyzed hydrolysis of glucose–glucose linkages and results in low viscosity. Antimicrobial preservatives are generally included to enhance the microbial stability of dispersions. Salting out is observed when high concentrations of electrolytes are added to methylcellulose dispersions. The viscosity of methylcellulose dispersions is also influenced by the presence of formulation excipients and drugs [25].

***Poloxamer*** Poloxamers are copolymers of ethylene oxide and propylene oxide. Different molecular weight grades that are different in physical form, solubility, and melting point are available. Poloxamer 124 is a colorless liquid, whereas poloxamers 188, 237, 338, and 407 are solids at room temperature. All poloxamer grades are freely soluble in water and form gels at higher concentrations. The pH of aqueous liquids ranges between 5 and 7.5 [33].

***Polyethylene Oxide*** Polyethylene oxide is a nonionic homopolymer of ethylene oxide synthesized by polymerization of ethylene oxide. It is available as a free-

flowing powder white to off-white in color with a slight ammonia odor. Various molecular weight grades of polyethylene oxide are commercially available. They swell in water and form viscous dispersions or gels based on the concentration and grade used. Inclusion of alcohol improves the rheological stability of polyethylene oxide dispersions [6].

***Polyvinyl Alcohol (PVA)*** PVA is a synthetic polymer prepared by hydrolysis of polyvinyl acetate. It is available as a granular powder which is odorless and white in color. Mixing with water at room temperature, heating for about 5 min at 90 °C, followed by cooling with constant mixing yield aqueous dispersions or gels. Higher viscosities can be obtained by using high-viscosity grades. Addition of borax improves the gelling properties of PVA, whereas inorganic salts destabilize these dispersions. The pH of PVA dispersions ranges between 5 and 8. Physical and chemical decompositions occur at lower and higher pH conditions. Incorporation of an antimicrobial preservative and storage at room temperature improve its stability [6].

***Povidone*** Povidone is a synthetic polymer consisting of 1-vinyl-2-pyrrolidinone units. It is available as a fine powder and appears white to creamy-white in color. Various molecular weight grades of povidone are available which differ in their degree of polymerization. Povidone is soluble in water and forms viscous solutions and gels based on the concentration and viscosity grade used. Decomposition occurs when dispersions are heated to about 150 °C. The pH of aqueous dispersions ranges from 3 to 7. The microbial stability of povidone aqueous dispersions can be increased by adding preservatives [6].

***Propylene Carbonate (PC)*** PC is prepared by reacting propylene chlorohydrin with sodium bicarbonate. It is available as a clear liquid with a faint odor. Mixtures of PC and propylene glycol are good solvents for corticosteroids in topical preparations. It is incompatible with strong acids, bases, and amines. The pH of 10% aqueous dispersion is 6.0–7.5 [34].

***Propylene Glycol Alginate (PGA)*** PGA is a propylene glycol ester of alginic acid obtained by treating alginic acid with propylene oxide. It is available as granular or fibrous powder which is odorless, tasteless, and white to yellowish-white in color. PGA is soluble in water and forms viscous colloidal dispersions. The viscosity of these dispersions is based on the concentration of PGA, temperature, and pH. Its aqueous solubility decreases at higher temperatures. The aqueous dispersions are acidic in nature and more stable at pH 3–6. Higher pH leads to saponification. As these dispersions are prone to microbial spoilage, antimicrobial preservatives are generally included [6].

***Sodium Alginate*** Sodium alginate is obtained by extraction of alginic acid from brown seaweed followed by neutralization with sodium bicarbonate. Alginic acid is composed of D-mannuronic acid and L-guluronic acid. It is available as a powder which is tasteless, odorless, and white to yellowish-brown in color. Sodium alginate forms viscous gels in water. Dispersing agents such as glycerol, propylene glycol, sucrose, and alcohol are added to improve dispersion. The presence of low concentration of electrolytes improves the viscosity, whereas at high concentrations salting

out takes place. The viscosity of gel is based on the concentration of sodium alginate, temperature, pH, and other additives. Various viscosity grades of sodium alginate are commercially available. Aqueous dispersions are stable at pH 4–10. Precipitation or decrease in viscosity is observed when the pH is below or above these values. Autoclaving or heating above 70 °C results in depolymerization and decrease in viscosity. Inclusion of preservative is essential to maintain the microbial stability of sodium alginate topical gels [35].

***Tragacanth*** Tragacanth is a polysaccharide polymer obtained from some *Astragalus* species. It is composed of two polysaccharides: bassorin (water insoluble) and tragacanthin (water soluble). It is available as odorless powder white to yellowish in color and possesses mucilaginous taste. Tragacanth swells about 10 times its weight in water and forms viscous solutions or gels. Tragacanth is usually added with vigorous stirring to avoid lump formation. Wetting agents such as glycerin, propylene glycol, and 95% ethanol are used in initial stages to improve wetting and dispersion of tragacanth in water. The viscosity of tragacanth dispersions is influenced by the processing temperature and formulation pH. High temperature usually increases the viscosity of gels. Tragacanth dispersions show higher viscosity at pH 8 and starts decreasing at higher pH. These gels usually contain preservatives such as benzoic acid or a combination of methyl and propyl parabens for effective preservation from microbial attack. The viscosity of tragacanth dispersions reduces in the presence of strong mineral and organic acids and sodium chloride [6].

### 3.6 Preparation and Packaging

Gels are relatively easier to prepare compare to ointments and creams. In addition to the gelling agent, medicated gels contain drug, antimicrobial preservatives, stabilizers, dispersing agents, and permeation enhancers. Some of the factors discussed below are essential to obtain a uniform gel preparation.

***Order of Mixing*** The order of mixing of these ingredients with the gelling agent is based on their influence on the gelling process. If they are likely to influence the rate and extent of swelling of the gelling agent, they are mixed after the formation of gel. In the absence of such interference, the drug and other additives are mixed prior to the swelling process. In this case, the effects of mixing temperature, swelling duration, and other processing conditions on the physicochemical stability of the drug and additives are also considered. Ideally the drug and other additives are dissolved in the swelling solvent, and the swelling agent is added to this solution and allowed to swell.

***Gelling Medium*** Purified water is the most widely used dispersion medium in the preparation of gels. Under certain circumstances, gels may also contain cosolvents or dispersing agents. A mixture of ethanol and toluene improves the dispersion of ethylcellulose, dichloromethane and methanol increase the viscosity of hydroxypropyl cellulose dispersions, alcohol improves the rheological stability of polyethylene oxide gels, and inclusion of glycerin, propylene glycol, sucrose, and alcohol improves the dispersion of sodium alginate dispersions. Borax is included in polyvinyl alcohol gels and magnesium oxide, zinc oxide, and glycerin are included in bentonite gel as

dispersing agents. Care should be taken to avoid the evaporation or degradation of these cosolvents and dispersing agents during the preparation of gels.

***Processing Conditions and Duration of Swelling*** The processing temperature, pH of the dispersion, and duration of swelling are critical parameters in the preparation of gels. These conditions vary with each gelling agent. For instance, hot water is preferred for gelatin and polyvinyl alcohol, and cold water is preferred for methylcellulose dispersions. Carbomers, guar gum, hydroxypropyl cellulose, poloxamer, and tragacanth form gels at weakly acidic or near-neutral pH conditions (pH 5–8). Gelling agents such as carboxymethyl cellulose sodium, hydroxypropylmethyl cellulose, and sodium alginate form gels over a wide pH range (4–10). Hydroxyethyl cellulose forms gel at alkaline pH condition. A swelling duration of about 24–48 h generally helps in obtaining homogeneous gels. Natural gums need about 24 h and cellulose polymers require about 48 h for complete hydration.

***Removal of Entrapped Air*** Entrapment of air bubbles in the gel matrix is a common issue, especially when the swelling process involves a mixing procedure or the drug and other additives are added after the swelling process. Positioning the propeller at the bottom of the mixing container minimizes this issue to a larger extent. Further removal of air bubbles can be achieved by long-term standing, low-temperature storage, sonication, or inclusion of silicon antifoaming agents. In large-scale production, vacuum vessel deaerators are used to remove the entrapped air.

***Packaging*** Being viscous and non-Newtonian systems, gels need high attention during packing into containers. Usually they are packed into squeeze tubes or jars made of plastic materials. Aluminum containers are also used when the product pH is slightly acidic. Pump dispensers and prefilled syringes are sometimes used for packing gels. As most of the gels contain an aqueous phase, preservation in airtight containers helps in protecting them from microbial attack. Usually they are preserved at room temperature and protected from direct sunlight and moisture.

In large-scale production, different mills, separators, mixers, deaerators, shifters, and packaging machines are used. Most of this equipment is similar to those discussed under ointments and creams. Figure 3 shows a "one-bowl" vacuum processing machine manufactured by FrymaKoruma-Rheinfelden (Romaco) for the preparation gels. Batch sizes ranging from 15 to 160 L are processed using this machine. It uses an extremely efficient high-shear rotor/stator system for homogenizing and a counterrotating mixing system for macromixing. The raw materials are drawn into the multichamber system of the homogenizer by vacuum and then mixed and pumped into the homogenizing zone. The product which enters the vessel is mixed, sheared, and recirculated. All the entrapped air is removed during the recirculation. The machine also has an insulated jacket for controlling the processing temperature.

### 3.7 Evaluation

Various pharmacopeial and nonpharmacopeial tests are carried out to evaluate the physicochemical, microbial, in vitro, and in vivo characteristics of gels. These tests are meant for assessing the quality of gel formulations and minimizing the batch-

**FIGURE 3** Vacuum processing machine used for preparation of gels. (Courtesy of FrymaKoruma-Rheinfelden, Switzerland.)

to-batch variations. Some of the tests recommended by the USP for gels include minimum fill, pH, viscosity, microbial screening, and assay. In some cases sterility and alcohol content are also specified. The USP also recommends storage for each compendial gel formulation. Table 9 shows the quality control tests and storage requirements that are specified for some pharmacopeial gels. The procedures for minimum fill, microbial screening, sterility, assay, in vitro drug release, and in vivo bioequivalence are similar to those of ointments and creams. The procedures for additional tests such as homogeneity, surface morphology, pH, alcohol content, rheological properties, bioadhesion, stability, and ex vivo penetration are described below.

***Homogeneity and Surface Morphology*** The homogeneity of gel formulations is usually assessed by visual inspection and the surface morphology by using scanning electron microscopy. Generally, the swollen gel is allowed to freeze in liquid nitrogen and then lyophilized by freeze drying. It is assumed that the morphologies of the swollen samples are retained during this process. The lyophilized hydrogel is gold sputter coated and viewed under an electron microscope.

***pH*** Many gelling agents show pH-dependent gelling behavior. They show highest viscosity at their gel point. Determination of pH is therefore important to maintain consistent quality. As conventional pH measurements are difficult and often give erratic results, special pH electrodes are used for viscous gels. Flat-surface digital

**TABLE 9 USP Specifications for Some Official Gels**

| Drug | Quality Control Tests | Packaging and Storage Requirements |
|---|---|---|
| Aminobenzoic acid | Minimum fill, pH (4.0–6.0), alcohol content, and assay | Tight, light-resistant containers |
| Benzocaine | Microbial limits, minimum fill, and assay | Well-closed containers |
| Benzoyl peroxide | pH (2.8–6.6), and assay | Tight containers |
| Betamethasone benzoate | Microbial limits, minimum fill, and assay | Tight containers; store at 25 °C, excursions permitted between 15 and 30 °C; protect from freezing |
| Clindamycin phosphate | Minimum fill, pH (4.5 and 6.5), and assay | Tight containers |
| Desoximetasone | Minimum fill, alcohol content, and assay | Collapsible tubes at controlled room temperature |
| Dexamethasone | Minimum fill and assay | Collapsible tubes; keep tightly closed; avoid exposure to temperatures exceeding 30 °C |
| Dyclonine hydrochloride | pH (2.0 and 4.0), and assay | Collapsible, opaque plastic tubes or in tight, light-resistant glass containers |
| Erythromycin | Minimum fill and assay | Tight containers |
| Fluocinonide | Minimum fill and assay | Collapsible tubes or tight containers |
| Hydrocortisone | Minimum fill and assay | Tight containers |
| Lidocaine hydrochloride | Sterility, minimum fill, pH (7.0–7.4), and assay | Tight containers |
| Metronidazole | Minimum fill, pH (4.0 and 6.5), and assay | Laminated collapsible tubes at controlled room temperature |
| Naftifine hydrochloride | Microbial limits, minimum fill, pH (5.5–7.5), content of alcohol, and assay | Tight containers |
| Salicylic acid | Alcohol content and assay | Collapsible tubes or tight containers, preferably at controlled room temperature |
| Sodium sulfide | pH (11.5–13.5) and assay | Tight containers at controlled room temperature or in a cool place |
| Stannous fluoride | Viscosity, pH (2.8–4.0), stannous ion content, total tin content, and assay | Well-closed containers |
| Tolnaftate | Minimum fill and assay | Tight containers |

pH electrodes from Crison, combination electrodes that contain a built-in temperature probe, a bridge electrolyte chamber and movable sleeve junction from Mettler, and combination pH puncture electrodes with spear-shaped tip from Mettler are some commercially available pH measurement systems for semisolid formulations.

***Alcohol Content*** Alcohol levels in some gel preparations are determined by gas chromatographic (GC) methods. Desoxymetasone gel USP and naftifine hydrochloride gel USP contain 18–24% and 40–45% (w/w) of ethyl alcohol, respectively. In a desoxymetasone gel, the sample is dissolved in methanol and injected into a gas

chramatograph for quantitative analysis. Isopropyl alcohol is used as an internal standard. In naftifine hydrochloride gel, *n*-propyl alcohol is used as an internal standard and water is used as the diluting solvent [15].

***Rheological Studies*** Viscosity measurement is often the quickest, most accurate, reliable method to charactreize gels. It gives an idea about the ease with which gels can be processed, handled, or used. Some of the commonly used tests for characterizing rheology of gels are yield stress, critical strain, and creep. *Yield stress* refers to the stress that must be exceeded to induce flow. This helps in characterizing the flow nature of non-Newtonian systems. *Critical strain* or *gel strength* refers to the minimum energy needed to disrupt the gel structure. When samples are subjected to increasing stress, viscosity is maintained as long as the gel structure is maintained. When the gel's intermolecular forces are overcome by the oscillation stress, the sample breaks down and the viscosity falls. The higher the critical strain, the better the physical integrity of gel systems. *Creep* or *recovery* helps in assessing the strength of bonds in a gel structure. This is assessed by determining relaxation times, zero-shear viscosity, and viscoelastic properties.

Based on the nature of the test material, different techniques are employed to measure the rheological parametrs of gels. Very sophisticated automatic equipment is commercially available for measurements. Cup-and-bob viscometers and cone-and-plate viscometers are widely used for viscous liquids and gels. They measure the frictional force that is created when gels start flowing. These viscometers are usually calibrated with certified viscosity standards before each measurement. General-purpose silicone fluids which are less sensitive to temperature change are used as standards. The viscosity of gels is affected by the experimental temperature and shear rate and the gels exhibit liquid- or solidlike properties. Hence the viscosity of these non-Newtonian systems are recorded at several shear rates under controlled temperatures. The USP specifies the operating conditions for each gel formulation. Commercially available viscometers include Brookfield rotational viscometers, Haake rheometers, Schott viscoeasy rotational viscometers, Malvern viscometers, and Ferranti-Shirley cone-and-plate viscometers.

***Bioadhesion*** This test is performed to assess the force of adhesion of a gel with biological membranes. The bioadhesive property is preferred for ophthalmic, nasal buccal, and gastroretentive gel formulations. Drugs applied as solutions, viscous solutions, and suspensions drain out from these biological locations within a short time and only a limited fraction of drug elicits the pharmacological activity. Products with higher bioadhesion thus help in increasing the contact time between drugs and absorbing surface and improve their availability. The bioadhesive properties of gels are measured using various custom-designed equipment. All the equipment, however, measures the force required to detach the gel from a biological surface under controlled experimental conditions (e.g., temperature, wetting level, contact time, contact force, surface area of tissue). A typical bioadhesion measurement system consists of a moving platform and a static platform. A tissue from a particular biological region is fixed onto these platforms and a known quantity of the test product is uniformly applied to the tissue surface of the lower static platform. The upper moving platform is allowed to contact with the product surface with a known contact force. After allowing for a short contact time, the moving platform is

separated from the product with a constant rate. The force required to detach the mucosal surface from the product is recorded. The analog signals generated by precision load cells are then converted to digital signals through data acquisition systems and processed using specific software programs.

***Stability Studies*** Being dispersed systems containing water in their matrix, gels are prone to physical, chemical, and microbial stability issues. Syneresis is a commonly observed physical stability problem with gels. It involves squeezing out dispersion medium due to elastic contraction of polymeric gelling agents. This results in shrinkage of gels. Syneresis can be determined by heating the gels to a higher temperature followed by rapid cooling using an ice water bath at room temperature. The sample is preserved at 4 °C for about a week, and water loss from the gel matrix is measured. Water loss is measured by weighing the mass of the gel matrix after centrifugation. Absence of syneresis indicates higher physical stability of gels. The chemical stability of drugs in the gel matrix is determined using stability-indicating analytical methods. Studies are conducted at accelerated temperature, moisture, and light conditions to determine the possible degradation of drug in the gel.

***Ex Vivo Penetration*** Ex vivo studies are carried out to examine the permeation of drug from gels through the skin or any other biological membrane. As with in vitro release studies, ex vivo penetration is conducted using vertical diffusion cells or modified cells with flow-through design. In this case, the receiver side is filled with phosphate buffer solution of pH 7.4 to simulate the biological pH of human blood. Skin samples from different animal sources such as rats, rabbits, pigs, and human cadavers are used for screening dermatological products. The stratum corneum layer of the skin is separated from the dermis before mounting onto the diffusion cells. The epidermis is separated by immersing the skin sample in normal saline or purified water which is maintained at 60 °C for 2 min followed by immersion into cold water for 30 s. Careful peeling helps in the separation of the epidermis layer from the dermis. This layer is mounted between the donor and receiver sides and studies are conducted after application of test gel over the surface of the stratum corneum in the donor side. Samples are withdrawn at different time intervals and analyzed for drug permeation by suitable analytical techniques.

### 3.8 Typical Pharmacopeial and Commercial Examples

Gels are becoming popular dosage forms for delivering various categories of drugs for treating dermatological, oral, ophthalmic, vaginal, and other conditions. Many dermatological gels are used for treating mild to moderate acne, eczema, dermatitis, allergies, rash, and psoriasis and for removal of common warts. Oral gels are available for relieving painful mouth sores, treating tooth decay, preventing tooth plaque, and relieving inflammation of the gums, and vaginal gels are available for treating certain type of vaginal infections (e.g., bacterial vaginosis). Some special types of gels are available for preventing or controlling pain during certain medical procedures, numbing and treating urinary tract inflammation (urethritis), and numbing mucous membranes. Table 10 shows some of the commercially available compendial gels.

**TABLE 10 Examples of Compendial/Commercial Gels**

| Drug[a] | Category | Indication | Commercial Products | Strength (%) |
|---|---|---|---|---|
| Benzocaine | Local anesthetic | In mouth to relieve pain or irritation caused by many conditions | Oratect Gel, Num Zit Gel | 7.5, 10 |
| Benzoyl peroxide | Keratolytic | Mild to moderate acne | Persa-Gel, 5 Benzagel 10 | 5.0, 10 |
| Betamethasone | Anti-inflammatory | Eczema, dermatitis, allergies, and rash | Diprolene | 0.05 |
| Clindamycin phosphate (vag) | Antibiotic | Certain types of vaginal infection (e.g., bacterial vaginosis) | Cleocin T | 1.0 |
| Desoximetasone | Anti-inflammatory | Eczema, dermatitis, allergies, and rash. | Topicort | 0.05 |
| Dyclonine hydrochloride | Antipruritic and local anesthetic | Relieve painful mouth sores | Dyclone | 0.5, 1.0 |
| Erythromycin | Antibiotic | Acne and skin infection due to bacteria | Erygel | 2.0 |
| Fluocinonide | Anti-inflammatory | Psoriasis, eczema, dermatitis, allergies, and rash | Lidex | 0.05 |
| Lidocaine hydrochloride | Local anesthetic | Prevent and control pain during certain medical procedures, numb and treat urinary tract inflammation (urethritis), and numb mucous membranes | Xylocaine, Anestacon | 2.0 |
| Metronidazole (vag) | Antifungal | Certain types of bacterial infections in the vagina | Metrogel | 0.75, 1.0 |
| Naftifine hydrochloride | Antifungal | Fungal infections of skin such as jock itch, athlete's feet, or ringworm | Naftin | 1.0 |
| Salicylic acid | Keratolytic | Removal of common warts | Sal-Plant Gel | 17.0 |
| Stannous fluoride | Fluoride | Treat tooth decay, prevent tooth plaque and inflammation of gums | Flo-Gel, Gel-Kam | 0.4 |
| Tolnaftate | Antifungal | Skin infections such as athlete's foot, jock itch, and ringworm | Tolnaftate | 1.0 |

[a]vag: vaginal gel.

## 4 REGULATORY REQUIREMENTS FOR SEMISOLIDS

Regulatory agencies such as the Center for Drug Evaluation and Research (CDER) have at the Food and Drug Administration (FDA) have set forth guidelines for various pharmaceutical activities to ensure the identity, strength, quality, safety, and efficacy of semisolid drug products. A manufacturer of semisolid formulations needs to fulfill these requirements at the time of filing for investigational new drug (IND), abbreviated new drug application (ANDA), or abbreviated antibiotic drug application (AADA). Standard chemistry, manufacturing, and control (CMC) tests are necessitated for all dermatological drug products. Additional information on polymorphic form, particle size distribution, and other characteristics is needed for submitting an NDA. When an ANDA for a semisolid product is filed, the manufacturer should meet the standards of compendial requirements if available and match the important in vitro and in vivo characteristics of the reference listed drug (RLD). If such information is not available, appropriate in vitro release methods are submitted to ensure batch-to-batch consistency. Even at later stages, if changes are made for an approved semisolid product with respect to its components, composition, equipment, process, batch size, and manufacturing site, the formulator should submit necessary details to the regulatory agency.

A typical guideline that defines the types and levels of scale-up and postapproval changes (SUPAC) is outlined in Table 11. Based on the type and level of change, the manufacturer needs to submit application and compendial product release requirements, executed batch records, accelerated and long-term stability data, identification and assay for new preservative, preservative effectiveness test at lowest specified level, validation methods to support absence of interference of preservative with other tests, in vitro release test, and in vivo bioequivalence data to the FDA. When changes are made with respect to the quality and quantity of excipients or crystallinity of drug, especially if the drug is in suspension, in vivo bioequivalence studies are recommended. As routine pharmacokinetic studies do not produce measurable quantities of drug in blood, plasma, urine, and other extracutaneous biological fluids, dermatopharmacokinetic (DPK) studies and pharmacodynamic or comparative clinical trials are recommended to establish bioequivalence of topical products. Table 12 shows specific requirements for various SUPAC levels. If bioavailability or bioequivalence data of a highest strength product are already available,

**TABLE 11 Description on SUPAC Guidelines for Nonsterile Semisolid Dosage Forms**

| Type of Change | Level | Description |
|---|---|---|
| Change in components and composition | 1 | Partial deletion or deletion of color, fragrance, or flavor; up to 5% excipient change in approved amount; change in supplier for structure forming or technical-grade excipient |
| | 2 | Excipient changes from 5 to 10%; change in supplier for structure forming excipient, which is not covered under level 1; change in technical grade of structure forming excipient; change in particle size of drug if drug is in suspension |
| | 3 | Qualitative and quantitative changes in excipient not covered under levels 1 and 2; any change in crystallinity of drug if drug is in suspension |

**TABLE 11** ***Continued***

| Type of Change | Level | Description |
|---|---|---|
| Change in preservative components and composition | 1 | Less than 10% quantitative change in preservative |
| | 2 | 10–20% quantitative change in preservative |
| | 3 | Deletion or more than 20% quantitative change in preservative; inclusion of a different preservative |
| Change in manufacturing equipment | 1 | Change to automated or mechanical equipment for transfer of ingredients; use of alternative equipment of same design and operating principles |
| | 2 | Use of alternative equipment of different design and operating principles; change in type of mixing equipment |
| Change in manufacturing process | 1 | Changes in process within approved application ranges; addition of formulation additives |
| | 2 | Changes in process outside approved application ranges; process of combining phases |
| Change in batch size | 1 | Batch size changes upto 10 times of pivotal clinical trial or biobatch |
| | 2 | Batch size changes above 10 times of pivotal clinical trial or biobatch |
| Change in manufacturing site | 1 | Changes within existing facility |
| | 2 | Changes within same campus or facilities in adjacent city blocks |
| | 3 | Change to different campus; change to a contract manufacturer |

**TABLE 12 SUPAC Requirements for Nonsterile Semisolid Dosage Forms**

| Parameter | Change Level | Requirements[a] | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | A | B | C | D | E | F | G | H | I | J | K |
| Change in components and composition | 1 | • | | • | | | | | | | | |
| | 2 | • | • | | • | | | • | | | | |
| | 3 | • | • | | | • | | • | • | | | |
| Change in preservative components and composition | 1 | • | | | | | | | | • | | |
| | 2 | • | | | | | | | | • | | |
| | 3 | • | • | | • | | | | | • | • | |
| Change in manufacturing equipment | 1 | • | | • | | | | | | | | |
| | 2 | • | • | | | | • | • | | | | |
| Change in manufacturing process | 1 | • | | | | | | | | | | |
| | 2 | • | • | | | | • | • | | | | |
| Change in batch size | 1 | • | • | • | | | | | | | | |
| | 2 | • | • | | • | | | | | | | |

**TABLE 12** ***Continued***

| Parameter | Change Level | Requirements[a] | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | A | B | C | D | E | F | G | H | I | J | K |
| Change in manufacturing site | 1 | • | | | | | | | | | | |
| | 2 | • | • | • | | | | | | | | • |
| | 3 | • | • | | | • | | • | | | | • |

[a]Only those highlighted with black circles:

A: Application/compendial product release requirement.
B: Executed batch records.
C: long-term stability data for first production batch.
D: 3-month accelerated stability data for 1 batch and long-term data for first production batch.
E: 3-month accelerated stability data for 1 batch and long-term data for first 3 production batches if significant information is available or 3-month accelerated stability data for 3 batches and long-term data for first 3 production batches if significant information is not available.
F: 3-month accelerated stability data for 1 batch and long-term data for first production batch if significant information is available or 3-month accelerated stability data for 3 batches and long-term data for first 3 production batches if significant information is not available.
G: In vitro release test.
H: In vivo bioequivalence.
I: Preservative effectiveness test at lowest specified preservative level.
J: Identification and assay for new preservative; validation methods to support absence of interference with other tests.
K: Location of new site.

in vitro release data are used to evaluate the in vivo bioequivalence of a lower strength product [16].

## REFERENCES


1. Ansel, H. C., Allen, L. V., and Popovich, N. G. (1999), *Pharmaceutical Dosage Forms and Drug Delivery Systems*, 7th ed., Lippincott Williams and Wilkins, Philadelphia, pp. 245–250.
2. Breit, J., and Bandmann, H.J. (1973), Dermatitis from lanolin, *Br. J. Dermatol.*, 88, 414–416.
3. Smolinske, S. C. (1992), *Handbook of Food, Drug, and Cosmetic Excipients*, CRC Press, Boca Raton, FL, pp. 225–229.
4. Barker, G. (1977), New trends in formulating with mineral oil and petrolatum. *Cosmet. Toilet.* 92(1), 43–46.
5. Davis, S. S. (1969), Viscoelastic properties of pharmaceutical semisolids I: Ointment bases. *J. Pharm. Sci.*, 58, 412–418.
6. Rowe, R. C., Sheskey, P. J., and Weller, P. J. (2003), *Handbook of Pharmaceutical Excipients*, 4 ed., Pharmaceutical Press and American Pharmaceutical Association, IL, pp. 16–18, 89–92, 260–263, 287–288, 417–418, 491–492, 508–513, 524–525, 618–619, 654–656, 679–684.
7. Hadia, I. A., Ugrine, H. E., Farouk, A. M., and Shayoub, M. (1989), Formulation of polyethylene glycol ointment bases suitable for tropical and subtropical climates I. *Acta Pharm. Hung.*, 59, 137–142.

8. Fisher, A. A. (1978), Immediate and delayed allergic contact reactions to polyethylene glycol, *Contact Dermatitis*, 4, 135–138.
9. Mores, L. R. (1980), Application of stearates in cosmetic creams and lotions, *Cosmet. Toilet.* 95(3), 79–84.
10. Mapstone, G. E. (1974), Crystalization of cetyl alcohol from cosmetic emulsions, *Cosmet. Perfum.*, 89(11), 31–33.
11. Eccleston, G. M. (1984), Properties of fatty alcohol mixed emulsifiers and emulsifying waxes, in Florence, A. T., Ed., *Materials Used in Pharmaceutical Formulations: Critical Reports on Applied Chemistry*, Vol. 6, Blackwell Scientific, Oxford, UK, pp. 124–156.
12. Kline, C. H. (1964), Thixcin, R-Thixotrope, *Drug Cosmet. Ind.*, 95(6), 895–897.
13. Cronin, E. (1967), Contact dermatitis from cosmetics, *J. Soc. Cosmet. Chem.*, 18, 681–691.
14. Forcinio, H. (1998), Tubes: The ideal packaging for semisolid products, *Pharm. Tech.*, 22, 32–36.
15. *The United States Pharmacopeia/The National Formulary*, 28th/23rd ed., U.S. Pharmacopeial Convention, Rockville, HD, 2005, pp. 2246–2255, 2378, 2434–2435, 2441, 2510–2512.
16. Guidance for industry, Nonsterile semisolid dosage forms: Scale-up and postapproval changes, U.S. Food and Drug Administration, May 1997, pp. 1–37.
17. Shah, V. P., Glynn, G. L., and Yacobi, A. (1998), Bioequivalence of topical dosage forms—methods of evaluation of bioequivalence, *Pharm. Res.*, 15, 167–171.
18. Allen, L. V. (2002), *The Art, Science, and Technology of Pharmaceutical Compounding*, American Pharmaceutical Association, Washington, DC, pp. 301–312.
19. Jeong, B., Bae, Y. H., and Kim, S. W. (1999), Thermoreversible gelation of PEG-PLGA-PEG triblock copolymers aqueous solutions, *Macromolecules*, 32, 7064–7069.
20. Miyazaki, S., Suisha, F., Kawasaki, N., Shirakawa, M., Yamatoya, K., and Attwood, D. (1998), Thermally reversible xyloglucan gels as vehicles for rectal drug delivery, *J. Controlled Release*, 56, 75–83.
21. Watanabe, K., Yakou, S., Takayama, K., Isowa, K., and Nagai, T. (1996), Rectal absorption and mucosal irritation of rectal gels containing buprenorphine hydrochloride prepared with water-soluble dietary fibers, xanthan gum and locust bean gum, *J. Controlled Release*, 38, 29–37.
22. Cohen, S., Lobel, E., Trevgoda, A., and Peled, T. (1997), A novel in situ—forming ophthalmic drug delivery system from alginates undergoing gelation in the eye, *J. Controlled Release*, 44, 201–208.
23. Carlfors, J., Edsman, K., Petersson, R., and Jornving, K. (1998), Rheological evaluation of Gelrite in situ gels for ophthalmic use, *Eur. J. Pharm. Sci.*, 6, 113–119.
24. Altagracia, M., Ford, I., Garzon, M. L., and Kravzov, J. (1987), A comparative mineralogical and physico-chemical study of some crude Mexican and pharmaceutical grade montmorillonites, *Drug Dev. Ind. Pharm.*, 13, 2249–2262.
25. Doelker, E. (1993), Cellulose derivatives, *Adv. Polym. Sci.*, 107, 199–265.
26. Lev, R., Long, R., and Mallonga, L. (1997), Evaluation of carrageenan as a base for topical gels, *Pharm. Res.*, 14(11), 42.
27. Daniels, R., Kerstiens, B., Tishinger-Wagner, H., and Rupprecht, H. (1986), The stability of drug absorbates on silica, *Drug Dev. Ind. Pharm.*, 12, 2127–2156.
28. Ruiz-Martinez, A., Zouaki, Y., and Gallard-Lara, V. (2001), In vitro evaluation of benzylsalicylate polymer interaction in topical formulation, *Pharm. Ind.*, 63, 985–988.
29. Ling, W. C. (1978), Thermal degradation of gelatin as applied to processing of gel mass, *J. Pharm. Sci.*, 67, 218–223.

30. Goldstein, A. M., Alter, E. N., and Seaman, J. K. (1973), Guar gum, in Whistler, R. L., Ed., *Industrial Gums*, 2nd ed., Academic, New York, pp. 303–321.
31. Gauger, L. J. (1984), Hydroxyethylcellulose gel as a dinaprostone vehicle, *Am. J. Hosp. Pharm.*, 41, 1761–1762.
32. Farley, C. A., and Lund, W. (1976), Suspending agents for extemporaneous dispensing: Evaluation of alternatives to tragacanth, *Pharm. J.*, 216, 562–566.
33. Cabana, A., Ait-Kadi, A., and Juhasz, J. (1997), Study of the gelation process of polyethylene oxide copolymer (poloxamer 407) aqueous solutions, *J. Colloid Interface Sci.*, 190, 307–312.
34. Burdick, K. H., Haleblian, J. K., Poulsen, B. J., and Cobner, S. E., (1973), Corticosteroid ointments: Comparison by two human bioassays, *Curr. Ther. Res.*, 15, 233–242.
35. Pavics, L. (1970), Comparison of rheological properties of mucilages, *Acta Pharm. Hung.*, 40, 52–59.

# LIQUID DOSAGE FORMS

Maria V. Rubio-Bonilla[1], Roberto Londono[1], and Arcesio Rubio[2]
[1]*Washington State University, Pullman, Washington*
[2]*Caracas, Venezuela*

**Contents**

1 Introduction
2 Generalities
  2.1 Dosage Form
  2.2 Liquid Dosage Form
  2.3 Dispersed Systems
  2.4 Solutions
  2.5 Manufacturing of Nonparenteral Liquid Dosage Forms
  2.6 Optimizing Drug Development Strategies
  2.7 Unit Operation or Batch
  2.8 Batch Management
  2.9 Steps of Liquids Manufacturing Process
  2.10 Protocols
3 Approaches
4 Critical Aspects of Liquids Manufacturing Process
  4.1 Physical Plant
  4.2 Equipment
  4.3 Particle Size of Raw Materials
  4.4 Compounding: Effects of Heat and Process Time
  4.5 Uniformity of Oral Suspensions
  4.6 Uniformity of Emulsions
  4.7 Microbiological Quality
  4.8 Filling and Packing
  4.9 Stability
  4.10 Process Validation
5 Liquid Dosage Forms
  References

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

## 1 INTRODUCTION

Liquid dosage forms are designed to provide the maximum therapeutic response in a target population with difficulty swallowing tablets and capsules and/or to produce rapid therapeutic effects. The major ingredient in most liquid dosage forms is water. While it is the safest and most palatable solvent option, water quality is significant for the stability of pharmaceutical dosage forms. Furthermore, the Food and Drug Administration (FDA) "Guide to Inspections of Dosage Form Drug Manufacturer's—CGMPRs" considers microbial contamination due to inappropriate design and control of purified water systems as the most common problem of liquid dosage forms. Solutions and dispersions studied in this chapter are chemically, microbiologically, and/or physically unstable systems that require a high level of organizational management of manufacturing processes in order to maintain a state of apparent stability, at least until the expiration date [1].

The pharmaceutical industry manufactures dosage forms in large-scale formulations. The decision to scale up is based on the economics of the production process related to costs of materials, personnel, equipment, and control [2]. To reduce costs of wastes and to obtain high-quality and efficacious drug products, the strategic plan to be applied during the process has to be developed carefully. In fact, the variables that affect product quality are identified and understood in the process instead of tested into the final product [3].

Commercial liquid dosage forms reach large-scale production after being preformulated at the laboratory level followed by formulation at the small scale and then at the pilot plant scale. Due to the complexity of the manufacturing process, scale-up from pilot to commercial production is not a simple extrapolation. The approaches to the four levels of production are different. Most of the formulation ingredients are analyzed, studied, and selected at the laboratory scale. While small-scale production is more focused on the liquid preparation procedure with higher amounts of ingredients, the main issues at the pilot plant scale are the design of infrastructure and reduction of costs. Commercial production introduces problems that are not a major issue on a small scale, for instance, materials handling and storage, pulverizing, mixing, dissipation of the generated heat during production, time control, personnel administration, and bottle-filling capabilities. Furthermore, purified water is essential for the manufacturing of these products as well as on-site packing capabilities [2].

The organization and advance of the pharmaceutical industry should be focused on three main points of project development based on quality by design (QbD): product objective (design of experiment, DoE), production resources (process analytical technology, PAT), and product acceptability (quality system) [3]. Manufacturers of liquid dosage forms must ensure safety, efficacy, stability, elegance, and acceptability of the final drug product while achieving development and clinical milestones [1]. Achieving the desirable clinical attributes of the product efficacy means ensuring potency stability by confirming the functionability of the manufacturing process and quality system. The stability and safety of the product are goals to be reached through chemical and microbiological stability by establishing and updating manufacturing and quality control protocols of product development. Time control and aesthetic considerations reflect the physical stability through product elegance and acceptability. Flavoring, sweetening, coloring, and texturing

are both challenges and opportunities. They are challenges because "no single correct method exists to solve significant problems of elegance"; they are opportunities because "they enable a pharmacist to prepare a product more easily accepted by the patient" [4]. Although the most important characteristic of a dosage form is efficacy, there are other characteristics that remain important subjects for the manufacturing of liquid dosage forms such as safety as well as chemical, physical, and microbiological stabilities. From a pharmaceutical point of view, stability problems are the main causes of safety complaints. Despite its significance, some companies decide to outsource stability services [5]. To solve or minimize stability problems in drug products, it is necessary to analyze and enhance the development of critical control points in each operation of the full manufacturing process as well as expected variances and tolerance limits.

Except some aqueous acids, water in aqueous solutions is an excellent media for microbiological growth, such as molds, yeast, and bacteria. Typical microorganisms affecting drug microbiological stability are *Pseudomonas, Escherichia coli Salmonella*, and *Staphylococcus* [1]. Deficient methods or an insufficient preservative system may be the principal causes of microbiological contamination in the pharmaceutical industry of liquid manufacturing [6].

Chemical instability reactions appear with or without microbiological contribution through reactions such as hydrolysis, oxidation, isomerization, and epimerization. Interactions between ingredients and ingredients with container closure materials are established as the principal causes of these reactions [1], for instance, the hydrolysis of cefotaxime sodium, the oxidation of vitamin C, the isomerization of epinephrine, and the epimerization of tetracycline [7].

In most cases, physical instabilities are consequences of previous chemical instabilities. Physical instabilities can arise principally from changes in uniformity of suspensions or emulsions, difficulties related to dissolution of ingredients, and volume changes [6]. For instance, some cases where physical stability has been affected are cloudiness, flocculence, film formation, separation of phases, precipitation, crystal formation, droplets of fog forming on the inside of container, and swelling of the container [8].

Although commercial oral solution and emulsion dosage forms rarely present bioequivalence issues, some bioequivalence problems have been reported for oral suspensions such as phenytoin [9]. The possibility of microbiological contamination and physicochemical instabilities during the manufacturing process also needs to be carefully considered. To approach the stability problems of liquid dosage forms, in this chapter, the main critical aspects during the manufacturing process are based on FDA inspection. From physical plant systems to batching management and packing, the potential sources of microbiological, chemical, and physical instabilities will be analyzed using definitions, case-by-case explanations, and practical examples.

Final product stability, which determines the therapeutic activity and uniformity among other characteristics of the final product, reflects the dynamic of the production process. Conceptualization of stability issues is important to determine the changes to enhance the design space as well as protocols of manufacturing and quality control [3]. An information technology (IT) solution like Enterprice Resource Planning (ERP) may support the pharmaceutical industry's current challenges of organization [10].

## 2 GENERALITIES

### 2.1 Dosage Form

According to the FDA: "A dosage form is the physical form in which a drug is produced and dispensed. In determining dosage form, FDA examines such factors as (1) Elegance: physical appearance of the drug product, (2) Stability: physical form of the drug product prior to dispensing to the patient, (3) Acceptability: the way the product is administered, (4) Efficacy: frequency of dosing, and (5) Safety: how pharmacists and other health professionals might recognize and handle the product" [11]. The term *dosage form* is different from "dose," which is defined as a specific amount of a therapeutic agent that can be taken at one time or at intervals.

### 2.2 Liquid Dosage Form

The physical form of a drug product that is pourable displays Newtonian or pseudoplastic flow behavior and conforms to its container at room temperature. In contrast, a semisolid is not pourable and does not flow at low shear stress or conform to its container at room temperature [12]. According to its physical characteristics, liquid dosage forms may be dispersed systems or solutions.

### 2.3 Dispersed Systems

Dispersed systems are dosage forms composed of two or more phases, where one phase is distributed in another [2]. If a dispersed system is formed by liquid phases, then it is known as an "emulsion." In contrast, the dispersed system is named a "suspension" when the liquid dosage form is accomplished by the distribution of a solid phase suspended in a liquid matrix. The solid phase of a suspension is usually the drug substance, which is insoluble or very poorly soluble in the matrix [12].

### 2.4 Solutions

A solution refers two or more substances mixed homogeneously [2]. Although solubility refers to the concentration of a solute in a saturated solution at a specific temperature, in pharmacy, solution liquid dosage forms are unsatured to avoid crystallization of the drug by seeding of particles or changes of pH or temperature [13]. The precipitation of drug crystals is one of the most important physical instabilities of solutions that may affect its performance [14]. Water is the most used solvent in solutions manufacturing; however, there are also some commercial nonaqueous solutions in the pharmaceutical market [1].

### 2.5 Manufacturing of Nonparenteral Liquid Dosage Forms

The manufacturing of liquid dosage forms with market-oriented planning includes the following stages with respect to special good manufacturing practice (GMP) requirements: planning of material requirements, liquid preparation, filling and packing, sales of drug products, vendor handling, and customer service [15]. From the viewpoint of product stability, each stage of the process includes critical batches

that are more decisive than others. Also, each decisive batch contains one or several unit operations that are more critical than others. The FDA inspection focuses on those critical unit operations to ensure the safety and stability of the liquid dosage forms [6].

### 2.6 Optimizing Drug Development Strategies

> According to Sokoll [16]: The phases of drug development include discovery, preclinical development, clinical development, filing for licensure, approval/licensure and post-approval. Discovery typically includes basic research, drug identification and early-stage process and analytical method development. . . . Emerging companies that review their pipeline objectively and strike a balance between properly resourcing and developing their lead candidates in the clinic while nurturing their next generation of drug candidates will have the best chance for success and sustainability.

### 2.7 Unit Operation or Batch

A "batch" job or operation is defined as a unit of work. Raw materials, semifinished drug products (bulk), and finished drug products are handled in batches. Each different type of material used during the process, such as product packing, should be managed by batches. This applies also to process aids and operation facilities [15].

### 2.8 Batch Management

The batch management of production simplifies the process and makes it easier to control the status of transformation between raw and final products [2]. Some of the data used to follow the material performance around and out of the product manufacturing process are batch-where-used-list, initial status, batch determinations, master data, and expiration date check [15].

The functionality of the overall process to manufacture liquid dosage forms depends on the successful linkage of one unit operation to another. To use mathematical formulations to scale up the manufacturing process, it is necessary to divide the process into stages, batches, and unit operations. Each single unit operation is scalable, but the composite manufacturing process is not. Production problems result from attempts to follow a process scale-up instead of a unit operation scale-up. By using mathematical formulations, it is possible to understand the level of similarity between two scale sizes. In addition, nonlinear similarities between two scale sizes might require the use of conversion factors to achieve an extrapolation point for the scale [2].

### 2.9 Steps of Liquids Manufacturing Process

Establishing short-term goals makes it easier to measure efficiency as well as evaluate the difficulties [2]. Based on these concepts, the problems of manufacturing liquid dosage forms can be approached as problems in one or more batches of the following process steps [6, 15]:

*Planning of Material Requirements* Research and development of protocols and selection of materials; acquisition and analysis of raw materials; physical plant

design, building, and installation; equipment selection and acquisition; personnel selection and initial training; and monitoring information system.

*Liquid Preparation* Research and development of protocols concerning liquid compounding; scale-up of the bulk product compounding; physical plant control and maintenance; equipment maintenance and renovation; continuous training of personnel and personnel compensation plan; and supervision of system reports.

*Filling and Packing* Research and development of protocols concerning filling and packing; scale-up of the finished drug product filling and packing; physical plant control and maintenance; equipment maintenance and renovation; continuous training of personnel and personnel compensation plan; and supervision of system reports.

*Sales of Drug Products* Research and development of protocols concerning product storage; distribution process; continuous training of personnel and personnel compensation plan; and supervision of system reports.

*Vendor Handling* Research and development protocols concerning precautions to maintain product stability; control of vendor stock; and sales system reports.

*Customer Service* Research and development of protocols concerning home storage and handling to maintain product stability; relations with health insurance companies and health care professionals; educational materials for patient counseling; and customer service system reports.

### 2.10 Protocols

Protocols are patterns developed by repeating procedures and fixing the identified problems each time that the procedure is followed. Therefore, protocols are dynamic entities that originally can be developed at a laboratory level but must be adjusted in every new step of the scal -up process. When the manufacturing process moves up in scale, the number of people affected by the protocol increases geometrically. Initially, the information can be obtained from library references, personal tests, interpersonal training, and previous laboratory protocols. However, when the production is scaled up, the information required to fine tune the process comes from monitoring the process itself [2].

## 3 APPROACHES

Quality by Design is a systemic approach that applies the scientific method to the process. QbD theory contains components of management, statistics, psychology, and sociology. The FDA's new century has identified the QbD approach as its "key component" based on process quality control before industry end results [3, 17].

The cooperation between industry members and regulators is increased when the industry explains clearly what it is doing and the agency can understand the formulation and production process. In these cases, regulatory relief appears when industry explores its issues and receives active guidance and programs from the FDA. The agency takes the role of facilitator, or even partner of the industry, in order to improve the strength of the process and formulation [3, 17].

To apply QbD as a systemic approach, the company starts by understanding, step by step, the space design, the design of the dosage form, the manufacturing process, and the critical process parameters to be controlled in order to reach the new building block which is the expectation of variances within those critical process parameters that can be accepted. This approach allows the establishment of priorities and flexible boundaries in the process [3]. Inflexible specifications allow uncontrolled small variances that can follow the butterfly effect of the theory of chaos by producing unpredictable large variations in the long-term behavior of the product shelf-life [18, 19]. In contrast, flexibility, with knowledge of potential variances, reduces changes in the approved spaces and manufacturing protocols [3, 17].

According to the FDA [6], critical parameters during the manufacturing process of nonparenteral liquid dosage forms may appear in the design of physical plant systems, equipment, protocols of usage and maintenance, raw materials, compounding, microbiological quality control, uniformity of suspensions and emulsions, and filling and packing [6].

Process isolation and installation of an appropriate air filtration system in the physical plant may reduce product exposition to chemical and microbiological contaminations. In addition, the use of a suitable dust removal system as well as a heating, ventilation, and air conditioning system (HVACS) may help to repress product chemical instabilities [6].

The equipment of sanitary design, including transfer lines, as well as appropriate cleaning and sanitization protocols may reduce chemical and microbiological contaminations in the final product. Chemical instabilities may be reduced by weighting the right amount of liquids instead of using a volumetric measurement, avoiding the common use of connections between processes, and using appropriate batching equipment [6].

Particle sizes of raw materials are critical to control dissolution in solutions as well as uniformity in suspensions and emulsions. Temperature control during compounding is important since heat helps to support mixing and/or filling operations, but, in contrast, high-energy mixers may produce adverse levels of heat that affect product stability. Too much heat may cause chemical and physical instabilities such as change of particle size or crystallization of drugs in suspensions, dissolution and potency loss of drugs in suspensions, oxidation of components, and activation of microbiological growth after degradation of compounds as well as precipitation of dissolved compounds in solution [20]. In addition, uniformity of suspensions depends on viscosity and segregation factors while solubility, particle size, and crystalline form determine uniformity of emulsions. Application of pharmaceutical GMP for product processes and storage assures microbiological quality. A deficient deionizer water-monitoring program and product preservative system facilitate microbial contamination. Filling uniformity is indispensable for potency uniformity of unit-dose products and depends on the mixing operation. Calibration of provided measuring devices and the use of clean containers will allow administering the right amount of the expected components in the liquid dosage form [6].

Principal product specifications are microbial limits and testing methods, particle size, viscosity, pH, and dissolution of components. Process validation requires control of critical parameters observed during compounding and scale-up. Product stability examination is based on chemical degradation of the active components and interac-

tions with closure systems, physical consequences of moisture loss, and microbial contamination control [6].

## 4 CRITICAL ASPECTS OF LIQUIDS MANUFACTURING PROCESS

### 4.1 Physical Plant

***Heating, Ventilation, and Air Conditioning System*** The manufacturer has to warrant adequate heating, ventilation, and air conditioning in places where labile drugs are processed [6].

The effect of long processing times at suboptimal temperatures should be considered at the production scale in terms of the consequences on the physical or chemical stability of individual ingredients and product. A pilot plant or production scale differs from laboratory scale in that their volume-to-surface-area ratio is relatively large. Thus, for prolonged suboptimal temperatures, jacketed vessels or immersion heaters or cooling units with rapid circulation times are absolutely necessary [2].

For heat-labile drugs, uncontrolled temperature increments can activate auto-oxidation chains when the drug product ingredients react with oxygen and generate free radicals but without drastic external interference. Vitamins, essential oils, and almost all fats and oils can be oxidized. A good example of a heat-labile drug solution is clindamycin, which has to be stored at room temperature and away from excess heat and moisture [19]. Auto-oxidation chains are finished when free radicals react with each other or with antioxidant molecules (quenching). The tocopherols, some esters of gallic acid, as well as BHA and BHT (butylated hydroxyanisole and butylated hydroxytoluene) are common antioxidants used in the pharmaceutical industry [1].

***Isolation of Processes*** To minimize cross-contamination and microbiological contamination, the manufacturer may develop special procedures for the isolation of processes. The level of facilities isolation depends on the types of products to be manufactured. For instance, steroids and sulfas require more isolation than over-the-counter (OTC) oral products [6]. To minimize exposure of personnel to drug aerosols and loss of product, a sealed pressure vessel must be used to compound aerosol suspensions and emulsions [21]. An example of cross-contamination with steroids was the controversial case of a topical drug manufactured for the treatment of skin diseases. High-performance liquid chromatography/ultraviolet and mass spectrometry (HPLC/UV, HPLC/MS) techniques were used by the FDA for the detection of clobetasol propionate, a class 1 superpotent steroid, as an undeclared steroid in zinc pyrithione formulations. The product was forbidden and a warning was widely published [22].

***Dust Removal System*** The efficiency of the dust removal system depends on the amount and characteristics of dust generated during the addition of drug substance and powdered excipients to manufacturing vessels [6]. Pharmaceutical industries usually generate some type of dust or fume during processing. Important factors for selecting dust collectors are maintenance, surrogate test, economics, and containment. In addition, reentrainment of the fine particles, vertical or horizontal position, efficiency, pressure resistant, service life time, as well as sealing capacity to work

through the bag are significant factors concerning filter selection of dust removal systems. Some examples of dust collection applications in the manufacture of liquid dosage forms are handling and pulverization of raw materials, spray dryers, and general room ventilation [23].

***Air Filtration System*** The efficiency of the air filtration system has to be demonstrated by surface or air-sampling data where the air is recirculated [6]. To monitor the levels of contamination in the air, there are commercial automatic samplers for microbiological contamination or gas presence. Air trace environmental samplers for pharmaceutical industries are based on the slit-to-agar impaction technique for the presence of viable microorganisms. Automatic samplers for compressed gas analyze the presence of a specified gas in $1\,m^3$ by absorbing air at a fixed flow rate for a sampling period of 1 h or a different adjusted time. These solutions to the sampling needs of the pharmaceutical industry are robust, require low maintenance, and are easy to use. This allows for validation of sampling data at the moment of application filling to support the process control. Sampling time and selection of microbiological growth media or analysis technique are important components to consider when developing a sampling plan [24].

### 4.2 Equipment

***Sanitary Design*** Pumps, valves, flowmeters, and other equipment should be easily sanitized. Some examples of identified sources of contamination are ball valves, packing in pumps, and pockets in flowmeters [6].

The sanitary design and performance of equipment make it accessible for inspection, cleaning, and maintenance. It has to be cleanable at a microbiological level and its performance during normal operations should contribute to sanitary conditions. The materials used in the design have to assure hygienic compatibility with other equipment, the product, the environment, other systems such as electrical, hydraulics, steam, air, and water, as well as the method and products used for cleaning and sanitation. The equipment should be self-draining to assure product or liquid collection. Small niches, for example, pits, cracks, corrosion, recesses, open seams, gaps, lap seams, protruding ledges, inside threads, bolt rivets, and dead ends, as well as inaccessible cavities of equipment such as entrap and curlers must be eliminated whenever possible; otherwise they have to be permanently sealed. Enclosures, for example, push buttons, valve handles, switches, and touch screens, should be prepared for a hygienic design of maintenance. Standards have been developed by the American Meat Institute [25].

***Standard Operating Procedures for Cleaning Production Equipments*** Current GMPs are defined as the basic principles, procedures, and resources required to guarantee an environment appropriate for manufacturing products of adequate quality [26]. To minimize cross-contamination and microbiological contamination, it is GMP for a manufacturer to create and pursue written standard operating procedures (SOPs) to clean and sanitize production equipment in a way that avoids contamination of in-progress and upcoming batches. When the drug is known as a potent generator of allergic reactions, such as steroids, antibiotics, or sulfas, cross-contamination becomes an issue of safety [20]. In addition, validation and data analysis procedures,

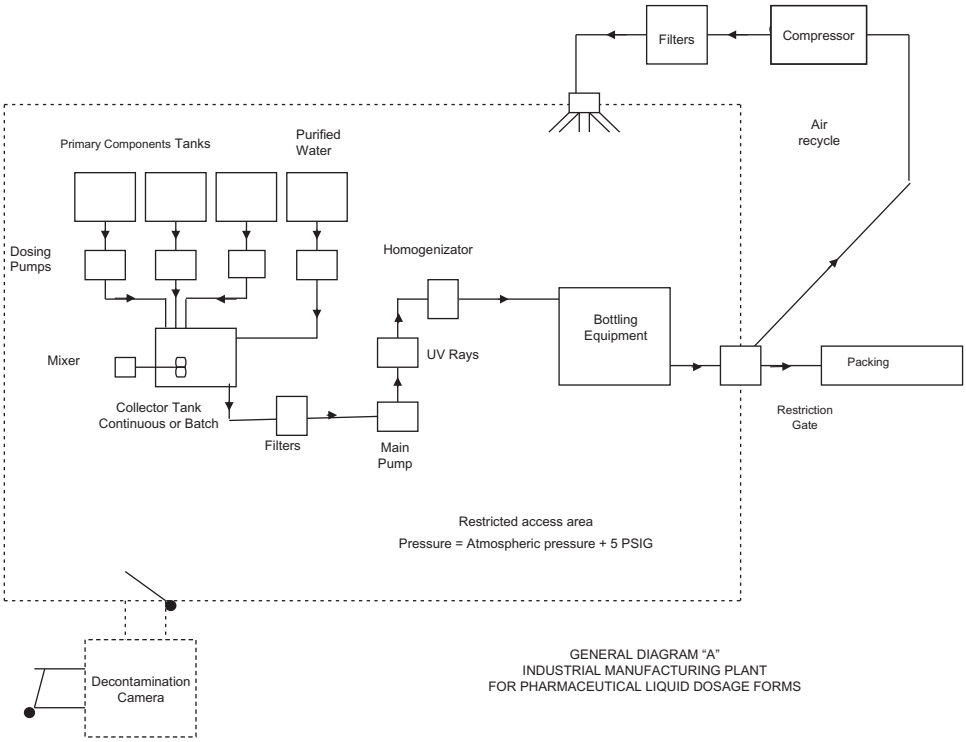


**FIGURE 1** Mixing and filling lines for pharmaceutical dosage forms. Positive indoor pressure of 5 psi over outdoor pressure assures constant airflow from inside to outside in order to reduce entrance of contaminating agents.

including drawings of the manufacturing and filling lines [6], are especially important for clean-in-place (CIP) systems, as indicated in Figures 1 and 2.

Many companies have problems with standardizing operating procedures for cleaning steps and materials used [6]. Appropriate SOPs are necessary to determine the scope of the problem in investigations about possible cross-contaminations or mix-ups. The best approach to validate a SOP is to test it, use it as a training tool, and observe the results obtained by different persons. This includes the worst-case situation in order to enhance the step-by-step writing methodology as well as standardizing the materials used. A typical SOP contains a header to present the SOP title, date of issue, date of last review, total number of pages, responsible person, and approval signature. Typically, a SOP includes position of responsible person, SOP purpose and scope, definitions, equipment and materials, safety concerns, step-by-step procedure, explanation of critical steps, tables to keep data, copies of forms to fill, and references [26]. The forms to keep the records must show the date, time, product, and lot number of each batch processed. However, the most important points of the SOP are equipment identity, cleaning method(s) with documentation of critical cleaning steps, materials approved for cleaning that have to be easily removable, names and position of persons responsible for cleaning and inspection, inspection methods, and maintenance and cleaning history of the equipment [20].

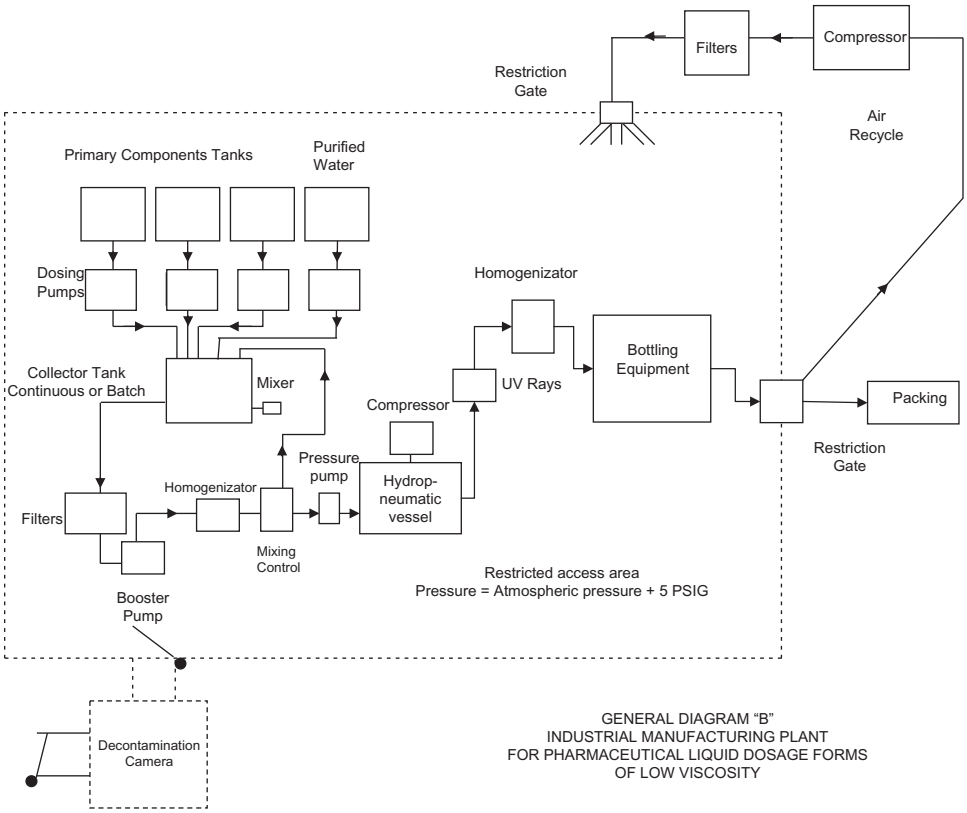


**FIGURE 2** Mixing and filling lines for pharmaceutical dosage forms. Using this hydro-pneumatic system, instead of the mechanical system in Figure 1, the liquid moves by the pressure generated in a compressed air tank.

***Cleaning and Sanitizing Transfer Lines*** Pipes should be hard, easily cleaned, and sanitized. To avoid moisture collection and microbiological contamination, hoses should be stored in a way that allows them to drain rather than be looped. For example, transfer lines are an important source of contamination when flexible hoses are handled by operators, lying on the floor, and after they are placed in transfer or batching tanks [6].

Heat is considered one of the most efficient physical treatments for sanitizing pharmaceutical equipment and could be used for sanitizing hoses that have already been cleaned. The recirculation of hot water at a temperature of 95°C for at least 100 min allows bacteria elimination [14].

Due to the important amount of insoluble residues left on piping and transfer lines after emulsion manufacturing, such as topical creams and ointments, equipment cleaning becomes difficult to address. To avoid cross-contamination, some manufacturers have decided to dedicate lines and hoses to specific products.

However, these decisions have to appear in the written production protocols and SOPs [20].

***Sampling Cleaned Surfaces for Presence of Residues*** The cleaning method is validated by sampling the cleaned surfaces of the equipment for the existence of residues. The equipment characteristics and residue solubility are factors to support the selection of the sampling method to be used [6]. There are two acceptable general types of sampling methods: direct surface sampling by swabbing of surfaces and rinse sampling with a routine production in-process control. Although surface residues will not be identical on each part of the surface, statistically the most advantageous is direct surface sampling because it allows evaluation of the hardest areas to clean as well as insoluble or "dried-out" residues by physical removal. The type of sampling material and solvent used for extraction from the sampling material should be validated in order to determine their impact on the test data. The second method, rinse sampling, is used for larger surfaces or inaccessible systems. Contaminants that are physically occluded and insoluble residues are disadvantages of the rinse sampling method. To validate this cleaning process, direct measurement of the contaminant in the rinse water has to be tested instead of a simple test for water quality. Routine production in-process control is used as indirect testing for large equipment that has to be cleaned by the rinse sampling method. The uncleaned equipment has to give an unacceptable result for the indirect test [27].

***Establishing Appropriate Limits on Levels of Postequipment Cleaning Residues*** Very low levels of residue are possible to be determined since technological advances offer more sensitive analytical methods. The manufacturer should know the toxicological information of the materials used and potential amounts of residues after exposure to the equipment surface. Accordingly, the manufacturer has to establish proper limits of residues after equipment cleaning and scientifically justify these limits. The established limits must be clinically and pharmaceutically safe, realistic, viable, and verifiable [20]. The sensitivity of the analytical method will determine the logic of the established limits since absence of residues could indicate a low sensitivity of the analytical method or a poor sampling procedure. Sometimes thin-layer chromatography (TLC) screening must be used in addition to chemical analyses. Some practical levels established by manufacturers include 10 ppm of chemicals, 1/1000 of the biological activity levels met on a normal therapeutic dose, and no visible residues of particles determined organoleptically [27].

***Connections*** Connectors and manifolds should not be for common use. For example, sharing connectors in a water supply, premix, or raw material supply tanks may be a source of cross-contamination [6].

***Time between Completion of Manufacturing and Initiation of Cleaning*** The time that may elapse from completion of a manufacturing operation to initiation of equipment cleaning should also be stated where excessive delay may affect the adequacy of the established cleaning procedure. For example, residual product may dry and become more difficult to clean [20].

SOPs are an example of deficiency in many manufacturers regarding time limitations between batch cleaning and sanitization [6]. Lack of communication between

departments responsible for the production at different levels is the main cause of time control problems. Typically each department, from human resources to finances, manufacturing, and warehouse, has its own computer system optimized for the particular ways that the department does its work. Therefore, time control becomes a primordial issue when labile materials are transferred from one department to another [28].

To facilitate communication between different departments, some useful softwares have been developed. For example, ERP is an integrated approach which may have positive payback if the manufacturer installs it correctly. An ERP is a type of software that can improve communication between planning and resources. The software attempts to integrate all departments and functions in a company onto a single computer system that can serve each particular need, such as finance, human resources, manufacturing management, process manufacturing management, inventory management, purchasing management, quality management, and sales management. Each department has its own software, except now the software is linked together, so that, for example, someone in manufacturing can look into the maintenance software to see if specific batch cleaning and sanitization have been scheduled or realized and someone in finance can review the warehouse software to see if a specific order has been shipped. The information is online and not in someone's heads or on papers that can be misplaced. People in different departments may see the same information, update it if they are allowed to do, and make right decisions faster. However, the software is less important than the changes companies make in the ways they work. Reorganization and training are the keys of ERP's success to fix integration problems. There are three different ways to install an ERP: big bang, franchising strategy, and slam dunk. Big bang is the most ambitious way whereby companies install a single ERP across the entire company. By the franchising strategy, departments do not share many common processes across, whereas slam dunk is focused on just a few key processes [28].

***Weight in Formulations*** Flow properties of liquids rarely vary due to their constant density at a constant temperature. Oral solutions and suspensions are formulated on a weight basis (gravimetry) in order to be able to measure the final volume by weight before filling and packing. Volumetric measurements of liquid amounts to be used for manufacturing liquid dosage forms have shown greater variability than weighted liquids. For instance, the inaccurate measurement of the final volume by using dip sticks or a line on a tank may cause further analytical errors and potency changes [6].

The importance of selecting gravimetry instead of volumetry to measure liquid amounts in the pharmaceutical industry of liquid dosage forms is well illustrated by the volume contraction of water–ethanol and volume expansion of ethyl acetate–carbon disulfide liquid mixtures as well as a CS2–ethyl acetate system. The National Formulary (NF) diluted alcohol is a typical example of the volume nonadditivity of liquid mixtures [29]. This solution is prepared by mixing equal volumes of alcohol [U.S. Pharmacopeia (USP)] *USP* and purified water (USP). The final volume of this solution is about 3% less than the sum of the individual volumes because of the contraction due to the mixing phenomenon [1]. In addition, molecular interactions of surfactants in mixed monolayers at the air–aqueous solution interface and in mixed micelles in aqueous media also cause some contraction of volume upon mixing [30].

***Location of Bottom Discharge Valve in Batching Tank*** The bottom discharge valve should be located exactly at the bottom of the tank. In some cases valves have been found to be several inches to a foot above the bottom of the tank [6].

For a tank suspected of having substantial deposits at the bottom, a fiber-optic camera can be inserted in the tank to provide a view and positive confirmation of the tank bottom condition. These camera and light vision systems are sanitary equipment able to provide a computational real-time visual inspection of the inside tank under process conditions or pressure vessel. In addition, they are used to control several parameters during the manufacturing process, such as product level and thickness, solids level, uniformity of suspensions, foam, and interface and/or cake detection [31].

***Batching Equipment to Mix Solution*** Ingredients of solutions have to be completely dissolved. For instance, it has been observed that some low-solubility drugs or preservatives can be kept in the "dead leg" below the tank, and the initial samples have reduced potency [6]. When there is inadequate solubility of the drug in the chosen vehicle, the dose is unable to contain the correct amount of drug in a manageable size unit, that is, one teaspoonful or one tablespoonful. Thus, ingredients as well as handling and storage conditions should be chosen to manage the problem [14].

In solutions, the most important physical factors that influence the solubility of ingredients are type of fluid, mixing equipment, and mixing operations. Generalized Newtonian fluids are ideal fluids for which the ratio of the shear rate to the shear stress is constant at a particular time. Unfortunately, in practice, usually liquid dosage forms and their ingredients are non-Newtonian fluids in which the ratio of the shear rate to the shear stress varies. As a result, non-Newtonian fluids may not have a well-defined viscosity [32].

When all the ingredients are miscible liquids, the combination and distribution of these components to obtain a homogeneous mixture are called blending. Whenever possible, ingredients should be added together and the impeller mixer often is located near the bottom of the vessel [21]. Mixing of high-viscosity materials requires higher velocity gradients in the mixing zone than regular blending operations. In fact, the fundamental laws of physics regarding the performance of Newtonian fluids in the production process may be studied using computational tools. For example, VisiMix is a software that is routinely used to calculate shear rates [2].

Finally, if it is determined that there is a bigger problem of insolubility coming from the formulation, then addition of cosolvents, surfactants, as well as the preparation of the ionized form of an acid or base, drug derivatization, and solid-state manipulation are approaches to manipulating the solubility of the drug [14].

***Batching Equipment to Mix Suspension*** In the case of suspensions, the flow necessary to overcome settling in a satisfactory suspension depends on the mixing equipment and is predicted by Stokes's law. Thus, to use the Stokes's law, suspensions are considered as Newtonian fluids if the percentage of solids is below 50%. Mixing equipment uses a mechanical device that moves through the liquid at a given velocity. Dispersing and emulsifying equipment is categorized as "high-shear" mixing equipment. The maximum shear rate with such equipment occurs very close to the

mixing impeller. Therefore, the diameter of the impeller and the impeller speed directly influence the power applied by the mixer to the liquid [21].

***Batching Equipment to Mix Emulsion*** The most common problems of mixing emulsions are removing "dead spots" of the mixture and scrapping internal walls of the mixer. Dead spots are quantities of ingredients that are not mixed and become immobile. Where dead spots are present, that quantity of the formula has to be recirculated or removed and not used. If the inside walls of the mixer keep residual material, operators should use hard spatulas to scrape the walls; otherwise the residual material will become part of the next batch. In both cases, the result may be nonuniformity. Stainless steel mixers have to include blades made of hard plastic, such as Teflon, to facilitate the scrapping of the mixer walls without damaging the mixer. Scrapper blades should be flexible enough to remove internal material but not too rigid to avoid damaging the mixer [20]. The mixing will be successful if the macroscale mixing offers sufficient flow of components in all areas in the mixing tank and the microscopic examination shows a correct particle size distribution [33].

### 4.3 Particle Size of Raw Materials

***Raw materials in Solution*** The types of raw materials used to be part of solutions are presented in Table 1. They have different purposes and can be cosolvents, electrolytes, buffers, antioxidants, preservatives, coloring, flavoring and sweetener agents, among others.

***Particle Size of Raw Materials in Solution*** Particle size is affected by the breaking process of the particle, crystal form, and/or salt form of the drug. The particle size can affect the rate of dissolution of raw materials in the manufacturing process. Raw materials of a finer particle size may dissolve faster because they have a larger surface area in contact with the solvent than those of a larger particle size when the product is compounded [6]. Mixing faster causes the particle to break down and dissolve more quickly. In addition, hydrated particles are less soluble than their anhydrous partners [37].

**TABLE 1 Solutions: pharmaceutical excipients**

| Purpose | Agent |
|---|---|
| Protecting the active product ingredients | - Buffers<br>- Antioxidants<br>- Preservatives |
| Maintaining the appearance | - Colorings<br>- Stabilizers<br>- Cosolvents<br>- Antimicrobial preservatives<br>- Electrolytes |
| Taste/Small Masking | - Sweeteners<br>- Flavorings |

*Source*: From ref. 4, 34, 35, 36

Solid drugs may occur as pure crystalline substances of definite identifiable shape or as amorphous particles without definite structure. In addition, when a drug particle is broken up, the total surface area is increased as well as its rate of dissolution. The amorphous form of a chemical is usually more soluble than the crystalline form while the crystalline form usually is more stable than the amorphous form [37]. Processing conditions used for providers to obtain raw materials can dramatically impact their quality and stability; for instance, the presence of different polymorphs may depend on the thermal history of freezing, concentration of solvents, and drying conditions [38]. The polymorphism of a crystalline form is the capacity of a chemical to form different types of crystals, depending on the conditions of temperature, solvents, and time followed for its crystallization. Among different polymorphs, only one crystalline form is stable at a given temperature and pressure. Over time, the other crystalline forms, called metastable forms, will be transformed into stable forms. Transformations longer than the shelf-life of metastable forms into stable forms of a drug are very common in final products and compromise its stability and efficacy to different extents depending on quality control [37].

While the metastable forms offer higher dissolution rates, many manufacturers use a particular amorphous, crystalline, salt, or ester form of a drug with the solubility needed to be dissolved in the established conditions, for instance, to prepare a chloramphenicol ophthalmic solution [39]. Thus, the selection of amorphous or crystalline form of a drug may be of considerable importance to facilitate the formulation, handling, and stability [37].

However, the dissolution rate of an equal sample of a slowly soluble raw material usually will increase with increasing temperature or rate of agitation as well as with reduce viscosity, changes of pH or nature of the solvent. In addition, other alternative mechanisms to enhance the solubility of insoluble drugs are: 1) hydrophilization: the reduction in contact angle or angle between the liquid and solid surface [40], which can be accessed by intensive mixing of the hydrophobic drug with a small amount of methylcellulose solution [41]; 2) the formation of microemulsions: by covering small particles with surfactants to obtain micromicelles that are visible only in the form of an opalescence; and, 3) the formation of complexing compounds: by adding a soluble substance to form soluble reversible complexes. However, the last method is used with some restrictions [42].

***Raw Materials in Suspension*** The types of raw materials used to be part of suspensions are presented in Table 2. They have different purposes and can be wetting agents, salt formation ingredients, buffers, polymers, suspending agents, flocculating agents, electrolytes, antioxidants, poorly soluble Active Product Ingredients, preservatives, coloring, flavoring and sweetener agents, among others.

***Particle Size of Drug in Suspension*** The physical stability of a suspension can be enhanced by controlling the particle size distribution [43]. Uncontrolled changes of drug particle size in a suspension affect the dissolution and absorption of the drug in the patient. Drug substances of finer particle size may be absorbed faster and bigger particles may not be absorbed. Aggregation or crystal growth is evaluated by particle size measurements using microscopy and a Coulter counter [21] or preferably techniques that allow samples to be investigated in the natural state. Allen [44] offers an academic and industrial discussion about particle characterization.

**TABLE 2 Suspensions: pharmaceutical excipients**

| Purpose | Agent |
|---|---|
| Facilitating the connection between Active Product Ingredient and vehicle | - Wetting agents particle size (>0.1 μm)<br>- Salt formation ingredients<br>- Sugars |
| Protecting the Active Product Ingredients | - Buffering–systems<br>- Polymers<br>- Antioxidants<br>- Poorly soluble drugs |
| Maintaining the suspension appearance | - Colorings<br>- Suspending agent<br>- Flocculating agent<br>- Antimicrobial preservatives<br>- Electrolytes |
| Masking the unpleasant taste/smell | - Sweeteners<br>- Flavorings<br>- Poorly soluble Active Product Ingredient |

*Source*: From ref. 4, 34, 35, 36

Powder properties and behavior, sampling, numerous potential particle size measuring devices, available equipment as well as surface and pore size are his principal themes.

Particles are usually very fine (1–50 μm). For instance, topical suspensions use less than 25 μm particle size [6]. The particle size of the drug is the most important consideration in the formulation of a suspension, since the sedimentation rate of disperse systems is affected by changes in particle size. Finer particles become interconnected and produce particle aggregation followed by the formation of nonresuspendable sediment, known as caking of the product. The two main causes of aggregation and caking are energetic bonding and bonding through shared material. A statistical wide distribution of particle sizes gives more compact packing and energetic bonding than narrower distributions. It has been observed that heat treatments can cause agglomeration of particles, not only due to energetic bonding but also by formation of crystal bridges. Also, when the application of shear forces to mix and homogenize the suspension uses too high energy inputs, then the probability for aggregation increases [43].

Examples of oral suspensions in which a specific and well-defined particle size specification for the drug substance is important are phenytoin suspension, carbamazepine suspension, trimethoprim and sulfamethoxazole suspension, and hydrocortisone suspension [6].

There are some useful methods to improve the physical stability of a suspension, such as decreasing the salt concentration, addition of additives to regulate the osmolarity, as well as changes in excipient concentrations, unit operations in the process, origin and synthesis of the drug substance, polymorphic behavior of the drug substance crystals, and other particle characteristics. However, methods based on changes of the particle properties and the surfactants used are the most successful [43].

**TABLE 3 Emulsions: pharmaceutical excipients**

| Purpose | Agent |
|---|---|
| Particle Size | - Solid particles (10 nanometers to 5 micrometers size) |
| | - Droplet particles (0.1–1.0 micrometers size) |
| Protecting the Active Product Ingredients | - Buffering-Systems |
| | - Polymers |
| | - Antioxidants |
| | - Distribution pattern (O/W, W/O) |
| Maintaining the appearance | - Colorings |
| | - Emulsifying agents |
| | - Penetration enhancers |
| | - Gelling agents |
| | - Stabilizers |
| | - Antimicrobial preservatives |
| Taste/smell Masking | - Sweeteners |
| | - Flavorings |
| | - Relation oil vs. water |

*Source*: From ref. 4, 34, 35, 36

To approach physical stability problems of suspensions, effectiveness and stability of surfactants as well as salt concentrations must be checked with accelerated aging. In addition, unit operations affecting particle size distribution, surface area, and surfactant effectiveness should be approached, taking into account that different types of distributions, for instance, volume or number weighted, give a different average diameter for an equal sample [43].

***Raw Materials in Emulsions*** The types of raw materials used to be part of emulsions are presented in table 3. They have different purposes and can be buffers, polymers, emulsifying agents, penetration enhancers, gelling agents, stabilizers, antioxidants, preservatives, coloring, flavoring and sweetener agents, among others.

***Particle Size in Emulsions*** When a solid drug is suspended in an emulsion, the liquid dosage form is known as a coarse dispersion. In addition, a colloidal dispersion has solid particles as small as 10 nm–5 μm and is considered a liquid between a true solution and a coarse dispersion [44].

### 4.4 Compounding: Effects of Heat and Process Time

***Oxygen*** Oxygen removal for processing materials that require oxygen to degrade is possible by methods such as nitrogen purging, storage in sealed tanks, as well as special instructions for manufacturing operations [6]. For instance, sealing glass ampules containing a liquid dosage form with heat under an inert atmosphere is a packing mechanism used to prevent oxidation. Some aspects of oxygen sensitiveness that should be taken into account are the necessity of water and headspace deoxygenation in ampules before sealing, the avoidance of multidose vials that facilitate oxygen contact with the product after opened, and rubber stoppers for vial sealing that are permeable to oxygen as well as release additives to catalyze oxidative reac-

tions. Rubber stoppers soften and get sticky over time because all rubber products degrade as sulfur bonds induced during vulcanization revert. Connors et al. [45] present the oxygen content of water at different temperatures and an interesting discussion of calculations for the case of captopril as an oxygen-sensitive drug.

***Dissolution of Drugs in Solutions*** Although some compounds, such as poloxamers, decrease their aqueous solubility with an increase in temperature [46], usually, drugs dissolve more quickly when the temperature increases because particle vibration is augmented and the molecules move apart to form a liquid. Chemical instabilities by oxidation due to high temperature or prolonged periods of heat exposure can occur when trying to increase the dissolution of poorly soluble raw materials. To control such instabilities, charts of time and amount of temperature treatments to dissolve materials as well as tests of dissolution are required [6]. In addition, precipitations and other reactions may occur between salts in solution and can be anticipated by using heat-of-mixing data and activation energy calculations for decomposition reactions. Connors et al. [45] provide examples of calculations about effects of temperature on chemical stability of pharmaceuticals in solution. Regarding the instability of the product, the reasons to limit temperature amounts can go from controlling final concentration changes to controlling burn-on/fouling when too-high temperatures are applied [45]. Usually salts are more soluble in water and alcohol than weak acids or bases. The reason salts are not always the best choice to increase the solubility of a drug is its permeability. Oral drug absorption depends not only on solubility and dissolution but also on permeability through the cellular membrane. Drugs have to be able to dissolve not only in the aqueous fluids of the body before reaching the intestinal wall but also in the lipophilic environment of the cellular membrane in order to reach the internal part of the cell and interfere with its functionability. Therefore, the cosolvent approach is essential if the drug presents problems in dissolving in the media. The dielectric constant of a solvent is a relative measure of its polarity. Comparing the hydroxyl–carbon ratio of the solvent molecule allows establishing the relative polarity of the cosolvent as determined by its dielectric constant [47]. Remington describes the formulations of some solutions, such as the ferrous sulfate syrup, amantadine hydrochloride syrup, phenobarbital elixir, and theophylline elixir [1].

***Potency of Drugs in Suspension*** To avoid degradation of the suspended drug substance by high temperature or prolonged periods of heat exposure, it is necessary to record the time and amount of temperature treatments on charts [6]. The rate of dissolution of a suspended drug increases with the increase in temperature. The potency stability of a suspended drug depends on the concentration of the dissolved drug since drug decomposition occurs only in solution [48]. The goal is to avoid the dissolution of suspensions. Changing the pH of the vehicle or replacing the drug with a less soluble molecule may result in enhanced potency stability of the suspended drug [48].

For instance, when the chemical stability of a suspension of ibuprofen powder and other ibuprofen–wax microspheres was studied with a modified HPLC procedure for three months, the amount of drug released from the microspheres was affected by the medium pH, type of suspending agent, and storage temperature without observing chemical degradation of the drug [49].

***Temperature Uniformity in Emulsions*** During the preparation of emulsions, heat may be increased as part of the manufacturing protocol or mixing operation system. Temperature measurements should be monitored and documented continuously using a recording thermometer if the temperature control is critical or using a hand-held thermometer if it is not a critical factor. Temperature may be critical in the manufacturing process depending on the thermosensitivity of the drug product and excipients as well as the type of mixer used. To guarantee the temperature uniformity during the mixing operation, manufacturers may consider the relation between the container size, mixer speed, blade design, viscosity of the contents, and rate of heat transfer [20].

Fong-Spaven and Hollenbeck [50] studied the apparent viscosity as a function of the temperature from 25 to 75 °C of an oil–water emulsion stabilized with 5% triethanolamine stearate (TEAS) using a Brookfield digital viscometer. They observed that the viscosity decreased when the temperature reached about 48 °C, but surprisingly viscosity increased to a small peak at 54 °C and then continued decreasing after that peak. The viscosity peak was attributed to a transitional gellike arrangement molecular structure of TEAS that is destroyed as soon as the temperature continues increasing, the TEAS crystalline form reappears, and viscosity again decreases [36].

***Microbiological Control*** To avoid chemical instabilities that yield microbiological and physical instabilities, as a result of high temperature or prolonged periods of exposure, it is necessary to record the time and amount of temperature treatments on charts [6].

***Product Uniformity*** Charts of storage and transfer operation times for the bulk product are required to control the risk of segregation. Transfers to the filling line and during the filling operation are the most critical moments to keep the suspension uniformity [6]. The implementation of an ERP for time scheduling is the best solution for time control and organization of resources. However, it could be difficult due to the reluctance of people to change [10]. The constant flow of the liquid through the piping, the constant mixing of the bulk product in the tank, as well as the transfer of small amounts near the end of the filling process to a smaller tank during the filling process may minimize segregation risks [6].

***Final Volume*** Excess heating produces variations of the final volume over time [6]. Although increasing solute concentration can elevate the boiling point and reduce evaporation of water, changes in drug concentration are undesirable because they yield different final products. Regarding the instability of the product, the reasons to limit temperature amounts can go from controlling final concentration changes to controlling burn-on/fouling when too-high temperatures are applied [36].

A solution is a liquid at room temperature that passes into the gaseous state when heated at very high temperature, forming a vapor with determined vapor pressure, through a process called vaporization. The kinetic energy is not evenly distributed between the molecules of the liquid. When the liquid is in a closed container at a constant temperature, the molecules with the highest kinetic energy leave the surface of the liquid and become gas molecules. Some of the gas molecules remain as gas and others condense and return to the liquid. When, at a determined temperature,

the rate of condensation equals the rate of vaporization, the equilibrium vapor pressure is reached. However, vapor pressure increases with increases in liquid temperature, resulting in more molecules leaving the liquid surface and becoming gas molecules [51].

***Storage*** Charts of time and temperature of storage are important to control the increased levels of degradedness [6]. Shelf life is defined as the amount of time in storage that a product can maintain quality and is equivalent to the time taken to reach 90% of the composition claim or have 10% degradation. The availability of an expiration date is assumed under specified conditions of temperature. Based on zero- and first-order reaction calculations, Connors et al. [45] show the estimation methods to determine the shelf life of a drug product at temperatures different from the one specified under standard conditions.

### 4.5 Uniformity of Oral Suspensions

Keeping the particles uniformly distributed throughout the dispersion is an important aspect of physical stability in suspensions. Based on Stokes's law for dilute suspensions where the particles do not interfere with one another, there are different factors that control the velocity of particle sedimentation in a suspension, for instance, particle diameter, densities of the dispersed phase and the dispersion medium, as well as viscosity of the dispersion medium [36]. Remington describes the formulation of trisulfapyrimidines oral suspension [1]. In addition, Lieberman et al. [42, 48] are also good sources of typical formulations for suspensions.

***Viscosity*** Depending upon the viscosity, many suspensions require continuous or periodic agitation during the filling process [6].

***Segregation in Transfer Lines*** When the stored bulk of a nonviscous product is transferred to filling equipment through delivery lines, some level of segregation is expected. The manufacturer has to write the procedures and diagrams for line setup prior to filling the product [6]. Delivery lines of suspensions increase the tendency of particles of the same size to assemble together. However, slightly increasing the global mixing in the lines can easily reverse the segregation without enhancing the global mixing [52]. Shear stress versus rate of shear can be plotted to determine the flow pattern of a specific suspension as pseudoplastic, Newtonian, or dilatant. The type of flow is determined by the slope of the plot. While shaking increases the yield stress and causes particles flow, the cessation of shear and rest rebuilds the order of the system. A good-quality suspension is known as a thixotrophic system and is obtained when the particles at rest avoid or show reduced sedimentation. The rheogram of a thixotrope system presents a typical hysteresis or curve representing different shear stresses over time [33].

***Quality Control*** The GMPs for suspensions include testing samples at different checkpoints in the procedure, at the beginning, middle, and end, as well as samples from the bulk tank. The uniformity will be successful only if, on microscopic analysis, the components are dispersed to the expected particle size distribution established by product development. Visual and microscopic examinations should consist of looking for verification of foam formation, segregation, and settling, although testing

for viscosity is important to determine agitation during the filling process. Samples used for tests should not be combined again with the lot [6, 33].

### 4.6 Uniformity of Emulsions

Remington describes the following three typical formulas of emulsions: type A gelatin, mineral oil emulsion (USP), and oral emulsion (O/W) containing an insoluble drug [1]. In addition, Lieberman et al. [42, 50] are also good sources of typical formulations for emulsions. The components of the emulsion system may present physical and chemical instabilities reflected on the distribution of an active ingredient, component migration from one phase to another, polymorphic changes in components, and chemical degradation of components [33].

***Solubility*** The soluble active ingredient should be added to the liquid phase that will be its carrier vehicle. Data of solubility have to be determined as part of the process validation. Potency uniformity has to be tested by demonstrating satisfactory distribution in the emulsified mix [20].

***Particle Size*** Regarding globule diameter in emulsions, the size–frequency distribution of particles in an emulsion over time may be the only method for determining stability [36]. Drug activity and potency uniformity of insoluble active ingredients depend upon control of particle size and distribution in the mix [6]. In addition, aggregation of the internal phase droplets, formation of larger droplets, and phase separation are categorized as emulsion system instabilities that are reflected in the particle size distribution of the emulsion. The measurement of particle size distribution over time allows the characterization of the emulsion stability and determines the rheological behavior of the emulsion. Well-accepted approaches to determine particle size distribution include microscopy, sedimentation, chromatography, and spectroscopy. However, these analyses are problematic in a multiphase emulsion [33].

***Crystalline Form*** Uncontrolled temperature or shear can induce changes in component crystallinity or solubility. For this reason, analytes originally present in each phase of the product should be counted as well possible interactions with the container or closure and the processing equipment analyzed. Some techniques used to obtain information about the emulsion system and its components are microscopic examination, macro- and microlaser Raman, and rheological studies [33]. The FDA guidance offers the following example: "in one instance, residual water remaining in the manufacturing vessel, used to produce an ophthalmic ointment, resulted in partial solubilization and subsequent recrystallization of the drug substance; the substance recrystallized in a larger particle size than expected and thereby raised questions about the product efficacy" [20].

### 4.7 Microbiological Quality

***Microbial Specifications*** These specifications are determined by the manufacturer. The USP Chapters 61, 62, and 1111 present the microbial limits to assess the significance of microbial contamination in a dosage form [53]. However, the USP

does not determine specific methods for water-insoluble topical products. The microbial specifications are presented as a manufacturer's document that details the methods to isolate and identify the organisms as well as the number of organisms permitted and action levels to be taken when limits are exceed and the potential causes are investigated [6]. The *Pharmaceutical Microbiology Newsletter* (PMF) presents several articles to discuss topics such as microbial identification, methods, data analysis, and preservation as well as topics related to USP and FDA regulations [54]. To minimize the differences about microbial limits and test methods, the USP is trying to harmonize the standards with the European Pharmacopoeia (EP) [55].

***Microbial Test Methods*** The selected microbial test methods determine specific sampling and analytical procedures. When the product has a potential antimicrobial effect and/or preservative, the spread technique on microbial test plates must be validated. In addition, the personnel performing the analytical techniques have to be qualified and adequately trained for this purpose [6].

Usually, total aerobic bacteria, molds, and yeasts are counted by using a standard plate count in order to test the microbial limits. The microbial limit test may be customized by performing a screening for the occurrence of *Staphylococcus aureus, Pseudomonas aeruginosa, Pseudomonas cepacia, Escherichia coli, and Salmonella* sp. [56].

***Investigation of Exceeded Microbiological Limits*** A high number of organisms may indicate deficiencies in the manufacturing process, such as excessive high temperature, component quality, inadequate preservative system, and/or container integrity. Information about the health hazards of all organisms isolated from the product has different meanings depending on the type of dosage form and group of patients to be treated. For instance, in oral liquids, pseudomonads are usually a high-risk contamination. Examples presented by the FDA are Nystatin antifungal suspension, used as prophylaxis in AIDS patients [57]; antacids, with which *P. aeruginosa* contamination can promote gastric ulceration [58]; and the presence of *Pseudomona putida*, which could indicate the presence of other significant contaminants such as *P. aeruginosa* [6].

***Deionizer Water-Monitoring Program*** Deionizing systems must be controlled in order to produce purified water, required for liquid dosage forms and USP tests and assays [1]. The monitoring program has to include the manufacturer's documentation about time between recharging and sanitizing, microbial quality and chlorine levels of feed water, establishment of water microbial quality specifications, conductivity monitoring intervals, methods of microbial testing, action levels when microbial limits are exceeded, description of sanitization and sterilization procedures for deionizer parts, and processing conditions such as temperature, flow rates, use and sanitization frequency, and regenerant chemicals for ion exchange resin beds [6, 59].

***Effectiveness of Preservative*** Manufacturing controls and shelf life must ensure that the specified preservative level is present and effective as part of the stability program [6]. Depending on the type of product, the selection of the preservative system is based on different considerations, such as site of use, interactions,

spectrum, stability, toxicity, cost, taste, odor, solubility, pH, and comfort. The USP and other organizations describe methods to validate the preservative system used in the dosage form. Compounds used as preservatives are alcohols, acids, esters, and quaternary ammonium compounds, among others. For instance, to preserve ophthalmic liquid dosage forms, these products are autoclaved or filtrated and require an antimicrobial preservative to resist contamination throughout their shelf life, such as chlorobutanol, benzalkonium chloride, or phenylmercuric nitrate [1].

### 4.8 Filling and Packing

***Constant Mixing during Filling Process*** Due to the tendency of suspensions to segregate during transport through transfer lines, special attention is required on suspension uniformity during the filling process. Appropriate constant mixing of the bulk to keep homogeneity during the filling process and sampling of finished products and other critical points are indispensable conditions to assure an acceptable quality level during the filling and packing process [20].

***Mixing Low Levels of Bulk Near End of Filling Process*** Constant mixing during the filling process includes mixing low levels of bulk near the end of the filling process. Large-size batches of bulk suspension require the transfer of the residual material to a smaller tank in order to assure appropriate mixing of components before filling and packing the containers [20].

***Potency Uniformity of Unit-Dose Products*** Products manufactured have to be of quality at least as good as the established acceptable quality level (AQL). The quality level should be based on the limits specified by the USP. However, when the bulk product is not properly mixed during filling and packing processes, liquid dosage forms, and specially suspensions, are not homogeneous and unit-dose products contain very different amounts of the active component and potency. For these reasons, finished products have to be tested to assure that the final volume and/or weight as well as the amount of active ingredient are within the specified limits [6].

***Calibration of Provided Measuring Devices*** Measuring devices consist of droppers, spoons for liquid dosage forms, and cups labeled with both tsp and mL. Measuring devices have to be properly calibrated in order to assure the right amount of ingredients per volume to be administered [6].

***Container Cleanliness of Marketing Product*** The previous cleanliness of containers filled with the product will depend on their transportation exposure, composition, and storage conditions. Glass containers usually carry at least mold spores of different microorganisms, especially if they are transported in cardboard boxes. Other containers and closures made with aluminum, Teflon, metal, or plastic usually have smooth surfaces and are free from microbial contamination but may contain fibers or insects [45]. Some manufacturers receive containers individually wrapped to reduce contamination risks and others use compressed air to clean them. However, the cleanliness of wrapped containers will depend on the provider's guarantee of the manufacturing process and compressed-air equipment may release vapors or oils that have to be tested and validated [6].

### 4.9 Stability

The typical stability problems are color change, loss of active component, and clarity changes for solutions; inability to resuspend the particles and loss of significant amounts of the active component for suspensions; and creaming and breaking (or coalescence) for emulsions [1]. These instabilities are usually related to the following:

*Active and Primary Degradant.* A liquid dosage form is stable while it remains within its product specifications. When chemical degradation products are known, for stability study and expiration dating, the regulatory requirements for the primary degradant of a active component are chemical structure, biological effect and significance at the concentrations to be determined, mechanism of formation and order of reaction, physical and chemical properties, limits and methods for quantitating the active component and its degradant molecule at the levels expected to be present, and pharmacological action or inaction [45]. Examples of drugs in liquid dosage forms that are easily degraded are vitamins and phenothiazines [6].

*Interactions with Closure Systems.* Elastomeric and plastic container and closure systems release leachable compounds into the liquid dosage form, such as nitrosamines, monomers, plasticizers, accelerators, antioxidants, and vulcanizing agents [44]. Each type of container and closure with different composition and/or design proposed for marketing the drug or physician's samples has to be tested and stability data should be developed. Containers should be stored upright, on their side, and inverted in order to determine if container–closure interactions affect product stability [6, 45].

*Moisture Loss.* When the containers are inappropriately closed, part of the vaporized solvent is released and the concentration and potency of the active component may be increased [6].

*Microbiological Contamination.* Inappropriate closure systems also increase the possibilities of microbial contamination when opening and closing containers [6].

### 4.10 Process Validation

***Objective*** Process validation has the objectives of identifying and controlling critical points that may vary product specifications through the manufacturing process [6].

***Amount of Data*** To validate the manufacturing process, the manufacturer has to design and specify in the protocol the use of data sheets to keep information about the control of product specifications from each batch in-process as well as finished-product tests. Some formats are common to different products, though each type of product has some specific information to be kept on special sheets. Thus, the amount of data varies from one type of product to another [6].

***Scale-Up Process*** Data obtained using special batches for the validation of the scale-up process are compared with data from full-scale batches and batches used for clinical essays [6].

***Product Specifications*** The most important specifications or established limits for liquid dosage forms are microbial limits and test methods, medium pH, dissolution of components, viscosity, as well as particle size uniformity of suspended components and emulsified droplets. Effectiveness of the preservative system depends on the dissolution of preservative components and may be affected by the medium pH and viscosity. In addition, dissolved oxygen levels are important for components sensitive to oxygen and/or light [6].

***Bioequivalence or Clinical Study*** In the patient, the general or systemic circulation is responsible for carrying molecules to different tissues of the body. To assure the expected bioactivity of a product, the amount of drug that reaches the systemic circulation per unit of time is analyzed and is known as bioavailability. Bioequivalence is the comparison of the bioavailability of a product with a reference product. While oral solutions may not always need bioequivalence studies because they are considered self-evidente, suspensions usually require bioequivalence or clinical studies in order to demonstrate effectiveness. However, OTC suspension products such as antacids are exempt from these studies [6].

***Control of Changes to Approved Protocol*** The manufacturing process of a specific product is validated and approved internally by the quality control unit and externally by the FDA. Any change in the approved protocol has to be documented to explain the purpose and demonstrate that the change will not unfavorably affect product safety and efficacy. Factors include potency and/or bioactivity as well as product specifications. However, the therapeutic activity and uniformity of the product are the main concerns after formulation and process changes [20].

## 5 LIQUID DOSAGE FORMS*

*Douche* A liquid preparation, intended for the irrigative cleansing of the vagina, that is prepared from powders, liquid solutions, or liquid concentrates and contains one or more chemical substances dissolved in a suitable solvent or mutually miscible solvents.

*Elixir* A clear, pleasantly flavored, sweetened hydroalcoholic liquid containing dissolved medicinal agents; it is intended for oral use.

*Emulsion* A dosage form consisting of a two-phase system comprised of at least two immiscible liquids, one of which is dispersed as droplets (internal or dispersed phase) within the other liquid (external or continuous phase), generally stabilized with one or more emulsifying agents. (Note: Emulsion is used as a dosage form term unless a more specific term is applicable, e.g. cream, lotion, ointment.).

*Enema* A rectal preparation for therapeutic, diagnostic, or nutritive purposes.

*Extract* A concentrated preparation of vegetable or animal drugs obtained by removal of the active constituents of the respective drugs with a suitable menstrua, evaporation of all or nearly all of the solvent, and adjustment of the residual masses or powders to the prescribed standards.

*The definitions in this section are from ref. 11.

*For Solution* A product, usually a solid, intended for solution prior to administration.

*For Suspension* A product, usually a solid, intended for suspension prior to administration.

*For Suspension, Extended Release* A product, usually a solid, intended for suspension prior to administration; once the suspension is administered, the drug will be released at a constant rate over a specified period.

*Granule, Effervescent* A small particle or grain containing a medicinal agent in a dry mixture usually composed of sodium bicarbonate, citric acid, and tartaric acid which, when in contact with water, has the capability to release gas, resulting in effervescence.

*Inhalant* A special class of inhalations consisting of a drug or combination of drugs, that by virtue of their high vapor pressure can be carried by an air current into the nasal passage where they exert their effect; the container from which the inhalant generally is administered is known as an inhaler.

*Injection* A sterile preparation intended for parenteral use; five distinct classes of injections exist as defined by the USP.

*Injection, Emulsion* An emulsion consisting of a sterile, pyrogen-free preparation intended to be administered parenterally.

*Injection, Solution* A liquid preparation containing one or more drug substances dissolved in a suitable solvent or mixture of mutually miscible solvents that is suitable for injection.

*Injection, Solution, Concentrate* A sterile preparation for parenteral use which, upon the addition of suitable solvents, yields a solution conforming in all respects to the requirements for injections.

*Injection, Suspension* A liquid preparation, suitable for injection, which consists of solid particles dispersed throughout a liquid phase in which the particles are not soluble. It can also consist of an oil phase dispersed throughout an aqueous phase, or vice-versa.

*Injection, Suspension, Liposomal* A liquid preparation, suitable for injection, which consists of an oil phase dispersed throughout an aqueous phase in such a manner that liposomes (a lipid bilayer vesicle usually composed of phospholipids which is used to encapsulate an active drug substance, either within a lipid bilayer or in an aqueous space) are formed.

*Injection, Suspension, Sonicated* A liquid preparation, suitable for injection, which consists of solid particles dispersed throughout a liquid phase in which the particles are not soluble. In addition, the product is sonicated while a gas is bubbled through the suspension and these result in the formation of microspheres by the solid particles.

*Irrigant* A sterile solution intended to bathe or flush open wounds or body cavities; they're used topically, never parenterally.

*Linament* A solution or mixture of various substances in oil, alcoholic solutions of soap, or emulsions intended for external application.

*Liquid* A dosage form consisting of a pure chemical in its liquid state. This dosage form term should not be applied to solutions.

*Liquid, Extended Release* A liquid that delivers a drug in such a manner to allow a reduction in dosing frequency as compared to that drug (or drugs) presented as a conventional dosage form.

*Lotion* An emulsion, liquid dosage form. This dosage form is generally for external application to the skin.

*Lotion/Shampoo* A lotion dosage form which has a soap or detergent that is usually used to clean the hair and scalp; it is often used as a vehicle for dermatologic agents.

*Mouthwash* An aqueous solution which is most often used for its deodorant, refreshing, or antiseptic effect.

*Oil* An unctuous, combustible substance which is liquid, or easily liquefiable, on warming, and is soluble in ether but insoluble in water. Such substances, depending on their origin, are classified as animal, mineral, or vegetable oils.

*Rinse* A liquid used to cleanse by flushing.

*Soap* Any compound of one or more fatty acids, or their equivalents, with an alkali; soap is detergent and is much employed in liniments, enemas, and in making pills. It is also a mild aperient, antacid and antiseptic.

*Solution* A clear, homogeneous liquid dosage form that contains one or more chemical substances dissolved in a solvent or mixture of mutually miscible solvents.

*Solution, Concentrate* A liquid preparation (i.e., a substance that flows readily in its natural state) that contains a drug dissolved in a suitable solvent or mixture of mutually miscible solvents; the drug has been strengthened by the evaporation of its nonactive parts.

*Solution, for Slush* A solution for the preparation of an iced saline slush, which is administered by irrigation and used to induce regional hypothermia (in conditions such as certain open heart and kidney surgical procedures) by its direct application.

*Solution, Gel Forming/Drops* A solution, which after usually being administered in a drop-wise fashion, forms a gel.

*Solution, Gel Forming, Extended Release* A solution that forms a gel when it comes in contact with ocular fluid, and which allows at least a reduction in dosing frequency.

*Solution/Drops* A solution which is usually administered in a drop-wise fashion.

*Spray* A liquid minutely divided as by a jet of air or steam.

*Spray, Metered* A non-pressurized dosage form consisting of valves which allow the dispensing of a specified quantity of spray upon each activation.

*Spray, Suspension* A liquid preparation containing solid particles dispersed in a liquid vehicle and in the form of coarse droplets or as finely divided solids to be applied locally, most usually to the nasal-pharyngeal tract, or topically to the skin.

*Suspension* A liquid dosage form that contains solid particles dispersed in a liquid vehicle.

*Suspension, Extended Release* A liquid preparation consisting of solid particles dispersed throughout a liquid phase in which the particles are not soluble; the suspension has been formulated in a manner to allow at least a reduction in dosing frequency as compared to that drug presented as a conventional dosage form (e.g., as a solution or a prompt drug-releasing, conventional solid dosage form).

*Suspension/Drops* A suspension which is usually administered in a dropwise fashion.

*Syrup* An oral solution containing high concentrations of sucrose or other sugars; the term has also been used to include any other liquid dosage form prepared in a sweet and viscid vehicle, including oral suspensions.

*Tincture* An alcoholic or hydroalcoholic solution prepared from vegetable materials or from chemical substances.

*Notes*:

1. A liquid is pourable; it flows and conforms to its container at room temperature. It displays Newtonian or pseudoplastic flow behavior.
2. Previously the definition of a lotion was "The term lotion has been used to categorize many topical suspensions, solutions, and emulsions intended for application to the skin." The current definition of a lotion is restricted to an emulsion.
3. A semisolid is not pourable; it does not flow or conform to its container at room temperature. It does not flow at low shear stress and generally exhibits plastic flow behavior.
4. A colloidal dispersion is a system in which particles of colloidal dimension (i.e., typically between 1 nm and 1 μm) are distributed uniformly throughout a liquid.

## REFERENCES


1. Crowley, M. M. (2005), Solutions, emulsions, suspensions, and extracts, in USIP, *Remington: The Science and Practice of Pharmacy*, Lippincott Williams & Wilkins, Philadelphia, pp. 745–774.
2. Block, L. H. (2002), Nonparenteral liquids and semisolids, in Levin, M., Ed., *Pharmaceutical Process Scale-Up*, Marcel Dekker, New York, pp. 57–94.
3. Spurgeon, T. (2007), Quality by design in solid dosage processes: How will QbD impact manufacturing? *Contract Pharm.*, 3–1.
4. Allen, L. V. *The art, Science, and Technology of Pharmaceutical Compounding*. American Pharmaceutical Association (APhA). Washington D.C., 2002. Pages 93, 95.
5. Qu, A., and Maglayo, A. (2003), Outsourcing stability services. *Contract Pharma*. 11–12. Available at: http://www.contractpharma.com/articles/2003/11/outsourcing-stability-services
6. *USA Department of Health and Human Services: Food and Drug Administration (FDA)*. Guides to inspections. Available at http://www.fda.gov/ora/Inspect_ref/igs/iglist.html

7. Kourounakis, P. N. (1994), *Advanced Drug Design and Development*, Taylor and Francis, p. 141.
8. Papich, M. G. (2005), Drug compounding for veterinary patients, *AAPS J.*, 7(2), E281–E287.
9. Wiberg, C. C., Leppik, I. E., and Cloyd, J. C. (2005), Lower phenytoin serum levels in persons switched from brand to generic phenytoin, *Neurol. J.*, 64(8), 1485–1486.
10. Brooks, K. (2007), Pharma IT outsourcing: ERP consolidates IT infrastructure, *Contract Pharm*, March 2007, available: http://www.contractpharma.com/articles/2007/03/pharma-it-outsourcing
11. *USA Department of Health and Human Services, Food and Drug Administration (FDA)*. Center for Drug Evaluation and Research (CDER) Data Standards Manual: Dosage Form (4/92) available at: http://www.fda.gov/cder/dsm/DRG/drg00201.htm.
12. Drugs@FDA glossary of terms, U.S. Department of Health and Human Services, Food and Drug Administration (FDA), available: http://www.fda.gov/cder/drugsatfda/glossary.htm#L.
13. DeLuca, P. P. (1992), Formulation of small volume parenterals, in Avis, K. E., Lieberman, H. A., and Lachman, L., Eds., *Pharmaceutical Dosage Forms: Parenteral Medications*, Vol. 1, Marcel Dekker, New York.
14. Allen, L. V., Popovich, N. G., and Ansel, H. C., *Ansel's Pharmaceutical Dosage Forms and Drug Delivery Systems*, Lippincott William & Wilkins, Baltimore, 2004, Chapters 2 and 5.
15. SAP Information Center, Best practices for automotive: Pharmaceutical industry, Liquid dosage forms, available: http://help.sap.com/.
16. Sokoll, K. (2007), Optimizing drug development strategies: Exploring methods to improve the process, *Contract Pharm*. Marzo 2007. Available: http://www.contractpharma.com/articles/2007/03/optimizing-drug-development-strategies
17. Carleton, F. J., and Agalloco, J. P. (1999), *Validation of Pharmaceutical Processes: Sterile Products*, Marcel Dekker, New York.
18. Stacey, R. D., Griffin, D., and Shaw, P. (2000), *Complexity and Management: Fad or Radical Challenge to Systems Thinking?* Taylor and Francis Group, Boca Raton, Fl.
19. Hardman, J. G., Limbrid, L. E., Molinoff, P. B., and Ruddon, R. W. *Goodman and Gilman's The Pharmacological Basis of Therapeutics*, McGraw-Hill, New York, 2006.
20. Scott, R. R. (1989), A practical guide to equipment selection and operating techniques, in Lieberman, H. A., Rieger, M. M., and Banker, G. S., Eds., *Pharmaceutical Dosage Forms: Disperse Systems*, Vol. 2, Marcel Dekker, New York, pp 1–71.
21. Reepmeyer, J. C., Revelle, L. K., and Vidavsky, I. (1998), Detection of clobetasol propionate as an undeclared steroid in zinc pyrithione formulations by high-performance liquid chromatography with rapid-scanning ultraviolet spectroscopy and mass spectrometry. *Journal of Chromatography A*, Vol. 828, Number 1, 18 December 1998, pp. 239–246(8).
22. FARR Air Pollution Control (2005), FARR dust collectors, HEPA BIBO filtration systems, Pharmaceutical Industry Standard On-line, available: http://www.pharmaceuticalonline.com/downloads/detail.aspx?docid=74B77B1C-0ED9-4E7B-A317-5DD0D23124B3&VNETCOOKIE=NO.
23. Biotrace International (2004), Air samplers. LIT055/002/29.06.04, available: http://www.biotrace.co.uk/uploads/documents/103.pdf.
24. 3-A Sanitary Standards, Technical Resource Center, resource papers, available: http://www.3-a.org/techresource/papers.htm.
25. Cook, Jr., J. L. (1998), *Standard Operating Procedures and Guidelines*, PennWell Books, Tulsa, OK, p. 318.

26. Guides to inspections: Emulsions, U.S. Department of Health and Human Services, Food and Drug Administration (FDA), available: http://www.fda.gov/ora/Inspect_ref/igs/iglist.html.
27. Guides to inspections: Validation of cleaning processes, U.S. Department of Health and Human Services, Food and Drug Administration (FDA), available: http://www.fda.gov/ora/Inspect_ref/igs/valid.html.
28. Thierauf, R. J., and Hoctor, J. J. (2006), *Optimal Knowledge Management: Wisdom Management Systems Concepts and Applications*, Idea Group, Hershey, PA.
29. Petruševski, V. M., and Metodija, Z. N. (2001), Volume non-additivity of liquid mixtures: Modifications to classical demonstrations, *Chem. Ed. J.*, 6, 3.
30. Zhou, Q., and Milton, R. (2003), Molecular interactions of surfactants in mixed monolayers at the air/aqueous solution interface and in mixed micelles in aqueous media: The regular solution approach, *Am. Chem. Soc.*, 19(11), 4555–4562.
31. CANTY Process Technology (2007), CANTY Lighting Systems, available: http://www.jmcanty.com/images/overview/16pagermaster.pdf.
32. Wikipedia, Non-Newtonian fluid definition, available: http://en.wikipedia.org/wiki/Non-Newtonian_fluid.
33. Lieberman, H. A., Rieger, M. M., and Banker, G. S. (1988), *Pharmaceutical Dosage Forms: Disperse Systems*, Vol. 1, Marcel Dekker, New York.
34. Gibson, M., *Pharmaceutical Preformulation and Formulation: A Practical Guide from Candidate Drug Selection to Commercial Dosage Form*, Interpharm, Boca Raton, FL, 2004, pp. 581.
35. Davidow, L. W., Davidow, L., and Thompson, J. E., *A Practical Guide to Contemporary Pharmacy Practice*, Lippincott Williams & Wilkins, 2003. pp. 704.
36. Sinko, P. J. *Martin's Physical Pharmacy and Pharmaceutical Sciences*, Lippincott Williams & Wilkins, Baltimore, MD., 2005, pp. 795. 47.
37. Nusim, S. (2005), *Active Pharmaceutical Ingredients: Development, Manufacturing, and Regulation*, Taylor and Francis Group, Boca Raton, FL.
38. Chongprasert, S., and Nail S. L. (1998), Influence of drug polymorphism on the physical chemistry of freeze-drying, Digital dissertations and theses, Purdue e-pubs, *ETD Collection for Purdue University*, available: http://docs.lib.purdue.edu/dissertations/AAI9939331/.
39. Borka, L. (1971), The stability of chloramphenicol palmitate polymorphs. Solid and solution phase transformations, *Acta Pharm. Suec.*, 9–8(4), 365–372.
40. Rosenholm, J. B. (2007), Wetting of surfaces and interfaces: A conceptual equilibrium thermodynamic approach, in *Colloid Stability: The Role of Surface Forces, Part II*, Tharwat F. Tadros Colloids and Interface Science Series, Vol. 2, Wiley-VCH, Weinheim.
41. Lerk, C. F., Lagas, M., Lie-a-Huen, L., Broersma, P., and Zuurman, K. (1979), In vitro and in vivo availability of hydrophilized phenytoin from capsules, *J. Pharm. Sci.*, 5–68(5), 634–638.
42. Rosoff, M. (1988), Specialized pharmaceutical emulsions, in Lieberman, H. A., Rieger, M. M., and Banker, G. S., Eds., *Pharmaceutical Dosage Forms: Disperse Systems*, Vol. 1, Marcel Dekker, New York, pp. 245–283.
43. Moorthaemer, B., Sprakel, J. (2006) Improving the stability of a suspension. Pharmaceutical Technology Europe, O1 February 2006. Available: http://www.ptemag.com/pharmtecheurope/article/articleDetail.jsp?id=306687
44. Allen, T. (1997), *Particle Size Measurement*, Chapmann and Hall, London.
45. Connors, K. A., Amidon G. L., and Stella V. J. (1986), *Chemical Stability of Pharmaceuticals: A Handbook for Pharmacists*, Wiley-Interscience, New York.

46. Miller, S. C., and Drabik, B. R. (1984), Rheological properties of poloxamer vehicles, *Int. J. Pharm.*, 18, 269.
47. Millard, J. W., Alvarez-Nunez, F. A., and Yalkowsky, S. H. (2002), Solubilization by cosolvents—Establishing useful constants for the log-linear model, *Int. J. Pharm.*, 245(1), 153–166 (14).
48. Nash, R. (1996), Pharmaceutical suspensions, in Lieberman, H. A., Rieger, M. M., and Banker, G. S., Eds., *Pharmaceutical Dosage Forms: Disperse Systems*, Vol. 2, Marcel Dekker, New York, pp. 1–46.
49. Adeyeye, C. M., and Price, J. C. (1997), Chemical dissolution stability and microscopic evaluation of suspensions of ibuprofen and sustained release ibuprofen-wax microspheres, *J. Microencapsul.*, 14(3), 357–377.
50. Fong-Spaven, F., and Hollenbeck, R. G. (1986), Thermal rheological analysis of triethanolamine-stearate stabilized mineral oil in water emulsions, *Drug Dev. Ind. Pharm.*, 12, 289.
51. Moelwyn-Hughes, E. A. (1961), *Physical Pharmacy*, Pergamon, New York, pp. 297.
52. Lan, C. W. (2005), Flow and segregation control by accelerated rotation for vertical Bridgman growth of cadmium zinc telluride: ACRT versus vibration, *J. Crystal Growth*, 274 (3–4), 379–386.
53. USP 29: U.S. Pharmacopeia and the National Formulary (USP 29/NF 24), United States Pharmacopeial Convention, Rockville, MD. 2006.
54. PMF Pharmaceutical Microbiology Forum Newsletters, available: http://www.microbiologyforum.org/news.htm.
55. Sutton, S. (2006), The harmonization of the microbial limits test—Enumeration, *Microbiol. Net.—PMF*, 12(3), 2–3.
56. Carstensen, J. T., and Rhodes, C. T. (2000), *Drug Stability: Principles and Practices*, Macel Dekker, New York, pp. 773.
57. Robinson, E. P. (2006), Pseudomonas aeruginosa contamination of liquid antacids: A survey, *J. Pharm. Sci.*, 60(4), 604–605.
58. Steer, H. W., and Colin-Jones, D. G. (1975), Mucosal changes in gastric ulceration and their response to carbenoxolone sodium. *Gut BMJ*, 16, 590–597.
59. Dreeszen, P. (2003), *Biofilm: Key to Understanding and Controlling Bacterial Growth in Automated Drinking Water Systems*, Edstrom Industries, Waterford, WI, pp. 19.

# CONTROLLED-RELEASE DOSAGE FORMS

Anil Kumar Anal
*Living Cell Technologies (Global) Limited, Auckland, New Zealand*

**Contents**

1 Introduction
2 Rationale
3 General Design Principle for Controlled-Release Drug Delivery Systems
4 Physicochemical and Biological Factors Influencing Design and Performance of Controlled-Release Formulations
  4.1 Physicochemical Factors
  4.2 Biological Factors
5 Controlled-Release Oral Dosage Forms
  5.1 Anatomical and Physiological Considerations
  5.2 Fundamentals of Controlled-Release Oral Dosage Forms
  5.3 Factors Influencing Oral Controlled-Release Dosage Forms
6 Design and Fabrication of Controlled-Release Dosage Forms
  6.1 Microencapsulation
  6.2 Nanostructure-Mediated Controlled-Release Dosage Forms
  6.3 Liposomes
  6.4 Niosomes
7 Technologies for Developing Transdermal Dosage Forms
8 Ocular Controlled-Release Dosage Forms
9 Vaginal and Uterine Controlled-Release Dosage Forms
10 Release of Drugs from Controlled-Release Dosage Forms
  10.1 Time-Controlled-Release Dosage Forms
  10.2 Stimuli-Induced Controlled-Release Systems
11 Summary
  References

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

## 1 INTRODUCTION

Therapeutic value and pharmaeconomic value have in recent years become major issues in defining health care priorities under the pressure of cost containment [1]. The improvement in drug therapy is a consequence of not only the development of new chemical entities but also the combination of active substances and a suitable delivery system. The treatment of an acute disease or chronic illness is mostly accomplished by delivery of one or more drugs to the patient using various pharmaceutical dosage forms. Tablets, pills, capsules, suppositories, creams, ointments, liquids, aerosols, and injections are in use as drug carriers for many decades. These conventional types of drug delivery systems are known to provide a prompt release of the drug. Therefore, to achieve as well as to maintain the drug concentration within the therapeutically effective range needed for treatment, it is often necessary to take this type of drug several times a day, resulting in the significant fluctuation in drug levels [2]. For all categories of treatment, a major challenge is to define the optimal dose, time, rate, and site of delivery. Recent developments in drug delivery techniques make it possible to control the rate of drug delivery to sustain the duration of therapeutic activity and/or target the delivery of drug to a specific organ or tissue. Many investigations are still going on to apply the concepts of controlled delivery for a wide variety of drugs [3].

## 2 RATIONALE

The basic rationale for controlled drug delivery is to alter the pharmacokinetics and pharmacodynamics of pharmacologically active moieties by using novel drug delivery systems or by modifying the molecular structure and/or physiological parameters inherent in a selected route of administration. It is desirable that the duration of drug action become more a design property of a rate-controlled dosage form and less, or not at all, a property of the drug molecules' inherent kinetic properties.

The rationale for development and use of controlled dosage forms may include one or more of the following arguments [4]:

- Decrease the toxicity and occurrence of adverse drug reactions by controlling the level of drug and/or metabolites in the blood at the target sites.
- Improve drug utilization by applying a smaller drug dose in a controlled-release form to produce the same clinical effect as a larger dose in a conventional dosage form.
- Control the rate and site of release of a drug that acts locally so that the drug is released where the activity is needed rather than at other sites where it may cause adverse reactions.
- Provide a uniform blood concentration and/or provide a more predictable drug delivery.
- Provide greater patient convenience and better patient compliance by significantly prolonging the interval between administrations.

However, there are also disadvantages attached to the use of controlled-release dosage forms. These include higher cost of manufacturing, unpredictability, poor in

vitro/in vivo correlation, reduced potential, and poor systemic availability in general and the effective release period is influenced and limited by the gastrointestinal (GI) residence time [5]. The transit time of a dosage form through the GI tract is dependent on the physical characteristics of the formulation as well as on physiological factors such as stomach emptying time and effect of food on the absorption process.

Only drugs with certain properties are suitable for controlled-release dosing. Characteristics that may make a drug unsuitable for controlled-release dosing include a long or short elimination half-life, a narrow therapeutic index, a large dose, low/slow solubility, extensive first-pass clearance, and time course of circulating drug levels different from that of the pharmacological effect. The ideal drug delivery should be inert, biocompatible, mechanically strong, comfortable for the patient, capable of achieving high drug loading, simple to administer, and easy to fabricate and sterilize [6]. A range of materials have been employed to control the release of drugs and other active substances. Controlled-release dosage forms have been developed for over four decades. One of the first practically used controlled-release oral dosage forms was the Spansule capsule, which was introduced in the 1950s. Spansule capsules were manufactured by coating a drug onto nonpareil particles and further coating with glyceryl stearate and wax. Subsequently, ion exchange resins were proposed for application as sustained-release delivery systems of accessible drug. Since then numerous products have been introduced and commercialized.

## 3 GENERAL DESIGN PRINCIPLE FOR CONTROLLED-RELEASE DRUG DELIVERY SYSTEMS

In the drug delivery system, the pharmacodynamics of active molecules becomes more a function of design and less one of inherent kinetic properties. Therefore, a deep understanding of the design of controlled-release systems of the pharmacokinetics and pharmacodynamics of the drug is required [7]. The conventional tablet or capsule provides only a single and transient burst of drug. A modification introduced to the molecular structure of the drug (often used to decrease the elimination rate) or a system for modified release rate is the common approaches used to increase the interval between two doses. The objective of both these approaches is to decrease the fluctuations in plasma levels during multiple dosing. This allows the dosing interval to increase without compromising the required dosage levels. If the half-life of a drug is less than 6 h or the passage time in the smaller intestinal track is decreased, there might not be enough time to allow proper absorption, thus making frequent dosing compulsory. For other routes, where the residence time is not a constraint, dosing intervals can be as long as months or even years.

A controlled-release drug delivery system serves primarily two functions [8]. First, it involves the transport of the drug to a particular part of the body. This may be accomplished in two ways, parenterally and nonparenterally. Second, the release of active ingredients occurs in a controlled manner, depending on the preparation of dosage forms. This determines the rate at which a drug is made available to the body once it has been delivered. Controlled drug delivery occurs when a biomaterial, either natural or synthetic, is judiciously combined with a drug or other active

agent in such a way that the active agent is released from the material in a predesigned manner. To be successfully used in controlled drug delivery formulations, a material must be chemically inert and free of leachable impurities. It must also have an appropriate physical structure, with minimal undesired aging, and be readily processable.

Controlled-release systems provide numerous benefits over conventional dosage forms. Conventional dosage forms are not able to control either the rate of drug delivery or the target area of administration and provide an immediate or rapid drug release. This necessitates frequent administration in order to maintain a therapeutic level. As a result, as shown in Figure 1, drug concentrations in the blood fluctuate widely. The concentrations of drug remain at a maximum value, which may represent a toxic level, or a level at which undersized side effects might occur, and a minimum value, below which the drug is no longer effective. The duration of therapeutic efficacy is dependent upon the frequency of administration, the half-life of the drug, and the release rate of dosage forms. In contrast, controlled-release dosage forms not only are able to maintain therapeutic levels of drug with narrow fluctuations but also make it possible to reduce the frequency of drug administration. The drug concentrations, as shown in Figure 1, released from controlled-release dosage forms fluctuate within the therapeutic range over a longer period of time. The plasma concentration profile depends on the preparation technology, which may generate different release kinetics, resulting in different pharmacological and pharmacokinetic responses in the blood or tissues.

The primary objectives of controlled drug delivery are to ensure safety and to improve efficacy of drugs as well as patient compliance. This is achieved by better control of plasma drug levels and less frequent dosing. For conventional dosage forms, only the dose ($D$) and dosing interval ($\tau$) can vary above which undesirable or side effects are elicited. As an index of this window, the therapeutic index (TI) can be used. This is often defined as the ratio of lethal dose ($LD_{50}$) to median effective dose ($ED_{50}$). Alternatively, it can be defined as the ratio of maximum drug concentration ($C_{max}$) in blood that can be tolerated to the minimum concentration ($C_{min}$) needed to produce an acceptable therapeutic response.

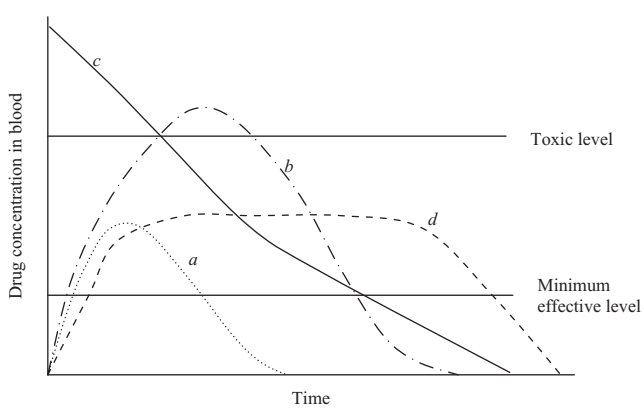


**FIGURE 1** Theoretical plasma concentration after administration of various dosage forms: ($a$) standard oral dose; ($b$) oral overdose; ($c$) IV injection; ($d$) controlled-release system.

Different types of modified release systems can be defined [4, 8]:

- Sustained release (extended release) that permits a reduction in dosing frequency as compared to the situation in which the drug is presented as a conventional form
- Delayed release when the release of the active ingredient comes sometimes other than promptly after administration
- Pulsatile release when the device actively controls the dosage released following predefined parameters

In general, the sustained-release dosage form is designed to maintain therapeutic blood or tissue levels of the drug for an extended period of time. This is accomplished by attempting to obtain zero-order release from the designed dosage form. Zero-order release constitutes drug release from the dosage form that is independent of the amount of drug in the delivery system at a constant release rate. Systems that are designed for prolonged release can also be attributed as achieving sustained-release delivery systems. Repeat-action tablets are an alternative method of sustained release in which multiple doses of drugs are contained within a dosage form and each dose is released at a periodic interval, while delayed-release systems may not be sustaining, since often the function of these dosage forms is to maintain the drug within the dosage form.

## 4 PHYSICOCHEMICAL AND BIOLOGICAL FACTORS INFLUENCING DESIGN AND PERFORMANCE OF CONTROLLED-RELEASE FORMULATIONS

A number of variables, such as drug properties including stability, solubility, partitioning characteristics. charge and protein binding behavior, routes of drug delivery, target sites, acute or chronic therapy, the disease, and the patient, must be considered to establish the criteria for designing controlled-release products [9]. The performance of a drug in its release pattern from the dosage form as well as in the body proper is a function of its properties. These properties can at times prohibit placement if the drug is in a controlled-release form, restrict the route of drug administration, and significantly modify performance for one reason or another. There is no clear distinction between physicochemical and biological factors since the biological properties of a drug are a function of its physicochemical properties while biological properties result from typical pharmacokinetic studies on the absorption, distribution, metabolism, and excretion (ADME) characteristics of a drug as well as those resulting from pharmacological studies.

### 4.1 Physicochemical Factors

Physicochemcial properties, such as aqueous solubility, partition coefficient and molecular size, drug stability, and protein binding, are those that can be determined from in vitro experiments.

***Ionization, p$K_a$, and Aqueous Solubility*** Most drugs are weak acids or bases. It is important to note the relationship between the p$K_a$ of the compound and the absorptive environment. Delivery systems that are dependent on diffusion or dissolution will likewise be dependent on the solubility of drug in the aqueous media. Since drugs must be in solution before they can be absorbed, compounds with very low aqueous solubility usually have the oral bioavailability problems because of limited GI transit time of the undissolved drug particles and they are limited at the absorption site. Unfortunately, for many of the drugs and bioactive compounds, the site of maximum absorption occurs at the site where solubility of these compounds is least.

The drug (e.g., tetracycline) for which the maximum solubility is in the stomach but high absorption takes place in the intestinal region may be poor candidates for controlled-release systems, unless the system is capable of retaining the drug in the stomach and gradually releasing it to the small intestine or unless the solubility is made higher and independent of the external environment by encapsulating those compounds in a membrane system. Other compounds, such as digoxin [10], with very low solubility, are inherently sustained, since their release over the time course of a dosage form in the gastrointestinal tract is limited by dissolution of the drug. Although the action of a drug can be prolonged by making it less soluble, this may occur at the expense of consistent and incomplete bioavailability.

The choice of mechanism for oral sustained/controlled-release systems is limited by the aqueous solubility of the drug. Thus, diffusional systems are poor choices for low aqueous-soluble drugs since the driving force for diffusion, the concentration in aqueous solution, will be low. The lower limit for the solubility of a drug to be formulated in a controlled-release system has been reported to be 0.1 mg/mL [11].

***Partition Coefficient and Molecular Size*** Following administration, drugs and other bioactive compounds must traverse a variety of membranes to gain access to the target area. The partition coefficient and molecular size influence not only the permeation of drug across biological membranes but also diffusion across or through a rate-controlling membrane or matrix. The partition coefficient is generally defined as the ratio of the fraction of drug in an oil phase to that of an adjacent aqueous phase. Drugs with extremely high partition coefficient (i.e., those that are highly oil soluble) readily penetrate the membranes but are unable to proceed further, while the excessive high aqueous-soluble compounds, having low oil/water partition coefficients, cannot penetrate the membranes. A balance in the partition coefficient is needed to give an optimum flux for permeation through the biological and rate-controlling membranes. The ability of drugs to diffuse through membranes, also known as diffusivity, is related to its molecular size by the following equation:

$$\log D = -s_V \log V + k_V = -s_{\mathrm{M}} \log M + k_{\mathrm{M}}$$

where $D$ is the diffusivity, $M$ is the molecular weight, $V$ is the molecular volume, and $s_V$, $s_{\mathrm{M}}$, $k_V$, and $k_{\mathrm{M}}$ are constants in a particular medium. Generally, there is smaller diffusivity with the denser medium.

***Drug Stability*** The stability of drug in the environment where it is to be exposed is an essential physicochemical factor to be considered before designing controlled

dosage forms [12]. For example, orally administered drugs are subjected to both acid–base hydrolysis and enzymatic degradation [13]. For drugs that are unstable in the stomach, the dosage forms can be designed in so that they can be placed in a slowly soluble form or have their release delayed until they reach the intestine. This type of approach can be ineffective and the drug may be unstable in the small intestine or undergo extensive gut-wall metabolism. To obtain better bioavailability for such types of drugs, which are unstable even in the intestine, a different route of administration (e.g., transdermal with controlled-release dosage forms) can be a better option [14]. A transdermal patch of nitroglycerin is a good example. The details for transdermal dosage forms will be described later in this chapter.

### 4.2 Biological Factors

A drug, being a chemical/biological agent or a mixture of chemical and biological agents, is recognized as a xenobiotic by the human body. Subsequently, the drug will be prevented from entering the body and/or eliminated after its entry. As a result, the defense mechanisms of the human body become barriers to the delivery of drugs. A drug may encounter physical, physiological, enzymatic, or immunological barriers on its way to the site of action. Hence, the design of controlled-release product should be based on a comprehensive picture of drug disposition. This would entail a complex examination of the ADME characteristics of the drug. The details of these effects on various controlled-release dosage systems will be given in the following sections of this chapter.

## 5 CONTROLLED-RELEASE ORAL DOSAGE FORMS

Oral drug delivery is the preferred route for drug administration because of its convenience, economy, and high patient compliance compared with several other routes. About 90% of the drugs are administered via the oral route [15, 16]. For the oral controlled administration of drugs, several research and development activities have shown encouraging signs of progress in the development of programmable controlled-release dosage forms as well as in the search for new approaches to overcome the potential problems associated with oral drug administration. Many oral drugs are perceived as "patient friendly" for compliance, often requiring that the medication only needs to be taken once a day. The most prescribed drugs in the United States that use oral drug delivery technologies include Lipitor (atorvastatin calcium), manufactured by Pfizer, and AstraZeneca's Toprol-XL (metoprolo succinate). These potential developments and recently developed approaches are discussed here along with an overview of GI physiology.

### 5.1 Anatomical and Physiological Considerations

Anatomically, the alimentary canal can be divided into a conduit region and digestive and absorptive regions. The conduit region includes the mouth, pharynx, esophagus, and lower rectum. The digestive and absorptive regions include the stomach, small intestine, and all parts of the large intestine except the very distal region.

**TABLE 1 Gastrointestinal Tract: Physical Dimensions and Dynamics**

| | | | Transit Time | |
|---|---|---|---|---|
| Region | Surface area ($m^2$) | pH | Fluid | Solid |
| Stomach | 0.1–0.2 | 1.2 | 50 min | 8 h |
| Small intestine | 100 | 6.8 | 2–6 h | 4–9 h |
| Large intestine | 0.5–1 | 7.5 | 2–6 h | 3–72 h |

The role of the stomach in drug and nutrition absorption is very limited, and it acts primarily as a reception area for oral dosage forms. Nonionic, lipophilic molecules of moderate size can be absorbed through the stomach only to a limited extent owing to the small epithelial surface area and the short duration of contact with the stomach epithelium in comparison with the intestine [17]. The transit time in the GI tract varies from one person to another and also depends upon the physical properties of the object ingested and the physiological conditions of the alimentary canal (Table 1). After passing through the stomach, the next organ that a drug or bioactive compound encounters is the small intestine. The intestinal epithelium is composed of absorptive cells (enterocytes) interspersed with goblet cells (specialized for mucus secretion) and a few enteroendocrine cells (that release hormones). The enterocytes of intestinal epithelium are the most important cells in view of the absorption of drugs and nutrients [18]. Histologically, colonic mucosa resembles the small intestinal mucosa, the absence of villi being the major difference [19]. The microvilli of the large intestine enterocytes are less organized than those of the small intestine. The resulting decrease in the surface area of the colon leads to a low absorption potential in comparison with the small intestine. However, the colonic residence time is longer than that for the small intestine, providing extended periods of time for the slow absorption of drugs and nutrients [20]. Figure 2 shows the various physiological processes encountered by an orally administered drug during the course of GI transit.

### 5.2 Fundamentals of Controlled-Release Oral Dosage Forms

Oral controlled drug delivery is a system that provides the continuous delivery of drugs at predictable and reproducible kinetics for a predetermined period throughout the course of GI transit [21]. Also included are systems that target the delivery of a drug to a specific region within the gastrointestinal tract (GIT) for either local or a systemic action. All the oral controlled drug delivery systems have limited utilization in the GI controlled administration of drugs if the systems cannot remain in the vicinity of the absorption site for the lifetime of the drug delivery. In the exploration of oral controlled-release dosage forms, one encounters three areas of potential challenges [22]:

1. *Drug Delivery System* To develop a viable oral controlled-release drug delivery system capable of delivering a drug at a therapeutically effective rate to a desirable site for the duration required for optimal efficacy.
2. *Modulation of GI Transit Time* To modulate the GI transit time so that the drug delivery system developed can be transported to a target site or to the

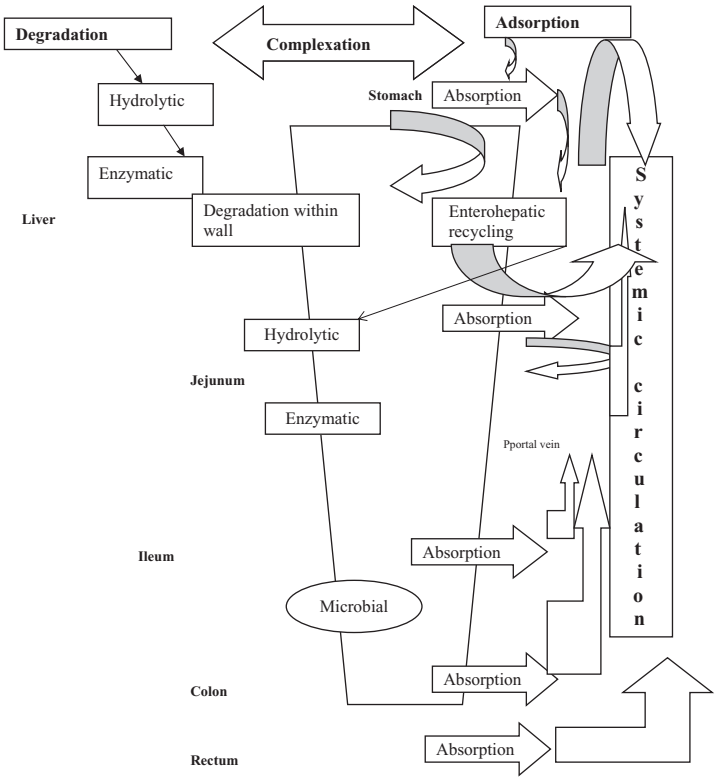


**FIGURE 2** Physical model illustrating various physiological processes during gastrointestinal transit.

vicinity of an absorption site and reside there for a prolonged period of time to maximize the delivery of a drug dose.

3. *Minimization of Hepatic First-Pass Elimination* If the drug to be delivered is subjected to extensive hepatic first-pass elimination, preventive measures should be devised to either bypass or minimize the extent of hepatic metabolic effect.

With most orally administered drugs, targeting is not a primary concern, and it is usually intended for drugs to permeate to the general circulation and perfuse to other body tissues, except it is medicated intentionally for localized effects in the GIT. There is a general assumption that increasing concentration at the absorption site will increase the rate of absorption and, therefore, increase circulating blood levels, which in turn promotes greater concentrations of drug at the site of action. If toxicity is not an issue, therapeutic levels can thus be extended as shown in Figure 3. In essence, "drug delivery" by these systems usually depends on release from specific types of dosage forms and permeation through an epithelial membrane to the blood. Still, the biological and physicochemical factors that come across play important roles in the design of such systems. The physicochemical properties have been described earlier in this chapter while biological factors involved with oral dosage forms will be described below.

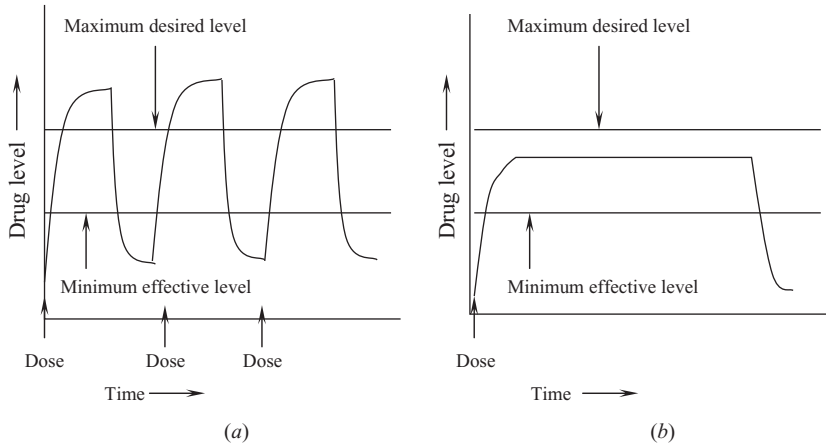


**FIGURE 3** Profile of drug level in blood: (*a*) traditional dosing of tablets; (*b*) controlled drug delivery dose.

The degree to which a delivery system can achieve standard release profiles for a variety of chemically and physically diverse, pharmaceutically active molecules is a measure of a delivery system's efficacy and flexibility (Figure 3). Among the most challenging profiles, linear, zero-order release of highly soluble actives over a 12–24-h period could be considered a reasonable performance standard against which delivery systems may be judged.

### 5.3 Factors Influencing Oral Controlled-Release Dosage Forms

***Biological Half-Life*** The usual goal of an oral controlled-release dosage form is to maintain therapeutic blood levels, as shown in Figure 3, over an extended period of time. A drug must be absorbed and enter the circulation at approximately the same rate at which it is eliminated. The elimination rate is quantitatively described by the half-life ($t_{1/2}$). Each drug has its own characteristic elimination rate, which is the sum of all elimination process, including metabolism, urinary excretion, and all other processes that permanently remove drug from the bloodstream.

Therapeutic compounds with short half-lives are excellent candidates for controlled/sustained-release preparations, since this can reduce dosing frequency [23]. In general, drugs with half-lives shorter than 2 h, such as furosemide or levodopa, are poor candidates for controlled-release preparations. Compounds with longer half-lives, such more than 8 h, also do not need to be in the form of controlled release, since their effect is already sustained. Digoxin, warfarin and phenytoin are some examples [24–26]. However, drugs having even longer half-lives can be used in other forms of modified release, such as pulsatile release.

***Gastrointestinal Tract and Absorption*** The design of a controlled-release dosage form should be based on a comprehensive picture of drug disposition. Both the pharmacokinetic property and biological response parameter have a useful range for the design of sustained- and controlled-release products. The potential problems inherent in oral controlled-release oral dosage forms generally relate to (i) interac-

tions between the rate, extent, and location that the dosage form releases the drug and (ii) the regional differences in GI physiology [27].

Total GI transit time in the normal population varies from 5 to 36 h, with an average total transit time of approximately 24 h [28]. There is still much more to explore about the physiological processes involved and factors that influence gastric emptying, intestinal transit, and colon residence. One of the major factors is food administration, which delays the "housekeeper wave," causing delay in gastric emptying. A high-fat meal may delay gastric emptying from 3 to 5 h, and the total GI transit delay is largely a function of delay in gastric emptying in this case. If we presume that the transit time of most drugs and devices in the absorptive areas of the GI tract is about 8–12 h, the maximum half-life for absorption should be approximately 3–4 h; otherwise, the device will pass out of the potential absorptive regions before drug release is complete. That many controlled-release products are having somewhat lesser bioavailability than their conventional dosage forms may be due to incomplete release of the dosage form or release at such a slow rate that the drug has passed the actual site of absorption. Compounds that demonstrate a true lower absorption rate constant will probably be poor candidates for controlled-release dosage forms [28].

An understanding of the behavior of dosage forms in the stomach has been gained largely from scintographic studies in which phases of a meal and formulations are labeled with different nucleotides, particularly technetium-99 and indium-111 [29]. Such studies have demonstrated that retention times of formulations in the stomach are dependent on the size of formulations and whether or not the formulation is taken with a meal. Enteric-coated or enteric matrix tablets may be retained for a considerable time if dosed with heavy meals or breakfast. Multiparticulate dosage forms will empty more slowly in the presence of food, and because the dosage forms mix evenly with the food, the entry into the small intestine will be strongly influenced by the caloric density and bulk of the ingested meal. The rate of gastric emptying, therefore, predicts the absorption behavior [29, 30]. However, the absorption of drugs from small, soft gelatin capsules is sometimes less predictable [31].

In the small intestine, contact time with the absorptive epithelium is limited, and the small intestinel transit time is 3.5–4.5 h [32]. The scintographic data show that many prolonged release products, particularly those intended for twice- or once-daily administration, actually release some of their drug contents in the colon where it may be absorbed into the systemic circulation for higher bioavailability. It is anticipated that conditions of dissolution, absorption, and metabolism in the distal portions of the intestine are different than in the proximal regions, due to differences in pH, lumen fluid, mucosal morphology, and motility.

For most formulations, colonic absorption represents the only real opportunity to increase the interval between doses. Transit through the lower part of the gut is quoted at about 24 h, but in reality only the ascending colonic environment has sufficient fluid to facilitate dissolution. In the cecum, the fermentation of soluble fiber produces fatty acids and gas [33]. The gas rises into the transverse colon and can form temporary pockets, restricting access of water to the formulation. Consequently, distal release of drug is associated with poor spreading, reduced surface area, and restricted absorption. In the colon, water availability is also low past the hepatic flexure, as the ascending colon is extremely efficient at water absorption [34].

## 6 DESIGN AND FABRICATION OF CONTROLLED-RELEASE DOSAGE FORMS

### 6.1 Microencapsulation

Microencapsulation has been the subject of massive research efforts since its inception around 1950. Today, it is the mechanism utilized by approximately 65% of all sustained-release systems [35]. The technique's popularity can be attributed mainly to its wide variety of applications. Hundreds of drugs have been microencapsulated and used as controlled-release systems. Some examples are Arthritis Bayer, Dexatrim Capsules, and Dimetapp Elixir.

Microencapsulation provides more efficient drug delivery because it increases the ability of the drug to interact with the body. The active ingredient of a drug is encapsulated into a particle that may be as small as 1 μm. The greatest feature of microencapsulation is the control provided by the choice of coating [36–38]. This control allows microencapsulation to be a controlled-release device. Microcapsules can be engineered to gradually release drugs to the body. To achieve this type of delivery, equilibrium is established which will monitor the liberation of medicine from those microcapsules. A microcapsule may be opened by many different means. Release mechanisms include fracture by heat, solvation, diffusion, and pressure [39]. A coating may also be designed to open specific areas of the body. A microcapsule containing drugs that will be consumed by GI fluids must not be fractured until after it passes through the stomach [36–39]. A coating can therefore be used that is able to withstand stomach acids and allow the drug to pass through the stomach.

#### *Methods of Microencapsulation*

1. *Air Suspension* This method, known as the Wurster process or fluidized-bed coating, involves dispersing solid particulate core materials in a supporting air stream and the spray coating of the suspended materials [40]. The design of the chamber and its operating parameters effect a recirculating flow of the particles through the coating zone of the chamber, where a coating material, usually a polymer solution, is sprayed onto the fluidized particles. The cyclic process is repeated until the desired coat thickness is obtained.

2. *Pan Coating* This process has been around for many decades and is commonly associated with sugar coating. It is essential that the particles be greater than 600 μm for effective coating. The process has been extensively utilized for the preparation of controlled-release beads.

The coating by this method is applied as a solution or as an atomized spray to the desired solid core material in the coating pan [41]. Warm air is passed over the coated materials as the coatings are applied in the coating pans to remove the coating solvent. Sometimes the final solvent removal is carried out in a drying oven.

3. *Multiorifice Centrifugal Process* This is a mechanical process involving the use of centrifugal forces to hurl a material particle through an enveloping microencapsulation membrane to effect mechanical encapsulation [42]. The microcapsules are then hardened and congealed. This method is capable of microencapsulating liquids and solids dispersed in a liquid.

4. *Coacervation Phase* Encapsulation by coacervation is the one of the more popular methods commonly studied. The process consists of three steps carried out under continuous agitation [43]:

Step 1 The core material is dispersed in a solution of coating polymer, the solvent for the polymer being the liquid manufacturing vehicle phase.

Step 2 Deposition of the coating, accomplished by controlled, physical mixing of the coating material and the core material in the manufacturing vehicle.

Step 3 Rigidization of the coating by thermal, cross-linking, or desolvation techniques to form self-sustaining microcapsules.

Since the core materials are microencapsulated while being dispersed in some liquid manufacturing vehicle, subsequent drying operations are usually required.

5. *Solvent Evaporation Technique [39, 44, 45]* The process is carried out in a liquid manufacturing vehicle. The core material to be encapsulated is dissolved or dispersed in the coating polymer solution. With agitation, the core coating material mixture is dispersed in the liquid manufacturing vehicle phase to obtain the appropriate microcapsule size. The microcapsules can be used in suspension form, coated onto substrates or isolated as powders. A schematic of the emulsification/solvent evaporation technique to prepare drug-loaded microparticles is shown in Figure 4.

6. *Spray Drying* Spray drying, by definition, is the transformation of feed from a fluid state to a dried particulate form by spraying the feed into a hot drying medium [44, 46]. The feed can be a solution, suspension, or paste. A schematic is shown in Figure 5.

Spray drying consists of four process stages:

(i) Atomization of feed into spray
(ii) Spray–air contact (mixing and flow)
(iii) Drying of spray (moisture evaporation)
(iv) Separation of dried product from the air

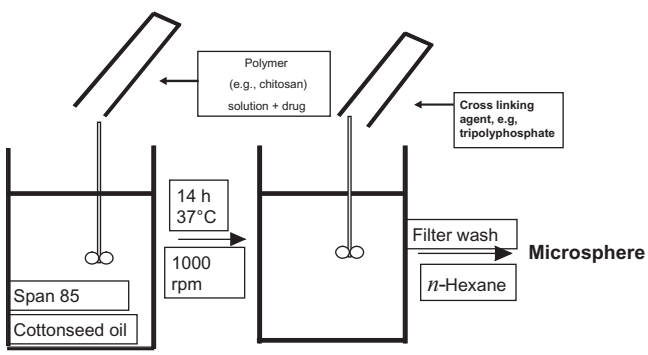


**FIGURE 4** Schematic of microspheres prepared by emulsification/solvent evaporation method.

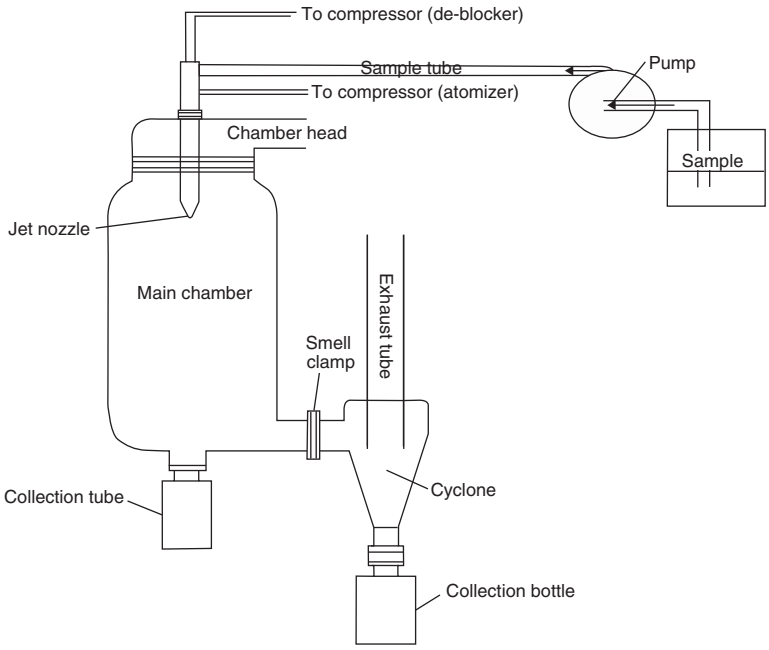


**FIGURE 5** Schematic of spray drying method.

7. *Spray Congealing [47]* Spray congealing is similar to spray drying in that it involves dispersing the core material in a liquefied coating substance. Coat solidification is accomplished by congealing the molten coating material or by solidifying a dissolved coating material by introducing the coat–core material mixture into a nonsolvent. Removal of the nonsolvent is then achieved by sorption, extraction, or evaporation techniques. Waxes, fatty acids and alcohol, and polymers and sugars, which are solids at room temperature but melt at high temperatures, are applicable to spray congealing methods.

### 6.2 Nanostructure-Mediated Controlled-Release Dosage Forms

The efficiency of drug delivery to various parts of the body is directly affected by particle size. Nanostructure-mediated drug delivery, a key technology for the realization of nanomedicine, has the potential to enhance drug bioavailability, improve the timed release of drug molecules, and enable precision drug targeting [48]. Nanoscale drug delivery systems can be implemented within pulmonary therapies, as gene delivery vectors, and in stabilization of drug molecules that would otherwise degrade too rapidly. Additional benefits of using targeted nanoscale drug carriers are reduced drug toxicity and efficient drug distribution.

Anatomic features such as the blood–brain barrier, the branching pathways of the pulmonary system, and the tight epithelial junctions of the skin make it difficult for drugs to reach many desired physiological targets. Nanostructured drug carriers will help to penetrate or overcome these barriers to drug delivery [49]. Greater uptake efficiency has also been shown for GI absorption and transcutaneous per-

meation, with particles around 100 and 50 nm in size, respectively. However, such small particles traveling in the pulmonary tract may help with delivery to the pulmonary extremities. For instance, the outer layers of the carrier architecture may be formulated to biodegradable as the carrier travels through the pulmonary tract. As the drug carrier penetrates further into the lung, additional shedding will allow the encapsulated drug to be released. Biodegradable nanoparticles of gelatin and human serum albumin (HSA) show promise as drug carriers.

Advantages of nanostructure-mediated drug delivery include the ability to deliver drug molecules directly into cells and the capacity to target tumors within healthy tissue [50]. The mechanisms of cellular uptake of external particulates include calthrin- and caveoli-mediated endocytosis, pinocytosis, and phagocytosis. However, phagocytosis may not play a role in the uptake of nanoscale particles because of the small size of such particles.

Nanoscale drug delivery architectures are able to penetrate tumors due to the discontinuous, or "leaky," nature of the tumor microvasculature, which typically contains pores ranging from 100 to 1000 nm in diameter. The microvasculature of healthy tissue varies by tissue type, but in most tissues, including the heart, brain, and lung, there are tight intercellular junctions less than 10 nm. Therefore, tumors within these tissue types can be selectively targeted by creating drug delivery nanostructures greater than the intercellular gap of the healthy tissue but smaller than the pores found within the tumor vasculature.

Various nanoscale architecture can be designed, including solid spheres, hollow spheres, tubes, porous particles, solid particles, and branched structures (Table 2). To achieve such nanostructures, different fabrication methods are used depending on the types of material. The methods used for nanoscale assembly include molecular self-assembly, bioaggregation, nanomanipulation, photochemical patterning, molecular imprinting, layer-by-layer electrsostatic deposition, and vapor deposition.

**TABLE 2 Nanoscale Drug Delivery Technologies**

| Drug Delivery Technology | Materials | Nanostructure Forms |
|---|---|---|
| Biologicals | Lipids, peptides, nucleic acids, polysaccharides, viruses | Vesicles, nanotubes, rings, nanoparticles, nanocapsules, nanospheres |
| Polymeric | poly(lactic acid), poly(glycolic acid), poly(alkylcyanoacrylate), poly(3-hydroxybutanoic acid), Poly(organophosphazene), poly(ethylene glycol), poly(caprolactone), poly(ethylene oxide), poly(amidoamine), poly(L-glutamic acid), poly(propylene imine) | Vesicles, nanospheres, nanoparticles, micelles, dendrimers |
| Silicon based | Silicon, silicon dioxide | Porous nanoparticles, nanoneedles |
| Carbon based | Carbon | Nanotubes |
| Metallic | Gold, silver, palladium, platinum | Nanoparticles, nanoshells |

***Manufacturing and Characterization of Nanoparticles/Nanocapsules/Nanospheres*** Production of nanoparticles of soft materials is much more difficult and challenging than that of hard materials because of the high stickiness of the former. The bulk pharmaceuticals are available in solids of large sizes, which often can be easily solubilized in solvent to obtain particular sizes. Hence, there are two extremes of sizes: molecular size (each particle containing one molecule) and larger sizes (e.g., each particle containing on the order of $10^{18}$ molecules). To obtain nanoparticles in the range of 50–300 nm of drug delivery, one requires on the order of $10^4$–$10^8$ molecules in each particle. This size has to be achieved from either solution-phase (single-molecule) or millimeter-size particles ($10^{18}$ molecules).

*Pearl/Ball Milling Technology for Production of Drug Nanocrystals* There are two different drug nanocrystal products, prepared by using pearl/ball milling technology. The Rapamune coated tablet is the more convenient formulation, introduced by Wyeth Pharmaceuticals in 2002. The Emend, introduced in 2003 by MSD, Sharp and Dohme Gmbh, is a capsule composed of sucrose, microcrystalline cellulose, hyperlose, and sodium dodecylsulfate [51].

Traditional equipment used for micronization of drug powders such as rotor-stator colloid mills or jet mills are of limited use for the production of nanocrystals. For example, jet milling leads to a drug powder with a size range of roughly 0.1–20 μm, containing only a very small fraction of about 10% in the nanometer range. However, it has been shown when running a pearl mill over a sufficiently long milling time that drug nanosuspensions can be obtained [52–54]. These mills consist of a milling container filled with fine milling pearls or larger sized balls. The container can be static and the milling material is moved by a stirrer; alternatively, the complete container is moved in a complex movement leading consequently to movement of the milling pearls.

The different milling materials available include traditional steel, glass, and zircon dioxide as well as new special polymers such as hard polystyrene. A general problem associated with this technology is the erosion from the milling materials during manufacturing [55]. Surfactants and stabilizers have to be added for the physical stability of the produced nanosuspensions. In the production process, the coarse drug powder is dispersed by high-speed stirring or homogenization in a surfactant/stabilizer solution to yield macro- and nanosuspensions. The choice of surfactant or stabilizers depends not only on the properties of the particles to be suspended (e.g., affinity of surfactant/stabilizer to the crystal surface) but also on the physical principles (e.g., electrostatic or steric stabilization) and the route of administration. In general, steric stabilization is recommended as it is less susceptible to electrolytes in the gut and blood.

There are number of pearl mills available on the market, ranging from laboratory-scale to industrial-scale volumes. The ability of large-scale production is an essential prerequisite for the introduction of a product to the market. One advantage of pearl mills, apart from their low cost, is their ability in scaling up.

*Nanoparticles/Nanoemulsions/Nanospheres Prepared by High-pressure Homogenization* High-pressure homogenization is a technology that has been applied for many years in various areas of the production of emulsions and suspensions. A distinct advantage of this technology is its ease in scaling up, even to very large volumes.

In the pharmaceutical industry, parenteral emulsions such as Intralipid and Lipofundin (mean droplet diameter 200–400 nm) are generally produced by this technology [56]. Typical pressures for the production of drug nanosuspensions are 1000–1500 bars (corresponding to 100–150 MPa); the number of required homogenization cycles varies from 10 to 20 depending on the properties of the drug. Most of the homogenizers used are based on the piston gap principle; an alternative can be jet stream technology [57].

In the piston gap homogenizer, the liquid is forced through a tiny homogenization gap, typically in the range of 5–20 μm (depending on the pressure applied and the viscosity of the dispersion medium). Using a Micron Lab 40, the suspension is supplied from a metal cylinder by a piston, and the cylinder diameter is approximately 3 cm. The suspension is moved by the piston having an applied pressure between 100 and 1500 bars. The piston gap homogenizer corresponds to a tube system in which the tube diameter narrows from 3 to 20 μm. The Microfluidizer (Microfluidics Inc.) is based on the jet stream principle [58]. Two streams of liquid collide, diminution of droplets or crystals is achieved mainly by particle collision, but occurrence of cavitation is achieved mainly by particle collision.

*Nanoparticles/Nanocapsules Obtained by Interfacial Polymerization* Nanoparticles/nanocapsules can be obtained by fast polymerization of a monomer at the interface between the organic and the aqueous phase of an emulsion. Alkylcyanoacrylates have been proposed for the preparation of both oil- and water-containing nanocapsules [59]. These monomers polymerize within a few seconds, initiated by hydroxyl ions from equilibrium dissociation of water or by nucleophilic groups of any compound of the polymerization medium.

1. *Formation of Nanocapsules/Nanospheres Containing Oil Core* This type of nanocapsule is preferred for the encapsulation of lipophilic and oil-soluble compounds. The general procedure for the preparation of oil-containing nanocapsules by interfacial polymerization of alkylcyanoacrylates consists of preparing a very fine oil-in-water (O/W) emulsion with an additional water-miscible organic solvent such as ethanol or acetone [60, 61]. These solvents are used to disperse the oil as very small droplets in the aqueous phase, which contains a hydrophilic surfactant. The solvent also serves as a vehicle for the monomer. Gallardo et al. [62] proposed a mechanism to explain nanocapsule formation. An important factor is that the organic solvent must be completely water miscible so that the formation of small enough oil droplets occurs spontaneously while the solvent is diffusing toward the aqueous phase and the water is diffusing toward the organic phase. The polymerization of monomer is also induced by contact with hydroxyl ions from the water phase, which should be very fast to allow efficient formation of a thin layer of coating around the oil droplet and thus achieve effective encapsulation of drugs.

The organic phase containing the oil, the monomer, and the bioactive compounds, dissolved in the water-miscible organic solvent, is injected into the aqueous phase containing water and a hydrophilic surfactant under strong magnetic stirring. The nanocapsules/nanospheres are formed immediately to give a milky suspension. The organic phase is then removed under reduced pressure using a rotary evaporator. The most commonly used materials for the preparation of oil-loaded nanocapsules are given in Table 3.

**TABLE 3 Main Components Used to Prepare Oil-Containing Nanocapsules/ Nanospheres by Interfacial Polymerization**

| Components | Examples |
|---|---|
| Oil | Miglyol, Lipiodol, benzylbenzoate |
| Monomers | Ethylcyanoacrylate, isobutylcyanoacrylate, *n*-butylcyanoacrylate, isohexylcyanoacrylate |
| Organic solvents | Ethanol, acetone, acetonitrile, *n*-butanol, isopropanol |
| Surfactants | Poloxamer 188, poloxamer 238, poloxamer 407, Triton X100, Tween 80 |

Nanocapsule/nanosphere size ranges between 200 and 350 nm were observed to be affected by both the oil–ethanol ratio and the oil–monomer ratio [63, 64]. It is also influenced by the particular oil, water-miscible organic solvent, and nonionic surfactant in the aqueous phase. The pH of the aqueous phase and the temperature also affect the size distribution.

2. *Nanocapsules/Nanospheres Containing Aqueous Core Obtained by Interfacial Polymerization* Nanocapsules/nanospheres with an aqueous core are a recent technology developed for the efficient encapsulation of water-soluble compounds, which are generally very difficult to include within carrier systems. This type of nanocapsule/nanosphere may also be obtained by interfacial polymerization, but in this case monomers are added to a water-in-oil (W/O) emulsion. Anionic polymerization of the cyanoacrylates in the water phase is initiated at the interface by nucleophiles such as hydroxyl ions in the aqueous phase, leading to the formation of nanocapsules/nanospheres with an aqueous core. A typical procedure as described by Lambert et al.[65] consists of preparing an aqueous phase composed of ethanol (20% v/v) in water (pH 7.4) which is emulsified in an organic phase containing Miglyol oil and Montane 80. Slow addition of cyanoacrylic monomer under mechanical stirring (about 4 h) allows the polymerization to occur. Thus the water droplets are surrounded by the polymer-forming nanocapsules with an aqueous core dispersed in an oily phase.

*Polymeric Nanocapsules/Nanospheres/Nanoparticles*

1. *Manufacturing by Interfacial Nanodeposition/Nanoprecipitation* The nanoprecipitation procedure generally consists of a water-miscible organic phase such as an alcohol or a ketone containing oil (with or without lipophilic surfactant) with an aqueous phase containing a hydrophilic surfactant. The polymer, which may be preformed synthetic, semisynthetic, or natural, is solubilized in the organic phase (or in a phase in which the polymer is soluble). After addition of the organic phase to the aqueous phase, the polymer diffuses with the organic solvent and is stranded at the interface between oil and water. The driving force for the formation of nanocapsules/nanospheres is the rapid diffusion of the organic solvent in the aqueous phase inducing interfacial nanoprecipitation of the polymer around droplets of the oily phase. The polymers which may be used to manufacture nanoparticles by this method include natural polymers such as gum arabic, chitosan, alginate, gelatin, ethylcellulose, hydroxypropyl methylcellulose (HPMC), and hydroxypropyl methylcellulose phthalate (HPMCP); semisynthetic polymers such as diacyl β-cyclodextrin; and synthetic polymers such as poly(D,L-lactide), poly(ε-caprolactone),

and poly(alkylcyanoacrylate), which are most commonly employed [65]. Similarly, a broad range of oils is suitable for the preparation of nanocapsules/nanospheres, including vegetable or mineral oils and pure compounds such as ethyl oleate and benzyl benzoate. The criteria for the selection of these compounds are the nontoxicity and the low solubility of the oil in the polymers and vice versa.

Both hydrophilic and lipophilic surfactants can be used to stabilize the polymeric nanoparticles. Generally the lipophilic surfactant is a natural lecithin of relatively low phosphotidylcholine content, whereas the hydrophilic one is synthetic: anionic (lauryl sulfate), cationic (quaternary ammonium), or more commonly nonionic [poly(oxyethylene)-poly(propylene)glycol]. Nanoparticles can be prepared in the absence of surfactants, but there are lots more chances to get aggregated during storage.

2. *Nanoparticles Obtained by Multiple Emulsion/Solvent Evaporation Method* The multiple emulsion/solvent evaporation method was initially developed for the preparation of microcapsules. This method consists in first dissolving the drug in an aqueous solution with or without a surfactant and the polymer in a volatile organic solvent that is not miscible to water. Polymers used for the formation of such types of particles have been mainly poly(lactide-*co*-glycolide) and poly(ε-caprolactone) [66, 67]. The inner water phase is then poured into the organic phase. This mixture is generally emulsified forming the first inner emulsion or the primary W/O emulsion, which is then mixed vigorously into an aqueous phase (outer water phase) that contains an emulsifier forming the water-in-oil-in-water (W/O/W) multiple emulsion. The resulting multiple nanoemulsion is continuously stirred and the solvent is allowed to evaporate, inducing precipitation of polymer and, thereby, the formation of solid drug loaded nanoparticles.

### 6.3 Liposomes

Liposomes were discovered in the mid-1960s [68] and were originally studied as cell membrane models. They have since gained recognition in the field of drug delivery. Liposomes are formed by the self-assembly of phospholipid molecules in an aqueous environment. The amphophilic phospholipid molecules form a closed bilayer sphere in an attempt to shield their hydrophobic groups from the aqueous environment while still maintaining contact with aqueous phase via the hydrophilic head group. The resulting closed sphere may encapsulate aqueous soluble drugs within the central aqueous compartment or lipid-soluble drugs within the bilayer membrane. Alternatively, lipid-soluble drugs may be complexed with other polymers (e.g., cyclodextrin) and subsequently encapsulated within the liposome aqueous compartment. The encapsulation within/association of drugs with liposomes alters the drug pharmacokinetics.

***Attractive Biological Properties of Liposomes*** [69]

- Liposomes are biocompatible.
- Liposomes can entrap hydrophilic bioactive compounds in their internal compartment and hydrophobic into the membrane.
- Liposome-incorporated bioactives are protected from the inactivating effect of external conditions yet do not cause undesirable side reactions.

**TABLE 4 Liposomal Drugs Approved for Clinical Application or Undergoing Clinical Evaluation**

| Active Drug | Product Name | Applications |
|---|---|---|
| Daunorubicin | DaunoXome | Sarcoma |
| Doxorubicin | Mycet | Breast cancer |
| | Doxil/Caelyx | Sarcoma, ovarian cancer, breast cancer |
| Amphotericin B | AmBisome | Fungal infections |
| Cytarabine | DepoCyt | Lymphomatous meningitis |
| Vincristine | Onco TCS | Non-Hodgkin's lymphoma |
| Lurtotecan | NX211 | Ovarian cancer |
| Nystatin | Nyotran | Topical antifungal agent |
| All-*trans* retinoic acid | Altragen | Leukemia, carcinomas |
| DNA plasmid encoding HLA-B7 and $\alpha_2$-microglobulin | Allovectin-7 | Metastatic melanoma |

- Liposomes provide a unique opportunity to deliver pharmaceuticals into cells or even inside individual cellular compartments.
- The size, charge, and surface properties of liposomes can be easily changed by adding new ingredients to the lipid mixture before liposome preparation and/or by variation of preparation methods.

The clinical applications of liposomes are well known (Table 4). The initial success achieved with many liposome-based drugs has fueled further clinical investigations. One of the drawbacks of the use of liposomes is the fast elimination from the blood and capture of liposomal preparations by the cells of the reticuloendothelial system (RES), primarily in the liver.

There are a number of different types of liposomal vesicles [69]:

- *Multilamellar Vesicles* These range in size from 500 to 5000 nm and consist of several concentric bilayers.
- *Small Unilamellar Vesicles* These are around 100 nm in size and are formed by a single bilayer.
- *Large Unilamellar Vesicles* These range in size from 200 to 800 nm.
- *Long Circulating Liposomes* Different methods have been suggested to achieve long circulation of liposomes in vivo, including coating the liposome surface with inert, biocompatible polymers, such as polyethylene glycol (PEG), which form a protective layer over the liposome surface and slow down its recognition by opsonins and therefore subsequent clearance of liposomes. An important feature of protective polymers is their flexibility, which allows a relatively smaller number of surface-grafted polymer molecules to create an impermeable layer over the liposome surface. These types of modified liposomes demonstrate dose-dependent, nonsaturable, long-linear kinetics, and increased bioavailability.
- *Immunoliposomes* To increase liposomal drug accumulation in the desired tissues and organs, the use of targeted liposomes with surface-attached ligands capable of recognizing and binding to cells of interest has been suggested [70]. Immunoglobulins (Ig) of the IgG class and their fragments are the most widely

used targeting moieties for liposomes, which can be attached, without affecting liposomal integrity or the antibody properties, by covalent bonding to the liposome surface or by hydrophobic insertion into the liposomal membrane after modification with hydrophobic residues.

### 6.4 Niosomes

The success achieved with liposomal formulations stimulated the search for other vesicle-forming amphiphiles. Nonionic surfactants were among the first alternative materials studied and a large number of surfactants have since been found to self-assemble into closed bilayer vesicles which may be used for drug delivery [71]. Anticancer niosomes are expected to accumulate within tumors. The niosomal encapsulation of methotrexate and doxorubicin increases drug delivery to the tumor and tumoricidal activity. Unlike nonstealth liposomes, doxorubicin niosomes (size 800 nm) possessing a triglycerol or doxorubicin niosomes (size 200 nm) possessing a muramic acid surface are not taken up significantly by the liver. As such, these triglycerol niosomes accumulate in the tumor. However, muramic acid vesicles do accumulate in the spleen. Uptake by the liver and spleen make niosomes ideal for targeting diseases manifesting in these organs. One such condition is leishmaniasis, and number of studies have shown that niosomal formulations of sodium stibogluconate improve parasite suppression in the liver, spleen, and bone marrow. Niosomes may also be used as depot systems for short-acting peptide drugs on intramuscular administration [72].

Niosomal antigens are potent stimulators of the cellular and humoral immune response. The formulation of antigens as a niosome in W/O emulsions further increases the activity of antigens. The controlled-release property of these types of emulsion formulations is responsible for enhancing the immunological responses.

## 7 TECHNOLOGIES FOR DEVELOPING TRANSDERMAL DOSAGE FORMS

Continuous intravenous infusion at a programmed rate has been recognized as a superior model of drug delivery not only to bypass the hepatic first-pass elimination but also to maintain a constant, prolonged, and therapeutically effective drug level in the body. A closely monitored intravenous infusion can provide both the advantages of direct entry of drugs into the systemic circulation and control of circulating drug levels. However, such a mode of drug delivery entails certain risks and therefore necessitates hospitalization of patients and close medical supervision of the medication. Recently there has been an increasing awareness that the benefits of intravenous drug infusion can be easily duplicated, without its potential hazards, by continuos transdermal drug administration through intact skin [73].

Advances in transdermal delivery systems (TDSs) and the technology involved have been rapid because of the sophistication of polymer science, which now allows incorporation of polymeric additives in TDSs in adequate quantity. Drugs with which transdermal therapy was pioneered include scopolamine, nitroglycerine, isosorbide dinitrite, clonidine, estradiol, nicotine, and testosterone [74].

***Advantages of Transdermal Drug Delivery System*** [75]

- Avoids GI absorption (pH effects, enzymatic activity, drug interactions)
- Substitute for oral route
- Avoids first-pass effect (drug deactivation by digestive and liver enzymes)
- Multiday therapy with a single application
- Extends the activity of the drugs with short half-lives
- Provides capacity to terminate drug effects rapidly
- Rapid identification of medication in emergency

***Limitations of Transdermal Drug Delivery*** [75]

- Not for all drugs
- Limited time that the patch can remain affixed
- Variable intra- and interindividual percutaneous efficiency absorption efficiency
- Variable adhesion to different skin types
- Skin rashes and sensitization
- Bacterial and enzymatic drug metabolism under the patch
- Complex technology/high cost

***Skin Site for Transdermal Drug Administration*** The skin is one of the most extensive and readily accessible organs of the human body. The skin of an average adult body covers a surface area of approximately $2\,m^2$ and receives about one-third of the blood circulating through the body. It is elastic, rugged, and, under normal physiological conditions, self-regenerating. It serves as a barrier against physical and chemical attacks and shields the body from invasion by microorganisms.

Microscopically the skin is a multilayered organ composed of, anatomically, many histological layers, but it is generally described in terms of three tissue layers: the epidermis, the dermis, and the subcutaneous fat tissue.

Microscopic sections of the epidermis show two main parts: the stratum corneum and the stratum germinativum. The stratum corneum forms the outermost layer of the epidermis and consists of many layers of compacted, flattened, dehydrated, keratinized cells in stratified layers. In normal stratum corneum, the cells have a water content of only approximately 20% compared to the normal physiological level of 70% in the physiologically active 10% (w/w) to maintain flexibility and softness. It becomes rough and brittle, resulting in so-called dry skin, when its moisture content decreases at a rate faster than can be resupplied from the underlying tissues. The stratum corneum is responsible for the barrier function of the skin. It also behaves as the primary barrier to percutaneous absorption. The thickness of this layer is mainly determined by the extent of stimulation of the skin surface by abrasion and weight bearing; hence thick palms and soles develop.

Several technologies have been successfully developed to provide rate control over the release and skin permeation of drugs. These technologies can be classified into four basic approaches which are described below [73, 75].

***Polymer Matrix Diffusion-Controlled Transdermal Drug Delivery (TDD) System*** In this approach, the drug reservoir is formed by homogeneously dispersing the drug solids in a hydrophilic or lipophilic polymer matrix, and the medicated polymer formed is then molded into medicated disks with a defined surface area and controlled thickness. This drug-reservoir-containing polymer disk is then mounted onto an occlusive baseplate in a compartment fabricated from a drug-impermeable plastic baking (Figure 6*a*).

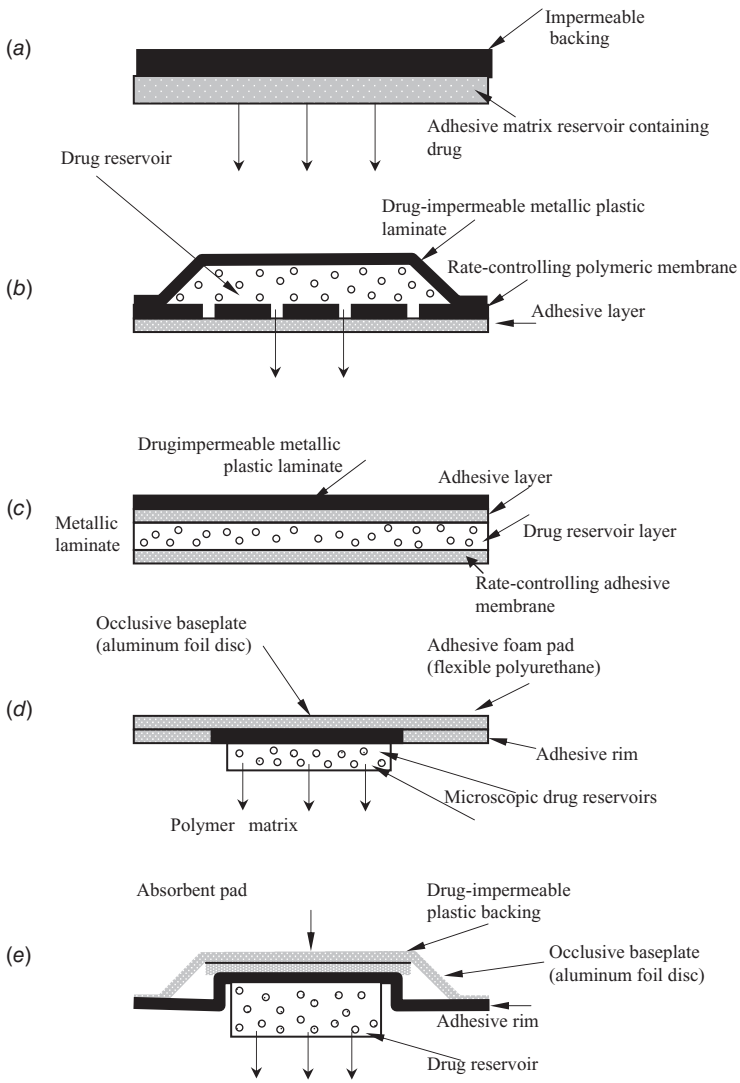


**FIGURE 6** Cross-sectional view of several TDSs: (*a*) poly(sebacic anhydride) (PSA) matrix device; (*b*) membrane-moderated TDS; (*c*) adhesive-controlled TDS; (*d*) microreservoir-type TDS; (*e*) matrix dispersion–type TDS.

***Polymer Membrane Permeation-Controlled TDD System*** In this system, the drug reservoir is sandwiched between a drug-impermeable backing laminate and a rate-controlling polymeric membrane (Figure 6*b*). The drug molecules are permitted to release only through the rate-controlling polymeric membrane.

***Drug Reservoir Gradient-Controlled TDD System*** The rate of drug release from this type of drug reservoir type is gradient controlled. In this system the thickness of the diffusion path through which the drug molecule diffuses increases with time (Figure 6*c*).

***Microreservoir Dissolution-Controlled TDD System*** This type of drug delivery system can be considered a hybrid of the reservoir and matrix dispersion-type drug delivery systems. In this approach, the drug reservoir is formed by first suspending the drug solids in an aqueous solution of water–miscible solubilizer (e.g., PEG), and then homogeneously dispersing the drug suspension with controlled aqueous solubility, in a lipophilic polymer, by high-shear mechanical force, to form thousands of unleachable microscopic drug reservoirs. This thermodynamically unstable dispersion is quickly stabilized by immediately cross-linking the polymer chains in situ, which produces a medicated polymer disk with a constant surface area and a fixed thickness. Mounting the medicated disk at the center of an adhesive pad then produces a TDD system (Figure 6*d* and *e*).

## 8 OCULAR CONTROLLED-RELEASE DOSAGE FORMS

The eye is unique in its therapeutic challenges. The eye drop dosage form is easy to instill but suffers from the inherent drawback that the majority of the medication it contains is immediately diluted in the tear as soon as the eye drop solution is instilled. Usually less than 10% of a topically applied dose is absorbed into the eye, leaving the rest of the dose to potentially absorb into the bloodstream, resulting in unwanted side effects [76]. The objectives of most of the controlled delivery system are to maintain the drug in the precorneal area and allow its diffusion across the cornea. Polymeric matrices seem to reduce the drainage significantly, but other newer methods of controlled-release dosage forms can also be used.

The sustained release of artificial tears has been achieved by a hydroxypropylcellulose polymer insert [77]. However, the best known application of diffusional therapy in the eye, Ocusert-Pilo, as shown in Figure 7, is a relatively simple structure with two rate-controlling membranes surrounding the drug reservoir containing

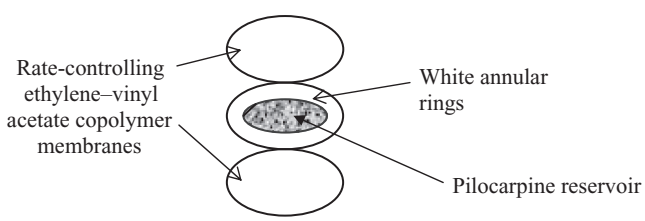


**FIGURE 7** Schematic of Ocusert intraocular device for controlled release of pilocarpine.

pilocarpine. The unit is placed in the eye and resides in the lowe cul-de-sac, just below the cornea. Since the device itself remains in the eye, the drug is released into the tear film. The advantage of such a device is that it can control intraocular pressure for up to a week. Controlled release is achieved with less drug and fewer side effects, since the release of drug is zero order. However, it is difficult to keep it in the eye for a longer time and can cause discomfort.

The prodrug administration is also getting attention as ocular controlled-release dosage forms. Since the corneal surface presents an effective lipoidal barrier, especially to hydrophilic compounds, it seems reasonable that a prodrug that is more lipophilic than the parent drug will be more successful in penetrating this barrier. Recently, dipivalyl epinephrine (Dividephrine), a dipivalyl ester of epinephrine, has been formulated [78]. Epinephrine itself is poorly absorbed owing to its polar characteristics and is highly metabolized. The prodrug form is 10 times as effective at crossing the cornea and produces substantially higher aqueous humor levels.

New sustained technologies are also gaining much interest in ocular delivery, as in other routes. Liposomes as drug carriers have achieved enhanced ocular delivery of certain drugs, antibiotics, and peptides. Prolonged delivery of pilocarpine can be achieved with a polymeric dispersion or submicrometer emulsions [79].

## 9 VAGINAL AND UTERINE CONTROLLED-RELEASE DOSAGE FORMS

Controlled-release devices for vaginal and uterine areas are most often for the delivery of contraceptive steroid hormones. The advantages are prolonged release, minimal side effects, and increased bioavailability. First-pass metabolism that inactivates many steroid hormones can also be avoided.

Therapeutic levels of the medroxyprogestrone vaginal ring have been achieved at a total dose that is one-sixth the required oral dose [80]. The sustained release of progesterone from various polymers given vaginally has also been found useful in cervical ripening and the induction of labor. A possible new use of the vaginal route is for long-term delivery of antibodies. When various antibodies, including monoclonal IgG, were administered from polymer vaginal rings in test animals, antibody concentrations remained high over a month in vaginal secretions and detectable in blood serum [81].

The hormone-releasing devices in uterus have a closer resemblance to controlled release because they involve the release of a steroid compound by diffusion [82, 83]. Progesterone, the active ingredient, is dispersed in the inner reservoir, surrounded by ethlene/vinyl acetate copolymer membrane. The release of progesterone from this system is maintained almost constant for about a year [84–86].

## 10 RELEASE OF DRUGS FROM CONTROLLED-RELEASE DOSAGE FORMS

There are three primary mechanisms by which active agents can be released from a delivery system: diffusion, degradation, and swelling followed by diffusion. Any or all of these mechanisms may occur in a given release system. Probable

**TABLE 5 Probable Mechanism of Drug Delivery from Hydrogels with Certain Environmental Conditions**

| Stimulus | Hydrogel | Mechanism |
|---|---|---|
| pH | Acidic or basic hydrogel | Change in pH, swelling–diffusion, erosion or burst release of drug |
| Ionic strength | Ionic hydrogel | Change in ionic strength, change in concentration of ions inside gel, change in swelling, release of drug |
| Chemical species | Hydrogel containing electron-accepting groups | Electron-donating compounds, formation of charge transfer complex, change in swelling, release of drug |
| Enzyme–substrate | Hydrogel containing immobilized enzymes | Substrate present, enzymatic conversion, product changes, swelling of gel, release of drug |
| Magnetic | Magnetic particles dispersed in alginate microspheres | Applied magnetic field, change in pores in gel, change in swelling, release of drug |
| Thermal | Thermoresponsive hydrogel | Change in temperature, change in polymer–polymer and water–polymer interactions, change in swelling, release of drug |
| Electrical | Polyelectrolyte hydrogel | Apply electric field, membrane charging, electrophoresis of charged drug, change in swelling, release of drug |
| Ultrasound irradiation | Ethylene–vinyl alcohol hydrogel | Ultrasound irradiation, temperature increase, release of drug |

mechanisms of drug release from controlled-release dosage forms are briefly described in Table 5. Diffusion occurs when a drug or other active agent passes through the polymer that forms the controlled-release device. The diffusion can occur on a macroscopic scale—as through pores in the polymer matrix—or on a molecular level—by passing between polymer chains [4, 87].

Table 5 describes the probable mechanisms of drug delivery from controlled-release dosage forms under contain environmental conditions. A polymer and active agent have generally been mixed to form a homogeneous system, also referred to as a matrix system. Diffusion occurs when the drug passes from the polymer matrix into the external environment. As the release continues, its rate normally decrease with this type of system, since the active agent has a progressively longer distance to travel and therefore requires a longer diffusion time to release. The drug release is accomplished only when the polymer swells. Because many of the potentially most useful pH-sensitive polymers swell at high pH values and collapse at low pH values, the triggered drug delivery occurs upon an increase in the pH of the environment. Such materials are ideal for systems such as oral delivery, in which the drug is not released at low pH values in the stomach but rather at high pH in the upper small intestine.

In reservoir systems, a reservoir—whether solid drug, dilute solution, or highly concentrated drug solution with in polymer matrix—is surrouned by a film or membrane of a rate-controlling material [5]. The only structure effectively limiting the

release of the drug is the polymer layer surrounding the reservoir. Since this polymer coating is essentially uniform and of a nonchanging thickness, the diffusion rate of the active agent can be kept fairly stable throughout the lifetime of the delivery system.

For the diffusion-controlled systems ddescribed thus far, the drug delivery device is fundamentally stable in the biological environment and does not change its size through either swelling or degradation [4]. It is also possible for a drug delivery system to be designed so that it is incapable of releasing its agent or agents until it is placed in an appropriate biological environment. Swelling-controlled-release systems are initially dry and, when placed in the body, will absorb water or other body fluids and swell. The swelling increases the aquenous solvent content with the formulation as well as the polymer mesh size, enabling the drug to diffuse through the swollen network into the external environment. Most of the materials used in swelling-controlled-release systems are based on hydrogels, which are polymers that well without dissolving when placed in water or other biological fluids.

## 10.1 Time-Controlled-Release Dosage Forms

To achieve a drug release which is independent of the environment (e.g., pH, enzymatic activity, intestinal motility), the lag time prior to the drug release has to be controlled primarily by the delivery system. The release mechanisms employed include bulk erosion of polymers in which drug release by diffusion is restricted, surface erosion of layered devices composed of alternation drug-containing and drug-free layers, osmotically controlled rupture, and enzymatic degradation of liposomes. The device environment may modulate the release profile of any of these systems and may depend on factors such as the amount of free moisture, regional blood flow, and various cellular activities at the site [88, 89].

***Systmes with Eroding or Soluble Barrier Coatings*** These types of delivery systems comprise reservoir devices coated with a barrier layer. the barrier dissolves or erodes after a specified lag period, after which the drug is released rapidly from the reservoir core. In general, the lag time prior to drug release from a reservoir-type device can be controlled by the thickness of the coating layer, for example, the Chronotropic systems, which consists of a drug-containing core layered with hydroxy propyl methyl cellulose (HPMC), optionally coated with an outer enteric coating. The lag time prior to drug release is controlled by the thickness and the viscosity grade of the HPMC layer. After erosion or dissolution of the HPMC layer, a distinct pulse was observed. The Chronotropic system [90, 91] is an oral dosage form designed to achieve time-controlled delivery. This system has been developed keeping in mind the interaction between GI fluid and coating polymer, which causes time- or site-controlled release. The reaction causes the liberation of drugs by the mechanism of swelling of polymer, increased permeability, and dissolution/erosion phenomena. This system probably works better for poorly water soluble drug because highly water soluble drugs could diffuse through the swollen HPMC layer prior to complete erosion.

The TIME CLOCK system for the oral dosage should enable fast and complete release of drug after a predetermined lag time [92]. A tablet was made containing the drug molecule and bulking agents (lactose, polyvinlpyrrolidine, corn starch, and

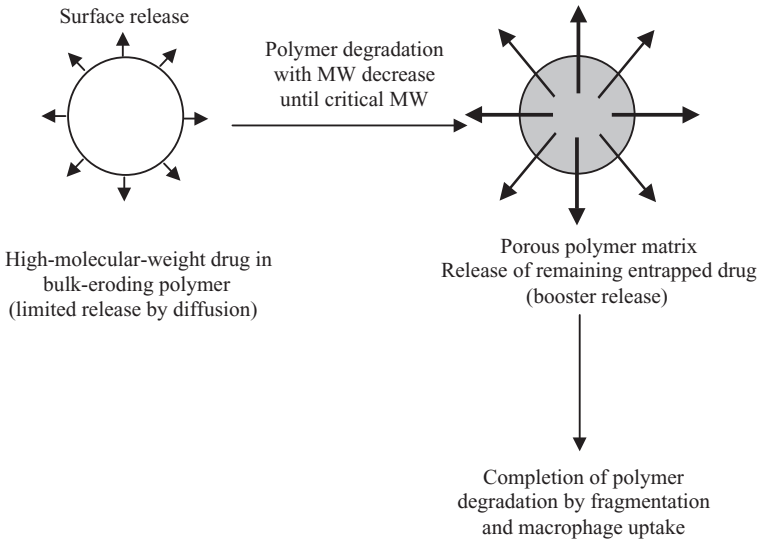


**FIGURE 8** Theoretical controlled release from a surface-eroding polymeric system. (Adapted from ref. 93 with permission of Elsevier Copyright 1999.)

magnesium stearate). This core was coated with a hydrophobic dispersion of carnauba wax, bees' wax, poly(oxyethylene) sorgitan monooleate, and HPMC in water. By altering the coating thickness, the lag time could be proportionally modulated. In vitro results indicated that the drug core was dissolved immediately after direct immersion in water and release was completed within 30 min, while a rapid release was observed after a certain lag time for the TIME CLOCK system with the hydrophobic coating. In vivo results revealed that drug disintergration was modulated by the coating thickness of the drug core as well as the food intake before drug administration. This approach may also be used to control the release onset time. Since the drug core is formulated with soluble ingredients, shell dissolution/distintegration becomes the key factor to control the lag time. Furthermore, drug release is independent of normal physiological conditions, such as pH, digestive state, and anatomical position at the time of release. This approach could be applicable for oral as well as for implant systems. Figure 8 illustrates the theoretical description of drug release from surface-eroding polymeric controlled-release dosage forms [93].

***Systems with Rupturable Coatings*** This class of reservoir-type pulsatile release system is based on rupturable coatings. The drug is released from a core after rupturing of a surrounding polymer layer, caused by a pressure buildup within the system, as shown in Figure 9 [93]. The pressure necessary to rupture the coating can be achieved with gas-producing effervescent excipients, inner osmotic pressure, or swelling agents.

An effervescent mixture of citric acid and sodium bicarbonate was incorporated in a tablet core coated with ethyl cellulose. The carbon dioxide development after water penetration into the core resulted in a pulsatile release after rupture of the coating, which was strongly dependent on the mechanical properties of the coating layer: The weak and nonflexible ethyl cellulose film ruptured sufficiently when

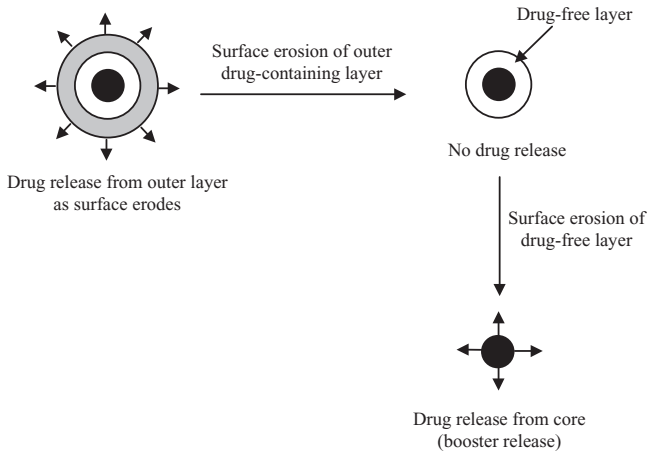


**FIGURE 9** Theoretical controlled release from an osmotically driven system. (Adapted from ref. 93 with permission of Elsevier Copyright 1999.)

compared with more flexible films. The lag time before release increased with increasing coating level and increasing hardness of the core tablet. The effectiveness of so-called superdistintegrants, which are highly swellable agents, was demonstrated for a capsule-based system consisting of a drug containing a core capsule, and swelling layer, and a rupturable polymeric layer. Croscarmellose, sodium starch glycolate, or low-substituted hydroxypropyl cellulose (L-HPC) were used as swelling substances, which resulted in a complete film rupture followed by a rapid drug release. The lag time is controlled by the composition of the outer polymer layer: Water-soluble polymers such as HPMC increase the permeability and therefore reduce the lag time. The swelling energy of several excipients decreased in the following order: croscamellose sodium > L-HPC > crospovidone > HPMC. Bothe solid and liquid drug formulations could be delivered with this system [94–96].

A novel capsule was made from ethyl cellulose for the time-controlled release of drugs in the colon [97]. Initially the ethyl cellulose capsule was prepared using a gelatin capsule with ethyl cellulose, followed by dissolution of the gelatin in water. The thickness of the ethyl cellulose capsule body was varied and the effect of wall thickness on the release of drugs in the capsules was investigated. Ethyl cellulose capsules contained a large number of mechanically made micropores (400 μm) at the bottom. Also located in the bottom of the capsule body was a swellable layer consisting of L-HPC. Above the swellable layer was the drug reservoir, which contained a mixture of the model drug, fluorescein, and a bulking agent, such as lactose or starch. The capsule was thus capped and sealed with a concentrated ethyl cellulose solution. After administration of drug-containing capsule, water molecules penetrated the capsule through the micropores in the bottom of the capsule body. Hydration and swelling of HPC induced an increase in the internal osmotic pressure, which resulted in the "explosion" of the capsule and a burstlike drug release was observed. By altering the thickness of the capsule, the lag time of the drug release could be altered. A similar approach for the pulsatile release of drug was reported in which a hydrostatic pressure was generated inside the capsules.

***Systems with Capsular Structure*** Several single-unit pulsatile dosage forms with a capsular design have been developed. Most of them consist of an insoluble capsular body which contains the drug and a plug which is removed after a predetermined lag time because of swelling, erosion, or dissolution.

The Pulsincap system consists of a water-insoluble capsule body filled with the drug formulation. The capsule half is closed at the open end with a swellable hydrogel plug. The dimension and the position of the plug can control the lag time prior to the release. In order to assure a rapid release of the drug content, effervescent agents or disintegrants can be included in the drug formulation, in particular with water-insoluble drugs. The system is coated with an enteric layer which dissolves upon reaching the higher pH region of the small intestine. This system comprises insoluble capsules and plugs. The plugs consist either of swellable materials, which are coated with insoluble but permeable polymers (e.g., polymethacrylates), or of erodible substances, which are compressed (e.g., HPMC, polyvinyl alcohol, polyethylene oxide) or prepared by congealing of melted polymers (saturated polyglycolated glycerides of glyceryl monooleate). The erosion of the plug can also be controlled enzymatically: A pectin plug can be degraded by pectinolytic enzymes being directly incorporated into the plug [98–100].

Linkwitz et al. [101] described the delivery of agents from osmotic systems based on the technology of an expandable orifice. The system is in the form of a capsule from which the drug is delivered by the capsule's osmotic infusion of moisture from the body. The delivery orifice opens intermittently to achieve a pulsatile delivery effect. The orifice forms in the capsule wall, which is constructed of an elastic material, preferably elastomer (e.g., styrene–butadiene copolymer), which strectches under apressure dufferential caused by the pressure rise inside the capsule as the osmotic infusion progessses. The orifice is small enough that when the elastic wall is relaxed, the flow rate of drug through the orifice is substantially zero, but when the elastic wall is stretched due to the pressure differential across the wall exceeding a threshold, the orifice expands sufficiently to allow the release of the drug at a phsiologically required rate. This osmotically driven delivery device as an implant can used in the anal–rectal passageway, in the cervical canal, as an artificial gland, in the vagina, as ruminal bolus, and the like.

A core-shelled cylindrical dosage form is available comprising a hydrophobic polycarbonate coating and a cylindrical core of alternating polyanhydride isolating layer and drug-loaded poly[ethyl glycinate) (benzyl amino acethydroxamate) phosphazene] (PEBP) layer for a programmable drug delivery system for single-dose vaccine and other related applications [102]. The pulsatile release of model compounds [fluorescein isothiocyanate (FITC)–dextran and myoglobin] whith a certain lag time (18–118 h) was achieved on the basis of the pH-sensitive degradation of PEBP and its cooperative interaction with polyanhydrides. In another experiment, Jiang and Zhu [103, 104] designed laminated devices comprising of polyanhydrides as isolating layers and pH-sensitive complexes of poly(sebacic anhydride)-*b*-polyethylene glycol (PSA-*b*-PEG) and poly(trimellitytylimdoglycine-*co*-sebacic anhydride)-*b*-polyethylene glycol [P(TMA-gly-*co*-SA)-*b*-PEG] as protein-loaded layers. The release of model proteins [bovine serum albumin (BSA) and myoglobin] showed a typical pulsatile fashion. The lag time prior to the release correlated with the hydrolytic druation of polyanhydrides, which varied from 30 to 165 h depending on polymer type and isolating layer thickness.

## 10.2 Stimuli-Induced Controlled-Release Systems

Several polymeric delivery systems undergo phase transitions and demonstrate marked swelling–deswelling change in reponse to environmental changes, including solvent composition ionic strength, temperature, electric fields, and light [105]. Responsive drug release from those systems results from the stimuli-induced changes in the gels or micelles, which maydeswell, swell, or erode in response to the respecive stimuli. The mechanisms of drug release include ejection of the drug from the gel as the fluid phase synerses out and drug diffusion along a concentration gradient.

***pH-Responsive Drug Release Dosage Forms*** pH-sensitive enteric coatings have been used routinely to deliver drug to the small intestine. These polymer coatings are insensitive to the acidic conditions of the stomach yet dissolve at the higher pH environment of the small intestine. This pH differential has also been attempted for colonic delivery purposes, although the polymers used for colonic targeting tend to have a threshold pH for dissolution that is higher than for those used in conventional enteric coating applications [104–106].

The synthesis and characterization of series of novel azo hydrogels for colon-targeting drug delivery have been described. The colon specificity is achieved dure to the presence of pH-sensitive monomers and azo cross-linking agents in the hydrogel structures. Most commonly, copolymers of methacrylic acid and methylmethacrylate that dissolve at pH 6 (Eudragit L) and pH 7 (Eudragit S) have been extensively investigated [106, 107]. This approach is based on the assumption that gastrointestinal pH increases progressively from the small intestine to the colon. The pH in the distal small intestine is usually around 7.5, while the pH in the proximal colon is closer to 6.

To overcome the premature release of drugs, a copolymer of methacrylic acid, methylmethacrylate, and ethyl acetate (Eudragit ES), which dissolves at a slower rate and at a higher threshold pH (7–7.5), has been developed [108]. The trn A series of in vitro dissolution studies with this polymer have highlighted clear benefits over the Eudragit S polymer for colon targeting. A gamma scintigraphy study comparing the in vivo performance of these various polymers revealed that Eudragit S (coating over the tablets) was superior to the older polymers in terms of retarding drug release in the small intestine, although, in some cases, the coated tablets did not break up at all. pH-sensitive delivery systems are commercially available for mesalazine (5-iminosalicylic cid) (Asacol and Salofalk) and budesonide (Budenofalk and Entocort) for the treatment of ulcerative colitis and Crohn's disease, respectively [109].

Natural polysaccharides are being used for the development of solid dosage forms for pH-dependent delivery and for targeting the release of drugs in colon [110]. Various major approaches utilizing polysaccharides are fermentable coating of the drug core, embedding of the drug in biodegradable matrix, and formation of drug–saccharide conjugate (prodrugs). A large number of polysaccharides have already been studied for their potential in these types of delivery systems, such as chitosan, alginate, pectin, chondroitin sulfate, cyclodextrin, dextrans, guar gum, inulin, amylose, and locust bean gum [111].

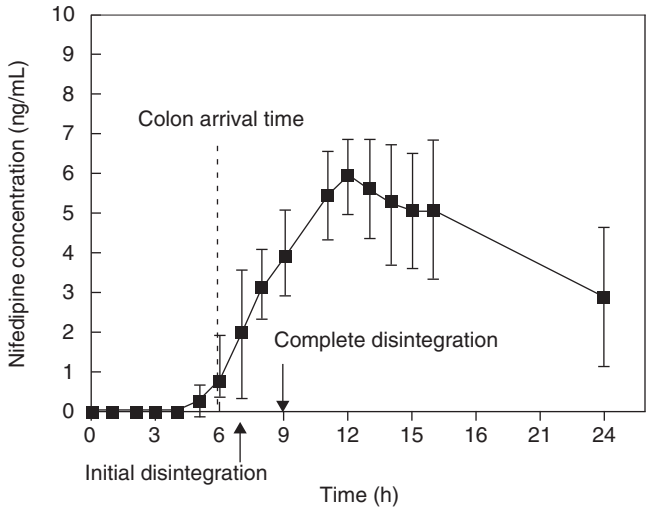


**FIGURE 10** Nifedipine plasma concentration–time profile from pectin–galactomannan-coated tablets. (Adapted from ref. 110 with permission of Elsevier Copyright 2002.)

A pectin-and-galactomannan coating was developed by Lee et al. [112]. It consists of a conventional tablet coated with pectin and galactomannan. The coating from aqueous solutions of pectin and galactomannan was shown to be strong, elastic, and insoluble in gastric fluid. Figure 10 shows the plasma concentration profile of nifedifine from pectin–galactomannan-coated tablets and associated in vivo transit and disintegration characteristics. The mean plasma concentration of nifedifine was negligible for more than 5 h postdose and then increased rapidly.

*CODES Technology* CODES is a unique colon-specific drug delivery technology that was designed to avoid the inherent problems associated with pH- or time-dependent systems [113, 114]. The design of CODES exploited the advantages of certain polysaccharides that are only degraded by microorganisms available in the colon [115]. This is coupled with a pH-sensitive polymer coating. Since the degradation of polysaccharides occurred only in the colon, this system exhibited the capability to achieve colon delivery consistently and reliably. As schematically presented in Figure 11, one typical configuration consists of a core table coated with three layers of polymer coatings. The first coating (next to the core tablet) is an acid-soluble polymer (e.g., Eudragit E) and an outer coating is enteric with a HPMC barrier layer in between the oppositely charged polymers. The polysaccharides, degradable by enteroorganisms, generate organic acid, including mannitol, maltose, lactulose, and fructooligosaccharides. During the transit through the GI tract, CODES remains intact in the stomach, but the enteric and barrier coatings disolve in the intestines. In vivo performance of CODES in beagle dogs was studied using acetaminophen as the model drug and lactulose as the matrix-forming excipient in the core tablet. Compared with enteric-coated tablet, the onset of acetaminophen release form CODES was delayed more than 3 h, as shown in Figure 12.

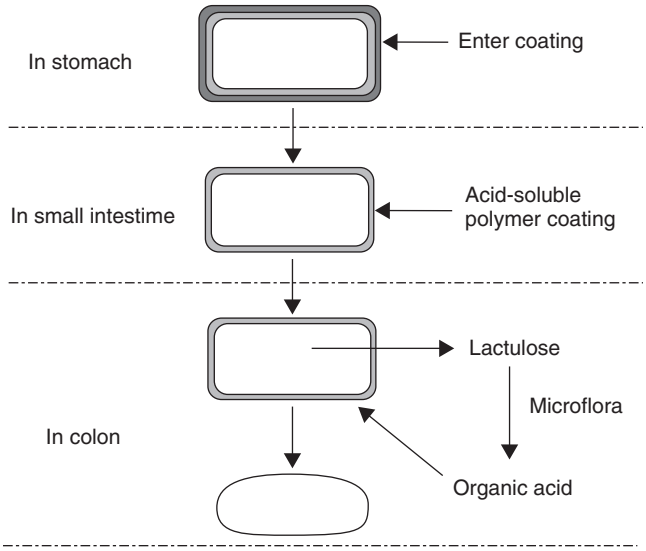


**FIGURE 11** Conceptual design of CODES technology. (Adapted from ref. 114 with permission of Elsevier Copyright 2002.)

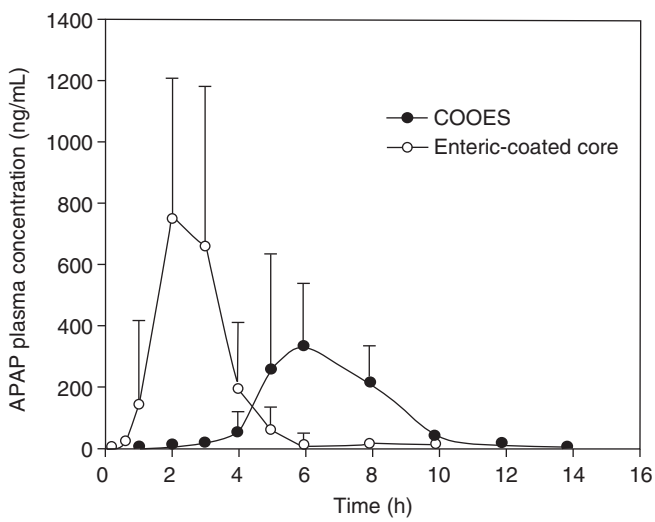


**FIGURE 12** Percentage of acetaminophen released from CODES and enteric-coated core tablets in beagle dogs. (Adapted from ref. 114 with permission of Elsevier Copyright 2002.)

***Thermoresponsive Drug Release Dosage Forms*** Temperature is the most widely utilized triggering signal for a variety of modulated or pulsatile drug delivery systems. The use of temperature as a signal has been justified by the fact that the body temperature often deviates from the physiological temperature (37°C) in the presence of pathogens or pyrogens. This deviation sometimes can be a useful stimulus that

achivates the release of therapeutic agents from various temperature-responsive drug delivery systems for diseases accompanying fever. The drug delivery systems that are responsive to temperature utilize various polymer properties, including the thermally reversible coil/globule transition of polymer molecules, swelling change of networks, glass transition, and crystalline melting.

Thermoresponsive hydrogels have been investigated as possible drug delivery carriers for stimuli-responsive drug delivery systems [116–118]. The common characteristics of temperature-sensitive polymers are the presence of hydrophobic groups, such as methyl, ethyl, and propyl groups. Of the many temperature-sensitive polymers, poly(*N*-isopropylacrylamide) (PIPPAm) is probably the most extensively used. PIPPAm cross-linked gels have shown thermoresponsive, discontinuous swelling/deswelling phases—swelling, for example, at temperatures below 32°C while shrinking above this temperature. A sudden temperature increase above the transition temperature of these gels resulted in the formation of a dense, shrunken layer on the gel surface, which hindered water permeation from inside the gel into the environment. Drug release from the PIPPAm hydrogels at temperatures below 32°C was governed by diffusion, while above this temperature drug release was stopped completely, due to the "skin layer" formation on the gel surface (on–off drug release regulation).

Swelling–deswelling kinetics of conventional cross-linked hydrogels are normally the reciprocal of the square of the gel dimension. This mobility of the cross-linked chains in the gel is affected by the surrounding chains and the swelling–deswelling phases of the gel are governed by the collective diffusions of the network chains. Thus, to accelerate structural changes of the gel in response to external stimuli, several approaches have been developed which form porous structure within the gel and decrease gel size. Kaneko et al. [119, 120] introduced a method to accelerate gel swelling–deswelling kinetics based on the molecular design of the gel structure by grafting the free mobile linear PIPPAm chains within the cross-linked PIPPAm hydrogels. These novel graft-type PIPPAm gels had the same transition temperature as conventional cross-linked PIPPAm gels and existed in the swollen state below the transition temperature, while above this temperature, they shrank. Adense skin layer formed on the conventional PIPPAm gels upon temperature change above the transition temperature, which limited the complete shrinkage of the gel. In contrast, the PIPPAm-grafted gels showed rapid deswelling kinetics without the formation of a skin layer on the gel surface. This is probably due to the rapid dehydration of the graft chains formed by hydrophobic aggregation on the three-dimensional cross-linked gel chains. The low-molecular-weight compounds released immediately from conventional PIPPAm gels after a temperature increase, after which the release was teminated due to the formation of a dense impermeable skin layer on the surface. In comparison, 65% of the drug was released in one burst from free PIPPAm-grafted hydrogels with a graft molecular weight (MW) of 9000 following the temperature increase. Graft-type gels with a molecular weight of 4000 showed oscillating drug release profiles. The release of high-molecular-weight compound (e.g., dextran, MW 9300) from PIPPAm graft-type gels was shown to burst after a temperature increase of 40°C. The difference in drug release profiles for two graft-type gels is probably due to the different strengths of aggregation forces between the formed hydrophobic cores within the graft-type gels. That is, the high-molecular-weight

graft chains formed more hydrophobic cores within the gels upon the temperature increase, which induced rapid gel deswelling.

Temperature-sensitive hydrogels can also be placed inside a rigid capsule containing holes or apertures. The on–off release is achieved by the reversible volume change of temperature-sensitive hydrogels. Such a device is called a squeezing hydrogel device because the drug release is affected by the hydrogel dimension. In addition to temperature, hydrogels can be made to respond to other stimuli, such as pH. In this type of system, the drug release rate was found to be proportional to the rate of squeezing of the drug-loaded polymer.

Clinical applications of thermosensitive hydrogels based on NIPAAm and its derivatives have limitations [121]. The monomers and cross-linkers used in the synthesis of the hydrogels are still not known to be biocompatible and biodegradable. The observation that acrylamide-based polymers activate platelets upon contact with blood, together with the unclear metabolism of poly(NIPAAm), requires extensive toxicity studies before clinical applications can merge.

Recently some studies have been conducted on anocomposite hydrogels for photothermally modulated drug delivery. Gold nanoshells can be designed to absorb light strongly at desired wavelengths, in particular, in the near infrared between 800 and 1200 nm, where tissue is relatively transparent [122]. When optically absorbing gold nanoshells are embedded in a matrix material, illuminating them at their resonance wavelength causes the nanoshells to transfer heat to their local environment. This photothermal effect can be used to optically modulate drug release from a nonshell polymer composite drug delivery system. To accomplish photothermally modulated release, the matrix polymer material must be thermally responsive. The authors observed the pulsatile release of insulin and other proteins in response to near-infrared irradiation when gold nanoshells were embedded in NIPAAm-*co*-acrylamide hydrogels [122].

***Electroresponsive Release*** An electric field as an external stimulus has advantages, such as the availability of equipment, which allow precise control with regard to the magnitude of current, duration of electric pulses, interval between pulses, and so on. Electrically responsive delivery systems are prepared from polyelectrolytes (polymers which contain relatively high concentration of ionizable groups along the backbone chain) and are thus pH responsive as well as electroresponsive. Under the influence of an electric field, electroresponsive hydrogels generally deswell or bend, depending on the shape of the gel that lies parallel to the electrodes, whereas deswelling occurs when the hydrogel lies perpendicular to the electrodes. Synthetic (e.g., acrylate and methacrylae derivatives) as well as naturally occurring polymers (including hyaluronic acid, chondroitin sulfate, agarose, carbomer, xanthan gum, and calcium alginate separately or in combination) have been used. Complex multicomponent gels or interpenetrating networks have been prepared in order to enhance the gels or interpenetrating networks have been prepared in order to enhance the gel's electroresponsiveness [123]. Electrically enhanced transdermal delivery provides the time-dependent delivery. Ionotophoresis, the electromigrational movement of charged molecules through the skin under a low-voltage and continuous driving force, has been reported for a number of bioactive compounds, such as leutinizaing hormone relesing hormone (LHRH), testosterone, and buserelin.

*Electronic Microelectromechanical System for Controlled Release* Electronic microelectromechanical devices are manufactured using standard microfabrication techniques that are used to create silicon chips for computers, and they often have moving parts or components that enable some physical or analytical function to be performed by the device. Microfabrication techniques, the same processing techniques used to make microprocessors for computers and other microelectronic devices, have been used increasingly to produce microscale devices whose primary functions are mechanical, chemical and optical in nature. Such devices are commonly referred to as *microelectromechanical systems* (MEMS) and are found in ink-jet printers, automotive applications, and microtube engines in the aerospace industry. MEMS for biological applications are classified as either microfluidic devices or nonmicrofluidic devices. The ultimate goal of MEMS is to develop a microfabricated device with the ability to store and release multiple chemical substances on demand by a mechanism devoid of moving its parts [124, 125]. A wide variety of microreservoirs, micropumps, cantilevers, rotors, channels, valves, sensors, and other structures have been fabricated, typically from the materials that have been demonstrated to be biocompatible and can be sterilely fabricated and hermetically sealed. The digital capabilities of MEMS may allow greater temporal control over drug release compared to traditional polymer-based systems, while the batch-processing techniques used in the microelectronics industry can lead to greater device uniformity and reproducibility than is currently available to the pharmaceutical industry. The use of MEMS for drug delivery necessitates the existence of drug depot or supply within or on the device. One straightforward approach to achieve this drug reservoir is the fabrication of silicon microparticles that contain an internal reservoir loaded with drug. These devices could be used for oral drug delivery, with release of the drug triggered by binding of a surface-functionalized molecule to cells in the digestive tract.

The completely implantable minipump made by Minimed has a pulsatile, radio-controlled injection rate through a catheter into the intraperitoneal region [126]. One study found that patients with the implantable pump did not differ from control subjects on any meansure of psychosocial function but that puump users monitored their blood glucose levels more frequently and had lower average blood glucose levels. Even though this type of device may improve patient's mbility and reduce infections by eliminating transcutaneous catheters, they may still be hampered by their size, cost, ability to deliver only drugs in solution, and limited stability of some drugs in solution at 37°C. Ikemoto and Sharpe [127] have developed a stepmotor micropump for the injection of nanoliter volumes of D-amphetamine solution into discrete brain regions of freely moving rats that was well tolerated. This micropump delivered a reliable volume of 50 nL per infusion over an hour at a rate of one infusion per minute.

Another development in MEMS technology is the microchip. The microchip consists of an array of reservoirs that extend through an electrolyte-impermeable substrate. The prototype microchip is made of silicon and contains a number of drug reservoirs. Each reservoir is sealed at one end by a thin gold membrane of material that serves as an anode in an electrochemical reaction and dissolves when an electric potential is applied to it in an electrolyte solution. The reservoirs are filled with any combination of drug or drug mixtures in any form (i.e., solid, liquid, or gel) through the opening opposite the anode membrane by ink-jet printing or microinjection and

are then sealed with a waterproof material. A cathode is also required for the electrochemical reaction to take place, and the cathode is usually made of the same conductive material as the anode to simplify the fabrication procedure. The device is submerged in an electrolyte solution containing ions and upon electric stimulation forms a soluble complex with the anode in its ionic form. When release is desired, an electric potential is applied between an anode membrane and a cathode, and the gold membrane anode is dissolved within 10–20 s and allows the drug in the reservoir to be released. This electric potential causes oxidation of the anode material to form a soluble complex with the electrolytes which when dissolves allowing release of the drug. Complex release patterns (such as simultaneous constant and pulsatile release) can be achieved from the microchips. The microchip has the ability to control both release time and release rate. The rate of release from a reservoir is a function of the dissolution rate of the materials in the reservoir, the diffusion rate of these materials out of the reservoir, or both. Therefore, the release rate from an individual reservoir can be tailored to a particular application by proper selection of the materials placed inside the reservoir [e.g., pure drug(s), drugs with polymers] [124, 125].

A microchip with insulin-filled reservoirs could eventually provide a better alternative for the treatment of insulin-dependent diabetes mellitus (IDDM) [125]. Because the microchip is capable of being programmed as well as integrated with other electronic devices, it is supposable that the microchip could be incorporated into a closed-loop biofeedback system. An electronic apparatus that continuously measures the blood glucose levels could provide the stimulus to the microchip and result in release of insulin into the bloodstream. Although such a system could still not perfectly mimic an endogenous system of healthy person, it could practically meet the needs of IDDM patients. Pulsatile release of synthetic gonadotropin–releasing hormone (GnRH) can be achieved with a programmed microchip. A subcutaneous implanted microchip containing 1000 drug reservoirs would be adequate to administer a month's worth of drug therapy. The implanted microchip would be a convenient means to achieve the desired pharmacotherapeutic outcome of ovulation without interfering with the patient's daily activities or causing phlebitis.

While microchip drug delivery would be the most technologically advanced delivery system, it has itself limited storage capacity for therapeutic drugs [125]. Because most applications of this technology require implantation within bodily tissues, the question arises, "What would be done when the chip runs out of drug?" Some sort of procedure would be required to retrieve the empty chip cartridge once it has emptied. Due to the limited quantity of drug that can be stored on one chip, this technology is only ideal for potent drugs. If a larger dose of a medication is required, the chip would not be adequate for dispensing larger quantities of drug.

***Magnetically Induced Release*** Magnetic carriers receive their magnetic response to a magnetic field from incorporated materials such as magnetite, iron, nickel, and cobalt. For biomedical applications, magnetic carriers must be water based, biocompatible, nontoxic, and nonimmunogenic. Earlier, Langer et al. [128] embedded magnetite or iron beads into a drug-filled polymer matrix and then showed that they could activate or increase the release of the drug from the polymer by moving a magnet over it or by applying an oscillating magnetic field. When the frequency of

the applied field was increased from 5 to 11 Hz, the release of BSA from ethylenevinylacetate copolymer (EVAc) matrices slowed in a linear fashion. The rate of release could be modulated by altering the position, orientation, and magnetic strength of the embedded materials as well as by changing the amplitude of frequency of the magnetic field. The micromovement within the polymer produced microcracks in the matrix and thus made the influx of liquid, dissolution, and efflux of the drug. Done repeatedly, this would allow the pulsatile delivery of insulin.

Another mechanistic approach based on magnetic attraction is the slowing down of oral drugs in the gastrointestinal system. This is possible by filling an additional magnetic component into capsules or tablets. The speed of travel through the stomach and intestines can then be slowed down at specific positions by an external magnet, thus changing the timing and/or extent of drug absorption into stomach or intestines. Slowing down the passage of magnetic liposomes with a magnet actually increased the blood levels of drug. Babincova et al. [129] developed magnetoliposomes for triggered release of drug. In their delivery systems, they entrapped dextran–megnetite and model drug 6-carboxyfluorescein in the liposomes and used laser to trigger the release of drug. The magnetite absorbs the laser light energy to heat the lipid bilayer above the gel–liquid crystal-phase transsition temperature $T_c$, which is 41°C for dipalmitoyl-phosphatidylcholine. Liposomes made from this lipid release their content as soon as the temperature reaches this level. They have also suggested that the absorption of laser energy by magnetite particles provides a means for localized heating and controlled release of liposome with a single laser pulse. This may have potential applications for selective drug delivery, especially to the eyes and skin. Even though the magnetic-modulated therapeutic approach is promising, it still needs very careful attention for a number of physical and magnetism-related properties. The magnetic force, which is defined by its field and field gradient, needs to be large and carefully shaped to activate the delivery system within the target area. The magnetic materials should be tissue stable and compatible.

### *Chemically Induced Release*

*Gluose-Responsive Insulin Release Device* A decrease in or the absence of insulin secretion from pancreatic islets is the cause of diabetes mellitus. An effective glucose-responsive insulin delivery system should be composed of a glucose-sensing component and an insulin-releasing component. The sensing component detects a change in the glucose level and produces a signal that affects the releasing component. The magnitude of the signal increases with increasing glucose concentration, and so does the rate of insulin release. Based on this principle, various polymer-based glucose-responsive delivery systems have been designed, most of which are hydrogels that can alter their volume and degree of hydration in response to glucose concentration. Several systems have already been developed which are able to respond to glucose concentration changes, such as glucose oxidase (GOD), which catalyzes glucose oxidation [130]. Glucosylated insulin bound to concanavalin (Con) A was released through exchange with external glucose, due to the difference in their binding constants. This system needs direct injection of microcapsules into the peritoneal cavity of patients, which may cause undesirable side effects arising from the immune response to Con A if Con A was directly exposed to immune systems after breakage

of microcapsules. Obaidat and Park [131] prepared a copolymer of acrylamide and allyl glucose. The side-chain glucose units in the copolymer were bound to Con A. These hydrogels showed a glucose-responsive, sol–gel phase transition dependent upon the external glucose concentration. The nonlinear dependence of this sol–gel phase transition with regard to the glucose concentration was due not only to the increased binding affinity of allyl glucose to Con A compared to native glucose but also to the cooperative interaction between glucose-containing copolymer and Con A. Kataoka et al. [132] developed glucose and thermoresponsive hydrogels using acrylamidophenylboronic acid and *N*-isopropylacrylamide (IPAAm). The obtained gels, containing 10 mol % phenylboronic acid moieties, showed a transition temperature of 22°C in the absence of glucose. Below this temperature, the gels existed in a swollen state. The introduction of glucose to the medium altered the transition temperature of the gels in such a way that the transition temperature increased with increasing glucose concentration to reach 36°C at 5 g/L glucose concentration. Boronic acid was in equilibrium between the undissociated and dissociated forms. With increasing glucose concentration, the equilibrium shifted to increase the amount of dissociated boronate groups and gels became more hydrophilic. Although all of the glucose-sensitive insulin delivery systems are elegant and highly promising, many improvements need to be made for them to become clinically useful. First of all, the response of these systems upon changes in the environment occurs too slowly. In clinical situations, these systems need to respond to ever-changing glucose concentrations at all times, requiring hydrogels that can respond reproducibility and with rapid-response onset times on a long-term basis. An additional constant is that all the components used in the systems must be biocompatible.

*Chemotactic Factor-Induced Controlled-Release Systems* With physical or chemical stress such as injury and broken bones, an inflammation reaction takes place at the injured site. At the inflammatory sites, inflammation-responsive phagocytic cells such as macrophages and polymorphonuclear cells play a role in healing the injury. During inflammation, hydroxyl radicals ($OH^{\bullet}$) are produced from the cells. Yui and co-workers [133, 134] developed inflammatory-induced hydroxyl radicals and designed drug delivery system which responded to the hydroxyl radicals and degraded in a limited manner. They used hyaluronic acid (HA), a linear aminopolysaccharide composed of repeating disaccharide subunits of *N*-acetyl-D-glucosamine and D-guluronic acid. In the body, HA is mainly degraded by hyaluronidase, or hydroxyl radicals. Degradation of HA via the enzyme is very low in a normal state of health. Degradation via hydroxyl radicals, however, is usually dominant and rapid when HA is injected at inflammatory sites. These authors prepared cross-linked HA with ethyleneglycol diglydylether or polyglycerol polygluycidalether. These HA gels degraded only when the hydroxyl radicals were generated through the reaction between the iron ($Fe^{2+}$) ions and the hydrogen peroxide in vitro. Thus, a surface erosion type of degradation was achieved. When microspheres were incorporated in the HA hydrogels as a model drug, these microspheres were released when hydroxyl radicals induced HA gel degradation. Furthermore, degradation of HA in vivo tests showed that HA gels are degraded only when inflammation was induced by surgical incision. Control HA gels were stable over 100 days. It is possible to treat locally in inflammatory diseases such as rheumatoid arthritis using anti-inflammatory drug incorporated in HA gels [135].

## 11 SUMMARY

As pharmaceutical scientists have increased their knowledge of pharmacokinetics and pharmacodynamics, it has become apparent that these factors can result in more efficacious drugs. The number of new drug entities appearing on the market yearly has declined and pharmaceutical manufacturers have shown a renewed interest in improving existing dosage forms and developing more sophisticated drug delivery systems, including those employing the principles of controlled drug release.

Current research in this area involves numerous new and novel systems, many of which have strong therapeutic potential. In this chapter, we have tried to emphasize the importance of oral routes as well as others, such as ocular, transdermal, intrauterine, and vaginal. The various microencapsulation, nanoencapsulation, and liposome technologies and the release of drugs and bioactive compounds from such products have been described.

## REFERENCES

1. Bhalla, H. L. (1999), Drug delivery research in India: A challenge and an opportunity, *J. Controlled Release*, 62, 65–68.
2. Breimer, D. D. (1999), Future challenges for drug delivery, *J. Controlled Release*, 62, 3–6.
3. Chien, Y. W. (1992), *Novel Drug Delivery Systems*, Marcel Dekker, New York, pp. 1, 2.
4. Brannon-Peppas, L. (1997), Polymers in controlled drug delivery, *Biomaterials*, 11, 1–14.
5. Kydoneieus, A. F. (1980), *Controlled Release Technologies: Methods, Theory, and Applications*, CRC Press, Boca Raton, FL.
6. Murano, E. (1998), Use of natural polysaccharides in the microencapsulation techniques, *J. Appl. Ichthyol.* 14, 245–249.
7. Junginger, H. E., and Verhoef, J. C. (1998), Macromolecules as safe penetration enhancers for hydrophilic drugs: A fiction, *Pharm. Sci. Today*, 1, 375–376.
8. Robinson, J. R., and Lee, V. H. L. (1987), *Controlled Drug Delivery: Fundamentals and Applications*, Marcel Dekker, New York.
9. Hofman, A. F., Pressman, J. H., Code, C. F., and Witztum, K. F. (1983), Controlled entry of orally-administered drugs: Physiological considerations. *Drug Dev. Ind. Pharm.*, 9, 1077.
10. Manninen, V., Melin, J., and Reissel, P. (1972), Tablet disintegration: Possible link with biological availability of digoxin, *Lancet*, 26, 490–491.
11. Fincher, J. H. (1968), Particle size of drugs and its relationship to absorption and activity, *J. Pharm. Sci.*, 57, 1825–1835.
12. Lee, V. H., and Robinson, J. R. (1987), *Sustained and Controlled Release Drug Delivery Systems*, Marcel Dekker, New York, p. 71.
13. Fix, J. A. (1996), Oral controlled release technology for peptides: Status and future prospectus. *Pharm. Res.*, 13, 1760–1763.
14. Chien, Y. W. (1983), Logics for transdermal controlled drug administration, *Drug Dev. Ind. Pharm.*, 9, 497–520.
15. Gandhi, R., Kaul, C. L., and Panchangula, R. (1999), Extrusion and spheronization in the development of oral controlled-release dosage forms, *Pharmaceutical Science and Technology Today*, 2, 160–170.

16. Anal, A. K. (2007), Time-controlled pulsatile delivery systems for bioactive compounds, Recent Patents Drug Deliv. *Formulat.*, 1, 73–79.
17. Chan, O. H., and Stewart, B. H. (1996), Physico-chemical and drug delivery considerations for oral drug bioavaibility. *Drug Discovery Today*, 1, 461–472.
18. Wilson, C. G. (2000), Gastrointestinal transit and drug absorption, in *Oral Drug Absorption: Prediction and Assessment*, Marcel Dekker, New York, pp. 1–9.
19. Knutson, L., Knutson, F., and Knutson, T. (2000), Permeability in the gastrointestinal tract, in *Oral Drug Absorption: Prediction and Assessment*, Marcel Dekker, New York, pp. 11–16.
20. Waterbeemd, H. V. (2000), Intestinal permeability: Prediction from theory, in *Oral Drug Absorption: Prediction and Assessment*, Marcel Dekker, New York, pp. 31–49.
21. Chien, Y. W. (1992), *Novel Drug Delivery Systems*, Vol. I, Marcel Dekker, New York, pp. 139–196.
22. Jantzen, G. M., and Robinson, J. R. (2000), Sustained and controlled release drug delivery systems, in *Modern Pharmaceutics*, Marcel Dekker, New York, pp. 501–528.
23. Welling, P. G. (1983), Oral controlled drug adminstration: Pharmacokinetic considerations. *Drug Dev. Ind. Pharm.*, 9, 1185–1125.
24. Hofman, A. F., Dressman, J. H., Code, C. F., and Witztum, K. F. (1983), Controlled entry of oral administered drugs: Physiological considerations, *Drug Dev. Ind. Pharm.*, 9, 1077–1083.
25. Fara, J. W. (1983), Gastrointestinal transit of solid dosage forms. *Pharm. Technol.*, 7, 23–26.
26. Meyer, J. H., Ohashi, H., Jehn, D., and Thomson, J. B. (1981), Size of particles emptied from the human stomach, *Gastroenterology*, 80, 1489–1496.
27. Welling, P. G., and Dobrinska, M. R. (1987), Dosing considerations and bioavailability assessment of controlled drug delivery systems, in *Controlled Drug Delivery: Fundamentals and Applications*, Marcel Dekker, New York, pp. 253–289.
28. Hinton, J. M., Lennard-Jones, J. E., and Young, A. C. (1969), Alimentary canal transit time for an inert object in humans, *Gut*, 10, 842.
29. Frier, M., and Perkins, A. C. (1994), Radiopharmaceuticals and the gastrointestinal tract, *Eur. J. Nucl. Med. Mol. Imaging*, 21, 1234–1242.
30. Basit, A. W., Podczeck, F., Newton, M., Waddington, W. A., Ell, P. J., and Lacey, L. F. (2004), The use of formulation technology to assess regional gastrointestinal drug absorption in humans, *Eur. J. Pharm. Sci.*, 21, 179–189.
31. Delie, F., and Blanco-Príeto, M. J. (2005), Polymeric particulares to improve oral bioavailability of peptide drugs, *Molecules*, 10, 65–80.
32. Walter, E., Kissel, T., and Amidon, G. E. (1996), The intestinal peptide carrier: A potential transport system for small peptide-derived drugs, *Adv Drug Deliv. Rev.*, 20, 33–58.
33. Scheline, R. R. (1973), Metabolism of foreign compounds by gastrointestinal microorganisms, *Pharmacol. Rev.*, 25, 451–523.
34. Edwards, C. (1997), Physiology of colorectal barrier. *Adv. Drug Deliv. Rev.*, 28, 173–190.
35. Whateley, T. L. (1992), *Microencapsulation of Drugs*, Harwood Academic, Geneva, Switzerland.
36. Anal, A. K., and Stevens, W. F. (2005), Chitosan-alginate multilayer beads for controlled release of ampicillin, *Int. J. Pharm.*, 290, 45–54.
37. Anal, A. K., Bhopatkar, D., Tokura, S., Tamura, H., and Stevens, W. F. (2003), Chitosan-alginate multilayer beads for gastric passage and controlled intestinal release of protein, *Drug Dev. Ind. Pharm.*, 29, 713–724.

38. Hui, H. W., Lee, V. H. L., and Robinson, J. R. (1987), Design and fabrication of oral controlled release drug delivery systems, in *Controlled Drug Delivery: Fundamentals and Applications*, Marcel Dekker, New York, pp. 373–421.
39. Anal, A. K., and Singh, H. (2007), Recent advances in microencapsulation of probioitcs for industrial applications and targeted delivery, *Trends Food Sci. Technol.*, 18, 240–251.
40. Mathir, Z. M., Dangor, C. M., Govender, T., and Chetty, D. J. (1997), In vitro characterization of a controlled-release chlorpheniramine maleate delivery system prepared by the air-suspension technique, *J. Microencapsul.*, 14, 743–751.
41. Deasy, P. B. (1991), Microencapsulation of drugs by pan and air suspension techniques, *Crit. Rev. Ther. Drug Carrier Syst.*, 8, 39–89.
42. O'Connor, R. E., and Schwartz, J. B. (1989), Extrusion and spheronization technology, in *Pharmaceutical Pelletization Technology*, Marcel Dekker, New York, pp. 187–215.
43. Gopferich, A., Alonso, M. J., and Langer, R. (1994), Development and characterization of microencapsulated microspheres, *Pharm. Res.*, 11, 1568–1574.
44. Anal, A. K., Stevens, W. F., and López, C. R. (2006), Ionotropic cross-linked chitosan microspheres for controlled release of ampicillin, *Int. J. Pharm.*, 312, 166–173.
45. Weidenauer, U., Bodmeier, D., and Kissel, T. (2003), Microencapsulation of hydrophilic drug substances using biodegradable polyesters. Part I: Evaluation of different techniques for the encapsulation of pamidronate di-sodium salt, *J. Microencapsul.*, 20, 509–524.
46. Cui, J. H., Goh, J. S., Park, S. Y., Kim, P. H., and Le, B. J. (2001), Preparation and physical characterization of alginate microparticles using air atomization method, *Drug Dev. Ind. Pharm.*, 27, 309–319.
47. Purvis, T., Vaughan, J. M., Rogers, T. L., Chen X., Overhoff, K. A., Sinswat, P., Hu, J., McConville, J. T., Johnston, K. P., and Williams, R. O. (2006), Cryogenic liquids, nanoparticles, and microencapsulation, *Int. J. Pharm.*, 4, 43–50.
48. Charman, W. N., Chan, H. K., Finnin, B. C., and Charman, S. A. (1999), Drug delivery: A key factor in realizing the full therapeutic potential of drugs, *Drug Dev. Res.*, 46, 316–327.
49. Kayser, O., Lemke, A., and Hernández-Trejo, N. (2005), The impact of nanobiotechnolgy on the development of new drug delivery systems, *Curr. Pharm. Biotechnol.*, 6, 3–5.
50. Roco, M. C. (2003), Nanotechnology: Convergence with modern biology and medicine, *Curr. Opin. Biotechnol.*, 14, 337–346.
51. Merck & Co. (2004), Drug Information: Emend, capsules, available: www.merck.com.
52. Liversidge, G. G., Cundy, K. C., Bishop, J. F., and Czekai, D. A. (1992), Surface modified drug nanoparticles, Un. St. Patent 5,145,684, Sterling Drug, New York.
53. Merisko-Liversidge, E., Liversidge, G. G., and Cooper, E. R. (2003), Nanosizing: A formulation approach for poorly water-soluble compounds, *Eur. J. Pharm. Sci.*, 18, 113–120.
54. Merisko-Liversidge, E., Sparpotdar, P., and Bruno, J. (1996), Formulation and antitumor activity evaluation of nanocrystalline suspension of poorly soluble anticancer drugs, *Pharm. Res.*, 13, 272–278.
55. Buchmann, S., Fischli, W., Thiel, F. P., and Alex, R. (1996), Aqueous microsuspension, an alternative intravenous formulation for animal studies, paper presented at the 42nd Annual Congress of the International Association for Pharmaceutical Technology, Mainz, p. 124.
56. Müller, R. H., and Heinemann, S. (1989), Surface modeling of microparticles as parenteral systems with high tissue affinity, in *Bioadhesion-Possibilities and Future Trends*, Wissenschaftliche Verlagsgesellschaft, Stuttgart, pp. 202–214.

57. Müller, R. H., and Peters, K. (1998), Nanosuspensions for the formulation of poorly soluble drugs: I: Preparation by a size-reduction technique, *Int. J. Pharm.*, 160, 229–237.
58. Maa, Y. F., and Hsu, C. C. (1998), Performance of sonication and microfluidisation for liquid-liquid emulsification, *Pharm. Dev. Technol.*, 4, 233–240.
59. Roller, J. M., Covereur, P., Robolt-Treupel, L., and Puisieux, F. (1986), Physicochemical and morphological characterization of polyisobutyl cyanoacrylates nanocapsules, *J. Pharm. Sci.*, 75, 361.
60. Chouinard, F., Kan, F. W., Leroux, J. C., Foucher, C., and Lenaerts, V. (1991), Preparation and purification of polyisohexylcyanoacrylate nanocapsules, *Int. J. Pharm.*, 72, 211.
61. Ammoury, N., Fessi, H., Devissaguet, J. P., Ouisieux, F., and Benita, S. (1989), Physicochemical characterization of polymeric nanocapsules and in vitro release evaluation of indomethacin as a drug model, *STP Pharma*, 5, 642.
62. Gallardo, M. M., Courraze, G., Denizor, B., Treupel, L., Couvreur, P., and Puisieux, F. (1993), Preparation and purification of isohexylcyanoacrylate nanocapsules, *Int. J. Pharm.*, 100, 55–64.
63. Chouinard, F., Buczkowski, S., and Lenaerts, V. (1994), Poly(alkylcyanoacrylate) nanocapsules: Physico-chemical characterization and mechanism of formation, *Pharm. Res.*, 11, 869.
64. Wohlgemuth, M., Machtle, W., and Mayer, C. (2000), Improved preparation and physical studies of polybutylcyanoacrylate nanocapsules, *J. Microencapsul.*, 17, 437.
65. Lambert, G., Fattal, E., Pinto-Alphandary, H., Gulik, A., and Couvereur, P. (2000), Polyisobutylcyanoacrylate nanocapsules containing an aqueous core the delivery of oligonucleotides, *Int. J. Pharm.*, 214, 13.
66. Couvreur, P., Barratt, G., Fattal, E., Legrand, P., and Vauthier, C. (2002), Nanocapsule technology: A review. *Crit. Rev. Therap. Drug Carrier Syst.*, 19, 99–134.
67. Quintanar-Guerrero, D., Allemann, E., Doelker, E., and Fessi, H. (1998), Preparation and characterization of nanocapsules from performed polymers by a new process based on emulsification-diffusion technique, *Pharm. Res.*, 15, 1056.
68. Ravi, K. M. N. (2000), Nano and microparticles as controlled drug delivery devices, *J. Pharm. Sci.*, 3, 234–258.
69. Torchilin, V. P. (2005), Recent advances with liposomes as pharmaceutical carriers, *Nature Rev.: Drug Discovery*, 4, 145–159.
70. Leserman, L. (2004), Liposomes as protein carriers in immunology, *J. Liposome Res.*, 14, 175–189.
71. Uchegbu, I. F., and Vyas, S. P. (1998), Non-ionic surfactant based vesicles (niosomes) in drug delivery, *Int. J. Pharm.*, 172, 33–70.
72. Uchegbu, I. F. (2000), Synthetic surfactant vesicles: Niosomes and other non-phospholipid vesicular systems, in *Drug Targeting and Delivery*, Vol. 11, Harwood Academic, Amsterdam.
73. Bodor, N., and Loftsson, T. (1987), Novel chemical approaches for sustained drug delivery, in *Controlled Drug Delivery: Fundamentals and Applications*, Marcel Dekker, New York, pp. 337–369.
74. Sugibayashi, K., and Morimoto, Y. (1994), Polymers for transdermal drug delivery systems, *J. Controlled Release*, 29, 177–185.
75. Chien, Y. W. (1987), Transdermal therapeutic systems, in *Controlled Drug Delivery: Fundamentals and Applications*, Marcel Dekker, New York, pp. 524–549.
76. Chien, Y. W. (1982), Ocular controlled release drug administration, in *Novel Drug Delivery Systems*, Marcel Dekker, New York, pp. 13–48.

77. Gelatt, K. N., Gum, G. G., williams, L. W., and Peiffer, R. L. (1979), Evaluation of a soluble sustained-release ophthalmic delivery unit in the dog, *Am. J. Vet. Res.*, 40, 702–704.
78. Chien, D. S., and Schoenwald, R. D. (1990), Ocular pharmacokinetics and pharmacodynamics of phenylephrine and phenylephrine oxazolidine in rabbit eyes, *Pharm. Res.*, 7, 476–483.
79. Vyas, S. P., Ramchandraiah, S., Jain, C. P., and Jain, S. K. (1992), Polymeric pseudolatices bearing pilocarpine for controlled ocular delivery, *J. Microencapsul.*, 9, 347–355.
80. Lyrenäs, S., Clason, I., and Ulmsten, U. (2001), In vivo controlled release of PGE2 from a vaginal insert (0.8 mm, 10 mg) during induction of labor, *Br. J. Obstet. Gynaecol.*, 108, 169–178.
81. Sherwood, J. K., Zeitlin, L., Whaley, K. J., Richard, A. C., and Saltzman, M. (1996), Controlled release of antibodies for long-term topical passive immunoprotection of female mice against genital herpes, *Nat. Biotechnol.*, 14, 468–471.
82. Nilsson, C., Lachteenmaki, P., and Luukkainen, T. (1980), Patterns of ovulation and bleeding with a low levonregesterol-releasing device, *Contraception*, 21, 225–233.
83. van Laarhoven, J., Kruft, M., and Vromans, H. (2002), In vitro release properties of etonogestrel and ethynylestradiol from a contraceptive vaginal ring, *Int. J. Pharm.*, 232, 163–173.
84. Rathbone, M. J., Macmillan, K. L., JöChle, W., Boland, M., and Inskeep, E. K. (1998), Controlled release products for the control of the estrous cycle in cattle, sheep, goats, deer, pigs, and horses, *Crit. Rev. Ther. Drug Carrier Syst.*, 15, 285–380.
85. Strix, J. (1999), German Patent Application DE98-19809243.
86. Duncan, R., and Seymour, L. (1989), Controlled Release Technolgies, Elsevier Advanced Technology, Amsterdam, 11.
87. Lee, E. S., Kim, S. W., Kim, S. H., Cardinal, J. R., and Jacobs, H. (1980), Drug release from hydrogel devices with rate-controlling barriers, *J. Membr. Sci.*, 7, 293–303.
88. El-Nokaly, M. A., Piatt, D. M., and Charpentier, B. A. (1993), *Polymeric Delivery Systems: Properties and Applications*, ACS Symposium Series, Washington, DC.
89. Hui, H. W., Lee, V. H. L., and Robinson, J. R. (1987), Design and fabrication of oral controlled release drug delivery systems, in *Controlled Drug Delivery: Fundamentals and Applications*, Marcel Dekker, New York, pp. 373–421.
90. Sangalli, M. E., Maroni, A., Zema, L., Busseli, C., Giordano, F., and Gazzaniga, A. (2001), *In vitro* and *in vivo* evaluation of oral system for time and/or site specific drug delivery, *J. Controlled Release*, 73, 103–110.
91. Sangalli, M. E., Maroni, A., Foppoli, A., Zema, L., Giordano, F., and Gazzaniga, A. (2004), Different HPMC viscosity grades as coating agents for an oral time and/or site-controlled delivery systems: A study on process parameters and in vitro performances, *Eur. J. Pharm. Sci.*, 22, 469–476.
92. Pozzi, F., Furlani, P., Gazzaniga, A., Davis, S. S., and Wilding, I. R. (1994), The TIME CLOCK® system: A new oral dosage form for fast and complete of drug after a predetermined lag time, *J. Controlled Release*, 31, 99–108.
93. Medlicott, N. J., and Tucker, I. J. (1999), Pulsatile release from subcutaneous implants, *Adv. Drug Deliv. Rev.*, 38, 139–149.
94. Bussemer, T., Peppas, N. A., and Bodmeier, R. (2003), Evaluation of the swelling, hydration and rupturing properties of the swelling layer of a rupturable pulsatile drug delivery system, *Eur. J. Pharm. Biopharm.*, 56, 261–270.
95. Bussemer, T., Peppas, N. A., and Bodmeier, R. (2003), Time-dependent mechanical properties of polymers coating used in rupturable pulsatile release dosage forms, *Drug Dev. Ind. Pharm.*, 29, 623–630.

96. Sungthongjeen, S., Puttipipatkhachorn, S., Paeratakul, O., Dashevsky, A., and Bodmeier, R. (2004), Development of pulsatile release tablets with swelling and rupturable layers, *J. Controlled Release*, 95, 147–159.
97. Niwa, K., Takaya, T., Morimoto, T., and Takada, K. (1995), Preparation and evaluation of a time-controlled release capsule made of ethylcellulose for colon delivery of drugs, *J. Drug Target.*, 3, 83–89.
98. Krogel, I., and Bodmeier, R. (1997), Pulsatile drug release from an insoluble capsule body controlled by an erodible plug, *Pharm. Res.*, 15, 474–481.
99. Krogel, I., and Bodmeier, R. (1999), Evaluation of an enzyme-containing capsular shaped pulsatile drug delivery system, *Pharm. Res.*, 16, 1424–1429.
100. Gohel, M. C., and Sumitra, M. (2002), Modulation of active pharmaceutical material release from a novel "tablet in capsule system" containing an effervescent blend, *J. Controlled Release*, 79, 157–164.
101. Linkwitz, A., Magruder, J. A., and Merrill, S. (1994), Osmotically driven delivery device with expandable orifice for pulsatile delivery effect, U.S. Patent 5,318,558.
102. Qui, L. Y., and Zhu, K. J. (2001), Design of core-shelled polymer cylinder for potential programmable drug delivery, *Int. J. Pharm.*, 219, 151–160.
103. Jiang, H. L., and Zhu, K. J. (1999), Preparation, characterization and degradation characteristics of polyanhydrides containing polyethlene glycol, *Poly. Int.*, 48, 47–52.
104. Jiang, H. L., and Zhu, K. J. (2000), Pulsatile protein release from a laminated device comprising of polyanhydride and pH-sensitive complexes, *Int. J. Pharm.*, 194, 51–60.
105. Anal, A. K. (2007), Stimuli-induced pulsatile or triggered release of bioactive compounds, Recent Patents Endocrine, *Metabolic Immune Drug Discovery*, 1, 83–90.
106. Rubinstein, A. (2005), Colonic drug delivery, *Drug Discovery Today: Technol.*, 2, 33–37.
107. Deshpande, A. A., Rhodes, C. T., Shah, N. H., and Mallick, A. W. (1996), Controlled-release drug delivery systems for prolonged gastric residence: An overview, *Drug Dev. Ind. Pharm.*, 22, 531–539.
108. Rubinstein, A. (1995), Approaches and opportunities in colon-specific drug delivery. *Curr. Rev. Ther. Drug Carrier Syst.*, 12, 101–149.
109. Basit, A., and Bloor, J. (2003), Perspectives on colonic drug delivery, *Pharmatech*, 185–190.
110. Vandamme, T. F., Lenoury, A., Charrueau, C., and Chaumeil, J.-C. (2002), The use of polysaccharides to target drugs to the colon. *Carbohydr. Poly.*, 48, 219–231.
111. Sinha, V. R., and Kumria, R. (2001), Polysaccharides in colon-specific drug delivery, *Int. J. Pharm.*, 224, 19–38.
112. Lee, S., Lim, C. B., Lee, S., and Pai, C. M. (1999), Colon selective drug delivery composition, WO 99/01115.
113. Watanabe, S., Kawai, H., Katsuma, M., and Fukul, M. (2002), Colon-specific drug release system, U.S. Patent 6,368,629.
114. Takemura, S., Watanabe, S., Katsuma, M., and Fukul, M. (2002), Gastrointestinal transit study of a novel colon delivery system (CODES™) using gamma scintigraphy, *Proc. Int. Symp. Controlled Release Bioactive Mater.*, 27, 445–446.
115. Li, J., Yang, L., Ferguson, S. M., Hudson, T. J., Watanabe, S., Katsuma, M., and Fix, J. A. (2002), In vitro evaluation of dissolution behaviour for a colon-specifc drug delivery system (CODESTM) in multi-pH media using USP apparatus II and III, *AAPS Pharm. Sci. Technol.*, 3, article 33, available: www.aapspharmascitech.org.
116. Bae, Y. H., Okano, T., and Kim, S. W. (1991), "On-off" thermocontrol of solute transport. I. Temperature dependence of swelling of *N*-isopropylacrylamide networks modified with hydrophobic components in water, *Pharm. Res.*, 8, 531–537.

117. Bae, Y. H., Okano, T., and Kim, S. W. (1991), "On-off" thermocontrol of solute transport. II. Temperature dependence of swelling of *N*-isopropylacrylamide networks modified with hydrophobic components in water, *Pharm. Res.*, 8, 624–628.
118. Okano, T., Bae, Y. H., Jacobs, H., and Kim, S. W. (1990), Thermally on-off switching polymers for drug permeation and release, *J. Controlled Release*, 11, 255–265.
119. Kaneko, Y., Sakai, K., Kikuchi, A., Yoshida, R., Sakurai, Y., and Okano, T. (1995), Influence of freely mobile grafted chain length on dynamic properties of comb-type grafted poly(*N*-isopropylacrylamide) hydrogels, *Macromolecules*, 28, 7717–7723.
120. Kaneko, Y., Sakai, K., Kikuchi, A., Sakurai, Y., and Okano, T. (1996), Fast swelling/deswelling kinetics of comb-type grafted poly(*N*-isopropylacrylamide hydrogels). *Macromol. Chem.*, 109, 41–53.
121. Bromberg, L. E., and Ren, E. S. (1998), Thepertaure-responsive gels and thermo gelling polymer matrices for protein and peptide delivery, *Adv. Drug Deliv. Rev.*, 31, 197–221.
122. Sershen, S. R., Westcott, S. L., Halas, N. J., and West, N. J. (2000), Temperature-sensitive polymer-nanoshell composites for photothermally modulated drug delivery, *J. Biomed. Mater. Res.*, 51, 293–298.
123. Yuk, S. H., Cho, S. H., and Lee, H. B. (1992), Electric current-sensitive drug delivery systems using sodium-alginate/polyacrylic acid composites, *Pharm. Res.*, 9, 955–957.
124. Santini, J. T., Cima, M. J., and Langer, R. (1999), A controlled-release microchip, *Nature*, 397, 335–338.
125. Santini, J. T., Richards, A. C., Schiedt, R., Cima, M. J., and Langer, R. (2000), Microchips as controlled-drug delivery devices, *Chem. Int. Ed.*, 39, 2396–2407.
126. Barrett, D. H., Davidson, P. C., Steed, L. J., Abel, G. G., Loman, K. E., and Saudek, C. D. (1995), Evaluation of the psychosocial impact of the minimed variable-rate implanatable insulin pump, *Med. J.*, 88, 1226–1230.
127. Ikemoto, S., and Sharpe, L. G. (2001), A head-attachable device for injecting anolitre volumes of drug solutions into brain sites of freely moving rats, *J. Neurosci. Meth.*, 110, 135–140.
128. Edelman, E., Kost, J., Bobeck, H., and Langer, R. (1985), Regulation of drug release from polymer matrices by oscillating magnetic fields, *J. Biomed. Mater. Res.*, 19, 67.
129. Babincová, M., Sourivong, P., Chorvát, D., and Babinec, P. (1999), Laser triggered drug release from magnetoliposomes, *J. Magnetism Magnetic Mater.*, 194, 163–166.
130. Kim, S. W., Pai, C. M., Makino, K., Seminoff, L. A., Holmberg, D. L., Gleeson, J. M., Wilson, D. E., Mack, E. J. (1990), Self-regulated glycosylated insulin delivery, *J. Controlled Release*, 11, 193–201.
131. Obaidat, A. A., and Park K. (1997), Characterization of protein release through glucose-sensitive hydrogel membranes, *Biomaterials*, 18, 801–806.
132. Kataoka, K., Miyazaki, H., Bunya, M., Okano, T., and Sakurai, Y. (1998), Totally synthetic polymer gels responding to external glucose concentration: Their preparation and application to on-off regulation of insulin release, *J. Am. Chem. Soc.*, 120, 12694–12695.
133. Yui, N., Okano, T., and Sakurai, Y. (1992), Inflammation responsive degradation of crosslinked hyaluronic acid gels, *J. Controlled Release*, 22, 105–116.
134. Yui, N., Nihira, J., Okano, T., and Sakurai, Y. (1993), Regulated release of drug microspheres from inflammation responsive degradable matrices of crosslinked hyaluronic acid, *J. Controlled Release*, 25, 133–143.
135. Kikuchi, A., and Okano, T. (2002), Pulsatile drug release control using hydrogels, *Adv. Drug Deliv. Rev.*, 54, 53–77.

# PROGRESS IN DESIGN OF BIODEGRADABLE POLYMER-BASED MICROSPHERES FOR PARENTERAL CONTROLLED DELIVERY OF THERAPEUTIC PEPTIDE/PROTEIN

SHUNMUGAPERUMAL TAMILVANAN*
*University of Antwerp, Antwerp, Belgium*

**Contents**

1 Introduction
2 Peptide/Protein-Loaded Microsphere Production Methods
  2.1 Phase Separation (A Traditional Technique)
  2.2 Double Emulsion (A Hydrous Technique)
  2.3 Spray Drying (An Anhydrous Technique)
  2.4 New Trends in Production Methods
3 Analytical Characterization of Peptide/Protein-loaded Microspheres
4 Immune System Interaction with Injectable Microspheres
5 Excipient Inclusion: Injectable Peptide/Protein-Loaded Microspheres
  5.1 Solubility- and Stability-Increasing Excipients
  5.2 Preservation-Imparting Excipients
6 Peptide/Protein Encapsulated into Biodegradable Microspheres: Case Study
  6.1 Vaccines
  6.2 Proteins
7 Conclusion
  References

*Current address: Department of Pharmaceutics, Arulmigu Kalasalingam College of Pharmacy, Anand Nagar, Krishnankoil, India

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

## 1 INTRODUCTION

At the cellular level, deoxyribonucleic and ribonucleic acids (DNA and RNA, respectively) serve as an endogenious vehicle not only to store genetic information but also to transfer genetic information from one generation to their offsprings of all known living organisms. In addition, utilizing the rule of complementary base pairing, the DNA undergoes replication and transcription processes to produce respectively a new double-stranded DNA molecule and a complementary single-stranded RNA molecule. Following the translation process, peptide and protein are synthesized/constructed in ribosomal subunits through peptidic linkages between available 20 amino acids. The peptide and protein thus constructed perform a wide variety of functions and each cell contains several thousands of different proteins. Peptide- and protein-mediated, important physiological and biological processes of the human body include ligands/hormones for signaling, enzymes for biotransformation reactions, receptors for pharmacological response elucidation, antibodies in immune system interactions, transcription, and translation. Hence these molecules play a vital role to ensure proper development and functioning of entire organs of the human body.

*Webster's New World Dictionary* defines a drug as "any substances used as a medicine or as an ingredient in a medicine." Indeed, peptides and recombinant proteins are highly potent, relatively macromolecular and promising therapeutic agents that emerged out from the significant development of biotechnic and biogenetic engineering technologies. Peptide and protein therapeutics include semisynthetic vaccines, monoclonal antibodies, growth factors, cytokines, soluble receptors, hormones, and enzymes. The advent of recombinant DNA technology allowed the possibility of the commercial production of proteins for pharmaceutical applications from the early 1980s and, in fact, manufacture of therapeutic proteins represented the first true industrial application of this technology [1]. During the 1980s the term *biopharmaceutical* became synonymous with *therapeutic protein produced by recombinant DNA technology* (or, in the case of a small number of therapeutic monoclonal antibodies, *by hybridoma technology*). Clinical evaluation of nucleic acid–based drugs used for the purposes of gene therapy and antisense technology commenced in the 1990s, and today the term biopharmaceutical also encompasses such (as-yet-experimental) drugs [2]. The first such recombinant therapeutic protein (insulin) was approved for general medical use only 24 years ago. Today there are in excess of 100 such products approved in some world regions at least, with 88 having received approval within the European Union (EU). This represents 36% of all new drug approvals since the introduction of the new centralized European drug approval system in 1995 [3]. Over the coming decade, therefore, in the region of a dozen new therapeutic proteins should, on average, gain regulatory approval each year. While EU figures are difficult to locate, the American Association of Pharmaceutical Researchers and Manufacturers (PhRMA) estimates that there are currently some 371 biotechnology medicines in development [4]. Out of these 371 biotechnology medicines, as estimated by PhRMA, more than 300 are protein based, with recombinant vaccines and monoclonal/engineered antibodies representing the two most promising categories. Incidentally, all 88 biopharmaceutical products currently approved within the EU are protein based. Of the proteins thus far approved, hormones and cytokines represent the largest product categories (23 and 18 products, respectively). Hormones approved include several recombinant insulins, displaying both native and modified amino acid

sequences. In addition, several recombinant gonadotrophins [follicle-stimulating hormone (FSH), luteinizing hormone (LH), and human chorionic gonadotrophins (hCG)] have been approved for the treatment of various forms of subfertility/infertility. Cytokines approved include a range of recombinant hematopoietic factors, including multiple erythropoietin-based products used for the treatment of anemia as well as a colony-stimulating factor aimed at treating neutropenia. Additional approved cytokines include a range of interferon-based products, mainly used to treat cancer and various viral infections, most notably hepatitis B and C, and a recombinant tumor necrosis factor (TNF) used as an adjuvant therapy in the treatment of some soft tissue cancers. Blood-related approved therapeutic proteins include a range of recombinant blood coagulation factors used to treat hemophilia, recombinant thrombolytics, and recombinant anticoagulants. Additional product categories include a range of subunit vaccines containing at least one recombinant component [mainly hepatitis B surface antigens (HBsAg)] and a variety of monoclonal antibody–based products indicated for the treatment/detection of various cancers or the prevention of organ transplant rejection. In summary, ailments that can be treated more effectively by this new class of therapeutic agents include cancers, autoimmune disease, memory impairment, mental disorders, hypertension, and certain cardiovascular and metabolic diseases [5, 6].

Poor absorption and easy degradation by endogenous proteolytic enzymes present in eye tissues, nasal mucosa, and gastro intestinal tract and low transdermal bioavailabilities due to relatively large size make the peptide/protein molecules to be administered only through parenteral routes either by multiple injections or infusion therapy in order to achieve desired therapeutic plasma levels for prolonged periods of time. Nevertheless, because of remarkably short half-lives within the in vivo arena, the therapeutic usuage of most of the peptide/protein is practically possible only through daily multiple injections under close medical supervision. Hence, the commercial success of peptides/proteins as therapeutic agents depends mainly on development of novel drug delivery systems which could potentially reduce the injection frequencies and thus eliminate the accompanying serious problem of patient compliance.

Among the several technologies that have been suggested for reducing injection frequencies of therapeutic peptide/protein, microspheres prepared from biodegradable polymers are widely recognized for controlled drug delivery following parenteral administration. Polyester polymers such as poly(lactic acid) (PLA), poly(glycolic acid) (PGA), and their copolymer poly(lactic acid-*co*-glycolic acid) (PLGA) are used routinely for the preparation of injectable microspheres after taking into consideration their well-known biocompatibility, controlled biodegradability, absorbability, and no toxicity of degradation products [7]. Furthermore, the PLGA types and related poly(hydroxyalkanonates) have a long history of medical and pharmaceutical use in fields as diverse as sutures, bone fixatives, artificial skins and cartilages, dental materials, materials for bone generation, drug delivery, and many others, as reviewed by Ueda and Tabata [8]. In conjunction with a long safety record of PLGA polymers, at least 12 different peptide/protein-loaded PLGA microsphere products are available in the market from nine different companies worldwide for the treatment of some life-threatening diseases (Table 1). In recent years, poly(ε-caprolactone) (PCL) has been investigated as an alternative to PLGA to make microspheres [9, 10]. A glimpse of ongoing research activities utilizing biodegradable polymer-based microspheres for various peptide/protein is shown in

**TABLE 1 Currently Marketed Preparations (Injectable Microspheres) Containing Peptide/Protein Molecules**

| Commercial Name | API | Polymer | Company | Indication |
|---|---|---|---|---|
| Lupron Depot | Leuprolide | PLGA or PLA | TAP | Prostate cancer, endometriosis |
| Enantone Depot | Leuprolide | PLGA or PLA | Takeda | Prostate cancer, endometriosis |
| Trenantone | Leuprolide | PLGA or PLA | Takeda | Prostate cancer, endometriosis |
| Enantone Gyn | Leuprolide | PLGA or PLA | Takeda | Prostate cancer, endometriosis |
| Sandostatin LAR | Octreotide | PLGA-glucose | Novartis | Acromegaly |
| Nutropin | Somatropin | PLGA | Genentech | Growth deficiencies |
| Trelstar Depot | Triptorelin | PLGA | Pfizer | Prostate cancer |
| Decapeptyl SR | Triptorelin | PLGA or PLA | Ipsen-Beaufour | Prostate cancer |
| Decapeptyl | Triptorelin | PLGA | Ferring | Prostate cancer |
| Suprecur MP | Buserelin | PLGA | Aventis | Endometriosis |
| Somatuline LA | Lanreotide | PLGA | Ipsen-Beaufour | Acromegaly |
| Parlodel LAR | Bromocriptine | PLGA-Glu | Novartis | Parkinsonism |

*Abbreviations:* PLA: polylactide; PLGA: poly(lactide-*co*-glycolide); API: active pharmaceutical ingredient; PLGA-Glu: poly(D,L-lactide-*co*-glycolide-D-glucose).

Table 2 (incorporating refs. 12–35). However, overcoming the propensity for peptides/proteins to undergo degradation processes during incorporation into the biodegradable microspheres or after injection into the body awaiting release is one of the key hurdles in bringing microencapsulated systems for these drugs to market. This partially explains the limited and only a countable number of formulations available on the market. Furthermore, irrespective of the various microencapsulation techniques adopted to prepare peptide/protein-loaded microspheres, several transfer-required processes such as filtration, centrifugation, and vacuum or freeze drying are necessary to obtain a final product, and these processes might be obstacles when scaling up the manufacturing technique to produce sufficient quantities of sterile material for clinical trial and, ultimately, commercialization [11].

This chapter encompasses investigations made progressively on the design of injectable peptide/protein-loaded PLGA microspheres. It covers an update on the state of art of the manufacturing of peptide/protein-loaded microspheres through both conventional and newer microencapsulation techniques, different analytical methods used for microsphere characterization, immune system interaction with microspheres following parenteral administration, and potential application of microspheres having therapeutic peptides/proteins. Special emphasis is given particularly on various instability problems and investigated mechanistic ways to obviate the possible instability problems of peptide/protein drug during microsphere preparation as well as its release from the microspheres. It should be added that although the chapter focuses mainly on PLGA microspheres, many of the destabilization mechanisms and stabilization approaches described herein can be valid to some extent for other polymeric delivery systems, too.

**TABLE 2 Injectable Peptide/Proteins/Vaccines Encapsulated in Biodegradable Microspheres**

| Peptides, Protein, Vaccine | Technique | Polymer | Reference |
|---|---|---|---|
| | *Vaccine* | | |
| SPf 66 malaria vaccine | Double emulsion | PLGA | 28 |
| Multivalent vaccines of *Haemophilus influenzae* type b (Hib), diphtheria toxoid (DT), tetanus toxoid (TT), pertussis toxin (PT) | Spray drying | PLGA | 29 |
| Rotavirus | Double emulsion | PLG | 30 |
| | *Polypeptides and Proteins* | | |
| Insulin | Double emulsion | PLA polyethylene glycol (PEG) | 12 |
| Recombinant human epidermal growth factor (rhEGF) | Double emulsion | PLA | 13 |
| Ribozyme | Double emulsion | PLA, PLGA | 14 |
| Vapreotide (somatostatin analogue) | Spray drying | PLGA | 15 |
| Insulinlike growth factor-1 (IGF-1) | Double emulsion | PLGA-PEG | 16 |
| Ornitide acetate leuteinizing hormone releasing hormone [(LHRH) antagonist] | Dispersion/solvent extraction/evaporation | PLA, PLGA | 17 |
| Vascular endothelial growth factor (VEGF) | Single emulsion | PLGA/PEG | 18 |
| Human chorionic gonadotropin (hCG) | Double emulsion | PLA, PLGA | 19 |
| Calcitonin | Double emulsion | PLGA | 20 |
| FITC-bovine serum albumin (BSA) | Double emulsion | Poly(ε-caprolactone) | 9 |
| Levonorgestrel and ethinylestradiol | Double emulsion | Poly(ε-caprolactone) | 10 |
| Recombinant human bone morphogenetic protein | Double emulsion | PLGA | 21–24 |
| Transforming growth factor beta | Double emulsion | PLGA or PLGA-PEG | 25–27 |
| Recombinant human erythropoietin (rhEPO) | Modified double emulsion | LPLG-PEO-LPLG | 31 |
| Protein-C | Double emulsion | PLA | 32 |
| Ovalbumin | Double emulsion | PLGA | 33 |
| Human serum albumin | Double emulsion | PLA | 34 |
| Bovine serum albumin | Nonaqueous oil-in-oil (o/o) emulsion | PLG | 35 |

*Abbreviations:* FITC: fluroscein isothiocyanate; PLG: poly(lactide-co-glycolide); LPLG-PEO-LPLG: copoly(l-lactic-co-glycolic acid-b-oxyethylene-b-l-lactic-co-glycolic adic); PEO: polyethylenenoxide.

## 2 PEPTIDE/PROTEIN-LOADED MICROSPHERE PRODUCTION METHODS

The development of delivery systems for therapeutic peptides/proteins depends on biophysical, biochemical, and physiological characteristics of these molecules, including molecular size, biological half-life, immunogenicity, conformational stability, dose requirement, site and rate of administration, pharmacokinetics, and pharmacodynamics [36]. Unlike conventional drug molecules, the unique conformational structure of peptidic/proteinic therapeutic agents poses a great challenge right from the beginning of the selection of suitable microencapsulation techniques to make microspheres. Table 3 lists the considerations to be taken before choosing a particular encapsulation technique. Apart from the traditional phase separation technique, other techniques suitable for peptide/protein-loaded microsphere production can be divided into two main categories: during microsphere preparation, those involved in utilizing a hydrous environment such as emulsion-based methods and those based on an anhydrous environment such as spray freeze drying, spray drying, freeze drying, grinding, jet milling, liquid-phase antisolvent precipitation, and supercritical $CO_2$-based methods [37–40]. In the following sections, the various production techniques to make injectable peptide/protein-loaded microspheres are briefly introduced; however, a detailed discussion is beyond the scope of this chapter.

### 2.1 Phase Separation (A Traditional Technique)

Polymer phase separation or coacervation is an excellent technique for the encapsulation of water-soluble drugs including peptide/protein into a final microsphere product [41]. The peptide/protein molecule is dispersed in solid form into solution containing dichloromethane and PLGA. Silicone oil is added to this dispersion at a defined rate, reducing the solubility of polymer in its solvent. The polymer-rich liquid phase (coacervate) encapsulates the dispersed peptide/protein molecules and embryonic microspheres are subjected to hardening and washing using heptane. The process is quite sensitive to polymer properties, and residual solvent is also an important issue. Decapeptyl [triptorelin, a luteinizing hormone releasing hormone (LHRH) analogue] [42] and Somatuline LA (lanreotide, a somatostatin analogue) [43] are microsphere commercial products developed by this technique (Table 1).

**TABLE 3 Factors in Selection of Microencapsulation Method to Prepare Peptide/Protein-Loaded Microspheres**

| |
|---|
| Optimal peptide loading |
| High microsphere yield |
| Batch content uniformity |
| Interbatch reproducibility |
| Peptide stability during preparation and release |
| Size uniformity |
| Adjustable release profile |
| Low burst release |
| Flowability of final product |
| Residual solvent and polymer monomer control |
| Sterilization (both aseptic and terminal) |

### 2.2 Double Emulsion (A Hydrous Technique)

Oil-in-water (o/w) and water-in-oil-water (w/o/w) are the two hydrous techniques representing respectively the single- and double-emulsion formation during microsphere preparation. However, the w/o/w technique is most commonly employed [44]. In this process, peptides/proteins in aqueous solution are emulsified with nonmiscible organic solution of polymer to form a w/o emulsion. Dichloromethane serves as organic solvent and the o/w primary emulsion is formed using either high-speed homogenization or ultrasonication. This primary emulsion is then rapidly transferred to an excess of aqueous medium containing a stabilizer, usually polyvinyl alcohol. Again homogenization or intensive stirring is necessary to initially form a double emulsion of w/o/w. Subsequent removal (by evaporation) of organic solvent by heat, vacuum, or both results in phase separation of polymer and core to produce microspheres. Instead of solvent evaporation, solvent extraction with a large quantity of water with or without a stabilizer can also be undertaken to yield microspheres containing peptide/protein. Although the w/o/w microencapsulation technique seems to be conceptually simple to carry out, the particle formation process is quite complicated, and a host of process parameters influence the properties of peptide/protein-loaded PLGA microspheres [45]. In spite of that, different peptides and proteins such as bovine serum albumin (BSA) or ovalbumin (OVA), insulin, recombinant human insulinlike growth factor-1 (rhIGF-1), recombinant human epidermal growth factor (rhEGF), human chorionic gonadotropin (hCG), protein C, recombinant human bone morphogenetic protein (rhBMP), and calcitonin, along with antigens and other therapeutically relevant proteins such as recombinant human erythropoietin (rhEPO), have been successfully encapsulated (see Table 2) by the w/o/w double-emulsion technique. Lupron Depot/Enantone Depot/Trenantone/Enantone Gyn (all having leuprolide acetate, a LHRH analogue) are very popular commercial microsphere products produced by this technique [46, 47], available both in the EU and United States, for the treatment of either prostate cancer of man or infertility (endometriosis) of women (Table 1).

### 2.3 Spray Drying (An Anhydrous Technique)

Spray drying is a rapid, convenient technique which can be conducted under aseptic conditions. First, a polymer—prevalently PLGA is applied—is dissolved in a volatile organic solvent such as dichloromethane or acetone. The protein is suspended as a solid or emulsified as aqueous solution in this organic solution by homogenization. After that, the resulting dispersion is atomized through a (heated) nozzle into a heated airflow. The organic solvent evaporates, thereby forming microspheres with dimensions of typically 1–100 μm. The microspheres are collected in a cyclone separator. For the complete removal of the organic solvent, a vacuum drying or lyophilization step can follow downstream.

The internal structure of the resulting polymeric microspheres depends on the solubility of the peptide/protein in the polymer before being spray dried leading to the formation of reservoir- or matrix-type products (see Figure 1). When the initial dispersion is solution, the final product obtained following spray drying is matrix or monolithic type, that is, polymer particles with a dissolved or dispersed nature of the active ingredient (defined as microspheres). Conversely, when the initial

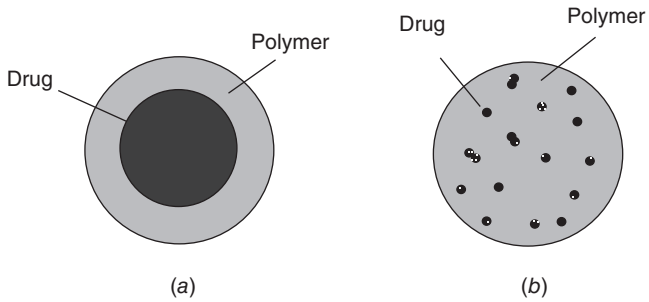


**FIGURE 1** Polymeric delivery systems: (a) reservoir systems (microcapsules); (b) matrix systems (microspheres).

dispersion is in suspension, the product obtained is reservoir type, that is, a distinct polymeric envelope/shell encirculating a liquid core of dissolved active ingredient (defined as microcapsules). Recombinant human erythropoietin [48] and bromocriptine mesylate, Parlodel Depot [49], are examples of microspheres (matrix type) obtained by the spray drying technique.

## 2.4 New Trends in Production Methods

Several issues such as reducing cost, reducing scale-up difficulties, improving protein stability, allowing for terminal sterilization, and eliminating the need for organic solvents during addition of the peptide/protein motivate the development of new methods to manufacture microspheres. Moreover conventional microencapsulation methods involve relatively harsh conditions that are not generally tolerated by peptide/protein molecules without stabilization. Therefore, new and improved processes shielding the peptide/protein from deleterious conditions have been proposed and evaluated.

***Modified Conventional Methods*** The w/o/w solvent evaporation or extraction is probably one of the most widely used methods for peptide and protein microencapsulation [44], despite its many drawbacks. Improvements and alternatives have therefore been proposed such as oil in water (o/w), *o/w (the asterisk including cosolvent) and oil in oil (o/o) [50].

Utilising a modified w/o/w method, the rhIGF-1 was encapsulated into PLGA microspheres after increasing the pH of the protein solution from 4.5 to 5.5–6.0, where rhIGF-1 formed a viscous gel [51]. High entrapment efficiency of fully bioactive protein was achieved, and 92–100% of pure, monomeric, and bioactive rhIGF-1 was released in vitro over 21 days. The lowering of the rhIGF-1 solubility at pH 5.5–6.0 probably restricted its conformational flexibility and changes upon exposure to the polymer solvent. Without pH adjustment, approximately 10–32% of rhIGF-1 was lost upon solvent exposure, due to degradation and aggregation. Elsewhere, a $w/o_1/o_2$ system was investigated for encapsulating different proteins and peptides, with the $o_1$ and $o_2$ phases consisting of acetonitrile/dichloromethane and liquid paraffin/Span 80, respectively [52]. The acetonitrile mediated the partial mixing of the w and $o_1$ phases and subsequent protein/peptide precipitation, which was a

prerequisite for microencapsulation. The proteins BSA, tetanus toxoid (TT), and lysozyme did precipitate at low acetonitrile concentration, resulting in efficient microencapsulation (more than 90%), while a decapeptide and a linear gelatine did not precipitate so rapidly, resulting in poor entrapment. TT and lysozyme released during the burst phase (15%) maintained their bioactivity, although lack of further release suggested aggregation within the microspheres.

Another approach consisted of dispersing the protein antigen in a mineral oil before encapsulation into PLGA microspheres by a $o_1/o_2/w$ method [53]. The mineral oil ($o_1$) was intended as a barrier to protect the antigen during emulsification with the polymer solution and from exposure to moisture during release. Over 92% of enzyme-linked imunosorbert assay (ELISA) reactive TT was released from the reservoir-type microspheres in a pulsatile pattern, proceeding with an initial burst and followed by a second release pulse between 14 and 35 or 35 and 63 days, depending on the polymer type used. The latter stage of release was ascribed to TT diffusion through the oily phase, once an appreciable loss of polymer mass had occurred. The authors claimed the mineral oil was the key to protecting the solid antigen during polymer erosion, where acidic degradants and moisture would otherwise have led to antigen inactivation.

To improve solvent extraction, a novel method using a static micromixer was presented where a $w_1/o$ dispersion (aqueous BSA in organic PLGA solution) is fed into an array of microchannels and the extraction fluid ($w_2$) into a second array of interdigitated channels [54]. The two fluids, transported separately through the channels, are discharged through an outlet slit where alternating fluid lamellae are formed with the $w_1/o$ fluid lamella disintegrating into microdroplets, which harden quickly to form microspheres. This process offers easy scale-up, methodological robustness, continuous production, and a simple setup, making it ideally suited for aseptic production, a strongly needed feature for microsphere vaccine formulations.

***ProLease Technology (Cryogenic Spray Drying)*** A variation of the conventional spray drying method is a cryogenic method which will described below. A novel low-temperature spraying technique (called ProLease technology) for preparing PLA and PLGA microspheres has been reported by Khan et al. [55] and the group at Alkermes [56, 57]. The method relies on the use of stabilizing and release controlling agents, low processing temperature, and nonaqueous microencapsulation. Typically, a protein powder is micronized, possibly with a stabilizer, by spray freeze drying and then suspended in an organic polymer solution. The suspension is atomized into a vessel containing liquid $N_2$ underlaid by frozen ethanol (extraction solvent). The atomized droplets freeze in the liquid $N_2$ and deposit on the surface of the frozen ethanol. As liquid $N_2$ evaporates, the frozen ethanol liquefies ($T_m$ approximately –110°C) so that the frozen polymeric droplets will transfer into the ethanol where the polymer solvent is extracted, yielding solid microspheres [58, 59]. To date, the ProLease system has been effectively applied to the encapsulation of zinc-complexed human growth hormone in PLGA microspheres, resulting in a one-month effect after one single injection [37, 57, 60]. As a reference, the recombinant human growth hormone (rhGH) was unstable in contact with ethyl acetate or dichloromethane [61]. The only protein-containing PLGA microspheres, Nutropin Depot, is produced by this novel technique. However, this product containing rhGH

marketed initially in the United States in 1999 was pulled from the market voluntarily by the manufacturer in June 2004 because of high costs of production and commercialization (http://www.gene.com/gene/news/press-releases/, accessed May 25, 2006).

ProLease technology was also used for encapsulating recombinant human vascular endothelial growth factor (rhVEGF) and rhIGF-1 [62, 63]. Both proteins were stabilized in aqueous solution, prior to spray freeze drying, and encapsulated (9–20% w/w) into PLGA microspheres. The microspheres also contained $ZnCO_3$ (3–6% w/w) as release modifier. The resistance of rhIGF-1 to aggregation and oxidation, determined from in vitro release studies, hardly changed. Protein, released in an almost pulsatile fashion over 21 days, was composed of predominantly monomeric rhIGF-1 with only minor amounts (~6%) of degradants forming toward day 21. Similarly, the integrity of rhVEGF dimer released over 21 days was good and its bioactivity remained largely unaffected, regardless of the extent of aggregation and degradation. In view of these studies, ProLease technology appears to have potential for sustaining antigen stability and release from microspheres.

***Techniques Using Supercritical Fluids*** Generally, the application of supercritical (SC) fluids for the encapsulation of peptides and proteins has been fueled by the recognition that the established methods implicate some drawbacks. The application of supercritical fluids, especially of supercritical carbon dioxide ($CO_2$), can minimize or even eliminate the use of organic solvents and renders work at moderate temperatures possible [64]. The term *supercritical* defines the area above the critical point, which specifies the final point of the liquid–gas phase transition curve. Beyond that critical point, isobar/isotherm alterations of pressure or temperature alter the density of the critical phase but do not lead to a separation into two phases. A density change is directly associated with a change of the solvent power, and thus the method features a high variability. Usually $CO_2$ is used as supercritical fluid due to its critical point ($T_c$ = 31.1°C, $P_c$ = 73.8 bars), which can be easily reached. That allows a moderate working temperature and leaves no toxic residues since it returns to the gas phase at ambient conditions. Two SC $CO_2$-based processes have been reported for the preparation of drug-loaded polymeric microspheres: first, the rapid expansion from supercritical solutions (RESS) process, whereby a SC $CO_2$ solution of an active agent and a polymeric carrier is rapidly expanded. This quickly transforms the SC $CO_2$ into a liquid that is a much poorer solvent, thereby precipitating the active agent–carrier mixture as small particles [65]. Second is the aerosol solvent extraction system (ASES), also referred as the gas antisolvent spray precipitation (GAS) process [66]. Here, a solution of the active agent and the polymeric carrier is sprayed into a chamber loaded with SC $CO_2$. The SC $CO_2$ extracts the solvent from the spray droplets and induces coprecipitation of the active agent and the polymeric carrier in the form of small, solvent-free particles [67, 68]. However, the use of organic solvents cannot be avoided, which is to be deemed a major disadvantage of both techniques.

In peptide/protein pharmaceuticals, the GAS process is predominantly applied for the preparation of microparticulate protein powders as an alternative to common drying processes. However, Winters et al. [69] reported an increase of β-sheet aggregates during the precipitation of lysozyme, trypsin, and insulin as a consequence of stress parameters such as organic solvent, pressure, and shear forces. One reason

why these methods were not credited as encapsulation techniques for protein within PLGA may be the tendency of several polymers to rapidly precipitate and agglomerate during the process [70].

ASES has been compared with conventional spray drying in terms of effects on the stability of the peptide tetracosactide [71]. Almost no intact peptide was recovered from spray-dried PLA particles, whereas the tetracosactide was well protected against oxidation during ASES (~94% unmodified peptide). In general, the particle formation step seems to be less detrimental to proteins than the loading step. For example, emulsification in an aqueous phase or spray drying of rhEPO/PLGA emulsions was mild compared to the first emulsification step [72]. Also, variation of the particle formation step (spray drying or coacervation) had a minor impact on diphtheria toxoid (DTd) antigenicity when compared to other process variables [73].

A serious limitation of GAS, ASES, and RESS for producing microspheres is the need of polymer types that form discrete crystalline domains upon solidification, such as l-PLA [74, 75]. The advantages of these methods offer (e.g., over spray drying) are the low critical temperatures for processing (34°C) and the avoidance of oxygen exposure during atomization, with both parameters being potentially important to peptide/protein stability.

***Ultrasonic Atomization*** Ultrasonic atomization of w/o dispersions is presently under investigation for preparing especially protein antigen containing microspheres. In one setup, the atomized antigen/polymer dispersion was sprayed into a nonsolvent where the polymer solvent was extracted, resulting in microspheres [76]. A comparable technique was proposed where the antigen or polymer dispersion was atomized into a reduced pressure atmosphere and the preformed microspheres hardened in a collection liquid [77]. Similarly, PLGA solutions were also atomized by acoustical excitation and the atomized droplets transported by an annular stream of a nonsolvent phase [aqueous polyvinyl alcohol (PVA)] into a vessel containing aqueous PVA [78]. Solvent evaporation and microsphere hardening occurred in the vessel over several hours. The main advantages of these atomization techniques encompass the possibility of easy particle size control and scale-up, processing at ambient or reduced temperature, and the suitability for aseptic manufacturing in a small containment chamber such as an isolator.

***In Situ Formed Injectable Microspheres*** All the encapsulation techniques discussed so far rely on the preparation of solid microspheres. However, a method for preparing a stable dispersion of protein containing semisolid PLGA microglobules has been reported [79]. Here, a protein dissolved in PEG 400 was added to a solution of PLGA in triacetin or triethyl citrate. This mixture, stabilized by Tween 80, was added dropwise and under stirring to a solution of Miglyol 812 or soybean oil, containing Span 80, resulting in a stable dispersion of protein inside semisolid PLGA microglobules. The microglobules remained in an embryonic state until mixed with an aqueous medium, so that the water-miscible components were extracted and protein containing matrix-type microspheres formed. Myoglobin was encapsulated and found to remain physically unchanged (circular dichroism analysis) after the process and during storage of the microglobular dispersion (15 days, 4°C).

***Preformed Porous Microspheres*** A new approach for attaining sustained release of protein is introduced involving a pore-closing process of preformed porous PLGA microspheres [80]. Highly porous biodegradable PLGA microspheres were fabricated by a single w/o emulsion solvent evaporation technique using Pluronic F127 as an extractable porogen. The rhGH was incorporated into porous microspheres by a simple solution dipping method. For its controlled release, porous microspheres containing rhGH were treated with water-miscible solvents in the aqueous phase for production of pore-closed microspheres. These microspheres showed sustained-release patterns over an extended period; however, the drug loading efficiency was extremely low. To overcome the drug loading problem, the pore-closing process was performed in an ethanol vapor phase using a fluidized-bed reactor. The resultant pore-closed microspheres exhibited high protein loading amount as well as sustained rhGH release profiles. Also, the released rhGH exhibited structural integrity after the treatment.

***Charged (Anionic and Cationic) PLGA Microspheres*** PLGA or any other type of microspheres can be readily decorated with positive or negative surface charges by simply preparing the particles by a w/o/w solvent evaporation/extraction process where the second water phase contains a cationic emulsion stabilizer [hexadecyltrimethylammonium bromide; poly(ethyleneimine); stearylamine] or an anionic emulsifier [sodium dioctyl-sulfosuccintate; sodium dodecyl sulfate (SDS)]. Such compounds attach tightly to PLGA surfaces during preparation and provide the necessary surface charge for ionic adsorption of counterions. It is known that a protein's surface charge depends on its isoelectric point (pI) and the pH of the medium in which it is dispersed. The use of particles with ionic surface charge offers several advantages over classical microencapsulation, among which the mild conditions for loading are probably the most attractive. PLGA microspheres with surface-adsorbed protein antigens and DNA have been highly efficient in inducing strong immune responses, as reviewed by Singh et al. [81] and Jilek et al. [82]. Nonetheless, it remains to be shown whether such particles are also suitable for eliciting long-term immunity after one or two injections.

Jabbal-Gill et al. [83, 84] noted the tendency for microencapsulated protein antigens to distribute heavily at the surface of PLGA microspheres and developed polymeric lamellar substrate particles (PLSP) by precipitating a highly crystalline poly(*l*-lactic acid)/organic solvent solution with water, followed by removal of remaining organic solvent with nitrogen purge. The particles, which can be sterilized by gamma irradiation and stored as a suspension for several months without changes to antigen absorption [84], possessed a large lamellar surface area and highly negative zeta potential (~–35 to –42 mV) and could adsorb significant amounts of antigen (up to 50 μg/mg microspheres) depending on pH, ionic strength, antigen–polymer ratio, and other factors. Release of protein antigen (TT) could be extended to over 1 month with minimal antigenic losses in released antigen, although most of the antigen was lost to the initial burst or to apparent irreversible adsorption (as indicated by the absence of reaching 100% release). Elevated antibody responses in mice were elicited using PLSP similar to one dose of aluminum adjuvant following subcutaneous administration of OVA at elevated doses (100 or 300 μg). Both immunoglobulin IgG1 and IgG2a antibody subtypes were of similar magnitude over 28 days in the PLSP/OVA groups, and cellular immunity was also observed following

immunization with a 38-kDa protein antigen against tuberculosis [85]. Similarly, Kazzaz et al. [86] created anionic PLGA microparticles by substituting the standard nonionic emulsifier PVA with anionic SDS during microsphere preparation. In addition to eliciting elevated antibody responses in mice relative to the soluble antigen, the adsorbed antigen elicited a potent cytotoxic T-cell (CTL) response, similar to that observed after infection from virus expressing the p55 gag and polymerase proteins. Moreover, the CTLs were formed from the more challenging intramuscular route but not significantly by the soluble antigen, even at elevated doses. The SDS-PLGA particles could also be gamma irradiated before adsorption and were shown to effectively boost antigen in nonhuman primates [87].

## 3 ANALYTICAL CHARACTERIZATION OF PEPTIDE/ PROTEIN-LOADED MICROSPHERES

An area requiring additional efforts is analytical characterization of peptides and proteins encapsulated in PLGA microspheres. The high complexity of the therapeutic peptides and proteins requires not only physicochemical methodologies but also immunochemical and biological techniques for the characterization and quality control of these substances. In general, the analytical methods can be broadly viewed from the following study perspectives: methods meant for microsphere product quality checking, methods used for peptide/protein stability identification inside the microspheres, and methods called for peptide integrity detection following liberation from the microspheres immediately upon placement in release medium either in vitro or in vivo. Therefore, in most cases, a combination of several analytical methods is necessary for a comprehensive characterization of the peptide/protein under investigation and for appropriate quality control of the product concerning identity, purity, and potency. However, some of the analytical methods have potentially appealing applications to interplay among the mentioned perspectives. In Table 4, a selection of widely used analytical methods is given, showing which technology is applicable for the testing of identity, purity, and potency of peptides and proteins. In addition, peptide/protein integrity evaluation is indeed likely to be affected by artefacts during the sample preparation before analysis and during the analysis itself. Therefore, artefacts might prevent the scientist from critically ascribing detected protein denaturation to manufacturing conditions [88].

In order to measure the extent of peptide/protein degradation within the carriers and during release, the encapsulated molecule has to be removed from the polymeric matrix. Moreover, for avoiding artefacts such as underestimation of drug content, recovery methods need to be tried by an empirical trial-and-error approach as each peptide/protein is different one from the other. Recovery methods so far reported include extraction-based method with the help of potentially deleterious organic solvents, hydrolysis of the polymer matrix with alkaline medium, dissolution of polymer matrix in an organic solvent, recovery of suspended insoluble protein by filtration [89], total protein quantification after complete digestion of carriers followed by amino acid analysis [90, 91], electrophoretic extraction of the protein using sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS–PAGE) [92, 93], and direct dissolution of both the polymer and the protein drug in a single liquid

**TABLE 4 Analytical Methods for Characterization and Quality Control of Pharmaceutical Peptides and Proteins**

| Methods | Indicated Usage/Checking | | |
|---|---|---|---|
| | Identity | Purity | Potency |
| *Physicochemical* | | | |
| Chromatography | | | |
| Reversed-phase high-performance liquid chromatography (HPLC, RP-1) | + | + | − |
| Ion exchange | − | + | − |
| Affinity | − | + | − |
| Size exclusion chromatography (SEC) | − | + | − |
| Spectroscopy | | | |
| Infrared spectroscopy | + | + | − |
| Raman spectroscopy | + | + | − |
| Fluorescence spectroscopy | − | + | − |
| Ultraviolet/visible (UV/VIS) spectroscopy | + | + | − |
| NMR spectroscopy | + | + | − |
| Mass spectrometry | + | + | − |
| Circular dichroism (CD) | + | + | − |
| Matrix-assisted light desorption ionization–time-of-flight (MALDI-TOF) mass spectrometry | + | + | − |
| Electrophoresis | | | |
| Capillary electrophoresis | − | + | − |
| SDS–polyacrylamide gel electrophoresis (PAGE) | + | + | − |
| Isoelectric focusing | − | + | − |
| *Immunochemical* | | | |
| Radioimmunoassay (RIA) | − | − | + |
| ELISA | − | − | + |
| Western blot | + | − | − |
| *Biological* | | | |
| In vivo assays | − | − | + |
| In vitro (cell culture) assays | − | − | + |

*Abbreviation:* SDS: sodium do decyl sulfate.

phase containing water-miscible organic solvents such as acetonitrile or dimethylsulfoxide (DMSO) [94, 95].

Following successful recovery of peptide/protein molecule from the microspheres, a simple spectrophotometric method does not always allow discrimination between the monomeric protein form and its aggregates. However, HPLC might separate these species and thus provides more accurate qualitative data [96]. But HPLC cannot quantify exclusively the amount of active protein antigen, as is the case with ELISA techniques [97]. Nowadays, Fourier transform infrared (FTIR) spectroscopy has become a popular, noninvasive method, as it is able to characterize the secondary structure of entrapped proteins [26, 95, 98–101]. Only recently, the integrity of their primary structure was evaluated, thanks to a new matrix-assisted laser

desorption ionization–time-of-flight (MALDI-TOF) mass spectrometry method [94, 102]. The method was shown to require little sample material and only a simple dissolution of the carrier was needed prior to the analysis. The MALDI-TOF allowed elucidation of a new degradation pathway, that is, peptide acylation within PLGA carriers resulting from a chemical interaction between peptide and degraded polymer [102]. Moreover, the method was also useful for quantification, and it should be underlined that no interference from PLGA was detected during the measurements. For all the reasons cited above, mass spectrometry should be considered one of the most promising methods for protein analysis inside polymeric carriers including microspheres. Using erythropoietin as an example, an exploratory and elaborative discussion was made on the analytical techniques used for the characterization and quality control of pharmaceutical peptides and proteins [103]. A similar discussion was also done on the analytical techniques critical to (as a part of) the quality assurance after process changes of the production of a therapeutic antibody [104].

With an increasing level of sophistication in the design of new protein antigens and adjuvants (including polymer controlled-release systems), efforts both in the United States and the EU are underway to respond with more appropriate regulations [105–107]. For example, the Committee for Proprietary Medicinal Products (CPMP), the primary scientific body in EU regulatory matters, is currently updating its "notes for guidance," which guide/direct industry and regulatory authorities on content and evaluation of marketing authorization applications for vaccines [105]. Early drafts of these updates include more rigorous guidelines for new non-aluminum-based adjuvants, including antigen stability requirements (see Sesardic and Dobbelaer [105] for a discussion). Similar discussions ongoing in the United States have attempted to standardize requirements of controlled-release parenterals [106, 107], including specifics regarding in vitro release assays and the need to account for 80% or more of the encapsulating agent during the release period.

## 4 IMMUNE SYSTEM INTERACTION WITH INJECTABLE MICROSPHERES

Since microspheres are capable of forming a drug depot, the encapsulated peptide or protein is being slowly released over days or months at the injection site. Interestingly, the size of microspheres plays an important role in immune response. Microspheres with sizes smaller than 10 μm can be directly taken up macrophages (and dendritic cells) through a phagocytosis mechanism while sizes greater than 10 μm need to undergo biodegradation before phagocytosis can occur [108]. It was shown that within a few days of intramuscular injection PLGA microspheres less than 10 μm are completely engulfed in a thin layer of connective tissue and thus evidenced infiltration by macrophages as a consequence of wound-healing response to injected particles [109]. It is feasible that the influx of these macrophages may cause degradation of the encapsulated protein and available protein released in the vicinity of the microspheres. Furthermore, it has been suggested that these macrophages are capable of producing proteolytic enzymes [110], which may result in the release and circulation of altered, inactive, or immunogenic forms of the encapsulated peptide or protein.

On the other hand, degradation, protein antigen release, location, and antigen presentation of microspheres larger than 10 μm are expected to be different from smaller ones. Larger microspheres can provide an extracellular depot for secondary immune responses by way of B-cell stimulation [111–113]. In both cases, upon administration of microspheres, a foreign-body response occurs resulting in an acute initial inflammation despite the excellent tissue compatibility and biodegradability properties of polymers such as PLGA. This initial inflammation is followed by the infiltration of small foreign-body giant cells and neutrophils [114]. These immune cells could consume the released peptide or protein and produce an immune response. However, if released protein is recognized as a self-protein (e.g., homologus), the probability of an immune response by these cells is reduced. It is therefore always essential to release the protein in its native conformation. The release of aggregated or denatured protein from the microspheres may, in fact, result in an unwanted immune response [115]. It should be added that systematic studies to explore the effects of tissue response on the bioavailability of incorporated peptide or protein drug have not appeared extensively in the literature, with a few exceptions as described below. Using a light microscopic technique, bumps containing residual amounts of microspheres were observed at the injection site two weeks after administration of TT-encapsulated PLGA microspheres to mice and guinea pigs [116]. These bump formations may be due to chronic reactions, long-term immunogenicity, and immunological priming of mice and guinea pigs against the injected polymeric microspheres. The immunogenicity of microsphere-encapsulated vaccines can be varied to some extent by changing the physicochemical properties of the microspheres, for example, size, surface properties, and release kinetics of the antigen from the microspheres [111]. An interesting review by Jiang et al. [117] details the various reports on the relationship between in vitro protein antigen stability and immunogenicity, modulation of cell-mediated immune responses, and different formulation approaches to achieve the appropriate immune response with microencapsulated vaccine antigens. There has been some debate, arising from some animal experiments, that the antigenicity does not directly correlate with immunogenicity. However, the stability of protein antigens is considered to play a significant role in the quality and magnitude of immune response for the controlled-release single-dose vaccines as degraded or nonantigenic proteins may not be able to provide a continuous boost for generation of protective levels of high-affinity antibodies.

## 5 EXCIPIENT INCLUSION: INJECTABLE PEPTIDE/ PROTEIN-LOADED MICROSPHERES

Peptide and protein molecules are highly prone to degradation mechanisms that can be divided into two classes: physical and chemical [118]. Whereas chemical degradation leads to the loss of the protein's primary structure through oxidation, deamidation, peptide bond hydrolysis, isomerization, disulfide exchange, and covalent aggregation, physical degradation refers to the changes in the higher order structure (secondary and above) mainly by noncovalent aggregation and precipitation. In particular, aggregates formation during the encapsulation process must be avoided because these aggregates always represent loss of therapeutic efficacy and

increased immunogenicity which can endanger the patient's health [119, 120]. The following few examples indicate the fragility of peptide and protein molecules due to physical or chemical degradation: Aggregation of insulin has been well characterized and depends on unfolding of the insulin molecules [121]; aggregation of lyophilized formulations of BSA, β-galactoglobulin, and glucose oxidase are attributed to disulfide interchange [118]; deamidation contributes to reduction in catalytic activity of lysozyme [122] and ribonuclease at high temperatures [123]; and peptide bond hydrolysis results in loss of activity of lysozyme when heated to 90–100°C [122]. A recent introduction to this list is formaldehyde-mediated aggregation pathway (FMAP) unique to formaldehyde-treated protein antigens such as TT [117, 124, 125].

The formulator of injectable microspheres for peptide and protein faces multiple challenges: (i) to maximize physical and chemical stability, (ii) to prolong biological half-life, (iii) to increase absorption, (iv) to decrease antigenicity, and (v) to minimize metabolism. Thus, it is quite obvious that the fabrication of peptide- and protein-loaded microspheres requires several kinds of excipients for effective stabilization or immobilization of encapsulated therapeutic molecules. Excipients of choice are included specifically for controlling protein degradation in microspheres due to (a) external and internal environmental changes, (b) manipulating the initial burst release, (c) preventing protein adsorption onto delivery devices, and (d) neutralizing the causative acidic microclimate formation due to the acids liberated by the biodegradable lactic/glycolic-based polymers. Therefore, it is generally best to find conditions to stabilize the protein before other aspects of the formulation, such as controlled-release characteristics, are optimized. Typically, the appropriate excipients for the protein under investigation are experimentally selected among various substances by screening. This tedious experimental screening is partly necessary due to the present inability to predict protein stability after addition of such excipients. Moreover, since individual entrapped peptides and proteins differ in terms of physicochemical properties and chemical/therapeutic function, each species is expected to demonstrate a different degree of sensitivity to stress and react differently to the same stabilization strategy. For example, a sugar, amino acid, or antacid excipient may be required to stabilize protein, each of which can increase water uptake in the polymer matrix leading to an increase in release rate. In the scenario in which controlled-release conditions are optimized before such a stabilizer has been identified, it is likely that upon addition of the new stabilizer the release kinetics may change enough to require reformulation. Certainly, there is a sharp contrast between encapsulating a highly water soluble protein [126] or a poorly soluble zinc–protein complex [37, 57]. Switching between these two cases would be expected to alter the requirements in the formulation necessary to attain the controlled-release function (e.g., low versus high polymer matrix permeability for the protein, respectively) because protein solubility in water may be important for any diffusion component of release. The principal stresses causing instability of encapsulated peptide/proteins in PLGA microspheres are elaborated in a book chapter [127] and in a jounal publication [128]. This subject was again reexamined in a review based on new findings since the previous book chapter by the same author [129]. An interesting review from the same research group was published on the biodegradable PLGA microparticles for injectable delivery of vaccine antigens [117], where they focused on mechanistic approaches to improve the stability of PLGA-encapsulated protein antigens.

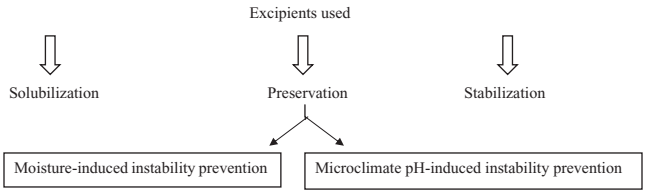


**FIGURE 2** Flow chart of excipients used to prevent/minimize protein instability problems.

Another review by Bilati et al. [130] also envisioned the strategic approaches for overcoming peptide and protein instability within biodegradable nano- and microparticles. The reader is also referred to related publications edited by Sanders and Hendren [131] and Senior and Radomsky [132] for information on excipients used in injectable peptide/protein-loaded formulations including microspheres. This section will not cover all excipients used in parenteral protein formulations because the aforementioned publications already do so. Rather this section highlights examples of synergestic and antagonistic interactions that have been reported mainly between the excipients and the peptide/protein drugs, especially before microsphere preparation, followed by a brief discussion of a major instability problem of proteins/peptides inside the microspheres. The published research paper is being organized according to major functions of parenteral excipients, namely, solubilization, stabilization, and preservation (see Figure 2) [133].

### 5.1 Solubility- and Stability-Increasing Excipients

The traditional approach is to solubilize directly the peptide/protein in organic solvents. This can be achieved by different means. Cleland and Jones [61] assumed that native protein conformation could be maintained by precipitating the protein at its pI. The molecule is then free of charge and can be readily solubilized in organic solvents. Conversely, an alternative concept is based on the freeze drying of the protein at a pH away from its pI value before formulating it. It was thought that this strategy could increase protein solubility and stability in various polar and water-miscible organic solvents such as DMSO [134, 135]. It should be noted that a preformulation procedure consisting of using spray freeze drying with a suitable excipient was able to stabilize BSA before encapsulation by a nonaqueous method [35]. Using the dissolution approach, lysozyme was successfully formulated but incomplete lysozyme release from microspheres was observed and ascribed to aggregation [136]. Protein solubility can also be increased via an ion-pairing mechanism. The protein is modified by adding an oppositely charged surfactant that binds to the protein, so as to obtain a neutral hydrophobic entity and thus reduce direct contact between the protein and the organic solvent. Positively charged proteins and negatively charged surfactants should be employed, since cationic surfactants might have toxic side effects. This technique was shown to improve lysozyme conformational stability after a hydrophobic complex between lysozyme and oleic acid [137, 138]. A new interesting concept is to encapsulate an aggregated protein in a reversibly dissociable form in order to avoid the formation of irreversible aggregates during processing and to promote the sustained release of the native monomeric form. Growth hormone was successfully formulated with this approach [139].

Cyclodextrins (CD) have emerged as very effective additive compounds for solubilizing hydrophobic drugs. In the parenteral dosage form area, modified cyclodextrins such as hydroxylpropyl-β-cyclodextrin and sulfobutylether-β-cyclodextrin have been reported to solubilize and stabilize many injectable drugs, including dexamethasone, estradiol, interleukin-2, and other proteins and peptides [140] without apparent compatibility problems [141]. In addition, CD-containing formulations (either 0.1 *M* sulfobutylether-β-cyclodextrin or 0.1 *M* hydroxylpropyl-β-cyclodextrin) were shown to cause less damage to venous epithelial cells at the site of injection compared with formulations containing organic cosolvents [142]. When CD were coentrapped in the internal aqueous phase, erythropoietin (EPO) covalent aggregate formation was significantly reduced during microsphere preparation by the double-emulsion method [72] and lysozyme stability was improved [88]. Although the precise mechanism is unclear, interactions between amino acids and the hydrophobic inner cavity of CD may play a role [143]. However, CD showed no protecting effect on insulinlike growth factor-1 (IGF-1) [144] and hepatitis B core antigen (HBcAg) [145] and even promoted the loss of superoxide dismutase activity at high CD concentrations [146]. By contrast, carboxymethylcellulose (CMC) did not efficiently stabilize HBcAg and GH against dichloromethane-induced denaturation [61, 145]. Various types (α, β, and γ) of CD were examined for encapsulating TT in PLGA microspheres [147], with γ-hydroxypropyl-cyclodextrin effectively increasing TT encapsulation. However, CD also showed low efficiency in retaining spray-dried TT antigenicity, probably due to antigenic epitopes being buried inside the molecular CD core [147].

Surfactants have the ability to lower surface tension of protein solutions and prevent protein adsorption and/or aggregation at hydrophobic surfaces such as PLGA. Among them, nonionic surfactants are generally preferred as ionic surfactants might bind to groups in proteins and cause denaturation. Tween 20 was shown to greatly reduce the rate of formation of insoluble aggregates of recombinant human factor XIII caused by both freeze thawing and agitation stresses [148]. Maximum protection occurs at concentrations close to the critical micelle concentration of Tween 20, independent of initial protein concentration. In another report, Tween 20 at a 1% (w/v) concentration caused precipitation of a relatively hydrophobic protein (*Humicola lanuginosa* lipase) by inducing nonnative aggregates [149]. Similarly, nonionic surfactants such as Tween 20 or 80 were not good stabilizers for lysozyme and rhGH against the unfolding effect of the water–dichloromethane interface. It has been assumed that both the hydrophilic (PEG chains) and hydrophobic (fatty acid chain) parts of the polysorbate molecules were preferentially partitioned in the dichloromethane phase, leading to low protection efficacy [61, 98]. Exchange of Tween 20 for a less hydrophobic surfactant, PEG 3350, provided almost complete rhGH recovery irrespective of protein concentration. However, an opposing trend was seen with EPO encapsulation in PLGA microspheres [72]. Encapsulated protein aggregates increased (~15%) with different PEG types codissolved in the $w_1$ phase. Conversely, when three nonionic surfactants of different hydrophilic–lipophilic balances (HLBs) were coencapsulated with insulin by the w/o/w double-emulsion method, only Tween 20 was able to improve insulin stability within particles and to limit formation of high-molecular-weight products during the sustained-release period [150].

Tween 80 is well known to protect proteins against surface-induced denaturation [151]. Tween 80 was demonstrated to reduce hemoglobin aggregation in solution by preventing the protein from reaching the air–liquid interface or the liquid–surface interfaces [152]. Polyoxyethylene surfactants such as Tween 80 can form peroxide impurities after long-term storage. Knepp et al. [153] concluded that Tween 80 and other nonionic polyether surfactants undergo oxidation during bulk material storage and subsequent use and the resultant alkyl hydroperoxides formed can contribute to the degradation of proteins. In such formulations, they further reported that thiols such as cysteine, glutathione, and thioglycerol were most effective in stabilizing protein formulations containing peroxide-forming nonionic surfactants.

The Pluronics, also known as poloxamers (e.g., poloxamer 188, British Pharmacopoeia standard) are a well-studied series of commercially available, nonionic, triblock copolymers with a central block composed of the relatively hydrophobic poly(propylene oxide) flanked on both sides by blocks of the relatively hydrophilic poly(ethylene oxide) [154, 155]. The Pluronics possess an impressive safety profile and are approved selectively by the Food and Drug Administration (FDA) for pharmaceutical and medical applications, including parenteral administration [156]. The strong safety profile, commercial availability, ease of preparation, and well-studied physical properties make the Pluronics particularly appealing for drug delivery purposes. They have been used in several patented protein formulations as stabilizers and sustained-release injectables in development as solubilizing and stabilizing agents [157]. However, poloxamers, like Tweens, can form peroxide impurities over time. Poloxamer 188 was successfully used when mixed with PLGA for prolonged release of active interferon-α (INF-α) [158], but such a formulation had no effect on BSA secondary structure compared to PLGA alone [35]. Poloxamer 188 was not effective in preventing nerve growth factor (NGF) aggregation during in vitro release from microspheres generated by spray drying [159]. Complex interactions between poloxamer, BSA, and PLGA were believed to have influenced BSA microencapsulation [160]. The gelling property of the amphiphilic poloxamer 407 was successfully employed for urease encapsulation. The protein was likely protected during the microsphere preparation by a hydrated gelled structure due to the hydrophilic polyoxyethylene chains [161]. EPO aggregates in PLGA microspheres decreased when poloxamer 407 was incorporated at a level of 10% (w/w) [72].

Interleukin-1α (IL-1α) was protected by phosphatidylcholine from damage during the double-emulsion process but underwent inactivation during microsphere incubation [162]. Sodium dodecyl sulfate significantly reduced insulin aggregation at the dichloromethane–water interface, whereas dodecyl maltoside did not, this surfactant being more efficient at air–water or solid–water interfaces [163]. It should be mentioned that surfactants are used along with sugars, proteins, and polymers effectively for solubilization and stabilization purposes of peptide/protein in microspheres. Bilati et al. [130] give an overview on various proteins and polymers that act as stabilizing excipients during the development of peptide/protein-loaded microspheres.

### 5.2 Preservation-Imparting Excipients

***Prevention/Minimization of Moisture-Induced Instability*** Moisture- and microclimate acid pH–induced instability (typically the aggregation) of the peptide/protein

encapsulated in PLGA microspheres has been monitored. Even several formulation strategies to inhibit these instability problems are being actively investigated. If the protein is expected to exist in the solid state within the PLGA polymer, the protein is remarkably prone to aggregation when formulated under conditions that allow moisture- and microclimate acid pH–induced instability. The two covalent aggregation mechanisms commonly described during exposure of the solid protein to moisture are the disulfide interchange/exchange [164] and the FMAP, which is operative for protein antigens that have been detoxified with formaldehyde exposure [124]. In the former pathway, the reaction is typically initiated by a thiolate ion on the protein or free thiolate ions that accompany β elimination of an intact disulfide [165]. Decreasing the concentration of the reactive species (e.g., lowering pH to favor the nonionized thiol, covalently blocking the thiol group, or oxidizing free thiols as they appear with divalent copper ion) has been shown to block this mechanism [165]. To inhibit the FMAP, strongly formaldehyde-interacting amino acids such as histidine and lysine [166] have been colyophilized with the formalinized protein antigen. On exposure to moisture, the amino acids appear to bind with the reactive Schiff base or equivalent electrophile [167] in the protein before a neighboring protein nucleophile can react to form an intermolecular cross-link [124]. Sorbitol has also been identified to inhibit the FMAP of TT at the maximal aggregating water content of the antigen, about 30 g $H_2O$/g protein [168], although whether this is a humectant effect [169] or a possible covalent reaction with the highly reactive electrophile in the antigen has not been determined. Several techniques have been developed to successfully bypass the destabilizing stress either by altering the role of water in the solid or immobilizing the protein or, alternatively, by directly inhibiting the aggregation. Clearly, one of the most significant findings in the field of peptide/protein stability in polymers is the success of the immobilization strategy of $Zn^{2+}$ precipitation, as performed with human growth hormone [37, 57, 170, 171]. The 2:1 mole ratio Zn–protein complex, which immobilizes the rhGH as a solid precipitate in a near-native state [99], has been shown to confer superior stability on the protein encapsulated in PLGA for a one-month release incubation. Since then, other proteins such as INF-α [172] and NGF [173] were also stabilized in PLGA microspheres by this approach. Another interesting approach originating in the patent literature is the precipitation of erythropoietin with salting-in salt, ammonium sulfate [174], which is a technique commonly used in protein processing. Other methods to alter the role of water in the reaction involve the addition of agents that alter the amount of water sorbed in the polymer and/or the activity of the water present. For example, both water-soluble salts (NaCl) and antacid excipients ($Mg(OH)_2$) are known to dramatically increase the amount of water sorbed in PLGAs, with the former due to osmosis and the latter to a complex effect of neutralizing acidic degradation products and end groups of the polymer (which also involves an osmotic component) [126]. In contrast, for a given moisture content, humectants such as sorbitol, which dissolve in water bound to the protein, reduce the available free water necessary to mobilize the protein or perform other roles in deleterious reactions [169].

The alternative to bypassing the deleterious role of moisture is to inhibit the aggregation mechanism directly. Several ways to accomplish this have been reported, particularly in the solid state and in the absence of the polymer. Well-referenced and useful book chapters by Johnson [175] and Carpenter and Chang [176] are

available to thoroughly focus on the importance of making a lyophilized powder before loading the peptide or protein into an injectable microspheres. It has been stated that, in comparison to protein solution, the protein in the solid state would be less susceptible to shear forces that occur during an emulsification procedure or denaturation at oil–water interfaces. However, special precautions should be taken during freeze drying because the drying process itself will expose the protein to destabilizing stresses. To circumvent this problem, cryo- and lyoprotectants and bulking agents are usually included along with a peptide or protein solution while it undergoes the drying stages of the lyophilization process.

*Cryo- and Lyoprotectants and Bulking Agents* Various mechanisms are proposed to explain why excipients serve as cryo- or lyoprotectants. The most widely accepted mechanism to explain the action of cryoprotection is the preferential exclusion mechanism [177]. Excipients that will stabilize proteins against the effects of freezing do so by not associating with the surface of the protein. Such excipients actually increase the surface tension of water and induce preferential hydration of the protein. Examples of solutes that serve as cryoprotectants by this mechanism include amino acids, polyols, sugars, and polyethylene glycol.

For lyoprotection, that is, stabilization of proteins during the drying stages of freeze drying and during storage in the dry state, two mechanisms are generally accepted. One is the water substitute hypothesis [178] and the other is the vitrification hypothesis [179]. Both are legitimate theories, but both also have exceptions; that is, neither fully explain the stabilization of proteins by excipients during dehydration and dry storage [180]. The water substitute hypothesis states that a good stabilizer is one that hydrogen bonds to the protein just as water would do where it presents and, therefore, serves as a water substitute. Sugars are good water substitutes. (It may at first appear contradictory that sugars can serve both as cryoprotectants because of being excluded from the surface of the protein and as lyoprotectants that hydrogen bond to the protein. However, keep in mind that the excluded solute concept involves a frozen aqueous system whereas the water substitute concept occurs in a dry system.) This is why many freeze-dried protein formulations contain sucrose or trehalose. Nevertheless, during a w/o/w procedure to prepare peptide/protein-loaded PLGA microspheres, sugars are often added to the inner aqueous phase. Trehalose was shown to partially improve the BSA secondary-structure protection within PLGA microspheres and to facilitate BSA monomer release [26]. Trehalose and mannitol had a significant effect on the recovery of soluble nonaggregated interferon-γ (INF-γ) and rhGH after emulsification and ultrasonication [61], whereas no or very little protecting effect on IGF-1 against these stress factors was observed [144]. No effect of trehalose, mannitol, and sucrose was observed against o/w interface-induced degradation of lysozyme, whereas lactose and lactulose significantly improved its structural stability and activity, mostly if these additives were also added to the second aqueous phase [88, 100]. Lysozyme and trypsin activity was not improved by addition of sucrose, which was unable to protect them from an emulsion-induced denaturation and from sonication [98, 181]. Mannitol and sucrose dissolved together in the inner aqueous phase had slight effect on NGF activity [182] and neither mannitol nor lactose improved HBcAg immunogenicity during dichloromethane/water emulsification [145]. Surprisingly, sucrose and trehalose even decreased urease bioactivity, showing the opposite effect to that

expected [161]. Coencapsulation of maltose reduced α-chymotrypsin aggregation [183]. With respect to microspheres generated by spray drying, trehalose was effective in retaining TT antigenicity [147] and in preventing BSA secondary-structure degradation [35]. Trehalose protected efficiently NGF during the processing but did not prevent its aggregation during in vitro release [159].

The vitrification hypothesis states that excipients that remain amorphous (glass formers) form a glassy matrix with the protein with the matrix serving as a stabilizer. Acceptance of this hypothesis requires formulators to determine glass transition temperatures of formulations to be freeze dried and to develop freeze-dry cycles that maintain drying temperatures below the glass transition temperature. Reports are available to indicate that excipient stabilizers, which are capable of undergoing crystallization during storage, caused degradation (typically aggregation and loss of potency) of the protein [176, 184, 185].

Freeze-dried formulations typically contain one or more of the following bulking agents: mannitol, lactose, sucrose, trehalose, dextran 40, and povidone. These excipients may also serve as cryo- and/or lyoprotectants in protein formulations. Fakes et al. [186] studied these bulking agents for moisture sorption behavior before and after freeze drying. Moisture uptake certainly can affect drug stability in the freeze-dried state, particularly with peptides and proteins. When selecting a bulking agent, these properties, particularly the tendency for moisture uptake, must be considered by the formulation scientist in developing an optimally stable freeze-dried formulation. Several excipients can serve as stabilizers for proteins that are unstable during the drying phases of freeze drying and/or during long-term storage in the dry state. Typically, additives that will crystallize during lyophilization (e.g., mannitol) or will remain amorphous but unable to hydrogen bond to the dried protein (e.g., dextran) are not effective lyoprotectants for proteins. Excipients that will crystallize during freeze drying will also be relatively ineffective, as was shown with sucrose in *H. lanuginosa* lipase formulations [149]. However, these authors also reported that sucrose crystallization could be inhibited by decreasing the mass ratio of sucrose to protein and by minimizing the moisture content that serves to decrease the glass transition temperature during storage. The reverse can also be true for certain small molecules. For example, excipients (mannitol or sodium bicarbonate) that promoted the crystallization of cyclophosphamide during freeze drying stabilized the final product whereas excipients that did not allow crystallization (e.g., lactose) destabilized the final product [187]. Costantino et al. [188] studied the effects of a variety of parenteral excipients on stabilizing human growth hormone in the lyophilized state. Mannitol, sorbitol, methyl *a*-D-mannopyranoside, lactose, trehalose, and cellobiose all provided significant protection of the protein against aggregation, particularly at levels (131:1 excipient-to-protein molar ratio) to potentially satisfy water binding sites on the protein in the dried state. At higher excipient-to-protein ratios, mannitol and sorbitol crystallized and were not as effective in stabilizing the protein compared with low levels in which they remained in the amorphous, protein-containing phase.

Reducing sugars may not be as effective as other bulking agents, cryoprotectants, or lyoprotectants because they may potentially react with proteins via the Maillard reaction. For example, glucose will form covalent adducts with side-chain amino acids lysine and arginine of human relaxin [189]. In addition, a significant amount of serine cleavage from the C terminal of the B chain of relaxin was formed when

glucose was used as the excipient. These reactions did not occur if mannitol and trehalose replaced glucose in the lyophilized formulation. Lactose will react with primary amines in the well-known Maillard-type condensation reaction to form brown-colored degradation products [190]. Thus, lactose is known to be incompatible with amine-containing compounds such as aminophylline, amphetamines, and amino acids/peptides. This reaction occurs more readily with amorphous lactose than crystalline lactose.

Hydrophilic additives such as glucose are known to increase the porosity of microspheres, causing an increase in permeability to mass transport and a higher burst. However, a significant reduction in initial burst release of a highly water-soluble model peptide, octreotide acetate, from poly(D,L-lactide-*co*-glycolide) microspheres by the coencapsulation of a small amount of glucose (e.g., 0.2% w/w) was reported [191]. Using the double emulsion–solvent evaporation method of encapsulation, the effect of glucose on initial burst in an acetate buffer pH 4 was found to depend on polymer concentration, discontinuous phase/continuous phase ratio, and glucose content. Extensive characterization studies were performed on two microsphere batches, ±0.2% glucose, to elucidate the mechanism of this effect. However, no significant difference was observed with respect to specific surface area, porosity, internal and external morphology, and drug distribution. Continuous monitoring of the first 24-h release of octreotide acetate from these two batches disclosed that, even though their starting release rates were close, the microspheres plus glucose exhibited a much lower release rate between 0.2 and 24 h compared to those without glucose. The microspheres plus glucose showed a denser periphery and a reduced water uptake at the end of the 24-h release, indicating decreased permeability. However, this effect at times was offset as glucose content was further increased to 1%, causing an increase in surface area and porosity. In summary, these authors concluded that the effects of glucose on initial burst are determined by two factors: (1) increased initial burst due to increased osmotic pressure during encapsulation and drug release and (2) decreased initial burst due to decreased permeability of microspheres [191].

Mannitol is probably the most widely used bulking agent in lyophilized formulations because of its many positive properties with respect to crystallinity, high eutectic temperature, and matrix properties. However, some lots of mannitol can contain reducing sugar impurities that were implicated in the oxidative degradation of a peptide in a lyophilized formulation [192]. Mannitol at or above certain concentrations and volumes in glass vials is well known to cause vial breakage because of the unique crystallization properties of mannitol-ice during the primary drying states of freeze drying [193, 194].

*Other Freeze-Dry Excipients* High-molecular-weight carbohydrates such as dextran have higher glass transition temperatures than peptides/proteins. Therefore, when mixed with proteins, the overall glass transition temperature presumably can be increased with resultant increases in protein storage stability. Typically, carbohydrates (sucrose, trehalose, or dextran) alone do not result in appreciable increases in the storage stability of proteins. However, combinations of disaccharide and polymeric carbohydrates do tend to improve protein storage stability [195]. However, singular carbohydrates (sucrose or trehalose at 60 m*M*) were also just as effective in stabilizing a model recombinant humanized monoclonal antibody as combinations of sucrose and mannitol or trehalose and mannitol. Interestingly, with this

model monoclonal antibody, mannitol alone at 60 m*M* provided less protection during storage than sucrose or trehalose alone. A specific sugar/protein molar ratio was sufficient to provide storage stability for this particular monoclonal antibody [196, 197].

Low-molecular-weight additives such as osmolytes (*N,N*-dimethylglycine, trehalose, and sucrose) or salts (sodium chloride, sodium phosphate, ammonium sulfate, and sodium citrate) were found to be highly effective in stabilizing keratinocyte growth factor, both against thermal denaturation and enhancing long-term storage stability [198]. Nevertheless, the stabilizing properties of osmolytes appear to be balanced between their binding to (deteriorating effect) and exclusion from (stabilizing effect) the peptide/protein surface. As binding or exclusion predominantly results from hydrophobic interactions, hydrogen bonding, and electrostatic interactions, the sum of the various interaction parameters are dissimilar for different proteins. Therefore, it becomes crucial to examine the individual nature of the additive toward each individual protein and to assess whether it will offer a stabilizing or destabilizing effect [199, 200].

Polyvinyl pyrrolidone (PVP) and glycine were found to stabilize lyophilized sodium prasterone sulfate whereas dextran 40 or mannitol did not [201]. PVP and glycine stabilized the pH of the reconstituted solution by neutralizing the acidic degradation product, sodium bisulfate, formed by the hydrolysis of prasterone sulfate. Dextran 40 or mannitol was ineffective because of no buffer capacity. Buffering agents, such as phosphate–citrate buffer and some neutral and basic amino acids (L-arginine, L-lysine, and L-histidine), also stabilized prasterone sulfate. L-Cysteine is an example of an amino acid that did not stabilize the drug, presumably because of its weak buffer capacity. Although the efficiency of proteinic additives for protein stabilization has been clearly demonstrated in several occasions even during encapsulation processes [31, 72, 98, 144], their use in pharmaceuticals is at present not desirable from a strictly regulatory point of view. Additionally, such agents might contribute to complicate all subsequent protein characterization within the formulation. Among these additives, albumins and gelatins are those mainly used for protection purposes. The protective effect of albumins against protein unfolding and aggregation has been extensively documented and is likely due to their surface-active properties (see Bilati et al. [130] for details).

***Prevention/Minimization of Microclimate pH-Induced Instability*** Evidence for acidification within degrading microspheres is investigated and local pH values between 1.5 and 4.7 are being reported [202–204]. Methods to measure microclimate pH in PLGA microspheres include (i) ensemble average measurements using electron paramagnetic resonance (EPR) [203, 204], nuclear magnetic resonance (NMR) [205], and potentiometry and (ii) direct visualization techniques such as confocal imaging of pH-sensitive dyes [206, 207]. In the EPR method, the constant of hyperfine splitting, $2aN$, was used to determine an average pH inside PLGA microspheres. Because the experiments relied on the mobility of spin-labeled protein, with an increase of the microviscosity in the later hours of the experiments, the spectra of EPR was changed and the signal-to-noise ratio decreased to prevent the measurement of pH throughout the release period [203]. The potentiometric measurements can give rapid values of pH for thin polymer films, and the pH of the thin water film between the electrode and polymer mimics the microclimate pH of aqueous pores inside the polymer-based drug delivery system. However, it is difficult to mimic

microclimate pH of a small-scale system, such as microspheres or nanospheres, which may have unique microstructures, excipient/drug distributions, and transport characteristics. Overall, the ensemble average measurements described above could give a general picture of microclimate pH at the macroscopic level. However, the microscopic level of the detection can only be achieved through direct visualization techniques, such as microscopic imaging. Shenderova et al. [207] first developed the confocal microscope imaging method to relate the microclimate pH with the fluorescent intensity. Because of the difficulty of controlling and predicting the fluorescein concentration in the aqueous pore inside the microsphere, the method was only semiquantitative. Fu et al. [206] improved the confocal microscopic imaging method by coencapsulating two dextran fluorescent dye (NERF and SNARF-1) conjugates inside microspheres and related the ratio of the two dye images with microclimate pH in order to eliminate the poorly controlled effects of dye concentration and pore distribution. However, both of the dyes emit in the green range (535 nm for NERF and 580 nm of SNARF), giving rise to poor resolution without a narrow-bandwidth detector. Because of the high noise-to-signal ratio from the ratio images, the prediction of pH is also expected to be semiquantitative. In order to overcome the aforementioned drawbacks in microclimate pH measurement, a new quantitative ratiometric method based on laser scanning confocal microscopic imaging was developed to create a pixel-by-pixel neutral range microclimate pH map inside PLGA microspheres [208]. This method was then applied to both acid-neutralized and nonneutralized PLGA microspheres during extended incubation in physiological buffer. In another study, the PLGA water-soluble acid distribution has been measured with prederivatization HPLC [209].

Ongoing acidification of the microsphere interior was shown to induce deamidation and covalent dimerization of nonreleased insulin [202]. Despite the evidence of acidification mentioned above, there is controversy on this subject. It has been pointed out that the sampling scheme has a significant impact on the degree of acidification; frequent replenishment of the release medium or the use of a dialysis bag can effectively prevent the acidification of the medium with subsequent reduced protein degradation [93, 210]. It is unsure, however, whether this also reflects the situation in vivo, in which the PLGA microspheres are often surrounded by a fibrous capsule that may reduce efflux of acidic degradation products from the PLGA matrix [93]. On the other hand, studies on rhGH-loaded PLGA microspheres showed a reasonable in vitro–in vivo correlation (IVIVC) only when a strong high-capacity buffer [200 m*M* *N*-(2-hydroxyethyl)piperazine-*N*′-2-ethanesulfonic acid (HEPES), pH 7.4] was used, which effectively minimized the pH drop [211].

As indicated by the prevention of acid-induced physical aggregation of BSA in an abstract [212], three principal ways have been identified thus far to avoid the formation of highly acidic microclimate regions in the PLGAs during protein release:

(i) Increasing the permeability of the polymer to facilitate escape of the water-soluble hydrolytic products of the PLGA polyester [125]
(ii) Decreasing the degradation rate of the polyester [213]
(iii) Coencapsulating additives to neutralize the weak acids formed by PLGA hydrolysis [126]

In addition, two more ways that are likely to favor a lowering of microclimate pH are elevated initial acid content in the polymer [214] and low-frequency release media exchange [206].

The concept of controlling polymer permeability is difficult because attempts to increase permeability can spoil the controlled-release function of the polymer and cause the encapsulated protein to be released too rapidly. For example, Jiang and Schwendeman [213] increased the permeability of slow-degrading PLA (molecular weight (MW) 145 kDa) by blending in PEG (MW 10 or 30 kDa) at 0, 10, 20, and 30%. Insoluble BSA aggregation in the PLA microspheres containing 4.5–5% w/w BSA was found in 0 and 10% PEG after a one-month incubation, but not in those preparations containing 20 or 30% PEG. The structural integrity of BSA was also intact in the stabilized formulations. However, between 10 and 30% PEG, the release rate of BSA increased rapidly and by 30% PEG, 60% of the protein encapsulated was released in only three days [213]. In contrast, an abstract [212] implied that 5% BSA encapsulated in a more permeable PLA (MW 77 kDa), the BSA formed <2% insoluble aggregates over one month. This result suggested strongly that in some instances the slow degradation rate of the non-glycolic-acid-containing PLA is sufficient to inhibit acid formation in the microclimate.

In instances in which it is desirable to increase permeability and/or decrease the hydrolytic rate of PLGAs, that is, where a highly water-soluble protein requires release for one month or longer, it becomes necessary to coencapsulate a basic additive. Antacids such as $MgCO_3$, $Mg(OH)_2$ or $ZnCO_3$ have been found to be particularly potent in preventing instabilty of acid-labile proteins [126, 215]. By means of thin films coating pH glass electrodes to measure directly the microclimate acidic environment in PLGA microspheres, the stabilization against insoluble acid-induced noncovalent BSA aggregation afforded by a series of antacid excipients has been correlated with the ability of the antacid to neutralize acidic pores in films of the same lot of PLGA coating pH glass electrodes [215]. Though much of the physical chemistry of microclimate pH adjustment with antacid additives is currently unclear, the strength of the base, the base solubility, and the association of the divalent cation with the carboxylate of the degradation products and/or polymer end groups appear to be important. For instance, Shenderova [216] has shown that from microclimate pH measurements in PLGA films coating pH glass electrodes, $MgCO_3$ and $Mg(OH)_2$ were found to be very similar under conditions which favor homogenous neutralization (i.e., high protein loading sufficient to make pores for the base to diffuse all regions of the polymer matrix), but $MgCO_3$ was found to increase microclimate pH higher than $Mg(OH)_2$. This result was consistent with the improved BSA stability in PLGA 50/50 microspheres when $MgCO_3$ was used in place of $Mg(OH)_2$ [126].

## 6 PEPTIDE/PROTEIN ENCAPSULATED INTO BIODEGRADABLE MICROSPHERES: CASE STUDY

Selected examples of therapeutic peptide and protein including vaccines which have been encapsulated into biodegradable polymer-based microspheres are discussed in this section. Besides what is mentioned below, many other proteins and vaccines have been encapsulated in biodegradable polymers, so a glimpse of ongoing

research on microsphere delivery systems using biodegradable polymers is shown in Table 2.

### 6.1 Vaccines

***Group B Streptococcus Vaccine*** Group B streptococcus (GBS) is the leading bacterial cause of neonatal sepsis and meningitis. Although antibiotic prophylaxis has decreased the infection rate, the best long-term solution lies in the development of effective vaccines. The GBS capsular polysaccharide (CPS) is a major target of antibody-mediated immunity. The feasibility of producing a GBS having the ability to produce both a local IgA immune response at the mucosal surface and humoral IgG response having capability of transplacental passive immunization was investigated [217]. Inactivated GBS antigen was encapsulated in PLGA by a w/o/w multiple-emulsion technique along with immunostimulatory synthetic oligodeoxynucleotides containing cytosine phosphate guanosine (CpG) as potent adjuvant [217]. Immunization of female mice with normal immune systems was done with these PLGA microspheres containing GBS type III polysaccharide and CpG adjuvant (PLGA/GBS/CpG) and results indicated a significantly higher GBS antibody response as compared to nonencapsulated GBS antigen or PLG-encapsulated GBS PS vaccine without the addition of the CpG.

***Diphtheria Toxoid (DT)*** Diphtheria is a communicable disease caused by *Corynebacterium diphtheriae* which colonizes and forms a pseudomembrane at the infection site. This pathogen produces a potent protein toxin, diphtheria toxin, which is responsible for the typical systemic toxemia. DT is required for active immunization against diphtheria. DT was encapsulated in different types of PLA and PLGA microspheres by spray drying and coacervation. Immunization of guinea pigs with DT microspheres made with relatively hydrophilic PLGA 50:50 resulted in specific and sustained antibody responses to alum adjuvanted toxoid in contrast to microspheres made with hydrophobic polymers where very low antibody responses were determined confirming the feasibility of microsphere vaccines to induce strong, long-lasting protective antibody responses after single immunization [218].

In an endeavor toward development of multivalent vaccines based on biodegradable microspheres, Peyre et al. [219] tested the immunological performance of several divalent microsphere formulations against tetanus and diphtheria. Microspheres were made by separate microencapsulation of tetanus and diphtheria toxoid in PLGA by either spray drying or coacervation. Guinea pigs were subcutaneously immunized by a single injection of the divalent vaccines or, for control, an equivalent dose of a licensed vaccine containing both antigens adsorbed on aluminum hydroxide. All microsphere formulations were strongly immunogenic, irrespective of particle size and hydrophobicity. Endpoint titers of ELISA antibodies, mainly of the IgG1 subtype, were comparable to those obtained after immunization with the licensed vaccine. The microspheres provided increasing levels of antibodies, during the 16 weeks of testing, and the antibodies were weakly polarized toward tetanus. The induced antibodies were also toxin neutralizing, as determined for both diphtheria (1–4 IU/mL) and tetanus (5–9 IU/mL) eight weeks after immunization. These neutralization levels were several orders of magnitude above the level considered

minimum for protection (0.01 IU/mL). When the animals were challenged with tetanus or diphtheria toxins six weeks after immunization, microsphere vaccines produced protective immunity that was comparable to or better than that induced by the licensed divalent vaccine. In conclusion, this study showed that a single administration of biodegradable microsphere vaccines provided protective immunity against diphtheria and tetanus and that this immunization approach might be feasible for multivalent vaccines. In a separate study, the same group have studied for the first time the fate of immunogenic fluorescent-labeled PLGA microspheres loaded with DT in vivo following a subcutaneous injection route [220].

A unique instability problem of DT is being foreseen when the DT would be encapsulated in PLGA microspheres along with a preservative such as thiomersal [221]. Thimerosal (TM)—also known as thiomersal, Merthiolate, or sodium ethylmercuri-thiosalicylate—is a water-soluble derivative complex of thiosalicylic acid (TSA) that has been used as bactericide in parenteral vaccines and ophthalmic products for decades. It has been reported that this preservative can be decomposed by oxidation to 2,2-dithiosalicylic acid, ethyl mercuric ion, 2-sulfenobenzoic acid, 2-sulfobenzoic acid, and 2-sulfinobenzoic acid [222]. Namura et al. [221] demonstrated in vitro that the TSA, produced after the reduction of TM by lactic acid, reduces the S–S bridge of the previously incubated DT. This reduction is immediately followed by blocking the two SH groups formed by the same TSA molecules. In light of these conclusions, it is necessary now to reinterpret the in vitro protein degradation–stabilization data in the presence of PLGA microsphere, mainly for those proteins which contain S–S. The authors propose that all the PLGA microsphere microencapsulation studies and protein structural considerations should be done in the absence of TM as preservative.

***Tetanus Toxoid (TT)*** Tetanus is considered a major health problem in developing and underdeveloped countries, with approximately one million new cases occurring each year. Tetanus is an intoxication manifested primarily by neuromuscular dysfunction. So vaccination is required for prevention of this disease. TT was encapsulated using PLGA with different molar compositions (50:50, 75:25) by the w/o/w multiple-emulsion technique and protein integrity was evaluated during antigen release in vitro in comparison to alum-adsorbed TT for in vivo induction of tetanus-specific antibodies [223]. TT microspheres elicited antibody titers as high as conventional alum-adsorbed TT, which lasted for 29 weeks, leading to the conclusion that TT microspheres can act as potential candidates for single-shot vaccine delivery systems.

The study by Determan et al. [224] focuses on the effects of polymer degradation products on the primary, secondary, and tertiary structure of TT, OVA, and lysozyme after incubation for 0 or 20 days in the presence of ester (lactic acid and glycolic acid) and anhydride [sebacic acid and 1,6-bis(*p*-carboxyphenoxy)hexane] monomers. The structure and antigenicity or enzymatic activity of each protein in the presence of each monomer was quantified. SDS-PAGE, circular dichroism, and fluorescence spectroscopy were used to assess/evaluate the primary, secondary, and tertiary structures of the proteins, respectively. ELISA was used to measure changes in the antigenicity of TT and OVA and a fluorescence-based assay was used to determine the enzymatic activity of lysozyme. TT toxoid was found to be the most stable in the presence of anhydride monomers, while OVA was most stable in the

presence of sebacic acid, and lysozyme was stable when incubated with all of the monomers studied.

Jaganathan et al. [225] compared the efficiency of microspheres produced from PLGA and chitosan polymers by using protein stabilizer (trehalose) and acid-neutralizing base [$Mg(OH)_2$]. The immunogenicity of PLGA- and chitosan microsphere–based single-dose vaccine was evaluated in guinea pigs and compared with multiple doses of alum-adsorbed TT. Results indicated that a single injection of PLGA and chitosan microspheres containing TT could maintain the antibody response at a level comparable to the booster injections of conventional alum-adsorbed vaccines. Both the PLGA- and chitosan-based stable vaccine formulations produced an equal immune response. Hence chitosan can be used to replace the expensive polymer PLGA. This approach should have potential application in the field of vaccine delivery.

The study by Kipper et al. [226] focuses on the development of single-dose vaccines based on biodegradable polyanhydride microspheres that have the unique capability to modulate the immune response mechanism. The polymer system employed consists of copolymers of 1,6-bis(*p*-carboxyphenoxy)hexane and sebacic acid. Two copolymer formulations that have been shown to provide extended-release kinetics and protein stability were investigated. Using TT as a model antigen, in vivo studies in C3H/HeOuJ mice demonstrated that the encapsulation procedure preserves the immunogenicity of the TT. The polymer itself exhibited an adjuvant effect, enhancing the immune response to a small dose of TT. The microspheres provided a prolonged exposure to TT sufficient to induce both a primary and a secondary immune response (i.e., high antibody titers) with high-avidity antibody production, without requiring an additional administration. Antigen-specific proliferation 28 weeks after a single immunization indicated that immunization with the polyanhydride microspheres generated long-lived memory cells and plasma cells (antibody-secreting B cells) that generally do not occur without maturation signals from T helper cells. Furthermore, by altering the vaccine formulation, the overall strength of the T-helper type 2 immune response was selectively diminished, resulting in a balanced immune response, without reducing the overall titer. This result is striking, considering free TT induces a T-helper type 2 immune response and has important implications for developing vaccines to intracellular pathogens. The ability to selectively tune the immune response without the administration of additional cytokines or noxious adjuvants is a unique feature of this delivery vehicle that may make it an excellent candidate for vaccine development.

Polylactide (PLA) polymer particles entrapping TT were evaluated in terms of particle size, antigen load, dose, and additional adjuvant for achieving high and sustained anti-TT antibody titers from single-point intramuscular immunization [227]. Admixture of polymer-entrapped TT and alum improved the immune response in comparison to particle-based immunization. High and long-lasting antibody titer was achieved upon immunization with 2–8-μm size microparticles. Microspheres within the size range 50–150 μm elicited very low serum antibody response. Immunization with very small particles (<2 μm) and with intermediate-size-range particles (10–70 μm) elicited comparable antibody response from single-point immunization but lower in comparison to that achieved while immunizing with 2–8-μm particles. Potentiation of antibody response on immunization of admixture of microspheres and alum was also dependent on particle size. These results indicate the need of

optimal particle sizes in micrometer ranges for improved humoral response from single-point immunization. Increasing antigen load on polymer particles was found to have a positive influence on the generation of antibody titers from particle-based immunization. Maximum peak antibody titer of ~300 μg/mL was achieved on day 50 upon immunization with particles having the highest load of antigen (94 μg/mg of polymer). Increase in dose of polymer-entrapped antigen resulted in concomitant increase in peak antibody titers, indicating the importance of antigen stability, particle size, and load on generating a reproducible immune response. Optimization of particle size, antigen load, dose, and use of additional adjuvant resulted in high and sustained anti-TT antibody titers over a period of more than 250 days from single-point immunization. Serum anti-TT antibody titers from single-point immunization of admixture of PLA particles and alum were comparable with immunization from two divided doses of alum-adsorbed TT.

**Vibrio Cholerae *(VC) Whole-Cell Vaccine*** Cholera, an acute intestinal infection caused by the bacterium *Vibrio cholerae*, produces an enterotoxin that causes a copious, painless, watery diarrhea that can quickly lead to severe dehydration and death if treatment is not promptly given. For prevention of cholera, cholera vaccine is usually given. VC was successfully entrapped in the PLGA microspheres by a double-emulsion method with trapping efficiencies up to 98%. The immnunogenic potential of VC-loaded microspheres physically mixed with or without amphotericin B was evaluated in adult mice by oral immunization in comparison to VC solution. The immunogenicity of VC-loaded microparticles mixed with amphotericin B in evoking *Vibrio*-specific serum IgG and IgM responses was higher than that of VC-loaded microparticles only [228, 229]. However, VC was loaded in different polymer compositions (50:50 PLGA, 75:25 PLGA, and PLA/PEG blended), the higher antibody responses and serum IgG, IgA, and IgM responses were obtained when sera from both VC-loaded 75:25 PLGA and PLA/PEG-blended microparticles immunized mice were titrated against VC solution [230].

***Japanese Encephalitis Virus (JEV)*** Japanese encephalitis is a disease that is spread to humans by infected mosquitoes in Asia. It is one of a group of mosquito-borne viral diseases that can affect the central nervous system and cause severe complications and even death. Vaccination is one of the ways of treating it. JEV vaccine was encapsulated in PLGA microspheres by a double-emulsion technique and influences of various process variables such as stirring rate, types and concentration of emulsifier, and polymer concentration were studied on size, size distribution, and biodegradation. The mean size of microspheres decreased with increasing speed, increasing concentration of emulsifier, and decreasing polymer concentration. Rate of biodegradation of nonporous microspheres was slower than that of porous microspheres, leading to the conclusion that PLGA microspheres can be used to apply oral vaccination through Peyers patches across the gastrointestinal tract (GIT) [231].

Several approaches to develop an improved JEV vaccine are in progress in various laboratories. Of these, immunization of mice with plasmid DNA encoding JEV envelope (E) protein has shown great promise. The technology, developed by Kaur et al. [232], involved the adsorption of DNA onto cetyltrimethyl-ammonium bromide (CTAB) containing cationic poly(lactide-*co*-glycolide) (PLG) microspheres.

The microsphere-adsorbed DNA induced a mixed Th1–Th2 immune response as opposed to Th1 immune responses elicited by the naked DNA.

JEV-loaded poly(lactide) (PLA) lamellar and PLG microspheres were successfully prepared with low-molecular-weight PLA by the precipitate method and with 6% w/v PLG in the organic phase, 10% w/v PVP, and 5% w/v NaCl in the continuous phase by using a w/o/w emulsion/solvent extraction technique, respectively [233]. The JEV incorporation, physicochemical characterization data, and animal results obtained in this study may be relevant in optimizing the vaccine incorporation and delivery properties of these potential vaccine targeting carriers.

***Hepatitis B Virus*** Hepatitis B is one of the most important infectious diseases in the world. Approximately 350 million people worldwide are chronic carriers of the hepatitis B virus (HBV), which accounts for approximately one million deaths annually. PLGA microspheres loaded with recombinant HBsAg were formulated using a double-emulsion technique. The pharmaceutical characteristics of size, surface morphology, protein loading efficiency, antigen integrity, release of HBsAg-loaded PLGA microspheres, and degradation of the polymer in vitro were evaluated [234–237]. Based on these findings in vitro and in vivo, it was concluded that HBsAg was successfully loaded into the PLGA microspheres, which can autoboost an immune response, and the HBsAg-loaded PLGA microsphere is a promising candidate for the controlled delivery of a vaccine.

### 6.2 Proteins

***Prolidase*** Deficiency of this enzyme results in chronic intractable ulcerations of the skin, particularly of lower limbs, since it is involved in the final stages of protein catabolism. To counteract the problem, the enzyme was encapsulated in PLGA microspheres by a double- or multiple-emulsion technique, in vitro and ex vivo evaluations were done, and the results indicated that microencapsulation stabilizes the enzymatic activity inside the PLGA microspheres resulting in both in vitro and ex vivo active enzyme release, hence opening the doors for the possibility of enzyme replacement therapy through microencapsulation [238]. Further evaluation from the same research group for prolidase-loaded PLGA microspheres is reported elsewhere [239, 240].

***Insulin*** Insulin is the most important regulatory hormone in the control of glucose homeostasis. The World Health Organization (WHO) has indicated that more than 50 million people around the world suffer from diabetes and require daily parenteral injections of insulin to stay healthy and live normally. For the treatment of type I diabetes insulin still is number one, with three subcutaneous injections to be taken per day. A controlled-release system for a long-term therapy of this disease is the need of the hour, as this can obviate the need for painful injection given a number of times to the diabetes patients. Insulin was encapsulated in blends of poly(ethylene glycol) with PLA homopolymer and PLGA copolymer by a w/o/w multiple-emulsion technique with entrapment efficiencies up to 56 and 48% for PLGA/ PEG and PLA/ PEG, respectively [12]. Insulin-loaded microspheres were capable of controlling the release of insulin for 28 days with in vitro delivery rates of 0.94 and 0.65 μg insulin/mg per particle per day in the first 4 days and steady release with a

rate of 0.4 and 0.43 μg insulin/mg per particle per day over the following 4 weeks, respectively, along with the extensive degradation of PLGA/ PEG microspheres as compared to PLA/ PEG blends which resulted in a stable particle morphology along with reduced fragmentation and aggregation of associated insulin.

Two types of injectable cationized microspheres were prepared based on a native gelatin (NGMS) and aminated gelatin with ethylenediamine (CGMS) to prolong the action of insulin [241]. Release of rhodamin B isothiocyanate insulin from CGMS was compared with that from NGMS under in vitro and in vivo conditions. Lower release of insulin from CGMS compared with that from NGMS was caused by the suppression of initial release. The disappearance of $^{125}$I-insulin from the injection site after intramuscular administration by NGMS and CGMS had a biphasic profile in mice. Almost all the $^{125}$I-insulin had disappeared from the injection site one day after administration by NGMS. The remaining insulin at the injection site after administration by CGMS was prolonged, with approximately 59% remaining after 1 day and 16% after 14 days. The disappearance of CGMS from the injection site was lower than that of NGMS. However, the difference in these disappearance rates was not great compared with those of $^{125}$I-insulin from the injection site by NGMS and CGMS. The time course of disappearance of $^{125}$I-CGMS from the injection site was similar to that of $^{125}$I-insulin by CGMS. The initial hypoglycemic effect was observed 1 h after administration of insulin by NGMS, and thereafter its effect rapidly disappeared. The hypoglycemic effect was observed 2–4 h after administration by CGMS and continued to be exhibited for 7 days. The prolonged hypoglycemic action by CGMS depended on the time profiles of the disappearance of insulin from muscular tissues, which occurs due to the enzymatic degradation of CGMS.

A novel controlled-release formulation was developed with PEGylated human insulin encapsulated in PLGA microspheres that produces multiday release in vivo [242]. The insulin is specifically PEGylated at the amino terminus of the B chain with a relatively low molecular weight PEG (5000 Da). Insulin with this modification retains full biological activity but has a limited serum half-life, making microencapsulation necessary for sustained release beyond a few hours. PEGylated insulin can be codissolved with PLGA in methylene chloride and microspheres made by a single o/w emulsion process. Insulin conformation and biological activity are preserved after PEGylation and PLGA encapsulation. The monolithic microspheres have inherently low burst release, an important safety feature for an extended-release injectable insulin product. In PBS at 37°C, formulations with a drug content of approximately 14% show very low (<1%) initial release of insulin over one day and near-zero-order drug release after a lag of three to four days. In animal studies, PEG-insulin microspheres administered subcutaneously as a single injection produced <1% release of insulin in the first day but then lowered the serum glucose levels of diabetic rats to values <200 mg/dL for approximately nine days. When the doses were given at seven-day intervals, steady-state drug levels were achieved after only two doses. PEG-insulin PLGA microparticles show promise as a once-weekly dosed, sustained-release insulin formulation.

Shenoy et al. [243] developed an injectable, depot-forming drug delivery system for insulin based on microparticles technology to maintain constant plasma drug concentrations over a prolonged period of time for the effective control of blood sugar levels. Formulations were optimized with two well-characterized biodegradable polymers, namely PLGA and poly-ε-caprolactone, and evaluated in vitro for

physicochemical characteristics, drug release in phosphate-buffered saline (pH 7.4), and evaluated in vivo in streptozotocin-induced hypoglycemic rats. With a large volume of internal aqueous phase during a w/o/w double-emulsion solvent evaporation process and high molecular weight of the polymers used, they could not achieve high drug capture and precise control over subsequent release within the study period of 60 days. However, this investigation revealed that upon subcutaneous injection the biodegradable depot-forming polymeric microspheres controlled the drug release and plasma sugar levels more efficiently than plain insulin injection. Preliminary pharmacokinetic evaluation exhibited steady plasma insulin concentration during the study period. These formulations, with their reduced frequency of administration and better control over drug disposition, may provide an economic benefit to the user compared with products currently available for diabetes control.

***Interferon $\alpha_{2a}$ (IFN $\alpha_{2a}$)*** Interferon $\alpha_{2a}$ is indicated for the treatment of adults with chronic hepatitis C virus infection who have compensated liver disease and have not been previously treated with interferon α. To improve the stability and loading efficiency of protein drugs, a new microsphere delivery system comprises calcium alginate cores surrounded by PELA [poly-*D,L*-lactide-poly-(ethylene glycol)]. Recombinant IFN$\alpha_{2a}$ as a model drug was entrapped within calcium alginate cores surrounded by PELA by a w/o/w multiple-emulsion technique [244]. Core-coated microspheres stabilized the IFN in the PELA matrix. The core-coated microspheres indicated high encapsulation efficiency and biological retention as compared to conventional PLGA microspheres. The extent of burst release reduced to 14% in core-coated microspheres from 31% in conventional microspheres, indicating a new approach for water-soluble macromolecular drug delivery.

## 7 CONCLUSION

From this chapter, it has become apparent that a number microencapsulation methods are available today for the preparation of microspheres on an industrial scale. In fact, parenteral drug delivery systems based upon biodegradable microspheres are a true success story for the concept of drug delivery. However, the production of biodegradable microspheres containing a stable therapeutic peptide or protein still remains a major challenge in terms of technical obstacles. Ideally, peptides/proteins of therapeutic interest should be studied case by case, so as to bring to the fore processing steps and stress factors which damage them. Continued efforts to establish methods for stable protein, especially antigen, delivery from microspheres may hopefully pave the way for future microsphere-based vaccines. Areas of further research should focus on the performance of peptide/protein-loaded microspheres under in vitro and in vivo conditions. Interestingly, the addition of medium-chain triglycerides (MCT) modifies/shifts the triphasic release pattern of leuprolide acetate-loaded PLGA microspheres to a more continuous release in vitro [245]. Alternatively, BSA-loaded PLGA microspheres were coated with a thermosensitive gel, Pluronic F127 (PF127) [246]. The results demonstrated that PF127, which gelled at 37°C, inhibited the initial burst release of BSA from microspheres effectively. It is anticipated that more efforts will be invested in the future to develop

novel ways to reduce the initial burst release of entrapped peptide/protein and to attain a more continuous release. In addition, an in vitro release model mimicking the fate of biodegradable microspheres applied through the parenteral route would be highly desirable. Also, new strategies to stabilize proteins in microspheres during manufacturing, shelf life, or in vivo could be of general interest. Moreover, the use of analytical techniques such as FTIR or MALDI-TOF mass spectrometry certainly constitutes a step forward for protein analysis in more appropriate conditions.

## REFERENCES


1. Brooks, G. (1998), *Biotechnology in Healthcare*, Pharmaceutical Press, London.
2. Walsh, G. (2002), Biopharmaceuticals and biotechnology medicines: An issue of nomenclature, *Eur. J. Pharm. Sci.*, 15, 135–138.
3. Walsh, G. (2003), Pharmaceutical biotechnology products approved within the European Union, *Eur. J. Pharm. Biopharm.*, 55, 3–10.
4. New biotechnology medicines in development report (2002), available: http://pharma.org, accessed May 25, 2006.
5. Banga, A. K., and Chien, Y. W. (1988), Systemic delivery of therapeutic peptides and proteins, *Int. J. Pharm.*, 48, 15–50.
6. Lee, V. H. L. (1987), Ophthalmic delivery of peptides and proteins, *Pharm. Technol.*, 11, 26–38.
7. Athanasiou, K. A., Niederauer, G. G., and Agrawal, C. M. (1996), Sterilization, toxicity, biocompatibility and clinical applications of polylactic acid/polyglycolic acid copolymers. *Biomaterials*, 17, 93–102.
8. Ueda, H., and Tabata, Y. (2003), Polyhydroxyalkanonate derivatives in current clinical applications and trials, *Adv. Drug Deliv. Rev.*, 55, 501–518.
9. Ramesh, V. D., Medlicott, N., Razzak, M., and Tucker, I. G. (2002), Microencapsulation of FITC-BSA into poly(e-caprolactone) by a water-in-oil-in-oil solvent evaporation technique. *Trends Biomater. Artif. Organs*, 15, 31–36.
10. Dhanaraju, M. D., Vema, K., Jayakumar, R., and Vamsadhara, C. (2003), Preparation and haracterization of injectable microspheres of contraceptive hormones, *Int. J. Pharm.*, 268, 23–29.
11. Tracy, M. A. (1998), Development and scale-up of a microsphere protein delivery system, *Biotechnol. Prog.*, 14, 108–115.
12. Yeh, M. K. (2000), The stability of insulin in biodegradable microparticles based on blends of lactide polymers and polyethylene glycol, *J. Microencapsul.*, 17, 743–756.
13. Han, K., Lee, K. D., Gao, Z. G., and Park, J. J. (2001), Preparation and evaluation of poly(L-lactic acid) microspheres containing rhEGF for chronic gastric ulcer healing, *J. Controlled Release*, 75, 259–269.
14. Jackson, J., Liang, L., Hunter, W., Reynolds, M., Sandberg, J., Springate, C., and Burt, H. (2002), The encapsulation of ribozymes in biodegradable polymeric matrices, *Int. J. Pharm.*, 243, 43–47.
15. Blanco-Prieto, M. J., Besseghir, K., Zerbe, O., andris, D., Orsolini, P., Heimgartner, F., Merkle, H. P., and Gander, B. (2000), In vitro and in vivo evaluation of a somatostatin analogue released from PLGA microspheres, *J. Controlled Release*, 67, 19–28.
16. Yuksel, E., Weinfeld, A. B., Cleek, R., Waugh, J. M., Jensen, J., Boutros, S., Shenaq, S. M., and Spira, M. (2000), De novo adipose tissue generation through long-term, local delivery of insulin and insulin like growth factor-1 by PLGA/PEG microspheres in an in vivo rat midel: A novel concept and capability, *Plastic Reconstruct. Surg.*, 105, 1721–1729.

17. Kostanski, J. W., Thanoo, B. C., and Deluca, P. P. (2000), Preparation, characterization and in vitro evaluation of 1- and 4-month controlled release ornitide PLA and PLGA microspheres, *Pharm. Dev. Technol.*, 5, 585–596.
18. King, T. W., and Patrick, Jr., C. W. (2000), Development and in vitro characterization of vascular endothelial growth factor (VEGF) loaded poly(DL-lactic-*co*-glycolicacid)/poly(ethylene glycol) microspheres using a solid encapsulation/single emulsion/solvent extraction technique, *J. Biomed. Mater. Res.*, 51, 383–390.
19. Zhu, K. J., Jiang, H. L., Du, X. Y., Wang, J., Xu, W. X., and Liu, S. F. (2001), Preparation and characterization of hCG-loaded polylactide or poly(lactide-*co*-glycolide) microspheres using a modified water-in-oil-in-water (w/o/w) emulsion solvent evaporation technique, *J. Microencapsul.*, 18, 247–260.
20. Prabhu, S., Sullivan, J. L., and Betageri, G. V. (2002), Comparative assessment of in vitro release kinetics of calcitonin polypeptide from biodegradable microspheres, *Drug Deliv.*, 9, 195–198.
21. Oldham, J. B., Lu, L., Zhu, X., Porter, B. D., Hefferan, T. E., Larson. D. R., Currier, B. L., Mikos, A. G., and Yaszemski, M. J. (2000), Biological activity of rhBMP-2 released from PLGA microspheres, *J. Biomech. Eng.*, 122, 289–292.
22. Woo, B. H., Fink, B. F., Page, R., Schrier, J. A., Jo, Y. W., Jiang, G., DeLuca, M. Vasconez, H. C., and DeLuca, P. P. (2001), Enhancement of bone growth by sustained delivery of recombinant human bone morphogenetic protein-2 in a polymeric matrix, *Pharm. Res.*, 18, 1747–1753.
23. Weber, F. E., Eyrich, G., Gratz, K. W., Maly, F. E., and Sailer, H. F. (2002), Slow and continuous application of human recombinant bone morphogenetic protein via biodegradable poly(lactide-*co*-glycolide) foamspheres, *Int. J. of Oral Maxillofacial Surg.*, 31, 60–65.
24. Bordem, M. D., Khan, Y., Attawia, M., and Laurencin, C. T. (2001), Tissue engineered microsphere-based matrices for bone repair: Design, evaluation and optimisation, *Biomaterials*, 23, 551–559.
25. Lu, L., Stamatas, G. N., and Mikos, A. G. (2000), Controlled release of transforming growth factor beta 1 from biodegradable polymer microparticles, *J. Biomed. Mater. Res.*, 50, 440–451.
26. Fu, K., Griebenow, K., Hsieh, L., Klibanov, A. M., and Langer, R. (1999), FTIR characterization of the secondary structure of proteins encapsulated within PLGA microspheres, *J. Controlled Release*, 58, 357–366.
27. Bezemer, J. M., Radersma, R., Grijpma, D. W., Dijkstra, P. J., van Blitterswijk, C. A., and Feijen, J. (2000), Microspheres for protein delivery prepared from amphiphilic multiblock copolymers. I. Influence of preparation techniques on particle characteristics and protein delivery, *J. Controlled Release*, 67, 233–248.
28. Rosas, J. E., Pedraz, J. L., Hernandez, R. M., Gascon, A. R., Igartua, M., Guzman, F., Rodrignuez, R., Cortes, J., and Patarroyo, M. E. (2002), Remarkably high antibody levels and protection against *P. falciparum* malaria in Aotus monkey after a single immunization of SPf66 encapsulated in PLGA microspheres, *Vaccine*, 20, 1707–1710.
29. Boehm, G., Peyre, M., Sesardie, D., Huskisson, R. J., Mawas, F., Douglas, A., Xing, D., Merkle, H. P., Gander, B., and Johansen, P. (2002), On technological and immunological benefits of multivalent single-injection microspheres vaccines. *Pharm. Res.*, 19, 1330–1336.
30. Sturesson, C., Artursson, P., Ghaderi, R., Johansen, K., Mirazimi, A., Uhnoo, I., Svensson, L., Alberstsson, A. C., and Carlfors, J. (1999), Encapsulation of rotavirus into poly(lactide-*co*-glycolide) microspheres, *J. Controlled Release*, 59, 377–389.

31. Morlock, M., Kissel, T., Li, Y. X., Koll, H., and Winter, G. (1998), Erythropoietin loaded microspheres prepared from biodegradable LPLG-PEO-LPLG triblock copolymers: Protein stabilization and in-vitro release properties, *J. Controlled Release*, 56, 105–115.
32. Zambaux, M. F., Bonneaux, F., Gref, R., Dellacherie, E., and Vigneron, C. (1999), Preparation and characterization of protein C loaded PLA nanoparticles, *J. Controlled Release*, 60, 179–188.
33. Cho, S. W., Song, S. H., and Choi, Y. W. (2000), Effects of solvent selection and fabrication on the characteristics of biodegradable poly(lactide-*co*-glycolide) microspheres containing ovalbumin, *Arch. Pharm. Res.*, 23, 385–390.
34. Li, X., Deng, X., and Huang, Z. (2001), In vitro protein release and degradation of poly-DL-lactide-poly(ethylene glycol) microspheres with entrapped human serum albumin: Quantitative evaluation of the factors involved in protein release phase, *Pharm. Res.*, 18, 117–124.
35. Carrasquillo, K. G., Stanley, A. M., Aponte-Carro, J. C., DeJesus, P., Costantino, H. R., Bosques, C. J., and Griebenow, K. (2001), Non-aqueous encapsulation of excipient stabilized spray freeze dried BSA into poly(lactide-*co*-glycolide) microspheres results in release of native protein, *J. Controlled Release*, 76, 199–208.
36. Lee, V. H. L. (1986), Peptide and protein drug delivery: Opportunities and challenges, *Pharm. Int.*, 7, 208–212.
37. Johnson, O. L., Cleland, J. L., Lee, H. J., Charnis, M., Duenas, E., Jaworowicz, W., Shepard, D., Shihzamani, A., Jones, A. J. S., and Putney, S. D. (1996), A month-long effect from a single injection of microencapsulated human growth hormone, *Nat. Med.*, 2, 795–799.
38. Costantino, H. R., Firouzabadian, L., Hogeland, K., Wu, C., Beganski, C., Carrasquillo, K. G., Cordova, M., Griebenow, K., Zale, S. E., and Tracy, M. A. (2000), Protein spray-freeze drying: Effect of atomization conditions on particle size and stability, *Pharm. Res.*, 17, 1374–1383.
39. Bustami, R. T., Chan, H. K., Dehghani, F., and Foster, N. R. (2000), Generation of microparticles of proteins for aerosol delivery using high pressure modified carbon dioxide, *Pharm. Res.*, 17, 1360–1366.
40. Alder, M., Unger, M., and Lee, G. (2000), Surface composition of spray-dried particles of bovine serum albumin/trehalose/surfactant, *Pharm. Res.*, 17, 863–870.
41. Ruiz, J. M., Tissier, B., and Benoit, J. P. (1989), Microencapsulation of peptides: A study of the phase separation of poly(D,L-lactic acid-*co*-glycolic acid) copolymers 50/50 by silicone oil, *Int. J. Pharm.*, 49, 69–77.
42. Redding, T. W., Schally, A. V., Tice, T. R., and Meyers, W. E. (1984), Long acting delivery systems for peptides: Inhibition of rat prostate tumors by controlled release of DTrp6 leutinizing hormone releasing hormone injectable microcapsules, *Proc. Nat. Acad. Sci. USA*, 81, 5845.
43. Heron, I., Thomas, F., Dero, M., Poutrain, J. R., Henane, S., Catus, F., and Kuhn, J. M. (1993), Traitement de l'acromegalie par une forme a liberation prolongee du lanreotide, un nouvel analogue de la somatostatine, *Presse Med.*, 22, 526–531.
44. O'Donnell, P. B., and McGinity, J. W. (1997), Preparation of microspheres by the solvent evaporation technique, *Adv. Drug Deliv. Rev.*, 28, 25–42.
45. Cleland, J. L. (1998), Solvent evaporation processes for the production of controlled release biodegradable microsphere formulations for therapeutics and vaccines, *Biotechnol. Prog.*, 14, 102–107.
46. Ogawa, Y., Yamamoto, M., Takada, S., and Shimamoto, T. (1988), Controlled release of leuprolide acetate from polylactic acid or copolymer ratio of polymer, *Chem. Pharm. Bull.*, 36, 1502–1507.

47. Ogawa, Y., Yamamoto, M., Okada, H., Yashiki, T., and Shimamoto, T. (1988), A new technique to efficiently entrap leuprolide acetate into microcapsules of polylactic acid or copoly(lactic/glycolic) acid, *Chem. Pharm. Bull.*, 36, 1095–1103.
48. Bittner, B., Morlock, M., Koll, H., Winter, G., and Kissel, T. (1998), Recombinant human erythropoietin (rHEPO) loaded poly(lactide-*co*-glycolide) microspheres: Influence of the encapsulation technique and polymer purity on microspheres characteristics, *Eur. J. Pharm. Biopharm.*, 45, 295–305.
49. Kissel, T., Brich, Z., Bantle, S., Lancranjam, I., NimmerFall, F., and Vit, P. (1991), Parentral depot systems as the basis of biodegradable polyesters, *J. Controlled Release*, 6, 27–34.
50. Herrmann, J., and Bodmeier, R. (1998), Biodegradable, somatostatin acetate containing microspheres prepared by various aqueous and non-aqueous solvent evaporation methods. *Eur. J. Pharm. Biopharm.*, 45, 75–82.
51. Singh, M., Shirley, B., Bajwa, K., Samara, E., Hora, M., and O'Hagan, D. (2001), Controlled release of recombinant insulin-like growth factor from a novel formulation of polylactide-coglycolide microparticles, *J. Controlled Release*, 70, 21–28.
52. Viswanathan, N. B., Thomas, P. A., Pandit, J. K., Kulkarni, M. G., and Mashelkar, R. A. (1999), Preparation of non-porous microspheres with high entrapment efficiency of proteins by a (water-in-oil)-in-oil emulsion technique, *J. Controlled Release*, 58, 9–20.
53. Sanchez, A., Gupta, R. K., Alonso, M. J., Siber, G. R., and Langer, R. (1996), Pulsed controlled-release system for potential use in vaccine delivery, *J. Pharm. Sci.*, 85, 547–552.
54. Freitas, S., Walz, A., Merkle, H. P., and Gander, B. (2003), Solvent extraction employing a static micromixer: A simple, robust and versatile technology for the microencapsulation of proteins, *J. Microencapsul.*, 20, 67–85.
55. Khan, M. A., Healy, M. S., and Bernstein, H. (1992), Low temperature fabrication of protein loaded microspheres, *Proc. Int. Symp. Controlled Release Bioactive Mater.*, 19, 518–519.
56. Herberger, J. D., Wu, C., Dong, N., and Tracy, M. A. (1996), Characterization of Prolease® human growth hormone PLGA microspheres produced using different solvents, *Proc. Int. Symp. Controlled Release Bioactive Mater.*, 23, 835–836.
57. Johnson, O. L., Jaworowicz, W., Cleland, J. L., Bailey, L., Charnis, M., Duenas, E., Wu, C., Shepard, D., Magil, S., Last, T., Jones, A. J. S., and Putney, S. D. (1997), The stabilization and encapsulation of human growth hormone into biodegradable microspheres, *Pharm. Res.*, 14, 730–735.
58. Gombotz, W. R., Healy, M., Brown, L. R., and Auer, H. E. (1990), Process for producing small particles of biologically active molecules, Pabst Patrea, Patent No. WO 90/13285.
59. Gombotz, W. R., Healy, M. S., and Brown, L. R. (1989), Very low temperature casting of controlled release microspheres, Enzytech Inc., Application No. 89-346143, U.S. Patent 5,019,400.
60. Costantino, H. R., Johnson, O. L., and Zale, S. F. (2004), Relationship between encapsulated drug particle size and initial release of recombinant human growth hormone from biodegradable microspheres, *J. Pharm. Sci.*, 93, 2624–2634.
61. Cleland, J. L., and Jones, A. J. S. (1996), Stable formulations of recombinant human growth hormone and interferon-g for microencapsulation in biodegradable microspheres, *Pharm. Res.*, 13, 1464–1475.
62. Cleland, J. L., Duenas, E. T., Park, A., Daugherty, A., Kahn, J., Kowalski, J., and Cuthbertson, A. (2000), Development of poly-(D,L-lactide-*co*-glycolide) microsphere formulations containing recombinant human vascular growth factor to promote local angiogenesis, *J. Controlled Release*, 72, 13–24.

63. Lam, X. M., Duenas, E. T., Daugherty, A. L., Levin, N., and Cleland, J. L. (2000), Sustained release of recombinant human insulin-like growth factor-I for treatment of diabetes. *J. Controlled Release*, 67, 281–292.
64. Ribeiro Dos Santos, I., Richard, J., Pech, B., Thies, C., and Benoit, J. P. (2002), Microencapsulation of protein particles within lipids using a novel supercritical fluid process, *Int. J. Pharm.*, 242, 69–78.
65. Debenedetti, P. G., Tom, J. W., Yeo, S., Lim, G. B. (1993), Application of supercritical fluids for the production of sustained delivery devices, *J. Controlled Release*, 24, 27–44.
66. Randolph, T. W., RAndolph, A. D., Mebes, M., and Yeung, S. (1993), Submicrometer-sized biodegradable particles of poly(L-lactic acid) via the gas antisolvent spray precipitation process, *Biotechnol. Prog.*, 9, 429–435.
67. Waßnus, W., Bleich, J., and Müller, B. W. (1991), Microparticle production by using supercritical gases, *Clin. Pharmacol.*, 13, 367.
68. Bleich, J., Mueller, B. W., and Wassmus, W. (1993), Aerosol solvent extraction system. A new microparticle production technique, *Int. J. Pharm.*, 97, 111–117.
69. Winters, M. A., Debenedetti, P. G., Carey, J., Sparks, H. G., Sane, S. U., and Przybycien, T. M. (1997), Long-term and high-temperature storage of supercritically-processed microparticulate protein powders, *Pharm. Res.*, 14, 1370–1378.
70. Bodmeier, R., Wang, H., Dixon, D. J., Mawson, S., and Johnston, K. P. (1995), Polymeric microspheres prepared by spraying into compressed carbon dioxide, *Pharm. Res.*, 12, 1211–1217.
71. Witschi, C., and Doelker, E. (1998), Peptide degradation during preparation and in vitro release testing of poly(L-lactic acid) and poly(DL-lactic-*co*-glycolic acid) microparticles, *Int. J. Pharm.*, 171, 1–18.
72. Morlock, M., Koll, H., Winter, G., and Kissel, T. (1997), Microencapsulation of rh-erythropoietin using biodegradable poly(D,L-lactide-*co*-glycolide): Protein stability and the effects of stabilizing excipients, *Eur. J. Pharm. Biopharm.*, 43, 29–36.
73. Bittner, B., and Kissel, T. (1999), Ultrasonic atomization for spray drying: A versatile technique for the preparation of protein loaded bio-degradable microspheres, *J. Microencapsul.*, 16, 325–341.
74. Breitenbach, A., Mohr, D., Kissel, T. (2000), Biodegradable semicrystalline comb polymers influence the microsphere production by means of a supercritical fluid extraction technique, *J. Controlled Release*, 63, 53–68.
75. Engwicht, A., Girreser, U., and Muller, B. W. (1999), Critical properties of lactide-*co*-glycolide polymers for the use of microparticle preparation by the aerosol solvent extraction system, *Int. J. Pharm.*, 185, 61–72.
76. Felder, C. B., Blanco-Príeto, M. J., Heizmann, J., Merkle, H. B., and Gander, B. (2003), Ultrasonic atomization and subsequent polymer desolvation for peptide and protein microencapsulation into biodegradable polyesters, *J. Microencapsul.*, 20, 553–567.
77. Freitas, S., Merkle, H. P., and Gander, B. (2004), Ultrasonic atomisation into reduced pressure atmosphere—envisaging aseptic spray drying for microencapsulation, *J. Controlled Release*, 95, 185–195.
78. Berkland, C., Kim, K., and Pack, D. W. (2001), Fabrication of PLG microspheres with precisely controlled and monodispersed size distributions, *J. Controlled Release*, 73, 59–74.
79. Jain, R. A., Rhodes, C. T., Railkar, A. M., Malik, A. W., and Shah, N. H. (2000), Controlled release of drugs from injectable in situ formed biodegradable PLGA microspheres: Effect of various formulation variables, *Eur. J. Pharm. Biopharm.*, 50, 257–262.

80. Kim, H. K., Chung, H. J., and Park, T. G. (2006), Biodegradable polymeric microspheres with "open/closed" pores for sustained release of human growth hormone, *J. Controlled Release*, 112, 167–174.

81. Singh, M., Kazzaz, J., Ugozzoli, M., Chesko, J., and O'Hagan, D. (2004), Charged poly(lactide-*co*-glycolide) microparticles as novel antigen delivery systems, *Expert Opin. Biolog. Ther.*, 4, 483–491.

82. Jilek, S., Merkle, H. P., and Walter, E. (2005), DNA-loaded biodegradable microparticles as vaccine delivery systems and their interaction with dendritic cells, *Adv. Drug Deliv. Rev.*, 57, 377–390.

83. Jabbal-Gill, I., Lin, W., Jenkins, P., Watts, P., Jimenez, M., Illum, L., Davis, S. S., Wood, J. M., Major, D., Minor, P. D., Li, X. W., Lavelle, E. C., and Coombes, A. G. A. (1999), Potential of polymeric lamellar substrate particles (PLSP) as adjuvants for vaccines, *Vaccine*, 18, 238–250.

84. Jabbal-Gill, I., Lin, W., Kistner, O., Davis, S. S., and Illum, L. (2001), Polymeric lamellar substrate particles for intranasal vaccination, *Adv. Drug Deliv. Rev.*, 51, 97–111.

85. Venkataprasad, N., Coombes, A. G. A., Singh, M., Rohde, M., Wilkinson, K., Hudecz, F., Davis, S. S., and Vordermeier, H. M. (1999), Induction of cellular immunity to a mycobacterial antigen adsorbed on lamellar particles of lactide polymers, *Vaccine*, 17, 1814–1819.

86. Kazzaz, J., Neidleman, J., Singh, M., Ott, G., and O'Hagan, D. T. (2000), Novel anionic microparticles are a potent adjuvant for the induction of cytotoxic T lymphocytes against recombinant p55 gag from HIV-1, *J. Controlled Release*, 67, 347–356.

87. Otten, G., Schaefer, M., Greer, C., Calderon-Cacia, M., Coit, D., Kazzaz, J., Medina-Selby, A., Selby, M., Singh, M., Ugozzoli, M., zur Megede, J., Barnett, S. W., O'Hagan, D., Donnelly, J., and Ulmer, J. (2003), Induction of broad and potent antihuman immunodeficiency virus immune responses in rhesus macaques by priming with a DNA vaccine and boosting with protein-adsorbed polylactide coglycolide microparticles, *J. Virol.*, 77, 6087–6092.

88. Kang, F., Jiang, G., Hinderliter, A., DeLuca, P. P., and Singh, J. (2002), Lysozyme stability in primary emulsion for PLGA microsphere preparation: Effect of recovery methods and stabilizing excipients, *Pharm. Res.*, 19, 629–633.

89. Sanchez, A., Villamayor, B., Guo, Y., McIver, J., and Alonso, M. J. (1999), Formulation strategies for the stabilization of tetanus toxoid in poly(lactide-*co*-glycolide) microspheres, *Int. J. Pharm.*, 185, 255–266.

90. Sharif, S., and O'Hagan, D. T. (1995), A comparison of alternative methods for determination of the levels of proteins entrapped in poly(lactide-*co*-glycolide) microparticles, *Int. J. Pharm.*, 115, 259–263.

91. Gupta, R. K., Chang, A. C., Griffin, P., Rivera, R., Guo, Y. Y., and Siber, G. R. (1997), Determination of protein loading in biodegradable polymer microspheres containing tetanus toxoid, *Vaccine*, 15, 672–678.

92. Lu, W., and Park, T. G. (1995), Protein release from poly(lactic-*co*-glycolic acid) microspheres: Protein stability problems, *PDA J. Pharm. Sci. Technol.*, 49, 13–19.

93. Park, T. G., Lu, W., and Crotts, G. (1995), Importance of in vitro experimental conditions on protein release kinetics, stability and polymer degradation in protein encapsulated poly(D,L-lactic acid-*co*-glycolic acid) microspheres, *J. Controlled Release*, 33, 211–222.

94. Bilati, U., Pasquarello, C., Corthals, C. L., Hochstrasser, D. F., Allémann, E., and Doelker, E. (2005), Matrix-assisted laser desorption/ionization time-of-flight mass spectrometry for quantitation and molecular stability assessment of insulin entrapped within PLGA nanoparticles, *J. Pharm. Sci.*, 94, 1–7.

95. van de Weert, M., Van't Hof, R., Van der Weerd, J., Heeren, R. M., Posthuma, G., Hennink, W.E., and Crommelin, D. J. (2000), Lysozyme distribution and conformation in a biodegradable polymer matrix as determined by FTIR techniques, *J. Controlled Release*, 68, 31–40.
96. Blanco, M. D., and Alonso, M. J. (1997), Development and characterization of protein-loaded poly(lactide-*co*-glycolide) nanospheres, *Eur. J. Pharm. Biopharm.*, 43, 287–294.
97. Johansen, P., Tamber, H., Merkle, H. P., and Gander, B. (1999), Diphtheria and tetanus toxoid microencapsulation into conventional and end-group alkylated PLA/PLGAs, *Eur. J. Pharm. Biopharm.*, 47, 193–201.
98. van de Weert, M., Hoechstetter, J., Hennink, W. E., and Crommelin, D. J. (2000), The effect of a water/organic solvent interface on the structural stability of lysozyme, *J. Controlled Release*, 68, 351–359.
99. Yang, T. H., Dong, A., Meyer, J., Johnson, O. L., Cleland, J. L., and Carpenter, J. F. (1999), Use of infrared spectroscopy to assess secondary structure of human growth hormone within biodegradable microspheres, *J. Pharm. Sci.*, 88, 161–165.
100. Perez, C., De Jesus, P., and Griebenow, K. (2002), Preservation of lysozyme structure and function upon encapsulation and release from poly(lactic-*co*-glycolic) acid microspheres prepared by the water-in-oil-in-water method, *Int. J. Pharm.*, 248, 193–206.
101. Jorgensen, L., Vermehren, C., Bjerregaard, S., and Froekjaer, S. (2003), Secondary structure alterations in insulin and growth hormone water-in-oil emulsions, *Int. J. Pharm.*, 254, 7–10.
102. Na, D. H., Youn, Y. S., Lee, S. D., Son, M. W., Kim, W. B., DeLuca, P. P., and Lee, K. C. (2003), Monitoring of peptide acylation inside degrading PLGA microspheres by capillary electrophoresis and MALDI-TOF mass spectrometry, *J. Controlled Release*, 92, 291–299.
103. Gilg, D., Riedl, B., Zier, A., and Zimmermann, M. F. (1996), Analytical methods for the characterization and quality control of pharmaceutical peptides and proteins, *Pharm. Acta Helv.*, 71, 383–394.
104. Brass, J. M., Krummen, K., and Moll-Kaufmann, C. (1996), Quality assurance after process changes of the production of a therapeutic antibody, *Pharm. Acta Helv.*, 71, 395–403.
105. Sesardic, D., and Dobbelaer, R. (2004), European union regulatory developments for new vaccine adjuvants and delivery systems, *Vaccine*, 22, 2452–2456.
106. Burgess, D. J., Crommelin, D. J. A., Hussain, A. S., and Chen, M. L. (2004), Assuring quality and performance of sustained and controlled release parenterals: EUFEPS workshop report, *AAPS PharmSciTech*, 6, 1–12.
107. Burgess, D. J., Hussain, A. S., Ingallinera, T. S., and Chen, M. L. (2002), Assuring quality and performance of sustained and controlled release parenterals: Workshop report, *AAPS PharmSciTech*, 4, (article 7).
108. Gupta, R. K., Chang, A. C., and Siber, G. R. (1998), Biodegradable polymer microspheres as vaccine adjuvants and delivery systems, *Dev. Biol. Stand.*, 92, 63–78.
109. Visscher, G. E., Robison, R. L., Maulding, H. V., Fong, J. W., Pearson, J. E., and Argentieri, G. J. (1985), Biodegradation of and tissue reaction to 50:50 poly(DL-lactide-*co*-glycolide) microcapsules, *J. Biomed. Mater. Res.*, 19, 349–365.
110. Schakenraad, J. M., Hardonk, M. J., Feijen, J., Molenaar, I., and Nieuwenhuis, P. (1990), Enzymatic activity toward poly(L-lactic) acid implants, *J. Biomed. Mater. Res.*, 24, 529–545.
111. Johansen, P., Men, Y., Merkle, H. P., and Gander, B. (2000), Revisiting PLG/PLGA microspheres: An analysis of their potential in parenteral vaccination, *Eur. J. Pharm. Biopharm.*, 50, 129–146.

112. Men, Y., Audran, R., Thomasin, C., Eberl, G., Demotz, S., Merkle, H. P., Gander, B., and Corradin, G. (1999), MHC class I-and class II-restricted processing and presentation of microencapsulated antigens, *Vaccine*, 17, 1047–1056.

113. Raychaudhuri, S., and Rock, K. L. (1998), Fully mobilizing host defense: Building better vaccines, *Nat. Biotechnol.*, 16, 1025–1031.

114. Visscher, G. E., Robinson, R. L., and Argentieri, G. J. (1987), Tissue response t biodegradable injectable microcapsules, *J. Biomater. Applic.*, 2, 118–131.

115. Cleland, J. L., Powell, M. F., and Shire, S. J. (1993), The development of stable protein formulations: A close look at protein aggregation, deamidation, and oxidation, *Crit. Rev. Therap. Drug Carrier Syst.*, 10, 307–377.

116. Gupta, R. K., Alroy, J., Alonso, M. J., Langer, R., and Siber, G. R. (1997), Chronic tissue reactions, long term immunogenicity and immunological priming of mice and guinea pigs to tetanus toxoids encapsulated in biodegradable polymer microspheres composed of polylactide-*co*-glycolide polymers, *Vaccine*, 15, 1716–1723.

117. Jiang, W., Gupta, R. K., Deshpande, M. C., and Schwendeman, S. P. (2005), Biodegradable poly(lactic-*co*-glycolic acid) microparticles for injectable delivery of vaccine antigens, *Adv. Drug Deliv. Rev.*, 57, 391–410.

118. Johnson, O. L., and Tracy, M. A. (1999), Peptide and protein drug delivery, in Mathiowitz, E., Ed., *Encyclopedia of Controlled Drug Delivery*, Vol. 2, Wiley, New York, pp. 816–833.

119. Braun, A., and Alsenz, J. (1997), Development and use of enzyme-linked immunosorbent assays (ELISA) for the detection of protein aggregates in interferon-alpha (IFN-alpha) formulations, *Pharm. Res.*, 14, 1394–1400.

120. Hochuli, E. (1997), Interferon immunogenicity: Technical evaluation of interferon-alpha 2a, *J. Interferon Cytokine Res.*, 17 (Suppl 1), 15–21.

121. Volkin, D. B., and Middaugh, C. R. (1992), in: Ahern, T. J., Manning, M. C., Eds., *Stability of Protein Pharmaceuticals Part A: Chemical and Physical Pathways of Protein Degradation*, Plenum, New York, pp. 109–134.

122. Manning, M. C., Patel, K., and Borchardt, R. T. (1989), Stability of protein pharmaceuticals, *Pharm. Res.*, 6, 903–918.

123. Zale, S. E., and Klibanov, A. M. (1986), Why does ribonuclease irreversibly inactivate at high temperatures? *Biochemistry*, 25, 5432–5443.

124. Jiang, W., and Schwendeman, S. P. (2000), Formaldehyde-mediated aggregation of protein antigens: Comparison of untreated and formalinized model antigens, *Biotechnol. Bioeng.*, 70, 507–517.

125. Jiang, W., and Schwendeman, S. P. (2001), Stabilization of a model formalinized protein antigen encapsulated in poly(lactide-*co*-glycolide)-based microspheres, *J. Pharm. Sci.*, 90, 1558–1569.

126. Zhu, G., Mallery, S. R., and Schwendeman, S. P. (2000), Stabilization of proteins encapsulated in injectable poly(lactide-*co*-glycolide), *Nat. Biotechnol.*, 18, 52–57.

127. Schwendeman, S. P., Cardamone, M., Klibanov, A., and Langer, R. (1996), Stability of proteins and their delivery from biodegradable polymer microspheres, in Cohen, S., and Bernstein, H., Eds., *Microparticulates Systems for the Delivery of Proteins and Vaccines*, Marcel Dekker, New York, pp. 1–49.

128. van de Weert, M., Hennink, W. E., and Jiskoot, W. (2000), Protein instability in poly(lactic-*co*-glycolic acid) microparticles, *Pharm. Res.*, 17, 1159–1167.

129. Schwendeman, S. P. (2002), Recent advances in the stabilization of proteins encapsulated in injectable PLGA delivery systems, *Crit. Rev. Therap. Drug Carrier Syst.*, 19, 73–98.

130. Bilati, U., Allémann, E., and Doelker, E. (2005), Strategic approaches for overcoming peptide and protein instability within biodegradable nano- and microparticles, *Eur. J. Pharm. Biopharm.*, 59, 375–388.

131. Sanders, L., and Hendren. W. (1998), *Protein Delivery: Physical Systems*, Plenum, New York.

132. Senior, J., and Radomsky, M. (2000), *Sustained-Release Injectable Products*, Interpharm, Denver.

133. Akers, M. J. (2002), Excipient-drug interactions in parenteral formulations, *J. Pharm. Sci.*, 91, 2283–2300.

134. Stevenson, C. L. (2000), Characterization of protein and peptide stability and solubility in non-aqueous solvents, *Curr. Pharm. Biotechnol.*, 1, 165–182.

135. Chin, J. T., Wheeler, S. L., and Klibanov, A. M. (1994), On protein solubility in organic solvents, *Biotechnol. Bioeng.*, 44, 140–145.

136. Park, T. G., Lee, H. Y., and Nam, Y. S. (1998), A new preparation method for protein loaded poly(D,L-lactic-*co*-glycolic acid) microspheres and protein release mechanism study, *J. Controlled Release*, 55, 181–191.

137. Yoo, H. S., Choi, H. K., and Park, T.G. (2001), Protein–fatty acid complex for enhanced loading and stability within biodegradable nanoparticles, *J. Pharm. Sci.*, 90, 194–201.

138. Quintanar-Guerrero, D., Allémann, E., Fessi, H., and Doelker, E. (1997), Applications of the ion-pair concept to hydrophilic substances with special emphasis on peptides, *Pharm. Res.*, 14, 119–127.

139. Kim, H. K., and Park, T. G. (2001), Microencapsulation of dissociable human growth hormone aggregates within poly(D,L-lactic-*co*-glycolic acid) microparticles for sustained release, *Int. J. Pharm.*, 229, 107–116.

140. Johnson, M. D., Hoesterey, B. L., and Anderson, B. D. (1995), Solubilization of a tripeptide HIV protease inhibitor using a combination of ionization and complexation with chemically modified cyclodextrins, *J. Pharm. Sci.*, 83, 1142–1146.

141. Brewster, M. E., Simpkins, J. W., Hora, M. S., Stern, W. C., and Bodor, N. (1989), The potential use of cyclodextrins in parenteral formulations, *J. Parenteral Sci. Technol.*, 43, 231–240.

142. Medlicott, N. J., Foster, K. A., Audus, K. L., Gupta, S., and Stella, V. J. (1998), Comparison of the effects of potential parenteral vehicles for poorly water soluble anticancer drugs (organic cosolvents and cyclodextrin solutions) on cultured endothelial cells (HUV-EC), *J. Pharm. Sci.*, 87, 1138–1143.

143. Matsuyama, K., El-Gizway, S., and Perrin, J. H. (1987), Thermodynamics of binding of aromatic amino acids to α, β, and γ cyclodextrins, *Drug Dev. Ind. Pharm.*, 13, 2687–2691.

144. Meinel, L., Illi, O. E., Zapf, J., Malfanti, M., Peter, M. H., and Gander, B. (2001), Stabilizing insulin-like growth factor-I in poly(D,L-lactide-*co*-glycolide) microspheres, *J. Controlled Release*, 70, 193–202.

145. Uchida, T., Shiosaki, K., Nakada, Y., Fukada, K., Eda, Y., Tokiyoshi, S., Nagareya, N., and Matsuyama, K. (1998), Microencapsulation of hepatitis B core antigen for vaccine preparation, *Pharm. Res.*, 15, 1708–1713.

146. Youan, B. B. C., Gillard, J., and Rollmann, B. (1999), Protein-loaded poly(ε-caprolactone) microparticles. III. Entrapment of superoxide dismutase by the (water-in-oil)-in water solvent evaporation method, *STP Pharma Sci.*, 9, 175–181.

147. Johansen, P., Men, Y., Audran, R., Corradin, G., Merkle, H. P., and Gander, B. (1998), Improving stability and release kinetics of microencapsulated tetanus toxoid by co-encapsulation of additives, *Pharm. Res.*, 15, 1103–1110.

148. Krielgaard, L., Jones, L. S., Randolph, T. W., Frokjaer, S., Flink, J. M., Manning, M. C., Carpenter, J. F. (1998), Effect of Tween 20 on freeze-thawing- and agitation-induced aggregation of recombinant human factor XIII, *J. Pharm Sci*., 87, 1593–1603.
149. Krielgaard, L., Frokjaer, S., Flink, J .M., Randolph, T. W., and Carpenter, J. F. (1999), Effects of additives on the stability of Humicola lanuginosa lipase during freeze-drying and storage in the dried solid, *J. Pharm. Sci*., 88, 281–290.
150. De Rosa, G., Iommelli, R., La Rotonda, M. I., Miro, A., and Quaglia, F. (2000), Influence of the co-encapsulation of different non-ionic surfactants on the properties of PLGA insulin-loaded microspheres, *J. Controlled Release*, 69, 283–295.
151. Chang, B. S., Kendrick, B. S., and Carpenter, J. F. (1996), Surface-induced denaturation of proteins during freezing and its inhibition by surfactants, *J. Pharm. Sci*., 85, 1325–1330.
152. Kerwin, B. A., Heller, M. C., Levin, S. H., and Randolph, T. W. (1998), Effects of Tween 80 and sucrose on the acute short term stability and long term stability storage at –20 degrees of a recombinant hemoglobin, *J. Pharm. Sci*., 87, 1062–1068.
153. Knepp, V. M., Whatley, J. L., Muchnik, A., and Calderwood, T. S. (1996), Identification of antioxidants for prevention of peroxide-mediated oxidation of recombinant human ciliary neurotrophic factor and recombinant human nerve growth factor, *J. Parenteral Sci. Technol*., 50, 163–171.
154. Nace, V. (1996), *Nonionic Surfactants: Polyoxyalkylene Block Copolymers*, Marcel Dekker, New York.
155. Kabanov, A., Batrakova, E., and Alakhov, V. (2002), Pluronic block copolymers as novel polymer therapeutics for drug and gene delivery, *J. Controlled Release*, 82, 189–212.
156. U.S. Food and Drug Administration (2003, Nov. 7), Inactive ingredient search for approved drug products, "poloxamer," available http:/www.accessdata.fda.gov/scripts/cder/iig/index.cfm, accessed May 21, 2006.
157. Johnston, T. P., and Miller, S. C. (1985), Toxicological evaluation of poloxamer vehicles for intramuscular use, *J. Parenteral Sci. Technol*., 39, 83–88.
158. Sanchez, A., Tobio, M., Gonzalez, L., Fabra, A., and Alonso, M. J. (2003), Biodegradable micro- and nanoparticles as long-term delivery vehicles for interferon-α, *Eur. J. Pharm. Sci*., 18, 221–229.
159. Lam, X. M., Duenas, E. T., and Cleland, J. L. (2001), Encapsulation and stabilization of nerve growth factor into poly(lactic-co-glycolic) acid microspheres, *J. Pharm. Sci*., 90, 1356–1365.
160. Nihant, N., Schugens, C., Grandfils, C., Jerôme, R., and Teyssie, P. (1994), Polylactide microparticles prepared by double emulsion/evaporation technique: I. Effect of primary emulsion stability, *Pharm. Res*., 11, 1479–1484.
161. Sturesson, C., and Carlfors, J. (2000), Incorporation of protein in PLG-microspheres with retention of bioactivity, *J. Controlled Release*, 67, 171–178.
162. Chen, L., Apte, R. N., and Cohen, S. (1997), Characterization of PLGA microspheres for the controlled delivery of IL-1a for tumor immunotherapy, *J. Controlled Release*, 43, 261–272.
163. Kwon, Y. M., Baudys, M., Knutson, K., and Kim, S. W. (2001), In situ study of insulin aggregation induced by water–organic solvent interface, *Pharm. Res*., 18, 1754–1759.
164. Costantino, H. R., Langer, R., and Klibanov, A. (1995), Aggregation of a lyophilized pharmaceutical protein, recombinant human albumin: effect of moisture and stabilization excipients, *BioTechnology*, 13, 493–496.
165. Costantino, H. R., Langer, R., and Klibanov, A. M. (1994), Solid-phase aggregation of proteins under pharmaceutically relevant conditions, *J. Pharm. Sci*., 83, 1662–1669.

166. Means, G. E., and Feeney, R. E. (1971), *Chemical Modification of Proteins*, Holden-Day, San Francisco.
167. Schwendeman, S. P., Costantino, H. R., Gupta, R. K., Siber, G. R., Klibanov, A. M., and Langer, R. (1995), Stabilization of tetanus and diphtheria toxoids against moisture-induced aggregation, *Proc. Nat. Acad. Sci. USA*, 92, 11234–11238.
168. Costantino, H. R., Schwendeman, S. P., Griebenow, K., Klibanov, A. M., and Langer, R. (1996), The secondary structure and aggregation of lyophilized tetanus toxoid, *J Pharm. Sci.*, 85, 1290–1293.
169. Hageman, M. J. (1988), The role of moisture in protein stability, *Drug Dev. Ind. Pharm.*, 14, 2047–2070.
170. Lee, H. J., Riley, G., Johnson, O., Cleland, J. L., Kim, N., and Charnis, M. (1997), In vivo characterization of sustained release formulations of human growth hormone, *J. Pharmacol. Exp. Ther.*, 281, 1431–1439.
171. Herbert, P., Murphy, K., Johnson, O., Dong, N., Jaworowicz, W., and Tracy, M. A. (1998), A large-scale process to produce microencapsulated proteins, *Pharm. Res.*, 15, 357–361.
172. Tracy, M. A., Bernstein, H., and Kahn, M. A. (2000), Controlled release of metal cation-stabilized interferon, U.S. Patent 6,165,508.
173. Lam, X. M., Duenas, E. T., and Cleland, J. L. (1998), Stabilization of nerve growth factor during microencapsulation and release from microspheres [abstract], *Proc. Int. Symp. Controlled Release Bioactive Mater.*, 25, 491.
174. Zale, S. E., Burke, P. A., Bernstein, H., and Brickner, A. (1997), Composition for sustained release of non-aggregated erythropoietin, U.S. Patent 5,674,534.
175. Johnson, O. L. (2000), Stabilization of proteins in solutions by lyophilization, in Wise, D. E., *Handbook of Pharmaceutical Controlled Release Technology*, Marcel Dekker, New York, pp. 693–749.
176. Carpenter, J. F., and Chang, B. S. (1996), Lyophilization of protein pharmaceuticals, in Avis, K. E., and Wu, V. L., Eds., *Biotechnology and Biopharmaceutical Manufacturing, Processing, and Preservation*, Interpharm, Buffalo Grove, IL, pp. 199–264.
177. Carpenter, J. F., and Crowe, J. H. (1988), The mechanism of cryoprotection of proteins by solutes, *Cryobiology*, 25, 244–255.
178. Arakawa, T., Prestrelski, S., Kinney, W., and Carpenter, J. F. (1993), Factors affecting short-term and long-term stabilities of proteins, *Adv. Drug Deliv. Rew.*, 10, 1–28.
179. Franks, F. (1990), Freeze drying: From empiricism to predictability, *Cryoletters* 11, 93–110.
180. Prestrelski, S. J., Tedeschi, N., Arakawa, T., and Carpenter, J. F. (1993), Dehydration-induced conformational changes in proteins and their inhibition by stabilizers, *Biophys. J.*, 65, 661–671.
181. Tabata, Y., Gutta, S., and Langer, R. (1993), Controlled delivery systems for proteins using polyanhydride microspheres, *Pharm. Res.*, 10, 487–496.
182. Péan, J. M., Boury, F., Venier-Julienne, M. C., Menei, P., Proust, J. E., and Benoit, J. P. (1999), Why does PEG 400 co-encapsulation improve NGF stability and release from PLGA biodegradable microspheres? *Pharm. Res.*, 16, 1294–1299.
183. Perez-Rodriguez, C., Montano, N., Gonzalez, K., and Griebenow, K. (2003), Stabilization of a-chymotrypsin at the $CH_2Cl_2$/water interface and upon water-in-oil-in-water encapsulation in PLGA microspheres, *J. Controlled Release*, 89, 71–85.
184. Izutsu, K., Yoshioka, S., and Teroa, T. (1993), Decreased protein-stabilizing effects of cryoprotectants due to crystallization, *Pharm. Res.*, 10, 1232–1237.
185. Izutsu, K., Yoshioka, S., and Teroa, T. (1994), Effect of mannitol crystallinity on the stabilization of enzymes during freeze drying, *Chem. Pharm. Bull.*, 42, 5–8.

186. Fakes, M. G., Dali, M. V., Haby, T. A., Morris, K. R., Varia, S. A., and Serajuddin, A. T. M. (2000), Moisture sorption behavior of selected bulking agents used in lyophilized products, *PDA J. Pharm. Sci. Technol.*, 54, 144–149.
187. Kovalcik, T. R., and Guillory, J. K. (1988), The stability of cyclophosphamide in lyophilized cakes. Part I. Mannitol, lactose, and sodium bicarbonate as excipients, *J. Parenterol Sci. Technol.*, 43, 80–83.
188. Costantino, H. R., Carrasquillo, K. G., Cordero, R. A., Mumenthaler, M., Hsu, C. C., and Griebenow, K. (1998), Effect of excipients on the stability and structure of lyophilized recombinant human growth hormone, *J. Pharm. Sci.*, 87, 1412–1420.
189. Li, S., Patapoff, T. W., Overcashier, D., Hsu, C., Nguyen, T. H., and Borchardt, R. T. (1996), Effects of reducing sugars on the chemical stability of human relaxin in the lyophilized state, *J. Pharm. Sci.*, 85, 873–877.
190. Hartauer, K. J., and Guillory, J. K. (1991), A comparison of diffuse reflectance FT-IR spectroscopy and DSC in the characterization of a drug-excipient interaction, *Drug Dev. Ind. Pharm.*, 17, 617–630.
191. Wang, J., Wang, B. M., and Schwendeman, S. P. (2004), Mechanistic evaluation of the glucose-induced reduction in initial burst release of octreotide acetate from poly(D,L-lactide-*co*-glycolide) microspheres, *Biomaterials*, 25, 1919–1927.
192. Dubost, D. C., Kaufman, M. J., Zimmerman, J. A., Bogusky, M. J., Coddington, A. B., and Pitzenberger, S. M. (1996), Characterization of a solid state reaction product from a lyophilized formulation of a cyclic heptapeptide: A novel example of an excipient-induced oxidation, *Pharm. Res.*, 13, 1811–1814.
193. Williams, N. A., Lee, Y., Polli, G. P., and Jennings, T. A. (1986), The effects of cooling rate on solid phase transitions and associated vial breakage occurring in frozen mannitol solutions, *J. Parenteral Sci. Technol.*, 40, 135–141.
194. Williams, N. A., and Dean, T. (1991), Vial breakage by frozen mannitol solutions: Correlation with thermal characteristics and effect of stereoisomerism, additives, and vial configuration, *J. Parenteral Sci. Technol.*, 45, 94–100.
195. Allison, S. D., Manning, M. C., Randolph, T. W., Middleton, K., Davis, A., and Carpenter, J. F. (2000), Optimization of storage stability of lyophilized actin using combinations of disaccharides and dextran, *J. Pharm. Sci.*, 89, 199–214.
196. Cleland, J. L., Lam, X., Kendrick, B., Yang, J., Yang, T., Overcashier, D., Brooks, D., Hsu, C., and Carpenter, J. F. (2001), A specific molar ratio of stabilizer to protein is required for storage stability of a lyophilized monoclonal antibody, *J. Pharm. Sci.*, 90, 310–321.
197. Lam, X. M., Patapoff, T. W., and Nguyen, T. H. (1997), Effect of benzyl alcohol on stability of recombinant human interferon gamma, *Pharm. Res.*, 14, 725–729.
198. Chen, B. L., and Arakawa, T. (1996), Stabilization of recombinant human keratinocyte growth factor by osmolytes and salts, *J. Pharm. Sci.*, 85, 419–422.
199. Arakawa, T., Kita, Y., and Carpenter, J. F. (1991), Protein–solvent interactions in pharmaceutical formulations, *Pharm. Res.*, 8, 285–291.
200. Zhou, Y., and Hall, C. K. (1996), Solute excluded-volume effects on the stability of globular proteins: A statistical thermodynamic theory, *Biopolymers*, 38, 273–284.
201. Sugimoto, L., Ishihara, T., Habata, H., and Nakagawa, H. (1981), Stability of lyophilized sodium prasterone sulfate, *J. Parenteral Sci. Techn.*, 35, 88–92.
202. Shao, P. G., and Bailey, L. C. (2000), Porcine insulin biodegradable polyester microspheres: Stability and in vitro release characteristics, *Pharm. Dev. Technol.*, 5, 1–9.
203. Mäder, K., Bittner, B., Li, Y., Wohlauf, W., and Kissel, T. (1998), Monitoring microviscosity and microacidity of the albumin microenvironment inside degrading microparticles from poly(lactide-coglycolide) (PLG) or ABA-triblock polymers containing hydropho-

bic poly(lactide-*co*-glycolide) A blocks and hydrophilic poly(ethyleneoxide) B blocks, *Pharm. Res*., 15, 787–793.

204. Brunner, A., Mader, K., and Gopferich, A. (1999), pH and osmotic pressure inside biodegradable microspheres during erosion, *Pharm. Res*., 16, 847–853.
205. Burke, P. A. (1996), Determination of internal pH in PLGA microspheres using $^{31}$P NMR spectroscopy, *Proc. Inter. Symp. Controlled Release Bioactive Mater*., 23, 133–134.
206. Fu, K., Pack, D. W., Klibanov, A. M., and Langer, R. (2000), Visual evidence of acidic environment within degrading poly(lactic-*co*-glycolic acid) (PLGA) microspheres, *Pharm. Res*., 17, 100–106.
207. Shenderova, A., Burke, T. G., and Schwendeman, S. P. (1999), The acidic microclimate in poly(lactide-*co*-glycolide) microspheres stabilizes camptothecins, *Pharm. Res*., 16, 241–248.
208. Li, L., and Schwendeman, S. P. (2005), Mapping neutral microclimate pH in PLGA microspheres, *J. Controlled Release*, 101, 163–173.
209. Ding, A. G., and Schwendeman, S. P. (2004), Determination of water-soluble acid distribution in poly(lactide-*co*-glycolide), *J. Pharm. Sci*., 93, 322–331.
210. Cleland, J. L., Mac, A., Boyd, B., Yang, J., Hsu, C., Chu, H., Mukku, V., and Jones, A. J. S. (1997), The stability of recombinant human growth hormone in poly(lactic-*co*-glycolic acid) (PLGA) microspheres, *Pharm. Res*., 14, 420–425.
211. Cleland, J. L., Duenas, E., Daugherty, A., Marian, M., Yang, J., Wilson, M., Celniker, A. C., Shahzamani, A., Quarmby, V., Chu, H., Mukku, V., Mac, A., Roussakis, M., Gillette, N., Boyd, B., Yeung, D., Brooks, D., Maa, Y. F., Hsu, C., and Jones, A. J. S. (1997), Recombinant human growth hormone poly(lactic-*co*-glycolic acid) (PLGA) microspheres provide a long lasting effect, *J. Controlled Release*, 49, 193–205.
212. Schwendeman, S. P. (2001), Stabilization of vaccine antigens encapsulated in PLGA microspheres [abstract 320], *Proc. Int. Symp. Controlled Release Bioactive Mater*., 28.
213. Jiang, W., and Schwendeman, S. P. (2001), Stabilization and controlled release of bovine serum albumin encapsulated in poly(D,L-lactide) and poly(ethylene glycol) microsphere blends, *Pharm. Res*., 18, 878–885.
214. Bittner, B., Witt, C., Mäder, K., and Kissel, T. (1999), Degradation and protein release properties of microspheres prepared from biodegradable poly(lactide-*co*-glycolide) and ABA triblock copolymers: Influence of buffer media on polymer erosion and bovine serum albumin release, *J. Controlled Release*, 60, 297–309.
215. Shenderova, A., Zhu, G., and Schwendeman, S. P. (2000), Correlation of measured microclimate pH with the stability of BSA encapsulated in PLGA microspheres abstract 0413], *Proc. Int. Symp. Controlled Release Bioactive Mater*., 27.
216. Shenderova, A. (2000), The microclimate in poly(lactide-*co*-glycolide) microspheres and its effect on the stability of encapsulated substances, Ph.D thesis, The Ohio State University.
217. Hunter, S. K., andracki, M. E., and Kreig, A. M. (2001), Biodegradable microspheres containing group B Streptococcus vaccine: Immune response in mice, *Am. J. Obstet. Gynecol*., 185, 1174–1179.
218. Johansen, P., Moon, L., Tamber, H., Merkle, H. P., Gander, B., and Sesardic, D. (1999), Immunogenicity of single-dose diphtheria vaccines based on PLA/PLGA microspheres in guinea pigs, *Vaccine*, 18, 209–215.
219. Peyre, M., Sesardic, D., Merkle, H. P., Gander, B., and Johansen, P. (2003), An experimental divalent vaccine based on biodegradable microspheres induces protective immunity against tetanus and diphtheria, *J. Pharm. Sci*., 92, 957–966.

220. Peyre, M., Fleck, R., Hockley, D., Gander, B., and Sesardic, D. (2004), In vivo uptake of an experimental microencapsulated diphtheria vaccine following sub-cutaneous immunisation, *Vaccine*, 22, 2430–2437.
221. Namura, J. A. M., Takata, C. S., Moroc, A. M., Politi, M. J., de Araujo, S., Cuccovia, I. M., and Bueno da Costa, M. H. (2004), Lactic acid triggers, in vitro, thiomersal to degrade protein in the presence of PLGA microspheres, *Int. J. Pharm.*, 273, 1–8.
222. Tan, M., and Parkin, J. E. (2000), Route of decomposition of thiomersal (Thimerosal®), *Int. J. Pharm.*, 208, 23–34.
223. Jung, T., Koneberg, R., Hungerer, K. D., and Kissel, T. (2002), Tetanus toxoid microspheres consisting of biodegradable poly(lactide-*co*-glycolide)- and ABA-triblock-copolymers: Immune response in mice, *Int. J. Pharm.*, 234, 75–90.
224. Determan, A. S., Wilson, J. H., Kipper, M. J., Wannemuehler, M. J., and Narasimhan, B. (2006), Protein stability in the presence of polymer degradation products: Consequences for controlled release formulations, *Biomaterials*, 27, 3312–3320.
225. Jaganathan, K. S., Rao, Y. U. B., Singh, P., Prabakaran, D., Gupta, S., Jain, A., and Vyas, S. P. (2005), Development of a single dose tetanus toxoid formulation based on polymeric microspheres: a comparative study of poly(D,L-lactic-*co*-glycolic acid) versus chitosan microspheres, *Int. J. Pharm.*, 294, 23–32.
226. Kipper, M. J., Wilson, J. H., Wannemuehler, M. J., and Narasimhan, B. (2006), Single dose vaccine based on biodegradable polyanhydride microspheres can modulate immune response mechanism, *J. Biomed. Mater. Res. Part A*, 76, 798–810.
227. Katare, Y. K., Muthukumaran, T., and Panda, A. K. (2005), Influence of particle size, antigen load, dose and additional adjuvant on the immune response from antigen loaded PLA microparticles, *Int. J. Pharm.*, 301, 149–160.
228. Yeh, M. K., Liu, Y. T., Chen, J. L., and Chiang, C. H. (2002), Oral immunogenicity of the inactivated *Vibrio cholerae* whole cell vaccine encapsulated in biodegradable microparticles, *J. Controlled Release*, 82, 237–247.
229. Yeh, M. K., Chen, J. L., and Chiang, C. H. (2002), Vibrio cholerae-loaded poly(DL-lactide-*co*-glycolide) microparticles, *J. Microencapsul.*, 19, 203–212.
230. Yeh, M. K., and Chiang, C. H. (2004), Inactive *Vibrio cholerae* whole-cell vaccine-loaded biodegradable microparticles: In vitro release and oral vaccination, *J. Microencapsul.*, 21, 91–106.
231. Khang, G., Cho, J. C., Lee, J. W., Rhee, J. M., and Lee, H. B. (1999), Preparation and characterization of Japanese encephalitis virus vaccine loaded poly(L-lactide-*co*-glycolide) microspheres for oral immunization, *Biomed. Mater. Eng.*, 9, 49–59.
232. Kaur, R., Rauthan, M., and Vrati, S. (2004), Immunogenicity in mice of a cationic microparticle-adsorbed plasmid DNA encoding Japanese encephalitis virus envelope protein, *Vaccine*, 22, 2776–2782.
233. Yeh, M. K., Coombes, A. G. A., Chen, J. L., and Chiang, C. H. (2002), Japanese encephalitis virus vaccine formulations using PLA lamellar and PLG microparticles, *J. Microencapsul.*, 19, 671–682.
234. Feng, L., Qi, X. R., Zhou, X. J., Maitani, Y., Wang, S. C., Jiang, Y., and Nagai, T. (2006), Pharmaceutical and immunological evaluation of a single-dose hepatitis B vaccine using PLGA microspheres, *J. Controlled Release*, 112, 35–42.
235. Jaganathan, K. S., Singh, P., and Prabakaran, D. (2004), Development of a single-dose stabilized poly(D,L-lactide-*co*-glycolide) microspheres-based vaccine against hepatitis B, *J. Pharm. Pharmacol.*, 56, 1243–1250.
236. Shi, L., Caulfield, M. J., Chern, R. T., Wilson, R. A., Sanyal, G., and Volkin, D. B. (2002), Pharmaceutical and immunological evaluation of a single-shot hepatitis B vaccine formulated with PLGA microspheres, *J. Pharm. Sci.*, 91, 1020–1035.

237. Singh, M., Li, X. M., McGee, J. P., Zamb, T., Koff, W., Wang, C. Y., and O'Hagan, D. T. (1997), Controlled release microparticles as a single dose hepatitis B vaccine: Evaluation of immunogenicity in mice, *Vaccine*, 15, 475–481.
238. Genta, I., Perugini, P., Pavanetto, F., Maculotti, K., Modena, T., Casado, B., Lupi, A., Iadarola, P., and Conti, B. (2001), Enzyme loaded biodegradable microspheres in vitro ex vivo evaluation, *J. Controlled Release*, 77, 287–295.
239. Perugini, P., Genta, I., Pavanetto, F., Modena, T., Maculotti, K., and Conti, B. (2002), Evaluation of enzyme stability during preparation of polylactide-*co*-glycolide microspheres, *J. Microencapsul.*, 19, 591–602.
240. Lupi, A., Perugini, P., Genta, I., Modena, T., Conti, B., Casado, B., Cetta, G., Pavanetto, F., and Iadarola, P. (2004), Biodegradable microspheres for prolidase delivery to human cultured fibroblasts, *J. Pharm. Pharmacol.*, 56, 597–603.
241. Morimoto, K., Chono, S., Kosai, T., Seki, T., and Tabata, Y. (2005), Design of novel injectable cationic microspheres based on aminated gelatin for prolonged insulin action, *J. Pharm. Pharmacol.*, 57, 839–844.
242. Hinds, K. D., Campbell, K. M., Holland, K. M., Lewis, D. H., Piché, C. A., and Schmidt, P. G. (2005), PEGylated insulin in PLGA microparticles. In vivo and in vitro analysis, *J. Controlled Release*, 104, 447–460.
243. Shenoy, D. B., D'Souza, R. J., Tiwari, S. B., and Udupa, N. (2003), Potential applications of polymeric microsphere suspension as subcutaneous depot for insulin, *Drug Dev. Ind. Pharm.*, 29, 555–563.
244. Zhou, S., Deng, X., He, S., Li, X., Jin, W., Wei, D., Zhang, Z., and Ma, J. (2002), Study on biodegradable microspheres containing recombinant interferon-alpha-2a, *J. Pharm. Pharmacol.*, 54, 1287–1292.
245. Luan, X., and Bodmeier, R. (2006), Modification of the tri-phasic drug release pattern of leuprolide acetate-loaded poly(lactide-*co*-glycolide) microparticles, *Eur. J. Pharm. Biopharm.*, 63, 205–214.
246. Wang, Y., Gao, J. Q., Chen, H. L., Zheng, C. H., and Liang, W. Q. (2006), Pluronic F127 gel effectively controls the burst release of drug from PLGA microspheres, *Die Pharma.*, 61, 367–368.

# LIPOSOMES AND DRUG DELIVERY

Sophia G. Antimisiaris,[1] Paraskevi Kallinteri,[2] and Dimitrios G. Fatouros[3]

[1]*School of Pharmacy, University of Patras, Rio, Greece*
[2]*Medway School of Pharmacy, Universities of Greenwich and Kent, England*
[3]*School of Pharmacy and Biomedical Sciences, Portsmouth, England*

**Contents**

1 Introduction
2 Liposome Structure and Characteristics
  2.1 Phospholipids: Structure Stability and Characterization of Lipid Membranes
  2.2 Physicochemical Properties of Liposomes
  2.3 Preparation of Liposomes
  2.4 Functionalization of Liposomes
3 In Vivo Distribution
  3.1 Conventional Liposomes
  3.2 Long-Circulating or PEGylated Liposomes
  3.3 Other Routes of Administration
4 Applications of Liposomes in Therapeutics
  4.1 Anticancer Drug Delivery
  References

## 1 INTRODUCTION

Liposomes are vesicles in which an aqueous volume is entirely surrounded by a phospholipid membrane and their size can range between 30 and 50 nm up to several micrometers. They can consist of one (unilamellar) or more (multilamellar) homocentric bilayers of amphipathic lipids (mainly phospholipids). Based on their lamellarity (number of lamellae)—and size—they are characterized as SUVs/LUVs

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

(small or large unilamellar vesicles) or MLVs (multilamellar vesicles). MLV liposomes are always large (at least cannot be considered small) and aqueous spaces exist in their center and also between their bilayers.

Liposomes were initially invented by Alec Bangham [1] to serve as a model for cell membranes in biophysical studies. In the 1970s they started to be investigated as promising drug carriers [2, 3]. The main advantages of liposomes as a drug delivery system are the following: (i) They have a very versatile structure which can be easily tairored in order to bear the properties needed for each specific application. (ii) They can accommodate any type of drug molecules either in their aqueous compartments (hydrophilic drugs) or in their bilayers (lipophilic drugs) or both (amphiphilic drugs). (iii) Last, but not least, they are nontoxic, nonimmunogenic, and fully biodegradable. Early attempts to use liposomes as a drug delivery system revealed the main limitation of the system and resulted in disappointment due to the fast nonspecifc clearance of liposomes from circulation by reticuloendothelial system (RES) cells [4]. However, it was realized that recognition by the RES macrophage system could be useful in antigen presentation, macrophage activation or killing, and elimination of parasitic infections (what is called passive targeting). From then on many different types of liposomes or methods to obtain longer circulation half-lives have been successfully invented by controlling the physicochemical characteristics of lipid bilayers and their interaction with the biological environment. Some of the basic methods to control the properties of liposomes are presented in Table 1. Today (Figure 1), several different liposome types are available, mainly conventional, long circulating (or "stealth," sterically stabilized, or PEGylated), targeted [or ligand bearing or immunoliposomes (when antibodies are used as targeting ligands)], cationic (for genetic material delivery), and deformable or elastic (or "transferosomes") [will be discussed in the paragraph about of skin delivery in Section 3.3]. As a consequence of these advances in liposome technology, a very broad range of liposome applications for drug delivery are being explored, some of which have resulted in life-saving products on the market or under late-stage clinical testing.

In this chapter we will provide information about the basic characteristics of liposomes staring from their building blocks, that is, phospholipids. After this, liposome structure, physicochemical properties, and stability, which are most important for their in vivo performance, will be discussed as well as methods used for liposome preparation, characterization, and stabilization. Following this first part which is more technological, we will move into the biological part and talk about the fate of conventional liposomes and sterically stabilized liposomes, as well as liposomal drugs, after in vivo administration by different routes [mainly intravenous (i.v.), intraperitoneal (i.p.), or subcutaneous (s.c.)] and also give some information about other possible routes for in vivo administration of liposomes. Finally, specific applications of liposomes in therapeutics will be presented, some in more detail, mainly for the therapy of different types of cancer.

## 2 LIPOSOME STRUCTURE AND CHARACTERISTICS

The main building blocks of most liposomal drug formulations are phospholipids. This chapter will start with an introduction of phospholipid structure and briefly

**TABLE 1 Methods and Results of Modifying Physicochemical Properties of Liposomes**

| Properties | Method | Result |
|---|---|---|
| Physicochemical properties | | |
| Size distribution | Sonication, extrusion, microfluidization | Control of circulation time, Increased extravasation |
| Membrane permeability | Lipid composition modification (cholesterol, thermotropic transitions) | Increased liposome stability, proton or temperature-induced sensitivity |
| Tendency for aggregation or fusion | Lipid composition modification, addition of cations | Formation of structures (as cochleates) that can carry antigens, deoxyribonucleic acid (DNA), vaccine formulations |
| Surface hydrophilicity | Steric stabilization by grafting hydrophilic molecules on liposome surface (linear dextrans [5], sialic acid–containing gangliosides [6], lipid derivatives of hydrophilic polymers [as PEG [7, 8], poly-*N*-vinylpyrrolidones [9], and polyvinyl alcohol [10]) | Increase of circulation time, modification of pharmacokinetics and tissue dispotition of liposomes and encapsulated drugs |
| Drug encapsulation efficiency | Active or remote loading [11] | Stable liposome encapsulation of drugs at high drug-to-lipid ratios, Many applications in drug delivery |
| Elasticity, rigidity | Introduction of detergents or edge activators and "skin lipids" (as ceramides) in liposome preparations [12] | Increased skin penetration and retention in skin which results in increased transdermal absortption of liposomal drugs |
| Surface properties, liposome stability, and other properties | Addition of spefific ligand on liposome surface, together with steric stabilization | Increased potential for efficient targeting |

describe the most important (for liposomes) physichochemical characteristics of lipids and lipid membranes. Afterward properties of liposomes and finally methods used for the preparation of liposomal drug formulations will be discussed.

### 2.1 Phospholipids: Structure Stability and Characterization of Lipid Membranes

Phospholipids are naturally occurring biomacromolecules that play an important role in the physiology of humans as they serve as structural components of biological membranes and support organisms with energy [13]. They are amphiphilic molecules with poor aqueous solubility and typically consist of two parts: a water-soluble group, the so called polar head, and an insoluble one, the backbone (Figure 2). The polar head group contains hydroxyl groups which are responsible for the surface

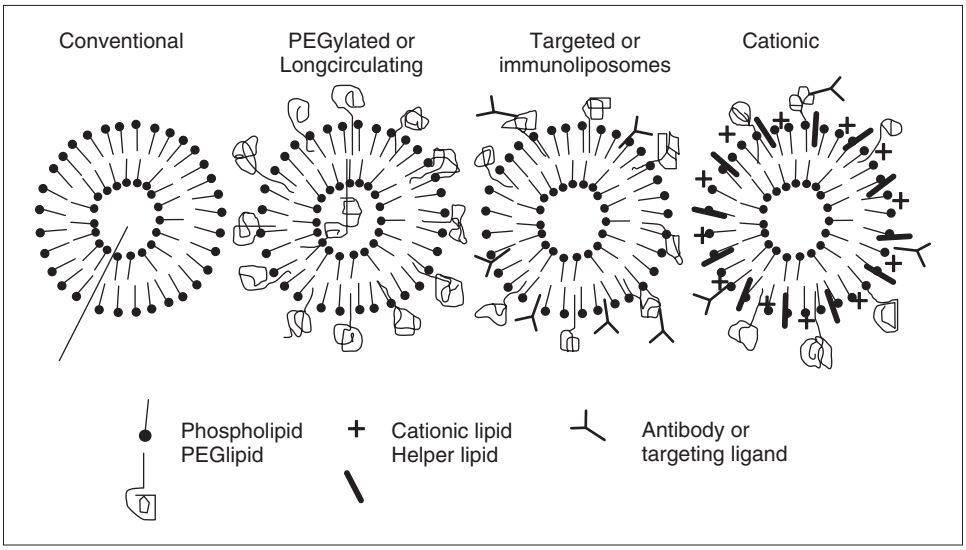


**FIGURE 1** Liposome types. *Conventional liposomes* are composed of phospholipids that form bilayers enclosing an aqueous compartment. Cholesterol may be included in the bilayer to increase membrane rigidity. Hydrophilic drugs can be encapsulated in the aqueous interior of the vesicles and lipophilic drugs can be included in their membranes. *Pegylated or long-circulating liposomes* have a surface coating of polyethylene glycol (PEG) molecules that permits liposomes to escape opsonization (coating with plasma proteins—opsonins—that make liposomes visible by RES macrophages). PEG-conjugated lipids are used for the preparation of this type of vesicle. *Targeted liposomes* or immunoliposomes are liposomes that in addition to a PEG coating (in most cases) have targeting moiety on their surface that directs them to the preferred target. This targeting moiety may be a sugar (i.e., galactose, to target cells with galactose receptors on their membranes) or other type of molecule or an antibody (usually monoclonal antibody), in which case the liposomes are characterized as *immunoliposomes. Cationic liposomes* are vesicles that consist of positively charged lipids (cationic lipids) which may form complexes with negatively charged DNA molecules and thus are used for gene delivery or targeting applications. For their preparation an additional lipid (helper lipid) is usually required. Cationic liposomes can also have PEG molecules on their surface (for longer circulation in the bloodstream) and/or targeting moieties. The last type of liposome is the so called *transformable liposome* or *elastic liposome*. (structure is presented and explained in Figure 10).

charge of the lipids that can be positively or negatively charged, zwitterionic, or noncharged. During liposome formation, these molecules arrange themselves by exposing their polar parts toward the water phase, while the hydrocarbon moieties (hydrophobic) adhere together in the bilayer. Two classes of lipids are mainly used for liposome preparation: double-chain polar lipids and sterols (mainly cholesterol). Such lipids form bilayers, in contrast to single-chain lipids (e.g., short-chain phosphatidylcholine) that form micelles, upon their dispersion in water. [14].

Double-chain lipids are either naturally occurring or synthesized in the laboratory. They consist of glycerol or sphingosine and a polar head containing a phosphor- or glyco-group. Glycerophopsholipids or phospholipids are the most popular, among the other lipids, for the preparation of liposomal dispersions [15]. Phosphatidylcho-

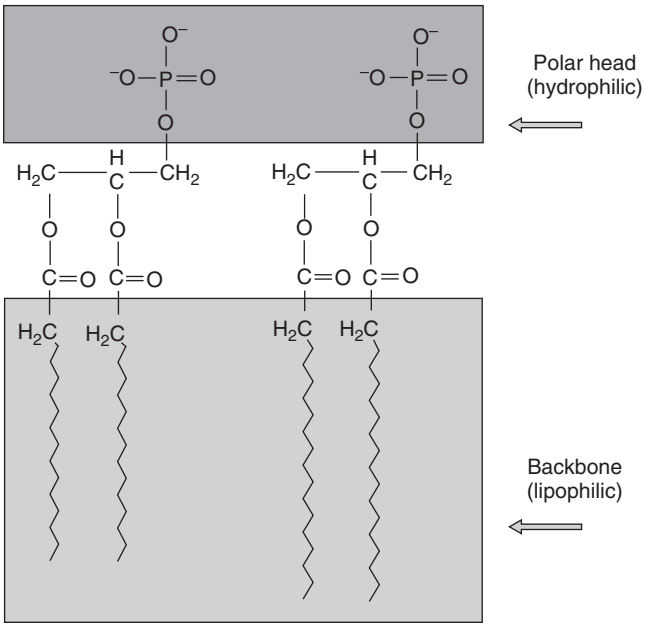


**FIGURE 2** Typical phospholipid structure.

line (PC) or lecithin, one of the main components of the liposomal bilayer, belongs to this group [1]. Phosphatidylcholine is zwitterionic in the range of physiological pH [16] and can be found in egg yolk. Other naturally occurring glycerophospholipids are the following: phosphatidylethanolamine (PE), which is isolated from brain lipids and is zwitterionic; phosphatidylserine (PS), which is found in bovine brains and posseses a negative charge; cardiolipin (CL), which is isolated from heart tissue or mitochondrial membranes and is negatively charged; phosphatidylglycerol (PG), which can be found in mitochondria or chloroplasts of mammalian cells and is negatively charged; and finally phosphatidylinositol (PI), a negatively charged lipid found in mammalian tissues [13].

Another group of naturally occurring lipids with applications in liposome technology is comprised of the sphingophospholipids (mainly sphingomyelin) which are derivatives of ceramides [17]. Sphingomyelin (SM) is found in the outer leaflet of plasma membranes [17] and has many similarities with PC since they both have the same zwitterionic polar group and two hydrophobic acyl chains.

Custom-made lipids can be produced by de- or reacylation of natural lipids. Commonly used phospholipids with polar heads containing myristoyl (14:0), palmitoyl (16:0), stearoyl (18:0) fatty acids are all classified by four-letter abbreviations, for example, DMPC, where DM stands for the number and type of fatty acids (di-myristoyl) and PC for the type of polar head (phosphatidylglycero-choline), and similarly DPPC and DSPC (Figure 2).

Positively charged lipids are capable of making complexes with deoxyribonucleic acid (DNA) (since it is negatively charged) and are currently very popular. Examples of such lipids are: *N*-[1-(2,3-dioleyloxy)propyl]-*N*,*N*,*N*-trimethylammonium chloride (DOTMA), 1,2-dioleoyl-3-trimethylammoniopropane (DOTAP),

and dicetyl phosphate (DCP), and dioctadecyldimethylammonium bromide (DODAB).

The last group of amphiphiles contains sterols that are present in the membranes of cells. The most popular among them is cholesterol (Chol), which can be easily incorporated in lipid bilayers, increasing their rigidity and making them less permeable, due to the interactions taking place with phospholipids in lipid membranes which result in modification of the lipid acyl-chain conformation.

Recently polyethyleneglycol (PEG, of varying molecular weight)–lipid conjugates have become commercially available and are frequently used in liposome applications. Aditionally, functionalized phospholipids exist for the covalent or non-covalent attachment of proteins, peptides, or drugs to the liposome surface. Most of these lipids fall into three major classes of functionality: Conjugation through amide bond formation, disulfide or thioether formation, or biotin/streptavidin binding.

Active lipids—mostly with anticancer activity—have also been added in liposome membrane for production of active liposomes. Examples of such lipids are ether lipids [18] and arsonolipids [19].

Several techniques are employed for the physicochemical characterization of lipid membranes, as summarized in Table 2. Thermal analysis, mainly differential scanning calorimetry (DSC), has been used extensively, offering information on the thermodynamics of various types of liposomes. The phase behavior of lipid components of membranes determines membrane fluidity. Each lipid has a characteristic lipid chain transition temperature, $T_m$. Changes in the structure of lipids occur below and above this temperature [20]. The temperature at which these changes occur depends on the head group, the chain length, and the degree and type of unsaturation of each lipid [21]. Using DSC studies in has been demonstrated that heat capacity curves are affected by the size of vesicles [20, 22], and can be modified by introduction of drugs [23, 24] or peptides [25, 26] in the lipid membranes (due to interactions between incorporated molecules and lipids).

**TABLE 2 Methods for Physicochemical Characterization of Lipid Membranes**

| Method | Information | References |
|---|---|---|
| Thermal analysis, mainly Differential scanning calorimetry (DSC) | Membrane fluidity<br>Lipid chain transition temperature, $T_m$ | 20–26 |
| Fluorescence spectroscopy | Phase transitions | 27, 28 |
| | Membrane dynamics | 29, 30 |
| Nuclear magnetic resonance (NMR) | Polymorphism | 31, 32 |
| | Lamellarity | 33 |
| | Membrane dynamics | 34, 35 |
| Electron paramagnetic resonance (EPR) | Fluidity of membranes | 36 |
| | Liposomal internal volume | 37 |
| | Membrane dynamics | 38 |
| | Membrane–drug interactions | 39 |
| Fluorescence quenching | Fusion processes | 40, 41 |
| X-ray diffraction | Structural information; thickness of the membrane and water layers | 42–45 |

## 2.2 Physicochemical Properties of Liposomes

The in vitro and in vivo performance of liposomes is highly dependent on their structural and surface properties. Liposome size and size distribution, surface charge (zeta potential), and trapping efficiency of the drug incorporated in the liposomes are important parameters that should be measured when developing a liposomal drug formulation. To obtain optimum performance of a liposomal preparation, parameters influencing both the liposome and the drug need to be carefully considered during early stages of development. In this chapter the most important physicochemical properties of liposomes will be discussed in terms of the way they may affect liposomal drug performance as well as the techniques used for their measurement.

***Liposome Size Distribution*** Liposomes have to be smaller than the vascular pore cutoff (380–780 nm) to extravasate and reach solid tumors [46]. Liposome size also plays an important role in complement activation and RES clearance of liposomes [47] and [48]. In general, vesicles that are larger than 100 nm require additional strategies for preventing surface opsonization and prolonging their circulation half-life (Table 1). Light scattering, field flow fractionation, microscopy, size exclusion chromatography, and turbidity are commonly used techniques for the physicochemical characterization of liposomal dispersions.

Quasi-elastic light-scattering or photon correlation spectroscopy is the most popular light-scattering technique. The Brownian motion of the particles causes fluctuations in the light intensity versus time. The hydrodynamic radius and the polydispersity index of liposomes can be easily determined from these studies [49–51]. Field flow fractionation is another approach to measure the particle size and the surface charge of liposomes [52–54]. The liposomes are exposed to a perpendicular field under laminal flow and their size and mass distribution can be determined. Size exclusion chromatography is a simple method to determine the size distribution of liposome dispersions [13, 55], especially on a routine basis, if other sophisticated equipment is not available. Samples with high heterogeneity are suitable to be analyzed with this method [55]. The selection of the proper gel, pretreatment of the column with sonicated vesicles to avoid any loss of material by adsorption, and use of isotonic buffer to avoid osmotic shock and frequent column calibration can secure the reproducibility of the method [13, 55].

Turbidity is a spectroscopic technique determining the optical density of colloidal particles. A wavelength between 350 and 500 nm is the first choice for such studies [56]. Turbidity measurements can offer important information on the kinetics of membrane–surfactant interactions since membrane solubilization changes reflect changes to the optical density of the dispersion [57–60]. It is also widely used as a technique to investigate liposome aggregation and fusion [61]. However the exact particle size of liposomes cannot be determined using turbidity techniques.

Optical microscopy is applicable for LUVs, MLVs, and especially giant liposomes [62, 63]. For instance, mechanical properties of the liposomal membrane can be studied by combining optical microscopy and a micropipette technique [64]. Electron microscopy can give information about both the morphology and homogeneity of liposomes. Quantification in terms of number can be carried out measuring at least 300 particles from different images [65]. Transmission electron microscopy (TEM) is a powerful technique since magnifications up to 200,000 and a resolution of approximately 1 nm can be achieved. Negative staining is a quite common

technique for visualization studies. A small amount of the sample is dried on a grid coated with carbon film. Then the film is coated with an electron-dense solution (e.g., tungsten molybdate). However the technique suffers from some drawbacks, such as artifacts due the fixation process or the extraction of lipid material by embedments [66, 67]. With freeze-fracture electron microscopy (FFEM) samples are quickly frozen and fractured. Compared with negative staining technique, FFEM has the advantage of preserving water-dependent lipid phases because no dehydration steps are involved. Therefore phase transitions, lipid polymorphism, or fusion processes can be visualized with this approach [68–71]. Recently cryogenic transmission electron microscopy (cryo-TEM) has been employed for visualization studies. With this technique very precise morphological assessment of liposome interior

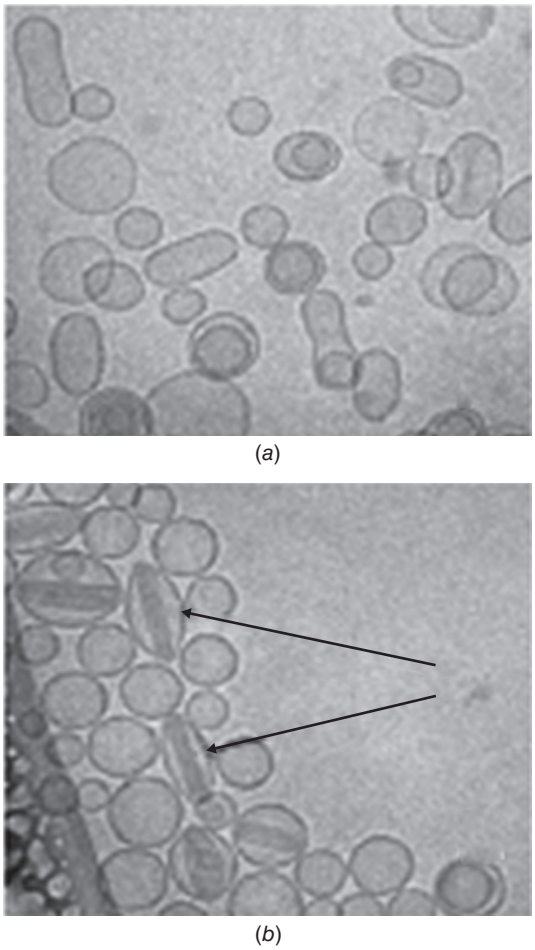


**FIGURE 3** Morphology of liposome of liposome by cryo-TEM (transmission electron microscopy). (*a*) Empty liposomes (liposomes with entrapped unbuffered $CuSO_4$ in the absence of drug). (*b*) Topotecan-encapsulating liposomes (drug was added to the empty liposomes of (*a*) to achieve a final drug-to-lipid ratio of 0.2 mol/mol, and the system was incubated at 20 °C). (Reproduced from ref. 72 with permission of Elsevier.)

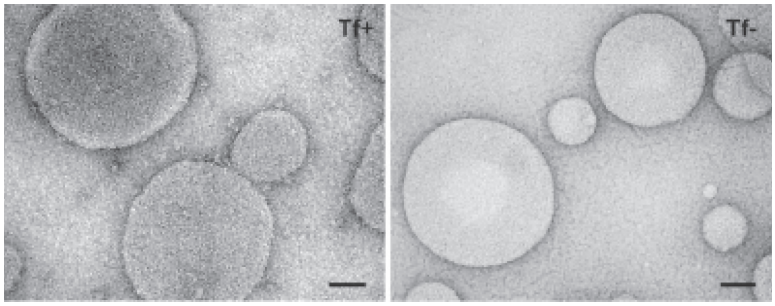


(*a*) (*b*)

**FIGURE 4** Morphological assessment of liposome surface coating using cryo-TEM. Liposomes that have been conjugated with transferrin display small particles on their surface (*a*), while control preparations appear to be smooth and undecorated (*b*). (Reproduced from ref. 73 with permission of Elsevier.)

(Figure 3) as well as liposome surface (Figure 4) can be carried out as recently demonstrated for Copper-Topexan encapsulating [72] and transferring-coated liposomes [73], respectively. Compared with the previous techniques the main advantage of cryo-TEM is avoidance of any fixation of the grid, which can create artifacts induced by staining and thus keeps the sample close to the original state [74, 75]. The samples for the cryo-TEM studies are prepared in a controlled environment vitrification system (CEVS). A small amount of the sample is placed on a grid-supported film. The grid is quenched in liquid ethane and it is vitrified. Then the samples are characterized with a TEM microscope. In a manner similar to the previous discussion, lipid polymorphism or fusion can be investigated [61, 76–79]. Finally scanning probe microscopy (SPM) including atomic force microscopy (AFM) has been recently applied in the liposome field [73, 80]. Some of the advantages of the technique are the high resolution in atomic dimensions (Figure 5 [73]), the production of three-dimensional images with high resolution, and the versatility of the operation conditions (vacuum air liquid). Aditionally, sample preparation does not involve any staining, freezing, embedding, or fracturing procedures.

***Surface Charge of Liposomes*** The electrical properties of liposomal surfaces can influence the physical stability of liposomal dispersions during storage as well as the behavior of liposomes in the biological milieu and their interaction with cells [47, 81].

Microelectrophoresis is used to measure the electrophoretic mobility or, in other words, the movement of liposomes under the influence of an electric field. From the electrophoretic mobility the electrical potential at the plane of shear or $\zeta$ (zeta) potential can be determined (by the Helmoholtz–Smoluchowski equation). From the zeta potential values the surface charge density ($\sigma$) can be calculated.

Aggregation of liposomes both in vitro and in vivo is one of their main stability problems. According to the Derjaguin-Landau-Verwey-Overbeek (DLVO) theory, or theory of colloidal stability, a colloidal system is stable if the electrostatic repulsion forces between two particles are larger than the attraction van der Waals forces. Therefore charged liposomal formulations are highly desirable. Manipulation of

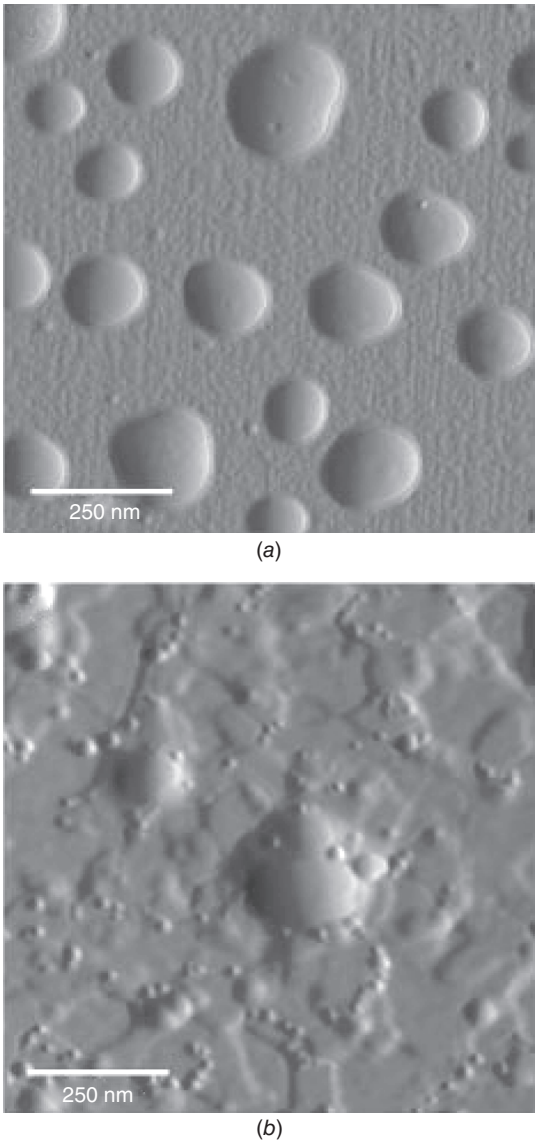


**FIGURE 5** Morphological assessment of liposome surface coating using atomic force microscopy (AFM). AFM images of liposomes prepared with DSPE–PEG 200–COOH. (*a*) Plain liposomal formulations. The liposomes show a smooth surface morphology. (*b*) Liposomes covalently modified with transferring. Small globular structures are visible at the surface. (Reproduced from ref. 73 with permission of Elsevier.)

liposome surface renders them stable against self-aggregation or nonspecific interactions, and this can be achieved by introducing other molecules on the surface, which may be natural molecules such as glycolipids [82, 83] or antibodies [84] or lectins [85] (that are usually grafted on the liposome surface by chemical linking).

As also mentioned in the introduction, the RES has been the "Achilles heel" to the delivery of liposomes by injection to the blood stream because of their rapid

uptake by the macrophages of the RES. Grafting of liposome surface with hydrophilic polymers has been used successfully as a method to protect the liposomes with a steric barrier which inhibits the adsorption of blood components [7, 86, 87]. It has been shown that *sterically stabilized* liposomes have significantly lower zeta potential values compared with conventional liposomes [87], indicating that the presence of PEG might shift the plane of shear away from the phosphate moieties. Thereby, zeta-potential modification can serve as proof of successful coating of liposomes by polymers.

***Drug Loading Efficiency and Techniques*** The amount of drug incorporated (for amphiphilic or lipophilic drugs) or encapsulated (for hydrophilic drugs) is a very important parameter that determines largely the selection of liposome type and components. Terms used to quantitate drug loading are drug–lipid (mol/mol) ratio (or D/L) and trapping or encapsulation efficiency (which is the percent D/L in the final liposomal formulation compared to the initial D/L used for liposome preparation). In many cases, the percent encapsulation is mentioned (percentage of drug encapsulated in relation with the amount of drug offered for encapsulation during liposome preparation); however, this number is practically useless and misdirecting, since it highly depends on the initial amount of drug offered to the specific lipid quantity, while the amount of lipid is not quantified.

Several parameters influence the encapsulation of drugs into liposomes. The types of vesicle and drug used play significant roles in the percentage of encapsulation. Hydrophobic drugs have higher loading efficiency in MLVs since they consist of high numbers of bilayers and low aqueous volumes [88–91]. In contrast, hydrophilic drugs have higher encapsulation in LUVs [91, 92]. Generally, for hydrophilic drugs, the percentage of encapsulation increases in the order of SUV $\leq$ MLV $<$ LUV [92].

The method used for liposome preparation has a significant impact on drug loading as well. Larger surface areas for the formation of thin films are preferable since they facilitate the hydration process of the bilayer, as demonstrated for doxorubicin (DOX) using five different hydration protocols [93]. The highest encapsulation was achieved when a thin film with large surface area was formed. Glass beads can be used for this purpose. From 8 to 10 times higher values for hydrophobic drug encapsulation were obtained after the addition of such beads [94]. The type of hydration (conditions and media used) also influences drug loading.

Drug encapsulation and stability of liposomes are also affected by the length of the acyl chain and the degree of saturation of the lipids used for their formation. As the acyl chain length of the lipid increases, so does the partitioning of hydrophobic drugs in the lipid membrane [95], as demonstrated for atenolol and propranolol in MLVs and SUVs [96]. The impact of the head group of different lipids used for the preparation of liposomes on encapsulation of citicoline was investigated by Puglisi et al. [97]. The fluidity of the membrane has also been demonstrated to influence encapsulation of different compounds [98–100].

When galactocerebroside was incorporated into liposomes, the percentage of encapsulation of mintoxantrone increased proportionally to the amount of glycolipid present in the membrane [101]. Numerous studies have demonstrated the impact of vesicle surface charge on drug loading efficiency, which is important for charge-bearing molecules. Encapsulation of hydroxycobalamin [102] and doxorubicin [103] was higher in negatively charged liposomes compared to neutral ones. In

a similar manner, higher amounts of calcitonin were encapsulated in positively charged compared to neutral liposomes [104] while the highest loading of indomethacin in liposomes was obtained in the following order: positively charged > negatively charged > neutral liposomes [105].

Cholesterol and α-tocopherol are used quite often to increase the rigidity and stability of liposomal membranes [88, 106–108]. In most cases cholesterol appears to improve the encapsulation of both hydrophilic and hydrophobic compounds. However, if the drug is lipophilic and partitions in the liposome membrane, there is a good chance that it might be displaced by adding increasing amounts of cholesterol in the bilayer (as observed in the case of dexamethasone encapsulating liposomes in our laboratory).

A novel system for enhancing the drug loading of lipophilic drugs combining liposomes and cyclodextrin–drug complexes by forming drug-in-cyclodextrins-in-liposome preparations has been proposed [109–111]. Cyclodextrins (CDs) are hydrophobic, cavity-forming, water-soluble oligosaccharides that can accommodate water-insoluble drugs in their cavities, increasing their water solubility. The basic intention is to encapsulate a stable water-soluble drug–cyclodextrin complex in the aqueous compartments of liposomes (Figure 6). As will be discussed below, this system can also serve as a method to increase the retention of lipophilic drugs in diluted liposome dispersions.

Finally, remote loading and active loading [11] are other methods used to achieve high trapping efficiency in liposome formulations, but unfortunately they can only be applied to a small number of drugs with specific physicochemical properties. This technique will be discussed below in Section 2.3.

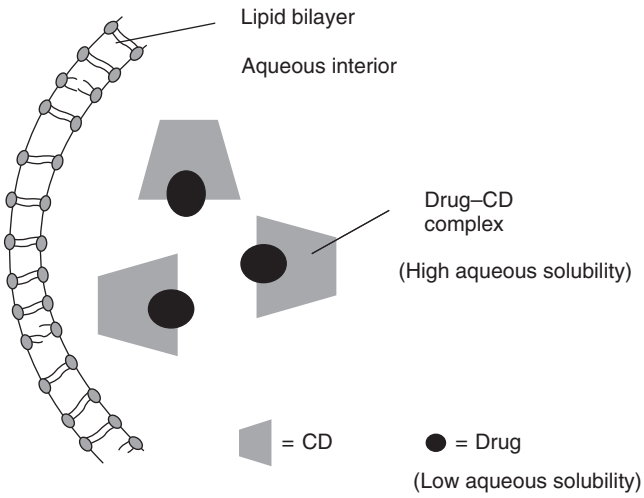


**FIGURE 6** Representation of drug in–cyclodextrin in liposome technique for encapsulation of lipophilic drugs in aqueous interior of vesicles. Drug molecules have low aqueous solubility and thus cannot be encapsulated in the aqueous compartment of the vesicle. However, the drug–cyclodextrin complex has high aqueous solubility and can thus be encapsulated in high concentrations in the vesicles.

***Stability of Liposomes*** A shelf life of at least two years is requested for pharmaceutical products. Therefore, chemical stability and physical stability are important parameters for the overall performance of liposomal formulations. Additionaly, another very important factor is the retention of encapsulated drug.

Several studies have sufficiently addressed the chemical degradation of liposomes during storage [112–115]. This is due to the hydrolysis of the phospholipids to fatty acids and 1- and 2-acyl-lysophospholipids. Further hydrolysis leads to the production of glycerol phospho compounds. Antioxidants (α-tocopherol), complexing agents [e.g., ethylenediaminetetraacetic acid (EDTA)], and inert atmosphere (e.g., nitrogen) are most commonly used to overcome this problem. Moreover, the presence of α-tocopherol can reduce the auto-oxidation of lipids, which is usually induced by light, metal ions, and temperature. Prevention of the chemical decomposition of the lipids by adding α-tocopherol in liposomal dispersions can increase the shelf life of liposomes [116]. Furthermore, the coexistence of cholesterol and α-tocopherol in the lipid bilayer can improve the antioxidant activity of tocopherol [117].

An alternative to circumvent problems related to the chemical decomposition of liposomes is their storage in the dry state (freeze dried). However, the protection against damage by freezing (cryoprotection) [118–120] and the protection against damage by dehydration (lyoprotection) [121, 122] require special attention for the proper storage of liposomes. Depending on the drug encapsulated in the liposomes and possible interactions between drug molecules and components of the lipid membrane, initial studies should be carried out in order to find the proper cryo- and lyoprotectant which preserves the integrity of the specific liposomal formulation.

The physical stability of liposome dispersions is mainly related with possible aggregation and leakage of the liposomal membrane. The size and surface charge play a significant role for the stability of liposomal dispersions, as has been discussed in the previous paragraph.

In addition to physical stability, the retention of drug in liposomal formulations is particularly important, not only during storage, but also during in vivo administration. Especially when targeted or long-circulating liposomes are used, association of the drug in the liposome carrier until the carrier reaches its biological target is a prerequisite for achieving any therepeutic benefit. However, in many cases, although the amount of drug loaded in liposomes is initially high, for different reasons most of the liposome-associated drug is rapidly released from the vesicles. The main causes for such behavior are different according to the physicochemical properties of the drug. *For hydrophilic drugs:* (i) The low integrity of liposome membranes after in vivo administration and contact with blood compomnents, which results in removal of some lipid molecules and concurrent opening of pores in the liposomal bialyer through which the loaded drug molecules may leak out, and (ii) the physical instability of liposomes that results in aggregation and fusion. and finally release of drug from liposomes. *For amphiphilic and lipophilic drugs*, the main problem is caused by the dilution of liposome dispersions, which usually occurs immediately after in vivo administration [by most routes, especially i.v. (diluted in 4 L of blood)]. This results in release of drug (since the drug can permeate membranes) until the drug saturates the full aqueous volume in which the liposome vesicles are diluted.

For hydrophilic drugs, the problem can be solved by increasing the rigidity of the liposome membrane while concurrently decreasing their tendency for aggregation, after proberly selecting the liposome lipid components. However, for membrane-permeable drugs, drug leaking upon dilution cannot be easily, if at all, confronted. Theoretically, one method may be to increase the affinity of the drug molecule with the lipid membrane (and at the same time decrease its aqueus solubility) by chemical modification (conjugation to a lipid or fatty acid). However, this is neither easy nor generally applicable. The association of liposomes with cyclodextrins has been recently proposed as a method to ensure high and stable entrapment of lipophilic drugs in aqueous compartments of liposomes [109–111, 123] (Figure 6), providing that the drug has high affinity for the cyclodextrin molecule and is not displaced from the cyclodextrin cavity by components of the lipid membrane (mostly cholesterol). In such cases, this approach may not increase drug retention, but encapsulation of drug may be substantially improved [111].

### 2.3 Preparation of Liposomes

***Liposome Preparation Techniques*** In most cases, liposomes are named by the preparation method used for their formation, Such as sonicated, dehydrated–rehydrated vesicle (DRV), reverse-phase evaporation (REV), one step, and extruded. Several reviews have summarized available liposome preparation methods [91, 124, 125]. Liposome formation happens spontaneously when phospholipids are dispersed in water. However, the preparation of drug-encapsulating liposomes with high drug encapsulation and specific size and lamellarity is not always an easy task. The most important methods are highlighted below.

*Thin Film Method* It was in 1964 that Alec Bangham introduced the "thin-film method" for liposome preparation [1, 126]. Lipids are dissolved in organic solvents (chloroform or mixtures with methanol) and the solvent is removed under a high-vacuum rotor evaporator forming a thin film on the walls of round-bottomed flask. Depending on the phase transitions of the lipids used for the preparation of liposomes, the aqueous phase for the rehydration should be prewarmed at temperatures above the phase transition of the lipid. After addition of the aqueous phase the thin film is detached from the flask walls by agitation and a highly heterogeneous population of MLVs is produced. Depending on the physicochemical properties of drugs, they can be introduced either in the thin film together with the lipids (lipophilic compound) or in the rehydration solution (hydrophilic compound).

*Sonication* This approach uses energy (ultrasound) and can be applied to a dispersion of MLVs [127] or to solid lipids mixed with aqueous solution. The flask with the liposome dispersion is placed in a bath sonicator or a probe sonicator (tip) is immersed in the tube containing the liposome dispersion. With the first setup it is difficult to reduce the size to the nanometer level since the energy produced by the bath sonicators is rather low. However, it has the advantage that there is no contact with the liposome dispersion. The position of the flask in the sonicator is equally important. It is easy to understand if it is in the right place from the noise produced by ultrasound waves. For instance, if foams are produced or there is no noise at all, that implies the sample is misplaced and finally the size of vesicles will not be reduced.

In the second approach (probe sonicator), the size of the liposomes can be reduced to nanometers and SUVs can be produced. As previously, the position of the probe plays an important role on the ability to minimize vesicle size. Because the energy produced from the transducer is high, overheating of the system is quite common; therefore, a water bath filled with ice is recommended. After sonication, fragments of Ti originated from the probe are scattered in the dispersion. Centrifugation for 4–5 min at 10.000 rpm will cause sedimentation of these fragments, giving clear liposome dispersions.

*Injection Methods*

*Ethanol Injection* Small unilamellar vesicles (with diameter of 30 nm) can be prepared with the ethanol injection technique [128]. Lipids are dissolved in ethanol and injected rapidly in the aqueous solution under stirring (final concentrations up to 7.5% (v/v) ethanol can be applied). The method is very easy, having the advantage of avoiding chemical or physical treatment of lipids. However, there is an extra step to remove ethanol and the concentration of vesicles produced is rather low. Also encapsulation of hydrophilic drugs is also low, due to the high volumes used.

*Ether Injection* The general principle of this method is the same as ethanol injection. The only difference is that the lipid is injected slowly in the aqueous solution that is warm [129]. Furthermore, the concentrations used in this case are somewhat higher (up to 10 m*M*) compared to the ethanol injection approach.

*Extrusion (Extruded Vesicles)* The extrusion method, which today is very popular for the production of homogenous vesicle samples of a predetermined size, was introduced by the group of D. Papahadjopoulos [65]. Multilamellar vesicles are extruded through filters with well-defined pores under pressure. Polycarbonate is the most commonly used material for these membranes, which have pore sizes from 30 nm up to several micrometers. For lipids with high melting point, the extrusion should be carried out above their phase transition temperature. The operating volumes are from 1 to 50 mL with a liposome dispersion concentration up to 150 m*M*. Generally, repeated extrusions reduce the number of lamellae and the produced liposomes are mainly unilamellar. High pressure can cause disruption of membranes. The reproducibility of the method is good; however, it is quite time consuming and membrane rupture problems occur quite often.

*French Press* With this approach MLV liposomes are introduced in a cell and a piston presses the dispersion [130]. Pressures up to 25,000 psi can be achieved and SUVs are produced. The main disadvantage of this technique is that it is not applicable for lipids with phase transition temperatures lower than 20 °C and the concentration of liposomes that can be used is relatively low (maximum 20 m*M*).

*Microfluidization* During microfluidization, MLVs are circulated with a pneumatic pump operating under high pressure through a prefilter and then to the interaction chamber [131]. From there, they are separated into two streams and they pass through defined microchannels under high velocity to a heat exchanger which is

connected with a water bath. This is repeated many times until the size of the liposomes is significantly reduced. They can operate with volumes from 0.1 up to 10 L of liposomes and with concentrations up to 300 m*M*, which is by far the highest capacity from all other methods. Small unilamellar vesicles less than 100 nm can be produced, but the population is not completely homogeneous.

*Reverse-Phase Evaporation* The REV method was developed by Szoka and Papahadjopoulos [132]. Lipids are dissolved in organic solvent and the solvent is removed with evaporation. The thin film is resuspended in diethyl ether (1 mL solvent/mL liposomes) followed by the addition of one-third of water and sonication in a bath sonicator for 1 min. This water-in-oil (w/o)-emulsion is evaporated until a dry gel is formed, and finaly the gel is broken by agitation and water addition. Sometimes this step is quite difficult. The remnants of the organic solvent are removed by evacuation and the resulting dispersion is REV liposomes.

*Dehydrated–Rehydrated Vesicles* DRV liposomes were developed by Kirby and Gregoriadis in 1984 [133] and are capable of encapsulating high amounts of aqueous soluble molecules under mild conditions (conditions that do not cause decomposition or loss of activity). The high entrapment ability of this type of liposomes is due to the fact that preformed, "empty" SUVs are disrupted during a freeze-drying step in the presence of the solute destined for entrapment. Subsequently, during controlled rehydration that is carried out in the presence of concentrated solution of the solute (to be encapsulated), the vesicles fuse into large oligolamellar vesicles entrapping high amounts of solute. The produced liposomes are multilamellar and their size is between 200 and 400 nm up to a few micrometers. Recently, with the addition of certain amounts of sucrose, DRV liposomes with diameter between 90 and 200 nm were obtained entrapping considerable proportions (up to 87) of the solute [134].

*Giant Vesicles* Large or giant liposomes have been developed by Reeves and Dowben [135]. Briefly, lipids are dissolved in chloroform/methanol 2:1 and dispersed on a piece of glass. Water is added for their rehydration; however, their population is quite heterogeneous. Other types of particle-encapsulating giant liposomes [136] can be prepared by applying a double-emulsion technique followed by a freeze–drying step.

*Detergent Depletion* With this method phospholipid–detergent mixed micelles are initially produced, and during controlled-rate detergent removal, liposomes are formed. The rate and method of detergent removal determine the size and size homogeneity of the liposomes produced. Gel filtration and dialysis are the most popular approaches [137, 138]. Although liposomes are produced under mild conditions (low temperature and low shear mechanical forces applied), this method suffers from low encapsulation efficiency of hydrophilic drugs.

*Large Unilamellar Vesicles from Cochleates* Large unilamellar vesicles can be produced with the "cochleate" approach [139]. Small unilamellar vesicles consisted from phosphatidylserine adopt a cochleate shape after addition of calcium. Addition of EDTA creates complexes with calcium, turning the cochleates to LUVs.

*One-Step Method* The "one-step method" has been introduced by Talsma et al. [140]. Lipid dispersions are hydrated at high temperatures in the presence of a steam of $N_2$. Liposomes between 200 and 500 nm can be prepared with this approach.

***Large-Scale Manufacturing*** Despite the fact that there are many methods to prepare liposomal dispersions not all of them are applicable for scaling up. In fact, scaling up to larger batches could be a monumental task. Among all the methods reviewed so far, microfluidization and homogenization are the most powerful methods to produce large quantities of liposomes. New homogenizers have a capacity of 1000 L/h and require a minimal sample volume of 2 L [141]. The fact that lipid concentrations up to 300 m*M* can be used secures high encapsulation capacity.

***Remode and Active Drug Loading Techniques*** The main advantages of these approaches are the high encapsulation efficiency and low leakage of the encapsulated material.

An in situ method for "remote" drug loading based on the development of a pH gradient across the internal and external water phases of the membrane has been established. A transmembrane pH gradient induces the uptake of charged drugs into liposomes. Drug encapsulation is based on its partitioning between the lipid and aqueous parts. This process is governed mainly by pH and to some extent by the ionic strength of the medium. Drugs that are weak bases can diffuse through the lipid membrane as unprotonated species. The presence of a proton gradient makes them more hydrophilic, allowing them to accumulate in the intraliposomal aqueous phase. Encapsulations up to 90% have been reported for doxorubicin [142–144] and vincristine [145].

An alternative but similar technique is based on an ammonium sulfate gradient used to obtain "active" loading of amphipathic weak bases into the aqueous compartment of liposomes. This has been used for active loading of anthracyclines, acridine orange, epirubicin, and doxorubicin [11, 146, 147] at very high efficiency (>90%). In the case of doxobubicide most of the intraliposomal drug is present in the aggregated state. Additionally, antracycline accumulation in liposomes is stabilized for prolonged periods of storage due to aggregation and gelation of antracycline sulfate salt.

Active entrapment and loading stability are dependent on liposome lipid composition, lipid quality, medium composition, and temperature as well as on the $pK_a$ and hydrophobicity of the base. The ammonium sulfate gradient approach differs from most other chemical approaches used for remote loading of liposomes, since it does not require liposomes with acidic pH interior or an alkaline extraliposomal phase.

In addition to the remote or active loading techniques mentioned above, metal complexation reactions have been demonstrated to achieve accumulation of doxorubicin in liposomes [148, 149]. Furthermore, copper–topotecan complexation has been recently seen to mediate drug accumulation into liposomes and is proposed as a methodology to prepare liposomal camptothecin formulations [72].

## 2.4 Functionalization of Liposomes

"Active targeting" is used to describe the specific liposomal drug localization achieved by grafting various moieties (antibodies, lectins, polymers, etc.) on the carrier surface.

There are a number of techniques available to attach the suitable ligand on the liposome surface, either by covalent or noncovalent coupling [150, 151].

These techniques should be fast, efficient, and reproducible, yielding stable nontoxic bonds, while the conjugated ligands should maintain the ability to recognize the target site with high binding affinity. Also, the coupling method should not affect the blood clearance of the formulation, colloidal stability, drug incorporation, and release in a negative way. For example, when antibodies were attached on the liposome surface where PEG molecules were grafted as well to ensure prolonged carrier retention in the blood, it was shown that the ligand binding efficiency on the bilayer was low as well as the binding ability to the target [152, 153]. The latter problem was opposed by attaching the ligand at the distal end of the PEG molecules already grafted on the liposomal bilayer [154].

***Covalent Binding of Ligands*** The majority of ligand coupling is achieved by covalent reactions with hydrophobic anchors. The procedure could be carried out in two patterns: Either the hydrophobic anchor is included already in the liposomal bilayer and the ligand interacts with the anchor on preformed liposomes [155] or the ligand–hydrophobic anchor conjugate in the form of micelles is mixed with the liposomes [156]. In the first instance, where a hydrophobic anchor is mediated between the liposomal surface and ligand, covalent attachment can occur via a thioether bond [157–161], via a disulfide bridge, between carboxylic acid and the primary amine group [162], via hydrazone, or via cross-linking between two primary amine groups (Figure 7).

The formation of the stable thioether bond is a reaction between thiol moieties of proteins mainly with maleimide groups. Usually, PE or PEG–PE or PEG–DSPE (distearoyl-phosphatidylethanolamine) has been functionalized with maleimido (Mal-), maleimido-phenylpropionate (MP), or pyridil-dithio-propionylamino (PDP) groups, which eventually will react with the thiol groups. Also, sometimes the ligand does not carry enough thiol groups or those are completely absent, so they have to be introduced using a heterobifunctional cross-linker, such as SPDP [*N*-hydroxysuccinimidyl 3-(2-pyridyldithio) propionate] or SATA (succinimidyl-*S*-acetylthioacetate), which introduce one amine group. Still, in the case of SATA, deacetylation using hydroxylamine is necessary to uncover the thiol group, while if SPDP is used, the produced disulfide bond has to be reduced to thiol groups with dithiothreitol (DTT). The former cross-linker is more preferable as only mild conditions are used to make the thiol group available.

It has been shown that attaching the ligand at the distal end of PEG molecules combines the advantages of a specific drug delivery system with steric stabilization for higher stability in the blood [163, 164]. However, Longmuir et al. showed that introducing a peptide from *Plasmodium* at the free end of $PEG_{3400}$–AP (aminopropane) was not capable of retaining the liposome stability, so $PEG_{5000}$–PE molecules were added to the liposomal composition resulting in PC/PE–$PEG_{5000}$/AP–$PEG_{3400}$-peptide liposomes with molar ratio of 86:10:4 [165]. The coupling efficiency of ligand attached on PEG was higher (60–70% or even 100% in some cases) compared to that achieved using, for example, *N*-(4′-4″-maleimidophenyl)butyrol)-dioleoylphosphatidylethanolamine (MPB–DOPE) (only 10%) [164, 166].

Antibodies, whole or fragments, have been attached successfully on liposomal surfaces and more commonly at the free end of PEG molecules, as has been reported

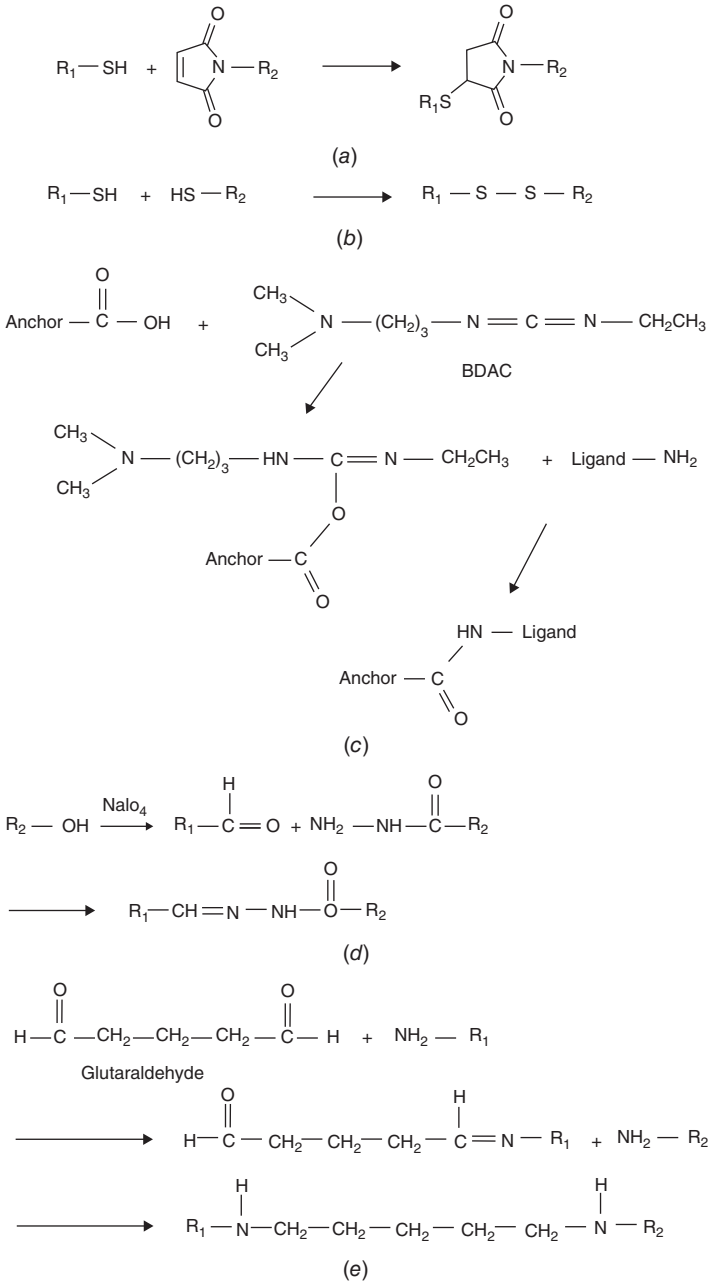


**FIGURE 7** Schematic of different coupling methods used: (*a*) reaction between meleimide and thiol functions; (*b*) formation of disulfide bond; (*c*) reaction between carboxylic acid and primary amine group; (*d*) reaction between hydrazide and aldehyde functions; (*e*) cross-linking between two primary amine functions. (Reproduced from ref. 150 with permission of Elsevier.)

in a number of studies since the interaction with the cells is much more favored [161]. Using Fab′ fragments is more advantageous as the Fc part which mediates MPS activation through a receptor, is omitted [157, 160]. Moreover, the distance of the antibody fragment from the liposomal surface is another important factor which determines the drug delivery system uptake by the cells, as reported by Mamot et al. [167]. Besides the latter has been reported elsewhere even if the coupling reaction is different from thioether formation [168].

Examples of peptides attached on liposomal surface via thioether bonds were TAT-peptide and antagonist-G on Mal–PEG2000–DSPE and PDP–PEG–DSPE, respectively [169, 170]. Both ligands exhibited significant increase in cell uptake.

Another possible way for ligand conjugation on liposomes is the formation of a disulfide bridge, which is quite unstable in serum and thus it is not used as much [171].

However, an amide bond formed between the carboxylic acid group on the liposome surface (DSPE–PEG–COOH) and the primary amine of the ligand is favored as the ligand modification is not necessary. According to this method, an acyl amino ether is produced in the presence of 1-ethyl-3-(dimethylaminopropyl) carbodiimide (EDAC) and *N*-hydroxysulfosuccinimide (NHS), which eventually will react with the primary amine of the ligand [150]. For example, Wartchow et al. improved the efficacy of a small integrin antagonist of the extracellular domain of the $\alpha v\beta 3$ integrin by grafting it on dextran-coated liposomes [172]. DPPE-succinate was included in the liposomal bilayer and a 3-amino-2-hydroxypropyl ether derivative of dextran was added to preformed liposomes in the presence of EDAC, while unreacted succinyl groups were converted to amides. The amino groups of dextran were succinylated and integrin antagonist (IA) was attached on the succinamidodextran liposomes in the presence of EDAC. The final IA–dextran liposomes had a size of 110 nm, which is attributed to dextran coating, as the liposomes without dextran were 60 nm in diameter. The antiangiogenic mechanism of IA–dextran liposomes as well as the apoptotic potency was proved after a series of studies.

Moreover, Voinea et al. attached antibodies against vascular cell adhesion molecule-1 (VCAM-1) overexpressed on activated human endothelial cells on liposomes with the intention of using them as drug carriers [162]. *N*-gluraryl-PE was used as membrane anchor for the antibody coupling via its free amino groups after its activation with carbodiimide. There is no necessity of antibody modification before the coupling reaction.

Also, transferrin was grafted onto liposomes containing *N*-glutaryl-PE activated with carboxidiimide with the final plan to use those as carriers for inhalation therapy for lung cancer [173]. Tfr liposome uptake was significantly higher from immortalized or cancer cell lines, but to reduce uptake from alveolar macrophages, PEG molecules are attached on the liposome surface. In general, transferrin is a glycoprotein that consists of a single chain of amino acids which has been coupled on the liposome surface for a number of applications because its receptor is overexpressed at malignant cells so a higher amount of transferrin binds on the cell membrane [174]. Its covalent attachment on liposomes takes place either by conjugation between transferrin-lipid and insertion of it on the preformed vesicles [174, 175] or preparation of liposomes with activated lipids and reaction with activated transferrin [176].

In addition, Torchilin et al. synthesized pNP (*p*-nitrophenylcarbonyl)–PEG–DOPE to enable protein coupling via its amino group at the distal end of PEG mol-

ecules on liposome surface in a quantitative manner at pH around 8.0 [177]. The only disadvantage mentioned is the hydrolysis of pNP groups from PEG–DOPE at pH higher than neutral (complete hydrolysis occurs in 1–2 h at pH around 8.0). Therefore, the coupling reaction between protein and pNP has to take place faster or at least at the same ratio. It was shown that 65% of pNP binds to ligand at pH 9, so the binding efficiency and time are adequate at the conditions studied. Also, the amount of pNP–PEG–DOPE was critical for successful protein binding; it was shown that 1 mol % of pNP–PEG–DOPE was enough to bind approximately 100 protein molecules. Incorporation of antibodies, lectins, avidin, and nucleosomes did not seem to alter the activity of those molecules even at high concentrations of pNP–PEG–DOPE. However, Savva et al. conjugated a genetically modified tumor necrosis factor (TNF) at the free end of the $PE_{3500}$ molecule by reacting the latter with NHS and DCC (*N*,*N*′-dicyclohexyl carbodiimide) to introduce the succinyl groups which will react with the phospholipid to produce DOPE–PEG–COOH [178]. Then, the carboxyl group of the derivatized PEGylated molecule located on the liposomal bilayer was activated with EDAC and NHS. Consequently, recombinant TNF was added in the liposome suspension for the final conjugation to occur at 4 °C where the overall coupling efficiency was approximately 55%. However, it was shown that the biological activity of TNF was reduced when attached on the liposomes as the degree of PEG modification increased irrespective of the PEG molecular weight (MW). Also, the formulation did not show the prolonged blood circulation expected due to the PEG presence. Those results were attributed to a number of reasons, including either damage of the protein during the coupling reaction or possible dissociation of the trimeric form of rKRKTNF to a monomeric less active form or cross-linking between PEG and rKRKTNF during the coupling reaction.

A hydrazone bond is an alternative way of antibody coupling onto liposomal surface so as to avoid the use of essential (for the recognition) amino groups at the coupling reaction (by using the maleimido method) or the risk of rapid clearance due to Fc moiety/segment [150, 179]. According to this method, the carboxylic group of the heavy chain of the antibody undergoes mild oxidation by sodium periodate or galactose oxidase to aldehyde groups. The oxidized product can be coupled either on the hydrazide-hydrophobic anchor inserted in the lipid bilayer [180] or at the free end of PEG molecules [181, 182]. Hydrazide cross-linkers used are, for example, *S*-(2-thiopyridyl)-L-cysteine hydrazide (TCPH), *N*-acetylmercaptobutyric hydrazide (AMBH), and 3-(2-pyridyldithio)propionic acid hydrazide (PDPH). TCPH is structurally closely related to PDPH and could be expected to behave in a similar manner [183]. Unprotected mercaptohydrazides such as AMBH are unsuitable since the free thiol function is susceptible to oxidation and may also reduce the disulfide bonds in immunoglobulin IgG at the concentrations required for conjugation. The disadvantage of this method is the low coupling efficiency (17% only) [166]. According to Ansell et al. [179], the drawbacks reported was the possible damage of some amino acid residues, such as methionine, tyrosine, and tryptophan, due to periodate. Therefore, antibodies sensitive to periodate treatment would be unsuitable candidates for the PDPH protocol. Also, the hydrazone bond might undergo hydrolysis after six weeks of storage, which would be a potential problem if the conjugate would not be used soon after preparation. However, it is possible to stabilize the bond using sodium cyanoborohydride to reduce the hydrazone linkage if long-term storage is required [179].

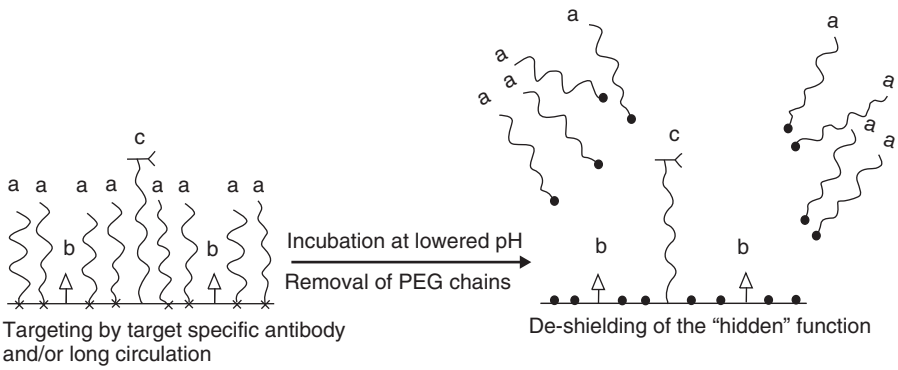


**FIGURE 8** Schematic for design of multifunctional drug delivery system (DDS) that includes pH-cleavable PEG-Hz-PE (a), temporarily "shielded" biotin or TATp (b), and monoclonal antibody (c) attached to surface of DDS via pH-noncleavable spacer. (Reprinted with permission from ref. 183. Copyright 2006 by the Americam Chemical Society.)

However, the labile nature of the hydrazone bond is used to formulate "smart" drug delivery systems where the basic idea is to introduce in parallel a pH-responsive ligand spacer in the lipid bilayer (which is $PEG_{5000}$–Hz–PE), a temporarily shielded biotin or TATp and mAb (monoclonal antibody) attached to the surface of the drug delivery system via a noncleavable bond (TATp–$PEG_{2000}$–PE) [183]. Such a system will be able to respond to environment stimuli such as pH changes, where, for example, at acidic pH (5.0–6.0) PEG5000 molecules will be detached from the carrier surface and biotin or TATp will be available to either bind to avidin or be internalized by the cells, respectively (Figure 8). The monoclonal antibody and biotin or TAT has been attached on pNP–PEG–PE. The produced mAb DDS (drug delivery system) demonstrated clear immunoreactivity toward the antigen. However, some affinity decrease was observed for the antibodies modified with the pNP–PEG–PE anchor and incorporated onto the immunoliposomes. Biotin binding to avidin was pH dependent with higher retention (75%) at pH 5.0, where the shielding PEG molecules were cleaved away. Significant increase of DDS uptake by the cells was achieved when TAT was incorporated on the liposome surface at pH 5.0.

An avidin–biotin system has been used to attach antibodies in the bilayer of DDSs. Xiao et al. developed a three-step strategy to improve the tumor-to-tissue ratio of anticancer agents [184]. Two antibodies specific for the CA-125 antigen that is highly expressed on NIH:OVCAR-3 cells were used. These cells were prelabeled with biotinylated anti-CA-125 antibody and fluoroscein isothiocyanate (FITC)–labeled streptavidin (SAv) prior to administration of biotinylated liposomes. Both antibodies were specifically bound to the cell surface of OVCAR-3 cells but not to SK-OV-3 cells, which do not express the specific antibody. Antibody biotinylation did not affect its immunoreactivity.

Schnyder et al. explored the targetability of biotinylated immunoliposomes to skeletal muscle cell line in vitro [185]. OX26 mAb binds to transferrin receptor and is covalently attached to streptavidin by introducing thiol groups using 2-iminothiolane (Traut's reagent). Immunoliposomes consisted of DSPC (5.2 μmol), cholesterol

(4.5 μmol), PEG–DSPE, (0.3 μmol), and linker lipid (bio-PEG–DSPE; 0.015 μmol). OX26 mAb–streptavidin was added to preformed liposomes in a 1:1 ratio. According to estimations, the average number of bio-PEG–DSPE molecules was 30, assuming that one 100-nm liposome contains 100,000 phospholipid molecules. Uptake experiments with muscle cell line using the OX26 mAb, fluorescence-labeled OX26–streptavidin, or fluorescent OX26–immunoliposomes demonstrated cellular uptake and accumulation within an intracellular compartment of the OX26 mAb and its conjugates.

All the methods described earlier consider that coupling of the ligand on the anchor already existed on the liposome surface. Another option is the ligand–anchor conjugate in the form of micelles to mix with the liposomes. According to that, anti-CD19 mAb was thiolated using Traut's reagent and reacted with Mal-PEG–DSPE in a micellar form with PEG–DSPE and molar ratio 4:1 [186]. Antibodies were coupled at the end of the PEG–DSPE. Consequently, micelles were incubated with preformed liposomes at molar ratio 0.05:1, respectively, for 1 h at 60 °C.

In another study, mAb 2C5 with nucleosome-restricted specificity, which recognizes specifically human brain tumor cells, was tested as a potential ligand candidate for liposome targeting to brain tumor cells [187]. The mAb was attached to pNP–PEG–DSPE and the formed micelles incubated with preformed liposomes. The 100–200 mAbs bound per single liposome of approximately 200 nm in diameter. A slight reduction in immunoreactivity was observed for a single antibody molecule for a number of reasons; the overall evaluation was sufficient target recognition and affinity due to multipoint attachment of immunoliposomes to the target via several antibody molecules. Indeed, the immunoliposomes showed threefold higher accumulation in the tumors compared to nonspecific carriers.

At this point it has to be emphasized that this method seems to be the most advantageous one because damaging chemical reactions are excluded as they happen at a different stage. Also, this method provides the flexibility of attaching a large variety of ligands on liposomes of various compositions loaded with different drugs. Apparently, targeted liposomes produced with this last technique have shown similarities in the in vitro drug leakage, cell association, and therapeutic efficacies to liposomes made by conventional coupling procedures.

***Noncovalent Binding of Ligands*** According to this procedure, the ligand is added to the lipid mixture during liposome preparation. Small molecules such as sugars have been attached on the liposome or lipoplex surface in this way. At first galactose, mannose, and fucose were modified to Gal–C4–Chol, Man–C4–Chol, and Fuc–C4–Chol and they were added in the lipid mixture of DSPC/Chol/Sugar–C4–Chol with a ratio 60:35:5. Chol was chosen due to the stability in the liposomal membrane while only one sugar was conjugated so the lipophilicity of the final glycolipid would not be altered considerably, and thus the stability of the latter in the liposomal membrane would be more secure [188]. After in vivo administration of 0.5% Gal, Man, and Fuc liposomes it was found that the ratio of their uptake from parenchymal/nonparenchymal liver cells was 15.1, 0.6, and 0.2, respectively. Also, in high doses, 5% Gal liposomes are taken up by nonparenchymal liver cells as well as the parenchymal ones, while they are capable of inhibiting the uptake of Fuc liposomes by nonparenchymal cells.

Even if this is a simple and mild technique, there is always a concern about ligand orientation, very low attachment efficiency achieved (4–40%), and the liposome aggregation often observed.

## 3 IN VIVO DISTRIBUTION

Successful treatment depends not only on the formulation characteristics but also on the route of administration. For example, the schistosomicidal drug tartar emetic incorporated in PEGylated liposomes was delivered either intraperitoneally or subcutaneously (27 mg Sb/kg) to mice infected with *Schisostoma mansoni* [189]. Indeed, 82 and 67% reduction levels of worm were obtained, respectively. However, the efficacy of the formulation given by either administrative route was not significantly different. The only difference was the slower liposome absorption by the subcutaneous route.

Also, the therapeutic effect of liposomal adriamycin (PC/Chol, 120 nm) was enhanced significantly after concurrent i.v. and s.c. administration to rabbits bearing VX2 tumors in the mammary gland [190]. The i.v. route significantly inhibits breast tumor and metastasis, while the s.c. route acts on local-regional lymph nodes. That was proved by slowed growth rates, decreased messenger ribonucleic acid (mRNA) expression of proliferating cell nuclear antigen, and extensive necrosis and apoptosis of tumor cells. Even if allergic reactions have not been reported after s.c. injection of liposomes, there is more to be done on systemic toxicity.

The therapeutic efficacy of paclitaxel is stronger after drug incorporation in magnetoliposomes injected either i.v. or s.c to an EMT-6 breast cancer mouse model [191]. The carrier manipulation due to the application of a magnetic field led to their increased accumulation to the tumor site. However, paclitaxel accumulation is slightly lower after s.c. administration, probably attributed to time delay during the drug transportation process to the circulation.

In another study by Wang et al., i.v. injection of liposomes carrying rat insulin promoter (RIK)–thymidine kinase (TK) was found to be less toxic to the liver than the i.p. injection of the same formulation to severe combined immunodeficient mice (SCID) [192]. The direct injection of the liposomes to abdominal cavity probably leads to higher local absorption and, thus, higher liver toxicity. In contrast, the i.v. injected volume is smaller (70 μL to 100 μL of i.p. injection), which is diluted fast as soon as it enters the body.

### 3.1 Conventional Liposomes

Conventional liposomes are those that do not carry any sterically stabilizing or targeting moieties on their surface. Their biodistribution depends strongly on their physicochemical properties (size, $\zeta$ potential, composition) and physiological and pathological conditions of the body [193]. Thus, conventional liposomes comprise the passive targeting of drug molecules.

***Intravenous Administration*** Liposomes administered intravenously face barriers such as the endothelial lining of the vasculature and the blood–brain barrier. Extravasation of the liposomes occurs only in organs such as liver, spleen, and bone

marrow (due to leaky fenestrae and loose junctions between the endothelial cells) and under certain pathological conditions (presence of tumors, inflammation, infection). Besides, neutral (uncharged) liposomes of size smaller than 100 nm show slow blood clearance compared to others of larger size and/or positive or negative charge (due to presence of opsonins) [194, 195]. Also, lipid exchange between liposomal carriers and plasma lipoproteins contributes to liposomal membrane rupture and consequently loss of the therapeutic substance [196].

Therefore, conventional liposomes are used mostly in treating the RES system or to mask the toxic side effects of anticancer drugs. Many anticancer drugs entrapped in liposomes have shown altered biodistribution and reduced toxicity. Plain SUV liposomes consisting of DSPC/Chol (known as "Stealth") in a molar ratio 2:1 or 1:1 have shown particularly promising vehicles as reported in a number of studies of animal models. They have undergone preclinical and clinical studies due to relatively low levels of RES uptake and the high level of tumor targeting exhibited [197–200]. Indeed, $^{111}$In-labeled DSPC/Chol liposomes have proven capable of targeting a number of tumors [201–206]. Although the significant RES uptake was a fact, about 45–50% of liposomes remained in the blood circulation 4 h after i.v. injection. Positive images on gamma camera were reported. In patients with recurrent high-grade gliomas, 1% of the injected liposomal dose was taken up by the tumor [205].

Moreover, anthracyclines have been formulated in conventional liposomes and are commercially available. Doxorubicin for i.v. use is commercially available in the form of Myocet, which consists of egg PC and cholesterol [207]. It is recommended in combination with cyclophosphamide for metastatic breast cancer. Thus, drug entrapment in liposomal vesicles may reduce the incidence of cardiotoxicity and lower the potential for local necrosis, but infusion reactions, sometimes severe, may occur. Hand–foot syndrome (painful, macular reddening skin eruptions) occurs commonly with liposomal doxorubicin and may be dose limiting. Daunorubicin also has general properties similar to those of doxorubicin. A liposomal formulation (DSPC/Chol, 2:1, size 45 nm) for i.v. use is licensed for patients with AIDS-related Kaposi's sarcoma [207, 208]. The plasma pharmacokinetics of DaunoXome differs significantly from the results reported for free daunorubicin hydrochloride. DaunoXome has a small steady-state volume of distribution of 6.4 L (probably because it is confined to vascular fluid volume) and clearance of 17 mL/min. These differences in the volume of distribution and clearance result in a higher daunorubicin exposure [in terms of plasma "Area on the curve" or "bioavailabity" (AUC)] from DaunoXome than with free daunorubicin hydrochloride. The apparent elimination half-life of DaunoXome (daunorubicin citrate liposome injection) is 4.4 h, far shorter than that of daunorubicin, and probably represents a distribution half-life. Although preclinical biodistribution data in animals suggest that DaunoXome crosses the normal blood–brain barrier, it is unknown whether DaunoXome crosses the blood–brain barrier in humans [207].

In addition, TAS-103 (a novel quinoline derivative, topoisomerase inhibitor) incorporated in DPPC/Chol (2:1) liposomes (size < 80 nm) enhanced the survival time of mice with Lewis lung carcinoma to 42 days in comparison to the 38.6 days of those treated with free TAS-103 [209]. The increases in lifetime were 45 and 58% for the free TAS-103 and liposomal TAS-103, respectively.

In another study, the antibiotic cefoxitine was incorporated in DMPC/Chol (2:1) liposomes prepared using the reverse-phase evaporation technique in order to

increase the efficacy of the drug characterized by a short half-life and poor intracellular diffusion [210]. It was shown that the cefoxitin levels achieved in liver and spleen 5 h postinjection were 6- and 16-fold higher than those observed after administration of free antibiotic. Also, the elimination rate through the kidney was slower.

***Intraperitoneal Administration*** Intraperitoneal administration has the biological and pharmacological advantage of creating direct exposure of the tumor, infection, or inflammation to the therapeutic agent. This drug delivery method increases the dose intensity within the peritoneal cavity [211–213]. Intraperitoneal administration of liposomal formulations of anticancer drugs is preferred to the i.v. one, due to the higher drug accumulation in the tumors and lower drug plasma concentration minimizing drug toxicity [214–216].

Size, liposomal composition, charge, drug density in the liposomal membrane, and preparation method are some of the important parameters which need to be considered carefully to design an efficient DDS. Sadzuka et al. assessed DOX-incorporating liposomes made by either DMPC or DSPC of a variety of sizes (150, 600, and 4000 nm) and surface charge (positive and negative) on the therapy of solid tumors and peritoneal dissemination in Ehrilch ascites carcinoma-bearing mice [214]. When using small negative liposomes, lipid composition did not affect the clearance or stability of liposomes in the abdominal cavity. However, for neutral liposomes, DSPC ones were found more effective for the treatment of the solid tumor due to the higher stability of those liposomes in comparison to DMPC ones. Thus DSPC exhibited longer plasma circulation. As for the effect of surface charge, the positive vesicles were cleared faster from the abdominal cavity until 1 h postinjection and then showed a slower clearance rate until 48 h, in opposition to the negative ones. Larger particles were found in abundance in the peritoneal cavity and stayed longer there, inducing toxicity due to liposomal membrane disruption and release of the anticancer drug. Overall, it was concluded that the larger liposomes were effective against peritoneal dissemination and the smaller ones against the solid tumor.

The same author evaluated the method of preparation by using DOX-encapsulating liposomes on the peritoneal dissemination of tumor in Ehrlich ascites carcinoma-bearing mice [217]. The liposomal carriers were made either with the method of Bangham et al. [1] (BLDOX), the pH gradient [144] (PLDOX), or the gelation method (GLDOX) [147]. It was shown that survival in the BLDOX group was significantly prolonged compared to that in the DOXsol (DOX as solution) group, while there was no effect on survival of the GLDOX group. BLDOX liposomes appeared to be less stable and released DOX in a higher degree than the other formulations. The latter seems to be of high importance due to increased DOX level in the abdominal cavity and enhancement of drug efficiency for the local therapy.

Also, positive outcome was achieved after i.p. administration of the liposomal formulation of an L-dopa prodrug derivative to rats [218]. It was shown that the level of dopamine in rat striatum was 2.5-fold higher to what was obtained after i.p. administration of L-dopa or the free prodrug itself.

***Subcutaneous Administration*** Liposomes given s.c. aim to target the lymphatic system for imaging, distribution of therapeutic agents, or vaccination [219, 220]. According to Oussoren and Storm, the determining factors influencing lymphatic

absorption are liposome size and site of injection [219]. Liposome charge, composition, or PEG coating does not have a significant effect on the fate of the liposome trip in the lymphatic system.

Liposomes injected s.c. that do not enter the bloodstream either enter the lymphatic capillaries or stay at the site of injection. In the first case, 1–2% of the injected liposomal formulation is captured by the lymphatic nodes 12 h postinjection. However, this depends on liposome size. Neutral vesicles smaller than 100 nm pass through the interstitium and then to the lymphatics a lot easier than the larger particles. Drug carriers remaining at the site of injection will release the entrapped molecule. Often, 40% of the injected dose of small liposomes (about 70 nm) is retained at the injection site. Therefore, liposome surface modification was attempted using nonspecific human antibodies and saccharides. Only saccharide-modified liposomes enhanced absorption from the injection site and enhanced lymph node uptake was in comparison to control liposomes [221]. Also, the specific site of injection is very important and species dependent. Taking the rat as animal model, s.c. injection in the dorsal foot or the footpad results in higher liposome uptake by the lymph nodes, in contrast to the flank as an injection point.

As mentioned earlier, the s.c. route for delivery of anticancer agents could prevent the metastatic spread of tumors that occurs often through the lymphatic system. However, a number of limiting factors, such as incomplete absorption of drug-loaded liposomes, which would increase, for example, the toxicity of the released drug at the surrounding tissue and the development of tumors in the regional lymph nodes could limit the therapeutic potential of liposomes.

In addition, for imaging studies, only liposome-encapsulated gadolinium was used successfully [222, 223]. In a more recent study, electron spin resonance (ESR) was applied successfully to investigate the integrity of MLV and the possibility of a depot effect after the s.c. injection in mice [224].

Also, Gregoriadis et al. evaluated the type and degree of immune response after s.c. injection of ovalbumin (OVA)–encoding plasmid DNA (2.5 or 10 μg) either alone or in liposomes, in mice [220]. Anti-OVA serum antibody titers were detected in animals immunized with 10 μg of liposomal DNA (after a single dose of antigen) and with both 2.5 and 10 μg of liposomal DNA (after two doses of antigen) [225]. However, the anti-OVA response was not detected using the DNA alone.

Similarly, significantly higher humoral responses were obtained after s.c. administration of either a lipid and/or a nonionic-based vesicle-entrapped plasmid for the nucleoprotein of $H_3N_2$ influenza virus in comparison to the naked DNA alone [226].

### 3.2 Long-Circulating or PEGylated Liposomes

The liposome biodistribution profile changes significantly when the vesicle surface [227–229] is coated with polymers, usually PEG. Longer blood circulation, lower liver uptake, and higher accumulation in tumors have been reported. The presence of the hydrophilic groups of PEG on the liposome surface provides electrostatic and steric repulsion between PEG-grafted liposomes. PEG molecules neutralize the surface charge of vesicles and thus prevent their opsonization. Also, liposome opsonization is reduced due to inability of opsonins to bind to hydrophilic surfaces. Moreover, the thickness of the PEG layer influences the interaction of the liposomes

with the macrophages. The thickness of the PEG layer depends on the PEG molecular weight and the amount (%) incorporated in the liposomal composition. For example (Figure 9), the fast clearance of positively charged stearylamine liposomes can be reversed by attachment of 6% mol of PEG with molecular weight of 750 or 5000 kDa [230]. In the case of phosphatidic acid–containing liposomes, only PEG 5000 can prolong blood circulation while phosphatidylserine-containing liposomes are eliminated fast due to the insufficient effect of 6% PEG 750 or 5000.

However, an optimum level of PEGylation (PEG 2000 kDa) was estimated for PC (1.85 mol %)/Chol (1 mol %) liposomes as to the effect on blood circulation [231]. As reported, after 5 mol % of PEG incorporation the accumulation in the liver was significantly decreased, while the minimum uptake by the spleen was achieved with 9.6 mol % of PEG insertion. The same authors showed that as the PEG amount grafted increased, the liposome accumulation in the heart increased. But above a 9.6 mol % of PEG, the circulation time was slightly decreased in blood and was increased in the liver and spleen. The uptake of liposomes by RES was even higher when 13.7% of PEG was present on the vesicle surface. As shown by another group, the optimum PEG amount required for liposome stabilization and prolonged half-life was 5–10 mol % DSPE–PEG 2000 [232]. Higher amounts than that led to disruption of the vesicles.

***Intravenous Administration*** The effect of PEG on liposome biodistribution presented in the previous paragraph is for i.v. administration of liposomes. Several recent examples for the effect of PEG coating of liposomal drug formulations on the biodistribution (and thus pharmacological outcome) for different drugs follow. DOX concentrations were estimated in various organs after i.v. administration of DOX-loaded liposomes (PC/Chol/PEG2000, molar ratio 55:40:5) in xenograft tumor-bearing mice [233]. Obviously, the encapsulation of DOX in conventional or PEGylated liposomes reduced the drug concentration in liver, heart, kidney, and stomach compared to the drug solution and prolonged the circulation half-life to 46.09 h, in contrast to 26.04 and 23.72 of conventional and free DOX, respectively. However, PEGylated DOX showed higher antitumor activity in comparison to that entrapped in conventional liposomes. In comparison with free DOX, the inhibition rate of both liposomal formulations was higher. Doxil is a commercially available formulation of DOX entrapped in HSPC/Chol/mPEG liposomes [234]. Thereafter, PEGylated liposomes are in common use in many applications. Covalent attachment of specific molecules (folate) at the free end of PEG molecules results in liposomes able to be recognized by specific receptors (folate receptors on cancer cells). There are a huge number of research papers on the active targeting of modified PEGylated liposomes [235–237] with very promising results.

***Intraperitoneal Administration*** The potential of PEGylated liposomes administered intraperitoneally has been evaluated for cancers located in the peritoneal cavity. For example, Syrigos et al. [238] studied the biodistribution of indium ($^{111}$In)–labeled PEGylated liposomes [Hydrogenated soya PC (HSPC)/Chol/PEG–DSPE] compared to free $^{111}$In via either i.p. or i.v. route to non-tumor bearing mice. The AUC of In-PEG liposomes was 74-fold higher than that of free indium. The relative ratio of the AUCs (RR-AUCs) for i.p. versus i.v. administration for peritoneum was 1.36 [423.6 vs. 311.3% individual dose (ID) g/h]. The blood AUC values for i.p. and

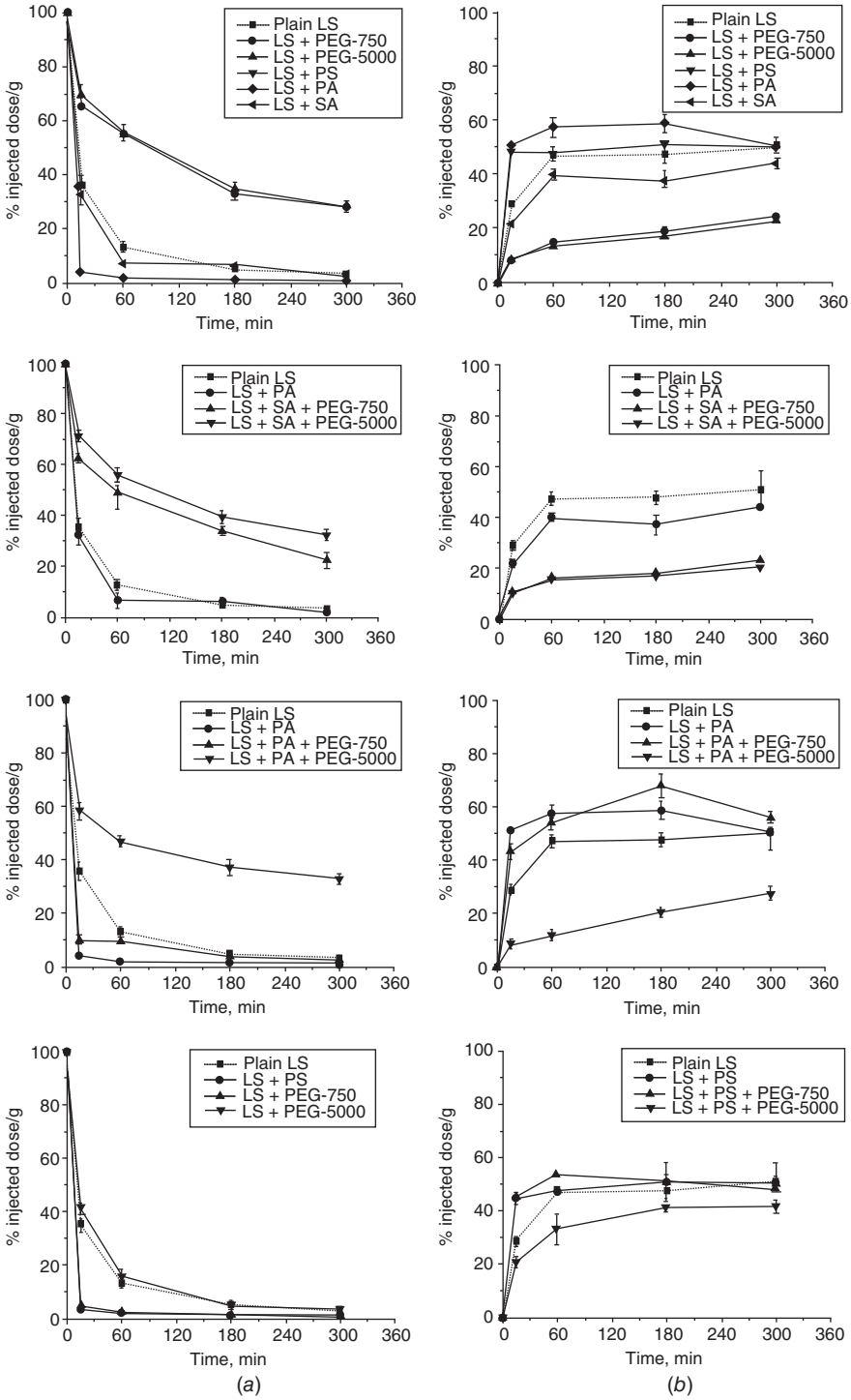


**FIGURE 9** (*a*) Clearance from circulation and (*b*) accumulation in liver and spleen of liposomes composed of lecithin (LS) mixed or not with charged lipids (PS = phosphaditylserine, PA = phosphatidic acid, SA = stearylamine) and bearing or not, a surface coating with PEG molecules after IV administration (PEG-750 and PEG-5000 = polyethylene glycol with molecular weights 750 and 5000). (Reproduced from ref. 230 with permission of Elsevier.)

i.v. administration were essentially the same (RR-AUC 1.03; 453.7 vs. 439.2% ID g/h) 18 h postinjection despite the delayed absorption of the liposomes from the peritoneal cavity. However, the relevant values for organs in the peritoneum were higher in case of i.p. administration. An increase in the range 1.2–5.1 was seen for organs such as stomach, pancreas, ileum, colon, gallbladder, ovary, and adrenal glands. This is an advantage compared to the i.v. administration because the drug can target both the primary site and any peritoneal deposits. The encapsulation of doxorubicin and cis-platin (small molecules with high toxicity and short half-lives) in PEGylated liposomes might increase the retention from the peritoneal cavity and reduce the drug toxicity.

Another study points out the vesicle size rather than the presence of PEG as a more determining factor to successfully tackle peritoneal cancers [239]. According to this study, the synergistic effect of Doxil after coadministration of PEG–SUV–interleukin-2 (IL-2) or MLV–IL-2 via either the i.p. or i.v. route to mice bearing M109 lung adenocarcinoma was studied. The cancer was inoculated i.p., resulting in multiple i.p. masses. Small PEGylated liposomes as vehicles for IL-2 for systemic treatment of metastatic lung cancer boosted the antitumor effect of Doxil to the same level achieved with soluble IL-2. In contrast, in the regional model, the most effective combination was Doxil with MLV–IL-2 liposomes. This is attributed to the retention and slow release of IL-2 in the peritoneal cavity due to the inability of MLVs to enter the circulatory system or the draining lymph vessels, whereas IL-2 in small liposomes or in soluble form escapes rapidly from the peritoneal cavity. Another possibility could be the enhanced immunostimulation results from the uptake of MLV–IL-2 by peritoneal macrophages as opposed to the stealth properties of PEGylated SUV–IL-2.

***Subcutaneous Administration*** As for the use of PEGylated liposomes subcutaneously, there is not much reported. In one paper the influence of the administration either s.c. or i.m. of mitoxantrone-loaded liposomes was studied. It was reported that mitoxantrone showed reduced irritation when the formulation was administered i.m. rather than s.c. However, when PEG was incorporated on the liposome surface, there was no apparent protective effect of the liposomes [240].

### 3.3 Other Routes of Administration

In general, the administration route plays an important role on the impact of the therapeutic treatment, and it is chosen according to the kind or purpose of the treatment (local or systemic), toxicity, and accessibility of the diseased area. In this section specific characteristics of other routes of administration that are currently receiving attention will be emphasized and the most recent developments with respect to liposomal drug applications will be presented.

***Pulmonary Drug Delivery*** Pulmonary epithelium offers many advantages for drug delivery due to easy access and the large surface area provided by alveoli [241]. Also, macromolecules can penetrate the lungs much easier and faster than other noninvasive routes avoiding the first-pass meabolism; therefore, many promising applications are being considered for the delivery of proteins and peptides. For a drug or drug delivery system to reach the lungs successfully, it has to be aerosolized with optimum aerodynamic particle diameter between 1 and 3 μm. The latter, in

combination with how the patient inhales determines if the drug particles deposit primarily in the conducting ways or in the alveoli.

Pulmonary delivery of liposomes has focused on the treatment of asthma, infectious diseases, genetic diseases (cystic fibrosis), and lung injury and lately on gene therapy.

Special attention has been paid on the physical characterization of liposome aerosols [242–244], including dry powder formulations [245], and cationic liposome DNA complexes [246, 247]. Corticosteroid therapy using liposomal formulations has focused on the development of aerosols containing beclomethasone dipropionate, triamcinolone acetonide, and triamcinolone acetonide phosphates [248–251]. Aminoglycosides have been considered as good candidates for pulmonary delivery because of their potency and their ability to directly target the lungs. When amikacin was encapsulated in liposomes [fluid or rigid state (chol containing)] and administered in sheep, the drug mean residence time (MRT) increased 5 times compared to the instilled solution (rigid liposomes gave 2 times higher MRT compared to fluid ones) [252]. In a similar study, prolonged retention of liposome-encapsulated tobramycin was reported after administration in *Pseudomonas aeruginosa*–infected rat lungs [253]. Liposomal formulation of antioxidants has also been investigated for pulmonary delivery [254]. Liposomes containing α-tocopherol prolonged the residence time of radioactively labeled glutathione [255]. Paraquat poisoning, which causes extended damage to lung tissues, was attenuated after pretreatment with α-tocopherol-containing liposomes [256]. Liposomes containing CuZn/superoxide dismutase and catalase were found to protect the lungs of premature rabbits when exposed to hyperoxia [257].

Drugs for the treatment of infections such as aspergillosis, tuberculosis, and anticancer therapy have been formulated in liposomes and tested in vivo by administration via the respiratory system [258–260]. More specifically, rifampicin used to treat pulmonary tuberculosis associated with AIDS has been incorporated in MLV liposomes (PC/Chol, 7:3 molar ratio) with the aim of increasiing its efficacy to macrophages and reduce side effects [258]. The rifampicin retention was found higher in PC/Chol/DCP (7:3:0.1) liposomes tested due to electrostatic interaction between DCP and the drug. Both liposomal formulations showed greater accumulation in the lungs when compared to the controls. In the case of free drug and after 0.5 h postadministration, only 39.12% of the administered dose was retained in the lungs and 29.84% of it was found in the serum. No drug was estimated in the lungs 24 h later. PC/Chol MLVs demonstrated higher lung accumulation (49.03%) after 0.5 h, but the overall distribution pattern was not much different to that of the free drug in solution with no drug estimation at the 24-h time point. It seems that they were rapidly passed to the systemic circulation and to the RES organs (liver and spleen). The negatively charged MLVs showed even higher drug accumulation (53.86% of the administered dose) while 4.14% of it was still present after 24 h. This is attributed to the interaction of the negatively charged liposomes with the scavenger receptors on alveolar macrophages [261]. Also, it has to be mentioned here that the ligands maleylated bovine serum albumin (MBSA) and *O*-steroyl amylopectin (O-SAP) were incorporated on the liposome surface because they are recognized by the scavenger receptors of the macrophages. The lung accumulation levels were 61.49 and 65.14% of the administered dose, respectively, after 0.5 h, while after 24 h they were 8.12 and 10.75%, respectively. The relative lung retention of the various formulations after 6 h of administration was 1.3 times for PC/Chol MLVs, 3.4 times

for PC/Chol/DCP MLVs, 4.53 times for MBSA-PC/Chol, and 4.76 for O-SAP–PC/Chol in comparison to plain drug solution administered by aerosolization.

The same group (Vyas et al.) studied the impact of amphotericin B entrapped in liposomes (MLVs: PC/Chol, 7:3 molar ratio) in the absence or presence of ligands [*O*-palmitoylated mannan (OPM) or *O*-palmitoylated pullulan (OPP)] on the vesicle surface as potential use for the treatment of aspergillosis [259]. Optimized formulation was the one of PC/Chol 7:3 molar ratio with the production of spherical MLVs of approximate mean vesicle size 2.56 μm and maximum entrapment efficiency 78.2%. Again the lung uptake of the ligand-appended liposomes was higher compared to the plain liposomes. The lung accumulation levels of OPM- and OPP-coated liposomes were 58.12 and 55.02% of the administered dose 5 h postadministration, respectively, while 24 h later the relevant lung retentions were 11.23 and 9.86%, respectively.

Methotrexate (MTX) (a folic acid antagonist) was entrapped into liposomes (PC/PI/Chol 2:1:1 molar ratio) so as to reduce nephrotoxicity and investigate the pharmacokinetics of the liposomal MTX [260]. Indeed, the latter showed increased MTX retention in the lungs while the biodistribution in spleen and kidney was less than that obtained with free MTX. Similar results were obtained by liposomalization of various anticancer drugs elsewhere [262–264]. For example, cytosine arabinoside was administered intratracheally to rats in the free or the liposomal form [262, 263]. Liposomal drug was effective into the lung but not other tissues, contrary to the free drug. Moreover, the antitumor properties of the anticancer 9-nitrocamptothecin (9NC) after its liposomalization were tested in three different human cancers xenografted s.c. in mice as well as murine melanoma and human osteosarcoma pulmonary metastases [264]. The liposomal form of anticancer drug inhibited the growth of subcutaneous tumors and metastatic pulmonary cancers given via the respiratory system. Intramuscularly administered liposomal drug exhibited some anticancer activity, but that achieved using the aerosol was superior. Thus, the liposomal 9NC aerosol was proved to be of high potential for the treatment of cancers throughout the body.

Recently the gene therapy of pulmonary diseases using liposomal formulations has attracted a lot of attention. Transfection of the lungs of animals with aerosolized cationic liposome–DNA complexes has been attempted [265, 266]. However, the transfection efficiency was rather low despite the large amounts of DNA used in these studies. Cationic liposomes from 2, 3-dioleyloxy-N-[2(sperminecarboxamido) ethyl]-N, N-dimethyl-1-propanaminium trifluoracetate (DOSPA), (±) N-(2-hydroxyethyl)-N, N-dimethyl-2, 3-bis(tetradecyloxy)-1-propanaminium bromide (DMRIE) mixed with DOPE, and DNA at fixed ratios of DNA/lipid 1:4 for DMRIE and 1:3 for DOSPA were tested with two different types of jet aerosols (Aerotech II and Puritan-Bennett 1600) [246]. The decrease in transfection activity was gradual with Puritan-Bennett 1600, in contrast with Aerotech II, which rapidly lost transfection efficiency. That was attributed to the increased throughput of the Aerotech II, resulting in more frequent cycling and therefore damage of the complex.

The impact of the zeta potential of the formulation was emphasized by Eastman et al. [247]. The maximal aerosol transfer efficiency of cationic lipid/DNA complexes was achieved in the presence of a salt concentration of 25 m*M*. The authors attributed that to the fact that the formulation kept its zeta potential between –40 and –50 mV. As a closing remark, one of the major problems and challenges of aerosol

delivery is the duration to deliver therapeutic doses of DNA to the lungs. This possibly could be overcome with dry powders avoiding volume limitations of aqueous dispersions.

***Oral Delivery*** The destructive effects of the conditions in the gastrointestinal tract (GIT), especially due to the presence of bile salts, are known and well established [267, 268]. The lipid composition of liposomes determines to a large extent the possibility of remaining intact under such conditions [58]. Quadachi et al. attempted to provoke IgA response from the M cells of Peyer's patches after oral administration of OVA-loaded MLV liposomes made by either soya PC or DSPC to a model of hypersensitivity to OVA Balb/C mouse [269]. Clearly, DSPC MLVs were much more stable as demonstrated in in vitro stability studies in simulating GIT media. However, liposome incorporation of OVA did not cause any significant impact on the reduction of hypersensitivity to that obtained with the free allergen. Surprisingly, the empty liposomes show some immunoresponse which was attributed to nonspecific stimulation.

The biodistribution of novel liposomal-like spherical carriers called Spherulites, consisting of PC/Chol/polyoxyethylene alcohol (43:4:3 w/w/w), was shown to be more promising [270]. These were prepared by shearing the phospholipid bilayer and labeling with $^{111}$In. Their integrity was demonstrated by an increase in radioactivity in the blood 1 h after oral administration to fasted rats, while no increase was seen for free label.

***Skin Delivery*** Dermal delivery of phospholipid-based vesicles first appeared in the literature in the early 1980s [271, 272]. Liposomes can play a dual role after their application to the skin: retention (and perhaps protection) of the active compound across the stratum corneum and acting as a penetration enhancer. The composition, size, and vesicle surface charge are parameters that can influence the transport rate of drugs contained in liposomal formulations across the skin [273]. Regarding their composition, liquid-state liposomes resulted in higher skin permeation rates compared with gel–liquid liposomes for progesterone and Triamcinolone (TRMA) [274, 275]. Moreover skin lipid liposomes have provided higher drug disposition in the deeper layers of the skin for corticosteroids and acyclovir [276, 277]. Neverthless, in another study [278] higher amounts of acyclovir in the skin were delivered from conventional lipid liposomes compared with liposomes containing skin-based lipids. These studies emphasize that a careful design is needed to define optimal compositions, depending also on the specific objective which may be either dermal (topical) delivery of the liposome-encapsulated drug (which is easier since only increased retention of the drug in the skin is desired) or transdermal delivery (systemic absorption of the drug), which is more complicated, and in general deeper penetration of the liposome carrier in the skin is required. What is also important as regards the final outcome (especially when transdermal delivery is the objective) is the type of final formulation with respect to the conditions applying, occlusive or nonocclusive.

Amphotericin B encapsulated in charged liposomes demonstrated 10-fold higher transport rates across the skin compared with neutral liposomes [279]. In a similar manner, the retention of acyclovir from positively charged vesicles was much higher compared with other formulations [277]. This could be attributed to the attraction

between the positively charged liposomes and the negative charge of the skin [277]. The pore size of the skin is approximately 0.3 nm; however it can be opened up to 40 nm without significant damage to its structure. To some extent, vesicles can transport across the skin via the follicular and transcellular routes. Therefore the size of the vesicles used for transdermal delivery is a crucial parameter for the overall performance of the formulation [280–282]. Several studies have also emphasized the impact of the size of the liposomes on the transport rate of active compounds across the skin [277, 280, 283, 284]. Biologically active macromolecules, including superoxide dismutase [285] and interferon-γ [286], have been successfully applied to skin in liposomal formulations. Generally it is believed that the main pathway for transdermal delivery of active compounds is either intercellular or paracellular. However, the appendage transport (e.g., follicular route) has attracted a lot of interest lately. The pilocebaceous units (hair follicle, hair shaft, sebaceous glands) can be used for drug targeting. Combinations of liposomes with DNA [287, 288] and monoclonal antibodies [289] have demonstrated that liposome composition, hair structure, and hair cycle play significant roles in the transfection of human hair follicles. Two different mechanisms have been suggested for the incorporation of active compounds to the hair shaft: (1) direct permeation of the vesicular formulations with the active compound to the hair shaft and (2) incorporation of the active compounds in the follicular matrix cells and then into the hair shaft as the matrix cells develop and differentiate into new hair shafts [290].

Recently, for the transdermal delivery of drugs using carrier systems, attention has been focused on the development of *transformable* [284, 285] or *elastic* vesicles [12]. These vesicles are liposomes that contain surfactants or in general "edge activators" in addition to phospholipids in their lipid membranes (Figure 10), a fact that

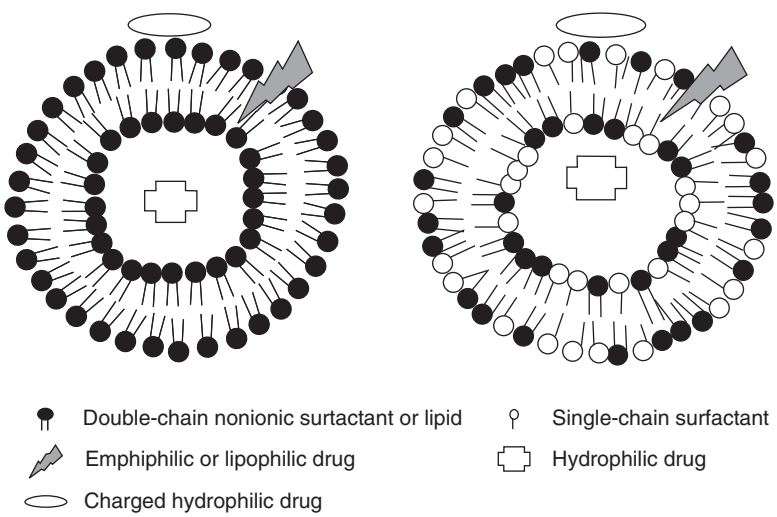


**FIGURE 10** Conventional liposomes/elastic vesicles. Charged hydrophilic, amphiphilic, and lipophilic drug molecules can be associated with the bilayers of the vesicles, whereas hydrophilic substances can also be entrapped in the vesicles. Rigid vesicles consist of double-chain nonionic surfactants or lipids in the presence or absence of cholesterol (left image). Elastic vesicles consist of double-chain surfactants or lipids and an edge activator. The edge is often a single-chain surfactant (right image). (Reprinted from ref. 273 with permission of Elsevier.)

increases their elasticity (ability to be deformed, without being disrupted, after applying pressure on them). Modified liposomes called ethosomes (containing alcohol) have shown increased skin permeability. Because of their structure, ethosomes are able to encapsulate and deliver through the skin highly lipophilic molecules such as cannabinoids, testosterone, and minoxidil as well as cationic drugs such as propranolol and trihexyphenidil or even plasmids and insulin [291]. Although the mechanism of the increased transdermal delivery of drug molecules still remains a controversial issue, mainly with respect to the depth at which intact vesicles can travel in the skin, many interesting results are being generated. However, due to the limited permeability of the skin membrane, physical enhancement mechanisms, including iontophoresis, electroporation, and ultrasound, have been used in combination with chemical enhancers (liposomes) to increase skin permeability. One of the methods, application of an electric current to the skin, has been shown to promote the transdermal transport of drugs by an additional driving force, namely, an electrical potential gradient across the skin [292]. Transdermal iontophoretic transport of a liposomal formulation across human cadaver skin was first reported for [Leu5] enkephalin [293]. Liposomes could penetrate into the skin. Enkephalin, when delivered iontophoretically from liposomes carrying positive or negative charge on their surface, was found to be the same or less than that of the controls; however, the degradation of enkephalin was less in liposome formulations as compared to controls, demonstrating that liposomes can protect peptides from the proteolytic environment of skin. However, when enoxacin was encapsulated in different liposome formulations and was transported (electrically assisted) across the skin, the drug transport results showed that the permeability of enoxacin released from liposomes was higher compared to that of free drug [294]. In vitro transdermal iontophoretic delivery of estradiol from ultradeformable liposomes, saturated aqueous solution [295], and conventional liposomes [296] has demonstrated the superiority of ultradeformable liposomes (Transfersomes) due to their lipid composition. Liposomal formulations of β blockers were iontophoresced in vivo to hairless rats [297]. Skin irritation was significantly reduced when a liposomal formulation of the propranolol base was used rather than the base itself, emphasizing another important role liposomes could play. When adriamycin was delivered via the hair follicles using various liposomes and iontophoresis combined with application of ionic liposomes, higher transport rates were obtained with the latter, emphasizing their synergistic effect [298]. Moreover skin electroporation was applied to enhance gene transfer into subcutaneous MC2 murine breast tumor skin in combination with cationic liposomes demonstrating significant transfection improvement [299]. However, when electoporation combined with estradiol-loaded liposomes were applied to skin, the estradiol skin penetration was not affected significantly. That was attributed to the antienhancer or retardant effect of liposomes [300]. In a mechanistic study, anionic phospholipids were found to enhance the transdermal transport of molecules by electroporation compared to cationic or neutral phospholipids, offering new insights to design better enhancers for transdermal drug and vaccine delivery [301]. Finally higher transport across the skin obtained after combined application of ultrasound and liposomal formulations of diclophenac demonstrated a synergistic effect [302].

***Ocular Delivery*** Ocular drug delivery has evolved into a great challenge and a subject of interest for many scientists with different backgrounds, including medical,

clinical, pharmaceutical, physical, chemical, biochemical, and toxicological sciences. For ailment of eye diseases, topical administration is preferred over systemic in order to avoid systemic toxicity, for rapid onset of action, and for decreasing the required dose.

The main route for intraocular absorption is across the cornea [303]. In terms of drug delivery, the cornea presents an effective barrier to the absorption of both hydrophilic and lipophilic compounds. Actually, the main constraints in topical ocular delivery are (i) poor ocular retention of conventional dosage forms [304] and (ii) poor corneal absorption. Various approaches have been developed to increase the bioavailability and duration of therapeutic action of ocular drugs. One such approach is based on the use of drug delivery systems [305, 306], which provide controlled and continuous delivery of drugs and can also provide improved (increased) residence time of the drug at the delivery site. Recently, intravitreal drug injection has evolved into a preferred administration method for therapy of disorders in the posterior segment of the eye [305]. The procedure is associated with a high risk of complications, particularly when frequent, repeated injections are required. Thus, sustained-release technologies are being proposed, and the benefits of using colloidal carriers in intravitreal injections are currently under investigation for posterior drug delivery.

Between the different types of particulate drug delivery systems, liposomes offer additional advantages for ophthalmic delivery, since they are completely biodegradable and relatively nontoxic and thus are well tolerated by the eye [305, 306]. Indeed, when using other types of colloidal systems, for example, nanoparticles consisting of polyalkyl cyanoacrylate, inflammation and damage of the corneal epithelium have been reported [307–309]. Another potential advantage of liposomes is their ability to come in intimate contact with the corneal and conjunctival surfaces. This results in increased probability of ocular drug absorption [310, 311].

The potential of liposomes in topical ocular drug delivery was first exploited in the 1980s by a number of research groups [303, 310–315]. As an example, higher levels of inulin were found in the cornea when it was encapsulated in liposomes as compared to its aqueous solution [301, 314], and this was attributed to the physical adsorption of lipid vesicles onto the epithelial surface of the membrane [315, 316]. More recently a number of liposomal applications for ocular delivery have been under investigation [305, 306] for anterior as well as posterior segment administration, as outlined in Table 3. Indeed, a large number of ophthalmic drugs used in cases of ocular surface disorders (such as dry eye syndrome) [317], keratitis and uveitis [318–327], and keratoplasty [328–331] have been studied in liposomal form, and in most cases the results were promising in terms of drug penetration and retention in the various ocular tissues (cornea, sclera, retina, and choroids), following subconjunctival administration. In some cases, detectable levels of drugs were found in ocular tissues up to 7 days after administration [305, 306].

As mentioned above, the ability to adsorb to the cornea and an optimal drug release rate have been defined as the two liposomal attributes most responsible for increasing bioavailability after topical ocular administration. A number of factors, including drug encapsulation efficiency, liposome size and charge, distribution of the drug within liposomes, stability of liposomes in the conjunctival sac and ocular tissues, their retention in the conjunctival sac, and most importantly their affinity toward the corneal surface and the rate of release of the encapsulated drug, have

**TABLE 3 Experimental Liposomal Preparations of Drugs for Anterior and Posterior Segment Administration**

| Drug Class | Anterior Segment Drugs | Posterior Segment Drugs |
|---|---|---|
| Antibiotics | Gentamicin<br>Norfloxacin<br>Tobramycin | Clindamycin<br>Gentamicin<br>Penicillin |
| Antifungals | Amphotericin B | Amphotericin B |
| Antivirals | Acyclovir<br>Idoxuridine | Ganciclovir<br>Trifluorothymidine (Trifluridine) |
| Steroids | Dexamethasone | |
| Immunosuppressives | Cyclosporine<br>FK506 (Tacrolimus) | Cyclosporine |
| Antimetabolites | 5-Fluorouracil (5-FU)<br>5-Fluoroorotate | 5-Fluorouracil (5-FU)<br>5-Fluorouridine (5-FUR)<br>Bleomycin<br>Cytosine arabinoside (Cytarabine)<br>Daunomycin<br>Daunorubicin<br>Etoposide (VP-16) |
| Platelet-aggregating agents | Adenosine diphosphate | Adenosine diphosphate |
| Photosensitive cytotoxic agents | | Verteporfin (BPD-MA) |
| Miscellaneous | Dichloromethylene diphosphonate (Clodronate)<br>Disulfiram<br>Immunoglobulins | Dichloromethylene diphosphonate (Clodronate) |

*Source*: S. Ebrahim, G. Peyman, and P. J. Lee, *Survey of Ophthalmology*, 50, 167–182, 2005.

been found to influence the effectiveness of liposomes in topical ocular drug delivery [310, 323, 332–335, 337]. Indeed, liposomal manipulation to increase corneal adherence has met with some success [310].

Positively charged liposomes seem to be preferentially captured at the negatively charged corneal surface as compared with neutral or negatively charged liposomes. Aditionally, cationic vehicles are expected to slow down drug elimination by the lacrymal flow both by increasing solution viscosity and by interacting with the negative charges of the mucus [334, 335]. Indeed, positively charged phospholipids yielded superior retention of liposomes at the corneal surface in rabbits [336]. Schaeffer et al. [310] worked with indoxole and penicillin G and reported that liposome uptake by the cornea is greatest for positively charged liposomes, less for negatively charged liposomes, and least for neutral liposomes. Positively charged unilamellar liposomes enhanced transcorneal flux of penicillin G across isolated rabbit cornea more than fourfold. Similar results were also obtained by others [336, 323]. By observing the morphology of corneal surface treated with liposomes, it was suggested that positively charged liposomes formed a completely coated layer on the corneal surface [323]. These liposomes bind intimately on the corneal surface,

leading to an increase in residence time and therefore to an increase in corneal absorption time.

With respect to vesicle size, larger particles are more likely to be entrapped under the eyelids or in the inner canthus and can thus remain in contact with the corneal and conjunctival epithelia for extended periods. Indeed, larger liposomes have been found to resist drainage at the inner canthus and are more bioavailable at the ocular surface [337]. However, for patient comfort, it is considered that solid particles intended for ophthalmic use should not exceed 5–10 μm diameter [338].

Assessment of ocular irritability of neutral or positively charged liposomes by the Draize test, histological examination, and the rabbit blinking test has also been reported in the literature [339]. The mean total score (MTS) of the Draize test was found to show a slight increase immediately following instillation of liposome preparations. However, it did not exceed the "practically nonirritating level," and no corneal histological alteration was observed by optical microscopy. Neutral liposome preparations were confirmed to be nonirritating; however, positively charged liposomes may cause initial pain or unpleasantness following instillation. Thus, althouth positive charge helps in improving the contact time with the cornea, at the same time it can lead to irritation. Additionally, the release rate of the drug is found to be more in neutral liposomes, while increased liposome size restricts solution drainage, thus prolonging contact time of the drug, but it can be increased within the limits of not inducing any irritancy. From all the above it is understood that in each case liposome properties have to be adjusted for best in vivo therapeutics results. As mentioned above, liposome stability is another important factor. Barber and Shek reported that increasing the cholesterol content of the liposomal membrane decreased the rate of tear-induced release of its contents [340].

The use of bioadhesive polymers (e.g., a polyacrylic acid, chitosan, hyaluronic acid) to prolong the residence time of an ocular preparation in the precorneal region, due to increased formulation viscosity, is another approach which can further improve liposomal drug delivery. In this respect "collasomes" (liposomes coupled to collagen matrices) increased the bioadhesive ability of liposomes, were well tolerated, and could be instilled safely and effectively by patients in the same fashion as ointments or drops [341]. It was also demonstrated that liposomes coated with collagen layer bound to cell monolayer with higher affinity [342]. Other approaches were used to increase the contact time of ocular liposomes, such as the case in which prolonged retention of liposomal suspension of oligonucleotide was achieved by dispersing liposomes within an in situ gel-forming medium [343–345]. Novel measures to further enhance adsorption of liposomes and increase penetration of the cornea included application of a natural lectin which promoted binding of ganglioside-containing liposomes to the cornea [346].

Liposomes can also be used as promising dosage forms for topical administration of immunosuppressive compounds for the treatment of ocular immune-mediated diseases [327]. Indeed, it was found that liposomes containing immunosuppressive compound FK506 were effective in delivering significantly higher drug concentrations to all ocular tissues and particularly aqueous humor and vitreous humor as compared to the oil formulation of the agent. Further, liposomes can be used to protect drug molecules from attack of metabolic enzymes present at the tear/corneal epithelium interface, as demonstrated for the *O*-palmitoyl prodrug of tilisolol [347].

As mentioned above, intravitreal injection of drugs should be used in many cases to achieve therapeutic intravitreal drug levels. This is especially true for cases of viral retinitis, such as cytomegalovirus (CMV) retinitis and acute retinal necrosis (ARN) which require intravitreal injection of antivirals, or for the treatment of bacterial and fungal endophthalmitis or proliferative vitreoretinopathy [305]. It still remains a controversial issue whether liposomes can reach the retina after intravitreal injections and which vesicle physicochemical characteristics should be preferred for such formulations.

In addition to conventional drugs, liposomes were also used for intravitreal administration of oligonucleotides in order to treat ocular viral infections such as herpes simplex virus or CMV [344]. Antisense oligonucleotides are poorly stable in biological fluids and their intracellular penetration is limited. Hence a system that is able to permit a protection of oligonucleotides against degradation and their slow delivery into the vitreous should be favorable for improving the therapeutic outcome in addition to patient compliance. It was found that lipid vesicles are able to protect oligonucleotides against degradation by nucleases [344, 345]. Furthermore, they increase the retention time of many drugs in the vitreous. Thereby, the use of liposomes for intravitreal administration is a very promising approach.

***Nasal Delivery*** Liposomes are able to decrease mucociliary clearance in the case of nasal administration due to their low viscosity. This is attributed to the incorporation of liposomal lipids in the membranes of the nasal epithelial cells, which results in the opening of pores in the paracellular tight junctions [348]. MLVs containing nifedipine administered via the nasal route could attain a constant plasma level [349]. Liposomal formulations of levonorgestrel containing carbopol or chitosan demonstrated prolonged contact time with the absorptive surfaces, resulting in increased bioavailability of the intranasally administered drug [350].

The nasal route may be highly promising for candidate vaccines against potential pathogens (and infection-related diseases such as cancers) that utilize this route of infection [351, 352]. In addition to the likelihood of increased patient compliance, immune responses elicited by nasal administration may be more predictable when compared with the vaginal route due to immunological changes in the female reproductive tract during the menstrual cycle [352]. Previous work with liposome-containing vaccines for nasal delivery demonstrated that liposomes can confer adjuvancy to the subunit influenza vaccine, but also empty liposomes administered 48 h prior to immunization resulted in immune stimulation, emphasizing that the properties and composition of the liposomes play a significant role facilitating the transport of the antigen across the membrane [353]. More recently, several promising liposome-based vaccines are being designed and investigated for delivery by the nasal route, such as the liposome-encapsulated plasmid DNA-encoding influenza virus hemagglutinin, which has been reported to elicit mucosal, cellular, and humoral immune responses after intranasal administration in Balb/C mice [354]. Additionally, nasal immunization studies using liposomes loaded with tetanus toxoid were perfomed, and it was found that intranasal administration of liposome-encapsulated vaccines can be an effective way for inducing mucosal immune responses [355]. Furthermore, intranasal immunization studies have been carried out with liposomes containing recombinant meningococcal opacity proteins [356] and with anthrax-protective antigen protein incorporated in liposome–protamine particles [357], both with promising results.

***Vaginal Delivery*** The vaginal route has been under investigation in the last years, especially for the topical delivery of drugs that are intended to act in the vagina, as contraceptives, microbicides and antibiotics, as recently reviewed [358]. In such cases the main advantage of using liposomes would be the controlled and sustained release of the drug at the site, which would result in a less frequent drug administration and improved patient compliance.

However, the major limitation of using liposomes topically and vaginally is the liquid nature of preparation. Nevertheless, several formulation characteristics should be optimized in order to achieve the needed rheological and mucoadhesice properties for maximum retention of the delivery system in the vagina. Research for the development of liposomal gels (gels that contain liposomes) with the required properties is currently ongoing [358–360].

It has been demonstrated that, by their incorporation in an adequate vehicle, such as carbopol resins, the original structure of vesicles is preserved [359], while liposomes are fairly compatible with gels made from polymers derived from such resins. A previous study has suggested application of liposomes containing antimicrobial drugs for the local therapy of vaginitis [359], while recently the design and in vitro evaluation of bioadhesive liposome gels containing clotrimazole and metronidazole, or acyclovir, were carried out [359, 360].

Some other applications are arising lately concerning liposomes and vaginal administration. As mentioned above (in the nasal administration section), mucosal surfaces serve as a gateway to disease. It was recently demonstrated that RNA interference can be used to manipulate mucosal gene expression in vivo. Using a murine model, it was shown by Zhang et al. [361] that direct application of liposome-complexed small interfering RNA (siRNA) mediates gene-specific silencing in cervicovaginal and rectal mucosa. A single vaginal or rectal administration of siRNA targeting hematopoietic or somatic cell gene products reduced corresponding messenger RNA (mRNA) levels by up to 90%. Additionaly, liposomal siRNA formulations proved nontoxic, did not elicit a nonspecific interferon response, and provided a means for genetic engineering of mucosal surfaces in vivo.

In addition, it was recently found that when human immunodeficiency virus type 2 (HIV-2) DNA vaccine were formulated with cationic liposomes [362], stronger immune responses in mice were observed compared with naked DNA alone. Using this knowledge, very recently a vaccine consisting of some HIV-2 genes (*tat*, *nef*, *gag*, and *env*) was formulated within cationic liposomes by Lochera et al. [363]. Baboons (*Papio cynocephalus hamadryas*) that were immunized by the intramuscular, intradermal, and intranasal routes with these expression constructs were challenged with HIV-$2_{UC2}$ by the intravaginal route, and the results of this study demonstrate that partial protection against HIV-2 vaginal challenge, as measured by reduced viral load, can be achieved using only a DNA vaccine formulation.

## 4 APPLICATIONS OF LIPOSOMES IN THERAPEUTICS

It is well known that liposomes have many applications in drug delivery. Initially, after the limitations of conventional liposomes were noticed, great effort was given

**TABLE 4 Currently Marketed Liposome-Based Products**

| Active Agent[a] | Application |
|---|---|
| Daunorubicin (DaunoXome, Gilead Sciences, Inc.) | Kaposi's sarcoma |
| Doxorubicin (Doxil/Caelyx, Ortho Biotech ProductsLP/Sequus Pharmaceuticals) | Kaposi's sarcoma |
| Amphotericin B (Ambisome/Abelcet, Fujisawa Healthcare, Wyeth Pharmaceuticals) | Fungal infections in immunocompromised patients |
| Doxorubicin (Myocet/Evacet, Sopherion/ Liposome Company) | Metastatic breast cancer |
| Hepatitis A virus envelope proteins (Epaxal, Berna Biotech) | Hepatitis A |
| Influenza virus (Inflexal V, Berna Biotech) | Influenza |
| Verteporfin (Visudyne, Novartis Ophthalmics) | Age-related macular degeneration |

*Source*: D. Felnerova, J. F. Viret, R. Gluck, and C. Moser, *Current Opinion in Biotechnology*, 15, 518–529, 2004.

[a]Product names and companies given in parentheses.

toward the therapy of parasitic diseases, due to the fact that the targeting of RES macrophages, the place were parasites are mainly located, was considered to be very easy and fast (usually mentioned as passive targeting). Indeed, even now, research is ongoing, such as the treatment of drug-resistant visceral leishmaniasis with liposomal amphotericin-B [364] or with sterically stabilized liposomes containing camptothecin [371]. Nevertheless, possibly due to the high cost of liposome manufacturing, in relation to other drug formulations, and the fact that the main need for such medicaments would be for third world countries, such products have not been marketed for these diseases, despite the fact that they offer therapeutics advantages. In addition, again due to the same etiology, a very small part of recent research efforts and money are devoted to such diseases.

A list of the marketed liposomal products is presented in Table 4 [364–366]. In addition, liposomes are currently being investigated for a variety of conventional and novel drugs: therapeutic agents, including antibiotics (as amikacin [367], vancomycin, and ciprofloxacin [368]); anticancer agents (e.g., paclitaxel [369] and cisplatin [370]; camptothecin and analogs [371–373]), biologics such as antisense oligonucleotide [374], DNA, and siRNA [375]; and muramyl tripeptide [376]. A list of most liposome-based products currently under clinical investigation is presented in Table 5. Additionally many products are currently being evaluated in preclinical studies. In many of the latest studies, the liposomes used have been surface modified with active targeting ligands to improve delivery of therapeutics to target cells [238, 377–379].

After investigating the recent literature, we have seen that most recent efforts connected with the use of liposome in therapeutics are mainly connected with the treatment of cancer. This is the reason why here we deal with the most common cancer types (brain, breast, lung, and ovarian cancer), and ongoing research and clinical treatments are discussed in more detail. However, we do not want to imply that the future of liposomes in drug delivery is limited to cancer therapy. Indeed, liposome structure, characteristics, and versatility are sure to find, in

**TABLE 5 Liposome-Based Products Currently under Clinical Testing**

| Active Agent or Product[a] | Application | Company and Trial Phase/Reference |
|---|---|---|
| Drug delivery | | |
| Caelyx | Bladder cancer | Schering Plough; approved EU |
| Doxil | Multiple myeloma | Schering Plough; III ALZA Pharm; III |
| | Bladder, liver cancer | ALZA Pharm; II, III |
| | Pancreatic cancer | ALZA Pharm; II, Sequuz; II |
| Doxorubicin combined with ATB | Prostate cancer | Pharmacia; III, Neopharm; III |
| | Bladder cancer | Neopharm; II, III Pharmacia; II, III |
| Myocet combined with ATB | Bladder cancer | Liposome; III |
| Liposomal ether lipid (TLC ELL12) | Bladder cancer | Liposome; I |
| | Lung, prostate, skin cancer | Elan Pharm; I |
| Platinum (Aroplatin) | Cervical, ovarian, kidney | Aronex Pharm; II |
| | Lung, pancreatic cancer | Aronex Pharm; I, II |
| Annamycin | Leukemia | |
| Paclitaxel | Head and neck cancer | Pharmacia; II, III |
| | Lung cancer | Neopharm; II, III |
| Vincristine (Onco-TCS) | Lung cancer | Inex Pharm; II |
| Topoisomerase inhibitor (OSI 211) | Lung, ovarian cancer | OSI Pharm; II |
| All-trans retinoic acid (ATRA-IV) | Lung cancer | Antigenics; II |
| Mitoxanthrone | Other cancers | Neopharm; II |
| Nystatin (Nyotran) | Leukemia (antifungal) | Pharmacia; II |
| | Lung (antifungal) | Aronex Pharm; II |
| | Prostate cancer (antifungal) | Abbott Laboratory; II |
| DNA delivery | | |
| Human leucocyte antigen (HLA) B and $\beta_2$ microglobulin plasmid DNA (Allovectin) | Head and neck cancer | Vical; II |
| Interleukin-2 plasmid DNA | Kidney, prostate cancer | Vical; II |
| Antisense toraf-1 (LerafON) | Leukemia | Neopharm; I |
| Antigen delivery: | | |
| MUC-1 peptide: BLP25 (human epithelial mucin peptide) | Lung cancer | Biomira; II<br>Merck; III |

*Source*: D. Felnerova, J. F. Viret, R. Gluck, and C. Moser, *Current Opinion in Biotechnology*, 15, 518–529, 2004.

[a]Commercial names are given in parentheses. ATB, antibiotika.

addition to those existing already, numberous applications in drug delivery in the future.

Recently, a multicomponent liposomal drug delivery system consisting of doxorubicin and antisense oligonucleotides targeted to MRP1 mRNA and BCL2 mRNA to suppress pump resistance and non–pump resistance, respectively, has been developed [379]. This liposomal system successfully delivered the antisense oligonucleotides and doxorubicin to cell nuclei, inhibited MRP1 and BCL2 protein synthesis, and substantially potentiated the anticancer action of doxorubicin by stimulating the caspase-dependent pathway of apoptosis in multidrug resistant human lung cancer cells.

### 4.1 Anticancer Drug Delivery

***Brain Tumors*** Brain tumors are classified as gliomas (astrocytomas, oligodendrogliomas, ependymomas) and primitive neuroectodermal tumors (PNET) (medulloblastoma and supratentorial PNETs). Approximately half of all primary brain tumors are gliomas, while 80% of those are astrocytomas and glioblastomas. Most chemotherapeutic drugs are toxic to the healthy tissue and have damaging side effects due to their nonspecific nature. Incorporation of those drugs in liposomes can enhance the therapeutic efficacy and reduce the toxicity. Drug delivery to the brain via the intravenous route has been a very challenging task due to the strict selectivity of the blood–brain barrier (BBB) as to the number and kind of molecules able to pass through.

The BBB is the tight junction formed between the cerebral endothelial cells (Figure 11). These cells are in close contact with astrocytes and pericytes connected over a basal membrane. Only small (MW < 400–600) lipophilic molecules can diffuse

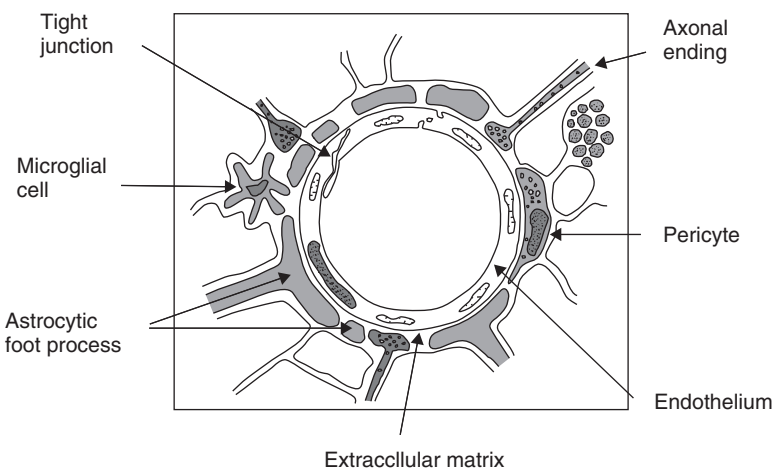


**FIGURE 11** Schematic of neurovascular unit/cell forming BBB (brain–blood barrier). (Reproduced from D. J. Begley, *Pharmacology & Therapeutics*, 104, 29–45, 2004, with permission by Elsevier.)

through the BBB, while the majority of the circulating drugs cannot access the brain. Nutrients and peptides pass through the BBB via receptor-mediated or carrier-mediated transport systems. These mechanisms are exploited in an attempt to deliver drug-loaded liposomes into the central nervous system (CNS) [380]. The most common are low-density lipoprotein (LDL) receptors, insulin receptors, and transferrin receptors.

Thus, mAb against the transferrin receptor OX26 mAb has been conjugated via a stable thioether bond to the end of the PEG chain inserted on the liposome surface. Tritiated daunomycin was incorporated in OX26 mAb–PEG liposomes and the formulation was given intravenously to rats. The brain volume of distribution of daunomycin increased with time and exceeded 200 μL/g 24 h after injection [381]. In contrast, the pharmacokinetics of the free drug and the drug loaded in conventional liposomes was much lower [382].

In another study, the OX26 mAb has been grafted on the PEG chains by the biotin streptavidin coupling [383]. Brain tissue distribution obtained using biotinylated immunoliposomes was the same with that reported in the previous work where the mAb was chemically linked on the distal end of PEG. Therefore, the coupling method has not had a great impact on the brain accumulation of immunoliposomes. However, accumulation of the biotinylated PEG immunoliposomes was quite high in tissues such as liver, spleen, heart, muscle, and kidney. The latter was attributed to either the fact that the OX26–biotinylated PEG immunoliposomes could pass through the BBB by an active transport system or the biotinidase activity, which could mediate cleavage of the targeting antibody from PEG and interfere with the tissue distribution of the formulation.

An increase in therapeutic efficacy and lower toxicity was reported with liposomes where the bradykinin analogue RMP-7 was chemically attached at the end of PEG molecules of PEGylated liposomes (approximate size 70 nm) [384]. RMP-7 exhibits high selectivity for the B2 receptor of the BBB endothelial cells, which "shrunk" and let the RMP-7–PEG liposomes to pass into the brain. Actually the mechanism used in that study was based on opening the tight junctions of the BBB. Liposome-incorporated nerve growth factor (NGF) concentration increased 10 times in comparison to free NGF, while they accumulated mainly in striatum, hippocampus, and cortex.

A different type of immunoliposome was developed using antinuclear autoantibodies with nucleosome (NS)–restricted specificity [187]. Anti-NS mAb 2C5 specifically recognizes human brain tumor cells. Evaluation of immunoliposomes 2C5–PEG–PC/Chol was carried out in nude mice bearing subcutaneous brain tumor (U-87 astrocytoma) and exhibited a threefold higher accumulation in the tumor in comparison to control (IgG–PEG liposomes).

Moreover, disialoganglioside ($GD_2$) is expressed in abundance by neuroectodermal cancer cells and in low levels by normal cells located mainly in cerebellum and peripheral nerves. The Fab′ fragment of the monoclonal antibody anti-$GD_2$ was grafted on PEG chains of sterically stabilized liposomes loaded with doxorubicin and their potential in treating neuroblastoma was assessed in nude mice with HTLA-230 xenografts [160]. Mice receiving intravenously the immunoliposomes showed significant improvement in long-term survival compared with other mice that received free DOX, free$GD_2$ Fab′, Fab′–PEG liposomes, PEG liposomes–DOX. The control mice died from metastatic disease, while the immunoliposome-treated

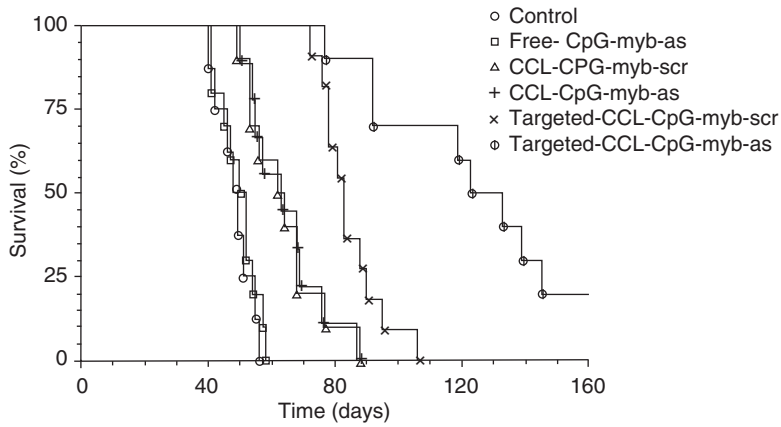


**FIGURE 12** Survival of NB-bearing nude mice after injection of oligonucleotides (free or encapsulated within liposomal formulations). Nude mice were injected intravenously with $3.5 \times 10^6$ HTLA-230 neuroblastoma cells. After 4 h each mouse received 50 μg of oligonucleotides either free or encapsulated in targeted or nontargeted liposomes. Control mice received HEPES-buffered saline. (Reprinted from ref. 385 with permission of Elsevier.)

group lived 4 months longer. The same ligand, anti-$GD_2$–Fab′, was used with liposomes carrying the antisense oligonucleotide c-myb [385]. After intravenous injection of the targeted liposomes into nude mice with HTLA-230 xenografts, significant prolonged survival times were obtained in comparison to controls (Figure 12). The suggested mechanism was downregulation of c-myb proto-oncogene expression.

Although all the active targeting liposomes mentioned earlier have not left the laboratory, nonspecific sterically stabilized liposomes are being tested in clinical trials. Doxorubicin is the anticancer agent which is used as standard therapy, and it has the most serious side effects (mucositis, cardiotoxicity), so its incorporation in liposomes and bioavailability enhancement are under scrutiny [386–388].

It has also been demonstrated that PEGylated liposomal doxorubicin, Caelyx, can cross the BBB with a consequent accumulation in primary and secondary brain lesions [389]. In 10 patients with metastatic brain tumors treated with radiolabeled Caelyx concurrent with radiotherapy, the accumulation of the liposomal doxorubicin was 7–13 times higher in the metastatic lesions compared to the normal brain.

Liposomal formulations, type of tumor, anticancer agent, delivery pathway, day of treatment, and general conclusions are given in Table 6. DaunoXome (liposomal daunorubicin) and Doxil (liposomal doxorubicin) have been proved to have good response in clinical trials, in the range of approximately 40%.

Also, according to Arnold et al. [390], the dose scheme of doxorubicin-loaded liposomes affects the drug accumulation in various tissues as well as the tumor. The authors reported that after repetitive doses of sterically stabilized liposomes (SSL)-DOX every week, the plasma half-life of the drug increased, the deposition in liver and spleen decreased, and peak concentrations of DOX in the heart were threefold

**TABLE 6 Recently Completed and Ongoing Clinical Trials of Low-Molecular-Weight Drug-Carrying Liposomes for Brain Tumors**

| Tumor | Number of Patients | Treatment | Delivery | Day of Treatment | General Conclusions |
|---|---|---|---|---|---|
| Recurrent tumor | 14 | Daunorubicin | IV | Once every 4 weeks | 6/14: Patients showed positive response |
| GBM | 8 | Daunorubicin | IV | 24 h before surgery | Concentration similar in tumor mass and peripheral regions |
| Pediatric glioma | 7 | Daunorubicin + free carboplatin and etoposide | IV | Daunorubicin: day 1 and 2; carboplatin and etoposide: day 1 | 5/7: Showed positive response with monthly treatment |
| GBM | 15 | Doxorubicin + radiation | IV | Doxorubicin: days 1, 21; radiation: days 1–21, 21–23 | 4/10: Patients completely responded |
| Solid tumor | 22 | Doxorubicin | IV | Once every 4 weeks | Phase I: dose-limiting toxicity: 70 mg/m$^2$ |
| Glioma | 40 | Daunorubicin + free tamoxifen | IV | PEG-dox: day 4 every 14 days | Phase II: response (including stabilization) 40%; tamoxifen: day 1 every 4 days |
| Glioma | 8 | $^{111}$In labeled | IV | Contrast only | Tumor uptake: 1.1% max tumor/brain contrast 7:5 |
| Glioma | 3 | Bleomycin | IT | Twice weekly for up to 6 weeks | All patients deteriorated, no toxicity |
| Recurrent meningeal malignancies | 15 | Cytarabine | IT | Once every 2 weeks for 2 courses; in positive response, second induction in 2 weeks after first dose | |
| Meningeal neoplasms | 100 | Cytarabine against free methotrexate | IT | Once every 2 weeks | Ongoing |

*Source*: From G. H. Huynh, D. F. Deen, and F. C. Szoka, Jr., *Journal of Controlled Release*, 110, 236–259, 2006. Reprinted with permission of Elsevier.

*Note*: GBM: glioblastoma; IV, intravenous; IT: intratumor; PEG: poly(ethylene glycol). See http://www.clinicaltrials.gov.

lower. These results were not obtained using free DOX. In addition, a significant increase in survival was achieved in animals treated weekly with SSL-DOX, while animals treated with free drug did not survive longer than the untreated controls.

In phase I clinical trials children with recurrent or refractory tumors previously treated with free doxorubicin were administered i.v. Doxil in various doses (40–70 mg/m$^2$) in order to determine the best tolerated one [391]. It was concluded that the maximum tolerated dose was 60 mg/m$^2$ because at the highest, 70 mg/m$^2$, some patients developed mucositis, so dose adjustment was necessary.

Administration of a combination of liposomal anthracyclines in parallel with other anticancer agents has been found of great advantage as they prolong the patients' survival. Caraglia et al. evaluated in a phase II study the use of combination Doxil with temozolomide in the treatment of brain metastases from brain tumors [392].

It is worth mentioning that temozolomide accumulates significantly in the brain after oral administration. It is well tolerated and therefore is considered a potential candidate for combination chemotherapy. Administration of 200 mg/m$^2$ of temozolomide for 5 days and liposomal doxorubicin 35 mg/m$^2$ on day 1 was performed on 19 patients. The overall response rate was 36.8% and the median overall survival was 10.0 months.

Moreover, Fiorillo et al. [393] studied the effect of a combination of liposomal daunorubicin, etoposide, and carboplatin administered to seven children with recurrent malignant supratentorial brain tumors as a second-line therapy. Chemotherapy consisted of infusion of liposomal daunorubicin on days 1 and 2 and infusion of etoposide and carboplatin on day 1 whereas courses were repeated every 3–4 weeks. After a total of eight courses, five of seven children evaluated were alive 12–64 months after diagnosis and 8–29 months from the start of the second-line chemotherapy. Of the seven children, three showed complete response, two partial responses, one stable disease, and one progressive disease. The time to the best response was 3–10 months, while the median time to progression was 23 months. The toxicity observed was minimum.

Boron neutron capture therapy (BNCT) is also of high interest in treating brain tumors, especially glioblastoma multiforme, due to the high degree of normal brain infiltration, the high histological complexity, and the heterogeneity of the cellular composition of the latter. This method is based on the nuclear reaction which occurs when boron-10 is irradiated with low-energy thermal neutrons, producing high linear energy transfer of α particles and lithium-7 nuclei [394]. Development of BNCT has been ongoing over the last 50 years and the greatest challenge is to achieve selective tumor targeting at a sufficient therapeutic dose with minimal toxicity. Various compounds are currently being used or have been used in BNCT (shown in Figure 1 of [394]), such as BPA [(L)-4-dihydroxy-borylphenylalanine] and BSH (sodium mercaptoundecahydro-*closo*-dodecaborate), which are the first most successfully used chemical compounds (so-called second-generation compounds) due to low toxicity, longer retention to the tumor site and tumor/brain, and tumor/blood ratios higher than 1. However, these drugs are not ideal, but they are safe after i.v. administration, so they are being used in clinical trials in Europe, the United States, Japan, and Argentina. The next group of advanced boron molecules (third generation) consists of stable boron or a cluster attached to a tumor-

targeting moiety, such as monoclonal antibodies or low-MW biomolecules with amphiphilic properties. Other anionic compounds show little specificity, so their potency increases when they are incorporated in either targeted or nontargeted liposomes [395, 396].

According to Feakes et al. [396], boron-loaded DSPC/Chol liposomes of 40 nm average size were prepared and injected i.v. in murine mice carrying EMT6 tumors. Those liposomes showed high tumor retention, while boron amount was at therapeutic levels through the entire course of the experiment (more than 15 μg B/g tumor). However, targeted liposomes would, in theory, assure higher boron accumulation than the nontargeted ones [397].

Due to the overexpression by glioma cells, the most potent ligands for glioma treatment are endothelial growth factor receptor (EGFR) [398], the *vIII* mutation of EGFR [399], platelet-derived growth factor (PDGFR) [400], and tenascin epitopes [401].

Liposomes are also used as carriers for gene delivery to gliomas while the cationic ones have demonstrated better interaction with cells in comparison to other types of liposomes [402, 403]. However, cationic liposomes suffer from toxicity, which varies according to cell type, duration of exposure, and density of cell culture. Antisense genes have been incorporated into liposomal carriers. For example, the EGFR antisense gene was packaged in PEGylated immunoliposomes [carrying human insulin receptor antibody (HIR)] [404]. It was reported [405] that the liposomes could cause 70–80% inhibition in human glioma cell growth. The same authors reported a 100% increase in survival time of mice with intracranial human brain cancer with weekly i.v. injections of antisense gene therapy directed at the human EGFR [405].

Moreover, double immunoliposomes were developed in order to treat intracranial human brain cancer in mice [406]. The mAbs used were the rat 8D3 mAb to the mouse transferrin receptor and the mAb against the HIR. RNAi (intereference RNA) is a new antisense gene therapy, where an expression plasmid encodes for a shRNA (short hairpin RNA). The shRNA is processed in the cell to an RNA duplex. The latter mediated RNAi. Indeed, weekly i.v. RNAi gene therapy reduced tumor expression of immunoreactive EGFR and caused an 88% increase in survival time of mice with advanced intracranial brain cancer.

In another study, the herpes simplex virus thymidine kinase (HSVtk) gene was evaluated as to its potency to increase the sensitization of ganciclovir (GCV) to glioma cell lines when that gene was incorporated in liposomes [407].

Although the efficiency of transfection was 18.6% in vivo after intratumoral injection of DNA liposomal complexes, the sensitivity to ganciclovir was improved as tumor weight induction was observed. In 2001, the FDA approved a clinical protocol relevant to liposomal gene therapy with the HSVtk/GCV system for the treatment of glioblastoma multiforme [408].

Hybrid vectors consisting of adeno-associated virus (AAV) vectors enclosed in liposomes lead to a10-fold increase in transduction efficiency compared to liposomes containing plasmid DNA and 6-fold increase compared to AAV vector alone [409].

However, alternative routes are currently used in the clinic according to the site/location and type of tumor to achieve higher therapeutic efficiency [410] because systemically administred drugs are not able to pass to cerebrospinal fluid (CSF),

whereas there is direct tumor exposure to the drug, characterized with increased drug concentration and half-life [410]. Thus, drugs can be administered in CNS via intrathecal, intraventricular, and intraparenchymal routes. Intrathecal and intraventricular routes result in a high drug concentration in the bulk CSF but with limited penetration to parenchyma. Thus, they are suited for treating meningeal and ventricular diseases. The intraparenchymal route assures the delivery in a local region in the parenchyma and so is useful for solid tumors and degenerative diseases that are surgically accessible.

DepoCyt, a slow-release liposomal cytosine–arabinoside, undergoes clinical trials, as in some previous preclinical and clinical tests it was shown of good potency for the treatment of meningeal malignancies. For example, DepoCyt was administered to children with refractory neoplastic meningitis via lumbar puncture using an Ommaya reservoir or intraventricularly [411]. That study demonstrated the safety and feasibility of using liposomal cytarabine in children older than three years at the recommended dose of 35 mg with concomitant administration of dexamethasone. The latter is given to reduce the side effects of liposomal cytarabine, which is mild headache and back or neck pain. In a relevant study on adults with lymphomatous meningitis, DepoCyt given once every two weeks at a dose of 50 mg yielded a response rate of 71%, whereas free cytarabine at a dose of 50 mg twice a week produced a response rate of only 15% [412]. In patients with solid tumor neoplastic meningitis, 50 mg DepoCyt given once every two weeks yielded a response rate in evaluable patients of 36%, whereas free methotrexate 10 mg given twice a week produced a response rate of 21%. However, a problem accompanies the direct delivery to the brain parenchyma: the limited diffusion coefficient of particles in general (liposomes, nanoparticles, viral vectors) from the injection site in the brain tissue [410]. Thus, intraparenchymal delivery of liposomal carriers is facilitated using convection-enhanced delivery (CED) to distribute the drug through a larger region in the tissue. CED utilizes a bulk flow mechanism to deliver and distribute macromolecules to clinically significant volumes of solid tissues providing a larger volume of distribution. A range of parameters affecting CED are connected with the physicochemical properties of liposomes, that is, size, surface charge, and steric stabilization [413].

According to investigations by MacKay et al. [413], the ideal liposome (or in general nanoparticle) for CED will be PEGylated and of less than 100 nm in diameter, have neutral or negative surface charge, and need a targeting ligand to direct the particle to the target cell. For example, mannosylated liposomes containing clodronate infused directly to the fourth ventricle of the rat brain were selectively taken up by macrophages but not from microglia [414]. Also, the lipid dose is important as it was shown that infusing a high total lipid concentration reduced the fraction of the dose taken up by perivascular cells in the brain.

However, MacKay and co-workers [413] carried out the previously mentioned study using healthy animals. Mamot et al. administered liposomes of either 40 or 90 nm with 25% mannitol in U-87 glioma xenograft animals [415]. In contrast to MacKay et al. [413], the small 40-nm liposomes infused via CED without mannitol achieved distribution over nearly all the tumor tissue. The addition of mannitol to CED further enhanced distribution, as liposomes accumulated throughout all sections of tumor and further penetrated beyond the tumor boundary into adjacent normal brain tissue. For the 90-nm liposomes, CED with mannitol resulted in more

than 50% tumor penetration of the total area evaluated, which also included extension into surrounding normal brain. Distribution was somewhat less extensive than with the 40-nm liposomes, though.

The efficacy of liposome administration with CED was demonstrated using real-time magnetic resonance imaging (MRI) of rat brain tumors [416]. Two types of liposomes were used: doxorubicin-loaded liposomes and gadolinium (Gd)–loaded (MRI contrast agent) liposomes. According to the authors, MRI facilitated the distinction of distribution between different tumor models, such as C6 gliomas and 9L-2 gliomas. Also, after CED of Doxil, the drug presence was identified in the tissue several weeks after a single administration, while the therapeutic response achieved was greater compared to systemic administration of Doxil. The fact that therapeutic liposomes can be coinfused with liposomal Gd to successfully monitor the distribution of the drug carrier in the nonhuman primate brain was shown by Saito and co-workers [417]. A 2:1 ratio between the volume of distribution and the volume of liposomes infused was found while liposomal Gd was still detectable 48 h postinfusion, confirming the previous finding of prolonged retention of liposomal Gd in the tissues and negligible toxicity in rat. With this study, the authors showed that CED is a technique enabling safe and extensive liposome distribution. Also, real-time MRI is a potentially useful tool to estimate the concentration and tissue half-lives of liposome-loaded drugs within target tissues.

However, CED might not always be successful as good technical skills are required (if the catheter is not placed properly, the liposomes will escape within the CSF). Also CED is an invasive technique and could cause inflammation and neurotoxicity and it is determined by many formulation characteristics and diffusional properties of the latter in brain tissue [418].

***Breast Cancer*** Breast cancer is the second leading cause of cancer deaths in women today (after lung cancer) and is the most common cancer among women, excluding nonmelanoma skin cancers [419]. According to the World Health Organization, more than 1.2 million people will be diagnosed with breast cancer this year worldwide. Breast cancer stages range from stage 0 (very early form of cancer) to stage IV (advanced, metastatic breast cancer). Each patient's individual tumor characteristics, state of health, genetic background, and so on, will impact her survival. In addition, levels of stress, immune function, will to live, and other unmeasurable factors also play a significant role in a patient's survival. The majority of women with breast cancer will undergo surgery as part of their cancer treatment (lumpectomy and mastectomy). In addition to surgery, some women will receive adjuvant (additional) treatment (chemotherapy, radiation therapy, and drug treatments) to stop cancer growth, spread, or recurrence. Occasionally women may be treated with chemotherapy, radiation, or drugs without having breast surgery.

From a clinical point of view and with relevance to liposomes, administration of either PEGylated (Doxil) or non-PEGylated liposomal doxorubicin (Myocet) has improved the safety and tolerance of patients with breast cancer by significantly reducing the main side effect of those drugs, cardiotoxicity. However, drug efficacy remains the same either incorporated or not in the liposomes as this is reported after conducting various clinical studies [420, 421]. For example, Myocet (M) received European approval for use in patients with metastatic breast cancer at a dose of

60 mg/m$^2$ in combination with cyclophosphamide (C) having shown equivalent efficacy at a phase III study [422]. Also, doxorubicin is given at a dose of 60 or 75 mg/m$^2$ [423]. Chan et al. [424] studied the efficacy and tolerability of Myocet in a dose of 75 mg/m$^2$ in combination with cyclophosphamide in 160 patients. A high incidence of neutropenia was obtained. Thus, that group confirmed that the use of 60 mg/m$^2$ is an appropriate choice. However, administering a combination of either Myocet and cyclophosphamide or epirubicin and cyclophosphamide to patients with metastatic breast cancer, the response rate was not significantly different, which showed that the use of the liposomal formulation just reduced the cardiotoxicity of doxorubicin.

In a phase I study, Myocet was administered in combination with docetaxel in 21 metastatic breast cancer patients with the aim of estimating the safety and maximum tolerated dose of Myocet in parallel with docetaxel [425]. The latter is proved highly active in metastatic breast cancer as well as doxorubicin. The maximum tolerated dose was 50 mg/m$^2$ of Myocet and 25 mg/m$^2$ of docetaxel. As reported in the previous study, neutropenia was the most common toxicity effect while some incidents of congestive heart failure were observed after a total doxorubicin dose of 540 mg/m$^2$. Moreover, a combination of Myocet (75 mg/m$^2$) with gemcitabine (350 mg/m$^2$) and docetaxel (75 mg/m$^2$) was injected intravenously in 44 patients with early breast cancer every three weeks for six cycles [426]. The overall clinical response rate was 80% and the pathological complete response was 17.5%. The toxicity of the regimen was moderate and, as expected, neutropenia and leukopenia were the most prominent side effects. The latter effects were well managed and treatment discontinuation was not required.

Using PEGylated liposomal doxorubicin (Caelyx), Keller et al. [427] compared the efficacy of the liposomal formulation with that of a common regimen in patients with taxane-refractory advanced breast cancer. The regimen scheme was Caelyx (50 mg/m$^2$ every 28 weeks) or vinorelbine (30 mg/m$^2$) or mitomycin C (10 mg/m$^2$ every 28 days) plus vinblastine (5 mg/m$^2$ at days 1, 14, 28, 42). Finally, progression free survival and overall survival were similar for Caelyx and the comparative regimen.

The efficacy and toxicity of Caelyx in combination with paclitaxel (Taxol) were investigated as a first-line therapy in 34 patients with advanced breast cancer in a multicentric phase II study [428]. Paclitaxel at a dose of 175 mg/m$^2$ and Caelyx (30 mg/m$^2$) were administered intravenously every 3 weeks. It was shown that the response rates of the combination were over 70% while the median time to treatment failure was 45 weeks. No significant clinical cardiotoxicity was observed and the usual side effects (mucositis, stomatitis, hand–foot syndrome) were treated accordingly.

In another phase II study, Caelyx was used in combination with cyclophosphamide (CP) as a first-line therapy in patients with metastatic or recurrent breast cancer [429]. Three different schemes were given in groups of patients: Caelyx 50 mg/m$^2$ intavenously on day 1 and CP 100 mg/m$^2$ (orally on days 1–14) every 28 days, 30 mg/m$^2$ Caelyx and 600 mg/m$^2$ i.v. on day 1 every 21 days, and 35 mg/m$^2$ Caelyx plus 600 mg/m$^2$ CP i.v on day 1 every 21 days. The responses were similar among the different groups (51%) of patients, but less side effects were observed in group 2, making that regimen the one of choice as a first-line therapy for patients

with metastatic or recurrent breast cancer. The median duration of response was 35.1 weeks.

Coleman et al. [430] demonstrated that using Caelyx in a dose format of 50 mg/m$^2$ every four weeks in patients with metastatic breast cancer is quite effective (objective partial responses 31%) while the main side effect was hand–foot syndrome.

From a research point of view, liposomes have been used in order to facilitate or alter the pharmacokinetics/bioavailability of various molecules. For example, ceramide is an antiproliferative and proapoptotic molecule produced after sphingomyelin metabolism [431]. Ceramide C6–loaded PEG(750)-C8 liposomes were injected intravenously in mice bearing syngeneic or human xenografts of breast adenocarcinoma. Administration of 36 mg/kg of liposomal C6 over a three-week period caused a sixfold decrease of tumor size in the case of syngeneic tumor–bearing mice. Liposomal C6 accumulated in caveolae and mitochondria, while a marked increase of apoptotic cells was observed. The PEGylated liposomal ceramide followed first-order kinetics in the blood and a steady-state concentration was achieved in tumor tissue (Figure 13). Also, a decrease in tumor size was obtained in the human xenograft model. Minimal toxicities were observed in tumor-bearing mice, suggesting that the bioactive concentration of C6 achieved in the tumor tissue is not active in normal tissues. Besides, contortrostatin (CN), a molecule isolated from snake venom, has been proved to interact with integrins on tumor cells and with newly growing vascular endothelial cells via its two Arg–Gly–Asp sites [432]. Thus, interactions between CN and integrins disrupt several steps critical to tumor growth: angiogenesis and metastasis. CN was incorporated in PEGylated liposomes with an encapsulation efficiency of approximately 80% while retaining full biological activity. Intravenous injection of approximately 100 μg of liposomal CN twice a week caused a significant reduction of 94% of microvascular density and hindered the tumor growth in the MDA-MB-435 xenograft model. Also, liposomal CN demonstrated prolonged circulation ($t_{1/2}$ values were 19 and 0.5 h for liposomal and nonencapsulated CN, respectively).

In another study, the authors modified the liposome composition by including polyoxy-ethylenedodecyl ether [$C_{12}$(EO)n] in the DMPC-made liposomes [433]. These vesicles caused activation of caspases 3, 8, 9 and eventually induction of apoptosis in MDA-MB-453 cells. Another idea is based on using methods to direct the drug loading carriers to the cancer site. More specifically, the paclitaxel is successfully used for the treatment of breast cancer among other types of cancers but has side effects such as neutropenia, peripheral neuropathy, and hypersensitivity reactions. Also, it is formulated with Cremophor EL due to its high lipophilicity. In order to decrease the drug toxicity and enhance the therapeutic potential of that drug, paclitaxel was incorporated in negatively charged magnetic liposomes and its efficacy was evaluated in vivo [191]. A magnetic field of suitable strength was used to direct the magnetoliposomes to the desired site. Indeed, after intravenous injection the AUC obtained for the magnetic carriers and the Cremophor EL/ethanol were 20.7 and 6.8 h·μg/mL, respectively. The liposomal paclitaxel concentration was much higher in the tumor in comparison to other organs while the concentration peak was reached sooner (19.85 μg/g at 0.25 h). Even after 8 h, paclitaxel concentration in the plasma was 2.71–0.33 μg/mL for the Cremophor EL formulation. The antitumor efficacy of magnetoliposomes

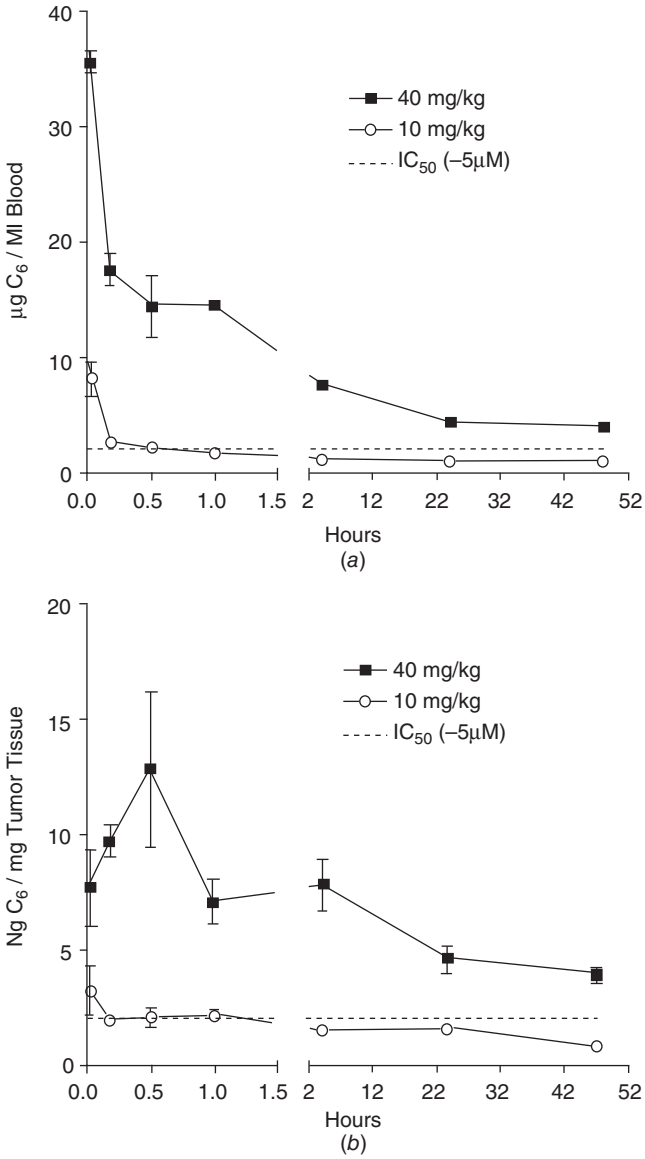


**FIGURE 13** Blood and tumor concentrations of bioactive ceramide–lipid C6 in tumor-bearing mice were maintained over a 48-h period: (*a*) 10- and 40-mg/kg doses of liposomal-C6 followed first-order kinetics with blood concentration exceeding in vitro $IC_{50}$ sustaibed at 48 h; (*b*) at these doses steady-state concentration of C6 in tumor tissue achieved at ~60 min. The 40-mg/kg dose maintained a concentration above the desired $IC_{50}$ up to 48 h. (Reprinted from ref. 431 with permission of American Association for Cancer Research.)

(expressed as change rate of tumor weight) after subcutaneous injection of 10 mg/kg drug was equivalent to that achieved with Cremophor EL 50.4 and 51.9%. When the administered dose was 20 mg/kg, the change rate of tumor weight was the highest, 60.5%.

Pulsed high-intensity focused ultrasound (HIFU) is another suggested method to enhance the delivery and therapeutic effect of Doxil in a murine breast cancer tumor model with the final aim of reducing the Doxil dose given during therapy [434]. However, the results were not particularly encouraging as there was not any significant difference between free and liposomal Doxil distribution in the tumor after HIFU exposure. The latter was possibly attributed to the ability of liposomes to extravasate easily through the leaky vasculature of the tumor, so the use of HIFU did not add any therapeutic or other advantage.

In an attempt to achieve active targeting, sigma receptor ligands were incorporated on to the liposomal surface. Sigma receptors are overexpressed in various human tumors, including breast cancer cells [435]. Haloperidol has shown high affinity for sigma receptors; thus it has been attached at the end of PEG molecules, which are protruding from the surface of cationic lipoplexes. Indeed, haloperidol–PEG lipoplexes showed more than 10-fold higher gene expression in MCF-7 (breast carcinoma) cells than control lipoplexes, while the presence of haloperidol or 1,3-ditolylguanidine suppressed the expression of the reporter gene [435].

Transferrin is another ligand attached on the lipoplex surface so to cause higher transfection efficiency, as has been reported [436]. Thus, Basma et al. [437] studied the effect of *cis*-diaminedichloro platinum (CDDP) in combination with bcl-2 antisense treatment on p53(+)MCF-7 and p53(–)MCF-7/E6 breast cancer cells using transferrin bcl-2 lipoplexes. The median inhibitory concentration ($IC_{50}$) for bcl-2 antisense delivered with lipoplexes plus transferrin was approximately 1.4 μ*M* for MCF-7 and 1.2 μ*M* for MCF-7/E6 cells (Figure 2 of ref. 437). In CDDP-treated cells, the $IC_{50}$ was approximately 5 μ*M* for both cell lines. In general, bcl-2 antisense delivered in the form of transferrin bcl-2 lipoplexes in combination with cisplatin induced cell death and apoptosis in a higher degree in MCF-7/E6 cells rather than MCF-7 cells. Cisplatin demonstrated higher caspase-8 activation compared to the targeted lipoplexes, which suggested that possibly caspase-8 is the major pathway for cancer cell death. G3139 is another oligonucleotide (ODN) capable of down-regulating bcl-2 and, its efficient delivery to breast cancer cells would potentially make it a successful candidate for antisense therapy [438]. Thus, the efficiency of different liposomal formulations for the lipid composition (DOTAP, DC-Chol, CCS)[1] with or without helper lipids DOPC, DOPE, and Chol and liposomal size [approximately 100 nm LUV or unsized heterolamellar vesicles (UHV)] on MCF-7 breast cancer cells was examined [438]. Out of 18 tested formulations, only the CCS-bcl-2 lipoplexes (UHV derived) have been effective, causing a larger than 50% decrease of cell growth in comparison to free ODN. The possible mechanism of action of the particular formulation is attributed to the presence of one primary and two secondary amines on the spermine moiety of the CCS molecule. Because of this, CCS lipoplexes are taken up by cells via adsorptive endocytosis and the secondary amines cause a "proton sponge effect" due to which ODN escapes from the endosomes and so has the ability to interact with the bcl-2 neutralizing it. Of course, the effect of the $L^+$/DNA– ratio is a determining factor along with the lipid composition and size.

[1]DOTAP: *N*-(1-(2,3-dioleoyloxy)propyl)-*N*,*N*,*N*-trimethylammonium chloride; DC-Chol: 3β[*N*-(*N′N′*-di methylaminoethane)carbamoyl]-cholesterol; CCS: ceramide carbomoyl spermine.

Diagnosis is always the first aim in effectively treating breast cancer, so the need to develop or reveal more tumor markers at an early stage of the cancer is absolutely critical [439]. Such are the circulating epithelial cells, the cyclins, and the urokinase-type plasminogen activator and plasminogen activator inhibitor, which indicate breast cancer or metastatic spread apart from the already existing markers estrogen receptor, progesterone receptor, and human epidermal growth factor receptor-2 (HER-2). Liposomal formulations have facilitated anticancer activity of highly toxic anticancer drugs as well as altered their biodistribution, while targeted drug-loaded liposomes have shown some promising results in the laboratory. Development of new drugs and use of particulate carriers to increase bioavailability could be one way forward in the battle for improvement of quality of life for patients suffering from breast cancer.

***Lung Cancer*** Lung cancer might be the most common form of cancer and the most common cause of death in both men and women, although it affects more men than women [440]. There are three main types of lung cancer, based on their appearance when examined by a pathologist: small cell carcinoma, squamous cell carcinoma, and adenocarcinoma. The latter two consist of non–small cell lung cancer. It is important to know which type of cancer a patient has because small cell cancers respond best to chemotherapy (anticancer medicines) whereas the other types (often referred to collectively as non–small cell cancer) are better treated with surgery or radiotherapy (X-ray treatment).

Surgery can cure lung cancer, but only one in five patients are suitable for this treatment. If the tumor has not spread outside the chest and does not involve vital structures such as the liver, then surgical removal may be possible, but only if the patient does not also have severe bronchitis, heart disease, or other illnesses. Small cell lung cancer is treated with chemotherapy. Non–small cell cancer may be treated with radiotherapy and chemotherapy (as part of a research trial) or with supportive care. Radiotherapy is either "radical" or "palliative."

Regarding drug treatment of lung cancer, Merck KGaA and Biomira will soon start phase III trial on their BLP-25 liposomal vaccine for patients with non–small cell lung carcinoma [441]. L-BLP25 is a synthetic MUC1 peptide vaccine. MUC1 is a mucinous protein expressed on the apical borders of normal epithelial cells. It is overexpressed and glycosylated on tumor cells, where it appears to be antigenically distinct from the normal protein. The liposomal vaccine can induce a MUC1-specific T-cell response. Median survival for patients with stage IIIB locoregional non–small cell lung cancer who received L-BLP25 in a phase IIb study was 30.6 months in the vaccinated group compared with 13.3 months for the unvaccinated group.

Caelyx is liposomal doxorubicin very well used as a treatment of choice for a number of cancers with good tolerability and antitumor activity, as has been demonstrated in many phase I or II clinical trials. One such example is the conduct of phase I study of Caelyx (PEGylated liposomal doxorubicin, 25–40 mg/m$^2$) in combination with cyclophosphamide (750–1000 mg/m$^2$) and vincristine (1.2 mg/m$^2$) every 21 days in patients with relapsed or refractory small cell lung cancer [442]. The suggested doses were CaelyxTM 35 mg/m$^2$, cyclophosphamide 750 mg/m$^2$, and vincristine 1.2 mg/m$^2$ intravenously every 21 days. This combination was well tolerated

while antitumor activity was observed for patients with relapsed small cell lung carcinoma with survival duration of 5 months. The latter is similar to what is achieved using only camptothecin analogues taxanes or vinorelbine (survival duration ≤6 months).

Also, paclitaxel is the most widely used anticancer agent for non–small cell lung cancer [420]. Liposomal encapsulated paclitaxel faces the problem of formulation due to the drug's high hydrophobicity. However, a phase I trial with liposomal paclitaxel reported dose-limiting toxicity at the dose of 150 mg/m$^2$/week. Besides, the whole blood clearance of paclitaxel was similar for liposomal and free paclitaxel.

Promising results that were referred to liposomal encapsulated paclitaxel easy to use (LEP-ETU; NeoPharm) consisted of DSPC/Chol/cardiolipin molar ratio 90:5:5, lipid/drug molar ratio 33:1, and paclitaxel concentration 2 mg/mL (mean liposome size 150 nm). In a phase I study of increasing doses of liposomal paclitaxel (135–375 mg/m$^2$) in 25 patients, the enhanced effectiveness was proved with much better tolerability and with 3 partial remissions and 11 patients with stable disease. Lurtotecan, a camptothecin analogue, was incorporated in PC/Chol (2:1 molar ratio) liposomes (size 50–100 nm) with a lipid/drug molar ratio 20:1. A remarkable 1500-fold AUC increase was obtained after administration in nude mice and several xenograft models.

9-Nitro-20(*S*)-camptothecin (9NC), another lipophilic camptothecin analogue, showed antitumor effects in mice and milder effect in humans after oral administration. Thus, the potency of 9NC-loaded di-lauryl-PC (DLPC) liposomes was investigated in patients with primary or metastatic lung cancer after aerosol administration (aerosol droplet size 1–3 μm) for five consecutive days per week [373]. Indeed, the most serious side effect of 9NC, hematological toxicity, did not appear in any of the patients, while the dose-limiting toxicity, chemical pharyngitis, was observed at a dose of 26.6 μg/kg/day. 9NC plasma levels, after aerosol administration (13.3 μg/kg/day), were similar to that given orally (2 mg/m$^2$), despite the lower dose administered in the first case. More specifically, $C_{max}$ and AUC were 76.7 ± 39.1 ng/mL and 275 ± 149 ng-h/mL, respectively, after aerosol administration and 111 ng/mL and 194.4 ± 108.4 ng-h/mL, respectively, after oral administration. The recommended dose for phase II studies is 13.3 μg/kg/day on a 1-h exposure for five consecutive days per week and for eight weeks, with 0.4 mg/mL 9NC concentration in the nebulizer.

Moreover, targeting moieties have been grafted on the vesicle surface, exploiting the fact that antigen/receptor overexpression on cancer cell membranes increases the specificity of the active substance. For example, antagonist-G is a hexapeptide which blocks the action of multiple mitogenic neuropeptides by interacting with their receptors [170]. Antagonist-G was chemically attached at the distal end of PEG molecules of stealth liposomes (HSPC/Chol/mPEG–DSPE/PDP–PEG–DSPE, 2:1:0.08:0.02 molar ratio) (SLG) loaded with doxorubicin. The antiproliferative activity of doxorubicin was estimated on the human variant small cell lung carcinoma (SCLC) H82 cell line. Indeed, 20- to 30-fold increase of both, binding, and internalization took place after SLG incubation on cells in comparison to stealth liposomes only, or PEGylated liposomes without the antagonist-G. The 0.03 μg of antagonist-G on the liposome surface was enough to cause 50% cell liposome association. Doxorubicin accumulation in the whole cell was 20-fold higher with SLG. Eventually, there was the suggestion that the main antagonist-G mechanism of

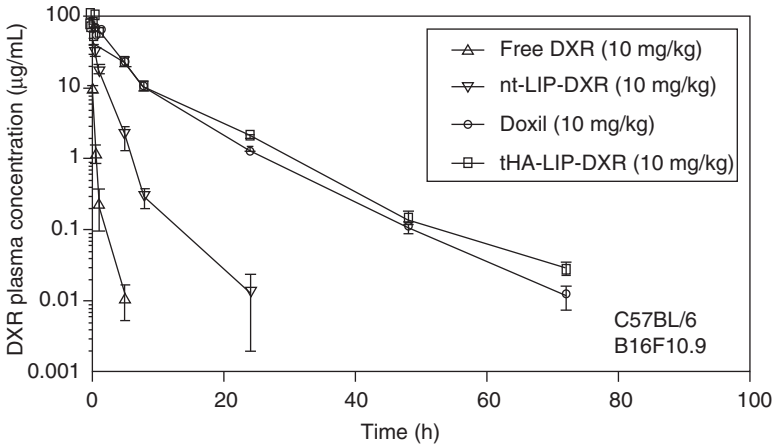


(*a*)

**FIGURE 14** Doxorubicin (DXR) plasma concentration (μg/mL) as function of time from dosing: (*a*) C57BL/6 mice inoculated (by intraveneous injection) with B16F10.9 cells; (*b*) healthy C57BL/6 mice; (*c*) BALB/c mice inoculated with C-26 cells (injected into right-hind footpad). A single dose of the selected formulation was injected into the tail vein. DXR formulations and doses are specified. (Reprinted from ref. 443 with permission of Neoplasia Press, Inc.)

action is to bind to vasopressin receptor, which is expressed abundantly on the SCLC cells.

Hyaluronan (HA) is another potential candidate for tumor targeting because it has been proven that hyaluronan receptors CD44 and RHAMM are overexpressed on several tumor types [443]. In this case, hyaluronan is used not only as a targeting moiety but also to prolong the circulation half-life of the vesicles in question; in other words, it replaces PEG molecules. HA was chemically attached on the vehicle surface (57 μg/μmol lipid), which was comprised of PC/PE/Chol (3:1:1) (HA–LIP). Doxorubicin was entrapped in 78 ± 5% encapsulation efficiency, while the ζ potential and size of the targeted liposomes were –13.1 ± 3.9 mV and 81 ± 13. Pharmacokinetic/biodistribution studies were carried out with Doxil (PEGylated liposomal doxorubicin) as a comparison to the new hyaluronan liposomes. Those studies plus therapeutic responses were tested on three mice models, one of which was C57Bl/6-bearing B16F10.9 lung metastasis. Clearly, ha-liposomal (HA–LIP)–DXR exhibited prolonged blood circulation similar to Doxil (approved formulation in the market), which indicated that the amount of attached HA on the vehicle surface is enough to stabilize and offer the steric stabilization required (Figure 14). Also, accumulation of HA–LIP–DXR was much more reduced in liver of C57Bl/6-bearing B16F10.9 lung metastasis mice, while DXR accumulation in the tumor site was threefold higher to that achieved using Doxil. Significant improvement was obtained in both metastatic burden and survival with Doxil and HA–LIP–DXR (Figure 15). Besides, the HA–LIP–DXR had positive results on all tumor types tested.

Manipulating the genetic material by injecting tumor-suppressing genes has been an alternative way in cancer treatment. The 3p *FUS1* gene is a tumor suppressor

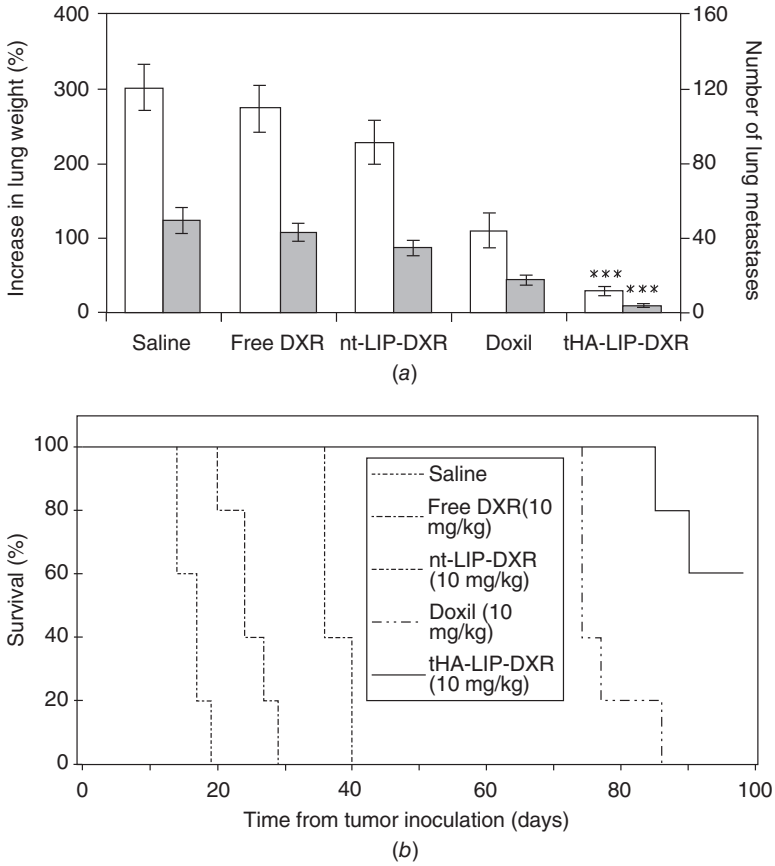


**FIGURE 15** Therapeutic responses of mice bearing B16F10.9-originating lung metastatic disease. Doxobubicin (DXR) formulations and doses are specified. Treatments were on days 1, 5, and 9 by injection of selected formulation to the tail vein. (*a*) Lung metastatic burden. Light-shaded bars are data for increase in lung weight; dark-shaded bars are data for number of lung metastasis: (∗∗∗) $P < 0.001$ compared with free drug. (*b*) Survival ($n = 5$). Each line connects the symbols representing the daily survival state of the group. (Reprinted from ref. 443 with permission of Neoplasia Press, Inc.)

gene which belongs to a 120-kb region of the 3p chromosome on the region 3p21.3. That 120-kb region is missing so injection of DOTAP/Chol lipoplexes with the *FUS1* gene could induce apoptosis and inhibit tumor growth [444]. A significant inhibition of lung metastatic tumor was obtained in animals (bearing subcutaneous lung tumor xenografts) treated with liposome–FUS1 DNA complex (total of six doses). Also, intravenous treatment of lung tumor-bearing animals with DOTAP/Chol–FUS1 complex prolonged the animal survival; 40% of the animals were still alive after approximately 100 days. Also, the antitumor potency of FUS1 was demonstrated to be superior to the one obtained with p53 as the same therapeutic effect was achieved by using less amount of *FUS1* gene in the lipoplexes in comparison to p53. This would be an advantage because lowering the DNA dose would result in much lower toxicity of DNA due to inflammatory response.

Moreover, the efficiency of gene therapy depends very much on the vector used to achieve the gene delivery, the properties and stability of the final lipoplexes in the presence or absence of serum, and the pharmacokinetics and biodistribution. On this basis, Li et al. studied the effect of lipoplex size on the lipofection efficiency of TFL-3/pDNA (plasmid-encoding luciferase) liposomes in the absence and presence of serum and investigated the correlation between in vitro and in vivo results using either B16BL6 murine melanoma cell line or the same cells injected i.v. in C57BL/6 mice in order to produce pulmonary metastases [445]. The authors incubated a range of different ratios of pDNA/TFL-3 (P/L) ranging from 8 to 120 P/L on the previously mentioned cell line in the absence and presence of serum. Serum had a dramatic effect on lipoplex size; in the absence of serum, the size increased as the pDNA amount increased and it reached a maximum value at the ratio 80 g/mol P/L, while the ζ potential decreased as the P/L ratio increased and at 80 g/mol P/L was approximately +2 mV. Maximum luciferase activity was observed with lipoplexes of 80 g/mol. In contrast, in the presence of serum, the size increased from 8 to 20 g/mol and decreased from 40 to 120 g/mol as the P/L increased. Maximum luciferase activity was obtained with 40 g/mol in the presence of serum. However, the in vitro transfection efficiency of the lipoplexes in the presence of serum at the highest point of 40 g/mol was twofold less than the one achieved at the peak of transfection efficiency at 80 g/mol in the absence of serum. However, the plasmid expression was higher with 8 and 80 g/mol lipoplexes in tumor-bearing animals in contrast to the in vitro findings in the presence of serum. This was attributed to less aggregate formation in the blood due to the lipoplex/serum ratio (50–100 μL serum/200 μL lipoplex dispersion) so there is no complete interaction between lipoplexes and serum proteins. Also, pulmonary gene expression was dependent on the time after cell inoculation. It was shown that gene expression takes place at different parts of the lung; for example, 8 g/mol P/L lipoplexes expressed luciferase in the cells surrounding the tumor, while 80 g/mol lipoplexes expressed the gene in the entire lung without any specificity. Possible explanations for the latter are the extended lung capillary bed and increased lipoplex uptake by tumor cells compared to the normal cells. Of course many other factors could contribute to the different patterns of gene expression, such as endosomal release, nuclear uptake, increased translation, and transcription.

All-trans retinoic acid (ATRA) has shown anticancer activity in a number of types of cancer cells [446]. However, some non–small cell lung carcinoma is resistant to ATRA probably due to the high lipophilicity of the molecule, which makes it unable to pass through the cellular membrane. Therefore, cationic liposomes (DOTAP/Chol 1:1 molar ratio) were used to incorporate and facilitate ATRA's action in the resistant NSCLC in A549 human lung cancer cells. The produced ATRA–DOTAP/Chol lipoplexes were 125 nm in diameter with 50 mV of ζ potential, while for the DSPC/Chol liposomes used as control these were 110 nm and –3 mV, respectively. The former lipoplexes exhibited higher uptake by the cells in comparison to DSPC/Chol due probably to the positive surface charge, which interacts to a higher degree with the negatively charged cell membrane. Thus, the apoptosis induced on these cells in the presence of ATRA lipoplexes was much higher than the one caused by ATRA only or ATRA–DSPC/Chol (Figure 16).

Lastly, although antisense therapy is considered a simple and efficient procedure, its products have never reached the market. This is due to the instability of phos-

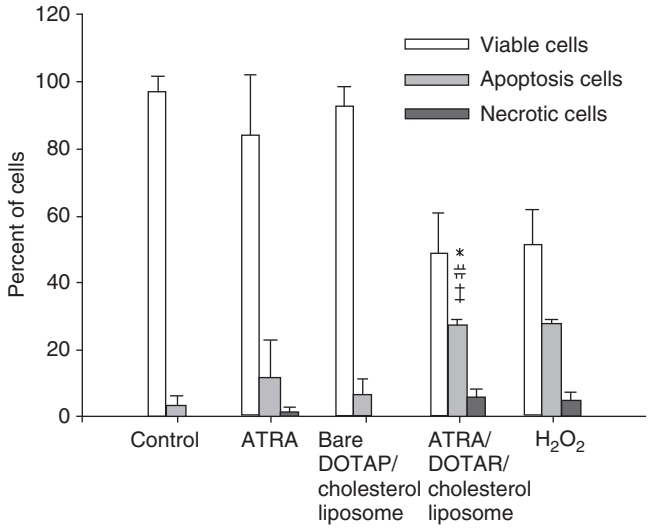


**FIGURE 16** Flow cytometric analysis of A549 cells treated with 1.0 μ*M* ATRA, bare DOTAP/cholesterol liposomes, or ATRA incorporated in DOTAP/cholesterol liposomes for 48 h in A549 cells. As a control. Cells were incubated with 1% DMSO and 400 μM $H_2O_2$. Significant differences: (*) $P < 0.05$ vs. control; (#) $P < 0.05$ vs. ATRA; (‡) $P < 0.05$ vs. DOTAP/cholesterol liposomes. (Reprinted from ref. 446 after permission of Elsevier.)

phodiester oligonucleotides (PO-ONs) in the cytoplasm, while the phosphothioate ONs demonstrate unwanted side effects [447]. In order to design better delivery systems for PO-ONs, the intracellular fate of the PO-ON and PS-ON lipoplexes with DOTAP/DOPE was investigated after their application on A549 cells. After endocytosis, endosomal localization, and endosomal escape of the lipoplexes, the ONs were localized in the nuclei. However, PO-ONs were degraded in the nucleus (degradation products diffuse out of the cells after 2 h) whereas PS-ONs were still intact and remained in the nucleus for 8 h. It is worth mentioning that PS-ONs were eliminated from the nucleus and were found in cytoplasmic granules, which indicated that the cell has a mechanism of elimination of intact PS-ONs. Also, it seemed that the amount of PO-ONs was quite important for the ONs' stability as the degradation was reduced after injecting 22 μ*M*, in contrast to complete degradation with injection of 2 μ*M*. From the study it was concluded that ON degradation happened after their release from the lipoplexes. Thus, for the successful delivery of ONs, they have to remain complexed with their carrier. Polyethylenimine (PEI) was suggested as a possible carrier due to its proton sponge effect, leading to endosomal rupture without releasing the ONs as well as graft-pDMAEMA–PEG [poly(2-dimethylamino)ethyl methacrylate-*co*-aminoethyl methacrylate–bearing polyethylene glycol chains]. According to the authors, probably cationic polymers are more efficient than cationic lipids, but lots of work still needs to be done on the matter.

***Ovarian Cancer*** Ovarian cancers start at the ovaries [448]. They can be either benign, and so never spread from the ovary, or malignant, in which case they can

metastasize to other parts of the body. Malignant cancers are classified into germ cell, stromal, and epithelial tumors. Germ cells produce eggs, stromal cells produce progesterone and estrogen, and epithelial cells cover the ovary. The majority of the malignant cases are epithelial cancer (almost 90%). As with most tumors, the usual treatment is surgery, chemotherapy, and radiotherapy.

Chemotherapy refers to drug administration with highly serious side effects, such as nausea, hand and foot rashes, mouth sores, and increased risk of infection, easy bruising, and so on. Therefore, liposomal carriers have been used in order to improve the drug's biodistribution and protect the patient from those side effects. The main anticancer drugs used to treat ovarian cancer are carboplatin and cisplatin, paclitaxel, topotecan, and lurtotecan. PEGylated liposomal doxorubicin has been approved as a regimen for patients with metastatic ovarian cancer refractory to both paclitaxel and platinum based-therapy [449].

A number of clinical trials have been reported and many refer to a combined therapeutic regimen of liposomes with other anticancer drugs. PEGylated liposomal doxorubicin was used in a phase II trial on patients with platinum/paclitaxel pretreated ovarian cancer with or without topotecan. Approximately 36% of patients showed stable disease [420]. A phase III study followed up the one already mentioned in 474 patients with pretreated epithelial ovarian cancer that failed or recurred upon platinum-based combination chemotherapy. The patients received either PEGylated liposomal doxorubicin (50 mg/m$^2$ for 1 day every 4 weeks) or topotecan (1.5 mg/m$^2$ for 1–5 days every 3 weeks). The response rates and the overall progression-free survival (PFS) were similar for the two formulations. The platinum-sensitive patients showed longer PFS and overall survival of 108 weeks against the 72 weeks obtained using topotecan. The interesting result of the study was the reduced cytotoxicity recorded for PEG–liposomal DXR compared to topotecan, proving that the former treatment improves patient quality of life.

A phase III study was conducted to compare PEG–liposomal DXR (50 mg/m$^2$ every 4 weeks) with paclitaxel (175 mg/m$^2$ every 3 weeks) using 214 patients with relapse after first-line platinum-based chemotherapy [420]. As previously, the response rates and PFS were not significantly different, but again the liposomal formulation was notably less toxic as fewer patients recorded with grade 4 adverse effects (17% compared to 71% for topotecan) and thus was more tolerable.

The most common toxic effects associated with liposomal DXR treatment were hand-and-foot syndrome and stomatitis, which can be handled by modifying the dose so there is no need for the regimen to cease. For example, administration of 40 mg/m$^2$ liposomal DXR every 4 weeks reduces the incidences of hand-and-foot syndrome and stomatitis compared to 50 mg/m$^2$ liposomal DXR every 4 weeks without loss of drug potency [450]. A series of clinical trials using patients with a variety of ovarian cancers in terms of characteristics (relapsed, refractory, platinum, paclitaxel resistant) were carried out. The optimized dose regimen for PEGylated liposomal doxorubicin was 10–12.5 mg/m$^2$ per week when it is used as a single therapy. Combining liposomal DXR with other anticancer drugs is an equally or more effective strategy due to the lower dosage regimen of the two formulations avoiding relevant toxicities.

While platinum compounds and paclitaxel comprise the first-line treatment in combination with PEGylated liposomal doxorubicin for patients with ovarian

cancer, vinorelbine has been used in combination with liposomes as a palliative second-line therapy for patients with refractory/resistant ovarian cancer to platinum–paclitaxel [451]. The best MTD (mean therapeutic dose) was 25 mg/m$^2$ at days 1 and 8 for vinorelbine and 30 mg/m$^2$ at day 1 for liposomal DXR of every 21 days, which was well tolerated with moderate hematologic and mild nonhematologic toxicities. A phase II study relevant to the clinical efficacy, toxicity, and pharmacokinetics of that combined therapeutic regimen was carried out by Katsaros et al. in 30 patients with platinum–paclitaxel pretreated recurrent ovarian cancer [452]. Caelyx (30 mg/m$^2$) and vinorelbine (30 mg/m$^2$) were administered every 3 weeks for six cycles. The regimen was proved of significant activity for patients pretreated with paclitaxel–platinum first-line therapy. The overall response rate was 37% with 10% of patients demonstrating stable disease. Vinorelbine bioavailability was higher under the current regimen. The overall survival was 9 months while the toxicity was mild and reversible. There were no treatment-related deaths and there were only 2 patients, one reported with grade 4 and the other with grade 3 hand–foot syndrome. Also, the toxicity due to liposomal formulation was much lower compared to that reported from a phase III study with the drug given as a single agent [453]. In another phase II study, the combination of liposomal doxorubicin and infused topotecan was studied in 27 patients with platinum-resistant ovarian cancer in two cohorts [454]. Liposomal DXR (30 mg/m$^2$ at day 1) and topotecan (1 mg/m$^2$ for 5 days) were infused and the cycle was repeated every 21 days. The overall response rate of the regimen was 28% and the median overall survival was 40 weeks. However, neutropenia and thrombopenia were observed at 70 and 41% of the patients. Therefore, the topotecan dose was reduced to 0.75 mg/m$^2$ and liposomal DXR was increased to 40 mg/m$^2$ in order to reduce the toxicity. After that, the cytotoxic effect due to liposomal doxorubicin increased and the bone marrow cytotoxicity remained the same despite the effectiveness of the regimen.

Lurtotecan is a more advanced camptothecin analogue with probably greater potency with regard to topotecan and therefore was encapsulated in liposomes to investigate the toxicity and pharmacokinetics [455]. In a multi-institutional open-label phase II study, 22 patients with topotecan-resistant ovarian cancer were administered liposomal lurtotecan in a dose of 2.4 mg/m$^2$ on days 1 and 8 every 21 days. Although the toxicity profile of the drug was lower, no response was observed. Others used a regimen of liposomal delivery at days 1 and 3 with more promising therapeutic results, but higher toxicity, too [456].

A variety of new molecules either in combination with liposomal doxorubicin or not are in development at the moment [457]. For example, a phase III study will be conducted to test the efficacy and safety of pattupilone versus PEG–liposomal DXR in taxane/platinum refractory/resistant patients with recurrent epithelial ovarian, primary fallopian, or primary peritoneal cancer. A phase III randomized study of Telcyta with Doxil/Caelyx versus Doxil/Caelyx has been planned in patients with platinum-refractory or platinum-resistant ovarian cancer. A phase II study relevant to side effects and best dose of ixabepilone combined with liposomal DXR will be assessed in patients with advanced ovarian epithelial, peritoneal cavity, or fallopian tube cancer or metastatic breast cancer.

Gemcitabine is a clinically active antineoplastic drug in platinum-refractory ovarian cancer. The efficacy and tolerability of the particular drug in combination with liposomal DXR were investigated in athymic mice bearing cisplatin-resistant human ovarian carcinoma [458].

Using two therapeutic regimens, either 80 mg/kg of gemcitabine and 15 mg/kg of liposomes or 20 mg/kg of gemcitabine and 6 mg/kg of liposomes, the same trend of response was observed, with some of the animals having complete tumor regression at the end of the study. The lack of toxicity and therapeutic efficacy observed makes that regimen promising for clinical trials.

In terms of drug development, a novel analogue of vitamin E assembled into liposomes was evaluated as a potent anticancer agent in combination with cisplatin in mice bearing human ovarian cancer xenografts [459]. The analogue is the 2,5,7,8-tetramethyl-2*R*-(4*R*, 8*R*, 12-trimethyltridecyl)chimoran-6-yloxyacetic acid (α-TEA), which has apoptotic properties to cancer and not the normal cells in a dose- and time-dependent manner. Liposomes were administered in the form of aerosol while cisplatin was injected intraperitoneally to mice with either small or large tumor volume. It was shown that the combined therapeutic scheme demonstrated the highest antitumor activity compared to α-TEA itself in both cases. Also, the current regimen significantly reduced micrometastases observed in lungs and lymph nodes, while the ratio of proliferating to apoptotic cells in the tumor was decreased due to induction of apoptosis. However, the specific apoptotic mechanism of the particular molecule needs to be elucidated.

By grafting folate molecules on the liposome surface, their accumulation is highly increased in macrophages located in tumor ascites fluid but not in solid tumors, which have been tested and the existence of functional folate receptors has been confirmed [460]. Folate receptors are overexpressed on epithelial tumor cells, and thus folate was attached on PEG molecules via cysteine (folate–cysteine–PEG3400–PE), which, consequently, was one of the components of the liposomal bilayer (PC/Chol/PEG–PE). The ligand-bearing liposomes (diameter of 65–90 nm) were injected i.p. in tumor-bearing mice. Significant folate-bearing liposome accumulation was obtained in the ascites fluid and more specifically by macrophages, which indicates macrophage activation. A reduced targeted liposome accumulation in the tumor could occur for a number of reasons, such as poor liposome penetration into solid tumor mass. Activated macrophages secrete immunosuppressive cytokines and angiogenic factors, so liposome targeting could possibly eliminate them during malignancy therapy.

Gene delivery is another approach trying to tackle the problem of cancer. Mutations of p53 tumor suppressor gene contribute to genetic abnormalities in ovarian cancer. Kim et al. developed a nonviral vector for delivery of p53 in ovarian cancer cells (OVCAR-3) [461]. The nonviral (liposomal) vector consisted of DOTAP, DOPE, and Chol in the molar ratio 1:0.7:0.3. High expression of p53 mRNA and proteins in OVCAR-3 cells indicated successful transfection of the lipoplexes to the cells used. The latter was indicated with the inhibition of cell growth obtained in OVCAR-3 cells due to apoptosis caused by that protein. Intratumoral injection of DDC/pp53-GFAP lipoplexes into mice bearing the OVCAR-3 cells clearly showed the tumor growth inhibition, which suggests the therapeutic efficiency of the particular lipoplexes.

The neutral 1,2-dioleoyl-*sn*-glycerol-3-phosphatidylcholine (DOPC) was used to make complexes with the siRNA targeting the oncoprotein EphA2 and then injected i.v. into ovarian tumor-bearing mice [462]. According to the authors, the significance of the particular study is the formulation of siRNA in liposomes with successful outcome after delivery of the loaded carrier, which showed significant reduced protein expression and tumor growth (Figure 17).

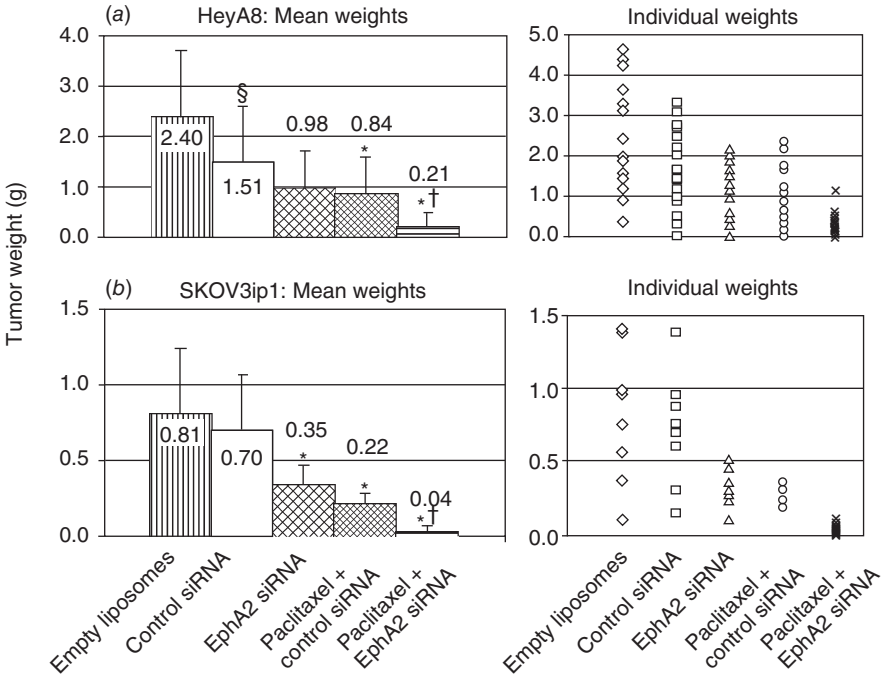


* $p<0.05$ compared to control siRNA

† $p<0.05$ compared to control siRNA plus paclitaxel

§ $p<0.05$ compared to empty liposomes

**FIGURE 17** Therapeutic efficacy of siRNA-mediated EphA2 down regulation. A and B nude mice were injected i.p. with (*a*) 2.5 × 105 HeyA8 cells or (*b*) 1.0 × 106 SKOV3ip1 cells and randomly allocated to one of five groups, with therapy beginning 1 week after cell injection. (a) Empty DOPC liposomes, (b) control siRNA in DOPC, (c) EphA2-targeting siRNA in DOPC, (d) paclitaxel + control siRNA in DOPC, or (e) paclitaxel + EphA2 siRNA in DOPC. siRNA liposomes were injected twice weekly at a dose of 150 μg/kg siRNA. Paclitaxel (100 μg) or vehicle (first three groups) was injected i.p. once weekly. (Reprinted from ref. 462 and with permission of the American Association for Cancer Research.)

## REFERENCES

1. Bangham, A. D., Standish, M. M., and Watkins, J. C. (1965), Diffusion of univalent ions across the lamellae of swollen phospholipids, *J. Mol. Biol.*, 13, 238–252.
2. Papahadjopoulos, D. (1978), Liposomes and their uses in biology and medicine, *Ann. N.Y. Acad. Sci.*, 408, 1.
3. Gregoriadis, G. (1978), Liposomes in therapeutic and preventive medicine: The development of the drug-carrier concept, *Ann. N.Y. Acad. Sci.*, 308, 343–370.
4. Kirby, C., Clarke, J., and Gregoriadis, G. (1980), Cholesterol content of small unilamellar liposomes controls phospholipid loss to high density lipoproteins in the presence of serum, *FEBS Lett.*, 111, 324–328.
5. Pain, D., Das, P. K., Ghosh, P. C., and Bachhawat, B. K. (1984), Increased circulatory half-life of liposomes after conjugation with dextran, *J. Biosci.*, 6, 811–816.

6. Allen, T. M., and Chonn, A. (1987), Large unilamellar liposomes with low uptake into the reticuloendothelial system, *FEBS Lett.*, 223, 42–46.
7. Papahadjopoulos, D., Allen, T. M., Gabizon, A., Mayhew, E., Matthay, K., Huang, S. K., et al. (1991), Sterically stabilized liposomes: Improvements in pharmacokinetics and antitumor therapeutic efficacy, *Proc. Natl. Acad. Sci. USA*, 88, 11460–11464.
8. Lasic, D. D., Martin, F. J., Gabizon, A., Huang, S. K., and Papahadjopoulos, D. (1991), Sterically stabilized liposomes: A hypothesis on the molecular origin of the extended circulation times, *Biochim. Biophys. Acta*, 1070, 187–192.
9. Torchilin, V. P., Levchenko, T. S., Whiteman, K. R., Yaroslavov, A. A., Tsatsakis, A. M., Rizos, A. K., et al. (2001), Amphiphilic poly-*N*-vinylpyrrolidones: Synthesis, properties and liposome surface modification, *Biomaterials*, 22, 3035–3044.
10. Takeuchi, H., Kojima, H., Yamamoto, H., and Kawashima, Y. (2001), Evaluation of circulation profiles of liposomes coated with hydrophilic polymers having different molecular weights in rats, *J. Controlled Release*, 75, 83–91.
11. Haran, G., Cohen, R., Bar, L., and Barenholz, Y. (1993), Transmembrane ammonium sulphate gradients in liposomes produce efficient and stable entrapment of amphipathis weak bases, *Biochim. Biophys. Acta*, 1151, 201–215.
12. Loan Honeywell-Nguyen, P., Wouter, H. W., Groenink, A., de Graaff, M., and Bouwstra, J. A. (2003), The in vivo transport of elastic vesicles into human skin: Effects of occlusion, volume and duration of application, *J. Controlled Release*, 90, 243–255.
13. Lasic, D. (1993), *Liposomes from Physics to Applications*, Elsevier, London.
14. Hauser, H. (2000), Short-chain phospholipids as detergents, *Biochim. Biophys. Acta*, 1508, 164–181.
15. Eibl, H. (1980), Synthesis of glycerophospholipids, *Chem. Phys. Lipids*, 26, 405–429.
16. Barenholz, Y., Thompson, T. E. (1999), Sphingomyelin: Biophysical aspects, *Chem. Phys. Lipids*, 102, 29–34.
17. Barenholz, Y. (2004), Sphingomyelin and cholesterol: From membrane biophysics and rafts to potential medical applications, *Subcell. Biochem.*, 37, 167–215.
18. Patel, G. B., and Sprott, G. D. (1999), Archaeobacterial ether lipid liposomes (archaeosomes) as novel vaccine and drug delivery systems, *Crit. Rev. Biotechnol.*, 19, 317–357.
19. Fatouros, D., Ioannou, P. V., and Antimisiaris, S. G. (2006), Novel nanosized arsenic containing vesicles for drug delivery: Arsonoliposomes, *J. Nanosci. Nanotechnol.*, 6, 2618–2687.
20. Biltonen, R. L., and Lichtenberg, D. (1993), The use of differential scanning calorimetry as a tool to characterize liposome preparations, *Chem. Phys. Lipids*, 64, 129–142.
21. Champan, D., Urbina, J., and Keough, K. M. (1974), Studies of lipid water systems using differential scanning calorimetry, *J. Biol. Chem.*, 249(8), 2512–2521.
22. van Osdol, W. W., Johnson, M. L., Ye, Q., and Biltonen M. (1991), Relaxation dynamics of the gel to liquid-crystalline transition of phosphatidylcholine bilayers. Effects of chain length and vesicle size, *Biophys. J.*, 59, 775–785.
23. van Osdol, W. W., Ye, Q., Johnson, M. L., and Biltonen, M. (1992), Effects of the anesthetic dibucaine on the kinetics of the gel-liquid crystalline transition of dipalmitoylphosphatidylcholine multilamellar vesicles, *Biophys. J.*, 63, 1011–1017.
24. Momo, F., Fabris, S., Bindoli, A., Scutari, G., and Stevanato, R. (2002), Different effects of propofol and nitrosopropofol on DMPC multilamellar liposomes, *Biophys. Chem.*, 95, 145–155.

25. Zhang, F., and Rowe, E. S. (1994), Calorimetric studies of the interactions of cytochrome c with dioleoylphosphatidylglycerol extruded vesicles: Ionic strength effects, *Biochim. Biophys. Acta*, 1193, 219–225.
26. Lo, Y. L., and Rahman, Y. E. (1995), Protein location in liposomes, a drug carrier: A prediction by differential scanning calorimetry, *J. Pharm. Sci.*, 84, 805–814.
27. Parente, R. A., and Lentz, B. R. (1985), Advantages and limitations of 1-palmitoyl-2-[[2-[4-(6-phenyl-trans-1,3,5-hexatrienyl)phenyl]ethyl]carbonyl]-3-sn-phosphatidylcholine as a fluorescent membrane probe, *Biochemistry*, 24, 6178–6185.
28. Metso, A. J., Zhao, H., Tuunainen, I., and Kinnnunen, P. K. (2004), Characterization of the main transition of dinervonoylphosphocholine liposomes by fluorescence spectroscopy, *Biochim. Biophys. Acta*, 1663, 222–231.
29. van Langen, H., Van Ginkel, G., Shaw, D., and Levine, Y. K. (1989), The fidelity of response by 1-[4-(trimethylammonio)phenyl]-6-phenyl-1,3,5-hexatriene in time-resolved fluorescence anisotropy measurements on lipid vesicles. Effects of unsaturation, head-group and cholesterol on orientational order and reorientational dynamics, *Eur. Biophys. J.*, 17, 37–48.
30. Davenport, L., and Targowski, P. (1996), Submicrosecond phospholipid dynamics using a long-lived fluorescence emission anisotropy probe, *Biophys. J.*, 71, 1837–1852.
31. de Kruijff, B., and Cullis, P. R. (1980), The influence of poly(L-lysine) on phospholipid polymorphism. Evidence that electrostatic polypeptide-phospholipid interactions can modulate bilayer/non-bilayer transitions, *Biochim. Biophys. Acta*, 601, 235–240.
32. Vasilenko, I., de Kruijff, B., and Verkleij, A. J. (1982), The synthesis and use of thion-phospholipids in 31P-NRM studies of lipid polymorphism, *Biochim. Biophys. Acta*, 685, 144–152.
33. Perkins, W. R., Mincey, S. R., Ostro, M. J., Taraschi, T. F., and Janoff, A. S. (1988), The captured volume of multilamellar vesicles, *Biochim. Biophys. Acta*, 943, 103–107.
34. Davis, J. H. (1983), The description of membrane lipid conformation, order and dynamics by 2H-NMR, *Biochim. Biophys. Acta*, 737, 117–171.
35. Wu, W. G., Dowd, S. R., Simplaceanu, V., Peng, Z. Y., and Ho, C. (1985), 19F NMR investigation of molecular motion and packing in sonicated phospholipid vesicles, *Biochemistry*, 24, 7153–7161.
36. Bertoli, E., Masserini, M., Sonnino, S., Ghidoni, R., Cestaro, B., and Tettamanti, G. (1981), Electron paramagnetic resonance studies on the fluidity and surface dynamics of egg phosphatidylcholine vesicles containing gangliosides, *Biochim. Biophys. Acta*, 647, 196–202.
37. Shin, Y. K., and Freed, J. H. (1989), Dynamic imaging of lateral diffusion by electron spin resonance and study of rotational dynamics in model membranes. Effect of cholesterol, *Biophys. J.*, 55, 537–550.
38. Lai, C. S., Joseph, J., and Shih, C. C. (1989), Molecular dynamics of antitumor ether-linked phospholipids in model membranes: A spin-label study, *Biochem. Biophys. Res. Commun.*, 160, 1189–1195.
39. Ondrias, K., Stasko, A., Marko, V., and Nosal, R. (1989), Influence of beta-adrenoceptor blocking drugs on lipid-protein interaction in synaptosomal membranes. An ESR study, *Chem. Biol. Interact.*, 69, 87–97.
40. Hoekstra, D. (1982), Fluorescence method for measuring the kinetics of $Ca^{2+}$-induced phase separations in phosphatidylserine-containing lipid vesicles, *Biochemistry*, 21, 1055–1061.
41. Hoekstra, D. (1990), Fluorescence assays to monitor membrane fusion: Potential application in biliary lipid secretion and vesicle interactions, *Hepatology*, 12, 61S–66S.

42. LeNeveu, D. M., Rand, R. P., Parsegian, V. A., and Gingel, D. (1976), Apparent modification of forces between lecithin bilayers, *Science*, 191, 399–400.
43. Franks, N. P. (1976), Structural analysis of hydrated egg lecithin and cholesterol bilayers. I. X-ray diffraction, *J. Mol. Biol.*, 100, 345–358.
44. Worcester, D. L., and Franks, N. P. (1976), Structural analysis of hydrated egg lecithin and cholesterol bilayers. II. Neutron diffraction, *J. Mol. Biol.*, 100, 359–378.
45. LeNeveu, D. M., and Rand, R. P. (1977), Measurement and modification of forces between lecithin bilayers, *Biophys. J.*, 18, 209–230.
46. Hobbs, S. K., Monsky, W. L., Yuan, F., Roberts, W. G., Griffith, L., Torchillin, V. P., et al. (1998), Regulation of transport pathways in tumor vessels: Role of tumor type and microenvironment, *Proc. Natl. Acad. Sci. USA*, 95, 4607–4612.
47. Drummond, D. C., Meyer, O., Hong, K., Kirpotin, D. B., and Papahadjopoulos, D. (1999), Optimizing liposomes for delivery of chemotherapeutic agents to solid tumour, *Pharmacol. Rev.*, 51, 691–743.
48. Devine, D. V., Wong, K., Serrano, K., Chonn A., and Cullis, P. R. (1994), Liposome-complement interactions in rat serum: Implications for liposome survival studies, *Biochim. Biophys. Acta*, 1191, 43–51.
49. Kolchens, S., Ramaswami, V., Birgenheier, J., Nett, L., and O'Brien, D. F. (1993), Quasi-elastic light scattering determination of the size distribution of extruded vesicles, *Chem. Phys. Lipids*, 65, 1–10.
50. Lesieur, S., Grabielle-Madelmont, C., Paternostre, M. T., and Ollivon, M. (1991), Size analysis and stability study of lipid vesicles by high-performance gel exclusion chromatography, turbidity, and dynamic light scattering, *Anal. Biochem.*, 1, 192, 334–343.
51. Matsuzaki, K., Murase, O., Sugishita, K., Yoneyama, S., Akada, K., Ueha, M., Nakamura, A., and Kobayashi, S. (2000), Optical characterization of liposomes by right angle light scattering and turbidity measurement, *Biochim. Biophys. Acta*, 31, 219–226.
52. Moon, M. H., and Giddings, J. C. (1993), Size distribution of liposomes by flow field-flow fractionation, *J. Pharm. Biomed. Anal.*, 11, 911–920.
53. Korgel, B. A., van Zanten, J. H., and Monbouquette, H. G. (1998), Vesicle size distributions measured by flow field-flow fractionation coupled with multiangle light scattering, *Biophys. J.*, 74, 3264–3272.
54. Lee, H., Williams, S. K., Allison, S. D., and Anchordoquy, T. J. (2001), Analysis of self-assembled cationic lipid-DNA gene carrier complexes using flow field-flow fractionation and light scattering, *Anal. Chem.*, 73, 837–843.
55. Grabielle-Madelmont, C., Lesieur, S., and Ollivon, M. (2003), Characterization of loaded liposomes by size exclusion chromatography, *J. Biochem. Biophys. Methods*, 56, 189–217.
56. Goni, F. M., and Alonso, A. (2000), Spectroscopic techniques in the study of membrane solubilization, reconstitution and permeabilization by detergents, *Biochim. Biophys. Acta*, 1508, 51–68.
57. Alonso, A., Villena, A., and Goñi, F. M. (1981), Lysis and reassembly of sonicated lecithin vesicles in the presence of Triton X-100, *FEBS Lett.*, 123, 200–204.
58. Kokona, M., Kallinteri, P., Fatouros, D., and Antimisiaris, S. G. (2000), Stability of SUV liposomes in the presence of cholate salts and pancreatic lipases: Effect of lipid composition, *Eur. J. Pharm. Sci.*, 9, 245–252.
59. Almog, S., Litman, B. J., Wimley, W., Cohen, J., Wachtel, E. J., Barenholz, Y., Ben-Saul, A., and Lichtenberg, D. (1990), States of aggregation and phase transformations in mixtures of phosphatidylcholine and octyl glucoside, *Biochemistry*, 29, 4582–4592.

60. Lopez, O., de la Maza, A., Coderch, L., Lopez-Iglesias, C., Wehrli E., and Parra, J. L. (1998), Direct formation of mixed micelles in the solubilization of phospholipid liposomes by Triton X-100, *FEBS Lett.*, 426, 314–318.
61. Fatouros, D. G., Piperoudi, S., Gortzi, O., Ioannou, P. V., and Antimisiaris, S. G. (2005), Physical stability of sonicated arsonoliposomes: Effect of calcium ions, *J. Pharm. Sci. US*, 94, 46–55.
62. Akashi, K., Miyata, H., Itoh. H., and Kinoshita, K., Jr. (1996), Preparation of giant liposomes in physiological conditions and their characterization under an optical microscope, *Biophys. J.*, 71, 3242–3250.
63. Shohda, K., Toyota, T., Yomo, T., and Saguwara, T. (2003), Direct visualization of DNA duplex formation on the surface of a giant liposome, *Chembiochem.*, 4, 778–781.
64. Henriksen, J. R., and Ipsen, J. H. (2004), Measurement of membrane elasticity by micropipette aspiration, *Eur. Phys. J. E. Soft Matter*, 14, 149–167.
65. Olson, F., Hunt, C. A., Szoka, F. C., Vail, W. J., and Papahadjopoulos, D. (1979), Preparation of liposomes of defined size distribution by extrusion through polycarbonate membranes, *Biochim. Biophys. Acta*, 557, 9–23.
66. Weibull, C., Christiansson, A., and Carlemalm, E. (1983), Extraction of membrane lipids during fixation, dehydration and embedding of Acholeplasma laidlawii-cells for electron microscopy, *J. Microsc.*, 129, 201–207.
67. Dermer, G. B. (1968), An autoradiographic and biochemical study of oleic acid absorption by intestinal slices including determinations of lipid loss during preparation of electron microscopy, *J. Ultrastruct. Res.*, 22, 312–325.
68. Hope, M. J., Walker, D. C., and Cullis, P. R. (1983), $Ca^{2+}$ and pH induced fusion of small unilamellar vesicles consisting of phosphatidylethanolamine and negatively charged phospholipids: A freeze fracture study, *Biochem. Biophys. Res. Commun.*, 110, 15–22.
69. Hope, M. J., Wong, K. F., and Cullis, P. R. (1989), Freeze-fracture of lipids and model membrane systems, *J. Electron. Microsc. Tech.*, 13, 277–287.
70. Burger, K. N., Nieva, J. L., Alonso, A., and Verkleij, A. J. (1991), Phospholipase C activity-induced fusion of pure lipid model membranes. A freeze fracture study, *Biochim. Biophys. Acta*, 1068, 249–253.
71. Meyer, H. W., and Richter, W. (2001), Freeze-fracture studies on lipids and membranes, *Micron.*, 32, 615–644.
72. Taggar, A. S., Alnajim, J., Anantha, M., Thomas, A., Webb, M., Ramsay, E., and Bally, M. B. (2006), Copper-topotecan complexation mediates drug accumulation into liposomes, *J. Controlled Release*, 114, 78–88.
73. Anabousi, S., Laue, M., Lehr, C. M., Bakowsky, U., and Ehrhardt. C. (2005), Assessing transferring modification of liposomes by atomic force microscopy and tramsmission electron microscopy, *Eur. J. Pharm. Biopharm*, 60, 295–303.
74. Bellare, J. R., Davis, H. T., Scriven, L. E., and Talmon, Y. (1988), Controlled environment vitrification system. An improved sample preparation technique, *J. Elec. Micr. Tech.*, 10, 87–111.
75. Dubochet, J., Adrian, M., Chang, J., Homo, J. C., Lepault, J., McDowell, A. W., and Schultz, P. (1988), Cryo-electron microscopy of vitrified specimens, *Q. Rev. Biophys.*, 21, 129–228.
76. Frederik, P. M., Stuart, M. C., and Verklej, A. J. (1989), Intermediary structures during membrane fusion as observed by cryo-electron microscopy, *Biochim. Biophys. Acta*, 979, 275–278.
77. Siegel, D. P., Green, W., and Talmon, J. (1989), Intermediates in membrane fusion and bilayer/nonbilayer phase transitions imaged by time-resolved cryo-transmission electron microscopy, *Biophys J.*, 56, 161–169.

78. Frederik, P. M., Burger, K. N., Stuart, M. C., and Verklej, A. J. (1991), Lipid polymorphism as observed by cryo-electron microscopy, *Biochim. Biophys. Acta*, 1062, 133–141.

79. Siegel, D. P., Green, W., and Talmon, J. (1994), The mechanism of lamellar-to-inverted hexagonal phase transitions: A study using temperature-jump cryo-electron microscopy, *Biophys. J.*, 66, 402–414.

80. Mozafari, M. R., Reed, C. J., Rostron, C., and Hasirci, V. (2005), A review of scanning probe microscopy investigations of liposome-DNA complexes, *J. Liposome Res.*, 15, 93–107.

81. Papadimitriou, E., and Antimisiaris, S. G. (2000), Interactions of PC/Chol and PS/Chol liposomes with human cells in vitro, *J. Drug Target.*, 5, 335–351.

82. Haywood, A. M., and Boyer, B. P. (1986), Ficoll and dextran enhance adhesion of Sendai virus to liposomes containing receptor (ganglioside GD1a), *Biochemistry*, 25, 3925–3929.

83. Shichijo, S., and Alving, C. R. (1986), Inhibitory effects of gangliosides on immune reactions of antibodies to neutral glycolipids in liposomes, *Biochim. Biophys. Acta*, 858, 118–124.

84. Ho, R. J., Rouse, B. T., and Huang, L. (1986), Target-sensitive immunoliposomes: Preparation and characterization, *Biochemistry*, 25, 5500–5506.

85. Bogdanov, A. A., Gordeeva, L. V., Torchilin, V. P., and Margolis, L. B. (1989), Lectin-bearing liposomes: Differential binding to normal and to transformed mouse fibroblasts, *Exp. Cell Res.*, 181, 362–374.

86. Woodle, M. C., and Lasic, D. D. (1992), Sterically stabilized liposomes, *Biochim. Biophys. Acta*, 1113, 171–199.

87. Woodle, M. C., Collins, L. R., Sponsler, E., Kossovsky, N., Papahadjopoulos, D., and Martin, F. J. (1992), Sterically stabilized liposomes. Reduction in electrophoretic mobility but not electrostatic surface potential, *Biophys. J.*, 61, 902–910.

88. Mayhew, E., Lazo, R., and Vail, W. J. (1989), Preparation of liposomes entrapping cancer chemotherapeutic agents for experimental in vivo and in vitro studies, in Gregoriadis, G., Ed., *Liposome Technology*, Vol. II, CRC Press, Boca Raton, FL, Chapter 2.

89. Fatouros, D. G., and Antimisiaris, S. G. (2001), Physicochemical properties of liposomes incorporating hydrochlorothiazide and chlorothiazide, *J. Drug Target.*, 9, 61–74.

90. Fatouros, D. G., and Antimisiaris, S. G. (2002), Effect of amphiphilic drugs on the stability and zeta-potential of their liposome formulations: A study with prednisolone, diazepam and griseofulvin, *J. Coll. Interf. Sci.*, 251, 271–277.

91. Szoka, F. C., and Papahadjopoulos, D. (1980), Comparative properties and methods of preparation of lipid vesicles (liposomes), *Ann. Rev. Biophys. Bioeng.*, 75, 4194–4199.

92. Kulkarni, S. B., Betageri, G. V., and Singh, M. (1995), Factors affecting microencapsulation of drugs in liposomes, *J. Microencapsul.*, 12, 229–246.

93. Amselem, S., Gabizon, A., and Barenholz, Y. (1990), Optimization and up scaling of doxorubicin-containing liposomes for clinical use, *J. Pharm. Sci.*, 79, 1045–1052.

94. Mezei, M., and Nugent, F. J. (1984), Method of encapsulating biologically active materials in multilmellar lipid vesicles, U.S. Patent 4, 485,054.

95. Ma, L., Ramachandran, C., and Weiner, N. D. (1991), Partitioning of a homologous series of alkyl *p*-amino benzoates in dipalmitoyl phosphatidyl choline liposomes: Effect of liposomes type, *Int. J. Pharm.*, 70, 209–218.

96. Betageri, G. V., and Parsons, D. L. (1992), Drug encapsulation and release from multilamellar and unilamellar liposomes, *Int. J. Pharm.*, 81, 235–241.

97. Puglisi, G., Fresta, M., La Rosa, C., Ventura, C. A., Panic, A. M., and Mazzonne, A. (1992), Liposomes as potential drug carrier for citicoline (CDP-choline) and the effect of formulation conditions on encapsulation efficiency, *Pharmazie*, 47, 211–215.
98. Betageri, G. V. (1993), Liposomal encapsulation and stability of dideoxynosine triphosphate, *Drug Dev. Ind. Pharm.*, 19, 531–539.
99. Fresta, M., Villari, A., Puglisi, G., and Cavallaro, G. (1993), 5-Fluorouracil: Various kinds of loaded liposomes: Encapsulation efficiency, storage stability and fusogenic properties, *Int. J. Pharm.*, 99, 145–156.
100. Elorza, B., Elorza, M. A., Frutos, G., and Chantres, J. R. (1993), Characterization of 5-flourouracil loaded liposomes prepared by reverse phases evaporation or freezing thawing extrusion methods: Study of release, *Biochim. Biophys. Acta*, 1153, 135–142.
101. Law, S. L., Chang, P., and Lin, C. H. (1991), Characteristics of mitoxantrone loading on liposomes, *Int. J. Pharm.*, 70, 1–7.
102. Alpar, O. H., Bamford, J. B., and Walters, V. (1981), In vitro incorporation and release of hydroxycobalamin by liposomes, *Int. J. Pharm.*, 7, 349–351.
103. Crommelin, D. J. A., and Van Bloois, L. (1983), Preparation and characterization of doxorubicin containing liposomes. Part 2. Loading capacity, long-term stability and doxorubicin bilayer interaction mechanism, *Int. J. Pharm.*, 17, 135–144.
104. Arien, A., Coigoux, C., Baquey, C., and Dupuy, B. (1993), Study of in vitro and in vivo stability of liposomes loading with calcitonin or indium in GIT, *Life Sci.*, 53, 1279–1290.
105. D'Silva, J. B., and Notari, R. E. (1982), Drug stability in liposomal suspensions: Hydrolysis of indomethacin, cyclocytidine and *p*-nitrophenyl acetate, *J. Pharm. Sci.*, 71, 1394–1398.
106. Kirby, C., Clarke, J., and Gregoriadis, G. (1980), Effect of the cholesterol content of small unilamellar liposomes on their stability in vivo and in vitro, *Biochem. J.*, 186, 591–595.
107. Kirby, C., and Gregoriadis, G. (1980), The effect of the cholesterol content of small unillamellar liposomes on the fate of their lipid components in vivo, *Life Sci.*, 27, 2223–2230.
108. Layton, D., and Trouet, A. (1980), A comparison of the therapeutic effects of free and liposomally encapsulated vincristine in leucemic mice, *Eur. J. Cancer*, 16, 945–951.
109. McCormack, B., and Gregoriadis, G. (1994), Drugs-in-cylcodextrins-liposomes: A novel concept in drug delivery, *Int. J. Pharm.*, 112, 249–258.
110. McCormack, B., and Gregoriadis, G. (1996), Comparative studies of the fate of free and liposome-entrapped hydroxypropyl-β-cyclodextrin-drug complexes after intravenous injection into rats: Implications in drug delivery, *Biochim. Biophys. Acta*, 1291, 237–244.
111. Fatouros, D. G., Hatzidimitirou, K., and Antimisiaris, S. G. (2001), Liposomes encapsulating prednisolone and prednisolone-cyclodextrin complexes: Comparison of membrane integrity and drug release, *Eur. J. Pharm. Sci.*, 13, 287–296.
112. Grit, M., and Crommelin, D. J. A. (1993), Analysis and hydrolysis kinetics of phospholipids in aqueous liposome dispersions, in Gregoriadis G., Ed., *Liposome Technology*, Vol. I, CRC Press, Boca Raton, FL, pp. 455–487.
113. Grit, M., and Crommelin, D. J. A. (1993), Chemical stability of liposomes: Implications for their stability, *Chem. Phys. Lipids*, 64, 3–18.
114. Zuidam, N. J., and Crommelin, D. J. A. (1995), Chemical hydrolysis of phospholipids, *J. Pharm. Sci.*, 84, 1113–1119.

115. Grit, M., de Smidt, J., Struijke, H. A., and Crommelin, D. J. A. (1989), Hydrolysis of natural soybean phosphatidylcholine in aqueous liposome dispersions, *Int. J. Pharm.*, 50, 1–6.
116. Hunt, C. A., and Tsang, S. (1981), α-Tocopherol retards autoxidation and prolongs the shelf life of liposomes, *Int. J. Pharm.*, 8, 101–110.
117. Fukuzawa, K., Chida, H., Akira, T., and Tsukatani, H. (1981), Autooxidative effect of α-tocopherol incorporation into lecithin liposomes on ascorbic acid $Fe^{++}$-induced lipid peroxidation, *Arch. Biochem. Biophys.*, 206, 173–180.
118. Talsma, H., van Steenbergen, M. J., and Crommelin, D. J. A. (1992), The cryopreservation of liposomes. 2. Effect of particle size on crystallization behavior and marker retention, *Cryobiology*, 29, 80–86.
119. Talsma, H., van Steenbergen, M. J., and Crommelin, D. J. A. (1991), The cryopreservation of liposomes. 3. Almost complete retention of a water-soluble marker in small liposomes in a cryoprotectant containing dispersion after a freezing/thawing cycle, *Int. J. Pharm.*, 77, 119–126.
120. Kristiansen, J. (1992), Leakage of a trapped fluorescent marker from liposomes: Effects of eutectic crystallization of NaCl and internal freezing, *Cryobiology*, 29, 575–584.
121. Crowe, L. M., and Crowe, J. H. (1988), Trehalose and dry dipalmitoloylphosphatidylcholine revisited, *Biochim. Biophys. Acta*, 946, 193–201.
122. Crowe, J. H., Leslie, S. B., and Crowe, L. M. (1994), Is vitrification sufficient to preserve liposomes during freeze-drying? *Cryobiology*, 31, 355–366.
123. Piel, G., Piette, M., Barillaro, V., Castagne, D., Evrard, B., and Delattre, L. (2006), Betamethasone-in-cyclodextrin-in-liposome: The effect of cyclodextrins on encapsulation efficiency and release kinetics, *Int. J. Pharm.*, 312, 75–82.
124. Woodle, M. C., and Papahadjopoulos, D. (1989), Liposome preparation and size characterization, *Meth. Enzym.*, 171, 193–217.
125. New, R. R. C., Ed. (1989), *Liposomes. A Practical Approach*, IRL Press, Oxford. UK.
126. Bangham, A. D., and Horne, R. W. (1964), Negative staining of phospholipids and their structured modification by surface active agents as observed in the electron microscope. *J. Mol. Biol.*, 8, 660–668.
127. Papahadjopoulos, D., and Watkins, J. C. (1967), Phospholipid model membrane 2: Permeability properties of hydrated liquid crystals, *Biochim. Biophys. Acta*, 135, 639–652.
128. Batzri, S., and Korn, E. D. (1973), Single bilayer vesicles prepared without sonication, *Biochm. Biophys. Acta*, 298, 1015–1019.
129. Deamer, D., and Bangham, A. D. (1976), Large volume liposomes by an ether vaporization method, *Biochm. Biophys. Acta*, 443, 629–634.
130. Barenholz, Y., Amselem, S., and Lichtenberg, D. (1979), A new method for preparation of phospholipid vesicles, *FEBS Lett.*, 99, 210–214.
131. Mayhew, E., Lazo, R., Vail, W. J., King, J., and Green, A. M. (1984), Characterization of liposomes prepared using microfluidizer, *Biochim. Biophys. Acta*, 775, 169–174.
132. Szoka, F. C., and Papahadjopoulos, D. (1978), Procedure for preparation of liposomes with large internal aqueous space and high capture by reverse phase evaporation, *Proc. Natl. Acad. Sci. USA*, 75, 4194–4198.
133. Gregoriadis, G., Da Silva, H., and Florence A. T. (1990), A procedure for the efficient entrapment of drugs in dehydration-rehydration liposomes, *Int. J. Pharm.*, 65, 235–242.
134. Zadi, B., and Gregoriadis, G. (2000), A novel method for high-yield entrapment of solutes into small liposomes, *J. Lipos. Res.*, 10, 73–80.
135. Reeves, J. P., and Dowben R. M. (1969), Formation and properties of thin phospholipids vesicles, *J. Cell. Phys.*, 73, 49–60.

136. Antimisiaris, S. G., Jayesekera, P., and Gregoriadis, G. (1993), Liposomes as vaccine carriers: Incorporation of soluble and particulate antigens in giant vesicles, *J. Immunol. Methods*, 166, 271–280.
137. Mimms, L. T., Zampighi, G., Nozaki, Y., Tanford, C., and Reynolds, J. A. (1981), Phospholipid vesicles formation and transmembrane protein incorporation using octyl glucoside, *Biochemistry*, 20, 833–840.
138. Hauser, H., Brunner, J., and Skrabal, P. (1976), Single bilayer vesicles prepared without sonication, *Biochim. Biophys. Acta*, 455, 322–331.
139. Papahadjopoulos, D., Vail, W. J., Jacobson, K., and Poste, G. (1975), Cochleate lipid cylinders formation by fusion of unilamellar lipid vesicles, *Biochim. Biophys. Acta*, 394, 483–491.
140. Talsma, H., Van Steenbergen, M. J., Borchert, J. C. H., Crommelin, D. J. A. (1994), A novel technique for the one-step preparation of liposomes and nonionic surfactant vesicles without the use of organic solvents. Liposome formation in a continuous gas stream: The "bubble" method, *J. Pharm. Sci.*, 83, 276–280.
141. Avestin, Inc., Ottawa, Canada, http://www.avestin.com/products.html, accessed Jan. 1, 2004.
142. Mayer, L. D., Bally, M. B., and Cullis, P. R. (1990), Characterization of liposomal systems containing doxorubicin entrapped in response to pH gradients, *Biochim. Biophys. Acta*, 1025, 143–151.
143. Nichols, J. W., and Deamer, D. W. (1990), Catecholamine uptake and concentration of liposomes maintaining pH gradients, *Biochim. Biophys. Acta*, 455, 269–278.
144. Harrigan, P. R., Wong, K. F., Redelmeier, T. E., Wheller, J. J., and Cullis, P. R. (1993), Accumulation of doxorubicin and other lipophilic amines into LUVs in response to transmembrane pH gradient, *Biochim. Biophys. Acta*, 1149, 237–244.
145. Boman, N. L., Mayer, L. D., and Cullis, P. R. (1993), Optimization of the retention and properties of vincristine in liposomal systems, *Biochim. Biophys. Acta*, 1152, 253–258.
146. Gabizon, A., Shiota, R., and Papahadjopoulos. D. (1989), Pharmacokinetics and tissue distribution of doxorubicin encapsulated in stable liposomes with long circulation times, *J. Nat. Can. Inst.*, 81, 1484–1488.
147. Gabizon, A., Barenholz, Y., and Bialer, M. (1993), Prolongation of the circulation time of doxorubicin encapsulated in liposomes containing a polyethylene glycol derivatized phospholipid: Pharmacokinetic studies in rodents and dogs, *Pharm. Res.*, 10, 703–708.
148. Cheung, B. C., Sun, T. H., Leenhouts, J. M., and Cullis, P. R. (1998), Loading of doxorubicin into liposomes by forming $Mn^{2+}$–drug complexes, *Biochim. Biophys. Acta*, 1414, 205–216.
149. Abraham, S. A., Edwards, K., Karlsson, G., MacIntosh, S., Mayer, L. D., McKenzie, C., and Bally, M. B. (2002), Formation of transition metal–doxorubicin complexes inside liposomes, *Biochim. Biophys. Acta*, 1565, 41–54.
150. Nobs, L., Buchegger, F., Gurny, R., and Allemann, E. (2004), Current methods for attaching targeting ligands to liposomes and nanoparticles, *J. Pharm. Sci.*, 93, 1980–1991.
151. Sapra, P., and Allen, T. M. (2003), Ligand-targeted liposomal anticancer drugs, *Prog. Lipid Res.*, 42, 439–462.
152. Klibanov, A. L., Maruyama, K., Beckerleg, A. M., Torchilin, V. P., and Huang, L. (1991), Activity of amphipathic poly(ethylene glycol) 5000 to prolong the circulation time of liposomes depends on the liposome size and is unfavorable for immunoliposome binding to target, *Biochim. Biophys. Acta*, 1062, 142–148.
153. Maruyama, K., Takizawa, T., Takahashi, N., Nagaike, K., and Iwatsuru, M. (1997), Targeting efficiency of PEG-immunoliposome-conjugated antibodies at PEG terminals, *Adv. Drug Deliv. Rev.*, 24, 235–242.

154. Maruyama, K., Takizawa, T., Yuda, T., Kennel, S. J., Huang, L., and Iwatsuru, M. (1995), Targetability of novel immunoliposomes modified with amphipathic poly(ethylene glycol)s conjugated at their distal terminals to monoclonal antibodies, *Biochim. Biophys. Acta*, 1234, 74–80.
155. Park, J. W., Benz, C. C., and Martin, F. J. (2004), Future directions of liposome- and immunoliposome-based cancer therapeutics, *Semin. Oncol.*, 31(6 Suppl 13), 196–205.
156. Ishida, T., Iden, D. L., and Allen, T. M. (1999), A combinatorial approach to producing sterically stabilized (Stealth) immunoliposomal drugs, *FEBS Lett.*, 460, 129–133.
157. Park, C. G., Thiex, N. W., Lee, K. M., Szot, G. L., Bluestone, J. A., and Lee, K. D. (2003), Targeting and blocking B7 costimulatory molecules on antigen-presenting cells using CTLA4Ig-conjugated liposomes: In vitro characterization and in vivo factors affecting biodistribuiton, *Pharm. Res.*, 20, 1239–1248.
158. Bartsch, M., Weeke-Klimp, A. H., Meijer, D. K. F., Scherphof, G. L., and Kamps, J. A. A. (2002), Massive and selective delivery of lipid-coated cationic lipoplexes of oligonucleotides targeted in vivo to hepatic endothelial cells, *Pharm. Res.*, 19, 676–680.
159. Chekhonin, V. P., Zhirkov, Y. A., Gurina, O. I., Ryabukhin, I. A., Lebedev, S. V., Kashparov, I. A., and Dmitriyeva, T. B. (2005), PEGylated Immunoliposomes directed against brain astrocytes, *Drug Deliv.*, 12, 1–6.
160. Brignole, C., Marimpietri, D., Gambini, C., Allen, T. M., Ponzoni, M., and Pastorino, F. (2003), Development of Fab′ fragments of anti-$GD_2$ immunoliposomes entrapping doxorubicin for experimental therapy of human neuroblastoma, *Cancer Lett.*, 197, 199–204.
161. Demirovic, A. M., Marty, C., Console, S., Zeisberger, S. M., Ruch, C., Jaussi, R., Schwendener, R. A., and Ballmer-Hofer, K. (2005), Targeting human cancer cells with VEGF receptor-2-directed liposomes, *Oncol. Rept.*, 13, 319–324.
162. Voinea, M., Manduteanu, I., Dragomir, E., Capraru, M., and Simionescu, M. (2005), Immunoliposomes directed toward VCAM-1 interact specifically with activated endothelial cells—a potential tool for specific drug delivery, *Pharm. Res.*, 22, 1906–1917.
163. Park, J. W., Hong, K., Kiprotin, D. B., Meyer, O., Papahadjopoulos, D., and Benz, C. C. (1997), Anti-HER2 immunoliposomes for targeted therapy of human tumors, *Cancer Lett.*, 118, 153–160.
164. Mercadal, M., Domingo, J. C., Petriz, J., Garcia, J., and de Madariaga, M. A. (1999), A novel strategy affords high-yield coupling of antibody to extremities of liposomal surface-grafted PEG chains, *Biochim. Biophys. Acta*, 1418, 232–238.
165. Longmuir, K. J., Robertson, R. T., Haynes, S. M., Baratta, J. L., and Waring, A. J. (2006), Effective targeting of liposomes to liver and hepatocytes in vivo by incorporation of a *Plasmodium* amino acid sequence, *Pharm. Res.*, 23, 759–769.
166. Hansen, C. B., Kao, G. Y., Moase, E. H., Zalipsky, S., and Allen, T. M. (1995), Attachment of antibodies to sterically stabilized liposomes: Evaluation, comparison and optimization of coupling procedures, *Biochim. Biophys. Acta*, 1239, 133–144.
167. Mamot, C., Ritschard, R., Kung, W., Park, J. W., Herrmann, R., and Rochlitz, C. F. (2006), EGFR-targeted immunoliposomes derived from the monoclonal antibody EMD72000 mediate specific and efficient drug delivery to a variety of colorectal cancer cells, *J. Drug Targe.*, 14, 215–223.
168. Kallinteri, P., Papadimitriou, E., and Antimisiaris, S. G., (2001), Uptake of liposomes which incorporate a glycopeptide fraction of asialofetuin by HepG2 cells, *J. Lipos. Res.*, 11, 175–193.
169. Fretz, M. M., Koning, G. A., Mastrobattista, E., Jiskoot, W., and Storm, G. (2004), OVCAR-3 cells internalize TAT-peptide modified liposomes by endocytosis, *Biochim. Biophys. Acta*, 1665, 48–56.

170. Moreira, J. N., and Gaspar, R. (2004), Antagonist G-mediated targeting and cytotoxicity of liposomal doxorubicin in NCI-H82 variant small cell lung cancer, *Braz. J. Med. Biol. Res.*, 37, 1185–1192.
171. Martin, F. J., Hubbell, W. L., and Papahadjopoulos, D. (1981), Immunospecific targeting of liposomes to cells: A novel and efficient method for covalent attachment of Fab′ fragments via disulfide bonds, *Biochemistry*, 20, 4229–4238.
172. Wartchow, C. A., Alters, S. E., Garzone, P. D., Li, L., Choi, S., DeChene, N. E., Doede, T., Huang, L., Pease, J. S., Shen, Z., Knox, S. J., and Cleland, J. L. (2004), Enhancement of the efficacy of an antagonist of an extracellular receptor by attachment to the surface of a biocompatible carrier, *Pharm. Res.*, 21, 1880–1885.
173. Anabousi, S., Bakowsky, U., Schneider, M., Huwer, H., Lehr, C. M., and Ehrhardt, C. (2006), In vitro assessment of transferrin-conjugated liposomes as drug delivery systems for inhalation therapy of lung cancer, *Eur. J. Pharm. Sci.*, *Jul 22*, 29, 367–374.
174. Lopez-Barcons, L. A., Polo, D., Llorens, A., Reig, F., and Fabra, A. (2005), Targeted adriamycin delivery to MXT-B2 metastatic mammary carcinoma cells by transferring liposomes: Effect of adriamycin ADR-to-lipid ratio, *Oncol. Rept.*, 14, 1337–1343.
175. Afzelius, P., Demant, E. J. F., Hansen, G. H., and Jensen, P. B. (1989), Covalent modification of serum transferrin with phospholipids and incorporation into liposomal membranes, *Biochim. Biophys. Acta*, 979, 231–238.
176. Stavridis, J. C., Deliconstantinos, G., Psallidopoulos, M. C., Armenakas, N. A., Hadjiminas, D. J., and Hadjiminas, J. (1986), Construction of transferrin-coated liposomes for on vivo transport of exogenous DNA to bone marrow erythroblasts in rabbits, *Exp. Cell Res.*, 164, 568–572.
177. Torchilin, V. P., Levchenko, T. S., Lukyanov, A. N., Khaw, B. A., Klibanov, A. L., Rammohan, R., Samokhin, G. P., and Whiteman, K. R. (2001), *p*-Nitrophenylcarbonyl-PEG-PE-liposomes: Fast and simple attachmentof specific ligands, including monoclonal antibodies, to distal ends of PEG chains via *p*-nitrophenylcarbonyl groups, *Biochim. Biophys. Acta*, 1511, 397–411.
178. Savva, M., Duda, E., and Huang, L. (1999), A genetically modified recombinant tumor necrosis factor-α conjugated to the distal terminals of liposomal surface grafted polyethyleneglycol chains, *Int. J. Pharm.*, 184, 45–51.
179. Ansell, S. M., Tardi, P. G., and Buchkowsky, S. S. (1996), 3-(2-Pyridyldithio)propionic acid hydrazide as a cross-linker in the formation of liposome-antibody conjugates, *Bioconjugate Chem.*, 7, 490–496.
180. Chua, M. M., Fan, S. T., and Karush, F. (1984), Attachment of immunoglobulin to liposomal membrane via protein carbohydrate, *Biochim. Biophys. Acta*, 800, 291–300.
181. Koning, G. A., Morselt, H. W., Velinova, M. J., Donga, J., Gorter, A., Allen, T. M., Zalipsky, S., Kamps, J. A., and Scherphof, G. L. (1999), Selective transfer of a lipophilic prodrug of 5-fluorodeoxyuridine from immunoliposomes to colon cancer, *Biochim. Biophys. Acta*, 1420, 153–167.
182. Harding, J. A., Engbers, C. M., Newman, M. S., Goldstein, N. I., and Zalipsky, S. (1997), Immunogenicity and pharmacokinetic attributes of poly(ethylene glycol)-grafted immunoliposomes, *Biochim. Biophys. Acta*, 1327, 181–192.
183. Sawant, R. M., Hurley, J. P., Salmaso, S., Kale, A., Tolcheva, E., Levchenko, T. S., and Torchilin, V. P. (2006), "SMART" drug delivery systems: Double-targeted pH-responsive. Pharmaceutical nanocarriers, *Bioconjugate Chem.*, 17, 943–949.
184. Xiao, Z., McQuarrie, S. A., Suresh, M. R., Mercer, J. R., Gupta, S., and Miller, G. G. (2002), A three-step strategy for targeting drug carriers to human ovarian carcinoma cells in vitro, *J. Biotechnol.*, 94, 171–184.

185. Schnyder, A., Krahenbuhl, S., Torok, M., Drewe, J., and Huwyler, J. (2004), Targeting of skeletal muscle in vitro using biotinylated immunoliposomes, *Biochem. J.*, 377, 61–67.

186. Allen, T. M., Mumbengegwi, D. R., and Charrois, J. R. (2005), Anti-CD19-targeted doxorubicin improves the therapeutic efficacy in murine B-cell lymphoma and ameliorates the toxicity of liposomes with varying drug release rates, *Clin. Cancer Res.*, 11, 3567–3573.

187. Gupta, B., Levchenko, T. S., Mongayt, D. A., and Torchilin, V. P. (2005), Monoclonal antibody 2C5-mediated binding of liposomes to brain tumor cells in vitro and in subcutaneous tumor model in vivo, *J. Drug Target.*, 13, 337–343.

188. Kawakami, S., Wong, J., Sato, A., Hattori, Y., Yamashita, F., and Hashida, M. (2000), Biodistribution characteristics of mannosylated, fucosylated, and galactosylated liposomes in mice, *Biochim. Biophys. Acta*, 1524, 258–265.

189. de Melo, A. L., Silva-Barcellos, N. M., Demicheli, C., and Frezard, F. (2003), Enhanced schistosomicidal efficacy of tartar emetic encapsulated in pegylated liposomes, *Int. J. Pharm.*, 255, 227–230.

190. Chen, J-H., Ling, R., Yao, Q., Li, Y., Chen, T. Wang, Z., and Li, K-Z. (2005), Effect of small-sized liposomal adriamycin administered by various routes on a metastatic breast cancer model, *Endocr.-Related Cancer*, 12, 93–100.

191. Zhang, J. Q., Zhang, Z. R., Yang, Q. Y., Qin, S. R., and Qiu, X. L. (2005), Lyophilized paclitaxel magnetoliposomes as a potential drug delivery system for breast carcinoma via parenteral administration: In vitro and in vivo, *Pharm. Res.*, 22, 573–583.

192. Wang, X. P., Yazawa, K., Templeton, N. S., Yang, J., Liu, S., Li, Z., Li, M., Yao, Q., Chen, C., and Brunicardi, F. C. (2005), Intravenous delivery of liposome-mediated nonviral DNA is less toxic than intraperitoneal delivery in mice, *World J. Surg.*, 29, 339–343.

193. Kamps, J. A. A. M., and Scherphof, G. L. (2004), Biodistribution and uptake of liposomes in vivo, *Methods Enzymol.*, 387, 257–266.

194. Moghimi, S. M., and Hunter, A. S. (2001), Recognition by macrophages and liver cells of opsonized phospholipid vesicles and phospholipid headgroups, *Pharm. Res.*, 18, 1–8.

195. Chiu, G. N., Bally, M. B., and Mayer, L. D. (2001), Selective protein interactions with phosphatidylserine containing liposomes alter the steric stabilization properties of poly(ethylene glycol), *Biochim. Biophys. Acta*, 1510, 56–69.

196. Harrington, K. J., Syrigos, K. N., and Vile, R. G. (2002), Liposomally targeted cytotoxic drugs for the treatment of cancer. *J. Pharm. Pharmacol.*, 54, 1573–1600.

197. Forssen, E. A., Coulter, D. M., and Proffitt, R. T. (1992), Selective in vivo localization of daunorubicin small unilamellar vesicles in solid tumours, *Cancer Res.*, 52, 3255–3261.

198. Forssen, E. A., Male-Brune, R., Adler-Moore, J. P., Lee, M. J., Schmidt, P. G., Krasieva, T. B., Shimizu, S., and Tromberg, B. J. (1996), Fluorescence imaging studies for the disposition of daunorubicin liposomes (DaunoXome) within tumour tissue, *Cancer Res.*, 56, 2066–2075.

199. Ogihara-Umeda, I., Sasaki, T., Kojima, S., and Nishigori, H. (1996), Optimal radiolabelled liposomes for tumour imaging, *J. Nucl. Med.*, 37, 326–332.

200. Sadzuka, Y., and Hirota, S. (1998), Does the amount of an antitumor agent entrapped in liposomes influence its tissue distribution and cell uptake? *Cancer Lett.*, 131, 163–170.

201. Turner, A. F., Presant, C. A., Proffitt, R. T., Williams, L. E., Winsor, D. W., and Werner, J. L. (1988), In-111-labelled liposomes: Dosimetry and tumour depiction, *Rdiology*, 166, 761–765.

202. Presant, C. A., Proffitt, R. T., Turner, A. F., Williams, L. E., Winsor, D., Werner, J. L., Kennedy, P., Wiseman, C., Gala, K., and McKenna, R. J. (1988), Successful imaging of human cancer with In-111-labelled phospholipid vesicles, *Cancer*, 62, 905–911.
203. Presant, C. A., Blayney, D., Proffitt, R. D., Turner, A. F., Williams, L. E., Nadel, H. I., Kennedy, P., Wiseman, C., Gala, K., and Crossley, R. J. (1990), Preliminary report: Imaging of Kaposi sarcoma and lymphoma in AIDS with indium-111-labelled liposomes, *Lancet*, 335, 1307–1309.
204. Kubo, A., Nakamura, K., Sammiya, T., Katayama, M., Hashimoto, T., Hashimoto, S., Kobayashi, H., and Teramoto, T. (1993), Indium-111-labelled liposomes: Dosimetry and tumour depiction in patients with cancer, *Eur. J. Nucl. Med.*, 20, 107–113.
205. Khalifa, A., Dodds, D., Rampling, R., Paterson, J., and Murray, T., (1997), Liposomal distribution in malignant glioma: Possibilities for therapy, *Nucl. Med. Commun.*, 18, 17–23.
206. Zucchetti, M., Boiardi, A., Silvani, A., Parisi, I., Piccolrovazzi, S., and D'Incalci, M. (1999), Distribution of daunorubicin and daunorubicinol in human glioma tumors after administration of liposomal daunorubicin, *Cancer Chemother. Pharmacol.*, 44, 173–176.
207. British National Formulary (BNF) Edition 51, 2006, Pharmaceutical Press, London.
208. http://www.gilead.com/pdf/daxpius.pdf#search='DaunoXome'. Gilead Sciences, Inc., 2007.
209. Shimizu, K., Takada, M., Asai, T., Kuromi, K., Baba, K., and Oku, N. (2002), Cancer chemotherapy by liposomal 6-[[2-(dimethylamino)ethyl]amino]-3-hydroxy-7H-indeno[2,1-*c*]quinolin-7-one dihydrochloride (TAS-103), a novel anticancer agent, *Biol. Pharm. Bull.*, 25, 1385–1387.
210. Wu, P. C., Tsai, Y. H., Liao, C. C., Chang, J. S., and Huang, Y. B. (2004), Tha characterization and biodistribution of cefoxitin-loaded liposomes, *Int. J. Pharm.*, 271, 31–39.
211. Alberts, D. S., Markman, M., Armstrong, D., Rothenberg, M. L., Muggia, F., and Howell, S. B. (2002), Intraperitoneal therapy for stage III ovarian cancer: A therapy whose time has come! *J. Clin. Oncol.*, 20, 3944.
212. Koga, S., Hamazoe, R., Maeta, M., Shimizu, N., Murakami, A., and Wakatsuki, T. (1988), Prophylactic therapy for peritoneal recurrence of gastric cancer by continuous hyperthermic peritoneal perfusion with mitomycin *C*, *Cancer*, 61, 232.
213. Speyer, J. L. (1985), The rationale behind intraperitoneal chemotherapy in gastrointestinal malignancies, *Semin. Oncol.*, 12, 23.
214. Sadzuka, Y., Hirota, S., and Sonobe, T. (2000), Intraperitoneal administration of doxorubicin encapsulating liposomes against peritoneal dissemination, *Toxicol. Lett.*, 116, 51–59.
215. Sadzuka, Y., Nakai, S., Miyagishima, A., Nozawa, Y., and Hirota, S. (1997), Effects of administered route on tissue distribution and antitumour activity of polyethyleneglycol-coated liposomes containing adriamycin, *Cancer Lett.*, 111, 77–86.
216. Marchettini, P., Stuart, A., Mohamed, F., Yoo, D., and Sugarbaker, P. H. (2002), Docetaxel: Pharmacokinetics and tissue levels after intraperitoneal and intravenous administration in a rat model, *Cancer Chemother. Pharmacol.*, 49, 499–503.
217. Sadzuka, Y., Hirama, R., and Sonobe, T. (2002), Effects of intraperitoneal administration of liposomes and methods of preparing liposomes for local therapy, *Toxicol. Lett.*, 126, 83–90.
218. Di Stefano, A., Carafa, M., Sozio, P., Pinnen, F., Braghiroli, D., Orlando, G., Cannazza, G., Ricciutelli, M., Marianecci, C., and Santucci, E. (2004), Evaluation of rat striatal L-dopa and DA concentration after intraperitoneal administration of L-dopa prodrugs in liposomal formulations, *J. Controlled Release*, 99, 293–300.
219. Oussoren, C., and Storm, G. (2001), Liposomes to target the lymphatics by subcutaneous administration, *Adv. Drug Deliv. Rev.*, 50, 143–156.

220. Gregoriadis, G., Bacon, A., Caparros-Wanderley, W., and McCormack, B. (2002), A role for liposomes in genetic vaccination, *Vaccine* 20, B1–B9.

221. Wu, M. S., Robbins, J. C., Bugianesi, R. L., Ponpipom, M. M., and Shen, T. Y. (1981), Modified in vivo behaviour of liposomes containing synthetic glycolipids, *Biochim. Biophys. Acta*, 674, 19–29.

222. Fujimoto, Y., Okuhata, Y., Tyngi, S., Namba, Y., and Oku, N. (2000), Magnetic resonance lymphography of profunded lymphnodes with liposomal gadolinium-diethylenetriamine penta-acetic acid, *Biol. Pharm. Bull.*, 23, 97–100.

223. Misselwitz, B., and Sachse, A. (1997), Interstitial MR lymphographyusing GD-carrying liposomes, *Acta Radiol. Suppl.*, 412, 51–55.

224. Moll, K. P., Stober, R., Hermann, W., Borchert, H. H., and Utsumi, H. (2004), In vivo ESR studies on subcutaneously injected multilamellar liposomes in living mice, *Pharm. Res.*, 21, 2017–2024.

225. Bacon, A., Caparrós-Wanderley, W., Zadi, B., and Gregoriadis, G. (2002), Induction of a cytotoxic T lymphocyte (CTL) response to plasmid DNA delivered by Lipodine™, *J. Liposome Res.*, 12, 173–183.

226. Perrie, Y., Barralet, J. E., McNeil, S., and Vangala, A. (2004), Surfactant vesicle-mediated delivery of DNA vaccines via the subcutaneous route, *Int. J. Pharm.*, 284, 31–41.

227. Klibanov, A. L., Maruyama, K., Torchilin, V. P., and Huang, L. (1990), Amphipathic polyethyleneglycols effectively prolong the circulation time of liposomes, *FEBS Lett.*, 268, 235–237.

228. Gabizon, A., and Papahadjopoulos, D. (1992), The role of surface charge and hydrophilic groups on liposome clearance in vivo, *Biochim. Biophys. Acta*, 1103, 94–100.

229. Ishida, O., Maruyama, K., Sasaki, K., and Iwatsuru, M. (1999), Size-dependent extravasation and interstitial localization of polyethyleneglycol liposomes in solid tumour-bearing mice, *Int. J. Pharm.*, 190, 49–56.

230. Levchenko, T. S., Rammohan, R., Lukyanov, A. N., Whiteman, K. R., and Torchilin, V. P. (2002), Liposome clearance in mice: The effect of a separate and combined presence of surface charge and polymer coating, *Int. J. Pharm.*, 240, 95–102.

231. Lee, C. M., Choi, Y., Huh, E. J., Lee, K. Y., Song, H. C., Sun, M. J., Jeong, H. J., Cho, C. S., and Bom, H. S. (2005), Polyethylene glycol (PEG) modified $^{99m}$Tc-HMPAO-liposome for improving blood circulation and biodistribution: The effect of the extent of PEG ylation, *Cancer Biother. Radiopharm.*, 20, 620–628.

232. Bradley, A. J., Devine, D. V., and Ansell, S. M. (1998), Inhibition of liposome-induced complement activation by incorporated poly(ethylene glycol)-lipids, *Arch. Biochem. Biophys.*, 357, 185–192.

233. Lu, W. L., Qi, X. R., Zhang, Q., Li, R. Y., Zhang, R. J., and Wei, S. L. (2004), A PEGylated liposomal platform: pharmacokinetics, pharmacodynamics and toxicity in mice using doxorubicin as a model drug, *J. Pharmacol. Sci.*, 95, 381–389.

234. Ortho Biotech Products, L. P., http://www.doxil.com/common/prescribing_information/DOXIL/PDF/DOXIL_PI_Booklet.pdf#search='DOXIL', 2003–2007.

235. Blume, G., Cevc, G., Crommelin, M. D., Bakker-Woudenberg, I. A., Kluft, C., and Storm, G. (1993), Specific targeting with poly(ethylene glycol)-modified liposomes: Coupling of homing devices to the ends of the polymeric chains combines effective target binding with long circulation times, *Biochim. Biophys. Acta*, 1149, 180–184.

236. Gabizon, A., Horowitz, A. T., Goren, D., Tzemach, D., Shmeeda, H., and Zalipsky, S. (2003), In vivo fate of folate-targeted polyethylene-glycol liposomes in tumor-bearing mice, *Clin. Cancer Res.*, 9, 6551–6559.

237. Maeda, N., Takeuchi, Y., Takada, M., Sadzuka, Y., Namba, Y., and Oku, N. (2004), Anti-neovascular therapy by use of tumour neovasculature-targeted long-circulating liposome, *J. Controlled Release*, 100, 41–52.

238. Syrigos, K. N., Vile, R. G., Peters, M., and Harrington, K. J. (2003), Biodistribution and pharmacokinetics of 111In-DTPA-labelled pegylated liposomes after intraperitoneal injection, *Acta Oncol.*, 42, 147–153.

239. Cabanes, A., Even-Chen, S., Zimberoff, J., Barenholz, Y., Kedar, E., and Gabizon, A. (1999), Enhancement of antitumor activity of polyethylene glycol-coated liposomal doxorubicin with soluble and liposomal interleukin 2, *Clini. Cancer Res.*, 5, 687–693.

240. Oussoren, C., Eling, W. M., Crommelin, D. J., Storm, G., and Zuidema, J. (1998), The influence of the route of administration and liposome composition on the potential of liposomes to protect tissue against local toxicity of two antitumour drugs, *Biochim. Biophys. Acta*, 1369, 159–172.

241. Patton, J. S., Fishburn, C. S., and Weers, J. G. (2004), The lungs as a portal of entry for systemic drug delivery, *Proc. Am. Thorac. Soc.*, 1, 338–344.

242. Niven, R. W., and Scheier, H. (1990), Nebulization of liposomes. I. Effects of lipid composition, *Pharm. Res.*, 7, 1127–1133.

243. Niven, R. W., Speer, M., and Scheier, H. (1991), Nebulization of liposomes. II. The effects of size and modeling of solute release profiles, *Pharm. Res.*, 8, 217–221.

244. Niven, R. W., Carvajal, M. T., and Scheier, H. (1992), Nebulization of liposomes. III. Effect of operating conditions, *Pharm. Res.*, 9, 515–520.

245. Taylor, K. M. G., Taylor, G., Kellaway, I. W., and Stevens, J. (1990), The stability of liposomes to nebulization, *Inter. J. Pharm.*, 58, 57–61.

246. Schwarz, L. A., Johnson, J. L., Black, M., Cheng, S. H., Hogan, M. E., and Waldrep, J. C. (1996), Delivery of DNA-cationic liposome complexes by small particle aerosol, *Hum. Gene Ther.*, 24, 35–36.

247. Eastman, S. J., Tousignant, J. D., Lukason, M. J., Murray, H., Siegel, C. S., Constantino, P., Harris, D. J., Cheng, S. H., and Scheule, R. K. (1997), Optimization of formulations and conditions for the aerosol delivery of functional cationic lipid: DNA complexes, *Hum. Gen. Ther.*, 8, 313–322.

248. Gonzalez-Rothi, R. J., and Schreier, H. (1993), Pulmonary delivery of liposomes, *J. Controlled Release*, 5, 149–161.

249. Waldrep, J. C., Gilbert, B. E., Black, M., and Knight, V. (1997), Operating characteristics of 18 different continuous-flow jet nebulizers with beclomethasone dipropionate liposome aerosol, *Chest*, 111, 316–323.

250. Hochhaus, G., Gonzalez-Rothi, R. J., Lukyanov, A., Derendorf, H., and Dallas Costa, T. (1995), Assessment of glycocorticoid lung targeting by ex vivo receptor binding studies in rats, *Pharm. Res.*, 12, 134–137.

251. Gonzalez-Rothi, R. J., Suarez, S., Hochhaus, G., Schreier, H., Lukyanov, A., Derendorf, H., and Dallas Costa, T. (1996), Pulmonary targeting of liposomal triamcinolone acetonide, *Pharm. Res.*, 13, 1699–1703.

252. Schreier, H., McNicol, K. J., Ausborn, M., and Soucy, D. W., Derendorf, H., Stecenko, A. A., and Gonzalez-Rothi, R. J. (1992), Pulmonary delivery of amikacin liposomes and acute liposomes toxicity in the sheep, *Int. J. Pharm.*, 87, 183–193.

253. Omri, A., Bealuc, C., Bouhajib, M., Montplaisir, S., Sharkawi, M., and Lagace, J. (1994), Pulmonary retention of free and liposome-encapsulated tobramycin after intratracheal administration in uninfected rats and rats infected with Pseudomonas aeruginosa, *Antimicr. Agents Chemother.*, 38, 1090–1095.

254. Shek, P. N., Suntres, Z. E., and Brooks, J. I. (1994), Liposomes in pulmonary applications: Physicochemical considerations, pulmonary distribution and antioxidant delivery, *J. Drug Targe.*, 2, 431–442.

255. Suntres, Z. E., and Shek, P. N. (1994), Incorporation of α-tocopherol in liposomes promotes the retention of liposome-encapsulated glutathione in the rat lung, *J. Pharm. Pharmacol.*, 46, 23–28.

256. Suntres, Z. E., and Shek, P. N. (1995), Intratracheally administered liposomal alpha-tocopherol protects the lung against long-term toxic effects of paraquat, *Biomed. Environ. Sci.*, 8, 289–300.

257. Walther, F. J., David-Cu, R., and Lopez, S. L. (1995), Antioxidant-surfactant liposomes mitigate hyperoxic lung injury in premature rabbits, *Am. J. Physiol.*, 269, L613–L617.

258. Vyas, S. P., Kannan, M. P., Jain, S., Mishra, V., and Singh, P. (2004), Design of liposomal aerosols for improved delivery of rifampicin to alveolar macrophages, *Int. J. Pharm.*, 269, 37–49.

259. Vyas, S. P., Quraishi, S., Gupta, S., and Jaganathan, K. S. (2005), Aerosolized liposome-based delivery of amphotericin B to alveolar macrophages, *Int. J. Pharm.*, 296, 12–25.

260. Doddoli, C., Ghez, O., Barlesi, F., D'Journo, B., Robitail, S., Thomas, P., and Clerc, T. (2005), In vitro and in vivo methotrexate disposition in alveolar macrophages: Comparison of pharmacokinetic parameters of two formulations, *Int. J. Pharm.*, 297, 180–189.

261. Fidler, I. J., Raz, A., Fogler, W. E., Kirsh, R., Bugleski, P., and Poste, G. (1980), Design of liposomes to improve delivery of macrophage-augmenting agents to alveolar macrophages, *Cancer Res.*, 40, 4460–4466.

262. Juliano, R. L., and McCullough, H. N. (1980), Controlled delivery of an antitumor drug: Localized action of liposome encapsulated cytosine arabinoside administered via the respiratory system, *Am. Soc. Pharm. Exp. Ther.*, 214, 381–387.

263. McCullough, H. N., and Juliano, R. L. (1979), Organ-selective action of an antitumor drug: Pharmacologic studies of liposome-encapsulated beta-cytosine arabinoside administered via the respiratory system of the rat, *J. Natl. Cancer Inst.*, 63, 727–731.

264. Knight, V., Kleinerman, E. S., Waldrep, J. C., Giovanella, B. C., Gilbert, B. E., and Koshkina, N. V. (2000), 9-Nitrocamptothecin liposome aerosol treatment of human cancer subcutaneous xenografts and pulmonary cancer metastases in mice, *Ann. N.Y. Acad. Sci.*, 922, 151–163.

265. Canonico, A. E., Conary, J. T, Meyrick, B. O., and Brigham, K. L. (1994), Aerosol and intravenous transfection of human a1-antitrypsin gene to lungs of rabbits, *Am. J. Respir. Cell Mol. Biol.*, 10, 24–29.

266. Stribling, R., Brunette, E., Liggitt, D., Gaensler, K., and Debs, R. (1992), Aerosol gene delivery in vivo, *Proc. Natl. Acad. Sci. USA*, 89, 11277–11281.

267. Rowland, R. N., and Woodley, J. F. (1980), The stability of liposomes in vitro to pH, bile salts and pancreatic lipase, *Biochim. Biophys. Acta*, 620, 400–409.

268. Nagata, M., Yotsuyanagi, T., and Ikeda, M. (1988), A two step model of disintegration kinetics of liposomes in bile salts, *Chem. Pharm. Bull.*, 36, 1508–1513.

269. Quadachi, S., Paternostre, M., andre, C., Genin, I., Thao, T. X., Puisieux, F., Devissaguet, J., and Barratt, G. (1998), Liposomal formulations for oral immunotherapy: In-vitro stability in synthetic intestinal media and in-vivo efficacy in the mouse, *J. Drug Targe.*, 5, 365–378.

270. Freund, O. (2001), Biodistribution and gastrointestinal drug delivery of new lipidic multilamellar vesicles, *Drug Deliv.*, 8, 239–244.

271. Mezei, M., and Gulasekharam, V. (1981), Liposomes: A selective drug delivery system for topical route of administration-gel dosage form, *J. Pharm. Pharmacol.*, 34, 473–474.
272. Mezei, M., and Gulasekharam, V. (1980), Liposomes: A selective drug delivery system for topical route of administration. I-lotion dosage form, *Life Sci.*, 26, 1473–14777.
273. Honeywell-Nguyen, P. L., and Bouwstra, J. A. (2005), Vesicles as a tool for transdermal and dermal delivery, *Drug Discovery Today: Technol.*, 2, 67–74.
274. Knepp, V. M., Hinz, R. S., Szoka, F. C., and Guy, R. H. (1988), Controlled drug release from a novel liposomal delivery system. I. Investigation of transdermal potential, *J. Controlled Release*, 5, 211–221.
275. Yu, H. Y., and Liao, H. M. (1996), Triamcinolone permeation from different liposome formulations through rat skin in vitro, *Int. J. Pharm.*, 127, 1–7.
276. Fresta, M., and Puglisi, G. (1997), Corticosteroid dermal delivery with skin lipid liposomes, *J. Controlled Release*, 44, 141–151.
277. Liu, H., Pan, W. S., Tang, R., and Luo, S. D. (2004), Topical delivery of acyclovir palmitate liposome formulations through rat skin in vitro, *Pharmazie*, 59, 203–206.
278. Egbaria, K., Ramanchandran, C., and Weiner, N. (1991), Topical application of liposomally entrapped cyclosporin evaluated by in vitro diffusion studies with human skin, *Skin Pharmacol.*, 4, 21–28.
279. Manosroi, A., Konganeramit, L., and Manosroi, J. (2004), Stability and transdermal absorption of topical amphotericin B liposome formulations, *Int. J. Pharm.*, 270, 279–286.
280. Aguillela, V., Kontturi, K., Murtomaki, L., and Ramirez, P. (1994), Estimation of the pore size and charge density in human cadaver skin, *J. Controlled Release*, 32, 249–257.
281. Cevc, G., Schatzlein, A., and Richardsen, H. (2002), Ultradeformable lipid vesicles can penetrate the skin and other semi-permeable barriers unfragmented. Evidence from double label CLSM experiments and direct size measurements, *Biochim. Biophys. Acta*, 564, 21–30.
282. Cevc, G. (1996), Transferosomes, liposomes and other lipid suspensions on the skin: Permeation enhancement, vesicle penetration and transdermal drug delivery, *Crit. Rev. Ther. Drug Carrier Syst.*, 13, 257–388.
283. de Plessis, J., Ramachandran, C., Weiner, N., and Muller, D. G. (1994), The influence of particle size of liposomes on the deposition of the drug into skin, *Int. J. Pharm.*, 103, 277–282.
284. Verma, D. D., Verma, S., Blume, G., and Fahr, A. (2003), Particle size of liposomes influences dermal delivery of substances into skin, *Int. J. Pharm.*, 258, 141–151.
285. Miyachi, Y., Imamura, S., and Niwas, Y. (1987), Decreased skin superoxide dismutase activity by a single exposure of ultraviolet radiation is reduced by liposomal superoxide dismutase pretreatment, *J. Invest. Dermatal.*, 89, 111–112.
286. Short, S. M., Rubas, W., Paasch, B. D., Mrsny, R. (1995), Transport of biologically active interferon-gamma across human skin in vitro, *Pharm. Res.*, 12, 1140–1145.
287. Cotsarelis, G. (2002), The hair follicle as a target for gene therapy, *Ann. Dermatal Venereol.*, 129, 841–844.
288. Raghavachari, N., and Fahl, W. E. (2002), Targeted gene delivery to skin cells in vivo: A comparative study of liposomes and polymers as delivery vehicles, *J. Pharm. Sci.*, 91, 615–622.
289. Balsari, A. L., Morelli, D., Menard, S., Veronesi, U., and Colnaghi, M. I. (1994), Protection against doxorubicin-induced alopecia in rats by liposome-entrapped monoclonal antibodies, *Res. Commun.*, 8, 226–230.

290. Weiner, N., Williams, N., Birch, G., Ramachandran, R., Shipman, C., Jr., and Flynn, G. (1989), Topical delivery of liposomally encapsulated interferon evaluated in a cutaneous herpes guinea pig model, *Antimicrob. Agents Chemoter.*, 33, 1217–1221.

291. Godin, B., and Touitou, E. (2003), Ethosomes: New prospects in transdermal delivery, *Crit. Rev. Ther. Drug Carrier Syst.*, 20, 63–102.

292. Kalia, Y. N., Naik, A., Garrison, J., Guy, R. H. (2004), Iontophoretic drug delivery, *Adv. Drug Deliv. Rev.*, 56, 619–658.

293. Vulta, N. B., Betageri, G. V., and Banga, A. K. (1996), Transdermal iontophoretic delivery of enkephalin formulated in liposomes, *J. Pharm. Sci.*, 85, 5–8.

294. Fang, J. Y., Sung, K. C., Lin, H. H., and Fang, C. L. (1999), Transdermal iontophoretic delivery of enoxacin from various liposome-encapsulated formulations, *J. Controlled Release*, 60, 1–10.

295. Essa, E. A., Bonner, M. C., and Barry, B. W. (2002), Iontophoretic estradiol skin delivery and tritium exchange in ultradeformable liposomes, *Int. J. Pharm.*, 240, 55–66.

296. Essa, E. A., Bonner, M. C., and Barry, B. W. (2004), Electrically assisted skin delivery of liposomal estradiol; phospholipid as damage retardant, *J. Controlled Release*, 95, 535–546.

297. Conjeevaram, R., Chaturvedala, A., Betageri, G. V., Sunkara, G., and Banga, A. K. (2003), Iontophoretic in vivo transdermal delivery of beta-blockers in hairless rats and reduced skin irritation by liposomal formulation, *Pharm. Res.*, 9, 1496–1501.

298. Han, I., Kim, M., and Kim, J. (2004), Enhanced transfollicular delivery of adriamycin with a liposome and iontophoresis, *Exp. Dermatol.*, 13, 86–92.

299. Wells, J. M., Li, L. H., Sen, A., Jahreis, G. P., and Hui, S. W. (2000), Electroporation-enhanced gene delivery in mammary tumors, *Gene Ther.*, 7, 541–547.

300. Essa, E. A., Bonner, M. C., and Barry, B. W. (2003), Electroporation and ultradeformable liposomes; human skin barrier repair by phospholipids, *J. Controlled Release*, 92, 163–172.

301. Sen, A., Zhao, Y. L., and Hui, S. W. (2002), Saturated anionic phospholipids enhance transdermal transport by electroporation, *Biophys. J.*, 83, 2064–2073.

302. Vyas, S. P., Singh, R., Asati, R. K. (1995), Liposomally encapsulated diclofenac for sonophoresis induced systemic delivery, *J. Microencapsul.*, 12, 149–154.

303. Ahmed, I., and Patton, T. F. (1987), Disposition of timolol and inulin in the rabbit eye following corneal versus noncorneal absorption, *Int. J. Pharm.*, 38, 9–21.

304. Mishima, S., Gasset, A. Klyce, S. D., and Baum, J. L. (1966), Determination of tear volume and tear flow, *Invest. Ophthalmol.*, 5, 264–276.

305. Ebrahim, S., Peyman, G., and Lee, P. J. (2005), Applications of liposomes in ophtalmology, *Sur. Opthalmol.*, 50, 167–182.

306. Kaur, I. P., Garg, A., Singla, A. K., and Aggarwal, D. (2004), Vesicular systems in ocular drug delivery: An overview, *Int. J. Pharm.*, 269, 1–14.

307. Zimmer, A. K., and Kreuter, J. (1991), Studies on the transport pathway of PBCA nanoparticles in ocular tissues, *J. Microencapsul.*, 8, 497–504.

308. Marchal-Heussler, Sirbat D., and Hoffman, M. (1993), Poly-E-caprolactone nanocapsules in carteolol ophthalmic delivery, *Pharm. Res.*, 10, 386–390.

309. Calvo, P., Thomas, C., and Alonso, M. J. (1994), Study of the mechanism of interaction of poly-ξ-caprolactone nanocapsules with the cornea by confocal laser scanning microscopy, *Int. J. Pharm.*, 103, 283–291.

310. Schaeffer, H. E., Brietfelter, J. M., and Krohn, D. L. (1982), Lectin-mediated attachment of liposomes to cornea: Influence on transcorneal drug flux, *Invest. Ophthalmol. Vis. Sci.*, 23, 530–533.

311. Dharma, S. K., Fishman, P. H., and Peyman, G. A. (1986), A preliminary study of corneal penetration of 125I-labelled iodoxuridine liposome, *Acta Ophthalmol. (Copenh.)*, 64, 298–301.

312. Smolin, G., Okumoto, M., Feiler, S., and Condon, D. (1981), Iodoxuridine-liposome therapy for herpes simplex keratitis, *Am. J. Ophthalmol.*, 91, 220–226.

313. Schaeffer, H. E., and Krohn, D. L. (1982), Liposomes in topical drug delivery, *Invest. Ophthalmol. Vis. Sci.*, 22, 220–227.

314. Stratford, R. E. J., Yang, D. C., Redell, M. A., and Lee, V. H. L. (1983), Effects of topically applied liposomes and disposition of epinephrine and inulin in albino rabbit eye, *Int. J. Pharm.*, 13, 263–272.

315. Stratford, R. E. J., Yang, D. C., Redell, M. A., and Lee, V. H. L. (1983), Ocular distribution of liposome encapsulated epinephrine and inulin in the albino rabbit, *Curr. Eye Res.*, 2, 377–386.

316. Lee, V. H. L., Takemoto, K. A., and Iimoto, D. S. (1984), Precorneal factors influencing the ocular distribution of topically applied liposomal inulin, *Curr. Eye Res.*, 3, 585–591.

317. Angelucci, D. (2001, Nov.), New solutions for dry eye, EyeWorld, available: http://eyeworld.org.pastissue.php.

318. Lin, H. H., Ko, S. M., Hsu, L. R., and Tsai, Y. H. (1996), The preparation of norfloxacin-loaded liposomes and their in-vitro evaluation in pig's eye, *J. Pharm. Pharmacol.*, 48, 801–805.

319. Frucht-Perry, J., Assil, K. K., Ziegler, E., et al. (1992), Fibrin-enmeshed tobramycin liposomes: Single application topical therapy of Pseudomonas keratitis, *Cornea*, 11, 393–397.

320. Assil, K. K., Frucht-Perry, J., Ziegler, E., et al. (1991), Tobramycin liposomes. Single subconjunctival therapy of pseudomonal keratitis, *Invest. Ophthalmol. Vis. Sci.*, 32, 3216–3220.

321. Barza, M., Doft, B., and Lynch, E. (1993), Ocular penetration of ceftriaxone, ceftazidime, and vancomycin after subconjunctival injection in humans, *Arch. Ophthalmol.*, 111, 492–494.

322. Pleyer, U., Grammar, J., Pleyer, J. H., et al. (1995), Amphotericin B—bioavailability in the cornea. Studies with local administration of liposome incorporated amphotericin B, *Ophthalmologe*, 92, 469–475.

323. Law, S. L., Huang, K. J., and Chiang, C. H. (2000), Acyclovir-containing liposomes for potential ocular delivery. Corneal penetration and absorption, *J. Controlled Release*, 63(1–2), 135–140.

324. Fresta, M., Panico, A. M., Bucolo, C., et al. (1999), Characterization and in-vivo ocular absorption of liposome-encapsulated acyclovir, *J. Pharm. Pharmacol.*, 51, 565–576.

325. Al-Muhammed, J., Ozer, A. Y., Ercan, M. T., and Hincal, A. A. (1996), In-vivo studies on dexamethasone sodium phosphate liposomes, *J. Microencapsul.*, 13, 293–306.

326. Pleyer, U., Elkins, B., Ruckert, D., et al. (1994), Ocular absorption of cyclosporine A from liposomes incorporated into collagen shields, *Curr. Eye Res.*, 13, 177–181.

327. Pleyer, U., Lutz, S., Jusko, W. J., et al. (1993), Ocular absorption of topically applied FK506 from liposomal and oil formulations in the rabbit eye, *Invest. Ophthalmol. Vis. Sci.*, 34, 2737–2742 [erratum appears *Invest. Ophthalmol. Vis. Sci.*, 34, 3481, 1993].

328. Torres, P. F., Slegers, T. P., Peek, R., et al. (1999), Changes in cytokine mRNA levels in experimental corneal allografts after local clodronate-liposome treatment, *Invest. Ophthalmol. Vis. Sci.*, 40, 3194–3201.
329. Slegers, T. P., van Rooijen, N., van Rij, G., and van der Gaag, R. (2000), Delayed graft rejection in pre-vascularised corneas after subconjunctival injection of clodronate liposomes, *Curr. Eye Res.*, 20, 322–324.
330. Van der Veen, G., Broersma, L., Dijkstra, C. D., et al. (1994), Prevention of corneal allograft rejection in rats treated with subconjunctival injections of liposomes containing dichloromethylene diphosphonate, *Invest. Ophthalmol. Vis. Sci.*, 35, 3505–3515.
331. Van der Veen, G., Broersma, L., Van Rooijen, N., et al. (1998), Cytotoxic T lymphocytes and antibodies after orthotropic penetrating keratoplasty in rats treated with dichloromethylene diphosphonate encapsulated liposomes, *Curr. Eye Res.*, 17, 1018–1026.
332. Lee, V. H., and Carson, L. W. (1986), Ocular disposition of inulin from single and multiple doses of positively charged multilamellar liposomes: Evidence for alterations in tear dynamics and ocular surface characteristics, *J. Ocul. Pharmacol.*, 2, 353–364.
333. Elorza, B., Elorza, M. A., Sainz, M. C., and Chantres, J. R. (1993), Comparison of particle size and encapsulation parameters of three liposomal preparations, *J. Microencapsul.*, 10, 237–248.
334. Felt, O., Furrer, P., Mayer, J. M., Plazonnet, B., Buri, P., and Gurny, R. (1999), Topical use of chitosan in ophthalmology: Tolerance assessment and evaluation of precorneal retention, *Int. J. Pharm.*, 180, 185–193.
335. Felt, O., Baeyens, V., Zignani, M., Buri, P., and Gurny, R., (1999), Mucosal drug delivery, in Mathiowitz, E., Ed., *The Encyclopedia of Controlled Drug Delivery*, Wiley, New York, pp. 605–626.
336. McCalden, T. A., and Levy, M. (1990), Retention of topical liposomal formulations on the cornea, *Experientia*, 46, 713–715.
337. Grass, G. M., and Robinson, J. R. (1988), Mechanisms of corneal drug penetration. I: In vivo and in vitro kinetics, *J. Pharm. Sci.*, 77, 3–14.
338. Burrow, J., Tsibouklis, J., and Smart, J. D., (2002), Drug delivery to the eye, *The Drug Delivery Company Report*, p. 4.
339. Taniguchi, K., Yamamoto, Y., Itakura, K., Miichi, H., and Hayashi, S. (1988), Assessment of ocular irritability of liposome preparations, *J. Pharmacobiodyn.*, 11, 607–611.
340. Barber, R. F., and Shek, P. N. (1990), Tear induced release of liposome entrapped agents, *Int. J. Pharm.*, 60, 219–227.
341. Kaufman, H. E., Steinemann, T. L., Lehman, E., Thompson, H. W., Varnell, E. D., Jacob-LaBarre, J. T., and Gebhardt, B. M. (1994), Collagen-based drug delivery and artificial tears, *J. Ocul. Pharmacol.*, 10, 17–27.
342. Yerushalmi, N., and Margalit, R. (1994), Bioadhesive, collagen-modified liposomes: Molecular and cellular level studies on the kinetics of drug release and on binding to cell monolayers, *Biochim. Biophys. Acta*, 1189, 13–20.
343. Bochot, A., Fattal, E., Gulek, A., and Aonarraze, G. (1998), Liposome dispersed within thermosensitive gel. A new dosage form for ocular delivery of oligonucleotide, *Pharm. Res.*, 15, 1364–1369.
344. Bochot, A., Couvreur, P., and Fattal, E. (2000), Intravitreal administration of antisense oligonucleotides: Potential of liposomal delivery, *Prog. Retin. Eye Res.*, 19, 131–147.
345. Bochot, A., Fattal, E., Boutet, V., Deverre, J. R., Jeamy, J. C., Chacun, H., and Couvreur, P. (2002), Intravitreal delivery of oligonucleotides by sterically stabilized liposomes, *Invest. Opthalmol. Vis. Sci.*, 43, 253–259.

346. Schmidt-Erfurth, U., Flotte, T. J., Gragoudas, E. S., et al. (1996), Benzoporphyrin-lipoprotein-mediated photodestruction of intraocular tumors, *Exp. Eye Res.*, 62, 1–10.
347. Kawakami, S., Yamamura, K., Mukai, T., Nishida, K., Nakamura, J., Sakaeda, T., Nakashima, M., and Sasaki, H. (2001), Sustained ocular delivery of tilisolol to rabbits after topical administration or intravitreal injection of lipophilic prodrug incorporated in liposomes, *J. Pharm. Pharmacol.*, 53, 157–161.
348. Harris, A. S., Svensson, E., Wagner, Z. G., Lethagen, S., and Nilsson, I. M. (1988), Effect of viscosity on particle-size, deposition, and clearance of nasal delivery systems containing desmopressin, *J. Pharm. Sci.*, 77, 405–408.
349. Vyas, S. P., Goswami, S. K., and Singh, R. (1995), Liposomes based nasal delivery system of nifedipine—development ad characterization, *Int. J. Pharm.*, 118, 23–30.
350. Shahiwala A, and Misra, A. (2004), Nasal delivery of levonorgestrel for contraception: An experimental study in rats, *Fertil. Steril.* 81(Suppl. 1), 893–898.
351. Kuper, H., Adami, H. O., and Trichopoulos, D. (2000), Infection as a major preventable cause of human cancer, *J. Intern. Med.*, 248, 171–183.
352. Alpar, H. O., Somavarapu, S., Atuah, K. N., and Bramwell, V. W. (2005), Biodegradable mucoadhesive particulates for nasal and pulmonary antigen and DNA delivery, *Adv. Drug Deliv. Rev.*, 57, 411–430.
353. De Haan, A., Renegar, K. B., Small, P. A., Wilschut, J., (1995), *Vaccine*, 13, 613–616.
354. Wang, D., Christopher, M. E., Nagata, L. P., Zabielski, M. A., Li, H., Wong, J. P., and Samuel, J. (2004), Intranasal immunization with liposome-encapsulated plasmid DNA encoding influenza virus hemagglutinin elicits mucosal, cellular and humoral immune responses, *J. Clin. Virol.*, 31(Suppl. 1), 99–106.
355. Tafaghodi, M., Jaafari, M.-R., and Abolghasem, S., and Tabassi, S. (2006), Nasal immunization studies using liposomes loaded with tetanus toxoid and CpG-ODN, *Eur. J. Pharm. Biopharm.*, 64, 138–145.
356. de Jonge, M. I., Hamstra, H. J., Jiskoot, W., Roholl, P., Williams, N. A., Dankert, J., van Alphen, L., and Van der Ley, P. (2004), Intranasal immunisation of mice with liposomes containing recombinant meningococcal OpaB and OpaJ proteins, *Vaccine*, 22, 4021–4028.
357. Sloat, B. R., and Zhengrong, C. (2006), Strong mucosal and systemic immunities induced by nasal immunization with anthrax protective antigen protein incorporated in liposome-protamine-DNA particles, *Pharm. Res.*, 23, 262–269.
358. Paveli , Z., Škalko-Basnet, N., and Schubert, R. (2001), Liposomal gels for vaginal drug delivery, *Int. J. Pharm.*, 219, 139–149.
359. Paveli , Z., Škalko-Basnet, N., and Jalšenjak, I. (2005), Characterisation and in vitro evaluation of bioadhesive liposome gels for local therapy of vaginitis, *Int. J. Pharm.*, 301, 140–148.
360. Paveli , Z., Škalko-Basnet, N., Filipovi -Gr i , J., Martinac, A., and Jalšenjak, I. (2005), Development and in vitro evaluation of a liposomal vaginal delivery system for acyclovir, *J. Controlled Release*, 106, 34–43.
361. Zhang, Y., Cristofaro, P., Silbermann, R., Pusch, O., Boden, D., Konkin, T., Hovanesian, V., Monfils, P. R., Resnick, M., Moss S. F., and Ramratnam, B. (2006), Engineering mucosal RNA interference in vivo, *Mol. Ther.*, 14, 336–342.
362. Hartikka, J., Bozoukova, V., Ferrari, M., Sukhu, L., Enas, J., Sawdey, M., et al. (2001), Vaxfectin enhances the humoral immune response to plasmid DNA-encoded antigens, *Vaccine*, 19, 1911–1923.
363. Lochera, C. P., Witta, S. A., Ashlocka, B. M., Polacinob, P., Hub, S. L., Shiboskic, S., Schmidtb, A. M., Agyb, M. B., anderson, D. M., Stapransd, S. I., Megedee, J., and Levy, J.

A. (2004), Human immunodeficiency virus type 2 DNA vaccine provides partial protection from acute baboon infection, *Vaccine*, 22, 2261–2272.

364. Davidson, R. N., Croft, S. L., Scott, A., Maini, M., Moody, A. H., and Bryceson, A. D. (1991), Liposomal amphotericin B in drug-resistant visceral leishmaniasis, *Lancet*, 337, 1061–1062.
365. Guaglianone, P., Chan, K., DelaFlor-Weiss, E., Hanisch, R., Jeffers, S., Sharma, D., et al. (1994), Phase I and pharmacologic study of liposomal daunorubicin (DaunoXome), *Invest. New Drugs*, 12, 103–110.
366. Gabizon, A., Peretz, T., Sulkes, A., Amselem, S., Ben-Yosef, R., Ben-Baruch, N., et al. (1989), Systemic administration of doxorubicin-containing liposomes in cancer patients: A phase I study, *Eur. J. Cancer Clin. Oncol.*, 25, 1795–1803.
367. Donald, P. R., Sirgel, F. A., Venter, A., Smit, E., Parkin, D. P., VandeWal, B. W., et al. (2001), The early bactericidal activity of a low-clearance liposomal amikacin in pulmonary tuberculosis, *J. Antimicrob. Chemother.*, 48, 877–880.
368. Kadry, A. A., Al-Suwayeh, S. A., Abd-Allah, A. R., and Bayomi, M. A. (2004), Treatment of experimental osteomyelitis by liposomal antibiotics, *J. Antimicrob. Chemother.*, 54, 1103–1108.
369. Nishiyama, N., and Kataoka, K. (2006), Current state, achievements, and future prospects of polymeric micelles as nanocarriers for drug and gene delivery, *Pharmacol. Therap.*, 112, 630–648.
370. Zamboni, W. C., Gervais, A. C., Egorin, M. J., Schellens, J. H., Zuhowski, E. G., Pluim, D., et al. (2004), Systemic and tumor disposition of platinum after administration of cisplatin or STEALTH liposomal-cisplatin formulations (SPI-077 and SPI-077 B103) in a preclinical tumor model of melanoma, *Cancer Chemother. Pharmacol.*, 53, 329–336.
371. Proulx, M. E., Desormeaux, A., Marquis, J. F., Olivier, M., and Bergeron, M. G. (2001), Treatment of visceral leishmaniasis with sterically stabilized liposomes containing camptothecin, *Antimicrob. Agents Chemother.*, 45, 2623–2627.
372. Giles, F. J., Tallman, M. S., Garcia-Manero, G., Cortes, J. E., Thomas, D. A., Wierda, W. G., et al. (2004), Phase I and pharmacokinetic study of a low-clearance, unilamellar liposomal formulation of lurtotecan, a topoisomerase 1 inhibitor, in patients with advanced leukemia, *Cancer*, 100, 1449–1458.
373. Verschraegen, C. F., Gilbert, B. E., Loyer, E., Huaringa, A., Walsh, G., Newman, R. A., et al. (2004), Clinical evaluation of the delivery and safety of aerosolized liposomal 9-nitro-20(s)-camptothecin in patients with advanced pulmonary malignancies, *Clin. Cancer Res.*, 10, 2319–2326.
374. Rudin, C. M., Marshall, J. L., Huang, C. H., Kindler, H. L., Zhang, C., Kumar, D., et al. (2004), Delivery of a liposomal c-raf-1 antisense oligonucleotide by weekly bolus dosing in patients with advanced solid tumors: A phase I study, *Clin. Cancer Res.*, 10, 7244–7251.
375. Chien, P. Y., Wang, J., Carbonaro, D., Lei, S., Miller, B., Sheikh, S., et al. (2004), Novel cationic cardiolipin analogue-based liposome for efficient DNA and small interfering RNA delivery in vitro and in vivo, *Cancer Gene Ther.*, 12, 321–328.
376. Asano, T., and Kleinerman, E. S. (1993), Liposome-encapsulated MTP-PE: A novel biologic agent for cancer therapy, *J. Immunother.*, 14, 286–292.
377. Dagar, S., Krishnadas, A., Rubinstein, I., Blend, M. J., and Onyuksel, H. (2003), VIP grafted sterically stabilized liposomes for targeted imaging of breast cancer: In vivo studies, *J. Controlled Release*, 91, 123–133.

378. Park, J. W., Hong, K., Kirpotin, D. B., Colbern, G., Shalaby, R., Baselga, J., et al. (2002), Anti-HER2 immunoliposomes: Enhanced efficacy attributable to targeted delivery, *Clin. Cancer Res.*, 8, 1172–1181.

379. Pakunlu, R. I., Wang, Y., Tsao, W., Pozharov, V., Cook, T. J., and Minko, T. (2004), Enhancement of the efficacy of chemotherapy for lung cancer by simultaneous suppression of multidrug resistance and antiapoptotic cellular defense: Novel multicomponent delivery system, *Cancer Res.*, 64, 6214–6224.

380. Pardridge, W. M. (1996), Vector-mediated drug delivery to the brain, *Adv. Drug Deliv.*, 36, 299–321.

381. Huwyler, J., Wu, D., and Pardridge, W. M. (1996), Brain drug delivery of small molecules using immunoliposomes, *Proc. Natl. Acad. Sci. USA*, 93, 14164–14169.

382. Huwyler, J., Yang, J., and Pardridge, W. M. (1997), Targeted delivery of daunomycin using immunoliposomes: Pharmacokinetics and tissue distribution in the rat, *J. Pharmacol. Exp. Ther.*, 282, 1541–1546.

383. Schnyder, A., Krahenbuhl, S., Drewe, J., and Huwyler, J. (2005), Targeting of daunomycin using biotinylated immunoliposomes: Pharmacokinetics, tissue distribution and in vitro pharmacological effects, *J. Drug Targeting.*, 13, 325–335.

384. Xie, Y., Ye, L., Zhang, X., Cui, W., Lou, J., Nagai, T., and Hou, X. (2005), Transport of nerve growth factor encapsulated into liposomes across the blood-brain barrier: In vitro and in vivo studies, *J. Controlled Release*, 105, 106–119.

385. Brignole, C., Marimpietri, D., Pagnan, G., Di Paolo, D., Zancolli, M., Pistoia, V., Ponzoni, M., and Pastorino, F. (2005), Neuroblastoma targeting by c-myb-selective antisense oligonucleotides entrapped in anti-GD2 immunoliposome: Immune cell-mediated antitumor activities, *Cancer Lett.*, 228, 181–186.

386. Siegal, T., Horowitz, A., and Gabizon, A. (1995), Doxorubicin encapsulated in sterically stabilized liposomes for the treatment of a brain tumor model: Biodistribution and therapeutic efficacy, *J. Neurosurg.*, 83, 1029–1037.

387. Sharma, U. S., Sharma, A., Chau, R. I., and Straubinger, R. M. (1997), Liposome-mediated therapy of intracranial brain tumors in a rat model, *Pharm. Res.*, 14, 992–998.

388. Aoki, H., Kakinuma, K., Morita, K., Kato, M., Uzuka, T., Igor, G., Takahashi, H., and Tanaka, R. (2004), Therapeutic efficacy of targeting chemotherapy using local hyperthermia and thermosensitive liposome: Evaluation of drug distribution in a rat glioma model, *Int. J. Hypertherm.*, 20, 595–605.

389. Koukourakis, M. I., Koukouraki, S., Fezoulidis, I., Kelekis, N., Kyrias, G., Archimandritis, S., and Karkavitsas, N. (2000), High intratumoural accumulation of stealth liposomal doxorubicin (Caelyx) in glioblastomas and in metastatic brain tumors, *Br. J. Cancer*, 83, 1281–1286.

390. Arnold, R. D., Mager, D. E., Slack, J. E., and Straubinger, R. M. (2005), Effect of repetitive administration of Doxorubicin-containing liposomes on plasma pharmacokinetics and drug biodistribution in a rat brain tumor model, *Clin. Cancer Res.*, 11, 8856–8865.

391. Marina, N. M., Cochrane, D., Harney, E., Zomorodi, K., Blaney, S., Winick, N., Bernstein, M., and Link, M. P. (2002), Dose escalation and pharmacokinetics of pegylated liposomal doxorubicin (Doxil) in children with solid tumors: A pediatric oncology group study, *Clin. Cancer Res.*, 8, 413–418.

392. Caraglia, M., Addeo, R., Costanzo, R., Montella, L., Faiola, V., Marra, M., Abruzzesse, A., Palmieri, G., Budillon, A., Grillone, F., Venuta, S., Tagliaferri, P., and Del Prete, S. (2006), Phase II study of temozolomide plus pegylated liposomal doxorubicin in the treatment of brain metastases from solid tumors, *Cancer Chemother. Pharmacol.*, 57, 34–39.

393. Fiorillo, A., Maggi, G., Greco, N., Migliorati, R., D'Amico, A., Del Basso De Caro, M., Sabbatino, M. S., and Buffardi, F. (2004), Second-line chemotherapy with the association of liposomal daunorubicin, carboplatin and etoposide in children with recurrent malignant brain tumors, *J. Neuro-Oncol.*, 66, 179–185.

394. Barth, R. F., Coderre, J. A., Graca, M., Vicente, H., and Blue, T. E. (2005), Boron neutron capture therapy of cancer: Current status and future prospects, *Clin. Cancer Res.*, 11, 3987–4002.

395. Hawthorne, M. F., Feakes, D. A., and Shelly, K. (1996), Recent results with liposomes as boron delivery vehicles from boron neutron capture therapy, in Mishima, Y., Ed., *Cancer Neutron Capture Therapy*, Plenum, New York, pp 27–36.

396. Feakes, D. A., Waller, R. C., Hathaway, D. K., and Morton, V. S. (1999), Synthesis and in vivo murine evaluation of Na4[1-(1′-B10H9)-6-SHB10H8] as a potential agent for boron neutron capture therapy, *Proc. Natl. Acad. Sci. USA*, 96, 6406–6410.

397. Carlsson, J., Kullberg, E. B., Capala, J., Sjoberg, S., Edwards, K., and Gedda, L. (2003), Ligand liposomes and boron neutron capture theory, *J. Neuro-Oncol.*, 62, 47–59.

398. Barth, R. F., Yang, W., Adams, D. M., Rotaru, J. H., Shukla, S., Sekido, M., Tjarks, W., Fenstermaker, R. A., Ciesielski, M., Nawrocky, M. M., and Coderre, J. A. (2002), Molecular targeting of the epidermal growth factor receptor for neutron capture therapy of gliomas, *Cancer Res.*, 62, 3159–3166.

399. Wikstrand, C. J., Cokgor, I., Sampson, J. H., and Bigner, D. D. (1999), Monoclonal antibody therapy of human gliomas: Current status and future approaches, *Cancer Metastasis Rev.*, 18, 451–464.

400. Hermanson, M., Funa, K., Koopmann, J., Maintz, D., Waha, A., Westermark, B., Heldin, C. H., Wiestler, O. D., Louis, D. N., von Deimling, A., and Nister, M. (1996), Association of loss of heterozygosity on chromosome 17p with high platelet derived growth factor alpha receptor expression in human malignant gliomas, *Cancer Res.*, 56, 164–171.

401. Akabani, G., Cokgor, I., Coleman, R. E., Gonzalez Trotter, D., Wong, T. Z., Friedman, H. S., Friedman, A. H., Garcia-Turner, A., Herndon, J. E., DeLong, D., McLendon, R. E., Zhao, X. G., Pegram, C. N., Provenzale, J. M., Bigner, D. D., and Zalutsky, M. R. (2000), Dosimetry and dose–response relationships in newly diagnosed patients with malignant gliomas treated with iodine-131-labeled anti-tenascin monoclonal antibody 81C6 therapy, *Int. J. Radiat. Oncol. Biol. Phys.*, 46, 947–995.

402. Yoshida, J., and Mizuno, M. (2003), Clinical gene therapy for brain tumors. Liposomal delivery of anticancer molecule to glioma, *J. Neuro-Oncol.*, 65, 261–267.

403. Felgner, P. L., Gadek, T. R., Holm, M., Roman, R., Chan, H. W., Wenz, M., Northrop, J. R., Ringold, G. M., and Danielson, M. (1987), Lipofection: A highly efficient, lipid-mediated DNA-transfection procedure, *Proc. Natl. Acad. Sci. USA*, 84, 7413–7414.

404. Zhang, Y., Jeong, Lee, H., Boado, R. J., and Pardridge, W. M. (2002), Receptro-mediated delivery of an antisense gene to human brain cancer cells, *J. Gene Med.*, 4, 183–194.

405. Zhang, Y., Zhu, C., and Pardridge, W. M. (2002), Antisense gene therapy of brain cancer with an artificial virus gene delivery system, *Mol. Ther.*, 6, 67–72.

406. Zhang, Y., Zhang, Y. F., Bryant, J., Charles, A., Boado, R. J., and Pardridge, W. M. (2004), Intravenous RNA interference gene therapy targeting the human epidermal growth factor receptor prolongs survival in intracranial brain cancer, *Clin. Cancer Res.*, 10, 3667–3677.

407. Zerrouqi, A., Rixe, O., Ghoumari, A. M., Yarovoi, S. V., Mouawad, R., Khayat, D., and Soubrane, C. (1996), Liposomal delivery of the herpes simplex virus thymidine kinase gene in glioma: Improvement of cell sensitization to ganciclovir, *Cancer Gene Ther.*, 3, 385–392.

408. Voges, J., Weber, F., Reszka, R., Sturm, V., Jacobs, A., Heiss, W. D., Wiestler, O., and Kapp, J. F. (2002), Clinical protocol; Liposomal gene therapy with the herpes simplex thymidine kinase gene/ganciclovir system for the treatment of glioblastoma multiforme, *Hum. Gene Ther.*, 13, 675–685.

409. Mizuno, M., and Yoshida, J. (1998), Improvement of transduction efficiency of recombinant adeno-associated virus vector by entrapment in multilamellar liposomes, *Jpn. J. Cancer Res.*, 89, 352–354.

410. Huynh, G. H., Deen, D. F., and Szoka, F. C., Jr. (2006), Barriers to carrier mediated drug and gene delivery to brain tumours, *J. Controlled Release*, 110, 236–259.

411. Bomgaars, L., Geyer, J. R., Franklin, J., Dahl, G., Park, J., Winick, N. J., Klenke, R., Berg, S. L., and Blaney, S. M. (2004), Phase I trial of intrathecal liposomal cytarabine in children with neoplastic meningitis, *J. Clin. Oncol.*, 22, 3916–3921.

412. Glantz, M., LaFollette, S., Jaeckle, K., Shapiro, W., Swinnen, L., Rozental, J., Phuphanich, S., Rogers, L., Gutheil, J., Batchelor, T., Lyter, D., Chamberlain, M., Maria, B., Schiffer, C., Bashir, R., Thomas, D., Cowens, W., and Howell, S. B. (1999), A randomized trial of a slow-release versus a standard formulation of cytarabine for the intrathecal treatment of lymphomatous meningitis, *J. Clin. Oncol.* 17, 3110–3116.

413. MakCay, A. J., Deen, D. F., and Szoka, F. C., Jr. (2005), Distribution in brain of liposomes after convection enhanced delivery; modulation by particle charge, particle diameter, and presence of steric coating, *Brain Res.*, 1035, 139–153.

414. Polfliet, M. M., Goede, P. H., van Kesteren-Hendrikx, E. M., van Rooijen, N., Dijkstra, C. D., and van De Berg, T. K. (2001), Amethod for the selective depletion of perivascular and meningeal macrophages in the central nervous system, *J. Neuroimmunol.*, 116, 188–195.

415. Mamot, C., Nguyen, J. B., Pourdehnad, M., Hadaczek, P., Saito, R., Bringas, J. R., Drummond, D. C., Park, J. W., and Bankiewicz, K. S. (2004), Extensive distribution of liposomes in rodent brains and brain tumors following convection-enhanced delivery, *J. Neuro-Oncol.*, 68, 1–9.

416. Saito, R., Bringas, J. R., McKnight, T. R., Wendland, M. F., Mamot, C., Drummond, D. C., Kiprotin, D. B., Park, J. W., Berger, M. S., and Bankiewicz, K. S. (2004), Distribution of liposomes into brain and rat brain tumor models by convection-enhanced delivery monitored with magnetic resonance imaging, *Cancer Res.*, 64, 2572–2579.

417. Saito, R., Krauze, M. T., Bringas, J. R., Noble, C., McKnight, T. R., Jackson, P., Wendland, M. F., Mamot, C., Drummond, D. C., Kiprotin, D. B., Hong, K., Berger, M. S., Park, J. W., and Bankiewicz, K. S. (2005), Gadolinium-loaded liposomes allow for real-time magnetic resonance imaging of convection-enhaced delivery in the primate brain, *Exp. Neurol.*, 196, 381–389.

418. Groothuis, D. R. (2000), The blood-brain and blood-tumor barriers: A review of strategies for increasing drug delivery, *Neuro-Oncol.*, 2, 45–59.

419. Imaginis Corp. (The Women's Health Resource), http://www.imaginis.com/breasthealth/treatment.asp, 1997–2007.

420. Hofheinz, R. D., Gnad-Vogt, S. U., Beyer, U., and Hochhaus, A. (2005), Liposomal encapsulated anti-cancer drugs, *Anti-Cancer Drugs*, 16, 691–707.

421. Gradishar, W. J. (2005), The future of breast cancer: The role of prognostic factors, *Breast Cancer Res. Treat.*, 89, S17–S26.

422. Batist, G., Ramakrishnan, G., and Rao, C. S. (2001), Reduced cardiotoxicity and preserved antitumor efficacy of liposome-encapsulated doxorubicin and cyclophosphamide compared with conventional doxorubicin and cyclophosphamide in a randomized, multicenter trial of metastatic breast cancer, *J. Clin. Oncol.*, 19, 1444–1454.

423. Paridaens, R., Van Aaelst, F., and Georgoulias, V. (2003), A randomized phase II study of alternating and sequential regimens of docetaxel and doxorubicin as first line chemotherapy for metastatic breast cancer, *Ann. Oncol.*, 14, 433–440.

424. Chan, S., Davidson, N., Juozaityte, E., Erdkamp, F., Pluzanska, A., Azarnia, N., and Lee, L. W. (2004), Phase III trial of liposomal doxorubicin and cyclophosphamide compared with epirubicin and cyclophosphamide as first line therapy for metastatic breast cancer, *Ann. Oncol.*, 15, 1527–1534.

425. Mrozek, E., Rhoades, C. A., Allen, J., Hade, E. M., and Shapiro, C. L. (2005), Phase I trial of liposomal encapsulated doxorubicin (Myocet™; D-99) and weekly docetaxel in advanced breast cancer patients, *Ann. Oncol.*, 16, 1087–1093.

426. Schmid, P., Krocker, J., Jehn, C., Michniewicz, K., Lehenbauer-Dehm, S., Eggemann, H., Heilmann, V., Kummel, S., Schulz, C. O., Dieing, A., Wischnewsky, M. B., Hauptmann, S., Elling, D., Possinger, K., and Flath, B. (2005), Primary chemotherapy with gemcitabine as prolonged infusion, non-pegylated liposomal doxorubicin and docetaxel in patients with early breast cancer: Final results of a phase II trial, *Ann. Oncol.*, 16, 1624–1631.

427. Keller, A. M., Mennel, R. G., Georgoulias, V. A., Nabholtz, J. M., Erazo, A., Lluch, A., Vogel, C. L., Kaufmann, M., Minckwitz, G., Henderson, G., Mellars, L., Alland, L., and Tendler, G. (2004), Randomized phase III trial of pegylated liposomal doxorubicin versus vinorelbine or mitomycin C plus vinblastine in women with taxane-refractory advanced breast cancer. *J. Clin. Oncol.*, 22, 3893–3901.

428. Vorobiof, D. A., Rapoport, B. L., Chasen, M. R., Slabber, C., McMichael, G., Eek, R., and Mohammed, C. (2004), First in line therapy with paclitaxel (Taxol®) and pegylated liposomal doxorubicin (Caelyx®) in patients with metstatic breast cancer: A multicentre phase II study, *Breast*, 13, 219–226.

429. Overmoyer, B., Silvermann, P. Holder, L. W., Tripathy, D., and Henderson, I. C. (2005), Pegylated liposomal doxorubicin and cyclophosphamide as first-line therapy for patients with metastatic or recurrent breast cancer, *Clin. Breast Cancer*, 6, 150–157.

430. Coleman, R. E., Biganzoli, L., Canney, P., Dirix, L., Mauriac, L., Chollet, P., Batter, V., Ngalula-Kabanga, E., Dittrich, C., and Piccart, M. (2006), A randomized phase II study of two different schedules of pegylated liposomal doxorubicin in metastatic breast cancer (EORTC-10993), *Eur. J. Cancer*, 42, 882–887.

431. Stover, T. C., Sharma, A., Robertson, G. P., and Kester, M. (2005), Systemic delivery of liposomal short-chain ceramide limits solid tumor growth in murine models of breast adenocarcinoma, *Clin. Cancer Ther.*, 11, 3465–3474.

432. Swenson, S., Costa, F., Minea, R., Sherwin, R. P., Ernst, W., Fujii, G., Yang, D., and Markland, F. S., Jr. (2004), Intravenous liposomal delivery of the snake venom disintegrin contortrostatin limits breast cancer progression, *Mol. Cancer Ther.*, 3, 499–511.

433. Nagami, H., Matsumoto, Y., and Ueoka, R. (2006), Induction of apoptosis by hybrid liposomes for human breast tumor calls along with activation caspases, *Biol. Pharm. Bull.*, 29, 380–381.

434. Frenkel, V., Etherington, A., Greene, M., Quijano, J., Xie, J., Hunter, F., Dromi, S., and Li, K. C. P. (2006), Delivery of liposomal doxorubicin (Doxil) in a breast cancer tumor model: Investigation of potential enhancement by pulsed-high intensity focused ultrasound exposure, *Acad. Radiol.*, 13, 467–479.

435. Mukherjee, A., Prassad, T. K., Rao, N. M., and Banerjee, R. (2005), Haloperidol-associated stealth liposomes, *J. Biol. Chem.*, 280, 16(22), 1561–1562.

436. Seki, M., Iwakawa, J., Cheng, H., and Cheng, P. W. (2002), p53 and PTEN/MMAC1/TEP1 gene therapy of human prostate PC-3 carcinoma xenograft, using transferrin-facilitated lipofection gene delivery strategy, *Hum. Gene Ther.*, 13, 761–773.

437. Basma, H., El-Refaey, H., Sgagias, M. K., Cowan, K. H., Luo, X., and Cheng, P. W. (2005), Bcl-2 antisense and cisplatin combination treatment of MCF-7 breast cancer cells with or without functional p53, *J. Biom. Sci.*, 12, 999–1011.
438. Meidan, V. M., Glezer, J., Salomon, S., Sidi, Y., Barenholz, Y., Cohen, J. S., and Lilling G. (2006), Specific lipoplex-mediated antisense against bcl-2 in breast cancer cells: A comparison between different formulations, *J. Liposome Res.*, 16, 27–43.
439. Gradishar, W. J. (2005), The future of breast cancer: The role of prognostic factors, *Breast Cancer Res. Treat.*, 89, S17–S26.
440. NetDoctor.co.uk, http://www.netdoctor.co.uk/diseases/facts/lungcancer.htm, 2007.
441. http://www.merck.de/servlet/PB/menu/1508710/index.html.
442. Leighl, N., Burkes, R. L., Dancey, J. E., Lopez, P. G., Higgins, B. P., Walde, P. L. D., Rudinskas, L. C., Rahim, Y. H., Rodgers, A., Pond, G. R., and Shepherd, F. A. (2003), A phase I study of pegylated liposomal doxorubicin (Caelyx™) in combination with cyclophosphamide and vincristine as second-line treatment of patients with small-cell lung cancer, *Clin. Lung Cancer*, 5, 107–112.
443. Peer, D., and Margalit, R. (2004), Tumor-targeted hyaluronan nanoliposomes increase the antitumor activity of liposomal doxorubicin in syngeneic and human xenograft mouse tumor models, *Neoplasia*, 6, 343–353.
444. Ito, I., Ji, L., Tanaka, F., Saito, Y., Gopalan, B., Branch, C. D., Xu, K., Atkinson, E. N., Bekele, B. N., Stephens, L. C., Minna, J. D., Roth, J. A., and Ramesh, R. (2004), Liposomal vector mediated delivery of the 3p FUS1 gene demonstrate potent antitumor activity against human lung cancer in vivo, *Cancer Gene Ther.*, 11, 733–739.
445. Li, W., Ishida, T., Okada, N., and Kiwada, H. (2005), Increased gene expression by cationic liposomes (TFL-3) in lung metastases following intravenous injection, *Biol. Pharm. Bull.*, 28, 701–706.
446. Kawakami, A., Suzuki, S., Yamashita, F., and Hashida, M. (2006), Induction of apoptosis in A549 human lung cancer cells by all-trans retinoic acid incorporated in DOTAP/cholesterol liposomes, *J. Controlled Release*, 110, 514–521.
447. Remaut, K., Lucas, B., Braeckmans, K., Sanders, N. N., Demeester, J., and De Smedt, S. C. (2006), Delivery of phosphodiester oligonucleotides: Can DOTAP/DOPE liposomes do the trick? *Biochemistry*, 45, 1755–1764.
448. American Cancer Society, Inc., http://www.cancer.org, 2007.
449. Johnston, S. (2004), Ovarian cancer: Review of the national institute for clinical excellence (NICE) guidance recommendations, *Cancer Inv.*, 22, 730–742.
450. Rose, P. (2005), Pegylated liposomal doxorubicin: Optimizing the dosing schedule in ovarian cancer, *Oncologist*, 10, 205–214.
451. Tambaro, R., Greggi, S., Iaffaioli, R. V., Rossi, A., Pisano, C., Manzione, L., Ferrari, E., Di Maio, M., Iodice, F., Casella, G., Laurelli, G., and Pignata, S. (2003), An escalating dose finding study of liposomal doxorubicin and vinorelbine for the treatment of refractory or resistant epithelial ovarian cancer, *Ann. Oncol.*, 14, 1406–1411.
452. Katsaros, D., Oletti, M. V., Rigault de la Longrais, I. A., Ferrero, A., Celano, A., Fracchioli, S., Donadio, M., Passera, R., Cattel, L., and Bumma, C. (2005), Clinical and pharmacokinetic phase II study of pegylated liposomal doxorubicin and vinorelbine in heavily pretreated recurrent ovarian carcinoma, *Ann Oncol.*, 16, 300–306.
453. Gordon, A. N., Fleagle, J. T., Guthrie, D., et al. (2001), Recurrent epithelial ovarian carcinoma: A randomized phase III study of pegylated liposomal doxorubicin versus topotecan, *J. Clin. Oncol.*, 19, 3312–3322.
454. Verhaar-Langereis, M., Karakus, A., van Eijkeren, M., Voest, E., and Witteveen, E. (2006), Phase II study of the combination of pegylated liposomal doxorubicin and topotecan in platinum-resistant ovarian cancer, *Int. J. Gynecol. Cancer*, 16, 65–70.

455. Seiden, M. V., Muggia, F., Astrow, A., Matulonis, U., Campos, S., Roche, M., Sivret, J., Rusk, J., and Barrett, E. (2004), A phase II study of liposomal lurtotecan (OSI-211) in patients with topotecan resistant ovarian cancer, *Gynecol. Oncol.*, 93, 229–232.

456. Calvert, A. H., Grimshaw, R., Poole, C., Dark, G., Swnerton, K., Gore, M., et al. (2002), Randomized phase II trial of two intravenous schedules of the liposomal topoisomerase I inhibitor, NX211, in women with relapsed epithelial ovarian cancer (OVCA): An NCIC CTG study, *Proc. Am. Soc. Clin. Oncol.*, 21, 208a.

457. National Cancer Institute, U.S. National Institutes of Health, http://www.cancer.gov/search/ResultsClinicalTrials, 2007.

458. Gallo, D., Fruscella, E., Ferlini, C., Apollonio, P., Mancuso, S., and Scambia, G. (2006), Preclinical in vivo activity of a combination gemcitabine/liposomal doxorubicin against cisplatin-resistant human ovarian cancer (A2780/CDDP), *Int. J. Gynecol. Cancer*, 16, 222–230.

459. Anderson, K., Lawson, K. A., Simmons-Menchaca, M., Sun, L., SAnders, B. G., and Kline, K. (2004), α-TEA plus cisplatin reduces human cisplatin-resistant ovarian cancer cell tumor burden and metastasis, *Exp. Biol. Med.*, 229, 1169–1176.

460. Turk, M. J., Waters, D. J., and Low, P. S. (2004), Folate-conjugated liposomes preferentially target macrophages associated with ovarian carcinoma, *Cancer Lett.*, 213, 165–172.

461. Kim, C. K., Choi, E. J., Choi, S. H., Park, J. S., Haider, K. H., and Ahn, W. S. (2003), Enhanced p53 gene transfer to human ovarian cancer cells using the cationic nonviral vector, DDC, *Gynecol. Oncol.*, 90, 265–272.

462. Landen, C. N., Jr., Chavez-Reyes, A., Bucana, C., SchmAndt, R., Deavers, M. T., Lopez-Berestein, G., and Sood, A. K. (2005), Therapeutic EphA2 gene targeting in vivo using neutral liposomal small interfering RNA delivery, *Cancer Res.*, 65, 6910–6918.

# BIODEGRADABLE NANOPARTICLES

Sudhir S. Chakravarthi and Dennis H. Robinson
*University of Nebraska Medical Center, College of Pharmacy, Omaha, Nebraska*

**Contents**

1 Introduction
  1.1 Classification of Nanoparticles
2 Natural Biodegradable Polymeric Nanoparticles
  2.1 Physical Properties of Natural Polymers and Methods Used to Prepare Nanoparticles
  2.2 Drug Delivery Applications and Biological Fate of Natural Polymeric Nanoparticles
3 Synthetic Biodegradable Polymeric Nanoparticles
  3.1 Synthetic Polymers: Physical Properties and Methods of Preparation of Nanoparticles
  3.2 Drug Delivery Applications and Biological Fate of Synthetic Biodegradable Polymers
4 Thermosensitive and pH-Sensitive Nanoparticles
  4.1 Physical Properties and Methods of Preparation
  4.2 Drug Delivery Applications and Biological Fate of Thermosensitive and pH–Sensitive Nanoparticles
5 Applications of Biodegradable Nanoparticles Other Than Drug Delivery
6 Physicochemical Characterization of Polymeric Nanoparticles
  6.1 Molecular Weight
  6.2 Hydrophobicity
  6.3 Glass Transition Temperature
  6.4 Particle Size and Particle Size Distribution
  6.5 Surface Charge and Zeta Potential
  6.6 Surface Hydrophilicity
  6.7 Drug Loading and Encapsulation Efficiency
  6.8 Drug Release
  6.9 Physical Stability of Polymeric Nanoparticles
7 Targeting Nanoparticles by Surface Conjugation with Ligands

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

8 Cellular Trafficking of Biodegradable Nanoparticles
9 Conclusions
References

## 1 INTRODUCTION

Interest in nanotechnology has increased exponentially in many scientific areas, including drug delivery, nanoimaging, and other medical-related applications. Nanoparticles can be fabricated in many different shapes and sizes using a wide range of organic and inorganic materials. However, by definition, these particles must be within the size range of 1–1000 nm. Because the use of nanoparticles in drug delivery and nanomedicine invariably requires parenteral administration, there has been, and continues to be, a major need for the use of polymeric carriers that are both biocompatible and biodegradable. This review will focus on the application of nanotechnology to deliver therapeutic or diagnostic agents using biodegradable polymeric nanoparticles for systemic, localized, or targeted delivery.

### 1.1 Classification of Nanoparticles

Depending on the method of preparation and resulting structure, nanoparticles are broadly classified as either matrix-type or encapsulated particles. Hence, a drug is either homogenously dispersed in the polymeric matrix or encapsulated within the core of the particle. In drug delivery applications, biodegradable polymers used to prepare nanoparticles are either natural or synthetic in origin (Table 1). Natural polymers, or biopolymers, include alginates, chitosan, cellulose, gelatin, gliadin, and pullulan. A recent review describes the applications of these polymers in gene delivery and tissue engineering [1]. Natural polymers may vary widely in their composition and therefore physicochemical properties. Such variability in properties may result in poor reproducibility in delivery characteristics, such as drug loading and release kinetics. Further, their purification from natural sources may be difficult. In contrast, synthetic polymers can be prepared with relatively precise properties such as molecular weight, solubility, and permeability characteristics. Examples of synthetic polymers used to make biodegradable nanoparticles include polylactide, poly(lactide-*co*-glycolide) (PLGA), polyanhydride, polycyanoacrylate, poly(ε-caprolactone), and polyphosphoester. As liposomes are more commonly prepared in the micrometer-size range, they were considered out of the scope of this review. Because there are numerous reviews of these polymers in the literature, their physicochemical properties, methods of synthesis, applications, and biological fate of each of these polymers are only briefly described in this chapter.

## 2 NATURAL BIODEGRADABLE POLYMERIC NANOPARTICLES

Natural polymers extracted and purified from plant and animal sources often vary significantly in their purity. For example, alginate is available in over 200 different

**TABLE 1 Classification of Natural and Synthetic Polymers and Their Methods of Purification or Synthesis**

| Polymer | Source | Method of Purification or Synthesis | Solubility |
|---|---|---|---|
| Sodium alginate | Natural (seaweed) | Alkali-based extraction | Water soluble (pH > 3), insoluble in organic solvents |
| Chitosan | Natural (crab shells) | Deacetylation of chitin | Soluble in aqueous solutions (low pH), insoluble in organic solvents |
| Gelatin | Natural (collagen) | Hydrolysis | Soluble in hot water (>34 °C), acetic acid, forms insoluble gel with water at room temperature, insoluble in organic solvents |
| Polysaccharides | Natural | Enzymatic reactions | Pullulans (soluble in water), dextrans (soluble in water) |
| Albumin | Natural (plants, animals) | Separation techniques (chromatography) | Soluble in water |
| Gliadin | Natural (wheat) | Alcohol extraction | Insoluble in water, soluble in ethanol |
| Poly(lactide) and Poly(lactide-*co*-glycolide) | Synthetic | Ring-opening polymerization | Insoluble in water, soluble in organic solvents |
| Poly(ε-caprolactone) | Synthetic | Anionic, cationic, free-radical, ring-opening polymerization | Soluble in select organic solvents such as chloroform, dichloromethane |
| Polyanhydrides | Synthetic | Melt condensation, ring-opening polymerization | Most polyanhydrides soluble in organic solvents, insoluble in water |
| Poly-alkylcyanoacrylates | Synthetic | Emulsion and interfacial polymerization | Soluble in organic solvents |
| Polyphosphoesters | Synthetic | Polyaddition, ring-opening polymerization | Available as water-soluble and water-insoluble types |

grades and is extracted from various sources that differ in molecular weight and the percentage and arrangement of guluronic and mannuronic acid blocks. Further, chitosan, poly-β(1-4-D-glucosamine), is available in grades varying in molecular weight, degree of deacetylation (from parent compound, chitin), and viscosity [2].

Some natural polymers may be chemically modified to tailor solubility properties. An example is the reaction of the free amino groups of chitosan to form the more water soluble derivative, methoxy-polyethylene glycol (PEG) chitosan [3]. Collagen is marketed as six different types, I–VI, depending on its source and physiological applicability. Similarly, the properties of gelatin are dependent on the method of preparation using acid- or base-catalyzed hydrolysis from collagen.

## 2.1 Physical Properties of Natural Polymers and Methods Used to Prepare Nanoparticles

***Sodium Alginate*** Alginates are primarily derived from the algae *Macrocystis pyrifera* and *Laminaria hyperborea*. These are linear, unbranched polymers containing β-(1–4)-linked mannuronic acid and α-(1–4)-linked guluronic acid residues that are either arranged in blocks, commonly called G blocks and M blocks, or alternate with each other. Alginates are hydrophilic, anionic polymers that vary in molecular weight, depending primarily on the G and M blocks. They are characterized by the ratio of guluronic and mannuronic acids, which can be quantified by ultraviolet (UV) spectrophotometry, gas chromatography, and high-performance liquid chromatography (HPLC) [4]. For example, the polymer obtained from *M. pyrifera* has an M/G ratio of 1.6. Alginate obtained from seaweed must be purified by one of several applicable alkali and acid treatment protocols [5, 6].

Alginate nanoparticles can be prepared using ionotropic gelation, emulsification/internal gelation, and emulsification/solidification methods. Ionotropic gelation results when the anionic alginate reacts with cationic ions or molecules such as calcium or poly-*l*-lysine. Gelation occurs when cations chelate the guluronic and mannuronic acid groups to produce an "egg-box" structure that encapsulates the drug. The size of the alginate particles is determined by the molar concentration of calcium or poly-*l*-lysine and the method of addition of these counterions to alginate [7]. In the emulsification/internal gelation method, the sodium alginate and an insoluble calcium salt are dispersed in a vegetable oil and the calcium ions are liberated to form an alginate gel when the pH of the dispersion is lowered [8, 9]. An advantage of the use of alginate polymers to deliver drugs is that nanoparticles are prepared in aqueous media and may be more suitable to formulate compounds that are unstable in organic solvents such as proteins and peptides. However, since the chelation to form the gel is reversible, a disadvantage of unmodified, alginate-based delivery systems is rapid drug release due to the collapse of the egg-box structure when exposed to monovalent ions in physiological media.

***Chitosan*** Chitosan is a nontoxic, biodegradable polymer obtained by hydrolysis of chitin, a natural polysaccharide that is a chief component of the crustacean exoskeleton. Unmodified chitosan is soluble in acidic media and has significant mucoadhesive properties.

Chitosan nanoparticles may be prepared using various methods, including emulsion cross-linking, coacervation–precipitation, spray drying, emulsion droplet coalescence, ionic gelation, reverse-micellar method, and sieving. A relatively recent review describes the methods and applications of chitosan nanoparticles in drug delivery [10]. Chitosan nanoparticles have also been prepared using water-soluble cross-linking agents such as carbodiimide with the size being controlled by changing

pH [11]. More monodisperse nanoparticles may be prepared using fractionated and deacetylated chitosan [12]. In general, the size of nanoparticles will depend on the molecular weight of chitosan, its concentration, and its surface charge [13]. The physicochemical properties of chitosan are determined by the solution pH and ionic strength [14].

***Gelatin*** Gelatin is obtained by either alkaline or acidic hydrolysis of collagen. It has a triple helical structure with a high content of glycine, proline, and hydroxyproline residues. Gelatin that is formed from alkaline treatment of collagen has more carboxyl groups and a lower isoelectric point than that derived from acidic hydrolysis [15]. The physicochemical properties of gelatin depend on the method of extraction and the extent of thermal denaturation that occurs during the purification.

Gelatin nanoparticles can be prepared by various methods, including chemical cross-linking, water-in-oil (w/o) emulsification, and desolvation. Gelatin is cross-linked with agents such as glutaraldehyde. Efficient cross-linking usually results in decreased rate of drug release. The w/o emulsification involves extruding a preheated, aqueous solution of gelatin into vegetable oils, such as corn or olive oils [15]. The two-step desolvation method involves the dropwise addition of a water-miscible nonsolvent such as acetone and ethanol [16]. While the use of collagen, the parent compound of gelatin, in drug delivery is rare, collagen nanoparticles have been used to deliver genes by exploiting the electrostatic interaction between the positively charged polymer and negatively charged deoxyribonucleic acid (DNA) [17].

***Polysaccharides*** The macromolecular polysaccharides that include pullulan, mannan, and dextran are the main constituents of the cellular glycocalyx and play an important role in cell–cell adhesion and the cell–cell recognition process [18]. Pullulan is a nonimmunogenic, nontoxic, water-soluble, linear, nonionic polysaccharide with $\alpha(1–4)$ and $\alpha(1–6)$ linkages with free hydroxyl groups for drug conjugation [19]. Pullulans are intracellularly synthesized and secreted by a fungus, *Aureobasidium pullulans* [20]. On the other hand, dextrans are anionic glucose polymers derived from sucrose with $\alpha(1–6)$ glucosidic linkage. A class of enzymes, glucansucrases, produced by two genera of lactic acid bacteria, namely, *Leuconostoc* and *Streptococcus,* catalyze the synthesis of dextrans from sucrose. The extraction of dextrans as well as their physical properties and drug delivery aspects has been reviewed [21, 22]. When coated with mannans, the biological response of both natural and synthetic polymeric nanoparticles may be changed [23].

To make pullulan nanoparticles, the polymer must first be made hydrophobic, typically by conjugating alkyl groups, cholesterol groups, or succinyl groups. Hydrophobic pullulans self-assemble to form stable hydrogel nanoparticles [24]. Alternately, pullulan nanoparticles can be formed by cross-linking reverse micelles of the polymer with glutaraldehyde [25]. Nanoparticles form when a solution of pullulan acetate in *N,N*-dimethyl acetamide is used and dialyzed with borate buffer [26]. Dextran is commonly conjugated to other polymers, such as PLGA, PEG, polystyrene, and poly(methyl methacrylate) for preparation of nanoparticles. Complex coacervation has been used to prepare dextran nanoparticles using oppositely charged polymers such as polyethyleneimine [27]. Although nanoparticle

formulation using mannans has not been reported, mannan-coated nanoparticles increased cell binding and macrophage uptake [28].

***Albumin*** Human serum albumin, the most abundant plasma protein, is a positively charged, multifunctional protein and is involved in transport, ligand binding, and enzymatic activities. Albumin is a globular protein containing approximately 585 amino acids in an α-helical tertiary structure. Several exogenous and endogenous compounds can covalently or reversibly bind to albumin, and because of the excellent adsorptive properties of human serum albumin, this polypeptide can be adsorbed onto the surface of polymeric nanoparticles [29]. In addition, as it is amphoteric, albumin can be used as a surfactant during the preparation of nanoparticles where it is irreversibly adsorbed onto the surface of biodegradable polymers such as poly(lactic acid) (PLA) and PLGA [30]. These protein–particle interactions are mainly driven by electrostatic forces further stabilized by hydrophobic forces [31].

Albumin nanoparticles can be prepared by controlled desolvation, pH-induced coacervation, and/or chemical cross-linking with glutaraldehyde. Briefly, the pH of an aqueous solution of albumin is raised to about 9.0 and nanoparticles are precipitated by adding a miscible cosolvent such as acetone [32]. A study reports attempts to optimize the desolvation method to prepare albumin nanoparticles of a more controlled particle size and narrower particle size distribution [33].

***Gliadin*** Gliadin, a glycoprotein derived from gluten, is extracted from wheat and separated by capillary electrophoresis [34, 35]. Gliadin is classified as ω-5, ω-1,2, α, and γ- type based on its structure and electrophoretic mobility [36]. The glycoprotein is water insoluble due to the presence of interpolypeptide disulfide bonds and hydrophobic interactions. A limitation on the use of gliadins is that patient sensitivity causes an autoimmune disorder called celiac disease.

Gliadin nanoparticles are prepared by the desolvation method by first pouring an organic solution of polymer into an aqueous phase such as physiological saline containing a surfactant stabilizer (e.g., Pluronic). The nanoparticles are formed by evaporating the organic solvent.

## 2.2 Drug Delivery Applications and Biological Fate of Natural Polymeric Nanoparticles

***Alginate Nanoparticles*** Alginate nanoparticles have been used to formulate a wide range of drugs. Because they are prepared in an aqueous environment under mild conditions, alginate nanoparticles are particularly suitable for formulating proteins, peptides, and oligonucleotides [37]. Further, in addition to being biodegradable, alginates are nonimmunogenic. To decrease the rate of exchange of cations such as $Ca^{2+}$ with monovalent ions in the dissolution medium, the anionic alginates are often treated with cationic molecules such as chitosan, poly-*l*-lysine, or tripolyphosphate. Some examples of the wide range of applications of alginate-based nanoparticles are described. Alginate nanoparticles prepared with tripolyphosphate were used in oral delivery [38]. A study of physical properties demonstrated that alginate–chitosan nanoparticles are suitable for the delivery of DNA [39]. Alginate-coated chitosan nanoparticles increased stability and decreased the burst release of

ovalbumin [40]. A study reported that chitosan-stabilized alginate nanoparticles increased bioavailability and sustained release of antifungal drugs compared to PLGA nanoparticles [41]. Although predominantly used for oral administration, inhaled alginate nanoparticles improved the bioavailability of antitubercular drugs [42]. In vivo, alginate nanoparticles accumulate in the Kupffer cells, parenchymal cells of liver, and phagocytes of spleen and lungs [43, 44]. Alginate nanoparticles have also been reported to be absorbed into Peyer's patches, suggesting that they may enhance targeting to the intestinal mucosa [45]. In the body, the alginates degrade by acidic hydrolysis of the guluronic and mannuronic segments [46].

***Chitosan Nanoparticles*** In addition to low-molecular-weight drugs and nutraceuticals, chitosan nanoparticles are widely used in the delivery of macromolecules such as DNA and small interfering ribonucleic acid (siRNA) [47]. Apart from sustaining the release of macromolecules, chitosan nanoparticles protect them from nucleases. Placebo chitosan nanoparticles exhibited antibacterial activity against several microbes, including *Escherichia coli* [48]. The surface of chitosan nanoparticles was hydrophobically modified with linoleic acid for delivery of trypsin [49]. Other applications of chitosan nanoparticles include lung [50] and ocular delivery [51]. The primary amine group at the 2 position can be modified to tailor chitosan for specific applications. For example, chemical conjugation of these amine groups to methoxy-PEG groups increased water solubility [52]. Thiolation of chitosan enhanced the mucosal permeation of the nanoparticles [53]. Hydrophobically modified glycol chitosans that self-assemble into nanoparticles have been used to deliver doxorubicin [54]. Targeting chitosan nanoparticles to folate receptors on the surface of cells enhanced the transfection efficiency of DNA [55]. *N*-Succinyl chitosan nanoparticles containing 5-fluorouracil demonstrated excellent activity against sarcoma tumors [56]. Self-assembled *N*-acetyl histidine–conjugated glycol chitosan nanoparticles were efficiently internalized into cells by adsorptive endocytosis [57]. No toxic effects have been observed with chitosan nanoparticles. Upon intravenous (i.v.) administration, chitosan nanoparticles accumulated in the liver with minimal concentrations in the heart and lung [54].

***Gelatin Nanoparticles*** Gelatin nanoparticles have been used as a delivery system for several drugs, including pilocarpine, hydrocortisone [58], methotrexate [59], paclitaxel [60], and chloroquine [61]. High protein loading and sustained release were achieved using composite gelatin and PLGA nanoparticles [62]. Surprisingly, placebo gelatin nanoparticles exhibited antimelanoma activity in vivo [63]. Primary amine groups of gelatin molecule can be chemically conjugated or cross-linked using bifunctional cross-linkers. This is demonstrated in the delivery of biotinylated peptide nucleic acid using avidin-cross-linked gelatin nanoparticles [64]. PEGylation of gelatin nanoparticles containing hydrophilic drugs prolonged circulation time in the body [65]. Thiolated gelatin nanoparticles produced effective transfection of plasmid DNA encoding the green fluorescent protein [66]. Following endocytic uptake by the cells, gelatin nanoparticles concentrate in the perinuclear region [65]. In vitro, the gelatin nanoparticles are efficiently internalized into macrophages and monocytes [67]. In tumor-bearing mice, PEGylated gelatin nanoparticles predominantly accumulated in the liver and tumor [68] while in dendritic cells they are primarily localized in the lysosomes [69]. In vivo, gelatin is degraded by proteases to

amino acids. Although cardiotoxicity and mild immunogenicity were reported with gelatin nanoparticles covalently coupled to doxorubicin, this was attributed to the coupling reagent glutaraldehyde [70].

***Pullulan and Dextran Nanoparticles*** Pullulan nanoparticles successfully delivered HER2 oncoprotein to induce humoral and cellular immune responses against HER2-expressing murine sarcomas [71]. Hydrophobic polysaccharides such as pullulan and mannan enable soluble proteins to induce cellular immunity and therefore may be a potential delivery vehicle for vaccines [72]. pH-sensitive pullunan nanoparticles prepared by conjugating sulfonamides are stable at physiological pH but aggregate and release the encapsulated drug when exposed to the lower tumor pH [73]. Coating of magnetic nanoparticles with pullulan enhanced their cellular uptake by endocytosis [74]. Amphotericin-loaded, dextran–polyethyleneimine nanoparticles were active against *Candida albicans* [27]. Similarly, insulin-containing dextran–polyethyleneimine nanoparticles prolonged the hypoglycemic effect in diabetic rats [75]. Hydrogels prepared from blends of polyvinyl alcohol and dextrans have been used as matrices to entrap PLGA nanoparticles [76]. In vitro, immunofluorescent staining illustrated that pullulan nanoparticles are internalized by active endocytosis [25]. However, other studies suggest that both absorption and internalization of pullulan nanoparticles are inhibited by coating them with dextran [77].

***Albumin Nanoparticles*** A significant development in the drug delivery of albumin nanoparticles has been the recent marketing of the commercial product Abraxane® for chemotherapy of breast cancer. This delivery system is prepared by nab™ technology, which involves noncovalent complexation of albumin with paclitaxel. Albumin nanoparticles have been used to deliver antisense oligonucleotides, interferon γ, and anticytomegaloviral drugs [78, 79]. Intravitreal injection of gancicyclovir-loaded albumin nanoparticles was attempted to prolong residence time in the eye [79]. PLA nanoparticles coated with albumin degrade more rapidly in the gastrointestinal (GI) region, resulting in efficient delivery of water-soluble drugs across the GI tract [80]. In addition, intra-arterial chemotherapy with paclitaxel-containing albumin nanoparticles effectively treated squamous cell carcinoma [81]. Human serum albumin–polyethyleneimine nanoparticles optimized transfection of the luciferase gene in human, embryonic, and epithelial kidney cells [82]. Conjugation of the cellular targeting agent, folic acid, resulted in increased cellular uptake of albumin nanoparticles compared to unmodified particles [83, 84]. Transferrin-conjugated PEGylated albumin nanoparticles demonstrated enhanced uptake into the brain tissues [85]. Glycyrrhizin was conjugated to the amine groups of albumin nanoparticles targeting hepatocytes [86]. After i.v. administration, albumin-coated PLA nanoparticles were distributed in the liver, bone marrow, lymph nodes, spleen, and peritoneal macrophages [87]. In the body, albumin nanoparticles are actively taken up by macrophages.

***Gliadin Nanoparticles*** The hydrophobic nature of gliadin makes this polymer ideal for delivery of hydrophobic compounds such as all-trans retinoic acid and vitamin E. Gliadin nanoparticles adhered to the stomach mucosa and significantly increased the bioavailability of carbazole [88]. Typically, encapsulation efficiency

and drug loading of gliadin nanoparticles are higher for hydrophobic drugs [89]. Gliadin nanoparticles were targeted to *Helicobacter pylori* by chemical conjugation of lectin glycoproteins to their surface, resulting in a twofold increase in inhibition of bacterial activity compared to unmodified nanoparticles [90]. Although gliadins adhere to the mucosa, internalization of gliadin nanoparticles into cells has not been reported. More definitive studies are required to fully understand the biological fate of these particles.

## 3 SYNTHETIC BIODEGRADABLE POLYMERIC NANOPARTICLES

A detailed description of the methods of polymerization and variables employed in polymer synthesis are beyond the scope of this review and can be found in many texts and review articles. The focus of this section is to provide an overview of the properties and methods used to prepare nanoparticles from each class of synthetic polymer.

### 3.1 Synthetic Polymers: Physical Properties and Methods of Preparation of Nanoparticles

***Poly(lactic acid) and Poly(lactide-*co*-glycolide)*** These poly-hydroxy acids are approved for human use by the Food and Drug Administration (FDA) and have been widely used to prepare biodegradable nanoparticles. PLA exists in optically active and inactive forms and is a semicrystalline, hydrophobic molecule that degrades in the body over a period of months. Conversely, poly(glycolic acid) is amorphous and hydrophilic and degrades more rapidly than PLA. In aqueous media, these polymers degrade by random hydrolysis of ester bonds that is autocatalyzed in acidic media to form lactic and glycolic acids [91]. The factors that affect the rate of hydrolytic degradation include type and composition of the polymer backbone, nature of pendent groups, molecular weight, pH, enzymes, and geometry of the delivery device.

The preparation and characterization of PLA and PLGA nanoparticles have been extensively reviewed elsewhere [92, 93]. Various techniques may be used to prepare PLA and PLGA nanoparticles, including simple and multiple emulsions, nanoprecipitation, gas antisolvent method, supercritical fluid technology, coacervation/phase separation, and spray drying [91]. Briefly, in the single-emulsion method, an organic solution of the polymer and drug is emulsified with an aqueous solution of surfactant such as polyvinyl alcohol (PVA). While PLA and PLGA nanoparticles containing hydrophobic drugs are prepared by the two-phase emulsion method, a w/o/w multiple-emulsion method is needed to encapsulate hydrophilic drugs. In the phase separation method, the addition of a nonsolvent precipitates or coacervates the polymer from solution to encapsulate the drug. The experimental variables for each protocol can be altered to influence the physicochemical properties, such as particle size, particle size distribution, morphology, and zeta potential [93]. The release of encapsulated drug from PLA and PLGA nanoparticles may occur by a combination of diffusion and polymer degradation at a rate that is influenced by properties of the polymer and nanoparticles and the environment. The surface of

both PLGA and PLA nanoparticles can be modified to target cells and organs by conjugation with ligands such as folates, transferrin, HIV-TAT, aptamers, heparin, and lectins. The negative zeta potential of PLA and PLGA nanoparticles can be altered by coating with cationic polymers such as chitosan and polyethyleneimine, which promote nanoparticle–cell interaction. As with liposomes, PEGylation of PLGA nanoparticles prolongs circulation times in the body.

***Poly(ε-caprolactone)*** Poly(ε-caprolactone) is a semicrystalline polymer synthesized by anionic, cationic, free-radical, or ring-opening polymerization [94]. It is available in a range of molecular weights and degrades by bulk hydrolysis autocatalyzed by the carboxylic acid end groups. The presence of enzymes such as protease, amylase, and pancreatic lipase accelerates polymer degradation [95]. The various methods of preparation of poly(ε-caprolactone) nanoparticles include emulsion polymerization, interfacial deposition, emulsion–solvent evaporation, desolvation, and dialysis. These methods and various applications are extensively reviewed [94].

***Polyanhydrides*** Polyanhydrides have a hydrophobic backbone with a hydrolytically labile anhydride linkage. These polymers widely vary in chemical composition and include aliphatic, aromatic, and fatty acid–based polyanhydrides. The rate of degradation depends on the chemical composition of the polymer. In general, aliphatic polyanhydrides degrade more rapidly than the aromatic polymer. Hence, copolymer blends with varying ratios of aliphatic-to-aromatic polyanhydrides can be synthesized to suit specific applications.

The synthesis and physical properties of polyanhydrides have been reviewed [96, 97]. Polyanhydride nanospheres are commonly prepared by the emulsion–solvent evaporation method using PVA as a stabilizer. However, as polyanhydrides are hydrolabile, they need to be flash frozen in liquid nitrogen and lyophilized immediately [98]. An example of their use to deliver drugs is entrapment of bovine zinc insulin by phase inversion nanoencapsulation [99]. Although not used to formulate nanoparticles, polyanhydride microspheres have been prepared using alternate techniques involving nonaqueous solvents, such as solid/oil/oil double emulsion and cryogenic atomization techniques [100]. The surface of polymeric nanoparticles can be modified for targeted delivery by reaction of the anhydride with an amino group to form an amide linkage with a ligand [101]. However, application of the ligand-conjugated nanoparticles for drug delivery is yet to be explored extensively.

***Poly(alkyl-cyanoacrylates)*** As poly(alkyl-cyanoacrylates) form strong bonds with polar substrates including the skin and living tissues, they exhibit bioadhesive properties. These polymers are synthesized by free-radical, anionic, or zwitterionic polymerization. As detailed in a recent review, poly(alkyl-cyanoacrylate) nanoparticles are prepared by emulsion polymerization, interfacial polymerization, nanoprecipitation, and emulsion–solvent evaporation methods [102].

***Solid–lipid Nanoparticles*** Solid–lipid nanoparticles (SLNs) are obtained by high-pressure homogenization of molten lipids in the presence of surfactant. The major advantage of using SLN is the ability to have high drug loading and prolonged stability with lipophilic compounds. In addition to drug delivery, SLNs have been used in dermatological and cosmetic preparations. The most common carriers used to

prepare SLNs are triglycerides, glycerides, fatty acids, steroids, and waxes and may contain a wide range of emulsifiers. Methods of preparation of SLNs include high speed hot and cold homogenization, ultrasound, emulsion–solvent evaporation, and microemulsion. The various parameters involved in the preparation of SLNs have been optimized and thoroughly reviewed and their physicochemical properties elucidated [103, 104].

***Other Synthetic Biodegradable Polymers*** Although well investigated for drug delivery, polyorthoesters, polyurethanes, and polyamides have found limited application as nanoparticles. A report documents the synthesis and characterization of polyorthoester nanoparticles [105].

### 3.2 Drug Delivery Applications and Biological Fate of Synthetic Biodegradable Polymers

**PLA/PLGA *Nanoparticles*** A wide range of hydrophilic and hydrophobic drugs, including low- and high-molecular-weight compounds, have been encapsulated into PLGA/PLA nanoparticles for a wide range of therapeutic applications and routes of administration, including oral, intravenous, intra-arterial, nasal, and inhalation delivery [92, 106]. Extensive reviews describing the application of PLGA nanoparticles in drug therapy are available [92, 107, 108].

After i.v. administration, the PLGA nanoparticles are removed from systemic circulation by the mononuclear phagocytic system in the liver [109]. PLGA nanoparticles enter cells by absorptive endocytosis and may escape the lysosomes to accumulate in cytoplasm [110, 111]. In the body, PLA and PLGA degrade into the monomers lactic and glycolic acids, which enter the citric acid cycle, where they are metabolized and eliminated as $CO_2$ and $H_2O$. Glycolic acid may also be excreted through the kidney [91]. Humoral response to these results in mild, acute, and chronic inflammation [112].

***Poly(ε-caprolactone) Nanoparticles*** As important applications of poly(ε-caprolactone) nanoparticles have been reviewed previously, only representative examples will be given [94]. Decreased cardiovascular adverse effects of cartelol was observed upon ophthalmic administration of poly(ε-caprolactone) nanocapsules [113]. Poly(ε-caprolactone) nanoparticles, nanocapsules, or nanoemulsions increased the ocular uptake of indomethacin [114]. The cytotoxicity of retinoic acid was enhanced when delivered in core-shell-type nanoparticles formed from poly(ε-caprolactone–polyethylene glycol) blends [115]. Alternately, these nanoparticles were also chemically modified with folic acid to target the folate receptors for enhanced cellular uptake [116]. Coating poly(ε-caprolactone) nanoparticles with polysaccharides such as galactose resulted in lectin-dependent aggregation, demonstrating the potential as a targeted delivery system to hepatocytes [117]. Stable complexes were formed between anionic DNA and chitosan-modified poly(ε-caprolactone) nanoparticles, demonstrating high transfection efficiency [118]. After i.v. administration, these particles are eliminated by macrophages of the reticuloendothelial system and biodegradation occurs by bulk scission of polymer chains [94]. However, dextran-coated poly(ε-caprolactone) nanoparticles lowered their uptake into macrophages [119].

***Polyanhydride Nanoparticles*** Polyanhydrides have been more commonly used to prepare microparticles than nanoparticles. However, the technology is adaptable for nanoparticles. The transfection efficiency of firefly luciferase DNA was enhanced when delivered in nanoparticles prepared from polyanhydride–lactic acid blends, demonstrating the potential application in gene delivery [120]. The degradation and elimination of polyanhydrides have been reviewed [97]. In vivo, the anhydride bond degrades to form diacid monomers that are eliminated from the body.

***Poly(alkyl-cyanoacrylate) Nanoparticles*** The applications of poly(alkyl-cyanoacrylate) nanoparticles have been reviewed elsewhere and therefore only representative examples are described [102]. Because of their adhesive properties, nanoparticles have the potential to prophylactically treat candidiasis of the oral cavity [121]. Not surprisingly, poly(alkyl-cyanoacrylate) nanoparticles have been used to deliver drugs to tumors [122]. Enhanced absorption and prolonged hypoglycemic effect were observed when insulin was delivered in poly(alkyl-cyanoacrylate) nanoparticles [121]. Nuclear accumulation of antisense oligonucleotides into vascular smooth muscle cells was increased when delivered using poly(alkyl-cyanoacrylate) nanoparticles [123]. Dextran-coated poly(alkyl-cyanoacrylate) nanoparticles lowered protein adsorption in the blood [124].

Poly(alkyl-cyanoacrylates) degrade by hydrolysis of the ester bond of the alkyl side chains to form water-soluble alkyl alcohol and poly(cyanoacrylic acid). After in vivo administration, poly(alkyl-cyanoacrylate) nanoparticles are predominantly distributed in the liver, spleen, and bone marrow where they are endocytosed into the cells to become localized in the lysosomes. However, the mechanisms of lysosomal escape have not been identified [102].

***Solid–Lipid Nanoparticles*** SLNs have been used to deliver small molecules and macromolecules such as DNA and peptides. The in vitro and in vivo applications of SLNs are reviewed elsewhere [125, 126]. The stability and oral bioavailability of insulin were enhanced when administered in wheat germ agglutinin–conjugated nanoparticles [127]. A polyoxyethylene stearate coat on the SLN confers stealth properties [128].

## 4 THERMOSENSITIVE AND pH-SENSITIVE NANOPARTICLES

### 4.1 Physical Properties and Methods of Preparation

***Thermosensitive Polymeric Nanoparticles*** Thermoresponsive "smart" polymers that change their physical characteristics, such as shape, surface properties, or solubility, in response to changes in temperature have been developed to target drugs [129]. The encapsulated drug is released when the nanoparticles are exposed to changes in temperature, such as body temperature or external heat source. For example, poly(*N*-isopropyl acrylamide), or poly(NIPAAm), is water soluble at room temperature but aggregates and is insoluble above its lower critical solution temperature (LCST), which typically ranges between 37 and 42 °C [129]. The methods used to prepare thermosensitive nanoparticles have been thoroughly reviewed [127,

129–131]. Interestingly, when heated above the LCST, poly(NIPAAm)–PEG block copolymers spontaneously self-assemble into nanoparticles whose size is controlled by the rate of heating [132].

***pH-Sensitive Polymeric Nanoparticles*** Enteric-coated polymers have long been used to protect drugs from the acidic pH in the stomach. Chitosan is a pH-sensitive polymer that is soluble only in acidic media. Similarly, the solubility of sulfonamide-modified pullulans is dependent on pH [133]. Chitosan–insulin nanoparticles are stable at low pH but dissociate at physiological pH, releasing insulin [134]. The pH of tumor interstitium is lower than the normal tissue. The pH-sensitive polyethylene oxide–poly(β-amino ester) microparticles containing paclitaxel significantly reduced tumor burden [135, 136]. Polyketals, a new generation of acid-sensitive polymers, degrade by acidic hydrolysis [137]. Low pH inside the endosomes facilitated the escape of pH-responsive plasmid–lipid nanoparticles, resulting in enhanced transfection efficiency [138].

### 4.2 Drug Delivery Applications and Biological Fate of Thermosensitive and pH-Sensitive Nanoparticles

Thermosensitive block copolymer nanoparticles containing doxorubicin increased cytotoxicity against Lewis lung carcinoma cells when activated by heating above the LCST [139]. Chitosan was chemically conjugated to NIPAAm/vinyl laurate copolymer to enhance gene transfection in mouse myoblast cells [140]. Upon i.v. administration, poly(NIPAAm) nanoparticles are taken up by the reticuloendothelial cells of the liver and mild inflammatory and fibrotic responses are observed [141].

After internalization into SKOV-3 (ovarian adenocarcinoma) cells, polyethylene oxide–modified poly(β-amino ester) nanoparticles rapidly disintegrated and released the drug in the low pH of the endosomes [142]. Intravenous administration of polyethylene oxide–modified poly(β-amino ester) nanoparticles containing paclitaxel significantly reduced tumor burden in mice with ovarian cancer [142]. *N*-Acetyl histidine–conjugated glycol chitosan nanoparticles were used to deliver drugs into the cytoplasm. These pH-responsive nanoparticles are endocytosed where their structural integrity is lost due to protonation of imidazoles, resulting in their endolysosomal escape [57]. As most of the pH-sensitive biodegradable polymers are blends of natural and synthetic polymers, they are degraded by mechanisms specific to individual polymers.

## 5 APPLICATIONS OF BIODEGRADABLE NANOPARTICLES OTHER THAN DRUG DELIVERY

Diagnosis and imaging are important applications of nanoparticles that are briefly described. The ability to encapsulate or conjugate fluorescent compounds into or onto biodegradable nanoparticles has been used extensively in imaging. Compounds that have been encapsulated into nanoparticles for imaging include gadolinium, fluorescein isothiocyanate (FITC)–dextrans, Bodipy, and the autofluorescent anticancer drug doxorubicin. Nanoparticles encapsulating radioactive ligands, such as

$^{99m}$Tc-labeled colloids and $^{111}$In, have been used in scintigraphic imaging [143]. In addition, fluorescent or radioactive moieties can be targeted by noncovalently or covalently tagging the nanoparticles through avidin–biotin conjugation and thiol formation [144]. In vitro imaging enables the dynamics of cellular internalization and localization of nanoparticles to be studied. For example, Bodipy—loaded PLGA nanoparticles have been used to study their cellular disposition in vitro [145] as well as the effect of storage temperature on their physical properties [146]. The biotinylated antibody, specific to the CD3 antigen on lymphocytes, was chemically conjugated to nanoparticles and their binding to leukemic and primary T lymphocytes investigated [143]. The instrumentation that facilitates imaging of nanoparticles includes confocal laser scanning microscopy, liquid scintigraphy, and flow cytometry.

## 6 PHYSICOCHEMICAL CHARACTERIZATION OF POLYMERIC NANOPARTICLES

The selection of polymer is critical to the performance, properties, and application of nanoparticles. Further, the physicochemical properties of the polymer will determine the surface properties of nanoparticles with polymer molecular weight, hydrophobicity, and glass transition temperature being particularly important. The surface properties that influence their biodistribution and cellular response include particle size, zeta potential, and surface hydrophilicity.

### 6.1 Molecular Weight

Many reviews and textbooks document the experimental methods available to determine number-average, weight-average, viscosity-average, and $z$-average molecular weights. The molecular weight of the polymer will influence many parameters during the preparation of nanoparticles as well as their properties such as drug loading and rate and extent of drug release.

### 6.2 Hydrophobicity

The experimental methods of determining hydrophobicity include interaction chromatography and two-phase partition using fluorescent or radiolabeled hydrophobic probes [147]. Generally, a more hydrophobic polymer degrades more slowly and releases drugs at a decreased rate. Hence, a blend of hydrophobic and hydrophilic polymers can be used to tailor drug release kinetics. Depending on the conditions of polymerization, mixtures of two or more monomers of different types yield block, alternate, or random cocopolymers, each of which will possess different hydrophobicities and consequently drug release kinetics. Additionally, the degree of hydrophobicity can change the mechanism of degradation. For example, by inserting hydrophilic ethylene glycol moieties within the hydrophobic backbone, polyanhydrides have been tailored to specifically degrade by bulk or surface erosion [148]. Incorporation of aromatic side chains generally increases hydrophobicity. A difference in hydrophobicity of individual polymers results in microphase separation of copolymers followed by a thermodynamic partition and altered release profile of encapsulated drugs [149].

### 6.3 Glass Transition Temperature

The morphology and physical properties of nanoparticles are affected by the glass transition temperature and physical state of the polymer or polymer blends. The glass transition temperature ($T_g$) is the temperature at which polymers undergo a change in heat capacity and transform their physical arrangement. The crystallinity or amorphous nature of the polymers can be altered by synthesizing polymer blends of varying ratios. The $T_g$ of a polymer is experimentally determined by differential scanning calorimetry (DSC). The $T_g$ is mathematically calculated using multidimensional lattice representations and statistical methods [150]. The various factors that influence the $T_g$ of polymers include molecular weight, composition and stereochemistry of the polymer backbone, type and length of pendent groups, additives such as copolymers, and plasticizers. Additives and copolymers may also be used to alter the $T_g$. For example, incorporation of mPEG significantly lowered the $T_g$ of PLA, resulting in rapid release of drug from the nanoparticles [151].

### 6.4 Particle Size and Particle Size Distribution Instrumentation

Particle size is a critical characteristic of nanoparticles and by definition differentiates them from microparticles. Particle size and particle size distribution play an important role in the biological performance of the nanoparticles. Important techniques for measuring particle size are photon correlation spectroscopy (PCS), electron microscopy, and atomic force microscopy, which have been comprehensively described [147]. Particle size and particle size distribution are determined by the method of preparation and experimental variables during manufacture. For example, in the emulsion–solvent evaporation method, the particle size is determined by controlling the energy of emulsification and the resulting droplet size of the internal phase. Control of particle size is also possible by altering experimental variables such as the volume and phase ratio of the internal and external phases and the concentration, type, and viscosity of the emulsifying agent. As an example, the experimental parameters that can influence the particle size and size distribution of PLGA nanoparticles include the method used, polymer concentration, surfactant concentration, stirring speed, ratio of aqueous and organic phases, and concentration of the emulsifier [152]. Methods of separation such as filtration and centrifugation can also influence particle size distribution. When chitosan nanoparticles are prepared using ionic gelation, the critical parameters for a narrow particle size distribution are molecular weight, degree of deacetylation, concentration and molar ratio of chitosan, and the presence and concentrations of counterions (e.g., tripolyphosphate) [12].

### 6.5 Surface Charge and Zeta Potential

The surface charge of nanoparticles is important because it determines the nature and extent of aggregation of colloids and their interaction with cells and other biological components within the body. The zeta potential is the potential at the solid–liquid interface and is commonly determined using light scattering [153]. Decreasing the zeta potential of nanoparticles below a critical value increases the rate and

extent of their aggregation, resulting in conglomerates with different physical properties. The surface charge can also be altered by chemical conjugation of ligands to nanoparticles. As the following examples demonstrate, surface charge may be modified to facilitate targeting. The degree of positive charge on chitosan–tripolyphosphate nanoparticles was increased by controlling the processing parameters, resulting in a stronger electrostatic interaction with negatively charged cell surfaces [13]. An increase in surface charge increased the adsorption of nanoparticles to plasma proteins [154]. Positively charged ligands such as chitosan as well as peptides neutralize the surface charge of negatively charged particles made using polymers such as PLGA [155]. Positively charged tripalmitin nanoparticles increased circulation time and higher blood concentrations of etoposide compared to negatively charged particles [156]. On the other hand, neutrally charged particles may be protected from opsonization [157].

### 6.6 Surface Hydrophilicity

The surface hydrophilicity of nanoparticles also influences the nature and extent of their interaction with cells and their behavior in the biological environment. The techniques for determining surface hydrophilicity have been described previously [147]. As hydrophobic nanoparticles are opsonized and eliminated by the mononuclear phagocytic system, surface hydrophilicity is an important parameter to ensure longer circulation times. The hydrophilicity to the nanoparticles may be modified using several methods, including adsorption of nonionic surfactants such as Poloxamer as well as adsorption or conjugation of hydrophilic polymers such as polysaccharides, polyacrylamides, PVA, poly(*N*-vinyl-2-pyrrolidone), PEG, polyoxamines, and polysorbates [157]. Stealth nanoparticles can be prepared by adsorbing or conjugating PEG to the particle surface, which protects them from opsonization [157]. The biological properties of stealth nanoparticles have been reviewed elsewhere [158].

### 6.7 Drug Loading and Encapsulation Efficiency

Drug loading or the weight of the drug encapsulated in the polymeric carrier is expressed as a percentage (w/w) of the delivery system. Encapsulation efficiency is the difference between the amount of the drug encapsulated into nanoparticles compared to the total amount added during preparation. The encapsulation efficiency is dependent on the properties of the polymer and excipients and the method of preparation of nanoparticles. For example, increased miscibility between the nanoparticle and water at lower temperatures and high rate of solvent evaporation at high temperatures result in the formation of an outer sphere wall, increasing encapsulation efficiency [159]. Efficient encapsulation depends on (1) physicochemical properties of the drug, such as solubility, hydrophilicity, and crystallinity; (2) physicochemical properties of the polymer, including molecular weight, hydrophobicity, drug–polymer interactions, and solubility parameter; and (3) variables involved in the preparation of the nanoparticles, such as drug–polymer weight ratio, solvents, method of preparation of particles, solvent evaporation rate, type as well as volume, and concentration of the surface-active agent used.

### 6.8 Drug Release

Many cumulative and differential dissolution methods are used to monitor the rate and extent of drug release. The rate of drug release from nanoparticles depends on the chemical properties of the polymer, properties of particles such as hydrophobicity and surface area, and environmental factors such as pH. Drug release may occur by one or a combination of the following: diffusion, dissolution, degradation, or swelling. Hence, drug release normally follows first- rather than zero-order kinetics [160–162]. A disadvantage of nanoparticles is that they may have a significant burst release (~40%) due to their high surface area–mass ratio. Surface treatment of nanoparticles can reduce the burst release of encapsulated drug [163, 164].

### 6.9 Physical Stability of Polymeric Nanoparticles

The main indication of physical instability of nanoparticles is irreversible aggregation. The principal factors that affect the extent of aggregation are type of polymer, zeta potential, duration and temperature of storage, and presence of electrolytes. Generally, homopolymers such as PLA and poly(ε-caprolactone) are more stable than copolymers. However, the stability of PLGA copolymers can be prolonged by appropriate storage [165]. The duration of storage can also be important. For example, poly(σ-caprolactone) nanoparticles aggregated after four months [166]. Another study compared the physical stability of PLGA nano- and microspheres after incubation at different temperatures and demonstrated that, while microspheres did not aggregate, the extent of aggregation of nanospheres increased as storage temperature increased [146]. Ideally, PLGA nanoparticles should be stored desiccated at 4 °C. Aqueous dispersions of SLN are stable for three years [167]. It is important to note that electrolytes can accelerate the aggregation and instability of nanoparticles. For example, multivalent ions caused solid–lipid nanoparticles to gel [168]. Physical stability can be improved by coating the surface of nanoparticles with hydrophilic polymers or surfactants [169]. As discussed previously, a reduction in the zeta potential below critical values causes flocculation of colloidal systems. For example, neutralization of the surface charge of colloidal systems by addition of cationic oligonucleotides resulted in aggregation [170].

## 7 TARGETING NANOPARTICLES BY SURFACE CONJUGATION WITH LIGANDS

Nanoparticles can be targeted to cells using specific or nonspecific ligands. Specific ligands can be covalently conjugated to the nanoparticles to target them to a selected site of action. Some examples of targeting ligands are folic acid [171], transferrin [172], lectin [173], and epidermal growth factor [174]. Nonspecific targeting is achieved by attaching ligands that alter the biodistribution of the nanoparticles. Examples of nonspecific targeting include the use of PEG for imparting stealth properties or hydrolytically cleavable linkers that protect the drug from degradation. Common methods of conjugation of ligands include carbodiimide coupling, glutaraldehyde conjugation, peptide bond formation, disulfide, and thiol linkages [175–178].

## 8 CELLULAR TRAFFICKING OF BIODEGRADABLE NANOPARTICLES

Biodegradable nanoparticles are internalized by one or more of the following mechanisms: phagocytosis, macropinocytosis, and clathrin- and caveolin-mediated endocytosis (Figure 1). While phagocytosis by macrophages eliminates nanoparticles from the body, efficient cellular uptake occurs when high-affinity receptors capture the nanoparticles through receptor-mediated endocytosis [179]. On the cell surface, nanoparticles activate caveolin, a dimeric protein, resulting in their internalization through caveolae. Clathrin-mediated endocytosis occurs when nanoparticles accumulate on the plasma membrane and clathrin-coated pits are formed to transport the nanoparticles into the cell, resulting in the formation of endosomes. Macropinocytosis is restricted to larger particles, such as nanoparticles typically greater than 800 nm. The mechanism by which particles enter cells depends on the composition of the nanoparticles, type of the cell, and particle size. For example, the uptake of chitosan nanoparticles into lung epithelial and Caco-2 cells was mediated, in part, by the clathrin-mediated pathway [180, 181]. However, PLGA nanoparticles of size 100 nm are internalized by the clathrin- and caveolin-independent pathway [110]. There are comprehensive reviews of the mechanisms responsible for the uptake of nanoparticles [182–184].

Nanoparticles that are internalized into cells by these mechanisms first enter the primary endosomes of the cell and are then transported into sorting endosomes. While some nanoparticles in the sorting endosomes are transported out of the cell by recycling endosomes, the remaining nanoparticles are transported into secondary endosomes that fuse with the lysosomes [107]. The surface charge of PLGA nanoparticles is reversed in the acidic lysosome, resulting in their escape into the cytoplasm [111]. A high external concentration of nanoparticles outside the cell prolongs their intracellular concentration within the cytoplasm [107].

Ligand-mediated endocytosis targets specific cell surface receptors and these nanoparticles are internalized by a receptor-mediated endocytic pathway. Examples of targeting transferrin, folate, lectins, and epidermal growth factor receptors are contained in the literature [185–187].

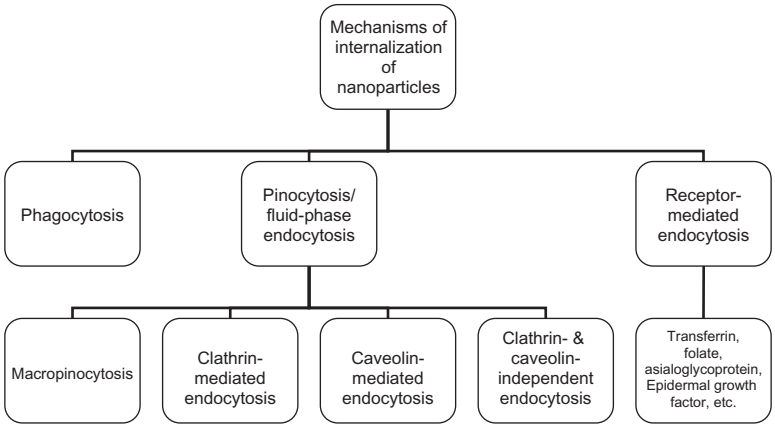


**FIGURE 1** Mechanisms of cellular internalization of biodegradable nanoparticles.

## 9 CONCLUSIONS

Biodegradable nanoparticles are a very active area of research in drug delivery, imaging, and diagnostics. This review has primarily focused on drug delivery applications. Natural and synthetic polymers used to prepare nanoparticles were discussed as well as their physicochemical properties that influence the biological performance of particles. Methods used to prepare and characterize the properties of nanoparticles have also been reviewed. Specific and nonspecific methods used to target nanoparticles to cells were mentioned as well as the mechanism for cellular and intracellular transport. As research into the various uses of biodegradable nanoparticles increases, so will our knowledge to further optimize their preparation and formulation and hence improve drug therapy and diagnosis.

## REFERENCES


1. Dang, J. M., and Leong, K. W. (2006), Natural polymers for gene delivery and tissue engineering, *Adv. Drug Deliv. Rev.*, 58(4), 487–499.
2. Singla, A. K., and Chawla, M. (2001), Chitosan: Some pharmaceutical and biological aspects—an update, *J. Pharm. Pharmacol.*, 53(8), 1047–1067.
3. Kulkarni, A. R., Lin, Y. H., Liang, H. F., Chang, W. C., Hsiao, W. W., and Sung, H. W. (2006), A novel method for the preparation of nanoaggregates of methoxy polyethyleneglycol linked chitosan, *J. Nanosci. Nanotechnol.*, 6(9–10), 2867–2873.
4. Sanchez-Machado, D. I., Lopez-Cervantes, J., Lopez-Hernandez, J., Paseiro-Losada, P., and Simal-Lozano, J. (2004), Determination of the uronic acid composition of seaweed dietary fibre by HPLC, *Biomed. Chromatogr.*, 18(2), 90–97.
5. Dusseault, J., Tam, S. K., Menard, M., Polizu, S., Jourdan, G., Yahia, L., and Halle, J. P. (2006), Evaluation of alginate purification methods: Effect on polyphenol, endotoxin, and protein contamination, *J. Biomed. Mater. Res. A.*, 76(2), 243–251.
6. Tonnesen, H. H., and Karlsen, J. (2002), Alginate in drug delivery systems, *Drug Dev. Ind. Pharm.*, 28(6), 621–630.
7. Rajaonarivony, M., Vauthier, C., Couarraze, G., Puisieux, F., and Couvreur, P. (1993), Development of a new drug carrier made from alginate, *J. Pharm. Sci.*, 82(9), 912–917.
8. Poncelet, D., Lencki, R., Beaulieu, C., Halle, J. P., Neufeld, R. J., and Fournier, A. (1992), Production of alginate beads by emulsification/internal gelation. I. Methodology, *Appl. Microbiol. Biotechnol.*, 38(1), 39–45.
9. Reis, C. P., Neufeld, R. J., Vilela, S., Ribeiro, A. J., and Veiga, F. (2006), Review and current status of emulsion/dispersion technology using an internal gelation process for the design of alginate particles, *J. Microencapsul.*, 23(3), 245–257.
10. Agnihotri, S. A., Mallikarjuna, N. N., and Aminabhavi, T. M. (2004), Recent advances on chitosan-based micro- and nanoparticles in drug delivery, *J. Controlled Release*, 100(1), 5–28.
11. Bodnar, M., Hartmann, J. F., and Borbely, J. (2005), Preparation and characterization of chitosan-based nanoparticles, *Biomacromolecules*, 6(5), 2521–2527.
12. Zhang, H., Oh, M., Allen, C., and Kumacheva, E. (2004), Monodisperse chitosan nanoparticles for mucosal drug delivery, *Biomacromolecules*, 5(6), 2461–2468.
13. Gan, Q., Wang, T., Cochrane, C., and McCarron, P. (2005), Modulation of surface charge, particle size and morphological properties of chitosan-TPP nanoparticles intended for gene delivery, *Colloids Surf. B. Biointerfaces*, 44(2–3), 65–73.

14. Lopez-Leon, T., Carvalho, E. L., Seijo, B., Ortega-Vinuesa, J. L., and Bastos-Gonzalez, D. (2005), Physicochemical characterization of chitosan nanoparticles: Electrokinetic and stability behavior, *J. Colloid Interface Sci.*, 283(2), 344–351.
15. Young, S., Wong, M., Tabata, Y., and Mikos, A. G. (2005), Gelatin as a delivery vehicle for the controlled release of bioactive molecules, *J. Controlled Release*, 109(1–3), 256–274.
16. Azarmi, S., Huang, Y., Chen, H., McQuarrie, S., Abrams, D., Roa, W., Finlay, W. H., Miller, G. G., and Lobenberg, R. (2006), Optimization of a two-step desolvation method for preparing gelatin nanoparticles and cell uptake studies in 143B osteosarcoma cancer cells, *J. Pharm. Pharmacal. Sci.*, 9(1), 124–132.
17. Lee, C. H., Singla, A., and Lee, Y. (2001), Biomedical applications of collagen, *Int. J. Pharm.*, 221(1–2), 1–22.
18. Sihorkar, V., and Vyas, S. P. (2001), Potential of polysaccharide anchored liposomes in drug delivery, targeting and immunization, *J. Pharm. Pharmacal. Sci.*, 4(2), 138–158.
19. Jeong, Y. I., Na, H. S., Oh, J. S., Choi, K. C., Song, C. E., and Lee, H. C. (2006), Adriamycin release from self-assembling nanospheres of poly(DL-lactide-*co*-glycolide)-grafted pullulan, *Int. J. Pharm.*, 322(1–2), 154–160.
20. Leathers, T. D. (2005), in Steinbuchel, A. R., (Ed.), *Polysaccharides and Polyamides in the Food Industry. Properties, Production, and Patents*, Vol. 1, Wiley-VCH, Weinheim, pp. 387–421.
21. Garach, V. (1975), The synthesis of dextrans and levans—a review of the literature, *Diastema*, 4(3), 25–28.
22. Mehvar, R. (2000), Dextrans for targeted and sustained delivery of therapeutic and imaging agents, *J. Controlled Release*, 69(1), 1–25.
23. Tizard, I. R., Carpenter, R. H., McAnalley, B. H., and Kemp, M. C. (1989), The biological activities of mannans and related complex carbohydrates, *Mol. Biother.*, 1(6), 290–296.
24. Hasegawa, U., Nomura, S. M., Kaul, S. C., Hirano, T., and Akiyoshi, K. (2005), Nanogel-quantum dot hybrid nanoparticles for live cell imaging, *Biochem. Biophys. Res. Commun.*, 331(4), 917–921.
25. Gupta, M., and Gupta, A. K. (2004), Hydrogel pullulan nanoparticles encapsulating pBUDLacZ plasmid as an efficient gene delivery carrier, *J. Controlled Release*, 99(1), 157–166.
26. Na, K., Lee, E. S., and Bae, Y. H. (2003), Adriamycin loaded pullulan acetate/sulfonamide conjugate nanoparticles responding to tumor pH: pH-dependent cell interaction, internalization and cytotoxicity in vitro, *J. Controlled Release*, 87(1–3), 3–13.
27. Tiyaboonchai, W., Woiszwillo, J., and Middaugh, C. R. (2001), Formulation and characterization of amphotericin B-polyethylenimine-dextran sulfate nanoparticles, *J. Pharm. Sci.*, 90(7), 902–914.
28. Cui, Z., Hsu, C. H., and Mumper, R. J. (2003), Physical characterization and macrophage cell uptake of mannan-coated nanoparticles, *Drug Dev. Ind. Pharm.*, 29(6), 689–700.
29. Quinlan, G. J., Martin, G. S., and Evans, T. W. (2005), Albumin: Biochemical properties and therapeutic potential, *Hepatology*, 41(6), 1211–1219.
30. Verrecchia, T., Huve, P., Bazile, D., Veillard, M., Spenlehauer, G., and Couvreur, P. (1993), Adsorption/desorption of human serum albumin at the surface of poly(lactic acid) nanoparticles prepared by a solvent evaporation process, *J. Biomed. Mater. Res.*, 27(8), 1019–1028.
31. Bousquet, Y., Swart, P. J., Schmitt-Colin, N., Velge-Roussel, F., Kuipers, M. E., Meijer, D. K., Bru, N., Hoebeke, J., and Breton, P. (1999), Molecular mechanisms of the adsorption

of a model protein (human serum albumin) on poly(methylidene malonate 2.1.2) nanoparticles, *Pharm. Res.*, 16(1), 141–147.

32. Lin, W., Garnett, M. C., Davis, S. S., Schacht, E., Ferruti, P., and Illum, L. (2001), Preparation and characterisation of rose Bengal-loaded surface-modified albumin nanoparticles, *J. Controlled Release*, 71(1), 117–126.
33. Langer, K., Balthasar, S., Vogel, V., Dinauer, N., von Briesen, H., and Schubert, D. (2003), Optimization of the preparation process for human serum albumin (HSA) nanoparticles, *Int. J. Pharm.*, 257(1–2), 169–180.
34. Mamone, G., Addeo, F., Chianese, L., Di Luccia, A., De Martino, A., Nappo, A., Formisano, A., De Vivo, P., and Ferranti, P. (2005), Characterization of wheat gliadin proteins by combined two-dimensional gel electrophoresis and tandem mass spectrometry, *Proteomics*, 5(11), 2859–2865.
35. Ezpeleta, I., Irache, J. M., Stainmesse, S., Chabenat, C., Gueguen, J., Popineau, Y., and Orecchioni, A. (1996), Gliadin nanoparticles for the controlled release of all-trans-retinoic acid, *Int. J. Pharm.*, 131(2), 191–200.
36. Wieser, H. (1996), Relation between gliadin structure and coeliac toxicity, *Acta Paediatr. Suppl.*, 412, 3–9.
37. Lambert, G., Fattal, E., and Couvreur, P. (2001), Nanoparticulate systems for the delivery of antisense oligonucleotides, *Adv. Drug Deliv. Rev.*, 47(1), 99–112.
38. Bodmeier, R., Chen, H. G., and Paeratakul, O. (1989), A novel approach to the oral delivery of micro- or nanoparticles, *Pharm. Res.*, 6(5), 413–417.
39. Douglas, K. L., and Tabrizian, M. (2005), Effect of experimental parameters on the formation of alginate-chitosan nanoparticles and evaluation of their potential application as DNA carrier, *J. Biomater. Sci. Polym. Ed.*, 16(1), 43–56.
40. Borges, O., Borchard, G., Verhoef, J. C., de Sousa, A., and Junginger, H. E. (2005), Preparation of coated nanoparticles for a new mucosal vaccine delivery system, *Int. J. Pharm.*, 299(1–2), 155–166.
41. Pandey, R., Ahmad, Z., Sharma, S., and Khuller, G. K. (2005), Nano-encapsulation of azole antifungals: Potential applications to improve oral drug delivery, *Int. J. Pharm.*, 301(1–2), 268–276.
42. Zahoor, A., Sharma, S., and Khuller, G. K. (2005), Inhalable alginate nanoparticles as antitubercular drug carriers against experimental tuberculosis, *Int. J. Antimicrob. Agents*, 26(4), 298–303.
43. Yi, Y. M., Yang, T. Y., and Pan, W. M. (1999), Preparation and distribution of 5-fluorouracil (125)I sodium alginate-bovine serum albumin nanoparticles, *World J. Gastroenterol.* 5(1), 57–60.
44. Ahmad, Z., Pandey, R., Sharma, S., and Khuller, G. K. (2006), Pharmacokinetic and pharmacodynamic behaviour of antitubercular drugs encapsulated in alginate nanoparticles at two doses, *Int. J. Antimicrob. Agents*, 27(5), 409–416.
45. Borges, O., Cordeiro-da-Silva, A., Romeijn, S. G., Amidi, M., de Sousa, A., Borchard, G., and Junginger, H. E. (2006), Uptake studies in rat Peyer's patches, cytotoxicity and release studies of alginate coated chitosan nanoparticles for mucosal vaccination, *J. Controlled Release*, 114(3), 348–358.
46. Holtan, S., Zhang, Q., Strand, W. I., and Skjak-Braek, G. (2006), Characterization of the hydrolysis mechanism of polyalternating alginate in weak acid and assignment of the resulting MG-oligosaccharides by NMR spectroscopy and ESI-mass spectrometry, *Biomacromolecules*, 7(7), 2108–2121.
47. Chen, L., and Subirade, M. (2005), Chitosan/beta-lactoglobulin core-shell nanoparticles as nutraceutical carriers, *Biomaterials*, 26(30), 6041–6053.

48. Qi, L., Xu, Z., Jiang, X., Hu, C., and Zou, X. (2004), Preparation and antibacterial activity of chitosan nanoparticles, *Carbohydr. Res.*, 339(16), 2693–2700.
49. Liu, C. G., Desai, K. G., Chen, X. G., and Park, H. J. (2005), Preparation and characterization of nanoparticles containing trypsin based on hydrophobically modified chitosan, *J. Agric. Food Chem.*, 53(5), 1728–1733.
50. Grenha, A., Seijo, B., and Remunan-Lopez, C. (2005), Microencapsulated chitosan nanoparticles for lung protein delivery, *Eur. J. Pharm. Sci.*, 25(4–5), 427–437.
51. Enriquez de Salamanca, A., Diebold, Y., Calonge, M., Garcia-Vazquez, C., Callejo, S., Vila, A., and Alonso, M. J. (2006), Chitosan nanoparticles as a potential drug delivery system for the ocular surface: Toxicity, uptake mechanism and in vivo tolerance, *Invest. Ophthalmol. Vis. Sci.*, 47(4), 1416–1425.
52. Saito, H., Wu, X., Harris, M., and Hoffman, A. (2003), Graft copolymers of poly(ethylene glycol) (PEG) and chitosan, *Macromol. Rapid Commun.*, 18(7), 547–550.
53. Bernkop-Schnurch, A. (2000), Chitosan and its derivatives: Potential excipients for peroral peptide delivery systems, *Int. J. Pharm.*, 194(1), 1–13.
54. Hyung Park, J., Kwon, S., Lee, M., Chung, H., Kim, J. H., Kim, Y. S., Park, R. W., Kim, I. S., Bong Seo, S., Kwon, I. C., and Young Jeong, S. (2006), Self-assembled nanoparticles based on glycol chitosan bearing hydrophobic moieties as carriers for doxorubicin: In vivo biodistribution and anti-tumor activity, *Biomaterials*, 27(1), 119–126.
55. Mansouri, S., Cuie, Y., Winnik, F., Shi, Q., Lavigne, P., Benderdour, M., Beaumont, E., and Fernandes, J. C. (2006), Characterization of folate-chitosan-DNA nanoparticles for gene therapy, *Biomaterials*, 27(9), 2060–2065.
56. Yan, C., Chen, D., Gu, J., and Qin, J. (2006), Nanoparticles of 5-fluorouracil (5-FU) loaded *N*-succinyl-chitosan (Suc-Chi) for cancer chemotherapy: Preparation, characterization—in-vitro drug release and anti-tumour activity, *J. Pharm. Pharmacol.*, 58(9), 1177–1181.
57. Park, J. S., Han, T. H., Lee, K. Y., Han, S. S., Hwang, J. J., Moon, D. H., Kim, S. Y., and Cho, Y. W. (2006), *N*-acetyl histidine-conjugated glycol chitosan self-assembled nanoparticles for intracytoplasmic delivery of drugs: Endocytosis, exocytosis and drug release, *J. Controlled Release.*, 115(1), 37–45.
58. Vandervoort, J., and Ludwig, A. (2004), Preparation and evaluation of drug-loaded gelatin nanoparticles for topical ophthalmic use, *Eur. J. Pharm. Biopharm.*, 57(2), 251–261.
59. Cascone, M. G., Lazzeri, L., Carmignani, C., and Zhu, Z. (2002), Gelatin nanoparticles produced by a simple W/O emulsion as delivery system for methotrexate, *J. Mater. Sci. Mater. Med.*, 13(5), 523–526.
60. Lu, Z., Yeh, T. K., Tsai, M., Au, J. L., and Wientjes, M. G. (2004), Paclitaxel-loaded gelatin nanoparticles for intravesical bladder cancer therapy, *Clin. Cancer Res.*, 10(22), 7677–7684.
61. Bajpai, A. K., and Choubey, J. (2006), Design of gelatin nanoparticles as swelling controlled delivery system for chloroquine phosphate, *J. Mater. Sci. Mater. Med.*, 17(4), 345–358.
62. Li, J. K., Wang, N., and Wu, X. S. (1997), A novel biodegradable system based on gelatin nanoparticles and poly(lactic-*co*-glycolic acid) microspheres for protein and peptide drug delivery, *J. Pharm. Sci.*, 86(8), 891–895.
63. Farrugia, C. A., and Groves, M. J. (1999), The activity of unloaded gelatin nanoparticles on murine melanoma B16-F0 growth in vivo, *Anticancer Res.*, 19(2A), 1027–1031.
64. Coester, C., Kreuter, J., von Briesen, H., and Langer, K. (2000), Preparation of avidin-labelled gelatin nanoparticles as carriers for biotinylated peptide nucleic acid (PNA), *Int. J. Pharm.*, 196(2), 147–149.

65. Kaul, G., and Amiji, M. (2002), Long-circulating poly(ethylene glycol)-modified gelatin nanoparticles for intracellular delivery, *Pharm. Res.*, 19(7), 1061–1067.
66. Kommareddy, S., and Amiji, M. (2005), Preparation and evaluation of thiol-modified gelatin nanoparticles for intracellular DNA delivery in response to glutathione, *Bioconjug. Chem.*, 16(6), 1423–1432.
67. Coester, C. J., Langer, K., van Briesen, H., and Kreuter, J. (2000), Gelatin nanoparticles by two step desolvation—a new preparation method, surface modifications and cell uptake, *J. Microencapsul.*, 17(2), 187–193.
68. Kaul, G., and Amiji, M. (2004), Biodistribution and targeting potential of poly(ethylene glycol)-modified gelatin nanoparticles in subcutaneous murine tumor model, *J. Drug Target.*, 12(9–10), 585–591.
69. Coester, C., Nayyar, P., and Samuel, J. (2006), In vitro uptake of gelatin nanoparticles by murine dendritic cells and their intracellular localisation, *Eur. J. Pharm. Biopharm.*, 62(3), 306–314.
70. Leo, E., Arletti, R., Forni, F., and Cameroni, R. (1997), General and cardiac toxicity of doxorubicin-loaded gelatin nanoparticles, *Farmaco*, 52(6–7), 385–388.
71. Gu, X. G., Schmitt, M., Hiasa, A., Nagata, Y., Ikeda, H., Sasaki, Y., Akiyoshi, K., Sunamoto, J., Nakamura, H., Kuribayashi, K., and Shiku, H. (1998), A novel hydrophobized polysaccharide/oncoprotein complex vaccine induces in vitro and in vivo cellular and humoral immune responses against HER2-expressing murine sarcomas, *Cancer Res.*, 58(15), 3385–3390.
72. Shiku, H., Wang, L., Ikuta, Y., Okugawa, T., Schmitt, M., Gu, X., Akiyoshi, K., Sunamoto, J., and Nakamura, H. (2000), Development of a cancer vaccine: Peptides, proteins, and DNA, *Cancer Chemother. Pharmacol.*, 46 (Suppl), S77–82.
73. Na, K., and Bae, Y. H. (2002), Self-assembled hydrogel nanoparticles responsive to tumor extracellular pH from pullulan derivative/sulfonamide conjugate: Characterization, aggregation, and adriamycin release in vitro, *Pharm. Res.*, 19(5), 681–688.
74. Gupta, A. K., and Gupta, M. (2005), Cytotoxicity suppression and cellular uptake enhancement of surface modified magnetic nanoparticles, *Biomaterials*, 26(13), 1565–1573.
75. Tiyaboonchai, W., Woiszwillo, J., Sims, R. C., and Middaugh, C. R. (2003), Insulin containing polyethylenimine-dextran sulfate nanoparticles, *Int. J. Pharm.*, 255(1–2), 139–151.
76. Cascone, M. G., Pot, P. M., Lazzeri, L., and Zhu, Z. (2002), Release of dexamethasone from PLGA nanoparticles entrapped into dextran/poly(vinyl alcohol) hydrogels, *J. Mater. Sci. Mater. Med.*, 13(3), 265–269.
77. Jaulin, N., Appel, M., Passirani, C., Barratt, G., and Labarre, D. (2000), Reduction of the uptake by a macrophagic cell line of nanoparticles bearing heparin or dextran covalently bound to poly(methyl methacrylate), *J. Drug Target.*, 8(3), 165–172.
78. Segura, S., Gamazo, C., Irache, J. M., and Espuelas, S. (2007), Interferon-γ loaded onto albumin nanoparticles: In vitro and in vivo activity against brucella abortus, *Antimicrob. Agents Chemother.*, 51(4), 1310–1314.
79. Irache, J. M., Merodio, M., Arnedo, A., Camapanero, M. A., Mirshahi, M., and Espuelas, S. (2005), Albumin nanoparticles for the intravitreal delivery of anticytomegaloviral drugs, *Mini. Rev. Med. Chem.*, 5(3), 293–305.
80. Landry, F. B., Bazile, D. V., Spenlehauer, G., Veillard, M., and Kreuter, J. (1998), Peroral administration of 14C-poly(D,L-lactic acid) nanoparticles coated with human serum albumin or polyvinyl alcohol to guinea pigs, *J. Drug Target.*, 6(4), 293–307.
81. Damascelli, B., Patelli, G. L., Lanocita, R., Di Tolla, G., Frigerio, L. F., Marchiano, A., Garbagnati, F., Spreafico, C., Ticha, V., Gladin, C. R., Palazzi, M., Crippa, F., Oldini, C.,

Calo, S., Bonaccorsi, A., Mattavelli, F., Costa, L., Mariani, L., and Cantu, G. (2003), A novel intraarterial chemotherapy using paclitaxel in albumin nanoparticles to treat advanced squamous cell carcinoma of the tongue: Preliminary findings, *Am. J. Roentgenol.*, 181(1), 253–260.

82. Rhaese, S., von Briesen, H., Rubsamen-Waigmann, H., Kreuter, J., and Langer, K. (2003), Human serum albumin-polyethylenimine nanoparticles for gene delivery, *J. Controlled Release*, 92(1–2), 199–208.
83. Zhang, L. K., Hou, S. X., Mao, S. J., Wei, D. P., Song, X. R., and Qiao, X. R. (2004), [Study on the tumor cell targetability of folate-conjugated albumin nanoparticles] *Sichuan Da Xue Xue Bao Yi Xue Ban*, 35(2), 165–168.
84. Zhang, L., Hou, S., Mao, S., Wei, D., Song, X., and Lu, Y. (2004), Uptake of folate-conjugated albumin nanoparticles to the SKOV3 cells, *Int. J. Pharm.*, 287(1–2), 155–162.
85. Mishra, V., Mahor, S., Rawat, A., Gupta, P. N., Dubey, P., Khatri, K., and Vyas, S. P. (2006), Targeted brain delivery of AZT via transferrin anchored pegylated albumin nanoparticles, *J. Drug Target.*, 14(1), 45–53.
86. Mao, S. J., Hou, S. X., Zhang, L. K., Jin, H., Bi, Y. Q., and Jiang, B. (2003), [Preparation of bovine serum albumin nanoparticles surface-modified with glycyrrhizin], *Yao Xue Xue Bao*, 38(10), 787–790.
87. Bazile, D. V., Ropert, C., Huve, P., Verrecchia, T., Marlard, M., Frydman, A., Veillard, M., and Spenlehauer, G. (1992), Body distribution of fully biodegradable [$^{14}$C]-poly(lactic acid) nanoparticles coated with albumin after parenteral administration to rats, *Biomaterials*, 13(15), 1093–1102.
88. Arangoa, M. A., Campanero, M. A., Renedo, M. J., Ponchel, G., and Irache, J. M. (2001), Gliadin nanoparticles as carriers for the oral administration of lipophilic drugs. Relationships between bioadhesion and pharmacokinetics, *Pharm. Res.*, 18(11), 1521–1527.
89. Duclairoir, C., Orecchioni, A. M., Depraetere, P., Osterstock, F., and Nakache, E. (2003), Evaluation of gliadins nanoparticles as drug delivery systems: A study of three different drugs, *Int. J. Pharm.*, 253(1–2), 133–144.
90. Umamaheshwari, R. B., and Jain, N. K. (2003), Receptor mediated targeting of lectin conjugated gliadin nanoparticles in the treatment of *Helicobacter pylori*, *J. Drug Target.*, 11(7), 415–423; discussion 423–424.
91. Jain, R. A. (2000), The manufacturing techniques of various drug loaded biodegradable poly(lactide-*co*-glycolide) (PLGA) devices, *Biomaterials*, 21(23), 2475–2490.
92. Bala, I., Hariharan, S., and Kumar, M. N. (2004), PLGA nanoparticles in drug delivery: The state of the art, *Crit. Rev. Ther. Drug Carrier Syst.*, 21(5), 387–422.
93. Astete, C. E., and Sabliov, C. M. (2006), Synthesis and characterization of PLGA nanoparticles, *J. Biomater. Sci. Polym. Ed.*, 17(3), 247–289.
94. Sinha, V. R., Bansal, K., Kaushik, R., Kumria, R., and Trehan, A. (2004), Poly-epsilon-caprolactone microspheres and nanospheres: An overview, *Int. J. Pharm.*, 278(1), 1–23.
95. Seppala, J. V., Helminen, A. O., and Korhonen, H. (2004), Degradable polyesters through chain linking for packaging and biomedical applications, *Macromol. Biosci.*, 4(3), 208–217.
96. Kumar, N., Langer, R. S., and Domb, A. J. (2002), Polyanhydrides: An overview, *Adv. Drug Deliv. Rev.*, 54(7), 889–910.
97. Jain, J. P., Modi, S., Domb, A. J., and Kumar, N. (2005), Role of polyanhydrides as localized drug carriers, *J. Controlled Release.*, 103(3), 541–563.
98. Pfeifer, B. A., Burdick, J. A., and Langer, R. (2005), Formulation and surface modification of poly(ester-anhydride) micro- and nanospheres, *Biomaterials*, 26(2), 117–124.

99. Carino, G. P., Jacob, J. S., and Mathiowitz, E. (2000), Nanosphere based oral insulin delivery, *J. Controlled Release*, 65(1–2), 261–269.
100. Torres, M. P., Determan, A. S., and Anderson, G. L., Mallapragada, S. K., and Narasimhan, B. (2007), Amphiphilic polyanhydrides for protein stabilization and release, *Biomaterials*, 28(1), 108–116.
101. Gao, J., Niklason, L., Zhao, X. M., and Langer, R. (1998), Surface modification of polyanhydride microspheres, *J. Pharm. Sci.*, 87(2), 246–248.
102. Vauthier, C., Dubernet, C., Fattal, E., Pinto-Alphandary, H., and Couvreur, P. (2003), Poly (alkylcyanoacrylates) as biodegradable materials for biomedical applications, *Adv. Drug Deliv. Rev.*, 55(4), 519–548.
103. Mehnert, W., and Mader, K. (2001), Solid lipid nanoparticles: Production, characterization and applications, *Adv. Drug Deliv. Rev.*, 47(2–3), 165–196.
104. Uner, M. (2006), Preparation, characterization and physico-chemical properties of solid lipid nanoparticles (SLN) and nanostructured lipid carriers (NLC): Their benefits as colloidal drug carrier systems, *Pharmazie*, 61(5), 375–386.
105. Heller, J., Barr, J., Ng, S. Y., Abdellauoi, K. S., and Gurny, R. (2002), Poly(ortho esters): Synthesis, characterization, properties and uses, *Adv. Drug Deliv. Rev.*, 54(7), 1015–1039.
106. Barichello, J. M., Morishita, M., Takayama, K., and Nagai, T. (1999), Encapsulation of hydrophilic and lipophilic drugs in PLGA nanoparticles by the nanoprecipitation method, *Drug Dev. Ind. Pharm.*, 25(4), 471–476.
107. Panyam, J., and Labhasetwar, V. (2003), Biodegradable nanoparticles for drug and gene delivery to cells and tissue, *Adv. Drug Deliv. Rev.*, 55(3), 329–347.
108. Avgoustakis, K. (2004), Pegylated poly(lactide) and poly(lactide-*co*-glycolide) nanoparticles: Preparation, properties and possible applications in drug delivery, *Curr. Drug Deliv.*, 1(4), 321–333.
109. Panagi, Z., Beletsi, A., Evangelatos, G., Livaniou, E., Ithakissios, D. S., and Avgoustakis, K. (2001), Effect of dose on the biodistribution and pharmacokinetics of PLGA and PLGA-mPEG nanoparticles, *Int. J. Pharm.*, 221(1–2), 143–152.
110. Qaddoumi, M. G., Gukasyan, H. J., Davda, J., Labhasetwar, V., Kim, K. J., and Lee, V. H. (2003), Clathrin and caveolin-1 expression in primary pigmented rabbit conjunctival epithelial cells: Role in PLGA nanoparticle endocytosis, *Mol. Vis.*, 9, 559–568.
111. Panyam, J., Zhou, W. Z., Prabha, S., Sahoo, S. K., and Labhasetwar, V. (2002), Rapid endo-lysosomal escape of poly(DL-lactide-*co*-glycolide) nanoparticles: Implications for drug and gene delivery, *FASEB J.*, 16(10), 1217–1226.
112. Shive, M. S., and Anderson, J. M. (1997), Biodegradation and biocompatibility of PLA and PLGA microspheres, *Adv. Drug Deliv. Rev.*, 28(1), 5–24.
113. Marchal-Heussler, L., Sirbat, D., Hoffman, M., and Maincent, P. (1993), Poly(epsilon-caprolactone) nanocapsules in carteolol ophthalmic delivery, *Pharm. Res.*, 10(3), 386–390.
114. Calvo, P., Vila-Jato, J. L., and Alonso, M. J. (1996), Comparative in vitro evaluation of several colloidal systems, nanoparticles, nanocapsules, and nanoemulsions, as ocular drug carriers, *J. Pharm. Sci.*, 85(5), 530–536.
115. Jeong, Y. I., Kang, M. K., Sun, H. S., Kang, S. S., Kim, H. W., Moon, K. S., Lee, K. J., Kim, S. H., and Jung, S. (2004), All-trans-retinoic acid release from core-shell type nanoparticles of poly(epsilon-caprolactone)/poly(ethylene glycol) diblock copolymer, *Int. J. Pharm.*, 273(1–2), 95–107.
116. Park, E. K., Lee, S. B., and Lee, Y. M. (2005), Preparation and characterization of methoxy poly(ethylene glycol)/poly(epsilon-caprolactone) amphiphilic block

copolymeric nanospheres for tumor-specific folate-mediated targeting of anticancer drugs, *Biomaterials*, 26(9), 1053–1061.

117. Cade, D., Ramus, E., Rinaudo, M., Auzely-Velty, R., Delair, T., and Hamaide, T. (2004), Tailoring of bioresorbable polymers for elaboration of sugar-functionalized nanoparticles, *Biomacromolecules*, 5(3), 922–927.
118. Haas, J., Ravi Kumar, M. N., Borchard, G., Bakowsky, U., and Lehr, C. M. (2005), Preparation and characterization of chitosan and trimethyl-chitosan-modified poly-(epsilon-caprolactone) nanoparticles as DNA carriers, *AAPS Pharm. Sci. Tech.*, 6(1), E22–30.
119. Lemarchand, C., Gref, R., Passirani, C., Garcion, E., Petri, B., Muller, R., Costantini, D., and Couvreur, P. (2006), Influence of polysaccharide coating on the interactions of nanoparticles with biological systems, *Biomaterials*, 27(1), 108–118.
120. Pfeifer, B. A., Burdick, J. A., Little, S. R., and Langer, R. (2005), Poly(ester-anhydride): poly(beta-amino ester) micro- and nanospheres: DNA encapsulation and cellular transfection, *Int. J. Pharm.*, 304(1–2), 210–219.
121. McCarron, P. A., Donnelly, R. F., Canning, P. E., McGovern, J. G., and Jones, D. S. (2004), Bioadhesive, non-drug-loaded nanoparticles as modulators of candidal adherence to buccal epithelial cells: A potentially novel prophylaxis for candidosis, *Biomaterials*, 25(12), 2399–2407.
122. Vauthier, C., Dubernet, C., Chauvierre, C., Brigger, I., and Couvreur, P. (2003), Drug delivery to resistant tumors: The potential of poly(alkyl cyanoacrylate) nanoparticles, *J. Controlled Release*, 93(2), 151–160.
123. Toub, N., Angiari, C., Eboue, D., Fattal, E., Tenu, J. P., Le Doan, T., and Couvreur, P. (2005), Cellular fate of oligonucleotides when delivered by nanocapsules of poly(isobu tylcyanoacrylate), *J. Controlled Release*, 106(1–2), 209–213.
124. Labarre, D., Vauthier, C., Chauvierre, C., Petri, B., Muller, R., and Chehimi, M. M. (2005), Interactions of blood proteins with poly(isobutylcyanoacrylate) nanoparticles decorated with a polysaccharidic brush, *Biomaterials*, 26(24), 5075–5084.
125. Wissing, S. A., Kayser, O., and Muller, R. H. (2004), Solid lipid nanoparticles for parenteral drug delivery, *Adv. Drug Deliv. Rev.*, 56(9), 1257–1272.
126. Manjunath, K., Reddy, J. S., and Venkateswarlu, V. (2005), Solid lipid nanoparticles as drug delivery systems, *Methods Find. Exp. Clin. Pharmacol.*, 27(2), 127–144.
127. Zhang, N., Ping, Q., Huang, G., Xu, W., Cheng, Y., and Han, X. (2006), Lectin-modified solid lipid nanoparticles as carriers for oral administration of insulin, *Int. J. Pharm.*, 327(1–2), 153–159.
128. Wang, Y., and Wu, W. (2006), In situ evading of phagocytic uptake of stealth solid lipid nanoparticles by mouse peritoneal macrophages, *Drug Deliv.*, 13(3), 189–192.
129. Chilkoti, A., Dreher, M. R., Meyer, D. E., and Raucher, D. (2002), Targeted drug delivery by thermally responsive polymers, *Adv. Drug Deliv. Rev.*, 54(5), 613–630.
130. Sakuma, S., Suzuki, N., Sudo, R., Hiwatari, K., Kishida, A., and Akashi, M. (2002), Optimized chemical structure of nanoparticles as carriers for oral delivery of salmon calcitonin, *Int. J. Pharm.*, 239(1–2), 185–195.
131. Piskin, E. (2004), Molecularly designed water soluble, intelligent, nanosize polymeric carriers, *Int. J. Pharm.*, 277(1–2), 105–118.
132. Neradovic, D., Soga, O., Van Nostrum, C. F., and Hennink, W. E. (2004), The effect of the processing and formulation parameters on the size of nanoparticles based on block copolymers of poly(ethylene glycol) and poly(*N*-isopropylacrylamide) with and without hydrolytically sensitive groups, *Biomaterials*, 25(12), 2409–2418.

133. Na, K., Lee, K. H., and Bae, Y. H. (2004), pH-sensitivity and pH-dependent interior structural change of self-assembled hydrogel nanoparticles of pullulan acetate/oligo-sulfonamide conjugate, *J. Controlled Release*, 97(3), 513–525.
134. Ma, Z., Yeoh, H. H., and Lim, L. Y. (2002), Formulation pH modulates the interaction of insulin with chitosan nanoparticles, *J. Pharm. Sci.*, 91(6), 1396–1404.
135. Lynn, D. M., Amiji, M. M., and Langer, R. (2001), pH-responsive polymer microspheres: Rapid release of encapsulated material within the range of intracellular pH, *Angew Chem. Int. Ed. Engl.*, 40(9), 1707–1710.
136. Potineni, A., Lynn, D. M., Langer, R., and Amiji, M. M. (2003), Poly(ethylene oxide)-modified poly(beta-amino ester) nanoparticles as a pH-sensitive biodegradable system for paclitaxel delivery, *J. Controlled Release*, 86(2–3), 223–234.
137. Heffernan, M. J., and Murthy, N. (2005), Polyketal nanoparticles: A new pH-sensitive biodegradable drug delivery vehicle, *Bioconjug. Chem.*, 16(6), 1340–1342.
138. Choi, J. S., MacKay, J. A., and Szoka, F. C. Jr. (2003), Low-pH-sensitive PEG-stabilized plasmid-lipid nanoparticles: Preparation and characterization, *Bioconjug. Chem.*, 14(2), 420–429.
139. Na, K., Lee, K. H., Lee, D. H., and Bae, Y. H. (2006), Biodegradable thermo-sensitive nanoparticles from poly(L-lactic acid)/poly(ethylene glycol) alternating multi-block copolymer for potential anti-cancer drug carrier, *Eur. J. Pharm. Sci.*, 27(2–3), 115–122.
140. Sun, S., Liu, W., Cheng, N., Zhang, B., Cao, Z., Yao, K., Liang, D., Zuo, A., Guo, G., and Zhang, J. (2005), A thermoresponsive chitosan-NIPAAm/vinyl laurate copolymer vector for gene transfection, *Bioconjug. Chem.*, 16(4), 972–980.
141. Weng, H., Zhou, J., Tang, L., and Hu, Z. (2004), Tissue responses to thermally-responsive hydrogel nanoparticles, *J. Biomater. Sci. Polym. Ed.*, 15(9), 1167–1180.
142. Devalapally, H., Shenoy, D., Little, S., Langer, R., and Amiji, M. (2007), Poly(ethylene oxide)-modified poly(beta-amino ester) nanoparticles as a pH-sensitive system for tumor-targeted delivery of hydrophobic drugs: Part 3. Therapeutic efficacy and safety studies in ovarian cancer xenograft model, *Cancer Chemother. Pharmacol.*, 59(4), 477–484.
143. Brigger, I., Dubernet, C., and Couvreur, P. (2002), Nanoparticles in cancer therapy and diagnosis, *Adv. Drug Deliv. Rev.*, 54(5), 631–651.
144. Mulder, W. J., Strijkers, G. J., van Tilborg, G. A., Griffioen, A. W., and Nicolay, K. (2006), Lipid-based nanoparticles for contrast-enhanced MRI and molecular imaging, *NMR Biomed.*, 19(1), 142–164.
145. De, S., Miller, D. W., and Robinson, D. H. (2005), Effect of particle size of nanospheres and microspheres on the cellular-association and cytotoxicity of paclitaxel in 4T1 cells, *Pharm. Res.*, 22(5), 766–775.
146. De, S., and Robinson, D. H. (2004), Particle size and temperature effect on the physical stability of PLGA nanospheres and microspheres containing Bodipy, *AAPS Pharm. Sci. Tech.*, 5(4), e53.
147. Alonso, M. J. (1996), Nanoparticulate drug carrier technology, Cohen, S., and Bennstein, H. Eds. *Microparticulate Systems for the Delivery of Proteins and Vaccines*, Marcel Dekkes, New York, pp. 203–242.
148. Torres, M. P., Vogel, B. M., Narasimhan, B., and Mallapragada, S. K. (2006), Synthesis and characterization of novel polyanhydrides with tailored erosion mechanisms, *J. Biomed. Mater. Res. A.*, 76(1), 102–110.
149. Shen, E., Pizsczek, R., Dziadul, B., and Narasimhan, B. (2001), Microphase separation in bioerodible copolymers for drug delivery, *Biomaterials*, 22(3), 201–210.

150. DiBenedetto, A. K. (2003), Prediction of the glass transition temperature of polymers: A model based on the principle of corresponding states, *J. Polym. Sci. Part B: Polym. Phys.*, 25(9), 1949–1969.

151. Dong, Y., and Feng, S. S. (2006), Nanoparticles of poly(D,L-lactide)/methoxy poly(ethylene glycol)-poly(D,L-lactide) blends for controlled release of paclitaxel, *J. Biomed. Mater. Res. A*, 78(1), 12–19.

152. Zweers, M. L., Grijpma, D. W., Engbers, G. H., and Feijen, J. (2003), The preparation of monodisperse biodegradable polyester nanoparticles with a controlled size, *J. Biomed. Mater. Res. B Appl. Biomater.*, 66(2), 559–566.

153. Kirby, B. J., and Hasselbrink, E. F. Jr. (2004), Zeta potential of microfluidic substrates: 2. Data for polymers, *Electrophoresis*, 25(2), 203–213.

154. Gessner, A., Lieske, A., Paulke, B., and Muller, R. (2002), Influence of surface charge density on protein adsorption on polymeric nanoparticles: Analysis by two-dimensional electrophoresis, *Eur. J. Pharm. Biopharm.*, 54(2), 165–170.

155. Kumar, M. N., Mohapatra, S. S., Kong, X., Jena, P. K., Bakowsky, U., and Lehr, C. M. (2004), Cationic poly(lactide-*co*-glycolide) nanoparticles as efficient in vivo gene transfection agents, *J. Nanosci. Nanotechnol.*, 4(8), 990–994.

156. Reddy, L. H., Sharma, R. K., Chuttani, K., Mishra, A. K., and Murthy, R. R. (2004), Etoposide-incorporated tripalmitin nanoparticles with different surface charge: Formulation, characterization, radiolabeling, and biodistribution studies, *AAPS J.*, 6(3), e23.

157. Owens, D. E., 3rd, and Peppas, N. A. (2006), Opsonization, biodistribution, and pharmacokinetics of polymeric nanoparticles, *Int. J. Pharm.*, 307(1), 93–102.

158. Moghimi, S. M., and Szebeni, J. (2003), Stealth liposomes and long circulating nanoparticles: Critical issues in pharmacokinetics, opsonization and protein-binding properties, *Prog. Lipid Res.*, 42(6), 463–478.

159. Freiberg, S., and Zhu, X. X. (2004), Polymer microspheres for controlled drug release, *Int. J. Pharm.*, 282(1–2), 1–18.

160. Siepmann, J., and Gopferich, A. (2001), Mathematical modeling of bioerodible, polymeric drug delivery systems, *Adv. Drug Deliv. Rev.*, 48(2–3), 229–247.

161. Grassi, M., and Grassi, G. (2005), Mathematical modelling and controlled drug delivery: Matrix systems, *Curr. Drug Deliv.*, 2(1), 97–116.

162. Kanjickal, D. G., and Lopina, S. T. (2004), Modeling of drug release from polymeric delivery systems—a review, *Crit. Rev. Ther. Drug Carrier Syst.*, 21(5), 345–386.

163. Fu, K., Harrell, R., Zinski, K., Um, C., Jaklenec, A., Frazier, J., Lotan, N., Burke, P., Klibanov, A. M., and Langer, R. (2003), A potential approach for decreasing the burst effect of protein from PLGA microspheres, *J. Pharm. Sci.*, 92(8), 1582–1591.

164. Dhoot, N. O., and Wheatley, M. A. (2003), Microencapsulated liposomes in controlled drug delivery: Strategies to modulate drug release and eliminate the burst effect, *J. Pharm. Sci.*, 92(3), 679–689.

165. Lemoine, D., Francois, C., Kedzierewicz, F., Preat, V., Hoffman, M., and Maincent, P. (1996), Stability study of nanoparticles of poly(epsilon-caprolactone), poly(D,L-lactide) and poly(D,L-lactide-*co*-glycolide), *Biomaterials*, 17(22), 2191–2197.

166. Molpeceres, J., Aberturas, M. R., Chacon, M., Berges, L., and Guzman, M. (1997), Stability of cyclosporine-loaded poly-sigma-caprolactone nanoparticles, *J. Microencapsul.*, 14(6), 777–787.

167. Freitas, C., and Muller, R. H. (1999), Correlation between long-term stability of solid lipid nanoparticles (SLN) and crystallinity of the lipid phase, *Eur. J. Pharm. Biopharm.*, 47(2), 125–132.

168. Freitas, C., and Muller, R. H. (1999), Stability determination of solid lipid nanoparticles (SLN) in aqueous dispersion after addition of electrolyte, *J. Microencapsul.*, 16(1), 59–71.

169. Santander-Ortega, M. J., Jodar-Reyes, A. B., Csaba, N., Bastos-Gonzalez, D., and Ortega-Vinuesa, J. L. (2006), Colloidal stability of Pluronic F68-coated PLGA nanoparticles: A variety of stabilisation mechanisms, *J. Colloid Interface Sci.*, 302(2), 522–529.

170. Zobel, H. P., Werner, D., Gilbert, M., Noe, C. R., Stieneker, F., Kreuter, J., and Zimmer, A. (1999), Effect of ultrasonication on the stability of oligonucleotides adsorbed on nanoparticles and liposomes, *J. Microencapsul.*, 16(4), 501–509.

171. Hattori, Y., Sakaguchi, M., and Maitani, Y. (2006), Folate-linked lipid-based nanoparticles deliver a NFkappaB decoy into activated murine macrophage-like RAW264.7 cells, *Biol. Pharm. Bull.*, 29(7), 1516–1520.

172. Sahoo, S. K., Ma, W., and Labhasetwar, V. (2004), Efficacy of transferrin-conjugated paclitaxel-loaded nanoparticles in a murine model of prostate cancer, *Int. J. Cancer*, 112(2), 335–340.

173. Gao, X., Tao, W., Lu, W., Zhang, Q., Zhang, Y., Jiang, X., and Fu, S. (2006), Lectin-conjugated PEG-PLA nanoparticles: Preparation and brain delivery after intranasal administration, *Biomaterials*, 27(18), 3482–3490.

174. Sokolov, K., Follen, M., Aaron, J., Pavlova, I., Malpica, A., Lotan, R., and Richards-Kortum, R. (2003), Real-time vital optical imaging of precancer using anti-epidermal growth factor receptor antibodies conjugated to gold nanoparticles, *Cancer Res.*, 63(9), 1999–2004.

175. Garnett, M. C. (2001), Targeted drug conjugates: Principles and progress, *Adv. Drug Deliv. Rev.*, 53(2), 171–216.

176. Nobs, L., Buchegger, F., Gurny, R., and Allemann, E. (2004), Current methods for attaching targeting ligands to liposomes and nanoparticles, *J. Pharm. Sci.*, 93(8), 1980–1992.

177. Kreuter, J. (1994), Drug targeting with nanoparticles, *Eur. J. Drug Metab. Pharmacokinet.*, 19(3), 253–256.

178. Jain, K. K. (2006), Nanoparticles as targeting ligands, *Trends Biotechnol.*, 24(4), 143–145.

179. Conner, S. D., and Schmid, S. L. (2003), Regulated portals of entry into the cell, *Nature*, 422(6927), 37–44.

180. Huang, M., Ma, Z., Khor, E., and Lim, L. Y. (2002), Uptake of FITC-chitosan nanoparticles by A549 cells, *Pharm. Res.*, 19(10), 1488–1494.

181. Ma, Z., and Lim, L. Y. (2003), Uptake of chitosan and associated insulin in Caco-2 cell monolayers: A comparison between chitosan molecules and chitosan nanoparticles, *Pharm. Res.*, 20(11), 1812–1819.

182. Liu, J., and Shapiro, J. I. (2003), Endocytosis and signal transduction: Basic science update, *Biol. Res. Nurs.*, 5(2), 117–128.

183. Steinman, R. M., Mellman, I. S., Muller, W. A., and Cohn, Z. A. (1983), Endocytosis and the recycling of plasma membrane, *J. Cell Biol.*, 96(1), 1–27.

184. Okamoto, C. T. (1998), Endocytosis and transcytosis, *Adv. Drug Deliv. Rev.*, 29(3), 215–228.

185. Hilgenbrink, A. R., and Low, P. S. (2005), Folate receptor-mediated drug targeting: From therapeutics to diagnostics, *J. Pharm. Sci.*, 94(10), 2135–2146.

186. Park, I. K., Seo, S. J., Akashi, M., Akaike, T., and Cho, C. S. (2003), Controlled release of epidermal growth factor (EGF) from EGF-loaded polymeric nanoparticles composed

of polystyrene as core and poly(methacrylic acid) as corona in vitro, *Arch. Pharm. Res.*, 26(8), 649–652.

187. Douglas, S. J., Davis, S. S., and Illum, L. (1987), Nanoparticles in drug delivery, *Crit. Rev. Ther. Drug Carrier Syst.*, 3(3), 233–261.

# RECOMBINANT *SACCHAROMYCES CEREVISIAE* AS NEW DRUG DELIVERY SYSTEM TO GUT: IN VITRO VALIDATION AND ORAL FORMULATION

Stéphanie Blanquet and Monique Alric
*Université d'Auvergne, Clermont-Ferrand, France*

**Contents**

1 Enginereed Microorganisms as Delivery Vectors to Human Gastrointestinal Tract
  1.1 What is the "Biodrug" Concept?
  1.2 Medical Applications
  1.3 Choice of Candidate Host Microorganisms
2 Evaluation of Scientific Feasibility of Biodrug Concept Using Yeast as Vector
  2.1 Approach
  2.2 Yeast Survival Rate in Simulated Gastrointestinal Conditions
  2.3 Yeast Heterologous Activity in Simulated Gastrointestinal Conditions
  2.4 Conclusion
3 Oral Formulation of Recombinant Yeasts
  3.1 Freeze Drying of Recombinant Model Yeasts
  3.2 Immobilization of Recombinant Model Yeasts in Whey Protein Beads
4 General Conclusion and Future Developments
  References

## 1 ENGINEREED MICROORGANISMS AS DELIVERY VECTORS TO HUMAN GASTROINTESTINAL TRACT

### 1.1 What Is the "Biodrug" Concept?

The development of recombinant deoxyribonucleic acid (DNA) technology has allowed the emergence of novel applications such as drug production directly in the

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

human digestive environment ("in situ") by ingested living recombinant microorganisms [1–3].

This new kind of vector offers several advantages over classical dosage forms. First, the microorganisms, by protecting the active compounds, can allow the administration of drugs known to be sensitive to digestive conditions when given in classical pharmaceutical formulations. Second, the regulation of gene expression (e.g., using an inducible promoter) makes it possible to target specific sites throughout the digestive tract (i.e., the absorption or reaction site of the drug) and to control drug release. Thus, similar therapeutic effects can potentially be obtained at lower doses [4, 5] and the degradation of the active compound by acid or proteases should be avoided upstream from its absorption or reaction site.

In the digestive tract, genetically modified microorganisms can either carry out a reaction of bioconversion or produce compounds of interest. The bioconversion reaction can lead either to the production of an active product or to the removal of undesirable compounds. The active compound produced in situ can be secreted in the digestive medium [5], be bound to the cells [6, 7], or accumulate inside the cells and be released in the digestive environment by cell lysis [8].

## 1.2 Medical Applications

The biodrug concept involves the use of orally administered recombinant microorganisms as a new drug delivery route to prevent or treat various diseases. The potential medical applications are numerous and can be classified in terms of bioconversion or production of active compounds. Validation studies have yet been conducted in animals (and even sometimes in human being), as described below and summarized in Table 1.

***Bioconversion*** Three main types of medical applications have been considered. First, recombinant microorganisms could be administered to perform "biodetoxication" in the gut [9]. Here, the objective is to increase the body's protection against environmental xeniobiotics, mainly those ingested with food (e.g., pesticides, procarcinogens, or chemical additives), by ingesting microorganisms expressing enzymes that play a major role in the human detoxication system [e.g., phase I xenobiotic metabolizing enzymes such as cytochrome P450 or phase II–like glutathione *S*-transferase (GST)]. Therefore, recombinant microorganisms could be used to prevent multifactorial diseases that have been associated with anomalies in human detoxication processes. For instance, a deficiency in GST-M1 has been correlated with an increased susceptibility to different cancers, endometriosis, and chronic bronchitis [10].

Second, modified microorganisms could correct errors of metabolism resulting from either gastric or intestinal enzyme deficiencies (e.g., lipase or lactase) [11] or organ failure (by removing urea in the case of kidney failure or ammonia in the case of liver failure) [12, 13]. This could constitute an alternative to current therapy such as renal dialysis, which is time consuming and uncomfortable for the patient.

The third potential application is the use of recombinant cells to control the activation of prodrug into drug directly in the digestive tract. This is of interest when the drug, but not the prodrug, is either toxic at high concentrations or damaged by digestive secretions [1].

**TABLE 1 Potential Medical Applications of Biodrug Concept and Their Validation in Animals**

| Biodrug Concept | Applications | Examples | Validation: Recombinant Microorganisms (Heterologous Gene) | Validation: Experimental Models | Validation: Effect |
|---|---|---|---|---|---|
| Bioconversion | Biodetoxication | Removal of benzo(*a*)pyrene [9] | *Escherichia coli* (P450 1A1 and glutathione *S*-transferase) | Stomach and duodenum of a mouse combined with mutagenesis assay (Ames assay) | Decrease of the mutagenic potential of B(a)P |
| | Correction of errors of metabolism | Correction of lipase deficiency [11] | *Lactococcus lactis* (lipase) | Pig with pancreatic deficiency, oral administration | Increase of lipid digestibility |
| | | Correction of urease deficiency [13] | *Escherichia coli* DH5 (urease) | Rat with renal failure, oral administration | Decrease of plasma uric acid |
| In situ production of active compounds | Synthesis of biological mediators | Secretion of Interleukine [5] | *Lactococcus lactis* (IL-10) | Mouse treated with dextran sulfate sodium, oral administration | Reduction of colitis symptoms |
| | Oral vaccines | Vaccination against bacteria [17] | *Lactobacillus plantarum* (*Helicobacter pylori* urease B subunit) | Mouse, intragastric immunization | Induction of immune response (anti-UreB Ig), partial protection against *H. felis* |
| | | Vaccination against virus [19] | Attenuated *Salmonella typhimurium* (SIV capsid antigen p27) | Macaques, oral administration | Induction of immune response (anti-p27 Ig) |
| | | Control of food allergy [20] | *Lactococcus lactis* (bovine β-lactoglobulin) | Model of mouse allergy, oral administration | Induction of BLG-specific Th1 response |

***In Situ Production of Active Compounds*** The first (and main) medical application derived from the biodrug concept is the development of oral vaccines [14]. In that case, the microorganisms will locally deliver the antigens to the digestive mucosa in order to stimulate an immune response (production of immunoglobulins) and ensure a protection against bacterial [7, 15–17] or viral [6, 18, 19] diseases or being used in the management of food allergy [4, 20]. For several immunological and practical reasons, these new vaccines represent a promising alternative to the traditional injectable ones [21]. In particular, oral immunization is the most efficient way to induce a protective local immune response at the site of pathogen contact. Recently, clinical trials have shown the vaccinal efficiency of different recombinant strains of attenuated *Salmonella typhi* in humans, but the patients were presenting undesirable side effects (diarrhea) [22].

Other medical applications involve the direct production in the digestive medium of various biological mediators, such as insulin, cytokines [23], or growth factors. In particular, new approaches to treating inflammatory bowel diseases (IBD) such as Crohn's disease, celiac disease, or ulcerative colitis have been considered [24]. Anti-inflammatory and immunosuppressive therapies are commonly used for the treatment of IBD. However, patients are often subject to unpleasant side effects owing to the high level of the drug in the body (systemic administration) and some of them remain refractory to such treatments. Steidler et al. [5] have investigated the potential of alternative therapeutics and shown the interest of delivering interleukin-10 (IL-10), a strong anti-inflammatory cytokine, in a localized manner by the action of recombinant *Lactococcus lactis*. Oral administration of the strain producing mouse IL-10 led to 50% curing of dextran sulfate sodium–induced colitis and prevented the onset of the pathology in mice [5]. The amount of IL-10 required to achieve the healing effect was 10,000-fold lower when the cytokine was in situ delivered by *L. lactis* compared to systemic treatment with anti-inflammatory drugs (e.g., dexamethasone). A recent study of the same author has reported that the treatment of Crohn's disease patients with *L. lactis* secreting human IL-10 was safe and allowed a decrease in disease activity [25].

## 1.3 Choice of Candidate Host Microorganisms

Recombinant bacteria, particularly lactic acid bacteria, have been mostly suggested as potential hosts for this new kind of drug delivery system [2, 3]. However, yeasts can offer advantages, especially when a eukaryotic environment is required for the functional expression of human genes. Moreover, the absence of bacterial sequences liable to promote gene transfer to host bacteria can be ensured using the efficient site-targeted homologous recombination machinery of yeasts for introduction of the heterologous gene into the yeast genome. Lastly, yeasts are not sensitive to antibacterial agents, allowing the simultaneous administration of the recombinant microorganisms and antibiotics. In this study, the common baker's yeast *Saccharomyces cerevisiae* was chosen owing to its "generally recognized as safe" (GRAS) status and its easy culture and genetic engineering. *Saccharomyces* spp. have already been used in humans, mainly in the treatment of intestinal functional disorders such as colitis [26] or antibiotic-associated diarrhea [27].

## 2 EVALUATION OF SCIENTIFIC FEASIBILITY OF BIODRUG CONCEPT USING YEAST AS VECTOR

### 2.1 Approach

The scientific feasibility of our approach was recently evaluated using recombinant *S. cerevisiae* expressing model genes and an original artificial digestive system simulating human gastrointestinal conditions. The survival rate and heterologous activity of the recombinant model yeasts were followed in this in vitro system.

***Recombinant Model Yeasts*** Three different yeast strains, all derived from the haploid strain W303-1B, were used to evaluate the scientific feasibility of the biodrug concept using yeast as vector.

The first one, WRP45073A1 (provided by Denis Pompon, CNRS, Gif-sur-Yvette, France), expresses the plant P45073A1 when grown in the presence of galactose [28]. P45073A1 was chosen as a model for a reaction catalyzed by a P450 owing to the nontoxicity and easy quantification of both substrate and product [29]. It catalyzes the 4-hydroxylation of *trans*-cinnamic acid into *p*-coumaric acid [cinnamate 4-hydroxylase (CA4H) activity]. In this model, the heterologous protein is an intracellular enzyme, and it was synthesized (induction of CYP73A1 by galactose during the last 12 h of culture) before yeast introduction into the artificial digestive system. The in situ CA4H activity of yeasts was quantified following the simultaneous introduction of recombinant yeasts and *trans*-cinnamic acid into the in vitro system.

The two other strains—WppGST$V_5H_6$ and Wpp$V_5H_6$—were genetically engineered [30] to secrete (i) a model protein derived from the commonly used reporter protein GST, named GST–$V_5H_6$ [molecular weight (MW) 31.5 kDa], and (ii) a model peptide, peptide–$V_5H_6$ (MW 5.6 kDa). The recombinant protein and peptide were expressed in fusion with the $V_5$ epitope ($V_5$) and the polyhistidine ($H_6$) tag to allow their immunological detection and make easier their purification, respectively. A leader sequence derived from that of the α-factor precursor was used to direct the secretion of the heterologous protein compounds into the extracellular medium [30]. In that case, recombinant yeasts and galactose, the inductor of the heterologous genes, were simultaneously introduced into the artificial digestive system to evaluate the yeast ability to initiate the synthesis and secrete protein compounds of various sizes, directly in the digestive environment.

***Artificial Digestive System TIM: Powerful In Vitro Tool*** The system TIM (TNO gastrointestinal tract model) is the in vitro model that at present time allows the closest simulation of in vivo dynamic physiological processes occurring within the lumen of the stomach and small intestine of humans [31]. It is composed of four successive compartments reproducing the stomach and the three parts of the small intestine: the duodenum, jejunum, and ileum (Figure 1). This dynamic, computer-controlled system has been designed to accept parameters and data from in vivo studies on human volunteers. The main parameters of digestion, such as pH, body temperature, peristaltic mixing and transport, gastric, biliary, and pancreatic secretions, and passive absorption of small molecules (e.g., nutrients, drugs) and water,

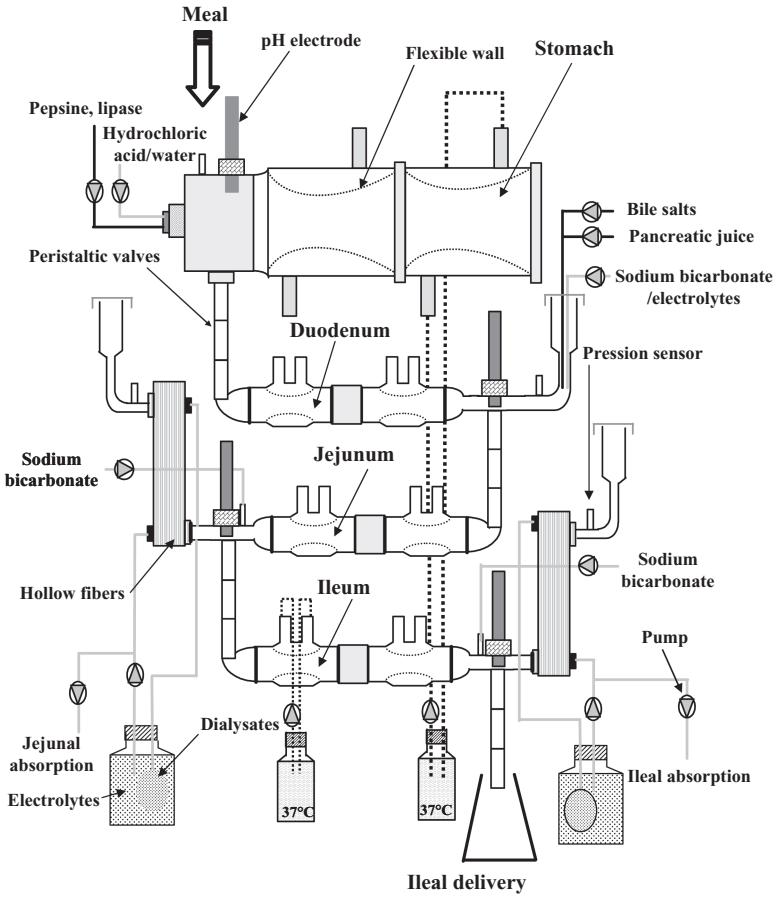


**FIGURE 1** Gastric and small intestinal system TIM.

are reproduced as accurately as possible (Table 2). This system has been previously described in detail [1, 29–31].

Compared with animal experiments, this in vitro system offers accuracy, reproducibility, easy manipulation, and the possibility of collecting samples at any level of the digestive tract and at any time during digestion with no ethical constraint. It has been validated by microbial, nutritional, and pharmaceutical studies. For instance, validation experiments demonstrate the predictive value of the TIM with regard to the survival rate of probiotic bacteria [32, 33], the digestibility of nutrients [31, 34], and the availability for absorption of minerals [35], vitamins [36], food mutagens [37], and drugs such as paracetamol [33, 38].

In the present study, the TIM was programmed to reproduce gastrointestinal conditions of the adult after the intake of a liquid meal, according to in vivo data [31, 39–41]. The initial "meal" (introduced into the artificial stomach at the beginning of digestion) was composed of (i) $10^{10}$ WRP45073A1 cells and 200 μmol of *trans*-cinnamic acid or (ii) $3 \times 10^{10}$ WppGSTV$_5$H$_6$ or WppV$_5$H$_6$ cells and galactose (40 g/L)

**TABLE 2 Digestive Parameters Reproduced in Gastrointestinal Model TIM and Their Simulation**

| | |
|---|---|
| pH | The pH is computer monitored and continuously controlled in each digestive compartment.<br>The fall of gastric pH is reproduced by adding hydrochloric acid.<br>The pH is kept to 6.5, 6.8, and 7.2 in the duodenum, jejunum, and ileum, respectively, by secreting sodium bicarbonate. |
| Temperature | The compartments are surrounded by water at body temperature (37 °C). |
| Peristaltic mixing | Peristaltic mixing is mimicked by alternate compression and relaxation of the flexible walls containing the chyme, following changes in the water pressure. |
| Dynamic of chyme transit | A mathematical model using power exponential equations [39] is used to reproduce gastric and ileal deliveries ($f = 1 - 2^{-(t/t1/2)\beta}$, where $f$ represents the fraction of meal delivered, $t$ the time of delivery, $t_{1/2}$ the half-time of delivery, and β a coefficient describing the shape of the curve).<br>Chyme transit is regulated by opening or closing the peristaltic valves that connect the compartments. |
| Volume | The volume in each compartment is monitored with a pressure sensor connected to the computer. |
| Digestive secretions | Simulated gastric, biliary, and pancreatic secretions are introduced into the corresponding compartments by computer-controlled pumps. |
| Absorption of small molecules and water | Semipermeable membrane units are connected to the jejunum and ileum to remove the products of digestion as well as water. |

in suspension in 300 mL of yeast culture medium. The parameters of in vitro digestion are summarized in Table 3.

### 2.2 Yeast Survival Rate in Simulated Gastrointestinal Conditions

At the end of in vitro digestion, yeast survival rate was evaluated by comparing the total ingested yeasts with the living yeasts recovered in both the ileal effluents of the TIM and the residual digestive content. After 240 or 270 min digestion (depending on the strain), 79.5 ± 12.1% ($n = 3$), 63.9 ± 2.4% ($n = 3$) and 75.5 ± 25.3% ($n = 4$) of the ingested WRP45073A1, $WppV_5H_6$, and $WppGSTV_5H_6$, respectively, were recovered in the ileal effluents (Figure 2). When the yeasts remaining in the residual chyme were added ($t$ = Tf), 95.6 ± 10.1% ($n = 3$), 83.1 ± 9.6% ($n = 3$), and 95.3 ± 22.7% ($n = 4$) survival percentages, respectively, were found, showing the high resistance of recombinant yeasts to gastric (pepsin and lipase) and small intestinal (bile salts and pancreatic juice) secretions and low gastric pH [29, 30]. This high survival rate was confirmed (Figure 2), no significant difference ($p < 0.05$) being observed during digestion between the ileal recovery profiles of recombinant yeasts (except for $WppV_5H_6$) and that of a nonabsorbable marker, blue dextran, added in the artificial stomach at the beginning of digestion, as previously described [31].

The survival rate of other microorganisms, such as lactic acid bacteria, has also been studied in the TIM. At the end of digestion, Marteau et al. [32] found a bacterial

**TABLE 3 Parameters of In Vitro Digestion in TIM When Simulating Gastrointestinal Conditions of Adult After Intake of Liquid Meal**

| | | |
|---|---|---|
| Gastric compartment | Initial volume | 300 mL |
| | Time (min)/pH | 0/6 |
| | | 20/4.2 |
| | | 40/2.8 |
| | | 60/2.1 |
| | | 90/1.8 |
| | | 120/1.7 |
| | Secretions | 0.25 mL/min pepsin (590 IU/mL) |
| | | 0.25 mL/min lipase (37.5 IU/mL) |
| | | 0.25 mL/min HCl 0.5 *M* if necessary |
| | Time of half emptying $t_{1/2}$ | 30 min |
| | ß coefficient | 1 |
| Duodenal compartment | Volume | 30 mL |
| | pH | Maintained at 6.5 |
| | Secretion | 0.5 mL/min bile salts (4% during first 30 min of digestion, then 2%) |
| | | 0.25 mL/min pancreatic juice ($10^3$ USP/mL) |
| | | 0.25 mL/min intestinal electrolyte solution |
| | | 0.25 mL/min $NaHCO_3$ 1 *M* if necessary |
| Jejunal compartment | Volume | 70 mL |
| | pH | Maintained at 6.8 |
| | Secretion | 0.25 mL/min $NaHCO_3$ 1 *M* if necessary |
| | Dialysis | 10 mL/min of jejunal fluid solution |
| Ileal compartment | Volume | 70 mL |
| | pH | Maintained at 7.2 |
| | Secretion | 0.25 mL/min $NaHCO_3$ 1 *M* if necessary |
| | Dialysis | 10 mL/min ileal fluid solution |
| | Time of half emptying $t_{1/2}$ | 160 min |
| | ß coefficient | 1.6 |

cumulative delivery from the ileum between 0 and 25% (depending on the tested strain). In this work, under similar experimental conditions, about 75% of ingested yeasts were recovered. Until now, the feasibility of the biodrug concept had been mainly evaluated with lactic acid bacteria (see Section 1.2). The high viability of *S. cerevisiae* in the digestive tract might favor the choice of yeasts over lactic acid bacteria as hosts for the development of biodrugs, particularly if the viability of the microorganisms is required for their in situ activity.

### 2.3 Yeast Heterologous Activity in Simulated Gastrointestinal Conditions

***Bioconversion*** The CA4H activity of WRP45073A1 yeasts was quantified measuring *p*-coumaric acid production by high-performance liquid chromatography. Control experiments showed that both *trans*-cinnamic and *p*-coumaric acids were stable under digestive conditions when no yeast was introduced into the TIM. In the presence of yeasts with no CA4H gene in their plasmid, no *p*-coumaric acid was produced, showing the specificity of the enzymatic reaction catalyzed by the recombinant model yeasts.

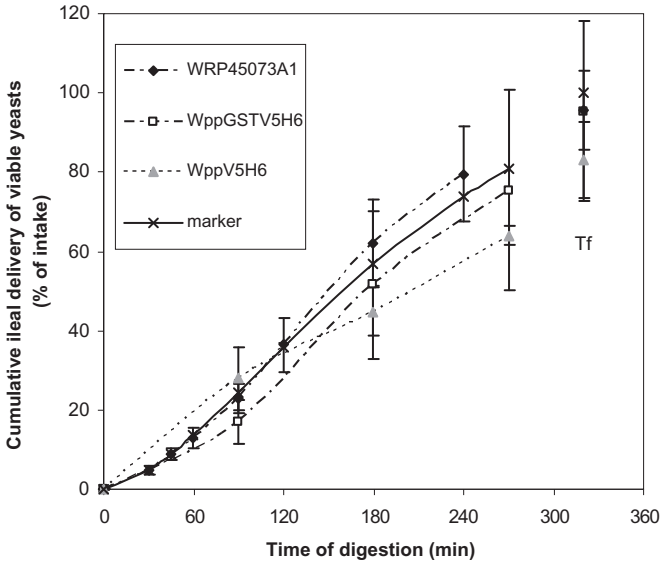


**FIGURE 2** Survival rate of three recombinant model yeasts in TIM. The cumulative ileal deliveries of viable yeasts and that of a nonabsorbable marker, bleu dextran, are represented. At the end of digestion, the percentages obtained in the cumulative ileal effluents (0–240 min or 0–270 min depending on strain) and in the residual digestive content are added ($t$ = Tf). Results are expressed as mean percentages ± SD ($n$ = 3 for WRP45073A1 and $WppV_5H_6$, $n$ = 4 for $WppGSTV_5H_6$) of intake.

At the end of digestion (240 min), 41.0 ± 5.8% ($n$ = 3) of initial *trans*-cinnamic acid was converted into *p*-coumaric acid (Figure 3*a*) [29]. By means of a computer simulation [29], in each compartment of the in vitro system, the amount of *p*-coumaric acid resulting from the CA4H activity of yeasts could be dissociated from that delivered by the previous compartment. After calculation, *trans*-cinnamic acid conversion rates of 8.9 ± 1.6%, 13.8 ± 3.3%, 11.8 ± 3.4%, and 6.5 ± 1.0% ($n$ = 3) were found in the stomach, duodenum, jejunum, and ileum, respectively (Figure 3*b*). The enzymatic reaction occurred throughout the artificial gastrointestinal tract, but mostly in the duodenum and jejunum. This could be explained by the fact that yeasts were no longer stressed by the acid pH of the stomach and could metabolize the *trans*-cinnamic acid that had previously easily entered the cells, owing to the low pH (*trans*-cinnamic acid is essentially in a cationic form which easily diffuses through the yeast membrane [42]). Also, previous studies have demonstrated that bile salts can favor enzymatic reactions [43]. The lower activity in the ileum might result from a decrease in the availability of *trans*-cinnamic acid owing to its previous conversion into *p*-coumaric acid in the upper digestive compartments. The computer simulation that was developed here should prove useful in future stages of the development of biodrugs, especially if a specific level of the digestive tract has to be targeted for drug action.

Further calculations were performed to quantify the specific enzymatic activity of the WRP45073A1 yeasts. Yeast specific activity in the TIM (from $0.05 \pm 0.04 \times 10^{-10}$ to $3.36 \pm 0.86 \times 10^{-10}$ μmol/cell/min, depending on the digestive compartment and the sampling time [29]) was close to that observed in classical batch cultures,

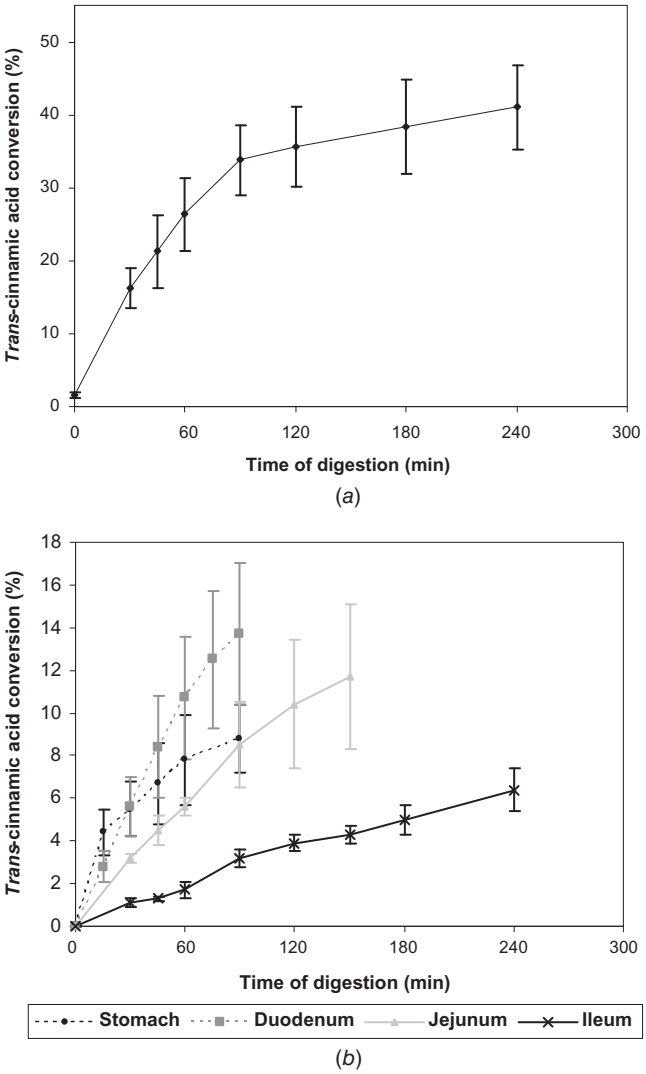


**FIGURE 3** CA4H activity of WRP45073A1 in TIM. The *trans*-cinnamic acid conversion rate was evaluated in (*a*) overall TIM and (*b*) each compartment of TIM. Results are expressed as mean percentages ± SD ($n = 3$) of ingested *trans*-cinnamic acid converted into *p*-coumaric acid. (Reprinted with permission from Blanquet et al., *Applied and Environmental Microbiology*, 69, 2889.)

which is very encouraging for a potential use of *S. cerevisiae* as a biodetoxication system to the gut.

***Secretion of Peptides or Proteins*** The production of the GST–$V_5H_6$ and peptide–$V_5H_6$ in the TIM was examined by Western blotting (data not shown). No signal was detectable during control digestions without yeast or with yeasts with no heterologous gene in their plasmid. The model protein and peptide were detected as early as 90 min after the yeast intake/gene induction in each compartment of the in vitro system and remained until 270 min of digestion in the lower part of the

small intestine [30]. No signal was detectable in the "meal" before introduction into the artificial stomach, showing the efficient initiation of protein compound synthesis and secretion by galactose in the digestive environment.

The amount of GST–$V_5H_6$ produced in each compartment of the TIM was quantified by enzyme-linked immunosorbent assay (ELISA). The protein concentrations in the digestive medium reached 15 ng/mL, the highest values being found in the jejunum and ileum from 150 min to the end of digestion (Figure 4*a*). The GST–$V_5H_6$ concentrations in the digestive environment were close to those measured in standard batch

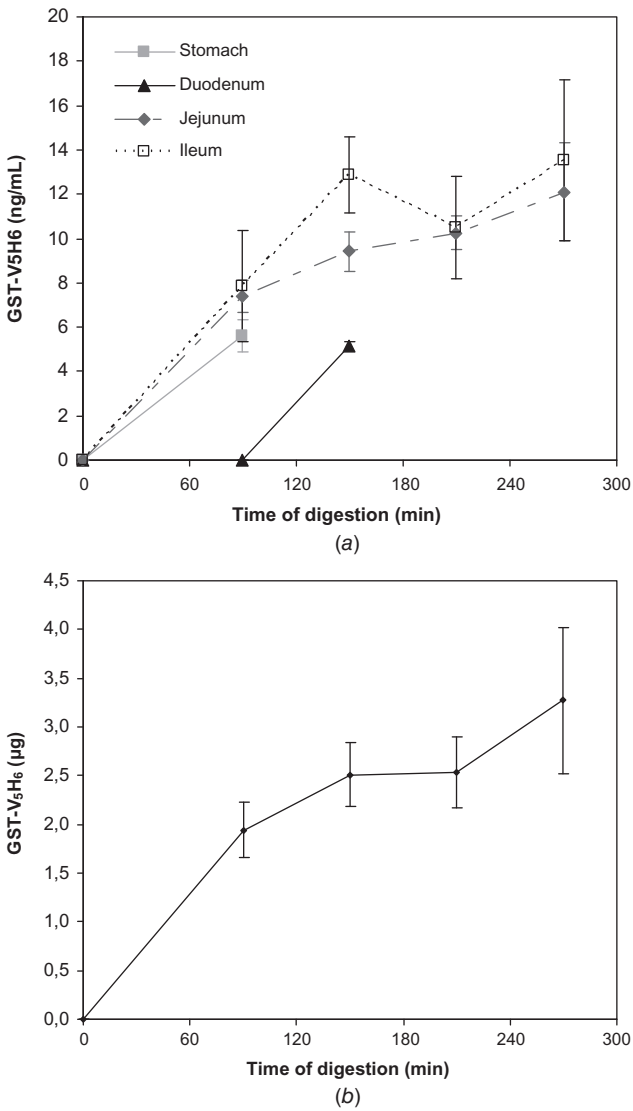


**FIGURE 4** Immunoenzymatic (ELISA) measurement of GST–$V_5H_6$ produced by WppG-$STV_5H_6$ in (*a*) different compartments of TIM (ng/mL) and (*b*) the overall TIM (µg). Error bars represent standard deviations ($n$ = 4). (Reprinted with permission from 110, Blanquet et al., *Journal of Biotechnology*, 45, 2006. Copyright 2006 by Elsevier.)

cultures. Therefore, the low secretion levels of GST–$V_5H_6$ may be only imputed to the large size of the protein and/or the genetic construction, but not to the particular digestive conditions. This hypothesis is consistent with the results of Zsebo et al. [44], who showed that *S. cerevisiae* is an efficient host for small polypeptide secretion, but not for larger proteins, which accumulate in the periplasmic place and cell wall of the yeast. Improved secretion levels might be obtained with another heterologous protein or a different expression vector. In the overall in vitro system, the total amount of GST–$V_5H_6$ regularly increased during digestion to reach 3.3 ± 0.7 μg ($n$ = 4) after 270 min digestion (Figure 4*b*).

To check that the GST–$V_5H_6$ recovered in the TIM was truly secreted by living recombinant yeasts and did not result from cell lysis, a control strain producing an intracellular form of the model protein was contructed (without the leader sequence) and tested in similar experimental conditions. Some GST–$V_5H_6$ was found in the ileum showing that cell lysis occurred in this compartment. At the end of digestion, the total amount of GST–$V_5H_6$ released by the control strain represented 30% (1 ± 0.2 μg, $n$ = 2) of the protein produced by WppGST$V_5H_6$, showing a significant contribution of cell lysis to the release of the protein in the TIM.

For the first time, the amount of heterologous proteins secreted by recombinant *S. cerevisiae* was evaluated throughout the upper digestive tract. Until now, the secretion efficiency of recombinant microorganisms had never been directly quantified throughout the digestive tract. The ability of recombinant strains to produce heterologous proteins in the digestive environment had been mainly demonstrated indirectly, following the biological effect of the protein: immune response (antibody production) [7, 15–17, 19, 20], growth improvement [45], or reduction in colitis symptoms [5]. Nevertheless, the ability of recombinant bacteria to initiate protein synthesis in situ has been reported in a few studies. Oozeer et al. [46] have shown that engineered *Lactobacillus casei* was able to initiate the synthesis of luciferase during its transit in the digestive tract of a human flora-associated mouse model. Steidler et al. [5] have demonstrated in mice the in situ synthesis of mouse IL-10 by recombinant *L. lactis*, the viability of these microorganisms being required to achieve their therapeutic effect (see Section 1.2). They have further documented this result by showing the de novo synthesis of IL-10 in the colon of IL-10 –/– mice. Moreover, these authors have quantified the amount of IL-10 produced by the engineered *L. lactis* in two animal models, but only in a limited part of their digestive tract : (i) 7 ng of mouse IL-10 was recovered in the colon of IL-10 –/– mice, but the interleukin was not detectable in other areas of the gastrointestinal tract [5] and (ii) about 470 pg/mL of human IL-10 was found in an ileal loop of a pig 4 h after injection of the recombinant bacteria [47]. Unlike what was previously obtained in the TIM with GST–$V_5H_6$, the concentrations of IL-10 found in the digestive tract of the animals were much lower than that recovered in batch cultures [5, 47].

## 2.4 Conclusion

For the first time, the ability of engineered *S. cerevisiae* to carry out a bioconversion reaction [29] and initiate the synthesis and secrete protein compounds of various sizes [30] was shown throughout the upper gastrointestinal tract in human simulated digestive conditions. The CA4H specific activity of WRP45073A1 and the secretion level of GST–$V_5H_6$ were surprisingly similar to that obtained in classical

batch cultures. This is particularly remarkable as the expression strategy of the model genes had not yet been adapted to the particular constraints of the digestive environment and promising for a future use of recombinant *S. cerevisiae* as host for biodrug development.

## 3 ORAL FORMULATION OF RECOMBINANT YEASTS

Once the scientific feasibility of the new drug delivery system was established, the development of pharmaceutical formulations allowing the oral administration of the genetically modified *S. cerevisiae* was considered. Ideally, these oral drug dosage forms would improve both the survival and the heterologous activity of yeasts in the digestive environment. The following works were carried out only with the strain expressing the model P450. In a preliminary step, the effect of a preservation technique (lyophilization) and an immobilization procedure (entrapment in whey protein beads) on the survival rate and heterologous activity of the model strain WRP45073A1 was assessed in simulated digestive conditions.

### 3.1 Freeze Drying of Recombinant Model Yeasts

***Freeze-Drying Conditions*** Freeze drying is a technique of dehydration commonly used for the formulation of drugs containing nonrecombinant *Saccharomyces* spp. [48–50]. Standard freeze-drying conditions derived from the literature [51–54] and our own experiments were applied for the lyophilization of the genetically modified model yeasts. The effect of cryoprotectants was further investigated because it appears as one of the most important parameters during lyophilization [51, 52]. Following galactose induction of the heterologous CYP73A1, yeasts in the beginning of their stationary growth phase ($10^9$ cells/mL) were lyophilized in suspensions of trehalose 10% w/v, maltose 10% w/v, lactose 10% w/v, or a mixture of 5% w/v milk proteins and 10% w/v trehalose. The parameters of lyophilization are summarized in Table 4 [55].

*Saccharomyces cerevisiae* WRP45073A1 survives freeze drying and yeast survival rates were dependent on the nature of the cryoprotectants: 13.1 ± 1.8%, 9.5 ± 6.0%,

**TABLE 4 Parameters of Lyophilization Used for Recombinant Model Yeast WRP45073A1**

| | |
|---|---|
| Recombinant yeasts | Growth phase: beginning of stationnary growth phase |
| | Cell concentration in freeze-drying flasks: $10^9$ cells/mL |
| Cryoprotectants | 10% w/v trehalose |
| | 10% w/v maltose |
| | 10% w/v lactose |
| | 5% w/v milk proteins and 10% w/v trehalose |
| | Control: physiological water |
| Freeze-drying conditions | Volume of sample: 5 mL |
| | Cooling rate: 1 °C/min |
| | Condenser plate temperature: –40 °C |
| | Time of lyophilization: 24 h |
| | Heating temperature (secondary drying): 23 °C |

7.7 ± 4.6%, and 7.1 ± 4.0% ($n$ = 5) for the milk protein–trehalose mix, lactose, maltose, and trehalose, respectively [55]. The protective effect of trehalose [51–53] and maltose [52] (but not that of lactose) compared to physiological water (survival rate 0.3 ± 0.2%, $n$ = 5) had already been shown in non genetically modified *S. cerevisiae*. Several mechanisms have been proposed to explain this protective effect. One hypothesis is related to the ability of these carbohydrates to form a glassy structure during drying, responsible for the long stability of biological materials [56]. The milk protein–trehalose mix led to a higher survival compared with trehalose alone ($p < 0.05$). A similar result had already been observed by Abadias et al. [54], who showed an improvement in the viability of another yeast, *Candida sake*, from 7 to 29% when 5% skim milk is used in combination with 10% trehalose.

***Influence of Cryoprotectants on Viability and Heterologous Activity of Lyophilized Yeasts in Simulated Gastrointestinal Conditions*** To evaluate the influence of cryoprotectants on both the survival rate and CA4H activity of WRP45073A1 in simulated gastrointestinal conditions, $10^{10}$ viable freeze-dried yeasts and 200 μmol of *trans*-cinnamic acid were simultaneously introduced into the TIM. Yeast cells were lyophilized in the presence of the milk protein–trehalose mix, trehalose, lactose, or maltose, as previously explained (see above). The freeze-dried samples were introduced into the artificial stomach suspended in 300 mL of yeast culture medium without any storage period. The number of viable cells introduced into the TIM was determined from previously obtained survival rates (cf. Section 3.1).

*Viability of Freeze-Dried Yeasts in TIM* Freeze-dried yeasts showed a high tolerance to gastric and small intestinal conditions. At the end of digestion (240 min), 61.5 ± 0.7%, 59.9 ± 3.8%, 56.3 ± 5.4%, and 55.6 ± 6.0% ($n$ = 3) of the ingested cells were recovered in the ileal effluents of the TIM, following freeze drying in the presence of the milk protein–trehalose mix, maltose, lactose, and trehalose, respectively (Figure 5). When the yeasts remaining in the residual chyme were added ($t$ = Tf), 84.7 ± 3.5%, 87.0 ± 6.4%, 83.7 ± 6.2%, and 70.7 ± 9.2% ($n$ = 3) survival percentages were found, respectively. Nevertheless, whatever the protective agent, the cumulative ileal delivery of freeze-dried yeasts remained significantly lower ($p < 0.05$) than that of the nonabsorbable marker (see Section 2.2) and that of nondried ("native") yeasts. The higher sensitivity to digestive conditions of freeze-dried cells compared with native ones may be linked to the damage caused to cells during drying and rehydration, resulting in increased membrane permeability [57].

In addition, no significant difference was observed between the ileal recovery profiles of yeasts with the various cryoprotectants, showing their lack of influence on the survival of freeze-dried WRP45073A1 in the TIM.

A few studies have evaluated the survival rate of freeze-dried *S. cerevisiae* spp. in human volunteers following their oral administration. Nevertheless, comparison between in vitro results in the TIM and these in vivo data is hampered by the fact that yeast survival had been evaluated only in feces (and not at the end of the ileum) after a single or multiple oral administration of the microorganisms. Klein et al. [49] found a fecal recovery of 0.12 ± 0.04% ($n$ = 8) after a single dose of 1 g of *S. boulardii* [$10^{10.4}$ colony-forming units (CFU)] to healthy volunteers, and Blehaut et al. [50] measured a steady-state fecal recovery of 0.36 ± 0.31% ($n$ = 8)

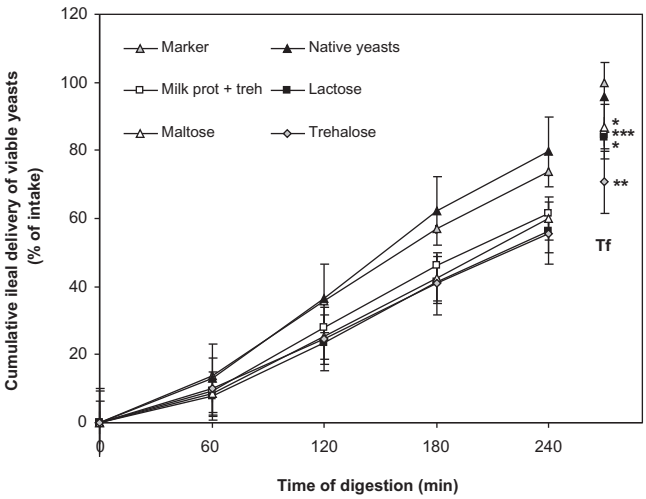


**FIGURE 5** Effect of cryoprotectants on survival rate of WRP45073A1 in TIM. The cumulative ileal deliveries of viable freeze-dried and native yeasts and that of the nonabsorbable marker are represented. At the end of digestion, the percentages obtained in the cumulative ileal effluents (0–240 min) and the residual digestive content are added ($t$ = Tf). Results are expressed as mean percentages ± SD ($n = 3$) of intake. Significantly different from the marker ($t$ = Tf) at $p < 0.05$ (*), $p < 0.01$ (**) and $p < 0.001$ (***). (Reprinted with permission from Blanquet et al., *European Journal of Pharmaceutics and Biopharmaceutics*, 61, 37, 2006. Copyright 2006 by Elsevier.)

after oral administration of 1 g of *S. boulardii* for 15 days. These survival rates in feces are much lower than those obtained in the ileal effluents of the TIM. However, a recent work [29] has shown that native WRP45073A1 yeasts are very sensitive to colonic conditions: Yeast viability in an artificial digestive system reproducing the human large intestine [58] was only 1.2 ± 0.4% ($n = 3$) after 12 h and no more yeast could be detected following 24 h fermentation.

*Heterologous Activity of Freeze-Dried Yeasts in TIM* The ability of freeze-dried recombinant yeasts to catalyze the bioconversion of *trans*-cinnamic acid into *p*-coumaric acid in the in vitro system was shown whatever the tested protectant. At the end of the experiment, conversion rates of 24.2 ± 1.0%, 17.7 ± 2.2%, 15.1 ± 3.3%, and 16.5 ± 0.7% ($n = 3$) were found for the yeasts lyophilized in the presence of milk proteins plus trehalose, maltose, lactose, and trehalose, respectively (Figure 6). During all the in vitro digestion, the CA4H activity of freeze dried cells remained significantly lower ($p < 0.01$) than that of native ones. This lower activity could result from the P45073A1 damage during freeze drying, probably aggravated by the membranous location of the enzyme. Among those tested, the cryoprotectant that allowed the highest CA4H activity was the milk protein–trehalose mix ($p < 0.05$).

***Conclusion*** Although the impact of freeze drying on both the survival rate and heterologous activity of yeasts in the artificial digestive system was found to be adverse, lyophilization appears to be a convenient technique for the dehydration of recombinant *S. cerevisiae*. Among the tested cryoprotectants, the association of milk proteins and trehalose was the most efficient to maintain the CA4H activity of

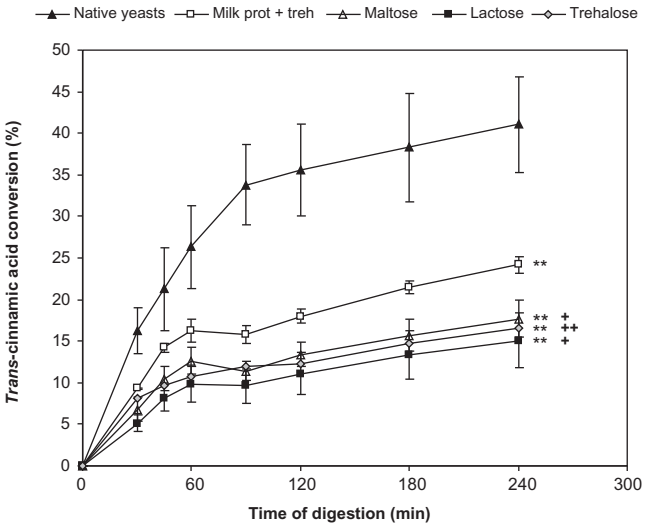


**FIGURE 6** Effect of cryoprotectants on CA4H activity of WRP45073A1 in TIM. The CA4H activity of freeze-dried and native yeasts was evaluated in the overall TIM. Values are expressed as mean percentages ± SD ($n$ = 3) of initial *trans*-cinnamic acid converted into *p*-coumaric acid. Significantly different from native yeasts ($t$ = 240 min) at $p < 0.01$ (**). Significantly different from milk proteins/trehalose group ($t$ = 240 min) at $p < 0.05$ (**+**) and $p < 0.01$ (**++**). (Reprinted with permission from Blanquet et al., *European Journal of Pharmaceutics and Biopharmaceutics*, 61, 38, 2006. Copyright 2006, by Elsevier.)

recombinant model yeasts in the stringent digestive conditions. This study also gives an example of the usefulness of the TIM in the prescreening of pharmaceutical excipients, such as cryoprotectants.

### 3.2 Immobilization of Recombinant Model Yeasts in Whey Protein Beads

***Cell Immobilization*** Among the available techniques, the entrapment in gel beads is frequently used for the immobilization of living cells in food sciences [59, 60] because of its simplicity and low cost. This technique has been recently extended to microorganisms with probiotic activity with the aim of increasing their survival in the human digestive environment and particularly in the stomach [61, 62].

Whey proteins have been recently considered a potential alternative to the commonly used alginate [59, 61–63] for the production of gel beads. They have been used to entrap drugs such as retinol [64] and living microorganisms such as bifidobacteria [65], but until now not yeasts. A new immobilization system using whey proteins was then developed for entrapping the recombinant model yeasts WRP45073A1 in order to ensure their oral administration [66].

The formation of beads is a two-step process based on the cold gelation of whey proteins in the presence of divalent cations, such as $Ca^{2+}$ [67]. Briefly, the whey protein isolate (WPI) solution (10% w/v in deionized water) was (i) adjusted at pH 7 to favor the apparition of negative charges implied in ionic bounds with $Ca^{2+}$ ions and (ii) heated (80 °C, 45 min) to denaturate the proteins. Recombinant cells in the beginning of their stationnary growth phase were suspended in a sterile solution of

10% w/v lactose (final concentration $10^9$ cells/mL) and added to denaturated WPI solution (7% v/v). The extrusion of the mixture through a needle led to the production of droplets forming gel beads in a calcium bath (0.1 *M* $CaCl_2$). This protocol allows the obtaining of spherical beads (diameter 2 605 ± 18 μm, $n = 3$) with an homogeneous distribution of yeasts through the matrix ($1.15 \times 10^6$ viable cells per bead) [66]. The lack of influence of the immobilizing procedure on the viability of yeasts was also shown [66].

***Gastric Digestion Protocol*** A most sought-after property of gel beads is their potential resistance to gastric conditions. Authors have already shown in vitro that beads resulting from the cold-induced gelation of a whey protein–oil emulsion [64] or a whey protein–polysaccharide mix [65] were gastroresistant. To further investigate the involvement of whey proteins in the gastroresistance of beads, the behavior of entrapped WRP45073A1 yeasts was followed in simulated gastric conditions.

The human gastric environment was reproduced using a simple in vitro model adapted from that initially developped by Yvon et al. [68]. The main parameters of gastric digestion are reproduced according to in vivo data: decrease of pH, pepsin supply, body temperature, mixing, and gastric emptying (Figure 7). This system was validated by studies on the digestability of milk proteins (unpublished data). In the present work, it was programmed to reproduce gastric conditions of the adult after the intake of a glass of milk. Initially, $10^{10}$ viable entrapped WRP45073A1 cells and 200 μmol of *trans*-cinnamic acid were simultaneously introduced into the artificial stomach, suspended in 300 mL of yeast culture medium. Table 5 summarizes the parameters of in vitro gastric digestion.

***Release of Yeasts from Beads in Simulated Gastric Conditions*** The release of entrapped WRP45073A1 cells from whey protein beads was followed in the artificial

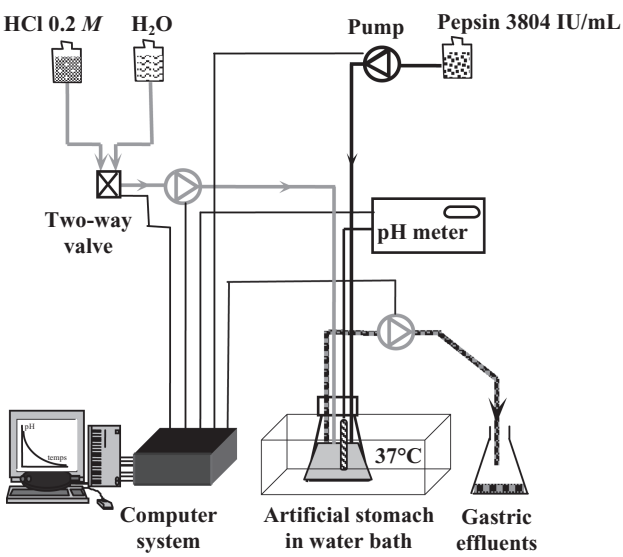


**FIGURE 7** Gastric digestive system.

**TABLE 5 Parameters of In Vitro Digestion in Artificial Stomach When Simulating Gastric Conditions of Adult after Intake of Glass of Milk**

| | |
|---|---|
| Initial "meal" | |
| Volume | 300 mL (constant during all digestion) |
| Pepsin | 75 IU/mL |
| Acidification | |
| Time (min)/pH | 0/6.5, 60/2.1<br>15/4.4, 75/1.9<br>30/3.2, 90/1.7<br>45/2.5, 120/1.6 |
| Flow rate | 2 mL/min HCl 0.2 *M* if necessary |
| Exponential base | $e = 1.04$ |
| Pepsin supply | |
| Time (min)/pepsine (IU/mL) | 0/77.6, 60/139.2<br>15/99.7, 75/146.7<br>30/117.1, 90/152.8<br>45/129.7, 120/163.0 |
| Flow rate | From 13 ($t = 0$) to 1 ($t = 120$ min) mL/min pepsin 3804 IU/mL |
| Exponential base | $e = 1.03$ |
| Gastric emptying | |
| Flow rate | From 10 ($t = 0$) to 3 ($t = 120$ min) mL/min |
| Exponential base | $e = 1.03$ |
| Time of digestion | 120 min |

stomach. During the first 60 min of gastric digestion, a few percentage points (2.2 ± 0.9%, $n = 3$) of initial entrapped yeasts was recovered in the gastric medium (Figure 8). This low percentage cannot be explained by cell death since control experiments showed the high survival rate of free yeasts (about 90%) during all the digestion [66]. These results are in agreement with those of Beaulieu et al. [64], who have observed that only 5–10% of the incorporated retinol was released from the whey protein–oil matrix following 30 min incubation in HCl 0.1 *M* and pepsin 24 mg/L. The low release of yeasts in the first hour of digestion indicates that the beads are resistant to acidification until pH 2 (cf. Table 5) and pepsin attack, which implies that they might cross the gastric barrier in humans.

From 60 min digestion, the percentage of released yeasts increased regularly to reach 39 ± 5% ($n = 3$) at 120 min. This phenomenon might be explained by two hypotheses: (i) a swelling of beads due to an increase in the acidity of the medium or (ii) a degradation of the matrix resulting from a raise in pepsin concentration. Complementary studies conducted to further evaluate the effect of pH and pepsin on beads have shown that pepsin has no effect on the protein matrix whatever the tested pH [66]. As already suggested by other authors [69], this resistance to enzyme attack might be due to the formation of hydrophobic interactions between aromatic amino acids of ß-lactoglobulin, the major whey protein. On the contrary, incubation at pH 2 led to an increase in the diameter of beads, certainly due to high electrostatic repulsive forces (between positive charges of protoned amino acids and $Ca^{2+}$), which induced a raise in the pore size. In conclusion, the release of yeasts observed from

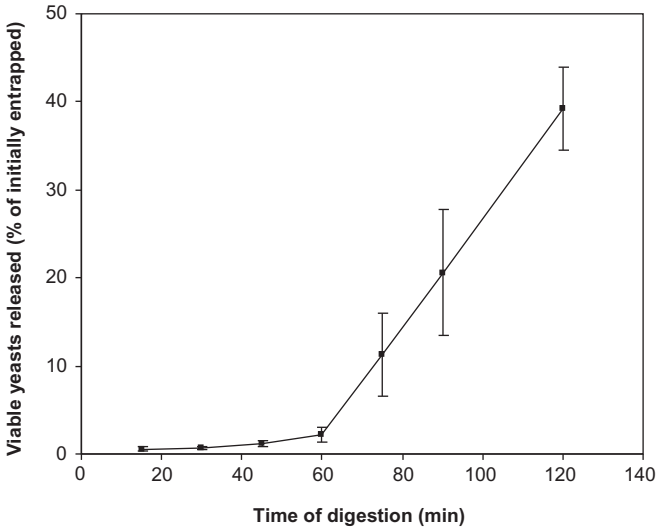


**FIGURE 8** Release of WRP45073A1 cells from whey protein beads in simulated gastric conditions. Results are expressed as mean percentages ± SD ($n = 3$) of initial entrapped yeasts. (Reprinted with permission from Hebrard et al., *Journal of Biotechnology*, in press. Copyright 2006 by Elsevier.)

60 min digestion is provoked by acidic conditions rather than by enzymatic degradation of beads.

***Influence of Entrapment on Heterologous Activity of Yeasts in Simulated Gastric Conditions*** In order to evaluate the influence of the entrapment process of the CA4H activity of WRP45073A1, the heterologous activity of free and entrapped yeasts was followed in the gastric system under similar experimental conditions. In both cases, *p*-coumaric acid was detected in the gastric medium as soon as 15 min after the beginning of the experiment (Figure 9). This implies that both *trans*-cinnamic and *p*-coumaric acid could diffuse through bead pores.

During all the digestion, the CA4H activity of entrapped yeasts was significantly ($p < 0.05$) higher than that of free ones (expected for $t = 15$ min and $t = 75$ min). At 120 min, 63.4 ± 1.6% ($n = 3$) of initial *trans*-cinnamic acid was converted into *p*-coumaric acid for immobilized yeasts versus 51.5 ± 1.8% ($n = 3$) for control yeasts. This phenomenon was particularly marked from 30 to 60 min of digestion when a very low amount of recombinant yeasts was released from beads. As suggested by Bienaimé et al. [70], beads might create a microenvironment (e.g., a buffer effect toward low pH or enzyme attack) favoring the heterologous activity of yeasts. In a similar way, the microenvironment resulting from the presence of an alginate matrix improves the invertase activity of recombinant *S. cerevisiae* in batch cultures [71].

***Conclusion*** This preliminary work reveals whey proteins as a convenient material for immobilizing recombinant yeasts. Gel beads were resistant to acidification until pH 2 and pepsin attack, suggesting that they should cross the gastric barrier in humans. Moreover, the presence of the protein matrix seemed to create "microconditions" that favor the heterologous activity of entrapped yeasts.

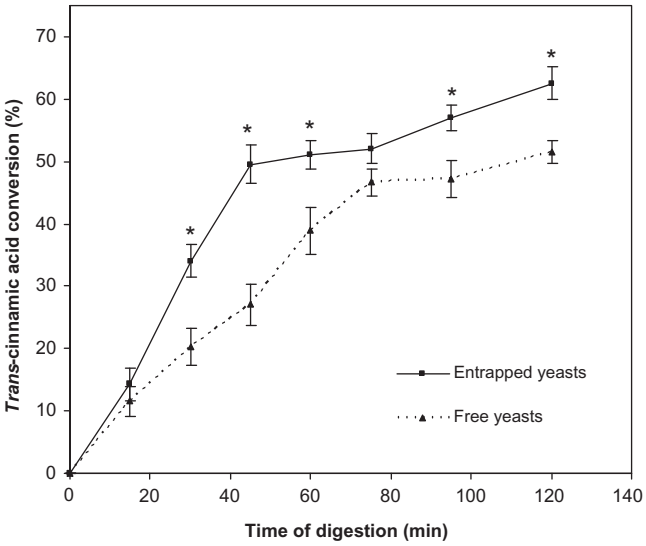


**FIGURE 9** Influence of immobilization on CA4H activity of WRP45073A1 in simulated gastric conditions. The CA4H activity of entrapped and free yeasts was evaluated in the artificial gastric system. Results are expressed as mean percentages ± SD ($n = 3$) of *trans*-cinnamic acid converted into *p*-coumaric acid. Significantly different from free yeasts at $p < 0.05$ (*). (Reprinted with permission from Hebrard et al., *Journal of Biotechnology*, in press. Copyright 2006 by Elsevier.)

## 4 GENERAL CONCLUSION AND FUTURE DEVELOPMENTS

Using genetically engineered microorganisms as new delivery vehicles to the gut is an important challenge for the development of innovative drugs. A potential application directly issued from the present work is the development of drug delivery systems based on orally administered yeasts carrying out a bioconversion reaction or secreting compounds directly in the human digestive tract.

Soon, the choice of candidate genes as well as the most appropriate dosage forms will be made according to the therapeutic target. Oral formulations will be optimized in order to (i) control the release of yeasts according to their action site in the gastrointestinal tract, (ii) maximize the heterologous activity of yeasts (by addition of the appropriate substrate and/or inductor), and (iii) ensure a stability of both yeasts and pharmaceutical dosage forms before administration to the patient. Of course, heterologous gene expression strategies have to be tailored for a safe use in humans, the presence of mobilizable vectors, antibiotic selection markers, and bacterial sequences liable to promote gene transfer to host microflora being prohibited. In addition, environmental confinement of recombinant cells has to be achieved by introducing a suicide process that triggers the elimination of the microorganisms upon leaving the digestive tract. Two types of biological containment systems may be considered [1, 3]: (i) the active system, which should provide control of the recombinant microorganism dissemination through the conditional production of a toxic protein [72–74], and (ii) the passive system, which could render the cell growth dependent on the complementation of an auxotrophy or other gene defects [75, 76].

Steidler et al. [47] have already developed and validated in pigs a passive containment system for the *L. lactis* expressing human IL-10 by deleting the thymidylate synthase gene which is essential for their growth (the resulting strain being dependent on thymidine or thymine).

The present study also shows the particular interest of the TIM in drug development and testing. This artificial system will constitute a powerful alternative to animal experimentation during all preclinical phases of biodrug development. The in vitro model can aid in the selection of pharmaceutical formulations, ensuring both the release of yeasts directly at their action site and their optimal activity. The efficiency of newly developed molecular tools (e.g., promoters, selection markers, and vectors) can also be evaluated to optimize the functionality of recombinant strains in the digestive environment. As the mucosal layer is not involved in the actual configuration of the TIM, this system may be used in combination with intestinal cells in culture (e.g., Caco-2) to study the mucosal transport and metabolism of the active compounds produced by recombinant yeasts. Moreover, experiments in a large intestinal model [58], complementary of the gastric and small intestinal system, could provide necessary data on the biological safety of engineered microorganisms. For example, the potential gene transfer to the human flora can be studied and the cell death outside of the digestive tract can be checked to ensure there is no dissemination in the environment.

This study opens up new opportunities in the development of new drug delivery vectors based on engineered living yeasts for the prevention or treatment of various diseases in human.

## REFERENCES


1. Blanquet, S., Marol-Bonnin, S., Beyssac, E., Pompon, D., Renaud, M., and Alric, M. (2001), The Biodrug concept: An innovative approach to therapy, *Trends Biotechnol.*, 19, 393–400.
2. Corthier, G., and Renault, P. (1999), Future directions for research on biotherapeutic agents: Contribution of genetic approaches on lactic acid bacteria, in Elmer, G. W., McFarland, L., Surawicz, C., Eds., *Biotherapeutic Agents and Infection Diseases*, Humana, Totowa, NJ, pp. 269–304.
3. Steidler, L. (2003), Genetically engineered probiotics, *Best Practise Res. Clin. Gastroenterol.*, 17, 861–876.
4. Chatel, J. M., Langella, P., Adel-Patient, K., Commissaire, J., Wal, J. M., and Corthier, G. (2001), Induction of mucosal immune response after intranasal or oral inoculation of mice with *Lactococcus lactis* producing bovine beta-lactoglobulin, *Clin. Diag. Lab. Immunol.*, 8, 545–551.
5. Steidler, L., Hans, W., Schotte, L., Neirynck, S., Obermeier, F., Falk, W., Fiers, W., and Remaut, E. (2000), Treatment of murine colitis by *Lactococcus lactis* secreting Interleukin-10, *Science*, 289, 1352–1355.
6. Schreuder, M. P., Deen, C., Boersma, W. J. A., Pouwels, P. H., and Klis, F. M. (1996), Yeast expressing hepatitis B virus surface antigen determinants on its surface: Implications for a possible oral vaccine, *Vaccine*, 14, 383–388.
7. Reveneau, N., Goeffroy, M. C., Locht, C., Chagnaud, P., and Mercenier, A. (2002), Comparison of the immune responses induced by local immunizations with recombinant

*Lactobacillus plantarum* producing tetanus toxin fragment C in different cellular locations, *Vaccine*, 20, 1769–1777.

8. Drouault, S., Corthier, G., Ehrlich, S. D., and Renault, P. (1999), Survival, physiology, and lysis of *Lactococcus lactis* in the digestive tract, *Appl. Environ. Microbiol.*, 65, 4881–4886.
9. Fahl, W. E., Loo, D., and Manoharan, H. (1999), Chemoprotective bacterial strains, International Patent WO 99/27953.
10. Baranov, V. S., Ivaschenko, T., Bakay, B., Aseev, M., Belotserkovskaya, R., Baranova, H., Malet, P., Perriot, J., Mouraire, P., Baskakov, V. N., Savitskyi, G. A., Gorbushin, S., Deyneka, S. I., Michnin, E., Barchuck, A., Vakharlovsky, V., Pavlov, G., Shilko, V. I., Guembitzkaya, T., and Kovaleva, L. (1996), Proportion of the GSTM1 0/0 genotype in some Slavic populations and its correlation with cystic fibrosis and some multifactorial diseases, *Hum. Genet.*, 97, 516–520.
11. Drouault, S., Juste, C., Marteau, P., Renault, P., and Corthier, G. (2002), Oral treatment with *Lactococcus lactis* expressing *Staphylococcus hyicus* lipase enhances lipid digestion in pigs with induced pancreatic insufficiency, *Appl. Environ. Microbiol.*, 68, 3166–3168.
12. Chang, T. M., and Prakash, S. (1998), Therapeutic uses of microencapsulated genetically engineered cells, *Mol. Med. Today*, 4, 221–227.
13. Prakash, S., and Chang T. M. S. (2000), In vitro and in vivo uric acid lowering by artificial cells containing microencapsulated genetically engineered *E. coli* DH5 cells, *Int. J. Artif. Organs*, 23, 429–435.
14. Pouwels, P. H., Leer, R. J., Shaw, M., Bak-Glashouwer, M. J. H. D., Tielen, F. D., Smit, E., Martinez, B., Jore, J., and Conway, P. L. (1998), Lactic acid bacteria as antigen delivery vehicles for oral immunization purposes, *Int. J. Food Microbiol.*, 41, 155–167.
15. Robinson, K., Chamberlain, L. M., Schofield, K. M., Wells, J. M., and Le Page, R. W. F. (1997), Oral vaccination of mice against tetanus with recombinant *Lactococcus lactis*, *Nat. Biotechnol.*, 15, 653–657.
16. Nayak, A. R., Tinge, S. A., Tart, R. C., McDaniel, L. S., and Briles, D. E., 3rd, and Curtiss, R. (1998), A live recombinant avirulent oral *Salmonella* vaccine expressing pneumococcal surface protein A induces protective responses against Streptococcus pneumoniae, *Infect. Immun.*, 66, 3744–3751.
17. Corthesy, B., Boris, S., Isler, P., Grangette, C., and Mercenier, A. (2005), Oral imunization of mice with lactic acid bacteria producing *Helicobacter pylori* urease B subunit partially protects against challenge with Helicobacter felis, *J. Infect. Dis.*, 192, 1441–1449.
18. Barron, M. A., Blyveis, N., Pan, S. C., and Wilson, C. C. (2006), Human dendritic cell interactions with whole recombinant yeasts: Implication for HIV-1 vaccine development, *J. Clin. Immunol.*, 26, 251–264.
19. Steger, K. K., Valentine, P. J., Heffron, F., So, M., and Pauza, C. D. (1999), Recombinant, attenuated *Salmonella typhimurium* stimulate lymphoproliferative responses to SIV capsid antigen in rhesus macaques, *Vaccine*, 17, 923–932.
20. Adel-patient, K., Ah-Leung, S., Creminon, C., Nouaille, S., Chatel, J. M., Langella, P., and Wal, J. M. (2005), Oral administration of recombinant *Lactococcus lactis* expressing bovine beta-lactoglobulin partially prevents mice from sensitisation, *Clin. Exp. Allergy*, 35, 539–546.
21. Mercenier, A., Müller-Alouf, H., and Grangette, C. (2000), Lactic acid bacteria as live vaccines, *Curr. Issues Mol. Biol.*, 2, 17–25.
22. Bumann, D., Hueck, C., Aebisher, T., and Meyer, T. F. (2000), Recombinant live *Salmonella spp*. for human vaccination against pathogens, *FEMS Immunol. Med. Microbiol.*, 27, 357–364.

23. Steidler, L. (2002), In situ delivery of cytokines by genetically engineered *Lactococcus lactis*, *Ant. Leeuwenh.*, 82, 323–331.
24. Steidler, L. (2001), Microbiological and immunological strategies for treatment of inflammatory bowel disease, *Microbes Infect.*, 3, 1157–1166.
25. Braat, H., Rottiers, P., Hommes, D. W., Huyghebaert, N., Remaut, E., Remon, J. P., Van Deventer, S. J., Neirynck, S., Peppelenbosch, M. P., and Steidler, L. (2006), A phase I trial with transgenic bacteria expressing interleukin-10 in Crohn's disease, *Clin. Gastroenterol. Hematol.*, 4, 754–759.
26. Lewis, S. J., and Freedman, A. R. (1998), Review article: The use of biotherapeutic agents in the prevention and treatment of gastrointestinal disease, *Aliment. Pharmacol. Ther.*, 12, 807–822.
27. Bergogne-Berezin, E. (2000), Treatment and prevention of antibiotic-associated diarrhea, *Int. J. Antimicrobial. Agents*, 16, 521–526.
28. Urban, P., Werck-Reichhart, D. H., Teutsch, G., Durst, F., Regnier, S., Kazmaier, M., and Pompon, D. (1994), Characterization of recombinant plant cinnamate 4-hydroxylase produced in yeast. Kinetic and spectral properties of the major plant P450 of the phenylpropanoid pathway, *Eur. J. Biochem.*, 222, 843–850.
29. Blanquet, S., Meunier, J. P., Minekus, M., Marol-Bonnin, S., and Alric, M. (2003), Recombinant *Saccharomyces cerevisiae* expressing a P450 in artificial digestive systems: A model for biodetoxication in the human digestive environment, *Appl. Environ. Microbiol.*, 69, 2884–2892.
30. Blanquet, S., Antonelli, R., Laforet, L., Denis, S., Marol-Bonnin, S., and Alric, M. (2004), Living recombinant *Saccharomyces cerevisiae* secreting proteins or peptides as a new drug delivery system in the gut, *J. Biotechnol.*, 110, 37–49.
31. Minekus, M., Marteau, P., Havenaar, R., and Huis in't Veld, J. H. J. (1995), A multicompartmental dynamic computer-controlled model simulating the stomach and small intestine, *ALTA*, 23, 197–209.
32. Marteau, P., Minekus, M., Havenaar, R., and Huis in't Veld, J. H. J. (1997), Survival of lactic acid bacteria in a dynamic model of the stomach and small intestine: Validation and the effects of the bile, *J. Dairy Sci.*, 80, 1031–1037.
33. Blanquet, S., Zeijdner, E., Beyssac, E., Meunier, J. P., Denis, S., Havenaar, R., and Alric, M. (2004), A dynamic gastrointestinal system for studying the behavior of orally administered drug dosage forms under various physiological conditions, *Pharm. Res.*, 21, 585–591.
34. Minekus, M., Jelier, M., Xiao, J. Z., Kondo, S., Iwatsuki, K., Kokubo, S., Bos, M., Dunnewind, B., and Havenaar, R. (2005), Effect of partially hydrolyzed guar gum (PHGG) on the bioaccessibility of fat and cholesterol, *Biosci., Biotechnol. Biochem.*, 69, 932–938.
35. Larsson, M., Minekus, M., and Havenaar, R. (1997), Estimation of the bioavailability of iron and phosphorus in cereals using a dynamic in vitro gastro-intestinal model, *J. Sci. Food Agric.*, 74, 99–106.
36. Verwei, M., Arkbage, K., Havenaar, R., Van den Berg, H., Witthöft, C., and Schaafsma, G. (2003), Folic acid and 5-methyl-tetrahydrofolate in fortified milk are bioaccessible as determined in a dynamic in vitro gastrointestinal model, *J. Nutr.*, 133, 2377–2383.
37. Krul, C. A. M., Luiten-Schuite, A., Baan, R., Verhagen, H., Mohn, G., Feron, V., and Havenaar, R. (2000), Application of a dynamic in vitro gastrointestinal tract model to study the availability of food mutagens, using heterocyclic aromatic amines as model compounds, *Food Chem. Toxicol.*, 38, 783–792.
38. Souliman, S., Blanquet, S., Beyssac, E., and Cardot, J. M. (2006), A level A in vitro/in vivo correlation in fasted and fed states using different methods: applied to solid immediate release oral dosage form, *Eur. J. Pharm. Sci.*, 27, 72–79.

39. Elashoff, J. D., Reedy, T. J., and Meyer, J. M. (1982), Analysis of gastric emptying data, *Gastroenterology*, 83, 1306–1312.
40. Heading, R. C., Tothill, P., McLoughlin, G. P., and Shearman, D. J. C. (1976), Gastric emptying rate measurement in man. A double isotope scanning technique for simultaneous study of liquid and solid components of a meal, *Gastroenterology*, 71, 45–50.
41. Chung, Y. C., Kim, Y. S., Shadchehr, A., Garrido, A., Macgregor, I. L., and Sleisenger, M. H. (1979), Protein digestion and absorption in human small intestine, *Gastroenterology*, 76, 1415–1421.
42. Castelli, F., Uccella, N., Trombetta, D., and Saija, A. (1999), Differences between coumaric and cinnamic acids in membrane permeation as evidenced by time-dependent calorimetry, *J. Agric. Food Chem.*, 47, 991–995.
43. Zarate, G., Chaia, A. P., Gonzalez, S., and Oliver, G. (2000), Viability and beta-galactosidase activity of diary propionibacteria subjected to digestion by artificial intestinal fluids, *J. Food Protection*, 63, 1214–1221.
44. Zsebo, K. M., Lu, H. S., Fieschko, J. C., Goldstein, L., Davis, J., Duker, K., Suggs, S. V., Lai, P. H., and Bitter, G. A. (1986), Protein secretion from *Saccharomyces cerevisiae* directed by the prepro-a-factor leader region, *J. Biolog. Chem.*, 261, 5858–5865.
45. Chen, C. M., Cheng, W. T. K., Chang, Y. C., Chang, T. J., and Chen H. L. (2000), Growth enhancement of fowls by dietary administration of recombinant yeast cultures containing enriched growth hormones, *Life Sci.*, 67, 2102–2115.
46. Oozeer, R., Goupil-Feuillerat, N., Alpert, C. A., Vande Guchte, M., Anba, J., Mengaud, J., and Corthier, G. (2002), *Lactobacillus casei* is able to survive and initiate protein synthesis during its transit in the digestive tract of human flora-associated mice, *Appl. Environ. Microbiol.*, 68, 3570–3574.
47. Steidler, L., Neirynck, S., Huyghebaert, N., Snoeck, V., Vermeire, A., Goddeeris, B., Cox, E., Remon, J. P., and Remaut, E. (2003), Biological containment of genetically modified *Lactococcus lactis* for intestinal delivery of human interleukin-10, *Nat. Biotechnol.*, 21, 785–789.
48. Harms, H. K., Bertele-Harms, R. M., and Bruer-Kleis, D. (1987), Enzyme-substitution therapy with the yeast *Saccharomyces cerevisiae* in congenital sucrase-isomaltase deficiency, *N. Engl. J. Med.*, 316, 1306–1309.
49. Klein, S. M., Elmer, G. W., McFarland, L. V., Surawicz, C. M., and Levy, R. H. (1993), Recovery and elimination of the biotherapeutic agent, *Saccharomyces cerevisiae*, in healthy human volunteers, *Pharm. Res.*, 10, 1615–1619.
50. Blehaut, H., Massot, J., Elmer, G. W., and Levy, R. (1989), Disposition kinetics of *Saccharomyces boulardii* in man and rat, *Biopharm. Drug Dispos.*, 10, 353–364.
51. Berny, J. F., and Hennebert, G. L. (1991), Viability and stability of yeast cells and filamentous fungus spores during freeze-drying: Effects of protectants and cooling rates, *Mycologia*, 83, 805–815.
52. Lodato, P., Segovia de Huergo, M., and Buera, M. P. (1999), Viability and thermal stability of a strain of *Saccharomyces cerevisiae* freeze-dried in different sugars and polymer matrices, *Appl. Microbiol. Biotechnol.*, 52, 215–220.
53. Diniz-Mendes, L., Bernardes, E., De Araujo, P. S., Panek, A. D., and Paschoalin, V. M. F. (1999), Preservation of frozen yeast cells by trehalose, *Biotechnol. Bioeng.*, 65, 572–578.
54. Abadias, M., Benabarre, A., Teixido, N., Usall, J., and Vias I. (2001), Effect of freeze-drying and protectants on viability of the biocontrol yeast Candida sake, *Int. J. Food Microbiol.*, 65, 173–182.
55. Blanquet, S., Garrait, G., Beyssac, E., Perrier, C., Denis, S., Hébrard, G, and Alric, M. (2005), Effects of cryoprotectants on the viability and activity of freeze-dried recombi-

nant yeasts as novel oral drug delivery systems assessed by an artificial digestive system, *Eur. J. Pharm. Biopharm.*, 61, 32–39.

56. Aquilera, J. M., and Karel, M. (1997), Preservation of biological materials under desiccation, *Crit. Rev. Food Sci. Nutr.*, 37, 287–309.
57. Beker, M. J., and Rapoport, A. I. (1987), *Advances in Biochemical Engineering/Biotechnology*, Springer, Berlin, pp. 128–171.
58. Minekus, M., Smeets-Peter, M., Bernalier, A., Marol-BonnIn, S., Havenaar, R., Marteau, P., Alric, M., Fonty, G., and Huis in't Veld, J. H. J. (1999), A computer-controlled system to simulate conditions of the large intestine with peristaltic mixing, water absorption and absorption of fermentation products, *Appl. Microbiol. Biotechnol.*, 53, 108–114.
59. Yadav, B. S., Rani, U., Dhamija, S. S., Nigam, P., and Singh, D. (1996), Process optimization for continuous ethanol fermentation by alginate-immobilized cells of *Saccharomyces cerevisiae* HAU-1, *J. Basic Microbiol.*, 36, 205–210.
60. Tsen, J. H., Lin, Y. P., and King, V. A. E. (2004), Fermentation of Banana media by using κ-carrageenan immobilized *Lactobacillus acidophilus*, *Int. J. Food Microbiol.*, 91, 215–222.
61. Lee, K. Y., and Heo, T. R. (2000), Survival of *Bifidobacterium longum* immobilized in calcium alginate beads in simulated gastric juices and bile salt solution, *Appl. Environ. Microbiol.*, 66, 869–873.
62. Chandramouli, V., Kailasapathy, K., Peiris, P., and Jones, M. (2004), An improved method of microencapsulation and its evaluation to protect *Lactobacillus* spp. in simulated gastric conditions, *J. Microbiol. Methods*, 56, 27–35.
63. Lamas, C. M., Bregni, C., D'Aquino, M., Degrossi, J., and Firenstein, R. (2001), Calcium alginate microspheres of Bacillus subtilis, *Drug Dev. Ind. Pharm.*, 27, 825–829.
64. Beaulieu, L., Savoie, L., Paquin, P., and Subirade, M. (2002), Elaboration and characterization of whey protein beads by an emulsification/cold gelation process: Application for the protection of retinol, *Biomacromolecules*, 3, 239–248.
65. Guérin, D., Vuillemard, J. C., and Subirade, M. (2003), Protection of bifidobacteria encapsulated in polysaccharide-protein gel beads against gastric juice and bile, *J. Food Prot.*, 66, 2076–2084.
66. Hébrard, G., Blanquet, S., Beyssac, E., Remondetto, G., Subirade, M., and Alric, M. (2006), Use of whey protein beads as a new carrier system for recombinant yeasts in human digestive tract. *J. Biotechnology*, 127, 151–160.
67. Hongsprabhas, P., and Barbut, S. (1998), $Ca^{2+}$-induced cold gelation of whey protein isolate: Effect of two-stage gelation, *Food Res. Int.*, 30, 523–527.
68. Yvon, M., Beucher, S., Scanff, P., Thirouin, S., and Pelissier, J. P. (1992), In vitro simulation of gastric digestion of milk proteins: Comparison between in vitro and in vivo data, *J. Agric. Food Chem.*, 40, 239–244.
69. Remondetto, G. E., and Subirade, M. (2003), Molecular mechanisms of $Fe^{2+}$-induced-β-lactoglobuline cold gelation, *Biopolymers*, 69, 461–469.
70. Bienaimé, C., Barbotin, J. N., and Nava-Saucedo, J. E. (2003), How to build an adapted and bioactive cell microenvironment? A chemical interaction study of the structure of Ca-alginate matrices and their repercussion on confined cells, *J. Biomed. Mater. Res., A*, 67, 376–388.
71. Rossi-Alva, J. C., and Miguez Rocha-Leaõ, M. H. (2003), A strategic study using mutant-strain entrapment in calcium alginate for the production of *Saccharomyces cerevisiae* cells with high invertase activity, *Biotechnol. Appl. Biochem.*, 38, 43–51.
72. Molin, S. (1993), Environmental potential of suicide genes, *Curr. Opin. Biotechnol.*, 4, 299–305.

73. Diaz, E., Munthali, M., De Lorenzo, V., and Timmis, K. N. (1994), Universal barrier to lateral spread of specific genes among microorganisms, *Mol. Microbiol.*, 13, 855–861.
74. Kaplan, D. L., Mello, C., Sano, T., Cantor, C., and Smith, C. (1999), Steptavidin-based containment systems for genetically engineered microorganisms, *Biomol. Eng.*, 16, 135–140.
75. Hols, P., Defrenne, C., Ferain, T., Derzelle, S., Delplace, B., and Delcour, J. (1997), The alanine racemase gene is essential for growth of *Lactobacillus plantarum*, *J. Bacteriol.*, 179, 3804–3807.
76. Fu, X., and Xu, J. G. (2000), Development of a chromosome-plasmid balanced lethal system for *Lactobacillus acidophilus* with thyA gene as selective marker, *Microbiol. Immunol.*, 44, 551–556.

# NASAL DELIVERY OF PEPTIDE AND NONPEPTIDE DRUGS

Chandan Thomas and Fakhrul Ahsan
*Texas Tech University Health Sciences Center, Amarillo, Texas*

**Contents**

1 Introduction
2 Nasal Anatomy and Physiology
  2.1 Structure of Nasal and Olfactory Mucosa
  2.2 Nasal Vasculature
  2.3 Enzymes and pH
3 Factors Influencing Nasal Drug Absorption
  3.1 Physiological Factors
  3.2 Nasal Mucociliary Clearance
  3.3 Pathological Condition of Nose
  3.4 Dose Volume and Site of Deposition
  3.5 Physicochemical Properties of Drugs
  3.6 Type of Delivery Device
4 Animal Models for Nasal Absorption Studies
5 Enhancement of Intranasal Drug Absorption
6 Nasal Delivery of Peptide and High-Molecular Weight Drugs
  6.1 Insulin
  6.2 Calcitonin
  6.3 Low-Molecular-Weight Heparins
  6.4 Azetirelin
  6.5 Growth Hormones
7 Nasal Delivery of Nonpeptide Molecules
  7.1 Morphine
  7.2 Benzodiazepines
  7.3 Buprenorphine
  7.4 Hydralazine
  7.5 Nitroglycerin
  7.6 Propranolol and Other β-Adrenergic Blocking Agents
  7.7 Sex hormones

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

7.8 17β-Estradiol ($E_2$)
7.9 Testosterone
8 Nose: Option for Delivery of Drugs to Central Nervous System
9 Nasal Delivery of Vaccines
9.1 Nasal Vaccines: Ideal Noninvasive Route
9.2 Immunity after Intranasal Immunization
9.3 Need for Adjuvants
References

## 1 INTRODUCTION

The delivery of drugs via the nasal route has been practiced since ancient times; for example, psychotropic and hallucinogenic agents have been used as snuff in many parts of the world for hundreds and possibly thousands of years. More recently, especially over the past two decades, intranasal drug delivery has shown great promise in various fields of medical practice. At present, a number of conditions (e.g., rhinitis, migraine, nasal congestion, and osteoporosis) are being treated successfully with nasal formulations. In addition, an ever-increasing number of nasally delivered, systemically acting drugs are in the pipeline. Recently this form of therapy received encouragement with the approval of FluMist (MedImmuneVaccines, Gaithersburg, MD), an intranasal vaccine against *Haemophilus influenzae,* the influenza virus. This vaccine is the first to be given by the nasal route as a mist rather than by injection. With its approval, many of the pharmaceutical companies, including some giants of the pharmaceutical industry, are increasingly looking toward the area of nasal drug delivery. The market for such therapy in 2005 is reported to have reached $2.4 billion. With 16 of the 20 major pharmaceutical companies conducting active programs in this area, the field of nasal drug delivery is expected to grow at an estimated 33% annually [1]. As reported by Koch [2], the U.S. drug delivery market in 2005 was somewhere in excess of $50 billion and has been predicted to be around $67 billion by the year 2009, whereas the nasal drug delivery market is expected to be valued at $9 billion by 2008 [1–3]. Nasal administration—as compared with injection and oral administration—is more feasible and convenient, especially in view of the rising number of peptide and protein therapeutics that are rapidly being developed. In particular, the possibility of delivering drugs to the brain by the nasal route is eliciting increased interest, especially due to the possibility of accessing or targeting the local receptors and also of circumventing the blood–brain barrier.

Drug delivery via the nasal route offers a number of advantages, the most important of which is the possibility of needle-free treatment. It also means that—in addition to the newly developed peptide- and protein-based drugs—this method is also suitable for a wide variety and perhaps most of the drugs that are currently in use. However, it is not only convenience that sets nasal drug delivery apart: This method also provides a rapid onset of action and high bioavailability.

Because of its rich vasculature and highly permeable structure, the nasal route can be used as an alternative to parenteral routes of delivery. It circumvents hepatic

first-pass metabolism and gut-wall enzyme-mediated degradation. It is also easily accessible for self-administration without the help of a health professional and there are no associated needle-stick hazards. Other advantages of nasal drug delivery systems include a rapid onset of action, reduced risk of overdose, and improved patient compliance. However, there are also several disadvantages, including the impermeability of the nasal mucosa to lipophilic and high-molecular-weight drugs, mucotoxicity associated with long-term use of some formulations, the requirement for an expensive delivery device, and possible dose inaccuracy.

In order to understand the delivery and absorption of drugs by the nasal route and appreciate the factors that may affect it, one must begin with a clear picture of the anatomy and physiology of the nose.

## 2 NASAL ANATOMY AND PHYSIOLOGY

### 2.1 Structure of Nasal and Olfactory Mucosa

The human nasal passage is about 12 cm long and runs from the nostrils to the nasopharynx (Figure 1). The nasal cavity is divided into right and left halves by a midline septum, or cartilaginous wall, that extends posteriorly into the nasopharynx (E). Each half of the nasal cavity consists of three well-separated regions: (A) the vestibule, (B) the olfactory region, and (C) the respiratory region. The vestibule is the most anterior part of the nasal cavity, which opens to the face through the

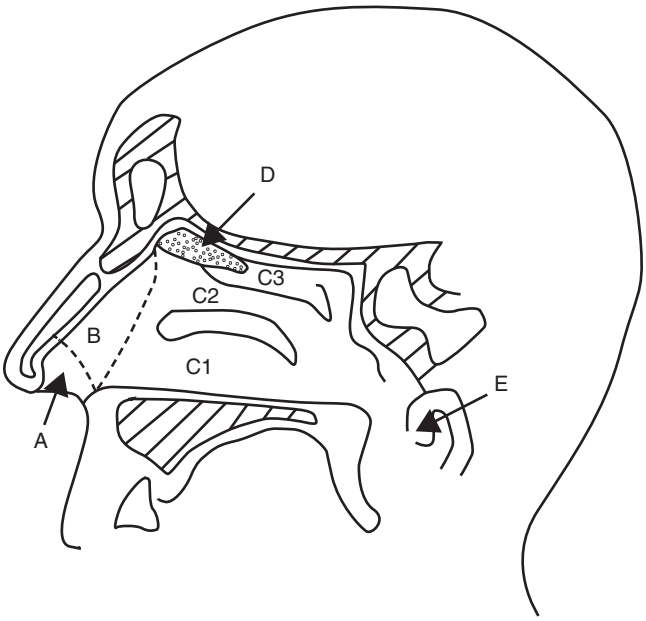


**FIGURE 1** Sagittal section of nasal cavity showing nasal vestibule (A), atrium (B), respiratory area and inferior turbinate (C1), middle turbinate (C2) and superior turbinate (C3), olfactory region (D), and nasopharynx (E). (Reproduced from ref. 5 with permission of Pharmaceutical Press.)

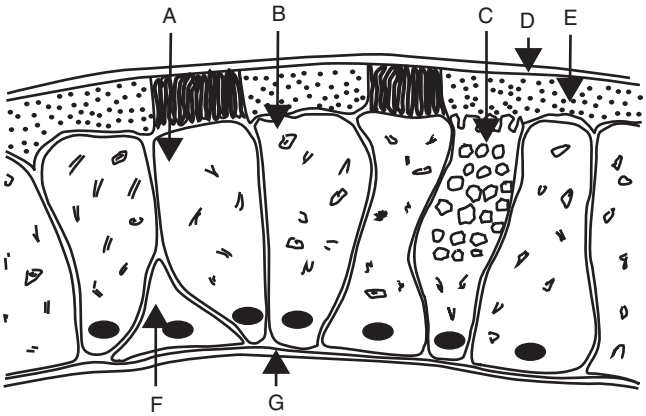


**FIGURE 2** Cell types of nasal epithelium showing ciliated cell (A), nonciliated cell (B), goblet cell (C), gel mucous layer (D), sol layer (E), basal cell (F), and basement membrane (G). (Reproduced from ref. 5 with permission of Pharmaceutical Press.)

nostrils. The respiratory region and the turbinates make up most of the nasal cavity; this region has lateral walls that divide it into three chambers, including the superior nasal turbinate (C3) at the top, the middle nasal turbinate (C2) below, and the inferior nasal turbinate (C1) at the bottom (Figure 1). The total volume of nasal cavity is 15 mL; its surface area is 150 $cm^2$ [4, 5]. Air generally travels into the nose in an approximately parabolic pattern from the vestibule to the nasopharynx. The hard palatine bone and soft palate form the floor of the nose and the roof of the mouth. The human nasal epithelial surface is covered mainly by stratified squamous, olfactory, and respiratory epithelia. The stratified epithelium lies in the anterior part of the nose and becomes pesudostratified columnar epithelium at the posterior part, which constitutes the respiratory epithelium (Figure 2). The olfactory epithelium covers the olfactory region in the upper part of the nasal cavity. The nasal epithelial surface is covered by a continuous layer of mucus secreted by various mucosal and submucosal glands. The mucous layer comprises the gel layer (D) and sol layer (E) as shown in Figure 2. The sol layer is a low-viscosity fluid surrounding the cilia, and the viscous gel layer covers the tips of the cilia—fine, hairlike structures that move in an organized fashion to ease the flow of mucus across the epithelial surface. There are also microvilli, and every ciliated cell carries some 100 cilia. Each ciliated and nonciliated cell has about 300 microvilli. Figure 2 also shows both nonciliated cells and basal cells [4–6].

The olfactory mucosa is discussed further on in relation to the delivery of drugs via the nose to the brain.

## 2.2 Nasal Vasculature

The nasal surface is supplied with a dense network of blood vessels by the external and internal carotid arteries. Blood from the anterior part of the nose is drained through the facial vein, but the nose's main blood supply drains through the sphenopalatine foramen into the pterygoid plexus or via the superior ophthalmic vein

[6]. The nasal blood vessels can be greatly dilated with blood to facilitate warming and humidification of inspired air in response to prevailing conditions. Nasal blood flow is very sensitive to a variety of agents applied topically and systemically.

### 2.3 Enzymes and pH

Nasal secretions contain a mixture of secretory materials from the goblet cells, nasal glands, and lacrimal glands. The main constituents of nasal secretions are water, with 2–3% mucin and 1–2% electrolytes. Nasal secretions also contain several enzymes, including lysozyme, cytochrome P450–dependent monooxygenases, steroid hydroxylases, proteases such as neutral endopeptidase, leucine aminopeptidase, aminopeptidase peroxidase, carboxypeptidase N, and protease inhibitors [7]. However, because most studies of peptide and protein degradation are carried out in homogenates of nasal tissue, the peptides and proteins can be exposed to both intracellular and extracellular enzymes; therefore, the data regarding peptide stability must be interpreted with caution. The normal pH of nasal secretions in adults ranges from about 5.5 to 6.5; in young children, it ranges from 5.0 to 6.7. The nasal pH can vary depending on pathological conditions such as allergic rhinitis and environmental conditions such as cold and heat [8].

## 3 FACTORS INFLUENCING NASAL DRUG ABSORPTION

The nasal absorption of drugs is influenced by a multitude of factors, including nasal physiology, nasal pathology, physicochemical properties of drugs, dosage forms, and delivery method. The presence of pathological conditions such as allergic rhinitis and the common cold further complicates nasal drug delivery. Moreover, the intimate contact between the nasal mucosa and the atmosphere leads to variability in absorption with changes in temperature and humidity.

### 3.1 Physiological Factors

***Nasal Blood Flow*** The nasal vasculature differs from the tracheobronchial tree due to the presence of (a) venous sinusoids, (b) arteriovenous anastosomes, and (c) the nasal vasculature, which shows cyclical changes of congestion, hence giving rise to the nasal cycle. In the nasal vasculature, the arterioles lack an internal elastic membrane, making the endothelial basement membrane continuous with the basement membrane of the smooth muscle cells. Also present are the fenestrated type of capillaries lying just below the surface epithelium and surrounding the glands. Because of this, the capillaries facilitate rapid movement of fluid through the vascular wall, allowing water to escape into the airway lumen. The conditioning of the inhaled air is greatly influenced by the nasal blood vessels. In essence, air is heated and humidified by the flow of nasal blood in the opposite direction to the incoming airflow. The nasal blood flow also controls the size of the nasal passage's lumen. Changes in ambient temperature and humidity, nasal administration of vasoactive drugs, nasal trauma, and compression of large veins in the neck may adversely affect blood flow in the nose [9, 10]. Other factors such as mood changes, hyperventilation, and even exercise can have an effect on the nasal blood flow and hence the nasal

passages [10–12]. In conclusion, any change in blood flow can alter the absorption of a nasally administered drug.

***Nasal Enzymes*** The enzymes present in the nasal cavity may be involved in the extensive enzymatic degradation of drugs administered nasally. The presence of the enzymes in the nasal epithelium acts as a defensive mechanism or a barrier against the entry of xenobiotics [9]. Examples include nasal decongestants, essences, anesthetics, nicotine, and cocaine, which are metabolized by the nose's P450-dependent monooxygenase system [13, 14]. The nasal mucosa also includes oxidative phase I and conjugative phase II enzymes. The phase I enzymes include aldehyde dehydrogenase, carboxyl esterase, and carbonic anhydrases; phase II enzymes include glucuronyl, sulfate, and glutathione transferases. 17β-Estradiol has shown significantly more conjugation when administered via the nasal route as compared to the intravenous route. A variety of other drugs have been shown to be metabolized by nasal enzymes, including progesterone, testosterone, and insulin [15, 16].

### 3.2 Nasal Mucociliary Clearance

Nasal mucociliary clearance is the transport of the mucous layer covering the nasal epithelium toward the nasopharynx by ciliary beating for its eventual discharge into the gastrointestinal tract. Nasal mucociliary clearance plays a very important role in the upper respiratory tract in preventing various noxious agents such as allergens, bacteria, viruses, and toxins from reaching the lungs. The ciliated cells of the nasal mucosa drive the movement of the mucus, and hence the physiological control of the ciliated cells and the rheological properties of the mucus determine the efficiency of the nasal mucociliary clearance system. In humans the normal mucociliary transit time is reported to be 12–15 min, and transit times of 30 min or more are likely to be an indication of impaired mucociliary clearance. Impairment of mucociliary clearance has been associated with longer contact times of various noxious agents as well as drugs with the nasal mucosa. On the other hand, increases in the mucociliary clearance rate decrease the contact between drug and the epithelium and consequently reduce drug absorption. Therefore nasal drug absorption can be augmented by the use of bioadhesive polymers or microspheres or by increasing the viscosity of the drug formulation [5, 17]. Hydroxypropyl methylcellulose, polyacrylic acid, and hyaluronan all enhance absorption by increasing nasal residence time [18–20]. The effect of mucociliary clearance may vary depending on the site of drug deposition. Ciliated epithelium is present in the middle and posterior parts of the turbinates, but there is little or no ciliary epithelium in the anterior regions of the nasal cavity [21, 22]. This is one of the reasons why a drug deposited in the posterior part of the nose is washed away more quickly than a drug deposited in the anterior site of the nasal cavity [23].

### 3.3 Pathological Condition of Nose

As mentioned earlier, the presence of nasal pathological conditions—such as allergic rhinitis, polyposis, and common colds—influences nasal drug absorption to a great extent. The majority of pathological conditions of the nose show bleeding, excessive secretion of mucus, nasal blockage, and crusting. It has been reported that

excessive nasal secretion may wash away a nasally administered drug before it can be absorbed [24]. Nasal drug absorption and distribution are also influenced by the presence of nasal polyps and blockage. Several studies however have suggested that the presence of nasal pathological conditions do not affect nasally administered peptide drugs. For example, buserelin and desmopressin absorption studies have shown similar nasal absorption profiles in normal subjects and in those suffering from colds or rhinitis [25, 26].

### 3.4 Dose Volume and Site of Deposition

A dose of 25–200 μL per nostril is what can be maximally accommodated by the human nose. A dose higher than this will be drained off and hence shows lower absorption. Some studies have reported that a 100-μL volume resulted in a larger deposition area. Hence, taking into account the volume of administration becomes very important for manufacturers of nasal drug delivery systems. The site of deposition of the nasal formulation may also affect the nasal absorption of drugs since the anterior part of the nose provides greater contact between the nasal epithelium and drug, but the mucociliary clearance mechanism of the posterior tends to remove drug more rapidly [27]. It has been found that the permeability of the posterior area is greater than that of the anterior portion, and hence, based on the formulation, drugs may be administered in either the anterior or posterior parts of the nose. The nasal adapter's spray-cone angle defines the width of the nasal spray pattern and thus plays an important role in determining the site of deposition in the nasal cavity. Changes in the cone angle of the adapter from 60° to 35° or 30° can produce a larger and more posterior deposition and therefore higher drug deposition in the ciliated area [10].

### 3.5 Physicochemical Properties of Drugs

The absorption of a drug across the nasal mucosa is a function of its physicochemical properties, such as molecular weight, lipophilicity, and water solubility, as seen with most of the mucosal routes of delivery. The majority of studies on the effects of drug lipophilicity on nasal absorption are rather conflicting. The effect of lipophilicity on the nasal absorption of barbituric acids has been investigated. It was found that drug absorption through the nasal mucosa increases with an increase in the partition coefficient. Interestingly, there was only a fourfold increase in absorption between phenobarbital and barbital despite the fact that the partition coefficient of phenobarbital was 40-fold higher than that of barbital [28]. Similarly, increases in nasal absorption have been seen for hydrocortisone, testosterone, and progesterone with increases in the partition coefficient. However, a hyperbolic—rather than a linear—relationship was observed between the in vivo nasal bioavailability of a series of progesterone derivatives and their octanol–water partition coefficients [29]. In contrast to this, Kimura et al. [30] showed that, for a series of quaternary ammonium compounds structurally related to tetraethylammonium chloride, nasal absorption was inversely related to the partition coefficient. All these studies suggest that a drug's lipophilicity may not be an appropriate indicator of the extent of its nasal absorption. Besides the drug lipophilicity another important factor most studied for its influence on nasal absorption is the aqueous solubility of a drug. This is because

the nasal mucosa is constantly kept moist by the nasal secretions and is well perfused with blood vessels. In addition to limiting drug absorption, it also is a limiting factor for the formulation of the drug. Further it is also important to understand the relation between the saturation solubility of the drug and drug absorption. However, the influence of drug solubility on absorption of drugs via the nasal route has not been significantly explored and needs much more attention. The relative effectiveness of nasal atropine and hyoscine was compared by Tonndorf et al. [31] by measuring each drug's capacity to arrest salivary secretion; they found that 0.65 mg of hyoscine, 40 times more soluble in water than atropine, is equivalent to 2 mg of atropine. Some authors have suggested that the aqueous pores in the nasal mucosa play a major role in absorption of hydrophilic drugs [32]. When nasal formulations are administered as inhaled powders or suspensions, drug dissolution rate becomes an important factor. Formulations deposited in the nostrils require proper dissolution for better absorption. Nasal mucociliary clearance may remove the drug if it remains undissolved in the nostrils. The size and shape of a drug molecule can affect its nasal absorption. Molecules with an average molecular weight less than 1000 Da are better absorbed nasally than higher molecular weight drugs whereas linear molecules are less effectively absorbed than compact ones [33]. The nasal and oral absorption of polyethylene glycol 600, 1000, and 2000 have been studied by Donovan et al. [34] in relation to molecular weight; they found that the greater the molecular weight, the less effective the absorption. This pattern of absorption was seen in the case of both the nasal and oral routes. The effect of water-soluble compounds such as 4-oxo-4*H*-1-benzopyran-2-carboxylic acid, *p*-aminohippuric acid, sodium cromoglycate, inulin, and dextran showing different molecular weights on the nasal absorption was studied by Fisher et al. [35]. A 43-fold decrease in the nasal absorption of the these compounds was observed with a 368-fold increase in the molecular weight [9]. Similarly, studies with 13 di-iodo-L-tyrosine-labeled dextran showed an inverse relationship between the percentage absorbed after nasal administration and the weight of the molecule; in fact, a 36-fold increase in molecular weight produced an 88-fold reduction in nasal absorption [36]. The effect of molecular weight on the nasal absorption of fluorescein isothiocyanate and diethylaminoethyl dextrans has been studied by Maitani et al. [37]. It was found that an inverse relationship between absorption and molecular weight existed for these compounds. However, since the nasal absorption of these compounds was low, enhancers were used; hence it is difficult to rule out the influence of the enhancers used on the extent of absorption obtained. The absorption of a drug after nasal administration is also influenced by the pH of a drug formulation as well as that of the nasal cavity—along with the $pK_a$ of the drug substance. Biological membranes form a major barrier to the transport of drugs into the bloodstream. There are a number of transport mechanisms by which drugs are transported across the biological membranes. These include transcellular, paracellular, and carrier-mediated transport mechanisms. The most important factors that influence the above-mentioned mechanisms are the pH, $pK_a$, and partition coefficient of the drug. The pH of a nasal formulation should be in the range of 4.5–6.5 in order to minimize nasal irritation. However, the drug's $pK_a$ must also be taken into account so as to maximize the drug's concentration in un-ionized form. The effect of pH on the nasal absorption of benzoic acid was studied by Hussain et al. [28], who showed that the absorption of benzoic acid is pH dependent.

### 3.6 Type of Delivery Device

Both the type of drug delivery system and the specific type of delivery device can affect drug absorption via the nasal route. The choice of delivery system depends mainly on the physiochemical properties of the drug, its desired site of action, and, more importantly, patient compliance and marketing aspects. The formulations most commonly used in nasal delivery are solutions, suspensions, gels, dry powders, and, most recently, nanoparticulate formulations.

Solutions are most commonly used for intranasal drug delivery. Such solutions may be used when the active ingredient is soluble in water or in some other vehicle approved by the Food and Drug Administration. At present, nasal solutions are available in the form of drops and sprays. Drops are the simplest and the most convenient nasal dosage form, and they are also easy to manufacture. However, their major drawback is that exact dosages cannot be administered with them. Another disadvantage is that they—like most solution-based medications—are vulnerable to microbial contamination; therefore, preservatives must often be added. These, in turn, have further disadvantages, as they may both cause irritation and hamper mucociliary clearance, thus decreasing compliance. Chemical stability is also often an issue with nasal drops.

Since the introduction of metered-dose inhalers, nasal solutions have increasingly been formulated as nasal sprays. Initially, aerosol-based systems containing chlorofluorocarbons were employed; however, the Montreal Protocol put an end to this. Thereafter, mechanical pumps or actuators were employed to deliver nasal formulations as sprays. These devices, using actuators, can precisely deliver as little as 25 μL and as much as 200 μL of a formulation. However, various factors must be considered in formulating the spray; these include viscosity, particle size, and surface tension, all of which may affect the accuracy of the dose administered.

Suspensions may also be used to deliver nasal formulations, though only rarely, since a number of complicating factors (e.g., particle size and morphology) must be considered. Suspensions offer the advantage of increasing residence time in the nasal cavity, thus possibly augmenting nasal bioavailability.

Gels are thickened solutions that may sometimes be used to deliver drugs via the nose, since they offer a number of advantages, such as reducing postnasal drip into the back of the throat and hence reducing the loss of the drug from the nasal cavity, anterior leakage, and the associated irritation. The use of gels is also reported to improve absorption and to mask the irritation associated with some ingredients by the addition of soothing agents and emollients. A vitamin $B_{12}$ (cyanocobalamin) nasal gel, Nascobal (Nastech Pharmaceutical, Kirkland, WA), is available in a metered-dose formulation. Several other drugs, such as insulin, are being studied with a view to formulating them as nasal gels [9, 27, 38]. Although nasal powders are more stable than other formulations, they are rarely used because they tend to irritate the nasal tissue. However, a powder form may be useful when the active ingredient cannot be formulated as a solution or suspension. With the development of refinements in technology, many researchers are exploring the use of nanoparticle-based formulations to deliver drugs nasally. The main advantage of these state-of-the-art formulations is that they ensure increased absorption as well as better compliance. Microsphere- and liposome-based formulations are being increasingly tested. Some of these studies are discussed in Sections 6 and 7.

The type of nasal device employed in delivering a drug formulation plays a major role in the efficacy of the treatment. In general, two types of delivery systems are used: mechanical pumps and pressurized aerosol containers. The properties of the drug to be used influence the selection of the system. The various types of delivery devices are described in the following sections.

***Unit-Dose Containers*** The unit-dose container offers a number of advantages. It is easy and convenient to carry and also does away with the need for preservatives, thus greatly increasing patient compliance. The unit-dose container is more accurate than the multidose container; metered-dose nasal sprays are still more accurate. The volume of drug held by such a container is usually determined by its filling volume, which greatly influences its accuracy. Another type of unit-dose container is based on an actuator and consists of a nasal adapter, or a small chamber that contains a piston [9]. Unit-dose containers are used mainly in emergencies (e.g., to manage pain), although they are not restricted to such use and may also be employed in instances where a single administration is required (e.g., vaccination). An example of a product that employs unit-dose containers is Imitirex (sumatriptan) (GlaxoSmithKline, Research Triangle Park, NC).

***Squeeze Bottles*** This is a smooth plastic bottle with a jet outlet. When the bottle is pressed, a certain volume of its contents is atomized as the air inside the bottle is pushed out. This type of device is vulnerable to contamination as ambient air rushes into it following the release of pressure. The squeeze bottle is used mainly to deliver decongestants and not vasoconstrictors, as in the latter case the dose administered would be difficult to control [9].

***Metered-Dose Nasal Pump Sprays*** Metered-dose pumps are the most widely used devices for the delivery of formulations via the nose. A number of commercially available products use this technology. The accuracy of the delivered dose is fairly high and makes it possible to administer dose volumes ranging from 25 to 200 μL. A metered-dose pump is made up of a container as well as a pump, valve, and actuator. The characteristics of the spray delivered will differ depending on the properties of the drug, the precompression mechanism, and the valve and pump selected. The length of the actuator is an important factor determining the deposition of the drug in the nose; the collection of residual drops on its tip will affect correct dosing.

***Airless and Preservative-Free Sprays*** There are now pumps that prevent the entry of air into a dispensing device after use, thus increasing the stability of numerous compounds that are vulnerable to oxidation; this innovation has also minimized the use of preservatives. The working principle of these pumps is operation against a vacuum using a collapsible bag and a sliding piston. This is possible because the vacuum created when a dose is dispensed is accompanied by a reduction in the volume of the container, either by deforming the container itself or by dragging the sliding piston out of it. These maneuvers have no influence on the system's efficiency and, in fact, provide an advantage in that the container can be held in any position without significantly compromising the accuracy of the dose dispensed. This system is particularly suitable for use with children and bedridden hospitalized patients.

Whenever systems without preservatives are used for single or double doses, they pose little risk of contamination. However, multidose systems are generally used over a longer period of time; therefore, unless their formulations also include preservatives, the chances of contamination are increased. Scientists from Erich Pfeiffer and Qualis Laboratorium (both in Germany) have reviewed the latest trends in preservative-free nasal sprays and report that it is possible to prevent microbial contamination via the orifice in two ways. The first is by introducing a chemical additive, such as a bacteriostatic agent, into the nasal actuator so that it comes into contact with both the medication and the environment. However, in the case of an open system, the formulation within the actuator can still be contaminated. The second approach is the use of a mechanism whereby the system is sealed behind the orifice, thus preventing microorganisms gaining access [9, 39].

Some innovative technologies are being developed by a variety of pharmaceutical firms. The following section touches briefly on the latest of these.

***Kurve Technology*** Kurve Technology has developed a unique system of controlled particle dispersion (CPD) by which it is possible to deliver drugs to the entire nasal cavity as well as the olfactory region and the paranasal sinuses. It uses the principle of *vortical* flow, by which inherent airflows of the nasal cavity are disrupted (Figure 3*a* and *b*). Its advantages include optimization of the size and trajectory of droplets, which makes it possible to saturate the nasal cavity. CPD also increases nasal residence time and reduces the deposition of compounds in the lungs and stomach, thus making the treatment more effective and efficient. ViaNase ID (Kurve Technology, Bothell, WA) is a CPD-powered electronic atomizer also developed by Kurve. Its advantages include generation of narrow droplet distribution between 3 and 50 μm and control of the atomization rate (i.e., the rate at which the droplets are generated and how rapidly they exit the device). CPD technology can be used to deliver both solutions and suspensions; currently, work is in progress to apply this principle to dry powders. Testing is also under way for the delivery of small and large molecules as well as peptides and proteins. Finally, CPD technology makes it possible to provide preservative-free packaging; unit-dose ampules; targeted deposition, as mentioned earlier; and monitoring of doses and compliance [40–42].

***OptiNose*** OptiNose AS (Oslo, Norway) has introduced the novel idea of bidirectional intranasal drug delivery, which delivers a drug while the patient exhales and thus is said to prevent lung deposition. It utilizes the concept that exhalation against resistance leads to closure of the soft palate, thus separating the nasal cavity from the mouth as well as cutting off communication between the cranial surface of the soft palate and the posterior margin of the nasal septum. When this occurs, the air can enter one nostril through the sealing nozzle, turn 180°, and finally exit through the other nostril in the reverse direction. This concept is utilized in breath-actuated bidirectional delivery; that is, the air is blown out of the container and the sealing nozzle is used to direct its flow of air into the nose. When this approach was compared with conventional nasal drug delivery in 16 healthy subjects using $^{99m}$Tc-labeled nebulized particles, it was found that bidirectional nasal delivery did, in fact, prevent deposition in the lungs. The single-use device is already developed and undergoing clinical testing for the delivery of a variety of compounds; a multidose liquid reservoir and powder delivery device are also being developed. The technol-

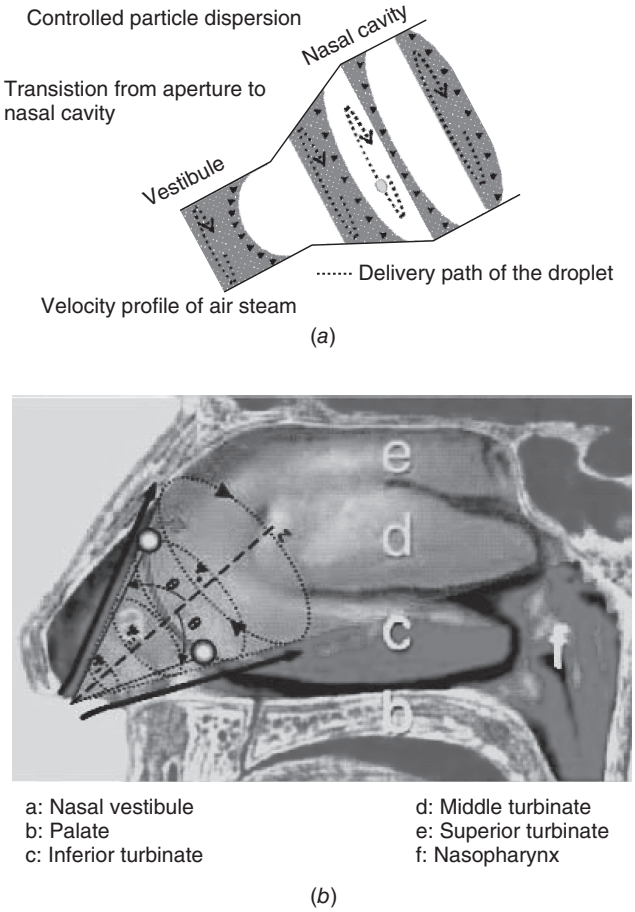


**FIGURE 3** (*a*) Vertical droplet flow created by controlled particle dispersion used in ViaNase ID (Kurve Technology, Bothell, WA). (*b*) Deposition pattern produced by controlled particle dispersion. (Reproduced from ref. 42 with permission from *Drug Delivery Technology.*)

ogy is also being tested in connection with vaccination by the nasal route, and vaccines against diphtheria and influenza have already been shown to improve local and systemic immune responses. Another area being explored is the delivery of drugs to the brain via the nose. A phase I clinical trial of midazolam has been carried out and has shown an onset and level of sedation comparable to that of intravenous administration. The duration of sedation was found to be longer with the OptiNose technology, but the bioavailability of the drug was found to be only 68%, as compared with 100% with intravenous administration. Figures 4*a* and *b* show the Optinose multidose liquid device and multiuse powder device being developed respectively [43, 44].

***DirectHaler*** The DirectHaler (Direct-Haler A/S, Copenhagen, Denmark) nasal delivery device takes advantage of the nasal anatomy and an innovation in device technology in order to improve nasal drug delivery and patient compliance. Similar

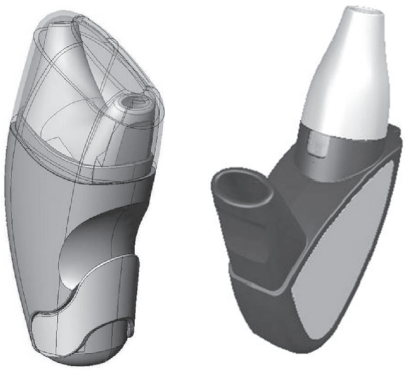


**FIGURE 4** (*a*) Optinose multidose liquid device and (*b*) Optinose multiuse powder device. (Reproduced with permission of Per Gisle Djupesland by personal communication.)

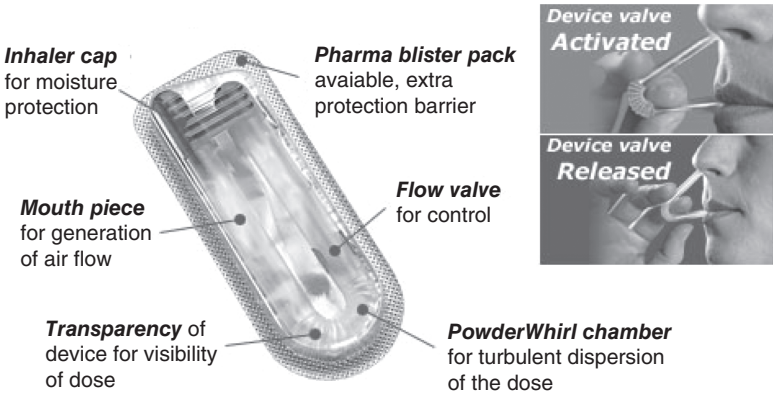


**FIGURE 5** DirectHaler Nasal delivery device. (Reproduced with permission of Troels Keldmann by personal communication.)

to OptiNose, this novel drug delivery device avoids lung deposition. It takes advantage of the fact that, when air is blown out of the mouth against a particular resistance, the oral and nasal cavity airway passage closes on its own. Hence when a patient blows air into the DirectHaler Nasal device, the nasal dry powder dose is delivered into the nostril (Figure 5). The DirectHaler is reported to solve most of the problems that are associated with existing drug delivery devices, including dripping of the liquid dose out of the nostril following its delivery, swallowing of the dose after the administration and hence low absorption, and other problems of contamination associated with the liquid and multidose formulations. The device is quite easy and cost effective to manufacture, fill, and assemble using the latest high-speed technology. The tube of the device is made by using extrusion and roll forming while the device cap is manufactured by injection molding. A modified high-speed capsule-filling machine is used to carry out the powder dose filling. DirectHaler has also been developed for the combination of oral and nasal drug delivery. For pulmonary delivery DirectHaler Pulmonary has been developed. Further, for the

complete targeting of the whole respiratory system, Novel DirectHaler Compliance system has been designed. The device has successfully undergone clinical trials for its use in nasal drug delivery and has shown patient compliance [45].

## 4 ANIMAL MODELS FOR NASAL ABSORPTION STUDIES

Nasal absorption studies can be carried out in mainly two types of animal models: the whole-animal model and the isolated organ perfusion model.

In vivo nasal absorption has been studied in a variety of mammals, including the rat, rabbit, dog, sheep, and monkey. Rats are the most widely used animals for testing intranasal drug delivery. They are easy to handle and simple and inexpensive to maintain. Use of the rat model for the nasal delivery of insulin was first described by Hirai and co-workers more than 25 years ago [46, 47]. The rat has been used in two basic models, in vivo and in situ—the latter also referred to as the ex vivo nasal perfusion model. Both models require anesthesia, though an in vivo model not requiring surgery has also been reported and is discussed later. In early studies of nasal absorption in rats, the animals were anesthetized and the passage of the nasopalatine tract was then sealed with acrylic glue to prevent the drainage of drug solution from the nasal cavity. The trachea was cannulated with a polyethylene tube, with another tube being inserted through the esophagus toward the posterior part of the nasal cavity. The drug solution was then delivered to the nose through either the nasal cavity or the esophageal tubing [8, 47]. Blood samples were then collected from the femoral vein.

However, a less traumatic and much more feasible rat model has recently been proposed by Pillion et al. and other groups [48, 49]. In this model, the rats are anesthetized and kept on their backs; then the drug solution is administered directly into the nose by inserting a pipette about 3–5 mm into each nostril. In some cases the drug is administered into only one nostril so as to prevent blockage. The rats then remain on their backs long enough for the formulation to come into contact with the nasal mucosa. Blood samples are then collected from the tail vein.

In support of the earlier models, it was argued that, by sealing the nasopalatine tract, the drug would necessarily be fully absorbed and transported into the circulation via permeation through the nasal mucosa. However, such nasal absorption with blockage of drainage from the nose to the mouth is not the normal physiological condition. Although the rat model has been used extensively by investigators throughout the world, application of the results of such studies to humans is very limited because of the small body size of these animals and significant interspecies differences. In fact, significant variability between the rat and the human was observed in studies of the bioavailability of insulin administered intranasally [50]. Furthermore, the use of anesthesia has raised concerns because of its potential to confound the test results.

In the rabbit model, drug solution is delivered by spray instillation into a nostril, keeping the rabbit's head in an upright position and allowing the rabbit to breathe normally. Blood samples are then collected from the ear vein. New Zealand White and Japanese White rabbits are most commonly used in such research. One of the advantages of the rabbit model is that the blood volume of these animals is large enough for multiple sampling and pharmacokinetic analysis [51].

Dogs, sheep, and monkeys can be kept conscious during nasal delivery to mimic the human [51]. Sheep, because of their large nostrils and docile nature, serve as excellent models for studies of this kind.

## 5 ENHANCEMENT OF INTRANASAL DRUG ABSORPTION

Lipophilic drugs or compounds have consistently been shown to be completely absorbed across the nasal mucosa; frequently, nasal absorption of these compounds is identical to that obtained with intravenous administration. In some reports, bioavailability after nasal administration reached almost 100% and $T_{max}$ was similar to that obtained with intravenous administration. For example, lipophilic drugs such as propranolol, naloxone, buprenorphine, testosterone, and 17α-ethynylestradiol have been reported to be completely or almost completely absorbed after nasal administration in animal models. However, the same is not true of polar molecules and certain low-molecular-weight drugs and also high-molecular-weight peptides and proteins. In these instances absorption enhancers have been employed, and this is the method most widely used to improve nasal drug absorption [52–54]. Such absorption enhancers belong to a variety of different chemical groups and may have one or multiple ways of enhancing the absorption. Absorption enhancers work by (1) altering the mucous layer by decreasing its viscosity or disrupting it; (2) altering the tight junctions by sequestering extracellular calcium ions, which are reported to be essential in maintaining the integrity of these junctions; (3) inhibiting mucosal enzymatic degradation; (4) reverse-phase micelle formation—in certain cases, reverse-phase micelles may be formed within the cell membranes, thus creating an aqueous pore through which the drug can pass; and (5) altering membrane fluidity, which can be achieved when there is disorder in the membrane phospholipid component or leaching of proteins from the membrane or by a combination of these mechanisms [52–56]. Table 1 lists some of the selected absorption enhancers based on the mechanisms mentioned above.

## 6 NASAL DELIVERY OF PEPTIDE AND HIGH-MOLECULAR-WEIGHT DRUGS

Protein and peptide delivery by means other than injection is currently receiving enormous attention due to the increasing number of biotechnology-based products being developed. There have been numerous attempts to design systems for oral peptide, protein, and gene delivery, but these have unfortunately met with limited success, thus providing an impetus for exploring alternative noninvasive delivery methods. As mentioned earlier, more and more research has been directed to nasal drug delivery because of the numerous advantages it offers. The following section deals with the delivery of peptide and protein drugs.

### 6.1 Insulin

Insulin is produced by the β cells of the islets of Langerhans in the pancreas. It is made up of two peptide chains, which have 21 and 30 amino acid residues,

**TABLE 1 Enhancement of Nasal Absorption**

| Type of Compound (*Absorption Enhancers*) | Examples | Mechanism of Absorption Enhancement |
|---|---|---|
| Bile salts and derivatives | Sodium deoxycholate, sodium glycocholate, sodium taurocholate, sodium taurodihydrofusidate, sodium glycodihydrofusidate | Alteration of tight junctions, membrane disruption, inhibition of mucosal enzymatic degradation, mucolytic activity |
| Chelating agents | Citric acid, EDTA, enamines, *N*-acyl derivatives of collagen, salicylates | Alteration of tight junctions |
| Enzyme Inhibitors | Amastatin, bestatin, camostat mesylate, boroleucine | Enzyme inhibition |
| Fatty acids and derivatives | Acylcarnitines, acylcholine, caprylic acid, capric acid, oleic acid, phospholipids, mono- and diglycerides, sodium laurate | Membrane disruption |
| Surfactants | Sodium lauryl sulfate, saponin, polyoxyethylene-9-lauryl ether, polyoxyethylene-20-lauryl ether, alkylmaltosides such as tetradecylmaltoside, dodecylmaltoside, and decylmaltoside | Membrane disruption |

respectively. These two chains are held together by disulfide linkages between cysteine residues. Insulin is an anabolic and anticatabolic peptide hormone with a molecular weight of about 6000. It was first used in the successful treatment of diabetes mellitus in 1922. Today insulin is widely used in the treatment of both insulin-dependent and non-insulin-dependent diabetes mellitus [57–59]. A nonparenteral formulation of insulin has been approved only recently; before this development, insulin treatment required the use of painful injections. Exubera (Pfizer Labs, New York, NY) is the first insulin and first biotechnology-based medicine for the treatment of a systemic disorder that can be administered without an injection. It was developed by Nektar (San Carlos, CA) and is now a registered trademark of Pfizer (New York, NY). It still poses a few concerns, especially its effect on the lungs of patients with asthma or chronic obstructive pulmonary disease. Because of the problems associated with the parenteral injection of insulin, many diabetic patients flatly refuse to accept insulin therapy. As to subcutaneous injections, these fail to attain a physiological pattern of insulin owing to their adverse pharmacokinetics, and normoglycemia is often not achieved [47]. Of all the routes so far studied apart from delivery via the lungs, the nasal route would appear to be the most advantageous for the delivery of insulin.

The nasal delivery of insulin was demonstrated as early as 1922 by Woodyatt [59]. Since then, numerous studies have focused on this methodology. Some of the early studies included absorption of insulin from the nasal mucosa in human diabetics, the use of an insulin sprayer that contained saponin, and insulin in ethylene glycol or trimethylene glycol applied to the nose in the form of drops or sprays; the last

of these methods demonstrated a significant fall in blood glucose levels in normal rabbits, dogs, and diabetic humans [60–62].

The enhancement of nasal absorption of insulin by hydrophobic bile salts has also been investigated. It was found that minor differences in the number, position, and orientation of the nuclear hydroxyl groups as well as alterations to side-chain conjugation can improve the adjuvant potency of bile salts. Moreover, the absorption of insulin positively correlated with an increase in the hydrophilicity of the steroid nucleus of the bile salts. In the presence of bile salts, nasal absorption of insulin reached peak levels within about 10 min, and some 10–20% of the dose was found to have been absorbed into the circulation. Marked increases in serum insulin levels were seen with sodium deoxycholate, the most lipophilic of the bile salts, whereas the least elevation—as well as least lowering of blood glucose levels—was seen with the most hydrophobic bile salt, sodium ursodeoxycholate [63].

Morimoto et al. [64] studied the nasal absorption of insulin using polyacrylic acid gel. When insulin was formulated with 0.1% w/v polyacrylic acid gel base (pH 6.5), the maximum hypoglycemic effect was seen 30 min following intranasal administration; in 1% w/v gel base, however, it took 1 h to reach the maximum effect. There was no effect of the pH (4.5, 6.5, and 7.5) of 0.1% w/v polyacrylic acid gel on the extent of nasal absorption.

Pillion and his group studied alkylmaltosides differing in chain length for their abilities to lower blood glucose levels when formulated with insulin [65]. Tetradecylmaltoside (TDM) was the most effective agent in producing the hypoglycemic effect, followed by dodecylmaltoside (DDM) and decylmaltoside (DM), all at concentrations of 0.060%. The onset of hypoglycemic action using these nasal drops was seen within 30 min and the maximum effect was obtained within 60–120 min. It was also demonstrated that insulin plus TDM at concentrations as low as 0.03% induced a hypoglycemic effect; however, insulin plus octylmaltoside (OM) failed to produce any hypoglycemic effect even at OM concentrations as high as 0.50%. Dodecylsucrose, which differs from DDM by only one carbohydrate residue, had a similar effect on blood glucose; however, decylsucrose was found to be less potent, and nonglucosides were able to enhance the nasal absorption of insulin only at concentrations ≥0.50% [65].

Insulin formulated with 0.06 or 0.125% hexadecylmaltoside produced a pronounced and rapid dose-dependent decrease in blood glucose levels after nasal administration. The effects of seven different alkylmaltosides were studied, and all the reagents (Figure 6) showed a similar maximal enhancement of insulin uptake when a concentration of 0.125% was employed. The figure demonstrates that TDM showed the greatest effect when concentrations of 0.03 and 0.06% were used.

Similar experiments were carried out using sucrose esters in nasal insulin formulations (Figure 7). It was observed that tetradecanoylsucrose and tridecanoylsucrose were more effective in stimulating insulin absorption as compared with decanoylsucrose and dodecanoylsucrose. But—compared with TDM at concentrations of 0.03%—the sucrose esters were less effective in promoting nasal absorption [66]. Sucrose cocoate (SL-40) is produced by the chemical esterification of coconut oil with sucrose; it has frequently been used in cosmetic and dental preparations as an excipient. When this excipient was formulated with insulin at 0.125, 0.25, and 0.5% concentrations, the associated plasma levels of insulin increased rapidly; whereas there was no enhancement of insulin plasma levels when insulin in saline was admin-

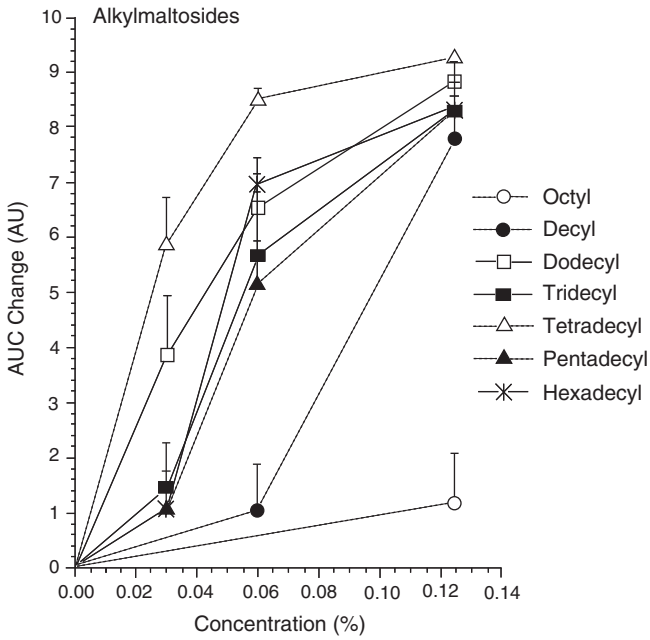


**FIGURE 6** Changes in area under the curve ($AUC_{0-120}$) for blood glucose values in rats that received 2 U insulin in presence of alkylmaltosides. Data represent mean change in $AUC_{0-120}$ in arbitrary units (AU) ± standard error of the mean (SEM) compared with rats that received insulin formulated without alkylmaltoside ($n = 3, 4$). (Reproduced from ref. 66 with permission of John Wiley & Sons.)

istered. The levels of insulin in plasma increased from a baseline value of 10 μU/mL to a level of 200 μU/mL after the nasal administration of sucrose cocoate-containing formulations; the $T_{max}$ was found to be 10 min (Figure 8) [67].

Chitosan is a linear cationic polysaccharide made up of copolymers of glucosamine and *N*-acetylglucosamine. It is commercially obtained by alkaline deacetylation of chitin [53, 68] and has been used for the nasal delivery of a number of drugs. The usefulness of chitosan in the enhancement of nasal absorption was reported first by Illum [69]. Later, Illum and his group also published experimental results indicating that solution formulations with 0.5% chitosan promoted the absorption of nasally administered insulin in rat and sheep [70].

The use of chitosan nanoparticles in the enhancement of the nasal absorption of insulin has also been investigated in rabbits. Chitosan nanoparticles were prepared by ionotropic gelation of chitosan and pentasodium tripolyphosphate (TPP). Two types of chitosan were used in the hydrochloride salt form (Seacure® 210 Cl and Protasan® 110 Cl). Insulin loaded in chitosan 210 Cl produced a significant increase in systemic absorption and also the greatest decline in the level of blood glucose, as much as 60% of basal levels; this result was found to be significantly different from that obtained with the insulin control solution as well as the insulin–chitosan solution [71]. A novel chitosan nanoparticle formulation was prepared by again employing ionotropic gelation of TPP and chitosan glutamate (A1), and postloaded insulin–chitosan nanoparticles (A2) were also prepared. Both these

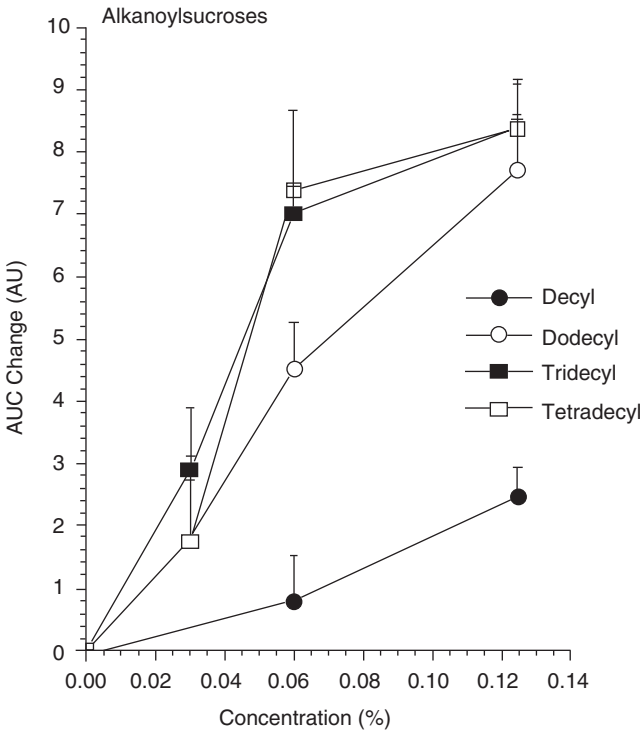


**FIGURE 7** Changes in area under the curve ($AUC_{0–120}$) for blood glucose values in rats that received 2 U insulin in presence of alkanoylsucroses. Data represent mean change in $AUC_{0–120}$ in arbitrary units (AU) ± SEM compared to rats that received insulin formulated without alkanoylsucroses ($n$ = 3, 4). (Reproduced from ref. 66 with permission of John Wiley & Sons.)

novel formulations were tested in rat and sheep models and also compared with the insulin–chitosan solution formulation (A3) and subcutaneous injection of insulin. In the rat model, it was found that A3 performed better than either A1 or A2. A1 showed a minimum blood glucose concentration ($C_{min}$) and time to reach $C_{min}$ ($T_{min}$) of 40% and 90 min, respectively. The $F_{dyn}$—calculated as individual area over the curve ($AOC_{IN\ or\ SC} \times dose_{SC}$/mean $AOC_{SC} \times dose_{IN\ or\ SC}$)×100—was found to be around 48%, whereas the insulin–chitosan nanoparticles showed $F_{dyn}$ values of 38 and 37 for the A1 and A2 formulations, respectively.

Since the concentrations of insulin to be administered in the sheep model would have been large, the insulin-loaded chitosan nanoparticles were not investigated in that model. However, the pharmacodynamics and pharmacokinetics of various insulin–chitosan preparations were compared with postloaded insulin–chitosan nanoparticles. It was found that chitosan solution and chitosan powder formulations were far better, with the chitosan powder formulation showing a bioavailability of 17% as against 1.3 and 3.6% for the chitosan nanoparticles and chitosan solution [72]. The effects of the concentration and osmolarity of chitosan and the presence of absorption enhancers in the chitosan solution on the permeation of insulin across the rabbit nasal mucosa in vitro and in vivo were investigated, and the same

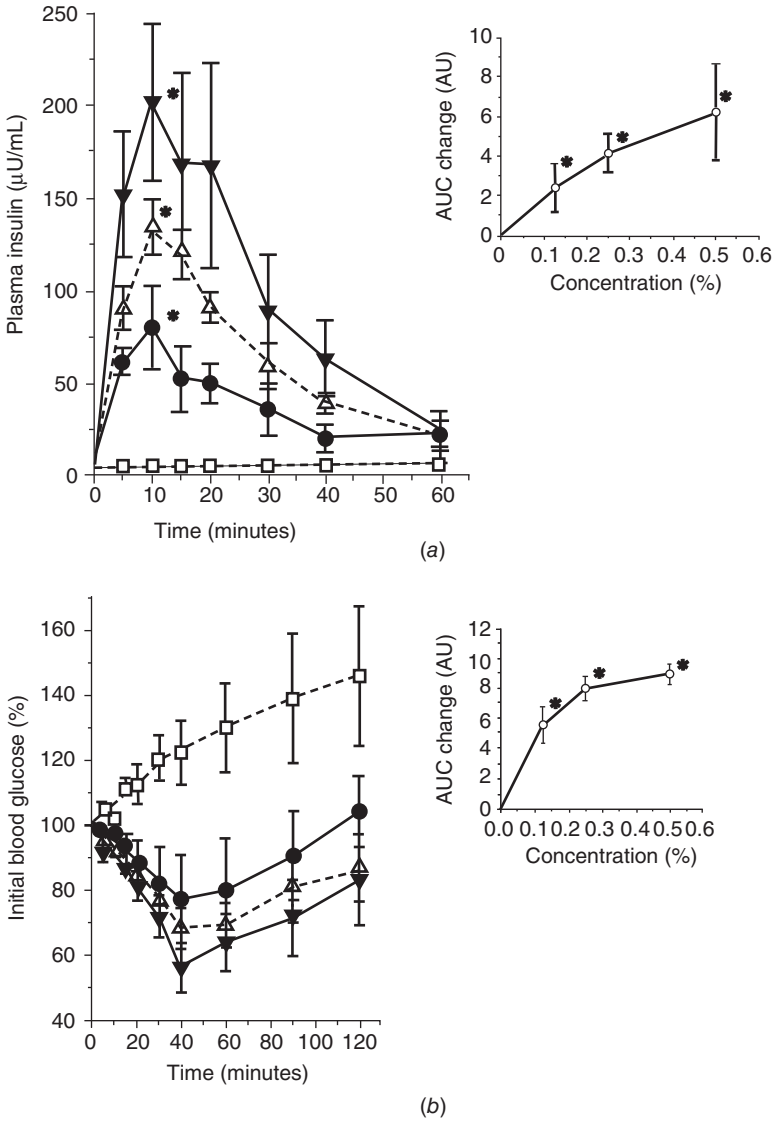


**FIGURE 8** Changes in (*a*) plasma insulin levels and (*b*) blood glucose levels after nasal administration of 0.5 U insulin formulated in saline (□) or 0.125% (•), 0.25% (▼), and 0.5% (%) sucrose cocoate. Blood glucose concentrations at time 0 (250–350 mg/dL) were normalized to a value of 100% in each animal. Data represent mean ± SEM, $n = 3$. Inserts represent changes in plasma insulin $AUC_{0-60}$ (*a*) and changes in blood glucose $AUC_{0-120}$ (*b*). (Reproduced from ref. 67 with permission of Elsevier.)

investigation was also carried out in rats. The results suggested that, by increasing the concentrations of chitosan from 0 to 1.5%, insulin permeation in vitro could be increased. As compared with insulin without chitosan, there was a 25-fold increase in the permeability coefficient of insulin with 1.5% chitosan. A similar increase in permeability was seen in the case of a hyperosmotic solution, and there was also an

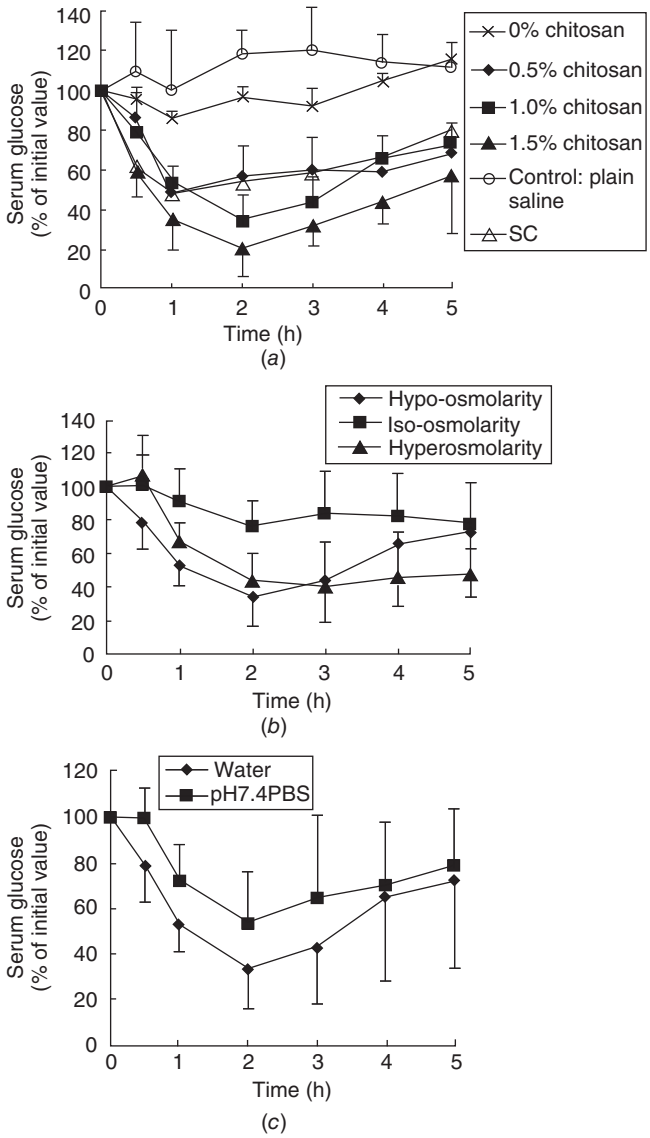


**FIGURE 9** Effect of (*a*) concentrations, (*b*) osmolarity, and (*c*) medium of chitosan solution on mean serum glucose concentrations after nasal administration of 10 IU/kg insulin to rats. Bars represent the standard deviation (SD) of five experiment. (Reproduced from ref. 73 with permission of Elsevier.)

increase in permeability when insulin was formulated in deionized water as compared with phosphate buffer 7.4 (Figure 9). An increase in permeability was also seen when 5% hydroxypropyl β-cyclodextrin (HPβ-CD) plus 1% chitosan was included with insulin as compared with chitosan alone. However, there was no statistical difference in permeability when insulin was formulated with 0.1% ethylenediaminetetraacetic acid (EDTA) and 1% chitosan. In the in vivo studies

in rats, insulin without chitosan did not show any reduction in blood glucose levels, but as the concentrations of chitosan were increased, a hypoglycemic effect was seen. With a formulation of 0.5% chitosan, the nadir in glucose levels was observed within 1 h after administration; with 1 and 1.5% chitosan–insulin formulations, the lowest levels were reached in about 2 h. In the case of isoosmotic formulations, the lowering of blood glucose levels was weaker as compared with the hypo- or hyperosmolar solutions; with deionized water and EDTA, effects similar to those of the permeability studies were seen in terms of lower serum glucose levels. When insulin was formulated with 1% chitosan, there was also a decrease in serum glucose levels, and a similar effect was seen with both 1% chitosan and 0.1% EDTA. A similar effect was also seen in the case of 5% Tween 80. Insulin formulated with 5% HPβ-CD and 1% chitosan was more effective in reducing serum glucose levels than when 5% HPβ-CD or 1% chitosan was used alone. These studies suggested that the concentrations, osmolarity, medium, and inclusion of absorption enhancers in chitosan solution influence absorption following the nasal delivery of insulin [73].

Varshosaz and co-workers recently explored the use of chitosan microspheres and chitosan gels in the nasal delivery of insulin [74]. They prepared microspheres of chitosan, and insulin was loaded into the microspheres. The formulations were administered through the nasal cavity by the method described earlier by Hussain et al. [47]. Varshosaz and colleagues found marked differences in the AUC of blood glucose reduction and the AUC of insulin concentration between the untreated controls and those animals that were treated either by intravenous administration and insulin-loaded chitosan (using ascorbyl palmitate as the cross-linking agent) or with microspheres. Serum glucose reduction in diabetic rats with nasal insulin–chitosan microspheres was around 67% in the intravenous group, and absolute bioavailability ($F_{abs}$) was around 44% [74]. Furthermore, the same researchers prepared insulin–chitosan gels containing different enhancers and investigated their nasal absorption. As seen in Figures 10 and 11, when EDTA was employed as an absorption enhancer in the chitosan gels and administered nasally, a significant increase in insulin absorption and a decrease in serum glucose levels by as much as

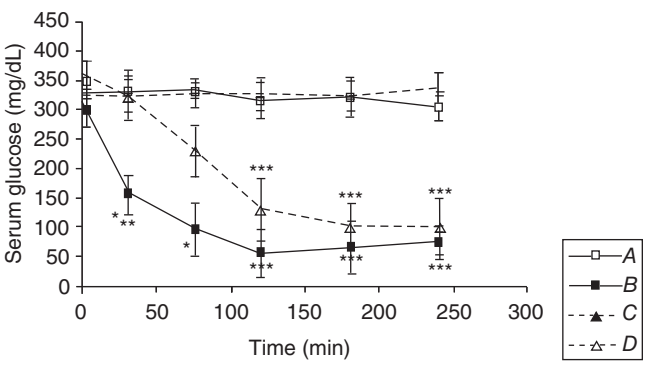


**FIGURE 10** Serum glucose level in four groups of diabetic rats ($n = 6$): *A*, untreated control group; *B*, intravenous administration of 4 IU/kg insulin; *C*, nasal administration of blank gel base; *D*, nasal administration of 100 μL/kg chitosan gel containing 4000 IU/dL insulin. (Reproduced from ref. 75 with permission of Taylor & Francis.)

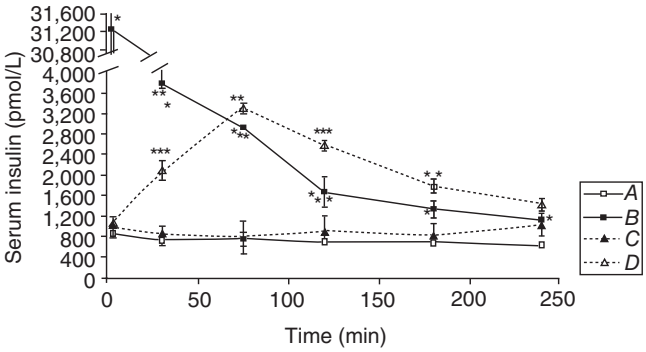


**FIGURE 11** Serum insulin levels of four groups of diabetic rats ($n = 6$): *A*, untreated control group; *B*, intravenous administration of 4 IU/kg insulin; *C*, nasal administration of blank gel base; *D*, nasal administration of 100 μL/kg chitosan gel containing 4000 μg/dL insulin. (Reproduced from ref. 75 with permission of Taylor & Francis.)

46% compared to the intravenous route of administration were obtained. The authors suggested that this formulation would be beneficial in the controlled delivery of insulin by the nasal route [75].

### 6.2 Calcitonin

Calcitonin is a 32–amino-acid peptide with a molecular weight of 3418 that is cleaved from a larger prohormone. It has a single disulfide bond, which causes the amino terminus to assume a ring shape. Calcitonin is a hormone that participates in calcium and phosphorus metabolism. The major source of calcitonin in mammals is the parafollicular or C cells in the thyroid gland; it is also synthesized in other tissues, including the lungs and intestinal tract. When serum calcium levels are elevated, calcitonin is released from the thyroid gland. Salmon calcitonin (sCT) is more potent and longer lasting than the mammalian form and hence is used clinically. Calcitonin's main action is to reduce the plasma concentration of calcium. At pharmacological doses, calcitonin brings about reduction in bone resorption. It is indicated for the treatment of postmenopausal osteoporosis in women with low bone mass relative to healthy premenopausal women. The marketed version of intranasal salmon calcitonin is Miacalcin (calcitonin-salmon) Nasal Spray (Novartis Pharma AS, Huningue, France). Up to now, this calcitonin treatment has been approved only for treatment in women. It has been reserved as a second-line treatment, since it reduces fracture risk less than do other available treatments for osteoporosis. Miacalcin Nasal Spray, first manufactured by Sandoz Pharmaceuticals, was approved by the FDA in 1995 and is now distributed by Novartis Pharmaceuticals in the United States.

Pontiroli et al. [76] looked at the intranasal absorption of calcitonin in normal subjects. Their study included six healthy volunteers who had no family history of endocrine or metabolic diseases. Human calcitonin (Cibacalcin; Ciba-Geigy) was administered intravenously or mixed with sodium glycocholate, a surfactant, in distilled water and instilled as nose drops. Plasma concentrations of calcitonin were found to be consistently higher when compared with intranasal administration of

the same dose of calcitonin. However, it was found that intranasal administration effected a reduction in the plasma concentrations of calcium to a similar extent as that seen when a similar dose of intravenous calcitonin was given. The data also showed that although a higher dose of calcitonin by the intranasal route brought about a higher plasma calcitonin concentration, there was no difference in the decrease in plasma concentrations when calcitonin was given at doses of 500 and 1000 μg.

Polyacrylic acid aqueous gel enhances the absorption of calcitonin after nasal as well as rectal administration. When [Asu1,7]-eel calcitonin (10 U/kg) was administered nasally in polyacrylic acid gel at a concentration of 0.1% w/v, a prominent hypocalcemic effect was seen in the first 30 min. Nasal administration of [Asu1,7]-eel calcitonin in saline had no hypocalcemic effect at the same dose when given by the nasal route. In addition to this, the effect of [Asu1,7]-eel calcitonin in the dose range of 1–10 U/kg has also been studied. The resulting data showed that a rapid reduction in plasma calcium concentrations can be achieved at doses of 5 and 10 U/kg; however, at doses of 1 U/kg only a small reduction in the plasma calcium concentration was observed, suggesting that polyacrylic acid gel can be used for the intranasal administration of peptides such as calcitonin. The possible side effects, however, were not known at the time the study was performed [76–78].

Commercially available sCT has been used mainly in the treatment of bone-related diseases such as Paget's disease, hypercalcemia, and osteoporosis. It can be used to alleviate pain due to its analgesic properties; hence there may be a need for its frequent administration. As reported in the literature, owing to chemical and enzymatic degradation, the polypeptide has a short half-life of only about 14 min. Hence its use calls for measures to improve its in vivo stability and overcoming the problem of rapid clearance. Polyethylene glycol (PEG) has been extensively used for this purpose in association with a variety of agents. Lee at al. [79] have prepared and characterized PEG-modified sCT and studied how blood clearance is affected with and without PEGylation. Succinimidyl carbonate mono-methoxy-polyethylene glycol (SC-mPEG) was prepared using mPEG, which has a molecular weight of 12,000, and bis(*N*-succinimidyl) carbonate as per the procedure of Miron and Wilcheck [80]. There are three main reactive sites for the activated PEG in sCT: the N-terminal amino group ($Cys^1$) and the ε-amino groups of two lysine residues ($Lys^{11}$ and $Lys^{18}$), as shown in Figure 12. The site of conjugation of the PEG on sCT determines the stability of the sCT in the face of enzymatic attacks; hence the three possible mono-PEG-sCTs—$Cys^1$-PEG-sCT, $Lys^{11}$-PEG-sCT, and $Lys^{18}$-PEG-sCT—can withstand the effects of proteolytic enzymes. It was found that, on PEGylation, the plasma half-life improved to 11.2 and 54.0 min for the mono-PEG-sCT and the di-PEG-sCT as compared with the non-PEGylated sCT, which showed a plasma half-life of about 4.7 min (Figure 13).

$Cys^1$-Ser-Asn-Leu-Ser-Thr-Cys-Val-Leu-Gly-$Lys^{11}$-Leu-Ser-Gln-Glu-Leu-His-$Lys^{18}$-Leu-Gln-Thr-Tyr-Pro-Arg-Thr-Asn-Thr-Gly-Ser-Gly-Thr-Pro-$NH_2$

**FIGURE 12** Primary structure of salmon calcitonin. Possible PEGylated sites are $Cys^1$, $Lys^{11}$, and $Lys^{18}$. (Reproduced from ref. 79 with permission of Taylor & Francis.)

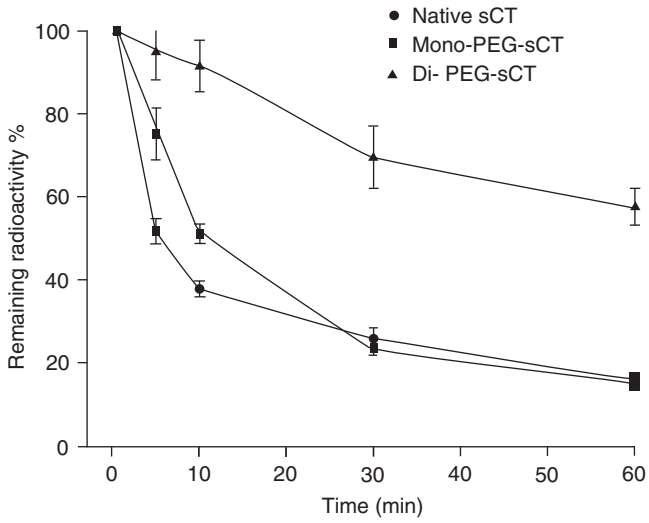


**FIGURE 13** Blood clearance of PEGylated salmon calcitonins in rat. (Reproduced from ref. 79 with permission of Taylor & Francis.)

The same group—Shin et al. [81] —did nasal absorption studies of a low-molecular-weight PEG (2000) instead of the 5000 and 12,000 they tried previously. They also used commercially available succinimidyl-propionated monomethoxy-poly(ethylene glycol)-2000 (SP-mPEG) for the chemical modification instead of synthesizing it, as in their previous studies. The PEGylation of sCT was done by mixing SC-mPEG and sCT, and this mixture was shaken at 25°C. The reaction mixture was stopped by using an excess of 1.0 *M* glycine solution. This conjugated mixture was then subjected to size exclusion chromatography. Radioiodination of the sCT and the PEG-sCT were carried out and the radiolabeled $^{125}$I-sCT and $^{125}$I-mono-PEG2000-sCT were then used for the nasal absorption studies.

Tissue distribution studies were also done in rats after nasal administration. As seen in Figure 14 and Table 2, it was found that the elimination half-life of the unmodified sCT was 199 min, whereas the SP-mPEG2000-modified sCT showed an increased terminal elimination with a half-life of 923 min. It was also found that the SP-mPEG2000-modified sCT took a significantly longer time to reach its maximum concentration, 520 min, as compared with the 77 min for the unmodified sCT, and the AUC was found to be 20,638 μg/min/mL, which is much higher than the 3650 μg/min/mL for the unmodified sCT. The authors reported that the increase in the terminal half-life observed could be due to a flip-flop phenomenon. Also, when the tissue distribution of the formulation was examined 12 h after administration, the highest radioactivity was found in the liver. The details of the biodistribution studies are as shown in Table 3.

The same group [82] further studied the stability of these mono-PEG2000-sCT and the unmodified sCT in the rat nasal mucosa. It was found that PEGylated sCT exhibited significant resistance against trypticlike and nonspecific enzymatic degradation. Ahsan et al. [49] showed that when sCT was formulated with alkylglycosides, bioavailability was enhanced following both nasal and ocular administration. Miacalcin (Novartis Pharma AS, Huningue, France) was used to prepare the formulation

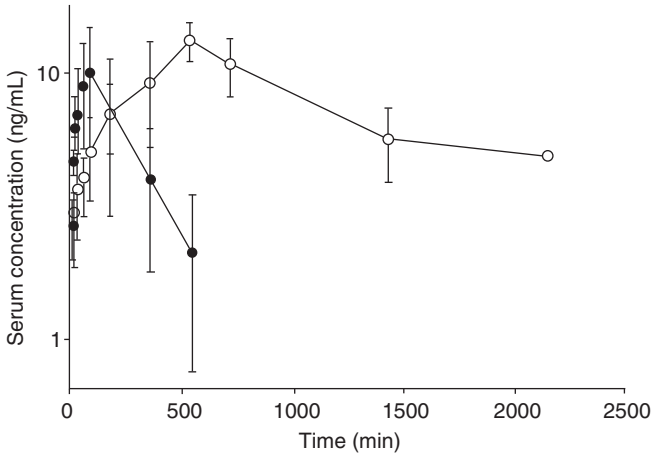


**FIGURE 14** Average serum concentration of intact sCT–time curves following nasal administration of unmodified sCT (•) and Mono-PEG2K-sCT (○) in rats ($n$ = 9 each). (Reproduced from ref. 81 with permission of The Pharmaceutical Society of Japan.)

**TABLE 2 Pharmacokinetic Parameters (Mean ± SD) of Unmodified sCT and Mono-PEG2k-sCT obtained after nasal administration to rats ($n$ = 9 each)**

| Parameter | Unmodified sCT | Mono-$PEG_{2k}$-sCT |
|---|---|---|
| $C_{max}$ (ng/mL) | 10.5 ± 4.7 | 12.9 ± 3.0 |
| $t_{max}$ (min) | 77 ± 22 | 520 ± 167[a] |
| $t_{1/2,\lambda z}$ (min) | 199 ± 97 | 923 ± 389[a] |
| $CI/F$ (mL/min) | 7.4 ± 5.2 | 1.3 ± 1.0[a] |
| $Vss/F$ (mL) | 1802 ± 811 | 1392 ± 450 |
| AUC (ng·min/mL) | 3650 ± 1894 | 20638 ± 9486[a] |
| AUC/D (ng·min/mL/ng) | 0.18 ± 0.09 | 1.03 ± 0.48[a] |
| $MRT_{\mu}a$ (min) | 314 ± 131 | 1505 ± 560[a] |

*Source*: From ref. 74.

[a]Significantly different from unmodified sCT ($p < 0.05$).

**TABLE 3 Extent of Total Radioactivity (Mean ± SD) in Various Body Organs after Nasal Administration of Unmodified sCT and Mono-PEG2k-sCT to Rats ($n$ = 9 each)**

| | Radioactivity (%) in Whole Organ[a] | |
|---|---|---|
| Tissue | Unmodified SCT | Mono-$PEG_{2k}$-sCT |
| Liver | 0.80 ± 0.41 | 1.03 ± 0.65 |
| Kidney | 0.30 ± 0.14 | 0.52 ± 0.24 |
| Lung | 0.18 ± 0.12 | 0.20 ± 0.12 |
| Heart | 0.06 ± 0.03 | 0.13 ± 0.05[b] |
| Spleen | 0.04 ± 0.02 | 0.10 ± 0.05 |
| Thyroid | 0.04 ± 0.03 | 0.05 ± 0.02 |

*Source*: From ref. 74.

[a]Determined 12 h after administration.

[b]Significantly different from unmodified sCT ($p < 0.05$).

in solutions containing different concentrations of tetradecylmaltoside and octylmaltodise. These formulations were administered as described by Ahsan et al. [49]. It was found that when calcitonin was formulated in saline, absorption after administration by the nasal route was negligible; a similar result was seen with 0.125% OM. However, when calcitonin was formulated with 0.125% TDM, absorption was found to be increased, and maximal absorption ($T_{max}$) was achieved after 10 min. The $AUC_{0-40}$ was found to be fourfold higher than with saline and OM formulations. The bioavailability was found to be 53% as compared with intravenously administered calcitonin at the same dose of 2.2 U. The $AUC_{0-40}$ was found to be 6250 pg/mL·min as compared with 3500 pg/mL·min when the concentration of TDM was increased to 0.25% from 0.125%. It was also found that in the absence of the absorption enhancer increasing the amount of calcitonin from 2.2 to 22 U did not increase absorption by the nasal route. However, when formulated in the presence of 0.25% TDM, there was a 23-fold increase in the relative bioavailability of calcitonin. Also, when calcitonin was formulated with 0.25% TDM and given as nose drops, there was a significant reduction in the plasma calcium concentration as compared with the saline formulation.

### 6.3 Low-Molecular-Weight Heparins

Low-molecular-weight heparins (LMWHs) are fragments of natural heparin that are obtained by either enzymatic degradation or chemical depolymerization of unfractionated heparins (UFHs). The molecular weight of LMWHs, a heterogeneous mixture of sulfated glycosaminoglycans, is about one-third that of UFHs. Owing to the variations in the distribution of molecular weights, they show differences in their affinity for plasma proteins, activity against factor Xa, and thrombin, as well as duration of activity. They have been proven to be useful in the treatment and prevention of deep vein thrombosis. Of late, LMWHs have been preferred to therapy with conventional heparin. However, one of the main disadvantages of the use of LMWHs on a regular, noninstitutional basis is that they must be delivered by the subcutaneous or intravenous route. There have been concerns regarding patient compliance, longer hospital stays, and the need for skilled health professionals for therapeutic drug monitoring and administration. This has prompted a number of researchers to seek alternative forms of delivery.

The nasal administration of LMWHs was investigated in Sprague-Dawley rats using TDM as an absorption enhancer [83]. TDM was used at concentrations of 0.125 and 0.25%. Lovenox (Aventis Pharmaceuticals, Bridgewater, NJ) (enoxaparin sodium injection), a commercially available LMWH, was prepared with 0.25% TDM and the nasal absorption was compared with and without TDM. It had been previously reported that a plasma anti–factor Xa level of 0.20 U/mL or higher is required for an antithrombotic effect to be considered therapeutic [83]. It was found that enoxaparin when formulated only in saline did not produce a therapeutic anti–factor Xa level. However, when enoxaparin was formulated with 0.25% TDM, a significant increase in the AUC and $C_{max}$ for the anti–factor Xa level was observed. Dalterparin, another commercially available LMWH, showed a similar response. When UFH was formulated with TDM, it produced an increase in the anti–factor Xa levels as compared with saline, but it was not in the therapeutic range (Figure 15).

The bioavailability ($F_{abs}$) of enoxaparin achieved by the subcutaneous route was found to be 83% of that achieved by the intravenous route. In the absence of TDM,

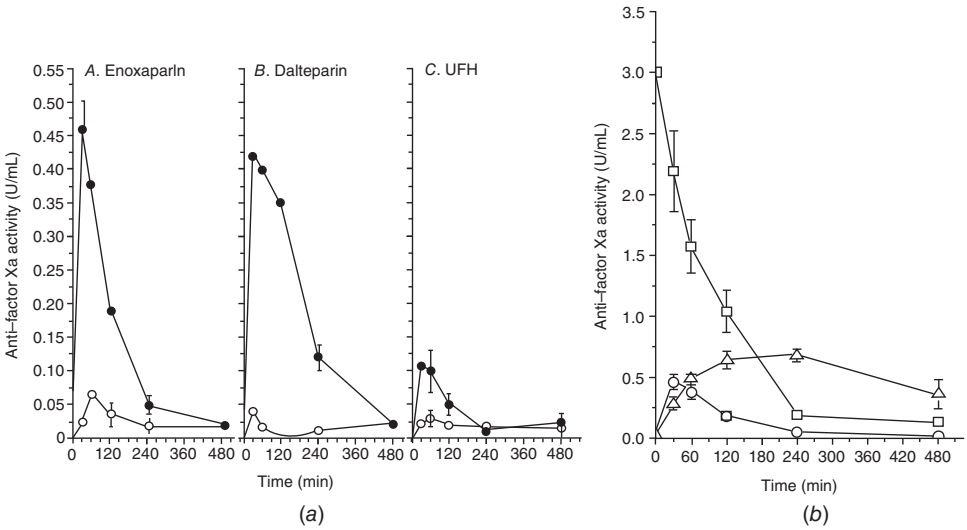


**FIGURE 15** (*a*) Nasal administration of 100 U of (*A*) enoxaparin, (*B*) dalteparin, or (*C*) UFH formulated with (•) and without (○) 0.25% tetradecylmaltoside. Data represent mean ± SEM, $n = 3$. Asterisks indicate results that are significantly different from those obtained with the drug formulated with saline, $P < 0.05$. (*b*) Administration of 100 U of enoxaparin via the subcutaneous (▲), intravenous (□), and nasal (○) routes. Nasal administration was performed with a formulation that included 0.25% TDM. Data represent mean ± SEM, $n = 3$. (Reproduced from ref. 83 with permission of John Wiley & Sons.)

the nasal $F_{abs}$ was 4.0%, but in the presence of 0.25% TDM, the $F_{abs}$ was 19% compared with the intravenous route. The relative bioavailability ($F_{rel}$) of nasal enoxaparin plus 0.25% TDM was found to be 23% compared with the subcutaneous route.

The use of alkanoylsucroses in the enhancement of nasal absorption of LMWHs was investigated by our group [84]. As seen in Figure 16*a*, enoxaparin plus 0.125% octanoylsucrose showed no significant increase in anti–factor Xa levels; even when the concentration was increased to 0.5%, it barely reached the therapeutic level of 0.2 U/mL. Similar results were reported for 0.125% decanoylsucrose; however, when the concentration was increased to 0.25 or 0.5%, there was an appreciable and rapid rise in anti–factor Xa levels (Figure 16*b*). The inability of octanoylsucroses and 0.125% decanoylsucrose to increase anti–factor Xa levels is attributed to the critical micellar concentration as reported by the authors [84]. In the case of dodecanoylsucroses, all the concentrations produced anti–factor Xa levels well above the therapeutic level required for an antithrombotic effect (Figure 16*c*).

Our group also compared the efficacy and potency of alkanoylsucroses with those of the well-known absorption enhancer 1% sodium glycocholate (Figure 16*d*); the results were similar to those seen with 0.5% dodecanoylsucrose. Also, 0.5% dodecanoylsucrose showed the highest increase in $C_{max}$, and it was found that an increase in the concentration of alkanoylsucroses led to a subsequent increase in the $C_{max}$. When the absolute and relative bioavailabilities of nasal LMWH plus 0.5% dodecanoylsucrose were compared with those of 1% sodium glycocholate, similar profiles were found [84].

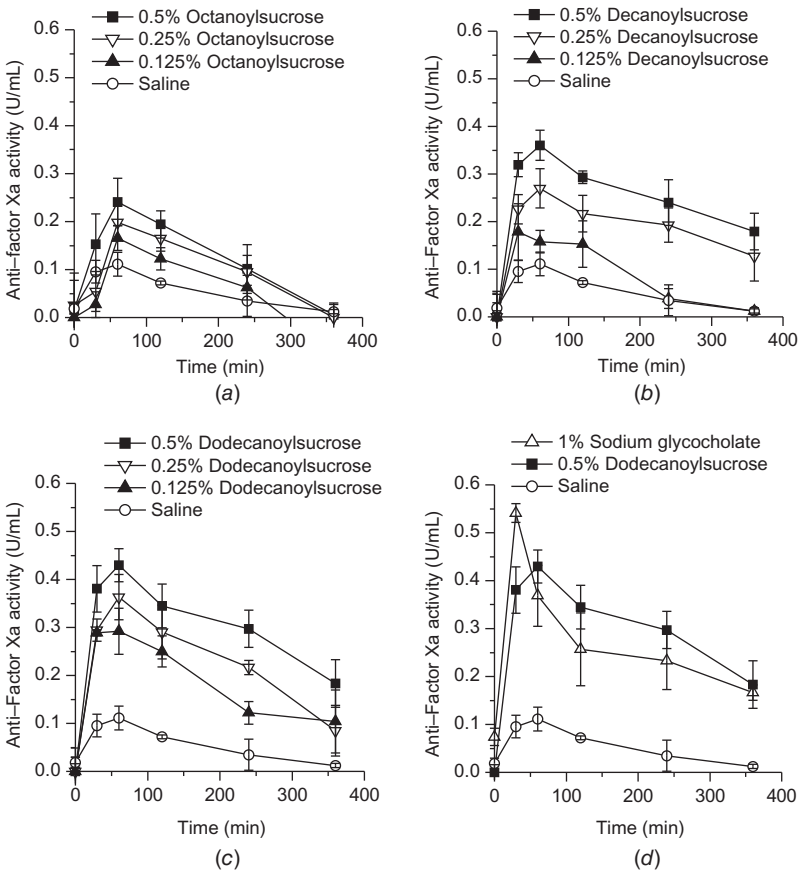


**FIGURE 16** Changes in anti–factor Xa activity after nasal administration of enoxaparin formulated in saline or in presence of different concentrations of the following: (*a*) octanoylsucrose; (*b*) decanoylsucrose; (*c*) dodecanoylsucrose; (*d*) sodium glycocholate and dodecanoylsucrose to anesthetized rats (enoxaparin dose, 330 U/kg). Data represent mean ± SEM, $n = 3, 5$. (Reproduced from ref. 84 with permission of Pharmaceutical Press.)

The authors also studied the influence of the chain length of alkylmaltosides on the nasal absorption of enoxaparin. The results indicated that increases in the concentration of alkylmaltosides increased the AUC for plasma anti–factor Xa; it was found that the absolute and relative bioavailabilities of enoxaparin increased by twofold with an increase in alkyl chain length from 8 to 14 carbons. Of all the alkylmaltosides, TDM was found to be the most potent absorption enhancer [85].

Furthermore, we have also shown the efficacy of cyclodextrins in enhancing absorption following the nasal delivery of LMWHs. Three different cyclodextrins were employed: β-cyclodextrins (β-CD), hydroxypropyl β-CD (HPβ-CD), and dimethyl β-CD (DMβ-CD). β-CD showed therapeutic levels of anti–factor Xa only at 2.5 and 5% β-CD, but there was no significant difference between the two concentrations, which was attributed to their solubility limit. In the case of HPβ-CD, neither 1.25 nor 2.5% produced an appreciable increase in anti–factor Xa levels;

only 5% HPβ-CD showed levels above 0.2 U/mL, which, however, was not significant. Unlike the other two cyclodextrins, 1.25 and 2.5% DMβ-CD showed a fourfold increase in AUC profiles. The studies reported that 5% DMβ-CD produced the greatest increase in the bioavailability of LMWHs, with an eightfold increase in the AUC profile. It was also found that the reduction in transepithelial electrical resistance (TEER) and changes in tight junction protein ZO-1 distribution facilitated by 5% DMβ-CD were much greater than with β-CD or HPβ-CD [86]. Recently our group has shown that positively charged polyethylenimines (PEI) increases the nasal absorption of LMWHs by reducing the surface negative charge of the drug. When PEI 1000 kDa was employed at a concentration of 0.25%, a four-fold increase in the absolute and relative bioavailabilities was observed in comparison with the control formulation of LMWH plus saline [87].

### 6.4 Azetirelin

Azetirelin is a novel analog of the tripeptide thyrotropin-releasing hormone (TRH). It was discovered in 1969 when two different groups of researchers, led by Guillemin and Schally, showed that the hypothalamic substance that causes the anterior pituitary gland to release thyrotropin (thyroid-stimulating hormone, or TSH) is L-pyroglutamyl-L-histidyl-L-prolineamide (L-pGlu-L-His-L-ProNH2), now called TRH. In azetirelin the pyroglutamyl moiety of the TRH is replaced by an (oxo-azetidinyl) carbonyl moiety. It has been reported that the inhibition of pentobarbital-induced sleep and reserpine-induced hypothermia due to azetirelin in mice, as opposed to TRH, are about 10–100 times more effective as well as 8–36 times longer lasting. Azetirelin is stable in plasma and degrades much more slowly than TRH in brain homogenates, thus showing improved pharmacological potency as well as efficacy over TRH. It is highly potent when given intravenously; however, when administered by the oral route, it shows very low bioavailability of only 2% [88–91].

Kagatani et al. [90] studied the effect of acylcarnitines as drug absorption enhancers for the nasal delivery of azetirelin in a rat model. A buffered azetirelin sample solution was administered intranasally, as described previously [47]. The nasal and oral absorptions of azetirelin were then compared. The $F_{abs}$ after nasal absorption was found to be 17.1%, which was 21 times greater than the 0.8% after oral administration. As reported above, a pilot study of oral azetirelin showed a bioavailability of about 2%. A bioavailability of about 20% was seen in the case of nasally administered TRH in humans as well as rats. The authors predicted that since azetirelin is an analog of TRH, its pharmacokinetic properties after nasal delivery in humans could also be about 20% [90, 91].

### 6.5 Growth Hormones

Recombinant human growth hormone (hGH) is a 22-kDa protein drug having 191 amino acids. It has been used to treat a number of conditions, including short stature in children, Turner syndrome, and chronic renal failure. It is said to play an important role in the metabolism of proteins, carbohydrates, and fats as well as electrolytes and hence influences weight and height. It has been reported that hGH secretion in humans is pulsatile, showing low basal serum levels in between peaks. It has been

suggested that this secretory pattern of hGH can be mimicked by the nasal route as opposed to painful subcutaneous injections [8, 92, 93].

In a pharmacokinetic-based study by Hedin et al. [94], hGH was administered with a nasal permeation enhancer, sodium tauro-24, 25-dihydrofusidate (STDHF), in patients deficient in growth hormone (GH) using a reprocessed lyophilized form of hGH. The lyophilized material was formulated with STDHF and all the subjects received the formulation by both the nasal and subcutaneous routes. The dose given by the subcutaneous route was a standard dose of 0.1 IU/kg body weight (BW), whereas three different doses (of 0.2, 0.4, and 0.8 IU/kg BW) of the nasal formulation were given. As compared with the subcutaneous route, all three nasal formulations showed a rapid increase in the plasma levels of hGH, with $T_{max}$ being reached 15–25 min after administration, as compared with 3–4 h in the case of the subcutaneous route. However, the $C_{max}$ was higher in the case of the latter route, and the nasal formulations touched baseline after 3–4 h, as compared with 14–18 h after subcutaneous delivery.

Several studies have shown that frequent doses of hGH are more beneficial than the total amount given as a single dose and that higher peaks of plasma hGH with low troughs are found in taller children; hence the nasal therapy for deficiencies in hGH would not only be more convenient but also offer advantages, including a rapid decrease of the peaks to zero levels and a mimicking of the pulsatile pattern of the endogenous hormone [94–97]. Nasal irritation studies also were carried out, indicating that the nasal formulations showed only local short-term irritation.

Didecanoylphosphatidylcholine (DDPC) and α-cyclodextrin (α-CD) were used as enhancers and reversibility studies were carried out in vivo in rabbits. Three different combinations were used: DDPC, α-CD, and DDPC plus α-CD for the nasal administration of hGH. Vermehren et al. [98] used intravenous hGH as the reference. When hGH was administered with α-CD as a powder, 23.6% bioavailability was seen, as compared with 18.1% $F_{rel}$ when given at the same time as two powders. When hGH only was given, a bioavailability of 8.3% was attained. DDPC plus hGH together showed a $F_{rel}$ of 22.3 and 21.5% when given as two powders, while simultaneous administration of DDPC plus α-CD and hGH as two powders showed a $F_{rel}$ of 14.3%. When dosed as one single powder, it showed a $F_{rel}$ of 31.9%. Reversibility of the enhancer effect was seen when a dose of hGH in enhancer-free formulation was given 30 min after dosing of the test formulations; this resulted in reduction in the AUC and $C_{max}$ by half.

Another group studied the enhancer effect of DDPC on the pharmacokinetics and the biological activity of nasally administered hGH in GH-deficient patients. Three different doses—0.05, 0.10, and 0.20 IU/kg with DDPC—were given by the nasal route and 0.10 and 0.015 IU/kg were given by the subcutaneous and intravenous routes. A short-lived serum GH peak was seen in the intravenous treatment, showing a peak value of around 128 μg/L, whereas the subcutaneous route showed a peak level of 13.98 μg/L and nasal doses showed peaks of about 3.26, 7.07, and 8.37 μg/L for the three treatment doses. The bioavailabilities of the nasal doses were found to be 7.8, 8.9, and 3.8%, respectively, as compared with the $F_{abs}$ of 49.5% for the subcutaneous dose. It was also found that the serum insulinlike growth factor 1 (IGF-1) increased only upon subcutaneous administration There was no change in the serum IGF protein binding protein 3 levels in any of the nasal doses or in the subcutaneous or intravenous doses [92].

Leitner et al. [93] tried to overcome the limitations posed by the low bioavailability of nasally delivered hGH due to the drug's high molecular weight and hydrophilicity. The strategy employed was to use dry polymer particles to enhance absorption, taking advantage of the fact that these polymers would form a gellike layer and hence be cleared slowly, giving a longer circulating drug. Due to the presence of covalent immobilization of sulfhydryl groups on the backbone of the thiolated polymers or thiomers, the permeation-enhancing properties of these thiomers were improved. They have also been reported to have mucoadhesive and enzyme-inhibitory properties, and the addition of glutathione (GSH) was found to increase their permeation-enhancing properties further [93, 99–103]. A polycarbophil–cysteine (PCP–Cys) microparticulate system was prepared using GSH and hGH, and this was added to the formulation. The in vitro release profiles of the PCP–Cys and unmodified PCP containing hGH proved to be similar. Three different formulations were tested for in vivo nasal absorption: PCP–Cys/GSH/hGH microparticles, PCP/hGH microparticles, and mannitol/hGH powder against a subcutaneous hGH. The $F_{rel}$ for mannitol/hGH powder was around 2.40%, the PCP–Cys/GSH/hGH microparticles showed a $F_{rel}$ of 8.11%, and PCP/hGH showed 2.70%, which represents a threefold increase in nasal uptake when thiomer/GSH is used in the formulation. The microparticulate formulation of PCP–Cys/GSH/hGH showed a sixfold higher plasma concentration when compared to the PCP–Cys/GSH/hGH gel formulation.

## 7 NASAL DELIVERY OF NONPEPTIDE MOLECULES

### 7.1 Morphine

Morphine is a potent narcotic analgesic used preoperatively and as an anxiolytic agent in pediatric patients; it is also used in the management of postoperative pain as well in moderate to severe pain in cancer. Oral morphine in solution, immediate-release, or controlled-release formulations shows a bioavailability of only about 20%. Morphine absorption in humans is poor, and only 10% bioavailability is obtained when it is given as a solution rather than intravenously. Chitosan has been reported to be a potent absorption enhancer and greatly improves the absorption of small polar molecules and peptides. Intranasal administration was carried out in a sheep model; various formulations were also tested in humans and comparisons versus the intravenous route were made. The sheep model was used for nasal delivery because it has been reported that testing in sheep is highly predictive of results in humans [104, 105]. In sheep studies, when morphine HCl solution was given nasally (control), the $C_{max}$ obtained was limited: 151 n$M$ with a $F_{rel}$ of 10%, with a $T_{max}$ of about 20 min, suggesting a slow rate of absorption. When 0.5% chitosan was formulated with morphine as a solution, the $C_{max}$ increased to 657 n$M$, and $F_{rel}$ was 26.6%. The rate of absorption was also increased, as evidenced by a $T_{max}$ of 14 min. With the formulation of chitosan into microspheres, the $C_{max}$ was 1010 n$M$, $T_{max}$ about 8 min, and $F_{rel}$ about 54.6%. A further increase in nasal absorption was seen when morphine was formulated as a powder consisting of starch microspheres and L-α-lysophosphatidylcholine (LPC), with $C_{max}$ being 1875 n$M$, $T_{max}$ about 10 min, and $F_{rel}$ about 75% (Figure 17). In the case of human phase I clinical trials, a dose of

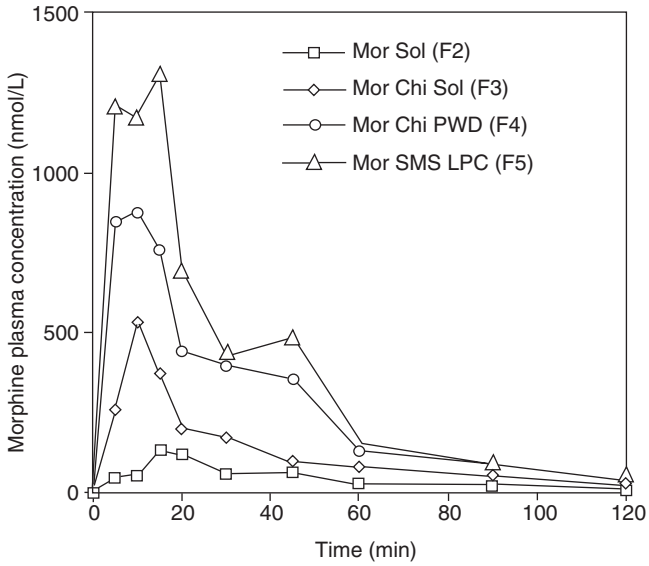


**FIGURE 17** Morphine plasma concentration after nasal administration of morphine formulations in sheep: Mor Sol, morphine solution; Mor Chi Sol, morphine solution containing chitosan; Mor Chi PWD, morphine chitosan powder; Mor SMS LPC, starch microspheres with lysophosphatidylcholine and morphine as a freeze-dried powder. (Reproduced from ref. 105 with permission of the American Society for Pharmacology and Experimental Therapeutics.)

10 mg morphine sulfate led to a mean $C_{max}$ of 336 n$M$ following 30 min of intravenous administration. A 0.5% chitosan and morphine HCl solution led to a $C_{max}$ of 98 n$M$ and a $T_{max}$ of about 16 min. The plasma half-life ($t_{1/2}$) in the case of nasal administration was found to be 2.98 h, as compared with 1.67 h via the intravenous route. The mean bioavailability with the nasal solution of morphine plus chitosan was 56%. Furthermore, the powder formulation comprising chitosan and morphine HCl showed a $C_{max}$ of 92 n$M$, $T_{max}$ of 21 min, $t_{1/2}$ of 2.72 h, and $F_{rel}$ of about 56% (Figure 18). The nasal formulations were reported to be well tolerated.

### 7.2 Benzodiazepines

Diazepam has been the standard or preferred option for the treatment of all types of seizures in both children and adults. However, it has disadvantages, including a short duration of action, so that in some cases diazepam must be given rectally in order to manage prolonged seizures. Moreover, its use in the community is restricted because of the need for privacy, especially in the case of adult patients. Finally, the intravenous route is also reported to be inconvenient, since nonprofessional caregivers may not be comfortable enough to administer the drug in this way. Midazolam has been reported to be clinically effective with both intravenous and oral administration for the induction of sedation and reduction of anxiety. Owing to the drawbacks mentioned above, intranasal formulations of benzodiazepines could be highly beneficial [106, 107].

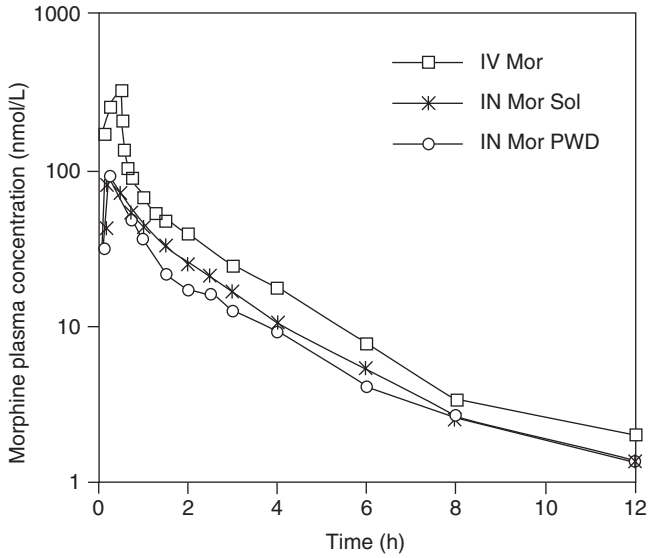


**FIGURE 18** Morphine plasma concentration in human volunteers after intravenous administration of morphine and after nasal administration of morphine as chitosan solution and powder formulations: Mor Chi Sol, morphine solution containing chitosan; Mor Chi PWD, morphine–chitosan powder; IN, intranasal. (Reproduced from ref. 105 with permission of the American Society for Pharmacology and Experimental Therapeutics.)

***Midazolam, Triazolam, and Flurazepam*** The feasibility of intranasal administration of midazolam, flurazepam, and triazolam has been studied and compared with oral absorption in dogs. There was a 3.4-fold increase in the $C_{max}$ after nasal administration, from 5.5–8.7 ng/mL to 17.4–30.0 ng/mL. The mean $t_{1/2}$ showed comparable values for both routes. The $T_{max}$ obtained after nasal administration of midazolam was found to be 15 min, as compared with the 15–45 min observed for oral dosing, while the $C_{max}$ after nasal administration was 6.5–20.3 ng/mL, as compared with 3.0–8.6 ng/mL observed for the oral route. Like midazolam and triazolam, flurazepam also showed a shorter half-life, 15 min, as compared with 15–45 min with oral administration. The $C_{max}$ for oral administration was 0.14–0.59 ng/mL; after nasal administration it was in the range of 2.6–11.1 ng/mL, a 16.4-fold increase. Since the gastrointestinal tract at bedtime is likely to be in the fed state, causing a twofold decrease in the absorption of midazolam and triazolam, the nasal route may be a better option for the treatment of amnesia, since these drugs cross the nasal mucosa effectively without the use of an absorption enhancer, as shown in these studies [108].

In situ nasal absorption studies of midazolam were carried out in rats. The effects of solution concentration, osmolality, and pH on nasal absorption were studied using the in situ perfusion technique. The absorption of midazolam was reported to be prevented at osmolalities in the range of 142–450 mOsm/kg; however, a hypoosmotic 3-mOsm/kg solution resulted in significant absorption, where the pH rose from 3.3 to 6.5. No lag time in absorption was observed when the solutions were buffered at a pH of either 5.5 or 7.4; however, at pH 3.3, no absorption was seen, suggesting

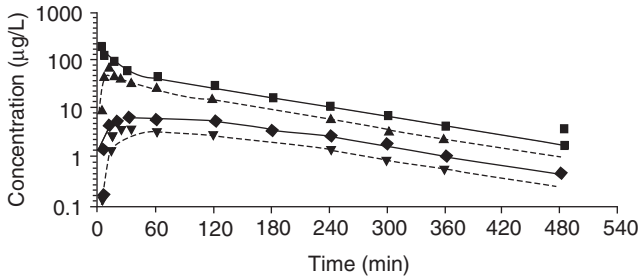


**FIGURE 19** Fit of composite model to concentration–time data for midazolam and 1-hydroxymidazolam in one volunteer. Solid lines indicate the time course of midazolam concentrations (■) and 1-hydroxymidazolam concentrations (◆) after intravenous administration. Dotted lines indicate the time course of midazolam concentrations (▲) and 1-hydroxymidazolam concentrations (▼) after intranasal administration. (Reproduced from ref. 111 with permission of Blackwell Publishing.)

that pH was the main factor determining the absorption of midazolam (Figure 19) [109].

The pharmacokinetics and pharmacodynamics of midazolam after nasal administration were investigated in healthy volunteers in two different studies in comparison with the intravenous route [110, 111]. Studies reported in 1997 demonstrated that intranasal midazolam was rapidly absorbed, with maximal concentration attained in the range of 10–48 min, with a mean of 25 min. These results were only the maximum concentration achieved and the time taken to reach this maximum. The maximum concentration reported after intranasal administration was in the range of 91.0 to 224.3 ng/mL, with a mean of 147 ng/mL. Bioavailability reported in this study [110] was about 50%, in line with the results obtained with oral administration, as reported in an earlier study [112]. Knoester et al. [111] also carried out similar studies using a concentrated intranasal spray in healthy volunteers. The concentrated preparation was prepared by mixing midazolam HCl in a mixture of water and propylene glycol, pH 4. A Spruyt Hillen (IJsselstein, Netherlands) intranasal device was used to deliver the dose. Besides nasal irritation lasting 10 min and teary eyes, no other discomfort was reported. Midazolam was rapidly absorbed on nasal administration, showing a maximum concentration of 72 ng/mL within 14 min. A mean bioavailability of 0.89 was obtained. It was reported that intraindividual basal electroencephalogram (EEG) activity after intranasal administration was comparable with that after intravenous administration (Figures 19 and 20).

***Diazepam*** As mentioned earlier, because of shortcomings of rectal administration, the nasal delivery of diazepam has gained interest. The nasal bioavailability of diazepam in sheep was estimated and further compared with results obtained earlier in humans and rabbits [106]; in this study, human and rabbit nasal bioavailability for the first 30 min was reported to be 37 and 54%, respectively [113]. Diazepam solubilized in PEG 300 was used for nasal administration via a modified nasal device, a Pfeiffer unit dose (Princeton, NJ). The sheep received the nasal formulations in a fixed standing position such that the head was slightly tilted back. It was found that the serum concentration after administration of a 7-mg solution of diazepam was

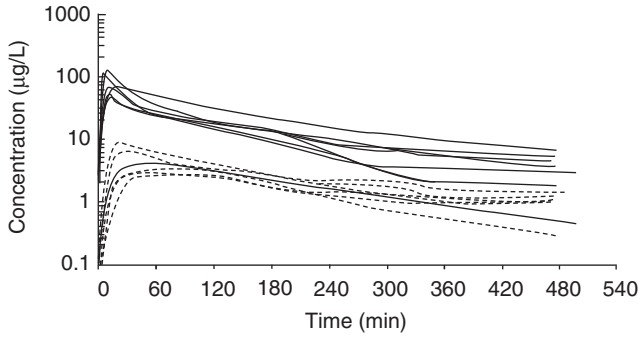


**FIGURE 20** Individual plasma concentration–time curves for midazolam (solid lines) and 1-hydroxymidazolam (broken lines) after intranasal administration of 5 mg midazolam. The bold curves represent the mean pharmacokinetic model fit to the data. (Reproduced from ref. 111 with permission of Blackwell Publishing.)

lower than that obtained with a 3-mg solution, suggesting a low nasal bioavailability. The bioavailability after the initial 30 min was found to be 15%, as compared to the earlier mentioned bioavailability in human and rabbit, and a $C_{max}$ of 934 ng/mL with a $T_{max}$ of 5 min. The difference in bioavailability between animal and human was larger when periods shorter than 0–60 min were used in the calculations. When the results among the three species were compared, it was found that the bioavailability in sheep was higher than that in humans during the early or initial phases, after which the reverse was observed. Lindhardt et al. [113] used a profile that took into account the observation period with respect to the rate of bioavailability and found that similar profiles with respect to rate were observable in relation to all of the three nasal formulations given in humans; moreover, an optimal correlation between sheep and rabbit was observed. The authors suggested that use of the jugular vein in sampling blood from sheep could have resulted in the low bioavailability.

### 7.3 Buprenorphine

Buprenorphine is a derivative of the morphine alkaloid thebaine and is a partial opioid antagonist. It exerts an agonistic effect on the μ-muscarinic receptors and an antagonistic effect on the κ type. Buprenorphine at lower doses produces sufficient agonist effect to enable opioid-addicted individuals to discontinue the use of opioids without experiencing withdrawal symptoms. It has been reported that an intravenous dose of 0.3 mg of buprenorphine is equivalent to 10 mg of morphine and that oral delivery of buprenorphine results in a bioavailability of only about 15% due to first-pass metabolism. In addition to the intravenous formulation, there is a sublingual formulation offering the advantage of avoiding the first-pass metabolism effect. A clinical trial of nasally administered buprenorphine was reported in 1989 [114]. Buprenorphine is highly lipophilic and hence easily absorbed across the nasal epithelium. The buprenorphine formulation used in the clinical trial was prepared in 5% dextrose solution and a Pfeiffer atomizing pump operated manually was used to deliver it. The mean $T_{max}$ and $C_{max}$ for the intranasal dose were 30.6 min and 1.77 ng/mL, respectively. A relative nasal bioavailability of about 48.2% was attained.

Butorphanol, an analog of buprenorphine, showed a nasal bioavailability of 70% and also a much lower $T_{max}$ after nasal absorption as compared with the sublingual and buccal routes [115]. Lindhardt et al. [106] compared buprenorphine formulated in 30% PEG-300 in sheep with that of the 5% dextrose formulation mentioned earlier. A unit-dose Pfeiffer device was again used to administer the formulation. It was found that nasal bioavailability in sheep was about 70% when buprenorphine was formulated in PEG-300 and approximately 89% when it was formulated with 5% dextrose. The rate of absorption was reported to be very fast, with a $T_{max}$ of 10 min; the $C_{max}$ was found to be 37 and 48 ng/mL for PEG-300 and dextrose, respectively. In sheep, the pharmacokinetics of buprenorphine showed a two-compartment model as compared to a three-compartment model in humans.

### 7.4 Hydralazine

Hydralazine is a vasodilator used in the treatment of malignant hypertension and hypertensive emergencies and is generally used in conjunction with other antihypertensive agents. Although the oral absorption is good, there is low oral bioavailability due to first-pass metabolism. Nasal absorption of hydralazine has been studied in rats using both in vivo nasal absorption and in situ nasal perfusion methods; the effect of surfactant and solution pH has also been reported. It was found that the nasal absorption of hydralazine was increased in the presence of surfactants such as sodium glycocholate and polyoxyethylene-9-lauryl ether. The nasal absorption of hydralazine was reported to be a pH-dependent passive process, with the absorption increasing as the pH was increased from 3.0 to 6.6. In nasal absorption studies in rats, peak levels of hydralazine were reached in 30 min at pH 3.0. The in situ absorption of hydralazine as a function of perfusion pH was also evaluated. The results of the in situ nasal perfusion studies demonstrated that hydralazine is eliminated from the nasal cavity and the perfusate by first-order kinetics. Even in the ionized form, the drug was well absorbed, and it was suggested that the aqueous channels in the nasal mucosa played an important role in the transport of hydralazine [116, 117].

### 7.5 Nitroglycerin

Nitroglycerin is delivered across the mucosal membranes in the management of acute ischemic conditions. Nitroglycerin carries out this function by arterial vasodilatation and venodilation, which leads to a decrease in both the preload and afterload and also improved coronary blood flow. The intranasal action of nitroglycerin, also called glyceryl trinitrate, appears to be similar [118] and brings about a reduction in myocardial oxygen consumption. Like that of hydralazine, the oral bioavailability of nitroglycerin is low; hence alternative routes of delivery and innovative delivery systems have been preferred, such as sublingual patches, ointments, or transdermal patches. The intravenous route ensures a rapid onset of action, but its preparation and standardization procedures make it costly. Intranasal nitroglycerin in various operative experiences has been found to have a rapid onset of action with predictable and consistent therapeutic effects. A peak level of nitroglycerin is reached 1–2 min after intranasal administration; it is barely detectable after 16 min. These studies were carried out in five patients who were undergoing elective coro-

nary artery bypass surgery. The plasma levels were reported to be similar with intravenous administration and better than with sublingual administration [118–120]. The pressor response to endotracheal intubation in both normotensive and hypertensive patients can be attenuated by intranasal nitroglycerin in operative settings. It has been reported that when nitroglycerin was given 30 s before the induction sequence in 40 hypertensive patients treated with β blockers, there was a blunted pressor response to intubation. In comparison to the placebo control group, the group that received nitroglycerin had a lower mean arterial pressure at 1, 3, and 5 min after induction. Thus intranasal nitroglycerin can be employed in the selective control of hypertension [118, 121, 122].

### 7.6 Propranolol and Other β-Adrenergic Blocking Agents

Propranolol is a nonselective β-adrenergic receptor blocking agent. It is clinically used in the management of hypertension and the treatment of angina pectoris. When given orally in humans, it has led to considerable variation in plasma drug levels. This, as well as its subsequently low bioavailability, is believed to be due to its extensive metabolism in the gut and in the liver. A study in the late 1970s showed that nasal absorption of propranolol at a dose of 1 mg in rats produced blood levels similar to those achieved with intravenous administration; however, the same dose administered orally resulted in very low blood levels [47, 123]. The feasibility of nasal absorption of propranolol in solutions and sustained-release formulations in rats and dogs was studied. The procedure described by Hirai et al. [46] was used to carry out the surgical operation in rats. In dogs, the formulations were given intranasally using a micropipette and syringe. The mean blood levels of propranolol by the nasal route were compared with those of the oral and intravenous routes in rats and dogs. As in the case of the results obtained in humans, oral administration of propranolol solutions resulted in low and variable drug levels in rats and dogs, whereas the nasal administrations of propranolol solution showed plasma drug levels that were similar to those achieved with intravenous administration. In the case of sustained-release formulations, it was found that there was an initial low level of drug; however, these levels were maintained for a longer time. The bioavailabilities obtained from the AUC were found to be identical, although the maximum blood levels in the case of sustained-release formulations were found to be much lower than with the propranolol solutions. A propranolol formulation containing 2% methylcellulose gels in humans was studied by the same group. Identical serum drug profiles were obtained after nasal administration as with intravenous administration [124].

The effect of intranasal propranolol on exercise parameters with the Bruce protocol [118, 125] was studied in 16 patients with chronic, stable, effort-induced angina pectoris. Propranolol was given as a single 5-mg/puff nasal spray to the patients. A mean plasma level of 20 ng/mL was obtained, and a significant increase in total exercise time was seen, from 460 to 530 s. This led to an increase in the time to 1 mm ECG as well as to the onset of angina. Both maximum heart rate and systolic blood pressure were lower than with placebo. This study demonstrated that propranolol in the form of a nasal spray elicited immediate β blockade and was useful in treating patients with angina pectoris, who showed improvement in exercise tolerance after receiving the drug [118, 125].

Although no adverse reactions have been reported with intranasal administration of propranolol, complications may occur, as ocular administration has produced some systemic side effects [118]. The influence of substrate lipophilicity on drug uptake by the nasal route was reported in humans. Alprenolol and metoprolol, β blockers with varying degrees of lipophilicity, were used. The findings from these studies demonstrate that the more hydrophilic drugs showed a lower bioavailability. Alprenolol showed rapid uptake into the systemic circulation by the nasal route and also a higher bioavailability [126, 127].

### 7.7 Sex Hormones

The low oral bioavailability of hormone-replacement drugs due to intestinal and first-pass metabolism requires the use of higher doses of these drugs, which are associated with many side effects. Parenteral administration of sex steroids as well as use of the transdermal route has been viewed as an alternative. However, the transdermal route has certain limitations, such as the visibility and palpability of the patch as well as possible skin irritation. These drawbacks have limited the use of this route. The intranasal route has therefore been considered as an alternative [128].

### 7.8 17β-Estradiol ($E_2$)

The most common form of estrogen in clinical use is 17β-estradiol. It is reported to reduce bone turnover, prevent postmenopausal bone loss, and decrease the risk of fracture in both early and late postmenopausal women. Lipophilic drugs such as sex steroids pose the problem of going into solution, thus leading to low estrogen levels. Hence a number of studies have attempted to solubilize 17β-estradiol using DMβ-CD for its intranasal delivery [129, 130]. An $E_2$ spray has been developed (S21400 or Aerodiol, Institut de Recherches Internationales Servier, France), which is estradiol that has been solubilized in water with randomly methylated β-CD. Estradiol delivered intranasally is rapidly absorbed by the nasal mucosa and shows maximum plasma levels within 10–30 min. Plasma levels return to 10% of the maximum plasma concentration within 2 h of administration and to untreated postmenopausal levels within 12 h. Hence after intranasal administration, estradiol has a pulsatile profile as compared to the more sustained plasma profile seen with the oral and transdermal routes. However, whether sustained levels are required for efficacy has not been determined, although intranasally delivered estradiol does increase serum estradiol to the same extent as is seen with oral administration [128, 130–134].

The short- and long-term effects of intranasal 17β-estradiol on bone marrow turnover and serum IGF-1 were studied in a double-blind placebo-controlled clinical trial and compared with oral 17β-estradiol. Some 425 Caucasian postmenopausal women with climacteric symptoms were studied. The nasal efficacy of estradiol was assessed using the Kupperman index (KI), which is a weighted evaluation of the incidence and severity of 11 menopausal symptoms summarized in a menopausal index as follows: hot flushes, the most heavily weighted (×4); night sweats (×2); and sleep disturbances and nervousness (each ×2). The lower weighted symptoms include depression, irritability, vertigo, fatigue, arthralgia, headache, tachycardia,

and vaginal dryness (each ×1). The highest possible score is 51 [128, 135–137]. Various markers of bone resorption—such as urinary type I collagen telopeptides, the formation of serum osteocalcin, serum type I collagen, N–terminal extension propeptide (PINP), and serum bone marrow alkaline phosphatase (BAP)—were determined at baseline and after 1, 3, and 15 months. Urinary-type collagen C telopeptides were considerably reduced in all the treated groups within 1 month, and this reduction continued even at 3 months. Neither serum osteocalcin nor PINP showed any change at 1 month; however, they were reduced at 3 months with oral dosage. There was an increase in the bone formation parameters at 1 month for the higher doses of intranasal estradiol, but no reduction was seen at 3 months. No significant change from placebo-treated groups was observed at the end of 3 months in the case of circulating IGF-I after intranasal estradiol, but a significant decrease was seen with oral estradiol. After a year of treatment with intranasal estradiol at a dose of 300 μg/day, resorption and formation markers decreased to premenopausal levels. This study concluded that normalization of bone turnover to premenopausal levels can be achieved following 1 year of treatment with intranasal 17β-estradiol [131].

### 7.9 Testosterone

Testosterone, the most potent natural male sex hormone, is generally given intramuscularly. It is absorbed well orally but is extensively metabolized in the liver and the gastrointestinal tract. Owing to fluctuations in serum levels of testosterone esters, other viable routes of delivery are being explored. Two transdermal patches have become available commercially. However, due to the limitations associated with this route, it is not preferred by patients, particularly because the site of application is the scrotum [8, 138]. In order to improve systemic bioavailability, the nasal absorption of testosterone has been evaluated versus intravenous and intraduodenal administration in rats. When given nasally, the concentration of testosterone in the circulation increased and peak levels were reached within 2 min; blood levels were similar to those seen with intravenous administration. Intraduodenal administration produced low blood levels. A bioavailability of 99% at 25 μg/dose was seen and 90% at 50 μg, but the intraduodenal route showed a very low bioavailability of 1% [8]. Hussain et al. [139] showed in rats that testosterone can be absorbed intranasally, and an elimination half-life of about 40 min was obtained. However, since the problem of solubility is an obstacle to preparing the formulation for nasal administration, the use of a prodrug was evaluated by the same group. A water-soluble ester of testosterone, testosterone 17β-*N*, *N*-dimethylglycinate hydrochloride (TSDG), is completely absorbed when given intranasally at much lower doses as compared with testosterone itself. After absorption, conversion of the prodrug to testosterone begins almost immediately, and the terminal elimination half-life of testosterone was found to be 55 min, which is similar to that obtained after intravenous administration. Peak plasma concentration was reached within 12 min for the lower dose (equivalent to 25 mg/kg of TS) and 20 min for the higher dose (50 mg/kg) by both routes. The AUC also showed similarity, suggesting the complete absorption of the nasally administered prodrug.

## 8 THE NOSE: OPTION FOR DELIVERY OF DRUGS TO CENTRAL NERVOUS SYSTEM

The possibility of delivering agents to the central nervous system (CNS) via the nose has long been known, an example being the sniffing of cocaine in order to produce a sense of euphoria, which is attained within 3–5 min [53]. In short, nasal administration is not only an exciting possibility in the field of drug delivery but may also be the means of solving delivery problems for the innumerable agents that cannot cross the blood–brain barrier (BBB). Such agents are therefore being developed by nanoparticle-based systems or by formulating them as prodrugs. The literature offers numerous examples demonstrating ways of delivering such drugs to the brain. In particular, use of a direct pathway, as in the case of cocaine, from the nasal cavity to the CNS has been suggested. Illum [140] and Chow et al. [141] have shown in animal models that, in the early time points after nasal administration, the brain concentration of cocaine was higher than when the drug was given by the intravenous route. The specific site through which nose-to-brain delivery is believed to take place is the olfactory region. In the early 1900s it was shown that the olfactory region was responsible for the uptake (or rather entry) of viruses, in particular the poliomyelitis virus, into the brain [142–146]. Further work in support of this theory then demonstrated the presence of the poliomyelitis virus in the cerebrospinal fluid (CSF) and also in the systemic lymphatics of the olfactory mucosa [147–150]. In the period from 1970 to 1990, there were many reports of nose-to-brain delivery across the olfactory epithelium for a number of different agents, including metals and tracer materials such as colloidal gold, cadmium, potassium ferricyanide, and iron ammonium citrate [151–154].

There are certain aspects of drug delivery that must be clearly understood in designing a nose-to-brain drug delivery system. For example, lipophilic drugs are absorbed across the nasal epithelium almost immediately and efficiently to enter the systemic circulation. Therefore such drugs will show little sign of direct nose-to-brain delivery. Drugs that are on the hydrophilic side or polar molecules will not be readily absorbed across the nasal mucosa; these molecules generally undergo paracellular transport as compared to transcellular transport in the case of lipophilic drugs. Such molecules have a higher chance of being taken up by the olfactory mucosa for delivery to the brain. In general, drugs travel from the nose to the brain via (1) drug internalization into the olfactory epithelium's primary neurons followed by the intracellular axonal transport to the olfactory bulb or (2) drug absorption by paracellular or transcellular pathways across the olfactory sustentacular epithelial cells, following which the drug enters the CSF or CNS. In the human nervous system, the olfactory region is the only site in direct contact with the surrounding environment. It has been reported that the intracellular axonal pathway takes a longer time to deliver drugs to the brain [140, 155–157]. Table 4 gives an account of some of the drugs/molecules that have been delivered from the nose to the brain in human and animal models. The direct delivery of drugs to the brain or CSF via the olfactory epithelium is discussed in the following paragraph, briefly describing the olfactory mucosa.

The olfactory region is mainly involved in the detection of smell, and the makeup and organization of the epithelial layer enhance the accessibility of air to the

**TABLE 4 Nose-to-Brain Delivery of Agents in Different Species**

| Drug/Molecule | Path Followed | Reference |
|---|---|---|
| | *Humans* | |
| Arginine-vasopressin | —[a] | 158 |
| Adrenocorticotrophin | —[a] | 159 |
| Cholecystokinin | —[a] | 160 |
| Diazepam | —[a] | 161 |
| Insulin | —[a] | 158, 162 |
| technetium-99m-diethylenetriaminepentacetic acid | Direct from nose to brain | 163 |
| Apomorphine | Nasal cavity → CSF | 164 |
| Melatonin/hydroxycobalamin | Nasal cavity → CSF | 165 |
| Melanocortin, vasopressin, and insulin | Nasal cavity → CSF | 166 |
| | *Rats* | |
| Zidovudine | Nasal cavity → CSF, nasal cavity → systemic circulation | 167 |
| Dextromethorphan HCl | Direct from nose to brain | 168 |
| Cephalexin | Nasal cavity → CSF, nasal cavity → systemic circulation | 169 |
| Sulfonamides | Nasal cavity → CSF, nasal cavity → systemic circulation | 170, 171 |
| Dextran | Nasal cavity → olfactory mucosa → CNS and also nasal cavity → systemic circulation | 172 |
| Dopamine | Direct from nose to brain | 173 |
| Cocaine | Direct from nose to brain | 141 |
| Dihydroergotamine | Direct from nose to brain | 174 |
| Insulin | Direct from nose to brain | 175 |
| Dopamine | Nasal cavity → CSF | 173 |
| Methotrexate | Nasal cavity → CSF | 176 |
| Nerve growth factor | Direct from nose to brain via olfactory pathway | 177 |
| WGA–HRP (wheat germ agglutinin–horseradish peroxidase) | Direct from nose to brain via olfactory nerve and bulb | 178 |
| Zolmitriptan | Direct from nose to brain | 179 |
| Leptin | Direct from nose to brain | 180 |
| Morphine | Direct nose to brain via olfactory pathway | 181 |
| Nimodipine | Direct nose to brain via olfactory bulb and nasal cavity → CSF | 182 |
| Meptazinol hydrochloride | Nasal cavity → CSF | 183 |
| Estradiol | Nasal cavity → CSF via olfactory neurons | 184 |

[a]Facilitated transport to brain based on functional evidence in humans.

neuronal components comprising the odorant detectors. The olfactory region is mainly located on the nasal septum and partly on the superior and middle turbinates. It occupies only a small region in humans of about 10 $cm^2$ in the roof of the nasal cavity, as compared to around 150 $cm^2$ in dogs. The olfactory epithelium is a

pseudostratified epithelium comprising three types of cells: olfactory receptor cells, supporting cells, and basal cells. The receptor cells are elongated bipolar neurons located in the middle stratum of the epithelium, interspersed among the sustentacular cells; the microvilli cover the supporting cells, which extend from the mucosal surface to the basal membrane; while the basal cells are located in the basal surface of the neuroepithelium. These basal cells go on to differentiate, becoming new receptor cells [140, 185, 186]. The surface of the nasal cavity measures about 180 $cm^2$ in humans as compared to about 10 $cm^2$ in rats, and the olfactory region is reported to constitute about 3% of the nasal cavity in humans and 50% in rats. Some other differences include the fact that in adult humans the volume of CSF is about 160 mL while it is only about 150 μL in rats; also, the CSF is replaced about three times daily in humans, whereas in rats it is replaced hourly. Hence, though there is sufficient evidence regarding nose-to-brain delivery, especially in rats and in some cases in humans, the impact of these factors on the interpretation of the results between the two species could be significant [155, 187].

## 9 NASAL DELIVERY OF VACCINES

The discovery of vaccines for smallpox, cholera, and typhoid and the variety of vaccines now available have led to a significant reduction in the mortality and morbidity due to many diseases, with smallpox being the first to have been completely eradicated and poliomyelitis targeted to be the next. At present, the World Health Organization is working toward the complete elimination of poliomyelitis throughout the world [188, 189]. However, since Jenner discovered the vaccine for smallpox more than two centuries ago [190], only some 50 vaccines have been approved for use, and few additional vaccines have been discovered. Most of those in current use are administered parenterally; they can induce only a systemic immune response, not mucosal immunity. Obviously the latter is very important in the prevention and treatment of infectious diseases, be they due to viral, bacterial, or parasitic pathogens that attack via the mucosal surfaces [190].

The criteria to be met in designing a vaccine formulation include the following: The vaccine should have the capacity to produce lifelong immunity, be able to act against the different strains and variants or the subtypes of the organisms, be effective in all age groups, be able to act quickly and also to induce immunity in the fetus when the mother is treated, be able to act effectively after a single treatment, and ideally be administered noninvasively. Finally, such a vaccine must be relatively inexpensive and remain active under a variety of conditions, especially not requiring the cold chain [191].

The following section addresses the need for needle-free vaccines with formulations based on safer adjuvant and delivery systems.

### 9.1 Nasal Vaccines: Ideal Noninvasive Route

When we talk about targeting the pathogens entering through the mucosal surfaces, the route that usually comes to mind is the oral route. However, this route has its drawbacks for several reasons, such as the fact that the antigen used is degraded along with the gastric contents; furthermore, there is also the difficulty of reaching the antigen-presenting cells [192]. In comparison to the oral route, nasal vaccination

has been shown to require a lower antigen dose, which is essential considering the cost of the recombinant agents that may be used as antigens. Nasal administration has been shown to induce immune responses in the respiratory and genital surfaces. When compared with the other mucosal routes such as vaginal or rectal administration, the nasal route is much more acceptable in terms of both accessibility and overall convenience. Hence the nasal route is increasingly being seen as ideal for the administration of vaccines.

The nasal route has traditionally been used as an effective route in the treatment of respiratory infections, the rationale having been to target the infectious agents at their port of entry. Since most infectious disease pathogens enter at various mucosal sites, the nasal route has attracted increased attention as an alternative route for the delivery of vaccines. A further advantage offered by the nasal route in that it is capable of inducing both systemic and mucosal immunity as compared to the parenteral route, which brings about only systemic immunity [190, 193, 194]. Nasal immunization can result in distant as well as local mucosal immunity because of the mucosal immune system's common properties. This means that it is possible, via nasal immunization, to attain adequate immunity at other mucosal sites such as the respiratory, intestinal, and genital; hence vaccines administered nasally will have an important role in the prevention of respiratory infections and, more importantly, in the treatment of sexually transmitted diseases [190–196].

The earliest vaccines were live attenuated, inactivated toxins, or inactivated toxoids. But with advances in molecular biotechnology, it has become possible to produce extremely pure vaccines. However, the main drawbacks of these vaccines are their poor immunogenicity, so that the use of adjuvants is often required to attain the necessary immunogenicity. A vaccine adjuvant is especially important in subunit vaccines, which is how most of the vaccines available today are supplied. Adjuvants can be best defined as particular agents that increase the immunity produced when they are coformulated and delivered with the vaccine antigen. The adjuvants are formulated in the vaccines so as to produce a longer lasting immunity. The more time the adjuvant takes to be eliminated from the system, the longer it is able to induce the required or intended lasting immune response.

Vaccine adjuvants may be classified operationally as delivery or immunostimulating adjuvants. Vaccine delivery adjuvants simply act as delivery agents, that is, they present the vaccine antigens to the antigen-presenting cells (APCs), whereas the immune-stimulating agents act by stimulating the APCs to elicit an appropriate immune response. This results from the stimulation of the Toll-like receptors, present on macrophages and dendritic cells, the activation of which shows increased antigen presentation and cytokine release. The combination of both types of adjuvants may lead to better formulations for nasal vaccines [196–199].

### 9.2 Immunity after Intranasal Immunization

An understanding of the respiratory tract and the immune response following nasal vaccination is necessary to understand how the antigen used in the vaccine interacts with the surfaces of the human body and how the different adjuvants may interact, modify, and aid in generating an immune response. The nose is a component of the upper respiratory tract, which is composed of the mouth, nasopharynx, and larynx. The nasal passages have an extensive surface area which is richly vascularized.

However, the nasal epithelium has little ability to break down drugs. The extensive mucosal surface of the nose has a lining of pseudostratified epithelium as well as cilia and the goblet cells involved in the secretion of mucus. The lymphoid tissue primarily involved in the mucosal immune responses is the mucosal-associated lymphoid tissue (MALT). The different regions of the respiratory tract that play an influencing role in the immune system are as follows:

- The epithelial surface, which comprises immunocompetent cells in the connective tissue
- The lymphoid tissues linked to the respiratory tract, which are categorized into three parts: larynx-associated lymphoid tissue (LALT), nose-associated lymphoid tissue (NALT), and bronchus-associated lymphoid tissue
- The lymph nodes that drain the respiratory tract

The NALT, which is the organized lymphoid structure in the nasal passages and occurs in abundance in the nasal mucosa, plays a significant role in the mucosal surface's defense against invading pathogens. The NALT is equivalent to Waldeyer's ring, which is made up of the adenoids or tonsils situated in the roof of the nasopharynx, bilateral lymphoid bands, the palatine tonsils, and the tonsil at the base of the tongue (the lingual tonsil). The NALT is the main target site for vaccine antigens in humans [193, 194, 197, 200, 201]. As the mucosal immune system develops, the NALT is involved in a variety of important functions in relation to the host defense mechanism, from being an impediment to drug absorption, serving as a guard against attacking pathogens and antigens, facilitating the uptake of antigens, eliciting the secretory antibody response, and inducing the immune response in the other distant mucosal surfaces due to the function of the common mucosal immune system. Owing to mucosal tolerance, it is also involved in preventing serious allergic responses to inhaled antigens [194, 201, 202].

On administration, the antigen interacts first with the inductive tissue of the MALT, thus initiating a primary response. The IgA serum cells are found in the effector sites of the MALT, and local immunity results from the production of the secretory IgA (s-IgA) response. As mentioned earlier, the NALT and the tonsils form the main inductive sites in rodents and humans. These are composed of M cells involved in the uptake of antigen and its presentation to the underlying lymphoid tissues, the antigen being taken up by the M cells and the APCs, consisting of dendritic cells, macrophages, and B cells; all these cells together with the T cells produce the cellular and humoral immune responses [194].

Vaccination by the nasal route produces a mucosal protection using mucus, the epithelial surface, and both innate and acquired immune responses. The innate defense mechanism plays a very important role in that it influences the type of acquired immune mechanism, which mainly responds on the basis of "immune memory." The ability to attain these responses is the main principle of attaining protection from infection.

### 9.3 Need for Adjuvants

As mentioned earlier, the subunit vaccines in particular require the use of adjuvants in order to initiate an immune response leading to protection. The subunit vaccines

**TABLE 5 Nasal Vaccination Delivery Systems Studied**

| Delivery System/Adjuvant | Antigen Employed | Remarks | Reference |
|---|---|---|---|
| Chitosan | Diphtheria toxin | Increased local as well as systemic effects | 208 |
| | Influenza virus | Immune response comparable to that of intramuscular administration | 209 |
| | Filamentous hemagglutinin and recombinant pertussis toxin (single or bivalent vaccine) | Chitosan stimulates mucosal and systemic effects | 209, 210 |
| IL-12 | Tetanus toxoid | IL-12-induced IgA response | 211 |
| | Influenza hemagglutinin and neuraminidase | Induces protective immunity | 212 |
| CpG motifs | Hepatitis B surface antigen (HBsAg) | Potent enhancement of systemic and mucosal immune responses | 213 |
| | Formalin-inactivated influenza virus | Enhances the serum IgG and s-IgA responses | 214 |
| Liposomes | Influenza subunit vaccine | Induction of systemic IgG and s-IgA responses | 215 |
| | Inactivated measles virus | Stimulation of mucosal and systemic responses | 216 |
| | Bovine serum albumin (BSA) | — | 217 |
| | *Streptococcus mutans* | — | 218 |
| | Influenza hemagglutinin and neuraminidase (when used in combination with heat-labile toxin (HLT) | Good response in presence of HLT | 219 |
| | Influenza, hepatitis B, tetanus toxoid | — | 220 |
| | *Yersinia pestis* | — | 221 |
| ISCOMs (immune stimulating complex) | Influenza subunit | Protective immunity to challenge | 222 |
| | Measles nucleoprotein | Induces cytotoxic T-cell response | 223 |
| | *Echinococcus* surface antigen | — | 224 |
| | Respiratory syncytial virus (RSV) envelope antigen | | 225 |
| Poly(lactic-*co*-glycolic acid (PLGA) microparticles | HBsAg | Strong systemic and mucosal immune responses | 226 |
| Cationic nanoparticles (SMBV) | HBsAg and β-galactosidase | Strong mucosal as well as systemic antibody and CTL responses | 227 |

**TABLE 5** ***Continued***

| Delivery System/Adjuvant | Antigen Employed | Remarks | Reference |
|---|---|---|---|
| HBcAg | HBsAg | Stimulates strong Th1 response | 228 |
| Cholera toxin | Group B streptococci | High levels of IgA in cervicovaginal secretions | 229 |
| | *Haemophilus influenzae* | Effective nasal vaccination | 230 |
| | Influenza virus | — | 231 |
| | Synthetic peptide of RSV | Complete protection | 232 |

available today, which are administered intramuscularly or subcutaneously, involve alum as an adjuvant. The drive for initiating further research into vaccine adjuvants has been stimulated by many factors, among the chief of which is that the aluminum-based adjuvants currently available have failed in many candidate vaccines or have not achieved the necessary immunity or induced a cytotoxic T-cell response. Nasally formulated vaccines, mainly subunit vaccines, currently being purified are less immunogenic and also cannot elicit the necessary T-cell response. As a result, research is now focusing on finding newer adjuvants for nasal DNA and subunit vaccines in order to attain specific immune responses as well as the necessary antibody subtype response. In addition to this, an adjuvant can help to reduce the dose of antigen required and also the number of doses needed to achieve mucosal immunity [203–207]. Despite the extensive research going on in the field of vaccine adjuvants, the only FDA-approved adjuvant for human use is alum. There are several other adjuvants, such as monophosphoryl lipid A (MPL), that have been approved in the European market; another, Corixa, is used as an adjuvant in Fendrix, the hepatitis B vaccine of GlaxoSmithKline Biologicals. The main hindrance to the approval of many adjuvants that reach clinical trials is their potential to elicit toxic side effects in clinical use. It has been reported that aluminum salts do induce some allergies in humans. As more purified and target-oriented or specific vaccines obtained by recombinant technology are being launched, it becomes more difficult for vaccine antigens alone to induce the necessary immune responses, as these recombinant antigens or synthetic peptides cannot jump start an immune response.

A number of adjuvants are awaiting approval for human use. The main impediment to the successful development of vaccine adjuvants is that their mechanism of action is not clearly understood. Table 5 offers a list of available nasal drug delivery systems and the various adjuvants that have been used in the development of nasal vaccines.

## REFERENCES

1. Southall, J., and Ellis, C. (2000), Developments in nasal drug delivery, *Innovat. Pharm. Technol.*, 110–115.

2. Koch, M. (2003), The growing market for nasal drug delivery, *Pharmatech*, 1–3.
3. Ruppar, D. (2006), Intranasal delivery: Sniffing out new sources for growth, *Drug Deliv. Technol.*, 6, 0–33.
4. Illum, L., and Fisher, A. N. (1997), Intranasal delivery of peptides and proteins, in Adjei, A. L. and Gupta, P. K., Eds., *Inhalation Delivery of Therapeutic Peptides and Proteins*, Marcel Dekker, New York, pp. 135–184.
5. Ugwoke, M. I., Verbeke, N., and Kinget, R. (2001), The biopharmaceutical aspects of nasal mucoahdesive drug delivery, *J. Pharm. Pharmacol.*, 53, 3–21.
6. Illum, L. (1999), Bioadhesive formulations for nasal peptide delivery, in Mathiowitz, E., Chickering, D. E., and Lehr, C.-M., Eds., *Bioadhesive Drug Delivery Systems, Fundamentals, Novel Approaches and Development*, Marcel Dekker, New York, pp. 507–539.
7. Washington, N., Washington, C., and Wilson, C. G. (2001), *Physiological Pharmaceutics, Barriers to Drug Absorption*, Taylor and Francis, London.
8. Chien, Y. W., Su, K. S. E., and Chang, S.-F. (1989), *Nasal Systemic Drug Delivery*, Marcel Dekker, New York.
9. Behl, C. R., Pimplaskar, H. K., Sileno, A. P., deMeireles, J., and Romeo, V. D. (1998), Effects of physicochemical properties and other factors on systemic nasal drug delivery, *Adv. Drug Deliv. Rev.*, 29, 89–116.
10. Mygind, N., and Dahl, R. (1998), Anatomy, physiology and function of the nasal cavities in health and disease, *Adv. Drug Deliv. Rev.*, 29, 3–12.
11. Olson, P., and Bende, M. (1985), Influence of environmental temperature on human nasal mucosa, *Ann. Otol. Rhinol. Lryngol.*, 94, 153–157.
12. Paulsson, B., Bende, M., and Ohlin, P. (1985), Nasal mucosal blood flow at rest and during exercise, *Acta Otolaryngol. (Stockh.)*, 99, 140–144.
13. Hadley, W. M., and Dahl, A. R. (1983), Cytochrome P450-dependent mono-oxygenase in nasal membrane of six species, *Drug Metab. Dispos.*, 11, 275–279.
14. Longo, V., Pacifici, G. M., Panattoni, G., Ursino, F., and Gervasi, P. G. (1989), Metabolism of diethylnitrosamine by microsomes of human respiratory nasal mucosa and liver, *Biochem. Pharmacol.*, 38, 1867–1869.
15. Brittebo, E. G., and Rafter, J. J. (1984), Steroid metabolism by rat nasal mucosa: Studies on progesterone and testosterone, *Steroid. Biochem.*, 20, 1147.
16. Smith, E. L., Hill, R. L., and Borman, A. (1958), Activity of insulin degraded by leucine aminopeptidase, *Biochem. Biophys. Acta.*, 29, 207–214.
17. Marttin, E., Schipper, N. G. M., Verhoef, J. C., and Merkus, F. W. H. M. (1998), Nasal mucosal clearance as a factor in nasal drug delivery, *Adv. Drug Deliv. Rev.*, 29, 13–38.
18. Nagai, T., Nishimoto, Y., Nambu, N., Suzuki, Y., and Sekine, K. (1984), Powder dosage form of insulin for nasal administration, *J. Controlled Release*, 1, 15–22.
19. Nagai, T., and Machida, Y. (1990), Bioadhesive dosage forms for nasal administration, in Lenaerts, V., and Gurny, R., Eds., *Bioadhesive Drug Delivery Systems*, CRC Press, Boca Raton, FL, pp. 169–178.
20. Pritchard, K., Lansely, A. B., Martin, G. P., Helliwell, M., Marriott, C., and Benedetti, L. M. (1996), Evaluation of bioadhesive properties of hyaluronan derivatives, detachment weight and mucociliary transport rate studies, *Int. J. Pharm.*, 129, 137–145.
21. Marttin, E., Verhoef, J. C., Romeijn, S. G., Zwart, P., and Merkus, F. W. H. M. (1996), Acute histopathological effects of bezalkonium chloride and absorption enhancers on rat nasal epithelium in vivo, *Int. J. Pharm.*, 141, 151–160.
22. Ennis, R. D., Borden, L., and Lee, W. A. (1990), The effects of permeation enhancers on the surface morphology of rat nasal mucosa: A scanning electron microscopy study, *Pharm. Res.*, 7, 468–475.

23. Hardy, J. G., Lee, S. W., and Wilson, C. G. (1985), Intranasal drug delivery by spray and drops, *J. Pharm. Pharmacol.*, 37, 294–297.
24. Proctor, D. F. (1985), Nasal Physiology in intranasal drug administrations, in Chien, Y. W., Ed., *Transnasal Systemic Medications*, Elsevier, Amsterdam, pp. 101–106.
25. Greiff, L., Venge, P., Andersson, M., Enander, I., Linden, M., Myint, S., and Persson, C. G. (2002), Effects of rhinovirus-induced common colds on granulocyte activity in allergic rhinitis, *J. Infect.*, 45, 227–232.
26. Larsen, C., Niebuhr, J. M., Tommerup, B., Mygind, N., Dagrosa, E. E., Grigoleit, H. G., and Malerczyk, V. (1987), Influence of experimental rhinitis on the gonadotropin response to intranasal administration of buserelin, *Eur. J. Clin. Pharmacol.*, 33, 155–159.
27. Kublik, H., and Vidgren, M. T. (1998), Nasal delivery systems and their effect on deposition and absorption, *Adv. Drug Deliv. Rev.*, 29, 157–177.
28. Hussain, A. A., Bawarshi-Nassar, R., and Huang, C. H. (1985), Physicochemical considerations in intranasal drug administration, in Chien, Y. W., Ed., *Transnasal Systemic Medication*, Elsevier, Amsterdam, pp. 121–138.
29. Corbo, D. C., Huang, Y. C., and Chien, Y. W. (1989), Nasal delivery of progestational steroids in ovarectomized rabbits. II. Effect of penetrant hydrophilicity, *Int. J. Pharm.*, 50, 253–260.
30. Kimura, R., Miwa, M., Kato, Y., Sato, M., and Yamada, S. (1991), Relationship between nasal absorption and physicochemical properties of quaternary ammonium compounds, *Arch. Int. Pharmacodyn. Ther.*, 310, 13–21.
31. Tonndorf, J., Chinn, H. I., and Lett, J. E. (1953), Absorption from nasal mucous membrane: Systemic effect of hyoscine following intranasal administration, *Ann. Otol. Rhinol. Laryngol.*, 62, 630–634.
32. Rogerson, A., and Parr, G. D. (1990), Nasal drug delivery, in Florence, A. T., and Salole, E. G., Eds., *Topics in Pharmacy, Routes of Drug Administration*, Northants, England, pp. 1–29.
33. McMartin, C., Hutchinson, E. F., Hyde, R., and Peters, G. E. (1987), Analysis of structural requirements for the absorption of drugs and macromolecules from the nasal cavity, *J. Pharm. Sci.*, 76, 535–540.
34. Donovan, M. D., Flynn, G., and Amidon, G. (1990), Absorption of polyglycols 600 through 2000: The molecular weight dependence of gastrointestinal and nasal absorption, *Pharm. Res.*, 7, 863–868.
35. Fisher, A., Brown, K., Davis, S., Parr, G., and Smith, D. A. (1987), The effect of molecular size on the nasal absorption of water-soluble compounds in the albino rat, *J. Pharm. Pharmacol.*, 39, 357–362.
36. Fisher, A., Illum, L., Davis, S., and Schacht, E. (1992), Di-iodo-L-tyrosine labeled dextrans as molecular size markers for nasal absorption in the rat, *J. Pharm. Pharmacol.*, 44, 550–554.
37. Maitani, Y., Machida, Y., and Nagai, T. (1989), Influence of molecular weight and charge on nasal absorption of water soluble compounds in the albino rat, *Int. J. Pharm.*, 40, 23–27.
38. D'Souza, R., Mutalik, S., Venkatesh, M., Vidyasagar, S., and Udupa, N. (2005), Nasal insulin gel as an alternative to parenteral insulin: Formulation, preclinical and clinical studies, *AAPS PharmSciTech.*, 6, E184–E189.
39. Bommer, R., Kern, J., Hennes, K., and Zwisler, W. (2005), Preservative-free nasal drug delivery systems, *Drug Deliv. Technol.*, 5.

40. Giroux, M. (2005), Controlled Particle Dispersion™: Effective nasal delivery from a versatile, flexible technology platform, in *Nasal Drug Delivery: Rapid Onset via Convenient Route*, ONdrugdelivery, pp. 13–15.
41. Hwang, P., Woo, E., and Fong, K., Intranasal deposition of nebulized saline: A radionuclide distribution study, paper presented at the 50th annual meeting of the American Rhinologic Society, New York, NY, Sept. 2004.
42. Giroux, M., Hwang, P., and Prasad, A. (2005), Controlled Particle Dispersion™: Applying vortical flow to optimize nasal drug deposition, *Drug Deliv. Technol.*, 5, 44–49.
43. Djupesland, P. G., and Watts, J. (2005), Breath-actuated bidirectional delivery sets the nasal market on a new course, ONdrugdelivery, pp. 20–23.
44. Djupesland, P. G., Skretting, A., Winderen, M., and Holand, T. (2004), Bi-directional nasal delivery of aerosols can prevent lung deposition, *J. Aerosol. Med.*, 17, 249–259.
45. Keldmann, T. (2005), Advanced simplification of nasal delivery technology: Anatomy + innovative device = added value opportunity, ONdrugdelivery, pp. 4–7.
46. Hirai, S., Yashiki, T., and Matsuzawa, T., Nasal absorption of drugs. Effect of surfactants on the nasal absorption of insulin in rats, paper presented at the 98th annual meeting of the Pharmaceutical Society of Japan, Apr. 1978.
47. Hussain, A., Hirai, S., and Bawarshi, R. (1980), Nasal absorption of propranolol from different dosage forms by rats and dogs, *J. Pharm. Sci.*, 69, 1411–1413.
48. Pillion, D. J., Ahsan, F., Arnold, J. J., Balusubramanian, B. M., Piraner, O., and Meezan, E. (2002), Synthetic long-chain alkyl maltosides and alkyl sucrose esters as enhancers of nasal insulin absorption, *J. Pharm. Sci.*, 91, 1456–1462.
49. Ahsan, F., Arnold, J. J., Meezan, E., and Pillion, D. J. (2001), Enhanced bioavailability of calcitonin formulated with alkylglycosides following nasal and ocular administration in rats, *Pharm. Res.*, 18, 1742–1746.
50. Merkus, F. W. H. M., Verhoef, J. C., Marttin, E., Romeijn, S. G., van der Kuy, P. H. M., Hermens, W. A. J. J., and Schipper, N. G. M. (1999), Cyclodextrins in nasal drug delivery, *Adv. Drug Deliv. Rev.*, 6, 41–57.
51. Gizurarson, S. (1990), Animal models for intranasal drug delivery, *Acta. Pharm. Nord.*, 2, 105–122.
52. Behl, C. R., Pimplaskar, H. K., Sileno, A. P., Xia, W. J., Gries, W. J., de Meireles, J. C., and Romeo, V. D. (1998), Optimization of systemic nasal drug delivery with pharmaceutical ecipients, *Adv. Drug Deliv. Rev.*, 29, 117–133.
53. Illum, L. (2002), Nasal drug delivery: New developments and strategies, *Drug Discovery Today*, 7, 1184–1189.
54. Hinchcliffe, M., and Illum, L. (1999), Intranasal insulin delivery and therapy, *Adv. Drug Deliv. Rev.*, 35, 199–234.
55. Illum, L., and Fisher, A. N. (1997), Intranasal delivery of peptides and proteins, in Adjei, A. L., and Gupta, P. K., Eds., *Inhalation Delivery of Therapeutic Peptides and Proteins*, Marcel Dekker, New York, p. 135.
56. Lansley, A. B., and Martin, G. P. (2001), Nasal drug delivery, in Hillery, A. M., Lloyd, A. W., and Swarbrick, J., Eds., *Drug Delivery and Targeting*, 2nd ed., Taylor & Francis, New York, p. 237.
57. Triplitt, C. L., Reasner, C. A., and Isley, W. (2005), Diabetes mellitus, in Dipiro, J. T., Talbert, R. L., Yee, G. C., Matzke, G. R., Wells, B. G., and Posey, L. M. Eds., *Pharmacotherapy: A Pathophysiological Approach*, 6th ed., McGraw-Hill, New York, p. 1333.
58. Bliss, M. (1933), The history of insulin, *Diabetes Care*, 16, 4–7.
59. Woodyatt, R. T. (1922), The clinical use of insulin, *J. Metab. Res.*, 2, 793–801.

60. Collens, W. S., and Goldzeiher, M. A. (1932), Absorption of insulin by nasal mucous membranes, *Proc. Soc. Exp. Biol. Med.*, 29, 756.
61. Major, R. H. (1935), The intranasal application of insulin, *J. Lab. Clin. Med.*, 21, 278.
62. Major, R. H. (1936), The intranasal application of insulin, *Am. J. Med. Sci.*, 192, 257.
63. Gordon, G. S., Moses, A. C., Silver, R. D., Flier, J. S., and Carey, M. C. (1985), Nasal absorption of insulin: Enhancement by hydrophobic bile salts, *Proc. Natl. Acad. Sci. USA*, 82, 7419–7423.
64. Morimoto, K., Morisaka, K., and Kamada, A. (1985), Enhancement of nasal absorption of insulin and calcitonin using polyacrylic acid gel, *J. Pharm. Pharmacol.*, 37, 134–136.
65. Pillion, D. J., Atchison, J. A., Gargiulo, C., Wang, R. X., Wang, P., and Meezan, E. (1994), Insulin delivery in nosedrops: New formulations containing alkylglycosides, *Endocrinology*, 135, 2386–2391.
66. Pillion, D. J., Ahsan, F., Arnold, J. J., Balusubramanian, B. M., Piraner, O., and Meezan, E. (2002), Synthetic long-chain alkyl maltosides and alkyl sucrose esters as enhancers of nasal insulin absorption, *J. Pharm. Sci.*, 91, 1456–1462.
67. Ahsan, F., Arnold, J. J., Meezan, E., and Pillion, D. J. (2003), Sucrose cocoate, a component of cosmetic preparations, enhances nasal and ocular peptide absorption, *Int. J. Pharm.*, 251, 195–203.
68. Illum, L., Jabbal-Gill, I., Hinchcliffe, M., Fisher, A. N., and Davis, S. S. (2001), Chitosan as a novel nasal delivery system for vaccines, *Adv. Drug Deliv. Rev.*, 51, 81–96.
69. Illum, L. (1992), Nasal delivery of peptides, factors affecting nasal absorption, in *Topics in Pharmaceutical Sciences*, Medpharm Scientific, Stuttgart, p. 71.
70. Illum, L., Farraj, N. F., and Davis, S. S. (1994), Chitosan as a novel nasal delivery system for peptide drugs, *Pharm. Res.*, 11, 1186–1189.
71. Fernández-Urrusuno, R., Calvo, P., Remuñán-López, C., Vila-Jato, J. L., and Alonso, M. J. (1999), Enhancement of nasal absorption of insulin using chitosan nanoparticles, *Pharm. Res.*, 16, 1576–1581.
72. Dyer, A. M., Hinchcliffe, M., Watts, P., Castile, J., Jabbal-Gill, I., Nankervis, R., Smith, A., and Illum, L. (2002), Nasal delivery of insulin using novel chitosan based formulations: A comparative study in two animal models between simple chitosan formulations and chitosan nanoparticles, *Pharm. Res.*, 19, 998–1008.
73. Yu, S., Zhao, Y., Wu, F., Zhang, X., Lü, W., Zhang, H., and Zhang, Q. (2004), Nasal insulin delivery in the chitosan solution: In vitro and in vivo studies, *Int. J. Pharm.*, 281, 11–23.
74. Varshosaz, J., Sadrai, H., and Alingari, R. (2004), Nasal delivery of insulin using chitosan microspheres, *J. Microencapsul.*, 21, 761–774.
75. Varshosaz, J., Sadrai, H., and Heidari, A. (2006), Nasal delivery of insulin using bioadhesive chitosan gels, *Drug Deliv.*, 13, 31–38.
76. Pontiroli, A. E., Alberetto, M., and Pozza, G. (1985), Intranasal calcitonin and plasma calcium concentrations in normal subjects, *Br. Med. J.*, 290, 1390–1391.
77. Morimoto, K., Morisaka, K., and Kamada, A. (1985), Enhancement of nasal absorption of insulin and calcitonin using polyacrylic acid gel, *J. Pharm. Pharmacol.*, 37, 134–136.
78. Morimoto, K., Akatsuchi, H., Aikawa, R., Morishita, M., and Morisaka, K. (1984), Enhanced rectal absorption of [Asu1, 7]-eel calcitonin in rats using polyacrylic acid aqueous gel base, *J. Pharm Sci.*, 73, 1366–1368.
79. Lee, K. C., Tak, K. K., Park, M. O., Lee, J. T., Woo, B. H., Yoo, S. D., Lee, H. S., and DeLuca, P. P. (1999), Preparation and characterization of polyethylene-glycol-modified salmon calcitonins, *Pharm. Dev. Tech.*, 4, 269–275.

80. Miron, T., and Wilcheck, M. (1993), A simplified method for the preparation of succinimidyl carbonate polyethylene glycol for coupling to proteins, *Biconj. Chem.*, 4, 568–569.
81. Shin, B. S., Jung, J. H., Lee, K. C., and Yoo, S. D. (2004), Nasal absorption and pharmacokinetic disposition of salmon calcitonin modified with low molecular weight polyethylene glycol, *Chem. Pharm. Bull.*, 52, 957–960.
82. Na, D. H., Youn, Y. S., Park, E. J., Lee, J. M., Cho, O. R., Lee, K. R., Lee, S. D., Yoo, S. D., Deluca, P. P., and Lee, K. C. (2004), Stability of PEGylated salmon calcitonin in nasal mucosa, *J. Pharm. Sci.*, 93, 256–261.
83. Arnold, J. J., Ahsan, F., Meezan, E., and Pillion, D. J. (2002), Nasal administration of low molecular weight heparin, *J. Pharm. Sci.*, 91, 1707–1714.
84. Yang, T., Mustafa, F., and Ahsan, F. (2004), Alkanoylsucroses in nasal delivery of low molecular weight heparins: In-vivo absorption and reversibility studies in rats, *J. Pharm. Pharmacol.*, 56, 53–60.
85. Mustafa, F., Yang, T., Khan, M. A., and Ahsan, F. (2004), Chain length-dependent effects of alkylmaltosides on nasal absorption of enoxaparin, *J. Pharm. Sci.*, 93, 675–683.
86. Yang, T., Hussain, A., Paulson, J., Abbruscato, T. J., and Ahsan, F. (2004), Cyclodextrins in nasal delivery of low-molecular-weight heparins: In vivo and in vitro studies, *Pharm. Res.*, 21, 1127–1136.
87. Yang, T., Hussain, A., Bai, S., Khalil, I. A., Harashima, H., and Ahsan, F. (2006), Positively charged polyethylenimines enhance nasal absorption of the negatively charged drug, low molecular weight heparin, *J. Controlled Release*, 115, 289–297.
88. Mason, G. A., Garbutt, J. C., and Prange Jr., A. J. (2000), Thyrotropin releasing hormone: Focus On neurobiology, available: http://www.acnp.org/g4/GN401000048/CH048.html.
89. Yamamoto, M., and Shimizu, M. (1987), Effects of new TRH analogue, YM-14673 on the central nervous system, *Naunyn. Schmiedebergs. Arch. Pharmacol.*, 336, 561–565.
90. Kagatani, S., Inaba, N., Fukui, M., and Sonobe, T. (1998), Nasal absorption kinetic behavior of Azetirelin and its enhancement by acylcarnitines in rats, *Pharm. Res.*, 15, 77–81.
91. Mitsuma, T., and Nogimori, T. (1984), Changes in plasma thyrotrophin-releasing hormone, thyrotrophin, prolactin and thyroid hormone levels after intravenous, intranasal or rectal administration of synthetic thyrotrophin-releasing hormone in man, *Acta Endocrinol.*, 107, 207–212.
92. Laursen, T., Grandjean, T., Jorgensen, J. O., and Christiansen, J. S. (1996), Bioavailability and bioactivity of three different doses of nasal growth hormone (GH) administered to GH-deficient patients: Comparison with intravenous and subcutaneous administration, *Eur. J. Endocrinol.*, 135, 309–315.
93. Leitner, V. M., Guggi, D., Krauland A. H., and Bernkop-Schnurch, A. (2004), Nasal delivery of human growth hormone: In vitro and in vivo evaluation of a thiomer/glutathione microparticulate delivery system, *J. Controlled. Release*, 100, 87–95.
94. Hedin, L., Olsson, B., Diczfalusy, M., Flyg, C., Petersson, A. S., Rosberg, S., and Albertsson-Wikland, K. (1993), Intranasal administration of human growth hormone (hGH) in combination with a membrane permeation enhancer in patients with GH deficiency: A pharmacokinetic study, *J. Clin. Endocrinol. Metab.*, 76, 962–967.
95. Albertsson-Wikland, K., and Rosberg, S. (1988), Analyses of 24-hour growth hormone (GH) profiles in children, *J. Clin. Endocrinol. Metab.*, 67, 493–500.
96. Clark, R. G., Jansson, J. O., Isaksson, O., and Robinson, I. C. A. F. (1985), Intravenous growth hormone: Growth responses to patterned infusions in hypophysectomized rats, *J. Endocrinol.*, 104, 53–61.

97. Eden, S. (1979), Age and sex related differences in episodic growth hormone secretion in the rat, *Endocrinology*, 105, 555–560.
98. Vermehren, C., Hansen, H. S., and Thomsen, M. K. (1996), Time dependent effects of two absorption enhancers on the nasal absorption of growth hormone in rabbits, *Int. J. Pharm.*, 128, 239–250.
99. Clausen, A. E., and Schnürch-Bernkop, A. (2000), In vitro evaluation of permeation enhancing effect of thiolated polycarbophil, *J. Pharm. Sci.*, 89, 1253–1261.
100. Schnürch-Bernkop, A., and Steininger, S. (2000), Synthesis and characterization of mucoadhesive thiolated polymers, *Int. J. Pharm.*, 194, 239–247.
101. Schnürch-Bernkop, A., Zarti, H., and Walker, G. F. (2001), Thiolation of polycarbophil enhances its inhibition of soluble and intestinal brush border membrane bound aminopeptidase N, *J. Pharm. Sci.*, 90, 1907–1914.
102. Clausen, A. E., Kast, C. E., and Schnürch-Bernkop, A. (2002), The role of glutathione in the permeation enhancing effect of thiolated polymers, *Pharm. Res.*, 19, 602–608.
103. Schnürch-Bernkop, A., Guggi, D., and Pinter, Y. (2004), Thiolated chitosans: Development and in vitro evaluation of a mucoadhesive, permeation enhancing oral drug delivery system, *J. Controlled Release*, 94, 177–186.
104. Illum, L. (1996), Animal models for nasal delivery, *J. Drug Target.*, 3, 717–724.
105. Illum, L., Watts, P., Fisher, A. N., Hinchcliffe, M., Norbury, H., Jabbal-Gill, I., Nankervis, R., and Davis, S. S. (2002), Intranasal delivery of morphine, *J. Pharmcol. Exp. Ther.*, 301, 391–400.
106. Lindhardt, K., Ólafsson, D. R., Gizurarson, S., and Bechgaard, E. (2002), Intranasal bioavailability of diazepam in sheep correlated to rabbit and man, *Int. J. Pharm.*, 231, 67–72.
107. Kyrkou, M., Harbord, M., Kyrkou, N., Kay, D., and Coulthard, K. (2006), Community use of intranasal midazolam for managing prolonged seizures, *J. Intellect. Dev. Disabil.*, 31, 131–138.
108. Lui, C. Y., Amidon, G. L., and Goldberg, A. (1991), Intranasal absorption of flurazepam, midazolam and triazolam in dogs, *J. Pharm. Sci.*, 80, 1125–1129.
109. Olivier, J. C., Djilani, M., Fahmy, S., and Couet, W. (2001), In situ nasal absorption of midazolam in rats, *Int. J. Pharm.*, 213, 187–192.
110. Burstein, A. H., Modica, R., Hatton, M., Forest, A., and Gengo, F. M. (1997), Pharmcokinetics and pharmacodynamics of midazolam after intranasal administration, *J. Clin. Pharmacol.*, 37, 711–718.
111. Knoester, P. D., Jonker, D. M., van der Hoeven, R. T. M., Vermeij, A. C., Edelbroek, P. M., Brekelmans, G. J., and de Haan, G. J. (2002), Pharmacokinetics and pharmacodynamics of midazolam administered as a concentrated intranasal spray. A study in healthy volunteers, *Br. J. Clin. Pharmcol.*, 53, 501–507.
112. Allonen, H., Ziegler, G., and Klotz, U. (1981), Midazolam kinetics, *Clin. Pharmacol. Ther.*, 30, 653–661.
113. Lindhardt, K., Gizurarson, S., Stefansson, S., Olafsson, D. R., and Bechgaard, E. (2001), Electroencephalographic effects and blood levels after administration of diazepam to humans, *Br. J. Clin. Pharmcol.*, 52, 1–12.
114. Eriksen, J., Jensen, N. H., Kamp-Jensen, M., Bjarno, H., Friis, P., and Brewster, D. (1989), The systemic availability of buprenorphine administered by nasal spray, J. Pharm. Pharmacol., 41, 803–805.
115. Shyu, W. C., Mayol, R. F., Pfeffer, M., Pittman, K. A., and Barbhiya, R. H. (1993), Biopharmaceutical evaluation of transnasal, sublingual and buccal disk dosage forms of butorphanol, *Biopharm. Drug Dispos.*, 14, 371–379.

116. Kaneo, Y. (1983), Absorption from the nasal mucous membrane: I. Nasal absorption of hydralazine in rats, *Acta. Pharm. Suec.*, 20, 379–388.
117. Hirai, S., Yashiki, T., Matsuzawa, T., and Mima, H. (1981), Absorption of drugs from the nasal mucosa of rats, *Int. J. Pharm.*, 7, 317.
118. Landau, A. J., Eberhardt, R. T., and Frishman, W. H. (1994), Intranasal delivery of cardiovascular agents: An innovative approach to cardiovascular pharmacotherapy, *Am. Heart J.*, 127, 1594–1599.
119. Hill, A. B., Bowley, C. J., Nahrwold, M. L., Knight, P. R., Krish, M. M., and Denlinger, J. K. (1981), Intranasal administration of nitroglycerin, *Anethesiology*, 54, 346–348.
120. Krishnamoorthy, R., and Mitra, A. K. (1998), Prodrugs for nasal drug delivery, *Adv. Drug Deliv. Rev.*, 29, 135–146.
121. Fossoulaki, A., and Kaniaris, P. (1983), Intranasal administration of nitroglycerin attenuates the pressor response to laryngoscopy and intubation of the trachea, *Br. J. Anaesth.*, 55, 49–51.
122. Grover, V. K., Sharma, S., Mahajan, R. P., and Singh, H. (1987), Intranasal nitroglycerin attenuates pressor response to tracheal intubation in beta-blocker treated hypertensive patients, *Anesthesia*, 42, 884–887.
123. Hussain, A., Hirai, S., and Bawarshi, R. (1979), Nasal absorption of propranolol in rats, *J. Pharm. Sci.*, 68, 1196.
124. Hussain, A. A., Foster, T., Hirai, S., Kashihara, T., Batenhorst, R., and Jones, M. (1980), Nasal absorption of propranolol in humans, *J. Pharm. Sci.*, 69, 1240.
125. Landau, A. J., Frishman, W. H., Alturk, N., Adjei-Poku, M., Fornasier-Bongo, M., and Furia, S. (1993), Immediate beta-adrenergic blockade and improved exercise tolerance with intranasal propranolol in patients with angina pectoris, *Am. J. Cardiol.*, 72, 995–998.
126. Duchateau, G. S. M. J. E., Zuidema, J., Albers, W. M., and Merkus, F. W. H. M. (1986), Nasal absorption of alprenolol and metoprolol, *Int. J. Pharm.*, 34, 131–136.
127. Duchateau, G. S. M. J. E., Zuidema, J., and Merkus, F. W. H. M. (1986), Bioavailability of propranolol after oral, sublingual and intranasal administration, *Pharm. Res.*, 3, 108–111.
128. Wattanakumtornkul, S., Pinto, A. B., and Williams, D. B. (2003), Intranasal hormone replacement therapy, *Menopause.*, 10, 88–98.
129. Hermens, W. A. J. J., Belder, C. W., Merkus, J. M., Hooymans, P. M., Verhoef, J., and Merkus, F. W. (1991), Intranasal estradiol administration to oophorectomized women, *Eur. J. Obstet. Gynecol. Reprod. Biol.*, 40, 35–41.
130. Hermens, W. A., Belder, C. W., Merkus, J. M., Hooymans, P. M., Verhoef, J., and Merkus, F. W. (1992), Intranasal administration of estradiol in combination with progesterone to oophorectomized women, *Eur. J. Obstet. Gynecol. Reprod. Biol.*, 43, 65–70.
131. Garnero, P., Tsouderos, Y., Marton, I., Pelissier, C., Varin, C., and Delmas, P. D. (1999), Effects of intranasal 17 β-estradiol on bone marrow turnover and serum insulin-like growth factor I in postmenopausal women, *J. Clin. Endocrinol. Metab.*, 84, 2390–2397.
132. Lievertz, R. W. (1987), Pharmacology and pharmacokinetics of estrogens, *Am. J. Obstet. Gynecol.*, 156, 1289–1293.
133. Studd, J., Pornel, B., Marton, I., Bringer, J., Varin, C., Tsouderos, Y., and Christansen, C. (1999), Efficacy and acceptability of intranasal 17 β-estradiol on menopausal symptoms: A randomized dose-response study, *Lancet.*, 353, 1574–1578.
134. Devissaguet, J. P., Brion, N., Lhote, O., and Deloffre, P. (1999), Pulsed estrogen therapy: Pharmacokinetics of intranasal 17 β-estradiol (S21400) in postmenopausal women and

comparison with oral and transdermal formulations, *Eur. J. Drug Metab. Pharmacokinet.*, 24, 265–271.

135. Wilkund, I., Karlberg, J., and Mattson, L. A. (1993), Quality of life of postmenopausal women on a regimen of transdermal estradiol therapy: A double blind placebo-controlled study, *Am. J. Obstet. Gynecol.*, 168, 824–830.
136. Wilkund, I., Holst, J., Karlberg, J., et al. (1992), A new methodological approach to the evaluation of quality of life in postmenopausal women, *Maturitas*, 14, 211–224.
137. Kupperman, H. S., Blatt, M. H. G., Weisbader, H., and Filler, W. (1953), Comparative clinical evaluation of estrogenic preparations by the menopausal and amenorrheal indices, *Endocrinology*, 13, 688–703.
138. Hussain, A. A., Al-bayati, A. A., Dakkuri, A., Okochi, K., and Hussain, M. A. (2002), Testosterone 17 β-*N,N*-dimethylglycinate hydrochloride: A prodrug with a potential for nasal delivery of testosterone, *J. Pharm. Sci.*, 91, 785–789.
139. Hussain, A. A., Kimura, R., and Huang, C. H. (1984), Nasal absorption of testosterone in rats, *J. Pharm. Sci.*, 73, 1300–1301.
140. Illum, L. (2000), Transport of drugs from the nasal cavity to the central nervous system, *Eur. J. Pharm. Sci.*, 11, 1–18.
141. Chow, H. S., Chen, Z., and Matsuura, G. T. (1999), Direct transport of cocaine from the nasal cavity to the brain following intranasal cocaine administration in rats, *J. Pharm. Sci.*, 88, 754–758.
142. Landsteiner, K., and Levaditi, C. (1910), Etude experimentale de la poliomyelite aigue (Maladiede Heine Medin), *Ann. Inst. Pasteur.*, 24, 833–878.
143. Leiner, C., and von Weisner, R. (1910), Experimentale untersuchungen uber poliomyelitis acuta, *Wien. Med. Wchnschr.*, 60, 2482.
144. Flexner, S. (1912), The mode of action of infection epidemic poliomyelitis, *JAMA*, 59, 1371.
145. Flexner, S., and Clark, P. F. (1912), A note on the mode of infection in epidemic poliomyelitis, *Proc. Soc. Exp. Biol. Med.*, 10, 1.
146. Clark, P. F., Fraser, F. R., and Amoss, H. L. (1917), The relation to the blood of the virus of epidemic poliomyelitis, *J. Exp. Med.*, 19, 223–233.
147. Fairbrother, R. W., and Hurst, E. W. (1930), The pathogenesis of the propagation of the virus in experimental poliomyelitis, *J. Path. Bact.*, 33, 17–45.
148. Faber, H. K., and Gebhardt, L. D. (1938), Localization of poliomyelitic virus during incubation period after intranasal instillation in monkeys, *Proc. Soc. Exper. Biol. Med.*, 30, 879–880.
149. Faber, H. K., and Gebhardt, L. D. (1938), Localizations of the virus of poliomyelitis. In the central nervous system during the pre-paralytic period, after intranasal inoculation, *J. Exp. Med.*, 57, 933–954.
150. Sabin, A. B., and Olitsky, P. K. (1936), Influence of pathway of infection on pathology of olfactory bulbs in experimental poliomyelitis, *Proc. Soc. Exp. Biol. Med.*, 35, 300–301.
151. Bodian, D., Morgan, I., and Schwerdt, C. E. (1950), Virus and host factors influencing the titer of Lansing poliomyelitis virus in monkeys, cotton rats and mice, *Am. J. Hyg.*, 51, 126–133.
152. Gopinath, P. G., Gopinath, G., and Kumar, T. C. A. (1978), Target site of intranasally sprayed substances and their transport across the nasal mucosa. A new insight into the intranasal route of drug delivery, *Curr. Ther. Res.*, 23, 596–607.
153. Hastings, L. (1990), Sensory neurotoxicology: Use of the olfactory system in the assessment of toxicity, *Neurotoxicol. Teratol.*, 12, 455–459.

154. Hastings, L., and Evans, J. E. (1991), Olfactory primary neuron as a route of entry for toxic agents into the CNS, *Neurotoxicology*, 12, 707–714.
155. Illum, L. (2004), Is nose-to-brain transport of drugs in man a reality? *J. Pharm. Pharmacol.*, 56, 3–17.
156. Mathison, S., Nagilla, R., and Kompella, U. B. (1998), Nasal route for direct delivery of solutes to the central nervous system: Fact or fiction? *J. Drug Target.*, 5, 415–441.
157. Thorne, R. G., and Frey, W. H. (2001), Delivery of neurotropic factors to the central nervous system, *Brain. Res.*, 692, 278–282.
158. Pietrowsky, R., Struben, C., Molle, M., Fehm, H. L., and Born, J. (1966), Brain potential changes after intranasal vs. intravenous administration of vasopressin: Evidence for a direct nose-to-brain pathway for peptide effects in humans, *Biol. Psychiatry*, 39, 332–340.
159. Derad, I., Willeke, K., Pietrowsky, R., Born, J., and Fehm, H. L. (1998), Intranasal angiotensin II directly influences central nervous regulation of blood pressure, *Am. J. Hypertens.*, 11, 971–977.
160. Pietrowsky, R., Thiemann, A., Kern, W., Fehm, H. L., and Born, J. (1966), A nose-to-brain pathway for psychotropic peptides: Evidence from a brain evoked potential study with cholecystokinin, *Psychoneuroendocrinology*, 21, 559–572.
161. Lindhardt, K., Gizurarson, S., Stefansson, S. B., Olafsson, D. R., and Bechgaard, E. (2001), Electroencephalographic effects and serum concentration after intranasal and intravenous administration of diazepam to healthy volunteers, *Br. J. Clin. Pharmacol.*, 52, 521–527.
162. Kern, W., Peters, A., Fruehwald-Schultes, B., Deininger, E., Born, J., and Fehm, H. L. (2001), Improving influence of insulin on cognitive functions in humans, *Neuroendocrinology*, 74, 270–280.
163. Okuyama, S. (1997), The first attempt at radioisotopic evaluation of the integrity of the nose-brain barrier, *Life Sci.*, 60, 1881–1884.
164. Quay, S. C. (2001), Successful delivery of apomorphine to the brain following intranasal administration demonstrated in clinical study, PRNewswire, July 18.
165. Merkus, P. (2003), Transport of non-peptide drugs from the nose to the CSF, paper presented at the Nasal Drug Delivery Meeting, Management Forum, London, Mar. 24–25.
166. Born, J., Lange, T., Kern, W., McGregor, G. P., Bickel, U., and Fehm, H. L. (2002), Sniffing neuropeptides: A transnasal approach to the human brain, *Nat. Neurosci.*, 5, 514–516.
167. Seki, T., Sato, N., Hasegawa, T., Kawaguchi, T., and Juni, K. (1994), Nasal absorption of zidovudine and its transport to cerebrospinal fluid in rats, *Biol. Pharm. Bull.*, 17, 1135–1137.
168. Char, H., Kumar, S., Patel, S., Piemontese, D., Iqbal, K., Waseem Malick, A., Salvador, R. A., and Behl, C. R. (1992), Nasal delivery of [14C] dextromethorphan hydrochloride in rats: Levels in plasma and brain, *J. Pharm. Sci.*, 81, 750–752.
169. Sakane, T., Akizuki, M., Yoshida, M., Yamashita, S., Nadai, T., Hashida, M., and Sezaki, H. (1991), Transport of cephalexin to the cerebrospinal fluid directly from the nasal cavity, *J. Pharm. Pharmacol.*, 43, 449–451.
170. Sakane, T., Akizuki, M., Yamashita, S., Sezaki, H., and Nadai, T. (1994), Direct drug transport from the nasal cavity to the cerebrospinal fluid: The relation to the dissociation of the drug, *J. Pharm. Pharmacol.*, 46, 378–379.
171. Sakane, T., Akizuki, M., Yamashita, S., Nadai, T., Hashida, M., and Sezaki, H. (1991), The transport of a drug to the cerebrospinal fluid directly from the nasal cavity: The relation to the lipophilicity of the drug, *Chem. Pharm. Bull.*, 39, 2456–2458.

172. Sakane, T., Akizuki, M., Taki, Y., Yamashita, S., Sezaki, H., and Nadai, T. (1995), Direct drug transport from the nasal cavity to the cerebrospinal fluid: The relation to the molecular weight of the drugs, *J. Pharm. Pharmacol.*, 47, 379–381.
173. Dahlin, M., Jansson, B., and Bjork, E. (2001), Levels of dopamine in the blood and brain following nasal administration to rats, *Eur. J. Pharm. Sci.*, 14, 75–80.
174. Wang, Y., Aun, R., and Tse, F. L. S. (1998), Brain uptake of dihydroergotamine after intravenous and nasal administration in the rat, *Biopharm. Drug Dispos.*, 19, 571–575.
175. Gizurarson, S., Trovaldsson, T., Sigurdsson, P., and Gunnarsson, E. (1996), Selective delivery of insulin into the brain: Introlfatory absorption, *Int. J. Pharm.*, 140, 77–83.
176. Wang, F., Jiang, X., and Lu, W. (2003), Profiles of methotrexate in blood and CSF following intranasal and intravenous administration to rats, *Int. J. Pharm.*, 263, 1–7.
177. Frey, W. H., Liu, J., Thorne, R. G., and Rahman, Y. E. (1995), Intranasal delivery of $^{125}$I-labelled nerve growth factor to the brain via the olfactory route, in Iqbal, K., Mortimer, J. A., Winbald, B., and Wisniewski, H. M., Eds., *Research Advances in Alzheimer's Disease and Related Disorders*, Wiley, New York, pp. 329–335.
178. Thorne, R. G., Emroy, C. R., Ala, T. A., and Frey, W. H. (1995), Quantitative analysis of the olfactory pathway for drug delivery to the brain, *Brain. Res.*, 692, 278–282.
179. Vyas, T. K., Babbar, A. K., Sharma, R. K., and Misra, A. (2005), Intranasal mucoadhesive microemulsions of zolmitriptan: Preliminary studies on brain-targeting, *J. Drug Target.*, 13, 317–324.
180. Fliedner, S., Schulz, C., and Lehnert, H. (2006), Brain uptake of intranasally applied radioiodinated leptin in wistar rats, *Endocrinology*, 147, 2088–2094.
181. Westin, U. E., Boström, E., Gråsjö, J., Hammarlund-Udenaes, M., and Björk, E. (2006), Direct nose-to-brain transfer of morphine after nasal administration to rats, *Pharm. Res.*, 23, 565–572.
182. Zhang, Q. Z., Zha, L. S., Zhang, Y., Jiang, W. M., Lu, W., Shi, Z. Q., Jhang, X. G., and Fu, S. K. (2006), The brain targeting efficiency following nasally applied MPEG-PLA nanoparticles in rats, *J. Drug Target.*, 14, 281–290.
183. Shi, Z., Zhang, Q., and Jiang, X. (2005), Pharmacokinetic behavior in plasma, cerebrospinal fluid and cerebral cortex after intranasal administration of hydrochloride meptazinol, *Life Sci.*, 77, 2574–2583.
184. Wang, X., Haibing, H., Leng, W., and Tang, X. (2006), Evaluation of brain-targeting for the nasal delivery of estradiol by the microdialysis method, *Int. J. Pharm.*, 317, 40–46.
185. Jones, N. (2001), The nose and paranasal sinuses physiology and anatomy, *Adv. Drug Deliv. Rev.*, 51, 5–19.
186. Graff, C. L., and Pollack, G. M. (2005), Nasal drug administration: Potential for targeted central nervous system delivery, *J. Pharm. Sci.*, 94, 1187–1195.
187. Sherwood, L. (1989), *Human Physiology. From Cells to Systems*, West Publishing, St. Paul, MN, Chapter 5.
188. Stanley, P. A., and Walter, O. A. (2004), in Plotkin, S. A., Orenstein, W. A., and Offit, P. A., Eds., *A Short History of Vaccination. Vaccines*, 4th ed., Saunders, Philadelphia, pp. 1–15.
189. Ogra, P. L., Faden, H., and Welliver, R. C. (2001, Apr.), Vaccination strategies for mucosal immune responses, *Clin. Microbiol. Rev.*, 14(2), 430–445.
190. Cui, Z. R., and Mumper, R. J. (2002), Intranasal administration of plasmid DNA-coated nanoparticles results in enhanced immune responses, *J. Pharm. Pharmacol.*, 54, 1195–1203.
191. Beverley, L. C. P. (2002), Immunology of vaccination, *Br. Med. Bull.*, 62, 15–28.

192. Illum, L., and Davis, S. S. (2001), Nasal vaccination: A non-invasive vaccine delivery method that holds great promise for the future, *Adv. Drug Deliv. Rev.*, 51, 1–3.
193. Davis, S. S. (2001), Nasal vaccines, *Adv. Drug Deliv. Rev.*, 51, 21–42.
194. Partidos, C. D. (2000), Intranasal vaccines: Forthcoming challenges, *PSTT*, 3, 273–281.
195. Mestecky, J., and McGhee, J. R. (1987), Immunoglobulin A: Molecular and cellular interactions in IgA biosynthesis and immune response, *Adv. Immunol.*, 40, 153–245.
196. Gill, J. I., Lin, W., Kistner, O., Davis, S. S., and Illum, L. ( 2001), Polymeric lamellar substrate particles for intranasal vaccination, *Adv. Drug Deliv. Rev.*, 51, 97–111.
197. Biesterveld, N. (2002), Nasal vaccine delivery, vaccine delivery adjuvants, Merck & Company, West Point, PA, a Literature review submitted to the Department of Chemistry, Lehigh University, Aug. 2002.
198. Gander, B. (2005), Trends in particulate antigen and DNA delivery systems for vaccines, *Adv. Drug Deliv. Rev.*, 57, 321–323.
199. Friede, M., and Aguado, T. M. (2005), Need for new vaccine formulation and potential of particulate antigen and DNA delivery system, *Adv. Drug Deliv. Rev.*, 57, 325–331.
200. Illum, L. (2003), Nasal drug delivery–possibilities, problems and solutions, *J. Controlled Release*, 87, 187–198.
201. Alpar, O. H., Somavarapu, S., Atuah, N. K., and Bramwell, W. V. (2005), Biodegradable mucoadhesive particulates for nasal and pulmonary antigen and DNA delivery. *Adv. Drug Deliv. Rev.*, 57, 411–430.
202. Mackay, M., Williamson, I., and Hastewell, J. (1991), Cell biology of Epithelia, *Adv. Drug Deliv. Rev.*, 7, 313–338.
203. Storni, T., Kundig, M. T., Senti, G., and Johansen, P. (2005), Immunity in response to particulate antigen-delivery systems, *Adv. Drug Deliv. Rev.*, 57, 333–355.
204. O'Hagan, T. D., and Rappuoli, R. (2004), Novel approaches to vaccine delivery, *Pharm. Res.*, 21(9), 1519–1530.
205. McNeela, A. E., and Mills, G. H. K. (2001), Manipulating the immune response: Humoral, versus, cell, mediated, immunity, *Adv. Drug Deliv. Rev.*, 51, 43–54.
206. Edelman, R. (2002), The development and use of vaccine adjuvants, *Mol. Biotechnol.*, 21, 129–148.
207. Singh, M., and O'Hagan, D. (1999), Advances in vaccine adjuvants, *Nat. Biotechnol.*, 17, 1075–1081.
208. McNeela, E. A., O'Connor, D., Gill, J. I., Illum, L., Davis, S. S., Pizza, M., Peppoloni, S., Rappuoli, R., and Mills, G. H. K. ( 2000), A mucosal vaccine against diphtheria: Formulation of cross reacting material ($CRM_{197}$) of diphtheria toxin with chitosan enhances local and systemic antibodies and Th2 responses following nasal delivery, *Vaccine.*, 19, 1188–1198.
209. Illum, L., Gill, J. I., Hinchcliffe, M., Fisher, N. A., and Davis, S. S. (2001), Chitosan as a novel nasal delivery system for vaccines, *Adv. Drug Deliv. Rev.*, 51, 81–96.
210. Jabbal-Gill, I., Fisher, A. N., Rappuoli, R., Davis, S. S., and Illum, L. (1998), Stimulation of mucosal and systemic antibody responses against Bordetella *pertussis* filamentous haemagglutinin and recombinant *pertussis* toxin after nasal administration with chitosan in mice, *Vaccine*, 16, 2039–2046.
211. Boyaka, P. N., Marinaro, M., Jackson, R. J., Menon, S., Kiyona, H., Jirillo, E., and McGhee, J. R. (1999), IL-12 is an effective adjuvant for induction of mucosal immunity, *J. Immunol.*, 162, 122–128.
212. Arulanandan, B. P., O'Toole, M., and Metzger, D. W. (1999), Intranasal interleukin-12 is a powerful adjuvant for protective mucosal immunity, *J. Infect. Dis.*, 180, 940–949.

213. McCluskie, M. J., and Davis, H. L. (2001), Oral, intrarectal and intranasal immunizations using CpG and non-CpG oligodeoxynucleotides as adjuvants, *Vaccine.*, 19, 413–422.

214. Moldoveanu, Z., Love-Homan, L., Huang, W. Q., and Krieg, A. M. (1998), CpG DNA, a novel immune enhancer for systemic and mucosal immunization with influenza virus, *Vaccine*, 16, 1216–1224.

215. de Haan, A., Geerligs, H. J., Huchshorn, J. P., van Scharrenburg, G. J., Palache, A. M., and Wilschut, J. (1995), Mucosal immunoadjuvant activity of liposomes: Induction of systemic IgG and secretory IgA responses in mice by intranasal immunization with an influenza subunit vaccine and coadministered liposomes, *Vaccine*, 13, 155–162.

216. de Haan, A., Tomee, J. F. C., Huchshorn, J. P., and Wilschut, J. (1995), Liposomes as an immunoadjuvant system for stimulation of mucosal and systemic antibody response against inactivated measles virus administered intranasally to mice, *Vaccine*, 13, 1320–1324.

217. Aramaki, Y., Fujii, Y., Yachi, K., Kikuchi, H., and Tsuchiya, S. (1994), Activation of systemic and mucosal immune response following nasal administration of liposomes, *Vaccine*, 12, 1241–1245.

218. Childers, N. K., Tong, G., Mitchell, S., Kirk, K., Russel, M. W., and Michalek, S. M. (1999), A controlled clinical study of the effect of nasal immunization with a *Streptococcus mutans* antigen alone or incorporated into liposomes on induction of immune responses, *Infect. Immun.*, 67, 618–623.

219. Gluck, U., Gebbers, J. O., and Gluck, R. (1999), Phase I evaluation of intranasal virosomal influenza vaccine with and without *Escherichia coli* heat-labile toxin in adult volunteers, *J. Virol.*, 73, 7780–7786.

220. Glück, R. (1999), Adjuvant activity of immunopotentiating reconstituted influenza virosomes (IRIVs), *Vaccine*, 17, 1782–1787.

221. Baca-Estrada, M. E., Foldvari, M., Snider, M., Harding, K., Kournikakis, B., Babiuk, L. A., and Griebel, P. (2000), Intranasal immunization with liposome-formulated *Yersinia pestis* vaccine enhances mucosal immune responses, *Vaccine*, 18, 2203–2211.

222. Lövgren, K., Kaberg, H., and Morein, B. (1990), An experimental influenza subunit vaccine (iscom): Induction of protective immunity to challenge infection in mice after intranasal or subcutaneous administration, *Clin. Exp. Immunol.*, 82, 435–439.

223. Hsu, S. C., Schadeck, E. B., Delmas, A., Shaw, M., and Steward, M. W. (1996), Linkage of a fusion peptide to a CTL epitope from the nucleoprotein of measles virus enables incorporation into ISCOMS and induction of CTL responses following intranasal immunization, *Vaccine*, 14, 1159–1166.

224. Carol, H., Nieto, A., Villacres-Eriksson, M., and Morein, B. (1997), Intranasal immunization of mice with *Echinococcus granulosus* surface antigens iscoms evokes a strong immune response, biased towards glucidic epitopes, *Parasite Immunol.*, 19, 197–205.

225. Hu, K. F., Elvander, M., Merza, M., Akerblom, L., BrAndenburg, A., and Morein, B. (1998), The immunostimulatory complex (ISCOM) is an efficient mucosal delivery system for respiratory syncytial virus (RSV) envelope antigens inducing high local and systemic antibody responses, *Clin. Exp. Immunol.*, 113, 235–243.

226. Jaganathan, K. S., and Vyas, S. P. (2006), Strong systemic and mucosal immune responses to surface-modified PLGA microspheres containing recombinant hepatitis B antigen administered intranasally, *Vaccine*, 24, 4201–4211.

227. Debin, A., Kravtzoff, R., Vaz Santiago, J., Cazales, L., Speranido, S., Melber, K., Janowicz, Z., Betbeder, D., and Moynier, M. (2002), Intranasal immunization with recombinant antigens associated with new cationic particles induces strong mucosal as well as systemic antibody and CTL responses, *Vaccine*, 20, 2752–2763.

228. Aguilar, J. C., Lobaina, Y., Muzio, V., Garcia, D., Penton, E., Iglesias, E., Pichardo, D., Urquiza, D., Rodeiguez, D., Silva, D., Petrovsky, N., and Guillen, G. (2004), Development of a nasal vaccine for chronic hepatitis B infection that uses the ability of hepatitis B core antigen to stimulate a strong Th1 response against hepatitis B surface antigen, *Immnol. Cell. Biol.*, 82, 539–546.

229. Hordness, K., Tynning, T., Brown, T. A., Haneberg, B., and Jonsson, R. (1997), Nasal immunization with group B streptococci can induce high levels of specific IgA antibodies in cervicovaginal secretions of mice, *Vaccine*, 15, 1244–1251.

230. Kurono, Y., Yamamoto, M., Fujihashi, K., Kodama, S., Suzuki, M., Mogi, G., McGhee, J. R., and Kiyono, H. (1991), Nasal immunization induces *Haemophilis influenzae*-specific Th1 and Th2 responses with mucosal IgA and systemic IgG antibodies for protective immunity, *J. Infect. Dis.*, 180, 122–132.

231. Hagiwara, Y., Komase, K., Chen, Z., Matsuo, K., Suzuki, Y., Aizawa, C., Kurata, T., and Tamura, S. (1999), Mutants of cholera toxin as an effective and safe adjuvant for nasal influenza vaccine, *Vaccine*, 17, 2918–2926.

232. Bastein, N., Trudel, M., and Simard, C. (1999), Complete protection of mice from respiratory syncytial virus infection following mucosal delivery of synthetic peptide vaccines, *Vaccine*, 17, 832–836.

# NASAL POWDER DRUG DELIVERY

Jelena Filipović-Grčić and Anita Hafner
*University of Zagreb, Zagreb, Croatia*

**Contents**

1 Introduction
2 Nasal Dry Powder Formulations
  2.1 Benefits Associated with Nasal Powder Drug Delivery
  2.2 Drug Powder or Drug/Polymer Powder Formulation for Nasal Drug Delivery?
  2.3 Powder Properties Affecting Nasal Deposition and Drug Delivery
3 Polymers in Nasal Powder Delivery System
4 Microspheres as Nasal Drug Delivery Devices
  4.1 Preparation Methods
  4.2 Microsphere Characterization
  4.3 Chitosan-Formulated Spray-Dried Microspheres
5 Toxicological Considerations
  References

## 1 INTRODUCTION

Intranasal drug administration has been practiced since ancient times. In Tibet extracts of sandalwood and aloewood were inhaled to treat emesis. Egyptians treated epistaxis and rhinitis using intranasal medication. North American Indians relieved headaches inhaling crushed leaves of *Ranunculus acris* [1]. Due to the rich vasculature and high permeability of nasal mucosa, the absorption rate and pharmacokinetics of nasally administrated drug are comparable to that obtained by intravenous drug delivery, while noninvasive nasal drug administration is more convenient to patients. As nasally administrated drugs avoid first-pass hepatic metabolism, improved bioavailability can be expected. However, rapid mucociliary

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

clearance reduces the residence time of nasal drug delivery system at the site of absorption. Dry powders have been shown to delay mucociliary clearance, thus prolonging the contact time between the drug delivery system and mucosa compared to liquid formulations. Most of the dry powder investigations are based on mucoadhesive swellable polymers as they can additionally improve drug absorption and bioavailability. Dry powder delivery systems such as microspheres are of special interest, offering the possibility of predictable and controlled drug release from the polymeric device [2, 3].

## 2 NASAL DRY POWDER FORMULATIONS

Liquid preparations are most frequently used nasal dosage forms at present. However, such preparations are characterized with short residence time in the nasal cavity, low drug concentration at the site of absorption, and problems linked to the chemical stability of the drug and the stability of the preparation. In the case of liquid formulations, drugs must be administered in small volumes. The maximum volume of a single dose in one nostril is about 200 μL. The volume of therapeutic dose should not exceed the capacity of the nasal cavity, as it would drain out of the nose. Thus, only low-dose or highly soluble drugs can be administered nasally in the form of a simple liquid formulation [4]. Dry powder formulations have been recognized as efficient nasal delivery systems offering numerous advantages over liquid formulations, such as avoidance of preservatives, improved formulation stability, and prolonged contact with the mucosa. For a powder formulation, the maximum quantity is approximately 50 mg, depending upon the bulk density of the material [5].

### 2.1 Benefits Associated with Nasal Powder Drug Delivery

A powder form was found to be more effective than liquid formulations in a number of investigations described in the literature [6–11]. Dry powders are characterized by prolonged residence time and higher drug concentration at the site of deposition as well as improved formulation stability with no requirement for preservatives [12]. Prolonged residence time of the powder delivery systems at the absorption site results in enhanced systemic bioavailability, compared to the liquid formulations. In the case of powders, higher drug concentration at the site of absorption causes rapid transmucosal diffusion and faster onset of action [13]. Most of the dry powder formulations are based on mucoadhesive swellable polymers (e.g., starch, dextran, chitosan) as they can additionally improve drug bioavailability, prolonging the residence time in the nasal cavity or even promoting drug absorption. Powder formulation with a water-insoluble and nonswellable drug carrier may also improve nasal bioavailability of the polar drugs. Ishikawa et al. [14] found that a nasal powder delivery system of elcatonin based on $CaCO_3$ significantly increased the systemic elcatonin bioavailability in rats and rabbits compared to the liquid formulation. Enhanced bioavailability has been primarily ascribed to the retardation of the clearance of the drug powder delivery system from the nasal cavity.

The use of dry powder formulations in nasal vaccine delivery has been extensively reviewed elsewhere [15–18]. The association of vaccines to some of the par-

ticulate systems has been proved to enhance the systemic and mucosal immune responses against the antigens [19–21]. Dry powder formulations for nasal vaccine delivery may also provide significant advantages with respect to stability shortcomings compared to conventional liquid intranasal and intramuscular formulations, which require frozen storage or refrigeration [22].

### 2.2 Drug Powder or Drug/Polymer Powder Formulation for Nasal Drug Delivery?

The drug candidate for nasal administration should possess a number of attributes, such as appropriate aqueous solubility and nasal absorption characteristics, minimal nasal irritation, low dose, no offensive odor or aroma, and suitable stability characteristics [23]. In the case of drug powder formulations it is possible to hide or alter the unfavorable characteristics of a drug using suitable polymers as drug carriers. Thus, improvement of the dissolution behavior of drugs of low aqueous solubility after incorporation in polymeric powder devices such as microspheres has been reported in the literature [24, 25]. The improvement of the drug dissolution rate from the microspheres has been ascribed to several factors, such as high microsphere surface–volume ratio, the hydrophilic nature of the polymer, and drug amorphization due to drug–polymer interaction and/or the microsphere preparation method [25, 26].

Nasally administrated polymer–drug powders were also characterized by improved drug absorption compared to pure drug powders [25, 27]. Teshima et al. [27] monitored changes in the plasma glucagons and glucose concentrations after nasal administration of the powder form of glucagon alone and glucagon mixed with the carrier, microcrystalline cellulose (MCC). Glucagon and glucose plasma concentrations remained unchanged after nasal administration of the powder form of glucagon alone while it increased after glucagon–MCC administration in an MCC content–dependent manner. Results of in vivo nasal administration of carbamazepine-loaded chitosan microspheres revealed an increase in carbamazepine concentration in serum compared to the pure carbamazepine powder [25]. Such an increase in drug absorption has been ascribed to both improved dissolution of carbamazepine and adhesion of the chitosan microspheres to the mucosal surface.

The influence of polymers on the drug stability in the powder formulation has also been reported [8]. Green coloration of polyacrylic acid powder dosage forms loaded with 60% apomorphine due to atmospheric drug oxidation upon storage has been observed. Dosage forms with lower drug loadings (and higher polymer content) showed no coloration, indicating the protective role of polymers against drug oxidation in powder formulations.

Powders intended for nasal administration have to be optimized in terms of particle size and morphology as these properties are related to potential irritation in the nasal cavity [23]. Certain procedures (e.g., spray drying process) can modify the particle size of the drug powder raw material, but in order to optimize the morphology and flowability properties of some pure drug powders, excipients need to be used. Sacchetti et al. [28] reported that the use of mannitol as a filler and hydroxypropylmethyl cellulose (HPMC) as a shaper of spray-dried caffeine microparticles modified the typical needle shape of spray-dried caffeine to a more convenient roundish shape. Further addition of polyethylene glycol (PEG) resulted in increased

cohesiveness between particles, producing agglomerates considered as beneficial for nasal deposition.

### 2.3 Powder Properties Affecting Nasal Deposition and Drug Delivery

Powder formulation efficiency is strongly dependent on two technological aspects of delivery: first, the complete release of the dose through the powder delivery device, that is, the quantitative aspect, and, second, the distribution of the powder on the nasal mucosa, that is, the qualitative aspect. For example, rhinitis treatment requires drug deposition over the entire mucosal area, while brain targeting requires deposition strictly at the roof of the nose. The qualitative aspect is defined by the aerodynamic properties of the powder cloud produced (puff) and is assessed by a photographic technique providing information on the spray pattern and cloud geometry. The parameters involved are shape, height, area, density, particle size, and velocity. For nasal impacting the powder cloud should remain as compact as possible to achieve an efficient shot of powder to the nasal mucosa, whereas for sedimentation a larger cloud would be preferred [12].

The quantitative aspect of nasal powder delivery is largely dependent on the type of delivery device. Devices for powder dosage forms, including insufflators, monodose and multidose powder inhalers, and pressurized metered-dose inhalers, are extensively reviewed elsewhere [29, 30]. Particles intended for nasal delivery should have good flow properties to be reproducibly filled in the dose reservoir and easily insufflated to obtain appropriate nasal deposition [28]. Since the particle size of the applied powder formulation has a major impact on its nasal deposition, the characterization of this parameter is very important. Only particles over 5 μm are deposited in the nostrils while smaller particles can be inhaled into the lower parts of the respiratory system [23]. For that reason, it is necessary to determine particle sizes not only prior to filling the nasal delivery device but also after its actuation, since the particle size of the formulation leaving the device depends on the device-disaggregating properties [29]. Furthermore, it is known that powder deposition in the anterior part of the nasal epithelium contributes to nasal absorption more than powder deposition in the posterior part, since ciliated cells cover mainly the posterior part of the nasal epithelium, providing a faster clearance of particles [31, 32]. Thus, deposition studies must include both the amount of deposited particles and the location of the particle deposition [29]. Distribution of the formulation in the nasal cavity should be evaluated considering two different deposition types: the initial deposition immediately after application and the secondary deposition, owing to translocation by mucociliary clearance [29]. Cornaz et al. [33] investigated the ability of microsphere delivery, studying the shape and area of the clouds of microspheres emitted from a nasal device. The puff areas of loaded particles were smaller compared to unloaded particles due to different particle porosities. Pringels et al. [31] investigated the influence of the deposition pattern and spray characteristics of nasal powder formulations on insulin bioavailability. It has been shown that posterior deposition of the powder formulation in the nasal cavity lowered insulin bioavailability. To study the spray characteristics, the shape and cross section of the emitted powder cloud (puff) were analyzed. It has been concluded that the powder bulk density of the formulation influenced the spray pattern. Comparing two powders with different bulk densities and particle sizes, it has been revealed that

the powder with higher bulk density and smaller particle size was more compact and was characterized by higher resistance to airflow, resulting in a slower spray time and larger spray pattern [31]. No influence of the powder bulk density and spray pattern on insulin bioavailability has been observed.

## 3 POLYMERS IN NASAL POWDER DELIVERY SYSTEM

It has been demonstrated that low absorption of drugs can be improved by using absorption enhancers or prolonging contact between drug and absorptive sites in the nasal cavity by delaying mucociliary clearance of the formulation. Some mucoadhesive polymers can serve both functions. They are typically high-molecular-weight polymers with flexible chains which can interact with mucin through hydrogen bonding, electrostatic, hydrophobic or van der Waals interactions [18, 34, 35].

The mucoadhesive polymers are often hydrophilic and swellable, containing numerous hydrogen bond–forming groups such as hydroxyl, carboxyl, or amine, which favors adhesion. When used in a dry form they attract water from the mucosal surface and swell, leading to polymer–mucus interaction, increased viscosity of polymer–mucus mixture, and reduced mucociliary clearance [34]. Beside the type of polymer functional groups, the mucoadhesive force of a polymer material is dependent on the polymer molecular weight, concentration, flexibility of the polymer chain, spatial conformation, contact time, environmental pH, and physiological factors such as mucin turnover and disease state [3]. There is a critical polymer molecular weight for each polymer type below or above which there is reduced adhesive power [34]. The mucoadhesive properties can also be affected by the degree of cross-linking of the polymer since mucoadhesion requires an adequate free chain length for interpenetration to occur. Hence, the more cross-linked the polymer, the less strong the mucoadhesive interaction [36]. Hydration and swelling present both polymer- and environment-related factors. Overhydration causes extended swelling, resulting in slippery mucilage formation [30]. The polymer concentration that is required for optimum mucoadhesion is different between gels and solid mucoadhesives. In the liquid state, an optimum concentration exists for each polymer for which best adhesion can occur, while with solid dosage forms, increased polymer concentration leads to increased mucoadhesive power [35]. Studies have shown that polymers with charge density can serve as good mucoadhesive agents, although their mucoadhesive properties are affected by the pH of the surrounding media. The presence of metal ions, which can interact with charged polymers, may also affect the adhesion process [35, 36]. It has also been reported that polyanion polymers are more effective bioadhesives than polycation polymers or nonionic polymers [37].

Research on nasal powder drug delivery has employed polymers such as starch, dextrans, polyacrylic acid derivatives (e.g., carbopol, polycarbophil), cellulose derivatives (microcrystalline cellulose, semicrystalline cellulose, hydroxypropylmethyl cellulose, hydroxypropyl cellulose, carboxymethyl cellulose), chitosan, sodium alginate, hyaluronans, and polyanhydrides such as poly(methyl vinyl ether-*co*-maleic anhydride) (PVM/MA). Many of these polymers have already been used as excipients in pharmaceutical formulations and are often referred to as first-generation bioadhesives [38–45]. In nasal dry powder a single bioadhesive polymer or a

combination of two or more polymers has been formulated as freeze-dried or spray-dried particles or micropheres.

Crystalline cellulose, hydroxypropyl cellulose, and Carbopol 934 have been studied in combination with lyophilized insulin as bioadhesive powder dosage forms for nasal delivery. Each formulation tested resulted in an decrease in plasma glucose level after nasal administration in dog and rabbit models. The most effective formulation, crystalline cellulose blended with insulin, decreased the plasma glucose level to 49% of the control value. In ternary systems the lyophilized Carbopol 934 and insulin blend with crystalline cellulose powder has been the most effective, leading to a hypoglycemia on the order of one-third of the effect obtained after intravenous injection of the same dose of insulin. The plasma glucose levels obtained in the volunteers after administration of the insulin–Carbopol–crystalline cellulose powder formulation were quite variable [38].

The various powder formulations were prepared by dry blending of octreotide with microcrystalline cellulose, semicrystalline cellulose, hydroxyethyl starch, cross-linked dextran, microcrystalline chitosan, pectin, and alginic acid [40]. Their potential to enhance the nasal absorption of the somatostatin analogue peptide octreotide was studied in vivo in the rat model. The powder mixtures were also characterized in vitro regarding calcium binding, water uptake, and drug release. The bioavailabilities obtained for all of the powder formulations were low, with the highest values for alginic acid and cross-linked dextran powder formulations (4.1 and 5.56%, respectively).

Callens and Remon [46] have shown improved nasal absorption of insulin in rabbits by using a bioadhesive powder formulation containing drum-dried waxy maize starch (mainly amylopectin) and Carbopol 974P. The bioavailability of 14% has been obtained. They have shown [47] that the initial advantage of a longer residence time of the powder formulation in the nasal cavity might turn into a disadvantage after multiple administration and impact bioavailability. They investigated the influence of eight daily administrations of two powder formulations to rabbits on the bioavailability and therapeutic effect of the insulin [47]. The first powder formulation consisted of a co-spray-dried mixture of Amioca starch and Carbopol 974P and the second one has been a physical mixture of drum-dried waxy maize starch and Carbopol 974P. By a single nasal administration to rabbits, absolute bioavailabilities of 17.8 and 13.4% have been obtained, respectively. The lower insulin bioavailabilities (4.4 and 3.6%, respectively) after multiple administrations were observed with both formulations, mainly due to the high viscosity of the bioadhesive powders in the nasal mucus, causing a physical barrier toward absorption and a strongly decelerated mucociliary clearance. Long residence times of the powder formulations were also reported by Ugwoke et al. [42, 48], who noticed nasal residence times of more than 24 h using powder formulations containing Carbopol 971P and carboxymethyl cellulose.

Rhinocort is a commercially available mucoadhesive transnasal powder preparation of beclomethasone dipropionate with hydroxypropyl cellulose (HPC) as a gel-forming drug carrier developed by Suzuki and Makino [49]. The HPC has been shown to promote the absorption of low-molecular-weight drugs, but it was not that effective with a peptide drug salmon calcitonin. Microcrystalline cellulose has been shown to be effective for the promotion of absorption of calcitonin in humans, producing about 10% bioavailability with rapid absorption onset [49]. In the study

of the effect of an HPC and MCC combination on the development of nasal powder preparations for peptide delivery, significant absorption enhancement of leuprolide, calcitonin, and fluorescein isothiocyanate (FITC)–dextran in rabbits has been obtained by the addition of 10–20% HPC to MCC. It has been suggested that MCC works as an absorption enhancer by causing a locally high concentration of drugs in the vicinity of the mucosa surface while HPC works to increase retention of drugs on the nasal mucosa due to its gel-forming property. In a comparative study [50] of a series of MCC nasal sprays and lyophilized powder formulations of ketorolac, the spray formulations have been shown to be better absorbed than powder formulations. The absolute bioavailability of ketorolac from a powder formulation has been 38%, and no significant differences in absorption between different powder formulations have been observed.

Lim et al. [51, 52] compared a number of mucoadhesive microspheres prepared by solvent evaporation composed of hyaluronic acid (HA), chitosan glutamate (CH), and a combination of the two with microcapsules of HA and gelatin prepared by complex coacervation. Some other polymers—such as alginates [53, 54], a natural polymer of low toxicity, irritability, and immunogenicity; epichlorohydrine cross-linked starch (Spherex) [55–61] and dextran (Sephadex) [62–65]; poly(lactide-*co*-glycolide) (PLGA) [66]; and the biocompatible and biodegradable copolymer of lactic and glycolic acids, which have also been approved by the Food and Drug Administration (FDA) [67]—have mainly been used in microspheres for nasal dry powder delivery and are referred to in more detail in the next section.

Recently, thiolated polymers or thiomers, a new generation of permeation-enhancing agents, have been introduced in the pharmaceutical literature. Thiomers are characterized by covalent immobilization of sulfhydryl groups on their polymeric backbone, which are responsible for improved permeation-enhancing properties combined with mucoadhesive and enzyme-inhibitory properties [68]. A further improvement of the permeation-enhancing effect of thiomers has been achieved by the addition of the permeation mediator glutathione [69]. The improvement of human growth hormone (hGH) bioavailability (8.11%) by intranasal administration of the microparticulate formulation composed of thiomer polycarbophil-cysteine (PCP-Cys) and permeation mediator glutathione has been shown. Evaluation of the effect of PCP-Cys on the ciliary beat frequency (CBF) of human nasal epithelial cells in vitro has shown no ciliotoxic effect [70].

Chitosan is a hydrophilic, biocompatible, and biodegradable polymer of low toxicity, and it has been extensively investigated for pharmaceutical and medical purposes. It has been included in the European Pharmacopoeia since 2002. Chitosan is a polysaccharide composed of *N*-acetyl-D-glucosamine (approximately 20%) and glucosamine (approximately 80%). It is derived by deacetylation of chitin, which after cellulose is the most abundant polymer found in nature. It is a polycation at acidic pH values where most of the amino groups are protonated and has an apparent $pK_a$ of 5.5 [71–74]. In the context of drug delivery, chitosan has been used for the preparation of microcapsules and microspheres with encapsulated small polar molecules, proteins, enzymes, DNA, and cells, as a nasal delivery system for insulin [75], as a system for oral vaccination, and as a stabilizing constituent of liposomes. Several studies have highlighted the potential use of chitosan as an absorption-enhancing agent due to its mucoadhesive properties and ability to open the tight junctions in the mucosal cell membrane [72].

Its biodegradability and low toxicity in humans have aided the recent increased interest in chitosan as an immunopotentiating agent. In vivo studies have demonstrated that chitosan powder and solution formulations are able to enhance the systemic and mucosal immune responses after nasal vaccine delivery [19, 22, 76].

The nasal absorption of insulin after administration in chitosan powder was the most effective formulation for nasal delivery of insulin in sheep compared to chitosan nanoparticles and chitosan solution [11]. Similarly, chitosan powder formulations have been shown to enable an efficient nasal absorption of goserelin in a sheep model where bioavailabilities of 20–40% were obtained depending on the nature of the formulation [9].

There has been a report on chitosan utility in improving the intranasal absorption of high-molecular-weight (>10-kDa) therapeutic protein. Chitosan glutamate powder blend or granules with recombinant hGH have been evaluated for intranasal administration in sheep. Relative to subcutaneous injection the nasal formulations produced bioavailabilities of 14 and 15%, respectively [77].

Various chitosan derivatives of enhanced solubility, mucoadhesive, and permeation properties were developed. *N*-Trimethyl chitosan chloride (TMC) is a quaternized derivative of chitosan with superior aqueous solubility over a broader pH range and penetration-enhancing properties under physiological conditions [78]. Carboxymethylated chitosan (CMChi) is a polyampholytic polymer able to form viscoelastic gels in aqueous environments. CMChi appears to be less potent compared with the quaternized derivative. Neither TMC nor CMChi have been found to provoke damage of the cell membrane, and therefore, they should not alter the viability of nasal epithelial cells [79].

Thiolated chitosans, chitosan thioglycolic acid conjugates, chitosan–cysteine conjugates, and chitosan-4-thio-butyl-amidine conjugates are new-generation polymers that are pH sensitive. The pH range at which the gelation and mucoadhesion of these polymers are optimal is within the physiological range (pH 5–6.5) of the nasal mucosa, but these polymers have been primarily investigated for oral drug delivery [80].

## 4 MICROSPHERES AS NASAL DRUG DELIVERY DEVICES

Developing an appropriate drug delivery system for a given drug can completely alter the drug's unfavorable properties, such as improve its effectiveness or reduce its side effects. Dry powder delivery systems such as microspheres are of special interest. In the last two decades they have been extensively studied with respect to nasal delivery and a considerable number of studies have been reported on that subject [3, 23].

In general, microspheres as specialized drug delivery systems represent spherical polymeric devices that are small in size (from 1 to 1000 μm), are characterized by high surface-to-volume ratio, and are able to provide targeted and predictable controlled release of the drug [3]. In the scope of nasal delivery, except for controlled drug release rate, microspheres are beneficial due to their broad surface area, which can provide extensive interaction with the mucin layer and protection of incorporated drug from enzymatic degradation in the nasal cavity [10, 41].

Microspheres prepared with bioadhesive polymers have some additional advantages; they assure much more intimate and prolonged contact with the mucous layer and improved drug absorption due to additional delay in mucociliary clearance. Bioadhesive microspheres can significantly improve patient compliance as all the advantages described lead to reduction in the frequency of drug administration [3, 74].

Bioadhesive microspheres that have been extensively studied for nasal drug delivery are water insoluble but they swell in contact with the mucosa. Swollen microspheres form a gellike system that adheres onto the mucus, retaining drug at the absorption site for prolonged periods [2, 81]. Swelling of the microspheres causes mucosal dehydration and reversible shrinkage of the cells, resulting in the temporary widening of the tight junctions and increased permeability of hydrophilic compounds, or more precisely, paracellular absorption of the drug [62]. Oechslein et al. [40] suggested that the opening of tight junctions could be related to the local decrease in $Ca^{2+}$ concentration as well, since the absorption-promoting effect of investigated particulate drug delivery systems correlated directly with their capability to bind $Ca^{2+}$. In order to additionally improve nasal drug absorption, bioadhesive particulate systems have been combined with biological absorption enhancers [55–57].

A number of studies of the nasal mucociliary clearance rate of microspheres confirmed their potential to retain drug at the absorption site longer than liquid formulations [58, 67, 81, 82, 83]. Methods to measure formulation clearance rates from the nasal cavity can be divided into three groups, differing in the detected substance [58]. The most exact method is gamma scintigraphy, which monitors the deposition and clearance of radiolabeled drug delivery systems. The second method involves mixing of a fluorescent dye with the formulation and monitoring the cumulative tracer amount in the pharynx. The third method is the saccharin test, in which saccharin is mixed with the formulation and the clearance rate is determined by the first perception of sweet taste [84]. The first study of the mucociliary clearance of microspheres using gamma scintigraphy was reported by Illum et al. [81]. They evaluated clearance rates of starch, dextran, and albumin microspheres: Three hours after nasal administration about 50% of albumin and starch microspheres and 60% of dextran microspheres were still detected at the site of deposition. Nasal clearance study of melatonin starch microspheres and the melatonin solution applied revealed that more than 80% of the starch microspheres remained in the nasal mucosa 2 h after administration, compared to only 30% for the melatonin solution [58]. The study of the clearance rate of alginate, PLGA, and Sephadex microspheres revealed that alginate and PLGA were suitable for nasal delivery as they had the best mucoadhesive properties [67]. It has also been shown that the limiting step of the mucociliary clearance of nasally administrated microspheres was their clearance from the initial deposition site. The same conclusion has been drawn for the Carbopol 971P and carboxymethyl cellulose microspheres [82]. Soane et al. [83] evaluated the clearance rate of chitosan microspheres and chitosan solutions, compared to the control solution from the nasal cavity in sheep, by gamma scintigraphy. They found that both chitosan systems had higher retention times compared to the control. Also, chitosan microspheres were cleared at a slower rate than the chitosan solution, with half times of clearance of 115 and 43 min, respectively. The nasal clearance rates found in the sheep model were similar to the clearance rates found in their previous

study carried out on human subjects [85], indicating that the sheep could be a suitable model for in vivo nasal clearance studies.

***Starch Microspheres*** Bioadhesive starch microspheres in the context of the nasal delivery system were first introduced by Illum et al. [81] in a study that examined human nasal mucociliary clearance. Since then starch microspheres have been shown to promote the nasal absorption of a number of drugs. Animal studies using the sheep model showed greatly improved absorption of gentamicin [59] and human growth hormone [56] when administered in combination with starch microspheres in a freeze-dried formulation. Similar findings have been reported for insulin [55, 60] and desmopressin [61] loaded starch microspheres compared to simple drug solutions administered to animal models. Biodegradable starch microspheres have also been investigated for nasal delivery of metoclopramide whereas enhanced bioavailability was achieved compared to nasal spray [86]. The bioadhesive starch microspheres were shown to act synergistically with the absorption enhancers improving the transport of insulin across the nasal membrane [57]. Recently, the potential of starch microspheres for the nasal delivery of melatonin was investigated [58]. An in vitro release study revealed a sustained drug release profile. Melatonin bioavailability after nasal administration of starch microspheres was high, 84%. A good correlation between the in vitro release profile and in vivo absorption has been observed.

The use of degradable starch microspheres has proved to be well tolerated in both experimental animals and humans. No alterations of nasal mucosa were detected after eight weeks of nasal administration of starch microspheres to rabbits. Additionally, a preliminary test on healthy volunteers also showed good acceptability [62]. Another study of healthy volunteers revealed no changes in mucociliary clearance or in the geometry of the nasal cavities after eight days of nasal administration of dry starch microspheres [87].

***Dextran Microspheres*** Similar to starch microspheres, Illum et al. [81] introduced dextran microspheres as a bioadhesive drug delivery system able to prolong the residence time in the nasal cavity. However, Rydén and Edman [65] reported dextran microspheres were not shown to improve nasal absorption of insulin in rats as insulin was too strongly bound to the Diethylaminoethyl (DEAE) groups to be released by a solution with an ionic strength corresponding to physiological conditions. In a later study it has been shown that the localization of insulin influenced the in vivo behavior of dextran microspheres [63]. The distribution of insulin at the surface or inside the dextran microspheres after the lyophilization loading process was determined by the cut-off limit of the microspheres. Microspheres with insulin left at the surface showed higher insulin absorption–enhancing effect than the microspheres with insulin inside the dextran matrix. Dextran microspheres have also been evaluated in vivo as a delivery system for octreotide [40] and in vitro for nicotine [33]. Ciliotoxicity studies performed in vitro on explants from rat trachea showed that dextran microspheres had no effect on the ciliary beat frequency [64]. The immediate recovery of the ciliary movement after dextran microspheres washing off indicated that the cilia were not damaged by dextran microspheres.

***Gelatin Microspheres*** Several studies characterizing gelatin microspheres as a nasal drug delivery system have been reported. Gelatin microspheres were shown

to swell readily in contact with nasal mucosa and to have good bioadhesive properties [65]. An in vitro release study using a Franz diffusion cell on levodopa-loaded gelatin microspheres showed prolonged drug release as compared to drug alone [88]. Negatively and positively charged gelatin microspheres intended for nasal and intramuscular delivery of salmon calcitonin were prepared by Morimoto et al. [10]. Both types of microspheres enhanced nasal absorption of salmon calcitonin compared to the solution. Positively charged gelatin microspheres seamed to exhibit greater enhancing effect on nasal absorption than negatively charged gelatin microspheres. Recently gelatin and gelatin–poly(acrylic acid) microspheres were studied with respect to oral and nasal delivery of oxprenolol [89]. Combining the gelatin with poly(acrylic acid) resulted in microspheres with improved bioadhesive properties. Also, nasal administration of gelatin–poly(acrylic acid) microspheres resulted in improved bioavailability of the drug compared to nasal administration of the drug solution.

***Polyacrylate Microspheres*** Cross-linked polyacrylate microspheres as nasal powder delivery systems have been investigated in several studies [39, 41, 90]. Microspheres were produced by spray drying and emulsification methods and their nasal drug delivery potential has been evaluated only in vitro. Carbopol 934P microspheres were shown to have the best bioadhesive properties compared to other hydrophilic microspheres prepared with polyvinyl alchohol, chitosan, and hydroxypropylmethyl cellulose [41]. Improved permeation-enhancing effect of polycarbophil microparticles was obtained when microparticles were prepared with the thiolated polycarbophil and the permeation mediator glutathione [68].

***Chitosan Microspheres*** As a specific chitosan-based delivery system, chitosan microspheres have been extensively studied and number of reports has verified their potential regarding nasal drug delivery [9, 21, 25, 41, 53, 73, 91, 92]. Chitosan microspheres have been prepared by the emulsification solvent evaporation method [51, 52, 93], emulsification cross-linking process [91], spray drying method [25, 53], and ionic gelation process [94]. They have been shown to significantly reduce mucociliary clearance from the nasal cavity of sheep and humans compared to solutions [83, 85]. The bioadhesive properties of chitosan microspheres were shown to be inversely proportional to particle size: Among chitosan microspheres in the size class between 50 and 200 μm, smaller microspheres appeared to swell faster than large microspheres, providing a more powerful mucoadhesive system [41].

A modulated release rate of drug from the swellable chitosan microspheres has been achieved with cross-linking agents such as glutaraldehyde [95], citric acid [92], ascorbic acid, or ascorbyl palmitate [91] that reacted with chitosan forming covalent bonds with chitosan amino groups. However, to maintain the bioadhesive properties of cross-linked chitosan microspheres, the amount of cross-linking agent should be optimized [95].

Chitosan molecular weight has also been reported to influence drug release. Jiang et al. [94] studied *Bordetella bronchiseptica* dermonecrotoxin (BBD) release from chitosan microspheres prepared by tripolyphosphate ionic gelation. It has been shown that the BBD release rate increased with chitosan molecular weight decrease. It has been explained by the weaker BBD interaction with chitosan of lower molecular weight and lower content of free amine groups, responsible for their interaction.

Chitosan microspheres were shown to enhance nasal bioavailability of several peptide drugs such as insulin and goserelin. A simple chitosan–insulin powder formulation provided about 20% of absolute insulin bioavailability in sheep [96]. Improved bioavailability (of 44%, in rats) was obtained when insulin was loaded into chitosan microspheres prepared with ascorbyl palmitate as cross-linking agent [91]. Chitosan microspheres have also been shown to improve nasal goserelin absorption providing about 40% bioavailability relative to goserelin intravenous application [9].

Krauland et al. [93] prepared the microparticles with thiolated chitosan (chitosan-TBA; chitosan–4-thiobutylamidine conjugate) intended for nasal peptide delivery. During the preparation process microparticles were stabilized by the formation of inter- and intramolecular cross-linking via disulfide bonds. Chitosan–TBA microparticles were characterized by improved swelling ability and displayed 3.5-fold higher insulin bioavailability compared to unmodified chitosan microparticles.

Besides the polymer derivatization, combining the polymers in microsphere preparations can result in improved drug delivery and absorption characteristics. Hyaluronic acid–chitosan microspheres appeared to improve the absorption of incorporated gentamicin compared to the individual polymers, assembling the mucoadhesive potential of both polymers and the penetration-enhancing effect of chitosan [51, 52].

Chitosan microparticulate systems have also been investigated for vaccine nasal delivery and have proven to induce strong systemic and mucosal immune responses [18, 21, 76].

### 4.1 Preparation Methods

The design of bioadhesive microspheres includes selection of the most suitable preparation method, considering the nature of the drugs and polymers used as well as the route of administration. A number of methods for the preparation of microspheres have been described in the literature [3, 97]. In the scope of nasal delivery, the first microspheres in use were starch and dextran microspheres, prepared by an emulsion polymerization technique employing epichlorohydrine as a cross-linking agent [55, 56, 59, 60, 62, 81]. Currently techniques based on solvent removal, such as solvent evaporation [41, 51, 66, 93, 98] and solvent extraction [88, 99], are most frequently in use. There are tree processes involved in such microensapsulation procedures: the preparation of emulsion, solvent removal, and separation of the particles obtained. Selection of the type of (oil-and-water) emulsion system (O/W, W/O, W/O/W, W/O/O, etc.) depends on the physicochemical properties of the drug and polymer used. After the preparation of stable emulsion, solvent is removed from the system at high or low temperature, at low pressure, or by addition of another solvent that enables the extraction of polymer solvent to the continuous phase. Hardened microspheres are then washed, centrifuged, and lyophilized.

Emulsion techniques are suitable for the preparation of microspheres intended for nasal delivery since they allow controlling the size of the particles. Freiberg and Zhu [97] reviewed solvent evaporation process parameters (e.g., polymer concentration, viscosity, stirring rate, temperature and percentage of emulsifying agent) affecting microsphere size. It can be assumed that the particle size is directly proportional to polymer concentration and inversely proportional to stirring rate and percentage of emulsifying agent [41, 88], while there is a nonlinear correlation between particle

size and process temperature. Yang et al. [100] reported that larger microspheres were produced at lower temperatures due to the higher viscosity of solution and at higher temperatures due to the higher solvent flow pressure moving more material from the microsphere center outward. In the same work encapsulation efficiency was also correlated with the temperature of solvent evaporation in the process of microsphere preparation. It was found that the highest encapsulation efficiencies occurred at the lowest and highest temperatures tested.

Recently, the spray drying method has been extensively used for the preparation of microspheres intended for nasal delivery [3, 25, 39, 44, 45, 53, 54, 101, 102]. Spray drying is a single-step procedure transforming liquid into dry particulate form (e.g., microparticles, microspheres, microcapsules) applicable to drugs and polymers of various solubility characteristics. It is a fast, simple, and reproducible technique that is easy to scale up [3]. Spray drying can be described as follows: The liquid is fed to the nozzle with a peristaltic pump, atomized by the force of the compressed air, and blown together with hot air into the chamber, where the solvent in the droplets is evaporated. The dry product is then collected in a collection bottle. Spray-dried microspheres are reported to have relatively low production yields, rarely higher than 50%. The loss of material during spay drying has been explained by the powder adhering to the cyclone walls, small amounts of materials processed in each batch, and loss of the smallest and lightest particles through the exhaust of the spray dryer apparatus, which lacks a trap to recover the lighter and smaller particles [25, 53, 103, 104].

Spray drying offers the possibility to control the particle size of the product. Microparticles of desirable size can be obtained by optimizing the spray drying process parameters, such as size of the nozzle, feeding pump rate, inlet temperature, and compressed airflow rate. In accordance with this, He et al. [105] reported that larger particles were formed at a larger size of nozzle and faster feeding pump rate, while smaller particles were formed at a greater volume of air input [105]. The size and other properties of spray-dried microspheres (e.g., morphology, density, shape, porosity, and flowability) can also be affected by the qualitative and quantitative composition of the liquid feed [28]. Thus, feed concentration has been reported to influence the particle size distribution, as spray drying of more concentrated liquid feeds resulted in the formation of larger particles [104, 106, 107].

### 4.2 Microsphere Characterization

Microspheres intended for nasal administration need to be well characterized in terms of particle size distribution, since intranasal deposition of powder delivery systems is mostly determined by their aerodynamic properties and particle sizes. Commonly used methods for particle size determinations described in the literature are sieving methods [108], light microscopy [58], photon correlation spectroscopy [66], and laser diffractometry [25, 41, 53, 93]. The morphology of the microparticles (shape and surface) has been evaluated by optical, scanning, and transmission electron microscopy [66, 95].

Determination of the zeta potential is an important part of microsphere characterization, as the zeta potential has a substantial influence on the adhesion of drug delivery systems onto biological surfaces [109]. For example, Jaganathan and Vyas [110] reported the reduction in the nasal clearance rate of PLGA microspheres modified with chitosan compared to unmodified PLGA microspheres due to the change in zeta potential from negative for PLGA microspheres to positive for surface-modified

PLGA microspheres. Methods to measure the zeta potential of microspheres are laser doppler anemometry [105] and photon correlation spectroscopy [110].

The physical state of the drug incorporated in a powder drug delivery system (e.g., degree of crystallinity and possible interactions with the polymer) is assessed by differential scanning calorimetry (DSC) or Fourier transform infrared (FTIR) spectroscopy. These observations can clarify the results of other parameter investigations, especially the results of in vitro drug release studies.

To predict microsphere performance in vivo, the swelling properties of nasal powder delivery systems need to be evaluated. Methods described in the literature are mostly based on the weight difference measurements between the dry and swollen powder [40]. Swelling properties of nasal powders such as water-absorbing capacity can be evaluated using a Franz diffusion cell [43, 107]. The swelling capacity may also be expressed as the volume expansion of the microspheres that is determined at equilibrium after placing the microspheres in water using a graduated cylinder [108]. Gavini et al. [53] determined the swelling properties of microspheres in vitro by laser diffractometry. That method allows us to evaluate the variation of particle size versus time.

In vitro evaluation of mucoadhesive properties is essential in the development of a nasal drug powder delivery system, since mucoadhesion is of great importance for the in vivo performance of formulation. A large number of in vitro and in vivo methods used to assess mucoadhesive properties of microspheres have been extensively described in the literature [3, 111, 112]. Many in vitro methods are based on the interaction of microspheres with mucin. Evaluation of that interaction can be performed using scanning and transmission electron microscopy [95] or photon correlation spectroscopy [45]. Scanning electron microscopy (SEM) provides the information on morphological changes on the microsphere surface in contact with mucin, while transmission electron microscopy confirms SEM results and reveals the ultrastructural features of the surface interactions between microspheres and mucin chains [95]. He et al. [102] evaluated the mucoadhesive properties of chitosan microspheres by measuring the amount of mucin adsorbed on the microspheres. Gavini et al. [53] evaluated the mucoadhesive properties of metoclopramide-loaded microspheres by determining the amount of microspheres that stuck to a filter paper saturated with mucin after exposure to the air stream. Vidgren et al. [39] used a tensiometer to measure the force required for the separation of two filter paper discs saturated with mucin and with the examined microspheres placed between them.

In the work reported by Witschi and Mrsny [54] mucoadhesion of dry powder microparticles was investigated using Callu-3 cells as a surrogate for human nasal epithelia: Microparticles were applied to the apical surface of cell sheets and at certain time points were washed with phosphate-buffered saline (PBS) to remove poorly adhering microparticles. Rango Rao and Buri [113] developed an in situ method to evaluate the bioadhesive properties of polymers and microparticles, based on washing off a mucous membrane covered with the formulation to be tested by simulated biological flow. The mucoadhesion of gelatin microspheres [10] was measured by an in situ nasal perfusion experiment. In the work reported by Lim et al. [51] the mucoadhesive properties of microspheres were evaluated by determining the mucociliary transport rate of the microparticles across an isolated frog palate. The boiadhesive properties of microspheres can be evaluated by the everted sac technique using a section of everted intestinal tissue or the CAHN dynamic contact angle analyzer [3].

Santos et al. [112] correlated these two methods and concluded that each method could be used alone as the relevant indicator of microsphere bioadhesion.

In vitro drug release experiments can be performed in order to characterize the release behavior of microparticles in general. For that purpose microparticles can be dispersed directly in the dissolution medium [51, 91] or a dynamic dialysis technique can be employed [58]. However, to obtain results comparable with the in vivo situation of nasal administration, it is necessary to provide experimental conditions similar to those encountered in the nasal cavity as nasally administrated powders are not being dispersed directly in the large quantity of liquid [8]. Such in vitro drug release experiments can be performed by a modified U.S. Pharmacopeia (USP) XXII rotating basket [8]. The drug-loaded powder formulation is weighed on a membrane filter placed between the filter holder and the cup and then immersed in the released medium. Thus, the membrane filter separates the donor and acceptor compartment but at the same time allows the powders to hydrate and to form a gel. Drug is released to the release medium after diffusion through the swollen gel of known surface area. Cornaz et al. [33] developed a special diffusion chamber that simulated the hydration conditions of the nasal mucosa to study the in vitro release of nicotine from dextran microspheres. A number of authors have used Franz diffusion cells for in vitro release experiments since that model provides conditions similar to those encountered in the nasal cavity and slow hydration of the microspheres [41, 43–45, 107].

### 4.3 Chitosan-Formulated Spray-Dried Microspheres

Chitosan, a biocompatible and biodegradable polycationic polymer with low toxicity, is known for its swelling ability and permeation-enhancing properties and represents a polymer of choice for the preparation of microspheres intended for nasal administration [74].

Spray drying has proved to be a suitable and simple technique for the preparation of chitosan microspheres with preserved chitosan properties, offering numerous advantages over other microencapsulation methods. It has been successfully used to entrap both hydrophilic and lipophilic drugs into the chitosan matrix, since a variety of colloidal systems (e.g., polymer solutions, emulsions, dispersions, suspensions) can be subjected to spray drying. Chitosan-based spray-dried microspheres have been prepared with chitosan alone (resulting in conventional microspheres) or in combination with another polymer (resulting in composed microspheres). Combining the polymers has been reported to result in microspheres with improved properties regarding surface characteristics, entrapment efficiency, or control over the drug release rate. Recently, several drugs, such as carbamazepine [25], propranolol hydrochloride [45], metoclopramide hydrochloride [53], and loratadine [103, 107], have been incorporated into chitosan-based nasal powder formulations produced by the spray drying-method. They were characterized in terms of encapsulation efficiency, morphology, size distribution, zeta potential, physical state of the drug, in vitro drug release behavior, and swelling and bioadhesive properties. Chitosan–ethyl cellulose composed microspheres improved loratadine entrapment compared to conventional chitosan microspheres, which influenced directly the microsphere surface characteristics: Loratadine was less present at the surface of the microspheres and consequently had less influence on their bioadhesive proper-

ties. Thus, although showing moderate swelling ability, loratadine-loaded composed microspheres were more bioadhesive than conventional chitosan microspheres. Composed microspheres showed good loratadine-sustained release potential in vitro, depending on the polymeric weight ratio and concentration of the spray-dried system [103, 107]. Gavini et al. [53] produced metoclopramide-loaded alginate and/or chitosan microspheres by the spray drying method. The results obtained revealed that complexation of chitosan with alginate in the microsphere preparation provided improved control of the drug release in vitro compared to chitosan alone. Despite the chitosan complexation, composed microspheres showed good bioadhesive properties. Ex vivo drug permeation tests carried out using sheep nasal mucosa showed higher drug permeation from chitosan-based microspheres than from alginate microspheres, confirming the well-known chitosan permeation-enhancing properties. Cerchiara et al. [45] developed spray-dried chitosan–poly(methyl vinyl ether-*co*-maleic anhydride) microparticles for nasal delivery of propranolol hydrochloride. Chitosan was combined with polyanhydride, able to enhance the formation of hydrogen bonds between the polymers and mucosal components through carboxylic acid groups generated after polyanhydride hydrolytical degradation. The swelling and bioadhesive properties of chitosan–polyanhydride microparticles increased in a pH-dependent manner. Both chitosan and chitosan–polyanhydride microparticles provided sustained propranolol hydrochloride release.

Microparticulate spray-dried delivery systems have shown great potential for nasal delivery of drugs characterized by poor water solubility. According to DSC analysis, the spray drying method together with the carriers seemed to promote the amorphization of loratadine [107] and carbamazepine [25]. Carbamazepine incorporated into chitosan microspheres was characterized by increased dissolution rate compared to carbamazepine raw material. It was explained not only by the promoted drug amorphization but also by the chitosan well-known dissolution rate enhancer properties and by the small size of microspheres (or high surface-to-volume ratio). Results of in vivo nasal administration of carbamazepine-loaded chitosan microspheres revealed a remarkable increase in carbamazepine concentration in serum compared to the pure carbamazepine powder [25]. Such an increase in drug absorption has been ascribed to both improved dissolution of carbamazepine and adhesion of the chitosan microspheres to the mucosal surface.

The mucoadhesive function of chitosan has also been employed in vaccine dry powder delivery. Alpar et al. [18] produced bovine serum albumin (BSA)–loaded chitosan microspheres using the spray drying method. It has been shown that the stability of encapsulated BSA was preserved in the microspheres prepared, indicating that spray drying was appropriate even for the preparation of antigen-loaded microspheres. BSA-loaded chitosan microspheres generated higher immune response than the free BSA, thus proving to be a suitable system for nasal antigen delivery.

## 5 TOXICOLOGICAL CONSIDERATIONS

The possible toxicological effects of dry powder formulation on the nasal mucosa, including local irritation, effect on mucociliary clearance, and epithelial damage and recovery rate [114], should be investigated in an early stage of its development. There were some attempts to define the categories of toxic effects as well of the constituents of nasal formulations according to their toxic potential. Thus, Hvidberg

et al. [115] introduced a scale of irritation in the nose (0, no irritation; 1, slight irritation; 2, acceptable; 3, unwilling to accept the treatment again) that was later used to evaluate the degree of nasal powder irritation after administration in human volunteers [27]. Ugwoke et al. [82] classified the degree of ciliary beat frequency change caused by nasally applied liquid formulation as follows: no effect (less than 10%), mild (10–20%), moderate (20–50%), and severe (more than 50%). Soon after, that classification was applied to a powder ciliotoxicity study [116]. Reversibility of ciliotoxic effect after washing out the tested compound was classified as well, resulting in three categories: reversible, partially reversible, and irreversible effects [82]. Merkus et al. [117] proposed the three categories of constituents of nasal formulations based on the recovery of ciliary beat frequency after the tested compound was washed out. The first category is cilio friendly, with ciliary beat frequency recovery of 75% or more; the second is cilio inhibiting, with recovery between 25 and 75%; and the third category is ciliostatic, with recovery of 25% or less.

Methods to evaluate the possible toxicological effect of formulations on the nasal mucosa described in the literature mainly refer to histopathological evaluation as a standard method for cytotoxicity evaluation [64] and a study on the release of marker enzymes [118]. In the work reported by Callens et al. [118] the possible toxicological effects of multiple starch and carbopol powder nasal administration were evaluated using rabbits by measuring the proteins and lactate dehyrogenase (LDH) release from the nasal mucosa. Contrary to the invasive in situ perfusion method, performable with anaesthetized animals [119], this method has been shown to be noninvasive, applicable to nonanaesthetized and nonsedated animals, and suitable for repeated measurement of the marker protein release on the same animal during a long-term administration study. A histopathological study has also been performed. In agreement with attempting to replace the use of vertebrates in scientific experiments with lower organisms such as invertebrates, plants, and microorganisms, Adriaens and Remon [120] developed a new mucosal toxicity screening method using the slug *Arion lusitanicus* as the model organism. The body wall of the slug resembles the nasal mucosa since it consists of a single-layered epithelium containing both ciliated and mucus-secreting cells. Callens et al. [118] successfully used that method for screening the mucosal irritation potential of bioadhesive starch and carbopol powder formulations. The possible toxicological effects of the powder formulations were evaluated by measuring the proteins LDH and alkaline phosphatase (ALP) release from the body wall of the slugs as well as the amount of mucus produced.

Among various nasal toxicity studies, ciliotoxicity studies are of special interest due to the importance of maintaining optimal ciliary beating to protect the lower respiratory system from infections. If the drug formulation inhibits the ciliary beating, such inhibition needs to be completely reversible upon formulation removal [116].

Methods to determine ciliary beat frequency and mucociliary transport in vitro and in vivo have been extensively reviewed elsewhere [84, 121]. In general, in vivo methods are more reliable for ciliotoxicity studies than in vitro methods and are essential to confirm the safety of nasal drug formulations. However, in vitro methods are more suitable for the large number of screening studies required during formulation development [122].

In vitro methods to measure ciliary beat frequency can be performed on explants of ciliated mucosa [64] or on different types of ciliated cells, such as ciliated chicken embryo trachea cells [123], nasal cell lines derived from carcinomas of epithelial origin (RPMI 2650, BT, NAS 2BL), or human lung adenocarcinoma cell line Calu-3

[124]. In vitro ciliotoxicity of dextran microspheres was evaluated on explants from rat trachea [64].

Human nasal epithelial cell cultures can serve as a relevant screening tool for prediction of nasal formulation toxicity in humans as long as the cells maintain differentiated morphological and biochemical characteristics of the original tissue [124]. Epithelial cells intended for initiation of primary cell culture should be taken from the regions of the nasal cavity where formulations are supposed to be deposited [125]. Considering toxicological investigations, human nasal epithelial cells cultured in an air–liquid interface system are the most promising ones at the moment, as in these culturing conditions cell differentiation closely resembles cell differentiation in vivo [124]. Monolayer immersion feeding and air–liquid interface cultures have already been used for ciliary beat frequency measurements. However, unstable ciliary activity and their short life span in the culture have proved to be the main shortcomings of these systems. Jorissen et al. [126] developed a suspension culture system of human nasal epithelium with actively beating cilia for several months. That system has been validated for ciliotoxicity investigations [122] and later applied in other studies [116, 127]. At this time it is the most successful model of human nasal epithelium culture useful for ciliary beat frequency determination [124].

In vivo methods to determine the toxicological effects of drug formulations on the nasal mucosa mainly refer to mucociliary clearance rate studies. Methods described in the literature are mostly based on the gamma scintigraphic technique, in which cleared formulations are labeled with the radiotracer [82]. In vivo scintigraphic evaluation of the nasal clearance of drug delivery systems requires a nondiffusible and stable radiotracer to prevent its absorption and decomposition. Another method described in the literature is the saccharin test [87], in which saccharin is mixed with the formulation and the clearance rate is determined by the first perception of sweet taste [84]. Gamma scintigraphy [31] is more relevant for the ciliary function monitoring than the saccharin test, since it investigates the whole mucosal surface, while the saccharin test investigates only the fastest flow rate [82].

A number of ciliotoxicity studies have pointed out a low correlation between the results obtained using different in vitro and in vivo methods [121]. The effects of nasal formulations on the ciliary beat frequency in vitro are usually more expressed than in vivo, since in vivo, cilia are partially protected by the mucous layer and investigated formulation is eventually cleared from the nasal cavity due to the mucociliary clearance mechanism. Also, toxic effects of the formulations on the cilia in vivo may be reversible due to the constant nasal mucosal cell turnover [121].

In conclusion, in order to make predictions regarding the safety of the nasal formulation on mucociliary clearance, both in vitro and in vivo studies have to be performed. It is also essential to determine long-term use effects in animals and in humans if the nasal formulation is intended for subchronic or chronic administration.

Dry powder formulations for nasal delivery of peptides and proteins have been investigated for the first time by Nagai and others [38]. Since then, much research work has been done on dry powders containing bioadhesive polymers for nasal drug administration. The bioavailability and duration of action of drugs administered by the nasal route are increased by the use of the principle of mucoadhesion and dry powder formulations. Research work on dry powder formulation containing bioadhesive polymers is summarized in Table 1.

**TABLE 1 Summary of Research Work on Nasal Dry Powder Formulations**

| Powder Formulation | Preparation Method | Drug | Polymer | Studies | Comments | Reference |
|---|---|---|---|---|---|---|
| Microparticles | Spray drying | BSA | Starch, alginate, chitosan, carbopol | In vitro | Chitosan microparticles provided most desirable characteristics for protein delivery | 54 |
| Microparticles | | Budesonide | Polymethacrylic acid–polyethylene glycol [P(MAA–PEG)] | In rabbits | Continuous drug release for at least 8 h<br>High bioavailability | 133 |
| Microparticles | Solvent evaporation | Gentamicin | Hyaluronic acid–chitosan | In rabbits | Polymers improved gentamicin absorption synergistically | 52 |
| Microparticles | Lyophilization | hGH | Polycarbophil–cysteine | In rats | Improved bioavailability | 68 |
| Microparticles | W/O emulsification solvent evaporation | Insulin | Chitosan–TBA | In rats | Improved bioavailability | 93 |
| Microparticles | Spray drying | Propranolol hydrochloride | CH-PVM/MA | In vitro | Sustained drug release | 45 |
| Microparticles, nanoparticles | W/O/W solvent evaporation | Model protein, tetanus toxoid | PLA-PEG | In rats | Size-dependent mucosal uptake | 66 |
| Microspheres | Ionic gelation with tripolyphosphate | Bordetella, bronchiseptica, dermonecrotoxin | Chitosan | In mice | Systemic and mucosal immune responses induced | 21 |
| Microspheres | Ionic gelation process with tripolyphosphate | Bordetella, bronchiseptica, dermonecrotoxin | Chitosan | In vitro | Chitosan molecular weight–related drug release profile | 94 |
| Microspheres | Spray drying | Carbamazepine | Chitosan hydrochloride, chitosan glutamate | In sheep | Increased drug dissolution rate and absorption | 25 |

**TABLE 1** ***Continued***

| Powder Formulation | Preparation Method | Drug | Polymer | Studies | Comments | Reference |
|---|---|---|---|---|---|---|
| Microspheres | W/O/O emulsion solvent evaporation | α-Cobrotoxin | PLGA/P(CPP: CEFB) | In rats | Increased strength and duration of antinociceptive effect | 98 |
| Microspheres | Emulsion polymerization | Desmopressin | Starch | In rats, in sheep | Increased bioavailability with LPC | 61 |
| Microspheres | W/O emulsification solvent evaporation | FITC–dextran | Carbopol 394P<br>Chitosan<br>HPMC<br>PVA | In vitro | Initial release at a constant rate<br>Chitosan microspheres exhibited size-dependent release effect | 41 |
| Microspheres | | FITC–dextran | Chitosan, Carbopol 934P | In rabbits | Improved bioavailability | 131 |
| Microspheres | Cross-linking with epichlorohydrin | Gabexate mesylate | Starch cyclodextrin | In vitro | Fast release rate | 108 |
| Microspheres | Solvent evaporation | Gentamicin | Chitosan, hyaluronan, gelatine | In vitro | Prolonged release, improved mucoadhesive properties | 51 |
| Microspheres | Emulsion polymerization | Gentamicin | Starch | In sheep, in rats | Improved bioavailability with microsphere/enhancer (LPC) system | 59 |
| Microspheres | Spray drying | Gentamicin sulfate | HPMC | In vitro | Modified drug release | 44 |
| Microspheres | Solvent evaporation | Heparin | Poly(lactic acid) | In rats | Sustained-release effect | 134 |
| Microspheres | Emulsion polymerization | hGH | Starch | In sheep | Enhanced nasal absorption with microspheres, rapid and higher absorption with microsphere/enhancer (LPC) system | 56 |
| Microspheres | Emulsion polymerization | Insulin | Starch, dextran | In rats | Rapid absorption | 62 |
| Microspheres | Emulsion polymerization | Insulin | Starch | In sheep | Improved bioavailability with microsphere/enhancer (LPC) system | 55 |

| | | | | | | |
|---|---|---|---|---|---|---|
| Microspheres | Emulsion polymerization | Insulin | Starch | In rats | Improved nasal absorption | 60 |
| Microspheres | Emulsion polymerization | Insulin | Dextran | In rats | Promoted absorption | 63, 65 |
| Microspheres | Emulsification–cross-linking | Insulin | Chitosan | In rats | Promising absolute bioavailability of 44% | 91 |
| Microspheres | Emulsification–solvent evaporation | Insulin | Hyaluronic acid ester | In sheep | Increased nasal absorption | 128 |
| Microspheres | W/O emulsification solvent extraction | Levodopa | Gelatin | In vitro | Initial fast release rate, followed by a second slower release rate | 88 |
| Microspheres | Spray drying | Loratadine | Chitosan, chitosan/ethylcellulose | In vitro | Moderate swelling behavior, sustained drug release | 107 |
| Microspheres | Emulsification–cross-linking | Melatonin | Starch | In rabbits | Increased residence time, rapid absorption rate, high absolute bioavailability | 58 |
| Microspheres | Spray drying | Metoclopramide | Sodium alginate, chitosan, sodium alginate/chitosan | In vitro | Controlled drug release, promising properties as nasal drug carriers | 53 |
| Microspheres | Emulsification–cross-linking | Metoprolol tartrate | Alginate | In rabbits | Improved therapeutic efficacy | 132 |
| Microspheres | W/O emulsification | Morphine | Chitosan, starch | In sheep | High bioavailability | 73 |
| Microspheres | Emulsion polymerization | Nicotine | Dextran | In vitro | Rapid release, good dispersion ability | 33 |
| Microspheres | W/O emulsification solvent extraction | Oxprenolol | Gelatin–poly(acrylic acid) | In rats | Slow-release drug delivery system with good adhesive characteristics | 89 |
| Microspheres | W/O emulsion cross-linking | Pentazocine | Chitosan | In rabbits | Matrix diffusion controlled delivery, improved bioavailability, in vitro/in vivo correlation | 129 |
| Microspheres | Emulsion solvent evaporation | Salbutamol | Chitosan | In rabbits | Prolonged and controlled release | 92 |

**TABLE 1** ***Continued***

| Powder Formulation | Preparation Method | Drug | Polymer | Studies | Comments | Reference |
|---|---|---|---|---|---|---|
| Microspheres | W/O emulsion cross-linking | Salmon calcitonin | Gelatin | In rats | Enhanced nasal absorption | 10 |
| Microsphers | Emulsion polymerization | Insulin | Starch | In sheep | Microspheres and absorption enhancers (LPC, GDC, and STDHF) acted synergistically to enhance absorption | 57 |
| Powder | Lyophilization | Apomorphine | Carbopol 971P, polycarbophil | In rabbits | Sustained release, improved bioavailability | 8 |
| Powder | Lyophilization | Apomorphine | Carboxymethyl cellulose | In rabbits | Sustained plasma level | 42 |
| Powder | Lyophilization | Apomorphine | Carbopol 971P, carboxymethyl cellulose | In rabbits | Increased residence time | 48 |
| Powder | Lyophilization | Apomorphine HCl | Carbopol 971P, Carbopol 974P, polycarbophil | In rabbits | In vitro release but not in vivo absorption has been influenced by drug loading | 130 |
| Powder | Spray drying | Cyanocobalamin | Microcrystalline cellulose, dextran, crospovidone | In rabbits | Improved bioavailability | 43 |
| Powder | Press-on force method | Glucagon | Microcrystalline cellulose | In human volunteers | Increased formulation stability, decreased irritability | 27 |
| Powder | — | Goserelin | Chitosan | In sheep | Improved bioavailability | 9 |
| Powder | Material mixing with a pestle | Insulin | Chitosan | In rats, in sheep | Improved bioavailability | 11 |
| Powder | Blending/ lyophilization | Insulin | Microcrystalline cellulose, hydroxypropyl cellulose, Carbopo 934 | In dogs | Decreased plasma glucose level | 38 |

| | | | | | | |
|---|---|---|---|---|---|---|
| Powder | Lyophilization | Insulin | Starch–Carbopol 974P, maltodextrin–Carbopol 974P | In rabbits | The highest absolute bioavailability obtained was 14.4% | 46 |
| Powder | Lyophilization | Ketorolac | Microcrystalline cellulose | In rabbits | Significantly lower bioavailability of drug from powders compared to spray formulation | 50 |
| Powder | — | Leuprolide, calcitonin, FITC–dextran | Hydroxypropyl cellulose, microcrystalline cellulose | In rabbits | Enhanced absorption | 49 |
| Powder | Manual blending using mortar and pestle | Morphine | Chitosan | In human volunteers | Rapid onset of pain relief, formulations well tolerated by patients | 73 |
| Powder | Dry blending | Octreotide | Dextran, microcrystalline cellulose, semicrystalline cellulose, hydoxyethyl starch, microcrystalline chitosan, pectin, alginic acid | In rats | Correlation between carrier calcium binding properties and their potential as nasal absorption enhancers for peptides | 40 |
| Powder, granules | Powder: manual mixing using mortar and pestle | Recombinant hGH | Chitosan | In sheep | Relative bioavailability of hGH from powder and granules have been 14 and 15%, respectively | 77 |

*Abbreviations:* PLGA: poly(lactide-co-glycolide); P(CPP:CEFB): poly[1,3-bis(p-carboxy-phenoxy) propane-co-p-(carboxyethylformamido) benzoic anhydride]; chitosan–TBA: chitosan–4-thiobutylamidine conjugate; LPC: lysophosphatidyl choline; hGH: human growth hormone; GDC: glycodeoxycholate sodium; STDHF: sodium taurodihydroxyfusidate

## REFERENCES


1. Quraishi, M. S., Jones, N. S., and Tomason, J. D. T. (1997), The nasal delivery of drugs, *Clin. Otolaryngol.*, 22, 289–301.
2. Pereswetoff-Morath, L. (1998), Microspheres as nasal drug delivery systems, *Adv. Drug Deliv. Rev.*, 29, 185–194.
3. Vasir, J. K., Tambwekar, K., and Garg, S. (2003), Bioadhesive microspheres as a controlled drug delivery system, *Int. J. Pharm.*, 255, 13–32.
4. Tirucherai, G. S., Pezron, I., and Mitra, A. K. (2002), Novel approaches to nasal delivery of peptides and proteins, *S. T. P. Pharm. Sci.*, 12, 3–12.
5. Davis, S. S. (1999), Delivery of peptide and non-peptide drugs through the respiratory tract, *Pharm. Sci. Technol. Today*, 2, 450–456.
6. Resta, O., Barbaro, M., and Carnimeo, N. (1992), A comparison of sodium cromoglycate nasal solution and powder in the treatment of allergic rhinitis, *Br. J. Clin. Practice*, 46, 94–98.
7. Schipper, N., Romeijn, S., Verhoef, J., and Merkus, F. (1993), Nasal insulin delivery with dimethyl beta cyclodextrin as an absorption enhancer in rabbits: Powder more effective than liquid formulations, *Pharm. Res.*, 10, 682–686.
8. Ugwoke, M. I., Exaud, S., Van Den Mooter, G., Verbeke, N., and Kinget, R. (1999), Bioavailability of apomorphine following intranasal administration of mucoadhesive drug delivery systems in rabbits, *Eur. J. Pharm. Sci.*, 9, 213–219.
9. Illum, L., Watts, P., Fisher, A. N., Jabbal-Gill, I., and Davis, S. S. (2000), Novel chitosan-based delivery systems for the nasal administration of a LHRH-analogue, *S. T. P. Pharm. Sci.*, 10, 89–94.
10. Morimoto, K., Katsumata, H., Yabut, T., Iwanaga, K., Kakemi, M., Tabata, Y., and Ikada, Y. (2001), Evaluation of gelatin microspheres for nasal and intramuscular administrations of salmon calcitonin, *Eur. J. Pharm. Sci.*, 13, 179–185.
11. Dyer, A. M., Hinchcliffe, M., Watts, P., Castile, J., Jabbal-Gill, I., Nankervis, R., Smith, A., and Illum, L. (2002), Nasal delivery of insulin using novel chitosan based formulations: A comparative study in two animal models between simple chitosan formulations and chitosan nanoparticles, *Pharm. Res.*, 19, 998–1008.
12. Colombo, P. (1999), Mucosal drug delivery, nasal, in Mathiowitz, E., Ed., *Encyclopedia of Controlled Drug Delivery*, J Wiley, New York, pp. 593–605.
13. Ishikawa, F., Murano, M., Hiraishi, M., Yamaguchi, T., Tamai, I., and Tsuji, A. (2002), Insoluble powder formulation as an effective nasal drug delivery system, *Pharm. Res.*, 19, 1097–1104.
14. Ishikawa, F., Katsura, M., Tamai, I., and Tsuji, A. (2001), Improved nasal bioavailability of elcatonin by insoluble powder formulation, *Int. J. Pharm.*, 224, 105–114.
15. Alpar, H. O., Eyles, J. E., Williamson, E. D., and Somavarapu, S. (2001), Intranasal vaccination against plaque, tetanus and diphteria, *Adv. Drug Deliv. Rev.*, 51, 173–201.
16. Vajdy, M., and O'Hagan, D. T. (2001), Microparticles for intranasal immunisation, *Adv. Drug Deliv. Rev.*, 51, 127–141.
17. Davis, S. S. (2005), The use of soluble polymers and polymer microparticles to provide improved vaccine responses after parenteral and mucosal delivery, *Vaccine*, 24, Suppl. 2, S7–S10.
18. Alpar, H. O., Somavarapu, S., Atuah, K. N., and Bramwell, V. W. (2005), Biodegradable mucoadhesive particulates for nasal and pulmonary antigen and DNA delivery, *Adv. Drug Deliv. Rev.*, 57, 411–430.

19. Van der Lubben, I. M., Verhoef, J. C., Borchard, G., and Junginger, H. E. (2001), Chitosan and its derivatives in mucosal drug and vaccine delivery, *Eur. J. Pharm. Sci.*, 14, 201–207.
20. Van der Lubben, I. M., Kersten, G., Fretz, M. M., Beuvery, C., Verhoef, J. C., and Junginger, H. E. (2003), Chitosan microparticles for mucosal vaccination against diphteria: Oral and nasal efficacy studies in mice, *Vaccine*, 21, 1400–1408.
21. Kang, M. L., Kang, S. G., Jiang, H.-L., Shin, S. W., Lee, D. Y., Ahn, J.-M., Rayamahji, N., Park, I.-K., Shin, S. J., Cho, C.-S., and Yoo, H. S. (2006), In vivo induction of mucosal immune responses by intranasal administration of chitosan microspheres containing *Bordetella bronchiseptica* DNT. *Eur. J. Pharm. Biopharm.*, 63, 215–220.
22. Huang, J., Garmise, R. J., Crowder, T. M., Mar, K., Hwang, C. R., Hickey, A. J., Mikszta, J. A., and Sullivan, V. J. (2004), A novel dry powder influenza vaccine and intranasal delivery technology: Induction of systemic and mucosal immune responses in rats, *Vaccine*, 23, 794–801.
23. Behl, C. R., Pimplaskar, H. K., Sileno, A. P., de Meireles, J., and Romeo, V. D. (1998), Effects of physicochemical properties and other factors on systemic nasal drug delivery, *Adv. Drug Deliv. Rev.*, 29, 89–116.
24. Genta, I., Pavanetto, F., Conti, B., Giunchedi, P., and Conte, U. (1995), Improvement of dexamethasone dissolution rate from spray-dried chitosan microspheres, *S. T. P. Pharm. Sci.*, 5, 202–207.
25. Gavini, E., Hegge, A. B., Rassu, G., Sanna, V., Testa, C., Pirisino, G., Karlsen, J., and Giunchedi, P. (2006), Nasal administration of carbamazepine using chitosan microspheres: In vitro/in vivo studies, *Int. J. Pharm.*, 307, 9–15.
26. Filipović-Grčić, J., Perissutti, B., Moneghini, M., Voinovich, D., Martinac, A., and Jalšenjak, I. (2003), Spray-dried carbamazepine-loaded chitosan and HPMC microspheres: Preparation and characterisation, *J. Pharm. Pharmacol.*, 55, 921–931.
27. Teshima, D., Yamauchi, A., Makino, K., Kataoka, Y., Arita, Y., Nawata, H., and Oishi, R. (2002), Nasal glucagon delivery using microcrystalline cellulose in healthy volunteers, *Int. J. Pharm.*, 233, 61–66.
28. Sacchetti, C., Artusi, M., Santi, P., and Colombo, P. (2002), Caffeine microparticles for nasal administration obtained by spray drying, *Int. J. Pharm.*, 242, 335–339.
29. Kublik, H., and Vidgren, M. T. (1998), Nasal delivery systems and their effect on deposition and absorption, *Adv. Drug Deliv. Rev.*, 29, 157–177.
30. Ugwoke, M. I., Verbeke, N., and Kinget, R. (2001), The biopharmaceutical aspects of nasal mucoadhesive drug delivery, *J. Pharm. Pharmacol.*, 53, 3–22.
31. Pringels, E., Callens, C., Vervaet, C., Dumont, F., Slegers, G., Foreman, P., and Remon, J. P. (2006), Influence of deposition and spray pattern of nasal powders on insulin bioavailability, *Int. J. Pharm.*, 310, 1–7.
32. Mygind, N., and Dahl, R. (1998), Anatomy, physiology and function of the nasal cavities in health and disease, *Adv. Drug Deliv. Rev.*, 29, 3–12.
33. Cornaz, A.-L., De Ascentis, A., Colombo, P., and Buri, P. (1996), In vitro characteristics of nicotine microspheres for transnasal delivery, *Int. J. Pharm.*, 129, 175–183.
34. Ugwoke, M. I., Agu, R. U., Verbeke, N., and Kinget, R. (2005), Nasal mucoadhesive drug delivery: Background, applications, trends and future perspectives, *Adv. Drug Deliv. Rev.*, 57, 1640–1665.
35. Smart, J. D. (2005), The basics and underlying mechanisms of mucosdhesion, *Adv. Drug Deliv. Rev.*, 57, 1556–1568.
36. Mathiowitz, E., and Chickering, D. E. (1999), Definitions, mechanisms and theories of bioadhesion, in Mathiowitz, E., Chickering, D. E., and Lehr, C.-M., Eds., *Bioadhesive*

*Drug Delivery Systems: Fundamentals, Novel Approaches and Development*, Marcel Decker, New York, pp. 1–10.

37. Lee, J. W., Park, J. H., and Robinson, J. R. (2000), Bioadhesive-based dosage forms: The next generation, *J. Pharm. Sci.*, 89, 850–866.
38. Nagai, T., Nishimoto, Y., Nambu, N., Suzuki, Y., and Sekine, K. (1984), Powder dosage form of insulin for nasal administration, *J. Controlled Release*, 1, 15–22.
39. Vidgren, P., Vidgren, M., Arppe, J., Hakuli, T., Laine, E., and Paronen, P. (1992), In vitro evaluation of spray-dried mucoadhesive microspheres for nasal administration, *Drug Dev. Ind. Pharm.*, 18, 581–597.
40. Oechslein, C. R., Fricker, G., and Kissel, T. (1996), Nasal delivery of octreotide: Absorption enhancement by particulate carrier systems, *Int. J. Pharm.*, 139, 25–32.
41. Abd El-Hameed, M. D., and Kellaway, I. W. (1997), Preparation and in vitro characterisation of mucoadhesive polymeric microspheres as intra-nasal delivery systems, *Eur. J. Pharm. Biopharm.*, 44, 53–60.
42. Ugwoke, M. I., Kaufmann, G., Verbeke, N., and Kinget, R. (2000), Intranasal bioavailability of apomorphine from carboxymethylcellulose-based drug delivery systems, *Int. J. Pharm.*, 202, 125–131.
43. García-Arieta, A., Torrado-Santiago, S., Goya, L., and Torrado, J. J. (2001), Spray-dried powders as nasal absorption enhancers of cyanocobalamin, *Biol. Pharm. Bull.*, 24, 1411–1416.
44. Hasçiçek, C., Gönül, N., and Erk, N. (2003), Mucoadhesive microspheres containing gentamicin sulfate for nasal administration: Preparation and in vitro characterisation, *Il Farmaco*, 58, 11–16.
45. Cerchiara, T., Luppi, B., Chidichimo, G., Bigucci, F., and Zecchi, V. (2005), Chitosan and poly(methyl vynil ether-co-maleic anhydride) microparticles as nasal sustained delivery systems, *Eur. J. Pharm. Biopharm.*, 61, 195–200.
46. Callens, C., and Remon, J. P. (2000), Evaluation of starch-maltodextrin-Carbopol 974 P mixtures for the nasal delivery of insulin in rabbits, *J. Controlled Release*, 66, 215–220.
47. Callens, C., Pringels, E., and Remon, J. P. (2003), Influence of multiple nasal administrations of bioadhesive powders on the insulin bioavailability, *Int. J. Pharm.*, 250, 415–422.
48. Ugwoke, M. I., Agu, R. U., Vanbilloen, H., Baetens, J., Augustijns, P., Verbeke, N., Mortelmans, L., Verbruggen, A., Kinget, R., and Bormans, G. (2000), Scintigraphic evaluation in rabbits of nasal drug delivery systems based on carbopol 971p and carboxymethylcellulose, *J. Controlled Release*, 68, 207–214.
49. Suzuki, Y., and Makino, Y. (1999), Mucosal drug delivery using cellulose derivatives as a functional polymer, *J. Controlled Release*, 62, 101–107.
50. Quadir, M., Zia, H., and Needham, T. E. (2000), Development and evaluation of nasal formulations of ketorolac, *Drug Deliv.*, 7, 223–229.
51. Lim, S. T., Martin, G. P., Berry, D. J., and Brown, M. B. (2000), Preparation and evaluation of the in vitro drug release properties and mucoadhesion of novel microspheres of hyaluronic acid and chitosan, *J. Controlled Release*, 66, 281–292.
52. Lim, S. T., Forbes, B., Berry, D. J., Martin, G. P., and Brown, M. B. (2002), In vivo evaluation of novel hyaluronan/chitosan microparticulate delivery systems for the nasal delivery of gentamicin in rabbits, *Int. J. Pharm.*, 231, 73–82.
53. Gavini, E.,Rassu, G., Sanna, V., Cossu, M., and Giunchedi, P. (2005), Mucoadhesive microspheres for nasal administration of an antiemetic drug, metoclopramide: In-vitro/ex-vivo studies, *J. Pharm. Pharmacol.*, 57, 287–294.

54. Witschi, C., and Mrsny, R. J. (1999), In vitro evaluation of microparticles and polymer gels for use as nasal platforms for protein delivery. *Pharm. Res.*, 16, 382–390.
55. Farraj, N. F., Johansen, B. R., Davis, S. S., and Illum, L. (1990), Nasal administration of insulin using bioadhesive microspheres as a delivery system, *J. Controlled Release*, 13, 253–261.
56. Illum, L., Faraj, N., Davis, S., Johansen, B., and O'Hagan, D. (1990), Investigation of the nasal absorption of biosynthetic human growth hormone in sheep: Use of a bioadhesive microsphere delivery system, *Int. J. Pharm.*, 63, 207–211.
57. Illum, L., Fisher, A. N., Jabbal-Gill, I., and Davis, S. S. (2001), Bioadhesive starch microspheres and absorption enhancing agents act synergistically to enhance the nasal absorption of polypeptides, *Int. J. Pharm.*, 222, 109–119.
58. Mao, S., Chen, J., Wei, Z., Liu, H., and Bi, D. (2004), Intranasal administration of melatonin starch microspheres, *Int. J. Pharm.*, 272, 37–43.
59. Illum, L., Farraj, N. F., Critchley, H., and Davis, S. S. (1988), Nasal administration of gentamicin using a novel microsphere delivery system, *Int. J. Pharm.*, 46, 261–265.
60. Björk, E., and Edman, P. (1988), Degradable starch microspheres as a nasal delivery system for insulin, *Int. J. Pharm.*, 47, 233–238.
61. Critchley, H., Davis, S. S., Farraj, N. F., and Illum, L. (1994), Nasal absorption of desmopressin in rats and sheep. Effect of a bioadhesive microsphere delivery system, *J. Pharm. Pharmacol.*, 46, 651–656.
62. Edman, P., Björk, E., and Rydén, L. (1992), Microspheres as a nasal delivery system for peptide drugs, *J. Controlled Release*, 21, 165–172.
63. Pereswetoff-Morath, L., and Edman, P. (1995), Dextran microspheres as a potential nasal drug delivery system for insulin—in vitro and in vivo properties, *Int. J. Pharm.*, 124, 37–44.
64. Pereswetoff-Morath, L., Bjurström, S., Khan, R., Dahlin, M., and Edman, P. (1995), Toxicological aspects of the use of dextran microspheres and thermogelling ethyl(hydroxyethyl) cellulose (EHEC) as nasal drug delivery systems, *Int. J. Pharm.*, 128, 9–21.
65. Rydén, L., and Edman, P. (1992), Effect of polymers and microspheres on the nasal absorption of insulin in rats, *Int. J. Pharm.*, 83, 1–10.
66. Vila, A., Sánchez, A., Évora, C., Soriano, I., McCallion, O., and Alonso, M. J. (2005), PLA-PEG particles as nasal protein carriers: The influence of the particle size, *Int. J. Pharm.*, 292, 43–52.
67. Tafaghodi, M., Tabassi, S. A. S., Jaafari, M. R., Zakavi, S. R., and Momen-nejad, M. (2004), Evaluation of the clearance characteristics of various microspheres in the human nose by gamma-scintigraphy, *Int. J. Pharm.*, 280, 125–135.
68. Leitner, V. M., Guggi, D., Krauland, A. H., and Bernkop-Schnürch, A. (2004), Nasal delivery of human growth hormone: In vitro and in vivo evaluation of a thiomer/glutathione microparticulate delivery system, *J. Controlled Release*, 100, 87–95.
69. Bernkop-Schnürch, A., Guggi, D., and Pinter, Y. (2004), Thiolated chitosans: Development and in vivo evaluation of a mucoadhesive permeation enhancing oral drug delivery system, *J. Controlled Release*, 94, 177–186.
70. Greimel, A., Dorly del Curto, M., D'Antonio, M., Palmberger, T., Sprinzl, G. M., and Bernkop-Schnürch, A. (2006), *In vitro* evaluation of thiomer microparticles for nasal drug delivery, *J. Drug Deliv. Sci. Technol.*, 16, 103–108.
71. Dodane, V., Khan, M. A., and Merwin, J. R. (1999), Effect of chitosan on epithelial permeability and structure, *Int. J. Pharm.*, 182, 21–32.

72. Singla, A. K., and Chawla, M. (2001), Chitosan: Some pharmaceutical and biological aspects—An update, *J. Pharm. Pharmacol.*, 53, 1047–1067.

73. Illum, L., Watts, P., Fisher, A. N., Hinchcliffe, M., Norbury, H., Jabbal-Gill, I., Nankervis, R., and Davis, S. S. (2002), Intranasal delivery of morphine, *J. Pharm. Exp. Ther.*, 301, 391–400.

74. Illum, L. (2003), Nasal drug delivery—Possibilities, problems and solutions, *J. Controlled Release*, 87, 187–198.

75. Illum, L., Farraj, N. F., and Davis, S. S. (1994), Chitosan as a novel nasal delivery system for peptide drugs, *Pharm. Res.*, 8, 1186–1189.

76. Illum, L., Jabbal-Gill, I., Hinchcliffe, M., Fisher, A. N., and Davis, S. S. (2001), Chitosan as a novel nasal delivery system for vaccines, *Adv. Drug Deliv. Rev.*, 51, 81–96.

77. Cheng, Y.-H., Dyer, A. M., Jabbal-Gill, I., Hinchcliffe, M., Nankervis, R., Smith, A., and Watts, P. (2005), Intranasal delivery of recombinant human growth hormone (somatropin) in sheep using chitosan-based powder formulations, *Eur. J. Pharm. Sci.*, 26, 9–15.

78. Thanou, M. M., Verhoef, J. C., Romeijn, S. G., Nagelkerke, J. F., Merkus, F. W. H. M., and Junginger, H. E. (1999), Effects of *N*-trimethyl chitosan chloride, a novel absorption enhancer, on Caco-2 intestinal epithelia and the ciliary beat frequency of chicken embryo trachea, *Int. J. Pharm.*, 185, 73–82.

79. Kotze, A. F., Lueßen, H. L., de Boer, A. G., Verhoef, J. C., and Junginger, H. E. (1998), Chitosan for enhanced intestinal permeability: Prospects for derivatives soluble in neutral and basic environments, *Eur. J. Pharm. Sci.*, 7, 145–151.

80. Bernkop-Schnürch, A., Hornof, M., and Guggi, D. (2004), Thiolated chitosans, *Eur. J. Pharm. Biopharm.*, 57, 9–17.

81. Illum, L., Jorgensen, H., Bisgaard, H., Krogsgaard, O., and Rossing, N. (1987), Bioadhesive microspheres as a potential nasal drug delivery system, *Int. J. Pharm.*, 39, 189–199.

82. Ugwoke, M. I., Agu, R. U., Jorissen, M., Augustijns, P., Sciot, R., Verbeke, N., and Kinget, R. (2000), Nasal toxicological investigations of Carbopol 971P formulation of apomorphine: Effects on ciliary beat frequency of human nasal primary cell culture and in vivo on rabbit nasal mucosa, *Eur. J. Pharm. Sci.*, 9, 387–396.

83. Soane, R. J., Hinchcliffe, M., Davis, S. S., and Illum, L. (2001), Clearence characteristics of chitosan based formulations in the sheep nasal cavity, *Int. J. Pharm.*, 217, 183–191.

84. Jones, N. (2001), The nose and paranasal sinuses physiology and anatomy, *Adv. Drug Deliv. Rev.*, 51, 5–19.

85. Soane, R. J., Frier, M., Perkins, A. C., Jones, N. S., Davis, S. S., and Illum, L. (1999), Evaluation of the clearence characteristics of bioadhesive systems in humans, *Int. J. Pharm.*, 178, 55–65.

86. Vivien, N., Buri, P., Balant, L., and Lacroix, S. (1994), Nasal absorption of metoclopramide administrated to man, *Eur. J. Pharm. Biopharm.*, 40, 228–231.

87. Holmberg, K., Björk, E., Bake, B., and Edman, P. (1994), Influence of degradable starch microspheres on human nasal mucosa, *Rhinology*, 32, 74–77.

88. Brime, B., Ballesteros, M. P., and Frutos, P. (2000), Preparation and in vitro characterization of gelatin microspheres containing Levodopa for nasal administration, *J. Microencapsul.*, 17, 777–784.

89. Preda, M., and Leucuta, E. (2003), Oxprenolol-loaded bioadhesive microspheres: Preparation and *in vitro/in vivo* characterization, *J. Microencapsul.*, 20, 777–789.

90. Kriwet, B., and Kissel, T. (1996), Poly(acrylic acid) microparticles widen intercellular spaces of Caco-2 cell monolayers: An examination by confocal laser scanning microscopy, *Eur. J. Pharm. Biopharm.*, 42, 233–240.

91. Varshosaz, J., Sadrai, H., and Alinagari, R. (2004), Nasal delivery of insulin using chitosan microspheres, *J. Microencapsul.*, 21, 761–774.
92. Jain, S. K., Chourasia, M. K., Jain, A. K., Jain, R. K., and Shrivastava, A. K. (2004), Development and characterization of mucoadhesive microspheres bearing salbutamol for nasal delivery, *Drug Deliv.*, 11, 113–122.
93. Krauland, A. H., Guggi, D., and Bernkop-Schnürch, A. (2006), Thiolated chitosan microparticles: A vehicle for nasal peptide drug delivery, *Int. J. Pharm.*, 307, 270–227.
94. Jiang, H.-L., Park, I.-K., Shin, N.-R., Kang, S.-G., Yoo, H.-S., Kim, S.-I., Suh, S.-B., Akaike, T., and Cho, C.-S. (2004), In vitro study of the immune stimulating activity of an athrophic rhinitis vaccine associated to chitosan microspheres, *Eur. J. Pharm. Biopharm.*, 58, 471–476.
95. Genta, I., Costantini, M., Asti, A., Conti, B., and Montanari, L. (1998), Influence of glutaraldehyde on drug release and mucoadhesive properties of chitosan microspheres, *Carbohydr. Polym.*, 36, 81–88.
96. Hinchcliffe, M., and Illum, L. (1999), Intranasal insulin delivery and therapy, *Adv. Drug Deliv. Rev.*, 35, 199–234.
97. Freiberg, S., and Zhu, X. X. (2004), Polymer microspheres for controlled drug release, *Int. J. Pharm.*, 282, 1–18.
98. Li, Y., Jiang, H. L., Zhu, K. J., Liu, J. H., and Hao, Y. L. (2005), Preparation, characterization and nasal delivery of a-cobrotoxin-loaded poly(lactide-*co*-glycolide)/polyanhydride microspheres, *J. Controlled Release*, 108, 10–20.
99. Ugwoke, M. I., Verbeke, N., and Kinget, R. (1997), Microencapsulation of apomorphine HCI with gelatin, *Int. J. Pharm.*, 148, 23–32.
100. Yang, Y.-Y., Chung, T.-S., Bai, X.-L., and Chan, W.-K. (2000), Effect of preparation conditions on morphology and release profiles of biodegradable polymeric microspheres containing protein fabricated by double-emulsion method, *Chem. Eng. Sci.*, 55, 2223–2236.
101. Conte, U., Giunchedi, P., Maggi, L., and Torre, M. L. (1994), Spray dried albumin microspheres containing nicardipine, *Eur. J. Pharm. Biopharm.*, 40, 203–208.
102. He, P., Davis, S. S., and Illum, L. (1998), In vitro evaluation of the mucoadhesive properties of chitosan microspheres, *Int. J. Pharm.*, 166, 75–68.
103. Martinac, A., Filipovic-Grčic, J., Voinovich, D., Perissutti, B., and Franceschinis, E. (2005), Development and bioadhesive properties of chitosan-ethylcellulose microspheres for nasal delivery, *Int. J. Pharm.*, 291, 69–77.
104. Giunchedi, P., Juliano, C., Gavini, E., Cossu, M., and Sorrenti, M. (2002), Formulation and in vivo evaluation of chlorhexidine buccal tablets prepared using drug-loaded chitosan microspheres, *Eur. J. Pharm. Biopharm.*, 53, 233–239.
105. He, P., Davis, S. S., and Illum, L. (1999), Chitosan microspheres prepared by spray drying, *Int. J. Pharm.*, 187, 53–65.
106. Pavanetto, F., Genta, I., Giunchedi, P., Conti, B., and Conte, U. (1994), Spray dried albumin microspheres for the intra-articular delivery of dexamethazone, *J. Microencapsul.*, 11, 445–454.
107. Martinac, A., Filipovic-Grčic, J., Perissutti, B., Voinovich, D., and Pavelic, Ž. (2005), Spray-dried chitosan/ethylcellulose microspheres for nasal drug delivery: Swelling study and evaluation of *in vitro* drug release properties, *J. Microencapsul.*, 22, 549–561.
108. Fundueanu, G., Constantin, M., Dalpiaz, A., Bortolotti, F., Cortesi, R., Ascenzi, P., and Menegatti, E. (2004), Preparation and characterization of starch/cyclodextrin bioadhesive microspheres as platform for nasal administration of Gabexate Mesylate (Foy(r)) in allergic rhinitis treatment, *Biomaterials*, 25, 159–170.

109. Berthold, A., Cremer, K., and Kreuter, J. (1996), Preparation and characterization of chitosan microspheres as drug carrier for prednisolone sodium phosphate as model for anti-inflammatory drugs, *J. Controlled Release*, 39, 17–25.

110. Jaganathan, K. S., and Vyas, S. P. (2006), Strong systemic and mucosal immune responses to surface-modified PLGA microspheres containing recombinant Hepatitis B antigen administrated intranasally, *Vaccine*, 24, 4201–4211.

111. Callens, C., Ceulemans, J., Ludwig, A., Foreman, P., and Remon, J. P. (2003), Rheological study on mucoadhesivity of some nasal powder formulations. *Eur. J. Pharm. Biopharm.*, 55, 323–328.

112. Santos, C. A., Jacob, J. S., Hertzog, B. A., Freedman, B. D., Press, D. L., Harnpicharnchai, P., and Mathiowitz, E. (1999), Correlations of two bioadhesion assays: The everted sac technique and the CAHN microbalance, *J. Controlled Release*, 61, 113–122.

113. Rango Rao, K. V., and Buri, P. (1989), A novel in situ method to test polymers and coated microparticles for bioadhesion, *Int. J. Pharm.*, 52, 265–270.

114. Agu, R. U., Jorissen, M., Kinget, R., Verbeke, N., and Augustijns, P. (2002), Alternatives to *in vivo* nasal toxicological screening for nasally-administrated drugs, *S. T. P. Pharm. Sci.*, 12, 13–22.

115. Hvidberg, A., Djurup, R., and Hilsted, J. (1994), Glucose recovery after intranasal glucagon during hypoglycaemia in man, *Eur. J. Clin. Pharmacol.*, 46, 15–17.

116. Ugwoke, M. I., Agu, R. U., Jorissen, M., Augustijns, P., Sciot, R., Verbeke, N., and Kinget, R. (2000), Toxicological investigations of the effects carboxymethylcellulose on ciliary beat frequency of human nasal epithelial cells in primary suspension culture and in vivo on rabbit nasal mucosa, *Int. J. Pharm.*, 205, 43–51.

117. Merkus, P., Romeijn, S. G., Verhoef, J. C., Merkus, F. W., and Schouwenburg, P. F. (2001), Classification of cilio-inhibiting effects of nasal drugs, *Laryngoscope*, 111, 595–602.

118. Callens, C., Adriaens, E., Dierckens, K., and Remon, J. P. (2001), Toxicological evaluation of a bioadhesive nasal powder containing a starch and Carbopol 974 P on rabbit nasal mucosa and slug mucosa, *J. Controlled Release*, 76, 81–91.

119. Hirai, S., Yashiki, T., Matsuzawa, T., and Mima, H. (1981), Absorption of drugs from the nasal mucosa of rat, *Int. J. Pharm.*, 7, 317–325.

120. Adriaens, E., and Remon, J. P. (1999), Gastropods as an evaluation tool for screening the irritating potency of absorption enhancers and drugs, *Pharm. Res.*, 16, 1239–1243.

121. Marttin, E., Schipper, N. G. M., Verhoef, J. C., and Merkus, F. W. H. M. (1998), Nasal mucociliary clearance as a factor in nasal drug delivery, *Adv. Drug Deliv. Rev.*, 29, 13–38.

122. Agu, R. U., Jorissen, M., Willems, T., Van Den Mooter, G., Kinget, R., and Augustijns, P. (1999), Effects of pharmaceutical compounds on ciliary beating in human nasal epithelial cells: A comparative study of cell culture models, *Pharm. Res.*, 16, 1380–1385.

123. Merkus, F. W. H. M., Schipper, N. G. M., Hermens, W. A. J. J., Romeijn, S. G., and Verhoef, J. C. (1993), Absorption enhancers in nasal drug delivery: Efficacy and safety, *J. Controlled Release*, 24, 201–208.

124. Dimova, S., Brewster, M. E., Noppe, M., Jorissen, M., and Augustijns, P. (2005), The use of human nasal in vitro cell systems during drug discovery and development, *Toxicol. in Vitro*, 19, 107–122.

125. Schmidt, M. C., Peter, H., Lang, S. R., Ditzinger, G., and Merkle, H. P. (1998), In vitro cell models to study nasal mucosal permeability and metabolism, *Adv. Drug Deliv. Rev.*, 29, 51–79.

126. Jorissen, M., Van der Schueren, B., Tyberghein, J., Van Der Berghe, H., and Cassiman, J. J. (1989), Ciliogenesis and coordinated ciliary beating in human nasal epithelial cells cultured in vitro, *Acta Oto-Rhino-Laryngol. Begica*, 43, 67–73.
127. Agu, R. U., Jorissen, M., Willems, T., Van Den Mooter, G., Kinget, R., Verbeke, N., and Augustijns, P. (2000), Safety assesment of selected cyclodextrins: Effect on ciliary activity using a human cell suspension model exhibiting in vitro ciliogenesis, *Int. J. Pharm.*, 193, 219–226.
128. Illum, L., Farraj, N. F., Fisher, A. N., Gill, I., Miglietta, M., and Benedetti, L. M. (1994), Hyaluronic acid ester microspheres as a nasal delivery system for insulin, *J. Controlled Release*, 29, 133–141.
129. Sankar, C., Rani, M., Srivastava, A. K., and Mishra, B. (2001), Chitosan based pentazocine microspheres for intranasal systemic delivery: Development and biopharmaceutical evaluation, *Pharmazie*, 56, 223–226.
130. Ugwoke, M. I., Sam, E., Van Den Mooter, G., Verbeke, N., and Kinget, R. (1999), Nasal mucoadhesive delivery systems of the anti-parkinsonian drug, apomorphine: Influence of drug-loading on in vitro and in vivo release in rabbits, *Int. J. Pharm.*, 181, 125–138.
131. El-Shafy, M. A., Kellaway, I. W., Taylor, G., and Dickinson, P. A. (2000), Improved nasal bioavailability of FITC-dextran (Mw 4300) from mucoadhesive microspheres in rabbits, *J. Drug Target.*, 7, 355–361.
132. Rajinikanth, P. S., Sankar, C., and Mishra, B. (2003), Sodium alginate microspheres of metoprolol tartarate for intranasal systemic delivery: Development and evaluation, *Drug Deliv.*, 10, 21–28.
133. Nakamura, K., Maitani, Y., Lowman, A. M., Takayama, K., Peppas, N. A., and Nagai, T. (1999), Uptake and release of budesonide from mucoadhesive, pH-sensitive copolymers and their application to nasal delivery, *J. Controlled Release*, 61, 329–335.
134. Yildiz, A., Okyar, A., Baktir, G., Araman, A., and Özsoy, Y. (2005), Nasal administration of heparin-loaded microspheres based on poly(lactic acid), *Il Farmaco*, 60, 919–924.

# AEROSOL DRUG DELIVERY

Michael Hindle
*Virginia Commonwealth University, Richmond, Virginia*

**Contents**

1 Introduction
2 Human Respiratory Tract and Aerosol Particle Deposition
  2.1 Human Respiratory Tract
  2.2 Mechanisms of Particle Deposition
  2.3 Pharmacokinetics
3 Therapeutic Indications for Aerosol Delivery
  3.1 Current Applications
  3.2 Future Applications
4 Aerosol Drug Delivery Devices
  4.1 Introduction
  4.2 Characteristics of Ideal Delivery Device
5 Metered Dose Inhalers
  5.1 Introduction
  5.2 Metered Dose Inhaler and HFA Reformulation
  5.3 Propellants
  5.4 Excipients
  5.5 Valves
  5.6 Actuators
  5.7 Canisters
  5.8 Breath Actuation
  5.9 Spacers
  5.10 Dose Counters
6 Dry Powder Inhalers
  6.1 Introduction
  6.2 Size Reduction and Particle Formation Technologies
  6.3 Drug–Lactose Formulations
  6.4 Dry Powder Inhaler Design
  6.5 Exubera
7 Nebulizers

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

8 Emerging Technologies
8.1 Soft Mist Aerosols
8.2 Respimat
8.3 AERx
8.4 Mystic
8.5 Capillary Aerosol Generator
8.6 Staccato
9 Conclusions
References

## 1 INTRODUCTION

Aerosol drug delivery to the lungs has long been the route of choice for the treatment of respiratory diseases, including asthma and chronic obstructive airway disease. Metered dose inhalers (MDIs), dry powder inhalers (DPIs), and nebulizers have been employed to successfully deliver a wide range of pharmaceuticals principally to the lungs for local action. However, with their unique characteristics, the lungs have now begun to be targeted as a means of noninvasive delivery of systemically acting compounds, including genes, proteins, peptides, antibiotics, and other small molecules [1, 2]. The primary function of the respiratory tract is gaseous exchange, transferring oxygen from the inspired air to the blood and removing carbon dioxide from the circulation. This pulmonary circulation offers rapid absorption and systemic distribution of suitable drugs deposited in the airways. Due to its anatomical structure, however, an important secondary role is the protection of the body from inhalation of foreign particles (including aerosol drug particles). The challenge of aerosol drug delivery is to overcome this highly effective barrier and accurately and reproducibly deliver aerosol drug particles in sufficient doses to their targeted sites within the lungs for either local action or systemic absorption. Effective aerosol drug delivery is tied to the aerosol inhaler that generates and delivers the respirable aerosol. This chapter will primarily focus on aerosol drug delivery devices, their development, and future prospects for pulmonary administration.

## 2 HUMAN RESPIRATORY TRACT AND AEROSOL PARTICLE DEPOSITION

### 2.1 Human Respiratory Tract

The human respiratory tract can be divided into three main regions: first, the upper airways, including the nose, mouth and throat (oropharnyx), and the larynx [3]. The conducting airways consist of the regions from the trachea to the respiratory bronchioles and have airway diameters between 0.6 and 20 mm. The alveolar region consists of respiratory bronchioles and alveolar sacs and has airway diameters between 0.2 and 0.6 mm. The lungs are a branching system which commences asymmetrically, dividing first at the base of the trachea. The left and right bronchi branch

dichotomously into the conducting airways. There are approximately 23 generations before the respiratory bronchioles give way to the alveoli, the site of gaseous exchange [4]. This branching produces a progressive reduction in airway diameter and also significantly increases the total surface area of the lower airways [3]. Another important characteristic with respect to drug delivery is the extensive vascular circulation. The blood vessels supplying the conducting airways are part of the systemic circulation. In contrast, the alveolar region is connected to the pulmonary circulatory pathway; drugs absorbed into this circulation will avoid first-pass hepatic metabolism effects.

### 2.2 Mechanisms of Particle Deposition

Aerosol particles are deposited in the lungs by three main mechanisms, and the site of deposition is dependent upon the predominating mechanism. Inertial impaction occurs because a particle traveling in an air stream has its own momentum (the product of its mass and velocity). As the direction of the airflow changes due to a bend or obstacle, the particle will continue in its original direction for a certain distance because of its inertia. Particles with a high momentum, due to high velocity or large size, are often unable to change direction before they impact on the surface in front of them [5]. Impaction of particles entering the mouth with a high velocity occurs either at the back of the mouth or at the bend where the pharynx leads to the trachea. Only a small fraction of particles greater than 15 μm will reach the trachea following mouth breathing. The majority, due to their size, will impact in the oropharyngeal region. Deposition by impaction will also occur as the trachea splits into the left and right bronchus. As the velocity of the particles decreases, inertial impaction becomes a less important mechanism of deposition in the smaller airways. Following the removal of larger particles in the upper airways by inertial impaction, gravitational sedimentation is the mechanism by which smaller particles (2–5 μm) are deposited in the respiratory bronchioles and alveoli. These particles settle under gravity and accelerate to a steady terminal velocity when the gravitational force is balanced by the resistance of the air through which it is traveling [6]. It is a time-dependent process which is aided by breath holding [7]. Brownian motion or diffusion is a mechanism which significantly affects only particles less than 0.5 μm in diameter. These particles are subjected to bombardment by surrounding gas molecules causing random movement of the particles. In this situation, the diffusivity of a particle is inversely proportional to its diameter. For an extensive mechanistic review of the area of particle deposition readers should consult Finlay (2001) [8].

Aerosol particle size and polydispersity are major determinants of the site and mechanism of pulmonary deposition. Fundamental deposition studies using monodisperse aerosols together with mathematical models have established the optimum aerosol particle size for lung deposition [9–12]. Aerosols larger than 10 μm will deposit in the oropharyngeal region and will not be inhaled. Particles less than 3 μm will be capable of penetrating into the alveolar region. Aerosols in the size range 3–10 μm will be distributed in the central and conducting airways [13]. A polydisperse aerosol containing a range of these particle sizes will allow deposition throughout the lungs. In theory, lung site deposition targeting should be possible by controlling the particle size of the inhalation aerosol [14]. However, a number of

other significant variables can affect deposition within the respiratory tract and these often confound any efforts at targeting [15]. The patient's respiratory cycle, both the rate and depth of breathing, will affect aerosol deposition, and this is also the source of large intersubject variability in deposition [16]. Slow and deep inhalations are required for deposition in the peripheral airways, and this is the technique often recommended for inhalation with the MDI [17]. A different technique may be required for DPIs, where the patient's inspiratory effort is often the powder dispersion and delivery force. Flow rates greater than 60 L/min are commonly employed for powder inhalers [18, 19]. A final respiratory maneuver can be employed to promote deposition; breath holding up to 10 s is generally recommended to enhance deposition by sedimentation [17]. Other parameters that will affect lung deposition are the disease state within the lungs and its effect on airway caliber together with the patient's age and airway morphology [20–27].

### 2.3 Pharmacokinetics

Once deposited on the surface of the airways, the particle is subject to absorption and clearance processes depending upon its physical properties and the site of deposition [28–30]. For example, a lipophilic small molecule deposited in the central airways would have a different pharmacokinetic profile than a 50-kDa macromolecule deposited in the alveolar region. The former may undergo mucociliary clearance following deposition on a ciliated epithelial cell. Following dissolution, lipophilic drugs may be transported across the epithelium by passive transcytosis, while hydrophilic compounds are taken up by other pathways such as via tight junctions and endocytosis. Having overcome the barrier of the epithelial layer, the drug is available for distribution into the systemic circulation or to its site of action. Finally, the drug may also be subject to metabolism within the airways. For the macromolecule deposited in the peripheral airways, the absorption rate has been shown to be dependent upon molecular size. Larger molecules are subject to active processes such as caveolae or vesicular transport across the cell. Diffusion remains the predominant mechanism for smaller lipophilic macromolecules. Insoluble molecules can be phagocytosed by alveolar macrophages and removed via the lymphatic system or the mucociliary escalator. The pharmacokinetics of inhaled drugs is complicated by the fact that a significant fraction of the delivered dose is deposited in the oropharnyx or removed from the lungs via mucociliary clearance and in both cases subsequently swallowed [31]. An often desired goal for a pulmonary formulation is prolonged action within the lung. Rapid clearance or metabolism results in short duration of action for most inhaled drugs. A number of approaches using formulation excipient additives have been investigated to increase the residency or prolong release of drug at its site of action within the lungs [32, 33]. Microspheres containing nanoparticles have been formulated as dry powders for inhalation offering sustained-release properties [34]. In addition, prodrugs which are activated locally within the lungs have been used in an alternative approach [35–37].

The pharmacokinetic process of absorption, distribution, metabolism, and excretion within the lungs is an enormous subject area and readers are referred to specific reviews for further details [38–43]. Of particular interest may be the subject of absorption enhancer methodologies for lung delivery, which is beyond the scope of this chapter [44].

## 3 THERAPEUTIC INDICATIONS FOR AEROSOL DELIVERY

### 3.1 Current Applications

Aerosolized drug delivery is currently used to deliver a limited range of therapeutic classes of compounds. These are mainly for asthma and chronic obstructive airway disease. These classes of compounds include short- and long-acting β-adrenoceptor agonist, corticosteroids, mast cell stabilizers, and muscarinic antagonists. Of recent note is the popularity of combination products. These have obvious advantages from a patient compliance perspective. In addition, certain combinations of drugs have shown synergistic therapeutics benefits when compared to the drugs given by separate inhalers [45]. Long-acting β-adrenoceptor agonists and corticosteroids formulated as combination products are available as both MDIs and DPIs [46]. Also recently introduced was a MDI formulation, the R enantiomer of albuterol, which is believed to be mainly responsible for bronchodilation in the racemic mixture [47]. Zanamivir is licensed in the United States as an inhaled antiviral agent for the treatment of influenza [48]. Recombinant human deoxyribonuclease (rhDNAase) is available as a nebulizer product for the treatment of cystic fibrosis, in which it acts to liquefy viscous lung secretions [49]. And recently, insulin was approved as an inhaled powder for glycemic control in type I and II diabetes (see Section 6.5) [50].

### 3.2 Future Applications

Research and development are presently underway covering a vast array of novel applications. Clark (2004) provides an extensive list of products and their current state of development [51]. A significant future advance will be the development of inexpensive, noninvasive, stable, single-dose vaccine delivery via the lungs [52]. Efforts in this area are being led by the World Health Organization in the Measles Aerosol Project, and in a separate project, the Grand Challenges in Global Health initiative has funded a program to further develop an inhalation aerosol measles vaccine. Delivery of the measles vaccine via the lungs has been demonstrated to be both safe and effective [53–58]. Now the challenge of each of these projects is to produce stable inhalation vaccine formulations to be delivered via inexpensive inhalers while maintaining both safety and efficacy [59]. The use of inhaled vaccinations in the event of a bioterrorism attack is also a potential application [60, 61].

The use of the inhalation route for the delivery of gene therapy is also an area of significant interest [60, 62]. Cationic liposomes and polymers together with adenoviral vectors containing the reporting genes have been aerosolized using nebulizers for the majority of clinical studies. However, there are a significant number of challenges that must be overcome before pulmonary gene delivery is deemed completely successful, the most important being low gene transfer efficiency at the cellular level. This problem is not unique to inhalation therapy. Inhalation of a recombinant adenovirus containing the cystic fibrosis transmembrane regulator (Ad2/CFTR) demonstrated the feasibility of this approach for the treatment of cystic fibrosis [63, 64]. However, the limited duration of transfection and low cellular uptake efficiency still remain a barrier to full utilization of this route [60, 65]. There

are a number of reviews that provide updates as to recent developments in this area [60, 66–70].

Given the success of delivering insulin, other peptides and proteins are being considered for pulmonary applications [71, 72]. Leuprolide is a nonapeptide which has been investigated as both an MDI and DPI formulation for the treatment of prostrate cancer [73–76]. Other hormones being investigated include calcitonin for the treatment of Paget disease and osteoporosis, parathyroid hormone to treat osteoporosis, growth hormone releasing factor for the treatment of pituitary dwarfism, and vasoactive intestinal peptide (VIP) for the treatment of pulmonary diseases [60, 77–81].

Other potential inhalation applications include drugs for both local and systemic delivery. Inhaled tobramycin is being investigated for the treatment of *Pseudomonas aeruginosa* exacerbations in cystic fibrosis [82, 83]. Liposomal ciprofloxacin is being developed as a first-line defense against biowarfare agents (e.g., anthrax) [61]. Inhaled cyclosporine has been shown to improve survival rates and extend periods of chronic rejection-free survival in lung transplant patients [84]. Apomorphine has been proposed as an inhalation formulation for the treatment of erectile dysfunction [85]. Aerosol delivery of chemotherapeutic drugs has been advocated for the treatment of lung cancer [86]. Morphine and fentanyl have been investigated for alternative routes of administering analgesics [87–90]. Heparin and low-molecular-weight heparins have been aerosolized and advocated for the treatment of emphysema and thrombosis [91–93]. Iloprost, a stable prostacyclin analog, has been aerosolized by nebulization for use in the treatment of pulmonary hypertension [94]. This list of potential new treatments approached via the inhalation route is not exhaustive; among the other compounds under investigation are $\alpha_1$-antitrypsin, sumatriptan, ergotamine, nicotine as replacement therapy, pentamidine, and ribavirin. Readers should be aware that a large number of these examples are proof-of-concept studies that may not get beyond in vitro experiments and animal studies.

## 4 AEROSOL DRUG DELIVERY DEVICES

### 4.1 Introduction

As can be seen from the previous section, aerosol drug delivery continues to be an area of intensive research and development for the pharmaceutical industry. Not only are new applications for the pulmonary route being investigated, but also new delivery technologies are under development. The reformulation of MDIs with hydrofluoroalkane (HFA) propellants together with the potential of using the inhalation route as a means of systemic administration has led to significant technological advances in delivery devices. In parallel to MDI research DPIs have been developed from breath-actuated single-dose devices to both multiple-dose inhalers and active-dispersion DPIs. There is an extensive literature detailing the fundamental mechanisms of powder dispersion aimed at improving pulmonary deposition from powder inhalers. In addition, novel particle production technologies have been developed that provide alternatives to the traditional micronized powder for formulation in both MDIs and DPIs. Nebulizer technology has evolved from previously

nonportable devices into high-efficiency, hand-held nebulizers that offer alternatives to the MDI and DPI for certain treatment regimes. Finally, novel soft mist inhalers that generate aerosols by solution atomization have emerged on the inhaler landscape. All this research has focused on improving aerosol deposition efficiency and reproducibility within the lungs, together with targeting the peripheral lungs for systemic absorption. The efforts of the last decade culminated in two significant events. First, the regulatory approval of Proventil HFA and QVAR, the first suspension and solution HFA MDIs, respectively. Second, in 2006, the U.S. and European regulatory authorities approved Exubera, an insulin DPI for the systemic treatment of type I and II diabetes. Exubera offered a noninvasive alternative to subcutaneous injections of insulin.

### 4.2 Characteristics of Ideal Delivery Device

With these developments, innovation continues toward development of the ideal inhaler. A number of authors have compiled lists of desired characteristics for an aerosol inhaler [95–97]. These can be grouped into patient-desired or industry-driven properties. From the patient's perspective the overriding requirement is a device that is simple to operate. This is becoming increasing difficult to achieve as evidenced by the intensive patient education initiative that is being planned for the launch of the Exubera insulin inhaler. Poor compliance and adherence to prescribed therapy may be related to patients' failure to use the inhaler correctly [98]. Inhalers should be portable and contain a large number of doses. The device should also give some indication to the patient when it is empty. The inhaler should be suitable for use by all of the population, especially children and the elderly. Ganderton (1999) cited that from a device perspective aerosol generation should be independent of the patient's inhalation and should continue for a substantial portion of the inspiratory cycle. This would minimize the reliance on coordinating inhalation and actuation of the device [99]. Breath-actuated devices have been developed to address this issue. In order to achieve lung deposition targeting, the particle size distribution of the aerosol should be capable of being altered depending upon the specific target region. For example, the central airways may be targeted with a 3–5 μm aerosol for the treatment of acute bronchoconstriction, while a smaller aerosol (1–3 μm) might be used for deep lung deposition and subsequent systemic absorption [99]. In addition, the dose should be delivered reproducibly with minimal oropharyngeal deposition, perhaps as a low-velocity aerosol. There should be a minimal number of small parts in the inhaler, and it should be robust and reliable when placed "in use." The manufacturer has the option of producing a disposable or refillable unit; however, the inhaler should protect the formulation from environment and not affect its stability.

Dolovich et al. (2005) have provided an extensive evidence-based evaluation of aerosol drug devices. They concluded that when selecting an inhalation delivery system the following should be considered: device and drug combination availability, clinical setting, patient age, the ability of the patient to use the device correctly, device use with multiple medications, cost and reimbursement, drug administration time, convenience in outpatient and inpatient settings, and patient and physician preference [100]. Other reviews have compared the benefits and disadvantages of inhalers from clinical and patients' perspectives [101, 102].

## 5 METERED DOSE INHALERS

### 5.1 Introduction

Since their development, MDIs have been widely used for pulmonary aerosol drug delivery [103]. Despite their recognized limitations, they remain the device of choice for many physicians around the globe. From a patient's perspective, they are light, portable, and robust and contain multiple doses of medication. They are also relatively simple to operate (press and fire); however, significant numbers of patients experience difficulties correctly using the MDI due to coordination problems [104]. To maximize lung drug deposition, actuation (pressing the MDI canister) by the patient must be coordinated with a slow, deep inhalation. Studies have reported that 51% of patients fail to operate the MDI correctly [104]. This leads to low lung deposition, high oropharyngeal deposition, and ultimately perhaps therapeutic failure. From the pharmaceutical industry perspective, the components are relatively inexpensive; however, the formulation and manufacturing have become increasingly complex. There are numerous studies describing the multifaceted and interactive effects of propellant [105–110], excipient [111–115], metering valve [110, 116], and actuator [116–119] on the aerosol particle size characteristics of the MDI [120, 121]. To date, the success of the MDI has relied in part on the potency and relative safety of the bronchodilators and corticosteroids commonly used for the treatment of respiratory disorders rather than its delivery efficiency. The relatively low and often variable aerosol deposition efficiency, only around 10–20% of the nominal dose being delivered to the lungs, is the challenge that is beginning to be addressed as the MDI looks to enter the next 50 years of aerosol drug delivery.

### 5.2 Metered Dose Inhaler and HFA Reformulation

The basic design and operation of the MDI has changed little over its lifetime. Aerosols are generated from a formulation of drug (0.1–1% w/w) either suspended or in solution in the liquefied propellant. The formulation is held under pressure in a canister.

Figure 1 shows the basic components of the MDI, consisting of a canister sealed with a metering valve which is inserted into a plastic actuator containing the mouthpiece. Aerosol generation takes place when the canister is pressed against the actuation sump by the patient. Actuation causes the outlet valve to open and the liquefied propellant formulation is released through the actuator nozzle and subsequently through the mouthpiece to the patient. Metered volumes between 20 and 100 μL are dispensed, and as the pressurized propellant is released, it forms small liquid droplets traveling at high velocity. These droplets evaporate to leave drug particles for inhalation [117]. Purewal and Grant (1998) have assembled a definitive reference source for issues relating to the design, manufacturing, and performance of MDIs [122].

The currently marketed MDIs may look similar to the devices that were first developed by Riker in 1950. However, due to the replacement of the ozone-depleting chlorofluorocarbon (CFC) propellants with HFA propellants, virtually all of the components of the MDI have been altered. In 1987, the Montreal Protocol was drawn up, leading to the eventual phase-out of CFC propellants. MDIs contain-

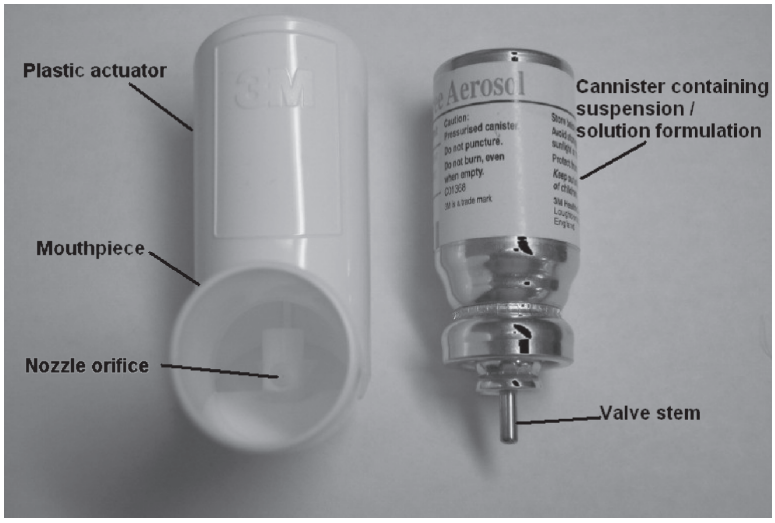


**FIGURE 1** Schematic of MDI.

ing CFC propellants were granted essential-use exemptions until viable alternatives became available. Therefore, with this impending withdrawal, a consortium of pharmaceutical companies (IPACT-I and IPACT-II) worked to identify and toxicologically test alternative propellants for MDIs. HFA 134a and HFA 227 were identified as viable alternatives and the task of reformulation began. At first look, it appeared that the most expeditious route to replacing a CFC product would be to produce a suspension HFA MDI with exactly the same in vitro characteristics as the CFC MDI. This would prove to be a time-consuming route [123, 124]. While some manufacturers focused on producing HFA products with identical characteristics to the current CFC versions to accelerate the pathway through clinical testing to market. Others undertook extensive research and development in the area of HFA formulation options, and this has led to the possibility of utilizing the MDI for both local and systemic administration. During this reformulation effort, the industry has taken the opportunity to address some of the other shortcomings of the MDI [125]. Among these issues were poor peripheral lung delivery, variable dose delivery, and limitations as to the dose capable of being delivered to the lung (typically about 1 mg) [126].

The replacement of CFC MDIs with inhalers formulated with the HFA propellants is now well underway in Europe. Although progress in the United States has been slower, with the introduction of suitable alternatives for albuterol inhalers, the FDA has ordered that CFC albuterol MDIs be withdrawn from the market by the end of 2008 [127]. Examples of reformulated products available in the United States include Ventolin HFA, which is a suspension albuterol sulfate formulation using HFA 134a alone. ProAir is an alternative albuterol sulfate product manufactured by Ivax which contains ethanol and HFA 134a. Xopenex HFA has recently been approved for marketing in the United States [128]. This product contains levalbuterol tartrate (R-albuterol enantiomer) together with HFA 134a, dehydrated alcohol, and oleic acid as a suspension formulation. Table 1 summarizes the HFA

**TABLE 1 Summary of HFA Metered Dose Inhaler Products Available in United States, June 2006**

| Product Name | Drug | Approval Date | Excipients | Type |
|---|---|---|---|---|
| Proventil HFA (3M) | Albuterol sulfate | August 1996 | HFA 134a, ethanol, oleic acid | Suspension |
| Ventolin HFA (GSK) | Albuterol sulfate | April 2001 | HFA 134a | Suspension |
| Proair HFA (IVAX) | Albuterol sulfate | October 2004 | HFA 134a, ethanol | Suspension |
| QVAR (3M) | Beclomethasone dipropionate | September 2000 | HFA 134a, ethanol, oleic acid | Solution |
| Flovent HFA (GSK) | Fluticasone propionate | May 2004 | HFA 134a, ethanol | Suspension |
| Atrovent HFA (BI) | Ipratropium bromide | November 2004 | HFA 134a, purified water, dehydrated alcohol, anhydrous citric acid | Solution |
| Xopenex (Sepracor) | Levalbuterol tartrate | March 2005 | HFA 134a, dehydrated alcohol, oleic acid | Suspension |

products currently available in the United States and their excipients. The following section will focus on the current options for formulation of drugs in HFA propellant systems and the challenges that are encountered as products are reformulated as HFA formulations.

### 5.3 Propellants

The CFC propellants primarily used in MDI formulations were CFC 11, 12, and 114. Blends of these propellants were held liquefied under pressures of 50–80 psig within the canister. Flocculated drug suspensions in CFC propellants were formulated using a surfactant (e.g., oleic acid and lecithin). In a suspension formulation, the aerosol particle size is dependent upon the initial micronized drug particle size (typically between 2 and 5 μm) and the evaporation of the propellant droplets. It has long been recognized that changes in CFC propellant vapor pressure result in changes in droplet size and velocity of the aerosol. Newman et al. (1982) showed that increasing the vapor pressure of the propellant blend in the MDI significantly increased whole-lung deposition and reduced oropharyngeal deposition [110].

The first challenge encountered during the reformulation with HFA propellants was the altered physicochemical properties of HFA 134a and HFA 227 compared to the CFC propellants [124]. Table 2 compares the physicochemical properties of the CFC and HFA propellants [124]. The increased polarity of HFA 134a and HFA 227 is illustrated by the increased dipole moments and dielectric constant. From a practical point of view, the altered solvency properties of the HFA propellant for the drug, excipient, water, and, surprisingly, components of the MDI have been the

**TABLE 2 Comparison of Physicochemical Characteristics of CFC and HFA Propellants**

| | CFC | | | HFA | |
|---|---|---|---|---|---|
| Property | 11 | 12 | 114 | 134a | 227 |
| **Thermodynamic** | | | | | |
| RRM | 137 | 121 | 171 | 102 | 170 |
| Boiling point, °C | 24 | –30 | 4 | –26 | –16 |
| Vapor pressure, 20°C, kPa | 89 | 566 | 182 | 572 | 390 |
| Enthalpy vap., 20°C, kJ/mol (J/g) | 25.1 (183) | 17.2 (142) | 22.1 (130) | 18.6 (182) | 19.6 (115) |
| $C_p$ liquid, 20°C, J/mol·K (J/g·K) | 120 (0.88) | 118 (0.98) | 168 (0.98) | 143 (1.41) | 210 (1.24) |
| **Polarity** | | | | | |
| Dielectric constant | 2.3 | 2.1 | 2.2 | 9.5 | 4.1 |
| Dipole moment (*D*) | 0.45 | 0.51 | 0.58 | 2.1 | 1.2 |
| Induced polarization | 2.8 | 2.3 | 3.2 | 6.1 | 6.1 |
| Solubility parameter (Hild. units) | 7.5 | 6.1 | 6.4 | 6.6 | 6.2 |
| Kauri-butanol value | 60 | 18 | 12 | 9 | 13 |
| log *P* octanol/water | 2.0 | 2.2 | 2.8 | 1.1 | 2.1 |
| Water solubility (ppm) | 130[a] | 120[a] | 110[a] | 2200[b] | 610[b] |
| **Liquid Phase** | | | | | |
| Density (g/cm$^3$) | 1.49 | 1.33 | 1.47 | 1.23 | 1.42 |
| Viscosity (mPa·s) | 0.43 | 0.20 | 0.30 | 0.21 | 0.27 |
| Surface tension (mN/m) | 18 | 9 | 11 | 8 | 7 |

*Source:* From ref. 124. [a]30° [b]25°.

major issues during reformulation. Suspension formulations of micronized drug in the liquefied CFC propellant blends with surfactants were replaced with either suspension or solution HFA formulations, depending upon the solubility of the individual drug in the HFA propellants. For suspension HFA formulations, however, it was observed that the conventional surfactants used in the CFC products were insoluble in the HFA propellants without the addition of a cosolvent (e.g., ethanol) [129, 130]. An alternative approach to produce suspension formulations was to develop a new class of surfactants suitable for use in the HFA systems [131]. It was also observed that some of the commonly used inhalation drugs were slightly soluble in the new propellants and therefore precluded their formulation as a suspension. The potential to formulate as a solution offered a number of advantages together with significant problems. Perhaps most importantly, changing from a suspension to a solution formulation altered the mechanism of aerosol particle formation. In the case of solution formulations, drug is dissolved in the liquefied propellant and a suitable cosolvent (if necessary) and particle formation takes place during evaporation of the propellant. This leads to much smaller particles being formed when propellant evaporation is complete. Stein and Myrdal (2006) recently described the MDI aerosol generation for solution formulations as a two-step process [132]. Droplet formation takes place as millions of atomized droplets are produced after the formulation exits the metering valve through the actuator sump. Initial droplet size is dependent upon the vapor pressure and surface tension of the formulation,

valve size, and actuator orifice diameter. The second step is an evaporative or "aerosol maturation" phase, as the propellant and cosolvents (e.g., ethanol) rapidly evaporate leaving inhalable drug particles. The final size of these particles is dependent upon the initial droplet size, the vapor pressure of the formulation mixture, and the proportion of nonvolatiles in the formulation [132].

Leach et al. (1998) compared the pulmonary deposition of a suspension CFC formulation of beclomethasone diproprionate with a solution HFA formulation [133]. The marketed CFC product had a mass median aerodynamic diameter (MMAD) of 3.5 μm compared to 1.1 μm for the solution formulation, reflecting the altered aerosol formation mechanism. The gamma scintigraphy profile for the solution formulation showed the drug and label to be diffusely deposited throughout the airways with approximately 55–60% deposited in the lungs. In contrast, the CFC product was deposited mainly in the mouth and throat (90–94%), with only 4–6% being deposited in the airways. In many ways, this study summarized the deficiencies of the suspension formulation CFC MDI and offered the alternative of improved delivery efficiencies with the solution HFA formulation. A significant conclusion from this and other studies supported the hypothesis that improved pulmonary deposition and reduced oropharyngeal losses of aerosols would allow reduction in the dose required by the patient to achieve the same therapeutic effect [108, 133, 134].

The altered solubility profile of the HFA propellant, while providing attractive characteristics for solution formulations, also provide significant challenges with respect to their interactions with the basic MDI components. Leachables are compounds that can be transferred from MDI component parts to the formulation during the shelf life of the product. Berry et al. (2003) postulated that MDI orientation could affect the amount of leachables that entered a formulation and affect the particle size distribution of aerosol [135]. Extensive efforts are now required for extractable and leachable testing of the component materials prior to formulation of an MDI. Another by-product of replacing the CFC propellants was to tighten the impurity specifications required for the new propellants. A proposed U.S. Pharmacopeia (USP) monograph for HFA 134a has now been published detailing the impurity profile [136].

Manufacturing processes for MDIs have also required adapting for the use with the new propellant system [137]. There are two main manufacturing processes used for MDIs: cold filling and pressure filling [138]. Cold filling requires cooling the propellants to below –50°F and filling at that temperature prior to crimping the valve onto the canister. Pressure-filling techniques for MDIs are most commonly employed. These can be accomplished in either a one- or two-step process. In the single-step process, the formulation is placed in a pressurized mixing vessel. The empty canister is purged with propellant to remove the air. The valve is then crimped onto the canister and the formulation is metered through the valve. The absence of a HFA propellant that was liquid at room temperature was a major difference compared to the process employed for CFC manufacturing. In the two-step process, the formulation (excluding the propellant) are mixed together to form a concentrate. Previously, liquefied CFC 11 was used in this step of the process. However, there is no suitable HFA propellant that is liquid at room temperature. Therefore, cosolvents such as ethanol and glyercol are employed during this step to form the product concentrate. The concentrate is metered into the empty canister. The valve

is then crimped onto the canister and the propellant is filled through the valve. Wilkinson (1998) provides an extensive history and review of the manufacturing procedures for MDIs [138].

### 5.4 Excipients

A number of excipients have been included in MDI formulations; however, the nature of the excipients has changed with the introduction of the HFA propellant aerosols. Oleic acid and sorbitan trioleate (SPAN 85) and lecithins were used in CFC suspension MDIs as suspending agents and valve lubricants [112]. Typical concentrations ranged from 0.1 to 2.0% w/w. Ethanol is now being used in HFA formulations as a cosolvent for suspension and solution formulations. The addition of ethanol to the formulation has a number of effects [121, 139]. Increasing the ethanol concentration has been shown to increase the initial droplet size [121]. In addition, ethanol can increase the hydrophilicity of the formulation and increase moisture uptake. Glycerol and polyethyleneglycol have also been added as cosolvents but also have the effect of increasing the residual droplet particle size due to their lower volatility [111]. In general, a relationship can be observed between the fraction of nonvolatile components (drug and nonvolatile excipients) in a solution HFA formulation and the final particle size of the aerosol. The MMAD was observed to be linearly proportional to the cube root of the nonvolatile concentration [119, 121]. Oligolactic acids (OLAs) have been investigated for their use in a variety of functions in HFA formulations. OLAs with repeating units of 6–15 units have been proposed as suspending agents [131]. They are readily soluble in both HFA 134a and 227. These molecules have also been shown to act as ion pair solubilizers for certain drugs (e.g., albuterol). The addition of ethanol to these OLA formulations synergizes the solubilizing effect [131].

A word of caution is required when considering introduction of novel excipients into any inhalation drug product formulation. Due to the unique toxicological challenges associated with administration and clearance from the lung, the qualification of novel excipients for inhalation has proven to be an expensive and time-consuming challenge. This has led to a limited number of compounds with an extensive "in-use" profile being commonly employed.

### 5.5 Valves

Metering valves are required to accurately meter and dispense the formulation upon MDI actuation. In addition, they perform an important contact closure role preventing moisture ingress and minimizing propellant evaporation. Figure 2 and Table 3 show the basic components of the metering valve. Currently, the most common valve type is the retention valve, consisting of a plastic metering chamber and two rubber gaskets. The remaining valve components are manufactured from plastic, metal, and elastomeric materials.

Material component evaluation and selection are critical steps in the development of a MDI formulation [140, 141]. The materials must be chemically resistant and compatible with all components of the formulation. Gaskets must have appropriate mechanical properties and work effectively as a seal, preventing leakage of the formulation and moisture ingress. While the basic components themselves have

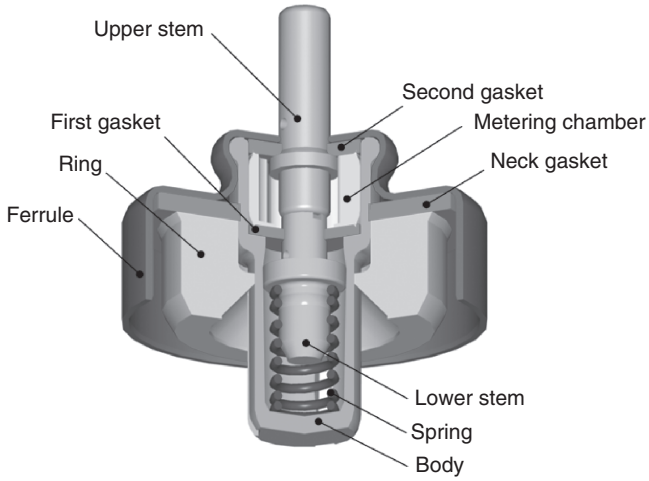


**FIGURE 2** Schematic of components of metering valve. (Courtesy of Valois Pharm.)

**TABLE 3 Summary of Components and Materials Used in Metered Dose Inhaler Valves**

| Component | Material |
|---|---|
| Metering chamber | Polyester |
| Core | Polyester/acetal |
| Core extension | Polyester/acetal |
| Body | Polyester/nylon |
| Seats/gaskets | EPDM/nitrile/butyl/chloroprene/bromobutyl |
| Spring | Stainless steel |
| Ferrule | Aluminium |

remained unchanged during the introduction of the HFA propellants, the materials used to manufacture the components have required significant adaptation. Nitrile was the most commonly employed elastomer in CFC MDIs; it has good mechanical and elastic properties. However, it has been shown to swell when in contact with HFA propellants and ethanol. Newer elastomers such as ethylene propylene diene monomer (EPDM), chloroprene, and bromobutyl are now used in HFA MDIs [142]. The ideal universal elastomer has yet to be developed and the newer materials must be assessed on a case-by-case basis for formulation compatibility and the desired moisture ingress characteristics. Among the many issues to be considered when screening materials are formulation–material compatibility, extractable profiles, and mechanical resistance. Manufacturers such as Valois, Solvay, and Bespak have extensive knowledge of drug/excipient/material component compatibility and should be used as the first point of reference when considering a MDI formulation project.

Another concern to formulators is the ingress of moisture into HFA-formulated MDIs [143]. HFA propellants have a higher moisture affinity compared to the CFC propellants, especially HFA 134a. In addition, the inclusion of ethanol in some formulations increases its hydrophilicity. Moisture entering the canister can have several effects; it may alter the physical or chemical stability of the formulation and aerosolization performance of HFA MDIs. Due to its lower volatility compared to

the other components, water may affect aerosol generation and alter aerosol particle size [144]. The increased water content may increase the solubility of suspended polar drug particles or decrease the solubility of hydrophobic compounds [145]. Corrosion in aluminum canisters may also increase over the shelf life of the product. Williams and Hu (2000) reported that HFA 134a had a greater tendency to take up moisture during storage than did HFA 227 [144]. The issue of moisture ingress during storage has led to certain HFA MDIs being stored in moisture-protecting pouches prior to initial use (e.g., Ventolin HFA). An alternative approach to minimize the effects of moisture ingress has been taken by SkyePharma, which has incorporated subtherapeutic doses of cromolyn sodium into its HFA MDI formulations. Cromolyn sodium is used as a hygroscopic excipient to scavenge any moisture that penetrates into the formulation. Cromolyn sodium has been used widely by inhalation over the past 30 years and has an excellent safety profile via the inhalation route. Burel et al. (2004) reported that for a HFA 134a MDI formulation the inclusion of a polyamide (nylon 66) molded ring around the valve body reduced both the initial water content and the final water content (6 months) when stored under stress conditions [40°C and 75% relative humidity (RH)]. A combination of a thermoplastic elastomer sealing gasket in the MDI valve and a polyamide ring produced the lowest water ingress under these stress conditions [143]. For formulations that might be susceptible to water-induced stability issues, HFA 227 may be considered a more suitable propellant than HFA 134a. Given the possibility of moisture ingress, there is also the issue of propellant leakage. Leak testing is among the array of in-process quality assurance tests that are required. These include assay of the suspension or solution, moisture level, consistency of filling of both the concentrate and the propellant, valve crimp measurements, quality of sealing, in-line leak testing under stress conditions, and performance of the valve.

Another significant issue encountered during use of MDIs was related to loss of prime and dose reproducibility [129, 146]. Loss of prime relates to the fact that in conventional capillary retention metering valves the dose is filled into the valve immediately following the last actuation. Capillary retention valves require priming with one or two sprays prior to their first use. In addition, if there is a significant interval between the actuations and the inhaler is stored upside down or on its side or shaken, then the metering valve may actually partially empty, resulting in a low and variable dose being delivered to the patient. A review of patient information leaflets indicated varying instructions on priming MDIs. This ranged from Atrovent CFC and Combivent CFC requiring priming with 3 sprays "after 24 hours of nonuse." Ventolin HFA and Proventil HFA both required priming with 4 sprays after "2 weeks of nonuse." Flovent CFC required priming with 4 sprays after "4 weeks of nonuse." Clearly, such instructions add to the complexity for patients using MDIs and also contribute to drug waste issues. Loss of prime is also a significant issue for breath-actuated MDIs, where the opportunity to prime the inhaler is not readily possible. A number of new valve designs have been developed to address this issue. The fast-fill, fast-empty valves offer a solution to the priming and loss of prime issues. In these valves [(e.g., 3M Shuttle valve (3M), 3M Face Seal valve (3M), ACT (Valois), and Easifill valve (Bespak)], the metering valve is only isolated from the formulation canister reservoir immediately prior to dose actuation. Therefore, the metering chamber can be emptied and refilled with a fresh dose from the reservoir simply by shaking the canister prior to use.

### 5.6 Actuators

Nonvolatile component concentration has previously been described as one of the primary determinants of the initial droplet size for HFA solution formulations [119]. Perhaps, equally important is the MDI actuator [147]. The actuator consists of the sump block into which the metered dose is immediately delivered during MDI actuation. As expansion and vaporization of the propellant take place, the aerosol exits the sump via the actuator nozzle and then is inhaled through the actuator mouthpiece. From a practical perspective, in general, reducing the size of the orifice diameter for HFA solution formulations produced a relatively slower spray emitted with less force compared to marketed CFC products [118]. The nozzle orifice diameter has been considered to be the most important, although not the only, actuator variable determining the particle size distribution of HFA solution formulations [118, 147, 148]. Recently, Smyth et al. (2006) described three critical components of the actuator that could affect the aerosol performance of a solution HFA formulation. In addition to the orifice diameter, sump depth (and hence the expansion chamber volume) together with orifice length was observed to have significant effects on the aerosol particle size distribution and should be considered for optimization with an HFA formulation [147]. It has also been recognized that the electrostatic charge of all components of the MDI and its formulation may affect the aerosolization properties of the aerosol spray [149, 150].

### 5.7 Canisters

Aluminum canisters are widely used in commercial MDI products mainly due to their inert characteristics. Other materials, including stainless steel and glass, can be employed depending upon the particular formulation characteristics. These canisters were usually uncoated. Changes to the canister may be required when the formulation interacts with the interior surface of the canister altering the chemical stability of the formulation. The presence of ethanol in HFA formulations has also increased the risk of metal corrosion. Drug migration or absorption to the metal components of the canister and also the metal valve components has also been reported [141]. The loss of drug to the walls of the canister will result in variability in the delivered dose from the MDI during the shelf life of the inhaler. The use of canister coating and anodized canisters has been advocated to mitigate this problem [141].

### 5.8 Breath Actuation

In order to overcome the problems associated with many patients' inability to coordinate actuating the MDI and inhaling, breath-actuated MDIs were developed [107]. These devices allow the MDI to be automatically actuated only when the patient commences inhaling through the mouthpiece. Of critical importance here is ensuring that the patient has sufficient inspiratory flow rate to trigger actuation. While these devices offer little improvement for patients with a good inhaler technique, it has been shown that patients with poor coordination did have significantly greater lung drug deposition when inhaling using a breath-actuated MDI [151]. The 3M Autohaler was the first device marketed using this technology [151, 152]. In Europe,

the Easibreathe and Autohaler breath-actuated MDIs are used to deliver β agonists and corticosteroids for the treatment of obstructive airway. Recently, in the United States, the MD Turbo has been launched, a device that allows patients to take their regular MDI canister and actuator and insert it into the MD Turbo. The MD Turbo acts as a generic breath actuator for a number of marketed MDIs and also incorporates a dose counter.

### 5.9 Spacers

Spacer devices have been developed as another alternative to overcome the problems associated with patients coordinating the beginning of their inspiratory effort with actuation of the MDI [153]. This problem is extenuated by the fact that the MDI emits a high-velocity, short-duration aerosol cloud. On actuation, the propellant spray is delivered into the spacer that often incorporates a one-way inhalation valve. The patient is now able to inhale the aerosol cloud. The large-volume spacers have an additional effect in that they allow evaporation of large propellant droplets prior to inhalation. These high-velocity droplets would previously have had a high probability of impacting in the patient's throat. Figure 3 shows the large number of spacer chambers that are available [154]. Spacers are advocated for use by children and elderly patients and people who experience difficulty coordinating actuation of the MDI. The use of spacers for the delivery of corticosteroids also minimizes oral deposition of the inhaled dose and therefore reduces the incidence of steroid-related side effects [155]. Both in vitro and clinical studies have shown the effectiveness of spacers with CFC MDIs [156–158]. It has been shown that electrostatic charge can have a significant effect on the performance of a spacer chamber and where possible the charge should be minimized to maximize drug delivery [159–161]. Finally, it should be noted that the use of any particular spacer–MDI combination should be evaluated at least in vitro to confirm the beneficial effect, especially when employed with solution-based HFA MDI formulations [162, 163].

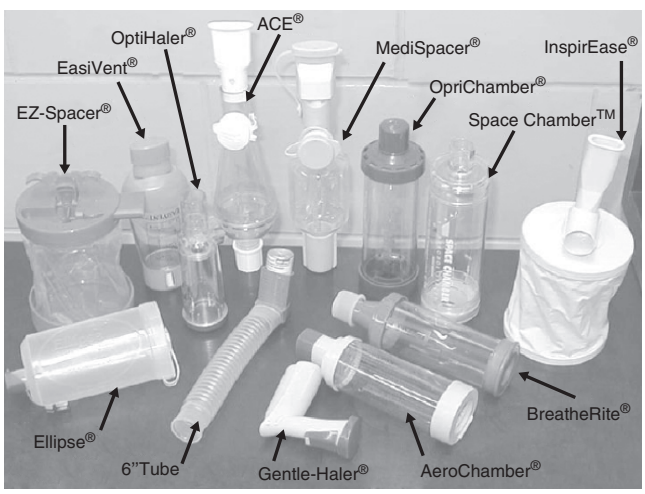


**FIGURE 3** Example spacer chambers available for use with MDIs. (Reproduced from ref. 154 with permission of *Pharmacotherapy*.)

### 5.10 Dose Counters

A guidance document from the Food and Drug Administration (FDA) recommends the addition of a dose counter to the MDI. This would overcome a long-standing problem with the MDI, the inability of a patient to accurately know the number of doses remaining in the canister [164]. Dose counters have been incorporated successfully into multiple-dose DPIs. In general, the counter should give a clear indication of when approaching end of life and the actual end of life. It should be either numeric or color coded. If numeric, it should count downward and should be 100% reliable and avoid undercounting [165].

## 6 DRY POWDER INHALERS

### 6.1 Introduction

Dry power inhalers have been in use for over 40 years. They were developed as an environmentally friendly alternative to the MDI. The early DPIs were simple in design, portable, but again, a relatively inefficient means of delivering drugs to the lungs for local action [166]. The Spinhaler, the first DPI, has been prescribed in Europe since the late 1960s. In general, the acceptance and use of DPIs is much greater in Europe than in the United States. However, with the reformulation efforts for MDIs, there are an increasing number of DPIs becoming available in the United States (Figure 4).

Research and development for dry powder inhalers have two main focuses: the optimization of the powder formulation for use in these inhalers and investigations of novel DPI device designs and technology. An enormous literature now exists in each of these areas; for more extensive reviews readers should consult refs. 33, 167, or 168.

**FIGURE 4** Example DPIs available in United States.

### 6.2 Size Reduction and Particle Formation Technologies

Dry powder inhaler formulations consist usually of either a drug-only formulation or an ordered mixture of drug and excipient, most commonly lactose monohydrate. In both cases, the first challenge is the production of drug particles with suitable size characteristics for inhalation (i.e., 1–5 μm). Traditionally, micronization or jet-milling methods have been employed as the method of choice for conventional small molecules. This method is identical to that employed for the production of fine particles for suspension MDIs. Using this method it is possible to produce primary particles between 1–5 μm. However, as a consequence of the particle size reduction there are a number of undesirable effects with respect to the powder properties. Micronized powders possess high intramolecular forces and are cohesive. They readily form aggregates that are difficult to disperse to the primary particles. Dispersion to its primary particle is essential for successful pulmonary deposition. In addition, they often possess high inherent electrostatic charges which cause particle adhesion to the components of the dry powder inhaler [169]. The high-energy micronization process also causes disruption of the crystal lattice and results in the formation of amorphous regions which may affect the long-term stability of the formulation [170]. Finally, it is not possible to control the drug particle morphology. Despite all of these problems, micronization remains the most common technique employed for respirable particle formation. Modifications to conventional micronization techniques have been investigated as alternative methods of particle size reduction [171–173].

A number of novel particle formation technologies now exist that are able to produce respirable drug particles for formulation in both DPIs and MDIs. Depending on the method of preparation, these particles offer unique and potentially advantageous physical and aerodynamic properties compared to conventional crystallization and micronization techniques. Some investigators have advocated that major improvements in aerosol particle performance may be achieved by lowering particle density and increasing particle size, as large, porous particles display less tendency to agglomerate than (conventional) small and nonporous particles. Also, large, porous particles inhaled into the lungs can potentially release therapeutic substances for long periods of time by escaping phagocytic clearance from the lung periphery, thus enabling therapeutic action for periods ranging from hours to many days [174].

Many of these techniques involve particle formation from solution formulations that contain novel excipients. Spray drying is the most advanced of these technologies and has been used to produce the powder formulation in the Exubera inhaler [175]. Various modifications of this basic technique, including co–spray drying with novel excipients, have been employed.

AIR particles are low-density lipid-based particles that are produced by spray drying lipid–albumin–drug solutions. These particles are characterized by their porous surface characteristics and large geometric diameter while having a low aerodynamic diameter [176, 177]. This technology has been used to produce porous particle powder formulations of L-dopa that have been investigated for the treatment of Parkinson's disease [178].

Pulmospheres are produced using a proprietary spray drying technique, with phosphatidylcholine as an excipient to produce hollow and porous particles with

low interparticulate forces. These particles have been formulated as suspended particles in HFA MDIs. In comparison with conventional suspension MDIs, the Pulmosphere MDI exhibited significantly higher fine particle fractions. This technology has been used to produce cromolyn sodium, albuterol sulfate and formoterol fumarate microspheres [179]. Pulmospheres powder formulations containing tobramycin and budesonide have also been tested clinically [83, 180].

Technosphere technology has been developed as an alternative porous particle for pulmonary delivery [181]. These porous microspheres are formed by precipitating a drug-diketopiperazine derivative from an acidic solution. Para-thryroid hormone (PTH) Technospheres have been investigated for the treatment of osteoporosis following aerosol delivery [182].

The use of supercritical fluid processing technology has also been widely used for its application in controlled microparticle formation. Conventional small molecules and proteins for inhalation have been generated and formulated as powders for inhalation. [183–186].

The application of pulmonary delivery of nanoparticles (<1 um) for pharmaceuticals remains to be developed [187–189].

### 6.3 Drug–Lactose Formulations

The most common means of overcoming cohesion problems is by incorporation of a carrier excipient. Lactose monohydrate is used most often; it is inert, cheap, widely available, and a GRAS (generally regarded as safe) non-toxic excipient. A significant area of research has been undertaken to optimize the critical parameters involved in the formulation of drug–lactose blends. Micronized drug is typically blended with lactose (50–100 μm) to produce an ordered mix. The blend ratio is fixed depending upon the dose of drug to be delivered and the mass of powder blend in each dosage unit (typically between 5 and 25 mg). The aerosolization properties of the blend are related to the adhesive forces between the drug and lactose together with the cohesive forces between the drug particles. Reproducible dispersion of the blend either by the dry powder inhaler (active DPI) or by the patients' inspiratory effort (passive DPI) is required. This allows the detached micronized drug to be inhaled and deposited in the respiratory tract while the larger lactose particles are deposited by inertial impaction in the oropharnyx.

Formulators have become increasingly aware of the criticality of the drug and lactose powder surface characteristics and their relationship to the aerosolization performance in a DPI [190, 191]. A number of investigators have shown in vitro the importance of controlling the size of the lactose and the amount of "fines" (lactose particles less than 5 μm in size) in the drug–lactose blend [192–194]. Inherent fines are present in all lactose powders, and the fines are usually adhered to the surface of the larger lactose particle. These fines are believed to occupy "active" or high-energy sites on the lactose particle surface. Occupation of these sites by the lactose fines prevents the micronized drug from adhering to these positions. This allows the drug to adhere to less active sites and become detached easier from the lactose surface during inhalation. Obviously any significant change in the quantity of fines present in the lactose may alter the distribution of the micronized drug on the lactose particle and therefore the aerosolization characteristics of the powder blend [193–195]. Batch-to-batch control of the fines content of inhalation lactose has been

recognized as critical to ensuring reproducible in vitro emitted and fine particle doses. Jones and Price (2006) have recently surveyed the literature in this area and provided a comprehensive review [196]. Modification of the surface characteristics of the lactose particle has been used as an alternative approach to control the adherence of drug particles to the lactose surface [197–200]. Alternative sugar carriers have also been investigated; these appear to possess many of the same performance-limiting characteristics as lactose [201]. Finally, tertiary additives have also been used to improve the aerosolization properties of DPI formulations [202]. The majority of the studies described above relate to in vitro testing of DPI formulation performance, and little is known about the clinical significance of these studies.

Moisture ingress into a powder formulation is a particular concern as it may significantly decrease the aerosolization performance of the formulation [203, 204]. Increased adhesion of particles is often seen following exposure to high-RH environments [205, 206]. Moisture ingress has also been shown to affect drug stability [170]. The pharmaceutical industry has used a number of approaches to protect powder formulations from the ingress of moisture during storage and for their "in-use" life. The Turbuhaler incorporates a desiccant in the base of the inhaler to keep the power reservoir free from moisture [207]. Unit-dose blisters used in the Diskus are sealed in a foil strip pack to protect each individual dose prior to inhalation [208]. It is also essential that the patient not exhale into the DPI immediately prior to inhaling the dose.

Electrostatic charge can also influence the performance of DPI formulations. A number of studies have investigated the interactions of drug and lactose particle charge with respect to aerosolization properties and drug retention by the plastic components within the inhaler [209–212].

### 6.4 Dry Powder Inhaler Design

***Inhalation Flow Rate*** The main function of a DPI is to facilitate dispersion and delivery of inhalable drug particles. An extensive patent and scientific literature exists describing the ever-increasing number of DPI device designs [33]. Powder dispersion in the early passive DPIs was provided in part by the inspiratory effort of the patient. This removed the necessity to coordinate patient inhalation with actuation and delivery of the dose (in contrast to MDIs). These passive DPIs were "breath actuated," with the patients' inspiratory effort dispersing, aerosolizing, and delivering the powder during the inhalation cycle. The airflow rate through the inhaler was determined by the inherent device resistance and the inspiratory force exerted by the patient [18]. Devices such as the Spinhaler, Rotahaler, and Diskhaler are low-resistance devices requiring relatively high inspiratory flow rates to disperse the powder formulations by turbulent deaggregation. These simple devices have low aerosolization efficiencies with only 5–20% of the dose being delivered to the lungs [166] The inhalation flow rate dependence of passive DPIs has been cited as a potential problem in their use, especially given the large intersubject flow rate variability within the patient population (especially for the young and older patients). In vitro testing revealed that for certain DPIs there was large variability in both the delivered dose and the aerodynamic particle size distribution as a function of the inhalation flow rate [203, 213–216]. Similar clinical studies also revealed a flow rate dependence for certain DPIs while others were observed to perform with a degree

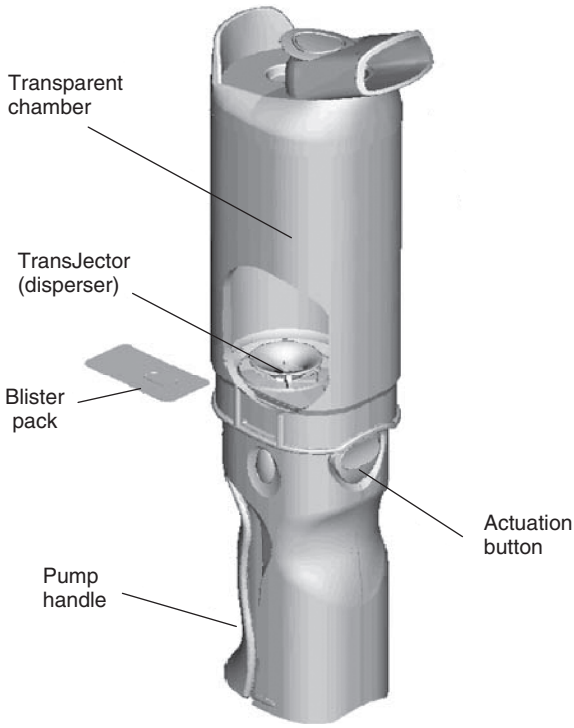


**FIGURE 5** Exubera Inhaler. (Reprinted from ref. 175. Courtesy of Mary Ann Liebert, Inc.)

of flow rate independence [214, 217–221]. When choosing a DPI, the effect of inhalation flow rate should be assessed on a case-by-case basis for each individual DPI, and readers should be aware of contradictory studies, especially when comparing in vitro and clinical performance. The Turbuhaler is one such example, where some in vitro studies show high variability; however, this is not reflected in clinical studies [215, 220, 222].

From these and many other studies it can be concluded that a desirable characteristic for any DPI is that its dose delivery performance is independent of inhalation flow rate. A second generation of DPIs have been developed that incorporate a combination of improved powder formulations, more effective turbulent dispersion within the inhaler, and in some cases an active dispersion mechanism. The Exubera inhaler releases a bolus of compressed air through the formulation and actively generates an aerosol cloud from the powder (Figure 5). The cloud is held within a reservoir chamber from which the patient then inhales the insulin dose [175]. Active dispersion improves device aerosolization efficiency, with greater than 50% of the dose being deposited in the lungs, while minimizing the reliance on the patients' inspiratory effort.

***Single- and Multiple-Dose DPIs*** Inhalation powder dose metering is one of the problems encountered by DPI formulators. The powder dose can range from 250 μg in the drug-only Pulmicort Turbuhaler formulation to 25 mg in the lactose-blended

Spinhaler formulations. In each case, accurate and reproducible metering of the powder is required for regulatory approval and therapeutic efficacy. This proved to be a technological challenge that was solved in a number of ways. Single-unit-dose inhalers were the first generation of DPIs, the unit dose being metered in the factory and subsequently loaded into the inhaler by the patient immediately prior to each dosing. Because metering takes place prior to batch release by the manufacturer, this allows for quality control and release testing, ensuring that dosage units were within acceptable criteria. Procedures such as capsule filling were common for early devices such as the Spinhaler and Rotahaler. This approach is still used by some of the newer devices being developed (e.g., Aerohaler and Cyclohaler) [185, 223]. While popular with the pharmaceutical industry, the single-unit-dose device required significant patient handling to load and empty the inhaler for each inhalation (unlike the MDI, which often contained up to 200 doses available for inhalation on demand). Two approaches were taken toward the design of multiple-dose DPIs; the multiple-unit-dose DPI (e.g., Diskhaler and Diskus) and the powder reservoir multidose DPI (e.g., Turbuhaler) [207]. For the multiple-unit-dose DPI, manufacturers sought to address the requirement for multiple doses while retaining the control of factory premetering. Perhaps the most successful DPI in this respect is the Diskus, in which the dose is premetered into a coiled foil covered strip containing individually sealed blister reservoirs [208]. Each blister is opened immediately prior to inhalation and up to 60 doses can be help in each foil strip. For the powder reservoir multidose DPI, volumetric dose metering of the powder takes place within the DPI immediately prior to inhalation in a manner analogous to MDIs. Among the devices that use this approach are the Turbuhaler, Clickhaler, Pulvinal, and Easyhaler [220, 224–226]. The Turbuhaler is used with a drug-only formulation (although lactose blends have also been used) that employs a proprietary powder agglomeration process to produce loosely bound aggregates that are easily dispersed by the patient's inhalation and by the turbulent flow path encountered in the DPI [227]. Besides the Diskus and Turbuhaler, there are four other devices currently available in the United States, the Asmanex Twisthaler, the Foradil Aerolizer, the Relenza Diskhaler, and the Spiriva Handihaler. Other devices in development include the Novolizer, a multidose, refillable, breath-actuated DPI that delivers up to 200 metered doses of drug from a single cartridge [228, 229]. The Ultrahaler offers yet another alternative DPI [230]. The Taifun inhaler, the JAGO inhaler, and the Airmax are other multidose DPIs [231–234].

### 6.5 Exubera

Systemic delivery of drugs via the lungs offers a noninvasive route of administration. Perhaps the most important and widely investigated molecule considered for this route has been insulin [235, 236]. Following over a decade of development, in January 2006, Pfizer and its partner Nektar received marketing approval for Exubera, their insulin DPI. This offered diabetics a noninvasive route of insulin administration rather than repeated subcutaneous injections [237]. Exubera has been indicated for the treatment of adult patients with diabetes mellitus for the control of hyperglycemia [238]. It has an onset of action similar to rapid-acting insulin analogs and has a duration of glucose-lowering activity comparable to subcutaneously administered regular human insulin [239]. Patton et al. (2004) provided an extensive review

of the clinical pharmacokinetics and pharmacodynamics of inhaled insulin [240]. In patients with type I diabetes, Exubera should be used in regimens that include a longer acting insulin [241]. In patients with type II diabetes, Exubera can be used as monotherapy or in combination with oral agents or longer acting insulins [242–244]. Studies revealed that the same level of blood sugar control was achieved following inhalation compared to subcutaneous injection, although different nominal doses were required due to lung bioavailability issues [245]. The therapeutic efficacy and safety of inhaled insulin appears to have been proven, although there are a significant number of issues with its administration via this route [246]. It has been noted that asthmatics absorb less insulin from the lungs than nonasthmatics. In addition, smokers absorb more insulin than nonsmokers. Small and reversible changes in pulmonary lung function have been observed in some studies with inhaled insulin. Each of these issues has led to the development of specific prescribing guidelines and an intensive physician/patient education program for the inhaled insulin product. The Exubera insulin formulation is a spray-dried, amorphous insulin powder containing 60% insulin in a buffered, sugar-based matrix [175].

Other pharmaceutical companies are also continuing to develop their own inhalation insulin products. Aradigm and NovoNordisk are using a liquid insulin formulation in combination with the AERx IDMS inhaler [247–251]. Alkermes and Lily are developing an insulin product derived from their research on geometrically large, low-density particles that are formed by a spray drying process incorporating a natural phospholipid. MannKind is using its Technosphere technology to produce low-density porous insulin particles. This formulation is delivered using the MedTone inhaler. Other companies working in this area include Kos Pharmaceuticals, Mircodose Technologies, Coremed, and Biosante.

## 7 NEBULIZERS

Nebulization of liquid formulations has long been established as an effective, if not efficient, means of pulmonary drug delivery. The basic principle of nebulizer aerosol generation has remained unchanged; however, a number of technological advances have been made which have improved efficiency and reduced variability. Aerosols that were previously delivered in a continuous inhalation mode over 5–15 min are now delivered only during the inspiratory cycle, thus reducing drug waste. In general, nebulizers convert a liquid into a fine droplet mist, either by means of a compressed gas (jet nebulizer) or by high-frequency sound (ultrasonic nebulizer) [252]. Ultrasonic nebulizers use a piezoelectric source within the formulation reservoir to induce waves at the surface of the nebulizer formulation. Interference of these waves induces the formation of droplets which are then carried in a flowing air stream that is passed over the formulation. These devices are not suitable for the nebulization of suspension formulations [253]. Rau (2002) also observed that ultrasonic nebulizers can increase the solution reservoir temperature and may cause drug degradation [254]. In the case of the jet nebulizer, an aerosol is produced by forcing compressed air through a narrow orifice which is positioned at the end of a capillary tube. The negative pressure created by the expanding jet causes formulation to be drawn up to the capillary tube from the reservoir in which it is immersed. As the liquid emerges from the tip of the capillary, it is drawn into the air stream and

broken up into droplets by the jet to produce an aerosol. Baffle structures within the nebulizer filter the large droplets from the aerosol by impaction and the deposited drug solution is recycled back into the drug reservoir [255]. Only the small aerosol droplets evade impaction on the baffles and are delivered to the patient for inhalation. Jet nebulizers can be categorized by function, for example, the DeVilbiss 646 is a conventional jet nebulizer with continuous drug output resulting in significant waste during exhalation. The Pari LC Plus system incorporated a valve system and operates as an active venturi jet nebulizer; although drug output is continuous, there is an increased output during inhalation. The patient's inspiratory effort increases the nebulizer airflow, thus increasing drug output for these breath-enhanced nebulizers. Finally, dosimetric jet nebulizers such as the Ventstream use a one-way valve system to emit aerosol only during inspiration and are also breath-enhanced nebulizers [256, 257]. It is this last type of nebulizer that offers the most significant advances in technology [258].

Jet nebulizers are commonly used in nonambulatory settings such as hospitals or the patient's home. In vitro studies comparing the performance of commercial nebulizers have concluded that there were large differences in drug delivery between nebulizers of different classes and even between nebulizers of apparently the same class [259–261]. The aerosolization performance of different nebulizers has been found to be dependent upon a number of factors, including the drug being aerosolized, the formulation fill volume, the compressed airflow rate, and breathing pattern [254, 260, 262, 263]. These parameters ultimately control the aerosol droplet size and rate of drug output [264]. However, probably the size of the conventional nebulizer, the duration of the treatment cycle (5–15 min), and the cost of the nebulizers are the main reasons that they are usually reserved for nonambulatory settings and remain less popular than the MDI and DPI.

Solutions or suspensions are available as nebulizer formulations. Due to the relative simplicity in formulating a liquid nebulizer formulation and because of the relatively large range of doses available for delivery, the nebulization method is often chosen as the aerosol method for proof-of-concept investigational studies.

Nebulizer technology continues to be developed to miniaturize and lower the cost of the devices while maintaining the quality of the aerosols generated. The Halolite incorporates adaptive aerosol delivery which monitors patients inspiratory cycle and delivers drug to patients during the first 50% of their inspiratory cycle [265–267]. The Pari eFlow is a hand-held device that uses a vibrating membrane nebulizer to generate a respirable aerosol [268]. Aerogen (now part of Nektar) has a range of nebulizer-based technologies, including the Aeroneb and Aerodose devices. Aerosols are generated as a liquid formulation passes through vibrating apertures [269, 270].

## 8 EMERGING TECHNOLOGIES

### 8.1 Soft Mist Aerosols

In recent years research has focused on a new method of pharmaceutical aerosol generation that involves passing a solution formulation through a nozzle or series of nozzles to generate a "soft mist" aerosol as a bolus dose [271]. Aerosol generation

is achieved by mechanical, thermomechanical or electromechanical processes depending upon the particular technology employed [272]. It is worth noting that these devices are bolus dose delivery inhalers, rather than the new continuous-generation nebulizers which generate aerosols by vibrating porous membranes at ultrasonic frequencies. Such devices include the eFlow and Aerodose, which were described earlier.

While the precise mechanism of soft mist aerosol generation may differ between inhalers, a number of common characteristics can be observed. They are propellant free and produce slow-moving aerosols over an extended duration with high in vitro fine-particle fractions compared to MDIs and DPIs. The aerosols are often generated from simple solution formulations containing pharmaceutically acceptable excipients. Water and ethanol are the most commonly employed vehicles for soft mist aerosols [273]. Perhaps the most simple and advantageous vehicle is water. There is often a well-known and established stability profile of many pharmaceuticals in aqueous solutions, accelerating the route to the clinic in any development program. Drug solubility can be manipulated by choice of water, ethanol, or mixtures of the two to increase formulation options and doses. In multidose reservoir-type devices, a preservative would be required to prevent microbial contamination. This is in addition to the current federal regulations that all aqueous-based drug products for oral inhalation must be manufactured to be sterile.

### 8.2 Respimat

The Respimat inhaler was recently launched in Germany as a combination product of fenoterol and ipratropium hydrobromide (Berodual) and was licensed for the treatment of chronic obstructive airway disease. A large body of literature now exists documenting the aerosol characteristics and clinical performance of the Respimat inhaler with a number of different drugs [274, 275]. Aerosolized formulations include the steroids budesonide and flunisolide in addition to the β agonist fenoterol as well as the commercially available combination product of fenoterol and ipratropium bromide [276–281].

The Respimat device is a multidose reservoir system that is primed by twisting the device base (Figure 6). This compresses a spring and transfers a metered volume of formulation from the drug cartridge to the dosing chamber. The metered volume is between 11 and 15 μl depending upon the drug formulation. When the device is actuated (in coordination with the patient's inspiration), the spring is released. This forces a micropiston into the dosing chamber and pushes the solution through the uniblock. The uniblock is the heart of the aerosol generation system and consists of a filter structure with two fine outlet nozzle channels. The uniblock produces two fine jets of liquid that converge at a precisely set angle and then collide. This collision aerosolizes the liquid to form an aerosol [282].

Aerosols generated from the Respimat inhaler have been characterized as having a prolonged aerosol cloud duration compared to MDIs and have a slower cloud velocity as measured using video camera imaging. Hochrainer et al. (2005) measured the cloud duration of the Respimat aerosol to be 0.2–1.6 s compared to less than 0.2 s for HFA and CFC MDIs. Aerosol velocities have been reported as less than 1 m/s for the Respimat, compared to 6–8 m/s for CFC MDI inhalers [283]. While a degree of patient coordination is required to actuate the Respimat and to inhale,

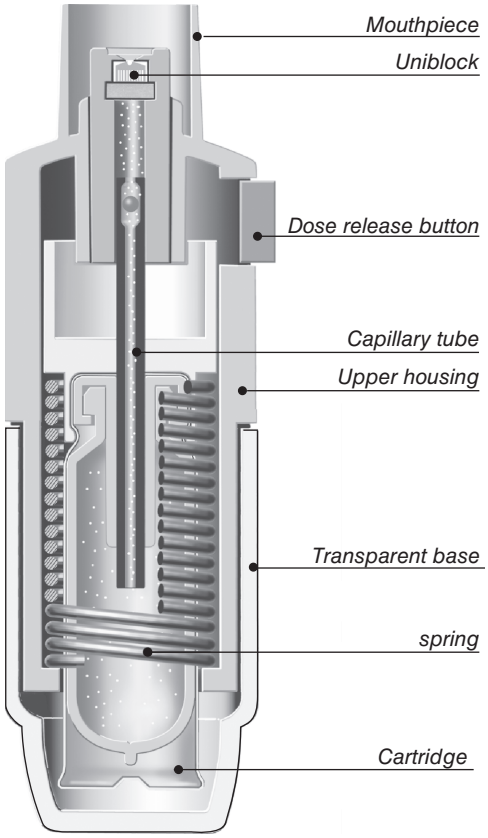


**FIGURE 6** Respimat Inhaler. (Courtesy of Boehringer Ingelheim.)

the longer duration of aerosol cloud generation makes this maneuver less critical than with MDIs.

Aqueous and ethanolic formulations have been employed with the Respimat and the in vitro aerosol performance determined. Zierenberg (1999) reported fine-particle fractions of 66% for an aqueous fenoterol formulation and 81% for an ethanolic flunisolide formulation. The respective MMADs were 2.0 ± 0.4 μm for the aqueous formulation and 1.0 ± 0.3 μm for the ethanolic formulation [284].

## 8.3 AERx

The AERx system was developed for the systemic delivery of insulin. Unit-dose aqueous solution formulations were produced in a blister strip design. The first-generation AERx device is a battery-operated device that guides the patient through the inhalation technique required to successfully deliver a dose. It can also monitor dose times and frequency together with the facility to download dosing data in the clinic. A number of macromolecules, including insulin, and traditional small molecules (e.g., morphine) have been investigated using the AERx technology [89, 251].

Aerosol generation using the AERx system is achieved by mechanically forcing a dose of the liquid formulation though a nozzle array in its disposable unit-dose blisters. The electronic version of the AERx inhaler guides the patient to inhale at the required flow rate. A cam-operated piston mechanism is actuated to compress the blister and extrude the dose as an aerosol through the nozzle array into warmed flowing air. The nozzle array consists of a number of laser-drilled holes. Nozzle design characteristics can be altered depending upon the formulation characteristics and the desired droplet particle size. The single-use nature of the blister avoids potential problems such as microbial contamination from a dosing solution reservoir and nozzle-clogging issues.

A number of prototype versions of the AERx system have been investigated. In general, the in vitro aerosol characteristics revealed that about 50–60% of the loaded dose was emitted from the device, of which over 90% was respirable. MMADs ranged from 1 to 3 μm depending upon the formulation and nozzle array [285]. In a scintigraphic study, lung deposition following inhalation from the AERx was 53.3% (expressed as a percentage of the radioactivity in the AERx blister) compared to 21.7% for an MDI [285].

A number of clinical studies delivering insulin to diabetic patients using the AERx system are currently ongoing. Hermansen et al. (2004) concluded that in type II diabetics, preprandial inhaled insulin via the AERx was as effective as preprandial subcutaneous insulin in achieving glycemic control [286]. Clinical studies with morphine revealed comparable analgesic efficacy for a matched dose of inhaled and intravenous morphine in a postsurgical pain model [251]. In addition, the AERx inhaler has been employed for the topical delivery of rhDNase to cystic fibrosis patients. A mean relative increase in forced expiratory volume in 1 s (FEV1) of 7.8% was observed after 15 days treatment compared to control [287].

### 8.4 Mystic

The Mystic inhaler offers a soft mist aerosol generated from solution or suspension formulations. Unlike the previously described soft mist inhalers which use purely mechanical forces to generate the aerosol, the Mystic inhaler applies an electric field to the formulation within the spray nozzle [288]. An electric charge builds on the fluid surface and, as the droplets exit the nozzle, the repelling force of the surface charge overcomes the surface tension of the droplets to form a soft mist droplet aerosol. This process is known as electrohydrodynamic aerosolization or electrospray. The particle size characteristics of the aerosol can be controlled by adjusting the physical and chemical characteristics of the formulation together with the formulation flow rate and electrical field properties. The inhaler consists of a number of components, a drug containment system, metering system, aerosol nozzle, power supply, and microprocessor, all enclosed in a housing. To date, Ventaira reports that the inhaler has been successfully employed to generate aerosols from small-molecule formulations (albuterol, triamcinolone, cromolyn, budesonide, and terbutaline) and macromolecules, including insulin [288].

### 8.5 Capillary Aerosol Generator

In the capillary aerosol generator (CAG) system, the aerosol is formed by pumping the drug formulation through a small, electrically heated capillary. Upon exiting the

capillary, the formulation is rapidly cooled by ambient air to produce an aerosol. The generated aerosol characteristics are dependent upon the formulation employed. Using propylene glycol as a condensing vehicle, drug containing condensation aerosols are generated [289]. When using water, ethanol, or combinations of both as noncondensing excipients, a stream of solid particles is delivered as a soft mist aerosol. In vitro studies using budesonide, cromolyn sodium, buprenophine, albuterol, and insulin have been performed to demonstrate various applications of the CAG technology. These studies are characterized by high emitted doses and high fine-particle fractions. Using noncondensing excipients, it is possible to produce aerosols with vastly different size characteristics, depending upon the required application.

### 8.6 Staccato

This technology utilizes a rapid heating technique to vaporize a thin film of drug. Following vaporization, the drug particles condense in the inhalation flow stream to form a respirable aerosol and are inhaled. Single- and multiple-dose breath-actuated inhalers are currently in development. As with any method involving heating of a formulation, drug degradation must be minimal. Rabinowitz et al. (2006) described the absorption of prochlorperazine from human lungs as similar to the pharmacokinetic profiles observed following intravenous administration [290, 291].

## 9 CONCLUSIONS

Pharmaceutical aerosol drug delivery has been established for over 50 years. Pulmonary administration remains the route of choice for local treatment of respiratory diseases. Over the past decade there have been changes in both the diseases treated by this route and the devices used for aerosol generation. Future advances will see pulmonary delivery of gene therapy and vaccines, together with improved drug targeting within the respiratory tract using novel inhalers.

## ACKNOWLEDGMENTS

The author would like to thank Suparna Das Choudhuri and Deepika Arora for their assistance and discussions during the preparation of this chapter. In addition, he is grateful to Guillaume Brouet (Valois Pharm), Michael Spallek (Boehringer Ingelheim), and Joanne Peart (RDD) for their help in obtaining figures and tables used in this chapter. Finally, the review of soft mist inhalers has previously been published in the Drug Delivery Company Report (Autumn/Winter 2004), and the author acknowledges PharmaVentures and the Drug Delivery Company Report, which allowed reproduction of an abridged form of this paper.

The author received a research grant from Chrysalis Technologies, a division of Philip Morris USA, for the development of the CAG technology.

## REFERENCES


1. Byron, P. R., and Patton, J. S. (1994), Drug delivery via the respiratory tract, *J. Aerosol Med.*, 7, 49–75.
2. Gonda, I. (2006), Systemic delivery of drugs to humans via inhalation, *J. Aerosol Med.*, 19, 47–53.
3. Douglas, R. B. (1987), The physiology of the lung, in Ganderton, D. and Jones, T. M., Eds., *Drug Delivery to the Respiratory Tract*, Ellis Horwood, Chichester, pp 13–26.
4. Weibel, E. R. (1963), *Morphometry of the Human Lung*, Academic, New York.
5. Gonda, I. (1990), Aerosols for delivery of therapeutic and diagnostic agents to the respiratory tract, *Crit. Rev. Ther. Drug Carrier Syst.*, 6, 273–313.
6. Stuart, B. O. (1973), Deposition of inhaled aerosols, *Arch. Intern. Med.*, 131, 60–73.
7. Palmes, E. D. (1973), Measurement of pulmonary air spaces using aerosols, *Arch. Intern. Med.*, 131, 76–79.
8. Finlay, W. H. (2001), Particle deposition in the respiratory tract, in Finlay, W. H., Ed., *The Mechanisms of Inhaled Pharmaceutical Aerosols*, Academic, London, pp 119–174.
9. Lippmann, M., and Albert, R. E. (1969), The effect of particle size on the regional deposition of inhaled aerosols in the human respiratory tract, *Am. Ind. Hyg. Assoc. J.*, 30, 257–275.
10. Foord, N., Black, A., and Walsh, M. (1975), Pulmonary deposition of inhaled particles with diameters in the range 2.5 to 7.5 micron, *Inhaled Part.*, 4(Pt. 1), 137–149.
11. Taulbee, D. B., and Yu, C. P. (1975), A theory of aerosol deposition in the human respiratory tract, *J. Appl. Physiol.*, 38, 77–85.
12. Yu, C. P., and Taulbee, D. B. (1975), A theory of predicting respiratory tract deposition of inhaled particles in man, *Inhaled Part.*, 4(Pt. 1), 35–47.
13. Pavia, D., and Thomson, M. L. (1976), The fractional deposition of inhaled 2 and 5 mum particles in the alveolar and tracheobronchial regions of the healthy human lung, *Ann. Occup. Hyg.*, 19, 109–114.
14. Bennett, W. D., Brown, J. S., Zeman, K. L., Hu, S. C., Scheuch, G., and Sommerer, K. (2002), Targeting delivery of aerosols to different lung regions, *J. Aerosol Med.*, 15, 179–188.
15. Dolovich, M. A. (2000), Influence of inspiratory flow rate, particle size, and airway caliber on aerosolized drug delivery to the lung, *Respir. Care*, 45, 597–608.
16. Bennett, W. D. (1988), Human variation in spontaneous breathing deposition fraction: A review, *J. Aerosol Med.*, 1, 67–80.
17. Hindle, M., Newton, D. A., and Chrystyn, H. (1993), Investigations of an optimal inhaler technique with the use of urinary salbutamol excretion as a measure of relative bioavailability to the lung, *Thorax*, 48, 607–610.
18. Clark, A. R., and Hollingworth, A. M. (1993), The relationship between powder inhaler resistance and peak inspiratory conditions in healthy volunteers-implications for in vitro testing, *J. Aerosol Med.*, 3, 99–110.
19. Broeders, M. E. A. C., Molema, J., Vermue, N. A., and Folgering, H. T. M. (2001), Peak inspiratory flow rate and slope of the inhalation profiles in dry powder inhalers, *Eur. Respir. J.*, 18, 780–783.
20. Phalen, R. F., and Oldham, M. J. (2001), Methods for modeling particle deposition as a function of age, *Respir. Physiol.*, 128, 119–130.
21. Segal, R. A., Martonen, T. B., Kim, C. S., and Shearer, M. (2002), Computer simulations of particle deposition in the lungs of chronic obstructive pulmonary disease patients, *Inhal. Toxicol.*, 14, 705–720.

22. Kim, C. S., and Hu, S. C. (1998), Regional deposition of inhaled particles in human lungs: Comparison between men and women, *J. Appl. Physiol.*, 84, 1834–1844.
23. Kim, C. S., and Kang, T. C. (1997), Comparative measurement of lung deposition of inhaled fine particles in normal subjects and patients with obstructive airway disease, *Am. J. Respir. Crit. Care Med.*, 155, 899–905.
24. Bennett, W. D., Zeman, K. L., and Kim, C. (1996), Variability of fine particle deposition in healthy adults: Effect of age and gender, *Am. J. Respir. Crit. Care Med.*, 153, 1641–1647.
25. Kim, C. S., Abraham, W. M., Garcia, L., and Sackner, M. A. (1989), Enhanced aerosol deposition in the lung with mild airways obstruction, *Am. Rev. Respir. Dis.*, 139, 422–426.
26. Kim, C. S., Lewars, G. A., and Sackner, M. A. (1988), Measurement of total lung aerosol deposition as an index of lung abnormality, *J. Appl. Physiol.*, 64, 1527–1536.
27. Palmes, E. D., Goldring, R. M., Wang, C., and Altshuler, B. (1970), Effect of chronic obstructive pulmonary disease on rate of deposition of aerosols in the lung during breath holding, *Inhaled Part.*, 1, 123–130.
28. Suarez, S., and Hickey, A. J. (2000), Drug properties affecting aerosol behavior, *Respir. Care*, 45, 652–666.
29. Groneberg, D. A., Witt, C., Wagner, U., Chung, K. F., and Fischer, A. (2003), Fundamentals of pulmonary drug delivery, *Respir. Med.*, 97, 382–387.
30. Labiris, N. R., and Dolovich, M. B. (2003), Pulmonary drug delivery. part I: Physiological factors affecting therapeutic effectiveness of aerosolized medications, *Br. J. Clin. Pharmacol.*, 56, 588–599.
31. Hindle, M., and Chrystyn, H. (1992), Determination of the relative bioavailability of salbutamol to the lung following inhalation, *Br. J. Clin. Pharmacol.*, 34, 311–315.
32. Niven, R. W. (2004), Modulated drug therapy with inhalation aerosols: Revisited, in Hickey, A. J., Ed., *Pharmaceutical Inhalation Aerosol Technology*, 2nd ed., Marcel Dekker, New York, pp 551–570.
33. Niven, R. W. (2002), Powders and processing: Deagglomerating a dose of patents and publications, in Dalby, R. N., Byron, P.R., Peart, J., and Farr, S. J., Eds., *Respiratory Drug Delivery VIII*, Davis Horwood International Publishing, Raleigh, NC, pp 257–266.
34. Cook, R. O., Pannu, R. K., and Kellaway, I. W. (2005), Novel sustained release microspheres for pulmonary drug delivery, *J. Controlled Release*, 104, 79–90.
35. Crooks, P. A., and Al-Ghananeem, A. M. (2004), Drug targeting to the lung: Chemical and biological considerations, in Hickey, A. J., Ed., *Pharmaceutical Inhalation Aerosol Technology*, 2nd ed., Marcel Dekker, New York, pp 89–154.
36. Sitar, D. S. (1996), Clinical pharmacokinetics of bambuterol, *Clin. Pharmacokinet.*, 31, 246–256.
37. Svensson, L. A. (1991), Mechanism of action of bambuterol: A beta-agonist prodrug with sustained lung affinity, *Agents Actions Suppl.*, 34, 71–78.
38. Taburet, A. M., and Schmit, B. (1994), Pharmacokinetic optimisation of asthma treatment, *Clin. Pharmacokinet.*, 26, 396–418.
39. Lipworth, B. J. (1996), Pharmacokinetics of inhaled drugs, *Br. J. Clin. Pharmacol.*, 42, 697–705.
40. Hochhaus, G., and Mollmann, H. (1992), Pharmacokinetic/pharmacodynamic characteristics of the beta-2-agonists terbutaline, salbutamol and fenoterol, *Int. J. Clin. Pharmacol. Ther. Toxicol.*, 30, 342–362.
41. Dahl, A. R., and Lewis, J. L. (1993), Respiratory tract uptake of inhalants and metabolism of xenobiotics, *Annu. Rev. Pharmacol. Toxicol.*, 33, 383–407.

42. Witek, T. J., Jr. (2000), The fate of inhaled drugs: The pharmacokinetics and pharmacodynamics of drugs administered by aerosol, *Respir. Care*, 45, 826–830.
43. Issar, M., Mobley, C., Khan, P., and Hochhaus, G. (2004), Pharmacokinetics and pharmacodynamics of drugs delivered to the lungs, in Hickey, A. J., Ed., *Pharmaceutical Inhalation Aerosol Technology*, 2nd ed., Marcel Dekker, New York, pp 215–252.
44. Hussain, A., Arnold, J. J., Khan, M. A., and Ahsan, F. (2004), Absorption enhancers in pulmonary protein delivery, *J. Controlled Release*, 94, 15–24.
45. Nelson, H. S., Chapman, K. R., Pyke, S. D., Johnson, M., and Pritchard, J. N. (2003), Enhanced synergy between fluticasone propionate and salmeterol inhaled from a single inhaler versus separate inhalers, *J. Allergy Clin. Immunol.*, 112, 29–36.
46. Dhillon, S., and Keating, G. M. (2006), Beclometasone dipropionate/formoterol: In an HFA-propelled pressurised metered-dose inhaler, *Drugs*, 66, 1475–1483.
47. Nowak, R., Emerman, C., Hanrahan, J. P., et al. (2006), A comparison of levalbuterol with racemic albuterol in the treatment of acute severe asthma exacerbations in adults, *Am. J. Emerg. Med.*, 24, 259–267.
48. Cass, L. M., Brown, J., Pickford, M., et al. (1999), Pharmacoscintigraphic evaluation of lung deposition of inhaled zanamivir in healthy volunteers, *Clin. Pharmacokinet.*, 36(Suppl. 1), 21–31.
49. Diot, P., Vecellio-None, L., Varaigne, F., Marchand, S., and Lemarie, E. (2003), Role of rhDNase in cystic fibrosis, *Rev. Mal. Respir.*, 20, S171–S175.
50. Anonymous (2006), First inhaled insulin product approved, *FDA Consum.*, 40, 28–29.
51. Clark, A. R. (2004), Pulmonary delivery technology: Recent advances and potential for the new millennium, in Hickey, A. J., Ed., *Pharmaceutical Inhalation Aerosol Technology*, 2nd ed., Marcel Dekker, New York, pp 571–591.
52. Roth, Y., Chapnik, J. S., and Cole, P. (2003), Feasibility of aerosol vaccination in humans, *Ann. Otol. Rhinol. Laryngol.*, 112, 264–270.
53. Sabin, A. B., Fernandez de Castro, J., Flores Arechiga, A., Sever, J. L., Madden, D. L., and Shekarchi, I. (1982), Clinical trials of inhaled aerosol of human diploid and chick embryo measles vaccine, *Lancet.*, 2, 604.
54. Sabin, A. B. (1983), Immunization against measles by aerosol, *Rev. Infect. Dis.*, 5, 514–523.
55. Sabin, A. B., Albrecht, P., Takeda, A. K., Ribeiro, E. M., and Veronesi, R. (1985), High effectiveness of aerosolized chick embryo fibroblast measles vaccine in seven-month-old and older infants, *J. Infect. Dis.*, 152, 1231–1237.
56. Sabin, A. B., Flores Arechiga, A., Fernandez de Castro, J., Albrecht, P., Sever, J. L., and Shekarchi, I. (1984), Successful immunization of infants with and without maternal antibody by aerosolized measles vaccine. II. Vaccine comparisons and evidence for multiple antibody response, *JAMA*, 251, 2363–2371.
57. Fernandez-de Castro, J., Kumate-Rodriguez, J., Sepulveda, J., Ramirez-Isunza, J. M., and Valdespino-Gomez, J. L. (1997), Measles vaccination by the aerosol method in Mexico, *Salud Publica Mex.*, 39, 53–60.
58. Bennett, J. V., Fernandez de Castro, J., Valdespino-Gomez, J. L., et al. (2002), Aerosolized measles and measles-rubella vaccines induce better measles antibody booster responses than injected vaccines: Randomized trials in Mexican schoolchildren, *Bull. World Health Organ.*, 80, 806–812.
59. LiCalsi, C., Maniaci, M. J., Christensen, T., Phillips, E., Ward, G. H., and Witham, C. (2001), A powder formulation of measles vaccine for aerosol delivery, *Vaccine*, 19, 2629–2636.
60. Laube, B. L. (2005), The expanding role of aerosols in systemic drug delivery, gene therapy, and vaccination, *Respir. Care*, 50, 1161–1176.

61. Blanchard, J. D. (2004), Pulmonary drug delivery as a first response to bioterrorism, in Dalby, R. N., Byron, P. R., Peart, J., Suman, J. D., and Farr, S. J., Eds., *Respiratory Drug Delivery IX*, Davis Healthcare International, River Grove, IL, pp 73–82.
62. Li, H. Y., Seville, P. C., Williamson, I. J., and Birchall, J. C. (2005), The use of amino acids to enhance the aerosolisation of spray-dried powders for pulmonary gene therapy, *J. Gene Med.*, 7, 343–353.
63. Joseph, P. M., O'Sullivan, B. P., Lapey, A., et al. (2001), Aerosol and lobar administration of a recombinant adenovirus to individuals with cystic fibrosis. I. Methods, safety, and clinical implications, *Hum. Gene Ther.*, 12, 1369–1382.
64. Perricone, M. A., Morris, J. E., Pavelka, K., et al. (2001), Aerosol and lobar administration of a recombinant adenovirus to individuals with cystic fibrosis. II. Transfection efficiency in airway epithelium, *Hum. Gene Ther.*, 12, 1383–1394.
65. Rochat, T., and Morris, M. A. (2002), Gene therapy for cystic fibrosis by means of aerosol, *J. Aerosol Med.*, 15, 229–235.
66. Garcia-Contreras, L., and Hickey, A. J. (2003), Aerosol treatment of cystic fibrosis, *Crit. Rev. Ther. Drug Carrier Syst.*, 20, 317–356.
67. Gautam, A., Waldrep, J. C., and Densmore, C. L. (2003), Aerosol gene therapy, *Mol. Biotechnol.*, 23, 51–60.
68. Anson, D. S., Smith, G. J., and Parsons, D. W. (2006), Gene therapy for cystic fibrosis airway disease—is clinical success imminent? *Curr. Gene Ther.*, 6, 161–179.
69. Griesenbach, U., Geddes, D. M., and Alton, E. W. (2006), Gene therapy progress and prospects: Cystic fibrosis, *Gene Ther.*, 13, 1061–1067.
70. Ziady, A. G., and Davis, P. B. (2006), Current prospects for gene therapy of cystic fibrosis, *Curr. Opin. Pharmacol*, 6, 515–521.
71. Cryan, S. A. (2005), Carrier-based strategies for targeting protein and peptide drugs to the lungs, *AAPS J.*, 7, E20–41.
72. Shoyele, S. A., and Slowey, A. (2006), Prospects of formulating proteins/peptides as aerosols for pulmonary drug delivery, *Int. J. Pharm.*, 314, 1–8.
73. Adjei, A., and Garren, J. (1990), Pulmonary delivery of peptide drugs: Effect of particle size on bioavailability of leuprolide acetate in healthy male volunteers, *Pharm. Res.*, 7, 565–569.
74. Adjei, A., Sundberg, D., Miller, J., and Chun, A. (1992), Bioavailability of leuprolide acetate following nasal and inhalation delivery to rats and healthy humans, *Pharm. Res.*, 9, 244–249.
75. Zheng, J. Y., Fulu, M. Y., Lee, D. Y., Barber, T. E., and Adjei, A. L. (2001), Pulmonary peptide delivery: Effect of taste-masking excipients on leuprolide suspension metered-dose inhalers, *Pharm. Dev. Technol.*, 6, 521–530.
76. Shahiwala, A., and Misra, A. (2005), A preliminary pharmacokinetic study of liposomal leuprolide dry powder inhaler: A technical note, *AAPS PharmSciTech.*, 6, E482–486.
77. Deftos, L. J., Nolan, J. J., Seely, B. L., et al. (1997), Intrapulmonary drug delivery of salmon calcitonin, *Calcif. Tissue Int.*, 61, 345–347.
78. Chan, H. K., Clark, A. R., Feeley, J. C., et al. (2004), Physical stability of salmon calcitonin spray-dried powders for inhalation, *J. Pharm. Sci.*, 93, 792–804.
79. Colthorpe, P., Farr, S. J., Smith, I. J., Wyatt, D., and Taylor, G. (1995), The influence of regional deposition on the pharmacokinetics of pulmonary-delivered human growth hormone in rabbits, *Pharm. Res.*, 12, 356–359.
80. Bosquillon, C., Preat, V., and Vanbever, R. (2004), Pulmonary delivery of growth hormone using dry powders and visualization of its local fate in rats, *J. Controlled Release*, 96, 233–244.

81. Ohmori, Y., Onoue, S., Endo, K., Matsumoto, A., Uchida, S., and Yamada, S. (2006), Development of dry powder inhalation system of novel vasoactive intestinal peptide (VIP) analogue for pulmonary administration, *Life Sci.*, 79, 138–143.
82. Ramsey, B. W., Pepe, M. S., Quan, J. M., et al. (1999), Intermittent administration of inhaled tobramycin in patients with cystic fibrosis. cystic fibrosis inhaled tobramycin study group, *N. Engl. J. Med.*, 340, 23–30.
83. Newhouse, M. T., Hirst, P. H., Duddu, S. P., et al. (2003), Inhalation of a dry powder tobramycin PulmoSphere formulation in healthy volunteers, *Chest*, 124, 360–366.
84. Iacono, A. T., Johnson, B. A., Grgurich, W. F., et al. (2006), A randomized trial of inhaled cyclosporine in lung-transplant recipients, *N. Engl. J. Med.*, 354, 141–150.
85. Staniforth, J. N. (2006), Nasal and pulmonary powder opportunities: New drugs and formulations for rapid systemic onset, in: Dalby, R. N., Byron, P. R., Peart, J., Suman, J. D., and Farr, S. J., Eds., *Respiratory Drug Delivery*, 2006, Davis Healthcare International, River Grove, IL, pp. 249–256.
86. Azarmi, S., Tao, X., Chen, H., et al. (2006), Formulation and cytotoxicity of doxorubicin nanoparticles carried by dry powder aerosol particles, *Int. J. Pharm.*, 319, 155–161.
87. Ward, M. E., Woodhouse, A., Mather, L. E., et al. (1997), Morphine pharmacokinetics after pulmonary administration from a novel aerosol delivery system, *Clin. Pharmacol. Ther.*, 62, 596–609.
88. Mather, L. E., Woodhouse, A., Ward, M. E., Farr, S. J., Rubsamen, R. A., and Eltherington, L. G. (1998), Pulmonary administration of aerosolised fentanyl: Pharmacokinetic analysis of systemic delivery, *Br. J. Clin. Pharmacol.*, 46, 37–43.
89. Otulana, B., Okikawa, J., Linn, L., Morishige, R., and Thipphawong, J. (2004), Safety and pharmacokinetics of inhaled morphine delivered using the AERx system in patients with moderate-to-severe asthma, *Int. J. Clin. Pharmacol. Ther.*, 42, 456–462.
90. Fulda, G. J., Giberson, F., and Fagraeus, L. (2005), A prospective randomized trial of nebulized morphine compared with patient-controlled analgesia morphine in the management of acute thoracic pain, *J. Trauma*, 59, 383–388; discussion 389–390.
91. Kohler, D. (1994), Aerosolized heparin, *J. Aerosol Med.*, 7, 307–314.
92. Bendstrup, K. E., Chambers, C. B., Jensen, J. I., and Newhouse, M. T. (1999), Lung deposition and clearance of inhaled (99m)Tc-heparin in healthy volunteers, *Am. J. Respir. Crit. Care Med.*, 160, 1653–1658.
93. Qi, Y., Zhao, G., Liu, D., et al. (2004), Delivery of therapeutic levels of heparin and low-molecular-weight heparin through a pulmonary route, *Proc. Natl. Acad. Sci. U S A*, 101, 9867–9872.
94. Olschewski, H., Rohde, B., Behr, J., et al. (2003), Pharmacodynamics and pharmacokinetics of inhaled iloprost, aerosolized by three different devices, in severe pulmonary hypertension, *Chest*, 124, 1294–1304.
95. Ganderton, D. (1997), General factors influencing drug delivery to the lung, *Respir. Med.*, 91(Suppl. A), 13–16.
96. Buck, H. (2001), The ideal inhaler for asthma therapy, *Med. Device Technol.*, 12, 24–27.
97. Virchow, J. C. (2005), What plays a role in the choice of inhaler device for asthma therapy? *Curr. Med. Res. Opin.*, 21(Suppl. 4), S19–S25.
98. Rau, J. L. (2005), Determinants of patient adherence to an aerosol regimen, *Respir. Care*, 50, 1346–1356; discussion 1357–1359.
99. Ganderton, D. (1999), Targeted delivery of inhaled drugs: Current challenges and future goals, *J. Aerosol Med.*, 12 (Suppl. 1), S3–S8.

100. Dolovich, M. B., Ahrens, R. C., Hess, D. R., et al. (2005), Device selection and outcomes of aerosol therapy: Evidence-based guidelines: American College of Chest Physicians/ American College of Asthma, Allergy, and Immunology, *Chest*, 127, 335–371.
101. O'Byrne, P. M. (1995), Clinical comparisons of inhaler systems: What are the important aspects? *J. Aerosol Med.*, 8(Suppl. 3), S39–46; discussion S47.
102. O'Callaghan, C., and Barry, P. W. (2000), How to choose delivery devices for asthma, *Arch. Dis. Child.*, 82, 185–187.
103. Newhouse, M. (1991), Advantages of pressurized canister metered dose inhalers, *J. Aerosol Med.*, 4, 139–150.
104. Crompton, G. K. (1990), The adult patient's difficulties with inhalers, *Lung*, 168(Suppl)., 658–662.
105. Harnor, K. J., Perkins, A. C., Wastie, M., et al. (1993), Effect of vapor pressure on the deposition pattern from solution phase metered dose inhalers, *Int. J. Pharm.*, 95, 111–116.
106. Leach, C. L., Davidson, P. J., Hasselquist, B. E., and Boudreau, R. J. (2002), Lung deposition of hydrofluoroalkane-134a beclomethasone is greater than that of chlorofluorocarbon fluticasone and chlorofluorocarbon beclomethasone: A cross-over study in healthy volunteers, *Chest*, 122, 510–516.
107. Leach, C. L., Davidson, P. J., Hasselquist, B. E., and Boudreau, R. J. (2005), Influence of particle size and patient dosing technique on lung deposition of HFA-beclomethasone from a metered dose inhaler, *J. Aerosol Med.*, 18, 379–385.
108. Leach, C. L. (1998), Improved delivery of inhaled steroids to the large and small airways, *Respir. Med.*, 92(Suppl. A), 3–8.
109. Williams, R. O., III, and Liu, J. (1998), Influence of formulation additives on the vapor pressure of hydrofluoroalkane propellants, *Int. J. Pharm.*, 166, 99–103.
110. Newman, S. P., Moren, F., Pavia, D., Corrado, O., and Clarke, S. W. (1982), The effects of changes in metered volume and propellant vapor pressure on the deposition of pressurized inhalation aerosols, *Int. J. Pharm.*, 11, 337–344.
111. Brambilla, G., Ganderton, D., Garzia, R., Lewis, D., Meakin, B., and Ventura, P. (1999), Modulation of aerosol clouds produced by pressurised inhalation aerosols, *Int. J. Pharm.*, 186, 53–61.
112. Vervaet, C., and Byron, P. R. (1999), Drug-surfactant-propellant interactions in HFA-formulations, *Int. J. Pharm.*, 186, 13–30.
113. Steckel, H., and Wehle, S. (2004), A novel formulation technique for metered dose inhaler (MDI) suspensions, *Int. J. Pharm.*, 284, 75–82.
114. Berry, J., Kline, L. C., Sherwood, J. K., et al. (2004), Influence of the size of micronized active pharmaceutical ingredient on the aerodynamic particle size and stability of a metered dose inhaler, *Drug Dev. Ind. Pharm.*, 30, 705–714.
115. Berry, J., Kline, L., Naini, V., Chaudhry, S., Hart, J., and Sequeira, J. (2004), Influence of the valve lubricant on the aerodynamic particle size of a metered dose inhaler, *Drug. Dev. Ind. Pharm.*, 30, 267–275.
116. Berry, J., Heimbecher, S., Hart, J. L., and Sequeira, J. (2003), Influence of the metering chamber volume and actuator design on the aerodynamic particle size of a metered dose inhaler, *Drug. Dev. Ind. Pharm.*, 29, 865–876.
117. Dunbar, C. A. (1998), Atomization mechanisms of the pressurized metered dose inhaler, *Part. Sci. Technol.*, 15, 253–271.
118. Gabrio, B. J., Stein, S. W., and Velasquez, D. J. (1999), A new method to evaluate plume characteristics of hydrofluoroalkane and chlorofluorocarbon metered dose inhalers, *Int. J. Pharm.*, 186, 3–12.

119. Lewis, D. A., Ganderton, D., Meakin, B. J., and Brambilla, G. (2004), Theory and practice with solution systems, in: Dalby, R. N., Byron, P. R., Peart, J., Suman, J.D., and Farr, S. J., eds., *Respiratory Drug Delivery IX*, Davis Healthcare International, River Grove, IL, pp. 109–115.
120. Polli, G. P., Grim, W. M., Bacher, F. A., and Yunker, M. H. (1969), Influence of formulation on aerosol particle size, *J. Pharm. Sci.*, 58, 484–486.
121. Stein, S. W., and Myrdal, P. B. (2004), A theoretical and experimental analysis of formulation and device parameters affecting solution MDI size distributions, *J. Pharm. Sci.*, 93, 2158–2175.
122. Purewal, T. S., and Grant, D. J. W., Eds. (1998), *Metered Dose Inhaler Technology*, CRC Press, Boca Raton.
123. Rogers, D. F., and Ganderton, D. (1995), Determining equivalence of inhaled medications, consensus statement from a workshop of the British association for lung research, held at Royal Brompton National Heart & Lung Institute, London 24 June 1994, *Respir. Med.*, 89, 253–261.
124. Meakin, B. J., Lewis, D. A., Ganderton, D., and Brambilla, G. (2000), Countering challenges posed by mimicry of CFC performance using HFA systems, in: Dalby, R. N., Byron, P. R., Peart, J., and Farr, S. J., Eds., *Respiratory Drug Delivery VII*, Serentec, Raleigh, NC, pp. 99–107.
125. Smyth, H. D., and Leach, C. L. (2005), Alternative propellant aerosol delivery systems, *Crit. Rev. Ther. Drug Carrier Syst.*, 22, 493–534.
126. Pritchard, J. N., and Genova, P. (2006), Adapting the pMDI to deliver novel drugs: Insulin and beyond, in: Dalby, R. N., Byron, P. R., Peart, J., Suman, J. D., and Farr, S. J., Eds., *Respiratory Drug Delivery*, 2006, Davis Healthcare International, River Grove, IL, pp. 133–141.
127. Jannick, P. (2006), CFC phase-out scenarios of pressurized metered dose inhalers: Current status, in: Dalby, R. N., Byron, P. R., Peart, J., Suman, J. D., and Farr, S. J., Eds., *Respiratory Drug Delivery*, 2006, Davis Healthcare International, River Grove, IL, pp. 789–792.
128. Anonymous (2006), A levalbuterol metered-dose inhaler (xopenex HFA) for asthma, *Med. Lett. Drugs Ther.*, 48, 21–22, 24.
129. Byron, P. R. (1994), Dosing reproducibility from experimental albuterol suspension metered-dose inhalers, *Pharm. Res.*, 11, 580–584.
130. Tansey, I. (1997), The technical transition to CFC-free inhalers, *Br. J. Clin. Pract. Suppl.*, 89, 22–27.
131. Stefely, J. D., Duan, D. C., Myrdal, P. B., Ross, D. L., Schultz, D. W., and Leach, C. L. (2000), Design and utility of a novel class of biocompatible excipients for HFA-based MDIs, in: Dalby, R. N., Byron, P. R., Peart, J., and Farr, S. J., Eds., *Respiratory Drug Delivery VII*, Serentec Press, Raleigh, NC, pp. 83–90.
132. Stein, S. W., and Myrdal, P. B. (2006), The relative influence of atomization and evaporation on metered dose inhaler drug delivery efficiency, *Aerosol Sci. Tech.*, 40, 335–347.
133. Leach, C. L., Davidson, P. J., and Boudreau, R. J. (1998), Improved airway targeting with the CFC-free HFA-beclomethasone metered-dose inhaler compared with CFC-beclomethasone, *Eur. Respir. J.*, 12, 1346–1353.
134. Seale, J. P., and Harrison, L. I. (1998), Effect of changing the fine particle mass of inhaled beclomethasone dipropionate on intrapulmonary deposition and pharmacokinetics, *Respir. Med.*, 92(Suppl. A), 9–15.
135. Berry, J., Kline, L. C., Hart, J. L., and Sequeira, J. (2003), Influence of the storage orientation on the aerodynamic particle size of a suspension metered dose inhaler containing propellant HFA-227, *Drug Dev. Ind. Pharm.*, 29, 631–639.

136. Anonymous (2005), Tetrafluoroethane, *Pharm. Forum.*, 31, 1672–1679.
137. Ashurst, I. C., Schultz, R. D., and Shurkus, D. D. (1996), Optimization and scale-up of a new pressure filling manufacturing process for albuterol/HFA 134a inhaler, in: Byron, P. R., Dalby, R. N., and Farr, S. J., Eds., *Respiratory Drug Delivery V*, Interpharm, Buffalo Grove, IL, pp. 294–296.
138. Wilkinson, A. (1998), The manufacture of metered dose inhalers, in: Purewal, T. S., and Grant, D. J. W., Eds., *Metered Dose Inhaler Technology*, CRC Press, Boca Raton, F.L., pp. 69–116.
139. Gupta, A., Stein, S. W., and Myrdal, P. B. (2003), Balancing ethanol cosolvent concentration with product performance in 134a-based pressurized metered dose inhalers, *J. Aerosol Med.*, 16, 167–174.
140. Tiwari, D., Goldman, D., Dixit, S., Malick, W. A., and Madan, P. L. (1998), Compatibility evaluation of metered-dose inhaler valve elastomers with tetrafluoroethane (P134a), a non-CFC propellant, *Drug Dev. Ind. Pharm.*, 24, 345–352.
141. Brouet, G., Robins, E., Hall, S., Butterworth, G., Hemy, J., and Turner, R. (2006), Developing new container closure options: A suppliers perpective, in: Dalby, R. N., Byron, P. R., Peart, J., Suman, J. D., and Farr, S. J., Eds., *Respiratory Drug Delivery 2006*, Davis Healthcare International, River Grove, IL, pp. 111–120.
142. Leone, P. (2006), Review of elastomer materials used in metering valves for pharmaceutical applications, paper presented at the International Rubber Conference IRC, Lyon, pp. 1–14.
143. Burel, S., Brouet, G., and Grandsire, T. (2004), Moisture uptake in pMDIs: The effect on valve component materials, paper presented at Drug Delivery to the Lungs XV, London, pp. 297–300.
144. Williams, R. O. III, and Hu, C. (2000), Moisture uptake and its influence on pressurized metered-dose inhalers, *Pharm. Dev. Technol.*, 5, 153–162.
145. Williams, R. O. III, and Hu, C. (2001), Influence of water on the solubility of two steroid drugs in hydrofluoroalkane (HFA) propellants, *Drug Dev. Ind. Pharm.*, 27, 71–79.
146. Ross, R. N. (1997), Loss of bronchodilator medication in priming a conventional metered dose inhaler: A cost of treating asthma, *Med. Interface*, 10, 141–146.
147. Smyth, H., Brace, G., Barbour, T., Gallion, J., Grove, J., and Hickey, A. J. (2006), Spray pattern analysis for metered dose inhalers: Effect of actuator design, *Pharm. Res.*, 23, 1591–1596.
148. Smyth, H. D., Brace, G., Barbour, T., Gallion, J., Grove, J., and Hickey, A. (2006), Actuator design variables for particle size modulation, in: Dalby, R. N., Byron, P. R., Peart, J., Suman, J. D., and Farr, S. J., Eds., *Respiratory Drug Delivery 2006*, Davis Healthcare International, River Grove, IL, pp. 857–860.
149. Peart, J., Orban, J. C., McGlynn, P., Redmon, M., Sargeant, C. M., and Byron, P. R. (2002), MDI electrostatics: Valve and formulation interactions which really make a difference, in: Dalby, R. N., Byron, P. R., Peart, J., and Farr, S. J., Eds., *Respiratory Drug Delivery XIII*, Davis Horwood International, Raleigh, NC, pp. 223–230.
150. Kwok, P. C., Glover, W., and Chan, H. K. (2005), Electrostatic charge characteristics of aerosols produced from metered dose inhalers, *J. Pharm. Sci.*, 94, 2789–2799.
151. Newman, S. P., Weisz, A. W., Talaee, N., and Clarke, S. W. (1991), Improvement of drug delivery with a breath actuated pressurised aerosol for patients with poor inhaler technique, *Thorax*, 46, 712–716.
152. Woodman, K., Bremner, P., Burgess, C., Crane, J., Pearce, N., and Beasley, R. (1993), A comparative study of the efficacy of beclomethasone dipropionate delivered from a breath activated and conventional metered dose inhaler in asthmatic patients, *Curr. Med. Res. Opin.*, 13, 61–69.

153. Terzano, C. (2001), Pressurized metered dose inhalers and add-on devices, *Pulm. Pharmacol. Ther.*, 14, 351–366.
154. Asmus, M. J., Liang, J., Coowanitwong, I., and Hochhaus, G. (2004), In vitro performance characteristics of valved holding chamber and spacer devices with a fluticasone metered-dose inhaler, *Pharmacotherapy*, 24, 159–166.
155. Goldberg, S., Einot, T., Algur, N., et al. (2002), Adrenal suppression in asthmatic children receiving low-dose inhaled budesonide: Comparison between dry powder inhaler and pressurized metered-dose inhaler attached to a spacer, *Ann. Allergy Asthma Immunol.*, 89, 566–571.
156. Dalby, R. N., Somaraju, S., Chavan, V. S., and Jarvis, D. (1998), Evaluation of aerosol drug output from the OptiChamber and AeroChamber spacers in a model system, *J. Asthma.*, 35, 173–177.
157. Newman, S. P., Millar, A. B., Lennard-Jones, T. R., Moren, F., and Clarke, S. W. (1984), Improvement of pressurised aerosol deposition with nebuhaler spacer device, *Thorax*, 39, 935–941.
158. Hindle, M., and Chrystyn, H. (1994), Relative bioavailability of salbutamol to the lung following inhalation using metered dose inhalation methods and spacer devices, *Thorax*, 49, 549–553.
159. Wildhaber, J. H., Waterer, G. W., Hall, G. L., and Summers, Q. A. (2000), Reducing electrostatic charge on spacer devices and bronchodilator response, *Br. J. Clin. Pharmacol.*, 50, 277–280.
160. Pierart, F., Wildhaber, J. H., Vrancken, I., Devadason, S. G., and Le Souef, P. N. (1999), Washing plastic spacers in household detergent reduces electrostatic charge and greatly improves delivery, *Eur. Respir. J.*, 13, 673–678.
161. Coppolo, D. P., Mitchell, J. P., and Nagel, M. W. (2006), Levalbuterol aerosol delivery with a nonelectrostatic versus a nonconducting valved holding chamber, *Respir. Care*, 51, 511–514.
162. Smyth, H. D., Beck, V. P., Williams, D., and Hickey, A. J. (2004), The influence of formulation and spacer device on the in vitro performance of solution chlorofluorocarbon-free propellant-driven metered dose inhalers, *AAPS PharmSciTech*, 5, E7.
163. Rahmatalla, M. F., Zuberbuhler, P. C., Lange, C. F., and Finlay, W. H. (2002), In vitro effect of a holding chamber on the mouth-throat deposition of QVAR (hydrofluoroalkane-beclomethasone dipropionate), *J. Aerosol Med.*, 15, 379–385.
164. Sander, N., Fusco-Walkert, S. J., Harder, J. M., and Chipps, B. E. (2006), Dose counting and the use of pressurized metered-dose inhalers: Running on empty, *Ann. Allergy Asthma Immunol.*, 97, 34–38.
165. Bradshaw, D. R. S. (2006), Developing dose counters: An appraisal based on regulator, pharma, and user needs, in: Dalby, R. N., Byron, P. R., Peart, J., Suman, J. D., and Farr, S. J., Eds., *Respiratory Drug Delivery 2006*, Davis Healthcare International, River Grove, IL, pp. 121–131.
166. Vidgren, M., Paronen, P., Vidgren, P., Vainio, P., and Nuutinen, J. (1990), In vivo evaluation of the new multiple dose powder inhaler and the rotahaler using the gamma scintigraphy, *Acta Pharm. Nord.*, 2, 3–10.
167. Smith, I. J., and Parry-Billings, M. (2003/2004), The inhalers of the future? A review of dry powder devices on the market today, *Pulm. Pharmacol. Ther.*, 16, 79–95.
168. Telko, M. J., and Hickey, A. J. (2005), Dry powder inhaler formulation, *Respir. Care*, 50, 1209–1227.
169. Clarke, M. J., Peart, J., Cagnani, S., and Byron, P. R. (2002), Adhesion of powders for inhalation: An evaluation of drug detachment from surfaces following deposition from aerosol streams, *Pharm. Res.*, 19, 322–329.

170. Ward, G. H., and Schultz, R. K. (1995), Process-induced crystallinity changes in albuterol sulfate and its effect on powder physical stability, *Pharm. Res.*, 12, 773–779.
171. Giry, K., Pean, J. M., Giraud, L., Marsas, S., Rolland, H., and Wuthrich, P. (2006), Drug/lactose co-micronization by jet milling to improve aerosolization properties of a powder for inhalation, *Int. J. Pharm*, 321, 162–166.
172. Rasenack, N., and Muller, B. W. (2004), Micron-size drug particles: Common and novel micronization techniques, *Pharm. Dev. Technol.*, 9, 1–13.
173. Steckel, H., Rasenack, N., Villax, P., and Muller, B. W. (2003), In vitro characterization of jet-milled and in-situ-micronized fluticasone-17-propionate, *Int. J. Pharm.*, 258, 65–75.
174. Edwards, D. A., Ben-Jebria, A., and Langer, R. (1998), Recent advances in pulmonary drug delivery using large, porous inhaled particles, *J. Appl. Physiol.*, 85, 379–385.
175. White, S., Bennett, D. B., and Cheu, S., et al. (2005), EXUBERA: Pharmaceutical development of a novel product for pulmonary delivery of insulin, *Diabetes Technol. Ther.*, 7, 896–906.
176. Vanbever, R., Mintzes, J. D., Wang, J., et al. (1999), Formulation and physical characterization of large porous particles for inhalation, *Pharm. Res.*, 16, 1735–1742.
177. Dunbar, C., Scheuch, G., Sommerer, K., DeLong, M., Verma, A., and Batycky, R. (2002), In vitro and in vivo dose delivery characteristics of large porous particles for inhalation, *Int. J. Pharm.*, 245, 179–189.
178. Bartus, R. T., Emerich, D., Snodgrass-Belt, P., et al. (2004), A pulmonary formulation of L-dopa enhances its effectiveness in a rat model of parkinson's disease, *J. Pharmacol. Exp. Ther.*, 310, 828–835.
179. Dellamary, L. A., Tarara, T. E., Smith, D. J., et al. (2000), Hollow porous particles in metered dose inhalers, *Pharm. Res.*, 17, 168–174.
180. Duddu, S. P., Sisk, S. A., Walter, Y. H., et al. (2002), Improved lung delivery from a passive dry powder inhaler using an engineered PulmoSphere powder, *Pharm. Res.*, 19, 689–695.
181. Pfutzner, A., Mann, A. E., and Steiner, S. S. (2002), Technosphere/Insulin—A new approach for effective delivery of human insulin via the pulmonary route, *Diabetes Technol. Ther.*, 4, 589–594.
182. Pfutzner, A., Flacke, F., Pohl, R., et al. (2003), Pilot study with technosphere/PTH (1-34)—A new approach for effective pulmonary delivery of parathyroid hormone (1-34), *Horm. Metab. Res.*, 35, 319–323.
183. Young, P. M., and Price, R. (2004), The influence of humidity on the aerosolisation of micronised and SEDS produced salbutamol sulphate, *Eur. J. Pharm. Sci.*, 22, 235–240.
184. Rehman, M., Shekunov, B. Y., York, P., et al. Optimisation of powders for pulmonary delivery using supercritical fluid technology, *Eur. J. Pharm. Sci.*, 22, 1–17.
185. Lobo, J. M., Schiavone, H., Palakodaty, S., York, P., Clark, A., and Tzannis, S. T. (2005), SCF-engineered powders for delivery of budesonide from passive DPI devices, *J. Pharm. Sci.*, 94, 2276–2288.
186. Schiavone, H., Palakodaty, S., Clark, A., York, P., and Tzannis, S. T. (2004), Evaluation of SCF-engineered particle-based lactose blends in passive dry powder inhalers, *Int. J. Pharm.*, 281, 55–66.
187. Pandey, R., Sharma, A., Zahoor, A., Sharma, S., Khuller, G. K., and Prasad, B. (2003), Poly (DL-lactide-*co*-glycolide) nanoparticle-based inhalable sustained drug delivery system for experimental tuberculosis, *J. Antimicrob. Chemother.*, 52, 981–986.
188. Dickinson, P. A., Howells, S. W., and Kellaway, I. W. (2001), Novel nanoparticles for pulmonary drug administration, *J. Drug Target.*, 9, 295–302.

189. Sham, J. O., Zhang, Y., Finlay, W. H., Roa, W. H., and Lobenberg, R. (2004), Formulation and characterization of spray-dried powders containing nanoparticles for aerosol delivery to the lung, *Int. J. Pharm.*, 269, 457–467.
190. Begat, P., Morton, D. A., Staniforth, J. N., and Price, R. (2004), The cohesive-adhesive balances in dry powder inhaler formulations I: Direct quantification by atomic force microscopy, *Pharm. Res.*, 21, 1591–1597.
191. Davies, M., Brindley, A., Chen, X., et al. (2005), Characterization of drug particle surface energetics and Young's modulus by atomic force microscopy and inverse gas chromatography, *Pharm. Res.*, 22, 1158–1166.
192. Larhrib, H., Zeng, X. M., Martin, G. P., Marriott, C., and Pritchard, J. (1999), The use of different grades of lactose as a carrier for aerosolised salbutamol sulphate, *Int. J. Pharm.*, 191, 1–14.
193. Zeng, X. M., Pandhal, K. H., and Martin, G. P. (2000), The influence of lactose carrier on the content homogeneity and dispersibility of beclomethasone dipropionate from dry powder aerosols, *Int. J. Pharm.*, 197, 41–52.
194. Lucas, P., Anderson, K., and Staniforth, J. N. (1998), Protein deposition from dry powder inhalers: Fine particle multiplets as performance modifiers, *Pharm. Res.*, 15, 562–569.
195. Begat, P., Morton, D. A., Staniforth, J. N., and Price, R. (2004), The cohesive-adhesive balances in dry powder inhaler formulations II: Influence on fine particle delivery characteristics, *Pharm. Res.*, 21, 1826–1833.
196. Jones, M., and Price, R. (2006), The influence of fine excipient particles on the performance of carrier-based dry powder inhalation formulations, *Pharm. Res.*, 23, 1665–1674.
197. Ming Zeng, X., Martin, G. P., Marriott, C., and Pritchard, J. (2001), The use of lactose recrystallised from carbopol gels as a carrier for aerosolised salbutamol sulphate, *Eur. J. Pharm. Biopharm.*, 51, 55–62.
198. Flament, M. P., Leterme, P., and Gayot, A. (2004), The influence of carrier roughness on adhesion, content uniformity and the in vitro deposition of terbutaline sulphate from dry powder inhalers, *Int. J. Pharm.*, 275, 201–209.
199. Zeng, X. M., Martin, G. P., Marriott, C., and Pritchard, J. (2001), Lactose as a carrier in dry powder formulations: The influence of surface characteristics on drug delivery, *J. Pharm. Sci.*, 90, 1424–1434.
200. El-Sabawi, D., Price, R., Edge, S., and Young, P. M. (2006), Novel temperature controlled surface dissolution of excipient particles for carrier based dry powder inhaler formulations, *Drug Dev. Ind. Pharm.*, 32, 243–251.
201. Steckel, H., and Bolzen, N. (2004), Alternative sugars as potential carriers for dry powder inhalations, *Int. J. Pharm.*, 270, 297–306.
202. Lucas, P., Anderson, K., Potter, U. J., and Staniforth, J. N. (1999), Enhancement of small particle size dry powder aerosol formulations using an ultra low density additive, *Pharm. Res.*, 16, 1643–1647.
203. Hindle, M., Jashnani, R. N., and Byron, P. R. (1994), Dose emissions from marketed dry powder inhalers: Influence of flow, volume and environment, in: Byron, P. R., Dalby, R. N., and Farr, S. J., Eds., *Respiratory Drug Delivery IV*, Interpharm, Buffalo Grove, IL, pp. 137–142.
204. Borgstrom, L., Asking, L., and Lipniunas, P. (2005), An in vivo and in vitro comparison of two powder inhalers following storage at hot/humid conditions, *J. Aerosol Med.*, 18, 304–310.
205. Price, R., Young, P. M., Edge, S., and Staniforth, J. N. (2002), The influence of relative humidity on particulate interactions in carrier-based dry powder inhaler formulations, *Int. J. Pharm.*, 246, 47–59.

206. Young, P. M., Price, R., Tobyn, M. J., Buttrum, M., and Dey, F. (2003), Investigation into the effect of humidity on drug-drug interactions using the atomic force microscope, *J. Pharm. Sci.*, 92, 815–822.
207. Borgstrom, L., Asking, L., and Thorsson, L. (2005), Idealhalers or realhalers? A comparison of diskus and turbuhaler, *Int. J. Clin. Pract.*, 59, 1488–1495.
208. Fuller, R. (1995), The diskus: A new multi-dose powder device—Efficacy and comparison with turbuhaler, *J. Aerosol Med.*, 8(Suppl. 2), S11–17.
209. Bennett, F. S., Carter, P. A., Rowley, G., and Dandiker, Y. (1999), Modification of electrostatic charge on inhaled carrier lactose particles by addition of fine particles, *Drug Dev. Ind. Pharm.*, 25, 99–103.
210. Byron, P. R., Peart, J., and Staniforth, J. N. (1997), Aerosol electrostatics. I: Properties of fine powders before and after aerosolization by dry powder inhalers, *Pharm. Res.*, 14, 698–705.
211. Carter, P. A., Rowley, G., Fletcher, E. J., and Stylianopoulos, V. (1998), Measurement of electrostatic charge decay in pharmaceutical powders and polymer materials used in dry powder inhaler devices, *Drug Dev. Ind. Pharm.*, 24, 1083–1088.
212. Murtomaa, M., Mellin, V., Harjunen, P., Lankinen, T., Laine, E., and Lehto, V. P. (2004), Effect of particle morphology on the triboelectrification in dry powder inhalers, *Int. J. Pharm.*, 282, 107–114.
213. Taylor, A., and Gustafsson, P. (2005), Do all dry powder inhalers show the same pharmaceutical performance? *Int. J. Clin. Pract. Suppl.*, 59, 7–12.
214. Cegla, U. H. (2004), Pressure and inspiratory flow characteristics of dry powder inhalers, *Respir. Med.*, 98(Suppl. A), S22–28.
215. Hindle, M., and Byron, P. R. (1995), Dose emissions from marketed dry powder inhalers, *Int. J. Pharm.*, 116, 169–177.
216. Hindle, M., and Byron, P. R. (1996), Impaction and impingement techniques for powder inhalers—Comparisons, problems and validation, in: Byron, P. R., Dalby, R. N., and Farr, S. J., Eds., *Respiratory Drug Delivery V*, Interpharm, Buffalo Grove, IL, pp. 263–272.
217. Pitcairn, G. R., Lim, J., Hollingworth, A., and Newman, S. P. (1997), Scintigraphic assessment of drug delivery from the ultrahaler dry powder inhaler, *J. Aerosol Med.*, 10, 295–306.
218. Pedersen, S., Hansen, O. R., and Fuglsang, G. (1990), Influence of inspiratory flow rate upon the effect of a turbuhaler, *Arch. Dis. Child.*, 65, 308–310.
219. Tukiainen, H., and Terho, E. O. (1985), Comparison of inhaled salbutamol powder and aerosol in asthmatic patients with low peak expiratory flow level, *Eur. J. Clin. Pharmacol.*, 27, 645–647.
220. Borgstrom, L. (1994), Deposition patterns with turbuhaler, *J. Aerosol Med.*, 7, S49–53.
221. Finlay, W. H., and Gehmlich, M. G. (2000), Inertial sizing of aerosol inhaled from two dry powder inhalers with realistic breath patterns versus constant flow rates, *Int. J. Pharm.*, 210, 83–95.
222. Borgstrom, L., Bengtsson, T., Derom, E., and Pauwels, R. (2000), Variability in lung deposition of inhaled drug, within and between asthmatic patients, with a pMDI and a dry powder inhaler, turbuhaler, *Int. J. Pharm.*, 193, 227–230.
223. Nielsen, K. G., Skov, M., Klug, B., Ifversen, M., and Bisgaard, H. (1997), Flow-dependent effect of formoterol dry-powder inhaled from the aerolizer, *Eur. Respir. J.*, 10, 2105–2109.
224. Koskela, T., Malmstrom, K., Sairanen, U., Peltola, S., Keski-Karhu, J., and Silvasti, M. (2000), Efficacy of salbutamol via easyhaler unaffected by low inspiratory flow, *Respir. Med.*, 94, 1229–1233.

225. Newhouse, M. T., Nantel, N. P., Chambers, C. B., Pratt, B., and Parry-Billings, M. (1999), Clickhaler (a novel dry powder inhaler) provides similar bronchodilation to pressurized metered-dose inhaler, even at low flow rates, *Chest*, 115, 952–956.
226. Meakin, B. J., Ganderton, D., Panza, I., and Ventura, P. (1998), The effect of flow rate on drug delivery from the pulvinal, a high-resistance dry powder inhaler, *J. Aerosol Med.*, 11, 143–152.
227. Newman, S. P., Moren, F., Trofast, E., Talaee, N., and Clarke, S. W. (1989), Deposition and clinical efficacy of terbutaline sulphate from turbuhaler, a new multi-dose powder inhaler, *Eur. Respir. J.*, 2, 247–252.
228. Kohler, D. (2004), The novolizer: Overcoming inherent problems of dry powder inhalers, *Respir. Med.*, 98(Suppl. A), S17–21.
229. O'Connor, B. J. (2004), The ideal inhaler: Design and characteristics to improve outcomes, *Respir. Med.*, 98(Suppl. A), S10–16.
230. Lim, J. G., Shah, B., Rohatagi, S., and Bell, A. (2006), Development of a dry powder inhaler, the ultrahaler, containing triamcinolone acetonide using in vitro–in vivo relationships, *Am. J. Ther.*, 13, 32–42.
231. McCormack, P. L., and Plosker, G. L. (2006), Inhaled mometasone furoate: A review of its use in persistent asthma in adults and adolescents, *Drugs*, 66, 1151–1168.
232. Pitcairn, G. R., Lankinen, T., Seppala, O. P., and Newman, S. P. (2000), Pulmonary drug delivery from the taifun dry powder inhaler is relatively independent of the patient's inspiratory effort, *J. Aerosol Med.*, 13, 97–104.
233. Keating, G. M., and Faulds, D. (2002), Airmax: A multi-dose dry powder inhaler, *Drugs*, 62, 1887–1895; discussion 1896–1897.
234. Iida, K., Leuenberger, H., Fueg, L. M., Muller-Walz, R., Okamoto, H., and Danjo, K. (2000), Effect of mixing of fine carrier particles on dry powder inhalation property of salbutamol sulfate (SS), *Yakugaku Zasshi*, 120, 113–119.
235. Patton, J. S., Fishburn, C. S., and Weers, J. G. (2004), The lungs as a portal of entry for systemic drug delivery, *Proc. Am. Thorac. Soc.*, 1, 338–344.
236. Patton, J. S. (2005), Unlocking the opportunity of tight glycaemic control: Innovative delivery of insulin via the lung, *Diabetes Obes. Metab.*, 7(Suppl. 1), S5–8.
237. Rosenstock, J., Cappelleri, J. C., Bolinder, B., and Gerber, R. A. (2004), Patient satisfaction and glycemic control after 1 year with inhaled insulin (exubera) in patients with type 1 or type 2 diabetes, *Diabetes Care*, 27, 1318–1323.
238. Bellary, S., and Barnett, A. H. (2006), Inhaled insulin: New technology, new possibilities, *Int. J. Clin. Pract.*, 60, 728–734.
239. Dunn, C., and Curran, M. P. (2006), Inhaled human insulin (exubera): A review of its use in adult patients with diabetes mellitus, *Drugs*, 66, 1013–1032.
240. Patton, J. S., Bukar, J. G., and Eldon, M. A. (2004), Clinical pharmacokinetics and pharmacodynamics of inhaled insulin, *Clin. Pharmacokinet.*, 43, 781–801.
241. Quattrin, T., Belanger, A., Bohannon, N. J., Schwartz, S. L., and Exubera Phase III Study Group (2004), Efficacy and safety of inhaled insulin (exubera) compared with subcutaneous insulin therapy in patients with type 1 diabetes: Results of a 6-month, randomized, comparative trial, *Diabetes Care*, 27, 2622–2627.
242. Barnett, A. H., Dreyer, M., Lange, P., and Serdarevic-Pehar, M. (2006), An open, randomized, parallel-group study to compare the efficacy and safety profile of inhaled human insulin (exubera) with glibenclamide as adjunctive therapy in patients with type 2 diabetes poorly controlled on metformin, *Diabetes Care*, 29, 1818–1825.
243. Rosenstock, J., Zinman, B., Murphy, L. J., et al. (2005), Inhaled insulin improves glycemic control when substituted for or added to oral combination therapy in type 2 diabetes: A randomized, controlled trial, *Ann. Intern. Med.*, 143, 549–558.

244. DeFronzo, R. A., Bergenstal, R. M., Cefalu, W. T., et al. (2005), Efficacy of inhaled insulin in patients with type 2 diabetes not controlled with diet and exercise: A 12-week, randomized, comparative trial, *Diabetes Care*, 28, 1922–1928.
245. Rave, K., Bott, S., Heinemann, L., et al. (2005), Time-action profile of inhaled insulin in comparison with subcutaneously injected insulin lispro and regular human insulin, *Diabetes Care*, 28, 1077–1082.
246. Davidson, M. B., Mehta, A. E., and Siraj, E. S. (2006), Inhaled human insulin: An inspiration for patients with diabetes mellitus? *Cleve. Clin. J. Med.*, 73, 569–578.
247. An, B., and Reinhardt, R. R. (2003), Effects of different durations of breath holding after inhalation of insulin using the AERx insulin diabetes management system, *Clin. Ther.*, 25, 2233–2244.
248. Boyd, B., Noymer, P., Liu, K., et al. (2004), Effect of gender and device mouthpiece shape on bolus insulin aerosol delivery using the AERx pulmonary delivery system, *Pharm. Res.*, 21, 1776–1782.
249. Cramer, J. A., Okikawa, J., Bellaire, S., and Clauson, P. (2004), Compliance with inhaled insulin treatment using the AERx iDMS insulin diabetes management system, *Diabetes Technol. Ther.*, 6, 800–807.
250. Farr, S. J., McElduff, A., Mather, L. E., et al. (2000), Pulmonary insulin administration using the AERx system: Physiological and physicochemical factors influencing insulin effectiveness in healthy fasting subjects, *Diabetes Technol. Ther.*, 2, 185–197.
251. Thipphawong, J., Otulana, B., Clauson, P., Okikawa, J., and Farr, S. J. (2002), Pulmonary insulin administration using the AERx insulin diabetes system, *Diabetes Technol. Ther.*, 4, 499–504.
252. Ferron, G. A., Kerrebijn, K. F., and Weber, J. (1976), Properties of aerosols produced with three nebulizers, *Am. Rev. Respir. Dis.*, 114, 899–908.
253. Nikander, K., Turpeinen, M., and Wollmer, P. (1999), The conventional ultrasonic nebulizer proved inefficient in nebulizing a suspension, *J. Aerosol Med.*, 12, 47–53.
254. Rau, J. L. (2002), Design principles of liquid nebulization devices currently in use, *Respir. Care*, 47, 1257–1275; discussion 1275–1278.
255. Nerbrink, O., and Dahlback, M. (1994), Basic nebulizer function, *J. Aerosol Med.*, 7, S7–11.
256. Nerbrink, O., Dahlback, M., and Hansson, H. C. (1994), Why do medical nebulizers differ in their output and particle size characteristics? *J. Aerosol Med.*, 7, 259–276.
257. Newnham, D. M., and Lipworth, B. J. (1994), Nebuliser performance, pharmacokinetics, airways and systemic effects of salbutamol given via a novel nebuliser delivery system ("ventstream"), *Thorax*, 49, 762–770.
258. Leung, K., Louca, E., and Coates, A. L. (2004), Comparison of breath-enhanced to breath-actuated nebulizers for rate, consistency, and efficiency, *Chest*, 126, 1619–1627.
259. Barry, P. W., and O'Callaghan, C. (1999), An in vitro analysis of the output of salbutamol from different nebulizers, *Eur. Respir. J.*, 13, 1164–1169.
260. O'Callaghan, C., White, J., Jackson, J., Barry, P. W., and Kantar, A. (2005), Delivery of nebulized budesonide is affected by nebulizer type and breathing pattern, *J. Pharm. Pharmacol.*, 57, 787–790.
261. Rau, J. L., Ari, A., and Restrepo, R. D. (2004), Performance comparison of nebulizer designs: Constant-output, breath-enhanced, and dosimetric, *Respir. Care*, 49, 174–179.
262. Raabe, O. G., Wong, T. M., Wong, G. B., Roxburgh, J. W., Piper, S. D., and Lee, J. I. (1998), Continuous nebulization therapy for asthma with aerosols of beta2 agonists, *Ann. Allergy Asthma Immunol.*, 80, 499–508.

263. Dolovich, M. B. (2002), Assessing nebulizer performance, *Respir. Care*, 47, 1290–1301; discussion 1301–1304.

264. Le Brun, P. P., de Boer, A. H., Heijerman, H. G., and Frijlink, H. W. (2000), A review of the technical aspects of drug nebulization, *Pharm. World Sci.*, 22, 75–81.

265. Byrne, N. M., Keavey, P. M., Perry, J. D., Gould, F. K., and Spencer, D. A. (2003), Comparison of lung deposition of colomycin using the HaloLite and the pari LC plus nebulisers in patients with cystic fibrosis, *Arch. Dis. Child.*, 88, 715–718.

266. Denyer, J., Nikander, K., and Smith N. J. (2004), Adaptive aerosol delivery (AAD) technology, *Expert Opin. Drug Deliv.*, 1, 165–176.

267. Nikander, K., Arheden, L., Denyer, J., and Cobos, N. (2003), Parents' adherence with nebulizer treatment of their children when using an adaptive aerosol delivery (AAD) system, *J. Aerosol Med.*, 16, 273–281.

268. Knoch, M., and Keller, M. (2005), The customised electronic nebuliser: A new category of liquid aerosol drug delivery systems, *Expert Opin. Drug Deliv.*, 2, 377–390.

269. Kim, D., Mudaliar, S., Chinnapongse, S., et al. (2003), Dose-response relationships of inhaled insulin delivered via the aerodose insulin inhaler and subcutaneously injected insulin in patients with type 2 diabetes, *Diabetes Care*, 26, 2842–2847.

270. Lipworth, B. J., Sims, E. J., Taylor, K., Cockburn, W., and Fishman, R. (2005), Dose-response to salbutamol via a novel palm sized nebuliser (aerodose inhaler), conventional nebuliser (pari LC plus) and metered dose inhaler (ventolin evohaler) in moderate to severe asthmatics, *Br. J. Clin. Pharmacol.*, 59, 5–13.

271. Hindle, M. (2004), Soft mist inhalers: A review of current technology, *The Drug Delivery Companies Report*, 13, 31–34.

272. Geller, D. E. (2002), New liquid aerosol generation devices: Systems that force pressurized liquids through nozzles, *Respir. Care*, 47, 1392–1404; discussion 1404–1405.

273. Patel, K. R., Pavia, D., Lowe, L., and Spiteri, M. (2006), Inhaled ethanolic and aqueous solutions via respimat soft mist inhaler are well-tolerated in asthma patients, *Respiration*, 73, 434–440.

274. Voshaar, T., Hausen, T., Kardos, P., et al. (2005), Inhalation therapy with respimat soft inhaler in patients with COPD and asthma, *Pneumologie*, 59, 25–32.

275. Kassner, F., Hodder, R., and Bateman, E. D. (2004), A review of ipratropium bromide/fenoterol hydrobromide (berodual) delivered via respimat soft mist inhaler in patients with asthma and chronic obstructive pulmonary disease, *Drugs*, 64, 1671–1682.

276. Pitcairn, G., Reader, S., Pavia, D., and Newman, S. (2005), Deposition of corticosteroid aerosol in the human lung by respimat soft mist inhaler compared to deposition by metered dose inhaler or by turbuhaler dry powder inhaler, *J. Aerosol Med.*, 18, 264–272.

277. Vincken, W., Dewberry, H., and Moonen, D. (2003), Fenoterol delivery by respimat soft mist inhaler versus CFC metered dose inhaler: Cumulative dose-response study in asthma patients, *J. Asthma*, 40, 721–730.

278. Newman, S. P., Brown, J., Steed, K. P., Reader, S. J., and Kladders, H. (1998), Lung deposition of fenoterol and flunisolide delivered using a novel device for inhaled medicines: Comparison of RESPIMAT with conventional metered-dose inhalers with and without spacer devices, *Chest*, 113, 957–963.

279. Newman, S. P., Steed, K. P., Reader, S. J., Hooper, G., and Zierenberg, B. (1996), Efficient delivery to the lungs of flunisolide aerosol from a new portable hand-held multidose nebulizer, *J. Pharm. Sci.*, 85, 960–964.

280. Schurmann, W., Schmidtmann, S., Moroni, P., Massey, D., and Qidan, M. (2005), Respimat soft mist inhaler versus hydrofluoroalkane metered dose inhaler: Patient preference and satisfaction, *Treat. Respir. Med.*, 4, 53–61.
281. Pavia, D., and Moonen, D. (1999), Preliminary data from phase II studies with respimat, a propellant-free soft mist inhaler, *J. Aerosol Med.*, 12(Suppl. 1), S33–39.
282. Dalby, R., Spallek, M., and Voshaar, T. (2004), A review of the development of respimat soft mist inhaler, *Int. J. Pharm.*, 283, 1–9.
283. Hochrainer, D., Holz, H., Kreher, C., Scaffidi, L., Spallek, M., and Wachtel, H. (2005), Comparison of the aerosol velocity and spray duration of respimat soft mist inhaler and pressurized metered dose inhalers, *J. Aerosol Med.*, 18, 273–282.
284. Zierenberg, B. (1999), Optimizing the in vitro performance of respimat, *J. Aerosol Med.*, 12(Suppl. 1), S19–24.
285. Farr, S. J., Warren, S. J., Lloyd, P., et al. (2000), Comparison of in vitro and in vivo efficiencies of a novel unit-dose liquid aerosol generator and a pressurized metered dose inhaler, *Int. J. Pharm.*, 198, 63–70.
286. Hermansen, K., Ronnemaa, T., Petersen, A. H., Bellaire, S., and Adamson, U. (2004), Intensive therapy with inhaled insulin via the AERx insulin diabetes management system: A 12-week proof-of-concept trial in patients with type 2 diabetes, *Diabetes Care*, 27, 162–167.
287. Geller, D., Thipphawong, J., Otulana, B., et al. (2003), Bolus inhalation of rhDNase with the AERx system in subjects with cystic fibrosis, *J. Aerosol Med.*, 16, 175–182.
288. Zimlich, W. C., Ding, J. Y., Busick, D. R., et al. (2000), The development of a novel electrohydrodynamic pulmonary drug delivery device, in Dalby, R. N., Byron, P. R., Peart, J., and Farr, S. J., Eds., *Respiratory Drug Delivery VII*, Serentec, Raleigh, NC, pp. 241–246.
289. Shen, X., Hindle, M., and Byron P. R. (2004), Effect of energy on propylene glycol aerosols using the capillary aerosol generator, *Int. J. Pharm.*, 275, 249–258.
290. Rabinowitz, J. D., Lloyd, P. M., Munzar, P., et al. (2006), Ultra-fast absorption of amorphous pure drug aerosols via deep lung inhalation, *J. Pharm. Sci*, 95, 2438–2451.
291. Rabinowitz, J. D., Wensley, M., Lloyd, P., et al. (2004), Fast onset medications through thermally generated aerosols, *J. Pharmacol. Exp. Ther.*, 309, 769–775.

# OCULAR DRUG DELIVERY

Ilva D. Rupenthal and Raid G. Alany
*The University of Auckland, Auckland, New Zealand*

**Contents**

1 Introduction
2 Challenges in Ocular Drug Delivery
  2.1 Anatomical and Physiological Considerations
  2.2 Pharmacokinetic Considerations
  2.3 Formulation Considerations
3 Formulation Approaches to Improve Ocular Bioavailability
  3.1 Conventional Dosage Forms
  3.2 Polymeric Delivery Systems
  3.3 Colloidal Delivery Systems
  3.4 Other Delivery Approaches
4 Conclusion
  References

## 1 INTRODUCTION

Due to the accessibility of the eye surface, topical administration of ophthalmic medications is the most common method for treating conditions affecting the exterior eye surface. However, the unique anatomy and physiology of the eye renders it difficult to achieve an effective drug concentration at the target site. Therefore, efficient delivery of a drug past the protective ocular barriers accompanied with minimization of its systemic side effects remains a major challenge.

Conventional eye drops currently account for more than 90% of the marketed ophthalmic formulations [1]. However, after instillation of an eye drop, typically less than 5% of the applied drug penetrates the cornea and reaches the intraocular

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

tissues. This is due to the rapid and extensive precorneal loss caused by drainage and high tear fluid turnover.

As a consequence, the typical corneal contact time is limited to 1–2 min and the ocular bioavailability is usually less than 10% [2]. Furthermore, there is an initial peak dose of the drug, which is usually higher than that needed for a therapeutic effect, followed by a sharp drop-off in concentration to subtherapeutic levels.

Various ocular delivery systems, such a ointments, suspensions, micro- and nano-carriers, and liposomes, have been investigated during the past two decades pursuing two main strategies: to increase the corneal permeability and to prolong the contact time on the ocular surface [3].

On the other hand, the most efficient method for drug delivery to the posterior chamber of the eye so far has been intravitreal injection. This chapter focuses on the topical application of drugs to the surface of the eye and discusses the most recent formulation approaches in this area.

## 2 CHALLENGES IN OCULAR DRUG DELIVERY

### 2.1 Anatomical and Physiological Considerations

In order to research and develop an effective ophthalmic delivery system, a good understanding of the anatomy and physiology of the eye (the globe) is necessary. Figure 1 shows a cross section through the human eye. This chapter will mainly focus on the precorneal area and the transport barriers present in the eye.

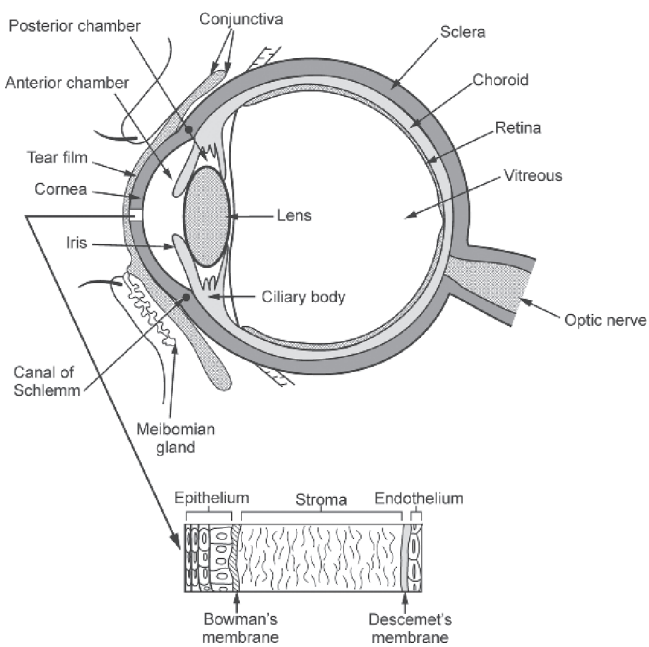


**FIGURE 1** Cross section of the human eye and cornea.

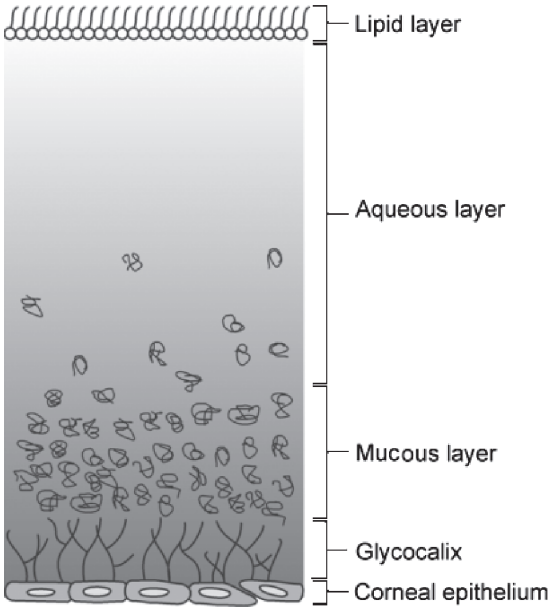


**FIGURE 2** Structure of the precorneal tear film.

### *Precorneal Area*

*Precorneal Tear Film* Corneal transparency and good visual function require a uniform eye surface. This is achieved by the tear film, which covers and lubricates the cornea and the external globe. It is about 7–8 μm thick and is the first structure encountered by topically applied drugs. The trilaminar structure of the tear film is shown in Figure 2.

Attached to the glycocalix of the corneal/conjunctival surface is a mucous layer, which consists mainly of glycoproteins. This layer is produced by the conjunctival goblet cells and the lacrimal gland. It plays an important role in the stability of the tear film as well as in the wetting of the corneal and conjunctival epithelium. The middle aqueous layer constitutes about 98% of the tear film and is mostly secreted by the main and accessory lacrimal glands [4]. It is composed of water, electrolytes, and various proteins such as lipocalin, lysozyme, and lactoferrin [5–8]. The outermost lipid layer is derived from the Meibomian and sebaceous Zeiss glands and prevents the evaporation of the tear fluid. It consists of sterol esters, triacylglycerols, and phospholipids and is spread over the aqueous layer during blinking.

*Nasolacrimal Drainage System* Figure 3 illustrates the nasolacrimal drainage system. The lacrimal gland, which is situated in the superior temporal angle of the orbit, is responsible for most of the tear fluid secretion. Secreted fluid is spread over the surface of the cornea during blinking and ends up in the puncta when the upper eye lid approaches the lower lid. The blinking process creates a suction mechanism which results in tears flowing through the lacrimal canaliculi into the lacrimal sac. Fluid from the lacrimal sac then drains into the 12-mm-long nasolacrimal duct, which empties into the inferior nasal passage. This passage is a highly vascular area

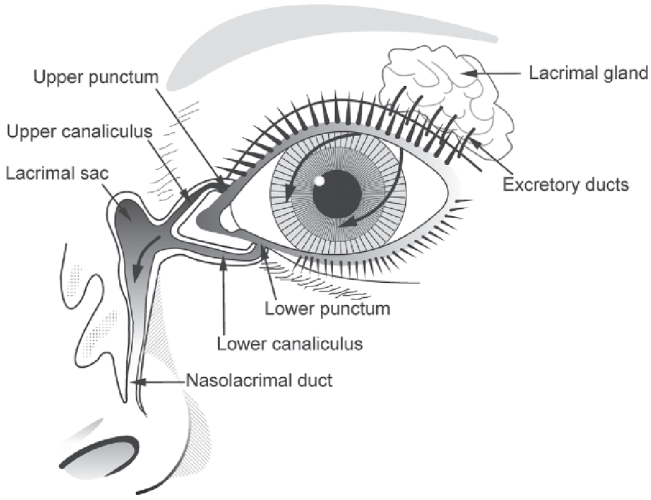


**FIGURE 3** Schematic of the nasolacrimal drainage system.

and is responsible for most of the systemic drug absorption and subsequent systemic side effects.

The cul-de-sac normally holds 7–9 μL of tear fluid, with the normal tear flow rate being 1.2–1.5 μL/min [4]. The loss from the precorneal area by drainage, tear fluid turnover, and noncorneal absorption plays an important role in determining the ocular bioavailability of a drug. As the drainage rate is much faster than the ocular absorption rate, most of the topically applied drug is eliminated from the precorneal area within the first minute [4].

Tear production can be divided into basal, reflex, and emotional tearing [9]. Reflex tearing can be induced by many pharmaceutical/formulation factors, including the drug itself as well as pH and tonicity of the ocular dosage form.

***Transport Barriers in the Eye*** Topical administration is the most common route for ocular drug delivery. Consequently, the cornea, conjunctiva, and sclera form the most essential barriers for drug penetration into the intraocular tissues.

*Cornea* The cornea is an important mechanical barrier protecting the intraocular tissues. It is considered to be the main pathway for ocular penetration of topically applied drugs. However, due to its unique structure, with the hydrophilic stroma sandwiched between the highly lipophilic epithelium and the less lipophilic endothelium, the penetration of compounds through the cornea depends on their *n*-octanol–water partition coefficient. Only drugs with a partition coefficient between 10 and 100 that show both lipid- and water-soluble properties can readily pass through the cornea.

The cornea is composed of five layers (see Figure 1): epithelium, Bowman's membrane stroma, Descemet's membrane, and endothelium.

The epithelium is the outermost layer and consists of five to six cell layers. These can be subdivided into one to two outermost layers of flattened superficial cells with microvilli on their anterior surface enhancing the cohesion and stability of the tear

film [10], two to three layers of polygonal wing cells, and a single layer of basal columnar cells, allowing for minimal paracellular transport. The superficial cells adhere to one another via desmosomes and the cells are encircled by tight junctions [11]. Due to the nature of the tight junctions, the epithelium represents the rate-limiting barrier for hydrophilic compounds.

Bowman's membrane is composed of a layer of collagen fibers which form a relatively tight and impermeable barrier against microorganisms and therefore protect the stroma.

The stroma makes up approximately 90% of the corneal thickness and is mainly composed of hydrated collagen fibrils. It is highly hydrophilic, porous, and can be considered as an open structure, as it allows free passage of hydrophilic substances with a molecular weight below 500,000 Da but acts as a diffusion barrier to all lipophilic drugs [4, 12].

Descemet's membrane is the basement membrane of the endothelial cells. It is comprised of collagen fibers arranged in a hexagonal pattern and embedded in a matrix.

The endothelium is a single layer of flattened polygonal cells with microvilli which increase the surface area for removal of waste and absorption of nutrients [12]. It plays an important role in the maintenance of corneal hydration and transparency via active ion and fluid transport mechanisms.

Since the cornea exhibits hydrophilic as well as lipophilic characteristics, it represents an effective barrier for diffusion of both hydrophilic and lipophilic substances [3, 13].

*Conjunctiva* The conjunctiva is a thin and vascular mucous membrane consisting of two to three layers of epithelial cells overlying a loose, highly vascular connective tissue. The tight junctions present on the apical surface of the epithelium act as the main barrier for drug penetration (molecules > 20,000 Da) across the tissue, although not as tight as the corneal epithelium, which is impermeable to molecules larger than 5000 Da [14, 15]. The conjunctiva covers the anterior surface of the globe (bulbar conjunctiva), with the exception of the cornea, and is folded at the fornix (fornix conjunctiva) to form the palpebral conjunctiva, which lines the inner surface of the eyelids. The bulbar conjunctiva represents the first barrier for permeation of topically applied drugs via the noncorneal route [16].

*Sclera* The sclera is the outermost firm coat of the eye that serves as a protective barrier for the sensitive inner parts. It is composed of the same type of collagen fibers as the corneal stroma. However, the fibers are arranged in an irregular network rather than a lattice pattern, which makes the tissue appear opaque compared to the transparent cornea. The white sclera constitutes the posterior five-sixths of the globe, whereas the transparent cornea comprises the anterior one-sixth [17].

*Iris–Ciliary Body* The ciliary body comprises the ciliary muscle, which mainly enables accommodation, and the ciliary processes, which produce the aqueous humor.

The iris separates the anterior and posterior chambers and consists of the pigmented epithelial cell layer, the iridial sphincter and dilator muscles, and the stroma. The amount of melanin present in the stroma determines the color of the iris: few

melanocytes exhibit blue, grey, or green, while many melanocytes are responsible for the brown appearance of the iris. There is often a considerable quantitative difference in drug response between light and heavily pigmented eyes [18]. The binding of drugs with melanin can decrease the aqueous humor concentration of free drug and is therefore likely to reduce the pharmacological response [19].

*Lens* The lens is the transparent biconvex structure situated behind the iris and in front of the vitreous. It plays an important role in the visual function of the eye and also enables accommodation together with the ciliary muscle. The lens is made up of slightly more than 30% protein (water-soluble crystallins) and therefore has the highest protein content of all tissues in the body [20]. The lens receives its nutrients from the aqueous humor and its transparency depends on the geometry of the lens fibres.

*Blood–Ocular Barriers* The blood–ocular barriers can be divided into the blood–aqueous barrier and the blood–retinal barrier.

The blood–aqueous barrier is located in the anterior part of the eye and is formed by the endothelial cells of the blood vessels in the iris and the nonpigmented cell layer of the ciliary epithelium [21]. It regulates the solute exchange between the blood and the intraocular fluid, preventing unspecific passage of solutes that could influence the transparency of the ocular tissues. The outward movement into the systemic blood circulation is less restricted, allowing especially small and lipophilic drug molecules to enter the uveal blood circulation [22]. These molecules are consequently removed more rapidly from the anterior chamber than larger, hydrophilic molecules, which are eliminated by the aqueous humor turnover only [23].

The blood–retinal barrier can be found in the posterior part of the eye. It prevents toxic molecules, plasma components, and water from entering the retina. It also forms a barrier for passage of systemically administered drugs into the vitreous, typically resulting in only 1–2% of the drug's plasma concentration in the intraocular tissues [24].

## 2.2 Pharmacokinetic Considerations

After topical application of an ophthalmic solution, the solution is instantly mixed with the tear fluid and then spread over the eye surface. However, various precorneal factors such as the drainage of the instilled solution, induced lacrimation, normal tear turnover, noncorneal absorption, drug metabolism, and enzymatic degradation limit the ocular absorption by shortening the contact time of the applied drug with the corneal surface [25]. As a result, typically less than 10% of the instilled dose is delivered into the intraocular tissues, whereas the rest is absorbed into the systemic circulation, leading to various side effects [3, 25, 26]. A summary of the drug deposition model in the eye after topical application as described by Lee and Robinson [27] is given below.

Upon instillation, the topically applied drug solution is instantly diluted by the resident tears, resulting in a significant decrease in the concentration gradient (driving force) and hence in the reduction of the transcorneal flux. Drainage of lacrimal fluid towards the nasolacrimal sac during blinking leads to a rapid elimination of the ocular solution via the canaliculi.

Any drug remaining on the ocular surface for a sufficient period of time can be absorbed into the anterior chamber via either the corneal or the conjunctival–scleral route.

Corneal absorption is considered to be the major penetration pathway for topically applied drugs. There are two mechanisms for absorption across the corneal epithelium, namely transcellular and paracellular diffusion [28]. Lipophilic drugs prefer the transcellular route while hydrophilic drugs penetrate primarily via the paracellular route. Transcorneal transport includes simple diffusion, facilitated diffusion, active transport, and endocytosis. Transport along the paracellular route is passive and is only limited by the pore size and the charge of the intracellular spaces. Only relatively small molecules can permeate through the pores. Negatively charged molecules permeate at a slower rate than positively charged and neutral ones [29]. In addition, a positive charge may also decrease the permeation due to the possible ionic interaction with the negatively charged carboxylic acid residues of the tight junction proteins [30].

The noncorneal route of absorption via the conjunctiva and the sclera is usually nonproductive, as most of the drug reaches the systemic circulation before gaining access to the intraocular tissues. As the surface area of the conjunctiva is much larger than that of the cornea and with the highly vascularized conjunctiva being more permeable, especially to larger hydrophilic molecules, drug loss through this route of absorption may be significant. Both transconjunctival and transnasal absorption via the nasolacrimal duct are generally undesirable, not only because of the loss of drug into the systemic circulation, but also because of the possible side effects [1].

### 2.3 Formulation Considerations

Irritation of the eye following the use of an ocular delivery system can be induced by a number of factors, including the instilled volume, the pH, and the osmolality of the formulation, as well as by the drug itself [31]. All these factors may induce reflex tearing and as a result increase lacrimal drainage. This is likely to reduce the ocular bioavailability of the drug and thus needs to be considered during the formulation process. In addition, general safety considerations such as sterility, ocular toxicity and irritation, and the amount of preservative used, need to be taken into account when formulating an ocular dosage form.

***Physicochemical Drug Properties*** On the one hand, factors such as the chemical nature or the concentration of a drug can cause irritation of the ocular tissues, inducing reflex tearing and therefore reducing the retention time of the formulation in front of the eye.

However, more important physicochemical properties in terms of ocular bioavailability are the ones that affect the corneal permeability of the active compound. These include the lipophilicity of the drug as reflected by its *n*-octanol–water partition coefficient [32], the molecular size and shape [33], the charge [34], and the acid–base properties as determined by its $pK_a$ [35].

According to Kaur and Smitha [36], the optimum lipophilicity for corneal absorption is found in drugs with an *n*-octanol–water partition coefficient between 10 and 100. For drugs with smaller partition coefficients (highly hydrophilic drugs), the lipophilic epithelium forms the rate-limiting barrier, whereas the hydrophilic stroma

represents the primary barrier for transcorneal diffusion of highly lipophilic drugs [1, 3, 25]. In general, more lipophilic drugs penetrate via the transcellular pathway while the more hydrophilic drugs enter the cornea via the paracellular route.

In the case of ionized compounds (weak acids and bases), drug permeation depends on the chemical equilibrium between the ionized and the un-ionized form, both in the delivery system itself and in the lacrimal fluid. In general, un-ionized molecules penetrate lipid membranes more readily than ionized ones.

Besides lipophilicity and degree of ionization, the charge of the drug molecule may have an effect on its penetration. Cationic drugs permeate the cornea more easily than anionic compounds, which are repelled by the negative charge of the mucin layer on the ocular surface as well as the negatively charged pores present in the corneal epithelium [1, 3, 25, 29]. However, a positive charge may also decrease permeation in some cases, due to possible ionic interactions between the positively charged molecules and the negatively charged carboxylic acid groups of the tight junction proteins [21, 30].

Finally, the molecular size of the drug has an effect on the corneal absorption. The cornea is impermeable to molecules larger than 5000 Da, whereas the conjunctival tissues allow compounds of up to 20,000 Da to penetrate [14, 15].

***Buffer Capacity and pH*** The normal pH of the tear fluid is 7.4. Ocular formulations should ideally be formulated between pH 7.0 and 7.7 to avoid irritation of the eye [31]. However, in most cases the pH necessary for maximal solubility or stability of the drug is well outside this range. The tear fluid has only a limited buffering capacity, which is mainly due to the dissolved carbon dioxide and bicarbonate. It is therefore recommended to formulate using buffers with a low buffering capacity to allow the tears to regain their normal pH more rapidly [31].

***Instillation Volume*** The cul-de-sac normally holds 7–9 μL [37], but can momentarily accommodate up to 30 μL without overflowing. Most commercial eye droppers however, deliver a volume of approximately 50 μL. The excess volume is rapidly removed, either by spillage from the conjunctival sac or through the puncta to the nasolacrimal drainage system, until the tears return to their normal volume [38]. Chrai et al. [39, 40] determined the influence of drop size on the rate of drainage of a solution instilled into the conjunctival sac of rabbits. The authors reported that the drainage process followed first-order kinetics and found that the rate of solution drainage from the conjunctival sac (as reflected by the elimination rate constant) was directly proportional to the instilled volume [40]. Similar observations were reported with other dosage forms such as suspensions [41] and liposomes [42]. Therefore, keeping the applied dose constant while decreasing the instilled volume substantially increases ocular bioavailability and decreases systemic absorption [3, 26, 43].

***Osmotic Pressure*** The osmolality of the lacrimal fluid is mainly dependent on the number of ions dissolved in the aqueous layer of the tear film and normally ranges between 310 and 350 mOsm/kg [44]. When an ophthalmic solution is instilled into the eye, it mixes with the tear fluid, resulting in an osmotic pressure that is dependent on the osmolality of the tears as well as that of the formulation and the amount of the formulation instilled. In general, hypotonic solutions are better tolerated by

the ocular tissues than hypertonic ones, which lead to increased lacrimation. If the tonicity of the formulation is lower than 260 mOsm/kg or higher than 480 mOsm/kg, the formulation becomes irritant [45], induces reflex tearing and blinking, and is therefore likely to reduce the bioavailability of topically instilled drugs.

## 3 FORMULATION APPROACHES TO IMPROVE OCULAR BIOAVAILABILITY

One of the major problems encountered with topical administration is the rapid precorneal loss caused by nasolacrimal drainage and high tear fluid turnover, which leads to drug concentrations of typically less than 10% of the applied drug. Approaches to improve the ocular bioavailability have been attempted in two directions: to increase the corneal permeability by using penetrations enhancers or prodrugs and to prolong the contact time with the ocular surface by using viscosity-enhancing or in situ gelling polymers. Table 1 summarizes conventional and novel ocular drug delivery approaches. Marketed ophthalmic delivery systems based on recent formulation approaches are listed in Table 2. An optimal ocular delivery systems would be administered in the form of an eye drop, causing neither blurred vision nor ocular irritancy, and would only need to be instilled once a day [1].

### 3.1 Conventional Dosage Forms

Conventional dosage forms such as solutions, suspensions, and ointments account for almost 90% of the currently accessible ophthalmic formulations on the market [1, 169]. They offer some advantages such as their ease of administration by the patient, ease of preparation, and the low production costs. However, there are also significant disadvantages associated with the use of conventional solutions in particular, including the very short contact time with the ocular surface and the fast nasolacrimal drainage, both leading to a poor bioavailability of the drug. Various ophthalmic delivery systems have been investigated to increase the corneal permeability and prolong the contact time with the ocular surface. However, conventional eye drops prepared and administered as aqueous solutions remain the most commonly used dosage form in ocular disease management.

***Solutions*** The reasons behind choosing solutions over other dosage forms are their favorable cost advantage, the simplicity of formulation development and production, and the high acceptance by patients [170]. However, there are also a few drawbacks, such as rapid and extensive precorneal loss, high absorption via the conjunctiva and the nasolacrimal duct leading to systemic side effects, as well as the increased installation frequency resulting in low patient compliance.

Some of these problems have been encountered by addition of viscosity-enhancing agents such as cellulose derivates, which are believed to increase the viscosity of the preparation and consequently reduce the drainage rate. The use of viscosity enhancers will be discussed later in this section.

***Suspensions*** Suspensions of the micronized drug (<10 μm) in a suitable aqueous vehicle are formulated, where the active compound is water insoluble. This is the

**TABLE 1 Summary of Conventional and Novel Ocular Drug Delivery Approaches**

| Drug | Formulation Approach | Polymers/Bases | References |
|---|---|---|---|
| Pilocarpine | Viscosity enhancer | Methylcellulose | [40] |
| Fluorescein | Viscosity enhancer + collagen shields | Collagen shields in methylcellulose vehicle | [46] |
| Ciprofloxacin | Viscosity enhancers + penetration enhancer | Carbopol 934P/HPMC + dodecylmaltoside | [47] |
| Pilocarpine | Viscosity enhancers, mucoadhesives, in situ gelling systems | Gellan gum, xanthan gum, HEC, HPMC, PVA | [48] |
| Progesterone | Mucoadhesive | Cross-linked acrylic acid | [49] |
| Levobetaxolol | Mucoadhesive | Polyacrylic acid | [50] |
| Tropicamide | Mucoadhesive | CMC, HPCL, HPCM, PVP, PVA | [51] |
| | | CMC-Na, HA-Na, PAA | [52] |
| Pilocarpine, tropicamide | Mucoadhesive | Hyaluronic acid, polyacrylic acid | [53] |
| Tertrahydrozoline | Mucoadhesive | Hyaluronic acid, polyacrylic acid, chitosan, gelatin | [54] |
| Tobramycin | Mucoadhesive | Chitosan | [55] |
| Tobramycin, ofloxacin | Mucoadhesive | Chitosan | [56] |
| Ofloxacin | Mucoadhesive | *N*-Trimethyl and *N*-carboxy-methyl chitosan | [57, 58] |
| Pilocarpine | Mucoadhesive | Carbopol 934P | [59] |
| | | Carbomer 974P and 1342 | [60] |
| | | PVA, PVP, dextran, HPMC, HEC, MC, PAA, Na-hyaluronate, Na-alginate, gellan gum, chitosan | [61] |
| Timolol | Mucoadhesive | Xyloglucan | [62] |
| | | Xanthan gum | [63] |
| | | Carrageenan, gellan gum | [64] |
| | | Carrageenan, locust bean gum, guar gum, xanthan gum, scleroglucan, xanthan gum, sodium alginate, β-cyclodextrin | [65] |
| | | Hyaluronic acid | [66] |
| | | HPMC, PVA, hyaluronic acid | [67] |
| | | Carbomer, hyaluronic acid | [68] |
| | | PAA, PVP | [69] |
| | In situ gelling system | Gellan gum | [70–78] |
| Indomethacine | In situ gelling system | Gellan gum | [79] |
| Pefloxacin mesylate | In situ gelling system | Gellan gum | [80] |
| Gatifloxacin | In situ gelling system | Alginate/HPMC | [81] |
| Ofloxacin | In situ gelling system | Carbopol 940/HPMC | [82] |
| Pilocarpine | In situ gelling system | Alginate | [83] |
| Carteolol | In situ gelling system | Alginate | [84] |
| Ciprofloxacin | In situ gelling system | Poloxamer/hyaluronic acid | [85] |

**TABLE 1** ***Continued***

| Drug | Formulation Approach | Polymers/Bases | References |
|---|---|---|---|
| Pilocarpine | In situ gelling system | Pluronic F127, MC, HPMC | [86] |
| | | Pluronic F127/carbopol | [87] |
| | | Pluronic F127, xyloglycan | [88] |
| Timolol | In situ gelling system | Pluronic F127, MC, HPMC, CMC | [89] |
| Doxorubicin | In situ gelling system | Chitosan/glycerophosphate | [90] |
| | | Pluronic F127 and F68, sodium hyaluronate | [91] |
| Pilocarpine, hydrocortisone | Nanoparticles | Gelatin | [92] |
| Hydrocortisone | Nanoparticles | Albumin | [93] |
| Pilocarpine | Nanoparticles | Albumin | [94, 95] |
| | | Poly(methyl)methacrylate–acrylic acid copolymer (Piloplex) | [96–98] |
| | | Cellulose acetate hydrogen phthalate (CAP) | [99–101] |
| Betaxolol | Nanoparticles | Poly-ε-caprolactone (PECL), poly (isobutyl)cyanoacrylate,polylactic-*co*-glycolic acid (PLGA) | [102] |
| Carteolol | Nanoparticles | Poly-ε-caprolactone (PECL) | [103] |
| Pilocarpine | Nanoparticles | Poly(butyl)cyanoacrylate (PBCA) | [104–107] |
| | | Poly(hexyl)cyanoacrylate (PHCA) | [104, 106, 107] |
| Ciclosporine | Nanoparticles | Chitosan | [108] |
| Indomethacine | Nanoparticles | Chitosan- and poly-L-lysine-coated poly-ε-caprolactone nanocapsules | [109] |
| Rhodamine | Nanoparticles | Chitosan- and PEG-coated poly-ε-caprolactone nanocapsules | [110] |
| Aciclovir | Nanoparticles | PEG-coated PLA nanospheres | [111] |
| Epinephrine | Nanoparticles | Poly-*N*-isopropylacrylamide | [112] |
| Ibuprofen | Nanoparticles | Eudragit RS100 | [113, 114] |
| Flurbiprofen | Nanoparticles | Eudragit RS100 and RL100 | [115] |
| Diclofenac | Nanoparticles | Eudragit RLPM and RSPM | [116] |
| Tobramycin | Nanoparticles | Various lipids [solid lipid nanoparticles (SLN)] | [117] |
| Gentamicin | Microspheres | Eudragit RS100 and RL100 | [118] |
| Piroxicam | Microspheres | Pectin | [119] |
| | | Albumin | [120] |
| Pilocarpine | Liposomes + mucoadhesive coating | Carbopol-coated liposomes | [121, 122] |
| Radioactive-labeled BSA | Liposomes + mucoadhesive coating | Chitosan-coated liposomes | [123] |
| Oligonucleotides (pdT16) | Liposomes + in situ gelling system | Liposomes in poloxamer 407 gel | [124–126] |
| Tropicamide | Liposomes + in situ gelling system | Liposomes in polycarbophil gel | [127] |

**TABLE 1** ***Continued***

| Drug | Formulation Approach | Polymers/Bases | References |
|---|---|---|---|
| Timolol | Niosomes | Chitosan- and carbopol-coated niosomes | [128] |
| Indomethacine | Submicrometer emulsions | Phospholipids, lauroamphodiacetate | [129] |
| Pilocarpine | Submicrometer emulsions | Mono-dodecylphosphoric acid | [130] |
| | Microemulsions | Lecithin-based microemulsions | [131] |
| Retinol | Microemulsions | Tween 60 and 80, soy bean lecithin, *n*-butanol, triacetin, PG | [132] |
| | | Crodamol EO, Crill 1 and Crillet 4 | [133] |
| Epinephrine | Prodrug | Dipivalyl epinephrine | [134, 135] |
| Pilocarpine | Prodrug | *O, O′*(Xylylene)bispilocarpic acid esters | [136] |
| Ganciclovir | Prodrug | Ganciclovir acyl ester | [137] |
| Atenolol, betaxolol, Timolol | Penetration enhancers | Polyoxyethylene alkyl ethers (Brij), bile salts | [138] |
| Cromoclycin | Penetration enhancers | EDTA | [33] |
| Carbonic anhydrase inhibitors | Cyclodextrins | β-Cyclodextrin, hydroxypropyl-β-cyclodextrin | [139] |
| Ciprofloxacin | Cyclodextrins | Hydroxypropyl-β-cyclodextrin | [140] |
| Pilocarpine | Cyclodextrins | Hydroxypropyl-β-cyclodextrin | [141] |
| | Ocular films | Hydroxypropyl cellulose | [142] |
| | | Poly(2-hydroxypropyl-methacrylate) | [143] |
| Pefloxacin mesylate | Ocular films | HPC, HPMC, PVP, PVA | [144] |
| | Ocular inserts | PVP, Eudragit RS and RL | [145] |
| Mitomycin C | Ocular inserts | Collagen implant | [146] |
| Pilocarpine | Ocular inserts | Collagen shield | [147] |
| | | PAA, PVP, HPMC | [148] |
| | | PVA, glyceryl behenate, xanthan gum, carrageenan, HPMC, HA; coated with Eudragit RS and RL | [149] |
| Timolol | Ocular inserts | HPC, coated with Eudragit RS and RL | [150] |
| Tilisolol | Ocular inserts | Poly(hydroxypropyl-methacrylate) | [151] |
| | | Poly(2-hydroxypropyl-methacrylate), polypropylene tape | [152] |
| Ciprofloxacin | Ocular inserts | Sodium alginate, Eudragit, polyvinyl acetate | [153] |
| Pradofloxacin | Ocular inserts | Hydrogel coating on thin metallic wire (OpthaCoil) | [154] |

**TABLE 1** ***Continued***

| Drug | Formulation Approach | Polymers/Bases | References |
|---|---|---|---|
| Fluorescein | Ocular inserts | HPMC lyophilisate on carrier strip | [155] |
| | | HPMC lyophilisate on poly(tetra fluoroethylene) carrier strip | [156] |
| | | Carbopol 974P, maize starch | [157, 158] |
| Gentamicin | Ocular inserts | CAP, carbomer, HPMC, HPC, EC | [159, 160] |
| Oxytetracycline | Ocular inserts | Silicone, PAA, PMA | [161] |
| Oxytetracycline, prednisolone, gentamicin | Ocular inserts | PEO 400, PEO 900 | [162] |
| Ofloxacin | Ocular inserts | PEO 200, 400, 900, and 2000 | [163] |
| | | PEO 400, Eudragit L100 | [164] |
| | | PEO 900, PEO 2000, chitosan-thiolated PAA | [165] |
| Fluorescein, diclofenac | Ocular inserts | | [166] |
| Gentamicin | Iontophoresis | Hydroxyethyl methacrylate | [167] |
| Pilocarpine, tropicamide | Dendrimers | Various poly(amidoamine) (PAMAM) dendrimers | [168] |

case for most of the steroids. It is assumed that the drug particles remain in the conjunctival sac, thus promoting a sustained release effect [171]. There have been many studies trying to prove this assumption, but none of them has revealed a pronounced prolonged release profile [35, 172, 173].

According to Davies [174], topical ophthalmic suspensions have a number of limitations compared to solutions. They need to be adequately shaken before use to ensure correct dosing, a process which can result in poor patient compliance. In addition, they need to be sterilized, which may cause physical instability of the formulation. Furthermore, the amount of drug required to achieve only a moderate increase in bioavailability is very high, rendering suspensions expensive in terms of their production costs [175].

The drug particle size plays the most important role in the formulation process of suspensions. Particles greater than 10 μm cause patient discomfort. As they are perceived as foreign substances, they cause reflex tearing in order to eliminate the particles from the ocular surface [176]. A study by Schoenwald and Stewart [177] showed the influence of the particle size of dexamethasone on its bioavailability. The in vivo dissolution rate decreased with increasing particle size to the point when particles were removed from the conjunctival sac before the dissolution was complete.

As a result, achieving a near-solution state with small particles that are easy to resuspend and show minimal sedimentation remains the goal when formulation of a suspension is unavoidable [176].

**TABLE 2 Marketed Ophthalmic Delivery Systems Based on Recent Formulation Approaches**

| Formulation Approach | Polymer/Base | Product | Company |
|---|---|---|---|
| Suspensions/ microparticulates | Carbomer ion exchange resin | Betoptic S | Alcon |
| Ointments | Wool fat, paraffin | Polyvisc | Alcon |
| | Liquid paraffin, white soft paraffin | LacriLube | Allergan |
| Viscosity enhancers/ mucoadhesives | Polyethylene glycol (PEG), propylene glycol (PG), HP-guar | Systane | Alcon |
| | Dextran, HPMC | Bion Tears | Alcon |
| | Carboxymethylcellulose sodium (CMC-Na) | Refresh Celluvisc | Allergan |
| | | Refresh Liquigel | Allergan |
| | Polyvinyl alcohol (PVA) | Liquifilm Tears | Allergan |
| | HPMC | Lacrigel | Sunways |
| | Carbomer | Viscotears | Novartis |
| | Carbomer, polyvinyl alcohol (PVA) | Nyogel | Novartis |
| | Hyaluronic acid (HA) | Hy-Drop | Bausch&Lomb Fidia Oftal |
| | Sodium hyaluronate | Vismed | TRB Chemedica |
| | Carbomer | Pilopine HS | Alcon |
| | Polyacrylic acid (PAA) | Fucithalmic | Leo Pharma |
| In situ gelling systems | Gellan gum | Timoptic XE | Merck |
| | Polycarbophil | DuraSite | InSite Vision |
| | Polyacrylic acid, poloxamer | Smart Hydrogel | Advanced Medical Solutions |
| Prodrugs | Dipivefrin hydrochloride (epinephrin prodrug) | Propine | Allergan |
| Ocular inserts | Alginic acid | Ocusert | Alza |
| | Hydroxylpropyl cellulose | Lacrisert | Merck |
| | Silicone elastomer | Ocufit SR | Escalon Medical |
| | Collagen shield | MediLens | Chiron |
| | | ProShield | Alcon |

***Ointments*** Ointments generally consist of a dissolved or dispersed drug in an appropriate vehicle base. They are the most commonly used semisolid preparations as they are well tolerated and fairly safe and increase the ocular bioavailability of the drug. The instilled ointment breaks up into small oily droplets that remain in the cul-de-sac as a drug depot. The drug eventually gets to the ointment–tear interface due to the shearing action of the eyelids [178].

Sieg and Robinson [35] compared the bioavailability of fluorometholone in a solution, a suspension, and an ointment. They found that the peak concentration

($c_{max}$) of the drug in the aqueous humor of rabbits was comparable with all three formulations, whereas the time to peak concentration ($t_{max}$) occurred much later with the ointment, leading to a significantly greater total bioavailability of the drug.

Overall, ophthalmic ointments offer the following advantages: reduced dilution of the medication via the tear film, resistance to nasolacrimal drainage, and an increased precorneal contact time [179, 180]. However, oily viscous preparations for ophthalmic use (such as ointments) can cause blurred vision and matting of the eyelids and may also be associated with discomfort by the patient as well as occasional ocular mucosal irritation. Ointments are therefore generally used in combination with eye drops, which can be administered during the day, while the ointment is applied at night, when clear vision is not required.

### 3.2 Polymeric Delivery Systems

Polymeric systems used for ocular drug delivery can be divided into three groups: viscosity-enhancing polymers, which simply increase the formulation viscosity, resulting in decreased lacrimal drainage and enhanced bioavailability; mucoadhesive polymers, which interact with the ocular mucin, therefore increasing the contact time with the ocular tissues; and in situ gelling polymers, which undergo sol-to-gel phase transition upon exposure to the physiological conditions present in the eye. However, there are no defined boundaries between the different groups and most polymers exhibit more than one of these properties.

***Viscosity-Enhancing Polymers*** In order to reduce the lacrimal clearance (drainage) of ophthalmic solutions, various polymers have been added to increase the viscosity of conventional eye drops, prolong precorneal contact time, and subsequently improve ocular bioavailability of the drug [40, 51, 181–184]. Among the range of hydrophilic polymers investigated in the area of ocular drug delivery are polyvinyl alcohol (PVA) and polyvinyl pyrrolidone (PVP), cellulose derivates such as methylcellulose (MC), and polyacrylic acids (carbopols).

Chrai and Robinson [40] evaluated the use of an MC solution of pilocarpine in albino rabbits and found a decrease in the drainage rate with increasing viscosity. Patton and Robinson [185] investigated the relationship between the viscosity and the contact time or drainage rate and demonstrated an optimum viscosity of 12–15 cps for an MC solution in rabbits. The influence of different polymers on the activity of pilocarpine in rabbits and human was reported by Saettone et al. [182]. Trueblood et al. [183] used lacrimal microscintigraphy to evaluate the corneal contact time for saline, PVA, and hydroxpropyl methylcellulose (HPMC) and observed the longest contact time for the formulation with HPMC as a viscosity-enhancing agent.

The ocular shear rate ranges from $0.03\,s^{-1}$ during interblinking periods to 4250–$28{,}500\,s^{-1}$ during blinking [186]. It has a great influence on the rheological properties of viscous ocular dosage forms and consequently the bioavailability of the incorporated drug [187]. Newtonian systems do not show any real improvement of bioavailability below a certain viscosity and blinking becomes painful, followed by reflex tearing, if the viscosity is too high [188]. While the viscosity of Newtonian systems is independent from the shear rate, non-Newtonian pseudoplastic or so-called

shear-thinning systems exhibit a decrease in viscosity with increasing shear rates. This pseudoplastic behavior is favorable for ocular drug delivery systems as it offers less resistance to blinking and therefore shows greater acceptance by patients than Newtonian systems of the same viscosity [189].

***Mucoadhesive Polymers*** Bioadhesion refers to the attachment of a drug molecule or a delivery system to a specific biological tissue by means of interfacial forces. If the surface of the tissue is covered by a mucin film, as is the case for the external globe, it is more commonly referred to as mucoadhesion.

In order to be an effective mucoadhesive excipient, polymers must show one or more of the following features [190]: strong hydrogen binding group, strong anionic charge, high molecular weight, sufficient chain flexibility, surface energy properties favoring spreading onto the mucus, and near-zero contact angle to allow maximum contact with the mucin coat.

The most commonly used bioadhesives are macromolecular hydrocolloids with numerous hydrophilic functional groups capable of forming hydrogen bonds (such as carboxyl, hydroxyl, amide, and sulfate groups) [191]. Hui and Robinson [49] were the first to demonstrate the usefulness of bioadhesive polymers in improving the ocular bioavailability of progesterone. Saettone et al. [53] evaluated a series of bioadhesive dosage forms for ocular delivery of pilocarpine and tropicamide and found hyaluronic acid to be the most promising mucoadhesive polymer. Lehr et al. [192] suggested that cationic polymers, which are able to interact with the negative sialic acid residues of the mucin, would probably show better mucoadhesive properties than anionic or neutral polymers. They investigated the polycationic polymer chitosan and demonstrated that the mucoadhesive performance of chitosan was significantly higher in neutral or slightly alkaline pH as it is present in the tear fluid.

However, according to Park and Robinson [193], polyanions are better than polycations in terms of binding and potential toxicity. In general, both anionic and cationic charged polymers demonstrate better mucoadhesive properties than non-ionic polymer, such as cellulose derivates or PVA [194, 195].

The mechanism of mucoadhesion involves a series of different steps. First, the mucoadhesive formulation needs to establish an intimate contact with the corneal surface. Prerequisites are either good wetting or swelling of the mucoadhesive polymer as well as sufficient spreading across the cornea. The second stage involves the penetration of the mucoadhesive polymer chains into the crevices of the tissue surface and also the entanglement with the mucous chains [196]. On a molecular level, mucoadhesion is a results of van-der-Waals forces, electrostatic attractions, hydrogen bonding, and hydrophobic interactions [36].

Mucoadhesive polymers increase the contact time of a formulation with the tear film and simulate the continuous delivery of tears due to a high water-restraining capacity. As such, they allow a decrease in the instillation frequency compared to common eye drops and are therefore useful as artificial tear products [1, 197, 198].

***In Situ Gelling Systems*** In situ gelling systems are viscous polymer-based liquids that exhibit sol-to-gel phase transition on the ocular surface due to change in a specific physicochemical parameter (ionic strength, temperature, pH, or solvent exchange). They are highly advantageous over preformed gels as they can easily be

instilled in liquid form but are capable of prolonging the residence time of the formulation on the surface of the eye due to gelling [199].

The principal advantage of in situ gelling systems is the easy, accurate, and reproducible administration of a dose compared to the application of preformed gels [198].

The concept of forming gels in situ (e.g., in the cul-de-sac of the eye) was first suggested in the early 1980s, and ever since then various triggers of in situ gelling have been further investigated.

Polymers that may undergo sol-to-gel transition triggered by a change in pH are cellulose acetate phthalate (CAP) and cross-linked polyacrylic acid derivates such as carbopols, methacrylates and polycarbophils. CAP latex is a free-running solution at pH 4.4 which undergoes sol-to-gel transition when the pH is raised to that of the tear fluid. This is due to neutralization of the acid groups contained in the polymer chains, which leads to a massive swelling of the particles. The use of pH-sensitive latex nanoparticles has been described by Gurny et al. [100, 200]. Carbopols have apparent $pK_a$ values in the range of 4–5 resulting in rapid gelation due to rise in pH after ocular administration.

Gellan gum is an anionic polysaccharide which undergoes phase transition under the influence of an increased ionic strength. In fact, the gel strength increases proportionally with the amount of mono- or divalent cations present in the tear fluid. As a consequence, the usual reflex tearing, which leads to a dilution of common viscous solutions, further enhances the viscosity of gellan gum formulations due to the increased amount of tear fluid and thus higher cation concentration [201]. Several studies have been performed comparing Gelrite formulations (low acetyl gellan gum) to conventional ophthalmic solutions of the same active compound. Shedden et al. [76] compared the plasma concentrations of timolol following multiple applications of Timoptic-XE and a timolol maleate solution. They found that a once-daily application of the in situ gelling formulation was sufficient to reduce the intraocular pressure to levels comparable to a twice-daily application of the solution, leading to better patient compliance as well as a reduction in systemic side effects.

Poloxamers or pluronics are block copolymers consisting of poly(oxyethylene) and poly(oxypropylene) units. They rapidly undergo thermal gelation when the temperature is raised to that of the ocular surface (32 °C), while they remain liquid at refrigerator temperature. Poloxamers exhibit surface active properties, but even if used in high concentrations (usually between 20 and 30%), Pluronic F127 was found no more damaging to the cornea than a physiological saline solution [202]. In order to reduce the total polymer concentration and achieve better gelling properties, several poloxamer combinations have been tested. Wei et al. [91] used a mixture of Pluronic F127 and F68 resulting in a more suitable phase transition temperature with a free-flowing liquid under 25 °C.

Combining thermal- with pH-dependent gelation, Kumar et al. [86] developed a combination of methylcellulose 15% and carbopol 0.3%. This composition exhibited a sol-to-gel transition between 25 and 37 °C with a pH increase from 4 to 7.4 [203]. A possible mechanism for the thermal effect could be the decrease in the degree of the methylcellulose hydration, while the polyacrylic acid can transform into a gel upon an increase in pH due to the buffering properties of the tear fluid [1].

### 3.3 Colloidal Delivery Systems

Colloidal carriers have been investigated as drug delivery systems for the past 30 years in order to achieve specific drug targeting, facilitate the bioavailability of drugs through biological membranes, and protect the drug against enzymatic degradation. Their use in topical administration and especially in ocular delivery however has only been studied for the last two decades [1, 3].

Colloidal carriers are small particulate systems ranging in size from 100 to 400 nm. As they are usually suspended in an aqueous solution, they can easily be administered as eye drops, thus avoiding the potential discomfort resulting from bigger particles present in ocular suspensions or from viscous or sticky preparations [38].

Most efforts in ophthalmic drug delivery have been made with the aim of increasing the corneal penetration of the drug. Calvo et al. [204] have shown that colloidal particles are preferably taken up by the corneal epithelium via endocytosis. It has also been stated by Lallemand and co-workers [205], that the cornea acts as a drug reservoir, slowly releasing the active compound present in the colloidal delivery system to the surrounding ocular tissues.

***Nanoparticles*** Nanoparticles have been among the most widely studied particulate delivery systems over the past three decades. They are defined as submicrometer-sized polymeric colloidal particles ranging from 10 to 1000 nm in which the drug can be dissolved, entrapped, encapsulated, or adsorbed [206]. Depending on the preparation process, nanospheres or nanocapsules can be obtained. Nanospheres have a matrixlike structure where the drug can either be firmly adsorbed at the surface of the particle or be dispersed/dissolved in the matrix. Nanocapsules, on the other hand, consist of a polymer shell and a core, where the drug can either be dissolved in the inner core or be adsorbed onto the surface [207].

The first nanoparticulate delivery system studied was Piloplex, consisting of pilocarpine ionically bound to poly(methyl)methacrylate–acrylic acid copolymer nanoparticles [44]. Klein et al. [1, 98] found that a twice-daily application of Piloplex in glaucoma patients was just as effective as three to six instillations of conventional pilocarpine eye drops. However, the formulation was never accepted for commercialization due to various formulation-related problems, including the nonbiodegradability, local toxicity, and difficulty of preparing a sterile formulation [208].

Another early attempt to formulate a nanoparticulate system for the delivery of pilocarpine was made by Gurny [99]. This formulation was based on pilocarpine dispersed in a hydrogen CAP pseudolatex formulation. Gurny and co-workers [101] compared the formed nanoparticles to a 0.125% solution of hyaluronic acid some years after their first investigation and found that the viscous hyaluronic acid system showed a significantly longer retention time in front of the eye than the pseudolatex formulation.

The most commonly used biodegradable polymers in the preparation of nanoparticulate systems for ocular drug delivery are poly-alkylcyanoacrylates, poly-ε-caprolactone, and polylactic-*co*-glycolic acid copolymers. Marchal-Heussler et al. [102, 103] compared the three particulate delivery systems using antiglaucoma drugs including betaxolol and cartechol. Results showed that poly-ε-caprolactone (nanospheres and nanocapsules) exhibited the highest pharmacological activity when

loaded with betaxolol. It seemed that the higher ocular activity was related to the hydrophobic nature of the carrier and that the mechanism of action seemed to be directly linked to the agglomeration of the particles in the conjunctival sac [1]. In general, nanocapsules displayed a much better effect than nanospheres probably due to the fact that the active compound was in its un-ionized form in the oily core and could diffuse faster into the cornea. Diffusion of the drug from the oily core of the nanocapsule to the corneal epithelium seems to be more effective than diffusion from the internal, more hydrophilic matrix of the nanospheres [1, 209].

In order to achieve a sustained drug release and a prolonged therapeutic activity, nanoparticles must be retained in the cul-de-sac and the entrapped drug must be released from the particles at a certain rate. If the release is too fast, there is no sustained release effect. If it is too slow, the concentration of the drug in the tears might be too low to achieve penetration into the ocular tissues [208]. The major limiting issues for the development of nanoparticles include the control of particle size and drug release rate as well as the formulation stability.

So far, there is only one microparticulate ocular delivery system on the market. Betoptic S is obtained by binding of betaxolol to ion exchange resin particles. Betoptic S 0.25% was found to be bioequivalent to the Betoptic 0.5% solution in lowering the intraocular pressure [208].

***Liposomes*** Liposomes were first reported by Bangham in the 1960s and have been investigated as drug delivery systems for various routes ever since then. They offer some promising features for ophthalmic drug delivery as they can be administered as eye drops but will localize and maintain the pharmacological activity of the drug at its site of action [1]. Due to the nature of the lipids used, conventional liposomes are completely biodegradable, biocompatible, and relatively nontoxic [1].

A liposome or so-called vesicle consists of one or more concentric spheres of lipid bilayers separated by water compartments with a diameter ranging from 80 nm to 100 μm. Owing to their amphiphilic nature, liposomes can accommodate both lipohilic (in the lipid bilayer) and hydrophilic (encapsulated in the central aqueous compartment) drugs [207]. According to their size, liposomes are classified as either small unilamellar vesicles (SUVs) (10–100 nm) or large unilamellar vesicles (LUVs) (100–300 nm). If more than one bilayer is present, they are referred to as multilamellar vesicles (MLVs). Depending on their lipid composition, they can have a positive, negative, or neutral surface charge.

Liposomes are potentially valuable as ocular drug delivery systems due to their simplicity of preparation and versatility in physical characteristics. However, their use is limited by instability (due to hydrolysis of the phospholipids), limited drug-loading capacity, technical difficulties in obtaining sterile preparations, and blurred vision due to their size and opacity [42].

In addition, liposomes are subject to the same rapid precorneal clearance as conventional ocular solutions, especially the ones with a negative or no surface charge [127]. Positively charged liposomes, on the other hand, were reported to exhibit a prolonged precorneal retention due to electrostatic interactions with the negative sialic acid residues of the mucin layer [2, 127, 208, 210–213].

There have been several attempts to use liposomes in combination with other newer formulation approaches, such as incorporating them into mucoadhesive gels or coating them with mucoadhesive polymers [210]. Mucoadhesive polymers inves-

tigated in this regard were poly(acrylic acid) (PAA), hyaluronic acid (HA), chitosan, and poloxamer [36, 43, 121–122, 214].

Durrani and co-workers [122] reported on the influence of a carbopol coating on the corneal retention of pilocarpine-loaded liposomes, and demonstrated a biphasic response with an initial low intensity followed by a sustained reaction.

Bochot et al. [124, 125] developed a novel delivery system for oligonucleotides by incorporating them into liposomes and then dispersing them into a thermosensitive gel composed of poloxamer 407. They compared the in vitro release of the model oligonucleotides pdT16 from simple poloxamer gels (20 and 27% poloxamer) with the ones where pdT16 was encapsulated into liposomes and then dispersed within the gels. They found that the release of the oligonucleotides from the gels was controlled by the poloxamer dissolution, whereas the dispersion of liposomes within 27% poloxamer gel was shown to slow down the diffusion of pdT16 out of the gel.

***Niosomes*** In order to circumvent some of the limitations encountered with liposomes, such as their chemical instability, the cost and purity of the natural phospholipids, and oxidative degradation of the phospholipids, niosomes have been developed. Niosomes are nonionic surfactant vesicles which exhibit the same bilayered structures as liposomes. Their advantages over liposomes include improved chemical stability and low production costs. Moreover, niosomes are biocompatible, biodegradable, and nonimmunogenic [215]. They were also shown to increase the ocular bioavailability of hydrophilic drugs significantly more than liposomes. This is due to the fact that the surfactants in the niosomes act as penetrations enhancers and remove the mucous layer from the ocular surface [209].

A modified version of niosomes are the so-called discomes, which vary from the conventional niosomes in size and shape. The larger size of the vesicles (12–60 μm) prevents their drainage into the nasolacrimal drainage system. Furthermore, their disclike shape provides them with a better fit in the cul-de-sac of the eye [26].

Vyas et al. [216] demonstrated that discomes entrapped higher amounts of timolol maleate than niosomes and that both niosomes and discomes significantly increased the bioavailability of timolol maleate when compared to a conventional timolol maleate solution.

***Microemulsions*** Microemulsions (MEs) are colloidal dispersions composed of an oil phase, an aqueous phase, and one or more surfactants. They are optically isotropic and thermodynamically stable and appear as transparent liquids as the droplet size of the dispersed phase is less than 150 nm. One of their main advantages is their ability to increase the solubilization of lipophilic and hydrophilic drugs accompanied by a decrease in systemic absorption [217]. Moreover, MEs are transparent systems thus enable monitoring of phase separation and/or precipitation. In addition, MEs possess low surface tension and therefore exhibit good wetting and spreading properties.

While the presence of surfactants is advantageous due to an increase in cellular membrane permeability, which facilitates drug absorption and bioavailability [218], caution needs to be taken in relation to the amount of surfactant incorporated, as high concentrations can lead to ocular toxicity. In general, nonionic surfactants are preferred over ionic ones, which are generally too toxic to be used in ophthalmic

formulations. Surfactants most frequently utilized for the preparation of MEs are poloxamers, polysorbates, and polyethylene glycol derivatives [219].

Similar to all other colloidal delivery systems discussed above it was hypothesized by numerous research teams that a positive charge (provided by cationic surfactants [220]) would increase the ocular residence time of the formulation due to electrostatic interactions with the negatively charged mucin residues. However, toxicological studies contradicted this assumption regarding the ocular effects, and so far there has been no publication demonstrating a distinct beneficial effect of charged surfactants incorporated into MEs.

Microemulsions can be classified into three different types depending on their microstructure: oil-in-water (o/w ME), water-in-oil (w/o ME), and bicontinuous ME. They have been investigated by physical chemists since the 1940s but have only gained attention as potential ocular drug delivery carriers within the last two decades.

Gasco and co-workers [221] investigated the potential application of o/w lecithin MEs for ocular administration of timolol, in which the drug was present as an ion pair with octanoate. The ocular bioavailability of the timolol ion pair incorporated into the ME was compared to that of an ion pair solution as well as a simple timolol solution. Areas under the curve for the ME and the ion pair solution respectively were 3.5 and 4.2 times higher than that of the simple timolol solution. A prolonged absorption was achieved using the ME with detectable amounts of the drug still present 120 min after instillation.

Various lecithin-based MEs were also characterized by Hasse and Keipert [131]. The formulations were tested in terms of their physicochemical parameters (pH, refractive index, osmolality, viscosity, and surface tension) and physiological compatibility (HET-CAM and Draize test). In addition, in vitro and in vivo evaluations were performed. The tested MEs showed favorable physicochemical parameters and no ocular irritation as well as a prolonged pilocarpine release in vitro and in vivo.

Muchtar and co-workers [129] prepared MEs with poloxamer 188 and soybean lecithin to deliver indometacin to the ocular tissues. They found a threefold increased indomethacin concentration in the cornea and aqueous humor 1 h post-instillation.

Beilin et al. [222] demonstrated a prolonged ocular retention of a submicrometer emulsion (SME) in the conjunctival sac using a fluorescent marker (0.01% calcein) as well as the miotic response of New Zealand Albino rabbits to pilocarpine. They found that the fluorescence intensity of calcein in SME was significantly higher than that of a calcein solution at all time points. Moreover, the pilocarpine SME exhibited a longer duration of miosis than the simple pilocarpine solution. It should be mentioned that SMEs are true emulsions, being different from MEs. They do not form spontaneously and are kinetically rather than thermodynamically stable. They generally require high-shear homogenization to form and are more susceptible to phase inversion. Furthermore, they are neither transparent nor translucent but rather turbid due their larger droplet size compared to MEs. While the two terms are used interchangeably in the the scientific literature, they actually refer to two distinct categories of dispersed systems and should be differentiated from each other.

The w/o MEs composed of water, Crodamol EO, Crill 1, and Crillet 4 were investigated as potential ocular delivery systems by Alany et al. [133]. It was hypothesized

that w/o MEs undergo phase transition into lamellar liquid crystals (LCs) upon aqueous dilution by the tears, prolonging the precorneal retention time due to an increase in the formulation's viscosity. HET-CAM studies revealed no ocular irritancy by the excipients used. Ocular drainage was evaluated via γ-scintigraphy and demonstrated a significantly higher precorneal retention of the tested microemulsions compared to an aqueous solution.

The use of MEs in ocular delivery is very attractive due to all the advantages offered by these formulations. They are thermodynamically stable and transparent, possess low viscosity, and thus are easy to instill, formulate, and sterilize (via filtration). Moreover, they offer the possibility of reservoir and/or localizer effects. All these factors, in addition to the ones previously mentioned, render MEs promising ocular delivery systems.

### 3.4 Other Delivery Approaches

Many other ocular delivery approaches have been investigated over the past decades, including the use of prodrugs, penetration enhancers, cyclodextrins, as well as different types of ocular inserts. In addition, iontophoresis, which is an active drug delivery approach utilizing electrical current of only 1–2 mA to transport ionized drugs across the cornea, offers an effective, noninvasive method for ocular delivery. Another more recent approach is the use of dendrimers in ocular therapy. Dendrimers are synthetic spherical molecules named after their characteristic treelike branching around a central core with a size ranging from 2 to 10 nm in diameter [223]. So far, PAMAM (polyamidoamine) has been the most commonly studied dendrimer system for ocular use [224, 225].

***Prodrugs*** Prodrugs are pharmacologically inactive derivates of drug molecules that require a chemical or enzymatic transformation into their active parent drug [226]. To be effective, an ocular prodrug should show an appropriate lipophilicity to facilitate corneal absorption, posses sufficient aqueous solubility and stability to be formulated as an eye drop, and demonstrate the ability to be converted to the active parent drug at a rate that meets therapeutic needs [227].

When considering ophthalmic drug molecules as prodrug candidates, the following factors need to be considered: the pathway and mechanism of ocular drug penetration, the functional groups of the drug candidate amenable to prodrug derivatization, and the enzymes present in the ocular tissues, which are necessary for prodrug activation [28].

The majority of ophthalmic prodrugs developed so far are chemically classified as ester. They are derived from the esterification of the hydroxyl or carboxylic acid groups present in the parent molecule. Of all enzymes participating in the activation of prodrugs, esterases, which are present in all anterior segment tissues except the tear film, have received the most attention [228, 229].

Prodrugs were introduced into the area of ocular drug delivery about 25 years ago [230], and steroids were probably the first ones to be utilized as prodrugs. However, the concept of prodrugs was not fully exploited until the introduction of dipivefrin (epinephrine prodrug) in the late 1970s. Kaback and co-workers [134] found that a 0.1% dipivalyl epinephrine solution lowered the intraocular pressure as effective as a 2% epinephrine solution, while significantly lowering the systemic

side effects. Wei et al. [135] compared the ocular penetration, distribution, and metabolism of epinephrine and dipivalyl epinephrine and found the partition coefficient of the later to be 100–600 times higher than that of epinephrine, therefore leading to a 10-times faster absorption into the rabbit eye.

Dipivefrin was the only commercially available ophthalmic prodrug at that time. However, numerous prodrug derivates have been designed to improve the efficacy of ophthalmic drugs ever since.

Jarvinen and co-workers [136] synthesized unique *O, O′*-(xylylene)bispilocarpic acid esters containing two pilocarpic acid monoesters linked with one moiety. The found that prodrug showed a two- to seven-fold higher corneal permeability than pilocarpine itself despite the high molecular weight.

Tirucherai et al. [137] formulated an acyl ester prodrug of ganciclovir. The increased permeability was associated with a linear increased susceptibility of the ganciclovir esters to the esterases present in the cornea.

So far, aims that have been achieved by using prodrugs include the modification of the drug's duration of action, reduction of the systemic absorption, and reduction of ocular and systemic side effects. Although prodrugs are commonly used to treat diseases of the anterior segment, there have also been attempts to treat conditions associated with the posterior segment of the eye.

***Penetration Enhancers*** The transport process across the corneal tissue is the rate-limiting step in ocular drug absorption. Increasing the permeability of the corneal epithelium by penetration enhancers is likely to enhance the drug transport across the corneal tissues and therefore improve ocular bioavailability of the drug.

Penetration enhancers act by increasing the permeability of the corneal cell membrane and/or loosening the tight junctions between the epithelial cells, which primarily restrict the entry of molecules via the paracellular pathway. Classes of penetration enhancers include surfactants, bile salts, calcium chelators, preservatives, fatty acids, and some glycosides such a saponin.

Surfactants are perceived to enhance drug absorption by disturbing the integrity of the plasma membranes. When present at low concentrations, surfactants are incorporated into the lipid bilayer, leading to polar defects in the membrane, which change the membrane's physical properties. When the lipid bilayer is saturated, micelles start to form, enclosing phospholipids from the membranes, hence leading to membrane solubilization [36]. Saettone et al. [138] found an increased corneal permeability for atenolol, timolol, and betaxolol by including 0.05% Brij 35, Brij 78, and Brij 98 into their formulations.

Bile salts are amphiphilic molecules that are surface active and self-associate to form micelles in aqueous solution. They increase corneal permeability by changing the rheological properties of the bilayer [231]. A number of bile salts such as deoxycholate, taurodeoxycholate, and glycocholate have been tested so far, and it was suggested, that a difference in their physicochemical properties (solubilizing activity, lipophilicity, $Ca^{2+}$ sequestration capacity) is probably related to their performance as permeability-enhancing agents [36].

Another class of penetration enhancers includes calcium chelators such as ethylenediaminetetraacetic acid (EDTA). These molecules induce $Ca^{2+}$ depletion in the cells. This leads to a global change within the cell and as a result loosens the tight junctions between superficial epithelial cells, thus facilitating paracellular transport

[138, 232]. Grass et al. [33] were among the first to emphasize the enhancing effects of chelating agents for ocular drug delivery. They found that 0.5% EDTA doubled the corneal absorption of topically applied glycerol and cromoclycin sodium.

Large numbers of penetration enhancers have been investigated to date. However, the unique structure of the corneal/conjunctival tissues requires caution. When selecting penetration enhancers for ocular delivery, their capacity to affect the integrity of the epithelial surfaces needs to be considered. Studies have shown that penetration enhancers themselves can penetrate ocular tissues, which could lead to potential toxicity. EDTA concentrations in the iris–ciliary body, for example, were found to be high enough to alter the permeability of the blood vessels in the uveal tract, therefore indirectly accelerating the drug's removal from the aqueous humor [233]. Similarly, benzalkonium chloride (BAC), a cationic surfactant which shows the highest penetration-enhancing effect among the currently used preservatives, was found to accumulate in the cornea for several days.

***Cyclodextrins*** Cyclodextrins were introduced into the area of ocular drug delivery in the early 1990s. They are a group of homologous cyclic oligosaccharides with a hydrophilic outer surface and a lipophilic cavity in the center. Their initial aim was to increase the solubility of lipophilic drugs by forming inclusion complexes. Cyclodextrin complexation generally results in improved wettability, dissolution, solubility, and stability in solution as well as reduced side effects [234].

It is assumed that cyclodextrins are too large and hydrophilic to penetrate biological membranes. However, they act as penetration enhancers by assuring a high drug concentration at the corneal surface, from where the drug then partitions into the ocular tissues [235].

Even though cyclodextrins drug complexes seem to decrease ocular toxicity of irritant drugs, cyclodextrins themselves may exhibit ocular toxicity and should therefore be screened by performing corneal sensitivity studies. Among all cyclodextrin derivates investigated, hydroxy-propyl-β-cyclodextrin showed the most favorable properties in terms of toxicity [1].

Nijhawan and Agarwal [140] investigated inclusion complexes of ciprofloxacin hydrochloride and hydroxy-propyl-β-cyclodextrin and found that the complexes exhibited better stability, biological activity, and ocular tolerance than the uncomplexed drug in solution.

Aktas et al. [141] showed an increased permeation of pilocarpine nitrate complexed with hydroxy-propyl-β-cyclodextrin using isolated rabbit cornea. They found a significant reduction in the pupil diameter compared to a simple aqueous solution of the same active compound.

Cyclodextrins improve chemical stability, increase the drug's bioavailability, and decrease local irritation. However, the improvement of ocular bioavailability seems to be limited by the very slow dissociation of the complexes in the precorneal tear fluid.

Several studies have shown that combinations of cyclodextrins drug complexes and viscosity enhancers can significantly improved ocular absorption [141, 236–237] and should therefore be further investigated.

***Ocular Inserts*** Solid ocular dosage forms such as films, erodible and nonerodible inserts, rods, and shields have been developed to overcome the typical pulse-entry-type drug release associated with conventional ocular dosage forms. This pulse entry

is characterized by a transient overdose, a relatively short period of appropriate dosing, followed by a prolonged period of underdosing. Ocular inserts were developed in order to overcome these disadvantages by providing a more controlled, sustained, and continuous drug delivery by maintaining an effective drug concentration in the target tissues and yet minimizing the number of applications [238].

Ocular inserts probably represent one of the oldest ophthalmic formulation approaches. In 1948 the British Pharmacopoeia described an atropine-in-gelatin wafer and ever since then numerous systems have been developed applying various polymers and different release principals. However, the difficulty of insertion by the patient, foreign-body sensation, and inadvertent loss of inserts from the eye make these systems less popular, especially among the elderly. Furthermore, the high cost involved in manufacture prevented the insert market from taking off [197].

Two products, Alza Ocusert and Merck Lacrisert, have been marketed, although Ocusert is no longer available. Ocusert is a membrane-controlled reservoir system for the treatment of glaucoma. It contains pilocarpine and alginic acid in the core reservoir, sandwiched between two transparent, lipophilic ethylenevinyl acetate (EVA) rate-controlling membranes, which allow the drug to diffuse from the reservoir at a precisely determined rate for a period of seven days. This system is nonbiodegradable and must therefore be removed after use. Lacrisert, on the other hand, is a soluble minirod of hydroxypropylmethyl cellulose without any active ingredient. The system is placed in the conjunctival sac, where it softens within an hour and completely dissolves within 14–18 h. Lacrisert stabilizes and thickens the precorneal tear film and prolongs the tear film break-up time, which is usually accelerated in patients with dry-eye syndrome (keratoconjunctivitis sicca) [239].

A number of ocular inserts using different techniques, namely soluble, erodible, nonerodible, and hydrogel inserts with polymers such as cellulose derivates, acrylates, and poly(ethylene oxide), have been investigated over the last few decades.

An example of a degradable matrix system is the pilocarpine-containing inserts formulated by Saettone et al. [148]. Pilocarpine nitrate and polyacrylic acid were incorporated into a matrix containing polyvinyl alcohol and two types of hydroxypropyl methylcellulose. It was shown that all inserts significantly increased the pharmacological effect (miotic response) compared to a solution of pilocarpine nitrate.

Sasaki et al. [151] prepared nondegradable disc-type ophthalmic inserts of β- blockers using different polymers. They found that inserts made from poly(hydroxypropyl methacrylate) were able to control the release of tilisolol hydrochloride.

Numerous studies have also been performed on soluble collagen shields. Collagen shields are fabricated from porcine scleral tissue, which has a similar collagen composition to that of the human cornea. Drug loading is typically achieved by soaking the collagen shield in the drug solution prior to application. Designed to slowly dissolve within 12, 24, or 72 h, collagen shields have attracted much interest as potential sustained ocular drug delivery systems over the last years [240].

## 4 CONCLUSION

Although conventional eye drops still represent about 90% of all marketed ophthalmic dosage forms, there have been significant efforts towards the development of new drug delivery systems.

Only a few of these new ophthalmic drug delivery systems have been commercialized over the past decades, but research in the different areas of ocular drug delivery has provided important impetus and dynamism, with the promise of some new and exciting developments in the field.

An ideal ophthalmic delivery system should be able to achieve an effective drug concentration at the target site for an extended period of time while minimizing systemic side effects. In addition, the system should be comfortable and easy to use, as the patient's acceptance will continue to play an important role in the design of future ocular formulations.

All delivery technologies mentioned in this chapter hold unlimited potential for clinical ophthalmology. However, each of them still bears its own drawbacks. To circumvent these, newer trends are directed toward combinations of the different drug delivery approaches. Examples for this include the incorporation of particulates into in situ gelling systems or coating of nanoparticles with mucoadhesive polymers.

These combinations will open new directions for the improvement of ocular bioavailability, but they will also increase the complexity of the formulations, thus increasing the difficulties in understanding the mechanism of action of the drug delivery systems.

Many interesting delivery approaches have been investigated during the past decades in order to optimize ocular bioavailability, but much remains to be learned before the perfect ocular drug delivery system can be developed.

## REFERENCES

1. Le Bourlais, C., Acar, L., Zia, H., Sado, P. A., Needham, T., and Leverge, R. (1998), Ophthalmic drug delivery systems—Recent advances, *Prog. Ret. Eye Res.*, 17(1), 33–58.
2. Le Bourlais, C. A., Treupel-Acar, L., Rhodes, C. T., Sado, P. A., and Leverge, R. (1995), New ophthalmic drug delivery systems, *Drug Dev. Ind. Pharm.*, 21(1), 19–59.
3. Jarvinen, K., Jarvinen, T., and Urtti, A. (1995), Ocular absorption following topical delivery, *Adv. Drug Deliv. Rev.*, 16(1), 3–19.
4. Sasaki, H., Yamamura, K., Mukai, T., Nishida, K., Nakamura, J., Nakashima, M., et al. (1999), Enhancement of ocular drug penetration, *Crit. Rev. Ther. Drug Carrier Syst.*, 16(1), 85–146.
5. Baeyens, V., and Gurny, R. (1997), Chemical and physical parameters of tears relevant for the design of ocular drug delivery formulations, *Pharm. Acta Helv.*, 72(4), 191–202.
6. Van Haeringen, N. J. (1981), Clinical biochemistry of tears, *Surv. Ophthalmol.*, 26(2), 84–96.
7. Nagyova, B., and Tiffany, J. M. (1999), Components responsible for the surface tension of human tears, *Curr. Eye Res.*, 19(1), 4–11.
8. Tiffany, J. M. (2003), Tears in health and disease, *Eye*, 17(8), 923–926.
9. Murube, J., Murube, L., and Murube, A. (1999), Origin and types of emotional tearing, *Eur. J. Ophthalmol.*, 9(2), 77–84.
10. Versura, P., Bonvicini, F., Caramazza, R., and Laschi, R. (1985), Scanning electron microscopy of normal human corneal epithelium obtained by scraping-off in vivo, *Acta Ophthalmol.*, 63(3), 361–365.

11. Klyce, S. D., and Crosson, C. E. (1985), Transport processes across the rabbit corneal epithelium: A review, *Curr. Eye Res.*, 4(4), 323–331.
12. Lens, A. (1999), *Ocular Anatomy and Physiology*, Slack, New York.
13. Ahmed, I. (2003), The noncorneal route in ocular drug delivery, in Mitra, A. K., Ed., *Ophthalmic Drug Delivery Systems*, 2nd ed., Marcel Dekker, New York, pp. 335–363.
14. Greaves, J. L., and Wilson, C. G. (1993), Treatment of diseases of the eye with mucoadhesive delivery systems, *Adv. Drug Deliv. Rev.*, 11(3), 349–383.
15. Huang, A. J. W., Tseng, S. C. G., and Kenyon, K. R. (1989), Paracellular permeability of corneal and conjunctival epithelia, *Invest. Ophthalmol. Vis. Sci.*, 30(4), 684–689.
16. Ahmed, I., and Patton, T. F. (1985), Importance of the noncorneal absorption route in topical ophthalmic drug delivery, *Invest. Ophthalmol. Vis. Sci.*, 26(4), 584–587.
17. Macha, S., Hughes, P. M., and Mitra, A. K. (2003), Overview of ocular drug delivery, in Mitra, A. K., Ed., *Ophthalmic Drug Delivery Systems*, 2nd ed., Marcel Dekker, New York, pp. 1–11.
18. Robinson, J. C. (1993), Ocular anatomy and physiology relevant to ocular drug delivery, in Mitra, A. K., Ed., *Ophthalmic Drug Delivery Systems*, Marcel Dekker, New York, pp. 29–58.
19. Mikkelson, T. J., Chrai, S. S., and Robinson, J. R. (1973), Altered bioavailability of drugs in the eye due to drug-protein interaction, *J. Pharm. Sci.*, 62(10), 1648–1653.
20. Stjernschantz, J., and Astin, M. (1993), Anatomy and physiology of the eye. Physiological aspects of ocular drug therapy, in Edman, P., Ed., *Biopharmaceutics of Ocular Drug Delivery*, CRC Press, Boca Raton, FL, pp. 1–25.
21. Hornof, M., Toropainen, E., and Urtti, A. (2005), Cell culture models of the ocular barriers, *Eur. J. Pharm. Biopharm.*, 60(2), 207–225.
22. Jumbe, N. L., and Miller, M. H. (2003), Ocular drug transfer following systemic drug administration, in Mitra, A. K., Ed., *Ophthalmic Drug Delivery Systems*, 2nd ed., Marcel Dekker, New York, pp. 109–133.
23. Urtti, A., and Salminen, L. (1993), Animal pharmacokinetic studies, in Mitra, A. K., Ed., *Ophthalmic Drug Delivery Systems*, Marcel Dekker, New York, pp. 121–136.
24. Duvvuri, S., Majumdar, S., and Mitra, A. K. (2003), Drug delivery to the retina: Challenges and opportunities, *Expert Opin. Biol. Ther.*, 3(1), 45–56.
25. Loftssona, T., and Jarvinen, T. (1999), Cyclodextrins in ophthalmic drug delivery, *Adv. Drug Deliv. Rev.*, 36(1), 59–79.
26. Kaur, I. P., Garg, A., Singla, A. K., and Aggarwal, D. (2004), Vesicular systems in ocular drug delivery: An overview, *Int. J. Pharm.*, 269(1), 1–14.
27. Lee, V. H., and Robinson, J. R. (1979), Mechanistic and quantitative evaluation of precorneal pilocarpine disposition in albino rabbits, *J. Pharm. Sci.*, 68(6), 673–684.
28. Lee, V. H. L., and Li, V. H. K. (1989), Prodrugs for improved ocular drug delivery, *Adv. Drug Deliv. Rev.*, 3(1), 1–38.
29. Rojanasakul, Y., and Robinson, J. R. (1989), Transport mechanisms of the cornea: Characterization of barrier permselectivity, *Int. J. Pharm.*, 55(2–3), 237–246.
30. Palmgren, J. J., Toropainen, E., Auriola, S., and Urtti, A. (2002), Liquid chromatographic-electrospray ionization mass spectrometric analysis of neutral and charged polyethylene glycols, *J. Chromatogr. A*, 976(1–2), 165–170.
31. Van Ooteghem, M. M. M. (1993), Formulation of ophthalmic solutions and suspensions. Problems and Advantages, in Edman, P., Ed., *Biopharmaceutics of Ocular Drug Delivery*, CRC Press, Boca Raton, FL, pp. 27–42.

32. Schoenwald, R. D., and Huang, H. S. (1983), Corneal penetration behavior of beta-blocking agents I: Physiochemical factors, *J. Pharm. Sci.*, 72(11), 1266–1272.
33. Grass, G. M., Wood, R. W., and Robinson, J. R. (1985), Effects of calcium chelating agents on corneal permeability, *Invest. Ophthalmol. Vis. Sci.*, 26(1), 110–113.
34. Liaw, J., Rojanasakul, Y., and Robinson, J. R. (1992), The effect of drug charge type and charge density on corneal transport, *Int. J. Pharm.*, 88(1–3), 111–124.
35. Sieg, J. W., and Robinson, J. R. (1975), Vehicle effects on ocular drug bioavailability: Evaluation of fluorometholone, *J. Pharm. Sci.*, 64(6), 931–936.
36. Kaur, I. P., and Smitha, R. (2002), Penetration enhancers and ocular bioadhesives: Two new avenues for ophthalmic drug delivery, *Drug Dev. Ind. Pharm.*, 28(4), 353–369.
37. Mishima, S., Gasset, A., Klyce, S. D., Jr, and Baum, J. L. (1966), Determination of tear volume and tear flow, *Invest. Ophthalmol.*, 5(3), 264–276.
38. Mainardes, R. M., Urban, M. C. C., Cinto, P. O., Khalil, N. M., Chaud, M. V., Evangelista, R. C., et al. (2005), Colloidal carriers for ophthalmic drug delivery, *Curr. Drug Targets*, 6(3), 363–371.
39. Chrai, S. S., Patton, T. F., Mehta, A., and Robinson, J. R. (1973), Lacrimal and instilled fluid dynamics in rabbit eyes, *J. Pharm. Sci.*, 62(7), 1112–1121.
40. Chrai, S. S., and Robinson, J. R. (1974), Ocular evaluation of methylcellulose vehicle in albino rabbits, *J. Pharm. Sci.*, 63(8), 1218–1223.
41. Sieg, J. W., and Triplett, J. W. (1980), Precorneal retention of topically instilled micronized particles, *J. Pharm. Sci.*, 69(7), 863–864.
42. Lee, V. H. L., Urrea, P. T., Smith, R. E., and Schanzlin, D. J. (1985), Ocular drug bioavailability from topically applied liposomes, *Surv. Ophthalmol.*, 29(5), 335–348.
43. Zimmer, A., and Kreuter, J. (1995), Microspheres and nanoparticles used in ocular delivery systems, *Adv. Drug Deliv. Rev.*, 16(1), 61–73.
44. Ludwig, A. (2005), The use of mucoadhesive polymers in ocular drug delivery, *Adv. Drug Deliv. Rev.*, 57(11), 1595–1639.
45. Ludwig, A., and Van Ooteghem, M. (1987), The influence of the osmolality on the precorneal retention of ophthalmic solutions, *J. Pharm. Belg.*, 42(4), 259–266.
46. Kaufman, H. E., Steinemann, T. L., Lehman, E., Thompson, H. W., Varnell, E. D., Jacob-LaBarre, J. T., et al. (1994), Collagen-based drug delivery and artificial tears, *J. Ocul. Pharmacol.*, 10(1), 17–27.
47. Ke, T.-L., Cagle, G., Schlech, B., Lorenzetti, O. J., and Mattern, J. (2001), Ocular bioavailability of ciprofloxacin in sustained release formulations, *J. Ocul. Pharmacol. Ther.*, 17(6), 555–563.
48. Meseguer, G., Buri, P., Plazonnet, B., Rozier, A., and Gurny, R. (1996), Gamma scintigraphic comparison of eyedrops containing pilocarpine in healthy volunteers, *J. Ocul. Pharmacol. Ther.*, 12(4), 481–488.
49. Hui, H.-W., and Robinson, J. R. (1985), Ocular delivery of progesterone using a bioadhesive polymer, *Int. J. Pharm.*, 26(3), 203–213.
50. Lele, B. S., and Hoffman, A. S. (2000), Insoluble ionic complexes of polyacrylic acid with a cationic drug for use as a mucoadhesive, ophthalmic drug delivery system, *J. Biomater. Sci. Polym. Ed.*, 11(12), 1319–1331.
51. Saettone, M. F., Giannaccini, B., Ravecca, S., La Marca, F., and Tota, G. (1984), Polymer effects on ocular bioavailability—The influence of different liquid vehicles on the mydriatic response of tropicamide in humans and in rabbits, *Int. J. Pharm.*, 20(1–2), 187–202.
52. Herrero-Vanrell, R., Fernandez-Carballido, A., Frutos, G., and Cadorniga, R. (2000), Enhancement of the mydriatic response to tropicamide by bioadhesive polymers, *J. Ocul. Pharmacol. Ther.*, 16(5), 419–428.

53. Saettone, M. F., Chetoni, P., Tilde Torracca, M., Burgalassi, S., and Giannaccini, B. (1989), Evaluation of muco-adhesive properties and in vivo activity of ophthalmic vehicles based on hyaluronic acid, *Int. J. Pharm.*, 51(3), 203–212.
54. Sandri, G., Bonferoni, M. C., Chetoni, P., Rossi, S., Ferrari, F., Ronchi, C., et al. (2006), Ophthalmic delivery systems based on drug-polymer-polymer ionic ternary interaction: In vitro and in vivo characterization, *Eur. J. Pharm. Biopharm.*, 62(1), 59–69.
55. Felt, O., Furrer, P., Mayer, J. M., Plazonnet, B., Buri, P., and Gurny, R. (1999), Topical use of chitosan in ophthalmology: Tolerance assessment and evaluation of precorneal retention, *Int. J. Pharm.*, 180(2), 185–193.
56. Felt, O., Baeyens, V., Buri, P., and Gurny, R. (2001), Delivery of antibiotics to the eye using a positively charged polysaccharide as vehicle, *Aaps Pharmsci*, 3(4), E34.
57. Di Colo, G., Zambito, Y., Burgalassi, S., Nardini, I., and Saettone, M. F. (2004), Effect of chitosan and of N-carboxymethylchitosan on intraocular penetration of topically applied ofloxacin, *Int. J. Pharm.*, 273(1–2), 37–44.
58. Di Colo, G., Zambito, Y., Burgalassi, S., and Saettone, M. F. (2003), Effects of chitosan and of its N-trimethyl and N-carboxymethyl derivatives on the ocular pharmacokinetics of ofloxacin in rabbits, paper presented at the 30th International Symposium on Controlled Release of Bioactive Materials, Glasgow, Scottland.
59. Davies, N. M., Farr, S. J., Hadgraft, J., and Kellaway, I. W. (1991), Evaluation of mucoadhesive polymers in ocular drug delivery. I. Viscous solutions, *Pharm. Res.*, 8(8), 1039–1043.
60. Edsman, K., Carlfors, J., and Harju, K. (1996), Rheological evaluation and ocular contact time of some carbomer gels for ophthalmic use, *Int. J. Pharm.*, 137(2), 233–241.
61. Hartmann, V., and Keipert, S. (2000), Physico-chemical, in vitro and in vivo characterisation of polymers for ocular use, *Pharmazie*, 55(6), 440–443.
62. Burgalassi, S., Chetoni, P., Panichi, L., Boldrini, E., and Saettone, M. F. (2000), Xyloglucan as a novel vehicle for timolol: Pharmacokinetics and pressure lowering activity in rabbits, *J. Ocul. Pharmacol. Ther.*, 16(6), 497–509.
63. Ceulemans, J., Vinckier, I., and Ludwig, A. (2002), The use of xanthan gum in an ophthalmic liquid dosage form: Rheological characterization of the interaction with mucin, *J. Pharm. Sci.*, 91(4), 1117–1127.
64. Verschueren, E., Van Santvliet, L., and Ludwig, A. (1996), Evaluation of various carrageenans as ophthalmic viscolysers, STP Pharma *Sci.*, 6(3), 203–210.
65. Albasini, M., and Ludwig, A. (1995), Evaluation of polysaccharides intended for ophthalmic use in ocular dosage forms, *Farmaco*, 50(Sept.), 633–642.
66. Snibson, G. R., Greaves, J. L., Soper, N. D., Prydal, J. I., Wilson, C. G., and Bron, A. J. (1990), Precorneal residence times of sodium hyaluronate solutions studied by quantitative gamma scintigraphy, *Eye*, 4(4), 594–602.
67. Snibson, G. R., Greaves, J. L., Soper, N. D., Tiffany, J. M., Wilson, C. G., and Bron, A. J. (1992), Ocular surface residence times of artificial tear solutions, *Cornea*, 11(4), 288–293.
68. Debbasch, C., De La Salle, S. B., Brignole, F., Rat, P., Warnet, J. M., and Baudouin, C. (2002), Cytoprotective effects of hyaluronic acid and Carbomer 934P in ocular surface epithelial cells, *Invest. Ophthalmol. Vis. Sci.*, 43(11), 3409–3415.
69. Oechsner, M., and Keipert, S. (1999), Polyacrylic acid/polyvinylpyrrolidone bipolymeric systems. I. Rheological and mucoadhesive properties of formulations potentially useful for the treatment of dry-eye-syndrome, *Eur. J. Pharm. Biopharm.*, 47(2), 113–118.

70. Shibuya, T., Kashiwagi, K., and Tsukahara, S. (2003), Comparison of efficacy and tolerability between two gel-forming timolol maleate ophthalmic solutions in patients with glaucoma or ocular hypertension, *Ophthalmologica*, 217(1), 31–38.
71. Rosenlund, E. F. (1996), The intraocular pressure lowering effect of timolol in gel-forming solution, *Acta Ophthalmol. Scand.*, 74(2), 160–162.
72. Stewart, W. C., Leland, T. M., Cate, E. A., and Stewart, J. A. (1998), Efficacy and safety of timolol solution once daily versus timolol gel in treating elevated intraocular pressure, *J. Glaucoma*, 7(6), 402–407.
73. Shedden, A., Laurence, J., and Tipping, R. (2001), Efficacy and tolerability of timolol maleate ophthalmic gel-forming solution versus timolol ophthalmic solution in adults with open-angle glaucoma or ocular hypertension: A six-month, double-masked, multicenter study, *Clin. Ther.*, 23(3), 440–450.
74. Carlfors, J., Edsman, K., Petersson, R., and Jornving, K. (1998), Rheological evaluation of Gelrite(R) in situ gels for ophthalmic use, *Eur. J. Pharm. Sci.*, 6(2), 113–119.
75. Schenker, H. I., and Silver, L. H. (2000), Long-term intraocular pressure-lowering efficacy and safety of timolol maleate gel-forming solution 0.5% compared with timoptic XE 0.5% in a 12-month study, *Am. J. Ophthalmol.*, 130(2), 145–150.
76. Shedden, A. H., Laurence, J., Barrish, A., and Olah, T. V. (2001), Plasma timolol concentrations of timolol maleate: Timolol gel-forming solution (TIMOPTIC-XE) once daily versus timolol maleate ophthalmic solution twice daily, *Doc. Ophthalmol.*, 103(1), 73–79.
77. Lindell, K., and Engstrom, S. (1993), In vitro release of timolol maleate from an in situ gelling polymer system, *Int. J. Pharm.*, 95(1–3), 219–228.
78. Rozier, A., Mazuel, C., Grove, J., and Plazonnet, B. (1989), Gelrite(R): A novel, ion-activated, in-situ gelling polymer for ophthalmic vehicles. Effect on bioavailability of timolol, *Int. J. Pharm.*, 57(2), 163–168.
79. Balasubramaniam, J., Kant, S., and Pandit, J. K. (2003), In vitro and in vivo evaluation of the Gelrite gellan gum-based ocular delivery system for indomethacin, *Acta Pharm.*, 53(4), 251–261.
80. Sultana, Y., Aqil, M., and Ali, A. (2006), Ion-activated, Gelrite-based in situ ophthalmic gels of pefloxacin mesylate: Comparison with conventional eye drops, *Drug Deliv.*, 13(3), 215–219.
81. Liu, Z., Li, J., Nie, S., Liu, H., Ding, P., and Pan, W. (2006), Study of an alginate/HPMC-based in situ gelling ophthalmic delivery system for gatifloxacin, *Int. J. Pharm.*, 315(1–2), 12–17.
82. Srividya, B., Cardoza, R. M., and Amin, P. D. (2001), Sustained ophthalmic delivery of ofloxacin from a pH triggered in situ gelling system, *J. Controlled Release*, 73(2–3), 205–211.
83. Cohen, S., Lobel, E., Trevgoda, A., and Peled, Y. (1997), A novel in situ-forming ophthalmic drug delivery system from alginates undergoing gelation in the eye, *J. Controlled Release*, 44(2–3), 201–208.
84. Demailly, P., Allaire, C., and Trinquand, C. (2001), Ocular hypotensive efficacy and safety of once daily carteolol alginate, *Br. J. Ophthalmol.*, 85(8), 921–924.
85. Cho, K. Y., Chung, T. W., Kim, B. C., Kim, M. K., Lee, J. H., Wee, W. R., et al. (2003), Release of ciprofloxacin from poloxamer-graft-hyaluronic acid hydrogels in vitro, *Int. J. Pharm.*, 260(1), 83–91.
86. Desai, S. D., and Blanchard, J. (1998), Evaluation of pluronic F127-based sustained-release ocular delivery systems for pilocarpine using the albino rabbit eye model, *J. Pharm. Sci.*, 87(10), 1190–1195.

87. Lin, H. R., and Sung, K. C. (2000), Carbopol/pluronic phase change solutions for ophthalmic drug delivery, *J. Controlled Release*, 69(3), 379–388.
88. Miyazaki, S., Suzuki, S., Kawasaki, N., Endo, K., Takahashi, A., and Attwood, D. (2001), In situ gelling xyloglucan formulations for sustained release ocular delivery of pilocarpine hydrochloride, *Int. J. Pharm.*, 229(1–2), 29–36.
89. El-Kamel, A. H. (2002), In vitro and in vivo evaluation of Pluronic F127-based ocular delivery system for timolol maleate, *Int. J. Pharm.*, 241(1), 47–55.
90. Wu, J., Su, Z.-G., and Ma, G.-H. (2006), A thermo- and pH-sensitive hydrogel composed of quaternized chitosan/glycerophosphate, *Int. J. Pharm.*, 315(1–2), 1–11.
91. Wei, G., Xu, H., Ding, P. T., Li, S. M., and Zheng, J. M. (2002), Thermosetting gels with modulated gelation temperature for ophthalmic use: The rheological and gamma scintigraphic studies, *J. Controlled Release*, 83(1), 65–74.
92. Vandervoort, J., and Ludwig, A. (2004), Preparation and evaluation of drug-loaded gelatin nanoparticles for topical ophthalmic use, *Eur. J. Pharm. Biopharm.*, 57(2), 251–261.
93. Zimmer, A. K., Maincent, P., Thouvenot, P., and Kreuter, J. (1994), Hydrocortisone delivery to healthy and inflamed eyes using a micellar polysorbate 80 solution or albumin nanoparticles, *Int. J. Pharm.*, 110(3), 211–222.
94. Zimmer, A. K., Zerbe, H., and Kreuter, J. (1994), Evaluation of pilocarpine-loaded albumin particles as drug delivery systems for controlled delivery in the eye I. In vitro and in vivo characterisation, *J. Controlled Release*, 32(1), 57–70.
95. Zimmer, A. K., Chetoni, P., Saettone, M. F., Zerbe, H., and Kreuter, J. (1995), Evaluation of pilocarpine-loaded albumin particles as controlled drug delivery systems for the eye. II. Coadministration with bioadhesive and viscous polymers, *J. Controlled Release*, 33(1), 31–46.
96. Mazor, Z., Ticho, U., Rehany, U., and Rose, L. (1979), Piloplex, a new long-acting pilocarpine polymer salt. B: Comparative study of the visual effects of pilocarpine and Piloplex eye drops, *Br. J. Ophthalmol.*, 63(1), 48–51.
97. Ticho, U., Blementhal, M., Zonis, S., Gal, A., Blank, I., and Mazor, Z. W. (1979), Piloplex, a new long-acting pilocarpine polymer salt. A: Long-term study, *Br. J. Ophthalmol.*, 63(1), 45–47.
98. Klein, H. Z., Lugo, M., Shields, M. B., Leon, J., and Duzman, E. (1985), A dose-response study of piloplex for duration of action, *Am. J. Ophthalmol.*, 99(1), 23–26.
99. Gurny, R. (1981), Preliminary study of prolonged acting drug delivery system for the treatment of glaucoma, *Pharm. Acta Helv.*, 56(4–5), 130–132.
100. Gurny, R., Boye, T., and Ibrahim, H. (1985), Ocular therapy with nanoparticulate systems for controlled drug delivery, *J. Controlled Release*, 2, 353–361.
101. Gurny, R., Ibrahim, H., Aebi, A., Buri, P., Wilson, C. G., Washington, N., et al. (1987), Design and evaluation of controlled release systems for the eye, *J. Controlled Release*, 6(1), 367–373.
102. Marchal-Heussler, L., Orallo, F., Vila Jato, J. L., and Alonso, M. J. (1992), Colloidal drug delivery systems for the eye: A comparison of the efficacy of three different polymers: Polyisobutylcyanoacrylate, polylacticcoglycolic acid, poly-epsilon-caprolactone, *STP Pharma Sci.*, 2, 98–104.
103. Marchal-Heussler, L., Sirbat, D., Hoffman, M., and Maincent, P. (1993), Poly(epsilon-caprolactone) nanocapsules in carteolol ophthalmic delivery, *Pharm. Res.*, 10(3), 386–390.
104. Harmia, T., Speiser, P., and Kreuter, J. (1986), Optimization of pilocarpine loading onto nanoparticles by sorption procedures, *Int. J. Pharm.*, 33(1–3), 45–54.

105. Diepold, R., Kreuter, J., Himber, J., Gurny, R., Lee, V. H., Robinson, J. R., et al. (1989), Comparison of different models for the testing of pilocarpine eyedrops using conventional eyedrops and a novel depot formulation (nanoparticles), *Graefes Arch. Clin. Exp. Ophthalmol.*, 227(2), 188–193.

106. Harmia, T., Speiser, P., and Kreuter, J. (1987), Nanoparticles as drug carriers in ophthalmology, *Pharm. Acta Helv.*, 62(12), 322–331.

107. Zimmer, A., Mutschler, E., Lambrecht, G., Mayer, D., and Kreuter, J. (1994), Pharmacokinetic and pharmacodynamic aspects of an ophthalmic pilocarpine nanoparticle-delivery-system, *Pharm. Res.*, 11(10), 1435–1442.

108. De Campos, A. M., Sanchez, A., and Alonso, M. J. (2001), Chitosan nanoparticles: A new vehicle for the improvement of the delivery of drugs to the ocular surface. Application to cyclosporin A, *Int. J. Pharm.*, 224(1–2), 159–168.

109. Calvo, P., Vila-Jato, J. L., and Alonso, M. J. (1997), Evaluation of cationic polymer-coated nanocapsules as ocular drug carriers, *Int. J. Pharm.*, 153(1), 41–50.

110. De Campos, A. M., Sanchez, A., Gref, R., Calvo, P., and Alonso, M. J. (2003), The effect of a PEG versus a chitosan coating on the interaction of drug colloidal carriers with the ocular mucosa, *Eur. J. Pharm. Sci.*, 20(1), 73–81.

111. Giannavola, C., Bucolo, C., Maltese, A., Paolino, D., Vandelli, M. A., Puglisi, G., et al. (2003), Influence of preparation conditions on acyclovir-loaded poly-*d,l*-lactic acid nanospheres and effect of PEG coating on ocular drug bioavailability, *Pharma. Res.*, 20(4), 584–590.

112. Hsiue, G.-H., Hsu, S.-h., Yang, C.-C., Lee, S.-H., and Yang, I.-K. (2002), Preparation of controlled release ophthalmic drops, for glaucoma therapy using thermosensitive poly-*N*-isopropylacrylamide, *Biomaterials*, 23(2), 457–462.

113. Pignatello, R., Bucolo, C., Ferrara, P., Maltese, A., Puleo, A., and Puglisi, G. (2002), Eudragit RS100(R) nanosuspensions for the ophthalmic controlled delivery of ibuprofen, *Eur. J. Pharm. Sci.*, 16(1–2), 53–61.

114. Bucolo, C., Maltese, A., Puglisi, G., and Pignatello, R. (2002), Enhanced ocular anti-inflammatory activity of ibuprofen carried by an eudragit RS100(R) nanoparticle suspension, *Ophthal. Res.*, 34(5), 319–323.

115. Pignatello, R., Bucolo, C., Spedalieri, G., Maltese, A., and Puglisi, G. (2002), Flurbiprofen-loaded acrylate polymer nanosuspensions for ophthalmic application, *Biomaterials*, 23(15), 3247–3255.

116. Khopade, A. J., and Jain, N. K. (1995), Self assembling nanostructures for sustained ophthalmic drug delivery, *Pharmazie*, 50(12), 812–814.

117. Cavalli, R., Gasco, M. R., Chetoni, P., Burgalassi, S., and Saettone, M. F. (2002), Solid lipid nanoparticles (SLN) as ocular delivery system for tobramycin, *Int. J. Pharm.*, 238(1–2), 241–245.

118. Safwat, S. M., and Al-Kassas, R. S. (2002), Evaluation of gentamicin-Eudragit microspheres as ophthalmic delivery systems in inflamed rabbit's eyes, STP Pharma *Sci.*, 12(6), 357–361.

119. Giunchedi, P., Conte, U., Chetoni, P., and Saettone, M. F. (1999), Pectin microspheres as ophthalmic carriers for piroxicam: Evaluation in vitro and in vivo in albino rabbits, *Eur. J. Pharm. Sci.*, 9(1), 1–7.

120. Giunchedi, P., Chetoni, P., Conte, U., and Saettone, M. F. (2000), Albumin microspheres for ocular delivery of piroxicam, *Pharm. Pharmacol. Commun.*, 6(4), 149–153.

121. Davies, N. M., Farr, S. J., Hadgraft, J., and Kellaway, I. W. (1992), Evaluation of mucoadhesive polymers in ocular drug delivery. II. Polymer-coated vesicles, *Pharm. Res.*, 9(9), 1137–1144.

122. Durrani, A. M., Davies, N. M., Thomas, M., and Kellaway, I. W. (1992), Pilocarpine bioavailability from a mucoadhesive liposomal ophthalmic drug delivery system, *Int. J. Pharm.*, 88(1–3), 409–415.
123. Henriksen, I., Green, K. L., Smart, J. D., Smistad, G., and Karlsen, J. (1996), Bioadhesion of hydrated chitosans: An in vitro and in vivo study, *Int. J. Pharm.*, 145(1–2), 231–240.
124. Bochot, A., Fattal, E., Gulik, A., Couarraze, G., and Couvreur, P. (1998), Liposomes dispersed within a thermosensitive gel: A new dosage form for ocular delivery of oligonucleotides, *Pharm. Res.*, 15(9), 1364–1369.
125. Bochot, A., Fattal, E., Grossiord, J. L., Puisieux, F., and Couvreur, P. (1998), Characterization of a new ocular delivery system based on a dispersion of liposomes in a thermosensitive gel, *Int. J. Pharm.*, 162(1–2), 119–127.
126. Bochot, A., Mashhour, B., Puisieux, F., Couvreur, P., and Fattal, E. (1998), Comparison of the ocular distribution of a model oligonucleotide after topical instillation in rabbits of conventional and new dosage forms, *J. Drug Target.*, 6(4), 309–313.
127. Nagarsenker, M. S., Londhe, V. Y., and Nadkarni, G. D. (1999), Preparation and evaluation of liposomal formulations of tropicamide for ocular delivery, *Int. J. Pharm.*, 190(1), 63–71.
128. Aggarwal, D., and Kaur, I. P. (2005), Improved pharmacodynamics of timolol maleate from a mucoadhesive niosomal ophthalmic drug delivery system, *Int. J. Pharm.*, 290(1–2), 155–159.
129. Muchtar, S., Abdulrazik, M., Frucht-Pery, J., and Benita, S. (1997), Ex-vivo permeation study of indomethacin from a submicron emulsion through albino rabbit cornea, *J. Controlled Release*, 44(1), 55–64.
130. Sznitowska, M., Zurowska-Pryczkowska, K., Dabrowska, E., and Janicki, S. (2000), Increased partitioning of pilocarpine to the oily phase of submicron emulsion does not result in improved ocular bioavailability, *Int. J. Pharm.*, 202(1–2), 161–164.
131. Haße, A., and Keipert, S. (1997), Development and characterisation of microemulsions for ocular application, *Eur. J. Pham. Biopharm.*, 43(2), 179–183.
132. Radomska, A., and Dobrucki, R. (2000), The use of some ingredients for microemulsion preparation containing retinol and its esters, *Int. J. Pharm.*, 196(2), 131–134.
133. Alany, R. G., Rades, T., Nicoll, J., Tucker, I. G., and Davies, N. M. (2006), W/O microemulsions for ocular delivery: Evaluation of ocular irritation and precorneal retention, *J. Controlled Release*, 111(1–2), 145–152.
134. Kaback, M. B., Podos, S. M., Harbin, T. S.Jr, ., Mandell, A., and Becker, B. (1976), The effects of dipivalyl epinephrine on the eye, *Am. J. Ophthalmol.*, 81(6), 768–772.
135. Wei, C. P., Anderson, J. A., and Leopold, I. (1978), Ocular absorption and metabolism of topically applied epinephrine and a dipivalyl ester of epinephrine, *Invest. Ophthalmol. Vis. Sci.*, 17(4), 315–321.
136. Jarvinen, T., Suhonen, P., Auriola, S., Vepsalainen, J., Peura, P., et al. (1991), *O*,*O′*-(1,4-Xylylene)bispilocarpic acid esters as new potential double prodrugs of pilocarpine for improved ocular delivery. Part 1. Synthesis and analysis, *Int. J. Pharm.*, 75(2–3), 249–258.
137. Tirucherai, G. S., Dias, C., and Mitra, A. K. (2002), Corneal permeation of ganciclovir: Mechanism of ganciclovir permeation enhancement by acyl ester prodrug design, *J. Ocul. Pharmacol. Ther.*, 18(6), 535–548.
138. Saettone, M. F., Chetoni, P., Cerbai, R., Mazzanti, G., and Braghiroli, L. (1996), Evaluation of ocular permeation enhancers: In vitro effects on corneal transport of four beta-blockers, and in vitro/in vivo toxic activity, *Int. J. Pharm.*, 142(1), 103–113.

139. Maestrelli, F., Mura, P., Casini, A., Mincione, F., Scozzafava, A., and Supuran, C. T. (2002), Cyclodextrin complexes of sulfonamide carbonic anhydrase inhibitors as long-lasting topically acting antiglaucoma agents, *J. Pharm. Sci.*, 91(10), 2211–2219.
140. Nijhawan, R., and Agarwal, S. P. (2003), Development of an ophthalmic formulation containing ciprofloxacin-hydroxypropyl-β-cyclodextrin complex, *Boll. Chim. Farm.*, 142(5), 214–219.
141. Aktas, Y., Unlu, N., Orhan, M., Irkec, M., and Hincal, A. A. (2003), Influence of hydroxypropyl beta-cyclodextrin on the corneal permeation of pilocarpine, *Drug Dev. Ind. Pharm.*, 29(2), 223–230.
142. Harwood, R. J., and Schwartz, J. B. (1982), Drug release from compression molded films: Preliminary studies with pilocarpine, *Drug Dev. Ind. Pharm.*, 8(5), 663–682.
143. Hsiue, G. H., Guu, J. A., and Cheng, C. C. (2001), Poly(2-hydroxyethyl methacrylate) film as a drug delivery system for pilocarpine, *Biomaterials*, 22(13), 1763–1769.
144. Bharath, S., and Hiremath, S. R. (1999), Ocular delivery systems of pefloxacin mesylate, *Pharmazie*, 54(1), 55–58.
145. Sultana, Y., Aqil, M., and Ali, A. (2005), Ocular inserts for controlled delivery of pefloxacin mesylate: Preparation and evaluation, *Acta Pharma.*, 55(3), 305–314.
146. Zimmermann, C., Drewe, J., Flammer, J., and Shaarawy, T. (2004), In vitro release of mitomycin C from collagen implants, *Curr. Eye Res.*, 28(1), 1–4.
147. Vasantha, V., Sehgal, P. K., and Rao, K. P. (1988), Collagen ophthalmic inserts for pilocarpine drug delivery system, *Int. J. Pharm.*, 47(1–3), 95–102.
148. Saettone, M. F., Giannaccini, B., Chetoni, P., Galli, G., and Chiellini, E. (1984), Vehicle effects in ophthalmic bioavailability: An evaluation of polymeric inserts containing pilocarpine, *J. Pharm. Pharmacol.*, 36(4), 229–234.
149. Saettone, M. F., Torracca, M. T., Pagano, A., Giannaccini, B., Rodriguez, L., and Cini, M. (1992), Controlled release of pilocarpine from coated polymeric ophthalmic inserts prepared by extrusion, *Int. J. Pharm.*, 86(2–3), 159–166.
150. Saettone, M. F., Chetoni, P., Bianchi, L. M., Giannaccini, B., Conte, U., and Sangalli, M. E. (1995), Controlled release of timolol maleate from coated ophthalmic mini-tablets prepared by compression, *Int. J. Pharm.*, 126(1–2), 79–82.
151. Sasaki, H., Tei, C., Nishida, K., and Nakamura, J. (1993), Drug release from an ophthalmic insert of a beta-blocker as an ocular drug delivery system, *J. Controlled Release*, 27(2), 127–137.
152. Sasaki, H., Nagano, T., Sakanaka, K., Kawakami, S., Nishida, K., Nakamura, J., et al. (2003), One-side-coated insert as a unique ophthalmic drug delivery system, *J. Controlled Release*, 92(3), 241–247.
153. Charoo, N. A., Kohli, K., Ali, A., and Anwer, A. (2003), Ophthalmic delivery of ciprofloxacin hydrochloride from different polymer formulations: In vitro and in vivo studies, *Drug Dev. Ind. Pharm.*, 29(2), 215–221.
154. Pijls, R. T., Sonderkamp, T., Daube, G. W., Krebber, R., Hanssen, H. H. L., Nuijts, R. M. M. A., et al. (2005), Studies on a new device for drug delivery to the eye, *Eur. J. Pharm. Biopharm.*, 59(2), 283–288.
155. Lux, A., Maier, S., Dinslage, S., Suverkrup, R., and Diestelhorst, M. (2003), A comparative bioavailability study of three conventional eye drops versus a single lyophilisate, *Br. J. Ophthalmol.*, 87(4), 436–440.
156. Dinslage, S., Diestelhorst, M., Weichselbaum, A., and Suverkrup, R. (2002), Lyophilisates for drug delivery in Ophthalmology: Pharmacokinetics of fluorescein in the human anterior segment, *Br. J. Ophthalmol.*, 86(10), 1114–1117.
157. Ceulemans, J., Vermeire, A., Adriaens, E., Remon, J. P., and Ludwig, A. (2001), Evaluation of a mucoadhesive tablet for ocular use, *J. Controlled Release*, 77(3), 333–344.

158. Weyenberg, W., Vermeire, A., Remon, J. P., and Ludwig, A. (2003), Characterization and in vivo evaluation of ocular bioadhesive minitablets compressed at different forces, *J. Controlled Release*, 89(2), 329–340.

159. Baeyens, V., Kaltsatos, V., Boisrame, B., Fathi, M., and Gurny, R. (1998), Evaluation of soluble Bioadhesive Ophthalmic Drug Inserts (BODI) for prolonged release of gentamicin: Lachrymal pharmacokinetics and ocular tolerance, *J. Ocul. Pharmacol. Ther.*, 14(3), 263–272.

160. Baeyens, V., Felt-Baeyens, O., Rougier, S., Pheulpin, S., Boisrame, B., and Gurny, R. (2002), Clinical evaluation of bioadhesive ophthalmic drug inserts (BODI(R)) for the treatment of external ocular infections in dogs, *J. Controlled Release*, 85(1–3), 163–168.

161. Chetoni, P., Di Colo, G., Grandi, M., Morelli, M., Saettone, M. F., and Darougar, S. (1998), Silicone rubber/hydrogel composite ophthalmic inserts: Preparation and preliminary in vitro/in vivo evaluation, *Eur. J. Pharm. Biopharm.*, 46(1), 125–132.

162. Di Colo, G., and Zambito, Y. (2002), A study of release mechanisms of different ophthalmic drugs from erodible ocular inserts based on poly(ethylene oxide), *Eur. J. Pharm. Biopharm.*, 54(2), 193–199.

163. Di Colo, G., Burgalassi, S., Chetoni, P., Fiaschi, M. P., Zambito, Y., and Saettone, M. F. (2001), Relevance of polymer molecular weight to the in vitro/in vivo performances of ocular inserts based on poly(ethylene oxide), *Int. J. Pharm.*, 220(1–2), 169–177.

164. Di Colo, G., Burgalassi, S., Chetoni, P., Fiaschi, M. P., Zambito, Y., and Saettone, M. F. (2001), Gel-forming erodible inserts for ocular controlled delivery of ofloxacin, *Int. J. Pharm.*, 215(1–2), 101–111.

165. Di Colo, G., Zambito, Y., Burgalassi, S., Serafini, A., and Saettone, M. F. (2002), Effect of chitosan on in vitro release and ocular delivery of ofloxacin from erodible inserts based on poly(ethylene oxide), *Int. J. Pharm.*, 248(1–2), 115–122.

166. Hornof, M., Weyenberg, W., Ludwig, A., and Bernkop-Schnurch, A. (2003), Mucoadhesive ocular insert based on thiolated poly(acrylic acid): Development and in vivo evaluation in humans, *J. Controlled Release*, 89(3), 419–428.

167. Frucht-Pery, J., Mechoulam, H., Siganos, C. S., Ever-Hadani, P., Shapiro, M., and Domb, A. (2004), Iontophoresis-gentamicin delivery into the rabbit cornea, using a hydrogel delivery probe, *Exper. Eye Res.*, 78(3), 745–749.

168. Vandamme, T. F., and Brobeck, L. (2005), Poly(amidoamine) dendrimers as ophthalmic vehicles for ocular delivery of pilocarpine nitrate and tropicamide, *J. Controlled Release*, 102(1), 23–38.

169. Lang, J. C. (1995), Ocular drug delivery conventional ocular formulations, *Adv. Drug Deliv. Rev.*, 16(1), 39–43.

170. Fitzgerald, P., and Wilson, C. G. (1994), Polymeric systems for ophthalmic drug delivery, in Dimitriuitra S., Ed., *Polymeric Biomaterials*, Marcel Dekker, New York, pp. 373–398.

171. Kupferman, A., Pratt, M. V., Suckewer, K., and Leibowitz, H. M. (1974), Topically applied steroids in corneal disease. 3. The role of drug derivative in stromal absorption of dexamethasone, *Arch. Ophthalmol.*, 91(5), 373–376.

172. Wilson, C. G., Olejnik, O., and Hardy, J. G. (1983), Precorneal drainage of polyvinyl alcohol solutions in the rabbit assessed by gamma scintigraphy, *J. Pharm. Pharmacol.*, 35(7), 451–454.

173. Olejnik, O., and Weisbecker, C. A. (1990), Ocular bioavailability of topical prednisolone preparations, *Clin. Ther.*, 12(1), 2–11.

174. Davies, N. M. (2000), Biopharmaceutical considerations in topical ocular drug delivery, *Clin. Exp. Pharmacol. Physiol.*, 27(7), 558–562.

175. Davies, N. M., Wang, G., and Tucker, I. G. (1997), Evaluation of a hydrocortisone/hydroxypropyl-[beta]-cyclodextrin solution for ocular drug delivery, *Int. J. Pharm.*, 156(2), 201–209.

176. Olejnik, O. (1993), Conventional systems in ophthalmic drug delivery, in Mitra, A. K., Ed., *Ophthalmic Drug Delivery Systems*, Marcel Dekker, New York, pp. 177–198.

177. Schoenwald, R. D., and Stewart, P. (1980), Effect of particle size on ophthalmic bioavailability of dexamethasone suspensions in rabbits, *J. Pharm. Sci.*, 69(4), 391–394.

178. MacKeen, D. L. (1980), Aqueous formulations and ointments, *Int. Ophthalmol. Clin.*, 20(3), 79–92.

179. Sieg, J. W., and Robinson, J. R. (1977), Vehicle effects on ocular drug bioavailability II: Evaluation of pilocarpine, *J. Pharm. Sci.*, 66(9), 1222–1228.

180. Greaves, J. L., Wilson, C. G., and Birmingham, A. T. (1993), Assessment of the precorneal residence of an ophthalmic ointment in healthy subjects, *Br. J. Clin. Pharmacol.*, 35(2), 188–192.

181. Lee, V. H., and Robinson, J. R. (1986), Topical ocular drug delivery: Recent developments and future challenges, *J. Ocul. Pharmacol.*, 2(1), 67–108.

182. Saettone, M. F., Giannaccini, B., Teneggi, A., Savigni, P., and Tellini, N. (1982), Vehicle effects on ophthalmic bioavailability: The influence of different polymers on the activity of pilocarpine in rabbit and man, *J. Pharm. Pharmacol.*, 34(7), 464–466.

183. Trueblood, J. H., Rossomondo, R. M., Wilson, L. A., and Carlton, W. H. (1975), Corneal contact times of ophthalmic vehicles. Evaluation by microscintigraphy, *Arch. Ophthalmol.*, 93(2), 127–130.

184. Patton, T. F., and Robinson, J. R. (1976), Quantitative precorneal disposition of topically applied pilocarpine nitrate in rabbit eyes, *J. Pharm. Sci.*, 65(9), 1295–1301.

185. Patton, T. F., and Robinson, J. R. (1975), Ocular evaluation of polyvinyl alcohol vehicle in rabbits, *J. Pharm. Sci.*, 64(8), 1312–1316.

186. Tiffany, J. M. (1991), The viscosity of human tears, *Int. Ophthalmol.*, 15(6), 371–376.

187. Van Ooteghem, M. (1987), Factors influencing the retention of ophthalmic solutions on the eye surface, in Saettone, M. F., Bucci, M., and Speiser, P., Eds., *Ophthalmic Drug Delivery: Biopharmaceutical, Technological and Clinical Aspects*, Vol. 11, Livinia, Padova, pp. 7–17

188. Ludwig, A., van Haeringen, N. J., Bodelier, V. M., and Van Ooteghem, M. (1992), Relationship between precorneal retention of viscous eye drops and tear fluid composition, *Int. Ophthalmol.*, 16(1), 23–26.

189. Greaves, J. L., Olejnik, O., and Wilson, C. G. (1992), Polymers and the precorneal tear film, *STP Pharma Sci.*, 2, 13–33.

190. Park, H., and Robinson, J. R. (1987), Mechanisms of mucoadhesion of poly(acrylic acid) hydrogels, *Pharm. Res.*, 4(6), 457–464.

191. Robinson, J. R., and Mlynek, G. M. (1995), Bioadhesive and phase-change polymers for ocular drug delivery, *Adv. Drug Deliv. Rev.*, 16(1), 45–50.

192. Lehr, C.-M., Bouwstra, J. A., Schacht, E. H., and Junginger, H. E. (1992), In vitro evaluation of mucoadhesive properties of chitosan and some other natural polymers, *Int. J. Pharm.*, 78(1–3), 43–48.

193. Park, K., and Robinson, J. R. (1984), Bioadhesive polymers as platforms for oral-controlled drug Delivery: Method to study bioadhesion, *Int. J. Pharm.*, 19(2), 107–127.

194. Meseguer, G., Gurny, R., Buri, P., Rozier, A., and Plazonnet, B. (1993), Gamma scintigraphic study of precorneal drainage and assessment of miotic response in rabbits

of various ophthalmic formulations containing pilocarpine, *Int. J. Pharm.*, 95(1–3), 229–234.

195. Thermes, F., Rozier, A., Plazonnet, B., and Grove, J. (1992), Bioadhesion: The effect of polyacrylic acid on the ocular bioavailability of timolol, *Int. J. Pharm.*, 81(1), 59–65.
196. Duchene, D., Touchard, F., and Peppas, N. A. (1988), Pharmaceutical and medical aspects of bioadhesive systems for drug administration, *Drug Dev. Ind. Pharm.*, 14(2–3), 283–318.
197. Calonge, M. (2001), The treatment of dry eye, *Surv. Ophthalmol.*, 45(Suppl. 2), S227–S239.
198. Zignani, M., Tabatabay, C., and Gurny, R. (1995), Topical semi-solid drug delivery: Kinetics and tolerance of ophthalmic hydrogels, *Adv. Drug Deliv. Rev.*, 16(1), 51–60.
199. Krauland, A. H., Leitner, V. M., and Bernkop-Schnurch, A. (2003), Improvement in the in situ gelling properties of deacetylated gellan gum by the immobilization of thiol groups, *J. Pharm. Sci.*, 92(6), 1234–1241.
200. Gurny, R., Ibrahim, H., and Buri, P. (1993), The development and use of in situ formed gels, triggered by pH, in Edman, P., Ed., *Biopharmaceutics of Ocular drug Delivery*, CRC Press, Boca Raton, FL, pp. 81–90.
201. Greaves, J. L., Wilson, C. G., Rozier, A., Grove, J., and Plazonnet, B. (1990), Scintigraphic assessment of an ophthalmic gelling vehicle in man and rabbit, *Curr. Eye Res.*, 9(5), 415–420.
202. Furrer, P., Plazonnet, B., Mayer, J. M., and Gurny, R. (2000), Application of in vivo confocal microscopy to the objective evaluation of ocular irritation induced by surfactants, *Int. J. Pharm.*, 207(1–2), 89–98.
203. Kumar, S., Haglund, B. O., and Himmelstein, K. J. (1994), In situ-forming gels for ophthalmic drug delivery, *J. Ocul. Pharmacol.*, 10(1), 47–56.
204. Calvo, P., Alonso, M. J., Vila-Jato, J. L., and Robinson, J. R. (1996), Improved ocular bioavailability of indomethacin by novel ocular drug carriers, *J. Pharm. Pharmacol.*, 48(11), 1147–1152.
205. Lallemand, F., Felt-Baeyens, O., Besseghir, K., Behar-Cohen, F., and Gurny, R. (2003), Cyclosporine A delivery to the eye: A pharmaceutical challenge, *Eur. J. Pharm. Biopharm.*, 56(3), 307–318.
206. Kreuter, J. (1990), Nanoparticles as bioadhesive ocular drug delivery systems, in Lenaerts, V., and Gurny, R., Eds., *Bioadhesive Drug Delivery Systems*, CRC Press, Boca Raton, FL, pp. 203–212.
207. Mainardes, R. M., and Silva, L. P. (2004), Drug delivery systems: Past, present, and future, *Curr. Drug Targets*, 5(5), 449–455.
208. Ding, S. (1998), Recent developments in ophthalmic drug delivery, *Pharm. Sci. Technol. Today*, 1(8), 328–335.
209. Kaur, I. P., and Kanwar, M. (2002), Ocular preparations: The formulation approach, *Drug Dev. Ind. Pharm.*, 28(5), 473–493.
210. Meisner, D., and Mezei, M. (1995), Liposome ocular delivery systems, *Adv. Drug Deliv. Rev.*, 16(1), 75–93.
211. El-Gazayerly, O. N., and Hikal, A. H. (1997), Preparation and evaluation of acetazolamide liposomes as an ocular delivery system, *Int. J. Pharm.*, 158(2), 121–127.
212. Fresta, M., Panico, A. M., Bucolo, C., Giannavola, C., and Puglisi, G. (1999), Characterization and in-vivo ocular absorption of liposome-encapsulated acyclovir, *J. Pharm. Pharmacol.*, 51(5), 565–576.

213. Law, S. L., Huang, K. J., and Chiang, C. H. (2000), Acyclovir-containing liposomes for potential ocular delivery: Corneal penetration and absorption, *J. Controlled Release*, 63(1–2), 135–140.

214. Alonso, M. J., and Sanchez, A. (2003), The potential of chitosan in ocular drug delivery, *J. Pharm. Pharmacol.*, 55(11), 1451–1463.

215. Carafa, M., Santucci, E., and Lucania, G. (2002), Lidocaine-loaded non-ionic surfactant vesicles: Characterization and in vitro permeation studies, *Int. J. Pharm.*, 231(1), 21–32.

216. Vyas, S. P., Mysore, N., Jaitely, V., and Venkatesan, N. (1998), Discoidal niosome based controlled ocular delivery of timolol maleate, *Pharmazie*, 53(7), 466–469.

217. Schmalfuss, U., Neubert, R., and Wohlrab, W. (1997), Modification of drug penetration into human skin using microemulsions, *J. Controlled Release*, 46(3), 279–285.

218. Bagwe, R. P., Kanicky, J. R., Palla, B. J., Patanjali, P. K., and Shah, D. O. (2001), Improved drug delivery using microemulsions: Rationale, recent progress, and new horizons, *Crit. Rev. Ther. Drug Carrier Syst.*, 18(1), 77–140.

219. Attwood, D. (1994), Microemulsions, in Kreuter, J., Ed., *Colloidal Drug Delivery Systems*, Marcel Dekker, New York, pp. 31–71.

220. Benita, S., and Levy, M. Y. (1993), Submicron emulsions as colloidal drug carriers for intravenous administration: Comprehensive physicochemical characterization, *J. Pharm. Sci.*, 82(11), 1069–1079.

221. Gasco, M. R., Gallarate, M., Trotta, M., Bauchiero, L., Chiappero, O., et al. (1989), Microemulsions as topical delivery vehicles: Ocular administration of timolol, *J. Pharm. Biomed. Anal.*, 7(4), 433–439.

222. Beilin, M., Bar-Ilan, A., Amselem, S., Schwarz, J., Yogev, A., and Neumann, R. (1995), Ocular retention time of submicron emulsion (SME) and the miotic response to pilocarpine delivered in SME, *Invest. Ophthalmol. Vis. Sci.*, 36(4), 166.

223. Esfand, R., and Tomalia, D. A. (2001), Poly(amidoamine) (PAMAM) dendrimers: From biomimicry to drug delivery and biomedical applications, *Drug Discov. Today*, 6(8), 427–436.

224. Cloninger, M. J. (2002), Biological applications of dendrimers, *Curr. Opin. Chem. Biol.*, 6(6), 742–748.

225. Patri, A. K., Majoros, I. J., Baker, J., and James R. (2002), Dendritic polymer macromolecular carriers for drug delivery, *Curr. Opin. Chem. Biol.*, 6(4), 466–471.

226. Stella, V. J., and Himmelstein, K. J. (1980), Prodrugs and site-specific drug delivery, *J. Med. Chem.*, 23(12), 1275–1282.

227. Lee, V. H. (1993), Improved ocular drug delivery by use of chemical modifications (prodrugs), in Edman, P., Ed., *Biopharmaceutics of Ocular Drug Delivery*, CRC Press, Boca Raton, FL, pp. 121–143.

228. Lee, V. H. (1983), Esterase activities in adult rabbit eyes, *J. Pharm. Sci.*, 72(3), 239–244.

229. Redell, M. A., Yang, D. C., and Lee, V. H. L. (1983), The role of esterase activity in the ocular disposition of dipivalyl epinephrine in rabbits, *Int. J. Pharm.*, 17(2–3), 299–312.

230. Jarvinen, T., and Jarvinen, K. (1996), Prodrugs for improved ocular drug delivery, *Adv. Drug Deliv. Rev.*, 19(2), 203–224.

231. Sasaki, H., Igarashi, Y., Nagano, T., Nishida, K., and Nakamura, J. (1995), Different effects of absorption promoters on corneal and conjunctival penetration of ophthalmic beta-blockers, *Pharm. Res.*, 12(8), 1146–1150.

232. Hochman, J., and Artursson, P. (1994), Mechanisms of absorption enhancement and tight junction regulation, *J. Controlled Release*, 29(Mar.), 253–267.

233. Grass, G. M., and Robinson, J. R. (1984), Relationship of chemical structure to corneal penetration and influence of low-viscosity solution on ocular bioavailability, *J. Pharm. Sci.*, 73(8), 1021–1027.
234. Szejtli, J. (1994), Medicinal applications of cyclodextrins, *Med. Res. Rev.*, 14(3), 353–386.
235. Sasaki, H., Yamamura, K., Nishida, K., Nakamura, J., and Ichikawa, M. (1996), Delivery of drugs to the eye by topical application, *Prog. Retin. Eye Res.*, 15(2), 583–620.
236. Jarvinen, K., Jarvinen, T., Thompson, D. O., and Stella, V. J. (1994), The effect of a modified beta-cyclodextrin, SBE4-beta-CD, on the aqueous stability and ocular absorption of pilocarpine, *Curr. Eye Res.*, 13(12), 897–905.
237. Jarho, P., Jarvinen, K., Urtti, A., Stella, V. J., and Jarvinen, T. (1996), Modified beta-cyclodextrin (SBE7-beta-CyD) with viscous vehicle improves the ocular delivery and tolerability of pilocarpine prodrug in rabbits, *J. Pharm. Pharmacol.*, 48(3), 263–269.
238. Sultana, Y., Jain, R., Aqil, M., and Ali, A. (2006), Review of ocular drug delivery, *Curr. Drug Deliv.*, 3(2), 207–217.
239. Ranade, V. V., and Hollinger, M. A. (2003), *Intranasal and ocular drug delivery. in Drug Delivery Systems*, 2nd ed., CRC Press, Boca Raton, FL, pp. 249–287.
240. Lee, V. H. (1990), New directions in the optimization of ocular drug delivery, *J. Ocul. Pharmacol.*, 6(2), 157–164.

# MICROEMULSIONS AS DRUG DELIVERY SYSTEMS

Raid G. Alany and Jingyuan Wen
*The University of Auckland Auckland, New Zealand*

**Contents**

1 Historical Background, Terminology, and Definition
2 Structure and Formation of Microemulsion Systems
3 Role of Cosurfactants/Cosolvents in Formation and Stabilization of Microemulsion Systems
4 Pharmaceutical Formulation of Microemulsions
  4.1 Selection of Microemulsion Ingredients
  4.2 Phase Behavior Studies
5 Techniques Used to Characterize Microemulsions and Related Systems
  5.1 Polarized Light Microscopy
  5.2 Transmission Electron Microscopy
  5.3 Electrical Conductivity Measurements
  5.4 Viscosity Measurements
  5.5 Other Characterization Techniques
6 Microemulsions as Drug Delivery Systems
  6.1 Oral Drug Delivery
  6.2 Transdermal Drug Delivery
  6.3 Parenteral Drug Delivery
  6.4 Ocular Drug Delivery
7 Concluding Remarks
  References

## 1 HISTORICAL BACKGROUND, TERMINOLOGY, AND DEFINITION

Hoar and Schulman coined the term *microemulsion* (ME) in 1943 to define a transparent system obtained by titrating a turbid oil-in-water (o/w) emulsion with a

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

**TABLE 1 Comparison of Emulsions and Microemulsions**

| Property | Emulsion | Microemulsion |
|---|---|---|
| Disperse-phase droplet size | 0.2–10 μm | Less than 0.2 μm |
| Visual appearance | Turbid to milky | Transparent to translucent |
| Stability | Thermodynamically unstable | Thermodynamically stable |
| Formation | Requires energy input | Spontaneous |

*Source*: From ref. 2.

medium-chain alcohol, namely hexanol. Since then the term has been used to describe systems comprising a nonpolar component, an aqueous component, a surfactant, and a cosurfactant. While a cosurfactant is usually present, a ME can be formulated without a cosurfactant, that is, using a single surfactant. It is important to point out that the term ME was (and occasionally is) used in the literature to describe various liquid crystalline systems (lamellar, hexagonal, and cubic), surfactant systems (micelles and reverse micelles), and even coarse emulsions that are micronized using external energy (submicrometer emulsions). To avoid such confusion, Danielsson and Lindmann [1] proposed the following definition: " Microemulsion is defined as a system of water, oil and amphiphile which is optically isotropic and a thermodynamically stable liquid solution." By this definition the following systems were excluded:

- Aqueous solutions of surfactants (micellar and nonmicellar) without additives or with water soluble nonelectrolytes as additives
- Liquid crystalline phases (mesophases)
- Coarse emulsions, including micronized coarse emulsions
- Systems that are surfactant free

The term ME is often incorrectly used in the literature to describe oil and water dispersions of small droplet size produced by prolonged ultrasound mixing, high-shear homogenization, and microfluidisation, that is, submicrometer emulsions. The major differences between a microemulsion and a coarse emulsion are shown in Table 1.

## 2 STRUCTURE AND FORMATION OF MICROEMULSION SYSTEMS

A ME system can be one of three types depending on the composition: oil in water (o/w ME), in which water is the continuous medium; water in oil (w/o ME), in which oil is the continuous medium, and water-and-oil bicontinuous ME, in which almost equal amounts of water and oil exist [3]. While the three types are quite different in terms of microstructure, they all have an interfacial amphiphile monolayer separating the oil and water domains.

The formation of a ME system can be explained using a simplified thermodynamic approach and with reference to the equation

$$\Delta G = \gamma \Delta A - T \Delta S \tag{1}$$

where $\Delta G$-free energy of ME formation
$\gamma$-interfacial tension at oil–water interface
$\Delta A$-change in interfacial area (associated with reducing droplet size)
$S$-system entropy
$T$-absolute temperature

The process of ME formation is associated with a reduction in droplet size, which results in an increase in the value of $\Delta A$ due to an overall increase in surface area that is associated with droplet size reduction. This is compensated by a very low interfacial tension that is normally achieved by using relatively high amphiphile concentrations. Furthermore, the process of ME formation is accompanied by a favorable entropy contribution (increased value of $\Delta S$) that is due to the mixing of the two immiscible phases, surfactant molecules partitioning in favor of the interface rather than the bulk and monomer–micelle surfactant exchange. The net outcome is a negative value for $\Delta G$ which translates into a spontaneous ME formation [4].

The simplest representation of the ME microstructure is with reference to the droplet model in which an interfacial film consisting of amphiphile (surfactant/cosurfactant) molecules surrounds the dispersed droplets (Figures 1*a* and *b*). The orientation of the amphiphile at the interface differs depending on the type of the ME. Whether an o/w or w/o ME forms is dependent to a great extent on the volume fraction of oil and water as well as the nature of the interfacial film as reflected by the geometry of the amphiphile molecules forming the film. It follows that the

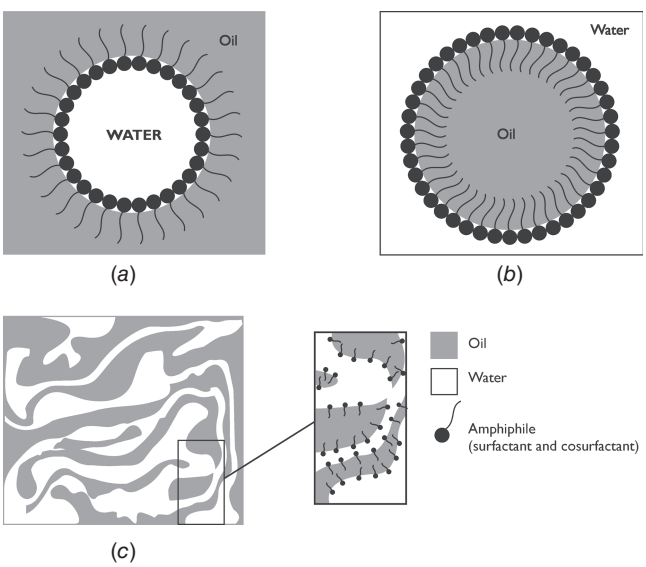


**FIGURE 1** Diagrammatic representation of different types of ME systems: (*a*) w/o ME; (*b*) o/w ME; (*c*) water-and-oil bicontinuous ME. Droplet diameter for (*a*) and (*b*) is typically less than 140 nm.

presence of o/w ME droplets is more likely to happen in systems where the oil volume fraction is low, whereas w/o ME droplets form when the water volume fraction is low and oil is present in abundance. Interestingly, in systems containing comparable amounts of water and oil, a bicontinuous ME may exist (Figure 1*c*). In such systems both oil and water exist as microdomains that are separated by an amphiphile-stabilized interface with a zero net curvature.

Mitchell and Ninham [5] extended the theory of self-assembly of surfactant molecules forming micelles and bilayers [6] to ME systems. Accordingly, if the volume of the surfactant is $v$, its head group surface area $a$, and its length $l$, it follows that when the critical packing parameter (CPP = $v/al$) has values between 0 and 1, o/w MEs are likely to be formed. On the other hand, when CPP is greater than 1, w/o MEs are favored. When using surfactants with critical packing parameters close to unity (CPP $\approx$ 1) and at approximately equal volumes of water and oil, the mean curvature of the interfacial film approaches zero and droplets may merge into a bicontinuous structure (Figure 1*c*). It should be noted, however, that this approach is based solely on geometric considerations and does not account for penetration of oil and cosurfactant molecules into the interface and the hydration of surfactant head groups.

The ratio of the hydrophilic and the hydrophobic groups of the surfactant molecules, that is, their hydrophile–lipophile balance (HLB), is also important in determining interfacial film curvature and consequently the structure of the ME. The HLB system has been used for the selection of surfactants to formulate MEs and accordingly the HLB of the candidate surfactant blend should match the required HLB of the oily component for a particular system; furthermore a match in the lipophilic part of the surfactant used with the oily component is favorable [7].

Shinoda and Kuineda [8] highlighted the effect of temperature on the phase behavior of systems formulated with two surfactants and introduced the concept of the phase inversion temperature (PIT) or the so-called HLB temperature. They described the recommended formulation conditions to produce MEs with surfactant concentration of about 5–10% w/w being (a) the optimum HLB or PIT of a surfactant; (b) the optimum mixing ratio of surfactants, that is, the HLB or PIT of the mixture; and (c) the optimum temperature for a given nonionic surfactant. They concluded that (a) the closer the HLBs of the two surfactants, the larger the cosolubilization of the two immiscible phases; (b) the larger the size of the solubilizer, the more efficient the solubilisation process; and (c) mixtures of ionic and nonionic surfactants are more resistant to temperature changes than nonionic surfactants alone.

## 3 ROLE OF COSURFACTANTS/COSOLVENTS IN FORMATION AND STABILIZATION OF MICROEMULSIONS

Cosurfactants are molecules with weak amphiphilic properties that are mixed with the surfactant(s) to enhance their ability to reduce the interfacial tension of a system and promote the formation of a ME [3]. Cosolvents have also been described as weak amphiphilic molecules that tend to distribute between the aqueous phase, the

oily phase, and the interfacial layer and act by making the aqueous phase less hydrophilic, the oily phase less hydrophobic, and the interfacial film more flexible and less condensed [9, 10].

Most single-chain surfactants do not sufficiently lower the oil–water interfacial tension to form MEs, nor are they of the right molecular structure (i.e., HLB) to act as cosolvents. To overcome such a barrier, cosurfactant/cosolvent molecules are added to further lower the interfacial tension between oil and water, fluidize the hydrocarbon region of the interfacial film, and influence the curvature of the film. Typically small molecules (C3–C8) with a polar head (hydroxyl or amine) group that can diffuse between the bulk oil and water phase and the interfacial film are suitable candidates [11].

All the abovementioned mechanisms are expected to facilitate the formation and stabilization of ME systems.

## 4 PHARMACEUTICAL FORMULATION OF MICROEMULSIONS

### 4.1 Selection of Microemulsion Ingredients

Pharmaceutically acceptable ME systems are formulated using at least GRAS- (generally regarded as safe) and preferably pharmaceutical-grade ingredients, that is, ones already in use in pharmaceutical formulation and devoid of serious side effects and toxicity in humans [12]. Nonionic and zwitterionic surfactants are among the most commonly used ingredients to formulate pharmaceutical MEs while vegetable oils, medium- and long-chain triglycerides, and esters of fatty acids are the most commonly used oils [2]. Among the range of nonionic surfactants used are sucrose esters [13], polyoxyethylene alkyl ethers [14], polyglycerol fatty acid esters [15], polyoxyethylene hydrogenated castor oil [16], and sorbitan esters [17]. Furthermore, systems formulated with zwitterionic phospholipids, particularly lecithin, have been widely investigated because of their biocompatible nature [9, 18–22].

The effect of the oily component on the phase behavior of o/w ME-forming systems formulated with nonionic surfactants was reported [23]. The authors showed that it is possible to formulate cosurfactant-free o/w ME systems suitable for use as drug delivery vehicles using either polyoxyethylene surfactants or amine-*N*-oxide surfactants. The major advantage of these ME systems is their ability to be diluted without destroying their integrity; however both classes of surfactants were shown to be sensitive to electrolytes.

The choice of cosurfactants to formulate pharmaceutically acceptable MEs is challenging as most of the cosurfactants investigated in fundamental ME research cannot be used for the development of pharmaceutically acceptable systems due to biocompatibility considerations. Among the pharmaceutically acceptable cosurfactants are ethanol [24], medium-chain mono- and diglycerides [25–28], 1,2-alkanediols [29, 30] and sucrose–ethanol combinations [31], alkyl monoglucosides, and geraniol [32]. Kahlweit et al. [29, 30] reported on the usefulness of certain 1,2-alkanediol cosurfactants as nontoxic substitutes to the physiologically incompatible short- and medium-chain alcohols. They suggested the possible use of these components for the formulation of nontoxic ME systems.

### 4.2 Phase Behavior Studies

Before a ME can be used as a drug delivery vehicle, the phase behavior of the particular combination of the candidate ingredients should be established. This is necessary due to the diverse range of colloidal and coarse dispersions that could be obtained when oil, water, and an amphiphile blend are mixed. Coarse emulsions, vesicles, lyotropic liquid crystals, and micellar systems are some examples. A variety of multiphase systems may coexist and the demarcations of the regional boundaries become important. One of the most suitable methods to study the phase behavior of such systems is to construct a ternary phase diagram using a Gibbs triangle (Figure 2). A ternary phase diagram can be constructed by two methods [33]:

- Titrating a mixture of two components with the third component
- Preparing a large number of samples of different composition

If all mixtures reach equilibrium rapidly, both methods give identical results. For mixtures that do not reach equilibrium quickly, the second method is recommended, as with the titration method the change in the ratio of components during titration may occur too fast, not allowing sufficient time to visually recognize phase changes [33].

As the formulation may contain more than three components, the complete phase behavior cannot be fully represented using a triangular diagram. However,

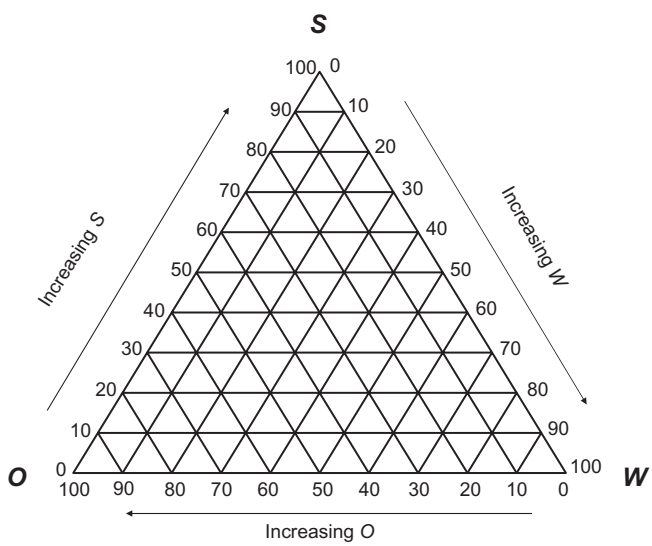


**FIGURE 2** Ternary phase diagram used to elucidate ME formation regions. Each of the three corners represents 100% of the individual components. Apex $S$ = 100% w/w surfactant (0% oil and water), apex $W$ = 100% w/w water (0% oil and surfactant), and apex $O$ = 100% w/w oil (0% water and surfactant). The three lines joining the corner points represent two-component systems. The area within the triangle represents all possible combinations of the three components.

the phase behavior of a four-component mixture at fixed pressure and temperature can be represented using a tetrahedron. Full characterization of such systems is a tedious task requiring a large number of experiments [34]. One acceptable approach for representing such systems is by fixing the mass ratio of two components (such as the two amphiphiles) and as such considered a single component. Such an approach, although regarded by many as an oversimplification of the systems, is yet acceptable for the purpose of phase behavior studies. Such systems are described as "pseudoternary" as they comprise more than three components (four or possibly five) yet are represented using a Gibbs triangle, which is a used to describe the phase behavior of a three-component system.

A novel approach to reduce the experimental effort associated with constructing pseudoternary phase diagrams is by using expert systems to predict the phase behavior of multicomponent ME-forming systems. Artificial neural networks have been investigated and were shown to be promising in phase behavior studies [17, 35, 36] as well as in the process of ingredient selection [37].

## 5 TECHNIQUES USED TO CHARACTERIZE MICROEMULSIONS AND RELATED SYSTEMS

The physicochemical and analytical techniques used to characterize MEs and related systems could be categorized into those used to:

- Elucidate the microstructure and monitor phase behavior changes
- Determine the droplet size of the disperse phase

The choice of a particular technique is limited by factors such as availability, feasibility, and the nature of the information sought. Pharmaceutical scientists are more focused on the usefulness of a particular ME system for a drug delivery application and the influence of the microstructure on that, rather than on the fundamental understanding of aspects such as microstructure and phase behavior. Polarized light microscopy is a readily available technique that could be used at the early formulation development stage to differentiate between isotropic and anisotropic systems. Transmission electron microscopy (TEM) is another available technique that has been shown to provide microstructural as well as size-related information on droplet and bicontinuous ME systems.

The main disadvantages of TEM applications (such as freeze fracture and cryo-TEM) are the lengthy and sophisticated experimental procedures associated with sample preparation and the possibility of creating artefacts during sample preparation. Other readily available and more user friendly techniques are electrical conductivity and viscosity measurements. Electrical conductivity measurements are widely used for their simplicity, feasibility, and sensitivity to structural changes in systems with increasing water content, particularly systems undergoing percolation transitions. Viscosity measurements, on the other hand, require more sophisticated instrumentation yet provide useful information on changes in the flow properties associated with structural changes of the systems. A brief overview of the key characterization techniques will follow.

### 5.1 Polarized Light Microscopy

When a mixture of oil, water, and surfactant(s) is examined under a polarized light microscope, the textures observed depend on the nature of the surfactant aggregate formed and the relative ratio of the comprising constituents. If the resulting aggregates are anisotropic, they tend to show strong birefringence and characteristic textures could be viewed when examined using a polarized light microscope. On the other hand, if the resulting aggregates are isotropic (as with ME systems or coarse emulsions), then polarized light microscopy would be of less value in disclosing structural information. Thermotropic and lyotropic liquid crystalline (LC) systems such as lamellar, hexagonal, and reverse hexagonal mesophases are anisotropic and exhibit characteristic textures when viewed under a polarizing light microscope. The only type of LC systems that is isotropic and would not display birefringence when viewed under a cross-polarizer is the cubic mesophase.

Polarized light microscopy is a simple technique to learn and use, readily available, and of great value to differentiate between various anisotropic LC systems. It is also of value to formulation scientists investigating amphiphile–oil–water mixtures with emphasis on colloidal systems in general and MEs in particular. This is mainly due to the fact that many LC systems may appear transparent to the naked eye and can be easily misinterpreted as isotropic ME systems. Thus it becomes essential when investigating systems of amphiphile–oil–water to confirm findings based on visual appearance with polarized light microscopic examination.

### 5.2 Transmission Electron Microscopy

Transmission electron microscopy was one of the earliest techniques used to investigate MEs [38, 39]. Freeze fracture along with replication is a sophisticated sample preparation method for TEM that requires careful attention to a variety of details to avoid formation of sample artefacts. The technique involves rapidly freezing the sample by immersing it in a cryogen (slush nitrogen, propane, ethane, and freons). For systems with volatile ingredients the freezing must be achieved rapidly to avoid phase separation or crystallization. Thus, high cooling rate and adequate environmental control of the samples before freezing are critical to prevent loss of volatile components. The frozen sample is then transferred to a vacuum chamber and split under vacuum by a fracturing device. The fractured surface is then shadowed with metal (usually platinum) deposited from one side. The shadowed surface is then coated with a layer of carbon that is directly deposited from above the specimen. The carbon layer is transparent when examined and forms a supportive backing for the shadowing metal deposited on the fractured surface. The specimen is then removed from the vacuum and treated with solvents of different polarities, leaving the metal carbon film as a replica of the fracture surface.

Although freeze-fracture TEM provides direct visualization of ME structures, it is not currently in wide use probably due to the experimental difficulties associated with the technique. The points to consider when preparing conventional TEM replicas are the physical and chemical sample properties, freezing, cleaving, etching, replication, cleaning, and mounting steps of the procedure.

Jahn and Strey [40] investigated systems with varying water-to-oil ratios at constant amphiphile concentration. The TEM images support the notion of a bicontinu-

ous network for systems containing comparable amounts of the aqueous and oil components. The author also showed images of w/o droplet ME systems which showed a reduction in number densities of the dispersed phase droplets upon dilution with the organic phase.

Freeze-fracture TEM combined with nuclear magnetic resonance and quasi-elastic light scattering was used to study the microstructure of surfactant–water systems and dynamics of o/w and bicontinuous ME systems [41]. The authors reported a rather abrupt transition from a discontinuous droplet (o/w) to bicontinuous (oil-and-water) microstructure occurring at low surfactant concentration, close to a three-phase region in the constructed phase diagram of pentaethylene glycol dodecyl ether, water, and octane [41].

Direct imaging of the ME microstructure using cryo-TEM involves directly investigating a thin proportion of the specimen in the frozen hydrated stage by using a cryo stage in the transmission electron microscope. Cryo-TEM was used in combination with nuclear magnetic resonance (NMR), small-angle X-ray diffraction, and small-angle neutron scattering to investigate four-component nonionic systems composed of 1-dodecane, octa-ethylene glycol mono dodecyl ether, *n*-pentanol, and water [42]. These authors reported on the existence of at least two different colloidal microstructures, swollen spherical micelles with a diameter of around 8 nm and lamellar structures. Both o/w and w/o MEs were also visualized using cryo-TEM [43].

In the pharmaceutical field, very little has been done to elucidate the microstructure of ME systems using electron microscopy. Bolzinger et al. [44] reported on bicontinuous sucrose ester-based ME systems for transdermal drug delivery. The microstructure of the investigated ME systems was viewed by freeze-fracture TEM. The authors showed images of a bicontinuous structure and reported that incorporating the anti-inflammatory drug niflumic acid into the system did not alter the ME microstructure. Alany et al. (2001) reported on the microstructure of ME systems formulated using a blend of two nonionic surfactants, ethyl oleate and water with and without 1-butanol. They described two distinct microstructures, namely droplet w/o and bicontinuous MEs [35]. Their TEM observations were complemented by electrical conductivity and viscosity measurements.

### 5.3 Electrical Conductivity Measurements

Electrical conductivity measurements can provide valuable information concerning the structure and phase behavior of ME systems [45]. Schulman et al. [38] measured the conductivity of ME systems, but only in a qualitative way, as they did not monitor changes in conductivity that are associated with phase changes.

Shah and Hamlin [46] studied the changes in electrical resistance associated with change in the water-to-oil ratio in an ME system during inversion into various LC systems and coarse dispersions. In such systems the electrical conductivity may exhibit both maxima and/or minima, reflecting changes in ion mobility caused by variation in viscosity. The most important feature of systems undergoing ME-to-LC transition is the gradual change in electrical conductivity with changing composition.

On the other hand, a large electrical conductivity transition has been observed in several w/o ME systems [47]. A well-known feature of w/o ME systems is the

steep rise in electrical conductivity as the water concentration increases [48]. This sudden change in electrical conductivity has been attributed to the percolation of spherical droplets (water droplets surrounded by an amphiphile shell, that is, water-swollen reverse micelles) in the oil phase [49]. The conductivity remains low up to a certain water volume fraction due to the nonconducting nature of the continuous phase of the w/o system. However, as the volume fraction of water reaches and exceeds the percolation threshold ($\varphi_p$), some of these conductive droplets begin to contact each other and form clusters which are sufficiently close to each other. This causes an efficient transfer of charge carriers between the dispersed droplets by charge hopping or transient merging of connected droplets resulting in an exponential increase of conductance from an almost zero value to much higher levels. It is reasonable to imagine a continuous pathway of water extending through the ME system, which by some authors is recognized as a sign of emergence of a bicontinuous structure [50]. Different methods were used to estimate the percolation threshold ($\varphi_p$) from the conductivity data by various investigators [35, 48, 51–54].

### 5.4 Viscosity Measurements

The viscosity of ME systems is also sensitive to structural changes and Newtonian flow is usually observed. The low viscosity of ME reflects the fluid character of the overall structure, which is a favorable feature for most ME applications. The pioneering work of Attwood et al. [55] on ME systems formulated using liquid paraffin, water, Span 60, and Tween 80 made reference to the equation

$$\eta_{rel} = a^{\phi/(1-\kappa\phi)} \tag{2}$$

where $\eta_{rel}$-relative viscosity
$a$-viscosity constant with theoretical value of 2.5 for solid spheres
$\varphi$-volume fraction of disperse phase
$\kappa$-hydrodynamic interaction coefficient

In the same study Attwood et al. [55] investigated the effect of increasing the surfactant concentration on the overall viscosity of an o/w ME system and obtained values for the viscosity constant $a$ of 3.19–4.17. The authors concluded that allowance for the hydration of the polyoxyethylene chain of the used surfactant reduced the value of the viscosity constant $a$ toward the theoretical value of 2.5 for a solid sphere. They also concluded that changing the ratio of the nonionic surfactants did not significantly affect the viscosity of the system.

Baker et al. [56] studied the viscosity of w/o ME systems containing water, xylene, sodium alkylbenzenesulfonate, and hexanol using Equation (1). They reported values of the viscosity constant $a$ of 3.3–6.0, which is above the theoretical value of 2.5 for a sphere with an increase in the surfactant concentration. This finding was attributed to the increase in the ratio of surfactant layer thickness to droplet core radius as the surfactant concentration was increased. However, deviation in values of the viscosity constant $a$ from the solid-sphere theoretical value of 2.5 could also be attributed to changes in the droplet shape or symmetry [57].

Viscosity studies have also been carried out to investigate the effect of the surfactant and cosurfactant concentrations as well as the surfactant–cosurfactant mass ratio on the hydration of the disperse-phase droplets for o/w ME systems [58]. A

study was conducted on systems composed of isopropyl myristate, water, polysorbate 80, and sorbitol. The results showed an increase in the viscosity constant and a decrease in the hydrodynamic interaction coefficient with decreasing surfactant–cosurfactant mass ratio. The increase in the viscosity constant resulted from greater hydrodynamic volume of droplets as well as the associated increase in the bound solvent layer of the droplet core radius from 7 to 22%.

Kaler et al. [50] reported on the viscosity changes in association with a percolative phenomenon for systems containing the commercial surfactant TRS 10–80, octane, tertiary amyl alcohol, and various brines. Their viscosity results were interpreted as evidence for a smooth transition from an oil-continuous to a bicontinuous one in which both oil and water span the sample. A second transition was observed and was attributed to a transition from a bicontinuous to a water-continuous system.

Borkovec et al. [59] also reported on a two-stage percolation process for the ME AOT (Aerosol OT, bis(2-ethylhexyl)sodium sulfosuccinate) system AOT–decane–water. The structural inversions were investigated using viscosity, conductivity, and electro-optical effect measurements. The viscosity results showed a characteristic profile with two maxima, which was interpreted as evidence for two symmetrical percolation processes: an oil percolation on the water-rich side of the phase diagram and a water percolation process on the oil-rich side.

Alany et al. [11, 35] reported on the phase behavior of two pharmaceutical ME systems showing interesting viscosity changes. The viscosity of both systems increased with increasing volume fraction of the dispersed phase to 0.15 and flow was Newtonian. However, formation of LC in one of the two systems, namely the cosurfactant-free system, resulted in a dramatic increase in viscosity that was dependent on the volume fraction of the internal phase and a change to pseudoplastic flow. In contrast, the viscosity of the bicontinuous ME was independent of water volume fraction. The authors used two different mathematical models to explain the viscosity results and related those to the different colloidal microstructures described.

### 5.5 Other Characterization Techniques

Among the other techniques used to characterize ME systems with emphasis on droplet size determination are optical techniques such as static and dynamic light scattering and nonoptical techniques such as small-angle X-ray scattering and small-angle neutron scattering [22, 45, 60–67], pulsed field gradient NMR [21, 42, 61, 68–71], and dielectric measurements [72–74]. One limitation of dynamic light scattering or so-called photon correlation spectroscopy is the need to either dilute or heat the ME system to overcome droplet–droplet interactions as well as microstructure changes that are associated with increased viscosity. Such techniques are valuable to obtain useful information regarding such interactions, but the downside would be the inevitable phase behavior changes that would render these techniques of some limited value in determining the original droplet size [75].

## 6 MICROEMULSIONS AS DRUG DELIVERY SYSTEMS

ME systems have been attracting increasing interest as vehicles for drug delivery via the various routes. Particular emphasis has been put on the oral, transdermal, ocular, and parenteral routes. Moreover, these systems have been investigated for

other applications that are relevant to the pharmaceutical, chemical, and biological sciences. The most significant development from a pharmaceutical perspective to this date would be undoubtedly the launch of the first oral cyclosporin A ME formulation, namely Sandimmnue Neoral. Other breakthroughs are likely to follow. Some of the key advantages related to ME systems include their thermodynamic stability, transparency, ease of preparation, low viscosity, and ultralow interfacial tension, to mention a few.

Furthermore, the presence of nanodomains of different polarity along with an interfacial surfactant/cosurfactnat film within the same systems allow for hydrophilic, lipophilic, and amphiphilic drugs to be accommodated together if needed. For o/w ME systems, their main advantage would be to improve the oral bioavailability of class two (II) drugs. These are drugs that display a dissolution-dependent bioabsorption [76]. The rationale would be to dissolve the poorly water soluble drug in the oil phase, thereby rendering it available in a molecular form ready for dissolution. However, it is important to remember that most drugs are insoluble in hydrocarbon and to some extent in vegetable oils, rather they would require more polar oils such as mixtures of mono- and diglycerides along the lines of Miglyol 840. Using such oils results in systems where the poorly soluble drug is dispersed as molecules in nanometer-sized oil droplets and as such are readily available for dissolution. Moreover, it is worth noting that the use of o/w MEs for delivery routes where extensive aqueous dilution is likely is less problematic than using w/o ME systems. This is because the external phase of an o/w ME will be diluted by water (being the main constituent of biological fluids) and will therefore retain its microstructure. Conversely, the potential use of w/o ME systems in the area of drug delivery where there is extensive aqueous dilution is rather complicated by the fact that such systems will be diluted by the aqueous biological fluids. This will result in droplet growth and subsequent phase changes with potential dose dumping. Despite the aforementioned disadvantages w/o MEs offer an option for the formulation of class III drugs. These drugs show permeability-limited bioabsorption in vivo [76]. Peptides/proteins, antisense oligonucleotides, deoxyribozymes, and small interfering ribonucleic acids (siRNAs) are good examples on class II drugs. Such molecules display little or no activity when delivered orally and are highly susceptible to the harsh gastrointestinal (GI) tract conditions and to their degrading enzymes in vivo. Water-in-oil MEs offer an exciting opportunity for optimizing the delivery of such molecules as they can be successfully incorporated in the internal (aqueous) phase and therefore are denied access to the harsh external conditions. Moreover the internal water droplets are likely to act as a nanoreservoir to control the mass transfer process of the loaded drug, that is, offer a mechanism for controlled release. In addition, the presence of surfactants/cosurfactants as constituents of the formulation can serve to increase membrane permeability, thereby improving drug absorption and possibly bioavailability. Bicontinuous MEs are the least investigated as drug delivery vehicles. They are highly fluid with low viscosity and possess ultralow interfacial tension. These properties render them potential candidates for topical and ocular drug delivery where their wetting and spreading properties come as an advantage.

The following section highlights some of the main drug delivery areas where ME systems have been researched as potential drug carriers.

### 6.1 Oral Drug Delivery

The most common method for drug delivery is through the oral route as it offers convenience and high patient compliance. However, recent advances in combinatorial chemistry is resulting in new molecules with very low water solubility. This leads to poor dissolution in the GI tract and subsequent erratic and unpredictable bioabsorption post–oral administration. Self-emulsifying drug delivery systems (SEDDSs) and self-micro-emulsifying drug delivery systems (SMEDDSs) offer an interesting option to optimize the delivery of such problematic drug molecules. SMEDDSs can be defined as isotropic, anhydrous systems comprising oil and surfactant that form o/w MEs upon mild agitation in the presence of water [77, 78]. The usefulness of SEDDSs in the area of oral drug delivery has been previously reported by Charman and co-workers, who showed improved pharmacodynamic properties of the investigational lipophilic drug WIN 54954 [79]. Systems comprising the medium-chain triglycerides Captex 355 and 800 in combination with Capmul MCM (medium chain mono- and di-glyceride mixture) and polyoxyethylene 20 sorbitan mono-oleate (Tween 80) were formulated by Constantinides et al. [26, 28, 80]. The authors reported on the improved bioavailability of calcein, RGD peptide, and a water-soluble marker when incorporated using a ME preconcentrate and w/o ME in comparison with an aqueous solution serving as a control. The same group also reported on the use of ME systems to increase oral absorption of poorly water soluble drugs [25]. The authors reported on the ability of the investigated systems to increase drug aqueous solubility and improve dissolution and oral bioavailability. Surprisingly, the formation of lamellar LC systems upon aqueous dilution of SEDDSs was reported to be a characteristic feature of the most effective SEDDSs [81]. Tenjarla summarized the factors that are likely to have an effect on the in vivo absorption of drugs from ME systems as being phase volume ratios, in vivo droplet size, partition coefficient of the drug between the two immiscible phases, the presence of a drug in an emulsified form or dispersed in oil, site or path of absorption, metabolism of the oil in the formulation, excipients that may act as absorption promoters, gastric emptying, and drug solubility in the ME excipients [2]. It was also reported that peptide uptake from MEs in the GI tract is dependent on particle size, type of ME lipid phase, digestability of lipid used, presence of bile slats, lipase, type of surfactants in ME, pH, and shedding of enterocytes [82].

The most remarkable story of success with ME research and development has to do with the oral delivery of cyclosporin A, marketed nowadays under the commercial name Sandimmune Neoral. Cyclosporin A is a cyclic peptide used post-transplantation surgery as an immune-suppressing agent. Unlike most peptides, cyclosporin A is hydrophobic and possesses very limited water solubility. The conventional cyclosporin oral formulation (Sandimmune) is in the form of a drug solution in olive oil along with ethanol and polyethoxylated oleic acid glycerides. Once given orally, the oily solution forms a coarse emulsion and as such behaves as a SEDDS with a bioabsorption process that is slow and incomplete. The net outcome is fluctuating drug plasma levels, poor and variable bioavailability, and pronounced inter- and intrapatient variability [83]. In an attempt to overcome some of these problems, Tarr and Yalkowsky [84] demonstrated that particle size reduction using high-shear homogenization can enhance absorption in rats. This improvement in oral absorption was attributed to the increased dosage form surface area that is

associated with droplet size reduction. The Neoral formulation is composed of a concentrated blend of two surfactants based on medium-chain partial glycerides along with an equivalent chain length triglyceride serving as oil, a cosolvent, and the drug and is described by the manufacturer as "microemulsion preconcentrate." Exposure of this "preconcentrate" to water results in the formation of a w/o ME that upon further dilution undergoes a phase change into an o/w ME. The improved in vivo performance of Neoral over Sandimmune has been demonstrated on multiple occasions [83, 85, 86]. Furthermore, the available pharmacokinetic data have been reviewed and it was concluded that Neoral offers better predictable and more extensive drug absorption than Sandimmune. Other poorly water soluble molecules that have been recently formulated in the form of SMEDDSs with the aim of improving bioavailability include simvastatin [87] and paclitaxel [88].

Water-in-oil microemulsions, on the other hand, offer an exciting opportunity to enhance the oral bioavailability of water-soluble peptide drugs. Because of their low oral bioavailability, peptide drugs are mostly available as parenteral formulations. However, parenteral peptides have an extremely short biological half-life and would therefore require multiple daily injections. This is likely to be problematic in chronic conditions (insulin for diabetes management is a good example) where patient compliance is likely to be an issue. Hydrophilic peptide drugs of this nature can be successfully accommodated into the internal aqueous phase of w/o ME systems where they are provided with protection from enzymatic degradation post–oral administration [89]. Furthermore, the presence of surfactant and some cosurfactants (such as medium-chain glycerides) can act to increase GI membrane permeability through interacting with the cell membrane bilayer and as such improve oral bioavaialibility [25, 26, 80, 89–92].

One major concern regarding the safety profile of ME systems intended for oral administration is the comparatively high amphiphile content. Both o/w and w/o ME systems are amphiphile-rich systems compared to conventional emulsions and would contain in the most conservative case up to 15–20% w/w surfactant–cosurfactant. This is further complicated by the limited models available to evaluate chronic toxicology in comparison to conventional oral dosage forms such as tablets [91].

### 6.2 Transdermal Drug Delivery

Transdermal drug delivery to the systemic circulation is one of the oldest routes that have been exploited using ME systems. This route offers distinct advantages compared to traditional routes by avoidance of first-pass metabolism, potential of controlled release, ease of administration, and possibility of immediate withdrawal of treatment when necessary [93]. Transdermal drug delivery aims at maximizing drug flux into the systemic circulation through the skin, whereas dermal drug delivery aims at targeting either the epidermis or the dermis of the skin. The key challenge in both cases is to provide sufficient increase in drug flux with minimal or no significant irreversible alteration to the skin barrier function [93].

Several studies have reported on the enhanced bioavailability of cutaneous drugs using o/w and w/o MEs compared to conventional emulsions, gels or solutions, mesophases, micellar and inverse micellar systems, and vesicles [93]. Moreover, a diverse range of drug molecules such as ketoprofen, apomorphine, estradiol, lidocaine [94–97], indomethacin and diclofenac [98], prostaglandin $E_1$ [99], aceclofenac

[100], vinpocetine [101], azelaic acid [102], methotrexate [103], piroxicam [104], triptolide [105], fluconazol [106], and ascorbyl palmitate [107] were incorporated into different ME systems.

Kantaria et al. [108, 109] reported on gelatin ME-based organogels (MBGs) as potential iontophoretic systems for the transdermal delivery of drugs. The microstructure of the proposed MBG was elucidated using small-angle neutron scattering where w/o ME droplets were entrapped in an extensive network of gelatin–water percolative channels. Theses MBGs were found to be electrically conducting and were shown to successfully deliver a model drug (sodium salicylate). Theses systems were formulated using pharmaceutically acceptable ingredients, including Tween 80 as a surfactant and isopropyl myristate as the oily component [108].

Transdermal delivery of proteins and/or DNA vaccines for needle-free immunization has been attracting increasing interest. Cui et al. [110] reported on ethanol-in-fluorocarbone (E/F) MEs for topical immunisation. The authors showed that plasmid DNA incorporated into E/F MEs was found to be stable. Furthermore, after topical application to the skin, significant enhancements in luciferase expression, antibody production, and T-helper type 1 based immune response compared to an aqueous or ethanolic solutions of DNA were observed [110].

One major concern with the topical application of ME systems is their biocompatibility and toxicity potential, mostly due to their high surfactant–cosurfactant content. Fundamental ME research utilizes ionic surfactants and medium-chain alcohols. While these ingredients are interesting from a physicochemical perspective, they pose serious biocompatibility and toxicity concerns [75]. Nonionic and zwitter ion–based surfactants (such as certain phospholipids) offer a more pharmaceutically acceptable alternative. Several research groups have been focusing on formulating ME systems using a single surfactant such as lecithin [98, 111, 112] or *n*-alkyl POE (polyoxyethylene ethers) [14, 112, 113]. This approach tends to compromise the phase behavior of ME-forming systems. This is usually seen when the ME region in the constructed ternary phase diagrams tends to become smaller in size. This translates to less choice in terms of ME composition and possibly stability. Alternative cosurfactants with improved biocompatibility and lower skin irritation potential have been recently introduced. Plurol isostearique has been shown to be compatible with a range of surfactants and oils and was capable of providing sizable ME regions.

In conclusion, topically applied MEs have been shown to significantly increase the cutaneous uptake of both lipophilic and hydrophilic drugs. The favorable properties of ME systems include the large concentration gradient (between vehicle and skin) due to the high drug solubilization power of ME systems without increasing drug affinity to the vehicle compared to conventional topical delivery systems [93]. Moreover, the penetration-enhancing properties of the individual surfactant/cosurfactant ingredient, ease of preparation and "infinite" physical stability, and good wetting and spreading properties make ME promising for future topical applications.

### 6.3 Parenteral Drug Delivery

Flubiprofen o/w ME systems were prepared and evaluated as vehicles for parenteral drug delivery [114]. These systems were formulated using POE 20 sorbitan monolaurate (Tween 20) as the surfactant and ethyl oleate as the oil phase. Flubiprofen

solubility in the o/w ME systems was eight times higher than that in an isotonic buffer; however, there was no significant differences in the pharmacokinetic parameters in rats between the ME formulation and the buffer [114]. Bicontinuous MEs designed for intravenous (i.v.) administration have been prepared and characterized [71]. The bicontinuous ME system underwent a phase change into an o/w emulsion upon aqueous dilution. In vitro investigations revealed small droplets with mean size radii of 60–200 nm. While solubilization studies were conducted using two drugs, namely felodipine and an antioxidant experimental drug (H 290/58), in vivo evaluations were conducted using the drug-free formulation. Doses of up to 0.5 mL/kg given i.v. to rats did not show any undesirable effects and had no significant effects on acid–base balance, blood gases, plasma electrolytes, arterial blood pressure, or heart rate [71].

Lecithin-based o/w MEs for parenteral use were formulated using polysorbate 80, IPM (Isopropyl myristate), lecithin, and water at different lecithin–polysorbate 80 weight ratios [115]. The formulated systems were shown to be highly stable and of minimal toxicity when evaluated in vitro. Phospholipid-based ME formulations of all-trans retinoic acid (ATRA) for parenteral administration were prepared and tested in vitro [116]. ATRA is effective against acute promyelocytic leukemia with highly variable oral bioavailability. Parenteral ME of ATRA was prepared using pharmaceutically acceptable ingredients, namely phospholipids and soybean oil. The inhibitory effect of ATRA on two human cancer cell lines (HL-60 and MCF-7) was not affected by incorporation into a ME formulation.

ME systems intended for parenteral application have to be formulated using nontoxic and biocompatible ingredients. The o/w ME systems would be suitable to improve the solubility of poorly water soluble drug molecules whereas w/o ME systems would be best suited for optimizing the delivery of hydrophilic drug molecules that are susceptible to the harsh GI conditions. Moreover, w/o systems can serve to prolong the release and mask any potential tissue irritation and site toxicity that are caused by intramuscular (i.m.) administration of hydrophilic drug molecules.

### 6.4 Ocular Drug Delivery

Aqueous solutions account for around 90% of the available ophthalmic formulations, mainly due to their simplicity and convenience [117]. However, extensive loss caused by rapid precorneal drainage and high tear turnover are among the main drawbacks associated with topical ocular drug delivery. Only 1–5% of the topically applied drug reaches the intraocular tissue with the remainder of the instilled dose undergoing nonproductive absorption via the conjunctiva or drainage via the nasolacrimal duct. This results in drug loss into the systemic circulation and provides undesirable systemic side effects [118]. Many strategies have been implemented to overcome such delivery challenges. These include the use of thermosetting in situ gelling polymer-based systems [119], nanoparticles, liposomes, and niosomes [120–123]. However, MEs offer a promising alternative as they comprise aqueous and oily components and can therefore accommodate both hydrophilic and lipophilic drugs. Moreover, they are transparent and thermodynamically stable and possess ultralow interfacial tension and therefore offer excellent wetting and spreading properties. Further advantages result form possible improvement of solubility and

stability of incorporated drugs with potential increase in bioavailability; hence these systems could be a suitable alternative to conventional ocular formulations. So far, only few investigators [124–127] have considered the use of MEs for ocular drug delivery. Their work was solely focused on o/w MEs as ocular delivery carriers. Recently Alany et al. [128] reported on w/o MEs formulated using a blend of two nonionic surfactants (Crillet 4 and Crill 1), an oily component (Crodamol EO), and water. These systems were shown to be capable of undergoing a phase change to lamellar liquid crystals upon aqueous dilution. The ocular irritation potential of the individual components and final formulations was assessed using a modified hen's egg chorioallantoic membrane test (HET-CAM), and the preocular retention was investigated in the rabbit eye using gamma scintigraphy. The authors demonstrated that the retention of ME systems was significantly greater than an aqueous solution. The rapid clearance of the w/o ME formulated with 10% water compared to the LC system indicated that phase change is less likely to take place in the rabbit eye [128]. It was also concluded that w/o MEs may be of value as vehicles for the ocular drug delivery of irritant hydrophilic drugs as they appear to have a protective effect when evaluated using a modified HET-CAM test [128]. The potential of bicontinuous ME systems as vehicles for ocular drug delivery is yet to be investigated. A recent review reported on the potential of MEs as ocular drug delivery systems; however, submicrometer emulsions and systems requiring energy input to prepare were also covered and classified as MEs [129].

## 7 CONCLUDING REMARKS

Microemulsions represent an exciting opportunity for pharmaceutical formulators and drug delivery scientists. They are easy to prepare and thermodynamically stable. Moreover, they can accommodate drugs of different physicochemical properties and protect those that are labile. They have the potential to increase the solubility of poorly water soluble drugs, enhance the bioavailability of problematic drugs, reduce patient variability, and offer an option for controlled drug release. A critical look at the current literature shows that exciting and promising research is taking place. It is only a matter of time before new ME-based products will find their way to the market following the successful introduction of Sandimmune Neoral.

## REFERENCES

1. Danielsson, I., and Lindman, B. (1981), The definition of microemulsion, *Colloid Surfaces*, 3, 391.
2. Tenjarla, S. (1999), Microemulsions: An overview and pharmaceutical applications [Review], *Crit. Rev. Ther. Drug Carrier Syst.*, 16(5), 461–521.
3. Attwood, D. (1994), Microemulsions, in Kreuter, J., Ed., *Colloidal Drug Delivery Systems*, Marcel Dekker, New York, pp 31–71.
4. Lawrence, M. J., and Rees, G. D. (2000), Microemulsion-based media as novel drug delivery systems, *Adv. Drug Del. Rev.*, 45(1), 89–121.
5. Mitchell, D., and Ninham, W. (1981), Micelles vesicles and microemulsions, *J. Chem. Soc. Faraday Trans.*, 77(2), 601–629.

6. Israelachvili, J. N., Mitchell, D., and Ninham, W. (1976), Theory of self-assembly of hydrocarbon amphiphiles into micelles and bilayers, *J. Chem. Soc. Faraday Trans. II*, 72, 1525–1568.
7. Prince, L. M. (1977), Formulation, in Prince, L. M., Ed., *Microemulsions: Theory and Practice*, Academic, New York, pp 33–49.
8. Shinoda, K., and Kuineda, H. (1973), Condition to produce so-called microemulsions: Factors to increase mutual solubility of oil and water solubilisers, *J. Colloid Interface Sci.*, 42, 381–387.
9. Aboofazeli, R., Lawrence, C. B., Wicks, S. R., and Lawrence, M. J. (1994), Investigations into the formation and characterization of phospholipid microemulsions. Part 3. Pseudo-ternary phase diagrams of systems containing water-lecithin-isopropyl myristate and either an alkanoic acid, amine, alkanediol, polyethylene glycol alkyl ether or alcohol as cosurfactant, *Int. J. Pharm.*, 111, 63–72.
10. Kahlweit, M., Busse, G., and Faulhaber, B. (1995), Preparing microemulsions with alkyl monoglucosides and the role of *N*-alkanols, *Langmuir*, 11(9), 3382–3387.
11. Alany, R. G., Rades, T., Agatonovic-Kustrin, S., Davies, N. M., and Tucker, I. G. (2000), Effects of alcohols and diols on the phase behaviour of quaternary systems, *Int. J. Pharm.*, 196(2), 141–145.
12. FDA (rev. Apr. 1, 2001), Food and drugs, Chapter I—Food and Drug Administration, in Department of Health and Human Services, Ed., *Code of Federal Regulations, Title 21* (Vol. 3), U.S. Government Printing Office, Washington, DC.
13. Thevenin, M. A., Grossiord, J. L., and Poelman, M. C. (1996), Sucrose esters cosurfactant microemulsion systems for transdermal delivery—assessment of bicontinuous structures, *Int. J. Pharm.*, 137(2), 177–186.
14. Malcolmson, C., Satra, C., Kantaria, S., Sidhu, A., and Lawrence, M. J. (1998), Effect of oil on the level of solubilization of testosterone propionate into nonionic oil-in-water microemulsions, *J. Pharm. Sci.*, 87(1), 109–116.
15. Ho, H. O., Hsiao, C. C., and Sheu, M. T. (1996), Preparation of microemulsions using polyglycerol fatty acid esters as surfactant for the delivery of protein drugs, *J. Pharm. Sci.*, 85(2), 138–143.
16. Kuineda, H., Hasegawa, Y., John, A. C., Naito, M., and Muto, M. (1996), Phase behaviour of polyoxyethylene hydrogenated caster oil in oil / water systems, *Colloid Surfaces*, 209–216.
17. Agatonovic-Kustrin, S., Glass, B. D., Wisch, M. H., and Alany, R. G. (2003), Prediction of a stable microemulsion formulation for the oral delivery of a combination of antitubercular drugs using ANN methodology, *Pharm. Res.*, 20(11), 1760–1765.
18. Aboofazeli, R., and Lawrence, M. J. (1993), Investigations into the formation and characterization of phospholipid microemulsions. Part 1. Pseudo-ternary phase diagrams of systems containing water-lecithin-alcohol-isopropyl myristate, *Int. J. Pharm.*, 93(May 31), 161–175.
19. Aboofazeli, R., and Lawrence, M. J. (1994), Investigations into the formation and characterization of phospholipid microemulsions. Part 2. Pseudo-ternary phase diagrams of systems containing water-lecithin-isopropyl myristate and alcohol: Influence of purity of lecithin, *Int. J. Pharm.*, 106, 51–61.
20. Aboofazeli, R., Patel, N., Thomas, M., and Lawrence, M. J. (1995), Investigations into the formation and characterization of phospholipid microemulsions. 4. Pseudo-ternary phase diagrams of systems containing water-lecithin-alcohol and oil—the influence of oil, *Int. J. Pharm.*, 125(1), 107–116.
21. Shinoda, K., Araki, M., Sadaghiani, A., Khan, A., and Lindman, B. (1991), Lecithin-based microemulsions; phase behaviour and microstructure, *J. Phys. Chem.*, 95, 989–993.

22. Saint-Ruth, H., Attwood, D., Ktistis, G., and Taylor, C. J. (1995), Phase studies and particle size analysis of oil-in-water phospholipid microemulsions, *Int. J. Pharm.*, 116, 253–261.
23. Warisnoicharoen, W., Lansley, A. B., and Lawrence, M. J. (2000), Nonionic oil-in-water microemulsions: The effect of oil type on phase behaviour, *Int. J. Pharm.*, 198(1), 7–27.
24. Park, K. M., Lee, M. K., Hwang, K. J., and Kim, C. K. (1999), Phospholipid-based microemulsions of flurbiprofen by the spontaneous emulsification process, *Int. J. Pharm.*, 183(2), 145–154.
25. Constantinides, P. P. (1995), Lipid microemulsions for improving drug dissolution and oral absorption: physical and biopharmaceutical aspects, *Pharm. Res.*, 12(Nov.), 1561–1572.
26. Constantinides, P. P., Lancaster, C. M., Marcello, J., Chiossone, D. C., Orner, D., Hidalgo, I., et al. (1995), Enhanced intestinal absorption of an RGD peptide from water-in-oil microemulsions of different composition and particle size, *J. Controlled Release*, 34, 109–116.
27. Constantinides, P. P., and Yiv, S. H. (1995), Particle size determination of phase-inverted water-in-oil microemulsions under different dilution and storage conditions, *Int. J. Pharm.*, 115(Mar. 7), 225–234.
28. Constantinides, P. P., Welzel, G., Ellens, H., Smith, P. L., Sturgis, S., Yiv, S. H., et al. (1996), Water-in-oil microemulsions containing medium-chain fatty acids/salts: Formulation and intestinal absorption enhancement evaluation, *Pharm. Res.*, 13(2), 210–215.
29. Kahlweit, M., Busse, G., Faulhaber, B., and Eibl, H. (1995), Preparing nontoxic microemulsions, *Langmuir*, 11(11), 4185–4187.
30. Kahlweit, M., Busse, G., and Faulhaber, B. (1996), Preparing microemulsions with alkyl monoglucosides and the role of alkanediols as cosolvents, *Langmuir*, 12, 861–862.
31. Joubran, R., Parris, N., Lu, D., and Trevino, S. (1994), Synergetic effect of sucrose and ethanol on formation of triglyceride microemulsions, *J. Disper. Sci. Technol.*, 15(6), 687–704.
32. Stubenrauch, C., Paeplow, B., and Findenegg, G. H. (1997), Microemulsions supported by octyl monoglucoside and geraniol. 1. The role of alcohol in the interfacial layer, *Langmuir*, 13, 3652–3658.
33. Bhargava, H. N., Narurkar, A., and Lieb, L. M. (1987), Using microemulsions for drug delivery, *Pharm. Technol.*, 11(Mar.), 46.
34. Bourrel, M., and Schechter, R. S. (1988), The R-ratio, in Bourrel, M., and Schechter, R. S. Eds., *Microemulsions and Related Systems*, Marcel Dekker, New York, pp 1–30.
35. Alany, R. G., Davies, N. M., Tucker, I. G., and Rades, T. (2001), Characterising colloidal structures of pseudoternary phase diagrams formed by oil/water/amphiphile systems, *Drug Dev. Ind. Pharm.*, 27(1), 33–41.
36. Alany, R. G., Agatonovic-Kustrin, S., Rades, T., and Tucker, I. G. (1999), Use of artificial neural networks to predict quaternery phase systems from limited experimental data, *J. Pharm. Biomed. Anal.*, 19(3–4), 443–452.
37. Richardson, C. J., Mbanefo, A., Aboofazeli, R., Lawrence, M. J., and Barlow, D. J. (1997), Prediction of phase behavior in microemulsion systems using artificial neural networks, *J. Colloid Interface Sci.*, 187(2), 296–303.
38. Schulman, J. H., Stoeckenius, W., and Prince, L. M. (1959), Mechanism of formaution and structure of microemulsions by electron microscopy, *J. Phys. Chem.*, 63, 1677–1685.
39. Schulman, J. H., Stoeckenius, W., and Prince, L. M. (1960), The structure of meylin figures and microemulsions as observed with the electron microscope, *Kolloid-Z*, 169, 170–179.

40. Jahn, W., and Strey, R. (1988), Microstructure of microemulsions by freeze fracture electron microscopy, *J. Phys. Chem.*, 92, 2294–2301.
41. Bodet, J. F., Bellare, J. R., Davis, H. T., Scriven, L. E., and Miller, W. G. (1988), Fluid microstructure transition from globular to bicontinuous midrange microemulsion, *J. Phys. Chem.*, 92, 1898–1902.
42. Regev, O., Ezrahi, S., Aserin, A., Garti, N., Wachtel, E., and Kaler, E. W., et al. (1996), A study of microstructure of a four component nonionic microemulsion by cryo-TEM, NMR, SAXS and SANS, *Langmuir*, 12, 668–674.
43. Krauel, K., Davies, N. M., Hook, S., and Rades, T. (2005), Using different structure types of microemulsions for the preparation of poly(alkylcyanoacrylate) nanoparticles by interfacial polymerization, *J. Controlled Release*, 106(1–2), 76–87.
44. Bolzinger, M. A., Thevenin, Carduner, T. C., and Poelman, M. C. (1998), Bicontinuous sucrose ester microemulsion: A new vehicle for topical delivery of niflumic acid, *Int. J. Pharm.*, 176(1), 39–45.
45. Kahlweit, M., Strey, R., Hasse, D., Kuineda, H., Schmeling, T., and Faulhaber, B., et al. (1987), How to study microemulsions, *J. Colloid Interface Sci.*, 118, 436–451.
46. Shah, D. O., and Hamlin, R. M. J. (1971), Structure of water in microemulsions: Electrical birefringence and nuclear magnetic resonance studies, *Sciences*, 171, 483–485.
47. Clausse, M., Nicholas-Morgantini, L., Zarbda, A., and Touraud, D. (1987), Water/sodium dodecylsulfate/1-pentanol/*n*-dodecanemicroemulsions realms of existence and transport properties, in Rosano, H. L., and Clausse, M., Eds., *Microemulsion Systems*, Marcel Dekker, New York, pp 387–425.
48. Lagues, M., and Sauterey, C. (1980), Percolation transition in water in oil microemulsions. Electrical conductivity measurements, *J. Phys. Chem.*, 84, 3503–3508.
49. Lagourette, B., Peyrelasse, J., Boned, C., and Clausse, M. (1979), Percolative conduction in microemulsion type system, *Nature*, 281, 60–62.
50. Kaler, E. W., Bennett, K. E., Davis, H. T., and Scriven, L. V. (1983), Toward understanding microemulsion microstructure I: A small-angle x-ray scattering study, *J. Phys. Chem.*, 79, 5673–5684.
51. Fang, J., and Venable, R. L. (1987), Conductivity study of the microemulsions system sodium dodecyl sulfate-hexylamine-heptane-water, *J. Colloid Interface Sci.*, 116, 269–277.
52. Gu, G., Wang, W., and Yan, H. (1996), Electric percolation of water-in-oil microemulsions: The application of effective medium theory to system sodium dodecylbenzenesulfonate (DDBS)/*n*-pentanol/*n*-heptane/water, *J. Colloid Interface Sci.*, 178(1), 358–360.
53. Mehta, S. K., and Bala, K. (1995), Volumetric and transport properties in microemulsions and the point of view of percolation theory, *Phys. Rev. E*, 51(6), 5732–5737.
54. Ray, S., Paul, S., and Moulik, S. P. (1996), Physicochemical studies on microemulsions: V. Additive effects on the performance of scaling equations and activation energy for percolation of conductance of water/AOT/heptane microemulsion, *J. Colloid Interface Sci.*, 183(1), 6–12.
55. Attwood, D., Currie, L. R. J., and Elworthy, P. H. (1974), Studies of solubilised micellar solutions, *J. Colloid Interface Sci.*, 46, 261–265.
56. Baker, R. C., Florence, A. T., Ottewill, R. H., and Tadros, T. F. (1984), Investigations into the formation and characterization of microemulsions. II. Light scattering conductivity and viscosity studies of microemulsions, *J. Colloid Interface Sci.*, 100(2), 332–349.
57. Florence, A. T., and Attwood, D. (1998), *Physicochemical Principles of Pharmacy*, 3rd ed., Macmillan, Basingstoke.

58. Ktistis, G. (1990), Viscosity study on oil-in-water microemulsions, *Int. J. Pharm.*, 61(June 30), 213–218.
59. Borkovec, M., Eicke, H. F., Hammerich, H., and Das Gupta, B. (1988), Two percolation process in microemulsions, *J. Phys. Chem.*, 92, 206–211.
60. Patel, N., Marlow, M., and Lawrence, M. J. (1998), Microemulsions: A novel pMD1 formulation, in *Drug Delivery to the Lungs IX*, AerosolSociety, London, pp 160–163.
61. Giustini, M., Palazzo, G., Colafemmina, G., Dealla Monica, M., Giomini, M., and Ceglie, A. (1996), Microstructure and dynamics of the water-in-oil CTAB/*n*-pentanol/*m*-hexane/water microemulsions: spectroscopic and conductivity study, *J. Phys. Chem.*, 100, 3190–3198.
62. Schurtenberger, P., Peng, Q., Leser, M. E., and Luisi, P. L. (1993), Structure and phase behaviour of lecithin-based microemulsions: A study of chanin length dependence, *J. Colloid Interface Sci.*, 156, 43–51.
63. Shioi, A., Harada, M., and Tanabe, M. (1995), Static light scattering from oil-rich microemulsions containing polydispersed cylindrical aggregates in sodium bis(2-ethylhexyl) phosphate system, *J. Phys. Chem.*, 99, 4750–4756.
64. Ktistis, G. (1997), Effect of polysorbate 80 and sorbitol concentration on in vitro release of indomethacin from microemulsions, *J. Disper. Sci. Technol.*, 18(1), 49–61.
65. Eugster, C., Rivara, G., Forni, G., and Vai, S. (1996), Marigenol-concentrates comprising Taxol and/or Taxan esters as active substances, *Panminerva Med.*, 38(4), 234–242.
66. Constantinides, P. P., and Scalart, J. P. (1997), Formulation and physical characterisation of water-in-oil microemulsions containing long versus medium chain length glycerides, *Int. J. Pharm.*, 58, 57–68.
67. Hantzschel, D., Enders, S., Kahl, H., and Quitzsch, K. (1999), Phase behaviour of quaternary systems containing carbohydrate surfactants-water-oil-cosurfactant, *Phys. Chem. Chem. Phys.*, 1, 5703–5701.
68. Olla, M., Monduzzi, M., and Ambrosone, L. (1999), Microemulsions and emulsions in DDAB/water/oil systems, *Colloids Surfaces A: Physicochem. Eng. Aspects*, 160, 392–401.
69. Carlfors, J., Blunte, I., and Schmidt, V. (1991), Lidocaine in microemulsion—a dermal-delivery system, *J. Disper. Sci. Technol.*, 12, 467–482.
70. Angelico, R., Palazzo, G., Colafemmina, G., Cirkel, P. A., Giustini, M., and Ceglie, A. (1998), Water diffusion and head group mobility in polymer-like reverse micelle: Evidence of a sphere-to-rod-sphere transition, *J. Phys. Chem. B*, 102, 2883–2889.
71. von Corswant, C., Thoren, P., and Engstrom, S. (1998), Triglyceride-based microemulsion for intravenous administration of sparingly soluble substances, *J. Pharm. Sci.*, 87(2), 200–208.
72. D'Angelo, M., Fioretto, D., Onori, G., Palmieri, L., and Santucci, A. (1996), Dynamics of water-containing sodium bis(2-ethylhexyl)sulfosuccinate (AOT) reverse micelles: A high-frequency dielectric study, *Phys. Rev. E*, 54(1), 993–996.
73. Feldman, Y., Kozlovich, N., Nir, I., and Garti, N. (1997), Dielectric spectroscopy of microemulsions, *Colloid Surfaces*, 128, 47–61.
74. Cirkel, P. A., van der Ploeg, J. P. M., and Koper, G. J. M. (1998), Branching and percolation in lecithin wormlike micelles studied by dielectric spectroscopy, *Phys. Rev. E*, 57, 6875–6883.
75. Alany, R. G. (2001), Microemulsions as vehicles for ocular drug delivery: Formulation, Physical-chemical characterisation and biological evaluation, Ph. D. thesis, University of Otago, Otago.

76. Amidon, G. L., Lennernas, H., Shah, V. P., and Crison, J. R. (1995), A theoretical basis for a biopharmaceutic drug classification: The correlation of in vitro drug product dissolution and in vivo bioavailability, *Pharm. Res.*, 12(3), 413–420.
77. Greiner, R. W., and Evans, D. F. (1990), Spontaneous formation of a water-continuous emulsion from water-in-oil microemulsion, *Langmuir*, 6, 1793–1796.
78. Shah, N. H., Carvajal, M. T., Patel, C. I., Infeld, M. H., and Malick, A. W. (1994), Self-emulsifying drug delivery systems with poly-glycolyzed glycerides for improving in vitro dissolution and oral absorption of lipophilic drugs, *Int. J. Pharm.*, 106, 15–23.
79. Charman, S. A., Charman, W. N., Rogge, M. C., Wilson, T. D., Dutko, F. J., and Pouton, C. W. (1992), Self-emulsifying drug delivery systems: Formulation and biopharmaceutic evaluation of an investigational lipophilic compound, *Pharm. Res.*, 9(1), 87–93.
80. Constantinides, P. P., Scalart, J. P., Lancaster, C., Marcello, J., and Smith, P. L., et al. (1994), Formulation and intestinal absorption enhancement evaluation of water-in-oil microemulsions incorporating medium-chain glycerides, *Pharm. Res.*, 11, 1385–1390.
81. Craig, D. Q. M., Barker, S. A., Banning, D., and Booth, S. W. (1995), An investigation into the mechanisms of self-emulsification using particle size analysis and low frequency dielectric spectroscopy, *Int. J. Pharm.*, 114(1), 103–110.
82. Ritschel, W. A. (1991), Microemulsions for improved peptide absorption from the gastrointestinal tract, *Methods Findings Exp. Clin. Pharm.*, 13(3), 205–220.
83. Holt, D. W., Mueller, E. A., Kovarik, J. M., van Bree, J. B., and Kutz, K. (1994), The pharmacokinetics of Sandimmun Neoral: a new oral formulation of cyclosporine, *Transplant. Proc.*, 26(5), 2935–2939.
84. Tarr, B. D., and Yalkowsky, S. H. (1989), Enhanced intestinal absorption of cyclosporine in rats through the reduction of emulsion droplet size, *Pharm. Res.*, 6(1), 40–43.
85. Kovarik, J. M., Mueller, E. A., van Bree, J. B., Tetzloff, W., and Kutz, K. (1994), Reduced inter- and intraindividual variability in cyclosporine pharmacokinetics from a microemulsion formulation, *J. Pharm. Sci.*, 83(3), 444–446.
86. Mueller, E. A., Kovarik, J. M., van Bree, J. B., Tetzloff, W., Grevel, J., and Kutz, K. (1994), Improved dose linearity of cyclosporine pharmacokinetics from a microemulsion formulation, *Pharm. Res.*, 11(2), 301–304.
87. Kang, B. K., Lee, J. S., Chon, S. K., Jeong, S. Y., Yuk, S. H., and Khang, G., et al. (2004), Development of self-microemulsifying drug delivery systems (SMEDDS) for oral bioavailability enhancement of simvastatin in beagle dogs, *Int. J. Pharm.*, 274(1–2), 65–73.
88. Yang, S., Gursoy, R. N., Lambert, G., and Benita, S. (2004), Enhanced oral absorption of paclitaxel in a novel self-microemulsifying drug delivery system with or without concomitant use of P-glycoprotein inhibitors, *Pharm. Res.*, 21(2), 261–270.
89. Sarciaux, J. M., Acar, L., and Sado, P. A. (1995), Using microemulsion formulations for drug delivery of therapeutic peptides, *Int. J. Pharm.*, 120, 127–136.
90. Swenson, E. S., and Curatolo, W. (1992), Intestinal permeability enhancement for proteins, peptides and other polar drugs: Mechanism and potential toxicity, *Adv. Drug Del. Rev.*, 8, 39–92.
91. Swenson, E. S., Milisen, W. B., and Curatolo, W. (1994), Intestinal permeability enhancement: Efficacy, acute local toxicity, and reversibility, *Pharm. Res.*, 11(8), 1132–1142.
92. Pouton, C. W. (1997), Formulation of self-emulsifying drug delivery systems for peptides: reduced plasma testosterone levels in male rats after a single injection, *Int. J. Pharm.*, 25, 47–58.
93. Kreilgaard, M. (2002), Influence of microemulsions on cutaneous drug delivery, *Adv. Drug Del. Rev.*, 54 Suppl 1, S77–S98.

94. Rhee, Y. S., Choi, J. G., Park, E. S., and Chi, S. C. (2001), Transdermal delivery of ketoprofen using microemulsions, *Int. J. Pharm.*, 228(1–2), 161–170.
95. Peira, E., Scolari, P., and Gasco, M. R. (2001), Transdermal permeation of apomorphine through hairless mouse skin from microemulsions, *Int. J. Pharm.*, 226(1–2), 47–51.
96. Peltola, S., Saarinen-Savolainen, P., Kiesvaara, J., Suhonen, T. M., and Urtti, A. (2003), Microemulsions for topical delivery of estradiol, *Int. J. Pharm.*, 254(2), 99–107.
97. Sintov, A. C., and Shapiro, L. (2004), New microemulsion vehicle facilitates percutaneous penetration in vitro and cutaneous drug bioavailability in vivo, *J. Controlled Release*, 95(2), 173–183.
98. Dreher, F., Walde, P., Walther, P., and Wehrli, E. (1997), Interaction of a lecithin microemulsion gel with human stratum corneum and its effect on transdermal transport, *J. Controlled Release*, 45(2), 131–140.
99. Ho, H. O., Huang, M. C., Chen, L. C., Hsia, A., Chen, K. T., and Chiang, H. S., et al. (1998), The percutaneous delivery of prostaglandin E1 and its alkyl esters by microemulsions, *Clin. Pharm. J.*, 50, 257–266.
100. Yang, J. H., Kim, Y. I., and Kim, K. M. (2002), Preparation and evaluation of aceclofenac microemulsion for transdermal delivery system, *Arch. Pharm. Res.*, 25(4), 534–540.
101. Hua, L., Weisan, P., Jiayu, L., and Hongfei, L. (2004), Preparation and evaluation of microemulsion of vinpocetine for transdermal delivery, *Pharmazie*, 59(4), 274–278.
102. Gasco, M. R., Gallarate, M., and Pattarino, F. (1991), In vitro permeation of azelaic acid from viscosized microemulsions, *Int. J. Pharm.*, 69(3), 193–196.
103. Alvarez-Figueroa, M. J., and Blanco-Mendez, J. (2001), Transdermal delivery of methotrexate: Iontophoretic delivery from hydrogels and passive delivery from microemulsions, *Int. J. Pharm.*, 215(1–2), 57–65.
104. Park, E. S., Cui, Y., Yun, B. J., Ko, I. J., and Chi, S. C. (2005), Transdermal delivery of piroxicam using microemulsions, *Arch. Pharm. Res.*, 28(2), 243–248.
105. Mei, Z., Chen, H., Weng, T., Yang, Y., and Yang, X. (2003), Solid lipid nanoparticle and microemulsion for topical delivery of triptolide, *Eur. J. Pharm. Biopharm.*, 56(2), 189–196.
106. El Laithy, H. M., and El-Shaboury, K. M. (2002), The development of Cutina lipogels and gel microemulsion for topical administration of fluconazole, *AAPS Pharm. Sci. Tech.*, 3(4), E35.
107. Jurkovic, P., Sentjurc, M., Gasperlin, M., Kristl, J., and Pecar, S. (2003), Skin protection against ultraviolet induced free radicals with ascorbyl palmitate in microemulsions, *Eur. J. Pharm. Biopharm.*, 56(1), 59–66.
108. Kantaria, S., Rees, G. D., and Lawrence, M. J. (1999), Gelatin-stabilised microemulsion-based organogels: Rheology and application in iontophoretic transdermal drug delivery, *J. Controlled Release*, 60(2–3), 355–365.
109. Kantaria, S., Rees, G. D., and Lawrence, M. J. (2003), Formulation of electrically conducting microemulsion-based organogels, *Int. J. Pharm.*, 250(1), 65–83.
110. Cui, Z., Fountain, W., Clark, M., Jay, M., and Mumper, R. J. (2003), Novel ethanol-in-fluorocarbon microemulsions for topical genetic immunization, *Pharm. Res.*, 20(1), 16–23.
111. Bonina, F. P., Montenegro, L., Scrofani, N., Esposito, E., Cortesi, R., and Menegatti, E., et al. (1995), Effects of phospholipid based formulations on in vitro and in vivo percutaneous absorption of methyl nicotinate, *J. Controlled Release*, 34(1), 53–63.
112. Malcolmson, C., and Lawrence, M. J. (1990), A comparison between nonionic micelles and microemulsions as a means of incorporating the poorly water soluble drug diazepam, *J. Pharm. Pharmcol.*, Suppl. 42, 6P.

113. Muller, M., Mascher, H., Kikuta, C., Schafer, S., Brunner, M., and Dorner, G., et al. (1997), Diclofenac concentrations in defined tissue layers after topical administration, *Clin. Pharmacol. Therap.*, 62(3), 293–299.
114. Park, K.-M., and Kim, C.-K. (1999), Preparation and evaluation of flurbiprofen-loaded microemulsion for parenteral delivery, *Int. J. Pharm.*, 181(2), 173–179.
115. Moreno, M. A., Ballesteros, M. P., and Frutos, P. (2003), Lecithin-based oil-in-water microemulsions for parenteral use: pseudoternary phase diagrams, characterization and toxicity studies, *J. Pharm. Sci.*, 92(7), 1428–1437.
116. Hwang, S. R., Lim, S.-J., Park, J.-S., and Kim, C.-K. (2004), Phospholipid-based microemulsion formulation of all-trans-retinoic acid for parenteral administration, *Int. J. Pharm.*, 276(1–2), 175–183.
117. Bourlais, C. L., Acar, L., Zia, H., Sado, P. A., Needham, T., and Leverge, R. (1998), Ophthalmic drug delivery systems—recent advances, *Prog. Retinal Eye Res.*, 17(1), 33–58.
118. Lang, J. C. (1995), Ocular drug delivery conventional ocular formulations, *Adv. Drug Del. Rev. Ocular Drug Del.*, 16(1), 39–43.
119. Miller, S. C., and Donovan, M. D. (1982), Effect of poloxamer 407 gel on the miotic activity of pilocarpine nitrate in rabbits, *Int. J. Pharm.*, 12(2–3), 147–152.
120. Fitzgerald, P., Hadgraft, J., Kreuter, J., and Wilson, C. G. (1987), A γ-scintigraphic evaluation of microparticulate ophthalmic delivery systems: Liposomes and nanoparticles, *Int. J. Pharm.*, 40(1–2), 81–84.
121. Calvo, P., Vila-Jato, J. L., and Alonso, M. J. (1997), Evaluation of cationic polymer-coated nanocapsules as ocular drug carriers, *Int. J. Pharm.*, 153(1), 41–50.
122. Pignatello, R., Bucolo, C., Ferrara, P., Maltese, A., Puleo, A., and Puglisi, G. (2002), Eudragit RS100 nanosuspensions for the ophthalmic controlled delivery of ibuprofen, *Eur. J. Pharm. Sci.*, 16(1–2), 53–61.
123. Vyas, S. P., Mysore, N., Jaitely, V., and Venkatesan, N. (1998), Discoidal niosome based controlled ocular delivery of timolol maleate, *Pharmazie*, 53(7), 466–469.
124. Gasco, M. R., Gallarate, M., Trotta, M., Bauchiero, L., Gremmo, E., and Chiappero, O. (1989), Microemulsions as topical delivery vehicles: Ocular administration of timolol, *J. Pharm. Biomed. Anal.*, 7(4), 433–439.
125. Gallarate, M., Gasco, M. R., Trotta, M., Chetoni, P., and Saettone, M. F. (1993), Preparation and evaluation in vitro of solutions and o/w microemulsions containing levobunolol as ion-pair, *Int. J. Pharm.*, 100(1–3), 219–225.
126. Hasse, A., and Keipert, S. (1997), Development and characterization of microemulsions for ocular application, *Eur. J. Biopharm.*, 43, 179–183.
127. Siebenbrodt, I., and Keipert, S. (1993), Poloxamer-Systems as potential ophthalmics II. Microemulsions, *Eur. J. Biopharm.*, 39, 25–30.
128. Alany, R. G., Rades, T., Nicoll, J., Tucker, I. G., and Davies, N. M. (2006), W/O microemulsions for ocular delivery: Evaluation of ocular irritation and precorneal retention, *J. Controlled Release*, 111(1–2), 145–152.
129. Vandamme, T. F. (2002), Microemulsions as ocular drug delivery systems: Recent developments and future challenges, *Prog. Retinal Eye Res.*, 21(1), 15–34.

# TRANSDERMAL DRUG DELIVERY

C. Scott Asbill and Gary W. Bumgarner
*Samford Unversity, Birmingham, Alabama*

**Contents**

1 Introduction
2 Physiology and Characteristics of Human Skin
  2.1 Transepidermal Water Loss and Occlusion
  2.2 Skin Lipids
3 Diffusion
4 Drug Candidates for Transdermal Drug Delivery
5 In Vitro Testing of Transdermal Devices and Drug Candidates
6 Transdermal Patch Design
  6.1 Membrane-Moderated Patches
  6.2 Adhesive Matrix
7 Commercially Available Patches
  7.1 Transderm Scop
  7.2 Catapres TTS
  7.3 Androderm
  7.4 Estradiol Transdermal Systems
  7.5 CombiPatch
  7.6 Duragesic
  7.7 Ortho Evra
  7.8 Oxytrol
  7.9 Emsam
  7.10 Daytrana
8 Chemical and Physical Approaches to Transdermal Delivery
  8.1 Chemical Penetration Enhancers
  8.2 Physical Enhancement Methods
9 The Future of Transdermal Drug Delivery
  References

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

## 1 INTRODUCTION

During the last decade there has been an explosion of drug candidates in clinical trials and a high number of dosage forms approved by the Food and Drug Administration (FDA) that utilize the transdermal route of drug delivery (Table 1). The interest in transdermals can be attributed to the many advantages offered by this route of delivery [1]. These advantages include maintaining steady-state drug concentrations (Figure 1), bypassing first-pass metabolism, mimicking an intravenous (IV) infusion, less frequent dosing, and increased patient compliance (Table 2).

The disadvantages of transdermal patches are patient allergies to adhesives found in transdermal patches, patch excipients that sometimes produce local irritation, potential for abuse or misuse, temperature that can affect the delivery of drugs from certain patches, and the inability of most drugs to be delivered transdermally (Table 3) [2].

In the 1960s a proposal was made that suggested the transdermal exposure of nitroglycerin to munitions workers resulted in tolerance of the workers to nitroglycerin. Long-term exposure of nitroglycerin in high doses can lead to tolerance. Workers who were abruptly removed from exposure to the nitroglycerin often experienced cardiac arrest and cases of sudden death were sometimes reported [3]. Soon after, a correlation was made between the physical and chemical properties of nitroglycerin and its significant transdermal penetration [4].

**TABLE 1 Selected FDA-Approved Transdermal Patches**

| Trade Name | Generic Name | Indication | Company |
|---|---|---|---|
| Androderm | Testosterone | Testosterone deficiency | Watson Pharma |
| Catapres TTS | Clonidine | Hypertension | Boehringer Ingelheim |
| Climara | Estradiol | Hormone replacement | Berlex |
| CombiPatch | Estradiol/norethindrone acetate | Hormone replacment | Novartis |
| Daytrana | Methylphenidate | Attention-deficit hyperactivity disorder | Shire Pharmaceuticals |
| Duragesic | Fentanyl | Pain managment | Janssen |
| Emsam | Selegiline | Depression | Somerset Pharmaceuticals |
| Estraderm | Estradiol | Hormone replacement | Novartis |
| Nicoderm CQ | Nicotine | Smoking cessation | GlaxoSmithKline |
| Nicotrol | Nicotine | Smoking cessation | Pfizer |
| Ortho Evra | Ethinyl estradiol/norelgestromin | Contraception | Ortho-Mcneil |
| Oxytrol | Oxybutynin | Urinary incontinence | Watson Pharma |
| Transderm Nitro | Nitroglycerin | Angina | Novartis |
| Transdrerm Scop | Scopolamine | Motion sickness | Novartis |
| Vivelle | Estradiol | Hormone replacement | Noven |
| Vivelle Dot | Estradiol | Hormone replacement | Noven |

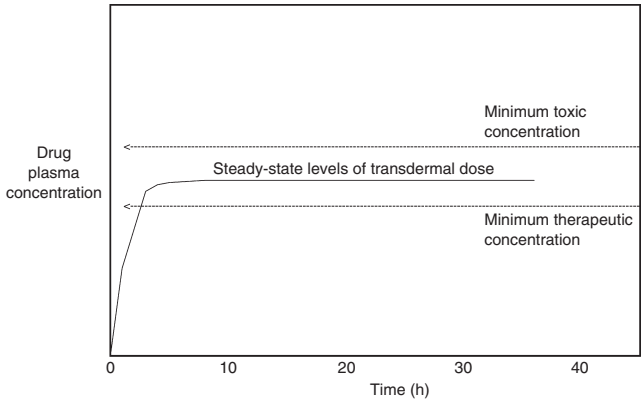


**FIGURE 1** Steady-state drug levels achieved by transdermal delivery.

**TABLE 2 Advantages of Transdermal Drug Delivery**

| |
|---|
| Excellent for drugs with short half-lives |
| Analogous to IV infusion |
| Route bypasses first-pass metabolism |
| Reduced side effects |
| Decreased dosing |
| Zero-order kinetics |
| Self-administration |
| Increased patient compliance |

**TABLE 3 Disadvantages of Transdermal Drug Delivery**

| |
|---|
| Allergies to adhesives |
| Product of local irritation |
| Potential for abuse or misuse |
| Temperature affects delivery |
| Drug needs to be potent and have desired physical and chemical properties |

Another historical perspective on transdermal drug delivery involved tobacco farmers receiving transdermal doses of nicotine from handling tobacco leaves. A paper published by Gehlback et al. in 1974 introduced the concept of transdermal nicotine [5]. It was common for farm workers who had direct contact with tobacco plants during rainy and humid conditions to exhibit symptoms often associated with nicotine poisoning. This condition is called green tobacco sickness and has affected many farm workers, particularly in states such as Kentucky and North Carolina, where the farming of tobacco is significant. Moisture on tobacco leaves from rain or dew may contain significant amounts of nicotine due to its high water solubility. Farm workers directly handling tobacco leaves often retained the moisture-

containing nicotine on their clothes or skin. Nicotine, because of its physical and chemical properties, is low molecular weight and has excellent transdermal penetration, and significant blood levels can lead to many adverse reactions, such as nausea and vomiting [6].

## 2 PHYSIOLOGY AND CHARACTERISTICS OF HUMAN SKIN

The skin is the largest organ in the human body and has many physiological functions. The skin serves to regulate overall body homeostasis, protect the body from external pathogens and chemicals, as well as control water loss from the body. The skin has three main layers. The epidermis, which is the outermost layer, is the thinnest layer of the skin and provides the most significant barrier function [7]. Beneath the epidermis, the dermis provides mechanical support to the skin and the third layer, immediately under the dermis, is a layer of subcutaneous fat called the hypodermis.

The epidermis consists of five principal layers and is an area of both intense biochemical activity and differentiation. These layers are the stratum corneum, stratum lucidum, stratum granulosum, stratum spinosum, and stratum basale. The stratum corneum (horny layer) is the uppermost layer of the epidermis and the skin. The stratum corneum is composed of dead keratinocytes, which are called corneocytes, and has an abundance of keratin and lipid structures [8]. The stratum corneum is considered the rate-limiting barrier for the diffusion of chemical compounds across the skin. The stratum lucidum (clear layer) is composed of two to three layers of dead flattened keratinocytes which appear translucent under a microscope and are present only in thick glabrous skin.

Stratum granulosum (granular layer) is an epidermal layer than consists of three to five layers of keratinocytes which have started to flatten out, and their nuclei and organelles have begun to disintegrate. Also, lipid and keratin granules have started to form inside the keratinocytes. Eventually the granules will give rise to the brick-and-mortar structure of the stratum corneum.

The stratum spinosum (prickly layer) is composed of several layers of keratinocytes which are starting to exhibit histological and biochemical changes that mark the beginning of the differentiation process. The shape of the keratinocytes has become irregular and enzymes responsible for lipid synthesis are present.

The stratum basale is the deepest layer of the epidermis and is composed mainly of keratinocytes with melanocytes making up approximately 10% of the cell population. The stratum basale is one cell layer thick and is a layer of rapid cell division where keratinocytes are rapidly dividing and giving rise to the uppermost layers of the epidermis.

The dermis is the largest layer of the skin. It is a region of strong and flexible connective tissue. The dermis consists of two primary layers, the papillary layer and the reticular layer. The papillary layer is the smallest layer of the dermis and is composed mainly of collagen and elastin fibers. The reticular layer is the largest layer of the dermis and is composed of mainly dense connective tissue. The layer of subcutaneous fat found directly beneath the dermis provides insulation and additional mechanical support to the skin.

### 2.1 Transepidermal Water Loss and Occlusion

Transepidermal water loss (TEWL) is a natural occurrence that takes place in the skin layers. TEWL is the result of movement of water from the deep skin layers across the epidermis into the outside atmosphere. It is a tightly regulated process that is controlled by the stratum corneum [9]. Occlusive topical bases and devices, such as transdermal patches, block TEWL and cause increased hydration of the skin. Hydration of the skin increases the permeation rates of compounds transdermally. The occluding effect of transdermal patches is an important mechanism that promotes increased diffusion of the compound across the skin into the systemic circulation [10].

### 2.2 Skin Lipids

As keratinocytes differentiate and move toward the stratum corneum, they synthesize an abundance of lipids, much of which are packaged into small organelles called lamellar granules [11]. When reaching the stratum corneum, the corneocytes eject these lipid granules forming a principal component of the brick-and-mortar structure of the stratum corneum. Lipids synthesized in the skin layers are thought to arrive from carbon sources derived from acetate obtained from the systemic circulation. The role of epidermal lipids in the brick-and-mortar structure of the stratum corneum and the barrier function of the skin is well established in the literature [12]. Many studies have suggested that when organic solvents such as dimethyl sulfoxide (DMSO) are used to dissolve epidermal lipids, an increase in skin permeability is found [13].

The major lipids found in the stratum corneum are ceramides, fatty acids, and cholesterol. Free fatty acids make up 10–15% of the lipid mass of the stratum corneum and predominantly consist of straight-chain saturated species ranging from 14 to 28 carbons in length. Cholesterol, a major lipid found in the stratum corneum, represents approximately 25% of the total stratum corneum lipid while cholesterol sulfate accounts for another 5%.

Ceramides are considered to be the largest group of lipids found in the stratum corneum, representing 50% of the total lipid weight. Six distinct ceramide fractions have been isolated and characterized [14]. Ceramide 1 is derived from the linoleate-rich acylglucosylceramide. Ceramide 2 consists of straight-chain saturated fatty acids amide linked to sphingosine and dihydrosphingosine bases. Ceramide 3 consists of saturated fatty acids amide linked to phytosphingosine, which has an additional hydroxyl group on carbon 4. Ceramides 4 and 5 both consist of α-hydroxyacids amide linked to sphingosines and dihydrosphingosines. Ceramide 6 contains α-hydroxyacids amide linked to phytosphingosines.

It has been suggested that these ceramides form a gel-phase membrane domain within the skin. Straight fatty acid chains as well as the small polar head groups on the ceramides are thought to produce a tightly packed domain which is less fluid and thereby less permeable than other liquid crystalline domains which are also present. Recent evidence using differential scanning calorimetery (DSC) and infrared absorption spectroscopy analyses verifies the presence of gel phases within the stratum corneum.

## 3 DIFFUSION

The diffusion of an active ingredient from a transdermal patch through the skin layers and into the systemic circulation can be attributed to Fick's law. Fick's law is best defined as a linear relationship between the flux of a chemical and the chemical's concentration gradient. The concentration gradient is the engine which triggers drug diffusion in all directions [15]. The chemicals move from a region of higher concentration to a region of lower concentration. Transdermal patches are typically loaded with large amounts of active ingredients in order to maximize diffusion. Once the drug is absorbed systemically the concentration gradient is maintained with "sink conditions" existing between the dosage form and the systemic circulation.

## 4 DRUG CANDIDATES FOR TRANSDERMAL DELIVERY

Several characteristics are important in determining which drugs are candidates for transdermal drug delivery. These are half-life, molecular weight, lipophilicity, and potency. Some pharmaceutical active ingredients have short biological half-lives whereas other have very long half-lives. For drugs such as estradiol with half-lives of less than 2 h, for the drug to be delivered orally, a frequent dosing schedule must be followed due to the rapid clearance of the drug from the systemic circulation. This high dosing frequency may be inconvenient and may lead to poor patient compliance. However, drugs with short half-lives are excellent candidates for transdermal drug delivery because the rate of drug delivery is constant and mimics the kinetics offered by an IV infusion. The steady-state drug delivery allows for less frequent dosing, which may result in better patient compliance.

Molecular weight is another property that must be considered when selecting potential candidates for transdermal drug delivery. Molecular weight is a major determinant in whether a molecule may pass through the restrictive barrier of the stratum corneum. Drugs that have low molecular weights have a better chance of penetrating the stratum corneum compared to high-molecular-weight compounds such as proteins and oligonucleotides, which are too large to passively diffuse across the stratum corneum. It has been proposed in the literature that compounds with a molecular weight of less than 400 daltons are potential candidates for transdermal drug delivery (Table 4) [16]. Lipophilicity of the drug is also an important factor concerning the chemical's ability to undergo transdermal delivery. Pharmaceutical active ingredients must have sufficient solubility in both the lipid portion and hydrophilic regions of the skin in order to have significant permeation. Drugs with a $\log P$ between –1.0 and 4.0 can potentially traverse the brick-and-mortar structure of the stratum corneum.

**TABLE 4 Physico Chemical Factors That Affect Permeation of Compounds**

| |
|---|
| Small molecular weight |
| Melting point <200°C |
| Suitable partition coefficient > $0.5 \times 10^{-3}$ cm/h |

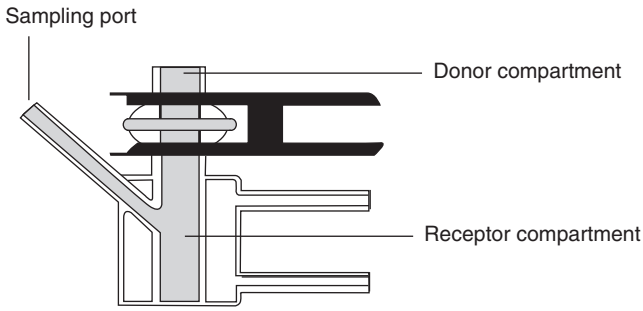


**FIGURE 2** Typical in vitro permeation cell.

Another important parameter when selecting transdermal candidates is the drug's potency. Only drugs that provide therapeutic effects at low steady-state plasma levels are viable candidates. The input rate needed for a transdermal patch flux can be determined from multiplying the clearance of the drug with the desired plasma concentration of the drug at steady state. This is a useful predictor when selecting the feasibility of using an active ingredient in a transdermal patch.

## 5 IN VITRO TESTING OF TRANSDERMAL DEVICES AND DRUG CANDIDATES

The ability to test the transdermal penetration of compounds in vitro has been investigated extensively over the last decade. Hundreds of studies have shown that the permeation rate of compounds can be accurately measured using several available in vitro diffusion cells and various types of skin models [17].

A typical in vitro permeation cell is composed of both a donor compartment and a receptor compartment (Figure 2). The model skin or membrane for permeation testing is placed between the donor and receptor compartments. The donor compartment is where the drug solution, ointment, or transdermal patch is applied. The compound then permeates across the membrane into the receptor compartment. The receptor compartment typically contains a buffered solution maintained at 37°C and is continuously stirred [18]. Often antibiotics and preservatives are added to the receptor compartment to restrict growth of microorganisms and to maintain the integrity of the skin model that is being utilized for the permeation study [19]. Samples are withdrawn from the receptor compartment at predetermined intervals for analysis. Following the withdrawal of samples from the receptor compartment, an equal amount of diffusion buffer is added to the volume.

## 6 TRANSDERMAL PATCH DESIGN

### 6.1 Membrane-Moderated Patches

Membrane-moderated patches have been utilized in many FDA-approved transdermal patches such as Duragesic, Transderm Scop, and Catapress TTS. This type of patch utilizes a rate-controlling membrane to precisely control the release of

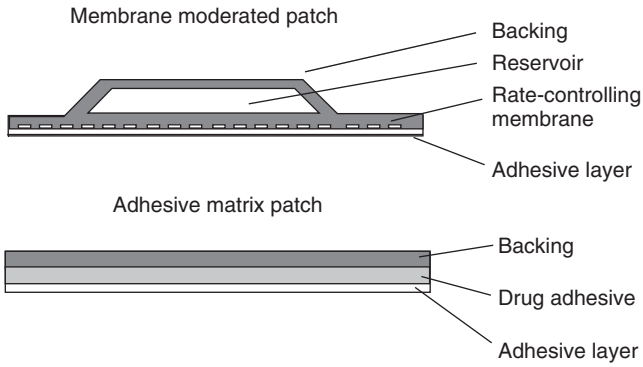


**FIGURE 3** Diagram of membrane-moderated and adhesive matrix transdermal patches.

active ingredient from the transdermal patch and provides release profiles of active ingredients that exhibit zero-order kinetics [20]. The layers of a membrane-moderated patch are as follows: The uppermost layer is an occlusive backing that is impermeable; next is the reservoir layer, which consists of active ingredient dissolved in an appropriate solvent such as mineral oil or ethyl alcohol. The next layer is a rate-controlling membrane that is typically composed of ethylene vinyl acetate copolymer. Underlying the rate-controlling membrane is a layer of adhesive and then finally the release liner (Figure 3). Commonly burst doses of the active ingredient are placed in the adhesive layer to saturate the skin layers and reduce the lag time to steady-state blood levels [21] (Figure 1).

### 6.2 Adhesive Matrix

Adhesive matrix patches are increasingly common, and most of the recently approved transdermal drug delivery devices utilize this type of technology [22–24]. Advantages of this type of system are easier to manufacture than membrane-moderated patches, smaller in size, and resistant to manufacturing defects which could lead to dose dumping. The layers of an adhesive matrix patch consists of the backing followed by a layer containing the active ingredient dissolved and mixed with adhesive and then the release liner [25] (Figure 3).

## 7 COMMERCIALLY AVAILABLE PATCHES

### 7.1 Transderm Scop

In 1978 Tranderm Scop was the first transdermal patch to receive FDA approval. Scopolamine, the active ingredient, is a belladonna alkaloid that is frequently used to treat motion sickness and nausea resulting from anesthetics and analgesics. Transderm Scop is a membrane-moderated patch that has a three-day life span. It has a circular shape with an area ($2.5\,cm^2$) approximately the size of a quarter [26].

### 7.2 Catapres TTS

Catapres TTS is the first and only transdermal patch approved by the FDA for the treatment of hypertension. Catapres TTS contains the active ingredient clondidine and was approved in 1985 [27]. Catapres TTS is a seven-day patch and is a membrane-moderated transdermal patch. The patch comes in three sizes, 0.1, 0.2, and 0.3 mg/day. In addition there is also a burst dose of clonidine in the adhesive layer. The presence of the burst doses provides an immediate-release dose of clonidine which promotes rapid systemic levels of the drug.

### 7.3 Androderm

Androderm is an FDA-approved membrane-moderated transdermal patch that delivers testosterone. Androderm is manufactured in two sizes and both are 24-h patches. The round Androderm patch contains 12.2 mg of testosterone and delivers a 2.5-mg dose over a 24-h period. The larger oval patch contains 24.3 mg of testosterone and delivers 5 mg over a 24-h period. The reservoir of Androderm contains testosterone gelled with alcohol and glycerin.

### 7.4 Estradiol Transdermal Systems

There are several FDA-approved estradiol patches currently on the market. These patches are three- to four-day patches used to treat symptoms associated with menopause. Estraderm is the only membrane-moderated estradiol system that is on the market. Estraderm is available in two sizes: 0.05 and 0.1-mg/day patches. Vivelle, Vivelle Dots, Alora, and Climara are examples of commercially available estradiol patches that utilize adhesive matrix patch design.

### 7.5 CombiPatch

CombiPatch is a three- or four-day patch that delivers both estradiol and norethindrone acetate. CombiPatch is available in two sizes: a 9-$cm^2$ patch that delivers 0.05 mg of estradiol per day and 0.14 mg of norethindrone acetate per day and a 16-$cm^2$ patch that delivers 0.05 mg of estradiol per day and 0.25 mg of norethindrone acetate per day. Estradiol is a lipophilic compound with a molecular weight of 272. The molecular weight of norethindrone acetate is 340. The design of the patch is considered an adhesive-matrix-type patch that consists of three layers. The backing is comprised of polyolefin and the adhesive layer contains a silicone adhesive, acrylate adhesive, estradiol, norethindrone acetate, oleic acid, and oleyl alcohol.

### 7.6 Duragesic

Duragesic is a transdermal patch that delivers the potent opioid analgesic fentanyl. The life span of the patch is three days and it is manufactured in five sizes: 12, 25, 50, 75, and 100 μg/h. Duragesic is a membraned-moderated patch and consists of four patch layers: a backing layer of polyester film, a drug reservoir that contains fentanyl, and U.S. Pharmacopeia (USP) alcohol gelled with hydroxyethyl cellulose,

a rate-controlling membrane made of ethylene-vinyl acetate copolymer, and a layer of silicone adhesive with a burst dose of fentanyl. Fentanyl has a molecular weight of 336.5 and is a lipophilic drug with significant transdermal permeation.

### 7.7 Ortho Evra

Ortho Evra was approved by the FDA in 2002 as the world's first transdermal contraceptive patch. Ortho Evra contains both ethinyl estradiol and norelgestromin as active ingredient. Ortho Evra is a three-layered patch that is a matrix-type design. The molecular weight of ethinyl estradiol is 296.41 and the molecular weight of norelgestromin is 327.47. The adhesives used in this system are polyisobutylene and polybutene. Ortho Evra is a seven-day patch which is cycled three weeks on and one week patch free.

### 7.8 Oxytrol

Oxytrol is a three- or four-day patch used for the treatment of urinary incontinence that was recently approved by the FDA. It delivers the active ingredient oxybutynin. Oxybutynin is an antispasmotic and anticholinergic agent with a molecular weight of 357. Oxybutynin is considered to be a lipophilic drug. Oxytrol is a matrix-type patch with a surface area of 39 $cm^2$ and contains 36 mg of oxybutynin. Oxytrol has an in vivo delivery rate of approximately 3.9 mg/day.

### 7.9 Emsam

Emsam is a once-a-day patch that delivers an active ingredient selegiline that is used to treat depression. Selegiline is an irreversible monoamine oxidase inhibitor and has a molecular weight of 187.30. Emsam is a matrix-type patch that has three layers and is available in three size. The 6-mg/day patch has a surface area of 20 mg/20 $cm^2$, the 9-mg/day patch has a surface area of 30 mg/30 $cm^2$, and the 12-mg/day patch has a surface area of 40 mg/40 $cm^2$.

### 7.10 Daytrana

Daytrana is a newly approved methylphenidate transdermal system. It is a 9-h adhesive matrix patch and comes in four sizes: a 12.5-$cm^2$ patch that has a delivery rate of 1.1 mg/h, an 18.75-$cm^2$ patch that has a delivery rate of 1.6 mg/h, a 25-$cm^2$ patch that delivers 2.2 mg/h, and a 37.5-$cm^2$ patch that delivers 3.3 mg/h.

## 8 CHEMICAL AND PHYSICAL APPROACHES TO TRANSERMAL DELIVERY

Due to the brick-and-mortar structure of the stratum corneum, the skin is a difficult layer to permeate across for most active pharmaceutical ingredients. Because of this diffusional barrier, new strategies have been developed to allow compounds to better penetrate the stratum corneum [28]. These strategies can be defined as either chemical or physical approaches to disrupting the barrier function of the skin.

**TABLE 5 Desired Properties of Chemical Permeation Enhancers**

| |
|---|
| Provide reversible effects in skin layers |
| Low systemic bioavailability |
| High stability and compatibility with formulations |
| Possess no pharmacological activity |
| Should be nontoxic and nonirritating |

**TABLE 6 Investigated Chemical Permeation Enhancers**

| | |
|---|---|
| Azone | Oleic acid |
| Ethano | Oxazolidnesl |
| Dimethyl sulfoxide | Propylene glycol |
| Fatty acids | Sodium lauryl sulfate |
| Lecithin | Terpenes |

### 8.1 Chemical Penetration Enhancers

Many approaches that have been investigated over the last several decades enhance the permeation of compounds across the skin using novel chemical compounds [29]. These permeation enhancers are compounds which partition into the stratum corneum and promote the passage of topically applied compounds across the skin layers by using three possible mechanisms of action (Table 5). The mechanisms by which these compounds enhance permeability of the skin have been previously described by Williams and Barry and are often referred to as the lipid–protein partitioning theory [30]. One proposed mechanism of action of permeation enhancers and the most common method by which chemical enhancers increase permeation is by fluidizing the intercellular lipid structures that are found within the stratum corneum. By interacting with and disorganizing these lipid structures, channels can be formed which allow the compound to better diffuse across the rate-limiting barrier of the stratum corneum.

Another method is based on the ability of some permeation enhancers to interact with intracellular proteins such as keratins inside corneocytes. The disassembly of these proteins structures within the corneocytes allows some compounds to transcellularly penetrate the stratum corneum. Also, some enhancers act as vehicles and cosolvents, increasing and promoting the partitioning of compounds into the stratum corneum (Table 6).

### 8.2 Physical Enhancement Methods

***Microneedles*** Microneedles are a new type of transdermal device that is receiving tremendous attention by pharmaceutical companies because of the potential for the transdermal delivery of high-molecular-weight compounds. This type of transdermal device contains microscopic needles or projections that can be loaded with active ingredient [31]. The projections when placed on the skin penetrate beyond the stratum corneum into the living epidermis. This allows the compound to bypass passage directly through the stratum corneum, which is the rate-limiting barrier for

passive diffusion. The drug can be loaded into the microneedles and then released into the lower skin levels or the microneedle patch may contain a reservoir and the drug will penetrate through channels in the stratum corneum that have been created by the microneedles. This allows compounds that normally would not passively diffuse across the stratum corneum, such as therapeutic proteins, to be delivered transdermally. Currently various formulations of microneedle patches are being investigated in clinical trials in the United States [32–34].

***Iontophoresis and Sonophoresis*** In the early 1900s it was discovered that some chemical compounds could be delivered into the systemic circulation across the skin using an electric current. This phenomenon was later described as iontophoresis. Iontophoresis occurs when an electric potential difference is created across the skin layers by an electric current and this gradient drives the penetration of both charged and uncharged drugs across the skin [35].

One of the earliest FDA-approved products that utilized iontophoresis was the iontophoretic delivery of pilocaripne as a method for diagnosing cystic fibrosis [36]. The sweat of individuals with cystic fibrosis contains large amounts of both sodium and chloride ions. The pilocarpine that is delivered into the skin promotes increased sweating, which can be easily collected and analyzed.

Another product that uses iontophoresis has been recently approved by the FDA. This device is called Ionsys and is an iontophoretic system that delivers fentanyl hydrochloride transdermally [37, 38]. This is a patient-controlled device that provides on-demand delivery of fentanyl for up to 24 h or 80 doses. This device contains 10.8 mg of fentanyl hydrochloride and is designed to deliver a 40-μg dose of fentanyl over a 10-min period upon activation of the dose button by the patient.

Another popular physical enhancement method that has routes in physical therapy and sports rehabilitation clinics is sonophoresis. Sonophoresis involves the use of ultrasound as a source of disrupting intercellular lipid structures in the stratum corneum [39, 40]. The sound waves produced by the device induce cavitation of the lipids found within the stratum corneum, which then opens channels and allows the chemical compound to easily penetrate the skin. This is a safe and reversible process that has received much attention in the literature and by pharmaceutical companies.

## 9 FUTURE OF TRANSDERMAL DRUG DELIVERY DEVICES

Through practical application of recombinant deoxyribonucleic acid (DNA) technology there has been a recent explosion of biotech drugs that are in clinical trials, and the FDA has approved many. Unfortunately, due to the high molecular weights and other physical and chemical characteristics of these macromolecules, many delivery obstacles exist that prevent these compounds from being delivered transdermally. Many physical and chemical approaches are being investigated that enhance the delivery of the biotech agents across the skin.

TransPharma-Medical, an Israeli-based pharmaceutical company, is investigating the transdermal delivery of human parathyroid hormone fragment for the treatment of osteoporosis in addition to the delivery of human growth hormone [41]. This technology utilizes a 1-$cm^2$ patch that creates small channels or holes in the stratum

corneum much like the microneedle technology that was previously discussed. The delivery of insulin has been studied for over a decade using transdermal systems. Unfortunately, due to the high molecular weight and dosing considerations with insulin, it has been difficult. However, insulin delivery transdermally has recent successes clinically using sonophoresis, permeation enhancers, iontophoresis, and microneedles [42, 43].

## REFERENCES


1. Robinson, D. H., and Mauger, J. W. (1991), Drug delivery systems, *Am. J. Hosp. Pharm.*, 48(10 Suppl. 1), S14–23.
2. Murphy, M., and Carmichael, A. J. (2000), Transdermal drug delivery systems and skin sensitivity reactions. Incidence and management, *Am. J. Clin. Dermatol.*, 1(6), 361–368.
3. Forman, S. A., Helmkamp, J. C., and Bone, C. M. (1987), Cardiac morbidity and mortality associated with occupational exposure to 1,2 propylene glycol dinitrate, *J. Occup. Med.*, 29(5), 445–450.
4. Ben-David, A. (1989), Cardiac arrest in an explosives factory worker due to withdrawal from nitroglycerin exposure, *Am. J. Ind. Med.*, 15(6), 719–722.
5. Gehlbach, S. H., et al. (1974), Green-tobacco sickness. An illness of tobacco harvesters, *JAMA*, 229(14), 1880–1883.
6. Curwin, B. D., et al. (2005), Nicotine exposure and decontamination on tobacco harvesters' hands, *Ann. Occup. Hyg.*, 49(5), 407–413.
7. Xhauflaire-Uhoda, E., et al. (2006), Dynamics of skin barrier repair following topical applications of miconazole nitrate, *Skin Pharmacol. Physiol.*, 19(5), 290–294.
8. Norlen, L., and Al-Amoudi, A. (2004), Stratum corneum keratin structure, function, and formation: The cubic rod-packing and membrane templating model, *J. Invest. Dermatol.*, 123(4), 715–732.
9. Fluhr, J. W., Feingold, K. R., and Elias, P. M. (2006), Transepidermal water loss reflects permeability barrier status: Validation in human and rodent in vivo and ex vivo models, *Exp. Dermatol.*, 15(7), 483–492.
10. Casiraghi, A., et al. (2002), Occlusive properties of monolayer patches: In vitro and in vivo evaluation, *Pharm. Res.*, 19(4), 423–426.
11. Wertz, P. W. (2000), Lipids and barrier function of the skin, *Acta Derm. Venereol. Suppl. (Stockh.)*, 208, 7–11.
12. Wertz, P. W. (1997), Integral lipids of hair and stratum corneum, *Exs*, 78, 227–237.
13. Astley, J. P., and Levine, M. (1976), Effect of dimethyl sulfoxide on permeability of human skin in vitro, *J. Pharm. Sci.*, 65(2), 210–215.
14. Wertz, P. W., et al. (1985), The composition of the ceramides from human stratum corneum and from comedones, *J. Invest. Dermatol.*, 84(5), 410–412.
15. Barry, B. W. (2002), Drug delivery routes in skin: A novel approach, *Adv. Drug Deliv. Rev.*, 54 (Suppl 1), S31–40.
16. Buchwald, P., and Bodor, N. (2001), A simple, predictive, structure-based skin permeability model, *J. Pharm. Pharmacol.*, 53(8), 1087–1098.
17. El-Kattan, A., Asbill, C. S., and Haidar, S. (2000), Transdermal testing: Practical aspects and methods, *Pharm. Sci. Technol. Today*, 3(12), 426–430.
18. Asbill, C., et al. (2000), Evaluation of a human bio-engineered skin equivalent for drug permeation studies, *Pharm. Res.*, 17(9), 1092–1097.

19. El-Kattan, A. F., Asbill, C. S., and Michniak, B. B. (2000), The effect of terpene enhancer lipophilicity on the percutaneous permeation of hydrocortisone formulated in HPMC gel systems, *Int. J. Pharm.*, 198(2), 179–189.
20. Krishnaiah, Y. S., and Bhaskar, P. (2004), Studies on the transdermal delivery of nimodipine from a menthol-based TTS in human volunteers, *Curr. Drug Deliv.*, 1(2), 93–102.
21. Misra, A., et al. (1997), Biphasic testosterone delivery profile observed with two different transdermal formulations, *Pharm. Res.*, 14(9), 1264–1268.
22. Burkman, R. T. (2004), The transdermal contraceptive system, *Am. J. Obstet. Gynecol.*, 190(4, Suppl), S49–53.
23. Sane, N., and McGough, J. J. (2002), MethyPatch Noven, *Curr. Opin. Invest. Drugs*, 3(8), 1222–1224.
24. Amsterdam, J. D. (2003), A double-blind, placebo-controlled trial of the safety and efficacy of selegiline transdermal system without dietary restrictions in patients with major depressive disorder, *J. Clin. Psychiatry*, 64(2), 208–214.
25. Chien, Y. W. (1985), Microsealed drug delivery systems: Fabrications and performance, *Methods Enzymol.*, 112, 461–470.
26. Parrott, A. C. (1989), Transdermal scopolamine: A review of its effects upon motion sickness, psychological performance, and physiological functioning, *Aviat. Space Environ. Med.*, 60(1), 1–9.
27. Burris, J. F., and Mroczek, W. J. (1986), Transdermal administration of clonidine: A new approach to antihypertensive therapy, *Pharmacotherapy*, 6(1), 30–34.
28. Trommer, H., and Neubert, R. H. (2006), Overcoming the stratum corneum: The modulation of skin penetration. A review, *Skin Pharmacol. Physiol.*, 19(2), 106–121.
29. Xu, P., and Chien, Y. W. (1991), Enhanced skin permeability for transdermal drug delivery: Physiopathological and physicochemical considerations, *Crit. Rev. Ther. Drug Carrier Syst.*, 8(3), 211–236.
30. Williams, A. C., and Barry, B. W. (1991), Terpenes and the lipid-protein-partitioning theory of skin penetration enhancement, *Pharm. Res.*, 8(1), 17–24.
31. Henry, S., et al. (1998), Microfabricated microneedles: A novel approach to transdermal drug delivery, *J. Pharm. Sci.*, 87(8), 922–925.
32. Giudice, E. L., and Campbell, J. D. (2006), Needle-free vaccine delivery, *Adv. Drug Del. Rev.*, 58(1), 68–89.
33. Aqil, M., Sultana, Y., and Ali, A. (2006), Transdermal delivery of beta-blockers, *Expert Opin. Drug Deliv*, 3(3), 405–418.
34. Cormier, M., et al. (2004), Transdermal delivery of desmopressin using a coated microneedle array patch system, *J. Controlled Release*, 97(3), 503–511.
35. Fischer, G. A. (2005), Iontophoretic drug delivery using the IOMED Phoresor system, *Expert Opin. Drug Deliv.*, 2(2), 391–403.
36. Parad, R. B., et al. (2005), Sweat testing infants detected by cystic fibrosis newborn screening, *J. Pediatr.*, 147(3, Suppl), S69–72.
37. Miaskowski, C. (2005), Patient-controlled modalities for acute postoperative pain management, *J. Perianesth. Nurs.*, 20(4), 255–267.
38. Sinatra, R. (2005), The fentanyl HCl patient-controlled transdermal system (PCTS): An alternative to intravenous patient-controlled analgesia in the postoperative setting, *Clin. Pharmacokinet.*, 44(Suppl 1), 1–6.
39. Mitragotri, S., and Kost, J. (2000), Low-frequency sonophoresis: A noninvasive method of drug delivery and diagnostics, *Biotechnol. Prog.*, 16(3), 488–492.

40. Mitragotri, S., Blankschtein, D., and Langer, R. (1996), Transdermal drug delivery using low-frequency sonophoresis, *Pharm. Res*, 13(3), 411–420.
41. Levin, G., et al. (2005), Transdermal delivery of human growth hormone through RF-microchannels, *Pharm. Res*., 22(4), 550–555.
42. Rastogi, S. K., and Singh, J. (2005), Effect of chemical penetration enhancer and iontophoresis on the in vitro percutaneous absorption enhancement of insulin through porcine epidermis, *Pharm. Dev. Technol*., 10(1), 97–104.
43. King, M. J., et al. (2002), Transdermal delivery of insulin from a novel biphasic lipid system in diabetic rats, *Diabetes Technol. Ther*., 4(4), 479–488.

# VAGINAL DRUG DELIVERY

José das Neves, Maria Helena Amaral, and
Maria Fernanda Bahia
*University of Porto, Porto, Portugal*

**Contents**

1 Introduction
2 The Human Vagina
  2.1 Anatomy
  2.2 Histology
  2.3 Physiology
  2.4 Childhood, Pregnancy, and Menopause
3 General Features of Vaginal Drug Delivery
  3.1 Advantages and Disadvantages of Vaginal Drug Delivery
  3.2 Permeability and Drug Absorption
  3.3 First-Uterine-Pass Effect
4 Vaginal Drug Delivery Systems
  4.1 Overview
  4.2 Excipients
  4.3 Solid Systems
  4.4 Semisolid Systems
  4.5 Liquid Systems
  4.6 Vaginal Rings
  4.7 Vaginal Films
  4.8 Medicated Vaginal Tampons
  4.9 Vaginal Foams
  4.10 Vaginal Sponges
  4.11 Other Strategies and Vaginal Drug Delivery Systems
  4.12 Packaging and Vaginal Applicators
5 Pharmaceutical Evaluation of Vaginal Drug Delivery Systems
  5.1 Legal and Official Compendia Requirements
  5.2 Drug Release and Permeability
  5.3 pH and Acid-Buffering Capacity
  5.4 Rheological Studies
  5.5 Textural Studies

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

5.6 Mucoadhesion
5.7 Vaginal Distribution and Retention
5.8 Safety and Toxicology
5.9 Other Characteristics
6 Clinical Usage and Potential of Vaginal Drug Delivery
6.1 Microbicides
6.2 Antimicrobials
6.3 Hormonal Contraceptives and Hormonal Replacement Therapy
6.4 Spermicides
6.5 Labor Inducers and Abortifacients
6.6 Proteins and Peptides
6.7 Vaccines
6.8 Other Uses
7 Acceptability and Preferences of Women Related to Vaginal Drug Delivery
8 Veterinary Vaginal Drug Delivery
9 Conclusions and Future Trends
References

## 1 INTRODUCTION

The vagina has been used for a long time as a route for drug administration, being as old as medicine and pharmacy themselves. Throughout the history of human civilization, vaginal administration of drugs has been practiced and recorded until the modern era. Some of the first records were found in Egypt, where the Kahun Papyrus, the oldest of the surviving medical papyri (ca. 1850 B.C.), included references to vaginal "preparations" containing substances such as mud, frankincense, oil, malachite, ass urine, myrrh, crocodile dung, honey, and sour milk, normally used in female genitalia–related conditions and contraception [1]. Latter papyri such as the Ramesseum Papyrus (ca. 1700 B.C.), the Ebers Papyrus (ca. 1550 B.C.), and the Greater Berlin Papyrus (ca. 1300 B.C.) also contained drug formulations to be administered in the vagina. Vaginal administration of drugs continued to be carried out by other civilizations from ancient Greece and Rome to the Middle Ages, in the Arabic and Oriental cultures, passing through the Renaissance, until our days [1, 2].

Although traditionally used for local action, some drugs can permeate the vaginal mucosa and reach the bloodstream in sufficient concentrations to have systemic effects. Current understanding of the vaginal anatomy, physiology, and pathophysiology is very considerable, contrasting with the still limited knowledge of the possibilities of vaginal drug delivery. Nonetheless, interest and contribution of pharmaceutical scientists toward vaginal drug delivery development have increased in the last years in response to the specific needs of this route of drug administration. Indeed, vaginal drug delivery has seen recent advances that make it very promising, in particular therapeutic fields such as the prevention of human immunodeficiency virus (HIV) and other sexually transmitted infections, contraception, hormone replacement therapy in menopausal women, and labor induction.

Therefore, this chapter discusses the main features of vaginal anatomy, physiology, and histology related to drug delivery as well as vaginal drug delivery systems and their evaluation. Also, past and current usage of the vagina as a route of drug administration, ongoing investigation, and promising strategies in this field are addressed.

## 2 THE HUMAN VAGINA

The knowledge of the human vagina's anatomy, histology, and physiology is fundamental in the design of drug delivery systems, helping researchers optimize vaginal products. Pharmaceutical scientists must comprehend the particularities of this organ in order to overcome its natural limitations and enhance its advantages over other drug delivery routes.

### 2.1 Anatomy

The organs of the female reproductive tract are classically divided into the internal and external genitalia [3]. The internal genital organs include the vagina, cervix, uterus, oviducts, ovaries, and surrounding supporting structures, as seen in Figure 1. The external genital organs, also known as the vulva or pudendum, are composed by the structures that surround the vaginal entrance, which are visible in the perineal area. These include the mons pubis, labia majora, labia minora, hymen, clitoris, vestibule, urethra, vestibular glands (Skene glands and Bartholin glands), and vestibular bulbs.

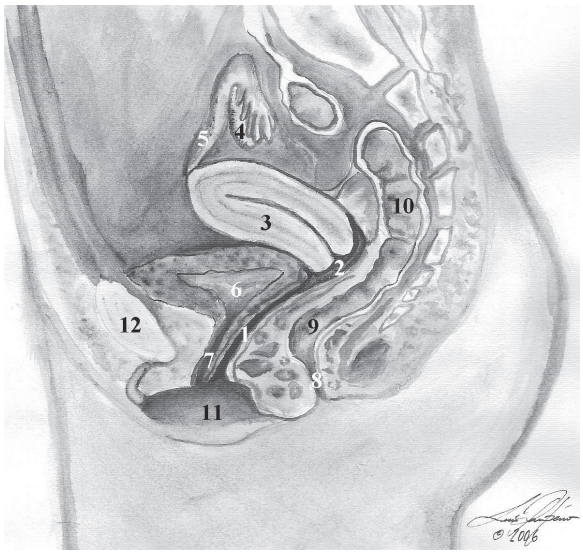


**FIGURE 1** Female reproductive system and related structures: vagina (1), cervix (2), uterus (3), ovary (4), fallopian tube (5), urinary bladder (6), urethra (7), anus (8), rectum (9), colon (10), vestibule (11), and pubic symphysis (12). (Courtesy of Luís Paupério.)

The vagina is a tubular, fibromuscular organ approximately 9 cm long, extending from the cervix to the vestibule, being positioned between the urinary bladder and the rectum (Figure 1). Because of the protrusion of the cervix into the upper part of the vagina, two deep recessions (anterior and posterior) are formed, called fornices, the posterior fornix being considerably larger than the anterior one. Also, two lateral fornices can be considered [3, 4]. In the adult woman, the anterior wall of the vagina is approximately 7.5 cm long, while the posterior wall extends by approximately 9 cm. The width of the vagina varies throughout its length, the transverse diameter being higher at the level of the fornices, decreasing progressively toward the vaginal entrance. The vagina is commonly misunderstood as a straight tube, descending from the cervix to the introitus. The axes of the upper and lower vagina are different, forming a slight S-shaped curve: The axis of the lower vagina in relation to a standing woman is vertical and posterior, while the upper part (from the pelvic diaphragm to the cervix) becomes more horizontal, with the final portion of the vagina curving toward the hollow of the sacrum [3, 5]. The angle between these two portions is approximately 130°C. This fact is important in drug delivery, taking into consideration that pharmaceutical formulations must be retained in the vagina, resisting to gravity forces. The walls of the vagina are covered by a mucosal tissue that forms a series of transverse folds denominated rugae (more prominent in the lower third of the vagina), thus increasing the available surface for absorption and allowing considerable distension of this organ during penile penetration and childbirth [4, 6]. Indeed, the surface area of the vaginal mucosa is an important parameter concerning its coverage by vaginal products or drug absorption. Results by Pendergrass et al. varied through a wide range of values (65.73–107.07 $mm^2$), with a mean surface area of 87.46 $mm^2$, although these results can be an understatement because of limitations in the determination method [7]. In a more recent study, Barnhart et al. found higher surface values which ranged from 103.9 to 165.0 $mm^2$. These variable results suggest that one volume of vaginal product may not be appropriate for all women [5]. The walls of the vagina are normally in apposition and flattened in the anteroposterior diameter. Thus the vagina has the appearance of the letter H in cross section, although a W shape may also be observed [8].

The vagina has a high blood supply, which is primarily provided by the vaginal branches of the internal iliac artery. To a minor extension, blood is also supplied through branches of the uterine, middle rectal, and internal pudendal arteries. Blood drainage from the vagina returns through vaginal veins (the perineum venous plexus), running parallel to the course of the arteries, which flow into the pudendal vein, internal iliac vein, and then the vena cava, thus bypassing the portal circulation. Blood supply increases with sexual excitement, particularly to the clitoris and vestibular bulbs, causing the erection of these structures along with the expansion and elongation of the vagina, allowing the accommodation of the male penis [4]. The lymphatic drainage is characterized by its wide distribution and frequent crossovers between both sides of the pelvis. Generally, the upper third of the vagina drains to the external iliac nodes, the middle third to the common and internal iliac nodes, and the lower third to the common iliac, superficial inguinal, and perirectal nodes [3].

Sensory innervation of the vagina is provided by the pudendal plexus and is particularly developed near the vaginal orifice. The nervous sympathetic innervation of the vagina is provided by the hypogastric plexus, while the nervous parasympa-

thetic innervation derives from the second and third sacral nerves. There is a lack of free nerve endings in the upper two-thirds of the vagina [3, 4]. Thus, objects placed in this area, such as vaginal drug delivery systems, are likely to be unperceived by women.

## 2.2 Histology

The vaginal tissue is composed of four distinct layers: stratified squamous epithelium, lamina propria, muscular layer, and tunica adventitia, the first two usually referred to as vaginal mucosa (Figure 2). The epithelial layer is about 150–200 μm thick, corresponding approximately to 30–45 layers of cells, being nonkeratinized. The epithelium presents different types of cells, making it possible to identify five different layers, namely the basal, parabasal, intermediate, transitional, and superficial layers. Epithelial cells are closely joined by numerous desmosomes and occasionally tight junctions, the last being particularly abundant in the basal layer. Deposits of a lipid material are found near the surface of the epithelium, acting as a permeability barrier to large water-soluble molecules. The lamina propria, or tunica, is composed of fibrous connective tissue richly supplied by small blood and lymphatic vessels. Although being a mucosal surface, the vaginal mucosa does not

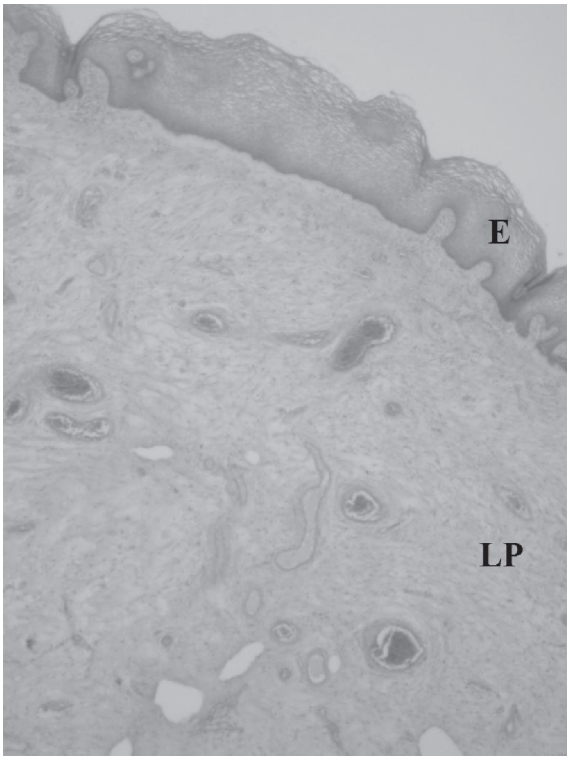


**FIGURE 2** Detail of human vaginal mucosa of fertile adult (H&E, ×40). Image shows stratified squamous epithelium (E), which is nonkeratinized, and lamina propria (LP), richly supplied by small blood vessels and deprived of glands (Courtesy of Manuel Dias.)

have mucus-secreting glands (Figure 2). The muscular layer has a large number of interlacing fibers, making it possible to distinguish an inner circular layer and an outer longitudinal layer. These muscle fibers allow substantial elongation during childbirth, increasing approximately five-fold. The tunica adventitia consists of connective tissue containing a large plexus of blood vessels, lymphatic vessels, and nerves [9].

The characteristics of the vaginal mucosa depend on sexual hormones undergoing changes throughout the menstrual cycle. During the proliferative phase, estrogens increase vaginal blood flow and the integrity of the mucosa, inducing the proliferation of the vaginal epithelium, peaking at approximately midcycle. Nonetheless, this change in epithelial thickness is small and probably clinically insignificant [10, 11]. In the secretory phase progesterone opposes those actions, although without significantly influencing the epithelium thickness [12].

## 2.3 Physiology

Functions of the vagina include receiving the erect penis and semen during coitus and ejaculation and serving as a passageway for fetus and menses to the outside of the body [4].

The vaginal milieu is typically acidic in healthy women during their fertile years, playing an important role in preventing the proliferation of pathogenic microorganisms. Vaginal pH is around 3.5–4.5, changing during the menstrual cycle being lower in the middle cycle and higher around menses [13]. Also, vaginal pH may be altered by several conditions, including the presence of semen (pH 7.2–8.0) or bacterial infections. In this last case vaginal pH usually increases to values of 5.0–6.5 [14]. These and other variations of vaginal pH may alter the efficacy of administered drugs through quite a few mechanisms, including variable release from the delivery system, alteration in drug absorption and/or metabolism, and modulation of drug activity in the cervicovaginal milieu [15]. *Lactobacillus* species are present in the vagina of healthy women, being responsible for its acidity. These bacteria use glycogen, which is synthesized by epithelial cells under estrogen's influence in order to produce lactic acid, thus reducing vaginal pH [16]. Besides their contribution to vaginal pH, lactobacilli also have the ability to regulate the proliferation of microorganisms, namely pathogenic species, by other mechanisms: adherence to the mucus, forming a barrier which prevents colonization by pathogens or competition for the receptors of the epithelial cells, and the production of antimicrobial compounds such as hydrogen peroxide, lactic acid, bacteriocin-like substances, and possibly biosurfactants [17]. Although lactobacilli are the main organisms responsible for vaginal acidity, recent research suggests that other acid-producing microorganisms, such as *Atopobium* sp., *Megasphaera* sp., and *Leptotrichia* sp., may also contribute to the acidity of the vaginal tract [18]. The composition of vaginal flora is complex, being influenced by hormonal changes, although lactobacilli levels of colonization seem to remain relatively constant during the menstrual cycle [19]. Also, other factors such as glycogen content, pH, sexual intercourse, medication, and immunity status are known to influence the vaginal ecosystem [20].

Although deprived of secreting glandules, the vaginal mucosa is covered with a watery acidic fluid. This fluid includes contributions from vaginal transudation, Bartholin and Skene glands, exfoliated epithelial cells, residual urine, and fluids from

the upper reproductive tract such as cervical mucus and endometrial and tube fluids. Vaginal fluid consists of 90–95% water, inorganic and organic salts, urea, carbohydrates, glycerol, mucins, fatty acids, albumins, immunoglobulins, enzymes, leukocytes, and epithelial debris. Although quite variable, normal daily production of vaginal fluid is estimated at around 6 mL, increasing at midcycle and decreasing around the menstruation period [21]. In fact, the menstrual cycle plays an important role in vaginal fluid characteristics, particularly on pH values, rheological properties, color (from milky white to transparent), and antimicrobial activity. Estrogens induce the production of vaginal fluid, leading to the lubrication of the mucosa. Decline of serum estrogen levels, as during prepubertal or postmenopausal periods, results in a reduction of vaginal moisture [22, 23]. This variability may influence drug release, dissolution, absorption, and removal and thus its activity. Also, vaginal fluid is selectively antimicrobial, lactic acid and to a lesser extent antimicrobial peptides and proteins (calprotectin, lysozyme, histones, and others) being partially responsible by the resistance of the normal vagina to colonization by exogenous microorganisms [24].

Cervical mucus is an important component of the vaginal fluid, although it presents substantial differences when compared with whole vaginal fluid, mainly in pH value (approximately 7.0). It is produced in the cervix, leaking down into the vagina, particularly during the three- to five-day interval that precedes ovulation. Its properties are also influenced by the menstrual cycle, particularly pH (range of 5.4–8.0) and antimicrobial activity [25]. Also, estrogens stimulate the secretion of abundant and fluid cervical mucus, while progesterone induces the formation of thick cervical mucus. These changes in viscosity influence the capacity of sperm and other substances, such as drugs, to pass the cervix and migrate to the uterus [22, 23].

Interaction between vaginal fluids and drug delivery systems is an important aspect that has to be managed during drug design, as it may influence the flow, retention, drug delivery kinetics, and bioactivity of vaginal formulations [21]. Since fluids present in the vaginal environment are difficult to obtain, simulated fluids have been developed in order to emulate their physical and chemical properties. Recently, a vaginal fluid simulant was proposed by Owen and Katz [21], whether Burruano et al. [26] developed a synthetic cervical mucus formulation. These fluids have been successfully used to evaluate vaginal formulations in vitro, being able to mimic with considerable accuracy the physiological fluids. Also, the standardization of a single composition provides the possibility of comparing results obtained by different investigators. Table 1 presents the main features of these two simulants.

Enzymes present in the vagina may influence drug delivery, as they can degrade and influence the permeability of the administered drugs. Although the enzymatic activity in the vagina is not as high as in other drug delivery sites, several enzymes can be found in the vaginal fluid and in different vaginal epithelium cells, namely succinic and lactic dehydrogenase, acid and alkaline phosphatases, β-glucuronidase, phosphoamidase, lactate dehydrogenase, aminopeptidase, and esterases [27, 28]. This enzymatic activity can limit drug bioavailability and decrease the stability of prolonged delivery formulations.

Understanding the immune mechanisms responsible for the defense of the female genital tract is of extreme importance concerning the development of vaccines that are effective against pathogens. The female genital tract has several mechanisms of defense against infectious agents, which appear complementary, additive, and even

**TABLE 1 Main Features of Vaginal Fluid Simulant[a] and Synthetic Cervical Mucus[b]**

| | Vaginal Fluid Simulant | Synthetic Cervical Mucus |
|---|---|---|
| **Composition** | NaCl (3.51 g), KOH (1.40 g), $Ca(OH)_2$ (0.222 g), bovine serum albumin (0.018 g), lactic acid (2.00 g), acetic acid (1.00 g), glycerol (0.16 g), urea (0.4 g), glucose (5.0 g), HCl (to adjust pH), and water (e.q. 1 L) | Guar gum (1.00%), dried porcine gastric mucine (type III) (0.50%), imidurea, (0.30%), methylparaben (0.15%), propylparaben (0.02%), dibasic potassium phosphate (0.26%), monobasic potassium phosphate (1.57%), and water (96.20%) |
| **pH** | 4.2 | 7.4 |

[a]From ref. 21.
[b]From ref. 26.

**TABLE 2 Immune Response of Female Genital Tract**

| | |
|---|---|
| Nonimmune | Passive |
| | Synthesis of protective mucus |
| | pH |
| | Epithelial barrier |
| | Active |
| | Inflammatory reaction |
| | Secretion of humoral soluble factors |
| Preimmune | Humoral response, cellular response |
| Immune | Humoral response, cellular response |

*Source*: From ref. 29.

synergistic. The immune response can be classified in three levels: nonimmune, preimmune, and acquired or specific (Table 2) [29]. Nonimmune response is very effective in limiting the infectious inoculum, while preimmune mechanisms hold up the infection long enough so that the immune response can be activated. These defense strategies are largely influenced by pathogenic agents presented in the vagina and the hormonal milieu [29, 30].

The genital tract is part of the common mucosal immune system, which comprises all mucosal tissues of the body, sharing similarities and common immunization mechanisms, being able to disseminate acquired immunity between them. Nonetheless, it has become clear that the genital tract presents several unique features that differentiate it from other mucosal sites [31]. One of the most important is the ability to only induce local immune responses. In fact, most of the antibodies are produced locally at the mucosa, while those derived from the circulation represent only a small fraction. Also, contrasting with other mucosal tissues, vaginal secretions contain more immunoglobulin (Ig) G than secretory IgA [30, 32]. Hence, immunity of the genital tract is conferred by local production of IgA and, to a less extent, by transudation of serum IgG, although cellular immune response should also be considered [33]. Local immune cell population in the vaginal mucosa includes Langerhans cells, macrophages, T cells, and neutrophils. Systemic immune response is also impor-

tant in the reinforcement of the acquired mucosal immunity or to take over when this one has faded, although they are independent from one another [29, 30].

Sexual hormones, particularly estrogens, also influence the regulation of local immunity, as IgA levels in genital secretions, antigen presentation by vaginal cells, and lymphocyte proliferation vary throughout the menstrual cycle [34, 35]. Estrogens can decrease the concentrations of immunoglobulins, while progestogens can increase their levels [36].

Alterations in vaginal physiology before, during, and after sexual intercourse are important factors in the performance of those formulations intended to be used in this period. During sexual arousal, genital vasocongestion occurs, leading the clitoris and the labia minora to become enlarged with blood and the vagina to increase in length and diameter as a result of relaxation of the smooth muscular wall. The vaginal canal is lubricated by secretions from the uterus, Bartholin and Skene glands, and a fluid that transudates from the subepithelial vascular tissues, being passively transported through the intercellular spaces. Engorgement of the vaginal wall raises the pressure inside the capillaries, thus increasing the transudation of plasma through the vaginal epithelium [6]. The resulting vaginal fluid is increased in quantity, less acidic, and more diluted, allowing penile penetration and thus preventing the male and female genitalia from becoming irritated [3, 4].

### 2.4 Childhood, Pregnancy, and Menopause

The vagina undergoes several lifetime changes that may influence vaginal formulation performance. These changes are particularly important when developing vaginal products that are intended to be used in specific situations, such as labor inducers in pregnant women or hormonal supplements in postmenopausal women.

The vagina of the newborn exhibits the influence of residual maternal estrogens, presenting a stratified squamous epithelium rich in glycogen and becoming colonized with lactic acid–producing microorganisms shortly after birth. By the fourth postnatal week these estrogenic effects disappear, and the vaginal epithelium loses its stratification and glycogen content, becoming much thinner and exhibiting alkaline or neutral pH because of acid-producing microorganism depletion. These characteristics remain throughout childhood, until puberty. By this time, the vagina experiences changes due to adrenal and gonadal maturation. This organ increases in size, vaginal fornices develop, cervicovaginal secretions start being produced, and vaginal milieu becomes acidic. Also, vaginal epithelium thickens and intracellular glycogen production increases [37].

When the fertile woman becomes pregnant, the connective tissue of the vulva, vagina, and perineum relaxes, and the muscle fibers of the vaginal wall increases in size. These alterations prepare the vagina for childbirth. During delivery, perineal and vaginal musculature relaxes and the vaginal rugae flatten, allowing full expansion of the vaginal tract, accommodating the passage of the newborn. Normal morphology and dimensions of the vagina are recovered after 6–12 weeks [37]. The vaginal blood supply is substantially increased during pregnancy, which can enhance systemic absorption of drugs. High levels of estrogens during pregnancy lead to thickening of vaginal epithelium and stimulation of glycogen production. This increased glycogen content promotes lactobacilli growth, consequently decreasing vaginal pH by the enhancement of lactic acid synthesis.

Menopausal women experience a decline in estrogens, which leads to several changes in the genital organs, namely in the vagina. Such changes include atrophy of the labia majora and shortening and loss of elasticity of the vaginal barrel. The number of epithelial cell layers decreases (from 8–10 in premenopausal to 3–4 in postmenopausal), leading to alterations in the barrier capacity and potential increase of mucosal damage and pain and burning sensation during sexual intercourse. Vaginal fluids decrease approximately 50% because of the Bartholin glands atrophy and a decrease in the number and maturity of vaginal cells. A decrease in the colonization by lactobacilli species is observed in menopausal women as a result of the reduction of vaginal glycogen levels, leading to a low production of lactic acid and consequently to increased vaginal pH. In postmenopausal women without estrogen treatment the vaginal pH is estimated to be 5.5–6.8 or even higher [37, 38].

## 3 GENERAL FEATURES OF VAGINAL DRUG DELIVERY

Vaginal drug delivery is mostly used in gender-specific conditions, although it can be a viable alternative for drugs usually administered by other routes. Also, traditionally problematic drugs from a delivery point of view (e.g., peptides) may find in the vaginal route an interesting and promising way for nonparental administration. Limitations and potentialities of vaginal drug administration are intimately connected to this route's idiosyncrasy. Hence, acquaintance of particular features of vaginal drug delivery is required.

### 3.1 Advantages and Disadvantages of Vaginal Drug Delivery

The administration of drugs through the vagina, and eventually their absorption, is a function for which this organ is not physiologically conceived. Nonetheless, the vaginal route of administration presents some advantages. Substances absorbed through the vaginal mucosa bypass the liver before entering systemic circulation, avoiding hepatic first-pass metabolism. Thus, drugs that undergo extensive hepatic first-pass metabolism can benefit from vaginal administration, usually requiring less amount of drug to achieve the same biological effects. Steroids used in hormone replacement therapy or contraception are a good example of molecules that are largely metabolized in the liver, with approximately 95% of orally administered estrogens undergoing hepatic metabolism. Also, these molecules are able to damage the liver when administered by the oral route, an event that can be minimized with vaginal administration. Gastrointestinal side effects are common for many oral administered drugs; the vaginal route may be an alternative to their administration, with the benefit of increased patient compliance. Additionally, vaginal enzymatic activity is lower when compared with gastrointestinal activity, lacking even some important enzymes enrolled in drug metabolism, such as trypsin and chymotrypsin. The vaginal route offers women the possibility of easy self-insertion and removal of drug delivery systems as well as avoidance of the pain, tissue damage, and eventual infection often associated with parenteral routes. Ocular and buccal administration sites frequently become irritated after prolonged contact with drug delivery systems; conversely, the vagina presents less sensitivity, allowing the presence of drug formulations for long periods of time. Although absorption of substances is

not a function of the vagina, features such as its relatively large surface area and rich blood supply contribute to this organ's high permeability to several drugs, allowing higher bioavailability of some active substances when compared to other routes [39–41]. On the other hand, the vaginal drug delivery route has some limitations. Gender specificity is the most important, as it restricts its use to females only. Others, such as misperceptions and cultural issues about genital manipulation and insertion of objects in the vagina, personal hygiene, influence with sexual intercourse, and variability in drug absorption related with menstrual cycle, menopause, and pregnancy, can also limit vaginal drug delivery route usage [41, 42].

### 3.2 Permeability and Drug Absorption

Although some substances are not desired to be absorbed, such as those targeted for local action, permeation of drugs through the vaginal wall into the bloodstream must occur if one seeks to obtain a systemic effect. Although vaginal administration of drugs has been performed for a long time, the capability of systemic drug absorption through this organ was not clarified until the early twentieth century by investigators such as Macht or Robinson [43, 44]. Both conducted independent experiments with substances such as potassium iodide, morphine, atropine, sodium salicylate, quinine, sucrose, and phenol red, which evidenced the permeability of this mucosal tissue, thus opening the possibility of systemic drug delivery through the vagina. Nonetheless, the currently acknowledged variability of vaginal drug absorption with the hormonal status of women, which can limit the potential for systemic drug delivery, was documented in the 1940s by authors such as Rock et al. These researchers reported that although the vaginal administration of drugs, namely penicillin, could provide therapeutic blood levels, the extension of their absorption was highly variable [45].

Recently, several in vitro experiments substantiated the potential of the human vaginal mucosa as a good administration route relating to the degree of permeation when compared with other mucosal surfaces. In fact, the vagina can be more permeable to some commonly used model substances, such as water, 17β-estradiol (Figure 3), arecoline, arecaidine, and vasopressin, than colonic or small intestinal mucosa, or at least as permeable as when compared to human buccal mucosa [46, 47].

In general, systemic drug absorption requires three steps: drug release from the delivery system, drug dissolution in the vaginal fluid, and permeation of the vaginal mucosa. Knowledge of the permeability characteristics of the vaginal mucosa is an important step when developing a pharmaceutical product, with both the physicochemical properties of chemical substances (e.g., chemical nature, degree of ionization, molecular weight and size, conformation, and oil/water partition coefficient) and the biophysicochemical nature of the tissue influencing absorption. The epithelial layer of the vaginal mucosa presents itself as the main permeability barrier for drug absorption. As the epithelium is hormonally dependent, its permeability also changes, usually decreasing with higher estrogen levels because of the induced membrane thickening. However, contradictory findings of enhanced vaginal absorption in postmenopausal women treated with estrogen have been reported [48]. These results can be explained by the increment of vaginal mucosa blood flow that is also induced by estrogen. The transport mechanism of most vaginal absorbed substances is simple diffusion. Lipophilic substances are absorbed through the

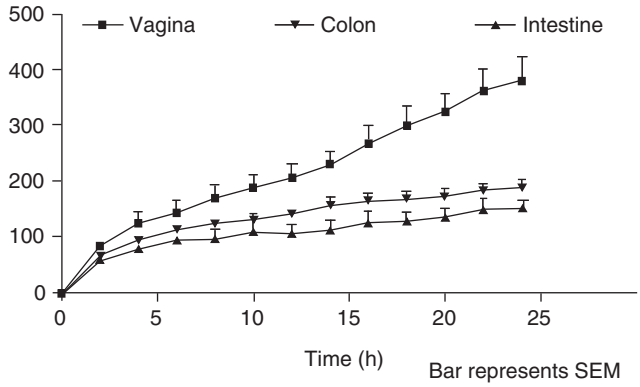


**FIGURE 3** Overall mean flux values of 17β-estradiol across human vaginal, colonic, and small intestinal mucosa, as determined by flow-through diffusion cells. Flux values ($J$) were calculated as $J = Q/(At)$, where $Q$ is quantity of 17β-estradiol crossing mucosa (in dpm), $A$ is mucosa area exposed (in $cm^2$), and $t$ is time of exposure (in min). SEM: standard error of the mean. (Reprinted with permission from P. van der Bijl and A. D. van Eyk, *International Journal of Pharmaceutics*, 261, 147–152, 2003. Copyright 2003 by Elsevier.)

intracellular (or transcellular) pathway, whereas hydrophilic substances are absorbed through the intercellular (or paracellular) pathway or across aqueous pores present in the vaginal mucosa [49]. Also, receptor-mediated transport mechanisms can be involved in the absorption of some substances.

The magnitude of the flux rate across the vaginal mucosa is mainly related to the molecular size and hydrophobicity of the permeating substances. In general, compounds with molecular weight over 300 Da have decreased flux rates, while hydrophobic properties usually increase permeation [50–52]. The influence of the penetrant hydrophobicity/hidrophilicity in the rate and extent of absorption through the vaginal mucosa was demonstrated by Corbo et al. [53, 54]. Experiments performed in rabbits showed that hydrophilicity influenced mucosal permeability of drugs such as progesterone, with increasing hidrophilicity leading to decrease in rate and extent of vaginal absorption. Nonetheless, one should keep in mind that a minimum degree of aqueous solubility is always required in order to ensure that the drug dissolves in the vaginal fluid. Also, pH, viscosity, and volume of this fluid change throughout the menstrual cycle, potentially influencing the extension of drug dissolution: Different pH values influence the drug degree of ionization and thus its solubility; increased viscosity of the vaginal fluid may enhance drug retention, although it may also present a barrier to drug absorption; and higher volumes of fluid benefit drug dissolution but also increase its clearance from the vagina.

However, permeability-enhancing strategies may be required, as many substances cannot permeate the vaginal mucosa in significant pharmacological levels. An interesting and helpful option is the use of several permeation enhancers which can increase drug absorption by interacting with epithelial tight junctions, providing a new intercellular penetration pathway [55]. Some of these substances, such as citric acid, benzalkonium chloride, laureth-9, lysophosphatidylglycerol, sodium taurodihydrofusidate, lysophosphatidylcholine, palmitoylcarnitine chloride, lysophosphatidylglycerol, and sodium glycodihydrofusidate, have been successfully tested [56–60].

Although this strategy may enhance permeability, it presents some disadvantages, particularly the possibility of mucosal damage. These unwanted effects are variable and not always observed; however, they may often be severe [61]. Also, some drugs achieve poor systemic levels after being delivered by the vaginal route, not because of poor permeation through the mucosa, but due to their fast inactivation by local enzymes, particularly when therapeutic peptides and proteins are considered. In these cases, enzymatic inhibitors may be a helpful solution. Vaginal peptidase inhibitors, such as ethylene diamine tetraacetic acid (EDTA), thimerosal, amastatin, bestatin, leuptin, and pepstatin A, were shown to be useful, promoting peptide absorption in rats, as they prevent drug degradation [60, 62]. At this point it is also important to notice that some substances can reduce the permeability of the vaginal mucosa [63], this possibility always being important when designing a drug delivery system. In addition, other approaches, such as the use of mucoadhesive polymers, in situ gelling formulations, or solubility enhancers, have been shown to be useful in improving vaginal permeability of several drugs [40].

### 3.3 First-Uterine-Pass Effect

The first-uterine-pass effect can be defined as a preferential transfer of vaginally administered drugs to the uterus. This effect is due to a countercurrent mode of exchange, with an upward vagina-to-uterus transport of substances absorbed in vaginal and lymphatic vessels and diffusing to nearby arteries [64]. Evidences of higher than expected uterine concentration after vaginal administration of drugs, namely progesterone, terbutaline, or danazol, led to the postulation and verification of this hypothesis [65, 66]. This effect can be of the utmost importance when the uterus is the desired locale for a drug to exert its effects, opening new therapeutic options for uterus-related conditions.

Further investigations showed that the placement of a formulation in different areas of the vagina dramatically influences the observation of the first-uterine pass-effect. Experimental findings suggest that this preferential transfer to the uterus is only observed when absorption occurs in the outer one-third of the vagina [67]. Thus, drugs intended to exert their effects in the uterus should not be inserted deeply in the vagina, as it is often recommended, instead they should be placed near the vaginal entrance.

## 4 VAGINAL DRUG DELIVERY SYSTEMS

### 4.1 Overview

A wide range of drug delivery systems have been used, although many of them are not specifically designed for intravaginal administration. Traditionally used vaginal drug delivery systems include solutions, ointments, creams, vaginal suppositories, and tablets. Recently, others, such as vaginal rings or vaginal films, have been developed. Also, several strategies and improvements have been tested in order to overcome natural limitations of drug delivery through this route, particularly low retention, limited absorption, and cyclic variations.

Most of the currently available vaginal formulations, particularly those that have been marketed for a longer period of time, have serious limitations such as poor

spreadability, messiness, and small capacity of retention in the vagina. Nonetheless, recent advances allowed circumventing some of the major difficulties that hold back the use of this route of drug delivery as a serious alternative to the most traditional ones, with the consequent increase of commercially available drug delivery systems [39]. Drug release of most traditional formulations is rapid, needing frequent administrations to sustain therapeutic drug concentrations. Thus, in recent years, sustained release has been a new approach to deliver several active substances through the vaginal route. Also, vaginal drug delivery systems should ensure either an adequate penetration of the drug within the mucosa, in order to enhance the local effects and reduce systemic absorption, or an ideal permeation of the active substances into the bloodstream in order to assure an effective systemic response.

Before formulators choose a delivery system for a selected drug, several issues should be taken into consideration: physicochemical properties of the active substance, intended effect of the active substance, required drug release profile, excipients to be used and their compatibility with the active substance and vaginal mucosa, women's preferences, and economical implications.

### 4.2 Excipients

When vaginal drug delivery is considered, formulators must select a number of suitable excipients in order to design a drug delivery system able to ensure the therapeutic success of the active substance(s). In fact, it is known that excipients used in vaginal formulations can influence the pharmacological performance of active substances, being able to improve or diminish their activity [68, 69]. The decision of which excipients to use depends to a great extent on the final dosage form and desired characteristics of the drug delivery system. Some excipients can influence drug delivery system performance by changing some properties, such as viscosity, mucoadhesion, and distribution [70]. Although these variations do not interfere directly with the pharmacological effect of the active substances, their availability and thus the formulation clinical outcome can be compromised. Thus, excipient selection must be performed with utmost caution, taking into consideration the quality, safety, and functionality aspects of these materials. Indeed, Garg et al. recently compiled a list of excipients that are currently approved or have already been investigated for vaginal administration [71].

Although by definition excipients are deprived of pharmacological effects, some have showed that this is not always true. For instance, chitosan, an excipient that has attracted a lot of interest in the formulation of vaginal drug delivery systems, exhibits antimycotic effects, particularly against the common vaginal pathogen *Candida albicans* [72]. Also, other polymers commonly used in tablets and capsules, such as cellulose acetate phthalate, have been investigated in the formulation of vaginal microbicides, due to their antiviral effects against HIV [73].

Some commonly used excipients can interact with vaginal and cervical fluids, altering their properties. These interactions should be taken into consideration when designing a drug delivery system, as they can influence in vivo performance. For example, small amounts of nonionic (e.g., polyethylene glycol) and cationic (e.g., polyvinylpyridine) polymers are able to modify the gel structure of the cervical mucus, altering its barrier properties, while like-charged molecules (e.g., polyacrylic acid) interact little with this biological fluid. This approach has been taken into

account, particularly as a new prevention strategy for pathogens that infect via the mucosa, as a new treatment option for diseases that affect the mucous layer itself or even as a strategy for systemic drug delivery routes [74, 75].

### 4.3 Solid Systems

Solid systems commonly administered by the vaginal route include tablets, capsules, and vaginal suppositories.

Tablets are frequently used as vaginal drug delivery systems, being inexpensive and easy to manufacture. They are also easily administered in the vagina, allowing a "clean" insertion that contrasts with the typical messiness of semisolid drug delivery systems. Although very similar to oral tablets, these systems have some particularities, such as being round or oval shaped and devoid of sharp edges, in order to avoid damage of mucosal tissue. These drug delivery systems are usually designed to rapidly release their active substances after being placed in the vagina. In fact, disintegration or dissolution problems, mainly due to the scarce amount of vaginal fluid, are important issues to be managed by formulators. This rapid release and solubility enhancement of the active substances can be important because of the rapid vaginal wash-off and low in situ retention. Increased and faster release of drug content has been achieved using effervescent tablets [76] or including specific excipients that can enhance its disintegration in vaginal fluids [77]. For instance, Karasulu et al. proposed an effervescent tablet made of a mixture of mucoadhesive microcapsules loaded with ketoconazole and effervescent granules [76]. This combination showed ability to improve retention with rapid onset of action. Additionally, other strategies have been used, such as inclusion complexes of poorly soluble drugs with cyclodextrins, in order to improve drug solubility, allowing a rapid onset of the pharmacological effect.

Although fast release of the active substances is a frequent goal, controlled-release tablets can be used in order to enhance their efficacy, because of their prolonged release, and prevent the irritation of the vaginal mucosa that may be caused by some drugs [78]. Nonoxynol-9, a commonly used microbicide and spermicide, known for its irritability when administered in the vagina, is a good example of a drug that can benefit from controlled release. Formulation of double-layer tablets (fast-release outer layer and slow-release core) obtained from coprecipitates of nonoxynol-9 with polyvinylpyrrolidone, can provide extended drug release, allowing a more prolonged spermicidal effect while reducing its irritating effect [79]. Also, controlled release prevents peaks in serum concentration of absorbable drugs, limiting possible systemic effects [80].

Vaginal tablets containing lactobacilli have been used in order to restore the normal vaginal flora. Formulation of these delivery systems requires specific proceedings in order to provide viability of lactobacilli and stability of the final product. Freeze drying of bacterial suspensions has been tested to obtain lyophilized powders for tablet production [81]. These powders were shown to be processable and tablet production was easy and reproducible. Also, the use of double-layer tablets (fast-release layer and slow-release layer) seems to be an interesting approach to lactobacilli administration.

It is common to use tablets designed for the oral route in order to deliver drugs through the vagina. Nonetheless, issues such as delivery system retention and

distribution and drug release can influence the final performance of the formulation, being preferable to use specifically vaginal designed drug delivery systems, or at least study the pharmacokinetics of oral tablets after vaginal administration [82].

Capsules, particularly soft capsules, have been used as vaginal drug delivery systems, but with modest popularity. These systems are relatively stable, particularly when compared with semisolid formulations or vaginal suppositories, being an adequate way to deliver liquid drugs within a solid dosage form.

Vaginal suppositories, also referred as ovules or pessaries, are ovoid-shaped, solid (but generally malleable) dosage forms specifically designed for vaginal administration. These systems usually weigh 2–3 g, although formulations with up to 16 g have been used in the past [83]. Vaginal suppositories have a long history of use as vaginal drug delivery systems, mainly in the management of local conditions. Major advantages are their reduced price and ease of manufacture. However, they present some inconveniences, such as messiness, low retention in the vagina, and poor stability, the last feature due to their temperature and moisture sensibility.

Vaginal suppositories are very close to rectal suppositories in terms of excipient nature and manufacturing process. Thus, they are usually prepared by fusion of the excipient(s) (referred as "base") and incorporation of the active substance(s), this mixture being subsequently poured into molds and allowed to solidify. Other methods, such as by compression, can also be used. Several substances have been utilized as bases for the formulation of vaginal suppositories: gelatin and glycerin, cocoa butter, semisynthetic glycerides, and polyethylene glycol, among others [83]. Composition of vaginal suppositories is importantly related to their melting or dissolution, thus influencing drug release profile. Generally, it can be stated that drug release rate increases as the melting temperature of a suppository decreases or as its dissolution time in vaginal fluids increases. Also, affinity of the drug for the base influences its release from vaginal suppositories: Greater release of drug is expected when there is less affinity between the active substance(s) and the base [84]. The melting temperature and melting process of vaginal suppositories can be characterized by several techniques, such as differential scanning calorimetry and viscosity and dilatometry methods, among others [85]. Specific pharmaceutical characterization of vaginal suppositories includes the determination of disintegration time and breaking hardness. Also, other standard quality control tests include appearance description, surface texture evaluation, pH determination, uniformity of content, and microbial limit testing [86, 87].

Recently, sustained-release vaginal suppositories have been developed in order to attain drug delivery systems with improved performance. Sustained release can reduce the number of administrations, thus improving patient compliance. A base composition consisting of a polymeric gum (carboxymethylcellulose and xanthan gum), a dispersing agent (colloidal silicone dioxide), and polyethylene glycol, referred as long acting, sustained release of spermicide (LASRS), has recently been studied by Zaneveld et al. in order to deliver contraceptives and microbicides [88]. Results showed that a LASRS base is able to spread quickly and evenly over the mucosa, being retained in place for prolonged periods of time and allowing long-lasting efficacy for several active drugs. Preliminary human trials have confirmed these results [89]. In another study, Mandal developed hydrophilic vaginal suppositories comprising mixtures of miconazole cross-linked with poly(vinyl alcohol) by

freeze thawing and different polyethylene glycols that were able to sustain release this antifungal drug for up to 108 h [90].

### 4.4 Semisolid Systems

Semisolid systems present several advantages over other drug delivery systems: They are easy to use and generally inexpensive and have good acceptability. Among their disadvantages, leakage has been one of the most disturbing, mainly because many conventional formulations are not mucoadhesive. The simplest way of dealing with this problem has been the recommendation for night administration, as the supine position diminishes leakage. Also, messiness and discomfort upon application and difficulties in dispensing an accurate dose are important limitations.

Once widely used, ointments have been largely substituted by creams and gels. Nonetheless, some of these drug delivery systems may still be encountered, particularly as hydrophilic bases.

Creams have been used for quite some time as vaginal drug delivery systems, particularly for the administration of sexual hormones and antimicrobials. The main advantage of creams over other semisolid systems is their ability to easily dissolve both hydrophobic and hydrophilic drugs. As most conventional creams do not possess bioadhesive properties, incorporation of bioadhesive polymers is an effective approach to improve their retention in the vagina. Recently, a new approach to vaginal drug delivery was developed using Site Release (SR) technology (KV Pharmaceutical, St. Louis, MO). This technology is based on bioadhesive controlled-release water-in-oil emulsions, being formulated as a vaginal cream (SR cream). The outer oily phase repels moisture (thereby resisting dilution) and retains the dispersed water phase containing the drug (allowing controlled release) [91]. The SR cream allows minimizing leakage and enhancing clinical outcome, requiring less total drug exposure per course of therapy. Site Release technology is currently available in the United States in two commercial products: one containing butoconazole nitrate 2% (Gynazole-1, Ther-Rx Co., St. Louis, MO) and the other containing clindamycin phosphate 2% (Clindesse, Ther-Rx Co.). Clinical findings demonstrated that a single application of Gynazole-1 makes it possible to achieve more rapid relief of vaginal candidiasis symptoms than standard oral therapy with fluconazole [92]. Similarly, Clindesse was show to be able to achieve prolonged local effective concentrations while presenting lower systemic bioavailability and thus less systemic adverse effects when compared with conventional formulations in the treatment of bacterial vaginitis [93]. Also, other drugs have been studied in order to further evaluate the potential of this versatile technology [94].

Since the pioneer work by Wichterle and Lim in the 1960s [95], gels have evolved greatly from simple formulations to advanced drug delivery systems. These systems were soon demonstrated to be good candidates to deliver drugs in the vagina, particularly because of their high bioavailability (mainly because of mucoadhesive properties), biocompatibility, spreadability, ease of usage, and economical savings [96]. Gels are extremely versatile, being used to deliver most of the currently used drugs through the vaginal route.

Recent advances in gel and polymer technology boosted research, opening new possibilities for vaginal drug delivery [97]. Indeed, the development of new and

improved gelling agents, particularly concerning to their mucoadhesive properties, has been of great importance. For example, polycarbophil (Noveon AA-1, Noveon, Cleveland, OH), a mucoadhesive polyacrilic acid polymer, has been widely used as a gelling agent in vaginal gel formulations. This polymer is acidic in nature, which can be useful in reducing the elevated pH associated with bacterial vaginosis [98]. Additionally, acidic polycarbophil gels may be used in the treatment of dry vagina and menopause-related stress incontinence [99].

Also, gel microemulsions have been recently reported as safe and devoid of mucosal toxicity drug delivery systems, presenting intrinsic spermicide activity and the possibility of improving vaginal bioavailability of poorly soluble antimicrobial agents [100].

Although most currently available vaginal gels rapidly release their active substance(s), they can also be formulated to achieve modified drug release profiles [101]. It is not clear how controlled release is achieved, but the analysis of most formulations that claim to possess this feature suggests a combination of dissolution and diffusion control [102]. Gels are also known to be promising drug delivery systems in protein and peptide administration through the vagina, proving to be adequate to accommodate and stabilize sensible molecules such as leuprolide [103].

### 4.5 Liquid Systems

Vaginal douching with liquids containing antimicrobial drugs such as povidone-iodine has been a common practice among women, with the intention of improving personal hygiene and treat vaginitis [104]. These liquids are almost immediately removed from the vagina after administration, thus being inadequate for controlled release. Although vaginal washing is frequently performed by women all over the world, its practice is discouraged, as it is associated with increased risk of acquiring HIV, particularly when soap or other substances rather than water are used [105, 106]. Also, bacterial vaginitis and other adverse reproductive health effects are possible when vaginal douching is a frequent practice [107, 108].

In addition, several solutions are utilized by gynecologists in their office practice. For example, glacial acetic acid solutions (3–5%) are used to identify cervical dysplasia during colposcopy, and Lugol solution is employed to perform Schiller's test (diagnosis of cervix cancer).

### 4.6 Vaginal Rings

Vaginal rings are doughnut-shaped drug delivery systems designed to provide controlled release of drugs. Developed systems are made of flexible, inert, and nonirritating polymeric materials, presenting different dimensions, usually 54–58 mm in diameter and 4–9.5 mm in cross-sectional diameter [109, 110]. Vaginal rings present several advantages particularly important for hormonal contraceptives delivery: (1) They do not require daily attention, allowing higher compliance than with daily dosage forms; (2) flexibility of current rings allow them to be easily inserted and removed by the woman herself, not requiring medical assistance as in the case of subcutaneous or intrauterine devices; (3) the continuous and prolonged delivery (three weeks to one year) of hormones avoids the high peak concentrations and

fluctuations seen with daily oral administration; (4) rings are not associated with adverse local effects, including cytological and normal flora changes; and (5) contraceptive rings may be removed from the vagina during sexual intercourse and up to 2 h, without compromising their pharmacological effect [110–112]. Although vaginal rings have been essentially investigated and used for the delivery of sexual hormones with contraceptive purposes or as hormone replacement therapy, these drug delivery systems can be also useful for the administration of other drugs such as bromocriptine mesylate, danazol, oxybutynin, antigens, and microbicides [113–117].

In the 1960s, first reports that implants made of polysiloxane, containing sexual steroids, could release their content at constant rates in saline solutions provided early information that led to the development of the first vaginal rings [118]. Vaginal rings were initially developed in the 1970s as contraceptives. The first system was composed of a silicone rubber ring containing medroxyprogesterone acetate as the active substance [119]. Nonetheless, the first vaginal rings have just recently reached the market, due to several unpredictable obstacles such as formulation difficulties, safety issues, and poor ovulation suppression [120, 121]. Table 3 presents some vaginal rings currently available in the market.

Vaginal rings may present several designs, as seen in Figure 4. The first vaginal rings were made of a homogeneous matrix containing the mixture of poly(dimethylsiloxane) (matrix-forming polymer) and the active drug, usually referred as a matrix design. Unfortunately, these rings showed an initial burst effect due to rapid release of the drug contained in the system's surface followed by persistent linear decrease of the drug release rate. This later phenomenon is related to the gradually thickening of a drug-depleted boundary between the inner drug-loaded region and the release surface, which is created by continuous drug release from the outer layers. Thus, their use in clinical practice was compromised, namely

**TABLE 3 Selected Vaginal Rings Currently Available**

| Commercial Name | Active Substance(s) | Clinical Indications | Availability | Company |
|---|---|---|---|---|
| Nuvaring | Etonogestrel and ethinyl estradiol | Contraception | United States, Europe, Brazil, Chile | Organon |
| Progering | Progesterone | Contraception in lactating women | Chile | Laboratorios Silesia |
| Femring | Estradiol acetate | Relief of vaginal and urogenital symptoms in menopausal women | United States, Netherlands | Warner Chilcott Laboratories |
| Estring | Estradiol | Relief of vaginal and urogenital symptoms in menopausal women | United States, Canada, Europe, South Africa | Pfizer |

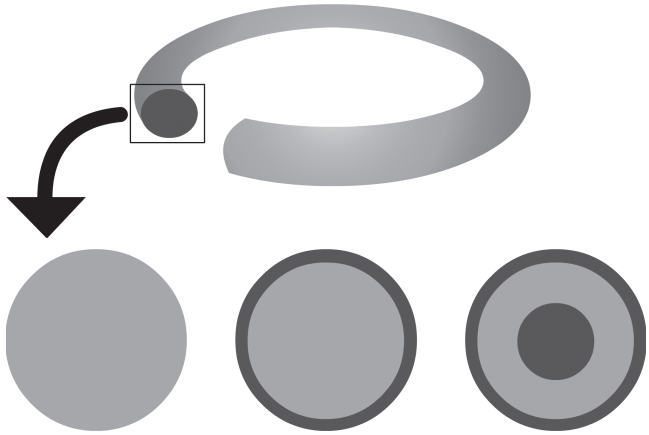

**FIGURE 4** Cross sections of three vaginal rings presenting different designs: matrix design (left), core or reservoir design (center), and sandwich or shell design (right). Light gray represents drug–polymer mixture and dark gray represents polymer only.

as contraceptive devices. Later, and in order to improve control of the drug release, a layer of poly(dimethylsiloxane) without active substance was applied over the core containing the drug, acting as a drug release–limiting sheath (core or reservoir design). This strategy allowed achieving a near zero-order drug release profile. The diffusion rate of reservoir design rings is dependent on the drug concentration in the core, its partition coefficient between the core and membrane, the thickness and surface area of the membrane, and the diffusion coefficient of the drug in the membrane. In order to achieve a constant release rate, the drug should be much more permeable through the core than through the membrane [122]. Rings with several independent reservoirs containing different drugs have been obtained, thereby allowing the administration of two or more active substances from the same device. Also, another design has been developed in order to overcome drug release drawbacks, comprising a core of poly(dimethylsiloxane), an intermediate layer of poly(dimethylsiloxane) containing the active substance, and an outer drug release–limiting membrane of poly(dimethylsiloxane) (sandwich or shell design). As the drug is closer to the releasing surface, this strategy is particularly suited for substances presenting poor polymer diffusion characteristics [123]. As with reservoir design rings, a near zero-order drug release profile is obtained.

Besides poly(dimethylsiloxane), other elastomeric polymers have been employed in the manufacturing of vaginal rings, such as poly(dimethylsiloxane/vinylmethylsiloxane), styrene–butadiene–styrene block copolymer, and poly(ethylene-*co*-vinyl acetate) [123–125]. In fact, poly(ethylene-*co*-vinyl acetate) (commonly referred as EVA) appeared in the mid 1990s as an alternative to poly(dimethylsiloxane), when the manufacturer of this last material stopped supplying it for human use, demonstrating it to be very suitable for the production of controlled-release systems.

At the laboratory scale, silicone vaginal rings are usually obtained by injection molding, where poly(dimethylsiloxane) is mixed with a polymerization catalyst and the drug, being subsequently injected in ring-shaped molds. The mixture is allowed to cure for a period of time at a preestablished temperature, which can range from

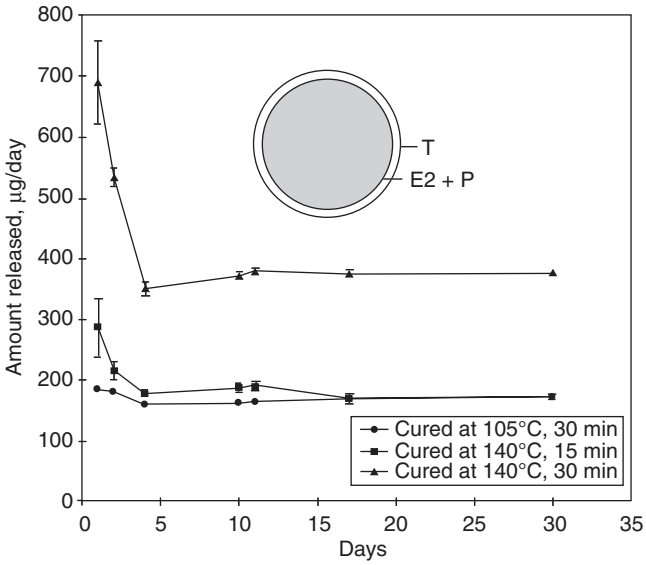


**FIGURE 5** Effect of curing conditions on in vitro release profile of estradiol (µg) from core design rings containing estradiol and progesterone: T: polymer-only outer sheath [poly(dimethylsiloxane/vinylmethylsiloxane)]; E2 + P: core sheath, comprising mixture of polymer [poly(dimethylsiloxane/vinylmethylsiloxane)], estradiol, and progesterone. (Reprinted with permission from S. I. Saleh et al., *Journal of Pharmaceutical Sciences*, 92, 258–265, 2003. Copyright 2003 by Wiley-Liss.)

room temperature to 150°C and over (higher temperatures allow increasing the speed of the ring curing). In fact, curing time and temperature should be optimized as they influence the final performance of the ring, particularly drug release (Figure 5). This step leads to the formation of a three-dimensional network by means of a cross-linking reaction between polymer chains [126]. Afterward, other layers can be added, a step that is usually performed by injection molding or a dipping process. Although vaginal rings produced at the laboratory scale are useful during preclinical and clinical experimentation, the pharmaceutical industry needs other manufacture solutions to allow large-scale and financially viable production of these drug delivery systems. This process scale-up requires proof of bioequivalence between rings obtained by both processes [109].

The most common process of obtaining vaginal rings in the pharmaceutical industry is hot-melt extrusion (or hot-melt spinning), where the polymer, either alone or mixed with drugs or other additives, is melted (usually between 105 and 120°C) and forced by single- or twin-extrusion screws to pass through a die. After leaving the die, the obtained coaxial fiber is cooled and cut, the obtained fragments being shaped as rings by gluing both ends with an adequate pharmaceutical adhesive. Figure 6 presents a simplified scheme of the manufacturing process of a reservoir design ring similar to one used to produce Nuvaring (NV Organon, Oss, The Netherlands), an EVA vaginal ring containing etonogestrel and ethiny lestradiol. The core [polymer and active substance(s)] and the surrounding membrane polymer mixtures are extruded separately through two single-screw extruders (coextrusion)

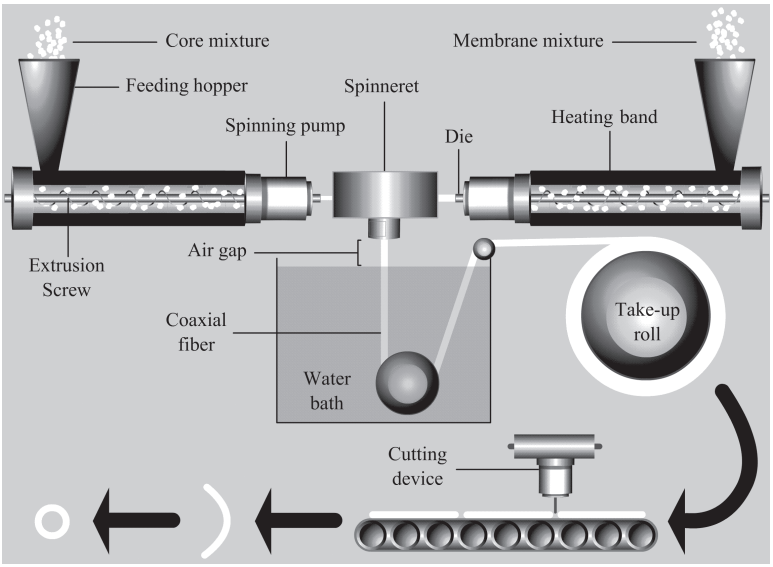


**FIGURE 6** Schematic of manufacturing process of reservoir design ring by hot-melt extrusion.

that are connected to a spinning pump. In these cylindrical pipelines the polymers are melted and extruded through a die at an accurate flow rate. Then, the core and membrane polymers are combined in a spinneret, forming a coaxial fiber. The ratio between flow rates of both spinning pumps determines the thickness of the membrane. After leaving the spinneret, this fiber is cooled, first by air exposure and then by immersion in a water bath. At this stage, the fiber diameter is adjusted to the desired value by elongation with take-up rolls. In fact, after leaving the die, the obtained fiber expands its diameter as a result of the viscoelastic behavior of the polymers used [122, 127].

The drug release profile is conditioned by the polymeric structure of the systems, which is influenced by several parameters such as polymer composition, melt spinning process variables (namely feeding of polymer mixture and spinning velocity, extrusion temperature, spinline stress, cooling rate, and drawing back elongating force), and storage conditions [122, 128]. Also, drug release characteristics are largely influenced by the active substances' molecular weight and diffusion coefficient and solubility in the polymer. For example, extremely hydrophobic drugs and molecules with molecular weight above 450 Da are poorly released from silicone. Modulation of the drug solubility by adding various excipients (e.g., propylene glycol, polyethylene glycol, gelatin, and fluid silicone) can be used to change the release profile [129, 130]. The physical state of the drug is also an important parameter related to drug release. Taking the example of steroids, these compounds can be either in a solid crystalline state or in a molecularly dissolved state. In the first case, the concentration of the drug is fixed by its saturation solubility, making it possible to control the release rate by the thickness and the permeability of an outer membrane. When two drugs are present, this concept cannot be used: Both drugs need to be completely dissolved, their release rates being controlled by their concentrations [123].

### 4.7 Vaginal Films

Vaginal films are polymeric drug delivery systems shaped as thin sheets, usually ranging from 220 to 240 μm in thickness. These systems are often square (approximately 5 cm × 5 cm), colorless, and soft, presenting a homogenous surface. Vaginal films have some advantages, such as portability, ease of application, long time of retention in the vagina, and good drug stability [131]. Once placed in the vagina, the fluid present in the mucosa hydrates the polymer, covering the mucosa with the active substance. This coating may also be helped during sexual intercourse due to the spreading motion of the male penis.

Vaginal films are produced with polymers such as polyacrylates, polyethylene glycol, polyvinyl alcohol, and cellulose derivatives. Proper combination of these polymer is essential to achieve adequate mucoadhesion and optimal drug release profiles. Vaginal films can be produced by casting [132], in which polymer solutions containing the active substance(s) are poured into adequate molds and dried until a thin, solid, and flexible polymeric sheet is formed. Afterward, the sheet is cut in small pieces (individual films) and peeled off.

Vaginal films have been mostly used as spermicides, although they present inferior contraceptive success than hormonal methods, condoms, and intrauterine devices. A contraceptive film containing 28% nonoxynol-9 in a polyvinyl alcohol base (VCF, Apothecus Pharmaceutical) is currently available in the United States. Alternative contraceptive films containing different drugs and film-forming polymers have also been investigated, in order to obtain more effective and acceptable formulations [133].

### 4.8 Medicated Vaginal Tampons

Vaginal tampons have been studied for their feasibility as vaginal drug delivery systems. First experiments used commercially available tampons that were impregnated with active substances by a simple dipping process [134]. Currently, a medicated vaginal tampon, approved as a medical device by the Food and Drug Administration (FDA), is available (Ela Tampon, Rostam, Israel). This bifunctional tampon contains a polymeric delivery system (strips) that absorbs menstrual fluid while gradually releasing lactic acid and citric acid. These two drugs act by preserving the acid vaginal milieu, preventing the proliferation of potential pathogenic bacteria [135].

### 4.9 Vaginal Foams

Vaginal foams have been tested to deliver drugs in the vagina, mainly microbicides or spermicides [136, 137]. The efficacy of these systems was shows to be limited when compared to other available options, leading to a decline of their use. Nonetheless, foams are easy to use, providing a good coverage of the vaginal mucosa with minimal leaking. Also, most foam bases are nonirritating, unlike some other conventional formulations which are reported to cause burning and itching. Thus, improving their formulation can be an interesting approach to obtaining new vaginal drug delivery systems. In fact, foams containing antimicrobials, local anesthetics, and hormones have the potential to gradually substitute several currently available dosage forms, namely creams and ovules.

### 4.10 Vaginal Sponges

Vaginal sponges were once widely used as vaginal contraceptives, the oldest reference to their use in the Talmud (c. 500 B.C.), where a sponge soaked in vinegar was recommended in order to prevent pregnancy [138]. Although these devices have been disapproved by some researchers, since their use encouraged colonization and proliferation of bacteria in the vagina, predisposing women to vaginitis and other genital disorders [139], sponges containing spermicides are still used as contraceptives. These contraceptive devices have the ability to deliver the active drugs while absorbing semen and blocking the cervical canal. Also, vaginal sponges are inexpensive devices; less messy than creams, gels, and foams; and easier to insert and remove than diaphragms [140]. Nonetheless, these systems have demonstrated less efficacy than other contraceptive methods, such as the diaphragm [141].

A vaginal sponge (Today, Whitehall Robins) made of soft polyurethane foam (diameter of 5.5 cm and 2.5 cm thick) and saturated with 1 g of nonoxynol-9 was authorized as a spermicide in the United States in 1983 [140]. Advantages of this formulation include the possibility of being used up to 24 h without requiring additional application of spermicide. Although in 1995 the manufacturer discontinued its production because of increased costs related to FDA guideline compliance, Allendale Pharmaceuticals purchased the rights to the sponge, reintroducing the product in U.S. and Canadian markets [142]. Similar sponges are available outside the United States containing benzalkonium chloride alone (Pharmatex, Innotech International) or a combination of nonoxynol-9 and sodium cholate (Protectaid, Pirri Pharma).

### 4.11 Other Strategies and Vaginal Drug Delivery Systems

Bioadhesion may be defined as the state in which two materials, at least one of which is biological in nature, are held together for extended periods of time by interfacial forces. Mucoadhesion is a particular case of bioadhesion where one of these materials is a mucous membrane or the mucus [143]. Mucoadhesive drug delivery systems can circumvent the poor retention that most traditional vaginal formulations present, improving residence time, and even their specific location in the mucosa. Prolonged time of contact and intimate interaction with the vaginal mucosa are able to increase drug absorption and bioavailability and thus its therapeutic effect. Currently used mucoadhesive materials are polymeric in nature, many of them previously used for other specific purposes (e.g., as gelling agents in vaginal gels). Several polymers, alone or in combination, have been used or investigated for both systemic and local vaginal drug delivery in order to obtain pharmaceutical systems with mucoadhesive properties (see Table 4) [144]. Virtually almost all drug delivery systems may benefit from these mucoadhesives, particularly gels, creams, tablets, vaginal suppositories, and films. To date, polyacrylates are the most used and explored mucoadhesive polymers, although others such as chitosan, carrageenan, or sodium alginate have proved to be advantageous. Also, some widely used mucoadhesives in vaginal formulations, such as polycarbophil (Noveon AA-1, Noveon, Cleveland, OH) present the advantage of being useful as controlled- and sustained-release matrices [145]. Recently, thiolated polymers, usually referred as thiomers, have been studied in vaginal mucoadhesive systems, exhibiting higher mucoadhesion than nonmodified

**TABLE 4 Examples of Mucoadhesive Polymers Used or Investigated in Formulation of Vaginal Drug Delivery Systems**

| Polymer Classes | Examples |
|---|---|
| Carrageenan | Iota-carrageenan |
| Cellulose derivatives | Sodium carboxymethylcellulose (NaCMC) |
| | Methylcellulose (MC) |
| | Hydroxypropyl methylcellulose (HPMC) |
| | Hydroxypropyl cellulose (HPC) |
| | Hydroxyethyl cellulose (HEC) |
| Chitosans | |
| Gelatin | |
| Gums | Pectin |
| | Tragacanth |
| | Dextran |
| | Xanthan |
| Hyaluronic acid and derivatives | Sodium hyaluronate |
| Polyacrylates | Carbopol 974P |
| | Polycarbophil (Noveon® AA-1) |
| Polyethylene glycols (PEGs) | PEG 600 |
| Povidone | |
| Alginates | Sodium alginate |
| Starch | |
| Sulfated polysaccharides | Cellulose sulfate |
| Thiomers | Carbopol® 974P–cysteine |
| | Chitosan–thioglycolic acid (TGA) |
| | Chitosan–4-thio-butylamidin-conjugates (TBA) |

*Note*: As reviewed in ref. 144.

polymers. Although thiomers can improve the mucoadhesion of the original polymers, they do not interfere with other advantageous characteristics such as the ability of controlled release [146, 147].

Environmentally sensitive drug delivery systems are characterized by their ability to respond to changes in pH, temperature, ionic strength, solvent composition, magnetic fields, light, and electric current, among others, most commonly used for the first two [148]. Thermoreversible systems are fluids that can be introduced into the body in a minimally invasive manner prior to solidifying or gelling within the local of administration because of temperature increase from room temperature to body temperature [149]. This innovative approach is particularly interesting in vaginal drug delivery, as a liquid system is easily introduced in the vaginal cavity, while the in situ formed gel facilitates retention. Indeed, some thermosensitive formulations have already been suggested to obtain vaginal drug delivery systems, poloxamer gels being the most extensively studied. A clotrimazole-containing vaginal thermosensitive gel (poloxamer 188 and polycarbophil) was demonstrated to be useful for the effective and convenient treatment of vaginal candidiasis in female rats, with the advantages of facilitating administration and reducing dosing interval when compared with conventional therapy [150].

The incorporation of drug-loaded liposomes in adequate formulations can improve their stability, allowing their applicability in vaginal drug delivery [151]. Additionally, entrapment of drugs in liposomes may improve their solubility and

availability at the site of administration, reducing the administered dose and systemic effects. In order to administer liposomes, adequate dosage forms must be selected. Polyacrylate gels have emerged as good candidates, showing the potential to accommodate liposomes for vaginal drug delivery [151, 152]. Incorporation of liposomes in these gel bases can be achieved by gently mixing both components, which is an advantage over other dosage forms. Also, these novel drug delivery formulations enable sustained drug release, combining both the liposome limiting release effect and the bioadhesive properties of these gels.

When liposomes are considered for vaginal administration, their stability should be assessed under in situ conditions presented by both pre- and postmenopausal women. Drug formulation can help improve their stability. Several gels made of Carbopol 974P NF were shown to be suitable in improving the stability of liposomes containing clotrimazole, metronidazole, or acyclovir in a vaginal environment simulating media when compared to liposomal dispersions [153, 154]. Also, compatibility between liposomes and vaginal mucosa (animal or human) should be checked [155]. Nonetheless, liposomes still have some stability-related problems that limit their shelf life. Formulation of liposomes as proliposomes can prevent these stability issues. Proliposomes are dry, free-flowing products which on addition of water disperse to form liposomal suspensions [156]. As proliposomes are dry powders, they can be formulated as solid dosage forms, which can be more convenient to administer in the vagina. Recent work by Ning et al. demonstrated that clotrimazole-containing proliposomes can be suitable for vaginal administration, allowing a rapid conversion to sustained-release liposomes upon contact with physiological fluids [157].

Niosomes (nonionic surfactant vesicles) have been experimented as controlled- and prolonged-release drug delivery systems to administer insulin through the vagina. These particles might be good carriers for protein delivery, showing the potential to enhance the effects of these drugs when compared to the vaginal administration of free insulin. Results obtained in rats showed that niosomes containing insulin can achieve comparable bioavailability with subcutaneous administration of this protein [158]. Also, experiments performed with vaginal gels containing both niosomes and liposomes have been shown to be a promising strategy for the prolonged release and safe vaginal administration of clotrimazole [159].

Microparticles and nanoparticles present some advantageous features, namely mucoadhesive properties. They have demonstrated some potential in vaginal drug delivery, particularly in the formulation of delivery systems for vaccines or peptides and proteins [160, 161]. Nonetheless, these particles have to be incorporated in adequate carrier systems in order to be delivered. This task has been shown to be complex, it being hard to achieve controlled-release and steady-release profiles.

Cyclodextrins are commonly used in pharmaceutics, their applicability being no exception in vaginal drug delivery. Drug/cyclodextrin complexes allowed enhancement of solubility and achievement of prolonged-release properties of drugs such as clotrimazole and itraconazole when included in suitable vaginal drug delivery systems such as gels, creams, or tablets [162–164].

Cervical barrier devices such as diaphragms can be modified to include a reservoir that releases active substances such as spermicides or microbicides. Although these devices have limited applicability, they can be particularly helpful in prevent-

ing pregnancy, as they combine a physical barrier device with a chemical spermicide that faces the vagina, after correct placement in the cervix [165]. Lee et al. presented a diaphragm made from silicone containing 35% nonoxynol-9 that was able to achieve controlled drug release depending on the device design and size [166]. The diaphragm was prepared by using compression molding in a single-cavity aluminum mold.

Patches can provide delivery of very toxic drugs in very limited areas of the vaginal mucosa without leaking to the circumventing tissue. McCarron et al. described a bioadhesive patch that delivered 5-aminovulinic acid to intraepithelial neoplasia lesions. The proposed delivery system is based on a poly(methylvinylether/maleic anhydride) matrix that contains the active drug, providing local retention of up to 4 h and enhancing efficacy without damaging healthy epithelium [167]. Patchs with bilayer design may also be an option for the treatment of these neoplasic lesions, where a bioadhesive drug-loaded matrix bonded to a drug-impermeable backing layer is able to prevent drug spillage to healthy tissues, promoting unidirectional and deep drug penetration. Woolfson et al. presented a 5-fluorouracil vaginal patch with a bilayer design, comprising a flexible polyvinyl chloride (PVC) emulsion as a backing layer and a drug-loaded bioadhesive film made of 2% Carbopol 981 and 1% glycerin as a plasticizer [168].

Propess (Ferring Pharmaceuticals) is a commercially available controlled-release vaginal insert presented as a thin macrogol 8000 matrix (rectangular in shape with radiused corners) containing 10 mg of dinoprostone being used for labor inducement. This drug-loaded matrix is included within a knitted polyester retrieval system that ends in a long tail to help retrieval at the end of the dosing interval. This insert is capable of releasing the active substance at a rate of 0.8 mg/h over 12 h, after being exposed to vaginal moisture [169]. Clinical studies indicate that Propess is as equally effective and safe in achieving cervical ripening as other commercially available vaginal drug delivery systems containing dinoprostone [170].

Multiple emulsions have been studied as drug vehicles for vaginal administration that are able to include several active substances in the different phases. An antimicrobial water–oil–water (W/O/W) multiple emulsion containing lactic acid in the internal aqueous phase, octadecylamine in the oily phase, and benzalkonium chloride in the external aqueous phase was tested, proving to stabilize the included active substances while providing adequate viscosity properties to vaginal administration [171]. Nonetheless, the structure of these formulations is destroyed at high shear rates (e.g., shear rates observed during coitus), losing or diminishing their intended activity and thus limiting their applicability.

Recently, an interesting approach for protein delivery using genetically engineered normal vaginal flora as delivery systems has been proposed. This strategy is based on the natural affinity that these microorganisms have to adhere tightly to the epithelial surface, providing a direct delivery of the drugs to the mucosa and thus minimizing enzymatic and bacterial degradation. Lactic acid bacteria, transformed with plasmids that contained a gene encoding for the therapeutic protein to be administered, have been used as delivery systems. Obtained results showed enhancement of protein delivery when compared with a conventional solution containing the same molecule. This strategy looks particularly appealing for the development of vaccines that induce mucosal immunization [172, 173].

### 4.12 Packaging and Vaginal Applicators

Vaginal packaging and applicators are an integral part of vaginal products. Packaging is designed to accommodate and protect formulations, while applicators should allow their convenient administration in the vagina. Several materials have been used to manufacture these devices, such as plastics (e.g., polypropylene and polyethylene) and nonlatex rubber. Besides compatibility, stability, and suitability issues, these materials should be selected regarding the final cost of packaging and applicators.

Applicators are intended to be introduced in the vagina, adequately deliver the product, and then be removed. Their design relates to safety (e.g., relationship with product purity and stability, avoidance of local trauma associated with insertion or use), efficacy (e.g., consistent delivery of the required amount of product in the intended location), and acceptability (comfort, ease of use, convenience, aesthetic appeal) [174]. In general, they can be divided as single-use or reusable applicators. Single-use applicators are usually prefilled, while reusable applicators are filled by women prior to vaginal insertion. Several applicator designs have been used, such as barrel-and-plunger and squeeze tube, but they all should be easy to insert, comfortable, and deprived of cutting edges. Also, some specific formulations, such as those intended for vaginal douching, require other types of applicators. Typically, squeeze plastic bottles with variable volumes (approximately 100–200 mL) are used.

## 5 PHARMACEUTICAL EVALUATION OF VAGINAL DRUG DELIVERY SYSTEMS

Evaluation of pharmaceutical systems is consensually recognized as an important component of their development and quality control. Although evaluation of general features (e.g., drug content) is also required for vaginal formulations, this section focuses only upon some of the most important parameters that are intimately related to drug delivery systems specifically designed to be administered by this route.

### 5.1 Legal and Official Compendia Requirements

There is a lack of well-defined guidelines and regulations for vaginal products in most countries as well as official compendia information and requirements on quality control and other important aspects of vaginal drug delivery systems [39]. In fact, the latest editions of the U.S. Pharmacopeia, European Pharmacopoeia, and Japanese Pharmacopeia include limited or even no information related to the quality control and evaluation of vaginal drug delivery systems. Thus, most of the currently used evaluation procedures require standardization, making it difficult to compare results obtained by different research groups.

### 5.2 Drug Release and Permeability

When formulating vaginal drug delivery systems, it is important to consider the release of the active substances, as different formulations can greatly affect the

release rate and ultimately their pharmacological effects. The choice of the dissolution method should be done on a case-by-case basis, while dissolution profiles can then be fitted to commonly used mathematical models, as reviewed by Costa and Sousa Lobo [175]. Conventional procedures and apparatus have been adapted taking into consideration vaginal physiological specifications such as pH, fluid volume, and temperature, among others. The experimental method may not inevitably imitate the vaginal environment, but it should test the main key performance indicators of the formulation. Although official methods are not available, some have been recommended for specific vaginal drug dosage forms [176]. For semisolid vaginal formulations, the Franz diffusion cell is considered the most promising apparatus for drug release investigation. In the case of hydrophilic vaginal suppositories, a basket apparatus, a paddle apparatus, or flow-through cells can generally be considered as suitable; for hydrophobic vaginal suppositories, modified flow-through cells would be preferable. Dissolution methods that use a basket instead of a paddle can be advantageous for vaginal solid formulations that tend to float, particularly those that may include a modification that prevents the formulations to form a cake inside the basket, limiting their dissolution [177, 178]. Drug release from vaginal rings is usually determined by placing these systems in conical flasks containing an adequate dissolution medium. Flasks are then placed in a water bath controlled at 37°C and shaken during the time of the assay, with samples being taken and release medium being replaced typically every 24 h [113, 129]. Correlation of these in vitro release tests with in vivo results proved to be satisfactory for the majority of the evaluated vaginal rings [125].

Absorption of drugs through the vagina is an important parameter to be evaluated, particularly when a systemic effect is required. Also, assessment of the absorption potential of drugs intended to locally exert their effects needs to be evaluated, as this event can lead to unwanted systemic effects. The evaluation of both formulated and unformulated drugs can be performed by in vitro or in vivo methods.

In vitro permeability studies have been performed in flow-through diffusion cells using either animal or human vaginal mucosa [179, 180]. After isolation and adequate treatment of vaginal mucosa specimens, small tissue disks are mounted in the apparatus in order to perform the permeation experiments. At the end of these procedures, routine histological examination of the used tissues can be performed in order to identify any changes in the normal structure of the vaginal mucosa. Several alterations can suggest potential mechanisms of permeation of tested drugs [181, 182]. Animal mucosa used in these experiments can be obtained from several species, namely rabbits and pigs. Also, porcine vaginal mucosa was demonstrated to be a good in vitro permeability model of human vaginal mucosa, particularly when hydrophobic substances are tested [50]. This feature can be explained because both mucosal tissues are very similar in many aspects, namely their lipid composition and histological structure. On the other hand, high molecular weight or charged molecules, such as oxytocin, may show different permeability profiles when tested with either porcine or human vaginal mucosa. Thus, researchers must be careful when interpreting and extrapolating results to human tissues because unexpected differences often occur. When human vaginal mucosa is used, samples are usually obtained from excess tissue removed from postmenopausal women after vaginal hysterectomy [181, 182]. Use of postmenopausal vaginal mucosa is advantageous as it is less altered (particularly in thickness) by hormonal stimulation, leading to more uniform

results. Nonetheless, these features do not reflect the normal histological architecture of fertile women, being able to significantly alter the permeability profiles for many drugs. Along with vaginal mucosal tissue, other model membranes, such as vaginal and cervical cell monolayer membranes, have been suggested in recent years in order to predict in vivo absorption [172].

It is also noteworthy that many of the in vitro results of vaginal permeability studies cited in the literature have limitations related to the experimental conditions that were used, particularly pH values at which they have been performed. Differences in permeability values are particularly expected when ionization characteristics change between experimental pH and vaginal pH [183]. However, in vitro results should only be considered as evidence that the vaginal mucosa is able to be permeated.

In vivo studies performed in animals are an important step before considering human experimentation. Animal species commonly used in vaginal permeability studies include rabbits, rats, and mice [53, 158, 184]. Although potentially more accurate in predicting human vaginal absorption of drugs, animal experimentation have some limitations. A major problem is the variability of the vaginal epithelium properties throughout the estrous cycle, thus influencing drug absorption [184, 185]. In order to minimize this variability and standardize the thickness of the epithelium, animals are usually ovariectomized. Also, vaginal enzymatic activity is an important parameter in choosing animal models. It is recommended that the enzymatic profile of such animal should be comparable to that of the woman. Taking this into consideration, rats and rabbits seem to be good models for vaginal permeability studies, particularly when protein and peptide drugs are considered [27].

### 5.3 pH and Acid-Buffering Capacity

As already referred, pH is an important parameter concerning the health and normal physiology of the vagina, being important that vaginal formulations do not interfere with its normal value. Also, the pH of the vagina can be elevated due to changes in its normal physiology (e.g., bacterial vaginitis) or the presence of semen. Vaginal formulations presenting good acid-buffering capacity have the potential to reestablish normal pH or to prevent it from rising.

The acid-buffering capacity of a vaginal formulation can be measured by simple titration with an inorganic alkali, such as sodium hydroxide. The physiologically relevant acid-buffering capacity can be defined as the amount of alkali required to elevate the pH from its initial value to the maximum desirable value when considering the healthy vagina [186]. Also, mixtures of vaginal formulations with semen may be useful in determining their buffering capacity, this proceeding being particularly helpful when testing products used during sexual intercourse.

Variations in vaginal pH of the vagina can influence drug stability, particularly when extreme values are observed. Thus, the adjustment of the formulation pH can also be important in order to assure maximum stability or pharmacological activity of the active substance(s). As an example, the administration of antibodies in the vaginal milieu can compromise their activity because of the acidic pH. Generally monoclonal human antibodies are more stable at pH 4–7, losing binding and neutralizing activity below pH 4 [187]. These findings underline the importance of pH buffering when delivering pH-sensible molecules such as antibodies.

### 5.4 Rheological Studies

Rheological properties of semisolid vaginal formulations are crucial to their suitability as drug delivery systems. They experience in vivo a wide range of shear rates (from less than $0.1\,s^{-1}$ to about $1000\,s^{-1}$), both steady and transient, while being diluted with vaginal fluids, which influences their rheological properties and hence their spreading and retention. Events such as passive seeping, sliding, squeezing between vaginal walls, and coitus, among others, influence the rheological performance of vaginal formulations [188]. Thus, knowledge of rheological properties of semisolid vaginal drug formulations may assist in improving their design, being helpful in the process of predicting which formulations can retain their structural stability over time, particularly in the physiological environment [189, 190].

Qualitative and quantitative composition of a semisolid vaginal formulation can strongly influence its rheological properties [191]. This fact is particularly important when considering the optimization of a drug-containing formulation and its placebo formulation. These two systems should only differ in the absence of the active substance(s). However, this small difference can sometimes originate different rheological properties that can greatly influence the formulation's performance and even the results of clinical trials [192].

Rheological properties of a formulation can be studied either by simple flow measurements or by dynamic oscillatory measurements, although the latter are preferable as they allow a complete characterization of both elastic and viscous components. Also, they are nondestructive and, if the strain is not too high, the sample is not disturbed [193]. As already noticed, in vivo conditions, particularly temperature and fluids that may be present in the vagina (vaginal fluid, cervical mucus, and semen), can influence the rheology of pharmaceutical formulations. Indeed, these factors have to be taken into account when formulating a vaginal drug delivery system. Thus, optimization should not only focus upon the rheology of undiluted material but also include mixtures of formulations and fluids that may be present in the vagina [194]. However, these biological fluids are not always available and considerable differences between individuals limit their use. In order to abbreviate these limitations, some simulants of vaginal fluid [21], cervical mucus [26], and semen [195] have been used.

### 5.5 Textural Studies

Textural profile analysis is a widely used analytical method based on the measurement of the forces involved during the compression/decompression of a probe in a sample of the product to be tested. From the obtained results, important parameters can be calculated, including hardness (force required to attain a given deformation), compressibility/spreadability (the work required to deform the product during the first compression cycle of the probe), and adhesiveness of the product. Besides their influence in the ease of removal from a container (e.g., vaginal applicators), the ease of application, or retention, among others, it is consensual that textural properties of formulations will influence their clinical performance. Therefore, it is important to fully characterize these properties during the formulation process [196]. Also, these textural parameters can be converted into rheological properties, such as shearing stress, shear rate, and viscosity, using dimensional analysis, allowing the comparison of results generated by both techniques [197].

### 5.6 Mucoadhesion

As previously discussed, development of mucoadhesive drug delivery systems is a promising strategy in order to enhance vaginal drug administration. Several methods for the in vitro evaluation of mucoadhesive properties of vaginal formulations may be found in the literature. Although these tests present some limitations, they can offer easy and valuable tools in the initial formulation and evaluation of vaginal drug delivery systems. Most methods can be categorized as tensile strength or shear stress tests, where the force needed to separate a model membrane attached to the formulation measures mucoadhesion. Different results can be obtained for the same sample because of the different types of forces involved in these two methods [198]. For this reason, both types of tests should be performed in order to achieve a more complete characterization of the mucoadhesive potential of formulations.

Alternative methods have also been shown to be useful. For example, mucoadhesive properties of semisolid vaginal formulations can be assessed by the rheological characterization of the mucoadhesive interface, based on the assumption that the interpenetration extension between polymer gels and their mixtures with mucin can be detected by measuring differences in rheological parameters. Also, the texture analysis of these formulations/mucin interfaces was demonstrated to be a useful technique in measuring bioadhesion, with results in the same rank as the ones obtained with the rheological technique [199]. Kast et al. tested vaginal tablets for their mucoadhesiveness by a simple method, comprising the use of a dissolution apparatus [146]. In this procedure, the formulation to be tested is attached to a vaginal mucosa that is fixed on a cylinder; the cylinder is then immersed in the dissolution vessel and rotated in an adequate testing solution at approximately 37°C. The mucoadhesion is evaluated by the detachment time of the formulation. An alternative method based on a modified balance, in which the tablet is attached to a vaginal mucosa fixed to one side plate, has also been proposed [200]. In this case, mucoadhesion is measured by the weight required to detach the formulation.

### 5.7 Vaginal Distribution and Retention

Vaginal mucosa coating by a formulation is an important parameter to be determined, since its action may depend on the effectiveness of this phenomenon. Nonetheless, knowledge about the distribution and retention of commercially available products is limited. After being administered, liquid, semisolid, or solid (after liquefaction) formulations can spread throughout the vagina, with part of the material being able to exit this tube, either to the exterior or to the upper genital tract [201]. This distribution is governed by physical forces that include gravity, normal forces from contacting tissues, surface tension, and shearing. Formulation flow due to these forces is affected by many factors, including formulation physical properties, amount of formulation applied, surface interactions, surrounding tissue properties, vaginal secretions, baseline dimensions of the vagina, ambulation and posture changes of the user, and sexual intercourse [202, 203]. Drug delivery systems intended for local effect should ideally spread evenly throughout the vaginal mucosa. Also, evaluation of the possible erosion of a product coating during its residence in the vagina, particularly during sexual intercourse, is extremely important. Thus, vaginal distribution and retention studies may clarify some questions, such as the amount of product to

cover evenly the vaginal surface, the time required to distribute to all areas and to be removed from the vagina, and the effect of daily activities such as ambulation and sexual activity in these phenomena.

Simple in vitro tests can be used to quantify the vaginal coating of a vaginal formulation. It is important that these tests model the natural history of a product in the vagina, from initial application and contact with fluids that may be present in the vagina to the period during and following sexual intercourse [204]. The distribution and retention of solid drug delivery systems can be evaluated by a simple method proposed by Ceschel et al. [205], where the formulations are placed in a vertical thermostated cellophane tube, the discharged liquid collected and measured throughout the time of the experiment. The amount of discharge liquid is related to the retention while the distribution of formulations can be assessed by dosing the amount of active substance(s) in different sections of the tube at the end of the experiment. Also, the contribution of gravity to the vaginal coating flows of vaginal semisolid drug delivery systems can be evaluated by a simple and objective method proposed by Kieweg et al. [203]. The proposed technique is based on the measurement of the flow behavior of a formulation sample after being placed in an inclined plane surface. Obtained results, together with mathematical models, can help formulators to select primary candidate formulations before in vivo tests commence.

In vivo assessment provides more reliable and complete information about the vaginal distribution and retention of drug delivery systems, although ethical and economical issues limit its applicability to routine evaluation. In recent years, imaging techniques, which have been used for other purposes for a long time, emerged as valuable tools for the evaluation of vaginal distribution and retention. Magnetic resonance imaging proved to be a helpful method, providing cross-sectional images of drug delivery systems that are administered in the vagina, both in animals and in humans, allowing precise reproducible data regarding the spread of vaginal formulations to be achieved [202]. As magnetic resonance imaging reflects the images of a chemical label, such as gadolinium, and not necessarily that of the drug delivery systems or carried drugs, it is necessary to perform the association of the chemical label with vaginal products and its validation [206]. Another interesting imaging technique that has been used in this type of assessment is gamma scintigraphy, shown to be a useful tool for evaluating and comparing the distribution, spreading, and clearance of vaginal delivery systems [207, 208].

Imaging methods, such as gamma scintigraphy and magnetic resonance imaging, are useful but have some limitations concerning their resolution, being unable to quantify or even identify the presence of vaginal coating layers of just a few hundred micrometers. In order to overcome such limitations an optical instrument capable of detecting coating layers as thick as 50 μm has been developed by Henderson and co-workers [209]. The device is inserted and remains stationary in the vagina, where both local video images and fluorescence intensity measurements of fluorescein-labeled formulations are obtained. Since the tube that is inserted in the vagina is shaped and sized like a phallus, the vaginal coating measured is analogous to that observed during sexual intercourse. Nonetheless, in vivo fluorescence-based methods present limitations, particularly because of the diffusion of the dye out of the formulation and photobleaching, limiting the interval over which measurements can be performed accurately. A new technique based on low-coherence interferometry that can overcome these difficulties is being developed, allowing extended time

studies to be performed [210]. This easy, label-free, high-resolution method uses broadband light in an interferometry scheme to achieve depth-resolved reflection measurements. Future studies focus on the development of an easy-to-use endoscopic device that may be used in clinical studies.

### 5.8 Safety and Toxicology

It is accepted that a new pharmaceutical product should be assessed for its effects on the vaginal mucosa before being approved by drug-licensing agencies. Nonetheless, many of the older products that have been used for a long time by women all over the world do not have this type of information available. Safety issues are particularly important when a vaginal drug formulation is used repeatedly, as in the case of microbicides, spermicides, and contraceptive vaginal rings. Also, it is important to test vaginal applicators for their safety as they are considered an integral part of vaginal products, being able to induce alterations in the mucosa [211].

Local effects assessment should include not only short-term but also long-term protocols, as some formulations are intended to be used for large periods, in order to assess their real impact on vaginal health. Although systemic exposure to drugs intended for topical action is expected to be minor, vaginal formulations should also be assessed for systemic effects due to possible absorption. Alterations in blood parameters and liver and renal function should be investigated [212]. Also, drugs and formulations to be administered through the vaginal route must be assessed for fertility and teratogenic effects in animal models, before entering clinical trialing and human use [213].

In vitro testing helps formulation scientists understand and predict the potential harmful effects of formulations to the vaginal mucosa. Although animal testing still needs to be performed, these in vitro methods can be of great interest in initial screening of new products and formulations, reducing the amount of animal testing required. Also and unlike animal testing, in vitro testing can often differentiate products that are very mild in terms of toxicity potential. Thus, toxicity of vaginal products can be assessed using simple tests with epithelial cell monolayers, where maintenance of membrane integrity in the presence of testing formulation indicates potential safety [214, 215]. Nonetheless, monolayer cell cultures lack histological and functional resemblance with native ectocervical and vaginal tissues, which limits the interpretation of the obtained results. In order to respond to this and other problems, Ayehunie et al. recently proposed a fast and highly reproducible three-dimensional organotypic vaginal–ectocervical tissue model that simulates the structure of the vaginal epithelium [216].

The standard preclinical test of local vaginal irritation and toxicity of pharmaceutical products, and the only one recommended by the FDA, is the rabbit vaginal irritation test [217, 218]. However, reproducibility problems and differences in vaginal physiology when compared to women limit the interpretation of results. Other animal models have also been used, namely primates, dogs, guinea pigs, pigs, mice, and rats [219–223]. Classic animal testing is limited because of the number of animals required, which makes testing burdensome, expensive, and ethically questionable, and because of differences between species, which may jeopardize the extrapolation of results to humans [224]. Thus, simpler toxicity tests performed in nonvertebrate organisms, such as gastropods, are interesting alternatives to

vertebrate animal testing. These tests also proved to be superior than in vitro testing, mainly because of the limitations that are intrinsic to simple cell culture models [225]. Recently, a simple in vitro test using slugs (*Arion lusitanus*) has been proposed as an alternative to vertebrates in order to screen new vaginal semisolid formulations for local tolerance early in the development process. The irritation potential is evaluated by mucus production, and protein and enzyme release (lactate dehydrogenase and alkaline phosphatase). Experimental results showed that the slug mucosa irritation test performance is comparable to the classically used rabbit vaginal irritation test [226].

When initiating human testing, symptoms and signs of genital irritation must be assessed [227]. These investigations should be performed comparing results between the formulation to be tested (vehicle plus active substance(s) and vehicle only) and formulations that are well known for their irritative effects. In the late 1980s and early 1990s colposcopy of the vagina and cervix began to be used in the in vivo safety assessment of vaginal products, becoming a standard technique [228]. The objective of this procedure is to detect epithelial changes, such as breaks in the epithelium, inflammation, or other not well characterized, that may be a consequence of vaginal products usage. Although very important, it presents several limitations such as costs, specific personnel training, and difficulty in understanding which colposcopic findings indicate risk [229]. Also, other techniques, such as Papanicolau stained smears or automated cytomorphometric analysis, have been used in order to assess the effect of formulations on the vaginal mucosa [230].

### 5.9 Other Characteristics

In addition to the discussed evaluation tests and methodologies, other characteristics of vaginal formulations may be assessed according to their individual specificities. For example, the compatibility of vaginal formulations with condoms is an important parameter to be determined, particularly when they are used during sexual intercourse. These studies are usually performed according to the American Society for Testing and Materials (ASTM) norm D3492-89 (Standard Specification for Rubber Contraceptives), where condoms are tested by accelerated testing for their tensile strength and elongation on break point after being exposed to vaginal formulations [231, 232]. Also, it is important to consider the effects of vaginal formulations that are used during sexual intercourse in the penis and the possibility of drug penetration through this organ. Although the human penis is covered with keratinized stratified epithelium, and thus the expected absorption should be less than that of the vaginal epithelium, it is always a possibility to be taken into account [233].

## 6 CLINICAL USAGE AND POTENTIAL OF VAGINAL DRUG DELIVERY

### 6.1 Microbicides

Microbicides (initially termed "virucides") are anti-infective drugs formulated for topical self-administration in the vagina before sexual intercourse in order to protect against HIV and other sexually transmitted pathogens [234]. Once a neglected

subject in the war against HIV and other sexually transmitted diseases, in the last decade microbicide investigations have gained important boosting and interest by the scientific community, being considered as a new approach for prevention [235]. Despite the fact that several microbicides are already in clinical trialing (see Table 5), currently there are no available products on the market. Also, 100% effective microbicides are not likely to be achieved, even though only partially effective, microbicides can be a big help in reducing the spread of HIV infection. Investigators estimate that the use of a 60% effective microbicide in only 20% of all coital acts could prevent approximately 2.5 million infections over a period of 3 years [236]. While waiting for the first-generation microbicides, preclinical research in new and improved microbicides is already in progress. Several candidates, such as PSC-RANTES [237], antimicrobial peptides [238], monoclonal antibodies [239], inhibitors of virus–cell fusion [240], and natural products [241], are currently being developed. Also, another strategy that seems to be gaining consensus among the

**TABLE 5 Microbicides Currently Undergoing Clinical Trialing**

| Mechanism of Action | Active Substance(s) | Candidate Product | Drug Delivery Systems | Developers | Clinical Trial Status |
|---|---|---|---|---|---|
| Vaginal defense enhancer | — | ACIDFORM | Gel | CONRAD/Instead | Phase I |
| | — | BufferGel | Gel | Reprotect | Phases II/III |
| Membrane disruptive agent/ surfactant | C31G | Savvy | Gel | CONRAD | Phase III |
| Entry/fusion inhibitor | Carrageenan (PC-515) | Carraguard | Gel | Population Council | Phase III |
| | Cellulose acetate 1,2-benzenedicarboxylate | Cellacefate/ CAP | Gel | Lindsey F. Kimball Research Institute / Dow Pharmaceuticals, Inc. | Phase I |
| | Cellulose sulfate | Cellulose Sulfate Vaginal Gel | Gel | CONRAD | Phase III |
| | Sodium lauryl sulfate | Invisible Condom | Gel | Laval University | Phases I/II |
| | Naphthalene 2-sulfonate polymer | PRO 2000/5 | Gel | Indevus Pharmaceuticals | Phase III |
| | SPL7013 | VivaGel | Gel | Starpharma | Phase I |
| Replication inhibitor | Tenofovir | PMPA Gel | Gel | Gilead Sciences | Phases II/IIb |
| | Dapivirine (TMC120) | TMC120 Vaginal Gel | Gel and vaginal ring | International Partnership for Microbicides | Phases I/II |
| | UC-781 | UC-781Gel | Gel | CONRAD | Phase I |
| Unknown mechanism | Extracts of *Azadirachta indica, Sapindus mukerossi*, and *Mentha citrata* | Praneem polyherbal | Vaginal tablet | Talwar Research Foundation | Phases II/IIb |

scientific community is the synergistic association of microbicides with different action mechanisms in order to improve protection [242].

Early expectations created around the possibility of using nonoxynol-9 (widely used as a vaginal contraceptive) as an effective microbicide, based on its in vitro efficacy against HIV, were frustrated in clinical trials [243]. Several hypotheses for this failure were brought up, an inadequate choice of drug delivery system being one of them. These results confirm that preformulation and formulation studies play an important role in microbicide rational design and development, being a big challenge to overcome. Although clinical development recommendations have been extensively reviewed [244, 245], pharmaceutical development algorithms for microbicides are yet to be defined. Nonetheless, it is known that preformulation parameters such as organoleptic characteristics, stability, permeability, inherent bioadhesion and retention features, and compatibility with excipients and condoms of candidate drugs, among others, play a crucial role when developing drug delivery systems containing microbicides [246]. After collection of this information, formulation studies are necessary in order to obtain a final product that fulfills microbicide objectives and requirements, namely safety, efficacy, acceptability, affordability, and regulatory duties [247]. As well as providing effective protection, microbicide formulations must also be safe on multiple exposures over time, chemically and physically stable, compatible with latex and other materials used in barrier devices, and affordable and acceptable to the end user. Ideally, they should be colorless, odorless, tasteless, and nonmessy [234, 248]. Tested microbicides have been formulated mostly as gels, although creams, vaginal rings, foams, sponges, vaginal suppositories, and films may also be considered. Nonetheless, alternative innovative options for the delivery of microbicides have been developed in recent years. For instance, the formulation of a safe and inexpensive "universal" drug carrier for microbicidal substances is an interesting strategy that may potentially ensure the efficacy of most currently researched molecules [249]. Another interesting approach for the delivery of microbicides was proposed by Chang et al. These researchers studied the possibility of using genetically modified comensal vaginal bacteria (*Lactobacillus jensenii*) to produce anti-HIV proteins [250]. In vitro experiments showed that this strategy can be a new step toward an effective microbicide formulation.

### 6.2 Antimicrobials

Vaginitis is a common condition in women which can be caused by bacteria, yeasts, or protozoa. Treatment of vaginitis has been achieved by oral or vaginal administration of antimicrobials, often with similar efficacy rates [251]. Several drugs are currently available for intravaginal treatment of bacterial (e.g., metronidazole, clindamycin), fungal (e.g., azoles, boric acid, nystatin), and protozoal (e.g., metronidazole) vaginitis [252]. Also, alternative vaginal therapies have been investigated in order to treat vaginitis. For instance, herbal formulations, particularly those containing essential oils, have been referred as potential antimicrobials for the treatment of both fungal and bacterial infections [253, 254]. Several vaginal drug delivery systems have been used in order to administer antimicrobial substances, particularly gels, creams, tablets, and vaginal suppositories. Current research in antimicrobial vaginal drug delivery systems is focused on more convenient single-dose

formulations that can achieve clinical cure while improving patient compliance [255–257].

Vaginal treatment may be advantageous when compared with oral treatment, as systemic adverse effects are less likely to occur [258]. For example, Cunningham and co-workers conducted a study where systemic levels of metronidazole after vaginal administration (5 g of a 0.75% metronidazole gel) were residual when compared to the oral administration of a 500-mg standard dose despite comparable clinical outcome [259]. This low level of vaginal absorption may be attributed to metronidazole's poor lipid solubility. Similar results have been obtained for clindamycin when comparing intravaginal and intravenous routes [260]. Also, contraindication during pregnancy and possible interference with oral contraceptives are situations that probably recommend vaginal administration of these therapeutic agents over oral administration in women requiring treatment [261].

As previously discussed, vaginal pH plays an important role in the normal physiology of the vagina, being elevated in bacterial vaginitis. Vaginal acidifiers, such as vitamin C in the form of tablets, can be effective in the treatment of vaginitis, with the advantage of maintaining or even improving the normal vaginal flora equilibrium, particularly in diabetic and pregnant women or those with recurrent bacterial vaginitis episodes [78, 262]. Also, other formulations such as acid-buffering gels were demonstrated to be potential helpers in the maintenance of a healthy vaginal milieu and as adjuvant of antimicrobials in the treatment of bacterial vaginosis [263, 264]. In fact, some gels may even be used in the treatment of bacterial vaginosis. For example, a mucoadhesive gel containing two polymers, polycarbophil and Carbopol 974P, was demonstrated to be effective in the treatment of bacterial vaginosis, even when compared with the clinical cure rate of vaginal metronidazole or tinidazole [265].

### 6.3 Hormonal Contraceptives and Hormonal Replacement Therapy

The vaginal route is one of the many available routes for estrogen and progestogen delivery. When compared to the oral and transdermal route, vaginal administration of these hormones presents the advantage of needing a smaller quantity of drug to achieve similar systemic effects with less relative variability [266, 267]. Also, the inconveniences associated with subcutaneous implants are circumvented. Thus, vaginal hormonal contraception and replacement therapy were soon considered as a possibility.

Hormonal contraception requires drug delivery systems that are able to achieve sustained blood levels of these substances, either by multiple administrations, as in the case of oral contraceptives, or by sustained drug release. Since daily vaginal administration would not be feasible and acceptable by most women, sustained-release systems were developed. This need led to the development of several devices, including vaginal rings [109]. Progestogen-only rings were initially developed, but menstrual bleeding problems and inefficacious control of ovulation led to discontinuation in several studies. Thus, combined estrogen–progestogen rings were the next natural step toward effective and acceptable formulations. Various combined estrogen–progestogen rings have proven to be highly effective as contraceptives, providing excellent inhibition of ovulation [41]. Also, serious vaginal lesions were shown to be unlikely to occur with short- and medium-term exposure

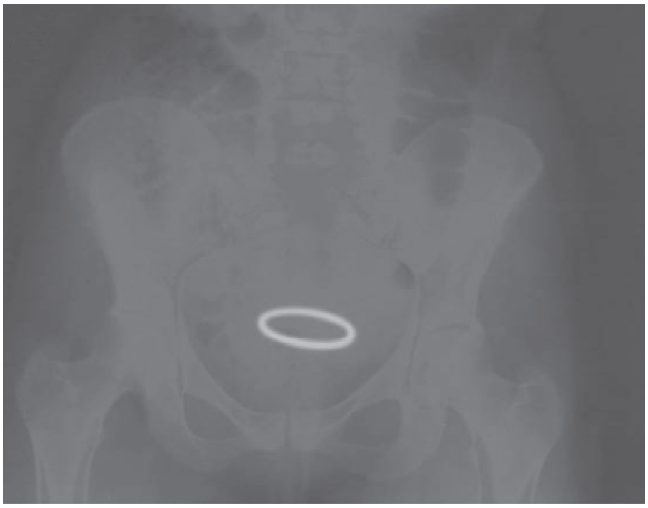

**FIGURE 7** X-ray image of vaginal ring after placement in human vagina. (Reprinted with permission from K. Malcolm et al., *Journal of Controlled Release*, 90, 217–225, 2003. Copyright 2003 by Elsevier.)

to currently available rings, mainly due to their improved flexibility and small dimensions [268]. Although other alternatives have also been studied, the most popular schedule for contraceptive vaginal rings is three weeks in, one week out. With this regimen the ring is inserted in the vagina on day 5 of the menstrual cycle and left in place for three weeks. On week 4 the ring is removed, allowing menstrual bleeding. A new ring (one-month ring) or the same one (over-one-month ring) is inserted after one week ring free [112]. Vaginal rings are easy to use, being self-administered by women. Once inserted in the vagina, the ring fits in the upper vagina and delivers the active substance(s) by contact with the vaginal mucosa (see Figure 7).

Despite all the investigational work already performed, only one contraceptive vaginal ring, Nuvaring (Organon), is available in the United States and Europe. This three weeks in–one week out vaginal ring was the first approved by the FDA, being available in the market since 2002. It is a flexible and transparent ring containing etonogestrel and ethinyl estradiol in a poly(ethylene-*co*-vinylacetate) matrix, with an outer diameter of 54 mm and a cross section of 4 mm [269]. When placed in the vagina it releases 120 μg/day of etonogestrel and 15 μg/day of ethinyl estradiol. Clinical studies showed that Nuvaring is as effective and reliable as commonly used oral contraceptives, being well tolerated, convenient, and highly acceptable to most women and their partners [270–272]. Moreover, this contraceptive ring is able to provide lower and more stable systemic exposure to estrogens than other contraceptive options, namely combined oral contraceptives or transdermal patches, thereby reducing drug-related side effects (Figure 8) [267]. Other contraceptive vaginal rings that are still in premarket research include a one-year contraceptive ring being developed by the Population Council, New York. This nestorone and ethinyl estradiol containing ring has been shown to be as efficacious as oral contraceptives when used up to 12 months on a 3 weeks in–1 week out regimen [273, 274].

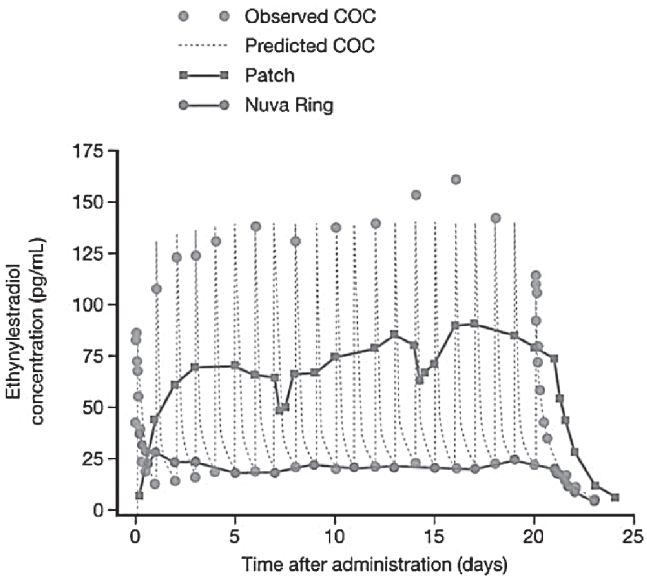


**FIGURE 8** Mean serum ethinyl estradiol concentration versus time for subjects treated during 21 days with Nuvaring ($n = 8$), transdermal contraceptive patch (Evra, OrthoMcNeil Pharmaceutical; releases 20 μg ethinyl estradiol and 150 μg norelgestromin daily; $n = 6$), and combined oral contraceptive (COC) (Microgynon, Schering AG; contains 30 μg ethinyl estradiol and 150 μg levonorgestrel; $n = 8$). Subject exposure to ethinyl estradiol in Nuvaring group was on average 3.4 and 2 times lower than for subjects in patch and combined oral contraceptive groups, respectively. (Reprinted with permission from M. W. van den Heuvel et al., *Contraception*, 72, 168–174, 2005. Copyright 2005 by Elsevier.)

Although conventional emergency contraception usually comprises the use of orally administered levonorgestrel (either alone or in combination with ethinyl estradiol), the vaginal route has been shown to be an efficacious alternative. The obtained hormonal plasma levels for the same oral dose are lower, but presumably high enough to prevent pregnancy [275]. Nonetheless, increases in the administered dose by the vaginal route have been suggested [276, 277].

Hormone replacement therapy has been a common practice for a long time in order to improve the quality of life of women suffering from acute symptoms related to menopause. Several formulations for oral, buccal, subcutaneous, parenteral, intrauterine, nasal, transdermal, or vaginal administration have been used for this purpose [278]. Although in recent years hormone replacement therapy has been associated with an increased risk of fatal breast cancer [279], some postmenopausal women may still benefit from this treatment. Estrogens have been used through the vaginal route for the treatment of vaginal symptoms associated with hormone decline in menopausal woman, such as dryness, dyspareunia, pruritus, irritation, discomfort, and atrophy and has been shown to be as effective as systemic therapy [280, 281]. Creams and vaginal suppositories were the first vaginal drug delivery systems to be used to deliver estrogens. Recently, estradiol rings for urogenital symptoms therapy (low-dose rings) or vasomotor plus vaginal symptoms relief (usually higher dose rings) were approved in some countries, providing a new option for the

administration of these drugs [282]. For instance, Estring (Pharmacia & Upjohn), a currently marketed vaginal ring containing low-dose estradiol (releases 7.5 μg/day when placed in the vagina), proved to be efficacious, well tolerated, and safe when used for up to a year in the treatment of urogenital symptoms in postmenopausal women [283, 284]. This silicone polymer ring has a diameter of 55 mm and a cross-sectional diameter of 9 mm and is used up to three months. Another similar ring containing estradiol acetate (Femring, Warner Chilcott), with a diameter of 56 mm and a cross-sectional diameter of 7.6 mm, is also currently available.

Estrogens and progestogens have also been administered through the vagina in order to manage other conditions. In fact, vaginally administered progesterone is commonly used for luteal-phase support in women undergoing assisted reproduction treatment, allowing optimal uterine concentrations without the high serum levels observed by other routes (oral and intramuscular), possibly due to the first-uterine-pass effect [285]. Several drug delivery systems such as capsules, tablets, vaginal suppositories, or gels have been used and have been shown to be equally efficacious in increasing the chance of becoming pregnant. Nonetheless, sustained-release formulations allow fewer administrations per day with lower doses [286]. Also, progesterone vaginal rings have been successfully used for luteal-phase support [287].

### 6.4 Spermicides

The vaginal use of spermicidal substances during sexual intercourse is perhaps the oldest method of contraception. However, the introduction of oral contraceptives and the intrauterine device in the 1960s led to the decay in their use. Since many of these substances also offer protection against sexually transmitted diseases, interest and investigation in this field have recently increased [288]. Also, the development of new potential spermicides, namely antibodies [289], contributed to further awareness. Currently used spermicides include nonoxynol-9, octoxynol, benzalkonium chloride, and chlorhexidine.

Nonoxynol-9 has been used for more than 30 years as a spermicide in over-the-counter vaginal products, such as semisolid formulations, sponges, foams, films, and others, in order to prevent pregnancy. Although relatively safe and effective, nonoxynol-9 formulations are still not able to achieve the same decrease in pregnancy risk obtained with hormonal methods [290]. Thus, some strategies have been used to enhance nonoxynol-9 effects while reducing its toxic effects. For instance, some synergistic associations with chelating agents that have themselves little spermicide activity, such as EDTA and ethylene glycol tetraacetic acid (EGTA), have been shown to be promising in this matter [291]. Also, coprecipitation of nonoxynol-9 with polyvinylpyrrolidone by a freeze-drying method can be useful, particularly when the formulation of a solid system is desirable, as this process is necessary to alter the chemical state (liquid to solid) of nonoxynol-9 [292, 293]. Using these coprecipitates, a tablet with an inner core that provides slow release of nonoxynol-9 after its fast release of the outer core was shown to be an efficient and safer way of delivering this spermicide in rabbits [294].

Semisolid formulations are often used as contraceptives, particularly gels. These drug delivery systems were demonstrated to be useful, namely in reducing the toxicity of nonoxynol-9 [295]. Adequate formulation of these products has also been

shown to be essential for their efficacy. For example, the contraceptive effect of spermicides can be enhanced by adequate consistency of the formulation. When the viscosity of a formulation increases, contraceptive efficacy may increase as a result of becoming more tenacious and more resistant to sperm migration, consequently decreasing its capacity of reaching the site of fertilization [296]. Other gel properties, such as pH and osmolarity, may also influence its spermicidal effects [297]. Moreover, vaginal rings and inserts have been proposed as adequate vehicles for the delivery of nonhormonal contraceptives [298, 299]. These systems may be advantageous due to their controlled and prolonged release properties.

### 6.5 Labor Inducers and Abortifacients

Cervical ripening for the induction of labor has been a common practice in modern obstetrics. Several drugs have been tested with this purpose, misoprostol, dinoprostone, and oxytocin the three most frequently used. These labor inducers are normally administered by the vaginal route as gels, tablets, suppositories, or inserts, all demonstrated to be effective and safe. Nonetheless, drug formulation and choice of drug delivery are important factors to be considered. In fact, different times to achieve cervical ripening have been observed in several studies due to different drug release profiles [300, 301].

Vaginal administration of prostaglandins can be useful in the termination of pregnancy. Misoprostol has been used to terminate unwanted pregnancies, demonstrating a relatively high efficacy, even when different regimens have been tested [302–304]. Indeed, improved results with vaginal administered misoprostol may be expected, because of higher and prolonged serum drug concentrations obtained, when compared with the oral route [305]. It is also noteworthy that since vaginal formulations are usually not available, oral tablets containing misoprostol (Cytotec, Pfizer) have been routinely used to administer this drug by the vaginal route. As these tablets are not specifically designed to be administered in the vagina, suboptimal clinical results may occur.

### 6.6 Proteins and Peptides

A few proteins and peptides, such as insulin, leuprolide, and salmon calcitonin, have been tested for their possible administration by the intravaginal route. In spite of the rich blood supply, relatively large surface area, and good permeability, vaginal absorption of peptides and proteins is influenced by hormonal-induced changes in the mucosa histology and enzymatic activity, thus limiting their administration through this route [306]. Nonetheless, rational design of adequate and innovative drug delivery carriers allowed considerable progress in protein and peptide vaginal delivery. For example, microspheres have shown good potential to deliver peptides and proteins, such as calcitonin, being able to increase drug stability and absorption. This enhancement of absorption is thought to be related with the intimate contact between the microspheres and the mucosa, resulting in high local concentrations at the site of absorption [307, 308]. Another tested approach to vaginal administration of proteins and peptides has been polymeric matrices, identical to the ones used in the design of vaginal rings. For instance, antibodies have been successfully administered in mice using poly(ethylene-*co*-vinyl acetate) disks [309]. Also, these and other

polymeric matrices can provide long-term (up to several years), controlled, and high-dose topical delivery of antibodies [310, 311].

Although considered a route that offers little or no real opportunities for insulin administration, mainly because of low and variable levels of absorption, efforts have been made to systemically deliver insulin through the vaginal mucosa. In the early 1980s, studies by Morimoto et al. showed that polyacrylic acid gels containing insulin were able to induce and maintain (up to 30 min) hypoglycemia when administered to rats and rabbits. However, sustained release was necessary to achieve longer time of hypoglycemia [312]. Later, Richardson et al. observed that insulin is almost not absorbed by the vaginal mucosa of ovariectomized rats in the absence of permeation enhancers, but the coadministration of substances such as sodium taurodihydrofusidate, polyoxyethylene-9-laurylether, lysophosphatidyl choline, palmitoylcarnitine chloride, and lysophosphatidyl glycerol significantly increased its absorption and consequently hypoglycemia [59]. Also, the use of mucoadhesive microspheres as delivery systems for insulin improved the absorption rate of this drug in sheep, particularly when associated with permeation enhancers [55]. Recently, Degim et al. developed vaginal chitosan gels as carriers for insulin. Studies performed in rabbits showed that chitosan gels containing 5% dimethyl-β-cyclodextrin as a penetration enhancer may provide longer insulin release, offering a potential alternative to the parenteral route [313]. Also, Ning et al. investigated the suitability of niosomes as insulin carriers for vaginal administration in rats. Results demonstrated that the bioavailability of insulin when administered through the vaginal route was comparable to that of the subcutaneous route [158].

### 6.7 Vaccines

Although once considered not to be a very promising approach, intravaginal vaccines have emerged in recent years as a potential noninvasive immunization strategy, particularly for the prevention of HIV transmission [314]. In fact, some animal and human experiments suggest that the obtained female genital tract immunization can be superior with vaginal administration of vaccines when compared to other routes, such as oral, nasal, or rectal [315–317]. Also, it is known that effective immunization against sexually transmitted diseases will require strong local genital tract as well as strong systemic antibody responses [318]. Mucosal vaccination presents several advantages over systemic immunization: improved safety profile, minimization of adverse effects, ease of administration, and potentially lower costs. However, limitations such as epithelium changes with menstrual cycle, which leads to reduced mucosa permeability and poor antigen presentation during certain stages, and inactivation of antigens by exposure to the vaginal environment can modulate the magnitude of immune response inducement [319].

Phosphate-buffered saline has been conveniently used as a vehicle for the delivery of antigens. Although these liquid formulations facilitate vaginal administration, they allow poor retention, controlled release, and protection of the delivered antigens. Thus, development of adequate vaginal drug delivery systems plays an important role in the success of vaginal immunization. Thermosensitive delivery systems with adequate mucoadhesiveness may be useful in enhancing the mucosal and systemic immune responses, as they can increase the exposure and contact of antigens to the vaginal mucosa [320]. The combination of thermosensitive polymers such as

poloxamers and mucoadhesive polymers such as polycarbophil were shown to be a good strategy in the development of mucoadhesive, thermosensitive, and controlled-release formulations for vaginal delivery of vaccines [321]. Also, antigen susceptibility to the vaginal environment can be circumvented by administering these molecules in protective systems, such as the one described by Shen et al., which comprises poly(ethylene-*co*-vinyl acetate) matrices containing plasmid DNA [322]. These disk-shaped devices, produced by a solvent evaporation technique described by Luo et al. [323], were shown to be effective in inducing local immunity in mice, protecting and providing controlled and sustained delivery of plasmid DNA. In addition, other strategies for vaginal antigen administration have also been proposed, such as the use of nonpathogenic bacterial vectors [324–326] and microspheres [327].

### 6.8 Other Uses

Many other drugs have been administered in the vagina for the management of either local or systemic conditions. Table 6 summarizes some of these reports.

## 7 ACCEPTABILITY AND PREFERENCES OF WOMEN RELATED TO VAGINAL DRUG DELIVERY

Usage of vaginal products is intimately related to taboos and presumptions associated with the knowledge and handling of genitalia. Results from a recent international survey showed that, despite improvements in recent decades, society's attitude toward the vagina and its use as a drug delivery route is not very open and not as open as women would like it to be. Also, this study revealed that the vagina is not commonly recognized as a possible drug delivery route, with only approximately 35% of women acknowledging this fact [342]. Although frequently overlooked, acceptability studies of vaginal products are important in predicting women's compliance and thus their effectiveness. Dedicated studies to women's acceptability and preferences toward vaginal drug delivery are not common, most information available resulting from parallel research to clinical trials. Also, most of these studies have focused on microbicides and spermicides, since their acceptance is decisive for their consistent use [343–345]. In fact, there can be a sense of guilt or negative feeling with this kind of vaginal product, contrasting with others that are prescribed by a physician for a specific gynecological problem, which seem to be deprived of moral issues due to the "legitimacy" of their use [346]. Acceptability studies should also be extended to sexual partners, particularly when evaluating products meant to be used during sexual intercourse [347]. Data collected from previous acceptability studies must be considered during the development of formulations in order to improve their suitability. For example, lubrication provided by a vaginal product seems to be important in determining its acceptability, as this feature may be regarded advantageous or not by different users [348]. Also, negative perceptions regarding product characteristics should be identified during clinical trials in order to improve formulations.

Although results found in the literature related to women's preferences toward the vaginal route of drug administration may vary significantly, particularly when different locations and cultures are considered, some general statements can be

**TABLE 6 Selected Drugs Administered in the Vagina**

| Drugs | Intended Use | Drug Delivery Systems | Comments | References |
|---|---|---|---|---|
| 5-Fluorouracil | Treatment of intravaginal warts | Gel | Demonstrated to be effective, safe, and tolerable | 328 |
| | Maintenance therapy of cervical dysplasias after standard excisional or ablative therapy | Cream | Reduction of recurrence was achieved | 329 |
| Bromocriptine | Therapy of hyperprolactinemia | Oral tablet and vaginal suppository | Proved to be effective and safe, without the adverse effects of oral administration; vaginal suppository obtained higher reduction in serum prolactin | 330 |
| Cabergoline | Therapy of hyperprolactinemia | Oral tablet | Proved to be effective and safe, without adverse effects of oral administration | 331 |
| Chlorhexidine | Prevention of peripartum infections of newborn | Aqueous solution | Proved to be useful | 332 |
| Danazol | Treatment of pelvic endometriosis | Ring and vaginal suppository | Effective without increased serum concentrations observed during oral therapy | 114 |
| Etoposide | Management of cervical dysplasias associated with human papilloma virus (HPV) | Vaginal suppository | Demonstrated to be safe and tolerable | 333 |
| Imiquimod | Treatment of high-grade vaginal intraepithelial neoplasia | Cream | Results suggest it can be an alternative conservative therapy | 334 |
| Indomethacin | Tocolysis in preterm labor | Rectal suppository | Vaginal administration proved to be more effective than conventional rectal plus oral administration | 335 |
| Lignocaine | Cervical anesthesia during insertion of tenaculum | Spray and gel | No difference in pain management between spray and gel | 336 |
| Morphine | Pain control as alternative to parenteral administration | Tablet and vaginal suppository | Requires close monitoring due to unpredictable bioavailability | 337 |
| Oxybutinin | Treatment of urge urinary incontinence | Insert | Proved to be effective and safe in rabbits | 338 |
| Propranolol | Control of tachycardia | Tablet | Obtained serum levels were within β-blocking range and comparable to those achieved by oral route | 339 |
| Sildenafil | Endometrial development for embryo implantation | Vaginal suppository | Enhanced endometrial development was achieved | 340 |
| Trichloroacetic acid | Treatment of low-grade vaginal intraepithelial neoplasia | Solution | Proved to be effective as well as inexpensive and easy to perform | 341 |

made. Ideally, formulations should be easy and comfortable to use, colorless, odorless, and messiness free. Also, products that are better retained in the vagina seem to be favored by women, since leakage is one of the most undesired feature of vaginal formulations [349]. Concerning preferred vaginal dosage forms, gels and creams seem to be the most popular among women. On the other hand, vaginal suppositories and tablets are among the most disliked dosage forms. Others, such as films, present ambiguous results, mostly related with difficulties during insertion [350, 351]. Vaginal rings have been shown to be highly acceptable by both women and their sexual partners, even during sexual intercourse. A trial conducted in North America and Europe showed that couples rarely felt the device during penile penetration, and when the ring was noticed, almost none of the partners seemed to mind [352]. Furthermore, insertion and removal of vaginal rings are judged to be easy by users [353].

Packaging and applicators may also influence women's choice. Products that are placed in the vagina by means of an applicator seem to be preferred because it facilitates administration and avoids direct touching of genitalia during insertion. Applicator characteristics, namely length, width, color, filling features (single-use or reusable applicators), and ease of usage, are also known to influence women's acceptability [351, 354].

## 8 VETERINARY VAGINAL DRUG DELIVERY

As with humans, veterinary vaginal drug administration has been performed for a long time, particularly for the treatment of local infections, traditionally involving the use of vaginal suppositories, liquid formulations, or gels. Advantages of this route include the avoidance of damage to the skin or to tissue that is associated with injections, minor stress inflicted to the animals, and possibility of ceasing drug delivery at will [355]. Nonetheless, the first major studies on veterinary drug delivery have been performed by Robinson in the 1960s with progestogen-impregnated polyurethane sponges [356]. Since then, the major use of the vaginal route has been the control of the estrus cycle in livestock by delivering progestogens and estrogens in a controlled fashion. Also, the administration of these hormones showed good results in treating reproductive disorders, such as ovarian quiescence, cystic ovary or cystic corpus luteum [357]. Estrus synchronicity is advantageous as it allows insemination of all or selected females in a herd or flock to occur during a single period of several hours or days [358]. With this purpose, several drug delivery systems have been developed, being generally based on polymeric matrices that are able to control release of drug content. These devices also present two common features: Retention is guaranteed by means of a gentle pressure applied to the mucosa and the existence of a mechanism (e.g., an attached string) that allows their easy removal at the end of the treatment. Table 7 presents a synopsis of these systems, as reviewed by Rathbone et al. [358, 359]. Also, some of these devices are shown in Figure 9. In addition to the vaginal administration of progestogens and estrogens, delivery of other therapeutic agents through this route for both local and systemic effects has been investigated, namely 1,25-dihydroxy vitamin D3 [360], lactic acid–producing lactobacilli [361], antimicrobials [362], local anesthetics [363], and vaccines [364].

**TABLE 7 Selected Veterinary Vaginal Drug Delivery Systems for Control of Estrus Cycle in Livestock**

| Drug Delivery Systems | Active Substance(s) | Matrix-Forming Polymers | Brief Description | Comments |
|---|---|---|---|---|
| Sponges | Progesterone, estradiol, and several progestins | Polyurethane | Polymeric devices impregnated with active substance(s), usually cylindrical shaped; alternative designs allow achieving zero-order release profile | Variable vaginal retention; inexpensive and simple to prepare |
| PRID (Progesterone Releasing Intravaginal Device, InterAg, Hamilton, NZ) | Progesterone | Poly(dimethylsiloxane) | Spiral-shaped device obtained by molding and curing polymer (containing active substance) onto stainless steel spiral by high-temperature (~190°C) injection molding; also may enclose hard gelatin capsule (containing estradiol) glued to device | Excellent vaginal retention |
| CIDR (Controlled Internal Drug Release, CEVA, Libourne, France) | Progesterone | Poly(dimethylsiloxane) | T-shaped nylon spine over which the polymer (containing active substance) is molded and cured by high-temperature injection molding; several types developed: CIDR-S (rabbit eared in shape), CIDR-G (slimmer, straight T-shaped), CIDR-B (similar to CIDR-G but with increased dimensions) | Excellent vaginal retention; only types G and B currently available |
| INVAS (Intravaginal Application System) | Progesterone | Poly(dimethylsiloxane) | Similar to CIDR but obtained by lower temperature (<120°C) method | Not commercially available |
| Rings | — | Poly(dimethylsiloxane) | Similar to human vaginal rings | Not commercially available; abandoned because of poor retention properties |
| Rajamehendran intravaginal device | 17β-estradiol/ progesterone | Poly(dimethylsiloxane) | Two C-shaped polymer tubes (impregnated with active substances) tied together, forming "umbrella-shaped" device | Excellent vaginal retention; not commercially available |
| IBD (Intelligent Breeding Device, Advanced Animal Technology, Hamilton, NZ) | Progesterone/ estradiol/ prostaglandin | Poly(dimethylsiloxane) | Composed by outer plastic sheath [designed to protect circuit board (which controls drug release) and two batteries, four polymeric drug reservoirs (a large one at base and three small ones at head of device)], retention mechanism, and tail | Allows release of active substances at different rates and at specific times |

*Source*: From refs. 358 and 359.

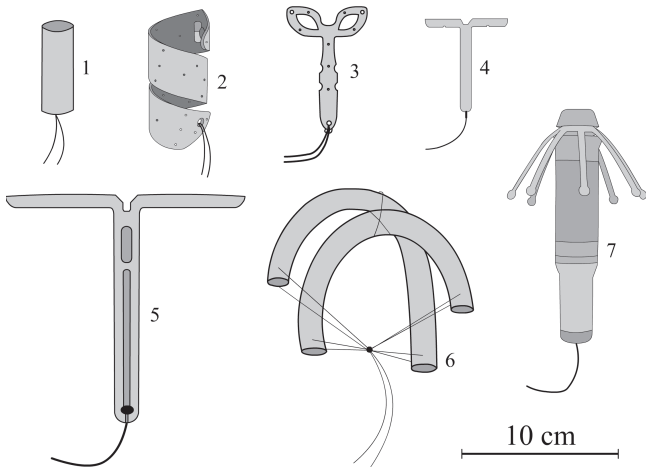


**FIGURE 9** Some vaginal veterinary drug delivery systems used or investigated to control estrus cycle in livestock: sponge (1), PRID (2), CIDR-S (3), CIDR-G (4), CIDR-B (5), Rajamehendran intravaginal device (6), and IBD (7).

In order to improve the performance of drug delivery systems, several strategies have been tested. For example, improvement of currently available progestogen- and estrogen-delivering devices and development of new ones have been a common gold of vaginal drug delivery investigators [365]. In other therapeutic field, Gavini et al. used bioadhesive chitosan microspheres compressed into tablets with several excipients to deliver acriflavine [362]. In vitro results demonstrated that these systems have good mucoadhesive properties, allowing increased residence time in the vagina. Recently, Cross et al. proposed an interesting electronic device that allows controlling drug release in response to in loco stimuli (e.g., vaginal temperature) or external commanding via radio wireless link [366]. This device is placed inside and behind the piston of a modified syringe, being able to control the production of hydrogen from a gas cell. The increased gas pressure behind the piston propels a viscous pharmaceutical vehicle that fills the syringe. This drug release and monitoring unit (DMU) have been shown to be a promising strategy in controlling the delivery of active substances to cows while simultaneously collecting physiological data in the vaginal environment.

## 9 CONCLUSIONS AND FUTURE TRENDS

As other routes of drug administration presented serious difficulties to deliver some active substances, the vagina emerged as a feasible alternative. Undoubtedly, the vaginal route of drug delivery has attracted the interest of the scientific community, particularly in the last few decades. The latest developments in this field, namely mucoadhesive formulations, vaginal rings, and other controlled-release drug delivery systems, have boosted research and clinical use of this once-neglected route of drug administration. Also, women's emancipation in the last century has slowly led to higher acceptability of vaginal formulations, as old fears and preconceived ideas are being demystified.

Much work remains to be done, particularly in specific fields, such as the delivery of macromolecular drugs (e.g., proteins and peptides) and other substances that are poorly absorbed through the vaginal mucosa. Another promising area that needs further investigation is vaginal administration of vaccines and microbicides. Indeed, in an era where HIV and other sexually transmitted diseases are an increasing concern, vaginal preventive strategies are required. Formulating scientists can contribute decisively to these objectives, as optimization of drug delivery systems seems to be essential. Issues such as poor vaginal distribution and retention, inadequate drug release, limited drug protection from vaginal "aggressors," and adverse effects of currently available drug delivery systems still need to be solved.

## ACKNOWLEDGMENTS

The authors would like to express their gratitude to Bruno Sarmento and Cláudia Carneiro for their kind review and useful comments on the manuscript.

## REFERENCES

1. O'Dowd, M. J. (2001), *The History of Medications for Women: Materia Medica Woman*, 1st ed., Taylor & Francis, London, pp 53–58.
2. O'Dowd, M. J., and Philipp, E. E. (2000), *The History of Obstetrics and Gynecology*, 1st ed., Taylor & Francis, London, pp 1–40.
3. Stenchever, M. A., Droegemueller, W., and Herbst, A. L. (2002), *Comprehensive Gynecology*, 4th ed., Mosby, St. Louis, MO, pp 40–44.
4. Van De Graff, K. (2001), *Human Anatomy*, 6th ed., McGraw-Hill, London, pp 725–753.
5. Barnhart, K. T., Izquierdo, A., Pretorius, E. S., Shera, D. M., Shabbout, M., and Shaunik, A. (2006), Baseline dimensions of the human vagina, *Hum. Reprod.*, 21, 1618–1622.
6. Berman, J. R., and Bassuk, J. (2002), Physiology and pathophysiology of female sexual function and dysfunction, *World J. Urol.*, 20, 111–118.
7. Pendergrass, P. B., Belovicz, M. W., and Reeves, C. A. (2003), Surface area of the human vagina as measured from vinyl polysiloxane casts, *Gynecol. Obstet. Invest.*, 55, 110–113.
8. Barnhart, K. T., Pretorius, E. S., and Malamud, D. (2004), Lesson learned and dispelled myths: Three-dimensional imaging of the human vagina, *Fertil. Steril.*, 81, 1383–1384.
9. Fawcett, D. W., and Raviola, E. (1994), *Bloom and Fawcett, A Textbook of Histology*, 12th ed., Chapman & Hall, New York, pp 857–858.
10. Patton, D. L., Thwin, S. S., Meier, A., Hooton, T. M., Stapleton, A. E., and Eschenbach, D. A. (2000), Epithelial cell layer thickness and immune cell populations in the normal human vagina at different stages of the menstrual cycle, *Am. J. Obstet. Gynecol.*, 183, 967–973.
11. Ildgruben, A. K., Sjoberg, I. M., and Hammarstrom, M. L. (2003), Influence of hormonal contraceptives on the immune cells and thickness of human vaginal epithelium, *Obstet. Gynecol.*, 102, 571–582.
12. Mauck, C. K., Callahan, M. M., Baker, J., Arbogast, K., Veazey, R., Stock, R., Pan, Z., Morrison, C. S., Chen-Mok, M., Archer, D. F., and Gabelnick, H. L. (1999), The effect of one injection of Depo-Provera on the human vaginal epithelium and cervical ectopy, *Contraception*, 60, 15–24.

13. Wagner, G., and Ottesen, B. (1982), Vaginal physiology during menstruation, *Ann. Intern. Med.*, 96, 921–923.
14. Caillouette, J. C., Sharp, C. F., Jr., Zimmerman, G. J., and Roy, S. (1997), Vaginal pH as a marker for bacterial pathogens and menopausal status, *Am. J. Obstet. Gynecol.*, 176, 1270–1275.
15. Ramsey, P. S., Ogburn, P. L., Jr., Harris, D. Y., Heise, R. H., and Ramin, K. D. (2002), Effect of vaginal pH on efficacy of the dinoprostone gel for cervical ripening/labor induction, *Am. J. Obstet. Gynecol.*, 187, 843–846.
16. Boskey, E. R., Cone, R. A., Whaley, K. J., and Moench, T. R. (2001), Origins of vaginal acidity: High D/L lactate ratio is consistent with bacteria being the primary source, *Hum. Reprod.*, 16, 1809–1813.
17. Boris, S., and Barbés, C. (2000), Role played by lactobacilli in controlling the population of vaginal pathogens, *Microbes Infect.*, 2, 543–546.
18. Zhou, X., Bent, S. J., Schneider, M. G., Davis, C. C., Islam, M. R., and Forney, L. J. (2004), Characterization of vaginal microbial communities in adult healthy women using cultivation-independent methods, *Microbiology*, 150, 2565–2573.
19. Larsen, B., and Galask, R. P. (1982), Vaginal microbial flora: Composition and influences of host physiology, *Ann. Intern. Med.*, 96, 926–930.
20. Woolfson, A. D., Malcolm, R. K., and Gallagher, R. (2000), Drug delivery by the intravaginal route, *Crit. Rev. Ther. Drug Carrier Syst.*, 17, 509–555.
21. Owen, D. H., and Katz, D. F. (1999), A vaginal fluid simulant, *Contraception*, 59, 91–95.
22. Widmaier, E. P., Raff, H., and Strang, K. T. (2004), *Vander, Sherman, Luciano's Human Physiology: The Mechanisms of Body Function*, 9th ed., McGraw-Hill, New York, pp 658–688.
23. Terranova, P. F. (2003), The female reproductive system, in Rhoades, R. A., and Tanner, G. A., Eds. *Medical Physiology*, 2nd ed., Lippincott, Williams & Wilkins, Philadelphia, pp 667–683.
24. Valore, E. V., Park, C. H., Igreti, S. L., and Ganz, T. (2002), Antimicrobial components of vaginal fluid, *Am. J. Obstet. Gynecol.*, 187, 561–568.
25. Eggert-Kruse, W., Botz, I., Pohl, S., Rohr, G., and Strowitzki, T. (2000), Antimicrobial activity of human cervical mucus, *Hum. Reprod.*, 15, 778–784.
26. Burruano, B. T., Schnaare, R. L., and Malamud, D. (2002), Synthetic cervical mucus formulation, *Contraception*, 66, 137–140.
27. Acarturk, F., Parlatan, Z. I., and Saracoglu, O. F. (2001), Comparison of vaginal aminopeptidase enzymatic activities in various animals and in humans, *J. Pharm. Pharmacol.*, 53, 1499–1504.
28. Gibbs, D. F., Labrum, A. H., and Stagg, B. H. (1968), Vaginal fluid enzymology. A new assay method with enzyme-potassium ratios, *Am. J. Obstet. Gynecol.*, 102, 982–988.
29. Belec, L. (2002), Défenses non immunes, pré-immunes et immunes du tractus génital féminin contre les infections, *J. Gynecol. Obstet. Biol. Reprod. (Paris)*, 31, 034302.1–034302.7.
30. Mestecky, J., and Fultz, P. N. (1999), Mucosal immune system of the human genital tract, *J. Infect. Dis.*, 179(Suppl. 3), S470–474.
31. Russell, M. W. (2002), Immunization for protection of the reproductive tract: A review, *Am. J. Reprod. Immunol.*, 47, 265–268.
32. Mestecky, J., Moldoveanu, Z., and Russell, M. W. (2005), Immunologic uniqueness of the genital tract: Challenge for vaccine development, *Am. J. Reprod. Immunol.*, 53, 208–214.

33. Hocini, H., Barra, A., Belec, L., Iscaki, S., Preud'homme, J. L., Pillot, J., and Bouvet, J. P. (1995), Systemic and secretory humoral immunity in the normal human vaginal tract, *Scand. J. Immunol.*, 42, 269–274.
34. Beagley, K. W., and Gockel, C. M. (2003), Regulation of innate and adaptive immunity by the female sex hormones oestradiol and progesterone, *FEMS Immunol. Med. Microbiol.*, 38, 13–22.
35. Wira, C. R., Rossoll, R. M., and Kaushic, C. (2000), Antigen-presenting cells in the female reproductive tract: Influence of estradiol on antigen presentation by vaginal cells, *Endocrinology*, 141, 2877–2885.
36. Franklin, R. D., and Kutteh, W. H. (1999), Characterization of immunoglobulins and cytokines in human cervical mucus: Influence of exogenous and endogenous hormones, *J. Reprod. Immunol.*, 42, 93–106.
37. Farage, M., and Maibach, H. (2006), Lifetime changes in the vulva and vagina, *Arch. Gynecol. Obstet.*, 273, 195–202.
38. Nilsson, K., Risberg, B., and Heimer, G. (1995), The vaginal epithelium in the postmenopause—Cytology, histology and pH as methods of assessment, *Maturitas*, 21, 51–56.
39. Vermani, K., and Garg, S. (2000), The scope and potential of vaginal drug delivery, *Pharm. Sci. Technol. Today*, 3, 359–364.
40. Bernkop-Schnurch, A., and Hornof, M. (2003), Intravaginal drug delivery systems, *Am. J. Drug Deliv.*, 1, 241–254.
41. Alexander, N. J., Baker, E., Kaptein, M., Karck, U., Miller, L., and Zampaglione, E. (2004), Why consider vaginal drug administration? *Fertil. Steril.*, 82, 1–12.
42. Hussain, A., and Ahsan, F. (2005), The vagina as a route for systemic drug delivery, *J. Controlled Release*, 103, 301–313.
43. Macht, D. I. (1918), On the absorption of drugs and poisons through the vagina, *J. Pharmacol. Exp. Ther.*, 10, 509–522.
44. Robinson, G. D. (1925), Absorption from the human vagina, *BJOG*, 32, 496–504.
45. Rock, J., Barker, R. H., and Bacon, W. B. (1947), Vaginal absorption of penicillin, *Science*, 105, 13.
46. van der Bijl, P., and van Eyk, A. D. (2003), Comparative in vitro permeability of human vaginal, small intestinal and colonic mucosa, *Int. J. Pharm.*, 261, 147–152.
47. van Eyk, A. D., and van der Bijl, P. (2004), Comparative permeability of various chemical markers through human vaginal and buccal mucosa as well as porcine buccal and mouth floor mucosa, *Arch. Oral Biol.*, 49, 387–392.
48. Villanueva, B., Casper, R. F., and Yen, S. S. (1981), Intravaginal administration of progesterone: Enhanced absorption after estrogen treatment, *Fertil. Steril.*, 35, 433–437.
49. Sassi, A. B., McCullough, K. D., Cost, M. R., Hillier, S. L., and Rohan, L. C. (2004), Permeability of tritiated water through human cervical and vaginal tissue, *J. Pharm. Sci.*, 93, 2009–2016.
50. van Eyk, A. D., and van der Bijl, P. (2005), Porcine vaginal mucosa as an in vitro permeability model for human vaginal mucosa, *Int. J. Pharm.*, 305, 105–111.
51. van der Bijl, P., and van Eyk, A. D. (2004), Human vaginal mucosa as a model of buccal mucosa for in vitro permeability studies: An overview, *Curr. Drug Deliv.*, 1, 129–135.
52. van der Bijl, P., Penkler, L., and van Eyk, A. D. (2000), Permeation of sumatriptan through human vaginal and buccal mucosa, *Headache*, 40, 137–141.
53. Corbo, D. C., Liu, J. C., and Chien, Y. W. (1989), Drug absorption through mucosal membranes: Effect of mucosal route and penetrant hydrophilicity, *Pharm. Res.*, 6, 848–852.

54. Corbo, D. C., Liu, J. C., and Chien, Y. W. (1990), Characterization of the barrier properties of mucosal membranes, *J. Pharm. Sci.*, 79, 202–206.
55. Richardson, J. L., Farraj, N. F., and Illum, L. (1992), Enhanced vaginal absorption of insulin in sheep using lysophosphatidylcholine and a bioadhesive microsphere delivery system, *Int. J. Pharm.*, 88, 319–325.
56. Okada, H., Yamazaki, I., Yashiki, T., and Mima, H. (1983), Vaginal absorption of a potent luteinizing hormone-releasing hormone analogue (leuprolide) in rats II: Mechanism of absorption enhancement with organic acids, *J. Pharm. Sci.*, 72, 75–78.
57. van der Bijl, P., van Eyk, A. D., Gareis, A. A., and Thompson, I. O. (2002), Enhancement of transmucosal permeation of cyclosporine by benzalkonium chloride, *Oral Dis.*, 8, 168–172.
58. Richardson, J. L., Thomas, N. W., and Illum, L. (1991), Recovery of the rat vaginal epithelium from the histological effects of absorptions enhancers, *Int. J. Pharm.*, 77, 75–78.
59. Richardson, J. L., Illum, L., and Thomas, N. W. (1992), Vaginal absorption of insulin in the rat: Effect of penetration enhancers on insulin uptake and mucosal histology, *Pharm. Res.*, 9, 878–883.
60. Sayani, A. P., Chun, I. K., and Chien, Y. W. (1993), Transmucosal delivery of leucine enkephalin: Stabilization in rabbit enzyme extracts and enhancement of permeation through mucosae, *J. Pharm. Sci.*, 82, 1179–1185.
61. Richardson, J. L., Minhas, P. S., Thomas, N. W., and Illum, L. (1989), Vaginal administration of gentamicin to rats. Pharmaceutical and morphological studies using absorption enhancers, *Int. J. Pharm.*, 56, 29–35.
62. Nakada, Y., Miyake, M., and Awata, N. (1993), Some factors affecting the vaginal absorption of human calcitonin in rats, *Int. J. Pharm.*, 89, 169–175.
63. van der Bijl, P., and van Eyk, A. D. (2001), Areca nut extract lowers the permeability of vaginal mucosa to reduced arecoline and arecaidine, *J. Oral Pathol. Med.*, 30, 537–541.
64. Cicinelli, E., Einer-Jensen, N., Galantino, P., Alfonso, R., and Nicoletti, R. (2004), The vascular cast of the human uterus: From anatomy to physiology, *Ann. N. Y. Acad. Sci.*, 1034, 19–26.
65. Cicinelli, E., Cignarelli, M., Sabatelli, S., Romano, F., Schonauer, L. M., Padovano, R., and Einer-Jensen, N. (1998), Plasma concentrations of progesterone are higher in the uterine artery than in the radial artery after vaginal administration of micronized progesterone in an oil-based solution to postmenopausal women, *Fertil. Steril.*, 69, 471–473.
66. Bulletti, C., de Ziegler, D., Flamigni, C., Giacomucci, E., Polli, V., Bolelli, G., and Franceschetti, F. (1997), Targeted drug delivery in gynaecology: The first uterine pass effect, *Hum. Reprod.*, 12, 1073–1079.
67. Cicinelli, E., Di Naro, E., De Ziegler, D., Matteo, M., Morgese, S., Galantino, P., Brioschi, P. A., and Schonauer, A. (2003), Placement of the vaginal 17beta-estradiol tablets in the inner or outer one third of the vagina affects the preferential delivery of 17beta-estradiol toward the uterus or periurethral areas, thereby modifying efficacy and endometrial safety, *Am. J. Obstet. Gynecol.*, 189, 55–58.
68. Thorgeirsdottir, T. O., Hilmarsson, H., Thormar, H., and Kristmundsdottir, T. (2005), Development of a virucidal cream containing the monoglyceride monocaprin, *Pharmazie*, 60, 897–899.
69. Thorgeirsdottir, T. O., Thormar, H., and Kristmundsdottir, T. (2005), The influence of formulation variables on stability and microbicidal activity of monoglyceride monocaprin, *J. Drug Deliv. Sci. Tech.*, 15, 233–236.

70. Kristmundsdottir, T., Sigurdsson, P., and Thormar, H. (2003), Effect of buffers on the properties of microbicidal hydrogels containing monoglyceride as the active ingredient, *Drug Dev. Ind. Pharm.*, 29, 121–129.
71. Garg, S., Tambwekar, K. R., Vermani, K., Garg, A., Kaul, C. L., and Zaneveld, J. D. (2001), Compendium of pharmaceutical excipients for vaginal formulations, *Pharm. Technol.*, 25, 14–24.
72. Knapczyk, J., Macura, A. B., and Pawlik, B. (1992), Simple test demonstrating the antimycotic effect of chitosan, *Int. J. Pharm.*, 80, 33–38.
73. Manson, K. H., Wyand, M. S., Miller, C., and Neurath, A. R. (2000), Effect of a cellulose acetate phthalate topical cream on vaginal transmission of simian immunodeficiency virus in rhesus monkeys, *Antimicrob. Agents Chemother.*, 44, 3199–3202.
74. Katz, D. F., and Dunmire, E. N. (1993), Cervical mucus: Problems and opportunities for drug delivery via the vagina and cervix, *Adv. Drug Deliv. Rev.*, 11, 385–401.
75. Willits, R. K., and Saltzman, W. M. (2001), Synthetic polymers alter the structure of cervical mucus, *Biomaterials*, 22, 445–452.
76. Karasulu, H. Y., Taneri, F., Sanal, E., Guneri, T., and Ertan, G. (2002), Sustained release bioadhesive effervescent ketoconazole microcapsules tabletted for vaginal delivery, *J. Microencapsul.*, 19, 357–362.
77. Knapczyk, J. (1992), Antimycotic buccal and vaginal tablets with chitosan, *Int. J. Pharm.*, 88, 9–14.
78. Petersen, E. E., and Magnani, P. (2004), Efficacy and safety of vitamin C vaginal tablets in the treatment of non-specific vaginitis. A randomised, double blind, placebo-controlled study, *Eur. J. Obstet. Gynecol. Reprod. Biol.*, 117, 70–75.
79. Digenis, G. A., Nosek, D., Mohammadi, F., Darwazeh, N. B., Anwar, H. S., and Zavos, P. M. (1999), Novel vaginal controlled-delivery systems incorporating coprecipitates of nonoxynol-9, *Pharm. Dev. Technol.*, 4, 421–430.
80. Notelovitz, M., Funk, S., Nanavati, N., and Mazzeo, M. (2002), Estradiol absorption from vaginal tablets in postmenopausal women, *Obstet. Gynecol.*, 99, 556–562.
81. Maggi, L., Mastromarino, P., Macchia, S., Brigidi, P., Pirovano, F., Matteuzzi, D., and Conte, U. (2000), Technological and biological evaluation of tablets containing different strains of lactobacilli for vaginal administration, *Eur. J. Pharm. Biopharm.*, 50, 389–395.
82. Bates, C. D., Nicoll, A. E., Mullen, A. B., Mackenzie, F., Thomson, A. J., and Norman, J. E. (2003), Serum profile of isosorbide mononitrate after vaginal administration in the third trimester, *BJOG*, 110, 64–67.
83. Prista, L. N., Alves, A. C., and Morgado, R. (1996), *Tecnologia Farmacêutica*, 4th ed., Fundação Calouste Gulbenkian, Lisboa, pp 1585–1592.
84. Ozyazici, M., Turgut, E. H., Taner, M. S., Koseoglu, K., and Ertan, G. (2003), In-vitro evaluation and vaginal absorption of metronidazole suppositories in rabbits, *J. Drug Target.*, 11, 177–185.
85. Bergren, M. S., Battle, M. M., Halstead, G. W., and Theis, D. L. (1989), Investigation of the relationship between melting-related parameters and in vitro drug release from vaginal suppositories, *J. Pharm. Biomed. Anal.*, 7, 549–561.
86. Kale, V. V., Trivedi, R. V., Wate, S. P., and Bhusari, K. P. (2005), Development and evaluation of a suppository formulation containing lactobacillus and its application in vaginal diseases, *Ann. N. Y. Acad. Sci.*, 1056, 359–365.
87. Mahaguna, V., McDermott, J. M., Zhang, F., and Ochoa, F. (2004), Investigation of product quality between extemporaneously compounded progesterone vaginal

suppositories and an approved progesterone vaginal gel, *Drug. Dev. Ind. Pharm.*, 30, 1069–1078.

88. Zaneveld, L. J., Waller, D. P., Ahmad, N., Quigg, J., Kaminski, J., Nikurs, A., and De Jonge, C. (2001), Properties of a new, long-lasting vaginal delivery system (LASRS) for contraceptive and antimicrobial agents, *J. Androl.*, 22, 481–490.
89. Ladipo, O. A., De Castro, M. P., Filho, L. C., Coutinho, E., Waller, D. P., Cone, F., and Zaneveld, L. J. (2000), A new vaginal antimicrobial contraceptive formulation: Phase I clinical pilot studies, *Contraception*, 62, 91–97.
90. Mandal, T. K. (2000), Swelling-controlled release system for the vaginal delivery of miconazole, *Eur. J Pharm. Biopharm.*, 50, 337–343.
91. Merabet, J., Thompson, D., and Saul Levinson, R. (2005), Advancing vaginal drug delivery, *Expert Opin. Drug. Deliv.*, 2, 769–777.
92. Seidman, L. S., and Skokos, C. K. (2005), An evaluation of butoconazole nitrate 2% site release vaginal cream (Gynazole-1) compared to fluconazole 150 mg tablets (Diflucan) in the time to relief of symptoms in patients with vulvovaginal candidiasis, *Infect. Dis. Obstet. Gynecol.*, 13, 197–206.
93. Levinson, R. S., Mitan, S. J., Steinmetz, J. I., Gattermeir, D. J., Schumacher, R. J., and Joffrion, J. L. (2005), An open-label, two-period, crossover study of the systemic bioavailability in healthy women of clindamycin phosphate from two vaginal cream formulations, *Clin. Ther.*, 27, 1894–1900.
94. del Palacio, A., Sanz, F., Sanchez-Alor, G., Garau, M., Calvo, M. T., Boncompte, E., Alguero, M., Pontes, C., and Gomez de la Camara, A. (2000), Double-blind randomized dose-finding study in acute vulvovaginal candidosis. Comparison of flutrimazole site-release cream (1, 2 and 4%) with placebo site-release vaginal cream, *Mycoses*, 43, 355–365.
95. Wichterle, O., and Lim, D. (1960), Hydrophilic gels in biologic use, *Nature*, 185, 117.
96. Justin-Temu, M., Damian, F., Kinget, R., and Van Den Mooter, G. (2004), Intravaginal gels as drug delivery systems, *J. Womens Health (Larchmt.)*, 13, 834–844.
97. das Neves, J., and Bahia, M. F. (2006), Gels as vaginal drug delivery systems, *Int. J. Pharm.*, 318, 1–14.
98. Milani, M., Molteni, B., and Silvani, I. (2000), Effect on vaginal pH of a polycarbophil vaginal gel compared with an acidic douche in women with suspected bacterial vaginosis: A randomized, controlled study, *Curr. Ther. Res. Clin. Exp.*, 61, 781–788.
99. Robinson, J. R., and Bologna, W. J. (1994), Vaginal and reproductive system treatments using a bioadhesive polymer, *J. Controlled Release*, 28, 87–94.
100. D'Cruz, O. J., and Uckun, F. M. (2001), Gel-microemulsions as vaginal spermicides and intravaginal drug delivery vehicles, *Contraception*, 64, 113–123.
101. Knuth, K., Amiji, M., and Robinson, J. R. (1993), Hydrogel delivery systems for vaginal and oral applications: Formulation and biological considerations, *Adv. Drug Deliv. Rev.*, 11, 137–167.
102. Venkatraman, S., Davar, N., Chester, A., and Kleiner, L. (2005), An overview of controlled release systems, in Wise, D. L., Ed., *Handbook of Pharmaceutical Controlled Release Technology*, Marcel Dekker, New York, pp 431–464.
103. Okada, H., Yamazaki, I., Ogawa, Y., Hirai, S., Yashiki, T., and Mima, H. (1982), Vaginal absorption of a potent luteinizing hormone-releasing hormone analog (leuprolide) in rats I: Absorption by various routes and absorption enhancement, *J. Pharm. Sci.*, 71, 1367–1371.
104. Rosenberg, M. J., Phillips, R. S., and Holmes, M. D. (1991), Vaginal douching. Who and why? *J. Reprod. Med.*, 36, 753–758.

105. McClelland, R. S., Lavreys, L., Hassan, W. M., MAndaliya, K., Ndinya-Achola, J. O., and Baeten, J. M. (2006), Vaginal washing and increased risk of HIV-1 acquisition among African women: A 10-year prospective study, *AIDS*, 20, 269–273.

106. Myer, L., Kuhn, L., Stein, Z. A., Wright, T. C., Jr., and Denny, L. (2005), Intravaginal practices, bacterial vaginosis, and women's susceptibility to HIV infection: Epidemiological evidence and biological mechanisms, *Lancet Infect. Dis.*, 5, 786–794.

107. Rajamanoharan, S., Low, N., Jones, S. B., and Pozniak, A. L. (1999), Bacterial vaginosis, ethnicity, and the use of genital cleaning agents: A case control study, *Sex. Transm. Dis.*, 26, 404–409.

108. Zhang, J., Thomas, A. G., and Leybovich, E. (1997), Vaginal douching and adverse health effects: A meta-analysis, *Am. J. Public Health*, 87, 1207–1211.

109. Sam, A. P. (1992), Controlled release contraceptive devices: A status report, *J. Controlled Release*, 22, 35–46.

110. Johansson, E. D., and Sitruk-Ware, R. (2004), New delivery systems in contraception: Vaginal rings, *Am. J. Obstet. Gynecol.*, 190, S54–59.

111. Timmer, C. J., and Mulders, T. M. (2000), Pharmacokinetics of etonogestrel and ethinylestradiol released from a combined contraceptive vaginal ring, *Clin. Pharmacokinet.*, 39, 233–242.

112. Sitruk-Ware, R. (2005), Vaginal delivery of contraceptives, *Expert Opin. Drug Deliv.*, 2, 729–736.

113. Acarturk, F., and Altug, N. (2001), In-vitro and in-vivo evaluation of a matrix-controlled bromocriptine mesilate-releasing vaginal ring, *J. Pharm. Pharmacol.*, 53, 1721–1726.

114. Igarashi, M., Iizuka, M., Abe, Y., and Ibuki, Y. (1998), Novel vaginal danazol ring therapy for pelvic endometriosis, in particular deeply infiltrating endometriosis, *Hum. Reprod.*, 13, 1952–1956.

115. Woolfson, A. D., Malcolm, R. K., and Gallagher, R. J. (2003), Design of a silicone reservoir intravaginal ring for the delivery of oxybutynin, *J. Controlled Release*, 91, 465–476.

116. Wyatt, T. L., Whaley, K. J., Cone, R. A., and Saltzman, W. M. (1998), Antigen-releasing polymer rings and microspheres stimulate mucosal immunity in the vagina, *J. Controlled Release*, 50, 93–102.

117. Malcolm, R. K., Woolfson, A. D., Toner, C. F., Morrow, R. J., and McCullagh, S. D. (2005), Long-term, controlled release of the HIV microbicide TMC120 from silicone elastomer vaginal rings, *J. Antimicrob. Chemother.*, 56, 954–956.

118. Dziuk, P. J., and Cook, B. (1966), Passage of steroids through silicone rubber, *Endocrinology*, 78, 208–211.

119. Mishell, D. R., Jr., Talas, M., Parlow, A. F., and Moyer, D. L. (1970), Contraception by means of a silastic vaginal ring impregnated with medroxyprogesterone acetate, *Am. J. Obstet. Gynecol.*, 107, 100–107.

120. Landgren, B. M., Johannisson, E., Masironi, B., and Diczfalusy, E. (1979), Pharmacokinetic and pharmacodynamic effects of small doses of norethisterone released from vaginal rings continuously during 90 days, *Contraception*, 19, 253–271.

121. Mishell, D. R., Jr., Lumkin, M., and Jackanicz, T. (1975), Initial clinical studies of intravaginal rings containing norethindrone and norgestrel, *Contraception*, 12, 253–260.

122. van Laarhoven, H., Veurink, J., Kruft, M. A., and Vromans, H. (2004), Influence of spinline stress on release properties of a coaxial controlled release device based on EVA polymers, *Pharm. Res.*, 21, 1811–1817.

123. van Laarhoven, J. A., Kruft, M. A., and Vromans, H. (2002), In vitro release properties of etonogestrel and ethinyl estradiol from a contraceptive vaginal ring, *Int. J. Pharm.*, 232, 163–173.

124. Vartiainen, J., Wahlstrom, T., and Nilsson, C. G. (1993), Effects and acceptability of a new 17 beta-oestradiol-releasing vaginal ring in the treatment of postmenopausal complaints, *Maturitas*, 17, 129–137.

125. Nash, H. A., Brache, V., Alvarez-Sanchez, F., Jackanicz, T. M., and Harmon, T. M. (1997), Estradiol delivery by vaginal rings: Potential for hormone replacement therapy, *Maturitas*, 26, 27–33.

126. Saleh, S. I., Khidr, S. H., Ahmed, S. M., Jackanicz, T. M., and Nash, H. A. (2003), Estradiol-progesterone interaction during the preparation of vaginal rings, *J. Pharm. Sci.*, 92, 258–265.

127. Metzger, A. P., and Matlack, J. D. (1968), Comparative swelling behavior of various thermoplastic polymers, *Polym. Eng. Sci.*, 8, 110–115.

128. Chokshi, R., and Zia, H. (2004), Hot-melt extrusion technique: A review, *Iran J. Pharm. Res.*, 3, 3–16.

129. Malcolm, K., Woolfson, D., Russell, J., Tallon, P., McAuley, L., and Craig, D. (2003), Influence of silicone elastomer solubility and diffusivity on the in vitro release of drugs from intravaginal rings, *J. Controlled Release*, 90, 217–225.

130. Lee, C. H., Bhatt, P. P., and Chien, Y. W. (1997), Effect of excipient on drug release and permeation from silicone-based barrier devices, *Int. J. Pharm.*, 43, 283–290.

131. Garg, S., Vermani, K., Garg, A., Anderson, R. A., Rencher, W. B., and Zaneveld, L. J. (2005), Development and characterization of bioadhesive vaginal films of sodium polystyrene sulfonate (PSS), a novel contraceptive antimicrobial agent, *Pharm. Res.*, 22, 584–595.

132. Yoo, J. W., Dharmala, K., and Lee, C. H. (2006), The physicodynamic properties of mucoadhesive polymeric films developed as female controlled drug delivery system, *Int. J. Pharm.*, 309, 139–145.

133. Mauck, C. K., Baker, J. M., Barr, S. P., Abercrombie, T. J., and Archer, D. F. (1997), A phase I comparative study of contraceptive vaginal films containing benzalkonium chloride and nonoxynol-9. Postcoital testing and colposcopy, *Contraception*, 56, 89–96.

134. Chien, Y. W., Oppermann, J., Nicolova, B., and Lambert, H. J. (1982), Medicated tampons: Intravaginal sustained administration of metronidazole and in vitro–in vivo relationships, *J. Pharm. Sci.*, 71, 767–771.

135. Brzezinski, A., Stern, T., Arbel, R., Rahav, G., and Benita, S. (2004), Efficacy of a novel pH-buffering tampon in preserving the acidic vaginal pH during menstruation, *Int. J. Gynaecol. Obstet.*, 85, 298–300.

136. Bernstein, G. S. (1974), Conventional methods of contraception: Condom, diaphragm, and vaginal foam, *Clin. Obstet. Gynecol.*, 17, 21–33.

137. Youssef, H., Crofton, V. A., Smith, S. C., and Siemens, A. J. (1987), A clinical trial of Neo Sampoon vaginal tablets and Emko foam in Alexandria, Egypt, *Contraception*, 35, 101–110.

138. Bullough, V. L., and Bullough, B. (1990), *Contraception: A Guide to Birth Control Methods*, Prometheus Books, Buffalo, NY.

139. Smith, C. B., Noble, V., Bensch, R., Ahlin, P. A., Jacobson, J. A., and Latham, R. H. (1982), Bacterial flora of the vagina during the menstrual cycle: Findings in users of tampons, napkins, and sea sponges, *Ann. Intern. Med.*, 96, 948–951.

140. Kafka, D., and Gold, R. B. (1983), Food and Drug Administration approves vaginal sponge, *Fam. Plann. Perspect.*, 15, 146–148.

141. Kuyoh, M. A., Toroitich-Ruto, C., Grimes, D. A., Schulz, K. F., and Gallo, M. F. (2003), Sponge versus diaphragm for contraception: A Cochrane review, *Contraception*, 67, 15–18.
142. Bowers, R., Ed. (2001), U. S. women are waiting for contraceptive sponge, *Contracept. Technol. Update*, 22, 6–7.
143. Smart, J. D. (2005), The basics and underlying mechanisms of mucoadhesion, *Adv. Drug Deliv. Rev.*, 57, 1556–1568.
144. Valenta, C. (2005), The use of mucoadhesive polymers in vaginal delivery, *Adv. Drug Deliv. Rev.*, 57, 1692–1712.
145. Hosny, E. A. (1993), Polycarbophil as a controlled release matrix, *Int. J. Pharm.*, 98, 235–238.
146. Kast, C. E., Valenta, C., Leopold, M., and Bernkop-Schnurch, A. (2002), Design and in vitro evaluation of a novel bioadhesive vaginal drug delivery system for clotrimazole, *J. Controlled Release*, 81, 347–354.
147. Valenta, C., Kast, C. E., Harich, I., and Bernkop-Schnurch, A. (2001), Development and in vitro evaluation of a mucoadhesive vaginal delivery system for progesterone, *J. Controlled Release*, 77, 323–332.
148. Peppas, N. A., Bures, P., Leobandung, W., and Ichikawa, H. (2000), Hydrogels in pharmaceutical formulations, *Eur. J. Pharm. Biopharm.*, 50, 27–46.
149. Ruel-Gariepy, E., and Leroux, J. C. (2004), In situ-forming hydrogels—Review of temperature-sensitive systems, *Eur. J. Pharm. Biopharm.*, 58, 409–426.
150. Chang, J. Y., Oh, Y. K., Kong, H. S., Kim, E. J., Jang, D. D., Nam, K. T., and Kim, C. K. (2002), Prolonged antifungal effects of clotrimazole-containing mucoadhesive thermosensitive gels on vaginitis, *J. Controlled Release*, 82, 39–50.
151. Pavelic, Z., Skalko-Basnet, N., and Schubert, R. (2001), Liposomal gels for vaginal drug delivery, *Int. J. Pharm.*, 219, 139–149.
152. Pavelic, Z., Skalko-Basnet, N., and Jalsenjak, I. (2004), Liposomal gel with chloramphenicol: Characterisation and in vitro release, *Acta Pharm.*, 54, 319–330.
153. Pavelic, Z., Skalko-Basnet, N., and Jalsenjak, I. (2005), Characterisation and in vitro evaluation of bioadhesive liposome gels for local therapy of vaginitis, *Int. J. Pharm.*, 301, 140–148.
154. Pavelic, Z., Skalko-Basnet, N., Filipovic-Grcic, J., Martinac, A., and Jalsenjak, I. (2005), Development and in vitro evaluation of a liposomal vaginal delivery system for acyclovir, *J. Controlled Release*, 106, 34–43.
155. Pavelic, Z., Skalko-Basnet, N., and Jalsenjak, I. (1999), Liposomes containing drugs for treatment of vaginal infections, *Eur. J. Pharm. Sci.*, 8, 345–351.
156. Payne, N. I., Timmins, P., Ambrose, C. V., Ward, M. D., and Ridgway, F. (1986), Proliposomes: A novel solution to an old problem, *J. Pharm. Sci.*, 75, 325–329.
157. Ning, M. Y., Guo, Y. Z., Pan, H. Z., Yu, H. M., and Gu, Z. W. (2005), Preparation and evaluation of proliposomes containing clotrimazole, *Chem. Pharm. Bull. (Tokyo)*, 53, 620–624.
158. Ning, M., Guo, Y., Pan, H., Yu, H., and Gu, Z. (2005), Niosomes with sorbitan monoester as a carrier for vaginal delivery of insulin: Studies in rats, *Drug Deliv.*, 12, 399–407.
159. Ning, M., Guo, Y., Pan, H., Chen, X., and Gu, Z. (2005), Preparation, in vitro and in vivo evaluation of liposomal/niosomal gel delivery systems for clotrimazole, *Drug Dev. Ind. Pharm.*, 31, 375–383.
160. Akagi, T., Kawamura, M., Ueno, M., Hiraishi, K., Adachi, M., Serizawa, T., Akashi, M., and Baba, M. (2003), Mucosal immunization with inactivated HIV-1-capturing

nanospheres induces a significant HIV1-specific vaginal antibody response in mice, *J. Med. Virol.*, 69, 163–172.

161. Bonucci, E., Ballanti, P., Ramires, P. A., Richardson, J. L., and Benedetti, L. M. (1995), Prevention of ovariectomy osteopenia in rats after vaginal administration of Hyaff 11 microspheres containing salmon calcitonin, *Calcif. Tissue Int.*, 56, 274–279.
162. Bilensoy, E., Rouf, M. A., Vural, I., Sen, M., and Hincal, A. A. (2006), Mucoadhesive, thermosensitive, prolonged-release vaginal gel for clotrimazole:beta-cyclodextrin complex, *AAPS PharmSciTech*, 7, E38.
163. Francois, M., Snoeckx, E., Putteman, P., Wouters, F., De Proost, E., Delaet, U., Peeters, J., and Brewster, M. E. (2003), A mucoadhesive, cyclodextrin-based vaginal cream formulation of itraconazole, *AAPS PharmSci*, 5, E5.
164. Ahmed, M. O., El-Gibaly, I., and Ahmed, S. M. (1998), Effect of cyclodextrins on the physicochemical properties and antimycotic activity of clotrimazole, *Int. J. Pharm.*, 171, 111–121.
165. Shihata, A. (2004), New FDA-approved woman-controlled, latex-free barrier contraceptive device "Fem Cap," *Int. Congr. Ser.*, 1271, 303–306.
166. Lee, C. H., Bagdon, R. E., Bhatt, P. P., and Chien, Y. W. (1997), Development of silicone-based barrier devices for controlled delivery of spermicidal agents, *J. Controlled Release*, 44, 43–53.
167. McCarron, P. A., Donnelly, R. F., Gilmore, B. F., Woolfson, A. D., McClelland, R., Zawislak, A., and Price, J. H. (2004), Phototoxicity of 5-aminolevulinic acid in the HeLa cell line as an indicative measure of photodynamic effect after topical administration to gynecological lesions of intraepithelial form, *Pharm. Res.*, 21, 1871–1879.
168. Woolfson, A. D., McCafferty, D. F., McCarron, P. A., and Price, J. H. (1995), A bioadhesive patch cervical drug delivery system for the administration of 5-fluorouracil to cervical tissue, *J. Controlled Release*, 35, 49–58.
169. Mazouni, C., Provensal, M., Menard, J. P., Heckenroth, H., Guidicelli, B., Gamerre, M., and Bretelle, F. (2006), Utilisation du dispositif vaginal Propess dans le déclenchement du travail: Efficacité et innocuité, *Gynecol. Obstet. Fertil.*, 34, 489–492.
170. Vollebregt, A., van't Hof, D. B., and Exalto, N. (2002), Prepidil compared to Propess for cervical ripening, *Eur. J. Obstet. Gynecol. Reprod. Biol.*, 104, 116–119.
171. Tedajo, G. M., Bouttier, S., Grossiord, J. L., Marty, J. P., Seiller, M., and Fourniat, J. (2002), In vitro microbicidal activity of W/O/W multiple emulsion for vaginal administration, *Int. J. Antimicrob. Agents*, 20, 50–56.
172. Kaushal, G., Trombetta, L., Ochs, R. S., and Shao, J. (2006), Delivery of TEM beta-lactamase by gene-transformed *Lactococcus lactis* subsp. *lactis* through cervical cell monolayer, *Int. J. Pharm.*, 313, 29–35.
173. Yao, X. Y., Yuan, M. M., and Li, D. J. (2006), Mucosal inoculation of *Lactobacillus* expressing hCGbeta induces an anti-hCGbeta antibody response in mice of different strains, *Methods*, 38, 124–132.
174. Vail, J. G., Cohen, J. A., and Kelly, K. L. (2004), Improving topical microbicide applicators for use in resource-poor settings, *Am. J. Public Health*, 94, 1089–1092.
175. Costa, P., and Sousa Lobo, J. M. (2001), Modeling and comparison of dissolution profiles, *Eur. J. Pharm. Sci.*, 13, 123–133.
176. Siewert, M., Dressman, J., Brown, C. K., and Shah, V. P. (2003), FIP/AAPS guidelines to dissolution/in vitro release testing of novel/special dosage forms, *AAPS PharmSciTech*, 4, E7.
177. Karasulu, H. Y., Hilmioglu, S., Metin, D. Y., and Guneri, T. (2004), Efficacy of a new ketoconazole bioadhesive vaginal tablet on *Candida albicans*, *Farmaco*, 59, 163–167.

178. Ondracek, J., Stoll, B., and Krifter, R. (1988), New basket dissolution method for vaginal suppositories, *Acta Pharm. Technol.*, 34, 169–171.
179. Chun, I. K., and Chien, Y. W. (1993), Transmucosal delivery of methionine enkephalin. I: Solution stability and kinetics of degradation in various rabbit mucosa extracts, *J. Pharm. Sci.*, 82, 373–378.
180. Bechgaard, E., Riis, K. J., and Jorgensen, L. (1994), The development of an Ussing chamber for isolated human vaginal mucosa, and the viability of the in vitro system, *Int. J. Pharm.*, 106, 237–242.
181. van der Bijl, P., van Eyk, A. D., and Thompson, I. O. (1998), Penetration of human vaginal and buccal mucosa by 4.4-kd and 12-kd fluorescein-isothiocyanate-labeled dextrans, *Oral Surg. Oral. Med. Oral Pathol. Oral Radiol. Endod.*, 85, 686–691.
182. van der Bijl, P., van Eyk, A. D., and Thompson, I. O. (1998), Permeation of 17beta-estradiol through human vaginal and buccal mucosa, *Oral Surg. Oral Med. Oral Pathol. Oral Radiol. Endod.*, 85, 393–398.
183. van der Bijl, P., van Eyk, A. D., van Wyk, C. W., and Stander, I. A. (2001), Diffusion of reduced arecoline and arecaidine through human vaginal and buccal mucosa, *J. Oral Pathol. Med.*, 30, 200–205.
184. Hsu, C. C., Park, J. Y., Ho, N. F., Higuchi, W. I., and Fox, J. L. (1983), Topical vaginal drug. delivery I: Effect of the estrous cycle on vaginal membrane permeability and diffusivity of vidarabine in mice, *J. Pharm. Sci.*, 72, 674–680.
185. Okada, H., Yashiki, T., and Mima, H. (1983), Vaginal absorption of a potent luteinizing hormone-releasing hormone analogue (leuprolide) in rats III: Effect of estrous cycle on vaginal absorption of hydrophilic model compounds, *J. Pharm. Sci.*, 72, 173–176.
186. Garg, S., Anderson, R. A., Chany, C. J., 2nd, Waller, D. P., Diao, X. H., Vermani, K., and Zaneveld, L. J. (2001), Properties of a new acid-buffering bioadhesive vaginal formulation (ACIDFORM), *Contraception*, 64, 67–75.
187. Castle, P. E., Karp, D. A., Zeitlin, L., Garcia-Moreno, E. B., Moench, T. R., Whaley, K. J., and Cone, R. A. (2002), Human monoclonal antibody stability and activity at vaginal pH, *J. Reprod. Immunol.*, 56, 61–76.
188. Owen, D. H., Peters, J. J., and Katz, D. F. (2000), Rheological properties of contraceptive gels, *Contraception*, 62, 321–326.
189. Chang, J. Y., Oh, Y. K., Choi, H. G., Kim, Y. B., and Kim, C. K. (2002), Rheological evaluation of thermosensitive and mucoadhesive vaginal gels in physiological conditions, *Int. J. Pharm.*, 241, 155–163.
190. Thorgeirsdottir, T. O., Kjoniksen, A. L., Knudsen, K. D., Kristmundsdottir, T., and Nystrom, B. (2005), Viscoelastic and structural properties of pharmaceutical hydrogels containing monocaprin, *Eur. J. Pharm. Biopharm.*, 59, 333–342.
191. Thorgeirsdottir, T. O., Thormar, H., and Kristmundsdottir, T. (2006), Viscoelastic properties of a virucidal cream containing the monoglyceride monocaprin: Effects of formulation variables: A technical note, *AAPS PharmSciTech*, 7, E44.
192. Owen, D. H., Peters, J. J., Katz, and D. F. (2001), Comparison of the rheological properties of Advantage-S and Replens, *Contraception*, 64, 393–396.
193. Madsen, F., Eberth, K., and Smart, J. D. (1998), A rheological examination of the mucoadhesive/mucus interaction: The effect of mucoadhesive type and concentration, *J. Controlled Release*, 50, 167–178.
194. Owen, D. H., Peters, J. J., Lavine, M. L., and Katz, D. F. (2003), Effect of temperature and pH on contraceptive gel viscosity, *Contraception*, 67, 57–64.
195. Owen, D. H., and Katz, D. F. (2005), A review of the physical and chemical properties of human semen and the formulation of a semen simulant, *J. Androl.*, 26, 459–469.

196. Jones, D. S., Irwin, C. R., Woolfson, A. D., Djokic, J., and Adams, V. (1999), Physicochemical characterization and preliminary in vivo efficacy of bioadhesive, semisolid formulations containing flurbiprofen for the treatment of gingivitis, *J. Pharm. Sci.*, 88, 592–598.

197. Jones, D. S., Lawlor, M. S., and Woolfson, A. D. (2002), Examination of the flow rheological and textural properties of polymer gels composed of poly(methylvinylether-*co*-maleic anhydride) and poly(vinylpyrrolidone): Rheological and mathematical interpretation of textural parameters, *J. Pharm. Sci.*, 91, 2090–2101.

198. Vermani, K., Garg, S., and Zaneveld, L. J. (2002), Assemblies for in vitro measurement of bioadhesive strength and retention characteristics in simulated vaginal environment, *Drug Dev. Ind. Pharm.*, 28, 1133–1146.

199. Tamburic, S., and Craig, D. Q. M. (1997), A comparision of different in vitro methods for measuring mucoadhesive performance, *Eur. J. Pharm. Biopharm.*, 44, 159–167.

200. El-Kamel, A., Sokar, M., Naggar, V., and Al Gamal, S. (2002), Chitosan and sodium alginate-based bioadhesive vaginal tablets, *AAPS PharmSci*, 4, E44.

201. Barnhart, K. T., Stolpen, A., Pretorius, E. S., and Malamud, D. (2001), Distribution of a spermicide containing Nonoxynol-9 in the vaginal canal and the upper female reproductive tract, *Hum. Reprod.*, 16, 1151–1154.

202. Barnhart, K. T., Pretorius, E. S., Shera, D. M., Shabbout, M., and Shaunik, A. (2006), The optimal analysis of MRI data to quantify the distribution of a microbicide, *Contraception*, 73, 82–87.

203. Kieweg, S. L., Geonnotti, A. R., and Katz, D. F. (2004), Gravity-induced coating flows of vaginal gel formulations: In vitro experimental analysis, *J. Pharm. Sci.*, 93, 2941–2952.

204. Geonnotti, A. R., Peters, J. J., and Katz, D. F. (2005), Erosion of microbicide formulation coating layers: Effects of contact and shearing with vaginal fluid or semen, *J. Pharm. Sci.*, 94, 1705–1712.

205. Ceschel, G. C., Maffei, P., Lombardi Borgia, S., Ronchi, C., and Rossi, S. (2001), Development of a mucoadhesive dosage form for vaginal administration, *Drug Dev. Ind. Pharm.*, 27, 541–547.

206. Barnhart, K., Pretorius, E. S., Stolpen, A., and Malamud, D. (2001), Distribution of topical medication in the human vagina as imaged by magnetic resonance imaging, *Fertil. Steril.*, 76, 189–195.

207. Brown, J., Hooper, G., Kenyon, C. J., Haines, S., Burt, J., Humphries, J. M., Newman, S. P., Davis, S. S., Sparrow, R. A., and Wilding, I. R. (1997), Spreading and retention of vaginal formulations in postmenopausal women as assessed by gamma scintigraphy, *Pharm. Res.*, 14, 1073–1078.

208. Chatterton, B. E., Penglis, S., Kovacs, J. C., Presnell, B., and Hunt, B. (2004), Retention and distribution of two 99mTc-DTPA labelled vaginal dosage forms, *Int. J. Pharm.*, 271, 137–143.

209. Henderson, M. H., Peters, J. J., Walmer, D. K., Couchman, G. M., and Katz, D. F. (2005), Optical instrument for measurement of vaginal coating thickness by drug delivery formulations, *Rev. Sci. Instrum.*, 76, 034302.

210. Braun, K. E., Boyer, J. D., Henderson, M. H., Katz, D. F., and Wax, A. (2006), Label-free measurement of microbicidal gel thickness using low-coherence interferometry, *J. Biomed. Opt.*, 11, 020504.1–020504.3.

211. Brache, V., Cohen, J. A., Cochon, L., and Alvarez, F. (2006), Evaluating the clinical safety of three vaginal applicators: A pilot study conducted in the Dominican Republic, *Contraception*, 73, 72–77.

212. Bagga, R., Raghuvanshi, P., Gopalan, S., Das, S. K., Baweja, R., Suri, S., Malhotra, D., Khare, S., and Talwar, G. P. (2006), A polyherbal vaginal pessary with spermicidal and antimicrobial action: Evaluation of its safety, *Trans. R. Soc. Trop. Med. Hyg.*, 100, 1164–1167.

213. D'Cruz, O. J., and Uckun, F. M. (2001), Lack of adverse effects on fertility of female CD-1 mice exposed to repetitive intravaginal gel-microemulsion formulation of a dual-function anti-HIV agent: Aryl phosphate derivative of bromo-methoxy-zidovudine (compound WHI-07), *J. Appl. Toxicol.*, 21, 317–322.

214. Krebs, F. C., Miller, S. R., Catalone, B. J., Fichorova, R., Anderson, D., Malamud, D., Howett, M. K., and Wigdahl, B. (2002), Comparative in vitro sensitivities of human immune cell lines, vaginal and cervical epithelial cell lines, and primary cells to candidate microbicides nonoxynol 9, C31G, and sodium dodecyl sulfate, *Antimicrob. Agents Chemother.*, 46, 2292–2298.

215. Dezzutti, C. S., James, V. N., Ramos, A., Sullivan, S. T., Siddig, A., Bush, T. J., Grohskopf, L. A., Paxton, L., Subbarao, S., and Hart, C. E. (2004), In vitro comparison of topical microbicides for prevention of human immunodeficiency virus type 1 transmission, *Antimicrob. Agents Chemother.*, 48, 3834–3844.

216. Ayehunie, S., Cannon, C., Lamore, S., Kubilus, J., Anderson, D. J., Pudney, J., and Klausner, M. (2006), Organotypic human vaginal-ectocervical tissue model for irritation studies of spermicides, microbicides, and feminine-care products, *Toxicol. In Vitro*, 20, 689–698.

217. Eckstein, P., Jackson, M. C., Millman, N., and Sobrero, A. J. (1969), Comparison of vaginal tolerance tests of spermicidal preparations in rabbits and monkeys, *J. Reprod. Fertil.*, 20, 85–93.

218. D'Cruz, O. J., and Uckun, F. M. (2001), Intravaginal toxicity studies of a gel-microemulsion formulation of spermicidal vanadocenes in rabbits, *Toxicol. Appl. Pharmacol.*, 170, 104–112.

219. Patton, D. L., Sweeney, Y. C., Tsai, C. C., and Hillier, S. L. (2004), *Macaca fascicularis* vs. *Macaca nemestrina* as a model for topical microbicide safety studies, *J. Med. Primatol.*, 33, 105–108.

220. Gray, J. E., Weaver, R. N., Lohrberg, S. M., and Larsen, E. R. (1984), Comparative responses of vaginal mucosa to chronic pyrimidinone-induced irritation, *Toxicol. Pathol.*, 12, 228–234.

221. D'Cruz, O. J., Erbeck, D., and Uckun, F. M. (2005), A study of the potential of the pig as a model for the vaginal irritancy of benzalkonium chloride in comparison to the nonirritant microbicide PHI-443 and the spermicide vanadocene dithiocarbamate, *Toxicol. Pathol.*, 33, 465–476.

222. D'Cruz, O. J., Waurzyniak, B., Yiv, S. H., and Uckun, F. M. (2000), Evaluation of subchronic (13-week) and reproductive toxicity potential of intravaginal gel-microemulsion formulation of a dual-function phenyl phosphate derivative of bromo-methoxy zidovudine (compound whi-07) in B(6)C(3)F(1) mice, *J. Appl. Toxicol.*, 20, 319–325.

223. Chvapil, M., Droegemueller, W., Owen, J. A., Eskelson, C. D., and Betts, K. (1980), Studies of nonoxynol-9. I. The effect on the vaginas of rabbits and rats, *Fertil. Steril.*, 33, 445–450.

224. CPMP/EMEA (1997), *Replacement of Animal Studies by In Vitro Models*, CPMP/EMEA, London.

225. Adriaens, E., and Remon, J. P. (1999), Gastropods as an evaluation tool for screening the irritating potency of absorption enhancers and drugs, *Pharm. Res.*, 16, 1240–1244.

226. Dhondt, M. M., Adriaens, E., Roey, J. V., and Remon, J. P. (2005), The evaluation of the local tolerance of vaginal formulations containing dapivirine using the Slug Mucosal

Irritation test and the rabbit vaginal irritation test, *Eur. J. Pharm. Biopharm.*, 60, 419–425.

227. Mauck, C. K., Weiner, D. H., Ballagh, S. A., Creinin, M. D., Archer, D. F., Schwartz, J. L., Pymar, H. C., Lai, J. J., Rencher, W. F., and Callahan, M. M. (2004), Single and multiple exposure tolerance study of polystyrene sulfonate gel: A phase I safety and colposcopy study, *Contraception*, 70, 77–83.
228. Mauck, C. K., Baker, J. M., Birnkrant, D. B., Rowe, P. J., and Gabelnick, H. L. (2000), The use of colposcopy in assessing vaginal irritation in research, *AIDS*, 14, 2221–2227.
229. WHO/CONRAD (2004), *Manual for the Standardization of Colposcopy for the Evaluation of Vaginal products, Update 2004*, CONRAD/WHO, Geneva.
230. van der Laak, J. A., de Bie, L. M., de Leeuw, H., de Wilde, P. C., and Hanselaar, A. G. (2002), The effect of Replens on vaginal cytology in the treatment of postmenopausal atrophy: Cytomorphology versus computerised cytometry, *J. Clin. Pathol.*, 55, 446–451.
231. Rosen, A. D., and Rosen, T. (1999), Study of condom integrity after brief exposure to over-the-counter vaginal preparations, *South Med. J.*, 92, 305–307.
232. Haineault, C., Gourde, P., Perron, S., Desormeaux, A., Piret, J., Omar, R. F., Tremblay, R. R., and Bergeron, M. G. (2003), Thermoreversible gel formulation containing sodium lauryl sulfate as a potential contraceptive device, *Biol. Reprod.*, 69, 687–694.
233. Ansbacher, R. (2004), Intravaginal medications, *Fertil Steril*, 82, 1474 (Letter).
234. Stone, A. (2002), Microbicides: A new approach to preventing HIV and other sexually transmitted infections, *Nat. Rev. Drug Discov.*, 1, 977–985.
235. Rowe, P. M. (1995), Research into topical microbicides against STDs, *Lancet*, 345, 1231.
236. Watts, C., and Vickerman, P. (2001), The impact of microbicides on HIV and STD transmission: Model projections, *AIDS*, 15(Suppl. 1), S43–44.
237. Lederman, M. M., Veazey, R. S., Offord, R., Mosier, D. E., Dufour, J., Mefford, M., Piatak, M., Jr., Lifson, J. D., Salkowitz, J. R., Rodriguez, B., Blauvelt, A., and Hartley, O. (2004), Prevention of vaginal SHIV transmission in rhesus macaques through inhibition of CCR5, *Science*, 306, 485–487.
238. Yedery, R. D., and Reddy, K. V. (2005), Antimicrobial peptides as microbicidal contraceptives: Prophecies for prophylactics—A mini review, *Eur. J. Contracept. Reprod. Health Care*, 10, 32–42.
239. Veazey, R. S., Shattock, R. J., Pope, M., Kirijan, J. C., Jones, J., Hu, Q., Ketas, T., Marx, P. A., Klasse, P. J., Burton, D. R., and Moore, J. P. (2003), Prevention of virus transmission to macaque monkeys by a vaginally applied monoclonal antibody to HIV-1 gp120, *Nat. Med.*, 9, 343–346.
240. Veazey, R. S., Klasse, P. J., Schader, S. M., Hu, Q., Ketas, T. J., Lu, M., Marx, P. A., Dufour, J., Colonno, R. J., Shattock, R. J., Springer, M. S., and Moore, J. P. (2005), Protection of macaques from vaginal SHIV challenge by vaginally delivered inhibitors of virus-cell fusion, *Nature*, 438, 99–102.
241. Neurath, A. R., Strick, N., Li, Y. Y., and Debnath, A. K. (2005), *Punica granatum* (pomegranate) juice provides an HIV-1 entry inhibitor and candidate topical microbicide, *Ann. N. Y. Acad. Sci.*, 1056, 311–327.
242. Liu, S., Lu, H., Neurath, A. R., and Jiang, S. (2005), Combination of candidate microbicides cellulose acetate 1,2-benzenedicarboxylate and UC781 has synergistic and complementary effects against human immunodeficiency virus type 1 infection, *Antimicrob. Agents Chemother.*, 49, 1830–1836.
243. Roddy, R. E., Zekeng, L., Ryan, K. A., Tamoufe, U., Weir, S. S., and Wong, E. L. (1998), A controlled trial of nonoxynol 9 film to reduce male-to-female transmission of sexually transmitted diseases, *N. Engl. J. Med.*, 339, 504–510.

244. International Working Group On Vaginal Microbicides (1996), Recommendations for the development of vaginal microbicides, *AIDS*, 10, 1–6.

245. Mauck, C., Rosenberg, Z., and Van Damme, L. (2001), Recommendations for the clinical development of topical microbicides: An update, *AIDS*, 15, 857–868.

246. Garg, S., Kandarapu, R., Vermani, K., Tambwekar, K. R., Garg, A., Waller, D. P., and Zaneveld, L. J. (2003), Development pharmaceutics of microbicide formulations. Part I: Preformulation considerations and challenges, *AIDS Patient Care STDS*, 17, 17–32.

247. Garg, S., Tambwekar, K. R., Vermani, K., Kandarapu, R., Garg, A., Waller, D. P., and Zaneveld, L. J. (2003), Development pharmaceutics of microbicide formulations. Part II: Formulation, evaluation, and challenges, *AIDS Patient Care STDS*, 17, 377–399.

248. Bax, R., Douville, K., McCormick, D., Rosenberg, M., Higgins, J., and Bowden, M. (2002), Microbicides—Evaluating multiple formulations of C31G, *Contraception*, 66, 365–368.

249. D'Cruz, O. J., Samuel, P., and Uckun, F. M. (2005), Conceival, a novel noncontraceptive vaginal vehicle for lipophilic microbicides, *AAPS PharmSciTech*, 6, E56–64.

250. Chang, T. L., Chang, C. H., Simpson, D. A., Xu, Q., Martin, P. K., Lagenaur, L. A., Schoolnik, G. K., Ho, D. D., Hillier, S. L., Holodniy, M., Lewicki, J. A., and Lee, P. P. (2003), Inhibition of HIV infectivity by a natural human isolate of *Lactobacillus jensenii* engineered to express functional two-domain CD4, *Proc. Natl. Acad. Sci. U.S.A.*, 100, 11672–11677.

251. Watson, M. C., Grimshaw, J. M., Bond, C. M., Mollison, J., and Ludbrook, A. (2002), Oral versus intravaginal imidazole and triazole anti-fungal agents for the treatment of uncomplicated vulvovaginal candidiasis (thrush): A systematic review, *BJOG*, 109, 85–95.

252. Sobel, J. D. (1997), Vaginitis, *N. Engl. J. Med.*, 337, 1896–1903.

253. SaiRam, M., Ilavazhagan, G., Sharma, S. K., Dhanraj, S. A., Suresh, B., Parida, M. M., Jana, A. M., Devendra, K., and Selvamurthy, W. (2000), Anti-microbial activity of a new vaginal contraceptive NIM-76 from neem oil (Azadirachta indica), *J. Ethnopharmacol.*, 71, 377–382.

254. Pina-Vaz, C., Goncalves Rodrigues, A., Pinto, E., Costa-de-Oliveira, S., Tavares, C., Salgueiro, L., Cavaleiro, C., Goncalves, M. J., and Martinez-de-Oliveira, J. (2004), Antifungal activity of Thymus oils and their major compounds, *J. Eur. Acad. Dermatol. Venereol.*, 18, 73–78.

255. Brown, D., Henzl, M. R., and Kaufman, R. H. (1999), Butoconazole nitrate 2% for vulvovaginal candidiasis. New, single-dose vaginal cream formulation vs. seven-day treatment with miconazole nitrate. Gynazole 1 Study Group, *J. Reprod. Med.*, 44, 933–938.

256. Faro, S., and Skokos, C. K. (2005), The efficacy and safety of a single dose of Clindesse vaginal cream versus a seven-dose regimen of Cleocin vaginal cream in patients with bacterial vaginosis, *Infect. Dis. Obstet. Gynecol.*, 13, 155–160.

257. Upmalis, D. H., Cone, F. L., Lamia, C. A., Reisman, H., Rodriguez-Gomez, G., Gilderman, L., and Bradley, L. (2000), Single-dose miconazole nitrate vaginal ovule in the treatment of vulvovaginal candidiasis: Two single-blind, controlled studies versus miconazole nitrate 100 mg cream for 7 days, *J. Womens Health Gend. Based Med.*, 9, 421–429.

258. Mikamo, H., Kawazoe, K., Izumi, K., Watanabe, K., Ueno, K., and Tamaya, T. (1997), Comparative study on vaginal or oral treatment of bacterial vaginosis, *Chemotherapy*, 43, 60–68.

259. Cunningham, F. E., Kraus, D. M., Brubaker, L., and Fischer, J. H. (1994), Pharmacokinetics of intravaginal metronidazole gel, *J. Clin. Pharmacol.*, 34, 1060–1065.

260. Borin, M. T. (1990), Systemic absorption of clindamycin following intravaginal application of clindamycin phosphate 1% cream, *J. Clin. Pharmacol.*, 30, 33–38.

261. Czeizel, A. E., Kazy, Z., and Vargha, P. (2003), A population-based case-control teratological study of vaginal econazole treatment during pregnancy, *Eur. J. Obstet. Gynecol. Reprod. Biol.*, 111, 135–140.

262. Polatti, F., Rampino, M., Magnani, P., and Mascarucci, P. (2006), Vaginal pH-lowering effect of locally applied vitamin C in subjects with high vaginal pH, *Gynecol. Endocrinol.*, 22, 230–234.

263. Amaral, E., Perdigao, A., Souza, M. H., Mauck, C., Waller, D., Zaneveld, L., and Faundes, A. (2006), Vaginal safety after use of a bioadhesive, acid-buffering, microbicidal contraceptive gel (ACIDFORM) and a 2% nonoxynol-9 product, *Contraception*, 73, 542–547.

264. Decena, D. C., Co, J. T., Manalastas, R. M., Jr., Palaypayon, E. P., Padolina, C. S., Sison, J. M., Dancel, L. A., and Lelis, M. A. (2006), Metronidazole with Lactacyd vaginal gel in bacterial vaginosis, *J. Obstet. Gynaecol. Res.*, 32, 243–251.

265. Fiorilli, A., Molteni, B., and Milani, M. (2005), Successful treatment of bacterial vaginosis with a policarbophil-carbopol acidic vaginal gel: Results from a randomised double-blind, placebo-controlled trial, *Eur. J. Obstet. Gynecol. Reprod. Biol.*, 120, 202–205.

266. Levine, H., and Watson, N. (2000), Comparison of the pharmacokinetics of crinone 8% administered vaginally versus Prometrium administered orally in postmenopausal women, *Fertil. Steril.*, 73, 516–521.

267. van den Heuvel, M. W., van Bragt, A. J., Alnabawy, A. K., and Kaptein, M. C. (2005), Comparison of ethinylestradiol pharmacokinetics in three hormonal contraceptive formulations: The vaginal ring, the transdermal patch and an oral contraceptive, *Contraception*, 72, 168–174.

268. Fraser, I. S., Lacarra, M., Mishell, D. R., Alvarez, F., Brache, V., Lahteenmaki, P., Elomaa, K., Weisberg, E., and Nash, H. A. (2000), Vaginal epithelial surface appearances in women using vaginal rings for contraception, *Contraception*, 61, 131–138.

269. Bjarnadóttir, R. I. (2003), Update on contraceptive vaginal rings, *Rev. Gyn. Pract.*, 3, 156–159.

270. Roumen, F. J., Apter, D., Mulders, T. M., and Dieben, T. O. (2001), Efficacy, tolerability and acceptability of a novel contraceptive vaginal ring releasing etonogestrel and ethinyl oestradiol, *Hum. Reprod.*, 16, 469–475.

271. Oddsson, K., Leifels-Fischer, B., de Melo, N. R., Wiel-Masson, D., Benedetto, C., Verhoeven, C. H., and Dieben, T. O. (2005), Efficacy and safety of a contraceptive vaginal ring (NuvaRing) compared with a combined oral contraceptive: A 1-year randomized trial, *Contraception*, 71, 176–182.

272. Dieben, T. O., Roumen, F. J., and Apter, D. (2002), Efficacy, cycle control, and user acceptability of a novel combined contraceptive vaginal ring, *Obstet. Gynecol.*, 100, 585–593.

273. Sivin, I., Mishell, D. R., Jr., Alvarez, F., Brache, V., Elomaa, K., Lahteenmaki, P., Massai, R., Miranda, P., Croxatto, H., Dean, C., Small, M., Nash, H., and Jackanicz, T. M. (2005), Contraceptive vaginal rings releasing Nestorone and ethinylestradiol: A 1-year dose-finding trial, *Contraception*, 71, 122–129.

274. Rad, M., Kluft, C., Menard, J., Burggraaf, J., de Kam, M. L., Meijer, P., Sivin, I., and Sitruk-Ware, R. L. (2006), Comparative effects of a contraceptive vaginal ring delivering a nonandrogenic progestin and continuous ethinyl estradiol and a combined oral contraceptive containing levonorgestrel on hemostasis variables, *Am. J. Obstet. Gynecol.*, 195, 72–77.

275. Devoto, L., Fuentes, A., Palomino, A., Espinoza, A., Kohen, P., Ranta, S., and von Hertzen, H. (2005), Pharmacokinetics and endometrial tissue levels of levonorgestrel after administration of a single 1.5-mg dose by the oral and vaginal route, *Fertil. Steril.*, 84, 46–51.

276. Mor, E., Saadat, P., Kives, S., White, E., Reid, R. L., Paulson, R. J., and Stanczyk, F. Z. (2005), Comparison of vaginal and oral administration of emergency contraception, *Fertil. Steril.*, 84, 40–45.

277. Kives, S., Hahn, P. M., White, E., Stanczyk, F. Z., and Reid, R. L. (2005), Bioavailability of the Yuzpe and levonorgestrel regimens of emergency contraception: Vaginal vs. oral administration, *Contraception*, 71, 197–201.

278. Yoo, J. W., and Lee, C. H. (2006), Drug delivery systems for hormone therapy, *J. Controlled Release*, 112, 1–14.

279. Beral, V. (2003), Breast cancer and hormone-replacement therapy in the Million Women Study, *Lancet*, 362, 419–427.

280. Cardozo, L., Bachmann, G., McClish, D., Fonda, D., and Birgerson, L. (1998), Meta-analysis of estrogen therapy in the management of urogenital atrophy in postmenopausal women: Second report of the Hormones and Urogenital Therapy Committee, *Obstet. Gynecol.*, 92, 722–727.

281. Crandall, C. (2002), Vaginal estrogen preparations: A review of safety and efficacy for vaginal atrophy. *J. Womens Health (Larchmt.)*, 11, 857–877.

282. Ballagh, S. A. (2004), Vaginal rings for menopausal symptom relief, *Drugs Aging*, 21, 757–766.

283. Henriksson, L., Stjernquist, M., Boquist, L., Cedergren, I., and Selinus, I. (1996), A one-year multicenter study of efficacy and safety of a continuous, low-dose, estradiol-releasing vaginal ring (Estring) in postmenopausal women with symptoms and signs of urogenital aging, *Am. J. Obstet. Gynecol.*, 174, 85–92.

284. Eriksen, B. (1999), A randomized, open, parallel-group study on the preventive effect of an estradiol-releasing vaginal ring (Estring) on recurrent urinary tract infections in postmenopausal women, *Am. J. Obstet. Gynecol.*, 180, 1072–1079.

285. Tavaniotou, A., Smitz, J., Bourgain, C., and Devroey, P. (2000), Comparison between different routes of progesterone administration as luteal phase support in infertility treatments, *Hum. Reprod. Update*, 6, 139–148.

286. Kleinstein, J. (2005), Efficacy and tolerability of vaginal progesterone capsules (Utrogest 200) compared with progesterone gel (Crinone 8%) for luteal phase support during assisted reproduction, *Fertil. Steril.*, 83, 1641–1649.

287. Zegers-Hochschild, F., Balmaceda, J. P., Fabres, C., Alam, V., Mackenna, A., Fernandez, E., Pacheco, I. M., Sepulveda, M. S., Chen, S., Borrero, C., and Borges, E. (2000), Prospective randomized trial to evaluate the efficacy of a vaginal ring releasing progesterone for IVF and oocyte donation, *Hum. Reprod.*, 15, 2093–2097.

288. Gupta, G. (2005), Microbicidal spermicide or spermicidal microbicide? *Eur. J. Contracept. Reprod. Health Care*, 10, 212–218.

289. Castle, P. E., Whaley, K. J., Hoen, T. E., Moench, T. R., and Cone, R. A. (1997), Contraceptive effect of sperm-agglutinating monoclonal antibodies in rabbits, *Biol. Reprod.*, 56, 153–159.

290. Raymond, E. G., Chen, P. L., and Luoto, J. (2004), Contraceptive effectiveness and safety of five nonoxynol-9 spermicides: A randomized trial, *Obstet. Gynecol.*, 103, 430–439.

291. Lee, C. H., Bagdon, R., and Chien, Y. W. (1996), Comparative in vitro spermicidal activity of chelating agents and synergistic effect with nonoxynol-9 on human sperm functionality, *J. Pharm. Sci.*, 85, 91–95.

292. Zavos, P. M., Correa, J. R., and Zarmakoupis-Zavos, P. N. (1998), Assessment of a tablet drug delivery system incorporating nonoxynol-9 coprecipitated with polyvinylpyrrolidone in preventing the onset of pregnancy in rabbits, *Fertil. Steril.*, 69, 768–773.

293. Fowler, P. T., Doncel, G. F., Bummer, P. M., and Digenis, G. A. (2003), Coprecipitation of nonoxynol-9 with polyvinylpyrrolidone to decrease vaginal irritation potential while maintaining spermicidal potency, *AAPS PharmSciTech*, 4, E30.

294. Zavos, P. M., and Correa, J. R. (1997), Vaginal delivery of new formulations of nonoxynol-9 coprecipitated with polyvinylpyrrolidone in rabbits. Comparisons between two formulation delivery systems, *Contraception*, 56, 123–127.

295. Gagne, N., Cormier, H., Omar, R. F., Desormeaux, A., Gourde, P., Tremblay, M. J., Juhasz, J., Beauchamp, D., Rioux, J. E., and Bergeron, M. G. (1999), Protective effect of a thermoreversible gel against the toxicity of nonoxynol-9, *Sex Transm. Dis.*, 26, 177–183.

296. El-Gizawy, S. A., and Aglan, N. I. (2003), Formulation and evaluation of metronidazole acid gel for vaginal contraception, *J. Pharm. Pharmacol.*, 55, 903–909.

297. Owen, D. H., Dunmire, E. N., Plenys, A. M., and Katz, D. F. (1999), Factors influencing nonoxynol-9 permeation and bioactivity in cervical mucus, *J. Controlled Release*, 60, 23–34.

298. Saxena, B. B., Singh, M., Gospin, R. M., Chu, C. C., and Ledger, W. J. (2004), Efficacy of nonhormonal vaginal contraceptives from a hydrogel delivery system, *Contraception*, 70, 213–219.

299. Saltzman, W. M., and Tena, L. B. (1991), Spermicide permeation through biocompatible polymers, *Contraception*, 43, 497–505.

300. Gauger, L. J., and Curet, L. B. (1991), Comparative efficacy of intravaginal prostaglandin E2 in the gel and suppository forms for cervical ripening, DICP *Ann. Pharmacother.*, 25, 456–460.

301. Chyu, J. K., and Strassner, H. T. (1997), Prostaglandin E2 for cervical ripening: A randomized comparison of Cervidil versus Prepidil, *Am. J. Obstet. Gynecol.*, 177, 606–611.

302. Carbonell, J. L., Rodriguez, J., Delgado, E., Sanchez, C., Vargas, F., Valera, L., Mari, J., Valero, F., Salvador, I., and Llorente, M. (2004), Vaginal misoprostol 800 microg every 12 h for second-trimester abortion, *Contraception*, 70, 55–60.

303. Carbonell, J. L., Rodriguez, J., Velazco, A., Tanda, R., Sanchez, C., Barambio, S., Chami, S., Valero, F., Mari, J., de Vargas, F., and Salvador, I. (2003), Oral and vaginal misoprostol 800 microg every 8 h for early abortion, *Contraception*, 67, 457–462.

304. Carbonell, J. L., Rodriguez, J., Aragon, S., Velazco, A., Tanda, R., Sanchez, C., Barambio, S., Chami, S., and Valero, F. (2001), Vaginal misoprostol 1000 microg for early abortion, *Contraception*, 63, 131–136.

305. Zieman, M., Fong, S. K., Benowitz, N. L., Banskter, D., and Darney, P. D. (1997), Absorption kinetics of misoprostol with oral or vaginal administration, *Obstet. Gynecol.*, 90, 88–92.

306. Richardson, J. L., and Illum, L. (1992), (D) Routes of delivery: Case studies: (8) the vaginal route of peptide and protein drug delivery, *Adv. Drug Deliv. Rev.*, 8, 341–366.

307. Richardson, J. L., Ramires, P. A., Miglietta, M. R., Rochira, M., Bacelle, L., Callegaro, L., and Benedetti, L. (1995), Novel vaginal delivery systems for calcitonin: I. Evaluation of HYAFF/calcitonin microspheres in rats, *Int. J. Pharm.*, 115, 9–15.

308. Rochira, M., Miglietta, M. R., Richardson, J. L., Ferrari, L., Beccaro, M., and Benedetti, L. (1996), Novel vaginal delivery systems for calcitonin: II. Preparation and characterization of HYAFF® microspheres containing calcitonin, *Int. J. Pharm.*, 144, 19–26.

309. Sherwood, J. K., Zeitlin, L., Whaley, K. J., Cone, R. A., and Saltzman, M. (1996), Controlled release of antibodies for long-term topical passive immunoprotection of female mice against genital herpes, *Nat. Biotechnol.*, 14, 468–471.

310. Saltzman, W. M., Sherwood, J. K., Adams, D. R., Castle, P., and Haller, P. (2000), Long-term vaginal antibody delivery: Delivery systems and biodistribution, *Biotechnol. Bioeng.*, 67, 253–264.

311. Saltzman, W. M., and Langer, R. (1989), Transport rates of proteins in porous materials with known microgeometry, *Biophys. J.*, 55, 163–171.

312. Morimoto, K., Takeeda, T., Nakamoto, Y., and Morisaka, K. (1982), Effective vaginal absorption of insulin in diabetic rats and rabbits using polyacrylic acid aqueous gel bases, *Int. J. Pharm.*, 12, 107–111.

313. Degim, Z., Degim, T., Acarturk, F., Erdogan, D., Ozogul, C., and Koksal, M. (2005), Rectal and vaginal administration of insulin-chitosan formulations: An experimental study in rabbits, *J. Drug Target.*, 13, 563–572.

314. Bogers, W. M., Bergmeier, L. A., Ma, J., Oostermeijer, H., Wang, Y., Kelly, C. G., Ten Haaft, P., Singh, M., Heeney, J. L., and Lehner, T. (2004), A novel HIV-CCR5 receptor vaccine strategy in the control of mucosal SIV/HIV infection, *AIDS*, 18, 25–36.

315. Livingston, J. B., Lu, S., Robinson, H., and Anderson, D. J. (1998), Immunization of the female genital tract with a DNA-based vaccine, *Infect. Immun.*, 66, 322–329.

316. Kozlowski, P. A., Cu-Uvin, S., Neutra, M. R., and Flanigan, T. P. (1997), Comparison of the oral, rectal, and vaginal immunization routes for induction of antibodies in rectal and genital tract secretions of women, *Infect. Immun.*, 65, 1387–1394.

317. Kozlowski, P. A., Cu-Uvin, S., Neutra, M. R., and Flanigan, T. P. (1999), Mucosal vaccination strategies for women, *J. Infect. Dis.*, 179(Suppl. 3), S493–498.

318. Wassen, L., Schon, K., Holmgren, J., Jertborn, M., and Lycke, N. (1996), Local intravaginal vaccination of the female genital tract, *Scand. J. Immunol.*, 44, 408–414.

319. Kozlowski, P. A., Williams, S. B., Lynch, R. M., Flanigan, T. P., Patterson, R. R., Cu-Uvin, S., and Neutra, M. R. (2002), Differential induction of mucosal and systemic antibody responses in women after nasal, rectal, or vaginal immunization: Influence of the menstrual cycle, *J. Immunol.*, 169, 566–574.

320. Han, I. K., Kim, Y. B., Kang, H. S., Sul, D., Jung, W. W., Cho, H. J., and Oh, Y. K. (2006), Thermosensitive and mucoadhesive delivery systems of mucosal vaccines, *Methods*, 38, 106–111.

321. Oh, Y. K., Park, J. S., Yoon, H., and Kim, C. K. (2003), Enhanced mucosal and systemic immune responses to a vaginal vaccine coadministered with RANTES-expressing plasmid DNA using in situ-gelling mucoadhesive delivery system, *Vaccine*, 21, 1980–1988.

322. Shen, H., Goldberg, E., and Saltzman, W. M. (2003), Gene expression and mucosal immune responses after vaginal DNA immunization in mice using a controlled delivery matrix, *J. Controlled Release*, 86, 339–348.

323. Luo, D., Woodrow-Mumford, K., Belcheva, N., and Saltzman, W. M. (1999), Controlled DNA delivery systems, *Pharm. Res.*, 16, 1300–1308.

324. Oggioni, M. R., Medaglini, D., Maggi, T., and Pozzi, G. (1999), Engineering the gram-positive cell surface for construction of bacterial vaccine vectors, *Methods*, 19, 163–173.

325. Medaglini, D., Oggioni, M. R., and Pozzi, G. (1998), Vaginal immunization with recombinant gram-positive bacteria, *Am. J. Reprod. Immunol.*, 39, 199–208.

326. Singh, S. R., Hulett, K., Pillai, S. R., Dennis, V. A., Oh, M. K., and Scissum-Gunn, K. (2006), Mucosal immunization with recombinant MOMP genetically linked with modified cholera toxin confers protection against *Chlamydia trachomatis* infection, *Vaccine*, 24, 1213–1224.

327. O'Hagan, D. T., Rafferty, D., Wharton, S., and Illum, L. (1993), Intravaginal immunization in sheep using a bioadhesive microsphere antigen delivery system, *Vaccine*, 11, 660–664.
328. Syed, T. A., Qureshi, Z. A., Ahmad, S. A., and Ali, S. M. (2000), Management of intravaginal warts in women with 5-fluorouracil (1%) in vaginal hydrophilic gel: A placebo-controlled double-blind study, *Int. J. STD AIDS*, 11, 371–374.
329. Maiman, M., Watts, D. H., Andersen, J., Clax, P., Merino, M., and Kendall, M. A. (1999), Vaginal 5-fluorouracil for high-grade cervical dysplasia in human immunodeficiency virus infection: A randomized trial, *Obstet. Gynecol.*, 94, 954–961.
330. Darwish, A. M., Hafez, E., El-Gebali, I., and Hassan, S. B. (2005), Evaluation of a novel vaginal bromocriptine mesylate formulation: A pilot study. *Fertil. Steril.*, 83, 1053–1055.
331. Motta, T., de Vincentiis, S., Marchini, M., Colombo, N., and D'Alberton, A. (1996), Vaginal cabergoline in the treatment of hyperprolactinemic patients intolerant to oral dopaminergics, *Fertil. Steril.*, 65, 440–442.
332. Stray-Pedersen, B., Bergan, T., Hafstad, A., Normann, E., Grogaard, J., and Vangdal, M. (1999), Vaginal disinfection with chlorhexidine during childbirth, *Int. J. Antimicrob. Agents*, 12, 245–251.
333. Garcia-Lopez, P., Coll, M., Cervera, E., Reyes-Vermot, L., Torres, M. A., Abrego-Perez, G., Hernandez-Pajaro, A. I., Castaneda-HernAndez, G., Mohar-Betancourt, A., and Meneses, A. (2006), The systemic absorption of etoposide after intravaginal administration in patients with cervical intraepithelial lesions associated with human papillomavirus infection, *Pharm. Res.*, 23, 378–383.
334. Diakomanolis, E., Haidopoulos, D., and Stefanidis, K. (2002), Treatment of high-grade vaginal intraepithelial neoplasia with imiquimod cream, *N. Engl. J. Med.*, 347, 374 (Letter).
335. Abramov, Y., Nadjari, M., Weinstein, D., Ben-Shachar, I., Plotkin, V., and Ezra, Y. (2000), Indomethacin for preterm labor: A randomized comparison of vaginal and rectal-oral routes, *Obstet. Gynecol.*, 95, 482–486.
336. Costello, M. F., Steigrad, S., and Collet, A. (2005), A prospective, randomised, single-blinded, controlled trial comparing two topical anaesthetic modalities for the application of a tenaculum to the cervix, *J. Obstet. Gynaecol.*, 25, 781–785.
337. Ostrop, N. J., Lamb, J., and Reid, G. (1998), Intravaginal morphine: An alternative route of administration, *Pharmacotherapy*, 18, 863–865.
338. Schroder, A., Levin, R. M., Kogan, B. A., Das, A. K., Kay, F., and Mahashabde, A. (2000), Absorption of oxybutynin from vaginal inserts: Drug blood levels and the response of the rabbit bladder, *Urology*, 56, 1063–1067.
339. Smith, R. C., and Berge, M. (1985), Systemic absorption after intravaginal propranolol, *Lancet*, 1, 112 (Letter).
340. Sher, G., and Fisch, J. D. (2002), Effect of vaginal sildenafil on the outcome of in vitro fertilization (IVF) after multiple IVF failures attributed to poor endometrial development, *Fertil. Steril.*, 78, 1073–1076.
341. Lin, H., Huang, E. Y., Chang, H. Y., and ChangChien, C. C. (2005), Therapeutic effect of topical applications of trichloroacetic acid for vaginal intraepithelial neoplasia after hysterectomy, *Jpn. J. Clin. Oncol.*, 35, 651–654.
342. Nappi, R. E., Liekens, G., and Brandenburg, U. (2006), Attitudes, perceptions and knowledge about the vagina: The International Vagina Dialogue Survey, *Contraception*, 73, 493–500.
343. Hammett, T. M., Mason, T. H., Joanis, C. L., Foster, S. E., Harmon, P., Robles, R. R., Finlinson, H. A., Feudo, R., Vining-Bethea, S., Jeter, G., Mayer, K. H., Doherty-Iddings, P., and Seage, G. R., 3rd (2000), Acceptability of formulations and application methods for

vaginal microbicides among drug-involved women: Results of product trials in three cities, *Sex Transm. Dis.*, 27, 119–126.

344. Mantell, J. E., Myer, L., Carballo-Dieguez, A., Stein, Z., Ramjee, G., Morar, N. S., and Harrison, P. F. (2005), Microbicide acceptability research: Current approaches and future directions, *Soc. Sci. Med.*, 60, 319–330.
345. Raymond, E., Alvarado, G., Ledesma, L., Diaz, S., Bassol, S., Morales, E., Fernandez, V., and Carlos, G. (1999), Acceptability of two spermicides in five countries, *Contraception*, 60, 45–50.
346. Hardy, E., de Padua, K. S., Hebling, E. M., Osis, M. J., and Zaneveld, L. J. (2003), Women's preferences for vaginal antimicrobial contraceptives. V: Attitudes of Brazilian women to the insertion of vaginal products, *Contraception*, 67, 391–395.
347. Bentley, M. E., Fullem, A. M., Tolley, E. E., Kelly, C. W., Jogelkar, N., Srirak, N., Mwafulirwa, L., Khumalo-Sakutukwa, G., and Celentano, D. D. (2004), Acceptability of a microbicide among women and their partners in a 4-country phase I trial, *Am. J. Public Health*, 94, 1159–1164.
348. Braunstein, S., and van de Wijgert, J. (2005), Preferences and practices related to vaginal lubrication: Implications for microbicide acceptability and clinical testing, *J. Womens Health (Larchmt.)*, 14, 424–433.
349. Aubeny, E., Colau, J. C., and Nandeuil, A. (2000), Local spermicidal contraception: A comparative study of the acceptability and safety of a new pharmaceutical formulation of benzalkonium chloride, the vaginal capsule, with a reference formulation, the pessary, *Eur. J. Contracept. Reprod. Health Care*, 5, 61–67.
350. Coggins, C., Elias, C. J., Atisook, R., Bassett, M. T., Ettiegnene-Traore, V., Ghys, P. D., Jenkins-Woelk, L., Thongkrajai, E., and VanDevanter, N. L. (1998), Women's preferences regarding the formulation of over-the-counter vaginal spermicides, *AIDS*, 12, 1389–1391.
351. Hardy, E., Jimenez, A. L., de Padua, K. S., and Zaneveld, L. J. (1998), Women's preferences for vaginal antimicrobial contraceptives. III. Choice of a formulation, applicator, and packaging, *Contraception*, 58, 245–249.
352. Novak, A., de la Loge, C., Abetz, L., and van der Meulen, E. A. (2003), The combined contraceptive vaginal ring, NuvaRing: An international study of user acceptability, *Contraception*, 67, 187–194.
353. Roumen, F. J., and Dieben, T. O. (1999), Clinical acceptability of an ethylene-vinyl-acetate nonmedicated vaginal ring, *Contraception*, 59, 59–62.
354. Hardy, E., de Padua, K. S., Jimenez, A. L., and Zaneveld, L. J. (1998), Women's preferences for vaginal antimicrobial contraceptives. II. Preferred characteristics according to women's age and socioeconomic status, *Contraception*, 58, 239–244.
355. Rothen-Weinhold, A., Dahn, M., and Gurny, R. (2000), Formulation and technology aspects of controlled drug delivery in animals, *Pharm. Sci. Technol. Today*, 3, 222–231.
356. Robinson, T. J. (1965), Use of progestagen-impregnated sponges inserted intravaginally or subcutaneously for the control of the oestrous cycle in the sheep, *Nature*, 206, 39–41.
357. Kim, S., Kengaku, K., Tanaka, T., and Kamomae, H. (2004), The therapeutic effects of a progesterone-releasing intravaginal device (PRID) with attached estradiol capsule on ovarian quiescence and cystic ovarian disease in postpartum dairy cows, *J. Reprod. Dev.*, 50, 341–348.
358. Rathbone, M. J., Macmillan, K. L., Bunt, C. R., and Burggraaf, S. (1997), Conceptual and commercially available intravaginal veterinary drug delivery systems, *Adv. Drug Deliv. Rev.*, 28, 363–392.

359. Rathbone, M. J., Macmillan, K. L., Inskeep, K., Burggraaf, S., and Bunt, C. R. (1998), Fertility regulation in cattle, *J. Controlled Release*, 54, 117–148.
360. Okura, N., Yamagishi, N., Naito, Y., Kanno, K., and Koiwa, M. (2004), Technical note: Vaginal absorption of 1,25(OH)2D3 in cattle, *J. Dairy Sci.*, 87, 2416–2419.
361. Otero, M. C., Morelli, L., and Nader-Macias, M. E. (2006), Probiotic properties of vaginal lactic acid bacteria to prevent metritis in cattle, *Lett. Appl. Microbiol.*, 43, 91–97.
362. Gavini, E., Sanna, V., Juliano, C., Bonferoni, M. C., and Giunchedi, P. (2002), Mucoadhesive vaginal tablets as veterinary delivery system for the controlled release of an antimicrobial drug, acriflavine, *AAPS PharmSciTech*, 3, E20.
363. Romano, J. E., and Benech, A. (1996), Effect of service and vaginal-cervical anesthesia on estrus duration in dairy goats, *Theriogenology*, 45, 691–696.
364. Loehr, B. I., Rankin, R., Pontarollo, R., King, T., Willson, P., Babiuk, L. A., and van Drunen Littel-van den Hurk, S. (2001), Suppository-mediated DNA immunization induces mucosal immunity against bovine herpesvirus-1 in cattle, *Virology*, 289, 327–333.
365. Grimmett, J. B., Hanlon, D. W., Duirs, G. F., and Jochle, W. (2002), A new intra-vaginal progesterone-releasing device (Cue-Mare™) for controlling the estrous cycle in mares, *Theriogenology*, 58, 585–587.
366. Cross, P. S., Kunnemeyer, R., Bunt, C. R., Carnegie, D. A., and Rathbone, M. J. (2004), Control, communication and monitoring of intravaginal drug delivery in dairy cows, *Int. J. Pharm.*, 282, 35–44.

# PHARMACEUTICAL PREFORMULATION: PHYSICOCHEMICAL PROPERTIES OF EXCIPIENTS AND POWDERS AND TABLET CHARACTERIZATION

BEOM-JIN LEE
*Kangwon National University, Chuncheon, Korea*

**Contents**

1 Introduction
2 Selection of Pharmaceutical Excipients
  2.1 Definitions and Goals
  2.2 Types of Pharmaceutical Excipients
  2.3 Characteristics of Pharmaceutical Excipients
  2.4 Selection Guideline of Pharmaceutical Excipients in Tablet Formulation
3 Drug–Excipient Compatibility
  3.1 Experimental Studies for Drug–Excipient Compatibility
  3.2 Analytical Methods for Drug–Excipient Compatibility
  3.3 Reaction Types and Stabilization Guidelines
4 Powder Characteristics
  4.1 Crystal Form and Habit
  4.2 Particle Size Distribution
  4.3 Flow Characteristics
  4.4 Density and Bulkiness
  4.5 Hygroscopicity
  4.6 Mixing
  4.7 Particle Size Reduction (Micronization and Milling)
  4.8 Compaction (Compressibility)
  4.9 Surface Area and Other Properties
5 Tablet Characterization
  5.1 Disintegration
  5.2 Dissolution

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

5.3 Weight Variation
5.4 Hardness or Breaking Strength
5.5 Friability
5.6 Content Uniformity
5.7 Tablet Thickness
5.8 Tablet Shape and Size
References

## 1 INTRODUCTION

Tablet is a major category of solid dosage forms which are widely used worldwide. Extensive information is required to prepare tablets with good quality at high standards. Based on preformulation studies, the optimal dosage forms are generally decided. When given orally, the solid dosage form tablet undergoes in vitro disintegration and dissolution followed by absorption through the gastrointestinal tract (GI). The in vivo biodistribution of drug which enters the systemic circulation then occurrs (Figure 1).

In this section, general preformulation approaches for tablet production are described (Figure 2). The physicochemical properties of drug and excipients, which are crucial factors at the beginning stages, are presented. The types and functions of excipients used for tablet formulation and drug–excipient incompatibility are also discussed. Because tablets are prepared by compression of the drug with powdered excipients, the physicochemical properties of excipients and their multiple functions under regulatory standards are very important. To prepare a drug substance for the final dosage forms, pharmaceutical excipients should be added. The guidelines for selection of excipients are also given, based on the type and function of excipient, drug–excipient compatibility, type of tablet, and manufacturing parameters. Drug–excipient incompatibility is a very important issue before and after tablet prepara-

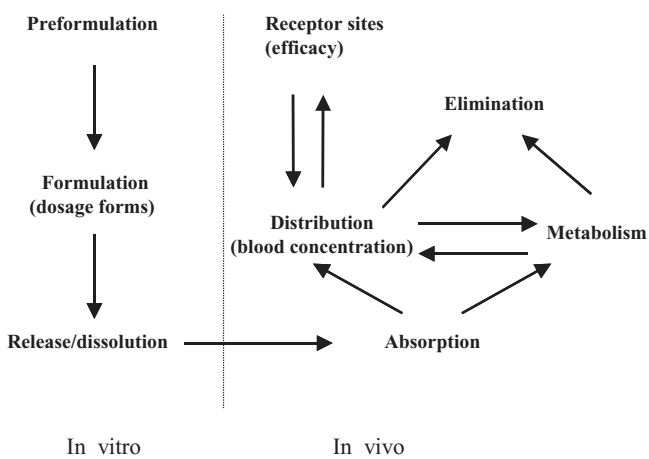


**FIGURE 1** In vitro and in vivo pathways of drug which enters systemic circulation.

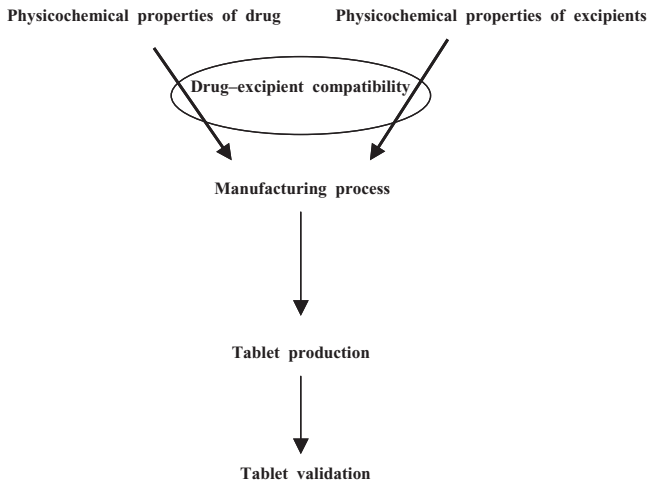


**FIGURE 2** General preformulation approaches for tablet production.

tion. According to global International Conference on Harmonisation (ICH) guidelines, we should cautiously choose the excipients in tablet formulations because the quality and sources are variable by supplier and by batch.

In general, a tablet is prepared by mixing, milling, and compression the drug–excipient mixtures using a tablet machine. Therefore, powder characteristics of drug and excipients, including particle size, flow characteristics, bulk density, hygroscopicity, mixing and milling, and compaction, are extensively discussed. If these physical properties are not fully understood, tablets with good quality at high standards are often impossible to produce.

The prepared tablet must be validated according to regulatory guidelines. In general, disintegration, dissolution, friability, hardness, and weight are characterized for quality validation of finished tablet. This biopharmaceutical preformulation information for tablet production will guide us to design the optimal tablet very efficiently in the laboratory and industrial companies.

## 2 SELECTION OF PHARMACEUTICAL EXCIPIENTS

### 2.1 Definitions and Goals

A tablet contains active ingredients as well as other substances known as excipients, which have specific functions. The types and functions of various excipients which are incorporated into tablet formulations are discussed in many textbooks [1–3].

A pharmaceutical excipient is defined as an inactive ingredient or any component other than the active ingredient added intentionally to the medicinal formulation or everything in the formulation except the active drug. Pharmaceutical excipients are also called additives, pharmaceutical ingredients, or inactive pharmaceutical ingredients. There are many reasons for selecting and adding these pharmaceutical excipients in formulations. In the preparation of various dosage forms, it is essential to combine pharmaceutical excipients with model drugs as adjuvants to prepare the

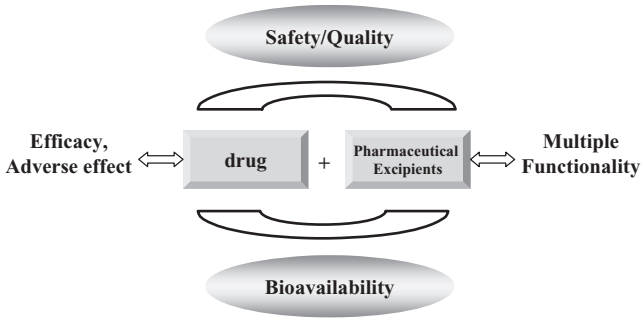


**FIGURE 3** Correlation and functions of drug combined with pharmaceutical excipients in dosage form design.

solid dosage forms, mainly tablets. The pharmaceutical excipients make the drug into the final dosage forms. Physicochemical properties such as solubility, stability, metabolism, and even bioavailability of drugs can be varied by the pharmaceutical excipients [4–6]. Figure 3 shows the correlation and functions of drugs combined with pharmaceutical excipients in dosage from designs. Pharmaceutical excipients are regarded as key ingredients not only to decide optimal dosage forms but also to change the physicochemical and biological parameters of drugs. With an aid of pharmaceutical excipients, drug efficacy can be maintained. Changes of other types of dosage forms for different routes of administration are also achieved. The excipients can also function for the preparation of dosage formulation during the manufacturing processes. Patient compliance and modified releases of drugs can also be achieved if the excipients are properly applied. The colorant makes the tablet distinguishable after the coating process.

However, utilization of these pharmaceutical excipients is limited by the regulatory guidelines to be satisfied in the dosage formulations [3, 7]. In general, the regulatory guidelines require the following conditions for the use of excipients in the dosage formulations: (a) no harmful or toxicological effect and listed GRAS (generally recognized as safe), (b) good stability with no drug–excipient incompatibility and by any impurities in the excipients, (c) no interference in quality validation and analytical tests, (d) satisfaction of regulatory issues and guidelines in all countries where the product is to be marketed, (e) no instability with primary packing materials, (f) ease of accessibility, distribution, and economical cost, (g) satisfaction for environmental issues, (h) be physiologically inert, (i) be physically and chemically compatible with the active substance and the other excipients in the formulation, and (j) no unacceptable microbiological burden.

The *Handbook of Pharmaceutical Excipients* [1] contains some details of functional tests carried out on a wide range of excipients. The excipients all have pharmacopeial monographs, but it is important to understand that compliance with a monograph does not indicate equivalence between different grades or suppliers.

## 2.2 Types of Pharmaceutical Excipients

To prepare a drug substance into a final dosage form, pharmaceutical excipients should be added. The *Handbook of Pharmaceutical Excipients* presents more than

230 monographs of excipients used in the dosage formulations [1]. The amount and type of pharmaceutical excipients are highly dependent on the final dosage forms. In the preparation of tablets, diluents or fillers are commonly added to increase the bulk of the formulation, binders to cause the adhesion of the powdered drug and pharmaceutical substances, antiadherents or lubricants to assist in the smooth tableting process, disintegrating agents to promote tablet break-up after administration, and coating agents to improve stability, control disintegration, or enhance appearance. Thus, the pharmaceutical excipients establish the primary features and physicochemical properties of the tablet, such as the physical form, stability, dissolution, taste, and overall appearance [3]. Table 1 presents examples of pharmaceutical excipients used in tablet formulations according to principal categories.

### 2.3 Characteristics of Pharmaceutical Excipients

***Filler (Diluent)*** Powder mixtures should be compacted to achieve the appropriate strength at a low compaction pressure in the tablet preparation. The compaction properties of a formulation will largely be governed by its major components. For a high-dose drug the drug itself will strongly influence the compaction, while for low-dose drugs the bulk size needs to be increased with an inactive ingredient

**TABLE 1 Pharmaceutical Excipients Used in Tablet Formulations**

| Excipient Type | Definition | Examples |
|---|---|---|
| Adsorbent | Agent capable of holding other molecules onto its surface by physical or chemical (chemisorption) means | Powdered cellulose, activated charcoal |
| Antioxidant | Agent that inhibits oxidation and thus is used to prevent deterioration of preparations by oxidative process | Ascorbic acid, ascorbyl palmitate, butylated hydroxyanisole, butylated hydroxytoluene, hypophosphorus acid, monothioglycerol, propyl gallate, sodium ascorbate, sodium bisulfite, sodium formaldehyde, sulfoxylate, sodium metabisulfite |
| Colorant | Used to impart color to tablet | FD&C red no. 3, no. 20, FD&C yellow no. 6, FD&C blue no. 2, D&C green no. 5, D&C orange no. 5, D&C red no. 8, caramel, ferric oxide, red |
| Encapsulant | Used to form thin shells for purpose of enclosing drug substance or drug formulation for ease of administration | Gelatin, cellulose acetate phthalate |
| Plasticizer | Used as component of film coating solutions to enhance spread of coat over tablets, beads, and granules | Diethyl phthalate, glycerin |

**TABLE 1** ***Continued***

| Excipient Type | Definition | Examples |
|---|---|---|
| Surfactant | Substance that adsorbs to surfaces or interfaces to reduce surface or interfacial tension; may be used as wetting agent, detergent, or emulsifying agents | Benzalkonium chloride, nonoxynol 10, oxtoxynol 9, polysorbate 80, sodium lauryl sulfate, sorbitan monopalmitate |
| Tablet antiadherent | Agent that prevents the sticking of tablet formulation ingredients to punches and dies during tablet production | Magnesium stearate, talc |
| Tablet binder | Substance used to cause adhesion of powder particles in tablet granulations | Acacia, alginic acid, carboxymethylcellulose sodium compressible sugar, ethylcellulose gelatin, liquid glucose, metylcellulose povidone, pregelatinized starch |
| Tablet diluent | Inert substance used as filler to create desired bulk, flow properties, and compression characteristics in preparation of tablets | Dibasic calcium phosphate, kaolin, lactose, mannitol, microcrystalline cellulose, powdered cellulose, precipitated calcim carbonate, sorbitol, starch |
| Tablet sugar and film coating excipient | Used to coat a formed tablet for purpose of protecting against drug decomposition by atmospheric oxygen or humidity, to provide desired release pattern for drug substance after administration, to mask taste or odor of drug substance, or for aesthetic purposes | Sugar: liquid glucose, sucroseFilm: hydroxyethyl cellulose, hydroxypropyl methylcellulose, methylcellulose, ethylcellulose<br>Enteric: cellulose acetate phthalate, shellac (35% in alcohol) |
| Tablet direct-compression excipient | Used in direct-compression tablet formulations | Dibasic calcium phosphate |
| Tablet disintegrant | Used in solid dosage forms to promote disruption of solid mass into smaller particles which are more readily dispersed or dissolved | Alginic acid, carboxymethylcellulose calcium, microcrystalline cellulose, polacrilin potassium, sodium alginate, sodium starch glycollate, starch |
| Tablet glidant | Agent used in tablet and capsule formulations to improve flow properties of powder mixture | Colloidal silica, corn starch, talc |
| Tablet lubricant | Substance used in tablet formulations to reduce friction during tablet compression | Calcium stearate, magnesium stearate, mineral oil, stearic acid, zinc stearate |
| Tablet opaquant | Used to render tablet coating opaque; may be used alone or in combination with colorant | Titanium dioxide |
| Tablet-polishing agent | Used to impart an attractive sheen to coated tablets | Carnauba wax, white wax |

**TABLE 2 Commonly Used Tablet Diluents**

| Diluent | Comments |
|---|---|
| Lactose | Available as anhydrous and monohydrate; anhydrous material used for direct compression due to superior compressibility |
| Microcrystalline cellulose | Originally direct-compression excipient, now often included in granulations due to its excellent compressibility |
| Dextrose, glucose | Direct-compression diluent, often used in chewable tablets |
| Sucrose | Was widely used as sweetener/filler in effervescent tablets and chewable tablets; less popular nowadays due to cariogenicity |
| Starch and derivatives | Versatile material that can be used as diluent binder, and disintegtant |
| Calcium carbonate | Brittle material |
| Dicalcium phosphate | Excellent flow properties; brittle material |
| Magnesium carbonate | Direct-compression diluent |

termed a diluent (or filler). High-dose formulations may also use a diluent to overcome compaction problems experienced with an active pharmaceutical substance.

There are a number of general rules for selecting a diluent. The selection of the diluent will depend on the type of processing and plasticity of materials to be used. A direct-compression formulation will require a diluent with good flow and compaction properties. If the material is extremely plastic, it is appropriated to add a diluent that compacts by brittle fracture; similarly, a brittle drug substance should be combined with a plastic filler. The solubility of the drug substance should also be considered. A soluble drug is normally formulated with an insoluble filler to optimize the disintegration and dissolution process. The hydrophilic excipients added in the formulation may also change drug solubility. Table 2 lists the more commonly used diluents in tablet formulation.

***Binder (Granulating Agent)*** Before tableting the powder mixture via direct compression, generally powders are granulated simply by adding water or an organic solvent to form liquid bridges followed by the drying process. This granulation process can make powders of larger particle size and that are more free flowing for tablet production. The most common method of adding binders is as a solution in the granulating fluid. It is also possible add synthetic polymers such as PVP and HPMC as powders and use water as the granulating agent. When the granulate dries, the crystallization of any solids that had dissolved in the liquid will form solid bonds between the particles [8]. Inclusion of granulating agents or binders to increase granule strength is necessary. Granulating agents are usually hydrophilic polymers that have cohesive properties that both aid the granulation process and impart strength to the dried granulate.

For a granulating agent to be effective, it must form a film on the particle surface and be selected on the basis of its spreading coefficient. The spreading coefficient is defined as the difference between the *work of adhesion* of the binder and the substrate and the *work of cohesion* of the binder. Commonly used granulating agents are listed in Table 3. The binder may vary the disintegration and dissolution. Binders form hydrophilic films on the surface of the granules, which can aid in the

**TABLE 3 Commonly Used Granulating (Binding) Agents**

| Granulating Agent | Normal Concentration (%) | Comments |
|---|---|---|
| Starch | 5–25 | Was once the most commonly used binder; starch has to be prepared as paste, which is time consuming |
| Pregelatinized starch | 0.1–0.5 | Cold-water soluble so easier to prepare than starch |
| Acacia | 1–5 | Requires preparation of past prior to use; can lead to prolonged disintegration times if used at too high a concentration |
| Polyvinylpyrrolidone (PVP) | 2–8 | Available in range of molecular weight/ viscosities; soluble in water and ethanol |
| Hydroxypropyl methylcellulose (HPMC) | 2–8 | Low-viscosity grades most widely used |
| Methylcellulose (MC) | 1–5 | — |

wetting of hydrophobic drugs. However, if added at too great concentrations, the films can form viscous gels on the granule surface and will retard dissolution.

***Disintegrant*** Tablets must have sufficient strength to withstand the stresses of subsequent manufacturing operations, such as the coating, packaging, and distribution process. However, once the tablet is taken by the patient, it must break up rapidly to ensure rapid dissolution of the active ingredient in immediate-release preparations. To overcome the cohesive strength produced by the compression process and to break down the tablet into the primary particles as rapidly as possible, the disintegrants are combined with other excipients during the tableting process. Starch was the first disintegrant used in tablet manufacture. Recently, so-called superdisintegrants, including croscarmellose sodium, sodium starch glycolate, and crospovidone, display excellent disintegration activity at low concentrations and have better compression properties than starches. Traditionally, swelling and rate of swelling have been regarded as the most important characteristics of disintegrants. With the aid of these superdisintegrants, sustained-release acetaminophen tablets with biphasic patterns were successfully established to mimic the bilayered Tylenol ER tablet [9]. As a general rule, soluble drugs are formulated with insoluble fillers to maximize the effect of disintegrants. The positioning of disintegrants within the intragranular and extragranular portions of granulated formulations can affect their water uptake and disintegration time.

Commonly used disintegrants are listed in Table 4. The greater the level of disintegrant, the faster the tablet will disintegrate. The compaction properties of many disintegrants, including starch, are not satisfactory and use of high concentration could also reduce tablet strength. Disintegrants are hygroscopic materials and will absorb moisture from the atmosphere, which could negatively affect the stability of moisture-sensitive drugs if the packaging does not provide adequate protection from the environment. Disintegrant activity can be affected by mixing with hydrophobic lubricants so that care needs to be taken to optimize the manufacturing

**TABLE 4 Commonly Used Disintegrants**

| Disintegrant | Normal Concentration (%) | Comments |
|---|---|---|
| Starch | 5–10 | Probably works by wicking; swelling minimal at body temperature |
| Microcrystalline cellulose | | Strong wicking action; loses disintegrant action when highly compressed |
| Insoluble ion exchange resins | | Strong wicking tendencies with some swelling action |
| Sodium starch glycolate[a] | 2–8 | Free-flowing powder that swells rapidly on contact with water |
| Croscarmellose sodium[a] | 1–5 | Swells on contact with water |
| Gums—agar, guar, xanthan | <5 | Swell on contact with water; form viscous gels that can retard dissolution, thus limiting concentration that can be used |
| Alginic acid, sodium alginate | 4–6 | Swell like gums but form less viscous gels |
| Crospovidone[a] | 1–5 | High wicking activity |

[a]Often mentioned as superdisintegrant.

process as well as the formulation. If the tablet contains a high proportion of a hydrophobic drug that has a high contact angle, a wetting agent or surfactant should be added to the formulation to modify disintegration time and subsequent dissolution of the drug from tablet. The most commonly used wetting agents are sodium lauryl sulfate and the polysorbates.

Most pharmacopeias include a disintegration test which can be applied to tablets and capsules and the detailed monograph is given in the pharmacopeias (see Section 5.1).

***Lubricant*** The use of a lubricant is essential to increase the free flow of powders and to prevent manufacturing disorders in the tablet production. The type and amount of lubricant are cautiously selected in the formulation. The order of addition and mixing time is also considered in the tableting process.

There are three types of lubricants employed in solid dosage form manufacture. The first class of lubricant is the *glidant*. The flow properties of a powder can be enhanced by the inclusion of a glidant. These are added to overcome powder cohesiveness. The two other classes of lubricant are *antiadherent* excipients, which reduce the friction between the tablet punch faces and tablet punches, and *die wall lubricant* excipients, which reduce the friction between the tablet surface and the die wall during and after compaction to enable easy ejection of the tablet. The level of a lubricant required in a tablet is formulation dependent and can be optimized using an instrumented tableting machine.

Commonly used lubricants are listed in Table 5. Talc is traditionally one of the most commonly used glidants, having the additional benefit of being an excellent antiadherent. The level of talc that can be added to a formulation is restricted by its hydrophobic nature, too high levels resulting in decreased wetting of the tablet and a subsequent reduction in the rate of dissolution. Fumed silicon dioxides are

**TABLE 5 Lubricants Commonly Used in Formulations**

| Lubricant | Typical Percent | Comments |
|---|---|---|
| | | *Glidants* |
| Talc | 1–5 | Fine, crystalline powder; Widely used as lubricant and diluent |
| Fumed silicon dioxide: Aerosil, Cab-O-Sil, Syloid | 0.1–0.5 | Has small particle size and large surface area for good flowability; used for adsorbent, antitacking agent disintegrant, and glidant |
| Starch | 1–10 | Mainly used for binder, disintegrant, and diluent but also used for glidant |
| Sodium lauryl sulfate | 0.2–2 | Anionic surfactant, luricant and wetting agent |
| | | *Boundary Lubricants* |
| Magnesium stearate | 0.2–2 | Hydrophobic, variable properties between suppliers |
| Calcium silicate | 0.5–4 | Hydrophobic |
| Sodium stearyl fumarate | 0.5–2 | Less hydrophobic than metallic stearates, partially soluble |
| Polyethylene glycol 4000 and 6000 | 2–20 | Soluble, poorer lubricant activity than fatty acid ester salts |
| Sodium lauryl sulfate | 1–3 | Soluble, also acts as wetting agent |
| Magnesium lauryl sulfate | 1–3 | Acts as wetting agent |
| Sodium benzoate | 2–5 | Soluble |
| | | *Fluid Lubricants* |
| Light mineral oil | 1–3 | Hydrophobic, can be applied to either formulation or tooling |
| Hydrogenated vegetable oil | 1–5 | Hydrophobic, used at higher concentrations as controlled-release agents |
| Stearic acid | 0.25–2 | Hydrophobic |
| Glyceryl behenate | 0.5–4 | Hydrophobic, also used as controlled-release agent |

perhaps the most effective glidants. These are materials with very small (10-nm) spherical particles that may achieve their glidant properties. They are available in a number of grades with a range of hydrophobic and hydrophilic forms and also commercially available under diverse brand names. Starch has also been used as a glidant. The use of large amounts of starch has also aided the disintegration properties.

Die-wall lubricants can be dived into two classes, *fluid and boundary lubricants*. Fluid lubricants work by separating moving surfaces completely with a layer of lubricant. These are typically mineral oils or vegetable oils, and they may be either added to the mix or applied directly to the die wall by means of wicked punches. The oily lubricants may have a mottled appearance in the tablet due to uneven distribution, poor powder flow due to their tacky nature, and reduced tablet strength. Fluid lubricants include stearic acid, mineral oils, hydrogenated vegetable oils, glyceryl behenate, paraffins, and waxes. Boundary lubricants work by forming a thin solid film at the interface of the die and the tablet. Metallic stearates are the most

widely used boundary lubricants. Such lubricants should have low shear strength and form interparticulate films.

Magnesium stearate is the most widely used lubricant. The magnesium stearate used in the pharmaceutical industry is not a pure substance but a mixture of magnesium salts of fatty acids, though predominantly magnesium stearate and magnesium palmitate. Despite its popularity, which is a reflection of its excellent lubricant properties, it has some problems associated with product consistency: its effect on tablet strength and its hydrophobicity. The U.S. Pharmacopeia (USP) requires that the stearate content should account for not less than 40% of the fatty acid content of the material, and the stearate and palmitate combined should account for not less than 90%. Within this definition, there are a range of materials to be supplied as magnesium stearate. For a given formulation, it is important that a single source of magnesium stearate be used for all batches to get product reliability.

The lubricant activity of magnesium stearate is related to its readiness to form films on the die wall surface. As a result, it has two consequences: a reduction in the ability of the powder to form strong compacts and, due to its hydrophobicity, a deleterious effect on the dissolution rate of the tablets. The hydrophobic surfaces created by magnesium stearate have been shown to reduce the rate of dissolution and bioavailability of several tablet formulations. When both lubricant and disintegrant are being added to a granulated formulation, the disintegrant should be blended with the granules prior to the addition of the lubricant to minimize the risk of forming a hydrophobic film around the disintegrant.

The third class of lubricant activity is the antiadherent. Some materials have adhesive properties and can adhere to the punch surfaces during compression. This will induce tablet disorders: sticking, with a film forming on the surface of the tablets, or picking, where solid particles from the tablet stick to the punch surface. Most die wall lubricants also have antiadherent actions, and in many formulations, the addition of a specific antiadherent will not be required separately. The antiadherent includes talc, maize starch, and microcrystalline cellulose.

***Coating Materials*** The core compressed tablet can be used by itself, but are additional coating process of the compressed tablet can be applied for several reasons: (a) protection of the drug from the environment (moisture, air, light) for stability reasons, (b) taste masking, (c) minimizing patient/operator contact with drug substance, particularly for skin sensitizers, (d) improving product identity and appearance, (e) improving ease of swallowing, (f) improving mechanical resistance and reducing abrasion and attrition during handling, and mostly (g) modifying release properties.

There are three main methods used to coat pharmaceutical tablets: *sugar coating, film coating*, and *compression coating*. Sugar coating has been the most commonly used method and involves coating tablets with sucrose. A sugar-coated tablet is water based and generally starts to break up in the stomach. This is a highly skilled and multistep process that is very labour intensive. This coating process results in a 50% increase of the final tablet weight and in a significant increase in tablet size. Traditionally, sugar coating has been performed in coating pans in which the tablets are tumbled in a three-dimensional direction. The pan is supplied with a source of warm air for drying and an extraction system to remove moist air and dust. The coating solution is distributed around the tablets by their tumbling action. A dusting

powder may be sprinkled onto the surface of the tablets during the drying phase to prevent the tablets from sticking together. The cycle of wetting and drying is continued until the desired amount of coating has been applied to the tablets. Typically, a sugar coating will consist of three types of coats: a sealing coat, a subcoat, and a smoothing coat. Traditionally, a seal coating of shellac dissolved in ethanol or synthetic water-resistant polymers such as cellulose acetate phthalate or polyvinylacetate phthalate is used. The subcoat is an adhesive coat on which the smoothing coat of the sharp corners of the tablet can be built. The subcoat is a mixture of a sucrose solution and an adhesive gum, such as acacia or gelatin, which rapidly distributes over the tablet surface. A dusting follows each application of solution with a subcoat powder containing materials such as calcium carbonate, calcium sulfate, acacia, talc, and kaolin. The smoothing coat consists of the majority of the tablet bulk and provides the tablet with a smooth finish. A colorant is also applied if needed. The coat consists of sucrose syrup which may contain starch or calcium carbonate. The coated tablets are usually transferred to a polishing pan and coated with a beeswax–carnauba wax mixture to provide a glossy finish to the surface.

Film coating involves the application of a polymer film to the surface of the tablet, gelatin capsules, and multiparticulate systems with a negligible increase in tablet size. The method of application of the coat differs from the sugar coating in that the coating suspension is sprayed directly onto the surface of the tablets, and drying occurs as soon as the coat hits the tablet surface. The tablet only receives a small quantity of coating solution at a time. The film coat can generally be affected by the following properties: a method of atomizing the coating suspension, the ability to heat large volumes of air (which heat the tablets and facilitate the rapid drying of the applied coat), and a method of moving the tablets that ensures all tablets are evenly sprayed.

The main methods of coating are modified conventional coating pans, side-vented pans, and fluid bed coating. The side-vented pan, now the most commonly used equipment for film coating, was designed to maximize the interaction between the tablet bed and the drying air. The mixing efficiency of the table, granules, or pellets is achieved by using appropriately designed baffles on the pan surface. Fluid-bed coating offers an alternative to pan coating and is particularly popular for coating multiparticulate systems. There are three methods by which the coating can be applied: top spraying, bottom spraying, and tangential coating. The pellets, granules, or tablets being coated are suspoended in an upward stream of air, maximizing the surface available for coating. The coating is applied by an atomizer, and this is dried by the fluidizing air.

Table 6 gives commonly used polymers for film coating of core tablet. With the exception of HPMC, the polymers are rarely used alone but are combined with other polymers to optimize the film-forming properties. A polymer for film coating will ideally meet the following criteria:

1. Solubility in the solvent selected for application: Currently the organic solvent are replaced with water as a suspension system, although certain types of film coatings may require organic solvents to be used. Commonly used solvents include alcohols (methanol, ethanol, and isopropanol), esters (ethyl acetate and ethyl lactate), ketones (acetone), and chlorinated hydrocarbons (dichloromethane and trichloroethane).

**TABLE 6 Commonly Used Polymers for Film Coating of Core Tablet**

| Polymer | Comments |
|---|---|
| Methylcellulose (MC) | Soluble in cold water, GI fluids, and a range of organic solvents |
| Ethylcellulose (EC) | Soluble in organic solvents, insoluble in water and GI fluids; used alone in modified-release formulations and in combination with water-soluble cellulose for immediate-release formulations |
| Hydroxyethylcellulose (HEC) | Soluble in water and GI fluids |
| Methyl hydroxyethylcellulose (MHEC) | Soluble in water and GI fluids; has similar film-forming properties to HPMC but is less soluble in organic solvents, which limited its popularity when solvent coating was the norm |
| Hydroxypropyl cellulose (HPC) | Soluble in cold water, GI fluids, and polar solvents; becomes tacky when dried, so is unsuitable for use alone, often used in combination with other polymers to optimize adhesion of coat |
| Hydroxypropyl methylcellulose (HPMC) | Soluble in cold water, GI fluids, alcohols, and halogenated hydrocarbons; excellent film former and the most widely used polymer; can be used with lactose to improve adhesiveness |
| Sodium carboxymethylcellulose (NaCMC) | Soluble in water and polar solvents |

2. Solubility in GI fluids: The solubility of polymers is dependent on its physicochemical nature and pH. Unless the coating is being applied for enteric coating, it should ideally be soluble across the range of pH values encountered in the GI tract.
3. Capacity to produce an elegant film even in the presence of additives such as plasticizers, pigments, and colorants.
4. Compatibility with film-coating additives and the tablet being coated.
5. Stability in the environment under normal storage conditions.
6. Freedom from undesirable taste or odor.
7. Lack of toxicity.

Enteric coating materials are also used to prevent release of the drug substance in the stomach if the drug is either an irritant to the gastric mucosa or unstable in gastric juice. Table 7 lists enteric coating polymers commonly used in tablet formulations. The choice of of enteric coating material depends on its solubility.

The third type of tablet coating is multiple-compression coating to make a bilayered, multilayered tablet (a layered tablet of two drugs) or a tablet within a tablet (a core of one drug and a shell of another). The multilayered tablet is prepared by initial compaction of a portion of the fill materials in a die followed by additional fill material and compression, depending on the number of fill materials. The layered tablet can provide some advantages. Each layer has different drug in a separate layer. The incompatible drug can be compressed simultaneously at different layers.

**TABLE 7 Enteric Coating Polymers Commonly Used in Tablet Formulations**

| Polymer | Solubility Profile | Comments |
|---|---|---|
| Shellac | Above pH 7 | Original enteric coating material, originally used in sugar-coated tablets; high pH required for dissolution may delay drug release; natural product which exhibits batch-to-batch variability |
| Cellulose acetate phthalate (CAP) | Above pH 6 | High pH required for dissolution a disadvantage; forms brittle films, so must be combined with other polymers |
| Polyvinylacetate phthalate (PVAP) | Above pH 5 | — |
| Hydroxypropyl methylcellulose phthalate (HPMCP) | Above pH 4.5 | Optimal dissolution profile for enteric coating |
| Polymers of methacrylic acid and its esters | Various grades available with dissolution occurring above pH 6 | — |

The staged release and improved appearance from the layered tablet are also possible. In the preparation of a tablet within a tablet, a special tableting machine is required to place the intended core tablet precisely within the die, which already contains some of the coating formulation to surround the fill materials. This multilayered tablet or tablet within a tablet is also useful to get the modified release, either immediately or in a sustained manner.

***Auxiliary Excipients*** Plasticizers are added during the tablet coating process. The film coating process involves two important stages: droplet formation and film formation. Film formation is a multistage process that involves wetting the tablet surface followed by spreading the film and eventually coalescence of the individual film particles into a continuous film. Most film-forming polymers have glass transition temperatures in excess of the temperatures reached during the coating process (typically 40–50 °C), so it is necessary to add plasticizers to the formulations which reduce the glass transition temperature. The choice of plasticizer is dependent on the type of polymer and its permanence and compatibility. Permanence is the duration of the plasticizer effect; the plasticizer should remain within the polymer film to retain its effect, so it should have a low vapor pressure and diffusion rate. Compatibility requires the plasticizer to be miscible with the polymer. Commonly used plasticizers include phthalate esters, citrate esters, triacetin, propylene glycol, polyethylene glycols (PEGs), and glycerol.

The pigments or opacifiers are also combined with the coating solution to get colored tablets. Insoluble pigments are normally preferred to soluble dyes for a number of reasons. Solid pigments produce a more opaque coat than dyes, protecting the tablet from light. The presence of insoluble particles in the suspension allows

the rate of solid application to the tablet to be increased without having an adverse effect on the viscosity of the coating suspension, improving productivity.

### 2.4 Selection Guideline of Pharmaceutical Excipients in Tablet Formulation

In the proper selection of pharmaceutical excipients for the formulation, there are numerous factors to be considered [7]. The type of excipient is highly dependent on the model compound and its intended dosage form. The preparation method of the dosage form and proper dosage regime are also factors.

The formulation parameters of the tablet are essential. The amount and type of excipients relative to drug contents should be considered to determine the size of the dosage form. The bulk density of excipients and drug and their filling doses in the die are important factors to be considered in the tablet formulation. Uniform distribution of drug or excipient is in trouble if its contents are low in the formulation, while the use of high contents are difficult to fill in the die and the tablet size is also larger. Excipients that are potentially incompatible with the drug should be avoided. The amount and type of reactive impurities in the excipients should also be established. Batch-to-batch and supplier-to-supplier variation in impurity levels is possible. Excipients with potential adverse interaction and any unwanted incompatibility with drug should be avoided in the formulation. For an ideal formulation, it is helpful to consider these parameters as well as drug–excipient incompatibility.

The global harmonization and standardization of pharmaceutical excipients are nowadays necessary in the formulation studies [3, 7]. Since all excipients in the market worldwide are supplied from many countries or companies with facilities in more than a single country, the quality of the excipients must be well documented and validated. Otherwise, the quality of the excipients should be varied batch to batch, factory to factory, and country to country. There are also hundreds of different brands and grades of excipients available, but it would be unrealistic for the formulator to expect to have a totally free choice. Most manufacturing companies select the excipients used in their factory on the basis of cost, availability, and performance. To establish the equivalency of excipients obtained from different sources, it is also necessary to perform some kind of functionality testing. Since the sources, origin, and manufacturer of the excipients are different, regulatory approval for the product is required in each country. The standards for each drug substance and excipient are contained in pharmacopeia. The four pharmacopeia with the largest international use are the United States Pharmacopeia (USP) and National Formulary (NF), British Pharmacopoeia (BP), European Pharmacopoeia (EP), and Japanese Pharmacopoeia (JP).

Unless global harmonization is established, analytical methods, testing criteria, and specification limits must be variable according to monographs of pharmacopeia from the different countries. For example, Table 8 lists the different specification limits for the viscosity of cellulose ether among three pharmacopeias. If global harmonization and standardization of pharmaceutical excipients are established, the marketing and sales of a single formulation are more facilitated worldwide. The manufacturing cost and research-and-development (R&D) cycles are also reduced. Most of all, the quality and bioequivalence of the drug products with the same formulation can be more validated since the regulatory approval of the pharmaceutical product is enhanced. The global harmonization is an ongoing effort by corporate representatives and international regulatory authorities.

**TABLE 8 Specification Limits for Viscosity of Cellulose Ether**

| Parameter | EP | USP XXII | JP XII |
|---|---|---|---|
| Concentration of solution, % | 2.0 | 2.0 | 2.0 |
| Temperature, °C | 20 ± 0.1 | 20 ± 0.1 | 20 ± 0.1 |
| Type of viscometer | Rotating viscometer | Capillary type, Ubbelohde | Capillary type, Ubbelohde |
| Shear rate, $S^{-1}$ | 10 | — | — |
| Unit | cP (Pa·s) | cP (Pa·s) | cSt |
| Sample preparation[a] | | | |

[a]Very similar in all three pharmacopeia.

**TABLE 9 Comparative Properties of Some Directly Compressible Fillers[a]**

| Filler | Compactibility | Flowability | Solubility | Disintegration | Hygroscopicity | Lubricity | Stability |
|---|---|---|---|---|---|---|---|
| Dextrose | 3 | 2 | 4 | 2 | 1 | 2 | 3 |
| Spray dried lactose | 3 | 5 | 4 | 3 | 1 | 2 | 4 |
| Fast-Flo lactose | 4 | 4 | 4 | 4 | 1 | 2 | 4 |
| Anhydrous lactose | 2 | 3 | 4 | 4 | 5 | 2 | 4 |
| Emdex (dextrates) | 5 | 4 | 5 | 3 | 1 | 2 | 3 |
| Sucrose | 4 | 3 | 5 | 4 | 4 | 1 | 4 |
| Starch | 2 | 1 | 0 | 4 | 3 | 3 | 3 |
| Starch 1500 | 3 | 2 | 2 | 4 | 3 | 2 | 4 |
| Dicalcium phosphate | 3 | 4 | 1 | 2 | 1 | 2 | 5 |
| Avicel (MCC) | 5 | 1 | 0 | 2 | 2 | 4 | 5 |

[a]Graded on a scale from 5 (good/high) to 1 (poor/low); 0 means none.

In the selection of excipients, formulation scientists should also remember that every excipient is not limited by a single function but rather can be used for many pharmaceutical applications in dosage forms. For example, cellulose and its derivatives (HPMC, EC, HPC) have been widely used as fillers, binders, suspending agents, and mainly controlled-release agents. Povidone, polymethacrylate, and cyclodextrins are also multifunctional excipients in formulations. Table 9 gives the physical properties of some directly compressible fillers. The grades of physical properties are variable among fillers used in tablet formulation. The simple and optimal formulation of drugs with multifunctional excipients can provide some advantages in tablet preparations.

## 3 DRUG–EXCIPIENT COMPATIBILITY

The potential for excipients to cause chemical and physical instability in drugs has been recognized for over 30 years. Drug compatibility studies have been used as an

approach for accepting/rejecting excipients for use in pharmaceutical formulations [1, 10–12]. In general, drug stability can be investigated under the stress condition according to the guideline of accelerated stability testing. The factors for stress condition usually include temperature, pH, light, moisture, agitation, gravity, packaging, and method of manufacture [13, 14]. Despite this fact, approaches for excipient selection in drug formulation are often empirical. The stability issues in the development of a drug must be considered at the early and late formulation stages. Nowadays, advanced analytical instrumentation makes it possible to more rapidly identify potential excipient-induced instability to select excipients. Excipients that exhibit incompatibility with the drug are "rejected" and not included in subsequent tablet formulation studies.

### 3.1 Experimental Studies for Drug–Excipients Compatibility

Compatibility studies are carried out by mixing drug with one or more excipients under some type of stress condition. It has also been suggested that aqueous suspensions of the drug and excipients or drug–excipient complexes provide a better model for tablet formulations. It has been recommended that small-scale formulations using the selected excipients (this may eventually be used for the eventual formulation) be prepared using experimental processes.

Table 10 provides examples of binary and factorial design for drug–excipient compatibility studies. The two conditions of each potential parameter can be included as a binary system. All potential parameters can also be combined for factorial design. For example, water is added to the drug–excipient mixture at reasonable temperatures (50 °C or less) and then both intact drug and degradation products are measured. The high-temperature and high-humidity conditions are usually used to obtain more rapid stability assessment of drug and excipients. Generally, physicochemical properties such as drug content, color, taste, related substances, thermal analysis, and high-performance liquid chromatography (PLC) studies of evaluated for drug–excipient compatibility studies using conventional analytical techniques.

### 3.2 Analytical Methods for Drug–Excipient Incompatibility

The key to the early assessment of instability in formulations is the availability of analytical methods to detect low levels of degradation products, generally less than 2%. With the aid of thermal analysis and chromatographic methods [HPLC and

**TABLE 10 Examples of Binary and Factorial Designs for Drug-Excipient Compatibility Studies**

| Binary Design | Factorial Design |
|---|---|
| Drug–excipient ratio 1:1 or 1:10 | Total number of formulation = $a \times b \times c \times d \times e$ |
| Water addition = yes/no | Type of drug–excipient ($a$) |
| Storing temperature 25 °C or 40 °C/75% | Amount of drug–excipient ($b$) |
| Storing time 1 or 4 weeks | Water addition ($c$) |
| | Storing temperature ($d$) |
| | Storing time ($e$) |

liquid chromatography/mass spectrometry (LC/MS)], it can assign at least tentative structures for most degradation products as well as the intact drug contents [15]. Excipient compatibility screening must provide more rapid identification of excipient-mediated instability that is detected in complete formulations.

Early compatibility studies relied on color change as an indication of incompatibility. Subsequently, there were numerous reports on the use of thermal analysis techniques such as differential thermal analysis (DTA) or differential scanning calorimetry (DSC) to detect drug–excipient incompatibilities. DTA and DSC have the advantage of rapid analysis. Generally, formation of new peaks and disappearance of drug peak by the endothermic or exothermic reaction are carefully investigated. Diffuse reflectance spectroscopy (DRS) can be also used to determine the ultraviolet (UV) absorption of drug on the surface of drug–excipient mixtures.

### 3.3 Reaction Types and Stabilization Guidelines

Understanding the degradation chemistry of drug with excipients is essential to select proper excipients in the formulation stages [16, pp 101–151]. Drug-excipient compatibility studies are crucial to decide optimal tablet formulation and to understand the possible mechanism in many cases [10, 12, 14]. Drug instability occurs by three types of reactions: hydrolysis, oxidation, and aldehyde–amine addition. Table 11 gives reaction types of chemical and physical instability.

Hydrolysis is the most common mechanism to induce drug–excipient incompatibility. Most hydrolysis reactions are catalyzed by acids and/or bases. The microscopic pH between drug and excipients is critical. The pH degradation rate profiles are good predictors for hydrolysis reactions in solid dosage forms. The pH of optimal stability in solution is similar to the "microscopic pH" in solid dosage forms. A shift

**TABLE 11 Reaction Types of Chemical and Physical Instability**

| Type of Instability | Order of Frequency |
|---|---|
| Chemical instability | |
| pH-dependent hydrolysis | 12 |
| Oxidation | |
| Air oxidation | 5 |
| Metal-catalyzed oxidation | 3 |
| Peroxides in excipients | 2 |
| Aldehyde–amine addition | |
| Aldose excipient | 2 |
| Formaldehyde from excipients | 1 |
| Michael addition with maleate salt | 1 |
| Intramolecular cyclization | 1 |
| Dimerization (Diels–Alder) | 1 |
| Racemization | 1 |
| Addition of ammonia residue from gelatin capsule | 1 |
| Esterification | 1 |
| Physical instability | |
| Evaporation of volatile free base | 2 |
| Reaction of methanesulfonic acid with disintegrants | 1 |
| Gelatin cross-linking by excipient impurities | 1 |

in only one pH unit can increase or decrease the reaction rate by a factor of 10. Although solid-state compatibility studies are commonly used, solution kinetic studies are generally conducted to select excipients since it is possible to study all three of these reactions in solution. Often solution kinetic studies allow the identification of more potential degradation products and hence the development of better stability-indicating assays. The pH of optimal stability is also important for selecting the appropriate salt form of the compound and excipients in the solid formulation. Acetylsalicylic acid (ASA) is a readily hydrolyzable drug and there are literature references to its rate of hydrolysis in aqueous solution.

The moisture content of the drug and excipients plays a critical role in their incompatibility by hydrolysis. Excipients such as starch and povidone have particularly high water contents (povidone contains about 28% equilibrium moisture at 75% relative humidity), which can increase the possibility of drug contact. Magnesium trisilicate causes increased hydrolysis of aspirin in tablet form because, it is thought, of its high water content. For these reasons, some scientists recommend inclusion of water in the samples for compatibility studies. Depending on the degree of hydrolytic susceptibility, different approaches for tablet formulation can be used to minimize hydrolysis. For compounds such as ASA that are readily hydrolyzable, direct compression or dry granulation is preferable rather than wet granulation. However, drug–excipient incompatibility still occurrs. Chemical interaction between moieties of drug and excipients may lead to increased decomposition. The transacetylation reaction between aspirin and paracetamol and also possible direct hydrolysis of the paracetamol can occur. The amount of free salicyclic acid at 37 °C in the tablets containing paracetamol increases by the addition of talc (0.5–1%). The stearate salts should be avoided as tablet lubricants if the active component is subject to hydroxide ion–catalyzed degradation. The degradative effect of the alkali stearates is inhibited in the presence of malic, hexamic, or maleic acid.

As a general rule, in selecting excipients, it is probably best to avoid hygroscopic excipients when formulating hydrolytically labile compounds. One of the most effective ways to stabilize a pH-sensitive drug is through adjustment of the microscopic pH of the formulation. Excipients with high pH stability and buffering agents are recommended. The equilibrium moisture content (hygroscopicity) at different relative humidities for a variety of drug and excipients would be a clue to selecting optimal formulations. If dry processing and the use of nonhygroscopic excipients still result in unacceptable rates of hydrolysis, use of a dessicant and/or moisture protective packaging can further increase stability against drug hydrolysis. The manufacturing process should be conducted under low-humidity conditions and not during the hot summer season to improve drug–excipient compatibility.

Oxidation reactions are complex and it is difficult to understand the reaction mechanism. The best approach is to avoid excipients containing oxidative reactants such as peroxides and metal ions. The air oxidation or metal ion–catalyzed oxidation can be tested after storing the samples in the solutions. The impurities in excipients such as povidone or as degradation products in PEGs are organic peroxides and are typically more reactive than hydrogen peroxide. A commercially available organic peroxide such as *tert*-butyl hydroperoxide is better to evaluate the susceptibility of a compound to peroxide oxidation rather than the hydrogen peroxide. Reactive impurities such as peroxides and ionic chemicals (talc and titanium oxide) in the excipients commonly may act as catalysts for oxidation of the drug.

Therefore, it is essential to remove any oxidative peroxides and ionic chemicals. Use of free-radical scavengers such as butylated hydroxyanisole (BHA) or butylated hydroxytoluene (BHT) can stabilize the oxidation reaction of the drug with excipients via a free-radical mechanism. However, uniform distribution of the low levels of these antioxidants in solid formulations is quite difficult. The BHA is added to both lovastatin and simvastatin tablets. If an oxidation reaction is catalyzed by metal ions or excipients with high transition metal ion contents (talc), chelating agents can also be used to bind trace metal ions. However, citrate is often the agent of choice because of the potential toxicity of many chelating agents. Oxidation is less likely to occur if the oxidizable group of phenols/catechols and secondary and tertiary amines is protonated. If some oxidation reactions are catalyzed by light, the formulation can be stabilized by a light-absorbing coating.

Aldehyde–amine addition is also a typical reaction type. The potential aldehyde interaction also makes sense to screen out a compound as part of the preformulation evaluation. All aldoses such as lactose or excipients such as starch and microcrystalline cellulose that have terminal aldose groups should be avoided with the excipients having primary/secondary amines. 5-(Hydroxymethyl)-2-furaldehyde (HMF) would be a good choice since it is a degradation product of many sugars (lactose) and celluloses, at least in trace levels. The Schiff's base[–CH=N–] is formed when the sugar aldehyde (lactone) and the primary/secondary amines are mixed. The isoniazide and trace level of HMF from lactose can also readily form the Schiff's base. It has been reported that the reaction of hydrazine hydrochloride and starch to form high-molecular-weight addition products gives high-molecular-weight products. The reaction of fluoxetine hydrochloride is more rapid with spray-dried lactose monohydrate.

Many excipients are acids or bases or have acidic or basic impurities. For this reason, the reaction of amines with aldehydes requires that the amine exist in the nucleophilic-free base, rather than the protonated cationic form. For example, the reaction of fluoxetine hydrochloride with lactose was much more rapid as potassium hydroxide was added to neutralize the hydrochloride salt. Michael addition between seproxetine and maleic acid from a tablet formulation following dissociation of the hydrochloride salt occurs because of the conversion of a salt to a free acid/base.

Finally, the physical instability that can occur is the cross-linking of gelatin or its capsules. Low levels of aldehyde impurities in excipients (starch, polysorbate 80, PEG, and rayon coils) and packaging materials have been reported to cause cross-linking of gelatin. Dissolution slowing is also more pronounced for wet granulation tablets than direct-compression tablets. Dissolution slowing appears to be due to hardening of the water-soluble excipients and a reduction of disintegration time.

Once suspect excipients have been eliminated from consideration, small-scale formulations with manufacturing processes such as granulation, drying, and compression can be used to assess whether interactions between excipients or processing conditions result in any unpredicted instability. The ratio of drug and excipients is also a critical factor to consider in incompatibility studies. In addition, the lower the drug dose, the greater the possibility of degradation by low-level impurities in the excipients.

In conclusion, drug–excipient compatibility studies have a key role at the early preformulation stages to select excipients or after formulation to help identify the mechanism of any detected instability [14]. An understanding of the potential physicochemical interactions of drug with known chemical reactivities of excipients and

their impurities will aid in the proper selection of excipients. Knowledge of trace impurities/additives in excipients and the consistency of these levels from batch to batch and supplies to supplier is also essential to select proper pharmaceutical excipients. Although the best solution kinetics and drug–excipient compatibility studies may be established, the possibility of unexpected instability because tablets and capsules are complex multicomponent systems must be remembered.

## 4 POWDER CHARACTERISTICS

Most drug and inactive excipients used in tablet formulation are in the solid state as amorphous powder or crystals of various morphological structures. There may be substantial differences in particle size, surface area, crystal morphology, wetting, and flowability as well as many physical properties of drug, excipients, and their blends [16]. Table 12 describes common micromeritic topics important to pharmaceutical preformulation.

Before their use in the solid dosage forms, it is necessary to understand and characterize the physical and chemical properties of drug, excipient, and their powder mixtures, including crystal habit, particles size, shape, flow characteristics, density, hygroscopicity, and compressibility and compaction [2, 3]. Hiestand noted that successful tableting operations require the selection of excipients that balance desirable physical, flow, and mechanical properties for tablet manufacturing [17]. The quantification of these properies using a unified approach is essential to the design and potimization of solid dosage formulations [1]. Instrumental analyses such as scanning electron microscopy (SEM), DSC, and powder X-ray diffraction (PXRD) can be very useful to characterize powder properties such as purity, polymorphism, salvation, degradation, drug–excipient compatibility, and other desirable characteristics.

### 4.1 Crystal Form and Habit

The morphology of a pharmaceutical solid is of importance since this property can influence the bulk powder properties. The six crystal systems are cubic, hexagonal, tetragonal, orthorhombic, monoclinic, and triclinic crystals. The observed overall

**TABLE 12 Micromeritic Topics Important to Pharmaceutical Preformulation**

| |
|---|
| Particle shape |
| Particle size distribution |
| Solid geometry (packing, density, porosity, void) |
| Surface characteristics (adsorption, area, surface energy, solubility) |
| Methods of determination |
| Chemical stability |
| Dynamics (flow rate, transport) |
| Particle separation |
| Processing (sieving, sedimentation, grinding, mixing, compaction) |
| Sampling |
| Drug release |

shape of crystal habits comprises plate, tabular, equant, columnar, blade, and avicular as well as dendrites, treelike pattern, and spherulites, tiny crystals radiating from a center [2, 18, 19]. Crystal form (crystal habit) as well as a noncrystalline amorphous form may affect drug stability, dissolution rate, flow, mechanical properties, and ability to mix with excipients. The amorphous form of a drug has the lowest melting point and usually the fastest dissolution rate, but it is most likely to react or degrade. Mechanical properties such as flowability, miscibility, particle strength, and cohesiveness often vary among the polymorphs, crystal shapes, or habits. Cohesiveness, the surface free energy effect that results in particle aggregation, may also be different for various polymorphs and habits.

### 4.2 Particle Size Distribution

The determination and control of particle size distribution are often very important in pharmaceutical preformlation since the drug safety, stability, and viability of the dosage form and manufacturing process can be significantly influenced [19]. Particle shape is also important in determining particle size. The particle size of materials is readily expressed in terms of its diameter according to the definition of particle size. As the degree of asymmetry of particles increases, however, so does the difficulty of expressing size in terms of a meaningful diameter [20]. Figure 4 shows different ways of defining a diameter. The "diameter" can be simply the longest or the shortest linear dimension of the crystal. If one can calculate the area of the particle, one may obtain the diameter $d_a$, which is the diameter of a circle with the same area as the particle.

If all the particles in a sample are of the same size, then the powder is monodisperse. Truly, the particles have more than one size in polydisperse samples. The monodisperse particle size distribution is more desirable than the polydisperse one. Therefore, the shape and surface area of the individual particles, the size ranges based on number or weight of particles, as well as the total surface area are variable. The commonly illustrated particle forms are sphere, rod, fiber, granular, cubical, flake, condensation floc, and aggregate.

The fineness of the powder is characterized by a number (e.g., a diameter $d$). Particles, of course, will have different shapes so that there are different ways of defining a diameter. The technique for obtaining $d_a$ given above has been used microscopically. More conventional is the so-called surface mean diameter, which, is the diameter of a sphere that has the same surface area as the particle. The so-called single-particle volume mean diameter is possible if there are instruments that can measure the volume of an odd-shaped particle. If the shape factor is indepen-

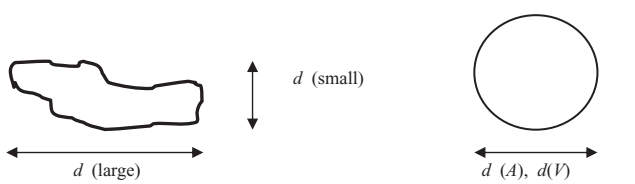


**FIGURE 4** Different ways of defining a diameter.

**TABLE 13 Particle Size Distribution of Powdered Cornstarch Samples**

| Particle Size Range (μm) | Number of Occurrences ($n$) | Percent Frequency | Cumulative Frequency (%) |
|---|---|---|---|
| 25–27 | 1 | 8.3 | 8.3 |
| 27–29 | 3 | 25.0 | 33.3 |
| 29–31 | 4 | 33.3 (33.4) | 66.7 |
| 31–33 | 3 | 25 | 91.7 |
| 33–35 | 1 | 8.3 | 100 |
| Total | 12 | 99.9 (100) | — |

dent of the size of the particle, then the shape is called isometric. Examples of isometric shapes are cube, sphere, and cylinder.

Suppose a sample of 12 particles of corn starch with fairly narrow particle size distribution (25–35 μm) were measured. In this case, the distribution of particles is very narrow and approximately distributed as normal or Gaussian pattern. To convert these numbers into frequencies, it is noted that there are 12 particles in total; that is, dividing each number by 12 and multiplying this by 100 will give the percent frequency, as shown in Table 13.

If the frequency is plotted as a function of the midpoint of the diameter ranges, then a frequency histogram is obtained. It is noted that for the noncumulative curve the midpoints of the intervals are used, but for the cumulative curve the interval endpoint is used. When the number or weight of particles lying within a certain size range is plotted against the size range or mean particle size, a *frequency distribution curve* is obtained However, due to the deviation from the normal distribution of particles, a lognormal distribution of the particle size is statistically plotted against the cumulative percent frequency on a probability scale, and a linear relationship is observed (Figure 5). Probability paper is a type of paper that straightens out this type of S-shaped curve. The logarithm of the particle size is equivalent to 50% on the probability scale, that is, the 50% size is known as the *geometric mean diameter* and the slope is given by the *geometric standard deviation* $\sigma_g$, which is also the quotient of the ratio: (84% undersize or 16% oversize)/(50% size) or (50% size)/(16% undersize or 84% oversize).

Many methods are available for determining particle size in pharmaceutical practice, including microscopy, sieving, sedimentation, and determination of particle volume [19]. Sieve analysis with U.S. standard sieves is widely used to determine the particle size distribution based on powder weight. Sieves are classified according to the number of openings (Table 14) and are generally made of wire cloth woven from brass, bronze, or other suitable wire.

The USP uses descriptive terms to characterize the particle size of a given powder, which are related to the proportion of powder that is capable of passing through the openings of standardized sieves of varying dimensions in a specified time period under the mechanical sieve shaker as follows:

Coarse (or a no. 20) powder: All particles pass through a no. 20 sieve and not more than 40% through a no. 60 sieve.

Moderately coarse (or a no. 40) powder: All particles pass through a no. 40 sieve and not more than 40% through a no. 60 sieve.

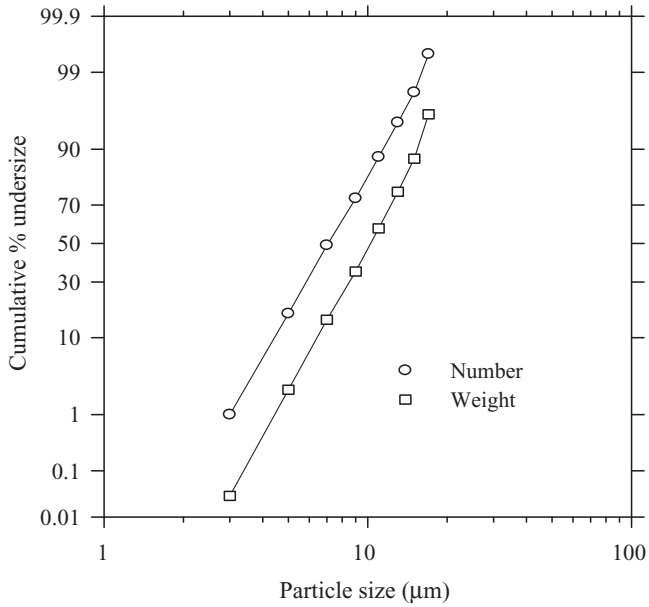


**FIGURE 5** Typical lognormal distribution of particles based on weight and number.

**TABLE 14 Openings of Standard Sieves**

| Sieve Number | Sieve Opening |
|---|---|
| 2 | 9.5 mm |
| 3.5 | 5.6 mm |
| 4 | 4.75 mm |
| 8 | 2.36 mm |
| 10 | 2.00 mm |
| 20 | 850 μm |
| 30 | 600 μm |
| 40 | 425 μm |
| 50 | 300 μm |
| 60 | 250 μm |
| 70 | 212 μm |
| 80 | 180 μm |
| 100 | 150 μm |
| 120 | 125 μm |
| 200 | 75 μm |
| 230 | 63 μm |
| 270 | 53 μm |
| 325 | 45 μm |
| 400 | 38 μm |

Fine (or a no. 80) powder: All particles pass through a no. 80 sieve. There is no limit as to greater fineness.

Very fine (or a no. 120) powder: All particles pass through a no. 120 sieve. There is no limit as to greater fineness.

Microscopy, in which the particles are sized through the use of a calibrated grid background or other measuring devices (range 0.2–100 μm). SEM can also readily measure the smallest particle size. The microscope allows the observer to view the actual particles, but it gives two-dimensional views. With the sedimentation rate, particle size is determined by measuring the terminal settling velocity of particles through a liquid medium in a gravitational or centrifugal environment (range 0.8–300 μm). Sedimentation rate may be calculated from the well-known Stokes equation. The sedimentation methods yield a particle size relative to the rate at which particles settle through a suspending medium, a measurement important in the development of emulsions and suspensions. With light energy diffraction (light scattering), particle size is determined by the reduction in light reaching the sensor as the particle, dispersed in a liquid or gas, passes through the sensing zone (range 0.2–500 μm). On the other hand, laser scattering utilizes a $H_3$–Ne laser, silicon photodiode detectors, and an ultrasonic probe for particle dispersion (range 0.02–2000 μm). The measurement of particle volume using a Coulter counter allows one to calculate an equivalent volume diameter, but no information on shape of the particles is available. Laser holography, in which a pulsed laser is fired through an aerosolized particle spray and photographed in three dimensions with a holographic camera, allows the particles to be individually imaged and sized (range 1.4–100 μm). The above methods and others may be combined to provide greater assurance of size and shape parameters. Automated particle size analyzers linked with computers are commercially available for data processing, distribution analysis, and printout.

Determination and control of particle size are often prerequisites in preformulation stages because the size distribution of excipients, drug, and their mixtures can influence safety, efficacy, stability, viability of dosage form, and manufacturing processes. Furthermore, the particle size of pharmaceuticals can affect uniform mixing, flow characteristics, formulation characteristics, dose-to-dose content uniformity, dissolution rate, and bioavailability of drug. Tablet characteristics such as porosity and flowability are highly affected by the particle size as well. The smallest particles induce electrostatic forces and aggregations while the larger particles show greater weight variations. The ideal size ranges of particles are usually 10–150 μm. Therefore, detailed information of the particle size of drug, excipients, and their blends should be required in tablet formulation as well as regulatory issues. Particle size is also important in the tableting field, since it can be very important for good homogeneity in the final tablet. The particle size should be consistent throughout the production to satisfy table formulation and regulatory demands.

### 4.3 Flow Characteristics

Good flow properties are a prerequisite for the successful manufacture of both tablets and powder-filled hard gelatin capsules. Proper fluidity of materials is required to be transported through the hopper of a tableting machine. The elongated particle shape and small particle size could cause high tablet weight variation, strength, unacceptable blend uniformity, and difficulty in filling containers and dies. Excipients with good flow characteristics and low cohesive powders should be more preferable in tablet production. Powder flow is affected by the numerous parameters, including purity, crystallinity, electrostatic forces, mechanical

properties (brittleness, elasticity), density, size, shape, surface area, moisture content, direction and rate of shear, storage container dimension, and particle–wall interaction [19].

It is a property of all powders to resist the differential movement between particles when subjected to external stresses. A bulk powder is somewhat analogous to a non-Newtonian liquid, which exhibits plastic flow and sometimes dilatancy if the particles being influenced by attractive forces. Accordingly, powders may be free flowing or cohesive ("sticky"). The resistance is to free flow is due to the cohesive forces between particles [18]. Three principal types of interparticular forces are forces due to electrostatic changing, van der Waals forces, and forces due to moisture. Electrostatic forces are dependent on the nature of the particles, in particular their conductivity. Van der Waals forces are the most important forces for most pharmaceutical powders. These forces are inversely proportional to the square of the distance between the two particles and hence diminish rapidly as particle size and separation increase. Powders with particles below 50 μm will generally exhibit irregular or no flow due to van der Waals forces. Particle shape is also important; for example, the force between a sphere and a plane surface is about twice that between two equal-sized spheres. At low relative humidity, moisture produces a layer of adsorbed vapor on the surface of particles. Above a critical humidity, typically in the range 65–80%, it will form water liquid bridges between particles. Where a liquid bridge forms, it will give rise to an attractive force between the particles due to surface tension or capillary forces. The role of the formulator is to ensure that the flow properties of the powder are sufficient to enable its use on modern pharmaceutical equipment, powder hoppers, and flow through orifices in the tablet production.

It is important that the powder flows from the hopper to the filling station of the tablet machine at an appropriate rate and without segregation occurring. There are two types of flow that can occur from a powder hopper: *core flow* and *mass flow* [2]. Figure 6 shows the two different powder flow patterns in hoppers. When a small amount of powder is allowed to leave the hopper, there is a defined region in which downward movement takes place and the top surface begins to fall in the center. A core flow hopper is characterized by the existence of dead spaces during discharge. A mass flow hopper is one in which all the material is in motion during discharge, in particular the areas adjacent to the hopper wall. As a small amount of powder is discharged, the whole bulk of the powder will move downward. Whether core flow or mass flow is achieved is dependent on the design of the hopper (geometry and wall material) and the flow properties of the powder.

Powder flow into orifices is also important when filling dies in tablet machines and in certain types of capsule-filling machines. For a given material, the flow into or through an orifice is dependent on the particle size (Figure 7). In general, as the particle size increases, the powder flow rate also increases. However, there is practically no flow if the particle size is below 50 μm or above 1200 μm. The Carr index gives us the guidance for powder flowability. A lower Carr index of excipients is more desirable for acceptable powder flow. At the lower end of the particle size range, cohesive forces will result in poor flow. Powders with particles below 50 μm will generally exhibit irregular or no flow due to van der Waals forces. As the particle size increases, the flow rate increases until a maximum is achieved, at an orifice diameter–particle diameter ratio of 20–30. As the particle size continues to increase,

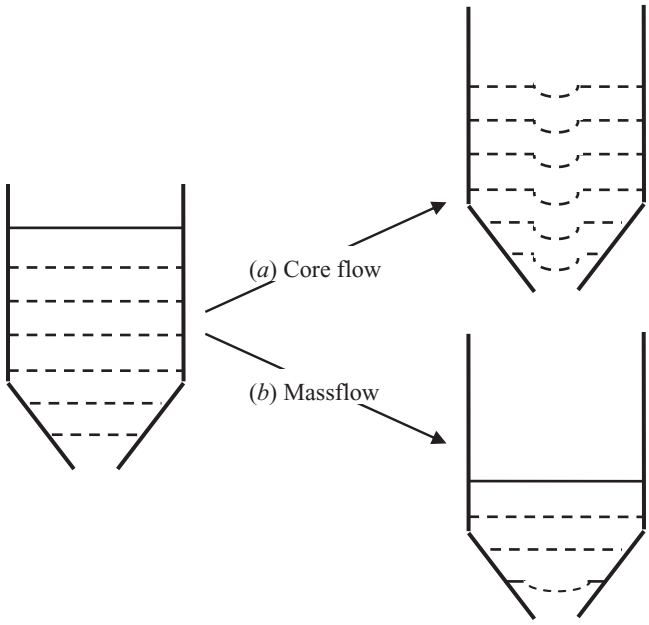


**FIGURE 6** Powder flow patterns in hoppers.

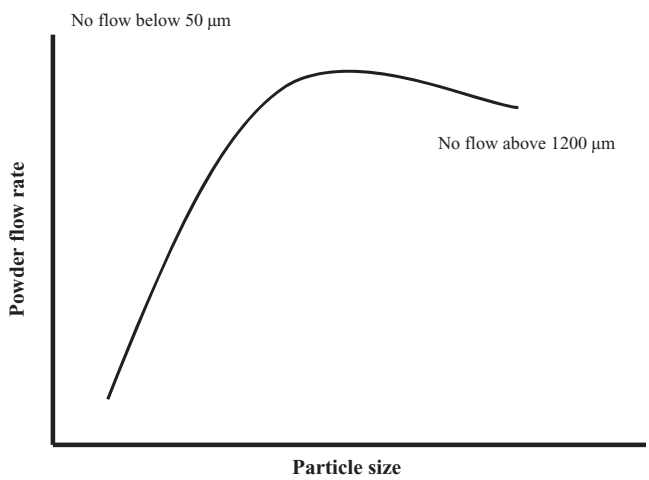


**FIGURE 7** Effect of particle size on rate of powder flow through orifice.

the rate decreases due to mechanical blocking or obstruction of the orifice. Flow will stop completely when the orifice–particle ratio falls below 6 and if the size is above 1200 μm.

There are several different methods available for determining the flow properties of powders. *Shear cell methods* provide an assessment of powder flow properties as a function of consolidation load and time. There are a number of types of shear cells available, the most common being the Jenike shear cell [21].

Common indices of flowability are the *Hausner ratio* and the *Carr index* (compressibility). The increase in bulk density of a powder is related to the cohesiveness of a powder. So measurement of the bulk density of a powder is essential to define the flow characteristics. Ratios of poured-to-tapped bulk densities are expressed in two ways to give indices of flowability:

$$\text{Hausner ratio} = \frac{\text{tapped bulk density}}{\text{poured bulk density}}$$

$$\text{Carr index (compressibility)} = \frac{100 \times (\text{tapped bulkdensity} - \text{poured bulk density})}{\text{poured bulk density}}$$

The Hausner ratio varies from about 1.2 for a free-flowing powder to 1.6 for cohesive powders. Carr index classifications for flowability [2] are listed in Table 15.

Compressibility indices are a measure of the tendency for arch formation and the ease with which the arches will fail and, as such, is a useful measure of flow. A limitation of the bulk density indices for flow characteristics is that they only measure the degree of consolidation; they do not describe how rapidly consolidation occurs. *Angle of repose* is a common method used to measure powder flow with small sample quantity. If powder is poured from a funnel onto a horizontal surface, it will form a cone due to gravitational forces. The angle between the sides of the cone and the horizontal is referred to as the angle of repose. So there is a correlation between powder flow and angle of repose. The relationship between the Carr index and angle of repose is now discussed. The angle of repose is a measure of the cohesiveness of the powder, as it represents the point at which the interparticle attraction exceeds the gravitational pull on a particle. A free-following powder will form a cone with shallow sides, and hence a low angle of repose, while a cohesive powder will form a cone with steeper sides. As a rough guide, angles less than 30° are usually indicative of good flow, while powders with angles grater than 40° are likely to be problematic.

The *avalanching behavior* of powder is also a measure of flowability. If a powder is rotated in a vertical disc, the cohesion between the particles and the adhesion of the powder to the surface of the disc will lead to the powder following the direction of rotation until it reaches an unstable situation where an avalanche will occur. After the avalanche, the powder will again follow the disc prior to a further avalanche.

**TABLE 15 Carr Index Classification and Powder Flowability**

| Carr Index (%) | Flow |
|---|---|
| 5–12 | Free flowing |
| 12–16 | Good |
| 18–21 | Fair |
| 23–35 | Poor |
| 33–38 | Very poor |
| >40 | Extremely poor |

Measurement of the time between avalanches and the variability in time is a measure of the flow properties of the powder.

If a powder flows poorly, the vibrator can be used, but it also causes powder segregation and stratification. The addition of glidant (occasionally lubricant) in the powder mixtures can readily increase flowability at the low concerntration. Talc or fumed silicon dioxide is an example of a glidant. If this is not sufficient to improve the flow, other means of flow improvement are necessary. There are two main factors that affect powder flow: particle size and particle shape. The more spherical a particle is, the better it flows. Small particles are very cohesive and hence do not flow well, but increasing the particle size will improve flow. With the aid of spray drying or spheronizers, particles become spherical.

In general, powder below 50 μm is not very free flowing because the cohesive forces below this size become stronger than the gravitational force, and flow through the orifice is prevented. This, of course, is a function of the size of the orifice, and flow might be possible in a larger orifice.

Particle size enlargement of the drug substance can be brought about by manipulation of the recrystallization step in the synthesis of the drug. To increase powder flow, particle size enlargement by slugging, roller compaction, and wet granulation can be used. If a powder is compressible but does not flow well, then slugging may be employed. In slugging, tablets are made of the poorly flowing substance on a high-duty, slowly operating machine into large dies. The dies are large so that the flow is sufficiently increased, but now the compression forces must be increased because the larger area dictates a larger force to attain a given pressure (the elastic limit being in stress units). In roller compacting, the powder is processed between two heavy-duty rollers, compacted between the rolls, and emerges as a compressed sheet, which is then milled. These two methods are necessary if the drug substance (e.g., aspirin) is sufficiently moisture sensitive and there are stability issues so it cannot be wet granulated. Otherwise, wet granulation is a frequently used method of particle enlargement for free-flowing powder.

### 4.4 Density and Bulkiness

***Density*** When a powder is poured into a container, the volume that it occupies depends on a number of factors, such as particle size, particle shape, and surface properties. In normal circumstances, it will consist of solid particles and interparticlulate air spaces (voids or pores). The particles themselves may also contain enclosed or intraparticulate pores. If the powder bed is subjected to vibration or pressure, the particles will move relative to one another to improve their packing arrangement. Ultimately, a condition is reached where further densification is not possible without particle deformation. The density of a powder is therefore dependent on the handling conditions to which it has been subjected, and there are several definitions that can be applied either to the powder as a whole or to individual particles.

Because particles may be hard and smooth in one case and rough and spongy in another, one must express densities with great care. Density is universally defined as weight per unit volume. Three types of densities—true density, particle density, and bulk density—can be defined, depending on the volume of particles containing microscopic cracks, internal pores, and capillary spaces.

The true density is the material itself exclusive of the voids and interparticular pores larger than molecular or atomic dimension in the crystal lattice. Particle (granular) density, determined by the displacement of mercury, which does not penetrate at ordinary pressure into pores smaller than 10 μm, is the mass of the particle divided by its volume. The different terms arise from the definition of volume: *True particle density* is when the volume measured excludes both open and closed pores and is a fundamental property of a material; *apparent particle density* is when the volume measured includes intraparticulate pores; *effective particle density* is the volume "seen" by a fluid moving past the particles. *Bulk density* (powder density) is the volume in a graduated cyclinder including both the particulate volume and the pore volume. The bulk density will vary depending on the packing of the powder. Based on the definition of volume, *minimum bulk density* is when the volume of the powder is at a maximum, caused by aeration, just prior to complete breakup of the bulk. *Poured bulk density* is when the volume is measured after pouring powder into a cylinder, creating a relatively loose structure. *Tapped bulk density* is, in theory, the maximum bulk density that can be achieved without deformation of the particles.

Based on the definition of density, two new terms are defined. *Porosity* is defined as the proportion of a powder bed or compact that is occupied by pores and is a measure of the packing efficiency of a powder and *relative density* is the ratio of the measured bulk density and the true density:

$$\text{Porosity} = 1 - \frac{\text{bulk density}}{\text{true density}}$$

$$\text{Relative density} = \frac{\text{bulk density}}{\text{true density}}$$

***Bulkiness*** The specific bulk volume, the reciprocal of bulk density, is often called *bulkiness* or *bulk*. It is an important consideration in the packaging and filling of powders for tablet production. The bulk density of calcium carbonate can vary from 0.1 to 1.3, and the lightest or bulkiest type would require a container about 13 times larger than the heaviest type. Bulkiness increases with a decrease in particle size. In a mixture of materials of different sizes, however, the smaller particles sift between the larger ones and tend to reduce the bulkiness.

To define bulkiness in detail, the porosity and density of powders should be understood. Suppose a powder, such as zinc oxide, is placed in a graduated cylinder and the total volume is noted. The volume occupied is known as the bulk volume $V_b$. If the powder is nonporous, that is, has no internal pores or capillary spaces, the bulk volume of the powder consists of the true volume of the solid particles plus the volume of the spaces between the particles. The volume of the spaces, known as the void volume $v$, is given by equation

$$v = V_b - V_p$$

where $V_p$ is the *true volume* of the particles.

The *porosity* or *voids* (ε) of a powder is also defined as the ratio of the void volume to the bulk volume of the packing given below. Porosity is frequently expressed in percent, $\varepsilon \times 100$:

$$\varepsilon = \frac{V_b - V_p}{V_b} = 1 - \frac{V_p}{V_b}$$

### 4.5 Hygroscopicity

The hygroscopicity of a drug and pharmaceutical substances is a potential parameter to be considered in tablet formulation. The moisture uptake rate is quite variable depending on the type of drug and excipients as well as the environmental conditions. So, a concise definition of hygroscopicity is not possible. Powders can absorb moisture by both capillary imbibition and swelling. The instantaneous water absorption prosperties of pharmaceutical excipients correlate with total surface area while the total absorption capacity correlates with powder porosity [22].

If drug and excipients are so hygroscopic, they can readily adsorb water until they deliquesce, or begin to dissolve. Moisture adsorption is important because adsorbed water can cause incorrect weighing and degradation of drug and/or excipients. The drug, excipients, and water reaction will continue as the amount of water increases. When a solid is placed in a room, moisture will condense onto it. If this occurs simply as a limited amount of adsorbed moisture, then the substance is not hygroscopic under these conditions. These conditions exist if the water vapor pressure in the surrounding atmosphere is lower than the water vapor pressure over a saturated solution of the solid in question. However, if the water vapor pressure in the atmosphere is higher than that of the saturated solution, there will be a thermodynamic tendency for water to condense upon the solid materials (drug and excipients).

The drug and pharmaceutical excipients adsorb or lose the moisture depending on the relative humidity in the atmosphere. The nonhygroscopic materials are not affected by the moisture and are in a equilibrium state. In general, solid dosage forms such as tablets or capsules should be hydrophilic because the solid materials must dissolve after swallowing. However, solid dosage forms must also be stable against physical and chemical factors.

The moisture uptake rates (MUR) can simply be obtained by weighing the sample after a given time (six days), but in such a case it is assumed that the moisture uptake is still in the linear phase. If, for instance, the weight gain is 5 mg per 10-g sample in six days, then the MUR is $5/(10 \times 6) = 0.083$ mg/g/day. If the MUR values are plotted versus relative humidity (RH) and the curve that intercepts the $x$ axis at 20% RH is obtained from a straight line, the compound can be stored without moisture pickup in atmospheres of less than 20% RH. In addition, the hygroscopicity of materials is indicated as follows:

$$\text{Loss of drying (LOD \%)} = \frac{\text{weight of water in sample}}{\text{total weight of wet sample}} \times 100$$

$$\text{Moisture content (MC \%)} = \frac{\text{weight of water in sample}}{\text{weight of dry sample}} \times 100$$

$$\text{Equilibrium moisture contents (EMC)} = \frac{\text{\% LOD}}{\text{\% LOD} + 100} \times 100$$

$$\text{Relative humidity (RH)} = \frac{\text{water vapor pressure in atmosphere}}{\text{saturated water vapor pressure}} \times 100$$

Depending on the hygroscopicity based on the EMC, various drug and excipients are classified in four groups (Table 16).

Figure 8 also shows an example of moisture uptake for four selected excipients as a function of relative humidity. Depending on the hygroscopicity of the exscipients, the uptake behaviors are quite variable. Excipients such as microcrystalline cellulose (MCC) and starch can pick up significant amounts of water at relatively low relative humidity. Since this water is not present as a hydrate, it is potentially free to interact with a drug.

**TABLE 16 Classification of Hygroscopicity and Example Pharmaceutical Excipients**

| Type I: Nonhygroscopic | Type II: Slightly Hygrocopic | Type II: Moderately Hygroscopic | Type IV: Very Hygroscopic |
|---|---|---|---|
| No MC change below RH 90% or less than 20% MC changes at RH 90% after 1 week storage | No MC change below RH 80% or less than 40% MC changes at RH 80% after 1 week storage | Not more than 5% MC change below RH 60% or less than 50% MC changes at RH 70% after 1 week storage | Increase of MC at RH 40–50% or more than 70% MC changes above RH 40% after 1 week storage |
| | | *Examples* | |
| Lactose USP | MCC | HPC | Povidone |
| Dicalcium phosphate | Sucrose, dextrose | HPMC | Sodium starch glycolate |
| Ethylcellulose | Poloxamer 188 | Bentonite | Polyplasdone XL |
| Magnesium stearate | PEG 3350 | Pregelatinized starch | Sorbitol |
| | CAP | Starch USP, Corn | CMC Na |
| | | Gelatin USP | |

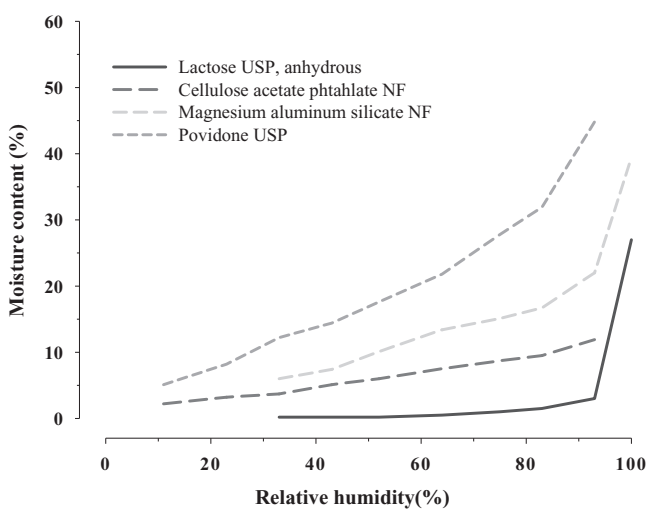


**FIGURE 8** Profiles of moisture uptake for four selected excipients as function of relative humidity.

The moisture content of drug can affect the cohesiveness of particles through hydrogen bonding or by changing surface energy effects. Water also can act as a plasticizing agent and can lower the glass transition temperature of amorphous polymorphs, thus allowing a rubbery state to exist at a lower temperature. Water also can take up intra- and intergranular pore spaces (capillaries) of powders by acting as an interparticle bridge through surface adsorption mechanisms.

The hygroscopicity of materials influences the flowing characteristics of drug and excipients by forming adsorption film with an aid of water as a solvent. Dissolution of drug and excipients from solid dosage forms can also occur. The crystallinity of troglitazone-PVP K30 solid disperisions can be changed by the water content [23]. During manufacturing operations, water's ability to act as a bridge between particles can improve the compression capabilities of powder masses and affect the tensile strength and hardness of the tablet. The milling or blending is also changed. Most of all, the chemical hydrolysis of labile materials is more facilitated, resulting in low stability during storage. Adsorbed water also can act locally as a solvent, and in many cases drug–excipient incompatibility occurs by dissolving drug or excipients in granulation or blended powder mass, as observed when aspirin tablets liberate a strong odor of free salicyclic acid. Any free, unbound water in the mass can migrate throughout the material mass and act as a reagent. A waterproof packaging or container is used for many moisture-sensitive materials.

### 4.6 Mixing

The mixing of powders is a key step in the manufacture of virtually all solid dosage forms. In general, particle size, shape, and surface energy are important factors for blending of drug and excipients in unit operations. The perfect mixing of powder is desirable, that is, a mixture in which the probability of finding a particle of a given component is the same at all positions in the mixture, but the powder mixing has a maximum degree of randomness (Figure 9). To determine the degree of mixing obtained in a pharmaceutical operation, it is necessary to reasonably sample the mixture and determine the variation within the mix statistically [2].

Uniform mixing of powdered materials occurs if they have similar particle size distributions and particle shapes. Spherical particles mix least well while plate and fiber shapes also do not mix well because they tend to clump. The more cohesive the material, the more difficult it is to mix that material with other materials. Similarly, cohesiveness between drug and excipient or among excipients may hinder the

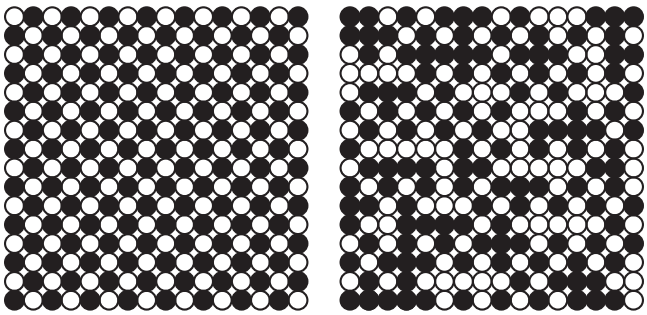

**FIGURE 9** Comparison of powder mixing: perfect mixing and random mixing.

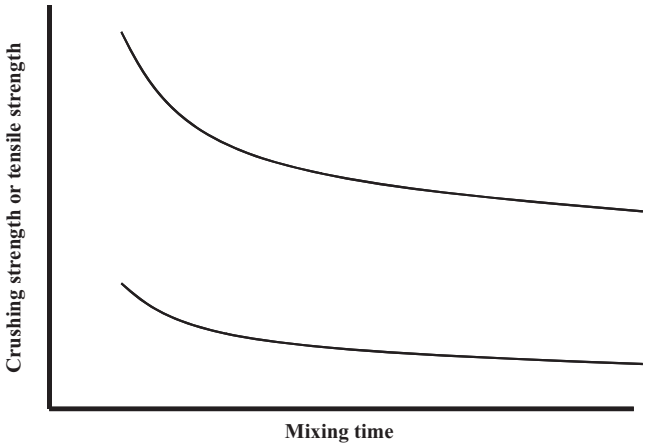


**FIGURE 10** Effect of mixing time on mechanical strength of excipients.

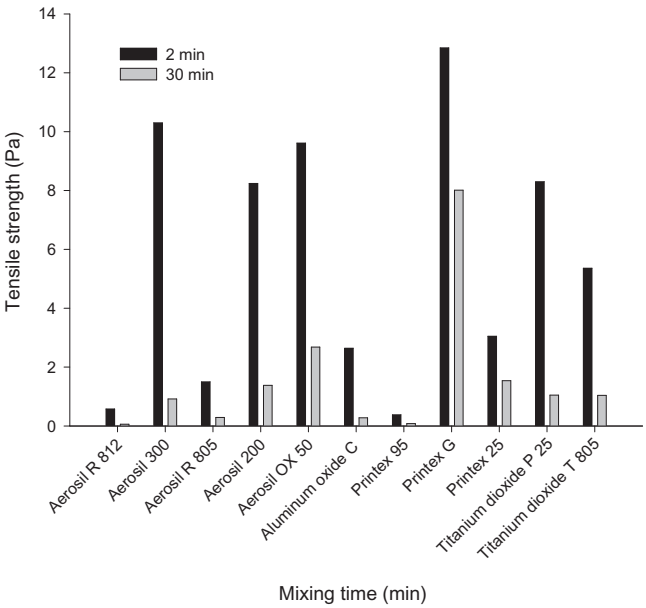


**FIGURE 11** Effect of mixing time glidant tensile strength.

successful blending process. Ideally, the uniform mixing of powders should be such that the weight of sample taken is similar to the weight that the powder mix contributes to the final dosage form.

Figure 10 shows the effect of mixing time on the mechanical strength of powders. As the mixing time increases, the tensile strength of powders (or curshing strength of the tablet) gradually decreases [24, 25]. While the strength decreases with increasing mixing time for all materials tested, the effect is far more marked for materials that deform plastically. For example, the glidant tensile strength invariably decreases as the mixing time (2 min vs. 3 min) increases (Figure 11). There is no direct correlation of tensile strength with primary particle size of glidant (Figure 12). However,

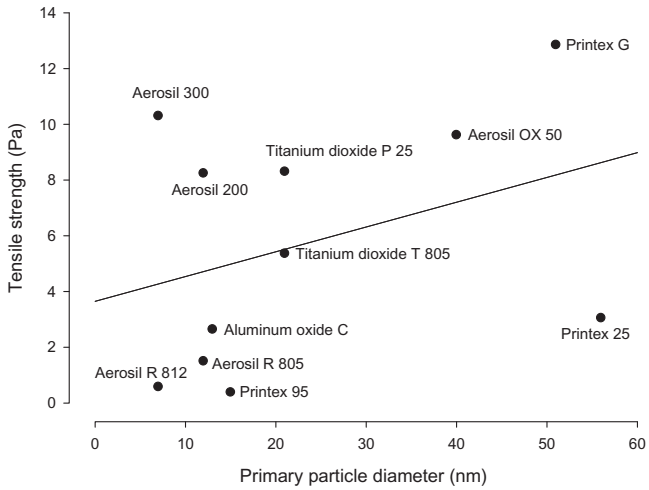


**FIGURE 12** Correlation of tensile strength with primary particle size of glidant.

this correlation was further improved when some outlying glidant (Aserosil 200 and 300 and Printex 25) were excluded ($R^2$ values of 0.1287–0.8616).

If a powder consisting of two materials both having identical physical properties is mixed for a sufficient time, random mixing will eventually be achieved. Unfortunately, most pharmaceutical powders consist of mixtures of materials with differing physical properties, such as size, shape, density, and surface area, leading to segregation among particles, where particles of similar properties tend to collect together in part of the powder. When segregating powders are mixed, as the mixing time is extended, the powders appear to unmix. The differences in particle size are the most important for segregation in pharmaceutical powders. One exception to overcome segregation is ordered mixing rather than random mixing. When one component of a powder mix has a very small particle size (less than 5 μm) and the other is relatively large, the fine powder may coat the surface of the larger particles, and the adhesive forces will prevent segregation, known as ordered mixing. This ordered mixing makes the powders produce greater homogeneity than by random mixing. The percolation of fine particles is also a factor. If the particles sizes are quite different, the smaller particles can drop easily and move to the bottom of powder, resulting in segregation. This segregation process can occur whenever movement of particles by vibration, shaking, and pouring takes place.

### 4.7 Particle Size Reduction (Micronization and Milling)

Mechanical attrition, that is, high-energy ball milling of powders, is a nonequilibrium processing method that has generated the reduced particle size and the formation of physically metastable materials. It can be used to modify materials by refining the microstructure, homogenizing the composition, extending solid solubility, creating matastable crystalline phases, or producing metallic glass. High-energy ball milling is both a processing method to reduce particle size and a route to the physical synthesis of metastable materials. Early in product development, when only small amounts of drug are available, comminution (grinding/mixing) may be carried out with a mortar and pestle. For lager batches, ball milling or micronization can be

used to reduce the particle size because comminution or grinding are not practical due to the length of time required [18].

There are numerous methods of particle size reduction, but their application is dependent on the intended particle size, particle distribution, cleaning convenience, operating cost, dust containment, temperature, and flexibility. The type of milling machine includes slurry, fluid energy (jet), universal, cone, and hammer. This size reduction process may reduce the risk of dissolution rate-limited bioavailability of drugs. Particle size reduction can be accomplished by using a hammer mall or a similar mill, but this process may only break up the larger crystal aggregates without significantly changing the distribution of smaller particle sizes. On the other hand, air jet mills, which impinge two streams of particles at a right angle to each other in high-velocity air streams, reduce particle size significantly within microsized ranges.

Ball milling was the most commonly used at the preformulation stage to reduce the particle size of small amounts of a compound. Ball mills reduce the size of particles through a combined process of impact and attrition. Usually they consist of a hollow cylinder that contains balls of various sizes which is rotated to initiate the grinding process. Micronized particles are typically less than 10 μm in diameter. The efficiency of the milling process is affected by rotation speed, number of balls, mill size, wet or dry milling, amount of powder, and length of time of milling. Although ball milling can effectively reduce the particle size of compounds, prolonged milling may be detrimental in terms of changes of compound crystal form from crystalline to polymorphic or amorphous form and stability.

Although ball milling on a large scale is possible, hammer milling is more preferable in the pharmaceutical industry. Powder is bled into the mill house via the hopper, and the rotating hammers impact with the powder. When this is fine enough to pass the screen, the powder will exit. The powder exiting will have a maximum particle size of that of the screen. The average particle size of the milled powder will be smaller, the smaller the feed rate, the more rapid the milling speed (rpm's), and the finer the screen. The knife has a blunt edge on one side and a knife edge on the other. Milling with the blunt edge forward gives rise to a smaller average particle size. The usual minimum particle range is about 50 μm. Large, heavy-duty hammer mills (micropulverizers) give much smaller particle size further down, typically to 20 μm. If particle sizes less than 5 μm are desired, micronizers are used.

If particle sizes in the micrometer range are required, then the attrition mill can be used. Here particles are bled into a chamber in which great turbulence has been created by two inlets of air at different pressure. The particles hit one another and are removed by centrifugal means and collected in a cyclone setup or in a bag above the cyclone. Such particles become highly electrically charged during the operation and also become very cohesive.

The milling process provides energy to the powders so that melting occurs if their melting point is sufficiently low. There is no guarantee that the original polymorph and habit will be regained upon resolidification. The milling and micronization process can also reduce the particle size of poorly soluble drugs so that the maximum surface area is exposed to enhance the solubility and dissolution properties. Although micronization of the drug offers the advantage of a small particle size and a lager surface area, it can also result in processing problems due to high dust, low density,

and poor flow properties. Indeed, micronization may be counterproductive, since the micronized particles may aggregate, which may decrease the surface area. Milling and micronization process also induce the changes of crystallinity of drug into amorphous form. It has been shown that the amorphous change of the crystalline structure can be achieved by vapor condensation, supercooling of melt, rapid precipitation from solution, and mechanical applications of a crystalline mass by milling or compaction [26].

Particle size reduction of the powder has produced defects on the surface that, if enough energy is imparted, leads to amorphous regions on the surface. In turn, these regions are found to have a greater propensity to adsorb water. The dissolution rate increases as the particle size of drug powder decreases due to its greater surface area for wetting.

It is known that ball milling and other types of milling can change the morphology of a solid, for example, make a crystalline compound amorphous, increase the surface energy of a solid, and distort the crystal lattice. For example, the crystalline solid state of sorbitol exhibits a complex polymorphism made of five different forms called A-, B-, Γ-, Δ-, and E-sorbitol [27]. The structures of these polymorphs have been identified by single-crystal X-ray diffraction. The structure of the Γ form is also the most stable and the most common polymorph of sorbitol. It was reported that the Γ form of sorbitol underwent a complete gransformation toward the A form upon ball milling and also was affected by milling time. The DSC and XRPD patterns are useful in identifying these phenomena, althought the low level of amorphous character cannot be detected by techniques such as XRPD and DSC. Crystal structures by the milling process as well as compaction forces have been shown [2].

### 4.8 Compaction(Compressibility)

Compressibility is the property of forming a stable and intact mass when pressure is applied. The manufacture of tablets involves the process of powder compaction or compressibility, the purpose of which is to convert a loose incoherent mass of powder into a single solid object. A protocol to examine the compression properties of flowing powder should be considered by the formulation scientist when selecting the excipient, the formulation type, and the manufacturing process, for example, direct compression or granulation for the intended solid dosage form. Acetaminophen is poorly compressible whereas lactose compresses well. In general, drug and excipients, including lubricant, are blended in a tumbler mixer for a period of time and then compressed into tablets in a hydraulic press. The crushing strength is also determined to test the compressibility of tablet at room temperature.

The compression of drug powder may change the crystal structure into a polymorphic form—likewise in the milling process. The PXRD patterns and DSC of a drug can detect these phenomena. Knowledge of the crystal structure of a drug is a prerequisite for compaction.

To fully understand the compaction behavior of a material, it is necessary to be able to quantify of its elasticity, plasticity, and brittleness. A powder in a container subjected to compressive force will undergo particle rearrangement [28]. The density of the bed will increase with increasing pressure at a characteristic rate. Brittle materials will undergo fragmentation, and the fine particles formed will percolate through the bed to give secondary packing. Plastically deforming materials will

distort to fill voids and may also exhibit void filling by percolation. When the limit of plastic deformation is reached, fracture occurs.

Many of the basic principles of compaction and the test methodologies are currently employed in pharmaceutical formulation. To characterize the compaction properties of a material or formulation, it must be possible to measure the relationship between the force applied to a powder bed and the volume of the powder bed. There are two principal types of compaction studies used to characterize materials: pressure–volume relationships and pressure–strength relationships. While ultimately it is the strength of a tablet that is important, the pressure–volume relationships provide information about the compaction properties of a material that allows an appropriate formulation to be developed. The instrumented tableting machine provides information for compaction that is directly relevant to production conditions. The compression profiles differ from those of rotary tableting machines used for commercial production. The profile of a single punch involves the powder bed being compressed between a moving upper punch and a stationary lower punch, while on a rotary machine, both punches move together simultaneously. A major advantage of instrumented machines is that they provide information not only on the compaction properties but also on flow and lubrication. The disadvantage of using instrumented rotary machines is the large quantity of materials required.

A large number of equations have been proposed to describe the relationship between pressure and volume reduction during the compaction process. The most widely used equation to describe the compaction of powders is the Heckel equation. Pharmaceutical powders do not produce perfect straight lines, and the type of deviation provides information about the compaction behavior of the material. A typical Heckel plot for a pharmaceutical powder is obtained showing a straight-line portion over a certain pressure range with a negative deviation at low pressures and a positive deviation at high pressures. The strength of tablets has traditionally been determined in terms of the force required to fracture a specimen across its diameter. The fracture load obtained is usually reported as a hardness value. Initially, most materials will demonstrate an increase in tensile strength proportional to the compaction pressure applied. As the compaction pressure increases, the tablet approaches zero porosity, and large increases in pressure are required to achieve small volume reductions [28]. Some materials will attain a maximum strength, and subsequent increases in pressure will produce weaker tablets. Other materials will also display an initial increase in strength proportional to the applied pressure, but the strength reaches a maximum before falling off sharply, resulting in capping or lamination in tablet production [28]. Figure 13 shows the correlation of hardness with compression pressure. Capping is the partial or complete removal of the crown of a tablet from the main body, while lamination is the separation of a tablet into two or more distinct layers. If the compressibility and flow of powders are good, it is possible to directly compress the powders into tablet. If either compressibility or flow is not satisfactory, roller compact, slugging, or we granulation can be utilized to improve compressibility.

### 4.9 Surface Area and Other Properties

The surface area of particles is related to particles size, as discussed previously. The surface area of powders affects the drug dissolution rate, powder flow, cohesiveness, and adsorption. Furthermore, the surface area of solid materials may also

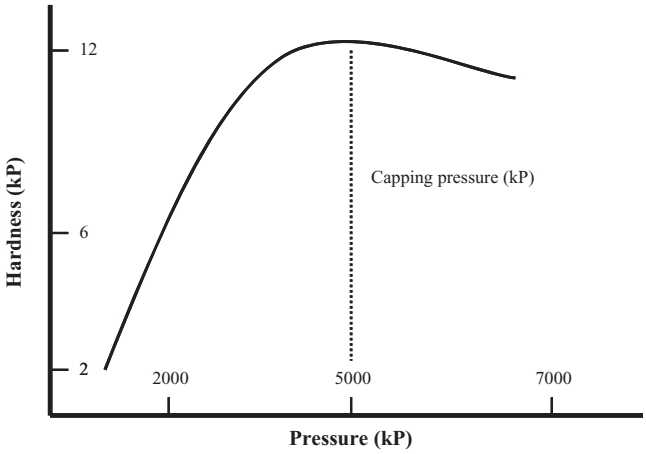


**FIGURE 13** Correlation of compression pressure with hardness.

influence the physicochemical properties, adsorption, dissolution, and bioavailability of drugs.

The particle size and surface area distributions of pharmaceutical powders can be obtained by microcomputerized mercury porosimetry. Mercury porosimetry gives the volume of the pores of a powder, which is penetrated by mercury at each successive pressure; the pore volume is converted into a pore size distribution. Two other methods, adsorption and air permeability, are also available that permit direct calculation of surface area. In the adsorption method, the amount of a gas or liquid solute that is *adsorbed* onto the sample of powder to form a monolayer is a direct function of the surface area of the sample. The air permeability method depends on the fact that the rate at which a gas or liquid permeates a bed of powder is related, among other factors, to the surface area exposed to the permeant. The determination of surface area is well described by the BET (Brunauer, Emmett, and Teller) equation.

The wetting behavior of powders is an also important factor for drug dissolution. If the wetting is not satisfactory, hydrophilic excipients (lactose) and surfactant (sodium lauryl sulfate or polysorbate) are combined in the powder mixtures. The contact angle is used as an index of wetting. The lower the contact angle, the better the wetting occurs.

The general appearance of a tablet with good visual identity and overall "elegance" are essential for consumer acceptance, quality control of lot-to-lot uniformity, general tablet-to-tablet uniformity, and monitoring trouble-free manufacturing. Control of the general appearance of a tablet involves the measurement of a number of attributes, such as a tablet's size, shape, color, presence or absence of and odor, taste, surface texture, physical flaws and consistency, and legibility of any identifying markings.

## 5 TABLET CHARACTERIZATION

In order to formulate the optimal tablet, various properties should be considered, including drug–excipient compatibility, flowability, lubricity, appearance, dissolution, and disintegration [2]. The prepared tablet must also meet physical specification and

quality standards according to the monograph of the pharmacopeia. In general, weight and its variation, content uniformity, thickness, hardness, friability disintegration, and dissolution should be considered for tablet validation [3]. These factors must be controlled during tablet production (in-process control) and are validated after the production to ensure the quality standards.

### 5.1 Disintegration

Immediate-release tablets should be readily disintegrated in the stomach when swallowed. This disintegration involves bursting apart the compact masses by aqueous fluid penetrating the fine pore structure of tablet. Disintegration testing is an important part of in-process control testing during production to ensure batch-to-batch uniformity, but its role in end-product testing has largely been superseded by dissolution testing because recently modified-release preparations are getting popular. It was recognized in the 1940s that tablets had to disintegrate in order for them to be bioavailable due to lack of biopharmaceutical information and primarily analytical limitations. Later, of course, in the 1950s and 1960s, the pharmaceutical scientist became aware of the importance of dissolution rates as well.

In general, for the medicinal agent in a tablet to become fully available for absorption, the tablet must first disintegrate and discharge the drug to body fluids for dissolution. The general manner in which a tablet disintegrates is as follows: (a) the tablet wets down, (b) the dissolution liquid penetrates the pore space, (c) the disintegrant absorbs water and swells, and (d) this swelling causes the tablet to break down into granules. Figure 14 shows the disintegration pathways of solid dosage forms for the dissolution and absorption of drugs. After the disintegration process, the solid dosage forms change into granules or smaller and fine particles ready for dissolution and absorption in the fluid.

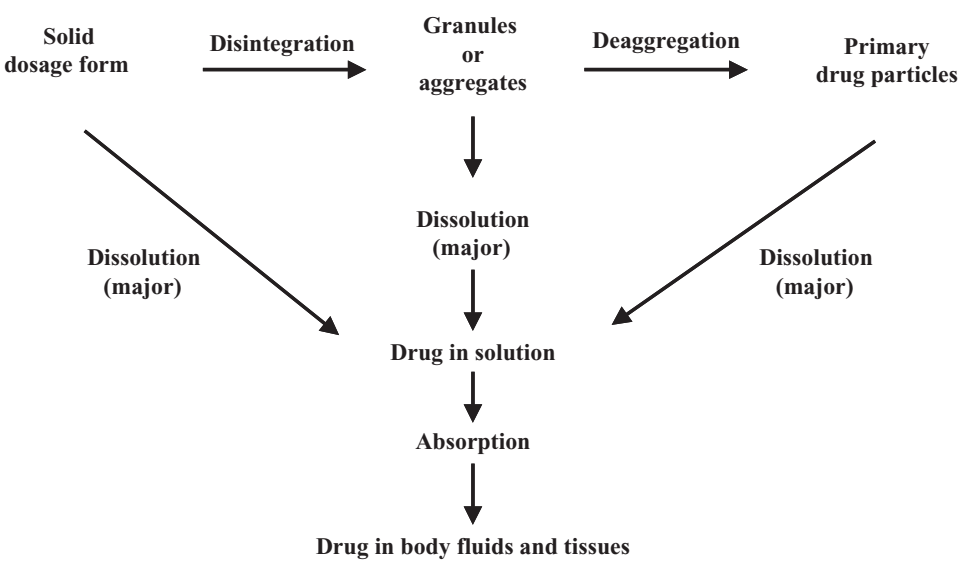


**FIGURE 14** Disintegration and dissolution pathways of solid dosage forms for absorption of drug.

Tablet disintegration is the important first step in the dissolution of the drug substance contained in immediate-release tablets but dissolution is more meaningful in case of many modified release products rather than disintegration. A number of formulation and manufacturing factors can affect the disintegration and dissolution of a tablet, including the particle size of the drug substance in the formulation; the solubility and hygroscopicity of the formulation; the type and concentration of the disintegrant, binder, and lubricant used; the manufacturing method, particularly the compactness of the granulation and the compression force used in tableting; and the in-process variables which may occur. Therefore, it is vitally important for batch-to-batch consistency to establish disintegration and dissolution standards and controls for both materials and processes.

Tablet disintegration also is important for those tablets containing medicinal agents (such as antacids and antidiarrheals) that are not intended to be absorbed but rather to act locally within the GI tract. In these instances, tablet disintegration provides drug particles with an increased surface area for localized activity within the GI tract.

It is evident that there are some correlations of physical parameters with tablet disintegration time. Figure 15 shows the correlation of water penetration force, disintegration force, disintegrant contents, and compression forces of the tablet with disintegration time of the tablet. As the water penetration force increases, the tablet disintegration force also increases, resulting in shorter disintegration time [29, 30]. The amount of disintegrant in the tablet also decreases the disintegration time. The disintegration time increases as tableting pressure increases below the critical capping pressure [18]. At very low pressures the penetration of liquid into the tablet is virtually unhindered (almost like pouring water into a breaker) but the pores will be too large to allow disintegrant swelling to cause stress and the disintegration time will decrease. Once the pores are sufficiently small, penetration of the liquid into the disintegrant becomes the limiting step, and the disintegration time will increase

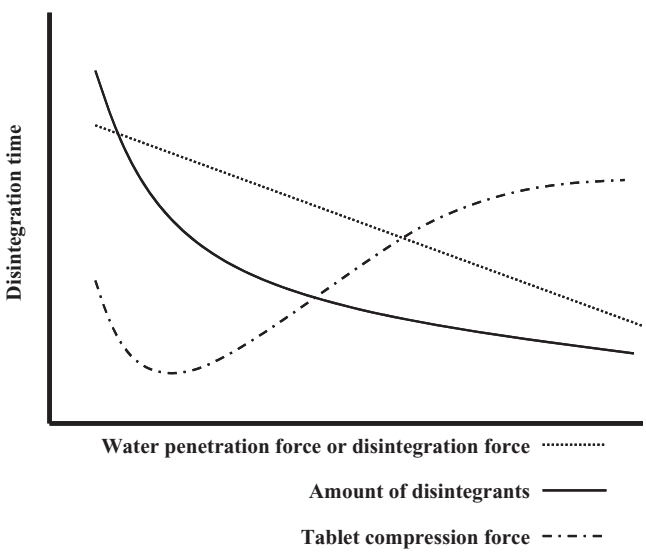


**FIGURE 15** Correlation of physical parameters on tablet disintegration time.

with increasing pressure. Disintegrants and lubricants are added to wet-granulated products after the granulation has been dried. Disintegration time increases as the amount of hydrophobic lubricant increases. The mixing time for the lubricant must be kept short because otherwise the lubricant may fluidize during the mixing step and lose part of the lubricant properties that are necessary for flow in the tablet die. If the disintegration is not satisfactory, numerous types of disintegrants are added in the tablet formulations, including starch, croscarmellose sodium, sodium starch glycolate and crospovidone known as superdisintegrants. In general, the swelling rate and water uptake are the most important properties of disintegrants.

All USP tablets must pass a test for disintegration, which is conducted in vitro using a testing apparatus. The detailed monograph for disintegration testing is described in the many pharmacopeias [2, 3].

The apparatus consists of a basket-rack assembly containing dimensions held vertically upon a 10-mesh stainless steel wire screen. During testing, a tablet is placed in each of the six tubes of the basket, and the mechanical device raises and lowers the basket in the immersion fluid at a frequency of between 29 and 32 cycles per minute, the wire screen always maintained below the level of the fluid. For uncoated tablets, buccal tablets, and sublingual tablets, water maintained at about 37 °C is used as the immersion fluid unless another fluid is specified in the individual monograph. For these tests, complete disintegration is defined as that state in which any residue of the unit, except fragments of insoluble coating or capsule shell, remaining on the screen of the test apparatus is a soft mass having no palpably firm core. Buccal tablets must disintegrate within the time set forth in the individual monograph, usually 30 min, but varying from about 2 min for nitroglycerin tablets to up to 4 h. If one or more tablets fail to disintegrate, additional tests prescribed by the USP must be performed. Enteric-coated tablets are also similarly tested, except that the tablets are permitted to be tested in simulated gastric fluid for 1 h with no sign of disintegration, cracking, or softening. They are then switched to the simulated intestinal fluid for the time stated in the individual monograph during which time the tablets disintegrate completely. If 1 or 2 of the 6 tablets fails to disintegrate completely, disintegration testing is repeated on 12 additional tablets, and not less than 16 of the total 18 tablets tested must disintegrate to meet the standards.

### 5.2 Dissolution

***Definitions*** Dissolution is the dynamic process by which drug is dissolved in a solvent (water) and is characterized by a rate (amount dissolved per unit time). In vitro dissolution testing of a tablet is very important for many reasons: It guides the formulation and product development process toward optimization of dosage forms for quality control and reliability. By conducting dissolution studies in the early stages of a product's development, the formulation compositions and manufacturing parameters are also tuned and monitored about the old and newly advanced tablet. The U.S. Food and Drug Administration (FDA) allows manufacturers to examine scale-up batches of 10% of the proposed size of the actual production batch or 100,000 dosage units, whichever is greater, by performing in vitro dissolution testing to assure bioequivalence from batch to batch and processing parameters. New drug applications (NDAs) submitted to the FDA contain in vitro dissolution data generally obtained from batches that have been used in pivotal clinical and/or bioavail-

ability studies and from human studies conducted during product development. Once the specifications are established in an approved NDA, they become official (USP) specifications for all subsequent batches and bioequivalent products. The dissolution testing is also used as a tool of SUPAC (scale-up postapproval change) and variation of equipment, location, and processing factors. In addition, dissolution testing is used as a tool to examine the short- and long-term stability of dosage forms. Release mechanism and parameters which change the dissolution are also studied. Most of all, one of main goals of in vitro dissolution testing is to provide reasonable prediction and correlation with the product's in vivo bioavailability. Differentiations in the formulations and other related variables may cause deviations from in vivo bioavailability data.

A system has been developed which relates combinations of a drug's solubility (high or low) and its intestinal permeability (high or low) as a possible basis for predicting the likelihood of achieving a successful in vivo–vitro correlation (IVIVC). The four classes based on BCS are (I): high solubility and high permeability, (II): low solubility and high permeability, (III): high solubility and low permeability, and (IV): low solubility and low permeability. In class I, dissolution testing can be used as a prognostic tool to predict in vivo biovailability.

***Equations*** Many dissolution equations are well described in the text. Most of the equation are based on the well-known Fick's law. Figure 16 shows a diagram of the concentration gradient between the matrix tablet and bulk fluid for dissolution.

From these gradient situations, the well-known Noyes–Whitney equation is given as

$$\frac{dc}{dt} = \frac{DAK}{Vh}(C_s - C_t)$$

where $dc/dt$ = rate of drug dissolution, where $dQ/dt = V\ dc/dt$

$V$ = volume of dissolution fluid

$D$ = diffusion rate constant

$A$ = surface area of dosage forms

$C_s$ = concentration of drug in stagnant layer

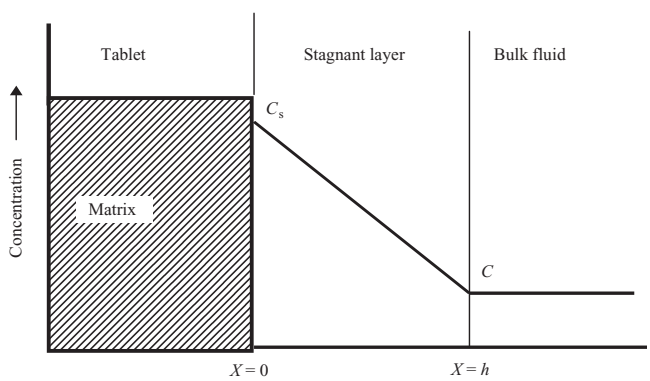


**FIGURE 16** Concentration gradient from the tablet between matrix and bulk fluid for dissolution. Cs, drug solubility; C, uniform concentration; h, thickness of stagnant film; X = diffusional path length.

$C_t$ = concentration of drug in bulk fluid at given time
$K$ = partition coefficient
$h$ = thickness of stagnant layer

If the bulk volume is large and the concentration of drug in the fluid is much lower than the drug solubility ($C_s >> C_t$), it is regarded as a sink condition. In this case, the equation is much simpler and the dissolution behaviors continuously occur because the chemical potential ($C_s - C_t$) approximates drug solubility ($C_s$).

In a matrix tablet, the following Higuchi equations are given depending on the polymeric structures of the homogenous and porous matrix [20]:

$$Q = \begin{cases} [D(2A - C_S)C_S t]^{1/2} & \text{for homogenous matrix} \\ [D\varepsilon(2A - \varepsilon C_S)C_S t]^{1/2}/\tau & \text{for porous matrix} \end{cases}$$

In the dissolution of granular powders, the Hixson–Crowell equation is also established as

$$Q_0^{1/3} - Q^{1/3} = kt \quad \text{for granular powders}$$

***Testing Method*** In addition to formulation and manufacturing controls, the method of dissolution testing also must be controlled to minimize important variables such as paddle rotational speed, vibration, and disturbances by sampling probes. The USP includes seven apparatus designs for drug release and dissolution testing of immediate-release oral dosage forms, extended-release products, enteric-coated products, and transdermal drug delivery devices:

Apparatus I: rotating basket method, 25–150 rpm (100 rpm)
Apparatus II: rotating paddle method, 25–150 rpm (50 rpm)
Apparatus III: reciprocating cylinder method, inner tube (5–40 dips/min), outer tube (300 mL)
Apparatus IV: flow-through method: 4–16 mL/min
Apparatus V: paddle over disc
Apparatus VI: cylinder method
Apparatus VII: reciprocating holder method

Detailed guidelines for dissolution testing are described in monographs of many pharmacopeias. The USP apparatus I and USP apparatus II are used principally for tablet dissolution testing. In USP apparatus I, the dosage unit is placed inside the basket. In USP apparatus II, the dosage unit is placed on the bottom in the vessel. In each test, a volume of the dissolution medium (500–900 mL in general) is placed in the vessel and allowed to come to 37 ± 0.5 °C. Then the stirrer is rotated at the specified speed (50–200 rpm). The samples of the medium are withdrawn for analysis of the proportion of drug dissolved. The tablet or capsule must meet the stated monograph requirement for rate of dissolution. For example, "not less than 85% of the labeled amount is dissolved in 30 minutes in case of immediate release tablet." In a floating tablet, the sinker can be used in the paddle method.

***Variables Affecting Dissolution*** In general, the dissolution profiles are highly dependent on the physicochemical properties, formulation, processing parameters, and testing conditions [2]. The physicochemical properties include particle size, surface area, crystal habit and polymorphism, solubility, molecular size, salt formation, p$K_a$, hydration, wetting, and surface tension. Physical factors such as viscosity, density, flocculation, and agglomeration are also considered. The formulation factors are also of importance. The amount and type of excipients and type of dosage forms play a key role in modifying dissolution behaviors. For example, Figure 17 gives the effect of lubricant and its mixing time on the dissolution rate of drugs. The presence of lubricant and its mixing time significantly changed tablet dissolution [31]. For poorly soluble drug, numerous pharmaceutical methods have been utilized to increase the dissolution rate of drug, including micronization, amorphous crystallization, spray drying, inclusion complex, microemulsion, and solid dispersion.

The processing parameters in the tablet preparation also change the dissolution, including temperature, mixing, milling, rotation speed, solvent, hardness, and surface area. The testing conditions are also important in modifying the dissolution of tablet. Therefore, the testing conditions are well defined by the regulations in the many pharmacopeias. The testing conditions include pH of the fluid, temperature, ionic strength, common ion effect, type of apparatus, rotation speed, volume size of dissolution fluid, analytical conditions, aeration, sample treatment, and mainly composition of dissolution media. Table 17 provides some of the media compositions suggested for in vitro dissolution testing of the tablet. These modified dissolution media can be used to achieve dissolution of drug under simulated in vivo conditions.

***In Vitro–In Vivo Correlation*** In vitro dissolution testing can provide a reasonable prediction of the product's in vivo bioavailability. For a high-solubility and high-permeability drug (class I), an IVIVC may be expected if the dissolution rate is slower than the rate of gastric emptying (the rate-limiting factor). In a low-solubility and high-permeability drug, drug dissolution may be the rate-limiting step for drug absorption and an IVIVC may be expected. In a high-solubility and low-

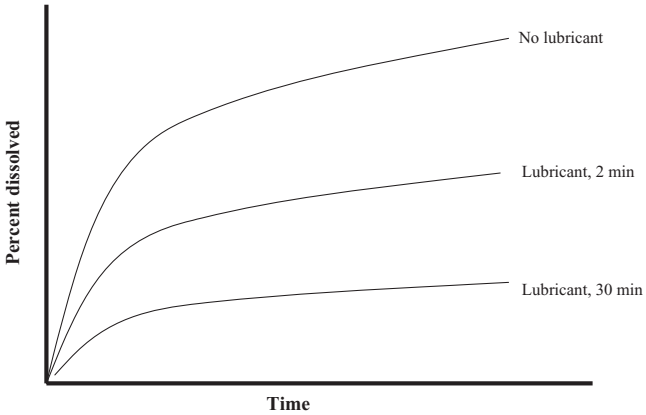


**FIGURE 17** Effect of lubricant and its mixing time on dissolution rate of drugs.

**TABLE 17 Suggested Media Compositions for In Vitro Dissolution Testing**

| Medium Type | Codes | Dissolution compositions |
|---|---|---|
| Water | W | Purified water |
| Gastric fluid (pH 1.2) | SGF | HCl–NaCl buffer |
| | SGF/TW | HCl–NaCl buffer/Tween 80 (1%) |
| | SGF/SLS | HCl–NaCl buffer/sodiun lauryl sulfate (1%) |
| | SGF/PEP | HCl–NaCl buffer/pepsin (0.32%) |
| Intestinal fluid (pH 6.8) | SIF | $KH_2PO_4$ buffer |
| | SIF/TW | $KH_2PO_4$ buffer/Tween 80 (1%) |
| | SIF/SLS | $KH_2PO_4$ buffer/sodiun lauryl sulfate (1%) |
| | SIF/Pan | $KH_2PO_4$ buffer/pancreatin (1%) |
| | SIF/fasted | 0.2 *M* $KH_2PO_4$ buffer (3.9 g)/3 m*M* Na-taurocholate, 75 m*M* lecithin/KCl (7.7 g)/NaOH (qs)/water (qs) |
| Intestinal fluid (pH 5.0) | SIF/fed | Acetic acid (8.65 g)/15 m*M* Na-taurocholate/3.75 m*M*, lecithin/KCl (15.2 g)/NaOH (qs)/water (qs) |

permeability drug, permeability is the rate-controlling step and only a limited IVIVC may be possible. In a drug with low solubility and low permeability, significant problems would be likely for oral drug delivery.

However, the in vivo GI condition is very complicated in terms of complex physiology and absorption process and is not simulated by the simple in vitro dissolution conditions [32]. Moreover, in vivo conditions are also complicated by food composition, type and composition of dosage formulations, relative rates of permeation, GI transit time, site of absorption, complexity of GI fluids such as pH, enzymes, bile and mucin, rate and capacity of metabolism by intestinal and hepatic enzymes, ethical difference, and patient conditions such as mood, disease, bed rest, and fasting volume of fluid given. For these reasons, in vitro dissolution is not always correlated with in vivo absorption, especially low-soluble and low-bioavailable drugs. To use dissolution testing as a prognostic tool for in vivo bioavailability, the dissolution fluid is simulated with in vivo GI condition as possible, compromising the biorelevant composition of dissolution fluid (see Table 17), gradient pH, proper stirring rate, and addition of lipid, enzyme, and surfactants.

For bioequivalence, dissolution profiles of two tablets are often compared. In this case, the difference factor and similarity factor are considered, defined by the equations

$$f_1 = \frac{\sum |R_t - T_t|}{\sum R_t} \times 100$$

$$f_2 = 50 \log \left\{ \left[ 1 + \frac{1}{n} \sum (R_t - T_t)^2 \right]^{-0.5} \times 100 \right\}$$

where $R_t$ and $T_t$ are the cumulative percents dissolved at each of the selected $n$ time points.

The difference factor $f_1$ is proportional to the average difference between the two profiles while the similarity factor $f_2$ is inversely proportional to the average squared difference between the two profiles and measures the closeness between the two profiles. The two dissolution profiles are identical if $f_2 = 100$. An average difference of 10% at all measured time points results in a $f_2$ value of 50. The FDA guideline states that $f_2$ values of 50–100 indicate similarity between two dissolution profiles.

### 5.3 Weight Variation

With a tablet designed to contain a specific amount of drug in a specific amount of tablet formula, the weight of the tablet being made is routinely measured to help ensure that the tablet contains the proper amount of drug. The quantity of fill placed in the die of a tableting press determines the weight of the resulting tablet. The volume of fill is adjusted to yield tablets of desired weight and content. The depth of fill in the tablet die must be adjusted to hold a predetermined volume of powder or granulation. Each tablet weight should be calculated if the amount of drug and other excipients such as diluent, disintegrant, and binder are decided.

During production, sample tablets are periodically removed for visual inspection and automated physical measurement known as in-process control (IPC). The USP provides some guidelines for weight variation. In the test, 10 uncoated tablets are weighed individually and the average weight is calculated. The tablets are assayed and the content of active ingredient in each of the 10 tablets is calculated assuming homogeneous drug distribution.

### 5.4 Hardness or Breaking Strength

Tablets require a certain amount of strength, or hardness and resistance to friability, to withstand mechanical shocks of handling in manufacture, packaging, and shipping. Adequate tablet hardness and resistance to powdering and friability are necessary requisites for consumer acceptance while immediate-release tablets should readily disintegrate in the stomach as quick possible. For this reason, the relationship of harness to tablet disintegration and drug dissolution has been described.

Tablet hardness has been defined as the force required to break a tablet in a diametric compression test. To perform this test, a tablet is placed between two anvils, and the crushing strength that just causes the tablet to break is recorded. Hardness is thus sometimes termed the tablet crushing strength. It is not unusual for a tableting press to exert as little as 3000 lb and as much as 40,000 lb of force in the production of tablets. Generally, the greater the pressure applied, the harder the tablets, although the formulation composition and manufacturing process may also change tablet hardness. In general, tablets should be sufficiently hard to resist breaking during normal handling or transportation but have no problem in disintegrating and/or dissolving after swallowing.

The hardness of a tablet, like its thickness, is a function of the die fill and compression force. At a constant die fill, the hardness values increase and thickness decreases as additional compression force is applied. This relationship holds up to a maximum value for hardness and a minimum value for thickness beyond which increases in pressure cause the tablet to laminate or cap, thus destroying the integrity of the tablet. At a constant compression force (fixed distance between upper

and lower punches), hardness increases with increasing die fills and decreases with lower die fills. The amount and mixing time of lubricants and excipients can affect tablet hardness. Large tablets require a greater force to cause fracture and are therefore "harder" than small tablets.

Special dedicated hardness testers or multifunctional systems are used to measure the degree of force (in kilograms, pounds, or arbitrary units) required to break a tablet. Devices to test tablet hardness include the Monsanto tester, the Strong-Cobb tester, the Pfizer tester, the Erweka tester, and the Schleuniger tester. A force of about 4 kg is considered the minimum requirement for a satisfactory tablet. Multifunctional automated equipment can determine tablet weight, hardness, thickness, and diameter. Unfortunately, these testers do not produce uniform results for the same tablet due to operator variation, lack of calibration, spring fatigue, and manufacturer variation.

### 5.5 Friability

Tablet hardness is not an absolute indicator of tablet strength since some formulations, when compressed into very hard tablets, tend to "cap" on attrition, losing their crown portions. Another measure of a tablet's strength is tablet friability. More powders, chips, and fragments can be produced during friability test if the tablet lacks proper strength and is manufactured in dirty processes during coating and packaging. The high friability causes lacks of elegance and consumer acceptance and even weight variation or content uniformity problems.

A tablet's durability may be determined using friability tester like Varian Friabilator testng apparatus. This apparatus determines the tablet's friability, or its tendency to crumble, by allowing it to roll and fall within the rotating apparatus. Normally, a preweighed tablet sample is placed in the friabilator, which is then operated for 100 revolutions. The tablets are then dusted and reweighed. Any loss in weight is determined. Resistance to loss of weight indicates the tablet's ability to withstand abrasion in handling, packaging, coating, and shipment. Compressed tablets that lose a maximum of not more than 0.5–1% of their weight are generally considered acceptable [3].

### 5.6 Content Uniformity

Tablet weight cannot be used as a potency indicator of its potency, except perhaps when the active ingredient is 90–95% of the total tablet weight. In tablets with smaller dosages, a good weight variation does not ensure good content uniformity, but a large weight variation precludes good content uniformity. The weight variation test would be a satisfactory method of determining the drug content uniformity of tablets. The content uniformity of the tablet is more important since the potency of tables is expressed on labels in terms of grams, milligrams, or micrograms. The content uniformity of tablets can be varied by three factors: (1) nonuniform distribution of the drug substance throughout the powder mixture or granulation, (2) segregation of the powder mixture or granulation during the various manufacturing processes, and (3) tablet weight variation.

By the USP method, 10 dosage units are individually assayed for their content uniformity according to the assay method described in the individual monograph.

The requirements for content uniformity are met if the amount of active ingredient in each dosage unit lies within the range of 85–115% of the label claim and the relative standard deviation is less than 6.0%. If one or more dosage units do not meet these criteria, additional tests as prescribed in the USP are required. Official compendia or other standards provide an acceptable potency range around the label potency. For highly potent, low-dose drugs such as digoxin, this range is usually not less than 90% and not more than 110% of the labeled amount. For most other larger dose drugs in tablet form, the official potency range that is permitted is not less than 95% and not more than 105% of the labeled amount.

### 5.7 Tablet Thickness

At a constant compressive load, tablet thickness varies with changes in die fill, particle size distribution and packing of the particle mix being compressed, and tablet weight, while with a constant die fill, thickness varies with variations in compression forces. The thickness of individual tablets may be measured with a micrometer, which permits accurate measurement and provides information on the variation between tablets. Other techniques employed in production control involve placing 5 or 10 tablets in a holding tray, where their total crown thickness may be measured with a sliding caliper scale. Tablet thickness should be controlled within ±5% variation of a standard value. Any variation in tablet thickness within a particular lot of tablets or between manufacturer's lots is not appropriate for consumer acceptance of the product. In addition, tablet thickness must be also controlled to facilitate packaging.

### 5.8 Tablet Shape and Size

The shape of the tablet alone can influence the choice of tableting machine used. Figure 18 shows some representative tablet shapes. The size, shape, and thickness

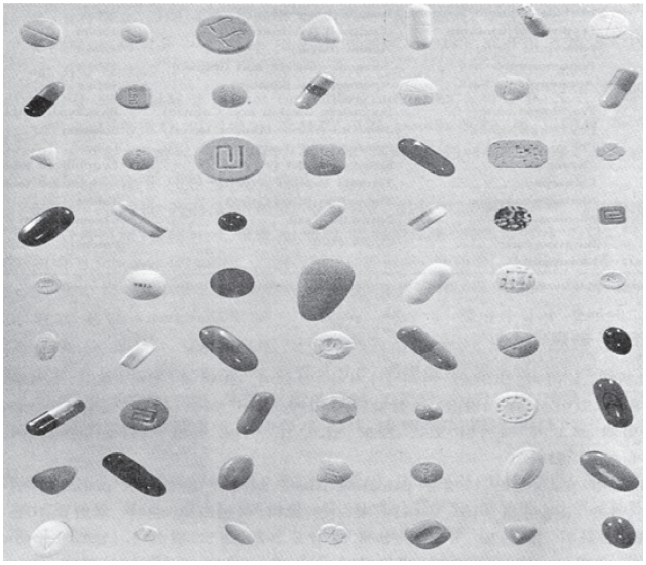

**FIGURE 18** Example tablet shapes.

are very changeable. Shaped tablets requiring slotted and sophisticated punches must be run at slower speeds to avoid manufacturing disorder as compared with round tablets using conventional punches. Because of the nonuniform forces involved during compression, the more convex the tablet surface, the more likely it is to cause capping or laminating problems. The more complicated shaped tablet requires the use of a slower tableting machine or one with precompression capabilities. The size and shape of the tablet are governed by the choice of tableting machine, the best particle size for the granulation, production lot sizes, and the best type of tablet processing, packaging operation, and cost to produce the tablet.

## REFERENCES


1. Kibbe, A. H. (2000), *Handbook of Pharmaceutical Excipients*, 3rd ed., American Pharmaceutical Association and Pharmaceutical Press, London.
2. Gibson, M. (2001), *Pharmaceutical Preformulation and Formulation, A Practical Guide from Candidate Drug Selection to Commercial Dosage Form*, HIS Health Group, Denver, CO.
3. Allen, L. V., Popovich, N. G., and Angel, H. C. (2005), *Pharmaceutical dosage forms and drug delivery systems*, Lippincott Williams & Wilkins, Baltimore, MD, USA.
4. Mountfield, R. J., Senepin, S., Schleimer, M., Walter, I., and Bittner, B. (2000), Potential inhibitory effects of formulation ingredients on intestinal cytochrome P450, *Int. J. Pharm.*, 211, 89–92.
5. Cornaire, G., Woodley, J., Hermann, P., Cloarec, A., Arellano, C., and Houin, G. (2004), Impact of excipients on the absorption of P-glycoprotein substrates in vitro and in vivo, *Int. J. Pharm.*, 278, 119–131.
6. Wang, S.-W., Monagle, J., McNulty, C., Putnam, D., and Chen, H. (2004), Determination of P-glycoprotein inhibition by excipients and their combinations using an integrated high-throughput process, *J. Pharm. Sci.*, 93, 2755–2767.
7. Karsa, D. R., and Stephenson, R. A. (1995), *Excipients and Delivery Systems for Pharmaceutical Formulations*, The Royal Society of Chemistry, Cambridge, pp. 1–34.
8. Farber, L., Tardos, G. I., and Michaels, J. N. (2003), Evolution and structure of drying material bridges of pharmaceutical excipients: Studies on a microscope slide, *Chem. Eng. Sci.*, 58, 4515–4525.
9. Cao, Q.-R., Choi, Y.-W., Cui, J.-H., and Lee, B.-J. (2005), Formulation, release characteristics and bioavailability of novel monolithic hydroxypropylmethylcellulose matrix tablets containing acetaminophen, *J. Controlled Release*, 108, 351–361.
10. Jackson, K., Young, D., and Pant, S. (2000), Drug-excipient interactions and their affect on absorption, *PSTT*, 3(10), 336–345.
11. Patel, H., Stalcup, A., Dansereau, R., and Sakr, A. (2003), The effect of excipients on the stability of levothyroxine sodium pentahydrate tablets, *Int. J. Pharm.*, 264, 35–43.
12. Verma, R. K., and Garg, S. (2004), Compatibility studies between isosorbide mononitrate and selected excipients used in the development of extended release formulations, *J. Pharm. Biomed. Anal.*, 35, 449–458.
13. Young, W. R. (1990), Accelerated temperature pharmaceutical product stability determinations, *Drug Dev. Ind. Pharm.*, 16(4), 551–569.
14. Waterman, K. C., and Adami, R. C. (2005), Accelerated aging: Prediction of chemical stability of pharmaceuticals, *Int. J. Pharm.*, 293, 101–125.

15. Simon, P., Veverka, M., and Okuliar, J. (2004), New screening method for the determination of stability of pharmaceuticals, *Int. J. Pharm.*, 270, 21–26.
16. Florence, A. T. and Attwood, D. (1998), *Physicochemical principles of pharmacy*, 3rd ed., Macmillan, London, pp. 5–35, 101–151.
17. Hiestand, E. N. (1989), The basis for practical applications of the tableting indices, *Pharm. Technol. Int.*, 8, 54–66.
18. Carstensen, J. T. (1993), *Pharmaceutical Principles of Solid Dosage Forms*, Technomic Publishing, Lancaster, PA.
19. Brittain, H. G. (1995), *Physical Characterization of Pharmaceutical Solids*, Marcel Dekker, New York.
20. Sinko, P. J. (2006), *Physical Pharmacy and Pharmaceutical Sciences*, Lippincott Williams and Wilkins, Baltimore, MD.
21. Carson, J. W. and Wilms, H. (2006), Development of an international standard for shear testing, *Powder Technol.*, 167, 1–9.
22. Hedenus, P., Mattsson, M. S., Niklasson, G. A., Camber, O., and Ek, R. (2000), Characterization of instantaneous water absorption properties of pharmaceutical excipients, *Int. J. Pharm.*, 141, 141–149.
23. Hasegawa, S., Hamaura, T., Furuyama, N., Kusai, A., Yonemochi, E., and Terada, K. (2005), Effects of water content in physical mixture and heating temperature on crystallinity of troglitazone-PVP K30 solid dispersions prepared by closed melting method, *Int. J. Pharm.*, 302, 103–112.
24. Mullarney, M. P., Hancock, B. C., Carlson, G. T., Ladipo D. D., and Langdon, B. A. (2003), The powder flow and compact mechanical properties of sucrose and three high-intensity sweeteners used in chewable tablets, *Int. J. Pharm.*, 257, 227–236.
25. Meyer, K., and Zimmermann, I. (2004), Effect of glidants in binary powder mixtures, *Powder Technol.*, 139, 40–54.
26. Hancock, B. C., and Zografi, G. (1997), Characteristics and significance of the amorphous state in pharmaceutical systems, *J. Pharm. Sci.*, 86(1), 1–12.
27. Willart, J., Lefebvre, J., Danède, F., Comini, S., Looten, P., and Descamps, M. (2005), Polymorphic transformation of the G-form of d-sorbitol upon milling: Structural and nanostructural analyses, *Solid State Communi.*, 135(8), 519–524.
28. Adolfsson, A., and Nystrom, C. (1996), Tablet strength, porosity, elasticity and solid state structure of tablets compressed at high loads, *Int. J. Pharm.*, 132, 95–106.
29. Caramella, C., Colombo, P., Conte, U., Ferrari, F., and Manna, A. L. (1986), Water uptake and disintegration force measurements: Towards a general understanding of disintegration mechanism, *Drug Dev. Ind. Pharm.*, 12, 1749–1766.
30. Pourkavoos, N., and Peck, G. E. (1993), The effect of swelling characteristics of superdisintegrants on the aqueous coating solution penetration into the tablet matrix during the film coating process, *Pharm. Res.*, 10(9), 1363–1371.
31. Lerk, C. F., Bolhuis, G. K., Smallenbroek, A. J., and Zuurman, K. (1982), Interaction of tablet disintegrations and magnesium stearate during mixing II. Effect on dissolution rate, *Pharm. Acta Helv*, 57, 282–286.
32. Gundert-Remy, U., and Moller, H. (1990), *Oral Controlled Release Products, Therapeutic and Biopharmaceutical Assessment*, Wissenschaftliche Verlagsgesellschaft mbH, Stuttgart, Germany, pp. 155–173.

# ROLE OF PREFORMULATION IN DEVELOPMENT OF SOLID DOSAGE FORMS

Omathanu P. Perumal and Satheesh K. Podaralla
*South Dakota State University, Brookings, South Dakota*

**Contents**

1 Introduction
2 Physical/Bulk Characteristics
  2.1 Crystallinity and Polymorphism
  2.2 Hydrates/Solvates
  2.3 Amorphates
  2.4 Hygroscopicity
  2.5 Particle Characteristics
  2.6 Powder Flow and Compressibility
3 Solubility Characteristics
  3.1 p$K_a$ and Salt Selection
  3.2 Partition Coefficient
  3.3 Drug Dissolution
  3.4 Absorption Potential
4 Stability Characteristics
  4.1 Solid-State Stability
  4.2 Solution-State Stability
  4.3 Drug–Excipient Compatibility
5 Conclusions
  References

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

## 1 INTRODUCTION

The advent of combinatorial chemistry and high-throughput screening (HTS) has exponentially increased the number of compounds synthesized and screened during the drug discovery phase. However, the overall efficiency of the drug discovery process is still exceedingly low (only 1 in 10,000 makes it to the market). Drug discovery is mostly driven by "activity screens" with little emphasis on "property screens." This is exemplified by the fact that 40% of attrition in drug discovery and development is attributed to poor biopharmaceutics and pharmacokinetic properties [1], which in turn are related to poor physicochemical properties. As a result, pharmaceutical companies have started to redesign their strategies by including property screens quite early in the discovery stage [2]. Preformulation is the study of fundamental properties and derived properties of drug substances. In other words, preformulation is the first opportunity to learn about the drug's physicochemical properties from the perspective of transforming a biologically active molecule to a "druggable" molecule. The type and extent of preformulation activities vary in a discovery and generic setting.

The main goal of a drug discovery program is to develop an orally deliverable molecule for obvious reasons of ease of manufacture and convenience of drug administration. More than 75% of the drug products in the market are oral formulations, of which more than half are solid dosage forms [3]. The "rule of five" developed by Christopher Lipinski [4] is one of the "physicochemical screens" to weed out molecules with poor physicochemical properties very early in the drug discovery process. According to Lipinski's rule, a drug will show poor oral absorption if it does not conform to any of the two physicochemical requirements listed in Table 1. The rule of five is applicable only to small molecules and it relates the chemical properties of the drug to its solubility and permeability characteristics. During the initial stages of drug discovery, the preformulation activities are mainly focused on developing a water-soluble compound for early activity studies and preclinical testing in animals. At this stage, the preformulation scientist is faced with the challenge of working with a limited quantity of compound (few milligrams) for testing a long list of physicochemical parameters. On the other hand, developing preclinical formulations can be quite a daunting task given the fact that toxicological studies require a high dose of drug (10–100 times above the effective dose) to be delivered in a small volume of the formulation. Preformulation activities increase as the molecule proceeds through the development phase. The "discovery and development phar-

**TABLE 1 Lipinski Rule of Five for Orally Active Compounds**

| Physicochemical Parameter | Lipinski rule |
|---|---|
| Molecular weight | Not more than 500 Da |
| log *P* | Not more than 5 |
| Hydrogen bond donors | Not more than 5 hydrogen bond donors expressed as the sum of OH's and NH's |
| Hydrogen bond acceptors | Not more than 10 expressed as the sum of OH's and NH's |

maceutics" documentation forms a significant portion of the investigational new drug application (IND) application and new drug application (NDA) filed to the U.S. Food and Drug Administration (FDA). In a generic setting, preformulation studies are mainly focused on developing a formulation that is bioequivalent to the innovator's product with the main objective of filing an abbreviated new drug application (ANDA). A strong preformulation team can generate intellectual property in the form of new salts, solid-state forms, or new stabilized formulations of the drug for an innovator and/or a generic manufacturer.

In the present chapter, the discussion is mainly focused on preformulation testing for oral solid dosage forms in a drug discovery setting. The chapter address the following goals of preformulation: (i) to gain knowledge about the physicochemical characteristics of the drug, (ii) to define the physical characteristics of the drug, (iii) to understand the stability characteristics of the drug, and (iv) to determine the compatibility of excipients with the drug. In this chapter, we have grouped all the parameters under three sections and discussed in a logical sequence for the convenience of the reader.

## 2 PHYSICAL/BULK CHARACTERISTICS

The bulk or physical characteristics of a drug substance are mainly dictated by its solid-state properties. Purity of the drug substance is a fundamental property that is characterized at the beginning of preformulation studies. In the initial stages of drug development, the drug is usually not very pure. Nevertheless, it is essential to know the purity of the material at hand using simple measurements such as melting point. This would serve to set drug specifications during later stages of drug development. Differential scanning calorimetry (DSC) requires very little sample (1–5 mg) and is a useful tool to estimate the purity of the compound. The drug sample is heated in a crucible, where the difference in heat between the sample and a reference crucible is seen as an endotherm or exotherm in the thermogram depending on whether heat is taken up or given up, respectively, by the sample. The integrated area under the endotherm or exotherm gives a measure of the heat or enthalpy involved in this process. Melting is seen as an endothermic event and the purity of the sample will govern the peak position, shape, sharpness, and heat of fusion ($\Delta H_f$). DSC is sensitive in detecting impurities to the extent of 0.002 mol % [5]. The DSC findings should be substantiated by a stability-indicating high-performance liquid chromatography (HPLC) assay. On the other hand, thin-layer chromatography (TLC) may be used to qualitatively detect the number of impurities in the drug sample. Impurity profiling is an important aspect of the drug development process, particularly for optimizing the synthetic process. The impurities can originate from many sources, including starting materials, intermediates, synthetic processes, or degradation reactions [6]. The regulatory guidelines stipulate that any impurity >0.05% of total daily dose (for drugs with a dose <2 g/day) or >0.15% of total daily dose (for drugs with a dose >2 g/day) should be evaluated for its safety [6]. Organoleptic properties such as color, taste, and odor are assessed qualitatively to set bulk drug specifications. If the drug has an unacceptable taste or odor, the chemistry group is advised to make a suitable salt form of the drug.

## 2.1 Crystallinity and Polymorphism

The majority of the drugs exist in crystalline form and are characterized by their crystal habit and crystal lattice. The crystal habit describes the external morphology of the crystal, including shape and size, while the crystal lattice describes the internal arrangement of molecules in the crystal (Figures 1*a* and *b*). Drug molecules arrange in more than one way in a crystal, and this difference in the internal arrangement of crystals is known as polymorphism. The polymorphs have the same elemental composition but differ in their physical, chemical, thermodynamic, stability, and spectroscopic properties. A crystal lattice represents the space in which molecules arrange in different ways. Organic molecules arrange in one or more of the seven possible crystal systems: triclinic, monoclinic, orthorhombic, tetragonal, rhombohedral, hexagonal, and cubic [7]. Each crystal system is characterized by its three-dimensional geometry and angles between the different crystal faces. The crystal lattice geometry is obtained using single-crystal X-ray diffractometry (XRD) and the details can be found elsewhere [7]. The difference in the crystal lattice of a drug arises as a result of the difference in packing of the molecules if the molecules are conformationally rigid (e.g., chlordiazepoxide) or due to the differences in conformation for flexible drug molecules (e.g., piroxicam). Although polymorphs differ in their internal crystal lattice, it may not be necessarily reflected in their external crystal habit (Figure 1*b*). In other words, a drug can exist in different crystal habits without any change in the internal crystal lattice (isomorphs).

Crystal habit is mainly dependent on crystal growth conditions [8]. For example, Figure 1*a* shows two different crystal habits for a given crystal lattice. A prismatic crystal habit will result if the growth is equal in all directions, while plates are formed if the growth is slow in one direction. Alternatively, needle-shaped crystals (acicular) are formed when the growth is slow in two directions. Thus, the crystal habits can

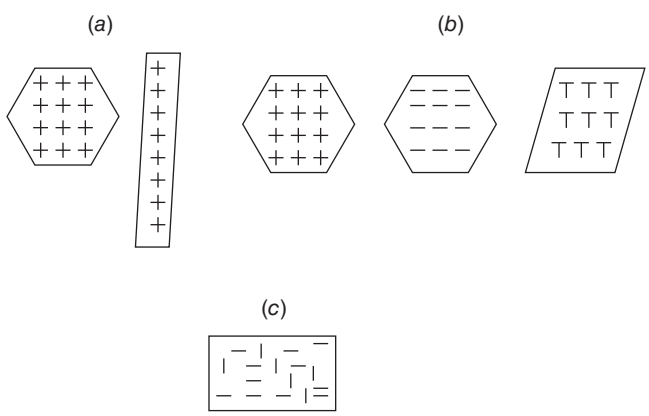


**FIGURE 1** Schematic of crystal habits, polymorphs and amorphous drug forms. (*a*) Two crystal habits are shown. The internal crystal lattice is the same while the external morphology is different. (*b*) In a crystal the molecules are arranged in a regular fashion. However, the arrangement may vary depending on how the molecules orient themselves in the internal crystal lattice. The internal crystal lattice is different in all the three polymorphic forms. The polymorphs may or may not differ in their external morphology. (*c*) Random arrangement of molecules in amorphous form.

be altered without any change in the internal crystal lattice by varying the crystallization conditions. The polarity of the crystallizing solvent mainly influences the crystal habit by preferentially adsorbing to one surface of the crystal face. Similarly, surfactants or additives are added to the crystallization medium to prevent or promote the growth of a specific crystal habit [8]. Crystal habits mainly differ in physicomechanical properties such as packing, flow property, compressibility, and tablettability. Acetaminophen crystallizes as polyhedral crystals when crystallized from water and as plates when crystallized from ethanol–water (Figure 2*a*). Both these crystal habits are isomorphic [9], that is, have the same internal crystal arrangement, since their melting points and heats of fusion were similar (melting point 178 °C and $\Delta H_f$ = 177 kcal/mol). The polyhedral crystals have better flow and

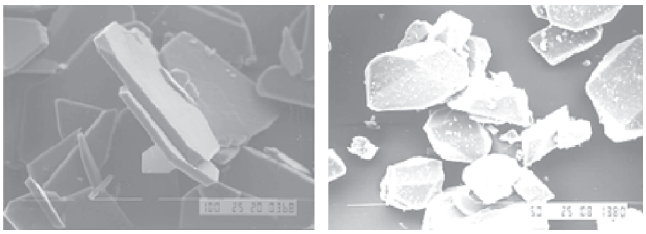

(*a*)

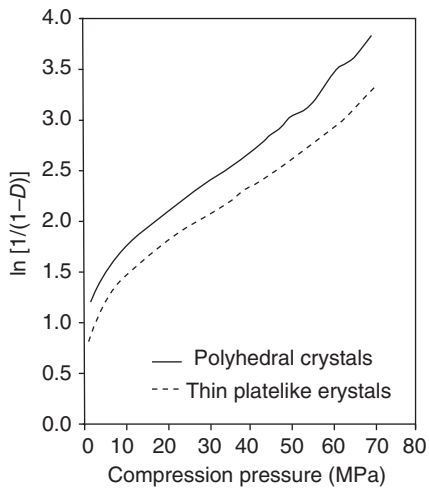


(*b*)

**FIGURE 2** Difference in crystal habit of acetaminophen and resultant difference in compressibility (*a*) Acetaminophen crystallizes as either platy crystals or polyhedral crystals depending on the solvent of crystallization. Both crystal habits have the same internal crystal lattice since they showed the same melting point. (*b*) Difference in compression behavior of two crystal habits. The *x* axis represents the compression pressure while the *y* axis represents the densification of the drug sample on compression. This plot is known as Heckel plot. The polyhedral crystal habit shows a higher densification implying better compressibility than plate like crystals. [From Garekani, H. A., Ford, J. L., Rubinstein, M. H., and Rajabi-Sahboomi, A. R., *International Journal of Pharmaceutics*, 187, 77–89, 1999. Reproduced with permission from Elsevier.)

compression properties than platy crystals, which were brittle and fragmented during tableting (Figure 2*b*). Crystal habits are characterized using optical and electron microscopy, but their internal crystal lattice should be confirmed using DSC, XRD, and spectroscopic techniques.

Polymorphs are generated by crystallizing the drug from various solvents. The solvents are usually those that are used in the synthesis and purification of the bulk drug but may also include solvents used in drug formulations [10]. By convention, polymorphs are named based on their order of discovery, such as forms I, II or A, B or α, β. In general form I is considered the most stable and least soluble form, while form II is considered the more soluble and least stable form. The least stable and more soluble polymorphic form is usually called the metastable form. They are not "unstable" but "metastable," because the least stable form can remain stable provided the conditions are controlled to prevent its conversion to the more stable polymorphic form. Polymorphs are characterized by their solubility and stability differences with respect to temperature (Figure 3). Thermodynamically, polymorphs are classified as enantiotropic or monotropic depending on their thermal reversibility from one form to another [11]. Enantiotropic polymorphs are reversible polymorphs, where one form (form I) is more stable at higher temperature, while the other form (form II) is stable at lower temperature. They are characterized by a transition point ($T_s$) below the melting points of both forms (Figure 3*a*). The transition point represents the temperature at which one form converts to another. In the temperature–solubility curve, this is represented by the intersection of the solubility curves of both forms; that is, at the transition temperature, both polymorphs have the same solubility. As shown in Figure 3*a*, form II can convert to form I at a temperature above $T_s$, while form I can convert to form II at a temperature below $T_s$. On the other hand, monotropic polymorphs are not reversible but can only convert from the metastable form to the stable form. Here, $T_s$ is higher than the melting point of both forms (Figure 3*c*). Both forms are stable in the entire temperature range below $T_s$.

The different polymorphs are generated based on their solubility differences in a given solvent. According to Ostwald's rule [12], the least stable or highly energetic form (form II, or metastable) will precipitate out first from a supersaturated solution followed by the stable or less energetic form (form I). Supersaturation is achieved by antisolvent addition or by altering the temperature. So, if the initial precipitate is separated rapidly, it would have predominantly the metastable form. Alternatively, the stable form can be melted and rapidly cooled to crystallize the metastable form. A stable or metastable polymorphic form is also used as a "seed" to preferentially grow and isolate the desired form during drug crystallization [13]. Several rules have been proposed to differentiate enantiotropic and monotropic polymorphs [11, 13]. A simple way to differentiate enantiotropic and monotropic polymorphs is the use of the heat–cool cycle in DSC [11]. As shown in Figure 3*b*, the enantiotropic polymorph is characterized by the appearance of solid–solid endothermic transition of form II to I followed by melting of form I. On cooling the melt of form I followed by reheating, the same thermogram is regenerated, proving the reversibility of the polymorphs. In monotropic polymorphs (Figure 3*d*), the thermogram is characterized by melting of metastable (form II) and recrystallization to form I followed by melting of form I. On cooling and reheating the sample, the transition and recrystallization peaks are not seen, indicating the irre-

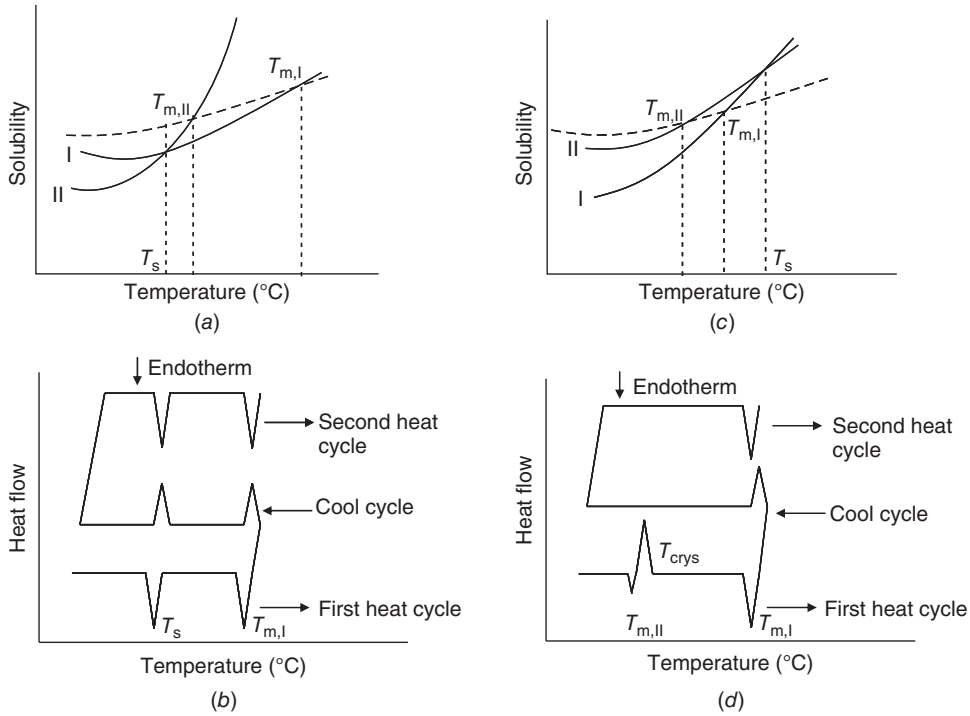


**FIGURE 3** Difference between enantiotropic and monotropic polymorphs. (*a*) Solubility of enantiotropic polymorphs as function of temperature. The dotted line indicates the melting curve. Form I is less soluble below the transition temperature (*T*s), while form II is more soluble above *T*s. Form I has a higher melting point (*T*m,I) than form II (*T*m,II). Below *T*s, form I is converted to form II and above *T*s form II coverts to form I. (*b*) Thermogram generated from heat–cool–heat cycle in DSC. In the first heating cycle two endotherms are seen corresponding to conversion of form II to form I and melting of form I, respectively. On cooling both events show up as exotherms and on second heating cycle both endotherms reappear, indicating thermal reversibility of enantiotropic polymorphic pairs. (*c*). Solubility of monotropic pairs as a function of temperature. The *T*s is above melting point of both forms. Forms I and II are stable in entire temperature range and their corresponding melting points are shown. (*d*) On heating in DSC, form II melts (*T*m,II) followed by recrystallization (*T*crys) and subsequent melting of form I (*T*m,I). In cooling cycle only melting of form I is seen as an exotherm and on reheating only one endotherm corresponding to form I is seen. This is typical of monotropic polymorphs which converts from form II to stable form I and not vice versa.

versible nature of these polymorphs. The heating rate in DSC is critical for characterizing the polymorphs, as a faster heating rate may not be able to identify the transition temperature, while a lower heating rate may lead to lower resolution of peaks. Therefore, it is a usual practice to generate DSC thermograms under different heating rates during polymorph characterization [11]. Also it is important to note that the sample preparation, particle size, and crucible type can affect the quality of the thermogram [5]. XRD is also another indispensable tool in identifying polymorphs. This is based on the differential scattering of X rays when passed

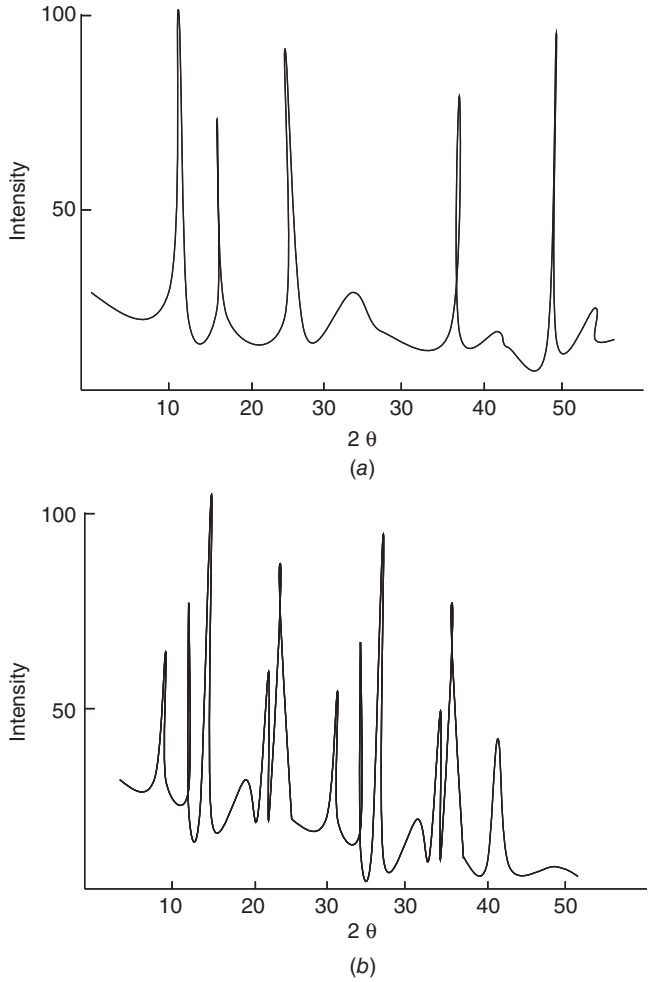


**FIGURE 4** Schematic X-ray diffractograms of two polymorphic forms of a hypothetical drug. The *x* axis represents the detection angle and the *y* axis represents the intensity of the peak. As can be seen, there is a difference in the diffractograms due to the difference in the internal crystal lattice of polymorphs. The different internal arrangement in a crystal deflects the X ray at different angles.

through a powder sample. Typically, on passing through a powder sample, X rays will tend to get diffracted at various angles, and at some angle of detection, the X rays diffracted from the different planes of the crystal converge to form an amplified signal, which is detected by a photomultiplier tube. The angles at which the XRD peaks are obtained are characteristic for a polymorph (Figures 4*a* & *b*). The sample should be uniformly spread to get a good X-ray diffractogram, as an improper sample preparation may lead to variation in intensities due to the preferred orientation of a crystal in the XRD sample holder [14]. Other techniques, such as infrared (IR) spectroscopy and solid-state nuclear magnetic resonauce (NMR), are also used to characterize the polymorphs and are listed in Table 2.

**TABLE 2 Techniques to Characterize Different Crystalline Forms**

| Technique | Applications |
|---|---|
| Thermal analysis | |
| Differential scanning calorimetry | Melting point, enthalpy of fusion, and crystallization; solid-state transformations |
| Thermogravimetric analysis | Stoichiometry of solvates and hydrates; identifying vaporization and volatilization |
| Hot-stage microscopy | Visualization of solid-state transformations and desolvation events |
| X-ray diffractometry | Identifying polymorphs; quantification of degree of crystallinity; crystal lattice geometry and solid-state transformations |
| Spectroscopy | |
| Infrared | Characterization of polymorphs based on functional groups; characterization of hydrates and solvates |
| Near infrared | In situ analysis of solid-state conversions; identification and quantification of polymorphs in dosage forms |
| Nuclear magnetic resonance | Useful to understand difference in molecular arrangement of polymorphs, hydrates, and solvates |

**TABLE 3 Difference in Solubilities of Polymorphs**

| Drug | Melting Point (°C) | Solubility Ratio[a] |
|---|---|---|
| Indomethacin | 157, 163 | 1.1 |
| Sulfathiazole | 177, 202 | 1.7 |
| Piroxicam | 136, 154 | 1.3 |

[a]Indicates ratio of solubility of low-melting polymorphic form to solubility of high-melting polymorphic form of drug.

Polymorphism has significant implications in the solubility, bioavailability stability, processing, packaging, and storage of solid drug substances [15–17]. The metstable polymorphic form may be used to improve the solubility of drug substances. Many drugs are known to exhibit polymorphism, particularly, steroids, barbiturates, anti-inflammatory drugs, and sulfonamides, which have a high probability of exhibiting polymorphism [15]. The existing knowledge on drug polymorphism is a good starting point for a preformulation scientist to anticipate polymorphs based on the drug chemistry. In some cases, polymorphism may provide an opportunity to improve the solubility of a drug. For example, form II of chloramphenicol palmitate has a higher dissolution rate resulting in significantly higher plasma concentration than form I when administered orally [15]. However, in many cases [16] the difference in solubility may not be significant enough to cause differences in oral bioavailability (Table 3). Although the polymorphs differ in their dissolution rates, it should be realized that once the drug goes into solution, they do not differ in their properties. If a drug's absorption is limited by its poor membrane permeability, then the difference in solubility of polymorphs may not impact its bioavailability. Similarly, if the drug dissolution is rapid in comparison to the gastrointestinal (GI) transit time, then the difference in polymorph solubility will not influence its bioavailability [16].

More than the presence of metastable polymorph, it is the conversion of the metastable to the stable form during processing, storage, or use that is of great concern to the pharmaceutical scientist [17]. The unpredictability in "conditions" that result in the generation and conversion of one polymorphic form to another mainly aggravates such a situation [18]. This is exemplified by ritonavir, which is a classical case of appearance of a "new polymorphic form" after the drug was marketed [19]. Ritonavir, an anti-retroviral, drug was introduced in the market as a polymorphic form I in soft gelatin capsule in 1996. Two years later, a new polymorphic form II appeared in the formulation due to some unknown reasons causing the drug to be less soluble. The drug manufacturer withdrew the product due to failure of the batches in dissolution tests. After extensive investigation and reformulation, the drug was reintroduced in the market in 1999.

Nonetheless, in spite of their unpredictability, a good preformulation team will be able to anticipate the different polymorphs during drug development. It should not be an impossible task given the recent advancements in high-throughput generation and characterization of polymorphs [20]. From an innovator's perspective, the identification and thorough characterization of multiple polymorphs during drug discovery can extend the patent life and delay the market entry of generic manufacturers. On the other hand, it is also an opportunity for the generic manufacturers to generate new polymorphs with better solubility and stability for gaining market entry. The patent dispute on ranitidine polymorphs is a good example in this regard [21]. Two polymorphic forms of ranitidine were patented by the innovator company and the generic manufacturers had to find an appropriate method to manufacture the desired polymorph without the accompanying impurity of the other polymorph. This provided an edge for the innovator to extend the drug's market exclusivity for a little longer than they would otherwise have had. If a pure polymorph cannot be generated, the extent of polymorphic impurity should be quantified and ensured from batch to batch. The preformulation scientist closely works with the synthetic chemist in setting specifications for polymorphs.

### 2.2 Hydrates/Solvates

In addition to drug molecules, solvent molecules also get incorporated in the crystal lattice, resulting in altered physicochemical properties. When the solvent is water, they are known as hydrates, while if it is any other solvent, they are known as solvates. They are also known as pseudopolymorphs or solvatomorphs. Hydrates are important in this regard as one-third of all marketed drugs are hydrates [13]. Depending on how the water is arranged inside the crystals, they are classified as isolated hydrates, channel hydrates, and ion-associated hydrates [13]. In isolated hydrates, the water molecules are separated from each other by the intervening drug molecules in the crystal lattice (e.g., cephadrine dihydrate). Channel hydrates result when water molecules are linked to one another forming a channel (e.g., theophylline monohydrate). The water molecules may be present either stoichometrically or nonstoichometrically within the crystal lattice. Ion-associated hydrates are typically seen when the water is metal ion coordinated (e.g., nedocromil zinc). Nonstoichometric channel hydrates are problematic due to the presence of diffusible water, which can easily move in and out of the crystal lattices [13, 22].

Hydrates or solvates are formed by crystallizing the drug in the presence of water or solvents. The hydrate formation is dictated by water activity in a given solvent

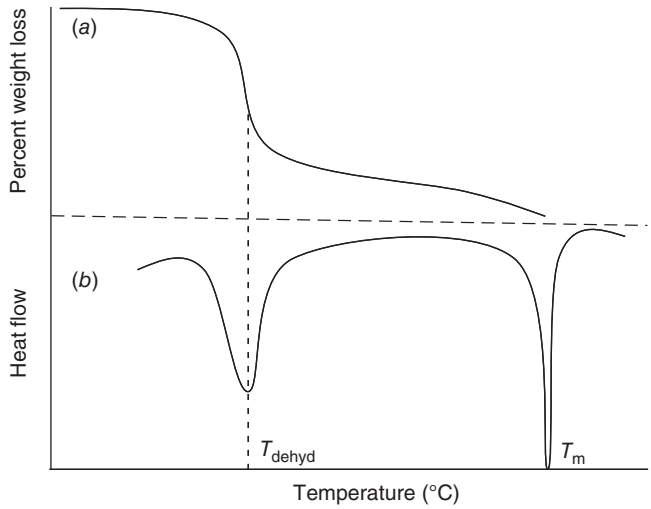


**FIGURE 5** Characterization of hydrate. (*a*) TGA thermogram of monohydrate. The thermogram shows weight loss as a function of temperature. The step in the thermogram shows weight loss due to dehydration of a hydrate. (*b*) DSC thermogram showing endotherm at corresponding temperature (*T*dehyd). The second endotherm indicates the melting point of the hydrate (*T*m).

[22]. Hydrates are characterized using gravimetric methods such as thermogravimetric analysis (TGA) or by Karl Fischer titrometry [23]. In TGA, the loss of water/solvate on heating a sample is recorded as a thermogram (Figure 5). The mass change due to dehydration is seen as a step loss in the TGA thermogram. Based on the weight of the initial sample and its elemental composition, the number of water molecules can be calculated. The TGA curve in combination with a DSC thermogram helps to differentiate hydrates from other thermal transitions. In Figure 5, the endotherm in the DSC thermogram corresponds to water loss as indicated by the TGA curve. The TGA can be coupled to an IR or mass spectrometer to characterize solvates. Thermal microscopy is a useful qualitative tool to visualize the release of water from the drug crystals as a function of temperature [23].

Hydrate formation and dehydration significantly influences the processing and storage of drug products [17]. Hydrates may take up further water or dehydrate to lose water. Dehydration of hydrates leads to several possibilities [24], as shown in Figure 6. Hydrates on dehydration can form isomorphic desolvates retaining the same crystal lattice as the hydrate but without the water. Alternatively, hydrates can lose water and become anhydrous crystals. They can also lose water, forming amorphates with the loss of crystal lattice. Higher hydrates can lose water to form lower hydrates, for example, pentahydrate converting to di- or monohydrate. The hydrates also can exhibit polymorphism or on dehydration can convert to a different polymorphic form [13, 17]. Such solid-state transformations are possible during processing, such as granulation, tableting, and storage [17]. In general, hydrates are less soluble than anhydrous forms while solvates are more soluble than ansolvates in water. Ampicillin trihydrate is a classical example which shows lower solubility and lower plasma concentration than anhydrous ampicillin [15]. Preformulation studies

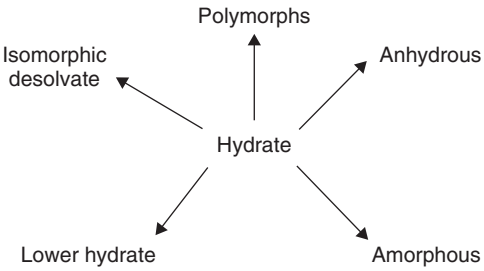


**FIGURE 6** Various possibilities that arise from dehydration of hydrate. A hydrate can dehydrate reversibly into various solid-state forms. It can dehydrate to form an anhydrous form of the drug or to a lower hydrate. Hydrate can also dehydrate to form an isomorphic desolvate where the crystal lattice is retained except for the absence of water. The crystal structure may also collapse on dehydration to form an amorphous form. Hydrates on dehydration can also result in different polymorphs.

provide valuable inputs to the formulator in selecting a suitable form of the drug. For example, ampicillin is hygroscopic and hence can be used in suspension dosage forms, while ampicillin trihydrate, which is non-hygroscopic, is used in solid dosage forms.

### 2.3 Amorphate

Unlike a crystalline drug, an amorphous form of the drug does not have a regular crystal lattice arrangement and the molecules are arranged in random order (Figure 1*c*). Glass is a typical amorphous substance and so amorphous drugs are also known as glasses [25]. Amorphous form is prepared through milling, rapid cooling of a melt, rapid precipitation using an antisolvent, rapid dehydration of a hydrate, spray drying, or freeze drying [26]. Some of the above methods may also unintentionally produce an amorphous form during processing of the crystalline form of the drug [17]. For instance, milling during dosage form manufacture may produce an amorphous form unintentionally, as in the case of indomethacin. The amorphous form does not show a melting point but is characterized by a glass transition temperature ($T_g$). This temperature indicates the conversion of the amorphous form from a rigid glassy state to a more mobile rubbery state. Above $T_g$ the amorphous form will tend to recrystallize and convert to the crystalline form, which then undergoes melting, as shown in Figure 7*a*. The $T_g$ for an amorphous drug can vary depending on the storage conditions and thermal history of the sample and is sensitive to moisture, pressure, and temperature [26]. The $T_g$ is seen only as a slight shift in the baseline due to a change in the specific heat capacity of the sample and is influenced by the heating rate in DSC [25]. In XRD, the amorphous form shows a shallow peak or halo, as opposed to sharp and intense peaks for a crystalline drug compound (Figure 7*b*).

The main advantage of amorphous form of the drug substance state is its significantly higher solubility than the crystalline form of the drug, primarily due to the excess surface energy [16, 27]. Therefore, conversion of a crystalline drug into an amorphous form is one of the strategies to increase drug solubility. Table 4 compares

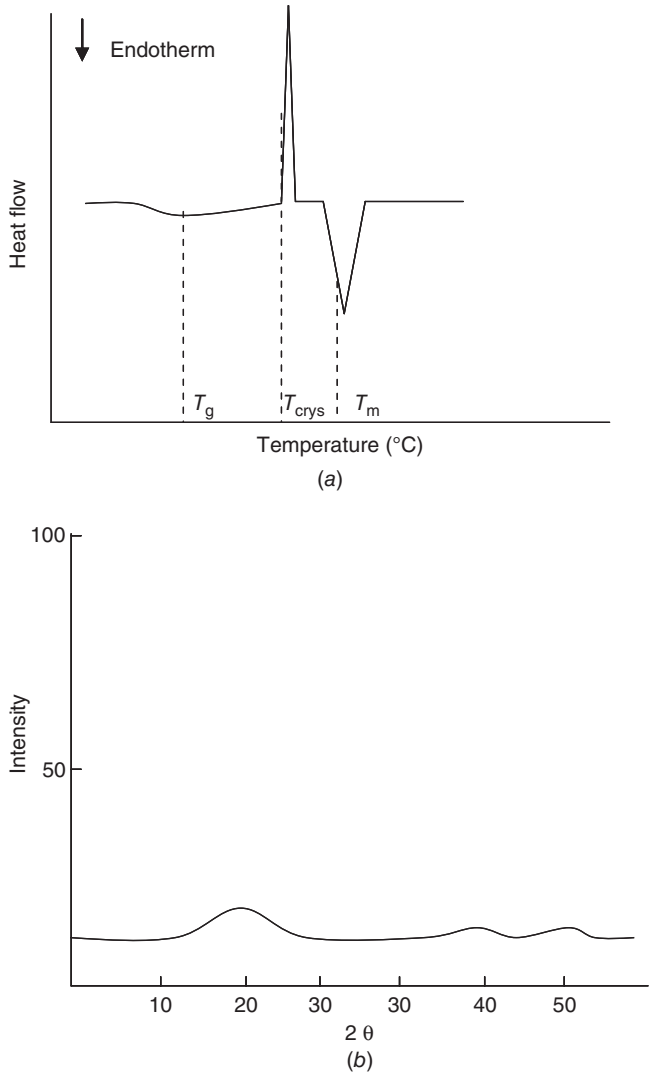


**FIGURE 7** Characterization of amorphous form. (*a*) DSC thermogram of amorphous substance. Thermogram is characterized by a glass transition temperature (*T*g) above which the amorphous form is mobile and recrystallizes (*T*crys) into a crystalline form which finally melts (*T*m). (*b*) Amorphous form that does not show any peaks in XRD as it does not have regular arrangement of molecules. Shallow peaks are indicative of an amorphous drug substance.

the solubility of amorphous and crystalline forms of a few drugs. However, the biggest challenge lies in the stabilization of the amorphous form to prevent it from converting to the less soluble crystalline form during storage and use. It should be noted that they can take up moisture to convert to a crystalline form or to a hydrate, resulting in decreased drug solubility. The moisture uptake can also lead to chemical degradation [26]. The stabilization of drug amorphates is usually accomplished by increasing the $T_g$ using polymers and thus restricting their molecular mobility and

**TABLE 4 Comparative Solubilities of Amorphous and Crystalline Forms of Drugs**

| Drug | Solubility Ratio[a] |
|---|---|
| Carbamazepine | 1.5–1.7 |
| Griseofulvin | 38–441 |
| Glibenclamide | 14 |

[a]Indicates ratio of solubility of amorphous form to solubility of crystalline form of drug.

chemical reactivity [26]. Particle size reduction, which is a very common processing step in dosage form manufacturing, can result in varying degrees of amorphous and crystalline forms in the drug. In such cases it is essential to quantify the degree of crystallinity using various analytical techniques such as DSC and XRD [28].

## 2.4 Hygroscopicity

Some solid drug substances have a tendency to absorb moisture from the atmosphere leading to physical and/or chemical instabilities. Hygroscopicity is the rate and extent of moisture adsorbed/absorbed by a solid substance. Solid drug substances may vary in their behavior to moisture and are classified as deliquescent, efflorescent, and effervescent. Deliquescent materials such as hydrochloride salts absorb moisture and become a liquid. Effervescent substances (e.g., a mixture of citric/tartaric acid and sodium bicarbonate) absorb moisture and release carbon dioxide. On the other hand, efflorescent substances such as hydrates may lose moisture depending on the relative humidity (RH). Therefore, it is important to study the moisture absorption behavior of drugs to choose the processing and storage conditions for the drug.

Hygroscopicity is measured by exposing the drug sample to various RH in a dessicator. The RH is maintained at a constant level by using salt solutions of varying concentrations (e.g., $KNO_3$, KCl) and the humidity is expressed with respect to the humidity of a saturated salt solution. Moisture sorption and desorption curves are generated to study the moisture uptake. The moisture absorption profile is generated by noting the increase in mass on exposure to varying RH and the desorption profile is generated by recording the change in weight with decreasing RH [23, 29]. This can be measured using a dynamic vapor sorption instrument. A typical sorption/desorption profile is shown in Figure 8 for an anhydrous and hydrate form of a hypothetical drug. As can been, for a hydrate the sorption and desorption profile is superimposable and is hence non-hygroscopic. On the other hand, the anhydrous form of the drug is hygroscopic and shows hysteresis on the sorption and desorption profile, indicating significant moisture uptake. Such profiles give useful clues to the preformulation scientist. Significant hysteresis is indicative of hydrate formation and can be used as a guide to evaluate the potential of hydrate formation. Further the profile also helps to differentiate hygroscopic and non-hygroscopic salts [29]. The profile also gives information on processing and storage conditions that can overcome solid-state transformations. For example, in Figure 8, it is seen that the anhydrous form does not take up moisture if RH is below 80% and it retains the moisture

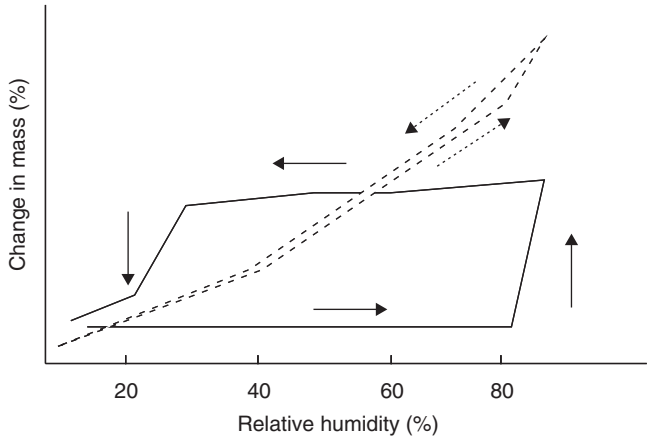


**FIGURE 8** Sorption and desorption profile of hydrate and anhydrous form of hypothetical drug. The solid lines represent the profile for an anhydrous form of the drug, while the broken lines represent the profile for a hydrate form of the drug. Anhydrous form of the drug does not take up moisture until it reaches 80% RH, and on reducing the RH, it does not lose moisture until it reaches RH of 20%. The hysteresis is indicative of hygroscopicity and significant moisture uptake. The hydrate form of the drug does not show hysteresis but both the sorption and desorption curve superimpose on each other indicating that it is non-hygroscopic.

until the RH is reduced below 20%. The study can also be used to extract kinetic and temperature information on moisture uptake by drug substances. The studies conducted during preformulation testing should be representative of the anticipated processing and storage conditions of the drug.

### 2.5 Particle Characteristics

Drug particle characteristics such as size, shape, and surface area impact the drug's processability and product performance. Particle size is the most influential among these as the other two properties can be related to it. When sufficient drug is available, the preformulation scientist characterizes the particle size and size distribution to set specifications for formulation and future drug lots. Table 5 lists the various methods used to measure particle size. The methods differ in their principle of operation and also in the range of particle sizes they can measure [30]. Usually, the gross particle morphology is characterized using a simple optical microscope and if required is further characterized using a scanning electron microscope. Light-scattering methods are commonly used to measure particle size due to their low sample requirements and ease of measurement. The instrument readout is in the form of a graph where the particle size is plotted against the percent frequency of particles in different particular size ranges. The results are used to set particle size specifications and understand polydispersity or multimodal particle size distribution in powders. Especially if the drug is potent, a narrow size distribution is desired to ensure drug homogeneity during formulation. The surface area of a powder bed is determined using the Brunauer, Emmett, and Teller (BET) method. In this method,

**TABLE 5 Methods for Particle Size Analysis**

| Method | Principle |
|---|---|
| Sieve | Sieve analysis utilizes a series or stack, or nest of electro brass or stainless steel sieves that have smaller mesh at the bottom followed by meshes that become progressively coarser toward the top of the series. Useful for measuring particles in size range 10–50,000 μm |
| Microscopy | Analysis is carried out on two-dimensional images (projected diameter) of particles which are assumed to be randomly oriented in three dimensions. Can measure particles in the size range 1–1000 μm. Electron microscopy is useful for analysis of particles in submicrometer range (0.01–100 μm). It also gives information on surface morphology and shape of the powder. |
| Sedimentation | Size analysis is based on sedimentation of particles as a function of their size due to gravitational pull or by using a centrifugal force. Can measure particles in the size range 0.01–100 μm. |
| Light scattering | This is based on the principle of light scattering of particles as a function of their hydrodynamic radius. Commonly used, as it requires a small sample size and is a rapid method for particle size measurement. It can be used to measure particles varying in size from 0.001 to 100 μm. |

$N_2$ gas is passed through a powder bed and the surface area is calculated based on the volume of gas coming out in the absence and presence of the powder [31].

Particle size reduction is an initial step in the development of any dosage form. The data generated during preformulation testing guides the formulator in deciding about size reduction. Particles with size >100 μm generally require size reduction and particles in the size range of 10–40 μm are generally acceptable for solid dosage forms [32]. As mentioned in the earlier section, particle size reduction may lead to partial or complete amorphization of a powder, and this factor should be taken into consideration. Also worth mentioning are the other solid-state transformations that may take place during milling, such as the conversion of one polymorphic form to another or the desolvation of a hydrate [17]. It is important to maintain the particle size distribution within a narrow range to avoid powder stratification and avoid flowability issues during capsule filling or tablet compression. Particle size reduction increases drug solubility due to the enormous increase in surface area with decreasing particle size. Griseofulvin is often widely quoted in the literature in this regard, where the bioavailability of this water-insoluble drug is increased 10 times on reducing the particle size [32]. On the other hand, particle size reduction may be counterproductive for some drugs such as nitrofurantoin. Particle size reduction of nitrofurantoin causes rapid drug absorption with an associated increase in its adverse effects. In contrast, the slowly dissolving macrocrystals of nitrofurantoin do not cause adverse effects [32]. Excessive particle size reduction (<30 μm) may also lead to static charge buildup on the surface, resulting in agglomeration and reduced powder flow or reduced drug solubility. If the drug is sensitive to moisture or oxygen, then the increased surface area associated with particle size reduction may accelerate degradative reactions [32]. Therefore, it is important to study the influence of particle size on drug solubility and stability during preformualtion testing.

### 2.6 Powder Flow and Compressibility

The derived properties of a drug substance play a critical role in deciding about the feasibility of a solid dosage form. These include the bulk density, flow properties, and compressibility of the drug powder. Powder flow is influenced by many solid-state properties, including crystal habit, bulk density, particle size, and shape. Bulk density is an important determinant of powder flow. It is the ratio of a known weight of the powder and its bulk volume. This is determined by pouring a weighed amount of the powder into a graduated cylinder and measuring its volume. Bulk density is particularly important for high-dose drugs, where the drug would occupy a major portion of the tablet or capsule dosage form. The true density of a powder is determined using a helium densitometer [31]. The volume of helium gas that passes through an empty tube is compared with the volume of helium passing through the tube when filled with a defined weight of the powder. The difference in the volume gives the true volume of the powder, which is then used to calculate the true density of the powder. This information can be used to calculate the porosity:

$$\text{Porosity} = \frac{\text{bulk volume} - \text{true volume}}{\text{bulk volume}} \times 100 \tag{1}$$

The porosity of a powder depends on particle size and shape; pharmaceutical powders vary in porosity from 30 to 50% [31]. Powder with varying particle size will give a porosity of less than 30%, where the small particles may occupy the pores in between the larger particles. On the other hand, powder aggregates lead to increased porosity and poor flow properties.

The powder flow is determined using the angle of repose or Carr's index (Figures 9*a* and *b*). The angle of repose is the angle that the powder makes with the horizontal surface when allowed to flow through a funnel (Figure 9*a*). This is based on the principle that powder flow is influenced by the relative influence of interparticle friction and the gravitational pull on the powder. Pharmaceutical powders have an angle of repose of 25°–40° [33], and in general, a good flowing powder will have a lower angle of repose (Figure 9*c*). The angle of repose using the funnel provides a good estimate of the influence of particle size, shape, and electrostatic interaction between the particles when the powder flows through the hopper in a tableting machine [34]. When there is a limited drug sample, Carr's index is used to estimate powder flow and compressibility. In this method, a small quantity of the powder is used to determine its bulk density and this is followed by determining the tapped density of the powder. After filling, the graduated cylinder is tapped on a hard surface (3–30 taps) until the powder consolidates and gives a constant volume (Figure 9*b*). Carr's index is calculated using the equation

$$\text{Carr's index} = \frac{\text{tapped} - \text{poured density}}{\text{tapped density}} \times 100 \tag{2}$$

This index is a good measure of powder consolidation and compressibility for predicting the feasibility of developing a tablet dosage form. A lower Carr's index is indicative of a good flowing powder. There is a good correlation between angle of repose and Carr's index [Figure 9*c*], and depending on the quantity of the drug

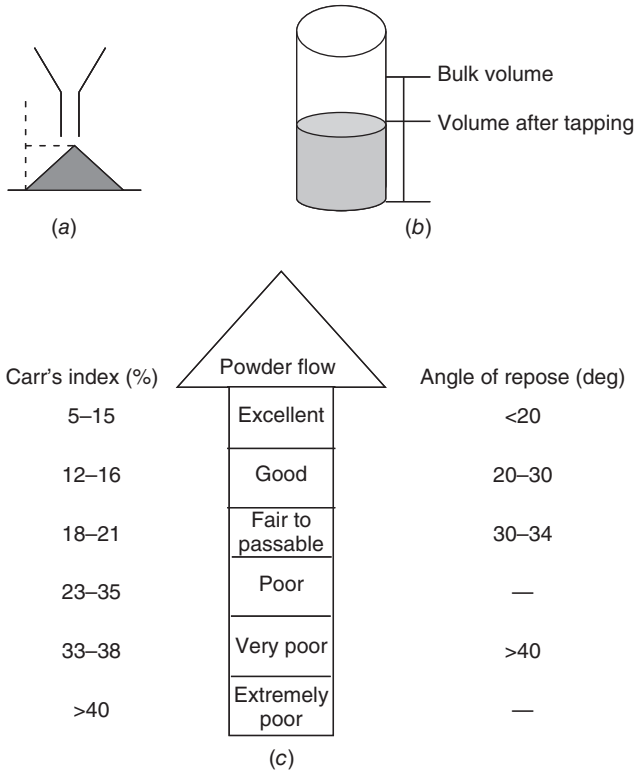


**FIGURE 9** Measurement techniques for powder flow. (*a*) Angle of repose is determined by pouring a powder through a funnel and noting the angle that the powder heap makes with the horizontal surface. (*b*) Carr's index is measured by pouring a known weight of the powder into a graduated cylinder and tapping it on a hard surface until the powder is consolidated. (*c*) Carr's index is measured as percent, while angle of repose is measured in degrees, and both methods show good correlation. Powder flow is classified based on either of the measurement methods.

available for preformulation testing, either method may be used to estimate the powder flow property. Poor flowing powders may require glidants to improve their flow property. Compressibility is studied by compressing the drug in a hydraulic or IR press, and such experiments give early warnings to the formulator about capping and lamination issues in tablets. Altogether, these preformulation tests give valuable information to the formulator in deciding excipients and processes.

## 3 SOLUBILITY CHARACTERISTICS

Drug solubility is one of the physicochemical parameters that receives lots of attention during preformulation testing. In the initial stages, solubility studies are usually kinetically determined, where the drug is placed in contact with the solvents and then the solubility is assessed using turbidometric methods almost instantaneously.

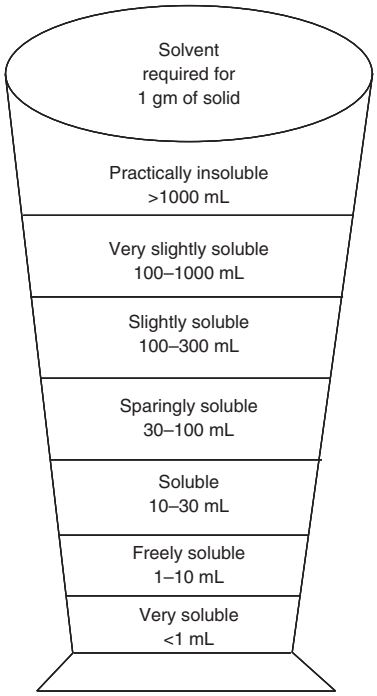


**FIGURE 10** Terminologies for drug solubility. The drug solubility is qualitatively described depending on how much solvent is required to solubilize 1 g of the drug.

A high-throughput solubility screen consists of dissolving the drug in a minimal volume (few microliters) of dimethyl sulfoxide (DMSO) and then adding different volumes of water until turbidity is observed. The appearance of turbidity is considered as a rough estimate of the drug's water solubility and turbidity is measured in a 96-well plate format in a turbidometer [35]. Solubility is qualitatively described in terms of how much solvent is required to solubilize 1 g of the drug and is shown in Figure 10. Kinetic solubility measurements are only rough estimates, as they do not take into account the solid-state transitions and should be followed up with equilibrium solubility studies later when more drug is available.

Equilibrium solubility is determined by placing an excess solid drug in a few milliliters of the solvent and shaking at 37 °C for 60–72 h until equilibrium is reached. One to three samples are withdrawn, filtered, and assayed for drug content. Equilibrium solubility helps to identify polymorphic or amorphous forms of the drug, which shows an apparently higher solubility in kinetic studies. Intrinsic solubility measurements are measured for ionic compounds to determine the inherent solubility of the un-ionized form of the drug. This would mean that the solubility of a weakly acidic drug is tested in an acid medium and a weakly basic drug is tested in an alkaline medium, where the drug would remain completely un-ionized. The studies are followed by determining drug solubility at different pH and are discussed later in this section.

A drug solubility of 1 mg/mL is usually considered acceptable to avoid dissolution rate–limited absorption in vivo. If the drug solubility is <1 mg/mL, salts should be considered, if the drug is ionic and cosolvents should be considered if the drug is nonionic [33]. This brings us to a discussion of the factors that influence drug solubility: temperature, pH, crystal form, and solvents. The effect of temperature on drug solubility depends on the heat of solution ($\Delta H_s$), which is the amount of heat given up or taken up when a drug goes into solution. The relationship between drug solubility $S$ and temperature $T$ is given by the van't Hoff equation, where $R$ and $C$ are constants:

$$\ln S = \frac{\Delta H_s}{RT} + C \tag{3}$$

Drugs which have $-ve$ $\Delta H_s$ (exothermic) generally decrease in solubility with increase in temperature (e.g., lithium salts and hydrochloride salts), while drugs which have $+ve$ $\Delta H_s$ (endothermic) usually increase in solubility with increase in temperature (e.g., most organic drugs). Nonelectrolytes have a $\Delta H_s$ of 4–8 kcal/mol and show significant increase in solubility when the temperature is increased. On the other hand, salts have a $\Delta H_s$ of –2–2 kcal/mol and are therefore less sensitive to temperature [34]. Solubility studies are conducted at 25, 37, and 50 °C. In addition to providing information on the drug solubility at the body or processing temperatures, it is also useful to understand polymorphic interconversions.

With respect to selecting cosolvents, one should consider drug polarity and solvent polarity. Usually the solvents include glycerol, propylene glycol, and ethanol. Other solubilization techniques such as complexation and surfactants can also be used to enhance the solubility of the drug [36]. However, the solubilization techniques used in preclinical testing may not be same as the final formulation used in clinical studies and marketing.

For ionic substances, salt formation is the preferred strategy for drug solubilization. The salt selection is crucial during early discovery, as any change will require repeating some of the earlier preclinical studies. Salts provide wider flexibility in modulating the drug properties without changing its activity. The salt formation is used to address drug solubility, stability, and processing issues [37]. Sometimes, salt formation may be used to deliberately reduce the solubility of drug to overcome unpleasant taste or stability of the drug. For example, clindamycin pamoate is used in place of hydrochloride salt to overcome the unpleasant taste of the latter. The various factors to be considered in salt selection are discussed in the next section. In addition to improving the solubility of the drug, solubility data guides the formulator to choose a suitable granulating solvent for a tablet dosage form.

### 3.1 p$K_a$ and Salt Selection

The majority of the pharmaceutical drugs are weak bases or weak acids. Among the marketed drugs, more than 75% are weak bases, 25% are weak acids, and 5% are nonionic [38]. Therefore, knowledge of p$K_a$ is useful for enhancing drug solubility and stability. The Henderson–Hasselbalch equation is used to describe the ionization of a weak acid or base:

$$\text{Percent ionized} = \begin{cases} \dfrac{100}{1+10^{(\mathrm{p}K_a-\mathrm{pH})}} & \text{for weak acid} \quad (4) \\ \dfrac{100}{1+10^{(\mathrm{pH}-\mathrm{p}K_a)}} & \text{for weak base} \quad (5) \end{cases}$$

The equations theoretically predict the ionization and solubility of a drug at a given pH. The p$K_a$, which is a characteristic property of an ionizable drug, defines the pH at which the drug is half ionized and half un-ionized. A weakly acidic drug will be predominantly in the un-ionized form two pH units below its p$K_a$ and predominantly ionized at two pH units above its p$K_a$. In a weakly basic drug, it is exactly the opposite of what is seen with an acidic drug. The p$K_a$ of a drug can be measured using potentiometry, solubility, conductometry, and spectroscopic techniques [35]. Potentiometry measures the change in potential when the drug is titrated with an acid or base and is suitable for drugs which have p$K_a$ of 3–10 [33]. If the drug is not water soluble, it is usually dissolved in a water-miscible solvent such as DMSO or methanol. In order to nullify the effect of cosolvents, measurements are made with various cosolvent concentrations and are plotted against p$K_a$. The intercept on the *y* axis gives the p$K_a$ at zero cosolvent concentration. Some drugs have more than one ionizable group, such as ampholytes, and are characterized by more than one p$K_a$ value. They are classified as ordinary or zwitterionic ampholytes [39]. In ordinary ampholytes, $\mathrm{p}K_{a,acidic} > \mathrm{p}K_{a,basic}$ (e.g., chlorambucil), while in zwitterionic ampholytes, $\mathrm{p}K_{a,acidic} < \mathrm{p}K_{a,basic}$ ampholytes (e.g., amino acids and proteins). The amino acids and proteins are characterized by their isoelectric point (pI), which is the pH at which the net charge is zero and is calculated using the formula

$$\mathrm{pI} = \frac{1}{2}(\mathrm{p}K_{a,acid} + \mathrm{p}K_{a,base}) \tag{6}$$

Drugs may also have more than two p$K_a$ values, such as polyprotic or polybasic compounds (e.g., minocycline), and such drugs exhibit a complex pH solubility profile. It is essential to know per se pH of the drug solution during preformulation studies. The pH is measured or theoretically calculated if the p$K_a$ and drug concentration *C* are known. The pH of a weak acid or the salt of a weak base and a strong acid can be calculated using the equation

$$\mathrm{pH} = \frac{(\mathrm{p}K_a - \log C)}{2} \tag{7}$$

Similarly, the pH of a weak base or a salt of weak acid–strong base can be calculated using the following equation, where p$K_w$ is the ionization constant of water:

$$\mathrm{pH} = \frac{(\mathrm{p}K_w + \mathrm{p}K_a + \log C)}{2} \tag{8}$$

Solubility of a drug is directly proportional to the extent of ionization of a drug in water. Therefore, p$K_a$ of the drug may also be determined by noting the change in solubility of a drug as a function of pH. It is customary to check the drug solubility

in the pH range usually encountered in the gastrointestinal tract (GIT; pH 1–8). The influence of pH on drug solubility can be estimated from the equations

$$S = \begin{cases} S_u[1+10(\mathrm{pH}-\mathrm{p}K_a)] & \text{for weak acid} \quad (9) \\ S_u[1+10(\mathrm{p}K_a-\mathrm{pH})] & \text{for weak base} \quad (10) \end{cases}$$

In this equation $S$ represents the total solubility of ionized and un-ionized forms of the drug, while $S_u$ represents the intrinsic solubility of un-ionized form of the drug. The pH solubility profile and the drug p$K_a$ are important for salt selection. In addition, the salt selection is governed by various factors as outlined in Table 6. As a rule of thumb, a strong acid is used to form a salt with a weakly basic drug and a strong base is used to form a salt with a weakly acidic drug. The probability of salt formation can be predicted from the relative p$K_a$ values of the drug and the counterions by using the equation [40]

$$\Delta \mathrm{p}K_a = \mathrm{p}K_{a,\mathrm{drug}} - \mathrm{p}K_{a,\mathrm{counterion}} \quad (11)$$

The probability of salt formation is high for a weak acid if $\Delta$p$K_a$ is negative, that is, p$K_{a,\mathrm{drug}}$ < p$K_{a,\mathrm{counterion}}$. Sodium phenytoin is an example, where phenytoin has a p$K_a$ of 8.3 and sodium has a p$K_a$ of 16. Similarly, a weak base will form a salt with acid counterion if $\Delta$p$K_a$ is positive, that is, p$K_{a,\mathrm{drug}}$ > p$K_{a,\mathrm{counterion}}$. In atropine sulfate, atropine has a p$K_a$ of 9.9 and sulfuric acid has a p$K_a$ of –3. It is important to mention that salts do not alter the intrinsic p$K_a$ of the drug but increase drug solubility by keeping pH on the ionization side of the drug's p$K_a$ (Figures 11*a* and *b*). The salt formation is a futile exercise if the p$K_a$ of a drug is <3 or >10 and other solubilization strategies have to be pursued [33].

Inorganic ions such as hydrochloride and sodium are the most frequently encountered species in pharmaceutical salts, primarily because of the ease of salt formation and their low molecular weights. They provide significant increase in drug solubility and at the same time strong acids/bases may also be hygroscopic, posing problems during processing and storage. In those cases, salts are formed using weaker organic anions or cations such as mesylate, besylate, and choline. Table 7 lists some of the commonly used inorganic and organic counterions used in pharmaceutical salts. The pH provided by the salts significantly influences the drug dissolution and subsequent drug absorption from the GIT. Though salts increase drug ionization and aqueous drug solubility, it is the un-ionized form of the drug that is absorbed through the membrane. According to the well-known pH partition hypothesis [33], a weakly

**TABLE 6 Factors to Consider during Salt Selection**

| |
|---|
| Relative p$K_a$ of the drug and the counterion |
| Common-ion effects |
| Crystal habit and crystallinity |
| Polymorphic conversions |
| Hygroscopicity |
| Chemical stability |
| Manufacturability |
| Toxicity |

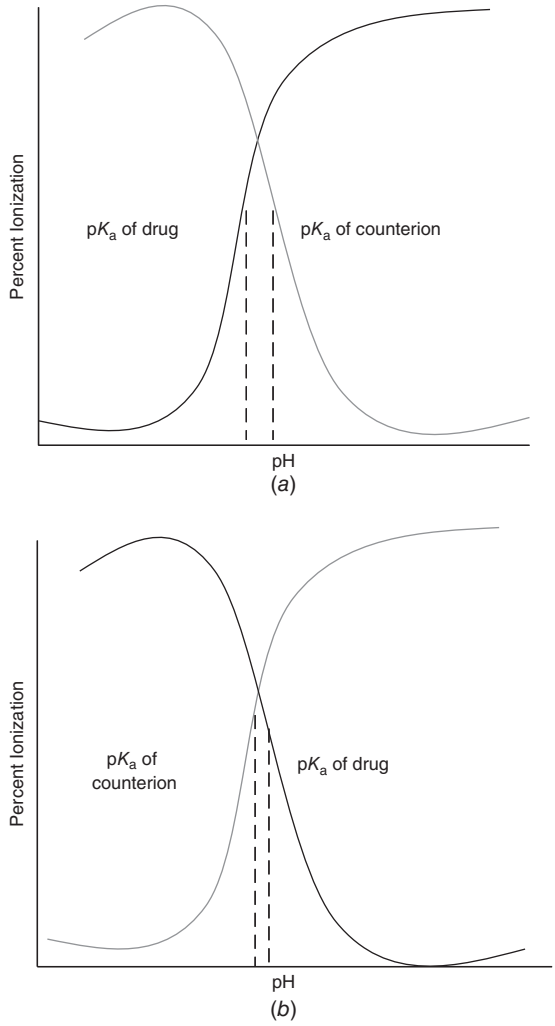


**FIGURE 11** Relative p*K*a of drug and salt-forming counterion. (*a*) For a weak acid, the p*K*a of the salt-forming counterion should be higher than the drug p*K*a to keep the pH on the ionization side. (*b*) For a weak base, the salt-forming counterion should have a p*K*a less than the drug's p*K*a to keep the pH in the ionization side.

**TABLE 7 Various Counterions Used to Form Drug Salts**

| Chemical class | Salt-Forming Counterions |
|---|---|
| Inorganic | Hydrochloride, hydrobromide, sulfate, nitrate, phosphate, sodium, potassium, calcium, and zinc |
| Organic | Triethanolamine, ethanolamine, lactic acid, maleic acid, citric acid, acetic acid, choline, ethanesulfonic acid, oleate, and stearate |

acidic drug is primarily absorbed from the stomach (pH 1–3), where it remains in un-ionized form. A basic drug is expected to be absorbed from the intestines (pH 6.5–8), where it would be in the un-ionized form. But in many cases, the absorption of drugs cannot be satisfactorily explained by this theory. This is understandable if one considers the fact that drug dissolution from a salt is mainly influenced by the surface pH of the dissolving drug (microenvironment) rather than the bulk pH of the GI fluids. For instance, the surface pH of weakly acidic drug is 1.5 times higher than the bulk pH in the stomach, and therefore the dissolution of the salt of a weak acid will be 100 times faster than the free-acid form of the drug in stomach [40]. Similarly, it is the free base rather than its salt form which will dissolve faster in the gastric pH. So it is possible to modulate the drug solubility of a pharmaceutical salt irrespective of its location in the GIT. Considering the fact that most of the drugs are absorbed from the intestine (due to large surface area), it is desirable to select a salt that is not completely ionized or unionized at the intestinal pH to have optimal dissolution and absorption.

An important phenomenon that is often overlooked during the salt selection process is the suppression of salt ionization in GI fluids due to the common-ion effect. This is particularly important with inorganic salts, where salt ionization can be suppressed by ions such as chloride and sodium which are abundant in GI fluids. A hydrochloride salt will ionize in solution, as shown in the Equation (12), but in gastric fluid, the presence of chloride ions suppresses the drug ionization [as shown by the thicker arrow in the reverse direction in Equation (13)] to maintain an equilibrium between the ionized and un-ionized form of the drug:

$$DH^{+}Cl^{-} \rightleftarrows DH^{+} + Cl^{-} \tag{12}$$

$$DH^{+}Cl^{-} \rightleftarrows DH^{+} + Cl^{-} \tag{13}$$

Hence it is important to study this effect during preformulation by testing the solubility of the salt in the presence and absence of sodium chloride. Although salts do not alter the pharmacological activity of the drug, safety is an important consideration in the selection of salts. From this perspective, salts are treated as a new molecule by the FDA. The safety of the salt is evaluated with respect to its route of administration and dose of the drug [37].

### 3.2 Partition Coefficient

In simple terms, the partition coefficient represents the relative solubility of a drug in a hydrophobic and a hydrophilic solvent. The hydrophilic solvent is usually water or buffer (pH 7.4), while the hydrophobic phase is usually *n*-octanol. The partition coefficient ($K_{o/w}$) is defined as the ratio of concentration of the drug in the organic phase ($C_o$) to drug concentration in the aqueous phase ($C_w$):

$$K_{o/w} = \frac{C_o}{C_w} \tag{14}$$

The choice of *n*-octanol is based on its ability to mimic the lipophilicity of the biological membranes [33] and further its solubility parameter ($\delta$ = 10.24; solubility

parameter is a measure of internal cohesive energy) falls within the solubility parameter range of most drugs (8–12.4). The partition coefficient is determined by dissolving the drug in one of the phases and shaking both the phases together in a flask for 30 min to achieve equilibration. Then the drug concentration is determined from one of the phases, usually the aqueous phase, and the drug concentration in the oil phase is determined by subtracting the drug concentration in the aqueous phase from the total drug that was added. This value when expressed in logarithmic form is known as log *P*. The phase volume of the two phases is 1 : 1 but, if the drug is less soluble in the aqueous phase, the ratio (water–octanol) is changed to 1 : 10 or 1 : 20 to have a measurable drug concentration in the aqueous phase [33]. It is important to saturate the phases with respect to the other solvent before starting the experiment to rule out the influence of solvent partitioning on drug distribution between the two phases.

Another important factor is the influence of drug ionization on the partition coefficient and this is particularly important when a buffer is used instead of water. The partition coefficient determined from Equation (14) is an apparent value rather than a true partition coefficient for ionic drugs. However, the true partition coefficient can be calculated from the apparent partition coefficient if the drug's p*K*a and the pH of the drug solution are known [39], as shown in the equations

$$\log P = \begin{cases} \log P_{\text{app}} - \log\left(\dfrac{1}{1+10^{(\text{pH}-\text{p}K_{\text{a}})}}\right) & \text{for weak acid} \quad (15) \\ \log P_{\text{app}} - \log\left(\dfrac{1}{1+10^{(\text{p}K_{\text{a}}-\text{pH})}}\right) & \text{for weak base} \quad (16) \end{cases}$$

During the initial stages of drug screening, the partition coefficient is calculated based on the chemical structure (*C* log *P*). This is done by assigning values to different fragments in the chemical structure [41]. The calculated values are only estimates, but they are useful to rank order a homologous series of compounds based on their lipophilicity for further lead optimization. Given the fact that drugs have to cross many biological membranes before reaching the site of action, the log *P* value has a significant influence on drug absorption, drug pharmacokinetics, and pharmacology. This is exemplified from the numerous structure–activity and structure–property relationships using log *P* [41]. The log *P* is important in assessing the oral absorption potential of a drug. If a drug has a low log *P* (<1), it is expected to have poor membrane permeability, while if a drug has a large log *P* (>5), it will be trapped in the lipophilic membrane. A log *P* of 1–5 is usually considered optimal for oral drug absorption [3]. For an ionic drug, the un-ionized form of the drug will be more lipophilic than the ionized form of the drug. Therefore, at any given pH in vivo, the relative proportion of ionized versus un-ionized form of the drug dictates drug dissolution and absorption through the membrane.

### 3.3 Drug Dissolution

Dissolution is the rate at which a solid drug goes into solution and is a critical determinant in the absorption of drugs from solid dosage forms. A drug has to go into solution before it can be absorbed. In vitro dissolution studies are a valuable

tool for determining the influence of different solid-state properties on drug dissolution and vis-à-vis predict in vivo drug dissolution and absorption. It is used to screen drugs that show dissolution rate–limited absorption. The factors that influence drug dissolution rate ($dc/dt$) can be understood from the well-established Noyes–Whitney equation [42]

$$\frac{dC}{dt} = \frac{kSC_s}{V} \tag{17}$$

where $k$ is the dissolution rate constant (cm/s), $C_s$ is the saturated solubility of the drug, $S$ is surface area of solid exposed to the solvent, and $V$ is the volume of the dissolution medium. The dissolution rate constant is a function of the diffusion coefficient $D$ of the drug through a stagnant aqueous layer adjacent to the dissolving surface and thickness $h$. Powder or particulate dissolution is carried out in a dissolution vessel in a specific volume of dissolution fluid (900 mL) which is stirred and maintained at 37 °C. Several configurations are available to study the dissolution of various dosage forms (Figures 12*a* and 12*b*). Various simulated physiological media are used to understand the in vivo behavior of drug and dosage forms [43]. The usual dissolution media include water, 0.1 N HCl, and pH 7.4 buffers. It is important to maintain sink conditions in the dissolution medium by keeping the drug concentration at 10% of saturated solubility of drug (to mimic in vivo). For poorly water soluble drugs, surfactants are often added to the dissolution medium for this purpose. During preformulation testing, particulate dissolution studies reveal the influence

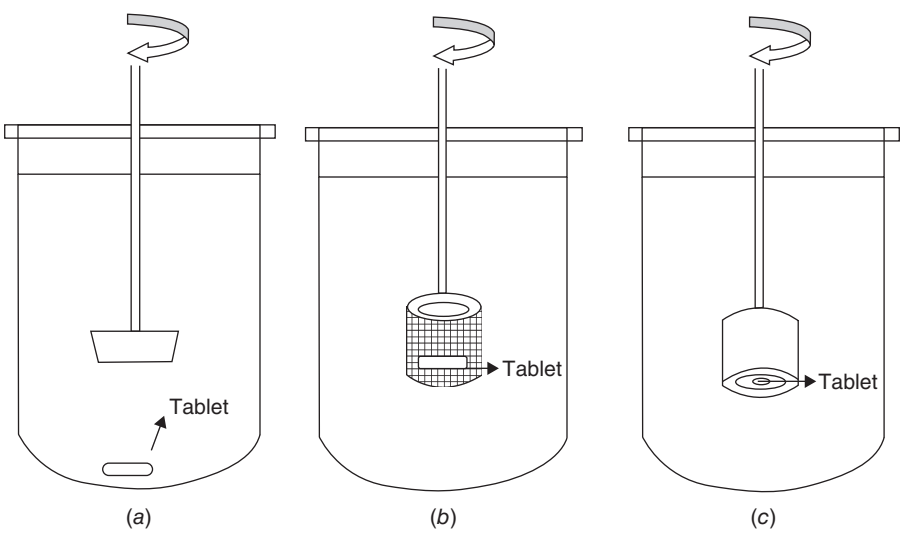


**FIGURE 12** Various drug dissolution methodologies. (*a*) In the paddle method, the tablet is placed in the dissolution vessel containing dissolution medium and the paddle is rotated at defined rpm, while the dissolution vessel is maintained at body temperature. (*b*) In the basket method, the tablet is kept inside a meshed basket and rotated. (*c*) For IDR studies, the tablet is kept inside a die cavity and only one face of the tablet is exposed to the dissolution medium.

of particle size, crystal habit, and wettability of a drug substance. A formulator, on the other hand, routinely uses dissolution testing as a quality control tool in the design of dosage forms. Various models have been developed to describe the release kinetics of conventional and modified release dosage forms [44].

Since powder dissolution is influenced by changing surface area, it is not useful for delineating the effects of polymorphs, hydrates, and pharmaceutical salts. Instead, the intrinsic dissolution rate (IDR) is used. The IDR studies are conducted at a constant surface area and hence the dissolution rate observed is only a function of the intrinsic solubility of the drug. The Noyes–Whitney equation is modified for IDR, where the surface area is kept constant, and Equation (17) reduces to

$$\frac{dC}{dt} = k_1 C_s \tag{18}$$

where $k_1$ is the intrinsic dissolution rate constant and $dC/dt$ is the intrinsic dissolution rate (mg·cm$^2$/s). For IDR studies, the drug (500 mg) is compressed in a hydraulic press (at 500 mPa) to a 13-mm disc. This disc is then loaded onto a holder which exposes only one surface of the disc to the dissolution medium (Figure 12*c*). The IDR is obtained by dividing the slope of the plot between the amount of drug dissolved and time by the area of solid exposed to the dissolution medium. The IDR predicts the influence of drug solubility on in vivo drug dissolution and absorption. A drug which has an IDR of 1 mg·cm$^2$/s will not generally show dissolution rate–limited absorption in vivo. However, if the IDR falls between 0.1 and 1 mg·cm$^2$/s, then further studies may be required to make a decision. Drugs with IDR < 0.1 mg·cm$^2$/s show dissolution rate–limited absorption in vivo, necessitating drug solubilization strategies [45]. With respect to pharmaceutical salts, IDR is used to understand the influence of surface pH on drug dissolution and absorption. The common-ion effect can be studied by including 0.1–0.15 *M* NaCl in the dissolution medium. Also IDR is useful to understand the difference in solubility of polymorph and amorphous forms. However, in some cases, the compression force used in making the IDR disk may by itself induce solid-state transformations [45]. DSC, IR, and XRD must be used to identify the drug's solid state before and after compression as well as at the end of the dissolution studies. A well-designed dissolution study is used as an early warning for drug molecules that would pose absorption problems in vivo.

### 3.4 Absorption Potential

The ultimate goal of any drug development program is to develop an orally absorbable compound. Solubility and permeability are the two most critical parameters that dictate oral absorbability of a molecule. All other parameters are directly or indirectly related to these two physicochemical properties. As can be seen from Figure 13, there are several barriers that a drug needs to overcome before reaching the systemic circulation. The oral absorption of a drug is mainly limited by drug dissolution and/or by the drug permeation across the GI membrane. Considering their importance, drug solubility and permeability are screened very early in the drug discovery process. Solubility studies are typically run in a high-throughput format using a turbidometric method as discussed earlier. Based on this primary

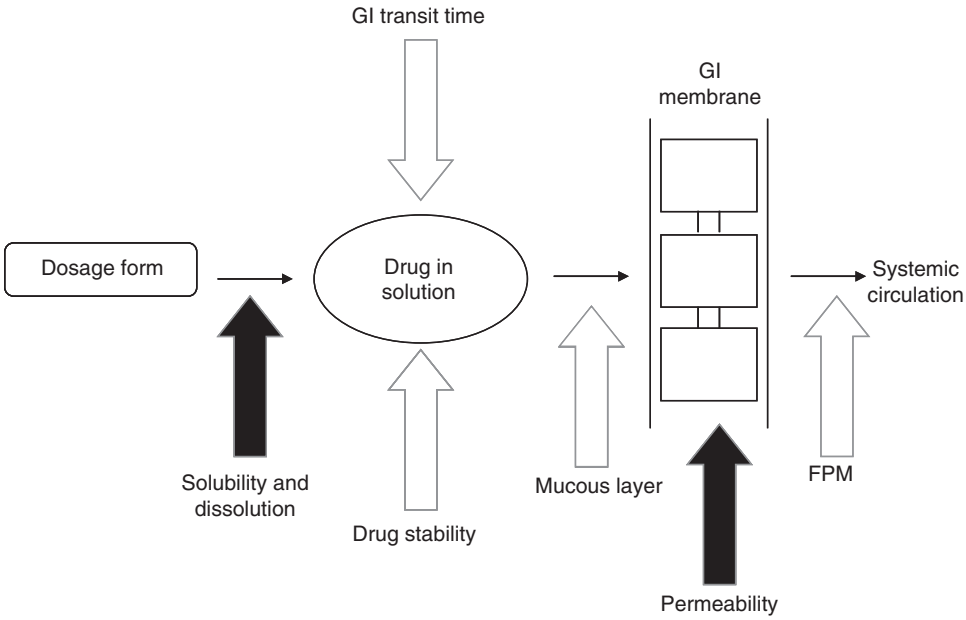


**FIGURE 13** Barriers to drug absorption. Drug from the dosage form should be soluble and dissolve in GI fluids before it can be absorbed. Drug dissolution is one of the major rate-limiting steps in drug absorption. Drug absorption may be affected if the drug is unstable in GI fluids. Furthermore, drug absorption will also depend on how long the drug resides in a particular region of the GIT. The drug has to diffuse through the highly viscous mucous layer before getting absorbed through the membrane. Membrane permeability is one of major the rate-limiting steps in absorption. After absorption the drug may be subject to first-pass metabolism (FPM) in the liver before reaching the systemic circulation. Dark arrows indicate that solubility and permeability are the most influential factors.

screen, detailed solubility and dissolution studies are carried out as the compound moves through different development phases (Figure 14*a*). A usual target for drug solubility during lead selection is 1 mg/mL for avoiding dissolution rate–limited absorption [46].

Drug permeability is commonly screened in the early discovery phase using CaCo-2 cell lines, which is a human colon carcinoma cell line. Using a 96-well format, the cell lines are used as a primary screen to rank order compounds based on permeability [4]. If designed properly, the cell culture studies can be used to understand the drug absorption mechanisms [47]. To avoid permeability-limited absorption in vivo, a drug should have a permeability coefficient of $2 \times 10^{-4}$–$4 \times 10^{-4}$ cm/s [48]. Inputs from the drug metabolism team (based on liver microsomal studies) can give clues on the drug's susceptibility to first-pass metabolism. Following cell culture studies, a select group of compounds are studied using isolated rat intestine and segmental absorption studies to understand the drug absorption mechanism and the site of drug absorption in the GIT (Figure 15*b* and 15*a*). This is further substantiated using in situ perfusion experiments in rats [49]. The details of the studies are depicted in Figure 15*b*. Some of these compounds are studied in whole animals of which a few may make it to clinical trials in humans.

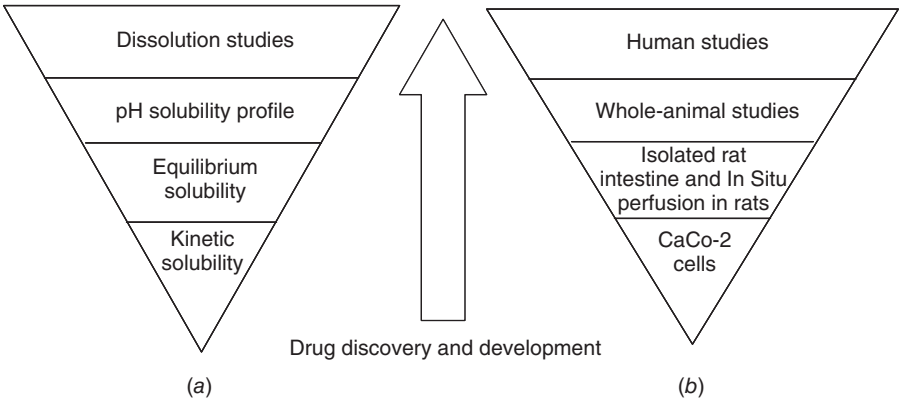


**FIGURE 14** Different methods to determine solubility and permeability during various stages of drug discovery and development. (*a*) In the initial stages the drugs are screened using kinetic solubility studies, which are later followed by equilibrium solubility studies, pH solubility profile, and dissolution studies during development phase. (*b*) Drug permeability is initially screened using CaCo-2 cell lines followed by rat intestinal studies. This is followed by pharmacokinetic studies in animals and finally the potential drug molecules are tested in humans in clinical trials.

Thus, the preformulation team in coordination with other discovery teams gets useful estimates on in vivo drug absorption. In addition to drug solubility and permeability, it is also important to consider the anticipated dose while assessing the absorption potential. A useful tool to optimize the drug's physicochemical properties is the maximum absorbed dose (MAD) model [46]. The model predicts the dose that would be absorbed based on the drug's solubility ($C_s$; solubility in intestinal pH of 6.5), absorption rate constant ($k_a$; usually obtained from rat permeability studies), physiological factors such as gastric transit time ($T_i$; approximated as 4.5 h), and intestinal fluid volume ($V_{int}$; approximated as 250 mL):

$$\mathrm{MAD} = C_s k_a V_{int} T_i \tag{19}$$

Using this model, the required solubility or permeability can be estimated for a given dose of the drug. For example, a drug with an anticipated human dose of 1 mg/kg (70 mg for a normal 70-kg subject) will require a solubility of 0.05 mg/mL provided the drug shows good permeability. Similarly, the absorption rate required to achieve the same dose for a drug with good solubility (1 mg/mL) is 0.001 $\mathrm{min}^{-1}$. The MAD model is helpful in guiding development teams on optimizing drug solubility and/or permeability.

A further refinement of this model led to the evolution of the biopharmaceutics classification system (BCS), which classifies the drugs into four classes depending on their solubility and permeability (Figure 16). The BCS is applicable only to the oral route of administration, and according to this classification, a drug is considered to be highly soluble if the highest dose of the drug is soluble in a glass of water (250 mL) covering the entire pH range in GIT from 1 to 7.5, and a drug is considered to be highly permeable if the drug has more than 90% oral bioavailability [48]. The model has been developed based on the solubility and permeability characteristics

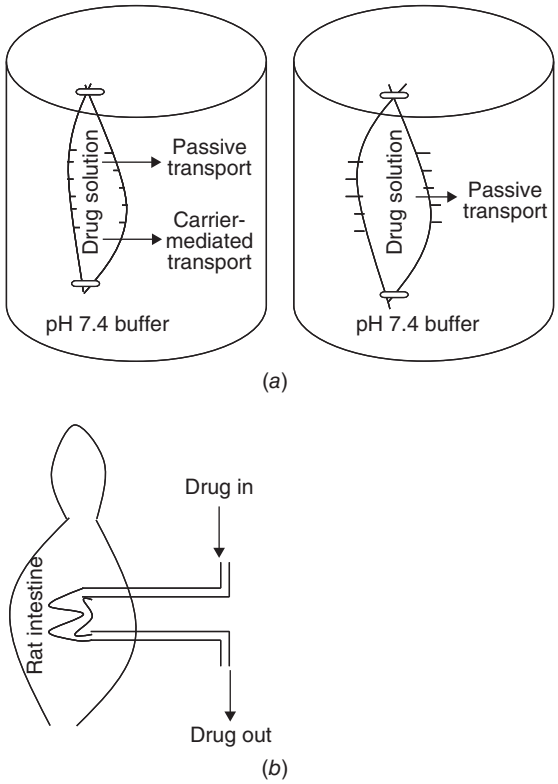


**FIGURE 15** Intestinal permeability studies. (*a*) Everted rat intestine is used to study the mechanism of drug absorption. Isolated rat intestine is filled with drug solution and tied at both ends and the drug permeation into the external medium is measured (left). In another set of experiment, the rat intestine is everted with the internal mucosal membrane facing outside (right) and the serosal side facing inward. If the drug is passively absorbed, then there would not be any differences in permeability in these two experiments. If the drug is transported by carriers, then drug permeation would be seen only in the first experiment, as the carriers are present only on the mucosal side. (*b*) In situ intestinal perfusion studies are conducted in an anesthetized animal and the drug solution is pumped through a tube and drug coming out on the other side of the intestine is measured. The drug is also measured by sampling from the jugular vein in the animal. This is useful to measure the dynamic drug absorption into systemic circulation.

of marketed drugs. However, BCS is bound to become an important biopharmaceutical tool in lead optimization in drug discovery setting and also at the same time serving as a useful guide to develop new formulations in a generic setting. From a regulatory perspective, BCS provides a scientific basis to grant biowaivers. As per BCS, the dissolution test can serve as a surrogate tool for costly and time-intensive bioequivalence studies for generic drugs which are highly soluble. This requires the establishment of a good in vitro–in vivo correlation. At present, the FDA grants biowaivers to generic manufacturers of immediate-release dosage forms for drugs in class I (highly soluble and highly permeable) provided they can prove the "dissolution equivalence" of their product to that of the innovator's drug product [50].

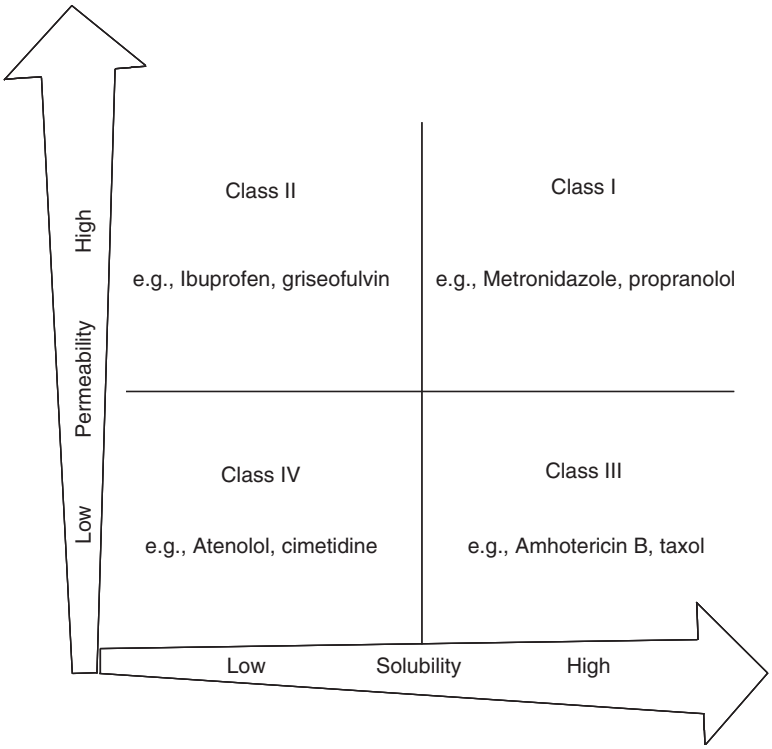


**FIGURE 16** Biopharmaceutics classification system. The drugs are classified based on drug solubility and drug permeability. A drug is considered to be highly soluble if the highest dose of the drug is soluble in 250 mL of water varying in pH from 1 to 7.8 (GIT pH range). A drug is considered highly permeable if more than 90% of the drug is bioavailable by oral route. Class I drugs are highly soluble and highly permeable, class II drugs are poorly soluble but highly permeable, class III drugs are highly soluble but poorly permeable, and class IV drugs have poor solubility and permeability.

The dissolution equivalence is tested using the statistical dissolution model, termed $f_2$ or similarity factor, and is described by the equation

$$f_2 = 50\log\left\{\left[1+\left(\frac{1}{N}\right)\sum_{t=1-n}(R_t+T_t)^2\right]^{-0.5}\times 100\right\} \tag{20}$$

where $N$ is the number of dosage form units (12 units are tested), $t$ is dissolution time points from 1 to $n$, $R_t$ is the percent drug dissolved for the reference drug product, and $T_t$ is the percent drug dissolved for the test product. Two dissolution profiles are considered similar, if $f_2 \geq 50$ and if the coefficient of variation does not exceed 20% at early dissolution time points (usually 10 min) and 10% at other time points in the pH range 1–7.5. However, if ≥85% of drug is dissolved in ≤15 min, then no comparison is required and the dissolution is based on a single time point determination [50].

The BCS paradigm can be used to develop strategies for enhancing drug solubility and/or permeability (Tables 8 and 9). Solubility enhancement may involve only

**TABLE 8 Solubility Enhancement Methods**

| Technique | Principle |
|---|---|
| Salt formation | Increases drug solubility by keeping the pH at which the drug is ionized. |
| Particle size reduction | Increase in surface area increases drug solubility. Particle size is reduced to nanodimensions (nanosuspensions) for increasing drug solubility. |
| Change of form | Crystalline drugs are converted to amorphous forms which show higher solubility than crystalline forms of the drug. |
| Cosolvents | Various water miscible cosolvents are used to increase the water solubility of drug. The cosolvents are selected based on the polarity of the drug. Common cosolvents include glycerol, ethanol, and propylene glycol. |
| Complexation | Drug is entrapped or complexed with excipients that can mask the lipophilic groups of the drug and enhance drug solubility in water. Cyclodextrins are commonly used to entrap hydrophobic drugs in the core, while the hydrophilic groups on the periphery help to solubilize the drug. |
| Surfactants | Surfactants are characterized by the presence of lipophilic and hydrophilic groups and form spherical structures known as micelles in water at a certain concentration. The hydrophobic drug is entrapped in the hydrophobic core of the micelle while the hydrophilic groups on the periphery help to solubilize the drug. |
| Disperse systems | The hydrophobic drug is dissolved in an organic solvent and in addition may also contain an emulsifier. On contact with the intestinal fluids, the drug is emulsified (microemulsions) by bile salts and is absorbed through the intestine. |

**TABLE 9 Permeability Enhancement Methods**

| Method | Mechanism |
|---|---|
| Transcellular transport | Sorption promoters can be used to enhance the transcellular transport in the intestine, including bile salts and fatty acid esters. They tend to fluidize the lipid bilayer and enhance drug permeation across the membrane. |
| Paracellular transport | Enhancement is achieved by modulating the tight junctions between the cells. Chelating agents such as ethylenediamine tetraacetic acid can chelate calcium ions and transiently open the tight junctions for drug transport. |
| Carrier-mediated transport | Nutrient transport carriers are utilized for drug transport. Prodrugs are designed to meet the structural requirements for carrier-mediated transport. |
| Blocking efflux pump | P-glycoprotein is a major efflux mechanism that pumps out drug from the intestinal cells back into the intestinal fluid. Several drugs and food substances are known to inhibit p-glycoprotein and enhance drug permeation. |

a physical intervention, as opposed to molecular modification for permeability enhancement. Generally [46], the permeability range of drugs varies by only 50-fold (0.001–0.5 $min^{-1}$) in comparison to drug solubility, which varies by six orders of magnitude (0.1 μg/mL–100 mg/mL). Hence, the formulator has greater flexibility in altering the drug's solubility in comparison to altering the drug's permeability. This

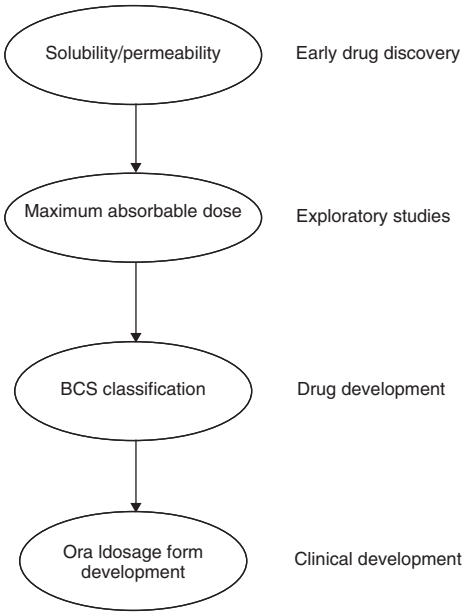


**FIGURE 17** Flow chart for determining absorption potential of a drug during drug discovery and development. Solubility and permeability studies from preclinical phase are used to calculate the maximum absorbable dose and, when correlated to BCS, this can provide information on its biopharmaceutics class. This is useful to estimate if the drug absorption would be dissolution and/or permeability limited for developing appropriate drug delivery strategies.

is evident form the fact [51] that the majority of the marketed drugs are highly soluble (>55% in classes I and III). Sometimes, enhancing the permeability by altering the drug's chemical structure may be counterproductive. This because of the associated increase in molecular weight that attenuates the permeability enhancement gained with structural modification. On the other hand, optimization of drug solubility may be more fruitful if the poor permeability is overcome by increasing drug solubility to provide a high drug concentration at the absorption site. However, this may be a difficult strategy if the dose is very high. The preformulation team should use the "appropriate tools" at every stage of the drug discovery and development process to guide or alert other development teams about drug solubility and permeability issues (Figure 17).

## 4 STABILITY CHARACTERISTICS

Drug stability is an essential component of preformulation testing. Establishing the stability of the drug under a variety of conditions is an expensive and time-consuming process. It cannot be overemphasized that the availability of a stability-indicating assay is the key to stability studies. The preformulation scientist works closely with the analytical method development team in developing a stability-indicating assay. During the early stages, a foolproof stability-indicating assay may

not be available. This is understandable considering the fact that the initial drug lots are not pure and the purity is improved as the molecule progresses to subsequent development stages. Therefore, the intention of the preformulation scientist is not to generate a thorough kinetic rate profile of the drug but to broadly define the conditions under which the drug would be stable. Only relevant stability data are generated during various phases for developing preclinical and clinical formulations [52]. The stability data are evaluated and conveyed to the formulation development team upfront if stabilization and additional packaging requirements are needed. Often the chemical structure of the drug can give clues on the drug's degradation pathway and its stability characteristics [53]. Table 10 lists some of the functional groups that are susceptible to common degradation pathways.

Stability test conditions are chosen keeping in mind the environment that the drug will encounter during drug development, processing, storage, and use (Table 11). One parameter is studied at a time while keeping all other parameters constant. Apart from classical stability studies, such as hydrolysis, oxidation, and photolysis, the preformulation scientist has to determine the stability of drugs in unconventional matrices such as animal feed used for toxicological studies [34]. The stability of a drug in the animal feed is complicated by the feed composition, including enzymes and vitamins among others. The moisture content in the feed may also vary with storage temperature. In such cases, it is appropriate to study the stability of the drug under the storage conditions encountered in the toxicological laboratory. Sensitive techniques such as liquid chromatography/mass spectrometry (LC/MS) are used to evaluate drug stability in such complex mixtures.

In general, most of the drugs undergo first-order degradation, while some drugs may follow zero-order kinetics and only a few drugs undergo second-order degradation [34]. The first- and zero-order reactions are readily differentiated by studying drug stability at two different initial drug concentrations. First-order kinetics will depend on initial drug concentration, while a zero-order reaction will be

**TABLE 10 Groups Susceptible to Common Degradation Pathways**

| Degradation Pathway | Functional Groups |
|---|---|
| Hydrolysis | Esters, lactones, amides, lactams, oximes, imides, and malonic ureas |
| Oxidation | Amines, sulfides, disulfides, sulfoxides, phenol anions, thiols, nitriles, and catechols |
| Photolysis | Aromatic hydrocarbons, aromatic heterocyclics, aldehydes, and ketones |

**TABLE 11 Stability Testing Conditions**

| Parameters | Conditions |
|---|---|
| Temperature | 5, 25, 30, 37, 40, and 60 °C |
| Moisture | 30, 45, 60, 75, and 90% RH |
| pH | 1, 3, 5, 7, and 9 at room and body temperature |
| Oxygen | Sparging with 40% oxygen or adding 100–200 ppm of hydrogen peroxide |
| Light | 1.2 million lux hours of exposure to visible light and 200 h/m$^2$ exposure to UV light (360–400 nm) |

independent of initial drug concentration. Accelerated stability studies are conducted to expedite the degradative reactions, where temperature is the commonly used accelerant. The influence of temperature on drug stability kinetics is described by the Arrhenius equation:

$$k = Ae^{-E_a/RT} \tag{21}$$

where $k$ is a reaction rate constant, $A$ is a frequency factor, $E_a$ is activation energy, $R$ is the gas constant, and $T$ is absolute temperature. According to the Arrhenius equation, every 10 °C rise in temperature increases the reaction rate by two- to five-fold [31]. The usual temperatures selected for early stability studies include 5, 25, 37, 40, and 60 °C to cover the temperatures encountered in processing, use, and storage of the drug product. Using the Arrhenius equation, the rate constant from higher temperatures can be extrapolated to determine the stability at room temperature [31]. The slope of the plot of the reciprocal of temperature and the rate constant gives the activation energy. The activation energy usually varies between 15 and 60 kcal/mol with a mean value of 19.8 kcal/mol [33]. A break in the line is usually indicative of change in the activation energy due either to change in the reaction order or the mechanism of degradation at higher temperature. In such cases, it becomes imperative to conduct detailed studies to understand the drug degradation mechanism. Some of the reactions seen at higher temperature may not be representative of the reactions at room temperature. Hence, short-term high-temperature studies should be supplemented with long-term real-time stability testing at room temperature. Additionally, the drug is also exposed to moisture, oxygen, and UV light (250–360 nm). The conditions used for stress studies may vary depending on the drug type and the drug development stage [54]. The stability studies in this chapter are discussed with respect to a solid dosage form which includes solid-state stability, limited liquid state stability, and drug–excipient compatibility.

### 4.1 Solid-State Stability

In general, solid-state reactions are slow, complex, and at times difficult to quantify. They may manifest as either physical and/or chemical instabilities. Physical instabilities include solid–solid transformations, desolvation of hydrates, and change in color [34]. On the other hand, chemical instabilities may involve a change in drug content as a result of hydrolysis, oxidation, or light-induced degradations [32]. The influence of temperature is studied by exposing the solid drug to increasing temperatures and also exposing the drug to various relative humidities at room temperature for two to eight weeks (Table 11). If substantial change is observed at higher temperatures, the drug samples stored at 5 °C are analyzed. If no degradation is seen at higher temperature, then none can be expected at room temperature. The results from higher temperature should be carefully interpreted. For instance, a hydrate may lose moisture at higher temperatures and make a drug unstable which otherwise would be stable at room temperature. Similarly, chlortetracycline hydrochloride converts from the β form to the α form at above 65% relative humidity, in contrast to <65% relative humidity, where no transformations are observed [32].

Oxidative degradation is studied by exposing the sample to an atmosphere of 40%. The oxygen is combined with heat to accelerate the reaction. The results

should be compared with samples stored in an inert atmosphere. Photolytic degradation is studied by exposing the sample to UV light (254 and 360 nm) for two to four weeks and observed for color fading and/or darkening [33]. Color fading may not always mean drug decomposition. It may be just a physical change which can be overcome by including dyes in the formulation. As indicated earlier, the solid-state reactions may be slow, sometimes generating only 1–5% of degradation products, which may be below the detection limit of HPLC [34]. A combination of qualitative and quantitative analytical tools is helpful to detect drug degradation. Nevertheless, accelerated stability studies provide an early warning of potential drug degradation and the preformulation scientist should use discretion in interpreting and sharing the results with other development teams.

### 4.2 Solution-State Stability

Detailed solution-state stability is of limited value if the final anticipated dosage form is a solid. However, solution-state stability studies can predict the stability of the drug in the granulating fluid and the GI fluids. Solution-state reactions are faster than the solid-state reactions. This is helpful in generating degradation products through so-called forced-degradation studies for toxicological screening and analytical method development. The studies are conducted by exposing the drug to extreme conditions such as 0.1 *N* HCl, 0.1 *N* NaOH, and water at 90 °C [34]. Forced-degradation studies are useful to qualify the safety of the degradation products if it exceeds 0.1 or 0.05% of total daily dose for drugs with <1 g or >1 g dose/day, respectively [6].

The pH rate profile is an important parameter that is studied in the solution state. In early preformulation studies, an approximate pH rate profile is generated, usually including the pH encountered in salt selection and in vivo [55]. The studies are later followed with a detailed pH profile in the whole pH range of 1–10. A typical pH rate profile is shown in Figure 18, which is useful to extract useful information on

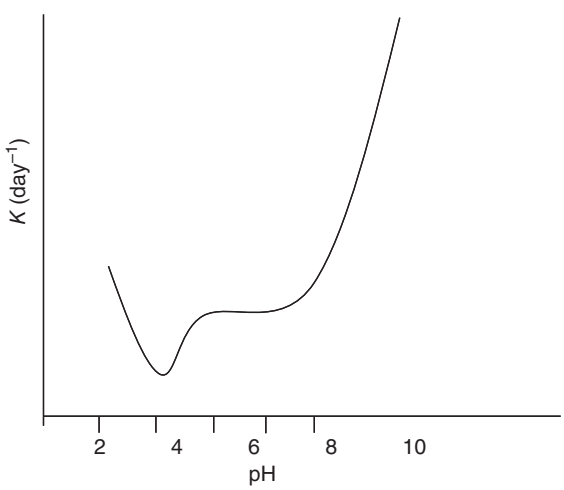


**FIGURE 18** Representative pH kinetic rate profile. The drug shows a minimal degradation at around 2–4 and the degradation rate is high in alkaline pH.

drug stability. The profile is used to predict if the drug degradation is catalyzed by hydronium or hydroxyl species (specific acid–base catalysis). The minimal point in the profile is indicative of the pH at which the drug is relatively stable. However, the buffer used in the study can by itself accelerate the reaction, which is referred to as general acid–base catalysis [39]. The influence of buffer is nullified by conducting the study at one to three buffer concentrations and then plotting the rate constant against the buffer concentration. The intercept on the *y* axis gives the rate constant at zero buffer concentration.

Oxidation reactions are studied by passing oxygen in the head space of the drug solution and comparing the drug degradation with a drug solution (Table 11) filled with an inert gas in the head space. The reduced solubility of oxygen at higher temperatures may lead to an apparently reduced reaction rate in comparison to lower temperature [34]. Light sensitivity is studied by exposing the drug solution in a clear-flint bottle to UV radiation and comparing the results with the drug solution in an amber-colored container [55].

The Solution-state studies should be extrapolated to the solid state with caution. The reaction in the solution state is usually done in a dilute drug solution, in contrast to reactions in the solid state, where the saturated drug solution at the surface undergoes a multiphase reaction [33]. Moreover, the reaction order may be different in the solution and solid states. Due to the excess solvent, reactions in the solution state are usually pseudo–first order as opposed to first-order or zero-order reactions in the solid state. In spite of these limitations, solution-state studies provide clues in selecting appropriate granulating solvent and in predicting in vivo drug stability. The pH–rate profile data are also useful to predict the solid-state stability of salt forms or the stability of a drug in the presence of acidic and basic excipients [55]. Further, the pH–rate studies predict the stability of drug and its salt in the gastrointestinal pH. This is illustrated by the example of erythromycin and its salts [32]. Erythromycin is rapidly inactivated in the acidic environment of the stomach. This is overcome by using insoluble erythromycin estolate, which is stable in the gastric pH, unlike the other salts, which are easily displaced by hydrochloric acid in the stomach.

### 4.3 Drug–Excipient Compatibility

Excipients are the backbone of a solid dosage form, and they function as diluents, binders, disintegrants, and fillers. The excipients are in intimate contact with the drug in a solid dosage form, therefore necessating the need for drug–excipient incompatibility testing. Since the formulator has a wide selection of excipients to choose from, it would be a daunting task for the preformulation scientist to screen all possible excipients. A general practice is to select those excipients that are routinely used in in-house manufacturing of dosage forms. At least two are selected from each class of functional excipients. Excipients that are known to cause potential incompatibilities (e.g., glucose and amines) with drugs are excluded from the study [34]. The excipients are intimately mixed with the drug in the ratio that is realistic for the desired solid dosage form. For example, diluents are mixed in the ratio 20:1, while other excipients are used in the ratio 1:5 with drug [32]. The drug–excipient mixtures are then stored in a tightly sealed container (ampoule or in a bottle, where the cap is sealed with wax) in the presence and absence of 5%

w/v moisture. The mixture is stored at 40 and 25 °C and analyzed for three months. The mixture is physically observed for caking, liquefaction, and gas formation and chemically analyzed for drug degradation. A variety of techniques are used for studying drug–excipient compatibility. DSC is a useful primary screen which can rapidly detect potential drug–excipient incompatibilities [33]. For DSC studies, the drug and excipient are mixed together without moisture and are analyzed within a few hours of preparation. Moisture is avoided, as it complicates the interpretation of thermograms. If there is no interaction, then the thermogram should be the sum of thermogram of drug and excipient. The thermogram is observed for peak position, peak appearance, peak shape, appearance of new peaks, disappearance of drug peaks, peak shift, and change in enthalpy values. It is obvious that a mixture of two substances will result in depression of melting point, but if the change is significant, then it may be indicative of eutectic formation. Some of the reactions seen at high temperatures in DSC may not be reflective of room temperature. Therefore, excipients that are suspected to show incompatibilities are screened further through stability studies using TLC and HPLC (Figure 19). TLC is used to qualitatively detect any new spots with the drug–excipient mixture. The spots should be compared with the control drug sample. This is important since in the early discovery stage the drug is not pure and contains impurities. The findings from TLC should be corroborated using HPLC. The studies may be further followed-up with stability studies using multicomponent mixtures. In addition, the drug and excipient may be tested by compressing in a hydraulic press or filling in a capsule to simulate the actual dosage form [32]. A well-designed drug–excipient preformulation study can thus help the formulator in judicious choice of excipients for the final dosage form.

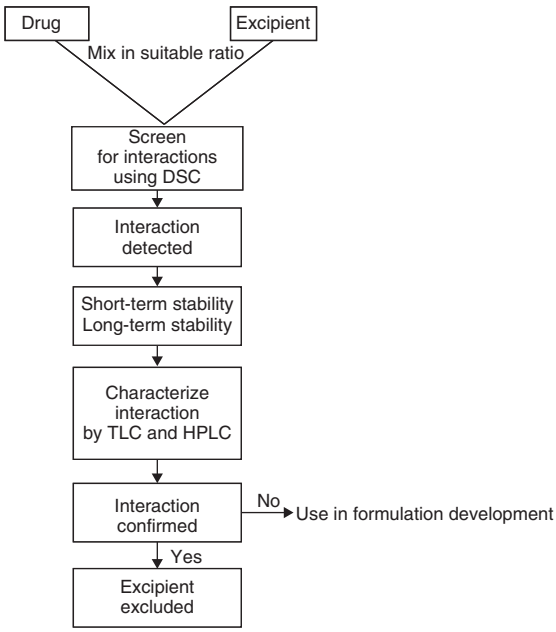


**FIGURE 19** Flow chart for studying drug–excipient compatibility.

## 5 CONCLUSIONS

Preformulation testing has a significant role in a drug discovery and development program, as it provides valuable feedback to the various discovery and development teams in enabling druggability during lead identification and optimization. The preformulation data may mean different things to different groups in the discovery/development phase (Figure 20). For the chemistry team, the feedback from preformulation testing provides clues to optimize the chemical structure with respect to solubility, permeability, and stability. Preformulation studies give inputs to the biology group for ensuring optimal drug exposure based on solubility, permeability, and stability data, in addition to developing preclinical formulations for pharmacokinetic and toxicological studies. The analytical team gets inputs from the preformulation group on developing stability-indicating assays and setting drug specifications. Once the lead is selected and as the molecule moves to the development phase, the preformulation group provides guidance to the bulk manufacturing, formulation, and clinical evaluation teams. The bulk manufacturing team uses the data generated from the preformulation studies on salts, polymorphic purity, and particle size specifications. It is the formulation team that utilizes the maximum data from the preformulation testing to design an appropriate dosage form. The physicochemical properties are utilized for improving the drug's solubility, improving the drug's permeability, developing stabilization strategies, selecting appropriate excipients, selecting processing conditions to design, and evaluating the final dosage form. The clinical evaluation team utilizes the preformulation data along with the preclinical animal studies to understand the drug's pharmacokinetics in humans through the MAD and BCS paradigms. Therefore, a strong preformulation team in a drug discovery setting is critical for optimizing the pharmaceutical properties of the drug. This can significantly reduce the attrition rate, time, and cost of discovering a new drug. On the other hand, the preformulation team in a generic setting is valuable to optimize or further enhance the efficacy of an existing drug by designing a new drug delivery system and thus giving a new life to an old drug. As opposed

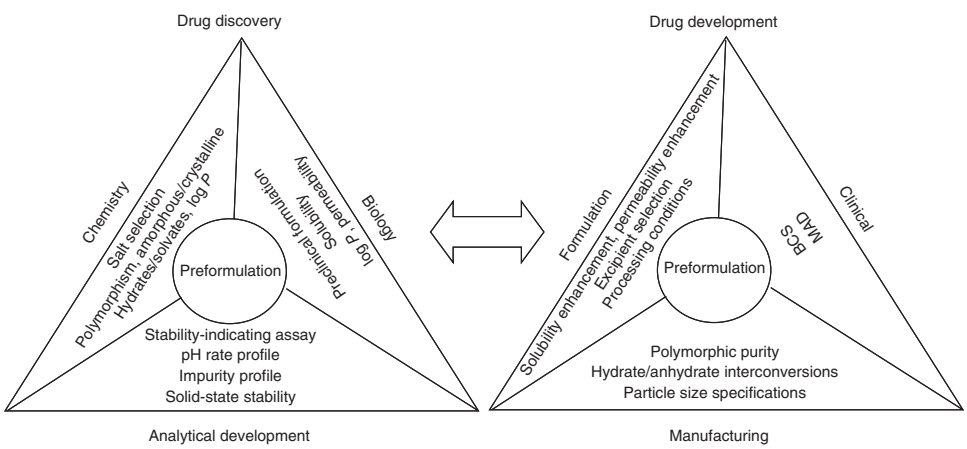


**FIGURE 20** Role of preformulation in supporting other discovery and development teams.

to discovering a new drug, a new drug delivery system can be designed at one-third of the cost and money involved in a new drug discovery process [56]. The availability of high-throughput property screens and predictive models is expected to improve the efficiency and maximize the output of preformulation data in the coming years. However, the main challenge lies in intelligent use of such tools to make sense from loads of data and communicate the appropriate information to the relevant discovery/development groups for getting the drug to the marketplace in time.

## REFERENCES


1. Waterbeemd, H. V., and Gifford, E. (2003), ADMET in silico modeling: Towards prediction paradise, *Nat. Rev. Drug Discov.*, 2, 192–204.
2. Kerns, E. H., and Di, L. (2003), Physicochemical profiling in drug discovery, *Drug Discov. Today*, 8, 316–323.
3. Lennarnas, H., and Abrahamsson, B. (2005), The use of biopharmaceutical classification of drugs in drug discovery and development: Current status and future extensions, *J. Pharm. Pharmacol.*, 57, 273–285.
4. Lipinski, C., Lombardo, F., Dominy B. W., and Feeney, P. J. (2001), Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings, *Adv. Drug Deliv. Rev.*, 46, 3–26.
5. Giron, D. (1998), Contribution of thermal methods and related techniques to the rational development of pharmaceuticals. Part I, *Pharm. Sci. Technol. Today*, 1, 191–199.
6. Ahuja, S. (2007), Assuring quality of drugs by monitoring impurities, *Adv. Drug Deliv. Rev.*, 59, 3–11.
7. Datta, S., and Grant, D. J. W. (2004), Crystal structures of drugs: Advances in determination, prediction and engineering, *Nat. Rev. Drug Discov.*, 42, 42–57.
8. Tiwary, A. K. (2004), Modification of crystal habit and its role in dosage form performance, *Drug Dev. Ind. Pharm.*, 27, 699–707.
9. Garekani, H. A., Ford, J. L., Rubinstein, M. H., and Rajabi-Sahboomi, A. R. (1999), Formation and compression characteristics of prismatic, polyhedral and thin plate-like crystals of paracetamol, *Int. J. Pharm.*, 187, 77–89.
10. Gu, C., Li, H., Gandhi, R. B., and Raghavan, K. (2004), Grouping solvents by statistical analysis of solvent property parameters: Implication to polymorph screening, *Int. J. Pharm.*, 283, 117–125.
11. Giron, D. (2001), Investigations of polymorphism and pseudopolymorphism in pharmaceuticals by combined thermoanalytical techniques, *J. Thermal Anal. Calor.*, 64, 37–60.
12. Threafall, T. (2003), Structural and thermodynamic aspects of Ostwald's rule, *Org. Proc. Res. Dev.*, 7, 1017–1027.
13. Vippagunata, S. R., Brittain, H. G., and Grant, D. J. W. (2001), Crystalline solids, *Adv. Drug Deliv. Rev.*, 48, 3–26.
14. Phadnis, N. V., Cavatur, R. K., and Suryanarayanan, R. (1997), Identification of drugs in pharmaceutical dosage forms by x-ray powder diffractometry, *J. Pharm. Biomed. Anal.*, 15, 929–943.
15. Haleblain, J., and McCrone, W. (1969), Pharmaceutical applications of polymorphism, *J. Pharm. Sci.*, 58, 911–929.

16. Huang, L., and Tong, W. (2004), Impact of solid state properties on developability assessment of drug candidates, *Adv. Drug Deliv. Rev.*, 56, 321–334.
17. Morris, K. R., Griesser, U. J., Eckhardt, C. J., and Stowell, J. G. (2001), Theoretical approaches to physical transformations of active pharmaceutical ingredients during manufacturing process, *Adv. Drug Deliv. Rev.*, 48, 91–114.
18. Dunitz, J. A., and Bernstein, J. (1995), Disappearing polymorphs, *Acc. Chem. Res.*, 28, 193–200.
19. Chemburkar, S. R., Bauer, J., Deming, K., Spiwek, U., Patel, K., Morris, J., Henry, R., Spanton, S., Dziki, W., Porter, W., Quick, J., Bauer, P., Donaubauer, J., Narayanan, B. A., Soldani, M., Riley, D., and McFarland, K. (2004), Dealing with the impact of ritonavir polymorphs on the late stages of bulk drug process development, *Org. Proc. Res. Dev.*, 4, 413–417.
20. Morissette, S. L., Almarsson, O., Peterson, M. L., Remenar, J. F., Read, M. J., Lemmo, A. V., Ellis, S., Cima, M. J., and Gardner, C. R. (2004), High-throughput crystallization: Polymorphs, salts, co-crystals and solvates of pharmaceutical solids, *Adv. Drug Deliv. Rev.*, 56, 275–300.
21. Davey, R. J. (2003), Pizzas, polymorphs and pills, *Chem. Commun.*, 13, 1463–1467.
22. Khankari, R. K., and Grant, D. J. W. (1995), Pharmaceutical hydrates, *Thermochim. Acta*, 248, 61–79.
23. Giron, D., Goldbrown, Ch., Mutz, M., Pfeffer, S., Piechon, Ph., and Schwab, Ph. (2002), Solid-state characterization of pharmaceutical hydrates, *J. Therm. Anal. Calor.*, 68, 453–465.
24. Byrn, S., Pfeffer, R., Ganey, M., Hoiberg, C., and Poochikian, G. (1995), Pharmaceutical solids: A strategic approach to regulatory considerations, *Pharm. Res.*, 12, 945–954.
25. Craig, D. Q. M., Royall, P. G., Kett, V. C., and Hopton, M. C. (1999), The relevance of the amorphous state to pharmaceutical dosage forms, *Int. J. Pharm.*, 179, 179–207.
26. Yu, L. (2001), Amorphous pharmaceutical solids: Preparation, characterization and stabilization, *Adv. Drug Deliv. Rev.*, 48, 27–42.
27. Hancock, B. L., and Parks, M. (2000), What is the true solubility advantage for amorphous pharmaceuticals? *Pharm. Res.*, 17, 397–404.
28. Shah, B., Kakamanu, V. K., and Bansal, A. K. (2006), Analytical techniques for quantification of amorphous/crystalline phases in pharmaceutical solids, *J. Pharm. Sci.*, 95, 1641–1665.
29. Giron, D. (2003), Characterization of salts of drug substances, *J. Thermal Anal. Calor.*, 73, 441–457.
30. Shekunov, B. Y., Chattopadhyay, P., Tong, H. H., and Chow, H. H. (2007), Particle size analysis in pharmaceutics: Principles, methods and applications, *Pharm. Res.*, 24, 203–227.
31. Sinko, P. (2006), *Martin's Physical Pharmacy and Pharmaceutical Sciences*, 5th ed., Lippincott Williams and Wilkins, Philadelphia, pp 533–560.
32. Wadke, D. A., and Jacboson, H. (1980), Preformulation testing, in Liberman, H. A., and Lachman, L., Eds., *Pharmaceutical Dosage Forms: Tablets*, Vol. 1, Marcel Dekker, New York, pp 1–59.
33. Wells, J. (2005), Pharmaceutical preformualtion, in Aulton, M. E., Ed., *Pharmaceutics, The Science of Dosage Form Design*, Churchill Livingstone, Edinburgh, pp 113–138.
34. Fiese, E. F., and Hagen, T. A. (1986), Preformulation, in Lachman, L., Liberman, H. A., and Kanig, J. A., Eds., *The Theory and Practice of Industrial Pharmacy*, 3rd ed., Lea and Febiger, Philadelphia, pp 171–196.

35. Krens, E. H. (2001), High-throughput physicochemical profiling for drug discovery, *J. Pharm. Sci.*, 90, 1838–1858.
36. Strickley, R. G. (2004), Solubilizing excipients in oral and injectable formulations, *Pharm. Res.*, 21, 201–229.
37. Bastin, R. J., Bowker, M. J., and Slates, B. J. (2000), Salt selection and optimization procedures for pharmaceutical new chemical entities, *Org. Proc. Res. Dev.*, 4, 427–435.
38. Haynes, D. A., Jones, W., and Motherwell, W. D. S. (2005), Occurrence of pharmaceutically acceptable anions and cations in the Cambridge structural databases, *J. Pharm. Sci.*, 94, 2111–2120.
39. Florence, A. T., and Atwood, D. (2005), *Physicochemical Principles of Pharmacy*, 4th ed., Pharmaceutical Press, London, pp 55–92.
40. Ando, H. Y., and Radebaugh, G. W. (2006), Property based drug design and preformulation, in Beringer, P., DerMarclerosian, A., Felton, L., Gelone, S., Gennaro, A. R., Gupta, P. K., Hoover, J. E., Popovich, N. J., Reilly Jr., W. J., and Hendrickson, R., Eds., *Remington's: The Science and Practice of Pharmacy*, Lippincott Williams and Wilkins, Philadelphia, pp 720–744.
41. Leo, A., Hansch, C., and Elkins, D. (1971), Partition coefficient and their uses, *Chem. Rev.*, 71, 524–616.
42. Dokoumetzidis, A., and Macheras, P. (2006), A century of dissolution research: From Noyes-Whitney to the biopharmaceutics classification system, *Int. J. Pharm.*, 321, 1–11.
43. Azaemi, A., Roa, W., and Lobenberg, R. (2007), Current perspectives in dissolution testing of conventional and novel dosage forms, *Int. J. Pharm.*, 328, 12–21.
44. Costa, P., and Lobo, J. M. S. (2001), Modeling and comparison of dissolution profiles, *Eur. J. Pharm. Sci.*, 13, 123–133.
45. Yu, L. X., Carlin, A. S., Amidon, G. L., and Hussain, A. S. (2004), Feasibility studies of utilizing disk intrinsic dissolution rate to classify drugs, *Int. J. Pharm.*, 270, 221–227.
46. Curatolo, W. (1998), Physical chemical properties of oral drug candidates in the discovery and exploratory development settings, *Pham. Sci. Technol. Today*, 1, 387–393.
47. Artursson, P., Palm, K., and Luthman, K. (2001), Caco-2 monolayers in experimental and theoretical predictions of drug transport, *Adv. Drug Deliv. Rev.*, 46, 27–43.
48. Amidon, G. L., Lennernas, H., Shah, V. P., and Crison, J. R. (1995), A theoretical basis for a biopharmaceutic drug classification: The correlation of in vitro drug product dissolution and in vivo bioavailability, *Pharm. Res.*, 12, 413–419.
49. Stewart, B. H., Chan, O. H., Jezyk, N., and Fleischer, D. (1997), Discrimination between candidates using models for evaluation of intestinal absorption, *Adv. Drug Deliv. Rev.*, 23, 27–45.
50. U.S. Department of Health and Human Services (DHHS) (2000, Aug.), Guidance for Industry: Waiver of in vivo bioavailability and bioequivalence studies for immediate release solid oral dosage forms based on a biopharmaceutics classification system, DHHS, Food and Drug Administration, Center for Drug Evaluation and Research, Washington, DC.
51. Yamashita, S., Yu, L. X., and Amidon, G. L. (2006), A provisional biopharmaceutical classification of the top 200 oral drug products in the United States, Great Britain, Spain and Japan, *Mol. Pharm.*, 3, 631–643.
52. Xue-Qing, C., Melissa, D. A., Christoph, G., and Gumundsson, S. O. (2006), Discovery pharmaceutics—Challenges and opportunities, *AAPS J.*, 8, E402–E408.
53. Guillory, J. K., and Poust, R. I. (2002), Chemical kinetics, in Banker, G. S., and Rhodes, C. T., Eds., *Modern pharmaceutics*, Marcel Dekker, New York, pp 139–166.

54. IFAMA (2003), ICH stability testing of new drug substances and products, International Federation of Pharmaceutical Manufacturers Associations (IFPMA), Geneva.
55. Carstensen, J. (2002), Preformulation, in Banker, G. S., and Rhodes, C. T., Eds., *Modern Pharmaceutics*, Marcel Dekker, New York, pp 167–185.
56. Pillai, O., Dhanikula, A. B., and Panchagnula, R. (2001), Drug delivery: An odyssey of 100 years, *Curr. Opin. Chem. Biol.*, 5, 439–446.

# TABLET DESIGN

Eddy Castellanos Gil,[1,2,3] Isidoro Caraballo,[2] and Bernard Bataille[3]

[1]*Center of Pharmaceutical Chemistry and University of Havana, Havana, Cuba*
[2]*University of Sevilla, Seville, Spain*
[3]*University of Montpellier 1, Montpellier, France*

**Contents**

1 Introduction
2 Tableting Cycle
3 Limitations for Direct Compression
4 Previous Granulation: Biopharmaceutical Versus Technological Properties
5 Tablet Design for Matrix System
  5.1 Controlled-Release Tablet by Direct Compression and Wet Granulation
6 Tablet Design with Natural Products
  6.1 Tablet Design from Aqueous Plant Extract
  6.2 Natural Product as Vehicle for Manufactured Tablets
  6.3 Natural Product as Vehicle for Controlled-Release System
  6.4 Mechanism of Soluble Principle Active Propranolol Hydrochloride and Lobenzarit Disodium from Dextran Tablets
7 Design Tools of Tablet Formulation
  7.1 MODDE 4.0
  7.2 iTAB
  7.3 Percolation Theory
  7.4 Artificial Neural Networks
8 Coating Systems
  8.1 Subcoating of Tablet Cores as a Barrier to Water
  8.2 Kollidon VA 64
  8.3 SEPIFILM
9 Development of Pharmaceutical Tablets Using Percolation Theory
  9.1 Case Study: Optimization of Inert Matrix Tablets for Controlled Release of Dextromethorphan Hydrobromide
  9.2 Critical Points of Hydrophilic Matrix Tablets

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

9.3 Case Study: Estimation of Percolation Thresholds in Acyclovir Hydrophilic Matrix Tablets
10 Ultrasound-Assisted Tableting (a New Perspective)
References

## 1 INTRODUCTION

The compression of powders—the fourth state of matter in the words of Hans Leuenberger—is a complex task due to the variability in particle size and shape, even for particles of the same component, the unknown distribution of the particles in the die, their different ability to flow, and the forces needed to create bonds between them.

Science-based formulation is nowadays a good strategy for the development of a pharmaceutical formulation. The new regulatory environment (process analytical technology) will transform science-based formulation in the next years.

Therefore, there is a need for statistical tools able to predict the behavior of a powder mixture when it is subjected to compression forces. The purpose of this chapter is to provide a brief description of the main theoretical aspects of tablet design and formulation together with practical examples of tablet development and characterization using different techniques.

## 2 TABLETING CYCLE

Compressing powder or granule into a tablet is one of the simplest and oldest ways of forming a product known to humans. Nowadays, as technology has advanced, more complex machines are used with different procedures, but the basic principle, the tabletting cycle, remains the same. In tablet design many factors have to be taken into account, such as the physicolchemical properties of active compound and excipients. An important role also has to be attributed to tableting machines.

Tablet press subclasses primarily are distinguished from one another by how the powder blend is delivered to the die cavity. Tablet presses can deliver powders without mechanical assistance (gravity) (e.g., Ronchii, Manesty, Stokes, and Colton machines), with mechanical assistance (power assisted) (e.g., Ronchii, Courtoy, Kilian, Manesty, Kikusui, Fette, and Hata machines), by rotational forces (centrifugal) (e.g., the Comprima machine), and in two different locations where a tablet core is formed and subsequently an outer layer of coating material is applied (compression coating) (e.g., Kilian Manesty, and Kikusui machines).

The basic unit of any tablet press is tooling consisting of two punches and a die, called a station. The upper and lower punches come together in the die that contains the tablet formulation. Principally, two different types of machines are used, the eccentric and the rotary press. The eccentric press produces about 50–130 tablets per minute. The rotary press has a multiplicity of stations arranged on a rotating table with the dies. A few or many thousands of tablets can be produced per minute. There are numerous models of presses, manufactured by a number of companies, ranging in size, speed, and capacity.

The eccentric press is widely used early in the development stage, because the tabletting machine and the tooling are inexpensive, it can be easily instrumented, little material is needed, and setting, servicing, cleaning, and tool changeover of the machine are easy. During the manufacturing process the tablet mixture is dosed by a hopper into the die. The position of the lower punch defines the volume of the subsequent tablet mass. The compression force is given by the position of the upper punch, which defines its immersion depth into the die, and the reagent force that is built up during the densification of the material. The compressed tablet is ejected by the lower punch.

During the compression process on an eccentric press, there are other pressure ratios at the upper and lower punches. The pressure at the upper punch is usually higher than the pressure at the lower punch. A part of the pressure is lost in the material and in the resulting radial friction force against the die wall during the compression [1, 2].

Figure 1 shows tablets (300 mg) of native dextran obtained from *Leuconostoc mesenteroides* B-512F (Sigma) with a 10% water (w/w) punch in an eccentric machine (Manesty) at 15 kN with tablet side (flat-flace diameter) 10 mm. Axial displacement of water was observed according to the change in color.

On a rotary press the filling of the die and the following compression process is done at the same time at different stations. The compression is carried out in the simplest case with two rolls touching the upper and lower punches and compressing the powder mixture. In contrast to the eccentric press, the upper and lower punches exert pressure on the tablet mixture from both sides at the same time.

Tablets compressed on a rotary press generally show a more consistent hardness when the upper and lower sides of the tablets are compared, where the upper side of tablets compressed on an eccentric press is usually harder than the lower side [3].

Figure 2 illustrates the difference in compression profiles of the upper and lower punches and the punch movement with fictitious rotary and eccentric presses.

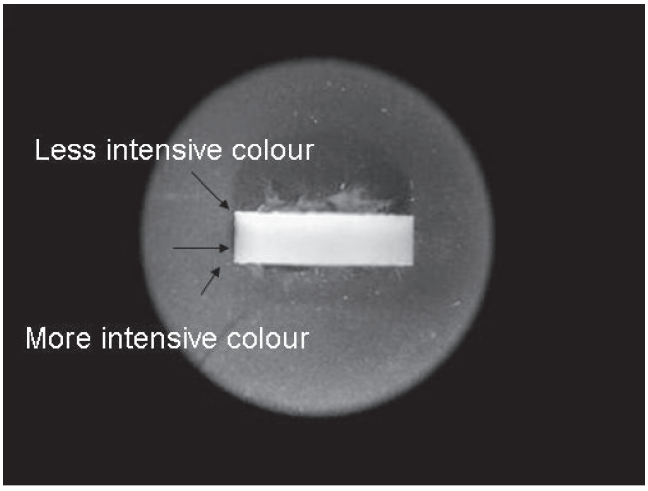


**FIGURE 1** Native dextran tablet press in eccentric machine shows variability of color due to different pressure between upper and lower punches.

Another reason, to use a rotary press rather than an eccentric press for tablet compression is the dwell time is usually shorter on a rotary press (see Figure 3).

All compression mixes have an optimum compressing speed. This is why tablet press manufacturers install variable speed controls for the rotor or turret and the powder feeding mechanism. Many compression mixes are speed sensitive and will not produce satisfactory tablets at inappropriate speeds. The dwell time, where maximum pressure is applied to the mix, is relative to the peripheral speed of the turret and the diameter of the punch head flat. Any air in the compression mix must be expelled to avoid laminating or splitting of the tablet. If air is compressed within the tablet, when the pressure applied by the punch is released, the compressed air expands and breaks the tablet. Precompression or dies with a taper in the bore will

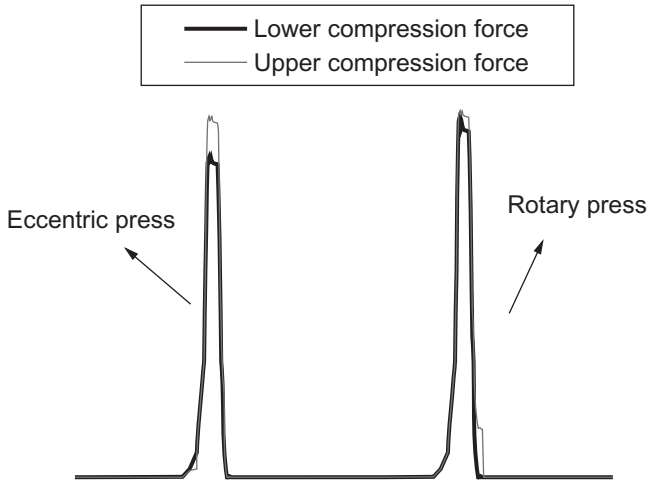


**FIGURE 2** Punch movements and compression profiles of upper and lower punches of rotary and eccentric presses.

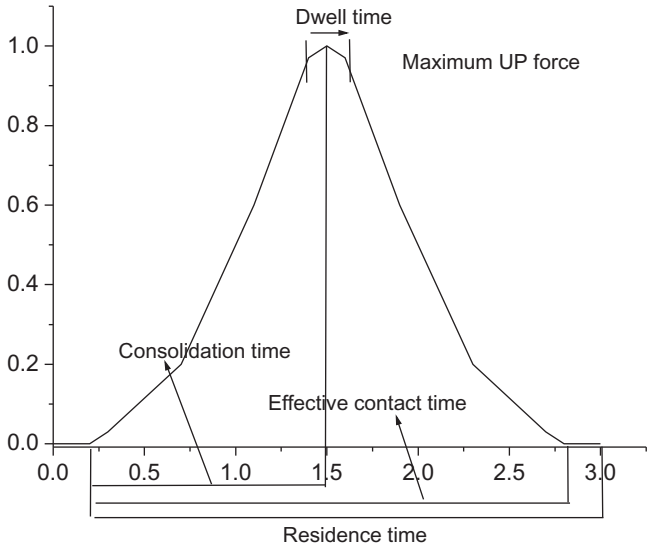


**FIGURE 3** Dwell time as function of tableting machine.

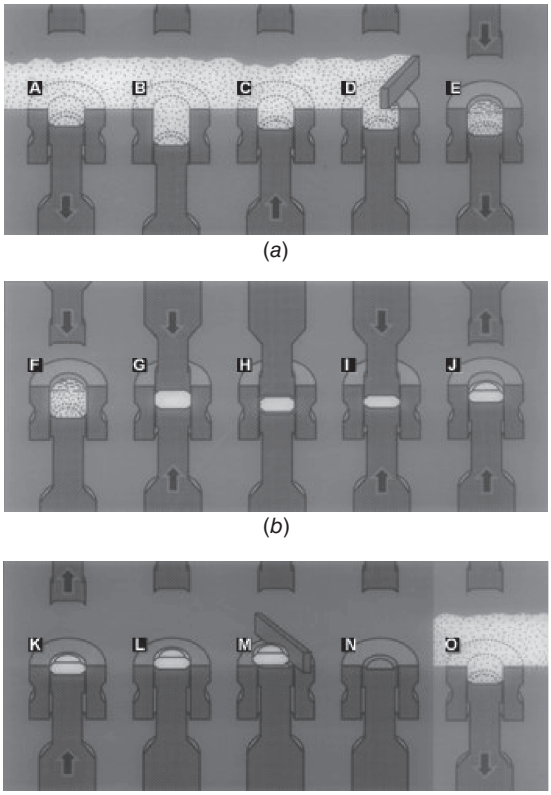


**FIGURE 4** Stages of tableting process: (*a*) filling; (*b*) compression; (*c*) ejection.

often reduce this problem. Rotary presses sometimes have two pairs of compression rolls (for precompression and compression). During precompression additional compression can take place and the absolute dwell time can be prolonged. In other words, there are numerous different tablet presses with various possibilities to carry out the compression process.

A direct correlation between the results of an eccentric press and a rotary press cannot always be drawn. In addition there are many different tablet presses with different settings and possibilities. These problems can be overcome by using a compaction simulator early in the development stage. An advantage of such a simulator is its versatility, that is, all types of presses can be simulated with small amounts of solid. The problem, however, is that such a simulator is very expensive.

The tableting cycle is well explained in the literature and is broken down into three stages (Figure 4):

1. *Filling* The volume of the granule is measured (Figure 4*a*):
   A: The lower punch is allowed to descend to its lowest point.
   B: The bore of the die is filled completely with powder.
   C: The lower punch is raised to a predetermined point so that excess.
   D: The powder is leveled by passing under a blade.

E: This ensures that the bore of the die is filled with as exact volume of the material to be used, and the next stage can begin.

2. *Compression* Pressure is applied to form the granule into a solid (Figure 4*b*):
   F: The upper punch is lowered into the bore of the die.
   G: Precompression gives the powder an initial "punch" to remove excess air.
   H: The powder is fully compressed.
   I: The correct pressure is reached.
   J: The upper punch is lifted out of the way ready for tablet ejection.
3. *Ejection* The tablet is ejected and the next tablet will be formed (Figure 4*c*):
   K: The lower punch begins to rise in the bore of the die lifting the tablet until step L is reached.
   L: Its base is level with the tap of the die.
   M: The tablet is pushed aside into the take-off chute by passing under a static blade.
   N: The lower punch moves to its lowest position ready for filling (**O**), similar to **A** and the entire cycle is repeated.

Guidance on compression levels for each tablet type (series 1, 2, and 3) and maximum punch pressures are given in Table 1:

**Series 1**: Flat flace, normal concave, shallow concave tablets
**Series 2**: Double radius, bevel and concave tablets
**Series 3**: Flat beveled edge, deep concave, ball or pill

In Figure 4 the tablets progress from start to finish from left to right. On an actual machine this will be determined by the direction in which the entire turret rotates in relation to the fixed item, such as the fill hopper precompressions and compression rollers. It is important to note that this direction may vary from machine to machine, but as a general rule, British, American, and some Asian machines rotate in a clockwise direction while European machines rotate anticlockwise (see Figure 5).

## 3 LIMITATIONS FOR DIRECT COMPRESSION

In tablet formulation, a range of excipient materials are normally required along with the active ingredient in order to give the tablet the desired properties. For example, the reproducibility and dose homogeneity of tablets are dependent on the properties of the powder mass. The tablet should also be sufficiently strong to withstand handling but should disintegrate after intake to facilitate drug release. The choice of excipients will affect all these properties:

1. *Filler* illers are used to make tablets of sufficient size for easy handling by the patient and to facilitate production. Tablets containing a very potent active

**TABLE 1 Guide Punch Pressures for Series 1, 2, and 3**

| Tablet Size | | Pressure (kN) | | |
|---|---|---|---|---|
| mm | in. | Series 1 | Series 2 | Series 3 |
| 3 | $\frac{1}{8}$ | 5 | 3 | 2 |
| 4 | $\frac{5}{32}$ | 10 | 5.6 | 3.7 |
| 5 | $\frac{3}{16}$ | 15 | 8.8 | 5.8 |
| 6 | $\frac{7}{32}$ | 22 | 12 | 8.5 |
| 7 | $\frac{9}{32}$ | 30 | 17 | 11 |
| 8 | $\frac{5}{16}$ | 40 | 22 | 15 |
| 9 | $\frac{11}{32}$ | 50 | 28 | 19 |
| 10 | $\frac{13}{32}$ | 60 | 35 | 23 |
| 11 | $\frac{7}{16}$ | 70 | 42 | 27 |
| 12 | $\frac{15}{32}$ | 90 | 50 | 33 |
| 13 | $\frac{1}{2}$ | 100 | 59 | 39 |
| 14 | $\frac{9}{16}$ | 120 | 67 | 46 |
| 15 | $\frac{19}{32}$ | 130 | 78 | 53 |
| 16 | $\frac{5}{8}$ | 160 | 90 | 60 |
| 17 | $\frac{21}{32}$ | 180 | 102 | 68 |
| 18 | $\frac{23}{32}$ | 203 | 114 | 74 |
| 19 | $\frac{3}{4}$ | 226 | 127 | 85 |
| 20 | $\frac{25}{32}$ | 251 | 141 | 94 |
| 25 | 1 | 393 | 221 | 147 |

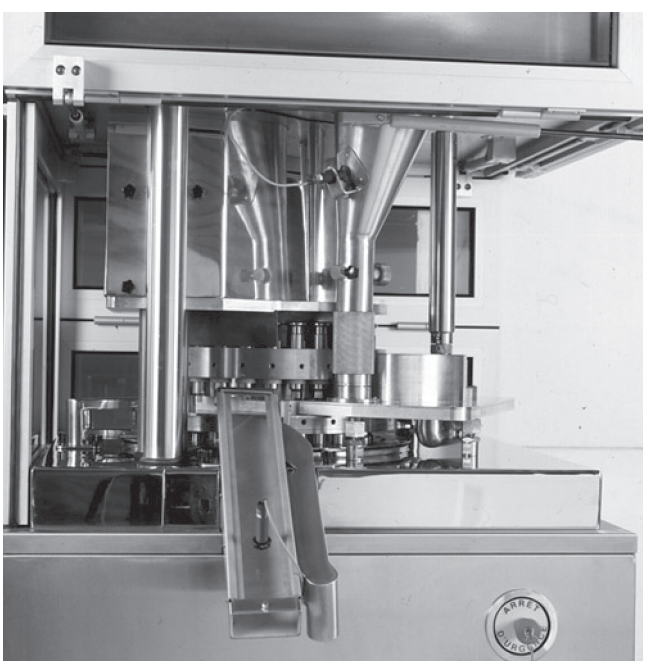


**FIGURE 5** Tableting machine (anticlockwise).

substance would be very small without excipients. A good filler will have good compactability and flow properties and acceptable taste and be non-hygroscopic and preferably chemically inert. It may also be advantageous to have a filler that fragments easily, since this counteracts the negative effects of lubricant additions to the formula [4].

2. *Binder* A material with a high bonding ability can be used as a binder to increase the mechanical strength of the tablet. A binder is usually a ductile material prone to undergo plastic (irreversible) deformation. Typically, binders are polymeric materials, often with disordered solid-state structures. Of special importance is the deformability of the peripheral parts (asperities and protrusions) of the binder particles [5].

This group of materials has the capacity of reducing interparticulate distances within the tablet, improving bond formation. If the entire bulk of the binder particles undergoes extensive plastic deformation during compression, the interparticular voids will, at least partly, be filled and tablet porosity will decrease. This increases contact area between the particles, which promotes the creation of interparticular bonds and subsequently increases tablet strength [6]. However, the effect of the binder depends on both its own properties and those of the other compounds within the tablet. A binder is often added to the granulation liquid during wet granulation to improve the cohesiveness and compactability of the powder particles, which assists the formation of agglomerates or granules. It is commonly accepted that binders added in dissolved form, during a granulation process, are more effective than those added in dry powder form during direct compression.

3. *Disintegrating Agent* A disintegrant is normally added to facilitate the rupture of bonds and subsequent disintegration of the tablets. This increases the surface area of the drug exposed to the gastrointestinal fluid; incomplete disintegration can result in incomplete absorption or a delay in the onset of action of the drug. There are several types of disintegrants, acting with different mechanisms: (a) promotion of the uptake of aqueous liquids by capillary forces, (b) swelling in contact with water, (c) release of gases when in contact with water, and (d) destruction of the binder by enzymatic action [7]. Starch is a traditional disintegrant; the concentration of starch in a conventional tablet formulation is normally up to 10% w/w. The starch particles swell moderately in contact with water, and the tablet disrupts. So-called superdisintegrants are now commonly used; since these act primarily by extensive swelling, they are effective only in small quantities [8] cross-linked sodium carboxymethyl cellulose (e.g., Ac-Di-Sol), which is effective in concentrations of 2–4%, is a commonly used superdisintegrant. Larger particles of disintegrants have been found to swell to a greater extent and with a faster rate than finer particles, resulting in more effective disintegration [9].

4. *Glidant, Antiadherent, and Lubricant* Glidants are added to increase the flowability of the powder mass, reduce interparticulate friction, and improve powder flow in the hopper shoe and die of the tableting machine. An antiadherent can be added to decrease sticking of the powder to the faces of the punches and the die walls during compaction, and a lubricant is added to decrease friction between powder and die, facilitating ejection of the tablet from the die. However, addition of lubricants (here used as a collective term and including glidants and antiadherents) can have negative effects on tablet strength, since the lubricant often reduces

the creation of interparticulate bonds [e.g., 4]. Further, lubricants can also slow the drug dissolution process by introducing hydrophobic films around drug and excipient particles [e.g., 10]. These negative effects are especially significant when long mixing times are required [11]. Therefore, the amount of lubricants should be kept relatively low and the mixing procedure kept short to avoid a homogenous distribution of lubricant throughout the powder mass. An alternative approach could then be to admix granulated qualities of lubricant [12].

5. *Flavor, Sweetener, and Colorant* Flavor and sweeteners are primarily used to improve or mask the taste of the drug, with subsequent substantial improvement in patient compliance. Coloring tablets also has aesthetic value and can improve tablet identification, especially when patients are taking a number of different tablets.

General instructions for the determination of tablet properties (e.g., hardness, disintegration, friability, dissolution profile, and stability) are contained in pharmacopeia [e.g., European Pharmacopoeia (Eur. Ph.) and U.S. Pharmacopeia (USP)].

In the manufacture of tablets it is important to define and appreciate the physical properties of the active substance, in particular particle size and flowability. The technology involved in direct compression assumes great importance in tablet formulations because it is often the least expensive, particularly in the production of generics that the active substance permits. The limiting factors are the physical properties of the active substance and its concentration in the tablets. Even substances such as ascorbic acid, which are not generally suitable for direct compression owing to the friability of the crystals, can normally be directly pressed into tablets at concentrations of 30–40%. However, this technique is not as suitable if the content of ascorbic acid is higher. This limit may be shifted upward by special direct-compression auxiliaries, for example, Ludipress (BASF).

Ludipress is derived from lactose, Kollidon 30, and Kollidon CL. It thus combines the properties of a filler, binder, disintegrant, and flow agent and also often acts as a release accelerator. By virtue of its versatility, formulations containing it are usually very simple. It can also be combined with almost all active substances with the exception of those that enter into a chemical interaction with lactose (Maillard reaction). Active substances (e.g., many analgetics) behave very differently with Ludipress when the dosage is extremely high. Acetylsalicylic acid (ASA) and metamizole can be pressed when little Ludipress has been added; ibuprofen requires a larger amount; and the fraction of Ludipress required in the tablets is too large for paracetamol (acetaminophen).

An alternative to the Ludipress grades is the outstanding dry binder Kollidon VA 64 together with excipients (e.g., calcium phosphate, microcrystalline cellulose, lactose, or starch) and a disintegrant (e.g., Kollidon CL). This combination even allows 500 mg of paracetamol to be pressed into good tablets with a weight of 700 mg.

No other dry binder has binding power and plasticity comparable to Kollidon VA 64. Plasticity, in particular, is an important parameter in direct compression. As can be seen in Figure 6 (99.5% binder + 0.5% magnesium stearate), this property of Kollidon VA 64 is not adversely affected by increasing the pressure. The beneficial

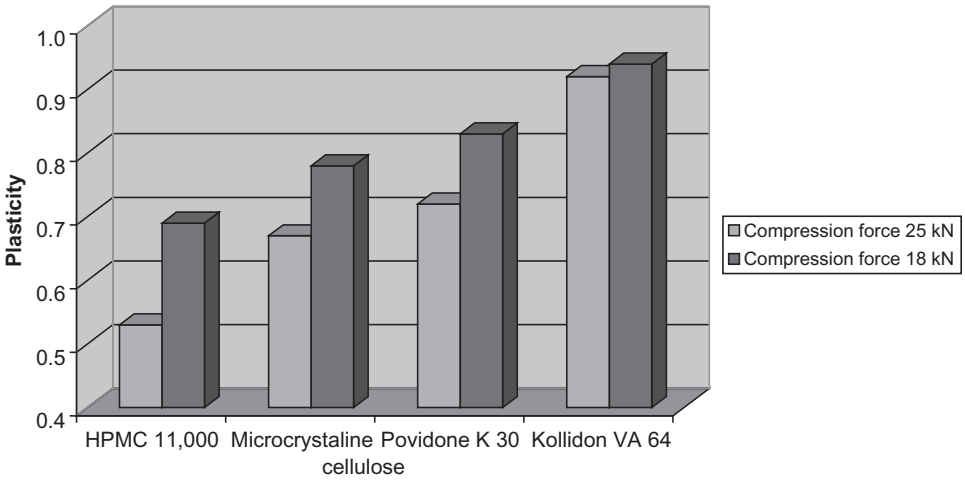


**FIGURE 6** Plasticity of dry binders in tablets.

properties of Kollidon VA 64 can also be exploited for the production of concentrated active substance that is subsequently used for direct tableting. Kollidon VA 64 and Ludipress can also be combined with one another.

### *Acetylsalicylic Acid–Acetaminophen–Caffeine Tablet (250 mg + 250 mg + 50 mg)*

*Formulation*

| | |
|---|---|
| Acetaminophen powder | 39.25% |
| Caffeine powder | 7.85% |
| ASA powder | 39.25% |
| Kollidon VA 64 | 9.4% |
| Kollidon CL | 3.15% |
| Aerosil 200 | 0.5% |
| Magnesium stearate | 0.6% |

*Manufacture by Roller Compaction* The active ingredients and Kollidon VA 64 were granulated in a Gerteis Roller compactor.

*Process Parameters*

| | |
|---|---|
| Force | 5.0 kN/cm |
| Crack | 2.9 mm |
| Looser | 5.0 rpm |
| Dosing screw | 90.5 rpm |
| Plug screw | 108.6 rpm |
| Granulator | 50 rpm |

*Tableting* The granules were passed together with magnesium stearate, Aerosil 200, and Kollidon CL through an 800-μm sieve, blended for 10 min in a Turbula

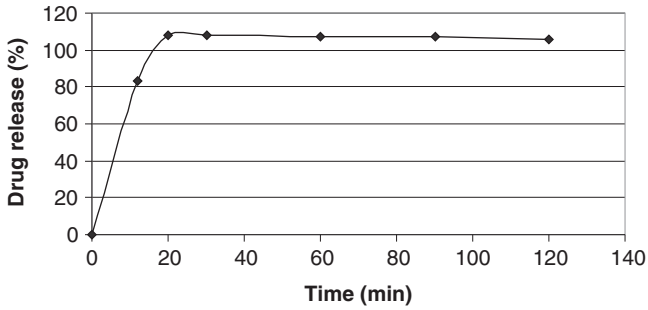


**FIGURE 7** Dissolution profile for acetylsalicylic acid–acetaminophen–caffeine tablet (250 mg + 250 mg + 50 mg): paddle 50 rpm, 37 °C deionized water.

mixer, and compressed into tablets with a force of about 12 kN (tablet press Korsch PH106, 30 rpm, compression force 11.5 kN)

*Tablet Properties*

| | |
|---|---|
| Weight | 629.7 mg |
| S rel | 1.4% |
| Diameter | 12 mm |
| Form | Biplanar |
| Hardness | 72 N |
| Friability | 2.75% |

*Dissolution* See Figure 7.

**Enteric Film Coating of Tablets** (Organic Solution)

*Formulation of Cores*

| Component | Percent/Tablet (w/w) |
|---|---|
| ASA | 33.4 |
| Ludipress | 49.5 |
| Avicel PH 102 | 16.6 |
| Magnesium stearate | 0.5 |
| Total | 100.0 |

*Manufacture by Direct Compression* All the components were mixed for 10 min, passed through an 0.8-mm sieve, and compressed into tablets on a rotary press with a rate of 40,000 tablets/h at a compression force of 15-kN. Core shape was convex with a diameter of 9 mm and the engraving BASF. Hardness of the tablets was about 60 N.

*Formulations of Coating Suspension* The formulation is for 5-kg cores (diameter 9 mm; 300 mg):

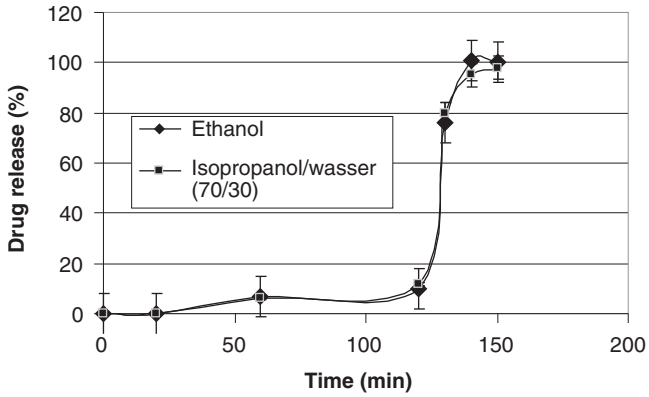


**FIGURE 8** Enteric film coating of tablets, acetylsalicylic acid (organic solution): paddle 100 rpm, 37 °C; 0–2 h: 0.08 *M* HCl, 2 h+: phosphate buffer Ph 6.8.

| | Parts by Weight [g] | |
|---|---|---|
| Component | No. 1 | No. 2 |
| Kollicoat MAE 100 P | 344.52 | 344.52 |
| Triethyl citrate | 34.45 | 34.45 |
| Ethanol | 4214.60 | — |
| Isopropanol/Wasser (70/30) | — | 4214.60 |
| Total | 4593.57 | 4593.57 |
| Polymer applied | 10.0 mg/cm$^2$ | |
| Content of polymer | 7.5% | |

*Preparation of Spray Suspension* Kollicoat MAE 100P and triethyl citrate were stirred into the solvent until complete dissolution.

*Process Parameters*

| | |
|---|---|
| Coating pan | 24 Accela Cota |
| Size of batch | 5 kg |
| Inlet air temperature | 50°C |
| Outlet air temperature | 35–38°C |
| Product temperature | 30–35°C |
| Inlet air rate | 70 m |
| Outlet air rate | 140 m |
| Spraying pressure | 2.0 bars |
| Nozzle diameter | 1.0 mm |
| Rate of spraying | 30 g/min |
| Time of spraying | 2.5 h |
| Preheating | 3 min |
| Final drying | 5 min |

*Dissolution:* (See Figure 8). Dissolution was done according to USP monograph "Aspirin Delayed-Release Tablets."

### *Beta Carotene Tablets (15 mg)*

*Formulation*

| | Formulation 1 | Formulation 2 |
|---|---|---|
| Beta carotene dry powder 10% | 160.0 g | 150.0 g |
| Ludipress | 240.0 g | — |
| Dicalcium phosphate, granulated with 5% Kollidon 30 | — | 175.0 g |
| Avicel PH 101 | — | 100.0 g |
| Kollidon CL | 6.0 g | 5.0 g |
| Aerosil 200 | — | 2.5 g |
| Talc | — | 20.0 g |
| Calcium arachinate | — | 2.5 g |
| Magnesium stearate | 2.0 g | — |

*Manufacturing (Direct Compression)* All components were mixed, passed through a 0.8-mm sieve, and pressed with a medium compression force.

*Tablet Properties*

| | Formulation 1 | Formulation 2 |
|---|---|---|
| Weight | 400 mg | 502 mg |
| Diameter | 12 mm | 12 mm |
| Form | Biplanar | Biplanar |
| Hardness | 59 N | 57 N |
| Disintegration | 12 min | 1 min |
| Friability | 0.1% | 0% |

*Chemical and Physical Stability (20–25°C)*

| | 6 Months | 12 Months |
|---|---|---|
| Formulation 1 | | |
| Loss of beta carotene | 3% | 4% |
| Hardness | 60 N | 59 N |
| Disintegration | 9 min | 7 min |
| Friability | 0.15% | 0.16% |
| Formulation 2 | | |
| Loss of beta carotene | 8% | 9% |

### *Diclofenac Na–Dispersion–Tablet (50 mg)*

*Formulation*

| | |
|---|---|
| Diclofenac Na | 50.0 mg |
| Avicel PH 102 | 143.8 mg |
| Kollidon CL | 50.0 mg |
| Aerosil 200 | 5.0 mg |
| Magnesium stearate | 1.0 mg |

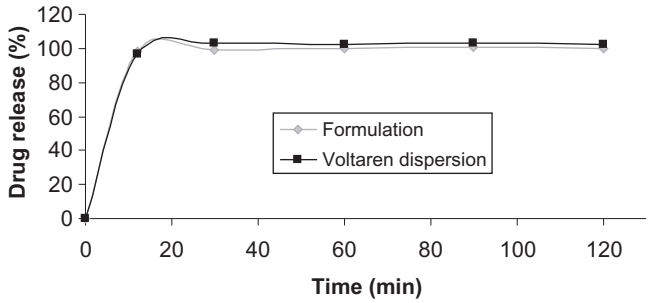


**FIGURE 9** Dissolution profile for diclofenac–Na dispersion tablet (50 mg): paddle 50 rpm, 37 °C phosphate buffer, pH 7.2.

*Procedure* The ingredients were mixed, passed through a 0.8-mm sieve, and compressed into tablets with a force of about 10 kN. (The tablet press was Korsch PH106, 30 rpm, compression force was 11.8 kN.

*Tablet Properties*

| | |
|---|---|
| Weight | 248.0 mg |
| S rel | 1.7% |
| Diameter | 10 mm |
| Form | bipla Nar |
| Hard Ness | 93 N |
| Friability | <0.1% |

*Dissolution* See Figure 9.

In direct–compression formulation, there is a wide particle size distribution. Usually, the active drug is at the fine end of the range. Such a wide particle size range can easily result in significant segregation. Five primary mechanisms are responsible for most particle segregation problems [13]. Of these, only three typically occur with pharmaceutical powders: sifting, entrainment of air, and entrainment of particles in an air stream.

Sifting is a process by which smaller particles move through a matrix of larger ones. It is by far the most common method of segregation. Sifting has been found to occur with particle size ratios as low as 1.3:1 or with a sufficiently large mean particle size (the tendency to segregate by sifting decreases substantially with particle size <500 μm. Free-flowing material and interparticle motion also caused segregation by sifting.

Two techniques can be used to decrease a material's segregation tendencies: change the material or change the design of the equipment.

***Lisinopril (5 mg)*** Reducing the ratio of excipient, lisinopril (2:1) tablets for direct compression can be obtained:

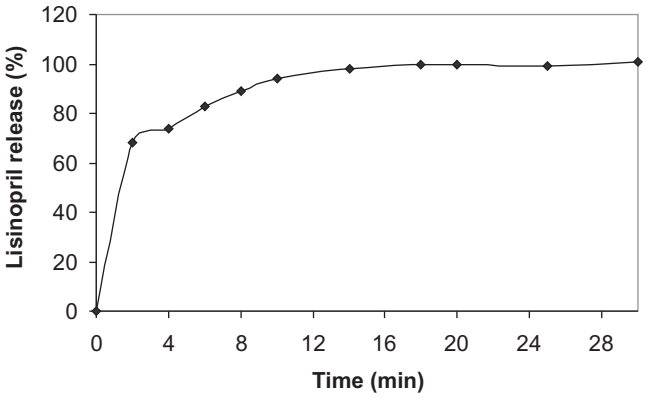


**FIGURE 10** Disolution profile for lisinopril 5 mg: paddle 50 rpm, 37 °C, HCl 0.1 *N*.

| Formulation | % |
|---|---|
| Lisinopril | 5.0 |
| Aerosil | 0.5 |
| Calcium phosphate dibasic | 19.8 |
| Starch 1500 | 2.0 |
| Magnesium stearate | 0.7 |
| Cellulose microcrystalline (PH-250) | 72.0 |

Lisinopril and Aerosil (<150 μm) were mixed for around 10 min. All other components were added and mixed for 15 min, passed through a 0.150-mm sieve, and pressed with 10 kN compression force.

Properties of 5-mg lisinopril tablets are as follows:

| | |
|---|---|
| Weight | 130 mg |
| Diameter | 8 mm |
| Form | Biplanar |
| Hardness | 98 N |
| Disintegration | 2–3 min |
| Friability | 0.05 |

The content uniformity of the formulation was measured at the beginning, middle, and end of the batch (50 kg) (Table 2).

Special attention should be given to the physical stability of the tablets manufactured by direct compression because some fillers/binders are known to soften or harden on storage.

## 4 PREVIOUS GRANULATION: BIOPHARMACEUTICAL VERSUS TECHNOLOGICAL PROPERTIES

Granulation is the process by which primary powder particles are made to adhere to form larger, multiparticle entities called granules. Pharmaceutical granules

**TABLE 2 Study of Uniformity for Formulation Lisinopril 5 mg (Batch 50 kg)**

| Number | 0–5 kg | 5–10 kg | 10–15 kg | 15–20 kg | 20–25 kg | 25–30 kg | 35–40 kg | 40–45 kg | 45–50 kg |
|---|---|---|---|---|---|---|---|---|---|
| **1** | 4.99 | 5.00 | 5.00 | 4.95 | 5.02 | 5.08 | 5.00 | 4.96 | 5.00 |
| **2** | 5.00 | 5.06 | 5.00 | 4.92 | 5.00 | 4.99 | 4.83 | 5.11 | 5.17 |
| **3** | 5.00 | 4.81 | 4.90 | 5.19 | 4.92 | 5.03 | 4.92 | 5.00 | 5.03 |
| **4** | 5.05 | 5.00 | 5.00 | 4.99 | 4.82 | 4.96 | 5.00 | 4.93 | 4.99 |
| **5** | 4.89 | 4.99 | 4.90 | 5.05 | 5.13 | 5.04 | 4.98 | 4.93 | 4.87 |
| **6** | 5.02 | 5.06 | 4.85 | 5.10 | 5.13 | 5.00 | 5.18 | 4.81 | 4.93 |
| **7** | 5.00 | 5.08 | 5.11 | 5.00 | 5.00 | 4.83 | 4.89 | 5.00 | 5.00 |
| **8** | 5.00 | 5.00 | 5.03 | 5.03 | 5.00 | 5.06 | 5.00 | 5.18 | 5.00 |
| **9** | 4.97 | 4.99 | 5.01 | 5.04 | 5.05 | 4.85 | 5.00 | 5.00 | 4.99 |
| **10** | 4.91 | 5.00 | 5.01 | 4.98 | 4.97 | 5.00 | 5.00 | 5.11 | 5.05 |

typically have a size range between 0.2 and 4.0 mm, depending on their subsequent use. In the majority of cases this will be in the production of tablets or capsules, when granules will be made as intermediate products and have a typical size range between 0.2 and 0.5 mm, but larger granules are used as a dosage form in their own right.

Granulation normally commences after initial dry mixing of the necessary powdered ingredients so that a uniform distribution of each ingredient through the mix is achieved. After granulation the granules either will be packed or may be mixed with other excipients prior to tablet compaction.

The principal reasons for granulation are as follows:

1. To prevent segregation of the constituents of the powder mix.
2. To improve the flow properties of the mix.
3. To improve the compaction characteristics of the mix

*Methods of Granulation* Granulation methods can be divided into two types: *wet methods*, which use a liquid in the process, and *dry methods*, in which no liquid is used. In a suitable formulation a number of different excipients will be needed in addition to the drug. The common types used are diluents, to produce a unit dose weight of suitable size, and disintegrating agents, which are added to aid the break-up of the granule when it reaches a liquid medium (e.g., on ingestion by the patient). Adhesives in the form of a dry powder may also be added, particularly if dry granulation is employed. These ingredients will be mixed before granulation.

1. *Dry Granulation* In the dry methods of granulation the primary powder particles are aggregated under high pressure. There are two main processes: Either a large tablet (*slug*) is produced in a heavy-duty tableting press (*slugging*) or the powder is squeezed between two rollers to produce a sheet of material (*roller compaction*). In both cases these intermediate products are broken down using a suitable milling technique to produce granular material, which is usually sieved to separate the desired size fraction. This dry method may be used for drugs

that do not compress well after wet granulation or those which are sensitive to moisture.

2. *Wet Granulation* Wet granulation involves the massing of a mix of dry *primary powder particles* using a *granulating fluid*. The fluid contains a solvent which must be volatile so that it can be removed by drying and be nontoxic. Typical liquids include water, ethanol, and isopropanol, either alone or in combination. The granulation liquid may be used alone or, more usually, as a solvent containing a dissolved *adhesive* (also referred to as a *binder* or *binding agent*) which is used to ensure particle adhesion once the granule is dry. Water is commonly used for economical and ecological reasons. Its disadvantages as a solvent are that it may adversely affect drug stability, causing hydrolysis of susceptible products, and it needs a longer drying time than do organic solvents. This increases the length of the process and again may affect stability because of the extended exposure to heat.

***Captopril (25 mg) + Hydrochlorothiazide (25 mg)***

*Formulation*

| | Formulation for 500 mg |
|---|---|
| Captopril | 5% |
| Hydrochlorothiazide | 5% |
| Lactose | 65% |
| Carboxyethylcellulose sodium | 10% |
| Ac-Di-Sol | 3% |
| Starch | 10% |
| Stearic acid | 2% |

*Manufacturing (Wet Granulation)* A mixture of all compounds (with 1.5% stearic acid) is granulated with solution 2-propanol (around 8% v/w), passed through a 0.8-mm sieve, and the rest (0.5% stearic acid) added and pressed with low compression force.

*Tablet Properties (Initial Time)*

| | |
|---|---|
| Weight | 500 mg |
| Diameter | 12 mm |
| Form | Normal concave |
| Hardness | 60 N |
| Disintegration | <4 min |
| Friability | <0.3% |
| Dissolution (captopril + hydrochlorothiazide) | |
| 30 min | 90.00% |
| 60 min | 100% |

*Stability of Three Batches (5 kg each) at 25 °C and 70% Relative Humidity (RH) during 12 Months*

Formulation Table 3 presents a comparison of lactose monohydrate and calcium phosphate.

**TABLE 3 Comparative Study of Lactose Monohydrate and Calcium Phosphate, Dibasic**

| | Formulation F1 | Formulation F2 |
|---|---|---|
| α-Methyldopa | 275 g (78%) | 275 g (78%) |
| Lactose monohydrate | 15.5% | — |
| Calcium phosphate, dibasic | — | 15.5% |
| Kollidon 30 | 4% | 4% |
| Isopropanol | 80 mL | 80 mL |
| Kollidon CL | 2% | 2% |
| Magnesium stearate | 0.5% | 0.5% |

| | 6 months | | | 12 months | | |
|---|---|---|---|---|---|---|
| Formulation | Batch 1 | Batch 2 | Batch 3 | Batch 1 | Batch 2 | Batch 3 |
| Hydrochlorothiazide (%) | 98.91 | 99.47 | 100.65 | 99.06 | 100.40 | 102.88 |
| Captopril (assay) (%) | 101.02 | 100.89 | 100.99 | 100.65 | 100.03 | 100.30 |
| Captopril disulfuric[a] | 0.62 | 0.71 | 0.69 | 0.69 | 0.79 | 0.72 |
| Weight | 500.12 | 501.23 | 499.65 | 499.00 | 500.33 | 500.14 |
| Hardness | 63 N | 67 N | 61 N | 60 N | 63 N | 61 N |
| Disintegration (min) | 3 | 3 | 4 | 4 | 4 | 4 |
| Friability | 1.98 | 1.15 | 1.12 | 1.71 | 1.12 | 1.12 |
| Dissolution | | | | | | |
| 30 min | 89.45 | 90.54 | 88.77 | 88.00 | 91.00 | 88.05 |
| 60 min | 100 | 100 | 100 | 100 | 100 | 100 |

[a]Captopril degradation product (%).

### *α-Methyldopa Tablet (250 mg)*

*Manufacturing (Wet Granulation)* A mixture of α-methyldopa with lactose or calcium phosphate (for formulations **F1** or **F2**, respectively) is granulated with isopropanol solution of Kollidon 30 and passed through a sieve, the dry granules are mixed with Kollidon CL and magnesium stearate, and pressed with medium compression force.

*Tablet Properties*

| | F1 | F2 |
|---|---|---|
| Weight | 361 mg | 362 mg |
| Diameter | 11 mm | 11 mm |
| Hardness | 118 N | 156 N |
| Disintegration | 5 min | 4 min |
| Friability | <0.1% | <0.1% |
| Dissolution | | |
| 10 min | 45% | 55% |
| 20 min | 82% | 90% |
| 30 min | 90% | 98% |

Calcium phosphate, dibasic offers high hardness and faster dissolution profile than lactose for α-methyldopa tablets in wet granulation.

## 5 TABLET DESIGN FOR MATRIX SYSTEM

The advantages of controlled-release systems include maintenance of drug levels within a desired range, the need for fewer administrations, optimal use of the drug in question, and increased patient compliance. Evaluation of matrix tablets is the same as for conventional formulations but the dissolution profile and stability have to be carefully studied. Numerous methods for development of matrix tablets can be used, such as direct compression, wet granulation, pelletization, and spheronization exclusion. Nevertheless, the potential disadvantages cannot be ignored: the possible toxicity or no biocompatibility of the materials used, undesirable by-products of degradation, any surgery required to implant or remove the system, the chance of patient discomfort from the delivery device, and the higher cost of controlled-release systems compared to traditional pharmaceutical formulations. The importance of matrix systems that they release bioactive component over an extended period of time has long been recognized in the pharmaceutical field. Matrix systems can be divided into three groups depending on the type of polymer formed:

1. *Inert Matrices* Polymers that after compression form an indigestible and insoluble porous skeleton [14] constitute the inert matrices. The main challenge in the preparation of these systems is to achieve, by means of a suitable design, total drug release from the device as well as adequate and precise drug release, guaranteeing the integrity of the matrix.

2. *Hydrophilic Matrices* Cellulose derivatives have been widely used in the formulation of hydrogel matrices for controlled drug delivery. Among them hydroxypropyl methylcellulose (HPMC) is the most extensively employed because of its ease of use, availability, and very low toxicity [15]. Drug release from these systems is controlled by the hydration of HPMC, which forms a gelatinous layer at the surface of the matrix through which the included drug diffuses.

Drug release from swellable matrix tablets is based on the glassy–rubbery transition of the polymer which occurs as a result of water penetration into the matrix. Therefore, the gel layer is physically delimited by the erosion (swollen matrix–solvent boundary) and swelling (glassy–rubbery polymer boundary) fronts.

Water-soluble drugs are released primarily by diffusion of dissolved drug molecules across the gel layer, while poorly water soluble drugs are released predominantly by erosion mechanisms.

The factors influencing the release of drugs from hydrophilic matrices include viscosity of the polymer, ratio of the polymer to drug, mixtures of polymers, compression pressure, thickness of the tablet, particle size, pH of the matrix, entrapped air in the tablet, solubility of the drug, the presence of excipients or additives, and the mode of incorporation of these substances.

3. *Lipid Matrices* These matrix tablets are formed with lipid polymers with low melting point. The drug is dissolved or solubilized in the melted lipid, such as cetyl

**TABLE 4 Theophylline Formulation by Direct Compression and Wet Granulation (mg/tablet)**

| | Direct Compression | Wet Granulation |
|---|---|---|
| Granulated theophylline | 264 | 264 |
| 90SH-4000SR HPMC (Metolose SR) | 64.5 | 64.5 |
| Mg stearate | 1.5 | 1.5 |
| Total | 330 | 330 |

alcohol, ceto-stearilic alcohol, and stearic acid. Solid lipid nanoparticles are an example of an innovative lipid matrix system.

### 5.1 Controlled-Release Tablet by Direct Compression and Wet Granulation

Theophylline is granulated in a fluid bed with Pharmacoat 606 3% (Shin Etsu, Metolose SR) as shown in Figure 11. Table 4 gives a comparison of direct compression (DC) and wet granulation (WG) using theophylline:

Diret compression: using a twin-cell mixer, theophylline and HPMC are mixed for 10 min; then Mg stearate is added and mixed for 2 min.

Tableting conditions for DC and WG: A rotary tableting machine (KIKUSUI) is used with 12 punches (punch size 10 mm diameter, 12 mm radius, compression force 98, 147, and 196 MPa; tableting speed 20, 40, and 60 $min^{-1}$).

Condition for WG: Granulation machine, vertical granulator FM-VG-05; charge 300 g; binder solution ethanol–water 8:2; agitation 600 (blade)/1000 (chopper) $min^{-1}$; granulation time 5 min.

Powder properties for compression: Bulk density 0.35 g/mL, tapped density 0.48, average particle size 122 μm.

Theophylline tablets made by both direct compression and wet granulation have been assessed. There is almost no difference between direct compression and wet granulation methods (see Figures 12–14) under the following conditions: appropriate formulation (sufficient level of HPMC in the tablet) and precise control of the wet granulation process. Direct compression using Metolose SR is recognized as a suitable process for matrix tablets.

## 6 TABLET DESIGN WITH NATURAL PRODUCTS

The development and production of tablets containing a high dose of active ingredients is a complex and extensive technological challenge. Dried plant extracts are often used as therapeutically active material in the manufacture of tablets. They are

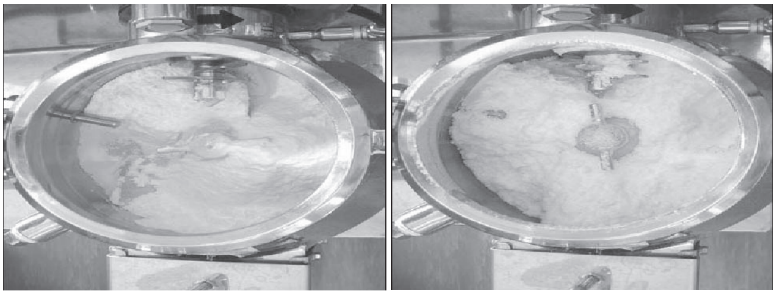

(*a*) (*b*)

**FIGURE 11** Granulation process: (*a*) before 90SH-4000SR; (*b*) after 90SH-4000SR.

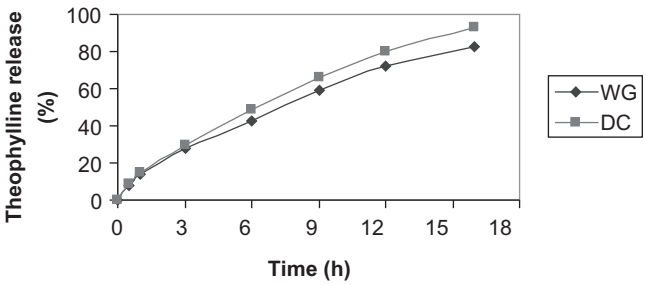


**FIGURE 12** Dissolution profile for theophylline tablets (DC and WG).

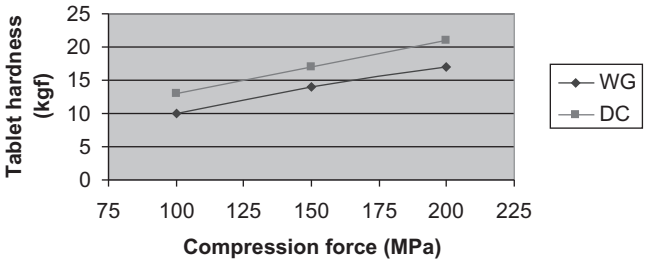


**FIGURE 13** Hardness of theophylline tablets (Dc and WG).

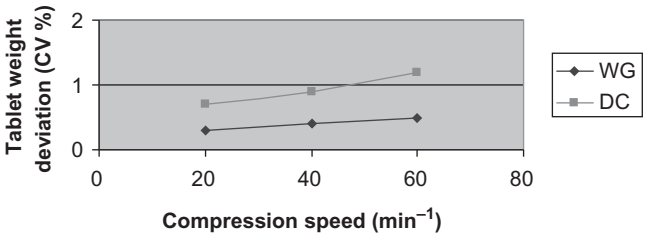


**FIGURE 14** Weight deviation for theophylline tablets (DC and WG).

often very fine, poorly compressible, and very hygroscopic powders. Tablets containing a high amount of spray-dried extract show prolonged disintegration times, affecting the release of active constituents [19]. Some alternatives have been proposed to minimize these problems. Granulation is the technique most often used to improve the technological properties of these products. However, because of the products' high hygroscopicity, extracts sometimes cannot be granulated using aqueous systems. Some reports have shown that the use of lubricants during direct compression of vegetable dried extracts increased the disintegration time. According to our experience and some previous work, tablets with high amounts of some lubricant such as aerosil (up to 25% w/w) and magnesium [16] stearate incorporated into the granules had shorter disintegration time than did tablets containing the powdered mixture.

Natural products can be used as plant extracts with pharmacological activity (e.g., *Mangifera indica* L., vallerian, aloe, *Cratoxylum pruniflorum*, microporous zeolite), excipient for direct compression or granulation (chitin, chitosan, and dextran), and controlled-release systems (cellulose and native dextran). For natural products the most important factor is the standardization of the extract because properties such as the amount of active substance can be changed from batch to batch. Factors such as the origin of the extract, geographic zone, and age of the tree could affect the properties of natural extracts.

### 6.1 Tablet Design from Aqueous Plant Extract

A bioactive product of natural origin has been developed from folk knowledge of Asian, Latin-American, European, and U.S. ethnic medicine. We developed an extract of the *M. índica* L. (mango) stem bark, obtained by decoction of some varieties grown in tropical and subtropical climates, that is used at present as an antioxidant nutritional supplement (Vimang). The aqueous extract was dried by atomization in a spray dryer until a brown solid with 10.5% (RSD = 0.9%) water content (measured by Karl Fischer) was achieved. Tablets obtained by wet granulation (plant extract, 300 mg/unit) were used for the applications. The product is a fine brown powder that has provend to be useful in the treatment of a large population sample presenting physical stress due to age or deteriorated physiological conditions caused by chronic diseases such as cancer, diabetes, or cardiovascular disorders [17]. Recent studies have shown that treatment with the extract provided significant protection against 12-*O*-tetradecanoylphorbol-13-acetate (TPA)–induced oxidative damage and better protection when compared with other antioxidants (Vitamin C, E and beta–carotene) [18]. Furthermore, the results indicate that this extract is bioavailable for some vital target organs, including liver and brain tissues, peritoneal cell exudates, and serum. Therefore, it was concluded that it could be useful to prevent the production of reactive oxygen species (ROS) and oxidative tissue damage in vivo. All these effects are likely due to the synergic action of several compounds, such as polyphenols, terpenoids, steroids, fatty acids, and microelements, which have been reported to be present in the extract [17].

**TABLE 5 Content of Mangiferin (μg/mg) in Natural Product Samples from *Mangifera indica* L. and assay Result of Pharmaceuticals**

| Sample | Content, μg/mg (RSD, %) | Sample | Content, μg/mg (RSD, %) |
|---|---|---|---|
| No.1 (batch 901) | 254 (0.7) | No.9 (batch 0201) | 125 (0.1) |
| No.2 (batch 903) | 195 (2.0) | No.10 (batch 0202) | 109 (1.5) |
| No.3 (batch E-1923) | 187 (1.6) | No.11 (batch 0203) | 116 (1.2) |
| No.4 (batch E-1924) | 180 (1.4) | No.12 (batch 0204) | 79 (2.4)* |
| No.5 (batch E-2032) | 206 (1.6) | No.13 (batch 0205) | 56 (0.5)* |
| No.6 (batch 0103) | 149 (5.7) | No.14 (batch 0206) | 55 (2.3)* |
| No.7 (batch 0104) | 162 (0.4) | No.15 (batch 0207) | 66 (1.8)* |
| No.8 (batch 0112) | 159 (0.3) | No.16 (batch 0208) | 49 (0.2)* |

| Pharmaceuticals from batch No.8 | Amount of Vimang® (mg) | Percentage of claimed content %, (RSD, %) |
|---|---|---|
| Tablets (batch A) | 299.91 | 99.97 (5.47) |
| Tablets (batch B) | 291.36 | 97.12 (3.14) |
| Tablets (batch C) | 310.51 | 103.51 (1.42) |

*RSD = relative standard deviation.

Mangiferin (1,3,6,7-tetrahydroxyxanthone-2-*C*-β-D-glucopyranoside), a *C*-glucosylxanthone, which was first isolated from the bark, branches, and leaves of *M. indica* L., has been found to be the major component of this extract. Mangiferin is a naturally occurring chemopreventive agent in rat colon carcinogenesis [19]; exerts antidiabetic activity by increasing insulin sensitivity; shows significant inhibitory effect on bone resorption; appears to act as a potential biological response modifier with antitumor, immunomodulatory, and anti–human immunodeficiency virus (HIV) effect; is capable of providing cellular protection as an antioxidant and a radical scavenger agent; is useful as an analgesic without adverse effects; and inhibits the late event in herpes simplex virus-2 replication [20–24].

The quality control of 16 batches of Vimang active ingredient [by high-performance liquid chromatography (HPLC) and the ultraviolet (UV) method] obtained from different regions of the country (batches 1–8 from the west and batches 9–16 from the east) and its pharmaceuticals (optimum formula) were investigated and are demonstrated in Table 5. Each sample was analyzed in triplicate and the average values are listed. All assay results fell between 100 and 300 μg of mangiferin per milligram of Vimang powder, except samples 12–16, which were rejected. The differences found are probably due to the fact that the mangiferin content in the plant varies with the season of the year and the zone where it was grown. The claimed contents of this natural product required by our producers are 85–115% for tablets.

Different tablet formulations were tested, but even superdisintegrants such as Ac-Di-Sol and CMCNa up to concentrations of 5% were not sufficient to disintegrate the tablets. With the use of pH-modified product such as $NaCO_2$ or canalling as NaCl the release of extract from the tablets changed dramatically. Use of lactose

**TABLE 6 Formulations by Wet Granulation**

| | F15 | F16 | F17 | F18 | F19 | F20 | F21 | F22 | F23 | F24 | F25 | F26 | F27 | F28 | F29 | F30 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Extract[a] | 50 | 50 | 50 | 50 | 50 | 50 | 50 | 50 | 50 | 50 | 50 | 50 | 50 | 50 | 50 | 50 |
| Lactose | — | — | 20[a] | 16[a] | 16[a] | 22[a] | 27[a] | 29[a] | 32[a] | 42[a] | 32[a] + 10 | 35[a] | 35[a] | 35[a] | 35[a] | 30[a] |
| CMM | 20[a] | 20[a] | | | | | | | | | | | | | | |
| NaCl | 20[a] | 15[a] + 5 | 10[a] | 5[a] | 5[a] | 5 | 5 | 5 | 5 | — | — | — | 5 | | | |
| PEG6000 | — | — | — | — | — | — | — | — | — | — | — | — | — | 5 | 5 | 5 |
| PEG400[a] | — | — | — | — | — | — | — | — | — | — | — | — | — | | 4 | |
| $Na_2CO_3$ | — | — | — | 5[a] + 5 | 10[a] | 5 | 5 | 5 | 5 | — | — | 5 | — | — | — | 5 |
| CMCNa | — | — | 5 | | | | | | | | | | | | | |
| PVP[a] | 5 | 5 | 5 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | | 4 |
| SDS | 2[a] | 1[a] + 1 | 1[a] + 1 | 1 | 1 | 2 | 2 | 2 | 2 | | | 2 | 2 | 2 | 2 | 2 |
| Talc | 1 | 1 | — | 1 | 1 | 2 | 2 | 2 | 2 | 4 | 4 | 4 | 4 | 4 | 4 | 4 |
| MgEs | 1 | 1 | 1 | 1 | 1 | | | | | | | | | | | |
| Aerosil | 1 | 1 | 5[a] + 2 | 5[a] + 2 | 5[a] + 2 | 5[a] | | | | | | | | | | |
| Acdisol | — | — | — | 5 | 5 | 5[a] | *5* | *3* | | | | | | | | |
| % Release In 30 Min | 62% | 65% | 71.5% | 89% | 85% | 88% | 100% IN 31 MIN | 100% IN 34 MIN | 100% IN 23 MIN | 46% | nF | 80% | 70% | 71% | 72% | 100% |

Abbreviations: CMM, cellulose microcrystalline; PEG, polyethylene glycal; CMCNa, sodium carboxymethyl cellulose; PVP, Kollidon K30; SDS, sodium docecyl sulfate; MgEs, magnesium stearate; NF, no flow ability.

[a]Internal phase.

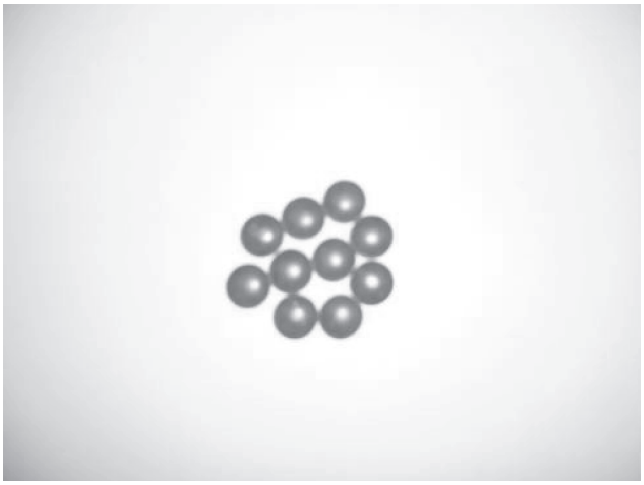

**FIGURE 15** Formulation F31.

**TABLE 7 Two Formulations for *mangifera indica* L. (Mango)**

| | Formulation F | Formulation F31 |
|---|---|---|
| | *Internal Phase* | |
| Extract *M. indica* L. | 50.0%[a] | 50.0%[a] |
| Lactose | | 31.0%[a] |
| Povidone (PVP K30) | 4.0%[a] | 4.0%[a] |
| Starch | 10.0%[a] | |
| CMM | 28.0%[a] | |
| Ac-Di-Sol | 5.0%[a] | |
| | *External Phase* | |
| SDS | — | 2.0% |
| $Na_2CO_3$ | — | 5.0% |
| PEG6000 | — | 5.0% |
| Talc | 2.5% | 2.5% |
| Magnesium stearate | 0.5% | 0.5% |

[a]Product added to internal phase.

was better than CMM, especially for flow properties. In all cases tablets always eroded and never disintegrated. All formulations studied are shown in Table 6.

**F23** and **F30** are shown in Figure 15. Two formulations, **F31** (obtained from the result of **F23** and **F30**) and F, were compared for water uptake and dissolution profile, as shown in Table 7.

The properties of formulations **F31** and **F** were measured in ENSLIN equipment in order to compare the kinetics of water uptake and to study the dissolution profile, as shown in Figures 16–19. As observed in Figure 19, the dissolution of plant extract after 60 min is very slow when pH-modified (**$NA_2CO_3$**) and sodium dodecyl sulfate (SDS) are not used (almost not change in color for formulation **F**).

The properties of formulation F31 are as follows: diameter 12 mm, normal concave; stability studies (see Table 8) under tropical condition, 30 °C and 75% RH with desiccant (silice 1 g for plastic bottle of polyethylene no. 8 for 60 tablets).

**FIGURE 16** Formulation F31 in ENSLIN equipment at initial time.

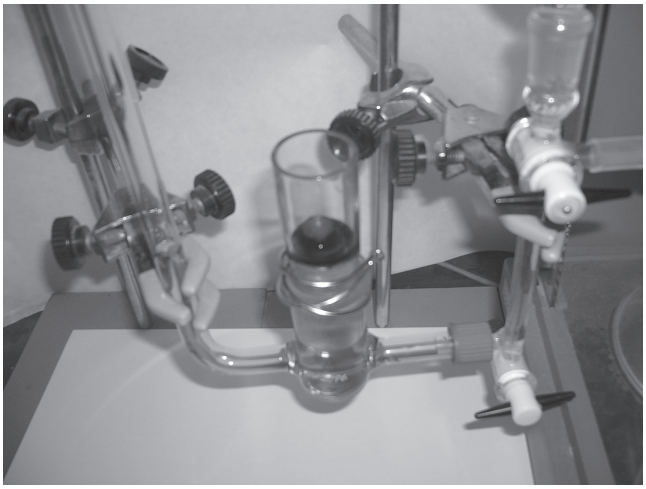

**FIGURE 17** Formulation F31 in ENSLIN equipment after 60 min.

### 6.2 Natural Product as Vehicle for Manufactured Tablets

Chitin was evaluated as a direct-compression vehicle using powder flow properties and the physicomechanical properties of the manufactured tablets, and it was proven that this natural polymer has suitable characteristics for being used for this end.

A comparative study of chitin obtained from lobsters, starch, and carboxymethyl chitosan as disintegrating agents was conducted. The influence of the method in the preparation of tablets on the disintegrating activity of both polymers was evaluated. Chitin proved to have good characteristics as a disintegrating agent independently

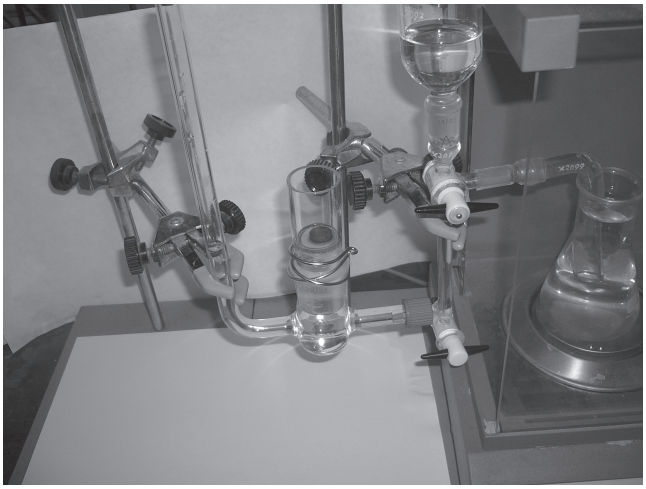

**FIGURE 18** Formulation F in ENSLIN equipment at initial time.

**FIGURE 19** Formulation F in ENSLIN equipment after 60 min.

**TABLE 8 Stability Studies for F31 under Tropical Condition**

| | Initial Time | 3 months | 6 months | 9 months | 12 months |
|---|---|---|---|---|---|
| **Weight** | 600.7 | 601.47 | 601.30 | 602.09 | 601.33 |
| **Assay** | 99.85 | 100.98 | 100.64 | 99.10 | 101.07 |
| **Disintegration, min** | 27 | 29 | 30 | 30 | 30 |
| **Hardness, N** | 59 | 60 | 64 | 60 | 60 |
| **Friability, %** | 0.89 | 0.74 | 0.80 | 0.77 | 0.81 |
| **Humidity, %** | 5.71 | 5.11 | 5.45 | 5.21 | 5.64 |

**TABLE 9 Formulation with Chitin, Chitosan, and Croscarmelose Sodium**

| | I | II | III | IV | V | VI | VII |
|---|---|---|---|---|---|---|---|
| Papaverin | 19.25 | 18.25 | 17.50 | 18.25 | 17.50 | 18.25 | 17.50 |
| Binder | 10.00 | 10.00 | 10.00 | 10.00 | 10.00 | 10.00 | 10.00 |
| Chitin | — | 1.00 | 1.50 | | | | |
| Chitosan | — | — | — | — | — | 1 | 1.50 |
| Croscarmelose sodium | — | — | — | 1.00 | 1.50 | | |
| Dibasic calcium phosphate | 70.00 | 70.00 | 70.00 | 70.00 | 70.00 | 70.00 | 70.00 |
| Magnesium stearate | 0.75 | 0.75 | 0.75 | 0.75 | 0.75 | 0.75 | 0.75 |

**TABLE 10 Results for Direct Compression**

| | I | II | III | IV | V | VI | VII |
|---|---|---|---|---|---|---|---|
| **Weight, mg** | 348 | 347 | 345 | 347 | 348 | 346 | 347 |
| **Release at 30 min, %** | <40 | 92.6 | 92.4 | 94.2 | 96.6 | 92.1 | 94.8 |
| **Disintegration, min** | >60 | 2.30 | 1.40 | 0.28 | 0.22 | 1.30 | 1.10 |
| **Hardness, kgf-Erw** | 3.92 | 3.99 | 3.97 | 3.97 | 4.03 | 3.99 | 3.92 |
| **Friability** | 2.44 | 2.24 | 2.10 | 2.01 | 1.97 | 1.89 | 1.94 |

**TABLE 11 Results for Wet Granulation**

| | FI | FII | FIII |
|---|---|---|---|
| **Weight, mg** | 346.50 | 348.20 | 349.00 |
| **Release at 30 min, %** | 81.00 | 86.00 | <70.00 |
| **Disintegration, min** | 23.70 | 23.00 | 45.30 |
| **Hardness, kgf-Erw** | 6.90 | 8.20 | 6.30 |
| **Friability** | 0.14 | 0.08 | 0.13 |

of the method used to make tablets (Tables 9–11). The disintegrating activity of carboxymethyl chitosan was affected by the granulation process.

Three formulas were prepared by wet granulation comparing starch, chitin, and chitosan as disintegrant (10%) formulations FI (10% starch), FII (10% chitin), and FIII (10% chitosan) (Table 11). Chitin has good properties as disintegrant. This product can be used for direct compression and wet granulation. The method for development (direct compresion or wet granulation) of tablets influences chitosan disintegrant properties.

### 6.3 Natural Product as Vehicle for Controlled-Release System

Dextrans are composed of chains of D-glucan ($1–20 \times 10^6$) with α-1,6 as the main-chain linkage and variable numbers of α-1,2, α-1,3, or α-1,4 branched-chain linkages. Dextran is synthesized from sucrose by dextransucrases, glucansucrases, and glucosyltransferases produced by *Leuconostoc* or *Streptococcus*. These bacteria growing in sugar juices produce dextran. High concentrations of dextran on solids (>1000 ppm) can result in severe financial losses to the sugar industry [25].

Dextran fractions obtained from enzymatic hydrolysis of native dextrans are supplied in molecular weights from 1000 to $2 \times 10^6$ Da. The molecular weight of the fraction is in most cases a key property and is defined in terms of the average molecular weight ($M_w$) and the number average molecular weight ($M_n$). The functionality of this raw material for controlling drug release is studied as a function of molecular weight. Fractions 43,000 (**F3**), 71,000 (**F2**), and 170,000 (**F1**) as native dextran $2 \times 10^6$ and $20 \times 10^6$ $M_w$ are used [26].

***Wet- and Dry-Weight Studies*** The method used was based on that of Tahara [15] and Jamzad [27]. The swelling and erosion of dextran polymers of differing molecular weights were examined by measuring the wet and subsequent dry weights of matrices. The experiment consisted of allowing the tablet (dextran alone) to dissolve in the medium (at the same condition described in drug release studies) for certain time periods (15, 30, 60 and 90 minutes) before being removed into a preweighed weighing boat. The excess dissolution medium was drained and blotted from around the tablet without touching it. The tablet and boat were then weighed to establish the wet weight of the tablet. The tablets were then dried to a constant weight in an oven at 105 °C. Each determination at each time point was performed in triplicate and mean values were expressed. The dissolution medium uptake per weight of dextran remaining was calculated at each time point for a particular matrix to correct for the effect of erosion and dissolution in the measurement of degree of dissolution medium uptake [Equation (1)]. Erode dextran was measured according to the equation described by Jamzad et al. [30] [Equation (2)]:

$$\text{Water uptake per unit polymer remaining (\%)} = \frac{\text{wet weight} - \text{dry weight}}{\text{dry weight}} \times 100 \tag{1}$$

$$\text{Mass polymer loss (\%)} = \frac{\text{original weight} - \text{remaining (dry) weight}}{\text{original weight}} \times 100 \tag{2}$$

The Davidson and Peppas model [Equation (3)] was applied to these data to study the mechanism and the rate of water uptake.

$$w = K_s t^n \tag{3}$$

where $w$ is the weight gain of the swelled matrix (water/dry polymer), $K_s$ the kinetic constant of water penetration, $t$ the penetration time, and $n$ the exponent which depends on the water penetration mechanism.

***Swelling and Erosion*** The change in wet weight, reflecting swelling, over time for the five polymer types is shown in Table 12. The higher molecular weight polymers showed the highest maximum average relative swelling, which occurred since the initial time with little erosion. In contrast, the lower molecular weight polymers (fractions **F1**, **F2**, and **F3**) exhibited minimal swelling and the erosion mechanism predominated. Consequently, tablets were dissolved very fast (100% before 45 min). These polymers and fractions show a wide range of viscosities, which cause differences in their swelling and erosion behaviors [26]. These results agree with results obtained by Sakar [C] and Walker [D] for HPMC polymer.

**TABLE 12 Swelling and Erosion Properties of Dextran Tablets as Funtion of Molecular Weight**

| | | Mass Polymer Loss (%) | | | |
|---|---|---|---|---|---|
| | Hardness (N) | 15 min | 30 min | 60 min | 90 min |
| DT, MW $2 \times 10^6$ | 431 | 11.48 | 14.59 | 27.473 | 36.224 |
| DT, MW $20 \times 10^6$ | 482 | 5.29 | 7.33 | 11.835 | 16.153 |
| F1 | 431 | 52.89 | 87.38 | | |
| F2 | 420 | 62.79 | 89.21 | | |
| F3 | 460 | 66.3 | 95.85 | | |

| | Percentage of Water Uptake | | | |
|---|---|---|---|---|
| | 15 min | 30 min | 60 min | 90 min |
| DT, MW $2 \times 10^6$ | 59.80 | 84.68 | 103.73 | 110.15 |
| DT, MW $20 \times 10^6$ | 110.03 | 142.94 | 178.50 | 210.02 |
| F1 | 31.61 | 33.14 | | |
| F2 | 17.61 | 26.89 | | |
| F3 | 16.82 | 17.65 | | |

*Note:* All values are referred to applied force, 14 kN, particle size for dextran 150–200 μm, tablet weight 300 ± 11 mg, dissolution media 1000 mL distilled water, temperature 37°C, 100-rpm paddle. Abbreviations: DT, dextran; MW, molecular weight.

For native dextran a linear relationship was seen between mass polymer loss and initial dissolution time. Native dextrans also showed the highest maximum dissolution medium uptake. Here, an increase in the molecular weight of dextran resulted in an increase in water uptake (native polymer with 10 times more than fraction **F1**, 13 times more than **F2**, and 15 times more than **F3** for the first 30 min) and less erosion. Anywhere the rate of water uptake per unit weight of polymer started to decline with last initial time and in consequence for longer periods of time, nonlinear dependence could be expected. Applying the Davidson–Peppas model [Equation (3)], a value of $n = 0.356$ ($r^2 = 0.998$) for native dextran was obtained ($r^2 = 0.984$). An inverse relationship between erosion rate constant and molecular weight was reported by Reynolds et al. [28]. Tahara et al. [18] reported that the lower viscosity HPMC (50-cps) polymer eroded faster than the 4000-cps polymer, consistent with the current work. Thus, the higher molecular weight native dextran polymers have higher intrinsic water-holding capacity and the matrices formed from such polymers are less prone to erosion than the lower molecular weight fractions.

### 6.4 Mechanism of Soluble Principle Active Propranolol Hydrochloride and Lobenzarit Disodium from Dextran Tablets

Soluble drugs are considered to be released by diffusion through the matrix and poorly soluble drugs are released by erosion of the matrix. Moreover, it is considered that factors affecting swelling and erosion of these polymers may account for differences between in vitro dissolution results and subsequent in vivo performance when hydrophilic matrix tablets are compared [15].

Lobenzarit disodium (LBZ) is a drug conceived for the treatment of rheumatoid arthritis. This drug produces an improvement of immunological abnormalities and

has a regulatory effect upon the antibody-producing system. Propranolol hydrochloride (PPL) is a β-adrenergic blocking agent, that is, a competitive inhibitor of the effects of catecholamines at β-adrenergic receptor sites. It is widely used in therapeutics for its antihypertensive, antiangorous, and antiarrhythmic properties. These two drugs are suitable candidates for the design of controlled-release delivery systems [25, 29]. According to their solubility in water they can be considered as high soluble (PPL) and soluble (LBZ) drugs.

A comparative study of the dissolution profile for PPL and LBZ was established as the analysis of a similarity factor defined as

$$f_2 = 50\log\left\{\left[1+\frac{1}{n}\sum_{t=1}^{n}(R_t - T_t)^2\right]^{-0.5}\times 100\right\} \tag{4}$$

In the above equation $f_2$ is the similarity factor, $n$ is the number of time points, $R_t$ is the mean percent drug dissolved of the reference formulation, and $T_t$ is the mean percent drug dissolved of the tested formulation.

The evaluation of similarity is based on the following conditions:

- A minimum of three time points
- Twelve individual values for every time point
- Not more than one mean value of >85% dissolved
- Standard deviation of the mean that is less than 10% from the second to last time point

An $f_2$ value between 50 and 100 suggests that two dissolution profiles are similar [30]. In this study experimental data corresponding to 30, 60, 90, 120, 180, 240, 300, 360, 420, and 480 min were considered.

Figure 20 shows dissolution profiles for tablets of PPL or LBZ from the native dextran DTB110-1-2 matrix system, respectively 1 : 1 (w/w). The value for relative standard deviation (CV) was less than 5% for all points measured ($n$ = 12).

The Higuchi and Hixson Crowell model as well as the nonlinear regression of Peppas and Peppas-Sahlin were employed to study the release data. Higuchi's slope

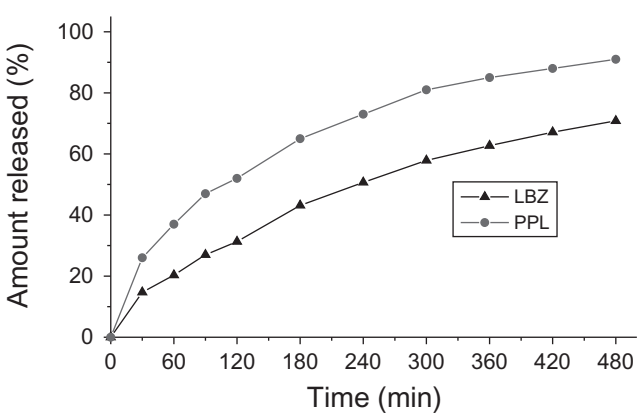


**FIGURE 20** Disolution profile for PPL and LBZ dextran tablets (direct compression): paddle 100 rpm, 37 °C, deionized water.

(3.179 and 4.500% $min^{-1/2}$ for LBZ and PPL, respectively), Korsmeyer's rate constant (1.195% $min^{-0.697}$ and 4.125% $min^{-0.540}$ for LBZ and PPL, respectively), the low relaxational constant $K_r$ (0.101% $min^{-0.898}$ for LBL and –0.040% $min^{-0.898}$ for PPL), compared with $K_d$ values (2.941 and 6.518% $min^{-0.449}$, respectively) of Peppas-Sahlin indicated the diffusional mechanism as predominant for soluble drugs since native dextran tablets. The influence of solubility of the drug can be observed for the hydrophilic matrix (release of PPL is faster than LBZ in correspondence with its solubility in water). The value of the diffusional exponent, 0.697 (by the Korsmeyer equation), for the less soluble drug corresponds to the increment of the influence of the erosion mechanism, in agreement with other authors. This can be also observed in the Peppas y Salhin equation where the negative value obtained for $K_r$ for the dissolution profile of PPL from dextran tablets should be interpreted in terms of a relaxation mechanism, which is insignificant compared to the diffusion process.

The dissolution profiles for LBZ:B110-1-2 and PPL:B110-1-2 tablets were also compared using similarity factor $f_2$. A value obtained for $f_2$ that is below 50 (37.48) indicates the influence of drug solubility in the dissolution profile. Furthermore, other parameters, such as dextran–drug ratio, particle size of polymer and drug, and influence of pH, have to be studied to obtain an optimum and robust formulation.

The mechanisms of drug release from dextran matrix occur in the early stage by polymer swelling, and the tablet thickness increases. Soon thereafter, polymer (and drug) dissolution starts occurring. The polymer dissolves because of chain disentanglement. Thus, there is a slow diminution of the thickness because of erosion until, finally, the tablet disappears (time > 480 min).

## 7 DESIGN TOOLS OF TABLET FORMULATION

Nowadays, most experimentation on tablet formulation development is still performed by changing the levels of each variable (factor) at a time, in an unsystematic way, keeping all other variables constant in order to study the effects of that specific variable on the selected response or to find the optimal conditions of a complete system. This methodology (trial and error) is based on a large number of experiments and often relies merely on the analyst's experience [31].

Statistical experimental design, also called design of experiments (DoE), is a well-established concept for planning and execution of informative experiments. DoE can be used in many applications. An important type of DoE application refers to the preparation and modification of mixtures. It involves the use of "mixture designs" for changing mixture composition and exploring how such changes will affect the properties of the mixture [32].

In the DoE approach, first process variables are "screened" to determine which are important to the outcome (excipient type, percentage, mixture time, etc.). The next step is *optimization*, when the best settings for the important variables are determined. In particular, response surface methodologies have been successfully applied in both drug discovery and development [33]. Advances in supporting software, automated synthesis instrumentation, and high-throughput analytical techniques have led to the broader adoption of this approach in pharmaceutical discovery and chemical development laboratories [34].

The benefits of using experimental design together with software to facilitate the formulation of a tablet for specific purposes, from screening to robustness testing, are well known. This technique has some advantages compared to the trial-and-error method. By applying a multivariate design for the screening experiments, many excipients are evaluated using comparatively few experiments.

The formulation work is generally based on designed experiments. Most of the experiments are fractional or full-factorial designs and are generated and evaluated in some cases with the center point replicated. The robustness of the formulation and batch-to-batch variation of the excipients and the active pharmaceutical ingredient can be evaluated with experimental designs on different occasions. Experimental design and optimization of the formulation can be performed with the use of software. Some of them have been useful in tablet design. MODDE (version 4.0 and 5.0, Umetri, Umeå, Sweden), iTAB [35], and TabletCAD are some examples.

### 7.1 MODDE 4.0

Tablet design for controlled-release propranolol hydrochloride was performed with the use of MODDE software [25]. A central composite design (one of the most used designs in pharmaceuticals) was applied to the optimization. This experimental design required 17 experiments ($2^k + 2k + 3$, where $k$ is the number of variables) including three center points. Three variables and five responses (according to USP 25 tolerances for the dissolution profile for propranolol hydrochloride extended-release capsule) were involved in the experimental design. The variables and their ranges studied are summarized in Table 13. The high and low values of each variable were defined based on preliminary experiments. The critical responses were $t_{100\%}$ and $t_{30\%}$ corresponding to the time when 100 and 30% of drug contained in the tablets is delivered to the dissolution medium because this system was developed to release drug in 24 h ($t_{100\%} \sim 24$ h) and to prevent an overdose for the first minutes ($t_{30\%} > 1.5$ h). The other responses were in the amount of PPL dissolved at 4, 8, and 14 h.

Table 14 shows results obtained for every formula development according to MODDE 4.0 software. The collected experimental data were fitted by a multilinear regression (MLR) model with which several responses can be dealt with simultaneously to provide an overview of how all the factors affect all the responses. The responses of the model, $R^2$ and $Q^2$ values, were over 0.99 and 0.93 for $t_{100\%}$ and 0.98 and 0.89 for $t_{30\%}$, respectively, implying that the data fitted well with the model. Here, $R^2$ is the fraction of the variation of the response that can be modeled and $Q^2$ is the fraction of the variation of the response that can be predicted by the model. The relationship between a response $y$ and the variables $x_i$, $x_j$, . . . can be described by the polynomial:

**TABLE 13 Levels of Formulation Variables (Central Composite Design)**

| Parameter | Low Value (–1) | Central Value (0) | High Value (+1) |
|---|---|---|---|
| Ratio DT–HPMC (w/w) | 1:1 | 4:1 | 7:1 |
| Cetyl alcohol (% w/w) | 10 | 15 | 20 |
| Ratio excipients–PPL (% w/w) | 30 | 50 | 70 |

**TABLE 14 Matrix of Central Composite Design and Results**

| Run Order | CEx–PPL | DT–HPMC | ce | $t_{100\%}$* | $t_{30\%}$* | $t_2$* | $t_3$* | $t_4$* |
|---|---|---|---|---|---|---|---|---|
| 10 | 30 | 1:1 | 10 | 10 | 0.5[a] | 71[a] | 90[a] | — [a] |
| 7 | 70 | 1:1 | 10 | 15 | 2.1[b] | 44[b] | 73[b] | 97[a] |
| 1 | 30 | 7:1 | 10 | 11 | 0.5[a] | 63[a] | 82[a] | — [a] |
| 11 | 70 | 7:1 | 10 | 14 | 2.2[b] | 40[b] | 76[b] | 99[a] |
| 9 | 30 | 1:1 | 20 | 14 | 1.3[a] | 49[b] | 70[b] | 99[a] |
| 8 | 70 | 1:1 | 20 | 16 | 2.2[b] | 46[b] | 62[b] | 96[a] |
| 5 | 30 | 7:1 | 20 | 14 | 1.6[b] | 49[b] | 70[b] | 99[a] |
| 15 | 70 | 7:1 | 20 | 16 | 2.4[b] | 43[b] | 60[b] | 96[a] |
| 3 | 30 | 4:1 | 15 | 20 | 1.2[a] | 54[b] | 79[b] | 94[b] |
| 16 | 70 | 4:1 | 15 | 26 | 2.6[b] | 39[b] | 56[b] | 72[b] |
| 17 | 50 | 1:1 | 15 | 15 | 1.6[b] | 45[b] | 63[b] | 98[a] |
| 4 | 50 | 7:1 | 15 | 16 | 1.7[b] | 43[b] | 65[b] | 96[a] |
| 2 | 50 | 4:1 | 10 | 21 | 1.4[a] | 52[b] | 71[b] | 93[b] |
| 12 | 50 | 4:1 | 20 | 24 | 2.2[b] | 45[b] | 63[b] | 82[b] |
| 6 | 50 | 4:1 | 15 | 24 | 2.1[b] | 48[b] | 65[b] | 85[b] |
| 14 | 50 | 4:1 | 15 | 23.5 | 2.2[b] | 49[b] | 66[b] | 86[b] |
| 13 | 50 | 4:1 | 15 | 24 | 2.1[b] | 48[b] | 66[b] | 86[b] |

*Note:* cEx–PPL, ratio of excipients to propranolol; DT–HPMC, ratio of native dextran to hydroxypropylmethylcellulose; ce, percentage of cetyl alcohol (w/w) in the tablets. Time $t_{100\%}$ is time (hours) when 100% of PPL is dissolved in dissolution medium and $t_{30\%}$ is time (hours) when 30% of PPL is dissolved in dissolution medium; $t_2$ is percent of PPL dissolved at 4 h, $t_3$ percent of PPL dissolved at 8 h, and $t_4$ percent of PPL dissolved at 14 h. Values presented are the average of eighteen replicates for each batch. [a]Inside the USP range. [b]Outside USP range.

$$y = \beta_0 + \beta_i x_i + \beta_j x_j + \beta_{ij} x_i x_j + \beta_{ii} x_i^2 + \beta_{jj} x_j^2 + \cdots + E$$

where $\beta_j$ are coefficients to be determined and $E$ is the overall experimental error. Figure 21 presents the dissolution profiles of all 17 trials generated from the central composite design.

The response surface plots formed by plotting the values for $t_{30\%}$ and $t_{100\%}$ as a function of the most important variables are shown in Figure 22, where the optimum condition obtained by the model can be seen. The optimum dextran (DT)–HPMC ratio of 4:1 (w/w) gave $t_{100\%}$ equal to 24 h.

With tablet formulations composed of a matrix excipient and PPL at a ratio ranging from 40:60 to 70:30 (w/w), the values for $t_{100\%}$ were satisfactory (around 24 h). However, the respective values for $t_{30\%}$ increased according to the ratio of matrix excipient and PPL ranging from 40:60 to 70:30 (w/w), showing that early drug release was demanded and the initial dose required for pharmacological effect could not be sufficient.

Sustained-release matrix tablets with good properties were obtained with a dextran–HPMC ratio of 4:1 (w/w), with a matrix excipient–PPL ratio of 60:40 (w/w), and with a cetyl alcohol amount of 15% (w/w). The hydrophilic polymers–PPL ratio of 60:40 (w/w) is more robust for any manufactured variability than 50:50 (w/w), because the central point of the design is near the lowest desired area (Figure 22). Under the optimal conditions, the mean value of hardness was 106 ± 3 N and the friability was less than 1% (0.2%).

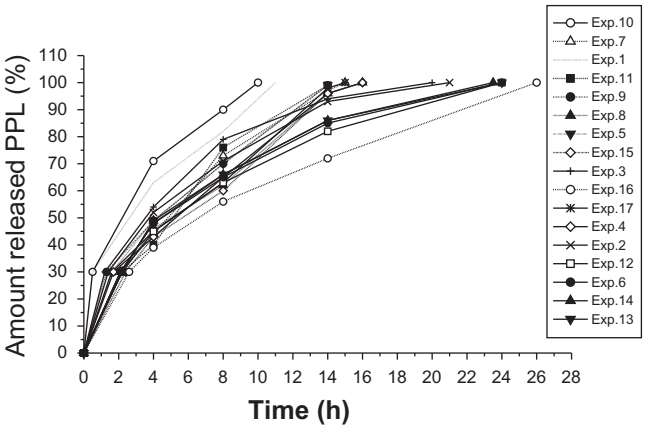


**FIGURE 21** Dissolution profile generated from central composite design.

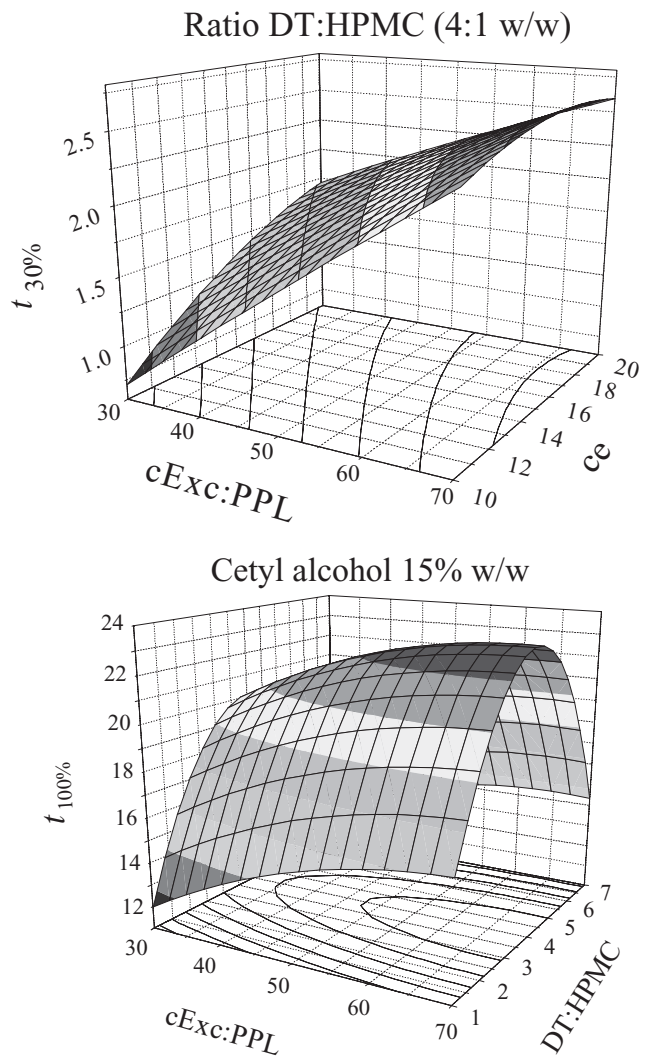


**FIGURE 22** Response surface plots formed by plotting values for $t_{30\%}$ and $t_{100\%}$.

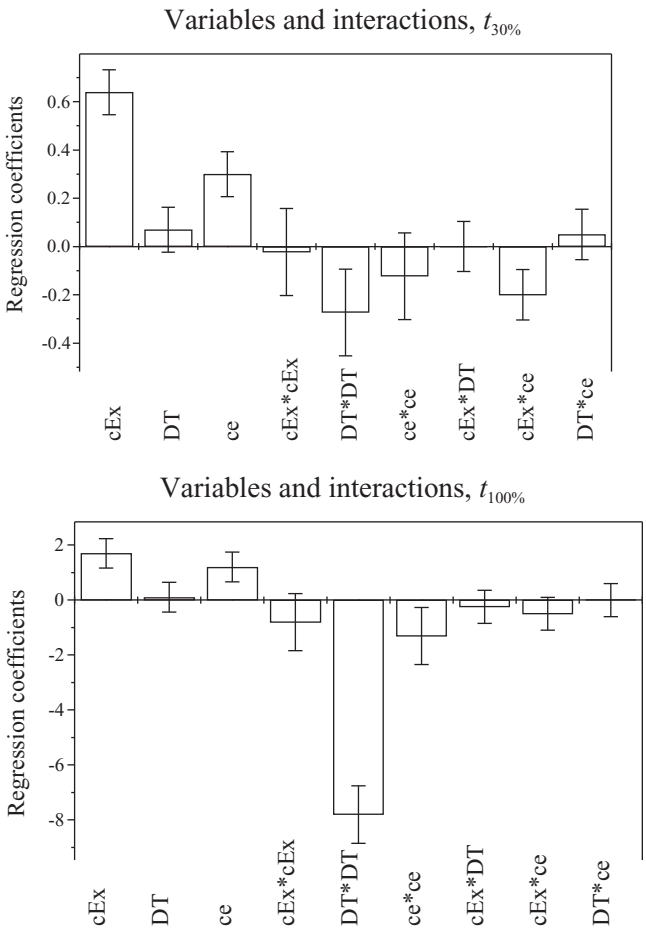


**FIGURE 23** Regression coefficients.

Cetyl alcohol (ce in Figure 23) has a significant positive effect on both responses in the range studied. This may be because the hydrophobic polymer prevents the fast release of PPL for the first few hours with an increase in the diffusional path-length of the drug because the swelling of hydrophilic polymers (DT and HPMC) retards the rate of release. Interaction of cetyl alcohol and hydrophilic polymers–PPL was observed for $t_{30\%}$. If a prolonged release rate is desired during this period, the ratio of hydrophilic polymers to cetyl alcohol can be increased, resulting in a decreased interspace volume after erosion of cetyl alcohol. In contrast to other products, such as lactose [15], cetyl alcohol as hydrophobic polymer can be increased. The viscosity and texture of the gel layer and some modifications in polymer–polymer and polymer–solvent interaction are present.

## 7.2 iTAB

iTAB is I Holland's new tablet design aid that calculates basic tablet parameters and stress analysis for "rounds" and "shapes" in three easy steps. iTAB is Windows

based, saving users from any extra set-up costs and making software upgrades easy.

Users are not required to possess any technical expertise or have any formal training, as iTAB is very user friendly and intuitive throughout. Using the drag-and-click feature, iTAB results in 2D drawings ready for production and automatically generated design reports:

Step 1 involves the selection of a specific tablet profile and modification of key parameters such as diameter, mass, volume, and surface area in real time. The iTAB safety zone ensures that unfeasible designs cannot be output from the system, meaning that only quality tablets can be produced.

Step 2 runs a real-time finite-element analysis (FEA) simulation against the design to give an indicative punch tip maximum force calculation.

Step 3 automatically produces a report summarizing key design data, for example, cup depth and surface area, that can be emailed directly to the I Holland team for further analysis, design work, and tooling production.

iTAB allows non–computer-aided design (CAD) and non-FEA users to simply and quickly design nonembossed tablets and obtain detailed information with the click of a button.

Core benefits include the avoidance of tablet manufacturing problems such as punch tip breakage from the outset. Immediate and accurate punch tip load calculations and 3D dynamic rotation of punch tip stress allow for instant decision making on design issues. iTAB is also unique in that it incorporates tooling and tablet design.

### 7.3 Percolation Theory

Leuenberger et al. introduced percolation theory in the pharmaceutical field in 1987 to explain the mechanical properties of compacts and the mechanisms of the formation of a tablet [36, 37]. Knowledge of the percolation thresholds of a system results in a clear improvement of the design of controlled-release dosage forms such as inert or hydrophilic matrices.

Percolation theory is a statistical theory that studies disordered or chaotic systems where the components are randomly distributed in a lattice. A cluster is defined as a group of neighboring occupied sites in the lattice, being considered an infinite or percolating cluster when it extends from one side to the rest of the sides of the lattice, that is, percolates the whole system [38].

Thus, a tablet is regarded simply as a heterogeneous binary system formed by the active principle and an excipient. As a function of their relative volume ratio, one or both components constitute a percolating cluster formed by particles of the same component that contact each other from one side to the other sides of the tablet.

In a binary pharmaceutical tablet (cylindrical lattice), the sites can be occupied by the component drug or excipient. The percolation threshold of the drug indicates at which concentration this substance dominates the drug/excipient system. The concept is very similar to the point where a component passes from being the inner

to being the outer phase of an emulsion. It is not surprising that the component becoming the "outer phase" or percolating phase will have more influence on the properties of the system.

Furthermore, the concentration point at which a component is starting to percolate the system is usually related to a change in the properties of the system, which will now be more affected by this component. This is known as a critical point. Close to the critical point important changes can take place, for example, changes in the release mechanism of the active agent and modification of the tablet structure (e.g., monolith versus a desegregating device).

An important difference between particulate solids and emulsions is that in the solids two components can percolate the system at the same time, that is, two components can act as the outer phase simultaneously. In this case the system is known as a bicoherent system.

***Study of Ternary Tablets*** Percolation theory has been developed for binary systems, however, drug delivery systems usually contain more than two components. The existence and behavior of the percolation thresholds in ternary pharmaceutical dosage forms have been studied [39] employing mixtures of three substances with very different hydrophilicity and aqueous solubility (Polyvinylpyrrolidone (PVP) cross-linked, Eudragit RS-PM, and potassium chloride).

After evaluation of the technological parameters and in vitro release behavior of the tablets, no sharp percolation thresholds were found in these ternary systems for the employed components separately. Nevertheless, a combined percolation threshold of the hydrophilic components was found, demonstrating that a multicomponent system can be reduced to a binary one using a discriminating property [39].

***Matrix Systems with Different Particle Sizes*** Another disadvantage in the application of percolation theory to the rationalization of the pharmaceutical design was the prerequisite of an underlying regular lattice. Usually, drug delivery systems contain substances with different particle sizes. Therefore, the particles cannot be considered as each occupying one lattice site.

This problem can be initially overcome using a volume ratio instead of a lattice site ratio, expressing the percolation thresholds as critical volume fractions [36, 40–42]. Nevertheless, the influence of the particle size of the components on the percolation threshold cannot be explained using a volume fraction; that is, from this point of view, tablets with the same excipient volume are equivalent independent of their particle size. A first qualitative study of the influence of particle size on the percolation threshold [43] demonstrated that this is in clear disagreement with experimental data.

According to percolation theory, the effect of a reduction in the drug particle size should be similar to an increase in the excipient particle size in a binary system: It may be expected that the relative particle size of the component, but not its absolute particle size, will determine the properties of the system.

A quantitative study of the influence of particle size on the percolation threshold employing inert matrix tablets prepared with KCl and Eudragit RS-PM as matrix-forming material [44–46] showed that experimental data are in agreement with this hypothesis.

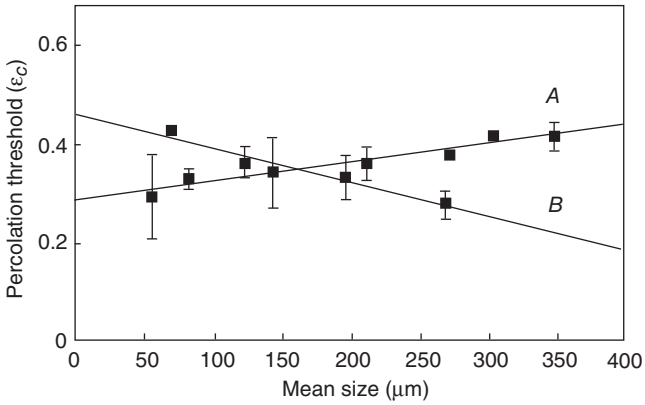


**FIGURE 24** Drug percolation threshold (mean ± SE) as function of mean particle size of drug (line *A*) and excipient (line *B*) employed.

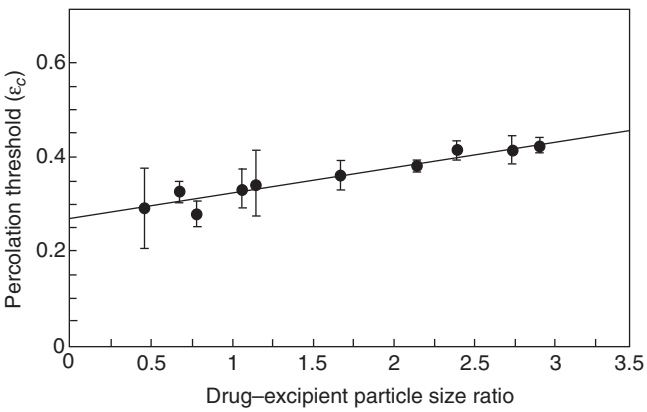


**FIGURE 25** Drug percolation threshold (mean ± SE) as function of drug–excipient particle size ratio employed.

As Figure 24 shows, a linear relationship was found in this study [44] between the mean drug particle size and the corresponding drug percolation threshold (Figure 24, line *A*). Furthermore, the excipient particle size exerts a contrary effect than the drug particle size (Figure 24, line *B*); that is, the larger the excipient particle size, the lower the drug percolation threshold [46].

In addition, when the obtained drug percolation thresholds were plotted as a function of the drug–excipient particle size ratio of the matrices (see Figure 25), a linear relationship was found between the drug percolation threshold and the relative drug particle size [46]. These results are in agreement with the above exposed theoretical model based on percolation theory.

One of the advantages of the proposed model versus classical theories is its ability to explain the changes in the release behavior of the matrices by means of a change in the critical points of the system (related to the drug and excipient percolation thresholds), which can be experimentally calculated, providing a scientific basis for the optimization of these dosage forms.

***Mechanical Properties*** The percolation approach was also employed to model the tensile strength of tablets [47, 48]. A critical tablet density was here understood as a minimal solid fraction needed to build a network of relevant contact points spanning the entire tablet. A rising tablet density led to a power law increase of the tensile strength showing an universal exponent $T_f = 2.7$.

It was shown that a power law based on percolation theory was suitable to fit the obtained tensile strength data of the binary matrix tablets studied. The best fitting was observed for a model where an initial tensile strength $\sigma_0$ was supposed [49]:

$$\sigma_t = k(\rho - \rho_c)^{2.7} + \sigma_0$$

The observed critical relative densities are understood as threshold values for the tensile strengths of the tablets. One practical consequence of these works is to avoid the manufacture of matrix tablets close to these critical densities. The formulation may not be robust in this critical range from the viewpoint of mechanical tablet stability.

### 7.4 Artificial Neural Networks

Artificial neural networks (ANNs) are computer programs designed to model the relationships between independent and dependent variables. They are based on the attempt to model the neural networks of the brain [50]. Functions are performed collectively and in parallel by the units, rather than there being a clear delineation of subtasks to which various units are assigned.

This methodology represents an alternative modeling technique that has been applied to pharmaceutical technology data sets, including tableting parameters [51]. The main advantage with respect to classical statistical techniques, such as response surface methodology, is that ANNs do not require the prior assumption of the nature of the relationships between input and output parameters, nor do they require the raw data to be transformed prior to model generation [51]. ANNs are capable of modeling complex, nonlinear relationships directly from the raw data.

The functional unit of ANNs is the perceptron. This is a basic unit able to generate a response as a funtion of a number of inputs received from others perceptrons. For example, the response value can be obtained as follows:

$$Y = \begin{cases} 1 & \text{if } W_0 I_0 + W_1 I_1 + W_b > 0 \\ 0 & \text{if } W_0 I_0 + W_1 I_1 + W_b \leq 0 \end{cases}$$

where $I_x$ is the input received from perceptron $x$ and $W_x$ the weight assigned to this input by the perceptron. The weights can be changed to adapt the answer to the desired one using a learning algorithm.

Usually complex structures with more than 15 layers are employed, called the multilayer perceptron (MLP). Some of the commercial programs which have been used to fit tableting parameters are INForm (Intelligensys, Billingham Teesside), CAD/Chem (AI Ware, Cleveland, OH), which is no longer commercially available, and the Neural Network Toolbox of MATLAB (MathWorks, Natick, MA).

## 8 COATING SYSTEMS

Coating processes have come to play an important role for the protection of substances prior to application or for their sustained release. Coatings on cores usually consist of a mixture of substances. The matrix formers are responsible for the stability of the coating structure, and they also determine the coating process. Depending on the type of matrix former on binder used, three coating categories can be distinguished:

*Coatings with Sucrose and Other Sugars* Permit application of copious amounts of mass to the core and are widely used in the manufacture of pharmaceuticals and confectionery.

*Hot Melts* Add a considerable amount of mass and are applied hot and solidify while cooling on the core. They are mainly used for confectionery. The most important raw materials are fats, mostly cocoa fat, polyethylene glycol (PEG) [52], and the sugar–alcohol mixture xylitol–sorbitol.

*Film Coatings* Require less material, forming thin membranes which largely follow the contours of the substrate, for example, scores and engravings. The partly pH-dependent solubility and selective permeability of coatings are affected by the film formers. Such films are sometimes also used as intermediate layers in sugar coatings.

Whether or not a core is suitable for coating depends on its hardness, shape, surface, size, heat sensitivity, and tendency to interact with the coating material. Moreover, since sugar and film coating processes involve very different techniques, they place different demands on the cores to be used [53]. Tablets used as cores must be biconvex in shape to prevent them from sticking together like coins in a roll.

The ideal tablets for sugar coating will have a pronounced convex curvature and a narrow band. The consumption of coating material depends very much on the tablet shape and increases sharply if the tablets are not round. Film coating supplies coated products in which the core surface (e.g., with notches, engravings, and defects) is faithfully reproduced.The films tend to chip at sharp edges or are particularly thin in these areas. For this reason slightly curved tablet cores are preferred for film coating [54].

Film coating of pharmaceuticals is a common manufacturing stage for the following reasons: (i) to provide physical and chemical protection for the drug, (ii) to mask the taste or color of the drug, or (iii) to control the release rate or site of the drug from the tablet. When a coating composition is applied to a batch of tablets or granules (or to a batch of liquid drops or even gas bubbles), the core surfaces become covered with a polymeric film that is formed as the surfaces dries. The major component in a coating formulation is a film-forming agent which ideally is a high-molecular-weight polymer that is soluble in the proper solvent (today, most preferably in aqueous-based media). The polymer forms a gel and produces an elastic, cohesive, and adhesive film coating.

In the pharmaceutical industry, organic-solvent-based film coatings have been used for over 40 years. In the 1990s, however, interest and demands in the use of aqueous-based film coating systems rapidly increased owing to the well-documented

drawbacks (unsafe, toxic, pollutive, and uneconomic) associated with organic-solvent-based coating systems. Consequently, and for the reasons mentioned above, much effort has been focused on the research and development of new aqueous-based film coating formulations. Nowadays, aqueous-soluble/dispersable polymers available on the market consist primarily of either cellulose polymers, PEGs, or acrylate copolymers. There are, however, some material-related limitations in using these polymers in aqueous-based film coatings. Consequently, application of new film formers such as chitosan, native starches, and special types of proteins for pharmaceuticals and foodstuffs has been increasingly studied.

The choice between sugars of film coatings depends not only on the desired coatings quality but also on the technical requirements, that is, on the economy of the process. The financial outlay for a selected technology often commits the manufacturer to this technology for a prolonged period of time. Film and sugar differ substantially in thickness and therefore also in the necessary mass of coating material.

The most important coating raw material is the film-forming polymer, which must be able to produce a coherent film on the substrate under the given process conditions [55, 56]. Second in importance is the solvent or dispersing system in which the polymer is applied to the surface and introduced to form a film. Other frequently used raw materials are plasticizers, glidants, fillers, and colorants. All these substances act together and influence the properties of the film.

The first decision to be made when developing a formulation concerns the desired function of the film. Depending on the requisite dissolution performance in physiological media, the film former to be tested—or several of them—is then selected from the available polymers. The economy of the coating process and the quality of the coated product depend on the correct calculation of the required amount of coating material. Empirical adjustment of new developments to existing products is therefore not recommended. The coating quantity needed for film coatings is directly related to the surface area of the core. Table 15 shows a simplified calculation of the surface area of tablets. It is based on the surface area of a cylinder circumscribing these shapes. All these data are taken from the literature. For sugar coating processes, the shape of the cores and end products have to be studied stereometrically. The sugar coating process balances irregularities, since the coating buildup does not follow the structure of the surface. Areas which require a high degree of rounding, such as high bands and very flat curvatures, attract more coating mass. Where the band height varies owing to manufacturing technique, this results in weight differ-

**TABLE 15 Calculation of Surface Area ($mm^2$) for Different Types of Cores**

| | Diameter (mm) | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Height (mm) | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 12 | 14 |
| 2 | 33 | 56 | 70 | 95 | 120 | 150 | | | | |
| 3 | 42 | 62 | 85 | 115 | 145 | 175 | 210 | 250 | 340 | |
| 4 | | 75 | 100 | 130 | 165 | 200 | 240 | 280 | 380 | 485 |
| 5 | | | | | 185 | 225 | 270 | 315 | 415 | 530 |
| 6 | | | | | | | 300 | 345 | 450 | 570 |
| 7 | | | | | | | | | 505 | 615 |
| 8 | | | | | | | | | | 660 |

ences between the individual sugar-coated tablets. This is one of the reasons why sugar-coated products always show a much wider weight distribution than film-coated tablets.

Several authors have made proposals for the most convenient size of tablet for sugar coating. A further rule has been established according to which the convex radius should be between 0.7 and 0.75 times the tablet diameter and the band height between 0.07 and 0.12 times that diameter. However, for calculation of the band height only a factor of 0.12 is advisable, since, otherwise, the minimum value of 1 mm will be fallen short of by far [53, 57–60]. For film coating the recommended convex radius is 1.5 times the tablet diameter (less curved tablets are preferred).

Numerous formulation are available for all film-forming polymers offered in the market. Thus, it is possible to dispense with many of the preliminary tests and concentrate development work on the special problems of the formulation in question. In this chapter, we will focus on SEPIFILM and Kollidon VA 64. Table 16 presents proven basic formulations reported on in the literature [61] and polymers used for the most important commercially available film formers, which also can be used as a basis for further tablet coating.

The major operational coating process parameters related to film coating are able to be measured and monitored continuously in some pan-type coaters. The inlet airflow rates influence the coating process and the subsequent quality of the coated tablets. Increasing the inlet airflow rate accelerated the drying of the tablet surface. At high inlet airflow rate, obvious film coating defects, that is, unacceptable surface roughness of the coated tablets, are observed and the loss of coating material increased.

Today advantages of aqueous film coating are well recognized and film coating technology is much developed to successfully perform these types of coatings. Process automation and monitoring of critical process parameters can be utilized to increase the overall process efficiency and predictability and to improve the homogeneity and reproducibility of the tablet batches. This will ensure high quality and safety of the final coated products, which are mandatory requirements of tablet manufacturing.

### 8.1 Subcoating of Tablet Cores as a Barrier to Water

As tablets are nowadays coated mostly with aqueous solutions or dispersions, it has become increasingly necessary to provide the tablet cores with a barrier layer prior to sugar or film coating. This is mainly to protect water-sensitive drugs against hydrolysis and chemical interactions, for example, between different vitamins, and to prevent the swelling of high-performance tablet desintegrants that are very sensitive even to small quantities of water. It can be especially useful when controlled-release systems with hydrophilic polymer are studied and the water contained can change the dissolution profile. Kollidon VA 64 also can be used to improve the adhesion of subsequent coatings by hydrophilization of the surface.

### 8.2 Kollidon VA 64

We are studying the ability of Kollidon VA 64 as a subcoating in a combined hydrophilic (dextran–HPMC)–hydrophobic (cetyl alcohol) matrix core prior to sugar

**TABLE 16 Basic Formulations for Film-Coated Tablets**

| *Formulation with Hydroxypropyl Methylcellulose (HPMC)* | | |
|---|---|---|
| | A | B |
| Oprady | 73.0% | — |
| Pharmacoat | — | 80.0% |
| PEG 6000 | — | 8.0% |
| Talc | 20.0% | 5.0% |
| Pigments included, $TiO_2$ | 7.0% | 7.0% |
| Solid content | 20.0% | 12.0% |
| Coating quantity, mg/cm$^2$ | 1–5 | 1–5 |

| *Formulation with Methacrylic Acid Copolymers* | | | |
|---|---|---|---|
| | A | B | C |
| Eudragit L100 | 5.0% | — | — |
| Eudragit S100 | — | 7.5% | — |
| Eudragit L30D-55 | — | — | 16.5% |
| PEG 6000 | 0.7% | 1.0% | 1.6% |
| Talc | 6.0% | 2.0% | 4.0% |
| Pigments included, $TiO_2$ | 3.3% | — | — |
| Isopropyl alcohol | 41.0% | 86.5% | — |
| Acetone | 41.0% | — | — |
| Water | 3.0% | 3.0% | 77.9% |
| Solid content | 14.0% | 10.5% | 22.1% |
| Coating quantity, mg/cm$^2$ | 2–4 | 2–4 | 3–5 |

| *Formulation with Hydroxypropyl Cellulose Acetate Succinate (HPMCAS)* | |
|---|---|
| Aqoat AS-MF | 10.0% |
| Triethyl citrate | 2.8% |
| Talc | 3.0% |
| Sorbitan sesquioleate | 0.0025% |
| Water | 84.2% |

| *Formulation with Ethylcellulose (EC)* | | |
|---|---|---|
| | A | B |
| Ethylcellulose | 5.0% | — |
| Aquacoat (30% solids) | — | 30.0% |
| PEG 6000 | — | 2.0% |
| Glycerol triacetate | 1.0% | — |
| Ethanol | 94.0% | — |
| Water | — | 68.0% |

| *Formulation with Carboxymethyl Ethylcellulose (CMEC)* | |
|---|---|
| Duodcell | 8.0% |
| Trisodium citrate | 0.70% |
| Tween 80 | 0.04% |
| Glycerol monocaprylate | 2.40% |
| Water | 88.86% |
| Coating quantity | 7 mg/cm$^2$ |

| *Formulation with Polyvinyl Acetate Phthalate (PVAP)* | |
|---|---|
| PVAP | 11.0% |
| PEG 400 | 1.0% |
| Ethanol | 66.0% |
| Water | 22.0% |

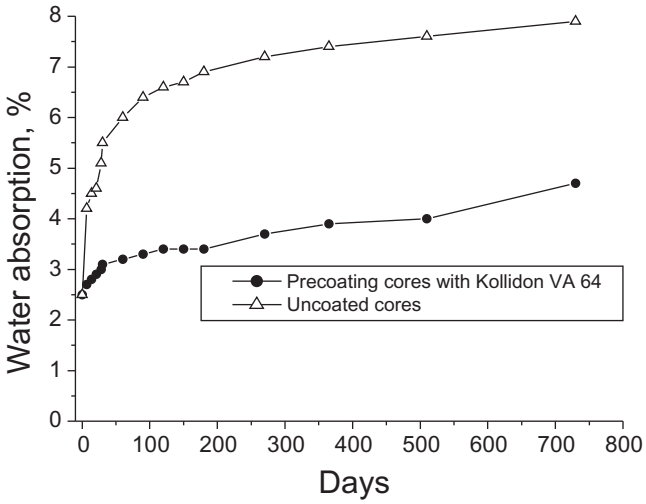


**FIGURE 26** Water absorption of precoating and uncoated cores of combinated dextran–HPMC matrix tables.

coating. The copovidone (i.e., Kollidon VA 64) not only increases the mechanical properties of the tablet (less friability) but also prevents the amount of water absorbed from the air in tropical and subtropical stability conditions (25 and 75% relative humidity).

Figure 26 shows a comparative study of uncoated tablet and sugar-coated tablet after a barrier to water with copovidone 0.5 mg/cm$^2$ of the warm tablet cores using a 10% solution in ethanol. During two years the coating tablets (with initial humidity 2.5%) remained stable and dissolution profiles were similar to the initial time with similarity $f_2$ = 82 observed and friability decreased from 0.25% (uncoated cores) to 0.02% (coated tablets).

Similar results are reported in the literature when Kollidon VA 64 is compared to povidone (Kollidon K 25, K 30, and 90 F). Copovidone absorbs about three times less water than the other soluble Kollidon K 25, 30, and 90F after seven days at 25 °C up to 80% relative humidity [62]. Kollidon VA 64 is manufactured by free-radical polymetization of 6 parts of vinylpyrrolidone and 4 parts of vinyl acetate in 2-propanol. A water-soluble copolymer with a chain structure is obtained. In contrast to the soluble grades of Kollidon, the number 64 is not a *K* value but the mass ratio of the two monomers, vinylpyrrolidone and vinyl acetate. The *K* value of Kollidon VA 64 is of the same order of magnitude as that of Kollidon 30. Synonyms for Kollidon VA 64 are copovidone, copovidonum, copolyvidone, copovidon, and PVP-VAc-copolymer [Eur. Ph., Japanese Pharmaceutical Excipients, and USP National Formulary (NF)] [63].

Copovidone forms soluble films independently of the pH value, regardless of whether it is processed as a solution in water or in organic solvents. He offers better plasticity and elasticity than other povidones. On the other hand, films are also less tacky. Kollidon VA 64 usually absorbs water, and it is seldom used as the sole film-forming agent in a formulation. Normally it is better to combine it with less hygroscopic substances such as cellulose derivates [54], shellac, polyvinyl alcohol (PVA),

PEG (e.g., Macrogol 6000), or sucrose. Others plasticizers such as triethyl citrate, triacetin, or phtalates are not required. The properties of coatings can be improved with combination copovidone–cellulose derivates [64, 65]. Cellulose polymers of high viscosity, such as HPMC 2910, are used in film coating. Spray suspension at 12% HPMC K4M offers values of viscosity above 700 mPa·s and can not be normally used because 250 mPa·s is considered the limit for spraying of a coating suspension. This value of viscosity can be reduced significantly up to 250 mPa·s if 60% HPMC is substituted by Kollidon VA 64 [66] and this leads us to apply this polymer concentration and therefore to economize the spraying procedure.

## 8.3 SEPIFILM

SEPIFILM and SEPISPERSE Dry are ready-to-use, immediate-release, film coating compositions designed for pharmaceutical and nutritional supplement applications. Based on a unique technology, SEPIFILM/SEPISPERSE Dry are granular forms offering the following benefits:

Easy handling: **no dust**
Easy mixing: **no lumps, no foam**
Homogeneous composition: **no segregation**

Most coating compositions are based on hypromellose (HPMC) as film-forming polymer and contains microcrystalline cellulose (MCC) (see Table 17).

SEPPIC was the first company to introduce MCC in coating formulations some 20 years ago. The use of MCC allows higher solid content and enhances the adhesion of the film to the tablet core, consequently improving logo definition. Different formulas are provided by Sepifilm, such as SEPIFILM LP, SEPIFILM 003 and 752, and SEPISPERSE Dry. The last one can be provided with SEPIFILM and Kollicoat IR.

***Moisture Protection of SEPIFILM LP*** Water vapor transmission rates were measured on free films, including titanium dioxide (Figure 27). SEPIFILM LP shows significantly lower moisture permeability compared to regular or PVA-based coating formulations. Removal of titanium dioxide (SEPIFILM LP clear) improves moisture resistance.

Like many other herbal extracts, valerian extract is very hygroscopic. Inclusion into tablets raises stability issues. Coatings reduce moisture absorption but often lead to tablet explosion or visual deterioration (black specs). Tablets formulated with 250 mg valerian extract, spray-dried lactose, and compressible starch were

**TABLE 17 General Formulation for SEPIFILM Coating Film**

| Film-forming agent | **Hypromellose** (former HPMC) |
|---|---|
| Binder | Microcrystalline cellulose |
| Hydrophobic plasticizer | Stearic acid (vegetable origin) |
| Colors | Pigments, lakes |

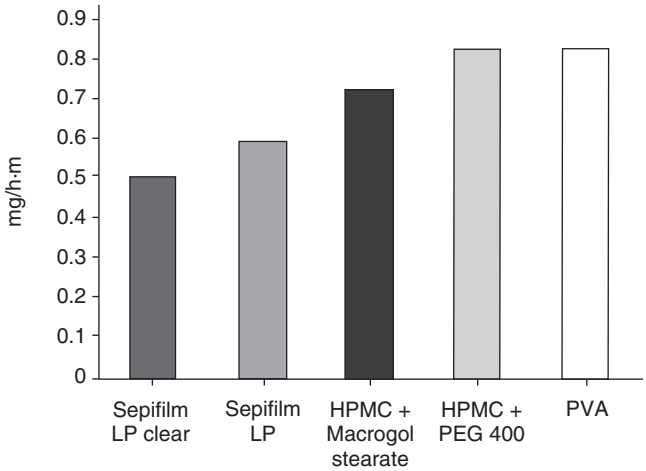


**FIGURE 27** Water vapor transmission rates of free films.

(*a*) (*b*)

**FIGURE 28** Tablets coated with (*a*) conventional hypromellose and (*b*) Sepifilm LP 770 white.

stored at 40 °C and 90% RH (relative humidity). Pictures were taken after one month storage (Figure 28).

***Tetrazepam*** Tetrazepam, a well-known muscular relaxant, undergoes chemical degradation when exposed to moisture and oxygen. 3-Ketotetrazepam is one of the main degradation substances. The amount of 3-ketotetrazepam, measured by HPLC, has been monitored under 25 °C/60 RH and 40 °C/75% RH aging conditions.

SEPIFILM LP 770 significantly improves the stability of tetrazepam. Uncoated tablets and tablets coated with a PVA-based formulation show a strong increase in the amount of degradation substance (Figure 29).

***Dissolution Profile*** SEPIFILM LP can efficiently improve the moisture barrier on moisture-sensitive active pharmaceutical ingredients (API) or hygroscopic cores. The breakthrough in this technology is that SEPIFILM LP does not modify the dissolution profile when compared to conventional coating (Figure 30).

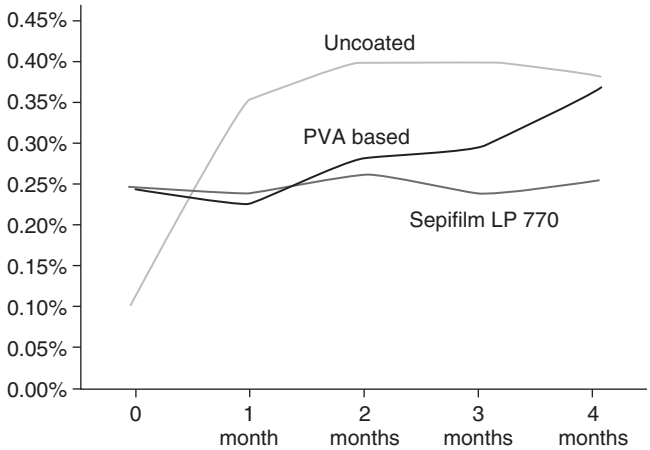


**FIGURE 29** Stability of tetrazepam: uncoated tablets and different tablets coated. Dosage of 3-ketotetrazepam in 50 mg tetrazepam tablets stored at 25 °C and 60% RH leads to partial degradation of tetrazepam at $T_0$. Initial amount of 3-ketotetrazepam is lower in uncoated tablets: cores are likely to absorb water during the aqueous coating process.

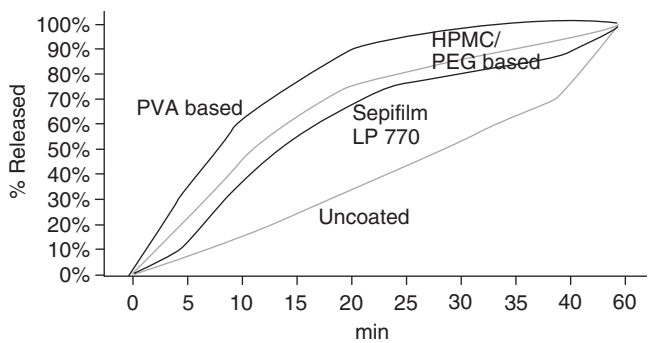


**FIGURE 30** Dissolution profile of tetrazepam, 3 months storage at 40 °C and 75% RH.

***SEPIFILM 003 and 752*** The association of cellulose with a film coating agent was originally patented by SEPPIC. Microcrystalline cellulose is probably one of the most extensively used excipients in pharmaceutical and nutritional products. Unlike other fillers, such as lactose, cellulose is inert, vegetable derived, and accepted worldwide and its shelf life is unlimited. Figure 31 shows the advantage of cellulose microcrystalline in film coating.

***Faster Film Coating Operations*** Cellulose is insoluble and does not increase viscosity. Dispersions with higher solid content can therefore be used and total spraying time is significantly decreased.

The maximum acceptable viscosity of a typical coating dispersion is 500 mPa·s, which corresponds to an 11% hypromellose 6 mPa·s dispersion. SEPIFILM can be dispersed up to 15% in order to reach the same viscosity (Figure 32).

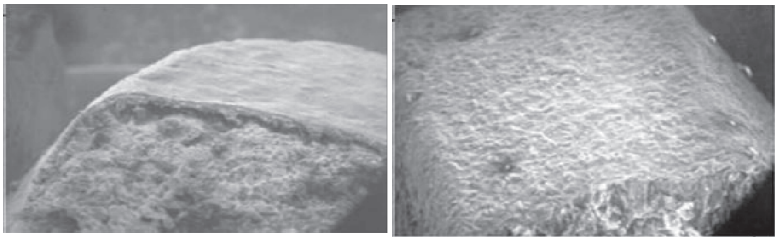

(*a*) (*b*)

**FIGURE 31** (*a*) Tablet coated with HPMC smooth film, medium discontinuity between film and core. (*b*) Tablet coated with SEPIFILM 752 white: clear edge perfectly coated, good adhesion of film to core and continuity between film and core.

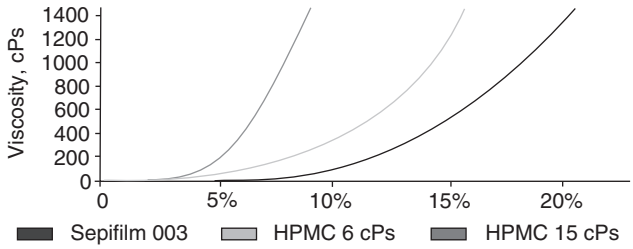


**FIGURE 32** Viscosity for deferments HPMC coating dispersion.

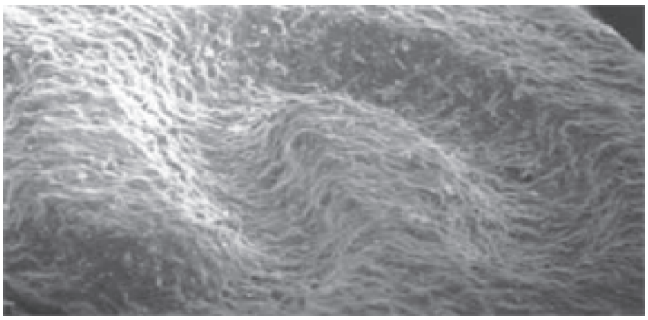

**FIGURE 33** Tablet surface and edges are smoothly covered whereas no logo bridging is observed.

***Enhanced Film Adhesion*** Comparative adhesion values were measured on 3% coated placebos using a modified crushing strength tester: Regular film formulation exhibited an adhesion value of 4 N. Addition of cellulose significantly improved film adhesion as it became impossible to remove film without disrupting tablets or film. Breakage occurred for an applied force of 6 N, which can be considered a minimum adhesion value (Figure 33).

***Ingredient Segregation Avoided*** Powders such as hypromellose and titanium dioxide exhibit dramatically different particle size distribution and density. Such particle heterogeneity may result in constituent segregation inducing film imperfections or color deviations on tablets.

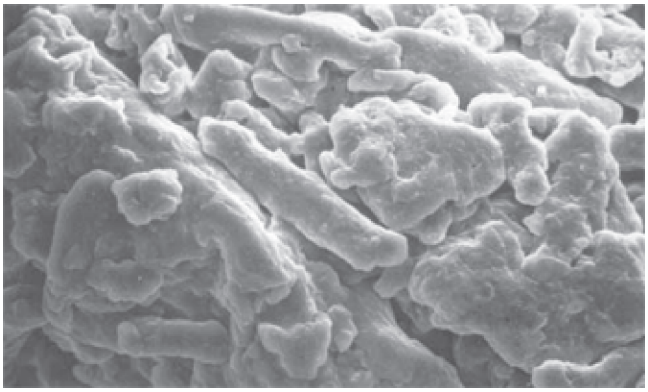

**FIGURE 34** Granules of SEPIFILM 1306 green (×350). Pigments, lakes, and cellulose fibers are thoroughly coprocessed into homogeneous granules.

As seen in (Figure 34), SEPIFILM rules out this weak feature of powdered coating agents as all constituents are closely bound together. Ingredient segregation is avoided and batch-to-batch consistency is therefore guaranteed.

The PVA–PEG graft copolymer Kollicoat IR is the new instant-release, aqueous coating polymer from BASF. Due to its low viscosity and excellent mechanical properties, it permits solid content of up to 30% and leads to fine and smooth tablet film coatings.

In 2005, SEPPIC and BASF Pharma Solutions began collaboration in marketing tablet film coating systems. From this collaboration, a new range of colored SEPIFILM coating systems based on Kollicoat (a registered trademark of BASF Aktiengesellschaft) polymers will be developed to meet individual customer needs. Kollicoat IR has already been approved in Europe as a finished drug in Germany (a reference member state in a mutual recognition procedure). Common uses in pharmaceuticals depend on regulations of each country and definition of uses. Actual regulations of a specific country and/or application should be checked before use.

For pharmaceuticals, this combination is based on 88% of Kollicoat® IR White and 12% of coloring system SEPISPERSE™ Dry.

### *Recommended Equipment*

A **propeller stirrer** is standard equipment even though a deflocculating blade is very efficient.

Deionized or distilled water at room temperature.

The **blade diameter** should be 3 times shorter than the tank's width.

The tank should be 1.5 times higher than wide. The blade should be slightly off center.

The blade should be positioned close to the bottom of the vessel.

**TABLE 18 SEPIFILM LP (with or without SEPISPERSE Dry)**

| | Manesty XL Lab01 (A*) | Manesty Accelacota (B) | IMA-GS HT/M (C) | Driacoater 500 (C) |
|---|---|---|---|---|
| Solid content | LP 770 at 12% | LP 014/Dry at 12% | LP 014/Dry at 11% | LP 014/Dry at 12% |
| Batch size, kg | 5 | 100 | 88.2 | 3 |
| Spraying rate | 15–27 mL/min | 200–300 g/min | 180–200 mL/min | 7–15 g/min |
| Inlet air temperature, °C | 60 | 70–75 | 62 | 55–60 |
| Outlet air temperature, °C | 47 | — | — | 42 |
| Bed temperature, °C | 43 | 40–45 | 36–40 | — |
| Atomizing air/pattern pressure, bars | 2 | 3.5–5 | 2.5 | 3 |
| Airflow, $m^3/h$ | 440 | 1800 | 1000 | 270 |
| Spraying time, min (% weight gain) | 76 (1% w.g.) | 60 (2% w.g.) | 90 (2.2% w.g.) | 70 (3% w.g.) |

### *Mixing Procedure*

*SEPIFILM Product Range*

1. Adjust rotation speed in order to create a vortex.
2. Quickly pour SEPIFILM into the vortex.
3. Reduce mixer speed to avoid drawing air into the liquid (risk of foam formation).
4. Increase speed again as viscosity builds up.

It is not necessary, though recommended, to keep stirring during the coating process.

Kollicoat® White + SEPISPERSE Dry

1. Adjust rotation speed to create a vortex.
2. Quickly pour SEPISPERSE Dry into the vortex.
3. Reduce mixer speed to avoid drawing air into the liquid. Stir for 15 min.
4. Pour Kollicoat IR White into the colored dispersion.
5. Stir an additional 5 min.

Some practical conditions for SEPIFILM and equipment machines are shown in Tables 18–20.

We used SEPIFILM LP in ranitidine core because this is a moisture-sensitive drug and can be a challenge to formulators because of its tendency to hydrolyze when exposed to humidity and/or high temperatures.

Cores of 300 mg that contained 50% ranitidine HCl, cellulose microcrystalline (PH-102) 27.25%, pregelatinized starch 22%, Aerosil 0.5%, and magnesium stearate

**TABLE 19 SEPIFILM Formulations 050 and 752**

| | Driacoater 500 | Glatt Coater 1000 |
|---|---|---|
| Solid content of dispersion | SEPIFILM 050 at 15% | SEPIFILM 752 at 20% |
| Batch size, kg | 3 | 80 |
| Spraying rate, g/min/kg | 7 (1 nozzle) | 3.1 (3 nozzles) |
| Rotation speed, rpm | 10 | 8 |
| Inlet air temperature, °C | 60 | 65 |
| Outlet air temperature, °C | 44 | — |
| Bed temperature, °C | 39–40 | 38–40 |
| Atomizing air/pattern pressure, bars | 3 | 3.5 |
| Airflow, $m^3$/h | 330 | 1800 |
| Spraying time, min (3% weight gain) | 85 | 55 |

**TABLE 20 Kollicoat IR White + SEPISPERSE Dry**

| | Manesty Accelacota 24 in., 300-mg Propranolol HCl Tablets | Manesty Premier 200, Placebo Tablets |
|---|---|---|
| Solid content of dispersion | IR + Dry at 20% | IR + Dry at 20% |
| Batch size, kg | 7 | 180 |
| Spraying rate | 5.3 g/min/kg | 300–375 mL/min |
| Inlet air temperature, °C | 60 | 60 |
| Outlet air temperature, °C | 38–43 | 45 |
| Bed temperature, °C | 39–41 | — |
| Atomizing air/pattern pressure, bars | 2/1 | 2.5/2.5 |
| Airflow, $m^3$/h | 220 | 2200 |
| Spraying time, min (%weight gain) | 37 (3.4% w.g.) | 80 (3% w.g.) |

0.5% were coated to a 3% weight gain similar to case **A*** (see Table 18). Tablet weight, diameter, thickness, hardness, and disintegration times were measured after coating. The film-coated tablets were packaged in bottles with a desiccant. Stability testing was conducted at 40 °C/75% RH for 12 months. Application of film coating (3% wg) resulted in a slight increase in tablet hardness (tablet breaking force 13–14 kp). Tablet disintegration time was not significantly affected by the film coating application (around 13 min) and 100% of the drug was released within 25 min compared to the USP limit of not less than 80% ($Q$) in 45 min. Figure 35 shows the dissolution profiles for the uncoated and coated tablets.

No significant changes were recorded for coated tablets after 12 months of storage for any property measured (Table 21). The stability of this formulation is partly due to the inclusion of pregelatinized starch in the formulation. Starch 1500 acts as a moisture scavenger and retains moisture in its complex structure of glucose polymer chains. A slight increase in tablet hardness was seen after storage (see Figure 36). No significant changes were seen in the disintegration time after storage.

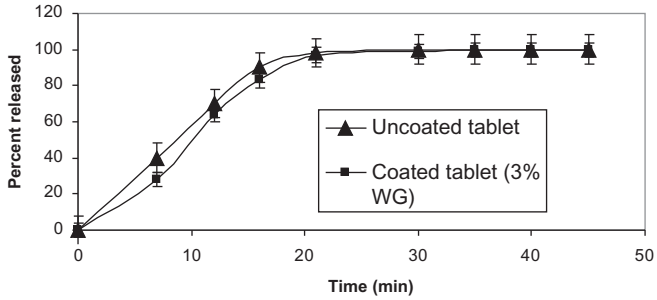


**FIGURE 35** Dissolution profile of coated and uncoated ranitidine HCl tablets.

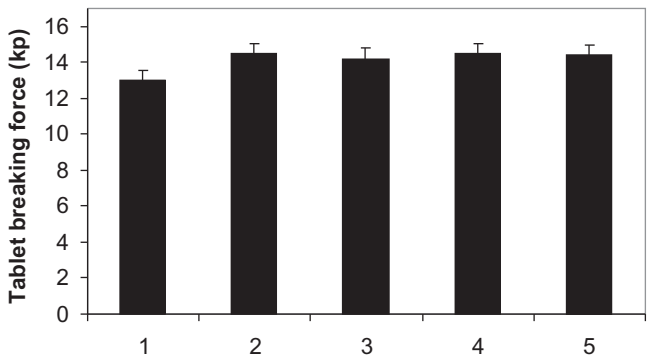


**FIGURE 36** Coated tablet hardness on storage, 40 °C, 75% RH storage conditions.

**TABLE 21 Stability Data Summary: Test**

| Test | USP Limit | Initial | 1 month | 3 months | 6 months | 12 months |
|---|---|---|---|---|---|---|
| Breaking force (kp) | NMT 1.00% | 13.7 | 14.4 | 14.2 | 14.5 | 14.4 |
| Friability, % | | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| Dissolution, T85%, min | NLT 85% in 45 min | 16 | 16 | 16 | 16 | 16 |
| Assay, % | 90–110 | 102 | 100 | 100 | 99 | 99 |
| Impurities, % | NMT 2% | 0.5 | 0.8 | 1.0 | 1.2 | 1.7 |

*Notes:* For 40 °C/75% RH storage conditions. Abbreviations: NMT, no more than; NLT, no less than.

### *Selected Coating Problems and Practical Solutions*

(a) *Defective Coatings* Caused by poor quality of the core or inadequate coating formulations.

(b) *Chipping* Solid content is too high or it is too brittle for want of plasticizer and the film does not adhere properly to the substrate surface (too lipophilic surface or surface lacking in porosity).

*Solution* Revise the formulation for core and coating.

(c) *Blistering* Drying or spraying is performed at high speed, and solvent may be retained in the film. They evaporate on postdrying and may then form blisters in the film.

*Solutions* Lower the inlet air temperature and reduce the spray rate. Check if adhesion of the film and core is adequate.

(d) *Cracks in Film or Along Edges* Caused by too much internal stress, owing to differences in the thermal expansion of film and core or also caused by the swelling of the core during the coating operation.

(e) *Embedded Particles* Particles broken off from the core are embedded in the film during spraying.

*Solutions* The core lacks mechanical stability. Check the formulation of the core.

(f) *Picking* The film surface contains substances that are not molecularly dispersed and start to melt at the core bed temperature of the film coating process. These substances (e.g., PEG, stearic acid) may interfere with the film-forming polymers and produce holes in the film surface [67].

*Solutions* Replace these substances or lower the core bed temperature. Decrease the speed of praying.

(g) *Dull Surfaces* The film will be dull and totally devoid of gloss if a coating process does not produce the requisite smooth surface. This happens if the spray droplets start to dry before reaching the cores and are too viscous to form a smooth film.

*Solutions* (1) Lower the inlet air temperature and reduce the atomizing air quantity or pressure. (2) Add substances that enhance film formation, for example, plasticizer or extra solvents.

(h) *Twinning* The cores permanently stick together.

*Solutions* Decrease the excessive spraying. Check the tablet shape and bands. If the tablets are plane or almost plane, this continues until many of them stick together.

(i) *Bridging* the film fails to follow the contours of the tablet over break lines or engravings and settles in these without adhering to the substrate. The bridging forces in the film exceed the interfacial forces between film and core.

*Solutions* change the tablet surface or add plasticizer.

## 9 DEVELOPMENT OF PHARMACEUTICAL TABLETS USING PERCOLATION THEORY

In 1991, Bonny and Leuenberger [40] explained the changes in dissolution kinetics of a matrix controlled-release system over the whole range of drug loadings on the basis of percolation theory. For this purpose, the tablet was considered a disordered system whose particles are distributed at random. These authors derived a model for the estimation of the drug percolation thresholds from the diffusion behavior.

Knowledge of the percolation thresholds and the related critical points of the system allows a rational optimization of the matrix formulation, avoiding the trial-and-error method usually employed in the pharmaceutical industry. The ideal formulation of an inert matrix, following percolation theory, must be above the drug percolation threshold (i.e., the drug plus the initial pores percolate the system). This fact guarantees the release of the total drug dose. On the other hand, the matrix must also contain an infinite cluster of excipient (i.e., the excipient must also be above its percolation threshold). This percolating cluster of excipient avoids the disintegration of the matrix during the release process and controls the drug release [43].

This kind of system, containing percolating clusters of both drug and excipient, is called a bicoherent system. Furthermore, in order to decrease the variability in the biopharmaceutical and mechanical behavior of the matrices, due to little change in the tablet composition, it is not convenient to formulate the matrices just at the percolation threshold. In this way, knowledge of the percolation thresholds of drug and excipient supposes an important decrease in the cost of the optimization process as well as in the time to market. The percolation thresholds of different pharmaceutical powders have already been estimated, including drugs such as morphine hydrochloride [68], naltrexone hydrochloride [69], dextromethorphan hydrobromide [70], and lobenzarit disodium [71] as well as matrix-forming excipients such as hydrogenated castor oil, ethylcellulose, and acrylic polymers.

### 9.1 Case Study: Optimization of Inert Matrix Tablets for Controlled Release of Dextromethorphan Hydrobromide

The objective of this work was to estimate the percolation thresholds of dextromethorphan hydrobromide and Eudragit RS-PM which characterize the release behavior of these inert matrices in order to rationalize the design of these controlled-release systems.

Dextromethorphan hydrobromide is an antitussive drug with no analgesic or addictive action. Its antitussive effect is similar to codeine. The recommended oral dose for adults is 10–30 mg three to six times a day, not to exceed 120 mg daily. It is absorbed rapidly and completely when taken orally with a lag time of 15–30 min [72].

In order to estimate the percolation threshold of dextromethorphan hydrobromide, the matrices were studied from different points of view:

1. *Release Profiles and Release Kinetics* Figure 37 shows the percentage of drug released from the studied matrices. As can be appreciated, a very similar behavior was observed for matrices containing up to 50% w/w of drug. This can be attributed to the swelling process (approximately 11% v/v) undergone by the matrices during the release assay. This process makes the influence of the percolation threshold on the release profiles less evident.

Higuchi's kinetic model and Peppas' nonlinear regression ($Q = a' + b't^k$) were employed to study the release data. The results obtained are shown in Table 22. As can be seen, the exponent $k$ underwent a change (0.4534–0.5472) between matrices containing 20 and 30% w/w of drug. Even if this is not an important change, it may be related to some changes in the matrix structure due to the drug percolation

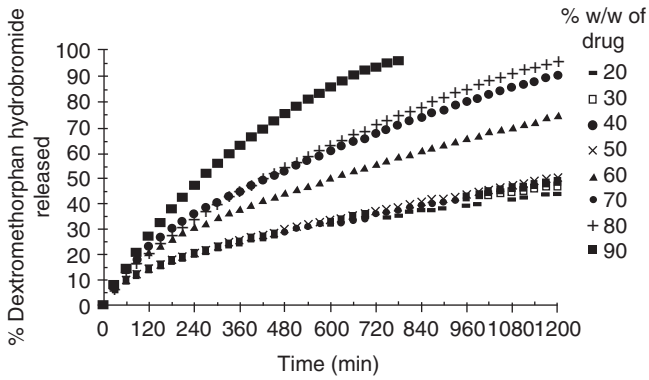


**FIGURE 37** Percentage of drug released vs. time for tablets prepared with different loadings of dextromethorphan hydrobromide.

**TABLE 22 Dissolution Data from Dextromethorphan–HBr/Eudragit RS-PM Matrices**

| Drug Load (% w/w) | $n$ | $Q-a+b\sqrt{t}$ | $Q = a^r + b^r t^k$ | | |
|---|---|---|---|---|---|
| | | $b \pm$ S.E. | $r$ | $k$ | $r$ |
| 20 | 248 | $1.43 \times 10^{-4} \pm 2.9 \times 10^{-7}$ | 0.999 | 0.45335 | 0.999 |
| 30 | 248 | $2.47 \times 10^{-4} \pm 4.7 \times 10^{-7}$ | 0.999 | 0.54724 | 0.999 |
| 40 | 248 | $3.38 \times 10^{-4} \pm 1.0 \times 1.0^{-6}$ | 0.998 | 0.58981 | 0.999 |
| 50 | 248 | $4.43 \times 10^{-4} \pm 2.0 \times 1.0^{-6}$ | 0.998 | 0.59792 | 0.999 |
| 65 | 241 | $8.60 \times 10^{-4} \pm 2.0 \times 10^{-6}$ | 0.999 | 0.54536 | 0.999 |
| 70 | 248 | $1.15 \times 10^{-3} \pm 2.0 \times 10^{-6}$ | 0.999 | 0.54697 | 0.999 |
| 80 | 241 | $1.45 \times 10^{-3} \pm 5.0 \times 10^{-6}$ | 0.999 | 0.57677 | 0.999 |
| 90 | 85 | $2.09 \times 10^{-3} \pm 2.9 \times 10^{-5}$ | 0.992 | 0.71910 | 0.999 |

threshold. The masking effect of the swelling process on the drug percolation threshold has to be taken into account.

2. *Estimation of Drug Percolation Threshold* The drug percolation threshold was calculated using the property β described by Bonny and Leuenberger [40]. This property is defined by the equation

$$\beta = \frac{b}{\sqrt{2A - \varepsilon C_s}} \tag{5}$$

where β is proportional to the square root of the effective diffusion coefficient $D_{eff}$, which is expected to obey, in the nearby of the percolation threshold, the scaling law

$$D_{eff} = kD_0(\varepsilon - \varepsilon_c)^{\mu}$$

where $D_0$ is the diffusion coefficient of the drug in pure solvent, $k$ a constant, ε the total porosity of the matrix (sum of initial porosity and porosity due to the dissolution of the drug), $\varepsilon_c$ the critical porosity or percolation threshold, and μ the critical

**TABLE 23 Calculation of Tablet Property β and Related Parameters in Matrices of Dextromethorphan Hydrobromide**

| Drug (% w/w) | $\varepsilon_0$ | $\varepsilon$ | $n^b$ | $F^b$ | Probability$^b$ | $A$ | $\beta \times 10^3$ |
|---|---|---|---|---|---|---|---|
| 20 | 0.145 | 0.300 | 248 | 248,833 | $9.9 \times 10^{-16}$ | 0.216 | 0.218 |
| 30 | 0.142 | 0.378 | 248 | 277,630 | $9.9 \times 10^{-16}$ | 0.329 | 0.305 |
| 40 | 0.129 | 0.453 | 248 | 88,245.6 | $9.9 \times 10^{-16}$ | 0.450 | 0.357 |
| 50 | 0.115 | 0.531 | 248 | 79,555.4 | $9.9 \times 10^{-16}$ | 0.579 | 0.413 |
| 65 | 0.091 | 0.657 | 241 | 126,817 | $9.9 \times 10^{-16}$ | 0.787 | 0.687 |
| 70 | 0.093 | 0.705 | 248 | 278,000 | $9.9 \times 10^{-16}$ | 0.851 | 0.887 |
| 80 | 0.083 | 0.799 | 241 | 97,483.4 | $9.9 \times 10^{-16}$ | 0.995 | 1.033 |
| 90 | 0.083 | 0.898 | 85 | 5,324.51 | $9.9 \times 10^{-16}$ | 1.133 | 1.395 |

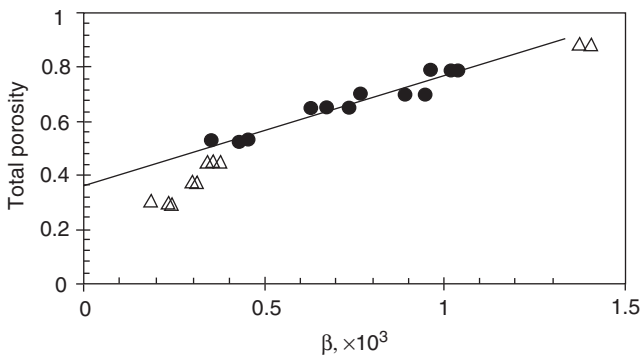


**FIGURE 38** Determination of drug percolation threshold. The circles represent the values selected for the regression, according to its linear behavior. These values correspond to drug loads between 50 and 80% w/w (three tablets per lot). Each point represents one experimental datum.

exponent for conductivity. This exponent has a value of 2.0 in 3D systems. The values of β as well as the parameters involved in its calculation are shown in Table 23.

The percolation threshold of dextromethorphan hydrobromide was estimated as the intersection with the $Y$ axis from a linear regression of the total porosity, ε, versus the property β (Figure 38). Following the method of Bonny and Leuenberger, only the β values above $p_{c1}$ showing a linear dependence on the total porosity (circles in Figure 38) are considered in the regression. The selected β values corresponded to matrices with 50–80% w/w of drug.

The estimated critical porosity is 0.3691 ± 0.0541, considering a 95% confidence interval ($P$ = 0.05). This range corresponds to a dextromethorphan hydrobromide content of between 23 and 36% w/w.

Estimation of the percolation threshold by visual methods is not very accurate, mainly due to extrapolation from 2D to 3D systems. Nevertheless, scanning electron microscopy was employed as an auxiliary technique in order to investigate the distribution of the particles of dextromethorphan hydrobromide in the matrices.

Figure 39 shows two scanning electron microscopy (SEM) micrographs corresponding to the tablet side facing the lower punch for matrices containing 20 and

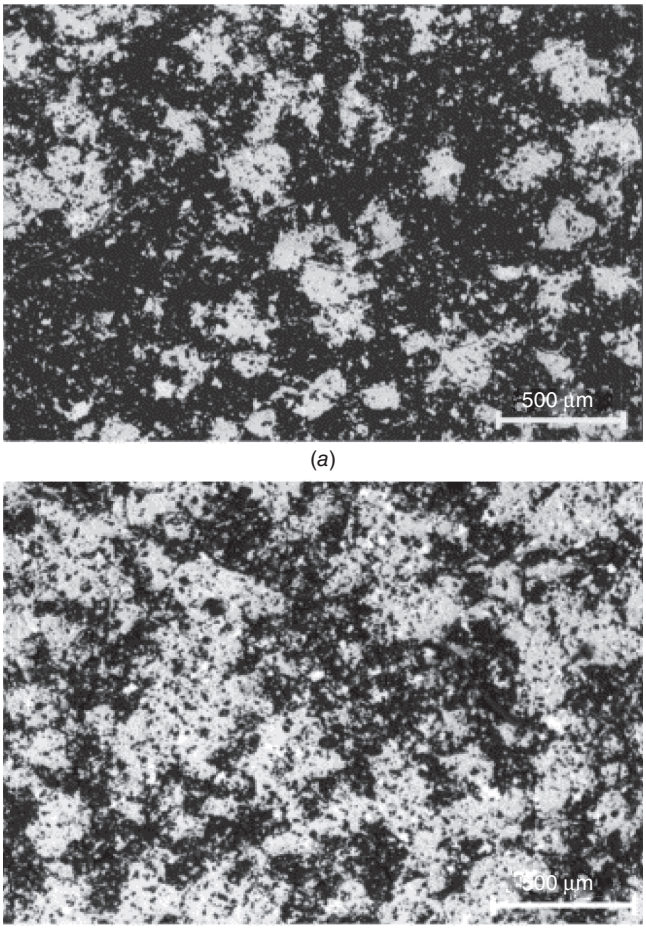


**FIGURE 39** SEM micrographs corresponding to bottom side of matrices using BSE detector. The light gray particles correspond to dextromethorphan–HBr and the dark gray particles to the excipient Eudragit RS-PM. (*a*) Matrices containing 20% w/w of drug. (*b*) Matrices containing 30% w/w of drug.

30% w/w of drug using backscattering electron (BSE) detector at the same magnification. In the tablet containing 30% of drug (Figure 39*b*), an infinite drug cluster can be observed. The drug particles (light-gray particles) begin to form a connective network from the left to the right and from the top to the bottom of the micrograph. In the tablet containing 20% w/w of drug (Figure 39*a*), the particles of the drug (light-gray particles) seem to form isolated groups in the matrix.

Therefore, considering both micrographs in Figure 39, a 2D geometric phase transition can be observed. Figure 39*a* shows the drug as gray particles on a black background, whereas in Figure 39*b* there is a black-on-gray array, with black particles corresponding to Eudragit RS-PM surrounded by a gray background formed by dextromethorphan hydrobromide particles.

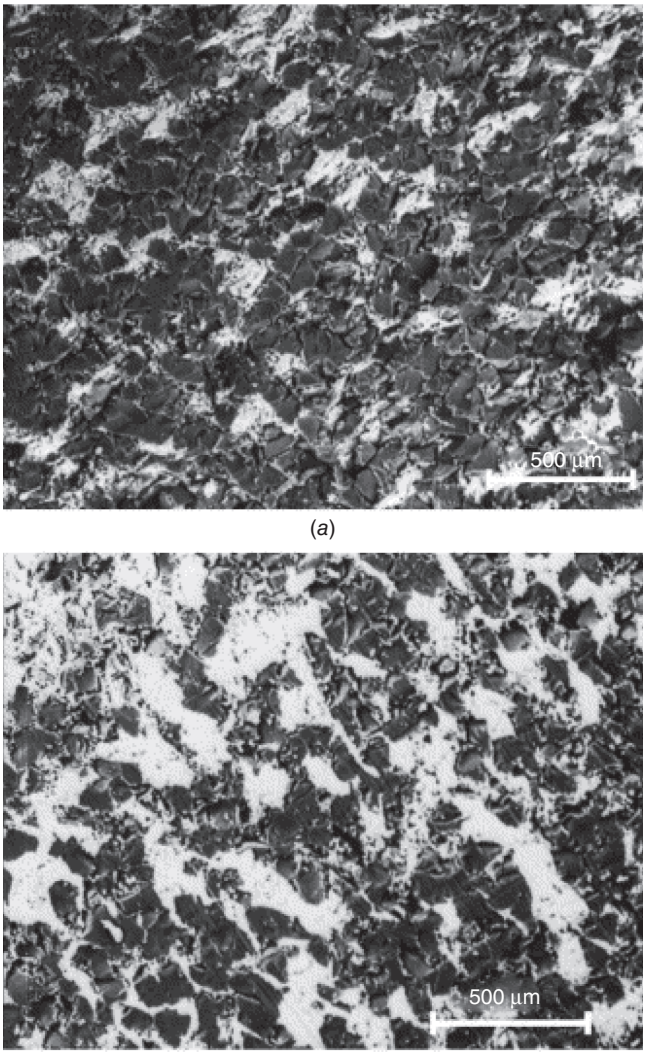


**FIGURE 40** SEM micrographs corresponding to cross section of tablets using BSE detector: (*a*) matrices containing 20% w/w dextromethorphan–HBr (light gray particles); (*b*) matrices containing 30% w/w of drug.

When the cross section of these matrices (20 and 30% w/w drug loading) was observed (Figure 40), the same pattern was found, changing from gray on black (Figure 40*a*, 20% w/w of drug) to black on gray (Figure 40*b*, 30% w/w of drug).

Therefore, according to the different methods employed, the drug percolation threshold in the studied matrices is expected to be between 20 and 30% w/w of dextromethorphan hydrobromide (total porosity between 30.0 and 37.8% v/v of drug).

3. *Estimation of Excipient Percolation Threshold.* In principle, for binary pharmaceutical systems, two percolation thresholds are expected: the drug percolation

threshold $p_{c1}$ and the excipient percolation threshold $p_{c2}$. The second is the point where the excipient ceases to percolate the system.

Nevertheless, in a previous study dealing with inert matrices of naltrexone–HCl [74], two different excipient percolation thresholds $p_{c2}$ were found for the matrix-forming excipient Eudragit RS-PM: the site percolation threshold related to a change in the release kinetics and the site-bond percolation threshold derived from the mechanical properties of the tablet, where the excipient failed to maintain tablet integrity after the release assay.

An evident change in the release kinetics between tablets containing 80 and 90% w/w of drug can be observed in Table 23 (from $k \approx 0.57$ to $k \approx 0.7$ in the Peppas equation). Therefore, the site percolation threshold of the excipient can be estimated between the matrices containing 80 and 90% w/w of dextromethorphan hydrobromide (10–20% v/v of excipient). Above this threshold, a percolating cluster of excipient particles exists. These particles are able to control the drug release kinetics, but their cohesion forces can be insufficient to maintain tablet integrity after the release assay.

Formulations containing more than 65% w/w of drug were unable to maintain tablet integrity after the 20-h release assay. According to this result, the site-bond percolation threshold of the excipient ranges between 65 and 70% w/w of drug, corresponding to 29.5 and 34% v/v of excipient. Above this percolation threshold, that is, for concentrations of excipient >34% v/v, there is a percolating cluster of excipient particles bound by sufficient forces to maintain tablet integrity after drug release.

In conclusion, according to percolation theory, the studied matrices should be formulated with drug content between 30 and 65% w/w (37.8–66% v/v of total porosity). These concentrations are optimal to ensure release of the total drug dose, to have controlled release of the drug, and to avoid disintegration of the matrix. In order to increase the robustness of the formulation, the limits of this range should be avoided.

### 9.2 Critical Points of Hydrophilic Matrix Tablets

Recently percolation theory is starting to be applied to the study of hydrophilic matrix systems. Figure 41 shows an example of the changes observed in several release parameters employed to estimate the critical point and the related percolation threshold in hydrophilic matrices prepared using KCl as the model drug [73].

Application of the percolation theory allows explanation of the changes in the release and hydration kinetics of swellable matrix-type controlled delivery systems. According to this theory, the critical points observed in dissolution and water uptake studies can be attributed to the excipient percolation threshold. Knowledge of these thresholds is important in order to optimize the design of swellable matrix tablets. Above the excipient percolation threshold an infinite cluster of this component is formed which is able to control the hydration and release rate. Below this threshold the excipient does not percolate the system and drug release is not controlled.

Miranda et al. demonstrated experimentally the influence of the particle size of the components on the percolation threshold in hydrophilic matrices as well as the importance of the initial porosity in the formation of the gel layer (sample-spanning cluster of excipient) [74].

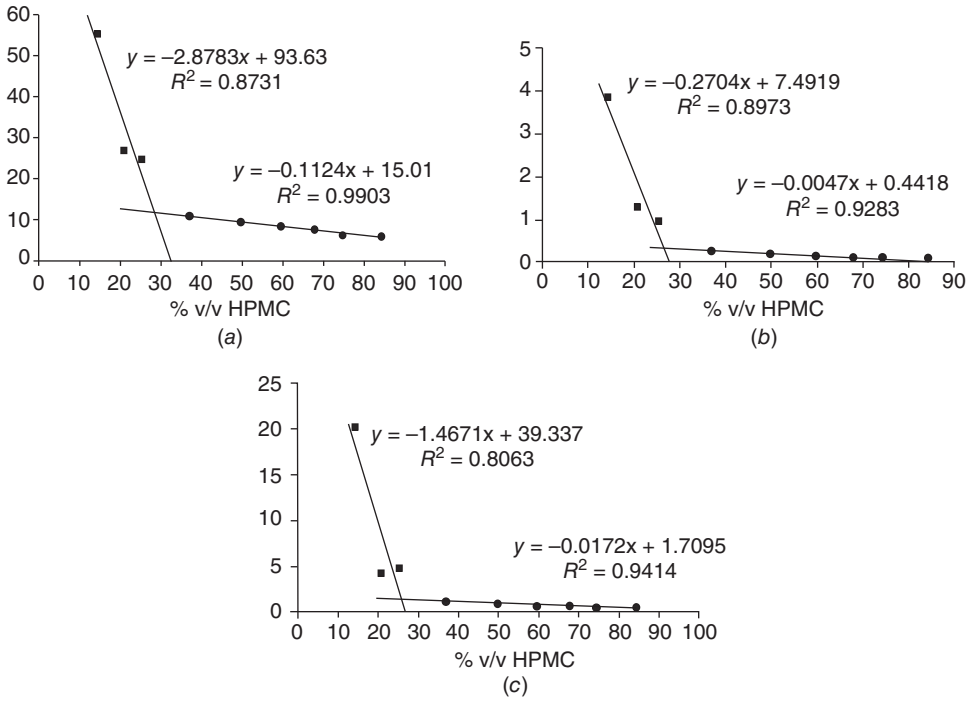


**FIGURE 41** (*a*) Higuchi slope; (*b*) normalized Higuchi slope; (*c*) relaxational constant of Peppas and Sahlin versus percentage of excipient volumetric fraction for batch A (50–100 μm KCl and 150–200 μm HPMC K4M).

## 9.3 Case Study: Estimation of Percolation Thresholds in Acyclovir Hydrophilic Matrix Tablets

The principles of the percolation theory were applied to design controlled-release matrix tablets containing acyclovir in order to estimate the percolation threshold of the excipient in acyclovir matrix tablets and to characterize the release behavior of these hydrophilic matrices in order to rationalize the design of these controlled-release systems.

Acyclovir is a potent inhibitory of viruses of the herpes group, particularly herpes simplex virus (HSV I and II) and herpes zoster varicella virus. Unfortunately, acyclovir has a short half-life (2–3 h), and the oral dosage form must be taken five times daily, which is very inconvenient for patients [75, 76]. Consequently, the aim of this study was to develop a controlled-release formulation of acyclovir that could be taken twice daily. The materials used to prepare the tablets were acyclovir (Kern Pharma, Tarrasa, Barcelona) and hydroxypropyl methylcellulose (Methocel K4M) (Colorcon) a hydrophilic cellulose derivative as the matrix-forming material.

Binary mixtures were prepared with varying drug contents (60, 70, 80, 90, and 95%) keeping constant the drug and excipient particle size. Table 24 gives the composition of the studied batches as well as the tablet thicknesses. The mixtures were compressed on an eccentric machine (Bonals A-300) without any further excipient. Cylindrical tablets with a mean dosage of 500 mg and a diameter of 12 mm were prepared at the maximum compression force accepted by the formulations.

**TABLE 24 Composition of Hydrophilic Matrices Prepared with Acyclovir/HPMC K4M (150–200 μm) and Percent HPMC plus Initial Porosity**

| Batch | Percent w/w | | Percent (v/v) HPMC + Initial Porosity | Tablet Thickness (mm) |
|---|---|---|---|---|
| | Acyclovir | HPMC K4M | | |
| A | 95 | 5 | 20.76 | 3.01 ± 0.056 |
| B | 90 | 10 | 26.41 | 3.25 ± 0.052 |
| C | 80 | 20 | 37.60 | 3.84 ± 0.051 |
| D | 70 | 30 | 45.11 | 4.68 ± 0.063 |
| E | 60 | 40 | 55.82 | 5.81 ± 0.056 |

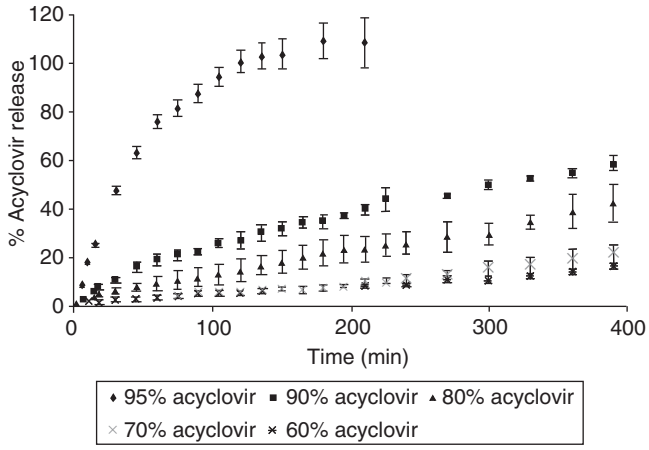


**FIGURE 42** Acyclovir release from matrix tablet with total drug content of 95, 90, 80, 70, and 60 prepared with acyclovir–HPMC K4M (150–200 μm) (mean ± SD, $n$ = 3).

The release profiles were measured with the USP 25 dissolution apparatus 2 (Turu Grau, model D-6) at 100 rpm in distilled water (900 mL) at 37 ± 0.5 °C for 12 h. Filtered samples taken at different times were determined for acyclovir content through ultraviolet absorption at $\lambda_{max}$ (242 nm).

Figure 42 shows the release profiles obtained from hydrophilic matrices formulated with acyclovir and HPMC K4M 150–200 μm.

In order to study the release mechanism of acyclovir from the tablets, the fitting of the drug release data to the following kinetic equations has been studied: zero-order equation, $Q = k_0 t$; Higuchi equation [77], $Q = k_H t^{1/2}$; Korsmeyer–Peppas equation [78], $Q = kt^n$; and Peppas-Sahlin equation [79], $Q = k_d^m + k_r^{2m}$, where $Q$ is the amount of drug remaining at time $t$, $k_0$ is the zero-order release constant, $k_H$ is the Higuchi rate constant, $k$ is the Korsmeyer–Peppas kinetic constant, $n$ is the exponent indicative of the release mechanism (for matrix tablets an $n$ value of 0.5 indicates diffusion control and an $n$ value of 1.0 indicates erosion or relaxation control [80], intermediate values suggest that at least two processes contribute to the overall release mechanism), $k_d$ is the diffusion rate constant, $k_r$ is the relaxation rate constant, and $m$ is the purely Fickian diffusion exponent for a device of any geometric shape which exhibits controlled release. In our case, the aspect ratios and exponent values ($m$) are shown in Table 25 [79]. The results obtained are shown in Table 26.

**TABLE 25 Aspect Ratios and Exponent Values (m) for Hydrophilic Matrices Studied**

| Batch | Aspect Ratio | Exponent (m) |
|---|---|---|
| A | 3.80 | 0.45 |
| B | 3.59 | 0.44 |
| C | 3.03 | 0.43 |
| D | 2.65 | 0.42 |
| E | 2.16 | 0.43 |

**TABLE 26 Values of Kinetic constants Derived with Selected Equations in Range 5–70% Acyclovir Release for All Batches Studied**

| **Batch** | | **A** | **B** | **C** | **D** | **E** |
|---|---|---|---|---|---|---|
| **Acyclovir, % w/w** | | **95** | **90** | **80** | **70** | **60** |
| Zero-order equation | $k_0$ | 1.222 | 0.122 | 0.096 | 0.057 | 0.042 |
| | $r^2$ | 0.984 | 0.974 | 0.994 | 0.995 | 0.987 |
| | Sum of squares total | 3545.9 | 10,352.6 | 6,531.4 | 5,149.2 | 1,268.0 |
| | Sum of squares residual | 53.7 | 278.3 | 51.8 | 24.8 | 16.4 |
| Higuchi equation | $k_H$ | 12.440 | 3.4167 | 2.8518 | 1.7683 | 1.4694 |
| | $r^2$ | 0.998 | 0.993 | 0.956 | 0.932 | 0.959 |
| | Sum of squares total | 3,545.9 | 10,352.6 | 12,367.8 | 5,149.2 | 1,268.0 |
| | Sum of squares residual | 7.0 | 66.9 | 483.6 | 362.4 | 54.6 |
| Korsmeyer–Peppas equation | $k_H$ | 3.167 | 0.290 | 0.254 | 0.027 | 0.041 |
| | $n$ | 0.782 | 0.843 | 0.856 | 1.114 | 1.008 |
| | $r^2$ | 0.999 | 0.994 | 0.998 | 0.998 | 0.997 |
| | Sum of squares total | 13,220.9 | 26,297.6 | 33,309.1 | 13,879.4 | 5,673.5 |
| | Sum of squares residual | 16.7 | 159.5 | 65.5 | 10.1 | 15.2 |
| Peppas and Sahlin equation | $k_d$ | 2.239 | 2.056 | 0.357 | –0.81 | –0.38 |
| | $k_r$ | 1.615 | 0.161 | 0.205 | 0.202 | 0.127 |
| | $r^2$ | 0.998 | 0.999 | 0.994 | 0.997 | 0.997 |
| | Sum of squares total | 49,567.0 | 13,220.9 | 28,716.1 | 13,879.4 | 5,673.5 |
| | Sum of squares residual | 42.9 | 23.3 | 165.0 | 43.7 | 15.2 |

*Notes:* $k_0$ (%/min), zero-order constant; $k_H$ (%/min$^{1/2}$), Higuchi's slope; $k$ (%/min$^n$), kinetic constant of Korsmeyer model; $n$, diffusional exponent; $k_d$ (%/min$^m$), diffusional constant of Peppas and Sahlin model; $k_r$ (%/min$^{2m}$), relaxational constant of Peppas and Sahlin model; $m$, diffusional exponent that depends on geometric shape of releasing device through its aspect ratio (see Table 25).

The analysis of the release profiles and the kinetic data indicate two different behaviors and a sudden change between them. In the first behavior, which corresponds to the matrices that release the drug at slow rates, the release was controlled by the fully hydrated gel layer. For these matrices, erosion of the hydrophilic gel structure has shown an important influence on drug release. This is indicated by the better fit of the drug release kinetics to the zero-order equation, the $n$ value of

Korsmeyer–Peppas equation near 1, and the higher value of the relaxation constant $k_r$ in comparison with the diffusion constant $k_d$ in the Peppas–Sahlin equation. Taking into account the drug solubility (2.5 mg/ML), prevalence of the erosion versus swelling mechanism can be expected. After the transition point, the tablets allow the free dissolution of the drug when they are exposed to the dissolution medium due to the fact that the gel layer is not established since the first moment and, in these conditions, this structure cannot control the drug release. The Korsmeyer release rate increases from 0.290 to 3.167% $min^{-1/2}$. For these matrices, according to the Higuchi ($r^2$ = 0.998), Korsmeyer ($n$ = 0.782), and Peppas–Sahlin ($k_r < k_d$) equations, drug release is governed by the diffusion process.

In hydrophilic matrices the drug threshold is less evident than the excipient threshold, which is responsible for the release control [73]. In order to estimate the percolation threshold of HPMC K4M, different kinetic parameters were studied: Higuchi rate constant, normalized Higuchi rate constant, and relaxation rate constant. The evolution of these release parameters has been studied as a function of the sum of the excipient volumetric percentage plus initial porosity. Recent studies of our research group have found the existence of a sample-spanning cluster of excipient plus pores in the hydrophilic matrix before the matrix is placed in contact with the liquid, clearly influences the release kinetics of the drug [73].

Figures 42–45 show changes in the different kinetic parameters: the Higuchi rate constant, normalized Higuchi rate constant, and relaxation rate constant. To estimate the excipient percolation threshold, these parameters were plotted versus the excipient volumetric fraction plus initial porosity.

The kinetic parameters studied show a nonlinear behavior as a function of the volumetric fraction of the excipient plus initial porosity. As an approximation for estimating the trend of the parameter, one regression line has been performed below and the other above the percolation threshold. The point of intersection between both regression lines has been taken as an estimation of the percolation threshold [73, 74].

As percolation theory predicts, the studied properties show a critical behavior as a function of the volumetric fraction of the components. A critical point has been found between 21 and 26% v/v of excipient plus initial porosity (see Table 24). This critical point can be attributed to the excipient percolation threshold.

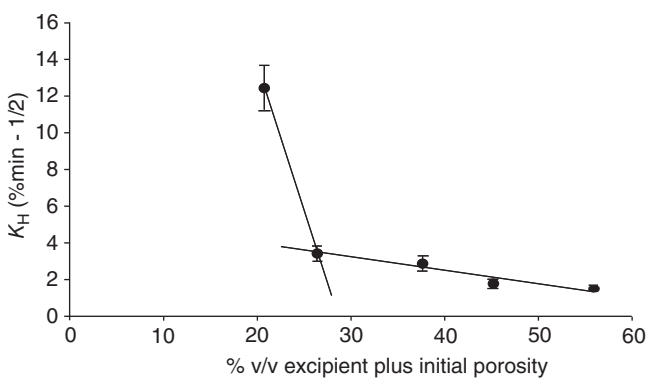


**FIGURE 43** Higuchi slope (mean ± SD, $n$ = 3) versus percentage of excipient volumetric fraction plus initial porosity for all batches studied.

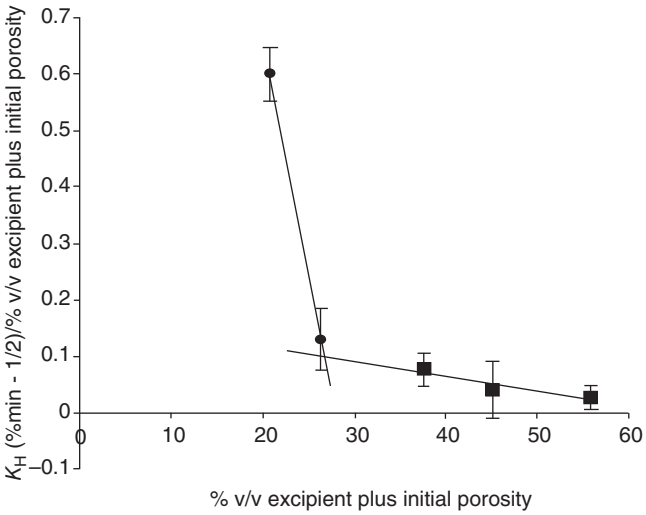


**FIGURE 44** Normalized Higuchi slope (mean ± SD, $n = 3$) versus percentage of excipient volumetric fraction plus initial porosity for all batches studied.

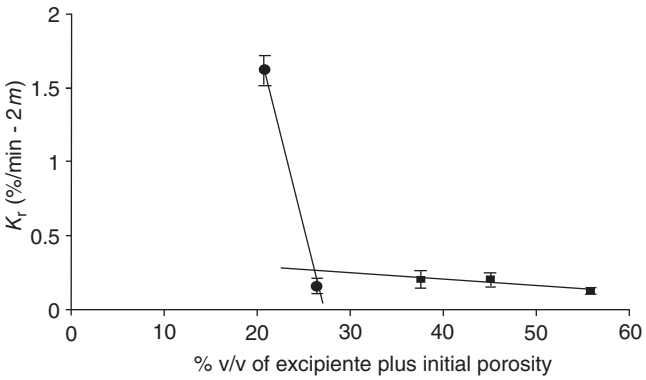


**FIGURE 45** Relaxational constant of Peppas–Sahlin (mean ± SD, $n = 3$) versus percentage of excipient volumetric fraction plus initial porosity for all batches studied.

The Effective Medium Approximation (EMA), based in some assumptions, allows us to employ linear regressions as an approximation of the behavior of a disordered system outside the critical range. Based on EMA theory, two linear regressions have been performed as an approximation for estimating the percolation threshold as the point of intersection between both regression lines (see Figures 43–45). The values of the excipient percolation thresholds estimated for all the batches studied, based on the behavior of the kinetic parameters, ranged from 25.99 to 26.77%.

Therefore, the results obtained from the kinetics analysis are in agreement with the release profiles, indicating a clear change in the release rate and mechanism between matrices containing 90 and 95% w/w of drug (5–10% w/w of excipient). The existence of a critical point can be attributed to the excipient percolation threshold. From the point of view of percolation theory, this means that above 10% w/w of HPMC K4M, the existence of a network of HPMC (able to form a hydrated layer from the first moment) controls the drug release.

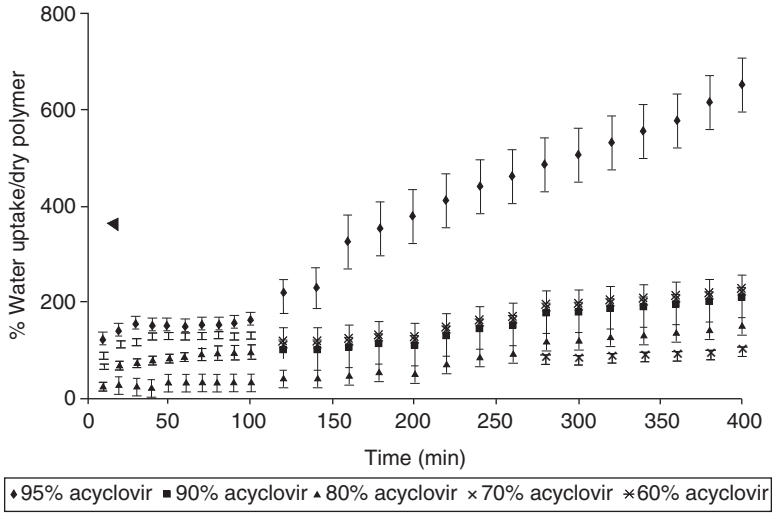


**FIGURE 46** Weight gain of systems as function of swelling time for matrix tablet with total drug content of 95, 90, 80, 70, and 60% prepared with acyclovir–HPMC K4M (150–200 μm) (mean ± SE, $n = 3$).

The process of water penetration into hydrophilic matrix tablets was also studied using a modified Enslin apparatus. This apparatus contains a fritter and a system to regulate the water level. When the tablet is placed on the fritter, the water is absorbed from a reservoir which is placed on a precision balance. The amount of water uptake at each time point was read from the balance as weight loss in the reservoir. Figure 46 shows the obtained release profiles.

An increase in the rate of water uptake can be observed when the HPMC concentration decreases. A critical point was found between 90 and 95% w/w of acyclovir. This range corresponds with the critical point observed in release profile studies. The water uptake data were subjected to the Davidson and Peppas model to calculate the rate of water penetration [81]. The results show a change in the water uptake constant between the matrices containing 90–95% w/w of acyclovir, which reflects the presence of the critical point previously observed.

Knowledge of the percolation threshold of the components of the matrix formulations contributes to improve their design. First, in order to develop robust formulations, that is, to reduce variability problems when they are prepared at industrial scale, it is important to know the concentrations corresponding to the percolation thresholds. The percolation thresholds correspond to formulations showing a high variability in their properties as a function of the volume fraction of their components. Therefore, in order to design robust dosage forms, the nearby of the percolation thresholds should be avoided.

Second, the excipient percolation threshold in hydrophilic matrices represents the border between a fast release of the drug (below the threshold) and a drug release controlled by the formation of a coherent gel layer (above the excipient percolation threshold). Therefore, knowledge of this threshold will allow us to avoid the preparation of a number of unnecessary lots during the development of a pharmaceutical formulation, resulting in a reduction of the time to market.

## 10 ULTRASOUND-ASSISTED TABLETING (A NEW PERSPECTIVE)

The compression of a powder is a complex process that is usually affected by different kinds of problems. These problems have been widely investigated and mainly concern the volume reduction and the development of a strength between the particles of the powder sufficient to ensure tablet integrity [82]. The application of ultrasonic energy shows a great ability to reduce and even avoid these problems [83]. Ultrasound refers to mechanical waves with a frequency above 18 kHz (the approximate limit of the human ear). In an ultrasound compression machine, this vibration is obtained by means of a piezoelectric material (typically ceramics) that acts as a transducer of alternate electric energy of different frequencies in mechanical energy. An acoustic coupler, or "booster," in contact with the transducer increases the amplitude of the vibration before it is transmitted (usually in combination with mechanical pressure) to the material to be compressed.

Ultrasound-assisted powder compression has been widely employed in metallurgy as well as in the plastic and ceramic industries [84]. The first references in the pharmaceutical industry are two patents in 1993 [85] and 1994 [86]. Since then, some papers have presented experimental data in this field [45, 87–92].

Two main objectives are pursued nowadays by means of the application of ultrasound-assisted compression:

1. Increase in the drug dissolution rate due to amorphization of the drug
2. Preparation of controlled-release dosage forms with thermoplastic excipients

As a consequence of the application of ultrasonic energy, the drug can lose its crystalline structure. This will result in an increase of the dissolution rate of the active substance, which can be very adequate for slowly dissolving drugs. Nevertheless, depending on the storage conditions, the drug can recover, at least partially, its crystallinity [89, 90].

To overcome this problem, it has been proposed to use an adequate excipient, preventing the recovery of the crystallinity, leading in some cases to the preparation of solid solutions into the die of the tableting machine.

Several analytical techniques, such as infrared (IR) spectroscopy, differential scanning calorimetry, HPLC, and thin-layer chromatography (TLC), have been used to investigate possible drug degradation due to ultrasonic energy. No important permanent modification of the drug has been found, with the exception of the loss of crystallinity [89, 90].

Concerning the design of controlled-release dosage forms, using a thermoplastic excipient (e.g., copolymers of acrylic and metacrylic acid), an important decrease in the release rate has been found for tablets compressed with the assistance of ultrasonic energy in comparison with traditional tablets.

Although the effects of the ultrasonic energy on the material are not completely clarified, this slow release rate has been attributed to different phenomena:

*Mechanical Pressure* This pressure is exerted by the punches of the ultrasound-assisted tableting machine. This is the main compression mechanism when low ultrasonic energies are employed (below 25 J in the mixtures studied by

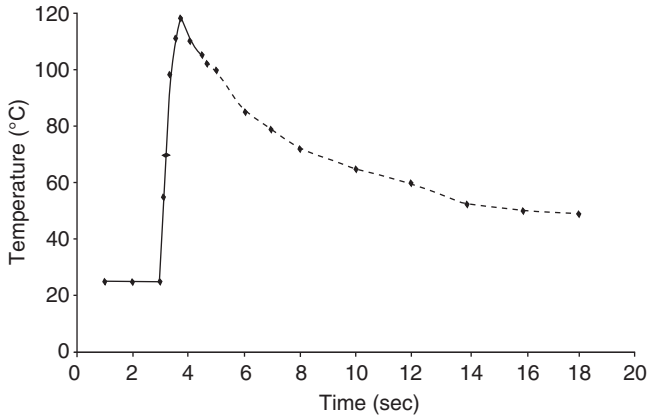


**FIGURE 47** Temperature profile inside compression chamber. (Courtesy of Tecnea Srl.)

Rodriguez et al. [87, 88]) or when the materials used are not thermoplastic. In these cases the machine acts as a multiple-impact mechanical press.

*Thermal Effects* Due to the poor conductivity for ultrasounds (low module of elasticity and high quantity of air trapped inside) usually exhibited by the materials included in pharmaceutical formulations, a fast decay of ultrasonic energy to thermal energy is obtained. This process has been studied, monitoring the temperature inside the compression chamber by means of a thermistor. In the studied mixtures [87, 88], a fast rise in temperature was obtained in tenths of a second followed by a relatively fast decrease (see Figure 47).

The peak temperature obtained for low ultrasonic energy (25 J) is below 80 °C, whereas for high energies (125–150 J) it is above 140 °C. In mixtures of ketoprofen with acrylic polymers [90], the increase in temperature was slightly lower. In this respect it must be mentioned that a recent modification of the ultrasound-assisted tableting machine that involves the suppression of Teflon isolators in contact with the powder must result in a faster decrease in temperature inside the compression chamber. Thermal effects can cause the total or partial fusion of some components of the formulation. Nevertheless, in the assayed controlled-release formulations, the components are usually below its melting points.

*Plastic Deformation* Plastic deformation results from the combination of thermal and mechanical effects. The thermoplastic excipient was subjected to a temperature above its glass transition temperature ($T_g$) and to a high-frequency mechanical pressure that can avoid the elastic recovery of the material.

*Sintering* The combination of temperature, pressure, and friction effects can result in the sintering of particles, so that the limits between them are no longer evident [46, 87].

Recent studies [91–93] have shown that, for one component of the system undergoing thermoplastic deformation, the continuum percolation model can be used to predict the changes in the system with respect to a traditional pharmaceutical

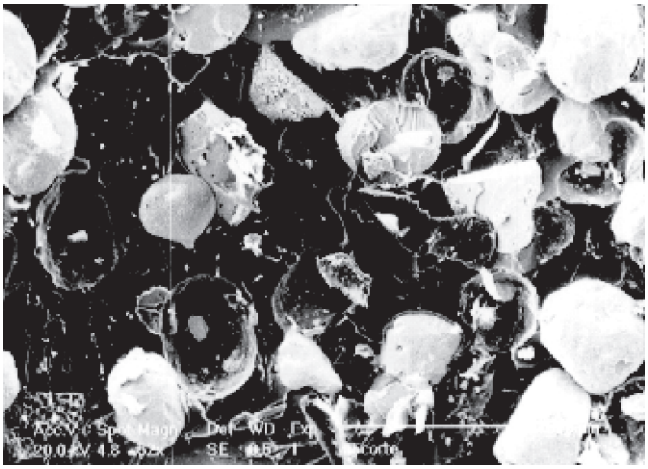

**FIGURE 48** SEM micrograph of matrix tablet containing potassium chloride as drug model and commercial acrylic–metacrylic copolymer. The white KCI particles are surrounded by an almost continuous dark gray mass of excipient. (Courtesy of M. Millán.)

dosage form. The continuum percolation model dispenses with the existence of a regular lattice underlying the system; therefore, the substance is not distributed into discrete lattice sites. This model deals with the volume ratio of each component and a continuum distribution function. The volume ratio is expressed as a space–occupation probability to describe the behavior of the substance [94, 95].

The continuum percolation model predicts an excipient percolation threshold around 16% v/v. This can explain the important decrease in the critical point corresponding to the excipient percolation threshold, a critical point that governs the mechanical and release properties of the matrix.

Ultrasound compaction lowers the percolation threshold of the thermoplastic excipient, resulting in a drastic reduction (about 50%) in the amount of matrix-forming excipient [98] needed to obtain the controlled-release system as well as in a better control of the drug release. The structure of the excipient inside the US-tablets does not correspond to a particulate system but to an almost continuous medium; therefore, there is no an excipient particle size inside these matrices (see Figure 48). Consequently, the percolation threshold of the active agent is higher than in traditional tablets. The insoluble excipient almost surrounds the active agent particles, slowing down the contact with the dissolution medium.

These facts can involve important advantages for the pharmaceutical industry, such as the preparation of controlled-release inert matrices containing high drug doses, with very little increase in the weight of the system. This fact is especially interesting when a high drug dose has to be included in the dosage form, as frequently occurs in controlled-release systems.

On the other hand, application of ultrasonic energy results in an increase in the temperature of the die during the compaction process. The consequences of this fact should be taken into account and cannot be neglected in the case of thermolabile drugs and/or excipients [87, 88].

Further research is needed in the area of ultrasound-assisted compression of pharmaceutical powders, including a higher number of drugs and excipients.

## REFERENCES


1. Unckel, H. (1945), Processes during compression of metal powders, *Archiv fuer das Eisenhuettenwesen*, 18, 161–167.
2. Toor, H. L., and Eagleton, S. D. (1956), Plug flow and lubrication of polymer particles, *Ind. Eng. Chem.*, 48, 1825–1830.
3. Ritschel, W. A., and Bauer-Brandl, A. (2002), *Die Tablette, Handbuch der Entwicklung, Herstellung und Qualitätssicherung*, Editio Cantor Verlag Aulendorf, (Württ), Germany 505.
4. De Boer, A. H., Bolhuis, G. K., and Lerk, C. F., (1978), Bonding characteristics by scanning electron microscopy of powders mixed with magnesium stearate, *Powder Technol.*, 20, 75–82.
5. Nyström, C., Alderborn, G., Duberg, M., and Karehill, P-G. (1993), Bonding surface area and bonding mechanism—Two important factors for the understanding of powder compactability, *Drug Dev. Ind. Pharm.*, 19, 2143–2196.
6. Mattsson, S., and Nyström, C. (2000), Evaluation of strength-enhancing factors of a ductile binder in direct compression of sodium bicarbonate and calcium carbonate powders, *Eur. J. Pharm. Sci.*, 10, 53–66.
7. Rudnic, E. M., and Kottke, M. K. (1999), Tablet dosage forms, in Banker, G. S., and Rhodes, C. T., Eds., *Modern Pharmaceutics*, 3rd rev. and expanded eds., Marcel Dekker, New York, pp. 333–394.
8. Pesonen, T., Paronen, P., and Ketolainen, J. (1989), Disintegrant properties of an agglomerated cellulose powder, *Int. J. Pharm.*, 57, 139–147.
9. Rudnic, E. M., and Kottke, M. K., Tablet dosage forms. In Banker GS, Rhodes CT Eds., *Modern Pharmaceutics* 3rd eds. New York: Marcel Dekker; 1999, pp 333–394.
10. Westerberg, M., and Nystrom, C. (1991), Physicochemical aspects of drug release. XII. The effect of some carrier particle properties and lubricant admixture on drug dissolution from tableted ordered mixtures, *Int. J. Pharm.*, 69, 129–141.
11. Bolhuis, G. K., and Hölzer, A. W., Lubricant sensitivity. VIII. Effect of third components on the film formation of lubricants. In Alberborn G, Nyström C Eds. *Pharmaceutical Powder Compactation Technology*, New York: Marcel Dekker; 1996, pp 550–555.
12. Johansson, M. E. (1984), Granular magnesium stearate as a lubricant in tablet formulations, *Int. J. Pharm.*, 21, 307–315.
13. Carson, J. W. (1998 July), Overcoming particle segregation in the pharmaceutical and cosmetic industries, paper presented at Interphex USA, New York.
14. Vila-Jato, J. L., Remuñán, M. C., Seijo, B., and Torres, D. (1997), Nuevas formas de administración de medicamentos, in Vila-Jato, J. L., Ed., *Tecnología Farmacéutica II: Formas Farmacéuticas*, Editorial Síntesis, Madrid.
15. Tahara, K., Yamamoto, K., and Nishihata, T. (1995), Overall mechanism behind matrix sustained release (SR) tablets prepared with hydroxypropyl methylcellulose 2910, *J. Controlled Release*, 35, 59–66.
16. Soares, L. A. L., et al. (2005), Dry granulation and compression of spray dried plant extracts, *AAPS PharmSciTech*, 6, E359–E366.
17. Núñez Sellés, A. J., Capote, H. R., Agüero, J., Garrido, G., Delgado, R., Martinez, G., Leon, O. S., and Morales, M. (2000), New antioxidant product derived from *Mangifera indica* L, *Abstracts of Papers*, paper presented at the 220th National Meeting of the American Chemical Society, New Orleans, LA, American Chemical Society, Washington, DC.

18. Sánchez, G. M., Re, L., Giuliani, A., Núnez-Sellés, A. J., Davison, G. P., and León-Fernández, O. S. (2000), Protective effects of *Magnifera indica* L. extract, mangiferin and selected antioxidents against TPA-induced biomolecules oxidation and peritoneal macrophage activation in mice, *Pharmacol. Res.*, 42, 565–573.
19. Yoshimi, N., Matsunaga, K., Katayama, M., Yamada, Y., Kuno, T., Qiao, Z., Hara, A., Yamahara, J., and Mori, H. (2001), The inhibitory effects of mangiferin, a naturally occurring glucosylxanthone, in bowel carcinogenesis of male F344 rats, *Cancer Lett*, 163, 163–170.
20. Ichiki, H., Miura, T., Masayoshi, I., Ishihara, E., Komatsu, Y., Tanigawa, K., Ichiki, H., and Okada, M. (1998). New antidiabetic compounds, mangiferin and its glucoside, *Biol. Pharm. Bull.*, 21, 1389.
21. Li, H., Miyahara, T., Tezuka, Y., Namba, T., Nemoto, N., Tonami, S., Seto, H., Tada, T., and Kadota, S. (1998), The effect of Kampo formulae of bone resorption in vitro and in vivo. I. Active constituents of Tsu-Kan-gan. *Biol. Pharm. Bull.*, 21, 1322.
22. Guha, S., Ghosal, S., and Chattopadhyay, U. (1996), Antitumor, immunomodulatory and anti-HIV effect of mangiferin, a naturally occurring glucosylxan thone, *Chemotherapy*, 42, 443–451.
23. Ghosal, S., Rao, G., Saravanan, V., Misra, N., and Rana, D. (1996), A plausible chemical mechanism of the bioactivities of mangiferin, *Indian J. Chem.*, 35B(6), 561–566.
24. Born, M., Carrupt, P. A., Zini, R., Bree, T., Tillement, J. P., Hostettmann, K., and Testa, B. (1996), Electrochemical behaviour and antioxidant activity of some natural polyphenols, *Helv. Chim. Acta*, 79, 1147.
25. Castellanos Gil, E., Iraizoz Colarte, A., Bataille, B., Pedráz, J. L., and Heinämäki, J. (2006), Development and optimization of a novel sustained-release dextran tablet formulation for propranolol hydrochloride, *Int. J. Pharm.*, 317, 32–39.
26. Castellanos Gil, E., Bataille, B., Iraizoz Colarte, A., Delarbre, J. L., El Ghzaoui, A., and Durand, D. (2008), A sugar cane native dextran as an innovative functional excipient for the development of pharmaceutical tablets. *Eur. J. Pharm. Biopharm* 68 (2), 319–329.
27. Jamzad, S., Tutunji, L., and Fassihi, R. (2005), Analysis of macromolecular changes and drug release from hydrophilic matrix systems, *Int. J. Pharm.*, 292, 75–85.
28. Reynolds, T. D., Gehrke, S. H., Hussain, A. S., and Shenouda, L. S. (1998), Polymer erosion and drug release characterization of hydroxypropyl methylcellulose matrices. *J. Pharm. Sci.*, 87, 1115–1123.
29. Miranda, A., Millán, M., and Caraballo, I. (2006), Study of the critical points in Lobenzarit disodium hydrophilic matrices for controlled drug delivery, *Chem. Pharm. Bull.* 54, 598–602.
30. Anonymous (1999, July), Note for guidance on quality of modified release products. Oral dosage forms and transdermal dosage forms, section 1 (quality), The European Agency for the Evaluation of Medicinal Products Human Medicines Evaluation (EMEA), London.
31. Kincl, M., Turk, S., and Vrecer, F. (2005), Application of experimental design methodology (DOE) in development and optimization of drug release method, *Int. J. Pharm.*, 291, 39–49.
32. Eriksson, L., Johansson, E., and Wikstrom, C. (1998), Mixture design—Design generation, PLS analysis, and model usage, *Chemometr. Intell. Lab. Syst.*, 43, 1–24.
33. Gooding, O. W. (2004), Process optimization using combinatorial design principles: Parallel synthesis and design of experiment methods, *Curr. Opin. Chem. Biol.*, 8, 297–304.

34. Congreve, M. S., and Jamieson, C. (2002), High-throughput analytical techniques for reaction optimization, *Drug Discovery Today*, 2, 139–142.
35. I HOLLAND (2006), ITab software, available: http://www.in-pharmatechnologist.com.
36. Leuenberger, H., Rohera, B. D., and Haas, C. (1987), Percolation theory—A novel approach to solid dosage form design, *Int. J. Pharm.*, 38, 109–115.
37. Holman, L. E., and Leuenberger, H. (1988), The relationship between solid fraction and mechanical properties of compacts—The percolation theory model approach, *Int. J. Pharm.*, 46, 35–44.
38. Stauffer, D., and Aharony, A. (1992), *Introduction to Percolation Theory*, 2nd ed., Burgess Science, London.
39. Caraballo, I., Fernandez Arévalo, M., Millán, M., Rabasco, A. M., and Leuenberger, H. (1996), Study of percolation thresholds in ternary tablets, *Int. J. Pharm.* 139, 177–186.
40. Bonny, J. D., and Leuenberger, H. (1991), Matrix type controlled release systems: I. Effect of percolation on drug dissolution kinetics, *Pharm. Acta Helv.*, 66, 160–164.
41. Bonny, J. D., and Leuenberger, H. (1993), Matrix type controlled release systems. II. Percolation effects in non-swellable matrices, *Pharm. Acta Helv.*, 68, 25–33.
42. Blattner, D., Kolb, M., and Leuenberger, H. (1990), Percolation theory and compactibility of binary powder systems, *Pharm. Res.*, 7, 113–117.
43. Caraballo, I., Fernandez-Arévalo, M., Holgado, M. A., and Rabasco, A. M. (1993), Percolation theory: Application to the study of the release behaviour from inert matrix system, *Int. J. Pharm.*, 96, 175–181.
44. Caraballo, I., Millan, M., and Rabasco, A. M. (1996), Relationship between drug percolation threshold and particle size in matrix tablets, *Pharm. Res.*, 13(3), 387–390.
45. Millán, M. (1998 Apr.), Estudio del Umbral de Percolación para la optimización de matrices inertes de liberación controlada, Ph.D. thesis, University of Seville, Seville, Spain.
46. Millán, M., Caraballo, I., and Rabasco, A. M. (1998), The role of the drug/excipient particle size ratio in the percolation model for tablets, *Pharm. Res.*, 15(2), 216–220.
47. Kuentz, M. T., Leuenberger, H., and Kolb, M. (1999), Fracture in disordered media and tensile strength of microcrystalline cellulose tablets at low relative densities, *Int. J. Pharm.*, 182, 243–255.
48. Kuentz, M. T., and Leuenberger, H. (2000), A new theoretical approach to tablet strength of a binary mixture consisting of a well and a poorly compactable substance, *Eur. J. Pharm. Biopharm.*, 49, 151–159.
49. Ramírez, N., Melgoza, L. M., Kuentz, M., Sandoval, H., and Caraballo, I. (2004), Comparison of different mathematical models for the tensile strength–relative density profiles of binary tablets, *Eur. J. Pharm. Sci.*, 22(1), 19–23.
50. Bourquin, J., Schmidli, H., van Hoogevest, P., and Leuenberger, H. (1997), Basic concepts of artificial neural networks (ANN) modeling in the application to pharmaceutical development, *Pharm. Dev. Technol.*, 2, 95–109.
51. Plumb, A. P., Rowe, R. C., York, P., and Brown, M. (2005), Optimisation of the predictive ability of artificial neural network (ANN) models: A comparison of three ANN programs and four classes of training algorithm, *Eur. J. Pharm. Sci.*, 25(4–5), 395–405.
52. Gans, E. H., and Chavkin, L. (1954), The use of polyethylene glycol in tablet coating, *Pharm. Ass. Sc. (Baltin)* Ed., 43, 483–485.
53. Porter, S. C., and Bruno, C. H. (1990), Coating of pharmaceutical solid-dosage forms, in Liebermann, H. A., Lachmann, L., and Schwarz, J. B., Eds., *Pharmaceutical Dosage Forms: Tablets*, Vol. 3, 2nd eds., Marcel Dekker, New York, Chapter 2.

54. Hess, H., and Janssen, H. J. (1969), Lacquered tablets and film coated tablets, *Pharm. Acta Helv.*, 44, 581.
55. Bindschädler, C., Gurny, R., and Dolker, B. (1983), Theoretical concepts regarding the formation of films from aqueous microdispersions and application to coating, Labo-Pharma, *Probl. Tech.*, 31, 389–394.
56. Radebaugh, G. W. (1990–1996), Film coating and film-forming materials: Evaluation, in Swarbrick, J., and Boylan, J. C., Eds., *Encyclopedia of Pharmaceutical Technology*, Vol. 6, Marcel Dekker, New York.
57. Moe, E. (1945), *Overtraeking og Dragering*, Dansk Farmaceutfoeningsverl, Copenhagen.
58. Hasegava, K., Ida, T., Saika, K., and Utsumi, I. (1965), Suitable dimensions of tablets for coating, *Yakugaku Zasshi*, 85, 796.
59. Selmeczi, B., and Smogyi, J. (1996), Suitable dimensions of tablets for coating, *Pharmacie*, 21, 604.
60. Bauer, K. H., Frömming, K. H., and Führer, C. (1993), *Pharmazeutische Technologie [Pharmaceutical Technology*], 4th ed., Thieme, Stuttgart.
61. Bauer, K. H., Lehmann, K., Osterwald, H. P., and Rothgang, G. (1998), *Coated Pharmaceutical Dosage Forms*, Medpharm Scientific Publisher, Stuttgart.
62. Bühler, V. (2003, Sept.), *Kollidon. Polyvinylpyrrolidone for the Pharmaceutical Industry*, 7th ed., BASF Pharma Ingredients, Ludwigshafen, Germany.
63. U.S. Pharmacopeial Forum (2002, May/June), Vol. 28. No. 3, 948–951.
64. Heng, P. W. S., Wan, L. S. C., and Tan, Y. T. F. (1996), Relationship between aggregation of HPMC coated spheroids and tackiness/viscosity/additives of the coating formulations, *Int. J. Pharm.*, 138, 57–66.
65. Tan, Y. T. F., Wan, L. S. C., and Heng, P. W. S. (1998), Evaluation of adhesion strength measurement for predicting aggregation propensity during fluidized bed coating, *S.T.P. Pharm. Sci.*, 8(3) 149–153.
66. Bühler, V. (2001), *Generic Drug Formulation*, 4th ed., BASF Fine Chemicals, Ludwigshafen, Germany.
67. Rowe, R. C., and Forse, S. F. (1983), Pitting—A defect on film-coated tablets, *Int. J. Pharm.*, 17(2–3), 347–349.
68. Melgoza, L. M., Caraballo, I., Alvarez-Fuentes, J., Millan, M., and Rabasco, A. M. (1998), Study of morphine hydrochloride percolation threshold in Eudragit® RS-PM matrices, *Int. J. Pharm.*, 170, 169–177.
69. Caraballo, I., Melgoza, L. M., Alvarez-Fuentes, J., Soriano, M. C., and Rabasco, A. M. (1999), Design of controlled release inert matrices of naltrexone hydrochloride based on percolation concepts, *Int. J. Pharm.*, 181, 23–30.
70. Melgoza, L. M., Rabasco, A. M., Sandoval, H., and Caraballo, I. (2001), Estimation of the percolation thresholds in dextromethorphan hydrobromide matrices, *Eur. J. Pharm. Sci.*, 12(4), 453–459.
71. Boza, A., Blanquero, R., Millán, M., and Caraballo, I. (2004), Application of a new mathematical method for the estimation of the mean surface area to calculate the percolation threshold of Lobenzarit disodium salt in controlled release matrices, *Chem. Pharm. Bull.*, 52(7), 797–801.
72. Reisine, T., and Pasternak, G. (1996), Analgésicos opioides y sus antagonistas, in Goodman, A., Hardman, J. G., Limbird, L. E., Molinoff, P. B., and Ruddon, R. W. Eds., *Las Bases Farmacológicas de la Terapéutica*, Vol. I, McGraw-Hill Interamericana, México, pp. 557–593.

73. Miranda, A., Millán, M., and Caraballo, I. (2006), Study of the critical points of HPMC hydrophilic matrices for controlled drug delivery, *Int. J. Pharm.*, 311, 75–81.
74. Miranda, A., Millán, M., and Caraballo, I. (2006), Study of the critical points in Lobenzarit disodium hydrophilic matrices for controlled drug delivery, *Chem. Pharm. Bull.*, 54, 598–602.
75. Tu, J., Wang, L., Yang, J., and Li, X. (2001), Formulation and pharmacokinetic studies of a cyclovir controlled-release capsules, *Drug Dev. Ind. Pharm.*, 27, 687–692.
76. Wangstaff, J. A., Faulds, D., and Goa, L. K. (1994), Acyclovir: A reappraisal of its antiviral activity pharmacokinetic properties and therapeutic efficacy, *Drugs*, 47(1), 153–205.
77. Higuchi, T. (1963), Mechanism of sustained-action medication. Theoretical analysis of rate of release of solid drugs dispersed in solid matrices, *J. Pharm. Sci.*, 52, 1145–1149.
78. Korsmeyer, R. W., Gurny, R., Doelker, E., Buri, P., and Peppas, N. A. (1983), Mechanisms of solute release from porous hydrophilic polymers, *Int. J. Pharm.*, 15, 25–35.
79. Peppas, N. A., and Sahlin, J. J. (1989), A simple equation for the description of solute release. 3. Coupling of diffusion and relaxation, *Int. J. Pharm.*, 57, 169–172.
80. Ford, J. L., Mitchell, K., Rowe, P., Armstrong, D. J., Elliot, P. N. C., Rostron, C., and Hogan, J. E. (1991), Mathematical modelling of drug release from hydroxypropylmethylcelluose matrices: Effect of temperature, *Int. J. Pharm.*, 71, 95–104.
81. Fuertes, I., Miranda, A., Millán, M., and Caraballo, I. (2006), Estimation of the percolation thresholds in acyclovir hydrophilic matrix tablets, *Eur. J. Pharm. Biopharm.*, 64(3), 336–342.
82. Leuenberger, H., and Rohera, B. D. (1986), Fundamentals of powder compression. II. The compression of binary powder mixtures, *Pharm. Res.*, 3, 65–74.
83. Levina, M., Rubinstein, M. H., and Rajabi-Siahboomi, A. R. (2000), Principles and application of ultrasound in pharmaceutical powder compression, *Pharm. Res.*, 17(3), 257–265.
84. Kromp, W., Trimmel, P., Prinz, F. B., and Williams, J. C. (1985), Vibratory compaction of metal powders, Mod. *Dev. Powder Metall.*, 15, 131–141.
85. Gueret, J.-L. H. (1993), Process for the compaction of a powder mixture providing an absorbent or partially friable compact product and the product obtained by this process, U.S. Patent 5,211,892.
86. Motta, G. (1994), Process for preparing controlled release pharmaceutical forms and the forms thus obtained, International Patent WO 94/14421.
87. Rodriguez, L., Cini, M., Cavallari, C., Passerini, N., Saettone, M. F., Fini, A., and Caputo, O. (1997), Physico-chemical properties of some materials compacted using an ultrasound-assisted tableting machine, in Rubinstein, M., Ed., *Proceedings 16th Pharmaceutical Technology Conference*, Athens, Greece, Vol. I, pp. 267–278.
88. Rodriguez, L., Cini, M., Cavallari, C., Passerini, N., Fabrizio Saettone, M., Fini, A., and Caputo, O. (1998), Evaluation of theophylline tablets compacted by means of novel ultrasound-assisted apparatus, *Int. J. Pharm.*, 170, 201–208.
89. Fini, A., Fernández-Hervás, M. J., Holgado, M. A., Rodriguez, L., Cavallari, C., Passerini, N., and Caputo, O. (1997), Fractal analysis of beta-cyclodextrin-indomethacin particles compacted by ultrasound, *J. Pharm. Sci.*, 86, 1303–1309.
90. Sancin, P., Caputo, O., Cavallari, C., Passerini, N., Rodriguez, L., Cini, M., and Fini, A. (1999), Effects of ultrasound-assisted compaction on ketoprogen/Eudragit S100 mixtures, *Eur. J. Pharm. Sci.*, 7, 207–213.
91. Caraballo, I., Millan, M., Fini, A., Rodriguez, L., and Cavallari, C. (2000), Percolation thresholds in ultrasound compacted tablets, *J. Controlled Release*, 69, 345–355.

92. Millán, M., and Caraballo, I. (2006), Effect of drug particle size in ultrasound compacted tablets: Continuum percolation model approach, *Int. J. Pharm.*, 310, 168–174.
93. Caraballo, I. (2001), Improvement on the release control using ultrasound assisted compression, Eurand Award 2001 "for Outstanding Novel Research in Oral Drug Delivery," organized by Eurand and the Controlled Release Society, 28th Meeting of the Controlled Release Society, San Diego, CA, June 26.
94. Efros, A. L. (1994), *Física y Geometría del desorden*, Hayka, Moscow, pp. 144–147.
95. Kuentz, M. T., and Leuenberger, H. (1998), Modified Young's modulus of microcrystalline cellulose tablets and the directed continuum percolation model, *Pharm. Dev. Technol.*, 3(1), 1–7.

## FURTHER READING

A. BASF Pharma ingredients generic drug formulations, 2004, 2005.
B. Baveja, S. K., Ranga Roa, K. V., and Padmalatha Devi, K. (1987), Zero-order release hydrophilic matrix tablets of β-adrenergic blockers, *Int. J. Pharm.*, 40, 223–234.
C. Bettini, R., Colombo, P., Massimo, G., Catellani, P. L., and Vitali, T. (1994), Swelling and drug release hydrogel matrices: Polymer viscosity and matrix porosity effect, *Eur. J. Pharm. Sci.*, 2, 213–219.
D. Campos-Aldrete, M. E., and Villafuerte-Robles, L. (1997), Influence of the viscosity grade and the particle size of HPMC on metronidazole release from matrix tablet, *Eur. J. Pharm. Biopharm.*, 43, 173–178.
E. Ferrero, C., Muñoz-Ruiz, A., and Jiménez-Castellanos, M. R. (2000), Fronts movements as a useful tool for hydrophilic matrix release mechanism elucidation, *Int. J. Pharm.*, 202, 21–28.
F. U.S. Pharmacopoeia 25 (2002), National Formulary 20, U.S. Pharmacopeial Convention, Rockville, MD.
G. Vázquez, M. J., Peres-Marcos, B., Gómez-Amoza, J. L., Martínez-Pacheco, R., Souto, C., and Concheiro, A. (1992), Influence of technological variables on release of drugs from hydrophilic matrices, *Drug Dev. Ind. Pharm.*, 18, 1355–1375.
H. Velasco, M. V., Ford, J. L., Rowe, P., and Rajabi-Siahboomi, A. R. (1999), Influence of drug: hydroxypropylmethylcellulose ratio, drug and polymer particle size and compression force on the release of diclofenac sodium from HPMC tablets, *J. Controlled Release*, 57, 75–85.
I. Tu, J., Wang, L., Yang, J., Fei, H., and Li, X. (2001), Formulation and pharmacokinetic studies of acyclovir controlled-release capsules, *Drug Dev. Ind. Pharm.*, 27(7), 687–692.

# TABLET PRODUCTION SYSTEMS

Katharina M. Picker-Freyer
*Martin-Luther-University Halle-Wittenberg, Institute of Pharmacy, Division of Pharmaceutics and Biopharmaceutics, Halle/Saale, Germany*

**Contents**

1 Introduction
2 Physics of Tablet Formation
  2.1 Tableting Process
  2.2 Final Formation of Tablet
3 Requirements for Tablet Production Systems
4 Tablet Manufacturing Process
  4.1 Filling
  4.2 Compression
  4.3 Ejection
5 Tableting Machines
  5.1 Single-Punch Tableting Machines
  5.2 Rotary Tableting Machines
  5.3 Application of Tableting Machines
6 Tableting Machine Simulators (Compaction Simulators)
  6.1 Hydraulic Compaction Simulators
  6.2 Mechanic Compaction Simulators
  6.3 Application of Tableting Machine Simulators
7 Instrumentation of Tableting Machines
  7.1 Force Measurement
  7.2 Displacement Measurement
  7.3 Temperature Measurement
  7.4 Measurement of Time
8 Analysis of Tableting Process
  8.1 Force–Time Analysis
  8.2 Displacement–Time Analysis
  8.3 Force–Displacement Analysis
  8.4 Force–Displacement–Time Analysis

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

9 Analysis of Final Tablet Formation
10 Complete Description of Process of Tablet Formation
11 Special Accessories of Tableting Machines
11.1 Optimization of Die Filling
11.2 Tablet Weight Control
11.3 Control of Mixing Homogeneity
11.4 Cleaning
12 Important Factors during Manufacturing Process
12.1 Climatization
12.2 Lubrication
12.3 Occurring Problems during Manufacturing
13 Future of Tablet Production Systems
References

## 1 INTRODUCTION

Tablet production systems can be defined as all machines which are able to produce tablets. They include tableting machines for production and research as well as tableting machine simulators, which are able to mimic the production processes of tableting machines of different size and velocity in order to facilitate scale-up.

Tablets have been produced for more than 150 years. The first tableting machine, developed by Brockedon in 1972 [1], was a manually operated single-punch machine. Currently high-speed tableting machines can produce more than a million tablets per hour.

However, the amount of drug in the early steps of development still makes the use of small tableting machines necessary. Thus, before final production a scale-up from small machines useful for the production of single tablets to high-speed machines is necessary. Since this scale-up is still based on "trial and error" [2], tableting machine simulators can be used to simulate different steps in the scale-up.

In research tableting machine simulators are often used since they allow precise and well-adjustable measurements during tableting with only a small amount of material.

The aim of this chapter is to give an overview of the different techniques used for production and research and further to show the possibilities, that instrumentation of tablet production systems gives in order to analyze the tableting process. The knowledge derived can be used for formulation development as well as to facilitate tablet production and scale-up. Applied techniques which are necessary for production in a good manufacturing practice (GMP) environment will also be discussed.

## 2 PHYSICS OF TABLET FORMATION

Tablets can be defined as two-phase systems which consist of a solid phase (the compressed powder) and the gaseous phase (the air). The solid phase forms a coher-

ent network inside the tablet and a defined form with precise outer dimensions results. Here we define the *process of tablet formation* as the transformation of the powder, a noncoherent solid phase, into the compact tablet, and this process lasts until no further changes are induced by tableting. "Tableting" is only part of the process of tablet formation. It is defined as the process by which the powder is transformed into a coherent form, the tablet. In conclusion tablet formation is the result of tableting and all the changes induced by tableting [3, 4].

## 2.1 Tableting Process

During the tableting, the powder in the die of the tableting machine is transformed by the influence of lower and upper punches into a coherent form—"the tablet" (Figure 1). During tableting different processes occur which are responsible for the cohesion of the tablet [5, 6]. First the particles are pushed together and reoriented in the die until they have arranged in the closest packing. This process is followed by elastic and/or plastic deformation of the particles, as can be observed by confocal laser microscopy [7]. Some materials show brittle fracture. During brittle fracture new particles are produced which can again deform or fracture. Under the influence of the applied forces the particles approach each other up to bonding [8]. The bonded particle collectives continue to deform. The mechanism of deformation and bonding is dependent on material properties and process conditions. Important bonding mechanisms are van der Waals forces, surface films, liquid or solid bridges, and mechanical interlocking [9].

The nature of the resulting bonds [9, 10] and the extent of particle approach [11] determine the cohesion of the final tablet and are responsible for the compactibility of the materials [8]. Further, the formation of bonds occurs not only during the

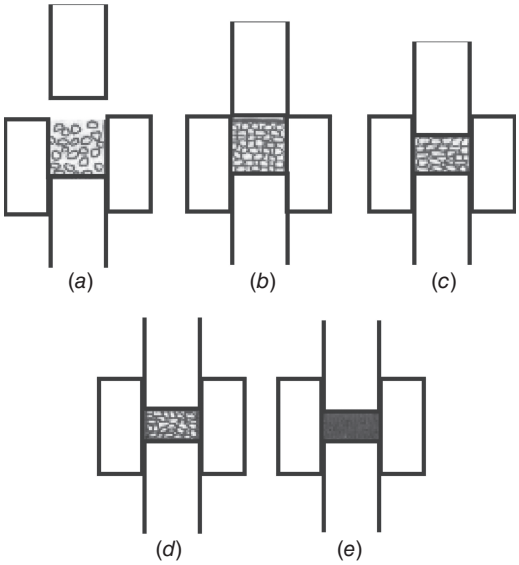


**FIGURE 1** Densification of powder bed: (*a*) particles in die; (*b*) reorientation of particles; (*c*, *d*) particle deformation and fracture; (*e*) tablet in die.

compression phase but also during decompression of the tablet [12, 13]. Hence Leuenberger and Ineichen [14] describe tableting as a percolation phenomenon.

Looking closer at the tableting process it becomes clear that the applied forces during tableting are not only the result of machine movement. When punches are moving in a machine without material, no force can be measured. The force only develops when the punches come into contact with the powder bed. The materials resist pressure deformation, and while the punches are moving, a counterforce builds up which is the measured force at the punches. In conclusion punch forces are determined by the material and as a result materials can be characterized by the measured punch forces.

### 2.2 Final Formation of Tablet

After ejection of the tablet from the die of the tableting machine, the tableting process is finished. However, decompression continues and thus the tablet formation process is not yet finished. The remaining stress inside the tablet can resolve and produce changes in the structure of the tablet. The tablet partially releases the energy gained during tableting. The tablet often shows the occurrence of relaxation, which is called elastic recovery or relaxation of the tablet [15, 16]. This relaxation is primarily axial relaxation and to a small part radial relaxation, as Newton and Rowley [17] figured out. Hiestand [15] called this event "compression born repulsion."

Van der Voort Maarshalk [18] showed that the axial relaxation of the tablet, that is, the change in tablet height, is dependent on the stored energy and hinders the formation of bonds. The elastic recovery [19–21] can be measured by measuring tablet height at different times after tableting. Simultaneously changes in the inner structure of the tablet can occur, for example, shifts of the crystal planes [22, 23]. The fusion of amorphous regions and the recrystallization of amorphous parts are assumed [24–26]. These structural changes are induced by the tableting process and also become visible when transformation of drugs into other polymorphic forms occurs [27–29]. However, even during final formation of the tablets, bonds are formed [15, 30, 31], and thus final tablet formation can be defined as a valuable part of the tablet formation process.

## 3 REQUIREMENTS FOR TABLET PRODUCTION SYSTEMS

Tablet production systems should be able to produce tablets in a reasonable time and without loss of material. Usually they consist of two punches and a die, as schematically shown in Figure 1. The material is placed into the die cavity, which is closed on the lower side by the lower punch. A tablet is formed when the powder is compressed by the punches as described above. The forces evolving to produce a tablet can be up to 80 kN for pharmaceutical purposes [32]; for the production of bigger tablets they are even higher. Usually they range between 10 and 30 kN. Thus punches and dies are produced from hardened steel.

To withstand the evolving forces, the die has to be fixed tightly in the machine, and for that purpose it is accurately fit in a die table consisting of steel. The die table is further friction locked to the machine frame.

Tablet production systems can be operated manually; however most of the systems nowadays operate automatically in order to produce a sufficient number of tablets. Further requirements depend on the type of machine, the production process, the operation mode of the machine, and the production rate of the machine.

## 4 TABLET MANUFACTURING PROCESS

The manufacturing process of tablets principally consists of three stages: filling the die with the powder, compressing the powder, and ejecting the tablet from the die.

### 4.1 Filling

Die filling a tableting machine is a volumetric process. In any case the die is filled to the upper edge of the die and the powder surface flushes with the surface of the die and the die table. The filling depth is determined by the height of the die cavity, the filling volume is determined by the diameter of the die hole and the filling depth. Both the diameter of the die hole and the filling depth are used to calculate the final filling volume.

Three different techniques exist for filling the die. The simplest is to use a filling shoe which moves back and forth over the die and fills the die to its upper edge. The surface is leveled off by the filling shoe.

Another possibility is a filling shoe which does not move. In this case the dies and the accessory lower punches move below the filling shoe and the accessory upper punches move above the filling shoe. In this case the filling volume ends at the upper edge of the die; however, to level off the surface, an additional scraper is necessary. In addition, many machines use the technique of lowering the lower punch and with it the powder bed before tableting [33]. This is most helpful to avoid dusting when the upper punch moves into the die.

The newest filling technology is to fill the die by centrifugal forces [34]. The material moves through specially shaped radial channels which approach the die from the side. In this case the powder volume is determined by the positions of the lower and upper punches in the die. However, for centrifugal filling an absolute must is a free-flowing powder. This is a great disadvantage for all non-free-flowing powders.

### 4.2 Compression

The compression event is the central stage of tablet production. Compression not only depends on the machine but to a great deal on the material properties of a tablet formulation. The principal stages of compression have been described above. Principally, from the machine manufacturing side two different possibilities of compression exist. Either the lower punch is closing the die from the bottom side and the upper punch moves downward for compression or both upper and lower punches move simultaneously toward each other and the powder is compressed from both sides. In the first case the surface hardness of the tablet is not the same on the upper

and lower sides; in the second case the surface hardness is the same on the upper and lower sides.

As already described, the tablet is not completely formed during compression [35, 36]. When the punches leave each other, the tablet relaxes during decompression, further relaxing when one punch leaves the tablet and continuing relaxation after ejection from the die. The process is called decompression as long as a force is measurable. Afterward the process is called elastic recovery or relaxation of the tablet.

### 4.3 Ejection

During ejection usually the lower punch moves upward to eject the tablet from the die and the upper punch has already left the die when the process of ejection starts. Only one machine is presently on the market which ejects the tablets downward at the bottom of the die [34]. After ejection from the die the tablets are collected.

## 5 TABLETING MACHINES

There exist tableting machines which operate in a different manner and can produce in between one single tablet and a million tablets per hour. The orientation of the particles in the machine and their rearrangement, densification, and deformation depend not only on the material but also on the tableting machine used.

The first machines for the production of tablets were simple hydraulic or manually operated presses. Later, eccentric and rotary tableting machines were developed. Today, eccentric tableting machines are only used for research, in early development, or for special applications.

The newest development for production is a machine which fills the dies by centrifugal force [34]. It is a special rotary tableting machine. Another innovation is a special machine which operates by ultrasound [37].

In the following the most important machine types and their working principles will be described.

### 5.1 Single-Punch Tableting Machines

Single-punch machines were the first tableting machines used at the end of the nineteenth century. The upper punch is lowered by a lever arm on the powder bed in the die and by reciprocating this procedure single tablets can be produced. Another possibility is to lower the upper punch by means of a screw. These manually operated tableting machines are no longer used.

Two other types are still in usage for special purposes. Hydraulically operated machines are used to produce tablets, for example, for Fourier transform infrared (FTIR) spectroscopy. In this case the upper punch moves hydraulically onto the powder bed and the tablet is formed in the die. The other possibility is an eccentric tableting machine, which will be described in the following.

***Eccentric Tableting Machines*** Eccentric tableting machines are still used in research, in early development, and for material characterization. They are the machines of choice when you want to use a single-punch tableting machine. As other single-punch machines, they work with one pair of punches. In principle, they have a mobile upper punch and a lower punch which is fixed during compression. The lower punch only moves for ejection of the tablet. The densification process is unilateral.

An eccentric valve is driven by a motor and this eccentric valve is responsible for the movement of the upper punch (Figure 2). The eccentric movement determines the operation and speed of the upper punch. The upper punch is driven by the eccentric valve into the die, which is closed on the bottom side by the lower punch, and the tablet is compressed in the die.

The punch forces for compression evolve due to contact with the powder. Thus the measured punch forces at the upper and lower punches result from the movement of the upper punch and the resistance of the powder bed toward deformation. From that it becomes clear that the measured upper punch force is usually higher than the lower punch force, and thus the upper punch force is the main force used for material characterization. The tablets show different hardness on upper and lower surfaces, as do all tablets produced by single-punch machines [38].

When the upper punch is lifted from the tablet surface, the lower punch moves the tablet upward for ejection out of the die. However, as is known nowadays, the lower punch also moves slightly before due to evolving forces.

In eccentric machines, die filling is performed by a filling shoe which moves back and forth above the die. When the upper punch moves upward above the die, the die is filled, and when the upper punch starts downward movement, the powder bed is leveled off. After ejection of the tablet the filling shoe pushes the tablet down from the die table. During this filling process with a moving filling shoe, demixing of the product mixture is more easily possible than with a fixed filling shoe. This is one major disadvantage of eccentric tableting machines.

The production rate varies usually between 10 and 60 tablets/min and is determined by the number of eccentric movements. An example of an eccentric tableting machine is given in Figure 3. Other machines are described by Ritschel and Bauer-Brandl [32].

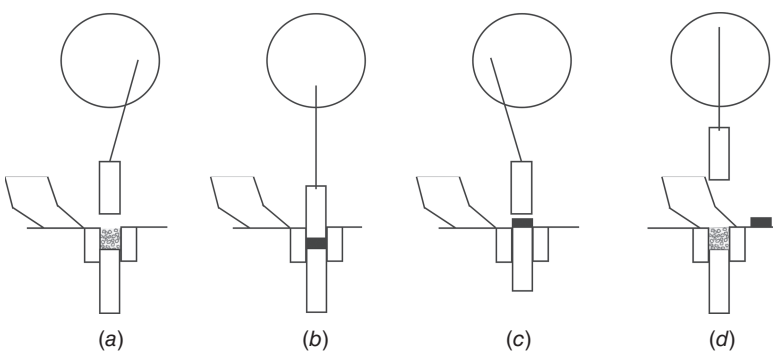


**FIGURE 2** Operation of eccentric tableting machine: (*a*) filling; (*b*) compression; (*c*) ejection; (*d*) pushing from die table.

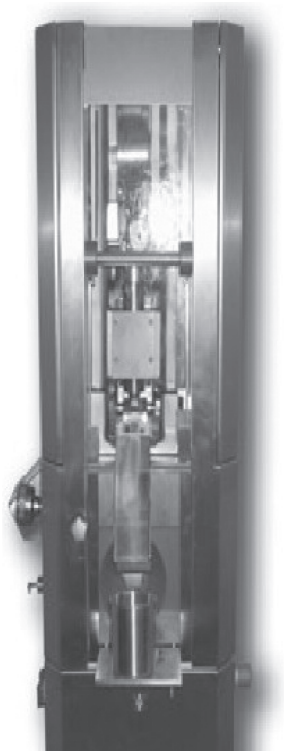

**FIGURE 3** Example eccentric tableting machine. (Courtesy of Korsch XP1.)

### 5.2 Rotary Tableting Machines

Rotary tableting machines are commonly used for tablet production. The principle of all rotary machines is the same, with one exception, which will be discussed separately. According to Konkel and Mielck [39], the information gained with eccentric and rotary machines complement each other.

Rotary tableting machines work with a number of punch and die sets which move in a circle. The dies are fixed in a round die table and the die table circulates. Together with the dies the lower and upper punches circulate on tracks. The lower punches close the dies. The densification process is bilateral since both punches pass the compression wheels and the force is evolving on the upper as well as on the lower side of the powder bed. The produced tablets show the same hardness on the upper and lower surfaces.

There are different stages of tablet production which happen simultaneously for several tablets. The central stage of tableting occurs when the punches pass the upper and lower compression wheels and the compression wheels determine the downward movement of the upper punch and the upward movement of the lower punch: The tablet is formed by the resulting punch forces. The compression wheels can be positioned either by a flexible swing or more seldom by an eccentric valve [40].

In Figure 4 this principal stage and the other stages of tablet production are visible. Except for the compression stage the upper punch is always in an upward position. The upper punch is above the die and the lower punch closes the die, determining the filling depth; then the dies pass below the filling shoe and the powder flows into the die up to its upper edge where the powder bed is leveled off by a scraper. Whereas the upper punch is still in an upward position, the lower punch is lowered slightly on the track to keep the whole powder volume during compression and to prevent dusting [33]. Now the punches start to pass the compression wheels and the main compression event occurs. Forced by the wheels, both the upper and lower punches move toward each other, compress the tablet, and leave each other again. The upper punch lifts into an upward position and the lower punch moves upward to eject the tablet. After ejection the tablet passes a scraper and it is pushed down from the die table. An example of a rotary machine is given in Figure 5.

Additionally often rotary machines are equipped with precompression wheels (Figure 4). In this case the punches pass the precompression wheels before the main compression events starts. Precompression has the same stages of compression as the main compression, but the applied forces are lower. Precompression is deemed to be helpful to avoid, for example, dusting, capping, or lamination. After precompression only a lower main compression force is necessary.

The number of dies and complementary punches of a rotary machine can vary between one and up to hundred. The number of punches and the rotation speed

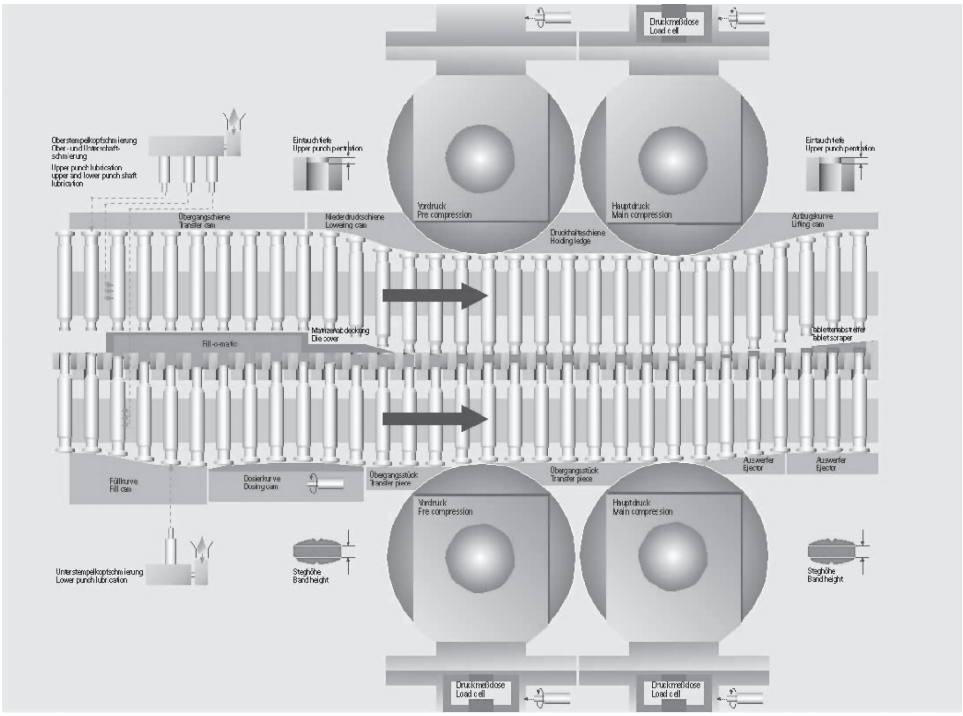


**FIGURE 4** Operation of rotary tableting machine with precompression. (Courtesy of Fette.)

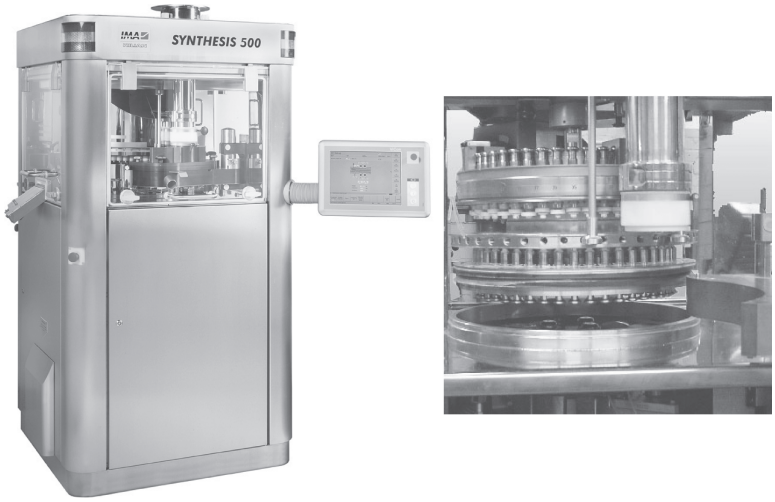


**FIGURE 5** Example rotary tableting machine: left, machine view; right, detail view into compression chamber. (Courtesy of Kilian Synthesis 500.)

of the die table determine tablet production. Fast rotary machines have up to 120 rpm of the die table [32]. However, the number of tablets produced per hour not only is determined by the number of punches and the rotation speed but also is limited by the deformation properties of the tableted material. Materials need some time for deformation, and if the time during one compaction cycle is not sufficient for compression and compaction of the material, no tablets result from the process.

In order to increase the production rate of rotary tableting machines, double-sided rotary machines were build which possess two pairs of compression wheels and two filling shoes. Thus during one rotation double the number of tablets are produced compared with a one-sided rotary machine. If these machines are equipped with precompression wheels one machine contains four pairs of wheels.

***Special Rotary Machines*** As already mentioned one special rotary machine works slightly different—called IMA Comprima (Figure 6) [34]. In this machine the material is filled by centrifugal force from the side directly into the die. The upper punch closes the die at the top and the lower punch closes the die at the bottom. However, when the given volume in the die is filled by the powder, both punches move downward until the die is completely closed. Then the compression process starts and the dies pass the compression wheels. After compression the tablet is ejected by the upper punch at the bottom of the die, contrary to all other rotary tableting machines which eject the tablet at the top of the die.

***High-Speed Rotary Tableting Machines*** High-speed rotary machines work with the same principles as all other rotary machines. They possess a huge number of punch and die sets and often two filling stations. Another possibility is to use punch and die sets which are able to produce several tablets simultaneously. Special tooling can be used for this purpose; however this is not the subject of this chapter. As

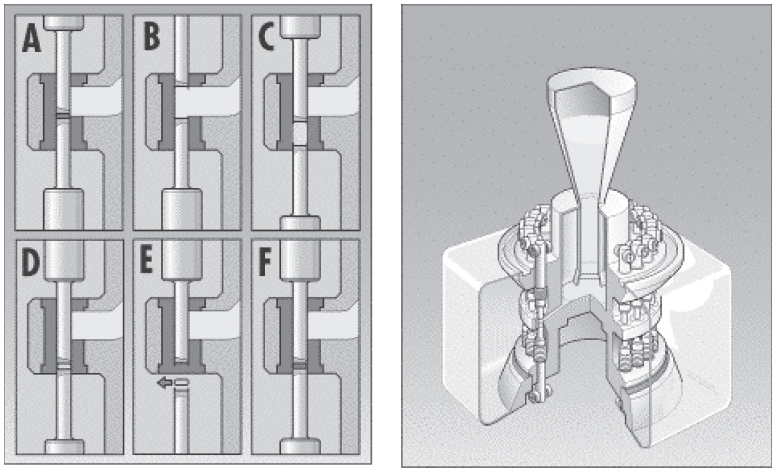


**FIGURE 6** Working principle of Comprima tableting machine with centrifugal filling: left, operation mode; right, centrifugal filling. (Courtesy of IMA Comprima.)

already mentioned production speed depends to a great extent on product properties. Excellent powder flow is essential since the dies have to be filled completely. This is also most essential for the IMA Comprima, which works with centrifugal force for die filling. Often this machine cannot be used since powder flow properties are not sufficient. To improve powder flow on conventional rotary machines special filling devices have been developed (Section 11.1).

### 5.3 Application of Tableting Machines

In summary, the single-punch tableting machines still being used are mainly eccentric tableting machines (mostly for research), whereas rotary machines with different production output are predominantly used for production; and for rotary machines in most cases it is not machine speed that determines the production rate but material flow and compression properties.

## 6 TABLETING MACHINE SIMULATORS (COMPACTION SIMULATORS)

Tableting machine simulators [41–48] have been developed in order to mimic tablet production systems with a very small amount of powder. Similar to eccentric tableting machines, tableting machine simulators use one pair of punches. Working only with a single pair of punches reduces the consumption of tableting materials and facilitates instrumentation for displacement measurement.

### 6.1 Hydraulic Compaction Simulators

The first compaction simulators developed were hydraulic [41–45]. The hydraulic system is electronically controlled. An example is given in Figure 7. Either compression force cycles or movement of the punches was freely adjustable. This allowed

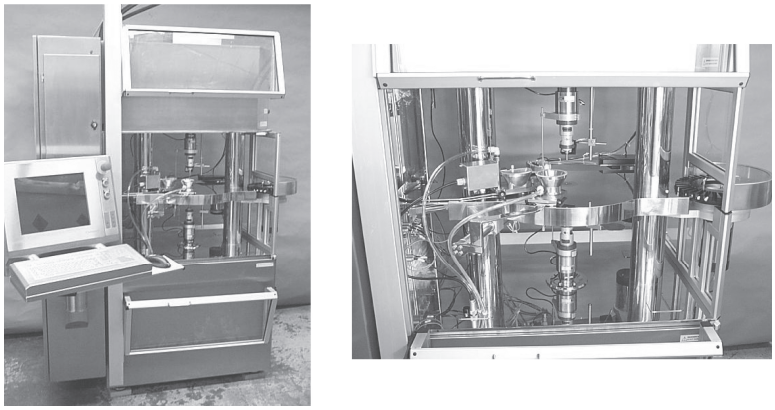

**FIGURE 7** Example hydraulically working compaction simulator: left, machine view; right, detail view into compression chamber. (ESH compaction simulator, Courtesy of Huxley Bertram.)

much variation and the primary aim was to mimic the densification process of a rotary machine and the mechanical factors influencing it.

For example, theoretically, from machine geometries, the force–time profile of a rotary tableting machine [49–52] can be deduced and calculated and the data are programmed into the compaction simulator. However, the force–time profile of a tableting machine could not be calculated. Too many factors influence the measured force, for example, the tableted material, the geometries of the machine, the machine wear time, tableting speed, and tableting tools. Similarly, the displacement–time profile of a tableting machine, especially a rotary tableting machine, is very difficult to calculate. It has been shown that calculation from machine geometries is only possible to a certain extent. Mainly the mechanics of a tableting machine cannot be completely simulated [53, 54]. Thus either an approximated displacement–time profile can be used for programming the compaction simulator or approximation of real punch movement is only possible using recorded data from real tableting machines.

Thus the simulation of tableting machines needs much effort and a real simulation is almost impossible because of the hydraulic control. Further the filling process of rotary tableting machines cannot be simulated since die filling is usually processed by a filling shoe moving forth and back.

However, hydraulical compaction simulators are still used in research for basic material characterization. They show the advantage of controlling speed exactly and of using low and high punch travel speeds, between 10 and 300 mm/s. Mostly a simple displacement profile is used for characterization (e.g., a saw tooth or a sine wave profile), and the evolving forces at the lower and upper punches are measured. Further the speed of the punches can be controlled separately and both punches move freely and independently from each other. Time intervals in which the punches stand still can be freely set. Thus lots of freedom for material characterization is possible and these compaction simulators are important tools.

Another advantage of compaction simulators is that only small amounts of material are necessary to produce a tablet. One single tablet can be produced at low as well as at high speed of the punches. This is important in order to evaluate defor-

mation properties of a formulation already in early dosage form development when only small quantities of the drug substance are available. Since nowadays the timelines for production of a new medicine are tight, this advantage of compaction simulators becomes more and more important.

### 6.2 Mechanical Compaction Simulators

More recently mechanical compaction simulators have been developed. The first was the linear mechanical rotary tableting machine simulator Presster (Figure 8), which was introduced in 1998 [55, 56]. It can mimic the mechanics of different rotary tableting machines and is called a linear rotary tableting machine replicator. The name Presster was combined from press and tester.

A single pair of punches moves linearly forth and back on a lower and an upper punch track. For tableting the punches pass the compression wheels which are equivalent in dimensions to those of rotary tableting machines used in practice. These compression wheels are exchangeable. Different machines are simulated by exchanging them.

The machine speed can be varied and different tableting machines are simulated by using similar dwell times between 5 and 80 ms. Special tests exhibited that rotary tableting machines can be simulated with a precision of 1–5% [57, 58]. One major disadvantage of the Presster is that it works with a moving filling shoe, and thus the filling process of rotary machine cannot be simulated.

In addition, the present model of the Presster possesses precompression wheels and thus, besides compression, precompression can be simulated. This is import

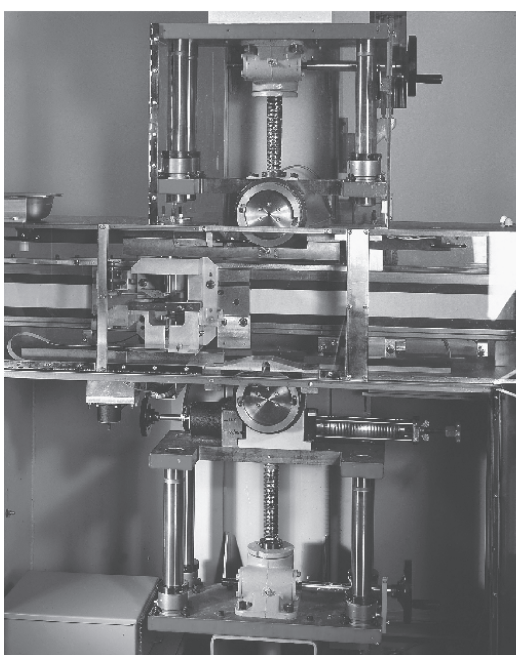

**FIGURE 8** Detail of Presster. (Courtesy of MCC Corp.)

when studying the effect of precompression on the final tablet properties. However, the time between precompression and main compression is determined by the Presster geometries since the positions of the precompression and main compression wheels are fixed.

The newest development for compaction simulation is a mechanical tableting machine simulator which operates with a cam. Thus it is called Stylcam (Figure 9) [59]. The cam is positioned on the lower compression wheel and allows the simulation of different tableting machines and their dwell times due to different acceleration of the punches. It was introduced in 2005.

With the Stylcam different dwell times are obtained by adjusting the speed of the compression wheels. Precompression is simulated by compressing a tablet twice. Thus the time interval between the precompression and main compression is freely adjustable. One further advantage of the Stylcam is that it works with a fixed filling shoe, as on a conventional rotary tableting machine. However, data on the precision of this instrument are not yet available.

### 6.3 Application of Tableting Machine Simulators

Using mechanical compaction simulators allows us to simulate the tableting process of rotary tableting machines to a greater extent than when using hydraulical compaction simulators. Thus they will be mainly used in formulation development and scale-up.

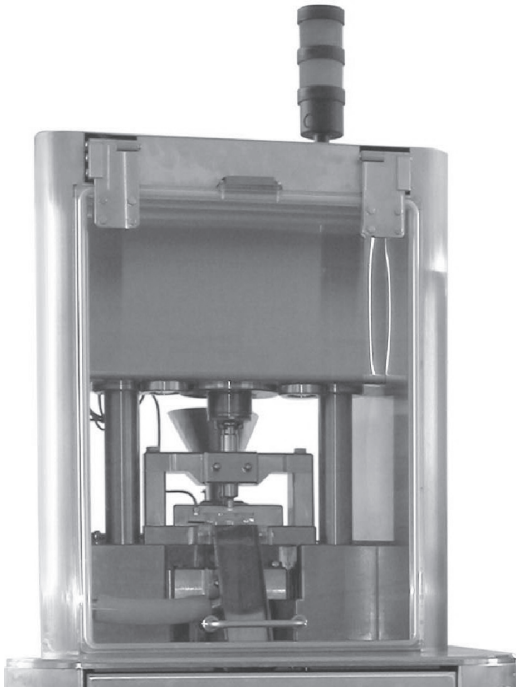

**FIGURE 9** Detail of Stylcam. (Courtesy of MedelPharm.)

However, for mechanical compaction simulators the movement of the punches is mechanically determined and, compared to hydraulic compaction simulators, not freely programmable. Thus, for basic material characterization and early formulation development, hydraulical compaction simulators can be advantageous.

## 7 INSTRUMENTATION OF TABLETING MACHINES

To describe the tableting process more precisely, tableting machines have been instrumented since the middle of the last century. Measured values are force, displacement, and temperature and they are always measured with dependence on time. Thus time is another variable.

### 7.1 Force Measurement

The first instrumentation of a tableting machine for measurement of upper punch force was performed by Brake [60]. Thus it was for the first time possible to visualize the compression process with regard to force development which results from the material stresses during tablet formation. Only shortly after that, similar measurements were published by another research group [61, 62].

Besides upper punch force, lower punch force, die wall force [63–65], ejection force [66], and tablet scraper force can be measured. Die wall force measurement will be discussed separately.

For measurement strain gages are mostly used. These strain gages consist of constantan. They are applied in eccentric tableting machines at the upper or lower punch holder and in rotary tableting machines at the machine frame or the compression roll pin. Alternatively piezoelectric crystals can be used which have to be placed inside the punches [40].

The sensitivity of force measurement is dependent on the distance between the force transducer and where the force occurs. Thus for force measurement, instrumentation of the punches is more advantageous than instrumentation of the machine frame. However, since punch and die sets have to be exchanged between different runs of the machines, instrumentation of the punch holder, the machine frame, or the roller pin is most widely spread.

The most often measured force is the upper punch force. For the eccentric machine it is the force which controls densification; for rotary tableting machines upper and lower punch forces have ideally the same values. Schmidt et al. [67] measured force with a single punch of a rotary tableting machine. Ejection force is visible as a small lower punch signal which occurs shortly after the end of one compaction cycle. It is measured by lower punch instrumentation but needs more resolution. A review of force measurement is given by Bauer-Brandl [68].

***Die Wall Force Measurement*** During compression of the powder the forces are evolving not only at the punches but also at the die wall [63–65]. Therefore die wall force measurement complements upper and lower force measurement. Since the compression process occurs axially, these radially evolving forces are smaller than the forces at the punches. Measurement of die wall force allows, for example, for indication on die wall friction, tablet capping, and lamination. Instrumentation for

die wall force measurement is difficult and different techniques have been developed [64, 69, 70]. The die can be instrumented axially or radially and strain gages or piezoelectric crystals can be used for measurement. Two main effects influence the measured signal: tablet height and tablet position. Related to this, the output signal can be nonlinear. Piezoelectric foils have been applied which possess the advantage of independence on tablet position [70].

One example most recently developed is a split die consisting of three sections (Figure 10) [64]. Integrating the sensing web in a thin middle layer isolates stress measurement to a narrow band around the tablet and gives much closer approximation to the true stress. Further die wall force measurement is linear and independent of tablet height and position as it is uncoupled from all other die wall stresses and strains. Further it is designed in the shape of a conventional die and can be mounted without modification into a die table.

## 7.2 Displacement Measurement

The first measurement of upper punch displacement was performed in the mid-1950s by Higuchi and co-workers [62, 71] with the aid of inductive displacement transducers. By the same instrumentation the movement of the lower punch can be visualized. Inductive transducers are mounted parallel to the punch and thus give information on punch position. Alternatively touchless measurement of displacement is possible. It is important that the transducers be positioned most closely to the punch in order to minimize the influence of machine deformation.

A measurement of displacement on a rotary tableting machine was presented in 1987 by Schmidt and Tenter [72]. Another possibility was presented by Matz and co-workes [73, 74]. Meanwhile touchless measurement systems for recording displacement were developed [73, 74]. For all measurements of displacement, correc-

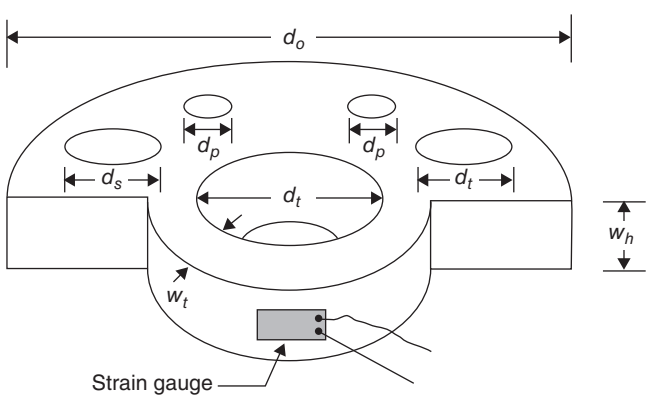


$d_i$: inner diameter $= \frac{3}{8}$ in.
$d_o$: outer diameter $= \frac{7}{8}$ in.
$d_s$: screw hole diameter $= \frac{5}{32}$ in.
$d_p$: alignment pin hole diameter $= \frac{1}{8}$ in.
$w_a$: web height $= \frac{1}{8}$ in.
$w_t$: web thickness $= \frac{1}{16}$ in., $\frac{1}{8}$ in.

**FIGURE 10** Construction details of split die. (Reproduced with permission from ref. 64.)

tions which take elastic punch and machine deformation into account are necessary. Müller and Caspar [75] showed problems which occur when machine and punch deformation are not taken into account. Krumme and co-workers gave an extensive description on this issue for eccentric tableting machines [76, 77]. From punch displacement measurements tablet height can be calculated, and this height can further be related to tablet density and porosity.

Again and again it was tried to derive displacement theoretically from machine geometry, punch geometry, and measured force [49–52]. However, until today this theoretical derivation has not been satisfying and thus experimental testing cannot be given up.

### 7.3 Temperature Measurement

Due to the forces evolving during tableting the compressed material can warm up. The evolving temperatures can be measured with different methods. In each case only approximated measurements are possible since either the measurement was not performed directly inside the tablet or additives were necessary, which can alter measurement.

Several methods have been applied to determine the temperature increase with thermal sensors which are installed in the punches, inside the die, or in the powder bed (Table 1) [78–87].

Most recently an analysis technique was developed which allows measurement of tablet temperature directly after ejection of the tablet on the machine (Figure 11) by an infrared sensor [85]. The temperature signal can be directly related to force and displacement measurement.

### 7.4 Measurement of Time

All measured variables can be determined with dependence on time. Thus the tableting process can be characterized for several variables with dependence on time.

Two time definitions are important: contact time and dwell time. Contact time can be defined as the time during which a contact of powder and punches is measurable, for example, when the force exceeds a certain limit of 100 N. Dwell time can be defined predominantly for rotary machines as the time during which the punch heads are completely under the compression wheels and thus the applied force is constant.

**TABLE 1 Methods to Determine Temperature During and Shortly After Compaction**

| |
|---|
| Measurement in punches [78–80] |
| Epoxide punches [81] |
| Calorimetric measurement [82, 83] |
| Infrared measurement [84, 85] |
| Measuring conductivity during tableting with conductive materials [86] |
| Energy calculations [78, 83, 84] |
| Determination of melting of materials with certain melting point [87] |

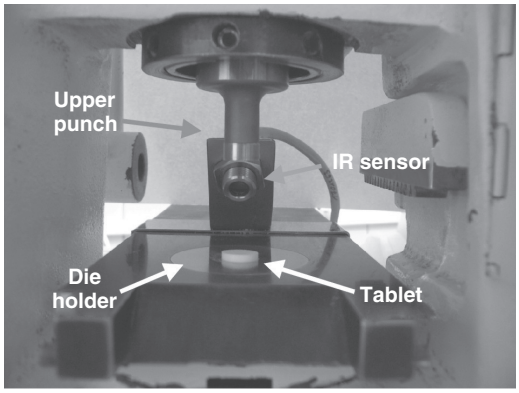


**FIGURE 11** Infrared sensor unit for measuring temperature directly after tableting [85] (Martin-Luther-University Halle-Wittenberg.)

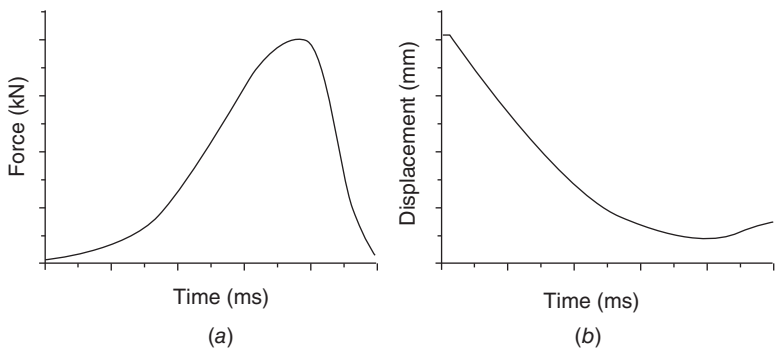


**FIGURE 12** Force–time and displacement–time profile for eccentric tableting machine.

Further for description of the tableting process, the entering of the punch into the die, the compression start at the begin of contact time, and the lifting of the upper punch from the tablet are important.

## 8 ANALYSIS OF TABLETING PROCESS

All measured variables, namely force, time, displacement, and temperature, can be combined differently and can be analyzed afterward. From the functional relations, conclusions can be drawn about the compression and compaction behavior of the materials.

The most basic analysis is the presentation of force versus time or displacement versus time. These curves are different for eccentric and rotary tableting machines. The data given in Figures 12 and 13 are valid for the contact time of the compaction cycle of one single tablet. Due to the eccentric-driven movement of the punches, the force–time curve can be described by a sharp peak at the maximum force evolving at the punches and the displacement–time curve can be described with a sharp peak at the minimum height of the powder bed. For curves of eccentric tableting

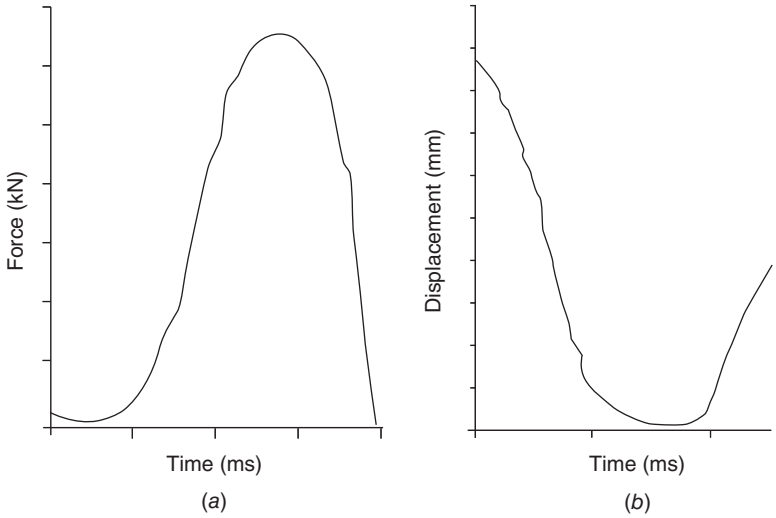


**FIGURE 13** Force–time and displacement–time profile for rotary tableting machine.

machines hardly any dwell time is measurable (Figure 12). In contrast, for rotary tableting machines, the force–time curve and the displacement–time curve are flatter at the maximum peak. This is the case due to the dwell time when the punch heads move completely between the compression wheels. The dwell time is indicated in Figure 13. For rotary tableting machines with precompression wheels additionally force–time curves and displacement–time curves for precompression can be recorded. They look similar to the curves from the main compression wheels with the exception that lower forces are applied. Precompression data will not be discussed in the following since the data can be treated similar to the data of the main compression event.

Besides this presentation of force and displacement versus time, which are data directly derived from the tableting machine, other more advanced methods are possible. Extensive reviews on the methods used can be found in the literature [54, 88, 89]. In the following only the most important aspects will be discussed.

### 8.1 Force–Time Analysis

One method to analyze tableting data is the use of force–time or pressure–time diagrams. They are easily recordable since displacement measurement is not necessary.

Some basic parameters can be directly read from the curves. For the force values upper and lower punch forces and ejection forces should be mentioned, and for the time values contact time should be mentioned. Deduced parameters such as pressure and normalized contact time can be calculated and further statistical data are often used for characterization (Table 2). Due to the different shapes of force–time curves from eccentric tableting machines compared with those from rotary tableting machines, some parameters can only be calculated from eccentric machine data and some can only be calculated from rotary machine data.

**TABLE 2 Parameters Directly Deduced from Force–Time Profiles**

| |
|---|
| Maximum upper punch force (pressure) |
| Maximum lower punch force (pressure) |
| Maximum ejection force (pressure) |
| Contact time |
| Normalized contact time |
| Time at maximum upper compression force |
| Time at maximum lower compression force |
| Maximum upper precompression force (pressure) |
| Maximum lower precompression force (pressure) |
| Precompression contact time |
| Normalized precompression contact time |
| Time at maximum upper precompression force |
| Time at maximum lower precompression force |

**TABLE 3 Parameters Calculated for Eccentric and Rotary Tableting Machines**

| Source | Parameters |
|---|---|
| Chilamkurti [90] | Area under curve (AUC) |
| | Height of curve at maximum upper punch force |
| | Width of curve at half-maximum upper punch force |
| | Slope during compression |
| | Slope during decompression |

Only eccentric machine data allow us to calculate the *R* value (maximum upper punch force/maximum lower punch force), which is an indication of friction. They also allow us to calculate the time difference between the maximum upper punch force and the maximum lower punch force. Only dwell time and the minimum force during the dwell time can be calculated for rotary tableting machine data. The rise time of rotary machines is defined as the time during the compression phase, and peak offset time is defined as the time difference between maximum pressure and vertical alignment of the punches. Further the inflection points during the compression and decompression phases are mostly only calculated for rotary machine data.

In addition, for force–time diagrams different methods to characterize the tableting process were developed. These methods can be divided in those applicable to force–time curves from eccentric and rotary tableting machines [90] and those applicable only to data from eccentric or rotary tableting machines (Tables 3–5).

One possibility to analyze the tableting process is to describe the areas under the curve during compression and decompression and to draw conclusions on plastic and elastic parts of deformation. Emschermann and Müller [91] applied this method to data from eccentric machines (Figure 14). Similar area comparisons were performed by the research group of Schmidt [92–94] for rotary machines (Figure 15). They tried to gain information on elasticity by calculating differences between the area under the plot in the compression phase and the area under the plot in the decompression phase. A sophisticated technique to interpret area data under one

**TABLE 4 Parameters Calculated for Eccentric Tableting Machines**

| Source | Parameters |
|---|---|
| Emschermann [91] (Figure 14) | Area under compression curve (compression area)<br>Area under decompression curve (decompression area)<br>Compression area/decompression area |
| Pressure–time function (modified Weibull function) [39, 96–98] (Figure 16) | $p = p_{\text{max, upper punch}}\left[\frac{1-t}{\beta}\right]^{\gamma} \exp\left[-\left(1-\frac{1-t}{\beta}\right)^{\gamma}\right]$<br>where $\beta$ = time parameter<br>$\gamma$ = asymmetry parameter |
| Modified Fraser–Suzuki function [99] | $f(t) = H \exp\left\{\left[\frac{-0.693}{A^2}\right] \times \left[\ln\left(\frac{\langle 1 + A \times (t_r - t)\rangle}{S}\right) \times 1.177\right]^2\right\}$<br>where $A$ = asymmetry parameter<br>$t_r$ = time parameter<br>$S$ = deviation of maximum |

**TABLE 5 Parameters Calculated for Rotary Tableting Machines**

| Source | Parameters |
|---|---|
| Tenter [72] | Area center of gravity |
| Vogel [92–94] (Figure 15) | $t_1$ = compression time<br>$t_2$ = time at the start of dwell time<br>$t_3$ = time at the half of dwell time<br>$t_4$ = time at the end of dwell time<br>$t_5$ = time at the end of one compression cycle<br>$A_1$ = area of densification<br>$A_4$ = area of decompression<br>$A_2$ = partial area<br>$A_3$ = partial area<br>$A_3/A_2$ = area quotient<br>$A_5$ = partial area<br>$A_6$ = partial area<br>$A_6/A_5$ = area quotient |

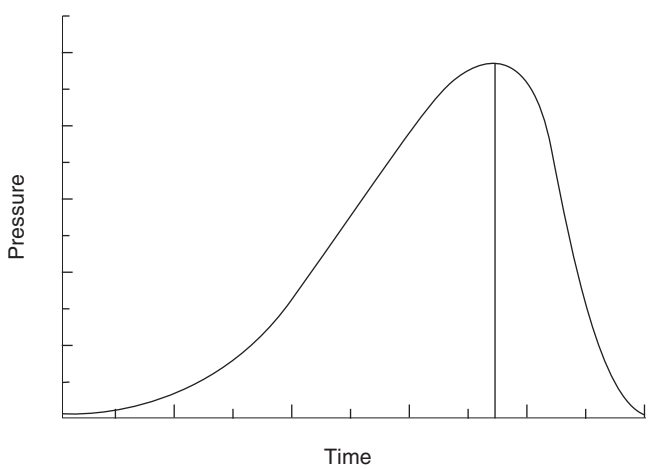


**FIGURE 14** Pressure–time curve analysis [91].

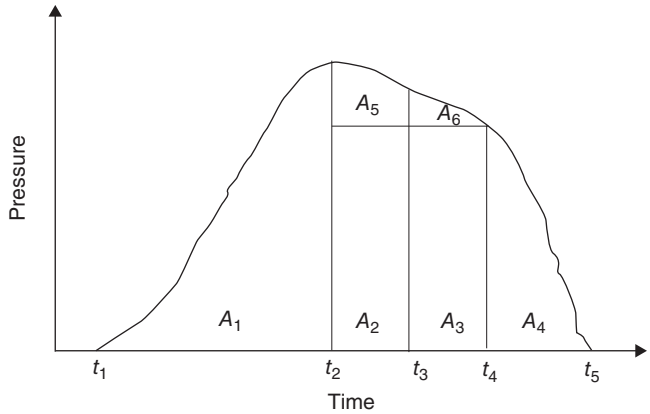


**FIGURE 15** Pressure–time curve analysis [92, 93].

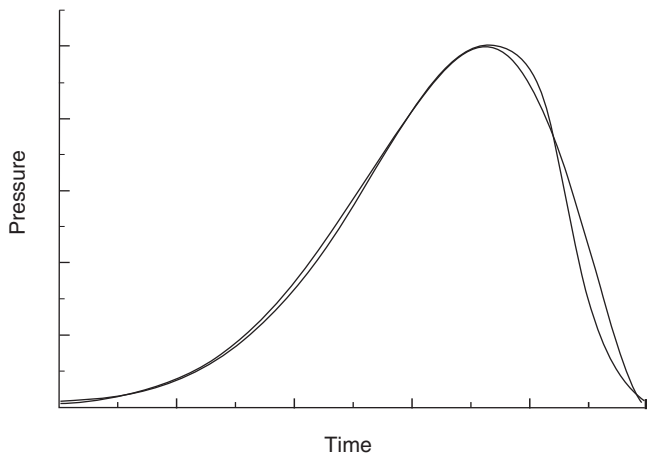


**FIGURE 16** Pressure–time curve analysis with pressure–time function [39, 96–98].

compaction cycle was developed. The parameters of interest are given in Table 5. Further advances were performed by Yliruusi and co-workers [95].

Another possibility of analysis is to fit different functions to the force–time data. The research group of Mielck [39, 96–98] described the densification behavior by the pressure–time function (Figure 16). The lower the values of the parameters β and γ, the more plastically the material deforms; the higher the values, the more elastically the material deforms. The parameter γ describes the asymmetry of the curve and β the time at maximum densification. A similar function, the Fraser–Suzuki function, which originates from chemical analytics was applied to tableting data [99]. It can also be used to derive parameters that describe the deformation behavior of materials. Information on the reversible and irreversible deformation of the material can be deduced.

**TABLE 6 Parameters Directly Deduced from Displacement–Time Profiles**

| |
|---|
| Maximum displacement |
| Minimum height of tablet |
| Minimum volume of tablet |
| Maximum density of tablet |
| Minimum porosity of tablet |
| Maximum relative density of tablet |
| Time at maximum displacement |
| Time at minimum height of tablet |

### 8.2 Displacement–Time Analysis

From displacement–time data a few parameters can be read. The most important data are given in Table 6. In addition the data allow us to calculate fast elastic recovery. The increase in tablet height from the minimum tablet height in the die up to the lifting of the upper punch from the tablet is called fast elastic recovery.

Since the travel of the punches is analyzed, the data also allow us to determine the speed of the punches at each point of the compaction cycle. Punch speed is an important parameter to compare different tableting events [100]. Maximum punch speed can be determined and used for characterization.

Only a few authors have tried to relate displacement with time. Ho and Jones [101] determined the slope of porosity over time (rise time). This slope was also used by Tsardaka [102] for analysis.

### 8.3 Force–Displacement Analysis

The most extensively used method to characterize the tableting process is the use of force and displacement measurements. Usually upper punch force and upper punch displacement are used. Models which relate force and displacement directly can be distinguished from those which analyze pressure and volume.

Further, some parameters can be directly read from the curves. The peak offset time of eccentric tableting machines is defined as the time difference between maximum displacement and maximum compression force [101, 103].

***Relation between Force and Displacement*** The first information on force–displacement analysis can be found in Führer [104] and Moldenhauer et al. [105]. Force–displacement diagrams (Figure 17) are used to calculate from the areas enclosed the work or energy necessary for tableting. The force–displacement profile includes compression and decompression of the powder to the tablet. The area between compression and decompression is the area of the compaction energy, often called the energy of plastic deformation ($E_P$) [106]. The area $E_E$ is the energy of elastic deformation. And the last area ($E_F$) to complete the triangle start of compression ($D_0$)–maximum force ($F_{max}$)–displacement at maximum pressure ($D_{PM}$) can be interpreted as the energy of friction ($E_F$) [107, 108]. The sum of all three energies is the total energy $E_T$ of the tableting machine. The energies can be displaced as an absolute value or relative to the total energy. Based on these values, Stamm and Mathis [109] developed the determination of plasticity $P$ as

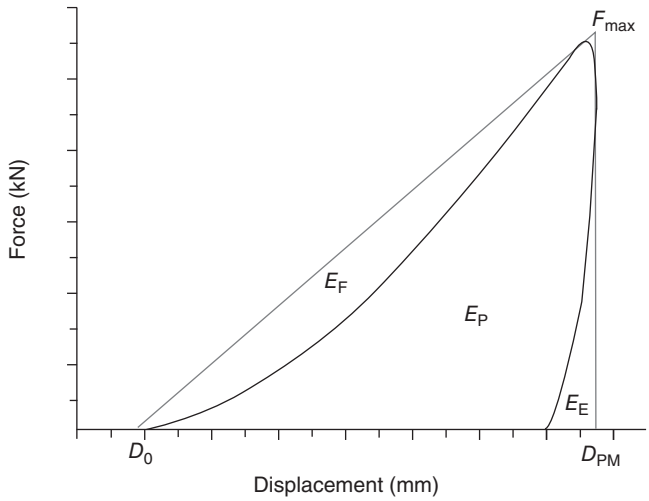


**FIGURE 17** Force–displacement diagram for energy analysis [104–108].

$$P = \frac{E_P}{E_P + E_E} \tag{1}$$

Other authors tried similar attempts [110, 111] and the developed methods were regarded to be very useful [27, 112–115]. Antikainen and Yliruusi [116] more recently tried to derive further parameters from the diagrams to enable a more complete characterization. An overview on the possibilities for force–displacement analysis is given by Ragnarson [117].

***Relation between Volume and Pressure*** The oldest method of this type of analysis is to establish a relation between the volume of the tablet and the force necessary to produce this volume [62, 71]. For exact description the height of the tablet is determined by displacement measurement and the accuracy of this measurement is extremely important. Further displacement measurement has to be corrected precisely for elastic deformation of the punches and the machine in order to use correct tablet heights.

From tablet height, volume, porosity, and relative density at different stages of densification can be deduced. These variables are plotted as a function of pressure. For analysis, for example, the equations of Heckel [118–120], Kawakita, [121, 122], Cooper and Eaton [123], Walker [124], Bal'shin [125], and Sønnergaard [126, 127] can be used. The equations are given in Table 7. A further overview of these and other equations used can be found in Celik [88].

The Heckel equation describes the densification process with first-order kinetics. A linear equation is obtained with a slope which is inversely proportional to the yield strength. The slope of the Heckel equation provides information on the plastic deformation of the powder. It has also been published that the slope of the Heckel equation can be correlated with the elastic modulus (Young's modulus).

**TABLE 7 Parameters Calculated from Force–Time Profiles**

| Source | Parameters |
|---|---|
| Heckel [118, 119] (Figure 18) | $-\ln \varepsilon = \ln\left(\frac{1}{1-D}\right) = Kp + A$<br>where $K$ = deformation parameter<br>$A$ = powder bed densification |
| Yield pressure [120] | $\text{Yield pressure} = \frac{1}{\text{Heckel slope}}$ |
| Yield strength [118] | $\text{Yield strength} = \frac{1}{3 \times \text{Heckel slope}}$ |
| Kawakita [121, 122] | $\frac{p}{C} = \frac{1}{ab} + \frac{1}{a}p$<br>where $a$ = porosity of powder bed<br>$b$ = compression parameter |
| Walker/Balshin [124, 125] | $100V_{\text{rel}} = 100 \times \frac{V}{V_\infty} = -W \log p + C$<br>where $W$ = compressibility coefficient |
| Sønnergard [126] | $V_{\text{rel}} = V_1 - W \log p + V_e e^{-p/p_m}$<br>where $W$ = compressibility coefficient |
| Cooper–Eaton [123] | $V^* = \sum_{i=1}^{n} a_i V_i^* = \frac{V_0 - V}{V_0 - V_\infty} = a_1 e^{-k_1/p} + a_2 e^{-k_2/p}$<br>where $k_1$ = deformation pressure for fraction part 1<br>$k_2$ = deformation pressure for fraction part 2<br>$a_1$ = fraction part 1 of deformation<br>$a_2$ = fraction part 2 of deformation |
| Cooper–Eaton (linearized) [123] | $\ln V^* = \ln \frac{V_0 - V}{V_0 - V_\infty} = -\frac{Q}{p} + \ln R$<br>where $Q$ = extent of compressibility<br>$R$ = sum of fraction parts |

The equation of Kawakita describes volume reduction with pressure in the form of a hyperbolic equation. Walker and Bal'shin [125] postulated a logarithmic relation between applied pressure and volume reduction, which was further modified by Sønnergard [126]. Cooper and Eaton [123] use an exponential function, which can also be linearized. Pressure thresholds for deformation mechanisms are determined. It should be noted that all of these equations and tableting models determine descriptive parameters.

The equation of Heckel is the most extensively used model and the underlying porosity–pressure plot is called a Heckel plot (Figure 18). The equation for the linear compression process follows fist-order kinetics (Table 7). Heckel

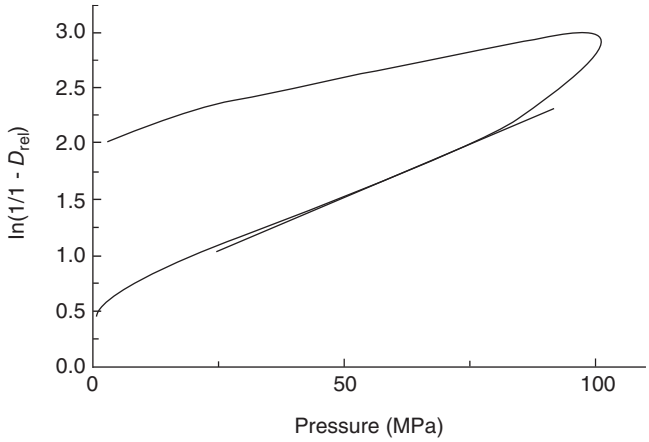


**FIGURE 18** Heckel plot [118, 119].

distinguished measurements which determine the volume of the tablet without pressure (zero pressure) [119] from those measurements which determine the volume with pressure [118]. The first method allows determination of the volume after release of the elastic energy; the second method allows a higher precision repeatability since it is often difficult to determine tablet height at a defined time after ejection. Some milliseconds can cause differences [35, 36]. Years later Sun and Grant [128] tried to determine the elastic part at pressure measurements. The experiments showed that deviations in Heckel plots at high pressures are dependent on the elasticity of the material.

The equation of Heckel has been discussed again and again. One main issue of critique is that pharmaceutical powders are not purely plastically deforming materials and thus particle size and deformation mechanisms influence the derived parameters [129, 130]. Already very small errors in displacement determination or the measurement of true density can induce huge errors in the derived parameters [75–77, 129, 131, 132]. Sønnergaard [126] referred the equation of Walker and Bal'shin for his characterization of materials. He criticized further that the yield strength derived from the Heckel equation is directly dependent on the true density of the powders [127].

Despite this critique of the Heckel equation, the analysis of Heckel plots has been intensively used for the description of powder compression [128, 133–136]. Gabaude et al. [136] stated that the analysis is quite useful when defining preconditions exactly and apply correct displacement measurement.

Since the development of the equation, it has been tried to derive further information from it. Rees and Rue [129] determined the area under the Heckel plot. Duberg and Nyström [137] used the nonlinear part for characterization of particle fracture. Paronen [138] deduced elastic deformation from the appearance of the Heckel plot during decompression. Morris and Schwartz [139] analyzed different phases of the Heckel plot. Imbert et al. [134] used, in analogy to Leuenberger and Ineichen [14], percolation theory for the compression process as described by the Heckel equation. Based on the Heckel equation, Kuentz and Leuenberger [135, 140] developed a new derived equation for the pressure sensitivity of tablets.

Tsardaka and co-workers [102, 141, 142] presented the Heckel plot with dependence on time and analyzed deformation in combination with elastic recovery. Additional areas to describe plasticity were determined from two-dimensional (2D) plots [129]. Finally, the three-dimensional (3D) model [143, 144] was developed by fitting a plane to a 3D data plot on the basis of normalized time, pressure, and porosity according to Heckel.

### 8.4 Force–Displacement–Time Analysis

Force, displacement, and time are the three most important parameters to characterize the compaction cycles of tableting materials. Even when Hoblitzell and Rhodes postulated a linear relationship between force–time and force–displacement data [145], this could not been exactly proved until today. Thus it is important to analyze these three measured data together.

From force–displacement–time curves some parameters can be directly deduced; for example, the power during tableting and the maximum power can be determined. Another important parameter which serves as a measure for viscoelasticity is the peak offset time [101, 103]. For single-punch tableting machines it is defined as the time difference between maximum displacement and maximum force.

In addition, advanced models as those calculating viscoelasticity and the 3D model have been developed. They will be described in the following.

***Viscoelasticity Models*** For characterization with viscoelasticity models, simulation models have been developed on the basis of Kelvin, Maxwell, and Voigt elements. These elements come from continuum mechanics and can be used to describe compression.

David and Augsburger [146] were the first to try this method of analysis. Further tests, for example, determination of complex functions based on methods of numerical mathematics, were performed by the research group of Müller [147, 148]. Although the results were helpful, for exact description the models became rather complicated and the derived parameters were complex. According to Bauer [149], these models have to be three dimensional for a reasonable description. Another similar approach was used by the research group of Rippie [12, 13]. They described the structure evolution in the tablet during decompression by the aid of vectorial 3D models and concluded that fracture and stress contribute to the final structure of the tablet.

Other research groups derived viscoelastic properties from creep experiments of the final tablet [150–154]. As Tsardaka and Rees [142] determined, stress relaxation follows a hyperbolic equation.

***3D Model*** Most recently another technique which uses force, displacement, and time has been developed. 3D modeling is a very useful method to characterize the tableting process [3, 46, 47, 155–159]. Force is expressed as pressure, time has to be normalized, and from displacement data the porosity according to Heckel [119] is calculated. It is the only possibility to combine these variables during analysis.

To describe the tableting process the three variables were presented in a 3D plot (3D data plot) and a plane was fitted to the data twisted at $t = t_{max}$ (Figure 19*a*). From the fitting process the parameters $d$ (time plasticity), $e$ (pressure plasticity),

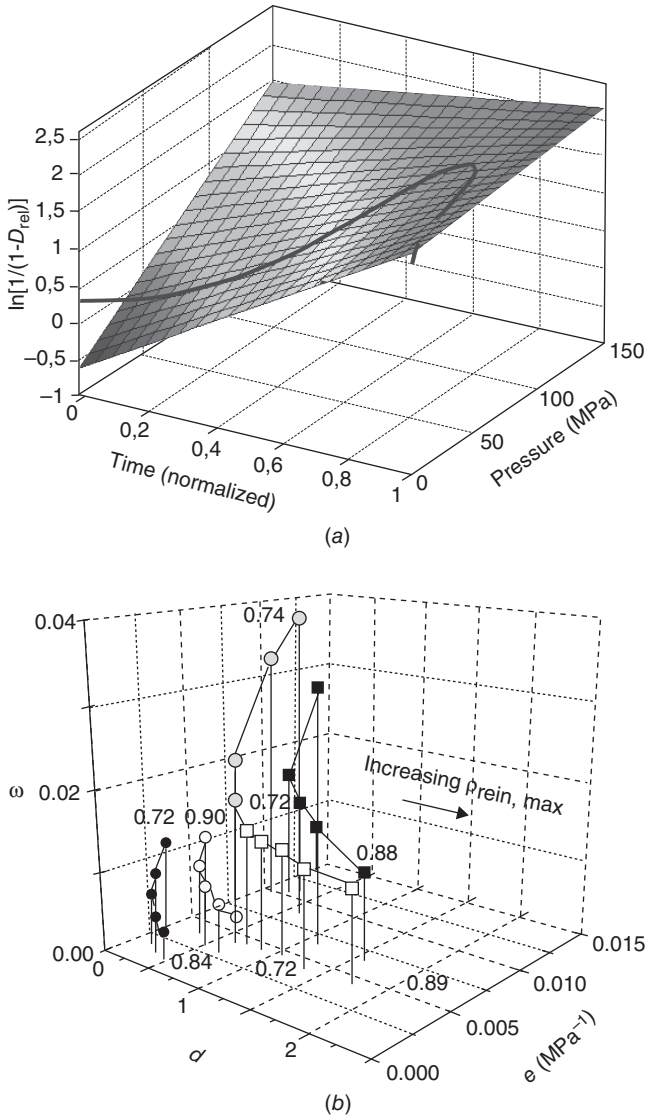


**FIGURE 19** (*a*) 3D data plot with fitted plane twisted at $t = t_{max}$ and (*b*) 3D parameter plot of (●) DCPD: dicalcium phosphate dihydrate, (○) spray-dried lactose, (□) MCC: microcrystalline cellulose, (●) theophylline monohydrate, and (■) HPMC: hydroxypropyl methylcellulose for data gained with an eccentric tableting machine [47].

and ω (twisting angle, which indicates fast elastic decompression) can be derived (Table 8).

The parameters of the fitted plane (time plasticity *d*, pressure plasticity *e*, and twisting angle ω) were also exhibited in a 3D plot and this plot is called the 3D parameter plot. This plot exhibits the compression behavior of the powder. It gives a simple yet characteristic description of the tableting properties. An example is given in Figure 19*b*.

**TABLE 8 Parameters Calculated from Force–Displacement–Time Profiles**

| Source | Parameters |
|---|---|
| 3D model [143, 144, 155] (Figure 19) | $z = \ln\left(\frac{1}{1-D_{rel}}\right) = [(t - t_{max})(d + \omega p_{max} - p)] + (ep) + (f + dt_{max})$ <br> $d = \frac{\delta \ln[1/(1-D_{rel})]}{\delta t}$ $e = \frac{\delta \ln[1/(1-D_{rel})]}{\delta p}$ $f = \ln\left(\frac{1}{1-D_{rel}}\right)$ <br> where $D_{rel}$ = relative density <br> $t$ = time <br> $p$ = pressure <br> $t_{max}$ = time at maximum pressure <br> $p_{max}$ = maximum pressure <br> $\omega$ = twisting angle at $t_{max}$, indicates fast elastic decompession <br> $d$ = time plasticity <br> $e$ = pressure plasticity |

Time plasticity $d$ describes the plastic deformation of the excipient according to time [160]. It is influenced by tableting speed [157]. With increasing time plasticity $d$ the powder deforms faster during tableting. Therefore, with increasing densification time plasticity increases. Pressure plasticity $e$ describes the pressure-dependent increase of density. The pressure plasticity correlates with the slope of the Heckel equation [144]. With increasing pressure plasticity $e$ the slope of the Heckel equation increases in the same direction and the necessary pressure for deformation (yield pressure [118] or yield strength [120]) decreases. The twisting angle $\omega$ is a measure of the material's elasticity and the ratio between compression and decompression. Thus, it indirectly describes fast elastic decompression during the tableting process. When $\omega$ increases, elasticity decreases. The twisting angle $\omega$ correlates with the elastic modulus [144]. In conclusion, materials which deform fast show high $d$ values, materials which deform easily and with low pressure show high $e$ values, and those which relax a lot show a lot elasticity and thus low $\omega$ values.

Thus the 3D model allows us to characterize the tableting process completely and to distinguish time-dependent information from pressure-dependent deformation and elasticity in one step of the analysis.

***Temperature Analysis*** The results gained by determination of temperature during and shortly after tableting vary strongly and depend on the method used.

A temperature increase of 5 K [79, 80, 161] could be determined with conventional punches; however, with epoxide punches the measured temperature increase was as high as 30 K [81]. By calorimetric measurement an increase of 10 up to 30 K [82] was determined and by infrared measurement the increase was 10–15 K [84].

Beissenhirtz [86] measured a temperature increase of 30 K indirectly by measuring conductivity which arose during tableting with conductive materials. Energy calculations indicate a temperature increase of more than 30 K caused by tableting

[78, 84, 83]. Most recently, partial melting of drugs could be analyzed for materials whose melting temperature is as high as 94 °C [87], and the reversible transgression of a glass transition temperature of 80 °C was determined [162].

All results indicated that temperature increase depends on the material. Further temperature increase during tableting can contribute to slight changes in material structure [85].

## 9 ANALYSIS OF FINAL TABLET FORMATION

This characterization of the process of tablet formation has to be completed by analyzing the changes induced by tableting.

Most important is the elastic recovery of the tablets which starts during decompression and is finished dependent on the material after several days. Elastic recovery can be defined as [163].

$$\mathrm{ER}(\%) = 100 \times \frac{H_t - H_{min}}{H_{min}} \tag{2}$$

where $H_{min}$ is the minimum height of the tablet under load and $H_t$ the height of the tablet at different times $t$ after tableting.

Elastic recovery gives information on the remaining elasticity of the materials which is only slowly released. Further it can indicate structural changes inside the materials and tablets. Structural changes induced by tableting have to be analyzed by physicochemical techniques, such as spectroscopic and thermoanalytical methods, X-ray diffraction, scanning electron microscopy, and transmission electron microscopy [35, 36, 85, 164–166]. The analyzed changes will help to better understand the process of tablet formation and identify the reasons for compactibility of materials. However, these changes are not the subject of this chapter.

## 10 COMPLETE DESCRIPTION OF PROCESS OF TABLET FORMATION

On the whole, the process of tablet formation can fully be described by combining the analysis of the tableting process with the final formation of the tablets. The methods which gives most detailed information of the whole process and simultaneously is a fast method is the 3D modeling technique in combination loith calculating the elastic recovery of the tablets. In addition, by combining both these methods and calculating general plasticity $P$ from time plasticity $d$, pressure plasticity $e$, twisting angle $\omega$, and elastic recovery ER, a more general tool for analysis of the process of tablet formation is available [3, 4].

Finally the crushing force of the tablets after relaxation gives information on the formed bonds inside the material and the compactibility. Compactibility has been described by Leuenberger [167]. For the future it can be expected that a prediction of compactibility as a result of the process of tablet formation is possible.

## 11 SPECIAL ACCESSORIES OF TABLETING MACHINES

In tablet production it is essential to control tablet weight and tablet homogeneity in order to ensure a uniform dosage form. For patient safety pharmacopeias demand that tablet weight is between certain limits. Tablet producers often set their more narrow specifications to ensure that they meet pharmacopeial specifications. To control tablet weight and tablet homogeneity an optimal product mixture, complete filling of the die, exact tooling, and tightly controlled machine conditions are necessary.

Another demand for patient safety and due to GMP regulations is to ensure the absence of impurities in the final dosage form, for example, residues from the previous product or residues from detergents. Thus the process of cleaning of machines has to be standardized and controlled. In addition optimized short cleaning times of the machine increase operating time for production. During the last years one innovation for tableting machines was the development of special accessories for cleaning in order to reduce standing times.

For special products (e.g., cytostatics or sterile products), it is necessary to produce tablets in a hermetically closed machine. For these products special containment solutions have been developed which allow the production in a hermetically closed machine or behind a wall. Most important is to separate the tablet production zone strongly from the mechanics of the machine.

In the following tablet filling devices, possibilities to control tablet weight and mixing homogeneity as well as advances in cleaning technology will be discussed.

### 11.1 Optimization of Die Filling

The basics of filling have been explained above: The filling shoe is moving back and forth, the filling shoe is fixed, or filling is centrifugally controlled. Two problems arise generally: Either the product demixes and tablet weight and content uniformity are no longer controlled or the die is not completely filled and thus tablet weight also varies.

Optimal filling of the die is determined to a great extent by the material, but the speed of the machine is also important. At low machine speeds the die is usually completely filled; at high machine speeds this becomes more difficult. Thus special filling devices using one or more paddles have been developed to improve filling. One example for a paddle feeder is given in Figure 20. Alternatively filling devices can be vibrated to improve feeding for materials with bad flow characteristics. An overview is given by Ritschel and Bauer-Brandl [32, 168].

Besides improved feeding, paddle feeders allow improved mixing uniformity since the formulation is mixed again shortly before feeding the die. This mechanical remixing is the only possibility to improve the homogeneity of the mixture. Demixing is a bigger problem for machines with a moving filling shoe than for those with a fixed filling shoe. Thus moving filling shoes are equipped with paddles as a standard. Fixed filling shoes need paddles usually only for materials with bad flowability or at high machine speeds.

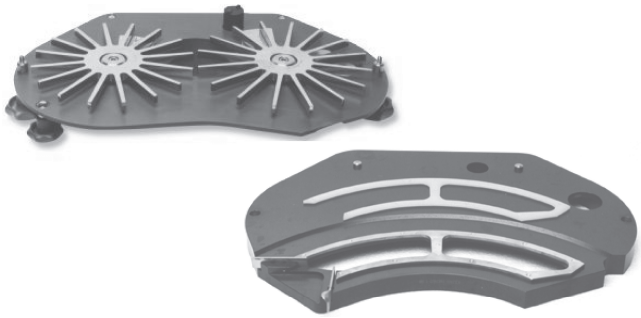

**FIGURE 20** Example paddle feeder. (Courtesy of Kilian.)

### 11.2 Tablet Weight Control

To control tablet weight different possibilities exist. The simplest method is to weigh at preset intervals (in-process control) a number of tablets manually and to adjust machine settings according to the results when necessary. For high-speed rotary machines automatically working weighing systems have been developed which determine tablet weight shortly after ejection [169, 170]. These systems can also determine tablet height and diameter, which are indirect measures for tablet uniformity. Simultaneously with tablet weight compression force drifts. A direct relation between compression force and tablet weight exists. Thus it is possible to monitor tablet uniformity also by control of compression force. For this purpose the machines are instrumented with strain gauges or piezoelectric force transducers. By control of compression, force changes in tablet weight can be directly detected.

Automatically working control systems are able to eject those tablets separately which fail the requirements (rejection mechanism), and they adjust the machine for die filling or compression force and collect only those tablets which meet the requirements. Two alternative principles for automatic tablet weight control and adjustment are possible, depending on the application and selected machine type. The principle of control of compression force is based on measurement of the final compression force under constant tablet height. This principle is used for all applications where tablet weight accuracy and constant tablet density are less critical.

The principle of control of displacement is based on measurement of tablet height variations under constant force. This principle is more accurate than the force control system. It is used for all applications where constant density of the produced tablets is critical.

Modern systems combine one of these control systems with automatic weighing of tablets. Weight control will automatically adjust the filling depth in order to keep tablet weight within specified tolerance limits.

### 11.3 Control of Mixing Homogeneity

The systems to control weight uniformity are not able to control uniformity of the mixture. When during filling of the die the tableted material demixes, tablet weight

usually tends to vary. However, these variations can be small and not easy to detect. To monitor mixing uniformity in the final tablet, most recently spectroscopic techniques such as Raman spectroscopy and near-infrared (NIR) spectroscopy have been used [171–173]. Special online sensors have been build into the machine and they measure the spectrum for each tablet. When the mixture is homogeneous, the appearance of the spectrum will always be the same or between certain limits. If not, the production can be stopped and adjusted as far as possible. Thus a further step in quality assurance of tablets has been made. This was partially caused by the process analytical technique (PAT) initiative of the U.S. Food and Drug Administration (FDA) [174].

### 11.4 Cleaning

To ensure product quality, cleaning is of utmost importance. Therefore different standardized cleaning technologies have been developed [175, 176]. Detailed information may be provided by manufacturers.

Usually, the most effective way to implement cleaning is to design it into a process which has to be performed after tableting. It involves the addition of spray systems, tank cleaners, nozzles, and seals into the tableting machine in order to automate the cleaning process. The automation converts the batch processes to a continuous operation of tableting cycles and cleaning cycles. Cleaning or washing in place means an advanced wash liquid preparation system which handles all filtering, preheating, mixing, and pumping of water, detergents, and demineralized water and provides continuous monitoring and control of cleaning parameters. Another possibility for standardized cleaning is the wash-off-line procedure. In this case exchangeable compression modules are especially designed for a fast product changeover. Different techniques have been developed to exchange the modules. Either carries or arms or lifting systems are applied or additionally used. The wash-off-line procedure increases production time by cleaning the exchangeable compression modules after it is removed from the tablet press. Thus production time is increased; however, two compression modules are necessary. In this case a special separate washing system is necessary.

## 12 IMPORTANT FACTORS DURING MANUFACTURING PROCESS

To run a tablet production process effectively, robustly and smoothly several factors have to be kept in mind. Of utmost importance for the process are environmental humidity during tablet production and adhesion forces between machine punches and dies. These and other factors can contribute to problems during manufacturing. In the following the relevance of climatization during tableting, the necessity and methods of lubrication, and frequently occurring problems during manufacturing will be discussed.

### 12.1 Climatization

Relative humidity (RH) in the production room influences the water content of the materials. This has to be kept in mind during tablet production.

When the materials sorb water, they deform differently compared to the status before sorption, and compressibility changes [98, 177, 178]. Furthermore the compactibility of the materials changes and tablets with a different crushing force and friability result [179]. Even the release from the tablets can be influenced [180]. However, the influence of water content on tableting and tablet properties depends on the material; for example, hydrophilic polymers are mostly influenced by RH. Further, the influence of RH during production is most decisive when production conditions change extremely. Smaller differences up to 10% RH do not influence a robust formulation [3]; however for critical formulations even these changes are of importance. Usually the conditions at one production site do not change from day to day, but great differences have been observed between different seasons of the year. Since tablet production and the final tablet quality should be the same throughout the year, often a humidity interval between 40 and 60% RH is used for tablet production. This is a first step. However, for characterization of material properties humidity control between 40 and 60% RH is not sufficient. Material properties cannot be compared when obtained at different conditions. In this case humidity control at a certain humidity with a precision of 2–5% is absolutely necessary [3].

## 12.2 Lubrication

Adhesion forces between the material and punches and dies result in sticking of the tablets at the punches and dies. These adhesion forces are further influenced by RH and this has also to be kept in mind. When the adhesion forces at the punches and dies are greater than the cohesion forces between the particles inside the tablet, the tablets stick at the punches and can cap.

To overcome this problem, lubrication is the method of choice [181, 182]. Two alternatives for lubrication exist: internal lubrication and external lubrication. Internal lubrication is performed by mixing the tableted product shortly before the tableting process with a solid lubricant. Thus the lubricant is not only at the surfaces of the final tablet but also inside the tablet. As a result, internal lubrication lowers bonding, and this is especially the case for plastically deforming materials [183–186]. The most frequently used and most effective material for internal lubrication is magnesium stearate [187–189]; however, other hydrophobic or amphiphilic lubricants are also possible [190–196]. Magnesium stearate has one major disadvantage: It shows a low solubility and remains as a solid after dissolution. Thus the search for other lubricants is ongoing.

When the lubricant should not be part of the tablet formulation, for example, when bonding properties of the drug are low, external lubrication is necessary [197]. For single-tablet production the punches and dies can be manually lubricated with a fluid. In production several methods have been developed to place the fluid on the surface of punches and dies [198]. Filaments applied at the punches to lubricate the die or special caps with fluid lubricant are possible solutions. However, external lubrication also has disadvantages [199].

## 12.3 Occurring Problems during Manufacturing

The most frequent problems occurring during the manufacture of tablets are high tablet weight variation, capping and lamination [200], and further picking and stick-

ing at punches and dies. Low product yield, low crushing force, and further tablet yams and chipping are other problems which have to be solved [33, 201].

High tablet weight variation can be reduced by using weight control systems. Further demixing of the tableted material has to be avoided, since demixing results in higher tablet weight variation and content uniformity can no longer be achieved. As already discussed, paddle feeders can be used to achieve mixing homogeneity for problematic products and further spectroscopic techniques can be used for control.

The problem of capping and lamination can be solved by increasing RH or adding wetting agents. Further either external lubrication may help. Picking and sticking of the tablets at punches and dies can be avoided by using lubricants as discussed above.

A low product yield is caused by loss of material during fast production processes. On rotary tableting machines this problem is solved by slightly lowering the lower punches before the compression event starts.

A low crushing force is often caused by the composition of the special formulation. If the formulation itself is not the reason for a low crushing force, compression force can easily be increased. Another possible explanation can be low humidity of the tableted product.

Finally tablets yam and chipping can occur before the tablets leave the die table of a rotary machine. Usually the lower punches or the tablet scraper of the tableting machine are not properly adjusted. Another reason can be low crushing force.

In conclusion, for smooth and perfect machine runs, product properties and machine conditions have to be tightly controlled.

## 13 FUTURE OF TABLET PRODUCTION SYSTEMS

In principle tablet production systems have remained the same throughout the last century. However, major improvements in instrumentation, data acquisition, and analysis techniques have been made. Nowadays more sophisticated data acquisition and analysis techniques are available which facilitate and improve interpretation of tableting data.

In order to facilitate scale-up, more sophisticated simulation systems can be thought of. They will be a real help for scale-up with small amounts of material as available in early development of formulations.

Further quality control of the tablets during tableting or shortly after has become more important. Recent trends show improvements for production in a GMP environment by isolating the production unit from the machinery. For the near future the implementation of the PAT initiative of the FDA is conceivable.

## REFERENCES

1. Brockedon, W. (1843), Shaping pills, lozenges, and black lead by pressure in dies, *British Patent* 9977.
2. Leuenberger, H. (1997), Research in solid dosage forms—An obsolete topic? *Pharm. Dev. Technol.*, 2, vii–viii.

3. Picker, K. M. (2002), *New Insights in the Process of Tablet Formation—Ways to Explore Soft Tableting*, Habilitationsschrift, Martin-Luther-University Halle-Wittenberg, Halle, Germany, and also Görich und Weiershäuser Verlag, Marburg.
4. Picker, K. M. (2004), Soft tableting: A new concept to tablet pressure sensitive drugs, *Pharm. Dev. Technol.*, 9(1), 107–121.
5. Train, D. (1957), Transmission of forces through a powder mass during the process of pelleting, *Trans. Inst. Chem. Eng.*, 35, 258–266.
6. Parrott, E. L. (1990), Compression in pharmaceutical dosage forms, in Lieberman, H. A., Lachman, L., and Schwartz, J. B., Eds., *Pharmaceutical Dosage Forms, Tablets. Bd.2*, Marcel Dekker, New York, pp. 201–243.
7. Guo, H. X., Heinämäki, J., and Yliruusi, J. (1999), Characterization of particle deformation during compression measured by confocal laser scanning microscopy, *Int. J. Pharm.*, 186, 99–108.
8. Hiestand, E. N. (1997), Mechanical properties of compacts and particles that control tableting success, *J. Pharm. Sci.*, 86, 985–990.
9. Rumpf, H. (1958), Grundlagen und Methoden des Granulierens, *Chem. Ing. Tech.*, 30, 144–158.
10. Joneja, S. K., Harcum, W. W., Skinner, G. W., Barnum, P. E., and Guo, J. H. (1999), Investigating the fundamental effects of binders on pharmaceutical tablet performance, *Drug. Dev. Ind. Pharm.*, 25, 1129–1135.
11. Adolfsson, A., Gustafsson, C., and Nyström, C. (1999), Use of tablet tensile strength adjusted to surface area and mean interparticulate distance to evaluate dominating bonding mechanism, *Drug Dev. Ind. Pharm.*, 25, 753–764.
12. Rippie, E. G., and Morehead, W. T. (1994), Structure evolution of tablets during compression unloading, *J. Pharm. Sci.*, 83, 708–715.
13. Hoag, S. W., and Rippie, E. G. (1994), Thermodynamic analysis of energy dissipation by pharmaceutical tablets during stress unloading, *J. Pharm. Sci.*, 83, 903–908.
14. Leuenberger, H., and Ineichen, L. (1997), Percolation theory and physics of compression, *Eur. J. Pharm. Biopharm.*, 44, 269–272.
15. Hiestand, E. N. (1997), Principles, tenets and notions of tablet bonding and measurements of strength, *Eur. J. Pharm. Biopharm.*, 44, 229–242.
16. Armstrong, N. A., and Haines-Nutt, R. F. (1972), Elastic recovery and surface area changes in compacted powder systems, *J. Pharm. Pharmacol.*, 24S, 135P–136P.
17. Newton, J. M., and Rowley, G. (1973), The influence of tablet weight on compaction pressure/tablet density relations, *J. Pharm. Pharmacol.*, 25, 767–768.
18. van der Voort Maarschalk, K. (1997), *Tablet relaxation, origin and consequences of stress relief in tablet formation*, Ph.D. thesis, Riksuniversiteit Groningen.
19. Aulton, M. E., Travers, D. N., and White, P. J. P. (1973), Strain recovery of compacts on extended storage, *J. Pharm. Pharmacol.*, 25, 79P–86P.
20. York, P., and Baily, E. D. (1977), Dimensional changes of compacts after compression, *J. Pharm. Pharmacol.*, 29, 70–74.
21. van der Voort Maarschalk, K., Zuurman, K., Vromans, H., Bolhuis, G. K., and Lerk, C. F. (1997), Stress relaxation of compacts produced from viscoelastic materials, *Int. J. Pharm.*, 151, 27–34.
22. Führer, C. (1975), *Kristallographische Aspekte bei der Tablettenherstellung*, Fortbildungskurs Arbeitsgemeinschaft für Pharmazeutische Verfahrenstechnik, Braunschweig.
23. Hüttenrauch, R. (1978), Molekulargalenik als Grundlage moderner Arzneiformung, *Acta Pharm. Technol.*, 24(S6), 55–127.

24. Elamin, A., Sebhatu, T., and Ahlneck, C. (1995), The use of amorphous model substances to study mechanically activated materials in the solid state, *Int. J. Pharm.*, 119, 25–36.
25. Sebhatu, T., Ahlneck, C., and Alderborn, G. (1997), The effect of moisture content on the compression and bond-formation properties of amorphous lactose particles, *Int. J. Pharm.*, 146, 101–114.
26. Sebhatu, T., and Alderborn, G. (1999), Relationships between the effective interparticulate contact area and the tensile strength of tablets of amorphous and crystalline lactose of varying particle size, *Eur. J. Pharm. Sci.*, 8, 235–242.
27. Moldenhauer, H., Kala, H., Zessin, G., and Dittgen, M. (1980), Zur pharmazeutischen Technologie der Tablettierung, *Pharmazie*, 35, 714–726.
28. Chan, H. K., and Doelker, E. (1985), Polymorphic transformation of some drugs under compression, *Drug Dev. Ind. Pharm.*, 11, 315–332.
29. Pirttimäki, J., Laine, E., Ketolainen, J., and Paronen, P. (1993), Effects of grinding and compression on crystal structure of anhydrous caffeine, *Int. J. Pharm.*, 95, 93–99.
30. Hiestand, E. N. (1991), Tablet bond. I. A theoretical model, *Int. J. Pharm.*, 67, 217–229.
31. Hiestand, E. N., and Smith, D. P. (1991), Tablet bond. II. Experimental check of the model, *Int. J. Pharm.*, 67, 231–246.
32. Ritschel, W. A., and Bauer-Brandl, A. (2002), *Die Tablette—Handbuch der Entwicklung, Herstellung und Qualitätssicherung*, Editio Cantor Verlag, Aulendorf Germany.
33. Bogda, M. J. (2002), Tablet compression: Machine theory, design, and process troubleshooting, in Swarbrick, J., and Boylan, J. C., Eds., *Encyclopedia of Pharmaceutical Technology*, Marcel Dekker, New York, pp. 2669–2688.
34. Hausmann, R., Kaufmann, H.-J., and Richter, K. (1996), Neue Tablettiertechnologie mit zentrifugaler Matrizenfüllung und vollautomatischer Reinigung, *Pharm. Ind.*, 58, 842–846.
35. Picker, K. M. (2000), The automatic micrometer screw, *Eur. J. Pharm. Biopharm.*, 49(2), 171–176.
36. Picker, K. M. (2001), Time dependence of elastic recovery for characterization of tableting materials, *Pharm. Dev. Technol.*, 6(1), 61–70.
37. Caraballo, I., Millan, M., Fini, A., Rodriguez, L., and Cavallari, C. (2000), Percolation thresholds in ultrasound compacted tablets, *J. Controlled Release*, 69, 345–355.
38. Aulton, M. E. (1981), Indentation hardness profiles across the faces of some compressed tablets, *Pharm. Acta Helv.*, 56, 133–136.
39. Konkel, P., and Mielck, J. B. (1997), Associations of parameters characterizing the time course of the tabletting process on a reciprocating and on a rotary tabletting machine for high-speed compression, *Eur. J. Pharm. Biopharm.*, 44, 289–301.
40. Sucker, H., Fuchs, P., and Speiser, P. (1991), *Pharmazeutische Technologie*, Thieme Verlag, Stuttgart.
41. Bateman, S. D., Rubinstein, M. H., Rowe, R. C., Roberts, R. J., Drew, P., and Ho, A. Y. K. (1989), A comparative investigation of compression simulators, *Int. J. Pharm.*, 49, 209–212.
42. Celik, M., and Marshall, K. (1989), Use of a compaction simulator system in tabletting research, *Drug Dev. Ind. Pharm.*, 15, 759–800.
43. Celik, M., and Lordi, N. G. (1991), The pharmaceutical compaction research laboratory and information center, *Pharm. Technol.*, 15, 112–116.

44. Rubinstein, M. H., Bateman, S. D., and Thacker, H. S. (1991), Compression to constant thickness or constant force: Producing more consistent tablets, *Pharm. Technol.*, 1, 150–158.
45. Celik, M., Ong, J. T., Chowhan, Z. T., and Samuel, G. J. (1996), Compaction simulator studies of a new drug substance: Effect of particle size and shape, and its binary mixtures with microcrystalline cellulose, *Pharm. Dev. Technol.*, 1, 119–126.
46. Picker, K. M. (2000), Three-dimensional modeling to determine properties of tableting materials on rotary machines using a rotary tableting machine simulator, *Eur. J. Pharm. Biopharm.*, 50(2), 293–300.
47. Picker, K. M. (2003), The 3D model: Comparison of parameters obtained from and by simulating different tableting machines, *AAPS PharmSciTech.*, 4(3) article 35.
48. Levin, M. (1999), *Theory and practice of tablet press simulation for process scale-up*, paper presented at the Arden House Conference Harsiman, New York, USA.
49. Rippie, E. G., and Danielsson, D. W. (1981), Viscoelastic stress/strain behaviour of pharmaceutical tablets: Analysis during unloading and postcompression periods, *J. Pharm. Sci.*, 70, 476–482.
50. Oates, R. J., and Mitchell, A. G. (1989), Calculation of punch displacement and work of powder compaction on a rotary tablet press, *J. Pharm. Pharmacol.*, 41, 517–523.
51. Oates, R. J., and Mitchell, A. G. (1990), Comparison of calculated and experimentally determined displacement on a rotary tablet press using both Manesty and IPT punches, *J. Pharm. Pharmacol.*, 42, 388–396.
52. Oates, R. J., and Mitchell, A. G. (1994), A new method of estimating volume during powder compaction and the work of compaction on a rotary tablet press from measurements of applied vertical forces, *J. Pharm. Pharmacol.*, 46, 270–275.
53. Pudipeddi, M., Venkatesh, G., Faulkner, P., and Palepu, N. (1993), Correlations between compaction simulator and instrumented Betapress, *Pharm. Res.*, 10S, S165.
54. Ruegger, C. D. (1996), An investigation of the effect of compaction profiles on the tableting properties of pharmaceutical substances, Ph.D. thesis, Rutgers University, Newark, NJ.
55. Levin, M., Tsygan, L., and Dukler, S. (1998), U.S. Patent 6,106.262, International Patent Application No. PCT/US98/27421.
56. MCC Corp. (2000), *The Presster Binder*, Technical Information, MCC Corp., East Hanover, New Jersey, USA.
57. Lamey, K. (2000), *Correlations between compaction simulators and rotary tableting machines*, paper presented at the Compaction Simulator's User Meeting, Loughbourough, UK.
58. Guntermann, A. (2005), *The Presster—A tablet press simulator*, paper presented at TabletTech, Brussels.
59. MedelPharm (2005), *Technical information*, Medel Pharm., Bourg-en-Bresee, France.
60. Brake, E. F. (1951), *Development of methods for measuring pressures during tablet manufacture*, M.S. thesis, Purdue University, West Lafayette, IN.
61. Higuchi, T., Arnold, R. D., Tucker, S. J., and Busse, L. W. (1952), The physics of tablet compression. I. A preliminary report, *J. Am. Pharm. Assoc.*, 41, 93–96.
62. Higuchi, T., Nelson, E., and Busse, L. W. (1954), The physics of tablet compression. III: Design and construction of an instrumented tableting machine, *J. Am. Pharm. Assoc.*, 43, 344–348.
63. Morehead, W. T., and Rippie, E. G. (1990), Timing relationships among maxima of punch and die-wall stress and punch displacement during compaction of viscoelastic solids, *J. Pharm. Sci.*, 79, 1020–1022.

64. Yeh, C., Altaf, S. A., and Hoag, S. W. (1997), Theory of force transducer design optimization for die wall stress measurement during tablet compaction: Optimization and validation of split-web die using finite element analysis, *Pharm. Res.*, 14, 1161–1170.
65. Khossravi, D., and Morehead, W. T. (1997), Consolidation mechanisms of pharmaceutical solids: A multi-compression cycle approach, *Pharm. Res.*, 14, 1039–1045.
66. Knoechel, E. L., Sperry, C. C., Ross, H. E., and Lintner, C. J. (1967), Instrumented rotary tablet machines I, *J. Pharm. Sci.*, 56, 109–115.
67. Schmidt, P. C., Tenter, U., and Hocke, J. (1986), Presskraft-und Weg-Zeit-Charakteristik von Rundlauftablettenpressen. 1. Mitt.: Instrumentierung von Einzelstempeln zur Presskraftmessung, *Pharm. Ind.*, 48, 1546–1553.
68. Bauer-Brandl, A. (1998), Qualifizierung der Kraftmessung an Tablettenpressen, *Pharm. Ind.*, 60, 63–69.
69. Hoag, S. W., Nair, R., and Muller, F. X. (2000), Force-transducer-design optimization for the measurement of die-wall stress in a compaction simulator, *Pharm. Pharmacol. Commun.*, 6(7), 293–298.
70. Laich, T., and Kissel, T. (1995), Axial die-wall force minimum. Influences and significance for elastic behavior of single components and binary mixtures of excipients, *Pharm. Ind.* 57(2), 174–182.
71. Nelson, E., Busse, L. W., and Higuchi, T. (1955), The physics of tablet compression: VII. Determination of energy expenditure in the tablet compression process, *J. Am. Pharm. Assoc.* Sc. Ed., 44, 223.
72. Schmidt, P. C., and Tenter, U. (1987), Force and displacement characteristics of rotary tableting machines, *Pharm. Ind.*, 49, 637–642.
73. Matz, C. (1999), *Evaluation einer IR-telemetrischen Kraft/Weg-Instrumentierung für Rundlauf tabletten pressen*. Differenzierung des Verformungsverhaltens direktkomprimierbarer Tablettierhilfsstoffe, Dissertation, Universität Freiburg, Freiburg, Germany.
74. Matz, C., Bauer-Brandl, A., Rigassi, T., Schubert, R., and Becker, D. (1999), On the accuracy of a new displacement instrumentation for rotary tablet presses, *Drug Dev. Ind. Pharm.*, 25, 117–130.
75. Müller, F., and Caspar, U. (1984), Viskoelastische Phänomene während der Tablettierung, *Pharm. Ind.*, 46, 1049–1056.
76. Krumme, M. (1992), *Entwicklung rechnergestützter Verfahren zur Kompressions-und Festigkeitsanalyse von Tabletten*, Dissertation, Freie Universität, Berlin.
77. Krumme, M., Schwabe, L., and Frömming, K. H. (1998), Development of computerised procedures for the characterization of the tableting properties with eccentric machines. High precision displacement instrumentation for eccentric tablet machines, *Acta Pharm. Hung.*, 68, 322–331.
78. Rankell, A. S., and Higuchi, T. (1968), Physics of tablet compression. XV. Thermodynamic and kinetic aspects of adhesion under pressure, *J. Pharm. Sci.*, 58, 574–577.
79. DeCrosta, M. T., Schwartz, J. B., Wigent, R. J., and Marshall, K. (2000), Thermodynamic analysis of compact formation; compaction, unloading, and ejection. I. Design and development of a compaction calorimeter and mechanical and thermal energy determinations of powder compaction, *Int. J. Pharm.*, 198, 113–134.
80. DeCrosta, M. T., Schwartz, J. B., Wigent, R. J., and Marshall, K. (2001), Thermodynamic analysis of compact formation; compaction, unloading, and ejection. II. Mechanical energy (work) and thermal energy (heat) determinations of compact unloading and ejection, *Int. J. Pharm.*, 213, 45–62.

81. Bogs, H., and Lenhardt, E. (1971), Zur Kenntnis thermischer Vorgänge beim Tablettenpressen, *Pharm. Ind.*, 33, 850–854.
82. Hanus, E. J., and King, L. D. (1968), Thermodynamic effects in compression of solids, *J. Pharm. Sci.*, 57, 677–684.
83. Führer, C., and Parmentier, W. (1977), Zur Thermodynamik der Tablettierung, *Acta Pharm. Technol.*, 23, 205–213.
84. Ketolainen, J., Ilkka, J., and Paronen, P. (1993), Temperature changes during tabletting measured using infrared thermoviewer, *Int. J. Pharm.*, 92, 157–166.
85. Picker-Freyer, K. M., and Schmidt, A. G. (2004), Does temperature increase induced by tableting contribute to tablet quality? *J. Therm. Anal. Cal.*, 77, 531–539.
86. Beissenhirtz, M. (1974), Ermittlung der Verformungscharakteristik von Tablettierstoffen mit Hilfe elektrisch leitfähiger Zusätze, Dissertation, Universität Bonn, Bonn, Germany.
87. Schmidt, J. (1997), *Direktablettierung niedrigschmelzender nichtsteroidaler Antirheumatika mit mikrokristallinen Cellulosen*, Dissertation, Universität Halle-Wittenberg, Wittenberg, Germany.
88. Celik, M. (1992), Overview over compaction analysis techniques, *Drug Dev. Ind. Pharm.*, 18, 767–810.
89. Alderborn, G., and Nyström, C. (1996), *Pharmaceutical Powder Compaction Technology*, Marcel Dekker, New York.
90. Chilamkurti, R. N., Rhodes, C. T., and Schwartz, J. B. (1982), Some studies on compression properties of tablet matrices using a computerized instrumented press, *Drug Dev. Ind. Pharm.*, 8, 63–86.
91. Emschermann, B., and Müller, F. (1981), Evaluation of force measurements in tablet manufacture, *Pharm. Ind.*, 43, 191–194.
92. Vogel, P. J., and Schmidt, P. C. (1993), Force-time curves of a modern rotary tablet machine. Part 2. Influence of compression force and tableting speed of the deformation mechanisms of pharmaceutical substances, *Drug Dev. Ind. Pharm.*, 19, 1917–1930.
93. Schmidt, P. C., and Vogel, P. J. (1994), Force-time-curves of a modern rotary tablet machine. Part 1. Evaluation techniques and characterization of deformation behavior of pharmaceutical substances, *Drug Dev. Ind. Pharm.*, 20, 921–934.
94. Schmidt, P. C., and Leitritz, M. (1997), Compression force/time-profiles of microcrystalline cellulose, dicalcium phosphate dihydrate and their binary mixtures—A critical consideration of experiments and parameters, *Eur. J. Pharm. Biopharm.*, 44, 303–313.
95. Yliruusi, J. K., Merkku, P., Hellen, L., and Antikainen, O. K. (1997), A new method to evaluate the elastic behavior of tablets during compression, *Drug Dev. Ind. Pharm.*, 23, 63–68.
96. Dietrich, R., and Mielck, J. B. (1985), Eignung der Weibull-Funktion zur Charakterisierung des zeitabhängigen Verformungsverhaltens von Tablettierhilfsstoffen, *Acta Pharm. Technol.*, 31, 67–76.
97. Dietrich, R., and Mielck, J. B. (1985), Parametrisierung des zeitlichen Verlaufs der Verdichtung bei der Tablettierung mit Hilfe der modifizierten Weibull-Funktion, *Pharm. Ind.*, 47, 216–220.
98. Picker, K. M. (1995), Hydrophile Matrixtabletten: Tablettierung und Freisetzung-unter besonderer Berücksichtigung der relativen Feuchte während der Herstellung, Dissertation, Universität Hamburg, Hamburg, Germany.
99. Shlieout, G., Wiese, M., and Zessin, G. (1999), A new method to evaluate the consolidation behavior of pharmaceutical materials by using the Fraser-Suzuki function, *Drug Dev. Ind. Pharm.*, 25, 29–36.

100. Armstrong, N. A., and Palfrey, L. P. (1989), The effect of machine speed on the consolidation of four directly compressible tablet diluents, *J. Pharm. Pharmacol.*, 41, 149–151.

101. Ho, A. Y. K., and Jones, T. M. (1988), Rise time: A new index of tablet compression, *J. Pharm. Pharmacol.*, 40, 74P.

102. Tsardaka, E. D. (1990), Viscoelastic properties and compaction behavior of pharmaceutical particulate material, Ph.D. Thesis, University of Bath, Bath, England.

103. Dwivedi, S. K., Oates, R. J., and Mitchell, A. G. (1991), Peak offset times as an indication of stress relaxation during tableting on a rotary tablet press, *J. Pharm. Pharmacol.*, 43, 673–678.

104. Führer, C. (1962), Über den Druckverlauf bei der Tablettierung in Exzenterpressen, *Dtsch. Apoth. Ztg.*, 102, 827–842.

105. Moldenhauer, H., Hünerbein, B., and Kala, H. (1972), Recording of pressure-path diagrams from an eccentric press using piezoelectric measurement, *Pharmazie*, 27, 417–418.

106. Dürr, M., Hanssen, D., and Harwalik, H. (1972), Kennzahlen zur Beurteilung der Verpressbarkeit von Pulvern und Granulaten, *Pharm. Ind.*, 34, 905–911.

107. de Blaey, C. J., and Polderman, J. (1970), Compression of pharmaceuticals. I. The quantitative interpretation of force-displacement curves, *Pharm. Weekblad*, 105, 241–250.

108. de Blaey, C. J., and Polderman, J. (1971), Compression of pharmaceuticals. II. Registration and determination of force-displacement curves using a small digital computer, *Pharm. Weekblad*, 106, 57–65.

109. Stamm, A., and Mathis, C. (1976), Verpressbarkeit von festen Hilfsstoffen für die Direkttablettierung, *Acta Pharm.*Technol., 24(S1), 7–16.

110. Dürr, M. (1976), Bedeutung der Energiebilanz beim Tablettieren für die Entwicklung von Tabletten rezepturen, *Acta Pharm. Technol.*, 22, 185–194.

111. Parmentier, W. (1978), Investigation on the interpretation of the Kraft-Weg diagram, *Pharm. Ind.*, 40, 860–865.

112. Lammens, R. F., Polderman, J., and De Blaey, C. J. (1979), Evaluation of force displacement measurements during powder compaction. Part 1. Precision and accuracy of force measurements, *Int. J. Pharm. Technol. Prod. Manuf.*, 1, 26–35.

113. Lammens, R. F., Polderman, J., Armstrong, N. A., and De Blaey, C. J. (1980), Evaluation of force displacement measurements during powder compaction. Part 2. Precision and accuracy of powder height and displacement measurements, *Int. J. Pharm. Technol. Prod. Manuf.*, 1, 26–35.

114. Krycer, I., and Pope, D. G. (1982), The interpretation of powder compaction data—A critical review, *Drug Dev. Ind. Pharm.*, 8, 107–347.

115. Ragnarsson, G., and Sjogren, J. (1985), Force displacement measurements in tableting, *J. Pharm. Pharmacol.*, 37, 145–150.

116. Antikainen, O. K., and Yliruusi, J. K. (1997), New parameters derived from tablet compression curves. Part 2. Force-displacement curve, *Drug Dev. Ind. Pharm.*, 23, 81–93.

117. Ragnarson, G. (1996), Force-displacement and network measurements, in Alderborn, G., and Nyström, C., Eds., *Pharmaceutical Powder Compaction Technology*, Marcel Dekker, New York, pp. 77–97.

118. Heckel, R. W. (1961), An analysis of powder compaction phenomena, *Trans. Metall. Soc. AIME*, 221, 1001–1008.

119. Heckel, R. W. (1961), Density-pressure relationships in powder compaction, *Trans. Metall. Soc. AIME*, 221, 671–675.
120. Hersey, J. A., and Rees, J. E. (1970), Deformation of particles during briquetting, paper presented at Proc. 2nd Particle Size Analysis Conference of the Society of Analytical Chemistry, Bradford, 33.
121. Kawakita, K., and Lüdde, K. H. (1970/71), Some considerations on powder compression equations, *Powder Tech.*, 4, 61–68.
122. Lüdde, K. H., and Kawakita, K. (1966), Die Pulverkompression, *Pharmazie*, 21, 393–403.
123. Cooper, A. R., and Eaton, L. E. (1962), Compaction behaviour of several ceramic powders, *J. Am. Ceram. Soc.*, 5, 97–101.
124. Walker, E. (1923), The properties of powders. Part VII. The influence of the velocity of compression on the apparent compressibility of powders, *Trans. Faraday Soc.*, 19, 614–620.
125. Bal'shin, M. Y. (1938), Contribution to the theory of powder metallurgical processes, *Metalloprom.*, 18, 124–147.
126. Sönnergaard, J. M. (2000), Impact of particle density and initial volume on mathematical compression models, *Eur. J. Pharm. Sci.*, 11, 307–315.
127. Sönnergaard, J. M. (1999), A critical evaluation of the Heckel equation, *Int. J. Pharm.*, 193, 63–71.
128. Sun, C., and Grant, D. J. W. (2001), Influence of elastic deformation of particles on Heckel analysis, *Pharm. Dev. Technol.*, 6, 193–200.
129. Rees, J. E., and Rue, P. J. (1978), Time-dependent deformation of some direct compression excipients, *J. Pharm. Pharmacol.*, 30, 601–607.
130. York, P. (1979), A consideration of experimental variables in the analysis of powder compaction behavior, *J. Pharm. Pharmacol.*, 31, 244–246.
131. Vachon, M. G., and Chulia, D. (1999), The use of energy indices in estimating powder compaction functionality of mixtures in pharmaceutical tableting, *Int. J. Pharm.*, 177, 183–200.
132. Krumme, M., Schwabe, L., and Frömming, K. H. (2000), Development of computerised prodedures for the characterisation of the tableting properties with eccentric machines: Extended Heckel analysis, *Eur. J. Pharm. Biopharm.*, 49, 275–286.
133. Humbert-Droz, P., Gurny, R., Mordier, D., and Doelker, E. (1983), Densification behavior of drugs having bioavailability problems, *Int. J. Pharm.*, 4, 29–35.
134. Imbert, C., Tchoreloff, P., Leclerc, B., and Couarraze, G. (1997), Indices of tableting performance and application of percolation theory to powder compaction, *Eur. J. Pharm. Biopharm.*, 44, 273–282.
135. Kuentz, M., and Leuenberger, H. (2000), A new approach to tablet strength of a binary mixture consisting of a well and a poorly compactable substance, *Eur. J. Pharm. Biopharm.*, 49, 151–159.
136. Gabaude, C. M. D., Guillot, M., Gautier, J. C., Saudemon, P., and Chulia, D. (1999), Effects of true density, compacted mass, compression speed, and punch deformation on the means yield pressure, *J. Pharm. Sci.*, 88, 725–730.
137. Duberg, M., and Nyström, C. (1986), Studies on the compression of tablets. XVII. Porosity-pressure curves for the characterization of volume reduction mechanisms, *Powder Technol.*, 46, 67–75.
138. Paronen, P. (1986), Heckel-plots as indicators of elastic properties of pharmaceuticals, *Drug Dev. Ind. Pharm.*, 12, 1903–1912.
139. Morris, L. E., and Schwartz, J. B. (1995), Isolation of densification regions during powder compression, *Drug Dev. Ind. Pharm.*, 21, 427–446.

140. Kuentz, M., Leuenberger, H., and Kolb, M. (1999), Fracture in disordered media and tensile strength of microcrystalline cellulose tablets at low relative densities, *Int. J. Pharm.*, 182, 243–255.

141. Tsardaka, K. D., Rees, J. E., and Hart, J. P. (1988), Compression and recovery behaviour of compacts using extended Heckel plots, *J. Pharm. Pharmacol.*, 40, 73P.

142. Tsardaka, K. D., and Rees, J. K. (1990), Relations between viscoelastic parameters and compaction properties of two modified starches, *J. Pharm. Pharmacol.*, 42, 77P.

143. Picker, K. M. (2000), A new theoretical model to characterize the densification behavior of tableting materials, *Eur. J. Pharm. Biopharm.*, 49(3), 267–273.

144. Picker-Freger, K. M. (2007), The 3-D Model: Experimental testing of the parameters d, e, and w and validation of the analysis, *J. Pharm. Sci.*, 96(5), 1408–1417.

145. Hoblitzell, J. R., and Rhodes, C. T. (1986), Preliminary investigations on the parity of tablet compression data obtained from different instrumented presses, *Drug Dev. Ind. Pharm.*, 12, 507–525.

146. David, S. T., and Augsburger, L. L. (1977), Plastic flow during compression of directly compressible fillers and its effect on tablet strength, *J. Pharm. Sci.*, 66, 155–159.

147. Müller, F., and Caspar, U. (1984), Viskoelastische Phänomene während der Tablettierung, *Pharm. Ind.*, 46, 1049–1056.

148. Müller, F. (1996), Viscoelastic models, in Alderborn, G., and Nyström, C., Eds., *Pharmaceutical Powder Compaction Technology*, Marcel Dekker, New York, pp. 99–132.

149. Bauer, A. D. (1991), Untersuchungen zur Prozessdatengewinnung, Viskoelastizität und Struktur von Tabletten, Dissertation, Universität Bonn, Bonn, Germany.

150. Staniforth, J. N., Baichwal, A. R., and Hart, J. P. (1987), Interpretation of creep behavior of microcrystalline cellulose powders and granules during compaction, *Int. J. Pharm.*, 40, 267–269.

151. Staniforth, J. N., and Patel, C. J. (1989), Creep compliance behavior of direct compression excipients, *Powder Technol.*, 57, 83–87.

152. Tsardaka, K. D., and Rees, J. E. (1989), Plastic deformation and retarded elastic deformation of particulate solids using creep experiments, *J. Pharm. Pharmacol.*, 41, 28P.

153. Malamataris, S., and Rees, J. E. (1993), Viscoelastic properties of some pharmaceutical powders compared using creep compliance, extended Heckel analysis and tablet strength measurements, *Int. J. Pharm.*, 92, 123–135.

154. Rees, J. E., and Tsardaka, K. D. (1994), Some effects of moisture on the viscoelastic behaviour of modified starch during powder compaction, *Eur. J. Pharm. Biopharm.*, 40, 193–197.

155. Picker, K. M., and Bikane, F. (2001), An evaluation of three-dimensional modeling of compaction cycles by analyzing the densification behavior of binary and ternary mixtures, *Pharm. Dev. Technol.*, 6(3), 333–342.

156. Picker, K. M. (2004), The 3D model: Explaining densification and deformation mechanisms by using 3D parameter plots, *Drug Dev. Ind. Pharm.*, 30(4), 413–425.

157. Picker, K. M. (2003), The 3D model: Does time plasticity represent the influence of tableting speed? *AAPS PharmSciTech.*, 4(4), article 66.

158. Hauschild, K., and Picker, K. M. (2003), The 3D model: Comparison with other data analysis techniques for tableting, *AAPS PharmSci*, 5(S1), W4304.

159. Hauschild, K., and Picker-Freyer, K. M. (2004), Evaluation of a new coprocessed compound based on lactose and maize starch for tablet formulation, *AAPS PharmSci*, 6(2), article 16.

160. Picker, K. M. (2002), New insights in the process of tablet formation—Ways to explore soft tableting, Martin-Luther-University Halle-Wittenberg and Görich und Weiershäuser Verlag, Marburg, Germany.
161. Wurster, D. E., Rowlings, C. E., and Creekmore, J. R. (1995), Calorimetric analysis of powder compression: I. Design and development of a compression calorimeter, *Int. J. Pharm.*, 116, 179.
162. Picker, K. M. (2003), The relevance of glass transition temperature for the process of tablet formation, *J. Therm. Anal. Cal.*, 73(2), 597–605.
163. Armstrong, N. A., and Haines-Nutt, R. F. (1972), Elastic recovery and surface area changes in compacted powder systems, *J. Pharm. Pharmacol.*, 24, 135–136.
164. Picker-Freyer, K. M. (2005), Carrageenans: Analysis of tablet formation and properties (Part 1), *Pharm. Technol. Eur.*, 17(8), 37–44.
165. Picker-Freyer, K. M. (2005), Carrageenans: Analysis of tablet formation and properties (Part 2), *Pharm. Technol. Eur.*, 17(9), 32–44.
166. Schmidt, A. G., and Picker-Freyer, K. M. (2006), Influence of mechanical activation on the micro structure of widely used excipients for solid dosage forms, paper presented at the 5th World Meeting of Pharmaceutics, Pharmaceutical Technology and Biopharmaceutics, Geneva.
167. Leuenberger, H., Hiestand, E. N., and Sucker, H. (1981), Contribution to the theory of powder compression, *Chemie Ing. Tech.*, 53(1), 45–47.
168. Bauer-Brandl, A. (2002), Tooling for tableting, in Swarbrick, J., and Boylan, J. C., Eds., *Encyclopedia of Pharmaceutical Technology*, Marcel Dekker, New York.
169. Conte, U., Colombo, P., Caramella, C., Ferrari, F., Gazzaniga, A., Guyot, J. C., La Manna, A., and Traisnel, M. (1988), Influence of tablet weight control systems during production on biopharmaceutical properties of the tablets, *Acta Pharm. Technol.*, 34(2), 63–67.
170. Cole, G. C. (1998), *Pharmaceutical Production Facilities: Design and Applications*, Taylor and Francis, London.
171. Lyon, R. C., Lester, D. S., Lewis, E. N., Lee, E., Yu, L. X., Jefferson, E. H., and Hussain, A. S. (2002), Near-infrared spectral imaging for quality assurance of pharmaceutical products: Analysis of tablets to assess powder blend homogeneity, *AAPS PharmSciTech*, 3(3), article 17.
172. Lai, C. K., Zahari, A., Miller, B., Katstra, W. E., Cima, M. J., and Cooney, C. L. (2004), Nondestructive and on-line monitoring of tablets using light-induced fluorescence technology, *AAPS PharmSciTech*, 5(1), article 3.
173. Lee, T. H., and Lin, S. Y. (2004), Microspectroscopic FT-IR mapping system as a tool to assess blend homogeneity of drug-excipient mixtures, *Eur. J. Pharm. Sci.*, 23(2), 117–122.
174. U.S. Food and Drug Administration, Center for Drug Evaluation and Research, Office of Pharmaceutical Science, Process Analytical Technology (PAT) Initiative, assessed on 12/20/2007, available: www.fda.gov/cder/OPS/PAT.htm.
175. Oeser, W. H., and Sander, A. (1992), *Pharma-Betriebsverordnung: Grundregeln für die Herstellung von Arzneimitteln (GMP)*, Wissenschaftliche Verlagsgesellschaft Stuttgart, Germany.
176. Hyde, J. M. (1985), New developments in CIP practices, *Chem. Eng. Prog.*, 81(1), 39–41.
177. Picker, K. M., and Mielck, J. B. (1998), Effect of relative humidity during tableting on matrix formation of hydrocolloids: Densification behaviour of cellulose ethers, *Pharm. Dev. Technol.*, 3(1), 31–41.

178. Hauschild, K., and Picker-Freyer, K. M. (2006), Evaluation of tableting and tablet properties of Kollidon SR: The influence of moisture and mixtures with theophylline monohydrate, *Pharm. Technol. Dev.*, 11(1), 125–140.
179. Kachrimanis, K., Nikolakakis, I., and Malamataris, S. (2003), Tensile strength and disintegration of tableted silicified microcrystalline cellulose: Influences of interparticle bonding, *J. Pharm. Sci.*, 92(7), 1489–1501.
180. Steendam, R., Eissens, A. C. E., Frijlink, H. W., and Lerk, C. F. (2000), Plasticization of amylodextrin by moisture. Consequences for drug release from tablets, *Int. J. Pharm.*, 204(1–2), 23–33.
181. Hanssen, D., Führer, C., and Schaefer, B. (1970), Evaluation of magnesium stearate as tablet lubricating agent by electronic pressure measurements, *Pharm. Ind.*, 32(2), 97–101.
182. Bolhuis, G. K., and Lerk, C. F. (1977), Film forming of tablet lubricants during the mixing process of solids, *Acta Pharm. Technol.*, 23(1), 13–20.
183. Zuurman, K., Van der Voort Maarschalk, K., and Bolhuis, G. K. (1999), Effect of magnesium stearate on bonding and porosity expansion of tablets produced from materials with different consolidation properties, *Int. J. Pharm.*, 179(1), 107–115.
184. Van Der Voort Maarschalk, K., and Bolhuis, G. K. (1998), Improving properties of materials for direct compaction. Part II, *Pharm. Technol. Eur.*, 10(10), 28–36.
185. Shotton, E., and Lewis, C. J. (1964), Effect of lubrication on the crushing strength of tablets, *J. Pharm. Pharmacol.*, 16S, 111–120.
186. Van Der Voort Maarschalk, K., and Bolhuis, G. K. (1999), Improving properties of materials for direct compaction, *Pharm. Technol.*, 23(5), 34–46.
187. Steffens, K. J., Mueller, B. W., and List, P. H. (1982), Tribological laws and results in tablet production. 7. Studies on magnesium stearate commercial products, *Pharm. Ind.*, 44(8), 826–830.
188. Hoelzer, A. W. (1984), Batch to batch variations of commercial magnesium stearates. Chemical, physical and lubricant properties, *Labo-Pharma*, 338, 28–36.
189. Miller, T. A., and York, P. (1985), Frictional assessment of magnesium stearate and palmitate lubricant powders, *Powder Technol.*, 44(3), 219–226.
190. Lindberg, N. O. (1972), Evaluation of some tablet lubricants, *Acta Pharm. Suec.*, 9(3), 207–214.
191. Alpar, O., Deer, J. J., Hersey, J. A., and Shotton, E. (1969), Possible use of poly(tetrafluo roethylene) (Fluon) as a tablet lubricant, *J. Pharm. Pharmacol.*, 21S, 6–8.
192. Shah, N. H., Stiel, D., Weiss, M., Infeld, M. H., and Malick, A. W. (1986), Evaluation of two new tablet lubricants—Sodium stearyl fumarate and glyceryl behenate. Measurement of physical parameters (compaction, ejection and residual forces) in the tableting process and the effect of the dissolution rate, *Drug Dev. Ind. Pharm.*, 12(8–9), 1329–1346.
193. N'Diaye, A., Jannin, V., Berard, V., Andres, C., and Pourcelot, Y. (2003), Comparative study of the lubricant performance of Compritol HD5 ATO and Compritol 888 ATO: effect of polyethylene glycol behenate on lubricant capacity, *Int. J. Pharm.*, 254(2), 263–269.
194. Staniforth, J. N. (1987), Use of hydrogenated vegetable oil as a tablet lubricant, *Drug Dev. Ind. Pharm.*, 13(7), 1141–1158.
195. Dawoodbhai, S. S., Chueh, H. R., and Rhodes, C. T. (1987), Glidants and lubricant properties of several types of talcs, *Drug Dev. Ind. Pharm.*, 13(13), 2441–2467.
196. Roescheisen, G., and Schmidt, P. C. (1995), Preparation and optimization of L-leucine as lubricant for effervescent tablet formulations, *Pharm. Acta Helv.*, 70(2), 133–139.

197. Jahn, T., and Steffens, K.-J. (2005), Press chamber coating as external lubrication for high speed rotary presses: Lubricant spray rate optimization, *Drug Dev. Ind. Pharm.*, 31(10), 951–957.

198. Laich, T., and Kissel, T. (1998), Automatic adaptation of lubricant quantity by control of an external lubrication. Tests on a reciprocating and on a rotary tablet press, *Pharm. Ind.*, 60(10), 896–904.

199. Otsuka, M., Sato, M., and Matsuda, Y. (2001), Comparative evaluation of tableting compression behaviors by methods of internal and external lubricant addition: Inhibition of enzymatic activity of trypsin preparation by using external lubricant addition during the tableting compression process, *AAPS PharmSci*, 3(3), article 20.

200. Jetzer, W., and Leuenberger, H. (1984), Determination of capping tendency in pharmaceutical active ingredients and excipients, *Pharm. Acta Helv.*, 59(1), 2–7.

201. Bauer, K. H., Frömming, K. H., and Führer, C. (2002) *Lehrbuch der pharmazeutischen Technologie*, Wissenschafteiche Verlagsgesellschaft, Stuttgart, Germany.

# CONTROLLED RELEASE OF DRUGS FROM TABLET COATINGS

Sacide Alsoy Altinkaya
*Izmir Institute of Technology, Urla-Izmir, Turkey*

**Contents**

1 Introduction
2 Tablet Coating Methods
3 Characterization of Tablet Coatings
4 Preparation of Asymmetric Membranes
5 Methods for Optimization of Tablet Coating Formulations
6 Applications
  6.1 Materials and Methods
  6.2 Results and Discussion
7 Conclusion
  References
  Appendix

## 1 INTRODUCTION

Controlled-release technology for drug delivery applications is designed to target the drug to particular places or cells in the body, to overcome certain tissue and cellular barriers, and to control the duration and level of the drug in the body within a speficied therapeutic window. This usually implies achieving a prolonged, zero-order release rate over the desired duration of drug delivery. Controlled-release dosage forms provide sustained drug release and require less frequent drug loading than conventional forms. Thus, the toxic side effects of the drug are minimized and

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

patient's convenience and compliance are improved. Controlled-release systems are usually classified into four categories based on the rate-limiting step of the release process [1]: (1) diffusion-controlled systems, (2) chemically controlled systems, (3) swelling-controlled systems, and (4) magnetically controlled systems. Diffusion-controlled systems are formulated in two main geometries: reservoirs and matrices. In matrix systems, the drug is generally uniformly distributed or dissolved throughout a polymer. The release kinetics from these types of systems depend on the quantity and the diffusivity of drug in the matrix and the geometry of the system. The release rates from matrix systems usually decrease with time due to an increase in the path length for diffusion of drug and, thus, their release characteristics are not generally zero order. In the case of chemically controlled systems, the drug is either distributed through a biodegradable polymer or chemically bound to a polymer backbone chain. The drug release from the biodegradable polymer is controlled by degradation of the polymer through penetration of water or a chemical reaction [1]. The drug attached to the polymer is released by hydrolytic or enzymatic cleavage. In swelling-controlled systems, the drug is dissolved or distributed in a glassy polymer matrix. As water penetrates the dry matrix, the polymer swells and its glass transition temperature decreases below the temperature in the matrix. As a result, the glassy–rubbery transition occurs and it allows the release of the drug. The release rate of the drug out of these systems is mainly controlled by the rate and degree of swelling. Alternatively, the drug release can also be controlled magnetically by dispersing the drug and small magnetic beads within a polymer matrix and exposing them to an oscillating external magnetic field. (The drug is released as the matrix is exposed to an oscillating external magnetic field.)

Reservoir systems consist of a drug-containing core surrounded by a polymer membrane. The release rate of drug is controlled by its diffusion through the membrane [1, 2]. In addition to diffusional release, osmotic pumping mechanisms contribute to the total drug release rate if either the drug is highly soluble or an osmotic agent is added to the active core [1–4]. Reservoir systems are prepared in different geometries, such as coated tablets, beads, particles, membrane-based pouches, and microcapsules. The main advantage of these systems is their ability to maintain zero-order release rates [1, 2]. This is usually achieved by loading the powdered form of drug at a level far above the solubility of drug. Then, the concentration of drug at the internal wall of the reservoir becomes its saturation concentration and zero-order release occurs until the drug is completely dissolved. In addition, drug loading can be higher for these systems compared to other controlled-release systems; thus, the cost of formulation is minimized and the drug is released at a higher rate for a longer period of time. The major disadvantage of the reservoir system is rupture of the rate-controlling membrane if it is subject to dose dumping.

Numerous studies exist in the literature on the drug release from tablet coatings. In the majority of these studies, diffusional drug release takes place from the coatings [5–17]. Various factors such as the type [8, 10, 15], and concentration of the coating material [7, 8, 11], the type and amount of the plasticizer in the coating solution [10, 12, 15], the thickness of the coating [5], the composition of the tablet core [6, 14], the particle size of the coating material [10, 18], and the weight gain of the coating [7, 8, 11, 13, 16] were considered during the formulation of tablet coat-

ings, and their effects on the drug release rates were investigated. The use of osmotic tablet coating systems are also described in the literature [3, 4, 17, 19–23]. Osmotic systems utilize osmotic pressure difference as a driving force, and in the simplest design they consist of an osmotic core containing drug with and without osmotic agents. Different studies have shown that the release rate of drugs from oral osmotic pump tablet coating systems is governed by various formulation variables such as solubility and osmotic pressure of core components [3, 4, 19–23], number [17] and size of the delivery orifice [3, 4, 20, 21, 23], drug loading [3, 4, 17] thickness of the coating [3, 4, 19, 21, 23], composition of the coating solution [3, 4, 17, 20, 21, 23], and weight of the coating [17, 20, 22]. The coatings prepared in most of these studies have dense structures with or without a hole drilled through the coating through which the drug is delivered. In some cases, drug delivery ports are formed by adding leachable materials to the coating [24, 25]. The main problem with these systems, in the absence of a hole, is an excessively prolonged drug release due to the low permeability of the coating. To increase permeability of the coatings, plasticizers and water-soluble additives were incorporated in the coating solution, and multilayer composite coatings [14, 15, 24, 26, 27] or multiple-compartment osmotic tablets were prepared [17, 22, 28, 29]. The permeability of the tablet coating systems were further improved by changing the structure of the coatings from dense to porous asymmetric ones [30–42]. Asymmetric membrane tablet coatings are characterized by a relatively thin, dense skin layer supported on a highly permeable, thicker, and porous sublayer. The permeabilities and the release rates of the drugs through the asymmetric-membrane capsule/tablet coatings were determined to be higher compared to conventional dense tablet coatings [30, 33, 34, 39]. The composition of the coating solution was found to have a significant effect on the structure of these type of coatings and thus on the release rate of the drug [30, 33, 35, 36, 39, 40–42]. In addition, it was reported that the drying condition had a significant impact on the structure of the asymmetric-membrane tablet coating and the release rate of the drug [41]. A review of the literature studies clearly indicates that asymmetric-type tablet coating is a new solution for developing controlled drug delivery systems, since the structure of these types of coatings can be varied easily by changing the preparation conditions without altering the coating material or significantly varying the coating thickness.

This work contains six sections in addition to the introduction. Sections 2– 5 present a brief review of manufacturing methods for the application of coating materials on tablets, characterization methods used to evaluate the uniformity and defects of the tablet coatings, the techniques commonly used for manufacturing asymmetric-type membranes, and modeling approaches employed for optimization of the tablet coating formulations. The aim of the last two sections is to demonstrate the advantages of asymmetric membrane tablet coatings with respect to their drug release properties and the factors that affect the morphology of these types of tablet coatings. To achieve this goal, the in vitro release of a model compound, theophylline, from asymmetric membrane tablet coatings is determined and the morphology of the coatings is examined. In addition, the dynamics of the phase inversion is quantified in terms of ternary-phase diagrams coupled with composition paths determined from a mathematical model developed previously by our group [43]. To draw meaningful and objective conclusions from experimental data and derive

an empirical expression for the release rate of drug, compositions of the coating solution are chosen using a statistical experimental design.

## 2 TABLET COATING METHODS

Four basic methods are commonly used for the application of coating materials on drug tablets: (1) pan coating, (2) fluidized-bed coating, (3) compression (press) coating, and (4) melt/dry coating.

***Pan Coating*** The pan coating process is the oldest form of pharmaceutical coating for manufacturing small, coated particles or tablets. In this process, the particles are tumbled in a pan which is rotated at an angle of usually 45° to the horizontal surface at a speed between 20 and 50 rpm [44]. Coating fluid is sprayed onto the particles from a above by means of an air jet. The hot air blown through the coater evaporates the fluid and dries the film coating. Pan coating is generally preferred to coat large tablets since they are exposed to mechanical damage in other coating operations. The uniformity of the coating applied to the tablets and defects on the coating are important issues from a practical point of view. Operating conditions such as drum speed, drum solids loading, the presence/absence of baffles, air velocity, and temperature influence the movement of tablets within the moving bed, the circulation and surface exposure times spent by a tablet in the bulk of the moving bed and on the surface of the bed, respectively, and the rate of drying, which all in turn determine the uniformity of the coating in a pan coating process [45]. Different experimental tools such as light emission from a single luminous tablet, photographic and manual counting, positron emission particle tracking, magnetic resonance imaging, and near real-time video imaging techniques [45–51] were used to study the motion of tablets in the drum. Simulation of particle movement using discrete-element modeling has also been used to study the movement of tablets [52–55]. In addition, statistical experimental design was utilized to identify the critical processing variables that affect coating uniformity and loading of active agent coated on tablets [56]. Inlet airflow, pan speed, inlet air temperature, coating time, atomization pressure, and fan pressure were investigated as the process variables. Among these variables, atomization pressure, pan speed, and duration of coating were found to be critical process variables with respect to uniformity of the coating.

***Fluidized-Bed Coating*** The basic principle of fluidized-bed coating is to suspend tablets in a moving hot-air stream in the upward direction during the coating process. The coating material is sprayed through a nozzle from the top, the side, or the bottom into the fluidized bed. The solvent in the solution is removed from the coating by the hot-air stream, which also carries the coated tablets/particles. Fluidized-bed coating provides better coating uniformity due to good solid–fluid mixing and minimizes formation of agglomerates. There are various types of fluidized-bed coating equipment, the most commonly used configuration being the Wurster column coater [44], in which a draft tube insert is placed coaxially in the bed to aid the circulation of particles. This column is not suitable for coating of large particles and tablets due to high erosion of solids associated with the higher velocity

needed to circulate them [57]. Typical operating variables that affect the performance of the fluid bed coating in a Wurster column are airflow rate, air and bed temperature, spray rate, gap height between the draft tube and the air distributor plate, atomizing air pressure, humidity, and solids charge. The product coating uniformity in the Wurster column coating process is primarily determined by the variation in coating per pass and the circulation time distribution. Radioactive particle tracing, magnetic particle tracing, and controlled single- or multiple-pass coating evaluation techniques were commonly used for detecting particle circulation time and distribution [57, 58]. In addition, several models have been developed to predict the amount of material coated on the particles [59–65].

***Compaction Coating*** In this process, the coating is compressed around a preformed core by using a special tablet press. Compaction coating is especially useful when the drug itself is unstable in the solution and precipitates from the solution in a less stable morphology. On the other hand, the large amount of coating material required limits the applicability of this technology. Another disadvantage of the process is poor adhesion at the coating–core interface which causes physical instability (i.e., friability) [66]. The Press coating process is not useful for coating relatively hard cores which provide essentially no compressibility.

***Melt/Dry Coating*** Film coating processes require water or organic solvent(s). The use of organic solvents causes environmental pollution and excessive cost of recovery while a long time is required to remove the aqueous solvent. Both hot-melt and dry coating techniques eliminate the use of solvents; as a result the processing times become much shorter and the cost of the process is reduced. Melt coating is possible for coating materials that have a low melting temperature and acceptable thermostability. The principal stages in the film formation during dry coating are softening, melting, and curing [67–70]. The process requires larger amount of plasticizer to partially soften and dissolve the polymer.

## 3 CHARACTERIZATION OF TABLET COATINGS

Tablet coatings are applied to improve tablet swallowability, mask unpleasant tastes and odors, protect the tablet core against water and oxygen, which can degrade the drug in the core, and control the release rate of the drug. The rate of dissolution and bioavailability of a drug are primarily influenced by the quantity and quality of the coating applied on the tablet. Thus, the characterization of the coatings becomes essential in terms of the uniformity and integrity of the coating. A number of instrumental methods ranging from liquid chromatography [71] to various noninvasive spectroscopic probes have been introduced and evaluated as a means of monitoring the coating process. Among these methods, laser-induced breakdown spectroscopy (LIBS) has been used as a rapid technique for assessing the uniformity of the coating thickness [72]; however, the destructive nature of the method has also been reported [73]. Near-infrared spectroscopy has been employed for determining the amount of coating applied [74–76]. The main disadvantage of this method has been identified as imprecision in the calibration and validation models due to uneven distribution of coating from tablet to tablet [77]. Microscopy

techniques provide direct and accurate measurement of the coating thickness but require laborious sample preparation and are therefore impractical for real-time process monitoring. Recently, Raman spectroscopy combined with multivariate data analysis has been reported as a feasible noninvasive technique to quantify tablet coating thickness and uniformity in the presence of a fluorescent ingredient in the coating formulation [77, 78]. Confocal laser scanning microscopy has been introduced as a novel technique for imaging the film–core interface and surface defects of film-coated tablets [79]. Surface roughness was also determined as an important factor in characterizing the tablet coatings using different imaging and roughness analytical techniques, including optical microscopy, scanning electron microscopy (SEM), laser profilometry, and atomic force microscopy (AFM) [80–84].

## 4 PREPARATION OF ASYMMETRIC MEMBRANES

Asymmetric membranes are usually produced by phase inversion techniques. In these techniques, an initially homogeneous polymer solution becomes thermodynamically unstable due to different external effects and the phase separates into polymer-lean and polymer-rich phases. The polymer-rich phase forms the matrix of the membrane, while the polymer-lean phase, rich in solvents and nonsolvents, fills the pores. Four main techniques exist to induce phase inversion and thus to prepare asymmetric porous membranes [85]: (a) thermally induced phase separation (TIPS), (b) immersion precipitation (wet casting), (c) vapor-induced phase separation (VIPS), and (d) dry (air) casting.

***Thermally Induced Phase Separation*** In the TIPS process, an initially homogeneous solution consisting of a polymer and solvent(s) phase separates due to a decrease in the solvent quality when the temperature of the casting solution is decreased. After demixing is induced, the solvent is removed by extraction, evaporation, or freeze drying.

***Immersion Precipitation (Wet Casting)*** A homogeneous polymer solution consisting of a polymer and solvent(s) is cast on a support and is immersed in a nonsolvent bath. During the immersion, casting solvent diffuses into the nonsolvent bath and, countercurrently, nonsolvent in the bath penetrates into the solution. The nonsolvent has a limited solubility in the polymer, and when it reaches its critical concentration in the solution, precipitation takes place. Then, the solvent and nonsolvent in the solution are extracted and film is annealed.

***Vapor-Induced Phase Separation*** During the VIPS process, phase separation is induced by penetration of nonsolvent vapor, into the homogeneous polymer solution consisting of polymer and solvent(s). Mass transfer is usually much slower than that in the wet casting process; thus, the VIPS process has been used to obtain membranes with symmetric, cellular, and interconnected pores [86, 87].

***Dry (Air) Casting*** In this process, the polymer dissolved in a mixture of a volatile solvent and a less volatile nonsolvent is cast on a support and exposed to an air

stream. During drying of the solution, fast solvent evaporation leads to a decrease in solubility of the polymer, then phase separation takes place.

## 5 METHODS FOR OPTIMIZATION OF TABLET COATING FORMULATIONS

Tablet coating formulation is composed of various formulation factors and process parameters. The formulation is usually divided into modeling and optimization phases. The modeling phase consists of preparing series of experimental formulations by varying the ingredients and process conditions systematically and measuring their properties. A detailed review of the literature indicates that the response surface method (RSM) has been widely used for modeling and choosing acceptable tablet coating and other pharmaceutical formulations. The RSM includes statistical experimental designs, multiple regression analysis, and optimization algorithms to search the best formulation for a given set of constraints. Full factorial, orthogonal, Box–Behnken, central composite, pseudorandom, and Plackett–Burman designs were used to investigate the effects of tablet core formulation, coating thickness, and process parameters such as mixing time and speed in the pan coating process, inlet airflow, pan speed, inlet air temperature, coating time, atomization pressure, and fan pressure in the fluidized-bed coating process [88–92]. Artifical neural networks (ANNs) have also been investigated as an alternative method for modeling the pharmaceutical formulations [93–105]. In the RSM, the pharmaceutical responses are predicted based on the second-order polynomial equation, which is usually limited to low levels. When a nonlinear relationship between formulation factors and response variables is observed, the ANN approach was shown to give better estimations of optimal formulations [105].

## 6 APPLICATIONS*

Previous sections have reviewed numerous studies in the literature investigating the release mechanisms from the tablet coating systems, their advantages/disadvantages as a control release system, and the methods used to characterize and optimize their formulation. In the following sections, the in vitro performance characteristics of the asymmetric membrane tablet coatings will be illustrated using release studies of a model compound theophylline. For this purpose, first the method used for preparing asymmetric membrane tablet coatings will be explained, then the results of the dissolution studies will be discussed.

### 6.1 Materials and Methods

***Preparation and Characterization of Tablet Coatings*** Tablet cores were prepared by compressing the drug without any excipient using a hydraulic press operated at 110 MPa. A stainless steel die with a diameter of 1.2 cm was used to produce 400-mg drug tablet cores. Asymmetric-membrane tablet coatings were applied

*This article was published in Biochemical Engineering Journal, 28, Sacide Alsoy Altinkaya, Hacer Yenal, In vitro drug release rates from asymmetric-membrane tablet coatings: Prediction of phase-inversion dynamics, 131–139, Copyright Elsevier (2006).

using a dip coating process (Dip Coater Nima, type D1L, serial no. 327). The tablets were dip coated in polymer solutions prepared by dissolving cellulose acetate (Aldrich) in a solution of acetone (Merck) and water. The rate of withdrawal of the tablets from the solution was adjusted to obtain similar final coating thicknesses for each coating formulation. Immediately after coating, tablets were rotated for even distribution of the viscous membrane solution, transferred into an environmental chamber (Angelantoni Industrie, Italy, Challenge series, model number CH250), and kept there for 2 h to allow for evaporation of both the solvent (acetone) and nonsolvent (water). The temperature and relative humidity of air in the environmental chamber were adjusted to 25°C and 50%, respectively. Tablets were allowed to dry further for a minimum of 24 h at room temperature prior to dissolution experiments. Faster evaporation of acetone and resulting increase in the concentration of water in the coating leads to a decrease in solubility of the cellulose acetate (CA); then phase separation takes place. Consequently, asymmetric-membrane structure forms on the tablet core. Morphology of the coatings was examined using a scanning electron microscope (Philips, XL-30SFEG). Samples were coated with gold palladium using a magnetron sputter coating instrument. The thickness of the dense skin layer, the overall porosity, and the average pore size were determined from image analysis of micrographs showing cross sections of the membranes.

***Dissolution Studies*** The release rate of theophylline from the tablets was determined by the U.S. Pharmacopeia (USP) XXIII dissolution methodology using a dissolution tester (Caleva 10ST). According to this standard, 900 ml of dissolution medium was placed in the vessel and the temperature was maintained constant at 37°C using a constant-temperature bath. Then, the tablets were immersed in the vessel and the solution was stirred at a speed of 50 rpm. To simulate the actual dissolution environment in the body, the pH of the dissolution medium was kept at 3 during the first 3.5 h by adding 8.5 vol % phosphoric acid to 900 mL distilled water and then increased to 7.4 by adding 5.3 *M* NaOH to the dissolution medium and kept at this value until the end of the experiment. To determine the quantity of drug released from the tablets, samples were taken periodically and assayed by ultraviolet (UV) spectrophotometry (Shimadzu UV-1601) at a wavelength of 272 nm. Dissolution experiments were performed on three tablets and the release profiles were reported as the arithmetic average of the three experimental runs.

***Statistical Experimental Design*** To determine the influence of the composition of the coating solution on the release rate of drug, experiments were statistically designed using a commercial software package called Design-Expert [106]. The system studied in this chapter consists of three components with the following compositional restrictions:

$$5 \le \omega_1 \le 15 \qquad 70 \le \omega_2 \le 90 \qquad 5 \le \omega_3 \le 15 \tag{1}$$

where $\omega_i$ is the weight percent of component $i$ and 1, 2, 3 represent CA, acetone, and water, respectively. Any composition outside these limits will probably fail to

**TABLE 1 Theophylline Release Rates from Asymmetric-Membrane Tablet Coatings, Results of Fitting Release Profiles to Zero-Order Kinetics, and Precipitation Times Determined from Model Predictions**

| Compositon (wt %) | | | | | |
|---|---|---|---|---|---|
| Cellulose Acetate | Acetone | Water | Release Rate (mg/min) | $R^2$ | Precipitation Time[a](s) |
| 15 | 80 | 5 | 0.036 | 0.9871 | 2671/2671 |
| 15 | 80 | 5 | 0.036 | 0.9864 | 2671/2671 |
| 5 | 90 | 5 | 0.45 | 0.9757 | 660/660 |
| 5 | 90 | 5 | 0.36 | 0.9864 | 660/660 |
| 15 | 70 | 15 | 0.036 | 0.9908 | 3374/3374 |
| 15 | 70 | 15 | 0.027 | 0.9925 | 3374/3374 |
| 5 | 80 | 15 | 0.36 | 0.9928 | 1000/1000 |
| 5 | 80 | 15 | 0.36 | 0.9876 | 1000/1000 |
| 5 | 85 | 10 | 0.27 | 0.9918 | 675/675 |
| 10 | 80 | 10 | 0.054 | 0.9958 | 1751/1751 |
| 15 | 75 | 10 | 0.054 | 0.9975 | 2314/2314 |
| 10 | 85 | 5 | 0.036 | 0.9887 | 2554/2554 |
| 12.5 | 77.5 | 10 | 0.036 | 0.9902 | — |
| 10 | 75 | 15 | 0.063 | 0.9889 | 1484/1484 |

[a]The first number corresponds to the precipitation time at the tablet–coating interface.

produce a successful asymmetric-membrane coating. In mixture experiments, the factors are the compositions of the mixture components, and the sum of the fractions of all components is equal to 1. Therefore the factor levels are mutually dependent. Thus, factorial experimental designs are not suitable for response surface modeling of mixtures since such designs require that the experimental treatment combinations be determined by independent adjustments of each component level. In addition, a standard response surface design cannot be used either due to the same constraints. Consequently using the constraint levels shown in Equation (1), a *D*-optimal design was generated by the Design-Expert software package. The 14 experimental formulations determined are shown in Table 1. The lower and upper limits on the weight fraction of each component are required to (a) obtain appropriate viscosity of the solution and coat the tablets uniformly and (b) induce phase separation, thus forming a porous membrane structure. These constraints were established based on preliminary dissolution experiments, available literature data, and the simulation results reported by Altinkaya and Ozbas [43].

Of the 14 formulations listed in Table 1, six experimental runs were required to fit the quadratic mixture model, four additional distinct runs were used to check for the lack of fit, and finally four runs were replicated to provide an estimate of pure error. Design-Expert used the vertices, the edge centers, the overall centroid, and one point located halfway between the overall centroid and one of the edge centers as candidate points. Additionally, four vertices of the design region were used as check points [106].

***Determination of Phase Diagrams and Composition Paths*** The dynamics of the membrane formation process is predicted by combining the kinetics and thermody-

namics of the system simultaneously. An appropriate thermodynamic model is necessary to construct the ternary-phase diagram and to formulate the boundary conditions of the kinetic model. Phase separation is considered to occur when a mass transfer path touches the binodal curve. In this study, a robust algorithm developed previously by our group was used to construct the phase diagram [43]. The algorithm utilizes Flory Huggins thermodynamic theory with constant interaction parameters. The composition paths were determined from the kinetic model equations which consist of coupled unsteady-state heat and mass transfer equations, film shrinkage, and complex boundary conditions. The details of both the thermodynamics and kinetic equations can be found in our previous study [43]. We have assumed that the kinetic model derived for a plane geometry can be used to predict the membrane formation process on a tablet surface. This assumption is fairly reasonable since the thickness of the coating is very small and, thus, the cylindrical geometry can be approximated as the plane geometry.

## 6.2 Results and Discussion

***Effect of Composition of Coating Solution*** The effect of changing the composition of the casting solution is well documented for asymmetric membranes prepared for separation applications [107–114]. However, there are relatively few quantitative studies illustrating the relationship between the composition of the casting solution and the drug release rate from the asymmetric-membrane coated tablets/capsules [30, 33]. To investigate such a relationship, the in vitro release profiles of the model drug theophylline were measured for the 14 formulations listed in Table 1 and they are shown in the Appendix in Figures A1–A10. To find out whether the drug release from the tablet coatings provides a zero-order release kinetics, each data set was fitted to a linear equation. The quality of the fitted model is determined by the coefficient of determination $R^2$ and it is defined as

$$R^2 = \frac{\sigma_y - \hat{\sigma}_y}{\sigma_y} \tag{2}$$

where

$$\sigma_y = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \bar{Y})^2 \tag{3}$$

$$\hat{\sigma}_y = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2 \tag{4}$$

denote the sample variance and the prediction error power, respectively. Additionally, $n$ is the number of experimental data points, $Y_i$ is the experimental observation, $\bar{Y}$ is the average of the experimental data points, and $\hat{Y}_i$ denotes the predicted value by the fitted model. The quantity $R^2$ lies between 0 and 1, and if the value is 1, it can be said that the fit of the model is perfect. High $R^2$ values listed in Table 1 for each data set indicate that there is an excellent linear relationship between the concentration of the drug and the release time; thus, all tablet coatings prepared

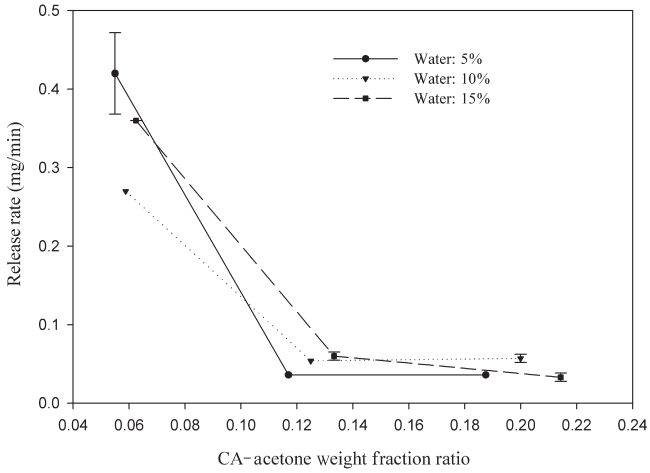


**FIGURE 1** Release rate of theophylline as function of cellulose acetate(P)–acetone(S) weight fraction ratios.

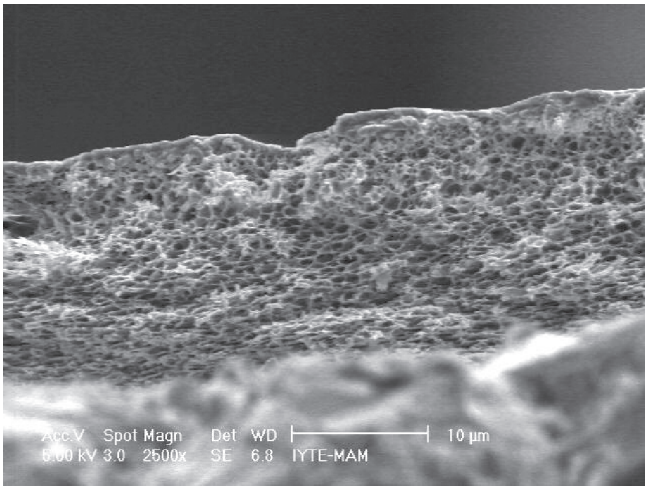


**FIGURE 2** SEM of cross section of asymmetric-membrane made with 5% water and CA–acetone weight ratio of 5/90, magnification 2500×.

can provide zero-order or near-zero-order drug release. The release rates for each coating formulation were estimated from the slope of the average release profiles and they are also listed in Table 1.

The drastic change in the release rates with the composition of the coating solution is shown in Figure 1. Within the experimental composition range covered, the highest release rate was observed in the case of the lowest CA (P: polymer) and the highest acetone (S: solvent) concentrations in the casting solution [polymer (P)–solvent (S) weight fraction ratio 5/90]. This is caused by the final coating structure consisting of a very thin and dense top skin layer, and highly porous lower sublayer, as illustrated in Figure 2. At the lowest level of polymer concentration (5%), the

thicknesses of the dense skin layers of the coatings are very small and similar to each other; thus, the structure of the lower sublayer becomes an important factor in determining the release rate of the drug. The porosity of the coating structure in the case of 90% acetone is so high that an increase in water concentration from 5 to 10% is not sufficient to produce a more porous final structure and, thus, the release rate of the drug decreases. A further increase of water concentration to 15% results in an opposite effect, which makes the release rate increase to a level of 0.36 mg/min due to the dominant effect of water concentration in increasing the porosity of the lower sublayer.

An increment in the CA concentration from 5 to 10% resulted in significantly lower release rates at all water concentrations because the structure of the coatings changed from porous to dense. Keeping the polymer concentration at 10% while changing the water concentration from 5 to 15% makes the release rates increase. This behavior is explained by the formation of more porous structures by adding more nonsolvent into the casting solution, which is in agreement with the observations of other groups [30, 33].

When the polymer concentration increases from 10 to 15%, no significant changes in release rates were observed. As a matter of fact, in the cases of 5% and 10% water concentrations, the release rates did not change at all. This is due to the unusual transport characteristics of asymmetric membranes which are complex functions of the properties of the different regions of the membrane. In addition to the thickness of the dense skin layer and the porosity of the lower sublayer, structural factors such as tortuosity, pore size, shape, and connectivity of the pores also strongly affect the rate of transport through the coating. The SEM pictures taken at high magnification which are shown in Figures 3*a* and 3*b* indicate that the tablet coating prepared with a P/S ratio of 10/85 has a uniform and narrow pore size distribution with regular elliptic pore shapes, while the other one (P/S: 15/80) has cylindrical pores with a wide pore size distribution, forming a connected pore network. As a result, even though the resistance of the dense skin layer of the coating made with P/S ratio of 15/80 is larger, its lower sublayer resistance is smaller due to the connected pore network. Consequently, the release rate of the theophylline becomes the same through both tablet coatings. Comparison of the scanning electron micrographs in Figures 4*a* and 4*b* indicates that the tablet coating made with the P/S ratio of 10/80 has a uniform pore size distribution, cylindrical pore shapes and high tortuosity while the tablet coating prepared with the P/S ratio of 15/75 has elliptic, irregular pore shapes and pores are isolated. Thus, the lower sublayer resistance of the coating made with the P/S ratio of 10/80 is larger due to the relatively higher tortuosity resulting in the same release rate with the coating prepared with the P/S ratio of 15/75.

At the 15% polymer concentration level, the release rate increases with the change in water concentration from 5 to 10%, which is mainly caused by the increased porosity. However, a further 5% increment in the water concentration (to 15%) made the release rate decrease back to the same level as in the case of 5% water concentration. Even though higher water concentration favors forming a more porous structure and a concomitant higher release rate, the acetone level which decreased below a critical value destroyed this mechanism.

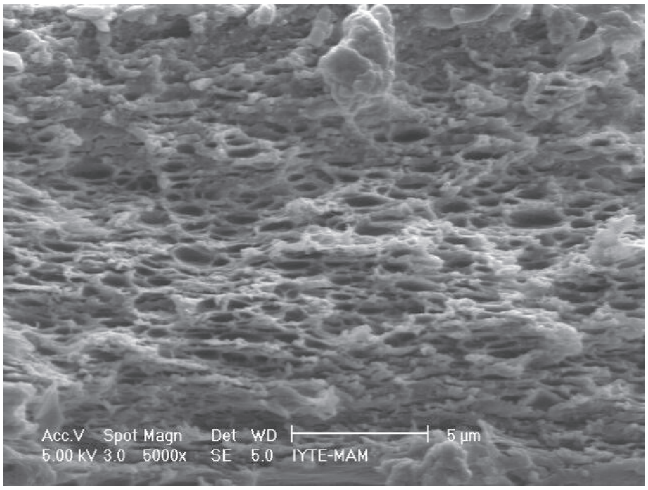


(*a*)

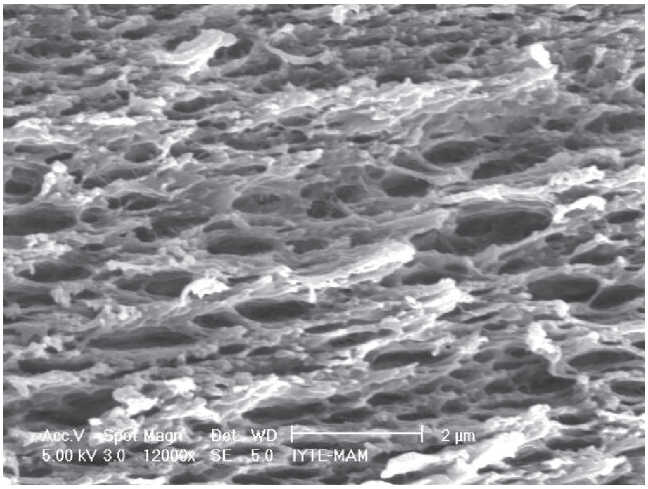


(*b*)

**FIGURE 3** SEM of cross section of asymmetric-membrane made with 5% water and CA/acetone weight ratios of: (*a*) 10/85, magnification 5000× (*b*) 15/80, magnification 12,000×.

Results of dissolution studies along with morphological observations clearly indicate that the drug release rate is strongly influenced by the morphology of the membrane. Thus, if one wishes to control the drug release characteristics of the delivery system, a quantitative understanding of the dynamics and morphology of the phase inversion process is required. The dynamics of the phase inversion process can be quantified in terms of the ternary-phase diagram coupled with the heat and mass transfer model equations. We have obtained information about the structure of the tablet coating by plotting the composition paths on the ternary-phase diagram and the polymer concentration profile at the moment of precipitation. As an illustra-

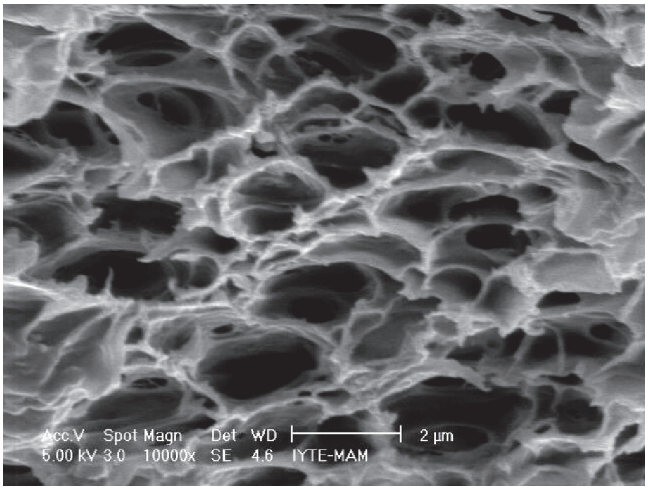


(*a*)

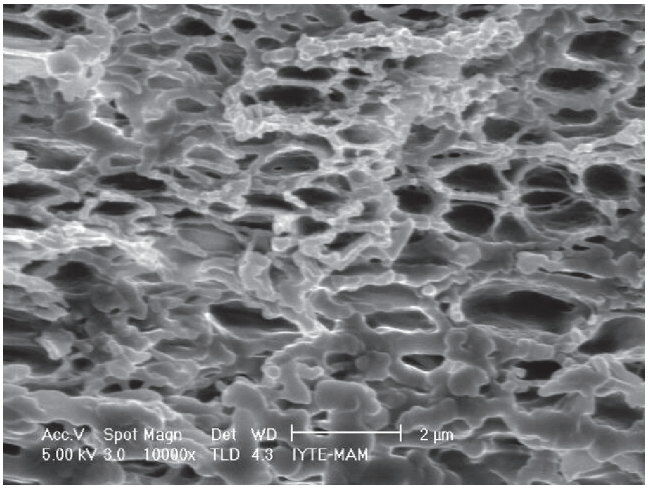


(*b*)

**FIGURE 4** SEM of the cross section of the asymmetric-membrane made with 10% water and CA–acetone weight ratios of (*a*) 10/80, magnification 10,000×, and (*b*) 15/75, magnification 10,000×.

tion, in Figure 5, concentration paths in time for the tablet coating prepared with 5% CA, 90% acetone, and 5% water are shown. According to this plot, the phase separation takes place since the concentration paths in time for the drug tablet–coating and the coating–air interface cross the binodal curve at the same time. In addition, the coating–air interface enters the phase envelope at a polymer volume fraction of 0.76, while the tablet–coating interface enters with a volume fraction of 0.023. These two observations imply that the coating will be porous and asymmetric in which the top layer is more dense than the lower sublayer, which was confirmed by the SEM picture in Figure 2. The predictions, have shown that phase separation was achieved for all coating formulations, supporting the morphological observa-

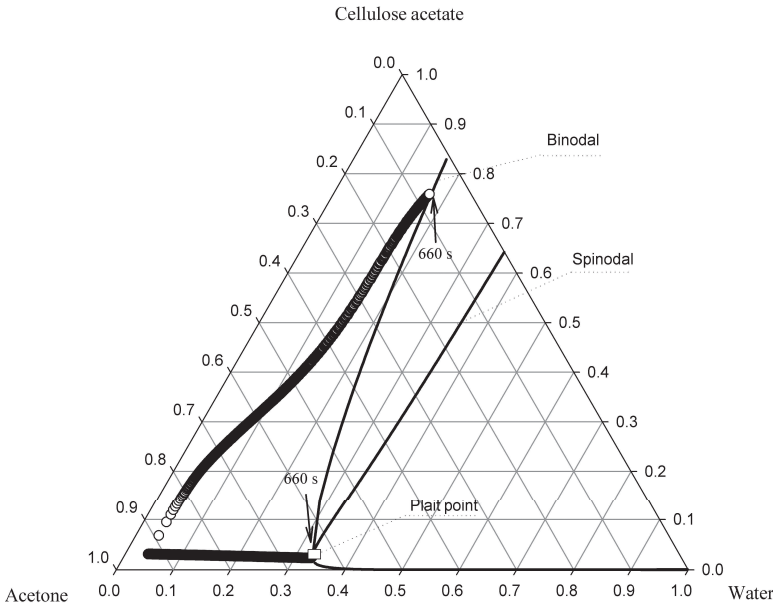


**FIGURE 5** Ternary-phase diagram and concentration paths for coating made with 5% CA dissolved in 90% acetone and 5% water.

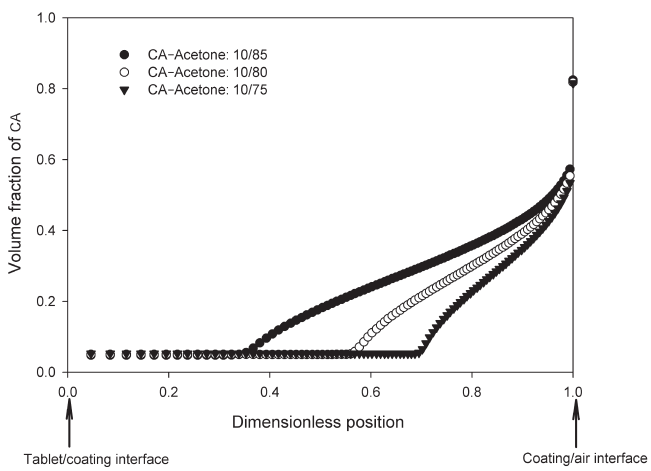


**FIGURE 6** Concentration profile of CA in membrane at moment of precipitation for different CA–acetone weight fraction ratios.

tions and the precipitation times calculated for each case (listed in Table 1). Model predictions can also be used to determine a rough thickness of the high-polymer-concentration region near the coating–air interface and the pore distribution of the sublayer structure when the polymer concentration profiles at the moment of phase separation are plotted. As an illustration, such a plot is shown in Figure 6 for coatings prepared with 10% CA in the casting solution. Examination of these profiles

leads to the following conclusions regarding the effect of increased water concentration: (1) The polymer concentration at the coating–air interface slightly decreases. (2) More uniform porosity distribution throughout the lower sublayer is favored. Also, porosity increases, which is in complete agreement with our release studies and the observations of other groups [30, 33].

***Effect of Evaporation Condition*** Previous studies on more traditional applications have investigated the effect of increased air velocity, that is, forced-convection conditions for a combination of dry/wet phase inversion techniques to produce defect-free, ultrahigh flux asymmetric membranes with ultrathin skin layers [115–117]. To investigate the effect of evaporation condition on the release rate of drug, tablets were dip coated with CA solution containing 10% CA, 80% acetone, and 10% water and allowed to dry by blowing air across the surface with a blower (forced convection). As a comparison, tablets coated with the same solution were air dried under natural free-convection conditions.

As illustrated in Figure 7, the release profiles of both tablet coatings show a linear behavior only at small times, and then exponential increases in concentrations were observed. Based on this behavior, the release profiles were fitted to an empirical equation as

$$C = k_0 t \lfloor 1 + e^{b_0 C t} \rfloor \tag{5}$$

where $k_0$ and $b_0$ are fitting parameters. The accuracy of Equation (5) for correlating the release rate data in Figure 7 was confirmed by high $R^2$ values, very close to 1 in both cases. Due to the presence of the second term in Equation (5), the release rate

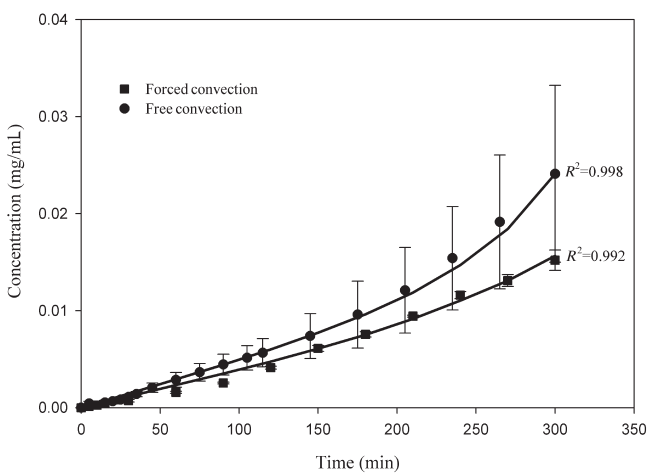


**FIGURE 7** Release of theophylline from tablet coatings made with 10% CA dissolved in 80% acetone and 10% water. Coated tablets were dried under free- and forced-convection conditions. The lines correspond to prediction from Equation (5) using $k_0 = 1.92 \times 10^{-5}$ mg/(mL·min); $b_0 = 0.1179$ mL/(mg·min) for forced convection and $k_0 = 2.41 \times 10^{-5}$ mg/(mL·min); $b_0 = 0.1174$ mL/(mg·min) for free convection.

of the drug, $dC/dt$, is no longer constant and its dependency on the concentration of drug in the dissolution medium, $C$, can be expressed as

$$\frac{dC}{dt} = \frac{k_0 \lfloor 1 + e^{b_0 Ct}(1 + b_0 Ct) \rfloor}{1 - k_0 b_0 t^2 e^{b_0 Ct}} \tag{6}$$

Using Equation (6), the release rate of theophylline from the tablets dried under forced- and free-convection conditions were determined as 0.047 and 0.078 mg/min, respectively. It shoud be noted that both of these values correspond to the arithmetic average of the release rates calculated at each average concentration level shown in Figure 7. The difference in release rates can be explained by comparing the scanning electron micrographs shown in Figures 8*a* and 8*b*. It can be seen that the cross-sectional morphology of the tablet coating dried under the forced-convection condition is dense and nonporous while a porous and asymmetric structure is observed for the tablet coating dried under free-convection conditions. In the dense coating, diffusional resistance to transport of the drug occurs through the overall thickness and is larger than that in the asymmetric porous coating; hence, a lower drug release rate is observed. To understand the effect of air velocity on the formation of the coating structure, we have utilized our model predictions. The composition paths plotted in Figure 9 indicate that, when the speed of air in the drying atmosphere is significantly increased, the rate of evaporation of solvent (acetone) increases dramatically and within a short time its concentration at the surface drops to zero. This situation leads to very strong diffusional resistance within the membrane solution and, thus, slow evaporation of the nonsolvent (water). Consequently, phase separation is never achieved and the resulting membrane structure becomes dense as supported by the SEM picture in Figure 8*a*.

***Effect of Nonsolvent Type*** A few studies in the literature have shown that various membrane morphologies can be obtained by changing the type of nonsolvent in the casting solution [118–120]. To investigate the effect of nonsolvent type, tablet coatings were prepared from a casting solution of CA in acetone as a solvent and octanol, formamide, glycerol, and hexanol as nonsolvents. The drastic change in the release rates of theophylline and percentage of dense skin layer at the surface of the tablet coating with nonsolvent type is shown in Figure 10. The decrease in the release rates is associated with the increase in the thickness of dense skin layer of the coating. The results illustrated in Figure 10 indicate that the membrane structure formation during the phase inversion process is strongly influenced by the type of nonsolvent, since each nonsolvent has different volatility, thermodynamic, and diffusion characteristics. The difference in the rates of evaporation and diffusion of nonsolvents and change in the miscibility gaps in the case of each nonsolvent lead to different mass transfer paths; consequently, the morphology of the resulting membrane structures significantly varies from porous symmetric to dense asymmetric ones.

***Statistical Analysis of Experimental Design*** The effect of the composition of the coating solution on the release rate of drug was investigated in detail with the 14 formulations listed in Table 1. The drug release rate was chosen as an appropriate

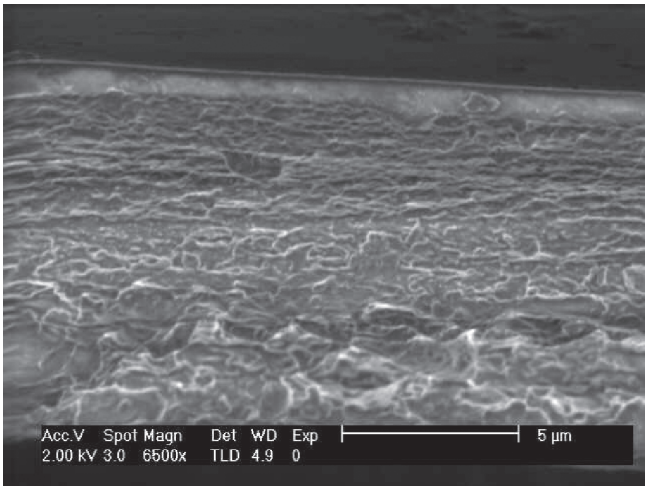


(*a*)

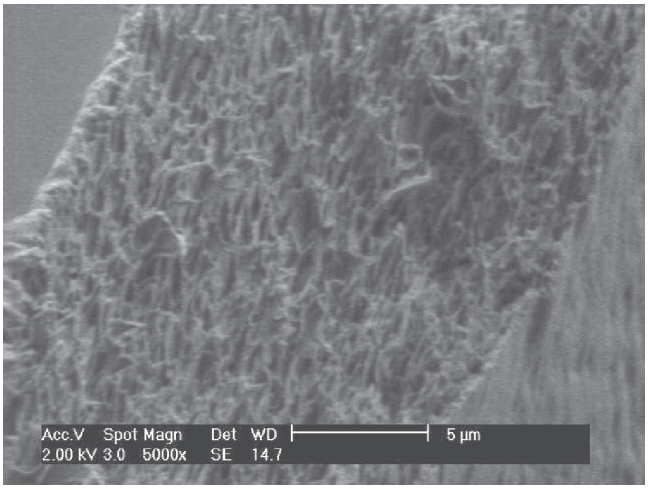


(*b*)

**FIGURE 8** SEM of cross section of asymmetric-membrane made with 10% CA dissolved in 80% acetone and 10% water: (*a*) coating solution was dried under forced-convection condition, magnification 6500×; (*b*) coating solution was dried under free-convection condition, magnification 5000×.

response variable since zero-order release was easily achieved for all tablet coatings prepared. The data in Table 1 were best fit by the special cubic equation

$$\begin{aligned}\text{Release rate} = {} & 70.94\omega_1 + 1.698\omega_2 + 37.55\omega_3 - 94.66\omega_1\omega_2 \\ & - 354.3\omega_1\omega_3 - 49.3\omega_2\omega_3 + 357.66\omega_1\omega_1\omega_3 \end{aligned} \tag{7}$$

Results of the regression analysis are given in Table 2. An excellent fit of the experimental release rate data to Equation (7) was confirmed by the high $R^2$ value of 0.9801. In addition to the $R^2$ values, the significance of Equation (7) and each

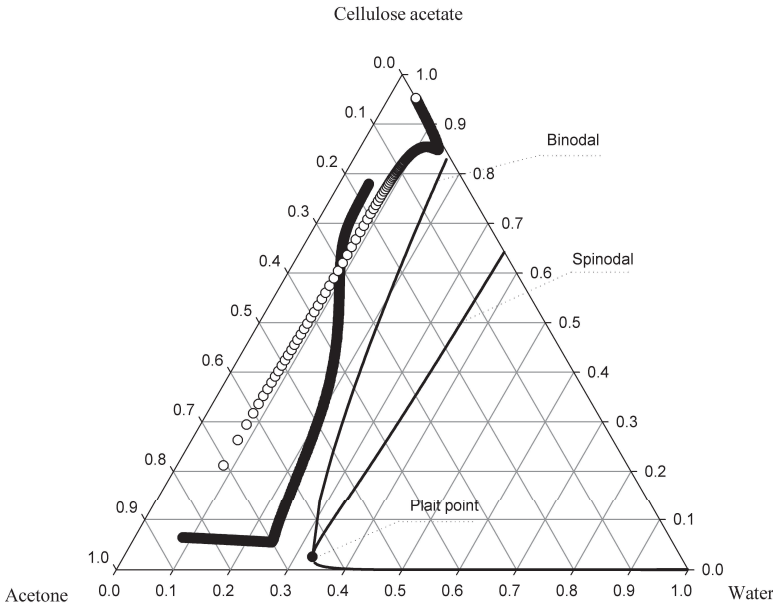


**FIGURE 9** Ternary-phase diagram and concentration paths for coating made with 10% CA dissolved in 80% acetone and 10% water. Coating solution was dried under forced-convection condition.

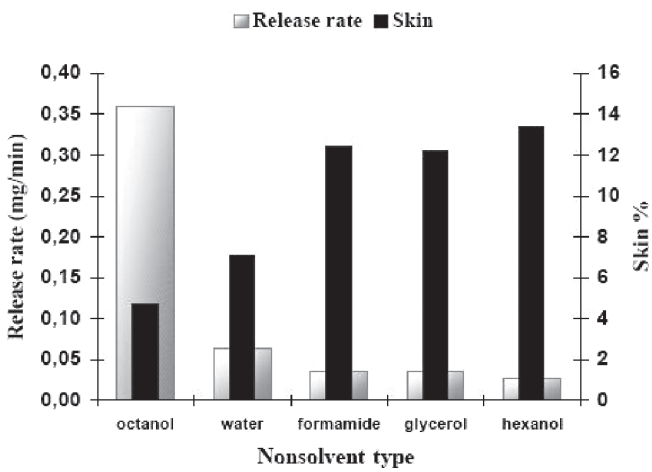


**FIGURE 10** Release rates of theophylline and percentage of dense skin layer of membranes as function of nonsolvent type. Polymer: CA solvent–acetone.

term in it to the prediction of the release rate of theophylline was evaluated by the $F$ statistic or $F$ value. The $F$ statistic is viewed as a ratio that expresses variance explained by the model divided by variance due to model error or experimental error and is defined as

$$F = \frac{R^2/k}{(1-R^2)/(n-k-1)} \tag{8}$$

**TABLE 2 Statistical Analysis of Release Rate Data**

| | | | $\alpha = 0.05$ | | $\alpha = 0.01$ | |
|---|---|---|---|---|---|---|
| Model Number | $R^2$ | $F$ values | $F_{critic}$[a] | $\Delta F$ | $F_{critic}$[a] | $\Delta F$ |
| 1 (full model) | 0.9801 | 42.21 | 4.21 | 38 | 8.26 | 33.95 |
| 2 ($\omega_1\omega_2\omega_3$) | 0.955 | 25.13 | 3.87 | 21.26 | 7.19 | 17.94 |
| 3 ($\omega_2\omega_3$) | 0.953 | 23.57 | 3.87 | 19.7 | 7.19 | 16.38 |
| 4 ($\omega_1\omega_3$) | 0.946 | 20.64 | 3.87 | 16.77 | 7.19 | 13.45 |
| 5 ($\omega_1\omega_2$) | 0.881 | 8.65 | 3.87 | 4.78 | 7.19 | 1.46 |

[a]Determined from statistical tables [121].

where $k$ is the number of variables in the model. Usually the computed value of $F$ is compared with the critical $F$ value, $F_{k,\,n-k,\,1-\alpha}$, where $\alpha$ is a preselected significance level. If the value of $F$ is substantially greater than the critical $F$ value ($F_{critic}$), that is, if $\Delta F$, the difference between $F$ and $F_{critic}$, is large, then the regression equation is considered useful in predicting the response. We have assessed the contribution of each interaction term by comparing the change in $\Delta F$ and $R^2$ values between the full model given in Equation (7) and reduced models. The reduced models were obtained by deleting a specific interaction term in the full model; for example, model 2 includes all terms in Equation (7) except the term involving $\omega_1\omega_2\omega_3$. The results in Table 2 indicate that all binary and ternary interaction terms in the full model are needed for accurate prediction of the release rate since the largest $\Delta F$ values are calculated for the full model for both significance levels, $\alpha = 0.05$ and $\alpha = 0.01$. According to the criterion mentioned above, among all interactions, CA–acetone ($\omega_1\omega_2$) was identified as the most influential factor on the response since the largest decrease in both $\Delta F$ and $R^2$ values compared to those of the full model were observed when the term $\omega_1\omega_2$ was deleted from the full model. This simply implies that changing the CA–acetone ratio in the coating formulation has the most significant effect on the release rate. Specifically, increasing the ratio of CA to acetone from 5/90 to 15/80 resulted in a decrease of the release rate from 0.45 to 0.036 mg/min since the porosity of the membrane decreases and the thickness of the dense skin layer increases. The ratio of the composition of CA to water ($\omega_1\omega_3$) was also found to be an important parameter on the release rate of drug as indicated by the second largest decrease in $R^2$ and $\Delta F$ values compared to those of the full model. Decreasing this ratio from 15/5 to 5/15 caused an increase in the release rate by a factor of 10 since the thickness of the dense skin layer significantly decreases. Based on the decrease in $\Delta F$ and $R^2$ values from those of the full model given in Table 2, the relative importance of each interaction term can be ranked as follows: $\omega_1\omega_2 > \omega_1\omega_3 > \omega_2\omega_3 > \omega_1\omega_2\omega_3$.

To further illustrate the simultaneous effects of the factors on the release rate of drug, a three-dimensional response surface plot based on Equation (7) was generated, as shown in Figure 11. As can be clearly seen from this figure, the release rate can be significantly varied just by tailoring the CA–acetone and CA–water ratios without changing the coating material. In addition, Figure 11 shows that a slight maximum in release rate is observed as the ratio of composition of acetone to water increases.

In order to validate the predictive capability of the empirical expression, two formulations with compositions given in Table 3 were selected randomly from the

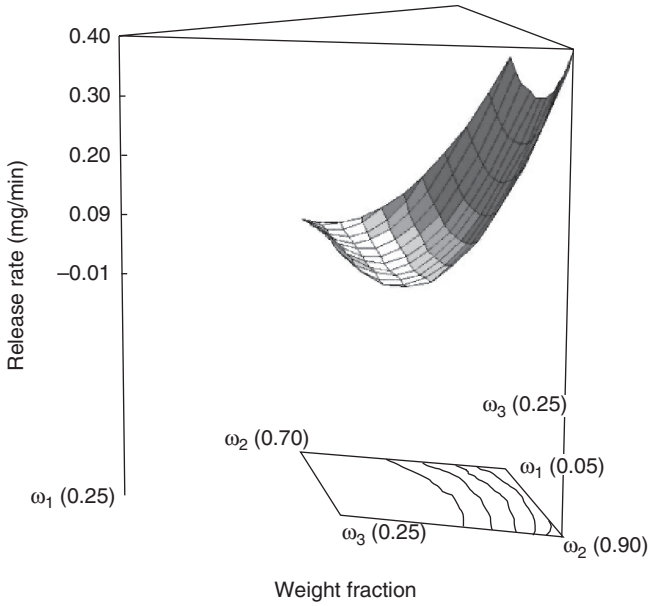


**FIGURE 11** Three-dimensional response surface plot of release rate as function of composition of coating solution.

**TABLE 3 Composition of Coating Solution Randomly Selected for Testing Predictive Capability of Empirical Expression**

| Compositon (wt %) | | | Experimental Release Rate (mg/min) | Predicted Release Rate (mg/min) | $R^2$ |
|---|---|---|---|---|---|
| Cellulose Acetate | Acetone | Water | | | |
| 7.5 | 82.5 | 10 | 0.09 | 0.125 | 0.9816 |
| 10 | 82.5 | 7.5 | 0.027 | 0.029 | 0.9858 |

experimental design region. Experimental release rates estimated from the slope of the release profiles and corresponding rates predicted from Equation (7) are also listed in Table 3. According to the results, the maximum absolute percentage difference between the experimental and predicted release rates is 3.5%. This value lies within the residuals obtained in deriving Equation (7). Based on this comparison, it is fair to conclude that the empirical expression derived in this study can be used as a tool to predict the release rate of theophylline for any composition within the experimental design region.

## 7 CONCLUSION

This chapter has considered the controlled release of drugs from tablet coating systems. These systems are still the preferred route of drug administration due to

their advantages of ease of administration, maintaining zero-order release rates, and better patient compliance; thus, they hold the major market share in the pharmaceutical industry along with the other formulations of oral drug delivery. In tablet coated systems, diffusion of drug through the coating is usually the rate-limiting step and the desired drug release rates are achieved by properly selecting the coating material and adjusting the morphology of the coating using suitable manufacturing methods. Coating characteristics such as glass transition temperature, crystalline content, and degree of cross-linking significantly determine the release rate of the drug. For a selected coating material, the structure of the coating plays a critical role in achieving useful release rates. The morpholology of the tablet coatings can be varied by incorporating plasticizer or water-soluble additives in the coating solution, by blending the polymers, or by applying multilayer composite coatings. Another approach is to apply asymmetric- and porous-type coatings. The morphological characteristics of these types of coatings, such as fraction of dense top layer and porous sublayer and size and shape of the pores, can be varied by optimizing the composition of the coating solution, evaporation conditions (temperature, relative humidity, and velocity of air), and the type of solvent or nonsolvent used in the coating solution. Asymmetric-type coatings can be used to facilitate osmotic delivery of drugs with low solubilities since high water fluxes can be achieved. These coatings allow us to control the release kinetics without altering the coating material or significantly varying the coating thickness.

Currently, considerable research efforts have been directed toward developing protein drug delivery systems due to discovery of numerous protein and peptide therapeutics. The delivery of protein drugs is usually limited to parenteral administration and frequent injections are required due to their short half-lives in the blood. In this regard, development of oral delivery systems is necessary for patient compliance and convenience. The challenge in the design of the oral delivery systems is that they should protect the incorporated drugs from chemical and enzymatic degradation until the drug reaches the delivery site. The protein drug should not be influenced by pH or bacteria and enzymes along the gastrointestinal (GI) tract and should be delivered at the desired site with a desired efficiency. To achieve site-specific delivery for protein- and peptide-based drugs, one of the strategies that has often been investigated is to coat the drug core with polymers that can respond to the stimuli of local environments such as pH and enzymes. For water-soluble protein drugs, such as insulin, an additional protective coating is usually required to isolate the drug from the surrounding water. Composite tablet coating materials which combine the enzymatic susceptibility and protective properties of polymers can be another solution for this problem. A systematic comparison study among various polymer combinations is required to select the appropriate coating materials for specific drugs. Not only the selection of the coating material but also the manufacturing technique for the preparation of tablet coatings plays a critical role and still remains one of the most challenging subjects in the controlled drug delivery area. Proteins are very sensitive to conditions that can occur during the coating process. Mechanical stresses during the preparation, exposure to a hydrophobic organic solvent, intermediate moisture level during hydration, and interaction between protein and polymer can easily inactivate the protein-based drug. Therefore, more research focused on both optimization of coating materials and manufacturing methods for encapsulating the protein and peptide drugs is necessary.

## REFERENCES


1. Langer, R. S., and Peppas, N. A. (1981), Present and future applications of biomaterials in controlled drug delivery systems, *Biomaterials*, 2, 201–214.
2. Ho, W. S. W., and Sirkar, K. K. (1992), *Membrane Handbook*, Van Nostrand Reinhold, New York, pp. 915–935.
3. Liu, L., Ku, J., Khang, G., Lee, B., Rhee, J. M., and Lee, H. B. (2000), Nifedipine controlled delivery by sandwiched osmotic tablet system. *J. Controlled Release*, 68, 145–156.
4. Liu, L., Khang, G., Rhee, J. M., and Lee, H. B. (2000), Monolithic osmotic tablet system for nifedipine delivery, *J. Controlled Release*, 67, 309–322.
5. Narisawa, S., Nagata, M., Ito, T., Yoshino, H., Hirakawa, Y., and Noda, K. (1995), Drug release behavior in gastrointestinal tract of beagle dogs from multiple unit type rate-controlled or time-controlled release preparations coated with insoluble polymer-based film, *J. Controlled Release*, 33, 253–260.
6. Moussa, I. S., and Cartilier, L. H. (1997), Evaluation of cross-linked amylose press-coated tablets for sustained drug delivery, *Int. J. Pharm.*, 149, 139–149.
7. Khan, M. Z. I., Prebeg, E., and Kurjakovic, N. (1999), A pH-dependent colon targeted oral drug delivery system using methacrylic acid copolymers: I. Manipulation of drug release using Eudragit® L100-55 and Eudragit® S100 combinations, *J. Controlled Release*, 58, 215–222.
8. Krögel, I., and Bodmeier, R. (1999), Floating or pulsatile drug delivery systems based on coated effervescent cores, *Int. J. Pharm.*, 187, 175–184.
9. Macleod, G. S., Fell, J. T., and Collett, J. H. (1999), An in vitro investigation into the potential for bimodal drug release from pectin/chitosan/HPMC-coated tablets, *Int. J. Pharm.*, 188, 11–18.
10. Thoma, K., and Bechtold, K. (1999), Influence of aqueous coatings on the stability of enteric coated pellets and tablets, *Eur. J. Pharm. Biopharm.*, 47, 39–50.
11. Fukui, E., Uemura, K., and Kobayashi, M. (2000), Studies on applicability of press-coated tablets using hydroxypropylcellulose (HPC) in the outer shell for timed-release preparations, *J. Controlled Release*, 68, 215–223.
12. Crotts, G., Sheth, A., Twist, J., and Ghebre-Sellassie, I. (2001), Development of an enteric coating formulation and process for tablets primarily composed of a highly water-soluble, organic acid. *Eur. J. Pharm. Biopharm.*, 51, 71–76.
13. Ofori-Kwakye, K., and Fell, J. T. (2003), Biphasic drug release from film-coated tablets, *Int. J. Pharm.*, 250, 431–440.
14. Goto, T., Tanida, O., Yoshinaga, T., Sato, S., Ball, D. J., Wilding, I. R., Kobayashi, E., and Fujimura, A. (2004), Pharmaceutical design of a novel colon-targeted delivery system using two-layer-coated tablets of three different pharmaceutical formulations, supported by clinical evidence in humans, *J. Controlled Release*, 97, 31–42.
15. Sundy, E., and Danckwerts, M. P. (2004), A novel compression-coated doughnut-shaped tablet design for zero-order sustained release, *Eur. J. Pharm. Sci.*, 22, 477–485.
16. Tarvainen, M., Peltonen, S., Mikkonen, H., Elovaara, M., Tuunainen, M., Paronen, P., Ketolainen, J., and Sutinen, R. (2004), Aqueous starch acetate dispersion as a novel coating material for controlled release products, *J. Controlled Release*, 96, 179–191.
17. Thombre, A. G., Appel, L. E., Chidlaw, M. B., Daugherity, P. D., Dumont, F., Evans, L. A. F., and Sutton, S. C. (2004), Osmotic drug delivery using swellable-core technology, *J. Controlled Release*, 94, 75–89.

18. Kim, I. H., Park, J. H., Cheong, I. W., and Kim, J. H. (2003), Swelling and drug release behavior of tablets coated with aqueous hydroxypropyl methylcellulose phthalate (HPMCP) nanoparticles, *J. Controlled Release*, 89, 225–233.
19. Okimoto, K., Rajewski, R. A., and Stella, J. V. (1999), Release of testosterone from an osmotic pump tablet utilizing (SBE)7m-ß-cyclodextrin as both a solubilizing and an osmotic pump agent, *J. Controlled Release*, 58, 29–38.
20. Khan, M. A., Sastry, S. V., Vaithiyalingam, S. R., Agarwal, V., Nazzal, S., and Reddy, I. K. (2000), Captopril gastrointestinal therapeutic system coated with cellulose acetate pseudolatex: Evaluation of main effects of several formulation variables, *Int. J. Pharm.*, 193, 147–156.
21. Lu, E-X., Jiang, Z-Q., Zhang, Q-Z., and Jiang, X-G. (2003), A water-insoluble drug monolithic osmotic tablet system utilizing gum arabic as an osmotic, suspending and expanding agent, *J. Controlled Release*, 92, 375–382.
22. Zhang, Y., Zhang, Z., and Wu, F. (2003), A novel pulsed-release system based on swelling and osmotic pumping mechanism, *J. Controlled Release*, 89, 47–55.
23. Liu, L., and Che, B. (2006), Preparation of monolithic osmotic pump system by coating the indented core tablet, *Eur. J. Pharm. Biopharm.*, 64, 180–184.
24. Zentner, G. M., Rork, G. S., and Himmelstein, K. J. (1985), The controlled-porosity osmotic pump, *J. Controlled Release*, 1, 269–282.
25. Baker, R. W., and Brooke, J. W. (1987), Pharmaceutical drug delivery system, U.S. Patent 4,687,660.
26. Theeuwes, F., and Ayer, A. D. (1978), Osmotic device having composite walls, U.S. Patent 4,077,407.
27. Theeuwes, F. (1978), Microporous-semipermeable laminated osmotic system, U.S. Patent 4,256,108.
28. Cortese, R., and Theeuwes, F. (1982), Osmotic device with hydrogel driving member, U.S. Patent 4,327,725.
29. Swanson, D. R., Burday, B. L., Wong, P. S. L., and Theeuwes, F. (1987), Nifedipine gastrointestinal therapeutic system, *Am. J. Med.*, 83, 3–10.
30. Herbig, S. M., Cardinal, J. R., Korsmeyer, R. W., and Smith, K. L. (1995), Asymmetric membrane tablet coatings for osmotic drug delivery, *J. Controlled Release*, 35, 127–136.
31. Cardinal, J. R., Herbig, S. M., Korsmeyer, R. W., Lo, J., Smith, K. L., and Thombre, A. G. (1997), Use of asymmetric membranes in delivery devices, U.S. Patent 5,612,059.
32. Cardinal, J. R., Herbig, S. M., Korsmeyer, R. W., Lo, J., Smith, K. L., and Thombre, A. G. (1997), Asymmetric membranes in delivery devices, U.S. Patent 5,698,220.
33. Wang, D-M., Lin, F-C., Chen, L-Y., and Lai, J-Y. (1998), Application of asymmetric TPX membranes to transdermal delivery of nitroglycerin, *J. Controlled Release*, 50, 187–195.
34. Thombre, A. G., Cardinal, J. R., DeNoto, A. R., Herbig, S. M., and Smith, K. L. (1999), Asymmetric-membrane capsules for osmotic drug delivery. Part I: Development of a manufacturing process, *J. Controlled Release*, 57, 55–64.
35. Thombre, A. G., Cardinal, J. R., DeNoto, A. R., and Gibbes, D. C. (1999), Asymmetric membrane capsules for osmotic drug delivery. Part II: In vitro and in vivo drug release performance, *J. Controlled Release*, 57, 65–73.
36. Thombre, A. G., DeNoto, A. R., and Gibbes, D. C. (1999), Delivery of glipizide from asymmetric-membrane capsules using encapsulated excipients, *J. Controlled Release*, 60, 333–341.

37. Lin, Y-K., and Ho, H-O. (2003), Investigations on the drug releasing mechanism from an asymmetric-membrane-coated capsule with an in situ formed delivery orifice, *J. Controlled Release*, 89, 57–69.
38. Meier, M. M., Kanis, L. A., and Soldi, V. (2004), Characterization and drug-permeation profiles of microporous and dense cellulose acetate membranes: Influence of plasticizer and pore forming agent, *Int. J. Pharm.*, 278, 99–110.
39. Prabakaran, D., Singh, P., Jaganathan, K. S., and Vyas, S. P. (2004), Osmotically regulated asymmetric capsular systems for simultaneous sustained delivery of anti-tubercular drugs, *J. Controlled Release*, 95, 239–248.
40. Wang, C-Y., Ho, H-O., Lin, L-H., Lin, Y-K., and Ming-Thau Sheu, M-T. (2005), Asymmetric membrane capsules for delivery of poorly water-soluble drugs by osmotic effects, *Int. J. Pharm.*, 297, 89–97.
41. Altinkaya, S. A., and Yenal, H. (2006), In vitro drug release rates from asymmetric-membrane tablet coatings: Prediction of phase-inversion dynamics, *Biochem. Eng. J.*, 28, 131–139.
42. Wang, G-M., Chen, C-H., Ho, H-O., Wang, S-S., and Sheu, M-T. (2006), Novel design of osmotic chitosan capsules characterized by asymmetric membrane structure for in situ formation of delivery orifice, *Int. J. Pharm.*, 319, 71–81.
43. Altinkaya, S. A., and Ozbas, B. (2004), Modeling of asymmetric-membrane formation by dry-casting method, *J. Membr. Sci.*, 230, 71–89.
44. Mathiowitz, M. (1999), *Encyclopedia of Controlled Drug Delivery*, Wiley, New York, pp. 302–303.
45. Sandadi, S., PAndey, P., and Turton, R. (2004), In situ, near real-time acquisition of particle motion in rotating pan coating equipment using imaging techniques, *Chem. Eng. Sci.*, 59, 5807–5817.
46. Prater, D., Wilde, J., and Meakin, B. (1980), A model system for the production of aqueous tablet film coating for laboratory evaluation, *J. Pharm. Pharmacol.*, 32(Suppl.), 90.
47. Leaver, T., Shannon, H., and Rowe, R. (1985), A photometric analysis of tablet movement in a side-vented perforated drum (Accela-Cota), *J. Pharm. Pharmacol.*, 37, 17–21.
48. Nakagawa, M., Altobelli, S., Caprihan, A., Fukushima, E., and Jeong, E. (1993), Non-invasive measurements of granular flows by magnetic resonance imaging, *Exper. Fluids*, 16, 54–60.
49. Parker, D., Broadbent, C., Fowles, P., Hawkesworth, M., and McNeil, P. (1993), Positron emission particle tracking—A technique for studying flow within engineering equipment, *Nucl. Instrum. Methods Phys. Res. A*, 326, 592–607.
50. Parker, D., Dijkstra, A., Martin, T., and Seville, J. (1997), Positron emission particle tracking studies of spherical particle motion in rotating drums, *Chem. Eng. Sci.*, 52, 2011–2022.
51. Wilson, K., and Crossman, E. (1997), The influence of tablet shape and pan speed on intra-tablet film coating uniformity, *Drug Dev. Ind. Pharm.*, 23, 1239–1243.
52. Yamane, K., Sato, T., Tanaka, T., and Tsuji, Y. (1995), Computer simulation of tablet motion in coating drum, *Pharm. Res.*, 12, 1264–1268.
53. Yamane, K., Nakagawa, M., Altobelli, S., Tanaka, T., and Tsuji, Y. (1998), Steady particulate flows in a horizontal rotating cylinder, *Phys. Fluids*, 10, 1419–1427.
54. Denis, C., Hemati, M., Chulia, D., Lanne, J., Buisson, B., Daste, G., and Elbaz, F. (2003), A model of surface renewal with application to the coating of pharmaceutical tablets in rotary drums, *Powder Technol.*, 130, 174–180.

55. Pandey, P., Song, Y., Kayihan, F., and Turton, R. (2006), Simulation of particle movement in a pan coating device using discrete element modeling and its comparison with video-imaging experiments, *Powder Technol.*, 161, 79–88.
56. Rege, B. D., Gawel, J., and Kou, J. H. (2002), Identification of critical process variables for coating actives onto tablets via statistically designed experiments, *Int. J. Pharm.*, 237, 87–94.
57. Turton, R., and Cheng, X. X. (2005), The scale-up of spray coating processes for granular solids and tablets, *Powder Technol.*, 150, 78–85.
58. Xu, M., and Turton, R. (1997), A new data processing technique for noisy signals: Application to measuring particle circulation times in a draft tube equipped fluidized bed, *Powder Technol.*, 92, 111–117.
59. Sherony, D. F. (1981), A model of surface renewal with application to fluid bed coating of particles, *Chem. Eng. Sci.*, 36, 845–848.
60. Wnukowski, P., and Setterwall, F. (1989), The coating of particles in a fluidized bed (residence time distribution in a system of two coupled perfect mixers), *Chem. Eng. Sci.*, 44, 493–505.
61. Choi, M. S., and Meisen, A. (1997), Sulfur coating of urea in shallow spouted beds, *Chem. Eng. Sci.*, 52, 1073–1086.
62. Maronga, J., and Wnukowski, P. (1997), Modeling of the three-domain fluidized bed particulate coating process, *Chem. Eng. Sci.*, 52, 2915–2925.
63. Nakamura, H., Abe, E., and Yamada, N. (1998), Coating mass distributions of seed particles in a tumbling fluidized bed coater. Part II. A Monte Carlo simulation of particle coating, *Powder Technol.*, 99, 140–146.
64. Cheng, X. X., and Turton, R. (2000), The prediction of variability occurring in fluidized bed coating equipment. II: The role of nonuniform particle coverage as particle pass through the spray zone, *Pharm. Dev. Technol.*, 5, 323–332.
65. KuShaari, K., Pandey, P., Song, Y., and Turton, R. (2006), Monte Carlo simulations to determine coating uniformity in a Wurster fluidized bed coating process, *Powder Technol.*, 166, 81–90.
66. Waterman, K. C., and Fergione, M. B. (2003), Press-coating of immediate release powders onto coated controlled release tablets with adhesives, *J. Controlled Release*, 89, 387–395.
67. Leong, K. C., Lu, G. Q., and Rudolph, V. (1999), A comparative study of the fluidized-bed coating of cylindrical metal surfaces with various thermoplastic polymer powders, *J. Mater. Proc. Technol.*, 89/90, 354–360.
68. Wulf, M., Uhlmann, P., Michel, S., and Grundke, K. (2000), Surface tension studies of levelling additives in powder coatings, *Prog. Org. Coat.*, 38, 59–66.
69. Belder, E. G., Rutten, H. J. J., and Perera, D. Y. (2001), Cure characterization of powder coatings, *Prog. Org. Coat.*, 42, 142–149.
70. Pfeffer, R., Dave, R. N., Wie, D., and Ramlakhan, M. (2001), Synthesis of engineered particulates with tailored properties using dry particle coating, *Powder Technol.*, 117, 40–67.
71. McLaren, D. D., and Hollenbeck, R. G. (1987), A high performance liquid chromatographic method for the determination of the amount of hydroxypropyl methylcellulose applied to tablets during an aqueous film coating operation, *Drug Dev. Ind. Pharm.*, 13, 2179–2197.
72. Mouget, Y., Gosselin, P., Tourigny, M, and B'echard, S. (2003), Three-dimensional analyses of tablet content and film coating uniformity by laser-induced breakdown spectroscopy (LIBS), *Am. Lab.*, 35, 20–22.

73. St-Onge, L., Kwong, E., Sabsabi, M., and Vadas, E. B. (2002), Quantitative analysis of pharmaceutical products by laser-induced breakdown spectroscopy, *Spectrochim. Acta, Part B*, 57, 1131–1140.
74. Kirsch, J. D., and Drennen, J. K. (1995), Determination of film-coated tablet parameters by near-infrared spectroscopy. *J. Pharm. Biomed. Anal.*, 13, 1273–1281.
75. Kirsch, J. D., and Drennen, J. K. (1996), Near-infrared spectroscopic monitoring of the film coating process, *Pharm. Res.*, 13, 234–237.
76. Anderson, M., Josefson, M., Langkilde, F., and Wahlund, K.-G. (1999), Monitoring of a film coating process for tablets using near infrared reflectance spectrometry, *J. Pharm. Biomed. Anal.*, 20, 27–37.
77. Romero-Torres, S., Pérez-Ramos, J. D., Morris, K. R., and Grant, E. R. (2006), Raman spectroscopy for tablet coating thickness quantification and coating characterization in the presence of strong fluorescent interference, *J. Pharma. Biomed. Anal.*, 41, 811–819.
78. Romero-Torres, S., Pérez-Ramos, J. D., Morris, K. R., and Grant, E. R. (2005), Raman spectroscopic measurement of tablet-to-tablet coating variability, *J. Pharm. Biomed. Anal.*, 38, 270–274.
79. Ruotsalainen, M., Heinämäki, J., Guo, H., Laitinen, N., and Yliruusi, J. (2003), A novel technique for imaging film coating defects in the film-core interface and surface of coated tablets, *Eur. J. Pharm. Biopharm.*, 56, 381–388.
80. Podczeck, F. (1998), Measurement of surface roughness of tablets made from polyethylene glycol powders of various molecular weight, *Pharm. Pharmacol. Commun.*, 4, 179–182.
81. Riippi, M., Antikainen, O., Niskanen, T., and Yliruusi, J. (1998), The effect of compression force on surface structure, crushing strength, friability, and disintegration time of erythromycin acistrate tablets, *Eur. J. Pharm. Biopharm.*, 46, 339–345.
82. Newton, M., Petersson, J., Podczeck, F., Clarke, A., and Booth, S. (2001), The influence of formulation variables on the properties of pellets containing a self-emulsifying mixture, *J. Pharm. Sci.*, 90, 987–995.
83. Seitavuopio, P., Rantanen, J., and Yliruusi, J. (2003), Tablet surface characterisation by various imaging techniques, *Int. J. Pharm.*, 254, 281–286.
84. Seitavuopio, P., Rantanen, J., and Yliruusi, J. (2005), Use of roughness maps in visualisation of surfaces, *Eur. J. Pharm. Biopharm.*, 59, 351–358.
85. van de Witte, P., Dijkstra, P. J., Van den Berg, J. W. A., and Feijen, J. (1996), Phase separation processes in polymer solutions in relation to membrane formation, *J. Membr. Sci.*, 117, 1–31.
86. Han, M., and Bhattacharyya, D. (1995), Changes in morphology and transport characteristics of polysulfone membranes prepared by different demixing conditions, *J. Membr. Sci.*, 98, 191–200.
87. Park, H. C., Kim, Y. P., Kim, H. Y., and Kang, Y. S. (1999), Membrane formation by water vapor induced phase inversion, *J. Membr. Sci.*, 156, 169–178.
88. Tobiska, S., and Kleinebudde, P. (2001), A simple method for evaluating the mixing efficiency of a new type of pan coater, *Int. J. Pharm.*, 224, 141–149.
89. Plumb, A. P., Rowe, R. C., York, P., and Doherty, C. (2002), The effect of experimental design on the modeling of a tablet coating formulation using artificial neural networks, *Eur. J. Pharm. Sci.*, 16, 281–288.
90. Rege, B. D., Gawel, J. and Kou, H. J. (2002), Identification of critical process variables for coating actives onto tablets via statistically designed experiments, *Int. J. Pharm.*, 237, 87–94.

91. Mueller, R., and Kleinebudde, P. (in press), Influence of scale-up on the abrasion of tablets in a pan coater, *Eur. J. Pharm. Biopharm.*
92. Wang, G-M., Chen, C-H., Ho, H-O., Wang, S-S., and Sheu, M-T. (press), Novel design of osmotic chitosan capsules characterized by asymmetric membrane structure for in situ formation of delivery orifice, *Int. J. Pharm.*
93. Hussain, A. S., Shivanand, P., and Johnson, R. D. (1994), Application of neural computing in pharmaceutical product development: Computer aided formulation design, *Drug Dev. Ind. Pharm.*, 20, 1739–1752.
94. Kesavan, J. G., and Peck, G. E. (1995), Pharmaceutical formulation using neural networks. *Proc.14th Pharm. Technol. Conf. (Barcelona)*, 2, 413–431.
95. Colbourn, E. A., and Rowe, R. C. (1996), Modeling and optimization of a tablet formulation using neural networks and genetic algorithms, *Pharm. Technol. Eur.*, 8, 46–55.
96. Bourquin, J., Schmidli, H., van Hoogevest, P., and Leuenberger, H. (1997), Basic concepts of artificial neural networks (ANN) modeling in the application to pharmaceutical development, *Pharm. Dev. Technol.*, 2, 95–109.
97. Bourquin, J., Schmidli, H., van Hoogevest, P., and Leuenberger, H. (1997b), Application of artificial neural networks (ANN) in the development of solid dosage forms, *Pharm. Dev. Technol.*, 2, 111–121.
98. Ebube, N. K., McCall, T., Chen, Y., and Meyer, M. C. (1997), Relating formulation variables to in vitro dissolution using an artificial neural network, *Pharm. Dev. Technol.*, 2, 225–232.
99. Bourquin, J., Schmidli, H., van Hoogevest, P., and Leuenberger, H. (1998), Comparison of artificial neural networks (ANN) with classical modeling techniques using different experimental designs and data from a galenical study on a solid dosage form, *Eur. J. Pharm. Sci.*, 6, 287–300.
100. Bourquin, J., Schmidli, H., van Hoogevest, P., and Leuenberger, H. (1998), Advantages of artificial neural networks (ANNs) as alternative modeling technique for data sets showing non-linear relationships using data from a galenical study on a solid dosage form, *Eur. J. Pharm. Sci.*, 7, 5–16.
101. Bourquin, J., Schmidli, H., van Hoogevest, P., and Leuenberger, H. (1998), Pitfalls of artificial neural networks (ANN) modeling technique for data sets containing outlier measurements using a study on mixture properties of a direct compressed dosage form, *Eur. J. Pharm. Sci.*, 7, 17–28.
102. Chen, Y., McCall, T. W., Baichwal, A. R., and Meyer, M. C. (1999), The application of an artificial neural network and pharmacokinetic simulations in the design of controlled-release dosage forms, *J. Controlled Release*, 59, 33–41.
103. Plumb, A. P., Rowe, R. C., York, P., and Doherty, C. (2002), The effect of experimental design on the modeling of a tablet coating formulation using artificial neural networks, *Eur. J. Pharm. Sci.*, 16, 281–288.
104. Plumb, A. P., Rowe, R. C., York, P., and Doherty, C. (2003), Effect of varying optimization parameters on optimization by guided evolutionary simulated annealing (GESA) using a tablet film coat as an example formulation, *Eur. J. Pharma. Sci.*, 18, 259–266.
105. Takayama, K., Fujikawa, M., Obata, Y., and Morishita, M. (2003), Neural network based optimization of drug formulations, *Adv. Drug Deliv. Rev.*, 55, 1217–1231.
106. Montgomery, D. C. (1997), *Design and Analysis of Experiments*, 4th ed., Wiley, New York, pp. 611–616.

107. Lai, J.-Y., Lin, F.-C., Wang, C.-C., and Wang, D.-M. (1996), Effect of nonsolvent additives on the porosity and morphology of asymmetric TPX membranes, *J. Membr. Sci.*, 118, 49–61.

108. Broadhead, K. W., and Tresco, P. A. (1998), Effects of fabrication conditions on the structure and function of membranes formed from poly(acrylonitrile-vinylchloride), *J. Membr. Sci.*, 147, 235–245.

109. Won, J., Park, H. C., Kim, U. Y., Kang, S., Yoo, S. H., and Jho, J. Y. (1999), The effect of dope solution characteristics on the membrane morphology and gas transport properties: PES/-BL/NMP system, *J. Membr. Sci.*, 162, 247–255.

110. Young, T-H., Lin, D.-J., Gau, J.-J., Chuang, W.-Y., and Cheng, L.-P. (1999), Morphology of crystalline Nylon-610 membranes prepared by the immersion–precipitation process: Competition between crystallization and liquid–liquid phase separation, *Polymer*, 40, 5011–5021.

111. Chuang, W.-Y., Young, T.-H., Chiu, W.-Y., and Lin, C.-Y. (2000), The effect of polymeric additives on the structure and permeability of poly(vinyl alcohol) asymmetric-membranes, *Polymer*, 41, 5633–5641.

112. Won, J., Lee, H. J., and Kang, Y. S. (2000), The effect of dope solution characteristics on the membrane morphology and gas transport properties. Part 2: PES/BL system, *J. Membr. Sci.*, 176, 11–19.

113. Fan, S.-C., Wang, Y.-C., Li, C.-L., Lee, K.-R., Liaw, D.-J., Huang, H.-P., and Lai, J.-Y. (2002), Effect of coagulation media on membrane formation and vapor permeation performance of novel aromatic polyamide membrane, *J. Membr. Sci.*, 204, 67–79.

114. Mohamed, N. A., and Al-Dossary, A. O. H. (2003), Structure–property relationships for novel wholly aromatic polyamide-hydrazides containing various proportions of *para*-phenylene and *meta*-phenylene units. Part III: Preparation and properties of semi-permeable membranes for water desalination by reverse osmosis separation performance, *Eur. Polym. J.*, 39, 1653–1667.

115. Pinnau, I., and Koros, W. (1991), Structures and gas separation property asymmetric polysulfone membranes made by dry, wet, and dry/wet phase-inversion, *J. Appl. Polym. Sci.*, 43, 1491–1502.

116. Sharpe, I. D., Ismail, A. F., and Shilton, S. J. (1999), A study of extrusion shear and forced convection residence time in the spinning of polysulfone hollow fiber membranes for gas separation, *Sep. Purif. Technol.*, 17, 101–109.

117. Ismail, A. F., Ng, B. C., and Abdul Rahman, W. A. W. (2003), Effects of shear rate and forced convection residence time on asymmetric polysulfone membranes structure and gas separation performance, *Sep. Purif. Technol.*, 33, 255–272.

118. Matsuyama, H., Teramoto, M., and Uesaka, T. (1997), Membrane formation and structure development by dry cast process, *J. Membr. Sci.*, 135, 271–288.

119. Jansen, J. C., Macchione, M., and Drioli, E. (2005), High flux asymmetric gas separation membranes of modified poly(ether ether ketone) prepared by the dry phase inversion technique, *J. Membr. Sci.*, 255, 167–180.

120. Jansen, J. C., Buonomenna, M. G., Figoli, A., and Drioli, E. (2006), Asymmetric membranes of modified poly(ether ether ketone) with an ultra-thin skin for gas and vapour separations, *J. Membr. Sci.*, 272, 188–197.

121. Kleinbaum, D. G., Kupper, L. L., and Muller, K. E., (1987), *Applied Regression Analysis and Other Multivariable Methods*, 2nd ed., Duxbury, Belmont, pp. 658–659.

## APPENDIX

The in vitro release profiles of the model drug theophylline are shown below in Figures A1 through A10.

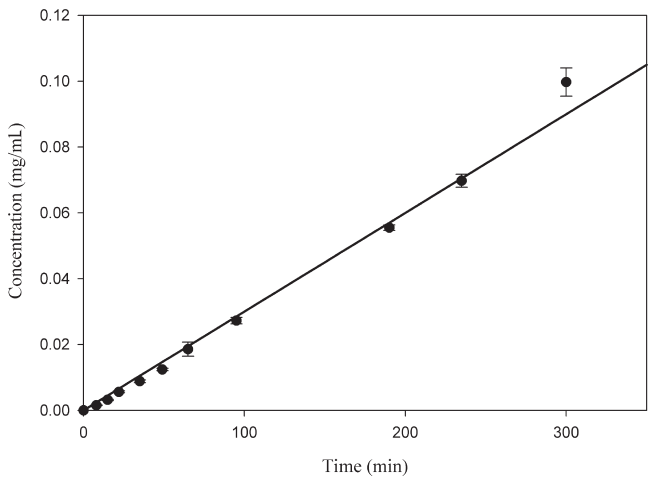


**FIGURE A1** Release of theohylline from tablet coatings made with 5% CA dissolved in 90% acetone and 5% water.

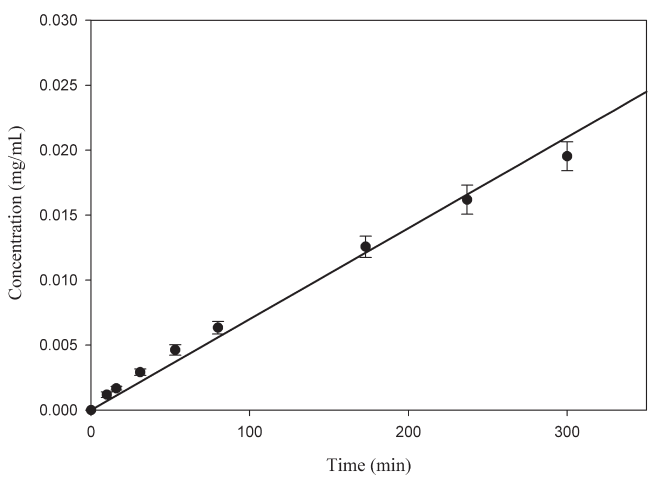


**FIGURE A2** Release of theohylline from tablet coatings made with 10% CA dissolved in 85% acetone and 5% water.

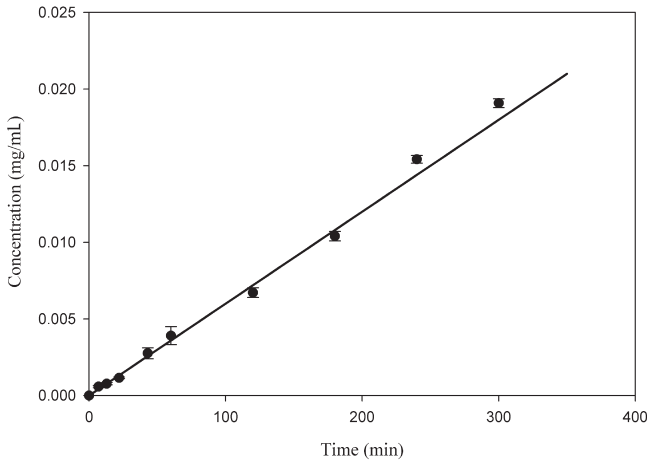


**FIGURE A3** Release of theohylline from tablet coatings made with 15% CA dissolved in 80% acetone and 5% water.

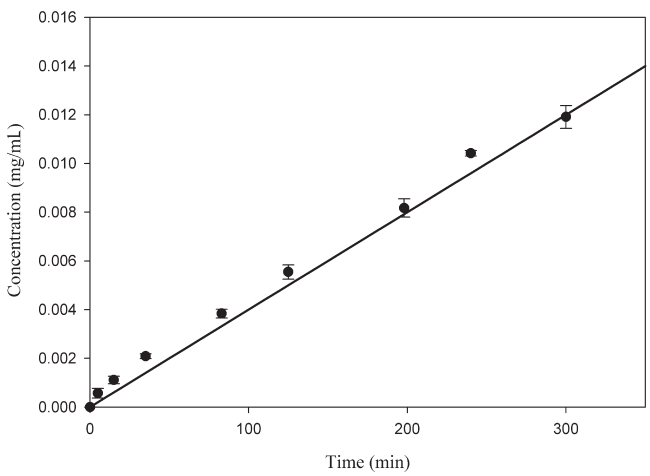


**FIGURE A4** Release of theohylline from tablet coatings made with 5% CA dissolved in 85% acetone and 10% water.

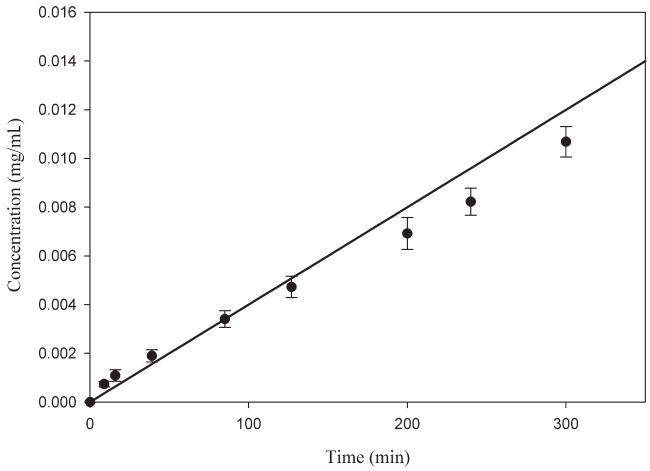


**FIGURE A5** Release of theohylline from tablet coatings made with 10% CA dissolved in 80% acetone and 10% water.

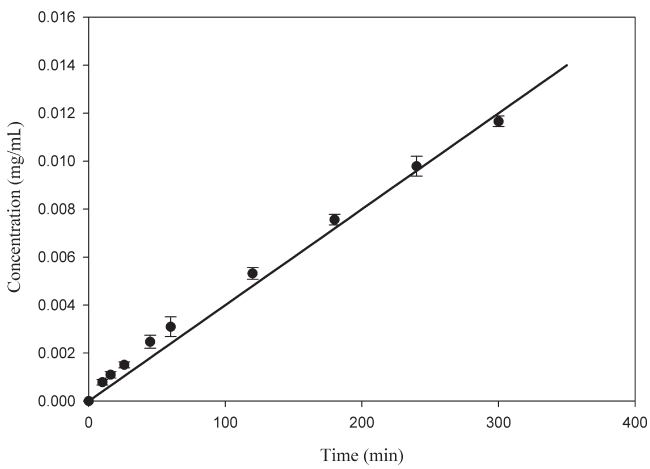


**FIGURE A6** Release of theohylline from tablet coatings made with 15% CA dissolved in 75% acetone and 10% water.

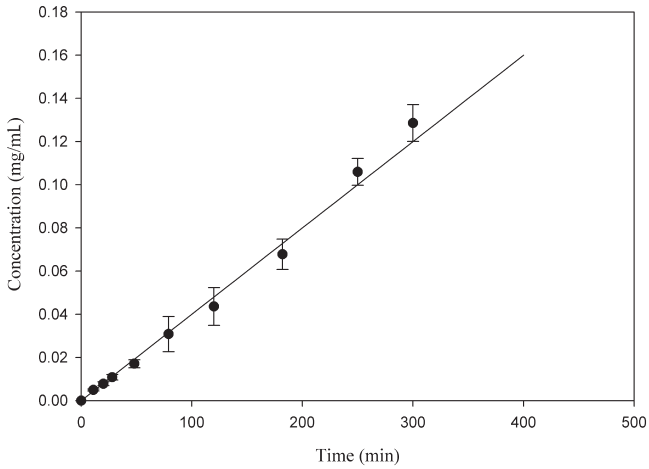


**FIGURE A7** Release of theohylline from tablet coatings made with 5% CA dissolved in 80% acetone and 15% water.

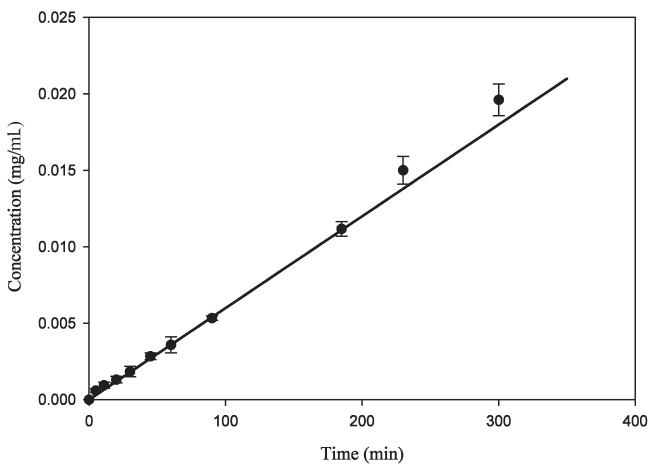


**FIGURE A8** Release of theohylline from tablet coatings made with 10% CA dissolved in 75% acetone and 15% water.

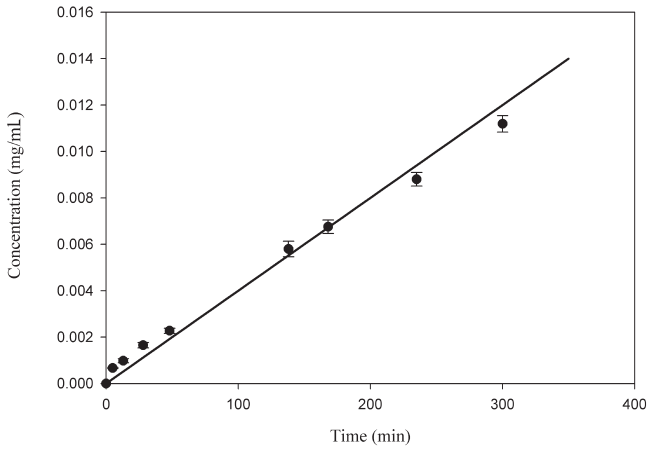


**FIGURE A9** Release of theohylline from tablet coatings made with 15% CA dissolved in 70% acetone and 15% water.

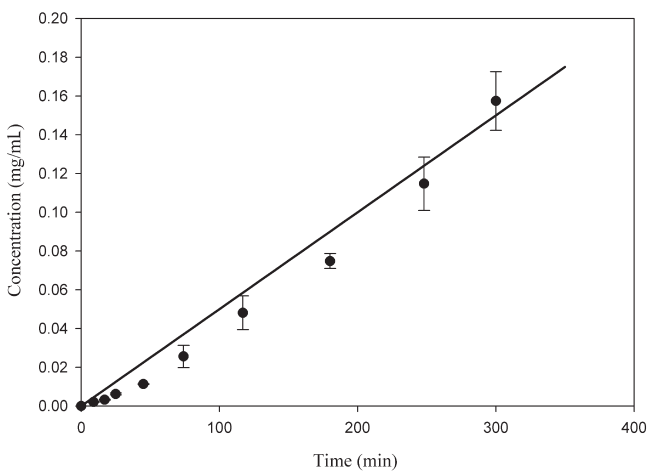


**FIGURE A10** Release of theohylline from tablet coatings made with 12.5% CA dissolved in 77.5% acetone and 10% water.

# TABLET COMPRESSION

Helton M. M. Santos and João J. M. S. Sousa
*University of Coimbra, Coimbra, Portugal*

**Contents**

1 Introduction
2 Theory of Particle Compaction
3 Compactibility
4 Tablet Compression
5 Equipment for Tablet Compression
6 Tablet Press Tooling
7 Tablet Engraving
8 Tablet Shape and Profile
9 Tablet Bisect
10 Problems during Tablet Manufacturing
10.1 Capping and Lamination
10.2 Picking and Sticking
10.3 Mottling
10.4 Weight and Hardness Variation
References

## 1 INTRODUCTION

Tablets are the most important pharmaceutical dosage from and their design has always been of great interest to pharmaceutical engineering. Since tablets are made by a process of die compaction, although commonly called tablet compression, many investigations have been involved in the task to describe the mechanisms involved in this process. Nevertheless, some considerations should be taken regarding the definitions of stages involved in tablet compression. Compression is one of two

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

stages involved in the compaction of a two-phase system due to the application of an external force. It is defined as the reduction in the bulk volume of a material as a result of gaseous phase [1, 2]. The second stage is consolidation, which is described as an increase in the mechanical strength of the material resulting in particle-to-particle interaction [1, 2].

It is suggested that four mechanisms are basically involved in the process of compression of particles: deformation, densification, fragmentation, and attrition. The process of compression is briefly described as follows: small solid particles are filled in a die cavity and a compression force is applied to it by means of punches and then the formed monolithic dosage form is ejected. The shape of the tablet is dictated by the shape of the die while the distance between the punch tips at the point of maximum compression governs the tablet thickness, and the punch tip configuration determines the profile of the tablet. The compression cycle in a conventional rotary tablet press will be described in detail in this chapter and is illustrated in Figure 1.

The physical and mechanical properties of tablets, such as density and mechanical strength, are significantly affected by the process. Since tablet compression relies on the ability of particulates to be compacted, the need to control the critical properties of the materials with respect to readiness or ability to compact is an important issue to the formulator.

In order to compress a powder or granulation product into a tablet of specific hardness, a defined compression force must be applied. As pointed out by Shlieout et al. [3], by compressing a constant mass of powder, any variation in the applied force causes a change in the measured force. In addition, the substance itself plays an important role, that is, if it is of good compressibility, then the force needed for compression would be low. It is well known that this compressibility will depend on powder characteristics such as crystal habit and thermodynamic behavior.

The structure and strength of tablets are often discussed in terms of the relationship between the properties of the particulate material and the properties of the formed tablet. The properties of a powder that control its evolution into a tablet

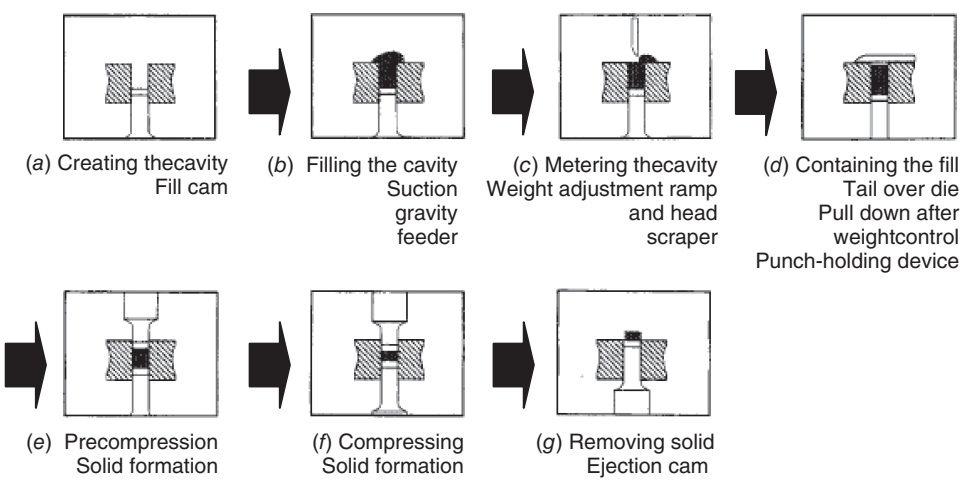


**FIGURE 1** Compression cycle. (Courtesy of Thomas Engineering.)

during compression, which will also relate to the fracture toughness and the tensile strength of the tablets, are the compression mechanics of the particles and their dimensions. Generally, all materials have the ability to store some elastic strain; however, its extent will greatly vary for different materials and will depend upon the intrinsic nature of the material. There are many instances where a brittle material, or its surface, reduces significantly its cohesion or adhesion compared to that of a ductile material [4].

## 2 THEORY OF PARTICLE COMPACTION

Basically, the process of tablet compression starts with the rearrangement of particles within the die cavity and initial elimination of voids. As tablet formulation is a multicomponent system, its ability to form a good compact is dictated by the compressibility and compactibility characteristics of each component. Compressibility of a powder is defined as its ability to decrease in volume under pressure, and compactibility is the ability of the powdered material to be compressed into a tablet of specific tensile strength [1, 2]. One emerging approach to understand the mechanism of powder consolidation and compression is known as percolation theory. In a simple way, the process of compaction can be considered a combination of site and bond percolation phenomena [5]. Percolation theory is based on the formation of clusters and the existence of a site or bond percolation phenomenon. It is possible to apply percolation theory if a system can be sufficiently well described by a lattice in which the spaces are occupied at random or all sites are already occupied and bonds between neighboring sites are formed at random.

The transitional repacking stage is driven by the particle size distribution and shape. This will determine the bulk density as the powder or granulation product is delivered into the die cavity. In this phase, the punch and particle movements occur at low pressure. The particles flow with respect to each other, with the finer particles entering the void between the larger particles, and thus the bulk density of the granulation is increased. Various techniques have been utilized to determine the degree of the two consolidation mechanisms in pharmaceutical solids (initial packing of the particles and elimination of void spaces), namely the rate dependency technique. By applying this technique, stress relaxation data based on the Maxwell model of viscoelastic behavior indicate virtually no rate dependency for elastic or brittle materials. There is also an increase in the calculated yield pressure with an increase in punch velocity for viscoplastic materials such as maize starch and polymeric materials. This is attributed to the reduction of time necessary for the plastic deformation process to occur [6]. For brittle materials such as magnesium and calcium carbonates there is no observed change in the yield pressure with increasing punch velocity [6].

When a force is applied to a material, deformation occurs. When this deformation completely disappears after cessation of the external force, further deformation occurs. Deformations that do not completely recover after release of the stress are known as plastic deformations. The force required to initiate a plastic deformation is known as the yield stress. When the particles are so closely packed that no further filling of the voids can occur, a further increase of the compressional force causes deformation at the points of contact. Both plastic and elastic deformation may occur,

although one type will predominate for a given material. The ability of materials to be compressed relies on their deformation behavior. The known extreme cases are as follows. For elastic bodies, the force applied to consolidation will be fully given back (action equals reaction). This is expressed as a completely elastic deformation. For plastic bodies, the force applied will be saved as energy in the body and will express no elastic deformation at all. During tablet building, these two processes never occur alone but occur only in combination, as mentioned before.

As the external force is increased, the stresses within the particles become great enough and cracks may occur. Fragmentation furthers densification with the infiltration of the smaller fragments into the void spaces being responsible for increasing the number of particles and formation of new and clean surfaces that are potential bonding areas. The mechanisms of fragmentation and plastic deformation are not independent since both phenomena modify particle size distribution, and larger particles do not act as small particles with respect to plastic deformation [7].

The bonding of particles is governed by different mechanisms. The three most considered theories are mechanical theory, intermolecular theory, and liquid surface film theory. The first theory assumes that under pressure the individual particles experience elastic, plastic, or brittle deformation and that the edges of the particles intermesh, forming a mechanical bonding. According to Parrot [2], intermolecular theory states that under pressure the molecules at the points of true contact are close enough so that van der Waals forces interact to consolidate the particles. Liquid surface film theory relies on the presence of a thin liquid film, which may be the consequence of fusion or solution, at the surface of the particle, induced by the energy of compression. Even tough the applied force is not high, it is locally transmitted to small areas of true contact so that a very high pressure will exist at the contact surfaces. This high pressure plays an important role in the melting point and solubility of the material and proves to be essential to bonding. It follows that after releasing the pressure, solidification of the fused material would form solid bridges between the particles. An important consideration has been proposed by Zuurman et al. [8] to explain the action of some excipients during this phase. One of these excipients is magnesium stearate, which is widely used as a lubricant in order to prevent tablets from sticking to the die and punches and minimize wear of tooling. It has been proven that magnesium stearate can have an adverse effect on bonding between particles. The decrease of tablet strength is always considered to be the result of reduction of interparticle bonding due to the addition of a lubricant.

The production of tablets with the desired characteristics depends on the stresses induced by elastic rebound and the associated deformation processes during decompression and ejection. Ideally, if only elastic deformation occurred, with the sudden removal of axial pressure the particles would return to their original form, breaking any bonds that may have been under pressure.

Finally, as the lower punch rises and pushes the tablet upward, there is a continued residual die wall pressure and considerable energy may be expanded due to the die wall friction. As the tablet is removed form the die, the lateral pressure is relieved, and the tablet undergoes elastic recovery with an increase in the volume of the portion removed from the die.

The compression cycle profiles may be used to characterize the consolidation mechanisms of powders as they help to characterize the extent of pressure distribution within the powder bed as well as the formed tablet. The compression behavior

of powder mixtures is usually characterized using the well-known Heckel equation [9, 10], which describes the relationship between the porosity of a compact and the applied pressure and is based on the assumption that the densification of the bulk powder in the die follows first-order kinetics:

$$\ln\frac{1}{1-\rho_r} = kP + A \tag{1}$$

where $\rho_r$ is the relative density of the compact at pressure $\sigma$, $P$ is the applied pressure, and $K$ and $A$ are constants. The constants $A$ and $k$ are determined from the intercept and slope, respectively, of the extrapolated linear region of a plot of ln(1/1 – $\rho_r$) versus $\sigma$ (compaction pressure). The Heckel constant $k$ is related to the reciprocal of the mean yield pressure, which is the minimum pressure required to cause deformation of the material undergoing compression. The intercept obtained from the slope of the upper portion of the curve is a reflection of the densification after consolidation. A large value of $k$ indicates the onset of plastic deformation at relatively low pressure. Thus, $K$ appears to be a material constant. The correlation between $k$ and the mean yield pressure $P_y$ gives Equation (2). The constant $A$ is related to the densification during die filling and particle rearrangement prior to bonding [11]:

$$k = \frac{1}{P_y} \tag{2}$$

A high $\rho_r$ value indicates that there will be a high volume reduction of the product due to particle rearrangement. The constant $A$ has been shown to be equal to the reciprocal of the mean yield pressure required to induce plastic deformation. A larger value for $A$ (low yield pressure) indicates the onset of plastic deformation at relatively low pressure, a sign that the material is more compressible.

The Heckel plot allows an interpretation of the mechanism of bonding. A nonlinear plot with small value for its slope (a small Heckel constant) indicates that the material undergoes fragmentation during compression. When the plot is linear, it indicates that the material undergoes plastic deformation during compression.

In addition to the Heckel approach, other techniques may be applied to the characterization of powder compression. One of these approaches was proposed by Cooper and Eaton [12]:

$$\frac{V_0 - V}{V_0 - V_s} = a_1 \exp\left(\frac{-k_1}{P}\right) + a_2 \exp\left(\frac{-k_2}{P}\right) \tag{3}$$

where $V$ is the volume of the compact at pressure $P$ ($m^3$), $V_0$ is the volume of compact at zero pressure ($m^3$), $V_s$ is the void-free solid material volume ($m^3$), $a_1$, $a_2$, $k_1$, and $k_2$ are the Cooper–Eaton constants.

The Kawakita equation [13] describes the relationship between volume reduction and applied pressure according to Equation (4), where $P$ is the applied pressure, $V_0$ is the initial bulk volume, $V$ is the volume at pressure $P$, $a$ and $b$ are the constants characteristic of the powder under compression, and $C$ is the degree of volume reduction [Equation (5)]:

$$\frac{P}{C} = \frac{P}{a} + \frac{1}{ab} \tag{4}$$

$$C = \frac{V_0 - V}{V_0} \tag{5}$$

In the Kawakita equation the particle density is not introduced in the calculations since the model operates on the relative change in volume, which gives the same result whether the relative or the absolute volume is used. The problem in the calculation of this equation is to find the correct initial volume $V_0$.

## 3 COMPACTIBILITY

Compactibility of a powdered mixture is defined as the ability of the material to be compressed into a tablet of a specified strength without changing its composition. Investigations have demonstrated that binary mixes of identical composition could have different organizations, depending on the surface energy and particle size of the fraction used. Actually, it has been demonstrated that it is possible to control the organization of binary mixes by modifying the particle sizes of the fractions blended if they have the appropriate surface energies [6].

Generally, only powders that form hard compacts under an applied pressure without exhibiting any tendency to cap or chip can be considered as readily compactible. The compactibility of pharmaceutical powders can be characterized by its tensile strength and indentation hardness, which can be used to determine three dimensionless parameters: strain index, bonding index, and brittle fracture index.

To calculate the work of compaction during tableting, it is necessary to have accurate values of force and punch displacement. Differences in the dynamics of powder densification between eccentric and rotary machines were pointed out by Palmieri et al. [14] after compression of microcrystalline cellulose, lactose monohydrate, and dicalcium phosphate dehydrate at different compression pressures. The effect of the longer dwell time of the rotary machine press on the porosity reduction after the maximum pressure is reached is more noticeable in a ductile material such as microcrystalline cellulose. It has been shown that Heckel parameters obtained in the rotary press are in some cases different from those recovered in the eccentric machine because of the longer dwell time, machine deflection, and punch tilting occurring in the rotary press, although theoretically they could better describe the material densification in a high-speed production rotary machine.

Williams and McGinity [15] studied and compared the compaction properties of microcrystalline cellulose from six different sources using tableting indices. It was demonstrated that storage of compacts at elevated humidity conditions prior to determining the tableting indices decreased the magnitude of the tensile strength, dynamic indentation hardness, and bonding index. Based on the differences in physicomechanical properties observed for the tableting indices, the authors stated that microcrystalline cellulose products from different sources are not directly interchangeable and showed that the tensile strength, indentation hardness, bonding index, and brittle fracture index for compacts composed of microcrystalline cellu-

lose in combination with either talc or magnesium stearate generally decreased as the amount of the lubricant was increased over the concentration range of 0–9%. Similar results were observed for admixtures of sodium sulfathiazole in combination with either talc or magnesium stearate. It was also demonstrated that the tensile strength, indentation hardness, and bonding index increased, and the brittle fracture index decreased as the percent of microcrystalline cellulose was increased in a mixture with sodium sulfathiazole.

The results of a study conducted by Muller and Augsburger [16] suggest that the pressure–volume relationship determined during powder bed compression is affected by the instantaneous punch speed profile of the displacement–time waveform for all materials studied, even though they deform by different mechanisms. It appears that the instantaneous punch speed profile of the particular displacement–time waveform is a confounding factor of Heckel analysis.

Moisute acts as a plasticizer and influences the mechanical properties of powdered materials for tablet compression. In the case of microcrystalline cellulose, at moisture levels above 5% the material exhibits significant changes consistent with a transition from the glassy state to the rubbery state [17]. The possible influence of moisture on the compaction behavior of powders was also analyzed by Gupta et al. [18]. This work evaluates the effect of variation in the ambient moisture on the compaction behavior of microcrystalline cellulose powder.

The work conducted by Gustafsson et al. [19] evaluated the particle properties and solid-state characteristics of two different brands of microcrystalline cellulose (Avicel PH101 and a brand obtained from the alga *Cladophora* sp.) and related the compaction behavior to the properties of the tablets. The difference in fibril dimension and, thereby, the fibril surface area of the two celluloses were shown to be the primary factor in determining their properties and behavior.

The compaction properties of pharmaceutical formulations can be studied experimentally using a variety of techniques, ranging from instrumented production presses to compaction simulators, and methods of analysis. The results are usually plotted as porosity–axial stress functions, which is of interest to compare different materials. However, there are some drawbacks on this type of evaluation. As mentioned by Cunningham et al. [20], this approach is deficient once it considers only the average stress along the direction of compaction, ignoring radial stress transmission and friction.

There have been some attempts to overcome the analysis of compaction problems, mostly by introducing numerical modeling approaches. The modeling approaches often used in compaction analysis are (a) phenomenological continuum models, (b) micromechanically based continuum models, and (c) discrete-element models. The parameters that should be analyzed when tableting is under development are as follows:

1. Understanding the formulation and compositional effects on the compaction process, including axial loading and unloading along with ejection
2. Determination of the stress distributions within the powder compact, including residual stresses
3. Optimization of the tablet tooling design

4. Estimation of the density distributions within a tablet that may influence dissolution or mechanical properties
5. Estimation of the compaction force necessary to obtain tablets having given properties
6. Taking into account the effect of the tablet material on the stress distribution on tooling to aid tool design
7. Assessment of the origin of defect or crack formation
8. Optimization of more complex compaction operations such as bilayer and trilayer tablets or compression-coated tablets

The demonstration of the validity of the continuum-based modelling approach to tablet compaction requires familiarity with fundamental concepts of applied mechanics. Under the theory of such a mechanism, powder compaction can be viewed as a forming event during which large irrecoverable deformation takes place as the state of the material changes from loose packing to near full density. Moreover, it is important to define the three components of the elastoplastic constitutive models which arose from the growing theory of plasticity, that is the deformation of materials such as powder within a die:

1. *Yield criterion*, which defines the transition of elastic to plastic deformation
2. *Plastic flow potential*, which dictates the relative amounts of each component of plastic flow
3. *Evolution of microstructure*, which in turn defines the resistance to further deformation

It is also known that the compression process can be described using static and dynamic models. In the case of static models, time is not considered, although it is a very important factor in the deformation process. The viscoelastic reactions are time dependent, especially for the plastic flow.

Recently Picker [21] proposed a three-dimensional (3D) model to help explaining the densification and deformation mechanism experienced by differently deforming materials during compression. According to the author, a single description of the processes during tableting is possible, and thus densification and deformation properties can be clearly distinguished with a single model. This issue has been investigated over the last years, and a comprehensive approach has been developed for the analysis of compaction using continuum mechanics principles. This approach is based on the following components:

1. Equilibrium equations (balance of forces transmitted through the material)
2. Continuity equation (conservation of mass)
3. Geometry of problem
4. Constitutive behavior of powder (stress–strain behavior)
5. Boundary conditions, including loading (e.g., displacement and velocity) and friction between tooling and powder
6. Initial conditions (e.g., initial relative density of powder)

Due to the significant nonlinearity in material properties and contact stresses, a typical powder compaction problem cannot be solved analytically without major simplifications, and thus a numerical approach is required.

The tableting properties of materials also depend on their deformation behavior. It is apparent that the tablet tensile strength is a strong function of the plastic work required for its formation but not a function of the elastic work recovered. Consequently, it is likely that strong and ductile interparticle functions, whose formation dissipated a significant plastic work, result in strong and tough compacts [4].

It should be mentioned that the material parameters do change with compaction and the use of constant material values, which is often applied, is not necessarily appropriate given the evolving microstructure of the deforming powder. The experimental characterization and accompanying analysis allow these material properties to be evaluated within the comprehensive framework of continuum mechanics, which can be useful in analyzing and predicting the effects of constitutive behavior, friction, geometry, loading schedule, and initial condition, for example, initial relative density and powder fill configuration.

Ruegger and Celik [22] investigated the effect of punch speed on the compaction properties of pharmaceutical powders with one particular objective: to separate out differences between the effect of the compression and decompression events. Tablets were prepared using an integrated compaction research system. The loading and unloading speeds were varied independently of one another. In general, when the compression speed was equal to the decompression speed, the tablet crushing strength was observed to decrease as the punch velocity increased. When the compression speed was greater than or less than the decompression speed, the results varied, depending on the material undergoing compaction. The authors also stated that the reduction of the unloading speed had a similar effect on the direct-compression ibuprofen; however, even greater improvement in the crushing strength was observed when the loading speed was reduced. As a major conclusion, it was demonstrated that the strength of tablets can be improved and some tableting problems such as capping can be minimized or prevented by modifying the rates of loading/unloading.

It is important to notice that, in the case of interacting materials, the compatibility of a binary mix will depend mostly on the compatibility of the percolating material [23]. Accordingly, several industrial applications can be made over these findings. In the development phase, it is possible to modify the formulation of interacting systems to increase the drug content without losing the compatibility of the mix, whereas in the production phase, it is possible to increase the compatibility of a poorly compatible active ingredient by sieving or preferably by milling an excipient with good compression qualities without changing the composition of the mixture.

## 4 TABLET COMPRESSION

The process of tablet compression is divided into three stages: filling, compression, and ejection of the tablet (see Figure 1).

During the first stage of a compression cycle the lower punch falls within the die, creating a cavity which will contain the powder or granulation product that flows

from a hopper. The fill volume is determined by the depth to which the lower punch descends in the die. At this moment the particles of the powder or granulation product flow with respect to each other, thus resulting in a close packing arrangement and the physical characteristics of the material (particle size, particle size distribution, density, shape, and individual particle surface properties) associated with process parameters such as flow rate and compression rate, and the relationship between the die cavity and the particle diameter will define the number of potential bonding points between the particles. The packing characteristic of the product to be compressed is greatly affected by the shape of the particles. Since the product to be compressed comprises components of different nature, the voidage of a closely packed system is considerably changed.

When the upper punch goes down, its tip penetrates the die, confining the powder or granulation product, letting the particle rearrangement stage to continue and initiating the compression stage as the compression force is applied. As a result, forces resulting from the compression force are transmitted through the interparticulate points of contact created in the previous stage. The porosity of the powder bed is gradually decreased, the particles are forced into intimate proximity to each other, and stress is developed at the interparticulate points of contact. Once the particles have formed contacts, they will deform plastically under the applied load. Deformation of the particles will be characterized by elastic, plastic, fragmentation, or a combination of these phenomena, which will depend on the rate and magnitude of the external applied load, the duration of locally induced stress, and the physical properties of the product under compression. When the particles are in sufficiently close proximity, they are bonded. Particles bond as a result of mechanical interlocking, which is described as entanglement of the particles, phase transition at the points of contact, and intermolecular forces, namely the van der Waals force, hydrogen bonding, and ionic bonding.

After formation of the tablet by application of a compression force follows the decompression stage, where the compression force is removed and the upper punch leaves the die. Then, the formed tablet undergoes a sudden elastic expansion followed by a viscoelastic recovery during ejection when the lower punch moves upward.

## 5 EQUIPMENT FOR TABLET COMPRESSION

The equipment employed for tablet compression is generally categorized according to the number of compression stations and dislocation mode. Therefore, eccentric model presses have only one compression station (one die and one pair of punches, upper and lower) while rotary models have multiple compression stations (each station with one die and one pair of punches, upper and lower). The basic difference between the two types of compression equipment is that for eccentric models the compression force applied during compression is due to the upper punch whereas for rotary models it is mainly applied by the lower punch.

A rotary tablet press machine (Figure 2) comprises a housing in which the compression set and subsets (upper and lower roller assemblies) are mounted, the turret head, the upper cams, the weight control assembly and the lower cams, the hopper,

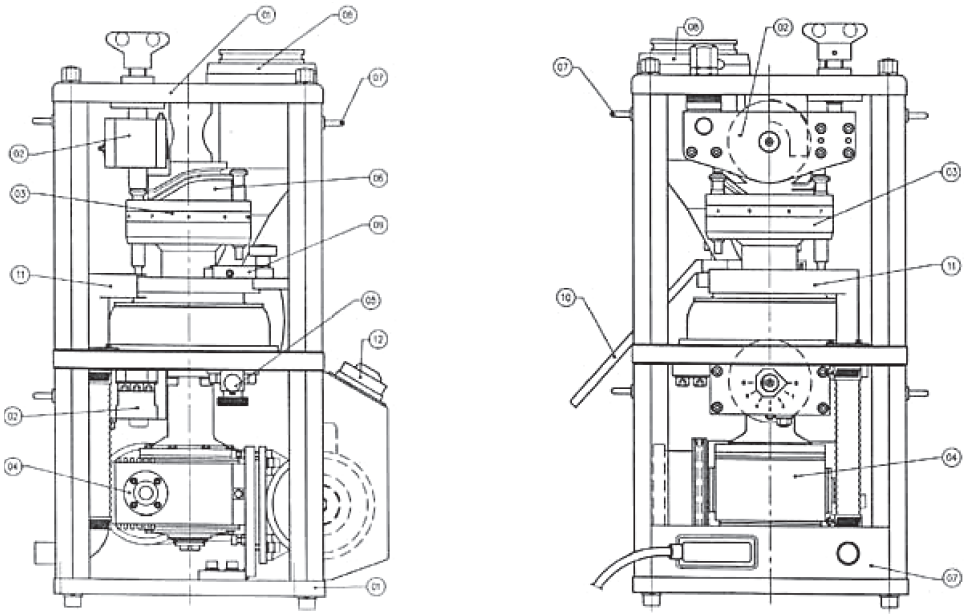

(*a*) (*b*)

**FIGURE 2** Rotary tablet press machine: (*a*) left-side view; (*b*) black-side view. (1) Cabinet, (2) compression, (3) turret, (4) gear, (5) weight control assembly and lower cams, (6) plate cams, (7) guarding, (8) hopper system, (9) feed frame assembly, (10) take-off chute, (11) aspiration assembly, (12) electrical system.

the feeder assembly, the take-off chute, the aspiration assembly, the gear box and the electrical unit, and the lubrication system.

The compression zone is located on the back side of the equipment and employs a maximum load force limited by the type of tooling being used. It is of paramount importance to note that, if a load force is applied over the indicated limit, the press unit will not function properly, resulting in premature wear or possible damage to the tooling. The compression set comprises the hopper and feeder system, the die table, the upper and lower compression rollers, the upper and lower turrets, the excess-material scraper, the tablet stripper, the recirculation channel, and the aspiration system.

The hopper is usually made of stainless steel and has the shape of a funnel to contain and deliver the product to be compressed. It may be provided with a window for the observation of the product level and may also be provided with low-level sensors that signal an alarm, shut off the engine, or activate the feeding mechanism to deliver the product when it falls below this level. The feeder system usually consists of three sections (in the case of force feeders) and is ideal for press performance at high speed. The first section of a force feeder system is where the hopper is connected and is responsible for the flow of the product from the hopper to the next sections. The second section is where the die cavities are filled to their maximum capacities, and the third section is where the weight control adjustment takes place. These sections contain paddle systems which prevent packing of the product. The

speed of the paddles is adjustable and should be synchronized with the die table in order to prevent tablet weight variation. Better adjustment of the paddle speed could be achieved when keeping the lowest standard deviation of the compression force. The feeder system height above the die table is usually kept between 0.05 and 0.10 mm. When the product to be compressed is of very fine particles, this height should be kept at 0.025 mm.

Presses are commonly equipped with a powder aspiration system which is connected to a vacuum source in order to remove excess powder from the die table. This assembly is essential for a high-speed press working for extended periods of time. Special attention must be taken when the powder product comprises an active ingredient of fine particle size. In this case, aspiration should be minimal in order to prevent loss of the active ingredient.

The compression subsets comprise the upper roller assembly and the lower roller assembly. The upper roller assembly is located on the roof plate of the press and utilizes an adjustment system for the regulation of the insertion depth. The lower roller assembly is located on the underside of the die table and utilizes a device for the regulation of the tablet edge thickness.

The turret head is fixed to the main shaft of the gear box. It is manufactured in two pieces (upper and lower) which guarantee the alignment between punches and dies. The gear box is mounted on the lower section under the die table and is responsible for transmitting the draft movement of the motor toward the turret head.

The upper cams are responsible for guiding the upper punches around the circumference of the turret head. It comprises the filling stage track, which guides the upper punches in an up position during its passage over the feeder system; the upper lowering cam, which guides the upper punches down in order to keep their tips covering the cavities (precompression position) and directs the upper punches to their compression stage; the upper compression roller, which guides the upper punches to their compression position; and the upper filling cam, which guides the upper punches back to the filling track.

The weight control assembly, which comprises the weight adjustment cam, is located in the lower section of the press and is regulated by an adjustment system. The lower cams are also located in the lower section of the press and comprise the preweight control (or fill cam), the weight adjustment cam, the lower lowering cam, and the ejection cam. The preweight control guides the lower punches to the full-fill position. The weight adjustment cam guides the lower punches up to the desired fill position. The lower lowering cam guides the lower punches to the precompression position. The ejection cam guides the lower punches and the formed tablets to the discharge position. It is recommended to operate the weight adjustment cam in the approximate center of the fill cam and because of this the fill cam is removable and available in different sizes having a range of approximately 10 mm with an increment range of 4 mm. The choice of the adequate fill cam for the operation of a tablet press with a particular product should be based on the density of the product. According to Figure 3, the fill cam can be adequately chosen when taking into account the density of the material to be compressed and thus the material column height in the die cavity.

Rotary tablet presses could be designed to be single, double, or triple sided. A single-sided press comprises one hopper, one set of compression rolls, and one take-off chute unit whereas double- and triple-sided presses comprise two and three each

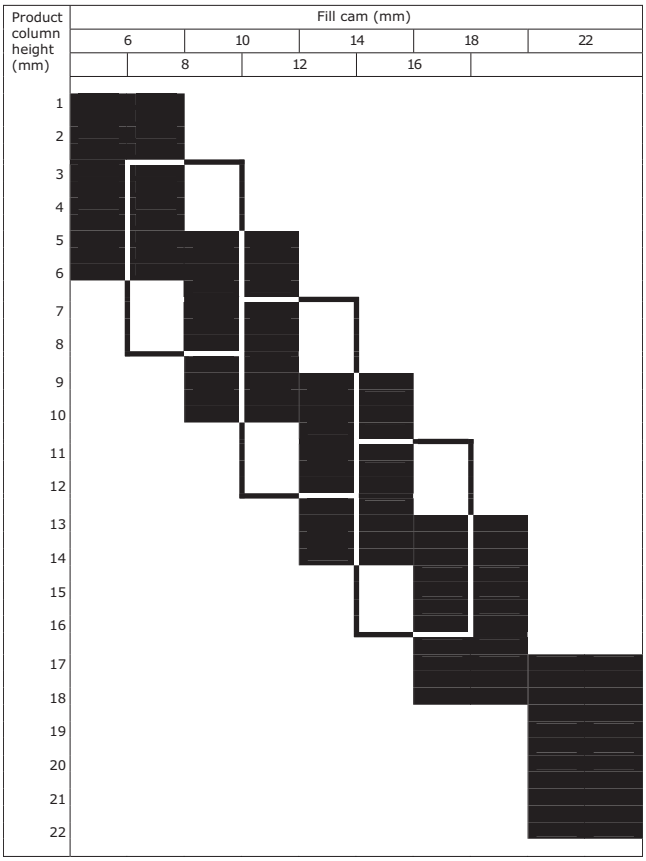


**FIGURE 3** Fill cam depth as function of product column height.

of these units, respectively. Irrespective of the design of the rotary tablet press, the compression cycle is described as follows (Figures 1 and 4).

The powder or granulation product contained in the hopper flows to the feeder which spread the product through a large area over the die table in order to provide enough time to fill the die cavity. The die cavity is created when the fill cam guides the lower punches to the full-fill position and enters the feeder area. Note that the die cavity is filled with an excess of the product at this stage of the cycle. Right after, the weight adjustment ramp and head guide the lower punches to the desired fill position. The excess of the product is removed by the scraper and is pushed back by the excess product stripper when entering the recirculation channel. At this stage the lower punches are guided to the first and second lower compression rols (precompression and main compression rolls, respectively) while the upper punches are guided by the upper lowering compression roll to the precompression position and to the compression position by the main compression roll. As the upper punch penetrates the die cavity until a predefined height, the main compression roll applies the compression load over the lower punch, compressing the product in the die. Soon after compression, the upper lifting cam allows the upper punch to leave the die cavity. Simultaneously, due to the ejection cam, the lower punch is pulled,

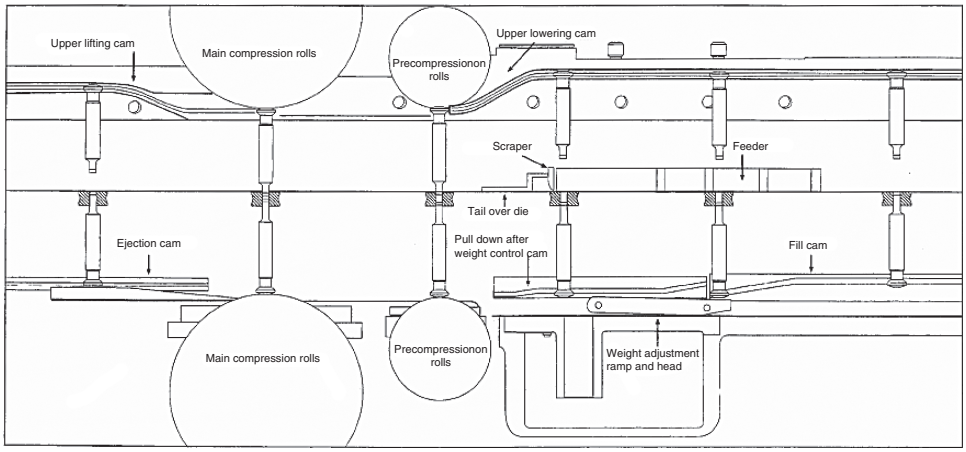


**FIGURE 4** Compression cycle on rotary tablet press. (Courtesy of Thomas Engineering.)

ejecting the formed tablet to the die table. The ejected tablets are then stopped by a scraper and allowed to escape through a chute and collected. At this time the fill cam geometry makes the lower punch go down and a new compression cycle begins.

## 6 TABLET PRESS TOOLING

Punches and dies are essential tools in the tableting process and therefore are critical to the quality of the tablets produced. Both tools are designed for long life under normal conditions of working, but, in spite of this, they are not proof against careless handing.

It is important for those working with a tablet press to be familiar with the terminology used in the industry concerning the punches and dies. Table 1 describes the commonly used terms related to press tooling. Some of the press tooling parts can be identified in Figure 5.

When considering a tableting operation, it is important not only to select the appropriate press tooling in terms of dimensional data but also to consider the material of which tools are made. Performance of the press tooling will in part be a function of the material selected for its manufacture. Usually the material and hardness of the compression tooling are left to the manufacturer's discretion. There are various types of steel available for the manufacture of press tooling. It is important to recognize the individual characteristics of the steel regarding its composition and the percentage of each constituent element. Usually only a small amount of alloying element is added to steels (usually less than 5%) for the purpose of improving hardness and strength corrosion resistance, stability at high or low temperatures, and control of grain size. Some of these elements are as follows:

1. *Carbon* Principal hardening element. As the carbon content increases, its hardness increases. Increases the tensile strength of the steel.

**TABLE 1 Tooling Terminology**

| | |
|---|---|
| Band | Area between opposing cup profiles formed by die wall |
| Bakelite tip relief | Undercut groove between lower punch tip straight and relief; assures sharp corner to assist in scraping product adhering to die wall |
| Barrel/shank | Surface controlled by turret punch guides to ensure alignment with die |
| Barrel-to-stem radius | Provides smooth transition from tip length to barrel |
| Barrel chamfer | Chamfers at ends of punch barrel, eliminates outside corners |
| Barrel flutes | Vertical slots machined into punch barrel to reduce bearing surface and assist in removing product in punch guides |
| Cup depth | Depth of cup from highest point of tip edge to lowest point of cavity |
| Die | Component used in conjunction with upper and lower punches; accepts product for ocmpaction and is responsible for tablet's perimeter size and configuration |
| Die bore | Cavity where tablet is made, shape and size determine the tablet |
| Die chamfer | Entry of die bore |
| Die groove | Groove around periphery of die to allow die to be fixed in press |
| Die height/ overall length | Overall height of die |
| Die lock | Mechanism used to lock die in position after it is installed in die table |
| Die outside diameter | Outside diameter of die, compatible with die pockets in press |
| Die taper | Gradual increase in die bore from point of compaction to mouth of bore, assists ejection |
| Head | End of punch which guides it through press cam track |
| Head/dwell flat | Flat area of head that receives full force of compression rolls at time that tablet is being formed |
| Inside head angle | Area of contact with lower cam and upper cam |
| Key | Prevents rotational movement of punches ensuring alignment to shaped and multihole dies |
| Keying angle | Relationship of punch key to tablet shape; position will be influenced by tablet shape, take-off angle, and turret rotation |
| Land | Area between edge of punch cup and outside diameter of punch tip |
| Neck | Relieved area between head barrel which provides clearance for die |
| Outside head flat angle/radius | Contact area with press cams and initial contact with pressure rolls |
| Overall length | Total punch length as measured from head flat to end of tip |
| Tip face/cup | Portion of punch tip that determines contour of tablet face including tablet embossing |
| Tip length | Straight portion of stem effective inside die bore |
| Tip relief | Portion of punch stem which is undercut or made smaller than punch tip straight; most common for lower punches in order to reduce friction from punch tip and die wall |
| Tooling set | Complete set of punches and dies to accommodate all stations in tablet press |
| Tooling station | Upper punch, lower punch, and die which accommodate one station in tablet press |
| Relief/undercut | Mechanical clearance between stem and die bore, sharp edge between tip straight and undercut areas acts to clean die |
| Stem | Area of punch opposite head which begins at end of barrel and extends to tip |
| Working length | Length of punch from bottom of cup to head flat; together, upper and lower working lengths control tablet thickness and weight; also known as overall length, bottom of cup (OLBC) |

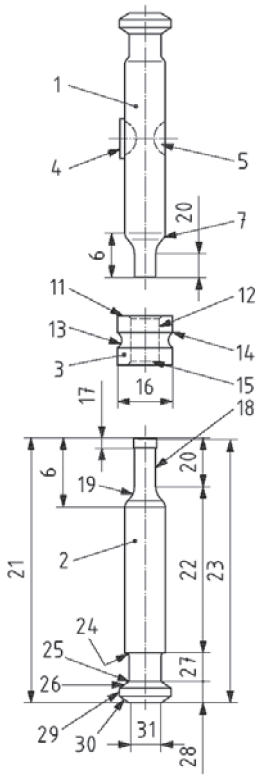


**FIGURE 5** Identification of common parts of press tooling (upper and lower punches and die) according to ISO 18084, 2005: (1) upper punch, (2) lower punch, (3) die, (4) key, (5) land, (6) stem, (7) barrel-to-stem chafer, (8) cup depth, (9) tip face, (10) blended land, (11) face, (12) bore, (13) die grove, (14) protection radius or shoulder, (15) chamfer or radius, (16) outer diameter, (17) tip straight, (18) relief, (19) barrel-to-stem radius, (20) working length of tip, (21) overall length, (22) barrel, (23) working length, (24) barrel-to-neck radius, (25) neck-to-head radius, (26) inside head angle, (27) neck, (28) head, (29) head outer diameter, (30) outside head angle, (31) head flat, (32) key orientation angle, (33) upper punch face key position, (34) barrel diameter.

2. *Manganese* Increases ductility and hardenability of the steel. Also increases the rate of carbon penetration during carbonizing and imparts excellent wear resistance.
3. *Nickel* Improves the toughness and impact resistance of the steel and mildly increases its hardness.
4. *Chromium* Increases the hardness of the steel and improves its wear or abrasion resistance. It helps to limit grain size. If added in amounts greater than 5%, it can impart corrosion and wear resistances.
5. *Molybdenum* Improves hardenability and increases tensile strength of the steel.
6. *Vanadium* Produces a fine grain size and improves fatigue strength of the steel, just like molybdenum.

7. *Tungsten* Is used in tool steels to maintain hardness at elevated temperatures.
8. *Copper* Increases corrosion resistance; nevertheless its content has to be controlled, otherwise the surface quality and hot-working behavior are compromised.

The carbon steels comprise alloying elements not exceeding the defined limits of 1% carbon, 0.6% cooper, 1.65% manganese, 0.4% phosphorus, 0.6% silicon, and 0.05% sulfur. For the alloy steels, the limits exceed those for the carbon steels and may also include elements not found in carbon steels. The alloy steels have a specific designation according to the American Iron and Steel Institute (AISI). Such designation is a four-digit number where the first digit stands for the class of the alloy (e.g., 1, carbon; 2, nickel-chromium; 3, molybdenum; 4, chromium), the second digit designates the subclass of the alloy, and the last two digits designate the amount of carbon in 0.01%. The stainless steels comprise at least 10% chromium with or without the addition of any other alloying element. The tool steels are carbon steel alloys with an excess fo carbides which impart hardness and wear resistance. According to the AISI, tool steels are grouped as water hardening (W), shock resisting (S), cold-work oil hardening (O), cold-work medium-alloy air hardening (A), cold-work high-alloy high chromium (D), low alloy (L), carbon tungsten (F), low-carbon mold steels (P1–P19), other mold steels (P20–P39), chromium-based hot work (H1–H19), tungsten-based hot work (H20–H29), molybdenum-based hot work (H40–H59), high-speed tungsten based (T), and high-speed molybdenum based (M).

The appropriate steel for press tooling should be selected based upon the toughness and wear resistance required by the application, and therefore it is mandatory to have satisfactory knowledge regarding the abrasiveness, corrosiveness, and lubricity of the product intended for compression as well as the desired dimensions of the punch tip. The toughness of the steel regards its ability to resist shock and its wear resistance regards the ability to resist physical damage or erosion due to product contact.

For a clear understanding, the Thomas Engineering *Press Tooling Manual* [24] states the following:

> Punches manufactured from high carbon/high chromium steel may exhibit improved wear resistance characteristics, however under extreme compression force, the cup may crack due to the brittle nature of the steel. Steels with lower carbon and chromium levels will act conversely. While these steels may be useful in some applications, the majority will require a more moderate balance of toughness and wear resistance. Steel selection for dies is not as critical. In most cases, high wear resistance steel is preferred.
>
> The bulk of pharmaceutical tablet press punches are manufactured from S1, S5, S7, or 408 (11% chromium, 8% nickel) tool steel. The S series steels provide a good combination of shock and wear resistance and have a proven record of performance in tableting operations. At one time, 408 or 3% nickel steel was the industry standard because of its superior shock resistance toughness. The S grades however, which have only a slight loss in ductibility by comparison, offer much improved wear characteristics and have all but replaced 408 as the preferred general purpose punch steel. A2, D2 and D3 are high carbon/high chromium steel used for their excellent wear resistance. Among all the steels commonly used for press tooling, D3 has the highest wear resistance. However,

its low toughness rating typically limits its use to dies only. D2 rates slightly lower in abrasion resistance than D3 but its increased toughness makes it suitable for punch use, provided the cup design is not too fragile. A2 is a compromise between the general purpose S grades and D2 in both toughness and wear. It can be used for punches as well as dies.

Tungsten carbide, while not actually a steel, is extremely wear resistant and is commonly used to line dies. Punch tips can be manufactured from tungsten carbide; however, the cost of tooling is quite high and restricted to applications where tip fracture due to high compression forces is not likely.

Ceramic materials such as partially stabilized zirconia can also be used as die liners. Ceramics offer high wear and corrosion resistance and lower tablet ejection forces than either steel or carbide due to their low coefficient of friction.

S1, S7 and 408 provide some protection against mildly corrosive materials. More severe corrosion problems however, demand the use of stainless steel (440C) tooling. From the standpoint of wear, 440C falls between the S and D grades of tool steel. Its low toughness rating (comparable to D3) requires a strong cup design if tip fracture problems are to be avoided.

One measure of tool steel quality is the rate of inclusions. Inclusion are unwanted impurities or voids and are present to some degree in all steels. After heat treatment, inclusions give rise to localized areas of stress concentration where microscopic cracks can later develop. Remelting of the steel at the foundry will further reduce a tool steel's level of impurities; therefore improving the quality of the steel and subsequently its performance in the tooling environment. In cases where punch tip fracture is a problem, tooling suppliers may recommend a remelted or premium grade of particular steel as a means of eliminating the problem.

Concerning the configuration of compression tooling, the most commonly used are the so-called B (19 mm, or 3/4 in.) and D (21 mm, or 1 in.) tooling types. Additionally, these two types are classified into three specifications: the North American TSM (*Tableting Specification Manual*) [25], the European Union (EU) standard, and the Japan Norm (JN). The North American TSM is used in the United States and is the only standard officially supported by the governing body and published by the American Pharmacists Association. The EU standard and the JN are generally used in Europe and the Far East, respectively. In spite of the existence of these standard specifications, there are tablet press manufacturers that use their own configurations for tooling which have the disadvantage of being restricted to a specific tablet press. Figures 6 and 7 depict the three standard configurations of compression tooling. In addition, the International Organization for Standardization (ISO) standard 18084:2005 [26] comprises specifications of the main dimensions, including tolerances and characteristics of punches and dies.

Regarding the importance of compression tooling to the performance of the tablet press and the quality of the tablets, it is of paramount importance that punches and dies are handled with care. The first criterion is the identification of the tooling; that is, punches and dies should be identified according to the standard and be designated by "upper punch without key," "upper punch with key," "lower punch with key," "lower punch without key," or "die," the reference of the standard (e.g., TSM, EU, JN, ISO), and the punch or die diameter. Punches and dies should also have a marking that includes at least the manufacturer's identification, the number of the punch in the series, and/or the identification number. Upon

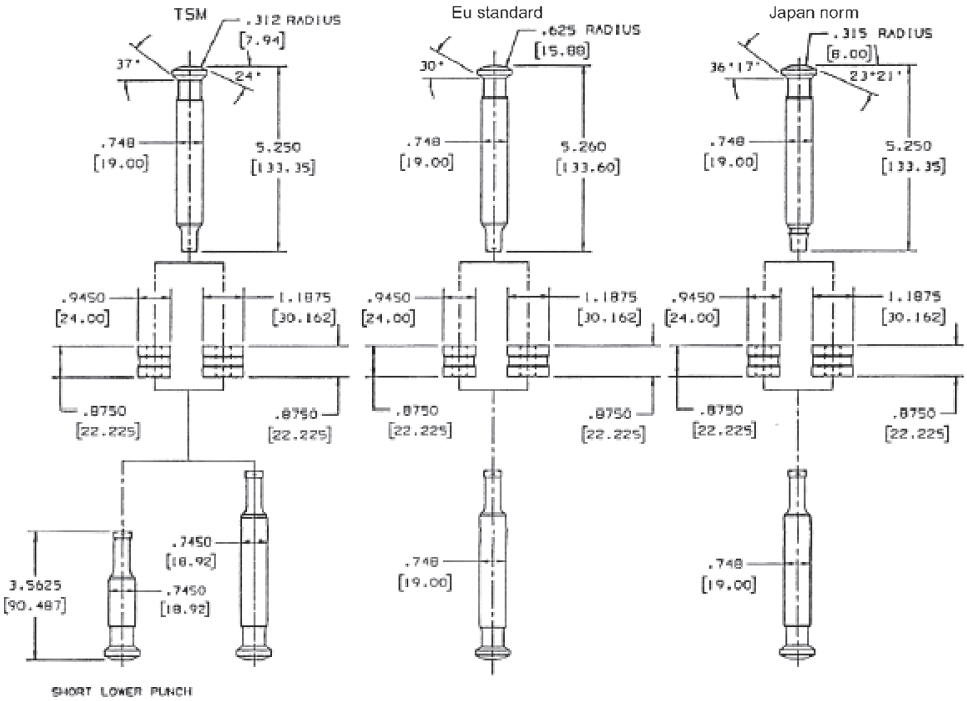


**FIGURE 6** Tooling standards configurations.

receipt, after manufacturing, and prior to inspection, the punches and dies should be carefully and thoroughly cleaned and dried. Then, tooling should be lightly oiled, packed, and stored in a dry, cool place.

Damage to the punches and dies should be avoided. Therefore, they should not be transported from place to place without protective package. During transportation, installation, and removal of tooling from the tablet press, cleaning, inspection, and storage, care must be taken to avoid hitting the tips of the punches.

The visual and dimensional inspection of punches and dies should be carried out periodically. Visual inspection should be performed each time punches and dies are installed in and removed from the tablet press. Under normal conditions, slight wear is to be expected. When abnormal or excess wear is detected, the cause should be immediately investigated, inspecting the cams or components which touch the affected area of the tool. The importance of visual inspection resides on the fact that it may ensure the optimum life of the punches and dies, performance of the tablet press, and consistency and appearance of the tablets. In addition to the visual inspection, it is also recommended that dimensional inspection be performed at specific intervals throughout the life span of the punches and dies. The dimensional inspection not only ensures the consistency of hardness, weight, and thickness of the tablets, but also proves to be critical in diagnosing potential and real problem areas with regard to the tableting process and press. A typical schedule for the dimensional inspection may be as follows: 50%, 75%, 85%, 90%, and 95% of the historical or projected life cycle of the punches and dies. Therefore, the history or data base should be maintained for each set of tooling. Nevertheless, there is no general agreement on what dimensions of the punches and dies should be included in a

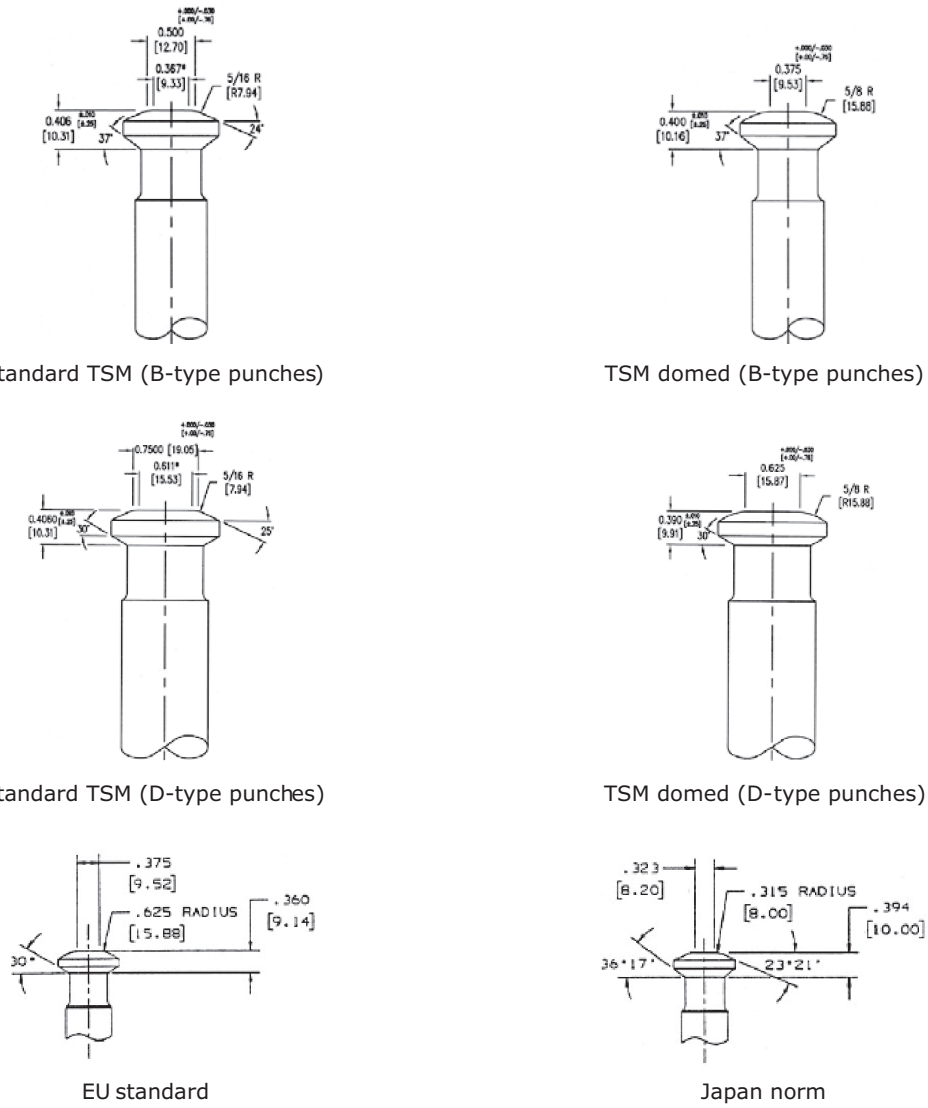


**FIGURE 7** Tooling head configurations.

dimensional inspection. Some believe that a 100% inspection should be carried out while others defend that only critical dimensions (e.g., working length, cup depth, and overall length) should be inspected, believing that measuring any other dimension is either unnecessary, since it rarely if ever changes and therefore is not worthy of the time and expense of measuring, or cannot be properly measured with current equipment and is better served by a visual inspection. What is important when inspecting compression tooling is that the dimensional values are consistent within the set and tolerances and within specifications. Before proceeding with inspection, the measuring instruments should be calibrated to be certain that the dimensional values obtained are accurate and true.

## 7 TABLE ENGRAVING

Engraving is the most common method for tablet surface marking identification. The engraving method could be embossed (letters or symbols are raised on the tablet surface and cut into the punch tip face) or debossed (letters or symbols are cut into the tablet surface and raised on the punch tip face). For engraving on the tablet surface some specifications should be considered: stroke width, angle of engraving, radius, depth, spacing, and engraving area.

Generally, for uncoated tablet application, a stroke width between 15 and 20% of the letter height having an engraving angle of 30° is recommended. The radius should be between 50.8 μm (0.002 in.) and a value derived from dividing the stroke width by 2 times the cosine of the engraving angle. It is important to note that radii smaller than 50.8 μm or exceeding the maximum value are difficult to machine since it may decrease depth and definition of engraving. The depth is a function of the engraving angles, stroke, and radius for a given tablet size and, as a general rule, the depth should not exceed 50% of the stroke width, or no less than 88.9 μm (0.0035 in.). Spacing between letters or symbols should be a minimum of one stroke width. The available engraving area is based upon letter distortion due to the curvature or radius of the cup and thus, as a general rule, letter distortion is defined by the ratio of the outside depth of the engraving to the specified depth. Generally, distortion is present when this ratio exceeds 1:3.

When engraving is considered for fillm-coated tablets, the recommended stroke width should be the same as recommended for uncoated tablets. The recommended engraving angle should be 35°. However, engraving angles up to 40° can be used in extreme applications to allow coating solution flow. For stroke widths of 203.2 μm (0.008 in.) or less, a 30° angle is recommended to maintain minimum engraving depth. The recommended radius should be between 76.2 and 152.4 μm and a value derived by dividing the stroke width by 2 times the cosine of the engraving angle. It is important to note that radii will be determined by the flowability of the coating solution and coating process. As a general rule, stroke depth should be at least 177.8 μm. However, shallower depths can be used, provided that the film coating process is properly developed taking this factor into account. The spacing between the letters and symbols and the available engraving area considerations are the same as for uncoated tablets. The following equations can be used to determine stroke width and engraving radii:

$$S = 0.15H \leftrightarrow 0.20H \tag{6}$$

$$R_{\min} = 50.8\,\mu\text{m} \qquad R_{\max} = \frac{S}{2\cos\theta} \tag{7}$$

## 8 TABLET SHAPE AND PROFILE

The more popular standard geometric shapes of tablets are the round and the caplet shapes. Other tablet shapes include the oval, elliptical, square, diamond, rectangular, and polygonal. The shape of tablets plays an important role in terms of aesthetics, process (printing, film coating, packaging, and shipping), and acceptability by the consumer (identification, help with swallowing).

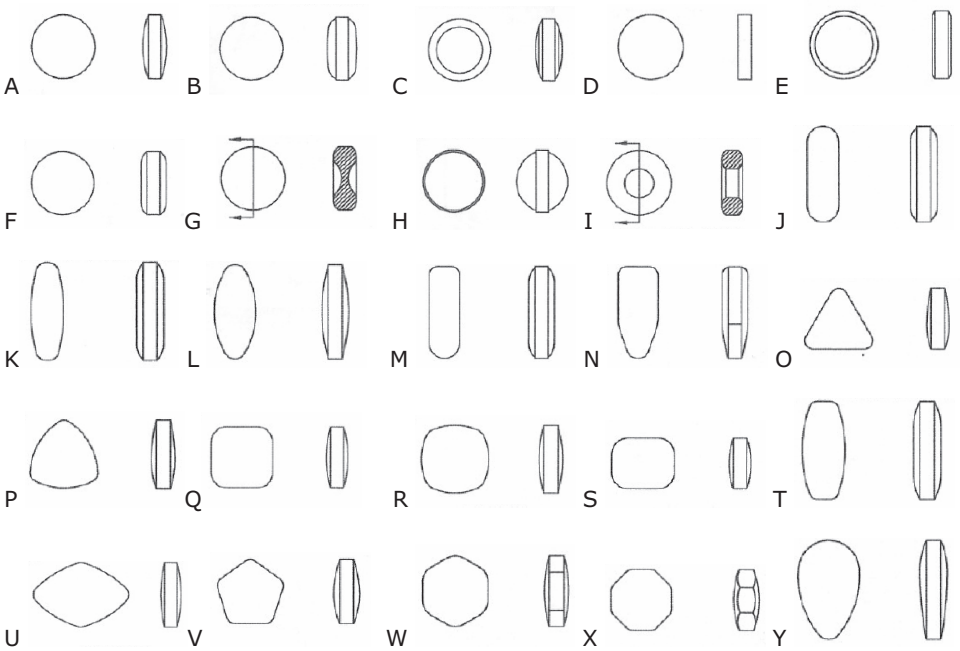


**FIGURE 8** Common tablet shapes and profiles: (A) standard convex, (B) compound cup, (C) convex beveled, (D) flat faced plain, (E) flat faced bevel edged, (F) flat faced radius edged, (G) lozenge, (H) modified ball, (I) core rod with hole in center, (J) capsule, (K) modified capsule, (L) oval, (M) bullet, (N) arrow head, (O) triangle, (P) arc triangle, (Q) square, (R) pillow or arc square, (S) rectangle, (T) modified rectangle, (U) diamond, (V) pentagon, (W) hexagon, (X) octagon, (Y) almond.

In terms of design, the profile of a tablet also plays an important role in the aesthetics, packaging, orientation for printing processes, and handling. The profile of a tablet is important in the film coating process and even in helping with the swallowing. Applying a bisect score onto the tablet surface enables the tablet to be easily divided into smaller dosage amounts.

Nowadays it is common for tooling suppliers to use software to provide 2D and 3D technical drawings of tablets and tooling. Such software may provide accurate details of tablets and tooling using only tablet dimensions as input and therefore enables fast evaluation by the manufacturing department prior to ordering prototypes. Figure 8 illustrates some tablet shapes and profiles.

Flat-face, bevelled-edge tablets have many advantages due to their flatness, which provides the most compact tablet weight per volume weight, uniform hardness since the compression force is exerted evenly on the cup face, and engraving with no distortion. This tablet profile proves to be ideal for small tablets, especially when engraved, although the engraving area may be limited by the 381 μm (0.015 in.) radius on the bevel. On the other hand, compression tooling displays an inherent weakness in the punch cup design at the point where the bevel edge meets the cup flat. Attention must be paid since these types of tablets cannot be coated as they will stick together, or twin.

Shallow and standard concave tablets have the great advantage of displaying a maximum allowable area available for engraving without distortion as a result of the moderate curvature of the cup profiles and the absence of a bevel. The shallow and standard cup configurations are the strongest profiles per punch tip diameter. In addition, such profiles allow consistent distribution of the compression force over the cup face due to the slight curves involved in the cup, thus contributing to the production of tablets of uniform hardness. Nevertheless, caution must be taken concerning the cup depth since when it approaches the cup edge it may be less than the depth of the engraving. The major disadvantage of these profiles may be due to the angle of the cup profile to the tablet sidewall, which may lead to chipping at the tablet edge during film coating or handling.

Caplet-shaped tablets are easier to swallow, aesthetically pleasing, and chipping at the tablet edge generally does not occur during film coating or handling due to the angle of the cup profile to the tablet sidewall. Nevertheless, the increased curvature of the cup reduces compression force by approximately 50% compared to the shallow and standard concave profiles. Distortion of engraving may also be a problem because of the more extreme curvature. During film coating caplet tablets have the potential of sticking together, or twinning, as the tablet sidewalls are parallel. This problem could be alleviated by applying a 76.2 mm (0.003 in.) drop (15.24–20.32 cm, or 6–8 in. radius) to the sidewalls.

The concave oval profile displays a maximum allowable area for engraving and a uniform distribution of the compression force over the cup face. Such an advantage is a consequence of the mild curvature of the cup profile and absence of a bevel. Structurally, concave oval tablets are the strongest of the non-round-shaped tablets. However, due to the angle of the cup profile to the tablet sidewall, chipping at the tablet edge may occur during film coating or handling.

The compound cup profile could be used to provide round or oval tablets. This profile provides a good tablet weight per volume but simultaneously presents a weak cup edge, thus being the weakest of all cup configurations. Because of this, the maximum compression force is limited to the minor cup radius on the round shapes and the minor cup radius on the minor side for the oval shapes. In addition, the available engraving area is limited to the blending point of the two radii.

## 9 TABLET BISECT

Usually known as score or break line, the tablet bisect has the purpose of easily breaking the tablet in predetermined small dosages. According to the TSM, the bisect types range from the most functional (the pressure sensitive, or type G) to the least functional (partial, or type H). Each bisect type has its own characteristic, as can be seen in Figure 9. Generally, the bisect is placed on the upper punch, especially when its depth exceed 40% of the cup depth, in order to avoid problems during ejection of tablets. Nevertheless, the bisect can be placed on the lower punch either when the upper punch is supposed to contain embossed characters or printing that makes difficult the existence of the bisect or when its depth does not exceed 40% of the cup depth. When it is desired to apply a bisect to the upper tablet's surface but there is interference of engraving or printing, then a modified bisect design should be considered.

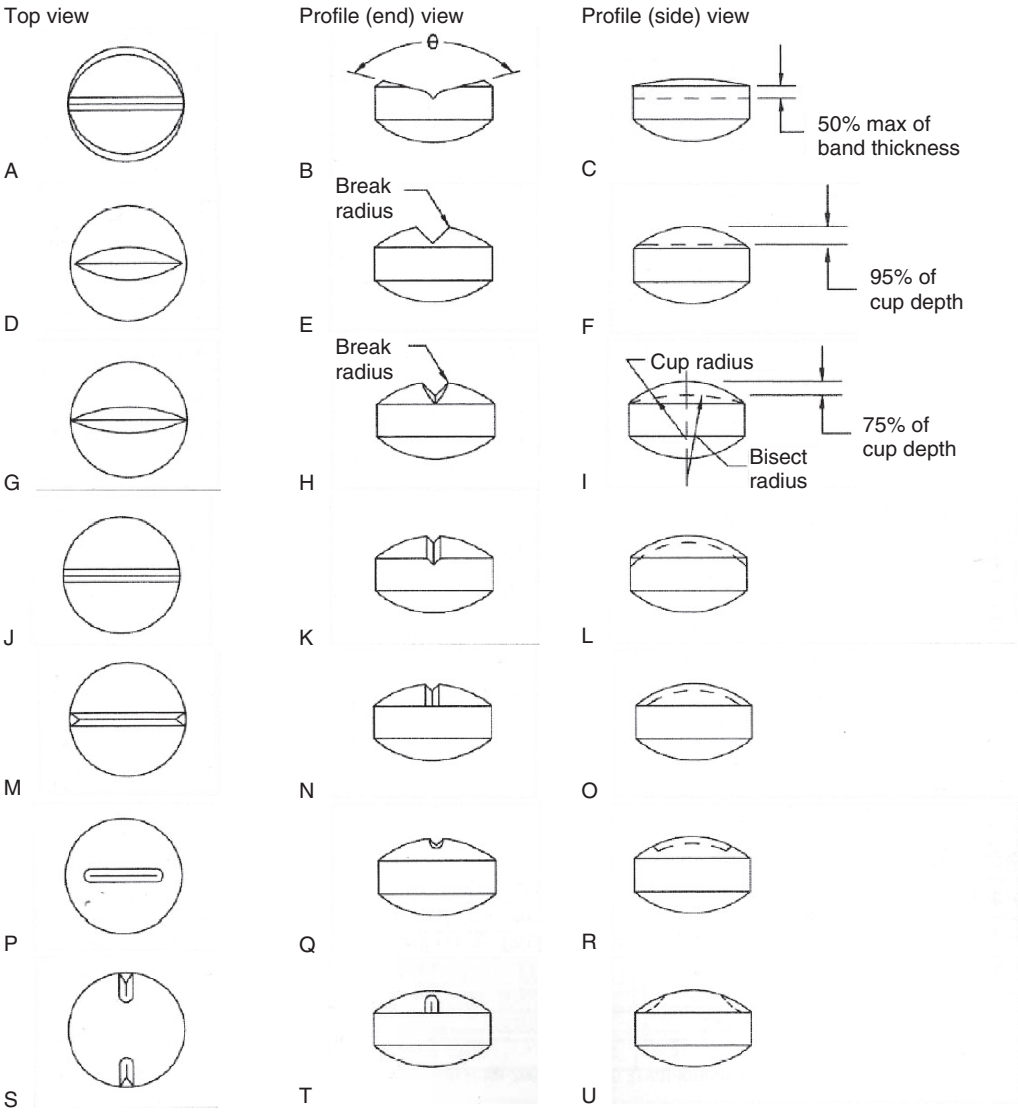


**FIGURE 9** Tablet bisect for concave tooling (according to TSM): (A, B, C) pressure-sensitive (type G), (D, E, F) cut through (type D) or European style, (G, H, I) decreasing (type C), (J, K, L) standard protruding (type A), (M, N, O) standard (type E), (P, Q, R) short (type B), (S, T, U) partial (type H).

When considering applying a bisect to a tablet's surface, careful attention should be taken with respect to the tablet's cup depth, band thickness, and hardness. Considering these aspects, the specifications for the bisect size are determined taking into account the tablet's size, engraving or printing, and desired bisect design.

The TSM acknowledges two different configurations of bisect for concave tablets: protruding and cut flush. The protruding configuration follows the curvature of a radiused cup and extends past the tip edge of the punch. The cut flush configuration

is the most popular bisect configuration since one may experience problems with the protruding configuration. This is explained by the fact that the protruding bisect may run into the tip edge of the lower punch if they become too close during the compression cycle of the press.

Among the bisect styles acknowledged by the TSM, the cut-through, also known as the European style, can only be applied on radiused cup designed tablets. Other styles are the standard, the short, and the partial bisects. Compared to the standard style, the cut-through style is said to have the advantage of letting patients better break the tablet into smaller subunits. On the other hand, because the cut-through is wider at the center, it decreases the available tablet surface area for engraving or printing.

## 10 PROBLEMS DURING TABLET MANUFACTURING

Due to either formulation or equipment, some problems can arise during the tablet compression process, such as capping and lamination, picking and sticking, mottling, double printing, weigh variation, and hardness variation. It is the early detection and accurate diagnosis of any of these flaws that can avoid tablet compression process failure and consequently improve its reliability, safety, reduce process downtime and the overall operating cost.

Often, some of the above-mentioned problems are not detected during the development of a particular tablet formulation, only appearing during scale-up as the processing speed is increased. Some of the problems experienced during tableting can be solved by shifting the formulation or alleviated by altering the tableting conditions.

### 10.1 Capping and Lamination

Capping and lamination are common problems that can be experienced during tableting. Capping is defined as the splitting of one or both lids of a tablet from its body. Lamination is a precursor to capping since it involves the occurrence of layers in a compact parallel to the punch face. Sometimes capping is noticed not during the process but during physical testing, such as friability and hardness.

An incipient theory proposed by Train [27] related lamination to radial elastic recovery of the compacted material during ejection. A once-accepted theory formulated that capping and lamination are the result of air entrapped in the tablet under pressure which tries to escape during ejection [28]. This theory is no longer widely accepted. Disagreement arises from the fact that some formulations cap or laminate even at low press speeds. Today, it is believed that the entrapped air may be related to capping but does not affect lamination.

A widely accepted theory for lamination presented by Long [29] and reformulated by Ritter and Sucker [30] attributes capping to the residual die wall pressure. This pressure is said to cause internal shear stresses in the tablet causing the propagation of cracks, which results in lamination or capping. The propagation of cracks can be prevented by plastic relaxation of shear stresses. Therefore, materials having sufficient plasticity may not be susceptible to lamination. Some properties of the powder mixture, such as moisture content, type and amount of the binder, and

particle size, are important formulation variables that could be assessed in order to impart plasticity, thus diminishing capping and lamination tendencies.

Normally, drugs such as paracetamol, mannitol, ibuprofen, phenazone, and mefenamic acid have poor compression properties and produce tablets that are weak and frequently exhibit capping. Materials that deform elastically or exhibit time dependence are more susceptible to capping and lamination and/or strength reduction, especially as tableting rate is increased. The effect of punch velocity is most marked when transferring a material from an eccentric to a rotary press or when scaling up to larger production size tablet presses.

In addition to the possible causes of capping and lamination discussed previously, one should also consider the possibility that shape of the tooling and tooling defects are sources of capping. In such cases the problem can simply be alleviated by repairing or altering press tooling.

Usually the process of capping can be evidenced as an increase in tablet height within a few seconds after tablets are ejected from the die.

A technique generally applied to characterize and prevent the capping and lamination of a material intended to be compacted is using the brittle fracture index (BFI). The BFI was designed by Hiestand et al. [31] and measures the ability of a material to relieve stress by plastic deformation around a defect. It is obtained by applying Equation (8) and compares the tensile strength of a tablet with a hole in its center ($T_0$), which acts as a built-in stress concentrator defect, with the tensile strength of a similar tablet without a hole ($T$), both at the same relative density:

$$\mathrm{BFI} = 0.5\left[\left(\frac{T_0}{T}\right) - 1\right] \tag{8}$$

It is said that a material showing a moderate to high BFI value (>0.5) is prone to laminate and cap during the process. A low value of BFI is desirable to minimize lamination and capping during tablet production.

Indentation hardness is another measure which finds wide application in the pharmaceutical industry for the assessment of capping and lamination tendency. The indentation hardness measurement employs an indentation hardness tester and is defined as the hardness of a material determined by either the size of an indentation made by an indenting tool under a fixed load or the load necessary to produce penetration of the indenter to a predefined depth. An instrumented indentation hardness tester can be employed for that purpose since it has the ability to measure the intender penetration ($H$) under the applied force ($F$) throughout the testing cycle and is therefore capable of measuring both plastic and elastic deformation of the material under test.

Another technique for the assessment of capping and lamination tendency which has been increasingly employed in the research-and-development phase of tablet manufacturing is acoustic emission. This technique relies on the fact that an abrupt change in stress within a material to be compacted generates the release of a transient strain energy designated as acoustic emission which results in a mechanical wave that propagates within and on the surface of a structure [5, 32]. Thus, this technique can discriminate between capped and noncapped tablets based on comparing the measured level of acoustic emission energy against a decision threshold.

If it is desirable to overcome capping and lamination during the tableting process, the use of ultrasound-assisted presses could be a reliable solution. However, use this technique is still very recent since reports in the scientific literature extend only over the last decade [33].

In general, when capping and lamination are possible problems during tablet manufacture, an option could be the slower removal of force during decompression. This could be useful since capping tendency increases with increasing rates of decompression. However, better improvements could be achieved if the compression and decompression events are treated separately. By determining the effect of reducing either the loading or unloading speeds on the individual materials, it could be possible to increase crushing strength and eliminate or minimize the incidence of capping and lamination to greater extents. Thus, there is the need for a machine that is capable of customizing compaction profiles so that each formulation can be manufactured under an optimum set of conditions.

### 10.2 Picking and Sticking

Picking refers to adherence of powder to the punch surface. It is more problematic when the punch surfaces are engraved with logos or letters such as B, A, or O in order to produce debossed tablets. Sticking occurs when powder tends to adhere to the die leading to the development of an additional pressure to surpass friction between the formed compact and the die wall. As a result, the produced tablets show a rough surface at their edges. Furthermore, sticking can cause picking or damage the press punches by blocking the free movement of the lower punches leading to an increase of compaction pressure.

Various approaches can be used to solve picking and sticking problems during tablet manufacturing, namely optimization of press tooling, process parameters, and formulation. Generally, it is important to find the optimal combination of formulation and process parameters, particularly when market image tablets are to be produced.

In relation to formulation adjustment, an antisticking agent (talc is commonly used for this purpose) can be added to the powdered formulation in order to eliminate picking and sticking during manufacturing. Colloidal silicon dioxide may be the right choice when picking is evident since this excipient can impart smoothness to the punch surfaces. However, when adding colloidal silicon dioxide to the powder formulation, it would be necessary to add an extra lubricant in order to avoid sticking and facilitate ejection of tablets from the dies. In addition to the need for an extra excipient in the powder formulation, press tooling may need to be adjusted to improve tableting. For the production of market image tablets, logo or letters on the punches should be as big as possible. Additionally, punch tips may be plated with chrome in order to give a smooth and nonadherent surface.

When a lubricant such as stearic acid or propylene glycol or any other raw material of low melting point is present in the powder formulation, the heat generated during tableting may cause softening of these ingredients, thus leading to sticking. To overcome this problem, it may be needed to refrigerate the powder load to be tableted or to equip the press machine with a cooling unit.

## 10.3 Mottling

Mottling is defined as an uneven coloration of tablets or nonuniformity of color over the tablet surface. One of the possible causes of mottling may be the difference in color between the active principle and excipients, but sometimes it may be the result of degradation of the active ingredient which imparts spot zones over the surface of the tablets.

Nonetheless, when colored compressed tablets are needed for aesthetic reasons, the foremost cause of mottling is dye migration to the periphery of granules during the drying process [34]. To overcome this problem, one should consider changing the solvent used for wet granulation or the binder agent, using a low drying temperature, or decreasing the particle size of the excipient. Another way to overcome mottling was demonstrated by Zagrafi and Mattocks [35] and suggests the inclusion of an adsorbent agent such as wheat or potato starch to the formulation. The adsorbent agent is said to adsorb the dye, retarding its migration then decreasing mottling.

## 10.4 Weight and Hardness Variation

Weight and hardness variation are common problems experienced when tableting. Tablet weight is mainly affected by factors such as powder variation, tablet press condition and tooling, and flow of powder on the tablet press.

Inconsistent powder or granulate density and particle size distribution are common sources of weight variation during tablet compression. Problems related to the density of the powder or granulate are often associated with overfilling of the die and recirculation of the product on the tablet press. A variation of particle size distribution of the powder or granulate can be the result of segregation due to transfer or static electricity. It might also vary because the product cannot withstand the handling and mechanical stress it undergoes before reaching the tablet press.

Weight variation can arise as a result of a poorly prepared or operated tablet press. To solve this problem, one should inspect the press performance. Attention must be taken when dealing with a new die table on a load tablet press. In such a case, operation of the tablet press must regard the up-and-down motion of the punches within 76.2 μm of the setting without neglecting the conditions of the pressure rolls and cams.

Inspection of the critical dimensions of tablet press tools is recommended. At least three dimensions of the upper and lower punches should be inspected: the working length, the cup depth, and the overall length. The working length is the key factor affecting tablet weight. Therefore, the length of each punch must be correct and identical. The cup depth and the overall length are not critical with regard to controlling tablet weight. Therefore comprehensive inspection and evaluation of the press tooling are essential to minimize deviation of tablet thickness, weight, and hardness.

During the course of a compression operation it is also important to not neglect the level of the product in the hopper. Head pressure is a critical factor related to the amount of product in the hopper. The more product present in the hopper, the greater the head pressure, and vice versa. Therefore, when the head pressure varies, so does the weight of the tablets. So, in order to maintain a constant head pressure,

thus reducing a potential variation of weight, compression should be conducted within a narrow range of the powder or granulate product in the hopper.

The fill cam is another factor that can have a profound effect on tablet weight. The choice of an adequate fill cam regarding some characteristics of the powder or granulate product allows the die cavity to be properly overfilled. Usually, in order to maintain consistent tablet weight during compression, it is recommended to overfill the cam by 10–30% of its volume. Basically, any tablet press part that is ultimately related to the powder product flow can have a mild or profound impact on weight control. It is important to remember that the scraper blade tends to become worn by die table rotation and powder product abrasion. Therefore, periodic inspection of its condition and replacement are recommended. Nevertheless, the scraper blade proper condition is important but also its adjustment since if it is not set up correctly, powder product may accumulate on the die table, leading to problems with weight control.

Tablet hardness variation is intimately related to weight variation and, accordingly, to the influence of compression variables such as dwell time, tablet thickness, and working length of the punches. Thus, to solve a hardness variation, consistency of the tablet weight must be checked first. If the predefined weight is achieved but hardness is out of limits, then precompression and compression forces should be adjusted while keeping tablet thickness within target limits. Although dwell time might be a source of hardness variation, adjustment of this parameter may be detrimental to the whole process since the compression rate is slowed. Occasionally, when the tablet weight target is kept within limits but hardness varies, the problem may be due to the formulation. As mentioned previously, the correct use of punches and dies is of paramount importance and periodic inspection is mandatory in order to ensure the compression process has not been compromised. So, when it becomes hard to achieve tablet hardness, it is recommended to first verify tablet weight and thickness consistency and then try to adjust the precompression. The choice to increase the tablet weight even if it is within limits or to reduce the tablet press speed is not convenient and should be used only when there are no more options.

## REFERENCES


1. Çelik, M. (1994), Compaction of multiparticulate oral dosage forms, in Ghebre-Sellassier, I., Ed., *Multiparticulate Oral Drug Delivery*, Marcel Dekker, New York, pp. 181–216.
2. Parrot, E. (1990), Compression, in Lieberman, H. L., and Schwartz, J. B., Eds., *Pharmaceutical Dosage Form*, 2nd ed., Marcel Dekker, New York, pp. 201–244.
3. Shlieout, G., Wiese, M., and Zessin, G. (1999), A new method to evaluate the consolidation behavior of pharmaceutical materials by using the Fraser-Suzuki function, *Drug. Dev. Ind. Pharm.*, 25(1), 29–36.
4. Mohammed, H., Briscoe, B. J., and Pitt, K. G. (2006), A study on the coherence of compacted binary composites of microcrystalline cellulose and paracetamol, *Eur. J. Pharm. Biopharm.*, 63(1), 19–25.
5. Leuenberger, H., and Leu, R. (1992), Formation of a tablet: A site and bond percolation phenomenon, *J. Pharm. Sci.*, 81(10), 976–982.
6. Khossravi, D., and Morehead, W. T. (1997), Consolidation mechanisms of pharmaceutical solids: A multi-compression cycle approach, *Pharm. Res.*, 14(8), 1039–1045.

7. Masteau, J. C., and Thomas, G. (1999), Modelling to understand porosity and specific surface area changes during tabletting, *Powder Technol.*, 101(3), 240–248.
8. Zuurman, K., Van der Voort Maarschalk, K., and Bolhuis, G. K. (1999), Effect of magnesium stearate on bonding and porosity expansion of tablets produced from materials with different consolidation properties, *Int. J. Pharm.*, 179(1), 107–115.
9. Heckle, R. W. (1961), An analysis of powder compaction phenomena, *Trans. Metall. Soc. AIME*, 221, 1001–1008.
10. Heckle, R. W. (1961), Density-pressure relationship in powder compaction, *Trans. Metall. Soc. AIME*, 221, 671–675.
11. Shivanand, P., and Sprockel, O. L. (1992), Compaction behavior of cellulose polymers, *Powder Technol.*, (69), 177–184.
12. Cooper, A. R., and Eaton, L. E. (1962), Compaction behavior of several ceramic powders, *J. Am. Ceram. Soc.*, 45(3), 97–101.
13. Kawakita, K., and Ludde, K. H. (1970), Some considerations on powder compression equations, *Powder Technol.*, 4, 61–68.
14. Palmieri, G. F., Joiris, E., Bonacucina, G., Cespi, M., and Mercuri, A. (2005), Differences between eccentric and rotary tablet machines in the evaluation of powder densification behaviour, *Int. J. Pharm.*, 298(1), 164–175.
15. Williams, R. O.3rd, and McGinity, J. W. (1989), Compaction properties of microcrystalline cellulose and sodium sulfathiazole in combination with talc or magnesium stearate, *J. Pharm. Sci.*, 78(12), 1025–1034.
16. Muller, F. X., and Augsburger, L. L. (1994), The role of the displacement-time waveform in the determination of Heckel behaviour under dynamic conditions in a compaction simulator and a fully-instrumented rotary tablet machine, *J. Pharm. Pharmacol.*, 46(6), 468–475.
17. Amidon, G. E., and Houghton, M. E. (1995), The effect of moisture on the mechanical and powder flow properties of microcrystalline cellulose, *Pharm. Res.*, 12(6), 923–929.
18. Gupta, A., Peck, G. E., Miller, R. W., and Morris, K. R. (2005), Influence of ambient moisture on the compaction behavior of microcrystalline cellulose powder undergoing uniaxial compression and roller-compaction: A comparative study using near-infrared spectroscopy, *J. Pharm. Sci.*, 94(10), 2301–2313.
19. Gustafsson, C., Lennholm, H., Iversen, T., and Nystrom, C. (2003), Evaluation of surface and bulk characteristics of cellulose I powders in relation to compaction behavior and tablet properties, *Drug. Dev. Ind. Pharm.*, 29(10), 1095–1107.
20. Cunningham, J. C., Sinka, I. C., and Zavaliangos, A. (2004), Analysis of tablet compaction. I. Characterization of mechanical behavior of powder and powder/tooling friction, *J. Pharm. Sci.*, 93(8), 2022–2039.
21. Picker, K. M. (2004), The 3D model: Explaining densification and deformation mechanisms by using 3D parameter plots, *Drug. Dev. Ind. Pharm.*, 30(4), 413–425.
22. Ruegger, C. E., and Celik, M. (2000), The effect of compression and decompression speed on the mechanical strength of compacts, *Pharm. Dev. Technol.*, 5(4), 485–494.
23. Barra, J., Falson-Rieg, F., and Doelker, E. (1999), Influence of the organization of binary mixes on their compactibility, *Pharm. Res.*, 16(9), 1449–1455.
24. *Press Tooling Manual* (2003), Thomas Engineering Inc. Hoffman Estates, IL, USA. pp. 1–38.
25. *Tableting Specification Manual*, 7th ed., American Pharmacists Association, p. 130.
26. *International Organization for Standardization (ISO)*, 18084:2005, 1st ed, ISO, Geneva, pp. 1–13.

27. Train, D. (1956), An investigation into the compaction of powders, *J. Pharm. Pharmacol.*, 8(10), 745–761.
28. Burlinson, H. (1968), *Tablets and Tabletting*, Heinemann, London.
29. Long, W. M. (1960), Radial pressures in powder compaction, *Powder Metall.*, 6, 73–86.
30. Ritter, A., and Sucker, H. B. (1980), Studies of variables that effect tablet capping, *Pharm. Tech.*, (3), 57–65, 128.
31. Hiestand, E. N., Bane, J. M., Jr., and Strzelinski, E. P. (1971), Impact test for hardness of compressed powder compacts, *J. Pharm. Sci.*, 60(5), 758–763.
32. Joe Au, Y. H., Eissa, S., and Jones, B. E. (2004), Receiver operating characteristic analysis for the selection of threshold values for detection of capping in powder compression, *Ultrasonics*, 42(1–9), 149–153.
33. Rodriguez, L., Cini, M., Cavallari, N., Passerini, N., Saettone, M. F., Monti, D., and Caputo, O. (1995), Ultrasound-assisted compaction of pharmaceutical materials, *Farm Vestn.*, (46), 241–242.
34. Armstrong, N. A., and Palfrey, L. P. (1989), The effect of machine speed on the consolidation of four directly compressible tablet diluents, *J. Pharm. Pharmacol.*, 41(3), 149–151.
35. Zografi, G., and Mattocks, A. M. (1963), Adsorption of certified dyes by starch, *J. Pharm. Sci.*, 52(Nov.), 1103–1105.

# EFFECTS OF GRINDING IN PHARMACEUTICAL TABLET PRODUCTION

Gavin Andrews, David Jones, Hui Zhai, Osama Abu Diak, and Gavin Walker
*Queen's University Belfast, Belfast, Northern Ireland*

**Contents**

1 Introduction
2 Milling Equipment
  2.1 Ball Mill
  2.2 Fluid Energy Mill
  2.3 Hammer Mill
  2.4 Cutting Mill
3 Powder Characterization Techniques
  3.1 Powder Sampling
  3.2 Particle Density and Voidage
  3.3 Particle Surface Area
  3.4 Particle Shape
4 Effect of Particle Size Reduction on Tableting Processes
  4.1 Wet Granulation Processes
  4.2 Mixing Processes
  4.3 Flowability of Pharmaceutical Powders
  4.4 Compression Processes
  References

## 1 INTRODUCTION

The importance of size reduction in relation to pharmaceutical active agents and excipients is well known, and the aim of this chapter is to identify methods for particle size reduction, discuss how particle size and shape are characterized, and

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

recognize the importance of controlling particle characteristics to ensure the success of pharmaceutical powder processing and the manufacture of elegant pharmaceutical products. An initial overview of the implications of size reduction within pharmaceutics and the importance of comminution in relation to variability of active pharmaceutical ingredient (API) surface area, efficacy, and ultimately dosing regimen required to maintain optimum therapeutic effects will be addressed. This will encompass examples from a diverse range of dosage forms, including oral, parenteral, and topical systems. The effects of particle size on the essential characteristics of powders intended for compression (tablets, capsules) such as fluidity and compressibility will be addressed. The need for uniformity of size and the effects of particle size distribution on the homogeneity of mixing/blending and in essence on the uniformity of APIs within the final manufactured dosage form will be highlighted.

## 2 MILLING EQUIPMENT

There are many factors that must be taken into consideration in choosing milling equipment. Some of these factors are related to required product specifications such as particle size distribution, but additionally, physical and chemical properties of the material such as particle shape and moisture content must also be taken into consideration. Furthermore, other factors that are related to production requirements (mill capacity and the required production rate) must be carefully balanced to ensure the correct choice of milling equipment.

### 2.1 Ball Mill

A ball mill consists of a hollow cylinder mounted such that it can be rotated on its horizontal longitudinal axis (Figure 1). The length of the ball mill is slightly greater than its diameter. A ball mill reduces particle size by subjecting particles to impact and attrition forces generated by moving steel balls or pebbles (grinding medium) that typically occupy 30–50% of the total volume of the mill. It is common for a ball mill to contain balls of different diameters that aid size reduction. Generally, larger diameter balls have a higher tendency to act upon coarse feed materials

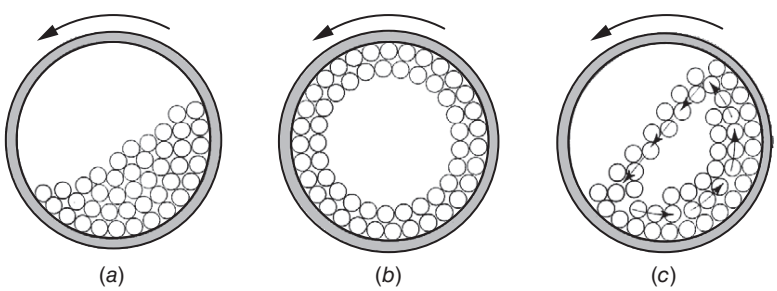


**FIGURE 1** Ball mill in operation showing correct cascade action.

whereas smaller diameter balls facilitate the formation of fine product by reducing void spaces between the balls.

The most important factors governing the performance of the mill and the achievement of the desired particle size are as follows:

1. Amount of material required for subsequent testing (sample volume)
2. Speed of rotation of ball mill

A high volume of powder feed produces a cushioning effect whereas small sample volumes cause a loss of efficiency and abrasive wear of the mill parts. The amount of material to be milled in a ball mill may be expressed as a material-to-void ratio (ratio of the volume of material to that of the void in the ball charge). As the amount of material is increased, the efficiency of a ball mill is increased until the void space in the bulk volume of ball charge is filled; then, the efficiency of milling is decreased by further addition of material.

Rotational speed is the most significant factor controlling the particle size specification. The optimum speed of rotation is dependent on mill diameter. At low angular velocities the balls move with the drum until the force due to gravity exceeds the frictional force of the bed on the drum, and the balls then slide back to the base of the drum. This sequence is repeated, producing very little relative movement of balls so that size reduction is minimal. At high angular velocities the balls are thrown out onto the mill wall by centrifugal force and no size reduction occurs. At about two-thirds of the critical angular velocity where centrifuging occurs, a cascading action is produced. Balls are lifted on the rising side of the drum until their dynamic angle of repose is exceeded. At this point they fall or roll back to the base of the drum in a cascade across the diameter of the mill. By this means, the maximum size reduction occurs by impact of the particles with the balls and by attrition.

The critical speed of a ball mill is the speed at which the balls just begin to centrifuge with the mill. Thus, at the critical speed, the centrifugal force is equal to the weight of the ball. At and above the critical speed, no significant size reduction occurs. The critical speed $n_c$ is given by the equation

$$n_c = \frac{76.6}{\sqrt{D}}$$

where $D$ is the diameter of the mill.

A larger mill reaches its critical speed at a slower revolution rate than a smaller mill. Ball mills are operated at from 60 to 85% of the critical speed. Over this range, the output increases with the speed; however, the lower speeds are for finer grinding. An empiric rule for the optimum speed of a ball mill is

$$n = 57 - 40 \log D$$

where $n$ is the speed in revolutions per minute and $D$ is the inside diameter of the mill in feet.

In practice, the calculated speed should be used initially in the process and modified as required.

The use of a ball mill is advantageous in that it may be used for both wet and dry milling and additionally can be successfully employed in batch and continuous operation. Also, the installation, operation, and labor costs involved in ball milling are extremely low in comparison to other techniques, which makes this technique economically favorable.

## 2.2 Fluid Energy Mill

Fluid energy milling acts by particle impaction and attrition that are generated by a fluid, usually air (Figure 2). Fluid energy mills can reduce the particle size to approximately 1–20 μm. A fluid energy mill consists of a hollow toroid that has a diameter of 20–200 μm, depending on the height of the loop, which may be up to 2 m. Fluid is injected as a high-pressure jet through nozzles at the bottom of the loop with the high-velocity air, giving rise to zones of turbulence into which solid particles are fed. The high kinetic energy of the air causes the particles to impact with other particles with sufficient momentum for fracture to occur. Turbulence ensures that the high levels of particle–particle collision produce substantial size reduction by impact and attrition.

The design of fluid energy mills provides an internal classification system according to their particle size in which the finer and lighter particles are discharged and the heavier, oversized particles, under the effect of centrifugal force, are retained until reduced to a significantly smaller size.

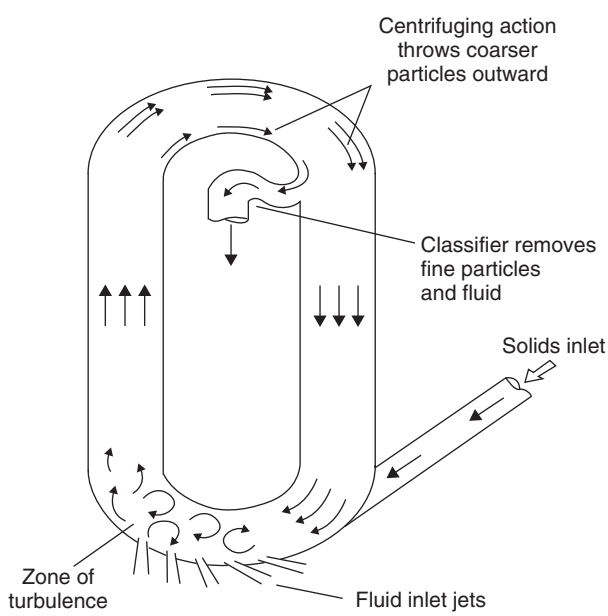


**FIGURE 2** Fluid energy mill.

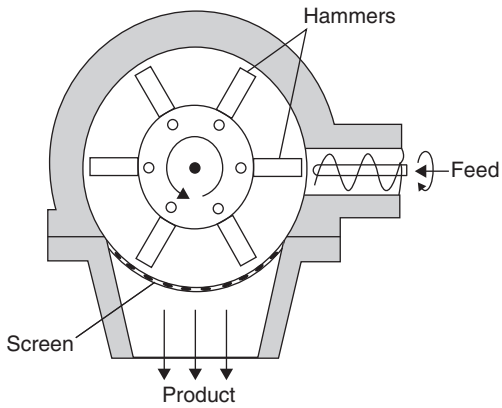


**FIGURE 3** Hammer mill.

## 2.3 Hammer Mill

The main mechanism of size reduction produced by a hammer mill is by impaction that is generated from a series of four or more hammers hinged on a central shaft and enclosed within a rigid metal case (Figure 3). During milling the hammers swing out radially from the rotating central shaft. The angular velocity of the hammers produce strain rates up to $80\,s^{-1}$, which are so high that most particles undergo brittle fracture. As size reduction continues, the inertia of particles hitting the hammers reduces markedly and subsequent fracture is less probable, so that hammer mills tend to produce powders with narrow particle size distributions. Particle retention within the mill is achieved using a screen, which allows only sufficiently milled particles (defined particle size) to pass through. Particles passing through a given mesh can be much finer than the mesh apertures, as particles are carried around the mill by the hammers and approach the mesh tangentially. For this reason, square, rectangular, or herringbone slots are often used. According to the purpose of the operation, the hammers may be square faced or tapered to a cutting edge or have a stepped form.

The particle size achieved may be controlled variation in the speed of the hammers and additionally by careful selection of the size and type of screen. During the operation of a hammer mill the speed of rotation is critical such that below a critical impact speed the rotor turns so slowly that a blending action rather than milling is obtained. Such operating conditions result in significant rises in temperature. Moreover, at very high speeds, there is the probability of insufficient time between successive passes of the hammers for a significant mass of material to fall from the grinding zone.

The hammer mill is particularly useful in achieving particles in the approximate size range of 20–40 μm and additionally in producing a particle size distribution that is extremely narrow. The equipment offers ease of use and high levels of flexibilty (speed and screen may be rapidly changed allowing rapid variation in achievable particle size), is easy to clean, and can be operated as a closed system, thus avoiding operator exposure to potent dusts and potential explosion hazards.

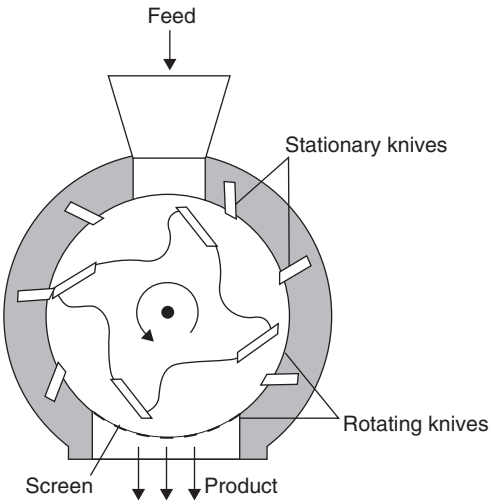


**FIGURE 4** Cutter mill.

### 2.4 Cutting Mill

Particle size reduction using a cutting mill involves successive cutting or shearing a sample using a series of knives attached to a horizontal rotor (Figure 4). This rotary motion pushes the sample against a series of stationary knives that are attached to the mill casing. Size reduction occurs by fracture of particles between the two sets of knives, which have a clearance of approximately a few millimetres. As with a hammer mill a screen is fitted at the base of the mill casing and acts to retain material until a sufficient degree of size reduction has occurred.

## 3 POWDER CHARACTERIZATION TECHNIQUES

### 3.1 Powder Sampling

Powdered materials are used in a wide range of industries, no more so than in the pharmaceutical industry wherein powders are used for the manufacture of a wide range of dosage forms, the two most common being tablets and hard gelatin capsules. Orally administered solid dosage forms are the preferred and most patient convenient, primarily because of the ease of administration and the convenience of handling. Pharmaceutically, orally administered solid dosage forms are generally more favorable because of increased stability in comparison to their liquid counterparts (suspensions, syrups) and the increased control they offer in manipulating drug dissolution in vivo to suit end-use requirements. Solid dosage forms administered via the oral route are an intricate blend of pharmaceutical excipients (diluents,

binders, disintegrants, glidants, lubricants, and flavors) and APIs. In order to successfully manufacture acceptable pharmaceutical products, these materials must be adequately mixed and/or granulated to ensure that the resultant agglomerates possess the required fluidity and compressibility and, in addition, avoid demixing during postgranulation processes. Moreover, the final characteristics of tablets or capsules such as drug dissolution rate, disintegration time, porosity, friability and hardness are significantly influenced by the properties of the powder blends used during their manufacture.

During product manufacture large volumes of powder blends are fed through production equipment/processes, and it is essential to be able to accurately determine, define, and control powder properties to ensure reproducible manufacture and product performance. Therefore the characterization of the physicochemical properties of powder blends is extremely important. It is well accepted that there are inherent difficulties in characterizing the entire mass of a bulk powder blend or process stream, so it is essential to remove and analyze discrete samples.

Sampling is a useful technique that allows an appropriate aliquot to be withdrawn from the bulk so as to collect a manageable amount of powder which is representative of the batch [3], in other words, every particle should have an equal chance of being selected [4]. However, there are many circumstances that may result in the selection of nonrepresentative samples and hence the definition of powder characteristics that are not a true estimation of the entire bulk powder. Typically, powder masses with an extremely wide particle size distribution or diverse physical properties are highly likely to be heterogeneous, which may result in high levels of variability and samples that do not represent bulk mass. Moreover, powder characteristics may change because of the attrition and segregation during transfer that can make sampling extremely difficult.

It is well accepted that two types of sampling errors are possible when removing small masses of powder from bulk [5].

1. Segregation errors, which are due to segregation within the bulk and can be minimized by suitable mixing and the use of a large number of incremental samples to form a larger test sample.
2. Statistical errors, which arise because the quantitative distribution in samples of a given magnitude is not constant but is subject to random fluctuations. Consequently, it is an example of a sampling error that cannot be prevented but can be estimated and indeed reduced by increasing the sample size.

Therefore, sampling procedures are of the greatest importance in order to reduce the effect of nonuniform size segregation and nonrandom homogeneity of a system to achieve statistically meaningful sampling results. Careful attention and faithful observance must be demonstrated and it is extremely important that sampling occurs when the powders are in motion [6] and samples are withdrawn from the whole stream for equal periods of time, rather than part of the stream for all of the time [3].

**TABLE 1 Stationary Bulk Sampling**

| Sampling Devices | Procedure of Sampling | Application and Characteristics |
|---|---|---|
| Low volume powder sampler (Figure 5*a*) | In operation the sampler is inserted into the product to be sampled. At a specific sampling depth the operator pushes down on the T bar, which opens the sampling chamber. When released the spring-loaded T bar will close the sampling chamber. | Used for small quantity of sample powders. The sampler has a sampling chamber volume approximately equal to 2 mL. |
| Pneumatic lance sampler (Figure 5*c*) | A gentle flow of air out of the nozzle allows the probe to move through the powder bed. At the site, the air is slowly reversed to draw up a sample, which is collected against a porous plate at the end of the probe [7]. | Minimizes powder disturbance and therefore is better than a sample thief, but bias still cannot be avoided [8]. |
| Scoop sampler | A single swipe of the scoop completely across the powder bulk collects the sample. Each collection should use opposite directions. | Suitable only for materials that are homogeneous within the limits set by the quantity of material taken by the scoop. It may be used for non-free-flowing or damp materials where instrumental methods are inappropriate [9]. |
| Thief/spear probe sampler (Figure 5*b*) | One or more cavities are stamped in a hollow cylinder enclosed by an outer rotating sleeve. The thief is inserted into sample with the cavities closed, once opened the sample fills the hole. The cavities are closed and the thief is withdrawn. It must be ensured that samples are withdrawn from different locations | Thief samplers belong to two main classes, side sampler (has one or more cavities along the probe) and end sampler (has a single cavity at the end of the probe), which are the most common used for stored non-flowing material [10]. |

There are a number of sampling techniques for particle sampling, which can be classified in many different ways. Here, particle sampling techniques are divided into three parts: stationary bulk sampling (Table 1 and Figure 5), flowing stream sampling (Table 2 and Figure 6), and subsampling (Table 3 and Figure 7). The sampling devices, procedures and application overview of the common used techniques in corresponding fields are shown as follow.

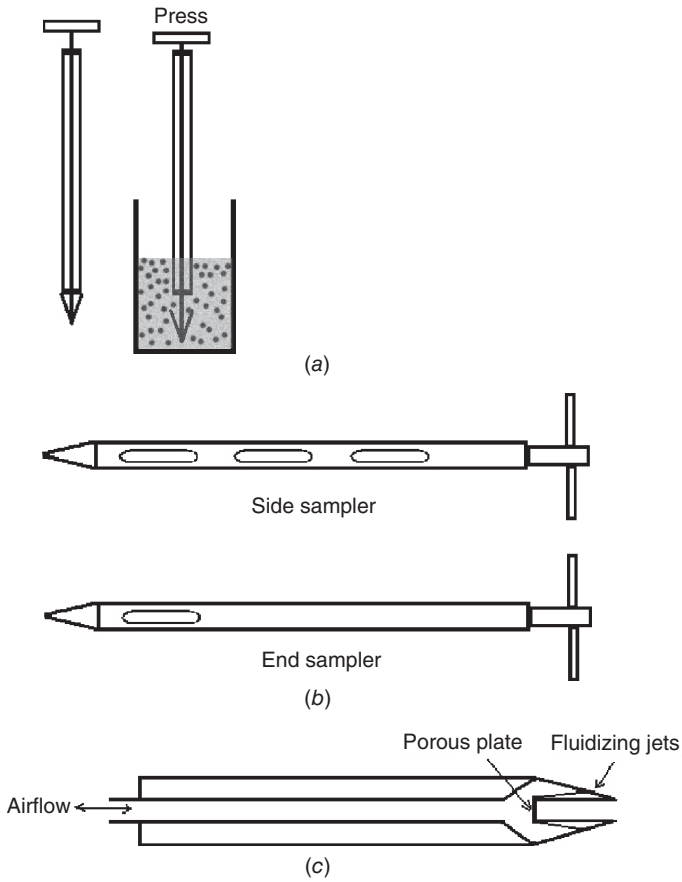


**FIGURE 5** Stationary bulk sampling: (*a*) low-volume powder sampler; (*b*) thief/spear probe sampler; (*c*) pneumatic lance sampler [7].

### 3.3 Particle Density and Voidage

Particle density may be defined as the total mass of the particle divided by its total volume; however, depending upon the different definitions of the total volume (or the different ways to measure the particle volume), there are various definitions of particle density in existence (see Table 4).

In order to get clear understanding of the subtle differences between the definitions of various particle density types, an illustration can be formed as shown in Figure 8.

***Particle Density Methods*** Density is defined as the ratio of mass to volume, so the density determination can be separated into two steps: measurement of mass and measurement of volume. Determining the mass of an object is rather straightforward; however, it is much more difficult to directly determine the volume of a solid. The volume of a solid object with a regular geometric shape may be calculated

**TABLE 2 Flowing Stream Sampling**

| Sampling Devices | Procedure of Sampling | Application and Characteristics |
|---|---|---|
| Auger sampler (Figure 6*a*) [Line sampler for stream] | A pipe with a slot is placed inside the process stream, permitting easy capture of powder through the process stream cross section when rotated. Samples are subsequently then delivered into a separate container by gravitational forces. [6] | While this is often used for stream sampling, it is difficult to collect a representative sample when stream is heterogeneous [10]. |
| Constant-volume sampler (Figure 6*b*) [Point sampler for stream] | Sampling occurs when the stream falls down through a pipe and a constant-volume container is inserted or withdrawn from the stream system. | Designed to extract a constant volume of homogeneous granular material for subsequent chemical analyses and is not suitable for withdrawing samples for physical analyses [11]. |
| Diverter sampler (Figure 6*c*) [Cross-sectional sampler for stream] | The whole stream is diverted by opening a sliding cover or pivoting an external flap in the bottom of a gravity-flow chutes or pipes or screw conveyors [12]. The samples could be removed to a low-angle laser light-scattering instrument then returned to the process stream [6]. | The process could be automated and highly suitable method for online particle size measurement. [7] |
| Full stream sampler (Figure 6*d*) [Cross-sectional sampler for stream] | Samples are withdrawn from conveyors, carried out only on the return stroke. | Extremely useful for dusty materials provided the trough extends the whole length of the stream and does not overfill [6]. |

mathematically; however, in most conditions, the shape of a particle is often irregular, especially in powder technology, which makes it extremely difficult to measure geometrically. Therefore, various methods have been developed to determine the volume of particles and powders. The two most in use in both laboratory and industrial settings are liquid and gas displacement methods. The different values of particle density can also be expressed in a dimensionless form, as "relative density" (or specific gravity), which is the ratio of the density of the particle to the density of water.

The discussion that follows will give an overview of the common methods used in particle density measurement.

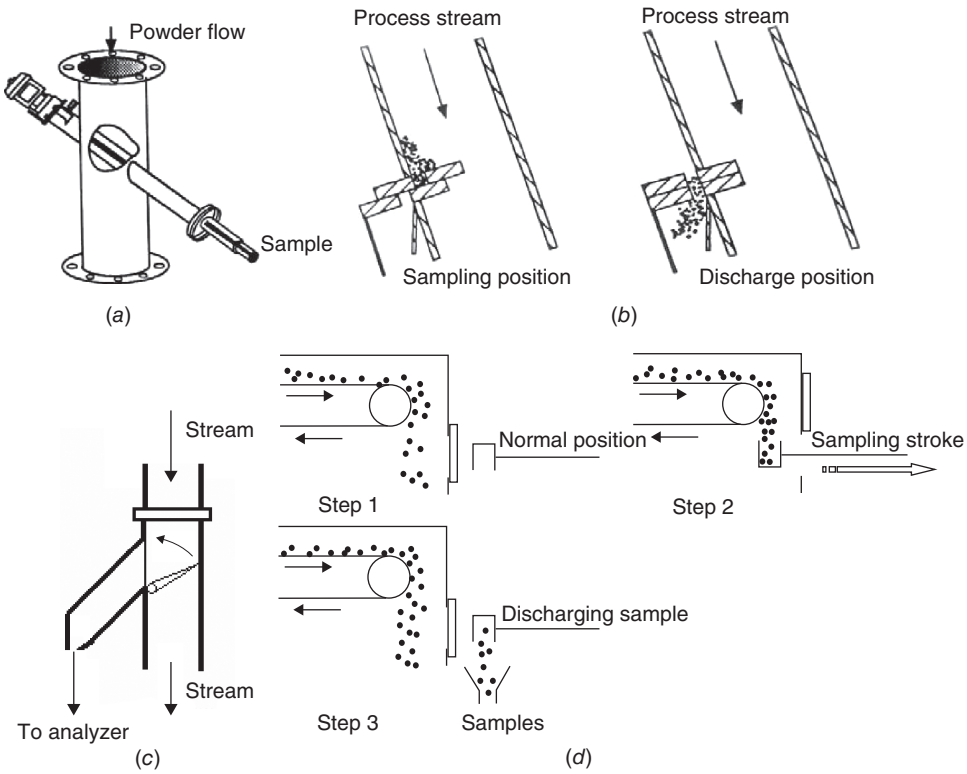


**FIGURE 6** Flowing stream sampling: (*a*) auger sampler [6]; (*b*) constant-volume sampler [6]; (*c*) diverter sampler; (*d*) full-steam sampler.

*Measurement of Particle Density*

1. *Liquid Pycnometry Method* There are several British standards that deal with liquid pycnometry applied to specific materials [18–23]. A pycnometer bottle is weighted empty (M1), and then full of liquid (M2). Following these two initial measurements, two subsequent measurements are made: a sample of powder approximately one-third of maximum container volume (M3) and the bottle filled to capacity containing the sample and water (M4). Great care is required in the final step to ensure that the liquid is fully wetted and all the air removed. Variations in recorded weight also arise depending on how much liquid escapes when the ground glass stopper is inserted in the liquid-filled container. It is extremely important that the liquid used in this procedure does not solubilize or react with the solid particles. Moreover, the solid particles must not absorb the selected fluid.

2. *Gas Pycnometry Method* Principally this method is similar to liquid pycnometry in that volume determination is achieved by detecting the pressure or volume change associated with the displacement of a gas (rather than liquid) by a solid object. Given that this method is largely dependent upon the diffusivity of the gas, helium is often used since it has a low molecular weight and a small atomic radius, allowing high diffusivity into small pores. Sample volumes are often displayed on a

**TABLE 3 Subsampling**

| Sampling Devices | Procedure of Sampling | Application and Characteristics |
|---|---|---|
| Coning and quartering (Figure 7*a*) | A cross-shaped cutter is used to separate the sample heap (which is first flattened at the top) into four equal parts. The segments are drawn apart and two opposite quadrants are combined together. This procedure is repeated at least 4 times until a small enough sample has been generated. | The first choice for non-free-flowing powders and nonflowing powders. Prone to operator bias as fine particles remain in the center of the cone and should never be used with free-flowing powders [13]. |
| Oscillating hopper sample divider (Figure 7*c*) | Hopper (paddle) oscillates and powder falls into two collectors placed under the hopper (paddle). | Used for small quantity of samples. Sample size can be controlled by monitoring time over each collector [7]. |
| Revolving sample splitter (Figure 7*f*) | The revolving feeder distributes the sample material equally (in time) over a number of radial chutes, assuming constant rotational speed [14]. | Very easy to perform and several versions are available that are suitable for free-flowing powders, dusty powders, and cohesive powders. Handling quantities can vary from 40 L to a few grams. |
| Riffle/chute splitter (Figure 7*e*) | The sample is introduced to a rectangular area, divided by parallel chutes leading to two separate receptacles [14]. | Well-accepted method for sample reduction that is highly suitable for free-flowing powders. Used to produce samples with a minimum volume of 5 mL. |
| Spinning riffler (Figure 7*d*) | a steady stream of powder is run into a rotating basket of containers [8]. | Useful in subsampling large samples [15]. Suitable for free-flowing materials [13]. |
| Table sampler (Figure 7*b*) | In a sampling table, powder flows down from the top of an inclined plane, holes and prisms splitting the powder. The powder that reaches the bottom of the plane is the sample. | Used for sample reduction with the advantages of low price and lack of moving parts. |

digital counter on the testing equipment [24]; however, such volumes are easily calculated using the pressure change and the ideal gas law, $PV = nRT$. The true density of the particle can be measured using this method if the particles have no closed pores, while the apparent particle density can be measured if there are any closed pores. Additionally, if open pores are filled with wax, envelope volumes may

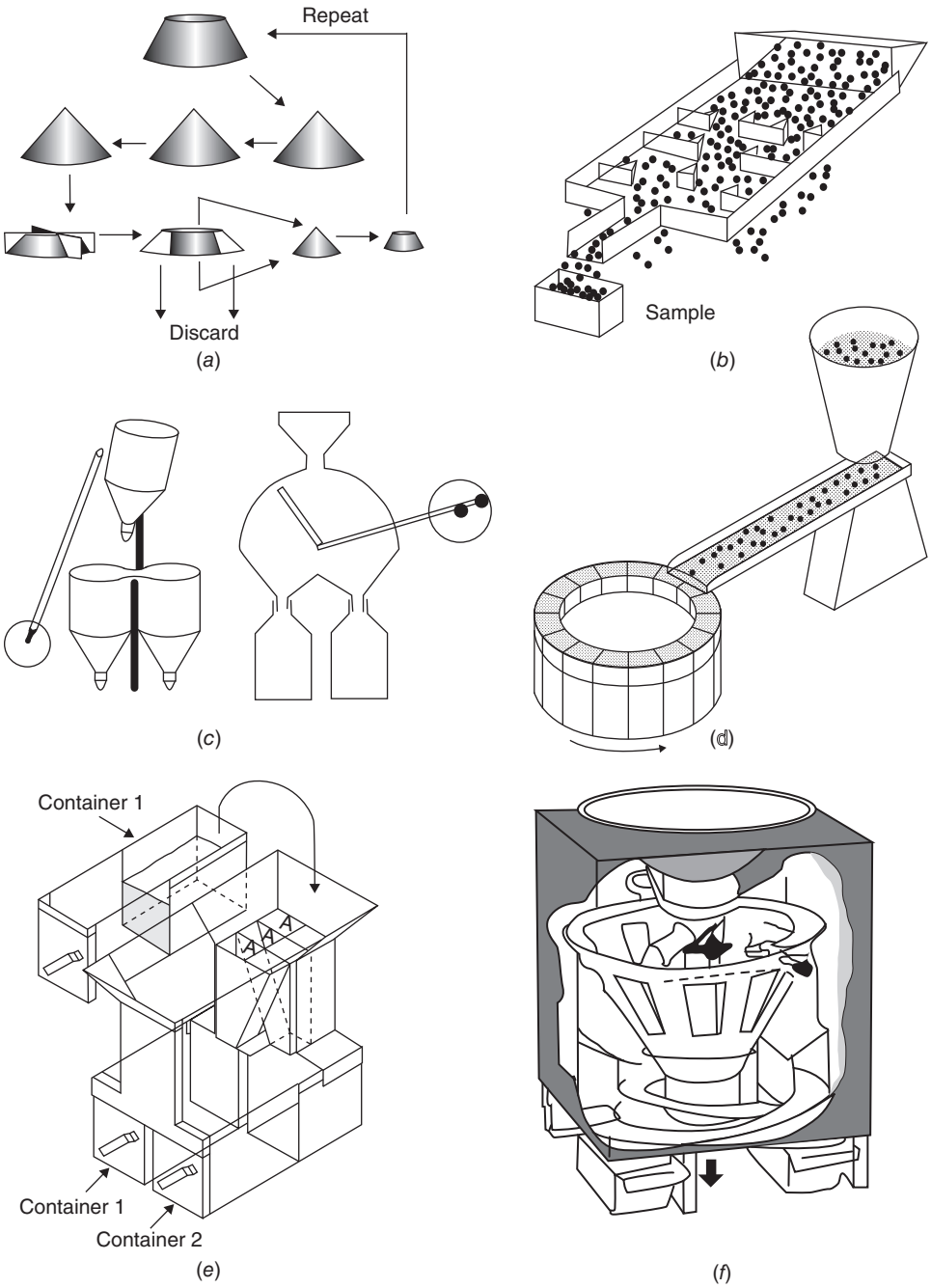


**FIGURE 7** Subsampling: (*a*) coning and quartering; (*b*) table sampler [6]; (*c*) oscillating hopper (paddle) sample divider [6]; (*d*) spinning riffler (BSI); (*e*) riffle/chute splitter (BSI); (*f*) revolving sample splitter [14].

**TABLE 4 Definitions of Density Terms**

| Density Types | Density Definitions | Volumes in Definition | | | |
|---|---|---|---|---|---|
| | | Solid Material Volume | Closed-Pore Volume | Open-Pore Volume | Interparticle Void Volume |
| Absolute powder density | Mass of powder per unit of absolute volume, which is defined as the solid matter after exclusion of all the spaces (pores and voids) (BSI) | ■ | □ | □ | □ |
| Apparent particle density | Mass of particles divided by its apparent particle volume, which is defined as the total volume of the particle, excluding open pores but including closed pores (BSI) | ■ | ■ | □ | □ |
| Apparent powder density | Mass of powder divided by its apparent powder volume, which is defined as the total volume of solid matter, including open pores and closed pores and interstices (BSI) | ■ | ■ | ■ | □ |
| Bulk density | Mass of the particles divided by the volume they occupy, which includes the space between the particles (ASTM) | ■ | ■ | ■ | ■ |
| Effective particle density | Mass of a particle divided by its volume, including open pores and closed pores (BSI) | ■ | ■ | ■ | □ |
| Envelope density | Ratio of the mass of a particle to the sum of the volumes of the solid in each piece and voids within each piece, which is, within close-fitting imaginary envelopes, completely surrounding each piece (ASTM) | ■ | ■ | ■ | □ |
| Skeletal density | Ratio of the mass of discrete pieces of solid material to the sum of the volumes of the solid material in the pieces and closed pores within the pieces (ASTM) | ■ | ■ | □ | □ |
| Tap density | Apparent powder density obtained under stated conditions of tapping (BSI) | ■ | ■ | ■ | ■ |
| True density | Mass of a particle divided by its volume, excluding open pores and closed pores (BSI) | ■ | □ | □ | □ |

*Note:* BSI = British Standards Institute [16], ASTM = American Society for Testing and Material [17]. ■, included; □, excluded.

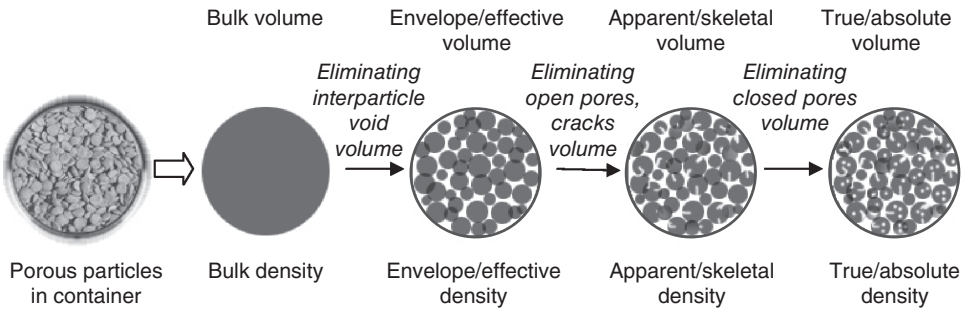


**FIGURE 8** Various density types. The density value increases from bulk density to true density while the volume value decreases from bulk volume to true volume.

be determined and the difference between envelope and apparent volume can yield the open-pore volume, which is sometimes used as a measure of porosity.

3. *Hydrostatic Weighing Method* The volume of a solid sample is determined by comparing the mass of the sample in air with the mass of sample immersed in a liquid with a known density. The volume of sample may be calculated using the difference between the two measured mass values divided by the density of the liquid. This method can be used to determine the bulk or apparent volume. It is extremely important that the suspending liquid does not interact with the powder under investigation.

4. *Float–Sink or Suspension Method* This method involves placing a solid sample into a liquid with known and adjustable density. The density of liquid is incrementally adjusted until the sample begins to sink–float (ASTM C729-75 [25]), or is suspended at neutral density in the liquid (ASTM C693-93 [26]). At the point of equilibrium the density of the sample is equal to the density of the liquid.

5. *Bed Pressure Drop Method* This technique is based on making measurements of bed pressure drop as a function of gas velocity at two voidages, when gas is passed through the bed of powder in the laminar flow regime [24]. During measurement pressure changes for at least four velocities must be measured. The effective particle density $\rho_p$ can be calculated using the equation

$$\frac{s_1}{s_2} = \left(\frac{\rho_{b1}}{\rho_{b2}}\right)\left(\frac{\rho_p - \rho_{b2}}{\rho_p - \rho_{b1}}\right)^3$$

where $s$ is the gradient of pressure drop with gas velocity, $\rho_b$ is the bulk density, $\rho_p$ is the particle effective density.

6. *Sand Displacement Method* The sand displacement method is another useful way of measuring the envelope density of a particle using fine sand as the displacement media. Sand is mixed with a known amount of particles, then the density of the sample particles can be determined from the difference of the bulk density between sand alone and that with samples.

7. *Mercury Porosimetry Method* Mercury is a nonwetting liquid that must be forced to enter a pore by application of external pressure. Consequently it is an extremely useful and convenient liquid for measuring the density of powders and/or particles. This method can measure the apparent and true density of one sample by

applying different pressures. At atmospheric pressure, mercury will resist entering pores smaller than about 6 μm in diameter, but at pressures of approximately 60,000 psi (414 MPa) mercury will be forced to enter pores with diameters as small as 0.003 μm [27].

*Measurement of Bulk Density* Bulk density is very important in determining the size of containers used for handling, shipping, and defining storage conditions for pharmaceutical powders and granules. It is a property that is also pertinent in defining the size of hoppers and receivers for milling equipment and for sizing blending equipment in the scale-up to pilot and to commercial production [28]. The concept of bulk density is the mass of particles divided by the bulk volume, which includes not only the envelope volume of particles but also the spaces between particles, so it should not be confused with particle density [24].

The most convenient method to measure bulk density is to fill the particles into a known volume container (usually cylindrical), level the surface, and weigh the particles in the container. The bulk density is calculated by the mass of the particles divided by the volume that can be read from the scale of the measuring cylinder. In order to minimize experimental errors, the container should be ideally at least 1 L in volume, and the ratio of length and diameter should be about 2:1. Also it is recommended to leave the sample for approximately 10 min to achieve an equilibrium volume (density) value before making readings.

Given that the bulk volume associated with the particle mass is a mixture of air and solid material, the bulk density value is highly dependent on sample history prior to measurement. Calculation of the tapped density can then be achieved by tapping the bulk powder a specified number of times (to overcome cohesive forces and remove entrapped air) to determine the tapped volume of the powder. The tapped and bulk density values can be used to define the flowability and compressibility of a powder using Carr's index and the Hausner ratio.

### 3.4 Particle Surface Area

Surface area is one of the most important characteristics in particle technology. Particles with a different surface area will express different physical properties that will subsequently affect many applications and ultimately final dosage form properties.

Similar to particle density, there are various definitions relating to particle surface area [16]:

1. *Adsorption surface area*: the surface area calculated from an adsorption method.
2. *BET surface area*: the surface area calculated from the Brunauer, Emmett, and Teller theory of multilayer adsorption of a gas on a solid surface.
3. *Calculated surface area*: the surface area of a powder calculated from its particle size distribution.
4. *Effective permeability mass-specific surface*: the effective volume-specific surface divided by the effective solid density, determined by permeametry.

5. *Effective permeability volume-specific surface*: the effective surface area divided by the effective solid volume, determined by permeametry.
6. *Permeability surface area*: the surface area of a powder calculated from the permeability of a powder bed under stated conditions.
7. *Specific surface area* ($S_w$): the surface area of a unit mass of material determined under stated conditions, where $S_w$ is usually expressed in centimeters squared per gram or meters squared per gram and can be used for quality control purposes [28].

***Particle Surface Area Determination Methods*** From the standard definitions of particle surface area, it can be seen that various determination methods are used for surface area measurement, such as adsorption (including Langmuir's equation for monolayer adsorption and the BET equation for multilayer adsorption), particle size distribution, and permeability methods. The different methods are rarely in agreement because the value obtained depends upon the procedures used and also on the assumptions made in the theory relating the surface area to the phenomena measured. The most common methods used for measuring particle surface area are described below.

1. *Gas Adsorption Method* Gas adsorption methods measure the surface area of particles/powders through measurement of the amount of gas adsorbed onto the sample surface. The methods can measure both external and internal surfaces (including open pores in the particles) and can yield physically meaningful average particle sizes with nonporous materials [24]. The amount of gas adsorbed depends upon the nature of the solid (adsorbent) and the pressure at which adsorption takes place. The amount of gas (adsorbate) adsorbed can be found by determining the increase in weight of the solid (gravimetric method) or the amount of gas removed from the system due to adsorption by application of the gas laws (volumetric method [6]). The adsorption used in this method is physical adsorption, which is a relatively weak interaction between samples and gases and therefore can be removed by evacuation.

In this method, a graph of the number of moles of gas adsorbed per gram of solid, at constant temperature, against the equilibrium gas pressure is called an adsorption isotherm. A point must be chosen on this isotherm corresponding to the completion of the adsorbed monolayer in order to calculate $S_w$ [29].

2. *Permeametry Method* This method is based on the fact that the flow rate of a fluid through a bed of particles depends on the pore space, the pressure drop across the bed, the fluid viscosity, dimensional factors such as the area of the bed, and specific surface area ($S_w$). The determination of permeability can be made either under continuous steady-state flow (constant flow rate) or under variable-flow (constant-volume) conditions.

All of the permeability methods are based on the Kozeny–Carman equation, which is used to calculate a surface area of a packed powder bed from its permeability. The Kozeny–Carman equation is expressed as [16]

$$S_k = \sqrt{\left(\frac{A\varepsilon^3 \Delta p}{K(1-\varepsilon)^2 L\eta q}\right)}$$

where $S_k$ = effective permeability volume-specific surface of powder assuming only viscous flow occurs in determination (Kozeny–Carman term)

$A$ = cross-sectional area of bed of powder perpendicular to direction of flow of air

$\varepsilon$ = porosity of bed of powder

$\Delta p$ = pressure difference across bed of powder

$K$ = Kozeny constant

$L$ = linear dimension of bed of powder parallel to direction of flow of air (commonly known as height of powder bed)

$\eta$ = viscosity of air at its temperature at time of determination

$q$ = rate of flow of incompressible fluid through bed of powder

The specific surface area calculated here only involves the walls of the pores of the bed and excludes the pores within the particles. Therefore, the surface area measured in this method can be much smaller than the total surface area measured by gas adsorption methods [24].

3. *Particle Size Distribution Method* The surface area of particles can be determined using particle size and particle shape values. The "equivalent spherical diameter" is used in this technique and many attempts to measure the surface area using this method have led to values that are significantly less than the true value (large deviations arising from inability to define particle shape due to surface irregularities and porosity). Surface area values calculated from particle size distribution methods will in effect establish the lower limit of surface area due to the implicit assumptions of sphericity or other regular geometric shapes and by ignoring the highly irregular nature of real surfaces [30].

Besides the three methods introduced above, there are many other methods of surface area determination: Any surface-dependent phenomenon can be used for such measurement [24]. Some available methods (mercury porosimetry, adsorption from solution, adsorption of dyes, chemisorption, density methods, and secondary ion mass spectroscopy) are explained in more detail elsewhere [6, 30, 31, 32].

### 3.5 Particle Shape

Particle behavior is a function of particle size, density, surface area, and shape. These interact in a complex manner to give the total particle behavior pattern [28]. The shape of a particle is probably the most difficult characteristic to be determined because there is such diversity in relation to particle shape. However, particle shape is a fundamental factor in powder characterization that will influence important properties such as bulk density, permeability, flowability, coatablility, particle packing arrangements, attrition, and cohesion [33–36]. Consequently it is pertinent to the successful manipulation of pharmaceutical powders that an accurate definition of particle shape is obtained prior to powder processing.

A number of methods have been proposed for particle shape analysis, including shape coefficients, shape factors, verbal descriptions, curvature signatures, moment invariants, solid shape descriptors, and mathematical functions (Fourier series

expansions or fractal dimensions); these are beyond the scope of this chapter but have been adequately described in other texts [37].

In the most simplistic means of defining particle shape, measurements may be classified as either macroscopic or microscopic methods. Macroscopic methods typically determine particle shape using shape coefficients or shape factors, which are often calculated from characteristic properties of the particle such as volume, surface area, and mean particle diameter. Microscopic methods define particle texture using fractals or Fourier transforms. Additionally electron microscopy and X-ray diffraction analysis have proved useful for shape analysis of fine particles.

### *Particle Shape Measurement*

1. *Shape Coefficients and Shape Factors* There are various types of shape factors, the majority based on statistical considerations. In essence this translates to the use of shape factors that do refer not to the shape of an individual particle but rather to the average shape of all the particles in a mass of powder. However, a method developed by Hausner [38] that uses three factors—elongation factor, bulkiness factor, and surface factor—may be used to characterize the shape of individual particles (Table 5).

2. *Determining Particle Shape by Fourier Analysis* Fourier transforms have been previously used to determine particle shape and the rollability of individual particles from the coefficients of the resulting series [39]. Moreover, fast Fourier transforms have been successfully used to determine coefficients and a particle "signature" by plotting $\ln An$ versus $\ln n$, where $An$ is the $n$th Fourier coefficient and $n$ is the frequency [29, 40, 41]. In brief, Fourier method consists of finding the centre of gravity of a particle and its perimeter, from which a polar coordinate system is set up. Amplitude spectra of a finite Fourier series in closed form are used as shape descriptors of each particle [42]. Several research papers have focused on the characterization of individual particle shape using Fourier grain analysis or morphological analysis [43–44]. The method has also been extended to the measurement of particle shapes in a blend [45] and to relate particle attrition rate in a milling operation to particle shape [46].

3. *Determining Particle Shape by Electron Microscopy* Electron microscopy has been used for the examination of fine powder dispersions and will provide information on particle shape perpendicular to the viewing direction. Standard shadowing procedures may be useful in obtaining information on shape in the third dimension. Scanning electron microscopy can give direct and valuable information on the shape of large particles [47].

4. *Determining Particle Shape by X-Ray Diffraction Broadening* The broadening of X-ray diffraction lines is primarily a measure of the departure from single-crystal perfection and regularity in a material and can therefore be used to characterize particle shape. This is the only method that gives the size of the primary crystallites, irrespective of how they are aggregated or sintered, and is of great value for determining the properties of fine powders [48, 49].

5. *Other Methods for Particle Shape Determination* Gotoh and Finney [50] proposed a mathematical method for expressing a single, three-dimensional body

**TABLE 5 Shape Coefficients and Shape Factors**

| Coefficients and Factors | Symbols | Definitions and Equations |
|---|---|---|
| Volume shape coefficient | $\alpha_v$ | $\alpha_v = \frac{V}{d^3}$<br>where $V$ = average particle volume<br>$d$ = mean particle diameter |
| Surface shape coefficient | $\alpha_s$ | $\alpha_s = \frac{S}{d^2}$<br>where $S$ = average particle surface<br>$d$ = mean particle diameter |
| Volume–surface shape coefficient | $\alpha_{vs}$ | $\alpha_{vs} = \frac{\alpha_s}{\alpha_v}$<br>where $\alpha_v$ = volume shape coefficient<br>$\alpha_s$ = surface shape coefficient |
| Shape factor | $\alpha_0$ | $\alpha_o = \alpha_v m \sqrt{n}$<br>where $\alpha_o$ = shape factor for equidimensional particle and thus represents part of $\alpha_v$ which is due to geometric shape only<br>$\alpha_v$ = volume shape coefficient<br>$m$ = flakiness ratio, or breadth/thickness<br>$n$ = elongation ratio, or length/breadth |
| Sphericity shape factor | $\Psi_w$ | Sphericity = (surface area of sphere having same volume as particle) / (surface area of actual particle) |
| Circularity shape factor | | Circularity = (perimeter of particle outline)$^2$ / $4\pi$(cross-sectional or projection area of particle outline) |

Source: From refs. 6 and 42.

by sectioning it as an equivalent ellipsoid with the same volume, surface area, and average projected area as the original body. Moreover, wedge-shaped photodetectors to measure forward light-scattering intensity have also been explored for determination of crystal shape [51]. More recently a technique referred to as time of transition (TOT) that was first introduced in 1988 has also been used for the analysis of particle size and shape [52, 53].

## 4 EFFECT OF PARTICLE SIZE REDUCTION ON TABLETING PROCESSES

Particle size plays a critical role in the efficacy of a drug product. It can impact not only bioavailability but also the efficiency and success of production process and ultimately the properties of the final dosage form.

### 4.1 Wet Granulation Processes

The particle size of an active pharmaceutical ingredient can have significant effect on the processing behavior of a formulation, such as granule growth during wet granulation and hence the resulting granule characteristics. The particle size of the starting material can affect the strength and deformability of moist granules and hence their behavior during the wet granulation process.

The effect of particle size on granule growth is a function of several interacting factors, the balance of which largely depends on the nature of the material and the experimental conditions. Differences in granule structure and porosity, resulting from changes in starting material particle size, can also affect other characteristics (e.g., compressibility) of the granulation.

Badawy et al. [57] studied the effect of DPC 963 (a nonnucleoside reverse transcriptase inhibitor) particle size on the granule growth, porosity, and compressibility of granules manufactured by a high-shear wet granulation process. It was found that DPC 963 granule growth in the high-shear granulator and the resulting granule compressibility and porosity were sensitive to relatively small changes in drug substance particle size. Decreasing the particle size resulted in more pronounced granule growth and enhanced the porosity and compressibility of the granulation. Higher pore volume for the granulation manufactured using the active ingredient with a smaller particle size may be the reason for its higher compressibility. The high granulation porosity resulted in an increased fragmentation propensity and volume reduction behavior of the granulation that led to increased compressibility. The more porous granulation has higher tendency to densify upon application of the compression force, resulting in closer packing of the particles.

### 4.2 Mixing Processes

Mixing may be defined as a unit operation that aims to treat two or more components, initially in an unmixed or partially mixed state, so that each unit of the components lies as nearly as possible in contact with a unit of each of the other components [2]. Whenever a product contains more than one component, mixing will be required in the manufacturing process in order to ensure an even distribution of the active component(s).

It is well accepted that mixing solid ingredients is usually more efficient and uniform if the active ingredient and excipients are approximately the same size, which ultimately provides a greater uniformity of dose [1]. Particle size and particle size distribution are important in the powder-mixing process since they largely determine the magnitude of forces, gravitational and inertial, that can cause interparticulate movement relative to surface forces, which resist such motion. As a consequence of high interparticulate forces, as compared with the gravitational forces, powders of less than 100 μm mean particle diameter sizes are not free flowing. Powders that have high cohesive forces due to interaction of their surfaces can be expected to be more resistant to intimate mixing than those whose surfaces do not interact strongly [2].

In moving from one location to another, relative to neighboring particles, a particle must surmount a certain potential energy barrier that arises from forces resisting movement. This effect is a function of both particle size and shape and is most

pronounced when high packing densities occur. Ideal mixing may be achieved when all the particles of the powder mix have similar size, shape, and density characteristics whereas segregation (demixing) may occur when powder blends are not composed of monosized near-spherical particles but contain particles that differ in size, shape, and density. Segregation is more likely to occur if the powder bed is subjected to vibration.

The main reason for segregation in powder blends is the difference in the particle size of the components of the particles contained within the blend. Due to the high diffusivity of small particles, such materials move through the voids between larger particles and so migrate to the lower regions of the powder mix. Moreover, during mixing operations, extremely fine particles have a high tendency to be forced upward by turbulent air currents as the powder blend tumbles and subsequently become isolated from the mixing process through continuous suspension above the blend. When mixing is stopped, these particles will sediment and form a layer on top of the coarser particles.

It is important to control the particle size distribution of pharmaceutical granules or powder blends because a wide size distribution can lead to a situation with a high probability of segregation. If this occurs within the hoppers of tablet machines, nonuniform products may be manufactured due principally to large weight variations. Tablet dies are filled by volume rather than weight, and consequently, the establishment of different regions within a hopper containing granules of different sizes (and hence bulk density) will contain a different mass of granules. This will lead to an unacceptable distribution of the active pharmaceutical content within the batch of finished product, even though the drug is evenly distributed by weight throughout the granules.

### 4.3 Flowability of Pharmaceutical Powders

Due to the relatively small particle size, irregular shape, and unique surface characteristics, many pharmaceutical powders have a high tendency to be extremely cohesive. This high level of cohesion results in "sticky" powders that have poor flowability, commonly resulting in large mass variability within the final product owing to unpredictable and variable filling of tablet dies.

Powders with different particle sizes have different flow and packing properties, which significantly alter the volume of powder expelled from manufacturing equipment during, for example, encapsulation or tablet compression. In order to avoid such problems, the particle sizes of the active pharmaceutical ingredient and other powder excipients should be defined and controlled during formulation so that problems during production are avoided. Most notably, powder flowability is of critical importance in the successful production of acceptable pharmaceutical dosage forms. High levels of flowability within pharmaceutical powders is not just important in the final stages of manufacture but is essential for many industrial processes, particularly mass transport.

Poor or uneven powder flow can result in excess entrapped air within powders, which may induce capping or lamination in specific high-speed tableting equipment. Moreover, uneven powder flow that is a direct result of the presence of excess fines within a powder blend will also promote increased particle–die wall friction, lubrica-

tion problems, and very importantly increased dust contamination hazards to operating personnel.

Although particle size is a significant factor controlling the flowability of pharmaceutical powders or granules, other factors must be considered. The presence of molecular forces between particle/granule surfaces increases the probability for cohesion and adhesion between solid particles. Cohesion may be defined as the attractive forces between like surfaces, such as component particles of a bulk solid, whereas adhesion may be defined as the attractive force between two unlike surfaces, for example, between a particle and a tablet punch. It is extremely important to appreciate that cohesive forces acting between particles in a powder bed are attributed mainly to short-range nonspecific van der Waals forces that are significantly altered as particle size and relative humidity change.

Cohesion and adhesion are phenomena that occur at the surface of a solid and hence particles with an extremely large surface area will have greater attractive forces than those with a smaller surface area. Consequently particle surface area will have a dramatic effect on the flowability of pharmaceutical powders. Typically, fine particles with very high surface-to-mass ratios will be more cohesive than larger particles, which are influenced more by gravitational forces. Particles larger than 250 μm are usually relatively free flowing, but as the size falls below 100 μm, powders become cohesive and flow problems are likely to occur. Powders having a particle size less than 10 μm are usually extremely cohesive and resist flow under gravity.

Although it has been previously stated that particles with similar particle sizes are desirable for pharmaceutical processes, a bulk powder mass with a narrow particle size distribution accompanied with dissimilar particle shapes can produce a bulk mass with inherently different flow properties, owing principally to differences in interparticle contact area.

### 4.4 Compression Processes

In general, the strength of a compressed powder depends on the inherent ability of the powder to reduce in volume during compression and the amount of interparticulate attraction in the final compact. The decrease in compact volume with increasing compression load is attributed normally to particle rearrangement, elastic deformation, plastic deformation, and particle fragmentation. Pharmaceutical materials normally consolidate by more than one of these mechanisms [58, 59]. Unmodified paracetamol crystals exhibit poor compressibility during compaction, resulting in weak and unacceptable tablets with a high tendency to cap [60]. Moreover the incidence of capping and lamination during production, following ejection of tablets from the die, depended on the plastic and elastic behaviors of the excipients used [61]. It has been suggested that materials undergoing plastic deformation, in contrast to elastic deformation, display enhanced bond formation and produce strong tablets.

The effect of particle size on the compression properties of paracetamol oral dosage forms has been previously reported [62]. Heckel analyses plots indicated that the predominant mechanism of compaction of paracetamol was fragmentation with larger particle fractions experiencing more fragmentation than the smaller particles. Furthermore, Heckel analysis also indicated that, for a given applied

pressure, the larger particles of paracetamol produced denser compacts than the smaller particles. The results of elastic–plastic energy ratios indicated that the majority of energy involved during compaction of paracetamol was utilized as elastic energy. This suggested a massive elastic deformation of paracetamol particles under pressure, resulting in weak and capped tablets. It was found that larger particles exhibited less elastic recovery and elastic energy compared to smaller particles. This was attributed to increased fragmentation of larger particles, resulting in increased bonding between particles due to the formation of more new, fresh, and clean particle surfaces.

## REFERENCES


1. Lachman, L., Lieberman, H. A., and Kanig, J. L. (1986), *The Theory and Practice of Industrial Pharmacy Textbook*, 3rd ed., Leo & Febiger, Philadelphia.
2. Aulton, M. E. (2002), *Pharmaceutics: The Science of Dosage Form Design Textbook*, 2nd ed., Churchill Livingstone, London.
3. Svarovsky, L. (1990), Characterization of powders, in Rhodes, M. J., Ed., *Principles of Powder Technology*, Wiley, Chichester, pp. 35–69.
4. Hawkins, A. E. (1990), Characterizing the single particle, in Rhodes, M. J., Ed., *Principles of Powder Technology*, Wiley, Chichester, pp. 9–34.
5. Sommer, K. (1981), Sampling error on particle analysis, *Aufbereit Tech.*, 22(2), 96–105.
6. Allen, T. (1997), *Particle Size Measurement*, 5th ed., Vol. 1, Chapman & Hall, London, pp. 1–62.
7. Venable, H. J., and Wells, J. I. (2002) Powder sampling, *Drug Dev. Ind. Pharm.*, 28(2), 107–117.
8. Kaye, B. H. (1997), *Powder Mixing*, Chapman & Hall, London, pp. 77–95.
9. British Standard, BS 3406-1 (1986), *Methods for Determination of Particle Size Distribution—Part 1: Guide to Powder Sampling*, British Standard Institution, London.
10. Bicking, C. A. (1964), Sampling, in Standen, A. Ed. *Kirk–OthmerEncyclopedia of Chemical Technology*, 2nd ed., Interscience Publishers, New York, pp. 744–762.
11. Hulley, B. J. (1970), Sampling and Sample Conditioning in On-line Fertilizer Analysis, *Chem. Eng.*, 77, 410–413.
12. Clarke, J. R. P., and Carr–Brion, K. G. (1996), Sampling systems for process analysers 2nd ed., The Bath Press, Avon, pp. 148–180.
13. Allen, T. (1964), Sampling and size analysis, *Silic. Ind*. 29(12), 509–515.
14. Petersen, L. (2004), Representative mass reduction in sampling—A critical survey of techniques and hardware, *Chemometr. Intell. Lab. Syst.*, 74, 95–114.
15. Crosby, N. T., and Patel, I. (1995), *General Principles of Good Sampling Practice*, Royal Society of Chemistry, Cambridge.
16. British Standard BS 2955 (1993), *Glossary of Terms Relating to Particle Technology*, British Standards Institution, London.
17. ASTM (1994), *Compilation of ASTM Standard Definitions*, 8th ed., American Society for Testing and Materials, Philadelphia, PA.
18. British Standard BS 812 (1995), *Testing Aggregates, Part 2: Methods of Determination of Density*, British Standards Institution, London.

19. British Standard 1016 (1980), *Methods for the Analysis and Testing of Coal and Coke, Part 13: Test Special to Coke*, British Standards Institution, London.
20. British Standard BS 1377 (1975), *Methods of Tests for Soils for Civil Engineering Purposes*, British Standards Institution, London.
21. British Standard BS 3483 (1974), *Methods for Testing Pigments for Paints, Part B8: Determination of Density Relative to Water at 4°C*, British Standards Institution, London.
22. British Standard BS 4550 (1987), *Methods for Testing Cement, Part 3.2: Density Test*, British Standards Institution, London.
23. British Standard BS 7755 (1998), *Soil Quality, Part 5: Physical Method, Section 5.3: Determination of Particle Density*, British Standards Institution, London.
24. Svarovsky, L. (1987), *Powder Testing Guide, Methods of Measuring the Physical Properties of Bulk Powders*, Published on behalf of the British Materials Handing Board by Elsevier Applied Science, London, pp. 3–33, 79–95.
25. ASTM C729-75 e1 (1995), *Standard Test Method for Density of Glass by the Sink-Float Comparator*, American Society for Testing and Materials, Philadelphia, PA.
26. ASTM C693-93 (1998), *Standard Test Method for Density of Plastics by the Density-Gradient Technique*, American Society for Testing and Materials, Philadelphia, PA.
27. Webb, P. (2001), *Volume and Density Determinations for Particle Technologists*, Micromeritics Instrument Corp, Georgia.
28. Lieberman, H., and Lachman, L. (1981), *Pharmaceutical Dosage Forms Tablets*, Vol. 2, Marcel Dekker, New York, pp. 112–150, 202–222.
29. Beddow, J. K., and Meloy, T. (1980), *Testing and Characterization of Powders and Fine Particles*, Heyden & Son, London, pp. 63–64.
30. Lowell, S., and Shields, J. E. (1984), *Powder Surface Area and Porosity*, 2nd ed., Chapman & Hall, London.
31. Beddow, J. K. (2000), *Particle Characterization in Technology*, Vol. I, CRC Press, Boca Raton, FL, 3–20.
32. Brunauer, S., Deming, L. S., Deming, W. S., and Teller, E. (1940), Adsorption of gases in multimolecular layers, *J. Am. Chem. Soc.*, 62, 1723–1732.
33. Fonner, D. E., Banker, G. S., and Swarbrick, J. (1966), Micromeritics of granular pharmaceutical solids. 1. Physical properties of particles prepared by 5 different granulation methods. *J. Pharm. Sci.*, 55, 181.
34. Ridgway, K., and Rupp, R. (1969), Effect of particle shape on particle properties, *J. Pharm. Pharmacol.*, 21, 30–39.
35. Shotton, E., and Obiorah, B. A. (1975), Effect of physical properties on compression characteristics, *J. Pharm. Sci.*, 64(7), 1213–1216.
36. Heyd, A., and Dhabbar, D., Particle shape effect on caking of coarse granulated antacid suspensions, *Drug Cosmet. Ind.* 125, 42–45.
37. Hawkins, A. E. (1993), *The Shape of Powder-Particle Outline*, Wiley, New York.
38. Hausner, H. H. (1967), *Characterization of the Powder Particle Shape in Particle Size Analysis*, Society for Analytical Chemistry, London.
39. Schwarcz, H. P., and Shane, K. C. (1969), Measurement of particle shape by Fourier analysis, *Sedimentology*, 13(3–4), 213–231.
40. Meloy, T. P. (1969), *Screening*, AIME, Washington, DC.
41. Meloy, T. P. (1977), Fast Fourier transforms applied to shape analysis of particle silhouettes to obtain morphological data, *Powder Technology*, 17, 27–35.

42. Ehrlich, R., and Full, W. E. (1984), Fourier shape analysis—a multivariate pattern recognition approach, in Beddow, J. K., Ed., *Particle Characterization in Technology*, Vol. II, Morphological Analysis, CRC Press, Boca Raton, FL.
43. Meloy, T. P., Clark, N. N., Durney, T. E., and Pitchumani, B. (1985), Measuring the particle shape mix in a powder with the cascadograph, *Chemical Engineering Science*, 40(7), pp. 1077–1084.
44. Alderliesten, M. (1991), Mean particle diameters. part II: standardization of nomenclature. *Part. Part. Syst. Charact.*, 8, 237–241.
45. Fairbridge, C., Ng, S. H., and Palmer, A. D. (1986), Fractal analysis of gas adsorption on syncrude coke, *Fuel*, 65, 1759–1762.
46. Shibata, T., and Yamaguchi, K. (1990), paper presented at the Second World Congress Particle Technology, Sept., Part 1, Kyoto, Japan.
47. Johari, O., and Bhattacharyya, S. (1969), The application of scanning electron microscopy for the characterization of powders, *Power Technol.*, 2, 335.
48. Hillard, J. E., Cohen, J. B., and Paulson, W. M. (1970), Optimum Procedures for determining ultra fine grain sizes, in Burke, J. J., Reed, N. L., and Weiss, V., Eds., *Ultrafine Grain Ceramics*, Syracuse University Press, Syracuse, New York, pp. 73.
49. Oel, H. J. (1969), Crystal growth in ceramic powders, Gray, T. J., and Frechette, V. D., Eds., *Kinetics of Reactions in Ionic Systems*, Plenum, New York, p. 249.
50. Gotoh, K., and Finney, J. L. (1975), Representation of the size and shape of a single particle, *Powder Tech.*, 12, 125–130.
51. Heffel, C., Heitzmann, D., Kramer, H., and Scarlett, B. (1995), paper presented at the 6th European Symp. Particle Size Characterization, Partec 95, Nurenberg, Germany.
52. Karasikov, N., Krauss, M., and Barazani, G. (1988), in Lloyd, P. J., Ed., *Particle Size Analysis*, Wiley, New York.
53. Manohar, B., and Sridhar, B. S. (2001), Size and shape characterization of conventionally and cryogenically ground turmeric (Curcuma domestica) particles, *Powder Technol.*, 120, 292–297.
54. Abrahamsen, A. R., and Geldart, D. (1980), Behaviour of gas-fluidized beds of fine powders, Part I. Homogeneous expansion, *Powder Technol.*, 26, 35–46.
55. Shibata, T., Tsuji, T., Uemaki, O., and Yamaguchi, K. (1994), *Am. Inst. Chem. Eng. Part* 1, 59, 95–100.
56. Realpe, A., and Velázquez, C. (2005), *Powder Technol.*, 169, 108–113.
57. Badawy, S. I., Lee, T. J., and Menning, M. M. (2000), Effect of drug substance particle size on the characteristics of granulation manufactured in a high-shear mixer, *AAPS PharmSciTech*, 1(4).
58. De Boer, A. H., Bolhuis, G. K., and Lerk, C. F. (1978), Bonding characteristics by scanning electron microscopy of powder mixed with magnesium stearate, *Powder Technol.*, 20, 75–82.
59. Duberg, M., and Nystrom, C. (1986), Studies on direct compression of tablets. 17. Porosity pressure curves for the characterization of volume reduction—Mechanisms in powder compression, *Powder Technol.*, 46, 67–75.
60. Krycer, I., Pope, D. G., and Hersey, J. A. (1982), The prediction of paracetamol capping tendencies, *J. Pharm. Pharmacol*, 34, 802–804.
61. Malamataris, S., Bin-Baie, S., and Pilpel, N. (1984), Plasto-elasticity and tableting of paracetamol, Avicel and other powders, *J. Pharm. Pharmacol*. 36, 616–617.
62. Garekani, H. A., Ford, J. L., Rubinstein, M. H., and Rajabi-Siahboomi, A. R. (2001), Effect of compression force, compression speed, and particle size on the compression properties of paracetamol, *Drug Dev. Ind. Pharm.*, 27(9), 935–942.

# ORAL EXTENDED-RELEASE FORMULATIONS

Anette Larsson,[1] Susanna Abrahmsén-Alami,[2] and Anne Juppo[3]
[1]*Chalmers University of Technology, Göteborg, Sweden*
[2]*AstraZeneca R&D Lund, Lund, Sweden*
[3]*University of Helsinki, Helsinki, Finland*

**Contents**

1 Introduction
  1.1 Background
  1.2 Biopharmaceutical Aspects on Oral ER Formulations
  1.3 Influence of Drug Properties
  1.4 Principles for Extended Drug Release
2 Insoluble Matrix Tablets
  2.1 Principles of Formulation and Release Mechanisms
  2.2 Manufacturing of Insoluble Matrix Tablets
3 Membrane-Coated Oral Extended Release
  3.1 Principles of Formulation and Release Mechanisms
  3.2 Manufacturing of Oral Membrane-Coated Systems
4 Hydrophilic Matrix Tablets
  4.1 Principles of Formulation and Release Mechanisms
  4.2 Manufacturing of Hydrophilic Matrix Tablets
5 Comparison and Summary of Different Technologies
6 Other Oral ER Formulations
  References

## 1 INTRODUCTION

### 1.1 Background

In order to achieve therapeutic effect, a drug needs to reach the right place in the body at the right time. For some drugs, this may be achieved by simple solutions or solid dosage forms with an instant drug release while, for others, one has to modify

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

the drug release. To understand the literature within the area of modified drug release, it is important to be aware of the standard terms used for dosage forms within this field. Malinowski and Marroum have summarized these terms in the book *Encyclopedia of Controlled Drug Delivery* [1]. The authors state that modified-release (MR) formulations refer to "dosage forms for which the drug release characteristics of time course and/or location are chosen to accomplish therapeutic or convenience objectives not offered by conventional dosage forms." One group of MR formulations is the delayed-release dosage form, which does not release the drug immediately after administration. One example of delayed-release formulation is the enteric-coated formulations. Another subgroup of MR formulations is the extended-release (ER) dosage forms, which are the focus of the present chapter. According to a definition from the U.S. Pharmacopeia (USP), ER formulations can be referred to as dosage forms that allow at least a twofold reduction in the dosing frequency compared to conventional dosage forms [2, 3].

The interest in oral ER formulations has dramatically increased in recent years. This can be seen in Figure 1, where the bars in the diagram correspond to the number of publications found in a search in the database SciFinder Scholar 2006 [4] that include the words *oral extended release*. This increase in publications confirms that there are many ongoing activities in this field. The expression oral extended release occurs in the database for the first time in 1954, when Yamanaka et al. utilized the slow dissolution rate of various salts of a drug (pyrimidine penicillin) to extend the period of time when the drug had a clinical effect [5]. In 1959 Robinson and Suedres made a formulation of sulfamethylthiadiazole together with hydrogenated castor oil, which was suspended in an aqueous vehicle, creating a formulation with extended drug release [6]. Later, in the late 1950s and early 1960s Sjögren and Fryklöf compressed active substances (e.g., pentobarbitone sodium and theophyl-

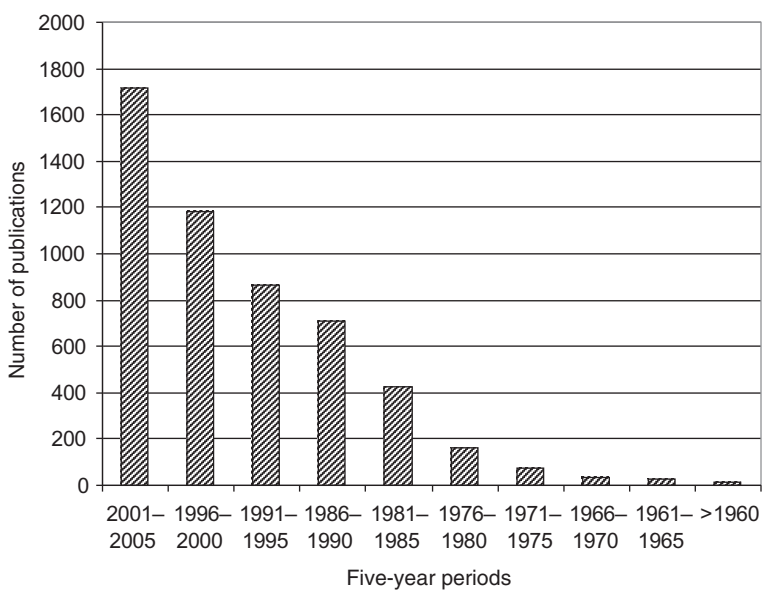


**FIGURE 1** Number of publications containing the words "oral extended release" identified in the database SciFinder Scholar 2006 for the five-year period 1960–2005 [4].

line) together with polyvinyl chloride (PVC) and obtained extended drug release from these insoluble matrix tablets [7]. In an early publication they showed that increased dose loadings or addition of channeling agents increased the drug release rate. It is also interesting to note that the most popular ways to prepare oral ER formulations today were already mentioned in an early review from the early 1970s [8].

## 1.2 Biopharmaceutical Aspects on Oral ER Formulations

The clinical effect of low-molecular-weight substances is often related to the concentration of the drug in the blood plasma. Classical blood plasma profiles for both immediate-release (IR) and ER formulations are shown in Figure 2. It is well known that a drug only has a clinical effect when the concentration in the blood plasma is above the minimum effective concentration (MEC). If the concentration of the active substance is above the maximum safe concentration (MSC), the side effects will be unacceptable. The interval between the MEC and MSC is called the therapeutic window or therapeutic range, and the time when the concentration is above the MEC is called the "duration" of the drug. One aim of ER formulations is to increase the time the substance is above its MEC by continuous release of the drug from the formulation. Under optimal conditions the rate-limiting step in the drug absorption process of an ER formulation is its release rate, which then can be directly related to the concentration of the drug in the blood plasma. When the drug release rate from an ER formulation is constant, the blood plasma concentration will be constant under ideal conditions, whereas ER formulations with time-

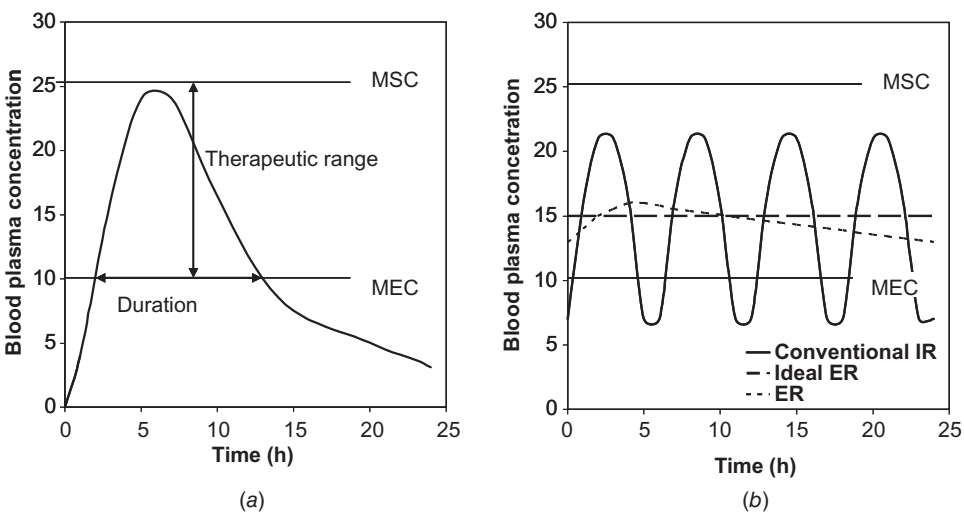


**FIGURE 2** (*a*) Schematic picture of blood plasma concentration profile after administration of a drug to an individual, including the MSC, MEC, therapeutic range, and duration. (*b*) Repeated administration of IR formulation (four times daily) of a drug with short pharmacokinetic half-time and administration once daily of an ideal ER formulation with constant drug release (broken line) or ER formulation with nonconstant drug release rate (dotted line).

dependent drug release rate may give rise to time variations in the concentration of the drug in the blood plasma (Figure 2).

The ER concept might offer several advantages, such as reduction in administration frequency, reduction of side effects, less irritation in the gastrointestinal tract, and improved patient compliance. Speers and Bonnano have also mentioned some economic aspects of ER formulations, such as the possibility to patent line extensions and to reduce manufacturing costs since fewer units are required to obtain the same effect [9]. On the other hand, ER formulations may have several drawbacks, for example, large variations in effect between patients due to varying physiological factors within the patient group, limited transit time for the ER formulation, drug stability problems during the gastrointestinal passage, and more severe complications such as dose dumping.

In order to obtain a clinical effect by an orally administered drug, it is, for example, required that the drug is (i) dissolved and released from its formulation, (ii) transported over the mucosal barrier, and (iii) has passed from the lumen to the systemic blood circulation without being metabolized by, for example, the lumen or the liver. The drug dissolution rate and the rate of absorption of the dissolved active substance as well as the relation between these processes are important, in particular the dissolution process since the absorption of undissolved drug particles can be disregarded.

The ER formulations can be a single-unit, monolithic system or multiple-unit systems containing many individual units with extended release. Multiple-unit systems consist of many small pellets and are normally produced by extrusion and spheronization or coating on inert spheres [2, 10]. The composition and ER mechanism can vary for multiple-unit systems, and some examples are membrane-coated reservoir systems and polymer- or lipid-based matrix systems, where the matrices can be made of both soluble and insoluble carriers [11–16].

From a biopharmaceutical point of view, the multiple-unit systems have many advantages, for example, a more consistent gastrointestinal transit compared to larger monolithic systems [17]. The gastrointestinal transit times for monolithic and multiple-unit systems were compared in a study by Abrahamsson et al. [17]. It was found that the gastric emptying time for the small multiple units was considerably shorter than that of larger monolithic systems (on average 3.6 and 9.6 h, respectively). The transit times through the small intestine were approximately equal, whereas the transit time in the colon for the multiple units was longer compared to the monolithic system, which was explained by different influence of the motility on the different systems. Another advantage with multiple-unit compared to monolithic systems is that the effects of dose dumping become less severe [2]. A breakage and instant drug release from one pellet will have considerably lower effect than breakage of one monolithic system.

There may also be development and manufacturing advantages and disadvantages with multiple-unit systems compared to monolithic systems. The dose for multiple-unit systems may be easier to adjust since one can readily increase the number of pellets in the formulation and thus increase the dose. One can adjust the drug release profiles for multiple-unit systems by mixing pellets with different release profiles. The multiple-unit systems offer the possibility to mix pellets containing different active substances, which can be an advantage for the patients who then only need to take one formulation at a time (containing more than one active

substance) instead of several formulations. Disadvantages with multiple-unit systems may also exist; for example, the time to develop the multiple-unit systems may well be longer than for monolithic systems. There are greater challenges in the scaleup procedure for the multiple-unit systems since several expensive and specialized types of equipment may be needed [18]. For film-coated formulations, additional changes in the drug release rate may be obtained upon storage as a consequence of aging of the film. However, this change can be suppressed by introducing a curing step in the production [10].

### 1.3 Influence of Drug Properties

Drug properties that are important to consider during development of IR tablets are metabolism, stability, permeability, and solubility [18, 19]. In the development of ER formulations, these aspects are also important, but in addition to IR formulations, they must be considered in relation to the different environments that the ER formulations meet during their passage through the gastrointestinal tract, and some of these aspects will be discussed briefly below. It can be mentioned that, based on these initial properties for drug candidates, Thrombre has constructed a feasibility assessment flow chart for ER formation development [18].

The pharmacokinetic half-life for a drug may give an indication of whether a conventional or ER formulation is to be chosen. For drugs with short biological $t_{50}$ (less than 1–2 h), devices that continuously release the active substance are required [2, 18]. For drugs with lifetimes longer than about 10 h, ER formulations may not add any benefits compared to IR formulations. However, for drugs with half-lives between these limits, ER formulations may be a good alternative to IR formulations.

The stability of a drug in the solid state or in aqueous solution is a critical parameter when selecting an appropriate manufacturing process. A drug in an oral ER formulation reaches aqueous environments with, for example, variations in pH (1–8), ionic strength, and bile salt concentration, which requires high chemical stability of the drug [19]. Furthermore, the substance should be stable not only against chemical degradation such as hydrolysis but also against enzymatic degradation (metabolism) during the passage from the lumen to the systemic blood circulation. The drug is released from oral ER formulations along the whole gastrointestinal tract. This implies that, in contrast to IR formulations, drug permeability must be good along the whole gastrointestinal tract for drugs in ER formulations [19]. Furthermore, the solubility and the dissolution rate of the drug are extremely important to consider, since these factors will directly influence the release rate for the drugs from ER formulations. The dissolution rate can be described as the flux $J$ of dissolved material from a drug particle and, according to the Noyes–Whitney equation, it is [20, 21]:

$$J = \frac{D}{h}(C_s - C_b) \tag{1a}$$

where $D$ is the drug diffusion coefficient, $h$ is the thickness of the stagnant diffusion layer around the particles, and $C_s$ and $C_b$ are the concentrations of the drug at the particle surface and in the surrounding bulk media, respectively. For substances in

their most stable solid-state form, the concentration $C_s$ close to the particle surface is equal to the saturated concentration. However, when a substance is in a different polymorphic state or in an amorphous state, $C_s$ can be larger than the saturation concentration.

Many drugs are weak acids or bases with one or several p$K_a$ values. According to the Henderson–Hasselbalch equation, the solubility of an acidic drug will depend on the p$K_a$ and pH as

$$\mathrm{pH} = \mathrm{p}K_a + \log \frac{S_i}{S_0} \tag{2}$$

where $S_i$ and $S_0$ are the concentrations of the drug's un-ionized and ionized forms, respectively. Since the pH varies along the gastrointestinal tract, the solubility and the dissolution rate [Equation (1a)] of the drug depend on the position of the drug in the gastrointestinal tract. Furthermore, for some drugs, such as indomethacin, the dissolution of the acid drug may lead to a changed microenvironmental pH within the stagnant layer and thus also influence the dissolution rate [22, 23].

## 1.4 Principles for Extended Drug Release

The main principles related to ER systems are as follows:

(i) Insoluble matrix formulations
(ii) Membrane-coated solid dosage forms including osmotic pump systems
(iii) Soluble hydrophilic matrix formulations

Below we will discuss each of these formulation principles in terms of basic release mechanisms and the advantages or drawbacks associated with the different formulations and manufacturing processes. However, drug release from all kinds of ER formulations starts with hydration of the formulation and water diffusion into the system. The presence of water in the formulation facilitates the start of the dissolution process of the drug, whereby the dissolved drug can be released from the formulations.

The driving forces for transport of water and drug are the differences in chemical potentials between the formulation and its surrounding. Due to the similarities in the driving forces for the dissolution of a drug and the release from an ER formulation, one can modify Equation (1a) to

$$\frac{dM_t}{dt} = JA = \frac{DA}{h}(C_s - C_b) \tag{1b}$$

where $M_t$ is the amount of active drug that is released at the time $t$, $D$ the drug diffusion coefficient, and $J$ the flux of the drug from the formulation. The other parameters in Equation (1a) have been adjusted to fit the drug release from the formulation and therefore $A$ becomes the surface area of the releasing system in Equation (1b) (e.g., the area of the membrane-coated tablet), $h$ the diffusion pathlength, and $C_s$ and $C_b$ correspond to the concentration of the dissolved active drug at the surface

of the drug particles/formulation and in the bulk solution surrounding the ER device, respectively. Depending on the exact type of ER system, some modifications in Equation (1b) may be needed to fully describe the drug release. One important factor in the equation is the difference in concentration ($C_s - C_b$), and $C_b$ is often assumed to be zero due to release under so-called sink conditions. For active substances in their most stable solid state, the remaining concentration, $C_s$, is equal to the saturation concentration in that medium. However, as mentioned above, $C_s$ can be oversaturated or depend on pH. This means that the drug release from formulations depends on the solid-state properties and $pK_a$ of the drug as well as the pH to which the formulation is exposed [24]. Since the pH varies along the gastrointestinal tract, the drug release will be dependent on the position of the formulation in this tract. Several attempts to avoid pH-independent drug release has been made, for example, by including buffers [25–27].

## 2 INSOLUBLE MATRIX TABLETS

### 2.1 Principles of Formulation and Release Mechanisms

The history of insoluble matrix tablets goes back to the beginning of the 1960s, when Hässle and Abbott developed the Duretter and the Gradumet, respectively [2]. Since then, many ER tablets based on this principle have been developed. Looking at the homepage of the U.S. Food and Drug Administration (FDA [28]) and searching for "insoluble matrix tablets" produces more than 140 hits, which indicates that this research area is still active.

The term *insoluble matrix tablet* refers to tablets in which the drug is embedded in an inert carrier that does not dissolve in the gastrointestinal fluids. The carrier material in insoluble matrix tablets can be based on insoluble lipids or polymers, both matrix builders whose function it is to keep the matrix together during the passage through the gastrointestinal tract and thus prolong the diffusion path of the drug before it is released from the formulation. The drug can be dispersed or dissolved or both in a matrix carrier (see Figure 3) and, depending on the formulation, different mechanisms can be regarded to take place:

- Dissolved drug in the matrix diffuses through the matrix.
- Dissolved drug in the matrix diffuses through pores in the matrix.
- Dispersed drug dissolves and diffuses through the matrix.
- Dispersed drug dissolves and diffuses through pores in the matrix.

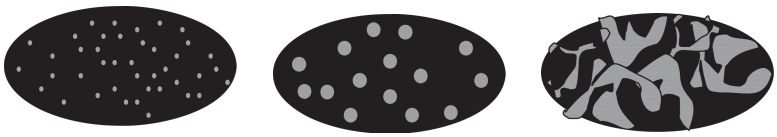

**FIGURE 3** Schematic pictures of insoluble matrix systems. Left: Drug (light gray) molecularly dissolved in carrier material (black); middle: drug particles dispersed in carrier material; right: drug particles dispersed in carrier material at higher drug loading, leading to continuous network of drug.

As early as 1963, Higuchi [29] derived an expression for drug release from insoluble matrix systems. In this historical paper, Higuchi derived two equations for two different geometries, the simple planar sheet matrix system with infinite area and spherical pellets. Furthermore, two special cases were treated, one where the matrix is a homogeneous matrix without pores and another where the matrix contains pores. In the system without pores, the drug is assumed to diffuse through the homogeneous matrix. For matrix systems with pores the efficiency in transport through liquid-filled pores is greater than through the solid matrix carrier. Therefore, the main contribution to the drug release is transport in the pores. This is gained by penetration of the surrounding medium into the pores, where it dissolves the drug. The dissolved drug can diffuse through the pores and be released at the surface of the matrix. The simplest theoretical treatment of the drug dissolved in the matrix carriers assumes the following [30]:

- There is no breakage of the matrix.
- There is no dissolution of the matrix.
- There is no resistance to drug transport in the boundary layer surrounding the device.
- There is no accumulation (e.g., adsorption) of the drug in the device.
- The saturated concentration $C_{s,m}$ of the drug in the matrix is constant during the process.
- The drug-loading concentration $C_0$ is larger than $C_{s,m}$.
- The drug concentration around the matrix is zero, $C_b = 0$ (sink conditions).
- The diffusion constant $D_m$ in the matrix is independent of the drug concentration.
- The partition coefficient $K$ between the matrix material and the surrounding release medium is independent of the drug concentration.

A handy derivation of the equation describing the release from planar homogeneous matrix systems can be found in a book by Wu [30]. It is derived for a sheet with the area $A$ and assumes that the concentration of dissolved drug inside the matrix is constant and equal to the saturation concentration $C_{s,m}$. Under these assumptions, the amount of drug, $M_t$, that is released at time $t$ can be predicted as

$$M_t = A[D_m K C_{s,m} t(2C_0 - C_{s,m})]^{0.5} \tag{3}$$

When the matrix contains a drug-filled network (Figure 3, right image), water can diffuse and dissolve the drug, and this creates a pore structure. The equation describing the drug release from matrices with networks is modified to include information about the created pore structure [Equation (4)]. This can be described by the porosity ε (the volume of the pores in proportion to the total volume of the device) and the tortuosity τ of the pores (a measure of how much the diffusion path is lengthened due to lateral excursions). Also the diffusion coefficient $D_m$ and $C_{s,m}$ in Equation (3) are replaced with $D$ and $C_s$ in Equation (4), corresponding to the diffusion coefficient and the solubility of the drug in the solution inside the pores, respectively:

$$M_t = A\left[\frac{D\varepsilon C_s t}{\tau}(2C_0 - \varepsilon C_s)\right]^{0.5} \tag{4}$$

The most common types of insoluble matrix tablets are those containing pores. From the equations above one can see that the drug-release depends on the solubility of the drug, the drug-loading concentration, and the diffusion coefficient, which is related to the molecular size of the drug. The area of the insoluble matrix tablet also affects the drug release and can be changed by altering the dimensions or the geometry of the tablet. The drug release from insoluble matrix tablets also depends on the porosity and pore structure of the tablet, and the drug release rate increases with increasing porosity.

A comparison of Equations (3) and (4) shows that, in both equations, the amounts of released drug are directly dependent on the area of the device, the square root of the time $t$, the drug-loading concentration $C_0$, the respective saturated drug concentrations, and the drug diffusion coefficients. In addition, the release rate (the time derivate of the amount of released material) depends on the square root of time and can be stated as

$$\frac{dM_t}{dt} \propto \frac{1}{\sqrt{t}} \tag{5}$$

As pointed out above, for ideal ER formulations, the rate-limiting step for drug absorption is the release rate from the ER formulation. Thus, since the release rate from an insoluble matrix system depends on time, the concentration of drug in the blood plasma will also be time dependent and not constant (Figure 2), which may be a therapeutic drawback. Another factor influencing the concentration of drug in the blood plasma is the gastrointestinal transit times. When the transit times of the formulations vary, the reproducibility between different administration occasions in one patient will be low, and furthermore, great variation in the patient group may be obtained. However, these conclusions are valid for all ER formulations based on matrix systems and not limited to insoluble matrix systems only.

The equations above are valid when no depletion of drug occurs inside the device. The equations for release rate will be much more complex when depletion of the drug can occur [30]. However, it has been shown that, when the amount of released material is less than approximately 60%, the release rate will depend on time as $t^{-0.5}$ [29, 30].

### 2.2 Manufacturing of Insoluble Matrix Tablets

Insoluble matrix tablets need a carrier, which can be a lipid- or polymer-based excipient [7, 31–36]. Some suggestions of carrier materials can be found in Table 1. The table also presents the number of hits found upon searching the FDA's homepage [37] for the number of times an excipient is registered as a component in oral extended, sustained, or controlled formulations. This list gives an indication of how often these excipients are commercially used in oral ER formulations but does not automatically tell us the exact formulation or exact mechanistic effect of the excipient. The choice of carrier material is important, and one should be aware of possible

**TABLE 1 FDA Registered Oral ER Formulations Containing Commonly Used Excipients in Insoluble Matrix Formulations**

| Excipient | Number of Hits on FDA Homepage | Content Interval (mg) |
|---|---|---|
| Lipid based | | |
| Carnauba wax | 9 | 46–300 |
| Stearyl alcohol | 4 | 25–244 |
| Glyceryl behenate | 3 | 15–51 |
| Castor oil | 2 | 23 |
| Cottonseed oil, hydrogenated | 2 | 58–402 |
| Cetyl alcohol | 2 | 44–59 |
| Paraffin | 2 | 50–150 |
| Stearic acid | 2 | 26–180 |
| Castor oil, hydrogenated | 1 | 295–410 |
| Vegetable oil, hydrogenated | 1 | 228 |
| Mineral oil | 1 | |
| Microcrystalline wax | 0 | |
| Insoluble polymer | | |
| Ethylcellulose | 9 | 15–309 |
| Ammonia methacrylate copolymer | 5 | 37–138 |
| Polyvinyl acetate | 1 | 46 |
| Polyethylene | 0 | |
| Inorganic | | |
| Calcium phosphate (dibasic) | 6 | 33–335 |

*Source*: http://www.accessdata.fda.gov/scripts/cder/iig/index.cfm.

exposure of lipid-based formulations to erosion, which can be the result of enzymatic degradation of the lipids [38]. This will of course also influence the drug release rate.

In the compositions of insoluble matrix systems, excipients other than the carrier material are needed to obtain products with processability and that meet requirements from the pharmacopedias. Examples of categories of excipients included in insoluble matrix tablets are binders, lubricants, glidants, colorant, taste maskers, and channeling agents. As mentioned above, the drug release rate can be regulated by the porosity in the insoluble matrix system. The properties and amounts of drugs and excipients that can create pores will have a large impact on the release rate. Examples of channeling agents are sugars, salts such as sodium chloride, and polyols [2]. The pore structure also depends on other factors such as the particle sizes of the excipients and the drug, the size and porosity of the granules, and the compaction pressure.

The choice of process steps depends on the properties of the drug and the chosen excipients. For insoluble matrix tablets one often mixes the active substance with the excipients. Either this mixture can be directly compressed to matrix tablets or the powder mixture can be exposed to a granulation technology to enlarge the particle sizes. One such technique is dry granulation, that is, compaction of the mixture in, for example, a roller compactor, followed by milling to desirable granule sizes. Another granulation technique is melt granulation, where the melted granulation liquid agglomerates the particles to granules. The most common granulation

method for insoluble matrix tablets is probably wet granulation [2, 7] with aqueous-based or organic granulation liquids. The wet granulated masses are dried in fluid bed driers or ovens. In order to increase the drying speed, microwaves can be used. The powder mixtures or granules are compressed in ordinary tableting machines [7]. However, it is an advantage if the compaction pressures can be carefully monitored, since this pressure may influence the porosity and thus the drug release. The final tablet can be coated to, for example, mask the taste.

Alternative production methods to the traditional compaction of powder to insoluble matrix tablets are available, some of which are based on melting technologies, but of course these methods rely on the ability of the carrier materials or additives to melt. The drug is commonly dissolved or dispersed in the melt. This melt can be filled into hard gelatin capsules [31] or it can be spray chilled by pressing it through a nozzle into a vessel containing solid carbon dioxide [39]. Hot-melt extrusion of polymer-based systems to form multiple-unit systems has been investigated. The carrier material in these cases can be, for example, Eudragit [40–42].

## 3 MEMBRANE-COATED ORAL EXTENDED RELEASE

### 3.1 Principles of Formulation and Release Mechanisms

One way to protect the drug from being directly released is to coat the system with an insoluble film. The drug is suspended or dissolved in a reservoir system which can consist of monolithic or multiple-unit systems. The MR films surrounding the reservoir will be insoluble and thus give the system extended drug release properties or they can become soluble by external trigging. The latter case is defined as a delayed-release formulation, which means that the formulation does not release directly after administration. This can be achieved, for example, by enteric coatings, where the film-forming materials are insoluble in aqueous solutions at low pH but soluble at high pH values. This results in delayed release from an enteric-coated formulation, since the pH is low in the stomach. When the units are transported into the intestine, the pH increases, the film dissolves, and the drug can be released immediately. This type of formulation will not be further discussed here, but the work of Hogan is recommended for further information on these systems [43, 44].

In ER reservoir systems, a membrane surrounds a reservoir of the drug, also called the core of the system. The membrane controls the drug release and the driving force is the difference in chemical potential over the membrane, which can be correlated with a concentration gradient over the membrane. The transport of the drug through the ER membranes can be divided into three different mechanisms [45–47]:

- Diffusion through the membrane
- Diffusion through pores and cracks in the membrane
- Osmotic transport through pores, cracks, or drilled holes

The overall drug release process for common membrane-coated systems has been shown to pass through three different time periods (Figure 4) [48]. During the initial

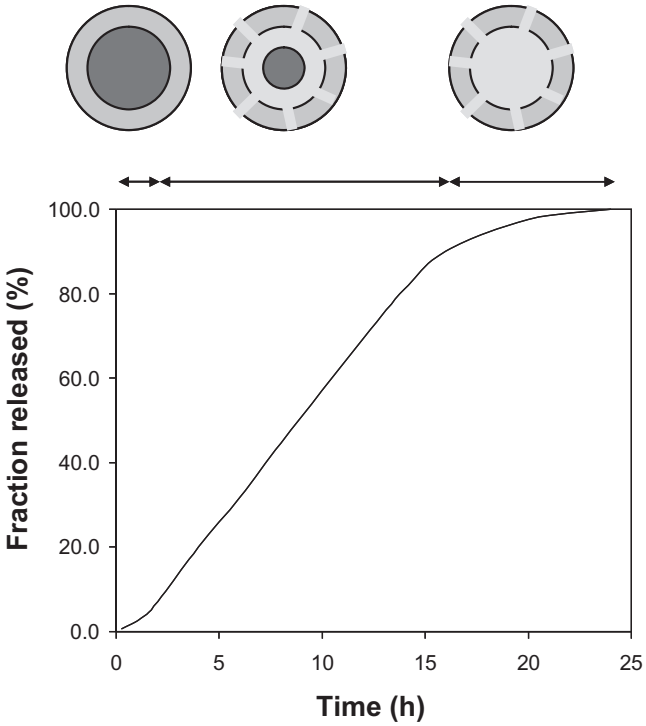


**FIGURE 4** Release from membrane-coated reservoir system, where the three different stages are depicted (initial lag period, "steady-state" period, and final depletion period). A schematic picture of a formulation with the drug reservoir (dark grey) surrounded by the membrane (middle grey) is presented at the top. The dissolution medium penetrates the membrane and dissolves the drug (light gray), and pores/cracks are formed through which the drug can be released.

period, the amount of released drug will be low. The water penetrates the membrane and at the same time starts to dissolve the material in existing pores, consisting of water-soluble sugars, salt, or polymers. The water that has penetrated the membrane and reached the core starts to dissolve the active substance in the reservoir [11]. The concentration of the active drug increases continuously until its saturation concentration has been reached and pseudoequilibrium between the solid material and saturated solution inside the membrane has developed. The osmotic pressure in the reservoir will depend on the concentrations of all dissolved species inside the membrane, and an increase of the osmotic pressure may lead to membrane rupture. After an initial lag period, the number of cracks and pores becomes constant, and the osmotic pressure and the concentration of dissolved species in the reservoir and in the membrane reach their steady-state levels. When these parameters, which are the driving forces for the drug release, are constant, the drug release rate will also be constant. Therefore, a second period with time-independent drug release rate will occur.

When the solid drug material inside the membrane is completely dissolved and the concentration of the dissolved drug decreases, a final period with declining

release rate will be entered. Ragnarsson et al. [49] have shown that the solid material disappears earlier as the drug solubility increases and that the third stage with time-dependent and decreasing release rate appears earlier compared to drugs with lower solubility.

The contribution of diffusion to the release process can be modeled by using Fick's first law. For diffusion of a substance through the membrane, it will turn out as (assuming sink conditions)

$$\frac{dM_t}{dt} = JA = \frac{D_m K_m A}{h}(C_s - C_b) = \frac{D_m K_m A C_s}{h} \tag{6}$$

where $K_m$ is the partition coefficient for the drug between the membrane and solution, $D_m$ is the diffusion coefficient in the membrane, $A$ is the area of the membrane, $h$ is the thickness of the membrane, and $C_s$ and $C_b$ are the concentrations on the inside of the membrane surface and in the bulk, respectively. The equation for diffusion through pores or cracks resembles Equation (6):

$$\frac{dM_t}{dt} = JA = \frac{D\varepsilon KA}{h\tau}(C_s - C_b) = \frac{D\varepsilon KAC_s}{h\tau} \tag{7}$$

where $\varepsilon$ and $\tau$ are introduced to describe the porosity and tortuosity in the membrane, respectively. The parameters $D_m$ and $K_m$ in Equation (6) are replaced in Equation (7) by $D$ and $K$, which are the diffusion coefficient in the solution inside the pores and the partition coefficient between the solution and materials surrounding the liquid-filled pores, respectively. Equations (6) and (7) depend on the concentration gradient over the membrane, and both are independent of the time.

The osmotic contribution to the drug transport is described by the so-called Kedem–Katchalsky equations (based on nonequilibrium thermodynamics) [50, 51]. A simplified version is

$$\frac{dM_t}{dt} = JA = \frac{AC_s L_p \sigma \Delta\Pi}{h} \tag{8}$$

where $\Delta\Pi$ is the osmotic pressure difference over the membrane and $L_p$ and $\sigma$ are the hydraulic permeability and the reflection, respectively.

A comparison of Equations (6)–(8) shows important similarities; they depend on the solubility of the drug, the area of the device, and the thickness of the membrane. This means that an increased solubility, larger area of the membrane, and thinner membranes will facilitate the drug release rate. This can be exemplified by a study by Ragnarsson and Johansson [11], who showed that, for different salt forms of metoprolol, an increased solubility also increased the drug release rate, which was predicted from the equations. Furthermore, Equations (6)–(8) show constant and time-independent release rates. This constant amount of released drug will be a biopharmaceutical benefit since it theoretically makes it possible to achieve a constant concentration of the drug in the blood plasma.

One special type of ER formulation based on coated reservoir systems is the so-called osmotic-controlled oral drug delivery system or osmotic pump. Pure osmotic systems have semipermeable membranes; that is, water can permeate the membrane but not other substances. These semipermeable membranes can be made of, for example, cellulose acetate [44], and for such formulations the dominating release mechanism is the osmotic pressure [Equation (8)]. The oldest formulation based on osmotic release was OROS from Alza Corporation [47]. In order to achieve drug release through the semipermeable membrane, a laser hole was drilled, but today many osmotic formulations instead use pores filled with water-soluble materials. The aspects of the formulation and different types of commercially available formulations are summarized in the review by Verma et al. [52]. Some advantages with osmotic pumps compared to other ER formulations such as hydrophilic matrix systems are (i) the time-independent drug release (often mentioned as zero-order release), (ii) the superior in vivo–in vitro correlation which facilitates further formulation development, and (iii) less variation between fasted and fed states. Potential drawbacks may be high initial development costs and lack of in-house competence. Drawbacks from an economic point of view may be the necessity to pay royalties and the need for special equipment associated with laser drilling technology.

### 3.2 Manufacturing of Oral Membrane-Coated Systems

The first step in the production of membrane-coated systems is to prepare the drug reservoir, the core, of the system. The process steps for producing the core depend on the size of the core. In monolithic membrane-coated systems, the core can be a filled capsule or a tablet which is produced in the traditional ways. This may include mixing of active substance and excipients, possibly granulation and drying, filling into capsules, or compaction into tablets. The production of cores for multiple-unit systems (often termed pellets) is more sophisticated and may be performed in different ways. One, and probably the most common way, is to produce cores for multiple-unit systems by extrusion and spheronizing [53–55]. An alternative methodology is to coat an inert core, e.g. glass or nonpareil beads, with the active substance and the excipients [11]. When the drug-containing core is manufactured, the process continues with the coating of the release-controlling membrane.

The composition of the core depends on the properties of the drug and the excipients, the chosen production chain, and whether the systems should be a monolithic or multiple-unit system. The compositions and production steps are reviewed by Tang et al. [56]. In general, the core will include the active substance together with the filler materials and, if necessary, solubilizers and lubricants/glidants. Classical filler materials are lactose and microcrystalline cellulose, but also other materials such as dextrose, mannitol, sorbitol, and sucrose can be used. However, it should be remembered that the dissolution of the filler material might influence the osmotic pressure. The effect of filler solubility has been investigated by Sousa et al. [54], who found a relation between the solubility of the filler materials and drug release rate. For filler materials with large water solubility, there is a great risk that the membrane will rupture due to the development of an excessive osmotic pressure, which will influence the drug release rate.

The choice of film-forming materials and film-coating techniques is critical for the drug release rate [54]. The ER membrane should remain intact during the release, which implies that it should not dissolve or erode. As film-forming material, water-insoluble substituted cellulose derivatives such as ethylcellulose have been suggested [53, 57] as well as synthetic polymers such as methylacrylates (e.g., Eudragit NE 30D, RS30D or RL30D, where NE stands for nonionic and RS/RL correspond to cationic polymers) [55]. A commercial technology platform is available under the name Eudramode, which is a platform for development of multiple-unit systems with extended drug release based on Eudragit [58, 59]. Other film-coating materials such as shellac and zein have been used, but a drawback with these naturally occurring materials is the variation in quality. To obtain a film with satisfying release properties, channeling agents such as the hydrophilic polymers hydroxypropyl methylcellulose (HPMC), hydroxypropyl cellulose (HPC), and polyethylene glycol (PEG) or other water-soluble materials such as sodium chloride or lactose may be used [2, 57]. To improve the mechanical properties of the film and thus avoid ruptures and cracks in the film, plasticizers may be added to the formulation. Examples of plasticizers are PEGs, diethyl phthalate, triacetin, mineral oils, glycerol, and chlorobutanol [60].

The film coating may be performed in different types of equipment. For coating of larger units, such as tablets or capsules, a rotating drum is often used, for example an Accela Coater, but other similar equipment may also be used [10, 53, 57, 61]. This type of equipment contains a perforated pan that rotates and thus mixes the bed of units. At the same time, a coating liquid is applied to the moving units by means of a spray gun, where the mixing of the units ensures uniform coverage of the coating. The film coating is dried by blowing a stream of hot air onto the surface of the tablets. For all types of coating processes, there are many parameters, such as the temperature and relative humidity of the inlet air, drum rotation speed, spraying rate, and droplet size of the coating liquid, that have to be adjusted in order to produce good coated films.

For coating smaller units, such as pellets, the fluid bed coating technique is used [56]. This is an attractive technique in which the starting material is placed in the coater and heated air is blown through a base plate. This leads to vigorous mixing of the pellet units. By changing the pressure of the incoming air stream, the material becomes suspended in the air. This happens when the bed starts to fluidize and the bed will then have fluidized properties similar to the properties of ordinary liquids. There are different designs of fluid bed: top spray coating, bottom spray coating (or Wurster coating), or tangential coating. They differ with regard to placement of the spray guns. In top spray coating, the liquid is sprayed from the top of the equipment and the droplets hit the particles moving in opposite direction. In the bottom spray coating, the sprayed liquid drops and particles flow in the same direction, which avoids the problem of blocking the spray guns that may occur in top spray equipment. In the more rarely tangential spray coating equipment the base plate rotates and the spray guns are spraying in a tangential direction to the spiral moving particles [10, 62].

A critical parameter for obtaining films with desirable properties is the creation of coating droplets, a process often referred to as atomization, that is, when a bulk liquid is dispersed in air to form a spray or a mist [61]. Atomization is done by letting pressurized air and bulk solution pass simultaneously through a nozzle (spray

gun). The air divides the droplets into smaller units, but in spite of the name atomization, the droplets are dispersed not at atomic scale but rather in the nano- to micrometer-scale range [61].

The mechanism for film formation depends on whether the polymer is dissolved or dispersed as small latex particles in the solution. For both technologies the film formation process starts with wetting and spreading of the coating droplets on the surface of the reservoir system [10]. For the case with polymer dissolved in solution, the film formation mechanism continues with evaporation of the solvent. This leads to an increase of the polymer concentration, and at a certain limit the polymers precipitate and a coated film is formed. In the case of ER films, the most commonly dissolved polymer is ethyl cellulose and the most commonly used solvent is ethanol but other organic solvents can also be used. The use of organic solvent is problematic from the SHE (safety, health, and environmental) point of view and therefore these aspects must be considered before starting to use organic-based coating process technologies. An alternative process methodology is to disperse particles of the film-forming polymer in an aqueous solution [63]. Film formation of dispersed particles undergoes the following steps: evaporation of water, close packing of the particles, deformation of the particles, and annealing of the particles by migration of individual polymer chains between the particles to form a coherent, smooth film coating. The first steps of the process occur in the coating equipment but the last, the annealing step, may continue days after the coated product has left the coating equipment.

## 4 HYDROPHILIC MATRIX TABLETS

### 4.1 Principles of Formulation and Release Mechanisms

Several recent informative review articles on hydrophilic matrix systems have been published [e.g., 24, 64]. Hydrophilic matrix tablets are composed of an active substance, a hydrophilic polymer, release modifiers, lubricants, and glidants. The technology goes back to the mid-1960s when Lapidus and Lordi [65, 66] and Huber et al. [67] determined the drug release from hydrophilic matrix systems. The release mechanism for this type of formulation starts with dissolution of hydrophilic matrix polymers and the formation of a highly viscous polymer layer around the tablet (Figure 5) [64]. This layer is often referred to as a gel layer even though it normally contains only physical entanglements and not chemical cross-linkers, which is traditionally required for gels. However, the gel layer surrounds the inner (more or less dry) part of the tablet and this part is called the core. In traditional hydrophilic tablets, the active substance particles are embedded in the matrix carrier. The dissolution process of the active substance can start when the carrier material has dissolved in water and formed the aqueous gel layer, since without exposure to water the active substance cannot dissolve. Therefore, the "dry" core will shield the active drug from dissolution, which is one reason for the extended drug release from this type of formulation.

The release process for hydrophilic matrix tablets can be schematically described as in Figure 5. The left side of the figure shows a hydrophilic tablet undergoing dissolution, swelling, and release. An interface between the solution and gel layer, here

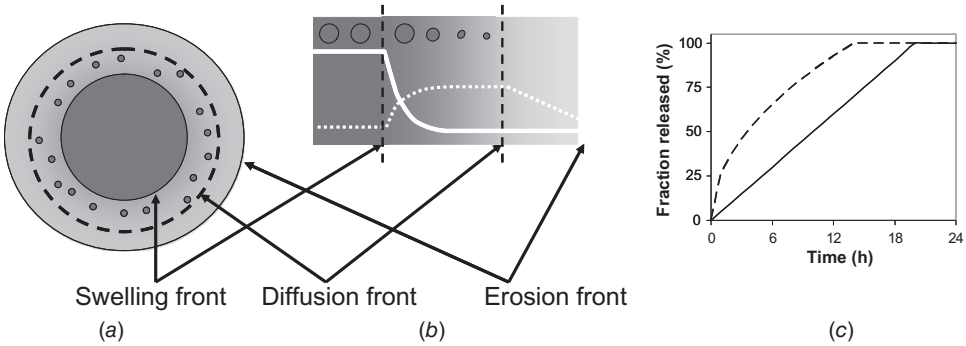


**FIGURE 5** (*a*) Hydrophilic matrix system shown with core (dark gray) and drug particles (small dark gray particles). The swelling, diffusion, and erosion fronts are depicted. (*b*) Dependencies of volume fraction polymer (solid line) and dissolved drug (broken line) as function of position in matrix system together with swelling, diffusion, and erosion fronts. Top shows how the solid drug particles diminish in size. (*c*) Examples of drug release as function of time for erosion-controlled ($n = 1$, solid line) and diffusion-controlled systems (broken line, $n = 0.5$).

called the erosion front, can be identified, and the polymer chains and drug molecules are released at this front. In the gel layer, the polymer concentration will decrease (Figure 5) from a highly concentrated solution at the swelling front, the interface between the gel layer and more or less dry core, to a diluted polymer solution at the erosion front [68]. Some authors have suggested that the polymer concentration at the erosion front is related to the overlap concentration [68–70]. At the same time, the water content in the gel layer gradually increases from the center of the tablet toward the erosion front, and thus the dissolution of the active substance particles can start already in the gel layer. However, the volume fraction of dissolved drug depends on the amount of available water, which in turn is a function of the position in the gel layer [64, 71]. Assuming that the drug saturation concentration is equal in water solutions and polymer gels, the volume fraction dissolved drug will correlate with the volume fraction water available, and the volume fraction drug will gradually increase with increasing distance from the core. Far from the core, the variations in the concentration of polymer and water are less pronounced and, as long as solid particles coexcite with the saturated drug solution, the volume fraction of the drug will theoretically be almost constant. A third front has also been introduced, the diffusion front, which corresponds to the position in the gel layer where all of the active substance has dissolved. Between the swelling and dissolution front, dissolved and undissolved drug particles will coexist, but between the dissolution and erosion fronts, only dissolved drug molecules occur. In this region, the diffusion of drug out from the matrix will give rise to a decrease in the volume fraction of the drug.

Achieving a mechanistic understanding of drug release from hydrophilic matrix tablets is not a trivial task since the release depends on the properties of both the polymer and the drug. The swelling process is directly related to the properties of the polymers, and this is an important factor for drug release, since the polymer swelling process can, for example, transport individual drug particles through the

gel layer, which has been shown by Adler et al. [72]. Macroscopically, polymer swelling can be observed as an increase in the size of the tablet. On a molecular level, the swelling depends on the dilution and transport of water into the gel layer. This transport is driven by changes in the chemical potential, and the main contribution is the increase of conformational entropy when the polymer chains are diluted [73]. The kinetics of the swelling process may vary. A faster polymer release rate compared to the swelling rate results in a movement of the erosion front toward the center of the tablet and the size of the tablet will diminish. Conversely, the tablet size increases when the polymer release is slower than the swelling process. When the swelling rate is on the same order as the polymer release rate, the position of the erosion front (i.e., tablet size) will remain constant.

The swelling front between the gel layer and amorphous (or semicrystalline) core material has traditionally been described as corresponding to a transition of the solid states of the polymers. The polymers in the core, for example in HPMC tablets, are in a glassy state, and the polymer material is transformed to a rubbery state due to the fact that water acts as a plasticizer and decreases the glass transition temperature [74]. This rubbery state can be regarded as a polymer solution, and therefore the glassy-to-rubbery state transition can be regarded as a dissolution process of the polymer, where the dissolution rate will determine the position of the swelling front. A commonly described special case for hydrophilic tablets is the so-called front synchronization, which is when the movements of the swelling front and the erosion front occur equally fast. This special case corresponds to a constant gel layer thickness.

Depending on the drug solubility and the dissolution, swelling, and release processes of the polymer, either of two different drug mechanisms can be observed: erosion- or diffusion-controlled drug release (Figure 5). One way to characterize these two mechanisms is to compare the drug and the polymer release. The erosion mechanism is characterized by equal release of the polymer and the drug, whereas when the release of the drug is faster than that of the polymer, this is called diffusion-controlled drug release [64]. The diffusion-controlled mechanism occurs when the diffusion front, the border between undissolved and dissolved drug, is displaced in the gel layer and the drug can efficiently diffuse out from the gel layer. The erosion mechanism dominates when the diffusion and erosion fronts overlap. This means that drug particles may be released from the surface of the gel layer. When this occurs, the drug particles will dissolve faster in the free solution than in the gel layer due to the fact that stirring is more efficient in the free solution, which results in a decreased thickness of the unstirred boundary layer around the particles and thus an increased dissolution rate [Equation (1a)]. Whether the release mechanism will be diffusion or erosion controlled depends on (i) the polymer release rate, which governs the position of the rate erosion front, and (ii) the drug dissolution rate, which governs the position of the diffusion front. The position of the diffusion front and the dissolution rate depend on the solubility of the drug. Lower solubility of the drug gives slower dissolution rates and hence the diffusion front can overlap more easily with the erosion front, which yields erosion-controlled drug release. On the other hand, large solubility of the drug will give diffusion-controlled drug release [75–78].

Traditionally, the drug release rate from hydrophilic matrix systems has been modeled as [79, 80]

$$Q = \int_0^t \frac{dM}{dt} dt = kt^n \tag{9}$$

where $Q$ is the accumulated amount of released drug and $k$ and $n$ are constants. The values of $n$ have been suggested to describe the drug release mechanisms. Release from a planar surface with $n = 1$ has been shown to correspond to erosion-controlled drug release and $n = 0.5$ to pure diffusion-controlled drug release. This is strictly only valid when the polymer release has $n = 1$, but this is often the case. In practice, for hydrophilic matrix systems, one often finds $n$ to be between 0.5 and 1, indicating that both the diffusion of the drug and the polymer erosion influence drug release. For other geometric shapes, such as tablet shapes, the limits for $n$ shift to 0.45 and 0.89 for diffusion- and erosion-controlled release, respectively [81–83].

Another popular way to describe the drug release is to characterize the influence of the relative contributions of erosion and diffusion to drug release as [84]

$$Q = at^m + bt^{2m} \tag{10}$$

where $a$, $b$, and $m$ are constants. The first factor in Equation (10) should represent the Fickian diffusional contribution and the second term to the erosion contribution to the drug transport. One mechanistic drawback with this approach is that it treats the diffusion and erosion processes as independent of each other, which they are probably not in any practical case.

The drug release from hydrophilic matrix tablets has been found to vary with the polymer parameters, the composition of the formulation, and the process parameters. Examples of important polymer-related parameters with significant influence on drug release are the viscosity and the hydrophilicity of the polymer, where polymers of larger viscosity grades give lower release rates and longer durations of the release [85] and an increased hydrophilicity gives larger swelling and faster drug release, which was found by comparing different degrees of substitution of HPMC [86]. The drug release rate also depends on the composition of the formulation. When components with high water solubility, such as lactose, are included in the matrix, the drug release increases, which can be seen as a dilution of the gel-forming material [69, 87]. Similarly, an increased amount of a soluble active substance also increases the release rate of drugs, probably also due to the corresponding decrease of the relative amount of hydrophilic polymer [88, 89]. The size and shape of the tablet also influence drug release, and the release rate increases as the area-to-volume ratio increases [90–93]. The influence of particle size on the drug release rate has also been investigated [89, 94, 95]. The size of polymer particles seems to have low influence on drug release when there is enough gelling polymer available to quickly form a coherent gel layer. In contrast, at low amounts of hydrophilic matrix polymer, gel formation may be too slow, which makes polymer particle size important. In this case, the matrix may disintegrate before it develops a coherent gel layer [95–97]. The effect of drug particle size on the drug release rate depends on the solubility of the substance. Varying the particle size of drugs with high solubility seems to have little influence [89], whereas the release rate may depend on the particle sizes of drugs with moderate solubility [24, 89].

The effects of process parameters on drug release have been discussed in the literature. Different granulation technologies, such as dry granulation [98–100] or wet granulation [101–103] (which includes fluid bed granulation [104, 105]), have been used. Also direct compression has been used for production of hydrophilic matrix tablets [95]. The effects on the choice of production steps may be critical when the relative fraction of the polymers is low in the formulations. The effect of the compaction pressure on drug release has also been studied. Several authors [89, 88, 106, 107] have found that, while the compaction pressure has a significant effect on the tensile strength of the tablets, it has a minimal influence on drug release. This can be due to the fact that, when a coherent gel layer is formed, only the parameters governing the performance of this gel layer are important, and since parameters such as porosity do not affect the gel layer, they are of low importance for the drug release rate.

### 4.2 Manufacturing of Hydrophilic Matrix Tablets

The traditional way of producing hydrophilic matrix tablets resembles the production of the core for membrane-coated tablets and insoluble matrix tablets. It includes a mixing step, possibly a granulation step, a compaction step, and sometimes a coating step. However, one large difference between the production of insoluble and soluble matrix systems is notable; the latter matrix type has strong interactions with water, which complicates the production steps when water is present [108]. Therefore, for hydrophilic matrix systems with large fractions of hydrophilic polymers, traditional wet granulation with water as granulation liquid may cause problems with formation of hard lumps [108]. To avoid this problem, a new technique using foam granulation has been suggested [109, 110]. During granulation, the foams will flow on the top of already foam-wetted particles, which may lead to superior distribution of the granulation liquid. An alternative wet granulation method is to use organic solvents such as ethanol as granulation liquid [111]. However, even if the production of granules with good compaction properties can be maintained in this way, it may, as already mentioned, be an advantage to consider the SHE aspects before choosing organic solvents as granulation liquid. Another alternative to wet granulation is to use dry granulation technologies such as roller compaction [98, 99, 100]. Although this technique has several advantages compared to traditional wet granulation with water, it may result in lower tensile strengths of the tablets [112]. An alternative to granulation technologies is direct compression. This can be done by purchasing special direct-compression qualities of the hydrophilic polymers. A drawback that always arises in relation to direct compression is the difficulty to achieve content uniformity of the tablets, a problem that increases with decreasing doses of the active drug. Therefore, special care should be taken with regard to formulations with low doses of active substance when direct compaction is used [2].

Water-soluble hydrophilic matrix systems may also be extruded, both to monolithic systems and to multiple-unit systems [12, 14–16]. Polyethylene oxide (PEO) and chloropheniramine maleate have, for example, been extruded to a monolithic unit [113]. This manufacturing method proved more feasible for mixtures between low- and high-molecular-weight PEO, since systems containing only high-molecular-weight PEO proved too viscous and difficult to extrude. It was also

**TABLE 2 FDA Registered Oral ER Formulations Containing Commonly Used Excipients in Hydrophilic Matrix Tablets**

| Hydrophilic Polymer[a] | Fraction of Excipient (%) | Number of Hits on FDA Homepage | Other Names of Excipients |
|---|---|---|---|
| Hydroxypropyl methyl cellulose [24] | 10–80 | 12 | Hypromellose, HPMC, MHPC |
| Hydroxypropylcellulose [120, 121] | 15–35 | 4 | HPC |
| Polyethylene oxide [68, 128] | 5–85 | 4 | PEO |
| Sodium alginate [134] | 5–50 | 3 | |
| Xanthan gum [125, 137–140] | 10–99 | 3 | |
| Hydroxyethyl cellulose [117–119] | 40–97.5 | 3 | |
| Carbomer [129] | 20–75 | 1 | Carbopol, polyacrylic acid |
| Methyl cellulose [121] | 5–75 | 1 | MC |

*Source*: http://www.accessdata.fda.gov/scripts/cder/iig/index.cfm.

[a]Examples of references using them are given.

shown that degradation of PEO to lower molecular weight might occur due to the high temperature used in the extruder. Extrusion to small pellets based on hydrophilic polymers hydroxyethyl cellulose (HEC) and HPMC has been performed, and isopropyl alcohol instead of water was used to avoid lump formation [12]. Also Carbopol [14–16], pectin [114], and xanthan [115, 116] have been extruded and used as matrix carrier materials for extended release.

The choice of hydrophilic polymer is one determining factor for the drug release rate, as discussed above. The most common type used for hydrophilic matrix systems is HPMC (hypromellose, hydroxypropyl methyl cellulose) [24, 83] but some alternatives are HEC [117–119], HPC [120, 121], methyl cellulose [121], xanthan [122–127], PEO [68, 128], Carbopol [129–131], pectin [131–133], and alginate [134–136]; see Table 2. HPMCs are available in several approved degrees of substitution [60]. To achieve different release durations, different viscosity grades of the polymers may be used (the higher the viscosity grade, the slower the release rate). If the desired viscosity is not available, one can mix two polymer grades with different viscosity grades [68].

## 5 COMPARISON AND SUMMARY OF DIFFERENT TECHNOLOGIES

The main principles for oral extended drug release reviewed here are the membrane-coated reservoir systems and hydrophilic or hydrophobic matrix systems. They all have advantages and drawbacks and Table 3 summarizes some aspects of the different formulation principles.

There are a number of commercial variants of the above-mentioned principles. Examples of commercial matrix-based formulations can be found in a review by Varma et al. [137]. One of the oldest commercial hydrophilic matrix systems is

**TABLE 3 Comparison between Three Drug Release Principles**

| Aspects | Insoluble Matrix Systems | Membrane-Coated Systems | Hydrophilic Matrix Systems |
|---|---|---|---|
| Drug solubility | Dose dependent [18] | Dose dependent [18] | Whole range |
| Release mechanism | Diffusion controlled | Diffusion and osmotic | Diffusion to erosion controlled |
| Release profile | $Q \propto t^{0.5}$ | $Q \propto t^{1}$ | $Q \propto t^{n}$, $0.4 < n < 1$ |
| Main release dependence (except drug properties) | Channeling components and processes parameters | Properties of membrane | Properties of polymer carrier |
| Composition alternatives | Many approved and functional excipients available<br>Pore-channeling excipients may determine drug release<br>Small changes in formulation may change release rate | Many approved and functional excipients available<br>Release can be changed by film and core compositions<br>Many polymers are based on natural material which may give large batch-to-batch variation | Many grades of available polymers exist<br>Easy to regulate drug release by means of composition and polymer properties<br>Many polymers are based on natural material, which may result in large batch-to-batch variation |
| Manufacturing aspects | Uses traditional production technologies and is inexpensive<br>Release depends on process parameters | Coating process can be used to control drug release<br>Film coating process dominates release, which may be sensitive in relation to many different process parameters | Release is mainly robust against process parameters<br>Granulation with water is difficult |
| In vivo–in vitro correlation | Release depends on fasted and fed state | Good | Release depends on fasted and fed state |

TIMERx, which is based on xanthan gum mixed with, for example, locus bean gum [138]. A mixture of these polymers has a special property; the different polymers interact strongly with each other and this interaction produces gels with large viscosity. Variants of osmotically controlled ER formulations, some of which are commercially available, are reviewed by Verma et al. [52].

## 6 OTHER ORAL ER FORMULATIONS

Here some new oral ER technologies based on principles other than those mentioned above will be presented. It is beyond the scope of this chapter to cover all systems and details thereof, and the interested reader is recommended to make use of the literature included in the references.

One promising technology is to let the active drug interact with the excipients, for example, by covalent binding between the drug and excipient. The azoaromatic cross-linkers between the drug and excipient can exemplify this. These special cross-linkers break due to bacterial degradation in the colon, but not until they are exposed to this bacterial flora [139–141]. This means that the drug is hindered from release before it reaches the colon. This is an example of site-specific delivery to the colon, an area that is reviewed by a special issue of the journal *Advanced Drug Delivery Reviews* 2005 (volume 57, number 2).

Another example of modifications of ER systems is when electrostatic interaction between charged drugs and excipients (with opposite charge) is used. This concept can be valuable for drugs with pH-dependent solubility. If, for example, the drug is a weak base, it will have a large solubility at low pH, and a major part of the drug may be rapidly released already in the stomach, which is not desirable. An introduction of polyions with opposite charge can result in electrostatic interaction between the drugs and the polyions, which can change the solubility or decrease the diffusion rate of the drug through the ER formulation. This concept is, for example, used when propranolol hydrochloride interacts with sodium carboxymethylcellulose (NaCMC) in HPMC matrices [142]. In addition, buffering of the ER systems by organic or inorganic buffers has been used to obtain pH-independent drug release [27, 143, 144]. A recent publication by Riis et al. showed that insoluble inorganic buffers such as magnesium hydroxide and magnesium oxide provided stable drug release over longer periods of time than when more soluble buffer systems were used [26].

Sophisticated systems based on responsive gels are promising alternatives in terms of oral extended release. The principle behind many of these systems is that the formulation should react and undergo some kind of transition due to a trigger signal [145]. In general, the triggers can be various factors such as temperature, light, pressure, electrosensitivity, or interactions with specific molecules such as glucose or antigens. For oral drug delivery systems, the triggers can also be a biological change in the gastrointestinal environment, such as a change in the bacteria flora, as mentioned earlier [139]. It can also be a physiochemical change, such as a pH change in the gastrointestinal tract. Hydrogels composed of copolymers of poly(acrylic acid) and covalently attached Pluronic surfactants is one such example that reacts on pH changes. The cross-linked microgels can be loaded with drug and tailored to collapse in low pH but swell and release the drug at high pH [146].

A novel method for producing ER formulations is a technology called three-dimensional printing (TheriForm technology) [147], which is similar to the one used in ordinary printers. The ink is here replaced with an active substance and carrier material. The layer-by-layer "printing" provides controlled placement of the active drug and thus of the release from the device [148].

Even if there have been advances in oral drug delivery technologies during the last 50 years, many highly sophisticated drug delivery systems have failed and have

not reached the market [149]. This can be due to several reasons and in many cases the formulation is probably not the cause. However, on occasion one may be tempted to agree with Rocca and Park in their review of prospects and challenges in the oral drug delivery: "breakthrough technologies are required to generate novel dosage forms raising drug delivery to higher level" (p. 52) [149].

## ACKNOWLEDGMENT

Sven Engström, Chalmers University of Technology, is acknowledged for his useful suggestions and comments.

## REFERENCES

1. Malinowski, H. J., and Marroum, P. J. (1999), *Encyclopedia of Controlled Drug Delivery*, vols. 1 and 2, *Food and Drug Administration Requirements for Controlled Release Products*, John Wiley & Sons, New York, vol. 1, pp. 381–395.
2. Aulton, M. E. (2002), *Pharmaceutics: The Science of Dosage Form Design*, Churchill Livingstone, Hartcourt Publisher, Edinburgh.
3. Marroum, P. J. (1997), Bioavialability/Bioequivalence for Oral Release Producs, Controlled Release Drug Delivery Systems. Paper presented at the 5th Int. Symp. Drug Del., East Brunswick, NJ, May 15–17.
4. SCHOLAR, SciFinder (2006), American Chemical Society, Washington, DC.
5. Yamanaka, K., Yamamoto, H., Kawamura, Y., and Ito, S. (1954), Pyrimidine pencillin, U.S. Patent 2,681,339.
6. Robinson, M. J., and Svedres, E. V. (1957), Sustained-release pharmaceutical preparations, U.S. Patent 2805977.
7. Sjögren, J., and Fryklöf, L-E. (1960), Duretter—A new type of oral sustained action preparation, *Farmacevtisk Revy*, 59, 171–179.
8. Sjögren, J. (1971), Studies on a sustained release principle based on inert plastic matrix, *Acta Pharm. Suecica*, 8, 153–168.
9. Speers, M., and Bonnano, C. (1999), Economic aspects of controlled drug delivery, in (Mathiowitz, E., Ed.), *Encyclopedia of Controlled Drug Delivery*, Wiley, New York, pp. 341–347.
10. Cole, G., Hogan, J., and Aulton, M., Eds. (1995), *Pharmaceutical Coating Technology*, Taylor & Francis, London, 1995.
11. Ragnarsson, G., and Johansson, M. O. (1988), Coated drug cores in multiple unit preparations. Influence of particle size, *Drug Dev. Ind. Pharm.*, 14, 2285–2297.
12. Chatlapalli, R., and Rohera, B. D. (1998), Physical characterization of HPMC and HEC and investigation of their use as pelletization aids. *Int. J. Pharm.*, 161, 179–193.
13. Kojima, M., and Nakagami, H. (2002), Development of controlled release matrix pellets by annealing with micronized water-insoluble or enteric polymers, *J. Controlled Release*, 82, 335–343.
14. Young, C. R., Dietzsch, C., Cerea, M., Farrell, T., Fegely, K. A., Rajabi-Siahboomi, A., and McGinity, J. W. (2005), Physicochemical characterization and mechanisms of release of theophylline from melt-extruded dosage forms based on a methacrylic acid copolymer, *Int. J. Pharm.*, 301, 112–120.

15. Neau, S. H., Chow, M. Y., and Durrani, M. J. (1996), Fabrication and characterization of extruded and spheronized beads containing Carbopol 974P, NF resin, *Int. J. Pharm.*, 131, 47–55.
16. Bommareddy, G. S., Paker-Leggs, S., Saripella, K. K., and Neau, S. H. (2006), Extruded and spheronized beads containing Carbopol 974P to deliver nonelectrolytes and salts of weakly basic drugs, *Int. J. Pharm.*, 321, 62–71.
17. Abrahamsson, B., Alpsten, M., Jonsson, U. E., Lundberg, P. J., Sandberg, A., Sundgren, M., Svenheden, A., and Toelli, J. (1996), Gastro-intestinal transit of a multiple-unit formulation (metoprolol CR/ZOK) and a non-disintegrating tablet with the emphasis on colon, *Int. J. Pharm.*, 140, 229–235.
18. Thrombre, A. G. (2005), Assessment of feasibility of oral controlled release in an exploratory development setting, *Drug Discov. Today*, 10, 1159–1166.
19. Li, S., He, H., Parthiban, L. J., Yin, H., and Serajuddin, A. T. M. (2005), IV-IVC considerations in the development of immediate-release oral dosage form, *J. Pharm. Sci.*, 94, 1396–1417.
20. Noyes, A. A., and Whitney, W. R. (1897), The rate of solution of solid substances in their own solutions, *J. Am. Chem. Soc.*, 19, 930–934.
21. Dokoumetzidis, A., and Macheras, P. (2006), A century of dissolution research: From Noyes and Whitney to the Biopharmaceutics Classification System, *Int. J. Pharm.*, 321, 1–11.
22. Mooney, K. G., Mintun, M. A., Himmelstein, K. J., and Stella, V. J. (1981), Dissolution kinetics of carboxylic acids. I: Effect of pH under unbuffered conditions, *J. Pharm. Sci.*, 70, 13–22.
23. Mooney, K. G., Mintun, M. A., Himmelstein, K. J., and Stella, V. J. (1981), Dissolution kinetics of carboxylic acids. II: Effects of buffers, *J. Pharm. Sci.*, 70, 22–32.
24. Li, C. L., Martini, L. G., Ford, J. L., and Roberts, M. (2005), The use of hypromellose in oral drug delivery, *J. Pharm. Pharmacol.*, 57, 533–546.
25. Tatavarti, A. S., and Hoag, S. W. (2006), Microenvironmental pH modulation based release enhancement of a weakly basic drug from hydrophilic matrices, *J. Pharm. Sci.*, 95, 1459–1468.
26. Riis, T., Bauer-Brandl, A., Wagner, T., and Krantz, H. (2007), pH-independent drug release of an extremely poorly soluble weakly basic acidic drug from multiparticulate extended release formulations, *Eur. J. Pharm. Biopharm.*, 65, 78–84.
27. Streubel, A., Siepmann, J., Dashevsky, A., and Bodmeier, R. (2000), pH-independent release of a weakly basic drug from water-insoluble and -soluble matrix tablets, *J. Controlled Release*, 67, 101–110.
28. U.S. Food and Drug Administration (FDA) (2006), available: http://www.fda.gov.search.html.
29. Higuchi, T. (1963), Mechanism of sustained-action medication. Theoretical analysis of rate of release of solid drugs dispersed in solid matrices, *J. Pharm. Sci.*, 52, 1145–1149.
30. Wu, X. S. (1996), *Controlled Drug Delivery Systems*, Technology Publishing, Lancaster, PA.
31. Jannin, V., Pochard, E., and Chambin, O. (2006), Influence of poloxamers on the dissolution performance and stability of controlled-release formulations containing Precirol ATO 5, *Int. J. Pharm.*, 309, 6–15.
32. Sinchaipanid, N., Junyaprasert, V., and Mitrevej, A. (2004), Application of hot-melt coating for controlled release of propranolol hydrochloride pellets, *Powder Technol.*, 141, 203–209.

33. Rodriguez, L., Caputo, O., Cini, M., Cavallari, C., and Greechi, R. (1993), In vitro release of theophylline from directly-compressed matrixes containing methacrylic acid copolymers and/or dicalcium phosphate dihydrate, *Farmaco*, 48, 1597–1604.
34. Ceballos, A., Cirri, M., Maestrelli, F., Corti, G., and Mura, P. (2005), Influence of formulation and process variables on in vitro release of theophylline from directly-compressed Eudragit matrix tablets, *Farmaco*, 60, 913–918.
35. Oezyazici, M., Goekce, E. H., and Ertan, G. (2006), Release and diffusional modeling of metronidazole lipid matrices, *Eur. J. Pharm. Biopharm.*, 63, 331–339.
36. Hamdani, J., Moes, A. J., and Amighi, K. (2003), Physical and thermal characterization of Precirol and Compritol as lipophilic glycerides used for the preparation of controlled-release matrix pellets, *Int. J. Pharm.*, 260, 47–57.
37. U.S. Food and Drug Administration (FDA) (2006), available: http://www.accessdata.fda.gov/scripts/cder/iig/index.cfm.
38. Mehnert, W., and Mader, K. (2001), Solid lipid nanoparticles. Production, characterization and applications, *Adv. Drug Deliv. Rev.*, 47, 165–196.
39. Savolainen, M., Herder, J., Khoo, C., Lovqvist, K., Dahlqvist, C., Glad, H., and Juppo, A. M. (2003), Evaluation of polar lipid-hydrophilic polymer microparticles, *Int. J. Pharm.*, 262, 47–62.
40. Zhu, Y., Shah, N. H., Malick, A. W., Infeld, M. H., and McGinity, J. W. (2006), Controlled release of a poorly water-soluble drug from hot-melt extrudates containing acrylic polymers, *Drug Dev. Ind. Pharm.*, 32, 569–583.
41. Young, C. R., Dietzsch, C., and McGinity, J. W. (2005), Compression of controlled-release pellets produced by a hot-melt extrusion and spheronization process, *Pharm. Dev. Technol.*, 10, 133–139.
42. Young, C. R., Koleng, J. J., and McGinity, J. W. (2002), Production of spherical pellets by a hot-melt extrusion and spheronization process, *Int. J. Pharm.*, 242, 87–92.
43. Hogan, J. E. (1995), Modified release coatings for pharmaceutics, in Cole, G., Hogan, J., and Aulton, M., Eds., *Pharmaceutical Coating Technology*, pp. 409–438.
44. Edgar, K. J., Buchanan, C. M., Debenham, J. S., Rundquist, P. A., Seiler, B. D., Shelton, M. C., and Tindall, D. (2001), Advances in cellulose ester performance and application, *Prog. Polym. Sci.*, 26, 1605–1688.
45. Hjärtstam, J. (1998), *Ethyl Cellulose Membranes Used in Modified Release Formulations*, Chalmers University of Technology, Göteborg, Sweden.
46. Savastano, L., Leuenberger, H., and Merkle, H. P. (1995), Membrane modulated dissolution of oral drug delivery systems, *Pharm. Acta Helv.*, 70, 117–124.
47. Theeuwes, F. (1975), Elementary osmotic pumps, *J. Pharm. Sci.*, 64, 147–157.
48. Ragnarsson, G., Sandberg, A., Johansson, M. O., Lindstedt, B., and Sjoegren, J. (1992), In vitro release characteristics of a membrane-coated pellet formulation. Influence of drug solubility and particle size, *Int. J. Ph+arm.*, 79, 223–232.
49. Ragnarsson, G., Sandberg, A., Jonsson, U. E., and Sjoegren, J. (1987), Development of a new controlled release metoprolol product, *Drug Dev. Ind. Pharm.*, 13, 1495–1509.
50. Kedem, O., and Katchalsky, A. (1961), *J. Gen. Phys*, 45, 143.
51. Kedem, O., and Katchalsky, A. (1958), *Biochem. Biophys. Acta*, 27, 229.
52. Verma, R. K., Krishna, D. M., and Garg, S. (2002), Formulation aspects in the development of osmotically controlled oral drug delivery systems, *J. Controlled Release*, 79, 7–27.
53. Elchidana, P. A., and Deshpande, S. G. (1999), Microporous membrane drug delivery system for indomethacin, *J. Controlled Release*, 59, 279–285.

54. Sousa, J. J., Sousa, A., Moura, M. J., Podczeck, F., and Newton, J. M. (2002), The influence of core materials and film coating on the drug release from coated pellets, *Int. J. Pharm.*, 233, 111–122.

55. Husson, I., Leclerc, B., Spenlehauer, G., Veillard, M., and Couarraze, G. (1991), Modeling of drug release from pellets coated with an insoluble polymeric membrane, *J. Controlled Release*, 17, 163–173.

56. Tang, E. S. K., Chan, L. W., and Heng, P. W. S. (2005), Coating of multiparticulates for sustained release, *Am. J. Drug Deliv.*, 3, 17–28.

57. Verma, R. K., Kaushal, A. M., and Garg, S. (2003), Development and evaluation of extended release formulations of isosorbide mononitrate based on osmotic technology, *Int. J. Pharm.*, 263, 9–24.

58. Ravishankar, H., Patil, P., Petereit, H-U., and Renner, G. (2005), EUDRAMODE: A novel approach to sustained oral drug delivery systems, *Drug Deliv. Technol.*, 5, 48–50, 52–55.

59. Ravishankar, H., Iyer-Chavan, J., Patil, P., Samel, A., and Renner, G. (2006), Clinical studies of terbutaline controlled-release formulation prepared using EUDRAMODE, *Drug Deliv. Technol.*, 6, 50–56.

60. Kibbe, A. H. (2000), *Handbook of Pharmaceutical Excipients*, Pharmaceutical, London.

61. Aulton, M. E., and Twitchell, A. M. (1995), Solution properties and atomization in film coating [of pharmaceuticals], *Pharm. Coat. Technol.*, 64–117.

62. Glatt (2006), available: www.glatt.com.

63. Harris, M. R., and Ghebre-Sellassie, I. (1989), Aqueous polymeric coating for modified-release pellets, *Drugs Pharm. Sci.*, 36, 63–79.

64. Colombo, P., Bettini, R., Santi, P., and Peppas, N. A. (2000), Swellable matrixes for controlled drug delivery: Gel-layer behavior, mechanisms and optimal performance, *Pharm. Sci. Technol. Today*, 3, 198–204.

65. Lapidus, H., and Lordi, N. G. (1966), Some factors affecting the release of a water-soluble drug from a compressed hydrophilic matrix, *J. Pharm. Sci.*, 55, 840–843.

66. Lapidus, H., and Lordi, N. G. (1968), Drug release from compressed hydrophilic matrixes, *J. Pharm. Sci.*, 57, 1292–1301.

67. Huber, H. E., Dale, L. B., and Christenson, G. L. (1966), Utilization of hydrophilic gums for the control of drug release from tablet formulations. I. Disintegration and dissolution behavior, *J. Pharm. Sci.*, 55, 974–976.

68. Koerner, A., Larsson, A., Piculell, L., and Wittgren, B. (2005), Tuning the polymer release from hydrophilic matrix tablets by mixing short and long matrix polymers, *J. Pharm. Sci.*, 94, 759–769.

69. Ju, R. T. C., Nixon, P. R., and Patel, M. V. (1995), Drug release from hydrophilic matrixes 1. New scaling laws for predicting polymer and drug release based on the polymer disentanglement concentration and the diffusion layer, *J. Pharm. Sci.*, 84, 1455–1463.

70. Ju, R. T. C., Nixon, P. R., Patel, M. V., and Tong, D. M. (1995), Drug release from hydrophilic matrixes. 2. A mathematical model based on the polymer disentanglement concentration and the diffusion layer, *J. Pharm. Sci.*, 84, 1464–1477.

71. Bettini, R., Colombo, P., Massimo, G., Catellani, P. L., and Vitali, T. (1994), Swelling and drug release in hydrogel matrixes: Polymer viscosity and matrix porosity effects, *Eur. J. Pharm. Sci.*, 2, 213–219.

72. Adler, J., Jayan, A., and Melia, C. D. (1999), A method for quantifying differential expansion within hydrating hydrophilic matrixes by tracking embedded fluorescent microspheres, *J. Pharm. Sci.*, 88, 371–377.
73. Treloar, L. R. G. (1975), *The Physics of Rubber Elasticity*, Clarendon, Oxford.
74. Siepmann, J., Kranz, H., Bodmeier, R., and Peppas, N. A. (1999), HPMC-matrices for controlled drug delivery: A new model combining diffusion, swelling, and dissolution mechanisms and predicting the release kinetics, *Pharm. Res.*, 16, 1748–1756.
75. Colombo, P., Bettini, R., Massimo, G., Catellani, P. L., Santi, P., and Peppas, N. A. (1995), Drug diffusion front movement is important in drug release control from swellable matrix tablets, *J. Pharm. Sci.*, 84, 991–997.
76. Bettini, R., Catellani, P. L., Santi, P., Massimo, G., Peppas, N. A., and Colombo, P. (2001), Translocation of drug particles in HPMC matrix gel layer: Effect of drug solubility and influence on release rate, *J. Controlled Release*, 70, 383–391.
77. Fu, X. C., Liang, W. Q., and Ma, X. W. (2003), Relationships between the release of soluble drugs from HPMC matrices and the physicochemical properties of drugs, *Pharmazie*, 58, 221–222.
78. Tahara, K., Yamamoto, K., and Nishihata, T. (1996), Application of model-independent and model analysis for the investigation of effect of drug solubility on its release rate from hydroxypropyl methyl cellulose sustained-release tablets, *Int. J. Pharm.*, 133, 17–27.
79. Korsmeyer, R. W., Gurny, R., Doelker, E., Buri, P., and Peppas, N. A. (1983), Mechanisms of solute release from porous hydrophilic polymers, *Int. J. Pharm.*, 15, 25–35.
80. Rinaki, E., Valsami, G., and Macheras, P. (2003), The power law can describe the "entire" drug release curve from HPMC-based matrix tablets: A hypothesis, *Int. J. Pharm.*, 255, 199–207.
81. Ritger, P. L., and Peppas, N. A. (1987), A simple equation for description of solute release. II. Fickian and anomalous release from swellable devices, *J. Controlled Release*, 5, 37–42.
82. Ritger, P. L., and Peppas, N. A. (1987), A simple equation for description of solute release. I. Fickian and non-Fickian release from non-swellable devices in the form of slabs, spheres, cylinders or disks, *J. Controlled Release*, 5, 23–36.
83. Siepmann, J., and Peppas, N. A. (2001), Modeling of drug release from delivery systems based on hydroxypropyl methylcellulose (HPMC), *Adv. Drug Deliv. Rev.*, 48, 139–157.
84. Peppas, N. A., and Sahlin, J. J. (1989), A simple equation for the description of solute release. III. Coupling of diffusion and relaxation, *Int. J. Pharm.*, 57, 169–72.
85. Gao, P., and Meury, R. H. (1996), Swelling of hydroxypropyl methylcellulose matrix tablets. 1. Characterization of swelling using a novel optical imaging method, *J. Pharm. Sci.*, 85, 725–731.
86. Mitchell, K., Ford, J. L., Armstrong, D. J., Elliott, P. N. C., Hogan, J. E., and Rostron, C. (1993), The influence of substitution type on the performance of methyl cellulose and hydroxypropyl methyl cellulose in gels and matrixes, *Int. J. Pharm.*, 100, 143–154.
87. Levina, M., and Rajabi-Siahboomi, A. R. (2004), The influence of excipients on drug release from hydroxypropyl methylcellulose matrices, *J. Pharm. Sci.*, 93, 2746–2754.
88. Ford, J. L., Rubinstein, M. H., and Hogan, J. E. (1985), Propranolol hydrochloride and aminophylline release from matrix tablets containing hydroxypropyl methyl cellulose, *Int. J. Pharm.*, 24, 339–350.

89. Ford, J. L., Rubinstein, M. H., and Hogan, J. E. (1985), Formulation of sustained-release promethazine hydrochloride tablets using hydroxypropyl methyl cellulose matrixes, *Int. J. Pharm.*, 24, 327–338.
90. Reynolds, T. D., Mitchell, S. A., and Balwinski, K. M. (2002), Investigation of the effect of tablet surface area/volume on drug release from hydroxypropyl methylcellulose controlled-release matrix tablets, *Drug Dev. Ind. Pharm.*, 28, 457–466.
91. Skoug, J. W., Borin, M. T., Fleishaker, J. C., and Cooper, A. M. (1991), In vitro and in vivo evaluation of whole and half tablets of sustained-release adinazolam mesylate, *Pharm. Res.*, 8, 1482–1488.
92. Sung, K. C., Nixon, P. R., Skoug, J. W., Ju, T. R., Gao, P., Topp, E. M., and Patel, M. V. (1996), Effect of formulation variables on drug and polymer release from HPMC-based matrix tablets, *Int. J. Pharm.*, 142, 53–60.
93. Ford, J. L., Rubinstein, M. H., McCaul, F., Hogan, J. E., and Edgar, P. J. (1987), Importance of drug type, tablet shape and added diluents on drug release kinetics from hydroxypropyl methyl cellulose matrix tablets, *Int. J. Pharm.*, 40, 223–234.
94. Heng, P. W., Chan, L. W., Easterbrook, M. G., and Li, X. (2001), Investigation of the influence of mean HPMC particle size and number of polymer particles on the release of aspirin from swellable hydrophilic matrix tablets, *J. Controlled Release*, 76, 39–49.
95. Velasco, M. V., and Ford, J. L., Rowe, P., and Rajabi-Siahhoomi, A. R. (1999), Influence of drug: Hydroxypropyl methyl cellulose ratio, drug and polymer particle size and compression force on the release of diclofenac sodium from HPMC tablets, *J. Controlled Release*, 57, 75–85.
96. Dabbagh, M. A., Ford, J. L., Rubinstein, M. H., and Hogan, J. E. (1996), Effects of polymer particle size, compaction pressure and hydrophilic polymers on drug release from matrixes containing ethyl cellulose, *Int. J. Pharm.*, 140, 85–95.
97. Mitchell, K., Ford, J. L., Armstrong, D. J., Elliott, P. N. C., Hogan, J. E., and Rostron, C. (1993), The influence of the particle size of hydroxypropyl methyl cellulose K15M on its hydration and performance in matrix tablets, *Int. J. Pharm.*, 100, 175–179.
98. Sheskey, P. J., Cabelka, T. D., Robb, R. T., and Boyce, B. M. (1994), Use of roller compaction in the preparation of controlled-release hydrophilic matrix tablets containing methyl cellulose and hydroxypropyl methyl cellulose polymers, *Pharm. Technol.*, 18, 132, 134, 136, 138, 140, 142, 144, 146, 148–150.
99. Sheskey, P. J., and Hendren, J. (1999), The effects of roll compaction equipment variables, granulation technique, and HPMC polymer level on a controlled-release matrix model drug formulation, *Pharm. Technol.*, 23, 90, 92, 94, 96, 98, 100, 102, 104, 106.
100. Sheskey, P., Pacholke, K., Sackett, G., Maher, L., and Polli, J. (2000), Effect of process scale-up on robustness of tablets, tablet stability, and predicted in vivo performance, *Pharm. Technol.*, 24, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52.
101. Timmins, P., Delargy, A. M., Howard, J. R., and Rowlands, E. A. (1991), Evaluation of the granulation of a hydrophilic matrix sustained-release tablet, *Drug Dev. Ind. Pharm.*, 17, 531–550.
102. Liu, C. H., Chen, S. C., Kao, Y. H., Kao, C. C., Sokoloski, T. D., and Sheu, M. T. (1993), Properties of hydroxypropyl methyl cellulose granules produced by water spraying, *Int. J. Pharm.*, 100, 241–248.
103. Herder, J., Adolfsson, A., and Larsson, A. (2006), Initial studies of water granulation of eight grades of hypromellose (HPMC), *Int. J. Pharm.*, 313, 57–65.
104. Dahl, T. C., and Bormeth, A. P. (1990), Naproxen controlled release matrix tablets: Fluid bed granulation feasibility, *Drug Dev. Ind. Pharm.*, 16, 581–590.

105. Nellore, R. V., Rekhi, G. S., Hussain, A. S., Tillman, L. G., and Augsburger, L. L. (1998), Development of metoprolol tartrate extended-release matrix tablet formulations for regulatory policy consideration, *J. Controlled Release*, 50, 247–256.
106. Liu, C. H., Kao, Y. H., Chen, S. C., Sokoloski, T. D., and Sheu, M. T. (1995), In-vitro and in-vivo studies of the diclofenac sodium controlled-release matrix tablets, *J. Pharm. Pharmacol.*, 47, 360–364.
107. Dahl, T. C., Calderwood, T., Bormeth, A., Trimble, K., and Piepmeier, E. (1990), Influence of physicochemical properties of hydroxypropyl methyl cellulose on naproxen release from sustained release matrix tablets, *J. Controlled Release*, 14, 1–10.
108. Shah, N. H., Railkar, A. S., Phuapradit, W., Zeng, F. W., Chen, A., Infeld, M. H., and Malick, A. W. (1996), Effect of processing techniques in controlling the release rate and mechanical strength of hydroxypropyl methyl cellulose based hydrogel matrixes, *Eur. J. Pharm. Biopharm.*, 42, 183–187.
109. Sheskey, P. J., and Keary, C. M. (2002), *Aqueous air foams*, U.S. Patent 7011702.
110. Keary, C. M., and Sheskey, P. J. (2004), Preliminary report of the discovery of a new pharmaceutical granulation process using foamed aqueous binders, *Drug Dev. Ind. Pharm.*, 30, 831–845.
111. Cao, Q-R., Choi, Y-W., Cui, J-H., and Lee, B-J. (2005), Effect of solvents on physical properties and release characteristics of monolithic hydroxypropylmethylcellulose matrix granules and tablets, *Arch. Pharm. Res.*, 28, 493–501.
112. Sheskey, P. J., and Williams, D. M. (1996), Comparison of low-shear and high-shear wet granulation techniques and the influence of percent water addition in the preparation of a controlled-release matrix tablet containing HPMC and a high-dose, highly water-soluble drug, *Pharm. Technol.*, 20, 80, 82, 84, 86, 88, 90, 92.
113. Zhang, F., and McGinity, J. W. (1999), Properties of sustained-release tablets prepared by hot-melt extrusion, *Pharm. Dev. Technol.*, 4, 241–250.
114. Urbano, A. P. A., Ribeiro, A. J., and Veiga, F. (2006), Design of pectin beads for oral protein delivery, *Chem. Ind. Chem. Eng. Q.*, 12, 24–30.
115. Fukuda, M., Peppas, N. A., and McGinity, J. W. (2006), Properties of sustained release hot-melt extruded tablets containing chitosan and xanthan gum, *Int. J. Pharm.*, 310, 90–100.
116. De Brabander, C., Vervaet, C., and Remon, J. P. (2003), Development and evaluation of sustained release mini-matrices prepared via hot melt extrusion, *J. Controlled Release*, 89, 235–247.
117. Baumgartner, S., Slamersek, V., and Kristl, J. (2003), Controlled drug delivery of hydrophilic drugs from cellulose ether matrix tablets: The influence of the drug molecule size on its release mechanism and kinetics, *Farm. Vestnik (Ljubljana, Slov.)*, 54, 359–360.
118. Genc, L., Bilac, H., and Guler, E. (1999), Studies on controlled release dimenhydrinate from matrix tablet formulations, *Pharm. Acta Helv.*, 74, 43–49.
119. Sinha Roy, D., and Rohera Bhagwan, D. (2002), Comparative evaluation of rate of hydration and matrix erosion of HEC and HPC and study of drug release from their matrices, *Eur. J. Pharm. Sci.*, 16, 193–199.
120. Vueba, M. L., Batista de Carvalho, L. A. E., Veiga, F., Sousa, J. J., and Pina, M. E. (2006), Influence of cellulose ether mixtures on ibuprofen release: MC25, HPC and HPMC K100M, *Pharm. Dev. Technol.*, 11, 213–228.
121. Vueba, M. L., Batista De Carvalho, L. A. E., Veiga, F., Sousa, J. J., and Pina, M. E. (2004), Influence of cellulose ether polymers on ketoprofen release from hydrophilic matrix tablets, *Eur. J. Pharm. Biopharm.*, 58, 51–59.

122. Talukdar, M. M., Michoel, A., Rombaut, P., and Kinget, R. (1996), Comparative study on xanthan gum and hydroxypropyl methyl cellulose as matrixes for controlled-release drug delivery I. Compaction and in vitro drug release behavior, *Int. J. Pharm.*, 129, 233–241.
123. Talukdar, M. M., and Kinget, R. (1997), Comparative study on xanthan gum and hydroxypropyl methyl cellulose as matrixes for controlled-release drug delivery. II. Drug diffusion in hydrated matrixes, *Int. J. Pharm.*, 151, 99–107.
124. Andreopoulos, A. G., and Tarantili, P. A. (2002), Study of biopolymers as carriers for controlled release, *J. Macromol. Sci. Phys.*, B41, 559–578.
125. Andreopoulos, A. G., and Tarantili, P. A. (2001), Xanthan gum as a carrier for controlled release of drugs, *J. Biomater. Appl.*, 16, 34–46.
126. Talukdar, M. M., Rombaut, P., and Kinget, R. (1998), The release mechanism of an oral controlled-release delivery system for indomethacin, *Pharm. Dev. Technol.*, 3, 1–6.
127. Talukdar, M. M., and Kinget, R. (1995), Swelling and drug release behavior of xanthan gum matrix tablets, *Int. J. Pharm.*, 120, 63–72.
128. Maggi, L., Segale, L., Torre, M. L., Ochoa Machiste, E., and Conte, U. (2002), Dissolution behaviour of hydrophilic matrix tablets containing two different polyethylene oxides (PEOs) for the controlled release of a water-soluble drug. Dimensionality study, *Biomaterials*, 23, 1113–1119.
129. Choulis, N., and Papadopoulos, H. (1975), Timed-release tablets containing quinine sulfate, *J. Pharm. Sci.*, 64, 1033–1035.
130. Huang, L-L., and Schwartz, J. B. (1995), Studies on drug release from a carbomer tablet matrix, *Drug Dev. Ind. Pharm.*, 21, 1487–1501.
131. El-Sayed, G. M., El-Said, Y., Meshali, M. M., and Schwartz, J. B. (1996), Kinetics of theophylline release from different tablet matrixes, *STP Pharma Sci.*, 6, 390–397.
132. Patel, G. N., Patel, G. C., Patel, R. B., Patel, S. S., Patel, J. K., Bharadia, P. D., and Patel, M. M. (2006), Oral colon-specific drug delivery: An overview, *Drug Deliv. Technol.*, 6, 62–71.
133. Liu, L. S., Fishman, M. L., Kost, J., and Hicks, K. B. (2003), Pectin-based systems for colon-specific drug delivery via oral route, *Biomaterials*, 24, 3333–3343.
134. Liew, C. V., Chan, L. W., Ching, A. L., and Heng, P. W. S. (2006), Evaluation of sodium alginate as drug release modifier in matrix tablets, *Int. J. Pharm.*, 309, 25–37.
135. Shilpa, A., Agrawal, S. S., and Ray, A. R. (2003), Controlled delivery of drugs from alginate matrix, *J. Macromol. Sci. Polym. Rev.*, C43, 187–221.
136. Ostberg, T., and Graffner, C. (1994), Calcium alginate matrixes for oral multiple unit administration: III. Influence of calcium concentration, amount of drug added and alginate characteristics on drug release, *Int. J. Pharm.*, 111, 271–282.
137. Varma, M. V. S., Kaushal, A. M., Garg, A., and Garg, S. (2004), Factors affecting mechanism and kinetics of drug release from matrix-based oral controlled drug delivery systems, *Am. J. Drug Deliv.*, 2, 43–57.
138. McCall, T. W., Baichwal, A. R., and Staniforth, J. N. (2003), TIMERx oral controlled-release drug delivery system, *Drugs Pharm. Sci.*, 126, 11–19.
139. Friend, D. R. (2005), New oral delivery systems for treatment of inflammatory bowel disease, *Adv. Drug Deliv. Rev.*, 57, 247–265.
140. Ghandehari, H., Kopeckova, P., and Kopecek, J. (1997), In vitro degradation of pH-sensitive hydrogels containing aromatic azo bonds, *Biomaterials*, 18, 861–872.
141. Akala, E. O., Kopeckova, P., and Kopecek, J. (1998), Novel pH-sensitive hydrogels with adjustable swelling kinetics, *Biomaterials*, 19, 1037–1047.

142. Takka, S., Rajbhandari, S., and Sakr, A. (2001), Effect of anionic polymers on the release of propranolol hydrochloride from matrix tablets, *Eur. J. Pharm. Biopharm.*, 52, 75–82.
143. Nie, S., Pan, W., Li, X., and Wu, X. (2004), The effect of citric acid added to hydroxypropyl methylcellulose (HPMC) matrix tablets on the release profile of vinpocetine, *Drug Dev. Ind. Pharm.*, 30, 627–635.
144. Varma, M. V. S., Kaushal, A. M., and Garg, S. (2005), Influence of micro-environmental pH on the gel layer behavior and release of a basic drug from various hydrophilic matrices, *J. Controlled Release*, 103, 499–510.
145. Qiu, Y., and Park, K. (2001), Environment-sensitive hydrogels for drug delivery, *Adv. Drug Deliv. Rev.*, 53, 321–339.
146. Bromberg, L. (2005), Intelligent hydrogels for the oral delivery of chemotherapeutics, *Exp. Opin. Drug Deliv.*, 2, 1003–1013.
147. Rowe, C. W., Wang, C-C., and Monkhouse, D. C. (2003), TheriForm technology, *Drugs Pharm. Sci.*, 126, 77–87.
148. Lee, K-J., Kang, A., Delfino, J. J., West, T. G., Chetty, D., Monkhouse, D. C., and Yoo, J. (2003), Evaluation of critical formulation factors in the development of a rapidly dispersing captopril oral dosage form, *Drug Dev. Ind. Pharm.*, 29, 967–979.
149. Rocca, J. G., and Park, K. (2004), Oral drug delivery: Prospects & challenges, *Drug Deliv. Technol.*, 4, 52–54, 57.

# CYCLODEXTRIN-BASED NANOMATERIALS IN PHARMACEUTICAL FIELD

Erem Bilensoy and A. Atilla Hincal
*Hacettepe University Faculty of Pharmacy, Ankara, Turkey*

**Contents**

1 Introduction
  1.1 Cyclodextrins in Pharmaceutical Field
2 Application of Cyclodextrins in Nanoparticles
  2.1 Incorporation of Drug–Cyclodextrin Complexes in Nanoparticulate Delivery Systems
  2.2 Preparation of Nanoparticles from Cyclodextrins
  2.3 Efficacy and Safety of Amphiphilic Cyclodextrin Nanoparticles
3 Conclusion
  References

## 1 INTRODUCTION

Cyclodextrins (CDs) have a wide range of application in the pharmaceutical field due to their unique structure, which allows them to include hydrophobic molecules in their apolar cavity and to mask the physicochemical properties of the included molecule. This results in the enhancement of drug bioavailability by improving aqueous solubility and the physical and chemical stability of the drug, masking undesired side effects such as irritation, taste, or odor, and overcoming compatibility problems or interactions between drugs and excipients.

Parallel to the increasing interest and successful licensing and commercialization of nanoparticulate pharmaceutical products, CDs have also been incorporated into nanoparticulate drug delivery systems for several purposes. This can be

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

achieved by two approaches: (1) complexation of active ingredient with an appropriate CD and entrapment into polymeric nanoparticles to solve problems arising from the drug's physicochemical properties or (2) modification of CDs to render an amphiphilic character to these molecules, which allows CDs to self-align into nanoparticles in the form of nanospheres, nanocapsules, solid lipid nanoparticles, nanosize liposomes, and nanosize vesicles with or without the presence of surface-active agents.

In light of current research, this chapter will deal with the following issues concerning the use of CDs and derivatives as nanomaterials for drug delivery: use of CDs (natural and synthetic) derivatives in the pharmaceutical field and application of CDs in nanoparticulate drug delivery systems. A major part of the chapter will be focused on new CD derivatives, amphiphilic CDs, and the characterization, efficacy, and safety of nanoparticles prepared from amphiphilic CDs.

## 1.1 Cyclodextrins in Pharmaceutical Field

***Natural Cyclodextrins*** Cyclodextrins are cyclic oligosaccharides obtained by the enzymatic degradation of starch. Major natural CDs are crystalline, homogeneous, nonhygroscopic substances which have a toruslike macroring shape built up from glucopyranose units, as seen in Figure 1 [1–3]. Cyclodextrins are named depending on the number of glucopyranose units. Major industrially produced CDs are named as follows; α-CD, possessing six units, β-CD, possessing seven units; and γ-CD, possessing eight units.

Natural CDs have been demonstrated to have a special structure; that is, glucose residues in the CD ring possess the thermodynamically favored $^4C_1$ chair conformation because all substituent groups are in equatorial position. Cyclodextrins behave like rigid compounds with two degrees of freedom: rotation at the glucosidic links C(4)–O(4) and C(1)–O(4) and rotations at the O(6) primary hydroxyl groups at the C(5)–C(6) band. As a consequence of this chair conformation, all secondary hydroxyl groups at C(2) and C(3) are located at the broader side of the CD torus in the equatorial position. Hydroxyl groups on C(2) point toward the cavity and hydroxyl groups on C(3) point outward. The primary hydroxyl groups at the C6 position are located at the narrower side of the torus. These hydroxyl groups ensure good water solubility for the natural CDs. The cavity of the torus is lined with a ring of C—H groups (C3), a ring of glucosidic oxygen atoms, and another ring of C—H groups (C5). Thus, the cavity of CDs exhibits an apolar character. This is accompanied by a high electron density and Lewis base property. The physicochemical characteristics and inclusion behavior of CDs are a direct consequence of these special binding conditions [4, 5]. Certain physicochemical characteristics of natural CDs are given in Table 1.

***Cyclodextrin Derivatives*** Natural CDs were reported to form total or partial inclusion complexes with several drugs to improve the aqueous solubility and stability under physiological or ambient conditions, reduce or mask completely the side effects associated with the included drug, and increase compatibility of the drug with other drugs in the formulation or excipients while improving patient compliance by masking the taste or odor of the active ingredient [6–8].

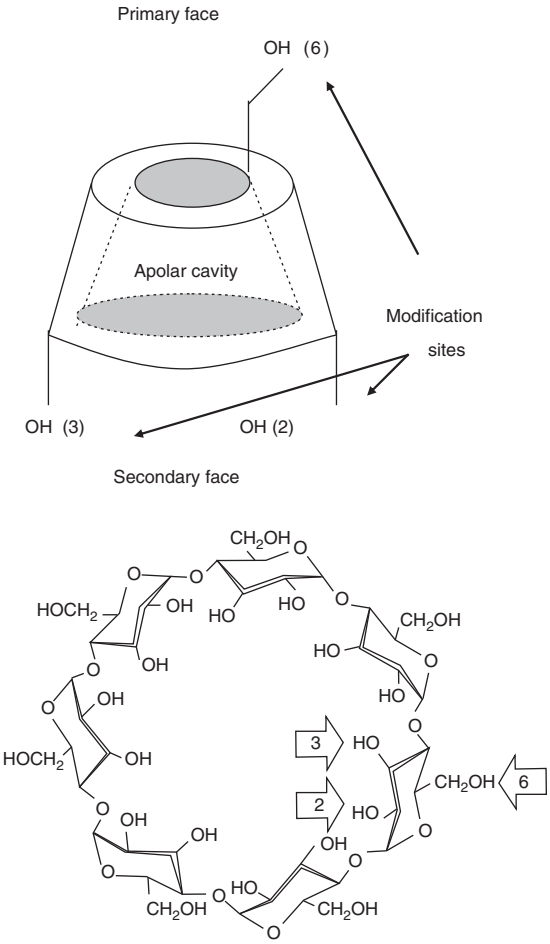


**FIGURE 1** Schematic representation of natural CD structure and modification sites.

**TABLE 1 Some Physicochemical Characteristics of Natural Cyclodextrins**

| Characteristics | α-CD | β-CD | γ-CD |
|---|---|---|---|
| Number of glucose units | 6 | 7 | 8 |
| Molecular weight, g/mol | 972 | 1135 | 1297 |
| Internal diameter, Å | 4.7–5.2 | 6.0–6.4 | 7.5–8.3 |
| External diameter, Å | 14.2–15.0 | 15.0–15.8 | 17.1–17.9 |
| Depth, Å | 7.9–8.0 | 7.9–8.0 | 7.9–8.0 |
| Solubility in water (25°C), g/L | 145 | 18.5 | 232 |
| Crystal water, w % | 10.2 | 13.2–14.5 | 8.13–17.7 |
| Approximative cavity volume in 1 mol CD, $Å^3$ | 174 | 262 | 472 |
| Melting point, °C | 250–260 | 255–265 | 240–245 |
| Half-life in 1 *M* HCl at 60°C, h | 6.2 | 5.4 | 3.0 |
| Crystal forms (from water) | Hexagonal plate | Monoclinic parallelogram | Quadratic prism |
| Partial molar volumes in solution, mL/mol | 611.4 | 703.8 | 801.2 |

In contrast to the advantageous nature of CDs for molecular inclusion, their surface makes it more difficult for the highly hydrophilic CD molecule to interact with lipophilic biological membranes. For this reason, natural CDs have been chemically modified to alter their water solubility, interaction with biological membranes, and drug release properties.

Two of the natural CDs are known to be parenterally unsafe due to nephrotoxic effects [9]. The etiology of the nephrotoxicity of α- and β-CD is unknown but is believed to be related to either CD uptake by kidney tubule cells resulting in disruption of intracellular function or the extraction of lipid membrane components by the CDs. The latter is suggested to be of validity since there seems to be a linear correlation between the ability of some CDs to disrupt cellular membranes and kidney nephrotoxicity [2, 6]. The ability of CDs to cause red blood cell hemolysis and membrane irritation seems also to correlate with their ability to extract lipid membrane components: cholesterol and phospholipids [10, 11].

Modification of natural CDs has been the aim of many research groups to improve safety while maintaining the ability to form inclusion complexes with various substrates. Some groups have also focused on improving the interaction between the pharmaceuticals and the CDs while others have attempted to prepare materials that can be chemically defined more precisely.

Methylated CDs are obtained by methylation of CDs on either all C2 secondary and C6 primary hydroxyl groups [dimethyl cyclodextrins (DIMEB)] or all the hydroxyl groups C2, C3, and C6 [trimethyl cyclodextrins (TRIMEB)]. Major disadvantage of methylated CDs is that their solubility decreases with increasing temperature and they are reported to be hemolytic [12]. This is a result of partial methylation of the hydroxyls of β-CD which leads to stronger drug binding but also to stronger hemolysis [12].

Hydroxypropylated CDs are statistically substituted derivatives because hydroxypropylation does not result in selective substitution as with methylation. While the reaction proceeds, the reactivity of the hydroxyl group changes, and this results in a mixture of products with various degrees of substitution. Their dissolution is endothermic so there is no decrease in solubility with increasing temperature [6, 13]. It is necessary to note that degree of substitution in hydroxypropylated CDs is inversely correlated with their inclusion capability [14, 15]. Hydroxyalkylated CDs are commercially available as tablets, ocular collyrs [16], and excipients under the trademarks Encapsin and Molecusol.

Sulfobutylether-β-cyclodextrins (SBE-β-CDs) are water soluble and parenterally safe. An advantage over hydroxypropylated-β-CDs (HP-β-CDs) is that higher sulfobutyl group substitution often results in higher drug complexation ability [17]. Inability of the SBE-β-CDs, especially the commercially available product $(SBE)_{7M}$-β-CD (Captisol), to form strong complexes with cholesterol and other membrane lipids, arising from their polyanionic nature causing Coulombic repulsions, results in a little or no membrane disruption [6, 13] and reduced hemolysis [18]. Captisol is used in parenteral and ocular systems as well as osmotic tablets and also as a freeze-drying excipient [19].

Branched CDs (mono- or di-glucosyl, maltosyl and glucopyranosyl α- and β-CDs) are more resistant to α-amylase than natural CDs. Natural β-CD and monoglucosyl-β-CD are stable in rat blood because they have no linear glycosidic bond. Branched CDs exert a lower hemolytic activity on human erythrocytes and are

weaker than natural CDs. Their inclusion capability is more or less drug dependent. Steroids were reported to show a slightly higher affinity to branched CDs than to natural CDs [19].

The water solubility of acylated CDs decreases proportionally to their degree of substitution, whereas their solubility in less polar solvents such as ethanol increases. Acylated CDs of hydrophobic nature are useful for controlling the release rates of water-soluble drugs [20].

Ionizable β-CDs possess interesting solubility properties, too. Solubility in water for these derivatives depends on the pH of the solution. Carboxymethylethyl-β-cyclodextrin (CME-β-CD) is prepared as an enteric-type carrier system. The presence of a carboxymethyl group causes a pH-dependent solubility range in water, meaning that it is only slightly soluble in the low-pH region, but freely soluble in neutral and alkaline regions due to the ionization of the carboxyl group ($pK_a$ about 3–4) [21]. Inclusion-forming capability of CME-β-CD is dependent on drug properties, including size, shape, and charge of the molecule.

Sulfated CDs are of anionic nature and are interesting from chemical and biological points of view because of their angiogenic and antiviral properties [22, 23]. Sulfated derivatives are also reported to have a heparin-like activity, resulting in increase in blood-clotting times, which limits their injectable dose [7].

Low-molecular-weight CD polymers (MW 3000–6000 Da) are soluble in water whereas high-molecular-weight CD polymers (MW > 10,000 Da) can only swell in water and form insoluble gels [24, 25]. Insoluble cross-linked bead polymers seem to be applicable as wound-healing agents for the treatment of wounds like burns or ulcers. Iodine has been complexed with such a CD polymer as an antiseptic wound healing agent [24].

## 2 APPLICATIONS OF CYCLODEXTRINS IN NANOPARTICLES

### 2.1 Incorporation of Drug–Cyclodextrin Complexes in Nanoparticulate Delivery Systems

Nanoparticles are of pharmaceutical interest due to their active and passive targeting properties and their ability to deliver poorly solube drugs and drugs with stability problems. Nanoparticles are considered more stable than liposomal delivery systems. However, a major drawback is associated with the drug-loading capacity of polymeric nanoparticles. Classical emulsion polymerization procedures result in considerably low drug-loading capacities. This results in the administration of excessive quantities of polymeric material which may impair the safety of the drug delivery system [26, 27].

For this reason, different techniques are used to improve the drug-loading properties of polymeric nanoparticles. Cyclodextrins are used for this reason to improve water solubility and sometimes the hydrolytic or photolytic stability of drugs for better loading properties [8]. Drug–CD complexes act to solubilize or stabilize active ingredients within the nanoparticles, resulting in increased drug concentration in the polymerization medium and increased hydrophobic sites in the nanosphere structure when large amounts of CDs are associated to the nanoparticles [27, 28].

The antiviral agent saguinavir was complexed to HP-β-CD to increase saquinavir loading into polyalkylcyanoacrylate nanoparticles by providing a soluble drug reservoir in the polymerization medium that is the basis of nanoparticle formation [29]. A dynamic equilibrium was observed between the complex, dissociated species of the complex, and the forming polymeric nanoparticle. During nanoparticle formation, free drug was believed to be progressively incorporated into a polymer network, driven by the drug partition coefficient between the polymer and polymerization medium. Simultaneous direct entrapment of some CD–drug complex was also suspected [28–30].

Incorporation of the steroidal drugs hydrocortisone and progesterone in complex with β-CD and HP-β-CD reduced the particle size for solid lipid nanoparticles (SLNs) below 100 nm. Steroids were demonstrated to be dispersed in the amorphous state. Compexation to CDs resulted in higher drug-loading properties for the more hydrophobic drug hydrocortisone and lower in vitro release for both drugs when they are complexed to CDs rather than their free form [31].

The in vivo behavior of nanoparticles obtained from drug–CD complexes was also evaluated. HP-β-CD addition in the polymerization medium of polyethylcyanoacrylate (PECA) nanospheres improved the subcutaneous absorption of metoclopramide in rats. PECA nanospheres with HP-β-CD provided the highest drug concentration and enhanced drug absorption compared with those with dextran or with drug solution. However, in addition to drug absorption from subcutaneous sites, HP-β-CD also enhanced the drug elimination by enhancing the drug absorption to reticuloendothelial tissues [32].

Progesterone complexed to HP-β-CD or DM-β-CD was loaded into bovine serum albumin (BSA) nanospheres. Dissolution rates of progesterone were significantly enhanced by complexation to CDs with respect to free drug. Nanospheres of 100 nm loaded with drug–CD complexes provided a pH-dependent release profile and good stability in an aqueous neutral environment [33].

In another approach, CD properties of complexation were combined with those of chitosan. Complexation with CD was believed to permit solubilization as well as protection for labile drugs while entrapment in the chitosan network was expected to facilitate absorption. Chitosan nanoparticles, including complexes of HP-β-CD with the hydrophobic model drugs triclosan and furosemide, were prepared by ionic cross-linking of chitosan with sodium tripolyphosphate (TPP) in the presence of CDs. Nanoparticles were then prepared by ionotropic gelation using the obtained drug–HP-β-CD inclusion complexes and chitosan. Cyclodextrin and TPP concentration largely affected particle size but the zeta potential remained unchanged with different parameters. On the other hand, drug entrapment increased up to 4 and 10 times by triclosan and furosemide, respectively. The release profile of nanoparticles indicated an initial burst release followed by a delayed release profile lasting up to 4 h [34].

Recently a CD–insulin complex was encapsulated in polymethacrylic acid–chitosan–polyether[polyethylene glycol (PEG)–propylene glycol] copolymer PMCP nanoparticles from the free-radical polymerization of methacrylic acid in the presence of chitosan and polyether in a medium free of solvents or surfactants. Particles had a size distribution of 500–800 nm. The HP-ß-CD inclusion complex with insulin was encapsulated into the nanoparticles, resulting in a pH-dependent release profile as seen in Figure 2. The biological activity of insulin was demonstrated with enzyme-

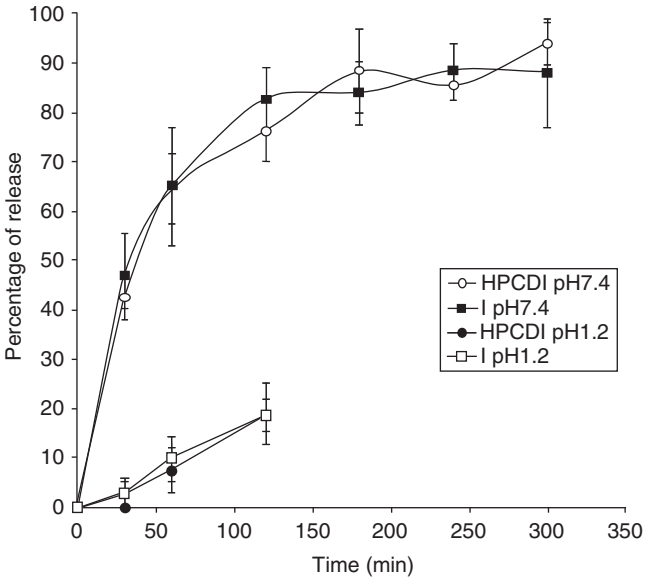


**FIGURE 2** pH-dependent release profile for insulin complexed to HP-ß-CD and encapsulated in nanoparticles. (Reprinted from S. Sajeesh and C. P. Sharma, *International Journal of Pharmaceutics*, 325, 147–154, 2006, Copyright 2006, with permission from Elsevier.)

linked immunosorbent assay (ELISA). Cyclodextrin complexed to insulin encapsulated into mucoadhesive nanoparticles was believed to be a promising candidate for oral insulin delivery [35].

## 2.2 Preparation of Nanoparticles from Cyclodextrins

Cyclodextrins are used as excipients in the preparation of nanoparticles by three approaches:

1. Preparing nanoparticles from polymers under the presence of CDs in the medium
2. Preparing nanoparticles from polymers incorporating or conjugated to CDs
3. Preparing nanoparticles directly from amphiphilic CDs

Nanoparticles consisting copolymers of aminoethylcarbamoyl-β-cyclodextrin (AEC-β-CD) and ethylene glycol diglycidyl ether (EGDGE) were prepared by an interfacial polyaddition reaction in a mini–emulsion system. By combining these two technologies, namely, cross-linking and modification of hydroxyl groups, a novel functional nanoparticle based on β-CD was introduced as a novel material of nanobiotechnology [36].

Nanoparticles prepared fromn CDs are promising targeted delivery systems. Transferrin, an iron-binding glycoprotein ligand for tumor targeting, was conjugated to β-CD polymers and adamantane–PEG5000 through carbohydrate groups self-assembled into sub-100-nm particles in a recent study [37]. A CD-containing polyca-

tion was used for nucleic acid condensation into nanoparticles [38] and the second component, adamantane-terminated modifier for stabilizing nanoparticles to minimize interactions with plasma and to target cell surface receptors, was incorporated in this system. The particles were demonstrated to mediate transferrin-mediated delivery of nucleic acids to cultured cells [37].

Transferrin-containing CD polymer-based nanoparticles were studied as nucleic acid delivery system that can be modified for targeted delivery of small interfering ribonucleic acid (siRNA) to cancer cells. Molecular studies showed that the siRNA CD nanoparticles reduced levels of Ewing's transcript by 80% and inhibited growth of cultured Ewing's tumor cell line. It was also reported that this delivery system indicated a lack of toxicity [39].

A new tadpole-shaped polymer was synthesized via a coupling reaction of PLA onto mono[6-(2-aminoethyl)amino-6-deoxy]-β-cyclodextrin (CDenPLA). A hydrophilic head consisting of the CD group was believed to bind proteins and the PLA tail gave the amphiphilic property [40]. BSA was incorporated into nanoparticles of CDenPLA using both nanoprecipitation and double-emulsion techniques, as can be seen in Figure 3 [40]. A similar process was used to couple PLGA onto amino-β-CD and ethylenediamino-bridged bis(β-CD) to afford amphiphilic conjugates forming nanoparticles with the nanoprecipitation technique. This approach was believed to be promising for protein delivery since BSA structure was unchanged

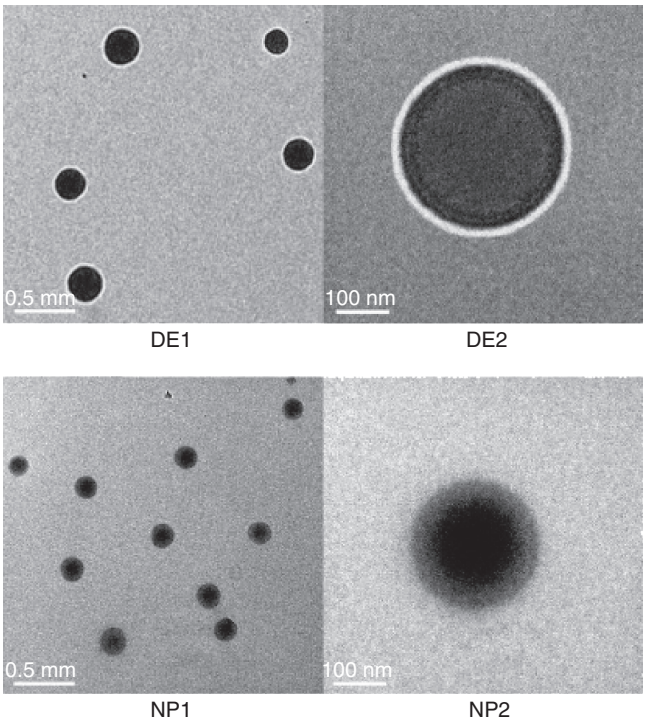


**FIGURE 3** TEM photographs of BSA-loaded ßCDen47PLA nanoparticles prepared with different techniques and their magnified images. (Reprinted from H. Gao, Y. W. Yang, Y. G. Fan, and J. B., Ma, *Journal of Controlled Release*, 112, 301–311, 2006. Copyright 2006, with permission from Elsevier.)

after encapsulation into nanoparticles and during its release. Nanoparticles were reported to be stable after freeze drying [40, 41].

Nanoassemblies were formed by mixing solutions of β-CD polymer and dextran hydrophopbically modified with alkyl chains (C12) and loaded with the model hydrophobic drug benzophenone. Nanoassemblies were characterized as 110–190 nm with relatively low drug-loading values ranging between 0.3 to 1.01% w/w. Authors suggested that hydrophobic model drug and hydrophobically modified dextran compete for the apolar CD cavity; however, benzophenone does not impede the hydrophobic dextran to interact with β-CD polymer to form supramolecular assemblies at the nanoscale [42].

Another group has worked on the oligo(ethylenediamino)-β-cyclodextrin modified gold nanoparticles (OEA-CD-NP) as a vector for DNA binding. Possessing many hydrophobic cavities at the outer space, OEACD-NP was believed to have a capability of carrying biological and/or medicinal substrates into cells. Presence of the CD moieties was suggested to be the key parameter in the high affinity to DNA for the gold nanoparticles. In addition, CD moieties were demonstrated to reduce the cytotoxicity of gold nanoparticles arising from the gold clusters that impair plasma membrane functions and lead to cell death [43].

Nucleic acid delivery was also studies by Park et al. using CD-based nanoparticles prepared from β-CD-modified poly(ethylenimine) (CD-PEI). The inclusion-forming capability of β-CD was used in order to immobilize the nanoparticles on solid surfaces (adamantine-modified self-assembled monolayers). CD-PEI nanoparticles were proposed as delivery systems onto solid surfaces to attain specific and high affinity loading. The interaction is schematized in Figure 4 [44].

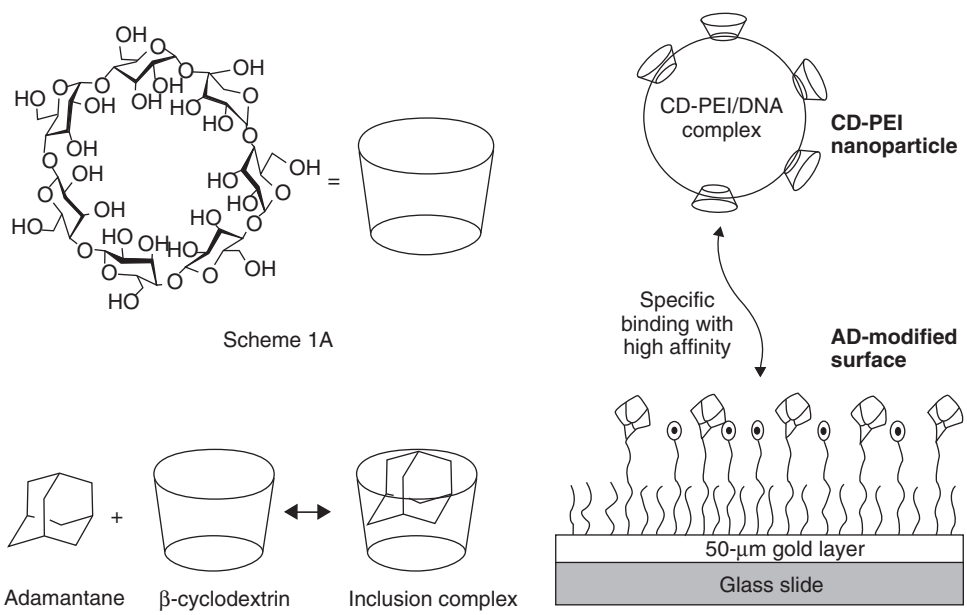


**FIGURE 4** Schematic representation of β-CD–adamantane inclusion complex formation and immobilization of CD nanoparticles on adamantine-modified surface. (Reprinted with permission from ref. 44. Copyright 2006 by the American Chemical Society.)

***Amphiphilic Cyclodextrins*** Nanoparticles have been obtained spontaneously from modified CDs of amphiphilic structure since the last decade. This approach differs from the previously discussed approaches in that amphiphilic CDs do not require the presence of another polymer or macromolecule or even surfactants.

Amphiphilic CDs have been synthesized to solve problems of natural CDs that limit their pharmaceutical applications. The main reasons for the synthesis of amphiphilic CDs are as follows:

1. Enhancement of interaction of CDs with biological membranes through a relative external hydrophobicity
2. Modification or enhancement of interaction of CDs with hydrophobic drugs arising from the high number of long aliphatic chains and by increasing the number of hydrophobic sites for possible interactions with hydrophobic molecules
3. Allowing self-assembly of CDs resulting in the spontaneous formation of nanosize carriers in the form of nanospheres and nanocapsules

The unique advantage of amphiphilic CDs is that they possess self-assembling properties that are sufficient to form nanoparticles spontaneously without the presence of a surfactant as well as the capability of including hydrophobic molecules in their cavity and within the long aliphatic chains [45–47]. Amphiphilic CDs can be classified according to their surface charge.

*Nonionic Amphiphilic Cyclodextrins* Nonionic amphiphilic CDs are obtained by grafting aliphatic chains of different length on the primary and/or secondary face of the CD glucopyranose unit. Different derivatives depicted in Figure 5 are named after their structure:

1. *Lollipop CDs* [48] are obtained by grafting only one aliphatic chain to 6-amino-β-CD.
2. *Cup-and-ball CDs* were synthesized by the introduction of a voluminous group such as the *tert*-butyl group which is linked to the end of the aliphatic chain in order to prevent self-inclusion of the pendant group [49, 50].

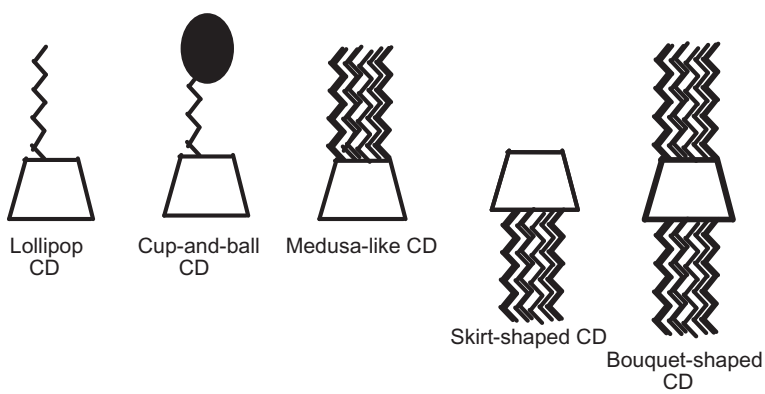


**FIGURE 5** Schematic representation of some nonionic amphiphilic CDs.

3. *Medusa-like CDs* are obtained by grafting aliphatic chains with length between C10 and C16 to all the primary hydroxyls of the CD molecule [51–53].
4. *Skirt-shaped CDs* consist of β- and γ-CDs per-modified with aliphatic esters (C2–C14) on the secondary face [54–57].
5. *Bouquet-shaped CDs* result from the grafting of 14 polymethylene chains to 3-monomethylated β-cyclodextrin, meaning seven chains on each side of the CD ring molecule [58]. Per(2,6-di-*O*-alkyl) CDs where the alkyl chain may be propyl, butyl, pentyl, 3-methylbutyl, or dodecyl also take part in the bouquet family [4].
6. *Cholesteryl CDs* were recently introduced as more complicated derivatives [59]. They have been designed assuming that CD is the hydrophilic head group and cholesterol is the hydrophobic part.

Interfacial properties of nonionic amphiphilic CDs have been demonstrated by different groups [60–62]. It was found that length, structure, and bond type of the aliphatic chain play important roles upon the surface-active characteristics of amphiphilic CDs. Alignment of the amphiphilic CD molecule at the air–water interface was demonstrated to be aliphatic chains perpendicular to and a CD ring parallel to the film [62]. Inclusion-forming capability of nonionic amphiphilic CDs also has been reported with various model molecules of a different nature. It was suggested that leaving the wider side of the cavity, that is, the secondary face, unsubstituted may facilitate entrance of the drug in the cavity of the amphiphilic CD [63–65].

*Cationic Amphiphilic Cyclodextrins* Recently cationic amphiphilic CDs were obtained and characterized carrying an amino group as an ionic group. Heptakis(2-ω-amino-*O*-oligo(ethyleneoxide)-6-hexylthio)-β-CD, a "stealth" cationic amphiphilic CD because of the oligoethylene glycol group it carries, was synthesized [66]. The structural properties of cationic amphiphilic CDs were believed to be due to the balance between hydrophobic tails such as thioalkyl chains and hydrophilic components such as ethylene glycol oligomers. The presence of ethylene glycol chains in particular was believed to increase the colloidal stability of the supramolecular aggregates formed by cationic amphiphilic CDs. Amphiphilic alkylamino-α- and β-CDs were also reported regarding their synthesis and characterization [67].

A series of polyamino-β-CDs have been synthesized by Cryan et al. [68] and complete substitution by amine groups at the 6-position. Neutral CDs have been shown to interact with nucleic acids and nucleotides and to enhance their transfection efficiency in vivo. Cationic CDs have shown even greater ability to bind nucleotides and enhance delivery by viral vectors. The major advantage of polycationic CDs and their nanoparticles is their enhanced ability to interact with nucleic acids combined with their self-organizational properties [68].

*Anionic Amphiphilic Cyclodextrins* Anionic amphiphilic CDs possess a sulfate group in their structure to render an anionic nature to the molecule. An efficient regiospecific synthetic route to obtain acyl-sulfated β-CDs was introduced in which the upper rim is functionalized with sulfates and the lower rim with fatty acid esters [69]. These derivatives were able to form aggregates in aqueous medium.

Sulfated amphiphilic α-, β-, and γ-CDs were demonstrated to form 1:1 inclusion complexes with the antiviral drug acyclovir. Noncovalent interactions between

acyclovir and nonsulfated amphiphilic CDs (nonionic amphiphilic CDs) appeared to take place both in the cavity of the CD and inside the hydrophobic zone generated by alkanoyl chains. However, in the case of sulfated anionic amphiphilic CDs, the interactions appear to take place only in the hydrophobic region of the alkanoyl chains [70].

Fluorine containing anionic β-CDs were first introduced by Granger et al. [71] functionalized at the 6-position by trifluoromethylthio groups. They exhibit an amphiphilic behavior at the air–water interface and are good candidates for a new class of amphiphilic carriers. Péroche et al. [72] described the synthesis of new amphiphilic perfluorohexyl- and perfluorooctyl-thio-β-CDs and their alkyl analogue, nonanethio-β-CD. The ability of these products to form nanoparticles was also investigated by photon correlation spectroscopy and imaging techniques such as scanning electron microscopy (SEM) and cryo–transmission electron microscopy (TEM).

Fluorophilic CD derivatives have been obtained as a result of combinations of CDs and a linear perfluorocarbon [73]. 2,3-Di-*O*-decafluorooctanoyl-γ-CD was obtained with a protection–deprotection synthetic method and characterized further by thin-layer chromatography (TLC), Fourier transform infrared (FTIR) spectroscopy, differential scanning calorimetry (DSC), elemental analysis, and time-of-flight mass spectrometry (OF-MS).

### 2.3 Safety and Efficacy of Amphiphilic Cyclodextrin Nanoparticles

Amphiphilic CDs yield nanoparticles spontaneously in the form of nanospheres or nanocapsules depending on the preparation technique. Nanoparticles have been manufactured using three different techniques. However, the nanoprecipitation technique is generally preferred since it is a simple technique resulting in unimodal distribution. The general preparation techniques for amphiphilic CD nanoparticles are as follows:

1. Nanoprecipitation [74–76]
2. Emulsion/solvent evaporation [77]
3. Detergent removal [78]

Nanocapsules are also prepared according to the same techniques. Amphiphilic CD and the oil Miglyol or benzyl benzoate are dissolved in suitable organic solvent (acetone, ethanol). The solution is poured into aqueous phase under constant stirring and the nanocapsules form spontaneously. Organic solvent is then evaporated. Resulting nanocapsules vary in size between 100 and 900 nm according to the preparation process and technological parameters [79].

Particle sizes of nanocapsules are mostly affected by the size of the oil droplet formed during the preparation along with the molar concentration and nature of amphiphilic CD. Nanospheres, on the other hand, are not significantly affected by amphiphilic CD concentration and can be formed with very high concentrations of amphiphilic CDs. The modification site of the CD (primary or secondary face) is influential for nanosphere size since modifications on the secondary face result in a larger surface area. The presence and concentration of a surfactant such as Pluronic F68 do not affect the particle size of nanospheres and nanocapsules [80]. Nano-

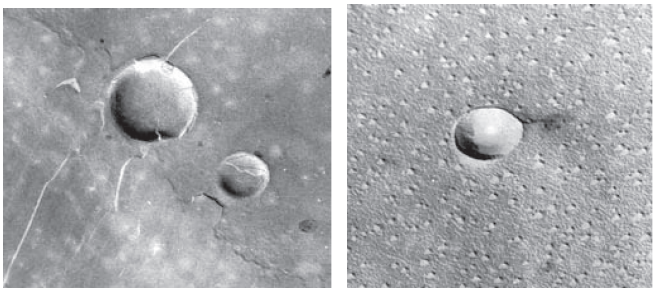

(*a*) (*b*)

**FIGURE 6** Cryo-TEM images of 6-*O*-CAPRO-β-CD nanospheres (*a*) and nanocapsules (*b*). [(*a*) Reprinted from E. Memisoglu, A. Bochot, M. Şen, D. Duchene, and A. A. Hincal, *International Journal of Pharmaceutics*, 252, 143–153, 2003. Copyright 2003 with permission from Elsevier.)

spheres and nanocapsules of amphiphilic CDs were imaged with different microscopic techniques such as cryo-TEM, atomic force microscopy (AFM), and scanning transmission microscopy (STM). SEM imaging, on the other hand, results in shrinkage or disruption of the nanoparticles due to electron bombardment. Figures 6*a* and 6*b* present TEM photomicrographs after freeze fracture for 6-*O*-CAPRO-β-CD nanospheres and nanocapsules, respectively [62, 66].

Drug loading into amphiphilic nanospheres and nanocapsules is governed by the loading technique used. Amphiphilic CD nanoparticles can be loaded with the following techniques:

1. *Conventional Loading* Drug solution is added to the organic phase during preparation.
2. *Preloading* Nanoparticles are prepared directly from preformed drug–amphiphilic CD complexes.
3. *High Loading* Nanoparticles are prepared directly from preformed drug–amphiphilic CD complexes and further loaded by the addition of drug solution in the organic phase.

A high-loading technique results in two- to threefold increase in drug entrapment. Other factors influencing drug loading to amphiphilic CD nanospheres are related to drug physicochemical properties such as drug–CD association constant $k_{1:1}$, representing the affinity of the drug to the CD cavity, oil/water partition coefficient, and aqueous solubility. The affinity of the drug to the CD cavity is correlated with drug-loading capacity. Lipophilic drugs interact both with the CD cavity and the long aliphatic chains situated on either the primary or the secondary face.

Drug release properties of amphiphilic CD nanospheres are affected by various parameters, including drug lipophilicity, drug–CD association constant, and loading technique with release profiles varying from 2 to 96 h depending on the above parameters. Nanocapsules, on the other hand, exert somewhat different drug release profiles that are mostly dependent on lipophilicity and aqueous solubility of the drug. Lipophilicity of the drug is inversely correlated with the rate of release, as seen in Figure 7 [81]. Nevertheless, preparing nanoparticles directly from preformed

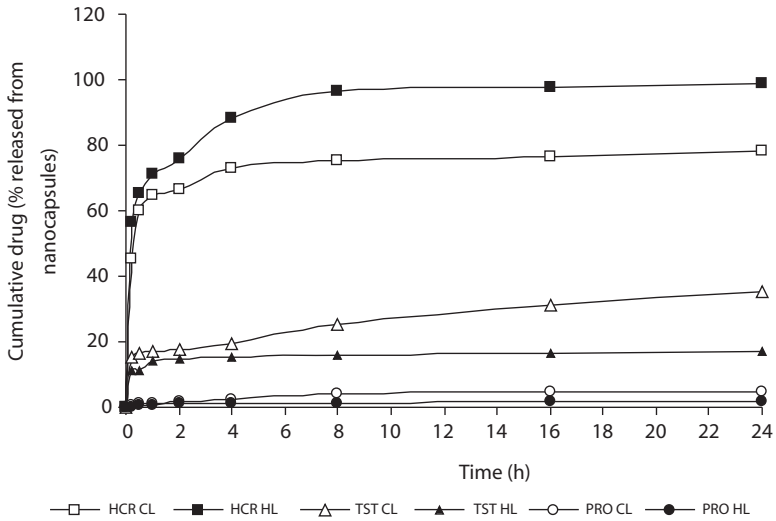


**FIGURE 7** In vitro release profiles of steroids with different physicochemical properties from β-CDC6 nanocapsules (HCR HL, hydrocortisone high loaded, HCR CL, hydrocortisone conventionally loaded; TST HL, testosterone high loaded; TST CL, testosterone conventionally loaded; PRO HL, progesterone high loaded; PRO CL, progesterone conventionally loaded).

inclusion complexes helped reduce the initial burst effect observed in general for nanospheres due to their very large surface area.

***Cancer Therapy*** Nanoparticles were first prepared with the concept of targeting colloidal carriers of nanosize to tumor tissues via the leaky vasculature in tumor regions. Since then nanoparticulate drug carriers have been associated with cancer therapy through passive and active targeting to cancer cells. Thus, amphiphilic CD nanoparticles were mainly focused on cancer therapy and its different aspects.

Tamoxifen, an antiestrogen drug used for the first-line and adjuvant therapy for metastatic breast cancer as long-term chemotherapy, has been incorporated into amphiphilic CD nanoparticles prepared using the amphiphilic CD, β-CDC6 seen in Figure 8 in order to reduce the severe side effects associated with the nonselective cytotoxicity of this drug. Tamoxifen citrate–loaded nanospheres and nanocapsules with approximately 65% entrapment efficiency liberated the drug with a controlled-release profile up to 6 h when the high-loading technique is used [82]. Anticancer efficacy of tamoxifen citrate–loaded nanospheres and nanocapsules was demonstrated to be equivalent to tamoxifen citrate solution in ethanol against MCF-7 human breast cancer cells. Transcription efficiency of the tamoxifen citrate nanocapsules and nanospheres was evaluated against MELN cells in the presence of 17-β-estradiol (E2) for the inhibition of E2-mediated luciferase gene expression. It was found that transcription efficiency of tamoxifen citrate–loaded nanospheres and nanocapsules were concentration dependent [83].

Paclitaxel, an anticancer drug with bioavailability problems arising from its very low aqueous solubility, its tendency to recrystallize when diluted in aqueous media,

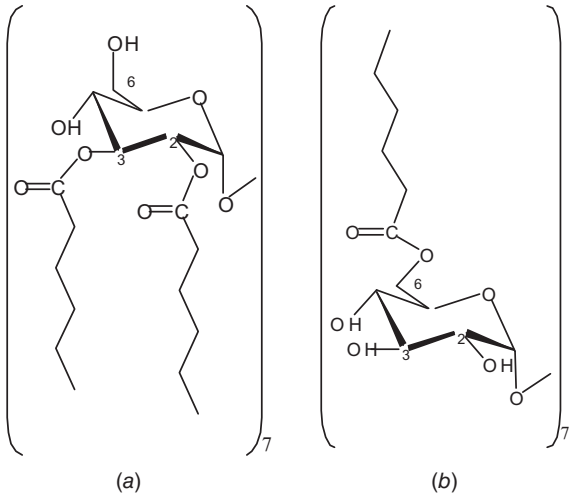


**FIGURE 8** Amphiphilic β-CD derivatives modified with 6C aliphatic esters on (*a*) secondary face, β-CDC6 and (*b*) primary face, 6-*O*-CAPRO-β-CD.

and solubilizers used in its commercially available injectable formulations, has been loaded into nanoparticles prepared from amphiphilic β-CD modified on the primary face with 6C aliphatic esters, 6-*O*-CAPRO-β-CD, seen in Figure 8. Paclitaxel-loaded 6-*O*-CAPRO-β-CD nanospheres and nanocapsules were characterized with a diameter of 150 nm for nanospheres and 500 nm for nanocapsules with high entrapment efficiencies. Blank nanoparticles were proven to be physically stable in aqueous dispersion for 12 months. The in vitro release of paclitaxel from nanoparticles was completed in 24 h [84]. Amphiphilic β-CD nanoparticles were compared to the commercial vehicle Cremophor EL in terms of hemolysis and cytotoxicity. 6-*O*-CAPRO-β-CD nanospheres in particular were found to be significantly less hemolytic than paclitaxel solution in the Cremophor vehicle on human erythrocytes. Cytotoxic effects of blank nanoparticles were assessed against L929 mouse fibroblast cells and a vast difference in cytotoxicity of up to 100-fold reduction was observed for amphiphilic CD nanoparticles.

Drug-loaded nanoparticles were also evaluated for their safety and efficacy. Paclitaxel-encapsulated 6-*O*-CAPRO-β-CD nanospheres and nanocapsules were evaluated for their physical stability in a one-month period in aqueous dispersion form with repeated particle size and zeta potential measurements and AFM imaging to evaluate recrystallization in aqueous medium. Paclitaxel-loaded amphiphilic CD nanoparticles were found to be physically stable for a period of one month whereas recrystallization occurs within minutes when diluted for intravenous (IV) infusion [85]. Finally, paclitaxel-loaded amphiphilic nanoparticles were demonstrated to show similar anticancer efficacy against MCF-7 cells when compared to paclitaxel solution in a cremophor vehicle [85].

Our group is currently working on the formulation of another potent anticancer drug, camptothecin, that is clinically inactive due to its very low water solubility and poor stability under physiological pH, which causes the drug to be converted from its active lactone form to its inactive carboxylate form. Two different amphiphilic

β-CD nanospheres, β-CDC6 and 6-*O*-CAPRO-β-CD, have succeeded in maintaining camptothecin in its active lactone form with considerable loading values and release profiles prolonged up to 96 h [86, 87].

Cationic amphiphilic CDs, heptakis[2-ω-amino-*O*-oligo(ethylene oxide] hexylthio-β-CD nanoparticles, have encapsulated anionic porphyrins (TPPS) by entangling these molecules within the aliphatic chains aligning both faces of the cationic amphiphilic CD. These nanoparticles were demonstrated to preserve the photodynamic properties of the entrapped photoactive agent. The photodynamic efficacy of the carrier/sensitizer nanoparticles was proven by in vitro studies on tumor HeLa cells showing significant cell death upon illumination with visible light [88].

***Oxygen Delivery*** Amphiphilic and fluorophilic β-CD derivatives perfluoro-β-CDs were used to prepare nanocapsules with a single-step nanoprecipitation technique. Highly fluorinated materials have multiple properties, such as repellance to water and oil, unique dielectric, rheological, and optical properties, as well as exceptional chemical and biological inertness. The fluorinated chains, due to their strong hydrophobic and flurorophilic character, impart unique properties to the vesicles, including enhanced particle size stability, prolonged intravascular persistence, and increased drug encapsulation capability. Thus, 2,3-di-*O*-decafluorooctanoyl-β-CD nanoparticles were believed to be a suitable carrier for oxygen solubilization and delivery. Oxygen delivery of perfluorinated amphiphilic CD nanocapsules was compared to water and showed a prolonged delivery of oxygen. Fluorophilic nanocapsules were believed to overcome fluorocarbon emulsions as oxygen carriers due to their higher number of particles in the colloidal solution which will permit a greater rate of dissolved oxygen [73].

***Oral Delivery*** Amphiphilic β-CD nanocapsules loaded with indomethacin have been evaluated in vivo. The nanocapsules have been applied to the rat model. It was reported that the gastrointestinal mucosa of the rat was significantly protected from the ulcerogenic effects of the active ingredient indomethacin in free form. Drug encapsulation yield in the nanocapsules were >98% and the drug content per CD unit was 7.5% w/w [89].

***Cytotoxicity*** The cytotoxicity of nanocapsules was investigated against L929 mouse fibroblast cells and human polymorphonuclear PMNC cells with MTT assay [90]. Cell viability values of different nanocapsule and nanosphere formulations on L929 and PMNC cells indicated that nonsurfactant β-CDC6 nanocapsules were less cytotoxic than nanocapsules containing surfactants. The cytotoxicity of the nanoparticles mostly arises from surfactant presence and was concentration dependent [90].

Nanospheres of β-CDC6 prepared without surfactant and with Pluronic F68 of varying concentrations between 0.1 and 1% were found to be slightly less cytotoxic than nanocapsules to both L929 and human PMNC cells. It was concluded that cytotoxicity increased with increasing concentration of surfactant and the most suitable percentage for surfactant if required was found to be 0.1% [80].

***Sterilizability*** Three different sterilization techniques—autoclaving, filtration, and gamma sterilization—were evaluated for amphiphilic CD nanoparticles of

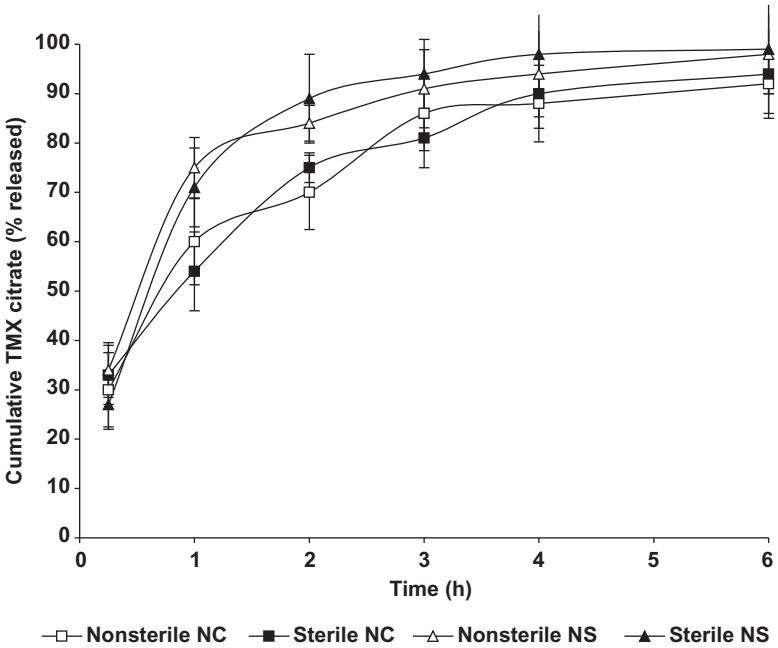


**FIGURE 9** In vitro release profile of tamoxifen from ß-CDC6 nanospheres and nanocapsules before and after gamma sterilization. (Reprinted from E. Memisoglu-Bilensoy and A. A. Hincal, *International Journal of Pharmaceutics*, 311, 203–208, 2006. Copyright 2006 with permission from Elsevier.)

β-CDC6 loaded with the model drug tamoxifen [90]. It was found that filtration was not suitable for injectable amphiphilic CD nanoparticles since nanoparticle sizes were too close to filter pore sizes of 0.22 μm. Autoclaving did not affect the nanoparticle yield but caused a significant increase in particle size and aggregates. Gamma irradiation realized with a dose of 25 kGy was demonstrated to be a suitable sterilization technique since no significant change was observed for mean diameter, zeta potential, drug entrapment efficiency, and in vitro release profiles for nimodipine-loaded β-CDC6 nanospheres and nanocapsules. The in vitro release profile of sterile and nonsterile nanospheres and nanocapsules of ß-CDC6 loaded with nimodipine is seen in Figure 9 [90].

## 3 CONCLUSION

Cyclodextrins have been involved in nanoparticulate drug delivery systems by increasing the solubility of the drug via complex formation, forming nanoparticles in the presence of another polymer/macromolecule, forming nanoparticles by conjugation to polymers, or modification of natural CDs to render this molecule an amphiphilic character. This chapter mainly focused on the potential of amphiphilic CDs as promising carriers for anticancer drugs with bioavailability problems, oxygen delivery for the treatment of ischemia, or the safe oral administration of drugs with gastrointestinal side effects.

Many new studies are available to modify amphiphilic CDs further by giving them "stealth" properties or targeting moieties such as transferring to enable the active targeting of CD-based nanoparticles to tumor tissues. Amphiphilic CD nanocarriers now emerge as promising delivery systems for poorly soluble anticancer drugs, DNA and oligonucleotide delivery, and photodynamic and targeted tumor therapy. These systems are proven to be nonhemolytic and noncytotoxic and are capable of prolonging the release of drugs with different properties.

## ACKNOWLEDGMENTS

The authors wish to thank the TUBITAK Turkish Council of Scientific and Technical Research, projects SBAG-CNRS-3 and SBAG-CD-66, and the Hacettepe University Research Fund, project 0202301005, for financial support of the amphiphilic cyclodextrin research carried out by our group at Hacettepe University, Faculty of Pharmacy, Department of Pharmaceutical Technology.

## REFERENCES

1. Albers, E., and Müller, B. W. (1995), Cyclodextrin derivatives in pharmaceuticals, *CRC Crit. Rev. Ther. Drug Carrier Syst.*, 12(4), 311.
2. Loftsson, T., and Duchene, D. (2007), Cyclodextrins and their pharmaceutical applications, *Int. J. Pharm.*, 329(1–2), 1–11.
3. Uekama, K. (2004), Design and evaluation of cyclodextrin-based drug formulation, *Chem. Pharm. Bull.*, 52(8), 900–915.
4. Wenz, G. (1991), Synthesis and characterization of some lipophilic per(2,6-di-*O*-alkyl) cyclomaltooligosaccharides, *Carbohydr. Res.*, 214, 257–265.
5. Duchene, D., and Wouessidjewe, D. (1992), Industrial uses of cyclodextrins and their derivatives, *J. Coord. Chem.*, 27, 223–236.
6. Thompson, D. O. (1997), Cyclodextrins as enabling excipients: Their present and future use in pharmaceuticals, *CRC Crit. Rev. Ther. Drug Carrier Syst.*, 14(1), 1–110.
7. Rajewski, R. A., and Stella, V. J. (1996), Pharmaceutical applications of cyclodextrins 2. In vivo drug delivery, *J. Pharm. Sci.*, 85(11), 1142–1169.
8. Loftsson, T., and Brewster, M. E. (1996), Pharmaceutical applications of cyclodextrins I. Drug solubilization and stabilization, *J. Pharm. Sci.*, 85(10), 1017–1025.
9. Frank, D. W., Gray, J. E., and Weaver, R. N. (1976), Cyclodextrin nephrosis in the rat, *Am. J. Pathol.*, 83, 367–382.
10. Szejtli, J., Cserhati, T., and Szögyi, M. (1986), Interactions between cyclodextrins and cell-membrane phospholipids, *Carbohydr. Polym.*, 6, 35–49.
11. Debouzy, J. C., Fauvelle, F., Crouzy, S., Girault, L., Chapron, Y., Göschl, M., and Gadelle, A. (1997), Mechanism of α-cyclodextrin induced hemolysis 2. A study of the factors controlling the association with serine-, ethanolamine- and choline-phospholipids, *J. Pharm. Sci.*, 87(1), 59–66.
12. Uekama, K., Hirayama, F., and Irie, T. (1991), Modification of drug release by cyclodextrin derivatives, in Duchene, D., Ed., *New Trends in Cyclodextrins and Derivatives*, Editions de Santé, Paris, pp. 409–446.

13. Stella, V. J., and Rajewski, R. A. (1997), Cyclodextrins: Their future in drug formulation and delivery, *Pharm. Res.*, 14(5), 556–567.
14. Müller, B. W., and Brauns, U. (1986), Hydroxypropyl-β-cyclodextrin derivatives: Influence of average degree of substitution on complexing ability and surface activity, *J. Pharm. Sci.*, 75, 571–572.
15. Yoshida, A., Yamamoto, M., Irie, T., Hirayama, F., and Uekama, K. (1989), Some pharmaceutical properties of 3-hydroxypropyl/ and 2,3-dihydroxypropyl β-cyclodextrins and their solubilizing and stabilizing ability, *Chem. Pharm. Bull.*, 37, 1059–1063.
16. Müller, B. W. (2000), Hydroxypropyl β-cyclodextrin in drug formulation, paper presented at the CRS Workshop What's New in Cyclodextrin Delivery? Paris, July 7–8, Article 3.
17. Zia, V., Rajewski, R., Bornancini, E. R., Luna, E. A., and Stella, V. J. (1997), Effect of alkyl chain length and substitution degree on the complexation of sulfoalkyl ether β-cyclodextrins with steroids, *J. Pharm. Sci.*, 86, 220–224.
18. Stella, V. J. (2000), $SBE_{7M}$-β-CD or Captisol®-possible utilizations, paper presented at the CRS Workshop What's New in Cyclodextrin Delivery? Paris, July 7–8, Article 4.
19. Ma, D. Q., Rajewski, R. A., and Stella, V. J. (1999), Thermal properties and processing of (sulfobutylether)-7M-β-cyclodextrin as a freeze-drying excipient in pharmaceutical formulations, *STP Pharma Sci.*, 9(3), 261–266.
20. Komiyama, M., Yamamoto, Y., and Hirai, H. (1984), Complex formation of modified cyclodextrins with organic salts in organic solvents, *Chem. Lett.*, 1081–1084.
21. Uekama, K., Horiuchi, Y., İrie, T., and Hirayama, F. (1989), *O*-Carboxymethyl-*O*-ethylcyclomaltoheptaose as a delayed-release type drug carrier. Improvement of the oral availability of diltiazem, *Carbohydr. Res.*, 192, 323–330.
22. Uekama, K., Hirayama, F., and Irie, T. (1998), Cyclodextrin drug carrier systems, *Chem. Rev.*, 98, 2045–2076.
23. Anand, R., Nayyar, S., Pitha, J., and Merril, C. R. (1990), Sulphated sugar alpha-cyclodextrin sulphate, a uniquely potent anti-HIV agent, also exhibits marked synergism with AZT, and lymphoproliferative activity, Antiviral Chem. *Chemother.*, 1, 41–46.
24. Friedman, R. (1991), Cyclodextrin-containing polymers, in Duchene, D., Ed., *New Trends in Cyclodextrins and Polymers*, Editions de Santé, Paris, pp. 157–178.
25. Sebillé, B. (1987), Cyclodextrin derivatives, in Duchene, D., Ed., *Cyclodextrins and Their Industrial Uses*, Editions de Santé, Paris, pp. 351–393.
26. Memişoğlu, E., Bochot, A., Şen, M., Duchene, D., and Hincal, A. A. (2003), Non-surfactant nanospheres of progesterone inclusion complexes with amphiphilic ß-cyclodextrins, *Int. J. Pharm.*, 251, 143–153.
27. Monza de Silveira, A., Ponchel, G., Puisieux, F., and Duchene, D. (1998), Combined poly (isobutylcyanoacrylate) and cyclodextrin nanoparticles for enhancing the encapsulation of lipophilic drug, *Pharm. Res.*, 15(7), 1051–1055.
28. Duchene, D., Ponchel, G., and Wouessidjewe, D. (1999), Cyclodextrins in targeting. Application to nanoparticles, *Adv. Drug Deliv. Rev.*, 36, 29–40.
29. Boudad, H., Legrand, P., LeBas, G., Cheron, M., Duchene D., and Ponchel, G. (2001), Combined hydroxypropyl-beta-cyclodextrin and poly(alkylcyanoacrylate) nanoparticles intended for oral administration of saquinavir, *Int. J. Pharm.*, 218, 113–124.
30. Challa, R., Ahuja, A., Ali, J., and Khar, R. K. (2005), Cyclodextrins in drug delivery. An updated review, *AAPS PharmSciTech*, 6(2), E329–E357.
31. Cavalli, R., Peira, E., Caputo, O., and Gasco, M. R. (1999), Solid lipid nanoparticles as carriers of hydrocortisone and progesterone complexes with β-cyclodextrins, *Int. J. Pharm.*, 182, 59–69.

32. Radwan, M. A. (2001), Preparation and in vivo evaluation of parenteral metoclopramide-loaded poly(alkylcyanoacrylate) nanospheres in rats, *J. Microencapsul.*, 18, 467–477.
33. Luppi, B., Cerchiara,T., Bigucci, F., Caponio, D., and Zecchi, V. (2005), Bovine serum albumin nanospheres carrying progesterone inclusion complexes, *Drug Deliv.*, 12, 281–287.
34. Maestrelli, F., Garcia-Fuentes, M., Mura, P., and Alonso, M. J. (2006), A new drug nanocarrier consisting of chitosan and hydoxypropylcyclodextrin, *Eur. J. Pharm. Biopharm.*, 69(2), 79–86.
35. Sajeesh, S., and Sharma, C. P. (2006), Cyclodextrin-insulin complex encapsulated polymethacrylic acid based nanoparticles for oral insulin delivery, *Int. J. Pharm.*, 325, 147–154.
36. Eguchi, M., Da, Y. Z., Ogawa, Y, Okada, T., Yumoto, N., and Kodaka, M. (2006), Effects of conditions for preparing nanoparticles composed of aminoethylcarbamoyl-β-cyclodextrin and ethylene glycol diglycidyl ether on trap efficiency of a guest molecule, *Int. J. Pharm.*, 311, 215–222.
37. Bellocq, N. C., Pun, S. H., Jensen, G. S., and Davis, M. E. (2003), Transferrin-containing, cyclodextrin polymer-based particles for tumor-targeted gene delivery, *Bioconjug. Chem.*, 14, 1122–1132.
38. Hwang, S., Bellocq, N., and Davis, M. (2001), Effects of structure of beta-cyclodextrin-containing polymers on gene delivery, *Bioconjug. Chem.*, 12, 280–290.
39. Hede, K. (2005), Blocking cancer with RNA interference moves toward the clinic, *J. Natl. Cancer Inst.*, 97(4), 626–628.
40. Gao, H., Wang, Y. N. Fan, Y. G., and Ma, J. B. (2006), Conjugates of poly(DL-lactic acid) with ethylenediamino or diethylenetriamino bridged bis(β-cyclodextrin)s and their nanoparticles as protein delivery systems, *J. Controlled Release*, 112, 301–311.
41. Gao, H., Wang, Y. N., Fen, Y. G., and Ma, J. B. (2007), Conjugation of poly(DL-lactide-*co*-glycolide) on amino cyclodextrins and their properties as protein delivery system, *J. Biomed. Mater. Rest. A*, 80A, 111–122.
42. Daoud-Mahammed, S., Ringard-Lefebvre, C., Razzouq, N., Rosilio, V., Gillet, B., Couvreur, P., Amiel, C., and Gref, R. (2007), Spontaneous association of hydrophobized dextran and poly-β-cyclodextrin into nanoassemblies, *J. Colloid Interf. Sci.*, 307(1), 83–93.
43. Wang, H., Chen, Y., Liu, X. Y., and Liu, Y. (2007), Synthesis of oligo(ethylenediamino)-β-cyclodextrin modified gold nanoparticles as a DNA concentrator, *Mol. Pharm.*, 4(2), 189–198
44. Park, I. K., van Recum, H. A., Jiang, S., and Pun, S. H. (2006), Supramolecular assembly of cyclodextrin-based nanoparticles on solid surfaces for gene delivery, *Langmuir*, 22, 8478–8484.
45. Munoz, M., Deschenaux, R., and Coleman, A. W. (1999), Observation of microscopic patterning at the air/water interface by mixtures of amphiphilic cyclodextrins—A comparison isotherm and Brewster angle microscopy study, *J. Phys. Org. Chem.*, 12, 364–369.
46. Duchene, D., Wouessidjewe, D., and Ponchel, G. (1999), Cyclodextrins and carrier systems, *J. Controlled Release*, 62, 263–268.
47. Memisoglu-Bilensoy, E., Bochot, A., Trichard, L., Duchene, D., and Hincal, A. A. (2005), Amphiphilic cyclodextrins and microencapsulation, in Benita, S., Ed, *Microencapsulation*, 2nd rev. ed., Taylor&Francis, New York, pp. 269–295.
48. Bellanger, N., and Perly, B. (1992), NMR investigations of the conformation of new cyclodextrins-based amphiphilic transporters for hydrophobic drugs: Molecular lollipops, *J. Mol. Struct.*, 15, 215–226.

49. Dodziuk, H., Chmurski, K., Jurczak, J., Kozminski, W., Lukin, O., Sitkowski, J., and Stefaniak, J. (2000), A dynamic NMR study of self-inclusion of a pendant group in amphiphilic 6-thiophenyl-6-deoxycyclodextrins, *J Mol. Struct.*, 519, 33–36.
50. Lin, J. (1995), Synthése des cyclodextrins amphiphiles et étude de leur incorporation dans des phases phospholipidiques, Ph.D. dissertation, University of Paris, Paris.
51. Kawabata, Y., Matsumoto, M., Tanaka, M., Takahashi, H., Irinatsu, Y., Tagaki, W., Nakahara, H., and Fukuda, K. (1983), Formation and deposition of monolayers of amphiphilic β-cyclodextrin derivatives, *Chem. Lett.*, 1933–1934.
52. Djedaini, F., Coleman, A. W., and Perly, B. (1990), New cyclodextrins based media for vectorization of hydrophilic drug, mixed vesicles composed of phospholipids and lipophilic cyclodextrins, in Duchene, D., Ed., *Minutes of the 5th International Symposium on Cyclodextrins*, Editions de Santé, Paris, pp. 328–331.
53. Liu, F. Y., Kildsig, D. O., and Mitra, A. K. (1992), Complexation of 6-acyl-*O*-β-cyclodextrin derivatives with steroids, effects of chain length and substitution degree, *Drug Del. Ind. Pharm.*, 18, 1599–1612.
54. Zhang, P., Ling, C. C., Coleman, A. W., Parrot-Lopez, H., and Galons, H. (1991), Formation of amphiphilic cyclodextrins via hydrophobic esterification at the secondary hydroxyl face, *Tetrahedron Lett.*, 32, 2769–2770.
55. Zhang, P., Parrot-Lopez, H., Tchoreloff, P., Basakin, A., Ling, C. C., De Rango, C., and Coleman, A. W. (1992), Self-organizing systems based on amphiphilic cyclodextrins diesters, *J. Phys. Org. Chem.*, 5, 518–528.
56. Memişoğlu, E., Charon, D., Duchene, D., and Hıncal, A. A. (1999), Synthesis of per(2,3-di-*O*-hexanoyl)-β-cyclodextrin and characterization amphiphilic cyclodextrins nanoparticles, in Torres-Labandeira, J., and Vila-Jato, J., Eds., *Proceedings of the 9th International Symposium on Cyclodextrins*, Kluwer Academic, Dordrecht, pp. 622–624.
57. Lesieur, S., Charon, D., Lesieur, P., Ringerd-Lefebvre, C., Muguet, V., Duchene, D., and Wouessidjewe, D. (2000), Phase behaviour of fully-hydrated DMPC-amphiphilic cyclodextrin systems, *Chem. Phys. Lip.*, 10, 127–144.
58. Canceill, J., Jullien, L., Lacombe, L., and Lehn, J. M. (1992), Channel type molecular structures, Part 2, synthesis of bouquet-shaped molecules based on a β-cyclodextrin core, *Helv. Chim. Acta*, 75, 791–812.
59. Auzely-Velty, R., Djedaini-Pilard, F., Desert, S., Perly, B., and Zemb, T. (2000), Micellization of hydrophobically modified cyclodextrins 1. Miceller structures, *Langmuir*, 16, 3727–3734.
60. Tchoreloff, P., Boissonnade, M. M., Coleman, A. W., and Baszkin, A. (1995), Amphiphilic monolayers of insoluble cyclodextrins at the water/air interface. Surface pressure and surface potential studies, *Langmuir*, 11, 191–196.
61. Ringard-Lefebvre, C., Bochot, A., Memişoğlu, E., Charon, D., Duchene, D., and Baszkin, A. (2002), Effect of spread amphiphilic β-cyclodextrins on interfacial properties of the oil/water system, *Coll. Surf. B Biointerf.*, 25, 109–117.
62. Memişoğlu, E., Bochot, A., Şen, M., Charon, D., Duchene, D., and Hincal, A. A. (2002), Amphiphilic β-cyclodextrins modified on the primary face: Synthesis, characterization and evaluation of their potential as novel excipients in the preparation of nanocapsules, *J. Pharm. Sci.*, 95(1), 1214–1224.
63. Bilensoy, E., Doğan, A. L., Şen, M., and Hincal, A. A. (2007), Complexation behaviour of antiestrogen drug tamoxifen citrate with natural and modified cyclodextrins, *J. Inclus. Phenom. Macroc. Chem.*, 57, 651–655.
64. Alexandre, S., Coleman, A. W., Kasselouri, A., and Valleton, J. M. (1996), Scanning force microscopy investigation of amphiphilic cyclodextrin Langmuir-Blodgett films, *Thin Solid Films*, 284–285, 765–768.

65. Tanaka, M., Ishizuka, U., Matsumoto, M., Nakamura, T., Yabe, A., Nakanishi, H., Kawabata, Y., Takahashi, H., Tamura, S., Tagaki, W., Nakahara, H., and Fukuda, K. (1987), Host-guest complexes of an amphiphilic β-cyclodextrin and azobenzene derivatives in Langmuir-Blodgett films, *Chem. Lett.*, 1307–1310.
66. Mazzaglia, A., Angelini, N., Darcy, R., Donohue, R., Lombardo, D., Micali, N., Sciortino, M. T., Villari, V., and Scolaro, L. M. (2003), Novel heterotropic colloids of anionic porphyrins entangled in cationic amphiphilic cyclodextrins: Spectroscopic investigation and intracellular delivery, *Chem. Eur. J.*, 9, 5762–5769.
67. Matsumoto, M., Matsuzawa, Y., Noguchi, S., Sakai, H., and Abe, M. (2004), Structure of Langmuir-Blodgett films of amphiphilic cyclodextrin and water-soluble benzophenone, *Mol. Cryst. Liq. Cryst.*, 425, 197–204.
68. Cryan, S. A., Holohan, A., Donohue, R., Darcy, R., and O'Driscoll, C. M. (2004), Cell transfection with polycationic cyclodextrin vectors, *Eur. J. Pharm. Sci.*, 21, 625–633.
69. Dubes, A., Bouchu, D., Lamartine, R., and Parrot-Lopez, H. (2001), An efficient regiospecific synthetic route to multiply substituted acyl-sulphated ß-cyclodextrins, *Tetrahedron Lett.*, 42, 9147–9151.
70. Dubes, A., Degobert, G., Fessi, H., and Parrot-Lopez, H. (2003), Synthesis and characterisation of sulfated amphiphilic alpha-, beta- and gamma-cyclodextrins: Application to the complexation of acyclovir, *Carbohydr. Res.*, 338, 2185–2193.
71. Granger, C. E., Feliz, C. P., Parrot-Lopez, H., and Langlois, B. R. (2000), Fluorine containing ß-cyclodextrin: A new class of amphiphilic carriers, *Tetrahedron Lett.*, 41, 9257–9260.
72. Péroche, S., Degobert, G., Putaux, J. L., Blanchin, M. G., Fessi, H., and Parrot-Lopez, H. (2005), Synthesis and characterization of novel nanospheres made from amphiphilic perfluoroalkylthio-ß-cyclodextrins, *Eur. J. Pharm. Biopharm.*, 60, 123–131.
73. Skiba, M., Skiba-Lahiani, M., and Arnaud, P. (2002), Design of nanocapsules based on novel fluorophilic cyclodextrin derivatives and their potential role in oxygen delivery, *J. Inclus. Phenom. Macroc. Chem.*, 44, 151–154.
74. Fessi, H. C., Devissaguet, J. P., Puisieux, F., and Thies, C. (1997), Process for the preparation of dispersible colloidal systems of a substance in the form of nanoparticles, U.S. Patent 5,118,528.
75. Wouessidjewe, D., Skiba, M., Leroy-Lechat, F., Lemos-Senna, E., Puisieux, F., and Duchene, D. (1996), A new concept in drug delivery based on "skirt-shaped cyclodextrins aggregates" present state and future prospects, *STP Pharma. Sci.*, 6, 21–26.
76. Lemos-Senna, E., Wouessidjewe, D., Lesieur, S., Puisieux, F., Couarrazze, G., and Duchene, D. (1998), Evaluation of the hydrophobic drug loading characteristics in nanoprecipitated amphiphilic cyclodextrins nanospheres, *Pharm. Dev. Technol.*, 3, 1–10.
77. Lemos-Senna, E., Wouessidjewe, D., Lesieur, S., and Duchene, D. (1998), Preparation of amphiphilic cyclodextrin nanospheres using the emulsion solvent evaporation method, influence of the surfactant on preparation and hydrophobic drug loading, *Int. J. Pharm.*, 170, 119–128.
78. Lemos-Senna, E. (1998), Contribution a l'etude pharmacotechnique et physicochimique de nanospheres de cyclodextrins amphiphiles comme transporteurs de principes actifs, Ph.D. thesis, University of Paris, Paris.
79. Skiba, M., Wouessidjewe, D., Fessi, H., Devissaguet, J. P., Duchene, D., and Puisieux, F. (1992), Preparation et utilizations des nouveau systemes colloidaux dispersibles a base de cyclodextrines sous forme de nanocapsules, French Patent 92-07285.
80. Memisoglu-Bilensoy, E., Doğan, A. L., and Hincal, A. A. (2006), Cytotoxic evaluation of injectable amphiphilic cyclodextrin nanoparticles on fibroblasts and polymorphonuclear cells: Surfactant effect, *J. Pharm. Pharmacol.*, 58(5), 585–589.

81. Memişoğlu-Bilensoy, E., Şen, M., and Hincal, A. A. (2006), Effect of drug physicochemical properties on in vitro characteristics of amphiphilic cyclodextrin nanospheres and nanocapsules, *J. Microencapsul.*, 23(1), 59–68.
82. Memişoğlu-Bilensoy, E., Vural, İ., Renoir, J. M., Bochot, A., Duchene, D., and Hincal, A. A. (2005), Tamoxifen citrate loaded amphiphilic β-cyclodextrin nanoparticles: In vitro characterization and cytotoxicity, *J. Controlled Release*, 104, 489–496.
83. Vural, İ., Memişoğlu-Bilensoy, E., Renoir, J. M., Bochot, A., Duchene, D., and Hincal, A. A. (2005), Transcription efficiency of tamoxifen citrate loaded β-cyclodextrin nanoparticles, *J. Drug Deliv. Sci. Technol.*, 15(5), 339–342.
84. Bilensoy, E., Gürkaynak, O., Ertan, M., Şen, M., and Hincal, A. A., Development of non-surfactant cyclodextrin nanoparticles loaded with anticancer drug paclitaxel, *J. Pharm. Sci.*, DOI: 10.1002/jps.2111, Published online.
85. Bilensoy, E., Gürkaynak, O., Doğan, A. L., Duman, M., and Hıncal, A. A., Safety and efficacy of amphiphilic cyclodextrin nanoparticles for paclitaxel delivery, *Int. J. Pharm.*, DOI: 10.1016/j.ijpharm.2007.06.05.
86. Çırpanlı, Y., Bilensoy, E., Çalış, S., and Hincal, A. A. (2006), Camptothecin inclusion complexes with natural and modified β-cyclodextrins, in Proceedings of the 33rd Annual Meeting and Exhibition on Controlled Release Society, Vienna, July, 22–26 pp. 988–989.
87. Çırpanlı. Y., Bilensoy, E., Çalış, S., and Hincal, A. A. (2007), Development of camptothecin loaded nanoparticles from amphiphilic β-cyclodextrin derivatives, paper presented at the Pharmaceutical Sciences World Congress PSWC, Amsterdam, April, 22–25.
88. Sortino, S., Mazzaglia, A., Scolaro, L. M., Merlo, F. M., Volveri, V., and Sciortino, M. T. (2006), Nanoparticles of cationic amphiphilic cyclodextrins entangling anionic porphyrins as carrier-sensitizer system in photodynamic cancer therapy, *Biomaterials*, 27, 4256–4265.
89. Skiba, M., Morvan, C., Duchene, D., and Puisieux, F., (1995), Evaluation of the gastrointestinal behaviour in the rat of amphiphilic β-cyclodextrin nanocapsules loaded with indomethacin, *Int. J. Pharm.*, 126, 275–279.
90. Memisoglu-Bilensoy, E., and Hincal, A. A. (2006), Sterile injectable cyclodextrin nanoparticles: Effects of gamma irradiation and autoclaving, *Int. J. Pharm.*, 311, 203–208.

# NANOTECHNOLOGY IN PHARMACEUTICAL MANUFACTURING

YIGUANG JIN
*Beijing Institute of Radiation Medicine, Beijing, China*

**Contents**

1 Introduction
2 Nanomaterials
  2.1 Types of Nanomaterials
  2.2 Manufacturing and Processing of Nanomaterials
3 Nanotechnology for Drug Delivery
  3.1 Nanocarriers
  3.2 Nanosuspensions
  3.3 Self-Assembled Drug Nanostructures
4 Nanomedicine
5 Perspective
  References

## 1 INTRODUCTION

Nanotechnology is the ability to produce and process nanosized materials or manipulate objects within the nanoscale. The nanoscale commonly indicates the range from 1 to 100 nm [1]. However, some scientists regard the nanoscale range from 1 to 200 nm [2], even to 1000 nm [3]. Making a comparison with a human hair, it is about 80,000 nm wide. Nanotechnology is a broad, highly interdisciplinary, and still evolving field which involves the production and application of physical, chemical, and biological systems. Nanotechnology is likely to have a profound impact on our economy and society in the early twenty-first century, perhaps comparable to that of information technology or advances in cellular and molecular biology. Science and engineering research in nanotechnology promises breakthroughs in areas such

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

as materials, manufacturing, electronics, medicine, health care, energy, environment, biotechnology, information technology, and national security. It is widely felt that nanotechnology will lead to the next industrial revolution [4].

The idea of nanotechnology was first presented by physicist Richard Feynman. His lecture entitled "Room at the Bottom" in 1959 unveiled the possibilities available in the molecular world. Because bulk matter is built of so many atoms, there is a remarkable amount of space within which to build. Feynman's vision spawned the discipline of nanotechnology, and his dream is now coming true [5]. Along with continually increasing multidisciplinary applications of nanotechnology, many new terms with nanotechnology characteristics appear, for example, *nanomechanics* [6], *nanooptics* [7], *nanoelectronics* [8], *nanochemistry* [9], *nanomedicine* [10], *nanobiotechnology* [5, 11], *nanolithography* [12], *nanoengineering* [13], *nanofabrication* [14], and *nanomanufacturing* [15]. A very broad sense term, *nanoscience* is often used. More and more new words with *nano* as a prefix will be created to fit for the nowadays nanoworld. In fact, applications of nanotechnology in medicine and biotechnology have made great progresses in the recent two decades.

All developed countries including the United States, Japan, and Europe invest a great deal of money in nanotechnology. The National Science Foundation (NSF) of the United States is a leading agency in the national nanotechnology initiative, funding nanotechnology investments at $373 million in 2007, an increase of 8.6% from 2006 and of nearly 150% since 2001 [16]. Developing countries such as China and India also invest a lot in this increasing field so as not to stay far behind developed countries. Cancer therapy and research are hottest applied fields of bionanotechnology. In 2004, the U.S. National Cancer Institute (NCI) launched a $144 million cancer nanotechnology initiative, and the investment increased largely in the following two years [17]. At the same time, investment from public resources or companies is much higher than that from governments.

The application of nanotechnology in pharmacy has a long history, before the prevalence of the nanoconcept. It was well known 50 years ago that very small drug particles have a high solubility in solvents, resulting from the too large surface area when particle size decreased to a very small level, that is, the nanoscale, although this scale had not been mentioned yet. In 1965, Banham created liposomes (lipid vesicles) consisting of phospholipids which had a small size, typically ranging from 10 nm to several micrometers. It was soon found that liposomes were excellent drug carriers, and more importantly they had site-specific distribution capability in vivo depending on their size. It is well known that nanosized liposomes are inclined to deposit in the mononuclear phagocyte system (MPS), including liver, spleen, lung, and marrow. Therefore, nanotechnology was introduced in drug delivery very long ago. Now various nanomaterials are used to deliver drugs, and some nanosystems delivering active agents are available on the market. Undoubtedly, nanotechnology plays a key role in future pharmaceutical development and pharmacotherapy.

## 2 NANOMATERIALS

### 2.1 Types of Nanomaterials

*Nanomaterial* is a general term. Although nanomaterials are defined as solid or liquid materials at the nanoscale, the nanoscale range remains unclear. Many scien-

tists regard materials that are one dimensional and 1 nm to less than 100 nm as nanomaterials. However, some scientists treat larger materials (e.g., less than 200 nm) as nanomaterials [2]. In spite of the different views, nanomaterials show unique characteristics that are different from those of bulk materials. Rapid development of nanotechnology in varied disciplines helps to create various kinds of nanomaterials. In terms of shape differences, nanomaterials can be classified as nanospheres, nanovesicles, nanoshells, nanotubes, nanohorns, nanofibers, nanowires, nanoribbons, nanorods, nanosticks, nanohelices, and so on, and they can appear in any shape imagined. In terms of state differences, nanomaterials can also be classified as nanoparticles with solid cores, nanoemulsions with liquid cores, and nanobubbles with air cores. Images of some nanomaterials of various shapes are shown in Figure 1.

*Nanoparticle* is the most usually used term, having a broad meaning. From a narrow sense, nanoparticles are always used to indicate all ball-like nanomaterials, and therein the term nanosphere is also used. Nanocapsules are core–shell nanoparticles, wherein trapped drugs are gathered in a core coated with a hard shell, though generally nanoparticles have uniformly dispersed drugs within the whole particle. *Nanovesicle* is not a familiar term, for example, liposomes have an inner phase and an outer phase (dispersing medium) that exist together in nanovesicles [18]. In light of drug nature, especially solubility, drugs are entrapped in an inner phase or bilayers (shells). In addition, *nanosuspension* often appears in the pharmaceutical field, meaning drug nanocrystal dispersion in liquid media [19]. Needle-shaped nanocrystals are more common than globe-shaped ones. Nanogels are newly developed based on hydrogels, being similar to nanoparticles after lyophilization [20]. Recently, a special kind of nanomaterial consisting of drugs was created for drug delivery, called self-assembled drug nanostructure (SADN), which is formed by the self-assembly of amphiphilic prodrugs in aqueous media [21, 22].

Some special nanomaterials are of great interest due to their unique properties. Dendrimers are versatile, well-defined, nanosized monodispersing macromolecules which are hyperbranched synthesized polymers constructed by repetitive monomer units. They are perfect nanoarchitectures with size from 1 nm to more than 10 nm depending on the synthesis generation. Drugs can be entrapped into the branches

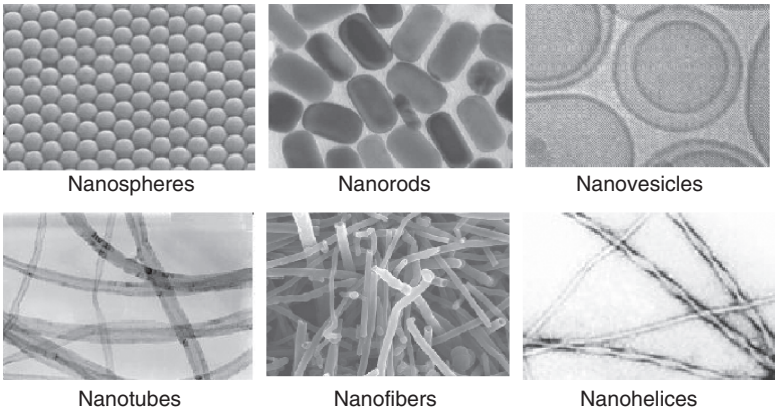


**FIGURE 1** Some typical nanomaterials.

of dendrimers or conjugated with them on the high reactive surfaces [23]. The first fullerene discovered was the buckyball, also known as buckminsterfullerene. It was discovered by Smalley, Curl, and Kroto in 1985 [24], who shared a Nobel Prize in 1996 for the discovery. Buckyball is roughly spherical cages of 60 carbon atoms ($C_{60}$) arranged in interlocking hexagons and pentagons, like the patches on a soccer ball. Fullerenes have attracted considerable research interest, partly because of their unique structures and further because, once suitably dissolved, they display a diverse range of biological activity [25]. Quantum dots (QDs) are semiconductor nanocrystals commonly consisting of CdSe or ZnS. Besides their utilization as electronic materials, QDs have recently been applied to biomedical areas after modification. The new generations of QDs have far-reaching potential for the study of intracellular processes at the single-molecule level, high-resolution cellular imaging, long-term in vivo observation of cell trafficking, tumor targeting, and diagnostics [26].

Although many types of nanomaterials are created continually, the most important and basic issues are nanoscale effects and the subsequent particular functions. Nanomaterials with varied shapes and components provide different platforms to achieve more functions. In the area of pharmaceutical manufacturing, people focus on the drug delivery function of nanomaterials. Furthermore, the rapid development of modern medicine has led to the belief that traditional drug dosage forms such as tablets, capsules, and injections may not treat some vital diseases well, perhaps not at all. Some advanced techniques developed in other disciplines should be considered to apply to medicine. Nanomaterials can load and deliver drugs in vivo as well as display special properties such as high dispersion, adhesive property, and site-specific distribution in vivo. Modified nanomaterials further possess new functions, for example, they may be thermally sensitive, pH sensitive, magnetically sensitive, and ultrasound sensitive.

Nanotechnology has a great effect on pharmaceutical manufacturing. The unique functions of nanomaterials promise considerable benefit to pharmacotherapy over traditional drug preparations. When drug-loaded nanomaterials go through the gastrointestinal tract, high dispersion and adhesion can lead to tight contact of nanomaterials with mucous membranes, enhancing drug absorption. Nanomaterials have been applied in all routes of administration, including oral, injection (intravenous, subcutaneous, intramuscular, intra-articular cavity, and other possible injection sites), intranasal, pulmonary inhalation, conjunctiva, topical, and transdermal, possibly showing various required effects. Some of the characteristics and pharmaceutical applications of nanomaterials are given in Table 1. More applications will continue to be developed.

## 2.2 Manufacturing and Processing of Nanomaterials

When material dimensions reach the nanoscale, quantum mechanical and thermodynamic properties that are insignificant in bulk materials dominate, causing these nanomaterials to display new and interesting properties. The manufacturing and processing of nanomaterials may become difficult due to the unique properties. The very small size of nanomaterials produces a very large surface-to-volume ratio, that is, a great number of molecules/atoms locate on surfaces. High surface energy leads to nanomaterials easily agglomerating to diminish energy unless enough hindrance

**TABLE 1 Characteristics and Applications of Some Nanomaterials in Pharmacy**

| Types of Nanomaterials | Characteristics | Applications in Pharmacy | References |
|---|---|---|---|
| Nanoparticles | Solid nanosized particles consisting of polymers, lipids, or inorganic materials spherically shaped most of the time, entrapped compounds dispersing in the whole particle | Loading all kinds of active agents, including drugs, vaccines, diagnostic agents, and imaging agents for good bioavailability, targeted delivery, and controlled release | 27, 28 |
| Nanocapsules | Core–shell nanoparticles with entrapped compounds gathering in the core | Loading all kinds of active agents for same aims as nanoparticles, possibly protecting entrapped agents | 29 |
| Liposomes | Lipid vesicles with entrapped compounds in inner phase or bilayers depending on physicochemical property | Loading all kinds of active agents for good bioavailability, targeted delivery, and controlled release | 30, 31 |
| Niosomes | Nonionic surfactant vesicles with similar property as liposomes | Loading all kinds of active agents for same aims as for liposomes | 32 |
| Nanoemulsions | Nanoscale emulsions | Loading drugs, as a method to prepare nanoparticles | 33, 34 |
| Polymeric micelles | Micelles consisting of amphiphilic polymers | Loading hydrophobic drugs in the core for solubilization, targeted delivery, and controlled release | 35, 36 |
| Nanogels | Nanosized hydrogels consisting of cross-linked hydrophilic polymers | Loading various compounds for controlled release or targeting | 20 |
| Dendrimers | Well-defined, nanosized, monodispersing macromolecules with hyperbranched structures | Loading all kinds of active agents for good bioavailability, targeted delivery, and controlled release | 23 |

**TABLE 1** ***Continued***

| Types of Nanomaterials | Characteristics | Applications in Pharmacy | References |
|---|---|---|---|
| Carbon nanotubes (CNTs) | Nanosized tubes as if rolling up a single layer of graphite sheet (single-walled CNTs; SWNTs) or by rolling up many layers to form concentric cylinders (multiwalled CNTs; MWNTs) with diameters of ~1 nm and large length–diameter ratio | Linking a wide variety of active molecules with functionalized CNTs | 37 |
| Fullerenes | Very tiny balls consisting of 60 carbon atoms with diameter of ~0.7 nm | Water-soluble carboxylic acid $C_{60}$ derivatives acting as antimicrobials, being linked to a variety of active molecules | 25, 38, 39 |
| Quantum dots | Tiny nanocrystals commonly consisting of semiconductor materials in the range of 2–10 nm, glowing upon ultraviolet (UV) light | Mainly as probes to track antibodies, viruses, proteins, or deoxyribonucleic acid (DNA) in vivo | 26, 40 |
| Nanosuspensions | Drug nanocrystals dispersing in aqueous media commonly stabilized by surfactants | Suitable for insoluble drugs to obtain good bioavailability and targeting | 19 |
| Self-assembled drug nanostructures | Nanostructures consisting of amphiphilic prodrugs | Suitable for hydrophilic drugs to obtain good bioavailability, targeting, and controlled release | 21, 22 |

prevents them from agglomeration. As a result, manufacturing and processing of nanomaterials become hard issues. Anyway, many successful methods have been found to manufacture stable nanomaterials.

"Top down" and "bottom up" are two basic ways to manufacture nanomaterials. From its apparent meaning, the top-down method starts with a bulk material and then breaks it into smaller pieces using mechanical, chemical, or other forms of energy. Microchip manufacturing is the most common example of the top-down approach to produce nanomaterials. While this is an efficient approach for some industries, the process is generally labor and cost intensive. In contrast, the bottom-up method produces nanomaterials from atomic or molecular species via chemical reactions or physicochemical interactions such as self-assembly, allowing the precur-

sor molecules/particles to grow in size. Self-assembly leads to gaining the lowest energy state of molecules and makes molecules reorient naturally to obtain ordered aggregates. Carbon nanotubes, liposomes, and the SADNs are examples of nanomaterials that are manufactured using the bottom-up approach. A deep understanding of chemical and physical properties of precursor molecules/particles is needed to design and manufacture nanomaterials using the bottom-up approach. Both top-down and bottom-up approaches can be performed in gas, liquid, supercritical fluid, solid state, or vacuum. Anyway, when bulk materials corrupt, energy is required, and certainly the obtained nanoscale materials stay at a higher energy state than their parents. Whereas in the bottom-up approach molecules self-assemble into ordered aggregates with controlled behavior. Considering the higher energy of self-assembling monomolecules dispersing in media, their aggregation should be an energy-diminishing procedure and proceed spontaneously (Figure 2).

One of the largest hurdles of nanomanufacturing is how to scale up production. In the laboratory, manufacturing nanomaterials is difficult enough as highly advanced tools and carefully clean environments are required. Therefore, scale-up manufacturing in factories becomes a great challenge, hard to achieve. The most successful mass nanomanufacturing to date has occurred with computer microprocessors where companies have been able to etch circuit boards at 65 nm or smaller. Most manufacturers are interested in the ability to control (a) particle size, (b) particle shape, (c) size distribution, (d) particle composition, and (e) degree of particle agglomeration. Neither the top-down nor bottom-up approach is superior at the moment. Each has its advantages and disadvantages. However, the bottom-up approach may have the potential to be more cost-effective in the future.

Clinical applications require that biomedical nanomaterials have good biocompatibility or biodegradability. Therefore, biodegradable polymers (synthetic or natural), small molecules such as lipids, and some bioabsorptive inorganic salts such as calcium phosphate are preferred. Other materials such as poly(ethylene glycol) (PEG) is eventually excreted from body so they can also be selected. Materials that are nonbiodegradable or not easily removed from the body, such as carbon nanotubes and quantum dots, should be carefully considered as drug carriers, although they have already been used to deliver drugs or genes. More importantly, before any nanomaterial can be used in a clinic, the acute and long-term toxicity and side effects must be estimated in detail. So a novel discipline, nanotoxicology, is of great interest [41]. In addition, problems of large-scale production of nanomaterials, for example, the uniformity and stability of products, cannot be ignored. Some nanomaterials, including liposomes, polymeric or lipid nanoparticles, nanosuspensions, and SADNs, are described in detail in the following sections. The common manufacturing methods of pharmaceutical nanomaterials are listed in Table 2, though some are only used in the laboratory.

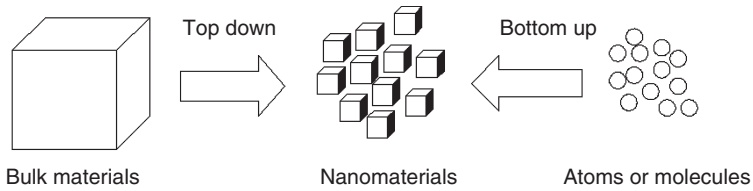


**FIGURE 2** Two basic methods to manufacture nanomaterials.

**TABLE 2 Manufacturing Methods of Some Nanomaterials in Pharmacy**

| Types of Nanomaterials | Materials Involved[a] | Manufacturing Methods | References |
|---|---|---|---|
| Nanoparticles | | | |
| Polymeric nanoparticles | Various natural polymers, e.g., albumin, gelatin, alginate, collagen, chitosan; biodegradable synthetic polymers, e.g., poly(lactic acid) (PLA), poly(lactide-*co*-glycotide) (PLGA), poly(ε-caprolactone) (PCL), poly(methyl methacrylate), and poly(alkyl cyanoacrylate); derivatives of cyclodextrin and starch; some modified polymers (e.g., PEGylated polymers) also used | Monomer polymerization, precipitation, solvent evaporation, salting out | 42–44 |
| Solid lipid nanoparticles (SLNs) | Mainly glycerides and fatty acids, surfactants also used | High-pressure homogenization, microemulsion technique, solvent evaporation | 27 |
| Inorganic nanoparticles | Calcium salts (e.g., calcium carbonate and calcium phosphate), gold | Precipitation | 45–47 |
| Nanocapsules | Various polymers, e.g., poly(*iso*-butylcyanoacrylate) (PIBCA), PLA, PLGA, PCL | Interfacial polymerization of monomers or interfacial nanodeposition of polymers | 29, 48, 49 |
| Liposomes | Phospholipids and cholesterol, phospholipid derivatives, e.g., PEG–polyethylene (PE), also added | Many methods used, mainly film hydration, reverse-phase evaporation, injection, freeze drying | 50 |
| Niosomes | Noionic surfactants, e.g., sorbitan monostearate (Span 60) | As for liposomes | 32 |
| Nanoemulsions | Oil and surfactants | High-pressure homogenization, ultrasonic emulsification, phase inversion | 34 |

**TABLE 2** ***Continued***

| Types of Nanomaterials | Materials Involved[a] | Manufacturing Methods | References |
|---|---|---|---|
| Polymeric micelles | Poloxamer-like block copolymers; PEG and lipophilic polymer copolymers; PEGylated lipids | Dialysis, emulsification, or film method | 35, 36 |
| Nanogels | Cross-linked hydrophilic copolymers, e.g., Pluronic–poly(ethylenimine) (PEI) and polyethylene oxide (PEO)–PEI | Covalent conjugation of polymers | 20, 51, 52 |
| Dendrimers | Dendritic macromolecules with repetitive moieties | Divergent or convergent synthesis | 23, 53 |
| Carbon nanotubes | Carbon, but only the water-soluble derivatives of CNTs used in pharmacy | CNTs formed by chemical vapor deposition (CVD) in presence of Fe catalyst, water-soluble CNT derivatives obtained by acid processing followed by conjugation with drugs | 54, 55 |
| Fullerenes ($C_{60}$) | Carbon, but only the water-soluble derivatives of $C_{60}$ used in pharmacy | $C_{60}$ obtained by arc discharge method using graphite electrodes or in a benzene flame, water-soluble $C_{60}$ derivatives obtained by acid processing followed by conjugation with drugs | 38, 56 |
| Quantum dots | Water-soluble derivatives of semiconductor materials (e.g., ZnS, PbS, CdSe, InP) used in pharmacy | QDs obtained via pyrolysis of organometallic precursors, water-soluble QD derivatives obtained by chemical reaction | 57–59 |
| Nanosuspensions | Pure drugs and stabilizers (including surfactants or polymers) | Precipitation, wet milling, homogenization | 19, 60, 61 |
| Self-assembled drug nanostructures | Polar drugs with proper conformation and lipids with long chains (e.g., glycerides, fatty acids, cholesterol) | Amphiphilic prodrugs obtained by synthesis, subsequently SADNs obtained by injection method | 21, 22 |

[a]Organic solvents may be involved and subsequently removed.

## 3 NANOTECHNOLOGY FOR DRUG DELIVERY

### 3.1 Nanocarriers

High-throughput screening technologies in drug discovery present an efficient way to find new potential active agents. But in recent years it has become evident that the development of new drugs alone is not sufficient to ensure progress in pharmacotherapy. Poor water solubility of potential active molecules, insufficient bioavailability, fluctuating plasma levels, and high food dependency are the major and common problems. Major efforts have been spent on the development of customized drug carriers to overcome the disappointing in vivo fate of those potential drugs. For drug carriers the followings are considered: nontoxicity (acute and chronic), sufficient drug-loading capacity, possibility of drug targeting, controlled-release characteristic, chemical and physical storage stability (for both drugs and carriers), and feasibility of scaling up production with reasonable overall costs. Nanocarriers have attracted great interest because they are desirable systems to fulfill the requirements mentioned above.

Over the past decade nanocarriers as nanoparticulate pharmaceutical carriers have been shown to enhance the in vivo efficiency of many drugs both in pharmaceutical research and the clinical setting, including liposomes, micelles, nanocapsules, polymeric nanoparticles and lipid nanoparticles. They perform various therapeutically or diagnostically important functions. More importantly, many useful modifications have been made, including the increased stability and half-life of nanocarriers in the circulation, required biodistribution, passive or active targeting into the required pathological zone, responsiveness to local physiological stimuli such as pathology-associated changes in local pH and/or temperature, and ability to serve as imaging/contrast agents for various imaging modalities (gamma scintigraphy, magnetic resonance imaging, computed tomography, ultrasonography). In addition, multifunctional pharmaceutical nanocarriers have already made a promising progress [62]. Some of those pharmaceutical carriers have already found their way into clinics, while others are still under preclinical investigation. This section presents two of the most promising nanocarriers, that is, liposomes and nanoparticles, especially their manufacturing, characteristics, and applications.

***Liposomes*** Liposomes (lipid vesicles) have a relative long history, first discovered by Banham in 1965 [63]. In the following decades, liposomes rapidly became a useful drug carrier. During the 1990s, many liposome-based drugs reached the market in the United States and Europe. The history of liposomes is the procedure of nanotechnology application to biomedicine. Phospholipids have particular structural conformation, leading to their self-assembly into bilayers with lipid chains inside and polar head groups outside during hydration. Importantly, phospholipids are the primary components of cell membranes so that liposomes have good biocompatibility without toxicity. The formation of liposomes is almost spontaneous, wherein a bottom-up procedure is involved [64]. When relatively free phospholipid molecules meet water, their polar head groups have affinity with water while lipid chains repulse water, which subsequently leads to their aggregation due to hydrophobic interaction, and then bilayers consisting of phospholipids are formed spontaneously. Closed vesicles are further formed by bilayer bending (Figure 3). Before phospho-

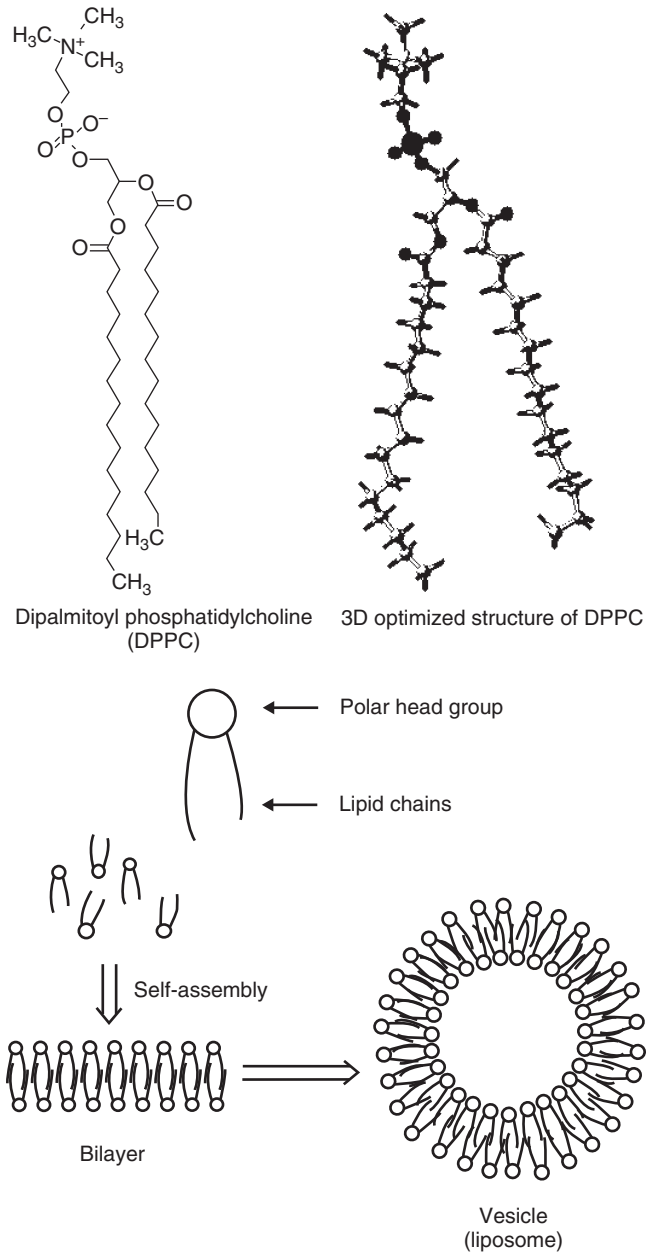


**FIGURE 3** Structures of phospholipids and formation of liposomes.

lipids become "free," bulk phospholipids must be dispersed throughout, forming a thin film, dissolution or emulsification, wherein additional energy is sometimes needed. Liposomes may have a size ranging from 10 nm to more than 10 μm mainly depending on composition and manufacturing approaches. A number of reports about the preparation of liposomes can be found in the literature and a detailed description of liposomes is in Chapter Cyclodextrin-Based Nanomaterials

in Pharmaceutical Field of this handbook. In this section the preparation, characteristics, and applications of nanosized liposomes are presented as well as some modified liposomes and recent progress.

Although various methods of manufacturing liposomes are reported, three types are usually involved: hydration of lipid film, interface aggregation of lipid molecules by emulsion-like process, and lipid solutions dispersing into nonsolvents by an injection-like process or controlled mixture. Practical methods are thin-film hydration [65], reverse-phase evaporation [66], ethanol injection [67], polyol dilution [68], double emulsions [69], proliposome method [70], and high-pressure homogenization [71]. Liposomes may have various morphologies related to manufacturing methods, mainly multilamellar vesicles (MLVs), large unilamellar vesicles (LUVs), and small unilamellar vesicles (SUVs). Liposomes can be further processed by sonication, detergent depletion, membrane filtration [72], and lyophilization [73] to make them finer and more uniform or stable. For example, MLVs are sonicated to SUVs.

The composition of liposomes is a key factor in their manufacturing. Phospholipids are major components of liposomes. In terms of resources, phospholipids are classified as natural, semisynthetic, and wholly synthetic phospholipids. Natural phospholipids also have different resources (e.g., soybean, egg yolk). In terms of polar head groups, phospholipids are classified as phosphatidylcholine (PC), phosphatidylethanolamine (PE), phosphatidylserine (PS), phosphatidylinositol (PI), phosphatidylglycerol (PG), and phosphatidic acid (PA), where PC and PE are the most used. Different polar head groups result in varied surface charged liposomes that then influence the stability and in vivo distribution. Because PS, PI, PG, and PA have negative charges, the liposomes containing them are negatively charged. Sometimes, other lipids such as *N,N′*-dioleoyl-*N,N′*-dimethylammonium chloride (DODAC) and stearylamine are mixed with phospholipids to prepare positively charged liposomes. Cholesterol is commonly used with phospholipids because cholesterol can make liposomal membranes stronger [50]. The mole percentage of cholesterol in the liposomal composition is commonly not more than 50%. Lecithin (an often used term in the lipid field) as a phospholipid from natural resources (e.g., soybean lecithin and egg lecithin) is often used to manufacture liposomes, which is actually a mixture composed of various kinds of phospholipids though PC dominates. The long-chain fatty acids constituting phospholipids also have many types, such as lauric (C12), myristic (C14), palmitic (C16), and stearic (C18). In general, unsaturated fatty acids occur in natural phospholipids. Dimyristoyl phosphatidylcholine (DMPC), dipalmitoyl phosphatidylcholine (DPPC), distearoyl phosphatidylcholine (DSPC), and dipalmitoyl phosphatidylethanolamine (DPPE) are the most common synthetic phospholipids. The length of the lipid chain influences the gel–liquid crystalline phase transition temperature ($T_c$) of phospholipids, wherein longer chained lipids lead to higher $T_c$. For example, DPPC has a $T_c$ of 41°C while DSPC has a $T_c$ of 54°C [50].

Drug entrapment is an important parameter in manufacturing liposomes which is influenced by many factors: the types, molecular weights, and physicochemical properties of drugs; the types, sizes, and compositions of liposomes; and the manufacturing methods. In addition, entrapped drugs may leak during storage. Drugs may be entrapped in one of two parts of liposomes, the inner phase or bilayers, depending on the physicochemical property of the drugs. Water-soluble drugs prefer the aqueous inner phase while lipophilic drugs prefer the hydrophobic environment of bilayers. Macromolecules such as peptides and proteins can adsorb onto bilayers,

wherein electrostatic interaction can influence entrapment. Those drugs insoluble in both water and oil are hard to entrap. Rather than common manufacturing approaches, more promising methods are adopted to improve drug entrapment. Ionic gradient methods can increase the entrapment efficacy of some ionic drugs, including the pH gradient method [74], the ammonium sulfate gradient method [75, 76], the acetate gradient method [77], and the manganese ion gradient method [78]. Lyophilization of liposomes is a good preservation method which can prevent entrapped drugs from leaking, liposome precipitation and agglomeration due to gravity and thermal movement, and possible hydrolysis of phospholipids (resulting in production of toxic lyso-phospholipids). Generally, lipophilic drugs have a high entrapment efficacy, though drug loading is limited, because they insert into bilayers tightly. Therefore, an efficient method increasing entrapment efficacy is to prepare the lipophilic derivatives of hydrophilic drugs [79].

The manufacturing of nanosized liposomes can be performed using the methods mentioned above. However, the small size of nanoliposomes is difficult to achieved by methods such as film hydration. Molecular self-assembly occurs in the injection method, and then the size and morphology of obtained liposomes can be well controlled. In fact, liposomes that result from the injection method are uniform and small enough, to the nanoscale, and usually SUVs are obtained. Because of the very low toxicity of ethanol, the ethanol injection method is usually used and is described as follows to show the process of manufacturing liposomes [50]. A scale-up manufacturing process of the ethanol injection method has been established [80–82]. The obtained liposome size is mostly less than 300 nm:

(a) Handling and storage of lipids is important. Store organic solutions of phospholipids in a sealed glass container layered with argon or nitrogen below –20°C, preferably at –78°C. When transferring a portion of the material, allow it to reach room temperature before opening the bottle. Saturated phospholipids, that is, lipids composed of completely saturated fatty acids, such as DPPC, are stable as powders. However, storage of these lipids as described above is highly recommended. Unsaturated phospholipids are extremely hygroscopic as powders, which will quickly absorb moisture and become gummy upon opening the storage container. Always dissolve such lipids in a suitable solvent (preferably chloroform) and store it in a glass container at –78°C.

(b) Prepare materials such as phospholipids, cholesterol, other additives, ethanol, injector, beaker, agitation machine, and evaporation device before manufacturing. Calculate the amount of these agents according to the request of the last products. A fine-gauge needle to a 1-mL glass syringe is preferred. Dissolve lipid components (including lipophilic drugs) in ethanol. Dissolve water-soluble drugs in water or aqueous media as dispersing media.

(c) Rapidly inject the ethanol solutions into agitated aqueous media with the tip under the surface. A homogeneous and almost transparent liquid will be obtained. Repeat this process and notice that the percentage of ethanol in the last product is not more than 7.5%. Collect all liquids and remove ethanol by evaporation, dialysis, or gel filtration. The last liposomal suspensions can be further concentrated through evaporating water. Sterilize them by autoclave. They may be lyophilized when needed.

Separation of nonentrapped drugs from liposomes (or purification of liposomes) is an important process after manufacturing. The size difference between liposomes and unincorporated materials is the basis of separation. Gel chromatography, dialysis, and centrifugation are usual approaches. The drug entrapment percentage of liposomes can be obtained after separation. The whole drugs in liposomal suspensions or the entrapped drugs can be determined by dissolving liposomes with organic solvents or solubilizing liposomes with detergents to release drugs. The morphology of liposomes can be investigated by negatively stained transmission electron microscopy, cryo-electron microscopy, or freeze-fracture electron microscopy. The size distribution of liposomes is usually analysized by photon correlation spectroscopy (laser light scattering) [50].

Beyond conventional liposomes, functional liposomes are designed to achieve various therapeutic effects. Conventional liposomes manufactured by natural phospholipids or commonly used synthetic phospholipids such as PC and PE are negatively charged. However, cationic liposomes can form complexes with peptides or nucleic acids through electrostatic interaction and prefer to adsorb onto the surfaces of cell membranes, subsequently improving interaction with cells and penetrating into cytosol or phagocytosis. Therefore, cationic liposomes have become a standard transfection agent in cell manipulation [83]. Furthermore, they become primary nonviral gene delivery carriers [84].

Liposomes show site-specific distribution in the MPSs after intravenous (IV) administration due to opsonization by the complement system [85]. The diseases in MPSs can benefit from the drug targeting. But this is a bad result for diseases in other tissues. Long-circulating liposomes are then developed for targeting to non-MPS tissues. The long-circulating effect results from hydrophilic polymers coated on liposomes. For example, the half-life of the long-circulating liposomes can be extended to 20 h in rat. They are also called sterically stable liposomes or Stealth liposomes. The lipid conjugate of PEG, PEG–DSPE, is commonly used and inserts into bilayers and hinders plasma protein adsorption. The enhanced permeability and retention (EPR) effect of solid tumors makes long-circulating liposomes a very useful tool for anticancer therapy [86]. However, in recent years it was reported that in most cases PEGylated liposomes were cleared very rapidly from circulation with repeated injection. But doxorubicin PEGylated liposome is an exception. The production of anti-PEG immunoglobulin (Ig) M following injection is the major reason, and the spleen also plays a key role [87]. However, a more recent case has appeared. A modified phospholipid–methoxy(polyethylene glycol) conjugate was recently synthesized through the methylation of phosphate oxygen moiety which could prevent PEGylated liposomes from being activated by a complement system in vivo followed by achieving a true long-circulating effect [88].

Other functional liposomes are mainly stimuli-responsive liposomes. The pH-sensitive liposomes contain pH-sensitive lipids such as 1,2-dioleoyl-*sn*-3-phosphatidylethanolamine (DOPE) showing an inverted hexagonal configuration in a low-pH environment and release entrapped drugs in the low-pH environment of tumor tissues due to liposomal membrane destabilization [89]. Temperature-sensitive liposomes are prepared from special lipids such as DPPC whose phase transition temperature ($T_c = 41°C$) is proper to perform clinical anticancer therapy. When up to $T_c$, the fluidity of liposomal membranes increases sharply, followed by

entrapped drugs releasing [90]. Some thermosensitive polymers can also be used to manufacture temperature-sensitive liposomes [91]. Magnetoliposomes load ultrafine magnetite, preferring to accumulate in the local tissue within the magnetic field [92]. Immunoliposomes load attached monoclonal antibodies to treat some severe diseases such as cancer [93].

Liposomes have been successfully applied to many drugs, diagnostic agents, imaging agents, transfection agents, vaccines, and so on. Liposomes have been tried in almost all routes of administration: oral, injection (intravenous, subcutaneous, intramuscular, intra-articular cavity, and other possible injection sites), intranasal, pulmonary inhalation, conjunctiva, topical, and transdermal. The most significant application field of liposomes is still anticancer therapy. After a long-time research for 30 years, some liposomal products have reached the market (Table 3). The major problems in manufacturing liposomes are scale-up production, efficient sterilization, and stable storage.

**TABLE 3 Liposomal Drugs Approved for Clinical Application**

| Drug | Product Name | Composition of Liposomes and Other Major Excipients | Indication | Company |
|---|---|---|---|---|
| Daunorubicin | DaunoXome | DSPC, cholesterol | Kaposi's sarcoma | Gilead Sciences |
| Doxorubicin | Mycet | Egg PC, cholesterol | Combinational therapy of recurrent breast cancer | Zeneus |
| Doxorubicin | Doxil/Caelyx | MPEG–DSPE, HSPC, cholesterol, ammonium sulfate, sucrose, histidine | Refractory Kaposi's sarcoma; ovarian cancer; recurrent breast cancer | Alza/SP Europe |
| Amphotericin B | AmBisome (lyophilized product) | HSPC, cholesterol, DSPG, α-tocopherol, sucrose, disodium succinate | Fungal infections | Gilead Sciences |
| Cytarabine | DepoCyt | DOPC, DPPG, cholesterol, triolein | Lymphomatous meningitis | SkyePharma |
| Morphine | DepoDur | DOPC, DPPG, cholesterol, tricaprylin, triolein | Pain following major surgery | SkyePharma |

MPEG = methyl PEG; HSPC = hydrogenated soy phosphatidylcholine; DSPG = disteroylphosphatidylglycerol; DOPC = dioleoylphosphatidylcholine; DPPG = dipalmitoylphosphatidylglycerol.

***Nanoparticles*** Nanoparticles attract much attention of pharmaceutical scientists because of their controllable manufacturing, uniform preparations, and low cost. The major difference between nanoparticles and liposomes is that the former has a solid core while the latter only has inner aqueous phase and thin bilayers. In addition, in the case of liposomes, the entrapped water-soluble drugs exist only in solutions of the inner phase, while lipophilic drugs are only limited in the small space of bilayers. Therefore, the drug-loading efficiency (drug–lipid ratio) of liposomes is always limited. In the case of nanoparticles, drugs exist in the solid state, and high drug loading is possibly achieved. Unlike liposomes, nanoparticles may be composed of various materials, and biodegradable materials are preferably used. Furthermore, modified materials based on traditional natural and synthetic materials are also frequently used to manufacture nanoparticles to achieve more functions, which then highly enlarges the lists of used materials. In addition, more manufacturing methods of nanoparticles are optional than liposomes. Therefore, nanoparticles are relatively ideal nanocarriers for most drugs.

Nanoparticles can be classified as three types, polymeric nanoparticles, lipid nanoparticles, and inorganic nanoparticles, depending on the major components. Polymeric nanoparticles are exploited earlier, while lipid nanoparticles are of great interest in recent years due to very good biocompatibility. The development of polymeric nanoparticles is highly related to polymer science. Besides a great deal of natural polymers, more and more biodegradable polymers are synthesized, which allows pharmaceutical scientists to have enough optional subjects. Solid lipid nanoparticles (SLNs) composed of solid lipids have a profound advantage of no biotoxicity [94]. Inorganic nanoparticles are currently exploited only a little [45–47], and the major problems may be their poor biodegradability and relatively low drug-loading efficiency.

*Polymeric Nanoparticles* Polymeric materials for manufacturing nanoparticles include synthetic poly(lactic acids) (PLA), poly(lactide-*co*-glycolide) (PLGA), poly(ε-caprolactone) (PCL), poly(methyl methacrylates), and poly(alkyl cyanoacrylates); natural polymers (albumin, gelatin, alginate, collagen), and modified natural polymers (chitosan, starch). Polyesters, alone and in combination with other polymers, are most commonly used for the formulation of nanoparticles. PLGA and PLA are highly biocompatible and biodegradable. They have been used since the 1980s for numerous in vivo applications (biodegradable implants, controlled drug release). The U.S. Food and Drug Administration (FDA) has approved PLGA for human therapy [95]. More recently, formulations based on natural polymers have been developed and are on the market. For example, a wonderful natural polymer, chitosan, has permeability enhancer abilities, allowing the preparation of organic solvent free mucoadhesive particles [42].

Nanoparticles of synthetic polymers are usually manufactured by dispersion of preformed polymers. Although many methods can be used, they may be classified as monomer polymerization, nanoprecipitation, emulsion diffusion/solvent evaporation, and salting out. An appropriate method is selected mainly depending on polymer and drug natures. Polymerization of polymer monomers has been developed usually using poly(alkyl cyanoacrylate) [96, 97]. Organic solvents are usually used in polymerization. A detailed description of this method is not provided here.

The nanoprecipitation method is commonly adopted to entrap lipophilic drugs, and low polydispersity is probably achieved [42]. In general, the organic solution containing drugs and polymers is added a nonsolvent to lead to polymers precipitating together with drugs. The size of formed nanoparticles can be adjusted by the polymer and nonsolvent amounts in the organic phase. Nanoparticles can be separated from solvents and unincorporated drugs with centrifugation followed by spray drying or freeze drying when needed. The stability and drug recovery yield of nanoparticles depend on the ratio of drugs to polymers [98]. Recently, this technique has also been used to entrap hydrophilic compounds into PLGA and PLA nanoparticles [99, 100], especially peptides and proteins [101].

Another common method to manufacture polymeric nanoparticles is the emulsion diffusion or solvent evaporation technique, which is used to entrap hydrophobic or hydrophilic drugs. Generally, the polymer and hydrophobic drugs are dissolved in a partially water miscible organic phase (e.g., benzyl alcohol, propylene carbonate, and ethyl acetate). The organic solution is emulsified in aqueous media containing a suitable surfactant [i.e., anionic sodium dodecyl sulfate (SDS), nonionic poly(vinyl alcohol) (PVA) or cationic didodecyl dimethyl ammonium bromide (DMAB)] under stirring. The diffusion of the organic solvent and the counterdiffusion of water into the emulsion droplets induce polymeric nanoparticle formation. The organic solvent is evaporated. Also hydrophilic drugs could be entrapped into a water-in-oil (W/O) emulsion containing polymers and then undergo the above process. Then a water-in-oil-in-water (W/O/W) emulsion is obtained. After evaporation of total organic solvent, the drug-loaded nanoparticles can be separated. Polymer nature, polymer concentration, solvent nature, surfactant molecular mass, viscosity, phase ratio, stirring rate, temperature, and flow of water all affect nanoparticle size [102].

The salting-out method is also used. Polymers are dissolved in water-miscible organic solvents such as acetone or tetrahydrofuran (THF). The organic phase is emulsified in an aqueous phase that contains the emulsifier and salts of high concentration. Typically, the salt solution used contains 60% (w/w) magnesium chloride hexahydrate or magnesium acetate tetrahydrate with a polymer-to-salt ratio of 1:3. In contrast to the emulsion diffusion method, no diffusion of solvents occurs due to the presence of high concentrated salts. The fast addition of pure water to the O/W emulsion under mild stirring reduces the ionic strength and leads to the migration of the organic solvent to the aqueous phase, inducing nanoparticle formation. The final step is purification by cross-flow filtration or centrifugation to remove the salting-out agent. Common salting-out agents are electrolytes (sodium chloride, magnesium acetate, or magnesium chloride) or nonelectrolytes, such as sucrose. Polymer concentration and molecular weight, stirring rate and time, and the nature and concentration of surfactant and solvent are all important parameters. This method would allow avoiding the use of organic chlorinated solvents and large amounts of surfactants [102]. Furthermore, formulation of nanoparticles with natural polymers is performed by ionic gelation (chitosan), coacervation (chitosan, gelatin), and desolvation (gelatin) [102, 103]. These mild methods have the advantage of producing organic solvent-free formulations.

Additional advantages can be obtained by changing nanoparticle surface properties, for example, good stability, mucoadhesion, and long circulation time. For example, the in vivo long-circulating effect is achieved either by coating

nanoparticle surfaces with hydrophilic polymers/surfactants or by incorporating biodegradable copolymers containing a hydrophilic moiety. Like long-circulating liposomes, PEG-containing polymers are frequently used to manufacture long-circulating nanoparticles. PEGylated copolymers (PLA–PEG, PLGA–PEG and PCL–PEG) are used [42, 104], and the long-circulating effect also results from the adsorption or covalent conjugation of some hydrophilic polymers with the hydrophobic surface of nanoparticles [105, 106]. Moreover, the active targeting of nanoparticles can be achieved by incorporating the conjugate of the polymer and target-directed molecule, such as (Arg-Gly-Asp) RGD, (trans-activator transcription) TAT peptides [107], and monoclonal antibody [108]. Recently, self-assembled nanoparticles have aroused great interest, consisting of amphiphilic macromolecules, such as hydrophobically modified glycol chitosan, which can also entrap drugs or peptides [109].

Before nanocarriers go into clinical applications, some issues must be considered, including drug-loading capacity, possibility of drug targeting, in vivo fate of the carrier (interaction with the biological surrounding, degradation rate, accumulation in organs), acute and chronic toxicity, scaling up of production, physical and chemical storage stability, and overall costs. A certain advantage of polymer systems is the wealth of possible chemical modifications. Possible problems of polymeric nanoparticles derive from residues of organic solvents used in the production process, polymer cytotoxicity, and the scaling up of production. Polymer hydrolysis during storage has to be taken into account and lyophilization is often required to prevent polymer degradation [94].

*Solid Lipid Nanoparticles* The outstanding advantage of lipid nanoparticles is perfect biocompatibility because their raw materials are the components of our body, preferring to be used or degraded by the body. Solid lipids are usually used as the major component of lipid nanoparticles—hence the name solid lipid nanoparticles. However, the used solid lipids generally become liquid at a high temperature to adapt to the preparation of SLNs. Compared with polymeric nanoparticles, the materials used for SLNs are simpler. The frequently used lipids are glycerides of various fatty acids, which also exist in the emulsions for parenteral nutrition. Large-scale production of SLNs can be achieved in a cost-effective and relatively simple way using high-pressure homogenization and microemulsion. Another useful method is solvent emulsification/evaporation. The SLN introduced in 1991 represents an alternative carrier system to traditional colloidal carriers, such as liposomes and polymeric nanoparticles. SLNs combine advantages of the traditional systems but avoid some of their major disadvantages [27]. The proposed advantages of SLNs include [94]:

- Possibility of controlled drug release and drug targeting
- Increased drug stability
- High drug payload
- Incorporation of lipophilic and hydrophilic drugs feasible
- No biotoxicity of the carrier
- Avoidance of organic solvents
- No problems with respect to large-scale production and sterilization

Solid lipids, emulsifiers, and water are generally the ingredients involved for manufacturing SLNs. The term *lipids* is used in a broader sense and includes triglycerides (e.g., stearin), partial glycerides (e.g., Imwitor), fatty acids (e.g., stearic acid), steroids (e.g., cholesterol), and waxes (e.g., cetyl palmitate). All categories of emulsifiers may be used to stabilize the lipid dispersion, and the combination of emulsifiers prevents particle agglomeration more efficiently. The choice of the emulsifier depends on the administration route and is more limited for parenteral administration.

High-pressure homogenization (HPH) has emerged as a reliable and powerful technique for the preparation of SLNs. Homogenizers of different sizes are commercially available from several manufacturers at reasonable prices. In contrast to other techniques, scaling up of HPH is out of the question in most cases. High-pressure homogenizers push a liquid with high pressure (100–2000 bars) through a narrow gap (in the range of a few micrometers). The fluid accelerates over a very short distance to very high velocity (over 1000 km/h). Very high shear stress and cavitation forces disrupt the particles down to the submicrometer range. Typical lipid contents are 5–10%, though higher lipid concentrations (up to 40%) may be used. Two general approaches of HPH, hot and cold homogenization, can be used for manufacturing SLNs (Figure 4) [27].

Microemulsions (transparently appearing with droplet size less than 100 nm) are thermodynamically stable systems, and the choice of optimal formulation containing oil, surfactant, cosurfactant, and oil–water ratio is key [110]. Generally, the solid lipid of low melting point (e.g., stearic acid) melts at a high temperature (e.g., 65–70°C), and then hot microemulsions are prepared using it. SLNs can be obtained after the hot microemulsions are rapidly cooled by injecting them into cold water (e.g., 2°C) under stirring. Emulsifiers in a formulation typically include Tween 20/60/80, lecithin, and sodium taurodeoxycholate, and coemulsifiers include alcohols and sodium monooctylphosphate. The typical volume ratios of hot microemulsions to cold water are from 1:25 to 1:50. The very low solid concentration of SLN suspensions is the disadvantage of the microemulsion method. Rapid temperature decrease in hot microemulsions is key to obtaining homogeneous and small-sized nanoparticles. A high temperature gradient can also ensure rapid lipid crystallization and prevent aggregation [94].

The solvent emulsification/evaporation method involves lipid precipitation in O/W emulsions. Solid lipids are dissolved in a water-immiscible organic solvent (e.g., cyclohexane) followed by emulsification in an aqueous medium. Upon evaporation of the solvent, the nanoparticle dispersion is formed due to lipid precipitation. Residue of organic solvents is the major problem of this method [94]. However, the microemulsion and solvent emulsification/evaporation methods can be performed conveniently in the laboratory without specific apparatuses.

During research of SLNs, some problems have continually appeared, for example, very low drug loads, drug expulsion during storage, and high water content of SLN dispersions. The $\alpha$ and $\beta'$ crystallines of higher energy state mainly appear in conventional SLNs manufactured by hot-homogenization technique. However, these crystallines prefer to transform to the more ordered $\beta$ modification of low energy state during storage. The high ordered degree improves the crystal imperfections, diminishing further to lead to drug expulsion. To solve this problem, a new kind of lipid nanoparticle was developed, called a nanostructured lipid carrier (NLC). NLCs

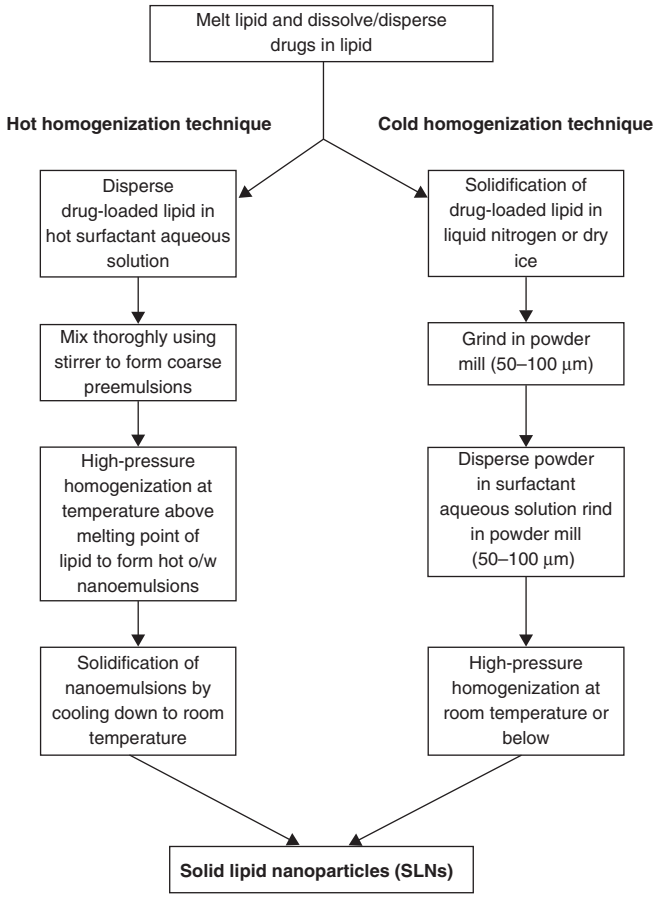


**FIGURE 4** High-pressure homogenization for manufacturing SLNs.

are composed of spatially very different lipid molecules, that is, solid lipids and liquid lipids (oils). The matrix remains solid at body temperature though its melting point is lower than one of the original solid lipids. No crystallization happens in NLCs so that the drug loads can be increased and the expulsion during storage is avoided [111].

New functions can be obtained by modifications of SLNs. Incorporation of Tween 80 and Poloxamer 188 can stabilize SLNs to achieve long-circulating or crossing blood–brain barrier effects [112]. Recently, novel nanoparticles called polymer–lipid hybrid nanoparticles (PLNs) were developed [113]. They can entrap cationic anticancer agents (e.g., doxorubicin) effectively by incorporation of an anionic lipophilic polymer into lipids to treat multidrug-resistant (MDR) cancers.

In general, SLNs are used to entrap hydrophobic drugs due to their lipid nature, but a few reports show that hydrophilic drugs can also be entrapped. A hydrophilic peptide, gonadorelin, and monostearin were dissolved in acetone and ethanol at 50°C followed by pouring the resultant organic solution into an aqueous medium containing 1% PVA under agitation to obtain peptide-loaded SLNs that were sub-

sequently separated by centrifugation. Up to 69% of gonadorelin was incorporated. The in vitro release of gonadorelin from SLNs was slow [114]. The W/O/W multiple-emulsion technique was also used to manufacture peptide-loaded SLNs. Insulin is a model peptide located in the inner water phase of the W/O/W emulsion, tripalmitin is the core of SLNs, and the surfaces are modified with PEG 2000–stearate. The insulin-loaded SLNs show good stability upon the low pH of the gastric medium and the pancreatic enzymes in intestinal medium [115].

*Perspective of Nanoparticles* As drug nanocarriers, nanoparticles have unique advantages: for example, high dispersing, adhesive property, targeting in vivo. Like liposomes, anticancer therapy is a major function of nanoparticles [116]. Easy modification of nanoparticles also makes them platforms to perform more functions, for example, delivering drugs across the blood–brain barrier (BBB) [117], lymphatic targeting [118], and gene delivery [119].

Abraxane is a successful paradigm of nanoparticle application. It is an albumin nanoparticle loading paclitaxel developed by American Pharmaceutical Partners (APP) and American BioScience. The outstanding advantage of Abraxane is no significant side effects, not like the traditional paclitaxel preparation with Cremophor EL (polyethoxylated castor oil) and ethanol. More nanoparticle products will reach the market in the future.

### *Other Nanocarriers*

*Nanoemulsions* Lipid nanoemulsions were introduced in the 1950s as parenteral nutrition. Vegetable oils (e.g., soy oil) or middle-chain triglycerides are used, typically occupying 10–20% of the emulsion. Other ingredients include phospholipids as stabilizers and glycerol as osmolar regulation agent. In recent years this system has been further developed to load lipophilic drugs and several formulations are commercialized. Examples are etomidate (Etomidat-Lipuro), diazepam (Diazepam-Lipuro and Stesolid), propofol (Disoprivan), and dexamethasone palmitate (Lipotalon). In comparison to previous, solubilization-based formulations of these drugs, reduction of the local and systemic side effects (e.g., pain during injection) has been achieved. The possibility of controlled drug release from nanoemulsions is restricted due to the small size and the liquid state of the carrier. Most drugs show a rapid release from them. Advantages of nanoemulsions include toxicological safety and a high content of the lipid phase as well as the possibility of large-scale production by high-pressure homogenization [94].

*Microemulsions* Microemulsions are nanoemulsions which are optically isotropic, transparent or translucent, low-viscous, and thermodynamically stable liquid solutions, mainly containing tiny liquid droplets less than 100 nm. The manufacturing of microemulsions as a self-formed system is relatively simple. They are bicontinuous systems essentially composed of water and oil with surfactant and cosurfactant separating. A very low interfacial tension to 0 mN/m is found in microemulsions despite the large oil–water interfacial areas. A prominent example is the Sandimmun Optoral/Neoral preconcentrate for microemulsions. Now microemulsions are usually limited to dermal and peroral applications due to their high surfactant content. Because they only exist in narrow regions of phase diagrams, they are very restricted in tolerance to quantitative formulation changes [120].

*Polymeric Micelles* Polymeric micelles composed of amphiphilic copolymers, that is, polymers consisting of a hydrophobic block and a hydrophilic block, are gaining increasing attention. They show high stability both in vitro and in vivo and good biocompatibility, and more importantly they can solubilize a broad variety of poorly soluble drugs in their inner core. Many of these drug-loaded micelles are currently at different stages of preclinical and clinical trials. Due to their hydrophilic shell and small size, they prefer to exhibit prolonged circulation times in vivo and can accumulate in tumor tissues. Polymeric micelles are formed by block copolymers consisting of hydrophilic and hydrophobic monomer units with the length of a hydrophilic block exceeding to some extent that of a hydrophobic one. If the length of a hydrophilic block is too high, copolymers exist in water as unimers (individual molecules), while molecules with very long hydrophobic block prefer to form structures with nonmicellar morphology, such as rods and lamellae. Diblock copolymers with an A–B structure and tri- or multiblock copolymers such as poly(ethylene oxide)–poly(propylene oxide)–poly(ethylene oxide) (PEO–PPO–PEO) (A–B–A) may self-organize into micelles. The amphiphilic copolymers commonly have the critical micelle concentration (CMC) values as low as $10^{-6}$ *M*, which is about two orders of magnitude lower than that of such surfactants as Tween 80. As potential drug carriers, the hydrophobic core of polymeric micelles generally consists of a biodegradable polymer such as poly(β-benzyl-*l*-aspartate) (PBLA), PLA, or PCL which serves as a reservoir for an insoluble drug, protecting it from contact with the aqueous environment. The core may also consist of a water-soluble polymer [e.g., poly(aspartic acid; P(Asp)], which is rendered hydrophobic by the chemical conjugation of a hydrophobic drug or is formed through the association of two oppositely charged polyions (polyion complex micelles). Another special group of polymeric micelles is formed by lipid-core micelles, that is, micelles formed by conjugates of soluble copolymers with lipids (e.g., PEG–PE) [35, 36].

Three methods are used to manufacture polymeric micelles: dialysis, emulsification, and film methods. In the dialysis method, the drug and copolymer are dissolved together in a water-miscible solvent (e.g., ethanol) followed by dialysis against water. During the process (possibly several days), the insoluble drugs are incorporated into the formed micellar core. In the emulsification method, an O/W emulsion is first prepared using an aqueous solution of the copolymer and the drug solution in a water-insoluble volatile solvent (e.g., chloroform). The drug-loaded micelle is formed as solvent evaporation. In the film method, the copolymer solution and the drug solution are dissolved separately in miscible volatile organic solvents and are mixed followed by evaporating solvents to form a polymer/drug film. The film is hydrated in water or buffers, and then the micelle is formed by intensive shaking. If the amount of a drug exceeds the solubilization capacity of micelles, the excess drug precipitates in a crystalline form and is removed by filtration. The loading efficiency for different compounds varies from 1.5 to 50% by weight. The major driving force behind self-association of amphiphilic polymers is the decrease of free energy of the system due to removal of hydrophobic fragments from the aqueous surroundings with the formation of micelle core stabilized by hydrophilic blocks exposed to water [35, 36].

Various drugs, for example, diazepam and indomethacin, doxorubicin, anthracycline antibiotics, and polynucleotides, were effectively solubilized in polymeric micelles. Also polymeric micelles can carry various reporter (contrast) groups and

become the imaging agents. Besides targeted drug delivery due to the EPR effect of tumor, specific polymeric micelles having stimuli-responsive amphiphilic block copolymers, targeting ligand molecules, or monoclonal antibody molecules are also manufactured [35, 36].

*Nanogels* Nanogels are colloidal stable particles made from hydrogels with nanosized hydrophilic polymeric networks. Hydrogels are the simple gels swelling strongly in aqueous media, typically composed of hydrophilic polymer components cross-linked into a network by either covalent (chemical cross-linking) or noncovalent (physical cross-linking) interactions. It is the cross-linking that provides for dimensional stability, while the high solvent content gives rise to the fluidlike transport properties. Cross-links are important to maintain the network structure of the hydrogels and prevent dissolution of the hydrophilic chains [121].

Two methods, emulsification–evaporation and the micelle/nanoparticle approach, are used to manufacture nanogels. In the former method, bis-activated PEG in dichloromethane is added dropwise to the aqueous solution of polyethylenimine (PEI) and then sonicated. The resulting white emulsion is evaporated in vacuum, producing a clear, slightly opalescent solution. This solution is stirred for less than one day at room temperature and much debris is separated by centrifugation. The nanogel suspension is obtained after dialysis against water [51]. This procedure is convenient except for using organic solvents, a vacuum evaporation step, and the obtained particles with a wide size distribution. Another method involves surface preactivated micelles or nanoparticles followed by reaction with other polymers on the surface in aqueous media. None of the organic solvents involved are of benefit. For example, a Pluronic block copolymer both ends of which are activated by 1,1′-carbonyldiimidazole is dissolved in water at a concentration above its CMC. A diluted aqueous solution of PEI is then added dropwise to the micellar solution, stirring overnight. During this procedure a covalently linked cationic polymer PEI layer is formed on the Pluronic micelles, producing nanogels with narrow size distribution. After dialysis the resulting nanogel suspension can be further lyophilized. Using this procedure, the nanogels based on Pluronic P85/PEG and F127/PEG are obtained with final yields of 55 and 70% by weight and average hydrodynamic diameters of 100 and 180 nm, respectively [20, 52].

Many drugs can be entrapped into nanogels, for example, valproic acid, nucleoside analogues, antisense oligonucleotides, adenosine triphosphate (ATP), and small interfering ribonucleic acid (siRNA). Because macromolecular drugs such as peptides and proteins need to locate in a hydrophilic environment to maintain their activity, the particular hydrophilic property of nanogels would be of benefit. Special functions such as cellular targeting, crossing the BBB, and controlled release may also be achieved by the nanogel technique. In addition, the nanogel materials should be biodegradable. In cationic nanogels, PEI and PEG are cross-linked via urethane bonds, usually considered as stable links. However, due to the presence of highly protonated PEI, hydrolysis of these bonds was significantly accelerated, and the polymer network of nanogels could rapidly degrade in aqueous solution at the physiological pH during a period of about two weeks [20].

*Dendrimers* Dendrimers attracted much attention after they were first investigated by Tomalia 20 years ago [122, 123], and they have become the star molecules

in recent years. Dendrimers possess perfect nanoarchitectures from 1 nm to more than 10 nm, consisting of repetitive chemical moieties with tree architecture. According to repetitive folds, dendrimers with the same basic cores are divided into a series of generations. The higher generation of dendrimers represents more repetitive units. Dendrimers are hyperbranched macromolecules that can be subdivided into three architectural parts: (a) the multivalent surface, with a high number of potential reactive sites; (b) the "outer shell" just beneath the surface, having a well-defined microenvironment protected from the outside by the dendrimer surface; and (c) the core, which in higher generation dendrimers is protected from the surroundings, creating a microenvironment surrounded by dendritic branches. Therefore, the interior of dendrimers is well suited for entrapment of guest molecules. The multivalent surfaces on a higher generation dendrimer can contain a very high number of functional groups. This makes the dendritic surfaces and outer shell well suited to host–guest interactions. Dendrimers can be tailored specifically for the desired purposes, for example, as dendritic sensors, drug vehicles, or even drugs [23].

Dendrimers are synthesized through a stepwise repetitive reaction sequence, wherein a convergent or divergent approach is used. On the one hand, the most divergent dendrimer syntheses require excess monomer loading and lengthy chromatographic separations, particularly at higher generations. On the other hand, convergent synthesis strategies are generally limited to the construction of only lower generation dendrimers due to the nanoscale steric issues that are encountered when attaching the dendrons to the molecular-level core [124]. Currently much of the work on dendrimers has been based on the commercially available Starburst poly(amidoamine) (PAMAM) dendrimers that are extensively studied as drug carriers. PAMAM may be synthesized from an ammonia or ethylenediamine core (EDA) by the divergent approach, involving Michael addition followed by amidation with methyl acrylate and resulting in the production of a dendrimer family ($G = 0$–$7$), and half-generation dendrimers are carboxyl terminated and full-generation dendrimers are amine terminated (e.g., $G = 5.0$, 5.3 nm in size) [122, 123].

Dendrimers have been evaluated as drug nanocarriers, gene transfection agents imaging agents, and nanodrugs [124]. Also many surface-modified dendrimers have been synthesized to obtain more functions such as active targeting and gene delivery. Dendrimers may be used as drugs for antibacterial and antiviral treatment and as antitumor agents. VivaGel, a topical water-based gel based on sulfonated naphthyl-modified poly(lysine) dendrimers, has been evaluated against human immunodeficiency virus (HIV) and other sexually transmitted diseases (STDs). The cationic dendrimers prefer to destabilize cell membranes and cause cell lysis and the cytotoxicity is generation dependent with higher generation dendrimers being the most toxic. The degree of substitution as well as the type of amine functionality is important, with primary amines being more toxic than secondary or tertiary amines. Another common dendrimer, poly(propylenimine) (PPI), shows similar behavior. However, anionic dendrimers show significantly lower cytotoxicity than cationic ones. PEG or fatty acid surface-modified dendrimers can reduce the cytotoxicity of cationic dendrimers [124].

*Carbon Nanotubes, Fullerenes, and Quantum Dots* Carbon nanotubes and fullerenes are carbon-based nanomaterials, and quantum dots are semiconductor nano-

crystals. All of them show hydrophobic property. The possibility of cytotoxicity of these materials with inorganic nature should not be ignored although low toxicity is shown [25, 40, 125, 126]. However, these seemingly good results may be partly attributed to their poor solubility in polar solvents, which subsequently makes investigation of their biological properties difficult. They hardly load any drugs unless the surface is modified hydrophilically. Because these nanomaterials are mainly produced in laboratories with the special devices, their modifications and subsequent pharmacological investigations are limited. However, a number of functional derivatives have been synthesized, and it is found that the modified products have potent and selective pharmacological effects on organs, cells, enzymes, and nucleic acids [25, 37, 57].

### 3.2 Nanosuspensions

Nanosuspensions of drugs are submicrometer colloidal dispersions of pure particles of drug which are stabilized by surfactants. A surprisingly large proportion of new drug candidates emerging from drug discovery programs are water insoluble, and therefore poorly bioavailable, leading to development efforts being abandoned. More than 40% of active substances during formulation development by the pharmaceutical industry are poorly water soluble. A substantial factor that prevents the development of such substances is the limited dissolution rate. Nanosuspensions are promising in addressing these so-called brickdust candidates. During the process of overcoming issues involving solubility, the additional pharmacokinetic benefits of drugs formulated in nanosuspensions come to be appreciated [19, 61].

Nanosuspensions can be used for those water-insoluble and oil-soluble compounds (high log $P$), although other lipidic carriers such as liposomes and emulsions can be used to formulate these compounds as well. However, nanosuspensions can be used to address other problems, such as compounds that are insoluble in both water and oil. Nanosuspensions can maintain the drug in a preferred crystalline state of size sufficiently small for pharmaceutical acceptability. For reasons of convenience to the patients, aqueous nanosuspensions can also be transformed to tablets or capsules after spray drying or freeze drying. Moreover, utilization of the dense, solid state confers an additional advantage of higher mass per volume loading. This is crucial when high dosing is required, for example, low-volume intramuscular and ophthalmic applications. Conventional approaches often attempt to solubilize insoluble drugs with the use of excessive amounts of cosolvents, but this often brings problems of toxicity. Besides, very large doses of drugs must be administered to animals when acute toxicity is investigated in preclinical research. As a result, the interference of toxic side effects caused by cosolvents cannot be ignored if using them [19].

Nanosuspensions are not nanocarriers so that what is emphasized during manufacturing is not materials but the manufacturing techniques. Only drugs and stabilizers (usually surfactants) participate in manufacturing nanosuspensions so that the process may be simple depending on drug instincts but sometimes it is not easy. The bottom-up and top-down approaches may be used in manufacturing nanosuspensions depending on drug nature and in-house devices.

Antisolvent precipitation is a bottom-up method wherein two phases are involved: the initial creation of crystal nuclei of drugs and the subsequent growth. Formation

of a stable suspension with the smallest particle size requires a high nucleation rate but low growth rate. Both process rates are dependent on temperature and supersaturation. The optimum temperature for nucleation might lie below that for growth. A high-supersaturation condition is achieved by adding small amounts of a water-miscible organic solution of the drug to the nonsolvent (water) under rapid mixing, which leads to spontaneous nucleation. At high-supersaturation levels, the crystal habit or external appearance is changed to a needlelike or dendritic morphology. These crystals are easily broken, forming new smaller nuclei. Rapidly grown crystals tend to be more imperfect and often incorporate impurities and dislocations. This effect is more pronounced for flexible molecules that have many degrees of freedom [19]. The presence of stabilizing surfactants is generally necessary to assist in forming submicrometer particles, and hydrophilic groups in the surfactants lead to rapid wetting of the high-surface-area particles in aqueous media, for example, in the case of oral administration. It is well known that the unprotective surfaces of nanoparticles show a high energy that leads to particle agglomeration. Therefore, the nanoparticles must be protected by, for example, steric hindrance and electrostatic pulsion. In the case of itraconazole (ITZ) nanosuspension manufacturing, a mixture solution of ITZ and Poloxamer (P407) in THF at room temperature was mechanically injected into a P407 aqueous solution at 3°C. Magnetic stirring was utilized to enhance heat and mass transfer. Nanosuspensions containing sub-300-nm particles were obtained with drug loads as high as 86% [60].

Top-down methods are also commonly used to manufacture nanosuspensions, including wet milling and homogenization. In pearl/ball milling, the active agent, in the presence of surface stabilizer(s), is comminuted by milling media. Particle size is determined by stress intensity and the number of contact points. The drug macrosuspensions are poured into a milling container containing milling pearls from, for example, glass, zircon oxide or special polymers such as hard polystyrene derivatives. The drugs are ground to nanocrystals between the pearls. The nanosuspension-derived products, Rapamune (sirolimus tablets) and Emend (aprepitant capsules), were approved by the FDA and launched in 2000 and 2003, respectively. They are manufactured by Elan's NanoCrytal technology using a proprietary wet-milling technique. A general problem of pearl mills is potential erosion of materials from the milling pearls leading to product contamination. A polymer as substitution may minimize erosion. Scaling up with pearl mills is possible; however, there is a certain limitation in the size of the mill due to its weight. Up to about two-thirds of the mill volume are the pearls lead to heavy weight of the machinery, thus limiting the maximum batch size [127].

Homogenization can be divided into two types. One is the forcing of a suspension under pressure through a narrow-aperture valve (microfluidization). The other is high-pressure homogenization of particles in water or other media (piston gap). Microfluidization is a jet stream principle. The suspension is accelerated and passes a specially designed homogenization chamber with a high velocity. In the Z-type chamber, the suspension changes the direction of its flow a few times, leading to particle collision and shear forces, while in the Y-type chamber, the suspension stream is divided into two streams colliding frontally. Sometimes it is necessary to pass through the microfluidizer many times to minimize particle size [127].

In piston-gap homogenization, suspension contained in a cylinder passes a very thin gap with an extremely high velocity. Bubbles of water vapor are produced for

compensation followed by collapsing in the valve. Cavitation-induced shock waves occur and crack the particles. Homogenization can also be utilized to further reduce the size of particles made by precipitation. Commonly dendritic crystals made by precipitation are more susceptible to rupture by the subsequent mechanical shock of homogenization. In addition, the mechanical energy supplied by the homogenizer can change the initially formed, unstable amorphous particles to a stable state through subsequent crystallization. The size of drug nanocrystals depends mainly on (a) power density of the homogenizer, (b) number of homogenization cycles, and (c) temperature. Another important determining factor for the final size of drug nanocrystals is the hardness of drugs. A relatively soft drug, paclitaxel, can reach 250 nm in size, which is less than harder drugs. The size should be homogeneous as achieved with a homogenizer to avoid physical destabilization. Stabilizers have an effect on long-term physical stability but not on maximum dispersity or the nanocrystal shape. Contamination from the production equipment is typically below 1 ppm, which is within a suitable range. Besides water, water-free media and water mixtures are used preferably due to advantages of easy evaporation or homogenization at higher temperature (with subsequent more cavitation). Oils, propylene glycol, and PEG with varieous molecular weights can be used. For PEG being solid at room temperature (e.g., PEG 1000, 6000), the obtained drug nanocrystals disperse in PEG particles at room temperature and can conveniently be put into hard capsules [19, 127].

The lyophilized drug nanosuspensions can be transferred to a final dry oral dosage form such as tablets or reconstituted prior to administration. Drug nanosuspensions can be directly used as parenteral products. A shelf life of up to three years was shown for selected nanosuspensions. Sterilization can be achieved by aseptic processing of previously sterilized components, membrane filtration for particles sufficiently small or for drugs that can withstand it, steam sterilization, or γ-irradiation.

### 3.3 Self-Assembled Drug Nanostructures

It is well known that liposomes are composed of amphiphilic phospholipids. The formation of liposomes is actually a procedure of molecular self-assembly. Furthermore, great amounts of amphiphilic compounds, natural or synthesized rather than phospholipid-like surfactants, can also self-assemble into ordered aggregates in aqueous media or organic solvents [128–130]. The formed aggregates are mostly nanoarchitectures with various shapes such as vesicle, rod, ribbon, fiber, tube, or helix [131–134]. They can remain relatively stable in certain environments. Many of them may become drug nanocarriers, such as liposomes, or even perform as active agents [109, 135, 136]. The research of self-assembled nanocarriers seems to go into the field of supramolecular chemistry. But these results also give us some useful information for developing new approaches of drug delivery. A novel idea may relate to why we do not try to construct a nanostructure from drugs themselves.

Twenty years ago the cardiovascular drug pindolol was conjugated with stearyl glycerol via succinyl as linker followed by forming maleate salt to obtain pindolol diglyceride. Vaizoglu and Speiser used the word “pharmacosomes” to describe the colloidal dispersions prepared from drug–lipid conjugates with or without additional surfactants [137]. Pindolol pharmacosomes (vesicle-like) were prepared from

pindolol diglyceride, which showed good stability and useful pharmacokinetic parameters. Unfortunately, no more detailed research is being done about pharmacosomes, possibly because no appropriate theory supports the new dosage form and no proper drugs and lipids are selected.

The idea of manufacturing nanostructures from drugs may be resourced from liposomes, pharmacosomes, and other molecular self-assemblies in supramolecular chemistry. More importantly, this novel idea resulted from long-term efforts to work on drug delivery and to solve the disadvantages of current drug carriers. Almost all current delivery systems (usually called carriers) passively load drugs so as to always lead to low entrapment efficiency and possible drug leakage in preparation, preservation, and transportation in vivo [30, 73], and these carriers might have been destroyed in vivo before reaching target sites. In addition, lipophilic biomembranes, including cell membranes, usually prevent hydrophilic drugs from entering into target sites. If carriers cannot override cell membranes except for endocytosis/phagocytosis by cells/macrophages, the loaded hydrophilic drugs are probably released only on target surfaces, not entering. In summary, a majority of drugs could not eventually reach and get into target sites due to poor properties of carriers and drugs.

A novel technology involving prodrug, molecular self-assembly, and nanotechnology was developed to address the problems of drugs and classical carriers. The nanostructures are formed by molecular self-assembly of amphiphilic prodrugs in aqueous media generally without additional excipients. The self-assembled drug nanostructures not only possess the amphiphilic property of monomolecular drugs, benefiting to cross biomembranes, but also deliver themselves in vivo without "carriers" and then prefer to release active parent agents with a sustained rate. They may overcome some deficiencies of traditional nanocarriers such as liposomes, for example, low efficiency of drug entrapment and loading, rapid drug leakage in vitro/in vivo, and bad stability [22].

*Self-assembled drug nanostructure* is not a proprietary term in pharmacy currently. Herein this term is defined as the ordered nanosized self-aggregates of amphiphilic drugs in aqueous media. It is abbreviated as SADN. Another term, *self-assembled drug delivery system* (SADDS), introduced by Jin [22] obviously includes SADN. Unfortunately, most current drugs do not occupy an amphiphilic and self-assembling nature [138], so they must be modified in chemical structures before manufacturing SADNs. Then the prodrug technique is selected.

In contrast to nanosuspension technology, SADN technology is mainly applied to hydrophilic or polar drugs. These drugs are rationally modified to their amphiphilic prodrugs by lipid derivation. Molecular self-assembly in aqueous media is the key to manufacturing SADNs. According to the principles of supramolecular chemistry, the amphiphilic molecules forming self-assemblies should have proper structural conformation. The morphology of assemblies also depends on the structure of amphiphiles and the surrounding environment. Some parameters, including the optimal head area $a_o$, the volume $v$ of fluid hydrophobic chain, and the maximum efficient chain length $l_c$, are used to describe the conformation of amphiphiles. The critical packing parameter (CPP), equal to $v/a_o l_c$, can be applied to direct self-assembly behavior. The amphiphiles prefer to form vesicles when the CPP is –1. Generally, single-chain lipids with small head group areas (e.g., SDS in a low-salt solution) are cone shaped, prone to form spherical micelles, while double-chain

lipids with large head group area and fluid chains (e.g., phosphatidylcholine) are truncated-cone shaped, prone to form flexible bilayers, vesicles [139]. Therefore, the lipids used for drug covalent conjugation are rationally selected from long-chain alkyl lipids, for example, fatty acids, lipid alcohols, lipid amines, long-chain glycolipids, and cholesterol. Furthermore, too large or small polar drugs are not appropriate for preparation of self-assembling prodrugs.

Antiviral nucleoside analogues such as acyclovir, didanosine, and zidovudine were used to prepare their long-chain glyceride or cholesteryl derivatives in Jin's laboratory [140–142]. All the derivatives showed amphiphilic property and some of them self-assembled into ordered aggregates in water. Amphiphilic prodrugs were subsequently used to manufacture self-assemblies using the bottom-up approach, such as liposomes, and the self-assembly may be driven by a hydrophobic interaction, hydrogen bonding, and so on [21, 143]. The monomolecular amphiphilic prodrug is prone to incorporate into the assemblies and not to depart so that almost no drugs leak from SADNs. The whole self-assemblies are nearly composed of amphiphilic drugs, leading to high drug loading. When SADNs reach targets in vivo, the continual dissociation of aggregates and the sustained degradation of prodrugs provide controlled drug release.

Acyclovir self-assembled nanoparticles as SADNs were manufactured which showed strong targeting effect in vivo (mainly in the MPSs) and sustained release at target sites [22]. Based on this paradigm, a general process to manufacture SADNs is as follows [21, 22]:

(a) To obtain an amphiphilic prodrug with proper molecular structure, stearyl glyceride was selected to conjugate with acyclovir. Succinyl acyclovir (SACV) was synthesized and subsequently conjugated with stearyl glyceride by acylation reaction. The amphiphilic prodrug stearyl-glyceride-succinyl acyclovir (SGSA) was obtained.
(b) The injection method was used to manufacture SADNs. SGSA was dissolved in the water-miscible solvent THF. The solution containing 5 mg/mL SGSA was slowly and continually injected into vortexed water under surface via a 100-μL microsyringe. A homogeneous and slightly opalescent suspension was obtained, which was acyclovir self-assembled nanoparticles (SANs).
(c) The organic solvent was removed from the suspension through evaporation by heating, and the suspension can further be concentrated by removing water under heating until the appropriate prodrug concentration is obtained. The concentrated suspension was transferred into ampoules and sealed. It may be sterilized by autoclave.
(d) Acyclovir SANs were characterized. They were cuboidlike shaped based on transmission electron microscopy and were nanoscale with an average size of 83 nm based on dynamic light scattering. The zeta potential of –31 mV indicated the nanoparticles had negative surface charge. Hydrophobic interaction of alkyl chains improves SGSA molecules to form bilayers, and then cuboidlike nanoparticles were achieved by layer-by-layer aggregation based on inter-bilayer hydrogen bonding. The gel–liquid crystalline phase transition was about 50°C, and the mechanism of configuration changes on phase transition was analyzed [144].

(e) The in vitro and in vivo behavior of acyclovir SANs was investigated. The SANs kept the physical state stable upon centrifugation or exposure of some common additives Autoclave and bath heat for sterilization hardly influenced the state of SANs. SGSA in SANs showed good chemical stability in weak acidic or neutral buffers, although they were very sensitive to alkaline solutions and carboxylester enzymes. The SANs were rapidly removed from blood circulation after bolus IV administration to rabbits and mainly distributed in liver, spleen, and lung followed by slow elimination in these tissues.

Because nucleoside analogues are important and plentiful agents in antiviral and anticancer therapy, other polar drugs can simulate the above process to manufacture SADNs. In addition, macrophages as the reservoirs of HIV or other viruses prefer to carry viruses throughout the whole body even to the central nervous system. How to deliver drugs to macrophages has become a key issue in antiviral therapy [145]. SADNs prefer to show macrophage-specific distribution. Therefore, the antiviral SADNs show the advantages of high drug loading, controlled release, and targeting macrophage, which may provide a useful and promising way to treat increasing viral diseases. In the future SADNs will be modified to get more functions, such as long circulating effect, pH sensitivity, and use in antiviral, anticancer, and gene therapy.

## 4 NANOMEDICINE

Nanomedicine is a concept with broad implications. According to the definition of the European Science Foundation (ESF), the field of nanomedicine is the science and technology of diagnosing, treating, and preventing disease and traumatic injury, of relieving pain, and of preserving and improving human health using molecular tools and molecular knowledge of the human body. It is perceived as embracing five main subdisciplines that in many ways are overlapping and underpinned by the following common technical issues: (a) analytical tools, (b) nanoimaging, (c) nanomaterials and nanodevices, (d) novel therapeutics and drug delivery systems, and (e) clinical, regulatory, and toxicological issues. The ESF's scientific forward look on nanomedicine warns that nanomedicine benefits will be lost without major investment and calls for a coordinated European strategy to deliver new nanotechnology-based medical tools for diagnostics and therapeutics [146]. From a view of narrow sense, nanomedicine can be defined as the use of nanoscale or nanostructured materials in medicine that have unique medical effects according to their structure. In addition, nanostructures up to 1000 nm in size are adopted because from a technical point of view the control of materials in this size range not only results in new medical effects but also requires novel, scientifically demanding chemistry and manufacturing techniques [147].

The increasing research in nanomedicine has led to many publications, accounting for about 4% of publications on nanotechnology research (about 34,300 documents in 2004) worldwide. Also commercialization efforts in nanomedicine are increasing. About 207 companies (including 158 small- and medium-size enter-

prises) visibly pursue nanomedicine activities and devote either all or a significant share of their business to the development of nanomedicines. A characterizing feature of nanotechnology is its enabling function to add new functionality to existing products, making them more competitive. For example, Ambisome (Gilead), a liposomal formulation of the fungicide Fungizone (Bristol-Myers Squibb) that shows reduced kidney toxicity, had total sales of $212 million in 2004. The total sales of the 38 identified nanomedicine products from all sectors of nanomedicine are estimated to be $6.8 billion in 2004. The market is predicted to further grow to ~$12 billion by the year 2012. Currently, nanomedicine is dominated by drug delivery systems, accounting for more than 75% of the total sales. Twenty-three nanoscale drug delivery systems are available on the market, but within this group, three polymer therapeutics alone account for sales of $3.2 billion: (i) Neulasta (pegfilagrastim; recombinant methionyl human granulocyte colony stimulating factor and PEG), (ii) Pegasys (PEGylated interferon α 2a), and (iii) PEG-Intron (PEGylated interferon α 2a), all protein therapeutics to which nanoscale polymer strings of PEG have been attached to reduce immunogenicity and to prolong plasma half-life. The most widely used nanotechnology product in the field of in vitro diagnostics is colloidal gold in lateral flow assays, rapid tests for pregnancy, ovulation, HIV, and other indications. Magnetic nanoparticles are also used for cell-sorting applications in clinical diagnostics. In the field of biomaterials, the commercial status of nanotechnology-based dental restoratives is most advanced. Furthermore, nanohydroxyapatite-based products for the repair of bone defects have been successfully commercialized. Nanotechnology-based contrast agents are a market with estimated sales of about $12 million. All of the marketed contrast agents consist of superparamagnetic iron oxide nanoparticles for magnetic resonance imaging. Nanostructured electrodes are used to improve the electrode tissue contact, and nanomaterials are used to increase the biocompatibility of implant housings. Pacemakers with nanostructured (fractal) electrodes are the only active implants currently on the market that contain a nanotechnology-enabled component [147].

In spite of the great success, the safety of nanomedicine is maintained as a worrying issue. A new discipline appears to exploit the toxicological problem in nanotechnology applications, called nanotoxicology. Nanotoxicology can be defined as safety evaluation of engineered nanostructures and nanodevices. Nanomaterials could be deposited in all regions of the respiratory tract after inhalation. The small size facilitates uptake into cells and transcytosis across epithelial and endothelial cells into the blood and lymph circulation to reach potentially sensitive target sites such as bone marrow, lymph nodes, spleen, and heart. Access to the central nervous system and ganglia via translocation along axons and dendrites of neurons has also been observed. Nanomaterials could also penetrate the skin via uptake into lymphatic channels [41]. Although possible damages of those biodegradable nanomaterials for drug delivery need consideration, too much fear is needless. Usually they would be ultimately degraded nearly without any trace. However, hard or nonbiodegradable materials, including carbon nanotubes, fullerenes, quantum dots, polystyrene, and metal nanoparticles, should be thoroughly investigated about their toxic effects on our body before clinical application.

## 5 PERSPECTIVE

Nanotechnology has had a great effect on pharmaceutical manufacturing and strongly improves it, rapidly progressing. No one suspects the key role nanotechnology will have in future pharmaceutical research and manufacturing. The continually increasing achievements in nanotechnology will result in exciting changes in the pharmaceutical industry. Now it has gone into an era of controlling the behavior of drugs in vitro/in vivo. Although some problems such as toxicity are not addressed, the tremendous advantages that result from nanotechnology are obvious. More and more potent medicines will be manufactured and diseases such as cancer, HIV, cardiovascular diseases, and nervous system diseases may well be cured or better treated in the future by nanomedicine technology.

## REFERENCES

1. Silva, G. A. (2006), Neuroscience nanotechnology: Progress, opportunities and challenges, *Nat. Rev. Neurosci.*, 7, 65–74.
2. Wang, Z. L. (2000), *Characteriaztion of Nanophase Materials*, Wiley-VCH, Weinheim, Germany, pp. 1–6.
3. Vasir, J. K., Reddy, M. K., and Labhasetwar, V. D. (2005), Nanosystems in drug targeting: Opportunities and challenges, *Curr. Nanosci.*, 1, 47–64.
4. Lane, N. F. (2004), in Bhushan, B., Ed., *Springer Handbook of Nanotechnology*, Springer-Verlag, Berlin.
5. Goodsell, D. S. (2004), *Bionanotechnology: Lessons from Nature*, Wiley-Liss, Hoboken, NJ, pp. 1–6.
6. Fritz, J., Baller, M. K., Lang, H. P., Rothuizen, H., Vettiger, P., Meyer, E., Guntherodt, H. J., Gerber, C., and Gimzewski, J. K. (2000), Translating biomolecular recognition into nanomechanics, *Science*, 288, 316–318.
7. Xu, X. H. N., Chen, J., Jeffers, R. B., and Kyriacou, S. (2002), Direct measurement of sizes and dynamics of single living membrane transporters using nanooptics, *Nano Lett.*, 2, 175–182.
8. Li, J., Papadopoulos, C., and Xu, J. (1999), Nanoelectronics—Growing Y-junction carbon nanotubes, *Nature*, 402, 253–254.
9. Sergeev, G. B. (2006), *Nanochemistry*, Elsevier, Amsterdam, The Netherlands.
10. Moghimi, S. M., Hunter, A. C., and Murray, J. C. (2005), Nanomedicine: Current status and future prospects, *FASEB J.*, 19, 311–330.
11. Whitesides, G. M. (2003), The "right" size in nanobiotechnology, *Nat. Biotechnol.*, 21, 1161–1165.
12. Lee, K. B., Park, S. J., Mirkin, C. A., Smith, J. C., and Mrksich, M. (2002), Protein nanoarrays generated by dip-pen nanolithography, *Science*, 295, 1702–1705.
13. Liu, G. Y., and Amro, N. A. (2002), Positioning protein molecules on surfaces: A nanoengineering approach to supramolecular chemistry, *Proc. Nat. Acad. Sci. U.S.A.*, 99, 5165–5170.
14. Quake, S. R., and Scherer, A. (2000), From micro- to nanofabrication with soft materials, *Science*, 290, 1536–1540.
15. Xia, Y. N., McClelland, J. J., Gupta, R., Qin, D., Zhao, X. M., Sohn, L. L., Celotta, R. J., and Whitesides, G. M. (1997), Replica molding using polymeric materials: A practical step toward nanomanufacturing, *Adv. Mater.*, 9, 147–149.

16. http://www.whitehouse.gov/omb/budget/fy2007/nsf.html 2007-1-5.
17. Service, R. F. (2005), Nanotechnology takes aim at cancer, *Science*, 310, 1132–1134.
18. Vauthey, S., Santoso, S., Gong, H., Watson, N., and Zhang, S. (2002), Molecular self-assembly of surfactant-like peptides to form nanotubes and nanovesicles, *Proc. Nat. Acad. Sci. U. S. A.*, 99, 5355–5360.
19. Rabinow, B. E. (2004), Nanosuspensions in drug delivery, *Nat. Rev. Drug Discov.*, 3, 1–12.
20. Vinogradov, S. V. (2006), Colloidal microgels in drug delivery applications, *Curr. Pharm. Des.*, 12, 4703–4712.
21. Jin, Y., Qiao, Y., Li, M., Ai, P., and Hou, X. (2005), Langmuir monolayers of the long-chain alkyl derivatives of a nucleoside analogue and the formation of self-assembled nanoparticles, *Coll. Surf. B: Biointerf.*, 42, 45–51.
22. Jin, Y., Tong, L., Ai, P., Li, M., and Hou, X. (2006), Self-assembled drug delivery systems. 1. Properties and in vitro/in vivo behavior of acyclovir self-assembled nanoparticles (SAN), *Int. J. Pharm.*, 309, 199–207.
23. Boas, U., and Heegaard, P. M. H. (2004), Dendrimers in drug research, *Chem. Soc. Rev.*, 33, 43–63.
24. Kroto, H. W., Heath, J. R., O'Brien, S.C., Curl, R. F., and Smalley, R. E. (1985), *Nature*, 318, 162–163.
25. Satoh, M., and Takayanagi, I. (2006), Pharmacological studies on fullerene (C60), a novel carbon allotrope, and its derivatives, *J. Pharm. Sci.*, 100, 513–518.
26. Michalet, X., Pinaud, F. F., Bentolila, L. A., Tsay, J. M., Doose, S., Li, J. J., Sundaresan, G., Wu, A. M., Gambhir, S. S., and Weiss, S. (2005), Quantum dots for live cells, in vivo imaging, and diagnostics, *Science*, 307, 538–544.
27. Muller, R. H., Mader, K., and Gohla, S. (2000), Solid lipid nanoparticles (SLN) for controlled drug delivery—A review of the state of the art, *Eur. J. Pharm. Biopharm.*, 50, 161–177.
28. Groneberg, D. A., Giersig, M., Welte, T., and Pison, U. (2006), Nanoparticle-based diagnosis and therapy, *Curr. Drug Targets*, 7, 643–648.
29. Couvreur, P., Barratt, G., Fattal, E., Legrand, P., and Vauthier, C. (2002), Nanocapsule technology: A review, *Crit. Rev. Ther. Drug Carrier Syst.*, 19, 99–134.
30. Barenholz, Y. (2001), Liposome application: Problems and prospects, *Curr. Opin. Coll. Interf. Sci.*, 6, 66–77.
31. Torchilin, V. P. (2005), Recent advances with liposomes as pharmaceutical carriers, *Nat. Rev. Drug Discov.*, 4, 145–160.
32. Uchegbu, I. F., and Vyas, S. P. (1998), Non-ionic surfactant based vesicles (niosomes) in drug delivery, *Int. J. Pharm.*, 172, 33–70.
33. Sadurni, N., Solans, C., Azemara, N., and Garcia-Celma, M. J. (2005), Studies on the formation of O/W nano-emulsions, by low-energy emulsification methods, suitable for pharmaceutical applications, *Eur. J. Pharm. Sci.*, 26, 438–445.
34. Solans, C., Izquierdo, P., Nolla, J., Azemara, N., and Garcia-Celma, M. J. (2005), Nano-emulsions, *Curr. Opin. Coll. Interf. Sci.*, 10, 102–110.
35. Jones, M.-C., and Leroux, J.-C. (1999), Polymeric micelles—A new generation of colloidal drug carriers, *Eur. J. Pharm. Biopharm.*, 48, 101–111.
36. Torchilin, V. P. (2007), Micellar nanocarriers: Pharmaceutical perspectives, *Pharm. Res.*, 24, 1–16.
37. Bianco, A., Kostarelos, K., and Prato, M. (2005), Applications of carbon nanotubes in drug delivery, *Curr. Opin. Chem. Biol.*, 9, 674–679.

38. Foley, S., Crowley, C., Smaihi, M., Bonfils, C., Erlanger, B. F., Seta, P., and Larroqueb, C. (2002), Cellular localisation of a water-soluble fullerene derivative, *Biochem. Biophys. Res. Commun.*, 294, 116–119.
39. Tsao, N., Luh, T.-Y., Chou, C.-K., Chang, T.-Y., Wu, J.-J., Liu, C.-C., and Lei, H.-Y. (2002), In vitro action of carboxyfullene, *J. Antimicrob. Chemother.*, 49, 641–649.
40. Ozkan, M. (2004), Quantum dots and other nanoparticles: What can they offer to drug discovery, *Drug Discov. Today*, 9, 1065–1071.
41. Oberdorster, G., Oberdorster, E., and Oberdorster, J. (2005), Nanotoxicology: An emerging discipline evolving from studies of ultrafine particles, *Environ. Health Perspect.*, 113, 823–839.
42. des Rieux, A., Fievez, V., Garinot, M., Schneider, Y.-J., and Preat, V. (2006), Nanoparticles as potential oral delivery systems of proteins and vaccines: A mechanistic approach, *J. Controlled Release*, 116, 1–27.
43. Lemos-Senna, E., Wouessidjewe, D., Lesieur, S., and Duchene, D. (1998), Preparation of amphiphilic cyclodextrin nanospheres using the emulsification solvent evaporation method. Influence of the surfactant on preparation and hydrophobic drug loading, *Int. J. Pharm.*, 170, 119–128.
44. Xiao, S., Tong, C., Liu, X., Yu, D., Liu, Q., Xue, C., Tang, D., and Zhao, L. (2006), Preparation of folate-conjugated starch nanoparticles and its application to tumor-targeted drug delivery vector, *Chin. Sci. Bull.*, 51, 1693–1697.
45. Sokolova, V. V., Radtke, I., Heumann, R., and Epple, M. (2006), Effective transfection of cells with multi-shell calcium phosphate-DNA nanoparticles, *Biomaterials*, 27, 3147–3153.
46. Ueno, Y., Futagawa, H., Takagi, Y., Ueno, A., and Mizushima, Y. (2005), Drug-incorporating calcium carbonate nanoparticles for a new delivery system, *J. Controlled Release*, 103, 93–98.
47. Souza, G., Christianson, D. R., Staquicini, F. I., Ozawa, M. G., Snyder, E. Y., Sidman, R. L., Miller, J. R., Arap, W., and Pasqualini, R. (2006), Networks of gold nanoparticles and bacteriophage as biological sensors and cell-targeting agents, *Proc. Nat. Acad. Sci. U.S.A.*, 103, 1215–1220.
48. Guinebretiere, S., Briancon, S., Lieto, J., Mayer, C., and Fessi, H. (2002), Study of the emulsion-diffusion of solvent: Preparation and characterization of nanocapsules, *Drug Dev. Res.*, 57, 18–33.
49. Hillaireau, H., Doan, T. L., Appel, M., and Couvreur, P. (2006), Hybrid polymer nanocapsules enhance in vitro delivery of azidothymidine-triphosphate to macrophages, *J. Controlled Release*, 116, 346–352.
50. New, R. R. C. (1990), *Liposomes: A Practical Approach*, Oxford University Press, Oxford, pp. 1–162.
51. Vinogradov, S., Batrakova, E., and Kabanov, A. (1999), Poly(ethylene glycol)-polyethylenimine NanoGel particles: Novel drug delivery systems for antisense oligonucleotides, *Coll. Surf. B: Biointerf.*, 16, 291–304.
52. Vinogradov, S. V., Kohli, E., and Zeman, A. D. (2005), Cross-linked polymeric nanogel formulations of 5′-triphosphates of nucleoside analogues: Role of the cellular membrane in drug release, *Mol. Pharm.*, 2, 449–461.
53. Bosman, A. W., Janssen, H. M., and Meijer, E. W. (1999), About dendrimers: Structure, physical properties, and applications, *Chem. Rev.*, 99, 1665–1688.
54. Gagner, J., Johnson, H., Watkins, E., Li, Q., Terrones, M., and Majewski, J. (2006), Carbon nanotube supported single phospholipid bilayer, *Langmuir*, 22, 10909–10911.

55. Kam, N. W. S., Liu, Z., and Dai, H. (2006), Carbon nanotubes as intracellular transporters for proteins and DNA: An investigation of the uptake mechanism and pathway, *Ang. Chem. Int. Ed.*, 45, 577–581.
56. Zakharian, T. Y., Seryshev, A., Sitharaman, B., Gilbert, B. E., Knight, V., and Wilson, L. J. (2005), A fullerene-paclitaxel chemotherapeutic: Synthesis, characterization, and study of biological activity in tissue culture, *J. Am. Chem. Soc.*, 127, 12508–12509.
57. Akerman, M. E., Chan, W. C. W., Laakkonen, P., Bhatia, S. N., and Ruoslahti, E. (2002), Nanocrystal targeting in vivo, *Proc. Nat. Acad. Sci. U.S.A.*, 99, 12617–12621.
58. Boulmedais, F., Bauchat, P., Brienne, M. J., Arnal, I., Artzner, F., Gacoin, T., Dahan, M., and Marchi-Artzner, V. (2006), Water-soluble pegylated quantum dots: From a composite hexagonal phase to isolated micelles, *Langmuir*, 22, 9797–9803.
59. Algar, W. R., and Krull, U. J. (2006), Adsorption and hybridization of oligonucleotides on mercaptoacetic acid-capped CdSe/ZnS quantum dots and quantum dot-oligonucleotide conjugates, *Langmuir*, 22, 11346–11352.
60. Matteucci, M. E., Hotze, M. A., Johnston, K. P., and Williams, III. R. O. (2006), Drug nanoparticles by antisolvent precipitation: Mixing energy versus surfactant stabilization, *Langmuir*, 22, 8951–8959.
61. Douroumis, D., and Fahr, A. (2006), Nano- and micro-particulate formulations of poorly water-soluble drugs by using a novel optimized technique, *Eur. J. Pharm. Biopharm.*, 63, 173–175.
62. Torchilin, V. P. (2006), Multifunctional nanocarriers, *Adv. Drug Deliv. Rev.*, 58, 1532–1555.
63. Banham, A. D., Standish, M. M., and Watkins, J. C. (1965), Diffusion of univalent ions across the lamellae of swollen phospholipids, *J. Mol. Biol.*, 13, 238–252.
64. Lasic, D. D., Joannic, R., Keller, B. C., Frederik, P. M., and Auvray, L. (2001), Spontaneous vesiculation, *Adv. Coll. Interf. Sci.*, 89–90, 337–349.
65. Bhalerao, S. S., and Harshal, A. R. (2003), Preparation, optimization, characterization, and stability studies of salicylic acid liposomes, *Drug Dev. Ind. Pharm.*, 29, 451–467.
66. Szoka, F., and Papahadjopoulos, D. (1978), Procedure for preparation of liposomes with large internal aqueous space and high capture by reverse-phase evaporation, *Proc. Nat. Acad. Sci. U. S. A.*, 75, 4194–4198.
67. Pons, M., Foradada, M., and Estelrich, J. (1993), Liposomes obtained by the ethanol injection method, *Int. J. Pharm.*, 95, 51–56.
68. Kikuchi, H., Yamauchi, H., and Hirota, S. A. (1994), Polyol dilution method for mass production of liposomes, *J. Liposome Res.*, 4, 71–91.
69. Wang, T., Deng, Y., Geng, Y., Gao, Z., Zou, J., and Wang, Z. (2006), Preparation of submicron unilamellar liposomes by freeze-drying double emulsions, *Biochim. Biophys. Acta*, 1758, 222–231.
70. Ning, M.-Y., Guo, Y.-Z., Pan, H.-Z., Yu, H.-M., and Gu, Z.-W. (2005), Preparation and evaluation of proliposomes containing clotrimazole, *Chem. Pharm. Bull.*, 53, 620–624.
71. Barnadas-Rodriguez, R., and Sabes, M. (2001), Factors involved in the production of liposomes with a high-pressure homogenizer, *Int. J. Pharm.*, 213, 175–186.
72. Taira, M. C., Chiaramoni, N. S., Pecuch, K. M., and Alonso-Romanowski, S. (2004), Stability of liposomal formulations in physiological conditions for oral drug delivery, *Drug Deliv.*, 11, 123–128.
73. Glavas-Dodov, M., Fredro-Kumbaradzi, E., Goracinova, K., Simonoska, M., Calis, S., Trajkovic-Jolevska, S., and Hincal, A. A. (2005), The effects of lyophilization on the stability of liposomes containing 5-FU, *Int. J. Pharm.*, 291, 79–86.

74. Liang, W., Levchenko, T. S., and Torchilin, V. P. (2004), Encapsulation of ATP into liposomes by different methods: Optimization of the procedure, *J. Microencapsul.*, 21, 251–261.
75. Haran, G., Cohen, R., Bar, L. K., and Barenholz, Y. (1993), Transmembrane ammonium sulfate gradients in liposomes produce efficient and stable entrapment of amphipathic weak bases, *Biochim. Biophys. Acta*, 1151, 201–215.
76. Zhigaltsev, I. V., Kaplun, A. P., Kucheryanu, V. G., Kryzhanovsky, G. N., Kolomeichuk, S. N., Shvets, V. I., and Yurasov, V. V. (2001), Liposomes containing dopamine entrapped in response to transmembrane ammonium sulfate gradient as carrier system for dopamine delivery into the brain of parkinsonian mice, *J. Liposome Res.*, 11, 55–71.
77. Hwang, S. H., Maitani, Y., Qi, X. R., Takayama, K., and Nagai, T. (1999), Remote loading of diclofenac, insulin and fluorescein isothiocyanate labeled insulin into liposomes by pH and acetate gradient methods, *Int. J. Pharm.*, 179, 85–95.
78. Abraham, S. A., Edwards, K., Karlsson, G., Hudon, N., Mayer, L. D., and Bally, M. B. (2004), An evaluation of transmembrane ion gradient-mediated encapsulation of topotecan within liposomes, *J. Controlled Release*, 96, 449–461.
79. Gulati, M., Grover, M., Singh, S., and Singh, M. (1998), Lipophilic drug derivatives in liposomes, *Int. J. Pharm.*, 165, 129–168.
80. Wagner, A., Vorauer-Uhl, K., and Katinger, H. (2002), Liposomes produced in a pilot scale: Production, purification and efficiency aspects, *Eur. J. Pharm. and Biopharm.*, 54, 213–219.
81. Wagner, A., Vorauer-Uhl, K., Kreismayr, G., and Katinger, H. (2002), The crossflow injection technique—An improvement of the ethanol injection method, *J. Liposome Res.*, 12, 259–270.
82. Wagner, A., Platzgummer, M., Kreismayr, G., Quendler, H., Stiegler, G., Ferko, B., Vecera, G., Vorauer-Uhl, K., and Katinger, H. (2006), GMP production of liposomes: A new industrial approach, *J. Liposome Res.*, 16, 311–319.
83. Woodle, M. C., and Scaria, P. (2001), Cationic liposomes and nucleic acids, *Curr. Opin. Coll. Interf. Sci.*, 6, 78–84.
84. Templeton, N. S. (2002), Cationic liposome-mediated gene delivery in vivo, *Biosci. Rep.*, 22, 283–295.
85. Ishida, T., Harashima, H., and Kiwada, H. (2002), Liposome clearance, *Biosci. Rep.*, 22, 197–224.
86. Winterhalter, M., Frederik, P. M., Vallner, J. J., and Lasic, D. D. (1997), Stealth® liposomes: From theory to product, *Adv. Drug Deliv. Rev.*, 24, 165–177.
87. Ishida, T., Ichihara, M., Wang, X., and Kiwada, H. (2006), Spleen plays an important role in the induction of accelerated blood clearance of PEGylated liposomes, *J. Controlled Release*, 115, 243–250.
88. Moghimi, S. M., Hamad, I., Andresen, T. L., Jorgensen, K., and Szebeni, J. (2006), Methylation of the phosphate oxygen moiety of phospholipid-methoxy(polyethylene glycol) conjugate prevents PEGylated liposome-mediated complement activation and anaphylatoxin production, *FASEB J.*, 20, 2591–2593.
89. Drummond, D. C., Zignani, M., and Leroux, J.-C. (2000), Current status of pH-sensitive liposomes in drug delivery, *Prog. Lipid Res.*, 39, 409–460.
90. Yatvin, M. B., Weinstein, J. N., Dennis, W. H., and Blumenthal, R. (1978), Liposomes and local hyperthermia: Selective delivery of methotrexate to heated tumors, *Science*, 202, 1290–1293.
91. Kono, K. (2001), Thermosensitive polymer-modified liposomes, *Adv. Drug Deliv. Rev.*, 53, 307–319.

92. Zhang, J. Q., Zhang, Z. R., Yang, H., Tan, Q. Y., Qin, S. R., and Qiu, X. L. (2005), Lyophilized paclitaxel magnetoliposomes as a potential drug delivery system for breast carcinoma via parenteral administration: In vitro and in vivo studies, *Pharm. Res.*, 22, 573–583.
93. Koning, G. A., Kamps, J. A. A. M., and Scherphof, G. L. S. (2002), Efficient intracellular delivery of 5-fluorodeoxyuridine into colon cancer cells by targeted immunoliposomes, *Cancer Detection Prevention*, 26, 299–307.
94. Mehnert, W., and Mader, K. (2001), Solid lipid nanoparticles: Production, characterization and applications, *Adv. Drug Deliv. Rev.*, 47, 165–196.
95. Shive, M. S., and Anderson, J. M. (1997), Biodegradation and biocompatibility of PLA and PLGA microspheres, *Adv. Drug Deliv. Rev.*, 28, 5–24.
96. Sommerfeld, P., Schroeder, U., and Sabel, B. A. (1997), Long-term stability of PBCA nanoparticle suspensions suggests clinical usefulness, *Int. J. Pharm.*, 155, 201–207.
97. Sommerfeld, P., Schroeder, U., and Sabel, B. A. (1998), Sterilization of unloaded polybutylcyanoacrylate nanoparticles, *Int. J. Pharm.*, 164, 113–118.
98. Chorny, M., Fishbein, I., Danenberg, H. D., and Golomb, G. (2002), Lipophilic drug loaded nanospheres prepared by nanoprecipitation: Effect of formulation variables on size, drug recovery and release kinetics, *J. Controlled Release*, 83, 389–400.
99. Govender, T., Stolnik, S., Garnett, M. C., Illum, L., and Davis, S. S. (1999), PLGA nanoparticles prepared by nanoprecipitation: Drug loading and release studies of a water soluble drug, *J. Controlled Release*, 57, 171–185.
100. Bilati, U., Allemann, E., and Doelker, E. (2005), Development of a nanoprecipitation method intended for the entrapment of hydrophilic drugs into nanoparticles, *Eur. J. Pharm. Sci.*, 24, 67–75.
101. Bilati, U., Allemann, E., and Doelker, E. (2005), Nanoprecipitation versus emulsion-based techniques for the encapsulation of proteins into biodegradable nanoparticles and process-related stability issues, *AAPS Pharm. Sci. Tech.*, 6, 74.
102. Astete, C. E., and Sabliov, C. M. (2006), Synthesis and characterization of PLGA nanoparticles, *J. Biomater. Sci. Polym. Ed.*, 17, 247–289.
103. Zwiorek, K., Kloechner, J., Wagner, E., and Coester, C. (2004), Gelatin nanoparticle as a new and simple gene delivery system, *J. Pharm. Pharm. Sci.*, 7, 22–28.
104. Cheng, J., Teply, B. A., Sherifi, I., Sung, J., Luther, G., Gu, F. X., Levy-Nissenbaum, E., Radovic-Moreno, A. F., Langer, R., and Farokhzad, O. C. (2007), Formulation of functionalized PLGA-PEG nanoparticles for in vivo targeted drug delivery, *Biomaterials*, 28, 869–876.
105. Hillery, A. M., and Florence, A. T. (1996), The effect of adsorbed poloxamer 188 and 407 surfactants on the intestinal uptake of 60-nm polystyrene particles after oral administration in the rat, *Int. J. Pharm.*, 132, 123–130.
106. Peracchia, M. T., Vauthier, C., Puisieux, F., and Couvreur, P. (1997), Development of sterically stabilized poly(isobutyl 2-cyanoacrylate) nanoparticles by chemical coupling of poly(ethylene glycol), *J. Biomed. Mater. Res.*, 34, 317–326.
107. Suk, J. S., Suh, J., Choy, K., Lai, S. K., Fu, J., and Hanes, J. (2006), Gene delivery to differentiated neurotypic cells with RGD and HIV Tat peptide functionalized polymeric nanoparticles, *Biomaterials*, 27, 5143–5150.
108. Aktas, Y., Yemisci, M., Andrieux, K., Gursoy, R. N., Alonso, M. J., FernAndez-Megia, E., Novoa-Carballal, R., Quinoa, E., Riguera, R., Sargon, M. F., Celik, H. H., Demir, A. S., Hincal, A. A., Dalkara, T., Capan, Y., and Couvreur, P. (2005), Development and brain delivery of chitosan-PEG nanoparticles: Functionalized with the monoclonal antibody OX26, *Bioconjugate Chem.*, 16, 1503–1511.

109. Park, J. H., Kwon, S., Nam, J.-O., Park, R.-W., Chung, H., Seo, S. B., Kim, I.-S., Kwon, I. C., and Jeong, S. Y. (2004), Self-assembled nanoparticles based on glycol chitosan bearing 5h-cholanic acid for RGD peptide delivery, *J. Controlled Release*, 95, 579–588.

110. Klier, J., Tucker, C. J., Kalantar, T. H., and Green, D. P. (2000), Properties and applications of microemulsions, *Adv. Mater.*, 12, 1751–1757.

111. Muller, R. H., Radtke, M., and Wissing, S. A. (2002), Solid lipid nanoparticles (SLN) and nanostructured lipid carriers (NLC) in cosmetic and dermatological preparations, *Adv. Drug Deliv. Rev.*, 54, S131-S155.

112. Goppert, T. M., and Muller, R. H. (2003), Plasma protein adsorption of Tween 80- and Poloxamer 188-stabilized solid lipid nanoparticles, *J. Drug Target.*, 11, 225–231.

113. Wong, H. L., Bendayan, R., Rauth, A. M., and Wu, X. Y. (2006), Simultaneous delivery of doxorubicin and GG918 (Elacridar) by new polymer-lipid hybrid nanoparticles (PLN) for enhanced treatment of multidrug-resistant breast cancer, *J. Controlled Release*, 116, 275–284.

114. Hu, F. Q., Hong, Y., and Yuan, H. (2004), Preparation and characterization of solid lipid nanoparticles containing peptide, *Int. J. Pharm.*, 273, 29–35.

115. Garcia-Fuentes, M., Torres, D., and Alonso, M. J. (2002), Design of lipid nanoparticles for the oral delivery of hydrophilic macromolecules, *Coll. Surf. B Biointerf.*, 27, 159–168.

116. Brannon-Peppas, L., and Blanchette, J. O. (2004), Nanoparticle and targeted systems for cancer therapy, *Adv. Drug Deliv. Rev.*, 56, 1649–1659.

117. Lockman, P. R., Mumper, R. J., Khan, M. A., and Allen, D. D. (2002), Nanoparticle technology for drug delivery across the blood-brain barrier, *Drug Dev. Ind. Pharm.*, 28, 1–13.

118. Nishioka, Y., and Yoshino, H. (2001), Lymphatic targeting with nanoparticulate system, *Adv. Drug Deliv. Rev.*, 47, 55–64.

119. Csaba, N., Caamano, P., Sanchez, A., Dominguez, F., and Alonso, M. J. (2005), PLGA: Poloxamer and PLGA:Poloxamine blend nanoparticles: New carriers for gene delivery, *Biomacromolecules*, 6, 271–278.

120. Bagwe, R. P., Kanicky, J. R., Palla, B. J., Patanjali, P. K., and Shah, D. O. (2001), Improved drug delivery using microemulsions: Rationale, recent progress, and new horizons, *Crit. Rev. Ther. Drug Carrier Syst.*, 18, 77–140.

121. Nayak, S., and Lyon, L. A. (2005), Soft nanotechnology with soft nanoparticles, *Ang. Chem. Int. Ed.*, 44, 7686–7708.

122. Tomalia, D. A., Baker, H., Dewald, J. R., Hall, M., Kallos, G., Martin, S., Roeck, J., Ryder, J., and Smith, P. (1985), A new class of polymers: Starburst-dendritic macromolecules, *Polym. J.*, 17, 117–132.

123. Tomalia, D. A., Baker, H., Dewald, J. R., Hall, M., Kallos, G., Martin, S., Roeck, J., Ryder, J., and Smith, P. (1986), Dendritic macromolecules: Synthesis of starburst dendrimers, *Macromolecules*, 19, 2466–2468.

124. Svenson, S., and Tomalia, D. A. (2005), Dendrimers in biomedical applications—Reflections on the field, *Adv. Drug Deliv. Rev.*, 57, 2106–2129.

125. Lacerda, L., Bianco, A., Prato, M., and Kostarelos, K. (2006), Carbon nanotubes as nanomedicines: From toxicology to pharmacology, *Adv. Drug Deliv. Rev.*, 58, 1460–1470.

126. Smart, S. K., Cassady, A. I., Lu, G. Q., and Martin, D. J. (2006), The biocompatibility of carbon nanotubes, *Carbon*, 44, 1034–1047.

127. Keck, C. M., and Muller, R. H. (2006), Drug nanocrystals of poorly soluble drugs produced by high pressure homogenisation, *Eur. J. Pharm. Biopharm.*, 62, 3–16.

128. Guerin, C. B. E., and Szleifer, I. (1999), Self-assembly of model nonionic amphiphilic molecules, *Langmuir*, 15, 7901–7911.

129. Choi, S. K., Vu, T. K., Jung, J. M., Kim, S. J., Jung, H. R., Chang, T., and Kim, B. H. (2005), Nucleoside-based phospholipids and their liposomes formed in water, *ChemBioChem*, 6, 432–439.

130. Snip, E., Shinkai, S., and Reinhoudt, D. N. (2001), Organogels of a nucleobase-bearing gelator and the remarkable effects of nucleoside derivatives and a porphyrin derivative on the gel stability, *Tetrahedr. Lett.*, 42, 2153–2156.

131. Engberts, J. B. F. N., and Hoekstra, D. (1995), Vesicle-forming synthetic amphiphiles, *Biochim. Biophys. Acta*, 1241, 323–340.

132. Giulieri, F., and Krafft, M. P. (2003), Tubular microstructures made from nonchiral single-chain fluorinated amphiphiles: Impact of the structure of the hydrophobic chain on the rolling-up of bilayer membrane, *J. Coll. Interf. Sci.*, 258, 335–344.

133. Giorgi, T., Lena, S., Mariani, P., Cremonini, M. A., Masiero, S., Pieraccini, S., Rabe, J. P., Samori, P., Spada, G. P., and Gottarelli, G. (2003), Supramolecular helices via self-assembly of 8-oxoguanosines, *J. Am. Chem. Soc.*, 125, 14741–14749.

134. Yanagawa, H., Ogawa, Y., Furuta, H., and Tsuno, K. (1989), Spontaneous formation of superhelical strands, *J. Am. Chem. Soc.*, 111, 4567–4570.

135. Zemel, A., Fattal, D. R., and Ben-Shaul, A. (2003), Energetics and self-assembly of amphipathic peptide pores in lipid membranes, *Biophys. J.*, 84, 2242–2255.

136. Chabaud, P., Camplo, M., Payet, D., Serin, G., Moreau, L., Barthelemy, P., and Grinstaff, M. W. (2006), Cationic nucleoside lipids for gene delivery, *Bioconjugate Chem.*, 17, 466–472.

137. Vaizoglu, M. O., and Speiser, P. P. (1986), Pharmacosomes—A novel drug delivery system, *Acta Pharm. Suec.*, 23, 163–172.

138. Schreier, S., Malheiros S. V. P., and de Paula, E. (2000), Surface active drugs: Self-association and interaction with membranes and surfactants. Physicochemical and biological aspects, *Biochim. Biophys. Acta*, 1508, 210–234.

139. Israelachvili, J. N. (1992), *Intermolecular and Surface Forces with Applications to Colloidal and Biological Systems*, Academic, London.

140. Jin, Y., Li, M., Tong, L., Wang, L., and Peng, T. (2003), Lipid derivatives of nucleoside analogues and their salts, China Patent CN1259331.

141. Jin, Y., and Ai, P. (2004), Cholesteryl derivatives of nucleoside analogues, China Patent CN1566130.

142. Jin, Y., Du, L., Xing, L., and Xin, R. (2006), Cholesteryl phosphoryl derivatives of nucleoside analogues, China Patent Application 2006101122957.

143. Jin, Y., Qiao, Y., and Hou, X. (2006), The effects of chain number and state of lipid derivatives of nucleosides on hydrogen bonding and self-assembly through the investigation of Langmuir-Blodgett films, *Appl. Surf. Sci.*, 252, 7926–7929.

144. Jin, Y. (2007), Effect of temperature on the state of the self-assembled nanoparticles prepared from an amphiphilic lipid derivative of acyclovir, *Coll. Surf. B Biointerf.*, 54, 124–125.

145. Aquaro, S., Calio, R., Balzarini, J., Bellocchi, M. C., Garaci, E., and Perno, C. F. (2002), Macrophages and HIV infection: Therapeutical approaches toward this strategic virus reservoir, *Antivir. Res*., 55, 209–225.
146. European Science Foundation (2005), Nanomedicine, An ESF-European Medical Research Council (EMRC) forward look report, European Science Foundation, Strasbourg.
147. Wagner, V., Dullaart, A., Bock, A.-K., and Zweck, A. (2006), The emerging nanomedicine landscape, *Nat. Biotechnol*., 24, 1211–1217.

# PHARMACEUTICAL NANOSYSTEMS: MANUFACTURE, CHARACTERIZATION, AND SAFETY

D. F. Chowdhury
*University of Oxford, Oxford, United Kingdom*

**Contents**

1 Definition
  1.1 Top-Down and Bottom-Up Approaches to Nanotechnology
2 Taxonomy of Nanomedicine Technologies
3 Nano–Pharmaceutical Systems
4 Description of Nanosystems
  4.1 Polymeric Systems
  4.2 Quantum Dots and Quantum Confinement
  4.3 Metal Nanoparticles and Surface Plasmon Resonance
  4.4 Self-Assembled Systems
  4.5 Nanostructures Based on Carbon
5 Manufacturing Technologies
  5.1 Nanoscale Assembly Methods
  5.2 Nano-structuring processes for polymeric materials
6 Characterization Techniques
  6.1 Nanoparticle Characterization Methods and Tools
  6.2 Scanning Probe Technologies
7 Toxicology Considerations
  7.1 Lung Toxicity
  7.2 Systemic Uptake
  7.3 Skin Permeation of Nanoparticles
References
Suggested Reading

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

## 1 DEFINITION

Nanotechnology is an enabling technology and one which is generally manifest at the primary level in the form of nanomaterials. The definition of nanotechnology therefore focuses on materials and how manipulation at the nanoscale leads to novel properties and therefore potentially new uses. The pharmaceutical industry has yet to adopt strict guidelines for what falls under the remit of nanotechnology, with numerous definitions in existence. For the purpose of this chapter, the current U.S. Food and Drug Administration (FDA) definition for nanotechnology as applied to pharmaceuticals is deemed most appropriate. The FDA describes nanotechnology as technology that includes the following [1]:

1. Research and technology development or products regulated by the FDA that are at the atomic, molecular or macromolecular levels and where at least one dimension that affects the functional behavior of the product is in the length scale range of approximately 1–100 nm
2. Creating and using structures, devices, and systems that have novel properties and functions because of their small and/or intermediate size
3. Ability to control or manipulate at the atomic scale

Nanotechnology is therefore essentially about understanding and manipulating materials at the atomic, molecular, and macromolecular level in a way that imparts properties to the material that would otherwise not exist either as individual atoms or as bulk processed macroscopic systems.

Properties that can be exploited to provide novel and unique properties to materials include surface and quantum effects, for example, van der Waals forces; electrostatic interaction; ionic, covalent and hydrogen bonding; and quantum confinement. Additionally nonconventional means of molecular assembly and atomic manipulation can lead to novel material properties. Control and exploitation of these effects can lead to new and useful changes to the thermal, magnetic, electrical, optical and mechanical, and biological and physicochemical properties of materials.

### 1.1 Top-Down and Bottom-Up Approaches to Nanotechnology

There are generally two approaches to nanotechnology, the top-down and bottom-up approaches. As the names suggest, the top-down approach utilizes ultraprecision machining and nanolithographic techniques among others to achieve very high definition structures with nanolevel accuracy, usually either by removing material from the surface of a larger structure until the desired structure with desired features is achieved or through deposition of material with almost atomic-scale precision and control. The bottom-up approach involves the assembly of atoms, molecules, or nanoscale components to assemble a larger structure within the nanoscale range. There are numerous methods by which this can be achieved, including conventional bulk chemical processing methods and exploitation of chemical and biological self-assembly techniques. The pharmaceutical industry is primarily involved in the application of nanomaterials rather than the discovery and development of new materials, though as this chapter will indicate, there are often areas of overlap between what is a new material and a construct of a novel material.

## 2 TAXONOMY OF NANOMEDICINE TECHNOLOGIES

A useful starting point would be to gauge the breadth of technologies falling under the classification of nanomedicine. Table 1 provides a means of classifying materials and processes derived from nanotechnology as relating to pharmaceuticals and medicine in general.

## 3 NANO–PHARMACEUTICAL SYSTEMS

Having gauged the huge scope of nanotechnology in medicine, the scope of this chapter is limited to pharmaceuticals. The term *pharmaceutical* is intended in the context of systems pertaining to drugs or dosage formulations. Nano–pharmaceutical systems generally imply products that may in their own right or in combination with another moiety bring about therapeutic benefit. They may also include engineered nanostructured systems that may act as a carrier for drugs or a delivery vehicle or a delivery system for drugs and therapeutic agents. The definition is extended to include imaging systems which may be used alone or in conjunction with therapeutic agents given the numerous nanosystems that have found application in diagnostics and imaging.

The various nanoscale architectures that can be achieved using nanotechnology include spheres (solid or hollow), tubes, porous particles, solid particles, and branched structures, and with the rapid evolution of lithographic techniques, three-dimensional objects of almost any desired shape can be achieved from both metals and polymers. Given the vast spectrum of materials of construction, size, shape, and form covered by nanosystems, a simple means of classification is needed for effective differentiation between systems. Nanosystems can be classified in a number of ways, for example, according to their elemental composition, according to size, structure, and function, or perhaps according to a structure–function relationship. It can be said that nanosystems fall into the broader category of nanostructures, which can be generalized into the following categories:

Particulate nanostructures
Capsular nanostructures
Crystalline nanostructures
Polymeric nanostructures

It can be readily appreciated that there will be an element of overlap between the broad categories above and in particular with evolving complex hybrid systems. The classification is intended as a point of reference and for ease of understanding the vast possibilities that exist with nanosystems without the need for constant reclassification as far as possible. These structures may be further differentiated according to their primary composition:

Organic
Inorganic
Organic/inorganic hybrid
Carbon based

**TABLE 1 Partial Nanomedicine Technologies Taxonomy**

- Raw nanomaterials
  - Nanoparticle coatings
  - Nanocrystalline materials
- Nanostructured materials
  - Cyclic peptides
  - Dendrimers
  - Detoxification agents
  - Fullerenes
  - Functional drug carriers
  - Magnetic resonance (MR) scanning (nanoparticles)
  - Nanobarcodes
  - Nanoemulsions
  - Nanofibers
  - Nanoparticles
  - Nanoshells
  - Carbon nanotubes
  - Quantum dots
- Artificial binding sites
  - Artificial antibodies
  - Artificial enzymes
  - Artificial receptors
  - Molecularly imprinted polymers
- Cell simulations and cell diagnostics
  - Cell chips
  - Cell stimulators
- DNA manipulation, sequencing, diagnostics
  - Genetic testing
  - Deoxyribonucleic acid (DNA) microarrays
  - Ultrafast DNA sequencing
  - DNA manipulation and control
- Tools and diagnostics
  - Bacterial detection systems
  - Biochips
  - Biomolecular imaging
  - Biosensors and biodetection
  - Diagnostic and defense applications
  - Endoscopic robots and microscopes
  - Fullerene-based sensors
  - Imaging (e.g., cellular)
  - Lab on chip
  - Monitoring
  - Nanosensors
  - Point-of-care diagnostics
  - Protein microarrays
  - Scanning probe microscopy
- Control of surfaces
  - Artificial surfaces, adhesive
  - Artificial surfaces, nonadhesive
  - Artificial surfaces, regulated
  - Biocompatible surfaces
  - Biofilm suppression
  - Engineered surfaces
  - Pattern surfaces (contact guidance)
  - Thin-film coatings
- Nanopores
  - Immunoisolation
  - Molecular sieves and channels
  - Nanofiltration membranes
  - Nanopores
  - Separations
- Biological research
  - Nanobiology
  - Nanoscience in life sciences
- Drug delivery
  - Drug discovery
  - Biopharmaceuticals
  - Drug delivery
  - Drug encapsulation
  - Smart drugs
- Synthetic biology and early nanodevices
  - Dynamic nanoplatform "nanosome"
  - Tecto-dendrimers
  - Artificial cells and liposomes
  - Polymeric micelles and polymersomes
- Nanorobotics
  - DNA-based devices and nanorobots
  - Diamond-based nanorobots
  - Cell repair devices
- Intracellular devices
  - Intracellular assay
  - Intracellular biocomputers
  - Intracellular sensors/reporters
  - Implant inside cells
- BioMEMS
  - Implantable materials and devices
  - Implanted bio—microelectromechanical systems (MEMSs), chips, and electrodes
  - MEMS/nanomaterial-based prosthetics
  - Sensory aids (e.g.,)
  - Microarrays
  - Microcantilever-based sensors
  - Microfluidics
  - Microneedles
  - Medical MEMS
  - MEMS surgical devices

**TABLE 1** ***Continued***

| | |
|---|---|
| Molecular medicine | Biotechnology and biorobotics |
| Genetic therapy | Biological viral therapy |
| Pharmacogenomics | Virus-based hybrids |
| Artificial enzymes and enzyme control | Stem cells and cloning |
| Enzyme manipulation and control | Tissue engineering |
| Nanotherapeutics | Artificial organs |
| Antibacterial and antiviral nanoparticles | Nanobiotechnology |
| Fullerene-based pharmaceuticals | Biorobotics and biobots |
| Photodynamic therapy | |
| Radiopharmaceuticals | |

*Source*: From ref. 2.

## 4 DESCRIPTION OF NANOSYSTEMS

There is clear evidence from Table 2 that some of the nanosystems indicated are based on conventional colloidal chemistry, and their characteristics are well established and understood. The descriptions below deal mainly with systems deemed nonconventional and cover some of the key novel properties derived from or utilized as part of their construction.

### 4.1 Polymeric Systems

Polymer-based systems offer numerous advantages, such as biocompatibility, biodegradability, and ability to incorporate functional groups for attachment of drugs. Drugs can be incorporated into the polymer matrix or in the cavity created by the polymeric architecture, from which the drug molecule can be released with an element of temporal control, and controlled pharmacokinetic profile with almost zero-order release achievable.

Dendrimers are large complex globular polymeric molecules [46] with well-defined chemical structure, size, and shape [47]. They consist of characteristic three-dimensional branched structures. The key components of dendrimers are their core, branches, and end groups, and precise control over these features is possible during the bottom-up synthesis process, thus allowing control over size composition, and final chemical reactivity. A more advanced form of dendrimer is the hyperbranched dendrimer, where precision control over the architectural construct is lost during the synthesis process [48]. Dendrimers are produced from monomers through an iterative sequence of reaction steps [49] using either convergent [50, 51] or divergent [52, 53] step growth polymerization [54–56] and have potential applications in gene and cancer therapy and drug delivery through complexation or encapsulation [57–59].

### 4.2 Quantum Dots and Quantum Confinement

Quantum dots are inorganic semiconductor nanocrystals that possess physical dimensions smaller than the exciton Bohr radius, giving rise to the unique phenom-

**TABLE 2 Classification of Nanostructures According to Composition and Perceived Applications**

| Composition | Type of Nanostructure | Applications |
|---|---|---|
| Organic | Polymer micelles [3, 4] | Drug delivery |
| | Polymeric spheres [5] | Drug delivery |
| | Polymer nanoparticles [6, 7] | Drug delivery |
| | Polymer vesicles/containers [8–10] | Drug delivery |
| | Lipid nanovesicles [11] | Drug delivery |
| | Lipid emulsions [12] | Drug delivery |
| | Ring peptides [13] | Drug delivery |
| | Lipid nanospheres [14, 15] | Drug delivery |
| | Lipid nanoparticles [16, 17] | Drug delivery |
| | Lipid nanotubes [18] | Drug delivery |
| | Peptide nanoparticles [19, 20] | Drug delivery |
| | Nanobodies [21] | Therapeutic, diagnostic |
| | Dendrimers [22–24] | Drug delivery, gene delivery |
| Inorganic | Palladium/platinum nanoparticles [25] | Drug delivery |
| | Silicon nanoneedles [26] | Drug delivery |
| | Porous silicon [27, 28] | Drug delivery |
| | Gold nanoparticles [29] | Drug delivery |
| | Iron oxide nanoparticles [30] | Imaging |
| | Gold nanoshells [31] | Imaging agent, thermal ablation |
| | Quantum dots [32] | Imaging, cell targeting |
| | Metallic nanoshells [33] | Imaging, thermal ablation |
| | Nanocrystals [34, 35] | Drug delivery |
| Organic/ inorganic hybrid | Nanocomposites [36] | Drug delivery |
| | Nanosphere–metallic particle composite [37] | Drug delivery, imaging |
| | Carbon nanotube clusters [38] | Drug delivery, imaging, thermal ablation |
| | Core–shell structures [39, 40] | Imaging, thermal ablation |
| Carbon based | Fullerenes [41–43] | Drug delivery, prodrug |
| | Carbon nanotubes [44, 45] | Drug delivery, imaging, thermal ablation |

enon known as quantum confinement. Quantum confinement is the spatial confinement of charge carriers (i.e., electrons and holes) within materials. It leads to unique optical and electrical properties that are not common for bulk solids.

Quantum dots have novel and unique optical, magnetic [60], and electronic properties, exceptional imaging properties due to high color intensity, with up to 20 fluorophores, high resistance to photobleaching, and narrow spectral line widths. Their size and composition allow for tunable emission that can be excited using a single wavelength [61–64]. These properties have led to uses as fluorescent imaging probes, detection of cell signaling pathways, and cell targeting. Low depth of light penetration and relatively high background fluorescence are the key limitations of quantum dots in in vivo clinical applications. Quantum dots are generally of the size rage 2–8 nm in diameter [65] and have large molar extinction coefficients [66], thus making them very bright in vivo probes.

Applications of quantum dots include optical detection of genes and proteins in animal models and cell assays and tumor and lymph node visualization through imaging [67, 68].

### 4.3 Metal Nanoparticles and Surface Plasmon Resonance

Surface plasmons, also known as surface plasmon polaritons or packets of electrons, are surface electromagnetic waves that propagate parallel along a metal–dielectric interface [69, 70]. Surface plasmons exist where the complex dielectric constants of the two media are of opposite sign. The excitation of surface plasmons by light of a wavelength matching the resonant frequency of the electrons is termed *surface plasmon resonance* (SPR) for planar surfaces and *localized surface plasmon resonance* where nanometer-sized metallic structures are concerned [71, 72].

Surface plasmon effects result in useful photothermal effects [73] and have been used to enhance the surface sensitivity of various spectroscopic measurements [74], including fluorescence, Raman scattering, and second-harmonic generation.

Metallic nanostructures exhibiting SPR are composed of a dielectric core and metallic shell, for example, gold sulfide dielectric core and gold shell. By varying the core–shell thickness ratio, the surface plasmon resonance is shifted from the visible to the infrared range [75], spanning a range that is mostly transparent to human tissue, that is, has a high physiological transmissivity. Additionally, control over the particle diameter allows control over light scattering and light absorption at particle diameters below approximately 75 nm. Potential applications of the photothermal effects of engineered nanoparticles include the following:

Controlled drug delivery [76]
Analysis of controlled drug release from a matrix [77]
DNA sensor [78, 79]
Deep tissue tumor cell thermal ablation [80]
Real-time assessment of drug action [81, 82]
Immunosensor applications [83, 84]

### 4.4 Self-Assembled Systems

Molecular self-assembly is a synthetic technique that has been widely used to produce nano- and microstructures in a quick and efficient manner. It has become all the more crucial to the formation of nanostructures due to the control attainable over the end product and the relative ease with which nanostructures of defined structure and function can be produced using bulk manufacturing methods.

The basic principle of self-assembly is based on the simultaneous coexistence of two parallel forces [85, 86], long-range repulsive forces and short-range attractive interactions.

The types of structures attainable using molecular self-assembly are referred to as micellar structures [87] and can take on various sizes and shapes:

Direct spherical micelles [88]
Inverse spherical micelles [89, 90]

Lamellar sheets [91]
Vesicles (hollow or concentric) [92]
Body-centerd-cubic [92]
Hexagonally packed cylinders/tubes [92]
Gyroids [93]
Hollow spheres [94]

These systems have found widespread use as drug delivery vehicles, and the more advanced nanosystems are termed *smart nano-objects* due to their ability to sense local variations in physiological conditions, such as pH and temperature, and respond to the stimulus accordingly.

### 4.5 Nanostructures Based on Carbon

***Nanotubes*** The structure of carbon nanotubes as observed by scanning tunneling microscopy is that of rolled grapheme sheets where endpoints of a translation vector are folded one onto another [95]. Single-walled carbon nanotubes (SWCNTs) were first reported by Iijima and Ichihashi [96] in 1993. Enormous interest in CNTs has centered around their unique properties, including high electrical conductivity, thermal conductivity, high strength and aspect ratio, ultralight weight, and excellent chemical and thermal stability.

The most common method for the production of carbon nanotubes is hydrocarbon-based chemical vapor deposition (CVD) [97] and adaptations of the CVD process [98, 99], where the nanotubes are formed by the dissolution of elemental carbon into metal nanoclusters followed by precipitation into nanotubes [100]. The CVD method is used to produce multiwalled carbon nanotubes (MWCNTs) [101] and double-walled carbon nanotubes (DWCNTs) [102] as well as SWCNTs [103]. The biomedical applications of CNTs have been made possible through surface functionalization of CNTs, which has led to drug and vaccine delivery applications [104, 105].

***Fullerenes*** Fullerenes were first discovered in 1985 [106] and are large molecules composed exclusively of carbon atoms and manifest physically in the form of hollow spherical cagelike structures. The cages are in the region of 7–15 Å in diameter with the most common form being $C_{60}$, though other forms exist too, such as $C_{70}$, $C_{76}$, and $C_{84}$, depending on the number of carbon atoms making up the cage. Fullerenes can be produced using combustion [107] and arc discharge methods [108].

Fullerenes offer numerous points of attachment and allow precise bonding of active chemical groups in three-dimensional (3D) conformations and positional control with respect to matching conjugated fullerene compounds with a given target. Water-soluble fullerenes have shown low biological toxicity both in vitro [109] and in vivo [110]. Some of the potential applications of fullerenes in pharmaceuticals include their use in neurodegenerative and other disease conditions where oxidative stress is part of the pathogenesis due to their powerful antioxidant properties [111] and in nuclear medicine for binding of toxic metals ions, increasing therapeutic potency of radiation therapy and reducing adverse events as fullerenes do not undergo biochemical degradation within the body. Fullerene applications in photodynamic tumor therapy have also been shown [112].

## 5 MANUFACTURING TECHNOLOGIES

There are a diverse range of technologies being applied to the manufacture of nanosystems for pharmaceutical applications. Some of these are derived from conventional pharmaceutical technologies, such as colloidal processing, and many have been adopted from the semiconductor industry, whereby precision spatial control is achieved over the production of nanosystems and particles using fabrication techniques. To add to this, new technologies are constantly evolving through the adaptation and amalgamation of existing technologies in different fields or through pure innovation leading to completely new processes. It is outside the remit of this chapter to cover in any depth all those manufacturing technologies that may be applied to pharmaceutical manufacturing. The summary in Table 3 provides a detailed synopsis of the different types of manufacturing processes and types of technologies for each process. This is followed by a brief introduction to some of the technologies, with the omission of silicon- and carbon-based fabrication processes, which are beyond the scope of this chapter, to provide the reader with a starting point for further detailed study and investigation into those processes and technologies that may be most suited to their particular product or concept.

**TABLE 3 Summary of Manufacturing Processes and Technologies for Producing Nanosystems**

| Manufacturing Process | Technology |
|---|---|
| Nanoscale assembly | Self-assembling micellar structures [113, 114] |
| | Bio-self-assembly and aggregation [115, 116] |
| | Nanomanipulation [117, 118] |
| | Soft lithography [119, 120] |
| | Molecular imprinting [121, 122] |
| | Layer-by-layer electrostatic deposition [123, 124] |
| | Chemical vapor deposition [125] |
| Nanostructuring processes for polymeric materials | Mold replication [126, 127] |
| | Colloidal lithography [128, 129] |
| | X-ray lithography [130] |
| | Interfacial polymerization [131, 132] |
| | Nanoprecipitation [133, 134] |
| | Emulsion solvent evaporation [135] |
| | Nanoimprinting [136] |
| | Electrospinning [137, 138] |
| Nanostructuring processes for silicon | Photolithographic fabrication |
| | X-ray lithography [130] |
| | Electron beam lithography [139] |
| | Chemical etching [140] |
| | Physical and chemical vapor deposition [141] |
| Nanostructuring processes for carbon | Electric arc discharge [142, 143] |
| | Laser ablation [144, 145] |
| | Chemical vapor deposition [146] |
| | Combustion [147] |

## 5.1 Nanoscale Assembly Methods[1]

***Self-Assembly through Micelle Formation*** Self-assembly at the nanoscale is deemed important to be able to produce commercially viable products and processes, since it offers a mode of bulk production with control over features such as size, shape, and morphology at the nanoscale. The basic principle of self-assembly is based on the simultaneous coexistence of two parallel forces:

Long-range repulsive interactions between incompatible domains
Short-range attractive interactions

If we take the example of an amphiphilic diblock copolymer, the polymer is composed of two blocks, a hydrophobic block and a hydrophilic block. When introduced to a solvent beyond a minimum concentration, the critical micelle concentration (CMC), the monomers begin to orientate such that the block that is soluble in the solvent orients itself toward the periphery, in contact with the continuous media, and the insoluble portion turns toward the core in an attempt to minimize contact with the continuous phase, thus leading to the formation of a micelle. The long-range repulsive forces arise from the relative solubilities of the blocks in the solvent, and the short-range attractive forces arise from the covalent link between the two blocks. The basic theory of micelle formation using block copolymers is outlined below since nanosystem and nano-object self-assembly is likely to be facilitated by such polymer systems, and similar principles will apply or aid toward developing self-assembling systems.

Key factors that affect micelle formation are as follows:

Equilibrium constant
Solvent type
Solvent quality
Critical micelle temperature (CMT)
Critical micelle concentration
Overall molar mass of the micelle, $M_W$
Micelle aggregation number, $Z$
Copolymer architecture
Relative block lengths
Relative geometries of copolymer blocks
Polymer composition
Core–corona interfacial tension

These key factors will influence the following micelle characteristics:

Hydrodynamic radius of micelles formed, $R_H$
Radius of gyration, $R_G$

[1]The description in this section has been summarized and adapted from J. Rodriguez-Hernandez et al., Toward "smart" nano-objects by self-assembly of block copolymers in solution, *Progress in Polymer Science*, 30 (2005), 691–724 [148].

Ratio of hydrodynamic radius, $R_H$, to radius of gyration, $R_H$
Micelle core radius, $R_C$
Micelle corona thickness, $C$

Micellar structures can be produced either by addition of the polymer solution or addition of the powdered material to the desired solvent and stirring at the optimum temperature and monomer concentration. The CMC can be determined by ultraviolet (UV) absorption or light scattering techniques such as static light scattering (SLS), dynamic light scattering (DLS), or small-angle X-ray scattering (SAXS). At the concentration at which monomers form micelles, there will be a radical drop in monomer concentration in the bulk.

The stability of micellar systems depends upon the ability to ensure the aggregated monomers do not deaggregate and that individual micelles do not coalesce to form larger aggregates. This is inevitable over a period of time, but steps can be taken to prolong the stability of the systems through various techniques, and those listed below are some of the methods used for spherical micellar systems:

Steric stabilization using emulsifiers and surfactants
Shell or core cross-linking
Viscosity-enhanced stabilization
Amine cross-linked stabilization
Thermodynamic stabilization

The types of systems that can be produced using micelle formation include spheres, shells, capsules, vesicles, clusters, and particles of various shapes and sizes, such as spheres, rods, planar structures, and layered structures. Further processing can be undertaken to add rate-controlling polymer membranes to the outer shell and to incorporate different molecules to the surface (e.g., for receptor recognition).

***Biological Self-Assembly Using DNA as Construction Tool*** This is a technique that has been adopted to produce 2D or 3D nanosystems by utilizing the base-pairing affinity of DNA [149, 150].

Biological self-assembly using DNA can be described as a process that allows the systematic assembly of molecules with high levels of precision and accuracy without external constraints or influences. This allows the construction of nanoscale objects to the desired structure, conformation, and composition very rapidly and without the need for complex processing techniques and conditions.

DNA is a copolymer composed of a phosphate and sugar backbone and four types of bases that branch off from the backbone, A (adenine), G (guanine), C (cytosine), and T (thymidine). During DNA replication two strands of DNA come together to form a helical structure through complementary base pairing which is highly specific, whereby thymidine pairs with adenine and guanine with cytosine. When strands of DNA come together where the ends of the strands are noncomplementary, a portion of the strand extends beyond the complementary base-paring region leading to an overhang, otherwise known as a "sticky end".

The natural mechanism of DNA base pairing can be used to assemble synthetic sequences of DNA molecules by synthesizing DNA molecules such that they form

stable branches [150, 151] with arms that form sticky ends that in turn can assemble to form supramolecular structures [152, 152a]. This approach may be used to produce complex 3D assemblies through sequential or layer-by-layer self-assembly and may incorporate other materials, such as particles and proteins [153–156]. The advantages of this method are as follows:

Specificity and geometry of intermolecular interactions that can be predicted
Precision control over the final structure at the nanoscale
Simple manufacturing process without external restrictions
Complex structures that can be built with defined topologies
Potential for creation of nanodevices

The types of structures that may be constructed are Branched planar/2D quadrilateral structures [157], cubes [158], octahedrons [159], and complex 2D and 3D periodic structures [160].

*DNA Synthesis for Nanoconstruction* Single strands of DNA, otherwise known as oligomers, are most commonly produced using a solid-support synthesis process [161, 162]. This is a cyclic process where each nucleotide is sequentially coupled to form a nucleotide chain (working from the 3′ end to the 5′ end). The 3′ end is initially covalently linked to a solid support and the nucleotide monomers are added sequentially. This is a well-established process and its key parameters and critical process steps are well documented in the literature [163, 164]. The DNA strands can be tailored according to the desired nanoconstruction scheme and target structure [165].

***Nano Manipulation*** As the name suggests, this is quite literally a technique for physically manipulating matter at the nanoscale. Scanning probe microscopy (SPM) techniques have been most widely used to achieve this using the scanning probe tip as an implement for assembling atoms, molecules, or nanoparticles according to the desired spatial conformation [166–169].

***Soft Lithography*** Lithography is essentially a process for printing features on a planar surface. Nanolithography tools, commonly referred to as soft lithography, allow precisely defined nanoscale features to be produced on a substrate, which can be removed from the substrate as free-standing 3D nano-objects. A number of techniques fall within the field of soft lithography, primarily for construction of micrometer-sized objects:

Replica molding
Micromolding in capillaries (MIMIC)
Microtransfer molding
Solvent-assisted microcontact molding (SAMIM)
Microcontact printing

Near-field phase shift lithography is a soft lithographic technique used to produce geometric shapes with size features at the nanoscale (approximately 40–80 nm). This involves the production of a polymer mask containing the desired pattern to be

replicated on the substrate, with nanoscale features usually patterned by X-ray or electron beam exposure. The mask is then placed on the surface of the substrate and exposed to near-field light, the intensity of which leads to replication of the pattern on the mask on to the substrate. Complex geometries, shapes, and features can be produced on the substrate which can subsequently be removed to give free-standing particles or objects [170, 171].

***Molecular Imprinting*** Molecular Imprinting is a process used to imprint or copy recognition sites from desired molecules on to polymer structures [172, 173]. The recognition sites can be produced on organic or inorganic polymers and inorganic materials such as silica and biomaterials such as proteins. A template molecule is dissolved in solvent with polymerizable monomers which undergo bond formation with the template molecule forming either noncovalent bonds through electrostatic interactions, hydrogen bonds or hydrophobic interactions, or reversible covalent bonds. The monomers are then polymerized to form a cast or semirigid polymeric structure which maintains the steric conformation of the molecule template and its recognition site upon removal of the template molecule. As a result, the molecular template affinity for molecules and analyte is mimicked by the "imprinted" polymer [174]. This has applications in chromatography and drug discovery and potential applications in targeted drug delivery.

***Layer-by-Layer Electrostatic Deposition*** Electrostatic deposition utilizes the electrostatic bonding affinities of materials imparted by their surface charge to build highly ordered multilayered films or structures on a substrate. The process involves the successive deposition of oppositely charged polyions, exploiting the Coulombic long-range electrostatic interactions between the oppositely charged molecules, allowing formation of multilayers over a large distance. This technique can be used to build multilayer composite films on particles incorporating molecular fragments such as polymer–polymer, polymer–organic, polymer bimolecular, and polymer–mineral composition [175–177].

***Chemical Vapor Deposition*** CVD is a crystal growth process whereby a solid material is deposited from the gas phase onto a controlled substrate using a suitable mixture of volatile precursor materials which react to produce the desired deposit on the substrate surface (Table 4). Types of films and structures that can be produced include the following:

Polycrystalline
Amorphous
Epitaxial silicon
Carbon fiber
Filaments
Carbon nanotubes
Silicon dioxide
Tungsten
Silicon nitride
Titanium nitride

**TABLE 4 Chemical Vapour Deposition Methods and Their Key Features**

| CVD Method | Key Features |
|---|---|
| Atmospheric pressure CVD [178, 179] | Operates at atmospheric pressure |
| Atomic layer CVD (atomic layer epitaxy) [180, 181] | High-precision film thickness and uniformity requirements |
| Aerosol-assisted CVD [182, 183] | For use with involatile precursors |
| Direct liquid injection CVD [184, 185] | High film growth rates possible |
| Hot-wire CVD [186] | High growth rate, low temperature, and use of inexpensive materials such as plastics as substrate |
| Low-pressure CVD [187] | Improved film uniformity |
| Metal organic CVD [188] | Uniform and conformal deposition |
| Microwave plasma-assisted CVD [189] | No external heating required |
| Plasma-enhanced CVD [190] | Reduced substrate temperatures can be used |
| Rapid thermal CVD [191] | Conformal coverage over high-aspect-ratio features is possible, i.e., improved control of interfacial properties |
| Remote plasma-enhanced CVD [192] | Excellent conformal coverage of complex structures<br>Can produce multilayer and graded layers with tailored functional group attachment |
| Ultrahigh vacuum CVD [193] | Reduced surface contamination |

### 5.2 Nanostructuring Processes for Polymeric Materials

Numerous microfabrication techniques have been used to produce a wide range of implantable and oral drug delivery systems using materials ranging from silicon, glass, silicone elastomer, and plastics. Fabrication techniques have rapidly evolved to produce nanoscale objects and therapeutic systems using polymeric materials as the substrate due to their biodegradable nature. There are a number of different synthetic polymer systems that have been developed for this type of application, and the most common ones are listed below:

Poly(D-lactic acid) (PDLA)
Poly(ε-caprolactone) (PCL)
Poly(vinyl alcohol) (PVA)
Polyalkylcyanoacrylates (PACA)
Poly(L-lactide) (PLLA)
Poly(lactide-*co*-glycolide) (PLGA)
Polymethylcyanoacrylate (PMCA)

Techniques for the production of micrometer-sized features using polymers are well established and apply primarily to device construction. The techniques listed

below focus primarily on attaining submicrometer, nanoscale features, and geometries using polymers such as those listed above.

***Nanomold Replication*** A physical mold is produced that has nanoscales on the order of tens or a few hundred nanometers. To achieve such fine features with precision and repeatability, electrodeposition is used to produce the molds, otherwise referred to as a nanostamp [194]. The stamp is then use as a master stamp to duplicate the image or object by casting or embossing the polymeric material.

***Colloidal Lithography*** Colloidal lithography is a process whereby an electrostatically self-assembled array of monodispersed colloidal nanospheres is used as a mask to construct nanoscale objects and features through deposition or etching processes. The monodisperse colloidal spheres, for example, surface-charged latex, self-organize or assemble into periodic arrays on the substrate, glass, for example, and do not aggregate due to the surface charge repulsion. This method can be used to produce 2D [195, 196] and 3D [197] nanostructures, arrays of rings, dots, honeycomb structures, pillars, and chemical patterns [198] with a high level of control over structure and conformation.

***Interfacial Polymerization*** Interfacial polymerization is a process whereby very thin films or membranes, on the order of nanometer thickness, are produced by reacting two monomers at the interface between two immiscible solutions [199]. Nanoparticles [200] and aqueous core capsules with very thin membranes have been produced using this method for drug delivery applications.

***Nanoprecipitation*** Nanoprecipitation is a self-assembly directed nanoparticle formation method. There are three key steps involved in this process: rapid micromixing of the solutes, the creation of a high level of supersaturation to instigate rapid nucleation and growth of precipitate, and the kinetic control and termination of growth using copolymer stabilizers. One of the drawbacks using this method is the poor incorporation of water-soluble drugs [201]. However, the main advantage associated with the production of nanospheres for drug delivery using this technique is the high degree of control attainable over particle size [202].

***Emulsion Solvent Evaporation*** The basic concept of the emulsion solvent evaporation technique producing nanoparticles is very straightforward. The particles are formed as an emulsion of a polymer–surfactant mixture and dispersed in an organic solvent. The solvent is then evaporated to leave behind the individual emulsion droplets which form stable free nanoparticles [203]. This method is far easier and more preferable over methods such as spray drying and homogenization and operates under ambient conditions and mild emulsification conditions. The size and composition of the final particles are affected by variables such as phase ratio of the emulsion system, organic solvent composition, emulsion concentration, apparatus used, and properties of the polymer [204].

***Nanoimprinting*** This is a lithographic technique similar to soft contact lithography discussed earlier, with the main difference being that nanoimprinting uses a

hard mold to produce nanoscale features down to sub-10 nm resolution [205] by directly imprinting onto the polymer surface at high temperatures. More recently, molds produced from carbon nanotubes have been used to achieve molecular-scale resolution. Molds are generally made using electron beam lithography; however, high-definition molds are produced using molecular beam epitaxy. Some of the technical issues associated with this technique include sticking, adhesion, and material transport during imprinting [206].

***Electrospinning*** Electrospinning is a process that uses electrostatic force to produce nanofibers from a charged polymer. An electrode is placed into a spinning polymer solution/polymer melt and the other electrode is attached to a collector plate. A high-intensity electric field is created by applying a high voltage such that the polymer solution is discharged as a jet, and during travel of this charged polymer jet toward the grounded collector plate, solvent evaporation leaves a charged polymer fiber which deposits on the collector plate [207]. These fibers have high specific surface areas and are highly flexible, and applications include the preparation of controlled drug release membranes [208].

## 6 CHARACTERIZATION TECHNIQUES

### 6.1 Nanoparticle Characterization Methods and Tools

A summary of some key properties that may be assessed as part of a characterization schedule for nanoparticles and nanostructures and a comprehensive but not exhaustive list of tools and techniques that may be used are presented in Table 5 [209–223]. The degree of characterization and method used will be determined by the intended application of the nanomaterials.

Characterization of micellar and supramolecular structures and their counterparts often require different or additional tools and techniques [224–232] and a summary is provided in Table 6 of various characterization parameters for micellar and supramolecular structures and components and analytical tools that may be applied.

### 6.2 Scanning Probe Technologies

Scanning probe microscopy has almost become synonymous with nanomaterial characterization [233]. This is a family of techniques that have evolved from the use of a sharp proximal probe to scan a surface in order to ascertain its properties down to atomic-scale resolution based on tip–surface interaction. There are two main SPM techniques, scanning tunneling microscopy (STM) [234, 235] and AFM [236]. Near-field scanning optical microscopy (NSOM) [237–239] also falls within the SPM family of techniques; however, this uses a subwavelength near-field light source as the scanning probe, achieving resolutions down to 50 nm, and is not discussed further here.

A host of techniques have evolved from STM and AFM, primarily involving adaptations to instrumentation depending on the material and parameter under

**TABLE 5 Characterization Parameters and Tools for Nanoparticles and Nanostructures**

| Characterization Parameter | Analytical Tool |
|---|---|
| Composition | Liquid chromatography, e.g., high-performance liquid chromatography (HPLC), size exclusion chromatography (SLC)–HPLC |
| | Field flow fractionation (FFF) |
| | UV–visible spectrophotometry |
| | Refractive Index |
| | Inductively coupled plasma–optical emission spectrometry (ICP–OES) |
| | Fourier transform infrared spectrometry |
| | Mass spectrometry |
| | X-ray fluorescence |
| | Extended X-ray absorption fine structure (EXAFS) spectroscopy |
| | X-ray absorption near edge (XANES) spectroscopy |
| Particle diameter | Static and dynamic laser light scattering |
| | Scanning probe technologies |
| Size distribution | Static and dynamic laser light scattering |
| | Photon correlation spectroscopy |
| Surface area | BET method (Brunauer, Emmett, and Teller method) |
| Porosity (pore size, volume, and distribution) | Physical gas sorption |
| | Chemical gas sorption |
| | Helium picnometry |
| | Mercury intrusion porometry (MIP) |
| Core–shell thickness | Small-angle scattering of polarized neutrons (SANSPOL) |
| Surface structure and morphology | Small-angle neutron scattering (SANA) |
| | Proton nuclear magnetic resonance ($^{1}$H NMR) spectroscopy |
| | Scanning electron microscoscopy (SEM), atomic force microscopy (AFM), energy dispersive X-ray (EDXA), transmission electron microscopy (TEM), scanning probe microscopy (SPM), auger electron spectroscopy (AES), X-ray diffraction (XRD), X-ray photoelectron microscopy (APS), X-ray photoelectron spectroscopy (XPS), Vertical scanning phase shifting interferometry |
| Surface charge density | Zeta potential using Electrostatic light scattering (ELS) |
| | Zeta potential using multifrequencyelectro acoustics |
| | Zeta potential using phase analysis light scattering (PALS) |
| Shape | Electron microscopy |
| | Scanning probe technologies |
| Concentration distribution | Energy dispersive X-ray spectrometry (EDS) combined with SEM or scanning |
| Crystallinity, Bulk | X-ray diffraction |
| Crystallinity, Local | TEM/–selected area diffraction (SAD) |
| | Differential scanning calorimetry (DSC) |
| Magnetic properties | Scanning probe technologies |
| Electrical properties | Scanning probe technologies |
| Optical properties | UV–visible Spectroscopy |

**TABLE 6 Characterization Parameters and Analytical Tools for Micellar and Supramolecular Structures**

| Characterization Parameter | Analytical Tool |
|---|---|
| Critical micelle concentration | Flurimetric methods<br>Static light scattering<br>Dynamic Light Scattering |
| Aggregation number | Fluorescence correlation spectroscopy |
| Radius of gyration, $R_G$ | Small-angle X-ray scattering |
| Hydrodynamic radius, $R_H$ | Photon correlation spectroscopy |
| Core/corona size, micelle structure, overall micelle size | Small-angle X-ray scattering |
| Overall shape, cross section | Small-angle neutron scattering |
| Size, shape, and internal structure | Transmission electron microscopy<br>Scanning probe technologies |
| Average molecular weight | Membrane and vapor pressure osmometry |
| Monitor equilibrium state, stability monitoring | Light-scattering methods |
| Structure elucidation, polymer architecture, polymer interactions | Nuclear magnetic resonance |

investigation. It should be noted that technological advances continue unabated and new techniques are constantly being developed within the scanning probe family to cater for the characterization of new and novel materials and nanoscopic constructs. Table 7 gives a current synopsis of these techniques.

***Scanning Tunneling Microscopy*** The scanning tunneling microscope was first described by Nobel Prize winners Binnig and Rohrer in 1982 [249] and consists of an atomically sharpened tip usually composed of tungsten, gold, or platinum–irridium. The tip is scanned within atomic distance (about 6–10 Å) of the sample under study under very high vacuum, and a bias voltage is applied between the sample and the scanning probe tip, resulting in a quantum mechanical tunneling current across the gap. The magnitude of the tunnelling current is related exponentially to the distance of separation and the local density of states (i.e., electron density in a localized region of a material) [250, 251].

The relationship between tunneling current and separation is given as:

$$I = C\rho_t\rho_s e^{d0.5}$$

where $I$ = tunnelling current
$C$ = constant (linear function of voltage)
$\rho_t$ = tip electron density
$\rho_s$ = sample electron density
$e^{d0.5}$ = separation (governed by exponential term)

The tip is scanned across the sample surface using a piezoelectric transducer in one of two modes, topographic mode or current mode. In the topographic mode a con-

**TABLE 7 Summary of Scanning Probe Technologies**

| SPM Technique | Property Measured |
|---|---|
| **Atomic force microscopy** | Visualisation and measurement of surface features |
| Noncontact AFM [240] | Insulating substrates, atomic resolution |
| | Molecular systems, atomic resolution |
| | Biocluster and biomolecular imaging |
| | Imaging and spectroscopic data in liquid environments |
| | Nanoscale charge measurement |
| | Nanoscale magnetic properties |
| Contact AFM [241, 242] | Topographic imaging of solid substrates |
| | Mechanical properties |
| | Local adhesive properties |
| Piezoresponse | Characterization and domain engineering of ferroelectric materials |
| Lateral force | Fine structural detail |
| | Transitions between components on surface, e.g., polymer composites |
| Scanning thermal | Defects in sample based on thermal differences |
| Intermittent AFM (tapping mode) [243] | Biological systems: DNA/RNA analysis, protein–nucleic acid complexes, molecular crystals, biopolymers, ligand–receptor binding |
| Phase imaging AFM | Two phase polymer blends |
| | Surface contaminants |
| | Biological samples |
| Lift mode AFM techniques [244–246] | Topography |
| Magnetic Force | Magnetic properties/regions |
| Electrical Force | Electrical properties |
| Surface Potential | Surface potential |
| Scanning Capacitance | Material capacitance |
| Force modulation | Elasticity |
| **Scanning tunnelling microscopy [247]** | Surface imaging |
| | Three-dimensional profiling with vertical resolutions to 0.1 Å |
| | Measurement of electronic and magnetic properties |
| | Surface electronic state |
| Spin-polarized STM/STS [248] | Mapping surface magnetism at atomic scale |

stant distance is maintained between the tip and sample surface using a feedback loop operated with the scanner. In the current mode, variations in current with changes in surface topography are monitored by switching off the feedback loop, thus providing a 3D image of the surface under study. The key features and limitation of STM are as follows:

*Features of STM* [252, 253]

Can undertake topographical imaging of surfaces with atomic-scale lateral resolution, down to 1 Å

There-dimensional profiling possible with vertical resolutions down to 0.1 Å

Wide range of materials can be analyzed

Surface electronic properties may be measured
Large field of view, from 1 Å to 100 μm
Vibrational isolation allows highly sensitive measurements to be undertaken
Ultrahigh vacuum (in the range $10^{-11}$ torr) minimizes sample contamination and reduces oxide layer growth, thus allowing for high sensitivity measurements

*Limitations of STM*

Can be difficult to differentiate between a composite of materials on the surface
Tip-induced desorption of surface molecules may occur
Ultrahigh vacuum requirements
Vibrational isolation requirements lead to increased installation costs
Low scanning speed

***Atomic Force Microscopy*** Atomic force microscopy is a direct descendant of STM and was first described in 1986 [254]. The basic principle behind AFM is straightforward. An atomically sharp tip extending down from the end of a cantilever is scanned over the sample surface using a piezoelectric scanner. Built-in feedback mechanisms enable the tip to be maintained above the sample surface either at constant force (which allows height information to be obtained) or at constant height (to enable force information to be obtained). The detection system is usually optical whereby the upper surface of the cantilever is reflective, upon which a laser is focused which then reflects off into a dual-element photodiode, according to the motion of the cantilever as the tip is scanned across the sample surface. The tip is usually constructed from silicon or silicon nitride, and more recently carbon nanotubes have been used as very effective and highly sensitive tips.

In noncontact-mode AFM the cantilever is oscillated slightly above its resonant frequency and the tip does not make contact with the sample surface but instead oscillates just above the adsorbed fluid layer on the surface, maintaining a constant oscillation. The resonant frequency of the cantilever decreases due to van der Waals forces extending from the adsorbed fluid layer. This changes the amplitude of oscillation, the variations of which are detected using sensitive alternating current (AC) phase-sensitive devices, providing topographical information. In contact mode, AFM the tip remains in contact with the sample surface, and the feedback loop maps the vertical vibrational changes. In tapping mode, the cantilever is oscillated above the sample surface such that it intermittently contacts the sample surface. The key features and limitations of AFM are a follows:

*Features of AFM* [255–257]

High scan speeds
Atomic-scale resolution possible
Rough sample surfaces can be analyzed
High lateral resolutions possible
Soft samples (e.g., biological tissue) can be measured

*Limitations of AFM* [255–257]

Potential for image distortion due to lateral shear forces (in contact mode)
May be reduced spatial resolution due to sample scraping (in contact mode)
Tapping mode has lower scan speed compared to contact mode, though there is less susceptibility for sample damage and image distortion

## 7 TOXICOLOGY CONSIDERATIONS

Nanomaterials may in their own right possess novel and useful properties or as a composite of the same or different materials to form larger useful structures. Safety consideration is therefore of paramount importance since completely inert materials have the ability to exhibit toxic effects by virtue of a reduction in their size and associated increase in surface area–mass ratio, let alone materials manipulated specifically to impart novel properties.

Two obvious routes of human contact with nanoparticulates are the skin and via inhalation. Given the size of the particles, there may be a propensity for absorption into the systemic circulation. In some cases the nanosystems are engineered to achieve enhanced systemic absorption. The established methodology for toxicological assessment of new materials should be adhered to, and the discussion below is intended only to touch upon some of the immediate safety concerns that should be understood and addressed when dealing with nanomaterials.

### 7.1 Lung Toxicity

The safety of ultrafine particles remains to be clearly elucidated and requires the collaborative input of toxicologists (animal, cellular, molecular), epidemiologists, clinicians (pulmonary, cardiovascular, neurological), and atmospheric scientists. There are several published studies to indicate that ultrafine particles pose a higher toxicity risk compared with their larger counterparts [258–261]. Figure 1 outlines the potential effects of ultrafine particles on respiratory mucosa, the cardiovascular system, and central and peripheral nervous systems, upon inhalation.

Hohfeld et al. [262] hypothesized the toxic effects of ultrafine particles to be attributable to the following:

- High efficiency of deposition in the alveolar region due to particle size
- Large surface area
- Decreased phagocytosis leading to interaction of the particles with the epithelium, resulting in the development of conditions such as chronic diffuse interstitial fibronodular lung disease
- Dislocation from the alveolar space, leading to potential systemic effects

The hypothesis affirms the need to characterize the material's physical and chemical properties, including morphological analysis. The latter has significant ramifications on the aerodynamics of the particulate matter and hence its ultimate disposition.

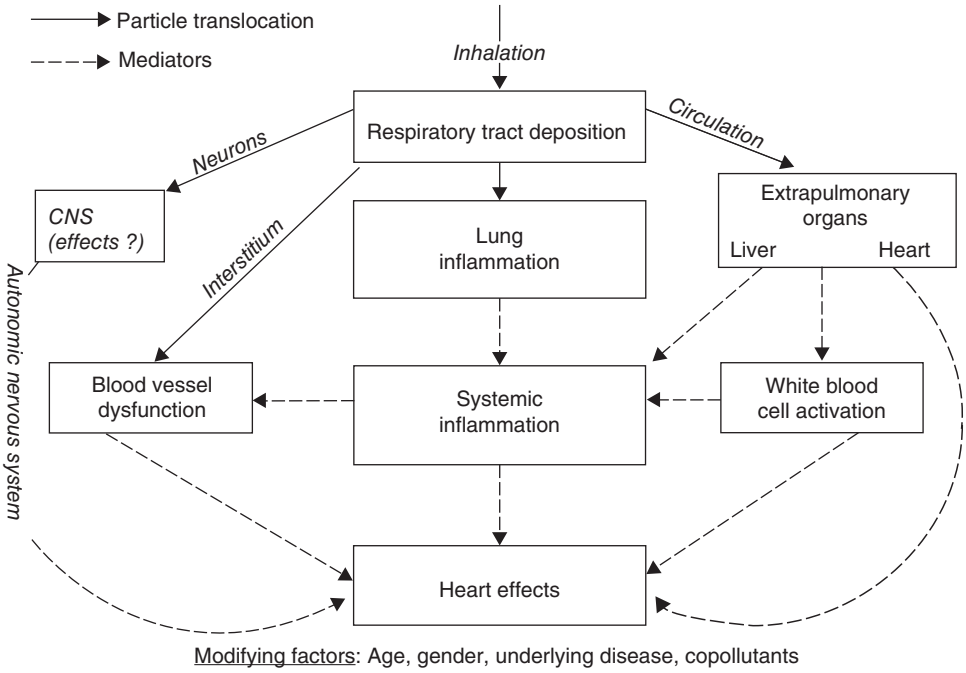


**FIGURE 1** Potential mechanisms of effects of inhaled ultrafine particles. (Reproduced with permission from G. Oberdörster, Inhaled nano-sized particles: Potential effects and mechanisms, paper presented at the Symposium, Health Implications of Nanomaterials, October 2004.)

Carbon nanotubes are a class of materials finding increasingly widespread applications in drug discovery and development and may be classed as a form of ultrafine material. It has been shown that single-wall carbon nanotubes do not produce any significant respirable aerosol levels due to agglomeration resulting from the very high surface area–volume ratio and associated electrostatic interaction between the nanotubes [263]. A good deal of research has focused on developing methods for the dispersion of nanotubes for further downstream processing for conversion to useful applications. The liquid dispersions do not pose the same level of hazards posed by the dry powder material, and most of the work in the pharmaceutical industry with carbon nanotubes is focused on liquid dispersions whereby nanotubes are being functionalized and conjugated with drugs and possibly other carriers for therapeutic intent [264].

However, it has at the same time been shown that dry powder carbon nanotubes can persist in the lungs and have the potential to induce inflammatory and fibrotic reactions, evident in the form of collagen-rich granulomas in the bronchi and interstitium [265]. This emphasizes the need for caution and further work to establish the exact cause of these effects given the propensity of nanotubes to agglomerate.

### 7.2 Systemic Uptake

Nanoparticles, by virtue of size, have a tendency to evade phagocytosis. Uptake into the systemic circulation is thus thought to be through diffusion and via the endo-

thelial cells, the epithelium, interstitium, and blood vessels. Translocation into the blood is thought to be through enhanced epithelial or endothelial permeability imparted by inflammatory mediators. Systemic hypercoagulation may be triggered by the inflammatory mediators in response to the diffusion of the nanoparticles through endothelium and vasculature [260].

### 7.3 Skin Permeation of Nanoparticles

The stratum corneum provides a formidable barrier to the entry of chemical and particulate matter into human tissue and systemic circulation. It provides a first-line defence to the ingress of foreign agents. However, there are indications that particles up to 1 μm are able to penetrate the skin's barrier and deposit in the epidermis where the antigen-presenting cells reside [266]. It follows therefore that submicrometer particles in the nanometer range have the potential to cross the stratum corneum and illicit an inflammatory response. Once again, however, the tendency of fine particles to agglomerate will to some extent inhibit penetration into the skin, in particular where the agglomerates are macroscopic. A correlation must however be drawn to establish any potential link between nanoparticle affinity for skin penetration and particle physical and morphological characteristics or indeed whether the novel and unique properties of the engineered particle in any way impart enhanced skin permeation properties and, if so, their nature and mechanisms.

## REFERENCES


1. FDA and Nanotechnology Products. U.S. Food and Drug Administration (FDA), available: http://www.fda.gov/nanotechnology/faqs.html, accessed June 2006.
2. Freitas, R. A., Jr. (2005), What is nanomedicine? *Nanomed. Nanotechnol. Biol. Med.*, 1, 2–9.
3. Yang, L., et al. (2000), Physicochemical aspects of drug delivery and release from polymer-based colloids, *Curr. Opin. Coll. Interf. Sci.*, 5(1–2), 132–143.
4. Bontha, S., et al. (2006), Polymer micelles with cross-linked ionic cores for delivery of anticancer drugs, *J. Controlled Release*, 114(2), 163–174.
5. Sphurti, V., et al. (2006), Nanofibers and spheres by polymerization of cyanoacrylate monomer, *Polymer*, 47(12), 4328–4332.
6. Qian, F., et al. (2006), Preparation, characterization and enzyme inhibition of methylmethacrylate copolymer nanoparticles with different hydrophilic polymeric chains, *Eur. Polym. J.*, 44(7), 1653–1661.
7. Choi, C., et al. (2006), Thermosensitive poly(*N*-isopropylacrylamide)-*b*-poly(ε-caprolactone) nanoparticles for efficient drug delivery system, *Polymer*, 47(13), 4571–4580.
8. Gross, M., and Maskos, M. (2005), Dye loading of unimolecular, amphiphilic polymeric nanocontainers, *Polymer*, 46(10), 3329–3336.
9. Brož, P., et al. (2005), Cell targeting by a generic receptor-targeted polymer nanocontainer platform, *J. Controlled Release*, 102(2), 475–488.
10. Santoso, S. S., et al. (2002), Structures, function and applications of amphiphilic peptides, *Curr. Opin. Coll. Interf. Sci.*, 7(5–6), 262–266.
11. Heyes, J., et al. (2006), Synthesis and characterization of novel poly(ethylene glycol)-lipid conjugates suitable for use in drug delivery, *J. Controlled Release*, 112(2), 280–290.

12. Seki, J., et al. (2004), A nanometer lipid emulsion, lipid nano-sphere (LNS®), as a parenteral drug carrier for passive drug targeting, *Int. J. Pharm.*, 273(1–2), 75–83.
13. Carlson, J. C. T., and Jena, S. S. (2006), Chemically controlled self-assembly of protein nanorings, *J. Am. Chem. Soc.*, 128(23), 7630–7638.
14. Perkins, W. R., et al. (2000), Novel therapeutic nano-particles (lipocores): Trapping poorly water soluble compounds, *Int. J. Pharm.*, 200(1), 27–39.
15. Fukui, H., et al. (2003), A novel delivery system for amphotericin B with lipid nanosphere (LNS®), *Int. J. Pharm.*, 265(1–2), 37–45.
16. Müller, R. H., et al. (2006), Oral bioavailability of cyclosporine: Solid lipid nanoparticles (SLN®) versus drug nanocrystals, *Int. J. Pharm.*, 317(1), 82–89.
17. Casadei, M. A., et al. (2006), Solid lipid nanoparticles incorporated in dextran hydrogels: A new drug delivery system for oral formulations, *Int. J. Pharm*, 325(1–2), 140–146.
18. Masuda, M., and Shimizu, T. (2004), Lipid nanotubes and microtubes: experimental evidence for unsymmetrical monolayer membrane formation from unsymmetrical bolaamphiphiles, *Langmuir*, 20(14), 5969–5977.
19. Gupta, K., et al. (2006), Nanoparticle formation from poly(acrylic acid) and oppositely charged peptides, *Biophys. Chem.*, 119(3), 303–306.
20. Costantino, L., et al. (2005), Peptide-derivatized biodegradable nanoparticles able to cross the blood–brain barrier, *J. Controlled Release*, 108(1), 84–96.
21. Cortez-Retamozo, V., and Backmann, N. (2004), Efficient cancer therapy with a nanobody-based conjugate, *Cancer Res.*, 64, 2853–2857.
22. Parekh, H. S., et al. (2006), Synthesis of a library of polycationic lipid core dendrimers and their evaluation in the delivery of an oligonucleotide with hVEGF inhibition, *Bioorg. Med. Chem.*, 14(14), 4775–4780.
23. Gupta, U., et al. (2006), A review of in vitro–in vivo investigations on dendrimers: The novel nanoscopic drug carriers, *Nanomed. Nanotechnol. Biol. Med.*, 2(2), 66–73.
24. Najlah, M., et al. (2006), Synthesis, characterization and stability of dendrimer prodrugs, *Int. J. Pharm.*, 308(1–2), 175–182.
25. Koga, T., et al. (2006), Fabrication of a switchable nano-surface composed of acidic and basic block-polypeptides, *Coll. Surf. Physicochem. Eng. Aspects*, 284–285, 521–527.
26. Yin, A., and Xu, A. (2006), Fabrication of highly-ordered and densely-packed silicon nano-needle arrays for bio-sensing applications, *Mater. Res. Soc. Symp. Proc.*, 900E.
27. Vaccari, L., et al. (2006), Porous silicon as drug carrier for controlled delivery of doxorubicin anticancer agent, *Microelectron. Eng.*, 83(4–9), 1598–1601.
28. Venkatesan, N., et al. (2005), Liquid filled nanoparticles as a drug delivery tool for protein therapeutics, *Biomaterials*, 26(34), 7154–7163.
29. Xu, Z. P., et al. (2006), Inorganic nanoparticles as carriers for efficient cellular delivery, *Chem. Eng. Sci.*, 61(3), 1027–1040.
30. Neuberger, T., et al. (2005), Superparamagnetic nanoparticles for biomedical applications: Possibilities and limitations of a new drug delivery system, *J. Magnet. Magnetic Mater.*, 293(1), 483–496.
31. Kim, J-H., and Lee, T. R. (2004), Discrete thermally responsive hydrogel-coated gold nanoparticles for use as drug-delivery vehicles, *Chem. Mater*, 16, 3647–3651.
32. Gao, X., et al. (2004), *In vivo* cancer targeting and imaging with semiconductor quantum dots, *Nat. Biotechnol.*, 22, 8.
33. Wu, C., et al. (2005), Metal nanoshells as a contrast agent in near-infrared diffuse optical tomography, *Opt. Commun.*, 253(1–3), 214–221.

34. Ostrander, K. D., et al. (1999), An in-vitro assessment of a NanoCrystal™ beclomethasone dipropionate colloidal dispersion via ultrasonic nebulization, *Eur. J. Pharm. Biopharm.*, 48(3), 207–215.
35. Muller, R. H., and Keck, C. M. (2004), Challenges and solutions for the delivery of biotech drugs—A review of drug nanocrystal technology and lipid nanoparticles, *J. Biotechnol.*, 113(1–3), 151–170.
36. Cui, F., et al. (2006), Preparation and characterization of mucoadhesive polymer-coated nanoparticles, *Int. J. Pharm.*, 316(1–2), 154–161.
37. Goodson, T., et al. (2004), Optical properties and applications of dendrimer-metal nanocomposites, *Int. Rev. Phys. Chem.*, 23, 1.
38. Panchapakesan, B. (2005), Tiny technology—Tremendous therapeutic potential, in *Oncology Issues, Nanotechnology: Part 2*.
39. Lin, C-R., et al. (2006), Magnetic behavior of core–shell particles, *J. Magnet. Magnetic Mater.*, 304(1), 34–36.
40. Layre, A., et al. (2006). Novel composite core-shell nanoparticles as busulfan carriers, *J. Controlled Release*, 111(3), 271–280.
41. Foley, S., et al. (2002), Cellular localisation of a water-soluble fullerene derivative, *Biochem. Biophys. Res. Commun.*, 294(1), 116–119.
42. Friedman, S. H., et al. (2002), Optimizing the binding of fullerence inhibitors of the HIV-1 protease through predicted increases in hydrophobic desolvation, *J. Med. Chem.*, 41, 2424–2429.
43. Zhu, Z., et al. (2003), Molecular dynamics study of the connection between flap closing and binding of fullerene-based inhibitors of the HIV-1 protease, *Biochemistry*, 42, 1326–1337.
44. Smart, S. K., et al. (2006), The biocompatibility of carbon nanotubes, *Carbon*, 44(6), 1034–1047.
45. Bianco, A., et al. (2005), Applications of carbon nanotubes in drug delivery, *Curr. Opin. Chem. Biol.*, 9(6), 674–679.
46. Luan, B., and Pan, C-Y. (2006), Synthesis and characterizations of well-defined dendrimer-like copolymers with the second and third generation based on polystyrene and poly(L-lactide), *Eur. Polym. J.*, 42, 1467–1478.
47. Tomalia, D. A., et al. (1990), Starburst dendrimers: Control of size, shape, surface chemistry, topology and flexibility, *Angew. Chem., Int. Ed. Engl.*, 29, 138.
48. Petkov, V., et al. (2005), 3D structure of dendritic and hyper-branched macromolecules by X-ray diffraction, *Solid State Commun.*, 134(10), 671–675.
49. Tomalia, D. A., et al. (1985), A new class of polymers: Starburst-dendritic macromolecules, *Polym. J.*, 17, 117–132.
50. Buhleier, E., et al. (1978), Cascade and nonskid-chain-like syntheses of molecular cavity topologies, *Synthesis*, 155–l58.
51. Tomalia, D. A., et al. (1991), Comb-burst dendrimer topology. New macromolecular architecture derived from dendritic grafting, *Macromolecules*, 24, 1435–1438.
52. Hawker, C. J., and Frechet, J. M. J. (1990), Preparation of polymers with controlled molecular architecture. A new convergent approach to dendritic macromolecules, *J. Am. Chem. Soc.*, 112, 7638–7647.
53. Miller, T. M., and Neenan, T. X. (1990), Convergent synthesis of monodisperse dendrimers based upon 1,3,5 trisubstituted benzenes, *Chem. Mater.*, 2, 346–349.
54. Hedrick, J. L., et al. (1998), Dendrimer-like star block and amphiphilic copolymers by combination of ring-opening and atom transfer radical polymerization, *Macromolecules*, 31, 8691–8705.

55. Hou, S. J., et al. (2003), Synthesis of water-soluble star-block and dendrimer-like copolymers based on poly(ethylene oxide) and poly(acrylic acid), *Macromolecules*, 36, 3874–3881.
56. Angot, S., et al. (2000), Amphiphilic stars and dendrimer-like architectures based on poly(ethylene oxide) and polystyrene, *Macromolecules*, 33, 5418–5426.
57. Reuter, J. D., et al. (1999), Inhibition of viral adhesion by sialic-acid-conjugated dendritic polymers, *Bioconjug. Chem.*, 10, 271–278.
58. Na, M., et al. (2006), Dendrimers as potential drug carriers. Part II. Prolonged delivery of ketoprofen by in vitro and in vivo studies, *Eur. J. Med. Chem.*, 41(5), 670–674.
59. Vandamme, Th. F., and Brobeck, L. (2005), Poly(amidoamine) dendrimers as ophthalmic vehicles for ocular delivery of pilocarpine nitrate and tropicamide, *J. Controlled Release*, 102(1), 23–38.
60. Zeng, H., et al. (2002), Exchange-coupled nanocomposite magnets via nanoparticle self-assembly, *Nature*, 420, 395–398.
61. Chan, W. C. W., et al. (2002), Luminescent quantum dots for multiplexed biological detection and imaging, *Curr. Opin. Biotechnol.*, 13, 40–46.
62. Qu, L. H., and Peng, X. G. (2002), Control of photoluminescence properties of CdSe nanocrystals in growth, *J. Am. Chem. Soc.*, 124, 2049–2055.
63. Murphy, C. J. (2002), Optical sensing with quantum dots, *Anal. Chem.*, 74, 520–526.
64. Parak, W. J., et al. (2003), Biological applications of colloidal nanocrystals, *Nanotechnology*, 14, 15–27.
65. Gao, X., et al. (2005), In vivo molecular and cellular imaging with quantum dots, *Curr. Opin. Biotechnol.*, 16, 63–72.
66. Leatherdale, C. A., et al. (2002), On the absorption cross section of CdSe nanocrystal quantum dots, *J. Phys. Chem. B*, 106, 7619–7622.
67. Gao, X. H., et al. (2004), In vivo cancer targeting and imaging with semiconductor quantum dots, *Nat. Biotechnol.*, 22, 969–976.
68. Xiao, Y., et al. (2003), Plugging into enzymes: Nanowiring of redox enzymes by a gold nanoparticle, *Science*, 299(5614), 1877–1881.
69. Liedberg, B., et al. (1995), Biosensing with surface plasmon resonance—How it all started, *Biosens. Bioelectron.*, 10, 1–4.
70. Chen, W. P., and Chen, J. M. (1981), Use of surface plasma waves for determination of the thickness and optical constants of thin metallic films, *J. Opt. Soc. Am.*, 71, 189–191.
71. Link, S., and El-Sayed, M. A. (1999), Spectral properties and relaxation dynamics of surface plasmon electronic oscillations in gold and silver nanodots and nanorods, *J. Phys. Chem.*, 103, 8410.
72. Jensen, T. R., et al. (2000), Nanosphere lithography: Tunable localized surface plasmon resonance spectra of silver nanoparticles, *J. Phys. Chem.*, 104, 10549.
73. Salmon, Z., et al. (1997), Surface plasmon resonance spectroscopy as a tool for investigating the biochemical and biophysical properties of membrane protein system. II. Application to biological system, *Biochim. Biophys. Acta*, 1331, 131–152.
74. Homola, J., et al. (2001), A novel multichannel surface plasmon resonance biosensor, *Sens. Actuators Chem.*, 76, 403–410.
75. Oldenburg, S. J. (1998), Nanoengineering of optical resonances, *Chem. Phys. Lett.*, 288, 243–247.
76. Sershen, S., et al. (2000), Temperature-sensitive polymer-nanoshell composites for photothermally modulated drug delivery, *J. Biomed. Mater. Res.*, 51(03), 293–298.

77. Muangsiri, W., and Kirsch, L. E. (2006), The protein-binding and drug release properties of macromolecular conjugates containing daptomycin and dextran, *Int. J. Pharm.*, 315(1–2), 30–43.
78. Rothenhausler, B., and Knoll, W. (1988), Surface-plasmon microscopy, *Nature*, 332, 615–617.
79. Piliarik, M., et al. (2005), A new surface plasmon resonance sensor for high-throughput screening applications, *Biosens. Bioelectron.*, 20, 2104–2110.
80. Averitt, R. D. (1997), Plasmon resonance shifts of Au-coated $Au_2S$ nanoshells: Insight into multicomponent nanoparticle growth, *Phys. Rev. Lett.*, 78, 4217–4220.
81. McDonnell, J. M. (2001), Surface plasmon resonance: Towards an understanding of the mechanisms of biological molecular recognition, *Curr. Opin. Chem. Biol.*, 5, 572–577.
82. Baek, S. H., et al. (2004), Surface plasmon resonance imaging analysis of hexahistidine-tagged protein on the gold thin film coated with a calix crown derivative, *Biotechnol. Bioprocess. Eng.*, 9, 143–146.
83. Oh, B. K., et al. (2003), Immunosensor for detection of *Legionella pneumophila* using surface plasmon resonance, *Biosens. Bioelectron.*, 18, 605–611.
84. Oh, B. K., et al. (2004), Surface plasmon resonance immunosensor for the detection of Salmonella typhimurium, *Biosens. Bioelectron.*, 19, 1497–1504.
85. Lehn, J. M. (1993), Supramolecular chemistry, *Science*, 260, 1762–1763.
86. Tanford, C. (1974), Theory of micelle formation in aqueous solutions, *J. Phys. Chem.*, 78, 24.
87. Desai, H., et al. (2006), Micellar characteristics of diblock polyacrylate–polyethylene oxide copolymers in aqueous media, *Eur. Polym. J.*, 42(3), 593–601.
88. Yoshii, N., and Okazaki, S. (2006), A molecular dynamics study of structural stability of spherical SDS micelle as a function of its size, *Chem. Phys. Lett.*, 425(1–3), 58–61.
89. Gochman-Hecht, H., and Bianco-Peled, H. (2006), Structure modifications of AOT reverse micelles due to protein incorporation, *J. Coll. Interf. Sci.*, 297(1), 276–283.
90. Luisi, P. L., et al. (1998), Reverse micelles as hosts for proteins and small molecules, *Biochim. Biophys. Acta*, 947(1), 209–246.
91. Munch, M. R., and Gast, A. P. (1998), Block copolymers at interfaces. 1. Micelle formation, *Macromolecules*, 21, 1360–1366.
92. Siegel, D. P. (1986), Inverted micellar intermediates and the transitions between lamellar, cubic, and inverted hexagonal lipid phases. II. Implications for membrane-membrane interactions and membrane fusion, *Biophys. J.*, 49(6), 1171–1183.
93. Sokolova, I., and Kievskya, Y. (2005), 3D Design of self-assembled nanoporous colloids, *Stud. Surf. Sci. Catal.*, 56, 433–442.
94. Zhang, Y., et al. (2004), Hollow spheres from shell cross-linked, noncovalently connected micelles of carboxyl-terminated polybutadiene and poly(vinyl alcohol) in water, *Macromolecules*, 37, 1537–1543.
95. Venema, L. C., et al. (2000), Atomic structure of carbon nanotubes from scanning tunneling microscopy, *Phys. Rev. B*, 61, 2991–2996.
96. Iijima, S., and Ichihashi, T. (1993), Single-shell carbon nanotubes of 1-nm diameter, *Nature*, 363, 603–605.
97. Vinciguerra, V., et al. (2003), Growth mechanisms in chemical vapour deposited carbon nanotubes, *Nanotechnology*, 14(6), 655–660.
98. Sonoyama, N., et al. (2006), Synthesis of carbon nanotubes on carbon fibers by means of two-step thermochemical vapor deposition, *Carbon*, 44, 1754–1761.
99. Bondi, S. N., et al. (2006), Laser assisted chemical vapor deposition synthesis of carbon nanotubes and their characterization, *Carbon*, 44, 1393–1403.

100. Gavillet, J., et al. (2002), Microscopic mechanisms for the catalyst assisted growth of single-wall carbon nanotubes, *Carbon*, 40(10), 1649–1663.
101. Lee, C. J., et al. (2002), Large-scale production of aligned carbon nanotubes by the vapor phase growth method, *Chem. Phys. Lett.*, 359(1–2), 109–114.
102. Endo, M., et al. (2005), "Buckypaper" from coaxial nanotubes, *Nature*, 433, 476.
103. Hata, K., et al. (2004), Water-assisted highly efficient synthesis of impurity-free single-walled carbon nanotubes, *Science*, 306(5700), 1362–1364.
104. Pantarotto, D., et al. (2003), Synthesis, structural characterization, and immunological properties of carbon nanotubes functionalized with peptides, *J. Am. Chem. Soc.*, 125(20), 6160–6164.
105. Kam, N. W. S., and Dai, H. (2005), Carbon nanotubes as intracellular protein transporters: Generality and biological functionality, *J. Am. Chem. Soc.*, 127(16), 6021–6026.
106. Kroto, H. W., et al. (1985), C60: Buckminsterfullerene, *Nature*, 318, 162–163.
107. Goel, A., et al. (2002), Combustion synthesis of fullerenes and fullerenic nanostructures, *Carbon*, 40, 177–182.
108. Kreatschmer, W., et al. (1990), Solid C60: New form of carbon, *Nature*, 347, 354–358.
109. Baierl, T., et al. (1996), Comparison of immunological effects of fullerene C60 and raw soot from fullerene production on alveolar macrophages and macrophage like cells in vitro, *Exp. Toxicol. Pathol.*, 48, 508–511.
110. Satoh, M., et al. (1997), Inhibitory effects of a fullerene derivative, dimalonic acid C60, on nitric oxide-induced relaxation of rabbit aorta, *Eur. J. Pharmacol.*, 327, 175–181.
111. Gharbi, N., et al. (2005), Fullerene is a powerful antioxidant in vivo with no acute or subacute toxicity, *NanoLetters*, 5, 2578–2585.
112. Tabata, Y., et al. (1997), Antitumor effect of poly(ethylene glycol)-modified fullerene, *Fullerene Sci. Technol.*, 5, 989–1007.
113. Signori, F., et al. (2005), New self-assembling biocompatible–biodegradable amphiphilic block copolymers, *Polymer*, 46(23), 9642–9652.
114. Lecommandoux, S., et al. (2006), Smart hybrid magnetic self-assembled micelles and hollow capsules, *Prog. Solid State Chem.*, 34(2–4), 171–179.
115. McNally, H., et al. (2003), Self-assembly of micro- and nano-scale particles using bio-inspired events, *Appl. Surf. Sci.*, 214(1–4), 109–119.
116. Bashir, R. (2001), Invited Review: DNA-mediated artificial nanobiostructures: State of the art and future directions, *Superlatt. Microstruct.*, 29(1), 1–16.
117. Hänel, K., et al. (2006), Manipulation of organic "needles" using an STM operated under SEM control, *Surf. Sci.*, 600(12), 2411–2416.
118. Guthold, M., et al. (1999), Investigation and modification of molecular structures with the nanoManipulator, *J. Mol. Graphics Modell.*, 17(3–4), 187–197.
119. Rogers, J. A., and Nuzzo, R. G. (2005), Recent progress in soft lithography, *Mater. Today*, 8(2), 50–56.
120. Choi, D-G., et al. (2004), 2D nano/micro hybrid patterning using soft/block copolymer lithography, *Mater. Sci. Eng.*, 24(1–2), 213–216.
121. Sellergren, B., and Allender, C. J. (2005), Molecularly imprinted polymers: A bridge to advanced drug delivery, *Adv. Drug Deliv. Rev.*, 57(12), 1733–1741.
122. van Nostrum, C. F. (2005), Molecular imprinting: A new tool for drug innovation, *Drug Discov. Today Technol.*, 2(1), 119–124.
123. Caruso, F., et al. (2000), Enzyme encapsulation in layer-by-layer engineered polymer multilayer capsules, *Langmuir*, 16, 1485–1488.

124. Caruso, F., et al. (1998), Nanoengineering of inorganic and hybrid hollow spheres by colloidal templating, *Science*, 282, 1111–1114.
125. Deotare, P. B., and Kameoka, J. (2006), Fabrication of silica nanocomposite-cups using electrospraying, *Nanotechnology*, 17, 1380–1383.
126. Lu, Y., and Chen, S. C. (2004), Micro and nano-fabrication of biodegradable polymers for drug delivery, *Adv. Drug Deliv. Rev.*, 56, 1621–1633.
127. Lu, Y., et al. (2005), Shaping biodegradable polymers as nanostructures: Fabrication and applications, *Drug Discov. Today Techno.*, 2(1), 97–102.
128. Michel, R., et al. (2002), A novel approach to produce biologically relevant chemical patterns at the nanometer scale: Selective molecular assembly patterning combined with colloidal lithography, *Langmuir*, 18, 8580–8586.
129. Hanarp, P., et al. (1999), Nanostructured model biomaterial surfaces prepared by colloidal lithography, *Nanostruct. Mater.*, 12(1), 429–432.
130. Di Fabrizio, E., et al. (2004), X-ray lithography for micro- and nano-fabrications at ELETTRA for interdisciplinary applications, *J. Phys. Condens. Matter*, 16, 3517–3535.
131. Couvreur, P., et al. (2002), Nanocapsule technology: A review, *Crit. Rev. Ther. Drug Carrier Syst.*, 19(2), 99–134.
132. Watnasirichaikul, S., et al. (2000), Preparation of biodegradable insulin nanocapsules from biocompatible microemulsions, *Pharm. Res.*, 17(6), 684–689.
133. Hitt, J., et al. (2002), Nanoparticles of poorly water soluble drugs made via a continuous precipitation process, in *Proceedings, American Association of Pharmaceutical Scientists*, annual meeting, Toronto.
134. Wang, Y., et al. (2004), Polymer coating/encapsulation of nanoparticles using a supercritical anti-solvent process, *J. Supercrit. Fluids*, 28, 85–99.
135. Perez, C., et al. (2001), Poly(lactic acid)-poly(ethylene glycol) nanoparticles as new carriers for the delivery of plasmid DNA, *J. Controlled Release*, 75(1–2), 211–224.
136. Dickert, F. L. (2003), Nano- and micro-structuring of sensor materials—From molecule to cell detection, *Synthetic Metals*, 138(1–2), 65–69.
137. Zeng, J., et al. (2003), Biodegradable electrospun fibers for drug delivery, *J. Controlled Release*, 92(3), 227–231.
138. Xu, X., et al. (2005), Ultrafine medicated fibers electrospun from W/O emulsions, *J. Controlled Release*, 108(1), 33–42.
139. Nicolau, D. V., et al. (1997), Bionanostructures built on e-beam-assisted functionalized polymer surfaces, *Proc. SPIE*, 3241.
140. Sinha, P. M., et al. (2004), Nanoengineered device for drug delivery application, *Nanotechnology*, 15, 585–589.
141. Maloney, J. M., et al. (2005), In vivo biostability of CVD silicon oxide and silicon nitride films, *Mater. Res. Soc. Symp. Proc.*, 872.
142. Sugai, T., and Yoshida, H. (2003), New synthesis of high-quality double-walled carbon nanotubes by high-temperature pulsed arc discharge, *NanoLetters*, 3(6), 769–773.
143. Du, F., et al. (2006), The synthesis of single-walled carbon nanotubes with controlled length and bundle size using the electric arc method, *Carbon*, 44(7), 1327–1330.
144. Tamir, S., and Drezner, Y. (2006), New aspects on pulsed laser deposition of aligned carbon nanotubes, *Appl. Surf. Sci.*, 252(13), 4819–4823.
145. Aratono, Y., et al. (2005), Formation of fullerene($C_{60}$) by laser ablation in superfluid helium at 1.5 K, *Chem. Phys. Lett.*, 408(4–6), 247–251.
146. He, C., et al. (2006), A practical method for the production of hollow carbon onion particles, *J. Alloys Compounds*, 425(1–2), 329–333.

147. Liu, T.-C., et al. (2006), Synthesis of carbon nanocapsules and carbon nanotubes by an acetylene flame method, *Carbon*, 44(10), 2045–2050.
148. Rodriguez-Hernandez, J., et al. (2005), Toward "smart" nano-objects by self-assembly of block copolymers in solution, *Prog. Polym. Sci.*, 30, 691–724.
149. Birac, J. J. (2006), Geometry based design strategy for DNA nanostructures, *J. Mol. Graphics Modell.*, 25, 470–480.
150. Samori, B., and Zuccheri, G. (2004), DNA codes for nanoscience, *Angew. Chem. Int. Ed. Engl.*, 44, 1166–1181.
151. Seeman, N. C. (1998), DNA nanotechnology: Novel DNA constructions, *Annu. Rev. Biophys. Biomol. Struct.*, 27, 225–248.
152. Seeman, N. C. (1997), DNA components for molecular architecture, *Acc. Chem. Res.*, 30, 347–391.
152a. Rothemund, P. W. K. (2006), Folding DNA to create nanoscale shapes and patterns, *Nature*, 440, 297–302.
153. Williams, K. A., et al. (2002), Nanotechnology: Carbon nanotubes with DNA recognition, *Nature*, 420, 761.
154. Niemeyer, C. M. (2004), Semi-synthetic DNA–protein conjugates: Novel tools in analytics and nanobiotechnology, *Biochem. Soc.Trans.*, 32, 51–53.
155. Asakawa, T., et al. (2006), Build-to-order nanostructures using DNA self-assembly, *Thin Solid Films*, 509(1–2), 85–93.
156. Högberg, B., et al. (2006), Study of DNA coated nanoparticles as possible programmable self-assembly building blocks, *Appl. Surf. Sci.*, 252(15), 5538–5541.
157. Seeman, N. C. (1982), Nucleic acid junctions and lattices, *J. Theor. Biol.*, 99, 237–247.
158. Chen, J., and Seeman, N. C. (1991), Synthesis from DNA of a molecule with the connectivity of a cube, *Nature*, 350, 631–633.
159. Li, X. J., et al. (1996), Antiparallel DNA double crossover molecules as components for nanoconstruction, *J. Am. Chem. Soc.*, 118, 6131–6140.
160. Fu, T. J., and Seeman, N. C. (1993), DNA double crossover structures, *Biochemistry*, 32, 3211–3220.
161. Zhang, Y., and Seeman, N. C. (1992), A solid-support methodology for the construction of geometrical objects from DNA, *J. Am. Chem. Soc.*, 114, 2656–2663.
162. Niemeyer, C. M., et al. (1999), DNA-directed immobilization: Efficient, reversible and site-selective surface binding of proteins by means of covalent DNA-streptavidin conjugates, *Anal. Biochem.*, 268, 5463.
163. Kallenbach, N. R., et al. (1983), An immobile nucleic aid junction constructed from oligonucleotides, *Nature*, 305, 829–831.
164. Ma, R. I., et al. (1986), Three-arm nucleic acid junctions are flexible, *Nucl. Acids Res.*, 14, 9745–9753.
165. Brucale, M., et al. (2006), Mastering the complexity of DNA nanostructures, *Trends Biotechnol.*, 24(5), 235–243.
166. Nakayama, Y. (2002), Scanning probe microscopy installed with nanotube probes and nanotube tweezers, *Ultramicroscopy*, 91(1–4), 49–56.
167. Lü, J. H. (2004), Nanomanipulation of extended single-DNA molecules on modified mica surfaces using the atomic force microscopy, *Coll. Surf. B Biointerf.*, 39(4), 177–180.
168. Decossas, S., et al. (2003), Nanomanipulation by atomic force microscopy of carbon nanotubes on a nanostructured surface, *Surf. Sci.*, 543(1–3), 57–62.
169. Ishii, Y., et al. (2001), Single molecule nanomanipulation of biomolecules, *Trends Biotechnol.*, 19(6), 211–216.

170. Rogers, J. A., et al. (1997), Using an elastomeric phase mask for sub-100 nm photolithography in the optical near field, *Appl. Phys. Lett.*, 70, 2658–2660.
171. Rogers, J. A., et al. (1998), Generating similar to 90 nanometer features using near-field contact-mode photolithography with an elastomeric phase mask, *J. Vac. Sci. Technol.*, 16, 59–68.
172. Wulff, G., et al. (1991), Racemic resolution of free sugars with macroporous polymers prepared by molecular imprinting. Selective dependence on the arrangement of functional groups versus spatial requirements, *J. Org. Chem.*, 56(1), 395–400.
173. Sellergren, B., et al. (1988), Highly enantioselective and substrate-selective polymers obtained by molecular imprinting utilizing noncovalent interactions—NMR and chromatrographic studies on the nature of recognition, *J. Am. Chem. Soc.*, 110, 5853–5860.
174. Lubke, C., et al. (2000), Imprinted polymers prepared with stoichiometric template-monomer complexes: Efficient binding of ampicillin from aqueous solutions, *Macromolecules*, 33, 5098–5105.
175. Tskruk, V. V., et al. (1997), Self-assembled multilayer films from dendrimers, *Langmuir*, 13(8), 2171–2175.
176. Schulze, K., and Kirstein, S. (2005), Layer-by-layer deposition of $TiO_2$ nanoparticles, *Appl. Surf. Sci.*, 246(4), 415–419.
177. Ji, J., and Shen, J. (2005), Electrostatic self-assembly and nanomedicine. Engineering in medicine and biology society, in *IEEE-EMBS 27th Annual International Conference Proc.*, Shanghai, pp. 720–722.
178. Beers, A. M., et al. (1983), CVD silicon structures formed by amorphous and crystalline growth, *J. Crystal Growth*, 64(3), 563–557.
179. Kyung, S., et al. (2006), Deposition of carbon nanotubes by capillary-type atmospheric pressure PECVD, *Thin Solid Films*, 506–507, 268–273.
180. Goodman, C., and Pessa, M. (1986), Atomic layer epitaxy, *J. Appl. Phys.*, 60(3), 65–81.
181. Suntola, T. (1992), Atomic layer epitaxy, *Thin Solid Films*, 216(1), 84–89.
182. Xia, C., et al. (1998), Metal-organic chemical vapor deposition of Sr-Co-Fe-O films on porous substrates, *J. Mater. Res.*, 13(1), 173.
183. Choy, K. L. (2003), Chemical vapour deposition of coatings, *Prog. Mater. Sci.*, 48(2), 57–170.
184. Kaul, A. R., and Seleznev, B. V. (1993), New principle of feeding for flash evaporation MOCVD devices, *J. Phys.*, 3, 375–378.
185. Douard, A., and Maury, F. (2006), Nanocrystalline chromium-based coatings deposited by DLI-MOCVD under atmospheric pressure from Cr(CO)6, *Surf. Coat. Technol.*, 200, 6267–6271.
186. Lau, K. K. S. (2001), Hot-wire chemical vapor deposition (HWCVD) of fluorocarbon and organosilicon thin films, *Thin Solid Films*, 395(1–2), 288–291.
187. Ren, Z. F., et al. (1998), Synthesis of large arrays of well-aligned carbon nanotubes on glass, *Science*, 282(5391), 1105–1107.
188. Han, B. W., et al. (2001), Growth of *in situ* $CoSi_2$ layer by metalorganic chemical vapor deposition on Si tips and its field-emission properties, *J. Vac. Sci. Technol. Microelectron. Nanometer Struct.*, 19(2), 533–536.
189. Kuo, T.-F., et al. (2001), Microwave-assisted chemical vapor deposition process for synthesizing carbon nanotubes, *J. Vac. Sci. Technol. Microelectron. Nanometer Struct.*, 19(3), 1030–1033.
190. McSporran, N., et al. (2002), Preliminary studies of atmospheric pressure plasma enhanced CVD (AP-PECVD) of thin oxide films, *Phys. IV France*, 12, 4–17.

191. Chang, J. P., et al. (2001), Rapid thermal chemical vapor deposition of zirconium oxide for metal-oxide-semiconductor field effect transistor application, *J. Vac. Sci. Technol. B Microelectron. Nanometer Struct.*, 19(5), 1782–1787.
192. Huang, Z. P., et al. (1998), Growth of highly oriented carbon nanotubes by plasma-enhanced hot filament chemical vapor deposition, *Appl. Phys. Lett.*, 73(26), 3845–3847.
193. Huczko, A. (2002), Synthesis of aligned carbon nanotubes, *Appl. Phys. A Mater. Sci. Process.*, 74(5), 617–638.
194. Lim, C.-Y., et al. (2004), Development of an electrodeposited nanomold from compositionally modulated alloys, *J. Appl. Electrochem.*, 34, 857–866.
195. Micheletto, R., et al. (1995), A simple method for the production of a two-dimensional, ordered array of small latex particles, *Langmuir*, 11, 3333–3336.
196. Boneberg, J., et al. (1997), The formation of nano-dot and nano-ring structures in colloidal monolayer lithography, *Langmuir*, 13, 7080–7084.
197. Blanco, A., et al. (2000), Large-scale synthesis of a silicon photonic crystal with a complete three-dimensional bandgap near 1.5 micrometres, *Nature*, 405, 437–440.
198. Denis, F. A., et al. (2004), Nanoscale chemical patterns fabricated by using colloidal lithography and self-assembled monolayers, *Langmuir*, 20, 9335–9339.
199. Cadotte, J. E. (1981), U.S. Patent 4,277,344. Interfacially synthesized reverse osmosis membrane.
200. Krauel, K., et al. (2005), Using different structure types of microemulsions for the preparation of poly(alkylcyanoacrylate) nanoparticles by interfacial polymerization, *J. Controlled Release*, 106(1–2), 76–87.
201. Govender, T., et al. (1999), PLGA nanoparticles prepared by nanoprecipitation: Drug loading and release studies of a water soluble drug, *J. Controlled Release*, 57(2), 171–185.
202. Chorny, M., et al. (2002), Lipophilic drug loaded nanospheres prepared by nanoprecipitation: Effect of formulation variables on size, drug recovery and release kinetics, *J. Controlled Release*, 83(3), 389–400.
203. Desgouilles, S., et al. (2003), The design of nanoparticles obtained by solvent evaporation: A comprehensive study, *Langmuir*, 19(22), 9504–9510.
204. Zambaux, M. F. (1998), Influence of experimental parameters on the characteristics of poly(lactic acid) nanoparticles prepared by a double emulsion method, *J. Controlled Release*, 50(1–3), 31–40.
205. Cao, H., et al. (2002), Fabrication of 10 nm enclosed nanofluidic channels, *Appl. Phys. Lett.*, 81(1), 174–176.
206. Scheer, H.-C., et al. (1998), Problems of the nanoimprinting technique for nanometer scale pattern definition, *J. Vac. Sci. Technol. B Microelectron. Nanometer Struct.*, 16(6), 3917–3921.
207. Kim, J-S., and Reneker, D. H. (1999), Polybenzimidazole nanofiber produced by electrospinning, *Polym. Eng. Sci.*, 39(5), 849–854.
208. Abidian, M. R., et al. (2006), Conducting-polymer nanotubes for controlled drug release, *Adv. Mater.*, 18, 405–409.
209. Arias, J. L., et al. (2006), Preparation and characterization of carbonyl iron/poly(butylcyanoacrylate) core/shell nanoparticles, *J. Collo. Interf. Sci.*, 299(2), 599–607.
210. Cui, F., et al. (2006), Preparation and characterization of mucoadhesive polymer-coated nanoparticles, *Int. J. Pharm.*, 316(1–2), 154–161.
211. Huang, C.-Y., and Lee, Y.-D. (2006), Core-shell type of nanoparticles composed of poly[(*n*-butyl cyanoacrylate)-*co*-(2-octyl cyanoacrylate)] copolymers for drug delivery

application: Synthesis, characterization and *in-vitro* degradation, *Int. J. Pharm.*, 325(1–2), 132–139.

212. Hirsjärvi, S., et al. (2006), Layer-by-layer polyelectrolyte coating of low molecular weight poly(lactic acid) nanoparticles, *Coll. Surf. B Biointerf.*, 49(1), 93–99.
213. Schubert, M. A., and Müller-Goymann, C. C. (2005), Characterisation of surface-modified solid lipid nanoparticles (SLN): Influence of lecithin and nonionic emulsifier, *Eur. J. Pharm. Biopharm.*, 61(1–2), 77–86.
214. Lo, C.-L. (2005), Preparation and characterization of intelligent core-shell nanoparticles based on poly(D,L-lactide)-*g*-poly(*N*-isopropyl acrylamide-*co*-methacrylic acid), *J. Controlled Release*, 104(3), 477–488.
215. Garcia-Fuentes, M. (2005), Design and characterization of a new drug nanocarrier made from solid–liquid lipid mixtures, *J. Coll. Interf. Sci.*, 285(2), 590–598.
216. Lochmann, D., et al. (2005), Albumin–protamine–oligonucleotide nanoparticles as a new antisense delivery system. Part 1: Physicochemical characterization, *Eur. J. Pharm. Biopharm.*, 59(3), 419–429.
217. Müller-Goymann, C. C. (2004), Physicochemical characterization of colloidal drug delivery systems such as reverse micelles, vesicles, liquid crystals and nanoparticles for topical administration, *Eur. J. Pharm. Biopharm.*, 58(2), 343–356.
218. Weyermann, J., et al. (2004), Physicochemical characterisation of cationic polybutylcyanoacrylat—Nanoparticles by fluorescence correlation spectroscopy, *Eur. J. Pharm. Biopharm.*, 58(1), 25–35.
219. Bootz, A., et al. (2004), Comparison of scanning electron microscopy, dynamic light scattering and analytical ultracentrifugation for the sizing of poly(butyl cyanoacrylate) nanoparticles, *Eur. J. Pharm. Biopharm.*, 57(2), 369–375.
220. Merroun, M., et al. (2007), Spectroscopic characterization of gold nanoparticles formed by cells and S-layer protein of *Bacillus sphaericus* JG-A12, *Mater. Sci. Eng.*, 27(1), 188–192.
221. Lu, Q. H., et al. (2006), Synthesis and characterization of composite nanoparticles comprised of gold shell and magnetic core/cores, *J. Magnet. Magnetic Mater.*, 301(1), 44–49.
222. Vdovenkova, T., et al. (2000), Silicon nanoparticles characterization by Auger electron spectroscopy, *Surf. Sci.*, 454–456, 952–956.
223. Bonini, M., et al. (2006), Synthesis and characterization of magnetic nanoparticles coated with a uniform silica shell, *Mater. Sci. Eng.*, 26(5–7), 745–750.
224. Pilon, L. N., et al. (2006), Synthesis and characterisation of new shell cross-linked micelles with amine-functional coronas, *Eur. Polym. J.*, 42(7), 1487–1498.
225. Castro, E., et al. (2006), Characterization of triblock copolymer $E_{67}S_{15}E_{67}$—Surfactant interactions, *Chem. Phys.*, 325(2–3), 492–498.
226. Tao, L., and Uhrich, K. E. (2006), Novel amphiphilic macromolecules and their in vitro characterization as stabilized micellar drug delivery systems, *J. Coll. Interf. Sci.*, 298(1), 102–110.
227. Gaucher, G., et al. (2005), Block copolymer micelles: Preparation, characterization and application in drug delivery, *J. Controlled Release*, 109(1–3), 169–188.
228. Yu, L., et al. (2006), Determination of critical micelle concentrations and aggregation numbers by fluorescence correlation spectroscopy: Aggregation of a lipopolysaccharide, *Anal. Chim. Acta*, 556(1), 216–225.
229. Kataoka, K., et al. (2001), Block copolymer micelles for drug delivery: Design, characterization and biological significance, *Adv. Drug Deliv. Rev.*, 47(1), 113–131.

230. Hagan, S. A., et al. (1996), Polylactide-poly(ethylene glycol) copolymers as drug delivery systems. 1. characterization of water dispersible micelle-forming systems, *Langmuir*, 12(9), 2153–2161.

231. Shin, I. G., et al. (1998), Methoxy poly(ethylene glycol)/epsilon-caprolactone amphiphilic block copolymeric micelle containing indomethacin. I. Preparation and characterization, *J. Controlled Release*, 51(1), 1–11.

232. Morishima, Y., et al. (1995), Characterization of unimolecular micelles of random copolymers of sodium 2-(acrylamido)-2-methylpropanesulfonate and methacrylamides bearing bulky hydrophobic substituents, *Macromolecules*, 28, 2874–2881.

233. Vansteenkiste, S. O., and Davies, M. C. (1998), Scanning probe microscopy of biomedical interfaces, *Prog. Surf. Sci.*, 57(2), 95–136.

234. Binnig, G., and Rohrer, H. (2000), Scanning tunneling microscopy, *IBM J. Res. Dev.*, 44(1–2), 279.

235. Chen, C. J., and Smith, W. F. (1994), Introduction to scanning tunneling microscopy, *Am. J. Phys.*, 62(6), 573–574.

236. Giessibl, F. J. (2005), AFM's path to atomic resolution, *Mater. Today*, 8(5), 32–41.

237. Durig, U., et al. (1986), Near-field optical-scanning microscopy, *J. Appl. Phys.*, 59(10), 3318–3327.

238. Ash, E. A., and Nichols, G. (1972), Super-resolution aperture scanning microscope, *Nature*, 237, 510.

239. Merritt, G. (1998), A compact near-field scanning optical microscope, *Ultramicroscopy*, 71(1–4), 183–189.

240. Albrecht, T. R., et al. (1991), Frequency modulation detection using high-*Q* cantilevers for enhanced force microscope sensitivity, *J. Appl. Phys.*, 69, 668–673.

241. Ohta, M., et al. (1995), Atomically resolved image of cleaved surfaces of compound semiconductors observed with an ultrahigh vacuum atomic force microscope, *J. Vac. Sci. Technol.*, 13(3), 1265–1267.

242. Binnig, G., et al. (1986), Atomic force microscope, *Phys. Rev. Lett.*, 56(9), 930–933.

243. Zhong, Q., et al. (1993), Fractured polymer/silica fiber surface studied by tapping mode atomic force microscopy, *Surf. Sci.*, 290(1–2), 688–692.

244. Stark, R. W., and Drobek, T. (1998), Determination of elastic properties of single aerogel powder particles with the AFM, *Ultramicroscopy*, 75(3), 161–169.

245. Müller, F., et al. (1997), Applications of scanning electrical force microscopy, *Microelectron. Reliabil.*, 37(10–11), 1631–1634.

246. Cappella, B., et al. (1997), Improvements in AFM imaging of the spatial variation of force–distance curves: On-line images, *Nanotechnology*, 8, 82–87.

247. Hamers, R. J. (1989), Atomic-resolution surface spectroscopy with the scanning tunneling microscope, *Annu. Rev. Phys. Chem.*, 40, 531–559.

248. Laiho, R., et al. (1997), Spin-polarized scanning tunnelling microscopy with detection of polarized luminescence emerging from a semiconductor tip, *J. Phys. Condens. Matter*, 9, 5697–5707.

249. Binnig, G., et al. (1982), Surface studies by scanning tunneling microscopy, *Phys. Rev. Lett.*, 49, 57–61.

250. Ciraci, S., and Tekman, E. (1989), Theory of transition from the tunneling regime to point contact in scanning tunneling microscopy, *Phys. Rev.*, 40, 11969–11972.

251. Doyen, G., et al. (1993), Green-function theory of scanning tunneling microscopy: Tunnel current and current density for clean metal surfaces, *Phys. Rev.*, 47, 9778–9790.

252. Fontaine, P. A., et al. (1998), Characterization of scanning tunneling microscopy and atomic force microscopy-based techniques for nanolithography on hydrogen-passivated silicon, *J. Appl. Phys.*, 84(1), 1776–1781.
253. Wiesendanger, R. (1994), Contributions of scanning probe microscopy and spectroscopy to the investigation and fabrication of nanometer-scale structures, *J. Vac. Sci. Technol. Microelectron. Nanometer Struct.*, 12(2), 515–529.
254. Binnig, G., et al. (1986), Atomic force microscopy, *Phys. Rev. Lett.*, 56, 930–933.
255. de Souza Pereira, R. (2001), Atomic force microscopy as a novel pharmacological tool, *Biochem. Pharmacol.*, 62(8), 975–983.
256. Rajagopalan, R. (2000), Atomic force and optical force microscopy: Applications to interfacial microhydrodynamics, *Coll. Surf. Physicochem. Eng. Aspects*, 174(1–2), 253–267.
257. Santos, N. C., and Castanho, M. A. R. B. (2004), An overview of the biophysical applications of atomic force microscopy, *Biophys. Chem.*, 107(2), 133–149.
258. Ferin, J., et al. (1992), Pulmonary retention of ultrafine and fine particles in rats, *Am. J. Respir. Cell Mol. Biol.*, 6, 535–542.
259. Li, X. Y., et al. (1999), Short term inflammatory responses following intratracheal instillation of fine and ultrafine carbon black in rats, *Inhal. Toxicol.*, 11, 709–731.
260. Oberdorster, G., et al. (1995), Association of particulate air pollution and acute mortality: Involvement of ultrafine particles? *Inhal. Toxicol.*, 7, 111–124.
261. Li, X. Y., et al. (1997), In vivo and in vitro pro-inflammatory effects of particulate air pollution (PM10), *Environ. Health Perspect.*, 105, 1279–1283.
262. Brüske-Hohlfeld, I., et al. (2004), Epidemiology of nanoparticles, in *Proc. of First International Symposium on Occupational Health Implications of Nanomaterials*.
263. Warheit, D. B., et al. (2004), Pulmonary bioassay toxicity study in rats with single wall carbon nanotubes, in *Proc. of First International Symposium on Occupational Health Implications of Nanomaterials*.
264. Klumpp, C., et al. (2006), Functionalized carbon nanotubes as emerging nanovectors for the delivery of therapeutics, *Biochim. Biophys. Acta*, 1758, 404–412.
265. Muller, J. (2004), Respiratory toxicity of carbon nanotubes, in *Proc. of First International Symposium on Occupational Health Implications of Nanomaterials*.
266. Tinkle, S. (2004), Dermal penetration of nanoparticles, in *Proc. of First International Symposium on Occupational Health Implications of Nanomaterials*.

## SUGGESTED READING

### General

Dai, L. (2004), *Intelligent Macromolecules for Smart Devices: From Materials Synthesis to Device Applications*, Springer.

Di Ventra, M., Evoy, S., and Heflin, J. R. (2004), *Introduction to Nanoscale Science and Technology*, Springer.

Lyshevski, S. E. (2005), *Nano- and Mirco-Electromechnical Systems: Fundamentals of Nano and Microengineering*, CRC Press, Boca Raton, FL.

Roco, M. C., and Bainbridge, W. S. (2003), *Converging Technologies for Improving Human Performance: Nanotechnology, Biotechnology*, Springer.

## Nano–Pharmaceutical Materials and Structures

Hirsch, A., and Brettreich, M. (2004), *Fullerenes: Chemistry and Reactions*, Wiley, Hoboken, NJ.

Cao, G. (2004), *Nanostructures & Nanomaterials: Synthesis, Properties & Applications*, Imperial College Press.

Haley, M. M., and Tykwinski, R. R. (2006), *Carbon-Rich Compounds: From Molecules to Materials*, Wiley, Hoboken, NJ.

Harris, P. J. (2001), *Carbon Nanotubes and Related Structures*, Cambridge University Press.

Kumar, C. S. S. R. (2005), *Biofunctionalization of Nanomaterials*, Wiley, Hoboken, NJ.

Kumar, C. S. S. R. (2006a), *Nanomaterials for Cancer Therapy*, Wiley, Hoboken, NJ.

Kumar, C. S. S. R. (2006b), *Biological and Pharmaceutical Nanomaterials*, Wiley, Hoboken, NJ.

Lazzari, M., Liu, G., and Lecommandoux, S. (2006), *Block Copolymers in Nanoscience*, Wiley, Hoboken, NJ.

Schmid, G. (2004), *Nanoparticles: From Theory to Application*, Wiley, Hoboken, NJ.

Tadros, T. F. (2005), *Applied Surfactants: Principles and Applications*, Wiley, Hoboken, NJ.

Vogtle, F. (2000), *Dendrimers III: Design, Dimension, Function*, Springer.

Zhang, J., Zhang, J., Wang, Z., Liu, J., and Chen, S. (2002), *Self-Assembled Nanostructures*, Springer.

## Manufacturing Technologies

Butt, H.-J., Graf, K., and Kappl, M. (2006), *Physics and Chemistry of Interfaces*, Wiley, Hoboken, NJ.

Champion, Y., and Fecht, H. (2004), *Nano-Architectured and Nanostructured Materials: Fabrication, Control and Properties*, Wiley, Hoboken, NJ.

Elimelech, M., Jia, X., Gregory, J., and Williams, R. (1998), *Particle Deposition & Aggregation: Measurement, Modelling and Simulation*, Elsevier, New York.

Hoch, H. C., Jelinski, L. W., and Craighead, H. G. (1996), *Nanofabrication and Biosystems*, Cambridge University Press.

Köhler, M., and Fritzsche, W. (2004), *Nanotechnology: An Introduction to Nanostructuring Techniques*, Wiley, Hoboken, NJ.

Komiyama, M., Takeuchi, T., Mukawa, T., and Asanuma, H. (2003), *Molecular Imprinting: From Fundamentals to Applications*, Wiley, Hoboken, NJ.

Kroschwitz, J. I. (1989), *Polymers: Biomaterials and Medical Applications*, Wiley, New York.

Kumar, C. S. S. R., Hormes, J., and Leuschner, C. (2005), *Nanofabrication Towards Biomedical Applications: Techniques, Tools, Applications, and Impact*, Wiley, Hoboken, NJ.

Osada, Y., and Nakagawa, T. (1992), *Membrane Science and Technology*, Marcel Dekker, New York.

Pierson, H. (1999), *Handbook of Chemical Vapor Deposition: Principles, Technologies and Applications*, William Andrew.

Ramakrishna, S., Fujihara, K., Teo, W., Lim, T., and Ma, Z. (2005), *An Introduction to Electrospinning and Nanofibers*, World Scientific.

Rotello, V. M. (2004), *Nanoparticles: Building Blocks for Nanotechnology*, Springer.

Sotomayor Torres, C. M. (2004), *Alternate Lithography: Unleashing the Potentials of Nanotechnology*, Springer.

Wise, D. L. (2000), *Handbook of Pharmaceutical Controlled Release Technology*, Marcel Dekker, New York.

## Characterization Methods

Binks, B. P. (1999), *Modern Characterization Methods of Surfactant Systems*, Marcel Dekker, New York.

Brittain, H. G. (1995), *Physical Characterization of Pharmaceutical Solids*, Marcel Dekker, New York.

Brown, W., Ed. (1996), *Light Scattering: Principles and Development*, Oxford University Press.

Buyana, T. (1997), *Molecular Physics*, World Scientific.

Chung, F. H., and Smith, D. K., Eds. (1999), *Industrial Applications of X-Ray Diffraction*, Marcel Dekker, New York.

Jena, P. B., and Hoerber, J. K. H., Eds. (2006), *Force Microscopy: Applications in Biology and Medicine*, Wiley, Hoboken, NJ.

Jenkins, R., and Jenkins, J. (1995), *Quantitative X-Ray Spectrometry*, Marcel Dekker, New York.

Meyer, E., Hug, H. J., and Bennewitz, R. (2003), *Scanning Probe Microscopy*, Springer.

Pethrick, R. A., and Viney, C., Eds. (2003), *Techniques for Polymer Organisation and Morphology Characterisation*, Wiley, Hoboken, NJ.

Rosoff, M. (2002), *Nano-Surface Chemistry*, Marcel Dekker, New York.

Sharma, A., and Schulman, S. G. (1999), *Introduction to Fluorescence Spectroscopy*, Wiley, New York.

Sibilia, J. P. (1988), *A Guide to Materials Characterization and Chemical Analysis*, Wiley, New York.

Watts, J. F., and Wolstenholme, J. (2004), *An Introduction to Surface Analysis by XPS and AES*, Wiley, Hoboken, NJ.

Wyckoff, R. W. G. (1971), *Crystal Structures*, Wiley, New York.

## Toxicology

Fan, A. M. (1996), *Toxicology and Risk Assessment: Principles, Methods & Applications*, Marcel Dekker, New York.

Gad, S. C. (2004), *Safety Pharmacology in Pharmaceutical Development and Approval*, CRC Press, Boca Raton, FL.

Gardner, D. E. (2005), *Toxicology of the Lung*, CRC Press, Boca Raton, FL.

Kumar, C. S. S. R. (2006), *Nanomaterials: Toxicity, Health and Environmental Issues*, Wiley, Hoboken, NJ.

Meeks, R. G. (1991), *Dermal and Ocular Toxicology: Fundamentals and Methods*, CRC Press, Boca Raton, FL.

# OIL-IN-WATER NANOSIZED EMULSIONS: MEDICAL APPLICATIONS

Shunmugaperumal Tamilvanan*
*University of Antwerp, Antwerp, Belgium*

**Contents**

1 Introduction
2 Generations of Oil-in-Water Nanosized Emulsions
  2.1 First-Generation Emulsion
  2.2 Second-Generation Emulsion
  2.3 Third-Generation Emulsion
  2.4 Unique Property of Third-Generation Emulsion
3 Preparation Methods for Drug-Free/Loaded Oil-in-Water Nanosized Emulsions
4 Excipient Inclusion: Oil-in-Water Nanosized Emulsions
5 Medical Applications of Oil-in-Water Nanosized Emulsions
  5.1 Parenteral Routes
  5.2 Ocular Routes
  5.3 Nasal Route
  5.4 Topical Route
6 Future Perspective
  References

## 1 INTRODUCTION

It is estimated that 40% or more of bioactive substances being identified through combinatorial screening programs are poorly soluble in water [1, 2]. Consequently, the drug molecules belonging to these categories cannot be easily incorporated into aqueous-cored/based dosage forms at adequate concentrations, and thus the clinical

*Current address: Department of Pharmaceutics, Arulmigu Kalasalingam College of Pharmacy, Anand Nagar, Krishnankoil, India

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

efficacy of highly lipophilic drugs is being impeded. Furthermore, it is well established that the pharmaceutical industry will face more difficulties in formulating and developing novel dosage forms of new chemical entities since 50–60% of these molecules are lipophilic in nature and often exhibit hydrophobic character. Among the different innovative formulation approaches that have been suggested for enhancing lipophilic drug absorption and then clinical efficacy, lipid-based colloidal drug delivery systems such as nanosized emulsions recognized particularly for overcoming the formulation and bioavailability-related problems of such drug molecules. Other nomenclatures are also being utilized often in the medical literature to refer to nanosized emulsions, including miniemulsions [3], ultrafine emulsions [4], and submicrometer emulsions [5, 6]. The term *nanosized emulsion* [7] is preferred because in addition to giving an idea of the nanoscale size range of the dispersed droplets, it is concise and avoids misinterpretation with the term *microemulsion* (which refers to thermodynamically stable systems). Hence, nanosized emulsions are heterogenous dispersions of two immiscible liquids [oil-in-water (o/w) or water-in-oil (w/o)], and they are subjected to various instability processes such as aggregation, flocculation, coalescence, and therefore eventual phase separation according to the second law of thermodynamics. However, the physical stability of nanosized emulsions can substantially be improved with the help of suitable emulsifiers that are capable of forming a mono- or multilayer coating around the dispersed liquid droplets in such a way as to reduce interfacial tension or increase droplet–droplet repulsion. Depending on the concentrations of these three components (oil–water–emulsifier) and the efficiency of the emulsification equipment/techniques used to reduce droplet size, the final nanosized emulsion may be in the form of o/w, w/o, macroemulsion, micrometer emulsion, submicrometer emulsion, and double or multiple emulsions (o/w/o and w/o/w). Preparation know-how, potential application, and other information pertinent to w/o emulsions [8], macroemulsions [9–11], microemulsions [12, 13], and multiple emulsions [14] are thoroughly covered elsewhere. In addition, some studies have compared the performance of different emulsified systems (macroemulsions, microemulsion, multiple emulsions, and gel emulsions) prepared with similar oils and surfactants for applications such as controlled drug release [15] or drug protection [16]. Similarly the state of the art of so-called oxygen carriers or perfluorocarbon emulsions, dispersions containing submicrometer/nanosized fluoroorganic particles in water, is also thoroughly covered in the literature [17–19] and readers can refer to these complete and interesting articles.

Possible usefulness as carriers stems from the nanosized emulsion's ability to solubilize substantial amounts of hydrophilic/hydrophobic drug either at the innermost (oil or water) phase or at the o/w or w/o interfaces. While hydrophilic drugs are contained in the aqueous phase of a w/o-type emulsion or at the w/o interface of the system, hydrophobic drugs could be incorporated within the inner oil phase of an o/w-type emulsion or at the o/w interface of the system. It appears that the choice of the type of emulsion to be used therefore depends, to a large extent, upon the physicochemical properties of the drug. Between w/o and o/w types, the o/w type of nanosized emulsions would be preferred in order to successfully exploit the advantages of an emulsion carrier system. Additionally, within the o/w type, simple modifications on surface/interface structures of emulsions can be made. For instance, incorporating an emulsifier molecule alone or in a specific combination that is capable of producing either positive or negative charges over the emulsified droplets surface will lead to the formation of surface-modified emulsions. Based on these

surface modifications, the o/w-type nanosized emulsions can be divided into cationic and anionic emulsions.

The o/w nanosized emulsions have many appealing properties as drug carriers. They are biocompatible, biodegradable, physically stable, and relatively easy to produce on a large scale using proven technology [20]. Due to their subcellular and submicrometer size, emulsions are expected to penetrate deep into tissues through fine capillaries and even cross the fenestration present in the epithelial lining in liver. This allows efficient delivery of therapeutic agents to target sites in the body. Not only considered as delivery carriers for lipophilic and hydrophobic drugs, nanosized emulsions can also be viewed nowadays as adjuvants to enhance the potency of deoxyribonucleic acid (DNA) vaccine. For instance, Ott et al. [21] prepared a cationic o/w emulsion based on MF59 (commercially termed Fluad), a potent squalene in water and a cationic lipid, 1,2-dioleoyl-*sn*-glycero-3-trimethylammonium propane (DOTAP). It is shown that an interaction of cationic emulsion droplets with DNA and the formed DNA-adsorbed emulsion had a higher antibody response in mice in vivo while maintaining the cellular responses equivalent to that seen with naked DNA at the same doses. Another example of o/w emulsion-based adjuvants resulting from U.S. patent literature is the Ribi adjuvant system (RAS) [22–24]. Depending on the animal species used, RAS can be classified into two types: one for use in mice, termed monophosphoryl-lipid A + trehalose dicorynomycolate emulsion (MPL + TDM emulsion), and another for use in rabbits, goats, and larger animals, called monophosphoryl-lipid A + trehalose dicorynomycolate + cell wall skeleton emulsion (MPL + TDM + CWS emulsion). Strikingly, the MPL + TDM and MPL + TDM + CWS emulsions are prepared based on 2% oil (squalene)–Tween 80–water. These adjuvants are derived from bacterial and mycobacterial cell wall components that have been prepared to reduce the undesirable side effects of toxicity and allergenicity but still provide potent stimulus to the immune system. Another example is the syntex adjuvant formulation (SAF) that contains a preformed o/w emulsion stabilized by Tween 80 and Pluronic L121 [25].

Keeping in mind the potential of o/w nanosized emulsions, the purpose of this chapter is to classify the emulsions and provide a short overview on the preparation of the new- (second- and third-) generation emulsions followed by a description of the unique property of the third-generation emulsion and examples of selected excipients used for emulsion preparation. Given a specific interest especially on the parenteral route and then on the ocular topical route, o/w nanosized emulsions for both routes share a common platform on strict criteria concerning the maximum globule size and requirement of sterility in the final emulsions. Nasal and topical routes are also covered based on published research works with nanosized emulsions. It is emphasized that the chapter focuses only on preformed nanosized emulsions (having size distribution ranging between 50 and 1000 nm with a mean droplet size of about 250 nm), which should not be confused with self-microemulsifying drug delivery systems or preformed microemulsions that are transparent, thermodynamically stable dosage forms.

## 2 GENERATIONS OF OIL-IN-WATER NANOSIZED EMULSIONS

In this chapter, the o/w nanosized emulsions are classified into three generations (see Figure 1) based on their development to ultimately make a better colloidal

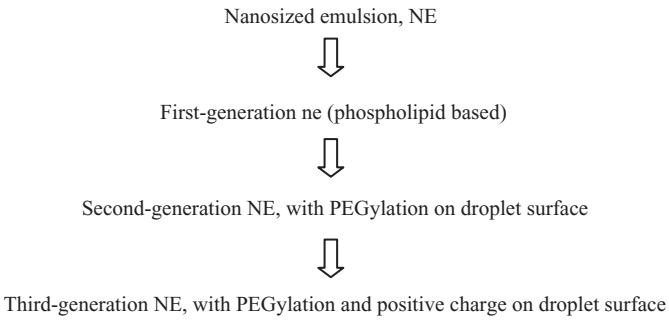


**FIGURE 1** Flow chart of three generations of emulsion.

carrier for a target site within the organs/parts of the body and eye, thus allowing site-specific drug delivery and/or enhanced drug absorption.

## 2.1 First-Generation Emulsion

To be healthy with a quality life style is every human's desire. According to documented Indian scriptures dating back to 5000 B.C., nutritional status has always been associated with health [26]. Because nutritional depletion due to either changes in the quality or amount of dietary fat intake or abnormalities in lipid metabolism results in immunosuppression and therefore host defense impairment, favoring increased infection and mortality rates.

Traditionally depletion in dietary fats in malnourished or hypercatabolic patients is compensated through intravenous feeding using a solution containing amino acids, glucose, electrolytes, and vitamins as well as nanosized emulsions. Structurally, an o/w-type emulsion is triglyceride (TG) droplets enveloped with a stabilizing superficial layer of phospholipids [27]. Emulsions for parenteral use are complex nutrient sources composed not only of fatty acids but also substances other than TG, such as phosphatidylcholine, glycerol, and α-tocopherol in variable amounts. The emulsions also had a complex inner structure and consisted of particles with different structures, namely, oil droplets covered by an emulsifier monolayer, oil droplets covered by emulsifier oligolayers, double-emulsion droplets, and possibly small unilamellar vesicles. Commercially available nanosized emulsions used as intravenous high-calorie nutrient fluids have particle size normally around 160–400 nm in diameter and, typically, their surfaces are normally negatively charged. Larger droplets can also be detected in commercially available emulsions [28]. Lutz et al. [29] reported that the mean diameter of particles in the 20% emulsions is larger than in the 10% emulsions.

The TG in nanosized emulsions may be presented structurally in long or medium chains respectively, named LCT and MCT. The mean diameter of particles in LCT emulsions is greater than that in MCT emulsions [29]. LCT contains fatty acid chains with 14, 16, 18, 20, and 22 carbon atoms and sometimes with double bonds. The number of double bonds present defines the fatty acids in LCT as saturated, monounsaturad, or polyunsaturated. If the first double bond is on carbon number 3, 6, or 9 from the methyl end of the carbon chain, then the fatty acid is *n*-3, *n*-6, or *n*-9, respectively. Purified soybean or safflower oil contains LCT with a high proportion

of *n*-6 polyunsaturated fatty acids whereas olive oil has LCT with *n*-9 monounsaturated fatty acids. Fish oil includes LCT with 20 or more carbon atoms where the first double bond is located between the third and fourth carbons from the methyl terminal of the fatty acids chain (omega-3 or *n*-3). On the other hand, MCT is derived from coconut oil and has saturated fatty acids with chains containing carbon atoms at the 6, 8, 10, or 12 positions. Both LCT and MCT either alone or MCT in combination with LCT are known for their long-term commercial acceptability in parenteral emulsions and are found in several U.S. Food and Drug Administration (FDA)–approved products. Also in Europe emulsion containing LCT/MCT enriched with fish oil became available for research. With MCT/LCT combinations in a specific ratio, nanosized emulsions appear to provide a more readily metabolizable source of energy [30]. However, LCT emulsion has been used in clinical practice for over 40 years. But for drug solubilization purpose, MCT is reported to be 100 times more soluble in water than LCT and thus to have an escalated solubilizing capability.

### 2.2 Second-Generation Emulsion

An easy and substantial association of lipophilic bioactive compound with the MCT or other vegetable oil–based emulsions, however, allows the emulsions to be used as vehicles/carriers for the formulation and delivery of drugs with a broad range of applications. These applications extend from enhanced solubilization or stabilization of the entrapped drug to sustained release and site-specific delivery. Hence the emulsions used for these applications are termed second-generation emulsions. Fittingly the o/w-type nanosized emulsions containing either positive or negative charge can be administered by almost all available routes, that is, topical, parenteral, oral, nasal, and even aerosolization into the lungs [31]. Despite differences in routes of administration, examples of commercially available emulsion-based formulations utilized for medical and nonmedical applications purposes are given in Table 1.

The lipid-induced enhancement in oral bioavailability of many drugs having poor water solubility is a well-known documented fact when the drugs are incorporated into emulsions [32, 33]. However, direct intravascularly or locally administered conventional first- and second-generation emulsions could be taken up rapidly by the circulating monocytes for clearance by reticuloendothelial cells (through organs such as the liver, spleen, and bone marrow) [34]. Furthermore, the extent of clearance is enhanced by the adsorption of opsonic plasma proteins onto emulsion surfaces. However, the oily hydrophobic particles of the emulsions can also be taken up by macrophages without the necessity of opsonization provided the oil phase is liberated from the emulsion through the destabilization process occurring inside the blood compartment immediately after emulsion administration intravascularly or locally. Although the core of o/w emulsions is indeed hydrophobic, the emulsion envelope is not. The exposure of the hydrophobic part to the aqueous medium will therefore destabilize the emulsion. Moreover, Sasaki et al. [35] have assumed that when the castor oil–based emulsions interact with the tears in the eye, the electrolytes in the tears elicit a physical emulsion instability resulting in some release of the oil. The electrolytes present in blood or cellular fluids can also cause a similar type of emulsion destabilization process, resulting in separation of the oil and water

**TABLE 1 Selected Marketed Medical and Nonmedical Emulsions**

| Parenteral Fat Emulsions (o/w Type) for Nutrition | | Registered Emulsions (o/w Type) Containing Drugs | | | |
|---|---|---|---|---|---|
| Product | Producer | Product | Drug | Producer | Application |
| Abbolipid/Liposyn | Abbott | Diazepam-Lipuro | Diazepam | Braun Melsungen | Intravenous |
| Intralipid | Pharmacia-Upjohn | Diprivan | Propofol | AstraZeneca | Intravenous |
| Lipofundin N or Endolipide | B. Braun | Etomidat-Lipuro | Etomidate | Braun Melsungen | Intravenous |
| Lipofundin MCT/LCT | — | Lipotalon (Limethason) | Dexamethasone palmitate | Merckle | Intra-arthr. |
| Medialipide/Vasolipid | B. Braun | Stesolid | Diazepam | Dumex | Intravenous |
| Medianut | B. Braun | Sandimmune | Cyclosporin A | Novartis | Oral |
| Lipovenos | Fresenius | Neoral | Cyclosporin A | Novartis | Oral |
| Ivelip/Salvilipid | Clintec/Baxter | Gengraf | Cyclosporin A | Abbott | Oral |
| Clinoleic | Clintec/Baxter | Norvir | Ritonavir | Abbott | Oral |
| Intralipos | Green Gross | Restasis | Cyclosporin A | Allergan | Ocular topical |
| Kabimix | Pharmacia-Upjohn | Refresh Endura | Drug free | Allergan | Ocular topical |
| Trivè 1000 | Baxter SA | Fluad (MF59) | Adjuvant | Chiron | Parenteral |

| Perfluorocarbon Emulsions (Fluorocarbon-in-Water Emulsions) | | | Selected Topical Formulations Based on o/w or w/o Emulsions | |
|---|---|---|---|---|
| Product | Producer | Application | Product | Producer |
| Fluosol DA | Green Gross, Osaka | Blood supplement or $O_2$ carrier | Daivonex cream and ointment | Laboratoire Leo |
| Imagent | Alliance Pharmaceutical | Contrast agent to image heart | Voltaren emulgel | Ciba-Geigy |
| Oxygent | Alliance Pharmaceutical | Blood supplement or $O_2$ carrier | EMLA cream | Astra, Swedan |

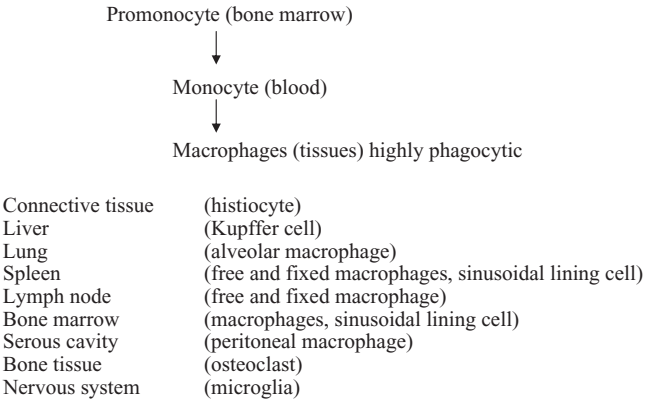


**FIGURE 2** Mononuclear phagocyte system.

phases from the parenterally administered emulsions. It is thus reasonable to say that the resultant oily hydrophobic particles of the emulsions would also be taken up by macrophages independent of an opsonization process. An opsonization process is the adsorption of protein entities capable of interacting with specific plasma membrane receptors on monocytes and various subsets of tissue macrophages (see Figure 2), thus promoting particle recognition by these cells. Classical examples of opsonic molecules include various subclasses of immunoglobulins [36, 37], complement proteins such as C1q and generated C3 fragments (C3b, iC3b) [38], apolipoproteins [36, 37], von Willebrand factor, thrombospondin, fibronectin [39], and mannose-binding protein. When given by other parentral routes, for example, intraperitoneally, subcutaneously, or intramuscularly, the majority of emulsion droplets enter the lymphatic system and eventually the blood circulation where particles behave as if given intravenously. Liver, spleen, and bone marrow uptake is significantly lower. Indeed, relative to the emulsion droplet size, lymph nodes take up a much greater (over 100-fold) proportion than any other reticuloendothelial system (RES) tissue.

There is increasing interest in developing injectable emulsions that are not cleared quickly from the circulation when they are designed to reach non-RES tissues in the vascular system or extravascular sites of action or to act as circulating drug reservoirs. Earlier approaches for making long-circulating emulsions concentrate mainly on changes in the oil phase of the emulsion such as MCT versus LCT [40], use of structured lipids (SLS) having medium-chain ($C_8$–$C_{10}$) fatty acids (SLM) and short-chain ($C_4$) fatty acids (SLS) [41], addition of sphingomyelin [42–45] and cholesterol [46] to the emulsion, and use of hydrogenated castor oil (HCO) with at least 20 oxyethylene units (HCO20) [47–52]. Using the further established formulation approaches by which the emulsion droplet surfaces could be altered might, however, be more realistic and even more useful for a wide array of drug-targeting purposes. Steric barrier or enhanced hydrophilicity effect exerted by a polyoxyethylene (POE) chain having surfactants when added as coemulsifier into the phospholipid-stabilized first-generation emulsion allows, to some extent, the passive/inverse drug targeting to the lung, kidneys, and areas of inflammation [53, 54]. Addition of POE-based surfactants into the otherwise amphipathic phospholipid-stabilized emulsion

is particularly effective against plasma protein adsorption onto emulsion surfaces because of the hydrophilicity and unique solution properties of POE-based surfactants, including minimal interfacial free energy with water, high aqueous solubility, high mobility, and large exclusion volume [54]. In addition, colloidal particles presenting hydrophilic surfaces with a low contact angle will be almost ignored by phagocytic cells [55], although emulsion particles are not supposed to be recognized as foreign by the body to some extent. Examples of POE chains containing surfactants employed so far in emulsions are Tween 80, Span 80, Brij, and Poloxamer 188 (commercially named Pluronic F-68 or Lutrol F-68). The effectiveness of these polymeric surfactant molecules to intercalate at the oil–water interface with strong bonding to the phospholipid molecules could also be checked/judged through an in vitro monolayer experiment [56].

In general, the modification of particulate carriers using amphipathic polyethylene glycol (PEG)–containing molecules results in a prolongation of their blood circulation time [57, 58]. A phosphatidylethanolamine derivative with polyethylene glycol (PEG–PE) is widely used to increase the plasma retention of particulate carriers such as liposomes [59–61], polystyrene microspheres [62], and nanospheres [63]. Therefore, similar to POE, the PEG–PE is also incorporated as a coemulsifier into emulsions (termed PEGylated emulsion) to augment its circulation half-life [64]. Liu and Liu [53] studied the biodistribution of emulsion particles coated with phosphatidylethanolamine derivatives with three different molecular weight PEGs (MW 1000, 2000, and 5000). Among them, PEG-2000 was able to prolong the circulation time of emulsion probably due to the increased hydrophilicity of the droplet surface and/or the formation of a steric barrier. However, Tirosh et al. [65] assumed, while characterizing the PEG-2000-grafted liposome by differential scanning calorimetry, densitometry, and ultrasound velocity and absorption measurements, that the steric stabilization is much more than increasing hydrophilicity. In addition, PEG-containing compounds also decrease the lipolysis of emulsion particles [47] and prevent the uptake by the mononuclear phagocytes [66].

A dipalmitoyl phosphatidylcholine (DPPC)–stabilized emulsion was prepared by Lundberg et al. [67] and the effect of addition of PEG–PE, polysorbate 80, or Pluronic F-68 on the metabolism of DPPC-stabilized emulsion was studied. Two different radioactive markers were used to investigate the fate of emulsion particles following injection into the tail vein of female BALB/c inbred mice. While $^{14}$C- triolein (TO) is susceptible to the action of lipoprotein lipase (LPL), $^{3}$H-cholesteryl oleate ether (CO ether) is not. Hence the removal of $^{14}$C-TO represents the triglyceride metabolism, whereas the other one is a core marker to represent whole particle removal by RES organ uptake. The emulsions with DPPC as sole emulsifier were rapidly cleared from the blood with only 10–11% of CO or TO left in circulation after 1 h. However, addition of PEG–PE gave a prolonged clearance rate, especially during first 3 h. A further addition of cosurfactant polysorbate 80 or Pluronic F-68 resulted in a marked extension of the circulation lifetime during the first 6 h. The notable effects of polysorbate 80 and Pluronic F-68 can apparently be attributed mainly to the decrease in droplet size, although an additional influence due to the increased hydrophilicity may not be ruled out.

The in vivo disposition of emulsions administered as nutrients (surface-unmodified first-generation emulsion) as well as administered as drug carriers (surface-modified second-generation emulsion) would depend on the particle prop-

erties, such as the size [68–71], zeta potential (see Sections 2.3 and 5), and compositions of phospholipids and oil phase (see the above paragraphs), which may vary among different products and the batches of each product. The size of particulate carriers such as liposomes is known to influence both the phagocytic uptake by the mononuclear phagocyte system (MPS) [68–70] and the binding of apolipoprotein (apo) to emulsions [71]. Furthermore, the particle size is a major determinant of the transfer to extravascular spaces from the blood compartment. The capillaries of the vascular system can be classified into three categories: continuous, fenestrated, and discontinuous (sinusoidal) [72]. Particulate carriers including nanosized emulsions are considered to pass through capillaries and reach extravascular cells only in organs having discontinuous capillaries such as liver, spleen, and bone marrow. In such tissues, the extravasation of particles should be regulated by their size since the largest pores in the capillary endothelium is reported to be about 100 nm [73]. In addition, tumor capillaries have unique characteristics in their structures and functions in comparison with normal tissues such as muscle [74, 75], which results in the enhanced distribution of particulate carriers to tumor tissues [76–78]. The distribution of emulsions within a tumor tissue was also regulated by the size of particulate carriers [79]. Obviously, because of the submicrometer size range (175–400 nm in diameter) of the emulsions, the more they circulate, the greater their chance of reaching respective targets. More specifically, growing solid tumors as well as regions of infection and inflammation have capillaries with increased permeability as a result of the disease process (e.g., tumor angiogenesis [74]). Pore diameters in these capillaries can range from 100 to 800 nm. Thus, drug-containing emulsion particles are small enough to extravasate from the blood into the tumor interstitial space through these pores [80]. Normal tissues, by and large, contain capillaries with tight junctions that are impermeable to emulsions and other particles of this diameter. This differential accumulation of emulsion-laden drug in tumor tissues relative to normal cells is the basis for the increased tumor specificity for the emulsion-laden drug relative to free (nonemulsion) drug. In addition, tumors lack lymphatic drainage and therefore there is low clearance of the extravasated emulsion from tumors. Thus, long-circulating lipid carriers, such as POE/PEG-coated nanosized emulsions, tend to accumulate in tumors as a result of increased microvascular permeability and defective lymphatic drainage, a process also referred to as the enhanced permeability and retention (EPR) effect [81]. Table 2 lists various formulation factors affecting the metabolism as lipoproteins, the recognition by the MPS, and the elimination from the blood circulation of both second- and third-generation nanosized emulsions after parenteral administration.

On the other hand, essential requirements of this "active" targeting approach include identification of recognition features (receptors) on the surface of the target and the corresponding molecules (ligands) that can recognize the surface. Indeed, emulsions with appropriate ligands anchored on their surface must be able to access the target, bind to its receptors, and, if needed, enter it. Furthermore, in order to bring the colloidal carrier closer to otherwise inaccessible pathological target tissues, homing devices/ligands such as antibodies and cell recognition proteins are usually linked somehow onto the particle surfaces. Various methods have been employed to couple ligands to the surface of the nanosized lipidic and polymeric carriers with reactive groups. These can be divided into covalent and noncovalent couplings. Noncovalent binding by simple physical association of targeting ligands to the

**TABLE 2 Formulation Factors Affecting Metabolism as Lipoproteins, Recognition by Mononuclear Phagocyte System (MPS), and Elimination from Blood Circulation of Second- and Third-Generation Nanosized Emulsions after Parenteral Administration**

| | Metabolism as Lipoproteins | | Recognition by MPS | | Elimination from Blood Circulation | |
|---|---|---|---|---|---|---|
| Factor | Poor | Extensive | Poor | Extensive | Slow | Rapid |
| Particle size | Large | Small | Small | Large | Small | Large |
| Emulsifier | DPPC | EYPC | DPPC | DSPC | DPPC | EYPC |
| | DSPC | — | SM | — | SM | DSPC |
| | SM | — | — | — | — | — |
| Coemulsifier | Poloxamers | — | Poloxamers | — | Poloxamers | — |
| | HCO-60 | — | HCO-60 | — | HCO-60 | — |
| | PEG–PE | — | PEG–PE | — | PEG–PE | — |
| | Polysorbates | — | Polysorbates | — | Polysorbates | — |
| | Solutol | — | Solutol | — | Solutol | — |
| Cationic lipid | SA/OA | — | SA/OA | — | SA/OA | — |
| Oil phase | LCT | MCT | — | — | LCT | MCT |
| | — | — | — | — | SLS | SLM |

*Note:* DPPC, dipalmitoylphosphatidylcholine; DSPC, distearoylphosphatidylcholine; SM, sphingomyelin; EYPC, egg yolk phosphatidylcholine; HCO-60, polyoxyethylene-(60)-hydrogenated castor oil; PEG–PE, phosphotidylethanolamine derivative with polyethylene glycol; SA, stearylamine; OA, oleylamine; LCT, long-chain triglyceride; MCT, medium-chain triglyceride; SLS, structured lipid with short-chain fatty acids, $C_8$–$C_{10}$; SLM, structured lipid with medium-chain fatty acids, $C_4$.

nanocarrier surface has the advantage of eliminating the use of rigorous, destructive reaction agents. Common covalent coupling methods involve formation of a disulfide bond, cross-linking between two primary amines, reaction between a carboxylic acid group and primary amine, reaction between maleimide and thiol, reaction between hydrazide and aldehyde, and reaction between a primary amine and free aldehyde [82]. For antibody-conjugated second-generation anionic emulsions, the reaction of the carboxyl derivative of the coemulsifier molecule with free amine groups of the antibody and disulfide bond formation between coemulsifier derivative and reduced antibody were the two reported conjugation techniques so far [83–85]. However, by the formation of a thio-ether bond between the free maleimide reactive group already localized at the o/w interface of the emulsion oil droplets and a reduced monoclonal antibody, the antibody-tethered cationic emulsion was developed for active targeting to tumor cells [86]. It should be added that the cationic emulsion investigated for tumor-targeting purpose belongs to the third-generation emulsion category (Section 2.3).

Apart from non-RES-related disease treatment through target-specific ligand conjugation, the second-generation emulsions may also be useful for RES-related disease treatment. Certain lipoprotein or polysaccharide moiety inclusion into the emulsions would help to achieve this concept. In general, uptake of small colloidal drug carriers by the phagocytotic mononuclear cells of RES in the liver can be exploited to improve the treatment of parasitic, fungal, viral and bacterial diseases such as, for example, leishmaniasis, acquired immunodeficiency syndrome (AIDS), and hepatitis B. The approach to use emulsions as a drug carrier against microbial

diseases is superior to free antimicrobial agents in terms of both distribution to the relevant intracellular sites and treating disseminated disease states effectively. As already discussed, conventional emulsion particles are capable of localizing in liver and spleen, where many pathogenic microorganisms reside.

Rensen et al. [87] demonstrated the active/selective liver targeting of an antiviral prodrug (nucleoside analogue, iododeoxyuridine) incorporated in an emulsion complexed with ligands such as recombinant apolipoprotein E (apoE) using the Wistar rat as animal model because its apoE–receptor system is comparable to that of humans [88]. Whereas the parent drug did not show any affinity for emulsion due to its hydrophilic property, derivatization with hydrophobic anchors allowed incorporation of at least 130 prodrug molecules per emulsion particle without imparting any effect on the emulsion structure and apoE association to emulsion droplets. The authors did not describe where the 130 prodrug molecules reside in the emulsion and what is the emulsion/medium partition coefficient of the prodrug. The prodrug molecules might have reasonably higher solubility in the oil or o/w interface of the emulsion possibly due to a high partition coefficient value. Plausibly, this high partition value for prodrug molecules will determine the kinetic parameter $k_{\text{off}}$ (desorption rate of an emulsion component from the assembly), as suggested by Barenholz and Cohen [89] for liposomal technology. Furthermore, without being bound by theory, the apoE component helps to disguise the emulsion particles so that the body does not immediately recognize it as foreign but may allow the body to perceive it as native chylomicrons or very low density lipoproteins (VLDL). The small size and the approximately spherical shape allow the emulsion particles to exhibit similar physicochemical properties to native chylomicrons or VLDL (hydrolyzed by LPL) whereas the incorporated prodrug remained associated with the emulsion remnant particles following injection into the blood circulation of the rat [87]. Because the carrier particles are not recognized as foreign, the systemic circulation of the drug increases, thus increasing the likelihood of drug delivery to the target tissues (up to 700 n$M$ drug concentration in liver parenchymal cells). Additionally, the clearance rate of the drug decreases, thereby reducing the likelihood of toxic effects of the drug on clearance tissues since accumulation of the drug in another part of the clearance tissues is reduced. Thus, specific organs may be targeted by using nanosized emulsion particles as described above due to target cells comprising high levels of specific receptors, for example, but not limited to apoE receptors.

To address this issue, the saccharide moieties of glycolipids and glycoproteins on the cell surface are considered to play an important role in various intercellular recognition processes. For instance, Iwamoto et al. [90] investigated the influence of coating the oil droplets in emulsion with cell-specific cholesterol bearing polysaccharide, such as mannan, amylopectin, or pullulan, on the target ability of those formulations. They observed a higher accumulation of mannan-coated emulsion in the lung in guinea pigs. Thus selective drug targeting through emulsion-bearing ligands not only leads to an improved drug effectiveness and a reduction in adverse reactions but also offers the possibility of applying highly potent drugs. Hence, the composition of the emulsion plays an important role concerning intercellular cell recognition processes and, indeed, cell recognizability is also being improved by incorporation of apoproteins or galactoproteins onto the emulsion particles to enhance their specificity [91].

Overall, although second-generation emulsion is usually used as a means of administering aqueous insoluble drugs by dissolution of the drugs within the oil phase of the emulsion, employing surface modification/PEGylation by the attachment of targeting ligands (apoE, polysaccharide, and antibody) onto the droplet surface of emulsions may be useful for both passive and active drug-targeting purposes. Thus receptor-mediated drug targeting using ligands attached to emulsions seems to hold a promising future to the achievement of cell-specific delivery of multiple classes of therapeutic cargoes, and this approach will certainly make a major contribution in treating many life-threatening diseases with a minimum of systemic side effects.

### 2.3 Third-Generation Emulsion

In order to increase cellular uptake, a cationization strategy is applied particularly on the surfaces of nonviral, colloidal carrier systems such as liposomes, nano- and microparticulates, and nanocapsules [92]. To make the surface of these lipidic and polymeric carrier systems a cationic property, some cationic lipids/polymers are usually added into these systems during/after preparation. But, adding only the cationic substances in phospholipids-stabilized first-generation emulsions does not help to obtain a physically stabilized emulsion for a prolonged storage period. However, using different cationic lipids as emulsifier and additional helper lipids as coemulsifier, for example, DOTAP, 1,2-dioleoyl-*sn*-glycero-3-phosphoethanolamine (DOPE), and 1-palmitoyl-2-oleoyl-*sn*-glycero-3-phosphoethanolamine-*N*-[poly(ethylene glycol)$_{2000}$] (PEG$_{2000}$PE), reports are available to prepare emulsions with positive charges on their droplet surfaces [93, 94]. Alternatively, inclusion of cation-forming substances such as lipids (stearyl or oleyl chain having primary amines) [95, 96], polymers (chitosan) [97, 98], and surfactants (cetyltrimethylammonium bromide) [99] during the preparation of second-generation emulsion allows the formation of a stabilized system with positive charges over on it. Further, the positive charge caused by stearylamine was also confirmed by a selective adsorption of thiocyanate. Its adsorption was correlated with increasing stearylamine concentration [95]. So, nanosized emulsion consisting of complex emulsifiers, that is, phospholipid–polyoxyethylene surfactant-cationized primary amine or a polymer combination, can conveniently be termed third-generation emulsion.

The extemporaneous addition of the solid drug or drug previously solubilized in another solvent or oil to the preformed first- and second-generation emulsions is not a favored approach as it might compromise the integrity of the emulsion. Since therapeutic DNA and single-stranded oligonucleotides or small interfering ribonucleic acid (siRNA) are water soluble due to their polyanionic character, the aqueous solution of these compounds need to be added directly to the preformed third-generation emulsion in order to interact electrostatically with the cationic emulsion droplets and thus associate/link superficially at the oil–water interface of the emulsion [100, 101]. During in vivo conditions when administered via parenteral, nasal, and ocular routes, the release of the DNA and oligonucleotides from the associated emulsion droplet surfaces should therefore initially be dependent solely on the affinity between the physiological anions of the biological fluid and the cationic surface of the emulsion droplets. For instance, the mono- and divalent anions con-

taining biological fluid available in the parenteral route is plasma and in ocular topical route is tear fluid, aqueous humor, and vitreous humor. Moreover, these biofluids contain a multitude of macromolecules and nucleases. There is a possibility that endogenous negatively charged biofluid components could dissociate the DNA and oligonucleotides from cationic emulsion. It is noteworthy to conduct, during the preformulation development stages, an in vitro release study for therapeutic DNA and oligonucleotide-containing emulsion in these biological fluids, and this type of study could be considered an indicator of the strength of the interaction that occurred between DNA or oligonucleotide and the emulsion particles. However, it is interesting to see what could happen when the third-generation emulsion is applied to in vitro cell culture models in the presence of serum. The serum stability of the emulsion–DNA complex was reported [102, 103]. Further interesting investigations using third-generation emulsions in gene delivery purposes are briefly summarized in a review article [104].

### 2.4 Unique Property of Third-Generation Emulsion

To enhance the drug-targeting efficacy of colloidal carriers such as nanospheres and liposomes, a PEGylation/cationization strategy is traditionally made over the surface of these carriers. While surface PEGylated colloidal carriers exhibit a prolonged plasma residence time through an escaping tendency from RES uptake following parenteral administration, surface-cationized colloidal carriers can facilitate the penetration of therapeutic agents into the cell surface possibly via an endocytotic mechanism. These two facts are proved in both liposomes and nanospheres when they possess separately the cationic and PEGylatic surface modifications on them. However, a cationic emulsion colloidal carrier system developed in Simon Benita's laboratory at Hebrew University of Jerusalem, Israel, differs significantly in such a way that it holds a combination of cationic and PEGylatic surface properties on it. Benita's group have prepared a novel cationic emulsion vehicle using a combination of emulsifiers consisting of Lipoid E 80, Poloxamer 188, and stearylamine and have found the formulation suitable for parenteral use, ocular application, nasal drug delivery, and topical delivery [105].

It has been reported in an ocular pharmacokinetic study of cyclosporin A incorporated in deoxycholic acid–based anionic and stearylamine-based cationic emulsions in rabbit that, when compared to anionic emulsion, the cationic emulsion showed a significant drug reservoir effect of more than 8 h in corneal and conjunctival tissues of the rabbit eye following topical application [106]. Since cornea and conjunctiva are of anionic nature at physiological pH [107], the cationic emulsion would interact with these tissues electrostatically to implicate the observed cyclosporin A reservoir effect. This hypothesis is supported, in principle, by an ex vivo study which showed that cationic emulsion carrier exhibited better wettability properties on rabbit cornea than either saline or anionic emulsion carrier [108].

Studies [109, 110] have shown that small changes in physical properties of emulsions can influence the elimination rate of these formulations from the blood. Indeed, an organ distribution study of stearylamine-based cationic or deoxycholic acid–based anionic nanosized emulsions and Intralipid, a well-known commercial anionic emulsion, containing $^{14}$C-CO was carried out following injection into the tail vein of male BALB/c mice (20–26 g) at a volume of 5 mL/kg [111, 112]. Since CO

(cholesteryl oleate) is one of the most lipophilic compounds used in biopharmacy and is not prone to degradation in the body (which remains within particles even after lipolysis of emulsion), its in vivo behavior can be regarded as reflecting that of the injected nanosized emulsion in the early phase [42, 113]. Following intravenous administration of the various emulsions having $^{14}$C-CO to BALB/c mice, the $^{14}$C-CO was found to accumulate in organs such as lung and liver. Furthermore, it was observed that the concentration of $^{14}$C-CO in the lung decreased but was again elevated over time for both the developed cationic and anionic emulsion formulations, with a concomitant decrease in the concentration of the radiolabeled compound in the liver. However, within the various emulsion distribution patterns observed in liver, a lower $^{14}$C-CO concentration was observed for stearylamine-based cationic emulsion when compared to Intralipid while for deoxycholic acid–based anionic emulsion the observed concentration of $^{14}$C-CO was relatively very low when compared to cationic emulsion and Intralipid. In addition, in comparison to both anionic emulsions, the stearylamine-based cationic emulsion elucidated a much longer retention time of $^{14}$C-CO in the plasma, clearly indicating a long circulating half-life for cationic emulsion in the blood. Thus, the cationic nanosized emulsion can be considered a stealth long-circulating emulsion.

The above two studies clearly described the unique characteristics of third-generation emulsion in enhancing ocular drug bioavailability; on the other hand, the same emulsion has the property of circulating for a longer time in blood following parenteral administration. Excess positive charge at the oil–water interface in conjunction with the projection of highly hydrophilic POE chain (due to the presence of Poloxamer 188) toward the aqueous phase of the o/w-type nanosized emulsion is the main reason behind the emulsion attaining its unique property, which is absent in first- and second-generation emulsions. However, a better understanding of the structure of the third-generation emulsion in terms of forces involved in its formation and stabilization must ultimately be obtained in an effort to provide a clearly understood physical basis for the uniqueness in its biological efficacy following parenteral and ocular administration.

It should be added that the use of stearylamine in intravenous administered emulsion might be problematic. Stearylamine is a single-chain amphiphile having relatively high critical micellar concentration, although the concentration used in the studied emulsion is much higher than the critical micellar concentration. Therefore, due to the dilution in plasma as well as plasma lipoproteins and blood cells, there is a high probability that the emulsion will lose its stearylamine almost instantaneously. To substantiate indirectly this issue, Klang et al. [114] showed the lack of potential induced toxicity of stearylamine-based cationic emulsion in animal models in vivo and Korner et al. [115] investigated the surface properties of mixed phospholipid–stearylamine monolayers and their interaction with a nonionic surfactant (poloxamer) in vitro. Despite the presence of the stearylamine, which may be suspected of being an irritant in pure form, in the emulsifier combination, the hourly instillation of stearylamine-based cationic emulsion vehicle into rabbit eye was well tolerated without any evidence of any toxic or inflammatory response to the ocular surface during the 5 days of the study (40 single-drop instillations between 8 AM and 4 PM each day) [114]. Following 0.2-, 0.4-, and 0.6-mL single-bolus injections of the same emulsion vehicle, representing a huge single administered dose of 30 mL/kg, no animal deaths were noted over a period of 30 days, apparently indicat-

ing the absence of marked acute toxicity [114]. Furthermore, the same stearylamine-based cationic emulsion vehicle did not cause acute neurotoxicity in rats when a continuous intravenous infusion (3.3 mL) for 2 h at a rate of 27.4 μL/min was administered through the jugular vein [114]. An another study from Benita's laboratory suggests that long-term subchronic toxicity examination of the rabbit eye (healthy) following thrice-daily single-drop topical instillation of the stearylamine-based emulsion elicited an almost similar nonirritating effect to eye tissues in comparison to the thrice-daily single-drop topical instillation of the normal saline–treated control rabbit eyes (unpublished data). Thus, overall results clearly indicated that the stearylamine was strongly bound at a molecular level to the mixed interfacial film formed by Lipoid E 80 and poloxamer 188 at the oil–water interface system [115]. Such an intercalation between the emulsifiers is responsible for emulsified oil droplet stability and, in fact, prevented the stearylamine from leaking and exerting any intrinsic possible local or systemic adverse effects in model animals.

## 3 PREPARATION METHODS FOR DRUG-FREE/LOADED OIL-IN-WATER NANOSIZED EMULSIONS

To get a better idea of how to formulate the nanosized emulsion delivery systems suitable for parenteral, ocular, percutaneous, and nasal uses, the reader is referred to more detailed descriptions of methods of nanosized emulsion preparation [6, 116]. A hot-stage high-pressure homogenization technique or combined emulsification technique (de novo production) is frequently employed in order to prepare nanosized emulsions with desired stability even after subjection to autoclave sterilization. Therefore, the steps involved in this technique in making blank anionic and cationic emulsions were arranged in the following order:

1. Weigh the oil- and water-soluble ingredients in separate beakers.
2. Heat both oil and water phases separately to 70°C.
3. Add the oil phase to the water phase and continue the heating up to 80°C with constant stirring to form a coarse emulsion.
4. Mix at high shear to make a fine emulsion.
5. Cool the fine emulsion formed in ice bath.
6. Homogenize the fine emulsion.
7. Cool the homogenized emulsion in ice bath.
8. Filter the emulsion using a 0.5 μm membrane filter.
9. Adjust the emulsion to 7 using 0.1 *N* hydrochloric acid or 0.1 *N* sodium hydroxide solution.
10. Pass nitrogen gas into the vials containing the emulsion.
11. Sterilize the emulsion using an autoclave.

The traditional droplet size–reducing steps involved during the preparation include constant mild stirring using a magnetic stirrer when initially mixing oil and water phases, rapid Polytron mixing at high speed, and final droplet size homogenization using a two-stage homogenizer valve assembly. The initial heating is vital for the

**TABLE 3 Typical Formula to Make o/w Anionic and Cationic Nanosized Emulsions**

| Oil Phase | Water Phase |
|---|---|
| Natural/semisynthetic oils | Poloxamer 188 |
| Phospholipid mixture | Glycerol |
| Stearylamine/oleylamine[a] | Double-distilled water |
| Deoxycholic acid/oleic acid[b] | |
| Vitamin E | |

[a]Necessary ingredient for cationic emulsions.
[b]Necessary ingredient for anionic emulsions.

effective solubilization of the respective oil and water phase components in their corresponding phases. Mixing the two phases with constant mild stirring and subsequently raising the temperature to 85°C are needed to form an initial coarse emulsion and to localize the surfactant molecules for better adsorption at the oil–water interface, respectively. A typical formula to make anionic and cationic nanosized emulsions is given in Table 3.

There are three different approaches to incorporate lipophilic drugs into the oil phase or at the o/w interface of the nanosized emulsions, namely, extemporaneous drug addition, de novo emulsion preparation, and an interfacial incorporation approach, which includes the recently developed SolEmul technology [117]. In principle, the lipophilic drug molecules should however be incorporated by a de novo process. Thus, the drug is initially solubilized or dispersed together with an emulsifier in suitable single-oil or oil mixture by means of slight heating. The water phase containing the osmotic agent with or without an additional emulsifier is also heated and mixed with the oil phase by means of high-speed mixers. Further homogenization takes place to obtain the needed small droplet size range of the emulsion. A terminal sterilization by filtration, steam, or autoclave then follows. The emulsion thus formed contains most of the drug molecules within its oil phase. This is a generally accepted and standard method to prepare lipophilic drug–loaded nanosized emulsions for parenteral, ocular, percutaneous, and nasal uses, as illustrated in Figure 3. This process is normally carried out under aseptic conditions and nitrogen atmosphere to prevent both contamination and potential oxidation of sensitive excipients.

## 4 EXCIPIENT INCLUSION: OIL-IN-WATER NANOSIZED EMULSIONS

In general, nanosized emulsions should be formulated with compatible vehicles and additives. The components of the internal and external phases of emulsion should be chosen to confer enhanced solubility and/or stability to the incorporated biologically active lipophilic drug. In addition, it should also be designed to influence biofate or therapeutic index of the incorporated drug following administration via parenteral, ocular, percutaneous, and nasal routes. In this section, general considerations concerning excipient selection and optimum concentrations are comprehen-

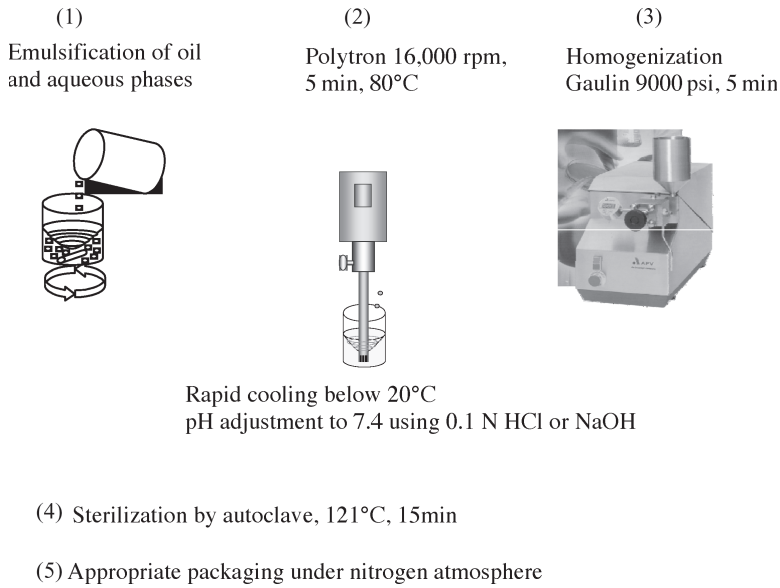


**FIGURE 3** Preparation of o/w nanosized emulsion (de novo method).

sively presented mainly in their relation to the oil phase, the aqueous phase, and the emulsifiers.

Prior to the formulation design of the emulsions, data are needed concerning the drug solubility in the oil vehicle. Additionally, prerequisite information is needed about compatibility of the oil vehicle with other formulation additives and the established ocular/skin tissues–oil vehicle matching before the dosage form can be prepared. Table 4 lists the common emulsion excipients and the oils suitable for dissolving or dispersing lipophilic drugs of ocular/parenteral interests. Since oils are triglycerides, care must be taken to minimize or eliminate oxidation. Therefore, antioxidants such as α-tocopherol (0.001–0.002% w/w) should be included in a typical emulsion formulation for medical applications. The final oil-phase concentration in emulsions meant for ocular use is now widely accepted to be at or below 5% w/w taking into account that the emulsion must be kept in a low-viscosity range of between 2 and 3 centipoises, which also is the optimal viscosity for ocular preparations [118]. However, for all other medical uses, the amount of oil may be varied but generally is within 5–20% w/w. Sometimes, a mixture of oils rather than a single oil is employed since drug solubilization in the oil phase is a prerequisite to exploiting the emulsion advantages. Jumaa and Müller [98, 119] reported the effect of mixing castor oil with MCT on the viscosity of castor oil. The oil combination at the ratio of 1:1 (w/w) led to a decrease in the viscosity of castor oil and simultaneously to a decrease in the interfacial tension of the oil phase. This was related to the free fatty acids contained in castor oil, which can act as a coemulsifier resulting in lower interfacial tension and, simultaneously, in a more stable formulation in comparison with the other oil phases. In addition to the digestible oils from the family of triglycerides, including soybean oil, sesame seed oil, cottonseed oil, and safflower oil, which are routinely used for making medical emulsions, alternative biocompatible

**TABLE 4 Excipients Used for Formulation of o/w Nanosized Emulsions**

| Oils | Emulsifiers | Cationic Lipids and Polysaccharide | Miscellaneous |
|---|---|---|---|
| Sesame oil | Cholesterol | Stearylamine | α-Tocopherol |
| Castor oil | Phospholipids (Lipoid) | Oleylamine | Glycerin |
| Soya oil | Polysorbate 80 and 20 (Tween 80 and 20) | Chitosan | Xylitol |
| Paraffin oil | Transcutol P | | Sorbitol |
| Paraffin light | Cremophor RH | | Thiomersal |
| Lanolin | Poloxamer 407 | | EDTA |
| Vaseline | Poloxamer 188 | | Methyl paraben |
| Corn oil | Miranol $C_2M$ and MHT | | Propyl paraben |
| Glycerin monostearate | Tyloxapol TPGS | | |
| Medium-chain monoglycerides | | | |
| Medium-chain triglycerides | | | |
| Squalene | | | |
| Vitamin E | | | |

*Note:* TPGS, α-tocopheryl-polyethylene glycol-1000-succinate; EDTA, ethylenediamine tetraacetic acid.

oils such as α-tocopherol and/or other tocols were also investigated for drug delivery purposes via o/w emulsions [120, 121]. But the emulsions formed from tocols are often considered as microemulsion systems with few exceptions being the nanosized emulsions [122, 123].

Unlike spontaneously forming thermodynamically stable microemulsion systems that require a high surfactant concentration (20% and higher), the kinetically stable nanosized emulsions can be prepared by using relatively lower surfactant concentrations. For example, a 20% o/w nanosized emulsion may only require a surfactant concentration of 5–10%. Traditionally, lecithins or phospholipids are the emulsifiers of choice to produce nanosized emulsions. However, emulsifiers of this kind are not suitable to produce submicrometer–sized emulsion droplets or to withstand the heat during steam sterilization. Therefore, additional emulsifiers preferably dissolved in the aqueous phase are usually included in the emulsion composition. A typical example of the aqueous soluble emulsifiers is nonionic surfactants (e.g., Tween 20) after taking into consideration their nonirritant nature when compared to ionic surfactants. The nonionic block copolymer of polyoxyethylene–polyoxypropylene, Pluronics F68 (Poloxamer 188), is included to stabilize the emulsion through strong steric repulsion. However, amphoteric surfactants Miranol MHT (lauroamphodiacetate and sodium tridecethsulfate) and Miranol $C_2M$ (cocoamphodiacetate) were also used in earlier ophthalmic emulsions [124]. It should be added that commercially available cyclosporin A–loaded anionic emulsion (Restasis) contains only polysorbate 80 and carbomer 1342 at alkali pH to stabilize the drug-loaded anionic emulsion. To prepare a cationic emulsion, cationic lipids (stearyl and oleylamine) or polysaccharides (chitosan) are added to the formulation. Strikingly, if chitosan is a choice of cation producing polysaccharide emulsifier molecules, there is no need to add amphoteric or nonionic surfactants to the phospholipid or lecithin-stabilized

emulsion [125]. Conversely, a cationic emulsion based on an association of poloxamer 188 and chitosan without lecithin was prepared and also showed adequate physicochemical properties regarding stability and charge effects [97, 98]. Oil-in-water emulsion compositions based on a tocopherol (or a tocopherol derivative) as the disperse phase have been described in a patent granted to Dumex [126]. The emulsion is intended for use with compounds that are sparingly soluble in water. Interestingly, the emulsifying agent used to make tocol-based emulsions are restricted to vitamin E TPGS.

Additives other than antioxidants such as preservatives (e.g., benzalkonium chloride, chlorocresol, parabens) are regularly included in emulsions to prevent microbial spoilage of multidose medical emulsions. α-Tocopherol is a good example of an antioxidant used to obtain a desirable stabilized emulsion under prolonged storage conditions. The presence of components of natural origin such as lecithin or oils with high calorific potential renders the emulsion a good medium to promote microbial growth when it is packed in multidose containers. Pharmaceutical products when distributed into multidose containers, especially for parenteral and ocular administrations, should be properly preserved against microbial contamination and proliferation during storage in normal conditions and proper use. Incorporation of preservatives in single-dose vials is also a common procedure if filtration is used as a sterilization method. Sznitowska et al. [127] studied the physicochemical compatibility between the lecithin-stabilized emulsion and 12 antimicrobial agents over two years of storage at room temperature. Preliminary physicochemical screening results indicate that addition of chlorocresol, phenol, benzyl alcohol, thiomersal, chlorhexidine gluconate, and bronopol should be avoided due to the occurrence of an unfavorable pH change followed by coalescence of lecithin-stabilized droplets of the emulsion. Furthermore, the efficacy of antimicrobial preservation was assessed using the challenge test according to the method described by the European Pharmacopoeia. Despite good physicochemical compatibility, neither parabens nor benzalkonium chloride showed satisfying antibacterial efficacy in the emulsion against the tested microorganisms and consequently did not pass the test. Therefore, higher concentrations of antimicrobial agents or their combination may be required for efficient preservation of the lecithin-stabilized emulsion probably because of unfavorable phase partitioning of the added antimicrobials within the different internal structures of the emulsion. It is interesting to note that benzalkonium chloride, a highly aqueous soluble drug, did not pass the standard challenge test even when incorporated in a cationic emulsion, particularly the third-generation category (unpublished data). This finding clearly indicates that the possible electrostatic attraction between the negatively charged lipid moieties of the mixed emulsifying film formed around the anionic emulsified oil droplets [127] and the quaternary cationic ammonium groups of the preservative is not the plausible cause for the reduced activity of the benzalkonium chloride. Thus, the possible intercalation of this surfactant in either the cationic or anionic interfacial mixed emulsifying film is likely to occur, preventing benzalkonium chloride from eliciting its adequate preservative action. Overall, it is preferable to formulate nanosized emulsions devoid of preservative agents and fill it in sterile single-dose packaging units to prevent potential contamination due to repeated use of multidose packaging. It should be pointed out that the two available ocular emulsion products (Refresh Endura and Restasis, Allergan, Irvine, CA) on the market are preservative free and packed in

single-use vials only. Currently there is no commercial parenteral emulsion which contains preservatives and research concerning the problem of preservation of nanosized emulsion is very limited [128–131].

## 5 MEDICAL APPLICATIONS OF OIL-IN-WATER NANOSIZED EMULSIONS

It has been shown in a number of studies that the incorporation of drug in o/w nanosized emulsions significantly increased the absorption of the drug when compared with the equivalent aqueous solution administered orally [132–135]. However, the use of emulsions for oral application is limited since other attractive alternatives, such as self-emulsifying oil delivery systems, which are much less sensitive and easy to manufacture, are available [136, 137]. Thus the potential of nanosized emulsions after administration with parenteral and traditional nonparenteral topical routes such as ocular, percutaneous, and nasal is covered in this section.

### 5.1 Parenteral Routes

The o/w nanosized emulsion formulations of lipophilic drugs, such as propofol, etomidate, dexamethasone palmitate, and diazepam, were already developed and marketed (Table 1). Furthermore, various research groups across the world are currently undertaking projects to exploit the potential of o/w emulsions for parenteral delivery of a myriad of investigational drugs as well as other lipophilic drugs by receptor-mediated targeting to cancer cells. The important technical and clinical points to keep in mind before the use of the emulsion systems for this kind of work are given below.

It has to be clear that, once diluted and injected (or administered in ocular and other routes), the emulsion stability and fate are determined by three measurable parameters. The first is the partition coefficient of each emulsion component (including added drugs and agents) between the emulsion assembly and the medium. To some extent this partition coefficient is related to oil–water and/or octanol–water partition coefficients. For example, it was well demonstrated that per component of which $\log P$ is lower than 8, the stability upon intravenous (IV) injection is questionable [42, 138]. The other two parameters are $k_{\text{off}}$, a kinetic parameter which describes the desorption rate of an emulsion component from the assembly, and $k_c$, the rate of clearance of the emulsion from the site of administration. This approach is useful to decide if and what application a drug delivery system will have a chance to perform well [89].

Stability in plasma is an important requirement for IV emulsions as flocculated droplets may result in lung embolism. It was found that tocol-based emulsions stabilized by sodium deoxycholate/lecithins flocculated strongly when mixed with mouse, rat, and sheep plasma and serum, whereas soya oil–based emulsions with the same emulsifiers did not [123]. It was hypothesized that this effect was caused by the adsorption of plasma proteins onto the tocol droplets (opsonization). Indeed, the steric stabilization of emulsions by incorporation of emulsifiers like poloxamer 188 or PEGylated phospholipids such as $PEG_{5000}PE$ proved to be effective in the stabilization of tocol-based emulsions in plasma. Conversely, in vitro studies were

conducted on plasma protein adsorption onto the blank second- and third-generation emulsion droplets [37, 139] to mimic the in vivo opsonization phenomenon responsible for the rapid clearance of the emulsion droplets from the blood. According to these authors, the adsorption of many protein species such as apoAs, apoCs, apoE albumin, fibrinogen, and gamma globulin onto the emulsion droplet surfaces is detectable by two-dimensional polyacrylamide gel electrophoresis.

### 5.2 Ocular Routes

For the eye, the method of drug delivery is important. However, when nanosized emulsion is used as a vehicle for ocular drug delivery purposes, both topical/local and intraocular routes of administration can be possible (though no data concerning intraocular drug delivery through emulsion are currently available). The o/w nanosized emulsions having both anionic and cationic charges provide a liquid-retentive carrier for ocular active agents, particularly when topically instilled into the eye. It is interesting to add here that thermodynamically stable and optically isotropic colloidal systems such as the w/o microemulsion is also designed nowadays for ocular topical. As delivery the w/o microemulsion system comprises both aqueous and oily components into its multistructure, it has the ability to incorporate considerable amounts of both hydrophilic and lipophilic drugs [140]. In fact, in comparison to ocular inserts or implants and semisolid ocular preparations, the liquid-retentive nature gives impetus to investigating further the emulsion-based ophthalmic drug delivery as it has the benefit of being comfortable to use for both ophthalmologists and patients. In addition, through topical instillation of emulsions possessing ocular active substances, the delivery of drug molecules even to the posterior portion of the eye might be of possible. In this context, the third-generation emulsion is being designed by adsorbing electrostatically the therapeutic oligonucleotides onto its surface for modulating functions of retinal pigment epithelium (RPE) cells effectively in order to treat blindness associated with age-related macular degeneration (AMD), proliferative vitreoretinopathy, retinal and choroidal neovascularization, and retinitis pigmentosa. To achieve this, it becomes necessary to know first the ocular protective mechanisms and other concomitant factors to be faced by emulsion droplets following ocular topical application. This point is further developed below.

Considerations of ocular drug delivery are not detailed in this section. Pertinent information concerning factors affecting drug permeation or retention as well as eye anatomy and physiology can be found in several reviews [141–146]. From a medical point of view, o/w nanosized emulsions for ophthalmic use aim at enhancing drug bioavailability either by providing prolonged delivery to the eye or by facilitating transcorneal/transconjunctival penetration. Drugs incorporated in o/w nanosized emulsions are lipophilic in nature, and depending on the extent of lipophilicity, either the corneal or the conjunctival/scleral route of penetration may be favored [147]. For the more lipophilic drugs the corneal route was shown to be the predominant pathway for delivering drugs to the iris, whereas the less lipophilic drugs underwent conjunctival/scleral penetration for delivery into the ciliary body [147]. Thus, transcorneal permeation has traditionally been the mechanism by which topically applied ophthalmic drugs are believed to gain access to the internal ocular structures. Relatively little attention has been given to alternate routes by which drugs may enter the eye. It was reported that drugs may be absorbed by the

noncorneal route and appeared to enter certain intraocular tissues through the conjunctiva/sclera [148–150]. Indeed when compared to the cornea, drug penetration through the conjunctiva has the advantage of a larger surface area and higher permeability, at least for drugs which are not highly lipophilic. Furthermore, the lasting presence of drug molecules in the lower conjunctival cul-de-sac of the eye could result in a reservoir effect. Nevertheless, the o/w nanosized emulsions more or less physically resemble a simple aqueous-based eye drop dosage form since more than 90% of the external phase is aqueous irrespective of the formulation composition. Hence, following topical administration, nanosized emulsions would probably face almost similar ocular protective events as encountered with conventional eye drops into the eye. The o/w nanosized emulsions are likely to be destabilized by the tear fluid electrolytic and dynamic action. Because of constant eyelid movements, the basal tear flow rate (1.2 μL/min), and the reflex secretion induced by instillation (up to 400 μL/min depending on the irritating power of the topical ocular solutions [35]), topical eye drop dosage forms are known for being rapidly washed out from the eye. Therefore, the water phase of the emulsion is drained off while, probably, the oil phase of the emulsion remains in the cul-de-sac for a long period of time and functions as a drug reservoir [35]. If the volume of instilled emulsion in the eye exceeds the normal lachrymal volume of 7–10 μL, then the portion of the instilled emulsion (one or two drops, corresponding to 50–100 μL) that is not eliminated by spillage from the palpebral fissure of conjunctiva is drained quickly via the nasolacrimal system into the nasopharynx. In other words, the larger the instilled volume, the more rapidly the instilled emulsion is drained from the precorneal area. Hence the contact time of the emulsion with the absorbing surfaces (cornea and conjunctiva) is estimated to be a maximum of a few minutes, well beyond the short residence time of conventional eye drops. In order to verify the extension of the residence time of the emulsion in the conjunctival sac, Beilin et al. [151] added a fluorescent marker to the formulations. One minute after the topical instillation into eye, 39.9 ± 10.2% of the fluorescence was measured for the nanosized emulsions whereas only 6.8 ± 1.8% for regular eye drops. In addition a study was carried out in male albino rabbits to compare the corneal penetration of indomethacin from Indocollyre (a marketed hydro-PEG ocular solution) to that of negatively and positively charged emulsions [108]. By this comparison, it was intended to gain insightful mechanistic comprehension regarding the enhanced ocular penetration effect of the emulsion as a function of dosage form and surface charge. The contact angle of one droplet of the different dosage forms on the cornea was measured and found to be 70° for saline, 38° for the anionic emulsion, and 21.2° for the cationic emulsion. Respectively, the values of the calculated spreading coefficient were –47, –8.6, and –2.4 mN/m. It can clearly be deduced, owing to the marked low spreading coefficient values elicited by the emulsions, that both nanosized emulsions had better wettability properties on the cornea compared to saline. The emulsion may then prolong the residence time of the drop on the epithelial layer of the cornea, thereby enabling better drug penetration through the cornea to the internal tissues of the eye, as confirmed by animal studies [108]. It is therefore believed that drug is not released from the oil droplet and equilibrates with the tears but rather partitions directly from the oil droplets to the cell membranes on the eye surface. Therefore, it is reasonable to consider that nanosized emulsions have a real advantage since they elicit an increased ocular residence time in com-

**TABLE 5 Selected List of o/w Nanosized Emulsions for Ocular Topical Drug Delivery**

| Emulsion Type | Drug Used | Reference |
|---|---|---|
| Anionic emulsion | $\Delta^8$-THC | 154 |
| | Pilocarpine base and indomethacin | 163 |
| | Adaprolol maleate | 161, 162 |
| | Indomethacin | 159, 160 |
| | Synthetic HU-211 and pilocarpine base | 124, 155, 156 |
| | Pilocarpine base | 157, 158, 164 |
| | Cyclosporin A | 168, 165–177 |
| Cationic emulsion | Piroxicam | 178 |
| | Indomethacin | 108 |
| | Miconazole | 112 |
| | Cyclosporin A | 106, 179 |

*Note:* $\Delta^8$-THC and synthetic HU-211 are derivatives of Cannabis sativa.

parison to conventional eye drops and will significantly improve the ocular drug bioavailability [152]. This is also confirmed in numerous cited papers that are listed in Table 5.

In spite of a relatively rapid removal of conjunctivally absorbed emulsion from the eye by local circulation, direct transscleral access to some intraocular tissues cannot be excluded, especially if an electrostatic attraction does occur between the cationic oil droplets of emulsion and anionic membrane moieties of the sclera, as shown by some authors [108]. There is no doubt that the drug absorption from emulsion through the noncorneal route needs to be investigated further as it may elicit useful information on the potential of nanosized emulsions to promote drug penetration to the posterior segment through a mechanism which bypasses the anterior chamber. In addition to the above-described protective and elimination mechanisms of the eye, nanosized emulsions remaining in the precorneal area may be subject to protein binding and to metabolic degradation in the tear film. In conjunction with blood plasma, although low, tear film, aqueous humor, and vitreous humor also have varying amounts of relatively detectable proteins such as albumin, globulin, and immunoglobulins (e.g., IgA, IgG, IgM, IgE) and the enzyme lysozyme. Additional studies (at least in vitro) are necessary to understand clearly the nanosized emulsion interaction with the ocular fluid components. Although it is unlikely to happen because of the low emulsion volume remaining in the conjunctival sac, the fluid dynamics may be moderately altered by the physical and chemical properties of nanosized emulsions, which include tonicity, pH, refractive index, interfacial charge, viscosity, osmolality, and irritant ingredients. Thus, formulations of ophthalmic drug products must take into account not only the stability and compatibility of a drug in the emulsion but also the influence of the emulsion on precorneal fluid dynamics. All of the concepts exposed in this section may ultimately result in transcorneal/conjunctival absorption of 1–2% or less of the drug applied topically through the emulsions. In summary, the rate of loss of drug/emulsion from the eye can be 500–700 times greater than the rate of absorption into the eye. Thus, conventional topical delivery using emulsions cannot achieve adequate intracellular concentrations of drugs or other substances such as oligonucleotides or genes for the treatment of endophthalmitis or other sight-threatening intraocular diseases (e.g., AMD).

In order to achieve a high concentration of drug within the eye using an emulsion delivery vehicle, an approach that bypasses physiological and anatomical barriers (e.g., blood–ocular) of the eye is a more viable and attractive option. One such approach is to administer emulsion through direct intraocular injections such as periocular (subconjunctival and sub-Tenon), intracameral, intravitreal, intracapsular, or subretinal. Moreover, it is likely that intraocularly administered emulsion is able to both significantly increase drug half-life and minimize intraocular side effects that appear following intraocular injection of drug alone. In general, drugs encapsulated within emulsion are less toxic than their free counterparts. Additionally, there is a possibility of obtaining slow drug release from an intraocularly injected emulsion. Taking into account the nonphagocytic character of neural retinal cells and the ability of RPE cells to take up large molecules, including oligonucleotides, the third-generation emulsion for intravitreal or subretinal injections is more likely to be a successful approach in future. Moreover, intravitreally administered drug molecules are able to bypass the blood–ocular barrier to achieve constant therapeutic levels in the eye while minimizing systemic side effects. However, the hyalocytes, the main cellular components of the vitreous, have been classified in at least one report [153] as macrophages and thus may play a role in the uptake of intravitreally injected emulsion. It should be added that no studies are focused so far on injecting emulsion intraocularly and significant work should be devoted to generate this novel idea into a fruitful solution in ophthalmic drug delivery applications.

Over the last decade, o/w nanosized emulsions containing either anionic or cationic droplets have been recognized as interesting and promising ocular topical delivery vehicles for lipophilic drugs. Complete details are available elsewhere [117]. As an overview of this topic, important results on emulsion-based ocular topical drug delivery are covered below and are listed in Table 5.

The in vivo data obtained from studies of early formulations confirm that o/w nanosized anionic emulsions can be effective topical ophthalmic drug delivery systems [154] with a potential for sustained drug release [155]. Naveh and coworkers [156] have also noted that the intraocular pressure (IOP)–reducing effect of a single, topically administered dose of pilocarpine-loaded anionic emulsion lasted for more than 29 h in albino rabbits whereas that of the generic pilocarpine lasted only 5 h. Zurowska-Pryczkowska et al. [157] studied how nanosized emulsion as a vehicle influences the chemical stability of pilocarpine and the effect the drug has on the physical stability of nanosized emulsions. In a subsequent paper [158] from the same group on in vivo evaluation using normotensive rabbits, it was shown that the nanosized emulsion formulated with pilocarpine hydrochloride at pH 5.0 could be indicated as a preparation offering prolonged pharmacological action (miotic effect) together with satisfactory chemical stability. However, the ocular bioavailability arising from such a formulation is not significantly improved when compared to an aqueous solution. Calvo et al. [159, 160] observed an improvement in indomethacin ocular bioavailability when the drug was incorporated in an emulsion dosage form with respect to the commercial aqueous drops following topical application into rabbit eye.

In order to verify the extension of the residence time of the emulsion in the conjunctival sac, Beilin et al. [151] added a fluorescent marker to the formulations, as mentioned previously. From that observation, it is reasonable to consider that an emulsion formulation has the real advantage of increasing ocular residence time in

comparison to eye drops. Anselem et al. [161] and Melamed et al. [162] prepared a nanosized emulsion containing adaprolol maleate, a novel soft β-blocking agent, and observed a delayed IOP depressant effect in human volunteers. A similar pharmacological effect was also observed in human volunteers by Aviv et al. [163] using pilocarpine base-loaded emulsion. Another randomized human trial conducted by Garty et al. [164] compared the activity of the pilocarpine base-laden nanosized emulsion instilled twice daily with a generic dosage form instilled four times a day to 40 hypertensive patients for seven days. No local side effects were observed. The IOP decreased 25% in both formulations during this time period. No significant difference was noticed between the two treatments. These results proved that the anionic emulsion extended the action of the drug and two daily administrations have the same result as four instillations of regular eye drops.

A novel nanosized anionic emulsion incorporating the immunomodulatory agent cyclosporin A was developed and the clinical efficacy was investigated for the treatment of moderate to severe dry-eye disease in humans [165–167]. The novel cyclosporin A ophthalmic dosage form represents a real breakthrough in the formulation of a complex, highly lipophilic molecule such as cyclosporin A within an o/w nanosized emulsion. Following thorough validation of this formulation through several clinical studies in various countries [165–175], an anionic o/w emulsion containing cyclosporin A 0.05% (Restasis, Allergan, Irvine, CA) was approved for the first time by the FDA, on December 23, 2002. In addition, this anionic emulsion having cyclosporin A is now available at pharmacies in the United States for the treatment of chronic dry-eye disease (available at www.restasis.com and www.dryeye.com). Furthermore, an over-the-counter (OTC) product that features an emulsion formula, Refresh Endura, is already launched in the U.S. market for eye-lubricating purposes in patients suffering from moderate to severe dry-eye syndrome.

The effect of Restasis on contact lens comfort and reducing dry-eye symptoms in patients with contact lens intolerance was evaluated in comparison to rewetting drops (carboxymethylcellulose 0.5%, Refresh Contacts) [176]. Both formulations were applied twice per day before and after lens wear. Symptoms were assessed by lens wear time, use of rewetting drops during lens wear, subjective evaluation of dryness, and completion of the ocular surface disease index questionnaire. The results of this pilot study indicate that cyclosporin 0.05% is beneficial for contact lens wearers with dry eye and reduces contact lens intolerance [176]. Furthermore, Sall et al. [177] have recently evaluated the efficacy of marketed artificial tears (Systane and Restasis) in relieving the signs and symptoms of dry eye when used as supportive therapy to a cyclosporin-based ophthalmic emulsion (i.e., Restasis + Systane vs. Restasis + Refresh). Significant differences were seen in favor of Restasis + Systane versus Restasis + Refresh for less ocular burning, stinging, grittiness, and dryness. Systane was better than Restasis + Refresh for less burning, dryness, and scratchiness. Results indicate that the choice of concomitant therapy used with Restasis has significant effects on outcome measures and both supportive therapies were compatible with Restasis [177].

When compared to either saline or anionic emulsions, the nanosized cationic emulsions were shown to enhance the ocular bioavailability of indomethacin [108], piroxicam [178], and cyclosporin A [106, 179] following one single-drop dose instillation into the rabbit eye (Figure 4). A significant drug reservoir effect was noted in the cornea and conjunctiva even for more than 8 h following the instillation [106].

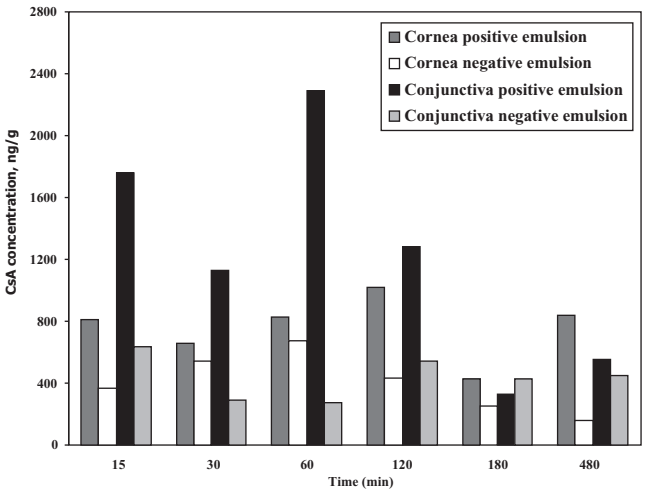


**FIGURE 4** Influence of emulsion surface charges cyclosporin A (CsA) concentrations in ocular surface tissues (cornea and conjunctiva) following one single-dose (50-μL) instillation of positively (cationic) and negatively (anionic) charged CsA-loaded nanosized emulsions into albino rabbit eye.

This long residence time may help reduce evaporation of the limited volume of natural tears present in patients with dry eye. This was probably due to the adhesion of the positively charged oil droplets to the negatively charged corneal surface moieties as a result of electrostatic attraction. This hypothesis was supported by data from an ex vivo study which showed that cationic emulsion exhibited better wettability properties on albino rabbit eye cornea than either saline or anionic emulsion [108]. Associated with Poloxamer and phospholipids, a cationic primary amine, stearylamine, has been used to obtain the above-described third-generation cationic emulsions. Additionally, a cationic emulsion based on an association of Poloxamer 188 and chitosan was prepared and also showed interesting physicochemical properties on stability and charge effects [97, 98]. Moreover, the stability and ocular tolerance following topical instillation into the eye of these cationic emulsion vehicles were investigated [98, 114]. The overall studies hence stress the effectiveness of nanosized cationic emulsion, which promotes ocular drug absorption via internalization possibly through an endocytic process [112].

### 5.3 Nasal Route

The nasal route is still receiving great attention due to a number of advantages over parenteral and oral administration [180], particularly when first-pass metabolism makes the drug ineffective. The approach of an o/w emulsion formulation of the drug may increase absorption by solubilizing the drug in the inner phase of the emulsion and prolonging contact time between emulsion droplets and nasal mucosa.

One of the first examples for systemic delivery of peptides concerned a lipid-soluble rennin inhibitor [181]. The peptide was solubilized in the oil phase of an o/w emulsion containing membrane adjuvants such as oleic acid and mono- and diglyc-

erides. Emulsion formulations have been proposed to simultaneously increase the solubility of the peptide and to enhance membrane permeability through interaction between the membrane and the oil components. Enhanced and prolonged in vivo nasal absorption of the rennin inhibitor was observed in emulsion formulation compared to aqueous suspension. From morphological studies, the emulsions did not provoke any significant changes to the nasal mucosa [181]. Such a formulation approach was also used for the administration of a steroidal male sex hormone testosterone [182]. The steroid was solubilized in the oil phase of the o/w emulsion and the ionic composition of the aqueous phase was modified in order to produce electrically positive, negative, and neutral droplets. Droplets with a surface charge led to better bioavailability than neutral droplets, but contrary to the above-described topical applicabilities of cationic emulsions over anionic emulsions, positively charged droplets did not provide the best results [182]. However, the emulsion approach was advantageous since it helped to overcome the solubility problem of the hydrophobic compounds.

In another study which does not involve peptide drugs, various emulsion formulations were prepared in order to modulate the partitioning of the drug between the aqueous phase and the oil phase [183]. The disappearance of a drug from the nasal cavity was determined by an in situ perfusion technique. When the drug was solubilized in the aqueous phase, the formulation did not have a significant effect on the drug disappearance rate. However, partitioning of the drug in the oil phase resulted in delaying absorption. It was suggested that oil droplets containing medium-chain triglycerides formed a pseudo–oily layer on the mucous membrane, which slowed down the drug disappearance from the nasal cavity [183]. Another interesting study reported nasal delivery of insulin formulated in both o/w and w/o emulsions [184]. As insulin partitions into the aqueous phase of the emulsion, the peptide is either incorporated within the continuous phase of the o/w emulsion or encapsulated in the aqueous droplets of the w/o emulsion. Following nasal perfusion experiments, plasma insulin concentration profiles showed enhanced insulin absorption when the peptide was formulated as an o/w emulsion compared to an aqueous solution. However, a w/o emulsion did not cause any significant increase in plasma insulin concentration. Delivery of insulin by administration of nasal drops also revealed a large dose-dependent increase in plasma insulin concentration. It also needs to be pointed out that the emulsifier mixture alone did not promote any absorption. It was suggested that insulin molecules probably were adsorbed at the surface of the oil droplets. Adhesion of the oil droplets on the mucosal membrane then induced a local increase in insulin concentration at the membrane surface. However, the number of droplets in contact with the surface had to be small enough. Otherwise, a stagnant oil layer is formed which acts as an additional barrier to the transport, as was observed with the w/o emulsion [183].

Other recent applications of emulsion formulation involve mucosal gene and vaccine delivery [185–187] and the preparation of polymeric microspheres by the w/o emulsification solvent extraction technique [188].

### 5.4 Topical Route

Many drugs exhibit low skin penetration, which results in poor efficacy. As opposed to common chemical skin penetration enhancers, organic solvents, which are

generally associated to some degree with skin irritation, toxicity and sensitization, a solvent-free topical vehicle based on drug entrapment in o/w emulsion droplets of submicrometer size is more efficacious in terms of percutaneous absorption and possibly devoid of adverse effects. In addition, the uniqueness of the large internal hydrophobic oil core of o/w nanosized emulsion droplets allows high solubilization capacity for water-insoluble topically active medicaments and also aids in carrying water, an excellent softener, to the skin.

The concept of using anionic nanosized emulsion vehicles for enhanced percutaneous absorption of nonsteroidal anti-inflammatory drugs (NSAIDs) and diazepam was clearly proven [189, 190]. NSAIDs and diazepam in a nanosized emulsion vehicle also demonstrated noticeable systemic activity. The o/w emulsion was tested for primary irritation in humans in a 48-h trial. Low irritancy and excellent human acceptance were observed, subsequently making the further development of a nanosized emulsion vehicle very attractive.

Even though emulsion vehicles increase dermal drug delivery of lipophilic drugs in humans, one of the problems for topical drug delivery is the difficulty of applying these vehicles to the skin because of their fluidity. Rheological properties are studied in transdermal formulations and different results are given. Realdaon and Ragazzi [191] have investigated different mechanical emulsifying conditions on o/w emulsion formulations containing methyl nicotinate. The influence of these procedures on rheological properties and in vivo availability of methyl nicotinate was evaluated. Even if various viscosities were obtained, differences between batches did not compromise drug availability. On the contrary, Welin-Berger and co-workers [192, 193] concluded in their study on nanosized emulsions containing model compounds that both release and permeation rates decrease when the apparent yield stress (i.e., the macroviscosity) increases by addition of gelling polymers. Because a topical anesthetic agent will induce a pain-suppressing anesthesia, the eutectic mixture of local anesthetics (EMLA) has proven to be a useful medication for children. It is an emulsion containing a mixture of lidocaine and prilocaine. This cream gives an effective deep sedation and can be applied half an hour to 1 h prior to the procedure. Local side effects with this emulsion are very mild [194, 195]. Systemic activity of nanosized emulsions containing diazepam was compared with regular creams or ointments by Schwarz et al. [190]. Their efficiency was tested on protection against pentamethylenetetrazole, which induces convulsive effects in mice. Diazepam applied topically in emulsion creams was strongly dependent on oil droplet size. Furthermore, nanosized emulsions increased transdermal drug delivery and prolonged protective activity for up to 6 h.

Many formulations for topical emulsions are available in the scientific literature, in patents, and on the market. Progresses made in the last years in this field are concentrated on the various aspects of drug release and the influence of droplet size.

Third-generation cationic emulsions were suggested as drug carriers for topical use in the skin. It was found that α-tocopherol-loaded cationic emulsion was able to prevent oxidative damage of cultured fibroblast cells [196]. In addition, the same cationic formulation was able to protect rat skin against oxidative stress induced by ultraviolet (UV) irradiation significantly better than either the corresponding anionic emulsion or the cationic blank emulsion, as measured with a noninvasive

evaluation of the lipid hydroperoxidation process of the rat skin. However, no difference was found between cationic or anionic nanosized emulsions of α-tocopherol as assessed by endogenous peroxyl radical scavenging ability. Taken together, these results suggest that the cationic emulsion allows the α-tocopherol to remain on the surface of the skin because of electrostatic interactions between the negatively charged sites of the superficial layers of the skin and the positively charged oil droplets. Although the extent of α-tocopherol incorporation into the skin is similar for both cationic and anionic emulsions, the prolonged skin surface residence time of the cationic emulsion allows an enhanced protective effect against oxidative stress. In contrast to these results, an in vitro percutaneous absorption study on hairless rat skin found that the antifungal drugs econazole and miconazole nitrate incorporated into a similar cationic emulsion formulation were more effective in terms of skin penetration than the corresponding anionic emulsion [197]. The enhanced rate of diffusion of these antifungal drugs through the skin by the cationic emulsion suggests a new approach for dermal drug penetration enhancement [197].

## 6 FUTURE PERSPECTIVE

Based on the performances in previous and present decades, o/w-type nanosized emulsions can conveniently be classified into three generations. First-generation emulsions are considered primarily as nutrient carriers to be administered via intravenous routes to bed-ridden patients. Second-generation emulsions start initially as drug carrier systems by solubilizing considerable amounts of lipophilic drugs at the oil phase or at the oil–water interface of the emulsion. This particular merit of emulsions is specifically exploited even commercially for both ocular and parenteral active drugs. Modifications made either in the oil phase or at the o/w interfacial film forming emulsifier molecules allow the emulsions to be able to escape from lipolysis by lipoprotein lipase, apo adsorption, and liver uptake. Such a surface-modified emulsion would prolong the circulation time in plasma and thereby an alteration in in vivo disposition of incorporated drugs following parenteral administration. Attachment of homing devices such as antibody and apoE to the surfaces of emulsions makes the selective/active delivering of emulsion-incorporated drugs to target sites such as a tumorized organ or hepatic system. Active targeting increases the affinity of the carrier system for the target site, while passive targeting minimizes the nonspecific interaction with nontargeted sites by the RES. Having together a positive charge and a steric stabilizing effect led to the development of third-generation emulsions that contain a unique property: plasma half-life prolongation and electrostatic adhesion to ocular surface tissues after topical instillation into eye. Furthermore, the third-generation emulsion shows a potential of carrying a wide range of lipophilic, amphiphilic, and polyanionic compounds, including DNA and oligonucleotides for transdermal and nasal routes. Accumulating knowledge thus suggests that constant progress in better understanding the principles and processes governing the various issues related to o/w nanosized emulsions has surely brought major improvements in the efficacy of parenteral or nonparenteral drug delivery systems.

## REFERENCES


1. Lipinski, C. A, Lombardo, F., Dominy, B. W., and Feeney, P. J. (2001), Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings, *Adv. Drug Deliv. Rev.*, 46, 3–26.
2. Lipinski, C. (2002), Poor aqueous solubility—An industry wide problem in drug discovery, *Am. Pharm. Rev.*, 5, 82–85.
3. El-Aasser, M. S., and Sudol, E. D. (2004), Miniemulsions: Overview of research and applications, *JCT Res.*, 1, 21–31.
4. Nakajima, H. (1997), Microemulsions in cosmetics, in Solans, C., and Kunieda, H., Eds., *Industrial Applications of Microemulsions*, Marcel Dekker, New York, pp. 175–197.
5. Benita, S. (1998), Introduction and overview, in Benita, S., Ed., *Submicron Emulsion in Drug Targeting & Delivery*, Harwood Academic, The Netherlands, pp.1–5.
6. Klang, S. H., Benita, S. (1998), Design and evaluation of submicron emulsions as colloidal drug carriers for intravenous administration, in Benita, S., Ed., *Submicron Emulsion in Drug Targeting & Delivery*, Harwood Academic, The Netherlands, pp. 119–152.
7. Tadros, T. F., Izquierdo, P., Esquena, J., and Solans, C. (2004), Formation and stability of nanoemulsions, *Adv. Coll. Interf. Sci.*, 108–109, 303–318.
8. Solans, C., Izquierdo, P., Nolla, J., Azemar, N., and Garcia-Celma, M. J. (2005), Nanoemulsions, *Curr. Opin. Coll. Interf. Sci.*, 10, 102–110.
9. Becher, P., and Schick, M. J. (1987), Macroemulsions, in Schick, M. J., Ed., *Non-Ionic Surfactants*, Surfactant Science Series Vol. 23., Marcel Dekker, Basel, pp. 435–493.
10. Kabalnov, A. (1998), Thermodynamic and theoretical aspects of emulsions and their stability, *Curr. Opin. Coll. Interf. Sci.*, 3, 270–275.
11. Stefan, A., Palazzo, G., Ceglie, A., Panzavolta, E., and Hochkoeppler, A. (2003), Water-in-oil macroemulsions sustain long-term viability of microbial cells in organic solvents, *Biotechnol. Bioeng.*, 81, 323–328.
12. Ceglie, A., Das, K. P., and Lindman, B. (1987), Microemulsion structure in four component systems for different surfactants, *Coll. Surf.*, 28, 29–40.
13. Attwood, D. (1994), Microemulsions, in Kreuter, J., Ed., *Colloidal Drug Delivery Systems*, Marcel Dekker, New York, pp. 31–71.
14. Hino, T., Kawashima, Y., and Shimabayashi, S. (2000), Basic study for stabilization of w/o/w emulsion and its application to transcatheter arterial embolization therapy, *Adv. Drug Deliv. Rev.*, 45, 27–45.
15. Gallarate, M., Carlotti, M. E., Trotta, M., and Bovo, S. (1999), On the stability of ascorbic acid in emulsified systems for topical and cosmetic use, *Int. J. Pharm.*, 188, 233–241.
16. ErÃfnofas, I., Csoka, I., Csany, E., Orosz, K., and Makai, M. (1998), in *Proceedings of 2nd World Meeting APGI/APV*, Paris, pp. 805–806.
17. Lowe, K. C. (1999), Perfluorinated blood substitutes and artificial oxygen carriers, *Blood Rev.*, 13, 171–184.
18. Spahn, D. R. (2000), Current status of artificial oxygen carriers, *Adv. Drug Deliv. Rev.*, 40, 143–151.
19. Krafft, M. P. (2001), Fluorocarbons and fluorinated amphiphiles in drug delivery and biomedical research, *Adv. Drug Deliv. Rev.*, 47, 209–228.
20. Fukushima, S., Kishimoto, S., Takeuchi, Y., and Fukushima, M. (2000), Preparation and evaluation of o/w type emulsions containing antitumor prostaglandin, *Adv. Drug Deliv. Rev.*, 45, 65–75.

21. Ott, G., Singh, M., Kazzaz, J., Briones, M., Soenawan, E., Ugozzoli, M., and O'Hagan, D. T. (2002), A cationic sub-micron emulsion (MF59/DOTAP) is an effective delivery system for DNA vaccines, *J. Controlled Release*, 79, 1–5.
22. Ribi, E. E. (1984), Refined detoxified endotoxin product, U.S. Patent 4,436,727, March 13.
23. Ribi, E., Schwartzman, S. M., and Cantrell, J. L. Refined detoxified endotoxin product, (1984), U.S. Patent 4,436,728, March 13.
24. Myers, K. R., and Truchot, A. T. (1990), Modified lipopolysaccharides and process of preparation, U.S. Patent 4,912,094, March 27.
25. Allison, A. C., and Byars, N. E. (1986), An adjuvant formulation that selectively elicits the formation of antibodies of protective isotypes and of cell-mediated immunity, *J. Immunol. Methods*, 95, 157–168.
26. Chandra, R. K. (1985), Grace A. Goldsmith Award lecture. Trace element regulation of immunity and infection, *J. Am. Coll. Nutr.*, 4, 5–16.
27. Shils, M. E. (1998), Parentral nutrition, in Shils, M. E., Olson, J. A., Shike, M. E., and Ross, O., Eds., *Modern Nutr. in Health and Disease*, 9th ed., Williams & Wilkins, Baltimore, PA, pp. 415–428.
28. Koster, V. S., Kuks, P. F. M., Langer, R., and Talsma, H. (1996), Particle size in parenteral fat emulsions, what are the true limitations? *Int. J. Pharm.*, 134, 235–238.
29. Lutz, O., Meraihi, Z., Mura, J. L., Frey, A., Riess, G. H., and Bach. A. C. (1989), Fat emulsion particle size: Influence on the clearance rate and the tissue lipolytic activity, *Am. J. Clin. Nutr.*, 50, 1370–1381.
30. Rubin, M., Harell, D., Naor, N., Moser, A., Wielunsky, E., Merlob, P., and Lichtenberg, D. (1991), Lipid infusion with different triglyceride cores (long-chain vs medium-chain/long-chain triglycerides): Effect on plasma lipids and bilirubin binding in premature infants, *J. Parenteral Enteral Nutr.*, 15, 642–646.
31. Mizushima, Y., Hoshi, K., Aihara, H., and Kurachi, M. (1983), Inhibition of bronchoconstriction by aerosol of a lipid containing prostaglandin E1, *J. Pharm. Pharmacol.*, 35, 397.
32. Craig, D. Q. M., Patel, M., and Ashford, M. (2000), Administration of emulsions to the gastrointestinal tract, in Nielloud, F., and Marti-Mestres, G., Eds., *Pharmaceutical Emulsions and Suspensions*, Marcel Dekker, New York, pp. 323–360.
33. Hauss, D. J. (2002), Oral lipid-based drug delivery—a case of implementation failing to keep up with innovation? *Am. Pharm. Rev.*, 5, 22–36.
34. Sakaeda, T., and Hirano, K. (1998), Effect of composition on biological fate of oil particles after intravenous injection of O/W lipid emulsions, *J. Drug Target.*, 6, 273–284.
35. Sasaki, H., Yamamura, K., Nishida, K., Nakamura, J., and Ichikawa, M. (1996), Delivery of drugs to the eye by topical application, *Prog. Ret. Eye Res.*, 15, 583–620.
36. Harnisch, S., and Müller, R. H. (1998), Plasma protein adsorption patterns on emulsions for parenteral administration: Establishment of a protocol for two-dimensional polyacrylamide electrophoresis, *Electrophoresis*, 19, 349–354.
37. Harnisch, S., and Müller, R. H. (2000), Adsorption kinetics of plasma proteins on oil-in-water emulsions for parenteral nutrition, *Eur. J. Pharm. Biopharm.*, 49, 41–46.
38. Szebeni, J. (2001), Complement activation-related pseudoallergy caused by liposomes, micellar carriers of intravenous drugs, and radiocontrast agents, *Crit. Rev. Ther. Drug Carrier Syst.*, 18, 567–606.
39. Price, M. E., Cornelius, R. M., and Brash, J. L. (2001), Protein adsorption to polyethylene glycol modified liposomes from fibrinogen solution and from plasma, *Biochim. Biophys. Acta*, 1512, 191–205.

40. Deckelbaum, R. J., Hamilton, J. A., Moser, A., Bengtsson-Olivecrona, G., Butbul, E., Carpentier, Y. A., Gutman, A., and Olivecrona, T. (1990), Medium-chain versus long-chain triacylglycerol emulsion hydrolysis by lipoprotein lipase and hepatic lipase: Implications for the mechanisms of lipase action, *Biochemistry*, 29, 1136–1142.
41. Hedeman, H., Brøndsted, H., Müllertz, A., and Frokjaer, S. (1996), Fat emulsions based on structured lipids (1,3-specific triglycerides): An investigation of the in vivo fate, *Pharm. Res.*, 13, 725–728.
42. Takino, T., Konishi, K., Takakura, Y., and Hashida, M. (1994), Long circulating emulsion carrier systems for highly lipophilic drugs, *Biol. Pharm. Bull.*, 17, 121–125.
43. Takino, T., Nakajima, C., Takakura, Y., Sezaki, H., and Hashida, M. (1993), Controlled biodistribution of highly lipophilic drugs with various parenteral formulations, *J. Drug Target.*, 1, 117–124.
44. Redgrave, T. G., Rakic, V., Mortimer, B.-C., and Mamo, J. C. L. (1992), Effects of sphingomyelin and phosphatidylcholine acyl chains on the clearance of triacylglycerol-rich lipoproteins from plasma. Studies with lipid emulsions in rats, *Biochim. Biophys. Acta*, 1126, 65–72.
45. Arimoto, I., Matsumoto, C., Tanaka, M., Okuhira, K., Saito, H., and Handa, T. (1998), Surface composition regulates clearance from plasma and triolein lipolysis of lipid emulsions, *Lipids*, 33, 773–779.
46. Handa, T., Eguchi, Y., and Miyajima, K. (1994), Effects of cholesterol and cholesteryl oleate on lipolysis and liver uptake of triglyceride/phosphatidylcholine emulsions in rats, *Pharm. Res.*, 11, 1283–1287.
47. Kurihara, A., Shibayama, Y., Mizota, A., Yasuno, A., Ikeda, M., and Hisaoka, M. (1996), Pharmacokinetics of highly lipophilic antitumor agent palmitoyl rhizoxin incorporated in lipid emulsions in rats, *Biol. Pharm. Bull.*, 19, 252–258.
48. Kurihara, A., Shibayama, Y., Yasuno, A., Ikeda, M., and Hisaoka, M. (1996), Lipid emulsions of palmitoylrhizoxin: Effects of particle size on blood dispositions of emulsion lipid and incorporated compound in rats, *Biopharm. Drug Dispos.*, 17, 343–353.
49. Lin, S. Y., Wu, W. H., and Lui, W. Y. (1992), In vitro release, pharmacokinetic and tissue distribution studies of doxorubicin hydrochloride (Adriamycin HCl) encapsulated in lipiodolized w/o emulsions and w/o/w multiple emulsions, *Die Pharmazie*, 47, 439–443.
50. Sakaeda, T., Takahashi, K., Nishihara, Y., and Hirano, K. (1994), O/W lipid emulsions for parenteral drug delivery. I. Pharmacokinetics of the oil particles and incorporated sudan II, *Biol. Pharm. Bull.*, 17, 1490–1495.
51. Yamaguchi, T., Nishizaki, K., Itai, S., Hayashi, H., and Ohshima, H. (1995), Physicochemical characterization of parenteral lipid emulsion: Influence of cosurfactants on flocculation and coalescence, *Pharm. Res.*, 12, 1273–1278.
52. Ueda, K., Yamazaki, Y., Noto, H., Teshima, Y., Yamashita, C., Sakaeda, T., and Iwakawa, S. (2003), Effect of oxyethylene moieties in hydrogenated castor oil on the pharmacokinetics of menatetrenone incorporated in O/W lipid emulsions prepared with hydrogenated castor oil and soybean oil in rats, *J. Drug Target.*, 11, 37–43.
53. Liu, F., and Liu, D. (1995), Long-circulating emulsions (oil-in-water) as carriers for lipophilic drugs, *Pharm. Res.*, 12, 1060–1064.
54. Lee, J. H., Lee, H. B., and Andrade, J. D. (1995), Blood compatibility of polyethylene oxide surfaces, *Prog. Polym. Sci.*, 20, 1043–1079.
55. Davis, S. S., and Hansrani, P. (1985), The influence of emulsifying agents on the phagocytosis of lipid emulsions by macrophages, *Int. J. Pharm.*, 23, 69–77.
56. Levy, M. Y., Benita, S., and Baszkin, A. (1991), Interactions of a non-ionic surfactant with mixed phospholipid-oleic acid monolayers. Studies under dynamic conditions, *Coll. Surf.*, 59, 225–241.

57. Harris, J. M., Martin, N. E., and Modi, M. (2001), Pegylation: A novel process for modifying pharmacokinetics, *Clin. Pharmacokinet.*, 40, 539–551.
58. Bhadra, D., Bhadra, S., Jain, P., and Jain, N. K. (2002), Pegnology: A review of PEG-ylated systems, *Die Pharm.*, 57, 5–29.
59. Klibanov, A. L., Maruyama, K., Torchilin, V. P., and Huang, L. (1990), Amphipathic polyethyleneglycols effectively prolong the circulation time of liposomes, *FEBS Lett.*, 268, 235–237.
60. Allen, T. M., Hansen, C., Martin, F., Redemann, C., and Yau-Young, A. (1991), Liposomes containing synthetic lipid derivatives of poly(ethylene glycol) show prolonged circulation half-lives in vivo, *Biochim. Biophys. Acta*, 1066, 29–36.
61. Woodle, M. C., and Lasic, D. D. (1992), Sterically stabilized liposomes, *Biochim. Biophys. Acta*, 1113, 171–199.
62. Dunn, S. E., Brindley, A., Davis, S. S., Davies, M. C., and Illum, L. (1994), Polystyrene-poly (ethylene glycol) (PS-PEG2000) particles as model systems for site specific drug delivery. 2. The effect of PEG surface density on the in vitro cell interaction and in vivo biodistribution, *Pharm. Res.*, 11, 1016–1022.
63. Gref, R., Minamitake, Y., Peracchia, M. T., Trubetskoy, V., Torchilin, V., and Langer, R. (1994), Biodegradable long-circulating polymeric nanospheres, *Science*, 263, 1600–1603.
64. Wheeler, J. J., Wong, K. F., Ansell, S. M., Masin, D., and Bally, M. B. (1994), Polyethylene glycol modified phospholipids stabilize emulsions prepared from triacylglycerol, *J. Pharm. Sci.*, 83, 1558–1564.
65. Tirosh, O., Barenholz, Y., Katzhendler, J., and Priev, A. (1998), Hydration of polyethylene glycol-grafted liposomes, *Biophys. J.*, 74, 1371–1379.
66. Papahadjopoulos, D., Allen, T. M., Gabizon, A., Mayhew, E., Matthay, K., Huang, S. K., Lee, K.-D., Woodle, M. C., Lasic, D. D., Redemann, C., and Martin, F. J. (1991), Sterically stabilized liposomes: Improvements in pharmacokinetics and antitumor therapeutic efficacy, *Proc. Nat. Acad. Sci. U. S. A.*, 88, 11460–11464.
67. Lundberg, B. B., Mortimer, B.-C., and Redgrave, T. G. (1996), Submicron lipid emulsions containing amphipathic polyethylene glycol for use as drug-carriers with prolonged circulation time, *Int. J. Pharm.*, 134, 119–127.
68. Allen, T. M., and Everest, J. M. (1983), Effect of liposome size and drug release properties on pharmacokinetics of encapsulated drug in rats, *J. Pharmacol. Exp. Ther.*, 226, 539–544.
69. Senior, J., Crawley, J. C. W., and Gregoriadis, G. (1985), Tissue distribution of liposomes exhibiting long half-lives in the circulation after intravenous injection, *Biochim. Biophys. Acta*, 839, 1–8.
70. Liu, D., Mori, A., and Huang, L. (1991), Large liposomes containing ganglioside GM1 accumulate effectively in spleen, *Biochim. Biophys. Acta*, 1066, 159–165.
71. Connelly, P. W., and Kuksis, A. (1981), Effect of core composition and particle size of lipid emulsions on apolipoprotein transfer of plasma lipoproteins in vivo, *Biochim. Biophys. Acta*, 666, 80–89.
72. Bundgaard, M. (1980), Transport pathways in capillaries—In search of pores, *Annu. Rev. Physiol.*, 42, 325–336.
73. Wisse, E. (1970), An electron microscopic study of the fenestrated endothelial lining of rat liver sinusoids, *J. Ultrastruct. Res.*, 31, 125–150.
74. Jain, R. K. (1987), Transport of molecules across tumor vasculature, *Cancer Metastasis Rev.*, 6, 559–593.
75. Takakura, Y., and Hashida, M. (1995), Macromolecular drug carrier systems in cancer chemotherapy: Macromolecular prodrugs, *Crit. Rev. Oncol./Hematol.*, 18, 207–231.

76. Gabizon, A., Price, D. C., Huberty, J., Bresalier, R. S., and Papahadjopoulos, D. (1990), Effect of liposome composition and other factors on the targeting of liposomes to experimental tumors: Biodistribution and imaging studies, *Cancer Res.*, 50, 6371–6378.
77. Wu, N. Z., Da, D., Rudoll, T. L., Needham, D., Whorton, A. R., and Dewhirst, M. W. (1993), Increased microvascular permeability contributes to preferential accumulation of Stealth liposomes in tumor tissue, *Cancer Res.*, 53, 3765–3770.
78. Yuan, F., Dellian, M., Fukumura, D., Leunig, M., Berk, D. A., Torchilin, V. P., and Jain, R. K. (1995), Vascular permeability in a human tumor xenograft: Molecular size dependence and cutoff size, *Cancer Res.*, 55, 3752–3756.
79. Nomura, T., Yamashita, F., Takakura, Y., and Hashida, M. (1995), *Proc. Int. Symp. Control Release Bioactive Mater.*, 22, 420–421.
80. Yuan, F., Leunig, M., Huang, S. K., Berk, D. A., Papahadjopoulos, D., and Jain, R. K. (1994), Microvascular permeability and interstitial penetration of sterically stabilized (stealth) liposomes in a human tumor xenograft, *Cancer Res.*, 54, 3352–3356.
81. Gabizon, A., Shmeeda, H., Horowitz, A. T., and Zalipsky, S. (2004), Tumor cell targeting of liposome-entrapped drugs with phospholipid-anchored folic acid-PEG conjugates, *Adv. Drug Deliv. Rev.*, 56, 1177–1192.
82. Nobs, L., Buchegger, F., Gurny, R., and Allemann, E. (2004), Current methods for attaching targeting ligands to liposomes and nanoparticles, *J. Pharm. Sci.*, 93, 1980–1992.
83. Song, Y. K., Liu, D., Maruyama, K. Z., and Takizawa, T. (1996), Antibody mediated lung targeting of long-circulating emulsions, *PDA J. Pharm. Sci. Technol.*, 50, 372–377.
84. Lundberg, B. B., Griffiths, G., and Hansen, H. J. (1999), Conjugation of an anti-B-cell lymphoma monoclonal antibody, LL2, to longcirculating drug-carrier lipid emulsions, *J. Pharm. Pharmacol.*, 51, 1099–1105.
85. Lundberg, B. B., Griffiths, G., and Hansen, H. J. (2004), Cellular association and cytotoxicity of anti-CD74-targeted lipid drug-carriers in B lymphoma cells, *J. Controlled Release*, 94, 155–161.
86. Goldstein, D., Nassar, T., Lambert, G., Kadouche, J., and Benita, S. (2005), The design and evaluation of a novel targeted drug delivery system using cationic emulsion-antibody conjugates, *J. Controlled Release*, 108, 418–432.
87. Rensen, P. C. N., Van Dijk, M. C. M, Havenaar, E. C., Bijsterbosch, M. K., Kruijt, J. K., and Van Berkel, T. J. C. (1995), Selective liver targeting of antivirals by recombinant chylomicrons—A new therapeutic approach to hepatitis B, *Nat. Med.*, 1, 221–225.
88. Mahley, R. W. (1988), Apolipoprotein E: Cholesterol transport protein with expanding role in cell biology, *Science*, 240, 622–630.
89. Barenholz, Y., and Cohen, R. (1995), Rational design of amphiphile-based drug carriers and sterically stabilized carriers, *J. Liposome Res.*, 5, 905–932.
90. Iwamoto, K., Kato, T., Kawahara, M., Koyama, N., Watanabe, S., Miyake, Y., and Sunamoto, J. (1991), Polysaccharide-coated oil droplets in oil-in-water emulsions as targetable carriers for lipophilic drugs, *J. Pharm. Sci.*, 80, 219–224.
91. Grolier, P., Azais-Braesco, V., Zelmire, L., and Fessi, H. (1992), Incorporation of carotenoids in aqueous systems: Uptake by cultured rat hepatocytes, *Biochim. Biophys. Acta*, 1111, 135–138.
92. Barratt, G. (2003), Colloidal drug carriers: Achievements and perspectives, *Cell. Mol. Life Sci.*, 60, 21–37.
93. Kim, T. W., Chung, H., Kwon, I. C., Sung, H. C., and Jeong, S. Y. (2001), Optimization of lipid composition in cationic emulsion as in vitro and in vivo transfection agents, *Pharm. Res.*, 18, 54–60.

94. Kim, Y. J., Kim, T. W., Chung, H., Kwon, I. C., Sung, H. C., and Jeong, S. Y. (2001), Counterion effects on transfection activity of cationic lipid emulsion, *Biotechnol. Bioproc. Eng.*, 6, 279–283.
95. Elbaz, E., Zeevi, A., Klang, S., and Benita, S. (1993), Positively-charged submicron emulsion—A new type of colloidal drug carrier, *Int. J. Pharm.*, 96, R1–R6.
96. Guilatt, R. L., Couvreur, P., Lambert, G., Goldstein, D., Benita, S., and Dubernet, C. (2004), Extensive surface studies help to analyse zeta potential data: The case of cationic emulsions, *Chem. Phys. Lipids*, 131, 1–13.
97. Calvo, P., Remuńá-López, C., Vila-Jato, J. L., and Alonso, M. J. (1997), Development of positively charged colloidal drug carriers: Chitosan-coated polyester nanocapsules and submicro-emulsions, *Coll. Polym. Sci.*, 275, 46–53.
98. Jumaa, M., and Müller, B. W. (1999), Physicochemical properties of chitosan-lipid emulsions and their stability during the autoclaving process, *Int. J. Pharm.*, 183, 175–184.
99. Samama, J. P., Lee, K. M., and Biellmann, J. F. (1987), Enzymes and microemulsions. Activity and kinetic properties of liver alcohol dehydrogenase in ionic water-in-oil microemulsions, *Eur. J. Biochem.*, 163, 609–617.
100. Teixeira, H., Dubernet, C., Puisieux, F., Benita, S., and Couvreur, P. (1999), Submicron cationic emulsions as a new delivery system for oligonucleotides, *Pharm. Res.*, 16, 30–36.
101. Choi, B. Y., Chung, J. W., Park, J. H., Kim, K. H., Kim, Y. I., Koh, Y. H., Kwon, J. W., Lee, K. H., Choi, H. J., Kim, T. W., Kim, Y. J., Chung, H., Kwon, I. C., and Jeong, S. Y. (2002), Gene delivery to the rat liver using cationic lipid emulsion/DNA complex: Comparison between intra-arterial, intraportal and intravenous administration, *Korean J. Radiol.*, 3, 194–198.
102. Yi, S. W., Yune, Y., Kim, T. W., Chung, H., Choi, Y. W., Kwon, I. C., Lee, E. B., and Jeong, S. Y. (2000), A cationic lipid emulsion/DNA complex as a physically stable and serum-resistant gene delivery system, *Pharm. Res.*, 17, 314–320.
103. Kim, Y. J., Kim, T. W., Chung, H., Kwon, I. C., Sung, H. C., and Jeong, S. Y. (2003), The effects of serum on the stability and the transfection activity of the cationic lipid emulsion with various oils, *Int. J. Pharm.*, 252, 241–252.
104. Tamilvanan, S. (2004), Oil-in-water lipid emulsions: Implications for parenteral and ocular delivering systems, *Prog. Lipid Res.*, 43, 489–533.
105. Benita, S. (1999), Prevention of topical and ocular oxidative stress by positively charged submicron emulsion, *Biomed. Pharm.*, 53, 193–206.
106. Abdulrazik, M., Tamilvanan, S., Khoury, K., and Benita, S. (2001), Ocular delivery of cyclosporin A II. Effect of submicron emulsion's surface charge on ocular distribution of topical cyclosporin A, *STP Pharma Sci.*, 11, 427–432.
107. Rojanasakul, Y., and Robinson, J. R. (1989), Transport mechanisms of the cornea: Characterization of barrier permselectivity, *Int. J. Pharm.*, 55, 237–246.
108. Klang, S., Abdulrazik, M., and Benita, S. (2000), Influence of emulsion droplet surface charge on indomethacin ocular tissue distribution, *Pharm. Dev. Technol.*, 5, 521–532.
109. Wretlind, A. (1981), Parenteral nutrition. *Nutr. Rev.*, 39, 257–265.
110. Davis, S. S. (1982), Emulsions systems for the delivery of drugs by the parenteral route, in Bundgaard, H., Bagger Hansen, A., and Kofod, H., Eds., *Optimization of Drug Delivery*, Munksgaard, Copenhagen, pp. 333–346.
111. Klang, S. H., Parnas, M., and Benita, S. (1998), Emulsions as drug carriers—Possibilities, limitations, and future perspectives, in Muller, R. H., Benita, S., and Bohm, H. L., Eds., *Emulsions and Nanosuspensions for the Formulation of Poorly Soluble drugs*, Medpharm, Stuttgart, pp. 31–65.

112. Yang, S. C., and Benita, S. (2000), Enhanced absorption and drug targeting by positively charged submicron emulsions, *Drug Dev. Res.*, 50, 476–486.
113. Takino, T., Koreeda, N., Nomura, T., Sakaeda, T., Yamashita, F., Takakura, Y., and Hashida, M. (1998), Control of plasma cholesterol-lowering action of probucol with various lipid carrier systems, *Biol. Pharm. Bull.*, 21, 492–497.
114. Klang, S. H, Frucht-Pery, J., Hoffman, A., and Benita, S. (1994), Physicochemical characterization and acute toxicity evaluation of a positively-charged submicron emulsion vehicle, *J. Pharm. Pharmacol.*, 46, 986–993.
115. Korner, D., Benita, S., Albrecht, G., and Baszkin, A. (1994), Surface properties of mixed phospholipid–stearylamine monolayers and their interaction with a non-ionic surfactant (poloxamer), *Coll. Surf. B BioInterf.*, 3, 101–109.
116. Benita, S., and Levy, M. Y. (1993), Submicron emulsions as colloidal drug carriers for intravenous administration: Comprehensive physicochemical characterization, *J. Pharm. Sci.*, 82, 1069–1079.
117. Tamilvanan, S., and Benita, S. (2004), The potential of lipid emulsion for ocular delivery of lipophilic drugs, *Eur. J. Pharm. Biopharm.*, 58, 357–368.
118. Lee, V. H. L., and Robinson, J. R. (1986), Review: Topical ocular drug delivery: Recent developments and future challenges, *J. Ocul. Pharmacol.*, 2, 67–108.
119. Jumaa, M., and Müller, B. W. (1998), The effect of oil components and homogenization condition on the physicochemical properties and stability of parenteral fat emulsions, *Int. J. Pharm.*, 163, 81–89.
120. Constantinides, P. P., Tustian, A., and Kessler, D. R. (2004), Tocol emulsions for drug solubilization and parenteral delivery, *Adv. Drug Deliv. Rev.*, 56, 1243–1255.
121. Constantinides, P. P., Han, J., and Davis, S. S. (2006), Advances in the use of tocols as drug delivery vehicles, *Pharm. Res.*, 23, 243–255.
122. Lundberg, B. B. (1997), A submicron lipid emulsion coated with amphiphathic polyethylene glycol for parenteral administration of paclitaxel (taxol), *J. Pharm. Pharmacol.*, 49, 16–21.
123. Han, J., Davis, S. S., Papandreou, C., Melia, C. D., and Washington, C. (2004), Design and evaluation of an emulsion vehicle for paclitaxel. I. Physicochemical properties and plasma stability, *Pharm. Res.*, 21, 1573–1580.
124. Muchtar, S., and Benita, S. (1994), Emulsions as drug carriers for ophthalmic use, *Coll. Surf. A Physicochem. Eng. Aspects*, 91, 181–190.
125. Ogawa, S., Decker, E. A., and McClements, D. J. (2002), Production and characterization of o/w emulsions containing cationic droplets stabilized by lecithin-chitosan membranes, *J. Agric. Food Chem.*, 51, 2606–2812.
126. Sonne, M. R. (2001), Tocopherol compositions for delivery of biologically active agents, USPTO Patent Full-Text and Image Database No. 6,193,985, Dumex, Copenhagen, pp. 1–21.
127. Sznitowska, M., Janicki, S., Dabrowska, E. A., and Gajewska, M. (2002), Physicochemical screening of antimicrobial agents as potential preservatives for submicron emulsions, *Eur. J. Pharm. Sci.*, 15, 489–495.
128. Jumaa, M., Furkert, F. H., and Müller, B. W. (2002), A new lipid emulsion formulation with high antimicrobial efficacy using chitosan, *Eur. J. Pharm. Biopharm.*, 53, 115–123.
129. Pongcharoenkiat, N., Narsimhan, G., Lyons, R. T., and Hem, S. L. (2002), The effect of surface charge and partition coefficient on the chemical stability of solutes in o/w emulsions, *J. Pharm. Sci.*, 91, 559–570.
130. Han, J., and Washington, C. (2005), Partition of antimicrobial additives in an intravenous emulsion and their effect on emulsion physical stability, *Int. J. Pharm.*, 288, 263–271.

131. Swietlikowska, D. W., and Sznitowska, M. (2006), Partitioning of parabens between phases of submicron emulsions stabilized with egg lecithin, *Int. J. Pharm.*, 312, 174–178.

132. Myers, R. A., and Stella, V. J. (1992), Systemic bioavailability of penclomedine (NSC-338720) from oil-in-water emulsion administered intraduodenally to rats, *Int. J. Pharm.*, 78, 217–226.

133. Ilan, E., Amselem, S., Weisspapir, M., Schwartz, J., Yogev, A., Zawoznik, E., and Friedman, D. (1996), Improved oral delivery of desmopressin via a novel vehicle: Mucoadhesive submicron emulsion, *Pharm. Res.*, 13, 1083–1087.

134. Palin, K. J., Phillips, A. J., and Ning, A. (1986), The oral absorption of cefaxitin from oil and water emulsion vehicles in rats, *Int. J. Pharm.*, 33, 99–104.

135. Kimura, T., Takeda, K., Kageyn, A., Toda, M., Kurpsaki, Y., and Nkayama, T. (1989), Intestinal absorption of dolichol from emulsions and liposomes in rats, *Chem. Pharm. Bull.*, 37, 463–466.

136. Charman, S. A., Charman, W. N., Rogge, M. C., Wildon, T. D., Dutko, F. J., and Pouton, C. W. (1992), Self-emulsifying drug delivery systems: Formulation and biopharmaceutical evaluation of an investigational lipophilic compound, *Pharm. Res.*, 9, 87–93.

137. Gursoy, R. N., and Benita, S. (2004), Self-emulsifying drug delivery systems (SEDDS) for improved oral delivery of lipophilic drugs, *Biomed. Pharmacother. Dossier: Drug Deliv. Drug Efficacy*, 58, 173–182.

138. Takino, T., Nagahama, E., Sakaeda (nee Kakutani), T., Yamashita, F., Takakura, Y., and Hashida, M. (1995), Pharmacokinetic disposition analysis of lipophilic drugs injected with various lipid carriers in the single-pass rat liver perfusion system, *Int. J. Pharm.*, 114, 43–54.

139. Tamilvanan, S., Schmidt, S., Muller, R. H., and Benita, S. (2005), In vitro adsorption of plasma proteins onto the surface (charges) modified-submicron emulsions for intravenous administration, *Eur. J. Pharm. Biopharm.*, 59, 1–7.

140. Alany, R. G., Rades, T., Nicoll, J., Tucker, I. G., and Davies, N. M. (2006), W/O microemulsions for ocular delivery: Evaluation of ocular irritation and precorneal retention, *J. Controlled Release*, 111, 145–152.

141. Stjernschantz, J., and Astin, M. (1993), Anatomy and physiology of the eye: Physiological aspects of ocular drug therapy, in Edman, P., Ed., *Biopharmaceutics of Ocular Drug Delivery*, CRC Press, Boca Raton, FL, pp. 1–25.

142. Lee, V. H. L. (1993), Precorneal, corneal and postcorneal factors, in Mitra, A. K. Ed., *Ophthalmic Drug Delivery System*, Marcel Dekker, New York, pp. 59–81.

143. Järvinen, K., Järvinen, T., and Urtti, A. (1995), Ocular absorption following topical delivery, *Adv. Drug Deliv. Rev.*, 16, 3–19.

144. Prausnitz, M. R., and Noonan, J. S. (1998), Permeability of cornea, sclera and conjunctiva: A literature analysis for drug delivery to eye, *J. Pharm. Sci.*, 87, 1479–1488.

145. Washington, N., Washington, C., and Wilson, C. (2001), Ocular drug delivery, in Washington, N., Washington, C., and Wilson, C., Eds., *Physiological Pharm., Barriers to Drug Absorption*, 2nd ed., Taylor & Francis, New York, pp. 250–270.

146. Ludwig, A. (2005), The use of mucoadhesive polymers in ocular drug delivery, *Adv. Drug Deliv. Rev.*, 57, 1595–1639

147. Chien, D. S., Homsy, J. J., Gluchowski, C., and Tang-Liu, D. D. (1990), Corneal and conjunctival/scleral penetration of *p*-aminoclonidine, AGN 190342, and clonidine in rabbit eyes, *Curr. Eye Res.*, 9, 1051–1059.

148. Ahmed, I., and Patton, T. F. (1985), Importance of the noncorneal absorption route in topical ophthalmic drug delivery, *Investigative Ophthalmol. Vis. Sci.*, 26, 584–587.

149. Kaur, I. P., and Smitha, R. (2002), Penetration enhancers and ocular bioadhesives: Two new avenues for ophthalmic drug delivery, *Drug Dev. Ind. Pharm.*, 28, 353–369.

150. Dey, S., Anand, B. S., Patel, J., and Mitra, A. K. (2003), Transporters/receptors in the anterior chamber: Pathways to explore ocular drug delivery strategies, *Expert Opin. Biol. Ther.*, 3, 23–44.

151. Beilin, M., Bar-Ilan, A., and Amselem, S. (1995), Ocular retention time of submicron emulsion (SME) and the miotic response to pilocarpine delivered in SME. *Investigative Ophthalmol. Vis. Sci.*, 36, S166.

152. Aiache, J. M., el Meski, S., Beyssac, E., and Serpin, G. (1997), The formulation of drug for ocular administration, *J. Biomater. Appl.*, 11, 329–348.

153. Maurice, D. M., and Mishima, S. (1984), Ocular pharmacokinetics, in Sears, M. L., Ed., *Pharmacology. of the Eye*, Springer Verlag, New York, pp. 20–116.

154. Muchtar, S., Almog, S., Torracca, M. T., Saettone, M. F., and Benita, S. (1992), A submicron emulsion as ocular vehicle for delta-8-tetrahydrocannabinol: Effect on intraocular pressure in rabbits, *Ophthalm. Res.*, 24, 142–149.

155. Naveh, N., Weissman, C., Muchtar, S., Benita, S., and Mechoulam, R. (2000), A submicron emulsion of HU-211, a synthetic cannabinoid, reduces intraocular pressure in rabbits, *Graefe's Arch. Clin. Exp. Ophthalmol.*, 238, 334–338.

156. Naveh, N., Muchtar, S., and Benita, S. (1994), Pilocarpine incorporated into a submicron emulsion vehicle causes an unexpectedly prolonged ocular hypotensive effect in rabbits, *J. Ocul. Pharmacol.*, 10, 509–520.

157. Zurowska-Pryczkowska, K., Sznitowska, M., and Janicki, S. (1999), Studies on the effect of pilocarpine incorporation into a submicron emulsion on the stability of the drug and the vehicle, *Eur. J. Pharm. Biopharm.*, 47, 255–260.

158. Sznitowska, M., Janicki, S., Zurowska-Pryczkowska, K., and Mackiewicz, J. (2001), In vivo evaluation of submicron emulsions with pilocarpine: The effect of pH and chemical form of the drug, *J. Microencapsul.*, 18, 173–181.

159. Calvo, P., Alonso, M. J., Vila-Jato, J. L., and Robinson, J. R. (1996), Improved ocular bioavailability of indomethacin by novel ocular drug carriers, *J. Pharm. Pharmacol.*, 48, 1147–1152.

160. Calvo, P., Vila-Jato, J. L., and Alonso, M. J. (1996), Comparative in vitro evaluation of several colloidal systems, nano-particles, nanocapsules, and nanoemulsions as ocular drug carriers, *J. Pharm. Sci.*, 85, 530–536.

161. Anselem, S., Beilin, M., and Garty, N. (1993), Submicron emulsion as ocular delivery system for adaprolol maleate, a soft β-blocker, *Pharm. Res.*, 10, S205.

162. Melamed, S., Kurtz, S., Greenbaum, A., Haves, J. F., Neumann, R., and Garty, N. (1994), Adaprolol maleate in submicron emulsion, a novel soft β-blocking agent, is safe and effective in human studies, *Investigative Ophthalmol. Vis. Sci.*, 35, 1387–1390.

163. Aviv, H., Friedman, D., Bar-Ilan, A., and Vered, M. (1995), Submicron emulsions as ocular drug delivery vehicles, U.S. Patent 5,496,811, March 5, 1996.

164. Garty, N., Lusky, M., Zalish, M., Rachmiel, R., Greenbaum, A., Desatnik, H., Neumann, R., Howes, J. F., and Melamed, S. (1994), Pilocarpine in submicron emulsion formulation for treatment of ocular hypertension: A phase II clinical trial, *Investigative Ophthalmol. Vis. Sci.*, 35, 2175.

165. Ding, S., Tien, W., and Olejnik, O. (1995), Nonirritating emulsions for sensitive tissue, U.S. Patent 5,474,979 (to Allergan), December 12.

166. Ding, S., and Olejnik, O. (1997), Cyclosporin ophthalmic oil/water emulsions. Formulation and characterization, *Pharm. Res.*, 14, S41.

167. Acheampong, A. A., Shackleton, M., Tang-Liu, DD-S., Ding, S., Stern, M. E., and Decker, R. (1999), Distribution of cyclosporin A in ocular tissues after topical administration to albino rabbits and beagle dogs, *Curr. Eye Res.*, 18, 91–103.
168. Stevenson, D., Tauber, J., and Reis, B. L. (2000), Efficacy and safety of cyclosporin A ophthalmic emulsion in the treatment of moderate-to-severe dry eye disease: A dose-ranging, randomized trial. The Cyclosporin A Phase 2 Study Group, *Ophthalmology*, 107, 967–974.
169. Sall, K., Stevenson, O. D., Mundorf, T. K., and Reis, B. L. (2000), Two multicenter, randomized studies of the efficacy and safety of cyclosporin ophthalmic emulsion in moderate-to-severe dry eye disease. CsA Phase 3 Study Group, *Ophthalmology*, 107, 631–639.
170. Turner, K., Pflugfelder, S. C., Ji, Z., Feuer, W. J., Stern, M., and Reis, B. L. (2000), Interleukin-6 levels in the conjunctival epithelium of patients with dry eye disease treated with cyclosporine ophthalmic emulsion, *Cornea*, 19, 492–496.
171. Kunert, K. S., Tisdale, A. S., Stern, M. E., Smith, J. A., and Gipson, I. K. (2000), Analysis of topical cyclosporin treatment of patients with dry eye syndrome: Effect on conjunctival lymphocytes, *Arch. Ophthalmol.*, 118, 1489–1496.
172. Brignole, F., Pisella, P. J., Saint-Jean, M. De., Goldschild, M., Goguel, A., and Baudouin, C. (2001), Flow cytometric analysis of inflammatory markers in KCS: 6-month treatment with topical cyclosporin A. *Investigative Ophthalmol. Vis. Sci.*, 42, 90–95.
173. Small, D. S., Acheampong, A., Reis, B., Stern, K., Stewart, W., Berdy, G., Epstein, R., Foerster, R., Forstot, L., and Tang-Liu, D. D.-S. (2002), Blood concentrations of cyclosporin A during long-term treatment with cyclosporin A ophthalmic emulsions in patients with moderate to severe dry eye disease, *J. Ocul. Pharmacol. Ther.*, 18, 411–418.
174. Galatoire, O., Baudouin, C., Pisella, P. J., and Brignole, F. (2003), Flow cytometry in impression cytology during keratoconjunctivitis sicca: Effects of topical cyclosporin A on HLA DR expression, *J. Francais d'Ophtalmol.*, 26, 337–343.
175. Tang-Liu, D. D., and Acheampong, A. (2005), Ocular pharmacokinetics and safety of ciclosporin, a novel topical treatment for dry eye, *Clin. Pharmacokinet.*, 44, 247–261.
176. Hom, M. M. (2006), Use of cyclosporine 0.05% ophthalmic emulsion for contact lens-intolerant patients, *Eye Contact Lens: Sci. Clin. Pract.*, 32, 109–111.
177. Sall, K. N., Cohen, S. M., Christensen, M. T., and Stein, J. M. (2006), An evaluation of the efficacy of a cyclosporine-based dry eye therapy when used with marketed artificial tears as supportive therapy in dry eye, *Eye Contact Lens: Sci. Clin. Pract.*, 32, 21–26.
178. Klang, S. H., Siganos, C. S., and Benita, S. (1999), Evaluation of a positively-charged submicron emulsion of piroxicam in the rabbit corneum healing process following alkali burn, *J. Controlled Release*, 57, 19–27.
179. Tamilvanan, S., Khoury, K., Gilhar, D., and Benita, S. (2001), Ocular delivery of cyclosporin A. I. Design and characterization of cyclosporin A-loaded positively charged submicron emulsion, *STP Pharma Sci.*, 11, 421–426.
180. Hussain, A. A. (1998), Intranasal drug delivery, *Adv. Drug Deliv. Rev.*, 29, 39–49.
181. Kararli, T. T., Needham, T. E., Schoenhard, G., Baron, D. A., Schmidt, R. E., Katz, B., and Belonio, B. (1992), Enhancement of nasal delivery of a rennin inhibitor in the rat using emulsion formulations, *Pharm. Res.*, 9, 1024–1028.
182. Ko, K. Y., Needham, T. E., and Zia, H. (1998), Emulsion formulations of testosterone for nasal administration, *J. Microencapsul.*, 15, 197–205.
183. Aikawa, K., Matsumoto, K., Uda, H., Tanaka, S., Shimamura, H., Aramaki, Y., and Tsuchiya, S. (1998), Prolonged release of drug from o/w emulsion and residence in rat nasal cavity, *Pharm. Dev. Technol.*, 3, 461–469.

184. Mitra, R., Pezron, I., Chu, W. A., and Mitra, A. K. (2000), Lipid emulsions as vehicles for enhanced nasal delivery of insulin, *Int. J. Pharm.*, 205, 127–134.
185. Kim, T. W., Chung, H., Kwon, I. C., Sung, H. C., and Jeong, S. Y. (2000), In vivo gene transfer to the mouse nasal cavity mucosa using a stable cationic lipid emulsion, *Mol. Cells*, 10, 142–147.
186. Jenkins, P. (1999), Mucosal vaccine delivery, *Exp. Opin. Ther. Patients*, 9, 255–262.
187. Benita, M. B., Oudshoorn, M., Romeijn, S., Meijgaarden, K. V., Koerten, H., Meulen, H. V. D, Lambert, G., Ottenhoff, T., Benita, S., Junginger. H., and Borchard, G. (2004), Cationic submicron emulsions for pulmonary DNA immunization, *J. Controlled Release*, 100, 145–155.
188. El-Hameed, M. D. A., and Kellaway, I. W. (1997), Preparation and in vitro characterization of mucoadhesive polymeric microspheres as intranasal delivery systems, *Eur. J. Pharm. Biopharm.*, 44, 53–60.
189. Friedman, D. I., Schwarz, J. S., and Weisspapir, M. (1995), Submicron emulsion vehicle for enhanced transdermal delivery of steroidal and nonsteroidal antiinflammatory drugs, *J. Pharm. Sci.*, 84, 324–329.
190. Schwarz, J. S., Weisspapir, M. R., and Friedman, D. I. (1995), Enhanced transdermal delivery of diazepam by submicron emulsion (SME) creams, *Pharm. Res.*, 12, 687–692.
191. Realdaon, N., and Ragazzi, E. (1998), Study of drug availability from o/w emulsions obtained in different manufacturing conditions, in Proceedings of 2nd World Meeting APGI/APV, 775–776.
192. Welin-Berger, K., and Bergenstahl, B. (2000), Inhibition of Ostwald ripening in local anesthetic emulsions—By using hydrophobic excipients in the disperse phase, *Int. J. Pharm.*, 200–202, 249–260.
193. Welin-Berger, K., Neelissen, J. A., and Bergenstahl, B. (2001), The effect of rheological behaviour of a topical anaesthetic formulation on the release and permeation rates of the active compound, *Eur. J. Pharm. Sci.*, 13, 309–318.
194. Dutta, S. (1999), Use of eutectic mixture of local anesthetics in children, *Ind. J. Pediatr.*, 66, 707–715.
195. Cordoni, A., and Cordoni, L. E. (2001), Eutectic mixture of local anesthetics reduces pain during intravenous catheter insertion in the pediatric patient, *Clin. J. Pain*, 17, 115–118.
196. Ezra, R., Benita, S., Ginsberg, I., and Kohen, R. (1996), Prevention of oxidative damage in fibroblast cell cultures and rat skin by positively-charged submicron emulsion of a-tocopherol, *Eur. J. Pharm. Biopharm.*, 42, 291–298.
197. Piemi, M. P., Korner, D., Benita, S., and Marty, J. P. (1999), Positively and negatively charged submicron emulsions for enhanced topical delivery of antifungal drugs, *J. Controlled Release*, 58, 177–187.

# GOOD MANUFACTURING PRACTICES (GMP) AND RELATED FDA GUIDELINES

James R. Harris
*James Harris Associates, Inc., Durham, North Carolina*

**Contents**

1 FDA Regulations: Real and Imagined
2 21 CFR 210 and 211: Current Good Manufacturing Practice for Finished Pharmaceuticals
3 Guidance for Industry: Quality Systems Approach to Pharmaceutical Current Good Manufacturing Practice Regulations
  3.1 CGMPS and the Concepts of Modern Quality Systems
  3.2 Quality Systems Model
4 Guidance for Industry: PAT—Framework for Innovative Pharmaceutical Development, Manufacturing, and Quality Assurance
  4.1 PAT Framework
5 Guidance for Industry: Part 11. Electronic Records; Electronic Signatures—Scope and Application
6 Guidance for Industry and FDA: Current Good Manufacturing Practice for Combination Products
7 Guidance for Industry: Powder Blends and Finished Dosage Units—Stratified In-Process Dosage Unit Sampling and Assessment
  7.1 Validation of Batch Powder Mix Homogeneity
  7.2 Verification of Manufacturing Criteria
8 Guidance for Industry: Immediate-Release Solid Oral Dosage Forms Scale-Up and Post-approval Changes (SUPAC)—Chemistry, Manufacturing and Controls, In Vitro Dissolution Testing, and In Vivo Bioequivalence Documentation
9 Other GMP-Related Guidance Documents

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

## 1 FDA REGULATIONS: REAL AND IMAGINED

A regulation is a law. In the United States, all federal laws have been arranged or codified in a manner that makes it easier to find a specific law. The *Code of Federal Regulations* (CFR) is a compilation of all federal laws published in the *Federal Register* by the executive departments and agencies of the federal government. This code is divided into 50 titles which represent broad areas of federal regulation. Each title is further divided into chapters. The chapters are then subdivided into parts covering specific regulatory areas. Changes and additions are first published in the *Federal Register*. Both the coded law and the *Federal Register* must be used to determine the latest version of any rule. All food- and drug-related laws are contained in Title 21 of the CFR. Each title of the CFR is updated annually. Title 21 is updated as of April 1 of each year.

Because virtually all of the drug regulations are written to state what should be done but do not tell how to do it, the Food and Drug Administration (FDA) also publishes guidance documents. These documents are intended to provide precisely what the name implies—guidance. In this context, *guidance documents are not law and do not bind the FDA or the public*. Manufacturers are not required to use the techniques or approaches appearing in the guidance document. In fact, FDA representatives have repeatedly stated that the regulations were not written to suggest how something should be done in order to encourage innovation. While following the recommendations contained in the guidance documents will probably assure acceptance (agency philosophy and interpretation may have changed since the guidance document was published), other approaches are encouraged. No matter how they choose to proceed, manufacturers should be prepared to show that their methods achieve the desired results.

A method used by the FDA to "float" new ideas is to discuss them at industry gatherings such as FDA-sponsored seminars or meetings of industry groups such as the Pharmaceutical Manufacturers Association (PMA), the Parenteral Drug Association (PDA), and the International Society of Pharmaceutical Engineering (ISPE). Again, it must be remembered that while these comments reflect current FDA thinking, they are simply thoughts and recommendations. They are not law.

Several industry groups also publish comments, guidelines, and so on, that put forth current thinking of the group writing the document. These publications are interesting and often bring out valuable information. However, it is important to remember that *these publications are not regulations or even official guidance documents.* If a firm chooses to follow the recommendations of such documents, they are probably following good advice. However, since the advice comes from a nonofficial source, firms should still be prepared to defend their actions with good scientific reasoning.

## 2 21 CFR 210 AND 211: CURRENT GOOD MANUFACTURING PRACTICE FOR FINISHED PHARMACEUTICALS

Parts 210 and 211 of CFR Title 21 are the laws defining good manufacturing practices for finished pharmaceutical products. All manufacturers must follow these regulations in order to market their products in the United States. When a firm files an application to market a product in the United States through a New Drug Application (NDA), abbreviated NDA, (ANDA), Biological License Application (BLA),

or other product application, one of the last steps in approving the application is a preapproval inspection of the manufacturing facility. A major purpose of this inspection is to assure adherence to the GMP regulations. Preapproval inspections are a part of every application approval. Thus, if a firm has 10 applications pending, it should expect 10 inspections. The fact that the manufacturing facility has already been inspected will not alter the need for another inspection.

The FDA also has the right to visit and inspect any manufacturing facility that produces a product or products sold in the United States. Such inspections are unannounced. A manufacturer must admit an inspector when he or she appears at that facility and must do so without undue delay.

*GMP requirements for manufacturers of pharmaceutical dosage forms* are discussed below. This information should not be considered to be an exact statement of the law. We have attempted to show intent and, occasionally, add some comments that will clarify how that particular regulation is interpreted. For precise wording of a regulation, refer to the CFR and then check the *Federal Register* to determine if there have been any changes since the last update.

### *General Provisions*

1. This section pertains to the manufacture of drug products for humans or animals.
2. These requirements will not be enforced for over-the-counter (OTC) drug products if the products and all their ingredients are ordinarily marketed and considered as human foods and which products may also fall within the legal definition of drugs by virtue of their intended use.

### *Organization and Personnel*

1. Responsibilities of quality control unit
   (a) A quality control unit *must* be a part of the facility organization.
   (b) This unit must be given responsibility and authority to approve or reject all components, drug product containers, closures, process materials, packaging material, labeling, and drug products, and the authority to review production records.
   (c) Adequate laboratory facilities for testing and approval or rejection of the above listed materials must be available.
   (d) The quality control unit is responsible for approving or rejecting all procedures or specifications that impact on the identity, strength, quality, and purity of the drug product.
   (e) Responsibilities and procedures applicable to the quality control unit must be written and these procedures must be followed.
2. Personnel qualifications
   (a) Every person involved in the manufacture, processing, packing, or holding of a drug product must have education, training, and experience that enable that individual to perform their duties. Employees must be trained in the particular operations that they perform and in Current GMPs (CGMPs). The GMP training must be conducted by qualified individuals and with sufficient frequency to assure that workers remain familiar with the requirements applicable to them.

(b) Persons responsible for supervision must have the education, training, and experience to perform their assigned functions in such a manner as to assure that the drug product has the safety, identity, strength, quality, and potency that it is represented to possess.
(c) There must be an adequate number of qualified personnel to perform the needed tasks.

3. Personnel responsibilities
   (a) Personnel shall wear clean clothing appropriate for the duties they perform. Protective apparel must be worn as necessary.
   (b) Personnel shall practice good sanitation and health habits.
   (c) Only personnel authorized by supervisory personnel shall enter those areas designated as limited-access areas.
   (d) Any worker considered to have an apparent illness or open lesions that may adversely affect safety or quality of drug products shall be excluded from direct contact with product, components, or containers.
4. Consultants that advise on the manufacture, processing, packing, or holding of drug products must have sufficient education, training, and experience to advise on the subject for which they are retained. The manufacturer must maintain records of name, address, and qualifications of any consultants and the type of service they provide.

### *Buildings and Facilities*

1. Design and construction features
   (a) Buildings should be of suitable size, construction location to facilitate cleaning, maintenance, and proper operations.
   (b) Space should be adequate for the orderly placement of equipment and materials to prevent mix-ups between different components, drug product containers and closures, labeling, in-process materials, or drug products and to prevent contamination.
   (c) The movement of components and product through the building must be designed to prevent contamination.
   (d) Operations should be performed within specifically defined areas having adequate control systems to prevent contamination or mix-ups during each of the following procedures:
      (i) Receipt, identification, storage, and withholding from use of components, drug product containers, closures, and labeling, pending the appropriate sampling, testing, and release for manufacturing or packaging.
      (ii) Holding rejected materials listed in (a) above.
      (iii) Storage of released components, drug product containers, closures, and labeling.
      (iv) Storage of in-process materials.
      (v) Manufacturing and processing operations.
      (vi) Packaging and labeling operations.
      (vii) Quarantine storage before release of drug products.
      (viii) Storage of drug products after release.
      (ix) Control and laboratory operations.

(x) Aseptic processing, which includes:
  (1) Floors, walls, and ceilings of smooth, hard surfaces that are easily cleanable.
  (2) Temperature and humidity controls.
  (3) An air supply filtered through High-Efficiency Particulate Air (HEPA) filters under positive pressure regardless of whether flow is laminar or nonlaminar.
  (4) A system for monitoring environmental conditions.
  (5) A system for cleaning and disinfecting the room and equipment to produce aseptic conditions.
  (6) A system for maintaining any equipment used to control the aseptic conditions.

(e) Operations relating to the manufacture, processing, and packing of penicillin must be performed in facilities separate from those used for other drug products for humans. *Note*: For all purposes of these GMP regulations, the FDA considers cephalosporins to be penicillin.

2. Adequate lighting should be provided in all areas.
3. Heating, ventilation, and air conditioning (HVAC)
   (a) Adequate ventilation is required in all areas.
   (b) Equipment for adequate control over air pressure, microorganisms, dust, humidity, and temperature must be provided when appropriate for the manufacture, processing, packing, or holding of a drug product.
   (c) When appropriate, air supplied to production areas should be filtered to avoid any possibility of contamination or cross-contamination.
   (d) Air-handling systems for the manufacture, processing, and packing of penicillin shall be completely separate from those for other drug products for humans.
4. Plumbing
   (a) Potable water should be supplied in a continuous positive-pressure system free from defects that could contribute to contamination of any drug product.
   (b) Potable water must meet the standards prescribed in the Environmental Protection Agency (EPA) Primary Drinking Water Regulations defined in 40 CFR Part 141.
   (c) Drainage must be of adequate size. Where connected directly to a sewer, an air break or other suitable mechanical device must be provided to prevent back-siphonage.
5. Sewage, trash, and other refuse in and from the building and immediate premises must be disposed of in a safe and sanitary manner.
6. Adequate washing facilities should be provided. This is to include hot and cold water, soap or detergent, air driers or single-service towels, and clean toilet facilities easily accessible to all work areas.
7. Sanitation
   (a) Any building used for manufacture, processing, packing, or holding of a drug product should be maintained in a clean and sanitary condition. Such buildings should be free of infestation by rodents, birds, insects, and other vermin.
   (b) Trash and organic waste matter should be held and disposed of in a timely and sanitary manner.

(c) Written procedures assigning responsibility for sanitation and describing in sufficient detail the cleaning schedules, methods, equipment, and materials to be used in cleaning the buildings and facilities are required. Such procedures must be followed.
(d) Written procedures for use of suitable rodenticides, insecticides, fungicides, fumigating agents, and cleaning and sanitizing agents are required and must be followed. These written procedures should be designed to prevent the contamination of equipment, components, product containers, closures, packaging, labeling materials, or drug products. Agent may not be used unless registered and used in accordance with the Federal Insecticide, Fungicide, and Rodenticide Act (7 U.S.C. 135).
(e) All sanitation procedures apply equally to contractors or temporary employees as to regular employees.

8. All buildings used for GMP-related purposes must be maintained in a good state of repair.

***Equipment***

1. Equipment should be of appropriate design, adequate size, and suitably located to facilitate operations for its intended use and for cleaning and maintenance.
2. Equipment construction
   (a) Equipment should be constructed so that surfaces that contact components, in-process materials, or drug products should not be reactive, additive, or absorptive so as to alter the safety, identity, strength, quality, or purity of the drug product beyond official or other established requirements.
   (b) Any substance required for operation such as lubricants or coolants shall not come into contact with drug products, containers, and so on, so as to alter the safety, identity, strength, quality, or purity of the drug product beyond established requirements.
3. Equipment cleaning and maintenance
   (a) Equipment and utensils should be cleaned, maintained, and sanitized at appropriate intervals to prevent malfunctions or contamination that would alter the drug product beyond the official requirements.
   (b) Written procedures must be established and followed for cleaning and maintenance of equipment and utensils used in the processing of a drug product. These procedures must include but are not limited to the following:
      (i) Assignment of responsibility for cleaning and maintaining equipment.
      (ii) Maintenance and cleaning schedules, including sanitizing schedules if appropriate.
      (iii) A sufficiently detailed description of the methods, equipment, and materials used in cleaning and maintenance operations and the methods of disassembling and reassembling equipment as a part of cleaning and maintenance.
      (iv) Removal or obliteration of previous batch identification.
      (v) Protection of clean equipment from contamination prior to use.
      (vi) Inspection of equipment for cleanliness immediately before use.

(vii) Records should be kept of maintenance, cleaning, sanitizing, and inspection of all processing equipment.

4. Automatic, mechanical, and electronic equipment
   (a) All such equipment, including computers or related systems that will perform a function to be used in any GMP-related activity, must be routinely calibrated, inspected, or checked according to a written program designed to assure proper performance. Written records must be maintained for all such activities.
   (b) Appropriate controls should be exercised to assure that changes in master production and control records or other similar records are made only by authorized personnel. Input to and output from such systems should be checked for accuracy.

   A backup file of data entered into a computer-related system must be maintained except where certain data such as calculations performed in connection with laboratory analysis are eliminated by computerization or other automated processes. In this situation, a written record of the program should be maintained along with validation data.
5. Filters for liquid filtration used as a part of the manufacture, processing, or packing of injectable drug products intended for human use must not release fibers into such products. Fiber-releasing filters may not be used unless it is not possible to manufacture the product without the use of such a filter. In this situation, an additional non-fiber-releasing filter of 0.22 μm maximum must be used after the fiber-releasing filtration. Use of an asbestos-containing filter is permissible only upon submission of proof to the appropriate FDA bureau that use of a non-fiber-releasing filter will compromise the safety or effectiveness of the drug product.

### *Control of Components and Drug Product Containers and Closures*

1. General requirements
   (a) There must be written procedures describing in sufficient detail the receipt, identification, storage, handling, sampling, testing, and approval or rejection of product components, containers, and closures. Of course, all such procedures must be followed. It is quite common and even more embarrassing to be cited for not following your own written procedures. *Note:* For the rest of this discussion, the term *components* will mean product ingredients, containers, closures, and so on.
   (b) All components listed above must be handled and stored in a manner that will prevent contamination.
   (c) Bagged or boxed components should be stored off the floor. Spacing should allow cleaning and inspection.
   (d) Every container of components must be identified with a distinctive code or lot number for each receival of that product. Even if the next receival is the same vendor lot number, it must be a new identifying number by the pharmaceutical manufacturer. Each lot must be appropriately identified as to its status (quarantined, approved, or rejected).

2. Receipt and storage of untested components
   (a) Upon receipt each container of components must be visually examined for appropriate labeling and any damage or contamination to the component container.
   (b) Components must be stored under quarantine until they have been tested as appropriate and released for use.
3. Testing and approval or rejection of components
   (a) Each lot of components shall be withheld from use until it has been sampled, tested, and released by the quality control unit.
   (b) Representative samples must be taken from every receival of every component. The number or amount of component to be sampled should be based on component appearance, statistical confidence levels, the past history of the supplier, and the quantity needed to analyze and reserve samples if required.
   (c) Sampling procedures
      (i) The component containers should be cleaned where necessary.
      (ii) The containers should be opened, sampled, and resealed in a manner designed to prevent contamination of the sample and remaining contents of the container.
      (iii) If appropriate, sterile equipment and aseptic sampling techniques should be used.
      (iv) Where sampling is done from various parts of a container, samples should not be composited for testing.
      (v) Containers from which samples have been taken must be marked to show that samples have been removed.
   (d) Examination and testing of samples
      (i) At least one test should be conducted on each lot of component drug product to verify identity.
      (ii) Each component must be tested for conformity with all appropriate written specifications for purity, strength, and quality if an ingredient or for conformity with written specifications for containers or closures.
      (iii) In lieu of the above testing by the manufacturer, a report of analysis may be accepted from the supplier provided that at least one specific identity test is conducted on the component by the manufacturer and provided that the manufacturer has established the reliability of the supplier's analyses through appropriate validation.
      (iv) When appropriate, components should be examined microscopically.
      (v) Each lot of a component that is liable to contamination with dirt, insect infestation, or other extraneous adulterant should be examined against established specifications for such contamination.
      (vi) Each lot of a component that is subject to microbial contamination that is contrary to its intended use should be subjected to microbiological tests before use.
   (e) If a lot of components meets the written specifications, it may be approved and released for use. Any lot of such material that does not meet such specifications must be rejected.
4. Use of approved components (including drug product containers and closures) must be rotated to assure that the oldest approved stock is used first.

5. Components must be retested and/or reexamined after storage for a long period of time or after exposure to the atmosphere, heat, or other condition that might adversely affect the component.
6. Rejected components should be identified and controlled under a quarantine system designed to prevent their use in manufacturing or processing.
7. Containers and closures
   (a) Containers and closures must not be reactive, additive, or absorbent so as to alter the drug beyond established acceptance criteria.
   (b) Container closure systems must provide adequate protection against foreseeable external factors in storage that can cause deterioration or contamination of the product.
   (c) Containers and closures should be clean and, if necessary, sterile and processed to remove pyrogens.
   (d) Standards or specification, methods of testing, and, if appropriate, sterilization and depyrogenation must be written and followed.

***Production and Process Controls***

1. Written procedures and procedure deviations
   (a) Written procedures for production and process control must be written and followed. These procedures should be designed to assure that the drug products have the identity, strength, quality, and purity they are represented to possess. These procedures must include all requirements given below and must be drafted, reviewed, and approved by the affected organizational units and reviewed and approved by the quality control unit.
   (b) When following the above identified procedures, all actions must be documented at the time of performance. Any deviations from the written procedure must be recorded and justified.
2. Charge-in of components—Written production and control procedures must include the following, which are designed to assure that the drug products produced meet all specifications and standards.
   (a) The batch must be formulated with the intent to provide not less than 100% of the labeled amount of active ingredient.
   (b) Components used must be weighed, measured, or subdivided appropriately. If a component is removed from its original container and placed in another, the new container should be identified with the following information:
      (i) Component name and/or item code.
      (ii) Receiving or control number.
      (iii) Weight or measure of material in the new container.
      (iv) Batch or lot number for which the component was dispensed, including its product name, strength, and lot number.
   (c) Weighing, measuring, or subdividing operations for all components must be adequately supervised. Each container of component dispensed to manufacturing must be examined by a second person to assure that:
      (i) The component was released by the quality control unit.
      (ii) The weight or measure is correct as stated in the batch production records.

(iii) The containers are properly identified and contain the quantity stated on the label.

(d) Addition of each component must be performed by one person and verified by a second person.

3. Actual yield and percentage of theoretical yield should be determined at the completion of each appropriate phase of manufacturing, processing, packaging, or holding. These calculations should be performed by one person and independently verified by a second individual.
4. Equipment identification
   (a) All compounding and storage containers, processing lines, and major equipment used during the production of a batch of a drug product must be properly identified at all times to indicate their contents and the phase of processing of the batch.
   (b) Major equipment should be identified by a distinctive identification that shall be recorded in the batch production record to indicate the specific equipment used. In cases where only one of a particular type of equipment exists in a given manufacturing facility, the name of the equipment may be used instead of creating a distinctive identification.
5. Sampling and testing of in-process materials and drug products
   (a) To assure batch uniformity and integrity, it is necessary to write and follow procedures that describe the in-process controls and tests or examinations that will be conducted on samples taken according to procedure. Procedures should be written to monitor the output and to validate the performance of those manufacturing processes that may be responsible for causing variability in the product being manufactured. These control procedures should include but are not limited to the following:
      (i) Tablet or capsule weight variation.
      (ii) Disintegraton time.
      (iii) Adequacy of mixing or blending to assure uniformity and homogeneity.
      (iv) Dissolution time and rate.
      (v) Clarity of solutions.
      (vi) pH of solutions.
   (b) In-process specifications for all characteristics must be consistent with the drug product final specifications and must be developed from previous acceptable product average and process variability data.
   (c) In-process materials should be tested for identity, strength, quality, and purity as appropriate. As a part of the production process, they must be approved for continued use or rejected by the quality control unit before production continues.
   (d) Rejected in-process materials must be identified and controlled under a quarantine system designed to prevent their use in manufacturing operations for which they have been found to be unsuitable.
6. When appropriate, time limits should be established for the completion of each phase of production. The purpose of this is to assure the quality of the drug product. Deviation from the established time limits may be acceptable if this deviation does not compromise the quality of the product. Any deviation must be documented, including the justification for such deviation.

7. Control of microbial contamination
   (a) To prevent the growth of objectionable microorganisms in products not required to be sterile, appropriate written procedures designed to prevent such growth should be written and followed.
   (b) If sterilization is a part of any procedure described in (a) above, this procedure must be validated.
8. Reprocessing
   (a) Written procedures describing any system used to reprocess batches that do not conform to the established standards must be written and followed.
   (b) Reprocessing must not be performed without the review and approval of the quality control unit.

### *Packaging and Labeling Control*

1. Materials examination and usage criteria
   (a) Written procedures describing in detail the receipt, identification, storage, handling, sampling, examination, and/or testing of labeling and packaging materials must be developed, approved, and followed. These materials must be representatively sampled, examined, or tested on receipt and accepted by the quality control unit before use.
   (b) Any materials that do not fully meet acceptance criteria must be rejected to prevent their use.
   (c) Records of each receival of each different label and packaging material must be maintained indicating receipt, examination or testing, and whether accepted or rejected.
   (d) Labels and other labeling materials for each different drug product, strength, dosage form, or quantity of contents must be stored separately with suitable identification. Access to the storage area must be limited to authorized personnel.
   (e) Obsolete and outdated labels, labeling, and other packaging materials must be quarantined and destroyed.
   (f) The use of gang-printed labels for different drug products or different strengths or different net contents is prohibited. The only exception to this rule is if labels from gang-printed sheets are adequately differentiated by size, shape, or color that will prevent mixing of labels.
   (g) If cut labeling is used, packaging and labeling operations must include one or more of the following special control procedures:
      (i) Dedication of a labeling and packaging line to each different strength of each different drug product.
      (ii) Use of appropriate electronic or electromechanical equipment to conduct a 100% examination for correct labeling during or after completion of the finishing operation.
      (iii) Use of visual inspection to conduct a 100% examination for correct labeling. If visual inspection is used, the inspection should be performed by one person and independently verified by a second individual.
   (h) Printing devices on or associated with the manufacturing line used to imprint labeling upon the drug product unit label or case must be monitored to assure

that the printing conforms to the print specified in the batch production record.

2. Issuance of labeling
   (a) Strict control should be exercised over the issuance of labeling for use in drug product labeling operations.
   (b) Labeling materials issued for a batch must be carefully examined for identity and conformity to the labeling specified in the batch production record.
   (c) Procedures should be written and followed for reconciliation of the quantities of labeling issued, used, destroyed, and returned. Procedures should require evaluation of discrepancies found between the number of packages finished and the amount of labeling issued if discrepancies outside narrow preset limits occur. Limits should be established on the basis of historical operating data. Labeling reconciliation is waived for either cut or roll labeling if a 100% examination for correct labeling is performed.
   (d) All excess labeling bearing a lot or control number must be destroyed.
   (e) Returned labeling should be maintained and stored in a manner to prevent mix-ups.
   (f) Written procedures should describe the control procedures used for the issuance of labeling.
3. There must be written procedures designed to assure that correct labels, labeling, and packaging materials are used. These procedures should incorporate the following features:
   (a) Prevention of mix-ups and cross-contamination by physical or spatial separation of operations on other drug products.
   (b) Identification and handling of filled drug product containers that are set aside and held in unlabeled condition for future labeling operations. Such procedures should be designed to prevent mislabeling individual containers, lots, or portions of lots. It is not necessary to apply identification to each individual container, but the procedure should be adequate to determine the name, strength, quantity of contents, and lot or control number of each container.
   (c) Identification of the drug product with a lot or control number that permits determination of the history of the manufacture and control of the batch.
   (d) Examination of packaging and labeling materials for suitability and correctness before issuing for use and before packaging operations. These examinations must be documented in the batch production record.
   (e) Inspection of the packaging and labeling facility immediately before use to assure that all drug products and labeling materials from the previous operation have been removed. Inspection results must be documented in the batch production record.
4. Tamper-evident packaging requirements for OTC human drug products
   (a) An OTC product (with the exception of a dermatological, dentifrice, insulin, or lozenge product) intended for retail sale is considered adulterated or misbranded or both if it is not packaged in a tamper-resistant package.
   (b) Requirements for a tamper-evident package
      (i) With the exceptions listed above, all OTC products must be packaged in a tamper-evident package if the product is accessible to the public while being held for sale. A tamper-evident package must have

one or more indicators or barriers to entry which, if breached or missing, can reasonably be expected to provide visible evidence to consumers that tampering has occurred: A tamper-evident package may involve an immediate container and closure system or a secondary container or carton system or a combination of systems intended to provide a visual indication of package integrity. The tamper-evident feature must be designed to and shall remain intact when handled in a reasonable manner during manufacture, distribution, and retail display.

(ii) In addition to the tamper-evident packaging feature described above, any two-piece hard gelatin capsule covered by this regulation must be produced using an acceptable tamper-evident technology.

(c) Labeling

(i) In order to alert consumers to the specific tamper-evident features used, each retained package of an OTC drug product covered by this regulation is required to bear a statement that:

(1) Identifies all tamper-evident features and any capsule-sealing technologies.

(2) Is prominently placed on the package.

(3) Is so placed that it will be unaffected if the tamper-evident feature of the package is breached or missing.

(ii) If the tamper-evident feature chosen to meet the requirement uses an identifying characteristic, that characteristic is required to be referred to in the labeling statement. For example, the labeling statement on a bottle with a shrink band could say *For your protection, this bottle has an imprinted seal around the neck*.

(d) A manufacturer or packer may request an exemption from the tamper-evident requirement. A request for exemption is required to be submitted in the form of a petition and should be clearly identified on the envelope as a "Request for Exemption from the Tamper-Evident Packaging Rule." This petition is required to contain the following:

(i) The name of the drug product or, if the petition seeks an exemption for a drug class, the name of the drug class and a list of products within that class.

(ii) The reasons that the drug product's compliance with the tamper-evident packaging and labeling requirements is unnecessary or cannot be achieved.

(iii) A description of alternative steps that are available or that the petitioner has already taken to reduce the likelihood that the product or drug class will be the subject of malicious adulteration.

(iv) Other information justifying an exemption.

(e) Holders of approved new drug applications for OTC drug products are required to provide the FDA with notification of changes in packaging and labeling to comply with the requirements of this section. Changes in packaging and labeling required by the regulation may be made before FDA approval. Manufacturing changes by which capsules are to be sealed require prior FDA approval.

(f) This section does not affect any requirements for "special packaging" as required under the Poison Prevention Packaging Act of 1970.

5. Drug product inspection
   (a) Packaged and labeled products must be examined during finishing operations to provide assurance that containers and packages in the lot have the correct label.
   (b) A representative sample of units should be collected at the completion of finishing operations and should be visually examined for correct labeling.
   (c) Results of these examinations must be recorded in the batch production records.
6. Expiration dating
   (a) All packaged drug products must carry an expiration date that has been determined from appropriate stability testing.
   (b) Expiration dates must be related to the recommended storage conditions stated on the label as determined by stability studies.
   (c) If the drug product is to be reconstituted at the time of dispensing, its label must carry expiration information for both the reconstituted and unreconstituted forms.
   (d) Expiration dates must appear on labeling in accordance with the requirements stated elsewhere in this regulation.
   (e) Homeopathic drug products are exempt from the requirements of this section.
   (f) Allergenic extracts that are labeled "No U.S. Standard of Potency" are exempt.
   (g) New drug products for investigational use are exempt provided that they meet appropriate standards or specifications as demonstrated by stability studies during their use in clinical investigations. If new drug products for investigational use are to be reconstituted at the time of dispensing, their labeling must bear expiration information for the reconstituted product.
   (h) Pending consideration of a proposed exemption published in the *Federal Register*, September 29, 1978, the requirements in this section will not be enforced for human drug products if their labeling does not bear dosage limitations and they are stable at least three years as supported by stability data.

### *Holding and Distribution*

1. Warehousing procedures
   (a) Written procedures describing the warehousing of drug products must be written and followed. These procedures should include:
      (i) Quarantine of drug products before release by the quality control unit.
      (ii) Storage of drug products under appropriate conditions of temperature, humidity, and light so that the quality of the drug products is not affected.
2. Distribution procedures
   (a) Written procedures concerning the distribution of drug products must be established and followed. These procedures should include:
      (i) A procedure that assures the distribution of the oldest approved stock first. Deviation from this procedure is acceptable if it is temporary and appropriate.

(ii) A system for documenting distribution so that distribution of each lot of drug product can be readily determined to facilitate its recall if required.

***Laboratory Controls***

1. General requirements
   (a) The establishment of any specifications, standards, sampling plans, test processes, or other laboratory control mechanism required by this part of the regulation, including any changes to the above must be drafted by the appropriate organizational unit and reviewed and approved by the quality control unit. All actions must be documented at the time of performance and any deviation must be recorded and justified.
   (b) Laboratory controls must include the establishment of scientifically sound and appropriate specifications, standards, sampling plans, and test procedures designed to assure that all materials conform to appropriate standards of identity, strength, quality, and purity. Laboratory controls should include:
      (i) Determination of conformance to written specifications for the acceptance of each lot within each shipment of raw materials. The specifications should include a description of the sampling and testing procedures used. Samples must be representative and adequately identified. These procedures must also require appropriate retesting of any material that is subject to deterioration.
      (ii) Determination of conformance to written specifications and a description of sampling and testing procedures for in-process materials.
      (iii) The calibration of instruments, apparatus, gauges, and recording devices at specified intervals in accordance with an established written program containing specific directions, schedules, limits for accuracy and precision, and provisions for remedial action in the event that the limits are not met. Any such devices that do not meet the established specifications must not be used.
2. Testing and release for distribution
   (a) Laboratory testing of each lot of drug product must be conducted to establish conformance to final specifications for the product. Testing must include identity and strength of each active ingredient. Where sterility and/or pyrogen testing are required on short-lived radiopharmaceuticals, batches may be released prior to completion of this testing provided that such testing is completed as soon as possible.
   (b) Each batch of product required to be free of objectionable microorganisms must be tested appropriately.
   (c) All sampling and testing plans must be described in written procedures that include the method of sampling and the number of units to be tested.
   (d) Acceptance criteria for the sampling and testing conducted by the quality control unit must be adequate to assure that the batch being tested meets all specifications. Appropriate statistical quality control criteria should be used. The statistical quality control criteria must include acceptance levels and/or rejection levels.

(e) The accuracy, sensitivity, specificity, and reproducibility of test methods used must be established and documented. Validation and documentation must be accomplished in accordance with this regulation.
(f) Drug products failing to meet established standards or specifications and any relevant quality control criteria must be rejected. Reprocessing may be performed, however, prior to acceptance and use, and reprocessed material must meet all standards, specifications, and other relevant criteria.

3. Stability testing
   (a) There must be a written testing program designed to assess the stability characteristics of every drug product. The results of such testing must be used to determine appropriate storage conditions and expiration dates. The written program must include:
      (i) Sample size and test intervals based on statistical criteria for each attribute examined.
      (ii) Storage conditions for sampled retained for testing.
      (iii) Reliable, meaningful and specific test methods.
      (iv) Testing of the product in the same container-closure system as the one in which the product is to be marketed.
      (v) Testing of drug products for reconstitution at the time of dispensing as well as after they are reconstituted.
   (b) An adequate number of batches of each drug product must be tested to determine appropriate expiration date. A record of such data must be maintained. Accelerated studies, combined with basic stability information on the components and drug product in its container-closure system may be used to project a tentative expiration date that is beyond the date supported by shelf life studies. However, there must be stability studies conducted including drug product testing at appropriate intervals until the tentative expiration date is verified.
   (c) The requirements for homeopathic drug products are as follows:
      (i) There must be a written assessment of stability based on testing or examination of the drug product for compatibility of the ingredients, and based on marketing experience with the drug product to indicate that there is no degradation of the product for the normal or expected period of use.
      (ii) Evaluation of stability must be based on the same container-closure system as the one in which the drug product is to be marketed.
   (d) Allergenic extracts that are labeled "No U.S. Standard of Potency" are exempt from the requirements of this section.
4. Special testing requirements
   (a) For each batch of drug product claimed to be sterile and/or pyrogen free, there must be appropriate laboratory testing to establish conformance to this claim. The test procedures must be in writing and must be followed.
   (b) For each batch of ophthalmic ointment, there must be appropriate testing to determine conformance to specifications regarding the presence of foreign particles and harsh or abrasive substances. The test procedures must be in writing and must be followed.
   (c) For each batch of controlled-release dosage form, there must be appropriate laboratory testing to determine conformance to the specifications for the rate

of release of each active ingredient. The test procedures must be in writing and must be followed.

5. Reserve samples
   (a) An identified reserve sample that is representative of each lot or of each shipment of each active ingredient must be retained. This reserve sample should contain at least twice the quantity needed for all tests required to determine whether the active ingredient meets its established specifications with the exception of sterility and pyrogen testing. The required retention time is as follows:
      (i) For an active ingredient in a drug product other than those described in paragraphs (b) and (c) below, the reserve sample must be retained for one year after the expiration date of the last lot of drug product containing that lot of active ingredient.
   (b) For an active ingredient in a radioactive drug product except for nonradioactive reagent kits, the reserve sample must be retained for:
      (i) Three months after the expiration date of the last lot of the drug product containing that lot of active ingredient if the expiration dating period of the drug product is 30 days or less.
      (ii) Six months after the expiration date of the last lot of the drug product containing that lot of active ingredient if the expiration dating period of the drug product is more than 30 days.
   (c) For an active ingredient in an OTC drug product that is exempt from bearing an expiration date, the reserve sample must be retained for three years after distribution of the last lot of drug product containing that lot of active ingredient.
   (d) A properly identified reserve sample that is representative of each batch of drug product must be retained and stored under conditions consistent with the product labeling. The reserve sample must be stored in the same immediate container closure system in which the drug product is marketed or in one that has essentially the same characteristics. The reserve sample consists of at least twice the quantity needed to perform all the required tests except those for sterility and pyrogens. Reserve samples from representative sample lots or batches selected by acceptable statistical procedures must be examined visually at least once a year for evidence of deterioration unless visual examination would affect the integrity of the reserve sample. Any evidence of reserve sample deterioration must be investigated. The results of the examination must be recorded and maintained with stability data concerning that drug product. Retention times are as follows:
      (i) For a drug product other than the exceptions noted above, the reserve sample must be retained for one year after the expiration date of the drug product.
      (ii) For a radioactive drug product, except for nonradioactive reagent kits, the retention sample must be retained for:
         (1) three months after the expiration date of the drug product if the expiration date is 30 days or less or
         (2) six months after the expiration date of the drug product if the expiration date is more than 30 days.

(iii) For an OTC drug product that is exempt from bearing an expiration date, the reserve sample must be retained for three years after the batch of drug product is fully distributed.

6. Animals used in testing components, in-process materials, or drug products for compliance with established specifications must be maintained and controlled in a manner that assures their suitability for their intended use. They must be identified and adequate records must be maintained showing the history of their use.
7. If a reasonable possibility exists that a nonpenicillin drug product has been exposed to cross-contamination with penicillin, the nonpenicillin drug product must be tested for the presence of penicillin. The drug product may not be marketed if a detectable level of penicillin is found when tested according to procedures specified in "Procedures for Detecting and Measuring Penicillin contamination in Drugs" which is incorporated in the regulation by reference.

***Records and Reports***

1. General Requirements
   (a) Any production, control, or distribution record that is associated with a batch of a drug must be retained for at least one year after the expiration date of the batch OR, for OTC drug products that do not have expiration dates, three years after complete distribution of the batch.
   (b) Records must be retained for all components, containers, closures, and labeling for the same time periods shown in (a) above.
   (c) All retained records or copies of these records must be readily available for authorized inspection at any time in the required retention period. Records must be available for inspection where the activities described therein occurred. Photocopying or similar reproduction by investigators must be permitted.
   (d) Retained records may be original records or true copies such as photocopies, microfilm, microfiche, or other accurate reproduction of the original.
   (e) Written records that must be retained must be maintained so that data contained therein can be used for evaluating the quality standards of each drug product to determine the need for changes in drug product specifications or manufacturing or control procedures. Such reviews should be conducted at least annually. Written procedures must be established and followed for these evaluations and must include provisions for:
      (i) A review of a representative number of batches, whether approved or rejected, and records associated with the batch.
      (ii) A review of complaints, recalls, returned or salvaged drug products, and investigations conducted under Section 211.192 of the GMP regulations for each drug product.
   (f) Procedures must be established to assure that the responsible officials of the firm are notified in writing of any investigations conducted under Sections 211.198, 211.204, or 211.208 of any recalls, reports of inspectional observations issued by the FDA, or any regulatory actions relating to GMP brought by the FDA.
2. A written record of major equipment cleaning, maintenance (except routine maintenance), and use must be included in individual equipment logs that show

the date, time, product, and lot number of each batch processed. The persons performing and double checking the cleaning and maintenance should date and sign or initial the log indicating that the work was performed. Entries in the log must be in chronological order.

3. Component, drug product container, closure, and labeling records must include the following:
   (a) The identity and quantity of each shipment of each lot of components, drug product containers, closures, and labeling. Also required are the identity of the supplier, the supplier's lot number(s), the receiving code, the date of receipt, and name and location of the prime manufacturer if different from the supplier.
   (b) The results of any test or examination performed and any conclusions derived from these results.
   (c) An individual inventory record of each component and a reconciliation of the use of each lot of such component. The inventory record must contain sufficient information to allow determination of any batch or lot of drug product associated with the use of each component.
   (d) Documentation of the examination and review of labels and labeling for conformance with established specifications.
   (e) The disposition of rejected materials.
4. Master production and control records

   Batch production and control records should be prepared for each batch of drug product produced and must include complete information about the production and control of that batch. These records must include:
   (a) A full and complete reproduction of the appropriate master production or control record. The copy must be checked for accuracy, dated, and signed.
   (b) Documentation that each significant step in the manufacture, processing, packaging, and holding of the batch was accomplished as prescribed, including:
      (i) Dates.
      (ii) Identity of individual major equipment used. This includes packaging lines.
      (iii) Complete and specific identification of each batch of component or in-process material used.
      (iv) Weight and measures of components used in the course of processing.
      (v) In-process and laboratory control results.
      (vi) Inspection of the packaging and labeling area before and after use.
      (vii) Documentation of the actual yield and the percentage of theoretical yield that this represents at critical stages of processing.
      (viii) Complete labeling control records, including specimens or copies of all labeling used.
      (ix) A description of drug product containers and closures.
      (x) Any sampling performed.
      (xi) Identification of the persons performing and directly supervising or checking significant steps in the operation.
      (xii) Any investigations conducted.
      (xiii) Results of examinations made.

5. All drug product production and control records, including those for packaging and labeling, must be reviewed and approved by the quality control unit to determine compliance with all established written procedures before a batch is released or distributed. Any unexplained discrepancy or the failure of a batch or any of its components to meet any of the established specifications must be thoroughly investigated. The investigation must be extended to other batches of the same drug product and other drug products that may have been associated with the specific fault or discrepancy. A written record of the investigation must be made and include the conclusions and any required follow-up.
6. Laboratory records
   (a) Laboratory records must include complete data derived from all tests needed to assure compliance with established specifications and standards. This includes examinations and assays as follows:
      (i) A description of the sample received for testing with identification of source. For example, location where the sample was obtained, quantity, lot number or other distinctive code, date the sample was taken, and the date that it was received for testing.
      (ii) A statement of each method used in the testing of the sample. The statement must indicate the location of data that establish that the methods used in the testing of the sample meet proper standards of accuracy and reliability as applied to the product tested. (If the method used is in the current revision of the U.S. Pharmacopeia (USP), National Formulary (NF), or other recognized standard reference or if it is detailed in an approved NDA, this statement will not be required.)
      (iii) A statement of the weight or measure of sample used for each test.
      (iv) A complete record of all data secured in the course of each test, including all graphs, charts, and spectra from laboratory instrumentation properly identified to the specific component and lot tested.
      (v) A record of all calculations performed in connection with the test, including units of measure, conversion factors, and equivalency factors.
      (vi) A statement of the results of tests and how the results compare with established standards of identity, strength, quality, and purity for the component tested.
      (vii) The initials or signature of the person who performed each test and the date the tests were performed.
      (viii) The initials or signature of a second person showing that the original records have been reviewed for accuracy, completeness, and compliance with established standards.
   (b) Complete records must be maintained of any modification of an established method employed in testing. These records must include the reason for the modification and verify that the modification produced results that are at least as accurate and reliable for the material being tested as the established method.
   (c) Complete records must be maintained of any testing and standardization of laboratory reference standards, reagents, and standard solutions.
   (d) Complete records must be maintained of the periodic calibration of laboratory instruments, apparatus, gauges, and recording devices.

(e) Complete records must be maintained of all stability testing performed in accordance with Section 211.166 of the regulation.

7. Distribution records must contain the name and strength of the product and description of the dosage form, name and address of the consignee, date and quantity shipped, and lot or control number of drug product. For compressed medical gas products, distribution records are not required to contain lot or control numbers.
8. Complaint files
   (a) Written procedures describing the handling of all written and oral complaints regarding a drug product must be established and followed. These procedures must include provisions for review by the quality control unit of any complaint involving the possible failure of a drug product to meet any of its specifications and a determination as to the need for an investigation. These procedures must include provisions for review to determine whether the complaint represents a serious and unexpected adverse drug experience which is required to be reported to the FDA.
   (b) A written record of each complaint must be maintained in a file designated for product complaints. The file may be maintained at another facility if the written records of such files are readily available for inspection at that other facility. Written reports involving a drug product must be maintained until at least one year after the expiration date of the drug product or one year after the date that the complaint was received, whichever is longer. In the case of certain OTC drug products lacking expiration dating because they meet the criteria for exemption, such written records must be maintained for three years after distribution of the drug product.
      (i) The written record must include the following information where known: the name and strength of the drug product, lot number, name of complainant, nature or complaint, and reply to the complainant.
      (ii) Where an investigation is conducted, the written record must include the findings of the investigation and follow-up. The record or a copy of the record of investigation must be maintained at the location where the investigation occurred.
      (iii) Where an investigation is not conducted, the written record must include the reason that an investigation was not considered to be necessary and the name of the responsible person making the determination.

### *Returned and Salvaged Drug Products*

1. Returned drug products—Returned drug products must be identified as such and held. If the conditions under which returned drug products have been held, stored, or shipped before or during the return or the condition of the drug product, its container, carton, or labeling is a result of storage or shipping casts doubt on the safety, identity, strength, quality, or purity of the drug product, the returned drug product must be destroyed unless examination testing or other investigation proves the drug product meets appropriate standards. Records of returned drug products must be maintained and must include the name and label potency of the drug

product dosage lot number, reason for the return, quantity returned, date of disposition, and ultimate disposition of the returned product. If the reason for a drug product being returned implicates associated batches, an investigation must be conducted. Procedures for the holding, testing, and reprocessing of returned drug products must be in writing and must be followed.

2. Drug product salvaging—Drug products that have been subjected to improper storage conditions, including extremes in temperature, humidity, smoke, fumes, pressure, age, or radiation due to natural disasters, fires, accidents, or equipment failures, must not be salvaged and returned to the marketplace. Whenever there is a question whether drug products have been subjected to such conditions, salvaging operations may be conducted only if there is (a) evidence from laboratory tests and assays that the drug products meet all applicable standards of identity, strength, quality, and purity and (b) evidence from inspection of the premises that the drug products and associated packaging were not subjected to improper storage conditions as a result of the disaster or accident. Organoleptic examinations are acceptable only as supplemental evidence that the drug products meet appropriate standards of identity, strength, quality, and purity. Records including name, lot number, and disposition must be maintained for drug products subject to this section.

## 3 GUIDANCE FOR INDUSTRY: QUALITY SYSTEMS APPROACH TO PHARMACEUTICAL CURRENT GOOD MANUFACTURING PRACTICE REGULATIONS

This guidance document was written by the FDA to help manufacturers implement what they consider to be modern quality systems and risk management approaches that will meet the requirements of the FDA's GMP regulations. The guidance describes what the FDA considers a comprehensive quality systems (QS) model. It also explains how manufacturers can be in full compliance with the GMP regulations by implementing such quality systems. The FDA does not intend this guidance to place new expectations on manufacturers nor does this replace the GMPs.

As is true with all guidance documents, this document does not establish legally enforceable responsibilities, but rather it describes the FDA's current thinking. Thus, this guidance should be viewed as a set of recommendations unless a regulation is cited.

The objective of this guidance is to describe a quality systems model and demonstrate how and where the elements of this model can fit within the requirements of the CGMP regulations. The philosophy being put forward is that *quality should be build into the product, and testing alone cannot be relied on to ensure product quality*.

### 3.1 CGMPS and the Concepts of Modern Quality Systems

The FDA believes that several key concepts are critical for any discussion of modern quality systems. The following concepts are used throughout this guidance as they relate to the manufacture of pharmaceutical dosage forms:

*Quality* For the purposes of this guidance, the phrase *achieving quality* means achieving the identity, strength, purity, and other quality characteristics designed to ensure safety and effectiveness.

*Quality by Design and Product Development* This means designing and developing a product and its associated manufacturing processes that will be used to ensure that the product consistently attains a predefined quality at the end of the manufacturing process.

*Quality Risk Management* This component of a quality systems framework can help guide the setting of specifications and process parameters for dosage form manufacturing, assess and mitigate the risk of changing a process or specification, and determine the extent of discrepancy investigations and corrective actions.

*Corrective and Preventative Action (CAPA)* This is a regulatory concept that focuses on investigating, understanding, and correcting discrepancies while attempting to prevent their recurrence. This model separates CAPA into three separate concepts:

- Remedial corrections of an identified problem
- Root cause analysis with corrective action to help understand the cause of the deviation and prevent recurrence of a similar problem
- Preventative action to prevent recurrence of similar problems

*Change Control* This process focuses on managing change to prevent unintended consequences.

*Quality Unit* While the GMPs refer to a quality unit, current industry practice is to divide the responsibilities of this unit between two groups:

- Quality control (QC) usually involves (a) assessing the suitability of incoming components and the finished products, (b) evaluating the performance of the manufacturing process, and (c) determining the acceptability of each batch for release and distribution
- Quality assurance (QA) involves (a) review and approval of all procedures related to manufacturing and maintenance, (b) review of records, and (c) auditing and performing/evaluating trend analyses.

*Six-System Inspection Model* The FDA's instruction manual for its investigators is a systems-based approach to inspection consistent with this guidance. The FDA defines six interlocked systems: (1) the quality system which encompasses all the other systems, (2) a materials system, (3) a production system, (4) a packaging and labeling system, (5) a facilities and equipment system, and (6) a laboratory controls system. The agency believes that use of this overall system approach will help firms achieve better control.

### 3.2 Quality Systems Model

This section was written to describe a model for use in pharmaceutical manufacturing that can supply the controls to consistently produce a product of acceptable quality. The model is described by four major factors:

- Management responsibilities
- Resources
- Manufacturing operations
- Evaluation

*Management Responsibilities* The FDA feels that a robust quality system model calls for management to play a key role in the design, implementation, and management of the quality system.

*Resources* Sufficient resources should be provided to create a robust quality system that complies with the GMP regulations. Senior management or a designee should be responsible for providing adequate resources.

*Facilities and Equipment* The technical experts who have an understanding of pharmaceutical science, risk factors, and manufacturing processes related to the product are responsible for defining specific facility and equipment requirements. The equipment must be qualified, calibrated, cleaned, and maintained to prevent contamination and product mix-ups. It is important to remember that the GMPs place as much emphasis on process equipment as on testing equipment while most quality systems focus only on testing equipment.

*Control Outsourced Operations* Quality systems call for contracts with outside suppliers that clearly describe the materials or service, quality specification responsibilities, and communication mechanisms.

*Manufacturing* There is an overlap between the elements of a quality system and the GMP regulation requirements for manufacturing operations. One should always remember that the FDA's enforcement programs and inspectional coverage are based on the GMPs. The FDA feels that the following factors are essential in a manufacturing quality system:

1. Design, develop, and document product and processes
2. Examine inputs
3. Perform and monitor operations
4. Address nonconformities

*Evaluation Activities* This includes the following activities:

1. Analyze data for trends
2. Conduct internal audits
3. Quality risk management
4. Corrective action
5. Preventative action
6. Promote improvements

## 4 GUIDANCE FOR INDUSTRY: PAT—FRAMEWORK FOR INNOVATIVE PHARMACEUTICAL DEVELOPMENT, MANUFACTURING, AND QUALITY ASSURANCE

This guidance is intended to describe a regulatory framework that the FDA chooses to call *process analytical technology*, or PAT. It is the FDA's hope that this will

encourage the voluntary development and implementation of innovative pharmaceutical development, manufacturing, and quality assurance. The FDA intended this guidance for a broad audience in different organizational units. To a large extent, the guidance discusses principles with the goal of highlighting opportunities and developing regulatory processes that encourage innovation.

Conventional pharmaceutical manufacturing is usually accomplished using batch processing with laboratory testing of samples at various stages of manufacturing to evaluate quality. The FDA believes that opportunities exist for improving the development, manufacturing, and quality assurance steps through innovation in product and process development, process control, and analysis.

Typically, the pharmaceutical industry has been reluctant to try something new due to the fear that the new approach will not find favor with the FDA. An FDA rejection would result in costly delays and processing revisions that industry is unwilling to risk. The FDA now says that this hesitancy is undesirable from a public health perspective and it would like to see more innovation introduced. According to the FDA, pharmaceutical manufacturing should be based on:

- The design of effective and efficient manufacturing manufacturing processes
- Product and process specifications based on an understanding of how formulation and process factors affect product performance
- Continuous real-time quality assurance
- Relevant regulatory policies and procedures tailored to accommodate the most current level of scientific knowledge
- Risk-based regulatory approaches that recognize:
  - The level of scientific understanding of how formulation and manufacturing process factors affect product quality and performance
  - The capability of process control strategies to prevent or mitigate the risk of producing a poor-quality product

It is the intent of this guidance to facilitate progress to this state. So far, the FDA's stated goal is not being met. FDA representatives have stated the agency's concern about the failure of industry to rush to implement change. However, the economies of change continue to favor the status quo.

### 4.1 PAT Framework

Quality should be built into pharmaceutical products through a comprehensive understanding of:

- Intended therapeutic objectives, patient population, route of administration, and pharmacokinetic characteristics of a drug
- Chemical, physical, and biopharmaceutic characteristics of a drug
- Design of a product and selection of product components and packaging based on drug attributes
- Design of manufacturing processes using principles of engineering, material science, and quality assurance to ensure acceptable and reproducible product quality and performance throughout a product's shelf life

***Process Understanding*** A process is considered to be well understood when all critical sources of variability are identified and explained, variability is managed by the process, and product quality attributes can be accurately and reliably predicted.

***Principles and Tools*** Pharmaceutical manufacturing often consists of a series of unit operations, each of which is intended to change certain properties of the materials being processed. To assure these changes are acceptable and reproducible, consideration should be given to the quality attributes of incoming materials and their acceptability for the given unit operation. Most current pharmaceutical processes are based on time-defined endpoints such as "blend for ten minutes." In some cases, these time-defined endpoints do not consider the effects of physical differences in raw materials. Processing difficulties can arise that result in the failure of a product to meet specifications even if the raw materials conform to established specifications. Use of PAT tools and principles can provide relevant information relating to physical, chemical, and biological attributes. The process understanding gained from this information will enable process control and optimization, address the limitation of the time-defined endpoints, and improve efficiency.

***PAT Tools*** There are many tools available that enable process understanding. These tools, when used within a system, can provide effective and efficient means for acquiring information to facilitate process understanding, continuous improvement, and development of risk mitigation strategies. Such tools are categorized as follows:

- Multivariate tools for design, data acquisition, and analysis
- Process analyzers
- Process control tools
- Continuous improvement and knowledge management tools

***Strategy for Implementation*** To enable successful implementation of PAT, flexibility, coordination, and communication with manufacturers are critical. The FDA believes that current regulations are sufficiently broad to accommodate these strategies. In the course of implementing the PAT framework, manufacturers may want to evaluate the suitability of a tool on experimental and/or production equipment and processes. It is recommended that risk analysis of the impact on product quality be conducted before installation. This can be accomplished within the facility's quality system without prior notification to the agency. Data collected using an experimental tool should be considered research data. If conducted in a production facility, it should be done under the facility's quality system. The FDA does not intend to inspect research data collected on an existing product for the purpose of evaluating the suitability of an experimental PAT tool. Its routine inspection of a firm's manufacturing process that incorporates a PAT tool for research purposes will be based on current regulatory standards.

The FDA has posted much of the information that firms will need in order to implement a PAT program on the Web at http://www.fda.gov/cder/ops/pat.htm.

All marketing applications, amendments, or supplements to an application should be submitted to the appropriate Center for Drug Evaluation and Research (CDER) or Center for Veterinary Medicine (CVM) division in the usual manner. In general,

PAT implementation plans should be risk based. The FDA has suggested the following possible implementation plans, where appropriate:

- PAT can be implemented under the facility's own quality system. CGMP inspections by the PAT team or PAT-certified investigator can precede or follow PAT implementation.
- A supplement [Changes Being Expected (CBE), Changes Being Expected in 30 Days (CBE-30), or Prior Approval Supplement (PAS)] can be submitted to the agency prior to implementation, and, if necessary, an inspection can be performed by a PAT team or PAT certified investigator before implementation.
- A *comparability protocol* can be submitted to the agency outlining PAT research, validation and implementation strategies, and time lines. Following approval of this *comparability protocol* by the agency, one or a combination of the above regulatory pathways can be adopted for implementation.

To facilitate adoption or approval of a PAT process, manufacturers may request a preoperational review of a PAT manufacturing facility and process by the PAT team by contacting the FDA Process Analytical Technology Team at PAT@cder.fda.gov. It should be noted that when certain PAT implementation plans neither affect the current process nor require a change in specifications, several options can be considered. Manufacturers should evaluate and discuss with the agency the most appropriate option for their situation.

## 5 GUIDANCE FOR INDUSTRY: PART 11. ELECTRONIC RECORDS; ELECTRONIC SIGNATURES—SCOPE AND APPLICATION

Of the many regulations written by the FDA, the least understood is undoubtedly 21 CFR Part 11. Rather than review the regulation itself, which is under review and possible revision, we will review the guidance for industry that FDA published in August 2003 to "aid" industry in their puzzlement. Depending on the source, it appears to be questionable as to whether this guidance document aids or confuses. It exists, however, and like it or not, understand it or not, the regulation must be followed.

The guidance indicates that the FDA's approach is based on three main components:

- The regulation will be interpreted narrowly. Fewer records will be considered subject to Part 11.
- Those records that are considered subject to Part 11 will be subject to enforcement discretion with regard to the requirements for validation, audit trails, record retention, and record copying in the manner described and with regard to all Part 11 requirements for systems that were operational before the effective date of this regulation.
- All predicate rule requirements will be enforced. This includes record and record-keeping requirements.

The FDA does intend to enforce all other provisions of Part 11, including certain controls for closed systems. The following controls and requirements will be enforced:

- Limiting system access to authorized individuals
- Use of operational system checks
- Use of authority checks
- Use of device checks
- Determination that persons who develop, maintain, or use electronic systems have the education, training, and experience to perform their assigned tasks
- Establishment of and adherence to written policies that hold individuals accountable for actions initiated under their electronic signatures
- Appropriate controls over systems documentation
- Controls for open systems corresponding to controls for closed systems
- Requirements related to electronic signatures

***Part 11 Records*** Under the narrow interpretation, the FDA considers Part 11 to be applicable to the following records or signatures in electronic format:

1. Records that are required to be maintained under predicate rule requirements and that are maintained in electronic format in place of paper format.
2. Records that are required to be maintainer under predicate rules, that are maintained in electronic format in addition to paper format, and that are relied on to perform regulated activities.
3. Records submitted to the FDA under predicate rules in electronic format. However, a record that is not itself submitted but is used in generating a submission is not a Part 11 record.
4. Electronic signatures that are intended to be the equivalent of handwritten signatures, initials, and other general signings required.

***FDA's Approach to Specific Part 11 Requirements***

1. *Validation* With respect to validation, the agency intends to exercise enforcement discretion regarding specific Part 11 requirements. However, compliance with all applicable predicate rules for validation is still expected. The FDA suggests an approach to validation be based on a justified and documented risk assessment and a determination of the potential of the system to affect product quality, safety, and record integrity.
2. *Audit Trail* The agency also intends to exercise enforcement discretion regarding specific requirements related to computer-generated, time-stamped audit trails and any corresponding requirements in Part 11. Compliance with all applicable predicate rule requirements related to documentation of date, time, or sequencing of events is still expected. It is also required to comply with rules for ensuring that changes to records do not obscure previous entries.
3. *Legacy Systems* The FDA intends to exercise enforcement discretion with respect to all Part 11 requirements for systems that otherwise were operational prior to August 20, 1997. Thus they do not intend to take enforcement action to enforce compliance with any Part 11 requirements if all of the following criteria are met for a specific system:

- The system was operational before the effective date.
- The system met all applicable predicate rule requirements before the effective date.
- The system currently meets all applicable predicate rule requirements.
- There is documented evidence and justification that the system is fit for its intended use.

4. *Copies of Records* Enforcement discretion will be applied with respect to specific Part 11 requirements for generating copies of records and any corresponding requirements in this part. An investigator should be provided with reasonable and useful access to records during an inspection. All records held by a manufacturer are subject to inspection.
5. *Record Retention* The FDA intends to exercise enforcement discretion with regard to the Part 11 requirements for the protection of records to enable their accurate and ready retrieval at any time throughout the records retention period.

## 6 GUIDANCE FOR INDUSTRY AND FDA: CURRENT GOOD MANUFACTURING PRACTICE FOR COMBINATION PRODUCTS

This document discusses the applicability of GMPs to combination products as defined under 21 CFR 3.2(e). Manufacturers must ensure that the product is not adulterated; the product possesses adequate strength, quality, identity, and purity; and the product complies with performance standards as appropriate. This guidance does not address technical manufacturing methods or make recommendations for manufacturers' selection of facilities used in manufacturing.

A combination product is a product composed of a drug and a device, a biological product and a device, a drug and a biological product, or a drug, a device, and a biological product. For the purposes of this document, a constituent part of a combination product is an article in a combination product that can be distinguished by its regulatory identity as a drug, device, or biological product.

For regulatory purposes, a combination product is assigned to an agency center or alternative organizational component that will have primary jurisdiction for its premarket review and regulation. Manufacturers will be required to use the applicable GMP for their products. Regulations that may apply are:

- GMP regulations for finished pharmaceuticals (21 CFR Parts 210 and 211).
- Quality system regulations for devices (21 CFR Part 820).
- The biological product regulations (21 CFR Parts 600–680) may also apply to the manufacture of drugs that are also biological products along with the drug provisions.

There are no GMP regulations specifically for combination products. Until such regulations are promulgated, the manufacture of each constituent part is governed by the regulations for that component.

The Office of Combination Products is available as a resource to sponsors throughout the lifecycle of a combination product. This office can be reached at

(301) 427-1934 or by E-mail at combination@fda.cov. Updated guidance documents are available at the office's Internet website, http://www.fda/gov/oc/combination.

## 7 GUIDANCE FOR INDUSTRY: POWDER BLENDS AND FINISHED DOSAGE UNITS—STRATIFIED IN-PROCESS DOSAGE UNIT SAMPLING AND ASSESSMENT

This guidance is intended to assist manufacturers in meeting the GMP requirements for demonstrating the adequacy of mixing to ensure uniformity of in-process powder blends and finished dosage units.

***Stratified Sampling*** In this process dosage units are sampled at predefined intervals and representative samples collected from specifically targeted locations in the compression/filling operations that have the greatest potential to yield extremes of drug concentration.

This guidance describes methods of sampling that might be used to demonstrate active ingredient homogeneity. These methods are put forward as suggestions and are not intended to be the only methods for meeting FDA requirements for demonstration of the adequacy of a powder mix.

***Assessment of Powder Mix Uniformity*** The following procedures are recommended:

1. Conduct blend analysis on batches by extensively sampling the mix in the blender and/or intermediate bulk containers.
2. Identify appropriate blending time and speed ranges, dead spots in blenders, and locations of segregation in intermediate bulk containers (IBCs).
3. Define the effects of sample size (1–10 times the dosage unit range) while developing a technique capable of measuring the true uniformity of the blend. Sample quantities larger than 3 times the dosage size can be used with adequate scientific justification.
4. Design blend-sampling plans and evaluate them using appropriate statistical analyses.
5. Quantitatively measure any variability that is present among the samples. Attribute the sample variability to either lack of uniformity of the blend or sampling error. Significant variances in the blend data within a given location can be an indication of one factor or a combination of factors such as inadequacy of blend mix, sampling error, or agglomeration. Significant between-location variance can indicate that the blending operation is inadequate.

***Correlation of Powder Mix Uniformity with Stratified In-Process Dosage Unit Data*** The following steps are recommended for correlation:

1. Conduct periodic sampling and testing of the in-process dosage units by sampling them at defined intervals and locations throughout the compression or filling process. Use a minimum of 20 appropriately spaced in-process dosage

unit sampling points. There should be at least 7 samples taken from each of these locations for a total minimum of at least 140 samples.

2. Take 7 samples from each additional location to further assess each significant event, such as filling or emptying of hoppers and IBCs, start and end of the compression or filling process, and equipment shutdown. This may be accomplished by using process development batches, validation batches, or routine manufacturing batches for approved products.
3. Significant events may also include observations or changes from one batch to another (e.g., batch scale-up and observations of undesirable trends in previous batch data).
4. Prepare a summary of the data and analysis used to correlate the stratified sampling locations with significant events in the blending process.
5. Compare the powder mix uniformity with the in-process dosage unit data described above.
6. Investigate any discrepancies observed between powder mix and dosage unit data and establish root causes. At least one troubleshooting guide is available that may be helpful with this task. Possible corrections may range from going back to formulation development to improve powder characteristics to process optimization. Sampling problems may also be negated by use of alternate state-of-the-art methods of in situ real-time sampling and analysis.

***Correlation of Stratified In-Process Samples with Finished Product*** The following steps are recommended:

1. Conduct testing for uniform content of the finished product using an appropriate procedure or as specified in the ANDA or the NDA for approved products.
2. Compare the results of stratified in-process dosage unit analysis with uniform content of the finished dosage units from the previous step. This analysis should be done without weight correction.
3. Prepare a summary of the data and analysis used to conclude that the stratified in-process sampling provides assurance of uniform content of the finished product.

### 7.1 Validation of Batch Powder Mix Homogeneity

This section describes sampling and testing the powder mix of demonstration and process validation batches used to support implementing the stratified sampling method described in this guidance.

The guidance document recommends that during the manufacture of demonstration and process validation batches, the following uniformity characteristics be assessed: (1) the powder blend, (2) the in-process dosage units, and (3) the finished product. Each attribute should be determined independently. It is further recommended that the following steps be used to identify sampling locations and acceptance criteria prior to the manufacture of the exhibit and/or validation batches:

1. Carefully identify at least 10 sampling locations in the blender to represent potential areas of poor blending. For example, in tumbling blenders (such as V-blenders, double cones, or drum mixers), samples should be selected from at least two depths along the axis of the blender. For convective blenders (such as a ribbon blender), a special effort should be made to implement uniform volumetric sampling to include the corners and discharge area (at least 20 locations are recommended to adequately validate convective blenders).
2. Collect at least three replicate samples from each location. Samples should meet the following criteria:
   - Assay one sample per location (number of samples $n = 10$, or $n = 20$ for ribbon blender).
   - RSD *(relative standard deviation)* of all individual results is 5.0%.
   - All individual results are within 10.0% (absolute) of the mean of the results.

It is also recommended that you not proceed any further with implementation of the methods described in this guidance until the criteria are met.

Sampling errors may occur in some powder blends, sampling devices, and techniques that make it impractical to evaluate adequacy of mix using only the blend data. In such cases, it is recommended that in-process dosage unit data be used in conjunction with blend sample data to evaluate blend uniformity.

Some powder blends may present an unacceptable safety risk when directly sampled. The safety risk, once described, may justify an alternate procedure. In such cases, process knowledge and data from indirect sampling combined with additional in-process dosage unit data may be adequate to demonstrate the adequacy of the powder mix. Data analysis used to justify using these alternate procedures should be described in a summary report that is maintained at the manufacturing facility.

### 7.2 Verification of Manufacturing Criteria

The assessment of powder mix uniformity and correlation of stratified in-process dosage unit sampling development procedures should be completed before establishing the criteria and controls for routine manufacturing. It is also recommend that the normality be assessed and that the RSD be determined from the results of stratified in-process dosage unit sampling and testing that were developed. The RSD value should be used to classify the testing results as either *readily pass* (RSD 4.0%), *marginally pass* (RSD 6.0%), or *inappropriate* for demonstration of batch homogeneity when RSD > 6.0%.

The FDA recommends that routine manufacturing batches be evaluated against the following criteria after completing the procedures described above to assess the adequacy of the powder mix and uniform content in the finished dosage form:

1. Standard criteria method (SCM)—This method is recommended when either of the following conditions is met:
   1.1. Results of establishing initial criteria are classified as *readily pass*.

1.2. Results of testing to the marginal criteria method (MCM) pass the criteria for switching to the SCM.
   1.2.1. *Stage 1 Test* To perform the stage 1 test, collect at least three dosage units from each sampling location, assay one dosage unit from each location, weight correct the results, and compare the results with the following criteria:
      1.2.1.1. RSD of all individual results is less than 5%.
      1.2.1.2. Mean of all results is 90–110% of target assay.

      If the results pass these criteria and the adequacy of mix and uniformity of dosage unit content for the batch are adequate, the SCM can be used for the next batch. If test results fail stage 1 criteria, extended testing to stage 2 is required.
   1.2.2. *Stage 2 Test* To perform the stage 2 test, assay the remaining two dosage units from stage 1 for each sampling location and compute the mean and RSD of data combined from both stage 1 and stage 2. Compare the results with the following criteria:
      1.2.2.1. For all individual results, the RSD should be less than 5.0%.
      1.2.2.2. Mean of all results is 90–110% of target assay.

      If results pass the above criteria, the adequacy of mix and uniformity of content for the batch are adequate and stage 1 can be used for the next batch. If test results fail the criteria, use the MCM described in the section below.

2. Marginal criteria method—The MCM can be used when either of the following conditions is met:
   2.1. Results of initial criteria establishment qualified as *marginally pass*.
   2.2. Results of initial criteria establishment qualified as *readily pass* or a batch was tested according to SCM and the test results failed both stage 1 and stage 2 criteria.
   2.3. If either of the above two criteria apply, use the weight corrected results from the stage 2 SCM analysis and compare this with the MCM criteria:
      2.3.1. For al individual results, the RSD is less than 6.0%.
      2.3.2. The mean of all results is 90.0–110.0% of target assay.
   2.4. It is acceptable to switch to the SCM when five consecutive batches pass the MCM criteria and result in RSD of less than 5.0%.

## 8 GUIDANCE FOR INDUSTRY: IMMEDIATE-RELEASE SOLID ORAL DOSAGE FORMS SCALE-UP AND POSTAPPROVAL CHANGES (SUPAC)—CHEMISTRY, MANUFACTURING AND CONTROLS, IN VITRO DISSOLUTION TESTING, AND IN VIVO BIOEQUIVALENCE DOCUMENTATION

This guidance provides recommendations to NDA and ANDA sponsors who intend to make changes to the product during the postapproval period. Changes include any change in components or composition of the product, the site of manufacture, the scale-up/scale-down of batch size, and/or the manufacturing process and/or equipment of an immediate-release oral formulation.

***Changes in Components (Excipients) and Composition*** Changes in the amount or source of drug substance are not addressed by this guidance. Changes in components or composition that have the effect of adding a new excipient or deleting an excipient are defined at level 3 except as described below:

1. Level 1 changes
   1.1. Level 1 changes are those that are unlikely to have any detectable impact on formulation quality and performance.
   1.2. Allowed changes (changes that can be made without prior FDA approval) are shown below. This is based on the assumption that the drug substance in the product is formulated to 100% of label potency. To be considered a level 1 change, the total additive effect of all excipient changes should not be more than 5% relative to the target dosage form weight.

| Excipient | Percentage of Excipient (W/W) Out of Total Target Dosage Form Weight |
|---|---|
| Filler | ±5 |
| Disintegrant | |
| Starch | ±3 |
| Other | ±1 |
| Binder | ±0.5 |
| Lubricant | |
| Calcium or magnesium Stearate | ±0.25 |
| Other | ±1 |
| Glidant | |
| Talc | ±1 |
| Other | ±0.1 |
| Film coat | ±1 |

   1.3. Test documentation
      1.3.1. Chemistry—Application/compendial release requirements and stability testing. For stability testing, one batch should be on long-term stability testing with data being reported in the annual report.
      1.3.2. Filing documentation—All information must be included in the annual report (including long-term stability data).
2. Level 2 changes
   2.1. Level 2 changes are those that could have a significant impact on formulation quality and performance. Tests and filing documentation for a level 2 change depend on three factors: (1) therapeutic range, (2) solubility, and (3) permeability. Therapeutic range is defined as either narrow or nonnarrow. Drug solubility and drug permeability are defined as either low or high. Changes in excipients, expressed as percent (w/w) of total formulation, greater than those listed for a level 1 change but less than or equal to the following percent ranges are acceptable level 2 changes:

| Excipient | Percentage of Excipient (w/w) of Total Target Dosage Form Weight |
|---|---|
| Filler | ±10 |
| Disintegrant | |
| Starch | ±6 |
| Other | ±2 |
| Binder | ±1 |
| Lubricant | |
| Ca or Mg stearate | ±0.5 |
| Other | ±2 |
| Glidant | |
| Talc | ±2 |
| Other | ±0.2 |
| Film coat | ±2 |

These percentages are based on the assumption that the drug substance in the finished product is formulated to 100% of labeled potency. The total additive effect of all excipient changes should not change by more than 10%.

All components in the formulation should have numerical targets that represent the nominal composition of the product on which any future changes in the composition of the product are based. Allowable changes in the composition should be based on the approved target composition and not on the composition based on previous level 1 or level 2 changes.

2.2. Test documentation

2.2.1. Chemistry

2.2.1.1. Application/compendial release requirements and batch records.

2.2.1.2. Stability testing—Test one batch with three months of accelerated stability data in supplement and on batch on long-term stability.

2.2.2. dissolution

2.2.2.1. High-permeability, high-solubility drugs—Dissolution of 85% in 15 min in 900 mL of 0/1 *N* HCl. If a drug product fails to meet this criterion, tests in 2.2.2.2 or 2.2.2.3 below should be performed.

2.2.2.2. Low-permeability, high-solubility drugs—Multipoint dissolution profile should be performed in the application/compendial medium at 15, 30, 45, 60, and 120 min or until an asymptote is reached. The dissolution profile of the proposed and currently used product formulations should be similar.

2.2.2.3. High-permeability, low-solubility drugs—Multipoint dissolution profiles should be performed in water, 0.1 *N* HCl, and USP buffer media at pH 4.5, 6.5, and 7.5 (five different profiles) for the proposed and currently accepted formulations. Adequate sampling should be performed at 15, 30, 45, 60, and 120 min until either 90% of drug from the drug product is

dissolved or an asymptote is reached. A surfactant may be used, but only with appropriate justification. The dissolution profile of the proposed and currently used product formulations should be similar.

2.2.3. In vivo bioequivalence documentation is not required for level 2. If the product does not meet any of the level 1 cases above, refer to level 3 changes.

2.2.4. Filing documentation—A prior approval supplement with all data including the accelerated stability data is required. This change should also be documented in the annual report along with the long-term stability data.

2.3. Level 3 changes

2.3.1. Level 3 changes are those that are likely to have a significant impact on formulation quality and performance. Tests and filing documentation vary depending on the following three factors: therapeutic range, solubility, and permeability. For example:

2.3.1.1. Any qualitative and quantitative excipient changes to a narrow therapeutic drug beyond the ranges specified in the level 1 table.

2.3.1.2. All other drugs not meeting the dissolution criteria under level 2.

2.3.1.3. Changes in the excipient ranges of low-solubility, low-permeability drugs beyond those listed in level 1.

2.3.1.4. Changes in the excipient ranges of all drugs beyond those listed in the level 2 table.

2.3.2. Test documentation

2.3.2.1. Chemical

(a) Application/compendial release requirements and batch records:

- Information available—One batch with three months accelerated stability data reported in a supplement and one batch on long-term stability reported in the annual report.
- Information NOT available—Up to three batches with three months accelerated stability data reported in the supplement and one batch on long-term stability data reported in annual report.

(b) Dissolution documentation—Case B dissolution profile as described in the table for level 2.

(c) In vivo bioequivalence documentation—Full bioequivalence study. This requirement may be waived with a verified acceptable in vivo/in vitro correlation.

2.3.2.2. Filing documentation—Prior approval supplement including accelerated stability data plus an annual report showing long-term stability data.

***Site Changes*** Site changes are changes in the location of manufacture for both company-owned and contract manufacturing facilities. A site change does not

include, for example, scale-up changes, changes in manufacturing equipment or a manufacturing process, and changes in Standard Operating Procedures (SOPs) or environmental changes. Each change must be considered separately.

1. Level 1 changes—A level 1 change consists of a site change within a single facility where the same equipment, SOPs, environmental conditions, and personnel are used and where no changes are made to the manufacturing batch records other than location of the facility and administrative changes.
   1.1. Required documentation—No documentation is required beyond the usual application/compendial requirements. No in vivo bioequivalence documentation is required.
   1.2. Filing requirements—Annual report.
2. Level 2 changes—A level 2 change is a site change within a contiguous campus or between facilities in adjacent city blocks where the same equipment, SOPs, environmental conditions and controls, and personnel common to both manufacturing sites are used. There must be no changes to the manufacturing batch records except for administrative information and the location of the facility.
   2.1. Required documentation
      2.1.1. Chemistry—Identify location of new site and updated batch records. No other documentation is required beyond application/compendial release requirements, although one batch produced at the new site should be placed on long-term stability and the data should be reported in the annual report. Dissolution data other than normal release requirements are not required nor is in vivo bioequivalence testing required.
      2.1.2. Filing documentation—A supplement should be filed showing the changes being effected. Long-term stability test data should be included in the annual report.
3. Level 3 changes—A level 3 change is a change in manufacturing site to a different campus. However, the same equipment, SOPs, environmental conditions, and controls should be used in the manufacturing process at the new site. No changes may be made to the manufacturing batch records except for administrative information, location, and language translation if needed.
   3.1. Documentation
      3.1.1. Chemistry—Location of new site and updated batch records.
      3.1.2. Stability
         3.1.2.1. If a significant body of data is available, one batch with three months accelerated stability data must be reported in a supplement. One batch should be on long-term stability with the stability data reported in the annual report.
         3.1.2.2. If a significant body of data is not available, up to three batches with three months accelerated stability data should be reported in the supplement. Up to three batches should be on long-term stability with these data being reported in the annual report.
      3.1.3. Dissolution—A multipoint dissolution profile should be performed in the application/compendial medium at 15, 30, 45, 60, and 120 min or until an asymptote is reached. The dissolution profile of the drug product at the current and proposed site should be similar.
      3.1.4. In vivo bioequivalence—None required.

3.2. Filing documentation required—Changes being effected should be identified in a supplement. Long-term stability data are reported in the annual report.

***Changes in Batch Size*** Postapproval changes in the size of a batch from the pilot scale used to manufacture product for clinical trials to larger or smaller commercial batch sizes require submission of additional information in the application. Scale-down below 100,000 dosage units is not covered by this guidance. All scale-up changes should be properly validated and, where needed, inspected by appropriate FDA personnel.

1. Level 1 changes—A change in batch size, up to and including a factor of 10 times the size of the pilot batch, is considered a level 1 change. However, (1) the equipment used must be of the same design and operating principles, (2) the product is manufactured in full compliance with the prevailing GMPs, and (3) the same formulation and manufacturing procedures are used as well as the same SOPs and controls.
   1.1. Chemistry documentation—(1) Application/compendial release requirements, (2) notification of change to the FDA and submission of updated batch records in the annual report, and (3) one batch should be on long-term stability with results being provided in the annual report.
   1.2. Dissolution documentation—None beyond application/compendial release requirements.
   1.3. In vivo bioequivalence—None.
   1.4. Filing documentation—Annual report with long-term stability data.
2. Level 2 changes—Level 2 consists of changes in batch size beyond a factor of 10 times the size of the pilot batch where (1) the equipment used to produce the pilot batches is of the same design and operating principles, (2) the product is manufactured in full compliance with the prevailing GMPs, and (3) the same formulation and manufacturing procedures are used as well as the same SOPs and controls.
   2.1. Chemistry—Application/compendial release requirements. Notification of change in batch size and submission of updated batch records to the FDA. One batch must be placed on accelerated stability testing and one on long-term stability.
   2.2. Dissolution—None beyond application/compendial release requirements.
   2.3. In vivo bioequivalence—None.
   2.4. Filing requirements—Must submit changes being effected in the supplement. Long-term stability data are reported in the annual report.

***Manufacturing*** Manufacturing changes may be either the equipment used in the manufacturing process or the process itself:

1. Equipment
   1.1. Level 1 equipment changes—This category includes change from the use of nonautomated or nonmechanical equipment to automated or mechanical equipment to move ingredients and a change to alternative equipment of the same design and operating principles of the same or different capacity.

1.1.1. Chemistry documentation—Application/compendial release requirements, notification of change, and submission of updated batch records. One batch should be placed on long-term stability.
1.1.2. Dissolution documentation—None other than application/compendial release requirements.
1.1.3. In vivo bioequivalence documentation—None.
1.1.4. Filing documentation—Annual report with long-term stability data.

1.2. Level 2 equipment changes—This type of change involves a change in equipment to a different design and different operating principles.
1.2.1. Chemistry documentation—Application/compendial release requirements, notification of change, and submission of updated batch records.
1.2.1.1. If a significant body of data are available, one batch with three months of accelerated stability data reported in the supplement and one batch on long-term stability with data reported in the annual report.
1.2.1.2. If a significant body of data are not available, submit up to three batches with three months accelerated stability data in the supplement and up to three batches on long-term stability with data reported in the annual report.
1.2.2. Dissolution documentation—A multipoint dissolution profile should be performed in the application/compendial medium at 15, 30, 45, 60, and 120 min or until an asymptote is reached. The dissolution profile of the drug product at the current and proposed site should be similar.
1.2.3. In vivo bioequivalence documentation—None.
1.2.4. Filing documentation—Prior approval supplement with justification for change; long-term stability data must be reported in the annual report.

2. Process changes
2.1. Level 1 process changes—This includes process changes such as changes in mixing times and operating speeds within application/validation ranges.
2.1.1. Chemistry documentation—None beyond application/compendial release requirements.
2.1.2. Dissolution documentation—None beyond application/compendial release requirements.
2.1.3. In vivo bioequivalence documentation—None.
2.1.4. Filing documentation—Annual report.

2.2. Level 2 process changes—Level 2 changes include process changes such as mixing times and operating speeds outside of application/validation ranges.
2.2.1. Chemistry documentation—Application/compendial release requirements; notification of change and submission of updated batch records. One batch on long-term stability.
2.2.2. Dissolution documentation—A multipoint dissolution profile should be performed in the application/compendial medium at 15, 30, 45, 60, and 120 min or until an asymptote is reached. The dissolution profile of the drug product at the current and proposed site should be similar.

2.2.3. In vivo bioequivalence documentation—None.
2.2.4. Filing documentation—A supplement with changes being effected. Long-term stability data should be reported in the annual report.

2.3. Level 3 process changes—Level 3 includes change in the type of process used in the manufacture of the product, such as a change from wet granulation to direct compression.
2.3.1. Chemistry documentation—Application/compendial release requirements. Notification of change and submission of updated batch records. Stability testing varies depending on the amount of data available:
2.3.1.1. Significant body of data available—One batch with three months accelerated stability data should be reported in the supplement; one batch should also be put on long-term stability with data being reported in the annual report.
2.3.1.2. No significant body of data available—Up to three batches with three months accelerated stability data should be reported in the supplement. Up to three batches should be on long-term stability with data being reported in the annual report.
2.3.2. Dissolution documentation—A multipoint dissolution profile should be performed in the application/compendial medium at 15, 30, 45, 60, and 120 min or until an asymptote is reached. The dissolution profile of the drug product at the current and proposed site should be similar.
2.3.3. In vivo bioequivalence documentation—An in vivo bioequivalence study should be performed. This may be waived if a suitable in vivo/in vitro correlation has been verified.
2.3.4. Filing documentation—A prior approval supplement must be filed with justification for the change. Long-term stability data should be submitted in the annual report.

## 9 OTHER GMP-RELATED GUIDANCE DOCUMENTS

This chapter has discussed the CGMP regulations and some of the more important guidances. There have been a number of additional guidance documents related to GMPs published by the FDA. These documents are all posted on the FDA website. They are listed below along with their URL:

- Current good manufacturing practice for combination products: http://www.fda.gov/cder/guidance/OCLove1dft.pdf
- Formal dispute resolution: Scientific and technical issues related to pharmaceutical CGMP: http://www.fda.gov/cder/guidance/5880fnl.pdf
- Questions and answers on current good manufacturing practices (cGMP) for drugs: http://www.fda.gov/cder/guidance/cGMPs/default.htm
- Powder blends and finished dosage units—Stratified in-process dosage unit sampling and assessment: http://www.fda.gov/cder/guidance/5831dft.pdf
- Sterile drug products produced by aseptic processing—Current good manufacturing practice: http://www.fda.gov/cder/guidance/5882fnl.pdf

- Current good manufacturing practice for medical gases: http://www.fda.gov/cder/guidance/3823dft.pdf
- General principles of process validation: http://www.fda.gov/cder/guidance/pv.htm
- SUPAC-IR: Immediate-release solid oral dosage forms: Scale-up and post-approval changes: Chemistry, manufacturing and controls, in vitro dissolution testing, and in vivo bioequivalence documentation: http://www.fda.gov/cder/guidance/cmc5.pdf
- SUPAC-IR/MR: Immediate release and modified release solid oral dosage forms manufacturing equipment addendum: http://www.fda.gov/cder/guidance/1721fnl.pdf
- SUPAC-MR: Modified release solid oral dosage forms scale-up and postapproval changes: Chemistry, manufacturing, and controls; in vitro dissolution testing and in vivo bioequivalence documentation: http://www.fda.gov/cder/guidance/1214fnl.pdf
- SUPAC-SS: Nonsterile semisolid dosage forms; scale-up and post-approval changes: Chemistry, manufacturing and controls; in vitro release testing and in vivo bioequivalence documentation: http://www.fda.gov/cder/guidance/1447fnl.pdf
- SUPAC-SS: Nonsterile semisolid dosage forms manufacturing equipment addendum: http://www.fda.gov/cder/guidance/1722dft.pdf

# ENFORCEMENT OF CURRENT GOOD MANUFACTURING PRACTICES

Kenneth J. Nolan
*Nolan & Auerbach, P. A., Fort Lauderdale, Florida*

**Contents**

1 Introduction and Background
2 Enforcement Players
3 FDA Enforcement Techniques
  3.1 Inspections
  3.2 After the Inspection: Form 483
  3.3 Recalls
  3.4 Warning Letter
4 Judicial Enforcement: Beyond the Warning Letter
  4.1 Introduction
  4.2 Civil Proceedings
  4.3 Criminal Proceedings
5 Conclusion

## 1 INTRODUCTION AND BACKGROUND

The legal authority for the Food and Drug Administration (FDA) to impose minimum manufacturing standards is set forth in the federal Food and Drug and Cosmetic Act (FDCA), 21 U.S.C. sec. 301 et seq. Section 351(a)(2)(B) of 21 U.S.C. requires manufacturers of drugs to operate in conformance with manufacturing regulations established by the FDA. The regulations are primarily contained in Title 21 of the U.S. *Code of Federal Regulations* (CFR), Parts 210 and 211, and are called the current good manufacturing practice (cGMP) regulations.

The cGMP regulations stem from congressional concern that impure and otherwise adulterated drugs might escape detection under a system predicated only on seizure of drugs shown to be *in fact* adulterated. That is, the U.S. Congress desired

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

to require manufacturers to utilize manufacturing practices designed to prevent pharmaceuticals from such defects as contamination, nonconforming bioavailability, or potency defects.

Congress stated the rationale for imposing cGMP on the pharmaceutical industry this way[1]:

> The manufacturing of drugs is a business that requires highly qualified and trained personnel, and special laboratory and other facilities and most careful internal manufacturing, packaging, and labeling controls. These requirements are necessary to the assurance that the drugs will be safe for the user and will have, and so far as possible retain, the identity, strength, quality, purity, and effectiveness that they purport to have.

The purpose of the cGMP requirement is to prevent injury and death "by building quality into the design and production of pharmaceuticals," [2] so that substandard prescription drugs do not jeopardize the health and safety of the patients.

The cGMPs require manufacturers to have adequately equipped manufacturing facilities, adequately trained personnel, precisely controlled manufacturing processes, appropriate laboratory controls, complete and accurate records and reports, appropriate finished product examination, and so on. Current GMPs are not "best practices"; rather, they establish threshold or *minimum* standards which must be satisfied in order for a pharmaceutical manufacturing operation to be compliant.

The cGMPs were modified only once between 1963 and 2002 with changes made in 1978 to update them in light of the current technology and also to describe the requirements more explicitly and with more specificity. Meanwhile, the intervening decades saw myriad advances in manufacturing science, engineering, and technology, including the development of better quality systems. These advances, combined with the desire to harmonize manufacturing standards in an increasingly globalized production environment, created the impetus to revamp the cGMPs again.

In August 2002, the FDA announced a comprehensive review of the pharmaceutical cGMPs. The agency identified its cGMP initiative "Pharmaceutical cGMPs for the 21st Century: A Risk-Based Approach." The FDA's articulated goals for the initiative, relevant to enforcement, were:

- The submission review program and the inspection program operate in a coordinated and synergistic manner.
- Regulation and manufacturing standards are applied consistently.
- FDA resources are used most effectively and efficiently to address the most significant health risks.

One of the major products of the cGMP initiative was issued by the FDA in September 2006 in a document entitled "Guidance for Industry—Quality Systems Approach to Pharmaceutical cGMP Regulations."[3] The FDA described the guid-

[1]H. R. Rep. No. 2464, 87th Cong., 2d Sess. 2 (1962). See also 1962 U.S. Cong. and *Admin. News*, p. 2884.
[2]FDA, Pharmaceutical cGMPs for the 21st century: A risk-based approach, Rockville, MD, August 21, 2002.
[3]The FDA's guidance documents advise the reader that they "do not establish legally enforceable responsibilities. Instead, guidances describe the [the FDA's] current thinking on a topic and should be viewed only as recommendations, unless specific regulatory or statutory requirements are cited."

ance as a comprehensive quality systems model which, if followed, would improve quality control and satisfy the requirements of the cGMP regulations. Quality systems and quality assurance are important parts of the cGMP modernization process because quality assurance problems have been the cGMP issues most frequently cited by FDA investigators in recent years.

Drugs which are manufactured not in accordance with any cGMP requirement, including the quality control and quality process mandates, are "adulterated" under the FDCA. Section 351 of 21 U.S.C. defines a drug as adulterated

> [if] the methods used in, or the facilities or controls used for, its manufacture, processing, packing, or holding do not conform to or are not operated or administered in conformity with current good manufacturing practice to assure that such drug meets the requirement of the act as to safety and has the identity and strength, and meets the quality and purity characteristics, which it purports or is represented to possess.

## 2 ENFORCEMENT PLAYERS

The FDA is obviously one of the most important regulatory agencies in the United States. It may also be characterized as the most important consumer protection agency in the world. Its decisions involving approval of drugs have a direct effect on testing, approval, access, and distribution of prescription drugs worldwide. As a regulatory agency in a largely scientific role, it is involved in shaping pharmaceutical science and drug access throughout the world. As a scientific agency, the FDA employs physicians, pharmacists, biologists, biochemists, engineers, biostatisticians, and other highly educated and specialized professionals.

But the FDA also has very important law enforcement responsibilities. The agency employs civil and criminal investigators, auditors, attorneys, and other enforcement professionals. One of the FDA's many enforcement functions is investigation, remediation, and prosecution of cGMP violations.

The FDA district offices operate under the auspices of the agency's Office of Regulatory Affairs (ORA). The ORA field organization is divided into five regional offices (northeast, central, southeast, southwest, Pacific). Each region includes district offices, of which there are 20 nationwide. Most district offices have three or four branches, including either a compliance branch or an enforcement branch. The branch offices are the primary regulatory contacts within the districts and act as the "eyes and ears" for FDA headquarters.

The FDA's Office of Criminal Investigations (OCI) is responsible for reviewing allegations which if proven would violate the U.S. criminal code, including potential violations of the cGMPs. The OCI investigators conduct such investigations as is deemed appropriate, sometimes in connection with other federal investigative agencies, including the FBI and the Office of Inspector General of the Department of Health and Human Services. If the OCI chooses not to recommend to the Department of Justice (DOJ)[4] that criminal indictment be pursued, then the district office is at liberty to pursue the matter through administrative or civil proceedings.

[4]The DOJ is under the direction of the attorney general of the United States. Its mission, relevant to this chapter, is to enforce federal statutes and uphold the rule of the law. It pursues violations brought to its attention by the FDA as well as other federal agencies.

Although the FDA's Office of General Counsel is involved with enforcement of both civil and criminal matters, cases involving court enforcement are handled by assistant U.S. attorneys (AUSAs), who are located in U.S. attorneys' offices located across the United States. U.S. attorneys are the local representatives of the DOJ; they are appointed by and serve at the discretion of the president, with advice and consent of the Senate. There are 93 U.S. attorneys, and they are located (by district) across the United States and its territories. Each U.S. attorney is the chief federal law enforcement officer of the United States within his or her particular district.

The AUSAs are the principal trial attorneys for the U.S. government. Each U.S. attorney exercises wide discretion in the use of his or her resources to further the priorities of the local jurisdiction. Discretion and expertise are big factors in case decisions. There may be significant disparity in the experience, interest, and capability of U.S. attorneys' offices with respect to their pursuit of cGMP violations.

The impact of this disparity is mitigated or eliminated by the expertise of the DOJ's Office of Consumer Litigation (OCL), which is charged with coordinating and supporting FDCA prosecutions nationwide. The OCR's attorneys exercise considerable influence over and discretion in deciding what to and what not to prosecute, thus fostering consistent prosecutive decision making. Many civil actions, particularly those seeking injunctive relief, cannot be brought by a U.S. attorney without OCL approval, minimizing the risk that an inconsistent policy position is taken by a U.S. attorney's office.

## 3 FDA ENFORCEMENT TECHNIQUES

### 3.1 Inspections

The FDA has the right to conduct surveillance inspections of manufacturing facilities for the purpose of enforcement. The goal of inspections is "to minimize consumers exposure to adulterated products."[5]

The FDCA, 21 U.S.C. 374, provides that the FDA is authorized to enter and "to inspect, at reasonable times and within reasonable limits and in a reasonable manner … all pertinent equipment, finished and unfinished materials, containers, and labeling" in the manufacturing or related facility. This statute further authorizes the inspection to "extend to all things therein (including records, files, papers, processes, controls, and facilities)" as long as the records, for example, are relevant to any potential adulteration or misbranding[6] or other FDCA violations. The statute denies the agency the right to review "financial data, sales data, pricing data, personnel data (other than data as to qualifications of technical and professional personnel performing functions)" and certain other types of documents.

Inspectors are required to notify the company that the inspection is occurring but need not provide their reasons. They may take samples and photographs related

[5]*Compliance Program Guidance Manual for FDA Staff: Drug Manufacturing Inspection Program*, 7356.002, available: www.fda.gov.

[6]Misbranding involves labeling a pharmaceutical product in a misleading way. See 21 U.S.C. 331(k).

to the subject of the inspection. It is a criminal offense to deny entry to FDA inspectors or other officials who have appropriately made attempts to conduct an inspection. [21 U.S.C. 331(f)]

In addition to the for-cause inspections, the FDCA mandates that the FDA routinely inspect a manufacturer's facilities for cGMP compliance every two years. This applies to domestic and foreign facilities which manufacture drugs for sale within the United States.[7] Unfortunately, this two-year mandate is rarely satisfied because the FDA's district offices, which are charged with the responsibility for the inspections, lack sufficient resources to conduct regular cGMP compliance inspections.

The FDA conducts two categories of facility inspections—surveillance inspections and compliance inspections. Surveillance inspections are periodic. Whether and when to inspect a particular manufacturing facility is decided in part by application of an analytical model to determine high risk sites. In late 2004, the FDA issued a report entitled "Risk-Based Method for Prioritizing cGMP Inspections of Pharmaceutical Manufacturing Sites—A Pilot Risk Ranking Model," which allows the agency to rank manufacturing plants' risk of noncompliance by using an analytical process to (1) pose a risk question, (2) identify potential hazards and risks, (3) characterize factors that can be used as variables for quantifying risk, and (4) mathematically combine the variables to yield an overall risk score. Since the publication of the report, the FDA has added adverse events reports data to the model. Surveillance inspections are supposed to involve audit coverage of two or more systems,[8] with mandatory coverage of the quality system.[9]

Compliance inspections are for the purpose of evaluating or verifying compliance corrective actions after a problem has been identified and regulatory action has been taken. Compliance inspections cover the areas found deficient and subjected to corrective actions. One type of compliance inspection is a "for-cause" inspection, which is conducted to investigate a specific problem that has come to the attention of the FDA. The sources that trigger a compliance inspection include field alert reports, industry complaints, and recalls.

In fiscal year 2005, the FDA field office conducted 1437 cGMP inspections, resulting in 15 warning letters, six injunctions, and one seizure. These enforcement actions are discussed later in this chapter. Data for the years 2000–2005 are set forth in Figures 1 and 2.

### 3.2 After the Inspection: Form 483

If the inspector determines that there are deviations from cGMP, he will complete a form FDA-483 (Inspectional Observations) detailing the violations. The findings are presented to the manufacturer, which is given an opportunity to respond. The FDA-483 advises:

[7]The other cGMP basic enforcement strategy is collection and analysis of drug samples during factory inspections as well as collecting and analyzing drug products in distribution.

[8]The FDA has separated the cGMP regulation into six systems: quality, facilities and equipment, production, materials, packaging and labeling, and laboratory controls.

[9]*Compliance Program Guidance Manual*, 7356.002, February 1, 2002.

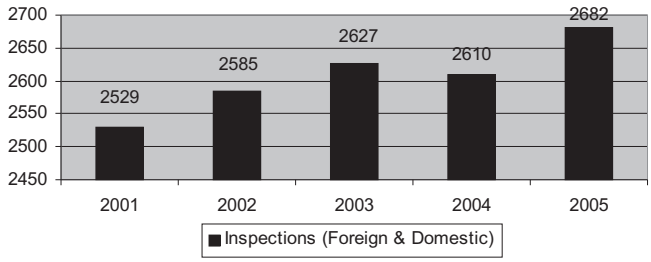


**FIGURE 1** CDER five-year Inspection data. (*Source*: FDA.)

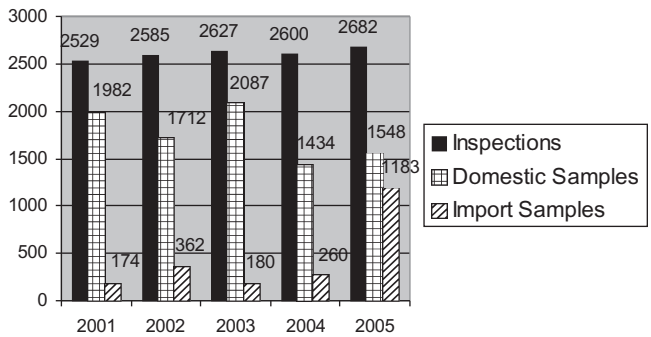


**FIGURE 2** Surveillance activity. (*Source*: FDA.)

> This document lists observations made by the FDA representative(s) during the inspection of your facility. They are inspectional observations, and do not represent a final Agency determination regarding your compliance. If you have an objection regarding an observation, or have implemented, or plan to implement, corrective action in response to an observation, you may discuss the objection or action with the FDA representative(s) during the inspection or submit this information to FDA at the address above. If you have any questions, please contact FDA at the phone number and address above.

Most manufacturers provide a written response to the FDA-483, either disputing the findings or addressing how they will correct the issues and how problems are to be corrected. Negotiations typically proceed for months or years until the inspectional problems and issues are resolved or the FDA elects to pursue elevated enforcement. The agency retains discretion to pursue elevated enforcement if it concludes that there is a significant risk of harm to patients, with such action being more likely where patient harm is more likely or more serious.

In addition to providing a form FDA-483, FDA investigators prepare an establishment inspection report (EIR), which is sent to FDA headquarters, which then evaluates the report and determines the corrective action, if any. The FDA then classifies the inspection as "no action indicated," "voluntary action indicated," or "official action indicated." The EIR contains much greater detail than contained in the 483 and is not provided to the manufacturer until after the inspection is deemed closed.

When the FDA conducted an analysis of past FDA-483 reports,[10] the two most reported violations were:

1. Violations of 21 CFR 211.100(b) (failure to follow and/or document production and process control procedures), occurring in over half of all of the 483's
2. Violations of 21 CFR 122d (failure to create adequate, written responsibilities and procedures for the quality control unit or failure to follow them), occurring in 42% of the 483's

The next eight violations, in order of prevalence, were as follows:

- Failure to have written procedures for production and process controls.
- Failure to have testing and release of drug product for distribution for determination of satisfactory conformance to the final specifications/identity and strength of each active ingredient prior to release.
- Batch production and control records were not prepared or are incomplete.
- Control procedures are not established to monitor the output/validate the performance of manufacturing processes that may be responsible for causing variability in the characteristics of the drug product.
- Employees were not given appropriate training.
- Laboratory controls do not include the establishment of scientifically sound and appropriate specifications/standards/sampling plans/test procedures.
- Drug product production and control records are not certified by the quality control unit to assure compliance with all established, approved written procedures before a batch is released or distributed.
- Procedures describing the handling of all written and oral complaints regarding a drug product either not established or not followed.

### 3.3 Recalls

Chapter 7 of the *Regulatory Procedures Manual* (March 2007, available at www.fda.gov) provides detailed instructions to FDA personnel regarding recalls. The FDCA does not authorize the FDA to "order" a manufacturer to recall a drug product.[11] In practice, however, the manufacturers or distributors of the drug products are encouraged to implement and carry out recalls voluntarily to fulfill their responsibility to protect the public. It is not uncommon for a company to discover that one of its products is defective and recall it entirely on its own; or the FDA informs a company of its findings that one of its products is defective and suggests or requests

[10]The data for the analysis were compiled by the FDA and derived from 614 Turbo EIR reports completed from 2001 to 2003. (FDA investigators enter their inspections observations on the FDA's Turbo EIR system. The Turbo's electronic format prompts investigators to select the specific cGMP violation in question and then to explain their findings uncovered during the inspection.)

[11]The FDCA gives authority to the FDA to order a recall in some cases involving infant formulas, biological products, and devices that present a "serious hazard to health," but not involving pharmaceuticals.

| | |
|---|---|
| Class I: | 18 |
| Class II: | 314 |
| Class III | 170 |
| **TOTAL:** | 502 |

**FIGURE 3** 2005 recall by class. [*Source*: Centre for Drug Evaluation and Research (CDER) 2005 Report to the Nation.]

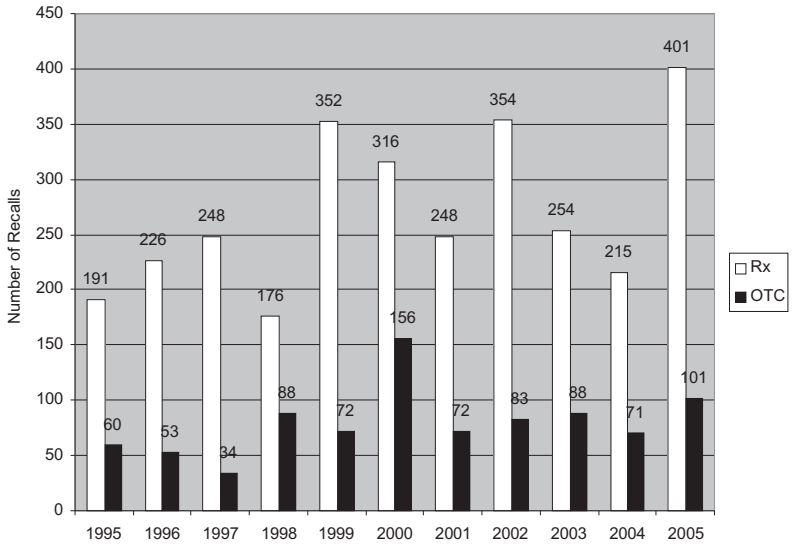


**FIGURE 4** Drug recalls. One firm had over 100 recalls in 2005, which caused a spike in the 2005 recall figures. (*Source*: CDER 2005 Report to the Nation.)

a recall.[12] Once a voluntary recall is initiated, the FDA generally follows the following protocol (Figures 3–5):

1. *Classify the Recall* The FDA reviews relevant information and then assigns a recall classification according to the level of health risk involved:
   - Class I recalls involve drug products in which the reason for recall predictably could cause serious health problems or death.
   - Class II recalls involve drug products which defect might cause a temporary health problem or pose only a slight threat of a serious nature.
   - Class III recalls involve products that are unlikely to cause any adverse health reaction but that violate FDA labeling or manufacturing regulations.
2. *Monitor and Audit the Recall* The FDA oversees a recall depending upon the health risk involved. For a class I recall, the FDA checks to make sure that the defective product has been recalled in full. In contrast, for a class III recall, FDA oversight may be to simply spot-check.

[12]If the company does not comply, then FDA can seek judicial enforcement under the FDCA.

- Miscellaneous cGMP deviations (other than below)
- Failed USP dissolution test requirements
- Microbial contamination of non-sterile products
- Lack of efficacy
- Impurities/degradation products
- Lack of assurance of sterility
- Lack of product stability
- Labeling: Label error on declared strength
- Misbranded: Promotional literature with unapproved therapeutic claims
- Labeling: Correctly labeled product in incorrect carton or package

**FIGURE 5** Top 10 reasons for drug recalls in fiscal year 2005. (*Source*: FDA.)

3. *Notification and Public Warning* Class I recalls almost always warrant a press release to the media. Classes II and III are not necessarily announced in the media, but all of them are included in the FDA's weekly enforcement report, posted at www.fda.gov/opacom/Enforce.html on the FDA's website.
4. *Termination* The FDA provides written notice to the recalling manufacturer on when the recall should be terminated.
5. *Noncompliance* If applicable, the FDA will take appropriate legal action if a manufacturer fails or refuses to timely complete a recall.

### 3.4 Warning Letter

A warning letter is intended to notify manufacturers about violations that the FDA has documented during its inspections or investigations. A warning letter will notify a responsible individual and/or firm that the FDA considers one or more products, practices, processes, or other activities to be in violation of the cGMPs. Warning letters should only be issued for violations of regulatory significance, that is, those that may actually lead to an enforcement action if the documented violations are not promptly and adequately corrected. A warning letter is one of the FDA's principal means of achieving prompt voluntary compliance.

Examples of situations in which the FDA may be expected to issue a warning letter include:

- An active pharmaceutical ingredient (API) batch fails to conform to established specifications and yet the manufacturer distributed it anyway.
- Deliberately blending API batches to dilute or hide noxious contaminant or filth or failing to determine actual yield and percentages of expected yields.
- Contamination of drugs with toxic chemicals, drug residues, airborne contaminants, or filth.
- Failing to comply with commitments in drug applications.
- Combining a batch that does not conform with critical attributes with a batch that does.
- Failing to demonstrate water used in the manufacturing process is suitable.
- Failing to validate water systems.

- Lacking a formal written program to validate an API validation process.
- Failing to demonstrate homogeneity of final blending operations.
- Failing to keep adequate batch records.
- Failing to have a formal process change control system in place.
- Using inadequate or unvalidated laboratory test methods.
- Packaging and labeling processes that could introduce a significant risk of mislabeling.
- Failing to test for residues of organic or inorganic solvents that may carry over to the API.
- Using incomplete stability studies to establish API stability for the intended period of use.

Warning letters detailing cGMP violations typically conclude with the following: "The article(s), (DRUG NAME), is (are) adulterated within the meaning of Section 501(a)(2)(B) of the Act, 21 U.S.C. 351(a)(2)(B), in that the methods used in, or the facilities or controls used for, its manufacture, processing, packing, or holding fails to conform to, or is not operated or administered in conformity with, cGMP regulations [21 CFR 210, 211]." The number of warning letters issued by the FDA concerning prescription and over-the-counter drugs has ranged from 130 letters in 2000 to 79 letters in 2005.

A warning letter is distinguishable from a notice of violation, also called an untitled letter. An untitled letter cites violations that do not meet the threshold of regulatory significance for a warning letter, but the FDA has a need nevertheless to communicate. Unlike a warning letter, an untitled letter does not include a warning statement that failure to take prompt correction may result in enforcement action and does not evoke a mandated FDA follow-up. Further, the untitled letter requests (rather than requires) a written response (from the manufacturer) within a reasonable amount of time (e.g., "Please respond within 45 days").

## 4 JUDICIAL ENFORCEMENT: BEYOND THE WARNING LETTER

### 4.1 Introduction

The FDA is likely to bypass sending a Warning Letter in certain circumstances. According to Chapter 4 of the FDA *Regulatory Procedures Manual*, the following violations are likely to result in an enforcement action without necessarily issuing a warning letter:

1. The violation reflects a history of repeated or continual conduct of a similar or substantially similar nature during which time the individual and/or firm has been notified of a similar or substantially similar violation.
2. The violation is intentional or flagrant.
3. The violation presents a reasonable possibility of injury or death.
4. Adequate notice has been given by other means and the violations have not been corrected or are continuing.

Written records are required to be kept as set forth in 211.180 to 211.208. Highlights are as follows:

§ 211.182 Equipment cleaning and use log.

A written record of major equipment cleaning, maintenance (except routine maintenance such as lubrication and adjustments), and use shall be included in individual equipment logs that show the date, time, product, and lot number of each batch processed.

§ 211.184 Component, drug product container, closure, and labeling records.

These records shall include the following:

(a) The identity and quantity of each shipment of each lot of components, drug product containers, closures, and labeling; the name of the supplier; the supplier's lot number(s) if known; the receiving code as specified in § 211.80; and the date of receipt. The name and location of the prime manufacturer, if different from the supplier, shall be listed if known.

(b) The results of any test or examination performed (including those performed as required by § 211.82(a), § 211.84(d), or §211.122(a)) and the conclusions derived therefrom.

§ 211.186 Master production and control records.

To assure uniformity from batch to batch, master production and control records for each drug product, including each batch size thereof, shall be prepared, dated, and signed (full signature, handwritten) by one person and independently checked, dated, and signed by a second person. The preparation of master production and control records shall be described in a written procedure and such written procedure shall be followed.

§ 211.188 Batch production and control records.

Batch production and control records shall be prepared for each batch of drug product produced and shall include complete information relating to the production and control of each batch.

§ 211.194 Laboratory records.

Laboratory records shall include complete data derived from all tests necessary to assure compliance with established specifications and standards, including examinations and assays ...

§ 211.198 Complaint files.

A written record of each complaint shall be maintained in a file designated for drug product complaints

**FIGURE 6** Written record highlights.

5. The violations, under Title 18 U.S.C. 1001, are intentional and willful acts that once having occurred cannot be retracted. Also, such a felony violation does not require prior notice. Therefore, Title 18 U.S.C. 1001 violations are not suitable for inclusion in warning letters.

In addition, actively deceiving the FDA is almost guaranteed to bring judicial enforcement actions. This includes false representations in the written record-keeping requirements or in written communication with the FDA. Manufacturing record-keeping requirements which give exposure to fraud liability are summarized in Figure 6. Potential violations include the following:

(a) Accepting and validating drug products that failed to meet established standards or specifications and any other relevant quality control criteria (i.e., dissolution rates, content uniformity, purity, potency) and then falsely recording the untruthful data as if the drug products did not fail
(b) Accepting and validating the stability characteristics of drug products and then falsely recording the untruthful data as if the drug products did not fail
(c) Documenting the examination and review of labels, when in truth and fact no review occurred (which results in inaccurate labels distributed with drugs)
(d) Falsely documenting any components of master production and central records
(e) Falsely documenting any component of the batch production and control records
(f) Falsely describing testing methods when no (or inadequate) testing methods were performed
(g) Failing to accurately make a written record of all written and oral complaints regarding a drug product and/or certifying that investigations were performed when they were not, falsely certifying that the findings were negative when they were not, and so on
(h) Falsifying records which would indicate manufacturing changes which require approval by the FDA
(i) False representations that contain statements of fact in correspondence sent to the FDA addressing violations in an inspector's form 483

The FDA typically initiates progressive enforcement, as described in Figure 7. Once the FDA and DOJ decide to bring enforcement action, the U.S. courts have held that the FDA's interpretation of its cGMPs is entitled to substantial deference. As long as the FDA's interpretation of its regulations are "reasonable" and "sensibly conforms to the purpose and wording of the regulations," courts are required to follow the FDA's interpretations.

### 4.2 Civil Proceedings

***Seizures*** If during an inspection of a facility the FDA inspector or employee making the inspection has reason to believe that a drug found in such facility is adulterated, such inspector or employee may order the drug detained for a reasonable period which may not exceed 20 days (unless the FDA institutes an action under Subsection 334(a) or an injunction, in which case a longer detention period may be authorized).

The FDCA expressly permits administrative seizure on the basis of an ex parte showing of reasonable belief [21 U.S.C. 334 (g)]. Seizure of a company's inventory deprives the company of both capital investment and potential profit.

If the FDA pursues relief beyond detainment, the United States can file a complaint for forfeiture directing the U.S. marshall to "seize" the pharmaceuticals (or take possession or place in constructive custody of the court). The theory in a complaint for forfeiture is that there is a violation of the law by the pharmaceutical

FDA enforcement mechanisms are often utilized progressively. A good example is the enforcement action against Glaxo SmithKline ("GSK"), which began in July 2002, identifying numerous significant cGMP violations found during a February/April 2002 inspection. A Warning Letter requested that the violations be corrected and stated that failure to correct the violations may result in regulatory action, including seizure and/or injunction. Although a limited follow-up FDA inspection in October 2002, found that some specific corrections were acceptable, the subsequent FDA inspections in November/December 2003 and September/November 2004, revealed continuing significant cGMP violations. FDA concluded that the firm's data and corrective plans were not adequate to correct the cGMP violations. GSK also initiated recall of some, but not all, lots of the two products. On March 4, 2005, in response to ongoing concerns about manufacturing quality, FDA and the DOJ initiated seizures of two GSK pharmaceuticals. The Agency initiated these seizures actions based on concerns that GSK's violation of manufacturing standards may have resulted in the production of poor quality drug products that could potentially pose risks to consumers. On April 28, 2005, FDA announced that GSK had signed a Consent Decree with FDA to correct manufacturing deficiencies at its Cidra, Puerto Rico, facility. The Consent Decree was initiated based on FDA's continued concerns that GSK's violation of manufacturing standards may have resulted in the production of drug products that could potentially pose risks to consumers.

**FIGURE 7** Progressive enforcement.

product itself. Accordingly, the government asks the court to condemn the article and declare forfeiture. Upon filing of the complaint, the clerk automatically issues a warrant. Thus, the FDA is able to obtain a warrant without review by a judicial officer or even a finding of probable cause.

There are three types of seizures: mass, open ended, and lot specific. A mass seizure is the seizure of all FDA-regulated products at an establishment/facility. Mass seizures might be conducted when all of the products are produced under the same conditions (e.g., nonconformance with cGMPs). An open-ended seizure is the seizure of all units of a specific product or products, regardless of lot or batch number, when the violation is expected to be continuous. An open-ended seizure may be conducted when a specific product extends to all lots or batches of a product but not to all of the products in the facility.

Following seizure of its drugs a manufacturer has three courses of action. First, it may do nothing, in which case the drug will be disposed of. Second, it can enter into a consent decree, admitting the violation, agreeing to pay costs, and seeking to destroy or rehabilitate the article. The consent decree will typically provide for (1) condemnation of the article as being in violation of the law; (2) a penal bond in approximately twice the retail value of the article under seizure; (3) provisions for payment of costs for storage and handling by the U.S. marshall and for supervision by the FDA before release of the product; and (4) a provision that the manufacturer will attempt to bring the article into compliance under the supervision of and to the satisfaction of the FDA.[13]

Third, it can contest the action. If the manufacturer contests the action, the case is then treated like any other civil case under the federal rules of civil procedure, and the government must prove its case by a preponderance of the evidence. The government must produce evidence, in support of its allegations, including proof of interstate shipment of the drug or its components. FDA employees may testify, but

[13] *Regulatory Procedures Manual*, Chapter 6-1-11, March 2007. This contemplates that seizure of a specific product(s) is the sole issue. More complex consent decrees are described hereinafter.

Major recent consent decrees are United States v. Abbott Labs., Consent Decree of Permanent Injunction filed Nov. 2, 1999; United States v. Various Articles of Drug Identified in Attachment A & Wyeth-Ayerst Labs., Consent Decree of Condemnation and Permanent Injunction filed Oct. 4, 2000; and United States v. Schering-Plough Corp., Consent Decree of Permanent Injunction filed May 20, 2002. To avoid giving manufacturers the wrong message by allowing them to keep on the market what FDA had determined to be produced in violation of the cGMPs, the FDA included three separate types of "disgorgement" payments in the Abbott, Wyeth and Schering consent decrees: (1) a lump sum payment (Abbott, $100 million; Wyeth, $30 million; and Schering, $500 million) (2) if the remedial work was not achieved by the deadline established in the decree, (i) a percentage of sales (Abbott, 16%; Wyeth, 18.5%; and Schering, 24.6%) and (ii) daily payments of a certain flat amount. Both were to be paid until compliance was achieved.

**FIGURE 8** Disgorgement.

also outside experts testify such as to the significance of failure to comply with cGMP requirements. If a decree of condemnation is entered (either after trial or by consent), the court may direct disposition of the article by destruction.

***Injunctions*** The FDCA expressly authorizes the courts to restrain and enjoin acts that are in violation of 21 U.S.C. 331, which includes prohibition of adulterated products. FDA policy provides that an injunction action is appropriate where:

(a) there is a current and definite health hazard or a gross consumer deception requiring immediate action to stop the violative practice;
(b) there are significant amounts of violative products owned by the same person in many locations, voluntary recall by the firm was refused or is significantly inadequate to protect the public, and seizures are impractical or uneconomical; or
(c) there are long-standing (chronic) violative practices that have not produced a health hazard or gross consumer fraud, but which have not been corrected through use of voluntary or other regulatory approaches.[14]

A complaint for injunction is typically accompanied by a motion for preliminary injunction.[15] The court schedules a court hearing to determine whether to grant a preliminary injunction, often very quickly and on short notice. The government's main focus at this preliminary stage will be to prove that there is a "substantial likelihood" that the defendant has been producing adulterated drugs in violation of 21 U.S.C. 331, by substantial noncompliance with the cGMPs. The government will also typically present evidence, if applicable, that the defendant has had a history of prior noncompliance with the FDCA and implementing regulations. No specific finding of irreparable harm is necessary as is required in the typical injunction, because the passage of the statute proscribing adulterated products has

[14]*Regulatory Procedures Manual*, March 2007.
[15]The government may also apply for a temporary restraining order (TRO) seeking immediate, temporary relief (for a period of 10 days, which may be extended for 10 additional days) prior to the hearing for preliminary injunction. The FDA will typically recommend a TRO when it believes that the violation is so serious that it must be controlled *immediately*.

been held itself to be an implied finding by Congress that violations will harm the public.

United States courts are imbued with authority to enjoin present and future violations of Section 331 based upon proof by the FDA that such violations have occurred and could recur. Factors that courts consider when determining whether there is a reasonable chance of future infractions include (1) the degree of scienter involved on the part of the defendant; (2) the isolated or recurrent nature of the infraction; (3) the defendant's recognition of the wrongful nature of his or her conduct; (4) the sincerity of the defendant's assurances against future violations; and (5) the nature of the defendant's violation. The court also considers whether the defendant voluntary ceased the challenged conduct, the genuineness of the defendant's efforts to conform to the law, the defendant's progress toward improvement, and the defendant's compliance with any recommendations made by the government.

Good faith is not a defense to the issuance of an injunction. Nor may a defendant successfully defend against the issuance of an injunction by asserting that the injunction would drive it out of business.

***Consent Decrees and Disgorgement*** A consent decree is a judgment (legal order) issued by the court that has been agreed to by the parties whereby the defendant agrees to stop illegal or improper activity as alleged by the government. Once court approval is obtained, the seizure or injunctive lawsuit, for instance, is dropped, and the government's remedy is then based upon any breach of the consent decree, itself which is enforceable by the court.

Consent decrees typically involve a defendant agreeing to address the areas of noncompliance in a manner satisfactory to the FDA within a certain amount of time. It can also provide for the hiring of an expert consultant to certify in detailed reports that the manufacturing facility, at periodic dates, is in full compliance with the cGMPs, and has adequate adverse-event controls, adequate training, and adequate recall procedures. It may also require the payment of money to the U.S. Treasury such as under the equitable remedy of "disgorgement," as described in Figure 8.

As part of a court action, the FDA will sometimes pursue "disgorgement." The purpose of disgorgement is to deprive the wrongdoer of ill-gotten gains as well as provide deterrence. The amount of disgorgement is not necessarily directly tied to restitution. In practice, the amount the FDA exacts is supposed to be enough to send a message but certainly does not provide for full disgorgement of profits of the drug product(s) at issue.

***False Claims Act*** The U.S. Civil False Claims Act, 31 U.S.C. 3729 et seq., is the government's principal means of redressing fraud by government contractors. The act has implications for cGMP violations because the United States (funding as it does the Medicare program, the state Medicaid programs, the Veterans Administration, the TRICARE program, and others) is the world's largest purchaser of prescription medications.

Nonetheless, the government has yet to bring a False Claims Act case which seeks damages for cGMP. One reason may be that the government has multiple other remedies within which to recover damages from noncompliant manufacturers, such as criminal fines and penalties and disgorgement.

*Qui tam* whistleblowers,[16] however, have already begun bringing such cases. Because the False Claims Act imposes liability on any government contractor which knowingly submits false claims to the United States or which uses false documents to get a false claim paid, a pharmaceutical manufacturer which knew or was recklessly indifferent to the fact that the manufacturing process was compromised by cGMP violations is in the same position as any other contractor which is required to conform to contractual or regulatory standards. The basis of liability under the False Claims Act is that false records have been generated which caused (false) claims for drugs to be paid by the United States.[17] The monetary damages result because the payor (in this case, the United States) is potentially paying for substandard drugs due to the cGMP violations—later covered up by false statements in documents required to be completed under the cGMP.

It makes sense, too: The cGMPs are a set of regulations which, by their very nature, are designed to ensure that drugs are manufactured in such a way that they meet the requirements of the federal Food, Drug and Cosmetic Act as to safety and have the identity and strength and meet the purity characteristics that they purport or are represented to possess. The major federally funded government health care programs, Medicare and Medicaid, operate under the express provisions that they will only pay for medical services and products that are "reasonable and necessary." Unsafe or ineffective drug products are neither reasonable nor necessary. Accordingly, as the theory goes, the United States suffers monetary damages if Medicare and Medicaid programs pay for unsafe or less effective products. These and other federally funded health care programs spend billions of dollars every year on pharmaceuticals.

False representations concerning minor or technical violations will not be the basis for FCA liability. Distribution of products that are not totally cGMP compliant (but have been falsely documented to be) does not necessarily result in unsafe (or subpotent) products. Substantial violations of the cGMP, later covered up in writing, however, could very well be the basis for FCA liability. The common thread through each violation is that the violation is severe enough so that the drug product that

[16]*Qui tam* is shorthand for the Latin phrase, *qui tam pro domino rege quam pro seipso*, meaning "He who is as much for the king as for himself." *Qui tam* statutes date back to thirteenth-century England. The actions were a means of enabling private parties to allege the king's interest and therefore gain access to the royal courts.

The *qui tam* provisions of the federal False Claims Act allow any citizen who has knowledge of fraud that has taken place against the government to bring a civil action in federal court in the name of the United States. In return for his or her efforts, the citizen is entitled to share in the proceeds of the recovery. The *qui tam* provisions raise the incentive for insiders to put the spotlight on the criminals, thereby providing the government with tangible and detailed evidence upon which to base an investigation and prosecution.

In 1986, Congress enacted amendments to the False Claims Act which strengthened the law and increased monetary awards. When hearings were held in 1985 and 1986, the climate was favorable for strengthened antifraud legislation, and Congress expected that most *qui tam* cases would involve defense contractor fraud. In the last decade, the majority of cases have instead been against the health care industry.

[17]Even so, factual questions will be raised, including: (1) Even with the false representations, was a false claim "caused" to be submitted? (2) Had the FDA known about the falsities, would it have enjoined the manufacturer from any further production, etc? (3) What about the false record or statement made the claims for such drugs false?

finally reaches the public is foreseeably and substantially less safe or less effective than if the cGMPs were not violated.

### 4.3 Criminal Proceedings

***Introduction*** Criminal prosecutions of violations of the FDCA are intended to further the goal of protecting the health and safety of the public. The FDA historically has not pursued criminal charges unless the defendant shows a continuous or repetitive course of violative conduct, with the exception of intentional violations, fraud, or danger to health.[18]

While the FDCA contains various prohibitions and restrictions which a drug company could violate, the most common FDCA violation arising out of cGMPs is charged by using 21 U.S.C. 331(a), which specifically prohibits introducing an adulterated[19] drug into interstate commerce. In addition to introducing an adulterated drug into interstate commerce, some other acts prohibited by Section 331(a) which could be involved in manufacturing violations include 331(e), which prohibits the refusal to allow access to records mandated elsewhere in the act and 331(f), which prohibits the refusal to allow inspection of production facilities (21 U.S.C. 374).

Commission of any act prohibited by Section 331 is a federal misdemeanor (21 U.S.C. 333). However, violations of Section 331(a) may be charged as felonies where there is intent to defraud or mislead or where the defendant previously has been convicted of a misdemeanor under the FDCA [21 U.S.C. 333(b)]. Federal misdemeanor charges are typically resolved in proceedings before U.S. magistrate judges, and federal felonies are resolved by U.S. district judges.

***Individual versus Corporate Liability*** Introducing an adulterated product into interstate commerce is a strict liability crime that can be enforced against individuals in positions of sufficient authority and responsibility as well as their company. Persons at risk are those who, at minimum,[20] fail to take adequate measures to *prevent* the cGMP violations. As such, warning letters and other communications are often directed at presidents and CEOs as well as their companies. As stated by the U.S. Supreme Court[21] in 1964, just two years after the FDCA as we know it was passed:

> Food and drug legislation, concerned as it is with protecting the lives and health of human beings, under circumstances in which they might be unable to protect themselves, often "dispenses with the conventional requirement for criminal conduct—awareness of some wrongdoing. In the interest of the larger good it puts the burden of acting at hazard upon a person otherwise innocent but standing in responsible relation to a public danger. . . ."

[18]A government review of recent FDA enforcement has suggested that adequate FDA enforcement activity is lacking. See "Prescription for Harm: The Decline in FDA Enforcement Activity," House Committee on Government Reform, June 2006.

[19]Failure to follow cGMP is the most common form of violating the prohibition against introducing an adulterated drug into interstate commerce.

[20]They may also be directly implicated in fraud and cover-ups.

[21]*United States v. Wiesenfeld Warehouse Co.*, 376 U.S. 86, 91 (1964).

In 1975, the Supreme Court made clear that individual responsibility is very important[22]:

> The [FDCA] imposes not only a positive duty to seek out and remedy violations when they occur but also, and primarily, a duty to implement measures that will insure that violations will not occur. The requirements of foresight and vigilance imposed on responsible corporate agents are beyond question demanding, and perhaps onerous, but they are no more stringent than the public has a right to expect of those who voluntarily assume positions of authority in business enterprises whose services and products affect the health and well-being of the public that supports them.

Manufacturing executives therefore carry a great liability burden. The government only need establish that the individual defendant failed to act on his or her own authority and that such an action could have prevented or corrected the violation. The individual need not have formed any intent to break any laws in order to be found guilty. What is relevant is, did the executive have the power to prevent the acts or omissions complained of? This includes a consideration of whether the executive could have prevented the acts or omission by the systems and processes alone. The job is not made easier to the extent that the cGMP regulations are open to varying interpretations or that the technology is constantly changing.

***Section 305 Proceedings*** Due to the nature of the inspection process, a company that the FDA deems is in violation of the cGMPs should not be surprised when a warning letter or more elevated enforcement techniques are implemented. Even so, the FDA sometimes issues a formal form of notice that criminal charges will be brought by what is called a Section 305 notice.

Section 305 of 21 U.S.C., the statutory basis for a 305 notice, seemingly requires that before any violation of the FDCA is reported to the DOJ for institution of a criminal proceeding, the target defendant must "be given appropriate notice and an opportunity to present his views, either orally or in writing, with regard to such contemplated proceeding." The U.S. Supreme Court has watered down this provision, holding that a notice under Section 305 is not a legal prerequisite to government prosecution.

In practice, then, the FDA only sometimes issues a 305 notice and conducts a 305 hearing when it is considering a misdemeanor prosecution. A very informal process, the manufacturer can approach it with as much or as little of a defense as counsel deems appropriate, as there are pros and cons to providing the government with the company's full defense at that juncture.

A prototype Section 305 notice appears in Figure 9.

***Grand Jury Proceedings*** If the government will be pursuing felony criminal charges against a manufacturing facility or persons associated with such facility, it will proceed by grand jury. The Fifth Amendment to the U.S. constitution requires that charges for all capital or "infamous" crimes be brought by an indictment

[22] *United States v. Park*, 421 U.S. 658 (1975).

In reply refer to:
Sample No.
Product

Firm Name and Individual
Street Address
City, State, Zip

Date

Investigation by this Administration indicates your responsibility for violations of the Federal Food, Drug, and Cosmetic Act, and other Federal Laws, as described in the attached Charge Sheet, with respect to the following:

[describes specifics of cGMP violations]

A meeting has been scheduled for (day, date, time) at (location), to give you an opportunity to present your views on this matter. The enclosed **INFORMATION SHEET** explains the purpose and nature of the meeting, and how you may reply. If no response is received on or before the date set, our decision on whether to refer the matter to the Department of Justice for prosecution will be based on the evidence in hand.
By direction of the Secretary of the Department of Health and Human Services:

Compliance Officer

Enclosures:
Legal Status Sheet (3)
Charge Sheet
Information Sheet
Regulations

**FIGURE 9** Section 305 notice.

returned by a grand jury. This has been interpreted by the U.S. courts to require that an indictment be used to charge federal felonies.

The activities, deliberations, or matters occurring before a grand jury are secret.[23] Strict adherence to grand jury secrecy is important to the integrity of the investigative process and ensures that the grand jury will be able to deliberate without outside pressure, to encourage people with information about a crime to come forward without fear of disclosure, and to protect the rights of the accused, specifically the innocent accused, from disclosure of the fact that he or she or it was investigated. Other than attorneys for the government, only the witness, interpreters when needed, and a court reporter are authorized persons permitted to be present while a grand jury is in session.

A grand jury's function is to determine whether there is probable cause to believe that a certain person(s) or company(ies) have committed a federal offense.[24] Prosecutors are permitted to appear before the grand jury and, in practice, conduct the grand jury proceedings. In general, the prosecutor is the one who makes the decision

[23]See Rule 6(e) Fed. R. Crim. Pro.

[24]The grand jury system is not presently used by countries outside the United States. The United Kingdom, New Zealand, Canada, and Australia, for instance, all have abolished the use of grand juries. See http://enwikipedia.org/wiki/Grand_jury.

as to which witnesses to call and what evidence should come before the grand jury. The prosecutor asks the witness questions and subsequently members of the grand jury may also question witnesses directly or through the prosecutor.

During the course of a grand jury investigation regarding cGMP violations/adulterated product allegations, for instance, the grand jury may hear testimony from not only federal investigators, federal agents, and federal inspectors but also former employees of the company (or current if a custodian of records) and/or experts in the pharmaceutical manufacturing field. These persons are considered witnesses. Witnesses are typically subpoenaed and may not refuse to appear before the grand jury or be subject to contempt charges.

In the federal grand jury system, a witness is not permitted to bring an attorney into the grand jury room. However, a witness is permitted to consult with his attorney outside the grand jury room even interrupting his own testimony.[25] It is typical for corporations such as pharmaceutical manufacturing companies to provide an attorney for any and all employees subpoenaed by a grand jury of which the manufacturer is a target.

A target is the person who is the focus of the grand jury investigation and is likely to be indicted. This company or person may receive a "target letter" from the grand jury which officially advises them of their jeopardy and serves as a formal warning of their status.

In practice, if a manufacturer is the target, the government will likely attempt to develop evidence by subpoena of persons and materials which will help prove culpability. It is likely the subpoenas will ask for correspondence, notes, and memos during a particular time period and involving a particular subject matter.

The grand jury may also issue a subpoena to the manufacturer's designated "custodian of records" for specific document production. The description of subpoenaed documents can include statements or charts of an organization, announcements, statements of policy and procedure, diaries, records of email, manufacturing logs, emails, travel vouchers, financial records and statements, correspondence, notes of conversations, and any other documents that relate to the manufacturing of certain drugs. A document subpoena may also request every writing or record of whatever type and description in the possession, custody, or control of the company that relates to a particular element of the criminal violation the grand jury is investigating. The request typically includes all handwritten, typed, printed, recorded, or transcribed records, including computer records tapes or disks.

The burden is on the government to prove that the crime was committed in the district in which the prosecution is brought. The grand jury should not consider a case unless venue lies in the district where the grand jury is sitting.[26] In the case of adulterated drugs, courts have generally held that it is proper to have venue in a district from which the defendant caused the unlawful introduction of goods into commerce, even though the physical shipment commenced from a different district.

[25]See Rule 6(d) Fed. R. Crim. Pro.; 28 U.S.C. secs. 515, 542, 547.
[26]See Rule 18 Fed. R. Crim. Pro.

***Form of Charges and Penalties*** A grand jury investigation may culminate in the return of an indictment. This means that the grand jury found probable cause to believe that a violation of law occurred. While the focus of the initial inquiry can surround adulterated drugs by virtue of failure to abide by the cGMPs, additional criminal violations may be charged, as, for instance, where there are actions to evade or mislead a grand jury. The end result could, for instance, include accusations of making false statements to the FDA and obstructing the FDA's or DOJ's investigation, in addition to the "adulterated" drug charges.

An indictment consists of a statement describing the time, place, and manner through which the defendant violated the law. Each violation of the law is set out in a separate count. A defendant charged by an indictment is entitled to a trial by jury, although this right can be waived. A defendant has the right to a trial by jury for any criminal offense punishable by imprisonment for more than six months.[27]

If the matter only involves a misdemeanor violation, the prosecutor charges "by information." The information is often referred to as a complaint. An information, like an indictment, is simply a pleading that accuses the defendant of committing crimes. The distinction between an information and an indictment is that a prosecutor can issue and file an information without the grand jury's participation or finding of probable cause but a grand jury must approve and return an indictment.[28]

The penalty for a violation of 21 U.S.C. 331(b) by violating the cGMPs resulting in the adulteration of drug products in interstate commerce is set forth in 21 U.S.C. sec. 333(a). Each separate count for violating the cGMPs (where a misdemeanor is charged) carries with it a possible imprisonment of not more than one year or a fine not to exceed $1000 or both. If the government charges a felony for violation of the cGMPs, then the penalties are imprisonment of not more than three years or a fine not to exceed $10,000 or both. Of course, there may be other charges with greater or lesser penalties which are not related to the adulterated drug charges.

The criminal fine amounts (but not the imprisonment durations), however, are superceded by the criminal fine amounts contained in a different federal statute enacted later. Section 3571 of 18 U.S.C. provides for much greater fines than those provided for within the FDCA itself. For a manufacturer convicted of a felony, the fine can be as much as $500,000; for a misdemeanor (not resulting in death), it could be $200,000. Fines for individuals include a maximum up to $150,000 for a felony and an amount up to $100,000 for a conviction of a misdemeanor not resulting in death. The statute also provides for a multiplier of 2 based upon a finding that the defendant derived a pecuniary gain from the offense. The FDA and DOJ are able to elevate the monetary recoveries against the manufacturers for violations of the cGMP under a civil disgorgement theory explained infra. Recoveries have been in the hundreds of millions in recent years, typically agreed to in a negotiated consent decree.

[27]See Sixth Amendment, U.S. Constitution.

[28]Sometimes, prosecutors are in communication with defendants and their counsel during the investigatory stage. If there are negotiations concerning a plea to a felony and they are successful, a defendant can waive his or her right to be indicted by a grand jury, and the prosecutor can charge them by information.

## 5 CONCLUSION

The diligent enforcement of good manufacturing practices is a cornerstone of the safety net for drugs in the United States. Congress, the courts and, manufacturers, most importantly, expect a degree of consistency and responsibility in enforcement policy over a statute as powerful and central to public health and safety as the FDCA. To the extent there is consistency and effective and evenhanded enforcement, it not only protects the public, but it provides a level playing field for those manufacturers who operate in accordance with the cGMPs.

# SCALE-UP AND POSTAPPROVAL CHANGES (SUPAC) REGULATIONS

Puneet Sharma, Srinivas Ganta, and Sanjay Garg
*University of Auckland, Auckland, New Zealand*

**Contents**

1 Introduction
2 Scientific and Regulatory Rationale for SUPAC
  2.1 Supporting Documents and Extent of Change
  2.2 Supporting Documents for Change in Specifications
  2.3 Comparability Protocols
  2.4 In Vitro–In Vivo Requirements
3 Regulatory Agencies and Guidelines
  3.1 FDA SUPAC Regulations
  3.2 Regulatory Guidance on SUPAC by Pharmaceutical Unit of EU
  3.3 Regulatory Guidance on SUPAC by Agencia Nacional de Vigilancia Sanitaria
4 Harmonization
5 GMP Issues: Change Control and Process Validation
  5.1 Change Control
  5.2 Process Validation
6 Conclusion

## 1 INTRODUCTION

Product development aims at formulating active drug ingredient in a palatable form. Technology transfer of a pharmaceutical product from research to the production floor (referred to as "shop floor") with simultaneous increase in production outputs is commonly known as scale-up. In simple terms, the process of increasing batch size is termed as scale-up. Conversely, scale-down refers to decrease in batch size in response to reduced market requirements.

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

Often, changing of scale from the research lab to the shop floor is fraught with problems. The basic reason for such problems is the usage of different processing equipment in research and on the shop floor. Moreover, insufficient information about the equipment, various requirements of process control, complexity of a particular pharmaceutical process which may have a several different unit operations, limited information about the behavior of ingredients at different scales, and adoption of trial-and-error methodology also add significantly to scale-up issues. Every product coming from research should be manufacturable and the process should be capable to demonstrate its ruggedness at the shop floor level. This statement points toward the criticality and significance of scale-up and technology transfer in a pharmaceutical development process. After successful accomplishment of technology transfer and validation activity, a product usually has a smooth run on large-scale production machines. Changes are being made in the manufacturing process and chemistry of a drug product following approval and continue throughout its life. Depending upon foreseen (or unforeseen) requirements, there can be changes in the raw materials, process, equipment or manufacturing site, and batch size which ultimately affect quality attributes of a drug or finished product. Therefore, there is a need to anticipate and fully evaluate the impact of any kind of change on the quality of a drug or finished product. There can be several reasons for these changes, such as changed market requirement affecting batch size, new source of raw material, change in manufacturing process, upgrades of packaging material, or shifting to a new analytical methodology.

The intensity of the adverse effect produced by a particular change depends on the type of dosage form. For example, a change in the inactive ingredient beyond a certain range will have more effect on a modified-release (MR) dosage form than it would on an immediate-release (IR) dosage form, where bioavailability is not rate limiting. Likewise, a change in the primary packaging of liquid parenteral may have more pronounced effect on its effectiveness than it would have on a solid dosage form. Hence, depending upon the intensity of change or the adverse effect it may have on the critical parameters of a dosage form, reporting requirements to regulatory authorities also vary.

A drug or drug product may experience many changes during its life cycle. These changes may have an adverse effect on the overall safety and effectiveness of the drug or drug product. After a number of changes over a long time period, the product coming to market may be completely different from the one that was approved. Hence, data submitted to regulatory authorities in support of a change must have a comparison record of the drug or drug product to the one that was approved initially. Documentation generated in support of any change to the approved drug or drug product is submitted to regulatory authority for review, and based on the benefit-to-risk ratio, the drug or drug product is approved. Depending upon the intensity of change, supporting documents are provided to the regulatory agency.

Regulatory authorities such as the U.S. Food and Drug Administration (FDA), the European Commission, the Agencia Nacional de Vigilancia Sanitaria (ANVISA) (in english the National Health Surveillance Agency—Brazil, and others require the pharmaceutical industries in respective countries to follow guidelines on scale-up and postapproval changes (SUPAC) to maintain the quality of the pharmaceutical produced. From time to time these guidelines are assessed so as to keep pace with

the technological advances and new guidelines are developed to reduce the burden on the pharmaceutical industry and regulatory authorities. Apart from these guidelines, there are other checkpoints within an industry to assure production of quality products, such as change control and validation exercises, which will be discussed in detail in this chapter. These operations are controlled through the principles of good manufacturing practices issued by regulatory authorities.

This chapter describes the regulations imposed by different regulatory authorities and measures taken by a pharmaceutical industry to assure quality and performance of pharmaceuticals. The FDA guidelines, being most descriptive, have been discussed at length. Other guidelines have been described in general terms and the interested reader is referred to the references or the regulatory websites for more specific details.

## 2 SCIENTIFIC AND REGULATORY RATIONALE FOR SUPAC

Guidelines pertaining to postapproval changes classify these changes in various categories depending upon the effect a particular change may have on the quality and performance of a drug or drug product. Irrespective of the terminologies used by regulatory agencies, in general terms, changes can be described as *mild, moderate*, and *major* and the extent of supporting document varies with the nature of the change. For example, U.S. FDA guideline "Changes to an Approved NDA or ANDA" describe these changes as mild changes that can be implemented immediately and filed in the next periodic report, moderate changes that can be implemented immediately, moderate changes that require 30 days notice before implementation, and major changes that require FDA approval before implementation [1]. Similarly, any changes in an approved drug or drug product under European Union (EU) domain type I (type IA and type IB) and II variation are filed prior to marketing products [2]. The therapeutics Good Administration—Australia (TGA) describes postapproval changes in three categories: nonassessable, self-assessable, and changes requiring prior approval [3].

### 2.1 Supporting Documents and Extent of Change

As per FDA guidelines, changes in excipients (%w/w) of total formulation not greater than 5% are considered minor and all information is provided in the annual report. However, changes likely to have significant effect on the quality and performance of a drug product calls for submission of a prior approval supplement on all information (in vitro dissolution and in vivo dissolution), including accelerated stability and long-term stability testing in the annual report [4]. Similarly, in EU guidelines, a change in the batch size of the finished product up to10-fold compared to the original batch size approved at the grant of the marketing authorization (or downscaling to 10-fold) has been defined as type IA and requires batch analysis data (in a comparative tabulated format) on a minimum of one production batch manufactured to both the currently approved and the proposed sizes. Batch data on the next two full production batches should be made available upon request and reported by the marketing authorization holder if outside specifications (with proposed action). However, for type IB (more than 10-fold), in addition to the above

data, a copy of an approved release and end-of-shelf-life specifications as well as the batch numbers (≥3) used in the validation study should be indicated or a validation protocol (scheme) be submitted and the number of batches used in the stability studies should be indicated.

### 2.2 Supporting Documents for Change in Specifications

Changes in any type of specification also need to be supported by documentation. In all the guidelines, relaxing an acceptance criterion or deleting any part of the specification is classified as a major change and hence extensive documentation is required, for example, submission of a prior approval supplement to the FDA or comparative table of current and proposed specifications and details of any new analytical method and validation data and batch analysis data on two production batches of the finished product for all tests in the new specification to EU. The specifications are benchmarks for comparison of performance of any product. For example, content uniformity specification of 90–110% assay limit of a 20-mg (average weight) tablet of a potent drug signifies the challenge in maintaining the uniformity of such a low-dose drug during the blending operation. Any relaxation in specification of this potent drug should be justified with extensive documentation to assure the performance. However, tightening of an acceptance criterion is considered as a minor level change and to have minimal potential for an adverse effect on the identity, quality, purity, or potency of a product.

### 2.3 Comparability Protocols

The FDA has introduced the concept of comparability protocols to expedite the process of approval after submission of supporting document for a particular change [5]. The protocol covers anticipated changes a product may experience during it shelf life. Its recently published draft guidance "Comparability Protocols—Chemistry, Manufacturing, and Controls (CMC) Information" describes the general principles and procedures to prepare comparability protocols. The FDA suggests a less stringent reporting category for any future change, where appropriate. Additionally, if a detailed comparability protocol is provided, the FDA is less likely to request additional supporting documents while comparing pre- and postapproval change, and this could also help in implementing a particular CMC change, thereby moving the product in the distribution line sooner. According to the FDA:

> A comparability protocol is a well-defined, detailed, written plan for assessing the effect of specific CMC changes in the identity, strength, quality, purity, and potency of a specific drug product as these factors relate to the safety and effectiveness of the product. A comparability protocol describes the changes that are covered under the protocol and specifies the tests and studies that will be performed, including the analytical procedures that will be used, and acceptance criteria that will be achieved to demonstrate that specified CMC changes do not adversely affect the product. The submission of a comparability protocol is optional.

A comparability protocol may be submitted with a new drug application (NDA), abbreviated new drug application (ANDA), or supplements to these applications.

Comparability protocols can have single or multiple changes provided that each change is discrete and specification of the acceptance criteria for a change is well defined.

### 2.4 In Vitro–In Vivo Requirements

Stability of a drug product, in vitro dissolution, and in vivo bioequivalence are prerequisites for performance of a drug product and play a key role in establishing the quality of a drug product after a postapproval change has been implemented. Any type of major change, for example, in the manufacturing process from dry granulation to wet granulation could affect the bioavailability and stability of a drug product. Careful selection of the dissolution condition can obviate the need for a costly bioequivalence study. Guidelines by the FDA [4] and ANVISA [6] take into consideration the solubility and permeability of a drug substance for selection of dissolution criteria for a particular drug product (immediate release or modified release) whereas guidelines by the EU and TGA recommend submitting comparison records between a particular number of manufacturing batches pre- and postapproval.

While categorizing a change or variation for its effect, sufficient consideration should be given to those parameters of a drug product which could affect its bioavailability. Critical parameters like the particle size of active ingredient or excipients, solid-state characteristics, and surface wettability may change during the process variation and could adversely affect product performance resulting in an altered dissolution profile. The effect would be more pronounced in drug products containing poorly soluble potent drugs and could have a deleterious effect on bioavailability. The FDA guideline considers recommendation of the Biopharmaceutic Classification System (BCS) regarding solubility and permeability characteristics to see whether any in vivo bioequivalence study is needed along with an in vitro dissolution study. In the same pattern, ANVISA places drugs in three categories for solid TM dosage form: case A, active substances with high permeability and high solubility; case B, active substances with low permeability and high solubility; and case C, active substances with high permeability and low solubility. As per the guideline, for alteration of registration due to excipient change, for level 2 alterations (that could cause significant impact on quality and performance) the following requirements should be met:

*Case A* "The required documentation must include the undertaking of the technical report and assessment of the results of the dissolution test, carried out as described in the Brazilian Pharmacopoeia and, in its absence, other codes authorized by the legislation in force. There must be dissolution of at least 85% of the active substance in up to 15 minutes, using 900 ml of HCl 0.1 *M*. In case this criterion is not complied with, the tests described for Cases B or C must be carried out."

*Case B* "The required documentation must include the undertaking of the technical report and assessment of the results of the dissolution profile employing Pharmacopeial conditions and removing samples from the medium at appropriate time points until the plateau is reached. The dissolution profile obtained must be similar to the profile of the unaltered formulation."

*Case C* "The required documentation must include the undertaking of the technical report and assessment of the results of the dissolution profile in five different conditions: distilled water, HCl 0.1 *M* and phosphate buffer pH 4.5, 6.5 and 7.5 for the proposed formulation and the previous formulation, without change. Samples of the dissolution medium must be removed at appropriate time points until 90% of the active substance is dissolved or the plateau is reached. A tensoactive may be used only when appropriately justified. The profile obtained must be similar to the profile of the unaltered formulation."

In addition, for level 2 change, no additional bioequivalence study is required if the proposed alteration matches with the situation for cases A, B, and C. However, if there is any deviation, then documentation containing the results and assessment of a new bioequivalence and/or bioavailability study [if proper in vitro/in vivo correlation (*ivivc*) has not been established] should be submitted as per the conditions mentioned in the level 3 alteration.

## 3 REGULATORY AGENCIES AND GUIDELINES

### 3.1 FDA SUPAC Regulations

The Food and Drug Administration Modernization Act (FDAMA) of 1997 (the Modernization Act) was passed on November 21. With FDAMA in effect, another Section 506A was added to the federal Food, Drug, and Cosmetic Act (the act) and Section 314.70 (21 CFR 314.70) and the section included recommendations for reporting categories (in terms of defined words) for any type of manufacturing changes to an approved application (NDA or ANDA). In accordance with the act, the FDA issued "Guidance for Industry: Changes to an Approved NDA or ANDA" (finalized in 2004). This guidance is a current standard for pharmaceutical manufacturers for making and reporting manufacturing changes to an approved application and for distributing a drug product made with such changes.

"SUPAC-IR: Immediate-Release Solid Oral Dosage Forms: Scale-Up and Post-Approval Changes: Chemistry, Manufacturing and Controls, In vitro Dissolution Testing, and In vivo Bioequivalence Documentation" (issued 1995) was the first attempt to provide the pharmaceutical industry with a clear-cut guideline covering the requirements for notification and submission of documentation to regulatory authorities pertaining to postapproval changes. This guideline was an outcome of (a) a workshop on the scale-up of IR products conducted by the American Association of Pharmaceutical Scientists with the U.S. Pharmacopoeia (USP) Convention and the FDA; (b) research conducted by the University of Maryland at Baltimore on the CMC of IR products; (c) drug categorization research on the permeability of drug substances at the University of Michigan and the University of Uppsala; and (d) SUPAC task force set up by the Centre for Drug Evaluation and Research (CDER) CMC coordination committee. Following the issuance, it became a benchmark for the industry. Two more guidances have been published on the same format (level of changes as defined by SUPAC IR) as of SUPAC for MR drug products (issued in 1997) [7] and nonsterile semisolid drug products (issued in 1997) [8]. "Guideline for Changes to Approved NDA or ANDA" supersedes any previous

guidelines which have information on reporting categories that is inconsistent with this guideline.

***Guideline to Industry: Changes to Approved NDA or ANDA*** "Guideline for Industry: Changes to Approved NDA or ANDA" provided reporting categories for various postapproval changes and relaxed certain requirements that were considered to have minimal or no impact on the drug product [1]. Moreover, it lessened the burden on regulatory authorities and companies as well. Four reporting categories provided in this guideline are as follows:

1. *Prior Approval Supplement* For a major change (substantial potential to have effect on quality and performance), a supplement has to be submitted to the FDA for approval before a product made with the change is distributed. There is also a provision for "Prior Approval Supplement: Expedite Review Requested" for public health reasons and if the delay in approval may cause any substantial concerns for the applicant.
2. *Supplement: Changes Being Effected (CBE) in 30 Days* For a moderate change (moderate potential to have effect on quality and performance), a supplement has to be submitted to the FDA for approval 30 days before a product made with the change is distributed.
3. *Supplement: Changes Being Effected in 0 Days* For some changes a supplement has to be submitted to the FDA and simultaneously the product made with the change can be distributed.
4. *Annual Report* For a minor change (minimal potential to have effect on quality and performance), all information has to be submitted to the FDA in the next annual review and the product made with the change can be distributed.

All three types of changes under this guidance have been categorized as follows:

(a) *Changes in Components and Composition* Any qualitative or quantitative changes in the components and composition of a drug product is considered as major changes. The current *Guideline to Industry: Changes to Approved NDA or ANDA* does not mention these in detail because of the complexity involved in the recommendations and therefore the SUPAC guideline has to be followed for any such type of changes and regarding documentation requirements for regulatory submission.

(b) *Changes in Manufacturing Sites* A change in a manufacturing site (for manufacturing, packaging, labeling of drug products, testing components, drug product containers, closures, packaging materials), either owned or contract site, of drug products from the one that is approved requires prior approval from the CDER. A prior approval supplement has to be submitted for a change to a site that does not have *a satisfactory CGMP inspection* for the *type of operation* to be performed. Further, changes in sites related to operations like labeling, secondary packaging, and testing are considered to have effect independent of drug product dosage form and therefore the

reporting categories for any of type of manufacturing site changes will be the same. However, changes in sites related to operations like manufacturing and primary packaging are considered to have effect that is dependent on dosage form and hence reporting categories may be different.

(c) *Changes in Manufacturing Process* Changes in the manufacturing process can have substantial effect on the identity, strength, quality, purity, or potency of a drug product and there may be a change in the efficacy of the drug product regardless of the testing of drug product for conformance for the approved specification.

(d) *Changes in Specifications* Specification, acceptance criteria, and regulatory analytical procedure are a part of every dossier submitted to regulatory agencies. Specifications are the standards, acceptance criteria are the limits for specifications, and the regulatory analytical procedure is used for testing a specification's acceptance criteria for the test substance that is approved by the regulatory authority. Alternative analytical procedure may be included in the application simultaneously with the main analytical procedure.

(e) *Changes in the Container Closure System* Effects related to changes in the container closure system are largely dependent on route on administration, the operation in which the container closure system is involved, and contact with the drug product. In some cases, there may be an effect in spite of the conformance of drug product with the approved specification.

(f) *Changes in Labeling* Changes in the package insert and package container label are included in the labeling changes and applicant must immediately revise all promotional labeling and drug advertisement in accordance with the change in the approved labeling.

(g) *Miscellaneous Changes* Apart from categories mentioned above, changes like stability protocol, expiration period, and addition of stability protocol or comparability protocol have been included in the miscellaneous category.

(h) *Multiple Related Changes* One change may lead to advertent or inadvertent incorporation of another change, for example, a change in the manufacturing site may lead to a change in the manufacturing equipment and manufacturing process or changes in packaging material may cause changes in stability protocol. For such combination changes, the CDER recommends submitting documents in accordance with the most stringent reporting category for the individual change.

***Scale-up and Postapproval Changes: Immediate-Release and Modified-Release Dosage Forms*** SUPAC guidelines categorized postapproval changes in terms of "levels" [4]. Three levels were defined depending upon the intensity of the adverse effect on the formulation. Level 1 signifies that the resulting effect on the quality would be minimal and less extensive documentation should be presented to the FDA in an annual review. Changes in accordance with level 2 could have significant effect on the quality and performance of the dosage form. Level 3 changes are most likely to affect the quality and performance of the dosage form and hence extensive

**TABLE 1 Different Cases and Respective Dissolution Conditions for Immediate-Release Solid Dosage Form**

| Case A[a] | Case B[b] | Case B[c] |
|---|---|---|
| Dissolution of 85% in 15 min in 900 mL of 0.1 *N* HCl. If a drug product fails to meet this criterion, the applicant should perform the tests described for case B or C. | Multipoint dissolution profile should be performed in the application/compendial medium at 15, 30, 45, 60, and 120 min or until an asymptote is reached. | Multipoint dissolution profiles should be performed in water, 0.1 *N* HCl, and USP buffer media at pH 4.5, 6.5, and 7.5 (five separate profiles) for the proposed and currently accepted formulations. Adequate sampling should be performed at 15, 30, 45, 60, and 120 min until either 90% of drug from the drug product is dissolved or an asymptote is reached. A surfactant may be used, but only with appropriate justification. |

[a]High-permeability, high-solubility drugs.
[b]Low-permeability, high-solubility drugs.
[c]High-permeability, low-solubility drugs.

documentation justifying those changes should be submitted to the FDA prior to distribution of the products made with these changes. Apart from describing these levels, recommendations were also made on the extent of CMC documentation, in vitro dissolution, and in vivo bioequivalence tests that need to be submitted. Each section in the guideline [(a) components and compositions, (b) site change, (c) scale-up/scale-down; and (d) manufacturing equipment and process] was categorized in terms of these three levels. Further, SUPAC IR also takes into consideration the therapeutic range, solubility, and permeability of the drug for defining any particular change. As per the guideline, three cases have been defined for the dissolution testing (as mentioned in Table 1). Moreover, changes in excipient limits for a narrow therapeutic range drug beyond that mentioned in level 1 have been recommended as level 3 changes and extensive documentation is required for justification. In the SUPAC guideline for MR dosage forms, changes have been described at the same three levels [7].

However, dissolution conditions have been distinguished quite reasonably between extended- and delayed-release dosage form (Table 2). For reporting any level 3 change, three-month accelerated stability data of three batches (significant body of information not available) or three-month accelerated stability data for one batch (significant body of information available) have to be submitted in a supplement along with long-term stability data for one batch in an annual review. A significant body of information has been defined in the guideline as availability of sufficient stability information of the product (stability data of five commercial batches). To provide a comparative outline, the guidelines for MR and IR dosage forms are described in Tables 3–8:

**TABLE 2 Dissolution Conditions for Modified-Release Dosage Form**

| Extended Release | Delayed Release |
|---|---|
| In addition to application/compendial release requirements, multipoint dissolution profiles should be obtained in three other media, for example, in water, 0.1 *N* HCl, and USP buffer media at pH 4.5 and 6.8 for the changed drug product and the biobatch or marketed batch (unchanged drug product). Adequate sampling should be performed, for example, at 1, 2, and 4 h and every 2 hours thereafter until either 80% of the drug from the drug product is released or an asymptote is reached. A surfactant may be used with appropriate justification. | In addition to application/compendial release requirements, dissolution tests should be performed in 0.1 *N* HCl for 2 h (acid stage) followed by testing in USP buffer media, in the range of pH 4.5–7.5 (buffer stage) under standard (application/compendial) test conditions and two additional agitation speeds using the application/compendial test apparatus (three additional test conditions). Multipoint dissolution profiles should be obtained during the buffer stage of testing. Adequate sampling should be performed, for example, at 15, 30, 45, 60, and 120 min (following the time from which the dosage form is placed in the buffer) until either 80% of the drug from the drug product is released or an asymptote is reached. The above dissolution testing should be performed using the changed drug product and the biobatch or marketed batch (unchanged drug product). |

(a) *Changes in Components and Compositions (Table 3)* The guideline for changes to approved NDA or ANDA does not define these changes in detail, and thus the SUPAC guideline has to be followed for reference and reporting. Changes in excipient levels are submitted as a *prior approval supplement* (with accelerated stability data) whereas any changes in the levels of colors or flavors are submitted in an annual review (long-term stability data). In MR dosage forms these changes have been logically categorized as (a) changes in excipient levels not affecting the release profile and (b) changes in excipient levels affecting the release profile. In level 2 changes for IR product and MR dosage forms for a non-narrow therapeutic drugs, three-month accelerated stability data of one batch (in MR dosage form for narrow therapeutic drugs three-month accelerated stability data of three batches) in a supplement and long-term stability data of one batch in an annual review should be submitted. Additionally, for delayed-release MR dosage forms of a narrow therapeutic range drug, the multipoint dissolution profile in the buffer stage of testing should be generated for changed and commercial product using the medium that is approved or in pharmacopeia. For extended-release MR dosage forms of a narrow therapeutic range drug, the multipoint dissolution profile should be generated for changed and commercial product using the medium that is approved or in pharmacopeia.

(b) *Changes in Manufacturing Site (Table 5)* A change in the manufacturing or packaging site (or a contract manufacturing location) that has been approved by the FDA in the original application has to be evaluated for its effect on the product quality and performance. These changes have been described in detail in current guideline changes to approved NDA or ANDA.

**TABLE 3 Changes in Nonrelease Controlling Components and Composition**

| Level | Classification | Therapeutic Range/Type of Drug | Test Documnetation | | | Filing Documentation |
|---|---|---|---|---|---|---|
| **I** | Complete or partial deletion of color/flavor<br>Change in inks, imprints<br>SUPAC-IR level 1 excipient ranges<br>No other changes | All drugs | Stability<br>Application/compendial requirements<br>No biostudy | | | Annual report |
| **II** | Change in technical grade and/or specifications<br>Higher than SUPAC-IR level 1 but less than level 2 excipient ranges<br>No other changes<br>Change in technical grade and/or specifications<br>Higher than SUPAC-IR level 1<br>No other changes | All drugs for MR<br>Depending upon therapeutic range solubility and permeability (as per BCS) for IR | **MR (ER):**<br>Notification and updated batch record<br>Stability<br>Application/ compendial requirements plus multipoint dissolution profiles in three other media (e.g., water, 0.1 *N* HCl, and USP buffer media at pH 4.5 and 6.8) until ≥80% of drug released or an asymptote is reached<br>Apply some statistical test (f2 test) for comparing dissolution profiles<br>No biostudy | **MR (DR):**<br>Notification and updated batch record<br>Stability<br>Application/compendial requirements plus multipoint dissolution profiles in additional buffer stage testing (e.g., USP buffer media at pH 4.5–7.5) under standard and increased agitation conditions until ≥80% of drug released or an asymptote is reached<br>Apply some statistical test (f2 test) for comparing dissolution profiles<br>No biostudy | **IR:**<br>Notification and updated batch record<br>Stability<br>Dissolution requirements: case A, case B, or case C<br>Apply some statistical test (f2 test) for comparing dissolution profiles<br>No biostudy | Prior approval supplement |

**TABLE 3** ***Continued***

| Level | Classification | Therapeutic Range/Type of Drug | Test Documnetation | | Filing Documentation |
|---|---|---|---|---|---|
| **III** | Higher than SUPAC-IR level 2 excipient ranges for MR and IR, change in excipient range for low solubility and low permeability drugs beyond level 1 | All drugs for MR and all drugs failing dissolution criteria for level 2 for IR | Updated batch record<br>Application/compendial (profile) requirements and as mentioned for level II<br>Stability<br>Biostudy or ivivc | Updated batch record<br>Dissolution profile as for level II<br>Stability<br>Biostudy or ivivc | Prior approval supplement |

*Note*: MR, Modified-release dosage form; ER, extended-release dosage form; DR, delayed-release dosage form; IR, immediate-release dosage form.

**TABLE 4 Changes in Release Controlling Components and Composition**

| Level | Classification | Therapeutic Range | Test Documentation | | Filing Documentation |
|---|---|---|---|---|---|
| **I** | ≤5% w/w change based on total release controlling excipient (e.g., controlled-release polymer, plasticizer) content<br>No other changes | All drugs | Stability<br>Application/compendial requirements<br>No biostudy | | Annual report |
| **II** | Change in technical grade and/or specifications<br>≤10% w/w change based on total release controlling excipient (e.g., controlled-release polymer, plasticizer) content<br>No other changes | Nonnarrow | **MR (ER):**<br>Notification and updated batch record<br>Stability<br>Application/compendial requirements plus multipoint dissolution profiles in three other media (e.g., water, 0.1 *N* HCl, and USP buffer media at pH 4.5 and 6.8) until ≥80% of drug released or an asymptote is reached<br>Apply some statistical test (f2 test) for comparing dissolution profiles<br>No biostudy | **MR (DR):**<br>Notification and updated batch record<br>Stability<br>Application/compendial requirements plus multipoint dissolution profiles in additional buffer stage testing (e.g., USP buffer media at pH 4.5–7.5) under standard and increased agitation conditions until ≥80% of drug released or an asymptote is reached<br>Apply some statistical test (f2 test) for comparing dissolution profiles<br>No biostudy | Prior approval supplement |
| | | Narrow | Updated batch record<br>Stability<br>Application/compendial (profile) requirements<br>Biostudy or ivivc | | Prior approval supplement |
| **III** | >10% w/w change based on total release controlling excipient (e.g., controlled-release polymer, plasticizer) content | All drugs | Updated batch record and stability<br>Application/compendial (profile) requirements<br>Biostudy or ivivc | | Prior approval supplement |

**TABLE 5 Site Changes**

| Level | Classification | Therapeutic Range | Test Documentation | | | Filing Documentation |
|---|---|---|---|---|---|---|
| **I** | Single facility<br>Common Personnel<br>No other changes | All drugs | Application/compendial requirements<br>No biostudy | | | Annual report |
| **II** | Same contiguous campus<br>Common personnel<br>No other changes | All drugs | **MR (ER):**<br>Identification and description of site change and updated batch record<br>Notification of site change<br>Stability<br>Application/ compendial requirements plus multipoint dissolution profiles in three other media (e.g., water, 0.1 *N* HCl, and USP buffer media at pH 4.5 and 6.8) until ≥80% of drug released or an asymptote is reached<br>Apply some statistical test (f2 test) for comparing dissolution profiles<br>No biostudy | **MR (DR):**<br>Identification and description of site change, and updated batch record<br>Notification of site change<br>Stability<br>Application/ compendial requirements plus multipoint dissolution profiles in additional buffer stage testing (e.g., USP buffer media at pH 4.5–7.5) under standard and increased agitation conditions until ≥80% of drug released or an asymptote is reached<br>Apply some statistical test (f2 test) for comparing dissolution profiles<br>No biostudy | **IR:**<br>Identification and description of site change and updated batch record<br>Notification of site change<br>Stability<br>Application/ compendial requirements<br>No biostudy | Changes being effected supplement (accelerated stability data for MR and no stability data for IR)<br>Annual report (long-term stability data for MR and IR) |
| **III** | Different campus<br>Different personnel | All drugs | Notification of site change<br>Updated batch record<br>Application/compendial (profile) requirements (as for level II)<br>Stability<br>Biostudy or ivivc | | Notification of site change<br>Updated batch record<br>Case B dissolution as for excipient change (level II)<br>Stability<br>No biostudy | Prior approval supplement (accelerated stability data) for MR and changes being effected supplement for IR<br>Annual report |

**TABLE 6 Changes in Batch Size: Scale-Up/Scale-Down**

| Level | Classification | Change | Test Documentation | | | Filing Documentation |
|---|---|---|---|---|---|---|
| **I** | Scale-up of biobatch(s) or pivotal clinical batch(s)<br>No other changes | ≤10× (all drugs) | Updated batch record<br>Stability<br>Application/compendial requirements<br>No biostudy | | | Annual report |
| **II** | Scale-up of biobatch(s) or pivotal clinical batch(s)<br>No other changes | >10× (all drugs) | **MR (ER):**<br>Updated batch record<br>Stability<br>Application/compendial requirements plus multipoint dissolution profiles in three other media (e.g., water, 0.1 *N* HCl, and USP buffer media at pH 4.5 and 6.8) until ≥80% of drug released or an asymptote is reached<br>Apply some statistical test (f2 test) for comparing dissolution profiles<br>No biostudy | **MR (DR):**<br>Updated batch record<br>Stability<br>Application/compendial multipoint dissolution profiles in additional buffer stage testing (e.g., USP buffer media at p*H* 4.5–7.5) under standard and increased agitation conditions until ≥80% of drug released or an asymptote is reached<br>Apply some statistical test (f2 test) for comparing dissolution profiles<br>No biostudy | **IR:**<br>Updated batch record<br>Stability<br>Case B dissolution as for excipient change (level II)<br>No biostudy | Changes being effected supplement (accelerated stability data)<br>Annual report (long-term stability data) |

**TABLE 7 Changes in Manufacturing: equipment**

| Level | Classification | Change | Test Documentation | | | Filing Documentation |
|---|---|---|---|---|---|---|
| **I** | Equipment changes<br>No other changes (all drugs) | Alternate equipment of same design and principle<br>Automated equipment | Updated batch record<br>Stability<br>Application/compendial requirements<br>No biostudy | | | Annual report |
| **II** | Equipment changes<br>No other changes (all drugs) | Change to equipment of a different design and operating principle | **MR (ER):**<br>Updated batch record<br>Stability<br>Application/ compendial requirements plus multipoint dissolution profiles in three other media (e.g., water, 0.1 *N* HCl, and USP buffer media at pH 4.5 and 6.8) until ≥80% of drug released or an asymptote is reached<br>Apply some statistical test (f2 test) for comparing dissolution profiles<br>No biostudy | **MR (DR):**<br>Updated batch record<br>Stability<br>Application/ compendial requirements plus multipoint dissolution profiles in additional buffer stage testing (e.g., USP buffer media at ph 4.5–7.5) under standard and increased agitation conditions until ≥80% of drug released or an asymptote is reached<br>Apply some statistical test (f2 test) for comparing dissolution profiles<br>No biostudy | **IR:**<br>Updated batch record<br>Stability<br>Case C dissolution as for excipient change (level II)<br>No biostudy | Prior approval supplement (accelerated stability data)<br>Annual report (long-term stability data) |

**TABLE 8 Changes in Manufacturing: Processes**

| Level | Classification | Change | Test Documentation | | | Filing Documentation |
|---|---|---|---|---|---|---|
| **I** | Processing changes affecting the nonrelease/release controlling excipients for MR<br>Changes within validation ranges (IR)<br>No other changes | Adjustment of equipment operating conditions (mixing times, operating speeds)—within approved application ranges | Updated batch record<br>Application/compendial requirements<br>No biostudy | | | Annual report |
| **II** | Processing changes affecting the nonrelease controlling excipients and/or the release controlling excipients<br>Processing changes outside validation ranges for IR<br>No other changes | Adjustment of equipment operating conditions (e.g. mixing times, operating speeds, etc.)<br>Beyond approved application ranges | **MR (ER):**<br>Updated batch record<br>Stability<br>Application/compendial requirements plus multipoint dissolution profiles in three other media (e.g. water, 0.1 *N* HCl, and USP buffer media at pH 4.5 and 6.8) until ≥80% of drug released or an asymptote is reached<br>Apply some statistical test (f2 test) for comparing dissolution profiles<br>No biostudy | **MR (DR):**<br>Updated batch record<br>Stability<br>Application/compendial requirements plus multipoint dissolution profiles in additional buffer stage testing (e. g., USP buffer media at pH 4.5–7.5) under standard and increased agitation conditions until ≥80% of drug released or an asymptote is reached<br>Apply some statistical test (f2 test) for comparing dissolution profiles<br>No biostudy | **IR:**<br>Notification of change<br>Updated batch record<br>Stability<br>Case B dissolution as for excipient change (level II)<br>No biostudy | Changes being effected supplement (accelerated stability data for MR)<br>Annual report (long-term stability data for MR and IR) |
| **III** | Processing changes affecting the nonrelease controlling excipients and/or the release controlling excipients | Change in the type of process used (e.g. from wet granulation to dry) | Updated batch record<br>Stability<br>Application/compendial (profile) requirements<br>Biostudy or ivivc | | Updated batch record—stability<br>Case B dissolution as for excipient change (level II)<br>No biostudy | Prior approval supplement (accelerated stability data for MR)<br>Annual report (long-term stability data for MR and IR) |

The FDA should be notified of the new location. For any type of moderate changes, accelerated stability data of one batch should be submitted with the CBE supplement for MR dosage forms and long-term stability data of one batch should be submitted in the annual review for IR as well as MR dosage forms. Additionally, the CBE should be submitted to the FDA in case of any moderate change. Stability requirements for any major change are the same as those mentioned in the previous section, that is, three-month accelerated stability data of three batches (significant body of information not available) or three-month accelerated stability data for one batch (significant body of information available) have to be submitted in a prior approval supplement along with long-term stability data for one batch in the annual review.

(c) *Scale-Up/Scale-Down (Table 6)* A change in the batch size of a drug product, either scale-up or scale-down, is likely to induce some changes in the operation parameters. This in turn can adversely affect product quality.

(d) *Changes in Manufacturing Equipment (Table 7) and Process (Table 8)* Any manufacturing changes in equipment and process are included in this section. For example, a change in the blending equipment from octagonal blender to double cone blender or a change in the granulation process from wet to dry granulation calls for submission of proper validation documentation for FDA approval. All these changes along with reporting categories have been described in current guidelines for changes to approved NDA or ANDA.

***Biowaivers*** In vitro and in vivo approaches are commonly used for establishment of bioavailability and bioequivalence. Dissolution studies are used as in vitro approaches and also serve as quality control tools for pharmaceuticals. Under certain circumstances, in vitro dissolution may also act as a surrogate marker for in vivo biostudy and enable the establishment of in vitro and in vivo bioequivalence.

"CDER Guidance for Industry: Waiver of In Vivo Bioavailability and Bioequivalence Studies for IR Solid Dosage Form Based on Biopharmaceutic Classification System (BCS)" recommends waiving an in vivo biostudy under specific circumstances. For example, a waiver of the in vivo biostudy of one or more lower strengths is acceptable based on the correlation data and in vivo bioequivalence of the higher strength, provided all strengths are proportionally equivalent in terms of active and inactive ingredients. A biostudy on a lower strength may also be requested based on safety reasons (as for mitrazapine tablets) and a biowaiver for highest strength is acceptable provided elimination kinetics is linear over a dose range, strengths are proportional, and comparative dissolution data of all strengths are acceptable.

The BCS classifies drugs in four classes:

Class I: high solubility, high permeability
Class II: low solubility, high permeability
Class III: high solubility, low permeability
Class IV: low solubility, low permeability

Dissolution, solubility, and permeability are the three fractors that control the bioavailability of a drug for an IR drug product. Provided the inactive excipient

does not control or modify the release and absorption of the active ingredient, the biostudy may be waived. According to the guideline, the solubility class is determined for the highest dose strength of a drug product for which a biowaiver has been requested. When the highest dose strength of a solid dosage form is soluble in 250 mL of water or less across a pH range of 1–7.5, it is considered as highly soluble. For determination of permeability class various in vivo methods like mass balance, absolute bioavailability, and intestinal perfusion approaches and in vitro methods like permeation studies using excised tissue or monolayer of cultured epithelial cells are used. When extent of absorption is greater than 90% of the administered dose in humans, it is considered as highly permeable. For a dissolution study, drug release should be evaluated in three media that are 0.1 N HCl or USP-simulated gastric fluid without enzymes, pH 4.5 buffer, and pH 6.8 buffer or USP-simulated intestinal fluid without enzymes. Rapidly dissolving drug products are those that dissolve more than 80% in 900 mL of the above-mentioned media in less than 30 min using USP apparatus at 100 rpm (or USP II apparatus of 50 rpm).

A biowaiver can be requested for the postchange products if it falls under class I of the BCS and displays a rapidly dissolving profile and there is a similarity (as determined by f2 test) between the pre- and postchanged drug product in all three media. For BCS class II drugs, a meaningful correlation (level A, B, or C correlation) between in vitro drug release and in vivo absorption also may be used for requesting the biowaiver. Deconvolution techniques are used for prediction of in vivo dissolution and absorption.

### 3.2 Regulations Guidance on SUPAC by Pharmaceutical Unit of EU

The pharmaceutical market in European countries is one of the largest in the world. To ensure that the EU promotes pharmaceutical trade and ensures safety, efficacy, and quality of medicinal products within the European member states, the pharmaceutical unit of the EU runs a series of information and communication projects, collectively called EUDRA projects. Out of these projects, the EUDRALEX pharmaceutical unit is responsible for making community pharmaceutical legislation, guidelines, and notices for applicants [9]. Under Volume 2, Section C, of Regulatory Guidelines (Pharmaceutical Legislation: Notice to Applicants) of Eudralex, "Guideline on Dossier Requirements for Type IA and Type IB Notifications" has been provided [2].

Regulations were introduced to lessen the administrative load on the authority and to simplify the procedure for granting a postapproval variation without negotiating any quality attribute of drug product [10]. Under these regulations, type IA and type IB were defined; also, clearcut terms were introduced for extension application, parallel/consequential notification/variation, and urgent safety restriction. For streamline operation of these regulations, four documents have been prepared:

(a) A procedural guidance for the member states (reference or concerned) and the applicant for notifications/variations in the mutual recognition procedure
(b) A procedural guidance for the applicant for notifications/variations in the centralized procedure

(c) A common application form which may be used for type IA and type IB notifications or type II variations in both the centralized and mutual recognition procedures
(d) A guideline on the documentation to be submitted for type IA and type IB notifications

All member states of the EU follow the same regulations for a change or "variation" in an already approved medical product. As per the guidance, three types of variations or changes have been identified—type I variation, which is further classified into types IA and IB and type II variations. The guidelines classify some specific changes in type IA or IB. It also provides specific data analysis required for variation and the types of document that need to be submitted to the regulatory authority. Any change that is not listed in this section is classified as type II variation.

According to Commission Regulation (EC, No. 1084/2003), type I variation has been defined as "A '*minor* variation' of type IA or type IB means a variation listed in Annex I, which fulfils the conditions, set out therein." Annex I of the regulation provides a list of changes and conditions (to be satisfied) to be classified as type IA or type IB variation and Annex II provides changes falling under the extension application category. Type II variations in proposed documentation are not type I or extension application. There is also a provision for "urgent safety restrictions." These are any temporary or provisional changes in the product summary characteristics, such as indications, posology, contraindications, warnings, target species, and withdrawal periods, as result of a new information that may cause significant safety concerns about the medicinal product [11].

Any change arising from the primary change has to be notified separately. Consequential changes form part of the same notification whereas parallel changes do not. A consequential change to type IA can only be another type IA whereas a consequential change to type IB can be type IA or type IB. All other variations should be submitted as Type II variations. "Guideline on Dossier Requirements for Type IA and Type IB Notifications" provides a complete list of all changes, conditions required to be met for the particular change, and documentation required by the regulatory authority [10].

### 3.3 Regulatory Guidance on SUPAC by Agencia Nacional de Vigilancia Sanitaria

Agencia Nacional de Vigilancia Sanitara (ANVISA) issues Brazil's generic drug policy. Under legislation for industry, Resolution RE N° 893, of May 29, 2003, is described in "Guide for Making Post-Registration Alterations, Inclusions and Notifications of Drug Products" [6]. This guideline describes postregistration changes as "alterations" and "inclusions" and also tells about the documentation and assays that need to be submitted in support of any type of change. As per the guideline, each type of alteration or inclusion has to be submitted separately and approved by the ANVISA before it can be implemented. Table 9 presents some examples for each category.

Under each category, certain requirements have to be met before its implementation. For example, for inclusion in the batch size, the company should notify, in alteration, if the included batch size is more than 10 times. The documentation that needs to be submitted includes the original proof of payment of fee or of exemption;

**TABLE 9 Examples of Different Categories**

| Postregistration Alterations | Postregistration Inclusions | Postregistration Notifications | Postregistration Cancellation |
|---|---|---|---|
| Labeling alteration | Inclusion of new commercial presentation | Temporary suspension of manufacture | Cancellation upon request of registration of drug presentation |
| Alteration of corporate name | Inclusion of new packing | Resumption of drug manufacture | Cancellation of drug registration |
| Alteration of date of expiry | Inclusion of new concentration already approved in country | | |
| Alteration of preservation conditions | Inclusion of new dosage form already approved in country | | |
| Alteration of synthesis path of drug | Inclusion of new therapeutic indication in country | | |
| Alteration of manufacturer of drug | Inclusion of manufacture site | | |
| Alteration of manufacturing site | Inclusion of manufacturer of drug | | |
| Alteration of excipient | Inclusion in the batch size | | |

a copy of the certificate of good manufacturing and control practices (CBPFC) issued by ANVISA; technical justification; production and quality control records of one batch of each strength of the product; a technical report; and a technical report and assessment of the dissolution profile.

## 4 HARMONIZATION

It is essential to evaluate the safety and quality of new or changed medical products before they reach the market. However, the need to set specific guidelines has been recognized at different times in different countries. For example, in the United States a tragic incident with a junior paracetamol formulation was the alarm to initiate guidelines for authorization of medical products. European countries followed this trend in the 1960s after the thalidomide incident. Since then there have been a large number of guidelines that have been put into place to evaluate medical products in terms of their quality, safety, and efficacy [12]. However, with the pharmaceutical industries becoming international and aiming for a worldwide market, there is a move toward internationally accepted guidelines and approval systems. In order for medical products to be marketed internationally, companies have found it necessary to duplicate many tests and studies that are time consuming and broad.

**TABLE 10 Changes in Specification of Excipients (Addition of New Test Limit): Comparison between Guidelines**

| Guidelines | Documentation | Type of Change |
|---|---|---|
| TGA | Details of the test method must be provided.<br>Appropriate validation data have been generated for the test method.<br>The limits proposed are based on batch analytical data and are in compliance with official standard and/or relevant accepted guidelines if applicable. | Self-assessable changes |
| EMEA | Comparative table of current and proposed specifications.<br>Batch analysis data on two production batches for all tests in the new specification.<br>Where appropriate, comparative dissolution profile data for the finished product on at least one pilot batch containing the excipient complying with the current and proposed specification. For herbal medicinal products, comparative disintegration data may be acceptable. | Minor change type IB requires approval |

Table 10 shows examples of documentation required by different countries when a postapproval change is made during the manufacturing process of medical products. When there are changes in the specification of an excipient, the documents required by the TGA and European Agency for Evaluation of Medicinal Products (EMEA) are variable. Furthermore this would indicate that the benefit of a patent/medical product might not reach globally. There are also many chances of making an error. For example, the 2003 recall of clotrihexal 100-mg vaginal tablets in New Zealand pharmacies was due to the fact that clotrihexal was packed according to TGA guidelines and thus its sale was prohibited in New Zealand. This resulted in much confusion and problems among patients and medical professionals. Harmonization is the process by which the pharmaceutical industries worldwide adopt the same laws and regulations. Harmonization is intended to assure the safety, quality, and efficacy of a medical product globally. The main goal of harmonization is to recognize and minimize the differences in the scientific requirements for medical product development within different regulatory agencies in different countries.

Harmonization activities focus on reducing and simplifying the types of studies that the pharmaceutical industries need to carry out in order to register a medical product in another country, protocols to be followed when performing these studies, techniques used to validate supporting data, and techniques used to perform risk assessment.

Harmonization reduces replications and unnecessary production and registration of new and changed products. The concept of harmonization was explored by European countries in the 1980s. The success of harmonization in these countries has demonstrated that it is practical and possible. Following harmonization in Europe the International Conference on Harmonization (ICH) was set up in 1990 [13]. Table 11 shows some of the harmonized rules that have been successfully developed by ICH.

**TABLE 11 Example of Quality Guidelines Harmonized by ICH**

| Quality Topic | Example of Guideline |
|---|---|
| Q1: Stability | Q1B: Photostability testing |
| Q2: Validation of analytical procedure | Q2A: Methodology |
| Q3: Impurity testing | Q3A: Impurities in new drug substances |
| Q4: Pharmacopoeias | Q4: Pharmacopoeial harmonization |
| Q5: Quality of biotechnological products | Q5A: Viral safety evaluation of biotechnological products |
| Q6: Specifications for new drug substance and products | Q6A: Acceptance criteria for new drug substances |
| Q7: GMP for pharmaceutical ingredients | Q7A: GMP for active pharmaceutical ingredients |

## 5 GMP ISSUES: CHANGE CONTROL AND PROCESS VALIDATION

Changes are unavoidable in a manufacturing setup. Manufacturers make changes at some stage of manufacturing during and after approval of a product. However, consistent quality of a drug product can only be assured through well-defined validation procedures. When a change is made in the manufacturing process of a drug product, sponsors are responsible for evaluating the effect of any change on the safety, efficacy, quality, stability, and potency of a drug product and ensuring that these properties are not influenced by the change. In a manufacturing setup, various disciplines like sales, marketing, medical, regulatory affairs, manufacturing, electrical, and technical services work together. Hence, any kind of change in one discipline will have direct consequences on other disciplines. Each company should have a procedure with regard to handling a change. Quality control and quality assurance departments usually keep track of various changes occurring in a GMP environment. Therefore, it is required that personnel performing the job are trained enough to assess the effect of any kind of change or variation and take appropriate action for its evaluation or control. Supporting data should be generated and once evaluated can confirm whether further clinical or nonclinical studies are required.

### 5.1 Change Control

When a change is made in a manufacturing setup, it is important to assess its impact. As a change can have impact on regulatory filing, manufacturing parameters, specifications, and technical services, it is important to consider the concerns and objections of various disciplines involved and only through well-defined standard operating procedures should it be properly validated, evaluated, and finally implemented. A properly defined order of evaluation of a change with strategic input of trained personnel is key to delivering a consistent quality product (Figure 1).

When a change is the processed, the manufacturer should have protocols in place with regard to assessing the change. Therefore, "control of change" is important. Control can be implemented effectively only through well-defined standard operating procedures. The main purpose of "change control" exercise is to have a

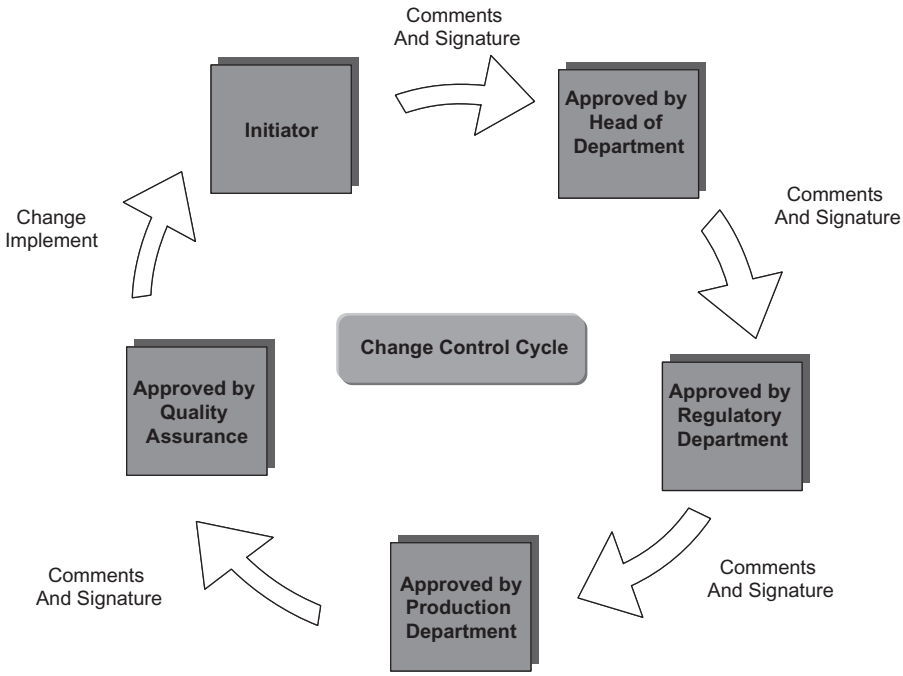


**FIGURE 1** Change control cycle for change in manufacturing process.

systematic process in place to accurately evaluate a change using specific tests. Moreover, it aims to measure the effects on quality safety and efficacy before a change is implanted. Change control and its evaluation through proper documentation should include [14]:

(a) Description and purpose of change
(b) Inputs from research and development (R&D) department
(c) Evaluation steps for impact assessment, such as evaluation of stability, validation requirements, and in vivo bioequivalence requirement
(d) Need and extent of regulatory documentation and approval
(e) Implementation schedule
(f) Clear definition of personnel authorized for change approval
(g) Monitoring protocol for change implementation and periodic review of impact

Following the informal proposal of a change, it should be reviewed by the responsible initiator, who will then generate a formal proposal [15]. The proposal should describe accurately what the change is concerned with, how to validate the change, and the time frame within which the change should be implemented. The final proposal should be reviewed and assessed by all functional groups involved. Once the change is approved, it can be implemented and the change cycle is completed. Figure 2 describes responsibilities of different departments of a pharmaceutical company,

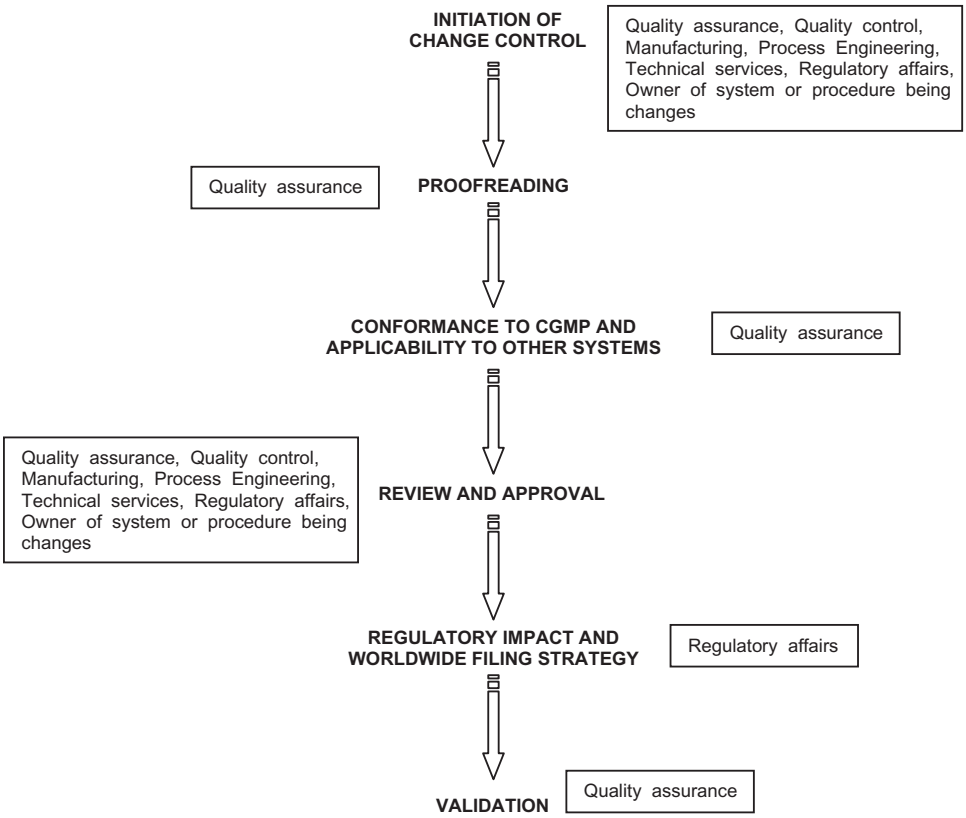


**FIGURE 2** Responsibilities of different disciplines of a pharmaceutical company in a change control procedure (modified from ref. 15).

in the change control procedure. Standard operating procedures (SOPs) for change control are an important part of any GMP audit. Hence it is important that it is implemented by trained and qualified personnel from appropriate disciplines.

After a change has been approved by all functional groups within the manufacturing setup and if it has no regulatory concerns, it can be implemented immediately. However, if the impact comes under any regulatory domain, the company may have to wait for regulatory approval.

## 5.2 Process Validation

Process validation is an important part in the implementation of a postapproval change. It establishes the documented evidence of conformance of a pharmaceutical operation in accordance with specifications. FDA "Guideline on General Principles of Process Validation" describes in detail the principles and practices of process validation and documentation required by the regulatory authority [13]. In general terms, process validation may be defined as the procedure which generates sufficient assurance and documented evidence that a particular operation is operating and producing drug products in accordance with the specifications and process controls.

Prospective validation, retrospective validation, concurrent validation, and revalidation are the four validation components. Prospective validation is performed before the distribution of drug products in the market or after the manufacturing of a drug product using revised changes that can affect product quality and characteristics. Retrospective validation is conducted for an established drug product whose manufacturing process is stable to ensure that the current pharmaceutical operation is performing as per the protocols and specification and yielding satisfactory product. Concurrent validation is conducted by monitoring in-process critical manufacturing parameters and end-product testing to ensure that the current manufacturing process is per the in-process control specifications. Revalidation is performed after changes to an approved drug product are implemented to ascertain that there is no adverse effect on the quality and performance of a drug product [16].

During a validation process, the products and processes are subjected to testing at extreme conditions of in-process limits and their performance is evaluated against the acceptance criteria. The parameters of different pharmaceutical operations are varied and product properties are recorded and evaluated (Figure 3). When it is found that adjustment is required, necessary actions are taken in consultation with R&D personnel. Generally, validation data of three production scale batches are compared to generate a high level of quality assurance.

Systematic documentation of the effect on the product attributes by varying various process parameters is very important in the validation process. The product development team, engineering and technical services, and production and regulatory departments are also consulted while making any process change or before finalizing any validation protocol or report. Depending on the "level" of change or degree of effect to be produced, the extent of the validation is determined.

Based on the validation requirements, samples are collected at different stages and submitted for analysis per the validation protocol. The data are finally compiled in the form of a validation report. A systematic validation protocol and validation report are the backbone of the validation process. Table 12 gives key components of any validation activity [16]. These protocols and reports should be verified and approved by the relevant functions.

Some changes are often made in the manufacturing process without prior notification, and hence it is advisable to consider revalidation at predetermined frequencies (or whenever an unusual behavior is noted).

When new equipment is purchased or there is a change in the manufacturing site, qualification exercises are performed as part of the validation process. Qualification (installation qualification, operation qualification, and performance qualification) for any equipment or facility is an extreme process which involves testing, verification, and documentation to assure that the particular equipment or facility is per the specification and meets the appropriate standards as defined by vendor and required by manufacturing and engineering personnel [14].

## 6 CONCLUSION

The global pharmaceutical industry is continuously growing in a rapidly changing and dynamic environment of the health care sector. New drugs and delivery systems

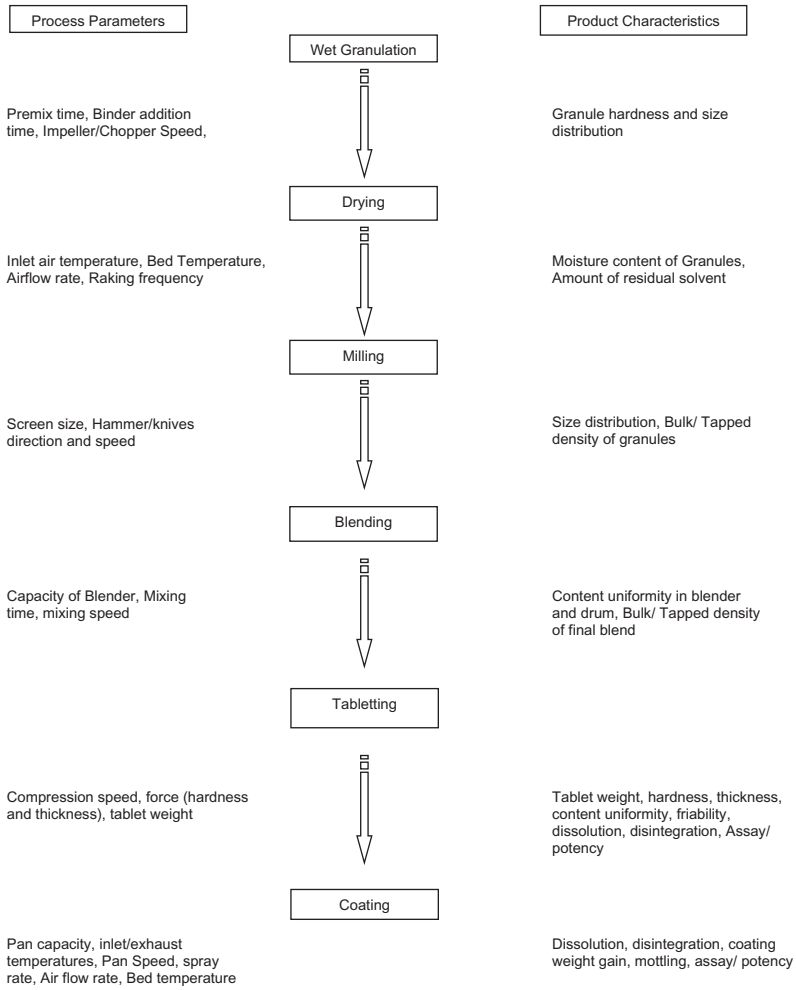


**FIGURE 3** Various process parameters and product characteristics associated with validation activity of typical coated tablet.

**TABLE 12 Key Components of Validation Activity**

| Validation Protocol | Validation Report |
|---|---|
| Purpose of study | Aim of study |
| Personnel responsibility | List of raw material used in study |
| Critical process steps | List of manufacturing equipment |
| Critical process parameters | Critical steps studied |
| Critical product parameters | Collected data and its analysis |
| Sampling plan | Acceptance criteria evaluation |
| Testing plan | Statistical analysis |
| Acceptance criteria | Recommendations by validation department |

surface each year in the market. To maintain the quality of new and existing drugs and delivery technologies, pharmaceutical operations are controlled by regulatory guidelines. The purpose of developing guidelines is to keep the health and safety of a person on the highest priority by delivering quality pharmaceuticals. Implementation of these guidelines and systematic follow-up of the effect of postapproval changes in the form of documentation are essential to safeguard against any possible failure of the whole system. Change control and validation ensure that there is no deleterious impact on the drug product characteristics. Anticipated changes incorporated in comparability protocols reduce significant risk of experiencing unpredictable adverse effects and help to introduce the product in less time. When an impact is anticipated, it should be properly discussed with R&D, process development, and other concerned departments for appropriate regulatory filing by following regulatory guidelines. Provided that these guidelines are followed properly, quality and performance of a drug product can be ensured.

## REFERENCES

1. U.S. Department of Health and Human Services, Food and Drug Administration, Centre for Drug Evaluation and Research (CDER), Guidance for industry: Changes to an approved NDA or ANDA, available: http://www.fda.gov/cder/guidance/3516fnl.pdf, accessed Apr. 15, 2006.
2. European Commission, Guideline on dossier requirements for type IA and type IB notifications: Pharmaceuticals: Regulatory framework and market authorizations, available: http://ec.europa.eu/enterprise/pharmaceuticals/eudralex/vol-2/c/gdvartypiab_rev0_200307.pdf, accessed Apr. 20, 2006.
3. Department of Health and Ageing, Therapeutic Goods Administration, Australian regulatory guidelines for prescription medicines. Appendix 12: Changes to the quality information of registered medicines: Notification. Self-assessment and prior approval, available: http://www.tga.gov.au/pmeds/argpmap12.pdf, accessed Apr. 12, 2006.
4. Food and Drug Administration, Centre for Drug Evaluation and Research (CDER), Guidance for industry: SUPAC-IR: Immediate-release solid oral dosage forms: Scale-up and post-approval changes: Chemistry, manufacturing and controls, in vitro dissolution testing, and in vivo bioequivalence documentation, available: http://www.fda.gov/cder/guidance/cmc5.pdf, accessed May 11, 2006.
5. Food and Drug Administration, Centre for Drug Evaluation and Research (CDER), Guidance for industry: Comparability protocols—Chemistry, manufacturing, and controls information (draft), available: http://www.fda.gov/cder/guidance/5427dft.pdf, accessed Apr. 15, 2006.
6. Brazilian Sanitary Surveillance Agency (ANVISA), Resolution: Guide for making post-registration alterations, inclusions and notifications of drug products, Brazil's generic drug policy, industry legislation, available: http://www.anvisa.gov.br/hotsite/genericos/legis/resolucoes/893_03re_e.htm, accessed Apr. 11, 2006.
7. U.S. Department of Health and Human Services, Food and Drug Administration, Centre for Drug Evaluation and Research (CDER), Guidance for industry: SUPAC-MR: Modified release solid oral dosage forms scale-up and postapproval changes: Chemistry, manufacturing, and controls; in vitro dissolution testing and in vivo bioequivalence documentation, available: http://www.fda.gov/cder/guidance/1214fnl.pdf, accessed May 5, 2006.

8. U.S. Department of Health and Human Services, Food and Drug Administration, Centre for Drug Evaluation and Research (CDER), Guidance for industry: SUPAC-SS: Nonsterile semisolid dosage forms; scale-up and post-approval changes: Chemistry, manufacturing and controls; in vitro release testing and in vivo bioequivalence documentation, available: http://www.fda.gov/cder/guidance/1447fnl.pdf, accessed Apr. 24, 2006.
9. Pharmaceutical Unit of European Commission at EUROPA, available: http://ec.europa.eu/enterprise/pharmaceuticals/pharmacos/docs/brochure/pharmaeu.pdf, accessed May 21, 2006.
10. Variations. Pharmaceuticals: Regulatory framework and market authorizations, Chapter 5, in *Procedures for Marketing Authorisation*, Vol. 2A, European Commission, available: http://ec.europa.eu/enterprise/pharmaceuticals/eudralex/vol-2/a/v2a_chap5_r1_2004-02.pdf, accessed Apr. 16, 2006.
11. Commission regulation (EC) No. 1084/2003 of June 3, 2003, concerning the examination of variations to the terms of a marketing authorisation for medicinal products for human use and veterinary medicinal products granted by a competent authority of a member state (Official Journal L 159, 27/6/2003, pp. 1–23), available: http://ec.europa.eu/enterprise/pharmaceuticals/eudralex/homev1.htm, accessed Apr. 10, 2006.
12. The ICH process for harmonisation of guidelines, available: http://www.ich.org/cache/compo/276-254-1.html, accessed May 15, 2006.
13. Food and Drug Administration, Centre for Drug Evaluation and Research (CDER), Guideline on general principles of process validation, available: http://www.fda.gov/cder/guidance/pv.htm, accessed Apr. 26, 2006.
14. Willig, S. H. (2001), Production and process controls, in Swarbrick, J., Ed., *Good Manufacturing Practices for Pharmaceuticals: A Plan for Total Quality Control from Manufacturer to Consumer*, Marcel Dekker, New York, pp. 99–138.
15. Waterland, N. H., and Kowtna, C. C. (2003), Change control and SUPAC, in Nash, R. A., and Wachter, A. H., Eds., *Pharmaceutical Process Validation*, Marcel Dekker, New York, pp. 699–748.
16. Ahmed, S. U., Naini, V., and Wadgaonkar, D. (2005), Scale-up, process validation and technology transfer, in Shargel, L., and Kanfer, I., Eds., *Generic Drug Product Development: Solid Oral Dosage Form*, Marcel Dekker, New York, pp. 95–136.

# GMP-COMPLIANT PROPAGATION OF HUMAN MULTIPOTENT MESENCHYMAL STROMAL CELLS

Eva Rohde, Katharina Schallmoser, Christina Bartmann, Andreas Reinisch, and Dirk Strunk
*Medical University of Graz, Graz, Austria*

**Contents**

1 Introduction
2 Acronyms and Definitions
  2.1 Mesenchymal Stromal Cells
  2.2 Somatic Stem Cell Therapy
  2.3 Good Manufacturing Practice
  2.4 Cell-Based Medicinal Products
  2.5 Human Platelet Lysate
3 Approaches
  3.1 Adherence to Principles of GMP in a Preclinical Developmental Process
  3.2 Efficient Standardized MSC Propagation Using Low Cell Seeding Density
  3.3 Superior MSC Proliferation Resulting from HPL-Driven as Compared to FBS-Driven Cultures
  3.4 Contamination Risks Can Be Minimized in Rational MSC Propagation Procedures
4 Testing Methods
  4.1 Safety and Efficacy of CBMP in Preclinical Stage
  4.2 Quality Controls During Cell Culture (In-Process Controls) and Final Product Release Criteria
  4.3 MSC Functionality and Potency Assays
5 Conclusion
References

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

## 1 INTRODUCTION

Somatic stem cell therapy (SCT) is a rapidly growing field that opens a broad spectrum of therapeutic options. The concept of regenerative SCT is based on the assumption that transplantation of adult human stem cells may support organ regeneration, modulate immunity, and regulate hematopoiesis. Transplantation of bone marrow (BM)–derived hematopoietic stem cells (SCs) for blood and immune system regeneration has been a clinical reality for almost 40 years. The existence of detectable numbers of mesenchymal and endothelial progenitors within blood and BM has promoted the readily harvestable hematopoietic tissue as a source of SCs for nonhematopoietic regenerative SCT (Figure 1).

Multipotent mesenchymal stromal cells (MSCs) are currently undergoing evaluation in a number of clinical trials (www.clinicaltrials.gov). These nonhematopoietic cells have been first described by Friedenstein et al. in a fibroblast colony-forming unit assay (CFU-F) based on low-density culture of adherent BM-derived cells [1–3]. Alternative sources for MSCs have been identified in a number of studies showing the successful isolation of fibroblast precursors from umbilical cord blood, placenta, umbilical cord, amniotic fluid, and adipose tissue [4–13]. To date, most experimental and clinical experience has been accumulated with BM-MSC [14–21]. Ex vivo expansion of these rare BM constituents (representing less than 1% of aspirated BM nucleated cells) is a prerequisite to achieve a reasonable MSC application dose of at least $2 \times 10^6$ MSCs/kg of the recipients' body weight. The majority of expansion procedures are currently based on the use of fetal bovine serum (FBS), which carries the risk of xenoimmunization and transmission of known (e.g., prions transmitting bovine spongiforme encephalopathia, BSE) and unknown pathogens. These risks could be avoided by developing MSC expansion protocols that use human alternatives which replace FBS.

The preclinical development of medicinal products in general bears high complexity due to the lack of fixed routines. Long term manipulations of cell-based medicinal products (CBMPs) may enhance the risks for undesirable effects in the course of ex vivo cell expansions. Safety concerns regarding the clinical application of ex vivo generated MSCs require a logistic environment providing an established good manufacturing practice (GMP) background embedded in a highly effective quality system. Demonstration of manufacturing and product consistency is achieved by applying rational in-process controls. Release criteria should ideally emerge from successful product development and optionally include the sterility, safety, purity, identity, and potency.[1] They must to be rapid, sensitive, and reliable and should retain some flexibility in type and timing of testing. The complexity and function of different CBMPs require an array of analytical procedures to adequately characterize the particular product (potency assays). Personalized (patient-specific) CBMPs differ from large drug batches in the pharmaceutical industry in terms of practicability in final product release in that they may require process-oriented rather than single-product potency testing. U.S. regulations demand that "tests for potency shall consist of either in vitro or in vivo tests, or both, which have been specifically

[1]U.S. legislation: 21 CFR 610, General biological products standards, CFR 610.10 Potency, CFR 600.3(s).

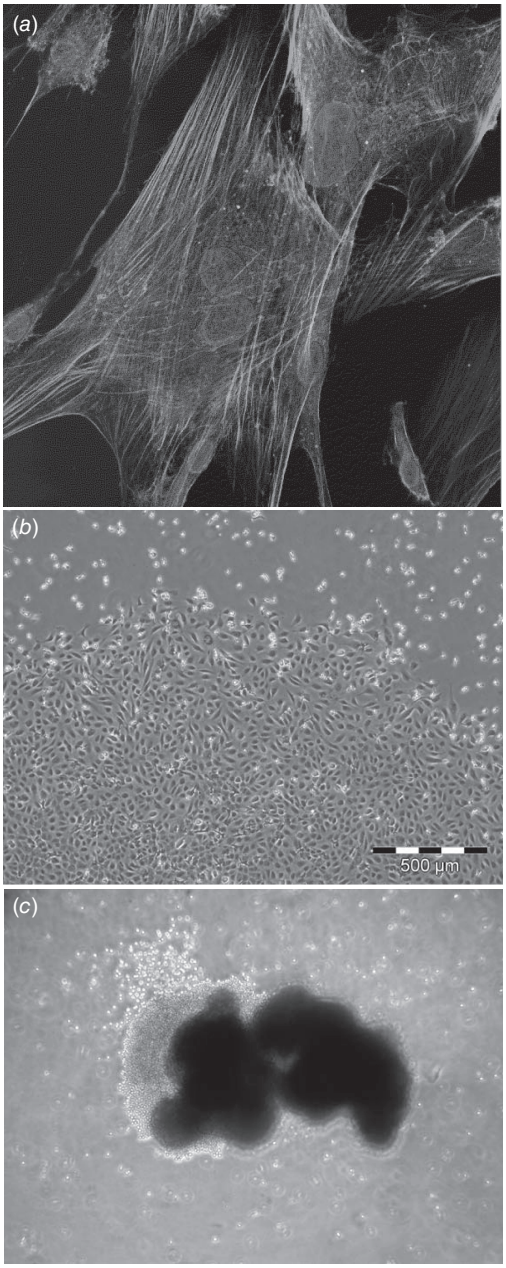


**FIGURE 1** Hematopoietic tissue-derived SC and progenitors. Hematopoietic tissue contains (*a*) mesenchymal and (*b*) endothelial in addition to (*c*) hematopoietic progenitor cells. (*a*) Adult human BM-derived MSCs were stained to visualize the actin cytoskeleton, mitochondria, and nuclei. (*b*) The periphery of an umbilical cord blood–derived endothelial progenitor cell (EPC) colony is depicted demonstrating typical cobble stone–like morphology. The entire colony was derived from a single UCB-EPC indicating impressive proliferation potential (more than 70,000 cells were obtained by harvesting single EPC-derived colonies indicating the completion of at least 16 population doublings). Less than 10 mL of adult BM [(*a*) and (*c*)] but at least 40 mL of UCB (*b*) were sufficient to generate appropriate numbers of cells for therapeutic purposes.

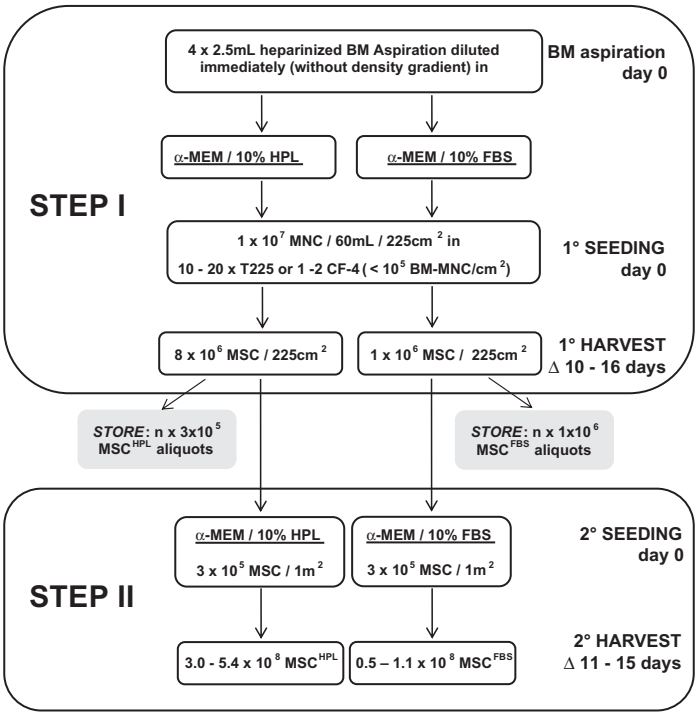


**FIGURE 2** GMP-compliant propagation of human MSCs. The summary of a two-step MSC production procedure shows seeding and harvest numbers of BM-MNC and resulting numbers of $MSC^{HPL}$ as compared to $MSC^{FBS}$. (Reproduced with permission from ref. 23.)

designed for each product so as to indicate its potency in a manner adequate to satisfy the interpretation of potency given by the definition in 21 CFR 600.3(s)." Functional analyses accompanying the expansion process development leading to a full product characterization and optimization of manufacturing steps are prerequisites that allow for the creation of a safe and effective CBMP.

This chapter demonstrates that rapid and standardized expansion of human MSCs to achieve a reasonable cell dose (i.e., $\geq 2 \times 10^6$/kg body weight of a 75-kg person corresponding to $\geq 1.5 \times 10^8$ MSCs) is feasible within less than four weeks. Replacing FBS with human platelet lysate (HPL) provides one strategy toward a safer CBMP (Figure 2). Appropriate preclinical development adhering to GMP principles will enhance safety in the course of a consecutive clinical evaluation of MSCs as a therapeutic agent.

## 2 ACRONYMS AND DEFINITIONS

### 2.1 Mesenchymal Stromal Cells

Adhesion of mononuclear cells from human bone marrow aspirates (BM-MNC) to tissue culture plastic and removal of nonadherent cells during the first days of culture selects for a population of proliferating spindle-shaped fibroblast-like non-

hematopoietic multipotent MSCs. Mesenchymal stromal cells can also be obtained from umbilical cord blood, umbilical cord, placenta, adipose tissue, and several fetal tissues. The minimum criteria for MSCs are defined in an ISCT (International Society for Cellular Therapy) position paper published in 2006 [22]. The MSCs have a high self-renewal potential and the capacity to be differentiated in vitro into progeny displaying an osteo-, chondro-, or adipogenic phenotype.

### 2.2 Somatic Stem Cell Therapy

The concept of regenerative SCT is based on experimental and early clinical observations indicating that the application of adult stem cells can improve organ regeneration after ischemic, toxic, or metabolic injury. Bone marrow harbors hematopoietic and mesenchymal stem cells and endothelial progenitor cells and is an easily accessible but not the sole, source of candidate cells to promote organ repair after systemic or local application. Regulation of hematopoiesis and immune modulation are the two established applications in the broad field of SCT with autologous and allogeneic stem and progenitor cells.

### 2.3 Good Manufacturing Practice

Good manufacturing practice is that part of the quality management system (QMS) that is concerned with the production and quality control of medicinal products (drugs) for human and veterinary use. It includes documentation, personnel training, facility, equipment, and process controls for the manufacture of pharmaceuticals.

### 2.4 Cell-Based Medicinal Products

Medicinal products containing viable cells are summarized under the umbrella term *cell-based medicinal products*. The term CBMP does not cover products containing nonviable cells or cellular fragments. The CBMPs may have much potential in the treatment of various diseases that to date have no cure. They are heterogeneous in terms of origin and type of cells and with regard to the complexity of the product. Cells may be self-renewing stem cells, more committed progenitors, or terminally differentiated cells exerting a specific regenerative function. Cells may be of autologous or allogeneic origin. Cells may be used alone or in combination with biomolecules, chemical substances, or structural materials that possibly potentiate their desired effects.

### 2.5 Human Platelet Lysate

Human platelet lysate can be obtained from buffy coat–derived platelet rich plasma. The platelet fraction is separated from the plasma and the white and red blood cell fraction by centrifugation steps and concentrated to a density of at least $1 \times 10^9$ platelets/mL. Platelets can either be activated with thrombin or lysed by repeated freeze–thaw cycles. Both mechanisms result in the release of growth factors and mitogens that are stored in intact platelets. Mediators released from platelets include, among others, epidermal growth factor (EGF), basic fibroblast growth factor (bFGF), platelet-derived growth factors (PDGFs), transforming growth factor (TGF-ß1), and insulinlike growth factor (IGF) [23, 24]. Perhaps HPL may replace FBS in many

cell culture systems that have previously been thought to strictly depend on the presence of FBS.

## 3 APPROACHES

### 3.1 Adherence to Principles of GMP in a Preclinical Developmental Process

The standardized MSC propagation should be conducted as a well-planned, consistently documented, and optimized procedure that also minimizes risks of microbiological, particulate and pyrogen contamination by reducing manipulation steps and manipulation time. According to current European legislation,[2] the principles of GMP should be applied to CBMP when they are manufactured for use in human subjects in phase 1 studies. These requirements do not apply to cellular or tissue-based medicinal products used in phase 1 studies according to U.S. legislation[3] or to products in the preclinical developmental phase. If it is expected that preclinical findings are to be translated into clinical use rather rapidly, it may be recommended to establish GMP-compliant technology during the preclinical developmental phase of any cell product. As a result, this ensures that products are consistently produced and controlled to meet the quality standards appropriate for their intended use or product specification. The GMP requirements are well described in "PIC/S Guide to Good Manufacturing Practice for Medicinal Products" and include the implementation of an efficiently running quality management system, dedicated areas for manufacture of sterile medicinal products complying with GMP, appropriately qualified and trained personnel, suitable equipment, correct materials, containers and labels, approved procedures and instructions, suitable storage and transport facilities, and a record-keeping system that allows the complete history of a medicinal product to be traced (See http://www.picscheme.org). It is a challenge to conduct preclinical research and development complying to GMP as procedures routinely turn out to be much more time and cost intensive than common laboratory-scale research. These circumstances can advance either the developmental progress at the expense of quality standards or vice versa. It should therefore be decided on a case-by-case basis how closely to adhere to GMP standards depending on the more or less stringent time schedule for the considered clinical use of a CBMP.

### 3.2 Efficient Standardized MSC Propagation Using Low Cell Seeding Density

The future use of MSCs in clinical studies may require very high absolute MSC numbers to gain appropriate cell doses ($>5 \times 10^6$/kg body weight) per patient compared to in vivo experimental models with small animals [20]. It is consequently advantageous to develop large-scale MSC expansion protocols that allow for the

[2]European legislation: Directive 65/65/EEC, Directive 75/318/EEC, Directive 75/318/EEC, Commission Communication on the Community marketing authorisation procedures for medicinal products (98/C229/03); Directive 2001/20/EC, EMEA/CHMP/410869/2006.

[3]U.S. legislation: 21 CFR 210; 21 CFR 211; 21 CFR 312.21; 21 CFR 312.22(a) and 21 CFR 312.23(a)(7)(i).

generation of up to $5 \times 10^8$–$10 \times 10^8$ MSCs from the limited starting volume of primary material.

The cell seeding density is of critical importance for the expansion rate of MSCs and must be defined for the primary seeding and the following passaging steps. Most experimental and clinical expansions described to date were started with a high seeding density of more than $1 \times 10^5$ BM-MNC/cm$^2$ [2, 14, 16]. For further passages pioneering studies showed that a very low seeding density between 0.5 and 10 MSCs/cm$^2$ selects for the expansion of a rapidly proliferating subpopulation of recycling SCs, termed RS cells [25–28]. This seeding density, referred to as "clonal density," would necessitate a theoretical growth area of from 2,000,000 to 100,000 cm$^2$ (from 200 to 10 m$^2$) to obtain a clinical quantity of $>1 \times 10^8$ MSCs from $1 \times 10^6$ starting MSCs within one passage. Plating 30–100 MSCs/cm$^2$ therefore is a reasonable compromise density requiring a more realistic growth area between 10,000 and 25,000 cm$^2$ (1 and 2.5 m$^2$). We have recently shown that the primary seeding of only 10 mL bone marrow aspirates on approximately 0.2 m$^2$ culture area for two weeks (culture step 1; BM diluted immediately after aspiration in culture medium without density gradient separation; removal of nonadherent cells at day 3) followed by an expansion on 2.5 m$^2$ (step 2) is sufficient to consistently generate at least $1.5 \times 10^8$ MSCs in FBS-supplemented medium within less than four weeks (Figure 2) [29]. This study furthermore corroborated earlier data on the inverse correlation of the seeding density to MSC proliferation (Figure 3) [25–28].

### 3.3 Superior MSC Proliferation Resulting from HPL-Driven as Compared to FBS-Driven Cultures

The most commonly used basic cell culture medium compositions for MSC propagation are minimum essential medium alpha (α-MEM) and low-glucose (1 g/L) Dulbecco's modified Eagle medium (DMEM-LG) supplemented with L-glutamin, antibiotics, and 5–20% FBS [14, 16, 19, 24, 25, 30]. Our experience with MSC propagation relates to the use of α-MEM supplemented with either FBS or HPL. In contrast to HPL that has been recognized only recently as a potent culture medium supplement [24], FBS is a well-known key medium supplement for cell culture and its role has been unchallenged for more than 50 years [31]. The common use of FBS in MSC cultures as a source of growth factors and mitogens bears the risk of transmission of known and unknown pathogens as well as xenoimmunization against bovine pathogens and should therefore be avoided for clinical use [32, 33].

In a recent study we analyzed the capacity of HPL to replace FBS in large-scale (clinical) MSC expansions and were able to demonstrate a superior propagation of MSC cultured with HPL ($MSC^{HPL}$) as compared to MSC derived from FBS-driven cultures ($MSC^{FBS}$) [23]. Figure 4 illustrates superior MSC proliferation at low plating density and higher population doublings (PDs) with HPL after a culture period of less than 14 days.

### 3.4 Contamination Risks Can Be Minimized in Rational MSC Propagation Procedures

Cell expansion is mainly performed according to labor-intensive time-consuming protocols using open systems that increase the risks of microbiological or particulate contamination and supplementation with potent antibiotics to control these prob-

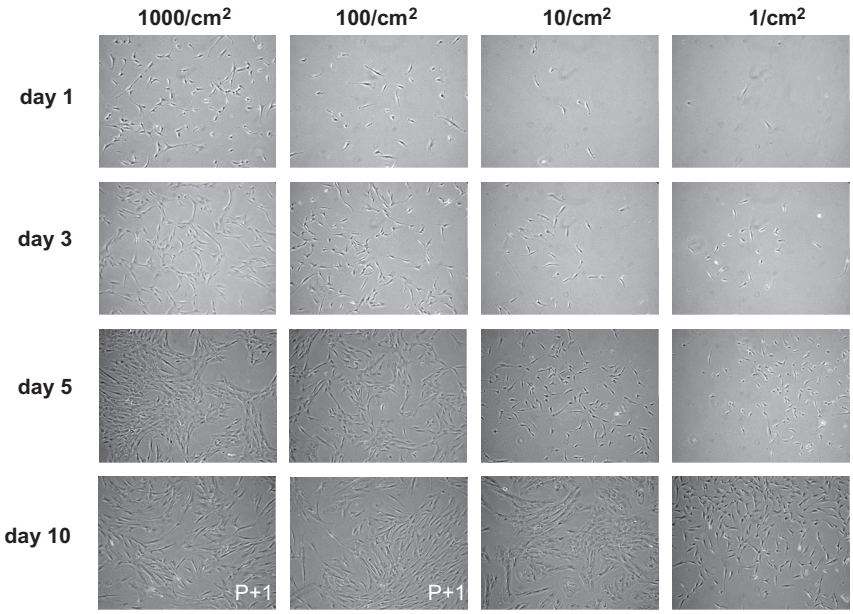


**FIGURE 3** Inverse correlation of seeding density to MSC proliferation. BM-derived MSCs derived from passage 2 were seeded at log fold deescalated density of 1000, 100, 10, and 1 $cm^{-2}$. Photographs were taken after 1, 3, 5, and 10 days of culture in α-MEM/10% FBS (original magnification 40×). In the case of MSC seeded at 100 and 1000 cells/$cm^2$ confluence necessitated trypsinisation between days 5 and 10 followed by reseeding at 100 and 1000 cells/$cm^2$, respectively, and is therefore indicated as P + 1.

lems. Avoiding the use of penicillin during clinical-scale cell propagation follows the rationale to reduce the risk of sensitization as well as anaphylactic precipitation. Thus, it may be worthwhile not using other antibiotics for the GMP-compliant MSC propagation. One approach to minimize potential contamination risks is to rigorously reduce handling in the course of MSC propagation to the absolute minimum of necessary steps. In our experience, the commonly used density gradient centrifugation step can be skipped prior to the primary cell seeding of the bone marrow aspirate. Immediate dilution of limited volumes (e.g., 10–20 mL) of heparinized BM aspirate into supplemented α-MEM medium for direct cell seeding does not result in a loss of MSC recovery [23]. Furthermore, the aforementioned low cell seeding density and the employment of an increased growth area in a simplified procedure together with the use of HPL in fact allow for an efficient production of high MSC numbers within one to two harvest-replating cycles. The relatively short ex vivo expansion time of less than three to four weeks may be helpful in reducing the cumulative risk of contamination.

## 4 TESTING METHODS

### 4.1 Safety and Efficacy of CBMP in Preclinical Stage

The preclinical developmental period should be used for the extensive characterization of the CBMP. Release criteria have to be defined and reasonable time frames

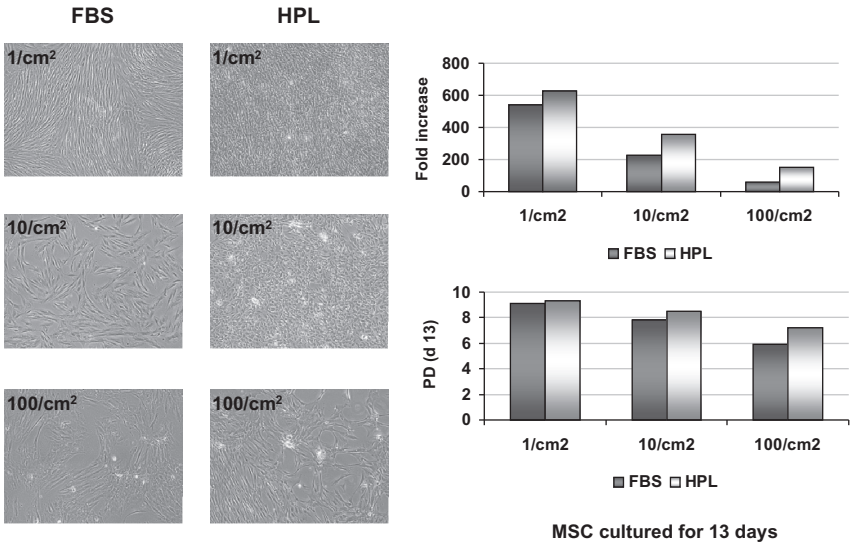


**FIGURE 4** MSC proliferation capacity depends on seeding density in xenogeneic FBS and HPL-supplemented cultures. The inverse correlation of MSC proliferation to their seeding density resulted in the formation of a confluent MSC layer in cultures starting with 1 MSC/cm$^2$ in α-MEM/10% FBS and cultures starting with 1–10 MSCs/cm$^2$ in α-MEM/10% HPL but not when initiating cultures with the respective higher seeding densities within less than two weeks. The calculated fold increase of the cell number and corresponding population doublings from a representative experiment harvested at day 13 are shown.

must be set to allow for a high safety and quality standard of the final cellular product. On the other hand, the logistic background should allow for a rapid release of the CBMP within a few hours due to the potential short shelf life of many cellular products. Ranges of cell purity, sterility, and absence of pyrogens and endotoxins are factors of utmost importance which must be determined. It is an inherent feature of CBMPs that product specifications must be adapted to the individual application. The challenge in the preclinical developmental phase is to find satisfactory answers to unresolved questions in terms of cell type, source, dose, and mode of application according to the particular target disease. Thus, in-process controls and definitive release criteria must be met by each CBMP. Since many CBMPs are personalized medicine, potency assays must be performed for selected representative products (e.g., before initiating a study and consecutively once per year).

### 4.2 Quality Controls During Cell Culture (In-Process Controls) and Final Product Release Criteria

***General Safety*** According the Food and Drug Administration (FDA), cellular therapy products are exempt from general safety testing [21 CFR 610.11(g)(1)].

***Cell Dose*** The preclinical stage can be used to determine the specifications for the minimum effective and maximum tolerable number of viable and functional cells. The optimum dose of cells to be administered still needs to be established [20].

***Viability*** Viability of MSCs can easily be determined immediately after trypsinization via trypan blue or 7-amino-actinomycin D (7-AAD) exclusion. According to the specifications developed from our cell culture studies, viability should be >90%. In selected exceptional cases a lower limit of 70% viability of total harvested cells may be acceptable.

***Microbiological Testing*** Sterility testing that detects fungal, anaerobic, and aerobic bacterial and mycoplasma contamination should be performed after each critical manipulation step during MSC culture that is prone to microbiological contamination [34]. The crucial bacterial sterility check at the end of last harvesting step cannot be evaluated prior to in vivo application if MSCs need to be applied immediately after propagation due to the duration of the cultures. Mycoplasma polymerase chain reaction (PCR) results can be obtained at the day of harvest within less than 6 h. MycoAlert® results are available within less than 1 h at the day of harvest. Definitive culture results to exclude mycoplasma contamination are available within two to three weeks and therefore are not applicable for CBMPs with a short shelf life that are planned to be administered immediately after production.

***Endotoxin and Pyrogenicity Testing*** Endotoxin measurement using the Limulus amebocyte lysate (LAL) assay is typically done as an alternative to pyrogenicity testing for early phase trials. For any parenteral drugs, except those administered intrathecally, the FDA recommends that the upper limit for endotoxin be 5 EU/kg body weight/dose. The LAL assay method can be applied to the safety evaluation of biological preparations according to existing regulations.[4] We use the LAL assay to substitute for the lengthy delay in microbiological data availability to obtain results prior to the clinical application of the final product within less than 2 h after harvest.

***Phenotypic Identity of MSC*** In addition to morphological identification by microscopy, the immunophenotypic characterization of MSCs can be done using a broad panel of fluorescence-conjugated antibodies directed against surface molecules. To date there is no specific marker uniquely defining MSCs. Therefore a profile is used to show the expression of certain markers and to exclude the contamination by cells expressing other marker profiles. Flow cytometry is recommended by the ISCT to reveal that MSCs stain positive for CD73, CD90, and CD105 and negative for HLA-DR, CD14, CD31, CD34, and CD45 (Figure 5) [22, 23]. Much more extensive phenotypic analyses have been performed without retrieving additional information about MSC type or function [35]. Gene expression profiling will hopefully result in a better definition of human MSCs [35–42].

### 4.3 MSC Functionality and Potency Assays

***Clonogenicity*** The self-renewal capacity of cells and the proportion of proliferating cells within a heterogeneous cell mixture can be evaluated using the CFU assay.

[4]Endotoxin testing, LAL, according to Eur. Pharm. 2.6.14 and Guideline on Validation of the Limulus Amebocyte Lysate Test as an End-Product Endotoxin Test for Human and Animal Parenteral Drug, Biological Products and Medical Devices, 1987, Sections I–IV, http://www.fda.gov/cber/gdlns/lal.pdf.

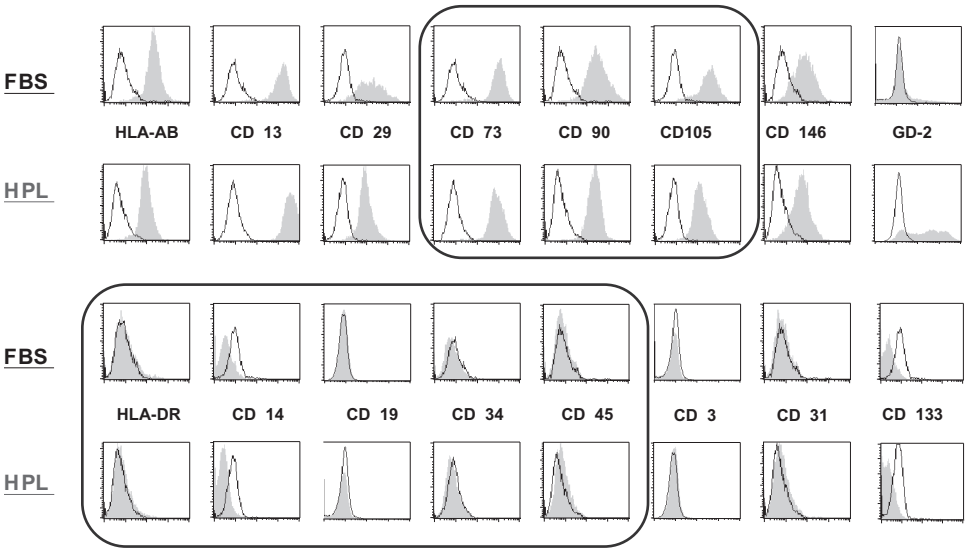


**FIGURE 5** Immune phenotype of human MSCs. Flow cytometric analysis of at least 10,000 viable MSCs was used to determine antibody reactivity (gray-filled histograms) compared to appropriately diluted isotype controls (black line). Phenotypic criteria require positivity (≥90%) for CD73, CD90, and CD105 and negativity (≤2%) for HLA-DR, CD14 (or CD11b), CD19 (or CD79α), CD34, and CD45. Absence of CD3+ T cells may be desirable in the case of GvHD treatment. Depending on the culture conditions, MSCs share reactivity with the anti-disialoganglioside antibody GD2 with neuroblastoma cells, melanoma, and small-cell lung cancer cells.

A tissue culture method allowing for the clone counting of cells was first describend in 1956 [43]. The introduction of bone marrow CFU assays led to the discovery of hematopoietic stem cells [44]. Fibroblast precursors existing within the hematopoietic system also have been evaluated with another specific CFU assay method introduced by Friedenstein in 1974 (CFU-F) [2]. We analyzed the clonal expansion capacity of MSC with the CFU-F method. Figure 6 shows differences in CFU-F appearance between $MSC^{HPL}$ and $MSC^{FBS}$. In the case of primary BM appropriate dilution is necessary to determine the CFU-F frequency (Figure 7). Once MSCs are enriched, the appropriate MSC seeding density recommended for CFU-F enumeration may range from 1 to 5 MSCs/$cm^2$ [2, 29].

***Osteo-, Chondro-, and Adipogenic Differentiation*** Isolated BM-derived MSCs were shown to differentiate along multiple mesechymal lineages in 1999 [45]. Evidence suggests MSCs can also express phenotypic characteristics of endothelial, neural, smooth muscle, skeletal myoblast, and cardiac myocyte cells [46]. The prototype pathways of MSC differentiation occur along osteogenic, chondrogenic, and adipogenic lineages and have been extensively demonstrated in a large number of publications [47]. This kind of potency assay may be performed regularly if bone or connective tissue repair is intended, although time limits do not enable immediate product release.

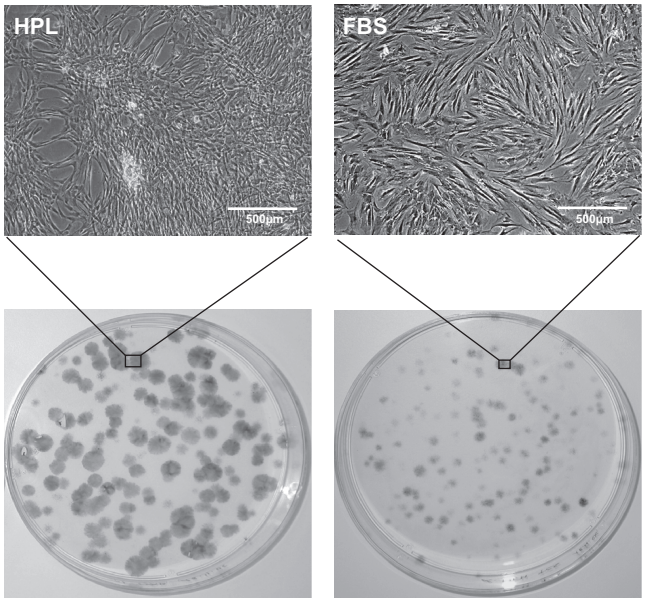


**FIGURE 6** Morphological evaluation of MSCs. CFU-F of $MSC^{HPL}$ compared to $MSC^{FBS}$ differ in size, morphology, and density (scale bar identifies magnification in the upper panel; colony photographs taken on day 12, 40× original magnification).

***Immune Modulatory Effects*** Mesenchymal stromal cells inhibit T-cell alloreactivity in mixed lymphocyte cultures (MLCs) or lymphocyte proliferation induced by mitogens, such as phytohemaglutinin (PHA) or concanavalin A [29, 48–51]. It is of note that high concentrations of MSCs (representing 10–40 MSCs per 100 responder lymphocytes) have an inhibitory effect while low MSC concentrations (0.1–1%) may stimulate lymphocyte proliferation in mixed lymphocyte cultures [50]. If MSCs are used for immunosuppressive therapies, these findings may imply that high doses of MSCs are needed to inhibit T-cell proliferation in patients with graft-versus-host disease following allogeneic bone marrow transplantation. The application of low MSC numbers could stimulate lymphocyte proliferation in vivo and hence result in an undesirable boost to graft-versus-host disease as an adverse reaction of the MSC therapy. It is not clear so far whether the precise number of T cells in a given MSC transplant needs to be determined to exclude a potential boost to alloreactivity. Immune modulation can be measured with carboxyfluorescein diacetate *N*-succinimidyl ester (CFSE) labeling of cells to quantify proliferation in response to allogeneic or mitogenic stimuli [52]. We analyzed the loss of CFSE fluorescent intensity indicating cell proliferation by flow cytometry after culturing CFSE-labeled MNCs in the absence or presence of different numbers of MSCs [53]. The immune regulatory capacity of $MSC^{HPL}$ and $MSC^{FBS}$ was studied by measurement of allogenic MNC proliferation after co-culturing pairs of MNC from three different donors with two independent $MSC^{HPL}$ and two other $MSC^{FBS}$ (Figure 8).

***Hematopoiesis Regulation*** Regulation of the behavior of early hematopoietic progenitor cells (HPCs) can be analyzed by MSC-HPC cocultures in vitro [54].

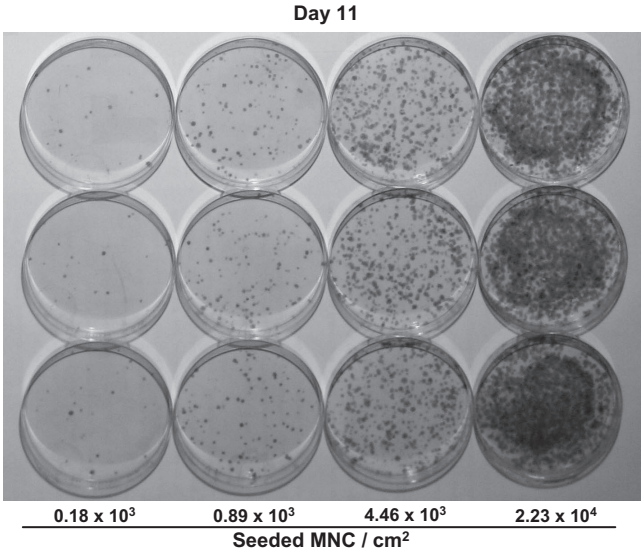


**FIGURE 7** CFU-F Evaluation of MSCs depends on BM seeding density. An appropriate dilution of the heparinised BM aspiration is needed for accurate enumeration of the primary CFU-F frequency as indicated in this representative experiment where whole heparinized BM was seeded corresponding to the respective measured BM-MNC number per square centimeters of growth area, cultured for 11 days at 37°C/humidified atmosphere/3% $O_2$/5% $CO_2$. Nonadherent cells were removed at day 3. CFU-F are visualized by Harris hematoxylin staining.

Liquid cultures of purified $CD34^+$ (HPC) with a preestablished MSC feeder layer result in the expansion of $CD34^+/38^+$ HPC and $CD34^+/38^-$ hematopoietic SCs and support the growth of mature hematopoietic total nucleated cell (TNC) progeny (Figure 9).

***Genetic Stability and Potential Tumorigenicity*** Genetic analysis of human MSCs is not well established. The significance of standard metaphase chromosome G banding is limited due to the low number of metaphases recovered during standard analyses. Advances in multicolor fluorescence in situ hybridization (FISH) and high-resolution array-based techniques may also soon be translated into practicable diagnostic tools in relation to CBMP safety in regenerative medicine [55].

Genetic instability can occur as a rare event after extended culture of mouse and human MSCs in FBS-supplemented medium [56, 57]. To test for potential in vivo tumor formation, MSCs derived from short-term clinical-scale expansions in FBS- or HPL-supplemented media were injected into immunocompromised athymic nude mice subcutaneously. Putative tumor formation was evaluated by histological analyses three months after injection of $2 \times 10^6$ and $2 \times 10^4$ MSCs and compared to controls that were injected 48 h prior to euthanasia. A primary cell deposit was visible immediately and 48 h after injection and MSCs could be recovered by conventional microscopic evaluation. However, none of 12 animals tested developed a macroscopic or microscopic detectable tumor over the 90-day observation period [23]. In this situation, genetic testing may be encouraged for prospective data acqui-

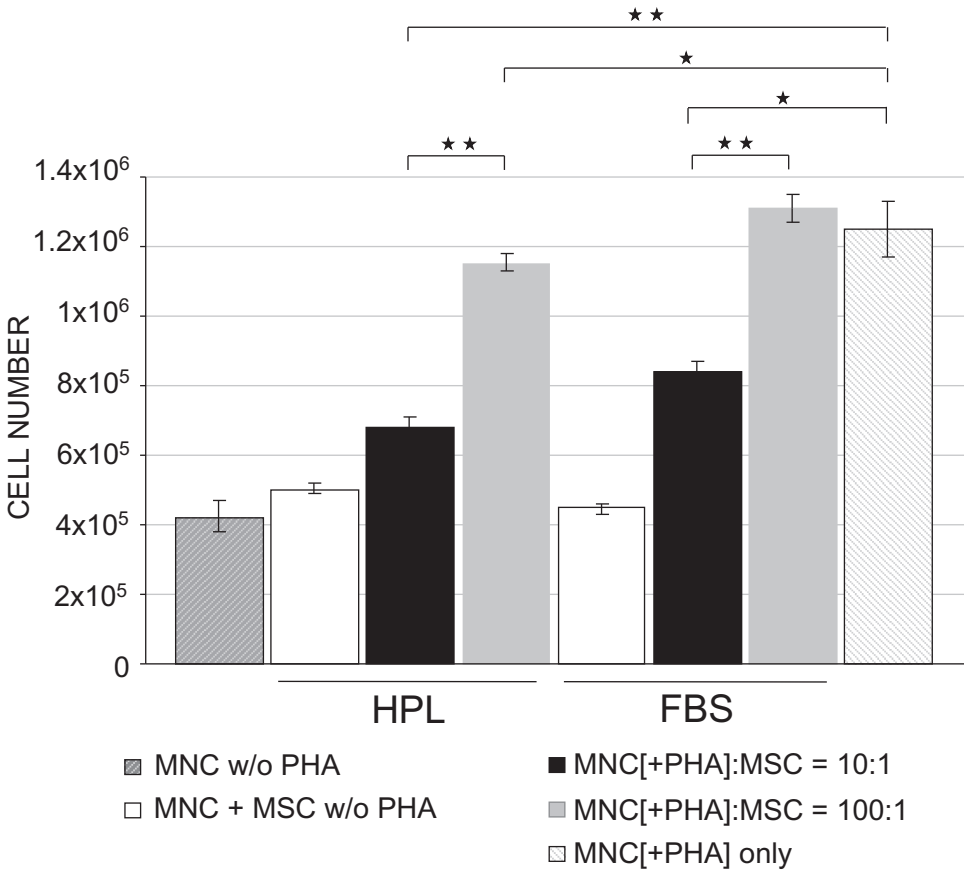


**FIGURE 8** MSC-mediated Immune Modulation. Allogeneic MNC proliferation (mean cell number ± SEM) was measured after co-culturing pairs of MNC from three different donors with two independent $MSC^{HPL}$ and two other $MSC^{FBS}$ as shown in the bar chart. MSC were added in a 1:10 ($3 \times 10^4$ MSC to $3 \times 10^5$ MNC/well; MNC[+PHA]:MSC = 10:1) or 1:100 (MNC[+PHA]:MSC = 10:1) ratio to test their influence on PHA-driven proliferation of MNC. MNC numbers were measured by flow cytometric MNC count using BD Truecount™ tubes. As a control numbers of mitogen stimulated MNC without additional MSC (MNC[+PHA] only) and background proliferation without PHA stimulation (MNC w/o PHA) are shown. MSC did not induce MNC proliferation (MNC + MSC w/o PHA). Significant differences are marked by asterisks (*$p < 0.05$ and **$p < 0.01$). (Figure reproduced with permission from reference 53)

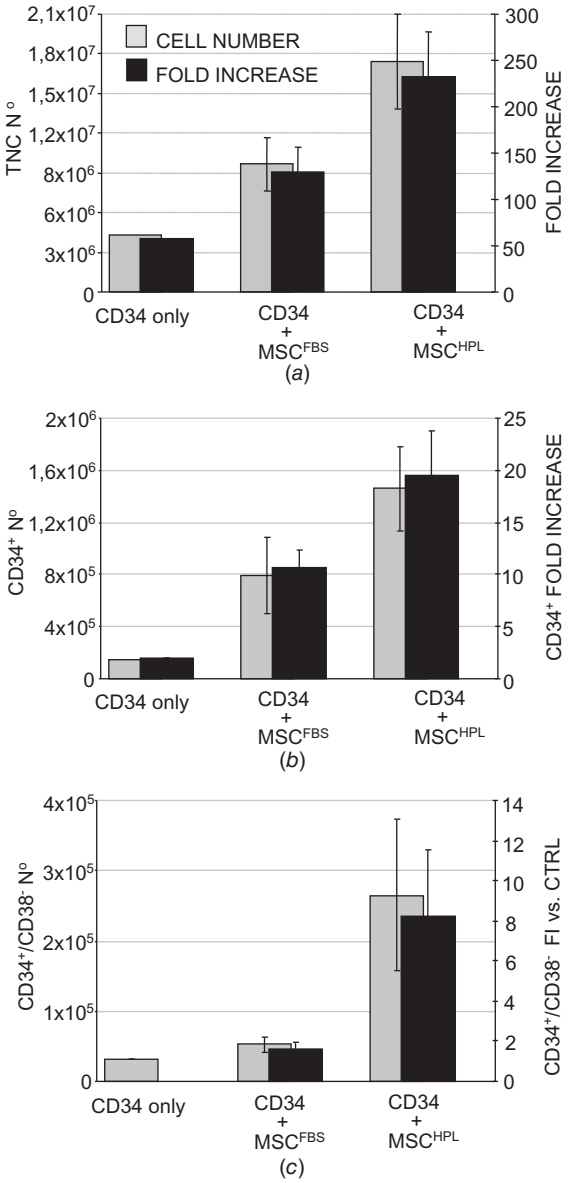


**FIGURE 9** MSC-mediated hematopoiesis regulation. (*a*) Umbilical cord blood (UCB)-derived sorted $CD34^+$ cells were expanded in cytokine-supplemented medium [Roswell Park Memorial Institue (RMPI)-1640/10% Fetal Bovine Serum (FBS)/Granulocyte and Macrophage Colony Stimulating Factor (GM-CSF)/Interleukin 3(IL-3)/Stem Cell Factor (SCF)/FMS-like tyrosin kinase 3 ligand (Flt-3L)] in the absence or presence of clinical-scale expanded MSCs. Gray bars show harvested total nucleated cell number (TNC N°.) and black bars show the fold increase (FI) of the TNC N°. compared to the starting $CD34^+$ cell number. (*b*) Harvested number of $CD34^+$ cells (gray bars) and fold increase (black bars) of $CD34^+$ cells after liquid culture with or without MSC support. (*c*) Harvested number (gray bars) and fold increase (black bars) of $CD34^+/CD38^-$ hematopoietic stem cells after liquid culture of $CD34^+$ cells with $MSC^{FBS}$ or $MSC^{HPL}$ support compared to cytokine-supplemented liquid cultures in the absence of MSCs. (Mean ± Standard Error of the Mean (SEM) of two independent expansions.) (Reproduced with permission from ref. 53.)

sition but is not considered as mandatory for product release in current MSC clinical trials.

## 5 CONCLUSION

There are considerable limitations of common pharmacological techniques used in determining the safety and efficacy of CBMPs at the preclinical stage. Conventional methods used in the pharmaceutical industry to develop pharmacological profiles and to determine the acute toxicity of drugs in animals as well as toxicity studies may not directly be translated to ex vivo generated cellular products. Nevertheless, it is inevitable that preclinical research and development of cellular products will be conducted under the guidance of either individual or consensus specifications and definitions that will continuously be improved. This approach will be helpful in developing successful therapeutic cellular agents.

## ACKNOWLEDGMENTS

This work was supported in part by The Adult Stem Cell Research Foundation (TASC$^{RF}$; C.B. and A.R.) and a young investigator fellowship of the Austrian Federal Ministry for Education, Science and Culture, bm:bwk (A.R.). The Austrian Nano-Initiative co-financed this work as part of the Nano-Health project (no. 0200), the sub-project NANO-STEM being financed by the Austrian Science Fund (FWF Project no. N211-NAN).

## REFERENCES

1. Luria, E. A., Panasyuk, A. F., and Friendenstein, A. Y. (1971), Fibroblast colony formation from monolayer cultures of blood cells, *Transfusion*, 11(6), 345–349.
2. Friedenstein, A. J., Deriglasova, U. F., Kulagina, N. N., et al. (1974), Precursors for fibroblasts in different populations of hematopoietic cells as detected by the in vitro colony assay method, *Exp. Hematol.*, 2(2), 83–92.
3. Kuznetsov, S. A., Friedenstein, A. J., and Robey, P. G. (1997), Factors required for bone marrow stromal fibroblast colony formation in vitro, *Br. J. Haematol.*, 97(3), 561–570.
4. Erices, A., Conget, P., and Minguell, J. J. (2000), Mesenchymal progenitor cells in human umbilical cord blood, *Br. J. Haematol.*, 109(1), 235–242.
5. Kogler, G., Sensken, S., Airey, J. A., et al. (2004), A new human somatic stem cell from placental cord blood with intrinsic pluripotent differentiation potential, *J. Exp. Med.*, 200(2), 123–135.
6. Bieback, K., Kern, S., Kluter, H., and Eichler, H. (2004), Critical parameters for the isolation of mesenchymal stem cells from umbilical cord blood, *Stem Cells*, 22(4), 625–634.
7. Romanov, Y. A., Svintsitskaya, V. A., and Smirnov, V. N. (2003), Searching for alternative sources of postnatal human mesenchymal stem cells: Candidate MSC-like cells from umbilical cord, *Stem Cells*, 21(1), 105–110.
8. Miao, Z., Jin, J., Chen, L., et al. (2006), Isolation of mesenchymal stem cells from human placenta: Comparison with human bone marrow mesenchymal stem cells, *Cell. Biol. Int.*, 30(9), 681–687.
9. In ‘t Anker, P. S., Scherjon, S. A., Kleijburg-van der Keur, C., et al. (2004), Isolation of mesenchymal stem cells of fetal or maternal origin from human placenta, *Stem Cells*, 22(7), 1338–1345.

10. In 't Anker, P. S., Scherjon, S. A., Kleijburg-van der Keur, C., et al. (2003), Amniotic fluid as a novel source of mesenchymal stem cells for therapeutic transplantation, *Blood*, 102(4), 1548–1549.
11. Gronthos, S., Franklin, D. M., Leddy, H. A., Robey, P. G., Storms, R. W., and Gimble, J. M. (2001), Surface protein characterization of human adipose tissue-derived stromal cells, *J. Cell Physiol.*, 189(1), 54–63.
12. Zuk, P. A., Zhu, M., Ashjian, P., et al. (2002), Human adipose tissue is a source of multipotent stem cells, *Mol. Biol. Cell*, 13(12), 4279–4295.
13. Kern, S., Eichler, H., Stoeve, J., Kluter, H., and Bieback, K. (2006), Comparative analysis of mesenchymal stem cells from bone marrow, umbilical cord blood, or adipose tissue, *Stem Cells*, 24(5), 1294–1301.
14. Lazarus, H. M., Haynesworth, S. E., Gerson, S. L., Rosenthal, N. S., and Caplan, A. I. (1995), Ex vivo expansion and subsequent infusion of human bone marrow-derived stromal progenitor cells (mesenchymal progenitor cells): Implications for therapeutic use, *Bone Marrow Transplant.*, 16(4), 557–564.
15. Koc, O. N., Peters, C., Aubourg, P., et al. (1995), Bone marrow-derived mesenchymal stem cells remain host-derived despite successful hematopoietic engraftment after allogeneic transplantation in patients with lysosomal and peroxisomal storage diseases. *Exp. Hematol.*, 27(11), 1675–1681.
16. Koc, O. N., Gerson, S. L., Cooper, B. W., et al. (2000), Rapid hematopoietic recovery after coinfusion of autologous-blood stem cells and culture-expanded marrow mesenchymal stem cells in advanced breast cancer patients receiving high-dose chemotherapy, *J. Clin. Oncol.*, 18(2), 307–316.
17. Lee, S. T., Jang, J. H., Cheong, J. W., et al. (2002), Treatment of high-risk acute myelogenous leukaemia by myeloablative chemoradiotherapy followed by co-infusion of T cell-depleted haematopoietic stem cells and culture-expanded marrow mesenchymal stem cells from a related donor with one fully mismatched human leucocyte antigen haplotype, *Br. J. Haematol.*, 118(4), 1128–1131.
18. Koc, O. N. Day, J., Nieder, M., Gerson, S. L., Lazarus, H. M., and Krivit, W. (2002), Allogeneic mesenchymal stem cell infusion for treatment of metachromatic leukodystrophy (MLD) and Hurler syndrome (MPS-IH), *Bone Marrow Transplant.*, 30(4), 215–222.
19. Le Blanc, K., Rasmusson, I., Sundberg, B., et al. (2004), Treatment of severe acute graft-versus-host disease with third party haploidentical mesenchymal stem cells, *Lancet*, 1, 363(9419), 1439–1441.
20. Lazarus, H. M., Koc, O. N., Devine, S. M., et al. (2005), Cotransplantation of HLA-identical sibling culture-expanded mesenchymal stem cells and hematopoietic stem cells in hematologic malignancy patients, *Biol. Blood Marrow Transplant.*, 11(5), 389–398.
21. Ringden, O., Uzunel, M., Rasmusson, I., et al. (2006), Mesenchymal stem cells for treatment of therapy-resistant graft-versus-host disease, *Transplantation*, 81(10), 1390–1397.
22. Dominici, M., Le Blanc, K., Mueler, I., et al. (2006), Minimal criteria for defining multipotent mesenchymal stromal cells. The International Society for Cellular Therapy position statement, *Cytotherapy*, 8(4), 315–317.
23. Schallmoser, K., Bartmann, C., Rohde, E., et al. (2007), Human platelet lysate can replace fetal bovine serum for clinical scale expansion of functional MSC, *Transfusion*. 2007 Aug; 47(8), 1436–1446.
24. Doucet, C., Ernou, I., Zhang, Y. Z., et al. (2005), Platelet lysates promote mesenchymal stem cell expansion: A safety substitute for animal serum in cell-based therapy applications. *J. Cell. Physiol.*, 205(2), 228–236.

25. Digirolamo, C. M., Stokes, D., Colter, D., Phinney, D. G., Class, R., and Prockop, D. J. (1999), Propagation and senescence of human marro stromal cells in culture: A simple colony-forming assay identifies samples with the greatest potential to propagate and differentiate, *Br. J. Haematol.*, 107(2), 275–281.
26. Colter, D. C., Class, R., DiGirolamo, C. M., Prockop, D. J. (2000), Rapid expansion of recycling stem cells in cultures of plastic-adherent cells from human bone marrow, *Proc. Nat. Acad. Sci. USA*, 97(7), 3213–3218.
27. Colter, D. C., Sekiya, I., and Prockop, D. J. (2001), Identification of a subpopulation of rapidly self-renewing and multipotential adult stem cells in colonies of human marrow stromal cells, *Proc. Nat. Acad. Sci. USA*, 98(14), 7841–7845.
28. Sekiya, I., Larson, B. L., Smith, J. R., Pochampally, R., Cui, J. G., and Prockop, D. J. (2002), Expansion of human adult stem cells from bone marrow stroma: Conditions that maximie the yields of early progenitors and evaluate their quality, *Stem Cells*, 20(6), 530–541.
29. Bartmann, C., Rohde, E., Schallmoser, K., et al. (2007), Two steps to functional MSC for clinical application, *Transfusion*. 2007 Aug; 47(8), 1426–1435.
30. Horwitz, E. M., Gordon, P. L., Koo, W. K., et al. (2002), Isolated allogeneic bone marrow-derived mesenchymal cells engraft and stimulate growth in children with osteogenesis imperfecta: Implications for cells therapy of bone, *Proc. Natl. Acad. Sci. USA*, 99(13), 8932–8937.
31. Pollard, J. W. (1989), Basic cell culture, in Pollard, J. W., Ed., *Animal Cell Culture*, Human, Clifton, NJ, pp. 1–2.
32. WHO (1997), Medicinal and other products and human and animal transmissible spongiform encephalopathies: Memorandum from a WHO meeting, *Bull. World Health Org.*, 75(6), 505–513.
33. European Union (2004), Note for guidance on minimising the risk of transmitting animal spongiform encephalopathy agents via human and veterinary medicinal products (EMEA/410/01 Rev. 2—October 2003) adopted by the Committee for Proprietary Medicinal Products (CPMP) and by the Committee for Veterinary Medicinal Products (CVMP), *Off. J. Europ. Union*, C24/26–C24/19.
34. Schallmoser, K., Rosin, C., Vormittag, R., et al. (2006), Specificities of platelet autoantibodies and platelet activation in lupus anticoagulant patients: A relation to their history of thromboembolic disease, *Lupus*, 15(8), 507–514.
35. Wagner, W., Wein, F., Seckinger, A., et al. (2005), Comparative characteristics of mesenchymal stem cells from human bone marrow, adipose tissue, and umbilical cord blood, *Exp. Hematol.*, 33(11), 1402–1416.
36. Ramalho-Santos, M., Yoon, S., Matsuzaki, Y., Mulligan, R. C., and Melton, D. A. (2002), "Stemness": Transcriptional profiling of embryonic and adult stem cells, *Science*, 298(5593), 597–600.
37. Boquest, A. C., Shahdadfar, A., Fronsdal, K., et al. (2005), Isolation and transcription profiling of purified uncultured human stromal stem cells: Alteration of gene expression after in vitro cells culture, *Mol. Biol. Cell*, 16(3), 1131–1141.
38. Shahdadfar, A., Fronsdal, K., Haug, T., Reinholt, F. P., and Brinchmann, J. E. (2005), In vitro expansion of human mesenchymal stem cells: Choice of serum is a determinant of cell proliferation, differentiation, gene expression, and transcriptome stability, *Stem Cells*, 23(9), 1357–1366.
39. Silva, W. A., Jr., Covas, D. T., Panepucci, R. A., et al. (2003), The profile of gene expression of human marrow mesenchymal stem cells, *Stem Cells*, 21(6), 661–669.
40. Panepucci, R. A., Siufi, J. L., Silva, W. A., Jr., et al. (2004), Comparison of gene expression of umbilical cord vein and bone marrow-derived mesenchymal stem cells, *Stem Cells*, 22(7), 1263–1278.

41. Jeong, J. A., Hong, S. H., Gang, E. J., et al. (2005), Differential gene expression profiling of human umbilical cord blood-derived mesenchymal stem cells by DNA microarray, *Stem Cells*, 23(4), 584–593.
42. Monticone, M., Liu, Y., Tonachini, L., et al. (2004), Gene expression profile of human bone marrow stromal cells determined by restriction fragment differential display analysis, *J. Cell. Biochem*, 92(4), 733–744.
43. Puck, T. T., and Marcus, P. I. (1956), Action of x-rays on mammalian cells. *J. Exp. Med.*, 103(5), 653–666.
44. Tepperman, A. D., Curtis, J. E., and McCulloch, E. A. (1974), Erythropietic colonies in cultures of human marrow, *Blood*, 44(5), 659–669.
45. Pittenger, M. F., Mackay, A. M., Beck, S. C., et al. (1999), Multilneage potential of adult human mesenchymal stem cells, *Science*, 284(5411), 143–147.
46. Pittenger, M. F., and Martin, B. J. (2004), Mesenchymal stem cells and their potential as cardiac therapeutics, *Circ. Res.*, 95(1), 9–20.
47. Delorme, B. C. S., and Charbord, P. (2006), The concept of mesenchymal stem cells, *Regenerative Med.*, 1(4), 497–509.
48. Di Nicola, M., Carlo-Stella, C., Magni, M., et al. (2002), Human bone marrow stromal cells suppress T-lymphocyte proliferation induced by cellular or nonspecific mitogenic stimuli, *Blood*, 99(10), 3838–3843.
49. Bartholomew, A., Sturgeon, C., Siatskas, M., et al. (2002), Mesenchymal stem cells suppress lymphocyte proliferation in vitro and prolong skin graft survival in vivo, *Exp. Hematol.*, 30(1), 42–48.
50. Le Blanc, K., Tammik, L., Sundberg, B., Haynesworth, S. E., and Ringden, O. (2003), Mesenchymal stem cells inhibit and stimulate mixed lymphocyte cultures and mitogenic responses independently of the major histocompatibility complex, *Scand. J. Immunol.*, 57(1), 11–20.
51. Tse, W. T., Pendleton, J. D., Beyer, W. M., Egalka, M. C., and Guinan, E. C. (2003), Suppression of allogeneic T-cell proliferation by human marrow stromal cells: Implications in transplantation, *Transplantation*, 75(3), 389–397.
52. Muller, I., Kordowich, S., Holzwarth, C., et al. (2006), Animal serum-free culture conditions for isolation and expansion of multipotent mesenchymal stromal cells from human BM, *Cytotherapy*, 8(5), 437–444.
53. Reinisch, A., Bartmann, C., Rohde, E. S. K., Bjelic-Radisic, V., Lanzer, G., Linkesch, W., and Strunk, D. (2007), A humanized system to propagate cord blood-derived mesenchymal stem cells for clinical application, *Regen. Med.* 2007 Jul; 2(4), 371–382.
54. Robinson, S. N., Ng, J., Niu, T., et al. (2006), Superior ex vivo cord blood expansion following co-culture with bone marrow-derived mesenchymal stem cells, *Bone Marrow Transplant.*, 37(4), 359–366.
55. Speicher, M. R., and Carter, N. P. (2005), The new cytogenetics: Blurring the boundaries with molecular biology, *Nat. Rev. Genet.*, 6(10), 782–792.
56. Rubio, D., Garcia-Castro, J., Martin, M. C., et al. (2005), Spontaneous human adult stem cell transformation, *Cancer Res.*, 65(8), 3035–3039.
57. Tolar, J., Nauta, A. J., Osborn, M. J., et al. (2007), Sarcoma derived from cultured mesenchymal stem cells, *Stem Cells*, 25(2), 371–379.

# NATIONAL GMP REGULATIONS AND CODES AND INTERNATIONAL GMP GUIDES AND GUIDELINES: CORRESPONDENCES AND DIFFERENCES

Marko Närhi and Katrina Nordström
*Helsinki University of Technology, Helsinki, Finland*

**Contents**

1 Introduction
2 National GMP Regulations and Codes
  2.1 United States
  2.2 Canada
  2.3 European Union
  2.4 East Asian Countries
  2.5 India
  2.6 Australia
  2.7 New Zealand
  2.8 South Africa
3 International GMP Guides and Harmonization
  3.1 World Health Organization
  3.2 Pharmaceutical Inspection Cooperation Scheme
  3.3 International Conference on Harmonization
  3.4 Association of Southeast Asian Nations (ASEAN)
  3.5 Mercado Comun del Sur (MERCOSUR)
4 Correspondences of the U.S. GMP Regulations with GMP Codes and Guidelines
  4.1 General Issues
  4.2 Organization and Personnel
  4.3 Buildings and Facilities
  4.4 Equipment
  4.5 Control of Components and Drug Product Containers and Closures
  4.6 Production and Process Controls

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

4.7 Packaging and Labeling Control
4.8 Holding and Distribution
4.9 Laboratory Controls
4.10 Records and Reports
4.11 Returned and Salvaged Drug Products
References

## 1 INTRODUCTION

The first predecessors of manufacturing and quality requirements, which later evolved into good manufacturing practices (GMPs), were issued in the 1940s in the United States by the Food and Drug Administration (FDA) [1]. In the general meeting of the World Health Organization (WHO) held in 1969, the World Health Assembly issued a recommendation for the introduction of GMPs [2]. Since then, most industrialized countries have passed laws on control procedures essential for the manufacture of drug products. In some countries GMPs are integrated into national legislation as a part of laws or regulations on production, distribution, marketing, and use of drug products (GMP regulations). In other countries, GMPs are separate guidelines outside the national drug legislation (GMP codes). In addition to national GMPs, also some international organizations and trade blocks have issued their own international GMP guidelines to harmonize the requirements for drug production in different countries. However, regardless of their origin, the main purpose of GMPs is to ensure that manufactured drug products have the safety, identity, potency, purity, and quality that they are presented to have [3]. To fulfill this aim, most GMPs usually cover quality management, personnel, premises, equipment, documentation, materials management, production and in-process controls, packaging and labeling of intermediate and finished products, laboratory controls, validation, and change controls [4].

## 2 NATIONAL GMP REGULATIONS AND CODES

### 2.1 United States

In the United States the production of drug products is controlled under the federal Food, Drug and Cosmetic Act, which states that a drug product will be deemed to be adulterated unless the methods used in or the facilities or controls used for its manufacture, processing, packaging, or holding conform to or are operated or administered in conformity with current GMP [5]. The actual GMP regulations are issued as a part of the Code of Federal Regulations and as such they are a federal law. The current set of GMP regulations is based on the 1978 revision [6, 7] of the original GMP regulations, which were first promulgated in 1963. The GMP regulations are updated every year in April [8]; however, no major changes have been implemented since 1978. As an addition to GMP regulations, the FDA also publishes other GMP-related guidance documents covering various issues of drug manufacturing [9]. On the other hand, although these documents reflect current views and

**TABLE 1 Contents of Part 211 of U.S. GMP Regulations [7]**

| Section | Subject |
|---|---|
| Subpart A | General provisions |
| Subpart B | Organization and personnel |
| Subpart C | Buildings and facilities |
| Subpart D | Equipment |
| Subpart E | Control of components and drug product containers and closures |
| Subpart F | Production and process controls |
| Subpart G | Packaging and labeling control |
| Subpart H | Holding and distribution |
| Subpart I | Laboratory controls |
| Subpart J | Records and reports |
| Subpart K | Returned and salvaged drug products |

expectations of the agency, they only provide guidance on principles and practices that are not legal requirements [1]. As a member of the International Conference on Harmonization of Technical Requirements for Registration of Pharmaceutical for Human Use (ICH), the United States has adopted the ICH guidance document Q7, *Good Manufacturing Practice Guide for Active Pharmaceutical Ingredients*, and published it as a guidance for industry document [10].

The U.S. GMP regulations are divided into two parts: 210 [6] and 211 [7]. Part 210, "Current Good Manufacturing Practice in Manufacturing, Processing, Packing or Holding of Drugs—General," provides the framework for the regulations [6], and Part 211, "Current Good Manufacturing Practice for Finished Pharmaceuticals," states the actual requirements. Part 211 is further divided into 11 subparts, which cover the requirements for personnel, premises, equipment, control of materials, production and process controls, packaging and labeling control, holding and distribution, laboratory controls, documentation, and returned and salvaged products [7]. The contents of Part 211 are presented in Table 1.

### 2.2 Canada

The production of drug products (drugs) in Canada is controlled under the Food and Drugs Act, which states that distributors and importers are not allowed to sell a drug product unless it has been manufactured according to the requirements of GMP. The principles of GMP are laid down by Division 2 in Part C of the Food and Drug Regulations, which is a part of the Food and Drugs Act [11]. The Health Products and Food Branch Inspectorate has also issued a guidance document (GMP code), which has been prepared to assist in the interpretation of GMP regulations. The current set of the Canadian GMP code was issued in 2002 and has not been revised since. It has been written with a view to harmonization with GMP standards of other countries and international organizations [WHO, Pharmaceutical Inspection Cooperation Scheme (PIC/S), ICH]. Canadian Healthcare authorities have also published several annexes to the basic GMP code covering topics such as GMP for medical gases, biological drug products, blood products, and production of investigational new drugs. In addition to the GMP code and its annexes, the Canadian

**TABLE 2 Contents of Canadian GMP Code [12]**

| |
|---|
| Introduction |
| Quality management |
| Glossary of terms |
| Regulation |
| Premises |
| Equipment |
| Personnel |
| Sanitation |
| Raw material testing |
| Manufacturing control |
| Quality control department |
| Packaging material testing |
| Finished product testing |
| Records |
| Samples |
| Stability |
| Sterile products |
| Medical gases |
| Annex A: Internationally Harmonized Requirements for Batch Certification |
| Annex B: Application for Alternate Sample Retention |
| Annex C: References |

authorities have also issued several other specific guidelines dealing with issues related to GMP and manufacturing methods [12].

As shown in Table 2 the Canadian GMP code can be divided into four chapters and three annexes. The first three chapters cover general issues such as scope and applicability of the code, definitions of used terms, and issues concerning quality management and GMP in general. GMP regulations and their application are presented in the fourth chapter ("Regulation"), which is divided into 14 subchapters covering the requirements for premises, equipment, personnel, sanitation, testing of components and packaging materials, testing of finished product, production control, quality control department, documentation, reserve samples, stability testing, and manufacture of sterile drug products and medical gases. Each subchapter contains the corresponding regulation according to regulations in Division 2 [11] issued with a rationale and interpretation to assist in their application. The annexes include requirements for batch certification, application form for alternate sample retention site, and references such as hyperlinks to Canadian laws concerning drug products and other GMP-related national and international guidelines [12].

### 2.3 European Union

The production of drug products (medicinal products) in the European Union (EU) is controlled under Directive 2001/83/EC of the European parliament and of the Council, which states that the holder of a manufacturing authorization for medicinal products is obliged to comply with good manufacturing practices as laid down by European Community law [13]. The principles and guidelines of GMP for medicinal products are stated by the Commission directive 2003/94/EC, which provides the

**TABLE 3 Contents of Part I of EU GMP Code Covering Basic Requirements for Manufacture of Drug Products [15]**

| Section | Subject |
|---|---|
| | Introduction |
| Chapter 1 | Quality management |
| Chapter 2 | Personnel |
| Chapter 3 | Premises and equipment |
| Chapter 4 | Documentation |
| Chapter 5 | Production |
| Chapter 6 | Quality control |
| Chapter 7 | Contract manufacture and analysis |
| Chapter 8 | Complaints and product recall |
| Chapter 9 | Self-inspection |
| | Glossary |

legal basis for GMP in the EU [14]. The actual GMP code with detailed written procedures is published in *The Rules Governing Medicinal Products in the European Union*, volume 4. The current set of the EU GMP code was first introduced in 1989 consisting of nine chapters covering the general requirements of GMP and one annex on the manufacture of sterile drug products. Since then the EU GMP code has been revised many times and several new annexes have been issued [15]. In addition to the GMP code, the EU has also published several other guidelines concerning the quality issues of drug production in *The Rules Governing Medicinal Products in the European Union*, volume 3A [16].

As shown in Tables 3–5, the EU GMP code is presented in two parts of basic requirements and 18 annexes. Part I, "Basic Requirements for Medicinal Products," covers GMP principles for the manufacture of drug products. It consists of nine chapters covering the requirements for quality management and control, personnel, premises, equipment, documentation, production, contract services, complaints, product recall, and self-inspection. Part II, "Basic Requirements for Active Substances Used as Starting Materials," covers GMPs for active substances used as starting materials. It is based on the ICH document Q7, *Good Manufacturing Practice Guide for Active Pharmaceutical Ingredients*, and was originally introduced in 2001 as Annex 18 of the EU GMP code. In the restructured revision of the EU GMP code issued in October 2005, Annex 18 was replaced with Part II. It consists of 19 chapters, which cover basic GMP issues related to quality management, personnel, premises, equipment, documentation, materials, production and process controls, packaging and labeling, storage and distribution, laboratory controls, validation, change control, complaints, recalls, contract services, co-operators, active pharmaceutical ingredients (APIs) manufactured by cell culture/fermentation, and APIs used in clinical trials. The annexes give more detailed specific guidance on the manufacture of sterile drug products, biological drug products, radiopharmaceuticals, veterinary drug products, medical gases, herbal drug products, oral liquids, external preparations (creams, ointments), aerosols, investigational new drugs, and blood and blood products. They also cover sampling of materials, computerized systems, use of ionizing radiation, qualification and validation, batch release, parametric release, reference, and retention samples [15].

**TABLE 4 Contents of Part II of EU GMP Code Covering Basic Requirements for Manufacture of Active Substances Used as Starting Materials [15]**

| Section | Subject |
|---|---|
| 1 | Introduction |
| 2 | Quality management |
| 3 | Personnel |
| 4 | Buildings and facilities |
| 5 | Process equipment |
| 6 | Documentation and records |
| 7 | Materials management |
| 8 | Production and in-process controls |
| 9 | Packaging and identification labeling of APIs and intermediates |
| 10 | Storage and distribution |
| 11 | Laboratory controls |
| 12 | Validation |
| 13 | Change control |
| 14 | Rejection and reuse of materials |
| 15 | Complaints and recalls |
| 16 | Contract manufacturers (including laboratories) |
| 17 | Agents, brokers, traders, distributors, repackers, and relabelers |
| 18 | Specific guidance for APIs manufactured by cell culture/fermentation |
| 19 | APIs for use in clinical trials |
| 20 | Glossary |

**TABLE 5 Annexes of EU GMP Code Covering Specific Guidance [15]**

| Section | Subject |
|---|---|
| Annex 1 | Manufacture of sterile medicinal products |
| Annex 2 | Manufacture of biological medicinal products for human use |
| Annex 3 | Manufacture of radiopharmaceuticals |
| Annex 4 | Manufacture of veterinary medicinal products other than immunological veterinary medicinal products |
| Annex 5 | Manufacture of immunological veterinary medicinal products |
| Annex 6 | Manufacture of medicinal gases |
| Annex 7 | Manufacture of herbal medicinal products |
| Annex 8 | Sampling of starting and packaging materials |
| Annex 9 | Manufacture of liquids, creams, and ointments |
| Annex 10 | Manufacture of pressurised metered-dose aerosol preparations for inhalation |
| Annex 11 | Computerized systems |
| Annex 12 | Use of ionizing radiation in manufacture of medicinal products |
| Annex 13 | Manufacture of investigational medicinal products |
| Annex 14 | Manufacture of products derived from human blood or human plasma |
| Annex 15 | Qualification and validation |
| Annex 16 | Certification by a qualified person and batch release |
| Annex 17 | Parametric release |
| Annex 19 | Reference and retention samples |

### 2.4 East Asian Countries

***Japan*** In Japan the production of drug products (drugs) is regulated under the Pharmaceuticals Affairs Law (PAL), which states that any drug manufacturer who plans to manufacture a drug product for sale in Japan must have a Japanese drug manufacturing license and comply with Japanese GMP requirements. The first regulations of Japanese GMP were introduced in 1974 as *The Standards for Manufacturing Control and Quality Control*. In 1979 PAL was partially revised and GMPs became legally binding [2].

PAL is managed and enforced via ministerial ordinances and notices, which are detailed regulations prepared by the Japanese government. The requirements for premises for drug manufacture are given in Ministry of Health, Labor and Welfare (MHLW) Ministerial Ordinance No. 73, 2005 *Regulations for Buildings and Facilities for Pharmacies, etc.* [originally Ministry of Health and Welfare (MHW) Ministerial Ordinance No. 2, 1961] [17], and the requirements for manufacturing and quality controls in MHLW Ministerial Ordinance No. 95, 2003 *Regulations for Manufacturing Control and Quality Control of Drugs* (originally MHW Ministerial Ordinance No. 3, 1994). As a member of the ICH Japan has adopted the ICH guidance document Q7, *Good Manufacturing Practice Guide for Active Pharmaceutical Ingredients*, and published it as Pharmaceutical and Food Safety Bureau (PFSB) Director-General Notification No. 1200, 2001 *Guidelines on GMP for Drug Substances*, which states the requirements for the manufacture of APIs. The requirements concerning imported drug products are given in MHLW Ministerial Ordinance No. 97, 2003 *Regulations for Importing/Retail Management and Quality Control of Drugs and Quasi-Drugs* (originally MHW Ministerial Ordinance No. 62, 1999). The requirements specifying manufacture of investigational products are given in PAB Notification No. 480, 1997 *Products and Standards for the Buildings and Facilities of Manufacturing Plants for Investigational Products* (Investigational Product GMP) [2].

***South Korea*** The production of drug products (drugs) in South Korea is regulated under the Pharmaceutical Affairs Law, which was first enacted in 1953 and has since been revised several times [18]. New drug approval and related activities are regulated in much the same way as in the United States and Japan. Korean GMP, which is often called KGMP, was initiated in 1984 and became mandatory in 1995 [19]. A drug manufacturer who intends to manufacture a drug product for sale in Korea must have approval from the Commissioner of the Korea Food and Drug Administration (KFDA). In order to require the license for manufacturing business the manufacturer has to prove the compliance of facility standards with KGMP [20].

***China*** China regulates the production of drug products (drugs) under the Drug Administration Law of the People's Republic of China, which states that a drug manufacturer has to conduct drug manufacture according to the GMP for pharmaceutical products formulated by the Drug Regulatory Department under the State Council on the basis of the Drug Administration Law [21]. In June 2004 GMP became mandatory in China and the State Drug Administration announced that local drug manufacturing establishments lacking approved GMP certification would not be allowed to continue the production of pharmaceuticals [22].

**TABLE 6 Contents of Part I of Indian GMP Regulations Covering Good Manufacturing Practices for Premises and Materials [24]**

| Section | Subject |
|---|---|
| 1 | General requirements |
| 2 | Warehousing area |
| 3 | Production area |
| 4 | Ancillary areas |
| 5 | Quality control area |
| 6 | Personnel |
| 7 | Health, clothing, and sanitation of workers |
| 8 | Manufacturing operations and control |
| 9 | Sanitation in the manufacturing premises |
| 10 | Raw materials |
| 11 | Equipment |
| 12 | Documentation and records |
| 13 | Labels and other printed materials |
| 14 | Quality assurance |
| 15 | Self-inspection and quality audit |
| 16 | Quality control system |
| 17 | Specification |
| 18 | Master formula records |
| 19 | Packing records |
| 20 | Batch packaging records |
| 21 | Batch processing records |
| 22 | Standard operating procedures (SOPs) and records |
| 23 | Reference samples |
| 24 | Reprocessing and recoveries |
| 25 | Distribution records |
| 26 | Validation and process validation |
| 27 | Product recalls |
| 28 | Complaints and adverse reactions |
| 29 | Site master file |

## 2.5 India

The production of drug products (drugs) in India is controlled under the Drugs and Cosmetics Rules (1945, last amended in 2005), which states that the holder of the license to manufacture drugs has to comply with the requirements of GMP as laid down in Schedule M [23]. Schedule M is a part of the Drugs and Cosmetics Rules and embodies the Indian GMP regulations [24], which are based on the 1982 version of WHO GMP guidelines [25].

As shown in Tables 6–8 the Indian GMP regulations consists of eight parts: I, IA, IB, IC, ID, IE, IF, and II. Part I covers the general requirements of GMP. It is divided into 29 chapters, which deal with the requirements for personnel, premises, equipment, sanitation, production and process controls, materials, documentation, quality management, validation, reserve samples, recalls, complaints, and self-inspection. Parts IA to IE cover specific requirements for the manufacture of different dosage forms regarding premises, equipment, and methods. Part IA deals with the require-

**TABLE 7 Contents of Parts IA, IB, IC, ID, IE, and IF of Indian GMP Regulations Covering Specific Guidance [24]**

| Section | Subject |
|---|---|
| Part IA | Specific requirements for manufacture of sterile products, parenteral preparations (small-volume injectables and large-volume parenterals) and sterile ophthalmic preparations |
| Part IB | Specific requirements for manufacture of oral solid dosage forms (tablets and capsules) |
| Part IC | Specific requirements for manufacture of oral liquids (syrups, elixirs, emulsions, and suspensions) |
| Part ID | Specific requirements for manufacture of topical products, i.e., external preparations (creams, ointments, pastes, emulsions, lotions, solutions, dusting powders, and identical products) |
| Part IE | Specific requirements for manufacture of metered-dose inhalers (MDIs) |
| Part IF | Specific requirements of premises, plant, and materials for manufacture of active pharmaceutical ingredients (bulk drugs) |

**TABLE 8 Contents of Part II of Indian GMP Regulations Covering Requirements of Plant and Equipment [24]**

| Section | Subject |
|---|---|
| 1 | External preparations |
| 2 | Oral liquid preparations |
| 3 | Tablets |
| 4 | Powders |
| 5 | Capsules |
| 6 | Surgical dressing |
| 7 | Ophthalmic preparations |
| 8 | Pessaries and suppositories |
| 9 | Inhalers and vitrallae |
| 10 | Repacking of drugs and pharmaceutical chemicals |
| 11 | Parenteral preparations |

ments for the manufacture of parenteral preparations; Part IB with the requirements for the manufacture of oral solid dosage forms such as tablets and capsules; Part IC with the requirements for the manufacture of oral liquids such as syrups, elixirs, emulsions, and suspensions; Part ID with the requirements for the manufacture of external preparations such as creams, ointments, pastes, emulsions, and lotions; and Part 1E with the requirements for the manufacture of inhalers. Part 1F covers specific requirements for the manufacture of APIs regarding buildings and facilities, utilities, equipment, controls, and containers. Part II of the Indian GMP regulations consist of detailed recommendations for the process equipment to be used in the manufacture of different dosage forms and requirements for the partition of the production area [24].

## 2.6 Australia

In Australia the production of drug products (medicinal products) is controlled under the Therapeutics Goods Act, which provides the Minister for Health and Aged Care the right to determine written principles including codes of GMP to be observed in the production of drug products for use in humans [26]. The Therapeutic Goods (Manufacturing Principles) Determination No. 2 of 2002 given by the minister states that drug products must be manufactured in compliance with the *Australian Code of Good Manufacturing Practice for Medicinal Products*, dated August 16, 2002 [27]. The current set of the Australian GMP code is based entirely on the PIC/S GMP guide version PH 1/97 (Rev. 3) published in 2002 with some minor modifications [28].

As shown in Table 9, the Australian GMP code consists of 9 chapters and 13 annexes. The chapters present the general requirements of GMP for the manufacture of drug products, the requirements for quality management and control, personnel, premises, equipment, documentation, production, contract services, complaints, product recall, and self-inspection. The annexes give specific guidance on the manufacture of sterile drug products, biological drug products, radiopharmaceuticals, medical gases, herbal drug products, oral liquids, external preparations (creams, ointments), aerosols, investigational new drugs, blood, and blood products. They also

**TABLE 9 Contents of Australian GMP Code [28]**

| Section | Subject |
|---|---|
| | Introduction |
| | Interpretation |
| Chapter 1 | Quality management |
| Chapter 2 | Personnel |
| Chapter 3 | Premises and equipment |
| Chapter 4 | Documentation |
| Chapter 5 | Production |
| Chapter 6 | Quality control |
| Chapter 7 | Contract manufacture and analysis |
| Chapter 8 | Complaints and product recall |
| Chapter 9 | Self-inspection |
| Annex 1 | Manufacture of sterile medicinal products |
| Annex 2 | Manufacture of biological medicinal products for human use |
| Annex 3 | Manufacture of radiopharmaceuticals |
| Annex 6 | Manufacture of medicinal gases |
| Annex 7 | Manufacture of herbal medicinal products |
| Annex 8 | Sampling of starting and packaging materials |
| Annex 9 | Manufacture of liquids, creams, and ointments |
| Annex 10 | Manufacture of pressurised metered-dose aerosol preparations for inhalation |
| Annex 11 | Computerized systems |
| Annex 12 | Use of ionizing radiation in the manufacture of medicinal products |
| Annex 13 | Manufacture of investigational medicinal products |
| Annex 15 | Qualification and validation |
| Annex 17 | Parametric release |
| | Glossary |

cover sampling of materials, computerized systems, use of ionizing radiation, qualification, and validation and parametric release [28].

Australia has not adopted Annexes 4, 5, 14, 16, and 18 of the PIC/S GMP guide. Annexes 4 and 5 cover the manufacture of veterinary drug products. Annex 14 covers the manufacture of products derived from human blood or human plasma, which is excluded from the Australian GMP code. Annex 16 is specific to the EU GMP code and Annex 18 is the ICH GMP guide for the manufacture of APIs, which Australia has adopted separately as a manufacturing principle [28].

### 2.7 New Zealand

The production of drug products (medicines) in New Zealand is controlled under the Medicines Act 1981, which states that a drug manufacturer is not allowed to manufacture drug products without a manufacturing license issued by the licensing authority. In order to obtain a manufacturing license the applicant must satisfy the licensing authority with respect to the proposed manufacturing premises and equipment, which must be suitable and adequate for the manufacture of drugs. Moreover, the applicant must show that adequate arrangements have been made or are to be made for the making, maintaining, and safekeeping of adequate records with reference to the drug products that are to be manufactured [29]. The authorities (Medsafe) require that any drug manufacturer who plans to manufacture drug products for sale in New Zealand must deliver evidence of GMP compliance for the manufacturing site. Copies of appropriate certificates, manufacturing licenses, or reports issued by a regulatory authority whose competence is recognized by Medsafe are accepted as proof of GMP compliance [30].

As shown in Table 10 New Zealand's own GMP code consists of five parts. The first part covers the manufacture of drug products and the second part the manufacture of blood products. Part 3 covers compounding and dispensing, including compounding of sterile drug products. Part 4 deals with wholesaling and Part 5 with product recalls. Parts 4 and 5 are combined in one document [31].

### 2.8 South Africa

South Africa controls the production of drug products (medicines) under the Medicines and Related Substances Control Act (Act 101 of 1965), which states that the Medicines Control Council may issue to a drug manufacturer a license to manufacture a drug product upon such conditions as to the application of such acceptable

**TABLE 10 Contents of New Zealand's GMP Code [31]**

| Section | Subject |
|---|---|
| Part 1 | Manufacture of pharmaceutical products |
| Part 2 | Manufacture of blood and blood products |
| Part 3 | Compounding and dispensing |
| Part 4 | Wholesaling of medicines and medical devices |
| Part 5 | Uniform recall procedure for medicines and medical devices |

**TABLE 11 Contents of South African GMP Code [34]**

| Section | Subject |
|---|---|
| | Introduction |
| Chapter 1 | Quality management |
| Chapter 2 | Personnel |
| Chapter 3 | Premises and equipment |
| Chapter 4 | Documentation |
| Chapter 5 | Production |
| Chapter 6 | Quality control |
| Chapter 7 | Contract manufacture and analysis |
| Chapter 8 | Complaints and product recall |
| Chapter 9 | Self-inspection |
| Annex 1 | Manufacture of sterile medicinal products |
| Annex 2 | Manufacture of biological medicinal products for human use |
| Annex 3 | Manufacture of radiopharmaceuticals |
| Annex 4 | Manufacture of veterinary medicinal products other than immunologicals |
| Annex 5 | Manufacture of immunological veterinary medical products |
| Annex 6 | Manufacture of medicinal gases |
| Annex 7 | Manufacture of herbal medicinal products |
| Annex 8 | Sampling of starting and packaging materials |
| Annex 9 | Manufacture of liquids, creams, and ointments |
| Annex 10 | Manufacture of pressurized metered-dose aerosol preparations for inhalation |
| Annex 11 | Computerized systems |
| Annex 12 | Use of ionizing radiation in the manufacture of medicinal products |
| Annex 13 | Manufacture of investigational medicinal products |
| Annex 14 | Manufacture of products derived from human blood or human plasma |
| Annex 15 | Qualification and validation |
| Annex 16 | Organisation and personnel |
| Annex 17 | Parametric release |
| | Glossary |

quality assurance principles and GMPs as the council may determine [32]. As a part of the license application the manufacturer must provide acceptable documentary proof of the ability to comply with GMP as determined by the council [33]. The current set of South African GMP code determined by the council is entirely based on the PIC/S GMP guide version PE 009-2, published in 2004 with some minor modifications [34].

As shown in Table 11 the South African GMP code consists of 9 chapters and 17 annexes. The chapters present the general requirements of GMP for the production of drug products covering the requirements for quality management and control, personnel, premises, equipment, documentation, production, contract services, complaints, product recall, and self-inspection. The annexes give specific guidance on the manufacture of sterile drug products, biological drug products, radiopharmaceuticals, veterinary drug products, medical gases, herbal drug products, oral liquids, external preparations (creams, ointments), aerosols, investigational new drugs, and blood and blood products. They also cover sampling of materials, computerized systems, use of ionizing radiation, qualification and validation, organization, and

personnel and parametric release. The original Annex 16, which is specific to the EU GMP code, has been replaced in South African GMP code with an annex covering organization and personnel. Nor has South Africa adopted Annex 18, which covers the ICH GMP guide for the manufacture of APIs, as it has been adopted separately as a manufacturing principle [34].

## 3 INTERNATIONAL GMP GUIDES AND HARMONIZATION

### 3.1 World Health Organization

The WHO was established in 1948 as a specialized agency of the United Nations (UN). Its purpose is to serve as the directing and coordinating authority for international health matters and public health. One of the main functions of the WHO is to provide objective and reliable information and advice in the field of human health, a task that it partly fulfills through WHO publications [35]. The first WHO draft text on GMP was prepared in 1967 and a revised version was published in 1968 as an annex of the twenty-second report of the WHO expert committee on specifications for pharmaceutical preparations. Over the years the WHO has issued several versions of its GMP guidelines as well as other guidelines related to the GMP and quality issues of the production of therapeutic products. The latest version of the WHO GMP guideline was published in 2003 as an annex of the WHO Technical Report 908 [36].

As shown in Table 12 the WHO GMP guideline is divided into five parts: introduction, general considerations, glossary, quality management in the drug industry, and references. The actual GMP guidelines are presented in the fourth part, which consists of 17 chapters covering the requirements for quality assurance and control, personnel, premises, equipment, sanitation, materials, validation, documentation, production, contract services, complaints, recalls, and self-inspection [36]. In addition to this guideline laying down the main principles of GMP, the WHO has also published several other guidelines covering specific requirements for components, quality of water for pharmaceutical use, APIs, excipients, sterile drug products, biological drug products, investigational drug products, herbal drug products, and radiopharmaceuticals (Table 13).

### 3.2 Pharmaceutical Inspection Cooperation Scheme

The Pharmaceutical Inspection Convention (PIC), which is the predecessor of PIC/S, was founded in 1970 by the European Free Trade Area (EFTA). The initial members comprised of the 10 EFTA member countries at that time. From the beginning one of the main goals has been the harmonization of GMP requirements as well as the promotion of mutual recognition of inspections and uniformity of inspection systems by training the inspectors, improving the exchange of information, and mutual confidence [46]. Originally PIC was a formal treaty between member countries and as such it also had a legal status. When countries outside Europe were seeking to join PIC, it became evident that, according to European law, individual EU countries that were members of PIC were not permitted to sign agreements with countries outside Europe. Only the European Commission, which itself was

**TABLE 12 Contents of WHO GMP Guideline Covering General Requirements of GMP for Manufacture of Drug Products [36]**

| | |
|---|---|
| Introduction | |
| General considerations | |
| Glossary | |
| Quality management in the drug industry: philosophy and essential elements | |
| Section | Subject |
| 1 | Quality assurance |
| 2 | Good manufacturing practices for pharmaceutical products (GMP) |
| 3 | Sanitation and hygiene |
| 4 | Qualification and validation |
| 5 | Complaints |
| 6 | Product recalls |
| 7 | Contract production and analysis |
| 8 | Self-inspection and quality audits |
| 9 | Personnel |
| 10 | Training |
| 11 | Personal hygiene |
| 12 | Premises |
| 13 | Equipment |
| 14 | Materials |
| 15 | Documentation |
| 16 | Good practices in production |
| 17 | Good practices in quality control |
| References | |

**TABLE 13 GMP-Related WHO Documents Covering Specific Guidance**

| Document | Subject |
|---|---|
| TRS 929, Annex 2 [37] | Requirement for the sampling of starting materials |
| TRS 823, Annex 1 [38] | Active pharmaceutical ingredients (bulk drug substances) |
| TRS 885, Annex 5 [39] | Pharmaceutical excipients |
| TRS 902, Annex 6 [40] | Sterile pharmaceutical products |
| TRS 834, Annex 3 [41] | Biological products |
| TRS 863, Annex 7 [42] | Investigational pharmaceutical products for clinical trials in humans |
| TRS 863, Annex 8 [43] | Herbal medicinal products |
| TRS 908, Annex 3 [44] | Radiopharmaceutical products |
| TRS 929, Annex 3 [45] | Water for pharmaceutical use |

not a member of PIC, was permitted to sign agreements. Consequently, a less formal and more flexible PIC/S was developed to continue the work of PIC. The PIC/S, which became operational in November 1995, is an informal arrangement without legal status between regulatory authorities instead of countries. The PIC and the PIC Scheme, operating together as PIC/S, provide an active and constructive cooperation in the field of GMP [47].

The current members of PIC/S are Australia, Austria, Belgium, Canada, Czech Republic, Denmark, Finland, France, Germany, Greece, Hungary, Iceland, Ireland,

Italy, Latvia, Liechtenstein, Malaysia, Netherlands, Norway, Poland, Portugal, Romania, Singapore, Slovak Republic, Spain, Sweden, Switzerland, and the United Kingdom. In addition, Estonia, the European Agency for the Evaluation of Medicinal Products (EMEA), UNICEF, and the WHO participate in PIC/S activities as observers [48]. Also many other regulatory authorities have shown interest in joining PIC/S, in particular Argentina, Brazil, Cyprus, Indonesia, Israel, Philippines, Slovenia, Thailand, the United States, Bulgaria, Estonia, Lithuania, Oman, Russia, South Africa, and the Ukraine [49].

To become a PIC/S member, a joining regulatory authority is required to go through a detailed assessment to prove that the authority has the arrangements and competence necessary to apply an inspection system equivalent to inspection systems of existing PIC/S members. To ensure that both new applicants and older members fulfill the same requirements, also existing members are reassessed on a regular basis. One of the main functions of PIC/S is to develop GMP guidance documents, which it carries out in close cooperation with the EU and relevant agencies thereof. Under this cooperation both parties have been able to adopt each others' documents, thus minimizing the duplication of effort in development of GMP-related documents. Among other highly informative guides on various aspects of GMP and quality issues [49], PIC/S has also published its own GMP guide (*Guide to Good Manufacturing Practice for Medicinal Products*), which is harmonized with the EU GMP code [50].

The latest revision of the PIC/S GMP guide (version PE 009-3) was issued in January 2006. As shown in Table 14, it consists of 9 chapters and 16 annexes. Chapters present the general requirements of GMP for the production of drug products covering the requirements for quality management and control, personnel, premises, equipment, documentation, production, contract services, complaints, product recall, and self-inspection. The annexes give specific guidance on the manufacture of sterile drug products, biological drug products, radiopharmaceuticals, veterinary drug products, medical gases, herbal drug products, oral liquids, external preparations (creams, ointments), aerosols, investigational new drugs, and blood and blood products. In addition, there are annexes covering the sampling of materials, computerized systems, use of ionizing radiation, qualification and validation, and parametric release [50].

Although the PIC/S GMP guide is harmonized with the EU GMP code and their contents are similar, there are some minor differences between them. Instead of the term *qualified person*, the PIC/S GMP guide uses the term *authorized person*. Furthermore, all references to EU directives have been deleted from the PIC/S GMP guide. Moreover, PIC/S has not adopted Annexes 16 and 18 of the EU GMP code. Annex 16 is specific to the EU GMP code covering the status of a qualified person in batch release and Annex 18 is the ICH GMP guide for the manufacture of APIs, which the PIC/S Committee has adopted as a stand-alone document (PE 007) [50].

### 3.3 International Conference on Harmonization

The ICH was established in 1990. Its main aim is to improve the efficiency of the drug development process and the registration of new drug products in its member countries through harmonization of national guidelines. This is a joint initiative

**TABLE 14 Contents of PIC/S GMP Guide [50]**

| Section | Subject |
|---|---|
| | Introduction |
| Chapter 1 | Quality management |
| Chapter 2 | Personnel |
| Chapter 3 | Premises and equipment |
| Chapter 4 | Documentation |
| Chapter 5 | Production |
| Chapter 6 | Quality control |
| Chapter 7 | Contract manufacture and analysis |
| Chapter 8 | Complaints and product recall |
| Chapter 9 | Self-inspection |
| Annex 1 | Manufacture of sterile medicinal products |
| Annex 2 | Manufacture of biological medicinal products for human use |
| Annex 3 | Manufacture of radiopharmaceuticals |
| Annex 4 | Manufacture of veterinary medicinal products other than immunologicals |
| Annex 5 | Manufacture of immunological veterinary medical products |
| Annex 6 | Manufacture of medicinal gases |
| Annex 7 | Manufacture of herbal medicinal products |
| Annex 8 | Sampling of starting and packaging materials |
| Annex 9 | Manufacture of liquids, creams, and ointments |
| Annex 10 | Manufacture of pressurised metered-dose aerosol preparations for inhalation |
| Annex 11 | Computerized systems |
| Annex 12 | Use of ionizing radiation in manufacture of medicinal products |
| Annex 13 | Manufacture of investigational medicinal products |
| Annex 14 | Manufacture of products derived from human blood or human plasma |
| Annex 15 | Qualification and validation |
| Annex 17 | Parametric release |
| | Glossary |

involving both regulators and industry as equal partners. The founders and current members of ICH, which represent the regulatory bodies and the research-based industry in the member countries, are the EU, European Federation of Pharmaceutical Industries and Associations (EFPIA), MHLW, Japan Pharmaceutical Manufacturers Association (JPMA), FDA, and Pharmaceutical Research and Manufactures of America (PhRMA). In addition to the actual member countries there are also observers who act as a link between ICH and non-ICH countries and regions. Current observers are the WHO, EFTA, Swissmedic (representing Switzerland), and Health Canada (representing Canada) [51].

Among other guidelines, ICH has also published a guide on GMP for APIs (Q7: *Good Manufacturing Practice Guide for Active Pharmaceutical Ingredients*). It is intended to provide guidance regarding GMP for the manufacture of APIs and to help ensure that APIs meet the quality and purity requirements that they are presented to possess. This covers APIs that are manufactured by chemical synthesis, extraction, cell culture/fermentation, recovery from natural sources, or any combination of these processes. Excluded are vaccines, medical gases, bulk-packaged drug

**TABLE 15 Contents of ICH GMP Guideline for API Production [52]**

| Section | Subject |
|---|---|
| 1 | Introduction |
| 2 | Quality management |
| 3 | Personnel |
| 4 | Buildings and facilities |
| 5 | Process equipment |
| 6 | Documentation and records |
| 7 | Materials management |
| 8 | Production and in-process controls |
| 9 | Packaging and identification labeling of APIs and intermediates |
| 10 | Storage and distribution |
| 11 | Laboratory controls |
| 12 | Validation |
| 13 | Change control |
| 14 | Rejection and reuse of materials |
| 15 | Complaints and recalls |
| 16 | Contract manufacturers (including laboratories) |
| 17 | Agents, brokers, traders, distributors, repackers, and relabelers |
| 18 | Specific guidance for APIs manufactured by cell culture/fermentation |
| 19 | APIs for use in clinical trials |
| 20 | Glossary |

products, radiopharmaceuticals, whole cells, whole blood and plasma, blood and plasma derivatives, and gene therapy APIs. However, APIs that are produced using blood or plasma as raw materials are included [52]. All ICH member countries have adopted this guideline: the EU in November 2000, Japan in November 2001, and the United States in September 2001 [53]. In addition, it has also been adopted by several other non-ICH countries such as Australia [28] and South Africa [34].

The basic structure of the ICH GMP guideline for API production is shown in Table 15. It consists of 19 chapters, which cover the requirements for quality management, personnel, premises, equipment, documentation, materials, production and process controls, packaging and labeling, storage and distribution, laboratory controls, validation, change control, complaints, recalls, contract services, cooperators, APIs manufactured by cell culture/fermentation, and APIs used in clinical trials [52].

### 3.4 Association of Southeast Asian Nations (ASEAN)

ASEAN was established in 1967 by Indonesia, Malaysia, Philippines, Singapore, and Thailand. Current members include also Brunei and Darussalam (joined in 1984), Vietnam (joined in 1995), Laos and Myanmar (joined in 1997), and Cambodia (joined in 1999). The aims and purposes of ASEAN involve cooperation in the economic, social, cultural, technical, educational, and other fields [54]. Among other cooperation schemes the ASEAN countries have also developed their own GMP guidelines, which were issued in 1984 [55].

### 3.5 Mercado Comun del Sur (MERCOSUR)

MERCOSUR was established in 1991 by Argentina, Brazil, Paraguay, and Uruguay to develop a common market between its member countries. Current members include also Bolivia and Chile (joined in 1996). One of the original aims was to harmonize the pharmaceutical legislation of the member countries. As a part of these harmonization activities MERCOSUR has developed its own GMP guidelines, which are based on WHO recommendations. In addition to the GMP guideline, MERCOSUR has also issued other GMP-related guides covering inspections, requirements for facilities, and quality control [56].

## 4 CORRESPONDENCES OF THE U.S. GMP REGULATIONS WITH GMP CODES AND GUIDELINES

The following sections deal with the correspondences and differences between the U.S. GMP regulations and the Canadian and EU GMP codes and the WHO GMP guideline. As the EU GMP code is harmonized with the PIC/S GMP guide, the correspondences between the EU GMP code and the U.S. GMP regulations cover also the correspondences between the U.S. GMP regulations and the PIC/S GMP guide as well as all other national GMPs that are based on the PIC/S GMP guide. Differences between the EU GMP code and the PIC/S GMP guide have been presented in Section 3.2.

### 4.1 General Issues

In the U.S. GMP regulations general issues related to the use and applicability of GMP regulations are presented in Part 210 [6], which consists of regulations 210.1, 210.2, and 210.3 and in Subpart A of Part 211 [7], which consists of regulations 211.1 and 211.3. Contents of Part 210 and Subpart A of Part 211 are presented in Table 16. Regulation 210.1 defines the status, 210.2 deals with the applicability, and 211.1 states the scope of the regulations. Definitions of terms used in the regulations are provided in regulation 210.3 and in regulation 211.3, which states that the definitions provided in regulation 210.3 apply also in Part 211.

***Correspondences in Canadian GMP Code*** In the Canadian legislation general issues related to the use and applicability of the GMP regulations and code are covered in the introduction of the GMP code [12] and in Divisions 1A [57] and 2

**TABLE 16 Contents of Part 210 and Subpart A of Part 211 of US GMP Regulations Covering General Issues Related to Use and Applicability of Regulations [6, 7]**

| Section | Subject |
|---|---|
| CFR 210.1 | Status of current good manufacturing practice regulations |
| CFR 210.2 | Applicability of current good manufacturing practice regulations |
| CFR 210.3 | Definitions |
| CFR 211.1 | Scope |
| CFR 211.3 | Definitions |

of the Part C of the Food and Drug Regulations [11]. Definitions for the GMP regulations are covered in regulation C.01A.001 of Division 1A [57] and in regulation C.02.002 of Division 2 [11]. Definitions for the GMP code are covered in the glossary of terms of the code [12].

***Correspondences in EU GMP Code*** In the EU legislation general issues related to the use and applicability of the GMP regulations and the code are covered in the Commission Directive 2003/94/EC [14] and in the introduction of the GMP code [15]. Definitions for the directive are covered in Article 2 of the directive [14] and definitions for the code in the glossary of the GMP code [15].

***Correspondences in WHO GMP Guideline*** In the WHO GMP guideline [36] general issues related to the use and applicability of the GMP guide are covered in section "General Considerations." Definitions for the GMP guide are covered in the glossary of the guideline.

### 4.2 Organization and Personnel

For GMP regulations in the United States issues related to organization and personnel are covered in Subpart B [7], which consists of regulations 211.22, 211.25, 211.28, and 211.34. The contents of Subpart B is presented in Table 17. Regulation 211.22 states the responsibilities and authorities of the quality control unit, including requirements for the resources. Regulation 211.25 deals with personnel qualifications covering the requirements for their education and experience and it also states the requirements for the training of the personnel. Regulation 211.28 states the responsibilities of personnel covering the requirements for the clothing and other protective apparel, personal sanitation and health habits, as well as personal health conditions. Furthermore, it states the requirements for the authorization for limited access. Regulation 211.34 deals with consultants and lays down the requirements for their education, training, and experience, including the requirements for documentation.

***Correspondences in Canadian GMP Code*** In the Canadian GMP code [12] issues related to organization and personnel are mainly covered in the interpretation of regulation C.02.006 (Personnel) and partly in the interpretations of regulations C.02.004 (Premises), C.02.008 (Sanitation), C.02.011 (Manufacturing Control), C.02.013 (Quality Control Department), C.02.015 (Quality Control Department), and C.02.024 (Records). Correspondences to regulation 211.22 are covered in

**TABLE 17 Contents of Subpart B of Part 211 of U.S. GMP Regulations Covering Organization and Personnel [7]**

| Section | Subject |
|---|---|
| CFR 211.22 | Responsibilities of quality control unit |
| CFR 211.25 | Personnel qualifications |
| CFR 211.28 | Personnel responsibilities |
| CFR 211.34 | Consultants |

Sections 1–5 of the interpretation of regulation C.02.015 and in Section 2 of the interpretation of regulation C.02.013. Sections 1–5 of the interpretation of regulation C.02.015 state the responsibilities of the quality control unit (quality control department), and Section 2 of the interpretation of regulation C.02.013 covers the requirements for resources. Correspondences to regulation 211.25 stating the requirements for the education, training, and experience of the personnel are covered in Sections 1–5 of the interpretation of regulation C.02.006. Correspondences to regulation 211.28 are covered in Section 6.3 of the interpretation of regulation C.02.004, Sections 1–2 of the interpretation of regulation C.02.008, Section 8 of the interpretation of regulation C.02.011, and Section 4 of the interpretation of regulation C.02.013. Section 1 of the interpretation of regulation C.02.008 states the health requirements and Section 2 the requirements for clothing, other protective apparel, and personal hygiene. Section 6.3 of the interpretation of regulation C.02.004, Section 8 of the interpretation of regulation C.02.011, and Section 4 of the interpretation of regulation C.02.013 cover the requirements regarding limited access. Correspondences to regulation 211.34 are covered in Section 6 of the interpretation of regulation C.02.006 and in Subsection 1.3.2 of the interpretation of regulation C.02.024. Section 6 of the interpretation of regulation C.02.006 states the requirements for the education, training, and experience of consultants and contractors and Subsection 1.2.3 of the interpretation of regulation C.02.024 the requirements for documentation.

***Correspondences in EU GMP Code*** In the EU GMP code [15] issues related to organization and personnel are mainly covered in Chapter 2 (Personnel) and partly in Chapters 3 (Premises and Equipment), 5 (Production), and 6 (Quality Control). Correspondences to regulation 211.22 are covered in Subchapters 2.6, 2.7, 6.1, and 6.2. Subchapters 2.6 and 2.7 deal with the responsibilities of the head of the quality control unit (quality control department) and 6.2 with the responsibilities of the quality control unit as a whole. Requirements for resources are covered in Subchapter 6.1. Correspondences to regulation 211.25 are covered in Subchapters 2.1, 2.4, and 2.8–2.12. Subchapters 2.1 and 2.4 deal with the requirements for personnel and Subchapters 2.8–2.12 with the requirements for their training. Correspondences to regulation 211.28 are covered in Subchapters 2.15, 2.16, 3.5, 3.21, 5.16, and 6.4. Subchapter 2.15 deals with requirements for personal health conditions and 2.16 with requirements for clothing and protection. Access limitations are covered in Subchapters 3.5, 3.21, 5.16, and 6.4. In the EU GMP code there is no correspondence to regulation 211.34, which covers the requirements for the use of consultants. However, Chapter 7 of the code deals with the requirements for the contract services in general.

***Correspondences in WHO GMP Guideline*** In the WHO GMP guideline [36] issues related to organization and personnel are mainly covered in Chapter 9 (Personnel) and partly in Chapters 10 (Training), 11 (Personal Hygiene), 16 (Good Practices in Production), and 17 (Good Practices in Quality Control). Correspondences to regulation 211.22 are covered in Subchapters 9.8, 9.10, 17.3, and 17.4. Subchapters 9.8 and 9.10 state the responsibilities of the head of the quality control unit and Subchapter 17.4 the responsibilities of the quality control unit as a whole. Subchapter 17.3 covers the requirements for resources. Correspondences to regula-

tion 211.25 are covered in Subchapters 9.2, 9.4, 9.7, and 10.1–10.4. Subchapters 9.2, 9.4, and 9.7 state the requirements for the personnel covering their education and experience and Subchapters 10.1–10.4 the requirements for the training. Correspondences to regulation 211.28 are covered in Subchapters 11.1–11.8, 9.5, and 16.7. Subchapters 11.1–11.5 state the requirements for the health conditions and personal hygiene and Subchapters 11.6–11.8 the requirements for the clothing and other protective apparel. Subchapters 9.5 and 16.7 cover the requirements for the limited access. Correspondences to regulation 211.34 are covered in Subchapter 10.6, which covers the requirements for the use of consultants.

### 4.3 Buildings and Facilities

In the United States GMP regulations on issues related to buildings and facilities are covered in Subpart C [7], which consists of regulations 211.42, 211.44, 211.46, 211.48, 211.50, 211.52, 211.56, and 211.58. Contents of Subpart C are presented in Table 18. Regulation 211.42 deals with design and construction features covering the requirements for the size, construction, and location of buildings used in the production. Furthermore, it states the requirements for the placement of equipment as well as the flow of materials and products and specifies operations, which have to be performed in separate or defined areas to prevent contamination or mix-ups. It also covers the special requirements for the facilities used in aseptic processing and facilities used in the production of penicillin. Regulation 211.44 states the requirements for lighting and 211.46 for ventilation, including the requirements for controls and air-handling systems. Furthermore, it states the special requirements for ventilation in the production of penicillin. Regulation 211.48 deals with plumbing covering requirements for the plumbing system, drains, and the quality of potable water. Regulation 211.50 deals with sewage, trash, and other refuse stating the requirements for their disposal. Regulation 211.52 covers the requirements for washing and toilet facilities. Regulation 211.56 deals with sanitation stating the requirements for the conditions to be maintained in the manufacturing facilities. It also states the requirements for handling of trash and organic waste. Furthermore, it states the requirements for the written procedures for sanitation operations and use of biocides, fumigating, cleaning, and sanitizing agents. It also states the

**TABLE 18 Contents of Subpart C of Part 211 of U.S. GMP Regulations Covering Buildings and Facilities [7]**

| Section | Subject |
|---|---|
| CFR 211.42 | Design and construction features |
| CFR 211.44 | Lighting |
| CFR 211.46 | Ventilation, air filtration, air heating and cooling |
| CFR 211.48 | Plumbing |
| CFR 211.50 | Sewage and refuse |
| CFR 211.52 | Washing and toilet facilities |
| CFR 211.56 | Sanitation |
| CFR 211.58 | Maintenance |

requirements for the use of biocides and the scope of sanitation procedures. Regulation 211.58 states the requirements for the maintenance of the buildings used in the production.

***Correspondences in Canadian GMP Code*** In the Canadian GMP code [12] issues related to buildings and facilities are mainly covered in the interpretation of regulation C.02.004 (Premises) and partly in the interpretations of regulations C.02.005 (Equipment), C.02.007 (Sanitation), C.02.009 (Raw Material Testing), C.02.011 (Manufacturing Control), and C.02.029 (Sterile Products). Correspondences to regulation 211.42 are covered in Sections 1, 2, 2.3, 6, 6.2, and 6.4 of the interpretation of regulation C.02.004, in Section 15 of the interpretation of regulation C.02.011, and in section "Premises" of the interpretation of regulation C.02.029. Sections 1 and 2 of the interpretation of regulation C.02.004 cover the requirements for the size, construction, and location of buildings used in the production. Section 6.2 of the interpretation of regulation C.02.004 and Section 15 of the interpretation of regulation C.02.011 state the requirements for the placement of equipment. Requirements for the flow of materials and products are covered in Section 6 of the interpretation of regulation C.02.004 and operations, which have to be performed in separate or defined areas in Sections 2.3 and 6.4 of the interpretation of regulation C.02.004. The requirements for the facilities used in aseptic processing are covered in section "Premises" of the interpretation of regulation C.02.029 and the requirements for the facilities used in the production of penicillin in Section 11.1 of the interpretation of regulation C.02.004. Correspondences to regulation 211.44 stating the requirements for the lighting are covered in Section 6.5 of the interpretation of regulation C.02.004. Correspondences to regulation 211.46 are covered in Sections 3.6 and 4 of the interpretation of regulation C.02.004. Section 3.6 of the interpretation of regulation C.02.004 states the requirements for the air-handling systems and Section 4 the requirements for the control of temperature and humidity. The specific requirements regarding the production of penicillin are covered in Section 11.1. Correspondences to regulation 211.48 are covered in Sections 3.5 and 7 of the interpretation of regulation C.02.004, Section 3.7 of the interpretation of regulation C.02.005, Section 4 of the interpretation of regulation C.02.009, and section "Water Treatment Systems" of the interpretation of regulation C.02.029. Section 7 of the interpretation of regulation C.02.004 states the requirements for the utilities and support systems, including supplies of purified water. Section 3.7 of the interpretation of regulation C.02.005 states the requirements for the operation of water purification, storage, and distribution equipment. Requirements for the quality of water are covered in Section 4 of the interpretation of regulation C.02.009 and in section "Water Treatment Systems" of the interpretation of regulation C.02.029. The requirements for drains are covered in Section 3.5 of the interpretation of regulation C.02.004. Correspondences to regulation 211.50 stating the requirements for the handling of sewage and refuse are covered in Section 2.6 of the interpretation of regulation C.02.007. Correspondences to regulation 211.52 stating the requirements for washing and toilet facilities are covered in Section 5 of the interpretation of regulation C.02.004. Correspondences to regulation 211.56 stating the requirements for sanitation are covered in Sections 1 and 2 of the interpretation of regulation C.02.007. The Canadian GMP code does not state any separate requirements for the handling of organic waste. General requirements for the handling of waste

materials are covered in Section 2.6 of the interpretation of regulation C.02.007. Correspondences to regulation 211.58 stating the requirements for the maintenance of the premises are covered in Section 9 of the interpretation of regulation C.02.004.

***Correspondences in EU GMP Code*** In the EU GMP code [15] issues related to buildings and facilities are mainly covered in Chapter 3 (Premises and Equipment) and partly in Annex 1 (Manufacture of Sterile Medicinal Products). Correspondences to regulation 211.42 are covered in the foreword of Chapter 3 and in Subchapters 3.6–3.8, 3.13, 3.22, 3.23, 3.26, and 3.33. The requirements for the size, construction and location of buildings used in the production are covered in the foreword of Chapter 3. Subchapter 3.8 states the requirements for the placement of equipment and Subchapter 3.7 for the flow of materials and products. Operations, which have to be performed in separate or defined areas, are specified in Subchapters 3.6, 3.13, 3.22, 3.23, 3.26, and 3.33. Annex 1 covers the requirements for facilities used in aseptic processing and Subchapter 3.6 the requirements for the facilities used in the production of penicillin. Correspondences to regulation 211.44 are covered in Subchapters 3.3 and 3.16, which state the requirements for lighting. Correspondences to regulation 211.46 are covered in Subchapters 3.3 and 3.12, which state the requirements for ventilation. The specific requirements for the production of penicillin are covered in Subchapter 3.6. Correspondences to regulation 211.48 are covered in Subchapters 3.10 and 3.11 and in Subsections 35 and 44 of Annex 1. Subchapter 3.10 states the requirements for the plumbing and Subchapter 3.11 the requirements for drains. Section 35 of Annex 1 covers the requirements for water treatment plants and distribution systems and Section 44 the requirements for the monitoring of water sources and water treatment equipment. More guidance on the quality of water is given in the EU guidance document *Note for Guidance on Quality of Water for Pharmaceutical Use* [58]. The EU GMP code does not have correspondence to regulation 211.50, which covers the requirements for the handling of sewage and other refuse. Correspondences to regulation 211.52 are covered in Subchapter 3.31, which covers the requirements for the facilities for washing and toilet purposes. Correspondences to regulation 211.56 are covered in Subchapters 3.2, 3.4, 3.43, and 4.26. Subchapters 3.2 and 3.4 cover the requirements for the conditions to be maintained in the manufacturing facilities. Subchapter 4.26 covers the procedures for cleaning and sanitization and Subchapter 3.43 the requirements for the sanitization of water pipes. The EU GMP code does not cover any separate requirements for the handling of organic waste. Correspondences to regulation 211.58 are covered in Subsection 3.2, which covers the requirements for the maintenance of the buildings used in the production.

***Correspondences in WHO GMP Guideline*** In the WHO GMP guideline [36] issues related to buildings and facilities are mainly covered in Chapter 12 (Premises) and partly in Chapters 3 (Sanitation and Hygiene), 14 (Materials), and 15 (Documentation). Correspondences to regulation 211.42 are covered in Subchapters 12.1, 12.2, 12.4, 12.5, 12.10, 12.14, 12.17, 12.19, 12.22–12.26, and 12.33. The requirements for the size, construction, and location of buildings are stated in Subchapters 12.1, 12.4, and 12.5. Subchapters 12.2 and 12.26 cover the requirements for the placement of equipment and Subchapters 12.10 and 12.25 the requirements for the flow of

materials and products. Operations, which have to be performed in separate or defined areas, are specified in Subchapters 12.14, 12.17, 12.19, 12.22–12.24, and 12.33. The requirements for the facilities used in aseptic processing are covered in Chapter 9 of Annex 6 of the WHO TRS 902 [40] and the requirements for the facilities used in the manufacture of penicillin are in Subchapter 12.24. Correspondences to regulation 211.44 are covered in Subchapters 12.8 and 12.32, which state the requirements for lighting. Correspondences to regulation 211.46 are covered in Subchapters 12.8 and 12.30, which state the requirements for ventilation. The specific requirements for the production of penicillin are covered in Subchapter 12.24. Correspondences to regulation 211.48 are covered in Subchapters 12.28, 12.29, and 14.6 and in Annex 3 of the WHO TRS 929 [45]. Subchapter 12.28 states the requirements for the plumbing and Subchapter 14.6 for the quality of water used in the production of drug products. More guidance on the quality of water is given in Annex 3 of the WHO TRS 929 [45]. The requirements for the drains are stated in Subchapter 12.29. Correspondences to regulation 211.50 are covered in Subchapters 14.44 and 14.45, which state the requirements for the handling of sewage and other refuse. Correspondences to regulation 211.52 are covered in Subchapter 12.12, which states the requirements for the facilities for washing and toilet purposes. Correspondences to regulation 211.56 are covered in Subchapters 3.1, 12.7, 12.9, 14.44–14.46, and 15.48. Subchapters 12.7 and 12.9 state the requirements for the conditions to be maintained in the manufacturing facilities and Subchapter 3.1 the general requirements for sanitation and hygiene. In the WHO GMP guideline there is no separate guidance on the handling of organic waste. General requirements for the handling of waste materials are stated in Subchapters 14.44 and 14.45. Subchapter 15.48 states the requirements for the written procedures for sanitation operations and Subchapter 14.46 for the use of rodenticides, insecticides, fumigating agents, and sanitizing materials. Correspondences to regulation 211.58 are covered in Subchapter 12.6, which states the requirements for the maintenance of the buildings used in drug production.

### 4.4 Equipment

For GMP regulations in the United States issues related to equipment are covered in Subpart D [7], which consists of regulations 211.63, 211.65, 211.67, 211.68, and 211.72. Contents of Subpart D are presented in Table 19. Regulation 211.63 states the requirements for the production equipment covering design, size, and location. Regulation 211.65 states the requirements for the construction of equipment cover-

**TABLE 19 Contents of Subpart D of Part 211 of U.S. GMP Regulations Covering Equipment [7]**

| Section | Subject |
|---|---|
| CFR 211.63 | Equipment design, size, and location |
| CFR 211.65 | Equipment construction |
| CFR 211.67 | Equipment cleaning and maintenance |
| CFR 211.68 | Automatic, mechanical, and electronic equipment |
| CFR 211.72 | Filters |

ing the characteristics of used materials and special requirements for the structure of the equipment. Regulation 211.67 deals with cleaning, maintenance, and sanitizing of equipment and utensils covering the requirements for the procedures for cleaning and maintenance operations. Regulation 211.68 deals with automatic, mechanical, and electronic equipment covering requirements for their calibration and inspection, including the requirements for the documentation of checks and inspections. Furthermore, it covers the requirements for the controls for computer or related systems, including the requirements for the maintenance of backup data. Regulation 211.72 covers the requirements for the filters for liquid filtration used in the manufacture of injectable products, including the specific requirements for the use of fiber-releasing and asbestos-containing filters.

***Correspondences in Canadian GMP Code*** In the Canadian GMP code [12] issues related to equipment are mainly covered in the interpretation of regulation C.02.005 (Equipment) and partly in the interpretation of regulation C.02.007 (Sanitation) and C.02.024 (Records). Correspondences to regulation 211.63 stating the requirements for the design, construction, and location of equipment used in the manufacture of drug products are covered in Sections 1 and 5 of the interpretation of regulation C.02.005. Correspondences to regulation 211.65 stating the requirements for the construction of equipment are covered in Sections 2.1–2.3 of the interpretation of regulation C.02.005. Correspondences to regulation 211.67 stating the requirements for the sanitation are covered in Sections 1, 2, and 3 of the interpretation of regulation C.02.007. Correspondences to regulation 211.68 are covered in Section 5.4 of the interpretation of regulation C.02.005 and in the foreword of the interpretation of regulation C.02.024. Section 5.4 of the interpretation of regulation C.02.005 states the requirements for the use of automatic, mechanical, and electronic equipment, including computerized systems, and the foreword of the interpretation of regulation C.02.024 the requirements for the maintenance of backup data. The Canadian GMP code does not have correspondence to regulation 211.72, which states the requirements for the filters for liquid filtration used in the manufacture of injectable products. Nor does it cover requirements for the use of fiber-releasing or asbestos-containing filters.

***Correspondences in EU GMP Code*** In the EU GMP code [15] issues related to equipment are mainly covered in Chapter 3 (Premises and Equipment) and partly in Chapter 4 (Documentation) and Annexes 1 (Manufacture of Sterile Medicinal Products) and 11 (Computerised Systems). Correspondences to regulation 211.63 are covered in Subchapter 3.34, which states the requirements for the design and location of equipment used in the manufacture of drug products. Correspondences to regulation 211.65 are covered in Subchapters 3.38 and 3.39, which state the requirements for the construction of equipment. Correspondences to regulation 211.67 are covered in Subchapters 3.36, 3.37, and 3.43, which cover the requirements for cleaning and sanitizing the manufacturing equipment. Correspondences to regulation 211.68 are covered in Subchapters 3.41 and 4.9 and Annex 11. Subchapter 3.41 states the requirements for the maintenance of measuring, weighing, recording, and control equipment and Subchapter 4.9 the requirements for the use of electronic data processing systems and the maintenance of backup data. Additional guidance on the use of computerized systems is given in Annex 11. Correspondences

to regulation 211.72 stating the requirements for filters for liquid filtration used in the sterile filtration are covered in Sections 84–87 of Annex 1. The EU GMP code does not have any separate guidance for the use of fiber-releasing or asbestos-containing filters.

***Correspondences in WHO GMP Guideline*** In the WHO GMP guideline [36] issues related to equipment are mainly covered in Chapter 13 (Equipment) and partly in Chapters 14 (Materials), 15 (Documentation), and 16 (Good Practices in Production). Correspondences to regulation 211.63 are covered in Subchapters 13.1 and 13.2, which state the requirements for the design, location, and installation of equipment used in the manufacture of drug products. Correspondences to regulation 211.65 are covered in Subchapters 13.9 and 14.3, which state the requirements for the construction of equipment. Correspondences to regulation 211.67 are covered in Subchapters 13.6, 13.8, 13.12, 16.17, 16.18, and 16.22, which state the requirements for cleaning and sanitizing the equipment. Correspondences to regulation 211.68 are covered in Subchapters 16.23 and 15.9. The requirements for the maintenance of measuring, weighing, recording, and control equipment and instruments are covered in Subchapter 16.23. Subchapter 15.9 states the requirements for the use of electronic data-processing systems, including the requirements for the maintenance of backup data. Correspondences to regulation 211.72 stating the requirements for the use of filters are covered in Subchapters 7.6–7.9 of Annex 6 of the WHO TRS 902 [40]. Subchapter 7.6 covers the requirements for asbestos-containing filters.

### 4.5 Control of Components and Drug Product Containers and Closures

In the United States issues related to control of components and drug product containers and closures are covered in Subpart E [7], which consists of regulations 211.80, 211.82, 211.84, 211.86, 211.87, 211.89, and 211.94. Contents of Subpart E are presented in Table 20. Regulation 211.80 defines the requirements for the procedures for the control of components, containers, and closures. It also states the requirements for their handling, storing, and identification. Regulation 211.82 covers the requirements for receipt and storage of untested components, containers, and

**TABLE 20 Contents of Subpart E of Part 211 of U.S. GMP Regulations Covering Control of Components and Drug Product Containers and Closures [7]**

| Section | Subject |
|---|---|
| CFR 211.80 | General requirements |
| CFR 211.82 | Receipt and storage of untested components, drug product containers, and closures |
| CFR 211.84 | Testing and approval or rejection of components, drug product containers, and closures |
| CFR 211.86 | Use of approved components, drug product containers, and closures |
| CFR 211.87 | Retesting of approved components, drug product containers, and closures |
| CFR 211.89 | Rejected components, drug product containers, and closures |
| CFR 211.94 | Drug product containers and closures |

closures. Regulation 211.84 deals with testing and approval or rejection of components, containers, and closures covering the requirements for sampling, testing, and release. Regulation 211.86 deals with the use of approved components, containers, and closures stating the requirements for the rotation of the storage. Regulation 211.87 states the requirements for the retesting of approved components, containers, and closures. Regulation 211.89 covers the requirements for the handling of rejected components, containers, and closures. Regulation 211.94 deals with drug product containers and closures covering the requirements for materials and the cleanliness of containers and closures. Furthermore, it states the requirements for container closure systems, standards and methods.

***Correspondences in Canadian GMP Code*** In the Canadian GMP code [12] issues related to control of components and drug product containers and closures are covered in interpretations of regulations C.02.009 (Raw Material Testing), C.02.010 (Raw Material Testing), C.02.011 (Manufacturing Control), C.02.014 (Quality Control Department), C.02.016 (Packaging Material Testing), and C.02.017 (Packaging Material Testing). Correspondences to regulation 211.80 stating the general requirements for the handling, storing, and identification of components (raw materials) and drug product containers and closures (packaging materials) are covered in Sections 1, 20, and 21 of the interpretation of regulation C.02.011. Correspondences to regulation 211.82 stating the requirements for the receipt, testing, and storage of untested components and drug product containers and closures are covered in Sections 16, 18, and 19 of the interpretation of regulation C.02.011. Correspondences to regulation 211.84 stating the requirements for testing and approval of components and drug product containers and closures are covered in Sections 6 and 7 of the interpretation of regulation C.02.009, Sections 1–8 of the interpretation of regulation C.02.010, Sections 1 and 2 of regulation C.02.016, Section 4 of its interpretation, and Section 1 of the interpretation of regulation C.02.017. Interpretations 6 and 7 of regulation C.02.009 and interpretations 1–8 of regulation C.02.010 cover the requirements for components. Sections 1 and 2 and interpretation 4 of regulation C.02.016 and interpretation 1 of regulation C.02.017 state the requirements for drug product containers and closures. In the Canadian GMP code there is no correspondence to regulation 211.86, which covers the requirements for the rotation of the storage. Correspondences to regulation 211.87 stating the requirements for the retesting of approved components are covered in Sections 8–10 of the interpretation of regulation C.02.009. For the retesting of drug product containers and closures the Canadian GMP code has no guidance. Correspondences to regulation 211.89 stating the requirements for the handling of rejected components and drug product containers and closures are covered in Section 14 of the interpretation of regulation C.02.011 and in Section 5 of the interpretation of regulation C.02.014. The Canadian GMP code does not have correspondence to regulation 211.94, which covers the requirements for containers and closure systems.

***Correspondences in EU GMP Code*** In the EU GMP code [15] issues related to control of components and drug product containers and closures are mainly covered in Chapter 5 (Production) and partly in Chapter 6 (Quality Control). Correspondences to regulation 211.80 are covered in Subchapters 5.2, 5.7, 5.10, 5.29, and 5.40–5.42, which cover the requirements for the handling, storing, and identification

of components (starting materials) and drug product containers and closures (primary packaging materials). Correspondences to regulation 211.82 are covered in Subchapters 5.5, 5.27, and 5.40, which state the requirements for the receipt, testing, and storage of untested components and drug product containers and closures. Correspondences to regulation 211.84 are covered in Subchapters 5.31, 5.40 and 6.11–6.22 and Annex 8. The general requirements for sampling and testing are covered in Subchapters 6.11–6.22. More guidance on sampling is given in Annex 8. The requirements for the approved use of components and drug product containers and closures are stated in Subchapters 5.31 and 5.40. Correspondences to regulation 211.86 are covered in Subchapter 5.7, which states the requirements for the storage conditions and rotation. Correspondences to regulation 211.87 are covered in Subchapters 5.29 and 5.40, which deal with the retesting of components and drug product containers and closures. Correspondences to regulation 211.89 are covered in Subchapter 5.61, which states the requirements for the handling of rejected components and drug product containers and closures. Correspondences to regulation 211.94 are covered in Subchapter 5.48, which states the requirements for drug product containers and closures.

***Correspondences in WHO GMP Guideline*** In the WHO GMP guideline [36] issues related to control of components and drug product containers and closures are covered in Chapters 14 (Materials), 16 (Good Practices in Production), and 17 (Good Practices in Quality Control). Correspondences to regulation 211.80 are covered in Subchapters, 14.5, 14.13, 14.14, 14.19–14.21, and 16.2, which state the requirements for the handling, storing, and identification of components (starting materials) and drug product containers and closures (primary packaging materials). Correspondences to regulation 211.82 are covered in Subchapters 14.4, 14.9–14.11, and 14.19, which state the requirements for receipt, testing, identification, and storage of untested components and drug product containers and closures. Correspondences to regulation 211.84 are covered in Subchapters 14.12, 14.15, and 17.7–17.17. The requirements for sampling and testing are covered in Subchapters 17.7–17.17 and 14.12. Subchapter 14.15 states the requirements for the approved use of components and drug product containers and closures. Correspondences to regulation 211.86 are covered in Subchapter 14.5, which states the requirements for the storage conditions and the rotation of the storage. Correspondences to regulation 211.87 are covered in Subchapter 14.13, which states the requirements for the retesting of approved components. The WHO GMP guideline does not cover the requirements for the retesting of drug product containers and closures. Correspondences to regulation 211.89 are covered in Subchapter 14.28, which states the requirements for the handling of rejected components and drug product containers and closures. Correspondences to regulation 211.94 are covered in Subchapter 16.19, which states the requirements for the drug product containers and closures.

### 4.6 Production and Process Controls

In the United States GMP regulations on issues related to production and process controls are covered in Subpart F [7], which consists of regulations 211.100, 211.101, 211.103, 211.105, 211.110, 211.111, 211.113, and 211.115. Contents of Subpart F are presented in Table 21. Regulation 211.100 states the requirements for procedures

**TABLE 21 Contents of Subpart F of Part 211 of U.S. GMP Regulations Covering Production and Process Controls [7]**

| Section | Subject |
|---|---|
| CFR 211.100 | Written procedures, deviations |
| CFR 211.101 | Charge-in of components |
| CFR 211.103 | Calculation of yield |
| CFR 211.105 | Equipment identification |
| CFR 211.110 | Sampling and testing of in-process materials and drug products |
| CFR 211.111 | Time limitations on production |
| CFR 211.113 | Control of microbiological contamination |
| CFR 211.115 | Reprocessing |

regarding production and process controls, including the requirements for the documentation and handling of deviations. Regulation 211.101 deals with the requirements for the charge-in of components. Regulation 211.103 states the requirements for the determination of yields. Regulation 211.105 covers requirements for the identification of processing equipment such as containers, processing lines, and major equipment used during manufacture. Regulation 211.110 states the requirements for in-process controls, including the testing and approval of in-process materials and handling of rejected in-process materials. Regulation 211.111 covers the requirements for the time limitations on production, including the handling of deviations from established limits. Regulation 211.113 covers the control of microbiological contaminations. Regulation 211.115 states the requirements for the reprocessing of batches that do not conform to standards or specifications.

***Correspondences in Canadian GMP Code*** In the Canadian GMP code [12] issues related to production and process controls are mainly covered in the interpretation of regulation C.02.011 (Manufacturing Control) and partly in the interpretations of regulations C.02.005 (Equipment), C.02.014 (Quality Control Department), and C.02.029 (Sterile Products). Correspondences to regulation 211.100 are covered in Sections 1–5 of the interpretation of regulation C.02.011. Interpretations 1–4 state the requirements for manufacturing processes and interpretation 5 for the handling of deviations. Correspondences to regulation 211.101 stating the requirements for charge-in of components are covered in Section 22 of the interpretation of regulation C.02.011. Correspondences to regulation 211.103 stating the requirements for the determination of yields including the handling of deviations from the expected yield are covered in Sections 6 and 7 of the interpretation of regulation C.02.011. Correspondences to regulation 211.105 stating the requirements for the identification of piping, containers, equipment, and rooms used in the manufacturing of drug products are covered in Section 3.5 of the interpretation of regulation C.02.005 and in Section 13 of the interpretation of regulation C.02.011. Correspondences to regulation 211.110 are covered in Sections 11 and 14 of the interpretation of regulation C.02.011 and in Section 5 of the interpretation of regulation C.02.014. Section 11 of the interpretation of regulation C.02.011 states the requirements for the in-process controls. The Canadian GMP code does not cover requirements for the testing of in-process materials. The handling of rejected materials is covered in Section 14 of the interpretation of regulation C.02.011 and in Section 5 of the interpretation

of regulation C.02.014. Correspondences to regulation 211.111 dealing with the requirements for the time limitations on production are covered in Section 24.7 of the interpretation of regulation C.02.011. Correspondences to regulation 211.113 are covered in the interpretation of regulation C.02.029, which deals with the manufacture of sterile products. Correspondences to regulation 211.115 stating the requirements for the reprocessing of batches that do not conform to specifications are covered in Sections 7–9 of the interpretation of regulation C.02.014.

***Correspondences in EU GMP*** In the EU GMP code [15] issues related to production and process controls are mainly covered in Chapter 5 (Production) and partly in Chapters 3 (Premises and Equipment), 4 (Documentation), and 6 (Quality Control). Correspondences to regulation 211.100 are covered in Subchapters 5.2, 5.15, and 5.22–5.24. The requirements for manufacturing processes are covered in Subchapters 5.2 and 5.22–5.24. Subchapter 5.15 states the requirements for handling of deviations from instructions or procedures. Correspondences to regulation 211.101 are covered in Subchapters 5.28–5.34, which state the requirements for the charge-in of components. Correspondences to regulation 211.103 are covered in Subchapters 5.8 and 5.39, which state the requirements for determination of yields, including the handling of deviations from the expected yield. Correspondences to regulation 211.105 are covered in Subchapters 3.42 and 5.12, which state the requirements for identification of piping, containers, equipment, and rooms used in the manufacture of drug products. Correspondences to regulation 211.110 are covered in Subchapters 3.17, 4.10, 4.12, 5.38, 5.61, and 6.18. The requirements for in-process controls are covered in Subchapters 3.17, 5.38, and 6.18. Subchapters 4.10 and 4.12 state the requirements for the specifications for in-process materials (intermediate products) and Subchapter 5.61 for handling of rejected materials. The EU GMP code does not cover separate guidance on testing and approval of in-process materials. General guidance on sampling and testing is given in Subchapters 6.11–6.22. Correspondences to regulation 211.111, which deals with the time limitations on production, are covered in Chapter 4.15. Correspondences to regulation 211.113 are covered in Subchapter 5.10 and in Annex 1, which cover the requirements for the control of microbiological contaminations. Correspondences to regulation 211.115 are covered in Subchapters 5.62 and 5.64, which state the requirements for the reprocessing of rejected batches.

***Correspondences in WHO GMP Guideline*** In the WHO GMP guideline [36] issues related to production and process controls are covered in Chapters 13 (Equipment), 14 (Materials), 15 (Documentation), 16 (Good Practices in Production), and 17 (Good Practices in Quality Control). Correspondences to regulation 211.100 are covered in Subchapters 16.1–16.3. Subchapters 16.1 and 16.2 state the requirements for the manufacturing operations and Subchapter 16.3 for the handling of deviations from instructions or procedures. Correspondences to regulation 211.101 are covered in Subchapters 14.12–14.18, which state the requirements for the charge-in of components. Correspondences to regulation 211.103 are covered in Subchapters 16.4 and 16.20, which state the requirements for the determination of yields, including the handling of deviations from the expected yield. Correspondences to regulation 211.105 are covered in Subchapters 13.3, 13.4, and 16.6, which state the requirements for the identification of piping, containers, equipment, and rooms used during pro-

duction. Correspondences to regulation 211.110 are covered in Subchapters 14.28, 15.20, 16.9, 16.16, and 17.8. Subchapters 16.9, 16.16, and 17.8 cover the requirements for the in-process controls and Subchapter 15.20 for the specifications for in-process materials (intermediate products). The requirements for the handling of rejected materials are stated in Subchapter 14.28. Correspondences to regulation 211.111, which deals with the time limitations on production, are covered in Chapter 15.23. Correspondences to regulation 211.113 are covered in Subchapters 16.10–16.14 and in Annex 6 of the WHO TRS 902 [40]. Subchapters 16.10–16.14 cover general requirements for the prevention of cross-contamination and bacterial contamination during production and Annex 6 general requirements for the manufacture of sterile drug products. Correspondences to regulation 211.115 are covered in Subchapters 14.29, 14.31, and 15.40, which state the requirements for the reprocessing of rejected batches.

### 4.7 Packaging and Labeling Control

For GMP regulations in the United States issues related to packaging and labeling control are covered in Subpart G [7], which consists of regulations 211.122, 211.125, 211.130, 211.132, 211.134, and 211.137. The contents of Subpart G is presented in Table 22. Regulation 211.122 deals with materials examination and usage criteria covering the requirements for the receipt, identification, storage, handling, sampling, testing, and approval of labeling and packaging materials, including documentation. Furthermore, it covers the requirements for the control of labeling, handling of obsolete and outdated labeling and packaging materials, and special requirements for different labeling methods. Regulation 211.125 states the requirements for the labeling issuance covering the testing of labeling materials, the control of discrepancy between the quantities of labeling issued, used, and returned, and the handling of excess and returned labeling. Regulation 211.130 states the requirements for the packaging and labeling operations covering the written procedures. Regulation 211.132 states the requirements for the tamper-evident packaging. Regulation 211.134 states the requirements for the inspections of packaged and labeled products covering sampling, examination, and documentation. Regulation 211.137 states the requirements for the expiration dates, including exemptions from the requirements.

**TABLE 22 Contents of Subpart G of Part 211 of U.S. GMP Regulations Covering Packaging and Labeling Control [7]**

| Section | Subject |
|---|---|
| CFR 211.122 | Materials examination and usage criteria |
| CFR 211.125 | Labeling issuance |
| CFR 211.130 | Packaging and labeling operations |
| CFR 211.132 | Tamper-evident packaging requirements for over-the-counter (OTC) human drug products |
| CFR 211.134 | Drug product inspection |
| CFR 211.137 | Expiration dating |

***Correspondences in Canadian GMP Code*** In the Canadian GMP code [12] issues related to packaging and labeling control are covered in the interpretations of regulations C.02.011 (Manufacturing Control), C.02.017 (Packaging Material Testing), C.02.016 (Packaging Material Testing), C.02.019 (Finished Product Testing), and C.02.027 (Stability). Correspondences to regulation 211.122 are covered in Sections 1, 16, 40, and 43–48 of the interpretation of regulation C.02.011, Sections 1, 8, and 9 of the interpretation of regulation C.02.017, and Sections 1 and 4–7 of the interpretation of regulation C.02.016. Sections 1, 16, 43, and 48 of the interpretation of regulation C.02.011 state the general requirements for the handling of packaging and labeling materials covering receipt and storage. Section 8 of the interpretation of regulation C.02.017 states the requirements for the identification of the packaging and labeling materials. The requirements for the testing of the packaging and labeling materials are covered in Sections 1 and 9 of the interpretation of regulation C.02.017. Sections 1 and 4 of the interpretation of regulation C.02.016 and Sections 6 and 7 of the interpretation of regulation C.02.017 state the requirements for the approval of packaging and labeling materials. Sections 44–47 of the interpretation of regulation C.02.011 cover requirements for the use of roll-fed labels, cut labels, gang printing, and the monitoring of the performance of printing. The requirements for the handling of obsolete and outdated packaging and labeling materials are covered in Section 40 of the interpretation of regulation C.02.011 and in Section 5 of the interpretation of regulation C.02.016. Correspondences to regulation 211.125 are covered in Sections 39 and 42 of the interpretation of regulation C.02.011 and in Section 8 of the interpretation of regulation C.02.017. Section 8 of the interpretation of regulation C.02.017 states the requirements for the examination of packaging and labeling materials. The requirements for the control and handling of discrepancy between the quantities of labeling issued, used, and returned are covered in Section 42 and the requirements for the handling of unused batch-coded packaging and labeling materials in Section 39 of the interpretation of regulation C.02.011. Correspondences to regulation 211.130 stating the requirements for the packaging and labeling operations are covered in Sections 29–38 of the interpretation of regulation C.02.011. In Canadian GMP code there is no correspondence to regulation 211.132 stating the requirements for the tamper-evident packaging. Correspondences to regulation 211.134 stating the requirements for the inspections of packaged and labeled products are covered in Section 1 of the interpretation of regulation C.02.019. Correspondences to regulation 211.137 stating the requirements for the expiration dates are covered in regulation C.02.027 and in Section 1 of its interpretation.

***Correspondences in EU GMP Code*** In the EU GMP code [15] issues related to packaging and labeling control are mainly covered in Chapter 5 (Production) and partly in Chapters 4 (Documentation) and 6 (Quality Control). Correspondences to regulation 211.122 are covered in Subchapters 4.11, 4.19, 4.21–4.23, 5.2, 5.40–5.43, and 5.50–5.52. Subchapters 4.19, 4.21–4.23, 5.2, and 5.40–5.42 cover the requirements for purchase, handling, control, storage, and identification of packaging and labeling materials. Specifications for packaging and labeling materials are stated in Subchapter 4.11 and the requirements for handling of outdated or obsolete packaging and labeling materials in Subchapter 5.43. Subchapter 5.51 covers the requirements for the use of cut-labels, off-line overprinting, and roll-feed labels. The requirements for

the control of the printing and labeling operations are stated in Subchapters 5.50 and 5.52. Correspondences to regulation 211.125 are covered in Subchapters 5.2, 5.56, and 5.57. Subchapter 5.2 states the general requirements for the handling of packaging and labeling materials. The requirements for the control of discrepancy between the quantities of labeling issued, used, and returned are covered in Subchapter 5.56 and the requirements for the handling of unused batch-coded packaging and labeling materials in Subchapter 5.57. Correspondences to regulation 211.130 are covered in Subchapters 5.2 and 5.44–5.49. Subchapter 5.2 states the general requirements for the handling of packaging and labeling materials and Subchapters 5.44–5.49 cover the requirements for the packaging and labeling operations. The European Community GMP code does not have correspondence to regulation 211.132, which covers the requirements for the tamper-evident packaging. Correspondences to regulation 211.134 are covered in Subchapters 5.54 and 6.3, which state the requirements for the control of packaged and labeled products. The EU GMP code does not have correspondence to regulation 211.137, which covers the requirements for expiration dates.

***Correspondences in WHO GMP Guideline*** In the WHO GMP guideline [36] issues related to packaging and labeling control are covered in Chapters 6 (Product Recalls), 12 (Premises), 14 (Materials), 15 (Documentation), 16 (Good Practices in Production), and 17 (Good Practices in Quality Control). Correspondences to regulation 211.122 are covered in Subchapters 12.21, 14.19–14.23, 15.18, 16.2, 17.14, and 17.16. Subchapters 6.2, 12.21, 14.19–14.21, 14.23, and 17.16 state the requirements for the purchase, handling, control, storage, and identification of packaging and labeling materials. The requirements for the approval of packaging and labeling materials are covered in Subchapters 17.14 and 15.18. Subchapter 14.20 states the requirements for the use of roll-feed and cut labels and Subchapter 14.22 the requirements for the handling of outdated and obsolete packaging and labeling materials. Correspondences to regulation 211.125 are covered in Subchapters 16.2, 16.34, and 16.35. Subchapter 16.2 states the general requirements for the handling of packaging and labeling materials and Subchapter 16.34 the requirements for the handling of discrepancy between the quantities of labeling issued, used, and returned. The requirements for the handling of unused batch-coded packaging and labeling materials are covered in Subchapter 16.35. Correspondences to regulation 211.130 are covered in Subchapters 16.25–16.30, which state the requirements for the packaging and labeling operations. The WHO GMP guideline does not cover correspondence to regulation 211.132, which covers the requirements for tamper-evident packaging. Correspondences to regulation 211.134 are covered in Subchapter 16.32, which states the requirements for the control of packaged and labeled products. Correspondences to regulation 211.137 are covered in Subchapter 17.24, which states the requirements for the determination of expiration dates and shelf-life specifications.

### 4.8 Holding and Distribution

In the United States GMP regulations on issues related to holding and distribution are covered in Subpart H [7], which consists of regulations 211.142 and 211.150. The contents of Subpart H is presented in Table 23. Regulation 211.142 states the

**TABLE 23 Contents of Subpart H of Part 211 of U.S. GMP Regulations Covering Holding and Distribution [7]**

| Section | Subject |
|---|---|
| CFR 211.142 | Warehousing procedures |
| CFR 211.150 | Distribution procedures |

requirements for the warehousing procedures covering quarantine and storage and regulation 211.150 for the distribution procedures covering distribution order and recalls.

***Correspondences in Canadian GMP Code*** In the Canadian GMP code [12] issues related to holding and distribution are covered in the interpretations of regulations C.02.004 (Premises), C.02.011 (Manufacturing Control), C.02.012 (Manufacturing Control), and C.02.019 (Finished Product Testing). Correspondences to regulation 211.142 stating the requirements for the quarantine and storage of products are covered in Sections 1 and 49 of the interpretation of regulation C.02.011, Section 11.4 of the interpretation of regulation C.02.004, and Section 2 of the interpretation of regulation C.02.019. Correspondences to regulation 211.150 stating the requirements for distribution and recalls are covered in Section 1 of the interpretation of regulation C.02.011 and Section 1 of the interpretation of regulation C.02.012.

***Correspondences in EU GMP Code*** In the EU GMP code [15] issues related to holding and distribution are covered in Chapters 4 (Documentation), 5 (Production), and 8 (Complaints and Product Recall). Correspondences to regulation 211.142 are covered in Subchapters 5.2, 5.58, and 5.60, which state the requirements for the storage and quarantine of products. Correspondences to regulation 211.150 are covered in Subchapters 4.25, 5.2, and 8.8–8.15, which state the requirements for distribution and recalls.

***Correspondences in WHO GMP Guideline*** In the WHO GMP guideline [36] issues related to holding and distribution are covered in Chapters 6 (Product Recalls), 14 (Materials), 15 (Documentation), and 16 (Good Practices in Production). Correspondences to regulation 211.142 are covered in Subchapters 14.4, 14.26, and 16.2, which state the requirements for the storage and quarantine of products. Correspondences to regulation 211.150 are covered in Subchapters 6.1–6.8, 15.45, and 16.2, which state the requirements for distribution and recalls.

### 4.9 Laboratory Controls

In the United States GMP regulations [7] issues related to laboratory controls are covered in Subpart I, which consists of regulations 211.160, 211.165, 211.166, 211.167, 211.170, 211.173, and 211.176. The contents of Subpart I is presented in Table 24. Regulation 211.160 states the requirements for the establishment of laboratory controls such as specifications, standards, sampling plans, and test procedures. Furthermore, it covers the requirements stated for the calibration of instruments, apparatus, gauges, and recording devices. Regulation 211.165 states the require-

**TABLE 24 Contents of Subpart I of Part 211 of U.S. GMP Regulations Covering Laboratory Controls [7]**

| Section | Subject |
|---|---|
| CFR 211.160 | General requirements |
| CFR 211.165 | Testing and release for distribution |
| CFR 211.166 | Stability testing |
| CFR 211.167 | Special testing requirements |
| CFR 211.170 | Reserve samples |
| CFR 211.173 | Laboratory animals |
| CFR 211.176 | Penicillin contamination |

ments for the laboratory testing of batches prior to release covering the requirements for sampling, testing, and approval. Furthermore, it states the requirements for the handling of rejected drug products. Regulation 211.166 states the requirements for stability testing, including the requirements for the determination of expiration dates and the requirements for stability testing of homeopathic drug products. Regulation 211.167 deals with special testing requirements covering sterile products, ophthalmic ointments, and controlled-release dosage forms. Regulation 211.170 states the requirements for reserve samples covering identification, quantity, retention time, and storage. Furthermore it covers the requirements for the deterioration investigations. Regulation 211.173 deals with laboratory animals covering the requirements for their maintenance and control. Regulation 211.176 states the requirements for the testing of penicillin contamination and the handling of penicillin contaminated drug product.

***Correspondences in Canadian GMP Code*** In the Canadian GMP code [12] issues related to laboratory controls are covered in the interpretations of regulations C.02.004 (Premises), C.02.009 (Raw Material Testing), C.02.011 (Manufacturing Control), C.02.014 (Quality Control Department), C.02.015 (Quality Control Department), C.02.016 (Packaging Material Testing), C.02.017 (Packaging Material Testing), C.02.018 (Finished Product Testing), C.02.025 (Samples), C.02.026 (Samples), C.02.027 (Stability), and C.02.028 (Stability). Correspondences to regulation 211.160 stating the general requirements for laboratory controls are covered in regulation C.02.009 and Sections 1–3 and 5–6 of its interpretation, regulation C.02.016 and Sections 1–3 of its interpretation, Section 1 of the interpretation of regulation C.02.017, regulation C.02.018 and Sections 1–5 of its interpretation, and Section 6.4 of the interpretation of regulation C.02.015. Correspondences to regulation 211.165 stating the requirements for the release for distribution including the testing of finished drug products and the handling of rejected drug products are covered in Sections 7 and 14 of the interpretation of regulation C.02.011, Sections 2 and 5 of the interpretation of regulation C.02.014, Section 3 of the interpretation of regulation C.02.015, and Section 2 of regulation C.02.018 and Sections 1 and 4 of its interpretation. Correspondences to regulation 211.166 stating the requirements for stability testing are covered in Section 1 of the interpretation of regulation C.02.027 and Sections 1 and 2 of the interpretation of regulation C.02.028. The Canadian GMP code does not cover separate requirements for the stability testing

of homeopathic drug products. Correspondences to regulation 211.167 stating the requirements for sterility testing are covered in Sections 1–4 of the interpretation of regulation C.02.029 (Sterile Products). The Canadian GMP code does not have any guidance covering the testing of ophthalmic ointments and controlled-release dosage forms. Correspondences to regulation 211.170 stating the requirements for reserve samples are covered in Section 1 of regulation C.02.025 and in regulation C.02.026 and Sections 1 and 3–5 of their interpretation. Correspondences to regulation 211.173 stating the requirements for laboratory animals are covered in Section 2.4 of the interpretation of regulation C.02.004. The Canadian GMP code does not have correspondence to regulation 211.176, which covers the requirements for the testing and handling of penicillin contamination.

***Correspondences in EU GMP Code*** In the EU GMP code [15] issues related to laboratory controls are covered in Chapters 1 (Quality Management), 4 (Documentation), 5 (Production), and 6 (Quality Control) and in Annexes 1 (Manufacture of Sterile Medicinal Products), 9 (Manufacture of Liquids, Creams, and Ointments), and 19 (Reference and Retention Samples). Correspondences to regulation 211.160 are covered in Subchapters 1.4, 4.2, 4.3, 4.10–4.13, 5.15, 6.7, and 6.18, which cover the general requirements for laboratory controls. Correspondences to regulation 211.165 are covered in Subchapters 4.22, 4.23, 5.61, 5.62, 6.3, 6.11, and 6.15, which state the requirements for the release for distribution, the testing of finished drug products, and the handling of rejected drug products. The EU GMP code does not have correspondence to regulation 211.166, which states the requirements for stability testing. However, there is a separate guideline, *Stability Testing on Active Ingredients and Finished Products* [59], which provides guidance on issues related to stability testing. Furthermore, Subchapters 6.23–6.33 cover the requirements for the on-going stability program. Correspondences to regulation 211.167 are covered in Annexes 1 and 9. Section 93 of Annex 1 covers the requirements for sterility testing and Annex 9 the requirements for ointments. In the EU GMP code there is no guidance on the testing of the controlled-release dosage forms. Correspondences to regulation 211.170 are covered in Subchapters 1.4 and 6.12 and Annex 19, which state the requirements for reserve samples. Correspondences to regulation 211.173 are covered in Subchapters 3.33 and 6.22, which state the requirements for the maintenance of animals. The EU GMP code does not have correspondence to regulation 211.176, which covers the requirements for the testing and handling of penicillin contaminations.

***Correspondences in WHO GMP Guideline*** In the WHO GMP guideline [36] issues related to laboratory controls are covered in Chapters 14 (Materials), 15 (Documentation), 16 (Good Practices in Production), and 17 (Good Practices in Quality Control). Correspondences to regulation 211.160 are covered in Subchapters 15.14–15.16, 15.18–15.21, 16.3, and 16.23, which state the general requirements for laboratory controls. Correspondences to regulation 211.165 are covered in Subchapters 14.28, 14.29, 15.13, 15.42, 17.7–17.13, 17.19, and 17.20, which state the requirements for the release for distribution covering the testing of finished drug products and the handling of rejected drug products. Correspondences to regulation 211.166 are covered in Subchapters 17.23–17.26, which state the requirements for stability testing. The WHO GMP guideline does not cover separate requirements

for the stability testing of homeopathic drug products. Correspondences to regulation 211.167 are covered in Annex 6 of the WHO TRS 902 [40], which states the requirements for sterility testing. The WHO GMP guideline does not cover any requirements for the testing of ophthalmic ointments or controlled-release dosage forms. Correspondences to regulation 211.170 are covered in Subchapter 17.22, which states the requirements for reserve samples. The WHO GMP guideline does not have correspondence to regulation 211.173, which covers the requirements for the maintenance of laboratory animals. Nor does it have correspondence to regulation 211.176, which covers the requirements for the testing of penicillin contaminations.

### 4.10 Records and Reports

In the United States issues related to records and reports are covered in Subpart J [7], which consists of regulations 211.180, 211.182, 211.184, 211.186, 211.188, 211.192, 211.194, 211.196, and 211.198. The contents of Subpart J is presented in Table 25. Regulation 211.180 states the general requirements for documentation covering maintenance, retention times, and availability of the records. Furthermore, it states the requirements for the annual quality standards evaluation. Regulation 211.182 states the requirements for individual equipment logs. Regulation 211.184 states the requirements for component, drug product container, closure, and labeling records. Regulation 211.186 states the requirements for master production and control records. Regulation 211.188 states the requirements for batch production and control records. Regulation 211.192 states the requirements for the review and approval of production and control records, including the requirements for the investigation of any unexplained discrepancies. Regulation 211.194 states the requirements for laboratory records, including the requirements for the documentation of modifications. Furthermore, it covers the requirements for the documentation of the testing and standardization of reference standards, reagents, and standard solutions; calibration of laboratory instruments and recording devices; and stability tests. Regulation 211.196 states the requirements for the distribution records. Regulation 211.198 states the requirements for the handling of complaints, including the maintenance and retention times of complaint files.

**TABLE 25 Contents of Subpart J of Part 211 of U.S. GMP Regulations Covering Records and Reports [7]**

| Section | Subject |
|---|---|
| CFR 211.180 | General requirements |
| CFR 211.182 | Equipment cleaning and use log |
| CFR 211.184 | Component, drug product container, closure, and labeling records |
| CFR 211.186 | Master production and control records |
| CFR 211.188 | Batch production and control records |
| CFR 211.192 | Production record review |
| CFR 211.194 | Laboratory records |
| CFR 211.196 | Distribution records |
| CFR 211.198 | Complaint files |

***Correspondences in Canadian GMP Code*** In the Canadian GMP code [12] issues related to records and reports are mainly covered in regulations C.02.021, C.02.022, C.02.023, and C.02.024 (Records) and in their interpretations and partly in the interpretations of regulations C.02.005 (Equipment), C.02.010 (Raw Material Testing), C.02.011 (Manufacturing Control), C.02.012 (Manufacturing Control), C.02.014 (Quality Control Department), C.02.015 (Quality Control Department), and C.02.017 (Packaging Material Testing). Correspondences to regulation 211.180 stating the general requirements for the maintenance of records, including periodic quality evaluation (self-inspection) and the retention time of the records are covered in regulations C.02.021, C.02.022, C.02.023, and C.02.024 and their interpretations and in Section 2 of the interpretation of regulation C.02.012. Correspondences to regulation 211.182 stating the requirements for individual equipment logs are covered in Section 5.5 of the interpretation of regulation C.02.005. Correspondences to regulation 211.184 stating the requirements for records to be kept on components, drug product containers, closures, and labeling are covered in Sections 4 and 5 of the interpretation of regulations C.02.020–24, Section 5 of the interpretation of regulation C.02.010, and Section 7 of the interpretation of regulation C.02.017. Correspondences to regulation 211.186 stating the requirements for the master production and control records (manufacturing and packaging master formulas) are covered in Sections 23–25 of the interpretation of regulation C.02.011 and Section 1.1 of the interpretation of regulations C.02.020–24. Correspondences to regulation 211.188 stating the requirements for the batch production and control records (manufacturing and packaging batch document) are covered in Sections 26, 27, 29, and 30 of the interpretation of regulation C.02.011 and Section 1.2 of the interpretation of regulations C.02.020–24. Correspondences to regulation 211.192 stating the requirements for review and approval of production and control records including investigation of batch deviations are covered in Section 2 of the interpretation of regulation C.02.014. Correspondences to regulation 211.194 stating the requirements for laboratory records are covered in Sections 6.4, 6.6, and 6.7 of the interpretation of regulation C.02.015. Correspondences to regulation 211.196 stating the requirements for distribution records are covered in Section 1.6 of the interpretation of regulation C.02.012 and Section 2.1 of the interpretation of regulations C.02.020–24. Correspondences to regulation 211.198 stating the requirements for the maintenance of complaint files including retention times are covered in Section 4 of the interpretation of regulation C.02.015, Section 3.1 of the interpretation of regulations C.02.020–24, and regulation C.02.023.

***Correspondences in EU GMP Code*** In the EU GMP code [15] issues related to records and reports are mainly covered in Chapter 4 (Documentation) and partly in Chapters 1 (Quality Management), 5 (Production), 6 (Quality Control), 8 (Complaints and Product Recall), and 9 (Self Inspection). Correspondences to regulation 211.180 are covered in Subchapters 4.1–4.9, 6.8, and 9.1–9.3, which state the general requirements for the maintenance of the records, including periodic quality evaluation (self-inspection) and retention times. Correspondences to regulation 211.182 are covered in Subchapters 4.28 and 4.29, which state the requirements for individual equipment logs. Correspondences to regulation 211.184 are covered in Subchapters 4.19 and 4.20, which state the requirements for the records to be kept on the receipt of components, drug product containers, closures, and labeling. Correspon-

dences to regulation 211.186 are covered in Subchapters 4.14–4.16, which state the requirements for the master production and control records (manufacturing formula, processing, and packaging instructions). Correspondences to regulation 211.188 are covered in Subchapters 4.17 and 4.18, which state the requirements for the batch production and control records (batch processing and packaging record). Correspondences to regulation 211.192 are covered in Subchapters 1.4, 4.3, 4.24, 5.8, and 5.39, which state the requirements for review and approval of production and control records, including the investigation of unexplained discrepancies. Correspondences to regulation 211.194 are covered in Subchapters 3.41, 6.7, 6.17, 6.20, and 6.21, which state the requirements for laboratory records. Correspondences to regulation 211.196 are covered in Subchapter 4.25, which states the requirements for distribution records. Correspondences to regulation 211.198 are covered in Subchapters 4.26 and 8.1–8.8, which state the requirements for the handling of complaints.

***Correspondences in WHO GMP Guideline*** In the WHO GMP guideline [36] issues related to records and reports are covered in Chapters 5 (Complaints), 8 (Self-Inspection and Quality Audits), 13 (Equipment), 14 (Materials), 15 (Documentation), 16 (Good Practices in Production), and 17 (Good Practices in Quality Control). Correspondences to regulation 211.180 are covered in Subchapters 8.1–8.6 and 15.1–15.9, which state the general requirements for the maintenance of the records, including periodic quality evaluation (self-inspection) and retention times. Correspondences to regulation 211.182 are covered in Subchapters 15.46 and 15.47, which state the requirements for individual equipment logs. Correspondences to regulation 211.184 are covered in Subchapters 15.32 and 15.33, which state the requirements for the records to be kept on the receipt of components, drug product containers, closures, and labeling. Correspondences to regulation 211.186 are covered in Subchapters 15.22–15.24, which state the requirements for the master production and control records (master formula and packaging instructions). Correspondences to regulation 211.188 are covered in Subchapters 15.25–15.30, which state the requirements for the batch production and control records (batch processing and packaging records). Correspondences to regulation 211.192 are covered in Subchapters 16.4, 16.20, and 17.21, which state the requirements for review and approval of production and control records covering also the requirements for the investigation of unexplained discrepancies. Correspondences to regulation 211.194 are covered in Subchapters 13.5, 14.34, 14.35, 14.41, 15.12, 15.42, 15.43, and 16.23, which states the requirements for laboratory records. Correspondences to regulation 211.196 are covered in Subchapter 15.45, which states the requirements for the distribution records. Correspondences to regulation 211.198 are covered in Subchapters 5.1–5.10, which state the requirements for the handling of complaints.

### 4.11 Returned and Salvaged Drug Products

In the United States GMP regulation [7] issues related to returned and salvaged drug products are covered in Subpart K, which consists of regulations 211.204 and 211.208. The contents of Subpart K is presented in Table 26. Regulation 211.204 states the requirements for the handling of returned drug products, including repro-

**TABLE 26 Contents of Subpart K of Part 211 of U.S. GMP Regulations Covering Returned and Salvaged Drug Products [7]**

| Section | Subject |
|---|---|
| CFR 211.204 | Returned drug products |
| CFR 211.208 | Drug product salvaging |

cessing and documentation. Regulation 211.208 states the requirements for drug product salvaging.

***Correspondences in Canadian GMP Code*** In the Canadian GMP code [12] issues related to returned and salvaged drug products are covered in the interpretation of regulation C.02.014 (Quality Control Department). Correspondences to regulation 211.204 stating the requirements for the handling of returned drug products are covered in Section 4 of the interpretation of regulation C.02.014. The Canadian GMP code does not have correspondence to regulation 211.208, which covers the requirements for drug product salvaging.

***Correspondences in EU GMP Code*** In the EU GMP code [15] issues related to returned and salvaged drug products are covered in Chapters 4 (Documentation) and 5 (Production). Correspondences to regulation 211.204 are covered in Subchapters 4.26 and 5.26, which state the requirements for the handling of returned drug products. The EU GMP code does not have correspondence to regulation 211.208, which covers the requirements for drug product salvaging.

***Correspondences in WHO GMP Guideline*** In the WHO GMP guideline [36] issues related to returned and salvaged drug products are covered in Chapter 14 (Materials). Correspondences to regulation 211.204 are covered in Subchapter 14.33, which states the requirements for the handling of returned drug products. The WHO GMP guideline does not have correspondence to regulation **211.208**, which covers the requirements for drug product salvaging.

## REFERENCES

1. Immel, B. K. (2001), A brief history of the GMPs for pharmaceuticals, *Pharm. Technol. No. Am.*, 25(7), 44–48.
2. Anonymous *Pharmaceutical Administration and Regulations in Japan*, Japan Pharmaceutical Manufacturers Association, available: http://www.jpma.or.jp/english/library/pdf/2005.pdf.
3. Vesper, J. L. (2003), So what are GMPs, anyway? *BioProcess Int.*, 1(2), 24–29.
4. Rosin, L. J. (2006), Regulatory affairs: If you didn't write it down, it didn't happen, *BioProcess Int.*, 4(3, Suppl), 16–23.
5. Anonymous (2002), Sec. 351: Adulterated drugs and devices, in *United States Code*, Title 21, Chapter 9, Subchapter V, Part A, U.S. Government Printing Office, Washington, DC, available: http://frwebgate.access.gpo.gov/cgi-bin/getdoc.cgi?dbname=browse_usc&docid=Cite:+21USC351.

6. Anonymous (2005), Part 210: Current good manufacturing practice in manufacturing, processing, packing, or holding of drugs: General, in *Code of Federal Regulations*, Title 21, Chapter I, U.S. Government Printing Office, Washington, DC, pp. 118–119, available: http://www.access.gpo.gov/nara/cfr/waisidx_05/21cfr210_05.html.
7. Anonymous (2005), Part 211: Current good manufacturing practice for finished pharmaceuticals, in *Code of Federal Regulations*, Title 21, Chapter I, U.S. Government Printing Office, Washington, DC, pp. 120–141, available: http://www.access.gpo.govnara/cfr/waisidx_05/21cfr211_05.html.
8. Grazal, J. G., and Earl, D. S. (1997), EU and FDA regulations: Overview and comparison, *Qual. Assur. J.*, 2, 55–60.
9. Anonymous (2006), Guidance page, available: http://www.fda.gov/cder/guidance/index.htm.
10. Anonymous (2001), Guidance for industry: Q7A Good manufacturing practice guidance for active pharmaceutical ingredients, Office of Training and Communications, available: http://www.fda.gov/cber/gdlns/ichactive.pdf.
11. Anonymous (2005), Division 2: Good manufacturing practices, in *Consolidated Statutes and Regulations*, Food and Drugs Act, Food and Drug Regulations, Part C, Department of Justice, Canada, available: http://laws.justice.gc.ca/en/f-27/c.r.c.-c.870/230079.html.
12. Anonymous (2002), Good manufacturing practices guidelines, Version 2, Health Products and Food Branch Inspectorate, available: http://www.hc-sc.gc.ca/dhp-mps/alt_formats/hpfb-dgpsa/pdf/compli-conform/2002v2_e.pdf.
13. Anonymous (2001), Directive 2001/83/EC of the European Parliament and of the Council, *Off. J. Eur. Union*, 44(L311), 67–128, available: http://europa.eu.int/eur-lex/pri/en/oj/dat/2001/l_311/l_31120011128en00670128.pdf.
14. Anonymous (2003), Commission directive 2003/94/EC, *Off. J. Eur. Union*, 46(L262), 22–26, available: http://europa.eu.int/eur-lex/pri/en/oj/dat/2003/l_262/l_26220031014en00220026.pdf.
15. Anonymous (2005), EU guidelines to good manufacturing practice, in *The Rules Governing Medicinal Products in the European Union*, Vol. 4, European Commission Enterprise and Industry Directorate-General, available: http://pharmacos.eudra.org/F2/eudralex/vol-4/home.htm.
16. Anonymous (1998), Quality and biotechnology, in *The Rules Governing Medicinal Products in the European Union*, Vol. 3A, European Commission Directorate General III, available: http://pharmacos.eudra.org/F2/eudralex/vol-3/home.htm.
17. Anonymous (2005), *Regulations for Buildings and Facilities of Pharmacies*, etc., MHLW Ministerial Ordinance, Yakuji Nippon Ltd., Tokyo, No.73.
18. Kim, Y.-O., Ha, K.-W., and Choi, K.-S. (2001), Safety evaluation for new drug approval in Korea, *Drug Info. J.*, 35(1), 285–291.
19. Shin, S.-G. (1998), Current status of clinical trials in the Republic of Korea, *Drug Info. J.*, 32(Suppl), 1217S–1222S.
20. Anonymous (2000), *Competition in the Pharmaceutical Industry—Republic of Korea*, Working Party No. 2 on Competition and Regulation, Committee on Competition Law and Policy, OECD Publishing, Paris.
21. Anonymous (2001), *Drug Administration Law of the People's Republic of China*, Order of the President of the People's Republic of China No. 45, available: http://www.sfda.gov.cn/cmsweb/webportal/W45649037/A48335975.html.
22. Deng, R., and Kaitin, K. I. (2004), The regulation and approval of new drugs in China, *Drug Info. J.*, 38(1), 29–39.

23. Anonymous (2005), The Drugs and Cosmetics Act and Rules, Ministry of Health and Family Welfare, Department of Health, available: http://cdsco.nic.in/html/Drugs&CosmeticAct.pdf.
24. Anonymous (2005), Schedule M: Good manufacturing practices and requirements of premises, plant and equipment for pharmaceutical products, in The Drugs and Cosmetics Act and Rules, Ministry of Health and Family Welfare, Department of Health, pp. 386–436, available: http://cdsco.nic.in/html/Drugs&CosmeticAct.pdf.
25. Venkateswarlu, M. (2006), Why do we need revision of schedule M, available: http://www.pharmabiz.com, 2006.
26. Anonymous (2006), Section 36: Manufacturing principles, in *Therapeutic Goods Act 1989*, Office of Legislative Drafting and Publishing, pp. 96–97, available: http://www.tga.gov.au/legis/index.htm#instruments.
27. Slater, T. (2002), Therapeutic goods (manufacturing principles) determination no 2 of 2002, *Commonwealth Austral. Gaz.*, GN 34, 2306–2307.
28. Anonymous (2002), Australian code of good manufacturing practice for medicinal products, Therapeutic Goods Administration, available: http://www.tga.gov.au/docs/pdf/gmpcodau.pdf.
29. Anonymous (2005), Medicines Act 1981, Parliamentary Counsel Office, available: http://www.legislation.govt.nz/browse_vw.asp?content-set=pal_statutes.
30. Anonymous (2001), Guidance notes for applicants for consent to distribute new and changed medicines and related products, in *New Zealand Regulatory Guidelines for Medicines*, Vol. 1, 5th ed., MedSafe, available: http://www.medsafe.govt.nz/downloads/vol1.doc.
31. Anonymous (2005), New Zealand code of good manufacturing practice for manufacture and distribution of therapeutic goods, available: http://www.medsafe.govt.nz/Regulatory/Guideline/code.htm.
32. Anonymous (2002), Medicines and Related Substances Control Act 101 of 1965, Medicines Control Council, available: http://www.mccza.com/showdocument.asp?Cat=27&Desc=Acts%20and%20Regulations.
33. Anonymous (2003), General regulations made in terms of the Medicines and Related Substances Act 1965 (Act no. 101 of 1965) as Amended, Government Notice, Department of Health, available: http://www.mccza.com/showdocument.asp?Cat=27&Desc=Acts%20and%20Regulations.
34. Anonymous (2005), Guide to Good Manufacturing Practice for Medicines in South Africa, Medicines Control Council, available: http://www.mccza.com/showdocument.asp?Cat=21&Desc=Guidelines%20-%20Good%20Manufacturing%20Practices.
35. Anonymous (2002), WHO Expert Committee on Specifications for Pharmaceutical Preparations: 37th Report, WHO Technical Report Series 908, World Health Organization, Singapore, available: http://whqlibdoc.who.int/trs/WHO_TRS_908.pdf.
36. Anonymous (2002), Annex 4: Good manufacturing practices for pharmaceutical products: Main principles, in WHO Expert Committee on Specifications for Pharmaceutical Preparations: 37th Report, WHO Technical Report Series 908, World Health Organization, Singapore, pp. 36–89, available: http://whqlibdoc.who.int/trs/WHO_TRS_908.pdf.
37. Anonymous (2005), Annex 2: Good manufacturing practices: Requirement for the sampling of starting materials (Amendment), in WHO Expert Committee on Specifications for Pharmaceutical Preparations: 39th Report, WHO Technical Report Series 929, World Health Organization, Singapore, pp. 38–39, available: http://whqlibdoc.who.int/trs/WHO_TRS_929_eng.pdf.
38. Anonymous (1992), Annex 1: Good manufacturing practices for pharmaceutical products: Good manufacturing practices for active pharmaceutical ingredients (bulk drug

substances), in WHO Expert Committee on Specifications for Pharmaceutical Preparations: 32th Report, WHO Technical Report Series 823, World Health Organization, Geneva, pp. 72–79, available: http://whqlibdoc.who.int/trs/WHO_TRS_823.pdf.

39. Anonymous (1998), Annex 5: Good manufacturing practices: Supplementary guidelines for the manufacture of pharmaceutical excipients, in WHO Expert Committee on Specifications for Pharmaceutical Preparations: 35th Report, WHO Technical Report Series 885, World Health Organization, Madrid, Spain, pp. 50–71, available: http://whqlibdoc.who.int/trs/WHO_TRS_885.pdf.
40. Anonymous (2002), Annex 6: Good manufacturing practices for sterile pharmaceutical products, in WHO Expert Committee on Specifications for Pharmaceutical Preparations: 36th Report, WHO Technical Report Series 902, World Health Organization, Singapore, pp. 76–93, available: http://whqlibdoc.who.int/trs/WHO_TRS_902.pdf.
41. Anonymous (1993), Annex 3: Good manufacturing practices for biological products, in WHO Expert Committee on Specifications for Pharmaceutical Preparations: 33th Report, WHO Technical Report Series 834, World Health Organization, Geneva, pp. 20–30, available: http://whqlibdoc.who.int/trs/WHO_TRS_834.pdf.
42. Anonymous (1995), Annex 7: Good manufacturing practices: Supplementary guidelines for the manufacture of investigational pharmaceutical products for clinical trials in humans, in WHO Expert Committee on Specifications for Pharmaceutical Preparations: 34th Report, WHO Technical Report Series 863, World Health Organization, Geneva, pp. 97–108, available: http://whqlibdoc.who.int/trs/WHO_TRS_863_(p1-p98).pdf (pp. 97–98); http://whqlibdoc.who.int/trs/WHO_TRS_863_(p99-p194).pdf (pp. 99–108).
43. Anonymous (1995), Annex 8: Good manufacturing practices: Supplementary guidelines for the manufacture of herbal medicinal products, in WHO Expert Committee on Specifications for Pharmaceutical Preparations: 34th Report, WHO Technical Report Series 863, World Health Organization, Geneva, pp. 109–113, available: http://whqlibdoc.who.int/trs/WHO_TRS_863_(p99-p194).pdf.
44. Anonymous (2002), Annex 3: Guidelines on good manufacturing practices for radiopharmaceutical products, in WHO Expert Committee on Specifications for Pharmaceutical Preparations: 37th Report, WHO Technical Report Series 908, World Health Organization, Singapore, pp. 26–35, available: http://whqlibdoc.who.int/trs/WHO_TRS_908.pdf.
45. Anonymous (2005), Annex 3: WHO Good manufacturing practices: Water for pharmaceutical use, in WHO Expert Committee on Specifications for Pharmaceutical Preparations: 39th Report, WHO Technical Report Series 929, World Health Organization, Singapore, pp. 40–58, available: http://whqlibdoc.who.int/trs/WHO_TRS_929_eng.pdf.
46. Anonymous (2006), Background to PIC, available: http://www.picscheme.org/indexnoflash.php?p=backg.
47. Anonymous (2006), Introduction, available: http://www.picscheme.org/indexnoflash.php?p=intro.
48. Anonymous (2006), List of PIC/S participating authorities (& observers), available: http://www.picscheme.org/indexnoflash.php?p=members.
49. Brunner, D. (2004), Pharmaceutical inspection co-operation scheme (PIC/S), *Qual. Assur. J.*, 8, 207–211.
50. Anonymous (2006), Guide to good manufacturing practice for medicinal products, PE 009–3, Pharmaceutical inspection co-operation scheme, available: http://www.picscheme.org/guides.php#.
51. Anonymous (2006), Structure of ICH, available: http://www.ich.org/cache/html/510-272-1.html.

52. Anonymous (2000), Good manufacturing practice guide for active pharmaceutical ingredients Q7, ICH harmonised tripartite guideline, ICH Steering Committee, available: http://www.ich.org/LOB/media/MEDIA433.pdf.
53. Anonymous (2006), Quality guidelines, available: http://www.ich.org/cache/compo/363-272-1.html.
54. Anonymous (2006), The founding of ASEAN, available: http://www.aseansec.org/7069.htm.
55. Anonymous (2006), Pharmaceuticals, available: http://www.aseansec.org/8657.htm.
56. Vernengo, M. J. (1998), Advances in pharmaceutical market integration in MERCOSUR and other Latin American countries, *Drug Info. J.*, 32(3), 831–839.
57. Anonymous (2005), Division 1A establishment licences, in *Consolidated Statutes and Regulations*, Food and Drugs Act, Food and Drug Regulations, Part C, Department of Justice Canada, available: http://laws.justice.gc.ca/en/f-27/c.r.c.-c.870/230049.html.
58. Anonymous (2002), Note for guidance on quality of water for pharmaceutical use, CPMP/QWP/158/01, Committee for Proprietary Medicinal Products, Quality Working Party, London, available: http://www.emea.eu.int/pdfs/human/qwp/015801en.pdf.
59. Anonymous (1998), Stability testing on active ingredients and finished products, in *The Rules Governing Medicinal Products in the European Union*, Vol. 3A, European Commission Directorate General III, pp. 143–151, available: http://pharmacos.eudra.org/F2/eudralex/vol-3/pdfs-en/3aq16aen.pdf.

# ANALYTICAL AND COMPUTATIONAL METHODS AND EXAMPLES FOR DESIGNING AND CONTROLLING TOTAL QUALITY MANAGEMENT PHARMACEUTICAL MANUFACTURING SYSTEMS

Paul G. Ranky,[1] Gregory N. Ranky,[2] Richard G. Ranky,[1] and Ashley John[1]

[1]*New Jersey Institute of Technology, Newark, New Jersey*
[2]*Public Research University of New Jersey, Newark, New Jersey*

**Contents**

1 Introduction
2 Flexible Pharmaceutical Manufacturing and Assembly System Design
3 Flexible Manufacturing Model Integrated with Design
4 Real-Time Operation Control
5 Innovative Design
6 Open Innovation Architecture
7 Generic, Object-Oriented Innovation Process Modeling Method and Sample Model
8 Systems Approach to Pharmaceutical Manufacturing Systems Management
9 Requirements Analysis for System Product, Process, and Service Design Innovation
10 Innovation Risk Analysis and Opportunity Method and Tool with Pharmaceutical Manufacturing System Applications
11 Open-Source Computational Statistical and Three-Dimensional Multimedia for Pharmaceutical Manufacturing System Innovation and Project Communication
12 RFID Applications
13 RFID Examples

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

14 RFID Integration Models for Digital Pharmaceutical Manufacturing and Assembly Supply Chains
15 Evaluation of Network Simulation Results
16 Summary
17 Complimentary Video on DVD
References

## 1 INTRODUCTION

Total quality management (TQM) and operation control in pharmaceutical manufacturing system design engineering is essential. TQM-focused pharmaceutical manufacturing system engineering involves the continual satisfaction of customer requirements at lowest cost by harnessing the efforts of everybody in the company. Quality assurance means sustaining a system that prevents defects. This includes quality control and quality engineering. Quality control means establishing and maintaining specified quality standards of products; quality engineering is the establishment and execution of tests to measure product quality and adherence to acceptance criteria.

This chapter explains the importance of reducing variation for the purpose of implementing total quality in every process of the pharmaceutical design and manufacturing enterprise. Furthermore, it represents a modular product, process, service design, implementation, and management approach to the introduction of various TQM methods, tools, technologies, and their management issues within a variety of small, medium, and large enterprises for the purpose of designing and controlling pharmaceutical manufacturing systems.

These aspects are very important, clearly illustrated by the fact that the U.S. Food and Drug Administration (FDA) has three classification levels for medical products:

- Class I products are passive devices that do not enter the patient's body or contact only the skin.
- Class II products are active devices or devices that are used to administer fluids to the patient's body.
- Class III products are implanted inside the patient's body.

The FDA is familiar with the complexity of designing pharmaceutical systems. To support this activity, there are several software tools that help product/process and system designers to achieve the above.

It should also be noted that the FDA expects design validation results to accompany some submissions. This is particularly true of class II and III devices. The agency expects such analysis results to match those obtained with established experimental methods. A number of software tools, including finite element analysis (FEA), motion and actuation simulation, computational fluid dynamics (CFD), in conjunction with the computer-aided design (CAD) used for the designs themselves and other solutions are available that help today's pharmaceutical/medical designer/medical manufacturing/assembly system designer to meet the complex requirements of the industry as well as the FDA. (The key, here, is to accept the important

principle that pharmaceutical design and manufacturing/assembly and even packaging must be an integrated approach.)

The main problems when applying a traditional quality management philosophy to any pharmaceutical design/manufacturing/assembly challenge include the following:

- This philosophy focuses on correcting mistakes after they have been made, rather than preventing them in the first place.
- It allows mistakes to be made. It actually builds them into every aspect of the system, typically costing around 20% of the turnover.
- It accepts that quality has to be sacrificed as the volume and the productivity go up.
- As viewed by accountants, it is an expensive add on item of the value chain.

However, modern thinking claims that, because TQM involves every person, aspect, and machine of the organization, it requires a total commitment. It is not a "test-and-fix" approach. It is a preventive system designed into every aspect of the world-class design, manufacturing, and service enterprise, including product design, manufacture, and management (and even in accounting terms costing somewhat less than conventional quality systems, i.e., typically around 10% of the turnover).

The fundamental goal of TQM and TQC (total quality management and control) is to program, measure, and keep process variability under control. Some of these methods discussed in this chapter are as follows:

- Pharmaceutical manufacturing system design methods and tools with examples
- Process modeling for designing and running pharmaceutical manufacturing systems
- Requirements analysis modeling for pharmaceutical manufacturing systems
- Risk analysis modeling for pharmaceutical manufacturing systems
- Dynamic modeling and network simulation for globally distributed pharmaceutical manufacturing systems and other methods and tools

## 2 FLEXIBLE PHARMACEUTICAL MANUFACTURING AND ASSEMBLY SYSTEM DESIGN

A flexible pharmaceutical manufacturing/assembly system, (FMS) is a highly automated, distributed feedback-controlled system of data, information, and physical processors, such as computer and manually controlled machines, cells, workstations, and robots, in which decisions have to be made often in real time. This is only possible if all information processors (including the human resources of such systems) are "well informed" and lean/flexible, meaning that they have the exact information at the exact time, format, and mode they need to allow responsible decision making within given time constraints. Note that this is a fundamentally different system design concept than that of the transfer line, operating on a fixed cycle time, and designed for large batch production [1–8].

When designing a flexible manufacturing/assembly system (FMS/FAS), the design team should consider the following steps:

1. Collect all current and possible future user and system requirements.
2. Analyze the system (i.e., the data processing and the FMS/ FAS hardware and software constraints).
3. Design an appropriate data structure and database for describing processors and their resources, such as machines, robots, and tools (and/or robot hands, probes, sensory-based inspection and assembly tools, etc.).
4. Specify and design programs and query routines and dialogues that are capable of accessing this database as well as communicating with the real-time production planning and control system of the FMS/FAS.
5. Design and integrate the system with the rest of the hardware and software, including on-line manuals, education, and training packages, preferably in interactive, engineering multimedia format.
6. Maintain the system and continuously learn for the benefit of the existing as well as future system designs.

Probably the most important questions to be answered before starting to design such a system are: Who is going to use it? For what purposes? With what data? How will it be used?

As an example, consider that tooling data in FMSs will typically be used by several subsystems as well as by human beings are as follows:

- Production planning subsystem
- Process control
- Part programming
- Tool preset and tool maintenance
- Tool assembly (manual or robotized)
- Stock control and material storage

By employing the above subsystems, the production planning system has to be informed in real time about the availability of tools in stock as well as about the current contents of the tool magazines of the machine tools (in the case of FASs the robot hands in the end-of-arm-tool magazines); otherwise it will not be able to generate a proper production schedule.

It must be noted that the real-time aspect is important because tools are changed in the magazines of machines (or cells), not only because they wear, but also because different part programs may need different sets of tools. (The actual tool-changing operation is done in most cases by manipulators or by robots. The tool magazine loading/unloading procedure is performed mostly by human operators, sometimes by robots or special-purpose mechanisms, such as a tool shuttle.)

Both the process control and the production planning systems have to update any changes and act in real time; otherwise the operation of the system can be disrupted.

From the FMS/FAS tooling and tool management points of view one must emphasize the links between the CAD system, in which the parts are designed (using design for manufacturing principles), and the computer-aided manufacturing (CAM) system, where the FMS part programs are written. Typically, an FMS part programmer analyzes the CAD output (i.e., the design drawings of the pharmaceuti-

cal products to be manufactured/assembled on the FMS), the fixturing, the different setup (i.e., work-mounting) tasks, as well as the necessary operations, their alternatives, the required tools, and finally a precedence list of the resources (i.e., the possible candidates of processing stations, or cells, or machines).

Real-time databases and software systems are also important, since they provide the reports and status information that are needed for the smooth operation of the FMS (in particular, its dynamic scheduler and other subsystems such as maintenance should be emphasized here) [4, 9–14].

## 3 A FLEXIBLE MANUFACTURING MODEL INTEGRATED WITH DESIGN

The output of the CAM system is a production rule base. This is the knowledge the FMS needs to produce each pharmaceutical product. In this production rule base, among others, tools are assigned to each operation. The tool codes are selected by the FMS process planner or automatically assigned by a process planning system and are obtained from the tool database.

On the basis of the requested tools a list is sent via the network to the tool preparation facility, or station, where the actual tools are prepared (i.e., assembled and preset) and stored in an appropriate way such that the material-handling system of the FMS can pick them up [12–21].

The tool preparation station also deals with other activities, among which the most important are as follows:

- Tool service and maintenance
- Tool assembly to orders (as it is necessary to replace worn tools)
- Tool preset, tool inspection and adjustment
- Real-time tool pickup and tool transportation organized to serve the needs of the real-time FMS

The tool preparation station receives its orders, initially originated by the CAD data processing system, via the FMS network and technically specified by the CAM system in the form of a production rule base. Order data arriving at the tool preparation station include the following:

- Part orders (consisting of part codes and quantities). Note that this is a very important data set for the real-time FMS dynamic scheduler too.
- Notification of when the parts are physically available for FMS processing, representing a due date for tool preparation.
- A priority order (note that this can change because of some real-time changes in the system, and thus this station must be able to cope with this task too).
- The portion of the production rule base describing the requirements regarding tool preparation.

The tool preparation station keeps in touch with the real-time FMS system, as well as with the rest of the system, by feeding back important tooling system-related data:

- Stock reviews (regarding tools)
- FMS status report (regarding tools)
- Part priority status reports (in case dynamic changes must be performed in the FMS which have an effect on tooling needs and tool preparation due dates)

## 4 REAL-TIME OPERATION CONTROL

The real-time part of the FMS operation control and management system must deal with the following tasks:

- It must handle the application of tools for a variety of processes as defined in the production rule base and assigned in real time to the FMS/FAS resources by the dynamic scheduler.
- It must provide data to control the transportation of tools and tool magazines within the FMS.
- It must provide information to perform and supervise tool changes and tool magazine changes at all levels.
- It must be notified of tool inspection results (e.g., if it finds a worn-out tool as a result of an inspection procedure, it must generate a command that instructs the tool magazine update system to change the tool in question in the appropriate tool magazine).
- It must provide information in the case of emergency.
- It must provide the necessary interfaces and data to perform diagnostic/recovery operations, preferably using diagnostic expert systems.

Finally, let us underline an important feedback loop starting at the real-time system and ending at the tool preparation station, which contains the real-time tool status, wear, and part priority information. These data are often useful to those people and/or system software systems that deal with the generation of the production rule base. It is also a very useful data set for FMS designers, since a lot of data which would previously have been lost will be saved in this way.

The most important operation control activities in FMS/FAS identify three levels at which simulation and optimization are required prior to or during FMS/FAS part manufacturing:

1. The factory level or business level handled by the business system of the computer integrated manufacturing (CIM) or, even broader, the enterprise resource management system
2. The FMS off-line level representing scheduling, simulation, and optimization activities prior to loading a batch or a single component on the FMS (handled sometimes by the CAM system, sometimes by the FMS part programming computer)
3. The real-time controlled level handled by the FMS/FAS operation control system, a dynamic scheduler with integrated tool management and multimedia support, representing a situation where the parts are already physically as well as logically in the real-time controlled environment

Due to its complexity, a truly integrated approach is required in designing a production rule base to provide the job description for the FMS dynamic scheduler. This is because the dynamic system relies heavily on the knowledge base as represented by the rule base, and an overly restrictive rule base will lead to inefficient, at times even wrong, decisions. In other words, such a structure should represent all the multilevel interactions and their possible precedence rules that relate to the manufacturing process planning and processing decisions in an FMS. This turns out to be a difficult task.

It should be underlined that the application of multimedia at this level is extremely beneficial in terms of part program preparation, teaching/training operators on setting up parts, fixtures, tools, machines, for troubleshooting, for regular maintenance, at the computer numerical control (CNC) level programming, robot programming, placement machine programming, programmable logic controller (PLC) programming, quality control, maintenance, and other tasks.

Most FMSs have some part-buffering capability. This may be not for scheduling reasons, but for technological, that is, process-planning, reasons (e.g., the part must cool before an accurate inspection procedure is performed). Some level of buffering is useful and necessary because of reliability reasons. (The actual number of buffer store locations should be established on the basis of simulation and experience.)

Cells often have some buffers too. The reason for this is that, by providing a part in the input queue of the cell just before the currently processed part is finished at the particular cell, the cell is kept running at its highest efficiency level, since time is only "wasted" for part changing. The other important point to note is that well-designed part buffers offer a direct access pickup/load facility, making the rescheduling process in the queues short, simple, and dynamic [18, 19, 21–27].

## 5 INNOVATIVE DESIGN

The key objective of this chapter is to describe a generic and systematic pharmaceutical manufacturing/assembly system design method that includes product, process, service systems, and even innovation project management architecture aspects of such systems.

This architecture must be simultaneously novel as well as compliant with set guidelines by the product/process design industry and the PMI (Project Management Institute), following International Organization for Standardization (ISO) 9000:2000 quality standards. Our tested pharmaceutical manufacturing system design solution integrates object-oriented process modeling, requirements and risk analysis, statistical methods, design of experiments, and three-dimensional (3D) interactive multimedia methods and tools which are 100% Web compatible. Furthermore, our methods and software tools are generic in that they can be applied not only to systems such as the pharmaceutical industries or automobile manufacturing but also to processes such as the oil business or services such as education.

A pharmaceutical manufacturing system design requires significant level of innovation. The broadest definition of innovation is the act of introducing something new to a society or community, whether a product or process. This is often confused with invention, which focuses more on specific objects. Within pharmaceuticals innovation can therefore include new business structures within the company,

manufacturing processes and quality control for the medications, and product materials. Process and service improvements can also qualify as innovation, but note that in this case services are usually counted as processes [4, 13, 21, 23, 28–32].

Discoveries such as the charting of new planets, land masses, or forms of life are not classified as innovations as they had existed before being observed by humans. When new species are introduced into a society and find a specific use, it can be classified as innovation. A pharmaceutical example of this is the antibiotic penicillin. Although it had existed as a fungal secretion, it was only within the past century that it was used to actively eliminate infectious bacteria. In initial analysis, however, it was not thought to survive long enough within the human body to be effective.

This brings a vital aspect of innovation, namely the ability to recognize alternative uses for existing processes or tools. This is difficult as unexpected changes within a system are usually labeled as mistakes or anomalies. The development of the Post-It note is an example of this. The original goal was to create a high-strength adhesive, and an extremely weak one was created by accident. Nonetheless, instead of simply disposing of it, the possibilities of this new substance were examined by technicians and managers alike, allowing the use of easily placed reminders for everyday usage. Possessing an ultraweak adhesive allows Post-It notes to be removed without damaging the surfaces that they are placed on, and they are available in a variety of colors and sizes.

The former example is a radical innovation, not only because it allowed significant changes in message reminders and adhesives, but also because it was completely unexpected. The diversification of Post-It notes into different sizes and colors is an example of incremental innovation, which involves step-by-step changes and improvements to existing products or processes. Radical innovations are far less common, though their effects are farther reaching over both society and history.

The general trend through human history has been one of learning to consciously recognize and direct innovation, particularly through combining science and technology. Human societies have often worked with certain processes even without fully understanding their effects or underlying ideas. Metallurgy shows this clearly, as iron, bronze, and gold have been used for millennia before the molecular structure could be seen and analyzed. Note also that, although our ancestors could not describe their chemical composition, these metals served a great many successful purposes. In these cases, the goals of innovation are highly pragmatic, as successful solutions are passed down and taught to future generations.

Those who innovate can therefore learn from working, viable solutions to begin their own practices. Those who continuously work with a fixed set of designs must be willing to experiment, test, and diversify their practices to avoid stratification, as innovation not only allows survival but also encourages prosperity.

The ability to innovate also involves learning from past mistakes, not just one's own. Mistakes and errors in practices can be both costly and dangerous but can be prevented from occurring successively if their causes are determined. This can be difficult because a near-miss scenario can be seen either as an infrequent event or as an averted disaster. The first reaction to this is usually to continue without changing current practices, allowing for similar mistakes to occur. Learning requires all levels of an organization to participate and create channels of communication to innovate effectively, as the inability to share experience denies new opportunities [13, 29–40].

## 6 OPEN INNOVATION ARCHITECTURE

Innovation as a process and the related research-and-development (R&D) project management are considered to be two of the most complex information systems and engineering architectures due to the large number of attributes, processes, and dynamic changes projects go through during their life cycle.

Following our integrated and simultaneously open architectural approach, we look at every innovation process and project as a system built of objects and classes of objects.

Then we look at the way the components of these systems interact with each other. Once we understand these behaviors, we follow our integrated system approach in terms of looking at the project management system as processes, trying to satisfy customer requirements and also representing risks.

We then embed this system model into a statistical analysis and 3D interactive multimedia framework (Figure 1). We use statistical methods to capture processes before they go out of control as well as to perform trend analysis, a great opportunity for innovation, and use 3D interactive multimedia and 3D visualization methods over the Web for communication purposes with global innovation team members. The emphasis on collaboration in today's competitive medical drug field requires these virtual environments to streamline team interaction. (Note that the active

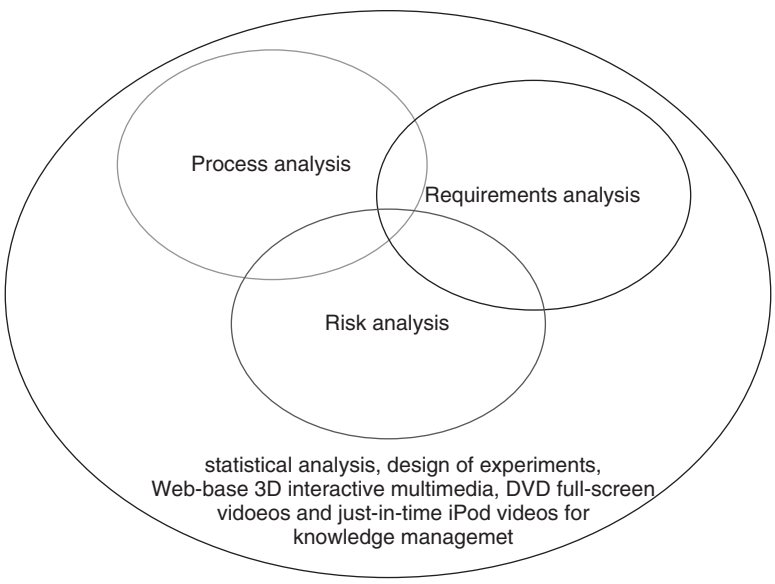


**FIGURE 1** When designing lean and flexible pharmaceutical manufacturing/assembly/packaging systems, one needs to analyze the required processes, customer, user, maintenance, quality, reliability, flexibility, lean, design requirements, and risks involved with any of the listed processes, all in a statistical framework. (Note that our 3D interactive multimedia and simulation framework supports integrated digital design and digital manufacturing system design principles, meaning that one should test all designs and systems first on the screen, and only if everything looks fine, in the real world.)

code spreadsheets and 3D objects referred to in this presentation are all part of Ranky's eLibrary and are available at http://www.cimwareukandusa.com.)

To illustrate the importance of the "openness" of our architecture, consider modern simulation/analysis tools by Parametric Technology Corporation (PTC) (Figure 2), and PLM (product life-cycle management) tools, such as the IBM/Dassault Systemes Delmia tools for pharmaceutical manufacturing system modeling and design (Figure 3) with sensory feedback processing (Figure 4).

Since these models can be designed, edited, run, and driven even over the limits, they can be extremely valuable sources for modeling in the digital domain, process analysis, requirements modeling, risk analysis, and even collecting statistical data and modeling breakdowns of complex systems.

Observe the FEA in Figure 2*a*. This is a torsional test of a pharmaceutical manufacturing machine element on the assembly line of a new medication packaging line. The line is still being tested and improved in the virtual environment, which greatly streamlines the refinement process. As can be seen by the von Mises stress distribution, the sharp edges of the shaft will need to be rounded with a fillet. These would also increase the distribution of the same stress and thus reduce the majority of the red zones (high stress) to blue or even green (low stress). Without using the virtual assembly line to test ideas before for the physical, the unexpected failure of this part could create delays or contamination of product or even harm human operators.

As can be seen, digital pharmaceutical manufacturing/assembly/packaging and factory design tools include not only machines but also advanced sensors, actuators, controls, material-handling systems, labeling machines, and even ergonomically realistic human models and operators performing real-world tasks in extremely realistic model factories. Simulations like these are not just pretty models; they actually save huge investments because the factories are not built until the models are satisfactory. Keep in mind that making changes in a physical factory costs time, money, and possibly production efficiency, even to just check a possible improvement. Virtual models can be simultaneously run thousands of times over a period of days, with hundreds of variables being optimized until the appropriate combination is chosen [30–36, 38, 40–44].

## 7 GENERIC, OBJECT-ORIENTED INNOVATION PROCESS MODELING METHOD AND SAMPLE MODEL

Understanding, modeling, and then following processes, procedures, and best practice reusable processes are essential for every business to stay at the top. The pharmaceutical manufacturing system "innovation business" is not exception.

Major international product/process design standards written and reviewed by thousands of leading researchers and companies around the world always help to create a model for complex problem-solving challenges such as innovation. Therefore this section discusses two of the eight quality management principles of the ISO 9000:2000 international quality standard and the way these rules should be applied to pharmaceutical manufacturing system designs.

We do this for the purpose of developing systematic innovation (with related project modeling skills) and reusable, tested pharmaceutical system design

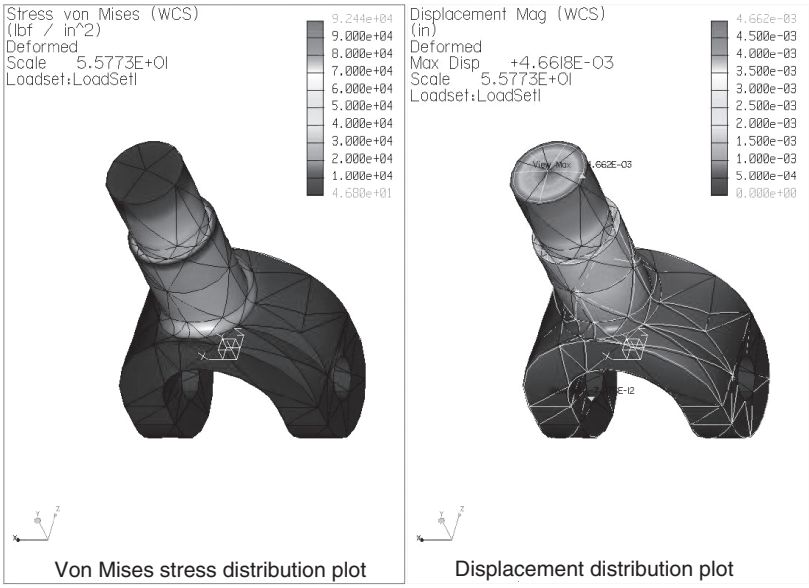


(*a*)

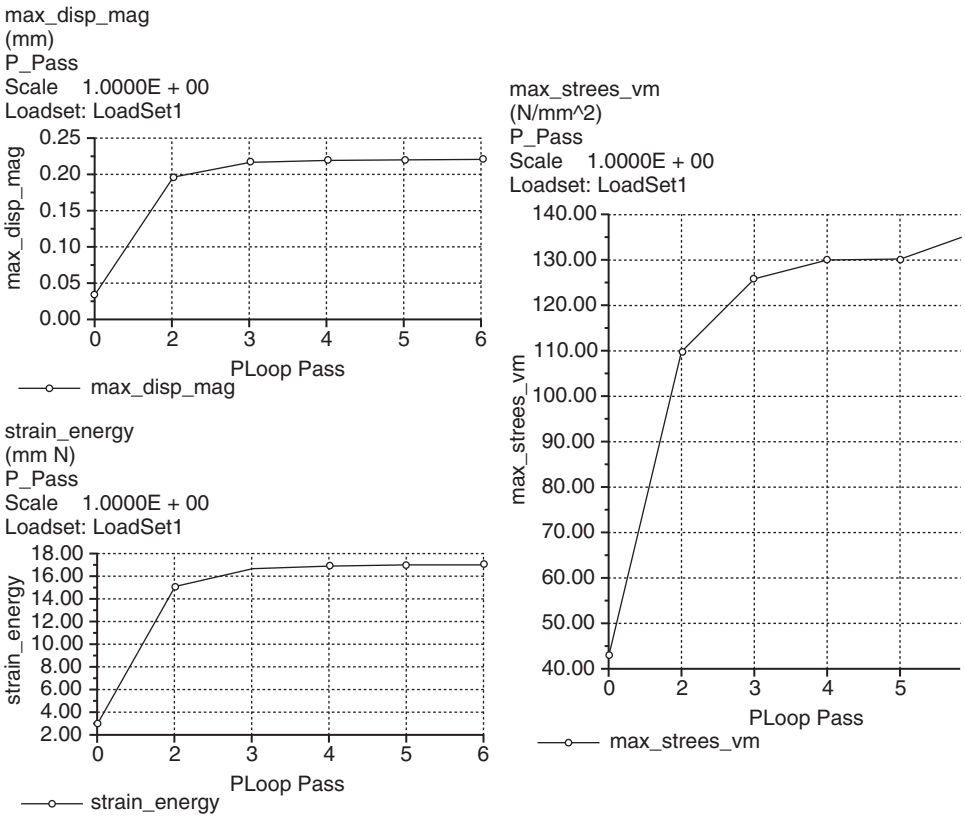


(*b*)

**FIGURE 2** Finite element torsional test of pharmaceutical manufacturing machine element on assembly line of new medication packaging line. The line is still being tested and improved in the virtual environment, which greatly streamlines the refinement process. As can be seen by the von mises stress distribution, the sharp edges of the shaft will need to be rounded with a fillet. These would also increase the distribution of the same stress and thus reduce the majority of the red zones (high stress) to blue or even green (low stress).

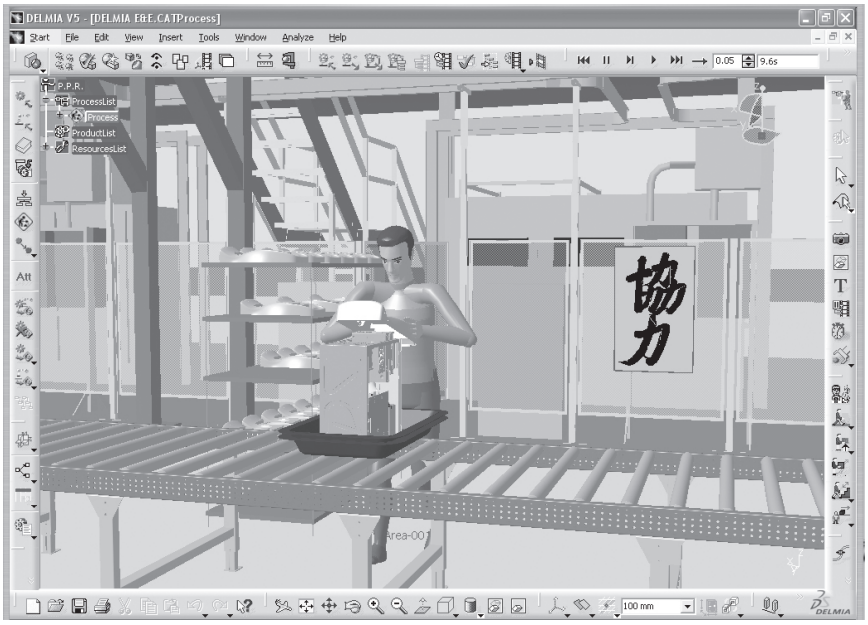


**FIGURE 3** Modern simulation/analysis and PLM tool: IBM/Dassault Systemes Delmia for pharmaceutical manufacturing system modeling and design. The benefits are huge, since the system can be built and tested in the digital domain. (Courtesy of IBM/Dassault Systemes Delmia, Inc.)

processes. We use our own object-oriented process modeling method, called CIMpgr. The nature of this method tansforms into UML models [the Unified Modeling Language of information technology (IT)] and also complies with international process modeling standards used in complex system modeling environments.

First, we will discuss a few important definitions that closely relate to ISO 9000:2000 (quality process modeling) standard principle 4:

- A process, or activity, can be defined as a transfer function with one or more inputs, outputs, controls, and resources that together all enable the variables to gain data and then fire.
- Transfer functions, when fired, create a transformation process. A transformation process in a project is made up of methods, steps, tasks, and various algorithms and processes that acquire and manipulate data and then turn it into system output(s). Note that the input data can describe material, human knowledge, technological standing, fiscal information, and others.
- The output of the process is a product that consists of specific technical and/or social products and services that conform to the sponsor's requirements.
- Processes, in terms of quality project management, have visibility, documentation, and traceability.
- In this context visibility, relates to whether we know and transparently (or graphically) see what methods and techniques, system process steps, and technologies are involved when creating the desired output. Do we know the

**FIGURE 4** Advanced sensors working in pharmaceutical assembly systems help real-time operation control and quality assurance system to test every product. (This is often referred to as the zero-defect policy designed into a system.) The luminescence sensor illustrated will detect a wide variety of invisible targets. This STEALTH-UV sensor was designed to sense the presence of invisible fluorescent materials contained in or added to many products. Users can detect the most difficult targets, including clear tamper-proof seals, clear labels, and invisible registration marks. This unique sensor is also ideal for solving many of today's toughest problems in product orientation, inspection, and verification. (Courtesy of TRI-TRONICS Co., Inc., www.ttco.com.)

sequence of these steps and the possible parallel process relationships? How does one process affect the other?

- Documentation means that the methods, steps, processes, and technologies are well specified and recorded according to agreed-upon standard specifications.
- Traceability means that the process steps as well as the output(s) can be traced back to actual customer requirements.
- Process capability can be defined as the ability of the production process to meet certain specifications and tolerances.
- Process discrepancy is the deviation of process settings from specifications.
- Process variability is the variation in dimensional or other measurable characteristics of output from a production process. (Note that in any project the ultimate goal is to stay within the predefined limits of process variability and, if possible and feasible, to reduce process variability, because this typically reduces risk too.)
- Variability can be expressed in terms of average range of standard deviation.
- A process variable is a process parameter that fluctuates in the manner of a random variable and hence requires surveillance.
- Process management means getting the activities and procedures that highly skilled and experienced managers carry in their heads into the open by means of a well-documented model, often referred to as the process model [40–46].

## 8 SYSTEMS APPROACH TO PHARMACEUTICAL MANUFACTURING SYSTEMS MANAGEMENT

Identifying, understanding, and managing interrelated processes as a system contribute to the organization's effectiveness and efficiency in achieving its objectives. (Note that every one of the key drivers listed below embed one or more innovation opportunities!)

Key drivers and achievable gains include the following:

- Processes that will achieve the desired results will become better integrated and aligned.
- Management and process owners will have the ability to focus their efforts on the key processes.
- Since the consistency, effectiveness, and efficiency of the organization will grow, the confidence of interested parties and collaborators in the organization will grow too.
- System structuring and fine-tuning will become possible to achieve the organization's objectives in the most effective and efficient way.
- Understanding the interdependencies between the processes of the system will yield good results.
- Structured (and object/component-oriented) modeling approaches that harmonize and integrate processes will become reality. Employees will understand them, follow them, and therefore reduce waste and increase quality in every process.
- The resistance created by cross-functional barriers will be reduced, providing a better understanding of the roles and responsibilities necessary for achieving common objectives.
- Organizational capabilities and the establishment of resource constraints prior to action will be better understood by all involved (and mostly by all those who have created the models).
- Targeting and defining how specific activities within a system should operate will become reality.
- Continually improving the system through measurement and feedback-controlled evaluation becomes possible due to the analytical and quantifiable approach of the process models. (Note that at its ultimate level this will lead to a real-time, feedback-controlled enterprise capable of reacting to dynamically changing market needs.)

After this introduction, let us show our object-oriented system components, following the above described ISO 9000:2000 principles, and how we can model complex innovation and related project management processes using them [40–54].

As a simple example, consider, that you are packaging a pharmaceutical product using a line that performs various process steps. Figure 5*a* illustrates one of these steps. It has input(s), output(s), control(s), and resource(s). These data

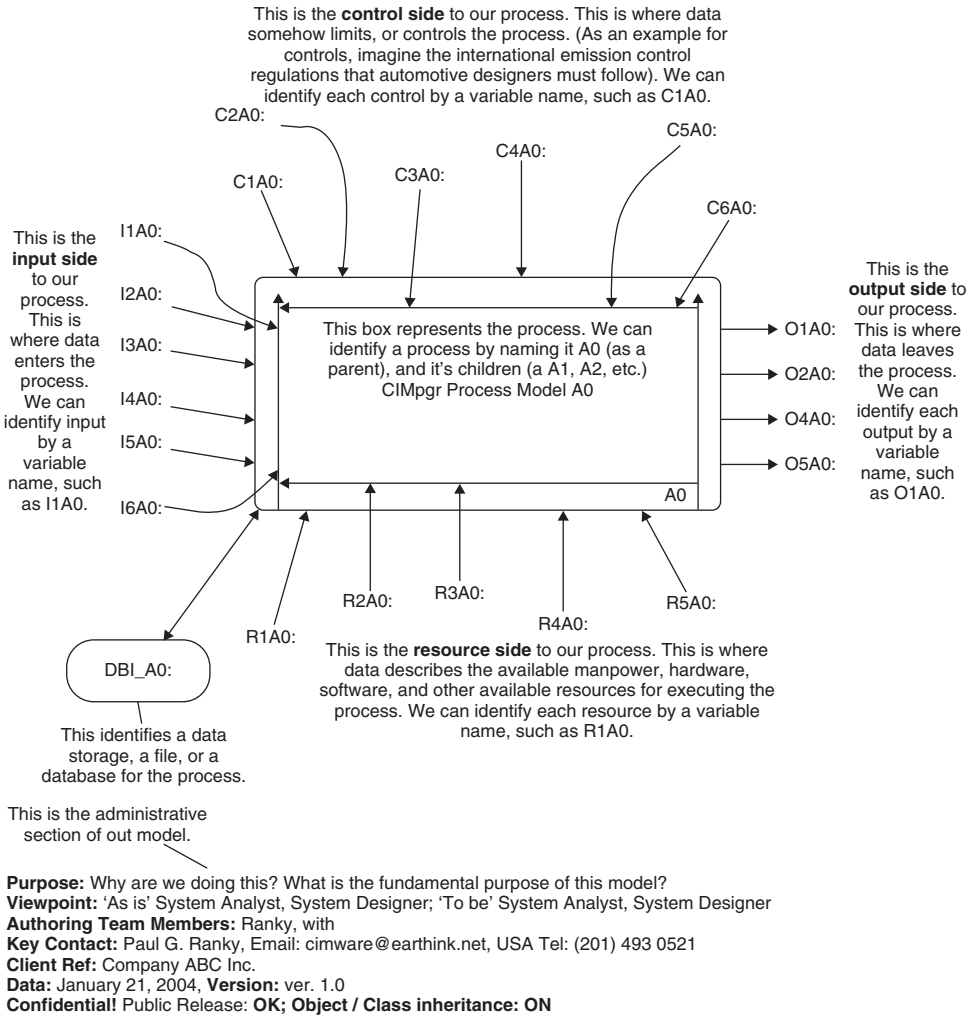


(*a*)

**FIGURE 5** Object-oriented process modeling method (CIMpgr) as applicable to pharmaceutical manufacturing/assembly/packaging system design.

types help to identify under what conditions the process should be executed by the pharmaceutical manufacturing system. (Definitions are offered in the diagrams.) We can also see the way the CIMpgr process maps into a UML diagram. This is important, since UML is the modeling language of the IT professionals who will program the PLCs and control systems for the lines. Figure 5*b* shows how the CIMpgr process maps into a UML diagram. We can see in Figure 5*a* how multiple processes have to interact as we design a pharmaceutical assembly system.

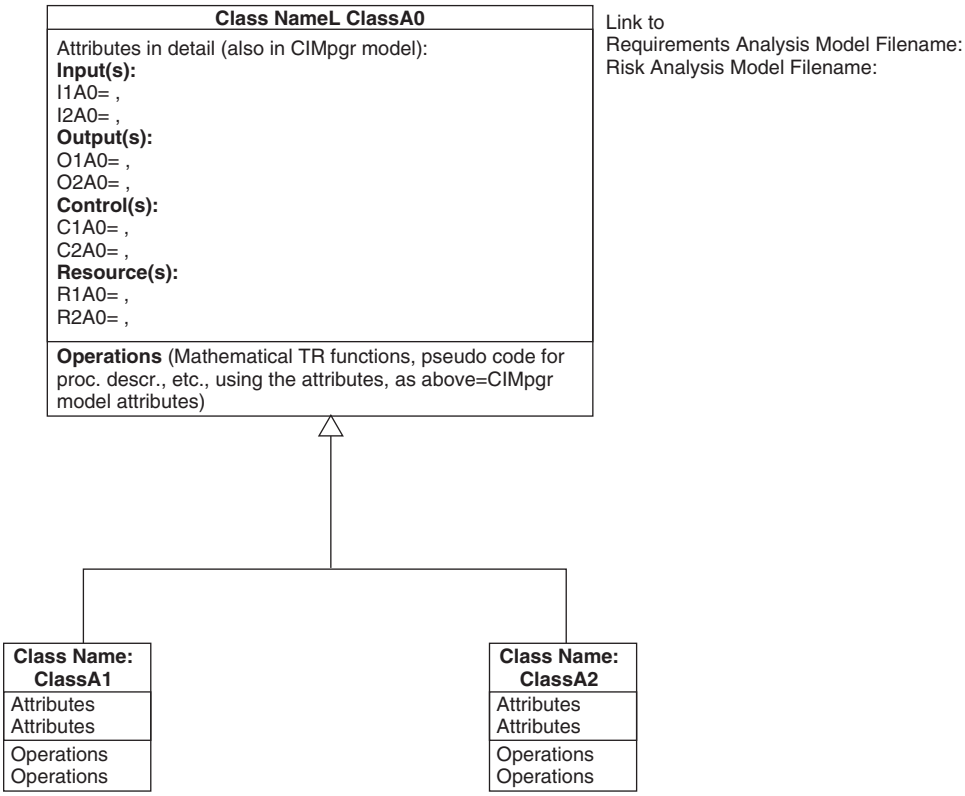


(*b*)

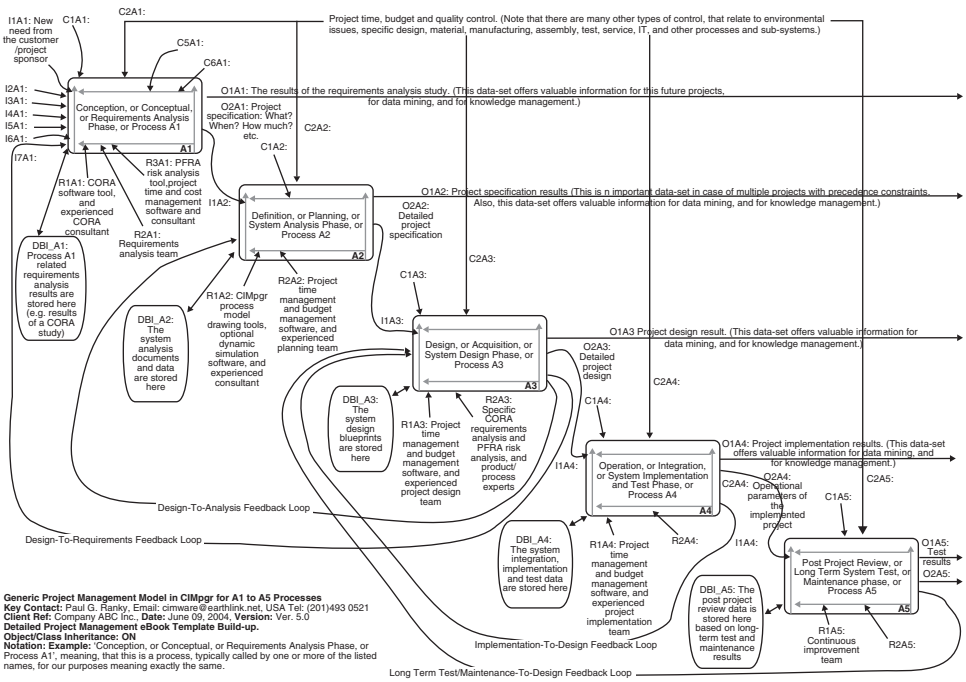


(*c*)

**FIGURE 5** *Continued*

## 9 REQUIREMENTS ANALYSIS FOR SYSTEM PRODUCT, PROCESS, AND SERVICE DESIGN INNOVATION

Processes in a successful innovation project must satisfy requirements set by the market, the sponsors, and/or the inventor's own dreams. Requirements analysis is considered to be one of the most important features of any innovative pharmaceutical manufacturing system project, because if done professionally, it helps to specify, research, and develop appropriate features and processes that customers need.

In our innovation project examples we have focused on generic needs and requirements, and our associated "customers" are the pharmaceutical R&D team members, managers, and operators in various industries.

In terms of our research approach, we have followed a proven method: Analyze the needs and the requirements, the demonstrated processes, and the methods and systems they try to or have to satisfy, and if you find a "gap", you have found an innovation opportunity. Note that when we search for this gap, it will simultaneously appear as a missing process in our CIMpgr model or as an existing process but missing attributes as well as a requirement in our CORA model (component-oriented requirements analysis) model:

- Analyze the actual methods presented. Find the core methodologies, the mathematical models, and the underlying engineering and/or other science foundation.
- Analyze the technologies involved. (How is science turned into a practical solution/engineering and/or computing technology?) Is there a need for a new, novel technology that has not been invented yet or applied in this field?
- Analyze and review the actual processes and the way the process flow is integrated. (Follow an object-oriented process analysis method, i.e., from concept to product.) Focus on the attributes of the processes. Note that by adding a new attribute you create new data types, with new information, and if your process can reason over this in a new way, new knowledge; therefore your combined CIMpgr and UML model becomes a new knowledge representation model too. This is important because innovation is formalized this way and can be communicated among global teams.
- Analyze potential alternative solutions. (A pharmaceutical manufacturing/assembly/packaging system must be very flexible these days, due to dynamically changing customer requirements and even operating conditions.)
- Analyze the benefits and the disadvantages of each process/solution.
- Design alternative methods, processes based on what you have experienced/seen and learned.
- Design an integrated system, based on what you have analyzed in this case.
- Work in a multidisciplinary team and exchange ideas.
- Understand the boundaries as well as the tremendous potential of new ideas and developments by working on this case (realize that in order to survive and win, you must add value) [54–57].

After this short introduction, we demonstrate our CORA spreadsheet solution with a real-world example (Figure 6).

**Object / Component Oriented Requirements Analysis Program for Networked Lean Manufacturing** by Paul G. Ranky © 1992–2006

© Paul G. Ranky, 2000–2002

| Engineering / Software Solutions Responding to Customer Requirements / Lean Manufacturing Manager: Customer Requirements ( Reflecting component object behavior related to customer needs) | | Importance Rating (1–5) | Fieldbus Network | Profibus Network | DeviceNet Network | Ethernet | Graphical CNC Progr. | On-site maint. Support | Off-site maint. Support | On-site Redundant Server | Off-site Redundant Server | Link to Factory prod. contr. | Link to Factory TQM/TQC | 2D videos/3D multimedia | Cell-based web Camera | PC-based CNC Controller | PC based workkstation contr. | PC based Cell controller |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| S.No | Describe the Requirement | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 |
| 1 | Reliability of data transfer for real-time access should be high | 5 | 9 | 9 | 9 | 9 | | | | 9 | 3 | | | | | 9 | 9 | 9 |
| 2 | Reliability of reporting process failure to the line manager's computer | 4 | 9 | 9 | 9 | 9 | | | | | | | | | | 9 | 9 | 9 |
| 3 | Ease of integration into a system (plug and play networking): important! | 4 | 9 | 9 | 9 | 9 | | | | | | | | | | 9 | 9 | 9 |
| 4 | Ease of machine programming (CNC machining / inspection) | 3 | | | | 9 | 9 | 9 | | | | | | | | 9 | 3 | 3 |
| 5 | Ease of changng CNC part programs (locally, and via the network) | 4 | 9 | 9 | 9 | 9 | 9 | 3 | 9 | | | | | 9 | 9 | 9 | 9 | 9 |
| 6 | Ease of adding new sensors to a workstation, CNC, or cell: high | 3 | 9 | 9 | 9 | 9 | | | | | | | | | | 9 | 9 | 9 |
| 7 | safety of operation: critical! | 5 | 9 | 9 | 9 | 9 | | 9 | 3 | 9 | 3 | | | | | 9 | 9 | 9 |
| 8 | Cost of change/extenaion/system expansion should be low | 4 | 3 | 3 | 3 | 3 | | 9 | 3 | 3 | 3 | | | | | 9 | 3 | 3 |
| 9 | Operator training needs and costs should be low | 3 | | | | | 9 | 9 | 3 | | | | | | | 9 | 3 | 3 |
| 10 | Network installation complexity and cost should be low | 3 | 9 | 9 | 9 | 9 | | | 3 | 3 | | | | | | 9 | 3 | 3 |

(*a*)

| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 20 | The system history database should be on the network | 5 | 3 | 3 | 3 | 3 | | | 3 | 3 | 9 | 9 | 9 | 3 | 3 | 9 | | 9 | | | | | | 4 | 0 | 0 |
| | **Target Values** (List here the parameters that specify engineering solutions accurately. If you don't know the range of the acceptable values, use our Taguchi Calculator Program for Designing an Experiment) | | Within 27 m sec | Within 27 m sec | Within 27 m sec | Within 27 m sec | GUI, Iconized, multimedia | Less than 3 minutes | Less than 24 hsr | 0 sec switch | Less than 30 sec switch | Refresh every 2 minutes | Refresh every 2 minutes | 320, 240 pxels or better | 320, 240 pxels or better | Win, Linux, Solans, or OSX | Response within 12 m sec | Response within 24 m sec | | | | | | | | |
| | **Absolute Importance Rating** | | 416 | 416 | 416 | 489 | 207 | 147 | 111 | 186 | 107 | 144 | 135 | 87 | 87 | 567 | 381 | 588 | 0 | 0 | 0 | 0 | | | | |
| | **Relative Importance Rating** | | 9.3 | 9.3 | 9.3 | 11 | 4.6 | 3.3 | 2.5 | 4.1 | 2.4 | 3.2 | 3 | 1.9 | 1.9 | 13 | 8.5 | 13 | 0 | 0 | 0 | 0 | | | | |
| | **Enter Our Product ratings (1 = low, 5 = High)** | | 4 | 5 | 5 | 4 | 5 | 3 | 4 | 3 | 4 | 5 | 5 | 5 | 4 | 5 | 4 | 5 | | | | | | | | |
| | **Enter Competitor A's ratings (1-5); Graph** | | 3 | 4 | 4 | 5 | 4 | 3 | 2 | 2 | 3 | 3 | 4 | 4 | 3 | 4 | 4 | 4 | | | | | | | | |
| | **Enter Competitor B's ratings (1-5); Graph** | | 4 | 4 | 4 | 4 | 4 | 5 | 4 | 4 | 3 | 2 | 3 | 2 | 3 | 4 | 3 | 4 | | | | | | | | |
| | **Enter Competitor C's ratings (1-5); Graph** | | 3 | 2 | 3 | 3 | 2 | 2 | 1 | 1 | 2 | 3 | 4 | 3 | 4 | 3 | 2 | 3 | | | | | | | | |

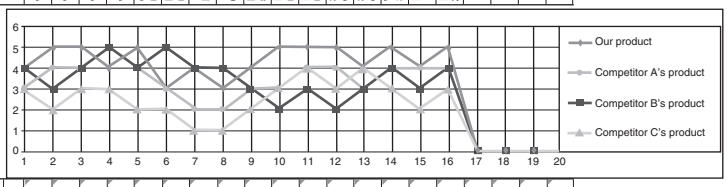


(*b*)

**FIGURE 6** CORA method. This is a spreadsheet-based tool, designed to analyze customer requirements. (A "customer" here can mean a pharmaceutical manufacturing line vendor, user, operators, maintenance engineers, and many others.) The key approach is that we create a correlation matrix and then evaluate the results using a quantitative, computational approach. This is much more accurate than just a simple structured list. Our method offers a list of all key requirements as well as the priorities for the pharmaceutical manufacturing system design team they should follow during the design process. (For more about this software tool, see http://www.cimwareukandusa.com.

## 10 INNOVATION RISK ANALYSIS AND OPPORTUNITY METHOD AND TOOL WITH PHARMACEUTICAL MANUFACTURING SYSTEM APPLICATIONS

Our failure risk analysis and opportunity method and iterative software tool, as part of our New Product & Process Innovation (NPPI) Tool Library, promotes systematic collaboration and team-oriented engineering thinking when a new pharmaceutical manufacturing system process and/or product are developed. (We call it "opportunity method" too, since most risks, if not all, offer new opportunities for innovation.) It is based on our generic process failure risk analysis method that could be applied to literally any process that involves risk—and innovation is a very risky process.

We follow a rule-based method when we analyze risk objects and components and their attributes. These plug-and-play rules can be different for different subjects, research areas, and industries. They can be designed and standardized for different industry sectors, enabling an analytical approach, systematic standardization, and accurate and predictable results.

Our risk analysis method and tools help the engineering management team to understand some of the following concerns:

- What could go wrong with the processes involved during the innovation project?
- How badly might it go wrong and what could the financial loss be?
- Which are the highest risk processes/operations when working on the product/process/service-related innovative design and project?
- What needs to be done to prevent failures?
- Which processes must be changed to reduce the risk of failure?
- What tools and fixtures are required to prevent failures and reduce the risk?
- What education is needed for participants, innovators, engineers, and process owners, such as line management and operators, to reduce or prevent failures?

After this introduction, we show the risk analysis system components, following the already described ISO 9000:2000 principles, and how we can model complex project management risks using them (Figure 7) (note that the active code spreadsheets and 3D objects are part of Ranky's eLibrary) [50–55].

## 11 OPEN-SOURCE COMPUTATIONAL STATISTICAL AND THREE-DIMENSIONAL MULTIMEDIA FOR PHARMACEUTICAL MANUFACTURING SYSTEM INNOVATION AND PROJECT COMMUNICATION

Since we follow an analytical, quantitative, and open-source computational approach, our pharmaceutical product/process and project management method and software toolset are implemented as (Internet browser readable) MS-Excel spreadsheets, integrated with several hyperlinks to the rule base and to optional 2D video and 3D virtual-reality objects for visualization.

| | | |
|---|---|---|
| Ranky 111601/DFRA_Ver.5 | **Disassembly Process Code** | Ranky PC DisassyCode: 05/07/97 |
| 11/16/01 | **Engineering Rerease Date or Process Methodology** | 5/7/97 |
| 9/19/01 | **Type of Product Disassembled** | Electro-mechanical |
| Rev.2.1.3. by Ranky | **Product Group Classifier** | Desktop PC |
| | **Engineering Release Date of the Product** | Estimated: 1993 |

| List/Identify the Parts/Components Retrieved in Each Disassembly Process Step | Process Time | Process Cost | Accumulated Process Cost | The DFRA Team Describes/Illustrates the Potential Disassembly Failure Mode and the Effect; the Risk of Failure | |
|---|---|---|---|---|---|
| | | 16.40 | | Proc. ID | Failure Mode(s) and Effect(s) |
| | [sec] | [USD] | [USD] | | |
| Painted metal PC cover (File: 3DMetalCover. mov) | 45 | 0.21 | 0.21 | ID 1.1 | Metal Cover scratched by slipped screwdriver |
| | | | | ID 1.2 | As PC Metal Cover is removed, internal parts are cratched |
| | | | | ID 1.3 | |
| Floppy drive, hard drive, mounted in a solid sheet metal bracket inside the PC (File: 3DFloppyHDassy. mov) | 137 | 0.62 | 0.83 | ID 2.1 | Floppy drive assy. screw removal can damage nother board |
| | | | | ID 2.2 | |
| | | | | ID 2.3 | Movie, illustrating floppy/HD assy. removal risks |
| Floppy drive (File: 3DFloppyDrive. mov) | 35 | 0.16 | 0.99 | ID 3.1 | Can damage Floppy Drive if assy. Is dropped |
| | | | | ID 3.2 | Can damage Hard Drive if assembly is dropped |
| | | | | ID 3.3 | |
| Hard drive (File: 3DHDobj. mov) | 65 | 0.30 | 0.28 | ID 4.1 | Can damage Hard Drive if assembly is dropped |
| | | | | ID 4.2 | |
| | | | | ID 4.3 | |
| Metal bracket holding the floppy and the hard drive (File: 3DFloppyHDBracket. mov) | 12 | 0.05 | 1.34 | ID 5.1 | |
| | | | | ID 5.2 | |
| | | | | ID 5.3 | |

(*a*)

| | | | |
|---|---|---|---|
| | Ranky PC DisassyCode: 05/07/97 | This DFRA Study Prepared By | Paul |
| ss | 5/7/97 | DFRA Team | NJIT |
| | Electro-mechanical | Responsible Organization/Department | NJIT |
| | Desktop PC | Comments | |
| oduct | Estimated: 1993 | | |

| ed ost | The DFRA Team Describes/Illustrates the Potential Disassembly Failure Mode and the Effect; the Risk of Failure | | Severity Rating | Detection Rating | Occurrence Rating | RPN Pr Nu |
|---|---|---|---|---|---|---|
| | Proc.ID | Failure Mode(s) and Effect(s) | | | | |
| | | | (1-10) | (1-10) | (1-10) | |
| | ID 1.1 | Metal Cover scratched by slipped screwdriver | 3 | 2 | 3 | |
| | ID 1.2 | As PC Metal Cover is removed, internal parts are cratched | 5 | 4 | 1 | |
| | ID 1.3 | | | | | |
| | ID 2.1 | Floppy drive assy, screw removal can damage mother board | 5 | 9 | 2 | |
| | ID 2.2 | | | | | |
| | ID 2.3 | Movie, illustrating floppy/HD assy. removal risks | | | | |
| | ID 3.1 | Can damage Floppy Drive if assy. Is dropped | 8 | 2 | 1 | |
| | ID 3.2 | Can damage Hard Drive if assembly is dropped | 8 | 2 | 3 | |
| | ID 3.3 | | | | | |
| | ID 4.1 | Can damage Hard Drive if assembly is dropped | 9 | 2 | 3 | |
| | ID 4.2 | | | | | |
| | ID 4.3 | | | | | |
| | ID 5.1 | | | | | |
| | ID 5.2 | | | | | |
| | ID 5.3 | | | | | |

(*b*)

**FIGURE 7** The process failure risk analysis (PFRA) tool is an analytical and computational tool using rule bases for evaluating process risks. It is an ideal method and tool for reducing costly failures. (For more about this software tool, see http://www.cimwareukandusa.com.

The reason for this is because we would like to offer our users the opportunity not just to understand the method and the coded logic, but also to be able to enjoy the 3D interactive graphics, the digital videos, the color images, and most importantly the active code spreadsheets. Along with any other imaginable visualization, this can be executed and experimented with using their own data.

In terms of statistical methods, our NPPI Tool Library has several statistical analysis tools to capture innovation opportunities at processes that are likely to drift and become out of control or processes that execute with random failure.

| | |
|---|---|
| **This DFRA Study Prepared By** | **Paul G Ranky NJIT/MERC** |
| **OFRA Team** | **NJIT/MERC CFRA Team** |
| **Responsible Organization/Department** | **NJIT/MERC** |
| **Comments** | |
| | |

| Severity Rating | Detection Rating | Occurrence Rating | RPN (Risk Priority Number) | Max. RPN | Tooling Factor | Clamping/ Fixturing Factor | Skill Factor | Any Other Factor You Define | Accumulated RPN | Risk Associated | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| (1-10) | (1-10) | (1-10) | | | 0.1-2.1=100% | 0.1-2.1=100% | 0.1-2.1=100% | 0.1-2.1=100% | | | |
| 3 | 2 | 3 | 18 | | | | | | | | |
| 5 | 4 | 1 | 20 | 20 | 1.40 | 1.00 | 1.20 | 1.00 | 33.60 | Low | |
| | | | 0 | | | | | | | | |
| 5 | 9 | 2 | 90 | | | | | | | | Mc |
| | | | 0 | 90 | 1.60 | 1.40 | 1.40 | 1.00 | 282.24 | HIGH | |
| | | | 0 | | | | | | | | |
| 8 | 2 | 1 | 16 | | | | | | | | Ha |
| 8 | 2 | 3 | 48 | 48 | 1.20 | 1.20 | 1.40 | 1.00 | 96.77 | Low | |
| | | | 0 | | | | | | | | |
| 9 | 2 | 3 | 54 | | | | | | | | |
| | | | 0 | 54 | 1.40 | 1.20 | 1.40 | 1.00 | 127.01 | Low | |
| | | | 0 | | | | | | | | |
| | | | 0 | | | | | | | | |
| | | | 0 | 0 | 1.00 | 1.00 | 1.00 | 1.00 | 0.00 | Low | |
| | | | 0 | | | | | | | | |

(*c*)

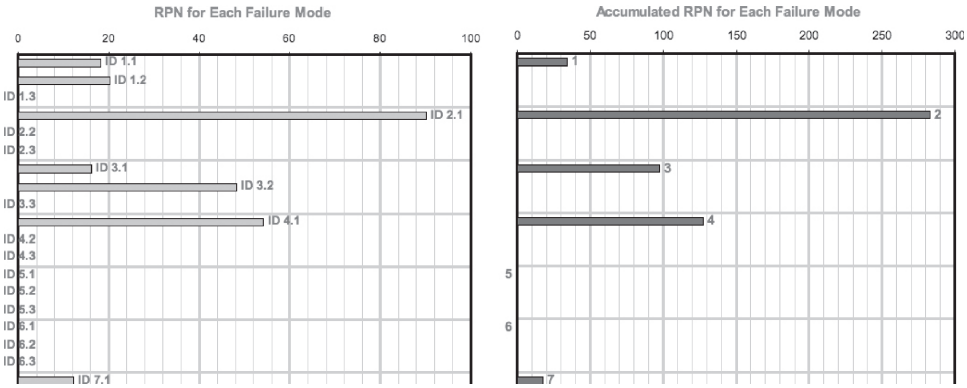


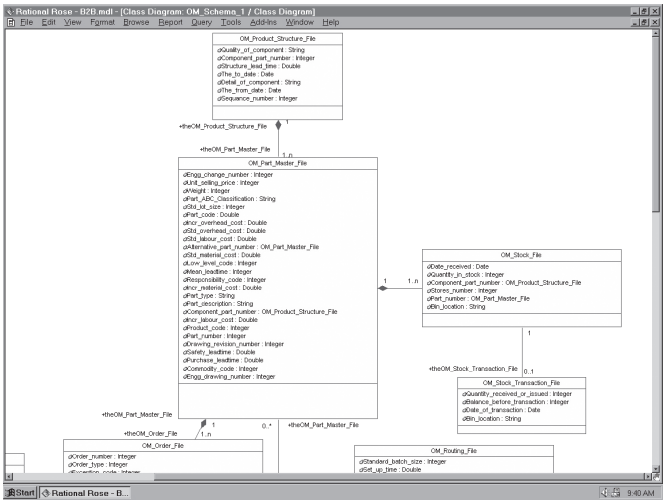


(*d*)

**FIGURE 7** *Continued*

For capturing such critical opportunities for innovation and process improvement, we use a range of control charts for drifting data analysis, Taguchi DOE (design of experiments) methods for developing the desired list of parameters for our engineering solutions in our requirements analysis method, and Weibull methods for process reliability analysis. As we progress, we plan to introduce further statistical and other tools to our NPPI Tool Library [56–70].

## 12 RFID APPLICATIONS

Radio-frequency identification (RFID) technologies are being adopted in the United States at a fast pace in pharmaceutical/assembly and packaging, in general manufacturing, warehousing, distribution, and global supply chain management. The market size for this technology is expected to rise from around $500 million in 2005 to about $4 billion in 2010. In this section we outline some of the main application areas with a focus on the pharmaceutical applications.

We also deal with the R&D opportunities and some digital pharmaceutical manufacturing systems with RFID information system modeling results. Furthermore we offer a generic factory assembly and tracking digital model for RFID integration, the most complicated task manufacturing systems engineers, industrial engineers, and IT experts have faced due to the mixed real-time as well as global traceability and messaging challenges one faces with RFID-tagged parts and shipments. RFID opportunities are great since with the appropriate IT infrastructure they help both major distributors and manufacturers as well as other logistics operations, such as in the health care system, defense industries, and others, dealing with complex, global supply chains in which products and product shipments must be traced and identified in a noncontact, wireless fashion using a computer network, because of cost, security, or safety or because parts are subject to corrosion or medicine is subject to quality degradation.

All of these requirements point to an automated, wireless-readable sensory-based identification method and network that offers more functionalities and is significantly "smarter" than the well-known bar code or the unified product code (UPC).

RFIDs are available as passive and/or active radio read/write sensor packages with active read (and often write) capabilities in relatively large areas (e.g., a large distribution center warehouse or a containership), all performed automatically, supervised by computers, and communicated in a wireless fashion over secure intranets. The attraction to a pharmaceutical assembly factory or a supply chain manager is that when the RFID network is integrated with the factories' material resource IT management systems, accurate information can be obtained on all tagged parts in close to real time throughout the entire supply chain. This can include the globally distributed factories as well as information about parts and assemblies during shipment, including in transit. This is why RFID represents great research and technology as well as huge business opportunities.

We introduce here some of the most important engineering and information systems management principles and challenges that RFID researchers, implementers, and users should keep in mind when developing such systems and/or planning

for such applications as well as offer an RFID digital factory integration model in UML [60–64, 70, 73].

## 13 RFID EXAMPLES

To set the scene, consider a large storage house for a variety of medications or their distribution center with thousands of boxes, parts, and assemblies that range from low cost to high value, on occasion even highly sensitive technology or perishable drugs that must be kept in certain environmental conditions, such as temperature, humidity, or pressure, for the entire period of the shipment and/or production/packaging operations.

Just-in-time delivery in an environment like this means that in order to build an order with variety or a new combination of treatments in a medical drug, every component must be in place on time and in good condition, which is a very difficult criterion to satisfy.

Obviously supply chains are global these days, and shipments are typically made by a variety of means, including cargo ships, air, rail, and trucks; all of these can be late or can get in trouble because of the weather, traffic, industrial disputes, or other reasons. Supply chain systems like this are very complex, because of the uncertainties in deliveries, parts and shipments are lost and/or stolen, goods get damaged during shipment, or the number of international ports and customs often take unpredictable time to check shipments with different levels of safety/security, and many other reasons.

There are many valid reasons why wireless, computer-networked, sensory-based part identification methods, tools, and technologies are being researched and deployed in industry. The application fields and opportunities are vast. The key driver is that even in chaotic, largely distributed, more stochastic than deterministic business environments, adaptive organizations and enterprises must react to demands quickly, else a competitor will take the business. Therefore they must reduce waste and improve efficiency at all fronts. The most important aspect of this strategy is to know exactly what parts they have in stock, exactly where these parts are, and in what condition/state of assembly or preparedness they are. Furthermore, major distributors dealing with complex, global supply chains must be able to trace their shipments in detail because of cost, security, safety, quality degradation (as in the case of temperature-, humidity-, and/or shock-sensitive components or drugs), or other reasons.

RFID technologies with the appropriate IT infrastructure help major distributors and manufacturers as well as other logistics operations such as the health care system, defense industries, and others deal with complex, global supply chains in which products and product shipments must be traced and identified in a noncontact, wireless fashion using a computer network.

All of the above-listed requirements point to an automated, wireless-readable, sensory-based identification method and network that offer more functionalities, and are significantly "smarter" than the well-known bar code or the UPC—hence the new popularity of RFID technology.

RFID tags carry a serialized tag data construct. As an example, a 64-bit class 0 tag offered by a supplier includes 64 bits of total user memory on the tag itself,

including a unique serial number. This number is encoded by the manufacturer and uniquely identifies up to 264 = 18,446,744,073,709,551,616 tagged items.

RFIDs are available as passive and/or active radio read/write sensor packages with active read (and often write) capabilities in relatively large areas (like a large distribution center warehouse or a containership), all performed automatically, supervised by computers, and communicated in a wireless fashion. The attraction to an assembly factory or a supply chain manager here is that when the RFID network is integrated with the factories' material resource IT management systems, accurate information can be obtained on all tagged parts at close to real time, all throughout the entire supply chain. This can include the globally distributed factories as well as information about parts and assemblies in shipment/in transit.

This is why RFID represents excellent research, technology, as well as big business opportunities. To illustrate this, consider research challenges such as remotely scanning and tracing products and parts in boxes on a cargo ship as it approaches national waters from international waters; tracing parts that are subject to corrosion and being used in agricultural or military equipment; medical drugs that are being counterfeited and repackaged and then shipped and imported illegally; or laptop computers that are dropped and damaged by accident. As a clear sign of the business opportunity, consider that according to a U.S. Department of Defense published presentation, RFID-enabled supply chain savings reached over U.S. $460 million in 2004 and the projections for 2010 are in excess of $4 billion!

## 14 RFID SYSTEM INTEGRATION MODELS FOR DIGITAL PHARMACEUTICAL MANUFACTURING AND ASSEMBLY SUPPLY CHAINS

In the U.S. manufacturing and assembly industry, many of the RFID pilot projects focus on achieving 100% read rates at speeds set by the widely used bar code technology. The focus for these projects is to achieve proper tag placement on cases and pallets as well as the proper configuration of pallets to enable 100% RFID tag read rates. This is a huge issue in pharmaceutical manufacturing and assembly in the fight to eliminate fake products and packages reaching the market! (See Figures 8–15.)

Based on the above-described requirements analysis, network planning, and server balancing reasoning, for our purposes, we have decided to follow a simple but powerful network architecture. In this architecture, we have included subnetworks. In terms of the way OPNET IT-Guru handles subnetworks, a subnetwork contains other network objects and abstracts them into a single object. A subnetwork can encompass a set of nodes and links to represent a physical grouping of objects (this can be a local-area network of CNC machines or robot PC controllers) or it can contain other subnetworks (e.g., including the material-handling system control of the line) [32, 34, 63, 66, 69–77].

Subnetworks within other subnetworks form the hierarchy of the network model. This hierarchy can then be extended as required to model the structure of the network. A subnetwork is considered the parent of the objects inside of it, and the objects are the children of the subnetwork. The highest level subnetwork in the network hierarchy does not have a parent, and therefore it is the top subnetwork,

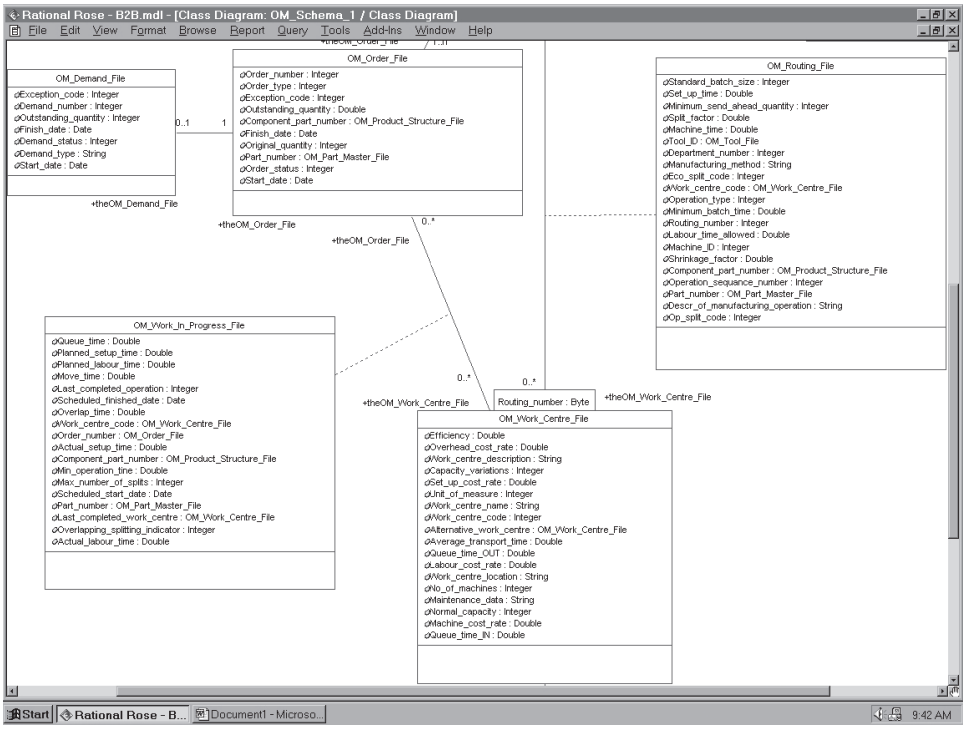


**FIGURE 8** UML model segment illustrating the way the stock file is integrated with the routing and tooling files, assuming that all parts and all tools are RFID tagged. UML models like this should be used prior to any implementation work to assess requirements, technology needs, and RFID integration challenges with the rest of the factory's IT infrastructure.

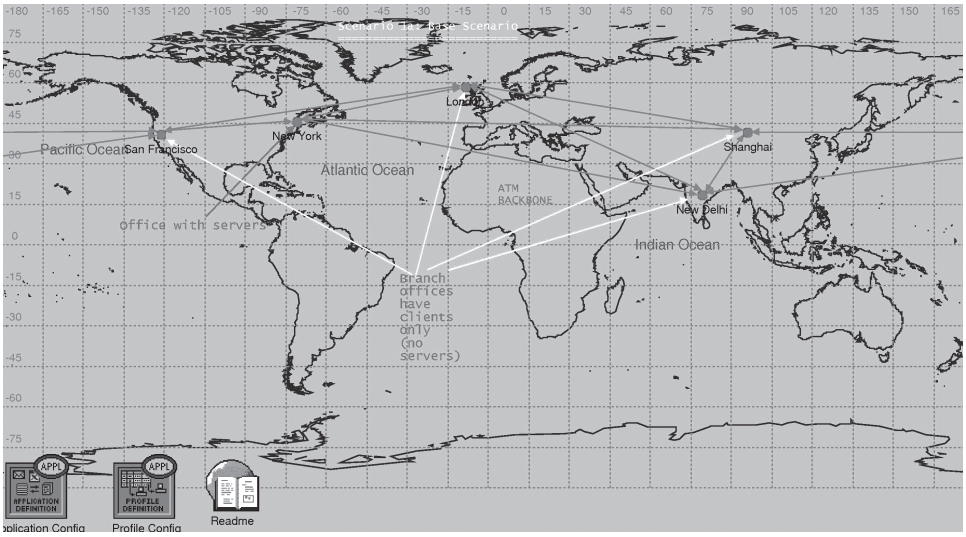


**FIGURE 9** Simulation network for distributed pharmaceutical manufacturing systems and their warehouses in U.S., Europe, India, and Asia. Model focuses on information and data management, the way the servers can cope with the task of tracking pharmaceutical product, and RFID data on a world wide basis. As a modeling tool we use OPNET, a professional network simulation tool.

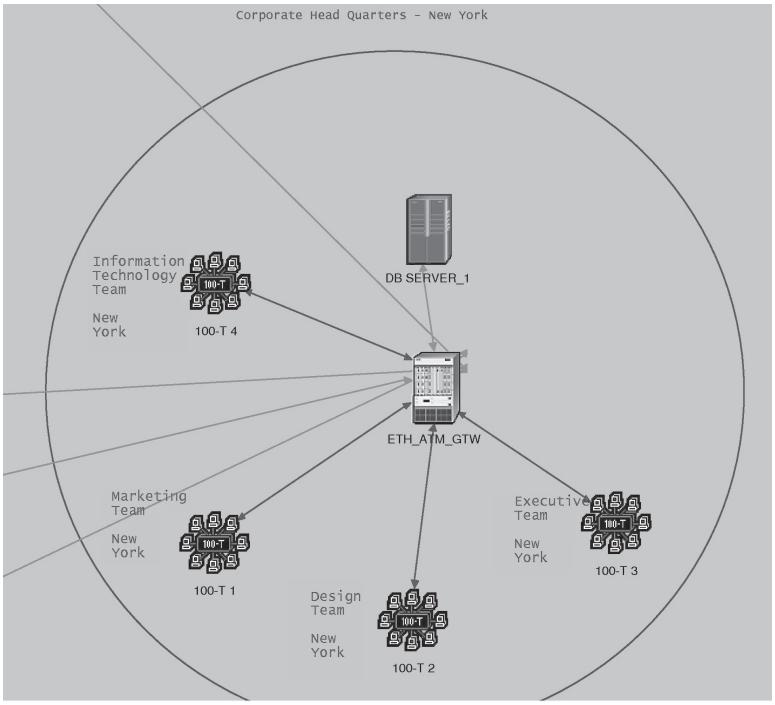


**FIGURE 10** Segment of simulation model illustrating corporate headquarters in New York. This is where we have our main servers in our distributed system. Modeling tool is OPNET.

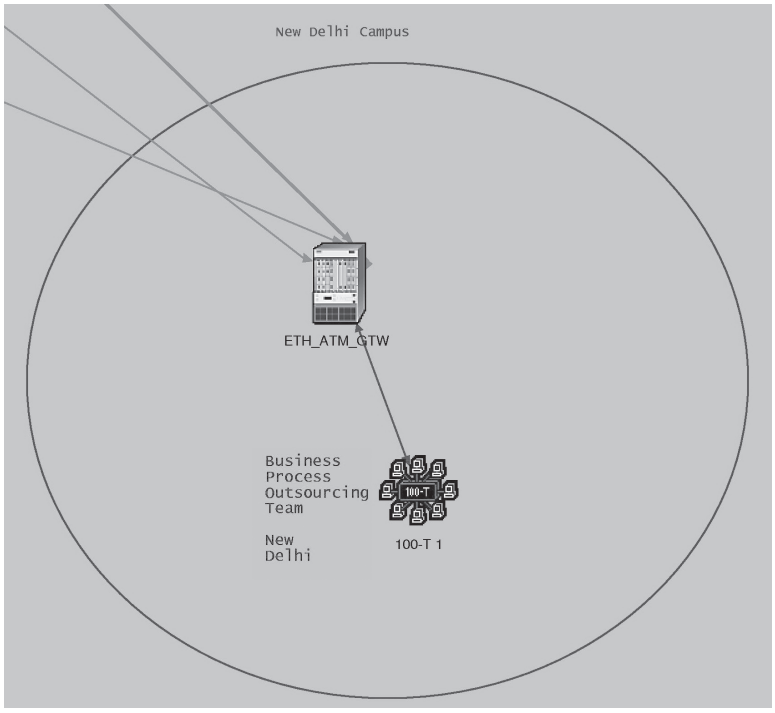


**FIGURE 11** Segment of simulation model illustrating New Delhi campus network. This is where the business process outsourcing team and related servers in our distributed system are located. Modeling tool is OPNET.

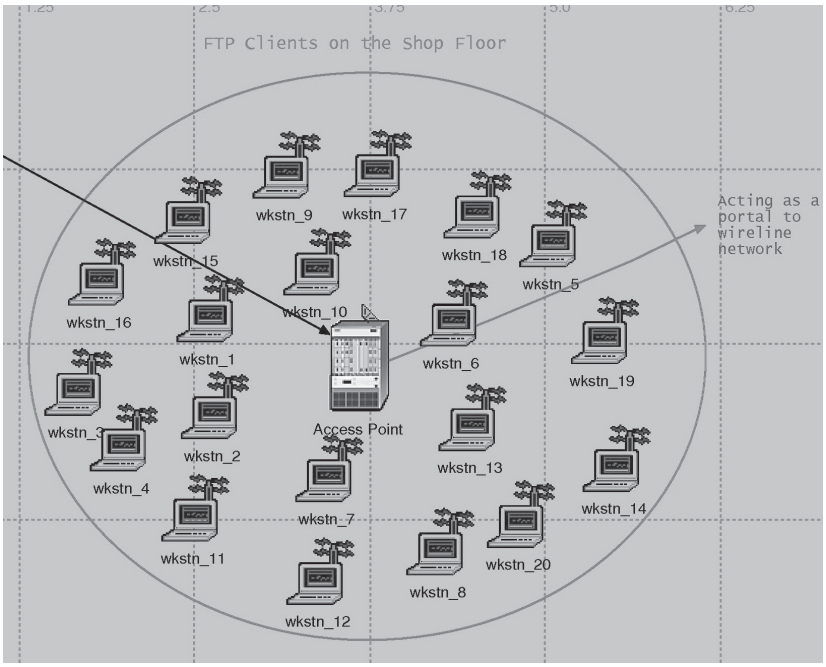


**FIGURE 12** Pharmaceutical company portals as a wireless network of a pharmaceutical manufacturing system. The power of the model is that we can simulate a shop-floor request, comment, or warning throughout the entire international network of globally distributed pharmaceutical companies, with all important functions and processes. This means that before any pharmaceutical manufacturing system is actually built, we can simulate the entire system in the digital domain, saving huge expense and time. Modeling tool is OPNET.

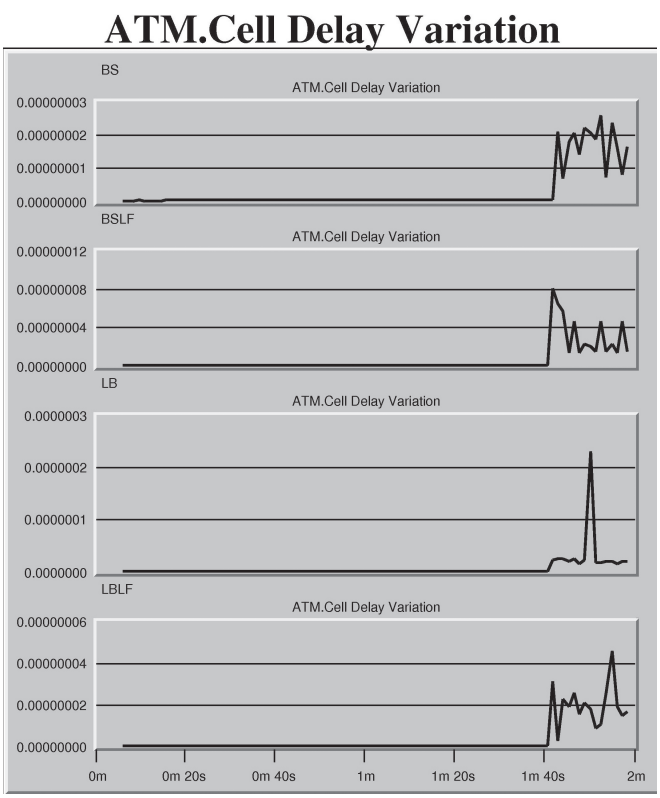


**FIGURE 13** Simulation diagram illustrating and confirming that the network system design from an ATM variation–response time point of view can cope with the demand. Modeling tool is OPNET.

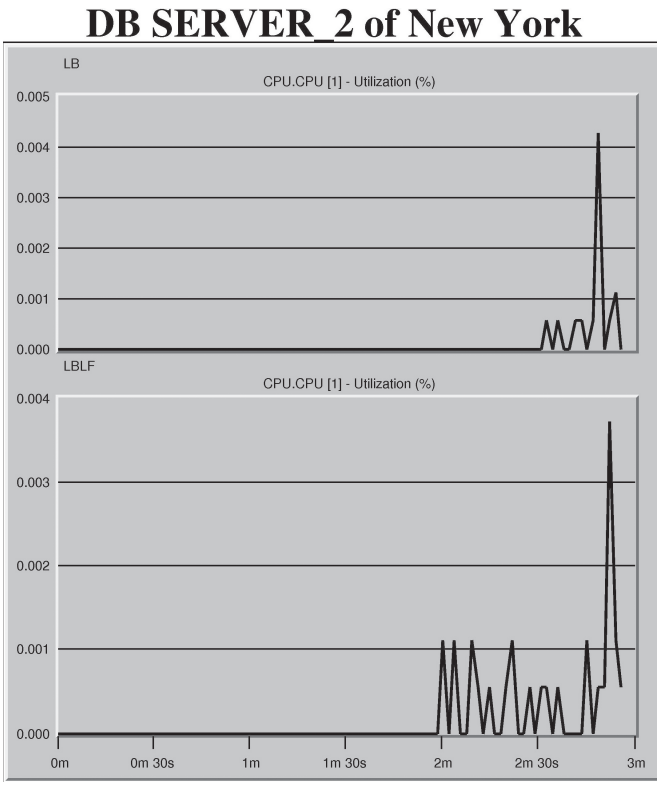


**FIGURE 14** Simulation diagram illustrating and confirming that the server balancing aspects of the network system design can cope with the demand. Modeling tool is OPNET.

or global subnetwork. Subnetworks can be created and interconnected within this top level or within other subnetworks. Subnetworks provide a powerful mechanism for manipulating complex networks by breaking down the system's complexity through abstraction.

Since in our pharmaceutical network simulation models we deal with packets, let us explain a few aspects of packet formats. Packets carry information and can be sent between transmitters and receivers. In our example, packets can carry robot programs when uploaded from the design/programming office servers to the robot lines and then to the individual CNCs, or robots, or parts of them if there is a need for an update, edit, quality control, production control, maintenance, and other data. (Packets can include mission-critical, "panic" related real-time data between the robot controller PCs and the line servers.)

Packets are data structures consisting of storage areas called fields and can either be formatted or unformatted. Formatted packets have fields designed according to a packet format which specifies the packets' field names, data types, sizes, and default values. Formatted packets can be read by corresponding communication protocols only. Unformatted packets have no predefined fields. In IT-Guru, packet formats are predefined and typically named according to the model in which they are intended to be used.

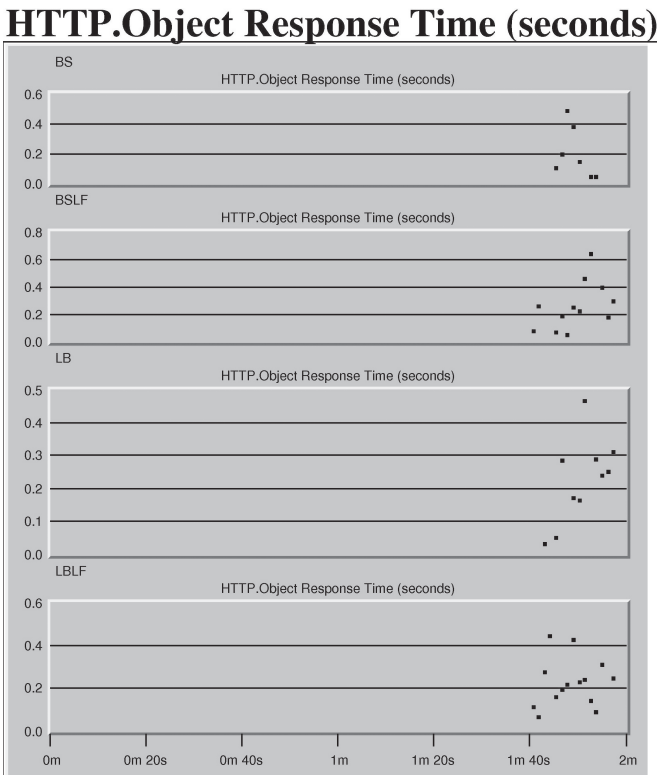


**FIGURE 15** Simulation diagram illustrating and confirming that the network system design from an object variation–response time point of view can cope with the demand. As an example, this is important if a pharmaceutical manufacturing system line manager in India wants to notify a manager in New York by sending an image object, a sound object, or a multimedia object of a machine in the line for quality evaluation. Modeling tool is OPNET.

## 15 EVALUATION OF NETWORK SIMULATION RESULTS

The goal of most simulation scenarios is to evaluate some aspect of a system's behavior or performance and to quantify, typically in terms of statistics, the results and then use the results for decisions. This requires a simulation environment with software tools that provide insight into a model's dynamic operation.

Based on IT-Guru's in-depth analysis, the pharmaceutical manufacturing system network engineering analyst can collect object, scenariowide, and global statistics as follows:

- Object statistics are collected from individual objects. They allow the network engineering analyst to evaluate the performance of specific network nodes or links (a single hub's Ethernet delay or a server balancing change, as in our example).
- Scenariowide object statistics are collected from all relevant objects in a network (e.g., Ethernet delay for every node). They allow the network engineering analyst to easily monitor the performance of all objects of a specific type.

- Global statistics are collected from the entire network. They represent results that apply to the network as a whole (such as global end-to-end delay) and let the designers and management analyze aspects of the network's overall performance.
- More specifically, IT-Guru offers the following types of statistics when analyzing networks:
  - Queue size
  - Available space
  - Overflow occurrences
  - Delay
  - Interarrival times
  - Packet sizes
  - Throughput
  - Utilization
  - Error rates
  - Collisions
  - Application-specific statistics defined by a model developer

Because there are many possible statistics to collect, the data files would quickly grow past practical use if the simulation program recorded them all. Therefore, the analyst must specifically select the statistics that are valuable for the particular study before running a simulation [71–79].

## 16 SUMMARY

In this chapter, we have presented the foundations of an analytical and simultaneously computational lean and flexible pharmaceutical manufacturing system design approach based on total quality standards. We have discussed why this approach is essential for pharmaceutical product, process, and manufacturing system designs.

As illustrated, based on simulation results, using the plotted graphs and screens, management can easily evaluate different design alternatives, machine and human behavior models, control systems, sensory feedback processing, and the need of a balanced server architecture, and even investigate "what if" scenarios further, without committing to major upfront investment.

We can clearly state that the time has come when pharmaceutical manufacturing systems can be designed and built in an entirely digital domain, saving huge amounts of capital and other related cost, and simultaneously increasing quality.

## 17 COMPLIMENTARY VIDEO ON DVD

To show real-world high-technology examples of pharmaceutical product, process, manufacturing, assembly, and packaging system designs, in action, something we cannot do in static, printed books, we have created a supplementary video, in high definition, and compressed onto a DVD. This professionally edited DVD supports

this chapter as an independent, self-contained publication illustrating advanced pharmaceutical and medical product, process, and manufacturing system designs, related quality assurance processes and solutions, and others, explained by industry experts. To find out more about this DVD, refer to Ranky, P. G., Ranky, G. N., and Ranky, R. G. (2006), *Design Principles and Examples of Pharmaceutical Manufacturing Systems (Product, Process, Lean and Flexible Manufacturing, Assembly and Packaging System Designs)*, Video on DVD, available: www.cimwareukandusa.com.

## REFERENCES


1. Ranky, P. G., A generic, analytical method to assess process-related risk with case studies, The Project Management Institute (PMI) Risk SIG and the Institute for International Research (IIR), paper presented at the Annual US National Project Risk Symposium, Houston, TX, May 22–May 26, 2006.
2. Ashley, S. (2004), Penny-wise smart labels, *Sci. Am*. 291(2), 30–31.
3. Ashton, P., and Ranky, P. G., (1999, Feb.), An advanced concurrent engineering research toolset and its application at Rolls-Royce motor cars, *ADAM (Adv. Des. Manufacturing),* available: http://www.cimwareukandusa.com, listed and indexed by the Association of Research Libraries, Washington, DC, and the Edinburgh Engineering Virtual Library, United Kingdom, Vol. 1.
4. Bradbrook, R. (2004), Wal-Mart and RFID, folding carton industry, *Printing News*, 31(4), 30–33.
5. Bradbrook, R. (2004), Procter and Gamble aim to be among the world leaders in RFID implementation, *Int. Paper Board Ind*., 47(8), 20–23.
6. Brzozowski, C. (2004), Tags, tickets & labels: New technologies emerge, *Printing News*, 153(3).
7. Ranky, P. G., An integrated PM approach, including: Process modeling, requirements analysis, risk analysis, statistical tools and 3D multimedia, presented at the Project Management Institute (PMI), New Jersey Chapter. Jan. 2006.
8. Ranky, P. G., Focus on RFID (radio frequency identification) methods, technologies and education, presented as part of the NCME Mission (National Center for Manufacturing Education), sponsored by NSF (National Science Foundation, USA) and industry, Jan. 2006.
9. Flaherty, M., Ranky, P. G., Ranky, M. F., Sands, S., and Stratful, S. (1999, Mar.), An engineering multimedia approach to servo pneumatic positioning, ADAM (Adv. Des. Manufacturing), available: http://www.cimwareukandusa.com, listed and indexed by the Association of Research Libraries, Washington DC, and the Edinburgh Engineering Virtual Library, United Kingdom, Vol. 1.
10. Glidden, R., Bockorick, C., etal. (2004), Design of ultra-low-cost UHF RFID tags for supply chain applications, *IEEE Commun. Mag*. 42(8), 140–151.
11. Graham-Rowe, D. (2004), Tags to banish forgetfulness, *New Scientist*, 183(2460), 19.
12. Graham-Rowe, D. (2004), Who's keeping tabs on your tags? *New Scientist*, 183(2462), 22.
13. Ho, K. L., and Ranky, P. G. (1999, Mar.), An object oriented approach to flexible conveyor system design, ADAM (Adv. Des. Manufacturing), available: http://www.cimwareukandusa.com, listed and indexed by the Association of Research Libraries, Washington DC, and the Edinburgh Engineering Virtual Library, United Kingdom, Vol. 1.

14. Knights, P. F., Henderson, E., and Daneshmend, L. K. (2004), Drawpoint control using radio frequency identification systems, *CIM Bull.*, 89(1003), 53–58.
15. Loose, D. C., and Ranky, P. G. (2007), *A Case-Based Introduction to IBM's Telematics Solutions*, interactive multimedia eBook with 3D objects, text, and videos in a browser-readable format on CD-ROM/intranet, available: http://www.cimwareukandusa.com, CIMware USA, Inc., and CIMware Ltd., United Kingdom; multimedia design and programming by P. G. Ranky and M. F. Ranky.
16. Mongeon, D. G. (2005, Feb.), RFID: A tool for the 21st century distribution, USA Department of Defense Conference Presentation, RFID Media Briefing, Washington DC.
17. Nadler, S. F., Ranky, P. G., and Ranky, M. (2002–2003), A 3D multimedia approach to the diagnosis of low back pain (Vol. 1, 18- and 40-year-old males), interactive 3D multimedia presentation on CD-ROM with off-line Internet support (650 Mbytes, approx. 150 interactive screens, 50 minutes of digital videos, 3D internal and external body tour, animation, and 3DVR objects), by CIMware (IEE and IMechE Approved Professional Developer); also in Multimedia design and programming by P. G. Ranky and M. F. Ranky.
18. Ranky, G. N., and Ranky, P.G. (2005), Japanese service robot R&D trends and examples, *Ind. Robot*, 32(6), 460–467.
19. Ranky, P. G., Caudill, R. J., Limaye, K., Alli, N., ChamyVelumani, S., Bhatia, A., and Lonkar, M. (2002, May), A Web-enabled virtual disassembly manager (web-VDM) for electronic product/process designers, disassembly line managers and operators, a UML (Unified Modeling Language) model of our generic digital factory, and some of our electronic support system analysis tools, ADAM with IT (Adv. Des. Manufacturing), available: http://www.cimwareukandusa.com, listed and indexed by the Association of Research Libraries, Washington DC, and the Edinburgh Engineering Virtual Library, United Kingdom, Vol. 3.
20. Ranky, P. G. (2006), Introduction to RFID—Radio frequency identification methods and solutions, *Assembly Automation*, 26(1), 28–33.
21. Ranky, P. G. (1999, Apr.), New trends in flexible, lean and agile manufacturing cells and systems, *ADAM (Adv. Des. Manufacturing)*, available: http://www.cimwareukandusa.com, listed and indexed by the Association of Research Libraries, Washington DC, and the Edinburgh Engineering Virtual Library, United Kingdom, Vol. 1.
22. Ranky, P. G. (2002), A method for planning industrial robot networks for automotive welding and assembly lines, *Ind. Robot: Int. J.*, 29(6), 530–537.
23. Ranky, P. G. (2003), A real-time manufacturing/assembly system performance evaluation and control model with integrated sensory feedback processing and visualization, *Assembly Automation*.
24. Ranky, P. G. (2003), A simulation method and distributed server balancing results of networked industrial robots for automotive welding and assembly lines, *Ind. Robot: Int. J.*, 30(2), 192–197.
25. Ranky, P. G. (2002), Advanced digital automotive sensor applications, *Sensor Rev.: Int. J.*, 22(3), 213–217.
26. Ranky, P. G. (2003), Advanced machine vision systems and application examples, *Sensor Rev.: Int. J.*, 23(3), 242–245.
27. Ranky, P. G. (2003), Collaborative synchronous robots serving machines and cells, *Ind. Robot: Int. J.*, 30(3), 213–217.
28. Ranky, P. G. (2004), Digital, Internet-enabled assembly line and factory modeling, *Assembly Automation*, 24(3), 247–253.
29. Ranky, P. G. (2004), Novel automated inspection methods, tools and technologies, *Assembly Automation: Int. J.*, 23(3), 252–257.

30. Ranky, P. G. (2003), Reconfigurable robot tool designs and integration applications, *Ind. Robot: Int. J.*, 30(4), 338–344.
31. Ranky, P. G. (2002), Smart sensors, *Sensor Rev.: Int. J.*, 22(4), 312–318.
32. Ranky, P. G. (1999, Jan.,) Some generic algorithmic solutions to the problem of dynamic scheduling in flexible manufacturing systems that operate globally, ADAM (Adv. Des. Manufacturing), available: http://www.cimwareukandusa.com, listed and indexed by the Association of Research Libraries, Washington DC, and the Edinburgh Engineering Virtual Library, United Kingdom, Vol. 1.
33. Ranky, P. G. (2001), Trends and R&D in virtual and robotized product disassembly, *Ind. Robot*, 28(6), 454–456.
34. Ranky, P. G. (2000, May), Some analytical considerations of engineering multimedia system design within an object oriented architecture, *Int. J. CIM*, 13(2), 204–214.
35. Ranky, P. G., and ChamyVelumani, S. (2003), A method, a tool (CORA), and application examples for analyzing disassembly user interface design criteria, *Int. J. CIM*, 16(4–5), 317–325.
36. Ranky, P. G., and ChamyVelumani, S. (2003), An analytical approach, a tool (DFRA) and application examples for assessing process-related failure risks, *Int. J. CIM*, 16(4–5), 326–333.
37. Ranky, P. G., and Nadler, S. F., A novel multimedia approach to low back pain diagnosis with internal and external 3D interactive body tours, paper presented at the 29th Annual Northeast Bioengineering Conference, New Jersey Institute of Technology, University Heights, Newark, NJ, Mar. 2003.
38. Ranky, P. G., and Nadler, S. F., A new, Web-enabled multimedia approach with 3D virtual reality internal and external body tours to support low back pain diagnosis, paper presented at the 4th Annual Faculty Best Practices Showcase in Kean University, NJ, Mar. 2003.
39. Ranky, P. G., and Ranky, M. F. (2000), A Dynamic Operation control algorithm with multimedia objects for flexible manufacturing systems, *Int. J. CIM*, 13(2), 245–263.
40. Ranky, P. G., Lonkar, M., and ChamyVelumani, S. (2003), eTransition models of collaborating design and manufacturing enterprises, *Int. J. CIM*, 16(4–5), 255–266.
41. Ranky, P. G., Morales, C., and Caudill, R. J., Lean Disassembly line layout, process and network simulation models and cases, based on real-world data, paper presented at the IEEE (USA) International Symposium on Electronics and the Environment and the IAER Electronics Recycling Summit, Boston, MA, May 19–22, 2003.
42. Ranky, P. G., Ranky, G. N., and Ranky, R.G. (2006), Examples of pharmaceutical product/process/manufacturing/assembly and packaging system designs, video on DVD, available: www.cimwareukandusa.com.
43. Ranky, P. G., Subramanyam, M., Caudill, R. J., Limaye, K., and Alli, N., A dynamic scheduling and balancing method and software tool for lean and reconfigurable disassembly lines, paper presented at the IEEE (USA) International Symposium on Electronics and the Environment and the IAER Electronics Recycling Summit, Boston, MA, May 19–22, 2003.
44. Ranky, P. G., 3D engineering multimedia cases. A customizable 3D Web-enabled library with reusable objects, paper presented at the ASEE (American Society of Engineering Educators) Mid-Atlantic Conference, Kean University, NJ, Apr. 2003.
45. Ranky, P. G., A 3D multimedia approach to biomedical engineering: Low back analysis, paper presented at the ASEE, American Society of Engineering Educators, U.S. National Meeting, Biomedical Engineering Division, Nashville, TN, June 2003.
46. Ranky, P. G., Ranky, G. N., and Ranky, R. G. (2006), Design principles and examples of pharmaceutical manufacturing systems (product, process, lean & flexible manufacturing,

assembly and packaging system designs), video on DVD, available: www.cimwareukandusa.com.

47. Ranky, P. G. (2001–2006), A 3D multimedia case: Component oriented disassembly failure risk analysis, an interactive multimedia publication with 3D objects, text and videos in a browser-readable format on CD-ROM/intranet available: http://www.cimwareukandusa.com, CIMware USA, Inc., and CIMware Ltd., United Kingdom; Multimedia design and programming by P. G. Ranky and M. F. Ranky (published 6 volumes of this main title with different risk analysis challenges explained).
48. Ranky, P. G. (2001–2005), A 3D multimedia case: component oriented disassembly user requirements analysis, an interactive multimedia eBook publication with 3D objects, text and videos in a browser-readable format on CD-ROM/intranet available: http://www.cimwareukandusa.com, CIMware USA, Inc., and CIMware Ltd., United Kingdom, Multimedia design and programming by P. G. Ranky and M. F. Ranky (published 7 volumes of this main title with different requirements analysis challenges explained).
49. Ranky, P. G., A 3D Web collaborative concurrent automotive engineering Method Based on our "distributed digital factory" and "digital car" models, paper presented at the Society of Automotive Engineers World Congress, Detroit, MI, Mar. 2003.
50. Ranky, P. G. (2003), A 3D Web-enabled, case based learning architecture and knowledge documentation method for engineering, information technology, management, and medical science/biomedical engineering, *Int. J. CIM*, 16(4–5). 346–356.
51. Ranky, P. G., A Biomedical Engineering case with 3D lower back interactive virtual anatomy tours inside and outside the human body with automated post-test student assessment, paper presented at the ASEE (American Society of Engineering Educators) Mid-Atlantic Conference, Kean University, NJ, Apr. 2003.
52. Ranky, P. G., A new approach for teaching and learning about engineering process failure risk analysis with IE (industrial engineering) case studies, paper presented at the ASEE, American Society of Engineering Educators, US National Meeting, Industrial Engineering Division, Nashville, TN, June 2003.
53. Ranky, P. G., A novel 3D Internet-based multimedia method for teaching and learning about engineering management requirements analysis, paper presented at the ASEE, American Society of Engineering Educators, US National Meeting, Engineering Management Education Division, Nashville, TN, June 2003.
54. Ranky, P. G., An interactive 3D multimedia problem-based library for manufacturing engineering technology education with Internet support, paper presented at the ASEE, American Society of Engineering Educators, US National Meeting, Engineering Technology Division, Nashville, TN, June 2003.
55. Ranky, P. G. (2003–2005), An introduction to alternative energy sources: Hybrid & fuel cell vehicles; an interactive multimedia eBook publication with 3D objects, text, and videos in a browser-readable format on CD-ROM/intranet, available: http://www.cimwareukandusa.com, CIMware USA, Inc., and CIMware Ltd.; United Kingdom, Multimedia design and programming by P. G. Ranky and M. F. Ranky, (2003–2005), Customer needs, wants & requirements analysis: Automotive exterior rearview mirror, an interactive multimedia eBook publication with 3D objects, text, and videos in a browser-readable format on CD-ROM/intranet, available: http://www.cimwareukandusa.com, CIMware USA, Inc., and CIMware Ltd., United Kingdom.
56. Ranky, P. G. (2003–2005), An introduction to digital factory & digital telematic car modeling with R&D and industrial case studies, an interactive multimedia eBook publication with 3D objects, text, and videos in a browser-readable format on CD-ROM/intranet, available: http://www.cimwareukandusa.com, CIMware USA, Inc. and CIMware Ltd., United Kingdom, Multimedia design and programming by P. G. Ranky and M. F. Ranky.

57. Ranky, P. G. (2005), An introduction to RFID, radio frequency identification methods and applications, DVD video, available: www.cimwareukandusa.com (approximately 30 min).
58. Ranky, P. G. (2005–2006), An introduction to RFID, radio frequency identification methods and applications with a total quality management and control focus, interactive browser-readable 3D eBook, available: www.cimwareukandusa.com, (approximately 30 min).
59. Ranky, P. G. (1999, Apr.), An object oriented system analysis and design method (CIMpgr) and an R&D case study, Adv. Des. Manufacturing, available: http://www.cimwareukandusa.com, listed and indexed by the Association of Research Libraries, Washington DC, and the Edinburgh Engineering Virtual Library, United Kingdom, Vol. 1.
60. Ranky, P. G., Computerized engineering assessment method based on 3D interactive multimedia, That students enjoy, paper presented at the ASEE, American Society of Engineering Educators, US National Meeting, Continuing Professional Development Division, Nashville, TN, June 2003.
61. Ranky, P. G. (1999), Design, manufacturing and assembly automation trends and strategies in China, *Assembly Automation*, 19(4), 301–305.
62. Ranky, P. G. (2003, Feb.), Designing a lean infrastructure; advanced machining cell design concepts, methods, architectures and cases, *Manuf. Eng., J. IEE*, 22–24.
63. Ranky, P. G. (2000), Engineering multimedia in CIM (computer integrated manufacturing), *Int. J. CIM*, 13(2), 169–171.
64. Ranky, P. G. (2003), eTransition in the multi-lifecycle CIM (computer integrated manufacturing) context, *Int. J. CIM*, 16(4–5), 229–234.
65. Ranky, P. G., Interactive 3D multimedia cases for engineering education with Internet support, ASEE, American Society of Engineering Educators, paper presented at the U.S. National Meeting, Computers in Education Division, Nashville, TN, June 2003.
66. Ranky, P. G., Interactive 3D multimedia cases for manufacturing engineering education with Internet support, paper presented at the ASEE, American Society of Engineering Educators, US National Meeting, Manufacturing Engineering Education Division, Nashville, TN, June 2003.
67. Ranky, P. G. (2002, Dec.), Introduction to concurrent engineering, an NSF (National Science Foundation, USA) sponsored Gateway Coalition streamed multimedia narrated web presentation, New Jersey Intitute of Technology, Public Research University, Newark, New Jersey.
68. Ranky, P. G. (2003–2005), Key R&D and eTransition trends in US and international collaborative design & manufacturing enterprises, an interactive multimedia eBook publication with 3D objects, text, and videos in a browser-readable format on CD-ROM/intranet, available: http://www.cimwareukandusa.com, CIMware USA, Inc., and CIMware Ltd., United Kingdom; Multimedia design and programming by P. G. Ranky and M. F. Ranky.
69. Ranky, P. G. (2000, Jan.), Modular fieldbus designs and applications, *Assembly Automation*, 20(1), 40–45.
70. Ranky, P. G. (2003), Network simulation models of lean manufacturing systems in digital factories and an intranet server balancing algorithm, *Int. J. CIM*, 16(4–5), 267–282.
71. Ranky, P. G., Rapid prototyping cases for integrated design and manufacturing engineering education with 3D Internet support, paper presented at the ASEE, American Society of Engineering Educators, US National Meeting, Design in Engineering Education Division, Nashville, TN, June 2003.
72. Roman, H. T., and Ranky, P. G. (2003–2005), A case-based Introduction to Service robotics, an interactive multimedia eBook publication with 3D objects, text, and videos in a browser-readable format on CD-ROM/intranet, available: http://www.cimwareukandusa.

com, CIMware USA, Inc., and CIMware Ltd., United Kingdom. Multimedia design and programming by P. G. Ranky and M. F. Ranky.

73. Romero, C., Department of logistics passive RFID initial implementation, paper presented at the USA Department of Defense Conference, RFID Media Briefing, Washington DC, Feb. 2005.
74. Sangoi, R., Smith, C. G., et al. (2004), Printing radio frequency identification (RFID) tag antennas using inks containing silver dispersions, *J. Dispersion Sci. Technol*. 25(4), 513–521.
75. Smith, K., Enabling the supply chain, paper presented at the USA Department of Defense Conference, RFID Media Briefing, Washington, DC, Feb. 2005.
76. Sugimoto, M., Kusunoki, F., Inagaki, S., Takatoki, K., and Yoshikawa, A. (2004), A system for supporting collaborative learning with networked sensing boards, *Syst. Comput. Jpn*., 35(9), 39–50.
77. Wilke, P., and Braunl, T. (2001), Flexible wireless communication network for mobile robot agents, *Ind. Robot*, 28(3), 220–233.

# ROLE OF QUALITY SYSTEMS AND AUDITS IN PHARMACEUTICAL MANUFACTURING ENVIRONMENT

Evan B. Siegel and James M. Barquest
*Ground Zero Pharmaceuticals, Inc., Irvine, California*

**Contents**

1 cGMP Regulations
  1.1 Duties of Quality Control Unit under cGMP Regulations
2 Quality Assurance Function
3 Quality Systems Approach
4 Management Responsibilities
5 Resources
6 Manufacturing Operations
  6.1 Design, Develop, and Document Product and Processes
  6.2 Inputs
  6.3 Perform and Monitor Operations
  6.4 Address Nonconformities
7 Evaluation Activities
  7.1 Trend Analysis
  7.2 Conduct Internal Audits
  7.3 Quality Risk Management
  7.4 Corrective and Preventive Actions
  7.5 Promote Improvement
8 Transitioning to Quality Systems Approach
9 Audit Checklist for Drug Industry
  9.1 Instructions for Using Audit Checklist
  References

By regulation, appropriate practice, and common sense, quality assurance (QA) is a critical function in the pharmaceutical manufacturing environment. The need for an independent unit to audit and comment on the appropriate application of

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

standard operating procedures, master batch records, procedures approved in product applications, and the proper functioning of the quality control (QC) unit is paramount. This helps assure that products are manufactured reliably, with adherence to approved specifications, and that current good manufacturing practices (cGMP) are maintained in conformance to regulation, both in the facility in general and the microenvironment of each product's manufacturing sequence.

Quality assurance personnel must have the appropriate training, experience, familiarization with the manufacturing facility and products, enforced independence from the production chain of command, and the ability to review adherence to procedures, policies, and agreed-upon approaches to manufacturing quality pharmaceuticals. This helps to provide both an environment and a manufactured product that can withstand Food and Drug Administration (FDA) inspection and support a firm's reputation for quality products.

The cGMP regulations establish requirements that are intended to provide a high level of assurance that the pharmaceutical products produced satisfy the strength, purity, potency, and other quality requirements established for the finished product to assure that it is fit for its intended use. Manufacturers must establish a quality control unit that is responsible for many of the quality-related activities required by the regulations. These regulations have not been substantially updated since 1978. Since then, the science and practice of quality assurance have substantially evolved to include the development of quality management systems and risk management approaches to better assure product quality and fitness for use. Pharmaceutical product manufacturers are increasingly interested in implementing a comprehensive quality management system (QMS) and employing risk management approaches because they allow them to apply newer quality management principles that they believe enable them to more effectively assure product quality and better allow harmonization with evolving international regulatory quality system requirements. The FDA has not changed the cGMP regulations but, as part of its Pharmaceutical CGMPs for the 21st Century Initiative, encourages this quality systems approach to cGMP compliance.

This Chapter describes outlines and discusses the regulations applicable to the QA function and unit, structure, function, charter, and application of the unit in the pharmaceutical manufacturing environment. In addition, it discusses additional quality-related responsibilities that may result when manufacturers move toward a quality systems approach to quality that incorporates current quality system models to further improve quality and harmonize with international quality system requirements.

The justification for, and execution of, the QA audit are also described, including preparation, key items of interest, a typical checklist of the audit itself, corrective and preventive actions following the audit, and suggested measures for assuring successful operation of the unit.

## 1 cGMP REGULATIONS

The cGMP regulations for the manufacture of pharmaceutical products are contained in Parts 210 and 211 of Title 21 of the *Code of Federal Regulations* (CFR) [1]. These regulations, as well as guidance documents and other FDA documents

pertaining to the regulation and FDA inspection of pharmaceutical product manufacturers, may be accessed on the FDA website at www.fda.gov. Part 210 specifies the scope and applicability of the cGMP regulations and defines terms used in the regulations. Part 210 also indicates that the regulations establish "minimum" cGMP requirements and that products that are not manufactured under cGMP are adulterated. Adulterated products and the persons responsible for the adulteration are subject to regulatory action by the FDA.

Part 211 contains specific good manufacturing practice requirements for finished pharmaceuticals and is divided into Subparts A–K as follows:

A. Scope
B. Organization and Personnel
C. Buildings and Facilities
D. Equipment
E. Control of Components and Drug Product Containers and Closures
F. Production and Process Controls
G. Packaging and Labeling Control
H. Holding and Distribution
I. Laboratory Controls
J. Records and Reports
K. Returned and Salvaged Drug Products

The cGMP regulations are written to address the primary potential sources of product variability. Subpart B establishes the quality control unit and the duties of that unit, establishes personnel requirements and addresses personnel practices (e.g. sanitation) intended to reduce the likelihood of product contamination. Subparts C and D establish requirements for buildings and facilities and equipment used in the manufacture, processing, packing, or holding of a drug product. Subparts E through H establish controls over the major processes associated with the production of a finished and packaged drug product that is ready to be shipped for distribution to users. Controls are established for incoming raw materials and components and continue through manufacturing, packaging, labeling, holding, and distribution of finished, packaged, labeled, and released drug product. Subpart I requires the establishment of scientifically sound and appropriate specifications, standards, sampling plans, and test procedures; requires instrument specifications and calibration; and establishes lot or batch testing and release requirements. Subpart J establishes documentation requirements including master and batch records, and Subpart K addresses the control and disposition of returned drug products and places limitations on the salvage of drug products that have been subjected to improper storage conditions (e.g., smoke, heat, fire, moisture).

### 1.1 Duties of Quality Control Unit under cGMP Regulations

The cGMP regulations assign specific duties to the quality control unit. The unit is required to have the responsibility and authority to approve or reject all components, drug product containers, closures, in-process materials, packaging material,

labeling, and drug products and the authority to review production records to assure that no errors have occurred or, if errors have occurred, that they have been fully investigated. The responsibilities of the unit extend to approving or rejecting drug products manufactured, processed, packed, or held by contract manufacturers. The organization must assure that the quality control unit has adequate laboratory facilities for the testing and approval (or rejection) of components, drug product containers, closures, packaging materials, in-process materials, and drug products.

In addition to duties associated with the approval of materials and finished products, the unit is also responsible for approving or rejecting all procedures or specifications impacting on the identity, strength, quality, and purity of the drug product. This includes review and approval of procedures for production and process control, including any changes to these procedures. These procedures, and the responsibilities and procedures applicable to the quality control unit within the organization, must be written and followed.

All specifications, standards, sampling plans, test procedures, or other laboratory control mechanisms, including any changes, must be in writing and reviewed and approved by the quality control unit.

Written procedures describing the handling of all written and oral complaints regarding a drug product are required. The quality control unit is responsible for reviewing any complaint involving the possible failure of a drug product to meet any of its specifications and, for such drug products, making a determination as to the need for an investigation in accordance with cGMP requirements. The review should include a determination if the complaint represents a serious and unexpected adverse drug experience, which is required to be reported to the FDA. A written record of each complaint must be maintained in a complaint file.

## 2 QUALITY ASSURANCE FUNCTION

The term *quality* is used in many industries and in everyday life and can have various meanings depending on context. For the purposes of discussion here quality means the product requirements or attributes that have a bearing on the product's specified requirements. Quality assurance activities are those processes and activities conducted to assure that a product or service consistently satisfies its requirements and is fit for its intended use. In the pharmaceutical manufacturing environment, this means the activities conducted to assure that the pharmaceutical product's identity, strength, purity, potency, and other quality attributes conform to approved specifications.

In the United States, cGMP requirements for the manufacture of drugs were established by regulation in 1978 and have not been substantially updated since then. The science and practice of quality assurance has substantially evolved since then to include the development of quality systems [2, 3] and risk management approaches [4] to better assure product quality and fitness for use. Pharmaceutical product manufacturers are increasingly interested in implementing these approaches because they allow the manufactures to apply newer quality management principles that they believe enable them to more effectively assure product quality and better allow harmonization with evolving international regulatory quality system requirements.

## 3 QUALITY SYSTEMS APPROACH

The systems approach to quality involves a coordinated approach to the management of quality-related activities as processes that work in conjunction with one another to provide assurance that the product meets its specified requirements. It involves:

- A management commitment to quality that is communicated throughout the organization
- Identifying quality requirements using risk management and other methods as appropriate
- Developing a quality policy, plan, objectives
- Establishing an organizational structure with identified responsibilities and authorities that allows quality objectives to be met
- Providing the resources needed to meet quality objectives
- Developing the required systems and processes
- Establishing methods for the ongoing objective evaluation of the performance of systems and processes including quality auditing
- Initiating corrective and preventive actions as needed to assure that quality objectives are consistently and reliably met

The use of risk management techniques in identifying product requirements, establishing processes and process control and monitoring methods, evaluating quality data, identifying appropriate corrective and preventive actions to address quality problems, and for other quality-related activities can increase the overall efficiency and effectiveness of the quality system.

The FDA has recognized the value of and encourages a risk based quality systems approach for the manufacture of pharmaceutical products. This is reflected in its Pharmaceutical CGMPs for the 21st Century Initiative. In association with this initiative the FDA has published reports and guidance documents that collectively provide information that can be used by pharmaceutical product manufacturers in implementing a quality systems and risk management approach to pharmaceutical cGMP regulations compliance [5–8]. In implementing this initiative, the FDA has made it clear that it does not impose new regulatory requirements on manufacturers. The FDA has provided information and guidance that is intended to serve as a bridge between the 1978 regulations and current quality systems by explaining how manufacturers implementing such systems can do so in full compliance with the cGMP regulations. This approach differs from that used by the FDA when it updated the cGMP regulations for medical devices to employ a quality systems approach. The 1996 Quality System Regulation updated the GMP requirements for finished medical device manufacturers to reduce the risk of inadequate device design and to harmonize them with international quality system standards that were in effect at that time [9]. These international standards have since been updated [10]; however the device quality system regulation remains consistent with modern quality system models.

In a modern quality system, the organizational unit responsible for quality-related activities within the organization generally has a central role in the

development and management of the overall quality system. These activities can include quality control, quality assurance, quality planning, and quality improvement. The cGMP regulations do not define or employ these terms, but the activities the regulations assign to the quality control unit fall within these definitions as currently defined [2, 8, 11].

Current quality system models involve quality-related activities and terms that are not included in the cGMP regulations. Further, quality as a professional discipline is evolving. It is, therefore, important for organizations adopting a quality systems approach to unambiguously define the terms and quality concepts they will be using and to include these definitions as appropriate in training all staff in the organization who will be involved in quality-related activities. This will help assure effective communication throughout the organization and with vendors and others (e.g., regulatory agencies, third-party auditors) who interact with the organization on quality-related matters. Regulatory definitions should be recognized, and the use of nonstandard or outdated terminology should be avoided to the extent possible. Incorporating terms and definitions by reference from pertinent standards and FDA guidance documents may be helpful. FDA guidance on the quality systems approach to pharmaceutical cGMP regulations (pharmaceutical QS guidance) includes the following definitions:

*Quality Assurance (QA)* Proactive and retrospective activities that provide confidence that requirements are fulfilled.

*Quality Control (QC)* The steps taken during the generation of a product or service to ensure that it meets requirements and that the product or service is reproducible.

*Quality Management (QM)* Accountability for the successful implementation of the quality system.

*Quality System (QS)* Formalized business practices that define management responsibilities for organizational structure, processes, procedures, and resources needed to fulfill product/service requirements, customer satisfaction, and continual improvement.

*Quality Unit (QU)* A group organized within an organization to promote quality in general practice.

The FDA notes in its pharmaceutical QS guidance document that many current quality system concepts correlate very closely with the cGMP regulations and that the activities required by the regulations are generally consistent with a quality systems approach. In this and other guidance documents, the FDA uses the term *quality unit* rather than *quality control unit* as defined in the cGMP regulations to refer to the organizational unit with responsibility for quality-related activities. In a modern quality systems model these quality-related activities may go beyond, but are not necessarily inconsistent with, those required by the cGMP regulation.

Use of the term quality unit is consistent with current quality management system models [2, 10], which are intended to assure that the various operations associated with all systems are appropriately planned, approved, conducted, and monitored, and because the cGMP regulations specifically assign the QU the authority to create,

monitor, and implement a quality system. The FDA cautions that such activities do not substitute for, or preclude, the daily responsibility of manufacturing personnel to build quality into the product. The FDA has specifically indicated that the overarching philosophy articulated in both the cGMP regulations and in robust modern quality systems is that quality should be built into the product, and testing alone cannot be relied on to ensure product quality.

Other cGMP-assigned responsibilities of the QU that are consistent with modern quality system approaches include the following:

- Ensuring that controls are implemented and completed satisfactorily during manufacturing operations
- Ensuring that developed procedures and specifications are appropriate and followed, including those used by a firm under contract to the manufacturer
- Approving or rejecting incoming materials, in-process materials, and drug products
- Reviewing production records and investigating any unexplained discrepancies

The FDA has stressed that the release of the pharmaceutical QS guidance document does not impose new regulatory requirements on manufacturers but encourages manufactures to adopt a quality systems approach to cGMP compliance because of the potential benefits. An appropriately designed and implemented quality system can do the following:

- Reduce the number of (or prevent) recalls, returned or salvaged products, and defective products entering the marketplace
- Harmonize the cGMP regulations to the extent possible with other widely used quality management systems, which is desirable because of the globalization of pharmaceutical manufacturing, and the increasing prevalence of drug–device and biologic–device combination products
- When coupled with manufacturing process and product knowledge and the use of effective risk management practices, handle many types of changes to facilities, equipment, and processes without the need for prior approval regulatory submissions
- Potentially result in shorter and fewer FDA inspections by lowering the risk of manufacturing problems
- Provide the necessary framework for implementing *quality by design* (building in quality from the product development phase and throughout a product's life cycle), continual improvement, and risk management in the drug manufacturing process

This suggests that even without making changes in the cGMP regulations, the FDA may be looking at them from a "new" quality systems perspective. The regulations include terms such as *adequate* and *appropriate* that may be subject to interpretation based on relevant technical or scientific capabilities and state-of-the-art knowledge. As these improve, the interpretation of what is adequate or appropriate

can change as well. Practically, most manufacturers are more than willing to adopt methods that can improve the quality and safety of their pharmaceutical products because it is cost effective in the long run [11] but may be reluctant to do so for fear of being considered out of compliance with the cGMP regulations. Current FDA efforts in this regard should serve to allay manufacturers' concerns in this area.

The major elements of the quality system model described in the FDA's pharmaceutical QS guidance document are consistent with existing quality system standards. These elements are as follows:

- Management responsibilities
- Resources
- Manufacturing operations
- Evaluation activities

## 4 MANAGEMENT RESPONSIBILITIES

Current quality system models assign management a major role in the deployment and operation of a successful quality system. In such systems, major management responsibilities include the following:

- Provide leadership by establishing a commitment to quality that is supported by all levels of management and is communicated throughout the organization
- Create an organizational structure with clearly defined responsibilities and authorities to perform quality functions associated with achieving quality objectives
- Building and documenting a quality system to meet specified quality and regulatory requirements and achieve quality objectives
- Establishing a quality policy and objectives, and quality plans that are aligned with the organization's strategic plans and communicate this throughout the organization
- Reviewing the system by establishing appropriate accountability systems within the organization to monitor and report quality data and system status to management and assure that appropriate corrective and preventive actions are taken in response to quality problems using effective change control procedures and documented

The cGMP does not specifically assign management responsibility for these actions, although actions of this nature are required by the regulation. Table 1 from the pharmaceutical QS guidance document shows this relationship.

Under a comprehensive quality system the QU can expect an expanded and more visible role within the organization with greater accountability to and interaction with upper management. The QU should ideally be independent of the other organizational units to assure clear delineation of responsibility and authority and avoid conflicts. In certain instances, such as auditing, independence or objectivity is central

**TABLE 1 21 CFR cGMP Regulations Related to Management Responsibilities**

| Quality System Element | Regulatory Citations |
|---|---|
| 1. Leadership | |
| 2. Structure | Establish quality function: §211.22(a) [see definition §210.3(b)(15)[ |
| | Notification: §211.180(f) |
| 3. Build QS | QU procedures: §211.22(d) |
| | QU procedures, specifications: §211.22(c), with reinforcement in: §§211.100(a), 211.160(a) |
| | QU control steps: §211.22(a), with reinforcement in §§211.42(c), 211.84(a), 211.87, 211.101(c)(1), 211.110(c), 211.115(b), 211.142, 211.165(d), 211.192 |
| | QU quality assurance; review/investigate: §§211.22(a), 211.100(a–b) 211.180(f), 211.192, 211.198(a) |
| | Record control: §§211.180(a–d), 211.180(c), 211.180(d), 211.180(e), 211.186, 211.192, 211.194, 211.198(b) |
| 4. Establish policies, objectives, and plans | Procedures: §§211.22(c–d), 211.100(a) |
| 5. System review | Record review: §§211.100, 211.180(e), 211.192, 211.198(b)(2) |

to the effectiveness of the audit process, and auditors therefore should not have direct responsibility over the areas being audited.

The cGMP regulations do not specify how the QU should be integrated into the overall organization but, in general, the QU should be structured to reflect management's strong commitment to quality and to facilitate achieving quality objectives. The structure (e.g., organizational relationship to other organizational units, reporting relationships) should provide clear lines of responsibility and authority that support the production, quality, and management activities necessary to achieve quality objectives. Different organizations may accomplish this in different ways; however, experience has been that placement of the quality function on the same level within the organizational hierarchy as other major organizational units (e.g., production) sends a clear message both within and outside the organization that top management has a strong commitment to quality.

The cGMP regulations require quality-related activities to be conducted during all phases of manufacturing from the acceptance of raw materials through batch release, packaging, and labeling. The regulations also require that all personnel, including those engaged in quality-related activities, have sufficient education, training, and experience or any combination thereof to enable them to perform their assigned functions. In a quality systems approach to cGMP compliance, the role of quality personnel can be significantly expanded to include internal quality auditing, expanded review and analysis of quality data, investigation of nonconformance, root cause analysis, risk analysis, and other quality-related activities. Many of these activities are likely to be conducted with personnel from other organizational elements such as manufacturing, material control, facilities, product development, or engineering staff. Quality staff should have sufficient scientific and technical knowledge and training (e.g., statistical methods, risk analysis) and knowledge of the product and manufacturing processes to effectively perform their assigned functions

and competently interact with personnel from other organizational elements as necessary.

## 5 RESOURCES

The appropriate assignment of resources is essential to the success of any endeavor, and this is particularly critical in a pharmaceutical manufacturing environment. Inadequate staffing, training, manufacturing equipment and facilities, environmental controls, analytical equipment, and other resources can be sources of variability leading to the production of product that does not meet specified requirements. Modern quality system standards specifically address the issue of resources by requiring the organization to determine and provide the human, infrastructure, and work environment resources necessary for the quality system. The cGMP regulations address the resource issue in provisions that are intended to assure the adequacy of personnel (including consultants), manufacturing facilities including contract facilities, equipment, and laboratory facilities. The QU has significant responsibility in this regard.

The FDA, in its pharmaceutical QS guidance document, discusses the need for adequate resources in developing, implementing, and managing a quality system that complies with the cGMP regulations. Management is responsible for identifying resource requirements and providing resources accordingly, including providing training that is appropriate to the assigned activities. Personnel should understand the impact of their activities on their assigned duties and be familiar with cGMP requirements and the organization's quality system. This is consistent with the generally accepted idea that a culture of quality within an organization requires personnel to understand quality concepts, the organization's quality and regulatory objectives, and how their assigned activities contribute to the achievement of these objectives and fit into the overall quality system. Management should establish a working environment that encourages problem solving and communication in identifying and acting upon quality-related issues. While the provision of resources is generally considered a management function, the QU and other organizational units should be involved in the identification of the resources required to achieve quality objectives, including regulatory compliance, the assessment of the adequacy of existing resources, evaluating the effect of personnel, facility, product, process, regulatory, and other changes on resource needs, and generally providing management the information needed to make necessary and appropriate resource decisions.

Current quality system models employ a risk-based and data-driven approach to the development of QS system requirements to assure their adequacy. The FDA notes that the cGMP regulations place as much emphasis on processing equipment as testing equipment and contain specific requirements for the qualification, calibration, cleaning, and maintenance of production equipment that may be a higher standard than most nonpharmaceutical quality system models. Organizations should always keep in mind that, while the FDA may be encouraging the adoption of a comprehensive quality system, any system developed must satisfy the requirements of the cGMP regulations.

**TABLE 2 21 CFR cGMP Regulations Related to Resources**

| Quality System Element | Regulatory Citation |
|---|---|
| 1. General arrangements | |
| 2. Develop personnel | Qualifications: §211.25(a)<br>Staff number: §211.25(c)<br>Staff training: §211.25(a–b) |
| 3. Facilities and equipment | Buildings and facilities: §§211.22(b), 211.28(c), 211.42–211.58, 211.173<br>Equipment: §§211.63–211.72, 211.105, 211.160(b)(4), 211.182<br>Lab facilities: §211.22(b) |
| 4. Control outsourced operations | Consultants: §211.34<br>Outsourcing: §211.22(a) |

Under a quality system model, the specification of facility and equipment requirements may be performed by technical experts (e.g., engineers, development scientists) who have an understanding of the pharmaceutical science, manufacturing processes, and risk factors associated with the product and its manufacture. The cGMP regulations require the QU to be responsible for reviewing and approving all initial design criteria and procedures pertaining to facilities and equipment and any subsequent changes. These requirements are not mutually exclusive; while they place ultimate responsibility for review and approval of these activities with the QU, the regulations do not preclude a cross-functional review involving persons with relevant expertise from multiple areas of the organization. A requirement of both the cGMP and current quality system models is that such review and approval be conducted by persons who are qualified by education, training and experience to do so.

In the control of outsourced operations, the cGMP regulations require that the QU approve or reject products or services provided under a contract. Under current quality system models, the organization must follow a formal vendor qualification process to qualify outsource providers and verify through inspection or other appropriate means that the provider is capable of meeting the requirements of the organization. To comply with the regulation, these operations should be conducted by the QU.

Table 2 compares the major elements of a quality systems approach to addressing resource issues with corresponding requirements in the CGMP regulations.

## 6 MANUFACTURING OPERATIONS

There is significant commonality between the requirements contained in current quality system models such as ISO 9001-2000 and the cGMP regulation requirements for manufacturing operations. The FDA has identified four major elements of a QS approach to manufacturing operations. These are identified and compared to the cGMP requirements in Table 3.

**TABLE 3 21 CFR cGMP Regulations Related to Manufacturing Operations**

| Quality System Element | Regulatory Citation |
|---|---|
| 1. Design and develop product and processes | Production: §211.100(a) |
| 2. Examine Inputs | Materials: §§210.3(b), 211.80–211.94, 211.101, 211.122, 211.125 |
| 3. Perform and monitor operations | Production: §§211.100, 211.103, 211.110, 211.111, 211.113<br>QC criteria: §§211.22(a–c), 211.115(b), 211.160(a), 211.165(d), 211.188<br>QC checkpoints: §§211.22 (a), 211.84(a), 211.87, 211.110(c) |
| 4. Address nonconformities | Discrepancy investigation: §§211.22(a), 211.100, 211.115, 211.192, 211.198Recalls: 21 CFR Part 7 |

### 6.1 Design, Develop, and Document Product and Processes

In a modern quality systems manufacturing environment, the significant characteristics of the product being manufactured should be defined and verified as meeting requirements from design to delivery, and control should be exercised over all changes. This is consistent with the requirements of the cGMP regulation that require quality and manufacturing processes and procedures, and changes to them, to be defined, approved, and controlled. The idea of controlling the design of both product and process is consistent with concepts included in the FDA Pharmaceutical cGMPs for the 21st Century Initiative to assure product safety that focus on the entire product life cycle. No amount of "downstream" control and testing can compensate for a design that results in a product or production process that is incapable of meeting the requirements necessary to assure that the product is safe and effective for its intended use. Documentation is required and can include the following:

- Resources and facilities used
- Procedures to carry out the process
- Identification of the process owner who will maintain and update the process as needed
- Identification and control of important variables
- Quality control measures, necessary data collection, monitoring, and appropriate controls for the product and process
- Any validation activities, including operating ranges and acceptance criteria
- Effects on related process, functions, or personnel

The cGMP regulations include specific packaging and labeling controls, so packaging and labeling requirement, processes, and controls should be included in a QS-based approach to product and process design and development.

Manufacturers and the FDA have expressed concern that existing regulatory requirements (e.g., the need to effect manufacturing process changes through the regulatory submission process) may be excessively rigid and not conducive to innovation regardless of the potential benefits. The FDA acknowledges that the reluctance to pursue potentially innovative changes in pharmaceutical manufacturing can be undesirable from a public health perspective and has published a process analytical technology (PAT) guidance document that is intended to address this by promoting the use of analytical tools to gain process understanding and meet regulatory requirements for validating and controlling manufacturing processes [7].

The PAT guidance document describes a voluntary approach to the design, analysis, and control of manufacturing processes that involves the timely (e.g., in-process) measurement of critical quality and performance attributes of raw and in-process materials and processes, with the goal of ensuring final product quality. The term *analytical* in PAT is broadly interpreted to include the integrated application of chemical, physical, microbiological, mathematical, and risk analysis as appropriate. One goal of PAT is to design and develop well-understood processes that will consistently ensure predefined quality at the end of the manufacturing process. This is consistent with a quality systems approach. PAT should ideally be initiated during the development stage and is intended to be integrated into existing regulatory processes with timely communication with the FDA a key element. The FDA has published the guidance document and other pertinent PAT information on its website at www.fda.gov. Companies interested in PAT methods should contact the FDA. FDA internal implementation of PAT includes the following:

- A PAT team approach of CMC review and cGMP inspections
- Joint training and certification of FDA PAT review, inspection, and compliance staff
- Scientific and technical support for the PAT review, inspection, and compliance staff

Process analytical technology is consistent with the quality systems approach in that it is based on science and engineering principles for assessing and mitigating risks related to poor product and process quality. In the PAT guidance, the FDA indicates that the desired state for pharmaceutical manufacturing may be characterized as follows:

- Product quality and performance are ensured through the design of effective and efficient manufacturing processes
- Product and process specifications are based on a mechanistic understanding of how formulation and process factors affect product performance
- Continuous real-time quality assurance
- Relevant regulatory policies and procedures are tailored to accommodate the most current level of scientific knowledge
- Risk-based regulatory approaches recognize the following:

- The level of scientific understanding of how formulation and manufacturing process factors affect product quality and performance
- The capability of process control strategies to prevent or mitigate the risk of producing a poor-quality product

Process analytical technology is consistent with a modern risk-based data-driven quality systems approach to cGMP compliance.

### 6.2 Inputs

Current QMS models adopt a process-oriented approach to the design and operation of a QMS as a system of interrelated processes, each with inputs and outputs, which are designed to function in a defined way. Some process outputs are inputs to other processes. This concept is easily applied and understood within the manufacturing environment because it is process oriented. Inputs to manufacturing processes include any material that goes into the final product, including materials purchased from vendors for use in manufacturing and in-process materials. Manufacturing operations generally involve multiple processes conducted in a defined manner to produce the finished product. Each process has a set of inputs and produces one or more outputs that may, in turn, be an input to a subsequent process. Each process has an input–output relationship such that changes or variation in one or more inputs will produce an attendant change in the output. Input specifications are established to assure that the final product meets its requirements. A robust quality system will ensure that all inputs to the manufacturing process are suitable for use by establishing quality controls for the receipt and acceptance from qualified vendors, production, storage, and use of all inputs.

The cGMP regulations require either testing or use of a certificate of analysis (COA) plus an identity analysis for the release of materials for manufacturing. The quality systems approach additionally calls for initial supplier qualification based on an objective evaluation and periodic auditing of suppliers based on risk assessment to verify the adequacy of suppliers' quality systems. During the audit, a manufacturer can observe the testing or examinations conducted by the supplier to help determine the reliability of the supplier's COA. Under a QMS model, the QU would normally be responsible for auditing suppliers as part of its overall responsibility for materials acceptance.

Change control involves the evaluation of proposed changes in a systematic way to determine how they would affect process outputs and ultimately the finished product and is an important element of current quality system models. The cGMP regulations require the QU to approve specifications, and certain changes require review and approval by the QU. Under a quality system model, changes to materials (e.g., specification, supplier, or materials handling) should be implemented through a formal change control system involving the documented competent review and approval of the proposed change prior to implementation and communication of changes as appropriate throughout the organization. Manufacturers should also consider how best to assure that changes made by suppliers in supplied materials that may affect the quality of the finished product can be identified and appropriately evaluated by the manufacturer. Such provisions should be included in supplier agreements where possible.

### 6.3 Perform and Monitor Operations

Both the cGMP regulations and quality system models call for the monitoring of critical processes that may be responsible for causing variability during production. The cGMP regulations require written production and process control procedures and specify process control activities that must be performed and documented. Current quality system models also require written procedures, process verification and validation as appropriate, the establishment of appropriate process control measures and documentation. Risk analysis methods and design and development data may be used to establish process control and monitoring requirements. A quality systems approach allows the manufacturer to more efficiently and effectively validate, perform, and monitor operations and ensure that the controls are scientifically sound and appropriate. Production and process controls should be designed to ensure that the finished products have the identity, strength, quality, and purity they purport or are represented to possess. A systems approach will consider all sources of variability from inputs, through manufacturing processes, packaging, labeling, and shipping to assure that the product that is delivered to the user meets quality requirements.

One important aspect of the quality systems approach is the ongoing collection and analysis of quality data to continuously evaluate quality system effectiveness. Historical data, process knowledge, and risk analysis methods can be applied to identify specific data requirements. Trending and other data analysis methods can allow identification of actual and potential sources of nonconformity so that appropriate corrective and preventive actions can be taken in accordance with established change control procedures.

The entire product life cycle should be addressed by the establishment of monitoring and continual improvement mechanisms in the quality system. Even well-defined or mature manufacturing processes may "drift" due to a host of factors including equipment and facility aging, changes or variation in raw materials, electrical power fluctuations, and environmental changes. Thus, process validation is not a one-time event but an activity that continues throughout a product's life. One major quality system objective should be to identify emerging quality problems before nonconformities occur. Trending of periodically collected environmental monitoring data may, for example, identify a slow but steady increase in airborne particulate levels that, if left unaddressed and the trend continues, could exceed a firm's internal environmental standards and adversely affect the product. Early identification of such problems allows an investigation to be initiated to identify the cause so that appropriate corrective and preventive actions can be taken in accordance with established change control procedures. After a change is implemented, its effectiveness should be objectively verified and affected processes revalidated if necessary.

### 6.4 Address Nonconformities

A key component in any quality system is appropriately responding to nonconformities (i.e., deviations from requirements established under the quality system for in-process material or final product quality attributes, process control parameters, records, procedures, etc.). Nonconformities may be detected during any stage of the

manufacturing process or during quality control activities. The cGMP regulations require an investigation to be initiated and that the investigation, conclusion, and follow-up be documented. A primary objective of any manufacturing quality system is to prevent nonconforming product from being produced and distributed. The complete response to nonconformities should be risk based and can include the following components:

- Assessment of how the nonconformity will affect the quality of the finished product (i.e., determination if the nonconformity has resulted, or could result, in product that does not meet its specified purity, potency, and quality characteristics).
- Determine any actions necessary to assure that product that does not meet its specified requirements is not produced and that appropriate steps are taken with regard to any nonconforming product that has been produced to assure that consumers are not harmed and that regulatory requirements are satisfied.
- Determine the cause of the nonconformity.
- Identify any actions needed to correct the cause and to prevent recurrence.
- Document the investigation, findings, and follow-up actions.
- Assess the effectiveness of follow-up actions.
- Repeat the cycle as needed.

A nonconformity may not result in the finished product failing to meets its requirements; however, investigation of the nonconformity may identify process or quality system deficiencies that require attention. For example, a small but unexpected deviation from a process control requirement (e.g., temperature, blending time) may not exceed the limit for which the process was initially validated and thus not be expected to adversely effect the finished product but could suggest an emerging process control or equipment issue that if not corrected could result in future product nonconformities. Similarly, nonconformities in the form of errors or omissions in production records or deviations from written procedures may not always result in product nonconformity but could suggest training, process design, or other issues that ought to be addressed. Thus the response to nonconformities should not be limited to a determination of the immediate impact on the finished product, but also consider its implications regarding overall quality system performance.

## 7 EVALUATION ACTIVITIES

The evaluation component of a QMS is intended to provide objective information and data that allow the organization to assess the conformity of the product, evaluate the performance of its quality system, and maintain and improve its effectiveness [10]. The cGMP regulations similarly require evaluation activities as shown in Table 4.

**TABLE 4 21 CFR cGMP Regulations Related to Evaluation Activities**

| Quality System Element | Regulatory Citation |
|---|---|
| 1. Analyze data for trends | Annual review: §211.180(e) |
| 2. Conduct internal audits | |
| 3. Risk assessment | |
| 4. Corrective action | Discrepancy investigation: §§211.22(a), 211.192 |
| 5. Preventive action | |
| 6. Promote improvement | §211.110 |

### 7.1 Trend Analysis

The cGMP regulations require review and analysis of certain quality data annually at least. Current quality system models emphasize data-based decision making and the use of appropriate statistical analysis methods [2, 11]. Trend analysis is one statistical tool specifically recommended by the FDA in its pharmaceutical QS guidance document that can be very valuable in monitoring processes and quality system performance to identify emerging problems and to assess the effectiveness of improvement efforts. Traditional statistical process control and other methods also provide valuable support in the objective and ongoing analysis of quality data and can be helpful in implementing real-time quality assurance practices as recommended by the FDA [7].

### 7.2 Conduct Internal Audits

Internal auditing is not specifically required by the cGMP regulations, but manufacturers have traditionally used internal audits as a self-assessment tool and to prepare for FDA inspections. The FDA has for some time recognized the value of internal auditing and encourages firms to conduct audits by, as a matter of policy, not reviewing internal audit results during inspections [12].

Current quality system models call for audits to be conducted at planned intervals to evaluate effective implementation and maintenance of the quality system and to determine if processes and products meet established parameters and specifications. International standards provide guidance on auditing [13]. Audit procedures should be developed and documented to ensure that the planned audit schedule takes into account the relative risks of the various quality system activities. Factors that can be incorporated into a risk-based approach to planning audit frequency and scope include the following [6]:

- Existing legal requirements (e.g., cGMPs)
- Overall compliance status and history of the company or facility
- Robustness of a company's quality risk management activities
- Complexity of the site
- Complexity of the manufacturing process
- Complexity of the product and its therapeutic significance
- Number and significance of quality defects (e.g., recall)

- Results of previous audits/inspections that can include prior internal audit results as well as regulatory (e.g., state, federal, or other regulatory agencies) and third-party audits
- Major changes of building, equipment, processes, and key personnel
- Experience with manufacturing of a product (e.g., frequency, volume, number of batches)
- Test results of official control laboratories

In general, auditors should not have direct responsibility over the matters being audited. Auditors should be trained in auditing methods and have sufficient technical knowledge to be able to evaluate the systems being audited using objective audit criteria [14]. Audit criteria may be based on applicable regulatory requirements, standards to which the quality system is intended to conform (e.g., ISO 9001-2000), and the specific requirements of the quality system being audited as indicated in quality system documents. Auditing criteria should be defined prior to the initiation of the audit.

Different audit approaches may be applied depending on the intended purpose and scope of the audit. A top-down approach first evaluates the overall structure of the quality system and its subsystems. Selected subsystems may be chosen for review. Systems identified and developed by the FDA in a six-system inspection model for the inspection of drug manufacturers [15] include the following:

- Overall quality system
- Facilities and equipment
- Materials system
- Production system
- Packaging and labeling
- Laboratory controls

Subsystems must be pertinent to the specific quality system being audited and may coincide with major elements of a standard to which the quality system is intended to conform or the major elements identified in the FDA pharmaceutical QS guidance. When using the top-down approach, the auditor will first review each subsystem to determine if the requirements that apply to that subsystem (e.g., regulatory requirements, the requirements of the standard) are met by defining, documenting, and implementing appropriate procedures. Once the auditor has verified that the requisite procedures are in place, he or she will review the associated records and other documents to verify that the procedures have been followed and documented and that the quality system is functioning effectively as designed and conforms to applicable regulatory requirements and standards. This approach allows for a systematic evaluation of each subsystem and can be as detailed as needed.

A bottom-up approach may be used to follow up on a specific quality problem identified from trend analyses, product nonconformities, adverse experiences, customer complaints, or other sources of quality data. Starting with quality records associated with the problem, the auditor will work his or her way up through the

quality system, examining the quality processes having a bearing on the quality problem. This approach is helpful in identifying quality system issues that may be associated with specific quality problems but does not readily allow evaluation of the entire quality system.

A combination approach may also be used that employs elements of top-down and bottom-up audits. This allows some level of assessment of the effectiveness of the overall quality system while evaluating the cause of specific quality problems.

Auditors should select the audit method most appropriate for their intended audit purpose. Initial quality system audits or regularly scheduled audits are likely candidates for the top-down approach, while audits conducted as part of a root cause analysis, for example, may best employ a bottom-up approach. The FDA employs a similar approach to inspections. Regular scheduled biennial inspections are more likely to employ a top-down methodology. *For cause* inspections conducted in response to a specific product issue such as a recall are more likely to employ a bottom-up approach. FDA investigators may employ a combination approach during biennial inspections if investigators are aware of specific quality problems that they wish to include in the inspection.

Auditing as described in QMS models is intended to assess the effectiveness of the overall quality system as designed and conformance to applicable standards. The overall quality system does not have to be covered in a single audit. Manufacturers may choose to employ a *rolling audit* approach in which specifically identified subsystems are chosen for evaluation in accordance with an approved audit schedule. Audit plans should be designed to effectively perform this assessment.

Compliance with cGMP requirements is also a major concern, and audit planning should include assessment of conformance to cGMP requirements and readiness for FDA inspections. Existing FDA guidance documents and compliance policy guides describe FDA inspectional approaches and policy and can be used for reference in developing audit plans [15–17]. It can be helpful to include mock FDA audits as part of an overall auditing regimen. Some firms prefer to use outside auditors for mock audits to better simulate the FDA inspection process. Mock audits are also useful for training purposes to prepare the organization for FDA inspections.

The audit plan should be consistent with written quality auditing procedures included in the quality manual or other quality system documentation. The plan should include or refer to the objective criteria to be used to evaluate conformance to requirements. The plan should include or refer to other documents that will be used during the audit, including previous audit reports. If the audit is to include the review of batch or production records, such review should be conducted in accordance with a specified sampling plan or other appropriate statistical rationale as specified in a firm's quality system procedures.

Manufacturers implementing a quality system that conforms to an existing standard may find it helpful to create a table or some other document that shows the relationship between cGMP requirements, requirements of the standard, and the element(s) of the manufacturer's quality system. Such a tool can help assure that all pertinent requirements are covered in the quality system design and that audit plans designs include assessment of all pertinent requirements.

Since current quality system models employ a systems approach, an audit checklist that is organized by subsystem may be helpful, as described in Table 5. The form would include appropriate document control information such as form

**TABLE 5 Example Audit Checklist**

| [Company Name] Quality System Audit Checklist | | | |
|---|---|---|---|
| Form: | Rev: | Date: | Approved: |
| Audit Date(s): | Refs: | | |
| Auditor: | | Title | Signature: |
| Requirement | cGMP Section Cross Reference | Conforms (Y/N/NA) | Objective Evidence and Comments |
| Subsystem 1 | | | |
| Requirement 1.1 | | | |
| Requirement 1.2 | | | |
| Subsystem 2 | | | |
| Requirement 2.1 | | | |
| Requirement 2.2 | | | |
| Subsystem 3 | | | |
| Requirement 3.1 | | | |
| Requirement 3.2 | | | |
| Subsystem N | | | |

identification, revision, and approval information. Companies may also wish to include reference information used in planning the audit such as previous audit reports, completed FDA Form 483 Inspectional Observations, third-party audit reports, and pertinent internal QS documents (e.g., audit procedures). Depending on the purpose of the audit, the subsystems may correspond to the six subsystems identified by the FDA for use by investigators in conducting cGMP inspections (i.e., quality, production, facilities and equipment, laboratory controls, materials, packaging and labeling) or the major elements of a quality system standard. Cross references between elements of the standard being used and the pertinent sections of the cGMP regulations may be included as appropriate. The audit form should allow entry of information regarding conformance or nonconformance to each requirement and have space for a description of pertinent findings.

The QMS models require periodic audits but do not specify audit frequency. Audit frequency must be determined based on the risk associated with the matters to be audited and other factors including results of previous audits and other quality data. Periodic audits should be conducted over the entire product life cycle and follow-up audits conducted as appropriate to verify that previously identified quality problems have been corrected in accordance with applicable quality system and regulatory requirements.

### 7.3 Quality Risk Management

The FDA has endorsed quality risk management as part of an overall quality systems approach to compliance with the cGMP regulations and achieving overall

quality system objectives [6]. Risk management methodologies permit management to assign priorities to activities or actions based on an assessment of the risk including both the probability of occurrence of harm and the severity of that harm.

Implementation of quality risk management includes assessing the risks, selecting and implementing risk management controls commensurate with the level of risk, and evaluating the results of the risk management efforts. In a manufacturing quality systems environment, risk management is used as a tool in the development of product specifications and critical process parameters. Used in conjunction with process understanding, quality risk management helps manufacturers effectively manage and control change.

A formal risk management process consists of several components:

- Risk assessment
  - Risk identification
  - Risk analysis
  - Risk evaluation
- Risk control
  - Risk reduction
  - Risk acceptance
- Risk communication
- Risk review

Risk assessment starts with risk identification, a systematic use of available information to identify hazards (i.e., events or other conditions that have the potential to cause harm). Information can be from a variety of sources including stakeholders, historical data, information from the literature, and mathematical or scientific analyses. Risk analysis is then conducted to estimate the degree of risk associated with the identified hazards. This is estimated based on the likelihood of occurrence and resultant severity of harm. In some risk management tools, the ability to detect the hazard may also be considered. If the hazard is readily detectable, this may be considered a factor in the overall risk assessment. Risk evaluation determines if the risk is acceptable based on specified criteria. In a quality system environment, criteria would include impact on the overall performance of the quality system and the quality attributes of the finished product. The value of the risk assessment depends on how robust the data used in the assessment process is judged to be. The risk assessment process should take into account assumptions and reasonable sources of uncertainty. Risk assessment activities should be documented.

Risk control starts with risk reduction, which includes any actions taken to eliminate or reduce the risk. Actions taken should be commensurate with the significance of the risk. If the risk has been reduced to an acceptable level, an affirmative decision can be made to accept the risk (risk acceptance). One question to ask is if new risks have been introduced as a result of the identified risks being controlled. Risk control measures should generally be conducted in accordance with change control procedures and documented.

Risk communication involves the communication of appropriate information about the risk to stakeholders (e.g., others involved in or affected by the quality

system including management, users, regulatory agencies). Risk communication should be documented. The included information might relate to the existence, nature, form, probability, severity, acceptability, control, treatment, detectability, or other aspects of risks to quality. Communication should be as appropriate and does not necessarily need to be carried out for each and every risk acceptance.

Risk review should be conducted to evaluate the outputs of the risk management process and repeated as necessary, based on new quality data or if there are process or product changes.

The Q9 Quality Risk Management guidance document [6] identifies a number of risk management tools that manufacturers can apply, including failure mode effects and criticality analysis (FMECA), hazard analysis and critical control points (HAACP), and preliminary hazard analysis (PHA), and provides examples of how quality risk management might be applied to quality management, development, materials management, production, and other operations within the organization.

### 7.4 Corrective and Preventive Actions

Corrective and preventive action (CAPA) is the term commonly used to describe the subsystem of a comprehensive quality system that deals with the systematic investigation, understanding, and response to quality issues including nonconformities. A corrective or preventive action may be initiated based on review and analysis of quality data from a variety of sources including adverse experiences, product complaints, quality audits, FDA inspections, third-party inspections, nonconforming materials reports, process control information, trend analyses, and other sources.

A corrective action is initiated to correct the cause of an identified nonconformity and to prevent it or similar problems from reoccurring. It may include initial and follow-up actions (e.g., conducted after root cause analysis). Current quality system models and the cGMP regulations emphasize corrective actions and require that actions be documented. Under current quality system models, preventive actions include actions taken in response to quality data to address the cause of potential nonconformities to prevent their occurrence. An effective CAPA system therefore includes both reactive and proactive components. The effectiveness of corrective and preventive actions should be evaluated using objective criteria when possible and the evaluation documented.

A firm's CAPA system and processes should be designed to analyze and respond to quality issues in a systematic way that is commensurate with the risk. The system should provide for the verification or validation of corrective and preventive actions to assure their effectiveness and to assure that actions do not adversely affect the finished product. The system should also assure that pertinent CAPA information is appropriately disseminated throughout the organization as necessary to assure the effective operation of the quality system and for management review.

### 7.5 Promote Improvement

Continual improvement is a requirement of existing quality system models such as ISO 9001-2000 in which the organization is required to continually improve the effectiveness of the quality management system through the use of the quality

policy, quality objectives, audit results, analysis of data, corrective and preventive actions, and management review. In adapting the ISO 9001-2000 standard to serve as a regulatory standard for medical device quality management systems, drafters of the ISO 13485 standard altered the requirement slightly to require the organization to "identify and implement any changes necessary to ensure and maintain the continued suitability and effectiveness of the quality management system through the use of the quality policy, quality objectives, audit results, analysis of data, corrective and preventive actions, and management review." The word *improvement* was deleted as not an objective of current regulatory standards, but the concept of continually monitoring the performance of the quality system and appropriately responding to quality data was retained.

The cGMP regulation does not specifically require continual improvement; however, the regulations are specific with regard to the sampling and testing of in-process materials and drug products, and failure to take reasonable action to reduce identified sources of variability may be of concern to FDA investigators. The FDA in its pharmaceutical QS guidance document encourages organizations to promote improvement through quality system activities and notes that it is critical for senior management to be involved. Process improvement, along with improvement of in-process controls, can render a manufacturing process more efficient and more robust. The end result can reduce costs and further prevent product failures and defects from occurring.

## 8 TRANSITIONING TO QUALITY SYSTEMS APPROACH

The cGMP regulations assign significant responsibilities to the organizational unit responsible for quality-related activities. Organizations implementing a quality system model will be responsible for additional quality-related activities including, but not necessarily limited to, conducting quality audits, analysis of quality data, risk assessment, and preventive actions based on review and analysis of quality data to prevent the occurrence of product nonconformities. In addition, management is required to provide requisite leadership by actively participating in the quality system and assuring that the quality system functions as intended. This is accomplished by establishing a quality policy and associated objectives, planning for quality, establishing an appropriate organization structure with designated responsibilities and authorities to appropriately carry out quality system requirements, providing appropriate resources and training, and periodically reviewing quality information and data, and assuring that the organization responds appropriately.

The organizational unit responsible for quality-related activities will in all likelihood have an even greater role within the organization, and roles and responsibilities throughout the organization are likely to change. Careful planning will be required to assure that the transition is effected smoothly with no adverse impact on product quality. Following are some points to consider in planning the transition:

- Create a transition team: A cross-functional team should be developed involving key managers and staff from throughout the organization to plan and

execute the transition. The transition team should have a clear understanding of its mission and the organizational objectives associated with the transition.

- Train the transition team: The decision to make the transition must come from management and management should assure that all individuals on the transition team receive proper training on quality systems requirements, risk management, and FDA's recommended approach to quality systems.
- Develop a transition plan: A transition plan, based on clearly defined objectives, should be developed by the transition team.
- Identify staffing requirements: The transition will likely affect individual job descriptions and create additional duties that will have to be addressed through the reassignment of staff, hiring new staff, and providing necessary training to all affected staff.
- Identify other resource needs: The plan should include a definition of resource requirements for planning and executing the plan.
- Define roles and responsibilities: the plan should clearly define the roles and responsibilities of those responsible for development and execution of the plan for quality system implementation as well as staff roles and responsibilities under the quality system.
- Consider organizational structure requirements: In order to function properly, persons responsible for quality-related activities must have the responsibility and associated authority defined and appropriately communicated within the organization.
- Conduct a gap analysis: The plan should conduct a gap analysis that identifies how the quality system model chosen can be effectively integrated with existing processes to create a quality system that conforms to the organization's quality objectives, meets regulatory requirements, and is consistent with other organizational requirements. The quality systems approach is intended to be somewhat flexible in application and can be tailored to specific organizational requirements. In order to function properly the quality system must be effectively integrated into the organization so that it is not viewed as an "add-on" or a set of extra requirements that prevent the "real" work from getting done.
- Consider benchmarking: If possible, arrange with other organizations that have successfully made the transition to meet with them, review their system, and discuss transition issues and how they were solved.
- Consult with experts: In addition to benchmarking, seeking assistance from persons familiar with quality systems can be very helpful, particularly when existing staff are relatively inexperienced with quality systems. It may be useful for one or more outside experts to work with the transition team on a regular basis as a coach or facilitator.
- Communicate regularly: Clear and ongoing communication within the transition team and with management is essential to effectively coordinate plan activities, report progress, resolve issues, and identify evolving resource needs.
- Sell the system: Successful implementation of a QS requires the active and informed participation of many individuals within the organization. Manage-

ment commitment should be clearly communicated and training provided so that affected staff understand basic quality system concepts and their role in the quality system.
- Validate the system.
- Maintain regulatory compliance.

## 9 AUDIT CHECKLIST FOR DRUG INDUSTRY

The checklist provided in Table 6 [15] is intended to aid in the systematic GMP audit of a facility that manufactures drug components or finished products.

The adequacy of any procedures is subject to the interpretation of the auditor. Therefore, the author accepts no liability for any subsequent regulatory observations or actions stemming from the use of this audit checklist.

### 9.1 Instructions for Using Audit Checklist

Before starting an on-site audit, plan the audit. Review past audits, note indications of possible problem areas and items, if any, that were identified for corrective action in a previous audit. If you are not already familiar with this facility, learn the type of product produced and how it is organized by personnel and function. What does your "customer," that is, your superior or senior facility management, expect to learn from this audit?

1. The checklist is to be used with a notebook into which detailed entries can be made during the audit.
2. While the checklist is to guide the auditor, it is not intended to be a substitute for knowledge of the GMP regulations.
3. Although a single question may be included about any requirement, the answer will usually be a multipart one since the auditor should determine the audit trail for several products that may use many different components. Enter details in you notebook and cross reference your comments with the questions.
4. At least three production batches should be selected for thorough analysis to include: (a) traceability of all components or materials used in the subject batches, (b) documentation of raw material or component, in-process, and finished goods testing for the subject product batches, and (c) warehousing and distribution records as they would relate to a possible recall.
5. Responses entered on the checklist should be consistent. "X" is recommended for "No"; a checkmark for "Yes"; "N/A" for not applicable to questions that do not apply. An asterisk and notebook page number should be entered on the checklist to identify where relevant comments or questions are recorded in your notebook.
6. The notebook used should be a laboratory-type notebook with bound pages. The notebook should be clearly labeled as to the audit type, date, and auditor(s). Many auditors prefer to use a notebook for a single audit so it may be filed with the checklist and the final report.

**TABLE 6 Audit Checklist**

| Question | Instructions/Questions (note any exceptions and comments in notebook) | Yes, No, or NA |
|---|---|---|
| **1.0** | **General Controls** | |
| | Does the facility and its departments (organizational units) operate in a state of control as defined by the GMP regulations? | |
| **1.1** | **Organizational & Management Responsibilities** | |
| 1.101 | Does this facility/business unit operate under a facility or corporate quality policy? | |
| 1.102 | §211.22(a) Does a Quality Assurance unit (department) exist as a separate organizational entity? | |
| 1.103 | §211.22(a) Does the Quality Assurance unit alone have both the authority and responsibility to approve or reject all components, drug product containers and closures, in-process materials, packaging materials, labeling, and drug products? | |
| 1.104 | §211.22 Does the QA department or unit routinely review production records to ensure that procedures were followed and properly documented? | |
| 1.105 | §211.22(b) Are adequate laboratory space, equipment, and qualified personnel available for required testing? | |
| 1.106 | If any portion of testing is performed by a contractor, has the Quality Assurance unit inspected the contractor's site and verified that the laboratory space, equipment, qualified personnel, and procedures are adequate? | |
| 1.107 | Date of last inspection: — | |
| 1.108 | §211.22(c) Are all QA procedures in writing? | |
| 1.109 | §211.22(c) Are all QA responsibilities in writing? | |
| 1.110 | Are all written QA procedures current and approved? (Review log of procedures) | |
| 1.111 | Are the procedures followed? (Examine records to ensure consistent record-keeping that adequately documents testing.) | |
| 1.112 | §211.25 Are QA supervisory personnel qualified by way of training and experience? | |
| 1.113 | §211.25 Are other QA personnel (e.g., chemists, analysts, laboratory technicians) qualified by way of training and experience? | |
| **1.2** | **Document Control Program** | |
| 1.201 | §211.22(a) Does the QA unit have a person or department specifically charged with the responsibility of designing, revising, and obtaining approval for production and testing procedures, forms, and records? | |
| 1.202 | §211.22(d) Does a written SOP, which identifies how the form is to be completed and who signs and countersigns, exist for each record or form? | |
| 1.203 | §211.165(a)(b)(c) Is the production batch record and release test results reviewed for accuracy and completeness *before* a batch/lot of finished product is released? | |

**TABLE 6** ***Continued***

| Question | Instructions/Questions (note any exceptions and comments in notebook) | Yes, No, or NA |
|---|---|---|
| **1.3** | **Employee Orientation, Quality Awareness, and Job Training** | |
| 1.301 | Circle the types of orientation provided to each new employee: (1) Company brochure. (2) Literature describing GMP regulations and stressing importance of following instructions. (3) On-the-job training for each function to be performed (*before* the employee is allowed to perform such tasks). (4) Other: enter in notebook. | |
| 1.302 | §211.25(a) Does each employee receive retraining on an SOP (procedures) if critical changes have been made in the procedure? | |
| 1.303 | Indicate how ongoing, periodic GMP training is accomplished. | |
| 1.304 | §211.25 is all training documented in writing that indicates the date of the training, the type of training, and the signature of both the employee and the trainer? | |
| 1.305 | §211.25 Are training records readily retrievable in a manner that enables one to determine what training an employee has received, which employees have been trained on a particular procedure, or have attended a particular training program? | |
| 1.306 | Are GMP trainers qualified through experience and training? | |
| 1.307 | §211.25(a) Are supervisory personnel instructed to prohibit any employee who, because of any physical condition (as determined by medical examination or supervisory observation) that may adversely affect the safety or quality of drug products, from coming into direct contact with any drug component or immediate containers for finished product? | |
| 1.308 | §211.28(d) Are employees required to report to supervisory personnel any health or physical condition that may have an adverse effect on drug product safety and purity? | |
| 1.309 | §211.25(a) Are temporary employees given the same orientation as permanent employees? | |
| 1.310 | §211.34 Are consultants, who are hired to advise on any aspect of manufacture, processing, packing or holding, of approval for release of drug products, asked to provide evidence of their education, training, and experience? | |
| 1.311 | §211.34 Are written records maintained stating the name, address, qualifications, and date of service for any consultants and the type of service they provide? | |
| **1.4** | **Plant Safety and Security** | |
| 1.401 | Does this facility have a facility or corporate safety program? | |
| 1.402 | Are safety procedures written? | |
| 1.403 | Are safety procedures current? | |

**TABLE 6** ***Continued***

| Question | Instructions/Questions (note any exceptions and comments in notebook) | Yes, No, or NA |
|---|---|---|
| 1.404 | Do employees receive safety orientation *before* working in the plant area? | |
| 1.405 | Is safety training documented in a readily retrievable manner that states the name of the employee, the type of training, the date of the training, and the name of the trainer and the signature of the trainer and the participant? | |
| 1.406 | Does this facility have a formal, written security policy? | |
| 1.407 | Is access to the facility restricted? | |
| 1.408 | Describe how entry is monitored/restricted: | |
| 1.409 | Is a security person available 24 hours per day? | |
| **1.5** | **Internal Quality/GMP Audit Program** | |
| 1.501 | Does this business unit/facility have a written quality policy? | |
| 1.502 | Is a copy of this quality policy furnished to all employees? | |
| 1.503 | If "yes" to above, when provided? — | |
| 1.504 | Is training provided in quality improvement? | |
| 1.505 | Does a formal auditing function exist in the Quality Assurance department? | |
| 1.506 | Does a written SOP specify who shall conduct audits and qualifications (education, training, and experience) for those who conduct audits? | |
| 1.507 | Does a written SOP specify the scope and frequency of audits and how such audits are to be documented? | |
| 1.508 | Does a written SOP specify the distribution of the audit report? | |
| **1.6** | **Quality Cost Program** | |
| 1.601 | Does this facility have a periodic and formal review of the cost of quality? | |
| 1.602 | Does this facility have the ability, through personnel, software, and accounting records, to identify and capture quality costs? | |
| 1.603 | Does this facility make a conscious effort to reduce quality costs? | |
| 2.0 | Design control | |
| | Not directly related to the drug regulation | |
| 3.0 | Facility control | |
| **3.1** | **Facility Design and Layout** | |
| 3.101 | §211.42(a) Are all parts of the facility constructed in a way that makes them suitable for the manufacture, testing, and holding of drug products? | |
| 3.102 | §211.42(b) Is there sufficient space in the facility for the type of work and typical volume of production? | |
| 3.103 | Does the layout and organization of the facility prevent contamination? | |
| **3.2** | **Environmental Control Program** | |
| 3.201 | The facility is NOT situated in a location that potentially subjects workers or product to particulate matter, fumes, or infestations? | |

**TABLE 6** ***Continued***

| Question | Instructions/Questions (note any exceptions and comments in notebook) | Yes, No, or NA |
|---|---|---|
| 3.202 | Are grounds free of standing water? | |
| 3.203 | §211.44 Is lighting adequate in all areas? | |
| 3.204 | §211.46 Is adequate ventilation provided? | |
| 3.205 | §211.46 Is control of air pressure, dust, humidity, and temperature adequate for the manufacture, processing, storage, or testing of drug products? | |
| 3.206 | §211.46 If air filters are used, is there a written procedure specifying the frequency of inspection and replacement? | |
| 3.207 | Are drains and routine cleaning procedures sufficient to prevent standing water inside the facility? | |
| 3.208 | §211.42(d) Does the facility have separate air-handling systems, if required, to prevent contamination? (MANDATORY IF PENICILLIN IS PRESENT!) | |
| **3.3** | **Facility Maintenance and Good Housekeeping Program** | |
| 3.301 | §211.56(a) Is this facility free from infestation by rodents, birds, insects, and vermin? | |
| 3.302 | §211.56(c) Does this facility have written procedures for the safe use of suitable (e.g., those that are properly registered) rodenticides, insecticides, fungicides, and fumigating agents? | |
| 3.303 | Is this facility maintained in a clean and sanitary condition? | |
| 3.304 | Does this facility have written procedures that describe in sufficient detail the cleaning schedule, methods, equipment, and material? | |
| 3.305 | Does this facility have written procedures for the safe and correct use of cleaning and sanitizing agents? | |
| 3.306 | §211.58 Are all parts of the facility maintained in a good state of repair? | |
| 3.307 | §211.52 Is sewage, trash, and other refuse disposed of in a safe and sanitary manner (and with sufficient frequency)? | |
| **3.4** | **Outside Contractor Control Program** | |
| 3.401 | §211.56(d) Are contractors and temporary employees required to perform their work under sanitary conditions? | |
| 3.402 | Are contractors qualified by experience or training to perform tasks that may influence the production, packaging, or holding of drug products? | |
| 4.0 | Equipment control | |
| **4.1** | **Equipment Design and Placement** | |
| 4.101 | §211.63 Is all equipment used to manufacture, process, or hold a drug product of appropriate design and size for its intended use? | |
| 4.102 | Are the following pieces of equipment suitable for their purpose: blender(s), conveyor(s), tablet, presses, capsule fillers, bottle fillers, other (specify)? | |
| 4.103 | Are the following pieces of equipment suitable in their size/capacity: blender(s), conveyor(s), tablet, presses, capsule fillers, bottle fillers, other (specify)? | |

**TABLE 6** ***Continued***

| Question | Instructions/Questions (note any exceptions and comments in notebook) | Yes, No, or NA |
|---|---|---|
| 4.104 | Are the following pieces of equipment suitable in their design: blender(s), conveyor(s), tablet, presses, capsule fillers, bottle fillers, other (specify)? | |
| 4.105 | Are the locations in the facility of the following pieces of equipment acceptable: blender(s), conveyor(s), tablet, presses, capsule fillers, bottle fillers, other (specify)? | |
| 4.106 | Are the following pieces of equipment properly installed: blender(s), conveyor(s), tablet, presses, capsule fillers, bottle fillers, other (specify)? | |
| 4.107 | Is there adequate space for the following pieces of equipment: blender(s), conveyor(s), tablet, presses, capsule fillers, bottle fillers, other (specify)? | |
| 4.108 | §211.65(a) Are machine surfaces that contact materials or finished goods nonreactive, nonabsorptive, and nonadditive so as not to affect the product? | |
| 4.109 | §211.65(b) Are design and operating precautions taken to ensure that lubricants or coolants or other operating substances do NOT come into contact with drug components or finished product? | |
| 4.110 | §211.72 Fiber-releasing filters are NOT used in the production of injectable products. | |
| 4.111 | §211.72 Asbestos filters are NOT used in the production of products. | |
| 4.112 | Is each idle piece of equipment clearly marked "needs cleaning" or "cleaned; ready for service"? | |
| 4.113 | Is equipment cleaned promptly after use? | |
| 4.114 | Is idle equipment stored in a designated area? | |
| 4.115 | §211.67(a)(b) Are written procedures available for each piece of equipment used in the manufacturing, processing, or holding of components, in-process material, or finished product? | |
| 4.116 | Do cleaning instructions include disassembly and drainage procedure, if required, to ensure that no cleaning solution or rinse remains in the equipment? | |
| 4.117 | Does the cleaning procedure or startup procedure ensure that the equipment is systematically and thoroughly cleaned? | |
| **4.2** | **Equipment Identification** | |
| 4.201 | §211.105 Are all pieces of equipment clearly identified with easily visible markings? | |
| 4.202 | §211.105(b) Are all pieces of equipment also marked with an identification number that corresponds with an entry in an equipment log? | |
| 4.203 | Does each piece of equipment have written instructions for maintenance that includes a schedule for maintenance? | |
| 4.204 | Is the maintenance log for each piece of equipment kept on or near the equipment? | |

**TABLE 6** ***Continued***

| Question | Instructions/Questions (note any exceptions and comments in notebook) | Yes, No, or NA |
|---|---|---|
| **4.3** | **Equipment Maintenance & Cleaning** | |
| 4.301 | §211.67(b) Are written procedures established for the cleaning and maintenance of equipment and utensils? | |
| 4.302 | Are these procedures followed? | |
| 4.303 | §211.67(b)(1) Does a written procedure assign responsibility for the cleaning and maintenance of equipment? | |
| 4.304 | §211.67(b)(2) Has a written schedule been established and is it followed for the maintenance and cleaning of equipment? | |
| 4.305 | Has the cleaning procedure been properly validated? | |
| 4.306 | §211.67(b)(2) If appropriate, is the equipment sanitized using a procedure written for this task? | |
| 4.307 | §211.67(b)(3) Has a sufficiently detailed cleaning and maintenance procedure been written for each different piece of equipment to identify any necessary disassembly and reassembly required to provide cleaning and maintenance? | |
| 4.308 | §211.67(b)(3) Does the procedure specify the removal or obliteration of production batch information from each piece of equipment during its cleaning? | |
| 4.309 | Is equipment cleaned promptly after use? | |
| 4.310 | Is clean equipment clearly identified as "clean" with a cleaning date shown on the equipment? | |
| 4.311 | §211.67(b)(5) Is clean equipment adequately protected against contamination prior to use? | |
| 4.312 | §211.67(b) Is equipment inspected immediately prior to use? | |
| 4.313 | §211.67(c) Are written records maintained on equipment cleaning, sanitizing, and maintenance on or near each piece of equipment? | |
| **4.4** | **Measurement Equipment Calibration Program** | |
| 4.401 | §211.68(a) Does the facility have approved written procedures for checking and calibration of each piece of measurement equipment? (Verify procedure and log for each piece of equipment and note exceptions in notebook with cross reference.) | |
| 4.402 | §211.68(a) Are records of calibration checks and inspections maintained in a readily retrievable manner? | |
| **4.5** | **Equipment Qualification Program** | |
| 4.501 | §211.63 Verify that all pieces of equipment used in production, packaging, and quality assurance are capable of producing valid results. | |
| 4.502 | §211.68(a) When computers are used to automate production or quality testing, have the computer and software been validated? | |
| 4.503 | Have on-site tests of successive production runs or tests been used to qualify equipment? | |
| 4.504 | Were tests repeated a sufficient number of times to ensure reliable results? | |

**TABLE 6** ***Continued***

| Question | Instructions/Questions (note any exceptions and comments in notebook) | Yes, No, or NA |
|---|---|---|
| 4.505 | §211.63 Is each piece of equipment identified to its minimum and maximum capacities and minimum and maximum operating speeds for valid results? | |
| 4.506 | Have performance characteristics been identified for each piece of equipment? (May be provided by the manufacturer but must be verified under typical operations conditions.) | |
| 4.507 | Have operating limits and tolerances for performance been established from performance characteristics? | |
| 5.0 | Material/component control | |
| **5.1** | **Material/Component Specification and Purchasing Control** | |
| | Although purchasing is not specifically addressed in the current GMP regulation, incumbent upon user of components and materials to ensure quality of product, material, or component. | |
| 5.101 | Has each supplier/vendor of material or component been inspected/audited for proper manufacturing controls? (Review suppliers and audits and enter names, material supplied, and date last audited in notebook.) | |
| **5.2** | **Material/Component Receipt, Inspection, Sampling, and Laboratory Testing** | |
| 5.201 | §211.80(a) Does the facility have current written procedures for acceptance/rejections of drug products, containers, closures, labeling, and packaging materials? (List selected materials and components in notebook and verify procedures.) | |
| 5.202 | §211.80(d) Is each lot within each shipment of material or components assigned a distinctive code so material or component can be traced through manufacturing and distribution? | |
| 5.203 | §211.82(a) Does inspection start with visual examination of each shipping container for appropriate labeling, signs of damage, or contamination? | |
| 5.204 | §211.82(b) Is the number of representative samples taken from a container or lot based on statistical criteria and experience with each type of material or component? | |
| 5.205 | §211.160(b) Is the sampling technique written and followed for each type of sample collected? | |
| 5.206 | Is the quantity of sample collected sufficient for analysis and reserve in case retesting or verification is required? | |
| | Verify that the following steps are included in written procedures unless more specific procedures are followed: | |
| 5.207 | §211.84(c)(2) Containers are cleaned before samples are removed. | |
| 5.208 | §211.84(c)(4) Stratified samples are not composited for analysis. | |
| 5.209 | §211.84(c)(5) Containers from which samples have been taken are so marked indicating date and approximate amount taken. | |

**TABLE 6** ***Continued***

| Question | Instructions/Questions (note any exceptions and comments in notebook) | Yes, No, or NA |
|---|---|---|
| 5.210 | Each sample container is clearly identified by material or component name, lot number, date sample taken, name of person taking sample, and original container identification. | |
| 5.211 | §211.84(d)(1)(2) At least one test is conducted to confirm the identity of a raw material (bulk chemical or pharmaceutical) when a Certificate of Analysis is provided by supplier and accepted by QA. | |
| 5.212 | If a Certificate of Analysis is not accepted for a lot of material, then additional testing is conducted by a written protocol to determine suitability for purpose. | |
| 5.213 | §211.84(d)(6) Microbiological testing is conducted where appropriate. | |
| **5.3** | **Material Component Storage and Handling** | |
| | Verify that materials and components are stored and handled in a way that prevents contamination, mixups, and errors. | |
| 5.301 | §211.42(b) Are incoming material and components quarantined until approved for use? | |
| 5.302 | Are all materials handled in such a way to prevent contamination? | |
| 5.303 | Are all materials stored off the floor? | |
| 5.304 | Are materials spaced to allow for cleaning and inspection? | |
| 5.305 | §211.122(d) Are labels for different products, strengths, dosage forms, etc., stored separately with suitable identification? | |
| 5.306 | Is label storage area limited to authorized personnel? | |
| 5.307 | §211.89 Are rejected components, material, and containers quarantined and clearly marked to prevent their use? | |
| **5.4** | **Inventory Control Program** | |
| 5.401 | §211.142 Are inventory control procedures written? | |
| 5.402 | Does the program identify destruction dates for obsolete or out-dated materials, components, and packaging materials? | |
| 5.403 | §211.150(a) Is stock rotated to ensure that the oldest approved product or material is used first? | |
| 5.404 | §211.184(e) Is destruction of materials documented in a way that clearly identifies the material destroyed and the date on which destruction took place? | |
| **5.5** | **Vendor (Supplier) Control Program** | |
| 5.501 | Are vendors periodically inspected according to a written procedure? | |
| 5.502 | Is the procedure for confirming vendor test results written and followed? | |
| 6.0 | Operational control | |

**TABLE 6** ***Continued***

| Question | Instructions/Questions (note any exceptions and comments in notebook) | Yes, No, or NA |
|---|---|---|
| **6.1** | **Material/Component/Label Verification, Storage, and Handling** | |
| 6.101 | §211.87 Do written procedures identify storage time beyond which components, containers, and closures must be reexamined before use? | |
| 6.102 | §211.87 Is release of retested material clearly identified for use? | |
| 6.103 | Are retesting information supplements originally obtained? | |
| 6.104 | Do written procedures identify steps in the dispensing of material for production? | |
| 6.105 | Do these procedures include (1) release by QC, (2) documentation of correct weight or measure, and (3) proper identification of containers? | |
| 6.106 | Does a second person observe weighing/measuring/dispensing and verify accuracy with a second signature? | |
| 6.107 | §211.101(c) Is the addition of each component documented by the person adding the material during manufacturing? | |
| 6.108 | §211.101(d) Does a second person observe each addition of material and document verification with a second signature? | |
| 6.109 | §211.125(a) Does a written procedure specify who is authorized to issue labels? | |
| 6.110 | §211.125(a) Does a written procedure specify how labels are issued, used, reconciled with production, returned when unused, and the specific steps for evaluation of any discrepancies? | |
| 6.111 | §211.125(d) Do written procedures call for destruction of excess labeling on which lot or control numbers have been stamped or imprinted? | |
| **6.2** | **Equipment/Line/Area Cleaning, Preparation, and Clearance** | |
| 6.201 | §211.67(b)(5) Do written procedures detail how equipment is to be checked immediately prior to use for cleanliness, removal of any labels, and labeling from prior print operations? | |
| 6.202 | §211.67(b)(3) Do written procedures detail any disconnection and reassembly required to verify readiness for use? | |
| **6.3** | **Operational Process Validation and Production Change Order Control** | |
| 6.301 | Have production procedures been validated? (Review selected procedures for validation documentation. Adequate?) | |
| 6.302 | §211.100(a) Does the process control address all issues to ensure identity, strength, quality, and purity of product? | |
| 6.303 | §§211.101(a) Does the procedure include formulation that is written to yield not less than 100% of established amount of active ingredients? | |

**TABLE 6** ***Continued***

| Question | Instructions/Questions (note any exceptions and comments in notebook) | Yes, No, or NA |
|---|---|---|
| 6.304 | §211.101(c) Are all weighing and measuring preformed by one qualified person and observed by a second person? | |
| 6.305 | §211.101(d) Have records indicated preceding policy been followed by presence of two signatures? | |
| 6.306 | §211.103 Are actual yields calculated at the conclusion of appropriate phases of the operation and at the end of the process? | |
| 6.307 | §211.103 Are calculations performed by one person? Is there independent verification by a second person? | |
| **6.4** | **In-Process Inspection, Sampling, and Laboratory Control** | |
| 6.401 | §211.110(a) Are written procedures established to monitor output and validate the performance of manufacturing procedures that may cause variability in characteristics of in-process materials and finished drug products? | |
| 6.402 | §211.110(c) Are in-process materials tested at appropriate phases for identity, strength, quality, purity, and are they approved or rejected by Quality Control? | |
| 6.403 | §211.160(b) Are there laboratory controls including sampling and testing procedures to assure conformance of components, containers, closures, in-process materials, and finished product specifications? | |
| **6.5** | **Reprocessing/Disposition of Materials** | |
| 6.501 | §211.115(a) Do written procedures identify steps for reprocessing batches? | |
| 6.502 | §211.115(b) Are quality control review and approval required for any and all reprocessing of material? | |
| 6.503 | Does testing confirm that reprocessed batches conform to established specification? | |
| 6.504 | Does a written procedure outline steps required to reprocess returned drug products (if it can be determined that such products have not been subjected to improper storage conditions)? | |
| 6.505 | Does Quality Control review such reprocessed returned goods and test such material for conformance to specifications before releasing such material for resale? | |
| 7.0 | Finished product control | |
| **7.1** | **Finished Product Verification, Storage, and Handling** | |
| 7.101 | §211.30 Do written procedures indicate how and who verifies that correct containers and packages are used for finished product during the finishing operation? | |
| 7.102 | §211.134(a) In addition, do written procedures require that representative sample of units be visually examined upon completion of packaging to verify correct labeling? | |
| 7.103 | §211.137(a) Are expiration dates stamped or imprinted on labels? | |
| 7.104 | §211.137(b) Are expiration dates related to any storage conditions stated on the label? | |

**TABLE 6** ***Continued***

| Question | Instructions/Questions (note any exceptions and comments in notebook) | Yes, No, or NA |
|---|---|---|
| 7.105 | §211.142(a) Are all finished products held in quarantine until QC has completed its testing and releases product on a batch-to-batch basis for sale? | |
| 7.106 | §211.142(o) Is finished product stored under appropriate conditions of temperature, humidity, light, etc. | |
| **7.2** | **Finished Product Inspection, Sampling, Testing, and Release for Distribution** | |
| 7.201 | §211.166 Has the formulation for each product been tested for stability based on a written protocol? (Containers must duplicate those used in final product packaging.) | |
| 7.202 | §211.166 Are written sampling and testing procedures and acceptance criteria available for each product to ensure conformance to finished product specifications? | |
| 7.203 | §211.170(a) Is a quantity of samples equal to at least twice the quantity needed for finished product release testing maintained as a reserve sample? | |
| 7.204 | §211.167(a) Are sterility and pyrogen testing performed as required? | |
| 7.205 | §211.167(b) Are specific tests for foreign particles or abrasives included for any ophthalmic ointments? | |
| 7.206 | §211.167(c) Do controlled release or sustained release products include tests to determine conformance to release time specification? | |
| **7.3** | **Distribution Controls** | |
| 7.301 | §211.150(a) Does a written procedure manage stocks to ensure that oldest approved product is sold first? | |
| 7.302 | §211.150(a) Are deviations to the policy above documented? | |
| 7.303 | §211.150(a) Does a written procedure identify the steps required if a product recall is necessary? | |
| 7.304 | Is the recall policy current and adequate? | |
| **7.4** | **Marketing Controls** | |
| 7.401 | The current regulation does not address marketing controls per se except that all finished products must meet their specifications. | |
| **7.5** | **Complaint Handling and Customer Satisfaction Program** | |
| 7.501 | §211.198(a) Are complaints, whether received in oral or written form, documented in writing, and retained in a designated file? | |
| 7.502 | §211.198(a) Are complaints reviewed on a timely basis by the Quality Control unit? | |
| 7.503 | §211.198(b)(1) Is the action taken in response to each complaint documented? | |
| 7.504 | §211.198(b)(3) Are decisions not to investigate a complaint also documented and the name of the responsible person documented? | |
| 7.505 | §211.198(b)(2) Are complaint investigations documented and do they include investigation steps, findings, and follow-up steps, if required? Are dates included for each entry? | |

7. The references to sections in the GMP regulation are for your convenience should a question arise. In some instances, two or more sections within the GMP regulation may have bearing on a specific subject. The headings in the GMP regulation will usually offer some guidance on the areas covered in each section.
8. A general suggestion for a successful audit is to spend most of your time on major issues and a smaller portion of your time on small issues. There may be observations that you may wish to point out to supervisory personnel that deserve attention but do not belong in an audit report because they are relatively insignificant. By the same token, too many small items suggests a trend of noncompliance and deserve attention as such. When citing these, be specific.

## REFERENCES


1. *U.S. Code of Federal Regulations* (CFR), Title 21, Part 211, Current good manufacturing practice for finished pharmaceuticals, available: http://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfcfr/CRFSearch.cfm?CFRPart=211, accessed Dec. 5, 2006.
2. American National Standards Institute (ANSI) (2000), Quality management system—Requirements, ANSI/ISO/ASQ Q9001-2000, ANSI, New York.
3. American National Standards Institute (ANSI) (2000), Quality management system—Fundamentals and vocabulary, ANSI/ISO/ASQ Q9000-2000, ANSI, New York.
4. International Organization for Standardization (ISO), Application of risk management of medical devices, ISO 14971:2000, ISO, Geneva.
5. U.S. Department of Health and Human services, U.S. Food and Drug Administration, Pharmaceutical cGMPs for the 21st century—A risk -based approach, Final Report—Fall 2004, September 2004, available: http://www.fda.gov/cder/gmp/gmp2004/GMP_finalreport2004.htm, accessed Dec. 5, 2006.
6. U.S. Department of Health and Human Services (DHHS), Food and Drug Administration (2006, June), Guidance for industry: Q9 Quality risk management, DHHS, Rockville, MD.
7. U.S. Department of Health and Human Services (DHHS), Food and Drug Administration (2004, Sept.), Guidance for industry: PAT—A framework for innovative pharmaceutical development, manufacturing, and quality assurance, DHHS, Rockville, MD.
8. U.S. Department of Health and Human Services (DHHS), Food and Drug Administration (2006, Sept.), Guidance for industry: Quality systems approach to pharmaceutical CGMP regulations, DHHS, Rockville, MD.
9. *U.S. Code of Federal Regulations* (CFR), Title 21, Part 820, Quality system regulation for medical devices, available: http://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfcfr/CFRSearch.cfm?CFRPart=820, accessed Dec. 5, 2006.
10. International Organization for Standardization (ISO), (2003), Medical devices—Quality management systems—Requirements for regulatory purpose, ISO 13485:2003, ISO, Geneva.
11. Juran J. M., and Godfrey, A. B., Eds. (1999), *Juran's Quality Handbook*, 5th ed., McGraw-Hill, New York.
12. FDA compliance policy guide section 130.000, FDA access to results of quality assurance program audits and inspections (CPG 7151.02), available: http://www.fda.gov/ora/compliance_ref/cpg/cpggenl/cpg130-300.html, accessed Dec. 5, 2006.

13. International Organization for Standardization (ISO), (2002), Guidelines for quality and/or environmental management systems auditing, ISO 19011:2002, ISO, Geneva.
14. The Global Harmonization Task Force, SG4, Training requirements for auditors (guidelines for regulatory auditing of quality systems of medical device manufacturers—Part 1: General requirements—Supplement 3), available: http://www.ghtf.org/sg4/inventorysg4/trainingfinal.pdf, accessed Dec. 5, 2006.
15. FDA compliance program guidance manual for FDA staff: Drug manufacturing inspections program (7356.002), 2/1/2002, available: http://www.fda.gov/cder/dmpq/compliance_guide.htm, accessed Dec. 5, 2006.
16. U.S. Department of Health and Human Services (DHHS), Food and Drug Administration (2004, Sept.), Guidance for industry: Sterile drug products produced by aseptic processing—Current good manufacturing practice, DHHS, Rockville, MD.
17. U.S. Department of Health and Human Services (DHHS), Food and Drug Administration (2001, Aug.), Guidance for industry: Q7A good manufacturing practice guidance for active pharmaceutical Ingredients, DHHS, Rockville, MD.

# CREATING AND MANAGING A QUALITY MANAGEMENT SYSTEM

Edward R. Arling, Michelle E. Dowling, and Paul A. Frankel
*Amgen, Inc., Thousand Oaks, California*

**Contents**

1 Introduction
2 Understanding a Quality Management System
  2.1 Defining Quality Management Systems
  2.2 Synthesis versus Analysis
  2.3 System versus Process
  2.4 Business Benefits of Establishing a Robust Quality Management System
  2.5 Industry and Regulatory Expectations
3 Management and Staff: Leadership and Support
  3.1 Outlining Benefits to the Enterprise
  3.2 Speaking Management Language
  3.3 Translating Benefits to Staff
  3.4 Ensuring Staff Support and Management Leadership
  3.5 Traps to Avoid
4 Establishing Quality Management System Scope
  4.1 Defining Business Requirements
  4.2 Integrating Quality Management System into Quality Plans
  4.3 Determining Process Resolution Requirements
  4.4 Scalability to Enterprise
5 System and Process Ownership: Roles and Responsibilities
  5.1 Quality Management System Ownership and Management
  5.2 Process Ownership
  5.3 Process Owner Selection
  5.4 Stakeholder/Process Owner Integration
  5.5 Decision Authority
  5.6 Industry Knowledge
  5.7 Regulatory Inspection and Audit Lead
  5.8 Subject Matter Expert
  5.9 Metric Ownership

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

5.10 Documentation Ownership
5.11 Training
5.12 Risk Management
5.13 Continuous Improvement and Project Management
5.14 Nonconformance / CAPA / Planned Deviation Ownership
6 Change Management/Communication
6.1 Managing Organizational Change
6.2 Communication
6.3 Feedback and Alignment
6.4 Training
7 Measuring Success through Meaningful Metrics
7.1 Performance Metric Development
7.2 Metric Review
7.3 Maturity Model
7.4 Meeting Process Maturity Requirements
8 Driving Continuous Improvement: Projects
8.1 Process Improvements
8.2 Process Improvement Proposal
8.3 Task versus Project
8.4 Project Metrics
9 Ensuring Ongoing Success
9.1 Establishing Mutual Goals
9.2 Rewards and Recognition
9.3 Ensuring Ongoing Program Continuity
9.4 Program Institutionalization
References

## 1 INTRODUCTION

The world's population continues to grow and the average life expectancy continues to increase. Pharmaceutical and biopharmaceutical products are more in demand as the population expands, requiring novel and specialized medications to treat common and debilitating diseases. The industry is challenged to rapidly discover and commercialize products to treat existing unmet medical needs and emerging threats as viruses mutate into new diseases that threaten the stability of the world as we know it.

At the same time, the global marketplace continues to increase its demand on the industry. Government, consumer, and wholesale buying pressures demand lower prices. Higher quality standards are expected by regulators and consumers. Competition continues to increase from generic, biosimilar, and counterfeit producers. Developing nations, with lower cost overheads, are developing economical production capabilities. Meanwhile, research and development costs are increasing.

This chapter will outline the concepts, benefits, and practical implementation steps for developing a comprehensive quality management system (QMS) that supports pharmaceutical and biopharmaceutical manufacturing operations. The material presented is universal in its utility, applicable to small and large companies, development, and commercial enterprises. A QMS is a proactive, structured approach

to supporting development and manufacturing operations. It includes all processes, metrics, management review, and continuous improvement activities. The QMS, as described in this chapter, is further supported through an active change management program and application of annual quality plans to ensure ongoing system sustainability.

A well-designed QMS, with mature, developed processes, provides the required infrastructure and support necessary for successful manufacturing operations. Integrated processes, proactively managed, that can be quickly modified to meet changing business and regulatory demands will support ongoing manufacturing operations and provide competitive advantage. This chapter provides guidance on creating and managing a robust QMS that supports manufacturing operations in the pharmaceutical and biopharmaceutical industry.

## 2 UNDERSTANDING A QUALITY MANAGEMENT SYSTEM

Every development, testing, manufacturing, packaging, warehouse, or distribution facility has its own unique role in producing an output or product for consumption by a customer somewhere in the pharmaceutical or biopharmaceutical supply chain. Each facility and organization is critically dependent upon several different processes that function interdependently producing the desired output. Organizations' survival and profitability are directly linked to the efficiency of design, execution, performance, and interrelational attributes of these processes. Throughout a product life cycle, from early discovery through development, scale up, clinical testing, product technology transfer, registration, approval, commercialization, and eventually product discontinuance, robust processes are the foundation supporting the successful enterprise.

Manufacturing support processes are discrete in their output, but interrelated in their overall effect. Weak or ill-defined processes have a diminishing overall effect on the organization and its product. It manifests itself as increased rework, rejected material, extended cycle times, delayed disposition, high nonconforming performance metrics, complaints, recalls, or other inabilities to meet customer or market demands. A comprehensive QMS may encompass all the processes supporting development and manufacturing. It includes the standards, policies, and procedures required to measure those processes for performance and maturity. It provides metrics necessary for leadership to perform risk-based prioritization and focus resources for business improvement and regulatory compliance.

Robust processes will have owners that have defined roles, responsibilities, and accountabilities. These process owners must be fully dedicated to their process. They must know their process capabilities and expectations, the interrelationship between their process and other processes and manage them like a business unto themselves. Functional management must support process owners, and leadership must understand and lead the QMS effort as an ongoing program, treating it as the integral part of the business that it is.

A QMS is an organizational approach consisting of people, interrelated processes, process inputs and outputs, and structured review programs that lead to ongoing continuous improvement. This complexity of processes requires a programmatic organization and management to effectively interrelate its components. A

QMS program office is required to provide the organizational benefits expected from well-managed processes and should be one of the first elements established when instituting the program.

A QMS and the processes comprising it are not the sole responsibility of the quality function or a single functional group. Inherently, these processes have no bounds in the organization. The concept must be owned, managed, or executed by all staff from leadership to the most entry-level manufacturing associate. A quality mindset must be part of every employee that contributes to the discovery, manufacture, packaging, testing, warehousing, and shipping of a product or output. A culture of quality and understanding of the processes in which personnel work are essential to advance the QMS to maximize benefits to the enterprise and remain competitive.

Instituting a QMS through a holistic approach that supports manufacturing operations has the potential to meet and exceed customer, patient, shareholder, and employee expectations. It requires a cross-functional team approach, with proactive management of all the processes responsible for manufacture, including functional support from development, manufacturing, analytical, engineering, and quality assurance. The development and maintenance of a tested, robust QMS requires time and resources. Full maturation of processes and organizational culture change may take, in some cases, years to fully implement and realize benefits, but worth the effort and time. Significant QMS issues should not be addressed as one-off fixes. Rather, action taken to remediate deficient processes should be approached as long-term corrections, addressing the root cause of the failed process, so they do not repeatedly plague the organization.

The ultimate responsibility for a robust, functional QMS lies with top management. The organization follows the leadership, and therefore, leadership must support a QMS that is specifically designed for the organization, be aware of and monitor its progress and contribution to the organization, and frequently support, guide, and maintain it. Doing so ensures viability of the QMS, and in turn the QMS will provide leadership the data and guidance necessary to effectively manage the organization.

### 2.1 Defining Quality Management Systems

The term *system* or *quality system* is used with surprising inconsistency throughout the pharmaceutical and biopharmaceutical industry and by government regulators. Even within a single company or within a department, the terms can be nebulous in their use and interpretation. System is often used to describe an individual process or unit operation. Often, the term system is used so narrowly as to describe an individual policy, standard, or even a single procedure.

Recent initiatives by global organizations such as ISO (International Organization for Standardization, www.iso.org) and ICH (International Conference on Harmonization, www.ich.org) are attempting to bring consistency in concept and standardization in definition to the QMS. In 2004, the Pharmaceutical Inspection Co-Operation Scheme (PIC/S, www.picscheme.org) issued its recommendation on Quality System Requirements for Pharmaceutical Inspectorates. The U.S. Food and Drug Administration (FDA) initiated inspection surveillance approaches based upon QMS organization and is another source of definition and interpretation.

Inconsistency in language and expectations continues to exist; however, efforts are progressing to minimize distinctions and globally harmonize efforts, structure, and language concerning quality systems.

According to Webster's dictionary, *system* is defined as a regularly interacting or interdependent group of items forming a unified whole; a group of interacting bodies under the influence of related forces . . . an assemblage of substances that is in or tends to equilibrium . . . a group of organs that, when together, perform one or more vital functions . . . an organization forming a network especially for distributing something or serving a common purpose . . . an organized set of doctrines, ideas, or principles usually intended to explain the arrangement or working of a systematic whole [1].

The vocabulary and definitions used in this chapter defines a quality management system as the compilation of all the processes required to support the manufacture, packaging, testing, release, and distribution of an active pharmaceutical ingredient (API) or drug product. It is aligned with that of the FDA Center for Drug Evaluation and Research (CDER) compliance program 7356.002, issued to investigators for the inspection of pharmaceutical and biopharmaceutical manufacturing plants (www.fda.gov/IOM 7356.002). The CDER inspection program subdivides the processes comprising the QMS into six subsystems: quality, facilities/equipment, production, materials control, laboratory controls, and packaging and labeling.

There are no specific CDER requirements as to which processes belong under each subsystem; however, one can easily follow the outline provided in 21 CFR Part 211, the regulations applicable to human drug product manufacture, to aid in the determination of processes likely to be inspected during a regulatory inspection (www.fda.gov). The FDA subdivides all the processes comprising a company's QMS into six subsystems to ensure adequate and varied coverage during inspections. See Figure 1. Using the same process organization structure and vocabulary as regulators provides an enterprise the advantage of more efficient inspection preparation and avoidance of miscommunication during and after regulatory inspections.

The CDER subsystem organization provides regulators and management the ability to focus attention to specific functional areas. Table 1 is an example of the processes, organized under appropriate subsystems, supporting a typical API or drug fill-and-finish operation. These subsystems are organized according requirements found in regulations used by investigators during inspections, 21 CFR Part 210, 211, and the unit operations and support processes necessary for production.

One size does not fit all situations. Each enterprise has the responsibility and latitude to design a QMS to meet its specific needs. Even facilities with very similar manufacturing operations may require different processes to support the business. Each manufacturing organization requires a customized set of processes which will comprise its QMS. The management group responsible for the QMS should be able to identify and justify the processes comprising the system. There is not a single set of processes that can be universally applied to all operations, as each organization is unique in its business, product output, organization, culture, as well as local and global regulatory and customer requirements.

Processes identified as part of the QMS can be organized into the appropriate CDER subsystems for the purpose of aligning with the methodology used during inspections. It also provides management the ability to determine areas of strength or opportunities for improvement within the QMS. Regulators will always include

**FIGURE 1** Subsystems and management relationship.

**TABLE 1 Quality Management System Subsystems and Processes**

| | |
|---|---|
| Quality | Facilities/equipment |
| Audits and inspections | Facility and equipment design |
| Management review | Equipment maintenance |
| Risk management | Equipment cleaning |
| Organization and personnel | Calibration |
| Training | Materials control |
| Document management | Supplier quality management |
| Change control | Sampling and inspection |
| Nonconformances | Receiving, warehouse, and storage |
| Corrective and preventative actions | Inventory management |
| Biological product deviation | Transport |
| Product disposition | Return and salvage |
| Validation | Laboratory controls |
| Production | Laboratory testing |
| Manufacturing | Sample management and sample plans |
| Process monitoring | Stability program |
| Environmental and gowning monitoring | Packaging and labeling |
| In-process controls | Labeling controls and approvals |
| Gowning | Package development |

a focus on the processes within the quality subsystem. Other subsystems will be reviewed during inspections based upon the type of inspection and compliance history of the enterprise. More information on how the FDA focuses inspections based on quality system and subsystem organization is available at the FDA website (www.fda.gov) or articles written on this subject [2].

To maximize the effect of a QMS, it should be designed to be scalable and transferable throughout the enterprise and easy to understand and execute. An adequately designed QMS results in increased efficiency, a compliant operation, and staff satisfaction.

### 2.2 Synthesis versus Analysis

With systems thinking, the whole is greater than the sum of its parts. Systems rely upon the interaction of several processes. An individual process has limited value on its own, regardless of the level of development it has achieved. Processes provide value to the system through synthesis with other processes.

In the early twentieth century, researchers began to recognize the existence of interdependent relationships and organizational patterns among seemingly discrete parts. It is the relationships that allow parts to function as a whole. The "perceived whole" is a system. Systems thinking involves considering the parts in the context of that whole. In systems thinking:

- Everything in a system is related to everything else in the system.
- The parts of a system work together to achieve the overall objective of the whole system.
- In addition to the immediate effects of an action, there will be other consequences that ripple through the system.
- Every change brings benefits and consequences.
- Changing or reinforcing patterns and relationships within a system is as necessary to achieving the goals of the system as changing or retaining the parts of the system.
- Systems are "living" entities that sustain themselves through self-regulating dynamic equilibrium and organize to respond to externally imposed change.

Viewing a QMS in this context is beneficial to organizational leadership and management responsible for the system. It puts into perspective the overall effect on an organization that is achievable by individual processes alone and what can be achieved and sustained through active management and the interaction between those processes.

### 2.3 System versus Process

Traditional industry paradigm has the Quality Department responsible for quality and the Manufacturing Department responsible for producing product. Inherent conflict exists in this model due to competing functional priorities. By building quality concepts and accountabilities into production processes responsible for production, quality becomes infused into the organization. Both Quality and

Manufacturing therefore share the common goal of supplying high-quality product through the efficient execution of their processes.

Historically, very few processes were regarded as "quality systems," and they were viewed as something owned by the Quality Department. These "systems" were in fact ill defined and nonrelated processes used to monitor or detect individual actions and activities occurring in the manufacturing environment. These systems were based on quality control (QC) type of responsibilities for testing quality into the product. Examples include raw material testing, in-process and finished-product testing, nonconforming material review, environmental monitoring, and release and distribution. Few were interrelated with other processes, actively supported by management, or reviewed by leadership for performance or compliance.

The QC monitoring processes described above, if supported, were limited in their ability to support improvements and could only lead to action that was reactive in nature. Process integration is weak or nonexistent. Neither process maturity and development nor proactive system management is achievable. In the past, QMS enhancement was viewed as an expense and not seen as a relational contributor to the value chain. Aware management now realizes, through regulatory action, penalty and fines, delayed product approvals, recalls, and the like that establishment of a comprehensive QMS is essential to survive in the current regulatory environment and remain competitive in the business environment.

With the advancement of quality assurance (QA) principles and concepts at the end of the last century, QMSs have evolved to be more proactive to include change control, supplier and internal auditing, risk management, lagging and leading metric collection, and review. Review of predictive metrics has become the basis for preventive action and continuous improvement programs. Today's competitive environment obligates leading manufacturers and world-class organizations to apply proactive system thinking to expand their focus to include all processes that support product quality, irregardless of the stage of development or manufacture. Early implementation of appropriate processes supports quality-by-design concepts and practice, within the framework of a QMS and ensures quality in all processes and provides the foundation for good investigations and continuous improvement.

A QMS should be comprised of all the processes supporting that business and include an effective management review of those process metrics. Management needs to be aware of and understand process performance through structured metrics review programs in order to take appropriate action, providing resources and capital to improve the QMS. This hierarchy is illustrated in Figure 2.

Processes supporting and applicable to pharmaceutical and biopharmaceutical manufacturing are easily determined by examining the business needs of the organization and the regulations governing them. A carefully designed QMS will consider the needs of the enterprise as a whole, as well as that of the individual unit operations comprising the enterprise. If the QMS design is comprehensive, it will provide significant value to global and local management. It will support staff by standardizing processes, requirements, and expectations and provide leadership meaningful and comparable metrics on system and process performance. Changes can be quickly facilitated and implemented when process modifications are required. A consistent representation of processes to regulators builds confidence and trust that the enterprise is capable to produce the product for which approval has been granted.

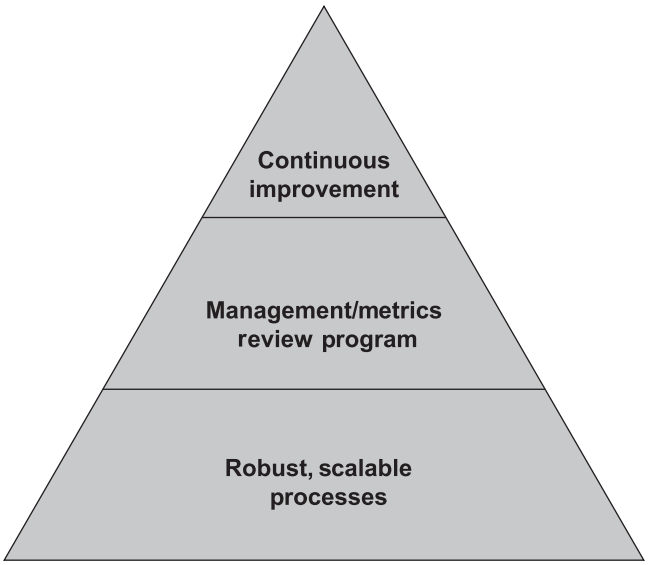


**FIGURE 2** Quality management system hierarchy.

### 2.4 Business Benefits of Establishing a Robust Quality Management System

The competitive nature of the pharmaceutical business demands capable and efficient processes supporting discovery, development, technology transfer and scale-up, and commercial manufacturing and distribution. Execution of efficient processes is the foundation for new and ongoing enterprises to be successful. It is the basis for successful manufacture and the bedrock upon which management and regulators can gauge the capability level of the enterprise. Providing patients with needed medicines in a timely, cost-efficient manner, without delay due to manufacturing or compliance issues, should be a primary driving force behind the pharmaceutical and biopharmaceutical industry.

Leadership may ask the question: Why implement a quality management system? The answer is that a well-designed system is necessary to establish a state of control to ensure that a high quality, safe, and efficacious product is produced and available for patients. Quality systems as described in the forthcoming ICH Q10 guidance is the logical complement to its predecessors, ICH Q8 (Product Development) and ICH Q9 (Risk Management) (www.ich.org). These three guidance documents build upon each other from quality-by-design activities in development through the entire product life cycle. When used together, the guidance documents maximize their benefits to the enterprise through better process understanding, less regulatory scrutiny, and increased freedom to operate. Together, these guidance's support more efficient product life-cycle management from discovery through development and commercialization.

Inefficient operations cost businesses untold amounts in financial and human capital. A poorly designed system coupled with inefficient processes may result in rework of development and commercialization activities, data integrity issues, inefficient use of resources, and delay in approval. Poorly designed processes may also

lead to loss of future revenue with business partners and have a negative regulatory consequence.

A recent study conducted by the Pharmaceutical Manufacturing Research Project, a joint venture by Georgetown University and Washington University in St. Louis business schools, collected data from 42 manufacturing facilities owned by 19 companies to determine factors that affected industry performance. The Final Benchmarking Report assessed performance in terms of manufacturing times, frequency of deviations from manufacturing standards, reasons for deviations, manufacturing yield, and rates of improvement for those metrics.

The study determined that improvements in manufacturing process could save industry more than $50 billion in manufacturing costs, which the researchers believe could result in lower drug prices and more money for R&D. The report received no industry or government funding [3].

Leadership is both challenged and rewarded for supporting the development of a robust QMS. On one hand, it takes time and resources to design and develop a comprehensive program. Immediate return on this investment is not usually forthcoming. Management is typically under pressure to deliver aggressive results in a short time period, which is counterintuitive to careful planning and long-range development. Conversely, proactively formalizing and supporting a robust QMS will, in the long run, ensure the operations freedom to operate (regulatory compliance) and deliver business efficiencies.

In the pharmaceutical and biopharmaceutical manufacturing industry, the perception of quality has dramatically changed over the past several years, and loss of market capitalization can be a direct correlation to this perception. Large pharmaceutical companies have gone from some of the world's most admired companies to losing significant percentage of their value, based on consumer, media, and investor perceptions of quality and ethics. Speaking at a recent Parenteral Drug Association (PDA)/FDA joint regulatory conference, Daniel Diermeier, IBM Distinguished Professor of Regulation and Competitive Practice, Northwestern University stated: "The perception of quality on the pharmaceutical value chain is greater than in other industries (auto, furniture, etc.). Patients cannot assess the quality of drugs as they can a car or hotel room. In healthcare, the 'value proposition' is higher than other industries and the Quality [Management] System is a critical subset of that perception" [4]. Dr. Diermeier goes on to suggest a QMS include processes for decision and detection to further protect the "value proposition" of the enterprise.

Enterprises lacking individual capable processes experience degrees of negative effects throughout the organization. This is true for processes that support discovery, development, manufacturing, or marketing. Recent examples of fines imposed by regulators for poor processes supporting the QMS are increasing (see Table 2). These costs are only indicative of the fine itself and do not include lost revenue, cost of consultancy for remediation, decreased shareholder value, and diminished staff morale and support. These costs are typically an order of magnitude or more greater than the fine itself.

A common misconception of pharmaceutical and especially smaller biopharmaceutical companies is that the implementation of a robust QMS is not required in areas other than commercial manufacturing. Small, biotech start-up companies also tend to delay the implementation of well-designed processes until they near the approval stage, focusing the organization instead for product approval or sale. This

**TABLE 2 Potential Financial Impacts**

| Company | Compliance Issue | Type of Impact | Cost to Business ($Mil) |
|---|---|---|---|
| A | Failure to follow procedure Inadequate training | Multiple 483 observations | <1 |
| B | Inadequate process definition, controls, and oversight | Warning letter | >1 |
| C | Repeat observations—direct product impact Failure to meet warning letter commitments | Consent decree | >100 |
| D | Plant shutdown | Direct fines product stock-out | >500 |

can become a costly miscalculation, as speed to market and limited capital demand processes supporting efficient development, clinical and regulatory submission processes be executed with minimal waste or rework.

Although QMSs are routinely identified with commercial manufacturing, it is critical to establish process parameters for discovery, development, and technology transfer, including scale-up, characterization of process, analytical methodology, and validation. Development activities are executed more efficiently through the application of robust processes and ultimately become the foundation for robust manufacturing. Failed development studies, inadequate comparability reports, clinical studies requiring repeated, or poorly supported analytical and process characterization contribute to delayed submissions and weak regulatory submission and inspection presentation. The identification of processes supporting these activities, owner identification and accountability and support will ensure success of the enterprise and reduce the anxiety and uncertainty that is inherent in development and approval activities.

Several opportunities exist for pharmaceutical and biopharmaceutical manufacturing plants to improve efficiency and cost savings, which ultimately validate the program's benefits and supports leadership in achieving their financial goals. Traditionally, the industry environment is heavily regulated and has been very risk adverse. These two elements combine to offer countless opportunities to improve inefficient and ill-defined processes, clarify process scope, define process owner accountabilities and responsibilities, and remediate process duplication or gaps. Performing inefficient processes for the sake of avoiding regulatory scrutiny or attempting to defend poorly characterized processes without adequate data and interpretation becomes self-defeating to the industry. Poor prioritization of work, ill-defined process relationships, and functional management interference or neglect may also contribute to inefficiency. Staff requires processes that are easy to execute, well integrated, and result in value-added activities. This can only be accomplished through the design and execution of efficient processes that are interrelated, bringing value to the enterprise, process owners, and stakeholders.

An example of a robust process is the design, development, and operation of a nonconformance process. Regulations require an operational process to identify, document, and correct nonconformances occurring in licensed pharmaceutical manufacturing facilities for approved products. Companies spend significant human

capital identifying, documenting, and tracking nonconformances. But how much is actually being done to remediate these nonconformances? Can the nonconformances be related to previously completed development or commercialization studies? Is the nonconformance process sufficiently related to an effective corrective or preventive action (CAPA) process? Does the preventive action interrelate efficiently with an efficient change control process to ensure proposed changes remain in compliance with registrations? Are the documentation and training processes sufficient to support approved changes? An adequately designed QMS will ensure the supporting processes are present and that functional and interrelationships established. A systems implementation provides a holistic approach, which results in both building effective individual processes and interrelating those processes to maximize their effect on the business, driving efficient and science-based activities.

Maintaining good manufacturing practice (GMP) compliance is essential for pharmaceutical and biopharmaceutical companies. Results of noncompliance are costly fines, loss of revenue, higher overhead costs, delayed approvals, and poor customer and regulatory perceptions. Poor compliance results from an inadequately designed QMS that lacked the processes and management review required to support the enterprise. Processes supporting compliance include self-audits, change control, document revision and approval, and staff training programs. Regular management review of these processes will ensure resources are allocated to appropriate initiatives and there should be no surprises during inspections. A well-designed QMS should prevent negative regulatory consequences. Efficient and compliant processes support lean manufacturing efforts through the documentation and understanding of processes. Management review of these processes ensures that leadership awareness, support, and action is taken by the organization when appropriate.

Figures 3 and 4 illustrate how a biennial document review process and document processing cycle time metrics faltered in their early stages due to lack of process ownership, definition, and management review. This situation presented a compliance risk to the organization and resulted in poor business efficiencies. Improve-

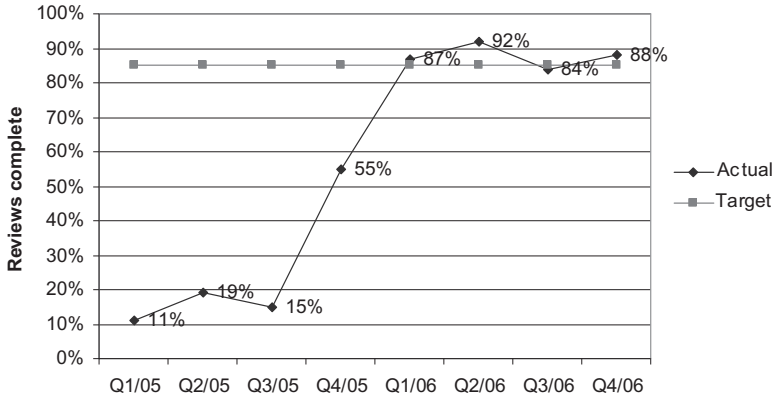


**FIGURE 3** Biennial document review process.

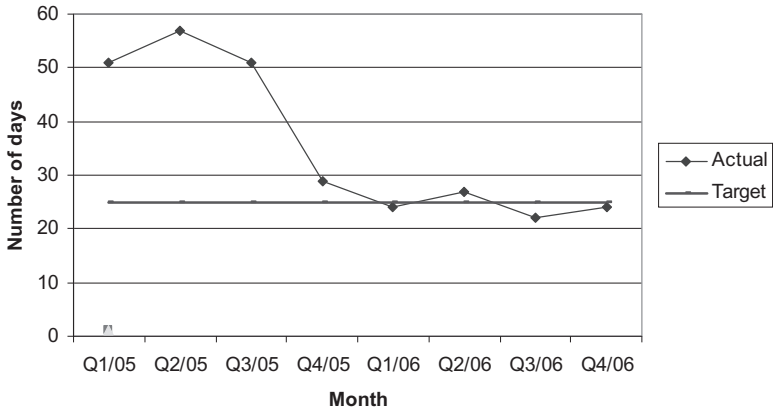


**FIGURE 4** Document review cycle time.

ment was attained by assigning a process owner who defined and improved each process, developed meaningful metrics, and presented those metrics to management. Management became aware of process performance, understood the compliance risk and business impact and took appropriate action to focus staff efforts to meet process requirements. Results were improved document review cycles, proactive compliance with internal procedures and regulatory requirements, and the satisfaction of knowing that no additional effort was required to achieve better business results and regulatory compliance.

### 2.5 Industry and Regulatory Expectations

While there are no requirements for a "quality system" in current FDA regulations applicable to pharmaceutical and biopharmaceutical manufacturing, regulatory agencies and industry trade organizations are increasingly recognizing the importance of robust, functioning quality systems in support of manufacturing the world's medicinal products. The FDA realizes not all quality principles are represented in current GMP regulations for drug products (21 CFR Part 211), which were last updated in 1978.

Quality management system issues and their association with risk management are common topics discussed in trade and regulatory seminars and conferences. Recent guidelines such as FDA "Quality Systems Approach to Current Good Manufacturing Practice Regulations" found on the FDA website and part of FDA's initiative titled "GMP's for the 21st Century" was written to complement existing regulations. While the FDA guidance may change or even become redundant with the issuance of ICH Q10, there is common intent among industry and government to advance quality management systems. According to Joe Famulare, Director DMPQ, FDA, the "FDA wanted to write a comprehensive Quality System model that would support and correlate with CGMP regulations. The guidance is consistent with defining a state of control; facilitate quality efforts, change control, Quality by Design, and risk management" [4].

In discussing quality systems at a recent industry conference on GMPs, Chris Joneckis of the FDA CBER (Center for Biological Evaluation and Research) had

this to say: "A robust Quality Management System makes a strong case for quality product. It is a win, win, win—for patient, industry and regulators. It benefits technology transfer, process control, monitoring, capability, improves manufacturing, fewer nonconformances and better quality of investigations. Regulatory benefits include enhanced Chemistry, Manufacturing, Controls (CMC) review, change control, and submission of postapproval changes" [5].

Regulatory and industry guidance documents have been generated in support of developing and organizing quality systems. In the late 1990s, the system-based inspection approach was formalized by the Center for Devices & Radiological Health (CDRH) of the FDA [6]. These regulations were codified as QSR, Quality Systems Regulations, and are included in Part 820 of the *Code of Federal Regulations* (CFR).

The CDER and CBER soon followed the CDRH approach and issued their own Compliance Program Guidance Manuals, 7356.002 [7] and 7345.848 [8], respectively, which were modeled on the CDRH QSR approach. The CDER and CBER are responsible for ensuring the biennial inspection of pharmaceutical and biopharmaceutical manufacturing facilities. The guidelines listed here are used by investigators during manufacturing inspections. Process owners and stakeholders as well as management and leadership should be familiar with these compliance manuals and how investigators plan to use them during inspections.

Current FDA inspectional surveillance, based on the models described above, requires investigators evaluate the processes within the subsystems defined by the QMS to determine compliance and risk to patient safety. This is different than the traditional approach of reviewing individual products during inspections. There is subtle, yet significant advantage to both the regulating agencies and compliant companies by using a system approach, as the inspections are designed to be faster and cover many product types during one inspection. Companies with compliant histories can benefit with nominal inspections, whereas companies with noncompliant histories will receive more regulatory scrutiny and possible regulatory action.

The movement by industry groups such as the ISO, which attempts to provide recognized standards for many industries, was also grounded in a systems approach with the publication and certification of ISO 9000 series and later with ISO 2000:9004 (www.iso.org), which is based on QMS establishment and eventually continuous improvements once processes become stable.

The ICH, a joint regulatory–industry initivative on international harmonization for drug development and approval, also recognizes the value and contribution of a quality systems approach through its guidance development on this topic (ICH Q10). The pharmaceutical and biopharmaceutical industry and regulatory agencies are collaborating to finalize the guidance sometime in 2008. ICH Q10 is focused on pharmaceuticals and is intended to align GMP requirements with a quality system approach. It will be applicable to drug substance and drug product, large and small molecule products, and harmonize one approach to quality systems. It also will complement ICH Q8 and ICH Q9. ICH Q10 contains a pharmaceutical context emphasizing a comprehensive approach; key elements included are management response and continuous improvement. Several ICH guidance documents are already adopted by regulatory agencies, such as ICH Q7A, for the manufacture of APIs. As these guidance documents are adopted, they often become the basis for regulatory expectations and inspections.

## 3 MANAGEMENT AND STAFF: LEADERSHIP AND SUPPORT

All manufacturing operations operate, to some extent, with elements and components of a quality management system. Those elements and processes may not be recognized or managed as though they are an integral part of a larger system and may be primarily reactive in nature. Significant time and resources are required to change an organization's culture and practices to move existing elements from a fragmented, reactive program to a defined structure that is proactively managed. The degree to which a program is proactively managed and supported by its leadership is directly related to the benefits experienced by the organization.

Three distinct levels of support are required for successful implementation of a QMS program: executive leadership, functional management, and operational staff. All three levels of the organization must support the effort to attain success. Delivering program understanding and benefits to each should be a priority to ensure acceptance and continuity. Motivating staff and leadership, through benefits and business results, is important to ongoing program sustainability.

Leadership requires capable and dedicated staff to design and maintain a dynamic QMS program. Leadership must embrace the program and support it throughout the organization. Functional management must understand the program in order to support it and direct its staff in execution of the program. Staff must understand what the program means to them and experience and realize the benefits in order to support it.

The quality organization must be seen as a partner in assuring product quality, not the department that disseminates quality. Within a QMS, certain processes are owned by the quality function, just as manufacturing, engineering, development, technical support, and facilities own processes within the system. All functional groups should have defined roles and responsibilities to ensure quality product is produced. Cross-functional support and delineation of responsibilities ensure quality is built into every process, and each process owner is ultimately responsible for his or her process output. Leadership that understands and embraces this concept will support and infuse a culture of quality throughout the organization, maximizing the probability of success and competitive advantage.

The organizations leadership, management, staff, and QMS program group must work together to develop and progress the QMS. A successful program should detail expected benefits for all stakeholders in the organization and provide ongoing results demonstrating functionality and utility.

### 3.1 Outlining Benefits to the Enterprise

Establishing a formal, structured QMS for an organization requires leadership approval, resources, and capital. Leadership support and approval is the place to initiate the program to ensure all program efforts are supported and the proposed system meets the business needs. This includes having dedicated resources that can focus their efforts to design and manage the program and operate and manage the processes.

Leadership has visibility to present business needs and budget and the vision and insight for the organizations' future. Quality management system design needs to fulfill present and future needs to be robust and value added. A gap analysis on

current business processes can help leadership understand where opportunities exist for improving processes. These gaps can be determined by analyzing the purpose of the organization and its ability to deliver quality results on time and on budget.

Manufacturing areas to examine for operational improvement are regulatory compliance, audit findings, rework, nonconformances, document revisions, disposition timeliness, complaints received, inventory on hand, equipment failures, manufacturing cycle times, employee turnover, and training opportunities. Additional areas targeted for improvement may come from benchmarking key manufacturing parameters against industry peers. Results of a gap analysis begin the dialogue regarding process performance and the need for process improvements. Leadership must be convinced there is opportunity for financial and competitive gain, and the resource investment to operate the QMS will be outweighed by the program benefits received.

Management at the highest level in the organization must understand, support, and lead the strategy to implement systems across the enterprise. More often than not, this requires some level of business transformation, a cultural and behavioral shift, and a certain level of risk. The risk associated in implementing change is minimal compared to that of not having a robust system, as outlined in the benefits section.

### 3.2 Speaking Management Language

Without upper management championing the establishment of systems, midlevel management will not support the effort, dedicate the time required, nor practice the behaviors essential to establish and maintain the processes. Leadership needs to be cognizant of the benefits and consequences of nonimplementation and be clear and unwavering in its support, delivering frequent consistent messaging to management and staff. Leadership requires tangible and intangible benefits to be convinced that the efforts are worthwhile and working and to regularly convey results to staff.

Tangible benefits should include metrics and improvements demonstrating process and system cost savings, compliant inspections and customer audits, faster product approvals and manufacturing throughput, less rejected material, reduced nonconformance issues, and more efficient continuous improvement and project implementation. Intangible benefits include improved staff morale, faster, more accurate transparent decision making, less employee turnover, increased staff accountability, and an enhanced culture of quality throughout the organization. The "feeling" conveyed by an organization that is reactive, stressed, and without well-structured processes is much different than that of a proactive organization with simple processes that are easily and successfully executed by trained staff.

Systems thinking allows decision making and process management to occur at the process owner level, not the functional management level. This is a cultural shift for many organizations but brings with it many benefits. Faster decision making, by subject matter experts is valuable to organizations. It can benefit both on a day-to-day, lot-to-lot basis as well as provide long-term strategic direction to leadership. Taking the burden off functional management and defining process owner responsibilities allows functional management to manage resource and personnel issues and not split time and attention between resources, personnel, technical, and process issues.

### 3.3 Translating Benefits to Staff

Similar to leadership and management requirements regarding system understanding and benefits, staff requires understanding prior to accepting the cultural changes that a system-based approach will bring to the organization. Once the program is initiated, tangible and intangible benefits must be realized and appreciated in order for staff to continually support the program. Staff support, through benefit realization and management direction, will ensure program execution, ultimately delivering the expected business results.

Transforming disparate processes into processes that are simple to understand, easy to execute, and provide a sense of accomplishment meet one of management's obligations to staff. Staff interest lies in the ability to perform their work, contribute to continuous improvement, and have a reasonable work–life balance. Finally, they want to be able to contribute to their careers, have defined career paths, and have attainable development goals for advancement. A well-designed quality management system can contribute to provide all these employee benefits.

Staff benefits should be designed into the QMS. An outline of expected benefits should be presented to staff to gain their support of the system initiative. Accomplishments should be advertised and rewarded. Establishing well-defined processes empowers employee involvement, participation, and contribution to the organization. It reinforces a culture of quality throughout the organization, and provides a conduit for their contribution.

### 3.4 Ensuring Staff Support and Management Leadership

Management's responsibility includes providing staff robust tools and processes necessary to accomplish their jobs efficiently. Complex, missing, or fragmented processes do not allow for easy operational execution, the ability to leave work at reasonable times, and may result in poor-quality output or rework. This type of environment quickly becomes dissatisfying to employees and results in poor morale, low efficiency, and ultimately lack of interest and loss of staff.

Staff empowerment allows pride in workmanship. Well-designed quality systems make clear to staff where decision authority and process accountability lies, provide clear expectations of the process and process owners, and provide personnel a clear development path to process ownership.

Clearly identified process attributes provide organizations more than tribal knowledge to pass onto the next process owner. They provide clear structure, process, and other attributes critical to the ongoing success of the enterprise. The organization becomes reliant on their system and processes not people's personal knowledge, which can be lost with staff turnover.

Ensuring leadership and staff support requires that a well-defined plan be designed and shared throughout the organization. A long-range plan, spanning several years may benefit the organization to maintain perspective and govern expectations. An annual quality plan should encompass all aspects of the QMS and contain detailed periodic goals and objectives. Progress against the quality plan needs to be advertised and celebrated. Quality plan leadership should be recognized for its efforts and accomplishments. Advertising wins and accomplishments in both small group and large settings should be designed into the communication and change management program.

**TABLE 3 Long-Term Strategic Vision**

| Year 1 | Year 2 | Year 3 | Year 4 | Year 5 |
|---|---|---|---|---|
| Gain management support<br>Create QMS office<br>Identify site processes and resources<br>Develop communication and change management plan | Implement program<br>Train management, process owners, QA, and support staff<br>Focus on maturing high-risk/impact processes<br>Reward and recognize QMS efforts | Indoctrinate remaining processes into program<br>Document and communicate cost/resource savings<br>Begin integrating processes across the organization | Focus on key projects based on QMS portfolio and management review<br>Provide ongoing training, communications, and change management | Adapt to changing business and regulatory environment<br>Provide leadership to industry on QMS paradigm |

Table 3 provides an outline of a long-term vision and goals for a quality management system. A long-term strategy provides leadership, management, and staff with an understanding of the program and anticipated timelines for implementation and benefit expectations. Annual quality plans become the short-term strategic milestone vehicle to achieve the long-term strategic vision.

### 3.5 Traps to Avoid

Several challenges and requirements present themselves when establishing a formal QMS. A primary requirement is a skilled team that understands the needs of the organization, regulatory, and customer requirements. It should have the skill, experience, and expertise to design a robust system and identify processes that support the enterprise. A mismatch of team skills with enterprise needs may result in a nonviable system that is not supported by leadership and staff, leading to failure and disuse over time.

Quality management system design must be well thought out and tested. Pilot programs are crucial to test system robustness and reliability, staff and management acceptance, and the ability to produce the desired results. Time spent in system design will pay dividends for years to come and increase staff support and critical mass throughout the enterprise supporting the program efforts. Avoid implementing any system or process design that has not been well thought out, does not have input from the stakeholders using the system, or has not been piloted prior to a full-scale implementation. Typically, a single opportunity exists to introduce a new program before staff and management either accept it or reject the ideas and concepts. Rebuilding interest and trust of a failed system is difficult. The probability for successful reintroduction is minimized. Taking sufficient precaution for correct implementation the first time is important.

Change management is another very important consideration when implementing a QMS because of the culture change required from the organization. Several resources can assist in managing change, and these should be incorporated into the system design. It is important to be cognizant that successful implementation requires change at all three levels of the organization; leadership, functional management, and operational staff. Each will need different messages, encouragement, rewards, and benefits. Consideration to deliver both tangible and intangible benefits to stakeholders is necessary.

Leadership support from the highest level is required. Middle management will not support an effort that is not supported by its leadership. Leadership must provide unwavering support, not provide mixed messages, continue to advertise and celebrate success, and support the program through rough times. Consistency in language and deeds from management supports understanding and appropriate risk taking by management and staff.

Functional management must also support system efforts and long-term strategies, to ensure that staff, who are critical to execution of the processes, know that their support and efforts are expected. Functional management send powerful messages to staff, and their support of the long-term plan and annual quality plan are essential. Specific system objectives, included in leadership, management, and staff goals reinforce the commitment and help ensure success of the program.

The system needs to remain flexible. Having a long-term plan and vision is necessary to provide a roadmap to the future. That roadmap may need to change as the business environment and enterprise needs change. The long-term plan and vision should be written at the level that it changes very little, but flexibility is maintained through the preparation of an annual quality plan that is capable of addressing temporal issues and business needs.

Prior to implementing any QMS initiative, one must understand what leadership, management, staff, and customers require. Knowing which processes are required to support customer needs and the impact of those processes upon each other is essential to system design. Developing process owners that understand their roles and deliverables in the organization, eliminating constraints so they may meet their goals is essential for success. Process owners must understand product and process priorities so significant benefits may be realized. These are important considerations in designing system and processes that support the organization, produce meaningful metrics, and demonstrate progress. Consideration of these important points prevents system initiatives from failing and interpreted as another burden to the already overburdened work and demands placed upon the organization.

## 4 ESTABLISHING QUALITY MANAGEMENT SYSTEM SCOPE

In many pharmaceutical and biopharmaceutical manufacturing operations, duplicity exists in some processes and gaps are present between others. Often it is unclear exactly what boundaries or scope constitute a process, the expected outputs, who are the customers, who is the owner, and who is responsible for continuous improvement. Duplicity is inefficient and costly. Examples include multiple layers of an organization performing data reviews as documentation or information moves through the value chain. Regulatory submissions for analytical validation are an

example, where raw data may be checked at the laboratory, supervisor, quality assurance, compliance, and regulatory group levels. On the other hand, gaps may exist where each functional group listed above assumes data verification is occurring with another group, and in fact there are gaps in data integrity. In this case, the result can be tremendously expensive if, upon regulatory inspection, errors are found and it appears data integrity issues are ubiquitous in a submission.

This section will discuss the importance of defining business requirements to ensure processes comprising the QMS are designed to support the enterprise, integrated into a quality plan, sufficiently defined to provide adequate resolution and are transferable and scalable throughout the enterprise.

### 4.1 Defining Business Requirements

The QMS and the processes that comprise it must be custom designed for the needs of the business. One size does not fit all situations. The requirements of an enterprise vary across sites and the phases of a product life cycle. A comprehensive system will ensure a holistic programmatic approach in its support to the enterprise. This does not mean that every phase of the product life cycle (discovery, development, commercial manufacturing) will utilize all the processes that comprise the system. Nor does it require that all commercial manufacturing sites will necessarily implement all processes. It does, however, provide a common platform and expectation for all processes, owners, metrics review programs, continuous improvement efforts, and the like when they are implemented.

The first step in designing a QMS is determining business needs and the processes required to support the enterprise. Important consideration must be given to ensure that all processes are included in the assessment. The assessment must include all activities that affect product quality at corporate, business, manufacturing, distribution, contractors, or joint venture sites. Processes controlling incoming materials from vendors, laboratory services, contractual support, and other inputs should also be included in the initial assessment.

Upon identification of the processes required to support the enterprise, the next step is to define exactly what is in and out of scope for each process. Mapping all the processes and their inter relationship with other processes will determine if any gaps or duplication exists in the system. Duplication may be warranted or eliminated. Gaps between processes require remediation. For example, a nonconformance process should have direct linkage into a corrective action process. A well-operating nonconformance process without an active, integrated corrective/preventive action process will yield little benefit to the organization and efforts expended on the nonconformance process will be nominal in their overall positive business impact.

This comprehensive approach allows for efficient integration between processes, different phases of product life cycle, and integration between different sites in the supply chain. This integration provides opportunity for efficiency in that process owners are integrated with each other's needs and expectations. Duplication of effort is avoided and efficiencies gained. Quality outputs from one process become reliable inputs into the next process. Management and leadership will have access and insight into compliance, infrastructure, and performance metrics of all processes on a comparable basis. This provides leadership the opportunity for risk-based resource allocation to appropriate areas of the enterprise.

Process mapping of the enterprise's requirements to supply product enables design of the processes required for the system. Staff, management, and leadership input into the business needs provide additional guidance into processes attributes.

### 4.2 Integrating Quality Management System into Quality Plans

A quality plan is required by the regulations governing medical devices (QSR) but can readily be adopted as a useful tool for pharmaceutical and biopharmaceutical manufacturing operations. A quality plan is the documented plan and goals for enhancing and advancing the QMS. It can provide the outline and requirements of the organization's purpose, mission, product, and business practices used to produce a quality product. A quality plan can detail the processes that comprise the QMS, the maturity level required for each process, organizational structure, and other requirements needed to meet the organization's purpose. Included in the quality plan are the elements of the business including location, size, products, and expectations. It also includes its structure and support functions, values, and other attributes of the organization.

An annual quality plan can be the detailed execution plan of the organization's long-term quality vision for the QMS. It provides staff and management the outline and goals for improving the QMS. It enables employees to see the big picture, how they fit into the organization, and the organization's expectations. Within the quality plan attributes of the QMS should be described, including functional management responsibility. This then becomes the foundation for further definition of processes, description of management review and responsibility, and continuous improvement programs. The preparation of a quality plan begins defining what is assumed to be known by all levels of the organization. It is the mechanism for ensuring requirements are addressed and gaps in the organization do not exist.

A quality plan may outline the organization's long-term (several years) and short-term (annual) goals through a risk-based approach to improving product quality. It is the foundation for the manufacturing structure and support processes. A quality plan ensures integration of personnel, their qualifications, product requirements, quality management system, and regulatory and compliance infrastructure. An example of an outline of a quality plan is in Table 4. Leadership review and approval of the quality plan is required to ensure that mission, scope, expectations, and division of labor in the organization is consistent and supported.

In larger organizations, site or suborganization-based quality plans can be designed to support the scaling of the QMS across all components of the enterprise. The individual site plans provide focus on process challenges that are more critical than at other sites due to variations in business and compliance environments. While the specific plans emphasize goals based on site priorities, they also connect the members of an organization to the mission of the greater QMS, as shown in Figure 5.

### 4.3 Determining Process Resolution Requirements

Leadership expects cost-efficient reliable results from their manufacturing operations. Management requires a capable workforce, equipment, facilities, and materials to manufacture the product. Employees require robust processes that are easy to

**TABLE 4 Elements of QMS Annual Plan**

| Element | Definition |
|---|---|
| Introduction | Purpose of plan and definitions for clarity |
| Plan | Planned activities for the calendar year |
| Goals | Specific/cascading goals of the site |
| Projects | Major projects in support of the goals |
| Metrics | Key metrics with defined targets |
| Approvals | Site/plant management |

**FIGURE 5** Scaling the QMS through site quality plans.

execute to perform their jobs. All this needs to be considered in the design of the processes that comprise the quality management system.

Complex processes may need to be managed as distinct subprocesses in order to provide process owners the ability to accomplish their work with specific focus and expertise. Management and leadership may require data and metrics on specific areas of the process that are not available if the process is too complex and large. Dividing a complex process into simpler, more manageable processes also allows for scalability and transferability throughout the organization.

Once processes have been defined for the enterprise, sufficient system resolution should be determined. This is accomplished by evaluating the ability of the process owner to manage and execute the process requirements. Another factor in this determination is the data and metrics needed from the process by management and leadership. An example of a complex process that benefits the organization by being managed through distinct subprocesses is validation.

Validation is a regulatory requirement and has become an industry standard for ensuring product consistently meets quality attributes and regulatory requirements. Validation requirements are woven throughout the manufacturing supply chain encompassing many different subprocesses. The validation process may best be

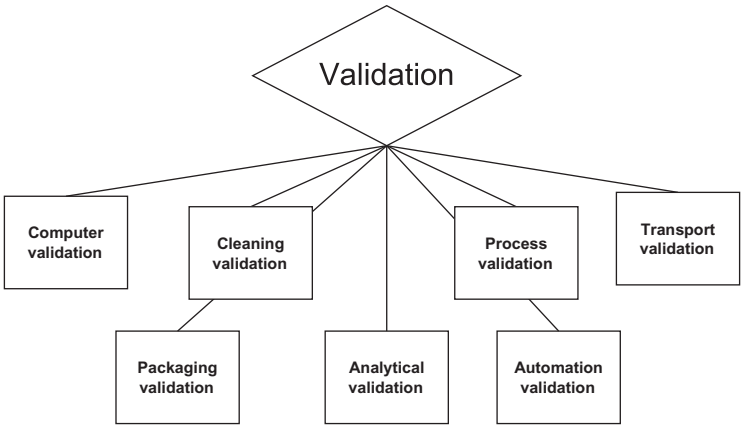


**FIGURE 6** Validation subprocesses.

managed by dividing it into manageable subprocesses. This allows for efficient management and execution of the subprocesses, and the metrics reported for those subprocesses are meaningful and specific. See Figure 6, which illustrates one potential organization of the validation subprocesses. Subprocesses contained within the validation system could be cleaning, computers, automation, analytical, packaging, process, transport validation, etc. Manufacturing is another example of a large, complex process that may best be subdivided to support better management and more meaningful metrics to management.

By dividing a larger process into manageable and specific subprocesses, management can assign appropriate subject matter expertise to lead and manage each subprocess. The metrics measuring subprocess performance can be uniquely reviewed, evaluated, and compared to similar subprocess metrics at other sites or companies. Valuable, meaningful comparisons can be obtained for process and subprocess performance that would otherwise be blinded or diluted, if they were summarized within the higher level process metrics.

An additional advantage of establishing subprocesses is that it affords the opportunity for rapid assimilation and transfer of the subprocess at various sites within the enterprise. An example of this is the comparison of a bulk manufacturing facility with that of a distribution center. Both will need to implement aspects of the subprocess "transport validation," however, the distribution center will not need to implement other subprocesses such as process or packaging validation. As the subprocess transport validation is designed and implemented at one site, that infrastructure and knowledge transfer to the other site is rapid, avoiding duplication of efforts. Sharing of information and expectations of the two sites becomes a common goal and format. Management can, therefore, compare transport validation needs and maturity levels between sites equally.

Once all the processes and subprocesses supporting an operation have been defined, another gap analysis may be conducted to ensure that there are no assumptions, and all required processes and subprocesses required to support the business are included in the scope of the QMS. This can easily be accomplished by listing all the business drivers for an operation and comparing that against the processes

established. Written definition of the process and subprocess scope is required. Stakeholders, owners, and users of the processes should be involved to ensure clear definition and understanding of process scope. Questions to be asked are: Are all the business needs addressed? Have all our activities and operations been included in the assessment? Is there any duplication in process expectations? Are there any gaps between process outputs and inputs? Once this evaluation has been concluded, it can be easily determined if any existing work process have been overlooked and the system requires further modification.

### 4.4 Scalability to Enterprise

Well-designed processes and subprocesses are scalable to the enterprise. A comprehensive design will allow for replication and comparison of processes and subprocesses between multiple sites. This allows for rapid implementation of new technology, sharing of best practices, and comparison of similar metrics to determine compliance, infrastructure, and performance. A comprehensive system allows for each unit operation or site within the enterprise to have the flexibility to apply applicable processes and subprocesses, yet continue operating within the defined structure of the QMS. For example, a manufacturing site may utilize almost all of the processes discussed in the validation process, whereas a distribution site may only utilize the transport process. Both sites, however, implement the same structure for the transport process, allowing for meaningful comparison of data and metrics and rapid implementation of any required changes to that process.

Well-designed quality management systems support structured organic growth and are valuable in evaluating and integrating manufacturing acquisition opportunities. Business and manufacturing management should utilize the QMS and its standards whenever evaluating external facilities for appraisal, approval, integration, or expansion. Meaningful metrics obtained from a QMS provides the standard to make critical decisions affecting multiple internal or external manufacturing capabilities.

Documented process structure provides rapid employee assimilations when transferring employees between sites. New employees, replacing existing process owners, are enabled to rapidly execute process responsibilities due to the abbreviated learning curve when processes have been well defined and documented. Systems designed as described here provide meaningful and comparable metrics for leadership to evaluate progress, compliance, and performance.

## 5 SYSTEM AND PROCESS OWNERSHIP: ROLES AND RESPONSIBILITIES

A well-designed QMS and the processes that comprise it require competent ownership with defined roles and responsibilities for program success. This combination ensures that the system and processes are established, maintained, improved, and remain current with industry practices and business expectations. Operational execution of the QMS and the processes comprising it will engage stakeholders, management, and leadership, provide business results, and support and ensure compliance.

### 5.1 Quality Management System Ownership and Management

The QMS is best owned at the highest level in the organization. At a minimum it should be owned at a level in the organization above manufacturing and quality. The owners' main responsibility is to champion the program and ensure organizational alignment. Regulatory investigators expect processes supporting manufacture are fully incorporated into the QMS. They also expect leadership to have significant knowledge of the operations and interact with investigators during inspections with some degree of familiarity with the processes supporting manufacture. At the conclusion of an inspection, regulators issue inspectional findings and, if appropriate, take regulatory action against the most senior member of the leadership group. Through high-level leadership's active involvement and ownership, the QMS program and enterprise will be successful.

As mentioned previously, the QMS is best managed by a group dedicated to the program. The QMS program office should have defined roles and responsibilities. In the FDA regulation 21 CFR Part 211.22, the responsibility of the quality unit is described. It is the only functional group in a manufacturing organization that has its job description codified in federal regulations. These responsibilities should not enable or dilute the responsibility for ensuring quality of other functional groups in the organization. All functional groups supporting manufacture should be applying their trade to the GMP world. Regulators expect the Quality Department to have oversight and approval of all processes affecting product quality. Program management is important because of the need for coordination and accountability to bring individual processes, long-term system strategy, yearly quality plans, and goals together to accomplish the program's objectives. These activities and benefits cannot be realized from individual process owners.

The program management of the QMS can be managed just as a process, with predefined expectations, metric collection, and management review, culminating in risk management application to continuous improvement programs. These metrics and improvement initiatives need to be vetted through leadership review and input to ensure alignment throughout the organization.

An outline for the roles and responsibilities for the QMS program office is illustrated in Table 5. By establishing the roles and responsibilities of the program office a defined point of contact and accountability is established for program execution. It establishes strong linkage and focus on the program objectives for process owners, training, functional management, and leadership. Similar program management structures are required at manufacturing sites and corporate functions to maximize benefits of the program through establishing common and specific goals and

**TABLE 5 QMS Program Office Roles and Responsibilities**

| Subject matter expert for QMS program |
|---|
| Develop and execute communication plan |
| Initial and ongoing training |
| Facilitate management review process |
| Identify process maturity goals and metrics |
| Develop long-term strategic vision |
| Create and execute annual action plan |

providing a platform for sharing best practices and knowledge. As organizations grow in complexity, additional management may be required to ensure that the elements of the QMS are integrated, functioning, and delivering the results expected.

### 5.2 Process Ownership

Designing a QMS that mandates and assigns process ownership to designated individuals is a significant strategic decision in the establishment of a successful quality management system. It provides efficiency, expertise, dedication to the process, and focused ownership for documentation, improvements, benchmarking, and compliance. Without defined and assigned process ownership functional management becomes the de facto process owner. This is problematic in that functional management is already overburdened with personnel and business management issues, unable to adequately focus and deliver the demanding process owner requirements required in today's manufacturing environment. Several processes typically are organized under an individual functional manager, further diluting focus, attention, and expertise if functional management is relied upon as a process owner.

### 5.3 Process Owner Selection

Process owner selection requires program management to establish defined criteria for the selection process. Criteria include the capability to perform process owner roles and responsibilities, including self-development and decision making. Empowered process owners are accountable for maintaining and executing the processes that management relies upon to deliver business results. This accountability ensures that staff, management, and leadership know who to solicit for answers to process-related questions and issues. It also provides the best representation to regulators, clients, and customers. Efficiencies are gained and current trends maintained with an active owner, with defined responsibilities.

Selection criteria may include attributes of technical, interpersonal, and management skills. The capabilities needed for different process will vary and should be considered in the selection process. At the end of the selection process, functional management and the process owner may consider inclusion of the process owner's roles and responsibilities into the process owner's job description. Personal goals and development activities should be based on improving the process owner's capabilities to manage the process and develop future process owners through active mentoring and talent development programs.

Process owners need to be dedicated to their process. They must be empowered and held accountable for all the attributes listed in their roles and responsibilities. Process owners may have ownership of more than one process and may have other job responsibilities, but it must be clear throughout the organization as to who has full authority for the process.

Process owners require a defined set of responsibilities to maintain a vibrant and effective process that continues to support product quality deliverables. Having roles and responsibilities defined provides owners with the structure and parameters

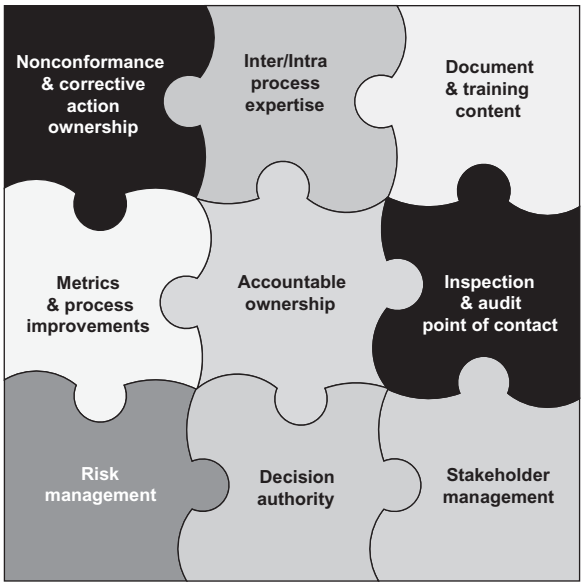


**FIGURE 7** Process owner roles and responsibilities.

needed to be effective. Examples of owner responsibilities include identification of stakeholders, defined decision authority, document ownership, nonconformance ownership, knowledge of regulations and industry trends, subject matter expertise, training content, metric ownership, and representation to internal auditors and external regulators. Identifying, training, and development of the process owner on his or her roles and responsibilities is similar to assembling the piece of a puzzle. If one piece is missing, the effectiveness of the process owner will be minimized (see Figure 7). Developing a sound methodology for process owner selection ensures objectivity and is critical to the success of the program. A brief discussion on each process owner roles and responsibilities follows.

### 5.4 Stakeholder/Process Owner Integration

Process owners must identify the stakeholders of their process and ensure design and output of the process meets stakeholder needs. Regular communication, interaction, and support are maintained with stakeholders through scheduled meetings to discuss process status and improvements. Any changes to the process are vetted through the stakeholder group. Typical stakeholders include the QA unit that is responsible for review and approval of the process components, suppliers to and receivers of the process (i.e., owners of other processes that interact with the process), customers, management, and leadership. Each stakeholder roles and responsibilities also require definition.

Including the key stakeholders in decisions affecting process design or changes ensures efforts by the process owner are applied correctly. Process robustness is dependent upon meeting business and customer needs, and the process owners require input and support from the stakeholder group.

### 5.5 Decision Authority

Each process owner requires a defined level of decision authority. This authority level delineates the bounds of decision making granted by the organization to the process owner. Business needs and risk assessment must be incorporated into the design of the decision authority granted to a process owner. Table 6 is an example of a decision authority matrix design for a process owner. It requires cross-functional management support to be effective.

Preparing a decision matrix that is shared and agreed upon by the stakeholder group and functional management ensures decisions are made and communicated quickly by appropriate persons. It removes the burden of making every technical process decision from functional management. It is important to outline the process owner's role in the decision-making process, as well as conditions for escalation. Effective process management is realized when the culture of an organization can support the outline of the decision matrix and not continually rely upon functional

**TABLE 6 Decision Authority Matrix**

| Decision Category | Definition | Decision Maker | Decision Support Required | Informed of Decision |
|---|---|---|---|---|
| Company standards | Process-related global standards relevant to all manufacturing sites | Corporate process owner | Site process owners<br>Site quality assurance counterpart | Process stakeholders<br>Impacted staff |
| Standard operating procedures | SOPs related to the specific process | Site process owner | Process stakeholders<br>Site quality assurance counterpart | Management review<br>Corporate quality assurance counterpart<br>Impacted staff |
| Training | Training on processes or procedures | Corporate and site process owners | Training<br>Technical system matter expert | Process stakeholders<br>Corporate quality assurance counterpart<br>Impacted staff |
| Site projects | All projects related to the existing process or the projected improved state of the process at a specific site | Site head | Process owner<br>Leadership team<br>Site project portfolio manager | Process stakeholders<br>Corporate quality assurance counterpart<br>QMS office |

management. If functional management continues to be relied upon and seen as the process decision makers, efforts and progress by the process owner will be nominal. The organization's culture must support the process owner at all levels of the enterprise for the owners to be successful.

### 5.6 Industry Knowledge

The *c* in *c*GMP represents the notion of current industry practices. Process owners must work to remain current with industry and regulatory trends affecting their process and its overall effect on the QMS and the business. Awareness of process capability and comparability with other like processes within and outside the pharmaceutical and biopharmaceutical industry is essential. Regulators will compare an owner's process against other similar processes in which they have experience when formulating value judgments. Benchmarking against similar processes provides process owners the data needed to determine adequacy of their process with industry peer groups.

Where technology, efficiency, performance, or compliance can be enhanced, it should be considered by an aware and informed process owner. Functional management cannot keep pace with the changes occurring with all the processes supporting manufacturing. Ensuring process owners dedicate sufficient time to keeping current with process-related external events will ensure process success. This may include review of industry periodicals, attendance at seminars and regulatory presentations, and routine self-evaluation and benchmarking against relative processes.

Often, the best examples of process efficiency can be found outside the pharmaceutical and biopharmaceutical industries. Other fields such as electronics, space, and software industries have evolved their documentation, training, quality, and change control systems to the point of best in class. These industries are more time sensitive to get product to market and have often evolved their processes to be efficient and decision processes to be very quick. Process owners may expand their knowledge by investigating other industries to find best practices and apply them internally.

### 5.7 Regulatory Inspection and Audit Lead

Process owners play a critical role during regulatory inspections and customer audits. Process owners are the best choice to represent process attributes and performance to interested parties. Process owners provide regulators and auditors with a capable, knowledgeable resource to represent the process and answer detailed questions. The process owner should be aware of process history, requirements, operations, exceptions, changes, and nonconformances. The process owner will have detailed knowledge of process operations, compliance, and be able to defend the process. Providing an accurate answer the first time to regulators and auditors is essential in building trust and representing competence.

Each process owner is required to work closely with his or her QA counterpart. This ensures design and operational issues are clearly reviewed and approved by a representative from the quality assurance function, a regulatory expectation. The quality assurance counterpart must be familiar with the process, understand documentation supporting the process, and able to convey what approval the Quality

Department has conveyed on the process and meaning of that approval. The QA counterpart to a process should also have defined and documented roles and responsibilities.

When teamed together, the process owner and the quality assurance counterpart for a process will make a favorable impression upon regulators, be able to explain all operations involving the process, supporting documentation, and any ongoing projects or process improvements. This pair is the best to evaluate and consider any deviations to the process or recommendations for continuous improvement.

In most all cases, the process owner and QA contact will posses more information about the process than regulators and be capable to defend the process design and operation. Should regulators have suggestions on process design or functionality, the process owner may consider them. If appropriate, any recommendations or observations made by regulators or auditors can be incorporated into the process design. However, it is critical for the process owner and stakeholders to evaluate proposed changes to avoid reactive management commitments, which could be deleterious to the efficient operation and output of the process.

## 5.8 Subject Matter Experts

Process owners, through their selection and development, become the subject matter experts for the process. It is more efficient for an enterprise to focus its expertise on individuals that have the authority and accountability described in this section, rather than dilute those attributes and accountabilities, thereby risking poor process execution and management.

As a subject matter expert on the process, the owner has the capacity to deliver results outlined in the list of owner responsibilities, mentor future process owners, assist in staff development, and accurately guide management in its strategy related to the process. The process owner's personal development of his or her process expertise is essential in delivering operational results and providing direction for future strategic changes to the process.

## 5.9 Metric Ownership

Process owners' responsibilities include determining appropriate metrics for their process. These metrics should include lagging and leading metrics that are meaningful to the process owner and management in determining performance, compliance, and infrastructure of the process.

The process owner should represent and interpret these metrics to the organizations leadership. The metric output from a process is the basis for management and leadership's action in resource deployment and approval of continuous improvement projects. Key operating parameters such as number of nonconformances and regulatory observations against the process should be tracked and factored into the maturity of the process.

Every process owner needs to base their continuous improvement plan for the process based upon metrics collected from the process output. The metrics must be designed to assist in these decisions and be readily available for review, presentation, and interpretation.

### 5.10 Documentation Ownership

Process owners are the most appropriate owners for all documentation supporting their processes. This includes having either direct ownership or controlling influence over guidance and execution documentation such as corporate policies and standards, local requirements and standard operating procedures (SOPs), logs, and records.

For the manufacturing process owner, this means owning the master manufacturing records and executed batch records, SOPs, use logs, and related training documents for their process. Combining responsibilities for process management and process ownership results in true accountability for the process owner. It also allows for progress and continuous improvement of the QMS. Removing questions of responsibility and accountability ensures integration between requirements (standards, policies, procedures) and execution (training, performance, and documentation).

### 5.11 Training

Assurance of adequate training for process users is an important responsibility of a process owner. Process owners must have a clear understanding of the requirements of their process and its operation. This understanding requires translation into executable training. Users must be able to understand and apply the training. Complicated processes coupled with ambiguous training will lead to confusion and an inability to properly execute a process, which eventually constitutes failure for the organization. A simple process, with easy to understand process steps, that are consistent with instructions and documentation requirements will support success, reduce production costs, minimize nonconforming events, and allow for employee satisfaction.

Process owners are subject matter experts and should influence and provide consulting for training on the process. They may also participate in training delivery. Ensuring adequate training on a process is a key goal for system efficiency and regulatory compliance. Process owners, capable of explaining the reasoning behind the process requirements, enhance the training experience for process users. Process owners should include effective presentation and training skill development into their personal development programs.

### 5.12 Risk Management

Process owners require basic understanding of risk management and its application to process design and continuous improvement prioritization. Several industry and regulatory resources exist, such as ICH Q9, that provide understanding on risk assessment, identification, control, methodology, and the overall risk management process. Process owners should be familiar with risk management techniques and tools and apply them to their process management when designing, executing, or managing improvement efforts for their process.

Risk management is especially important for the presentation of process improvement proposals to management where resources are required. The ability to quantify risk and demonstrate continuous improvement benefits is essential to project

and resource approval. Risk analysis, management, and presentation constitute guiding leadership to work on the right things at the right time and then improve it.

## 5.13 Continuous Improvement and Project Management

Instituting quality-by-design efforts early in the design of a process should negate the need for major process improvements. However, over time, due to business needs, regulatory changes, or technology improvements, processes will require some form of change to ensure compliance or performance enhancement. As part of executing and maintaining their processes, owners need to collect and report performance metrics to management and staff. These metrics will inevitably direct attention to opportunities for improvement that require capital and human resources. Process owners are the best leaders of continuous improvement projects due to their intimate knowledge of the process and accountability for process output.

Managing or leading a continuous improvement project requires process owners be knowledgeable in project management and team-leading skills. Improvement projects typically require cross-functional support and expertise from areas such as information systems, project management, manufacturing, engineering, and development. It is essential that continuous improvement efforts name the process owner as the project lead to ensure the required output of the project meets the process owner, stakeholders, and enterprise needs. Often, projects are completed and declared a success, delivering a substandard result that the process owner, users, and stakeholders find inadequate to meet process requirements.

Upon completion of continuous improvement projects, process owners' responsibilities include monitoring the changes made to the process to determine the impact of improvements. Metrics monitoring changes to the process, pre- and post-implementation, should be incorporated into existing performance metrics and reported during regular management reviews.

## 5.14 Non Conformance/CAPA/Planned Deviation Ownership

An important barometer of process performance is the number of nonconformances, corrective and preventive actions taken, and planned deviations initiated against a process. These types of process artifacts must be known and owned by the process owner and stakeholders. The process owner must consider these process metrics for evaluation of and changes to process design, training, documentation, and performance.

Nonconformances may fall into the category of manpower, machinery methods, materials, etc. Employees not following procedures or unable to execute required steps of the process indicate a poorly designed process requiring modification and/or improved training. Machinery failures often indicate poor qualification, validation, calibration, or maintenance programs. Unexpected results or outcomes are indicative of poor process design, characterization, or a break down between processes.

Although planned deviations are frowned upon by many in the industry and regulators, there are times when temporary changes to a process must be employed to support the business. Permanent changes must be made through a formal change

control process. When the use of a planned deviation is required, the affected process owner should be aware of and own the change. This provides owner control over the duration and extent of the change to the process and provides data for possible consideration in making a permanent change to the process. A planned deviation should be rare and monitored closely as it affects previously established standards, expectations, and training.

Process owners must be capable to evaluate and interpret the effect of nonconformances and planned deviations on their systems. Process owners can evaluate the need and lead efforts for corrective or preventive action, ensuring adequate corrections and improvements are implemented. An effective QMS ensures deviations from approved processes are owned and adequately investigated by the process owner's and ultimately approved by their quality assurance counterpart. The knowledge of these events is the basis and foundation for the process owners to make a risk-based evaluation on whether or not process changes are required, documentation or training require modification, or continuous improvement efforts are warranted.

A well-designed QMS will include identified process owners with defined roles and responsibilities. Process owners require support from management, their customers, stakeholders, and quality assurance. Accountability and decision-making parameters will empower process owners to drive execution and improvements to their process, delivering the business results expected. Without these process owner attributes and support, minimal results will be achieved, and functional management will be burdened with and assume the responsibility for making decisions that should be in the hands of capable process owners.

## 6 CHANGE MANAGEMENT/COMMUNICATION

Establishing and maintaining an effective QMS, as this chapter describes, requires a significant cultural shift. Many employees and functional management will find the business transformation of defining processes, assigning ownership, delegating authority, and responsibility for process performance within the QMS is a significant change in business conduct. The most significant change results from the shift of control in process expertise and decision-making authority from functional management to process owner. Significant business transformation may result by assigning responsibility and accountability to the process owner, and management's support of process owners who drive continuous process improvement.

In *The Second American Revolution*, Rockefeller describes the conservatism of organizations: "An organization is a system, with a logic of its own, and all the weight of tradition and inertia. The deck is stacked in favor of the tried and proven way of doing things and against the taking of risks and striking out in new directions" [9]. If an organization is not already practicing principles of delegation, process ownership, established metric collection, management review, and continuous improvement, barriers within the organization will need to be addressed and broken down in order to establish new behaviors. These barriers to change will exist within and between functions, functional management and staff, and possibly between companies and regulators.

Although expected benefits are significant when implementing a QMS and the end result desirable for employees and management, describing the desired state and motivating personnel to change and implement new behaviors contains significant challenges. A successful business transformation requires a robust change management and communication plan that includes support for all staff affected.

### 6.1 Managing Organizational Change

Integration of the skill sets of human resources, training, and change management groups will significantly augment efforts toward cultural change and acceptance. Often personality profiling tools are effective to gauge the organization's preferences, learning styles, and adoption tendencies. These types of tools should be considered in the overall change management program, used where applicable, and program modifications made based upon their results.

The first and critical step in developing a successful change management plan is to obtain initial support from the corporation's leadership and functional management. Without this support the QMS will not gain critical mass and may not deliver the desired effects or changes. To acquire this support, the implementation team must put together a strong business case that speaks to the leadership's needs and wants. The business case must include a risk assessment against compliance and the benefits of the financial gains. It is important to be honest, consider current system status and future requirements, and include a long-term strategy that addresses costs and benefits. The change management plan must include frequent and repetitive communications, to all levels of the organization, of the cost/benefits and successes expected and realized by the program.

Functional management support is also critical to the success of a new program such as a quality management system. Any time a staff member is asked to take on a new role or responsibility, he or she needs to be supported by the functional manager as well as leadership within the organization. Corporations are resource limited and necessarily need to continually prioritize where to allocate resources. Staff will only take on roles or responsibilities that they believe are supported by their functional manager in an effort to successfully meet their perceived immediate goals. Quantifiable support from leadership and functional management can be directly correlated to the success or failure of the QMS program.

Significant work is involved in training new process owners, functional managers, leadership, support organizations, and actualizing their new behaviors. A support system must be in place for the process owners, stakeholders, and management to guide and reinforce the new behaviors and maintain the process effectiveness. It is preferable that this support system be established through a dedicated team that can be fully attentive to all their needs. Without a single source to lead the efforts, diversity in interpretation and implementation will dilute the program, within different functions and sites, and its effectiveness and outcomes will be diminished.

Establishing an organization to lead the systems initiative is important. That organization requires management, standards, and parameters similar to managing an individual quality process. It requires roles and responsibilities be established,

metrics be determined, collected, reviewed and acted upon, and receive management and leadership visibility and support. The QMS program is best organized as a function within the Quality Department and be regarded as an ongoing program, not a short-term project or effort with limited shelf life. The group must be led by competent persons who are familiar with quality concepts and applications, regulatory expectations and requirements, needs of the enterprise, good communication and influencing skills, and are flexible and enduring.

### 6.2 Communication

Trying to get people to comprehend a vision of an alternative future is also a communications challenge of a completely different magnitude from organizing them to fulfill a short-term plan. It is much like the difference between a football quarterback attempting to describe to his team the next two or three plays versus his trying to explain to them a totally new approach to the game to be used in the second half of the season. Aligning the organization to accept and implement a system-based approach requires careful messaging coupled with management support and results.

Messages are not necessarily accepted just because they are understood. Another big challenge for leadership is credibility and getting people to believe the message. Aligning words and deeds supports the worthiness and credibility of the messaging. People have learned from experience that even if they correctly perceive important external changes and then initiate appropriate actions, they are vulnerable to someone higher up who does not like what they have done. Reprimands can take many forms: "That's against policy," or "We can't afford it," or "Shut up and do as you're told" [10].

Having established a dedicated team that provides overall program management, it is imperative that the team outline a strategic plan for presentation to leadership. Without a vision and long-term plan, which is supported by the enterprise leadership, quality system initiatives will become difficult. The plan needs to be comprehensive in nature, yet broad enough to convey purpose, mission, and benefits at a high enough level to be understood and supported. An outline such as this provides framework and direction for the program management team and leadership. It also guides the program management team to developing annual goals and quality plans that fit into the overall strategy and provide momentum and results to the organization.

Annual quality plans should be prepared by the QMS program office that address the long-term strategy and intermediate goals that come to surface during program implementation. Training, changes in regulatory requirements, metric-driven projects, and special circumstances warranting process changes such as implementation of new technology or programs should be included into the annual quality plan.

Long-term strategy documents and annual quality plans require leadership and functional management support and approval. These documents must be reviewed and discussed with the leadership of the organization, modified to meet the business and regulatory requirements, and then have full support through upper management approval. In this way, the goals are being led by top leadership and management and not any individual group in the organization. Once top leadership signs

onto the program, it can be shared throughout the organization in a number of ways.

If leadership can support the long-term strategy and annual quality plans to accomplish the vision, then the foundation for change management and cultural shift is in place. Leadership will need to continually discuss the need for systems implementation, in front of a variety of audiences. This includes leadership staff meetings, management, and employee meetings. The importance of leadership support cannot be overlooked. Without consistent visible leadership and management support process owners and staff will revert to old behaviors, become reactive, and perhaps unrelated in their process integration efforts. Leadership needs to require aspects of the program be included into functional management annual objectives with defined deliverables outlined and evaluated. Likewise, functional management should require staff to include appropriate aspects and objectives of the QMS program into their individual goals and work to accomplish them.

### 6.3 Feedback and Alignment

Managing the changes required to fully implement a QMS can include several forms of communication and feedback. A detailed annual communications plan can aid the QMS's group in identifying specific target groups, methods, and frequencies of communications, messaging types, and feedback mechanisms to monitor progress for program modifications. Table 7 is an example of an annual communications plan that supports efforts to keep internal audiences informed, aligned, and engaged.

Each target audience requires specific messaging that connects with its needs. Failure to get the appropriate message, that is, what is the program bringing to them, will minimize support for the program. This plan should include face-to face and written communications addressing multiple audiences and media types. Face-to-face meetings can include presentations to steering committees, process owners, functional departments, and all staff meetings. Written communications can include sitewide communications, poster sessions, and newsletters. The communications should speak to all audiences—"what's in it for me?" Topics can include leadership's commitment (direct quotes or actions taken); spotlight on successes (real-life stories from process owners); impact to the site (process improvements or risk mitigation); and progress to the program (metrics and successes). The progress and success of the QMS cannot be overcommunicated.

Another useful tool to help the message and modify the program is the use of a feedback survey. If properly designed and distributed to a defined set of stake-

**TABLE 7 Communication Plan**

| Vehicle | Communication Type | Frequency | Date |
|---|---|---|---|
| Functional metrics meeting | Face to face | Monthly | First week of month |
| Management interviews | Face to face | Annually | January 1–31 |
| QMS newsletter | Written | Quarterly | First week of quarter |
| All staff meeting | Presentation | Semiannually | March & September |
| Poster session | Written/face to face | Annually | July |

holders and employees, the survey can provide valuable insight into how staff and management view the program, its progress, and suggestions for modifications. If surveys are distributed electronically and offer only one-way communication, the benefits may be limited as the respondents are limited in their ability to fully convey their impressions or offer effective feedback. An electronic feedback survey may be a first good step in understanding the thoughts and concerns of the stakeholders.

Another suggestion or follow-up to the electronic survey is to utilize focus groups that have the ability to interact with the program questioners. This two-way conversation, verbal dialogue, allows further understanding of the program by the participants that follows with more meaningful feedback to the program administrators. Focus groups should be selected at different levels within the organization, including process owners, stakeholders and users of the system, leadership, functional management, and the general populace of employees. Focus groups provide valuable input into programs that the program administrators may be unaware of and can provide program redirection.

Once suggestions are received on the program, it is essential to consider and incorporate those ideas and modifications that make sense to implement. Those changes need to be communicated and seen by the focus group members to ensure that their time and effort has not been wasted and their suggestions have been heard. This is one of the best ways to spread the word about the QMS program and garner grassroots support.

### 6.4 Training

A training plan should be developed to identify the needs of the staff and affected functional areas required to support the successful implementation of a QMS. It is the responsibility of the corporation to adequately support staff with training and tools when staff is expected to take on new roles, responsibilities, or behaviors. The training plan should consist of targeted training for general staff, process owners, and functional management of the process owners.

At a minimum, all staff should be introduced to the purpose, goals, and requirements of the QMS. This training should be a high level explanation of the program looking to gain understanding and support for the program by communicating why it is important and what are the risks of not adopting the program. This can be accomplished by instructor-led training or an electronic, Web-based learning module depending on the size of the corporation.

Process owners require more comprehensive levels of training to fully understand their role and responsibilities within the program. Process owner training should teach key concepts and tools that owners will need to evaluate and support their processes. This training can be done in a phased approach to support the elevation and advancement of a process within the organization's chosen maturity model.

Training should be provided to offer functional managers supervising process owners a thorough understanding of the QMS. Training should address new roles and responsibilities of staff, time demands on process owners, overall program timelines, and impact to functional areas. The acceptance and support of functional management is critical to successful implementation of a QMS.

Managing organizational change demands a well-written strategy, skill set, and resources to ensure changes that come with system implementation are understood, supported, and maintained. Starting with high-level overviews of system design, benefits and timelines for implementation are the foundation for management understanding and support. Detailed annual quality plans can be the tactical vehicle for program implementation. Leadership support through understanding and approval of the annual quality plan, inclusion of program objectives into management goals, and frequent verbal and visual support of the program are essential to success. Building the program infrastructure is a significant undertaking. Inclusion of a comprehensive training, communication, and change management plan should be built into the overall goals of the program and routinely evaluated and delivered.

## 7 MEASURING SUCCESS THROUGH MEANINGFUL METRICS

Successful implementation of a comprehensive QMS can be determined by the establishment of a meaningful metrics program. The purpose of a metrics program is twofold: first, to allow an organization to evaluate its progress toward meeting its goals in an objective, data-driven manner and, second, to monitor the performance of each process to ensure continuous improvement. By evaluating metrics for the QMS and its processes, the enterprise has the knowledge and understanding of the overall health of its system and processes and can develop strategies based on risk for continuous improvement of the system and processes.

Once the metrics program is in place, the system and process metrics require visibility to process owners, upper management, and stakeholders. Process owners require understanding of the metrics' trends, issues, and associated risks. Stakeholders must work with the process owner to identify and propose process improvement opportunities. Leadership is accountable to understand the issues and associated risks and responsibly apply resources for remediation efforts.

### 7.1 Performance Metric Development

Quality and business indicating metrics should also be reviewed on a routine basis. These may include the following:

- Quality indicating: ability to meet quality standards and procedures
- Supply: ability to meet demand
- Cost: savings as well as avoidance
- Safety: near misses and incidents against process

The guiding principle of metric development is to have a stable system or process to collect, review, and draw conclusions. All metrics should be developed with stakeholders input taking into account the requirements and needs of the customers. This includes the touch points of the downstream quality processes. Without this input and understanding metrics may be developed within a silo and hold little value, causing both frustration at the leadership as well as the staff level. Without proper

design, metrics may become a check box activity that results in minimal or no action by management to support efforts by a process owner.

Metrics can either fall into one of two categories: lagging or leading indicators. Both types are important to the process owner and management. Lagging indicators are metrics that represent the process's ability to deliver results or outputs. They indicate the performance of the system in the past. They can assist process owners, management, and leadership in determining if goals have been met, objectives attained, or existing standards or expectations have been met. Leading metrics focus on the inputs and suppliers of a process. These metrics are important indicators to proactively allow owners and management to take action on a process prior to violating a standard, objective, or goal. A successful process owner will understand the relationship of leading metrics and their affect on the lagging metrics and process. Metrics need to be designed to meet the needs of the organization, be simple to track and present, and be regularly reviewed.

### 7.2 Metric Review

Ignorance of system and process performance leads to inefficiency, poor compliance, and low employee morale. It is good business practice to have regular review of process metrics to gauge the health and output of the system and processes that drive the organization.

Process owners should be aware of all the metrics affecting their process and have a conduit to present the critical metrics to upper management. There are examples in the industry where process owners responsible for execution of a process are not aware of the metrics being collected, if any are, and have no basis for judging the adequacy of their process or its performance.

Regulatory agencies hold management accountable for the operations of an organization. It is the fiduciary responsibility for process owners to share the output and performance of the operation with management and be able to explain and interpret those metrics. Management has the responsibility to know the operations, its performance, and take appropriate action to ensure compliance with government, industry, and company policies and regulations.

Regulations require an annual product review be conducted of pharmaceutical products to determine and assess changes made to processes that may affect product quality. However, good industry practices would mandate quarterly or monthly review for faster detection, decision, and action. Reviews need to include metrics on key operating parameters and critical quality attributes to ensure product safety and efficacy. Several other key business metrics also benefit the organization and should be included in the metrics review program. The metrics collected should easily provide the process owner and management with an indication if the process is in control and delivering the desired results. If not, the process owner needs to present management a proposal to pursue continuous improvement opportunities and be able to describe required changes necessary to realize process enhancement.

### 7.3 Maturity Model

A maturity model is a useful management tool to determine process status and provides a standard in which to value processes. It provides a standard in determin-

ing the overall robustness and progression of a process and assists in the determination of resource prioritization. It provides the basic framework to apply risk management in determination of process development. For example, development of high-risk systems, such as aseptic filling where high patient and business risk exist, should be developed to a higher maturity than other processes with less patient or regulatory risk. Business demands placed on the pharmaceutical and biopharmaceutical industry limit resources in development, quality, and manufacturing requiring wise deployment of these resources to the areas that can best benefit the organization.

An example of a maturity model can be seen in Figure 8. This example provides the QMS program group and leadership the ability to evaluate processes based on an objective standard. It is divided into five general levels, moving from informal, unstructured to best in class. It includes specific deliverables for each level of the model to be completed before a process can be considered to have achieved that level. This model can also be divided into distinct subcategories, for instance, infrastructure, performance, and compliance, which are depicted in Table 8. Each subcategory can be designed to provide meaningful information to the process owners

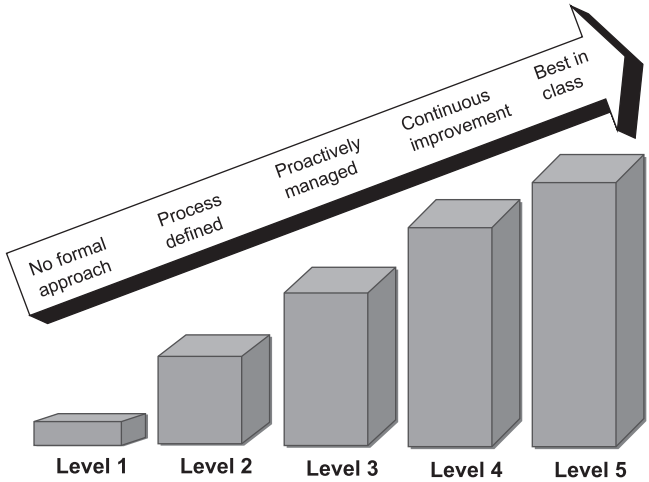


**FIGURE 8** Maturity model overview. (*Source*: Adapted from Capability Maturity Model Integration, www.sei.cmu.edu.)

**TABLE 8 Example Maturity Model**

| Theme | Level 1 (No Formal Approach) | Level 2 (Process Defined) | Level 3 (Proactively Managed) | Level 4 (Continuous Improvement) | Level 5 (Best in Class) |
|---|---|---|---|---|---|
| Compliance | → | → | → | → | → |
| Infrastructure | → | → | → | → | → |
| Performance | → | → | → | → | → |

*Source*: Adapted from Capability Maturity Model Integration, www.sei.cmu.org.

and management. The maturity model is an excellent metric to measure development of the QMS and focus leadership in deployment of resources.

The subcategories of the model demonstrate, through defined attributes that must be in place, specific areas required of a robust process. The infrastructure category includes a capable owner of the process is in place and a quality assurance counterpart is identified, the process owner has a strong understanding of the process flow, scope, process boundaries, suppliers, customers, and roles and responsibilities. The goal is to develop a highly integrated process that is fully transferable and scalable.

Compliance is a key process attribute for a process in the pharmaceutical and biopharmaceutical industry. Process maturity determination related to compliance can include documentation such as standards and SOPs, number of observations written against the process from internal audits, supplier audits, regulatory inspections, nonconformances, and a risk assessment on the process against patient safety and efficacy. Training programs are also required as part of the process compliance. Audit and inspection observations written against a process are key metrics indicating maturity. Processes that can meet high maturity level for compliance represents a well-managed process that is consistently delivering a compliant and quality output.

Performance metrics determine process performance, preferably against predetermined standards or expectations. Performance metrics should be indicators as to the health and robustness of the system. Performance metrics may include cycle turn around time, time to disposition from end of manufacture, and a risk assessment against business drivers. The purpose is to raise target performance objectives, developing a strategic approach, reducing variability, and improving efficiencies. Efficiencies gained in process performance contribute to the business needs.

Advantages to utilizing a maturity model are that it provides a useful methodology for the QMS program group and leadership to evaluate, grade and provide process owners a goal for process development. Again, using a risk-based mindset, the entire inventory of process can be evaluated by leadership to determine where to place resources and to what maturity level each process is best positioned to support the enterprise.

Maturity-level goals are best made by process owners, the QMS program group, and leadership. It is recommended that all processes are assessed against risk to the customer and the business. This allows the QMS to prioritize the processes and identify which of the processes need to be elevated to a higher maturity level in the maturity model. Upon completion of the risk assessment, results should be reviewed by leadership to determine if the processes have been prioritized appropriately and meet the corporation's goals. This feedback forum will ensure that leadership supports process owners as they endeavor to achieve a higher level of process maturity. A well-designed QMS will allow for two-way conversation between the leadership and the process owners. It is as important for the leadership to communicate priorities to the process owners as well as having the process owners communicate issues and concerns that need to be addressed to the leadership. This will improve the alignment of priorities between the leadership and process owners. This integration ensures that the corporation is working on the right things at the right time with the right people.

### 7.4 Meeting Process Maturity Requirements

A dedicated team or review board should be developed to review and approve all maturity-level deliverables upon completion of the attributes for the current level. This review board's purpose is to ensure that all deliverables meet a consistent level of quality and documentation. This board can provide feedback to process owners or QMS program group to communicate best practices and lessons learned.

A well-designed metric review program is essential to the success of the QMS. The program should include metrics for the QMS, process maturity-level assessment and process performance, infrastructure, and compliance metrics. These metrics are the basis for evaluating system progress against long-term vision and annual quality plans. The metrics provide leadership and process owners specific and objective data to determine program goal achievement. Leadership will have visibility and comparability of process performance within and between sites and have risk-based data to support their deployment of resources in addressing business issues.

## 8 DRIVING CONTINUOUS IMPROVEMENT: PROJECTS

Pharmaceutical and biopharmaceutical companies are under significant pressure to deliver consistent quality product as well as drive the overall product cost down. The goal of implementing ICH Q8, Q9, and ultimately Q10 is to characterize processes based on risk assessments and improve them through a well-designed QMS. There are regulatory and business drivers to continually improve the QMS processes by building in quality and improve process efficiency. The regulatory agencies are now focused on ensuring systems are in place that protect the public health by assuring both the safety and efficacy of products. Understanding manufacturing processes, through well-designed characterization studies, is one of the most efficient and effective methods to ensure process efficiency. To meet business and consumer demands as well as regulatory guidance and expectations, the implementation of continuous improvement through risk-managed evaluations of manufacturing processes is expected.

### 8.1 Process Improvements

A quality management system's process should follow a standard Six Sigma process improvement life cycle that includes the following steps: define (process and metrics), measure and control (identify problems and issues), analyze (analyze problems and issues), and improve (implement) circling back to measure and control [11]. An example of a process improvement life cycle can be seen in Figure 9.

The basic foundation of continuous improvement begins with a process owner who fully understands the process and recognizes how the process impacts other processes within the QMS. Understanding this cause-and-effect relationship between processes requires close integration between process owners and stakeholders. This integration is critical throughout the entire life cycle of a process, from design through development and management.

Prior to process improvements the process must be well-defined and predictable. This does not mean that the process or output is desirable but instead well under-

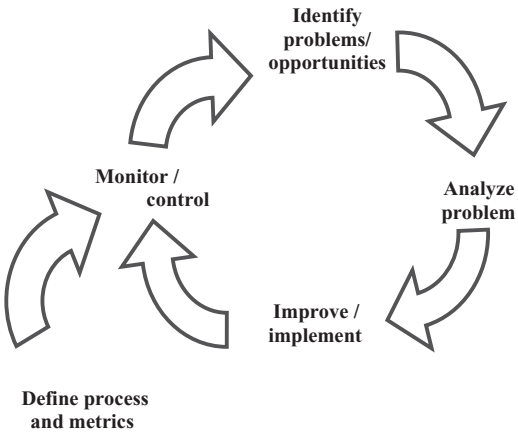


**FIGURE 9** Continuous improvement process.

**TABLE 9 Nine-Block Risk Assessment Matrix**

| | | Severity | | |
|---|---|---|---|---|
| | | Minor | Major | Severe |
| Frequency | Probable | Medium | High | High |
| | Occasional | Low | Medium | High |
| | Remote | Low | Low | Medium |

stood and predictable. It is through metrics, trends, and risk assessments that issues and concerns should be evident. Process owners can use the management review forum to present a proposal for process improvements.

### 8.2 Process Improvement Proposal

The process owner with stakeholders will need to provide a process improvement proposal if the issue or change requires prioritization due to funds or additional resources from the enterprise. The proposal should include, at minimum, the problem and or opportunity statement, impact to the site based on risk, and proposal of an action and/or project, including both cost and resource requirements.

During the development of the proposal, the process owner should consider requesting subject matter experts to assist with the development of the problem statement, risk assessment, and cost avoidance or savings. Many times a process owner's core competencies align closely with the process but may lack business or project management skills. The process owner may need assistance to clearly articulate to the leadership what the benefits are to accept the proposed change versus the risks for not adopting the proposal.

Risk assessment tools such as a nine-block risk assessment (Table 9) or a failure mode and effect analysis (FMEA) are available to assist the process owner with the evaluation of the process or issue to better understand and communicate the

probability of failure and the severity of a process issue. These analyses assist with the prioritization of issues and identification of specific actions required to mitigate risks or the identification of contingency plans for issues that are not mitigated. In a pharmaceutical and biopharmaceutical environment it is important that all risk tools are completed assessing impact to product safety and efficacy as well as the business drivers. Any steps or issues in the process that can negatively impact the safety and efficacy of the product require immediate elevation to leadership and must be addressed immediately. If the proposal requires prioritization, the process owners should clearly identify potential cost savings or avoidance through a cost of quality model to further engage senior leadership. The combination of risk and costs is an effective way to gain leadership support and attention.

Leadership's role in the process improvement proposal is to understand the issue or opportunity, understand associated risk(s), and approve or redirect a proposed action or project and provide appropriate funding and or resources. As action items and proposals are approved and initiated, the progress should be monitored on a routine basis to ensure appropriate progress is made.

## 8.3 Task versus Project

Process improvements may be conducted by the completion of a task or a project. A task is an activity that can be completed by the process owner with minimal cost and/or resources over a short period of time. A project is defined as temporary work to provide a product or service that is beyond the process owner's support. In general, a project requires more than one full-time equivalent (FTE), crosses over multiple functional organizations, and the duration of the effort spans over a longer period of time. Improvement status, updates, and issues should be discussed on a regular basis by a management forum or steering committee. Tasks and projects should be prioritized based on the risk against patient safety and efficacy and compliance.

If the process improvement meets the requirements of a project, a project manager should be identified. Formal project management allows for a holistic and integrated approach to the change. The project manager should not replace the process owner but ensure that the issues are identified, prioritized, and resources are applied, milestones are met, issues escalated and resolved, and progress reported. The process owner needs to be the project lead with the stakeholders or steering committee, providing support and guidance. This allows the process owner to focus on the issues and improvements (their core competencies) and allows the project manager to move the project forward in a methodical manner. During the project it is critical that success is defined and measured.

## 8.4 Project Metrics

Project metrics should be identified to measure the actual benefit of the change versus the expected result following the implementation. Many times, corporations implement a change and move on to the next project without fully understanding whether or not the changes achieved the desired result. A project that does not achieve the expected benefits can lead to an ineffective process, conflicts with associated touch points with other processes, or frustration from staff and customers.

Applying a systems-based approach to continuous improvement of the QMS, utilizing formal risk management tools benefits the overall efficiency of the organization. Process owners are accountable and empowered to drive continuous improvements. Metrics are utilized to identify trends, issues, and opportunities. Stakeholders are engaged throughout the process, and management is involved in the prioritization and staffing of the task or project. The processes are continually managed and evaluated. Continuous improvements based on risk allow the organization to apply resources and money to the most critical projects that will make the most impact. As process improvements are implemented, staff will benefit from a predictable, lean process allowing them to focus on the proactive nature of their work as opposed to the high stress of reacting to the issue of the day. The process owner will gain credibility as he or she demonstrate the ability to ensure that the right people are making the right decisions in a timely manner and that process improvements are addressing systemic process problems and not superficially addressing issues that will resurface again.

## 9 ENSURING ONGOING SUCCESS

Building infrastructure to establish and maintain a quality management system requires resources and resolve from leadership and staff. The current pharmaceutical and biopharmaceutical global and regulatory environment requires an organization invested in developing and maintaining a robust system and processes meeting the organization's requirements for producing quality product. Future competition, shorter time to market, efficient development, and first-pass approval expectations exacerbate the need for robust processes. The global marketplace continues its pressure on industry to deliver lifesaving and life-style changing medicines faster and cheaper.

### 9.1 Establishing Mutual Goals

Companies that have designed, developed, and established QMSs and processes that are simple to execute, easy to understand, and deliver the business and regulatory results will have competitive advantage over their industry peers. They will be faster and more efficient at adapting new technologies, assimilating new organizations through merger and acquisitions, able to apply adequate resources to appropriate business needs, and most importantly quickly modify and adapt to changing marketplace demands. Dependence on people, fragmented procedures, or tribal knowledge, rather than integrated, functional processes, will bring undesirable results to all levels of the organization.

Ensuring ongoing success requires establishing mutual goals for the organization from the beginning. These goals must satisfy the needs of the business, the employees, and the shareholders. Well-designed processes with accountable ownership that have been established through discussion, design, and support of leadership, functional management, operational stakeholders, and general staff provide the foundation for common shared needs (Figure 10). If anyone of these groups is not considered, nominal support and eventually failure of the program can be expected.

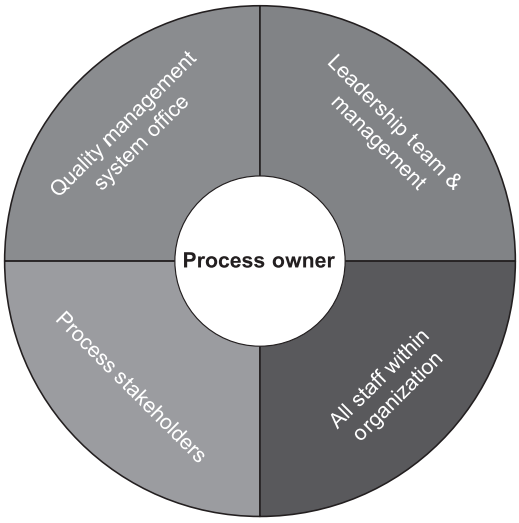


**FIGURE 10** Process owner support model.

These shared goals need to be memorialized through documentation of the program. This includes outlining the long-term objectives of the program, benefits required to be achieved for stakeholders, and annual quality plans for achieving milestone goals and success. Program success comes from leadership support, robust system design, adequate training for employees, and meaningful metrics to measure performance and continuous improvement efforts.

Mechanisms to determine stakeholder feedback on program acceptance, clarity, and improvement opportunities need incorporation into the ongoing maintenance of the QMS. Focus groups are one method of obtaining this type of information. Another opportunity exists with regulatory inspections and customer audits. Taking appropriate action to implement program changes and enhancements while recognizing contributors will ensure stakeholder support and participation.

Mutual goals will drive success of the program and provide the reference for why the system approach is needed and the benefits it can bring. There is no better situation than having an entire organization aligned around the business design, and executing against it, while supporting each other.

### 9.2 Rewards and Recognition

Process owners' responsibilities are significant. Owners need to be selected from predetermined criteria that are discriminatory in nature. Process owners are the drivers of the operations and therefore need to be recognized for their special efforts and responsibilities. This recognition can take many forms. A significant distinction in base qualifications and rewards is a valuable incentive to becoming a process owner. Ongoing development for process owners is another incentive and reward for the process owner. In addition to the financial and tangible rewards, being recognized by the organization to have the confidence of management is also another form of reward and recognition. Inevitably, processes contain waste and

inefficiencies, thereby providing another opportunity for owners to improve their process and be recognized for that improvement.

Public recognition of system program and process owner accomplishments is essential. This can easily be accomplished through regular review sessions, at metric review meetings, through staff meetings and updates, poster sessions, newsletters, and departmental meetings. Simple recognition and small gifts are appreciated and reinforce management's support and commitment to the program. Process owners and stakeholders are the most influential group to spread the word on the usefulness of the program and must be cultured to ensure ongoing success.

Studies indicate financial rewards alone cannot provide employee satisfaction and retention. High employee turnover costs companies tremendous financial and competitive resources. Many employees faced with equal or higher pay but unsatisfying work will move onto another company or position. A poorly integrated QMS with complicated processes is often the foundation for that dissatisfaction. To repeat work, lose valuable time, or deliver substandard product does not satisfy today's highly educated and competitive worker in the pharmaceutical and biopharmaceutical industry. The cost to recruit, replace, relocate, and retrain employees is significant. Avoidance of these costs can be used as a partial basis for support of the program.

### 9.3 Ensuring Ongoing Program Continuity

Accomplishments of a comprehensive QMS program should be shared between locations and be consistent. Common, competent leadership for the enterprise will ensure consistency. A consistent QMS program also allows for transfer of staff between sites with little or no training and assimilation requirements. Divergent evolution will dilute the QMS effort and support. Flexibility to execute is important, however, caution must be exercised to restrict diverging language, interpretation, and philosophy. Within a short time of a global execution, efficiencies will be quickly realized. Ensuring consistency also increases the number of process users with similar experiences and leverages focus for process improvements and therefore support.

Regulators and customers require assurance in consistency of pharmaceutical and biopharmaceutical manufacturing operations. Today's manufacturing supply chains require multiple sites in varying locations to produce a product. Quality systems must be perceived as an integral part of the value chain. This requires that all sites be compliant in their operations and systems. Strong areas in one location do not make up for weak or absent systems in another location. Fines are levied and business is made or lost based on the individual site or weakest link in the supply chain. Management must have a mechanism to measure its processes, and a comprehensive QMS is the mechanism to demonstrate capability.

### 9.4 Program Institutionalization

Program institutionalization is realized with time. All levels of the organization need to recognize and verbalize that the quality management system approach is the way business is conducted. This way of doing business will become part of the culture to the point at which it is second nature to leadership, management,

and staff. Regulators and customers will recognize the benefits, as do the shareholders and patients.

## REFERENCES


1. *Webster's New Collegiate Dictionary*, ninth edition, 1986, p.1199.
2. Arling, E. R. (2004), Integrating QSIT into quality plans, *Biopharm. Int.*, June, 44–46, 48, 50–52.
3. *Drug Industry Daily*, Oct. 12, 2006, Vol. 5, No. 200, Washington Business Information.
4. 2006 PDA/FDA joint regulatory conference, Sept. 11, 2006, Washington, DC.
5. Joneckis, C. (2006), Ph.D. presentation at 11th Annual GMP by the Sea, Aug. 28–30, Cambridge, MD.
6. Quality systems regulations CDRH, available: www.fda.gov.
7. CPGM 7356.002, CDER, available: www.fda.gov.
8. IOM biological inspections 7345.848 CBER, available: www.fda.gov.
9. Rockefeller, J. D., III (1973), *The Second Revolution*, Harper-Row, New York.
10. Kotter, J. P. (2001), Leadership insights, *Harvard Bus. Re.*, p. 29.
11. George, M. (2004), *Lean Six Sigma Pocket Tool Book*, McGraw-Hill Professional, New York, p. 4.

# QUALITY PROCESS IMPROVEMENT

Jyh-hone Wang
*University of Rhode Island, Kingston, Rhode Island*

**Contents**

1 Diagnosing a Process
  1.1 Introduction
  1.2 Basic Tools for Diagnosing a Process
2 Stabilizing and Improving a Process
  2.1 Introduction
  2.2 Control Charts for Attributes
  2.3 Control Charts for Variables
  2.4 Special Control Charts
3 Improving Performance of a Process
  3.1 Introduction
  3.2 Process Capability and Improvement Studies
  Bibliography

## 1 DIAGNOSING A PROCESS

### 1.1 Introduction

Quality process improvement starts with a diagnostic journey where problems are identified. Remedial activity will be taken and the process will be continuously monitored afterward. The common activities taken in the diagnostic journey are analyzing symptoms, formulating hypotheses, testing hypotheses, and identifying causes. Table 1 describes basic tools for the diagnostic journey. A description of them is given in Section 1.2.

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

**TABLE 1 Basic Quality Process Improvement Tools during Process Diagnosis**

| | Basic Tools for Quality Process Improvement | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Common Activities to Diagnose Cause | Cause–Effect Diagram | Pareto Chart | Histogram | Scatter Diagram | Normal Probability Plot | Flow Diagrams | Data Collection | Box Plot | Stratification |
| Analyzing symptoms | • | • | • | • | • | • | ○ | • | ○ |
| Formulating hypotheses | • | ○ | • | ○ | • | ○ | ○ | • | ○ |
| Testing hypotheses | • | • | ○ | • | • | ○ | • | ○ | • |
| Identifying cause(s) | • | • | • | • | ○ | ○ | ○ | ○ | • |

*Note*: (•) major; (○) minor.

## 1.2 Basic Tools for Diagnosing a Process

***Cause-and-Effect Diagram*** A cause-and-effect diagram relates potential causes of a problem to their effects. This is a tool that could be very useful in diagnosing a process. It focuses on the possible causes of a specific problem in a structured and systematic way. The following steps are suggested for constructing a cause-and-effect diagram:

1. Define the problem (effect).
2. Write problem on the right side and draw an arrow from the left to the right side.
3. Brainstorm the main categories of causes of problems and draw major branch arrows to the main arrow.
4. For each major branch, detailed causal factors (subcauses) are drawn as subbranches.
5. Write sub-subcauses branching off the subcauses.
6. Ensure all the items that may be causing the problem are indicated in the diagram.

Figure 1 shows a cause-and-effect diagram which is used to identify causes to yield a problem in a biopharmaceutical manufacturing process. Possible main causes and subcauses are identified. Once the causes are identified, other tools are employed to determine the contribution of various causes to the effect. Actions are taken to eliminate or minimize the impact of these causes.

***Pareto Chart*** The Pareto principle suggests a problem (effect) can be attributed to relatively few causes. In quantitative terms, 80% of the problems come from 20% of the causes (machines, raw materials, operators, etc.); therefore effort aimed at the right 20% can solve 80% of the problems. A Pareto chart includes three basic elements: (1) the causes to the total effect, ranked by the magnitude of the contribution; (2) the frequency of each cause; and (3) the cumulative-percent-of-total effect of

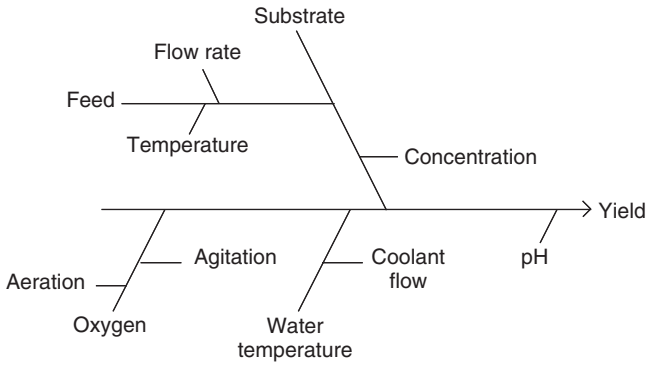


**FIGURE 1** Cause-and-effect diagram.

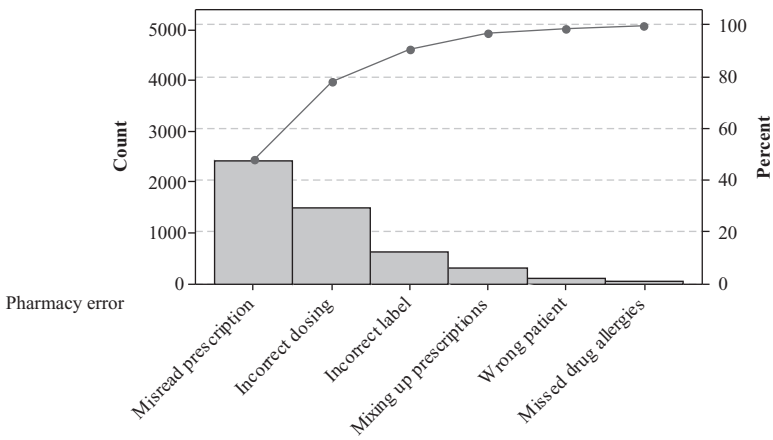


| | | | | | | |
|---|---|---|---|---|---|---|
| Count | 2452 | 1520 | 632 | 320 | 108 | 62 |
| Percent | 48.1 | 29.8 | 12.4 | 6.3 | 2.1 | 1.2 |
| Cum % | 48.1 | 78.0 | 90.4 | 96.7 | 98.8 | 100.0 |

**FIGURE 2** A Pareto chart showing pharmacy errors.

the ranked causes. Figure 2 gives an example of a Pareto chart which exhibits errors found in a pharmacy store chain in one month.

***Histogram*** A histogram is a graphic summary of variation in a set of data. Data are clustered into categories and the values of individual clusters are plotted to give a series of bars. For illustration, Table 2 presents 40 observations on the shelf life of a certain drug and their frequency distribution. Figure 3 gives a histogram for the drug shelf life data.

***Scatter Diagram*** A scatter diagram is a basic tool to identify the potential relationship between two variables. Scatter diagrams are similar to line graphs in that they use horizontal and vertical axes to plot data points. However, they have a very specific purpose. Scatter diagrams show how much one variable is affected by another. The relationship between two variables is called their correlation. The

**TABLE 2 Drug Shelf Life (days)**

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 102.2 | 104.1 | 103.5 | 104.5 | 103.2 | 103.7 | 103.0 | 102.6 |
| 103.4 | 101.6 | 103.1 | 103.3 | 103.8 | 103.1 | 104.7 | 103.7 |
| 102.5 | 104.3 | 103.4 | 103.6 | 102.9 | 103.3 | 103.9 | 103.1 |
| 103.3 | 103.1 | 103.7 | 104.4 | 103.2 | 104.1 | 101.9 | 103.4 |
| 104.7 | 103.8 | 103.2 | 102.6 | 103.9 | 103.0 | 104.2 | 103.5 |

| Range | Midpoint | Frequency | Cumutative % |
|---|---|---|---|
| $101.5 \le x < 102.0$ | 101.75 | 2 | 5.00 |
| $102.0 \le x < 102.5$ | 102.25 | 2 | 10.00 |
| $102.5 \le x < 103.0$ | 102.75 | 5 | 22.50 |
| $103.0 \le x < 103.5$ | 103.25 | 15 | 60.00 |
| $103.5 \le x < 104.0$ | 103.75 | 8 | 80.00 |
| $104.0 \le x < 104.5$ | 104.25 | 6 | 95.00 |
| $104.5 \le x < 105.0$ | 104.75 | 2 | 100.00 |

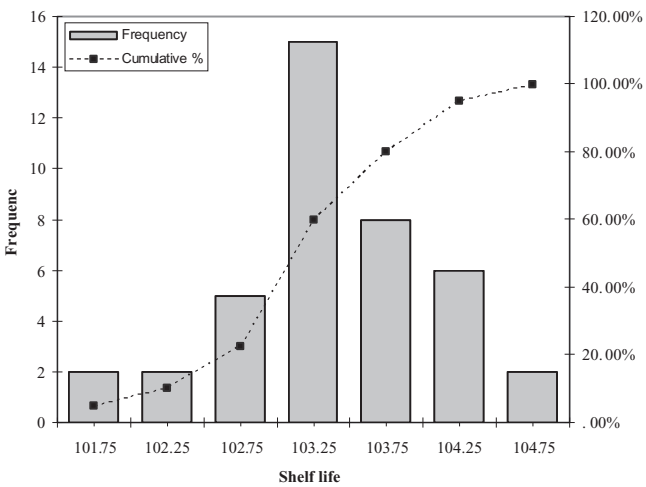


**FIGURE 3** Histogram of drug shelf life.

closer the data points come when plotted to making a straight line, the higher the correlation between the two variables. If the data points make a straight line going from the origin out to high $x$ and $y$ values, then the variables are said to have a positive correlation. If the line goes from a high value on the $y$ axis to a high value on the $x$ axis, the variables have a negative correlation. Figure 4 gives a few examples of scatter diagrams.

***Normal Probability Plot*** The normal probability plot is a graphical technique for assessing whether or not a data set is approximately normally distributed. The data are plotted against a theoretical normal distribution in such a way that the points form an approximate straight line. Departures from this straight line indicate departures from normality. The normal probability plot is important for quality process improvement since many other tools require the normality assumption. A normal

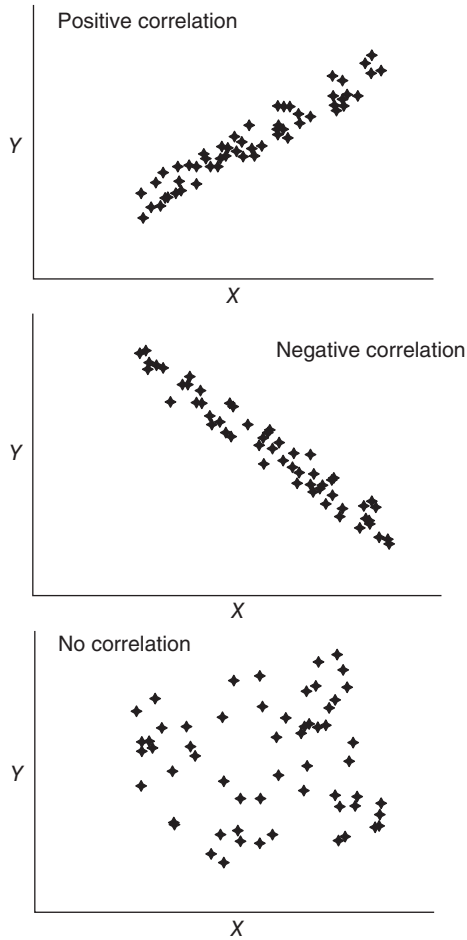


**FIGURE 4** Scatter diagrams.

probability plot of the drug shelf life in Table 2 is given in Figure 5. As seen from the plot, the observations follow a straight line and are contained in the 95% confidence interval. It can thus be said that the shelf life of this drug follows a normal distribution.

### *Other Tools*

*Box Plot* This plot is useful when analyzing the pattern of the data. It displays several important features of data such as central tendency, variability, departure from symmetry, and presence of outliers.

*Flow Diagrams* A process flow diagram can be used to study and understand the process.

*Data Collection* Data are essential for making a proper evaluation of the current process. Tools for data collection include checklists and data sheets.

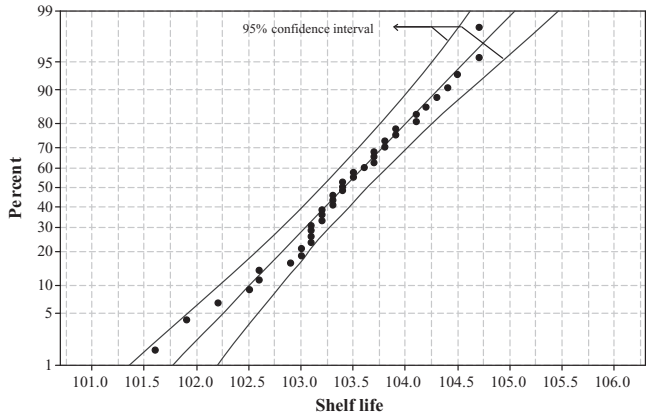


**FIGURE 5** Normal probability plot of drug shelf life with 95% confidence interval.

*Stratification* This technique is used to separate data into groups based on categories or characteristics. It is the basis for the application of other tools or it can be used with other data analysis tools such as scatter diagrams.

## 2 STABILIZING AND IMPROVING A PROCESS

### 2.1 Introduction

***Basic Concepts of a Control Chart*** The control chart is one of the main tools for quality process improvement. It is used to assess the nature of variation in a process and to facilitate the forecasting and management of a process. Values of the quality characteristic are plotted against the sample number or time, as shown in Figure 6. The centerline represents the process average. The upper and lower control limits (UCL and LCL) are usually chosen as three standard deviations (SDs) above and below the centerline so they can be used to detect "out-of-control" situations without causing mamy false alarms. An out-of-control situation is usually signaled by a plotted point falling outside the control limits or a cluster of plotted points forming an abnormal pattern.

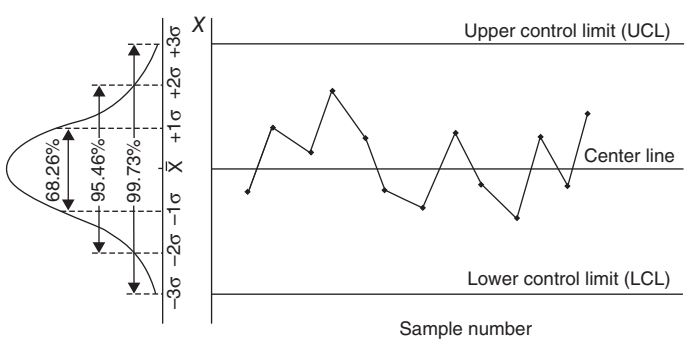


**FIGURE 6** Normal curve-based control chart.

Plotted points on a control chart are usually based on data collected from samples in a process. After a sufficient number of samples are collected and the data are plotted on a control chart, the stability of the process can be evaluated. A stable process is "in control" while an out-of-control process is unstable. Depending on the type of quality characteristic, control charts can be divided into two groups: variable control charts and attribute control charts. Variable control charts are used to monitor quality characteristic that are continuously varying in nature; attribute control charts are used to monitor those quality characteristics that are not numerically measurable. The determination of the centerline and control limits are described in Sections 2.2 and 2.3 with respect to different types of control charts.

***Applications of Control Charts*** Control charts serve to direct management attention toward special causes of variation in a process when they appear. In evaluating control charts, the following symptoms could indicate a process that is out-of-control:

- *Outlier* One or more point(s) that fell outside the control limits.
- *Run* A series of plotted points above or below the centerline.
- *Trend* A continual rise or fall of plotted points.
- *Cyclicity* A pattern that repeats itself over time.

The following steps are usually followed in a control chart's development and application:

- Determine a "base period" for initial control chart development.
- Collect sample data from the base period.
- Calculate the parameters for the control chart, that is, centerline and control limits.
- Plot collected sample points on the chart with the centerline and control limits.
- Determine whether the chart parameters can be used to monitor the process; revise the parameters if necessary.
- Collect ongoing samples and continue monitoring the process using the developed control chart.
- Conduct periodic audits on the parameters of the control chart.

Variable control charts are widely applied in many manufacturing and nonmanufacturing settings. They can be used to monitor, for example, the inside diameter of an aircraft bearing, the moisture content of a drug tablet, the net weight of a pharmaceutical product, the processing time of phone inquiries, and the satisfaction level of customers. The latter two are examples of nonmanufacturing applications.

Attribute control charts are less used compared to variable control charts. When it is not possible or practical to measure the quality characteristic of a product, attribute control charts are often applied. Examples of their application include monitoring the fraction of nonconforming of a certain sensor production, the number of defective diodes in an electronic assembly, the number of imperfections in textile

production, the fraction of defective batches in a biomedical manufacturing production, and the number of errors found in a pharmacy store.

In most applications, the choice between a variable control chart and an attribute control chart is clear-cut. In some cases, the choice will not be obvious. For instance, if the quality characteristic is the softness of an item, such as the case of pillow production, then either an actual measurement or a classification of softness can be used. Quality managers and engineers will have to consider several factors in the choice of a control chart, including cost, effort, sensitivity, and sample size. Variable control charts usually provide more information to analysts but cost more to implement and use. Attribute control charts are less sensitive and expensive but usually requires large samples to reach certain statistical significance.

## 2.2 Control Charts for Attributes

Control charts based on attribute data include the $p$ chart, $np$ chart, $c$ chart, and $u$ chart. The former two are applied when fraction nonconforming or number of nonconforming is a concern, and the latter two are used to deal with the nonconformities. Most pharmaceutical manufacturing industries employ one or more of these charts.

**p *Control Chart*** A $p$ chart can be used to monitor the fraction nonconforming of a process. *Fraction nonconforming* is defined as the ratio of the number of nonconforming items in a population to the total number of items in that population. In pharmaceutical manufacturing, an item will be classified as nonconforming if it fails to conform to standards on one or more attributes, for example, fill volumes of vials, moisture content, hardness, and solubility.

Let us suppose a random sample of $n$ items is selected and examined from a process running with a stable nonconforming rate $p$ and $D$ units of nonconforming items are found; then $D$ is a random variable following a binomial distribution with parameters $n$ and $p$. If the true fraction nonconforming, $p$, is known, then the parameters of the $p$ chart are

$$\begin{aligned} \text{UCL} &= p + 3\sqrt{\frac{p(1-p)}{n}} \\ \text{Centerline} &= p \\ \text{LCL} &= p - 3\sqrt{\frac{p(1-p)}{n}} \end{aligned} \tag{1}$$

In practice, the fraction nonconforming, $p$, is unknown most of the time and is thus needed to be estimated from the sample data. An estimated $p_i$ can be calculated for the $i$th sample collected and an average $\bar{p}$ value can be obtained as an arithmetic average of those individual $p_i$ found from the $m$ samples:

$$\bar{p} = \frac{\sum_{i=1}^{m} D_i}{mn} = \frac{\sum_{i=1}^{m} p_i}{m} \tag{2}$$

The $\bar{p}$ can then be used in place of $p$ in Equation (1) in the application. It should be noted that the $\bar{p}$ value needs to be assessed periodically to assure its representativeness of the average process fraction nonconforming.

**np *Control Chart*** An *np* chart is used to monitor the number of nonconforming items produced in a process. Very similar to the *p* chart, the parameters of an *np* chart are

$$\begin{aligned} \text{UCL} &= np + 3\sqrt{np(1-p)} \\ \text{Centerline} &= np \\ \text{LCL} &= np - 3\sqrt{np(1-p)} \end{aligned} \tag{3}$$

As in the *p* chart, if the actual *p* value is not available, $\bar{p}$ can be used in the calculation.

**c *Control Chart*** A *c* chart can be used to monitor the number of nonconformities (defects) per inspection unit. An inspection unit can be a single unit of product, a batch of multiple products, or a certain measured volume (weight) of product. Many pharmaceutical manufacturing processes are lot based where raw material or semi-product passes from one process to the next. For example, an inappropriately coated tablet in a coating process can be considered as a nonconformity (defect) where an inspection unit might be defined as 1 kg of the tablet.

Suppose an inspection unit of a certain product is selected and examined from a process running with a stable nonconformity rate *c* per inspection unit and *X* nonconformities are found. Then *X* is a random variable following a Poisson distribution with parameter *c*. If the true nonconformity level *c* is known, then the parameters of the *c* chart are

$$\begin{aligned} \text{UCL} &= c + 3\sqrt{c} \\ \text{Centerline} &= c \\ \text{LCL} &= c - 3\sqrt{c} \end{aligned} \tag{4}$$

If the actual nonconformity level *c* is unknown, it can be estimated by using average *c* values obtained from *m* inspection units collected in a base period:

$$\bar{c} = \frac{\sum_{i=1}^{m} C_i}{m} \tag{5}$$

The $\bar{c}$ can then be used in place of *c* in Equation (4) in the application. Since it is possible to obtain a negative LCL using Equation (4), a value of zero should be used in that case.

**u *Control Chart*** A *u* chart is used to monitor the rate of nonconformities. The rate of nonconformities ($u$) is the number of nonconformities ($x$) in an inspection unit divided by the number of physical units ($n$) inspected (e.g., 100 ft of pipe, 100 items in a batch). Similar to the *c* chart, the parameters of a *u* chart are

$$\begin{aligned} \text{UCL} &= u + 3\sqrt{\frac{u}{n}} \\ \text{Centerline} &= u \\ \text{LCL} &= u - 3\sqrt{\frac{u}{n}} \end{aligned} \tag{6}$$

If the actual $u$ value is not available, $\bar{u}$ can be used in Equation (6).

**Example 1** A medical device manufacturer is concerned about the nonconforming (defective) and the nonconformity (defect) produced in its recently set-up production line. Twenty batches of this medical device were randomly selected from the production line. Each batch contained 100 units. Each unit is inspected and is classified as either "conforming" or "nonconforming." During the inspection, the number of nonconformities (defects) was also counted. The data collected are shown in Table 3.

**TABLE 3 Nonconforming and Nonconformity Counts of 20 Batches of Medical Device**

| Batch number | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Nonconformings | 3 | 2 | 4 | 2 | 5 | 2 | 1 | 2 | 0 | 5 | 2 | 4 | 1 | 3 | 6 | 0 | 1 | 2 | 3 | 2 |
| Nonconformities | 9 | 7 | 13 | 8 | 6 | 8 | 10 | 10 | 7 | 10 | 12 | 9 | 11 | 15 | 8 | 12 | 11 | 8 | 7 | 15 |

Based on the data in Table 3, the average fraction nonconforming, $\bar{p}$, can be obtained as 2.5%; the average nonconformity per batch, $\bar{c}$, is 9.8; and the average nonconformity per unit, $\bar{u}$, is 0.098. The resulting control charts are shown in Figures 7–10. These charts indicate that the process is in control and thus the parameters established here can be used to monitor future productions.

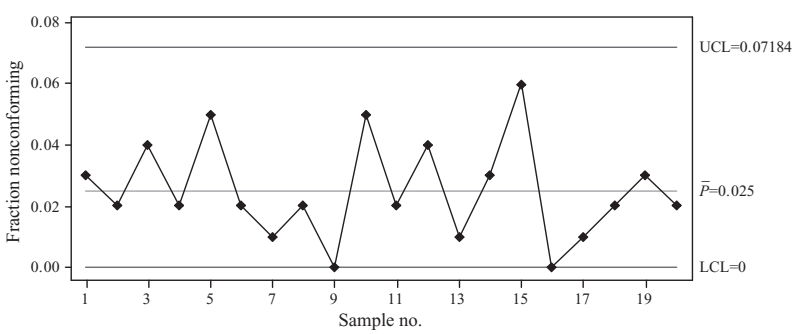


**FIGURE 7** $p$ Chart for medical device manufacturing example.

## 2.3 Control Charts for Variables

Control charts based on variable sample data include the $\bar{x}$ chart and the $s$ chart. When dealing with a numerically measurable quality characteristic, the $\bar{x}$ chart is usually employed to monitor the process average and the $s$ chart is used to monitor the process variability. When there is only one observation in each sample, the individual measurement chart ($I$ chart) and moving range chart (MR chart) are used to monitor the process average and variability. It should be noted that due to the poor

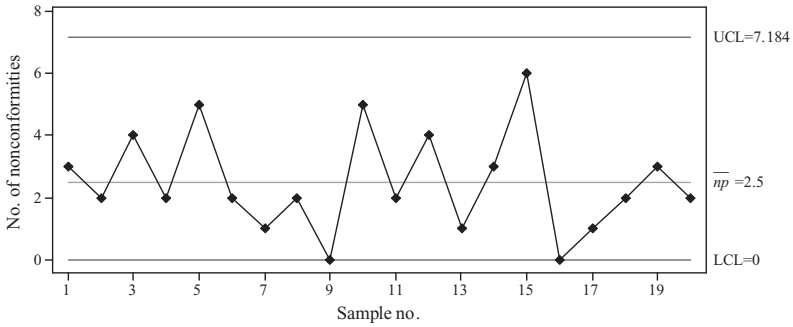


**FIGURE 8** *np* Chart for medical device manufacturing example.

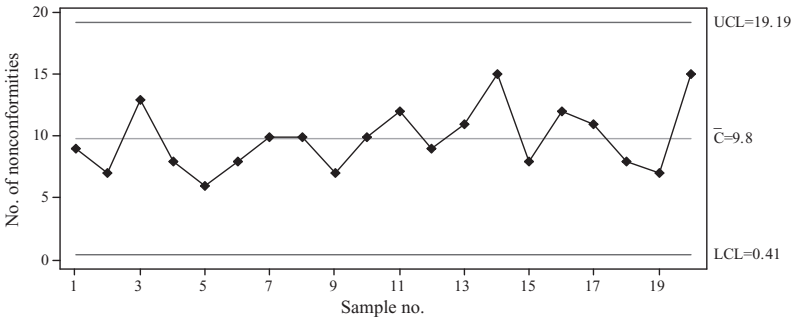


**FIGURE 9** *c* Chart for medical device manufacturing example.

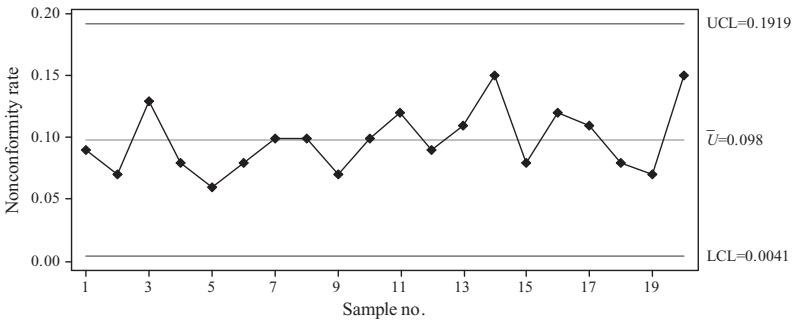


**FIGURE 10** *u* Chart for medical device manufacturing example.

sample size, the *I* and MR charts are less sensitive in detecting if the process is out of control than the $\bar{x}$ and *s* charts.

**$\bar{x}$ *Control Chart*** An $\bar{x}$ chart is used to monitor the process average. It is usually used in pharmaceutical manufacturing where multiple units are collected in each sample (e.g., a sample of multiple tablets formed by dry powders or wet granules.) Due to contamination risk and cost of sampling (including product loss due to

sample volumes and incurred labor cost of laboratory analysis), the sample size is usually kept small.

Sample means $\bar{x}$ are plotted on the $\bar{x}$ chart. Assume that random samples of $n$ items are collected and examined from a stable process with a process mean $\mu$ and standard deviation $\sigma$. Then $\bar{x}$ can be considered as a random variable following a normal distribution with mean $\mu_{\bar{x}}$, and standard deviation $\sigma_{\bar{x}}$, where

$$\mu_{\bar{x}} = \mu \qquad \text{and} \qquad \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \tag{7}$$

If the true process mean $\mu$ and standard deviation $\sigma$ are known, then the parameters of the $\bar{x}$ chart are

$$\begin{aligned} \text{UCL} &= \mu_{\bar{x}} + 3\sigma_{\bar{x}} = \mu + 3\frac{\sigma}{\sqrt{n}} \\ \text{Centerline} &= \mu_{\bar{x}} = \mu \\ \text{LCL} &= \mu_{\bar{x}} - 3\sigma_{\bar{x}} = \mu - 3\frac{\sigma}{\sqrt{n}} \end{aligned} \tag{8}$$

Since $\mu$ and $\sigma$ are not usually known, estimators of them can be obtained from the sample means ($\bar{x}$) and sample standard deviations ($s$) of the $m$ samples collected in the base period:

$$\begin{aligned} \text{Estimator of } \mu &= \bar{\bar{x}} = \frac{\sum_{i=1}^{m} \bar{x}_i}{m} \\ \text{Estimator of } \sigma &= \frac{\sum_{i=1}^{n} s_i / m}{4(n-1)/4n-3} = \frac{\bar{s}}{c_4} \end{aligned} \tag{9}$$

Using the estimators, the parameters of the $\bar{x}$ chart are now

$$\begin{aligned} \text{UCL} &= \bar{\bar{x}} + 3\frac{\bar{s}/c_4}{\sqrt{n}} = \bar{\bar{x}} + A_3\bar{s} \\ \text{Centerline} &= \bar{\bar{x}} \\ \text{LCL} &= \bar{\bar{x}} - 3\frac{\bar{s}/c_4}{\sqrt{n}} = \bar{\bar{x}} - A_3\bar{s} \end{aligned} \tag{10}$$

The common values of constants $c_4$ and $A_3$ are tabulated in Table 4 for sample sizes from 2 to 10. Like other control charts, the values of $\bar{\bar{x}}$ and $\bar{s}$ should be periodically verified to assure that they can be used to derive good estimators for the process average and process standard deviation.

**s *Control Chart*** An $s$ chart is used to monitor the process variability. Since it is equally important to ascertain that the variability and the mean of a process are in control, an $s$ chart is usually used in conjunction with the $\bar{x}$ chart. Sample standard deviations are plotted on the $s$ chart. Consider $s$ as a random variable with mean $\mu_s$ and standard deviation $\sigma_s$. Then the parameters of the $s$ chart can be stated as

**TABLE 4 Values of Constants in Variable Control Chart Parameters**

| $n$ | $A_3$ | $c_4$ | $B_3$ | $B_4$ | $d_2$ | $d_3$ |
|---|---|---|---|---|---|---|
| 2 | 2.659 | 0.798 | 0 | 3.267 | 1.128 | 0.853 |
| 3 | 1.954 | 0.886 | 0 | 2.568 | 1.693 | 0.888 |
| 4 | 1.628 | 0.921 | 0 | 2.266 | 2.059 | 0.880 |
| 5 | 1.427 | 0.940 | 0 | 2.089 | 2.326 | 0.864 |
| 6 | 1.287 | 0.952 | 0.030 | 1.970 | 2.534 | 0.848 |
| 7 | 1.182 | 0.959 | 0.118 | 1.882 | 2.704 | 0.833 |
| 8 | 1.099 | 0.965 | 0.185 | 1.815 | 2.847 | 0.820 |
| 9 | 1.032 | 0.969 | 0.239 | 1.761 | 2.970 | 0.808 |
| 10 | 0.975 | 0.973 | 0.284 | 1.716 | 3.078 | 0.797 |

$$\begin{aligned} \text{UCL} &= \mu_s + 3\sigma_s \\ \text{Centerline} &= \mu_s \\ \text{LCL} &= \mu_s - 3\sigma_s \end{aligned} \tag{11}$$

In practice, the parameters of the $s$ chart can be estimated using $\bar{s}$ as

$$\begin{aligned} \text{UCL} &= \bar{s} + 3\frac{\bar{s}}{c_4}\sqrt{1-c_4^2} = B_4\bar{s} \\ \text{Centerline} &= \bar{s} \\ \text{LCL} &= \bar{s} - 3\frac{\bar{s}}{c_4}\sqrt{1-c_4^2} = B_3\bar{s} \end{aligned} \tag{12}$$

If the LCL calculation results in a negative value, use zero as the LCL.

**Example 2** In a Pet Tabs (pet vitamin tablets) production, the pharmaceutical manufacturer is using milling and micronizing machines to pulverize raw materials into fine particles. These finished particles are combined and processed further in mixing machines. The mixed ingredients are then pressed into tablets, dried, and sealed in packages. A normally distributed quality characteristic, moisture content, is monitored. Samples of $n = 4$ tablets are taken from the manufacturing process every hour. The data after 25 samples have been collected are shown in Table 5.

From these data, it is found that $\bar{\bar{x}} = 10.254$ and $\bar{s} = 0.926$. Using Equations (10) and (12), the parameters of the $\bar{x}$ and $s$ charts are found as:

| | $\bar{x}$ Chart | $s$ Chart |
|---|---|---|
| UCL | 11.761 | 2.098 |
| Centerline | 10.254 | 0.926 |
| LCL | 8.747 | 0 |

The control charts are shown in Figure 11. The $\bar{x}$ and $s$ charts show that the process is in control and thus the parameters established here can be used to monitor future productions.

**TABLE 5 Moisture Content (%) of 25 Samples of Pet Tabs**

| | Observations | | | | | |
|---|---|---|---|---|---|---|
| Sample Number | 1 | 2 | 3 | 4 | $\bar{x}$ | s |
| 1 | 7.84 | 11.01 | 10.14 | 9.41 | 9.600 | 1.343 |
| 2 | 10.51 | 9.1 | 9.52 | 10.83 | 9.990 | 0.814 |
| 3 | 9.74 | 10.39 | 9.62 | 11.16 | 10.228 | 0.708 |
| 4 | 10.71 | 11.41 | 10.71 | 8.63 | 10.365 | 1.203 |
| 5 | 9.93 | 10.95 | 8.99 | 10.73 | 10.150 | 0.889 |
| 6 | 9.94 | 10.27 | 9.35 | 9.42 | 9.745 | 0.438 |
| 7 | 12.11 | 9.72 | 8.89 | 9.75 | 10.118 | 1.387 |
| 8 | 9.61 | 8.93 | 11.12 | 8.75 | 9.603 | 1.077 |
| 9 | 9.17 | 10.87 | 9.97 | 10.79 | 10.200 | 0.798 |
| 10 | 11.41 | 10.39 | 8.83 | 12.19 | 10.705 | 1.451 |
| 11 | 8.43 | 9.48 | 10.56 | 10.2 | 9.668 | 0.939 |
| 12 | 9.92 | 10.13 | 9.66 | 8.21 | 9.480 | 0.868 |
| 13 | 8.39 | 9.94 | 10.4 | 8.69 | 9.355 | 0.967 |
| 14 | 10.42 | 10.27 | 10.94 | 10.91 | 10.635 | 0.341 |
| 15 | 10.98 | 12.57 | 11.14 | 8.97 | 10.915 | 1.481 |
| 16 | 9.73 | 10.05 | 12.82 | 12.43 | 11.258 | 1.592 |
| 17 | 11.36 | 8.91 | 10.08 | 10.55 | 10.225 | 1.024 |
| 18 | 9.42 | 11.12 | 9.01 | 10.52 | 10.018 | 0.973 |
| 19 | 10.15 | 10.08 | 10.12 | 9.88 | 10.058 | 0.122 |
| 20 | 11.73 | 11.1 | 10.75 | 9.94 | 10.880 | 0.746 |
| 21 | 11.52 | 9.11 | 9.88 | 11 | 10.378 | 1.087 |
| 22 | 11.29 | 10.43 | 11.6 | 11.74 | 11.265 | 0.588 |
| 23 | 9.39 | 12.96 | 11.42 | 10.28 | 11.013 | 1.541 |
| 24 | 10.26 | 9.59 | 9.33 | 9.26 | 9.610 | 0.456 |
| 25 | 11.25 | 10.65 | 11.06 | 10.63 | 10.898 | 0.307 |

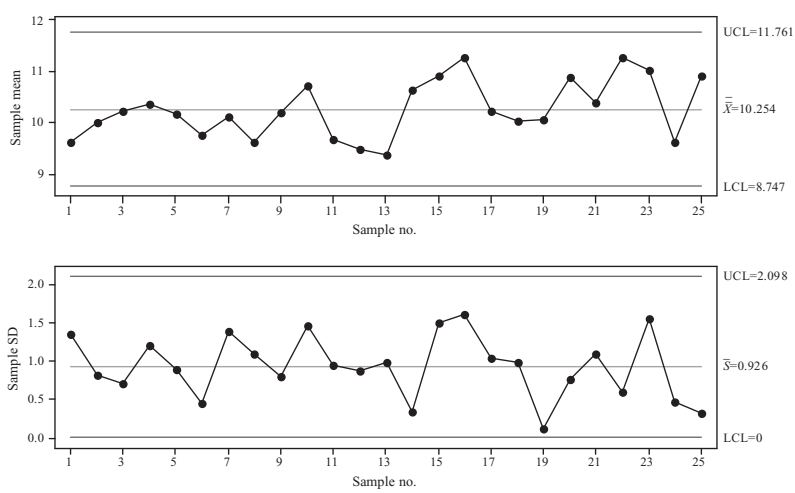


**FIGURE 11** $\bar{x}$ and $s$ charts for Pet Tab manufacturing example.

***Individuals Control Charts*** In some chemical and biopharmaceutical manufacturing processes involving lengthy and expensive procedures, it is not feasible to form a sample of size greater than one because only one product or one batch is available each time. When the sample size used for statistical process monitoring is limited to one, individual control charts, *I* and MR charts, are needed.

The *I* chart is serving the same function as the $\bar{x}$ chart except that now $x$ is the value of the individual measurement. Assuming that $x$ follows a normal distribution with mean $\mu$ and standard deviation $\sigma$, the theoretical parameters of the *I* chart are

$$\begin{aligned} \text{UCL} &= \mu + 3\sigma \\ \text{Centerline} &= \mu \\ \text{LCL} &= \mu - 3\sigma \end{aligned} \tag{13}$$

The process average $\mu$ can be estimated by $\bar{x}$, which is

$$\hat{\mu} = \bar{x} - \frac{\sum_{i=1}^{m} x_i}{m} \tag{14}$$

Since only individual measurements are available, moving ranges need to be calculated for the estimation of process standard deviation $\sigma$. A $k$-point moving range, $\text{MR}_k$, can be calculated as

$$\text{MR}_k = \max(x_i, \ldots, x_{i+k}) - \min(x_i, \ldots, x_{i+k}) \tag{15}$$

For $m$ individual measurements, there are $m - k$ $\text{MR}_k$ available, and the process standard deviation $\sigma$ can be estimated as

$$\hat{\sigma} = \frac{\overline{\text{MR}}_k}{d_2} = \frac{\sum_{i=1}^{m-k} \text{MR}_{ki} / m - k}{d_2} \tag{16}$$

The estimated process mean and standard deviation can be used to calculate the practical parameters for the *I* chart in Equation (13). The constant $d_2$ value is determined by $k$ and can be found by using $k$ as $n$ in Table 3. Common $k$ values can range from 2 to 5.

The MR chart is used to monitor process variability. Considering $\text{MR}_k$ as a random variable with a mean of $\mu_{\text{MR}_k}$ and a standard deviation of $\sigma_{\text{MR}_k}$, the theoretical parameters of the MR chart can then be stated as

$$\begin{aligned} \text{UCL} &= \mu_{\text{MR}_k} + 3\sigma_{\text{MR}_k} \\ \text{Centerline} &= \mu_{\text{MR}_k} \\ \text{LCL} &= \mu_{\text{MR}_k} - 3\sigma_{\text{MR}_k} \end{aligned} \tag{17}$$

Since $\mu_{\text{MR}_k}$ and $\sigma_{\text{MR}_k}$ are not usually available, they can be estimated as

$$\hat{\mu}_{\text{MR}_k} = \overline{\text{MR}}_k \qquad \text{and} \qquad \hat{\sigma}_{\text{MR}_k} = \frac{d_3}{d_2}\overline{\text{MR}}_k \tag{18}$$

The constant value of $d_3$ can also be found in Table 3. If a negative LCL was obtained, use zero.

## 2.4 Special Control Charts

The control charts discussed earlier are very useful in the diagnostic aspects of quality process improvement. They can be used to stabilize a process by identifying out-of-control situations. After the process is stabilized and brought in control, further improvement of the process can be achieved by using some special control charts such as the cumulative sum (CUSUM) control chart and the exponentially weighted moving average (EWMA) control chart. These control charts can be used when "small shifts" in a process are of interest.

***CUSUM Control Chart*** A CUSUM chart provides an efficient way of detecting small shifts in the mean of a process (<1/2 σ). For larger shifts (>1/2 σ), the $\bar{x}$ chart is usually used. The CUSUM chart incorporates information contained in a sequence of sample points. It keeps track of the cumulative sum of the deviations between each sample point (a sample mean) and a target value. Unlike the $\bar{x}$ chart, which often bases its out-of-control decision on just the most recently collected sample, the CUSUM calculated for a sample point carries the "history" prior to that sample. For example, a sequence of sample points above the centerline can trigger an out-of-control signal although all of them stayed well below the UCLs of the $\bar{x}$ chart.

There are two forms of the CUSUM chart, the tabular form and the V-mask form. Due to its practicality, the tabular form is more preferred in industrial settings. The tabular CUSUM accumulates deviations from a target value (or a known process mean $\mu_0$). Deviations above that target value are cumulated as a one-sided upper CUSUM ($C^+$) and deviations below the target value are cumulated as one-sided lower CUSUM ($C^-$):

$$\begin{aligned} C_i^+ &= \max[0, \bar{x}_i - (\mu_0 + k\sigma_{\bar{x}}) + C_{i-1}^+] \\ C_i^- &= \max[0, (\mu_0 - k\sigma_{\bar{x}}) - \bar{x}_i + C_{i-1}^-] \end{aligned} \tag{19}$$

where $C_0^+ = C_0^- = 0$.

The parameter $k$ is called the allowance and is usually determined as the magnitude of the shift to be detected in terms of $\sigma_{\bar{x}}$. If either $C_i^+$ or $C_i^-$ exceeds a decision interval $h$, the process is considered out of control. In other words, the value of $h$ is considered a UCL and $-h$ is considered an LCL. Its centerline is always at zero. A reasonable value for $h$ is five times the process standard deviation σ.

***EWMA Control Chart*** An EWMA control chart plots weighted moving average values for variables data. A weighting factor is chosen by the user to determine how older data points affect the mean value compared to more recent ones. Because the EWMA chart uses information from all samples, it is a good alternative to the CUSUM chart in detecting smaller process shifts.

The EWMA for sample $i$ ($z_i$) is plotted on the chart and is defined as $z_i = \lambda\bar{x}_i + (1 - \lambda)z_{i-1}$, where $z_0 = \mu_0$. The constant λ defines the weight assigned to the current

sample ($0 < \lambda \leq 1$) and $1 - \lambda$ is the weight assigned to earlier samples. Parameters of the EWMA are

$$\begin{aligned} \text{UCL} &= \mu_0 + L\sigma_{\bar{x}}\sqrt{\frac{\lambda}{2-\lambda}[1-(1-\lambda)^{2i}]} \\ \text{Centerline} &= \mu_0 \\ \text{LCL} &= \mu_0 - L\sigma_{\bar{x}}\sqrt{\frac{\lambda}{2-\lambda}[1-(1-\lambda)^{2i}]} \end{aligned} \tag{20}$$

where $L$ is a design parameter that defines the width of the control limits. The choice of $L = 3$ and $0.05 < \lambda \leq 0.25$ is reasonable. The control limits will become wider when the sample number $i$ is getting larger and finally reach constant values as

$$\begin{aligned} \text{UCL} &= \mu_0 + L\sigma_{\bar{x}}\sqrt{\frac{\lambda}{2-\lambda}} \\ \text{Centerline} &= \mu_0 \\ \text{LCL} &= \mu_0 - L\sigma_{\bar{x}}\sqrt{\frac{\lambda}{2-\lambda}} \end{aligned} \tag{21}$$

**Example 3** The data in Example 2 are now analyzed by CUSUM and EWMA charts. Table 6 shows calculated CUSUM and EWMA values. The value of $h$ in CUSUM is chosen as 5 times the standard deviation of $\bar{x}$ ($\hat{\sigma}_{\bar{x}} = 0.5027$) and the value of $k$ is chosen as 0.5. The $C_i^+$ and $C_i^-$ are calculated using a target value $\mu_0 = 10$. The CUSUM chart is shown in Figure 12. The value of $\lambda$ in EWMA is chosen as 0.2 and $L$ is chosen as 3. The UCL and LCL for individual samples are shown in Table 6 and the EWMA chart is shown in Figure 13. Although the $\bar{x}$ and $s$ charts in Figure 6 indicate that the process is in control, both CUSUM and EWMA gave out-of-control signals at sample point 22. A small process shift has occurred after sample 21.

## 3 IMPROVING PERFORMANCE OF A PROCESS

### 3.1 Introduction

***Basic Concepts*** After a process is diagnosed, corrected, and brought into statistical control, the next question is "How can the performance of a process be improved?" To answer this question, quality managers and engineers need first measure the present process performance. This measurement can be achieved through a process capability study which gauges the ability of a process to produce products according to the specifications. A process can achieve a state of statistical control but still exhibit a poor capability due to the variability in the process. It will be necessary to reduce variability to improve the process capability. Designed experiments based on statistical principles can offer helps toward reduction of variability and optimization of the process. Employing designed experiments, intentional changes can be made in various places in the process; results gathered from these experiments can lead to further process improvement and bring it to the next level. This section presents commonly

**TABLE 6 CUSUM and EWMA Values for Pet Tabs Example**

| | | CUSUM | | EWMA | | |
|---|---|---|---|---|---|---|
| Sample Number | $\bar{x}$ | $C_i^+$ | $C_i^-$ | $z_i$ | UCL | LCL |
| 1 | 9.600 | 0.000 | 0.149 | 9.920 | 10.302 | 9.698 |
| 2 | 9.990 | 0.000 | 0.000 | 9.934 | 10.386 | 9.614 |
| 3 | 10.228 | 0.000 | 0.000 | 9.993 | 10.432 | 9.568 |
| 4 | 10.365 | 0.114 | 0.000 | 10.067 | 10.458 | 9.542 |
| 5 | 10.150 | 0.012 | 0.000 | 10.084 | 10.475 | 9.525 |
| 6 | 9.745 | 0.000 | 0.004 | 10.016 | 10.485 | 9.515 |
| 7 | 10.118 | 0.000 | 0.000 | 10.036 | 10.491 | 9.509 |
| 8 | 9.603 | 0.000 | 0.146 | 9.950 | 10.495 | 9.505 |
| 9 | 10.200 | 0.000 | 0.000 | 10.000 | 10.498 | 9.502 |
| 10 | 10.705 | 0.454 | 0.000 | 10.141 | 10.500 | 9.500 |
| 11 | 9.668 | 0.000 | 0.081 | 10.046 | 10.501 | 9.499 |
| 12 | 9.480 | 0.000 | 0.350 | 9.933 | 10.501 | 9.499 |
| 13 | 9.355 | 0.000 | 0.744 | 9.817 | 10.502 | 9.498 |
| 14 | 10.635 | 0.384 | 0.000 | 9.981 | 10.502 | 9.498 |
| 15 | 10.915 | 1.047 | 0.000 | 10.168 | 10.502 | 9.498 |
| 16 | 11.258 | 2.054 | 0.000 | 10.386 | 10.502 | 9.498 |
| 17 | 10.225 | 2.027 | 0.000 | 10.354 | 10.502 | 9.498 |
| 18 | 10.018 | 1.794 | 0.000 | 10.286 | 10.502 | 9.498 |
| 19 | 10.058 | 1.600 | 0.000 | 10.241 | 10.502 | 9.498 |
| 20 | 10.880 | 2.229 | 0.000 | 10.368 | 10.502 | 9.498 |
| 21 | 10.378 | 2.355 | 0.000 | 10.370 | 10.503 | 9.497 |
| 22 | 11.265 | 3.369 | 0.000 | 10.549 | 10.503 | 9.497 |
| 23 | 11.013 | 4.130 | 0.000 | 10.642 | 10.503 | 9.497 |
| 24 | 9.610 | 3.489 | 0.139 | 10.435 | 10.503 | 9.497 |
| 25 | 10.898 | 4.135 | 0.000 | 10.528 | 10.503 | 9.497 |
| | | | $h = 2.513$ | $\lambda = 0.2$ | | |
| | | | $k = 0.5$ | $L = 3$ | | |

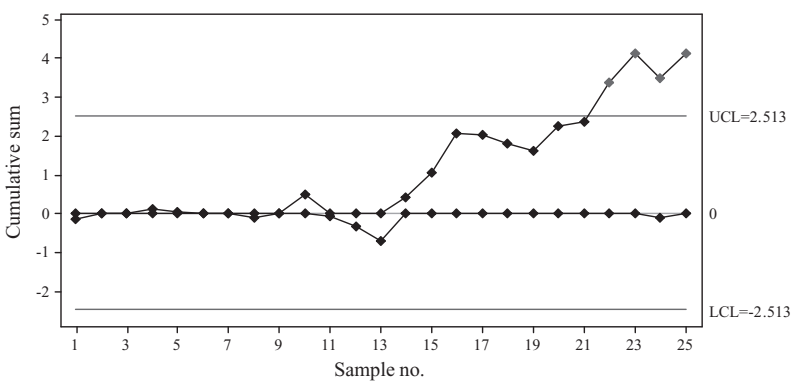


**FIGURE 12** CUSUM chart for Pet Tabs manufacturing example.

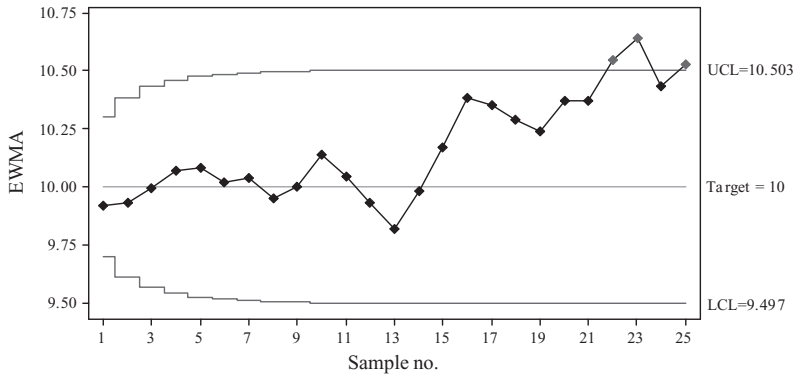


**FIGURE 13** EWMA chart for Pet Tabs manufacturing example.

used methods in process capability studies. Design-of-experiment techniques can be found elsewhere in this handbook and in many other textbooks.

***Specification Limits, Control Limits, and Natural Tolerance Limits*** To conduct a process capability study, it is important to distinguish the *specification limits* of a product, the *control limits* of the process producing the product, and the *natural tolerance limits* (NTLs) of the product. In general, specification limits are given by customers or prescribed by in-house design engineers before production. A product that failed to meet the specifications is a nonconforming product. Control limits are usually determined by samples collected from a process during a base period. A sample point that fell outside the control limits will trigger an out-of-control state; however, a product produced in the out-of-control state is not necessarily a nonconforming product. It should also be noted that a sample point plotted in a control chart usually represents a statistic of the sample such as the sample mean. In other words, a single product that fell outside of the control limits will neither cause the process to be out of control nor become nonconforming. The variability of products produced can usually be described by its natural tolerance limits. It is commonly acceptable that the ±3 standard deviations from the process mean be used as the natural tolerance limits.

**Example 4** Following Example 2, the specification limits are specified as 10.00 ± 2.00, where:

Nominal or target value ($\mu_0$) = 10.00
Upper specification limit (USL) = 10.00 + 2.00 = 12.00
Lower specification limit (LSL) = 10.00 − 2.00 = 8.00

The control limits for the $\bar{x}$ chart are:

Center line ($\bar{\bar{x}}$) = 10.254
Upper control limit (UCL) = 11.761
Lower control limit (LCL) = 8.747

Using the $\bar{\bar{x}} \pm 3\hat{\sigma}$ natural tolerance limits, they can be obtained as:

Process mean ($\bar{\bar{x}}$) = 10.254
Upper natural tolerance limit (UNTL) = 13.270
Lower natural tolerance limit (LNTL) = 7.238

The relationships among the three sets of limits are illustrated in Figure 14. As can be seen from this figure, the current process is not centered at its nominal value and its specification limits are tighter than its natural tolerance limits. Due to this, a portion of manufactured products (~5.4%) will not be able to conform to the specifications.

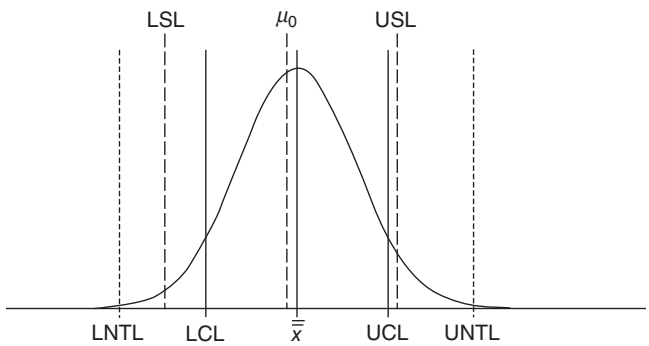


**FIGURE 14** Specification limits, control limits, and natural tolerance limits for Pet Tabs manufacturing example.

### 3.2 Process Capability and Improvement Studies

***Process Capability Indices*** Process capability indices provide a quantitative measure to assess the ability of a process to produce products that meet the specifications. A commonly used process capability index, denoted as $C_p$, can be calculated as

$$C_p = \frac{\text{USL} - \text{LSL}}{6\sigma} \tag{22}$$

where USL is the upper specification limit, LSL is the lower specification limit, and $\sigma$ is the process standard deviation. Since $\sigma$ is not usually known, it can be estimated by $\hat{\sigma} = \bar{s}/c_4$. A $C_p = 1$ means that the process is just capable. If the process is centered at its nominal value, it will produce 2,700 nonconforming products out of one million (PPM). The target value $C_p$ is usually set at 1.33 for an existing process and 1.50 for a new process.

It should be noted that the $C_p$ value could not indicate the proper process capability if the process is not centered since $C_p$ does not account for where the process mean is with respect to the specifications. To alleviate this issue, another process capability index, $C_{pk}$, is used:

$$C_{pk} = \min(C_{pu}, C_{pl})$$

where

$$C_{pu} = \frac{\text{USL} - \mu}{3\sigma} \quad \text{and} \quad C_{pl} = \frac{\mu - \text{LSL}}{3\sigma} \tag{23}$$

The μ value can be estimated by $\bar{\bar{x}}$ and the σ value can be estimated as discussed earlier. In general, a process is considered "centered" at the nominal value of the specifications when $C_p = C_{pk}$ and "off centered" when $C_p < C_{pk}$. The relationships between $C_p$ and $C_{pk}$ are further illustrated in Figure 15 where the process mean has shifted from $\mu_0$ to $\mu_0 + 2\sigma$ to $\mu_0 + 4\sigma$. As noted from the figure, $C_p$ remains the same regardless of the shift but $C_{pk}$ is significantly reduced.

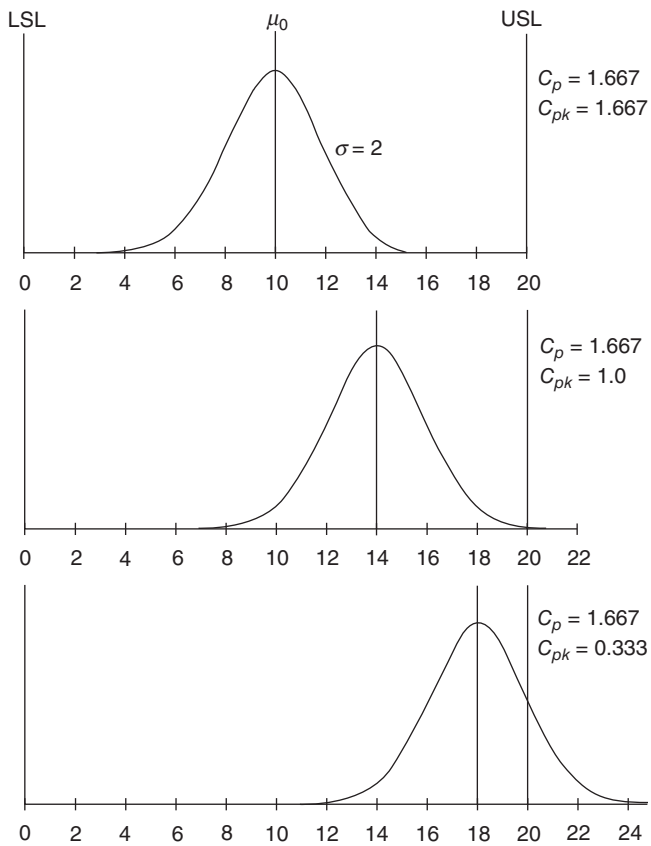


**FIGURE 15** Relationships between $C_p$ and $C_{pk}$.

If one-sided specifications are used, one-sided process capability can also be defined by Equation (23) where $C_{pu}$ is for upper specification and $C_{pl}$ for lower specification.

***Interpretation and Improvement of Process Capability*** Evaluation and interpretation of process capability represent an important step in process quality

improvement. A process must have its source of instability eliminated before it can be improved. Results obtained from process capability studies can help determine whether the process is stable and meeting its specifications. It should be noted that a valid process capability study is based on the normality assumption of the process. The normality assumption will need to be checked before proceeding to the next step.

Conclusions regarding whether the process is centered at the target and is meeting the specifications can be drawn from the process capability study. When $C_p = C_{pk}$, the process is centered. When $C_p$ has a value of 1.0 or greater, the process is capable of producing products meeting specifications; otherwise, it is not capable.

**Example 5** Following Example 2, $C_p$ and $C_{pk}$ are calculated as

$$C_p = \frac{\text{USL} - \text{LSL}}{6\sigma} = \frac{12 - 8}{6 \times 1.0054} = 0.6633$$

$$C_{pk} = \min(C_{pu}, C_{pl}) = \min(0.5790, 0.7476) = 0.5790$$

where

$$C_{pu} = \frac{\text{USL} - \mu}{3\sigma} = \frac{12 - 10.254}{3 \times 1.0054} = 0.5790$$

$$C_{pl} = \frac{\mu - \text{LSL}}{3\sigma} = \frac{10.254 - 8}{3 \times 1.0054} = 0.7476$$

Figure 16 shows the histogram of the data in relation to the specifications. The $\bar{x}$ and $s$ charts in Figure 11 show that the process is in statistical control. However, since $C_p < C_{pk}$, the process is not centered. With a $C_{pk}$ value of 0.579, it is expected to have 53,711 nonconforming Pet Tabs manufactured out of one million parts in this production line.

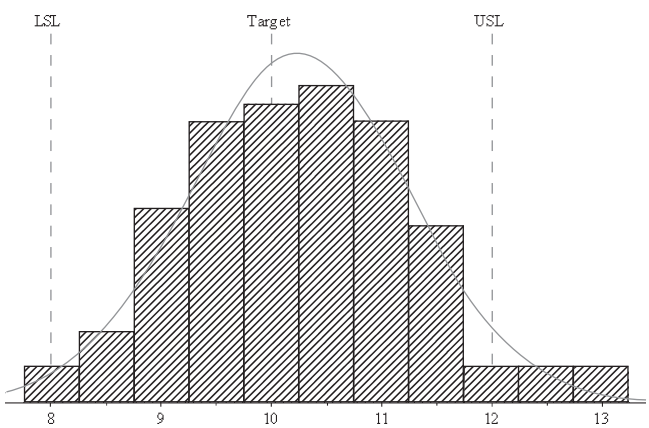


**FIGURE 16** Process capability plot for Pet Tabs manufacturing example.

To improve the process capability, the process needs to be centered first. This usually involves adjusting the process settings. Cause-and-effect diagrams, Pareto charts, and other tools discussed earlier in the chapter can be employed to find causes to the "off-centering" problems. After the process is brought back to its nominal (10), the total nonconforming Pet Tabs produced will be dropped to 46,673 PPM. This is still far from the 2700 PPM for a "just capable" process ($C_p = 1$), not to mention reaching the goal of 63 PPM at $C_p = 1.33$.

To further improve the process capability, the variability needs to be reduced. This can be achieved by designed experiments. Design of experiment (DOE) is a systematic approach that allows engineers and managers to make intentional changes in some process settings and assess the effects of those changes. An experiment can be designed in this example by varying a few key process settings such as drying time, mixing time, and temperature. Through a series of experimentations, optimum settings are found for these process variables and the variability of the process is reduced by 50%. With this reduction in process variability, the process is now exhibiting a $C_p$ of 1.265 with 694 PPM. This example highlights the benefits of process improvement. The move from an off-centered state to a centered state resulted in a reduction of process fall-out by 13.1%. With designed experiments, the process variability was cut in half and the process fall-out was significantly reduced by 98.5%.

## BIBLIOGRAPHY


Aft, L. S. (1997), *Fundamentals of Industrial Quality Control*, 3rd ed., CRC Press, Boca Raton, FL.

DeVor, R. E., Chang, T. H., and Sutherland, J. W. (2006), *Statistical Quality Design and Control*, 2nd ed., Prentice-Hall, Upper Saddle Brook, NJ.

Gitlow, H., Gitlow, S., Oppenheim, A., and Oppenheim, R. (1989), *Tools and Methods for the Improvement of Quality*, Irwin, Boston.

Grant, E. L., and Leavenworth, R. S. (1996), *Statistical Quality Control*, 7th ed., McGraw-Hill, New York.

Ishikawa, K. (1982), *Guide to Quality Control*, 2nd ed., Asian Productivity Organization, Tokyo, Japan.

Juran, J. M., and Godfrey, A. B. (1998), *Juran's Quality Handbook*, 5th ed., McGraw-Hill, New York.

Ledolter, J., and Burrill, C. W. (1998), *Statistical Quality Control: Strategies and Tools for Continual Improvement*, Wiley, New York.

Montgomery, D. C. (2004), *Design and Analysis of Experiments*, 6th ed., Wiley, New York.

Montgomery, D. C. (2001), *Introduction to Statistical Quality Control*, 4th ed., Wiley, New York.

Ryan, T. P. (2000), *Statistical Methods for Quality Improvement*, 2nd ed., Wiley, New York.

Smith, G. M. (2003), *Statistical Process Control and Quality Improvement*, 5th ed., Prentice-Hall, Upper Saddle Brook, NJ.

Summers, D. C. S. (2006), *Quality*, 4th ed., Prentice-Hall, Upper Saddle Brook, NJ.

Tague, N. R. (2005), *Quality Toolbox*, ASQ Quality Press, Milwaukee.

Thompson, J. R., and Koronacki, J. (2001), *Statistical Process Control: The Deming Paradigm and Beyond*, 2nd ed., Chapman & Hall, New York.

# CASE FOR PROCESS ANALYTICAL TECHNOLOGY: REGULATORY AND INDUSTRIAL PERSPECTIVES

ROBERT P. COGDILL
*Duquesne University, Center for Pharmaceutical Technology, Pittsburgh, Pennsylvania*

**Contents**

1 Introduction
2 Basis for Process Analytical Technology
  2.1 Process Analytical Chemistry
  2.2 Quality Management
  2.3 Lean Manufacturing
3 Historical Factors Limiting Implementation of PAT
  3.1 Real and Perceived Technological Barriers
  3.2 Lack of Economic Incentive
  3.3 Regulatory Disincentives
4 FDA Twenty-First-Century cGMPs Initiative
  4.1 Conception of the Initiative
  4.2 Risk-Based Orientation
  4.3 Quality Systems Approach
  4.4 Science-Based Policies
  4.5 International Collaboration
5 PAT Evolution in Pharmaceutical Manufacturing
  5.1 Process Understanding
  5.2 PAT Principles and Tools
  5.3 Strategy for Implementation
6 PAT Implementation Process
  6.1 Preparation
  6.2 Assessment
  6.3 Analyze
  6.4 Control
  6.5 Release Philosophy
  6.6 Optimization

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

7 Perspectives on the Impact of PAT
References

## 1 INTRODUCTION

The implementation of process analytical technology (PAT) is occurring in what is perhaps the most exciting period of change in pharmaceutical manufacturing of the past three decades. A host of technological, regulatory, and market forces have converged during the last five years, yielding new opportunities for innovation in the development and operation of pharmaceutical production processes. A major driving force for change is the Food and Drug Administration (FDA) initiative to implement a modern, risk-based framework for regulation and oversight of pharmaceutical manufacturing [1]. The objectives of this section are to outline the historical background of process analytics, to provide an overview of PAT in the pharmaceutical industry and the business drivers for change, to summarize the FDA's new initiative and the PAT guidance [2], and to present a basic plan for PAT implementation. While the focus of this chapter is PAT, it should be kept in mind that PAT is an important part of the much broader and risk-based paradigm introduced by the twenty-first-century current good manufacturing practices (cGMPs) initiative.

## 2 BASIS FOR PROCESS ANALYTICAL TECHNOLOGY

Despite the fact that the FDA's PAT framework (and guidance) began to take form just ahead of the creation of the twenty-first-century cGMPs initiative in 2001, it is well known that several of the core concepts were pioneered decades ago by other manufacturing industries such as fine chemicals, semiconductors, petroleum, and consumer products. The main concepts that differentiate PAT from the traditional industrial pharmacy skill set (including pharmaceutical and materials science, chemistry, and engineering) are process analytical chemistry (PAC) and advanced manufacturing science (Figure 1).

For the purpose of this discussion, the term *manufacturing science* is meant to describe the science and technology related to modern innovations in the design and management of manufacturing processes. Since it is neither practical nor necessary to cover all aspects of modern pharmaceutical manufacturing science in detail, the following sections are intended to introduce two specific topics which are popular in the current industrial vernacular but are not covered in detail in the pharmaceutical literature: quality management systems and "lean" manufacturing.

### 2.1 Process Analytical Chemistry

Process analytical chemistry generally describes the science and technology associated with displacement of laboratory-based measurements with sensors and instrumentation positioned closer to the site of operation. Although industrial process analyzers have been in use for more than 60 years [3], the modern period of PAC essentially began with the formation of the Center for Process Analytical Chemistry (CPAC) in 1984 [4]. As described by Callis, Illman, and Kowalski [5], the goal of

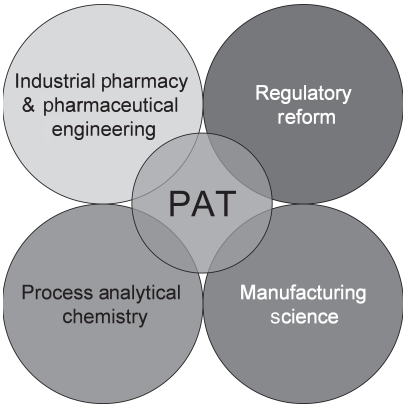


**FIGURE 1** Multidisciplinary components of PAT in pharmaceutical manufacturing.

PAC is to "supply quantitative and qualitative information about a chemical process" for monitoring, control, and optimization: they went on to define five "eras" of PAC: (1) off line, (2) at line, (3) online (4) inline, and (5) noninvasive, which describe the evolution of sensor technologies. In addition, they discussed the importance of issues beyond chemical sensing, such as sampling, extraction of information from data (chemometrics), integration with process controls, as well as the sociological aspects of PAC deployment (e.g., gaining the trust of plant operators).

The industrial PAC movement has been bolstered by two decades of advances in materials science, electronics, and chemometrics. Since the inception of CPAC, the pace of innovation in sensors, instrumentation, and analytics has quickened dramatically. The development of more robust, sensitive photodetector materials, microelectromechanical systems (MEMSs), and fiber optics and the perpetual advancement of computing power (as predicted by Moore's law) have both increased the performance and reduced the cost of PAC. As a result, PAC is now a critical part of routine operations within the realm of industrial chemistry. Many general reviews on the subject of PAC (and PAT) have been published [6–10]. A series of literature reviews on the subject of PAC have been published regularly in *Analytical Chemistry*.

The first review [11] listed manuscripts published between 1987 and 1992, covering seven specific topics (general PAC, chromatography, optical spectroscopy, fiber optics, mass spectrometry, chemometrics, and flow injection analysis), along with a section on needs for the future of PAC; in all, the first review included 507 references. Subsequent reviews were published in 1995 [12], 1999 [13], 2001 [14], 2003 [15], and 2005 [16]. The review series is an essential resource for scientists seeking information on specific PAC methods; in total, 2650 references covering more than 16 topics were catalogued by the authors.

Currently there are three major consortia involving university, government, and industrial partners—CPAC, the Measurement & Control Engineering Center (MCEC), and the Control Theory and Applications Centre (CTAC)—along with an annual conference, the International Forum on Process Analytical Chemistry (IFPAC), and numerous online resources that are devoted to issues related to process analytics [16]. In parallel with the FDA's initiative, the term *process*

*analytical chemistry* is gradually being replaced in the industrial vernacular by *process analytical technology*. This reflects the expansion of the field as the importance of physical characterization, risk analysis, and manufacturing science is recognized.

## 2.2 Quality Management

Many of the quality improvement goals for implementation of PAT in the pharmaceutical industry have been achieved by companies in other industries, such as automobile production and consumer electronics, as a direct result of adopting principles of quality management. The lineage of modern quality management can be traced to the work of Walter Shewhart, a statistician for Bell Laboratories in the mid-1920s [17]. His observation that statistical analysis of the dimensions of industrial products over time could be used to control the quality of production laid the foundation for modern control charts. Shewhart is considered to be the father of statistical process control (SPC); his work provides the first evidence of the transition from product quality (by inspection) to the concept of quality processes [18, 19].

Shewhart's methodologies were adopted and expanded by W. Edwards Deming [20] and Joseph M. Juran [21], who are credited with the birth of the "total quality" (TQC) approach in Japan following World War II. Successors to the total quality movement include management by objectives (MBO) (1960's), Crosby's zero-defects (ZD) movement (1970s), the American incarnation of total quality management (TQM) (1970s–1980s), quality circles (1970s), quality function deployment (QFD) (1980s), the International Organization for Standardization (ISO) 9000 series (1987), and the Malcolm Baldridge National Quality Award (1987–present). The most recent major quality management methodology, Six Sigma ($6\sigma$) [22, 23], pioneered by Motorola, has become immensely popular because of the litany of corporate CEOs (e.g., Thomas Galvin, Jack Welch) who have openly credited their internal $6\sigma$ initiatives for dramatic improvements in bottom-line performance. All of these quality movements [24], however, as well as PAT, are related to the principles of Shewhart, Deming, Juran, Crosby, Taguchi [25, 26], and others, in that they are based on systematic methods for understanding the sources of variability in processes and minimizing their impact on product quality.

The so-called DMAIC (define, measure, analyze, improve, and control) methodology is a common framework used by improvement teams in many industries to apply the concepts of quality management to systematically identify, prioritize, and eliminate the root cause of quality problems. A variant of DMAIC, known as DMADV (define, measure, analyze, design, and verify), is sometimes used when a process or operation requires complete redesign to bring about the desired quality improvement and is a central concept of the DFSS (design for six sigma) movement. The origins of DMAIC, DMADV, DFSS, and other various quality management cycles can be traced to the "Shewhart cycle" of (1) plan, (2) do, (3) study, and (4) act [24].

Arguably, the most important aspects of quality management for PAT are the concepts of quantitative process performance characterization using process capability indices as universal descriptors, which form the basis of the "measure" and "analyze" portions of the DMAIC model. Process capability indices consider simultaneously both process variability and process specifications to determine whether

the process is "capable" [27]. A process is said to be capable if the quality measurements for nearly all samples are within the specification limits. A common version of the process capability index, $C_{pk}$, is calculated according to

$$C_{pk} = \min\left[\frac{\mathrm{USL} - \mu}{3\sigma}, \frac{\mu - \mathrm{LSL}}{3\sigma}\right]$$

where μ and σ are the mean and standard deviation and USL and LSL are the upper and lower specification limits, respectively, for a product quality measurement. Process capability indices are useful for process improvement studies because they transform diverse measures of quality (e.g., weight, concentration, rate) into dimensionless units, thereby allowing investigators to pinpoint major sources of variation in a process (operations which have the lowest $C_{pk}$ scores) when many measurement systems and quality attributes are involved.

The process capability index, $C_{pk}$, is related to the so-called "process sigma" such that a 6σ process corresponds to a $C_{pk}$ of exactly 2.00, or 2.0 defective parts per billion (PPB), assuming $\sim N(0, \sigma)$ quality variance distribution (can alternative calculation for process sigma estimates 3.4 defective parts per million for a 6σ process). Examples of the correspondence between $C_{pk}$, process sigma, and defect rate for $\sim N(0, \sigma)$ distributions are shown in Figure 2. The process capability (based on observed yield) of pharmaceutical manufacturers has been cited by some benchmark studies to be roughly 0.7 (2.1σ) [28].

While industrial benchmarks clearly indicate that pharmaceutical manufacturers have many opportunities to improve quality control, direct comparison with other industries may be somewhat misleading. As opposed to such industries as semiconductor manufacturing, where defective parts are often readily apparent at some

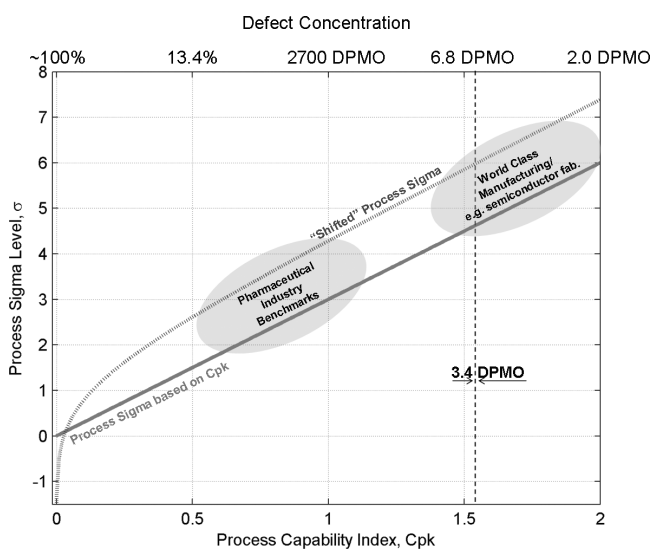


**FIGURE 2** Graphical illustration of the correspondence between the defects per million opportunity (DPMO) process capability ($C_{pk}$) and process sigma (assuming normally distributed quality variation).

point in the value chain (i.e., the device built from the part will fail), drug products suffer from a high degree of ambiguity in their quality specifications.

For example, finished-product release specifications such as content uniformity are rarely correlated to clinical evidence; rather, they are set according to compendial test standards. Furthermore, the functional relationship between in-process material characteristics and finished-product quality is seldom known at a high level; hence, the assigned in-process specifications for some operations may over- or underestimate the true level of process capability. As the level of process understanding in the pharmaceutical industry increases, development of science- and evidence-based in-process and release specifications will improve the reliability of $C_{pk}$ as a tool for process characterization.

For further information, the NIST/SEMATECH *Handbook of Engineering Statistics*, which is freely available online [23], and the American Society for Quality (www.ASQ.org), are excellent sources for background information and technical details related to quality management.

### 2.3 Lean Manufacturing

In contrast to quality management systems, which have clear parallels with PAT (i.e., reduction of quality variation), the links between PAT and lean manufacturing are less direct. In fact, while quality management systems are concerned with process analysis of quality variation, lean flow path management is concerned with process analysis of production time variation. Furthermore, the core concepts of lean manufacturing, however, provide the technology platform which the pharmaceutical industry will use to derive gains in production efficiency from the adoption of PAT. Without considering the impact of PAT on production efficiency [i.e., the return on investment (ROI) from implementing PAT], industry would have very little impetus to voluntarily embrace PAT. The following paragraphs are intended to provide a brief introduction to lean manufacturing; later portions of this chapter will discuss the business drivers for implementation of PAT.

Lean manufacturing, or "lean," is often misunderstood (not unlike TQM or 6σ); for some people, lean business initiatives conjure "slash-and-burn" management tactics to reduce workforce levels or shut down low-productivity operations. In fact, lean manufacturing has been characterized as "an amalgam of methodologies including industrial engineering, just-in-time (JIT) (Osadas's) 5-S's, TQC, continuous quality improvement (CQI), Visual Control, Total Productive Maintenance (TPM), Quality Circles, and Kaizen" [24].

The origins of lean manufacturing are often ascribed to the creation of the Toyota Production System (TPS) by the Toyota Motor Corporation. However, the history of lean manufacturing can be traced back to industrial developments which occurred more than 150 years before TPS. The foundation for modern manufacturing was laid by Eli Whitney in 1798; while Whitney is best known for his invention of the cotton gin, it is his invention of interchangeable parts and uniform production which revolutionized mass production (www.EliWhitney.org).

Nearly a century later, Frederick W. Taylor introduced the concepts of time study and standardized work, coining the term *scientific management*. It was not until 1908, with Henry Ford's introduction of the Model T, that the value of lean manufacturing was recognized worldwide. Henry Ford is considered by some to be the

first practitioner of JIT manufacturing; furthermore, his manufacturing system has been described as the inspiration for TPS [29]. More recently Ford Motor Company has developed a modernized version of Henry Ford's original system, the Ford Production System [24], which borrows heavily from TPS.

As a discipline of manufacturing science, lean manufacturing is a technical philosophy focused on the reduction of seven types of waste, or "muda," in manufacturing: overproduction, waiting, transport, inappropriate processing, unnecessary inventory, excess motion, and defects. The transformation of a process to lean operation is accomplished using many tools and strategies. Arguably, the most important mechanism for change is to replace traditional "make to forecast" or "push" production scheduling with "pull" strategies, such as "kanban" cards. The principles of lean have been applied with success in manufacturing and service industries, as well as governmental entities. Not unlike quality management, there are literally hundreds of books describing the various tools and techniques used to apply lean methodologies. The Society of Manufacturing Engineers (www.SME.org) maintains publications, conferences, and a technical community devoted to production management and is a good first source for more information on lean manufacturing.

Compared with other industries, pharmaceutical manufacturers have been relatively late to adopt lean manufacturing; consequently, pharmaceutical cycle times are extremely long when compared with other industries [30, 31]. By comparing the ratio of total cost of goods sold (COGS) to inventory value for the top 22 publicly traded branded, generic, and biotech pharmaceutical companies to the reported figures for other process industries, a rough indication can be gained of how much less effectively pharmaceutical manufacturers manage their supply chains (Figure 3). Furthermore, it would not be difficult for most industrial pharmaceutical scientists to find common examples of each of the "seven wastes" in a typical pharmaceutical manufacturing facility.

Admittedly, there are some constraints intrinsic to the industry which may ultimately prevent pharmaceutical manufacturers from achieving "world-class" supply chain and manufacturing performance. Furthermore, application of lean and quality management tools to pharmaceutical manufacturing is proving to be a unique challenge. A recent survey of 1500 pharmaceutical manufacturing professionals indicated that, while more than half of the companies surveyed have implemented lean, 6σ, or operational excellence, less than half of those programs have yielded satisfactory results [32].

While these data seem to suggest that lean manufacturing is not suited to pharmaceutical manufacturing, it is important to consider that most lean methodologies (e.g., TPS) were developed for high-volume production of uniform products. Although many "blockbuster" drugs are produced in dedicated facilities or in plants specializing in only a few products, it is quite common for pharmaceutical manufacturers to produce many products in a single plant, having a high proportion of shared equipment. Traditional lean methods, such as kanban cards, are difficult to manage in a complex, "high-mix" production environment. In order to solve these limitations, innovative software algorithms for "flow path management" [33] have been developed to simulate, design, and optimize pharmaceutical production processes according to lean manufacturing principles.

Furthermore, the effectiveness of lean manufacturing is limited by variability in the cycle time (C/T) for individual unit operations as well as by the finite risk of

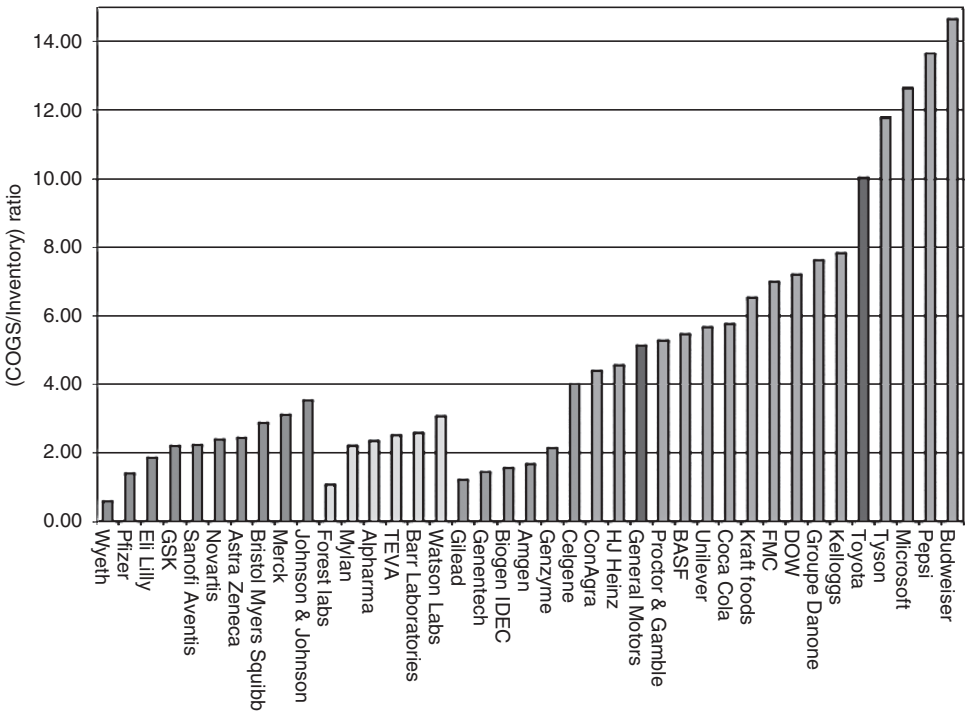


**FIGURE 3** Ratio of total COGS to reported inventory value. The ratio of COGS to inventories is a rough indicator of supply chain velocity. A large ratio implies that inventories are small relative to COGS and are turned over frequently. Toyota Motors, for example, which is well known for effective supply chain management, has a much higher ratio of COGS to inventories than General Motors.

batch failure during production; this is true regardless of the complexity of the flow path or of the degree to which equipment is shared. Pharmaceutical manufacturers cope with such risks by building up long production queues to accumulate work in process (WIP) ahead of unit operations. While this helps to improve capacity utilization and overall equipment effectiveness (OEE), it decreases efficiency by consuming working capital and increasing the intensity of overhead operations (required to finance, transport, and warehouse WIP). In order to gainfully implement lean, pharmaceutical manufacturers must first minimize C/T variation and risks to product quality.

Finally, it is well known that a significant portion of the typical production C/T reported by industry is consumed by the delay between completion of a unit operation, sampling, analysis, reporting, and in-process or finished-product release. In some scenarios, PAT will enable manufacturers to release finished products to the market immediately, with no delay for manual, offline testing; this is the so-called real-time release (RTR) benefit of PAT. Without PAT and RTR, the effectiveness of lean strategies in reducing C/T will be limited by the maximum rate of product inspection and release. Thus, it is critical for pharmaceutical manufacturers to deploy PAT and lean in parallel if real gains in process performance are to be realized. The lean–PAT concept is quite similar to lean–6σ, or "fusion management" [24].

## 3 HISTORICAL FACTORS LIMITING IMPLEMENTATION OF PAT

Despite the evidence of fiscal and competitive benefits enjoyed by the various industries which have embraced process analytics, pharmaceutical companies have been notoriously restrained in their efforts to deploy PAT. Indeed, the pharmaceutical industry has slipped so far behind peer industries that a well-known *Wall Street Journal* article from 2003 [34] characterized the manufacturing prowess of drug makers as lagging "far behind potato-chip and laundry-soap makers." While the declaration was shocking to many, it was, nonetheless, an accurate assessment. Before indicting the industry for gross negligence, however, it is important to consider the various factors which have acted over time to create the current state of affairs.

Over the years, dozens of excuses have been provided for the industry's lack of manufacturing innovation; many of the reasons are well known and have been published elsewhere [35]. For the sake of simplicity, the factors limiting the adoption of PAT can be distilled into three categories: real and perceived technological barriers, lack of economic incentive, and regulatory disincentives.

### 3.1 Real and Perceived Technological Barriers

Despite the fact that near infrared spectroscopy (NIR) has been used industrially for decades [36], there has been hesitance to accept and trust "new" process analytical measurement technologies as equivalent or superior to traditional methods. For example, when a discrepancy between online NIR and laboratory analyses is observed, it is rare that the destructive reference methods are ever targeted as the source of error, despite the fact that NIR is often the more precise method. The hesitance to trust more advanced, multivariate tools (which are perhaps less directly understood) has certainly been a detriment to progress in deploying PAT.

Similar concerns persist with regard to chemometrics (multivariate data analysis), information technology (IT), and advanced controls. One reason for such behavior may be the practice of calibrating and validating PAT sensors by correlating their signals to traditional, laboratory-based reference methods and characterizing performance in terms of prediction error [37–39]. It is a truism of statistics that, no matter how sensitive or accurate the PAT sensor may be in detecting quality variation, the performance of the reference method will always limit the level of perceived accuracy. A much more accurate depiction of the performance of PAT sensors compared to reference techniques would be to compare analytical figures of merit, such as signal-to-noise ratio (S/N) or analytical sensitivity, which explicitly account for measurement precision [40, 41].

Even though the perceptions of PAT instrumentation have begun to improve, companies continue to worry that the intensity of product quality sampling afforded by PAT sensors will result in negative consequences, such as increased inspection and investigations. In other words, many companies continue to "fear what they will find" if they begin to analyze their operations more closely. Prior to the introduction of rapid, nondestructive quality monitoring tools, there were few alternatives for efficient quality assurance except to rely on batch release criteria, such as the well-known U.S. Pharmacopeia (USP) ⟨905⟩ procedure, which were based on extremely limited sampling (i.e., assay 10 individual dosage units from a 30-unit sample of a production-scale batch).

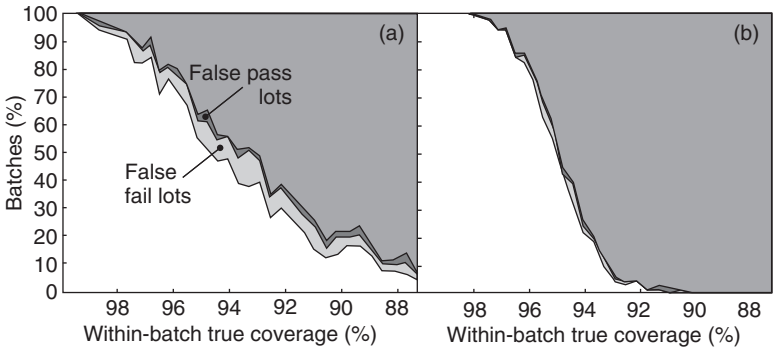


**FIGURE 4** Comparison of operational characteristic (OC) curves for the USP ⟨905⟩ (*a*) and PAT-based (*b*) release strategies generated by Monte Carlo simulation. The USP OC curve (*a*) is based on the assumption of 2% RSD measurement precision; the PAT OC curve (*b*) assumes NIR measurement of 800 tablets with 0.9% measurement precision; both curves were estimated using the same simulated populations of one million tablets having varying levels of quality uniformity. Each curve consists of four regions: the regions above and below the sigmoid curve correspond to proportions of batches accurately passed or rejected based on the release criteria. Along the sigmoid curve are regions related to the rates of false batch failure (lower side of curve) and false batch acceptance (upper side of curve). The jagged nature of the curves is related to the limitations imposed by finite iterations. The slope of the curves demonstrates the superior specificity (or "tunability") of release tests optimized for PAT systems.

Despite the fact that the operating characteristic (OC) curve of the USP ⟨905⟩ test guarantees a significant portion of each batch will have poor quality before batch rejection is probable [42, 43], companies have become comfortable with their odds. Process analytical monitoring tools such as NIR spectroscopy, which are capable of high-speed sampling in line, online, or at line, have been perceived as an additional burden on the rate of successful batch release.

By forgoing real-time, pervasive quality monitoring, however, companies incur significant opportunity costs in at least three ways. First, without continuous monitoring there are few feasible opportunities for implementing RTR; time delays related to offline release testing are one of the most significant factors limiting supply chain velocity in pharmaceutical manufacturing. Second, while there is some potential for "discovering" a greater number of batches which do not meet release criteria, statistical simulations suggest that potentially fewer batches will be rejected when larger sample sizes are considered. In other words, when the impact of measurement imprecision and the true distribution of quality characteristics are considered, traditional release testing methods pose finite risks of failing passable batches (which otherwise should have passed) because the limited sample does not adequately represent the characteristics of the population (Figure 4). Finally, and perhaps most importantly, traditional sampling techniques are an effective barrier to continuous improvement; based on fundamentals of statistical theory, it can be shown that samples of at least hundreds of individuals are required to detect incremental changes in process capability (Figure 5). Hence, even if a company were to investigate potential process improvements, only process capability changes of improbable magnitude would be recognized with statistical confidence.

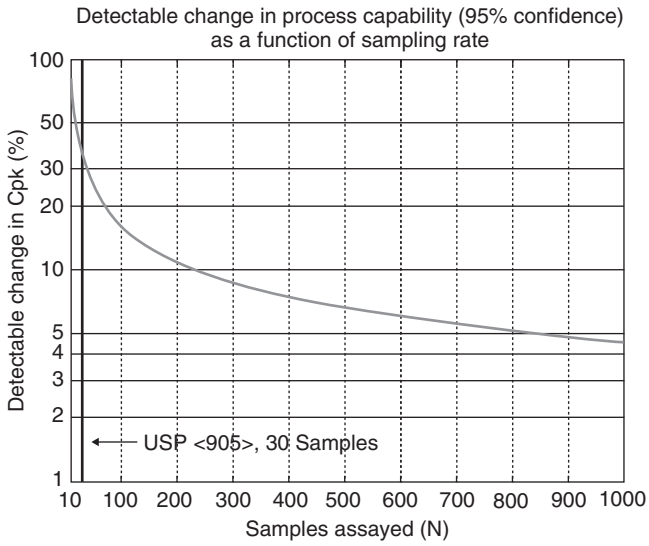


**FIGURE 5** Relationship between sampling rate and effective resolution of process capability assessment. The curve is based on the width of the confidence intervals for estimation of mean and variance. The relationship shown does not consider the effect of reference measurement precision, which would further reduce the ability to discern changes in process capability.

### 3.2 Lack of Economic Incentive

A common refrain within the industry has been that there simply is not sufficient financial return from investment in process analytics or manufacturing technology upgrades to justify spending. In some respects, this is a valid argument. Historically, many of the industries which have justified significant investment in process analytics utilized continuous manufacturing; it is far more difficult to efficiently control continuous processes (relative batch production systems) without real-time process analytics [35]. Hence, while the pharmaceutical industry has been able to choose, many other manufacturers have been forced to integrate PAT into their operations.

Since pharmaceutical investment in PAT continues to be an option rather than a priority for most companies, arguments justifying PAT spending are forced to compete with other spending initiatives for capital. During each planning cycle, company managers must decide whether to allocate additional capital toward diverse opportunities, such as greater research and development (R&D), improvements in manufacturing capabilities, or additional forces in sales and marketing [i.e., selling, general and administrative (SG&A)]. For any particular project to be funded, expected returns must not only exceed the company's cost of capital [i.e., weighted average cost of capital (WACC)], winning projects may be required to exceed the company's expected return on invested capital (ROIC) or at least provide expected returns in excess of other investment alternatives. A recent academic case study of the potential financial returns on investment (ROI) in PAT and lean manufacturing in the pharmaceutical industry show, however, that many pharmaceutical manufacturers could ultimately benefit tremendously by improving manufacturing performance [44].

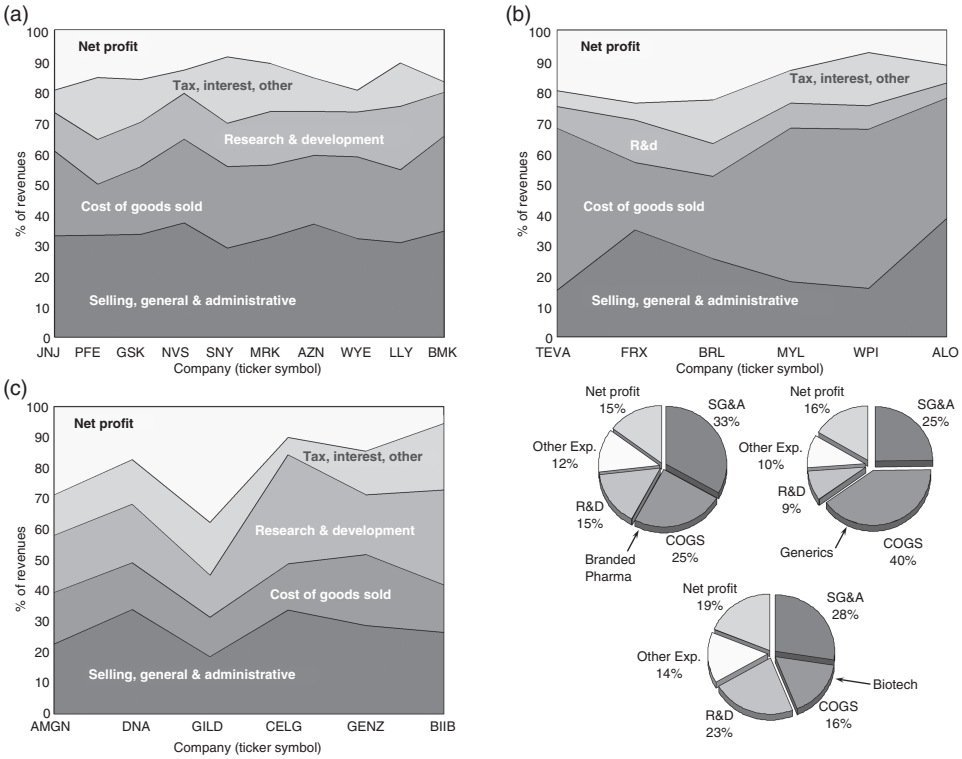


**FIGURE 6** Distribution of the components of revenue (FY2005 annual data) for branded (*a*), generic (*b*), and biotech (*c*) drug manufacturers. Companies are arranged according to market capitalization (as of November 2006).

Unfortunately, proponents of PAT are only just beginning to develop the methods to quantify all of the potential opportunities for ROI. Furthermore, it is important to consider the relative level of risk posed by investment in PAT (as opposed to other alternatives). Unlike investments in sales or marketing, there remains considerable uncertainty in the industry regarding the likelihood of achieving ROI projections or the prospect of PAT investment creating new problems. For these reasons, management teams have typically found it easier to justify spending in R&D and marketing instead of PAT or manufacturing reforms.

Besides concerns over the likelihood and magnitude of returns on PAT investments, it is often cited that manufacturing and optimizing the cost of production have simply not been a priority in the industry; manufacturing has often been viewed as a cost rather than a value-generating component. The distribution of corporate expenditures has been provided as evidence in support of this theory (Figure 6). Based on corporate annual income statements from 2005, the average expenditure on R&D and SG&A among the top-10 branded pharmaceutical companies (by market capitalization, November 7, 2006) was nearly double their reported cost of goods sold. Another take on this theory is that institutional and individual investors (who own the pharmaceutical companies and supply the capital for their operation) and the boards of directors elected by them look favorably on the expansion of R&D and marketing investment while taking a more myopic view on the importance of manufacturing. It has sometimes been said that Wall Street rewards

(pharmaceutical companies) for innovation in discovery and replication in manufacturing [45]. It is not completely coincidence, for example, that Merck's appointment of its president of manufacturing, Richard T. Clark, to chief executive in May 2005, which, according to financial journalists, "disappointed investors" who apparently would have preferred someone with a "research and development background" [46], marked the beginning of a nearly 25% loss in market capitalization over the next six months.

While the various reasons discussed for the pharmaceutical industry's tepid approach to PAT and manufacturing reform are plausible, they are likely secondary to the real and perceived risks posed by the regulatory uncertainty surrounding innovation in manufacturing. For example, it is well known that many companies were beginning to use PAT tools long before the FDA's initiative, which suggests that the economic benefits of process analytics have been recognized internally for some time. In response to the fear that their use of new technologies would spur additional investigations by the FDA, however, some of these companies operated in a "Don't use, or don't tell" manner with regard to PAT [45].

### 3.3 Regulatory Disincentives

The real and perceived fear of regulatory noncompliance has arguably been one of the most important factors explaining the industry's reluctance to pursue manufacturing innovation [1, 2]. While the first 25 years of pharmaceutical GMP have been effective in ensuring the safety of prescription drug products for consumers, it has been achieved at the expense of innovation and flexibility. Without the ability to adjust processes to account for changes in materials, operating conditions, or the level of process understanding, process analytics are of nearly no value since there is no capacity to act on new information (besides material/batch rejection).

Furthermore, companies who dared to make changes or implement new technologies, whether conventional process improvements, new unit operations, or process analytics, were met with extensive supplemental documentation, FDA inspection, and the finite risk of production delays. Ultimately, the potential for regulatory action stifled the industry's desire to pursue technologies which might have seemed extraordinary, such as real-time analytics or chemometrics. Finally, without the benefits conferred by the PAT guidance and risk-based cGMPs initiative, industry rarely had incentive to formally analyze the risk of established processes out of fear that what they might discover would be used against them in regulatory or legal actions.

## 4 FDA TWENTY-FIRST-CENTURY cGMPs INITIATIVE

The observation that the state of cGMP at the beginning of the twenty-first-century was stifling innovation in pharmaceutical manufacturing did not go unnoticed by the FDA, which also saw opportunity in remodeling the regulatory framework. Since many changes, even minor operational modifications, required prior approval from the agency prior to implementation, regulators were swamped with thousands of supplements every year. Resources were stretched between processing of supplements, review and approval of new facilities, processes and documentation, and inspection; all the while, the FDA was being squeezed by external constraints on

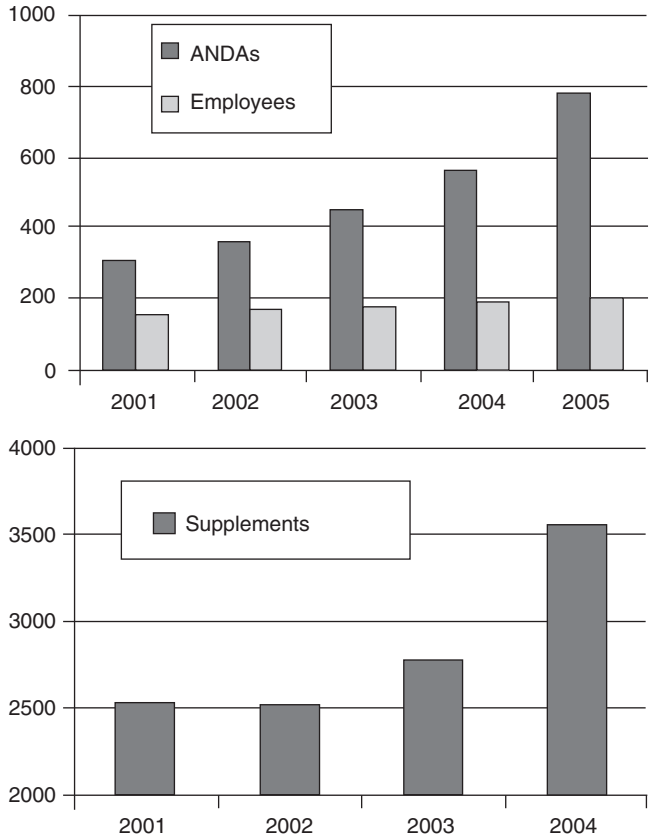


**FIGURE 7** Trends in FDA workload and staffing resources. (Adapted from L. X. Yu, Implementation of quality-by-design: Question-based review, Drug Information Association (DIA) 42nd Annual Meeting, Philadelphia, PA, 2006.)

budget growth (Figure 7). As of 2001, FDA regulators were so burdened that they were unable to meet statutory biennial GMP inspections. Finally, the load of supplements, reviews, and inspections were acting as a significant drag on the advancement to market of new pharmaceutical therapies.

### 4.1 Conception of the Initiative

The agency began a public dialogue on the state of pharmaceutical manufacturing and FDA regulation during discussions with the Advisory Committee for Pharmaceutical Science (ACPS) in July 2001, followed by further discussion within the FDA Science Board meetings in November 2001 and April 2002 [47]. A significant focus of the discussions was the impact of the regulatory framework on innovation, quality, and efficiency as well as opportunities for change. A new, risk-based paradigm which rewards innovative producers through opportunities for "regulatory relief" began to take shape, displacing the notion of regulatory compliance as a force for innovation. The new paradigm offered advantages to the FDA, as well, in that the level of inspection resources could be prioritized and allocated according to risk, thereby easing

the strain on FDA resources. These changes signaled an evolution of what seemed to be an adversarial FDA–industry relationship toward greater cooperation.

While the pharmaceutical incarnation of the term PAT was formally introduced during these meetings [48], a significant portion of the concepts which define the core of PAT in pharmaceutical science were presented by industrial and academic scientists, many of whom had been building support for and working on these issues within their organizations for years. Industrial and academic presentations included topics such as total quality management [49], new technologies for pharmaceutical manufacturing [50], and QbD [51], among others.

In August 2002, the agency announced the Pharmaceutical cGMPs for the 21st Century initiative (or "the initiative"), which began a two-year effort undertaken by a number of multidisciplinary working groups within the FDA, as well as the cGMP steering committee, to assess the current regulatory structure and define the agency's new vision for risk-based regulation of manufacturing and product quality. The new initiative, which was intended to modernize the FDA's regulation of pharmaceutical quality for human, veterinary, and select human biological products, sought to reform the pharmaceutical as well as the chemistry, manufacturing, and controls (CMC) programs, with the following specific objectives:

- Encourage the early adoption of new technological advances by the pharmaceutical industry.
- Facilitate industry application of modern quality management techniques, including implementation of quality systems approaches, to all aspects of pharmaceutical production and quality assurance.
- Encourage implementation of risk-based approaches that focus both industry and agency attention on critical areas.
- Ensure that regulatory review, compliance, and inspection policies are based on state-of-the-art pharmaceutical science.
- Enhance the consistency and coordination of the FDA's drug quality regulatory programs, in part, by further integrating enhanced quality systems approaches into the agency's business processes and regulatory policies concerning review and inspection activities.

The result of the working groups' assessment enabled the development of the new framework embodied by the finalized twenty-first-century cGMPs as well as the associated components, such as the PAT guidance. Throughout the assessment and development, and continuing during the "implementation phase" of the initiative, the following set of guiding principles has been maintained:

- Risk-based orientation
- Science-based policies and standards
- Integrated quality systems orientation
- International cooperation
- Strong public health protection

The final report on the results and future plans for the initiative were released in September 2004. The report efficiently describes the motives, origins, development

process, and mechanisms for implementing and evaluating the initiative and can be found posted on the Center for Drug Evaluation and Research (CDER) Office of Pharmaceutical Science (OPS) website (http://www.fda.gov/cder/OPS/).

Since it would be impractical to accurately describe all of the important aspects of the report within this space, the following sections are intended to detail some of the concepts and guiding principles of the initiative which are particularly important for understanding PAT. The organization of this summary is intended to efficiently describe selected concepts of the agency's twenty-first-century cGMPs and is not intended to mirror the structure or totality of the associated FDA documentation. All who are actively engaged in pharmaceutical manufacturing or are interested in PAT are encouraged to read the final report [1], which should be considered a primary source for direction.

### 4.2 Risk-Based Orientation

The FDA's adoption of a risk-based orientation for regulation is the most important aspect of the twenty-first-century cGMPs. It is a common misconception that the agency's initiative describes a new set of practices for the industry. In fact, while the FDA is committed to encouraging innovation in the industry, the twenty-first-century cGMPs initiative is entirely focused on changing the agency's regulatory framework so that quality and innovation are rewarded with reduced oversight. Now that the agency has entered the implementation phase of the initiative, many of the previous regulatory disincentives have been eliminated. In other words, pharmaceutical companies are currently free to voluntarily choose whether or not to pursue innovative changes in their development, operation, and quality assurance of manufacturing processes such as PAT.

***Risk-Based Prioritization of cGMP Inspections*** The mechanism by which the FDA will encourage the industry to join in implementing the new methods is provided by the risk-based algorithm for prioritizing cGMP inspections. Incidentally, risk-based site selection is the same mechanism which will allow the agency to optimally allocate its limited oversight resources to achieve the greatest public health impact. Operational efficiency is a major component of the FDA's plans for the future. The key to the risk-based site selection program is the agency's risk-ranking model, which has been deployed as a pilot program since the beginning of its 2005 fiscal year.

The model is based on a hierarchical risk-ranking and risk-filtering method whereby a *site risk potential* (SRP) is estimated as a function of the weighted potentials for each of three *top-level components* of site risk—product, facility, and process (Figure 8). The risk potential for each of the three top-level components is calculated as a function of selected risk factors which are relevant to the component (specific to the site). A set of subcategories are defined for each top-level component; each subcategory is comprised of individual risk factors. The initial model weights (the actual risk scores at the lowest level) were optimized using a combination of empirical evidence and expert judgment. Examples of potential risk factors for each top-level component (and associated subcategories) were provided in a report which describes the first iteration of pilot risk-ranking model in detail [52].

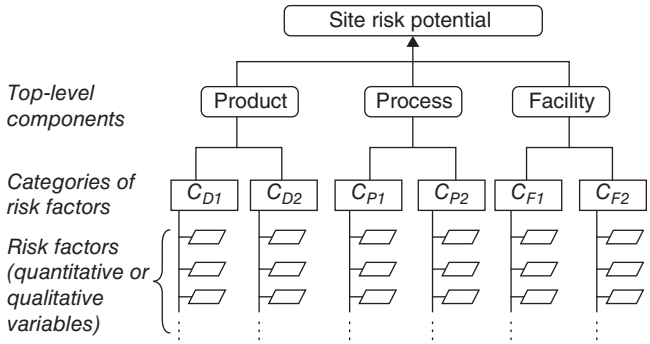


**FIGURE 8** Schematic of FDA's pilot risk-ranking model for calculation of site risk potential.

The results from the first iteration of the risk-ranking model demonstrated the capability of the model to spread SRP scores for the purpose of filtering. Future iterations of the risk-ranking model will be generated by correlating predicted site risk potentials with data gathered by traditional oversight activities (e.g., cGMP compliance inspections) and adjusting the risk factor weights to maximize the effectiveness of SRP prediction (similar to multivariate linear regression). The selection of risk factors included in the first iteration of the model was based on the availability of data. Some proposals for future iterations of the model include incorporating factors such as systems for continuous assessment of process capability as indicators of the site's level of process understanding and control. Certainly, as the model is updated to capture the benefits of new best practices in manufacturing, such as PAT, the risk ranking will begin to provide effective incentive for producers to pursue innovation.

### 4.3 Quality Systems Approach

According to the FDA staff manual guide [53], a *quality system* is a "set of formal and informal business practices and processes that focus on customer needs, leadership vision, employee involvement, continual improvement, informed decision making based on real-time data and mutually beneficial relationships with external business partners to achieve organizational outcomes." Based on this description, PAT should be considered to be an important tool for supporting a quality management system. As stated earlier, one of the FDA's objectives in undertaking the initiative was to integrate quality systems and risk management approaches into its existing programs with the goal of encouraging industry to adopt modern and innovative manufacturing technologies, including industrial deployment of quality management systems such as those described earlier in this chapter (e.g., ISO 9000).

In September 2006, the FDA released "Guidance for Industry: Quality Systems Approach to Pharmaceutical cGMP Regulations" [54]. The guidance is intended to "help manufacturers implementing modern quality systems and risk management approaches to meet the requirements of the Agency's cGMP regulations," in particular, Parts 210 and 211. In developing the guidance, the Quality System Guidance

**FIGURE 9** FDA's six-system inspection model.

Development (QS) working group "mapped" the relationship between cGMP regulations and various quality system models both internal and external to the FDA. Their result is a comprehensive model which allows producers seeking to implement their own quality management systems to quickly identify those aspects of quality systems which are, and are not, correlated with cGMP.

The QS guidance begins by defining critical concepts of modern quality systems, including quality, QbD and product development, quality risk management, corrective and preventative action (CAPA), change control, the "quality unit," and the six-system inspection model. The discussion of the quality unit describes its relationship with the concepts of quality control (QC) and quality assurance (QA) and the relationship between the quality unit and the other units within the pharmaceutical manufacturing organization. The six-system inspection model is described as a blueprint for how compliance inspections will be organized under the new quality systems approach and should be considered a template for internal verification of compliance within pharmaceutical organizations adopting quality management systems (Figure 9).

The majority of the QS guidance is devoted to describing the essential components of modern quality systems, including four major factors which must be addressed: management responsibilities, resources, manufacturing operations, and evaluation activities. Each factor is described in detail, including aspects which overlap with cGMP regulations (for each factor there is a table listing the related regulatory citations). In particular, the manufacturing section describes aspects of quality systems (and related cGMPs) which are closely related to PAT, including raw materials analysis, operations monitoring, and procedures for addressing nonconformities. Finally, the guidance includes many important references and related guidance documents which should be considered by companies seeking to implement a quality management system.

### 4.4 Science-Based Policies

Continuous improvement, which the agency describes as an "essential element in a modern quality system," is aimed toward improving efficiency by "optimizing a process and eliminating wasted efforts in production" [1]. One of the unintended consequences of the regulatory system (prior to the new initiative) had been the suppression of nearly all opportunities for continuous improvement in manufacturing once a pharmaceutical product has been approved for market. Changes to formulations and processes needed to be justified regarding their impact on product quality, often requiring time-consuming postapproval supplements. Producers in most other modern industries (many of whom deal with public safety risks on par with or exceeding those managed by the pharmaceutical industry) make it a practice to continuously fine tune and adjust their operations to maximize quality and efficiency. Pharmaceutical manufacturers, on the other hand, have largely been constrained to treat demonstrated processes as if they were set in stone.

While there is some logic to limiting the scope and pace at which changes can be made to processes, there is obvious fallacy in the idea that the first approved configuration for a drug manufacturing operation will be optimal, especially considering the enormous financial and ethical pressures on process development teams to quickly bring new drug therapies to market. This realization spurred the agency to begin the process of developing *science-based policies and standards to facilitate innovation*, which currently includes three new updated guidance documents: "Sterile Drug Products Produced by Aseptic Processing—cGMP" [55], the PAT guidance, and the draft guidance on comparability protocols. Each guidance document encourages voluntary adoption of new technologies in pharmaceutical manufacturing by defining modern, science-based regulatory mechanisms which enable producers to implement strategic improvements with opportunities for more efficient regulatory compliance.

***Comparability Protocols*** In fact, pharmaceutical manufacturers have always had the option to explore changes to their production processes. The difference between the old regulatory paradigm and the twenty-first-century cGMPs initiative is that producers who seek to improve the quality and efficiency of their processes will be able to implement changes much more quickly while spending significantly fewer resources to maintain compliance. The key to achieving these benefits is demonstrating that there is sufficient understanding of the process and changes to be made and that implementation of the improvements poses very little risk to consumers.

A new mechanism for implementing process changes, which reflects the inclination for science-based policies, is detailed in the FDA's draft guidance "Comparability Protocols—Chemistry, Manufacturing, and Controls (CMC) Information". A *comparability protocol* (CP) is a "well-defined, detailed, written plan for assessing the effect of specific CMC changes in the identity, strength, quality, purity, and potency of a specific drug product as these factors relate to the safety and effectiveness of the product" [56]. Submission of a CP by a producer is optional and may be used to facilitate changes in a manufacturing process, analytical procedures, manufacturing equipment or facilities, or container closure systems or for implementation of PAT.

The benefit for producers submitting a CP is that, upon approval of a CP, "the FDA can designate, where appropriate, a reduced reporting category for future

reporting of CMC changes covered by the approved CP". For example, changes that otherwise would require submission, review, and acceptance of a *postapproval supplement* (PAS) might be designated as *annual report* (AR) changes if they were provided for in an approved CP. The CP is one of the mechanisms by which the FDA intends to reduce the number of supplements requiring review. Additionally, the CP was designed to facilitate free flow of communication with the agency, thereby reducing the risk that process changes will lead to unexpected regulatory shutdown or delay.

***Process Validation*** In agreement with the pursuit of science-based policies, the FDA has begun to revise the 1987 "Guideline of General Principles of Process Validation" and in March 2004 released a revision of the compliance policy guide (CPG) (Section 490.100) "Process Validation Requirements for Drug Products and Active Pharmaceutical Ingredients Subject to Pre-Market Approval" [52]. The current revisions are designed to support continuous improvement and replace the notion of "three-batch" validation. The CPG describes the concept that, after having identified and established control of all critical sources of variability, conformance batches are prepared to demonstrate that under normal conditions and operating parameters the process results in the production of acceptable product. However, the CPG does not describe how many conformance batches are required; rather, the manufacturer is expected to provide "sound rationale" for the procedure they choose to follow in demonstrating validation.

The ambiguity in the revised (CPG) regulations may seem to signify that manufacturers would need to undertake even more extensive validation exercises when in fact the CPG contains language providing a pathway for batch release to market distribution concurrent with the manufacture of initial conformance batches or with a single conformance batch [57]:

> Advanced pharmaceutical science and engineering principles and manufacturing control technologies can provide a high level of process understanding and control capability. Use of these advanced principles and control technologies can provide a high assurance of quality by continuously monitoring, evaluating, and adjusting every batch using validated in-process measurements, tests, controls, and process endpoints. For manufacturing processes developed and controlled in such a manner, it may not be necessary for a firm to manufacture multiple conformance batches prior to initial distribution.

Interpretation of the CPG suggests that implementation of PAT can be an important consideration for streamlining process validation. Finally, a quotable interpretation of the new science-based paradigm suggests that (instead of validating the process) producers should "control the process, and validate the controls." Beyond revision of the CPG, FDA is expected in the near future to release draft guidance on process validation, which will be closely aligned with concepts associated with PAT, QbD, and the rest of the 21st century cGMPs.

### 4.5 International Collaboration

Recognizing the current realities of the global marketplace, the FDA has made coordination with international regulatory partners a priority of the twenty-first-century cGMPs initiative. By increasing its collaboration with international health and regulatory partners, the FDA has been able to leverage its resources through

increased sharing of information and harmonization of activities. The International Conference on Harmonization of the Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH) (www.ich.org) has been the dominant mechanism for international cooperation among pharmaceutical regulatory authorities in Europe, Japan, and the United States.

A consensus vision statement was drafted at the July 2003 ICH meeting with regard to the objective of the ICH in harmonizing the efforts of regulatory bodies to establish quality systems approaches in their operations: "Develop a harmonized pharmaceutical quality system applicable across the life cycle of the product emphasizing an integrated approach to quality risk management and science."

Three consensus guidelines define the core of the ICH's involvement in harmonization of pharmaceutical quality systems—Q8: Pharmaceutical Development, Q9: Quality Risk Management, and Q10: Pharmaceutical Quality Systems (in addition, each of the guidance documents cites critical areas of overlap with Q6A: Specifications: Test Procedures and Acceptance Criteria for New Drug Substances and New Drug Products: Chemical Substances).

***Q8: Pharmaceutical Development*** According to the ICH Q8 guideline [58], the aim of pharmaceutical development is to "design a quality product and the manufacturing process to deliver the product in a reproducible manner." While QbD is not specifically mentioned in the guideline, the intent of the ICH Q8 expert working group (EWG) was to describe a system that would provide incentive for manufacturers to incorporate aspects of QbD and continuous improvement throughout the product life cycle. In achieving this goal, the guideline they produced describes the suggested contents for Section 3.2.P.2 of a regulatory submission in the ICH M4 common technical document (CTD) [59] and the FDA electronic common technical document (eCTD) [60].

The pharmaceutical development and quality overall summary (QOS) sections of the CTD (Figure 10) provide pharmaceutical scientists with dedicated channels to present regulators with the relevant knowledge and process understanding gathered during the development of a new product (which can be updated to support new knowledge gained over the life cycle of the product following approval). The knowledge communicated within these sections are important considerations for justification of a lower site risk potential (i.e., SRP, with regard to risk-based inspection) and for facilitation of efficient, question-based review (QbR) [61]. Question-based review is another mechanism by which the agency intends to streamline the regulatory process as well as reward producers for adopting best practices in quality management.

In addition to facilitating risk-based oversight, the content of the pharmaceutical development and QOS sections of the CTD are critical to enabling continuous improvement and flexible operation. The information and knowledge communicated within these sections provide scientific understanding to support the establishment of a manufacturing design space, in-process and release specifications, and manufacturing controls.

As described within the Q8 guideline, a design space is the "multidimensional combination and interaction of input variables and process parameters that have been demonstrated to provide assurance of quality." So long as process control is maintained within the bounds of the design space, operating parameters can be adjusted to improve product quality or manufacturing efficiency. Based on the

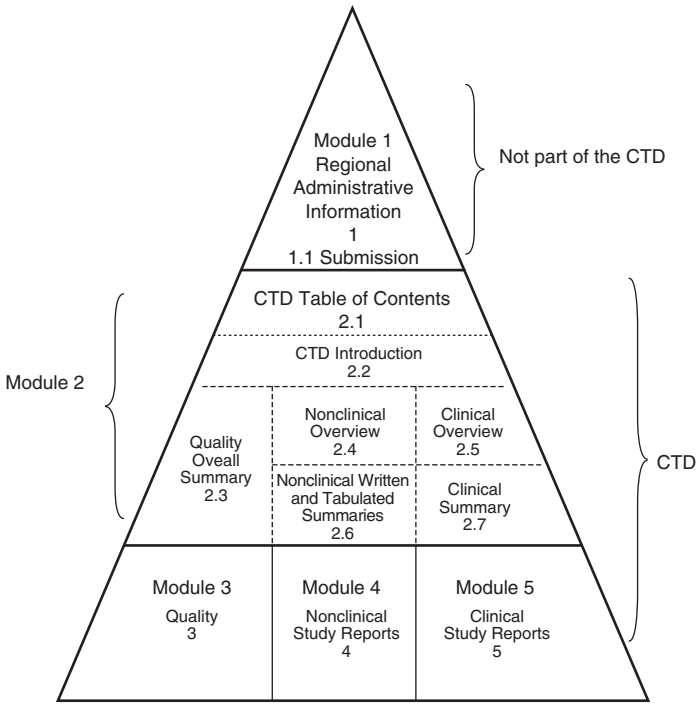


**FIGURE 10** Schematic illustration of the ICH M4 common technical document (CTD); the contents of the Quality Overall Summary (2.3) and Quality (3) modules are most relative to PAT.

current definition, operation outside of the established design space would initiate a regulatory postapproval change process. Thus, complete and accurate communication of the knowledge supporting a company's design space is vital for a company to maximize productivity while maintaining regulatory compliance. Furthermore, with the new communication pathways in place, companies have incentive to pursue manufacturing studies beyond marketing approval to expand their design space or to update specifications and controls. In addition to the product under review, if appropriate, experiences gained from the development (and manufacture) of similar drug products may be included.

***Q9: Quality Risk Management*** The second working group (ICH Q9 EWG) is trying to better define the principles by which risk management will be integrated into decisions by regulators and industry regarding quality, including cGMP compliance. In November 2005, the Q9 EWG released the "Step 4" version of the Q9 guideline which defines the two primary principles of quality risk management, provides a model for the quality risk management process (Figure 11), and describes the terminology and tools for risk assessment and management. In addition, the document includes a concise reference list for more detailed information on risk management methods, such as failure mode effect and criticality analysis (FMECA), which are important tools for prioritized implementation of PAT. While it is not intended to be a "how to" manual for risk management, the Q9 guideline is a valu-

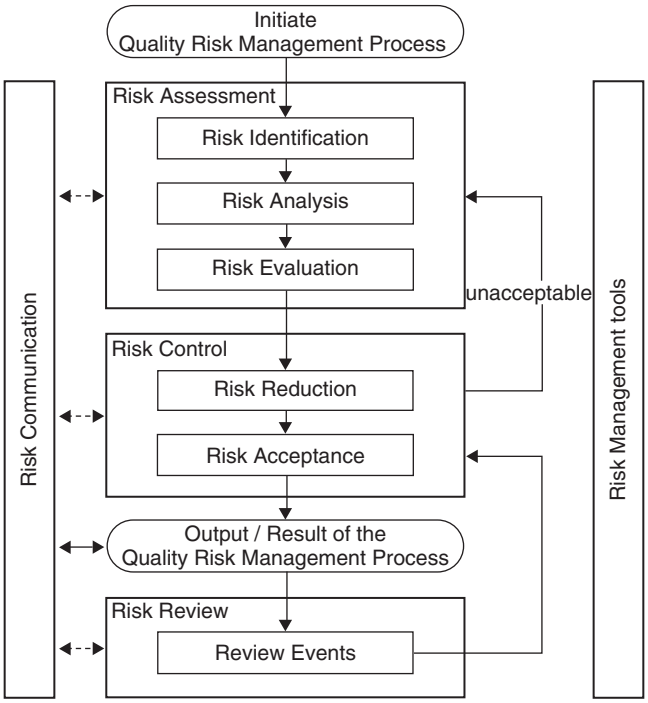


**FIGURE 11** Schematic of quality risk management process described within ICH Q9.

able information source for companies seeking to incorporate quality risk management into their operations [62].

***Q10: Pharmaceutical Quality Systems*** While the Step 2 document for the third tripartite guideline, Q10: Pharmaceutical Quality Systems, has not yet been released, the final concept paper has been available since 2005 [63]. Similar to the manner by which the FDA's quality systems approach guidance mapped the relationship between cGMPs and other industrial quality management systems, the Q10 guideline is anticipated to serve as a bridge between the approaches to quality systems taken by the different regional regulations, thereby helping to achieve global harmonization of quality systems. The guideline is expected to strengthen and complement issues covered in Q6A, Q8, and Q9 and will provide a foundation for a pharmaceutical quality system based on elements from the ISO 9001 and 9004 standards. The guideline is also expected to develop harmonized definitions for issues critical to PAT, including continuous improvement activities, data-gathering methods, and the approach to measurement system validation.

## 5 PAT EVOLUTION IN PHARMACEUTICAL MANUFACTURING

Though it may be tempting to characterize PAT as a revolutionary change in pharmaceutical manufacturing, history will likely show that the beginning of the twenty-first-century cGMPs initiative and the development of the PAT guidance mark the

beginning of a period of rapid evolution in pharmaceutical manufacturing which will extend far into the future. Even though the twenty-first-century cGMPs initiative is more extensive (with regard to changing the relationship between the FDA and the pharmaceutical industry), interest in the PAT guidance and the opportunities it presents for the industry were initially much greater. More recently, perhaps in parallel with some changes in leadership in the agency, there has been a palpable shift of emphasis toward QbD, which was barely mentioned in many of the twenty-first-century cGMPs documents. It is important to keep in mind that, just as most industries have seen a parade of "new" quality systems initiatives over the years since Shewhart's first methods were published, the principles upon which PAT and QbD are built, such as robust process design, quality monitoring, and effective controls, will persist regardless of the name of the initiative. Furthermore, as with PAT, QbD is not a new concept. Indeed, Dr. Genichi Taguchi, who has been credited by some as the father of QbD, began applying QbD in pharmaceutical manufacturing while working as a statistical consultant for Morinaga Pharmaceuticals Company of Japan from 1947–1949 [25].

The PAT guidance is unique when compared with typical FDA guidance documents in that it is not instructive or limiting per se; rather, the guidance describes the principles and tools upon which the PAT *framework* is built, with the goal of "highlighting opportunities and developing regulatory processes that encourage innovation." The FDA's goal in developing the PAT guidance was to eliminate the specter of regulatory uncertainty which has been identified as a major factor limiting innovation in pharmaceutical manufacturing. The guidance works with existing regulations and was designed to be consistent with the agency's twenty-first-century cGMPs initiative. Furthermore, the guidance emphasizes that the decision on the part of manufacturers to work with the agency to implement PAT is voluntary. Since the guidance is not prescriptive in nature, it neither describes "how to do PAT" nor identifies any particular practice or technology as "approved for PAT."

### 5.1 Process Understanding

The agency considers PAT to be a "system for designing, analyzing, and controlling manufacturing through timely measurements of critical quality and performance attributes of raw and in-process materials and processes, with the goal of ensuring final product quality." Based on this definition, it would be practical to consider PAT to be an expansion of PAC; PAT builds on the measurement and control aspects of PAC by incorporating additional emphasis on QbD and process understanding.

According to the PAT guidance, a process is generally considered well understood when:

1. All critical sources of variability are identified and explained.
2. Variability is managed by the process.
3. Product quality attributes can be accurately and reliably predicted over the design space established for materials used, process parameters, manufacturing, environmental, and other conditions.

Furthermore, according to the guidance, the ability to predict "reflects a high degree of process understanding."

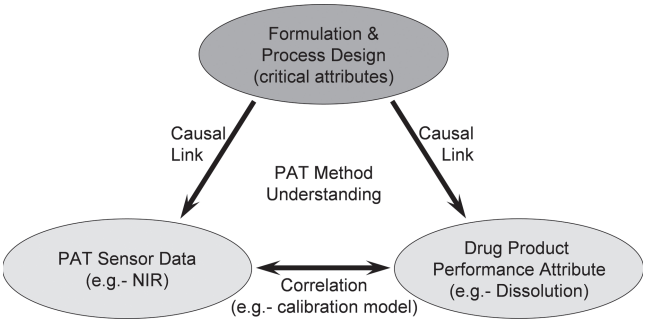


**FIGURE 12** Illustration of aspects of method understanding which must be in place to justify product performance measurements using indirect and/or nondestructive analyses.

Possession of a predictive model (for product quality attributes) alone does not necessarily constitute process understanding, however. A relatively common example would be prediction of material or product performance characteristics using multivariate measurements, such as prediction of tablet dissolution rate using NIR spectroscopy. Multiple researchers have demonstrated that (in some cases) it is possible to predict drug release from tablets in vitro using nondestructive NIR spectra by generating a calibration model for dissolution rate. Without demonstrating at least mechanistic understanding of the physicochemical feature (correlated to dissolution rate) being detected by NIR, the calibration model would constitute nothing more than pattern recognition (Figure 12) [64]. While such a calibration may be useful, without greater insight as to the basis for correlation, it would not likely be a useful demonstration of process understanding.

***Design Space and Quality by Design*** The concept of a multidimensional space of acceptable operating conditions, or design space, is perhaps one of the most important aspects of the twenty-first-century cGMPs which facilitates continuous improvement. In a PAT-enabled environment, the process design space must provide evidence of QbD [65] and should be the mathematical medium by which process understanding and real-time control decisions are communicated (Figure 13).

The current ICH Q8 definition of design space, unfortunately, offers little guidance with regard to the aspects of a process design space which are required for implementation. As a result, a variety of interpretations of what constitutes a suitable process design space have recently surfaced among industry participants. One of the most popular misconceptions is that an effective design space for a process or unit operation can be determined by the common trajectory of PAT measurements (i.e., "process signature") related to product batches known to have acceptable quality (i.e., "golden path"). While such data are useful for monitoring, they are nothing more than a modern version of "3-batch" process validation. Golden paths or process trajectories are not sufficient for control since 1) the path itself is not necessarily predictive and 2) such controls would imply that a process is limited by its historical path in the space of process parameters. Originally, the term process signature was defined as a multivariate process measurement, that is, NIR spectrum, which contained features useful for describing the impact of the process on the chemical and physical aspects of the processed material [38].

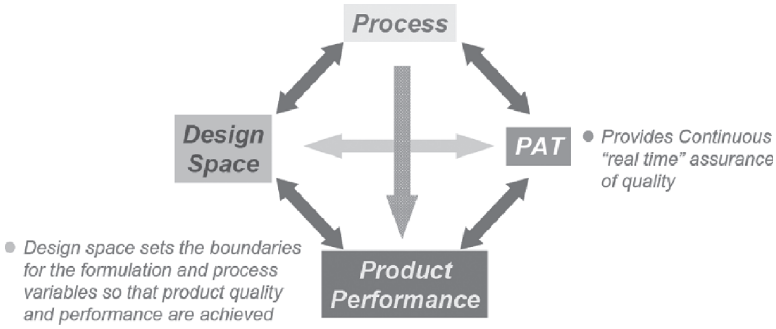


**FIGURE 13** Interrelation between design space, PAT, and process control in a manufacturing system based on quality-by-design. (*Source*: R. C. Lyon, Process monitoring of pilot-scale pharmaceutical blends by near-infrared chemical imaging and spectroscopy, Eastern Analytical Symposium (EAS), Somerset, NJ, 2006.)

While it is perhaps too early to posit a conclusive standard for pharmaceutical process design space development, the following minimum criteria should be achieved for a process design space to be suitable for process control:

- The process design space should be expressed in the form of a mathematical model which *quantitatively* links *process capability*, quality of input materials, and process operating parameters.
- Relevant critical-to-quality product attributes should be considered by the design space model (e.g., content uniformity, bioavailability, stability).
- Borrowing from a famous quote by Albert Einstein, the (design space) model should be as complex as necessary (for accurate prediction), but no less.
- Product attributes that are superfluous or are not known to be critical to quality should not be considered by the design space model (there should not be a penalty for monitoring such parameters, however).
- In the same way that in vitro–in vivo correlation (IVIVC) is required to be granted a biowaiver for implementation of postapproval changes, the ability of the design space model to predict the quality of finished goods must be validated prior to implementation.
- If the accuracy of the design space model cannot be established a priori with statistical significance within portions of the parameter hyperspace, operation in such regimes should initiate supplementary quality assurance (inspection) activities until the design space model can be updated and revalidated.
- If unacceptable product quality is observed during operation within a region of the design space expected to yield acceptable quality, the design space should be considered unsuitable for process control (due to drift or the appearance of new factors in the parameter space) until the missing factor(s) can be identified and incorporated into the model and the model is revalidated.

If such a *model-based process design space* includes a sufficient portion of the factors affecting product quality variance, the process *control space* can be projected to define the bounds of normal operation. Based on this definition, the control

model algorithm for each operation in the manufacturing process would be generated from a subset of the control space spanned by the material qualities and processing parameters which impact that operation. Each unit operation control model seeks to adjust process parameters in a timely manner in response to changes in raw material (feedforward) or finished-product (feedback) quality. In other words, *control* the process and *validate* the controls.

The mathematical linkage of the design space, process, and control models enables continuous optimization of product quality by seeking the optimal point within the control space. As the level of process understanding increases or as processing conditions evolve, factors might be added or removed from the design space and the process and control models updated. Furthermore, by considering other factors such as yield, efficiency, or C/T as a function of the variables spanned by the process design space, the process might be co-optimized for quality and profitability.

It is likely that many pharmaceutical manufacturing operations are not understood in a way that product quality variance can be fully described in functional form (e.g., transfer functions); attaining such a level of manufacturing knowledge should be a goal for the industry. Using functional representations of process understanding as the basis set for a process design space, rather than historical performance, offers many operational advantages:

- *Efficient Process Development* While the current definition of design space does not preclude the incorporation of knowledge from other products and processes, model-based knowledge representation offers a more robust framework for incorporation of external or *a priori* information. Even though the level of quality expected by a particular combination of input and process parameters from another product is not likely to transfer to a new product or process (in absolute terms), the functional relationships which predict quality may be quite similar. Furthermore, model-based design space development enables direct incorporation of first principles and mechanistic knowledge, which might significantly reduce the complexity of experimental designs required for process development since significant terms may be identified in silico.
- *Quality by Design* The incorporation of functional relationships between inputs, parameters, and product quality (or efficiency), which inherently imply magnitude and directionality, enables the use of a process design space as a tool for multiobjective process optimization. Furthermore, the model-based representation of knowledge is compatible with concepts of risk management, enabling more flexible operation since the risk associated with extrapolation could be predicted.
- *Control System Development* Model-based design space development offers an ideal segue between process and control development. Quite literally, a model-based design space would provide the template for development of feedforward process control models. Moreover, development of a process design space using a model-based framework would facilitate control system validation and identification of science-based, in-process, and release specifications.

- *Scaling and Technology Transfer* Within the current system for process development, it is common to use designed experiments (i.e., DOE) where some input variables are product specific (e.g., excipient "grade") or process parameters are device dependent (e.g., chopper speed, damper angle). In a model-based paradigm, however, a process design space would ideally be generated using product- and device-independent units which have more basic physical meaning (e.g., modulus, viscosity, energy, or work). Designing and describing production processes in fundamental terms or, perhaps, standardized dimensionless units would facilitate scaling and transfer of design space and process control models to similar manufacturing processes that are based on the same physical operating principles.

Academic research is currently underway to further develop the model-based design space concept. Working within the limits of the current system, though, producers who are able to demonstrate process understanding or are willing to invest in a PAT system to facilitate their development of process understanding can use the tools and provisions of the framework to pursue innovation and continuous improvement with more efficient regulatory oversight (i.e., the ability to make changes without supplemental review). The PAT framework is described as consisting of two components: (1) a set of scientific principles and tools supporting innovation and (2) a strategy for regulatory implementation that will accommodate innovation. The following paragraphs will describe selected aspects of both components in detail.

### 5.2 PAT Principles and Tools

Central to the PAT framework is the acceptance that certain physical and mechanical attributes of pharmaceutical ingredients are not necessarily well understood and that even processes which have achieved significant process understanding are subject to a finite level of stochastic variation. Thus, the core of the PAT guidance is allocated to describing the principles and tools, such as process analyzers and risk analysis, which producers can employ to augment process understanding and mitigate latent risks to product quality.

***PAT Tools*** The guidance describes four categories of PAT tools:

- Multivariate tools for design, data acquisition, and analysis
- Process analyzers
- Process control tools
- Continuous improvement and knowledge management tools

Since each of the four categories draws upon methods and technology which are already established in other fields such as PAC, the discussion of each category within the guidance is focused on aspects which are unique or significant to pharmaceutical manufacturing, such as *process signature* [2]. Furthermore, in keeping with the spirit of the framework as a catalyst for innovation, the agency made an effort to avoid mention of any particular tool or technology in the final version of

the PAT guidance. The PAT tools section of the guidance does, however, include cross-references to relevant portions of current regulations which should be considered by a manufacturer developing a PAT strategy or system.

***Standards for Pharmaceutical Applications of PAT*** During the early stages of developing the PAT framework, the agency was aware that the lack of international standards was a significant impediment to regulatory coordination and implementation of PAT in the global pharmaceutical industry. In 2003, the FDA's PAT team worked with ASTM International to form Technical Committee E55 on Pharmaceutical Application of Process Analytical Technology. The E55 committee addresses issues related to process control, design, and performance as well as quality acceptance/assurance tests for the pharmaceutical manufacturing industry. Stakeholders in the committee include manufacturers of pharmaceuticals and pharmaceutical equipment, federal agencies, design professionals, professional societies, trade associations, financial organizations, and academia (www.ASTM.org).

As of mid-2006, there were three subcommittees of E55: PAT system management, PAT system implementation and practice, and PAT terminology. The PAT team has been represented on E55 committees with a goal to ensure that standards developed are aligned with the PAT guidance and acceptable to the FDA. To date, one active standard has been published, while 16 additional standards have been proposed. The ASTM International provides another venue for international cooperation (consistent with the twenty-first-century cGMPs initiative); the definitions of PAT (in the FDA guidance and ASTM E55) as well as other concepts are being incorporated into the ICH Q8 guidance.

***Real-Time Release (RTR)*** The PAT guidance defines RTR as "the ability to evaluate and ensure the acceptable quality of in-process and/or final product based on process data." Whereas finished products are typically released for marketing only after sampling, inspection (i.e., laboratory-based QC testing), and review, implementation of an RTR system enables release of finished products concurrent with the completion of manufacturing operations. Practically speaking, RTR is one of the most significant, tangible benefits for producers who implement PAT, because it can facilitate dramatic reductions in process C/T.

Real-time release is considered by the guidance to be comparable to *alternative analytical procedures* for final product release and is defined within the guidance as an extension of parametric release. The defining characteristic of RTR is that it considers simultaneously the degree to which material attributes and process parameters are measured and controlled during manufacturing. It was not intended that RTR be implemented by simply installing a rapid measurement system at the end of a manufacturing process; such uses for PAT tools would be tantamount to inspection and would do nothing to improve quality management.

The guidance does suggest, however, that it may be feasible to implement RTR without finished-product quality monitoring by using "the combined process measurements and other test data gathered during the manufacturing process." Similar language is found in the USP general notices, where it is suggested that data derived from "[validation studies and] in-process controls may provide greater assurance that a batch meets a particular monograph requirement than analytical data derived from an examination of finished units drawn from that batch." It would not be

difficult to create a system more capable of detecting quality variation than current methods based on inspection. Recent statistical analyses [42] have demonstrated that, for determining batch quality, the traditional USP ⟨905⟩ method of content uniformity testing may indeed have little more statistical power than a coin toss until more than 5% of the product exceeds specification limits (corresponds to within-batch $C_{pk}$ of approximately 0.65, only slightly worse than has been observed in a recent industry benchmarking study [28]).

On the other hand, deployment of an RTR system without finished-product monitoring would require the manufacturer to demonstrate a very high level of process understanding based on, for example, their development of a comprehensive design space and/or a well-validated process model. Even though it may be feasible to implement RTR without end-of-process monitoring, a well-designed PAT system will typically include some form of final product quality monitoring as a means for mitigating latent risk and creating strategic redundancy in process controls and as an additional tool to bolster process understanding.

### 5.3 Strategy for Implementation

One of the FDA's goals for the PAT guidance is to "tailor the Agency's usual regulatory scrutiny to meet the needs of PAT-based innovations that (1) improve the scientific basis for establishing regulatory specifications, (2) promote continuous improvement, and (3) improve manufacturing while maintaining or improving the current level of product quality." Recognizing that the achievement of this goal requires a unique interface between regulators and manufacturers seeking to implement PAT, a strategy for implementation based on the integrated systems approach was developed. An objective of the strategy for implementation is to facilitate clear, effective, and meaningful communication between the agency and industry, for example, in the form of meetings or informal communication.

In practice, the strategy breaks with traditional industry–FDA modes of communication; whenever PAT is concerned, it is anticipated that regulators will communicate directly with the pharmaceutical scientists and engineers involved with development and operation of the PAT system rather than indirectly via a department of regulatory affairs. The components of the agency's regulatory strategy include:

- A PAT team approach for CMC review and cGMP inspections
- Joint training and certification of PAT review, inspection, and compliance staff
- Scientific and technical support for the PAT review, inspection, and compliance staff
- Recommendations provided within the PAT guidance

***PAT Team Approach*** FDA's assembly of the PAT team was one of the most significant incentives for the industry to pursue manufacturing innovation as described in the twenty-first-century cGMPs initiative and the PAT guidance. The PAT team was put in place to ensure that industrial PAT applications were handled with expediency and accuracy by scientists familiar with the most up-to-date PAT methods.

At one point the PAT team included more than 20 scientists, including investigators, compliance officers, reviewers, training coordinators, and a policy development team. More recently the agency has begun steps to "sunset" the PAT team, the duties of which will ultimately be handled by FDA staff trained in PAT systems. A comprehensive scientific training program was developed for the PAT team with guidance from the ACPS PAT subcommittee. Initial training began in January 2006, with plans for further training to be provided by faculty at Duquesne and Delaware Universities [47].

***Research Data Provision*** In developing the PAT guidance, the FDA recognized that, even with the guidance in place, manufacturers seeking to evaluate the suitability or potential value of new technologies for process control may be hesitant, figuring that such data will be subject to cGMP inspection, thereby increasing their liability with respect to regulatory actions. To allay these fears, the agency included a statement which applies to investigational deployment of new technologies [2]:

> Data collected using an experimental tool should be considered research data. If research is conducted in a production facility, it should be under the facility's own quality system. . . . FDA does not intend to inspect research data collected on an existing product for the purpose of evaluating the suitability of an experimental process analyzer or other PAT tool. FDA's routine inspection of a firm's manufacturing process that incorporates a PAT tool for research purposes will be based on current regulatory standards (e.g., test results from currently approved or acceptable regulatory methods). Any FDA decision to inspect research data would be based on exceptional situations similar to those outlined in Compliance Policy Guide Sec. 130.300. Those data used to support validation or regulatory submissions will be subject to inspection in the usual manner.

## 6 PAT IMPLEMENTATION PROCESS

The PAT guidance identifies three possible plans for companies seeking to implement PAT:

- PAT can be implemented under the facility's own quality system; cGMP inspections by the PAT team or PAT-certified investigator can precede or follow PAT implementation.
- A changes being effected (CBE), CBE in 30 days (CBE-30), or prior approval (PAS) supplement can be submitted to the agency prior to implementation, and, if necessary, an inspection can be performed by a PAT team or PAT-certified investigator before implementation.
- A comparability protocol (CP) can be submitted to the agency outlining PAT research, validation, and implementation strategies and time lines. Following approval of this comparability protocol by the agency, one or a combination of the above regulatory pathways can be adopted for implementation.

Reflecting its nonprescriptive nature, the three implementation plans are essentially the only "how to" portions of the PAT guidance. This leaves industrial (and academic) scientists and engineers with the burden of determining how best to proceed

in the deployment of a PAT system. Despite the fact that some pioneering companies have been incorporating aspects of PAT in their operations since long before the start of the FDA's twenty-first-century cGMPs initiative, there continues to be significant diversity in their approaches to implementation. While perhaps the ambiguity (in how best to proceed) has slowed the uptake of PAT to some degree, in the long run, the latitude is preferable since the optimal path of implementation will likely be unique for most facilities.

With regard to drug manufacturers' implementation of PAT, a list of 10 questions has been presented which provides an initial checklist for companies seeking approval of their plans [10, 66]:

1. Is this a PAT system?
2. Does it have aspects of design, measurement, and manufacturing control?
3. Are PAT principles and tools used?
4. Which tools specifically are used for manufacturing control?
5. How are continuous improvement and knowledge management performed?
6. What risk-based approach has the company taken—assessment, prevention, and management?
7. How are the PAT systems integrated?
8. What kind of RTR is being proposed or used?
9. What regulatory process is being considered?
   a. Can the companies' quality systems manage the PAT change?
   b. Are the submission proposals appropriate and justified?
10. What are the critical aspects that will be evaluated during site visits/inspections?

Drawing from aspects of the DMAIC model, as well as the risk-based orientation and quality systems approach espoused by the FDA's twenty-first-century cGMPs initiative, the Duquesne University Center for Pharmaceutical Technology (DCPT) has proposed a six-phase, iterative cycle for process improvement based on PAT (Figure 14). While there are certainly many acceptable variants of this strategy, some of which have begun to appear in conferences and the industrial literature, any successful PAT deployment, large or small, will most likely include some combination of these elements. In addition, each project phase will necessarily include one or more modules of training. Finally, while the project phases are presented as being discrete, most of the phases will overlap to some degree. In particular, consideration of the objectives for control, release strategies, and plans for continuous improvement should begin, along with management buy-in, early in the cycle.

### 6.1 Preparation

The preparation phase is arguably the most critical step in the path toward PAT implementation. Process analytical technology projects are inherently multidisciplinary, requiring acceptance and buy-in from corporate divisions which sometimes

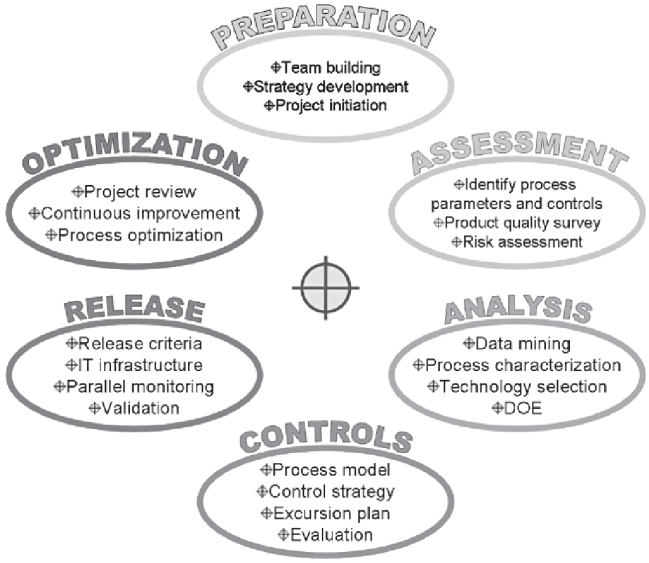


**FIGURE 14** PAT implementation cycle with examples of associated activities for each phase.

operate with rather divergent goals and procedures. Most importantly, those who are seeking to initiate a PAT project will need to obtain management buy-in at a level high enough in the corporate structure to ensure sufficient resources will be available and that the company will be committed to positive change. During the preparation phase, a PAT team having a diverse background and critical skills should be assembled, and formal planning of the project should begin, including selection of the product and process to address. Ideally, dialogue with the FDA PAT team should begin early in the preparation phase.

### 6.2 Assessment

The PAT guidance clearly states that industrial implementations should be risk based. Soon after the PAT team and objective have been identified, the project should commence with a formal risk assessment. The risk assessment should be focused on identifying and characterizing the failure modes which present risks to product quality; the outcome of the risk assessment will provide a means prioritizing the allocation of PAT resources and a baseline for review of the effect of PAT in mitigating risks to quality.

### 6.3 Analyze

The "analyze" phase of the project consists of the activities which are typically associated with PAC, including identification and assessment of potential sensor technologies, method development, qualification, and validation. In addition, designed experiments (DOE) or data-mining exercises may be performed to

generate process understanding or to support PAT goals. Plans for the IT infrastructure, sampling protocols, and development of controls should also be considered.

### 6.4 Control

The implementation of controls begins as each new analytical method or technology is deployed. Controls may be as simple as automated termination of a unit operation upon reaching an endpoint. With greater process understanding, more complex controls can be deployed, including feedback (e.g., control of punch force during tablet compaction, control of temperature or airflow during fluid bed processing) or feedforward controls (e.g., adjustment of process parameters based on incoming raw-material quality). The development and implementation of controls should also consider operating procedures for adverse situation management and should initiate a reassessment of risk to determine the suitability of controls.

### 6.5 Release Philosophy

For PAT projects including implementation of RTR or some modification of a preexisting release mechanism for an approved process, additional method development and validation procedures will be required. The real-time release decision will typically be determined by a process model, which can be a mathematical equation or algorithm within the control system; furthermore, the IT system must accurately convey the release decision and supporting data to downstream operations (i.e., warehouse, logistics), upstream operations (i.e., production scheduling, accounting), or the facility information repository. The interwoven IT and scientific components require an integrated systems approach to development, validation, deployment, and operation. Finally, implementation of PAT systems enables redefinition of product quality acceptance criteria for release; the task of identifying robust release criteria suitable for large sample sizes, for example, continues to merit examination [43].

### 6.6 Optimization

The optimization phase of the project provides an opportunity to assess the performance of the PAT system relative to the goals of the project as well as the level of latent risk in the system. Ideally, with the PAT s2ystem in place, the level of process understanding will be improving as more data are collected for every batch. The added insight into the operation may yield new opportunities for improving quality or efficiency or for solving similar problems with another product. The key to success in the optimization stage is realizing that it is only the beginning of continuous improvement.

## 7 PERSPECTIVES ON THE IMPACT OF PAT

PAT and the twenty-first-century cGMPs initiative have clearly made an impact within the pharmaceutical and associated industries. Significant sums of capital are now flowing in new directions to meet the challenges and opportunities pre-

sented by the changes. Some people within the industry, however, question whether there will be much of a long-term impact, citing the litany of new eras in the industry (and their careers) that turned out to be more of the same. With just a bit of observation, though, it is not hard to see that it really is different this time.

The modern pharmaceutical manufacturing industry finds itself in a difficult situation that perhaps few anticipated just 10 or 15 years ago. The rate of new blockbuster drug approvals has continued to wane, while new drug therapies become inexorably more expensive to discover and develop. Despite the fact that the market for drug sales has never been larger, drug company profit margins are shrinking while consumers, feeling that pharmaceutical company profits are unjust, have reached new lows in their opinion of the industry. A recent survey by the Kaiser Family Foundation placed pharmaceutical companies just above oil and tobacco companies, and right below health management organizations (HMOs), in terms of public opinion [67]. Entities of significant magnitude in both the public and private sector are increasingly applying pressure to capture an even greater portion of the industry's compensation. Indeed, there is no shortage of industrial and financial publications which have chronicled the pharmaceutical industry's troubles [34, 45, 68, 69].

The pharmaceutical industry is fortunate, perhaps, to follow (rather than lead) most other major industries in adopting truly automated controls, process analytics, quality management, and lean manufacturing. The performance of pharmaceutical companies relative to the benchmarks for world-class manufacturers provides a roadmap for improvement. If the pharmaceutical industry, as a whole, were able to at least approach the benchmarks for world-class manufacturing performance (by implementing PAT), the savings returned to consumers and shareholders would be immense (Figure 15). Finally, the returns on investment in PAT are not limited to major producers. Estimates based on recent benchmarks suggest that, by successfully transforming operations through the deployment of PAT and lean, a typical small or mid-sized pharmaceutical manufacturer could improve operating margins by up to 600 basis points [44].

Forces which are out of the industry's control are providing more reasons than ever before to seek efficiency in pharmaceutical manufacturing, and the FDA is doing its part to clear the way. While the pharmaceutical industry has likely been unjustly cast as a culprit behind America's fiscal crisis in health care, the industry has ample opportunity to change for the benefit of patients as well as investors.

## ACKNOWLEDGMENTS

The author would like to thank the following reviewers for their input, which was essential to the quality of this manuscript: James K. Drennen, III, Ph.D., Director, Duquesne University Center for Pharmaceutical Technology, Senior Consultant, Strategic Process Control Technologies; Robbe C. Lyon, Ph.D., Deputy Director, Division of Product Quality Research FDA/CDER; D. Christopher Watts, Ph.D., Team Leader, Standards & Technology, FDA/CDER/OPS; and Tom Knight, Founder & Chief Strategy Officer, Invistics Corp.

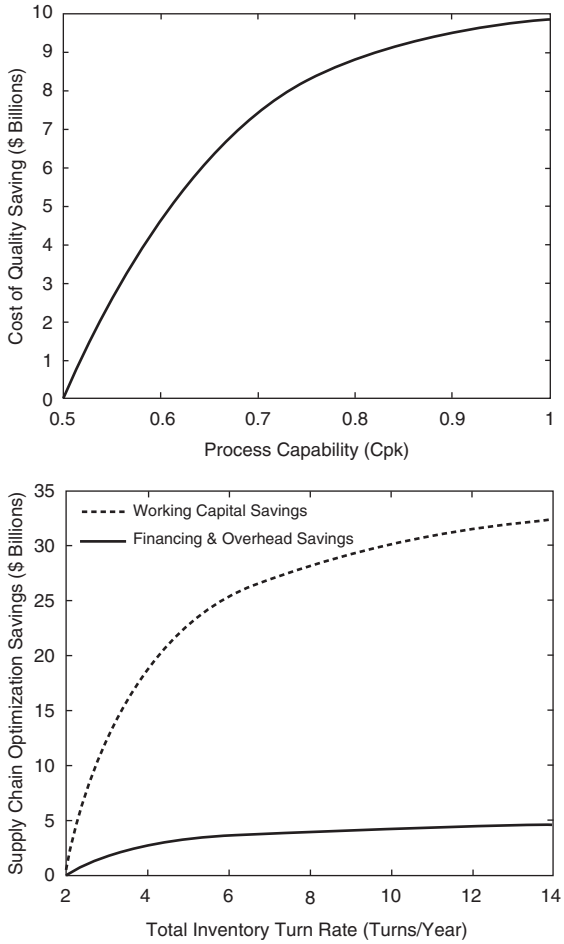


**FIGURE 15** Potential financial returns from deployment of PAT and lean. The curves are calculated based on the aggregate COGS and inventories reported in the 2005 annual reports of the top 16 branded and generic pharmaceutical manufacturers (according to market capitalization). It is important to keep in mind that working capital savings are a one-time-only benefit, while cost of quality and inventory financing and overhead savings represent on-going returns on investment. Furthermore, while the curves may overestimate savings because of innacuracies in benchmark data or the limits on the opportunities for PAT implementation, they do not account for numerous other potential pathways for returns from PAT such as capacity increase, labor productivity enhancement, reduction of QC expense, or decreased time to market.

## REFERENCES


1. U.S. Department of Health and Human Services (2004), Pharmaceutical CGMPs for the 21st century—A risk-based approach, Final report, Food and Drug Administration, Rockville, MD.
2. U.S. Department of Health and Human Services (2004), Guidance for industry: PAT—A framework for innovative pharmaceutical development, manufacturing, and quality assurance, Food and Drug Administration, Rockville, MD.
3. Clevett, K. J. (1986), *Process Analyzer Technology*, Wiley, New York.
4. Illman, D. L. (1986), CPAC: An industry—university cooperative research center for process analytical chemistry, *TrAC Trends Anal. Chem.*, 5, 164.
5. Callis, J. B., Illman, D. L., and Kowalski, B. R. (1987), Process analytical chemistry, *Anal. Chem.*, 59, 624A.
6. Balboni, M. L. (2003), Process analytical technology: concepts and principles, *Pharm. Technol*.
7. Koch, M. V. (2006), Optimizing the impact of developments in micro-instrumentation on process analytical technology: a consortium approach, *Anal. Bioanal. Chem.*, 384, 1049.
8. Küppers, S., and Haider, M. (2003), Process analytical chemistry—future trends in industry, *Anal. Bioanal. Chem.*, 376, 313.
9. Küppers, S., and Haider, M. (2006), Process analytical chemistry, *Anal. Bioanal. Chem.*, 384, 1034.
10. Hinz, D. C. (2006), Process analytical technologies in the pharmaceutical industry: the FDA's PAT initiative, *Anal. Bioanal. Chem.*, 384, 1036.
11. Beebe, K. R., et al., (1993), Process analytical chemistry, *Anal. Chem.*, 65, 199R.
12. Blaser, W. W., et al., (1995), Process analytical chemistry, *Anal. Chem.*, 67, 47R.
13. Workman, J. J., et al., (1999), Process analytical chemistry, *Anal. Chem.*, 71, 121R.
14. Workman, J. J., et al., (2001), Process analytical chemistry, *Anal. Chem.*, 73, 2705.
15. Workman, J. J., Koch, M. V., and Veltkamp, D. J. (2003), Process analytical chemistry, *Anal. Chem.*, 75, 2859.
16. Workman, J. J., Koch, M. V., and Veltkamp, D. J. (2005), Process analytical chemistry, *Anal. Chem.*, 77, 3789.
17. Gluckman, P., Roome, D. R., Deming, W. E., and Delavigne, K. (1993), *Everyday Heroes of the Quality Movement: From Taylor to Deming, the Journey to Higher Productivity*, 2nd ed., Dorset house, New York.
18. Shewhart, W. A. (1980), *Economic Control of Quality of Manufactured Product*, ASQC Quality Press, Milwaukee, WI.
19. Shewhart, W. A. (1986), in Deming W. E. ed., *Statistical Method from the Viewpoint of Quality Control*, Dover, New York.
20. Scherkenbach, W. W. (1991), *The Deming Route to Quality and Productivity: Road Maps and Roadblocks*, ASQC Quality Press, Milwaukee, WI.
21. Juran, J. M. (2003), *Architect of Quality: The Autobiography of Dr. Joseph M. Juran*, McGraw-Hill, New York.
22. Bhote, K. R. (2002), What is Six Sigma? *The Ultimate Six Sigma*. Vol 1. New York: AMACOM; 9–14.
23. Galvin, R.W. (2002), Forward. *The Ultimate Six Sigma*. Vol 1. New York: AMACOM; xxi–xxii.
24. Marash, S. A., Berman, P., and Flynn, M. (2004), *Fusion Management: Harnessing the Power of Six Sigma, Lean, ISO 9001:2000, Malcom Baldridge, TQM and Other Quality Breakthroughs of the Past Century*, QSU Pub., Fairfax, VA.

25. Ealey, L.A. (1988), *Quality by design. Taguchi Methods and U.S. industry*. Dearborn, MI: ASI Press.
26. Taguchi, G., Chowdhury, S., and Wu, Y. (2005), *Taguchi's quality engineering handbook*. Hoboken, NJ: John Wiley & Sons.
27. e-Handbook of Statistical Methods. *NIST/SEMATECH*. Available at: http://www.itl.nist.gov/div898/handbook/.
28. Macher, J., and Nickerson, J. (2006), Pharmaceutical manufacturing research project: Final benchmarking report, Georgetown University, McDonough School of Business and Washington University Olin School of Business, Washington DC.
29. Levinson, W. A. (2002), *Henry Ford's Lean Vision*, Productivity Press, New York.
30. Leiper, K. (2006), paper presented at the The Heidelberg PAT Conference, Heidelberg, Germany.
31. Lewis, N. A. (2006), A tracking tool for lean solid-dose manufacturing, *Pharm. Technol.* 30, 94–108.
32. Roumeliotis, G. (2006), Lean proves mean in drug manufacturing, available: in-PharmaTechnologist.com.
33. Gerecke, G., and Knight, T., Improving performance and reducing cycle time using flow path management: A case study, *Pharm. Eng.* 21.
34. Abboud, L., and Hensley, S. (2003), Factory shift: New prescription for drug makers: Update the plants, *Wall Street J.*, Sept. 3, p. 1.
35. Cooley, R. E., and Egan, J. C. (2004), The impact of process analytical technology (PAT) on pharmaceutical manufacturing, *Am. Pharm. Rev.*, 7, 62–68.
36. Cogdill, R. P. (2006), in Brittain, H. G., Ed., *Spectroscopy of Pharmaceutical Solids*, Taylor & Francis Group, New York, pp. 313–412.
37. Cogdill, R. P., et al., (2005), Process analytical technology case study, Part II: Development and validation of quantitative for tablet API content and hardness, *AAPS Pharm. Sci. Tech.*, 6, Article 38.
38. Cogdill, R. P., et al., (2005), Process analytical technology case study, Part I: Feasibility studies for quantitative NIR method development, *AAPS Pharm. Sci. Tech.*, 6, Article 37.
39. Cogdill, R. P., Anderson, C. A., and Drennen, J. K. (2005), Process analytical technology case study, Part III: Calibration monitoring and transfer, *AAPS Pharm. Sci. Tech.*, 6, Article 39.
40. Lorber, A. (1986), Error propagation and figures of merit for quantification by solving matrix equations, *Anal. Chem.*, 58, 1167.
41. Braga, J. W. B., and Poppi, R. J. (2004), Figures of merit for the determination of the polymorphic purity of carbamazepine by infrared spectroscopy and multivariate calibration, *J. Pharm. Sci.*, 93, 2124.
42. Lunney, P., and Drennen, J. K. I. (2005), A prevention based strategy for quality control using PAT, *NIR News*, 16, 7.
43. Sandell, D., Diener, M., Vukovinsky, K., Hofer, J., and Pazdan, J. (2006), Development of a content uniformity test suitable for large sample sizes, *Drug. Info. J.*, 40, 337.
44. Cogdill, R. P., Knight, T. P., Anderson, C. A., and Drennen, J. K. (2007), The financial returns on investments in process analytical technologies and lean manufacturing: Benchmarks and case study, *J. Pharm. Innov.*, 2(1–2), 38–50.
45. du Pre Gauntt, J. (2005), Quality manufacturing: A blockbuster opportunity for pharmaceuticals, Economist Intelligence Unit.
46. Steyer, R. (2005), Merck names clark CEO, available: TheStreet.com, accessed May 5, 2005.

47. Watts, D. C. (2006), PAT—An FDA paper presented at the The Heidelberg PAT Conference, Heidelberg, Germany.
48. Hussain, A. S. (2001), Emerging science issues in pharmaceutical manufacturing: Process analytical technologies, paper presented at the Science Board Presentations to FDA, Rockville, MD.
49. Chisholm, R. S. (2001), TQMS, statistically based in-process control with real time quality assurance, the AstraZeneca total quality management strategy, paper presented at the Science Board Presentations to FDA, Rockville, MD.
50. Raju, G. K. (2001), Pharmaceutical manufacturing: New technology opportunities, paper presented at the Science Board Presentations to FDA, Rockville, MD.
51. Scherzer, R. H. (2002), Quality by design: A challange to the pharma industry, paper presented at the Science Board Presentations to FDA, Rockville, MD.
52. U.S. Department of Health and Human Services (2004), Risk-based method for prioritizing cGMP inspections of pharmaceutical manufacturing sites—A pilot risk ranking model, Food and Drug Administration, Rockville, MD.
53. U.S. Department of Health and Human Services (2006), SMG 2020—FDA quality system framework for internal activities, Food and Drug Administration, Rockville, MD.
54. U.S. Department of Health and Human Services (2006), Guidance for industry: Quality systems approach to pharmaceutical cGMP regulations, Food and Drug Administration, Rockville, MD.
55. U.S. Department of Health and Human Services (2004), Guidance for industry: Sterile drug products produced by aseptic processing—Current good manufacturing practice, Food and Drug Administration, Rockville, MD.
56. U.S. Department of Health and Human Services (2003), Guidance for industry: Comparability protocols—Chemistry, manufacturing, and controls information, draft guidance, Food and Drug Administration, Rockville, MD.
57. U.S. Food and Drug Administration (FDA) (2004, Mar.), Sec. 490.100 Process validation requirements for drug products and active Pharmaceutical ingredients subject to premarket approval (CPG 7132c.08), FDA, Rockville, MD.
58. International Conference on Harmonization (ICH) (2004), Q8: Pharmaceutical development, International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use, ICH, Geneva.
59. International Conference on Harmonization (ICH) (2004), M4: Organisation of the common technical document for the registration of pharmaceuticals for human use, International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use, ICH, Geneva.
60. U.S. Department of Health and Human Services (2006), Guidance for industry: Providing regulatory submissions in electronic format—Human pharmaceutical product applications and related submissions using the eCTD specifications, Food and Drug Administration, Rockville, MD.
61. Yu, L. X. (2006), Implementation of quality-by-design: Question-based review, paper presented at the Drug Information Association 42nd Annual Meeting, Philadelphia, PA.
62. International Conference on Harmonization (ICH) (2005), Q9: Quality risk management, International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use, ICH, Geneva.
63. International Conference on Harmonization (ICH) (2005), Q10: Pharmaceutical quality systems, final concept paper, International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use, ICH, Geneva.

64. Hussain, A. S. (2006), Quality by design and bioequivalence/bioavailability assessment, paper presented at the The Heidelberg PAT Conference 2006, Heidelberg, Germany.
65. Lyon, R. C., and Hammond, S. (2006), Process monitoring of pilot-scale pharmaceutical blends by near-infrared chemical imaging and spectroscopy, paper presented at the Eastern Analytical Symposium, Somerset, NJ.
66. D'Sa, A. (2005), Process analytical technology (PAT): regulatory process, review and inspection, paper presented at the 19th International Forum on Process Analytical Technology—IFPAC 2005, Arlington, VA.
67. Views on prescription drugs and the pharmaceutical industry, The Kaiser Family Foundation, 2005.
68. Arlington, S., Barnett, S., Hughes, S., Palo, J., and Shu, E. (2002), Pharma 2010: The threshold of innovation, IBM Business Consulting Services, Somers, NY.
69. Arlington, S., et al., (2005), The metamorphosis of manufacturing, IBM Business Consulting Services, Somers, NY.

# PROCESS ANALYTICAL TECHNOLOGY

Michel Ulmschneider and Yves Roggo
*F. Hoffmann-La Roche Ltd, Basel, Switzerland*

**Contents**

1 Basic Concepts and Impact
- 1.1 Definition
- 1.2 What Motivated PAT?
- 1.3 Root-Cause Analysis and Process Control
- 1.4 When to Introduce PAT
- 1.5 PAT Enhances Process Understanding
- 1.6 Changing Current Practice Using PAT
- 1.7 Promoting Physical Pharmacy and Pharmaceutical Sciences
- 1.8 Data Mining
- 1.9 Data Warehousing
- 1.10 Data-Mining Methods for Pharmaceutical Processes
- 1.11 Data-Mining Practice
- 1.12 Comments about Data Mining
- 1.13 PAT Methods
- 1.14 Conclusion

2 Vibrational Spectroscopy
- 2.1 Introduction
- 2.2 IR Spectroscopy Theory
- 2.3 Mechanical Model of IR Vibration
- 2.4 Quantum Mechanical Model
- 2.5 Anharmonicity
- 2.6 Structure Elucidation Using MIRS
- 2.7 Extending Use of MIRS
- 2.8 Raman Spectroscopy
- 2.9 Introducing NIRS
- 2.10 Benefits of NIRS
- 2.11 Introducing MIR/NIR Chemical Imaging
- 2.12 Design of MIR Instruments
- 2.13 Conclusion

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

3 Chemometrics
3.1 Introduction
3.2 From Univariate to Multivariate Regression
3.3 Sample Quality and Data Error
3.4 Mathematical Preprocessing of Spectroscopic Data
3.5 Preprocessing NIR Data
3.6 Mathematical Pretreatment and Transformation
3.7 Principal-Component Analysis
3.8 PCA Practice for NIRS
3.9 Pattern Recognition
3.10 SIMCA Classification
3.11 Regression
3.12 Multiple Linear Regression
3.13 PCR and PLS Regression
3.14 Regression Practice in NIRS
3.15 Some Pitfalls
3.16 Example Analytical Applications of NIRS
3.17 Conclusion
Bibliography

## 1 BASIC CONCEPTS AND IMPACT

### 1.1 Definition

Process analytical technology (PAT) is one of the objectives contained in the *Initiative for Pharmaceutical cGMPs for the 21st Century* published by the Food and Drug Administration (FDA). In a few words and according to the FDA's guideline, PAT can be defined as a system for designing, analyzing, and controlling pharmaceutical manufacturing through the measurement of critical quality and performance parameters. The measurements performed on raw and in-process materials or process parameters are intended to enhance final product quality.

Process analytical technology encourages technological innovation, specifically the adoption of new analytical techniques by the pharmaceutical industry designed to improve the understanding and control of manufacturing processes. Both the FDA and industry experts expect benefits over conventional manufacturing practices: higher final product quality, increased production efficiency, decreased operating costs, better process capacity, and fewer rejects. Correspondingly, fundamental changes are also expected within the regulatory framework. The future of pharmaceutical production will require innovative technological approaches and more science-based processes. PAT will boost collaboration between research and development (R&D) and manufacturing departments inside companies and increase overall efficiency. Approvals and inspections will increasingly focus on scientific and engineering principles. As a result, regulators will set higher expectations for new products from the outset.

### 1.2 What Motivated PAT?

Preliminary discussions of PAT concepts between the FDA and certain pharmaceutical companies already active in this field date back to the late 1990s. In September

2004 the FDA released a document for the industry entitled "PAT Guidance for Industry: A Framework for Innovative Pharmaceutical Development, Manufacturing, and Quality Assurance." PAT is clearly anchored in FDA corporate culture.

Pharmaceutical companies are facing growing demands for increased productivity and reduced manufacturing costs. They also have to meet the evolving need for higher quality standards and higher drug expectations. At the same time the quest for new active substances remains a significant issue. Reducing the attrition rate among selected candidates will bring more new medicines onto the market. In terms of drug marketing, the goal is to improve formulations so as to offer patients innovative and more efficient solutions, and thus achieve commercial success or breakthrough. By prioritizing science-based design and introducing novel or improved process techniques, backed by the generation of increased critical data throughout a drug's life cycle, the aim of the emerging PAT strategy is to direct the drug industry toward these essential goals.

Because they have been used for many years, a variety of existing experimental methods and manufacturing processes are considered well established. They are trusted to generate few errors and make only modest contributions to process variation. Due to their longevity, they continue to be widely used in recent drug developments. Improvements in existing technologies are always possible and are constantly being made. However, this makes it difficult to consider or identify potential technological alternatives without critical review or a voluntary management decision to replace well-established techniques. The FDA noticed that nearly all recent drug developments lacked the possibility of enhancing and extending process capabilities toward newer or alternative technologies. More specifically, the FDA wanted to encourage drug manufacturers to achieve more innovation and improve risk management when releasing new medicines on the market.

### 1.3 Root-Cause Analysis and Process Control

When a quality problem arises in present-day production, it is increasingly difficult to identify the root cause. Thorough understanding of process and product performance often comes up against knowledge barriers, whether due to the escalating documentation burden, lack of time, or loss of expertise. The goal of PAT is to enhance process control and understanding so that procedures can be performed differently and more efficiently. The PAT initiative facilitates and encourages the introduction of innovative approaches. It makes it possible to consider shifting from validation to continuous verification. The next step is effective real-time release with continuous processing as an alternative to the conventional batch-after-batch production scheme.

### 1.4 When to Introduce PAT

Building quality into a pharmaceutical product has to be considered from the very beginning of the product's life. Essential preconditions are the equal involvement of—and seamless communication between—R&D and manufacturing. One purpose of PAT is to provide a motivating framework to bring quality into a product from the outset. It is thus essential for it to be involved in the R&D phase. If product quality requirements are understood and implemented from the beginning,

root-cause analysis of quality or process failure after scale-up to commercial manufacturing will be much easier. This is why PAT could play an even more important role in the design and analysis of manufacturing processes, enabling performance control to be based on timely measurement of well-described critical processing data.

Data processing needs should also be considered in the context of overall process analysis strategy to meet emerging requirements for the speed and volume of data collection. Real-time analysis supported by knowledge management requires collecting and gathering all production batch information, for example, by data warehousing. Thus, a PAT data management strategy based on online process analysis or data mining can be set up long before generating large sets of measurement data. Historical data analysis should aim to cover method development, method validation, and ongoing performance monitoring, as well as routine results for a given manufacturing process.

## 1.5 PAT Enhances Process Understanding

Process analytical technology can greatly enhance process understanding. In fact, introducing PAT can act as a key driver to better process knowledge. The expected steps in implementing the PAT approach are the collection of online, in-line, and at-line data (Figure 1) on critical attributes, extraction of information, and analysis of process status data, ending with closure of the loop by dynamic process control. Innovating during development, applying cutting-edge techniques, and process modeling whenever possible, all contribute to a more fundamental exploration of the science behind the process. It is important to realize that PAT is not only the straightforward introduction of additional analytical techniques into a process but also the development of methods to predict future behavior according to given settings of the critical parameters. That means being able to predict final product quality. For example, while implementing the process, it is important to explore all sources of component variation as well as their effect on the finished product in order to select which quality parameters (i.e., attributes) have to be measured for optimal and realistic process control.

Science, engineering, and control technologies can provide a very high level of process understanding and control capability. A process is well understood when all

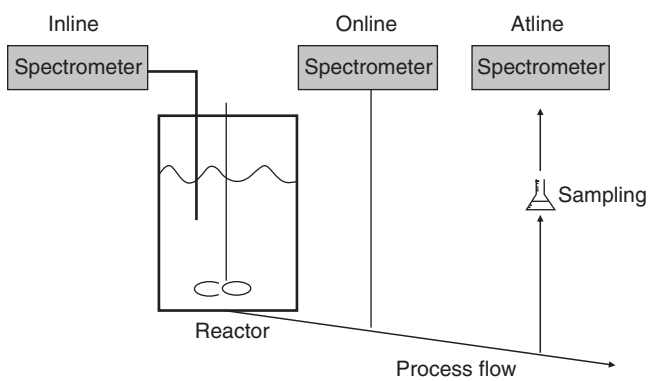


**FIGURE 1** In-line, online, and at-line process measurements.

critical sources of variability are identified and explained. The process should be robust enough to manage this variability. It is also expected that critical quality attributes can be accurately and reliably predicted in an adequate design space when other unexpected variables are encountered (e.g., change of raw material supplier).

### 1.6 Changing Current Practice Using PAT

An approach integrating R&D and manufacturing will enhance process understanding and make acceptable risk management possible. By establishing transferable process models, it will be possible to develop and implement adequate measurement technologies that match process needs rather than vice versa. More efficient and cost-effective technology transfers will facilitate process knowledge, continuous process verification, and compliance, thereby enhancing final product quality. Better process understanding makes it possible to operate by continuous process verification instead of three-batch validation. Measurement technique selection and integration occur very early. Accumulated pertinent knowledge is readily available through data-mining techniques to confirm or control processing. A series of dynamic closed control/compliance loops at the process steps identified as critical will increase confidence in final product quality. In addition knowledge accumulated over time will provide a basis for immediate and rapid intervention in the event of deviation or failure.

A typical illustration of a PAT approach to quality improvement is the use of near-infrared spectroscopy (NIRS) to qualify excipients and active principles just before they enter the production process, for example, in dispensing. As discussed in the next part, near-infrared (NIR) spectra are informative about product structure and overall quality. Because with substances such as excipients the quality range was investigated at some time in the past and fixed into a calibration, NIR measurement can provide simultaneous nondestructive confirmation of the predominant physical and chemical parameters. This is an effective method of reducing uncertainties about possible causes of failure or poor quality during production. Each time a given excipient fails its quality requirements at the moment of use, immediate action can be taken. Control is possible before the risk of failure is increased. Such an approach is complementary to container-wise identification of materials on delivery to a warehouse.

### 1.7 Promoting Physical Pharmacy and Pharmaceutical Sciences

Process analytical technology supposes a more science-based approach to pharmaceutical processes. As a matter of fact, it underlines the observed weakness in formal knowledge of the physical phenomena behind pharmaceutical processes. The physics is less well understood than the chemistry. Conventional physics has moved increasingly into the field of activity of engineers and technologists. Formal approaches are lacking. As a consequence, much highly valuable knowledge of physical phenomena is dispersed across various disciplines. Expertise in physics is often purely technological rather than being formalized and integrated into a specific discipline.

Just as the boundaries of physics and chemistry once merged to create physical chemistry, there is an opportunity now for assembling complementary

scientific knowledge from various disciplines. It is a major challenge to improve understanding through in-depth investigation of the physical phenomena behind pharmaceutical processes. This objective motivates the enforcement of physical pharmacy to improve process understanding through a grounding in theoretical physics.

One major issue is the science and technology of solid particles and powders: characterization, size and shape analysis, processing understanding, and so forth. Others include particle formation and fluid–particle separation, mixture stability, and understanding and simulating the dynamics of powder mixtures. For example, the compaction state of powders and mixtures may change rapidly depending on storage time and conditions. Time to use is not always under control and unexpected changes may occur. Stirring a mixture of two free-flowing powders of different size may result in segregation rather than improved mixture quality. The flow properties of powders depend not only on intrinsic characteristics of the different materials, such as particle size distribution, particle shape, and surface properties, but also on external conditions, such as humidity or compaction status. Further areas of interest include liquid drops, emulsions and colloids, bubbles, and polymers, as well as surface properties, surface analysis, interfacial and electrostatic phenomena, surface reactivity, wet chemistry properties, and solubility.

## 1.8 Data Mining

Complex processes generate large volumes of data over time. As ever-increasing volumes are collected and stored, the gap between buried information and usable accessible knowledge can quickly expand if care is not taken. Data mining extracts new knowledge out of accumulated observations and thus provides a basis for decision making and action. How to turn understanding of buried knowledge to best use? How to extract operational feedback from preexisting, but latent, dormant empirical knowledge? Such questions precede any data-mining project.

As a multidisciplinary technique, data mining sits at the interface between statistics, mathematics, and computer science. It is a collection of methods for detecting regularities and patterns and for extracting knowledge from massive databases using conventional and advanced analytical tools. Another approach to data mining is to view it as the multivariate modeling of a real environment on the basis of multidimensional and accumulated historical data. Thus, data mining is similar to explorative data analysis. It is driven by the data itself. However, it must be considered as different from conventional statistics due to the huge volume of processed data, far above the megabyte scale. Beyond this critical database dimension, most conventional statistical packages exceed their operational limit. Data mining can also be performed without the help of professional statisticians. It runs according to semiautomatic procedures, which makes it widely attractive and more likely to be used in an industrial environment.

Such situations are characteristic of pharmaceutical processes which accumulate a variety of historical data without consideration of pertinence. Accumulation is systematic and exhaustive. However, cross-links between data sources or types may not be established, leading to irrelevant and undetected redundancies. Reliability of the collected data is not clearly established over time and variations may not be detected.

## 1.9 Data Warehousing

The 1990s saw the development of data warehouses. An ideal data warehouse is a collection of historical data varying with time, organized by topic, aggregated in a unique database, and stored in a way that facilitates decision making (Figure 2). Three main functions are required to manage data warehouses. First, the data must be collected or else accessed by an alternative method, for example, as preexisting databases or files. Second, the data warehouse requires management and control tools. Only then can the third function operate, namely data analysis for the purpose of decision making and new knowledge. Dedicated information management tools mediate all external, operational, and historical data to the warehouse. Decisional information management components are used to extract and visualize the data warehouse information. Online analysis processing (OLAP) consists of the real-time analysis and visualization of the historical data. Data mining involves the extraction of rules and models constructed from the collected data.

Online analysis processing mainly comprises the interactive exploration of multidimensional data sets, or data cubes, which are manipulated by operations from matrix algebra, for example, slice-and-dice, roll-up, and drill-down. Computing performance is related to data warehouse size and also data quality, for example, missing data, unsharpness, and redundancy. The multidimensionality issue is critical for extracting pertinent information and selecting the results to be stored and visualized.

The data-mining tools now incorporated in much commercial software are a set of techniques and algorithms for exploring large databases in order to extract semantic links pertinent to event explanation and new knowledge acquisition. The more general goal of data mining is to extract rules and models for understanding connections and assisting decision making. There are numerous fields of application: risk analysis, manufacturing trends, raw material management, maintenance, process validation, development, quality control, and so forth. The idea behind data mining consists in introducing or proposing rules associated with likelihood coefficients established from a large set of existing (i.e., historical) data. The techniques used

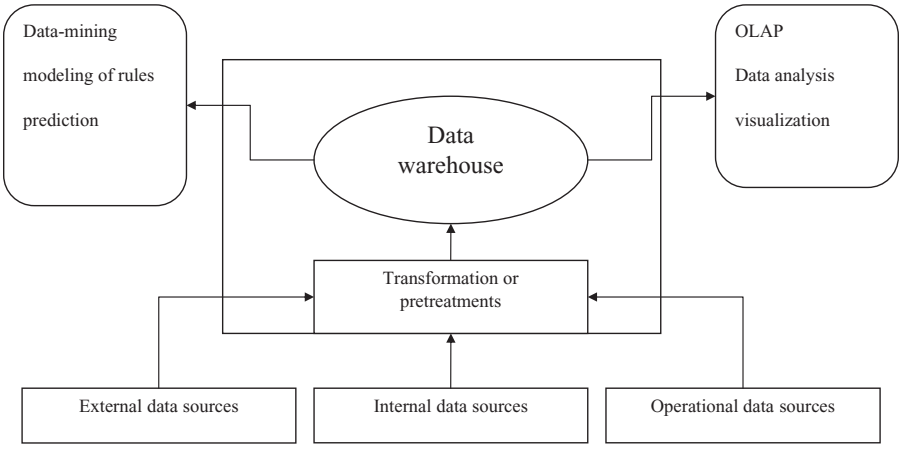


**FIGURE 2** Schematic structure of a data warehouse.

are drawn from the fields of artificial intelligence and numerical and statistical data analysis, for example, functional modeling, learning machines, neuronal networks, Bayesian networks, support vector machines, modeling of associations, and explanatory rules, classifications, and segmentations. Their computing complexity derives from the dramatic up scaling from database to data warehouse level (from megabase to petabase, i.e., $10^6$–$10^{15}$).

## 1.10 Data-Mining Methods for Pharmaceutical Processes

The data warehouse is a central repository of data accumulated over time from various origins: quality control, quality assurance, production, development, and the like. The accumulated data represent a potential gold mine, conferring competitive advantage by facilitating understanding of pharmaceutical process and optimizing it in the light of buried empirical knowledge.

Data mining is used to extract previously unexploited data and knowledge. Its potential for acquiring knowledge and generating explanatory rules can overcome the loss of data or underused accumulated data. There are two ways of proceeding. The first is proactive or directed, for example, hypothesis testing. Particular groupings or features are suspected, and verification or confirmation of identity is sought. The second is reactive or undirected, consisting of simple data exploration. Groupings are unknown, properties undetected or latent, and patterns unidentified. Alternative terms for these approaches are supervised and unsupervised learning, respectively. Top-down and bottom-up approaches complement one another. For example, the confirmatory tools of supervised learning can be used to verify and certify the quality of the discoveries obtained using the exploratory approach.

What can be obtained using data-mining tools? Here is a short list of achievable goals:

- Data characterization to extract or determine descriptors or indicators, for example, by generalizing, summarizing, or grouping
- Establishment of associative and explanatory rules
- Classification (supervised learning) of items or objects in classes according to a given probability
- Clustering of data items (unsupervised learning) in classes, after establishing class limits inductively from existing data sets
- Detection of similarities in time series
- Pattern recognition

Data from external and internal sources is integrated, aggregated, or associated in time series. Data items may contain errors or the data may be missing, unsharp, redundant, or contradictory. A language with operators and variables is required to establish models. Validity levels also have to be defined using suitable optimization and validation criteria. In addition, a search method is required to extract the data from the data warehouse and prepare it for analysis.

Data mining can, therefore, be considered as a three-step operation. Prior to any analysis, the collected data is preprocessed to integrate the warehouse, and some verification is performed to maintain the data level: for example, integration,

aggregation, or grouping of data from different internal and external sources. The data is then selected and data mining performed applying the appropriate algorithms or models. Results are visualized and interpreted for experts in the field.

## 1.11 Data-Mining Practice

Data mining is part of an action process known as a business intelligence chain (Figure 3). Data mining is a flexible solution to the recurrent problem of how to derive knowledge from data. The source for data mining is the existence of a large but buried data set. The corresponding data analysis is an intellectual method that applies only if integrated into the current operational process. Hypothesis testing, knowledge acquisition, and the generation of explanatory rules are directed by active collaboration between different process actors. Data mining is teamwork that requires expertise in various areas, such as information technology (IT), database management, and data analysis. However, the methods are available in commercial packages and may not require the expertise of traditional statisticians. It is the computer which is responsible for discovering patterns or identifying rules or features. In summary, data mining is a logical loop involving the following steps:

- Business understanding
- Precise setting of the data-mining project, for example:
  - Definition of realistic objectives
  - Field of treated data
  - Inventory of available or usable data
- Data preparation
  - Extraction from internal or external sources
  - Verification and correction
  - Pretreatment
- Warehouse construction
- Modeling, for example:
  - Description and visualization
  - Affinity grouping
  - Rules of association, explanatory rules
  - Clustering

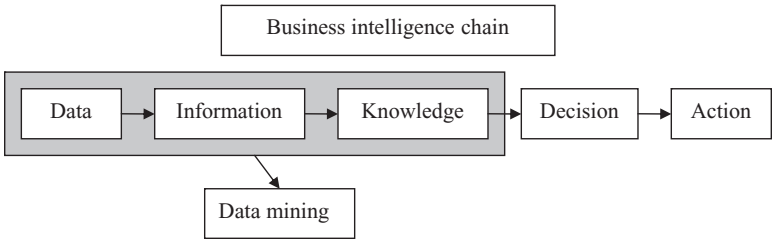


**FIGURE 3** Place of data mining in the decision chain.

  - Classification
  - Estimation
  - Prediction
- Evaluation and comparison of models
- Documentation and presentation of results
- Deployment for action
- Back to business understanding.

## 1.12 Comments about Data Mining

Data mining provides an explanatory analysis from a confirmatory analysis. It is tempting to extract maximal value from available resources such as any kind of accumulated data. But maximal efficiency requires critical insight into the expertise actually buried in data collections or warehouses. The goal of data exploration is to access the buried data to acquire the knowledge that will make explanation, prediction, or estimation possible. That is why data mining requires team effort from data specialists, users, information technologists, and specialists in the relevant field (in this case, pharmaceutical process). It also requires senior management support throughout the organization. Mining is a matter of good practice according to established rules but also a challenge for innovative mathematical techniques. Not all patterns or rules found by data mining are interesting, although the results should remain logical and actionable by experts in the relevant field. Because the algorithms involved tend to be complex and the data volume is huge, software implementation together with the level of information technology are major considerations.

Data mining is driven by the accumulated data but always directed at solving a process, business, or research problem. The results are designed to make it easier to reach a diagnosis or make a decision. They are only likely to be useful in context: that is, they are not simply numbers and graphics but an aid to insight for experts in the relevant field. Also, no single mining technique is equally applicable. A range of different methods or algorithms should be considered, as no one particular technique will work equally well or outperform all other techniques on all problems. Nor will the value of an analytical technique exceed that of the data upon which it is based.

## 1.13 PAT Methods

Almost any existing analytical method can serve the objectives of PAT. Many online applications already exist. With newer techniques, like NIR imaging or matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) mass spectrometry, there are technological problems about performing online or inline analytics. Implementing a given analytical technique close to or during process does not always provide better process understanding. Attributes which are not informative should not be measured at all and are not worth the burden of complex process implementation.

Use of the various techniques listed in Table 1 depends on process requirements. The validity of a given technique or analytical application is challenged by every

**TABLE 1 Analytical Methods for PAT**

| Method | Description | Online Application | Chemical Identification | Pharmaceutical Application Examples |
|---|---|---|---|---|
| Infrared, near-infrared, and Raman spectroscopy | Vibrational spectroscopy (discussed in this chapter) | × | × | Reaction monitoring<br>Polymorphism<br>Content determination<br>Process monitoring (drying, granulation, blending) |
| Hyperspectral imaging | Vibrational spectroscopy coupled with a spatial analysis (cf. chemical imaging chapter) | | × | Chemical compound distributions<br>Counterfeit detection |
| UV–Vis spectroscopy | Photoelectron spectroscopy | × | × | Color measurement<br>Dissolution testing<br>Cleaning validation (ppm-level detection) |
| Terahertz spectroscopy | Far-infrared spectroscopy; 3D imaging | | × | Polymorphism<br>Coating integrity and thickness<br>API distribution possible |
| Laser-induced breakdown spectroscopy | Plasma generated by a laser pulse and detection of the emitted light (destruction of sample) | | × | Drug development<br>Process troubleshooting |
| Laser diffraction | Interaction of a laser beam with particles and detection of the scattered light | × | | Particle size determination |
| Effusivity | Combines thermal conductivity, density, and heat capacity | × | | Mixing, blending, granulation monitoring |
| Acoustic methods | Active or passive | × | | Solid, semisolid, and high viscose sample<br>High shear granulation monitoring Crystallization monitoring |

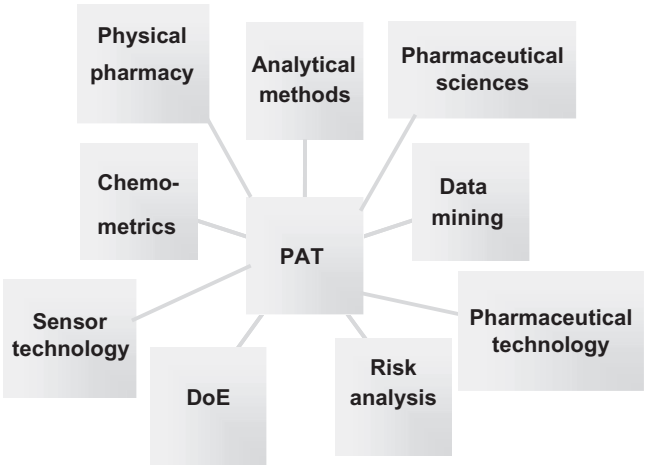


**FIGURE 4** PAT constellation (DoE, design of experiments).

technological advance or new analytical technique. Innovation continuously drives optimization of overall process performance.

## 1.14 Conclusion

Process Analytical Technology can be viewed as a constellation placing greater or less emphasis on a given activity depending on the current problem or situation (Figure 4). There is no written rule or straightforward path to progress through PAT. Experience and expertise are necessary, together with a good knowledge of the pharmaceutical environment. Once a pharmaceutical company has decided to implement PAT, continuous management support for the development and maintenance of PAT-related activities is critical. It is a strategic and necessary step for the future success of PAT to encourage, stimulate, and initiate scientific collaboration and interaction as well as the relevant education and training. Better understanding and control of chemical and pharmaceutical processes are greatly needed, as well as the development of advanced measurement tools and data analysis methods.

A summary of PAT benefits follows:

- Immediate action if quality is not met
- Better process control and understanding
- Less uncontrolled variation and less production waste
- Better and more stable products
- Data collection and improved historical knowledge

Process analytical technology continuously improves product quality, extends the acquired knowledge base for new projects, and shortens time to market.

## 2 VIBRATIONAL SPECTROSCOPY

### 2.1 Introduction

Modern infrared (IR) spectroscopy is a versatile tool applied to the qualitative and quantitative determination of molecular species of all types. Its applications fall into three categories based on the spectral regions considered. Mid-IR (MIR) is by far the most widely used, with absorption, reflection, and emission spectra being employed for both qualitative and quantitative analysis. The NIR region is particularly used for routine quantitative determinations in complex samples, which is of interest in agriculture, food and feed, and, more recently, pharmaceutical industries. Determinations are usually based on diffuse reflectance measurements of untreated solid or liquid samples or, in some cases, on transmittance studies. Far-IR (FIR) is used primarily for absorption measurements of inorganic and metal-organic samples.

Within the electromagnetic spectrum (Figure 5), the IR region ranges from 12,800 to 10 $cm^{-1}$ or from 0.78 to 1000 μm. The IR domain is conveniently subdivided into NIR, MIR, and FIR, respectively, with the following limits:

| | | |
|---|---|---|
| Near | 0.78–2.5 | 4000–12,800 |
| Mid | 2.5–50 | 200–4000 |
| Far | 50–500 | 20–200 |

Methods and applications differ with the IR subregion considered. Academia and analytical chemists commonly consider MIR as the default region of interest. Current MIR instruments are completely different from traditional grating spectrophotometer technology. The generalization of Fourier transform (FT)–based spectrometers in the early 1980s lowered instrument prices and increased the number and types of MIR applications, in particular thanks to the use of interferometers in improving signal-to-noise ratios and detection limits. IR applications were originally limited to qualitative organic analysis. Almost from the outset, absorption MIR became a well-established application for structure elucidation. Organic chemists were trained in the visual and direct interpretation of MIR spectra. Nowadays mid-IR spectroscopy (MIRS) tends to be more viewed as a useful tool for the quantitative analysis of complex samples by absorption and emission spectrometry, which may require calibration and data pretreatment.

Near-IR measurements can be performed similarly to those using dedicated ultraviolet (UV) or visible spectrophotometers. Historically, the most important

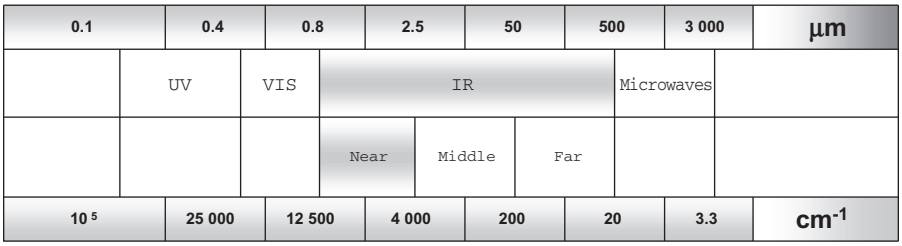


**FIGURE 5** Limits and designation of the spectroscopic domains.

application was quantitative analysis in the food and feed industries. Only more recently have the chemical and pharmaceutical industries shown increasing interest in the NIR range. The major reason for the delay is in the type of information delivered. All observed bands result from overtones or combinations of overtones originating in the fundamental MIR region of the spectrum. Because the measurement method is nondestructive, samples are measured with little or no specific preparation. NIR spectra contain chemical and physical information on the sample. Direct interpretation is limited, if not impossible, meaning that multivariate data processing is routinely required to extract the relevant information. This led most analytical chemists to ignore the potential of NIR. Until the early 1990s, NIR spectrophotometers tended to be the dispersive type based on diffraction gratings. Subsequent technological advance has brought FT and diode array instruments. Filter instruments remain used for ultrarapid measurement of material composition in the food and feed industries.

Being at the edge of the IR region, FIR is believed to have less industrial potential. This is partly due to unresolved experimental and technological difficulties. FIR may provide relevant information, but at the cost of disproportionate effort. Routine use in the pharmaceutical environment is not anticipated in the near future, and for this reason we shall not discuss FIR further.

The most recent developments in IR/NIR technology include imaging large sample surfaces, nondestructive analysis of solids by attenuated total reflectance (ATR), and photoacoustic measurement. Instrument performance continues to increase, with particular respect to reliability and modularity. Spectrometer downsizing, speed of measurement, and mobility no longer represent critical challenges. However, what has really expanded the scope of MIR applications, and use of the full NIR region, has been the constant increase in computing power. The field of application of IR spectroscopy is moving toward the quantitative analysis of complex samples in various measurement modes. These types of samples are characteristic of the pharmaceutical industry. Noninvasive spectral sampling using light probes is at last making in situ analytics attractive, for example, for performing online real-time measurements.

Infrared microscopy was introduced in the early 1980s. Two microscopes, an ordinary optical microscope and an FT IR instrument with reflection optics, were combined. The optical microscope is used to visually locate the spot of interest. The spot is then irradiated with the IR or NIR beam. There are numerous applications for noninvasive measurement, including of contaminants, particles, imperfections, and for fiber identification. Chemical imaging systems (CIS) are a refinement of the technique. Spectra are collected from adjacent areas (pixels) on a larger surface. In practice, an imaging breakthrough became possible after moving away from pixel-after-pixel scanning. CIS flexibility and speed of acquisition improved with the introduction of new detectors, for example, focal plane array (FPA) detectors. Multiple IR/NIR spectra (up to many thousand) are scanned in a single step on the sample surface. With image analysis algorithms and fast computers, current NIR/IR imaging techniques hold fresh promise for resolving quality problems.

### 2.2 IR Spectroscopy Theory

In a typical IR absorption spectrum of an organic substance (Figure 6), the ordinate is transmittance and the abscissa is the wavenumber. A linear wavenumber scale is

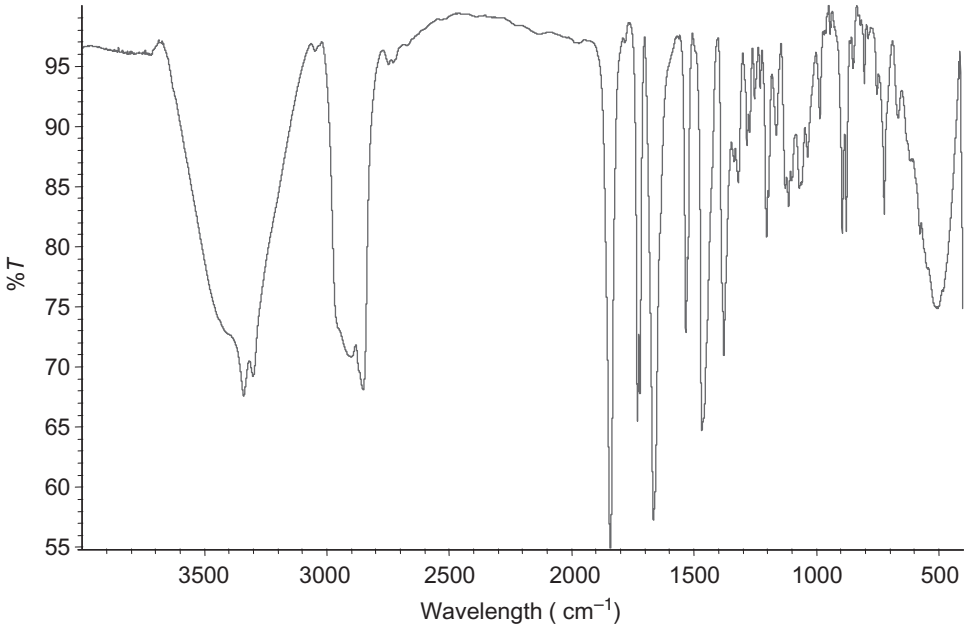


**FIGURE 6** Typical example of an infrared absorption spectrum.

preferred because of the linear relationship between wavenumber and energy and frequency. The frequency of an absorbed radiation is the molecular vibrational frequency actually responsible for the observed absorption.

Infrared absorption, emission, or reflection for molecular species can be explained by assuming transitions from one rotational or vibrational energy state to another. IR radiation is not energetic enough to produce electronic transitions similar to those resulting from UV, visible (Vis), or X-ray radiation. Absorption of IR radiation is limited to molecular species with small energy differences between various vibrational and rotational states. In order to absorb IR radiation, a molecule must undergo a net change in dipole moment as a consequence of its vibrational or rotational motion. Under these circumstances an alternating electrical field interacts with the molecule and causes changes in the amplitude of one of its motions. The dipole moment is determined by the magnitude of the charge difference and the distance between the two charge centers. In addition, regular fluctuation in dipole moment occurs, and a field is established which interacts with the electrical field associated with the incident radiation. If the radiation frequency exactly matches a natural vibrational frequency of the molecule, a transfer of energy takes place that changes the amplitude of molecular vibrations and absorption of radiation results. Similarly, the rotation of asymmetric molecules around their centers of mass results in periodic dipole fluctuations which interact with radiation. Homonuclear species are not concerned and such compounds cannot absorb in the IR.

The amount of energy required to cause a change in energy level is approximately equivalent to radiation of 100 $cm^{-1}$ or less. The relative positions of atoms in a molecule fluctuate continuously, and multiple types of vibrations and rotations about the bonds in the molecule are possible. Exact analysis of all movements becomes

impossible for molecules comprising several atoms. Not only do larger molecules have more vibrating possibilities, but intercenter interactions occur that must be taken into account. Vibrations may be of the stretching and bending variety. Stretching vibration involves a continuous change in interatomic distance along the axis of the bond between the atoms. Bending vibration is characterized by a change in the angle between two bonds and comes in four types: scissoring, rocking, wagging, and twisting. All vibration types may be possible in a molecule containing more than two atoms. In addition, vibration interaction or coupling may occur if the vibrations involve bonds to a single central atom with a change in the characteristic of the vibrations concerned.

### 2.3 Mechanical Model of IR Vibration

Infrared spectra result from light absorption by organic molecules. The easiest way to describe vibrational spectroscopy from a theoretical perspective is to consider the isolated vibrations of a mechanical model called the harmonic oscillator. Atomic stretching vibration behavior can be approximated by a mechanical model consisting of two masses, $m_1$ and $m_2$, connected by an ideal spring. Displacement of one such mass along the spring axis results in harmonic motion. Many fundamental frequencies may be calculated by assuming that band energies arise from the vibration of the ideal diatomic harmonic oscillator (Figure 7), obeying Hooke's law, that is,

$$\upsilon = \frac{1}{2}\pi\sqrt{\frac{k}{u}}$$

where $\upsilon$ is the vibrational frequency, $k$ the classical force constant, and $u = m_1 m_2/(m_1 + m_2)$, the reduced mass of the two atoms.

The model provides a good description of true diatomic molecules and is not far from the average value of two atoms stretching within a polyatomic molecule. The corresponding potential-energy curve is the typical parabola illustrated in Figure 8. This approximation gives the average vibration frequency of the bond. For example, the reduced masses for C—H, O—H, and N—H are 0.85, 0.89, and 0.87. These figures are similar, so the frequencies would be quite similar too. However, the electron-withdrawing and -donating properties of neighbors within molecules act

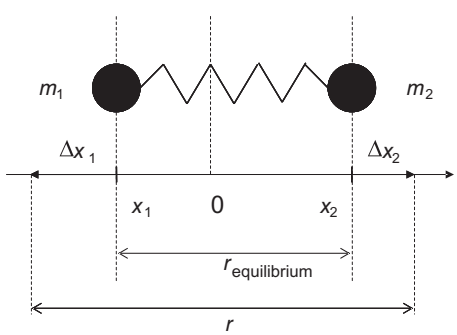


**FIGURE 7** Ideal diatomic harmonic oscillator.

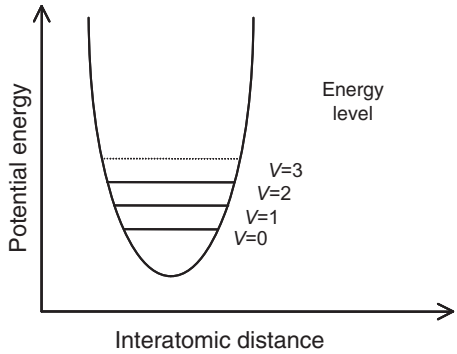


**FIGURE 8** Energy diagram of the ideal diatomic oscillator.

on the observed band strength, length, and frequency. An average value is of little use in structural determinations and these differences cause a real spectrum to develop. The force constant $k$ is a measure of the stiffness of the chemical bond and is the equivalent of the force constant of the spring in the harmonic model. The $k$ values vary widely and cause energy differences which can both be calculated and utilized in spectral interpretation. It has been possible to evaluate some force constants for various types of chemical bonds by IR spectroscopy. Generally, $k$ has been found to range between $3 \times 10^2$ N/m and $8 \times 10^2$ N/m for most single bonds (average: $5 \times 10^2$ N/m). Double and triple bonds are found to have $k$ values two and three times this average, respectively. In practice, these average experimental values can be used to estimate the wavenumbers of fundamental absorption peaks, that is, peaks of the transition from the ground state to the first excited state, for a variety of bond types.

Classical mechanics does not apply to the atomic scale and does not take the quantized nature of molecular vibration energies into account. Thus, in contrast to ordinary mechanics where vibrators can assume any potential energy, quantum mechanical vibrators can only take on certain discrete energies. Transitions in vibrational energy levels can be brought about by radiation absorption, provided the energy of the radiation exactly matches the difference in energy levels between the vibrational quantum states and provided also that the vibration causes a fluctuation in dipole.

## 2.4 Quantum Mechanical Model

Unlike the classical spring model for molecular vibrations, there are not an infinite number of energy levels. Instead of a continuum of energies, there are discrete energy levels described by quantum theory. The time-independent Schrödinger equation is solved using the vibrational Hamiltonian for a diatomic molecule. Values for the ground state ($\upsilon = 0$) and succeeding excited states can be calculated by solving the equation (Figure 8). Absorption of a photon of the correct energy can cause the molecule to change between vibrational energy levels. At room temperature only the ground state has a significant population, and so transitions due to absorption at these temperatures occur from the ground state. Transitions between ground state to energy level 1 give the fundamental absorption if this leads to a

change in molecular dipole moment. Transitions between ground state and energy level 2 or above give overtones. Transitions between multiple states can occur and give rise to combination bands.

A simplified version of the energy levels may be written for the energy levels of a diatomic molecule:

$$E_\upsilon = \left(\upsilon + \frac{1}{2}\right)\frac{h}{2\pi}\sqrt{\frac{k}{u}} \qquad \upsilon = 0, 1, 2, \ldots$$

in which Hooke's law terms can be seen. Rewritten using the quantum term $h\bar{V} = (h/2\pi)\sqrt{k/u}$, the equation reduces to

$$E_\upsilon = \left(\upsilon + \frac{1}{2}\right)h\bar{V} \qquad \upsilon = 0, 1, 2, \ldots$$

In the case of polyatomic molecules, the energy levels become quite numerous. Ideally, one can treat such a molecule as a series of diatomic, independent, harmonic oscillators and the above equation can be generalized:

$$E(\upsilon_1, \upsilon_2, \upsilon_3, \ldots) = \sum_{i=1}^{3N-6}\left(\upsilon_i + \frac{1}{2}\right)h\bar{V}_i \qquad \upsilon_1, \upsilon_2, \upsilon_3, \ldots = 0, 1, 2, 3, \ldots$$

Any transition of an energy state from 0 to 1 in any one of the vibrational states ($\upsilon_1$, $\upsilon_2$, $\upsilon_3$, …) is fundamental and allowed by selection rules. Where the transition is from the ground state to $\upsilon_i = 2, 3$, and so on and all others are zero, it is known as the first overtone, the second overtone, and so on. Transitions from the ground state to a state for which $\upsilon_i = 1$ and $\upsilon_j = 1$ simultaneously are combinations. Other combinations, such as $\upsilon_i = 1$, $\upsilon_j = 1$, $\upsilon_k = 1$, or $\upsilon_i = 2$, $\upsilon_j = 1$, and so forth are also possible. Typically, NIR spectra will contain these overtones and combinations derived from the fundamental vibrations which appear in the MIR. Overtones and combinations are not allowed, but appear as weak bands due to anharmonicity or Fermi resonance. As a rule, overtones occur at one-half and one-third of the fundamental absorption wavelength or 2 and 3 times the frequency. The majority of overtone peaks arise from the R—H stretching and bending modes because the dipole moment is high: O—H, C—H, S—H, and N—H are strong NIR absorbers and form most NIR bands. Since most absorption is repeated in the NIR range, this region is likely to be used to identify a molecule, as with MIR. As a consequence, IR bands are traditionally used to identify functional groups which have characteristic frequencies. NIR spectra are more overlapping, and, although bands can be identified, they cannot be placed in relation to the rest of the molecule. NIR spectra are, therefore, mainly used to confirm the identity of a material, as for true identification.

As given from the quantum mechanics equations, the energy for transition from energy levels 1 to 2 or 2 to 3 should be identical to that for transition from 0 to 1. Furthermore, quantum theory states that the only transitions that can take place are those for which, according to vibrational quantum theory, the vibrational quantum number changes by unity. This is the so-called selection rule.

So far we have illustrated the classic and quantum mechanical treatment of the harmonic oscillator. The potential energy of a vibrator changes periodically as the distance between the masses fluctuates. In terms of qualitative considerations, however, this description of molecular vibration appears imperfect. For example, as two atoms approach one another, Coulombic repulsion between the two nuclei adds to the bond force; thus, potential energy can be expected to increase more rapidly than predicted by harmonic approximation. At the other extreme of oscillation, a decrease in restoring force, and thus potential energy, occurs as interatomic distance approaches that at which the bonds dissociate.

In theory, the wave equations of quantum mechanics can be used to derive near-correct potential-energy curves for molecular vibrations. Unfortunately, the mathematical complexity of these equations precludes quantitative application to all but the very simplest of systems. Qualitatively, the curves must take the anharmonic form. Such curves depart from harmonic behavior by varying degrees, depending on the nature of the bond and the atom involved. However, the harmonic and anharmonic curves are almost identical at low potential energies, which accounts for the success of the approximate methods described.

Anharmonicity leads to deviations of two kinds. At higher quantum numbers, $\Delta E$ becomes smaller, and the selection rule is not rigorously followed; as a result, transitions of $\Delta\pm2$ or $\pm3$ are observed. Such transformations are responsible for the appearance of overtone lines at frequencies approximately two or three times that of the fundamental line; the intensity of overtone absorption is frequently low, and the peaks may not be observed. Vibrational spectra are further complicated by the fact that two different vibrations in a molecule can interact to give absorption peaks with frequencies that are approximately the sums or differences of their fundamental frequencies. Again, the intensities of combination and difference peaks are generally low.

It is ordinarily possible to deduce the number and kinds of vibrations in simple diatomic and triatomic molecules and determine whether these vibrations contain several types of atoms as well as bonds; for these molecules, the multitude of possible vibrations gives rise to IR spectra that are difficult, if not impossible, to analyze. The number of possible vibrations in a polyatomic molecule can be calculated as follows. Three coordinates are needed to locate a point in space; fixing $N$ points requires $3N$ coordinates. Each coordinate corresponds to one degree of freedom for one of the atoms in a polyatomic molecule; for this reason, a molecule containing $N$ atoms is said to have $3N$ degrees of freedom. A molecule features three types of motion. First, the motion of the entire molecule through space; second, the rotational motion of the entire molecule around its center of gravity; and, third, the vibrations of each of its atoms relative to the other atoms. Since all atoms in the molecule move in concert through space, definition of translational motion requires three of the $3N$ degrees of freedom. Another 3 degrees of freedom are needed to describe the rotation of the molecule as a whole. The remaining $3N - 6$ degrees of freedom involve interatomic motion and hence represent the number of possible vibrations within the molecule. In a linear molecule 2 degrees of freedom suffice to describe rotational motion. Thus, the number of vibrations for a linear molecule is $3N - 5$. Each of the $3N - 6$ or $3N - 5$ vibrations is a normal mode. For each normal mode of vibration there is a potential energy relationship. In addition, to the extent that a vibration approximates harmonic behavior, the differences between the

energy levels of given vibrations are the same; that is, a single absorption should appear for each vibration in which there is a change in dipole.

However, fewer experimental peaks may be observed than would be expected from the theoretical number of normal modes. Fewer peaks can be found when the symmetry of the molecules is such that no change in dipole results from a particular vibration. The energies of two or more vibrations can be identical or nearly identical. In some cases absorption intensity is too low to be detected by ordinary means. It may also happen that the vibrational energy is in a wavelength region which is beyond the range of the instrument.

Conversely, more peaks may be found than expected from the number of normal modes. This is the typical situation that concerns the NIR domain. Overtone peaks at two or three times the frequency of a fundamental peak, or addition combination bands at approximately the sum or difference of two fundamental frequencies, are sometimes encountered. The energy of a vibration and thus the wavelength of its absorption peak may be influenced by, or coupled with, other vibrators in the molecule. A number of factors influence the extent of such coupling. Vibration coupling is a common phenomenon. As a result, the position of an absorption peak corresponding to a given organic functional group cannot always be specified exactly. While interaction effects may lead to uncertainties in the identification of functional groups contained in a compound, it is this very effect that provides the unique features of an IR absorption spectrum that are so important for the positive identification of a specific compound.

## 2.5 Anharmonicity

The ideal harmonic oscillator is a somewhat limited model. As the oscillating masses get very close, real compression forces—which are neglected in calculations—fight against the bulk of the spring. As the spring stretches, it eventually reaches a point where it loses its shape and fails to return to its original coil. This ideal case is shown in Figure 9. The barriers at either end of the cycle are approached in a smooth and orderly fashion. Likewise, in molecules, the respective electron clouds of the two bound atoms limit approach by the nuclei during the compression step, creating an energy barrier. At extension of the stretch, the bond eventually breaks when the vibrational energy level reaches the dissociation energy. The barrier at smaller dis-

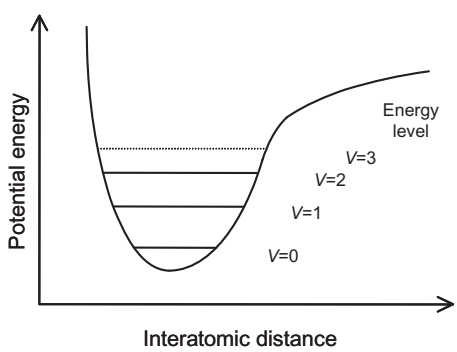


**FIGURE 9** Energy diagram of the anharmonic diatomic oscillator.

tances increases at a rapid rate, while the barrier at the far end of the stretch slowly approaches zero (Figure 9). The shape of the potential energy curve is typical of an anharmonic oscillator.

Energy levels in the anharmonic oscillator are not equal, although they become slightly closer as energy increases. This phenomenon can be seen in the following equation:

$$E_\upsilon = \left(\upsilon + \frac{1}{2}\right) h W_e - \left(\upsilon + \frac{1}{2}\right)^2 W_e X_e + \text{higher terms}$$

where $W_e = \frac{1}{2}\pi(K_e/u)^{1/2}$ is the vibrational frequency, $W_e X_e$ the anharmonicity constant, $K_e$ the anharmonicity force constant, and $u$ the reduced mass of the two atoms. In practice, anharmonicity is between 1 and 5%. Thus, the first overtone of a fundamental vibration set, for example, at 3500 nm would be

$$\upsilon = \frac{3500}{2} + (3500 \times [0.01, 0.02, \ldots])$$

Depending on structural or steric conditions, the number may range from 1785 to 1925 nm for this example. However, it would generally appear at 3500/2, plus a relatively small shift to a longer wavelength. As forbidden transitions, the overtones are between 10 and 1000 times weaker than the fundamental bands. Thus, a band arising from bending or rotating atoms would have to be in its third or fourth overtone to be seen in the NIR region of the spectrum. For example, a fundamental carbonyl stretching vibration at 1750 $cm^{-1}$ or 5714 nm would have a first overtone at approximately 3000 nm, a weaker second overtone at 2100 nm, and a third very weak overtone at 1650 nm. The fourth overtone, at about 1370 nm, would be so weak as to be useless. These figures are based on an illustrative 5% anharmonicity constant.

The detailed examination of the spectra of simple molecules is a direct source to determine the characteristic NIR frequencies for selected vibration modes. For qualitative and quantitative analyses there is the requirement to interpret as much as possible the NIR spectrum. Although interpretation of spectra in a manner analogous to MIR is not conceivable, attempts exist to define and categorize observed NIR frequencies. Examples of reported frequencies for aliphatic hydrocarbons are given in the following list:

| | |
|---|---|
| 8547 $cm^{-1}$ | C—H second overtone in CH=CH |
| 8474 $cm^{-1}$ | C—H group in cis olefins |
| 7700–9000 $cm^{-1}$ | C—H second overtone |
| 8696 $cm^{-1}$ | Second overtone of $CH_2$ antisymmetric stretching |
| 8285 $cm^{-1}$ | Second overtone of $CH_2$ symmetric stretching |
| 1080–1140 $cm^{-1}$ | Second overtone olefin |
| 7692, 8237, 8576 $cm^{-1}$ | C—H stretching second overtone in $CH_2$ |

### 2.6 Structure Elucidation Using MIRS

Mid-IR absorption and reflectance spectroscopy is typically used for determining the structure of organic and biochemical species. When used in conjunction with

other analytical methods, such as mass spectroscopy, nuclear magnetic resonance, and elemental analysis, IR spectroscopy usually achieves positive species identification. Spectra are obtained after sample preparation, usually involving dilution of the analyte. Sample handling is the difficult and time-consuming part of the analysis. Organic samples exhibit numerous IR absorption peaks used for qualitative structure confirmation. First, presumptive functional groups are identified by examining their frequency region from about 3600 to 1200 $cm^{-1}$. As mentioned earlier, the frequency at which an organic functional group absorbs radiation can be approximated from the atomic masses and bond forces between them. These group frequencies are not totally invariant because of interactions with other vibrations. However, such interaction effects are small, and a range of frequencies can be assigned within which it is highly probable that the absorption peak for a given functional group will be found. Group frequencies are listed in correlation charts, which serve as a starting point in the identification process.

Second, the spectrum of the unknown is compared with the spectra of reference compounds featuring all the functional groups found in the first step. The fingerprint region from $1200^{-1}$ to 600 $cm^{-1}$ is extremely useful because small differences in structure and constitution produce significant changes in the appearance and distribution of absorption peaks in this region. Most single bonds give rise to absorption bands at these frequencies. Because their energies are about the same, strong interaction occurs between neighboring bonds. The absorption bands are thus composites of these various interactions and depend upon the overall skeletal structure of the molecule. Exact interpretation in this region is seldom possible because of spectral complexity. On the other hand, it is this complexity that leads to uniqueness and the consequent usefulness of the region in final identification. A close match between two spectra in the fingerprint region constitutes almost conclusive compound identification.

In employing group frequencies it is essential that the entire spectrum rather than a small isolated portion be considered and interrelated. Correlation charts serve only as a guide for further and more careful study. Catalogs of IR spectra that assist in qualitative identification by providing comparison and reference spectra for a large number of pure compounds are commercially available on electronic media. Optimized search systems for identifying compounds from IR spectral databases and algorithms for the matching step produce rapid and reliable potential hits.

## 2.7 Extending Use of MIRS

Organic and inorganic molecular species (except homonuclear molecules) absorb in the IR region. IR spectroscopy has the potential to determine the identity of an unusually large number of substances. Moreover, the uniqueness of a MIR spectrum confers a degree of specificity which is matched or exceeded by relatively few other analytical methods. This specificity has found particular applications for the development of quantitative IR absorption methods. However, these differ from quantitative UV/Vis techniques in their greater spectral complexity, narrower absorption bands, and the technical limitations of IR instruments. Quantitative determinations obtained from IR spectra are usually inferior in quality and robustness to those obtained with UV/Vis and NIR spectroscopy. In addition, univariate or linear cali-

bration curves require meticulous attention to numerous details. One cause of failure is the frequent nonadherence to Beer's law due to the inherent complexity of IR spectra, featuring overlapping absorption peaks or disturbance by stray radiation. Analytical uncertainties cannot be reduced to a level which is comparable to other methods, despite considerable effort or care.

Diffuse-reflectance MIRS has found a number of applications for dealing with hard-to-handle solid samples, such as polymer films, fibers, or solid dosage forms. Reflectance MIR spectra are not identical to the corresponding absorption spectra, but sufficiently close in general appearance to provide the same level of information. Reflectance spectra can be used for both qualitative and quantitative analysis. Basically, reflection of radiation may be of four types: specular, diffuse, internal, and attenuated total.

Specular reflection is encountered when the reflecting medium is a smooth polished surface. The angle of reflection is identical to the incident angle of the radiation beam. If the surface is IR absorbent, the relative intensity of reflection is less for wavelengths that are absorbed than for wavelengths that are not. Thus, the plot of reflectance $R$, defined as the fraction of reflected incident radiant energy versus the wavelength (or wavenumber) appears similar to a transmission spectrum for the sample.

Diffuse-reflectance spectra are obtained directly from powder samples after a minimum of preparation. In addition to the time saved, measurement is nondestructive, leaving the sample intact for further analysis. The widespread use of diffuse reflectance was only possible with the introduction of the FT technique. Reflected radiation from powders is too low to be measured at medium resolutions or inadequate signal-to-noise ratios. Diffuse reflectance (Figure 10) occurs when a beam of radiation strikes the surface of a finely divided powder. With this type of sample, specular reflection occurs at each plane surface. However, since there are many of these surfaces and they are randomly oriented, radiation is reflected in all directions. The intensity of the reflected radiation is independent of the viewing angle. If peak locations are identical in reflectance and transmittance spectra, relative peak heights differ considerably. For example, minor transmittance peaks generally appear larger in reflectance spectra.

Internal-reflection spectroscopy is used to obtain IR spectra of hard-to-handle or hard-to-prepare samples such as solids with limited solubility, films, pastes, adhesives, and powders. Reflection occurs when a beam of radiation passes from a denser to a less dense medium. The fraction of incident beam which is reflected increases as the angle of incidence becomes larger. Beyond a certain critical angle, reflection is complete. During the reflection process the beam penetrates a small distance into

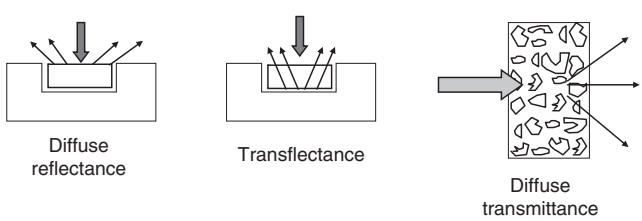


**FIGURE 10** Diffuse reflectance, transflectance, and transmittance measurements.

the less dense medium before reflection occurs. The depth of penetration varies from a fraction of a wavelength up to several wavelengths and depends on the wavelength of incident radiation, the refraction indices of the two materials, and the angle of incident beam with respect to the interface.

Attenuated total reflection (ATR) is the most common reflectance measurement modality. ATR spectra cannot be compared to absorption spectra. While the same peaks are observed, their relative intensities differ considerably. The absorbances depend on the angle of incidence, not on sample thickness, since the radiation penetrates only a few micrometers into the sample. The major advantage of ATR spectroscopy is ease of use with a wide variety of solid samples. The spectra are readily obtainable with a minimum of preparation: Samples are simply pressed against the dense ATR crystal. Plastics, rubbers, packaging materials, pastes, powders, solids, and dosage forms such as tablets can all be handled directly in a similar way.

## 2.8 Raman Spectroscopy

When radiation passes through a transparent medium, a fraction of the beam scatters in all directions. A small fraction of the scattered radiation differs from the incident beam, showing shifts in wavelength determined by the chemical structure of the molecules in the medium. The same types of quantized vibrational changes associated with IR absorption occur, and the difference in wavelengths between incident and scattered radiations corresponds to wavelengths in the MIR. The Raman scattering spectrum and IR absorption spectrum for a given species are very similar. Figure 11 illustrates a typical Raman spectrum. IR is generally the method of choice, but in some cases Raman spectroscopy offers more information about certain types of organic compounds. For example, it is sensitive to conformational

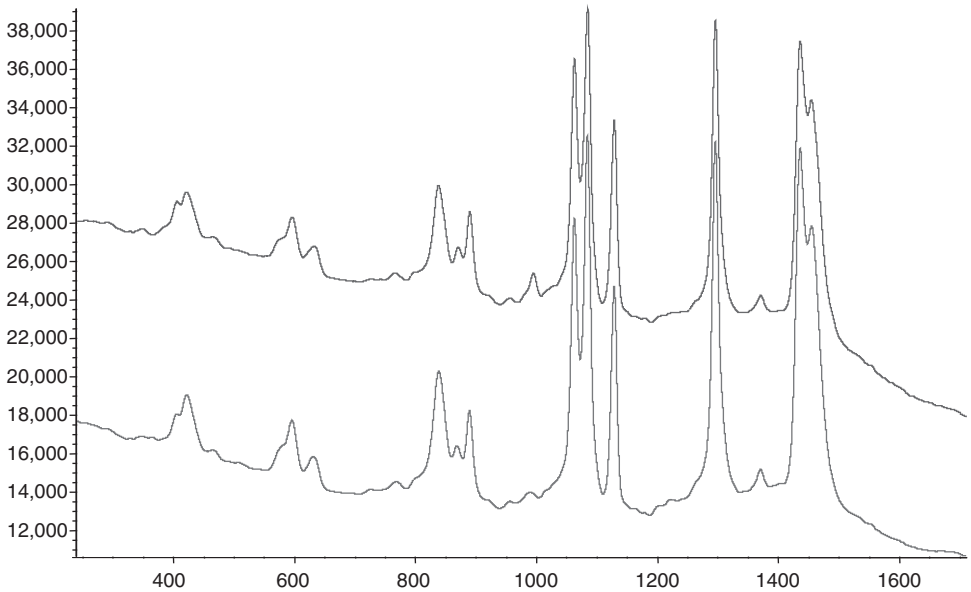


**FIGURE 11** Example plot of two Raman spectra (two polymorphic forms of an excipient).

and environmental information. Peak overlap in compound mixtures is less likely, and quantitative determinations are easier. In particular, accurate quantitative determination can be performed on very small samples. Despite these advantages, Raman spectroscopy has not yet been exploited due to the rather high cost of the instruments.

There are differences between the kinds of groups that absorb in the IR and those that are Raman active. Parts of Raman and IR spectra are complementary, each being associated with a different set of vibrational modes within a molecule. Other vibrational modes may be both Raman and IR active. The intensity or power of a Raman peak depends in a complex way on the polarizability of the molecule, the intensity of the source, and the concentration of the active group, as well as other factors. Raman intensities are usually directly proportional to the concentration of the active species.

In Raman spectroscopy, the excitation radiation occurs at a wavelength distant from any absorption peaks of the analyte. The mechanism which leads to Raman spectra is different from that of MIR spectra, although dependent upon the same vibrational modes. IR absorption requires a change in dipole moment or its associated charge distribution. Only then can radiation of the same frequency interact with the molecule to promote an excited vibrational state. In contrast, scattering involves momentary distortion of the electron cloud distributed around a bond in a molecule, followed by reemission of the radiation as the bond returns to its normal state. In the distorted form the molecule is temporarily polarized. A dipole is momentarily induced which disappears upon relaxation and reemission. Thus, the Raman activity of a given vibrational mode may differ markedly from its IR activity. For example, a homonuclear molecule has no dipole moment either in equilibrium or when stretched, and IR absorption of radiation at the exciting frequency cannot occur. On the other hand, the polarizability of the bond between the two atoms of such a molecule varies periodically in phase with the stretching vibrations, reaching a maximum at the greatest separation and a minimum at the closest position. A Raman shift corresponding in frequency to that of the vibrational mode results. Raman shift magnitude is independent of excitation wavelength. Thus shift patterns are identical regardless of the laser used for excitation.

In MIR spectroscopy water always causes interference, which is not the case with Raman scattering. Thus, Raman spectra can be obtained directly from aqueous solutions. In addition, glass or quartz cells can be employed. The development of Raman spectroscopy was closely associated with the availability of readily usable laser beams. Raman spectra are obtained by irradiating the sample with a laser source of visible or NIR monochromatic radiation. During irradiation, the scattered radiation is acquired at some angle (e.g., 90°) with a suitable device. Raman lines are 0.001% or less intense than the source and more difficult to detect than IR spectra. Raman measurement can be restricted by fluorescence or impurities in the sample. This problem has been partly solved by the use of NIR laser sources, which operate at longer wavelengths. Much higher power can irradiate the sample without causing photodecomposition or simply heating. NIR lasers are not energetic enough to populate a significant number of fluorescence-producing excited electronic energy states in most molecules. Fluorescence is less intense or nearly nonexistent.

Scattered radiation is of three types: Stokes, anti-Stokes, and Rayleigh. The wavelength of Rayleigh scattering is identical to that of the excitation source and

significantly more intense than either of the other two types. By convention Raman spectra are plotted with the abscissa defined as the difference in wavenumbers between the observed radiation and that of the source. As anti-Stokes lines are less intense than the corresponding Stokes lines, only this part of the spectrum is used. However, when fluorescence occurs and interferes with the observations of the Stokes lines, the anti-Stokes part of the spectrum proves helpful, despite the lower intensities involved.

## 2.9 Introducing NIRS

Until recently the NIR region was not considered particularly useful for the spectroscopy of organic compounds. The combination and overtone bands of molecular vibrations occur in the relatively narrow region of 750–3000 nm compared to the fundamental bands occurring at 2800–50,000 nm. This was considered as a drawback. In addition it was observed that NIR bands severely overlapped, were difficult to resolve, and, once resolved, difficult to interpret. If samples were not dried prior to NIR analysis, the changes in hydrogen bonding due to the effects of sample temperature, ionic strength, and analyte concentration could also complicate interpretation of frequently overlapping NIR spectral bands. Changes in hydrogen bonding produce band shifts as well as flattening or broadening of band shapes. The overtone and combination molecular absorptions found within the NIR region are inherently much less intense than the fundamental IR absorptions. Thus, the changes in absorbance in the NIR region are quite small with respect to changes in concentration. The relatively small extinction coefficients due to combination and overtone NIR bands severely restrict the allowable noise levels and stability of any NIR instrument that is used for quantitative work. Sample presentation and the relatively straightforward reflectance measurement involved aspects which were not expected with traditional IR spectroscopy.

If analytical information is obtained with better resolution in the IR region, why should a chemist be interested in NIRS? Numerous difficulties inherent in the use of qualitative NIRS led to its rejection by analytical chemists. Karl Norris, an engineer at the U.S. Department of Agriculture, demonstrated the potential value of the NIR region for quantitative work by making measurements of agricultural products in the 1960s. The basic idea was to provide various research and production facilities with online NIR measurements of agricultural products. NIRS has now become widespread in the chemical and pharmaceutical industry, with the publication of multiple practical applications and a massively increased presence in specialized journals.

## 2.10 Benefits of NIRS

The NIR region is of great interest for pharmaceutical applications. NIRS is fast, nondestructive, and cost effective. Samples require no preparation and can be measured as such, intact and available for further analysis. NIRS can be performed in-, on-, and offline. Also, glass fiber optics can be used to perform remote analysis, thus bringing radiation directly to the sample. Many more advantages can be cited when considering the practical use of NIR in a pharmaceutical process, depending on the particular objective.

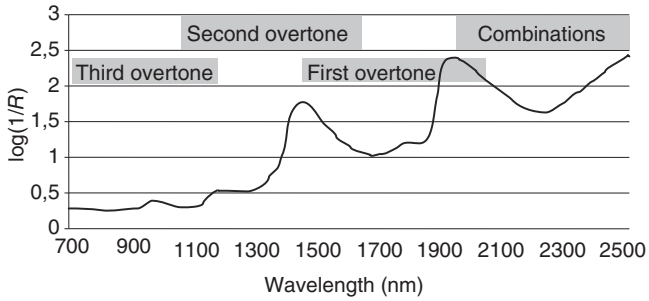


**FIGURE 12** Typical NIR spectrum.

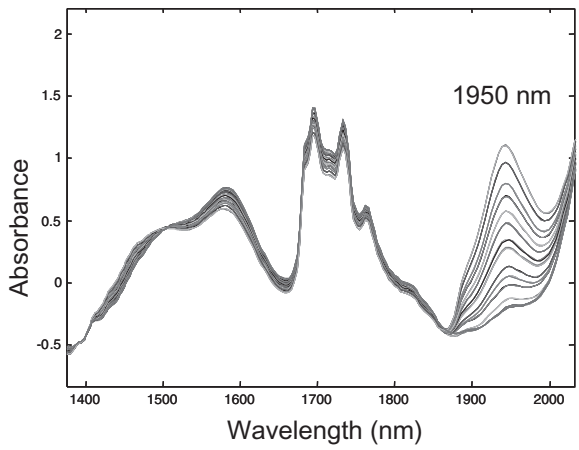


**FIGURE 13** Collection of NIR spectra. The varying water band near 1950 nm can be clearly identified.

As already mentioned, absorption bands in this region are overtones or combination bands of fundamental stretching bands that occur in the 3000–1700 $cm^{-1}$ region (Figure 12). The bonds involved are usually C—H, N—H, and O—H. Because the bands are overtones or combinations, their molar absorptivities are low and detection limits are around 0.1%.

Sample handling is simplified as glass can be used for windows, lenses, and any other optical components. In addition, the laser source is easily focused on small sample area. Very small samples can be investigated without time-consuming preparation. It is also possible for the source radiation to be transmitted through optical fibers. The fiber-optic probe can be in contact with the sample or immersed in it. The probe consists of input fibers surrounded by several collection fibers that transport the scattered radiation to the monochromator. This makes it possible to collect spectra directly under relatively adverse conditions.

In contrast to its MIR counterpart, an important application of NIRS is the routine quantitative determination of species, such as water, proteins, hydrocarbons, and fats, for example, in food or feed products, but also in the petroleum and chemical industries. Figure 13 illustrates a collection of spectra from a pharmaceutical product with varying water content. It shows that a quantitative application can be

developed once a given physical parameter varies, which can then be specifically modeled. The spread of the NIR technique in the pharmaceutical industry was encouraged by advances in computer technology and data handling. Both diffuse reflectance and transmittance measurements (Figure 10) are used, although diffuse reflectance much more widely because of its ease of use. Many spectrometers were specifically designed for the NIR range. Instrument variety is wide and much more varied than in the MIR range. The most sophisticated are dual-beam grating, diode-array, and FT instruments. Simpler filter instruments are still available and remain valuable tools. These instruments can basically be viewed as similar to those used for UV/Vis analysis. Tungsten–halogen lamps with quartz windows serve as sources. Detectors are usually lead sulfide (PbS) or arsine–gallium (AsGa) detectors and many instruments are designed to operate from 180 to 2500 nm.

The NIR spectra are less useful for identification and more useful for quantitative analysis of compounds containing functional groups made up of hydrogen bonded to carbon, nitrogen, or oxygen. Such compounds can often be determined with an accuracy and precision equivalent to UV/Vis spectroscopy, rather than with MIR spectroscopy. Many currently implemented applications concern the determination of water in variety of samples. Reflectance NIRS has become a reliable tool for the quantitative determination of constituents in solids. The first use of this fast technique was for the determination of protein, moisture, starch, oil, lipids, and cellulose in feed and food products. Sample handling, measurement procedure, and data treatment were established with these first applications. Typically in NIR the solid sample is irradiated with one or more narrow bands or the full range of radiation from 1 to 2.5 μm. Diffuse reflectance occurs in which the radiation penetrates the surface layer of the particles, excites the vibrational modes of the analyte molecule, and is then scattered in all directions. A reflectance spectrum is thus produced that is dependent upon the sample composition. In this case the ordinate is the logarithm of the reciprocal of reflectance $R$, $\log(1/R)$, where $R$ is the ratio of the intensity of radiation reflected from the sample to reflectance from a standard reflector, such as finely ground barium sulfate or magnesium oxide. The typical reflectance band at 1940 nm is a water peak used for moisture determination, as can be seen in Figure 13.

Many diffuse-reflectance instruments are available. Some employ several interference filters to provide narrow bands of radiation. Others are equipped with grating monochromators. Ordinarily, calibration is often a stringent requirement as samples must be acquired of the material for analysis that contain the range of analyte concentrations likely to be encountered. It may be useful to grind solid samples to a reproducible particle size. Equations are developed and used for the analysis. Once method development has been completed and validated, solid samples can be analyzed in a few minutes. Accuracy and precision are reported to be of 1 to 2% relative.

In addition to the somewhat empirical and difficult development of NIR applications, thorough documentation must be produced. NIR methods have to comply with the current good manufacturing practice (cGMP) requirements used in the pharmaceutical industry. Various regulatory aspects have to be carefully considered. For example, NIR applications in classification, identification, or quantification require extensive model development and validation, a study of the risk impact of possible errors, a definition of model variables and measurement parameters, and

comprehensive data analysis. Further, clearly defined operating procedures and user training are required for routine analysis. General regulatory requirements also require valid documentation to have been maintained on the life cycle of the NIR model, spectrometer, computer, and the like.

## 2.11 Introducing MIR/NIR Chemical Imaging

Chemical compound homogeneity is an important issue for pharmaceutical solid forms. A classical spectrometer integrates the spatial information. In solid form analysis, use of a mean spectrum on a surface can be a drawback. For example, in the pharmaceutical industry it is important to map the distribution of active ingredients and excipients in a tablet so as to reveal physical interaction between the compounds and help to solve homogeneity issues. Spectroscopic imaging techniques that visualize chemical component distribution are thus of great interest to the pharmaceutical community.

Vibrational hyperspectral imaging is the most recent development to combine chemical information from spectroscopy with spatial information in the sample. In principle, hyperspectral images can be collected using single-point detectors, that is, classical scanning or mapping with microscopes. The commercialization of FPA detectors promoted imaging using a more rewarding analytical method. Array detectors with multiple detector elements measure all pixels on the mapped surface simultaneously, thus drastically reducing recording time, and provide a uniform background to improve the signal-to-noise ratio. A complete spectrum is acquired for each pixel, meaning that a hyperspectral image is in fact a data cube, that is, a three-dimensional (3D) matrix. Hyperspectral imaging provides spatial and spectral information as well as qualitative and quantitative information.

Imaging techniques are typically effective when applied to pharmaceutical tablets in order to explore qualitative aspects or visually address process issues: dissolution, polymorph distribution, moisture content determination, active pharmaceutical ingredient (API) localization and characterization, content uniformity, blending, and granulation. Conventional spectroscopy cannot provide this kind of information. Chemical compound identification or particle size determination can also be estimated by imaging. Selection of an imaging technique will depend on several criteria, such as spatial and spectral resolutions, time of measurement, and wavelength range.

The CISs are rapidly becoming more popular and reliable as their field of application broadens. This is mainly due to the production of surface images by multipoint scanning and mapping. Hyperspectral imaging has proven its potential for qualitative analysis of pharmaceutical products and can be used when spatial information becomes relevant for an analytical application. Even if online applications and regulatory method validation require further development, the power of CIS in quality control and PAT needs no further demonstration, whatever the wavelength domain or method of spectra collection.

## 2.12 Design of MIR Instruments

Different types of IR instruments are available: dispersive grating spectrophotometers, FT instruments, and nondispersive photometers. Until the 1980s, and the

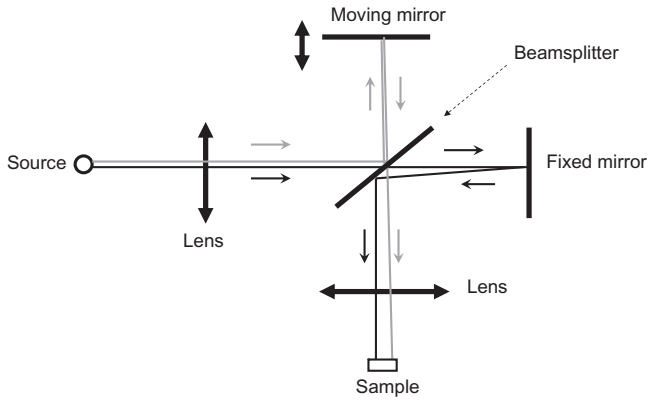


**FIGURE 14** Michelson interferometer.

introduction of more reliable interferometers, the dispersive type was the most widely used. FT spectrophotometers are preferred for many applications because of their speed, reliability, and convenience. They have largely displaced other equipments in the analytical laboratories. Source radiation is split into two beams whose path lengths can be varied periodically to produce interference patterns. The reversed FT algorithm is then used for data recovery. Early FT spectrometers required frequent adjustments of the optical parts, but their unique characteristics of speed, high resolution, sensitivity, and wavelength precision and accuracy made them indispensable. Recent instrument design is much more reliable and easier to adjust. Typically, FT spectrometers are based on the Michelson interferometer (Figure 14), although other types of optical systems are also encountered. Laboratory FT spectrometers are usually proposed as single-beam instruments. For determining transmittance or absorbance spectra with this type of instrument a reference interferogram, air, for example, must first be measured and stored. The analyzed sample is then placed in the radiation path and the process repeated. Transmittance at various frequencies is obtained by computing the ratio of the sample and reference spectra.

Over most of the MIR spectral range, FT instruments appear to have better signal-to-noise ratios than do good-quality dispersive instruments by more than one order of magnitude. The enhanced signal-to-noise ratio can, of course, be traded for rapid scanning, producing spectra within a few seconds. Interferometric instruments also feature high resolutions (less than $0.1\,cm^{-1}$), high accuracy, and reproducible frequency determinations. FT instruments provide a much larger energy throughput (one or two orders of magnitude) than do dispersive instruments, which are limited in throughput by the necessity of narrow slit widths. However, this advantage is partially offset by the lower sensitivity of fast-response detectors required for interferometry.

To perform chemical imaging of sample surfaces, FT spectrometers can be coupled with a microscope or macrochamber with an FPA detector. CIS are available for Raman, NIR, and MIR spectroscopy. Figure 15 illustrates an optical arrangement for chemical imaging.

Instrument design depends on how measurements are performed (Figure 10). The ideal instrument has both transmittance and reflectance capabilities. For

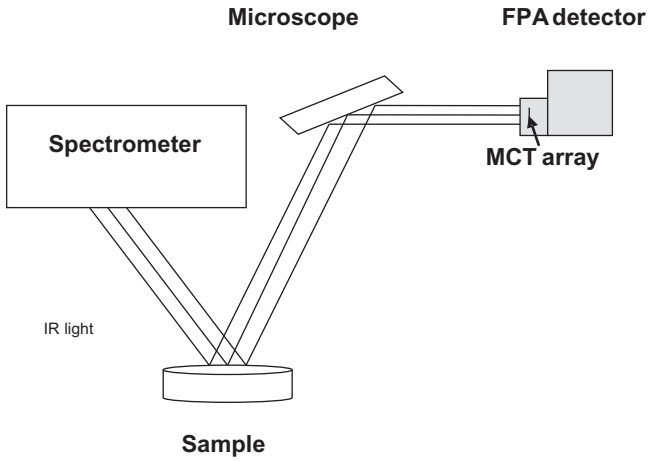


**FIGURE 15** IR microscope coupled with a FPA detector.

example, reflectance measurements penetrate only 1–4 mm into the surface of solid samples. This shallow penetration of energy into a sample brings greater variation when measuring nonhomogeneous samples than transmittance techniques. In transmittance measurements, the entire path through the sample is integrated into the spectral measurement, thereby reducing error due to sample nonhomogeneity. Transmittance is suitable for measuring through compact samples, like intact tablets, but surface scattering induces a loss of transmitted energy with the net effect being a decrease in the signal-to-noise ratio. In some circumstances, particle size is so small that most of the energy striking the sample is scattered back. If the particle size is small enough, the instrument will not transmit enough energy through the sample for the detectors to record a signal. In transmittance, higher frequencies (800–1400 nm) are used to increase the depth of penetration into the sample. But higher frequency energy is more susceptible to surface scattering than lower frequency energy. Transmittance measurements must therefore be optimized based on the relationships between the frequencies used for measurement, surface scatter, and sample path length.

The first NIR spectrometers utilized a tilting filter concept. This concept was refined into a spinning system with a range of filters mounted in an encoder wheel for greater positioning accuracy (wavelength reproducibility) and greater reliability. The use of interference filters represents another type of instrument design, for which prespecified discrete interference filters are manufactured. The filters are mounted in a turret and rotated slowly to different positions during measurement scanning. These instruments are rugged, provided the calibrations are well established and specific to the analyte concerned. If the wrong interference filters are selected for the specific application, successful calibration is impossible. Systems currently exist configured from 6 to 44 discrete wavelength interference filters, in particular for routine application in the food and feed industries.

Dispersive, grating, scanning NIR instruments have been available since the late 1970s. These instruments varied in optical design, but all shared tungsten–halogen source lamps, a single monochromator with a holographic diffraction grating, and uncooled lead sulfide detectors. This design dates back to the early 1980s. Figure 16

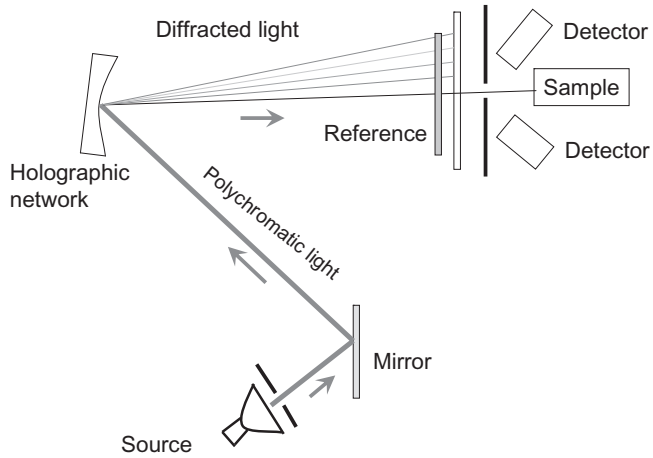


**FIGURE 16** Predispersive grating spectrometer for diffuse reflectance.

shows a typical predispersive monochromator-based instrument where the light is dispersed before striking the sample, with a detector system for diffuse reflectance. The monochromatic light beam illuminates the sample at 0° (normal incidence), and the detectors collect the reflected light at 45°. Two to six detectors can be used, generally lead sulfide detectors for measurement in the 1100- to 2500-nm region. A computer is required for data processing, calibration, and storage. The spectrum is the difference between the raw reflectance measurement of the sample and the raw reflectance measurement of the reference material. Reflectance is converted to absorbance using the function absorbance = $-\log_{10}$ (reflectance), written as $\log 1/R$. Raw transmittance is converted to absorbance using the expression $\log 1/T$.

Postdispersive design is an optical arrangement allowing the transmission of more energy in either a single fiber-optic strand or fiber-optic bundle. Radiation is piped through the fiber-optic strand or bundle where it strikes the sample and returns to the dispersive element, a grating. After striking the grating, the light is separated into the various wavelengths before striking the detector(s).

The integrating sphere is still a common sample presentation geometry for NIR measurements. The use of integrating spheres dates back to the first commercial photometers. Their greatest advantage is that a detector placed at an exit port of the sphere is unlikely to lose energy. However, as detector technology improves, the advantages of integrating spheres for energy collection are no longer critical. Some special applications still require the use of spheres attached to fiber-optic bundles. The use of a sphere allows internal photometric referencing—a sort of double-beam instrument. The application of FT techniques to NIR dates back to the late 1980s. Transmittance/diffuse-reflectance FT-NIR instruments are widespread. As already mentioned, FT spectrometers differ from scanning spectrometers in that the recorded signal is an interferogram (details in the MIR section of this chapter). Other designs include diode array detectors (Figure 17) and NIR emitting diode sources. Acousto-optic tunable filters (AOTF) are devices based on diffraction. The NIR filter is in fact a transparent crystal in which an ultrasonic wave field is created. Thus, the selected wavelength is a function of the field intensity. AOTF scanning speed is

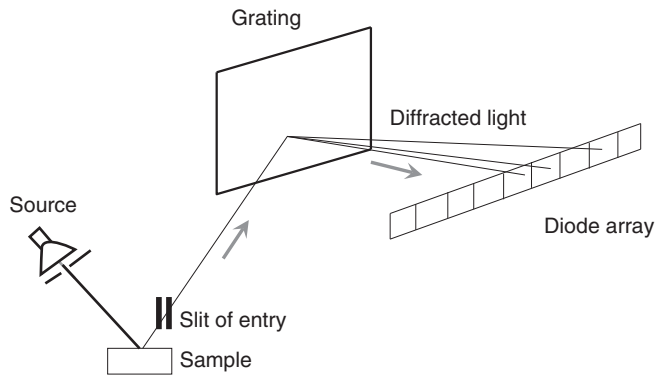


**FIGURE 17** Diode-array spectrometer.

measured in microseconds. This kind of spectrometer is considered very fast but requires extra care to perform reliably in a pharmaceutical process environment. Other techniques are available, such as ultrafast-spinning interference filter wheels, interferometers with no moving parts, and tunable laser sources.

## 2.13 Conclusion

To summarize, instruments for vibrational spectroscopy can be categorized by the optimization of their optical design for a specific type of sample presentation and for the solution of a multiplicity of measurement problems. For development purposes, the instrument of choice is generally one with broad capabilities. Once an application has been clearly defined, an instrument suited to the specific sample presentation geometry or sample type will give the most reliable results. However, at this stage it is worth mentioning that due to the optical design and the mode of measurement, it can be very tricky to transfer calibrations or analytical methods from one instrument to another. For example, calibration transfer is a recurrent issue in NIRS and concerns all instrument designs. This is important if planning to perform measurements of a given sample attribute at different locations with different instruments, for example, laboratory versus process. The issue is more stringent with complex pharmaceutical samples where the signal from the analyte or property of interest strongly interacts with other parameters.

The choice of a given spectroscopy, that is, MIR, NIR, or Raman, has great importance for improving the monitoring and understanding of pharmaceutical processes. In Table 2 the major characteristics of use of MIR, NIR, and Raman spectroscopy are summarized. The most appropriate strategy is implemented according to the necessity and requirements of the manufacturing or analytical operations. In the selection of the spectroscopic technique, attention has to focus on criteria like chemical information content, location of the measurement, speed or real-time ability, robustness, ease of use, sample preparation, sample destruction, and the like. Multiconstituent analysis requires levels of accuracy which are comparable to those of the primary reference methods. Nowadays, a large number of technologies and numerous manufacturers exist. With this in mind, a given choice of equipment will probably match most significant criteria for the analytical problem of interest and

**TABLE 2 Comprison of MIR/NIR Roman Spectroscopic Techniques**

| Spectroscopy | NIR | IR | Raman |
|---|---|---|---|
| Spectral range ($cm^{-1}$) | 12,000–4,000 | 4,000–400 | 4,000–50 |
| Signal intensity | ++ | +++ | + |
| Microscopic analysis | No | Yes | Yes |
| Fiber-optic interfacing | Yes | Yes (limited length) | Yes |
| Sampling through glass | Yes | No | Yes |
| Qualitative application | Yes | Yes | Yes |
| Quantitative application | Yes | Difficult | Yes |
| Instrument robustness | +++ | + | ++ |
| Chemical Interpretation | Chemometrics | Direct | Direct |

at the same time require to accept some inherent measurement limits or restrictions in use.

## 3 CHEMOMETRICS

### 3.1 Introduction

Chemometrics is a term which defines a discipline of chemistry involving mathematics and computer science in order to derive information from data of various type, origin, and complexity. Typical applications include relating the concentration of some analyte found in a sample to the sample's spectral data or identifying some physical or chemical characteristics of the sample. Although performing chemometrics without a computer is effectively impossible, the basic multivariate mathematic calculation has been known since the early twentieth century.

Near-IR spectroscopy is a typical application for chemometrics, for example, how to relate the NIR spectra, obtained at little expense, of the contents of a closed sterile vial with measurement of the concentration of a particular compound in the powdered mixture found by a reference method without opening the sample. To achieve this, the relationship between two data matrices has to be found and a quantitative calibration calculated. This task is a little different depending on whether the data has been generated using statistical experimental design (i.e., designed data) or has simply been collected, more or less at random, from a given population (i.e., nondesigned data). With designed data matrices, the variables are orthogonal by construction. Traditional statistical methods such as analysis of variance (ANOVA) and multiple linear regression (MLR) are well suited to find the regression model in orthogonal data tables. In nondesigned data matrices the variables are seldom orthogonal but are more or less collinear. MLR is likely to fail in these circumstances, and the use of projection techniques such as principal-component regression (PCR) or partial least squares (PLS) is recommended. A regression model can then be used to predict new, that is, unknown, samples. Prediction is a useful technique because it can be used in place of costly and time-consuming measurements.

Another example is the noninvasive identification of species from NIR absorbance spectra. Classification is simply a matter of finding out whether new samples

are similar to classes of samples that have been used to make the model. If a new sample fits a particular model well, it is said to be a member of that class. Many analytical tasks fall into this category. For example, raw materials like excipients may be sorted according to "good and bad quality," finished products classified into "grades A, B, C," and so on. Principal-component analysis (PCA) is a typical mathematical procedure for resolving such sets of data into orthogonal components whose linear combinations approximate the original data with a desired degree of accuracy. As successive components are calculated, each accounts for the maximum possible amount of residual variance in the set of data. In NIRS, the data usually consist of large sets of recorded spectra, and the number of components will be smaller than or equal to the number of known variables or the number of spectra.

### 3.2 From Univariate to Multivariate Regression

In spectroscopy the simplest method of quantitative calibration is based on a single independent variable, for example, wavelength, since a sample attribute such as analyte concentration is a linear function of absorbance at a given wavelength. Modeling of the concentration requires conventional least-squares fitting. A straight line is fitted through a set of data points to minimize the sum of squares of deviations between estimated and known data points. In this approach, a wavelength is selected when it shows a high degree of correlation between concentration and absorbance. Correlation is an indicator of goodness of fit between concentration and absorbance and of how well the calibration describes the data set. The linear relationship has a practical advantage in that it permits direct and visual estimation of goodness of fit, thus enhancing the analyst's trust in his data collection. Where pharmaceutical samples are concerned, the linear approach rapidly reaches its limits and another approach is required.

Multiple linear regression extends linear regression to one wavelength by least squares, with more than one wavelength selected to perform a calibration. The complexity of the test samples is again an issue with this approach. MLR requires independent variables for its ability to explain the data set. As pharmaceutical samples comprise a complex matrix in which species interact to various degrees in the NIR range, it is impossible to find appropriate wavelengths to select. Observed absorbance values are linked as they may describe related behaviors in the spectral data set. It is typical of NIR spectra of mixtures that collinearity is found among the wavelengths and MLR will never perform a usable linear calibration.

Partial least squares multivariate regression uses many wavelengths without the limitation due to collinearity. In fact, collinearity makes PLS the tool of choice when considering segments or entire NIR spectra. PLS treats constituent data and spectral data on the same basis, without statistical prerequisites of standard distribution or noncollinearity. Errors are assumed equally and in both data sets. Spectral and constituent data are modeled simultaneously according to an iterative algorithm. This training step is required for the model to learn to predict the sample attribute of interest. For each iteration a part of the spectral data and a part of the corresponding constituent data are combined until the data set is optimally described. The nonexplained part of the data set is made up of residuals, which are a measure of the quality of the modeling. The original data is combined in factors or principal components. Coefficients are calculated—loadings for spectral data and scores for

constituent data—to indicate the extent of original data involvement in the computation of each factor. Finally, the amount of variance modeled, that is, the explained part of the data, is maximized for each factor and the residuals are minimized.

The raw data is normally preprocessed to some extent before PLS calibration, as illustrated below. A critical step in PLS modeling is the selection of the number of factors. Selecting too few factors will provide an inadequate explanation of variability in the training data set, while too many factors will cause overfitting and instability in the resulting calibration. Optimal factor number is estimated during validation of the calibration, which is part of the PLS algorithm. Either an additional and independent data set is used (external validation in one computation step) or the training data set is split into subsets for continuous internal validation at each iterative step (cross-validation). In addition, the resulting PLS calibration must prove able to predict unknown samples. It is also good practice to challenge the resulting calibration in a way that enhances trust in future results. In no way are PLS calibrations black boxes despite the fact that visualization is far from evident when compared to basic linear regression.

### 3.3 Sample Quality and Data Error

In an ideal case, a given resulting calibration equation provides an explicit value for the concentration of an unknown analyte in terms solely of measurable absorbances. Only the concentrations of the desired analyte need be known during calibration. The values for other sample constituents need not be determined, although in order to develop a good calibration equation, variations of these values should also be evenly distributed. Indeed, in some cases, even the number of other compounds in the calibration samples may not be known. Using stepwise multiple regression, features in the spectra which correlate most closely with the analyte concentration are selected for a particular sample set. Once optimal calibrations are computed, the NIR instrument can be used to predict unknown samples for the determination of the quantity of desired analyte. Thus, regression analysis is a method to develop the relationship (i.e., regression calibration equation) between several spectral features and the constituent being investigated. For diffuse reflectance measurements, the effects of extraneous physical phenomena are superimposed on the absorbance readings and have to be taken into account. The equations can be formulated to account for these phenomena by implicitly including corrections for their effects in the calibration coefficients.

One problem that chemometrics cannot deal with is the appropriate selection or preparation of samples. Careful sample selection increases the likelihood of extracting useful information from spectral data. Whenever it is possible to actively experiment with selected variables or parameters, the quality of the results is increased. The critical part is deciding which variables can be changed and what limits to fix on their variation. Appreciation of which parameters may influence data depends mainly on the experience of NIR operators but is also a matter for specialists. Each problem will generate and implicate a specific set of more or less usable variables. The key to success lies in gaining experience in dealing with numerous and different problems. Experimental design could ideally help to generate the data indicating which parameters or design variables influence output or response variables.

Understanding the interactions between design variables while identifying the optimal conditions for extracting the relevant information should in principle minimize the number of experiments required, thus reducing cost. An experimental design program offers appropriate design methods and encourages good experimental practice by performing a small number of useful experiments that span the important variations. Experimental design will not be discussed here, but only the way in which data should be collected and organized in practice for multivariate regression. Typically, with spectral data, the problem is to find out which variables are really important for variation in the data matrix. The usual tasks when thinking about modeling a response variable are to find out which variables are necessary for describing the samples adequately, which samples are similar to each other, and whether the data set contains groups of samples. A good way of finding this information is by decomposing the spectral data matrix into a structured part and a noise part, typically by using PCA. Another problem is how to establish a regression model between a spectral data matrix and a response variable matrix.

Before performing multivariate data analysis, statistical analysis of sample response values or spectral data may help to check data quality. Descriptive statistics summarize the distribution of one or two variables at a time. They are not supposed to say much about the data structure, but they are useful for obtaining a quick look at each separate variable before starting an analysis. One-way statistics, that is, mean, standard deviation, variance, median, minimum, maximum, lower, and upper quartile, can be used to explore the data set and detect out-of-range values, abnormal spread, or asymmetry. These statistics reveal anything suspect in a data table and indicate whether a transformation might be useful. Two-way statistics, for example, correlation, show how variations in two variables may be linked in a data table. Checking these statistics is also useful to detect out-of-range values and outliers.

Experience in NIRS practice shows that the critical step(s) for successful implementation vary with each specific problem to solve. However, the following procedure is recommended when analyzing nondesigned data. First, investigate the origin and availability of the data. In formulating the study problem, define the precise objective of data collection and the expected analysis results. The data collection should span appropriate variation in the explored variable(s) or attribute(s). If the naturally available data do not span the expected variation, prepare or measure samples with corresponding experimental data. Raw spectra may have to be transformed and mathematical pretreatments performed. Calibrate and validate the model either using PCA or PLS. Explore and challenge the way the calibrations behave on real samples or data, for example, validate the method according to current pharmaceutical usage and requirements.

There are three types of data error: random error in the reference laboratory values, random error in the optical data, and systematic error in the relationship between the two. The proper approach to data error depends on whether the affected variables are reference values or spectroscopic data. Calibrations are usually performed empirically and are problem specific. In this situation, the question of data error becomes an important issue. However, it is difficult to decide if the spectroscopic error is greater than the reference laboratory method error, or vice versa. The noise of current NIR instrumentation is usually lower than almost anything else in the calibration. The total error of spectroscopic data includes

sample-induced errors which can be much greater than the instrument's noise level. Such errors include particle size effects, variations in the packing density of powders, effects of impurities, and effects due to changing physical characteristics of the sample (e.g., crystallinity). Practical experience shows that in many cases sample-induced error in the optical data remains small, prompting the empirical assumption that optical data error is always smaller than reference laboratory error.

### 3.4 Mathematical Preprocessing of Spectroscopic Data

As discussed above, the greatest source of error in NIR calibration is usually reference laboratory error, sample nonhomogeneity, and nonrepresentative sampling in the learning (training) set or calibration set population. Instrument quality and equation selection account for only a fraction of the variance or error attributable to NIR analytical technique in current routine application.

When radiation is reflected from solid matter surfaces, diffuse and specular reflected energies are superimposed. The intensity of diffusely reflected energy depends on the angles of incidence and observation but also on sample packing density, crystalline structure, refractive index, particle size distribution, and absorptive qualities. Thus, in practice, an ideal diffusely reflecting surface can only be approximated, even with the finest possible grinding of the samples. There are always coherently reflecting surface regions acting as elementary mirrors whose reflection obeys Fresnel's formulas. Radiation returning to the sample surface from the interior can be assumed to be largely isotropic, thus meeting the requirements of the Beer–Lambert law. It is assumed that radiant energy is continuously removed from the incident NIR beam and converted to thermal vibrational energy of atoms and molecules. Decrease in the intensity of the diffusely reflected light depends on the absorption coefficient of the sample. The absorption coefficient ($K$), when taken as the ratio $K/S$, where $S$ is the scattering coefficient, is proportional to the quantity of absorbing material in the sample. In the Kubelka–Munk theory, reflectance ($R$) is related to absorption ($K$) and the scattering coefficient ($S$) by the equation

$$\frac{K}{S} = \frac{(1-R)^2}{2R} = F(R)$$

Diffuse reflectance $R$ is a function of the ratio $K/S$ and proportional to the addition of the absorbing species in the reflecting sample medium. In NIR practice, absolute reflectance $R$ is replaced by the ratio of the intensity of radiation reflected from the sample and the intensity of that reflected from a reference material, that is, a ceramic disk. Thus, $R$ depends on the analyte concentration. The assumption that the diffuse reflectance of an incident beam of radiation is directly proportional to the quantity of absorbing species interacting with the incident beam is based on these relationships. Like Beer's law, the Kubelka–Munk equation is limited to weak absorptions, such as those observed in the NIR range. However, in practice there is no need to assume a linear relationship between NIRS data and the constituent concentration, as data transformations or pretreatments are used to linearize the reflectance data. The most used linear transforms include log $1/R$ and Kubelka–Munk as mathemati-

cal pretreatments. To some extent PCR, PLS, and multilinear regression compensate for nonlinearity. Calibration equations can be developed which compensate to some extent for the nonlinear relationship between analyte concentrations and log $1/R$ or Kubelka–Munk transformed data. If a matrix absorbs at different wavelengths than the analyte, Kubelka-Munk can prove a useful linearization method for spectroscopic data. If the matrix absorbs at the same wavelength as the analyte, log $1/R$ will prove a better choice for relating reflectance to concentration. This transformation is basically well suited for diffuse reflectance NIR spectra of most mixtures with absorbing matrices. Plots of $F(R)$ versus concentration are less linear than plots of log $1/R$ versus concentration.

### 3.5 Preprocessing NIR Data

When generating calibration equations using samples of known composition, the independent variable is represented by the spectroscopic readings (i.e., log $1/R$) at specific wavelengths, while the concentration of the analyte of interest, for example, determined by traditional laboratory techniques, is the dependent variable.

Spectral raw data may have a distribution or shape that is not optimal for analysis. Background effects, baseline shifts, measurements in different conditions, different variances in interfering variables and the like can complicate information extraction using multivariate methods. It is important to minimize the noise introduced by such effects. Preprocessing operations include centering, weighting, and numerous mathematical transformations. Mean centering consists of subtracting the average spectra from each individual spectrum. This ensures that all results will be interpretable in terms of variation around the mean considered as the model center. This is useful with models in which a linear relationship between spectral data and response data is supposed to go through the origin. Depending on the kind of information to be extracted from the spectral data, weights based on the standard deviation (i.e., square root of the variance, which expresses the variance in the same unit as the original variable) may be used for scaling. This may be a typical pretreatment for PCA, PLS, or PCR calibrations which are projection methods based on finding directions of maximum variation, thus depending on the relative variance of the variables. A possible weighting option is the 1/SD standardization, which gives all variables the same variance, that is, 1. In this case all variables are given the same chance to influence estimation of the components. This is advisable if the variables are measured with different units, have different ranges, or are of different types. It is also possible to fix a constant weight for each variable manually. Weighting involves stretching and shrinking by measuring a position relative to the extremes in the actual spectral data table. By standardizing the spectra, variation in the data set is performed relative to the extremes in the data table. However, this procedure emphasizes the relative influence of unreliable or noisy attributes.

### 3.6 Mathematical Pretreatment and Transformation

A wide range of transformations can be applied to spectral data before they are analyzed. The main purpose of transformations is to make the latent variables better available for powerful analysis. One of the most widely used is logarithmic transformation, which is especially useful to make skewed variables more symmetrically

distributed. It is also indicated when the measurement error in a variable increases proportionally with the level of that variable. Taking the logarithm will achieve uniform precision over the whole range of variation. This particular application is also called variance stabilization. In case of limited asymmetries, a square root can be sufficient. Smoothing is relevant for variables which are themselves a function of some underlying variable, for instance, of time. It is also one of the first operations performed on recorded NIR spectra. It removes as much noise as possible without degrading important information content. Moving point average (MPA) is a classic smoothing method which replaces each observation with an average of the adjacent observations including it. The number of observations to average is a variable parameter. Polynomial smoothing, also called Savitzky–Golay smoothing, involves least-square fitting of a polynomial equation to a window of $n$ sequentially selected spectral data points. If the polynomial order is less than the number of data points, the polynomial cannot pass through all selected data points, and the least-square fit gives a smoothed approximation to the original window. Normalization is a family of transformations which are computed sample wise. Here the purpose is to scale samples so as to improve specific properties. Mean normalization is the classic algorithm. It consists in dividing each row of a data matrix by its average, thus neutralizing the influence of possible hidden factors. It is equivalent to replacing the original variables by a profile centered on 1. Only the relative values of the variables are used to describe the sample, and the information carried by their absolute level is dropped. It is indicated in the specific case where all variables are measured in the same unit, and their values are assumed to be proportional to a factor which cannot be directly taken into account in the analysis. Maximum normalization is an alternative procedure which divides each row by its maximum absolute value instead of the average. The maximum value becomes +1 and the minimum value becomes –1. In range normalization, each row is divided by its range, that is, maximum value minus minimum value and the curve span becomes 1.

More specific transformations for spectroscopic data are the reflectance to absorbance, absorbance to reflectance, and reflectance to Kubelka–Munk transformations. Multiplicative scatter correction (MSC) is an additional transformation method used to compensate for additive and/or multiplicative effects in spectral data. MSC was originally designed to deal with multiplicative scattering alone. However, MSC successfully treats a number of similar effects such as path length problems, offset shifts, and interference. The idea behind MSC is to remove the two effects—amplification, which is multiplicative, and offset, which is additive—from the spectral data table to prevent them from dominating the table's information content. Derivation is typically relevant for spectral data that are a function of some underlying variable influencing absorbance at various wavelengths. Derivatives are also a simple but powerful technique for magnifying fine structure in raw spectra lacking structure, which is common in NIRS. By increasing the order of derivation, band structure resolution is increased. The inherent drawback is a decrease in the signal-to-noise ratio, which in the particular case of NIR spectra is not considered, since smoothing while performing a derivation may limit the effect of noise. On the other hand, it will tend to hide some weaker spectral features. The main advantage of the second derivative as usually performed is that band structure is maintained: Peak maxima still correspond while peak shape and resolution are improved. The Savitzky–Golay algorithm permits computation to higher order derivatives, includ-

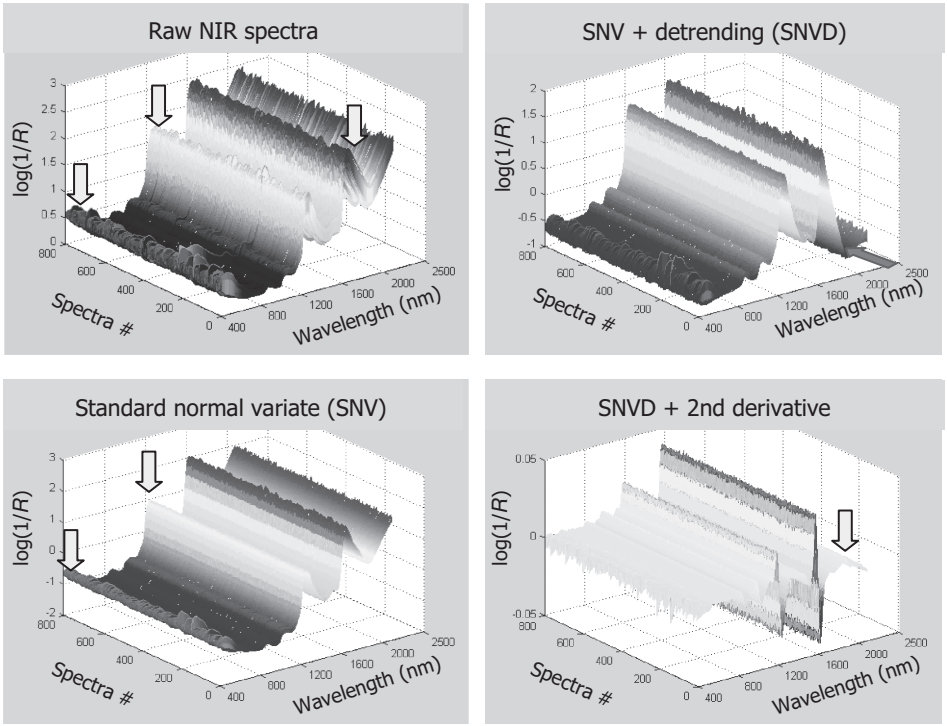


**FIGURE 18** Examples of combined spectral preprocessing.

ing a smoothing factor which determines how many adjacent variables will be used to estimate the polynomial approximation for derivation. Norris derivation is an alternative algorithm for computing first derivatives only. A baseline correction algorithm is the standard normal variate (SNV) method, which does not affect overall spectra layout. Averaging samples in case of replicates or variables for variable reduction, for example, to reduce the number of latent variables, is a method of obtaining more stable and more readily interpretable results. Figure 18 illustrates the effect of selected combined transformations to a collection of spectral data.

### 3.7 Principal-Component Analysis

Large data tables contain an amount of information which is partly hidden because the data complexity prevents ready interpretation. This is typical of NIR spectra collections. PCA is a projection method used to visualize all the information contained in the data table. It can be used to show in what respect one sample differs from another, which variables contribute most to this difference, and whether these variables contribute in the same way and are correlated or independent of each other. It also reveals sample patterns or groupings. In addition, it quantifies the amount of useful information, as opposed to noise or meaningless variation, contained in the data table. Principal components are defined only for the data set from which they were computed. They may also hold for other data of identical type, but this is not guaranteed, and it is certainly not true for different types of data.

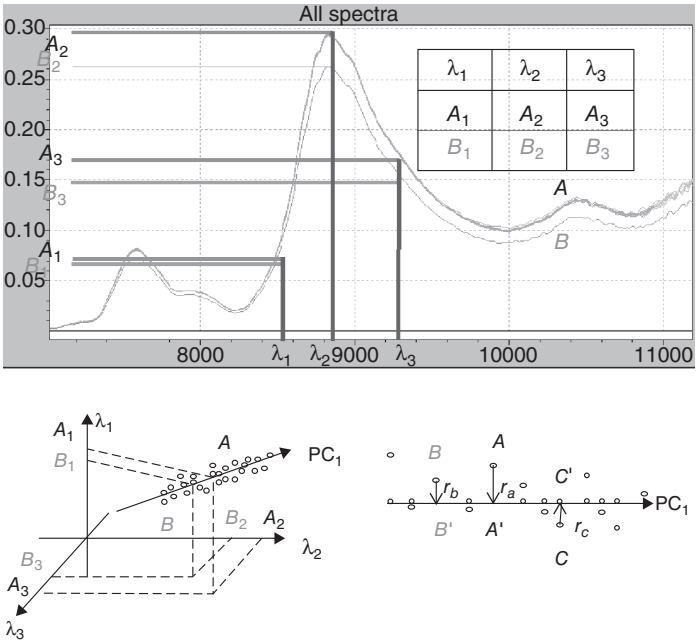


**FIGURE 19** Geometric illustration of PCA. Each spectrum is plotted as a point, whose coordinates are the intensities collected at three different wavelengths. For a given collection of spectra, the first principal component is the direction in space which covers the greatest variation of the corresponding point set.

The PCA modeling forms the basis for several classification and regression methods. The underlying idea is to replace a complex multidimensional data set by a simpler version having fewer dimensions, but still fitting the original data closely enough to be considered a good approximation. Extracting information from a data table consists of exploring variations between samples, that is, finding out what makes a sample different from, or similar to, another. Two samples can be described as similar if they have close values for most variables. From a geometric perspective, in the case of close coordinates in the multidimensional space of variables, the two points are located in the same area. Likewise, two samples can be described as different if their values differ greatly with respect to at least some variables. The two points have different coordinates and are located far away from each other in the multidimensional space. An illustration of this geometrical concept is proposed in Figure 19.

The principle of PCA consists of finding the directions in space—known as principal components (PCs)—along which the data points are furthest apart. It requires linear combinations of the initial variables that contribute most to making the samples different from each other. PCs are computed iteratively, with the first PC carrying the most information, that is, the most explained variance, and the second PC carrying most of the residual information not taken into account by the previous PC, and so on. This process can go on until as many PCs have been computed as there are potential variables in the data table. At that point, all between-sample variation has been accounted for, and the PCs form a new set of axes having two

advantages over the original set. First, the PCs are orthogonal to each other. Second, they are ranked so that each carries more information than any of those following. It thus prioritizes their interpretation, starting with the first PCs. This method of generation ensures that the new set of axes is more suitable for interpreting the data structure. Usually only the first PCs contain pertinent information, with later PCs being more likely to describe noise. In practice, only the first PCs are examined rather than the whole raw data table: Not only is it less complex, but it also ensures that noise is not mistaken for information. If all PCs were retained, there would be no approximation at all and no gain in simplicity either. Deciding on the number of components to retain in a PCA model is a compromise between simplicity, completeness, and effectiveness. The PCA model is only an approximation of reality.

Each component of a PCA model is characterized by three complementary sets of attributes, that is, variances, loadings, and scores, respectively. The importance of a given PC is expressed by its variance. Loadings describe the relationships between variables, while scores describe sample properties. Variances are error measures which tell how much information is taken into account by successive PCs. The way they vary with the number of components can be studied to decide how complex the model should be. Residual variance designates the variation in the data that remains to be explained once the current PC has been taken into account, while explained variance, often measured as a percentage of total variance in the data, measures the proportion of variation in the data accounted for by the current PC. These variances can be considered either for a single variable or sample or for the whole data. They are computed as mean square variations, corrected for the remaining degrees of freedom.

Loadings describe the data structure in terms of variable correlations. Each variable has a loading on each PC. This reflects how much the variable contributed to that PC and how well that PC reflects the variation of the variable considered. In geometric terms, loading is the cosine of the angle between the variable and the current PC, with a value that ranges by definition between −1 and +1. The smaller the angle, the greater the link between variable and PC, and the greater the loading value. Variables with high loadings (i.e., close to +1 or −1) for a given PC contribute greatly to the meaning of that particular PC. Thus, in studying correlations between variables, the loadings indicate their respective angles in the multidimensional space. For instance, if two variables have high loadings along the same PC, their angle is small, which in turn means that the two variables are highly correlated. If both loadings have the same sign, the correlation is positive; when one variable increases, so does the other. If negative, the variables are anticorrelated.

Scores describe the data structure in terms of sample patterns and emphasize differences or similarities. Each sample has a score on each PC, which is the coordinate of the sample on the PC. Once the information carried by a PC has been interpreted with help of the loadings, the score of a sample along that PC can be used to characterize a given sample. It describes the major features of the sample, relative to the variables with high loadings on the same PC. Samples with close scores along the same PC are considered as similar because they have close values for the corresponding variables. Conversely, samples with greatly dissimilar scores differ greatly from each other with respect to those variables.

### 3.8 PCA Practice for NIRS

As already mentioned, any multivariate analysis should include some validation, that is, formal testing, to extrapolate the model to new but similar data. This requires two separate steps in the computation of each model component: calibration, which consists of finding the new components, and validation, which checks how well the computed components describe the new data. Each of these two steps needs its own set of samples: calibration samples or training samples, and validation samples or test samples. Computation of spectroscopic data PCs is based solely on optic data. There is no explicit or formal relationship between PCs and the composition of the samples in the sets from which the spectra were measured. In addition, PCs are considered superior to the original spectral data produced directly by the NIR instrument. Since the first few PCs are stripped of noise, they represent the real variation of the spectra, presumably caused by physical or chemical phenomena. For these reasons PCs are considered as latent variables as opposed to the direct variables actually measured.

A convenient analogy for understanding latent variables is reconstructing the spectrum of a mixture from the spectra of the pure chemicals contained in the mixture. The spectra of these pure chemicals would be the latent variables of the measured spectrum because they are not directly accessible in the spectrum of the mixture. However, PCs are not necessarily the spectra of the pure chemicals in the mixtures representing the samples. PCs represent whatever independent phenomena affect the spectra of the samples composing the calibration set. If one sample constituent varies entirely independently of everything else, and this constituent has a spectrum of its own, then one of the PCs will indeed represent the spectrum of that constituent. It is most unusual for any one constituent to vary in a manner that is exactly independent of any other. There is inevitably some correlation between the various constituents in a set of specimens, and any PC will represent the sum of the effects of these correlated constituents. Even if full independence is accomplished, there is dependence in that the sum of all constituents must equal 100%. Consequently, the PC representing that source of independent variability will look like the difference between the constituent of interest and all the other constituents in the samples. The spectrum of the constituent considered could be extracted mathematically, but the PCs will not look exactly like the spectrum of the pure constituent.

Once samples have been collected and the corresponding NIR spectra stored in suitable files, building and using a PCA model involves three steps: selecting the appropriate preprocessing procedure(s), running the PCA algorithm and diagnosing the model, and interpreting the loading and score plots. Once the model is built, it is important to assess its quality before using it for interpretation. There are two steps in diagnosing a PCA model. The variances must be checked to determine how many components the model should include and to estimate how much information the selected components take into account. It is important to verify the variances calculated during validation. Then it is advisable to look for outliers, that is, samples that do not fit into the general pattern.

Total explained variance measures how much of the original variation in the data is described by the model. It expresses the proportion of structure found in the data by the model. Total residual and explained variances show how well the model fits

the data. Models with small total residual variance (close to 0) or large total explained variance (close to 100%) explain most of the variation in the data. With simple models residual variance falls to zero with few components. If this is not the case, it means that there may be a large amount of noise in the data. Alternatively, it may also mean that the data structure is too complex to be accounted for by only few components. Variables with small residual variance and large explained variance for a particular component are well explained by the relevant model. Variables with large residual variance for all or the few first components have a small or moderate relationship with other variables. If some variables have much larger residual variance than other variables for all components or the first components, they may be excluded in a new calculation. This may produce a model that is easier to interpret. Calibration variance is based on fitting the calibration data to the model. Validation variance is computed by testing the model on data not used in building the model.

Outliers may sometimes account for large residual variance. An outlier is a sample which looks so different from the others that either it is not well described by the model or it influences the model too much. In practice, true spectral outliers are first considered to be samples whose spectral characteristics are not represented within the specified sample set. At least one of the model components may focus on trying to describe only this particular sample, even if this is irrelevant to the more important structure present in the other samples. In PCA, outliers can be detected by using different plots or analysis. For example, score plots show sample patterns according to one or two components. It is easy to identify a sample lying far away from the others. Such a sample is likely to be an outlier. Residuals measure how well samples or variables fit the model determined by the components. A sample with a high residual is poorly described by the model, which, however, fits the other samples quite well. Such a sample does not fit among the samples well described by the model and can be considered an outlier. Deciding to exclude or retain such spectral outliers solely on mathematical criteria is not necessarily correct. The outlier may be considered as not part of the group intended for use as a calibration set. However, such outliers may indicate additional characteristics not taken into account during initial sample selection.

### 3.9 Pattern Recognition

Pattern recognition can be classified according to the distinction between supervised and unsupervised techniques. Unsupervised methods such as cluster analysis classify data without calibration and are based solely on the collected sample data. Supervised classification uses the spectral data and some class membership information. Therefore, mathematical models are computed in a first step with a calibration set containing spectra and class information. This model is then applied to predict new sample classes. Feature extraction methods such as PCA or wavelet compression are often applied before cluster analysis. PCA is valuable for extracting features or visualizing the data set. Another benefit of PCA is to reduce the number of wavelengths. Many clustering algorithms are in use. Nonhierarchical methods include Gaussian mixture models, $K$ means, and fuzzy $C$ means, each of which can be subdivided into hard and soft clustering methods. For example, hard clustering by $K$ means would declare one given item as one class membership, whereas soft

clustering by fuzzy *C* means would assign fractional membership degrees of each cluster.

Current methods for supervised pattern recognition are numerous. Typical linear methods are linear discriminant analysis (LDA) based on distance calculation, soft independent modeling of class analogy (SIMCA), which emphasizes similarities within a class, and PLS discriminant analysis (PLS-DA), which performs regression between spectra and class memberships. More advanced methods are based on nonlinear techniques, such as neural networks. Parametric versus nonparametric computations is a further distinction. In parametric techniques such as LDA, statistical parameters of normal sample distribution are used in the decision rules. Such restrictions do not influence nonparametric methods such as SIMCA, which perform more efficiently on NIR data collections.

### 3.10 SIMCA Classification

Classification is useful when the response considered is a category variable that can be interpreted in terms of several classes to which a sample may belong. The main goal of classification is to assign new samples reliably to preexisting classes, but classification results can also be used as a diagnostic tool to identify the most important variables to retain in the model or to find outliers. Applications include predicting whether a pharmaceutical product meets specified quality requirements, in which case the result is simply a binary response, or more generally testing or confirming the identity of a substance. PCA and discriminant analysis are techniques that have found extensive use in NIR analysis for that purpose. SIMCA, a multivariate technique optimized for NIR data analysis, combines PCA models for each defined class in the training set. The approach can be applied to a more general class of problems than simple classification, for example, identification. PCA is performed on a data set requiring qualitative analysis. After computing the PCs, scores are calculated and used to perform qualitative analysis by surrounding each region of multidimensional space containing each group's scores with a surface. Conceptually, this enclosing surface can be modeled as an ellipsoid and distances to these clouds are computed using Mahalanobis metrics based on the PC scores.

In practice, the optimal number of PCs should be chosen for each class model separately, according to a suitable validation scheme. Each model is checked for possible outliers and improved as much as possible, like any PCA model. Before using the models to predict class membership for new samples, the specificity of the models should be verified, that is, checked for class overlap and adequate interclass distance. Once each class has been modeled, and class overlap is not excessive, new samples can be fitted to each model. Unknown samples are then compared to the class models, and assigned to classes according to their similarity to the training samples. The modeling stage implies that enough samples have been identified as members of each class to allow reliable models to be built. Accurate sample description also requires a sufficient number of variables. Actual classification is based on statistical tests performed on Mahalanobis distances between sample and model. With each unknown sample, all variable values are computed using the model scores and loadings, before being compared to the actual values. The residuals are then combined into a measure of the object-to-model distance. The scores are also used to measure the sample's distance from the model center, known as leverage. Finally,

both object-to-model distance and leverage are taken into account to decide the classes to which the sample belongs. Any sample belonging to a class should have a small distance to the class model.

### 3.11 Regression

Historically, one motivation for performing NIRS was to develop fast and noninvasive quantitative analysis. To achieve this, it is not sufficient to extract PCs from the data. There must be a regression method relating these PCs to the constituent, analyte, or physical property for which the calibration is performed. Regression concerns all methods attempting to fit a model to the observed data. The fitted model may be used to describe the relationship between two groups of variables or to predict values of unknown samples. If $X$ and $Y$ are the two data matrices involved in regression, the purpose is to compute a $Y = f(X)$ model, which tries to explain, or predict, the variations in the $Y$ variable(s) from those in the $X$ variable(s). The link between $X$ and $Y$ is explored through a common set of samples from which both $X$ and $Y$ values have been collected and are clearly known. Building a regression model involves collecting variable values for selected samples and fitting a mathematical relationship to the corresponding spectral data. For example, spectroscopic measurements are performed on solutions with known concentrations of a given compound. Regression is used to relate the concentration to the spectrum. Once the regression model is build, the unknown concentration for new samples can be indirectly predicted using the spectroscopic measurements as predictors. The advantage appears obvious if one considers noninvasive and nondestructive measurement by NIR. If the concentration is difficult or expensive to measure directly, spectroscopic analysis offers an alternative and much cheaper method of determination. Thus the indication for using regression as a predictive tool is its potential for performing fast and low-cost measurement as a substitute for more expensive or time-consuming alternatives.

The data may require appropriate preprocessing before a regression can be built and used. The calibration step must be followed by a validation step, that is, the model must be checked for its efficiency in predicting independent data. Once the number of components has been selected based on the calibration and validation variances, the model is diagnosed by interpreting the loading and score plots (for PCR and PLS), the loading weight plots (for PLS), and $B$ coefficients (PCR), and examined for the prediction of new data.

### 3.12 Multiple Linear Regression

Classic univariate regression uses a single predictor, which is usually insufficient to model a property in complex samples. Multivariate regression takes into account several predictive variables simultaneously for increased accuracy. The purpose of a multivariate regression model is to extract relevant information from the available data. Observed data usually contains some noise and may also include irrelevant information. Noise can be considered as random data variation due to experimental error. It may also represent observed variation due to factors not initially included in the model. Further, the measured data may carry irrelevant information that has little or nothing to do with the attribute modeled. For instance, NIR absorbance

spectra may contain information relative to solvents, processing path, instrument status, light probes, and the like, in addition to the analyte concentration for measurement. A good regression model should be able to pick out only the relevant information while leaving irrelevant variation aside.

Multiple linear regression is a method based on ordinary least-squares regression. It involves matrix inversion which leads rapidly to collinearity issues if the variables are not linearly independent. In MLR, all $X$ variables participate in the model independently of each other, and their covariations are not taken into account. $X$ variance is not meaningful in this context. Variable independence is also an essential precondition. Further, to perform the inversion, MLR requires more samples than predictors and no missing values in the data table. If the data table complies with these conditions, MLR will approximate the response values by linear combination of predictor values, yielding regression coefficients known as $B$ coefficients. Other results are predicted $Y$ values, residuals with error measures, and ANOVA. It is noteworthy that MLR is the only multivariate method for which formal statistical tests of significance for regression coefficients are available. To evaluate the goodness of the model, diagnostic tools are associated with the regression coefficients. The standard error is an estimate of the precision of a given coefficient. A Student's $t$ value can be computed and compared to a reference $t$ distribution, which in turn indicates a significance level or $p$ value. It shows the probability of a $t$ value equal to or larger than the observed value if the true value of the regression coefficient were 0. Predicted $Y$ values are computed for each sample by applying the model equation with the estimated $B$ coefficients to the observed $X$ values. For each sample, the residual is the difference between the observed $Y$ value and predicted $Y$ value. The only relevant measure of how well the MLR model performs is provided by the $Y$ variances. Residual $Y$ variance is the variance of the $Y$ residuals. It expresses how much variation remains in the observed response after the modeled part is removed. It is an overall measure of misfit, that is, the error made when fitted $Y$ values are computed as a function of $X$ values.

### 3.13 PCR and PLS Regression

Multivariate regression is better suited to fit a relationship between spectroscopic data and the variable to estimate. PC regression (PCR) is a two-step procedure which first decomposes the $X$ matrix by PCA, then fits an MLR model, using the PCs instead of the raw data as predictors in the regression step. MLR and PCR model one $Y$ variable at a time, while a PCR model using all PCs gives the same solution as MLR. Partial least squares or projection to latent structures (PLS) will model both $X$ and $Y$ matrices simultaneously to find the latent or hidden variables in $X$ that will best predict the latent or hidden variables in $Y$. The difference between PCR and PLS lies in the algorithm. PLS components are similar to PCs and are also referred to as PCs. PLS1 deals with only one response variable at a time (like MLR and PCR). PLS2 handles several responses simultaneously. PCR and PLS are projection methods, like PCA. Model components are extracted in such a way that most information is carried by the first PC, then the second PC, and so on. At a certain point, the variation modeled by any new PC becomes mostly noise. Residual variances are used to determine the optimal number of PCs modeling useful information while avoiding overfitting. PLS uses both independent and dependent variables

to find the regression model. It switches iteratively between $X$ and $Y$. PLS usually needs fewer PCs than PCR to reach the optimal solution because the focus is on the dependent variables. Results of PCR modeling are given as scores, loadings, predicted $Y$ values, residuals, error measures, and $B$ coefficients. Results of PLS modeling are given as $T$ scores and $U$ scores, $P$ loadings and $Q$ loadings, loading weights, predicted $Y$ values, residuals, and error measures.

The PLS scores are interpreted in the same way as PCA scores since they are the sample coordinates along the model components. The additional feature in PLS is that two different sets of components are considered, summarizing variations in the $X$ space or $Y$ space. PLS loadings express the relatedness of each $X$ and $Y$ variable to the model component. $T$ scores are the coordinates of data points located in the $X$ space that describe the part of structure in $X$ which is most predictive for $Y$. $U$ scores summarize the part of structure in $Y$ which is explained by $X$ along a given model component. The relationship between $T$ scores and $U$ scores is a model of the relationship between $X$ and $Y$ along a specific component, and it can be visualized for diagnostic purposes. It follows that loadings will be interpreted differently in the $X$ space and $Y$ space. $P$ loadings are similar to PCA loadings. They express how much each $X$ variable contributes to a specific model component. Directions determined by the projections of $X$ variables are used to interpret the meaning of the location of a projected data point on a $T$ score plot, in terms of variations in $X$. $Q$ loadings express the direct relationship between $Y$ variables and $T$ scores. Thus, the directions determined by the projections of $Y$ variables by means of $Q$ loadings can be used to interpret the meaning of the location of a projected data point on a $T$ score plot in terms of sample variation in $Y$. When plotted on a single graph, $P$ and $Q$ loadings make it possible to interpret $T$ scores by considering variations in both $X$ and $Y$. In contrast to PCA loadings, PLS loadings are not normalized, so that $P$ loadings and $Q$ loadings do not share a common scale. Thus, only their directions can be interpreted, not their lengths. The residuals should be randomly distributed and free from systematic trends. The most useful residual plots are $Y$ residuals versus predicted $Y$, and $Y$ residuals versus scores plots.

Where there is more than one $Y$ variable, PLS2 is designed for interpretation of all the variables simultaneously. It is often argued that PLS1 or PCR are better predictors; this is usually true if there are strong nonlinearities in the data, in which case modeling each $Y$ variable separately according to its own nonlinear features might perform better than trying to build a common model for all $Y$'s. On the other hand, if the $Y$ variables are somewhat noisy, but strongly correlated, PLS2 will model the whole information by excluding more noise.

As in PCA, outliers may influence modeling and should be detected. In regression, there are many ways a sample can be defined as an outlier. It may be an outlier according to $X$ variables only or to $Y$ variables only, or to both. It may also not be an outlier for either separate set of variables but become an outlier for ($X$, $Y$) regression.

### 3.14 Regression Practice in NIRS

Calibration is the fitting stage: The main data set, containing only the calibration samples set, is used to compute model parameters such as PCs, regression coefficients, and the like. The models must be validated to get an idea of how well a

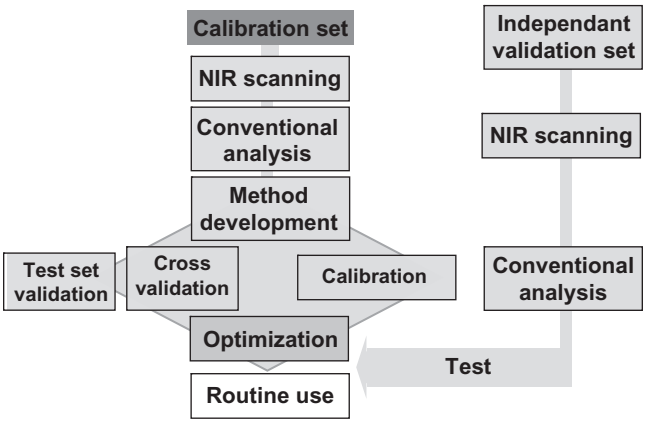


**FIGURE 20** Typical working flowchart of the development of a PLS calibration with NIR spectra.

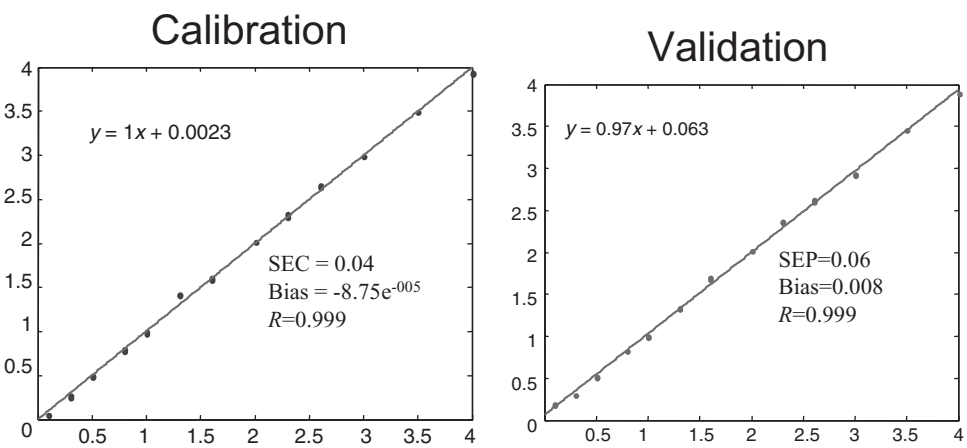


**FIGURE 21** Typical output of a PLS calibration and validation.

regression model performs if used to predict new, unknown samples. A test set consisting of samples with known variable values is used. Only the $X$ values are fed into the model, from which response values are predicted and compared to the known, true response values. The model is validated if prediction residuals are low. Validating a model means checking how well the model performs on real new data. As a regression model is usually made to perform predictions on future unknown samples, validation must estimate the uncertainty of prediction. If the uncertainty is reasonably low, the model can be considered usable. The steps required for complete modeling are illustrated in Figure 20. And in Figure 21 the end result obtained from a typical chemometric package performing calibration and validation is shown.

Independent (external) test-set validation and cross-validation are the most current methods of estimating prediction error. External test-set validation is based on testing the model on a subset of available samples, which will not be involved in

**TABLE 3 Descriptors Used to Estimate Performance of Calibration**

$$\sigma_y = \sqrt{\frac{\sum_{j=1}^{m}(y_j - \bar{y})^2}{m-1}} \qquad \mathrm{cov}(y', y) = \sqrt{\frac{\sum_{j=1}^{m}(y_j - \bar{y})(y'_j - \bar{y}')}{m-1}} \qquad R = \frac{\mathrm{cov}(y', y)}{\sigma_{y'}\sigma_y}$$

$$\mathrm{SEC} = \sqrt{\frac{\sum_{j=1}^{m}(y_j - y'_j)^2}{m-1-q}} \qquad \mathrm{SEP} = \sqrt{\frac{\sum_{j=1}^{n}(y_j - y'_j)^2}{n}} \qquad \mathrm{Bias} = \frac{\sum_{j=1}^{n}(y_j - y'_j)}{n}$$

where $y_j$ = reference value for the sample $j$
$y'_j$ = predicted value for the sample $j$
$m$ = number of samples in calibration
$n$ = number of samples in validation
$q$ = PC number

computing the model components. For example, the global data table is split into two subsets. The calibration set contains all samples used to compute the model components, using both $X$ and $Y$ values. The test set contains all the remaining samples, for which $X$ values are fed into the model once a new component has been computed. The predicted $Y$ values are then compared to the observed $Y$ values, yielding a prediction residual that can be used to compute a validation residual variance, or a measure of the uncertainty of future predictions, called root mean square error of prediction (RMSEP). This value is calculated for each modeled response and indicates the average uncertainty that can be expected when predicting $Y$ values for new samples, expressed in the same units as the $Y$ variables. Table 3 gives a formula for RMSEP. The results of future predictions can then be presented as the predicted values ±RMSEP. This measure is valid provided that the new samples are similar to those used for calibration. Otherwise, the prediction error might be much higher. No assumption about statistical error distribution is made for modeling. As a consequence, prediction error cannot be given as a proper statistical interval estimate (e.g., twice the standard deviation, etc.). RMSEP is a practical average prediction error and if both calibration and validation sample sets are representative of future samples to predict, it is a good error estimate.

With cross-validation, the same samples are used both for model estimation and testing. Cross-validation represents an alternative way of utilizing samples for validation if their number is small or moderate. The method consists in leaving out a few samples from the calibration data set and calibrating the model on the remaining data points. The left-out sample values are then predicted and the corresponding prediction residuals computed. The process is repeated with another subset of the calibration set, and so on until every object has been left out once. All prediction residuals are combined to compute the validation residual variance and RMSEP. Full cross-validation leaves out only one sample at a time. Segmented cross-validation leaves out a whole group of samples at a time. Segmented cross-validation is faster, but segment selection requires some care since it should feature unique information. For example, samples which can be considered as replicates of each other should not be present in different segments.

In the case of an independent test set the file should contain 20–40% of the full data. The calibration and test set must cover the same population of samples as

possible. Replicate measurements should never be present in both the calibration and test sets. This risk exists with a random selection which is proposed in current NIR software. If it is the simplest way to select a test set, it leaves the selection to the computer. Manual selection is recommended since it gives full control over the selection of a test set.

Multiple regression programs also calculate auxiliary statistics, designed to help decide how well the calibration fits the data, and how well it can be expected to predict future samples. For example, two of these statistics are the standard error of calibration (SEC) and the multiple correlation coefficient ($R$). The SEC (also called standard error of estimate, or residual standard deviation) and the multiple correlation coefficient indicate how well the calibration equation fits the data. Their formulas are given in Table 3.

The SEC has the same units as the dependent variables, and it reflects the differences between the instrumental value for the analyte of interest and the reference laboratory value. It expresses the modeling error and cannot be used to estimate future prediction errors. In the absence of instrumental error, SEC is only the measure of reference laboratory error. It is an indication of whether calculation using the calibration equation will be sufficiently accurate for the purposes for which it has been generated. In practice, SEC is less accurate than the reference laboratory, even in the absence of instrumental error, if the wavelength range used as independent variables in the calibration does not account for all interference in the samples or if other physical phenomena are present. The detection limit and sensitivity, or signal-to-noise, defines instrument performance for a specific NIR application. Detection limits can usually be approximated for any NIRS method as equal to three times the SEC for the specified application. Sensitivity for quantitative NIR methods can typically be evaluated as the slope of the calibration line for the concentration of the examined analyte ($y$ axis), versus the change in optical response ($x$ axis), between samples of varying concentration. Sensitivity, from a purely instrumental point of view, is expressed as a signal-to-noise or peak height ratio for a particular compound versus peak-to-peak noise at some absorbance (usually zero). However, in a practical sense, the above considerations do not really matter in NIR spectroscopy. This is due to the fact that the applications developed for practical NIR are mostly based on empirical calibration methods and that the calibration is specific to the problem of interest. It is a characteristic of calibration equations using multivariate mathematical techniques to compensate for the common variations found in noisy chemical samples and imperfect instrumental measurements. This is why properly calculated and validated NIR calibration models are robust and work extremely well.

The multiple correlation coefficient $R$ is a dimensionless measure of how well the calibration fits the data. $R$ can have values between $-1$ and $+1$, but in a calibration situation only positive values exist. A value close to zero indicates that the calibration fails to relate the spectra to the reference values. As the correlation coefficient increases, the spectra become better and better predictors of the reference values. Because the multiple correlation coefficient is dimensionless, it is a useful way of comparing data or results with different units, and that are difficult to compare in other ways. However, its value gives no indication of how well the calibration equation can be expected to perform on future samples.

### 3.15 Some Pitfalls

The first step before any multivariate data analysis is sampling. The selection or preparation of a set of calibration samples is a critical issue. For example, the analyst must collect or prepare samples which span the complete range of constituent concentrations as evenly distributed as possible. It is usual for random sample selection to cause the models to fit most closely to the mean concentration samples. Samples at high or low concentration levels will not influence the slope and intercept in the case of multivariate regression. Ideally, even concentration distribution will allow the model to minimize the residuals at the extremes and at the center with relatively equal weighting. Calibration sets must not only uniformly cover an entire constituent range, they must also be composed of a correctly distributed number of sample types. For example, ideal calibration sets are composed of more than 10–15 samples per analytical term. These samples ideally have widely varying composition evenly distributed across the calibration range. Last but not least, spectroscopic measurements should be performed under as identical conditions as possible between calibration samples and routine samples. A recurrent difficulty is the effect of moisture within a solid or powdered sample. The presence or absence of water in a sample influences the extent of hydrogen bonding, which affects both band position and width in the complete NIR domain. If a calibration model is developed from samples featuring a wide range of the component of interest but a small range in moisture, it will only be useful for samples with that narrow moisture range. Such a calibration may not be robust enough for routine application. To summarize, each calibration problem involves slightly different aspects.

Paying attention to finer details may be obvious, as unexpected error sources can affect the quality of computed models. Errors occur, for example, in sample preparation and measurement. Parameters of potential influence include temperature differences in samples or the instrument while recording the data, calibration sample instability, instrument noise and drift, changes in instrument wavelength setting, nonlinearity, stray light effects, particle size differences, concentration-dependent color differences, residual solvent interaction, and nonhomogeneous samples. The reference method may not measure the same components as the spectroscopic method. Controlling for these aspects may sound overly rigorous and may initially dampen enthusiasm for NIRS. However, multicomponent problems are inherently complex and require the management of several variables simultaneously in order to develop a usable calibration. The ultimate goal of successful calibration is to calculate a mathematical model with the calibration samples which is most sensitive to changes in the modeled parameter and least sensitive to all other noncalibration related factors, such as physical, chemical, and instrumental variables. Every case must be evaluated in terms of the chemical and physical data carried by the calibration samples and the information which the analyst wishes to obtain.

### 3.16 Example Analytical Applications of NIRS

Assuming NIR spectra have been recorded, once a set of pharmaceutical samples has been analyzed with high precision by some analytical reference method, the concentration of an analyte of interest clearly determined or the identity of a given

compound confirmed, the sample data can then be filed into a training set to generate a calibration for subsequent predictions. To form a usable training set, the samples should evenly span the concentration range for the analyte of interest or the expected variations in quality. An obvious pitfall is to develop calibrations that only use sample sets with uneven constituent distributions or too narrow variations of the attribute of interest. In that case the model will most closely fit the dominating samples in the calibration set. The calibration will be highly weighted to the mean value and behave poorly against variations in further samples. Conversely, an ideally evenly distributed calibration set will weight the calibration model equally across the entire concentration range. A properly developed calibration model of this kind will perform most accurately with samples at high and low concentrations.

Despite these sampling issues, which are important for NIR spectroscopists in the pharmaceutical industry, the potential of NIR applications remains intact. NIR is rapid and nondestructive, requiring little or no sample preparation. It can monitor concentrations of several chemical species and physical parameters simultaneously. Its speed and ease are a major bonus in many analytical or monitoring processes. All types of materials are concerned, from solid or liquid raw materials (e.g., excipients), active pharmaceutical ingredients (API), intermediate synthesis products, intermediate formulations in the form of powders, slurries, granulates, or pellets, up to final dosage forms such as capsules, tablets, or lyophilized substances. NIR measurements can be performed in close or direct contact with the sample, in both the laboratory or directly online or inline in the production plant, in order to obtain analytical information rapidly and save time. Transmittance measurements have become an alternative to the conventional reflectance spectroscopy of pharmaceuticals. The important difference is that transmittance NIR samples a volume, whereas reflectance NIR merely samples the surface of solid samples. This has the advantage of more representative values for less homogenous samples like tablets or capsules. On the other hand, more attention has to be paid to sample presentation with respect to stray light and light scattering.

Traditionally, pharmaceutical excipients have been characterized in the laboratory by viscosity, pH in dispersion/solution, water content, ash content, constituent amounts, particle size, and so on. In reality, there are also other subtle properties that are not covered by these parameters but that nevertheless affect the properties of a drug formulation. Many of these variations can be extracted from NIR reflectance spectra in addition to identity testing by other methods (Figure 22). The raw material spectra and results from other analytical methods can then be combined to predict the impact of raw material quality variations on final batch quality. This conformity test by NIR can be achieved by recording numerous raw material batches of known quality and calculating an envelope of acceptability around the mean spectrum. Conform raw material is only qualified when it lies within the threshold values of the envelope at each wavelength. This test would fail poor materials with high impurities and high water levels.

Most pharmaceutical production is performed in batches, both when synthesizing active compounds and manufacturing pharmaceutical formulations. NIR can measure the final state in batch production in terms of spectral similarity, using SIMCA, spectral correlation, or spectral distance. Individual batch development can also be monitored using PCA. Several parameters of the same product can be

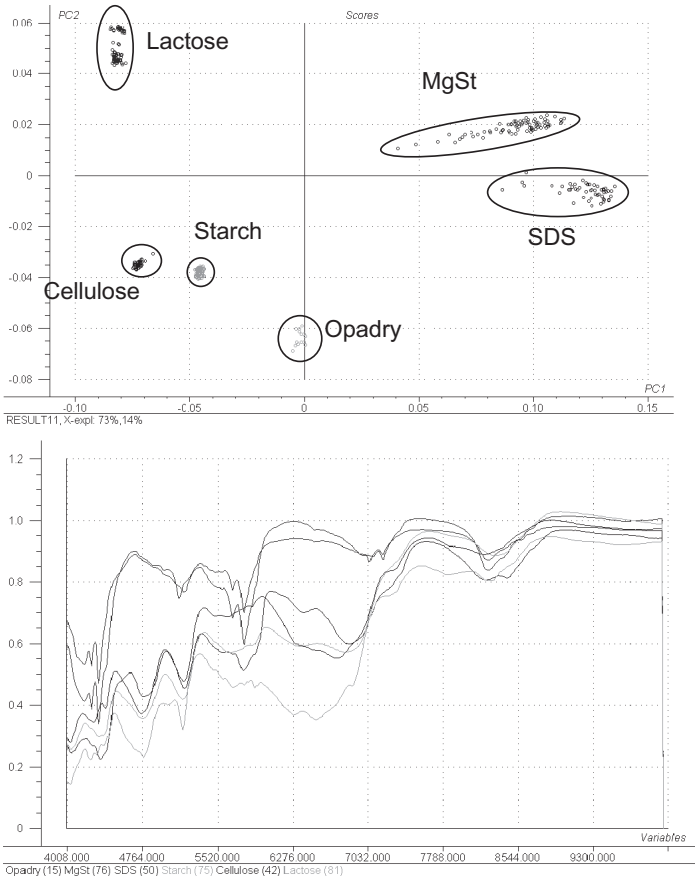


**FIGURE 22** Excipient library with the result of a PCA analysis for identity testing.

determined quantitatively using PLS models. A current basis for comparison includes processing time as a *Y* variable in a PLS model based on several batches considered good for release. Limits for an accepted batch can then be calculated and compared with online NIR measurements during the batch process.

Drying of drug substances is an important step in the production of active materials. It is usually performed in an agitated vacuum dryer, carefully controlled to ensure a suitably dry product and to minimize the risk of overdrying which could damage the high cost material. Some drug substances are hydrates, which present special requirements for dryer control. If the waters of crystallization are removed, the batch may need reprocessing or may even be lost. The traditional means of monitoring and controlling drying operations is to use the slight changes in temperature within the dryer to estimate an endpoint and then to sample the volatile content for laboratory analysis. There is often no distinct temperature endpoint. Some products will also retain residual solvent within the crystals. Sampling vacuum dryers is difficult as the vacuum often has to be released, and the laboratory analysis can take several hours. The use of an NIR fiber-optic probe, inserted directly into the plant scale dryer, can provide real-time analysis of dryer contents and can thus

be considered as a valuable alternative. Differences in physical and chemical properties between polymorphs of an active compound (e.g., solubility, dissolution rate, chemical reactivity, resistance to degradation, bioavailability) are highly significant for the pharmaceutical industry. Effective methods are required not only to control the content in the active compound but also to detect and determine the undesirable form. NIRS can be used in some circumstances to confirm a low percentage of undesirable crystalline state in the amorphous polymorph of a compound. The underlying methods are both qualitative (e.g., classification) and quantitative. The simplicity, expeditiousness, and reliability of the NIR method make it a promising tool for controlling the polymorphic purity of the amorphous phase.

## 3.17 Conclusion

In many pharmaceutical companies, quality control departments already use NIRS to identify formulations. Figure 23 illustrates a PLS calibration for the active content determination in a low-dose tablet. Once identity testing is passed, it is straightforward to consider as a next step the determination of active content in intact tablets. Thus, qualitative and quantitative analysis can be performed by acquiring a single NIR spectrum per sample. Two analytical techniques are replaced by one—nondestructive—NIR measurement. For this purpose near-infrared spectroscopy is a fast and powerful alternative to traditional analysis, which only remains necessary as reference analytics.

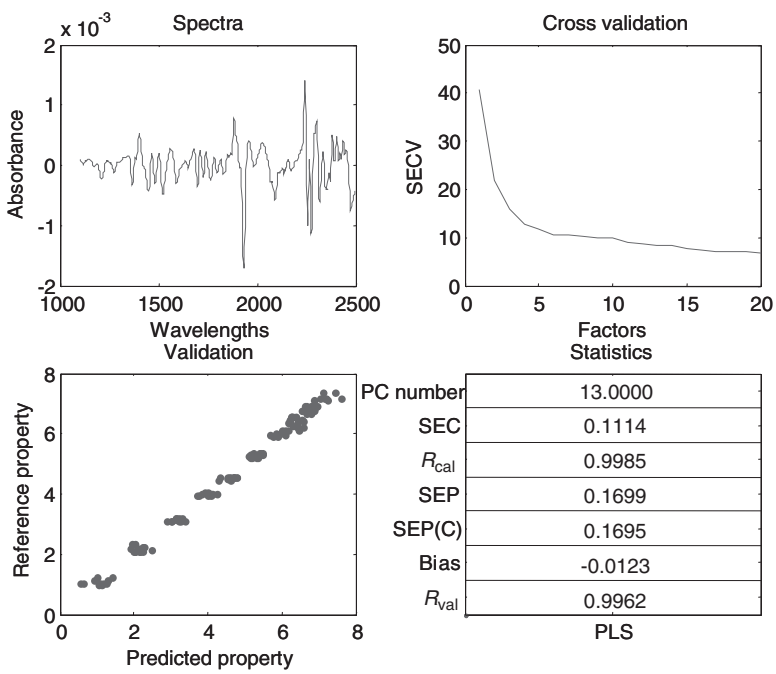


**FIGURE 23** Quantification of the API content in a tablet and overview of the corresponding PLS regression.

The viability of analytical methods obtained by combining vibrational spectroscopy—mainly NIRS—and chemometrics, once validated, can be proved on numerous industrial examples. However, dominant application field of chemometrics is not the pharmaceutical industry, and most accumulated experience is found in the agricultural and food industries. Best practice in chemometrics is independent of the analytical problem or working field. To solve complex pharmaceutical problems by using a large band of spectroscopic methods and at the same time optimizing sample handling, multivariate data analysis algorithms are highly recommended. Thus, education of chemists, pharmacists, and analysts with regards to chemometrics remains a requirement to catch from the beginning the usefulness of multivariate data analysis applied to spectroscopic data.

## BIBLIOGRAPHY


### Section 1

Bakeev, K. A., Ed. (2005), *Process Analytical Technology*, Blackwell, London.

Giodici, P. (2003), *Applied Data Mining. Statistical Methods for Business and Industry*, Wiley, Hoboken, NJ.

*Initiative for Pharmaceutical cGMPs for the 21st Century*, FDA, Washington, DC.

Martin, A. (1993), *Physical Pharmacy*, 4th ed., Lippincot Williams & Wilkins. Philadelphia, PA.

### Section 2

Colthup, N. B., Daly, L. H., and Wiberley, S. E. (1990), *Introduction to Infrared and Raman Spectroscopy*, 3rd ed., Academic.

The Proceedings of NIR-97, *J. Near-Infrared Spectros*., 6 (1–4), NIR Publications, 1998.

Rouessac F., and Rouessac, A. (2000), *Chemical Analysis. Modern Instrumentation Methods and Techniques*, Wiley-VCH, New York.

Siesler, H. W., Ozaki, Y., Kawata, S., and Heise, H. M. Eds (2002), *Near-Infrared Spectroscopy. Principles, Instruments, Applications*, Wiley-VCH, New York.

Williams, P., and Norris, K. (2001), *Near-Infrared Technology in the Agriculture and Food Industries*, 2nd ed. American Association of Cereal Chemists, St Paul, MN.

### Section 3

Brereton, R. G. (2003), *Chemometrics: Data Analysis for the Laboratory and Chemical Plant*, Wiley, Hoboken, NJ.

Kramer, R. (1998), *Chemometric Techniques for Quantitative Analysis*, Marcel Dekker, New York.

Manly, B. F. J. (2000), *Multivariate Statistical Methods. A Primer*, 2nd ed., Chapman & Hall/CRC, New York.

Mark, H. (1991), *Principles and Practice of Spectroscopic Calibration*, Wiley, New York.

Martens, H., and Martens, M. (2001), *Multivariate Analysis of Quality: An Introduction*, Wiley, New York.

Massart, D. L., Vendeginste, B. G. M., Buydens, L. M. C., De Jong, S., Lewi, P. J., and Smeyers-Verbeke, J. (1997/1998), *Handbook of Chemometrics and Qualimetrics, Parts A and B*, Elsevier, Amsterdam.

Naes, T., Isaksson, T., Fearn, T., and Davies, T. (2002), *A User-Friendly Guide to Multivariate Calibration and Classification*, NIR Publications, Chichester.

**Peer Review Articles**

Chalus, P., Roggo, Y., Walter, S., and Ulmschneider, M. (2005), Near-infrared determination of active substance content in intact low-dosage tablets, *Talanta*, 66, 1294–1302.

Inschneider, A. (2002), Validierung von chemometrischen Methoden am Beispiel multivariater Datenanalyse in der Nah-Infrarot Spektroskopie, in *Handbuch der Validierung*, Kromidas, S. Ed., Wiley-VCH, Weinheim, pp. 302–307.

Roggo, Y., Edmond, A., Chalus, P., and Ulmschneider, M. (2005a), Infrared imaging for qualitative analysis of pharmaceutical solid forms and trouble shooting, *Anal. Chim. Acta*, 535, 79–87.

Roggo, Y., Jent, N., Edmond, A., Chalus, P., and Ulmschneider, M. (2005b), Characterizing process effects on pharmaceutical solid forms using near-infrared spectroscopy and infrared imaging, *Eur. J. Pharm. Biopharm.*, 61(1–2), 100–110.

Roggo, Y., Roeseler, C., and Ulmschneider, M. (2004), Near-infrared spectroscopy for qualitative comparison of pharmaceutical batches, *J. Pharm. Biomed. Anal.*, 36, 777–786.

Russell, F., and Ulmschneider, M. (2004), Dissolution testing by means of NIR transmittance spectroscopy, *Am. Pharm. Rev.*, July/August.

Sukowski, L., and Ulmschneider, M. (2005), In-line process analytical technology on qualitative NIR modelling: An innovative approach for the pharmaceutical quality control, *PharmInd*, 67(7), 830–835.

Ulmschneider, M., Barth, G., Reder, B., Vögel, A., and Schilling, D. (2000a), A transferable basic library for the identification of active substances using near-infrared spectroscopy, *Pharm. Ind.*, 62(4), 301–304.

Ulmschneider, M., Barth, G., and Trenka, E. (2000b), Building transferable cluster calibrations for the identification of different solid excipients with near-infrared spectroscopy, *Pharm. Ind.*, 62(5), 374–376.

Ulmschneider, M., and Pénigault, E. (2000a), Assessing the transfer of quantitative NIR-calibrations from a spectrometer to another one, *Analusis*, 28, 83–87.

Ulmschneider, M., and Pénigault, E. (2000b), Direct identification of key-intermediates in containers using Fourier-Transform near-infrared spectroscopy through the protective polyethylene primary packaging, *Analusis*, 28, 136–140.

Ulmschneider, M., and Pénigault, E. (2000c), Non-invasive confirmation of the identity of tablets by near-infrared spectroscopy, *Analusis*, 28, 336–346.

Ulmschneider, M., Wunenburger, A., and Pénigault, E. (1999), Using near-infrared spectroscopy for the non invasive identification of five pharmaceutical active substances in sealed vials, *Analusis*, 27, 854–856.

# CHEMICAL IMAGING AND CHEMOMETRICS: USEFUL TOOLS FOR PROCESS ANALYTICAL TECHNOLOGY

Yves Roggo and Michel Ulmschneider
*F. Hoffmann-La Roche Ltd, Basel, Switzerland*

**Contents**

1 Introduction
2 Definition of Hyperspectral Imaging
3 Hyperspectral Image Acquisition and Instrumentation
  3.1 Principles of Hyperspectral Image Acquisition
  3.2 Spectroscopic Instrumentation
4 Chemometric for Imaging
  4.1 Data Preprocessing
  4.2 Pixel Classification
5 Practical At-Line and Offline Chemical Imaging
  5.1 Practical Tools for Distribution Map Analysis
  5.2 Wavelength Selection and Chemical Interpretation
  5.3 Unsupervised Classification for Process Troubleshooting
  5.4 Supervised Classification, NIR Imaging, and Process Development
  5.5 Further Developments: Online Analysis by Hyperspectral Imaging
6 Conclusions
  References

## 1 INTRODUCTION

Chemical compound homogeneity is an important issue for pharmaceutical solid forms. A classical spectrometer [1–3] integrates spatial information. However, use of a mean spectrum on a surface can be a drawback in solid-form analysis. For example, in the pharmaceutical industry, it is important to map the distribution of active ingredients and excipients in a tablet as this reveals physical interaction

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

between components and helps to solve homogeneity issues—hence the increasing number of spectroscopic imaging studies on the visualization of chemical component homogeneity [4–9].

Vibrational hyperspectral imaging (chemical imaging) is a recent development that combines the chemical information from spectroscopy with spatial information. In principle, it is possible to collect hyperspectral images with single-point detectors, that is, classical mapping with microscopes. However, array detectors measure all pixels simultaneously, reduce recording time, provide uniform background, and improve the signal-to-noise ratio [4]. A complete spectrum is acquired for each pixel, meaning that a hyperspectral data set is in fact a data cube, that is, a three-dimensional (3D) matrix.

This chapter begins by defining hyperspectral imaging, then presents details of instrumentation and image formation, and concludes by describing some chemometric image analysis tools with the help of pharmaceutical examples.

## 2 DEFINITION OF HYPERSPECTRAL IMAGING

This chapter considers three types of digital image. The first is binary (or black and white), where the pixel value can be either 0 or 1. The second is monochromatic (e.g., gray scale), which can be presented as a two-dimensional (2D) $X \times Y$ array describing the distribution of light intensity, where $X$ and $Y$ are the numbers of digitization steps (i.e., pixels) along the two spatial directions. Each pixel has a value between 0 (black) and 255 (white). The third is color, which can be described in the red–green–blue (RGB) space, for example, with three planes (for red, green, and blue, respectively) as a 3D $X \times Y \times 3$ matrix. Each pixel has a value between 0 and 255 for red, green, and blue, generating $255^3$ possible colors for such images (known as 24-bit images).

By analogy, a hyperspectral data set is defined by at least 50 planes; an absorbance map is acquired for each wavelength in the spectral range. If the wavelengths number less than 50, the term *multispectral imaging* is used.

With chemical imaging, a new type of data structure needs to be analyzed. Chemical imaging experiments yield a 3D $X \times Y \times \lambda$ matrix or data cube, where $X$ and $Y$ are the spatial dimensions and $\lambda$ the spectral dimension. One spectrum per pixel is recorded and selection of a wavelength will show an absorbance picture of the sample [10] (Figure 1).

The data cube combines spectral and spatial information and therefore includes the requisite statistics for spectral classifications. However, new chemometric strategies have to be applied to interpret chemical imaging results.

## 3 HYPERSPECTRAL IMAGE ACQUISITION AND INSTRUMENTATION

### 3.1 Principles of Hyperspectral Image Acquisition

This section discusses three modes of acquisition: mapping, array detection, and fiber bundles.

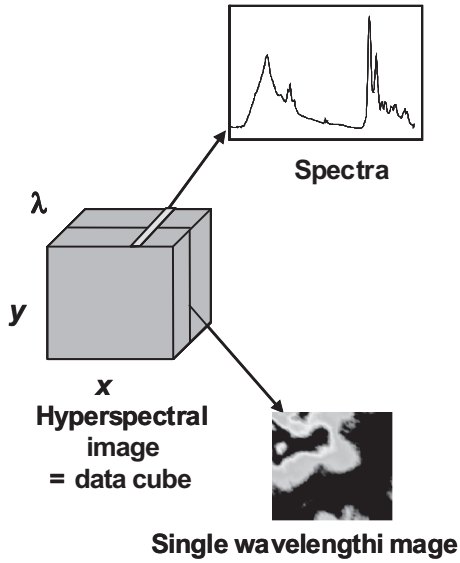


**FIGURE 1** Data structure of chemical image.

***Mapping*** Historically, mapping [11] was the first method used to acquire hyperspectral data cubes, in particular with Raman spectroscopy and infrared (IR) microscopy. The image is created pixel by pixel in a "step-and-acquire" mode: A spectrum is measured at one point of the sample, and then the sample moves to the next measurement position and another spectrum is acquired. The process is iterative for all positions in the area that define the image.

The drawback is the measurement time, which depends on the number of pixels. Spectrometer manufacturers therefore developed line mapping, in which samples are scanned line by line, thereby reducing the acquisition time. These devices were developed for Raman, IR, or near-IR (NIR) spectroscopy (with diode array detectors). However, due to the moving stage, this kind of imaging principle is only suitable for at-line applications.

***Focal Plane Array Detectors*** Focal plane array (FPA) optical detectors are composed of several thousand individual detector elements forming a matrix of pixels. As their name indicates, they are placed in the spectrometer's focal plane. They can be manufactured to be sensitive to ultraviolet (UV), visible (Vis), NIR, or IR radiation. Recent developments in optics have produced cooled and uncooled FPAs with different numbers of pixels from $64 \times 64$ up to $1024 \times 1024$ and different spectral ranges of detection (from 1000 to 12,000 nm). The different types of FPA include those built with indium antimonide (InSb), platinum silicide (PtSi), indium gallium arsenide (InGaAs), and mercury cadmium telluride (HgCdTe) [12]. The mercury cadmium telluride (MCT) detector has become the dominant IR FPA through its entire range coverage.

***Fiber Bundles*** Fiber bundles are used for Raman imaging. Several optical fibers are grouped together, each analyzing a specific sample area [13]. A 3D data cube is

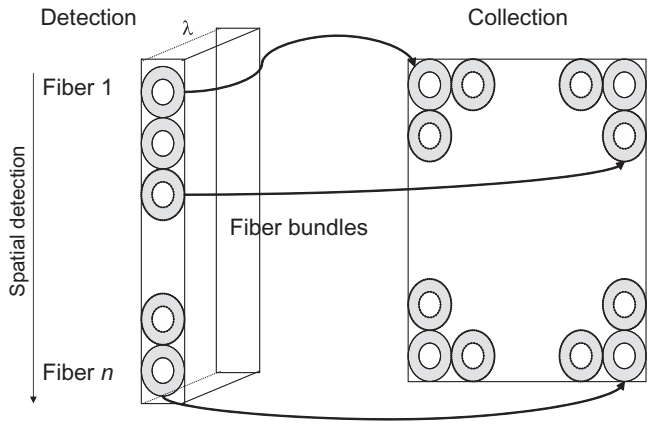


**FIGURE 2** Principle of fiber bundles. (Adapted from A. D. Gift, J. Ma, K. S. Haber, B. L. McClain, and D. Ben-Amotz, *Journal of Raman Spectroscopy*, 30, 757–765, 1999.)

collected using a 2D charge-coupled device (CCD) detector. At the collection point the fibers are arranged in a circular pattern or square array in order to analyze defined surfaces on the samples (Figure 2). At the detection end, the fibers are aligned in order to detect the signal: The first CCD detector dimension is used for spatial information and the second for the spectral region. The technical difficulty is to assign the signal position on the detector end with the pixel position in the image. This technique's main advantage is fast image acquisition.

### 3.2 Spectroscopic Instrumentation

The imaging detectors, whether for point mapping, line scanning, or array detection, can be coupled with different types of spectrometers. Instrument types are classified by wavelength selection modality into imaging Fourier transform (FT) and tunable filter (TF) spectrometers, both of which are presented below, and dispersive spectrometers. FT imaging systems are classical laboratory instruments while TF spectrometers are compact and robust systems for chemical imaging.

***FT Spectrometers*** FT spectrometers (Figure 3) differ from scanning spectrometers by the fact that the recorded signal is an interferogram [14] (see Chapter Role of Preformulation in Development of Solid Dosage Forms). They can be coupled to a microscope or macrochamber with an FPA detector. FT chemical imaging systems (CISs) are available for Raman, NIR, and IR spectroscopy. However, they can only be considered as research instruments. For example, most IR imaging systems are FT spectrometers coupled to microscopes. This type of spectrometer allows the acquisition of spectra in reflection, attenuated total reflection (ATR), or transmission mode.

***TF Systems*** A TF is a device whose spectral transmission can be controlled by applying a voltage or acoustic signal. There are two main TF devices: acousto-optical TF (AOTF), based on diffraction, and liquid crystal TF (LCTF), based on birefringence. An AOTF is a transparent crystal in which an ultrasonic wave field is created,

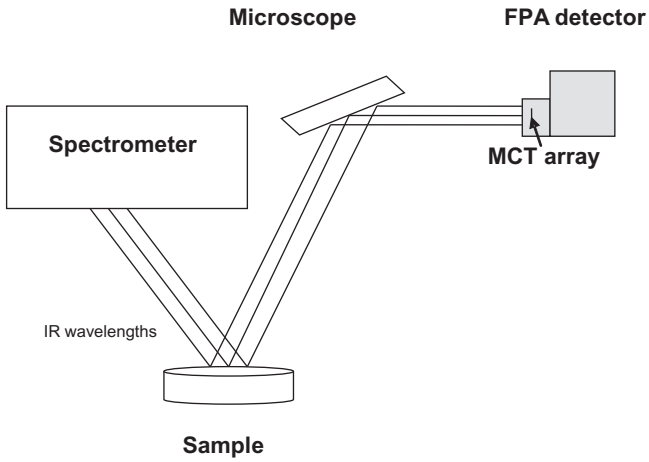


**FIGURE 3** IR microscope linked to FPA detector.

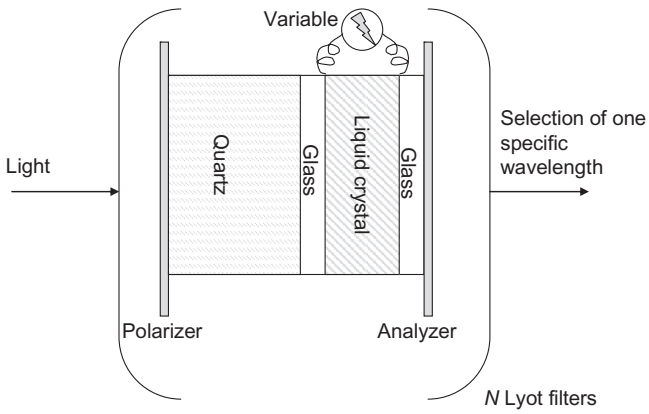


**FIGURE 4** Lyot filter: LCTF element. (Adapted from N. Gat, *Proceedings SPIE*, 4056, 50–64, 2000.)

that is, the selected wavelength is a function of field intensity. An LCTF is built using several Lyot filters (Figure 4): The incoming light is first polarized for one Lyot filter element, then the birefringent crystal introduces a phase difference (δ) between the rays of light, and the light finally passes through the second "analyzer" polarizer, which selects the transmitted wavelength. The disadvantage of an LCTF compared to an AOTF is in speed: AOTFs scan in microseconds, LCTFs in milliseconds [15].

With TFs, samples can be scanned wavelength by wavelength. Single-wavelength images are then grouped into a data cube. Specific wavelengths (e.g., specific for particular chemical compounds) can be selected to reduce acquisition time. Commercial TF devices (Figure 5) are available for NIR and Raman imaging. TF imaging systems appear suitable for process analytical technology (PAT) applications in that they can be installed online due to their fast acquisition times and simple and robust mode of image acquisition.

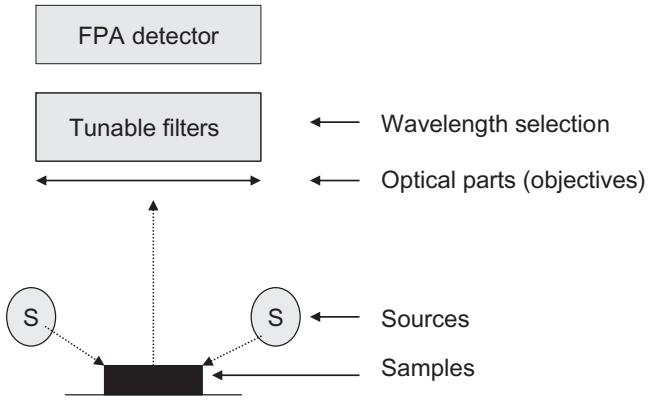


**FIGURE 5** Acquisition by LCTF NIR imaging system. (Adapted from N. Lewis, J. Schoppelrei, E. Lee, and L. Kidder, in *Process Analytical Technology*, Editor K. A. Bakeev Blackwell, London., 2005, pp. 187–225.)

## 4 CHEMOMETRIC FOR IMAGING

Hyperspectral imaging systems generate a large amount of data which has to be processed in order to find the relevant information. Extraction methods must therefore be developed to display images of chemical compound homogeneity ("distribution maps"). The main steps in data cube analysis (Figure 6), presented in sequence below, begin with data preprocessing (to improve information quality), proceed with classification (to identify the main chemical compounds for each pixel and display the distribution maps), and terminate in the deployment of tools for distribution map analysis.

### 4.1 Data Preprocessing

Preprocessing enhances chemical information and removes noise and scattering effects [14]. Its specificity resides in the fact that data cubes can be preprocessed in both the wavelength and spatial dimensions.

Classical spectral preprocessing, described in Section 1, comprises normalization, smoothing, and baseline correction. Some CIS provide intensity as reflectance. The data are therefore converted to absorbance. Another type of spectral preprocessing can be added based on prior chemical knowledge: wavelength selection. The less informative spectral ranges are removed in order to reduce computation time and improve convergence of chemometric algorithms.

Image preprocessing techniques are also useful. Thus bad pixels (e.g., those without signals or outliers) can be removed and several data cubes grouped together for simultaneous analysis in order to simplify comparison. The hyperspectral image can also be masked to select only the regions of interest. Manuals describe other image preprocessing methods, such as spatial smoothing, contrast enhancement, deblurring, and filtering [16].

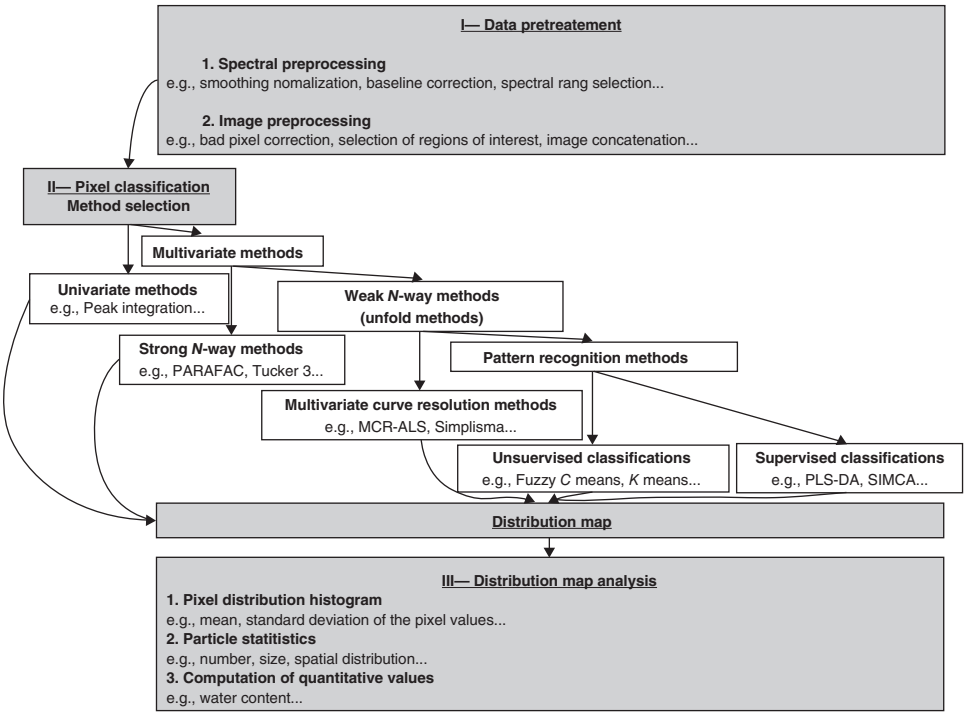


**FIGURE 6** Image analysis workflow: processing chain.

## 4.2 Pixel Classification

After data preprocessing, classification is a critical step in distribution map extraction. Various algorithms can be applied. Although an exhaustive list is beyond the scope of this handbook, methods can be grouped to give an overview of data cube processing solutions (Figure 6). Chemometric techniques are grouped by dichotomy according the following criteria: univariate or multivariate analysis, strong or weak *N*-way methods, multivariate curve resolution or pattern recognition techniques, and supervised or unsupervised classifications.

***Univariate and Multivariate Methods*** Univariate methods belong to classical spectroscopy: Depending on the chemical structure of the compound, a specific wavelength is selected to compute the distribution map. Peak height, area, and ratio between two peaks are used to display a false color picture in which the lowest value is displayed in black or blue and the highest in white or red. Univariate methods are the simplest ways of obtaining distribution maps. However, selection of compound-specific wavelengths can be difficult, especially when samples are complex mixtures. The advantage of hyperspectral imaging is that full spectra are available in the data cube. Drawing on all the information contained in the data sets rather than on a few wavelengths improves chemical map extraction. The techniques used in this case are termed multivariate. Multivariate image analysis (MIA) based on chemometrics generally improves distribution map extraction. Several types of MIA

methods are presented below: strong *N*-way, multivariate curve resolution, and pattern recognition methods.

***Multivariate Image Analysis: Strong and Weak Multiway Methods*** Strong and weak *N*-way methods analyze 3D and 2D matrices, respectively. Hyperspectral data cube structure is described using chemometric vocabulary [17]. A two-way matrix, such as a classical NIR spectroscopy data set, has two modes: O object (matrix lines) and V variables (matrix columns). Hyperspectral data cubes possess two object modes and one variable mode and can be written as an OOV data array because of their two spatial directions.

*Strong Multiway Methods* Strong multiway methods analyze data cubes directly, without any matrix rearrangement, whereas weak multiway methods require a prior unfolding step. Examples of strong multiway techniques used for hyperspectral imaging analysis include parallel factor analysis (PARAFAC) and Tucker 3 [three-way principal-component analysis (PCA)] [18]. Their main advantage is that they take into account the correlations between image pixels in the OO modes: Unfold methods do not use pixel spatial proximity, resulting in neglect of some data cube information.

Even if strong *N*-way methods are used to reduce image noise, compress data, and improve data cube visualization, weak multiway methods are more often used as they facilitate classification using classical single-point spectra.

Classical chemometric methods, that is, the classification and regression presented in Section 1, are also applied to hyperspectral images. However, $X \times Y \times \lambda$ matrices have to be unfolded into $(X \times Y) \times \lambda$ matrices before processing. In other words, the three-way OOV array is unfolded into a classical two-way OV matrix.

*Weak Multiway Methods* Figure 7 shows the three steps in weak N-way analysis: Unfold the data cube, perform the selected chemometric methods, and refold the matrix in order to display distribution maps. Weak *N*-way analysis comprises two main variants:

*Multivariate curve resolution* (MCR) has been in common use for 30 years in the analytical chemistry community [high-performance liquid chromatography (HPLC), FTIR, UV, NIR, and Raman]. It refers to self-modeling mixture analysis. Of the various methods, simple-to-use interactive self-modeling mixture analysis (SIMPLISMA) [19] and MCR alternating least squares (MCR-ALS) [20] appear the most successful with hyperspectral imaging. Their aim is to extract the chemical compound natures (so-called pure spectra) and concentration profiles from multicomponent systems. Their main advantage is that they are calibration free, that is, no prior knowledge is required.

*Pattern recognition* can be classified according to several parameters. Below we discuss only the supervised/unsupervised dichotomy because it represents two different ways of analyzing hyperspectral data cubes. Unsupervised methods (cluster analysis) classify image pixels without calibration and with spectra only, in contrast to supervised classifications. Feature extraction methods [21] such as PCA or wavelet compression are often applied before cluster analysis.

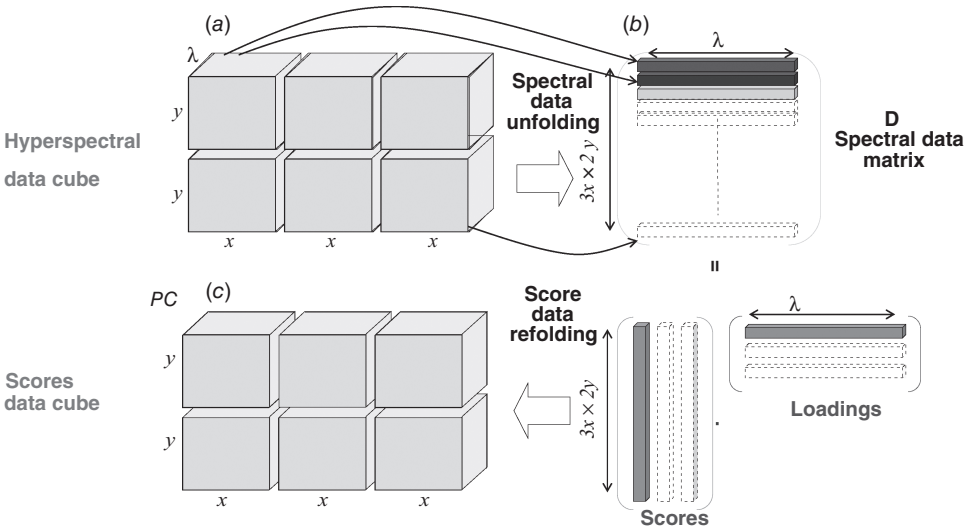


**FIGURE 7** Application of weak *N*-way methods: (*a*) preprocessing; (*b*) unfold matrix analysis; (*c*) refolding. Multivariate curve resolution and pattern recognition techniques.

PCA is used to extract features and visualize data. The most important PCA application is the reduction in the number of wavelengths.

Examples of nonhierarchical clustering [22] methods include Gaussian mixture models, *K* means, and fuzzy *C* means. They can be subdivided into hard and soft clustering methods. Hard classification methods such as *K* means assign pixels to membership of only one cluster whereas soft classifications such as fuzzy *C* means assign degrees of fractional membership in each cluster.

Supervised classifications use spectral and class membership information [23]. Mathematical models are first computed with a calibration set containing spectral and class information. This model is then applied to predict new sample classes. Supervised pattern recognition algorithms fall into three main categories of contrasting pairs. The first comprises methods based on discrimination [e.g., linear discriminant analysis (LDA)] and those emphasizing similarity within a class [e.g., soft independent modeling of class analogy (SIMCA)]. The second comprises linear and nonlinear methods such as neural methods. The third comprises parametric and nonparametric computations. In parametric techniques such as LDA, the decision rules use statistical parameters of normal sample distribution. In summary, classical supervised classification methods are distance-based LDA, SIMCA, and PLS discriminant analysis (PLS-DA), which performs regression between spectra and class memberships.

Reference spectra choice is critical when applying supervised pattern recognition methods. The first solution is to use pure compound spectra as references. The drawback is that mixture spectra in data cubes often differ from the reference spectra. Applying the model may therefore give wrong results. The second solution, suitable in a few studies, is to select image pixels where only one compound is present in order to obtain the calibration sets.

## 5 PRACTICAL AT-LINE AND OFFLINE CHEMICAL IMAGING

Having reviewed the methods, we will now illustrate their application to pharmaceutical situations, with particular respect to univariate methods, PCA, and supervised classification (PLS-DA).

### 5.1 Practical Tools for Distribution Map Analysis

***Image Comparison*** A solution for distribution map analysis is to compare images. References such as pure compounds or original samples can be included in the image. For example, references and samples can be measured simultaneously if the field of view is large enough; otherwise two data cubes can be concatenated, that is, grouped together. After the distribution map has been extracted, it can be readily interpreted simply by image comparison. In the example shown in Figure 8, the aim was to detect counterfeits. The original samples were compared with the suspected counterfeit. After PCA extraction, the differences between the two groups were clearly detected. The counterfeits had no active pharmaceutical ingredient (API) and the excipients were not identical. A self-calibrating comparison was then performed with NIR imaging for fast counterfeit detection.

***Pixel Distribution and Particle Size Determination*** Quantitative parameters can also be computed to analyze a distribution map. The first tool (Figure 9) is to display the pixel histogram and calculate classical statistics such as mean, standard

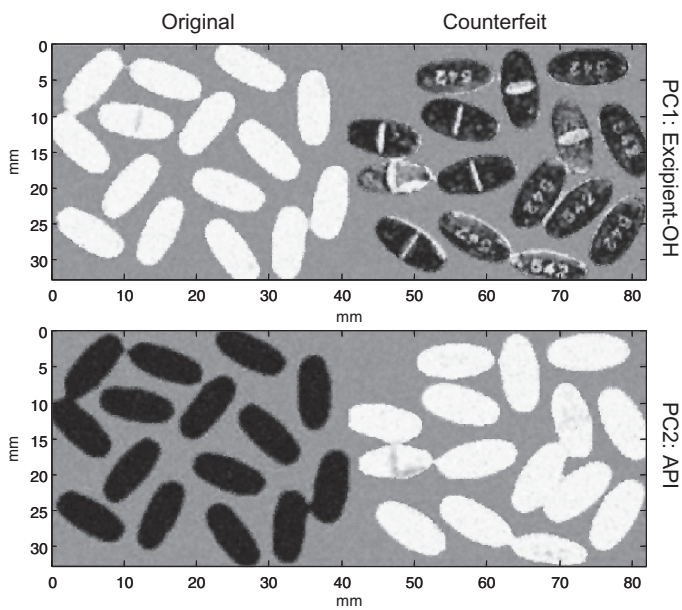


**FIGURE 8** Self-calibrating image comparison for counterfeit identification: score images. White: higher score. Black: lower score.

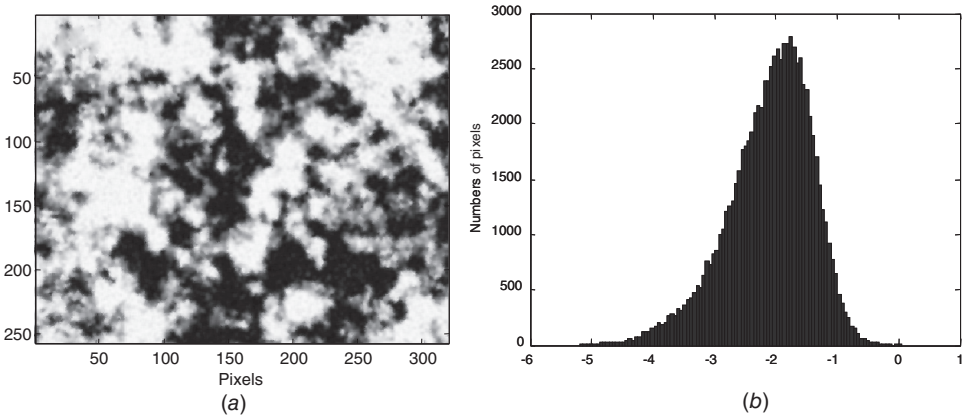


**FIGURE 9** Image and associated pixel histogram (white: higher absorbance, black: lower absorbance).

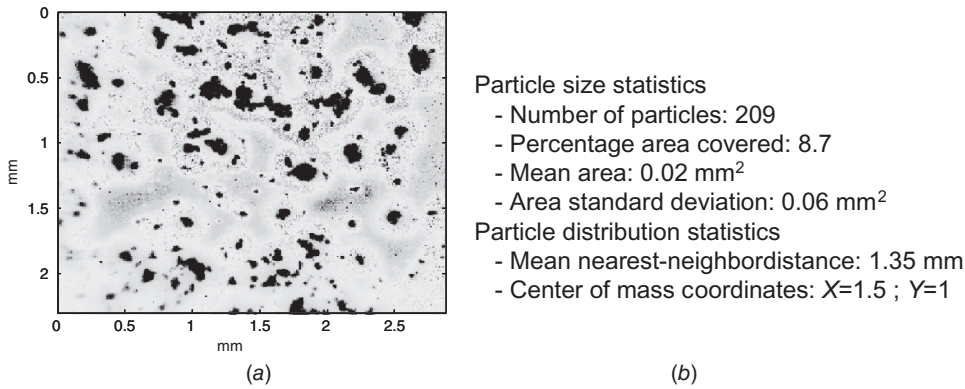


**FIGURE 10** Particle size computation with imaging techniques: (*a*) powder NIR image (particles are in black); (*b*) statistics.

deviation, and the parameters of normal distribution, that is, skew and kurtosis. For example, a higher mean may be explained by a higher content of a chemical compound, and a lower standard deviation may be synonymous with greater homogeneity. The second tool is to compute particle size and obtain information about spatial particle distribution (Figure 10). Several quantitative parameters are obtained, such as particle number, particle size, and percentage area covered by the particles. Thus the particle distribution statistics provide a measure of sample homogeneity.

### 5.2 Wavelength Selection and Chemical Interpretation

The common aim of the Raman and NIR spectroscopy examples below was to display the API and excipient localization on the tablet surface in order to assess distribution homogeneity and characterize the solid-state properties of the API.

**Example 1** API Mapping and Raman Spectroscopy

**Method**

A Raman microscope (Renishaw, 785-nm laser, spectral range 800–100 $cm^{-1}$) with a line-mapping detector (21 pixels/line) was used to analyze a solid dosage form. The image size was 105 × 88 pixels, that is, 325 μm × 270 μm, and acquisition time was about 40 min. Spectra were smoothed and normalized. Peak heights were determined for the three main compounds—API, lactose, and cellulose (Figure 11)—in order to create distribution maps.

**Results and Discussion**

Raman bands are narrow and specific peak selection is simple. One of the simplest classification methods, that is, univariate peak height, could thus be successfully applied. The drawbacks of Raman spectroscopy are long acquisition time and the effect of line scanning in the images (Figure 11). Another issue is fluorescence. Some chemical compounds fluoresce (e.g., cellulose), obscuring the relevant chemical signal and making the distribution map more difficult to extract.

Thus Raman imaging is a useful tool for detecting small API particles on the surface of pharmaceutical solid forms. It may even be the most suitable chemical imaging technique for API mapping due to its low spatial resolution (up to 0.5 μm/pixel) and the polymorphism of the spectral information.

**Example 2** Tablet Reconstruction by NIR Imaging

**Method**

The tablet was cut lengthwise using a trimmer to leave a plane surface with the tablet coating removed (Figure 12). Sample and references were analyzed using a chemical imaging NIR spectrometer (Sapphire, Malvern) with the following acquisition parameters: detector size 320 × 256 array, spectral range 1100–2450 nm, and spatial resolution 40 μm/pixel. Acquisition time was about 5 min.

**Results and Discussion**

After the second derivative, wavelengths could be selected giving contrast images and displaying the localization of mannitol, API, and crospovidone (Figures 13 and 14) with the NIR images obtained at specific wavelengths. In this example, chemometrics can help define concentration maps by being able to use the information present in the whole spectra and not only at specific wavelengths. For example, the MCR-ALS technique was able to extract five compounds whereas wavelength selection could only display three components due to peak overlapping in the NIR range. However, simpler methods are also fast and easy to use. Because the aim was to localize the API, the wavelength selection method provided the expected results.

Images were displayed at specific wavelengths (peak-height method). The single-wavelength images were then binarized, in an operation similar to the transformation of a gray-scale image into a black-and-white image. A color image resulted (Figure 14). The red channel was associated with the API, the blue channel with mannitol, and the green channel with crospovidone. Tablet reconstruction was then possible. This highlights the main advantage of the spectroscopic technique: the large area of analysis, meaning that the images are more representative of the sample.

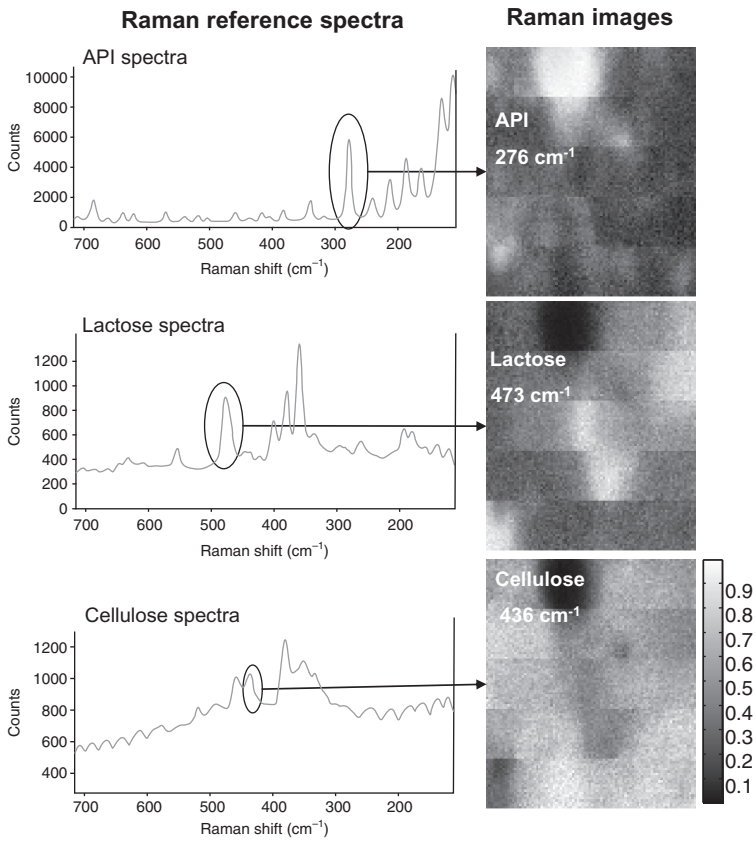


**FIGURE 11** Raman images at specific wavelengths and reference spectra (image size 325 μm × 270 μm; white: higher absorbance, black: lower absorbance).

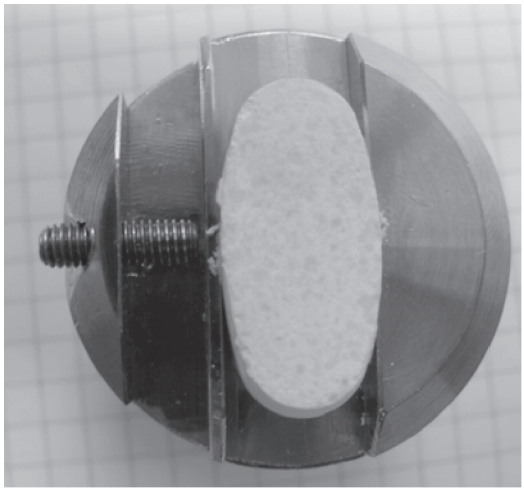

**FIGURE 12** Sample and cutting holder.

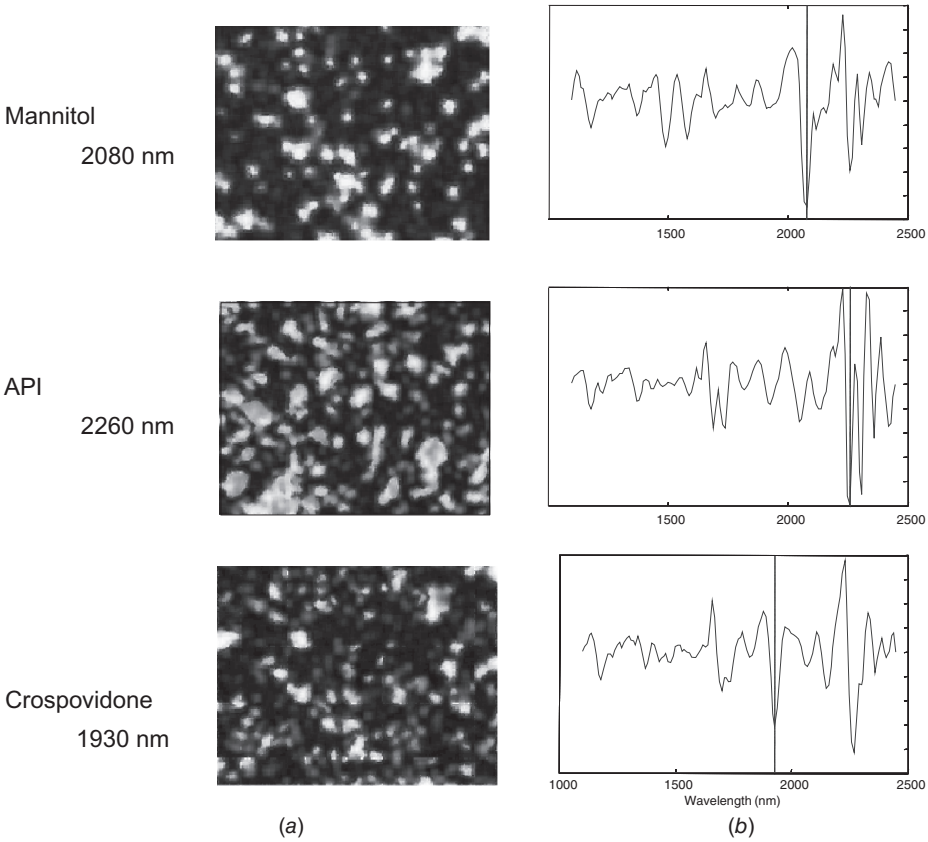


**FIGURE 13** Images at component-specific wavelengths and second derivative spectra of high-absorption pixel (image size: 1.02 cm × 1.3 cm; white: higher absorbance, black: lower absorbance).

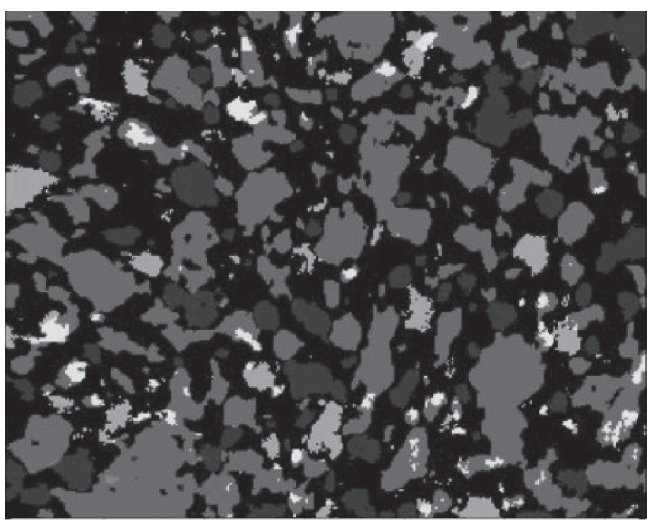

**FIGURE 14** Tablet reconstruction. Gray (normally red): API. Dark gray (normally blue): mannitol. Light gray (normally green): crospovidone. Black: other. Image size 1.02 cm × 1.3 cm.

### 5.3 Unsupervised Classification for Process Troubleshooting

The aim of the example was to compare tablets with respect to "good" and "bad" dissolution properties using IR imaging and to identify the root causes of bad dissolution [24].

***Methodology*** Six samples were analyzed: three passed the dissolution test (good samples) and three failed (bad samples). Several ingredients were used as references: Avicel (cellulose), the API, magnesium stearate, and poloxamer. Dissolution was tested after NIR imaging measurement to obtain the wet chemical references.

An Equinox 55 spectrometer coupled to a Hyperion 3000 microscope equipped with a 64 × 64 MCT FPA detector (Bruker, Ettlingen, Germany; www.brukeroptics.com) was used to acquire IR spectra between 3900 and 900 $cm^{-1}$ at 16 $cm^{-1}$ resolution (i.e., 376 data points) under $N_2$ purge. The pixel-grouping binning function was applied to improve the signal, and a 16 × 16 pixel image was finally acquired. The number of scans was 20 and the area of analysis per FPA measurement was 270 μm × 270 μm. Two FPA images were recorded per tablet, that is, a 32 × 16 pixel image per tablet. The three data cubes of the good samples were concatenated in the $Y$ dimension to obtain a 32 × 48 image. The same 32 × 48 image was produced for the bad samples. The two sets of good and bad samples were then concatenated in the $X$ dimension to obtain a 64 × 48 image. The final data cube was a 64 × 48 × 376 matrix.

***Results and Discussion*** In the PCA results (Figures 15 and 16), loadings were interpreted using raw material reference spectra. The first loading was attributed to poloxamer, the second to magnesium stearate (Figure 16), the third to the API, and the fourth to Avicel. However, the other ingredients could also have contributed to the loadings because PCA loadings are not pure component spectra. PCA extracts orthogonal signals, which is not the case for reference spectra. In particular, the PC3 loading attributed to the API could have been contaminated by other components, especially magnesium stearate. This confirms the drawback of feature extraction methods.

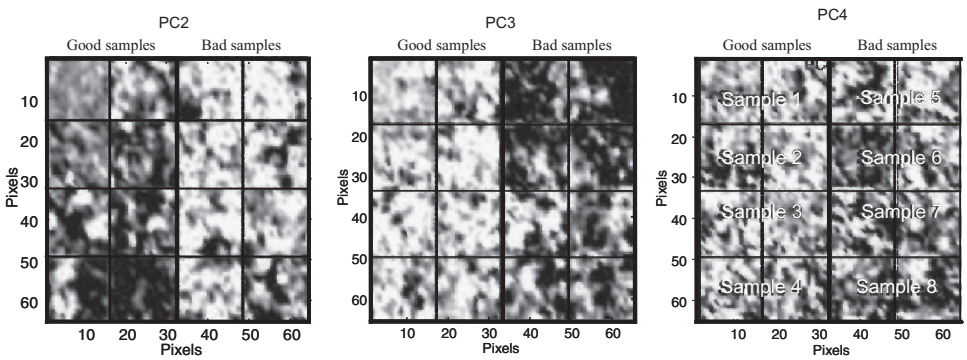


**FIGURE 15** PCA score images: PC2: magnesium stearate; PC3: API; PC4: Avicel; PC1: poloxamer, not discriminant and not displayed (white: higher absorbance, black: lower absorbance).

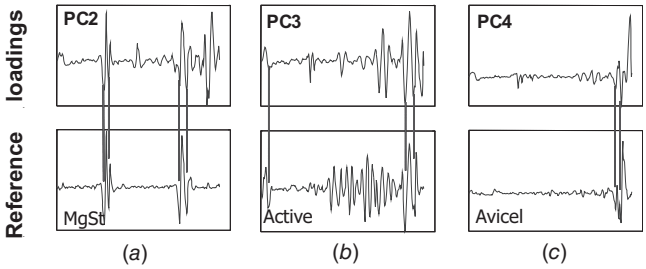


**FIGURE 16** PCA loadings and comparison with reference spectra: (*a*) PC2, magnesium stearate; (*b*) PC3, active ingredient; (*c*) PC4, Avicel.

This PCA image study showed the differences between the two data sets, that is, it separated the good and bad dissolution samples. The main differences were due to the distribution of magnesium stearate and the API. No differences were observed in the spatial distribution of poloxamer or Avicel. Magnesium stearate is hydrophobic, thus protecting the tablet core from moisture and hence slowing dissolution. When a sample had more active ingredient on the surface, the dissolution properties were increased.

Peak height is a classical method of interpreting IR spectra. Its main advantage is in the selection of specific wavelengths. The disadvantage is that the bands often overlap and finding a specific band becomes problematic. PCA solves the problem of wavelength selection. Its main advantage is that it reduces the number of variables, that is, the number of images to analyze. The disadvantage lies in the interpretation of the loadings, which differ from pure substance spectra. Interpretation can be complicated by the fact that several chemical species may contribute to one loading.

### 5.4 Supervised Classification, NIR Imaging, and Process Development

The aim of this study was to use NIR imaging to solve granulation issues in new formulation development. Undesired powder agglomerations developed during the granulation step (Figure 17). Imaging was applied to characterize the agglomeration structure.

***Method*** The sample contained starch, API, Avicel, crospovidone, and sodium lauryl sulfate. Sample and references were analyzed in triplicate by NIR imaging (Spectral Dimensions, 20 coadds, spectral range 1100–2450 nm). Full image size was $320 \times 256$ pixels or $4.1 \times 3.3$ mm. The NIR images were interpreted, and sample raw materials mapped, using PLS classification with five loadings (based on the reference spectra for starch, API, Avicel, crospovidone, and sodium lauryl sulfate).

***Results and Discussion*** The PLS model identified all five chemical species. PLS multivariate analysis showed (Figure 18) that the core contained Avicel and API and that the periphery contained starch and crospovidone. The solution to the granulation issue was to add a premixing step to avoid agglomeration. NIR imaging proved useful for improving process understanding.

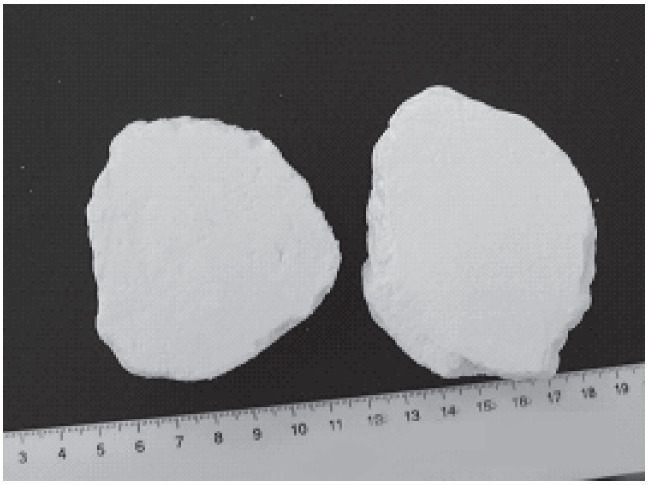

**FIGURE 17** Powder agglomeration: visible picture.

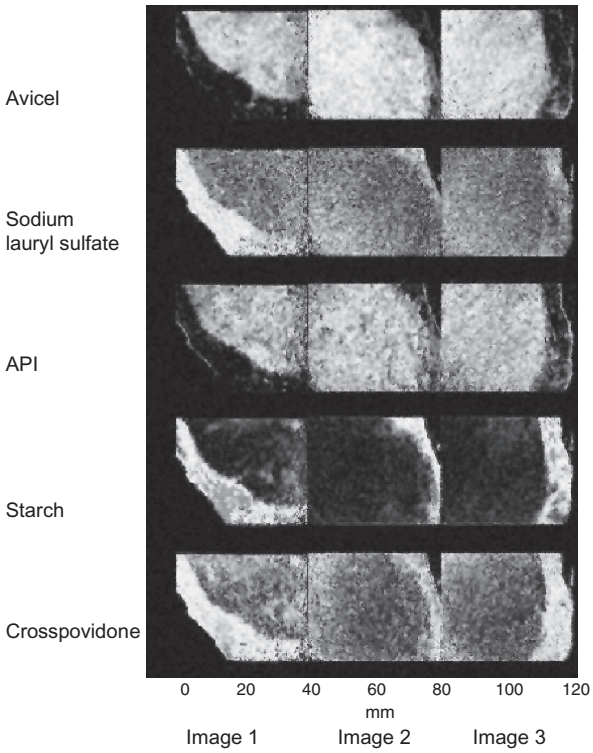


**FIGURE 18** PLS classification image (three replicates and five chemical compounds). Simulation of granulation issues.

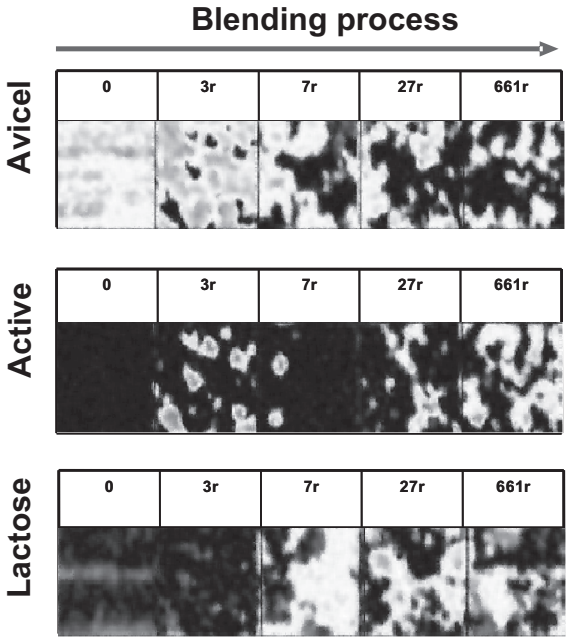


**FIGURE 19** IR imaging for blend monitoring (r = blender rotation; white: higher absorbance, black: lower absorbance).

The advantage of supervised classification is that it avoids a wavelength selection step or the interpretation of PCA loading. It also rapidly extracts several chemical components. These methods provide accurate results provided the sample spectra are similar to the reference spectra. In our case the powder agglomeration was heterogeneous and the layers had a high content of excipient, making it possible to apply supervised classification.

### 5.5 Further Developments: Online Analysis by Hyperspectral Imaging

The two main online applications of hyperspectral imaging are blending endpoint determination (Figure 19) and capsule control (Figure 20). Other applications may be possible, such as content uniformity, determination of dissolution properties, and water quantification, but are not described in this chapter.

In blending, the main imaging advantage over single-point spectroscopy is the ability to analyze a larger area. We studied blends of three compounds (Figure 19), selecting single wavelengths for Avicel, API, and lactose. Only Avicel was present in the sample at the beginning of the experiment. Avicel and API were detected after three blender rotations; the two maps were complementary. All compounds in the mixture were homogeneous after 27 rotations. Timed distribution map analysis is thus a blending monitoring tool. Our example shows the IR images. However, NIR imaging has proved highly successful in the literature [25].

In capsule control, NIR radiation penetrates the shell, enabling the filling to be checked. The capsules in our example (Figure 20) were blue and opaque. However,

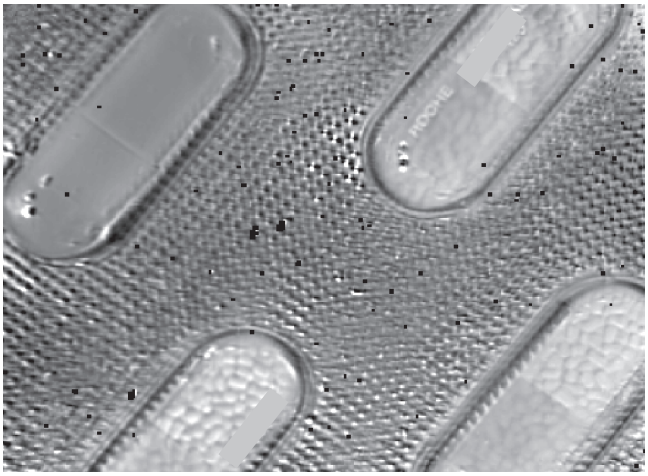

**FIGURE 20** Capsule control by NIR imaging (image size 33 mm × 41 mm). Simulation of empty capsule in blister.

**TABLE 1 Instrumentation Comparison**

| Parameters | FT System | Dispersive | Liquid Crystal Tunable Filter |
|---|---|---|---|
| **Spectroscopy** | NIR, IR, Raman | Raman, NIR | Raman, NIR |
| **Image acquisition** | Mapping, FPA detector | Mapping | FPA detector |
| **Measurement mode** | Reflexion, transmission, ATR | Reflexion, transmission | Reflexion |
| **Image size** | Microscopic, macroscopic | Microscopic | Microscopic, macroscopic |
| **Acquisition time** | Slow | Slow | Fast |

NIR imaging was able to detect empty capsules and monitor fillings, including pellet appearance.

## 6 CONCLUSIONS

Hyperspectral imaging provides information that is spatial and spectral and both qualitative and quantitative. It can map chemical compound distribution and determine particle size. Such information cannot be obtained using classical spectroscopy. We have applied the method to the solution of quality control and process problems affecting pharmaceutical tablets: dissolution, polymorph distribution, moisture content determination, API localization and characterization, content uniformity, blending, and granulation. The choice of imaging technique is based on several criteria: spatial and spectral resolution, measurement time, and wavelength range. Tables 1 and 2 summarize the main advantages and disadvantages of the different methods.

**TABLE 2 Comparison of Chemical Imaging Methods**

| Parameters | Raman Imaging | Infrared Imaging | Near-Infrared Imaging |
|---|---|---|---|
| **Instrumentation robustness** | LCTF: +++<br>other: + | + | +++ |
| **Spectral information specificity** | +++ | +++ | ++ |
| **Constituant map extration** | Easy | Easy | Need chemometrics |
| **Pharmaceutical application examples** | Polymorph screening, small API particle detection | Unknown particle identification | Tablet reconstruction, blend homogeneity, process troubleshooting, counterfeit identification |

Spectral imaging is a complex and multidisciplinary field. The introduction of new FPAs is making it increasingly powerful and attractive. It has proven potential in qualitative pharmaceutical analysis and can be used when spatial information becomes relevant for an analytical application. Even if online applications and regulatory method validation require further study, the potential contribution of imaging to quality control and PAT needs no further demonstration.

## ACKNOWLEDGMENTS

We would like to thank Anton Fischer (QC Manager, Hoffmann-La Roche) and Rolf Altermatt (Section Head, Hoffmann-La Roche) for providing the resources that made this study possible and Christelle Gendrin (Ph.D. student, Hoffmann-La Roche) for her greatly appreciated help.

## REFERENCES

1. Roggo, Y., Roeseler, C., and Ulmschneider, M. (2004), Near infrared spectroscopy for qualitative comparison of pharmaceutical batches, *J. Pharm. Biomed. Anal.*, 36, 777–786.
2. Chalus, P., Roggo, Y., Walter, S., and Ulmschneider, M. (2005), Near-infrared determination of active substance content in intact low-dosage tablets, *Talanta*, 66, 1294–1302.
3. Roggo, Y., Duponchel, L., and Huvenne, J.-P. (2003), Comparison of supervised pattern recognition methods with McNemar's statistical test: Application to qualitative analysis of sugar beet by near-infrared spectroscopy, *Anal. Chim. Acta*, 477, 187–200.
4. El-Hagrasy, A. S., Morris, H. R., D'Amico, F., Lodder R. A., and Drennen, J. K. (2001), Near-infrared spectroscopy and imaging for the monitoring of powder blend homogeneity, *J. Pharm. Sci.*, 90, 1298–1307.
5. Beleites, C., Steiner, G., Sowa, M. G., Baumgartner, R., Sobottka, S., Schackert, G., and Salzer, R. (2005), Classification of human gliomas by infrared imaging spectroscopy and chemometric image processing, *Vibrat. Spectrosc.*, 38, 143–149.

6. Chalmers, J. M., Everall, N. J., Schaeberle, M. D., Levin, I. W., Neil Lewis, E., Kidder, L. H., Wilson, J., and Crocombe, R. (2002), FT-IR imaging of polymers: An industrial appraisal, *Vibrat. Spectrosc.*, 30, 43–52.
7. Juan, A. D., Tauler, R., Dyson, R., Marcolli, C., Rault, M., and Maeder, M. (2004), Spectroscopic imaging and chemometrics: A powerful combination for global and local sample analysis, *TrAC Trends Anal. Chem.*, 23, 70–79.
8. Lewis, E. N., Schoppelrei, J., and Lee, E. (2004), Near-infrared chemical imaging and the PAT initiative, *Spectroscopy*, 19, 26–36.
9. Lewis, E. N., Lee, E., and Kidder, L. H. (2004), Combining imaging and spectroscopy: Solving problems with near infrared chemical imaging, *Microscopy Today*, 12, 8–12.
10. Lewis, N., Schoppelrei, J., Lee, E., and Kidder, L. (2005), Near-infrared chemical imaging as a process analytical tool, in Bakeev, K. A., Ed., *Process Analytical Technology*, Blackwell, London, pp. 187–225.
11. Lewi, P. J. (2005), Spectral mapping, a personal and historical account of an adventure in multivariate data analysis, *Chemomet. Intell. Lab. Syst.*, 77, 215–223.
12. Tran, C. D. (2003), Infrared multispectral imaging: Principles and instrumentation, *Appl. Spectrosc. Rev.*, 38, 133–153.
13. Gift, A. D., Ma, J., Haber, K. S., McClain, B. L., and Ben-Amotz, D. (1999), Near-infrared Raman imaging microscope based on fiber-bundle image compression, *J. Raman Spectrosc.*, 30, 757–765.
14. Burns, D. A., Ciurczak, E. W. (2001), *Handbook of Near-Infrared Analysis*, 2nd ed., rev. and expanded, CRC Press, New York.
15. Gat, N. (2000), Imaging spectroscopy using tunable filters: A review, *Proc. SPIE*, 4056, 50–64.
16. Russ, J. (2002), *The Image Processing Handbook*, 4th ed., CRC Press, London.
17. Huang, J., Wium, H., Qvist, K. B., and Esbensen, K. H. (2003), Multi-way methods in image analysis—Relationships and applications, *Chemomet. Intell. Lab. Syst.*, 66, 141–158.
18. Bro, R. (2003), Multivariate calibration: What is in chemometrics for the analytical chemist? *Anal. Chim. Acta*, 500, 185–194.
19. De Braekeleer, K., and Massart, D. L. (1997), Evaluation of the orthogonal projection approach (OPA) and the SIMPLISMA approach on the Windig standard spectral data sets, *Chemomet. Intell. Lab. Syst.*, 39, 127–141.
20. Tauler, R. (1995), Multivariate curve resolution applied to second order data, *Chemomet. Intell. Lab. Syst.*, 30, 133–146.
21. Naes, T., and Martens, H. (1991), *Multivariate Calibration*, Wiley, New York.
22. Massart, D. L., Vandeginste, B. G. M., Buydens, L. M. C., Jong, S. D., Lewi, P. J., and Smeyers-Verbeke, J. (1997), *Handbook of Chemometrics and Qualimetrics: Part A, Data Handling in Science and Technology, 20A*, Elsevier, Amsterdam.
23. Massart, D. L., Vandeginste, B. G. M., Deming, S. M., Michotte, Y., and Kaufman, L. (2003), *Chemometrics: A Textbook, Data Handling in Science and Technology*, 2, Elsevier, Amsterdam.
24. Roggo, Y., Edmond, A., Chalus, P., and Ulmschneider, M. (2005), Infrared hyperspectral imaging for qualitative analysis of pharmaceutical solid forms, *Anal. Chim. Acta*, 535, 79–87.
25. Lyon, R. C., Lester, D. S., Lewis, E. N., Lee, E., Yu, L. X., Jefferson, E. H., and Hussain, A. S. (2002), Near-infrared spectral imaging for quality assurance of pharmaceutical products: Analysis of tablets to assess powder blend homogeneity, *AAPS PharmSciTech* [Electronic Resource], 3, E17.

# PERSONNEL TRAINING IN PHARMACEUTICAL MANUFACTURING

David A. Gallup, Katherine V. Domenick, and Marge Gillis
*Training and Communications Group, Inc., Berwyn, Pennsylvania*

**Contents**

1 Overview
- 1.1 Background
- 1.2 Training Requirements
- 1.3 Good Training Practices and Pharmaceutical Manufacturing
- 1.4 What Is Competency-Based Training?
- 1.5 Why Is Competency-Based Training So Important?

2 Developing a Training Plan: Strategy to Ensure Training Compliance in a Pharmaceutical Manufacturing Facility
- Section 1 Training Organization
- Section 2 Training Programs
- Section 3 Training Development
- Section 4 Training Implementation
- Section 5 Training Recordkeeping

References

## 1 OVERVIEW

### 1.1 Background

By law and by ethical commitment, companies that manufacture pharmaceutical products must make sure that what they produce is safe and effective. Ensuring that pharmaceutical manufacturing personnel possess the competencies necessary to perform their jobs correctly and efficiently is critical to a safe and successful

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

manufacturing process. Developing required skills, providing discrete knowledge, and instilling an ethical and responsible approach to work are critical to training in an environment centered on good manufacturing practices.

This chapter includes the following types of pharmaceutical manufacturing organizations:

- Active pharmaceutical ingredient (API) or bulk manufacturers
- Biotechnology manufacturers
- Traditional manufacturers, solid and liquid dose
- Vaccine manufacturers

This chapter reviews the requirements for training personnel in a pharmaceutical manufacturing environment; it then focuses on how to develop and implement a training strategy that ensures pharmaceutical manufacturers are in compliance with the mandated *Code of Federal Regulations* (CFR), the current Food and Drug Administration (FDA) quality systems draft guidance document and good training practices (GTPs).

## 1.2 Training Requirements

The training requirements for personnel working in a pharmaceutical manufacturing environment are specified in (21 CFR) 211.25 [1]:

> (a) Each person engaged in the manufacture, processing, packing or holding of a drug product shall have education, training, and experience, or any combination thereof, to enable that person to perform the assigned functions. Training shall be in the particular operations that the employee performs and in current good manufacturing practices (including the current good manufacturing practice regulations in this chapter and written procedures required by these regulations) as they relate to the employee's functions. Training in current good manufacturing practice shall be conducted by qualified individuals on a continuing basis and with sufficient frequency to assure that employees remain familiar with cGMP requirements applicable to them.
>
> (b) Each person responsible for supervising the manufacture, processing, packing, or holding of a drug product shall have the education, training, and experience, or any combination thereof, to perform assigned functions in such a manner as to provide assurance that the drug product has the safety, identity, strength, quality, and purity that it purports or is represented to possess.
>
> (c) There shall be an adequate number of qualified personnel to perform and supervise the manufacture, processing, packing, or holding of each drug product.

In summary, the *Code of Federal Regulations* states that pharmaceutical manufacturing personnel must be trained as follows:

1. To do their specific jobs
2. In current good manufacturing practices (cGMPs)
3. On written procedures (required in 21 CFR 211:80)

The regulation also goes on to specify that training must fulfill the following conditions:

1. Be provided by qualified individuals
2. Be conducted on a frequent, continuing basis

Finally, the code states that manufacturing supervisors must be trained in the same manner as their employees.

In essence, the code specifies that pharmaceutical manufacturing companies must train their employees using qualified trainers; it also states that supervisors must have the same training as "qualified personnel." What the regulations do not say is *how to* conduct this training.

According to John W. Levchuck, then of the FDA, "The FDA has not published a guideline establishing acceptable procedures for personnel training, nor is a guideline being planned. Neither has the FDA specified strict training requirements" [2]. Levchuck's statement may be on its way to being updated. The latest pharmaceutical manufacturing training direction comes from the FDA's draft guidance document "Guidance for Industry: Quality Systems Approach to Pharmaceutical Current Good Manufacturing Practice Regulations" [3]. Published in 2004 this guidance document states in its develop personnel section:

> Under a quality system, senior management is expected to support a problem-solving and communicative organizational culture. Managers are expected to encourage communication by creating an environment that values employee suggestions and acts on suggestions for improvement. Management is also expected to develop cross-cutting groups to share ideas to improve procedures and processes.
>
> In the quality system, it is recommended that personnel be qualified to do the operations that are assigned to them in accordance with the nature of, and potential risk to quality presented by, their operational activities. Under a quality system, managers are expected to define appropriate qualifications for each position to help ensure individuals are assigned appropriate responsibilities. Personnel should also understand the impact of their activities on the product and the customer (this quality systems parameter is also found in the CGMP regulations, which identify specific qualifications, i.e., education, training, and experience or any combination thereof; see 21 CFR 211.25(a) & (b)). Under a quality system, continued training is critical to ensure that the employees remain proficient in their operational functions and in their understanding of CGMP regulations. Typical quality systems training would address the policies, processes, procedures, and written instructions related to operational activities, the product/service, the quality system, and the desired work culture (e.g., team building, communication, change, behavior).
>
> Under a quality system (and the CGMP regulations), training is expected to focus on both the employees' specific job functions and the related CGMP regulatory requirements.
>
> Under a quality system, managers are expected to establish training programs that include the following:
>
> - Evaluation of training needs
> - Provision of training to satisfy these needs
> - Evaluation of effectiveness of training
> - Documentation of training and/or re-training

> When operating in a robust quality system environment, it is important that supervisory managers ensure that skills gained from training be incorporated into day-to-day performance.

The quality systems draft guidance document reinforces the training requirements contained in 21 CFR 211.25 and goes on to add several key items. These include an evaluation of training needs, an evaluation of training effectiveness, and a documentation requirement. Once again, however, the document does not describe how training is to be designed, conducted, or evaluated. In addition, the guidance document does not provide details on the specific training information to be collected, how it should be stored, or how long the documentation records should be retained.

Managers have started to view training as a business-critical component in achieving improved performance and compliance with regulations, but at the same time they may have questions about how to direct training efforts. In the absence of firm guidelines for training, many in the industry have interpreted FDA commentary and audit results as supporting a competency-based approach to training, with validated and reliable training programs that produce measurable performance outcomes. Work responsibilities and tasks are specified in standard operating procedures, guidelines, batch records, employee directives, and protocol. Training must ensure that all employees know of the existence of all these documents, how to access them, and how they are used to direct work. Further, "qualified personnel" must demonstrate that they have read and understand these documents and can perform work as directed by them.

## 1.3 Good Training Practices and Pharmaceutical Manufacturing

Good training practices and a competency-based approach to training may first have been introduced to the general manufacturing industry in 1961 with the publication of a booklet by Robert Mager [4]. In this work, Mager tells educators and trainers how to follow a systematic approach to develop training materials and programs so students and trainees know what competencies they should possess after training. The idea of developing systematic, competency-based training in the pharmaceutical industry has its roots in an article published in 1982 by Ronald Tetzlaff [5], then of the FDA, which builds on the work done by Mager and others in the training field by explaining how a systematic approach to training program design is the best way to build effective, consistent training for employees in pharmaceutical manufacturing.

## 1.4 What Is Competency-Based Training?

Competency-based training is training aimed at ensuring that certain competencies (skills and knowledge) are achieved. Competency-based training measures trainees' mastery of materials through a test or skill demonstration or both to ensure that trainees have acquired competencies. For example, a competency-based program for tablet operators may be designed to ensure that operators can follow all steps in equipment start-up, operation, shut-down, and troubleshooting. The test to ensure that trainees acquired thoses competencies during training might include labeling

a diagram of the equipment and performing a demonstration of how to start up, operate, shut down, and troubleshoot the equipment.

### 1.5 Why Is Competency-Based Training So Important?

Competency-based training is important because it fulfills FDA training guidelines while at the same time meeting business needs. Businesses need a trained, competent workforce capable of manufacturing products effectively and efficiently. How effectively does pharmaceutical manufacturing respond to the challenge? According to one report, "A lack of trained and experienced technical and production staff will have an impact on more than half of the world's biopharmaceutical developers and contract manufacturers in the next five years, and impact their ability to meet demand, according to a survey of 100 international biopharmaceutical manufacturers and contract manufacturing organizations" [6].

From a compliance perspective, interest in training is also starting to rise: "The U.S. FDA is paying more attention to drug manufacturers' training programs under its quality system approach to inspections. While personnel training is not the hottest compliance topic in pharmaceutical GMP literature, it is a frequent U.S. FDA investigator observation, is appearing on a growing percentage of warning letters, and has been involved in some of the highest profile FDA regulator actions in recent years" [7]. So competency-based training makes sense from both business needs and compliance requirements. A well-thought-out, efficiently executed training plan can help ensure pharmaceutical companies not only meet the intent of regulations but also show a return on money invested in training.

## 2 DEVELOPING A TRAINING PLAN: STRATEGY TO ENSURE TRAINING COMPLIANCE IN A PHARMACEUTICAL MANUFACTURING FACILITY*

Ensuring training compliance within a pharmaceutical manufacturing facility requires the facility to establish and implement a comprehensive training plan. Set up and executed properly, this plan will encompass and meet all of the training requirements of:

- 21 CFR 210.25
- "Guidance for Industry: Quality Systems Approach to Pharmaceutical Current Good Manufacturing Practice Regulations"
- Good training practices

Managers in many organizations have an idea of what they want their training to look like, but few have a well-defined, written training plan to help them get there. A training plan is a roadmap that guides and integrates the training process throughout the facility. This plan helps ensure a facility accomplishes its goals consistently, according to its standards of quality, production, cost, and safety. The plan should

*The material in this section was originally published in 1999 in the *PDA Journal* [8]. The information is updated and expanded and examples of the material described are presented.

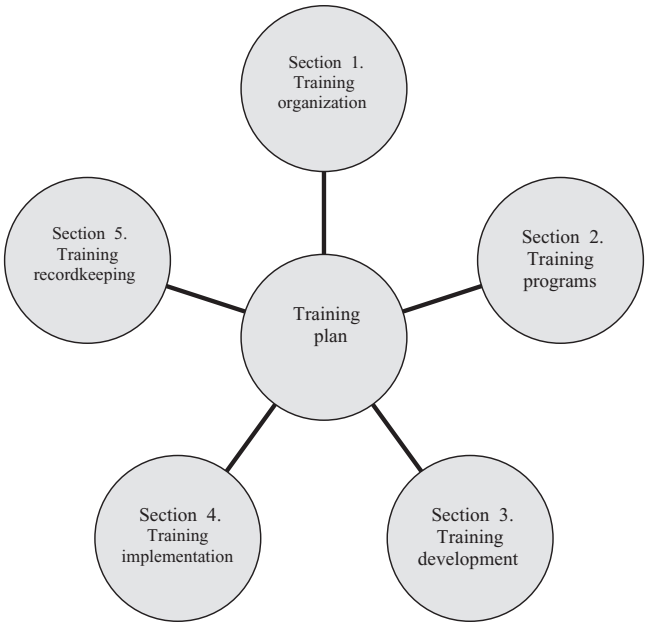


**FIGURE 1** Training plan schematic.

carry the full weight and authority of other corporate directives with the highest levels of management support for training as an essential component in the manufacture of safe and effective health care products.

Why develop a training plan rather than a standard operating procedure (SOP) that covers training? Some managers believe that such a procedure is "required" [9]. Remember, though, that developing a training SOP—or any SOP—mandates training on that SOP. Ideally, the training plan should be more comprehensive than what is normally captured in an SOP. The training plan is a document that guides the development of training programs and materials and the way those programs and materials are implemented. A training plan includes training policy and goes beyond a procedure. And because it is not a procedure, it need not be trained.

A training plan might look like the schematic in Figure 1. The plan has five sections.

*Note*: The facility should have a master training plan that covers the entire site. Each department should develop its own training plan, including the training needed to qualify personnel for all job functions. The departmental plan should provide a schedule of training and retraining to ensure the availability of a sufficient number of qualified personnel to perform all job functions at all times.

A checklist for ensuring all components of a training plan are identified is shown in Figure 2.

## Section 1 Training Organization

This section of the training plan calls for the development of a training philosophy and mission statement. These provide the overall foundation for a strong training

1. Training organization
1.1. Develop a training philosophy
1.2. Develop a training mission statement
1.3. Develop a training organization
1.4. Develop a training support network

2. Training programs
2.1. Identify all training programs/personnel to be trained
2.2. Establish personal qualification pathways
2.3. Training program materials

3. Training development
3.1. Establish a training program design model
3.2. Ensure development of valid training programs
3.3. Establish a change control methodology

4. Training implementation
4.1. Develop an implementation schedule
4.2. Establish a training failure response process

5. Training recordkeeping
5.1. Establish a recordkeeping system
5.2. Training record requirements

**FIGURE 2** Training plan checklist.

organization. In addition, a training organization must be established with persons designated to support the training effort throughout the facility.

***1.1 Develop a Training Philosophy*** Once the decision is made to develop a training plan, those involved need to formulate a training philosophy. Most organizations have developed philosophies around production, quality, cost, safety, and other aspects of their operation. Ask an employee what the company's philosophy is on meeting production quotas, and he or she can probably sum it up in a few words. But how many organizations have developed a well-articulated training philosophy? The answer is very few.

A training philosophy that acknowledges training as a critical component in achieving corporate business objectives is a key element in establishing the credibility of training in an organization. A training philosophy represents an organization's commitment to training. It tells employees, suppliers, customers, and other stakeholders the organization is committed to ensuring its people have the skills and knowledge needed to compete in an increasingly competitive environment. A training philosophy should acknowledge that training is conducted not only to meet regulatory requirements but also because properly trained personnel contribute to the quality of the operation and financial success of the company. Figure 3 provides an example of a training philosophy.

***1.2 Develop a Training Mission Statement*** The training philosophy provides the basis for the training mission statement. The mission statement identifies the overall goals of organizational training and emphasizes the importance of establishing personnel standards of performance. It also positions training as critical to quality

We are committed to a strong training organization as supported by our mission statement. Our training philosophy includes the following elements:

- Commitment to preparing all employees to do their jobs competently and efficiently to produce products that consistently meet or exceed customer requirements for safety and efficacy
- Commitment to continuous training of all employees throughout the organization so they may reach their full human potential

**FIGURE 3** Sample training philosophy.

Our training mission statement is based on our training philosophy. We will…

- Support all corporate objectives in the business plan by providing excellent training programs to all employees
- Evaluate the efficacy of those programs on an ongoing basis
- Grow and develop our employees in all areas of their lives, including interpersonal and technical areas
- Commit to support the organization as a top-tier growth company that can adapt to change in an ongoing regulated environment
- Provide ongoing training to ensure all personnel are in compliance with corporate and regulatory guidelines
- Provide resources for employee growth while meeting the organization's quality, productivity, safety, and financial goals

**FIGURE 4** Sample training mission statement.

outcomes. Representatives from every department should participate in drafting a company training mission statement to ensure it reflects their interests. A sample training mission statement is presented in Figure 4.

Once the training philosophy and mission statement are developed, they need to be communicated to all employees to solicit feedback and support. The communication should focus on the direction the company is taking and the role managers will play in ensuring training is carried out effectively in the organization. A clearly worded philosophy and mission statement set the groundwork for building a plan that integrates training into all areas of personnel development.

***1.3 Develop a Training Organization*** If it has not already been set up, a training organization needs to be established by the corporate or facility management team.

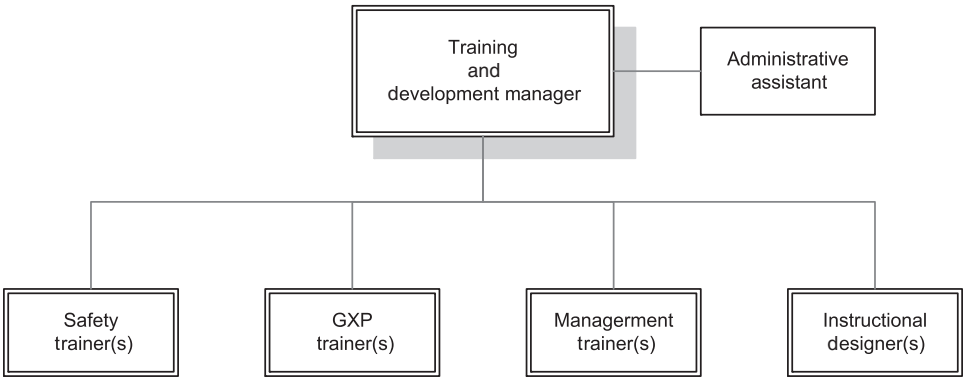


**FIGURE 5** Sample training department structure.

Our experience has shown that training is most effective when the training organization or department is a facility function—on the same level as other major departments within the manufacturing facility. A typical organization chart for the training department within a pharmaceutical manufacturing facility might look like the one in Figure 5.

Roles and responsibilities should be defined in the plan as follows. Depending upon the size of the facility, some of the roles may be combined.

*Training and Development Manager* Since the goal of training is to ensure that personnel are capable of performing their jobs to produce consistent, quality products, the training function must support all other functions and be supported by them. The training and development manager must have lines of communication with all other functions, including quality, safety, production, information systems, and laboratory. Although many organizations agree on this in principle, the practice is likely to be different.

The training and development manager should direct, monitor, and build support for the training effort throughout the organization. The position should meet regularly with site management to review the training plan of action and the progress to date and ensure the training effort supports business and compliance goals. He or she also directs training coordinators and assists in communicating and developing training plans and outcome assessments for the departments to ensure they are implemented effectively.

This function should extend to departmental training to ensure that systems exist in each department to meet specific training requirements. The training and development manager should develop guidelines to be followed by each department for complying with the corporate training policy. These guidelines will identify the training coordinators, instructional designers, departmental trainers, and subject matter experts (SMEs), as well as persons responsible for scheduling and recording participation in training.

The training and development manager should be responsible for at least the following:

- Developing, maintaining, and administering the facility training plan
- Identifying training needs for each department to support site business objectives and compliance efforts
- Developing training plans and schedules to meet organizational training needs
- Reviewing training plans and schedules to ensure support for training from all departments
- Establishing metrics to measure training effectiveness
- Directing and monitoring training coordinators in executing training plans

*Safety Trainer(s)* Safety trainers are those individuals that provide safety training to the manufacturing employees. The safety trainers are responsible for at least the following:

- Working with instructional designers to develop safety training programs
- Presenting required safety training to appropriate departments, such as:
  - Lockout/tagout
  - Confined space entry
  - Hearing protection
  - Respiratory protection
  - Bloodborne pathogens
- Identifying ongoing safety training needs

*Management Trainer(s)* Management trainers are individuals that provide "soft skill" training programs to the other employees at the pharmaceutical manufacturing site. The management trainers are responsible for at least the following:

- Working with instructional designers to develop management development programs
- Presenting management development programs, such as:
  - Leadership
  - Problem solving/decision making
  - Coaching and counseling
  - Time management

*GXP Trainer(s)* GXP trainers are primarily responsible for presenting training programs on current regulations within the pharmaceutical manufacturing industry. The GXP trainers are responsible for at least the following:

- Working with instructional designers to develop GXP training programs
- Presenting initial GXP training during new employee orientation
- Presenting ongoing GXP topics in:
  - Good manufacturing practices
  - Good documentation practices
  - Good laboratory practices

*Instructional Designers* The instructional designers develop training programs as directed by the training coordinators. Instructional designers serve as a liaison between the training coordinators, training and development manager, departmental supervisors, and SMEs. Instructional designers are responsible for at least the following:

- Identifying appropriate SMEs for training projects, in coordination with department training coordinators
- Developing competency-based training programs according to established principles of instructional systems design
- Developing evaluations and assessment measures
- Developing train-the-trainer programs for department trainers

***1.4 Develop a Training Support Network*** To function properly, a training department needs to establish a network for training support within each functional department of the manufacturing facility. Within each functional department, three roles should be identified to help support the training effort. Sometimes these roles can be performed by one person; sometimes the roles are performed by individuals within the department. Typically, the roles are not full-time positions. Figure 6 provides a list of departments and the training support roles necessary for each department; you may want to add to the list. Then, each support role is described.

*Training Coordinators* Training coordinators are the liaison between functional departments and the training department. They are responsible for communicating with departmental training needs to the training and development manager. Training coordinators also serve as the liaison between departmental SMEs, departmental trainers, and instructional designers. They work to identify training needs and develop training programs that support the plans set out by the training and development manager with site management approval. They are also responsible for building support within the departments for executing training plans. Training

| Department | Training Coordinator(s) | Subject Matter Experts | Departmental Trainers |
|---|---|---|---|
| • Administration and finance | | | |
| • Buildings and grounds | | | |
| • Facilities and engineering | | | |
| • Human resources | | | |
| • Information systems | | | |
| • Manufacturing | | | |
| • Regulatory compliance | | | |
| • Research and development | | | |
| • Safety and security | | | |
| • Shipping and receiving | | | |

**FIGURE 6** Departments/training support roles.

coordinators may serve more than one department within a facility. Training coordinators are responsible for at least the following:

- Consulting with department management to identify training needs
- Communicating departmental training needs and progress of training to the training and development manager
- Developing departmental training plans and specifying training programs as directed by the training and development manager and department management
- Directing instructional designers in developing training programs, evaluations, and assessments

*Subject Matter Experts* Subject matter experts are individuals who possess a broad base of knowledge about the area where they work. These individuals are typically employees who have been with the company and department a significant amount of time—typically five years or more. Subject matter experts are responsible for at least the following:

- Working with training coordinators to identify departmental training needs
- Working with instructional designers to design and develop training programs
- Working with departmental trainers to implement training programs

*Departmental Trainers* The departmental trainers should be selected for their skill and interest in training and ability to perform their jobs. Some SMEs may become departmental trainers—but only with the stipulation that they want to learn how to become a trainer. Departmental trainers are responsible for at least the following:

- Working with instructional designers to design and develop training programs
- Training new hires and incumbents on job-specific skills and regulatory issues

In order to meet the 21 CFR 211.25 requirements, departmental trainers must complete a train-the-trainer program. An outline of topics to be covered in a typical train-the-trainer program is shown in Figure 7.

## Section 2 Training Programs

This section includes a listing of all personnel to be trained as well as a listing of suggested training programs that need to be provided at the manufacturing facility. In addition, employee qualification pathways are discussed and illustrated.

***2.1 Identify All Training Programs/Personnel to be Trained*** Figure 8 provides a sample list of personnel who need to be trained and the training programs they need to work effectively and efficiently at a pharmaceutical manufacturing facility; you will want to add to the list.

1. Introductions
2. Program objectives
3. Trainer responsibilities
4. Training process: trainee characteristics
5. Training process: adult learning theory
6. Training process: what makes successful training?
7. What makes a successful trainer?
8. Training techniques
9. Plan
10. Prepare
11. Present
12. Feedback and coaching
13. Using an OJT checklist
14. Exercise 1— OJT with SOP
15. Exercise 2— OJT checklist

Review and wrap up

**FIGURE 7** Sample train-the-trainer program outline.

***2.2 Establish Personal Qualification Pathways*** Good business practice, as well as CFR 21 211.25, mandates that each employee possess the education, training, and experience to enable him or her to perform assigned functions in a safe and effective manner. The training plan must contain a policy allowing personnel to demonstrate their qualifications. This should cover training and supervisory personnel as well as those directly involved with operations.

A qualification pathway should include all of the training programs necessary to qualify employees in job functions. A sample personal qualification pathway for a compression technician is shown in Figure 9.

***2.3 Training Program Materials*** While there are basically three methods for delivering training available—classroom led, self-instruction, or a combination of the two—training materials should include:

1. A method of imparting knowledge or desired attitudes. This type of training typically occurs in a classroom-led or self-instructional system (paper or computer based).
2. A method of assessing if that knowledge or attitude transfer took place, that is some form of assessment.

| Management Training Programs | |
|---|---|
| **Training Programs** | **Personnel to be Trained** |
| | |
| Leadership | Supervisors and above |
| | |
| Coaching and counseling | Supervisors and above |
| | |
| Problem solving/decision making | Supervisors and above |
| | |
| Coaching and counseling | Supervisors and above |

| Specialty Training Programs | |
|---|---|
| **Training Programs** | **Personnel to be Trained** |
| | |
| Bloodborne pathogens | Appropriate personnel |
| | |
| Conducting effective investigations | Appropriate personnel |
| | |
| Drug-labeling regulations | Appropriate personnel |
| | |
| Electronic signatures and batch records | Appropriate personnel |
| | |
| FDA inspections | Appropriate personnel |
| | |
| Hazcom | Appropriate personnel |
| | |
| SOP writing | Appropriate personnel |
| | |
| Surveillance monitoring | Appropriate personnel |
| | |
| Working with contract research organizations | Appropriate personnel |
| | |
| Validation concepts | Appropriate personnel |

**FIGURE 8** Training programs/personnel to be trained.

If the training requires employees to be competent in a particular task or skill, then these materials must also be developed:

3. A structured checklist that allows different trainers to demonstrate the same way of doing a particular task to each trainee.
4. An evaluation checklist that can be used by a trainer (it is suggested that the trainer who provided the initial training not do the competency evaluation) to determine if a trainee is competent in performing the assigned skill.

*Frequency of Training* In general, providing refresher or retraining on a routine basis is not necessary. Some training programs are mandated to be repeated, or refreshed, usually by federal, state, or local government, that is, some safety training.

| Department: Manufacturing | | Job title: Compression operator |
|---|---|---|

| Job Task | Job Duties | Training Start Date | Training Completion Date |
|---|---|---|---|
| Completing documentation | Master batch record (MBR) | | |
| | | | |
| Performing cleaning procedures | Minor equipment cleaning | | |
| | Major cleaning of detachable parts | | |
| | Major cleaning of work centers | | |
| | | | |
| Operating equipment—basic | Fette | | |
| | Killian | | |
| | Courtoy | | |
| | | | |
| Operating equipment—advanced | Compression theory | | |
| | | | |
| operating support equipment | Use and cleaning of friability tester | | |
| | Use and cleaning of hardness tester | | |
| | Use and cleaning of micrometer | | |
| | Use and cleaning of metal detectors | | |
| | Use of weigh checkers | | |
| | | | |
| Transporting materials | Transferring sacks to pallets | | |
| | Use and handling of totes | | |
| | Staging totes/sacks for coating | | |
| | | | |

**FIGURE 9** Sample personal qualification pathway.

Other training programs may be mandated to be repeated by a company—such as sexual harassment. Retraining may be necessary when a person has been absent from a job for a period of time due to illness, pregnancy, other job assignment, and so on. In that case, the company may require retraining after a certain period of missed job function. In addition, retraining or training needs may be indicated when situations such as the following occur:

- Out-of-specification product (quality issues)
- Excessive waste
- Decrease in production

Periodic audits of work performance may identify training needs as well. However, our experience has been that retraining is typically identified through situations similar to those mentioned above.

## Section 3 Training Development

Employees need job-specific training in environment, SOPs, safety, GMP regulations and awareness, and technical skills. The training plan should ensure a systematic approach to develop and implement competency-based training for all programs and materials.

***3.1 Establish a Training Program Design Model*** Effective competency-based training is the result of applying a systematic process to training program design. This process involves following certain well-defined steps to develop a training program that meets both the trainee's and the organization's needs. The most reliable method of training and qualifying people in the safe and effective performance of work is using an instructional system design model. One instructional systems

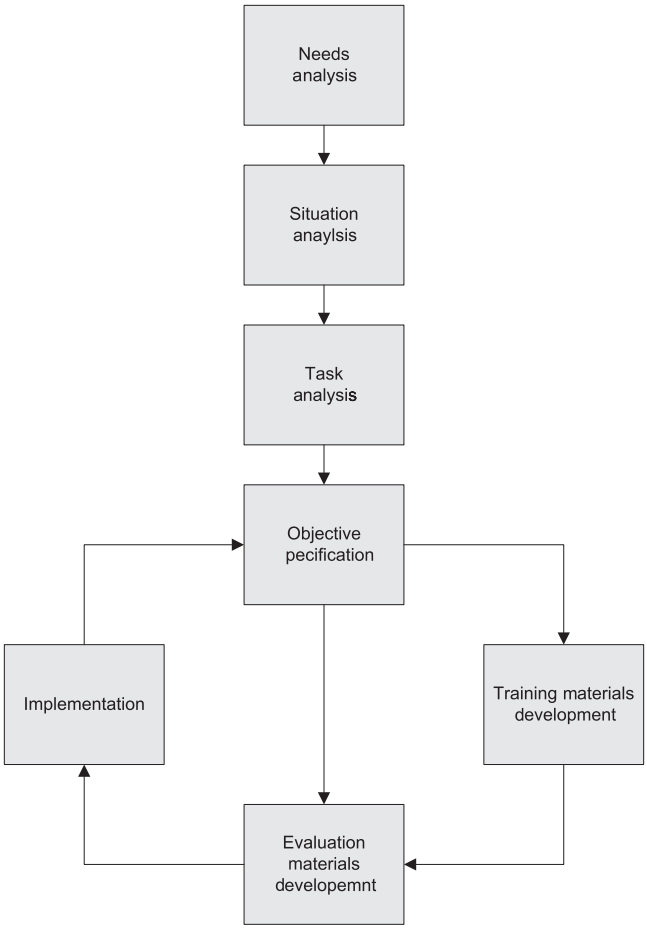


**FIGURE 10** Training program design model.

design model, the training program design model (TPDM) developed by Gallup and Griffin, is shown in Figure 10.

***3.2 Ensure Development of Valid Training Programs*** Validity may be defined as something that accomplishes that which it purports to accomplish. A training program is considered valid when it accurately imparts the competencies required to do a job. How can a training program be validated? The way to ensure program validation is to follow a training program design module and the following process:

1. Conduct a task analysis that identifies discrete bits of knowledge and specific job competencies; have SMEs sign off on completed task analyses indicating that all discrete bits of knowledge and specific tasks required to complete a job have been identified.
2. Write measurable performance objectives that state exactly what a trainee is to know or do after training has been completed; have SMEs sign off on com-

pleted performance objectives indicating that they agree that the level of stated performance is acceptable.

3. Develop assessment instruments that match up with the performance objectives; have SMEs sign off on assessments indicating they agree that the instruments adequately measure trainee competency.

Assess trainee's knowledge and skills using the approved assessment instruments; ensure that assessments are conducted in the same manner for all employee groups.

***3.3 Develop a Change Control Policy and Procedure*** A change control policy helps ensure personnel are prepared to carry out new and revised policies and procedures effectively. An effective change control policy should include directions for regular reviews and revision of training materials and programs as well as scheduling training for new and incumbent employees on all pertinent SOPs. The plan must also include a procedure for responding to developments in operations, processes, and documentation.

## Section 4 Training Implementation

***4.1 Develop an Implementation Schedule*** The plan must include a schedule to ensure that training begins at initial employment and is ongoing. The schedule should include time frames for completing job qualification training and address training required for changes in process and new or increased performance expectations. Developing a training implementation schedule can be accomplished by adding to the matrix developed in Section 2, "Identify All Training programs/Personnel to be Trained." This revised matrix is shown in Figure 11.

***4.2 Establish a Training Failure Response Process*** The training plan should describe a process for dealing with personnel who do not pass qualifying evaluations. It should list steps for isolating the cause of the failure and differentiate between discrepancies in training program design and inappropriate candidates for the job. It should include publishing the protocol for dealing with test failure. A typical failure response for a task-based process is shown in Figure 8. After a person completes a training program covering the knowledge to perform a specific task, an assessment is given. If that assessment is passed at a predetermined score, the person goes on to complete the skill-based portion of the training. Once the skill portion is completed, the person is observed as they perform the skill. If they perform it correctly, they are certified or qualified to perform the specific process operation. If they do not pass the skill demonstration check, they must repeat the training again. The number of times they are allowed to repeat the training and assessment should be part of the training failure response process. Figure 12 illustrates a basic failure response process.

## Section 5 Training Recordkeeping

The FDA does not specify training documentation/recordkeeping requirements in the *Code of Federal Regulations*, but it has made them an industry requirement by virtue of precedent-setting industry demands.

| Specialty Training Programs | | |
|---|---|---|
| **Training Programs** | **Personnel to be Trained** | **Training Scheduled** |
| | | |
| Bloodborne pathogens | Appropriate personnel | As needed |
| | | |
| Conducting effective investigations | Appropriate personnel | As needed |
| | | |
| Drug-labeling regulations | Appropriate personnel | As needed |
| | | |
| Electronic signatures and batch records | Appropriate personnel | As needed |
| | | |
| FDA inspections | Appropriate personnel | As needed |
| | | |
| Hazcom | Appropriate personnel | As needed |
| | | |
| SOP writing | Appropriate personnel | As needed |
| | | |
| Surveillance monitoring | Appropriate personnel | As needed |
| | | |
| Working with contract research organizations | Appropriate personnel | As needed |
| | | |
| Validation concepts | Appropriate personnel | As needed |

**FIGURE 11** Sample implementation schedule.

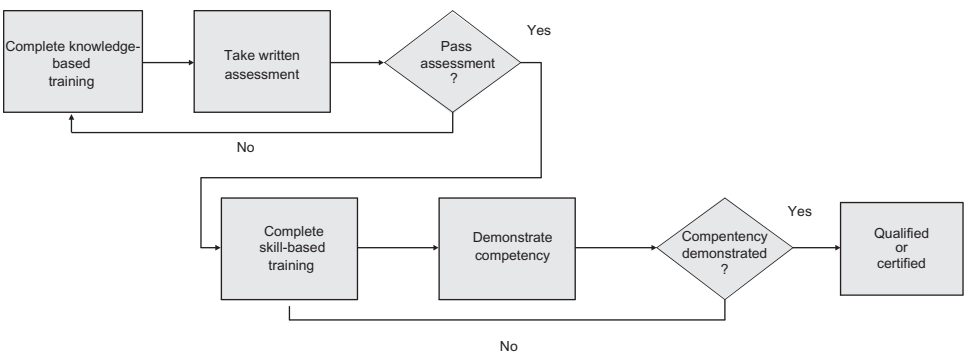


**FIGURE 12** Basic failure response process.

Discussions with training managers at pharmaceutical manufacturing companies audited by the FDA suggest that it is through documentation/recordkeeping that the FDA "backs into" an audit of a company's training. This typically happens in one of three ways:

1. Auditors ask to see a SOP and observe an employee perform the procedure. Depending on the employee's performance, the auditors may ask to review the documents or records of the observed employee's training.

2. Through investigation of production records, auditors spot deviations or out-of-specification issues. This in turn may lead to an examination of an operator's work and qualifications for the job. The auditors may ask how or what additional training was conducted and request the records that prove training was provided.
3. Auditors identify processes or work practices that appear to be performed incorrectly during a routine tour of a manufacturing facility. The auditors may ask to see the training materials or training records associated with individuals working in the area.

Further, training managers note that if the training department can produce training documentation quickly and in an organized fashion, auditors are more likely to view the overall training effort as effective. At the same time, when solid training documentation cannot be produced quickly, a more in-depth review of the training system may occur [10]. As a result of these requirements to produce training records, it is critical that an efficient training recordkeeping system be in place at the pharmaceutical manufacturing facility.

***5.1 Establishing a Recordkeeping System*** Any documentation system may meet the requirements of "produce … quickly and present in an organized fashion."

Three types of methods are commonly used for training documentation/recordkeeping in the pharmaceutical manufacturing industry: paper systems, electronic systems with paper backup, and stand-alone electronic systems. Some pharmaceutical companies use a paper-based documentation system. Electronic systems with paper backup, also known as learning management systems (LMSs), range from elaborate programs such as IsoTrain, SAP, Plateau, and Registrar to departmental databases created in Microsoft Access. Paper backup typically includes at least sign-in sheets and sometimes evaluation instruments. An LMS is essentially a database made up of multiple tables to store discrete units of information. These might include employee information, course information, evaluation data, and training programs or materials. These tables can be merged and queried to produce specific reports. Report capabilities allow users to compare, analyze, question, and project needs.

***5.2 Training Record Requirements*** The following basic information should be included in any training recordkeeping system:

- Employee name
- Employee identification
- Personal qualification pathway, including training programs that need to be completed and targeted completion dates
- Results of training programs completed

In addition, the training program materials, including trainee guides and assessment instruments, should be readily accessible—either electronically or in a paper-based filing system.

## REFERENCES


1. U.S. Food and Drug Administration (FDA) (2003), Current good manufacturing practice for finished pharmaceuticals, *Code of Federal Regulations*, Title 21, Part 211.25, FDA, Rockuille, MD.
2. Levchuck, J. W. (1991), Training for GMPs, *J. Parenter. Sci. Technol.*, 45(6), 270–275.
3. U.S. Food and Drug Administration (FDA) (2004), Guidance for industry: Quality systems approach to pharmaceutical good manufacturing practice regulations, FDA, Rockuille, MD.
4. Mager, R. (1962), *Preparing Instructional Objectives*, Center for Effective Performance, Atlanta, GA.
5. Tetzlaff, R. (1982), A systematic approach to GMP training, *Pharm. Technol.*, 6(11), 42–51.
6. Langer, E. S. (2004), Training day for International Biopharma Manufacturers, *Contract-Pharma*, 6(2), 46.
7. Morris, W. (2006), Personnel training: A growing compliance concern, *PDA Newslett.*, XLII(4), 1–38.
8. Gallup, D. A., Beauchemin, K. A., and Gillis, M. (1999), A comprehensive approach to compliance training in a pharmaceutical manufacturing facility, *PDA J. S. Technol.*, 53(4), 163–167.
9. Vesper, J. L. (2000), Defining your GMP training program with a training procedure, *BioPharm.*, 13(7), 28–32.
10. Gallup, D. A., Beauchemin, K. A., Gillis, M., Altopedi, D. and Manor, J. (2003), Selecting a training documentation/recordkeeping system in a pharmaceutical manufacturing environment, *PDA J. Sci. Technol.*, 57(1), 49–55.

# ORIGIN OF CONTAMINATION

Denise Bohrer
*Universidade Federal de Santa Maria, Santa Maria, Brazil*

**Contents**

1 Introduction
2 Endogenous Contaminants
  2.1 Raw Materials
  2.2 Additives
  2.3 Decomposition of Formulation Constituents
3 Exogenous Impurities
  3.1 Residual Solvents
  3.2 Containers
  3.3 Delivery Systems
  3.4 Particulate Matter
4 Concluding Remarks
  References

## 1 INTRODUCTION

There are many ways to approach the issue of contamination. Although the origin of contaminants in pharmaceutical products can easily be identified, definitions on this theme can be looked at from different angles.

In a first approach, the definition of contamination in pharmaceutical products can be referred to in terms of related substances and process contaminants: While *related substances* are structurally related to the drug substance, *process contaminants* are introduced during manufacturing or handling procedures. These two categories include all types of contaminants but do not define them.

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

The U.S. Pharmacopeia and National Formulary (USP-NF 27) [1] presents definitions of terms related to contaminants in *impurities in official articles* <1086> and *ordinary impurities* <466>. In these monographs, impurities are classified into inorganic, organic, biochemical, isomeric, and polymeric, and definitions are given in terms of foreign substances, toxic impurities, concomitant components, signal impurities, and ordinary impurities. According to these definitions, *foreign substances* are not the consequence of the synthesis or preparation of the drug, yet they are introduced by contamination or adulteration. *Concomitant components* are characteristic of bulk chemicals and are not considered impurities in the pharmacopeial sense. Ordinary impurities are species in bulk chemicals that are innocuous and do not present significant undesirable biological activity in the given amounts.

Different from foreign impurities, which theoretically are not present and, therefore, are not included when monograph tests and assays are selected, *toxic impurities* and *signal impurities* may arise from the synthesis, preparation, or degradation of compendial articles and, differently from *ordinary impurities*, may present undesirable biological activity, even as minor components. In this context, the following topics should be taken into account:

- Source of a drug substance: natural, synthetic, biotechnological
- Ratio impurity/drug substance and its toxicological effects
- Pharmacology of the impurity

Other factors that should also be taken into account when limits for impurity levels in bulk substances are to be set are:

- Route of administration
- Dose
- Target population
- Duration of the therapy

Not included in *process contaminants*, but equally an external source, are those impurities arising during packaging, during drug reconstitution or admixing, and also during the administration to the patient. Very important under this category is the particulate matter in intravenous formulations, mainly infusion solutions due to the large volume administered.

Figure 1 presents an overview of impurity sources, both endogenous and exogenous, considering the steps at which these can occur. It also summarizes how "the origin of contamination" will be addressed in this chapter. It is divided into endogenous and exogenous contaminants. The first part, endogenous contaminants, is devoted to water since it is the main raw material, if not in the drug form at least in any step of its preparation. Following water, the contamination arising with raw materials will be treated, considering the presence of concomitants (natural impurities of raw materials) and by-products, synthesized along with the drug. Also important while dealing with endogenous contaminants are those formed through inadvertent drug decomposition. Exogenous source of contaminants—aid materials used during drug or formulation preparation, containers, and delivery systems—will also be considered.

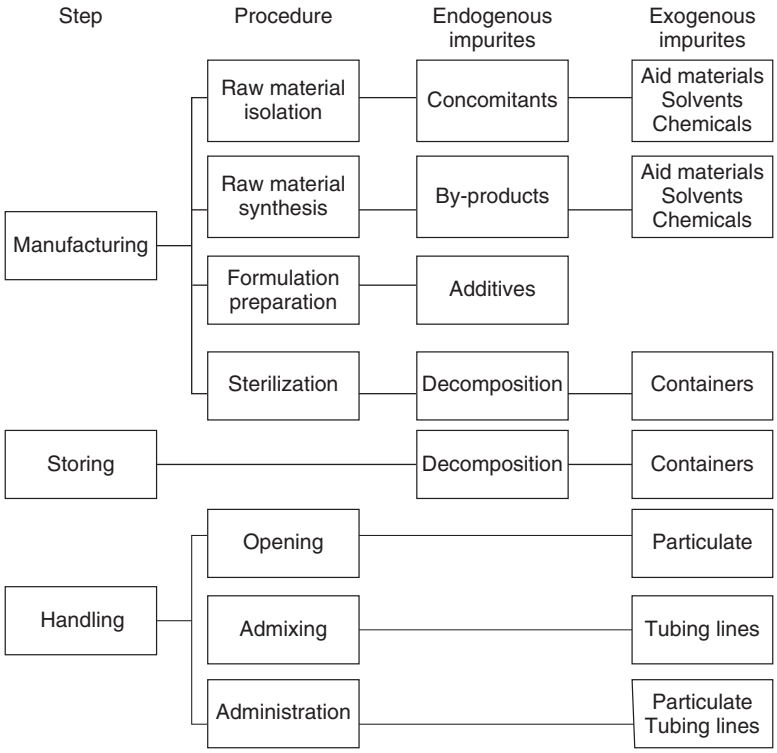


**FIGURE 1** Schematic representation of an overview on impurity sources.

## 2 ENDOGENOUS CONTAMINANTS

Endogenous contaminants can already be present as concomitants in raw substances, whether formed during drug synthesis as a by-product or arising from reactions among formulation constituents. The generation of contaminants in formulations after processing and packaging is generally promoted by external agents such as light, ultraviolet (UV) radiation, heat, or air, causing unexpected reactions between constituents.

Pharmacopeial compendia establish a couple of tests that should be carried out with raw material batches to determine their level of impurities. Concomitants usually detected are heavy metals, chloride, or sulfated ash. The presence of by-products varies from raw material to raw material. Examples can be both innocuous substances such as monohydrogen phosphate, which is an impurity in dihydrogen phosphate salts and glycine, which is an impurity in alanine, and toxic substances, such as the classical enantiomer S of thalidomide.

### 2.1 Raw Materials

Raw materials employed in the pharmaceutical industry may have two different origins. They are either naturally occurring substances or synthesized drugs. Among the natural products are active ingredients from plant extracts or animals, chemicals,

and biotechnological products. As vast as the variety of different raw materials is the array of impurities. This section will deal with the impurities already present in raw materials as concomitants or those arising during drug synthesis. Concomitants include the usual impurities considered by pharmacopeial compendia, such as heavy metals and arsenic, for which limits are prescribed in all monographs. By-products arising during drug synthesis may or may not be a problem when present as contaminants.

***Water*** Water is the primary raw material in pharmaceutical formulations. It is the most used vehicle since it is the major component of the human body. For many products, it is the main component, and, even in those containing non-water-soluble substances, water must be present. Depending on the product and the form of administration, lipophilic drugs are prepared as water–oil emulsions.

The amount and level of contaminants or impurities in water for pharmaceutical purpose depend on its use. Since water is used in all industries and scientific work, international and national standard authorities have established water quality parameters for all types of applications. Health-related water standards are given by organizations such as the World Health Organization (WHO) [2], the Environmental Protection Agency (EPA) [3], and the American Society for Testing and Materials Standards (ASTM) [4] in the United States and by pharmacopeial compendia when the aim is specifically related to water for pharmaceutical products for human and veterinary consumption.

Standards for water quality are similar among pharmacopeias (USP [1], BP [5], DAB [6], Ph. Eur. [7], IP [8]), with only a few minor differences in the allowed levels for chemical contaminants. Pharmacopeias classify water in three categories: *purified water*, *highly purified water*, and *water for injection*. Whatever the process for water purification, the raw material is always drinking water. Drinking water parameters, given by governmental regulatory agencies, include a large number of chemicals. They comprise not only naturally occurring substances but also a series of chemicals that may be present from anthropogenic activities in natural waters such as benzene, ethylene diamine tetraacetic acid (EDTA), and cadmium. Table 1 shows the guidelines for drinking water quality adopted by WHO [2].

The water purification process adopted by pharmaceutical industries must be able to furnish water with the quality parameters presented in Table 2.

The ability to achieve a guideline value depends on a number of factors, including:

- Concentration of the chemical in the raw water
- Nature of the raw water
- Treatment processes

Although there are no levels for them, pesticides, for example, should not be present in water for pharmaceutical purposes.

There are treatments available for the pharmaceutical industry for purifying water, which should be chosen according to the degree of purity necessary. They are listed in Table 3 according to their degree of complexity: the higher the ranking, the more complex the process.

**TABLE 1 WHO Guideline Values for Chemicals in Drinking Water and EPA National Drinking Water Standards [2, 3]**

| Contaminant | Guideline Value (mg/L) |
|---|---|
| *Primary Standards* | |
| Acrylamide | 0.0005 |
| Alachlor | 0.002 |
| Antimony | 0.006 |
| Arsenic | 0.010 |
| Asbestos (fibers > 10 μm) | 7 million fibers per liter |
| Atrazine | 0.003 |
| Barium | 2 |
| Benzene | 0.005 |
| Benzo(*a*)pyrene (PAHs) | 0.0002 |
| Beryllium | 0.004 |
| Boron | 0.5 |
| Bromate | 0.010 |
| Cadmium | 0.005 |
| Carbofuran | 0.04 |
| Carbon tetrachloride | 0.005 |
| Chloramines (as $Cl_2$) | 4.0 |
| Chlordane | 0.002 |
| Chlorine (as $Cl_2$) | 4.0 |
| Chlorine dioxide (as $ClO_2$) | 0.8 |
| Chlorite | 1.0 |
| Chlorobenzene | 0.1 |
| Chromium (total) | 0.1 |
| Copper | 1.3 |
| Cyanide (as free cyanide) | 0.2 |
| 2,4-DB | 0.07 |
| Dalapon | 0.2 |
| 1,2-Dibromo-3chloropropane (DBCP) | 0.0002 |
| *o*-Dichlorobenzene | 0.6 |
| *p*-Dichlorobenzene | 0.075 |
| 1,2-Dichloroethane | 0.005 |
| 1,1-Dichloroethylene | 0.007 |
| Cis-1,2-Dichloroethylene | 0.07 |
| Trans-1,2-Dichloroethylene | 0.1 |
| Dichloromethane | 0.005 |
| 1,2-Dichloropropane | 0.005 |
| Di(2-ethylhexyl) adipate | 0.4 |
| Di(2-ethylhexyl) phthalate | 0.006 |
| Dinoseb | 0.007 |
| Dioxin (2,3,7,8-TCDD) | 0.00000003 |
| Diquat | 0.02 |
| EDTA | 0.6 |
| Endothall | 0.1 |
| Endrin | 0.002 |
| Epichlorohydrin | 0.0004 |
| Ethylbenzene | 0.7 |
| Ethylene dibromide | 0.00005 |

**TABLE 1** ***Continued***

| Contaminant | Guideline Value (mg/L) |
|---|---|
| Fluoride | 4.0 |
| Formaldehyde | 0.9 |
| Glyphosate | 0.7 |
| Haloacetic acids (HAA5) | 0.060 |
| Heptachlor | 0.004 |
| Heptachlor epoxide | 0.0002 |
| Hexachlorobenzene | 0.001 |
| Hexachlorocyclopentadiene | 0.05 |
| Lead | 0.015 |
| Lindane | 0.0002 |
| Mercury (inorganic) | 0.002 |
| Methoxychlor | 0.04 |
| Nickel | 0.02 |
| Nitrate (measured as nitrogen) | 10 |
| Nitrite (measured as nitrogen) | 1 |
| Oxamyl (Vydate) | 0.2 |
| Pentachlorophenol | 0.001 |
| Picloran | 0.5 |
| Polychlorinated biphenyls (PCBs) | 0.0005 |
| Selenium | 0.05 |
| Simazine | 0.004 |
| Styrene | 0.1 |
| Tetrachloroethylene | 0.005 |
| Thallium | 0.002 |
| Toluene | 1 |
| Total trihalomethanes (TTHMs) | 1 |
| Toxaphene | 0.003 |
| 2,4,5-TP (Silvex) | 0.05 |
| 1,2,4-Trichlorobenzene | 0.07 |
| 1,1,1-Trichloroethane | 0.2 |
| 1,1,2-Trichloroethane | 0.005 |
| Trichloroethylene | 0.005 |
| Uranium | 0.030 |
| Vinyl chloride | 0.002 |
| Xylenes (total) | 10 |
| *Secondary Standards* | |
| Aluminum | 0.05 to 0.2 |
| Chloride | 250 |
| Copper | 1.0 |
| Fluoride | 2.0 |
| Foaming agents | 0.5 |
| Iron | 0.3 |
| Manganese | 0.05 |
| Silver | 0.10 |
| Sulfate | 250 |
| Total dissolved solids | 500 |
| Zinc | 5 |

**TABLE 2 Water Quality for Pharmaceutical Purposes (Pharmacopeial Standards)**

| | Water Grade | | |
|---|---|---|---|
| Parameter | Purified Water | Highly Purified Water | Water for Injection |
| Conductivity (at 20°C) $\mu S\,cm^{-1}$ | 4.3 | 1.1 | 1.1 |
| Total organic carbon (TOC) (mg/L) | 0.5 | 0.5 | 0.5 |
| Nitrates (ppm) | 0.2 | 0.2 | 0.2 |
| Aluminum (μg/L) | 10 | 10 | 10 |
| Heavy metals (ppm)[a] | 0.1 | 0.1 | 0.1 |
| Chloride | Pass/fail | — | 0.5 ppm |
| Sulfate | Pass/fail | — | Pass/fail |
| Ammonium (ppm) | 0.2 | — | 0.2 |
| Calcium/magnesium | Pass/fail | — | Pass/fail |
| Residue evaporation mg/100 mL | 1 | — | 0.4 (volume ≤ 10 mL)<br>0.3 (volume > 10 mL) |

[a]Measured as lead.
— = not informed.

**TABLE 3 Ranking of Complexity of Water Treatment Processes for Chemicals**

| Ranking | Process |
|---|---|
| 1 | Distillation |
| 2 | Ion exchange |
| 3 | Reverse osmosis |
| 4 | Membrane filtration |

Table 4 presents the quality parameters of water obtained by the purification processes listed in Table 3.

The effectiveness of each treatment in removing the contaminants listed in Tables 1 and 2 are given in Table 5 [2].

Even when water complies with quality parameters as a raw material, it can present some impurities after being turned into a pharmaceutical product. Table 6 presents the level of some contaminants found in water for injection (WFI). Since the raw material should have passed in the quality test, contaminants either were below the allowed concentration level or were introduced after packaging. Contaminants introduced after packaging most likely originate from the packaging materials. Section 3.2 discusses containers as sources of contamination.

In summary, water can be a source of contaminants. If the raw material (drinking water) complies with the quality parameters established by authorities, contaminants still present can be eliminated by usual water purification processes available to the pharmaceutical industry. While distillation and reverse osmosis provide water with the quality specifications for *purified water* and *highly purified water*, *WFI* is generally obtained by membrane filtration (associated with another purification process) not only because of chemical contamination but mainly because of sterility requirements.

**TABLE 4 Quality Parameters of Water after Different Purification Treatments**

| Purification Process | Conductivity (at 20°C) μS $cm^{-1}$ | Total Organic Carbon (mg/L) | Nitrates (ppm) | Aluminum (μg/L) | Heavy Metals (ppm)[a] | Ammonium (ppm) | Calcium/ Magnesium (mg/L) | Residue Evap. mg/100 mL (as silicate) |
|---|---|---|---|---|---|---|---|---|
| None (tap water example) | 240 | 10 | <10 | <200 | 1 | 1 | 35 | 1 |
| Distillation (single) | 10.2 | 0.03 | — | — | 0.5–1 | 0.01 | 1–3 | 0.5–1 |
| Distillation (double) | 2.1 | 0.06 | — | — | 0.1–0.8 | 0.01 | 0.3–0.1 | 0.1–0.7 |
| Ion exchange | 2–30 | — | — | — | <0.01 | — | — | 1 |
| Reverse osmosis | 10–25 | 0.03 | — | — | <0.04 | 0.4 | 1.6 | 0.1 |
| Membran filtration | 0.056 | 0.01 | — | — | <0.01 | <0.01 | — | <0.01 |

[a]Measured as lead — not informed.

**TABLE 5 Effectiveness of Each Treatment in Removing the Contaminants**

| | Treatment | | | |
|---|---|---|---|---|
| Contaminants | Distillation (%) | Ion Exchange | Reverse Osmosis (%) | Membran Filtration (%) |
| Ions | >70 | >80% | Monovalent > 95<br>Polyvalent > 97 | >80 |
| Organic compounds | >80 | Not efficient | >99 | >80 |
| Particles | >80 | Not efficient | >99 | >99 |

*Source*: From refs. 2 and 9.

**TABLE 6 Contaminants Found in Water for Injection**

| Parameter | Sample | Content (μg/L) | Reference |
|---|---|---|---|
| Aluminum | Sterile water, Abbott | <5 | 10 |
| | Sterile water, McGaw | <5 | 10 |
| | Sterile water, Travenol | <5 | 10 |
| | Aqua ad injectabilia, Braun, 50 mL | 1 | 11 |
| Arsenic | Water for injection, EMS | 39.3 | 12 |
| | Water for injection, Geyer | 30.9 | 12 |
| Zinc | Sterile water | 13.9 | 13 |
| Silicate | Sterile water, Geyer | 280 ± 13 | 14 |

**TABLE 7 Main Inorganic Impurities Listed in Pharmacopeial Compendia and Their Limits**

| Impurity | Limit | Impurity | Limit |
|---|---|---|---|
| Sulfated ash | 0.01–1% | Heavy metals | 1–50 ppm |
| Chloride | 10–500 ppm | Arsenic | 1–4 ppm |
| Sulfate | 50–400 ppm | Aluminum | 0.2–1 ppm |
| | 0.1–0.6% | Cadmium | 5–10 ppm |
| Fluoride | 3 ppm | Chromium | 0.05–10 ppm |
| Bromide | 50 ppm | Copper | 0.1 ppm |
| Oxalate | 100–350 ppm | Iron | 2–100 ppm |
| Phosphate | 25–400 ppm | Lead | 0.1–50 ppm |
| Sulfite | 15 ppm | Nickel | 0.2–1 ppm |
| Ammonium | 200 ppm | Silver | 250 ppm |
| | | Zinc | 10–30 ppm |

*Note*: Observed: Some impurities are related to certain products only. For example, the limit for silver is in cysplatin and sulfite in sugars.

***Concomitants*** Concomitants can be considered impurities present in naturally occurring, nonsynthesized raw materials. They may either present toxic effects, as with arsenic, or be as harmless as chloride ions. An overview of usual concomitants and their limits cited in pharmacopeial compendia are listed in Table 7.

The presence of concomitants in pharmaceutical products, although inevitable, may not exactly be a problem. The presence of magnesium in calcium salts is very

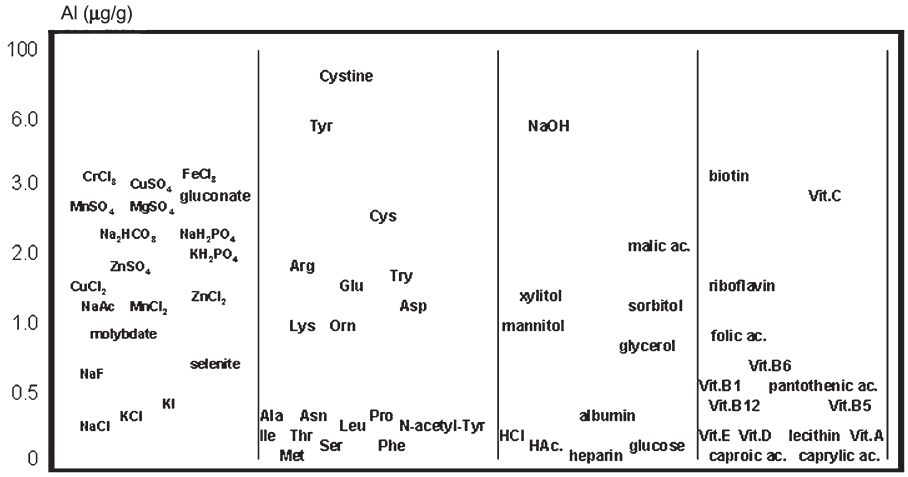


**FIGURE 2** Aluminium present as impurity in substances used in parenteral nutrition [15].

common due to the similarity between these cations, both alkaline earth metals with a very similar chemical behavior. The consequence is that most raw materials containing calcium also contain small amounts of magnesium. The same occurs between sodium and potassium, so that the limit for potassium as an impurity in sodium chloride (according to the BP monograph) is 550 ppm or 0.55 mg potassium per gram sodium salt.

If species such as sodium, potassium, chloride, and sulfate are tolerable in a fairly large range, other toxic species such as arsenic and lead have narrower limits. For arsenic and lead, the limits in pharmacopeial compendia are generally set between 1 and 10 ppm; in fact, these limits are high for such harmful species.

Studies showing the presence of impurities in raw materials are not very frequent, mainly because the quality of raw materials is certified by guarantee bulletins, and products are supposed to be within the quality attested by the certificate.

Two studies with raw materials used to prepare parenteral formulations were carried out to show their content of aluminum and arsenic [15, 16]. It is possible to see in Figures 2 and 3 that aluminum and arsenic were present in all investigated raw materials. There were also different levels of contamination among the substances. While salts such as NaCl and KCl presented low aluminum contaminations, phosphates, gluconate, and also citric acid were relatively contaminated. The authors attributed this difference to the affinity of aluminum to the latter substances. Arsenic showed a more uniform distribution of contamination. With the exception of the amino acid tyrosine, the arsenic level in all substances was below 1 µg/g, not exceeding the limits prescribed by pharmacopeias.

Since analyses to check the presence of impurities in pharmaceutical products are generally carried out with finished products, it is not possible to attribute the presence of contaminants to the raw materials. Unless the raw material itself is checked, many other sources can aggregate impurities to the final product.

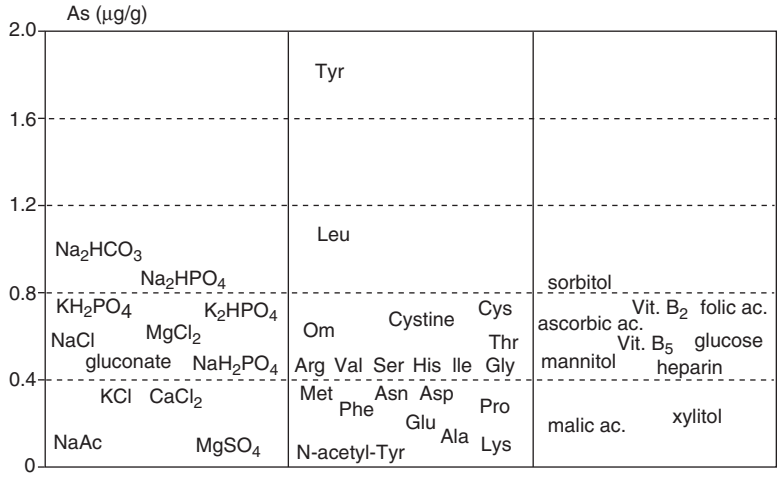


**FIGURE 3** Arsenic level in substances used in formulations for parenteral nutrition [16].

***By-products*** By-products are perhaps the most difficult impurities to summarize since each drug has its own by-products that may appear as impurities. They are synthesized along with active ingredients and are generally difficult to separate, owing to their similarity to the drug of interest. Most are isomeric species, differing from each other by the presence of only a small group or just by the position of a hydrogen atom. Even more difficult to separate and, therefore, purify are chiral isomers. Modern drugs that contain only one chiral isomer as the active ingredient are yet more difficult. Pharmacopeias usually present in the monographs the by-products that can exist as impurity of an active ingredient.

## 2.2 Additives

Additives are all formulation constituents other than the active ingredient. Although additives could be classified into excipients and vehicles (excipients for solid preparations and vehicles for liquid ones), there are several other agents used in pharmaceutical formulations with specific functions such as preservatives, sweeteners, coatings, colorants, antioxidants, surfactants, emulsifying agents, and flavors. Since they comprise a vast amount of products, this section will deal with additives for compounding pharmaceutical products for internal use only [17, 18].

Table 8 lists additives and the possible impurities they might present. The limits for these impurities were taken from manufacturers' specification bulletins. To compound Table 8, products of pharmacopeial grade (USP, BP, Ph. Eur., DAB) were selected from qualified suppliers (Merck, Aldrich, Sigma, Fluka, EMS, Riedel deHaën).

## 2.3 Decomposition of Formulation Constituents

The most relevant endogenous contamination arising from the decomposition of formulation constituents is the generation of peroxides in vitamins, amino acids, and lipid emulsions by action of light and air, that is oxygen.

**TABLE 8 Additives for Pharmaceutical Formulations and Impurities and Their Limits According to Manufacturers (Merck, Aldrich, Sigma, Fluka, EMS, Riedel deHaën)**

| Additive | Impurity | Limit |
|---|---|---|
| | *Solvents* | |
| Ethanol | Acetone and isopropyl alcohol | ≤0.01% |
| | Acetaldehyde and acetal | ≤0.001% |
| | Benzene | ≤0.0002% |
| | Methanol | ≤0.01% |
| | Total of other impurities | ≤0.03% |
| Glycerol | Chloride | ≤0.0010% |
| | Sulfate | ≤0.0010% |
| | Halogen compounds (as Cl) | ≤0.0030% |
| | Heavy metals | ≤0.0005% |
| | Arsenic | ≤0.0001% |
| | Calcium | ≤0.0001% |
| | Cadmium | ≤0.0001% |
| | Mercury | ≤0.0001% |
| | Ammonium | ≤0.0005% |
| | Lead | ≤0.0002% |
| | 1,2,4-Butantriol | ≤0.2% |
| | Residual solvents, class 3[a] | ≤0.5% |
| | Aldehydes | ≤10 ppm |
| | Sulfated ash | ≤0.01% |
| Polyethylene glycol | Dioxane | 10 ppm max |
| | Ethylene glycol and diethylene glycol | 0.4% max |
| | Ethylene glycol (class 2) | 620 ppm max |
| | Ethylene oxide | 1 ppm max |
| | Formaldehyde (HCHO) | 30 ppm max |
| | Heavy metals | 20 ppm max |
| | Sulfated ash | 0.2% max |
| | *Preservatives* | |
| Benzyl alcohol | Peroxide value[a] | ≤5 |
| | Benzaldehye | ≤0.15% |
| | Cyclohexylmethanol | ≤0.10% |
| | Benzene | ≤0.0002% |
| | Chlorobenzene | ≤0.01% |
| | Toluene | ≤0.01% |
| Benzoic acid | Sulfate | ≤0.02% |
| | Heavy metals | ≤0.001% |
| | Arsenic | ≤0.0001% |
| | Cadmium | ≤0.0010% |
| | Copper | ≤0.0010% |
| | Mercury | ≤0.0001% |
| | Lead | ≤0.0005% |
| | Zinc | ≤0.0010% |
| | Halogen compounds (as Cl) | ≤0.01% |
| | Toluene | ≤890 ppm |

**TABLE 8** ***Continued***

| Additive | Impurity | Limit |
|---|---|---|
| Chlorobutanol | Chloride | ≤0.01% |
| | Chloroform | ≤60 ppm |
| | Residual solvents, class 3[a] | ≤0.5% |
| | Sulfated ash | ≤0.1% |
| Methylparaben | Heavy metals | ≤0.001% |
| | Arsenic | ≤0.0003% |
| | Cadmium | ≤0.001% |
| | Copper | ≤0.001% |
| | Mercury | ≤0.0001% |
| | Lead | ≤0.0005% |
| | Zinc | ≤0.001% |
| | Methanol | ≤0.3% |
| | Sulfated ash | ≤0.05% |
| Methylparaben sodium salt | Chloride | ≤0.03% |
| | Sulfate | ≤0.03% |
| | Heavy metals | ≤0.001% |
| | Arsenic | ≤0.0003% |
| | Cadmium | ≤0.001% |
| | Copper | ≤0.001% |
| | Mercury | ≤0.0001% |
| | Lead | ≤0.0005% |
| | Zinc | ≤0.001% |
| Propylparaben sodium salt | Chloride | ≤0.03% |
| | Sulfate | ≤0.03% |
| | Heavy metals | ≤0.001% |
| | Arsenic | ≤0.0003% |
| | Cadmium | ≤0.001% |
| | Copper | ≤0.001% |
| | Mercury | ≤0.0001% |
| | Lead | ≤0.0005% |
| | Zinc | ≤0.001% |
| | 1-Propyl alcohol | ≤0.5% |
| | Sulfated ash | 34–36% |
| Propylparaben | Heavy metals | ≤0.001% |
| | Arsenic | ≤0.0003% |
| | Cadmium | ≤0.001% |
| | Copper | ≤0.001% |
| | Mercury | ≤0.0001% |
| | Lead | ≤0.0005% |
| | Zinc | ≤0.001% |
| | Residual solvents, class 3[a] | ≤0.5% |
| | Sulfated ash | ≤0.05% |
| Potassium sorbate | Arsenic | ≤0.0003% |
| | Cadmium | ≤0.001% |
| | Copper | ≤0.001% |
| | Mercury | ≤0.0001% |
| | Lead | ≤0.0002% |
| | Zinc | ≤0.001% |
| | Aldehydes (as acetaldehyde) | ≤0.15% |
| | Residual solvents, class 3[a] | <0.5% |

**TABLE 8** ***Continued***

| Additive | Impurity | Limit |
|---|---|---|
| Sodium benzoate | Chloride | ≤0.02% |
| | Sulfate | ≤0.01% |
| | Total chlorine | ≤0.03% |
| | Heavy metals | ≤0.001% |
| | Arsenic | ≤0.0001% |
| | Cadmium | ≤0.001% |
| | Copper | ≤0.001% |
| | Mercury | ≤0.0001% |
| | Lead | ≤0.0002% |
| | Zinc | ≤0.001% |
| Sorbic acid | Heavy metals | ≤0.0010% |
| | Arsenic | ≤0.0003% |
| | Cadmium | ≤0.001% |
| | Copper | ≤0.001% |
| | Mercury | ≤0.0001% |
| | Lead | ≤0.0002% |
| | Zinc | ≤0.001% |
| | Aldehydes (as acetaldehyde) | ≤0.15% |
| | Residual solvents, class 3[a] | ≤0.5% |
| | Sulfated ash | ≤0.2% |
| | *Antioxidants* | |
| Ascorbic acid | Heavy metals | ≤0.001% |
| | Ign. residue | ≤0.05% (as $SO_4$) |
| | Chloride | ≤50 mg/kg |
| | Sulfate | ≤20 mg/kg |
| | Copper | ≤5 mg/kg |
| | Iron | ≤2 mg/kg |
| L(+)-Ascorbyl palmitate | Heavy metals | ≤0.001% |
| | Arsenic | ≤0.0003% |
| | Copper | ≤0.0025% |
| | Lead | ≤0.001% |
| | Zinc | ≤0.0025% |
| | Sulfated ash | ≤0.1% |
| | Residual solvents, class 3[a] | ≤0.5% |
| Butyl hydroxyanisole | Arsenic | 0.0003% max |
| | Heavy metals | 0.001% max |
| | Lead | 0.0005% max |
| | Mercury | 0.0001% max |
| | 3-*tert*-Butyl-4-methoxyphenol | 10% max |
| | Hydroquinone | 0.2% max |
| | Sulfated ash | 0.01% max |
| Butyl hydroxytoluene | Arsenic | 0.0003% max |
| | Heavy metals | 0.001% max |
| | Lead | 0.0005% max |
| | Mercury | 0.0001% max |
| | Residual solvents, class 2 (MeOH)[a] | 0.2% max |
| | Sulfated ash | 0.002% max |

**TABLE 8** ***Continued***

| Additive | Impurity | Limit |
|---|---|---|
| Formaldehyde sulfoxylate sodium | Iron | 0.0025% max |
| | Sodium sulfite | 5.0% max |
| | Residual solvents, class 2 (MeOH)[a] | 0.3% max |
| Phosphoric acid | Volatile acids (as $CH_3COOH$) | ≤0.001% |
| | Chloride | ≤0.0005% |
| | Fluoride | ≤0.0010% |
| | Nitrate | ≤0.0003% |
| | Phosphite and hypophosphite (as $H_3PO_3$) | ≤0.02% |
| | Sulfate | ≤0.005% |
| | Heavy metals | ≤0.001% |
| | Arsenic | ≤0.0002% |
| | Cadmium | ≤0.00010% |
| | Copper | ≤0.002% |
| | Iron | ≤0.005% |
| | Mercury | ≤0.0001% |
| | Potassium | ≤0.005% |
| | Sodium | ≤0.03% |
| | Lead | ≤0.0010% |
| | Zinc | ≤0.002% |
| Sodium bisulfite | Arsenic | 0.001% max |
| | Heavy metals | 0.003% max |
| | Iron | 0.005% max |
| Sodium metabisulfite | Chloride | ≤0.01% |
| | Heavy metals | ≤0.001% |
| | Thiosulfate | ≤0.02% |
| | Arsenic | ≤0.0002% |
| | Iron | ≤0.001% |
| | Mercury | ≤0.0001% |
| | Lead | ≤0.0005% |
| | Selenium | ≤0.0006% |
| Sodium thiosulfate | Sulfate and sulfite (as $SO_4$) | ≤0.2% |
| | Heavy metals | ≤0.001% |
| Tocopherol | Heavy metals | ≤0.001% |
| | Arsenic | ≤0.0003% |
| | Copper | ≤0.0025% |
| | Mercury | ≤0.0001% |
| | Lead | ≤0.0005% |
| | Zinc | ≤0.0025% |
| | Methanol | ≤3000 ppm |
| | Pyridine | ≤200 ppm |
| | Toluene | ≤890 ppm |
| | Sulfated ash | ≤0.1% |
| | *Vehicles* | |
| Cellulose powder | Ether-soluble substances | ≤0.15% |
| | Water-soluble substances | ≤1.0% |
| | Heavy metals | ≤0.001% |
| | Sulfated ash | ≤0.3% |

**TABLE 8** ***Continued***

| Additive | Impurity | Limit |
|---|---|---|
| Gelatine | Sulfur dioxide ($SO_2$) | ≤0.004% |
| | Heavy metals | ≤0.001% |
| | Arsenic | ≤0.00008% |
| | Chromium | ≤0.001% |
| | Iron | ≤0.003% |
| | Zinc | ≤0.003% |
| | Peroxide (as $H_2O_2$) | ≤0.001% |
| | Ash | ≤2.0% |
| Lactose | Heavy metals | ≤0.0005% |
| | Arsenic | ≤0.0001% |
| | Copper | ≤0.0025% |
| | Lead | ≤0.00005% |
| | Zinc | ≤0.0025% |
| | Sulfated ash | ≤0.1% |
| Starch | Reducing matter (as maltose) | max 0.7% |
| | Sulfated ash | max 0.4% |
| Sorbitol | Chloride | ≤0.002% |
| | Sulfate | ≤0.006% |
| | Heavy metals | ≤0.0005% |
| | Arsenic | ≤0.00013% |
| | Nickel | ≤0.0001% |
| | Lead | ≤0.00005% |
| | Related substances (mannitol) | ≤2.0% |
| | Reducing sugars after hydrolysis/ total sugar (as glucose) | ≤0.5% |
| | Reducing sugars (as glucose) | ≤0.11% |
| | Sulfated ash | ≤0.02% |
| Sucrose | Chloride | ≤0.0035% |
| | Sulfate | ≤0.005% |
| | Sulfite (as $SO_2$) | ≤0.0010% |
| | Heavy metals | ≤0.0005% |
| | Arsenic | ≤0.0001% |
| | Lead | ≤0.00005% |
| | Sulfated ash | ≤0.02% |
| | Residual solvents, class 3[a] | ≤0.5% |
| Talc | Heavy metals | ≤0.004% |
| | Aluminum | ≤2.0% |
| | Arsenic | ≤0.0003% |
| | Calcium | ≤0.9% |
| | Iron | ≤0.25% |
| | Lead | ≤0.0005% |
| | Asbestos (according to Ph. Eur.) | Not detectable |
| Zinc oxide | Chloride | ≤0.005% |
| | Sulfate | ≤0.02% |
| | Arsenic | ≤0.0005% |
| | Cadmium | ≤0.001% |
| | Iron | ≤0.001% |
| | Lead | ≤0.005% |

**TABLE 8** ***Continued***

| Additive | Impurity | Limit |
|---|---|---|
| | *Chelating Agents* | |
| EDTA | Heavy metals | ≤0.001% |
| | Calcium | ≤0.001% |
| | Iron | ≤0.001% |
| | Magnesium | ≤5 ppm |
| | Nitrilotriacetic acid | ≤0.1% |
| | Sulfated ash | ≤0.1% |
| | *Buffers* | |
| Boric Acid | Sulfate | ≤0.04% |
| | Heavy metals | ≤0.0015% |
| Disodium hydrogen phosphate | Chloride | ≤0.02% |
| | Fluoride | ≤0.001% |
| | Sulfate | ≤0.05% |
| | Heavy metals | ≤0.001% |
| | Arsenic | ≤0.0002% |
| | Cadmium | ≤0.0001% |
| | Iron | ≤0.002% |
| | Mercury | ≤0.0001% |
| | Lead | ≤0.0004% |
| Sodium dihydrogen phosphate | Chloride | ≤0.005% |
| | Fluoride | ≤0.001% |
| | Hydrogen phosphate ($HPO_4$) | ≤0.5% |
| | Sulfate | ≤0.01% |
| | Heavy metals | ≤0.0005% |
| | Arsenic | ≤0.0002% |
| | Cadmium | ≤0.0001% |
| | Iron | ≤0.001% |
| | Lead | ≤0.0004% |
| | Mercury | ≤0.0001% |

[a]Residual solvents; peroxide value: mmol peroxide/L.

Lipid emulsions are essential components of parenteral nutrition. However, due to the amount of polyunsaturated fatty acids (PUFA), it is possible that chemical degradation occurs, forming hydroperoxides.

The exposure of lipid emulsions to room light and spotlight simulating phototherapy conditions in a neonatal intensive care unit for 24 h increases the level of hydroperoxides in these formulations by up to 60 times [19]. Lipid admixtures stored in ethyl vinyl acetate (EVA) bags presented a significant formation of lipid hydroperoxide. After one month under stressful conditions (gas-permeable container, 40°C), the peroxide value (PV = millimoles of peroxide per liter) was 450 times higher compared with controls (nitrogen-overlaid glass bottles) [20]. The results presented in Table 10 show that in contrast to lipids stored in closed glass containers gassed with nitrogen, significant and time-dependent peroxide formation occurred in all-in-one (AIO) plastic bags. Moreover, the lower PV of the lipid stored in V90 bags (polypropylene–polyamide 7 : 3, double layered) compared with EVA bags

(single layered) is indicative of the decreased oxygen permeability of the double-layered material and therefore the stabilization of the PV, although the increase of the PV occurred slowly in Intralipid samples stored in V90 bags.

Differences were found in the PV of Intralipid (LCT = long-chain triglyceride) and Lipofundin (MCT = medium-chain triglyceride) samples. Initial PV in Intralipid (0.02 mmol/L) was lower than in Lipofundin (0.10 mmol/L), whereas in the latter the formation rate of PV was slower over time. This behavior is due to the double PUFA content of Intralipid.

Polyunsaturated fatty acids are oxidized by enzymatic and nonenzymatic pathways. Nonenzymatic oxidation is a free-radical mediated peroxidation. It is a chain reaction providing a continuous supply of free radicals that initiate further peroxidation. The whole process can be depicted as follows [21]:

$$RH + X^{\bullet} \rightarrow R^{\bullet} + XH$$

$$R^{\bullet} + O_2 \rightarrow ROO^{\bullet}$$

$$ROO^{\bullet} + RH \rightarrow ROOH + R^{\bullet}$$

The reaction is initiated by any existing free radical ($X^{\bullet}$), by light, or by metal ions that react with lipids (RH) generating peroxy radicals ($ROO^{\bullet}$) and finally lipid hydroperoxides (ROOH). Besides hydroperoxides, malondialdehyde (MDA), ethane, and pentane can also be formed if PUFA with three or more double bonds and if ω-3 and ω-6 PUFA are present, respectively [14]. The formation of pentane (2 μmol/L) and MDA (10 μmol/L), along with hydroperoxides was observed in a lipid emulsion (Intralipid 10%) [22]. Malondialdehyde was also found in parenteral nutrition admixtures for newborn infants, in 12 samples analyzed, MDA levels varied from 1632 to 14,679 nmol/L [23]. Pironi and colleagues [24] also found MDA in fat emulsions. They compared three 20% lipid emulsions containing different amounts of PUFA and α-tocopherol, 24 hours after admixing them to make AIO solutions. They concluded that lipid peroxide generation is directly related to PUFA content and inversely related to the α-tocopherol/PUFA ratio of the emulsion. Moreover, MDA was also formed in these samples, presenting an increased concentration in the soybean-oil-containing emulsions when compared with the olive oil emulsion. Table 10 shows the PV and MDA measured in these samples.

Peroxide formation has also been attributed to a prooxidant action of vitamin E on PUFA, as vitamin E is another component present in lipid emulsions [19, 25]. Lipid oxidation occurring despite high concentrations of vitamin E (tocopherol) may appear surprising since the latter is generally regarded as the most efficient chain-breaking, lipid-soluble antioxidant [26]. Studies have revealed, however, that for vitamin E to be an efficient antioxidant in an isolated lipoprotein emulsion, it requires a suitable "co-antioxidant," such as vitamin C [27, 28]. The co-antioxidant eliminates the vitamin E radical formed in the interaction of the vitamin with initiating radical oxidants, which can have an adverse effect by promoting lipid peroxidation.

Through the measurement of triglyceride hydroperoxides in Intralipid samples stored in glass and plastic syringes exposed to direct light or after being wrapped with aluminum foil, it was possible to demonstrate that Intralipid is highly oxidizable

**TABLE 9 Peroxidation of Intralipid 20% and Lipofundin MCT 20% during Storage: Influence of Container Material, Light Exposure, and Temperature**

| Exposure, Condition, Time (day) | Nutrimix 2/3 Bag (EVA) 20–27°C Daylight[a] | Nutrimix 2/3 Bag (EVA) 20–27°C Light Protetion | 3-chamber Bag (V 90) 20–27°C Daylight[a] | Control sample Closed Glass Bottle 20–27°C Daylight[a] |
|---|---|---|---|---|
| Intralipid 20% | | | | |
| 1 | | | | 0.02 ± 0.006[c] |
| 5 | 0.47 ± 0.016[b] | 0.06 ± 0.004[b,c] | 0.33 ± 0.011 | |
| 8 | | | | 0.02 ± 0.005 |
| 9 | 0.84 ± 0.010 | 0.05 ± 0.007 | 0.50 ± 0.017 | |
| 14 | 1.46 ± 0.020 | 0.17 ± 0.002 | 0.60 ± 0.014 | |
| 15 | | | | 0.02 ± 0.001 |
| 19 | 1.87 ± 0.027 | 0.27 ± 0.017 | 0.76 ± 0.018 | |
| 22 | 2.48 ± 0.038 | 0.40 ± 0.013 | 0.92 ± 0.025 | |
| 28 | | | | 0.02 ± 0.005 |
| 29 | 2.93 ± 0.052 | 0.52 ± 0.023 | 1.24 ± 0.032 | |
| *Lipofundin MCT 20%* | | | | |
| 1 | | | | 0.11 ± 0.006[c] |
| 5 | 0.57 ± 0.018[b] | 0.19 ± 0.017[d,e] | 0.64 ± 0.014 | |
| 8 | | | | 0.10 ± 0.018 |
| 9 | 0.88 ± 0.024 | 0.35 ± 0.002 | 0.69 ± 0.010 | |
| 14 | 1.31 ± 0.048 | 0.54 ± 0.025 | 0.64 ± 0.021 | |
| 15 | | | | 0.08 ± 0.003 |
| 19 | 1.59 ± 0.064 | 0.61 ± 0.018 | 0.67 ± 0.036 | |
| 22 | 1.99 ± 0.076 | 0.84 ± 0.007 | 0.70 ± 0.054 | |
| 28 | | | | 0.10 ± 0.032 |
| 29 | 2.48 ± 0.040 | 0.99 ± 0.044 | 0.69 ± 0.032 | |

*Source*: Form ref. 20.

[a,b] $P < 0.001$ (Nutrimix 2/3 with light protection).

[c,d] $P < 0.002$ (Nutrimix 2/3 with light protection versus control samples).

**TABLE 10 Lipid Peroxide (LPX) and MDA Concentration in Lipid Emulsions**

| | LPX (μM/L) | | | | MDA (μM/L) | | | |
|---|---|---|---|---|---|---|---|---|
| Emulsions | Bottles ($n$ = 6) | No fat PN ($n$ = 6) | AIO–T0 ($n$ = 6) | AIO–T24 ($n$ = 6) | Bottles ($n$ = 6) | No fat PN ($n$ = 6) | AIO–T0 ($n$ = 6) | AIO–T24 ($n$ = 6) |
| Soybean oil | 28.8 ± 29.9 | <2 | <2 | 11.3 ± 15.0 | 8.8 ± 3.1 | 4.9 ± 0.8 | 5.0 ± 0.3 | 6.9 ± 1.0 |
| Soybean oil + MCT | 1.7 ± 0.5 | <2 | <2 | 5.2 ± 8.8 | 3.7 ± 0.4 | 3.4 ± 0.4 | 4.1 ± 0.3 | 7.8 ± 1.8 |
| Olive oil | 4.1 ± 2.8 | <2 | <2 | 5.5 ± 6.4 | 0.7 ± 0.1 | 5.4 ± 0.3 | 5.3 ± 0.3 | 3.3 ± 0.4 |

*Source*: From ref. 24.

*Note*: Values are mean ± standard deviation of six bags (one sample per bag, analyzed in triplicate).
AIO, all-in-one mixture immediately after the addition of lipid emulsion; AIO–T24, all in one mixture 24 h after the addition of lipid emulsion.

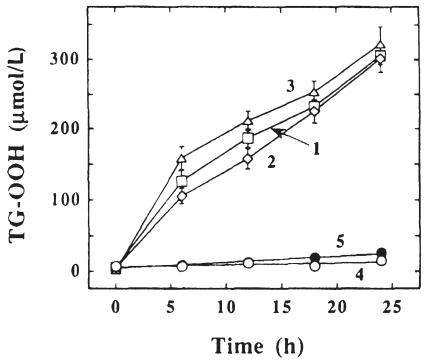


**FIGURE 4** Kinetics of accumulation of triglyceride hydroperoxide in Intralipid exposed to phototherapy light under conditions that mimic the situation in an NICU (neonatal intensive care unit). A solution of 20% Intralipid was dispensed into a glass reservoir containing plastic tubing (line 3), with aluminum foil (line 4), or a plastic reservoir after supplementation with 1 mmol/L sodium ascorbate (line 5) and exposed to phototherapy light for up to 24 h. For all samples the light fluxes measured were 11.7–25.9 μW/cm$^2$ per nanometer for 24 h. At the time points indicated, aliquots were withdrawn, extracted, and their hexane extracts analyzed for the presence of triglyceride hydroperoxides and α-tocopherol as described in the methods section [19].

under routine clinical conditions, in spite of its vitamin E content (≅60 μmol/L γ-tocopherol and 20 μmol/L each of the α and δ isomers).

The problem is particularly evident for neonates undergoing phototherapy. Besides the fact that lipids are used extensively to supply premature babies with calories, bags and extension sets are exposed for longer periods of time to the intense radiation of spotlights due to the low infusion rates of neonatal solutions. Figure 4 depicts the evolution of triglyceride hydroperoxide formation in a 20% Intralipid emulsion under different conditions of light exposure. It shows that lipid oxidation occurs despite the high concentration of vitamin E present in the samples. Decomposition can be prevented by wrapping containers and tubing sets in aluminum foil or by adding ascorbate to the infusate.

Lipid peroxides are also able to react with other components of parenteral nutrition admixtures (trace elements), causing a drop in pH with the subsequent potential for physical–chemical instability [29]. Table 11 shows the peroxide value and the pH drop in a pure lipid emulsion and a lipid-containing AlO admixture stored in EVA bags under different conditions of temperature and light exposure in the presence and absence of trace elements.

Peroxide formation has also been observed in multivitamin solutions for parenteral nutrition. Lavoie and co-workers [30] have studied the action of light, air, and composition on the stability of multivitamin formulations, and also total parenteral nutrition (TPN) admixtures containing and not containing vitamins and fatty acids. They analyzed the generation of peroxide in multivitamin solutions and in TPN for adults and neonates. The analysis of multivitamin solutions for enteral use revealed the presence of peroxides at the initial opening of the bottle. The levels were higher in Poly-Vi-Sol (vitamin A, Vitamin D, and vitamin C, vitamin $B_1$, riboflavin, and

**TABLE 11 Peroxide Value and pH Drop in Pure Lipid Emulsion and AIO Admixture Stored in EVA Bags under Different Conditions of Temperature and Light Exposition in Presence and Absence of Trace Elements**

| Sample | Conditions | Parameters | Without Trace Elements | With Trace Elements |
|---|---|---|---|---|
| AIO | 2–8°C | PV (mmol/L) | 0.04 | 0.19 |
| | Light protected 29 days | pH drop | 0.01 | 0.02 |
| AIO | 20–30°C | PV (mmol/L) | 0.52 | 1.92 |
| | Light exposed 29 days | pH drop | 0.03 | 0.11 |
| Intralipid 20% | 40°C | PV (mmol/L) | 2.77 | 18.04 |
| | Light protected 14 days | pH drop | 0.77 | 1.54 |

*Note*: PV: peroxide value.

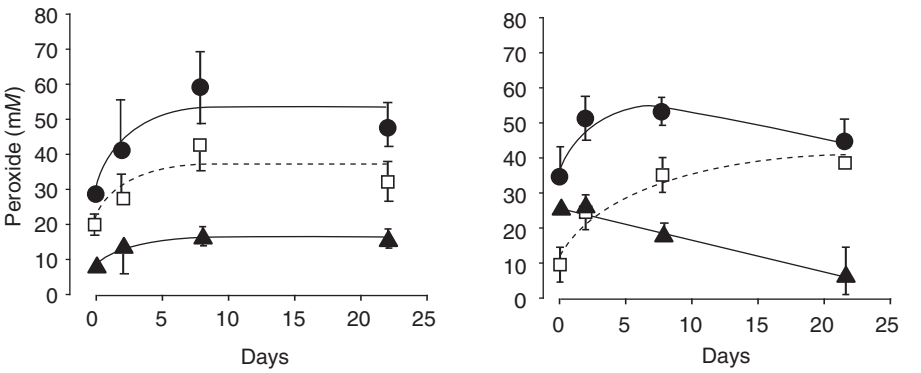


**FIGURE 5** Total peroxide (circles), catalase-resistant peroxides (squares), and $H_2O_2$ as the difference (triangles) were measured in three lots of oral multivitamins without riboflavin (Tri-Vi-Sol) and three lots of oral multivitamins with 0.6 mg of riboflavin (Poli-Vi-Sol) by time after the initial opening of the bottle. Compared with the preparation without riboflavin, the level of $H_2O_2$ was initially higher ($P < 0.01$) in (Poli-Vi-Sol), and that level dropped over time ($P < 0.05$). In both multivitamin preparations, the levels of catalase-resistant and total peroxide rose ($P < 0.05$) until day 8. Data represent the mean ± standard error of the mean. Variations too small relative to the symbol are not shown [30].

niacinamide) than in Tri-Vi-Sol (vitamin A, vitamin D, and vitamin C) (Fig. 5) [30].

The authors attributed the difference in the peroxide content of the preparations to riboflavin. This vitamin catalyses the photoinduced reaction between ascorbate and oxygen, which leads to hydrogen peroxide generation. The drop in hydrogen peroxide concentration over time observed with Poly-Vi-Sol is explained by the transformation of hydrogen peroxide into a more reactive species, which in turn could react with other components of the formulation. In contrast, the increase in catalase-resistant peroxides in the Tri-Vi-Sol preparation suggests a greater contribution of air in the peroxide generation. In comparison to the formulation for parenteral use, the enteral multivitamin presented levels of 100 times higher, where catalase-resistant peroxides represented the bulk of peroxides. To study the effect of air and light on the generation of peroxides in multivitamins for TPN, the authors

limited the experiment to fat-free TPN since lipids are also able to generate peroxides [31]. In neonate solutions, the peroxide concentration ranged from 190 to 300 μmol/L in sets unprotected from light in contrast to the values for sets protected from light, which ranged from 60 to 130 μmol/L. In the adult formulation, peroxide reached less than one-tenth of the concentration measured in the neonatal preparation (Fig. 6). The authors attributed the difference to the nutrient composition, as the adult formulation was four times lower in vitamins and higher in final amino acid and glucose concentrations. Amino acids and glucose contributed to a decrease in the concentration of peroxides [32].

It is possible to see in Figure 6 that protecting solutions from light and air decreases peroxide generation in these formulations. However, because of the high peroxide concentration in neonatal solutions, protection from air during the infusion is not efficient.

The action of vitamins in the presence of lipids was also investigated by these researchers [33]. The generation of peroxide was compared between TPN preparations left at dark and at daylight for 6 h. Four different formulations containing and not containing vitamins and lipids were included in the experiment. The results, presented in Figure 7 allowed for the conclusion that lipids had a significant but minor additive effect compared with vitamins in generating peroxides. Contamination of TPN by air during compounding accounted for the photoinduced generation of peroxide in TPN solutions. It was, however, more effective to protect the solution from light exposure than to avoid contact with oxygen.

Since ascorbate reduces photooxidation of lipid emulsions and multivitamin preparations (see Figure 4) [19], Lavoie et al. [34] studied the formation of oxidative by-products of vitamin C in multivitamins exposed to light. They found that the loss of ascorbic acid in photoexposed multivitamin preparations was associated with the generation of products other than dehydroascorbate and 2,3-diketogulonic acid, which are the usual products of vitamin C oxidation. The authors showed that hydrogen peroxide at concentrations found in TPN solutions induced the transformation of dehydroascorbate into new, biologically active compounds that had the potential to affect lipid metabolism. They believe that these species have peroxide and aldehyde functions [35].

Since air (oxygen) is one of the factors responsible for peroxide generation in lipid emulsions and vitamin solutions, Balet et al. [36] compared multilayered versus single-layered (EVA) bags in terms of oxidation of parenteral nutrition solutions. They measured PV, α-tocopherol, and ascorbic acid at the moment of admixing and after 6 and 14 days in 24 parenteral solutions. Admixtures in multilayered bags showed less oxidation than in EVA bags; no important difference was observed in α-tocopherol content, but just after 6 days storage, ascorbic acid and dehydroascorbic acid disappeared in the EVA bags.

Sodium metabisulfite is an antioxidant agent widely used in pharmaceutical preparations to reduce or prevent oxidation. There are some studies, however, that have shown that metabisulfite, under specific conditions, may have indirect oxidant properties. Baker et al. [37] demonstrated that sulfite propofol emulsion, but not EDTA propofol emulsion, underwent chemical changes during a simulated intravenous infusion. Compounds were identified as propofol oxidation products. The increase of propofol oxidation products demonstrated that sulfite from metabisulfite created a strong oxidant environment when air was introduced. Lavoie et al. [38]

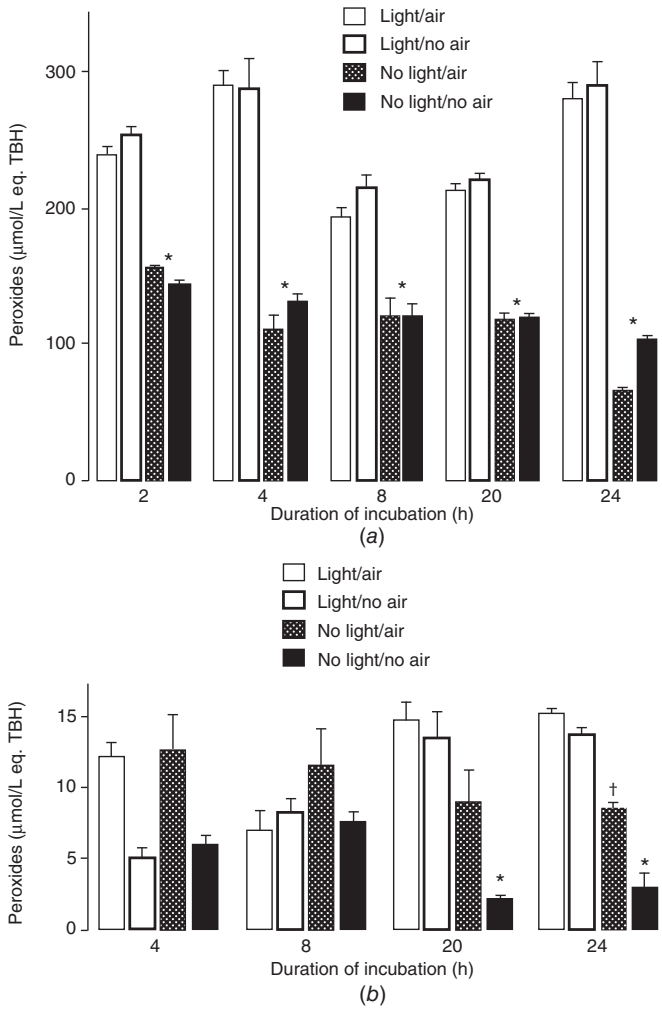


**FIGURE 6** (*a*) Peroxide concentration in a neonatal fat-free solution of parenteral nutrition run through four different infusion sets: with or without air inlet, protected and unprotected from light (lux). Photoprotection was associated with a significantly lower peroxide content (indicated by *; $P < 0.05$). Protection from light resulted in a greater prevention against the generation of peroxides than did protection from air contamination. (*b*) Peroxide concentration in an adult solution of fat-free parenteral nutrition run through four different infusion sets: with or without air inlet, protected and unprotected from light (lux). The concentration of peroxides is 10-fold lower than in the neonatal solution. Protection from light and air was not associated with an overall effect on peroxide generation. Protection from light and air resulted in a significant decrease in peroxide generation at 20 and 24 h of infusion (indicated by *; $P < 0.05$), whereas photoprotection alone resulted in a significant difference at 24 h (indicated by +; $P < 0.05$). Results are expressed as mean ± standard error of the mean (SEM); $n = 3$. TBH = *tert*-butylhydroperoxide [31].

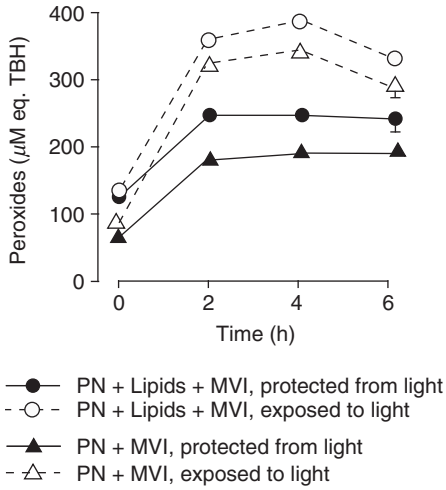


**FIGURE 7** Influence of a lipid emulsion and daylight on peroxide levels in freshly prepared solutions of parenteral nutrition containing multivitamins (PN + MVI and PN + Lipid + MVI). (PN = parenteral nutrition; MVI = multi vitamin preparation.) The data represent the mean ± SEM, $n = 3$; the variations are not depicted because of their small size relative to the symbols. The peroxide content rose significantly over time ($P < 0.001$), and exposure to daylight had a significant effect on peroxide generation ($P < 0.001$) [33].

showed that sulfite has been able to cause the oxidation of lipids. The reaction of sulfite with oxygen may lead to several sulfite-derived oxidant species [39].

Besides the oxidation to sulfate ($SO_3^{2-} \rightarrow SO_4^{2-} + 2e^-$), sulfite may undergo a one-electron oxidation leading to the formation of the reactive sulfite radical ($SO_3^{2-} \rightarrow SO_3^{\bullet -} + e^-$), which reacts with oxygen forming two strong oxidant species: sulfite peroxyl ($SO_3^{\bullet -} + O_2 \rightarrow SO_3OO^{\bullet -}$) and sulfate radical ($SO_3^{\bullet -} + O_2 \rightarrow SO_4^{\bullet -}$). These species and the sulfite radical itself react as oxidants, turning sodium metabisulfite into a prooxidant agent, depending on the circumstances.

Peroxidation and free-radical formation should be considered as important aspects of pharmaceutical stability and quality of parenteral nutriton and intravenous drugs. Peroxidation and free-radical formation depend on environmental factors, such as storage conditions and container material, but are also influenced by formulation components or additives such as tocopherols and metabisulfite. Since the generation of these harmful species occurs generally at the time of use, manufacturing quality controls fail in demonstrating their existence.

## 3 EXOGENOUS IMPURITIES

### 3.1 Residual Solvents

Residual solvents are organic volatile chemicals that remain in active substances, excipients, and other pharmaceutical products after processing. In spite of their toxic properties, solvents play an important role in the production of pharmaceuticals, during the synthesis, separation, or purification, and their use cannot be avoided. Solvents in this category do not include those used as excipients.

**TABLE 12 General Characteristics of Residual Solvents**

| Class | Action | Characteristics |
|---|---|---|
| 1 | Solvents to be avoided | Strongly suspected human carcinogens and environmental hazards |
| 2 | Solvents to be limited | Nongenotoxic animal carcinogens or possible causative agents of other irreversible toxicity such as neurotoxicity or teratogenicity |
| 3 | Solvents with low toxic potential | No health-based exposure limit is needed; solvents with permitted daily exposure of 50 mg or more per day |

*Source*: From ref. 40.

Residual solvents are organic volatile chemicals that were not completely removed by practical manufacturing techniques and may, therefore, be contaminants in pharmaceutical products.

The International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH) [40] has adopted impurities guidelines for residual solvents that prescribe limits for the amount of residual solvents allowed.

Residual solvents are divided into three classes. Class 1 solvents are those known to cause toxic effects and should be avoided in the production of active substances and excipients. Class 2 solvents present less severe toxicity than class 1, and class 3 solvents have such low toxic potential that exposure limits are not necessary. Table 12 presents the general characteristics of the solvents included in each class, and Table 13 lists the solvents and their concentration limit in pharmaceutical products.

### 3.2 Containers

The primary function of packaging is to provide adequate protection. Pharmacopeial compendia have established requirements for containers based on drug form characteristics. Thus, while for capsules and tablets the requirements are generally related to the design of the container (e.g., tight, well closed), for injectables, and ophthalmic, and inhalation products, materials of construction are also addressed, as compatibility is a very important issue for this kind of dosage. Considering the purpose of this chapter, only those packaging materials whose interaction with the formulation is a factor will be focused on here.

Packaging components for pharmaceuticals are basically made of glass and polymeric materials such as plastics and rubbers.[1] In spite of this simple classification, glass, plastic, and rubber are not elementary materials but rather complex mixtures.

The evaluation of the chemical stability of a packaging component depends on the likelihood of packaging component–dosage form interaction and is usually

[1]Although metallic packaging is also used for pharmaceutical products, its use is restricted to blisters' supports, which have no contact with the formulation, and tubes for ointments.

**TABLE 13 Solvents Included in Each Class and Their Concentration Limit in Pharmaceutical Products**

| Solvent | Concentration Limit (ppm) | Solvent | Concentration Limit (ppm) |
|---|---|---|---|
| *Class 1* | | *Class 3* | |
| | 2 | Acetic acid | — |
| Benzene | 4 | Acetone | — |
| 1,2-Dichloroethane | 5 | Anisole | — |
| 1,1-Dichloroethane | 8 | 1-Butanol | — |
| 1,1,1- Trichloroethane | 1500 | 2-Butanol | — |
| | | Butyl acetate | — |
| *Class 2* | | *tert*-Butyl methyl ether | — |
| Acetonitrile | 410 | Cumene | — |
| Chlorobenzene | 360 | Dimethyl sulfoxide | — |
| Chloroform | 60 | Ethanol | — |
| Cyclohexane | 3880 | Ethyl acetate | — |
| 1,2-Dichloroethene | 1870 | Ethyl ather | — |
| Dichloromethane | 600 | Ethyl formate | — |
| 1,2-Dimethoxyethane | 100 | Formic acid | — |
| *N*,*N*-Dimethylacetamide | 1090 | Heptane | — |
| *N*,*N*-Dimethylformamide | 880 | Isobutyl acetate | — |
| 1,4-Dioxan | 380 | Isopropyl acetate | — |
| 2-Ethoxyethanol | 160 | Methyl acetate | — |
| Ethylene glycol | 620 | 3-Methyl-1-Butanol | — |
| Formamide | 220 | Methyl ethyl ketone | — |
| Hexane | 290 | 2-Methyl-1-propanol | — |
| Methanol | 3000 | Pentane | — |
| 2-Methoxyethanol | 50 | 1-Pentanol | — |
| Methyl butyl ketone | 50 | 1-Propanol | — |
| Methylcyclohexane | 1180 | 2-Propanol | — |
| *N*-methylpyrrolidone | 4840 | Propyl acetate | — |
| Nitromethane | 50 | Tetrahydrofuran | — |
| Pyridine | 200 | | |
| Sulfolane | 160 | | |
| Tetralin | 100 | | |
| Toluene | 890 | | |
| 1,1,2-Trichloroethene | 80 | | |
| Xylene | 2170 | | |

*Source*: From ref. 40.
— = not limited.

carried out by exposing a sample of the packaging component to an appropriate solvent at elevated temperatures. The resulting extract should be analyzed for extractables. Thus, for glass samples the released alkalinity should be evaluated, whereas for plastics and rubbers the amount of extractables, as well as their nature, are to be determined. The elevated temperature has the purpose of increasing the extraction rate and simulating in a short period of time a longer exposure time at room temperature. The solvent used for the extraction test should present the same

**TABLE 14 Glass Classification According to Pharmacopeial Prescription and Their Pharmaceutical Usages**

| Glass Type | General Description | Uses |
|---|---|---|
| I | Highly resistant borosilicate glass | Parenteral preparations |
| II | Treated soda–lime glass | Acidic or neutral parenteral preparations |
| III | Soda–lime glass | Not for parenteral preparations |
| NP | Soda–lime glass | Oral or topic use |

**TABLE 15 Typical Composition of Glass Containers According to the Pharmacopeial Classification**

| Component | Type I, Borosilicate (%) | Types II, III, IV, Soda–Lime (%) |
|---|---|---|
| $SiO_2$ | 70 | 73 |
| $B_2O_3$ | 10 | — |
| $Na_2O$ | 9 | 14 |
| $Al_2O_3$ | 6 | 2 |
| BaO | 2 | — |
| $K_2O$ | 1 | — |
| CaO | 1 | 7 |
| MgO | 0.5 | 4 |
| ZnO | 0.5 | — |

*Source*: From Corning, Life Science Catalogue: Technical information Description of Glasses Used in Corning Labware.

propensity to interact with the packaging material as the dosage form. Although it is desirable that the dosage form itself be used in the test, pharmacopeial compendia prescribe tests with standard solvents such as purified water, drug vehicle, and isopropyl alcohol.

Even when impurities and degradation profiles of a drug substance have been established and containers comply with guidelines, some unexpected drug–container interactions can occur during the sterilization procedure or shelf life.

This section includes the most important container materials: glass, plastic, and elastomers and addresses the contamination issues for injectable formulations.

***Glass Containers*** Glass containers for packaging pharmaceutical preparations should meet stability specifications, which are related to the presence of certain components in the structure of the glass and to the formulation type and use. According to pharmacopeial prescriptions, four different types of glass may be used for storing pharmaceutical preparations. Table 14 presents the classification of glass containers in accordance with the product's nature and pharmaceutical usage.

Though mostly made of silica, the different glass types are obtained by adding or subtracting certain glass components. Table 15 lists the principal components of pharmaceutical glasses, based on the criteria presented in Table 14. Type I glass must have at least 10% boron oxide and a higher concentration of aluminum oxide than the ordinary soda–lime glass to improve its resistance. The reduced amount of sodium oxide lessens the solubility of type I glass in water. Type II glass is made of

common soda–lime glass after a dealkalinization process on the inside surface at the time of manufacture. In this process, known as sulfur treatment, sulfur dioxide is introduced as newly formed bottles pass from the forming machine, and it reacts at the surface of the glass to form sodium sulfate, according to the equation:

$$SO_2 + \frac{1}{2}O_2 + Na_2O \xrightarrow[\text{(glass)}]{} Na_2SO_4$$

The container is delivered with a haze of sodium sulfate, which should be rinsed away before filling.

Stability test for glass containers prescribed by pharmacopeial compendia is limited to the action of water on the glass surface to measure the alkalinity released (water attack). After autoclaving the container filled with water at 121°C for 60 min, 100 mL of the resulting extract is titrated with 0.02 *N* $H_2SO_4$. The volume of the acidic titrant should not be greater than a specific value based on the container's capacity. Briefly, types I and II glass containers with a capacity of less than 10 mL should consume less than 2.0 mL titrant, whereas for type III glass containers the limit is 20 mL. For containers with a capacity of between 20 and 500 mL, the titrant consumption should be from 0.8 to 0.2 mL for types I and II glass and from 8.0 to 2.2 mL for type III glass. For non-parenteral (NP) glass, no limit of titrant is established.

The released alkalinity is, however, not enough to show the actual stability of a glass container. Glass is not as inert as it appears, and water is not the only substance that attacks the glass surface. In spite of this, the only test prescribed at USP for toxic impurities in glass is arsenic (<661> containers, USP 27); type I glass containers should be assayed according to the water attack procedure and present not more than 0.1 ppm As. Though not listed as a glass constituent, As may be used as a fining agent in the glass industry (see below) and, therefore, may be present in the glass structure.

Studies have shown that even type I glasses are not completely inert. Bohrer et al. [41] showed that the hydrolytic resistance of glass containers is diminished in the presence of some substances commonly present in infusion solutions. Table 16 presents the released alkalinity and also the concentration of some glass constituents (silicate, sodium, boron, aluminum) in the aqueous extract obtained after the hydrolytic resistance test.[2] These data show that, despite complying with pharmacopeial prescription, glass can release its constituents by action of species in the solutions and even in pure water. Basic solutions of bicarbonate and gluconate presented the highest level of glass constituents extracted, confirming the potential of basic solutions to attack and dissolve the glass network. Glucose and citric acid interacted with the glass surface, selectively extracting not only Al but also Cu and Pb (traces also present in the glass). This specific action of citrate and glucose could be related to their metal-complexing ability.

The presence of arsenic as a contaminant in pharmaceutical products is a pharmacopeial concern. Practically all monographs present limits for arsenic, ranging from 0.1 to 3 ppm. Bohrer et al. have shown that arsenic is a ubiquitous contaminant

[2]The hydrolytic resistance test described in this chapter is the "powdered glass test" and not the "water attack test" as described above; see Appendix <661>, USP 27 [1].

**TABLE 16 Volume $H_2SO_4$ Solution Consumed in Titration of Solutions before and after Heating in Contact with Glass Mass for 30 min at 121°C and Species Extracted from Glass during the USP Powdered Glass Test Carried out with Clear Ampoules in Presence of Different Solutions**

| | Volume 0.02 $N$ $H_2SO_4$ (mL) | | Species Concentration (mg/L) ± SD ($n$ = 3) | | | | | |
|---|---|---|---|---|---|---|---|---|
| Sample | Before Heating | After Heating | Na | Silicate | Borate | Al | Cu | Pb |
| NaCl | [a] | 0.56 ± 0.01 | — | 8.30 ± 2.01 | 2.22 ± 0.11 | 0.15 ± 0.01 | 0.09 ± 0.02 | 0.03 ± 0.01 |
| KCl | [a] | 0.69 ± 0.01 | 14.42 ± 1.51 | 9.96 ± 1.71 | 1.58 ± 0.03 | 0.19 ± 0.01 | 0.14 ± 0.08 | 0.04 ± 0.02 |
| $CaCl_2$ | [a] | 0.45 ± 0.02 | 13.41 ± 1.33 | 7.31 ± 2.08 | 2.00 ± 0.11 | 0.18 ± 0.01 | 0.06 ± 0.01 | 0.02 ± 0.01 |
| $MgCl_2$ | [a] | 0.13 ± 0.01 | 15.25 ± 1.64 | 6.54 ± 0.53 | 3.25 ± 0.21 | 0.16 ± 0.01 | 0.09 ± 0.03 | 0.02 ± 0.01 |
| Sodium gluconate | 2.68 ± 0.22 | 1.83 ± 0.05 | — | 246.0 ± 18.7 | 1.91 ± 0.09 | 2.04 ± 0.02 | 0.49 ± 0.07 | 0.05 ± 0.01 |
| Sodium hydrogen phosphate | [a] | 10.22 ± 0.91 | — | 10.21 ± 3.41 | 2.66 ± 0.09 | 4.93 ± 0.12 | 0.05 ± 0.01 | 0.05 ± 0.02 |
| Potassium hydrogen phosphate | [a] | 8.78 ± 0.95 | 18.38 ± 2.34 | 45.32 ± 3.76 | 1.73 ± 0.06 | 4.30 ± 0.07 | 0.13 ± 0.04 | 0.04 ± 0.02 |
| Sodium bicarbonate | 24.3 ± 0.41 | 12.68 ± 1.03 | — | 271.7 ± 23.2 | 5.98 ± 0.14 | 5.22 ± 0.24 | 0.88 ± 0.09 | 0.06 ± 0.03 |
| Citric acid | [a] | 1.92 ± 0.04 | 19.82 ± 1.90 | 16.46 ± 0.87 | 1.44 ± 0.10 | 6.36 ± 0.53 | 0.49 ± 0.03 | 0.10 ± 0.05 |
| Glucose | [a] | 1.08 ± 0.05 | 13.86 ± 1.22 | 11.10 ± 0.88 | 3.18 ± 0.17 | 6.34 ± 0.33 | 0.28 ± 0.01 | 0.15 ± 0.03 |

*Note*: Concentration of the solutions: 0.01 mol/L [41].

[a]pH ≤ 7.

in raw materials [16], but its presence in final products occurs in concentrations much higher than in the corresponding raw substances [12]. According to the authors, the main source of arsenic in injectable formulations is glass containers [16]. As mentioned above, glass can contain arsenic because arsenic oxide(III) is a fining agent to improve glass transparency [42], a feature especially important for solutions for intravenous administration that must be subjected to visual inspection of the content before use [43]. Arsenic oxide(III) is supposed to react with potassium nitrate in the glass melt to release oxygen and nitrogen oxides. These gases form large bubbles that rapidly rise to the surface, stirring the bath and sweeping small bubbles formed by the decomposition of batch materials. Table 17 shows the arsenic levels measured by these authors in solutions for parenteral nutrition. Among these products, sodium bicarbonate and calcium gluconate are the most contaminated, presenting levels even higher than the limit of 0.1 mg/L for arsenic in infusion solutions.

Other elements such as chromium, barium, and zinc have also been found in solutions for parenteral nutrition, and, though not stated in the studies, the origin of these elements is probably the glass packaging. Since glass is an inorganic material, inorganic species may arise as contaminants from glass containers. Exceptions could be Zn and Ba, which, besides being present in type I glasses, are components of plastic additives as well. Table 18 summarizes inorganic elements as contaminants in different infusion solutions and their respective concentrations.

Data on Al are being presented separately because of its special behavior and toxicity. Since the discovery in 1976 by Alfrey and co-workers [48] that Al can accumulate in patients with reduced renal function causing toxic manifestations including neurological diseases, much has been done to reveal and reduce aluminum contamination sources. Nowadays, Al is also considered toxic for patients with normal renal function who receive parenteral nutrition, mainly preterm infants [49]. Recognized sources of Al include water used for dialysis, solutions for total parenteral nutrition, and oral aluminum-containing compounds. Though additives and raw materials are sources as well, the most important source of Al in solutions for parenteral nutrition is the storage of parenteral preparations in glass containers. Figure 8 shows the Al level in different kinds of solutions stored in plastic and glass containers for a period of three months. While not more than 50 μg/L Al was leached from polyethylene (PE) containers, over 3000 μg/L Al was found in solutions stored in glass containers, confirming that glass is a valid and continual source of Al.

The same authors also showed that though glass is the source, Al leaching depends on the nature of the substance in contact with its surface [51, 52]. In an experiment with glass and an ion exchanger containing Al attached, complexing agents and also amino acids were able to extract Al from both sources, but following an extraction yield related to the affinity of the ligand for the metal. The results of this interaction, measured over 2 months, can be seen in Figure 9. Based in this conclusion, the notable contamination by Al in gluconate and phosphate solutions can be attributed to the affinity of these species with Al. They are able to selectively withdraw Al from the glass network when in contact with the glass containers (see Table 20).

Figure 10 shows the result of one year of storage of different amino acids, components of parenteral solutions in type I glass containers. Not all the solutions were contaminated by Al, but only those whose amino acids presented affinity with this element.

**TABLE 17 Arsenic Species Present as Contaminant in Commercial Parenteral Solutions**[a]

| Product | Total As (μg/L) | As(V) (μg/L ± RSD) | As(III) (μg/L ± RSD) |
|---|---|---|---|
| KCl 19.1% (10) | 41.3 | 41.3 ± 1.6 | n.d. |
| KCl 10% (4) | 23.6 | 23.6 ± 0.8 | n.d. |
| NaCl 20% (10) | 43.8 | 40.9 ± 5.7 | 2.9 ± 0.2 |
| NaCl 20% (4) | 15.9 | 15.9 ± 3.2 | n.d. |
| Sodium acetate 2 meq/mL (1) | 46.1 | 41.8 ± 2.0 | 4.3 ± 0.6 |
| Sodium phosphate 0.5 mol/L (10) | 36.7 | 36.7 ± 2.5 | n.d. |
| Sodium bicarbonate 8.4% (5) | 248.6 | 227.8 ± 0.5 | 20.8 ± 2.0 |
| Sodium bicarbonate 8.4% (9) | 198.3 | 147.2 ± 1.2 | 51.1 ± 0.8 |
| Calcium gluconate 10% (11) | 73.1 | 46.8 ± 4.1 | 26.3 ± 2.3 |
| Calcium gluconate 10% (9) | 92.7 | 72.8 ± 0.5 | 19.9 ± 1.9 |
| Calcium gluconate 10% (4) | 239.6 | 176 ± 7.2 | 63.6 ± 2.0 |
| Magnesium sulfate 50% (11) | 15.7 | 15.7 ± 0.2 | 5.7 ± 0.2 |
| Magnesium sulfate 50% (10) | 33.5 | 12.2 ± 1.2 | 21.3 ± 1.8 |
| Magnesium sulfate 50% (9) | 53.8 | 39.9 ± 5.6 | 13.9 ± 3.0 |
| Glucose 25% (6) | 21.8 | 16.5 ± 2.6 | 5.3 ± 0.6 |
| Glucose 25% (6) | 18.6 | 14.0 + 2.2 | 4.6 ± 0.7 |
| Vitamins (Tiaminose[b]) | 103.7 | 86.1 ± 0.9 | 17.6 ± 3.2 |
| Vitamins (Dextrovitase[c]) | 10.2 | 10.2 ± 0.2 | 3.8 ± 0.1 |
| Vitamins (Frutovena[d]) | 61.7 | 21.5 ± 1.7 | 40.2 ± 6.7 |
| Amino acids 10% (3) | 2.5 | 2.5 ± 0.1 | n.d. |
| Amino acids 10% (1) | 15.4 | 15.4 ± 2.3 | n.d. |
| Amino acids 10% (2) | 41.0 | 41.0 ± 0.9 | n.d. |
| Amino acids 8% (2) | 21.1 | 16.7 ± 1.2 | 4.4 ± 0.4 |
| Amino acids 8% (2) | 94.7 | 87.9 ± 5.7 | 6.8 ± 1.0 |
| Lipid emulsion 10% (2) | 0.9 | 0.9 ± 0.3 | n.d. |
| Lipid emulsion 20% (2) | 1.7 | 1.7 ± 0.6 | n.d. |
| Heparin 5000 UI/mL (7) | 56.7 | 17.2 ± 2.8 | 39.5 ± 2.3 |
| Heparin 5000 UI/mL (8) | 79.4 | 79.4 ± 4.9 | 23.3 ± 1.2 |

[a]Number in parentheses refer to product brands: (1) B. Braun, (2) Fresenius, (3) Baxter, (4) Halex Istar, (5) JP, (6) Merck, (7) EMS, (8) Elkins Sinn, (9)Ariston, (10) Geyer, (11) Hipolabor. RSD, relative standard deviation; n.d., not detected (below LOD).
[b]Glucose 3 g, ascorbic acid 0.25 g, thiamin chloridrate 0.015 g in 10 mL.
[c]Glucose 2 g, ascorbic acid 2 g, pyridoxine chloridrate 20 mg, nicotinamide 30 mg, riboflavin in 20 mL.
[d]Frutose 5 g, ascorbic acid 1 g, pyridoxine chloridrate 20 mg, sodium pantothenate 10 mg, riboflavin 4 mg in 20 mL.

*Source*: Form ref. 16.

Table 20 presents the Al level in parenteral formulations, collected by different authors in different countries. Not all of them mention glass packaging as responsible for Al contamination, but it seems to have been the major source. These data also confirm that the nature of the formulation components plays a key role in selective Al leaching.

These results show that significant variations in the contaminant and level of contamination depend on the formulation constituents, brand, and mostly the nature of the packaging material.

**TABLE 18 Metallic Contaminants Found in Commercial Solutions for Parenteral Nutrition**

| | Element Concentration (μg/L) | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Sample | Zn | Cr | Fe | Pb | As | Mn | Ba | Sn | Ge | Cu | Cd | Ref. |
| TPN bag | 9.1 | — | — | — | — | — | — | — | — | — | — | 13 |
| Amino acids | 60–4,70 | n.d. | — | — | — | — | — | — | — | — | — | 44 |
| L-Cysteine HCl | 32,000–86,000 | 110–230 | — | — | — | — | — | — | — | — | — | 44 |
| NaCl, KCl, Na acetate | 350–560 | 20–230 | — | — | — | — | — | — | — | — | — | 44 |
| Ca gluconate | 280–2,380 | — | — | — | — | — | — | — | — | — | — | 44 |
| Phosphate | 910–2,330 | 390–440 | — | — | — | — | — | — | — | — | — | 44 |
| Ca gluconate | 47–244 | — | 237–655 | — | — | — | — | — | — | — | — | 45 |
| TPN | 233–703 | — | 84 | — | — | — | — | — | — | — | — | 45 |
| Standard adult TPN formula, Baxter | — | 86.8 | — | 0.6 | 288.0 | 309.9 | 11.0 | 0.5 | 5.5 | — | — | 46 |
| Standard adult TPN formula, Abbott | — | 25.8 | — | 1.1 | 65.0 | 109.9 | 16.0 | 0.4 | 5.5 | — | — | 46 |
| Standard adult TPN renal formula, Abbott | — | 99.8 | — | 0.6 | 298.0 | 259.9 | 28.0 | 2.2 | 28.9 | — | — | 46 |
| Standard adult TPN formula | — | 139.8 | — | 0.4 | 65.9 | 299.9 | 22.0 | 1.4 | 15.9 | — | — | 46 |
| Standard adult TPN formula, Baxter | — | 15.8 | — | 0.7 | 61.0 | 47.9 | 73.0 | 0.6 | 9.2 | — | — | 46 |
| Standard adult amino acid solution, Fresenius | 131.2 | — | — | 4.99 | — | — | — | — | — | 2.96 | 0.35 | 47 |
| Standard adult amino acid solution, B. Braun | 88.92 | — | — | 16.80 | — | — | — | — | — | 6.66 | 4.37 | 47 |
| Standard adult amino acid solution, Baxter | 1.39 | — | — | 4.39 | — | — | — | — | — | 40.81 | 0.71 | 47 |

*Note*: n.d. = not detected, — = not measured.

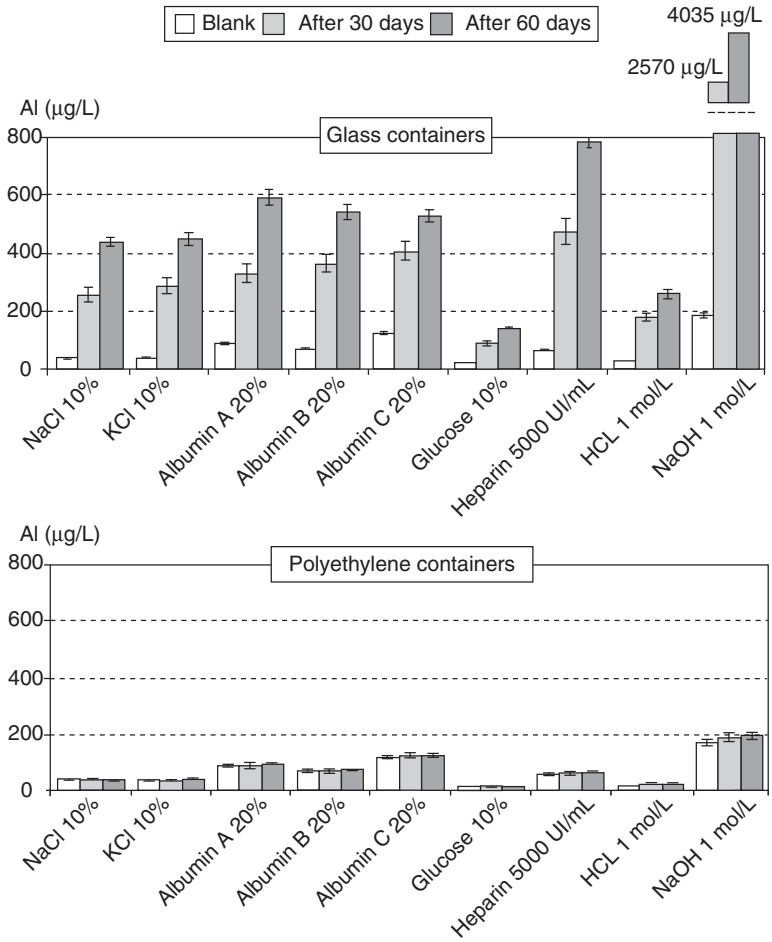


**FIGURE 8** Aluminum extracted from type II glass containers and from polyethylene containers by action of NaCl, KCl, albumin, glucose, heparin, HCl, and NaOH solutions after 30 and 60 days storage at room temperature. The three different albumins are: A, bovine (Merck); B, bovine (Reagen), and C, egg (Sigma) [50].

***Plastic Containers*** Plastics are organic and polymeric in nature. A polymer is a large molecule built up by the repetition of small and simple chemical units. The repeated unit of the polymer is usually equivalent or nearly equivalent to the monomer or the starting material from which the polymer is formed. The structural units of the polymers most used to manufacture plastic containers, along with their uses for pharmaceutical purposes, are shown in Table 21.

Modern polymer technology is founded on catalysis, and catalytic methods are extensively used in the production of plastics. Catalysts, since they only catalyse reactions, do not count as polymer constituents but may be present as impurities in the polymeric material. Table 22 lists the usual catalysts used for the polymerization of the polymers mentioned above, which can be found as contaminants in formulations stored in plastic containers.

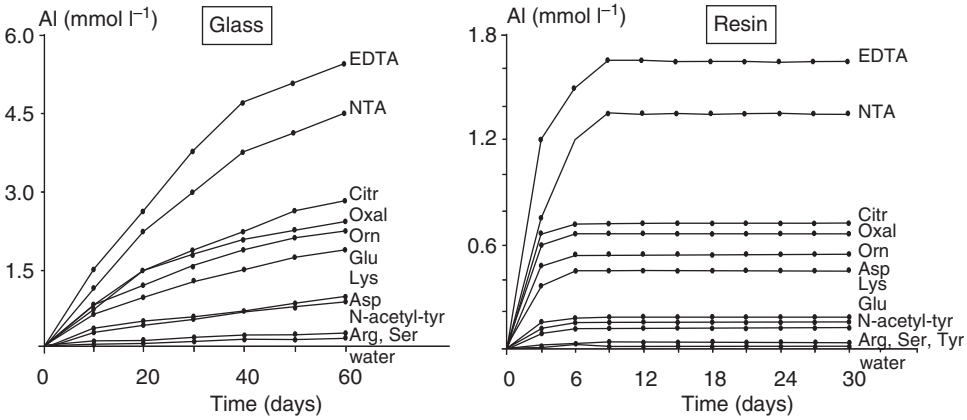


**FIGURE 9** Aluminum extracted from glass particles (18 mesh) and Al-form exchanger (18 mesh) as function of time by action of some amino acids and complexing agents. Concentration of ligands: 0.05 mol/L [51].

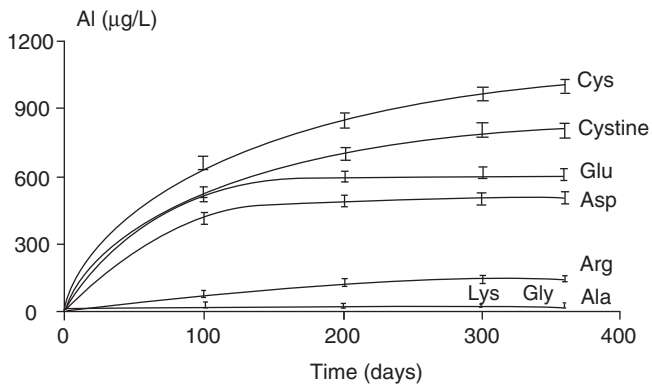


**FIGURE 10** Aluminum leached from glass containers by amino acids as a function of time at room temperature. Amino acid concentration: 0.028 mol/L [53].

Other contaminants that can originate from plastic containers are the additives necessary to turn the raw polymer into adequate containers. While PE may be used without any additive, the other plastics are virtually useless alone but are converted into highly serviceable products by combining them with other substances or materials. The additives most commonly found in plastics used for pharmaceutical products are antioxidants, heat stabilizers, lubricants, plasticizers, fillers, and colorants. These additives can be in liquid, solid, or fine particle forms and are used in amounts varying from less than 1% to more than 50% of the plastic mass. The additives necessary for each of the selected types of polymers are described in Table 23.

Additives authorized for use in plastics for pharmaceuticals, along with their limits in the polymer mass (according to BP and Ph. Eur.), are summarized in Table 24.

Since additives, together with the polymers, provide a large variety of substances, and the leachability of such components cannot be predicted a priori,

**TABLE 19 Aluminum Present as Contaminant in Commercial Parenteral Solutions, and Al Present in Container Materials**

| Product | Al Solution ± SD (μg/L) | Container | Al Container (%) |
|---|---|---|---|
| NaCl 20% | 149 ± 10 | Glass ampoule | 1.43 |
| | 13 ± 4 | Polyethylene | 0.04 |
| KCl 10% | 68 ± 6 | Glass ampoule | 1.25 |
| | 23 ± 5 | Polyethylene | 0.06 |
| Magnesium sulfate 50% | 560 ± 85 | Glass ampoule | 1.25 |
| | 380 ± 288 | Glass ampoule | 1.43 |
| Sodium acetate 2 meq/mL | 45 ± 7 | Glass ampoule | 2.14 |
| | 17 ± 8 | Polyethylene | 0.05 |
| Potassium phosphate 2 meq/mL | 988 ± 76 | Glass ampoule | 1.98 |
| | 1325 ± 142 | Glass ampoule | 2.45 |
| Sodium phosphate 0.5 *M* | 933 ± 88 | Glass ampoule | 1.65 |
| | 879 ± 203 | Glass ampoule | 2.05 |
| Calcium gluconate 10% | 5621 ± 1165 | Glass ampoule | 1.51 |
| | 5960 ± 62 | Glass ampoule | 2.21 |
| Sodium bicarbonate 8.4% | 833 ± 141 | Glass bottle | 0.99 |
| | 922 ± 102 | Glass bottle | 1.03 |
| Oligoelements[a] | 1129 ± 33 | Glass ampoule | 2.14 |
| Oligoelements[b] | 1854 ± 744 | Glass ampoule | 2.21 |
| Amino acids 10% | 164 ± 6 | Glass bottle | 0.82 |
| | | Rubber closure | 3.91 |
| Amino acids 10% | 116 ± 30 | Glass bottle | 0.76 |
| | | Rubber closure | 4.23 |
| Amino acids 10% | 93 ± 23 | Glass bottle | 0.84 |
| | | Rubber closure | 5.34 |
| Amino acids 10% | 65 ± 13 | Glass bottle | 0.89 |
| | | Rubber closure | 5.70 |
| Amino acids 10% | 23 ± 8 | Plastic bag | 0.01 |
| Glucose 50% | 13 ± 1 | Polyethylene | 0.04 |
| | 293 ± 14 | Glass ampoule | 1.15 |
| Glucose 25% | 9 ± 3 | Polyethylene | 0.08 |
| | 370 ± 23 | Glass ampoule | 1.87 |
| Albumin 20% | 644 ± 58 | Glass flask | 0.67 |
| | | Rubber closure | 4.06 |
| | 149 ± 23 | Glass flask | 0.66 |
| | | Rubber closure | 3.99 |
| Heparin 5000 UI/mL | 732 ± 23 | Glass ampoule | 2.88 |
| | 738 ± 54 | Glass ampoule | 3.03 |

*Note*: All solutions were within their guaranteed period of shelf life.
*Source*: From ref. 52.
*Suppliers*: Abbott, Ariston, Aster, Baxter, Behring, B.Braun, Darrow, Fresenius, Fujisawa, Gayer, Halex Istar, Hypofarma, Santisa, Roche, Zenalb.

[a]Composition: 22.0 mg $ZnSO_4$, 6.3 mg $CuSO_4$, 2.46 mg $MnSO_4$, 102.5 mcg $CrCl_3$ per ampoule.
[b]Composition: 8.8 mg $ZnSO_4$, 1.60 mg $CuSO_4$, 123.04 mcg $MnSO_4$ 20.50 mcg $CrCl_3$ per ampoule.

**TABLE 20 Aluminium Present in Commercial Solutions for Parenteral Administration**

| Product | Specification | Brand | Al concentration µg/L | Reference |
|---|---|---|---|---|
| *Electrolytes* | | | | |
| NaCl | 10% | Abbott | 3 | 10 |
| | 10% | Commercial Polfa | 14 | 54 |
| | 5.85% | Braun | 1 | 11 |
| | 1 M | Kabi | 22 | 11 |
| | 10% | Abbott | 43 ± 8* | 15 |
| | 10% | Elkins Sinn | 78 ± 7* | 15 |
| KCl | 15.0% | Commercial Polfa | 18 | 54 |
| | — | Abbott | 4 (5–11) | 10 |
| | — | Elkins Sinn | 3 | 10 |
| | 7.45% | Braun | <0.6 | 11 |
| | 7.45% | Kabi | 12 | 11 |
| | 7.45% | Braun | 33 ± 5* | 15 |
| | 10% | Abbott | 97 ± 8* | 15 |
| | 10% | Aster | 23 ± 5* | 15 |
| Ca chloride | — | Elkins Sinn | 15 (12–19) | 10 |
| | — | Abbott | 5 | 10 |
| | 0.5 M | Boehringer | 27 | 11 |
| | 1 N | Kabi | 224 | 11 |
| Na acetate | 2 meq/ml | Braun | 17 ± 8* | 15 |
| K acetate | — | Abbott | <5 | 10 |
| Mg sulphate | 50% | Abbott | 5 | 10 |
| | 50% | Braun | 606 | 11 |
| | — | Geyer | 560 ± 85* | 15 |
| | — | Ariston | 380 ± 288* | 15 |
| Na phosphate | — | Invenex | 2236 (2026–2370) | 10 |
| | 0.5 M | Abbott | 933 ± 88* | 15 |
| | 0.5 M | Geyer | 430 ± 177* | 15 |
| K phosphate | — | Invenex | 92 | 10 |
| | — | Abbott | 2189 (2069–2301) | 10 |
| | — | Braun | 188 | 11 |
| | — | Kabi | 2826 | 11 |
| | 2 meq/l | Fresenius | 1021 ± 188* | 15 |
| | 2 meq/l | Braun | 988 ± 76* | 15 |
| | 2 meq/l | Aster | 332 ± 27* | 15 |
| Ca gluconate | 10% | Elkins Sinn | 3973 (1095–5565) | 10 |
| | 10% | Lyphomed | 2245 (2000–2586) | 10 |
| | 10% | Braun | 4734 | 11 |
| | 10% | Fresenius | 6549 | 11 |
| | 10% | Pharma Hameln | 4421 | 11 |
| | 10% | Commercial Polfa | 4400 | 54 |
| | 10% | Elkins Sinn | 3987 ± 993* | 15 |
| | 10% | Braun | 4530 ± 1072* | 15 |
| | 10% | Ariston | 5960 ± 62* | 15 |
| | 10% | Halex Istar | 6781 ± 1842* | 15 |

**TABLE 20** ***Continued***

| Product | Specification | Brand | Al concentration µg/L | Reference |
|---|---|---|---|---|
| *Trace elements (TE)* | | | | |
| Zinc chloride | — | Abbott | 99 (81–123) | 10 |
| TE | Tracitrans | Fresenius | 994 | 11 |
| Pediatric TE | Ped-el | Pharmacia | 3000 | 54 |
| Pediatric TE | Peditrace | Pharmacia | 130 | 54 |
| Pediatric TE | Ped-element | Darrow | 1423 ± 68* | 15 |
| TE | Ad-element | Darrow | 3574 ± 237* | 15 |
| TE | Tracitrans | Fresenius | 5712 ± 988* | 15 |
| *Amino acids* | | | | |
| | Freamine 8.5% | McGaw | 12 (5–24) | 10 |
| | Travasol 10% | Travenol | 7 (6–8) | 10 |
| | HepatAmine 8% | McGaw | 22 | 10 |
| | Aminoplasmal 10% | Braun | 55 | 11 |
| | Aminopaed 10% | Kabi | 38 | 11 |
| | Aminosteril 8% | Fresenius | 17 | 11 |
| | Primene 10% | Clintec | 120 | 54 |
| | Aminomel 12.5% | Clintec | 121 | 54 |
| | Vaminolact 6.5% | Pharmacia | 30 | 54 |
| | Aminoplasmal L10 | Braun | 160 ± 48* | 15 |
| | Pediamino 10% | Braun | 116 ± 30* | 15 |
| | Aminoped 10% | Fresenius | 195 ± 27* | 15 |
| | Nefroamino AEH | Braun | 272 ± 66* | 15 |
| | Primène 10% | Baxter | 65 ± 13* | 15 |
| *Carbohydrates* | | | | |
| | Glucosteril 70% | Fresenius | 9 | 11 |
| | Glucose 40% | Fresenius | 20 | 11 |
| | Glucose 50% | Schiwa/Hormonchemie | 18 | 11 |
| | Glucose 20% | Schiwa/Hormonchemie | 3 | 11 |
| | Glucose 10% | Schiwa/Hormonchemie | <0.6 | 11 |
| | Dextrose (10, 20, 50%) | Abbott | <5 | 10 |
| | Dextrose (10, 50%) | McGaw | <5 | 10 |
| | Dextrose (5, 10, 50%) | Travenol | <5 | 10 |
| | Glucose 20% | Commercial Polfa | 16 | 54 |
| | Glucose 25% | Darrow | 9 ± 3* | 15 |
| | | Equiplex | 9 ± 2* | 15 |
| | Glucose 50% | Fresenius | 15 ± 3* | 15 |
| | | Darrow | 17 ± 3* | 15 |
| | | J.P. Ind. Farm | 8 ± 2* | 15 |
| | | B. Braum Lab. | 13 ± 1* | 15 |
| | | Ariston | 11 ± 4* | 15 |

**TABLE 20** ***Continued***

| Product | Specification | Brand | Al concentration µg/L | Reference |
|---|---|---|---|---|
| *Lipids* | | | | |
| | Intralipid 10% | Kabi Vitrum | <5 | 10 |
| | Intralipid 20% | Kabi Vitrum | <5 | 10 |
| | Intralipid 10% | Pfrimmer Kabi | 5 | 11 |
| | Intralipid 20% | Pfrimmer Kabi | 7 | 11 |
| | Lipofundin 20% | Braun | 35 | 11 |
| | Lipofundin 20% | Braun | 14 | 11 |
| | Intralipid 20% | Pharmacia | 30 | 54 |
| | Lipofundin 20% | Braun | 180 | 54 |
| | Lipofundin 10% | Braun | 29 ± 6* | 15 |
| | Lipofundin 20% | Braun | 34 ± 7* | 15 |
| *Vitamins* | | | | |
| | Vitamin C 500 mg | EMS | 3443 ± 233* | 15 |
| | Vitamin B12 1mg | Bunker | 92 ± 23* | 15 |
| | B complex | Hipolabor | 1089 ± 127* | 15 |
| | B complex | Ariston | 709 ± 65* | 15 |
| | Multivitamin MVI | ICN | 588 ± 63* | 15 |
| *Others* | | | | |
| | Albumin 20% | — | 190.4 | 55 |
| | Albumin 20% | Behring | 235 ± 12* | 15 |
| | Heparin 1000 U/mL | — | 211.7 | 55 |
| | Heparin 5000 U/mL | Fujisawa Inc. | 732 ± 23* | 15 |
| | Heparin 5000 U/mL | Cristália | 72 ± 6* | 15 |

—: not informed.
*mean value ± standard deviation of three samples of the same lot.

pharmacopeial compendia have determined procedures to investigate biological and physical–chemical properties of plastics. In general, tests do not include the measurement of the polymeric material or the additives themselves but the biological reactivity (toxicity screening) of their extracts. Extracts are generally obtained by autoclaving the plastic container filled with water at 121°C for 30 min or at 100°C for 2 h. Only containers for parenteral and ophthalmic preparations have a more controlled regulation since their risk of toxicity is more significant, and therefore assays for plastic additives are foreseen in some pharmacopeial compendia [5].

Regulatory guidelines for plastics for pharmaceutical containers also set limits for impurities other than additives. These are inorganic species (metallic and non-metallic) that can also be present in plastic containers, and their determination serves as a quality criterion for the plastic material. They may or may not be related

**TABLE 21 Polymeric Materials, Their Structural Units, and Uses in Pharmaceutical Products**

| Polymer | Monomers | Uses in Biotechnological Products |
|---|---|---|
| Polyethylene (PE) | Ethylene | Dry dosage forms<br>Noninjectable aqueous solutions<br>Intravenous aqueous infusions |
| Polypropylene (PP) | Propylene | Dry dosage forms<br>Noninjectable aqueous solutions<br>Intravenous aqueous infusions |
| Polyvinyl chloride (PVC) | Vinyl chloride | Dry dosage forms<br>Noninjectable aqueous solutions<br>Intravenous aqueous infusions<br>Blood and blood components<br>Tubing for blood |
| Polyethylene vinyl acetate (EVA) | Ethylene and vinyl acetate | Intravenous aqueous infusions<br>Tubing for parenteral nutrition preparations |
| Polyethylene terephthalate (PET) | Terephthalic acid or dimethyl terephthalate and ethylene glycol | Dry oral dosage forms<br>Liquid oral dosage forms |

*Source*: Data from refs. 5 (BP) and 56 (FDA).

**TABLE 22 Catalysts Used for Polymerization Reaction of Plastics Materials Listed in Table 21**

| Polymer | Catalysts |
|---|---|
| Polyethylene (PE) | $TiCl_4 + Al(C_2H_5)_3$; $CrO_3/SiO_2$, $MoO_3/Al_2O_3$ |
| Polypropylene (PP) | $\alpha$-$TiCl_3 + Al(C_2H_5)_3$ |
| Polyvinyl chloride (PVC) | — |
| Polyethylene vinyl acetate (EVA) | — |
| Polyethylene terephthalate (PET) | — |

*Source*: From ref. 57.

**TABLE 23 Overview on Additives Commonly Used in Plastic Manufacturing**

| | Additives | | | | | |
|---|---|---|---|---|---|---|
| Polymer | Antioxidant | Heat Stabilizer | Lubricant | Plasticizer | Filler | Colorant |
| PE | × | — | × | — | × | × |
| PP | × | — | × | — | × | — |
| PVC | — | × | × | × | × | × |
| EVA | × | — | × | — | × | × |
| PET | — | — | × | — | × | × |

*Source*: From ref. 58.

**TABLE 24 Acceptable Additives Allowed in Plastics for Pharmaceuticals According to the BP and Ph. Eur. and Their Limits**

| Additive Number | Additive Name | Polymer | Limit (%) |
|---|---|---|---|
| | *Plasticizer* | | |
| 1 | Di(2-ethylhexyl)phthalate | PVC | 40 |
| 2 | Zinc octanoate | PVC | 1 |
| 3 | *N′N′*-diacylethylenediamines | PVC | 1 |
| 4 | Epoxidized soya oil | PVC | 10 |
| 5 | Epoxidized linseed oil | PVC | 10 |
| | *Antioxidants* | | |
| 7 | Butylhydroxytoluene | PE, PP, EVA | 0.125 |
| 8 | Ethylene bis[3,3-bis[3-(1,1-dimethylethyl) -4-hydroxyphenyl]butanoate] | PE, PP | 0.3 |
| 9 | Pentaerythrityl tetrakis[3-(3,5-di-*tert*-butyl-4-hydroxyphenyl)propionate] | PE, PP<br>EVA | 0.3<br>0.2 |
| 10 | 2,2′,2″,6,6′,6′-hexa-*tert*-butyl-4,4′,4″ [(2,4,6-trimethyl-1,3,5-benzenetriyl) -trismethylene]triphenol | PE, PP<br>EVA | 0.3<br>0.2 |
| 11 | Octadecyl 3-(3,5-di-*tert*-butyl-4 -hydroxyphenyl)propionate | PE, PP<br>EVA | 0.3<br>0.2 |
| 12 | tri(2,4-di-*tert*-Butylphenyl) phosphite | PE, PP<br>EVA | 0.3<br>0.2 |
| 13 | 1,3,5-tris(3,5-di-*tert*-Butyl-4-hydroxybenzyl) -*s*-triazine-2,4,6(1*H*,3*H*,5*H*)-trione | PE, PP | 0.3 |
| 14 | 2,2′-bis(octadecyloxy)-5,5′-spirobi [1,3,2-dioxaphosphinane] | PE, PP | 0.3 |
| 15 | Dioctadecyl disulfide | PE, PP | 0.3 |
| 16 | Didodecyl 3,3′-thiodipropionate | PE, PP | 0.3 |
| 17 | Dioctadecyl 3,3′-thiodipropionate | PE, PP | 0.3 |
| 18[a] | 2,4-bis(1,1-Dimethylethyl)phenyl biphenyl-4,4′-diyldiphosphonite<br>2,4-bis(1,1-Dimethylethyl)phenyl biphenyl-3,4′-diyldiphosphonite<br>2,4-bis(1,1-Dimethylethyl)phenyl biphenyl-3,3′-diyldiphosphonite<br>2,4-bis(1,1-Dimethylethyl)phenyl biphenyl-4-ylphosphonite<br>2,4-bis(1,1-Dimethylethyl)phenyl phosphite<br>2,4-bis(1,1-Dimethylethyl)phenyl 4′-[bis[2,4-bis (1,1-Dimethylethyl)phenoxy]phosphanyl] biphenyl-4-ylphosphonate<br>2,4-bis(1,1-Dimethylethyl)phenol | PE, PP | 0.1 |
| | *Lubricants and Fillers* | | |
| 19 | Stearic acid | | |
| 20 | Oleamide | EVA | 0.5 |
| 21 | Erucamide | EVA | 0.5 |
| 22[a] | Copolymer of dimethyl butanedioate and 1-(2-hydroxyethyl)-2,2,6, 6-tetramethylpiperidin-4-ol | PE, PP | 0.3 |

**TABLE 24** ***Continued***

| Additive Number | Additive Name | Polymer | Limit (%) |
|---|---|---|---|
| 23 | Hydrotalcite | PE, PP | 0.5 |
| 24 | Alkanamides | PE, PP | 0.5 |
| 25 | Alkenamides | PE, PP | 0.5 |
| 26 | Sodium silico-aluminate | PE, PP | 0.5 |
| 27 | Silica | PE, PP | 0.5 |
| 28 | Sodium benzoate | PE, PP | 0.5 |
| 29 | Fatty acid esters | PE, PP | 0.5 |
| 30 | Trisodium phosphate | PE, PP | 0.5 |
| 31 | Liquid paraffin | PE, PP | 0.5 |
| 32 | Zinc oxide | PE, PP | 0.5 |
| 33 | Magnesium oxide | PE, PP | 0.5 |
| 34 | Calcium stearate | PE, PP | 0.5 |
| 35 | Zinc stearate | PE, PP | 0.5 |
| 36 | Talc | PP | 0.5 |
| | *Colorants* | | |
| 37 | Titanium dioxide | PE, PP | 4 |
| 38 | Ultramarine blue | PVC | n.i. |

*Note*: PE and PP for parenteral and ophthalmic preparations may contain at most three oxidants; n.i.: not informed.

[a]Only for nonparenteral preparations.

to any of the additives listed in Table 24. For example, the investigation of aluminum in PE and polypropylene (PP) is related to the use of aluminum-based catalysts for the attainment of the polymers; barium as a contaminant in a plastic container is indicative of the presence of a barium soap stabilizer, possibly used in plastic compounding. The tests, including chemicals such as sulfated ash residue and heavy metals, are listed in Table 25 with the respective limits for each species in the different plastic materials.

While not mandatory from regulatory guidelines, much research has been carried out to investigate the extractability of plastic additives in contact with a variety of pharmaceutical formulations, mainly those for parenteral use. The research concentrates on the extractability of plasticizer phthalates, mainly di-2-ethylhexylphthalate (DEHP) from polyvinyl chloride (PVC) into the blood, blood components, and infusion solutions. The purpose for these studies lies in its, up to now, controversial hazardous effects on humans. The amount of additive necessary to turn rigid PVC into a flexible material (40% m/m) and the absence of chemical bonds between the polymer and the plasticizer make it a potentially extractable species.

Limits for DEHP in formulations stored in PVC bags are not set in pharmacopeial compendia. The BP prescribes a standard test whose results should show that the amount of DEHP extracted by action of ether on PVC bags does not exceed 40% of the polymer mass. This test is valid for PVC for all uses: Dry dosage forms, noninjectable aqueous solutions, intravenous aqueous infusions, blood and blood components, and tubing used in sets for blood and blood components. On the other

**TABLE 25 Inorganic Impurities and Their Limits in Extracts of Plastic Containers for Pharmaceuticals**

| Impurity Number | Impurity | Polymer | Use | Limit |
|---|---|---|---|---|
| 1 | Heavy metals | PE, PP | Injectable, ophthalmic | 2.5 ppm |
| | | PVC | Aqueous parenteral, blood, tubing | 50 ppm |
| 2 | Sulfated ash | PE, PP | Injectable, ophthalmic | 0.2%, 1% |
| | | EVA | Parenteral container and tubing | 1.2% |
| | | PET | Oral forms | 0.5% |
| 3 | Aluminum | PE, PP | Injectable, ophthalmic | 1 ppm |
| | | PET | Oral forms | 1 ppm |
| 4 | Chromium | PE, PP | Injectable, ophthalmic | 0.05 ppm |
| 5 | Titanium | PE, PP | Injectable, ophthalmic | 1 ppm |
| | | PET | Oral forms | 1 ppm |
| 6 | Vanadium | PE, PP | injectable, ophthalmic | 0.1 ppm |
| | Zinc | PE, PP | Injectable, ophthalmic | 1 ppm |
| | | PVC | Aqueous parenteral, blood | 0.2% |
| | | PET | Oral forms | 1 ppm |
| 7 | Zirconium | PE | Injectable, ophthalmic | 0.1 ppm |
| 8 | Barium | PVC | Aqueous parenteral, blood, tubing | 5 ppm |
| | | PET | Oral forms | 1 ppm |
| 9 | Cadmium | PVC | Aqueous parenteral blood, tubing | 0.6 ppm |
| 10 | Calcium | PVC | Aqueous parenteral, blood | 0.07% |
| 11 | Tin | PVC | Aqueous parenteral, blood, tubing | 20 ppm |
| 12 | Ammonium | PVC | Blood | 2 ppm |
| 13 | Chloride | PVC | Blood | 0.4 ppm |
| 14 | Antimonium | PET | Oral forms | 1 ppm |
| 15 | Cobalt | PET | Oral forms | 1 ppm |
| 16 | Germanium | PET | Oral forms | 1 ppm |
| 17 | Manganese | PET | Oral forms | 1 ppm |

*Source*: Adapted from Ph. Eur [ref. 7].

hand, the Food and Drug Administration (FDA) launched a nonregulatory publication entitled "Safety Assessment of Di-2-ethylhexylphthalate (DEHP) Released from PVC Medical Devices" [59], where concerns about its toxicity are discussed in detail. This publication intends to provide risk managers with the information necessary to make decisions about the safety of DEHP exposure from medical devices. Conclusions of the study were reached by calculating the dose of DEHP received by patients undergoing different medical procedures. For the pharmaceutical industry, what is truly relevant are concerns about the presence of DEHP in intravenous (IV) infusion fluids and drugs and in parenteral nutrition because of their storage in plastic bags and the delivery of these preparations through PVC tubing sets. The study concluded that there is little to no risk posed by patient exposure to the amount of DEHP released from PVC IV bags following the infusion of crystalloid fluids (e.g., normal saline, Ringer's lactate). There is a small risk, however,

**TABLE 26 Concentration of DEHP in Infusion Solutions stored in Plastic containers**

| Sample | Volume (mL) | Time of storage | DEHP (μg/L) | Ref. |
|---|---|---|---|---|
| 0.9% saline | 1000 | 5 months | 7 | 60 |
| | | 14 months | 24 | |
| 0.9% saline | 500 | 0.5 month | 13 | 60 |
| | | 3 months | <4 | |
| 0.9% saline | 100 | 6 months | 8 | 60 |
| 0.9% saline | 1000 | — | 8 | 60 |
| 5% glucose | 1000 | 5 months | 4 | 60 |
| | | 12 months | 4 | |
| 5% glucose | 500 | 0.5 month | 34 | 60 |
| | | 3 months | 7 | |
| ACO bag | 100 | 12 months | 7 | 61 |
| Travenol bag | 100 | 12 months | 5 | 61 |
| Multivitamins | | 48 h (45 °C) | 21 | 62 |
| TPN—3.85% lipid | 2200– | 24 h | 330 | 63 |
| TPN—3.85% lipid | 2200– | 1 week (4 °C) | 450 | 63 |
| TPN—2.50% lipid | 2200– | 24 h | 230 | 63 |
| TPN—2.50% lipid | 2200– | 1 week (4 °C) | 260 | 63 |
| TPN—1.85% lipid | 650 | 24 h | 300 | 63 |
| TPN—1.85% lipid | 650– | 1 week (4 °C) | 360 | 63 |
| TPN—1.00% lipid | 800– | 24 h | 240 | 63 |
| TPN—1.00% lipid | 800– | 1 week (4 °C) | 270 | 63 |
| 5% glucose–5% paclitaxel | | 1 h | 6,600 | 64 |
| 5% glucose–5% paclitaxel | | 2 h | 18,500 | 64 |
| 5% glucose–10% paclitaxel | | 1 h | 29,400 | 64 |
| 5% glucose–10% paclitaxel | | 2 h | 56,600 | 64 |

posed by the exposure to the amount of DEHP released from PVC bags used to store and administer drugs that require a pharmaceutical vehicle for solubilization. The dose of DEHP received by adult patients receiving TPN is estimated to be less than the tolerable intake (TI),[3] suggesting that there is little need for concern about DEHP-mediated effects in these patients. The dose of DEHP received by neonates undergoing TPN is uncertain. Depending on the data used to derive the TI/dose ratio, neonates may be at an increased risk of DEHP-mediated adverse effects.

In fact, studies have shown that, depending on the constituents in the formulation, the extractable DEHP can be much higher than the expected. Our knowledge of the nature of the polymer and its additives, as well as of standard migration, is not great enough to predict container–formulation interactions. Migration is a unique phenomenon, and predictions based only on the structure and physical or chemical properties may not be effective for solving interaction problems. Table 26 shows the level of DEHP found in different kinds of parenteral formulations stored in PVC bags. It is clear that the amount of DEHP leached into the formulation depends on the nature of its constituents. While in saline solutions the amount did not exceed

[3]The tolerable intake values for DEHP are outlined in the International Organization for Standardization ISO/DIS 10993-17 standard, "Method for the Establishment of Allowable Limits for Leachable Substances."

24 μg/L in 14 months, in oleaginous vehicles, it reached over 300 μg/L in 2 h, which corroborates with FDA studies. Due to the lipophilic character of DEHP, its migration process from the packaging into hydrophilic preparations is limited to a random dissolution on the contacting surface. On the other hand, in lipophilic preparations such as lipid emulsions, DEHP is encouraged to leave the polymer surface through a dissolution process into the lipophilic milieu.

There are drug products whose interaction with PVC bags and infusion sets are so high that they must include labeling precautions for use with PVC containers. These drugs include antineoplastics such as paclitaxel, docetaxel, tacrolimus, and teniposide, and others such as ciprofloxacin, cefoperazone sodium, fluconazole, metronidazole HCl, cimetidine, and propofol [64, 65].

Although differences in the methods of the studies shown in Table 26 (time of contact, volume of solution, temperature, etc.) do not allow for direct comparisons, the data presented suggest that the leachability of DEHP is highly dependent on the nature of the pharmaceutical formulation. The question of migration of DEHP into solutions is relevant not only to lipophilic vehicles but to drugs such as paclitaxel and etoposide, which are able to extract much more DEHP than simple lipophilic lipidic emulsions do.

Attempts to substitute PVC containers have shown, however, that non-PVC containers may also contain plasticizer components resulting in the release of some DEHP into the preparation. A study conducted by Sautou-Miranda and co-workers showed that [65] leaches rapidly from coextruded and triple-layered IV tubing into etoposide infusion solution.

Although most studies focus on DEHP, it is not the only additive that can be leached from plastic materials. Scalia et al. [66] investigated the migration of antioxidants from polyolefinic plastics into oleaginous vehicles. The study concentrated on two antioxidants, Irganox 1010 and Irgafos 168, used in PE and PP polymers.[4] The study was carried out by dipping plastic sheets containing 0.15% of antioxidants into a mixture of five oils (caprylic, capric triglyceride, cyclomethicone, dicaprylyl ether, isohexadecane, and $C_{12–15}$ alkyl benzoate) commonly used as pharmaceutical vehicles, and storing the oil mixture in bottles manufactured with the polymers. The results showed that the amount of antioxidant leached into the oleaginous mixture varied remarkably in relation to the polymeric material, and decreases in the order EP > RACO > PP > high-density polyethylene (HDPE) (RACO and EP are ethylene–propylene copolymers). Table 27 shows the percentage of antioxidants left in the polyolefin bottles filled with the mixture and stored at 25°C for 1 year. The authors concluded that the migration of both antioxidants is related to the polymer crystallinity and structure: The higher the crystallinity, the lower the additive release. They suggested that PP and HDPE are satisfactory for the manufacture of oil.

***Elastomeric Closures*** Closures for pharmaceutical products are generally made of polymeric materials, which may be of either a synthetic or natural origin. While brittle closures such as screw caps are made of conventional thermoplastics with a single composition, elastomeric closures are made of complex mixtures of many

[4]Pentaerythrityl tetrakis(3,5-di-*tert*-butyl-4-hydroxyphenyl)propionate (Irganox 1010) and tris(2,4-di-*tert*-butylphenyl)phosphite (Irgafos 168), additives 9 and 12 in Table 25.

**TABLE 27 Percentage of Antioxidants Left in Polyolefin Bottles Filled with Mixture and Stored at 25 °C for 1 year**

| | Antioxidant Left (%) | |
|---|---|---|
| Polyolefin Bottles | Irganox 1010 | Irgafos 168 |
| HDPE | 92.7 ± 5.7 | 47.8 ± 5.2 |
| PP | 76.3 ± 7.1 | 47.0 ± 4.1 |
| RACO | 57.9 ± 6.3 | 36.4 ± 4.7 |
| EP | 0 | 0 |

*Source*: From ref. 66.
EP: ethylene-propulene amorphous copolymer blend; RACO: ethylene-*co*-propylene random copolymer.

ingredients. Because elastomers or rubbers can be molded into an almost limitless variety of permanent shapes and are easily penetrated by needles and reseal after needle withdrawal, they are primary closures for parenteral vials and for preparations intended for repeated use.

The polymeric materials usually used to manufacture rigid closures are practically the same as those seen under plastic containers (Section 3.2). The same impurities are therefore to be expected in these packaging components. On the other hand, though made of polymeric materials, elastomeric closures present a different structure. In the manufacture of rubber, elastomer, the chief component, is combined with other chemicals to produce a material with specific properties that meet target needs, such as its above-mentioned ability to reseal on repeated use. Table 28 lists the common elastomers used in the pharmaceutical industry and their monomeric structures.

The substances listed in Table 28 correspond to the basic structure of elastomeric closures. The other components in rubber formulations are curing or vulcanizing agents, accelerators, activators, antidegradants, plasticizers, fillers, and pigments. The most common additives used to compound rubber for the pharmaceutical industry are listed in Table 29. The amount of each component may vary from rubber to rubber, and, depending on the component, the amount can reach more than 50% of the total mass of a formulation. While accelerators are used in amounts of around 1%, fillers may make up more than 50% of the formulation mass.

Table 30 shows four typical pharmaceutical rubber formulations based on natural, halobutyl, ethylenepropylenediene (EPDM), and silicone elastomers.

The presence of such a variety of components makes elastomeric closures a potential source of contaminants. Polymeric materials present a reasonable inertness, however, the possibility of the other components reaching the drug formulation should be considered. Moreover, these ingredients, though intended for pharmaceutical purposes and therefore meeting pharmacopeial requirements, may still present their own impurities as contaminants. Carbon black is usually an impure material, and it can contain polynuclear aromatic hydrocarbons [68], which can be extracted into the packaged drug. Clays contain metallic impurities that can also be extracted into the drug formulation or even react with a drug constituent.

Pharmacopeial monographs do not set limits for additives, as they do for plastic polymers, and rubber closure (BP, Appendix XIX, USP ⟨381⟩) tests are limited to

**TABLE 28 Common Elastomers Used in the Pharmaceutical Industry**

| Common Name | Chemical Name | Structure |
|---|---|---|
| Butyl rubber | Poly(isobutylene-isoprene) | $\left[ -(CH_2-C(CH_3)_2)_{50}-CH_2-C(CH_3)=CH-CH_2- \right]_n$ |
| Halobutyl rubber | Halogenated poly(isobutylene-isoprene) | $\left[ -(CH_2-C(CH_3)_2)_{65}-CH-C(CH_3)-CH(X)-CH_2- \right]_n$ X = Cl or Br |
| Ethylene–propylene rubber | Poly(ethylene-propylene) | $\left[ -(CH_2-CH_2)_3-CH_2-CH(CH_3)- \right]_n$ |

| | | |
|---|---|---|
| Ethylene–propylene–diene rubber | Poly(ethylene-propylene-diene) | $\left[ -(CH_2-CH_2)_{15}(CH_2-\underset{\displaystyle CH_3}{\underset{\vert}{CH}})_5\text{diene}- \right]_n$ |
| Silicone rubber | Polydimethylsiloxane | $\left[ -\overset{\displaystyle CH_3}{\underset{\displaystyle CH_3}{\overset{\vert}{\underset{\vert}{Si}}}}-O \right]_n$ |
| Urethane rubber | Adipic acid–ethylene glycol polyester | $HO-(CH_2)_2\left[ O-\overset{\displaystyle O}{\overset{\Vert}{C}}-(CH_2)_4-\overset{\displaystyle O}{\overset{\Vert}{C}}-O-(CH_2)_2 \right]_n OH$ |
| Fluoroelastomers | Poly(tetrafluoroethylene) | $\left[ -\overset{\displaystyle F}{\underset{\displaystyle F}{\overset{\vert}{\underset{\vert}{C}}}}-\overset{\displaystyle F}{\underset{\displaystyle F}{\overset{\vert}{\underset{\vert}{C}}}}- \right]_n$ |
| Natural rubber | *cis*-1,4-Polyisoprene | $\left[ -CH_2-\underset{\displaystyle CH_3}{\underset{\vert}{C}}=CH-CH_2- \right]_n$ |

**TABLE 28** ***Continued***

| Common Name | Chemical Name | Structure |
|---|---|---|
| Polyisoprene rubber | *cis*-1,4-Polyisoprene | $\left[ -CH_2-\underset{\displaystyle CH_3}{\underset{\vert}{C}}=CH-CH_2- \right]_n$ |
| Neoprene rubber | Polychloroprene | $\left[ -CH_2-\underset{\displaystyle Cl}{\underset{\vert}{C}}=CH-CH_2- \right]_n$ |
| Styrene butadiene rubber | Poly(butadiene-styrene) | $\left[ -(CH_2-CH=CH-CH_2)_4-CH_2-\underset{\displaystyle C_6H_5}{\underset{\vert}{CH}}- \right]_n$ |
| Nitrile rubber | Poly(butadiene-acrylonitrile) | $\left[ -(CH_2-CH=CH-CH_2)_5-(CH_2-\underset{\displaystyle CN}{\underset{\vert}{CH}})_2- \right]_n$ |
| Polybutadiene | Polybutadiene | $\left[ -CH_2-CH=CH-CH_2- \right]_n$ |

*Source*: From ref. 67.

**TABLE 29 Ingredients Other Than Elastomers Used to Compound Rubbers for Pharmaceutical Industry**

| Curing Agents | Accelerators | Activators | Antidegradants | Plasticizers | Fillers | Pigments |
|---|---|---|---|---|---|---|
| Sulfur | Hexamethylene tetramine | Zinc Oxide | Amines | Paraffinic waxes | Carbon Black | Carbon black |
| Sulfur-containing chemicals | Ditiocarbamates | Stearic acid | Hindered phenols | Silicon oil | Alumnum silicates (clays) | Titanium dioxide |
| Peroxides | Sulfonamides | | Waxes | Paraffinic oils | Magnesium silicate (talc) | Iron oxide |
| Cadmium oxide | Thiuram | | | Naphthenic oils | Barium sulfate | Chromium oxide |
| Magnesium oxides | Thiazole | | | Organic phosphates | Zinc oxide | Organic dyes |
| Zinc oxide | | | | | Silica | |

*Source*: From ref. 67.

**TABLE 30 Composition of Four Typical Rubbers Used as Closures for Pharmaceutical Formulations Based on Natural, Halobutyl, EPDM, and Silicone Elastomers**

| | Rubber Type | | | |
|---|---|---|---|---|
| Ingredient | Red Rubber | Gray Rubber | Gray Rubber | Black Rubber |
| Elastomer | Natural rubber | Halobutyl rubber | Dimethylpolysiloxane polymer | EPDM |
| Filler | Alumnum silicate | Alumnum silicate | Silica | Carbon black |
| Plasticizer | Paraffinic oil | Naphthenic oil | — | Naphthenic oil |
| Pigment | Iron oxide | Titanium dioxide<br>Carbon black | Carbon black | — |
| Activator | Zinc oxide<br>Stearic acid | — | — | Zinc oxide<br>Stearic acid |
| Accelerator | Thiuram<br>Thiazole | Thiuram | — | Thiuram<br>Zinc dithiocarbamate |
| Antidegradant | Butylated hydroxytoluene | Butylated hydroxytoluene | — | — |
| Curing agent | Sulfur | Zinc oxide | 2,4-Dichlorobenzoyl peroxide | Sulfur |

*Source*: From ref. 67.

showing the content of sulfur, sulfated ash, volatile sulfides, and extractable zinc, ammonium, and heavy metals.

Extractability tests prescribed by other regulatory agencies [FDA, Parenteral Drug Association (PDA)] for closures for drug packaging [69, 70] are also limited to the amount of extractable residues or tests to evaluate the in vivo reaction of the extractable residue when the material fails in the in vitro tests.

The tests, however, are neither qualitative (in the sense of showing which substances can be extracted) nor specifically quantitative since they are conceived to show only the total amount of extractable as a residue. USP ⟨381⟩ "Elastomeric Closures for Injection," for example, recommends the calculation of the weight of the residue after evaporating the solvent (purified water, drug vehicle, or isopropyl alcohol) used for extraction. Tests in vivo are recommended only when the material does not meet the requirements of the in vitro tests.

These tests, however, are unable to show further contamination in the final product since the extractability also depends on the interaction between the container and the formulation constituents. In spite of possible interactions, very little is known about the leachability of closure constituents through the direct contact of closures with formulations.

Mennermaa and colleagues [71] determined the composition of three different types of rubber stoppers used to seal parenteral solutions. After immersing the stoppers in 0.9% NaCl solution and autoclaving at 121°C for 15 min, the analysis of the aqueous extract was carried out by proton-induced X-ray emission. Table 31 presents the elemental composition of the rubber stoppers investigated in micrograms/gram dry weight. The concentration of zinc varied from sample to sample by a factor of 40,000 (5–20,000 ppm). Titanium, Fe, and Br were present in all stoppers, and even Pb (2 ppm) was found in one extract.

**TABLE 31 Elemental Composition of Rubber Stoppers Based on Element Concentration in Aqueous Extract**

| | Amount (μg/g) | | |
|---|---|---|---|
| Element | Sample 1, Old Formulation of Bromolutyl Siliconized Rubber | Sample 2, New Formulation of Bromolutyl Siliconized Rubber | Sample 3, New Formulation of Chlorolutyl Siliconized Rubber |
| Ti | 500 | 1,600 | 200 |
| Fe | 7,500 | 8,000 | 4,500 |
| Cu | 10 | 10 | — |
| Zn | 50 | 5 | 20,000 |
| Br | 100 | 7,500 | 50 |
| Pb | 2 | — | — |

*Source*: From ref. 71.

**TABLE 32 Nature of Compounds Identified by GC-MS in Headspace and Soxhlet Extracts of Two Rubbers under Study**

| | Compound Identified in Extracts Obtained with | |
|---|---|---|
| | Headspace | Soxhlet |
| Alkanes | $C_5$ and higher | $C_9$–$C_{30}$ |
| Oligomers | $T_{max}$ (elution) = 235°C | $T_{max}$ (elution) = 280°C |
| Aromatics | 92 < MW < 132 | 168 < MW < 182 |
| Fatty acids | No | 228 < MW < 352 |
| Esters | No | Yes |
| 2,6-di-*tert*-Butyl-*p*-cresol | Yes | Yes |
| 2,6-di-*tert*-Butyl-*p*-benzoquinone | Yes | Yes |
| Ketone | MW = 198 | MW = 198 |

*Source*: From ref. 72.
*Note*: MW, molecular weight.

Schoenmakers et al. [72] analyzed two representative commercial rubbers by gas chromatography–mass spectrometry (GC–MS) and detected more than 100 different compounds. The rubbers, mixtures of isobutylene and isoprene, were analyzed after being cryogenically grinded and submitted to two different extraction procedures: a Sohxlet extraction with a series of solvents and a static-headspace extraction, which entailed placing the sample in a 20-mL sealed vial in an oven at 110°C for 5, 20, or 50 min. Although these are not the conditions to which pharmaceutical products are submitted, the results may give an idea of which compounds could be expected from these materials. Residual monomers, isobutylene in the dimeric or tetrameric form, and compounds derived from the scission of the polymeric chain were found in the extracts. Table 32 presents an overview of the nature of the compounds identified in the headspace and Soxhlet extracts of the polymers. While the liquid-phase extraction was able to extract less volatile compounds, the headspace technique was able to show the presence of compounds with low molecular mass

**TABLE 33 Accumulation of Extractable from Synthetic Polyisoprene Rubber after Autoclaving for 1 h**

| Compound | Concentration in Solution (mg/L) |
|---|---|
| Aniline | 1.64 |
| Diphenylguanidine | 11.76 |
| Dedenzylamine | 2.12 |
| Triisopropanolamine | 3.73 |

*Source*: From ref. 73.

in the extract. A complete screening of the extracted substances can be seen in Figure 11, where the graphics depict the abundance of the different classes of compounds detected in the extracts of different solvents for both rubbers. Not only oligomers, but also phthalates, phenols, and acidic compounds, were found in the extracts, mainly from rubber P1.

Jenke [73] studied the extractability of aniline, diphenylguanidine, dedenzylamine, and triisopropanolamine from a synthetic polyisoprene rubber similar to the material used in pharmaceutical applications. Rubber samples were autoclaved (121°C) in contact with water or NaCl 0.9% solution for 1 h. Table 33 presents the concentration of each compound in solution after the extraction procedure using 2 g rubber material. Extraction profiles ranged between 1.64 and 3.73 mg/L, with the exception of diphenylguanidine, whose extraction yield reached 11.76 mg/L.

### 3.3 Delivery Systems

Pharmaceutical formulations also have direct contact with plastic materials through delivery sets, transfer tubing, and devices, as well as during the phases of their production, via gaskets, filters, and transportation.

Even though there is no specific pharmacopeial prescription for testing these materials, they should present the same profile stipulated for container and closures. In the case of tubing lines, which are also made of elastomeric materials, an extractability test for elastomeric closures is recommended, mainly for those made of PVC.

It is more difficult to generalize about the leachability of plastic tubing materials than it is for plastic containers or closures themselves since it depends not only on the nature of the plastic material but also on the flow rate, temperature, solvent, and time of contact with the tubing.

Studies conducted by different authors on the release of chemical substances from medical devices, mainly those used for infusing solutions, show that these are potential sources of contamination for pharmaceutical formulations. One of the most studied is diethylhexyl phthalate, the same plasticizer found in PVC infusion bags to give flexibility. The same concerns about the use of PVC bags for the storage of lipids or lipophilic formulations are valid for tubing.

Table 34 presents the amount of DEHP leached from typical PVC infusion lines from a delivery set of 2.25 m kept at 27°C. A sample volume from 8 to 140 mL was perfused in 24 h through the lines and collected for analysis. It was possible to see that while amino acids did not promote the migration of DEHP from PVC tubing,

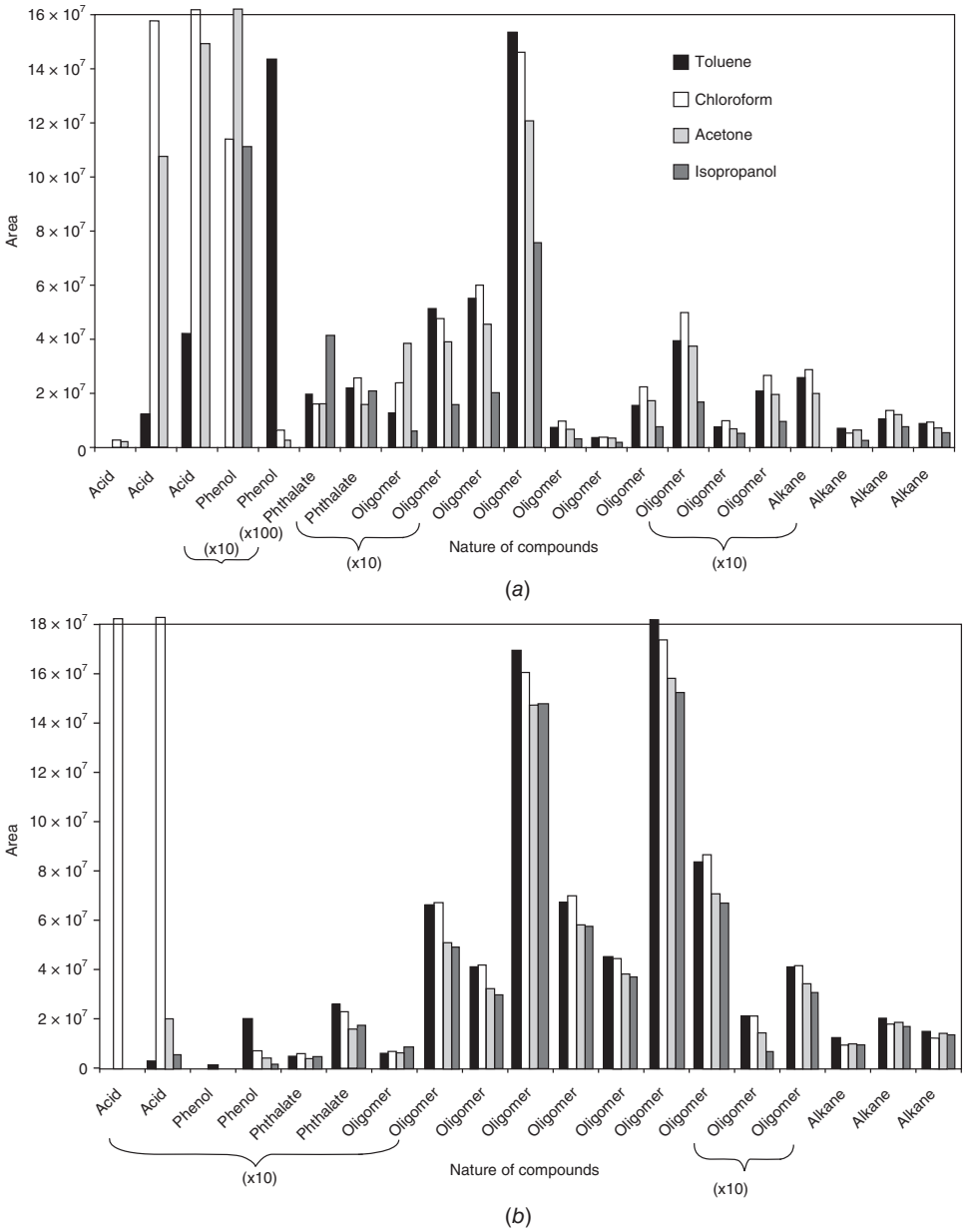


**FIGURE 11** Influence of the solvent used for the Soxhlet extractions of two different commercial rubber closures (P1) and (P2) on the area of the chromatographic peaks of different classes of compounds detected in GC–MS [72].

the lipophilic lipid emulsion and propofol promoted the extraction of large amounts of DEHP into the preparations.

An investigation of DEHP extractability from PVC tubing into a large variety of pharmaceutical formulations used for different purposes was carried out by

**TABLE 34 DEHP Loaded in Infusion Solutions after Exposure to PVC Infusion Lines**

| Sample | Volume (mL) | Number of Samples | DEHP (μg/mL) | Range DEHP (μg/mL) |
|---|---|---|---|---|
| Amino acid–glucose | 140 | 06 | 0.31 ± 0.56 | 0.0–1.05 |
| Lipid emulsion | 24 | 10 | 422.78 ± 47.39 | 329.15–490.40 |
| Midazolam | 24 | 03 | 0.90 ± 0.38 | 0.55–1.30 |
| Propofol infusion | 10 | 10 | 654.87 ± 96.49 | 423.85–736.10 |
| Fentanyl | 28.8 | 10 | 3.63 ± 0.92 | 1.95–5.05 |
| Imipenem | 8 | 03 | 0 | –0.10–0.05 |

*Source*: From ref. 74.

Haishima and colleagues [75]. The authors divided the IV drugs into five groups according to the properties of their active ingredients and additives. Group 1 included drugs that were practically insoluble or insoluble in water and contained additives such as surfactants, oils, glycerine, ethanol, or benzyl alcohol. Group 2 included drugs equally insoluble in water but soluble in acidic or basic solutions. Formulations slightly soluble in water were placed in group 3, and formulations very soluble or freely soluble in water were classified in groups 4 and 5. The difference between group 4 and 5 was the presence of additives suspected to induce DEHP migration in group 4. Tests were carried out by placing the drug formulations into PVC tubing of 10 cm in length with 2.13 mm inner diameter for 1 h at room temperature under shaking. As shown in Table 35, products included in group 1 promoted the release of large amounts of DEHP, with the exception of insulin and dinoprost, probably due to the absence of oleaginous excipients. No significant DEHP release was observed by most of the other drugs assigned to groups 2 through 5, with the exception of human serum albumin and antithrombin III (both group 4) and phenytoin solution (group 2), probably due to the presence of propylene glycol and ethanol in the formulation. These results confirm the affinity of DEHP to lipophilic media.

Due to their utilization modus, many studies conducted with tubing materials are essentially kinetic, where tubing length, time of contact, flow rate, and temperature are important parameters.

Kambia et al. [76] studied the kinetics of DEHP migration into TPN solutions from PVC tubing. They quantified the amount of DEHP leached into two types of emulsions 24 h after preparation and storage at 4 °C. The authors concluded that the extraction depended on the lipid content of the formulations and the flow rates. Figure 12 shows the DEHP concentration time course obtained from bags and outlet tubing during 10–11 h simulated infusion of TPN after 24-h storage at 4°C.

To overcome the problem of DEHP leaching into parenteral infusions containing lipophilic components, a triple-layered tubing material has been used. This type of tubing is made of a PVC outer layer, while its innermost layer, which comes into contact with the drug solution, is made of inert PE. In spite of this arrangement, it has been shown that, depending on the preparation, DEHP from the PVC outer layer may be released in the infusion solution.

Sautou-Miranda et al. [77] showed that DEHP is rapidly leached from PVC, coextruded, and also triple-layered IV tubing by action of an etoposide infusion

**TABLE 35 DEHP Released from PVC Tubing into Intravenous Drug Preparations after 1 h Contact at Room Temperature**

| Active Ingredient | Concentration | Additives | DEHP (μg/L) | SD (μg/L) |
|---|---|---|---|---|
| | | *Group 1* | | |
| Cyclosporin | 500 μg/mL | Polyoxythene castor oil, ethanol | 27,363.9 | 384.8 |
| Tacrolymus hydrate | 10 μg/mL | Absolute ethanol, HCO-60 | 4,091.9 | 31.9 |
| Propofol | 10 mg/mL | Soybean oil, concentrated glycerin, egg yolk lecithin, edetate | 19,451.2 | 852.5 |
| Flurbiprofen axetil | 10 mg/mL | Soybean oil, egg yolk lecithin, concentrated glycerin | 17,838.5 | 821.6 |
| Vitamins—fat soluble | Whole amounts of Sorbita was mixed with PN-Twin no. 2 (2.2 L) | Sodium citrate, sodium pyrosulfite, sodium thioglycollate, HCO-60, benzyl alcohol, polysorbate 80 | 1,157.1 | 5.1 |
| Menatetrenone | 5 mg/mL | Aminoethylsulfonic acid, sesame oil, soybean lecithin, D-sorbitol, concentrated glycerin | 8,457.5 | 62.9 |
| Insulin human | 40 units/mL | Concentrated glycerin, *m*-cresol | 281.6 | 6.0 |
| Dinoprost | 2 mg/L | — | 185.8 | 17.3 |
| Miconazole | 1 mg/L | HCO-60 | 30,098.3 | 423.3 |
| Diazepam | 5 mg/L | Propylene glycol, ethanol, benzyl alcohol, sodium benzoate, benzoic acid | 2,008.8 | 257.6 |
| Prednisolone sodium succinate | 10 mg/L | Sodium carbonate, sodium hydrogen phosphate, sodium dihydrogen phosphate | 915.6 | 182.3 |

**TABLE 35** ***Continued***

| Active Ingredient | Concentration | Additives | DEHP (μg/L) | SD (μg/L) |
|---|---|---|---|---|
| Prednisolone sodium succinate | 1 mg/L | Sodium carbonate, sodium hydrogen phosphate, sodium dihydrogen phosphate | 407.1 | 2.4 |
| | | *Group 2* | | |
| Famotidine | 20 mg/L | L-Aspartic acid, D-manitol | 166.0 | 0.9 |
| Droperidol | 2.5 mg/mL | *p*-Oxymethyl benzoate, *p*-oxypropyl benzoate | 171.0 | 0.6 |
| Droperidol | 50 μg/mL | *p*-Oxymethyl benzoate, *p*-oxypropyl benzoate | 167.4 | 24.6 |
| Sivelestat sodium hydrate | 1 mg/L | D-Mannitol, sodium hydroxide, propylene glycol, ethanol | 885.7 | 10.6 |
| Phenytoin | 50 mg/L | | | |
| Methotrexate | 0.2 mg/mL | Sodium chloride, sodium hydroxide | 372.8 | 6.8 |
| Haloperidol | 5 mg/mL | Glucose, lactic acid, sodium hydroxide | 50.6 | 2.5 |
| Epinephrine | 0.25 mg/mL | Chorobutanol, sodium hydrogen sulfite, hydrochloric acid, sodium chloride | 290.3 | 24.6 |
| | | *Group 3* | | |
| Methilergometrine maleate | 0.2 mg/mL | — | 462.7 | 4.2 |
| Vecuronium bromide | 2 mg/mL | D-Mannitol | 192.7 | 1.5 |
| Panipenem betamipron | 5 mg/mL | — | 237.0 | 1.2 |
| Mynocycline hydrochloride | 1 mg/mL | — | 150.0 | 8.9 |
| Nicardipine hydrochloride | 0.1 mg/mL | D-Sorbitol | 211.6 | 24.0 |
| Bromhexine hydrochloride | 2 mg/mL | Glucose | 174.9 | 23.7 |
| Ceftazidime | 10 mg/mL | Sodium carbonate | 301.0 | 0.5 |
| Fluconazole | 1 mg/mL | — | 210.5 | 0.2 |
| Aspoxicillin | 50 mg/mL | Sodium chloride | 296.7 | 2.6 |
| Carbazochrome sodium sulfonate | 0.05 mg/mL | Sodium hydrogen sulfite, D-sorbitol, propylene glycol | 246.1 | 3.0 |

| | | | | |
|---|---|---|---|---|
| | | *Group 4* | | |
| Oxytocin | 0.01 units/mL | Chlorobutanol | 423.1 | 0.8 |
| Hydroxyzine hydrochloride | 0.05 mg/mL | Benzyl alcohol | 430.8 | 144.4 |
| Ranitidine hydrochloride | 0.1 mg/mL | Phenol | 197.9 | 29.5 |
| Human immunoglobulin G | 50 mg/mL | D-Sorbitol | 243.9 | 14.3 |
| Panthenol | 250 mg/mL | Benzyl alcohol | 412.1 | 18.2 |
| Human serum albumin | 250 mg/mL | Sodium *N*-acetyl tryptophan, sodium caprylate, sodium hydrogen carbonate | 10,080.8 | 84.1 |
| Human antithrombin III | 25 units/mL | Sodium chloride, sodium citrate, D-mannitol | 2,008.2 | 21.8 |
| Nitroglycerin | 0.5 mg/mL | D-Mannitol | 267.6 | 8.9 |
| Sulpyrine | 2.5 mg/mL | Benzyl alcohol | 302.8 | 3.8 |
| Erythromycin lactobionate | 2.5 mg/mL | Benzyl alcohol | 92.2 | 0.7 |
| Clindamycin phosphate | 3 mg/mL | Benzyl alcohol | 274.9 | 4.0 |
| | | *Group 5* | | |
| Imipenem cilastatin sodium | 5 mg/mL | Sodium hydrogen carbonate | 205.1 | 1.6 |
| 5% glucose | 50 mg/mL | — | 284.6 | 4.8 |
| Ferric oxide, saccharated | 0.4 mg/mL | — | 244.5 | 5.5 |
| Maltose, sodium chloride, magnesium chloride, potassium dihydrogenphosphate, sodium acetate | — | — | 262.8 | 5.0 |
| Atropine sulfate | 0.5 mg/mL | — | 200.7 | 5.1 |
| Ampicillin sodium | 10 mg/mL | — | 262.3 | 6.8 |
| Aminophyline | 0.5 mg/mL | Ethylenediamine | 301.1 | 4.0 |
| Fosfomycin sodium | 20 mg/mL | Glucose solution | 289.6 | 6.7 |
| Calcium gluconate | 85 mg/mL | — | 179.4 | 4.3 |
| Cefazolun sodium hydrate | 10 mg/mL | — | 215.1 | 0.9 |
| Amino acids, electrolytes | — | Sodium hydrogen sulfite | 328.5 | 5.0 |
| Suxamethonium chloride | 2 mg/mL | — | 228.6 | 2.1 |
| Ioversol | 320 mg/mL | — | 404.0 | 79.5 |
| *l*-Isoprenaline hydrochloride | 1 μg/mL | Sodium hydrogen sulfite, L-cysteine hydrochloride | 326.3 | 8.6 |

*Source*: From ref. 75.

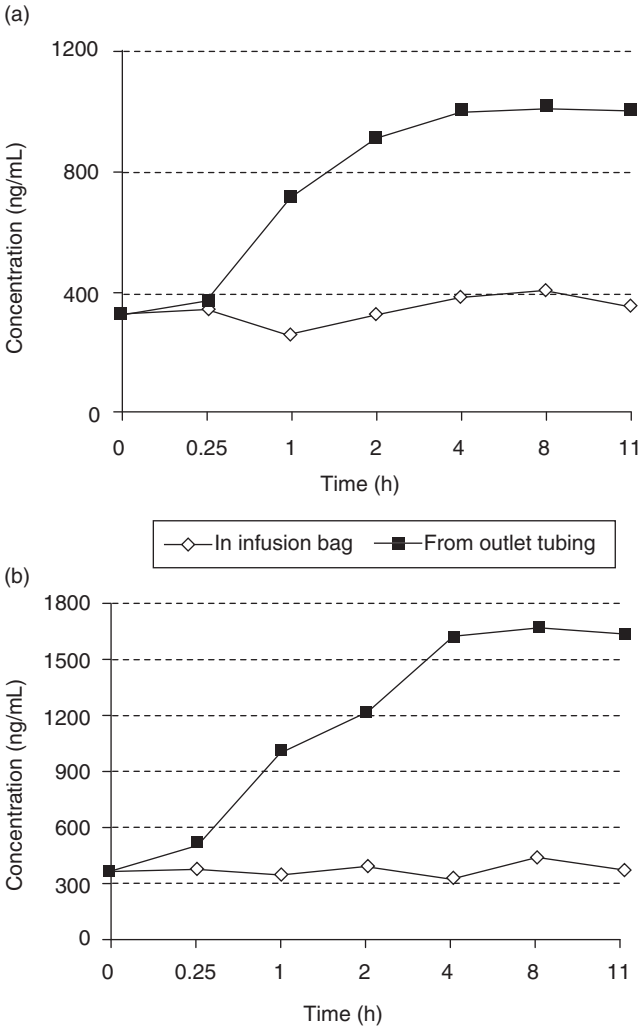


**FIGURE 12** Comparative kinetics of DEHP leachability during simulated infusion of TPN [76]. (a) Kinetics of DEHP leachability during simulated infusion of TPN 24 hours after reconstitution of the preparation (n = 2 bag). Formula 1: infusion volume of 2200 mL (flow rate 177 mL/h, lipid concentration 3.85%). (b) Kinetics of DEHP leachability during simulated infusion of TPN 24 hours after reconstitution of the preparation (n = 2 bag). Formula 2: infusion volume of 650 mL (flow rate 46 mL/h, lipid concentration 1.85%).

solution. Table 36 presents the concentration of DEHP leached from tubing into etoposide infusion solutions as a function of time of exposure for two tubing lengths. The results showed that fast and considerable leaching of DEHP occurred, even when the PVC component had no direct contact with the solution. DEHP leaching was greatest for the tubing made only of PVC, but it also occurred with the coextruded (PE + PVC) and triple-layered tubes, despite their claims to prevent such leaching. The authors concluded that either DEHP was present in the internal PE layer or it migrated rapidly through the other polymer layers.

The kinetics of DEHP extraction from triple-layered tubing was studied by these authors by varying tube length, flow rate, and drug concentration. Figure 13 shows

**TABLE 36 Concentration of DEHP Leached from Tubing into Etoposide Infusion Solutions as a Function of the Time of Exposure**

| Tubing Length and Type | Mean ± SD DEHP Concentration (μg/mL) | | | | | |
|---|---|---|---|---|---|---|
| | 0 h | 1 h | 2 h | 3 h | 4 h | 6 h |
| 25 cm | | | | | | |
| PE[b] | <0.5 | <0.5 | <0.5 | <0.5 | <0.5 | <0.5 |
| PVC[b,c] | <0.5 | 22.59 ± 4.21 | 35.05 ± 2.51 | 49.88 ± 4.78 | 58.57 ± 1.03 | 73.51 ± 3.51 |
| Triple layered[d] | <0.5 | 18.92 ± 2.23 | 31.49 ± 3.78 | 44.19 ± 5.38 | 54.83 ± 4.28 | 61.99 ± 1.25 |
| PVC[e] | <0.5 | 19.93 ± 1.90 | 33.01 ± 1.87 | 46.82 ± 1.98 | 55.46 ± 3.15 | 66.10 ± 1.23 |
| Coextruded[f] | <0.5 | 18.64 ± 1.38 | 28.77 ± 2.33 | 39.37 ± 3.47 | 48.13 ± 2.21 | 54.60 ± 2.53 |
| 50 cm | | | | | | |
| PE[b] | <0.5 | <0.5 | <0.5 | <0.5 | <0.5 | <0.5 |
| PVC[b,c] | <0.5 | 51.67 ± 3.63 | 82.00 ± 2.65 | 112.69 ± 4.89 | 131.67 ± 3.65 | 155.22 ± 2.35 |
| Triple layered[d] | <0.5 | 39.85 ± 0.49 | 59.73 ± 3.04 | 81.51 ± 1.57 | 85.01 ± 2.17 | 98.72 ± 2.33 |
| PVC[e] | <0.5 | 45.38 ± 2.08 | 73.27 ± 0.96 | 94.65 ± 0.88 | 117.53 ± 3.43 | 143.41 ± 11.39 |
| Coextruded[f] | <0.5 | 39.98 ± 0.74 | 59.61 ± 1.49 | 78.03 ± 0.52 | 82.47 ± 3.56 | 107.83 ± 9.68 |

*Source*: From ref. 77.

[b]Manufactured by Vycon, Ecouan, France.
[c]Tubing length 80 cm.
[d]Consisted of an outer layer of PVC, a middle layer of EVA, and an inner layer of PE.
[e]Manufactured by Cair, Civrieux d'Azergues, France.
[f]PVC and PE tubing.
DEHP: diethyhexylphthalate; PE: polyethylene; PVC: polyvinyl chloride.

the amount of DEHP leached with the volume of solution infused. As could be expected, higher drug concentrations and tubing lengths increased the amount of DEHP released. Similarly, the slower the flow rate was, the higher the migration of the additive.

As mentioned under plastic containers (Section 3.2), DEHP is not the only additive that can be leached from plastic materials. Jenke et al. [78] studied the composition of extractables from plastic tubing used in pharmaceutical production facilities. Eight tubing materials made of silicone and neoprene rubber were characterized by their extractable substances. The authors investigated not only organic but also inorganic extractables. Extracts were obtained by static experiments using both water and ethanol. The tubing was cut and autoclaved in water at 121°C for 1 h, or lengths of tubing were filled with 100% ethanol and kept at 55°C for 24 h. Metal determination was carried out by inductively coupled plasma–atomic emission spectroscopy (ICP-AES), which included 29 elements. Besides the elements listed in Table 37, Be, Co, Cr, Cd, Se, V, Ge, Pb, and Bi were also measured but not detected since they either presented a concentration not different from the blank sample or below the lowest quantity determinable (LQD). It is important to mention that the ICP-AES is not a very sensitive technique. LQD values for elements ranging between 0.01 and 0.1 mg/L are relatively high levels for contaminants such as Cd or Be. This means that these elements could possibly be present in the extracts but not detected using this technique. In general, all tested tubing materials presented metals that were extracted either by water or by ethanol, although water, due to its greater capability to interact with metallic ions, promoted higher extraction yields. All tubing materials contained extractable Ca, Mg, Zn, and

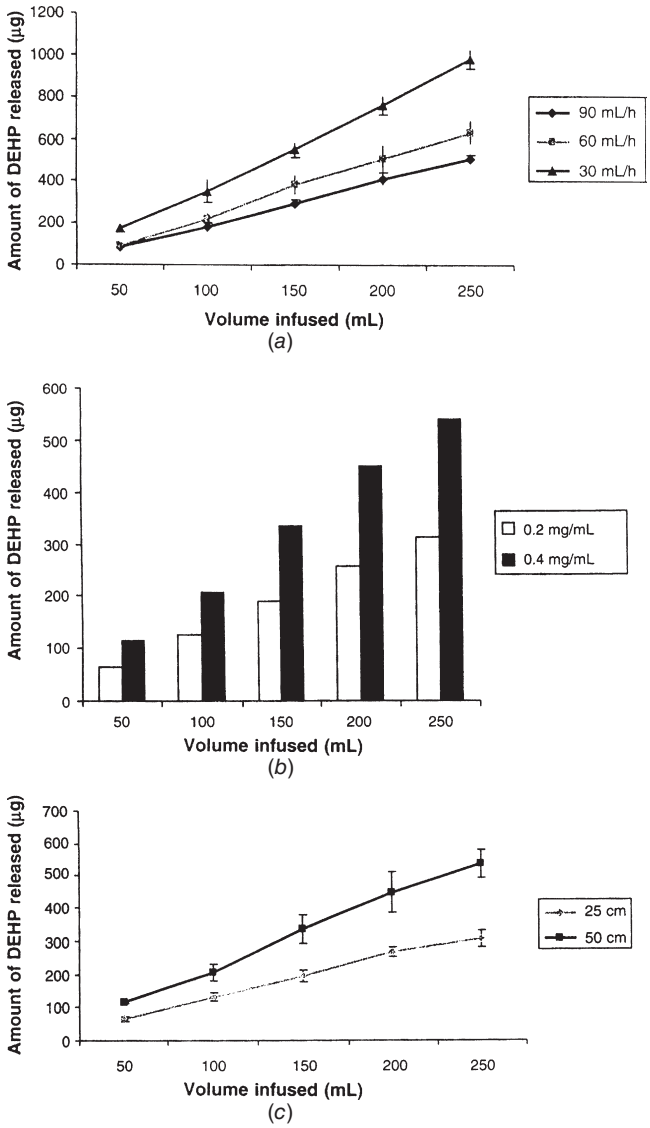


**FIGURE 13** (*a*) Cumulative amount of DEHP leached from 50- and 80-cm PVC-only tubing after infusion of 0.4-mg/mL etoposide solution using various flow rates. (*b*) Cumulative amount of DEHP leached from 50- and 80-cm triple-layered tubing after infusion of 0.2 and 0.4 mg/mL etoposide solution with a flow rate of 30 mL/h. (*c*) Amount of DEHP leached from 50- and 80-cm triple-layered tubing after infusion of 0.4-mg/mL etoposide solution with a flow rate of 30 mL/h [77].

B. Elevated concentrations of other metals were also extracted from samples 3 and 5, probably due to the tubing's reinforced embedded metallic wire.

The authors also measured silicon and both organic and inorganic carbon. While inorganic carbon is indicative of the presence of carbonates (earth carbonates as additives), organic carbon is related to organic extractables. As could be expected,

**TABLE 37 Levels of Total Carbon (Organic and Inorganic), and Silicone and Metals Extracted of the Tubing Materials with Water or Ethanol by Heating at 121°C for 1 h**

| Tubing Material | Total Inorganic Carbon (μg/g) | Total Organic Carbon (μg/g) | Si (μg/g) Ethanolic Extract | Si (μg/g) Aqueous Extract | Metal Concentration in Aqueous Extract | | |
|---|---|---|---|---|---|---|---|
| | | | | | <0.5 μg/mL | 0.5–1.0 μg/mL | >1 μg/mL |
| Silicone 1 | 0 | 14.0 | 765 | 101 | B, Mg, Zn | | |
| Silicone 2 | 0.9 | 38.1 | 1360 | <0.2 | Ca | B, Mg, Mn Mo, Ti, Zr. Sn, Zn, Sb, Li, Ag, Ni | |
| Silicone 3 (embedded wire) | 2.3 | 250 | 1860 | 66.0 | Ca, Ba, Mn, Mg, Al, Cu | — | B, Fe, Zn, |
| Silicone 4 | 0.2 | 34.0 | 1120 | 130 | Ca, B, Fe, Mg, Al, Zn | — | — |
| Silicone 5 (embedded wire) | 0 | 49.9 | 1300 | 87.3 | Ca, B, Mn, Fe, Mg, Zn, Sb | — | — |
| Silicone 6 | 0 | 46.9 | 739 | <0.2 | Mn, Fe, Mg, Zn, Cu, Sb, Ni | B | Ca |
| Santoprene 1 | 10.1 | 16.6 | — [a] | <0.2 | Ca, Ba, B, Mg, Zn, | — | — |
| Santoprene 2 | 4.6 | 13.5 | Not tested | 0 | Ca, Mo, Ti, Zr, Mn, Zn, Li, Ag, Ni | — | Mg |

*Source*: From ref. 78.

[a]Tube disintegration.

Santoprene extracts contained higher concentrations of inorganic carbon and little, if any, extractable silicon. On the other hand, the amount of silicone in extracts of silicone polymers was very high. From Table 37, it is possible to see that these materials, though all made of silicone, are very different each other. While about 100 μg/g Si was found in the aqueous extract from sample 1, no measurable Si was detected in the same extracts from samples 2 and 6. As could be expected, silicone extractables were much higher in ethanolic extracts, being that 739 μg/g Si from sample 6 was the lowest amount leached.

The analysis of aqueous and ethanolic extracts by techniques such as liquid chromatography–mass spectrometry (LC–MS) and CG–MS showed that all silicon materials contained essentially the same peaks, that is, the same compounds were extracted. The distribution, however, varied from material to material. The primary organic extractables from silicone tubing were homologous series of silicone oligomers with the structural formula $[(CH_3)_2SiO]_n$. Direct matches with an MS spectral library allowed the identification of oligomers with $n = 5$ to $n = 25$. However, while in sample 3, the peaks were at the lower range of $n$, the oligomer distribution for sample 2 was shifted toward higher molecular weights, which are less soluble and therefore less extractable. Figure 14 illustrates the differences in the ethanolic extracts of materials 2 and 3.

Results of Santoprene extracts are quite different from those of silicone. GC–MS chromatograms for water extracts of Santoprene tubing show peaks associated with C8 and C9 acids, phthalates, and other organic substances. Figure 15 shows the GC–MS chromatograms of static ethanol extracts of Santoprene tubing materials, and Table 38 lists the compounds identified in Figure 14, based on matches of MS spectrum peaks.

### 3.4 Particulate Matter

Particulate contamination can be classified as intrinsic or extrinsic depending on its origin. Intrinsic contaminants arise from the manufacturing, packaging, transport, and storage of solutions; extrinsic contaminants are introduced or generated during drug reconstitution and administration to the patient.

The presence of particulate matter, while undesirable in any pharmaceutical product, is truly a problem in intravenous and ophthalmic preparations. Most common particulates in intravenous preparations are glass fragments, from the opening of glass ampoules, particles from rubber stoppers and intravenous equipment, and particles from plastic syringes, providing that the manipulation of such solutions is carried out in a controlled clean area. Air in a controlled area, in the immediate proximity of exposed sterilized containers/closures and filling/closing operations is appropriate when it has a particle count of no more than 3520 in a size range of 0.5 μm per cubic meter. This corresponds to a Class 100 air cleanliness level and is the regular environmental condition for manipulating sterile preparations.

Pharmacopoeial compendia prescribe the examination of particulate contamination for injections and infusion solutions and consider particulate contamination as the presence of extraneous mobile undissolved particles, other than gas bubbles, unintentionally present in the solution. They limit the number of particles according to their size and the volume of the preparation.

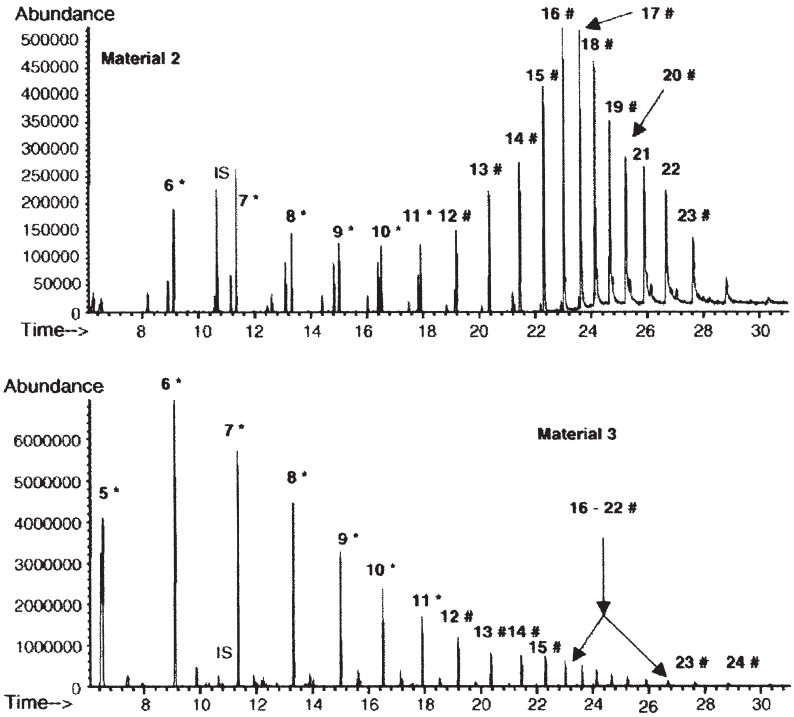


**FIGURE 14** GC–MS chromatograms of the static ethanol extracts of silicone tubing materials (underivatised). The chromatograms from all the silicone materials were similar (same peaks but different relative sizes), and thus chromatograms from two different silicones are shown. A majority of the peaks are attributable to silicone oligomers. Peaks associated with cyclic oligomers are identified by the number of repeating units, *n*: e.g., $[(CH_3)_2SiO]_n$. Peaks denoted with * produced exact compound matches versus a library of mass spectra, while peaks denoted by # produced a library match to the right compound class (cyclic oligomer) but wrong specific oligomer. Small peaks at 7.95, 10.32, and 11.93 min were linked to 5–7 member linear silicone oligomers. IS = internal standard (dimethyl phthalate) [78].

Generally, particulate matter is eliminated by filtration, but the word *elimination* or even the concept of *being substantially free of*, as appear in some compendia, are levels practically impossible to reach. In fact, complying with pharmacopeia prescription is limited to a removal of particles above a certain size, since the elimination of particles becomes increasingly more difficult the smaller the particle is. Elimination difficulty seemingly increases in an exponential fashion as the size decreases.

Pharmacopeial count limits for particulates in parenteral solutions is given in Table 39. The limit depends on the method used for the determination and also on the volume of the sample. Two different procedures for the determination are generally proposed: light obscuration particle count test (LO) and microscopic particle count test (M), since neither is applicable to all kinds of samples.

It is possible to display the relationship between size and number of particles graphically. Groves [79] reported a logarithmic relationship between the number of

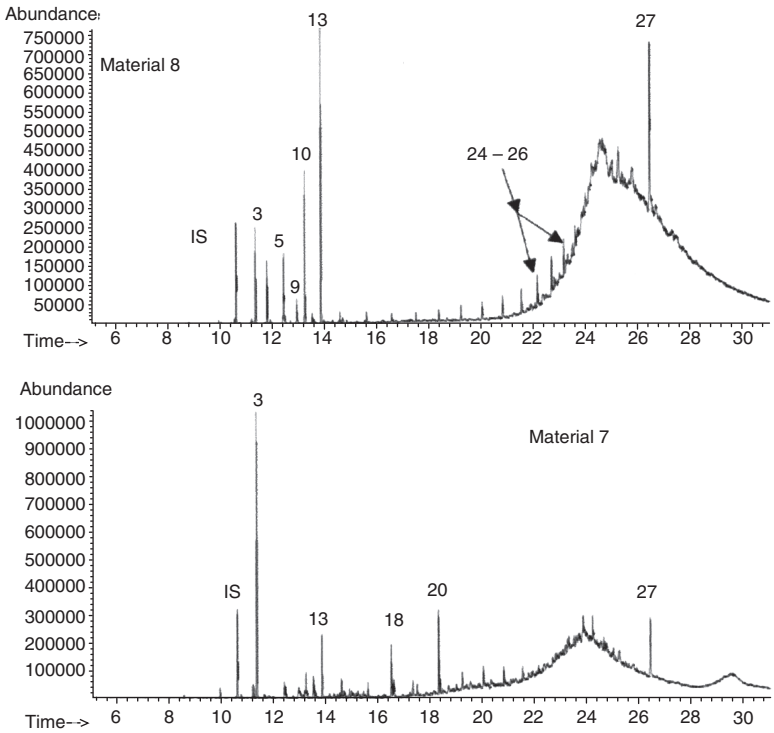


**FIGURE 15** GC–MS chromatograms of the static ethanol extracts of Santoprene tubing materials (underivatized). The chromatograms from these two Santoprene materials were quite different from those of the silicone materials (Figure 14). IS = internal standard (dimethyl phthalate). See Table 39 for the tentative peak identifications. Only those peaks with recorded spectral library matches are noted for each sample, although retention times and patterns may suggest some additional peak identifications [78].

particles and their diameter in a solution. Thus, using the limits prescribed by BP, the graphical representation would be as shown in Figure 16.

The author, however, considers that if the distribution number/size of particles follows a random pattern, graphically represented in Figure 16, it cannot be considered contamination. On the other hand, if the solution no longer contains random identities but rather is dominated by particles such as starch grains from a stopper composition, glass shards, from a delaminating container, skin fragments, or clothing fibers, a nonrandom behavior would occur and be defined as "contamination." In other words, only positive deviation of the log-log particle size distribution curve, constructed using, for example, pharmacopoeial limits, would be considered contamination. Figure 17 graphically displays the particulate matter found in 5% dextrose solutions. The smooth curve represents the BP specification and the two others represent the samples. One product is within the specification (2G) and the other is not (2AL). According to the author, while in the 2G sample carbon black particles with sizes between 1 and 3 μm were identified, in the 2AL sample, besides carbon black (1–5 μm), some other particles such as starch grains, lacquer flakes, and rubber were found [80].

**TABLE 38 Peak Identification in Ethanol Extracts of Santoprene Tubing Samples 7 and 8**

| Peak Number | Tentative Compound Identification | Present in Material |
|---|---|---|
| 3 | 2,4-di-*t*-Butylphenol | 7, 8 |
| 5 | 4-(1,1,3,3-Tetramethylbutyl)phenol | 8 |
| 6 | Isomer of octyl phenol (TMS) | 8 |
| 7 | Hexadecane | 7 |
| 8 | Isomer of octyl phenol (TMS) | 8 |
| 9 | 4-Methyl-6-*tert*-octyl phenol | 8 |
| 10 | Isomer of decyl phenol | 7, 8 |
| 11 | Isomer of nonyl phenol (TMS) | 8 |
| 12 | Heptadecane | 7 |
| 13 | Isomer of undecyl phenol | 7, 8 |
| 14 | Octadecane | 7 |
| 15 | Isomer of decyl phenol (TMS) | 8 |
| 16 | Nonadecane | 7 |
| 17 | Isomer of undecyl phenol (TMS) | 7, 8 |
| 18 | Hexadecanoic acid, ethyl ester | 7 |
| 19 | Cyclohexadecane, heneicosane | 7 |
| 20 | Octadecanoic acid, ethyl ester | 7 |
| 21 | Docosane | 7 |
| 22 | Tetratriacontane, 9-methyl-nonadecane | 7, 8 |
| 23 | Tetracosane | 7 |
| 24 | Pentacosane | 7, 8 |
| 25 | Hexacosane, nonadecane | 7, 8 |
| 26 | Heptacosane | 8 |
| 27 | Irganox 1076 | 7, 8 |

*Source*: From ref. 78.

**TABLE 39 Limits for Particulate Contamination in Infusion Solutions Established by Pharmacopeial Compendia**

| Compendia | Volume | Particle Size (μm) | LO Limit | M Limit |
|---|---|---|---|---|
| BP | SVP | ≥10 | 6000/container | 3000/container |
| | | ≥25 | 600/container | 300/container |
| | LVP | ≥10 | 25/mL | 12/mL |
| | | ≥25 | 3/mL | 2/mL |
| USP | SVP | ≥10 | 6000/container | — |
| | | ≥25 | 600/container | — |
| | LVP | ≥10 | 25/mL | 12/mL |
| | | ≥25 | 3/mL | 2/mL |
| IP | SVP, LVP | — | One or more particles in more than one container | — |

*Note*: The limits are related to the method used for the determination: LO = light obscuration particle count test, M = microscopic particle count test, IP = International Pharmacopeia.

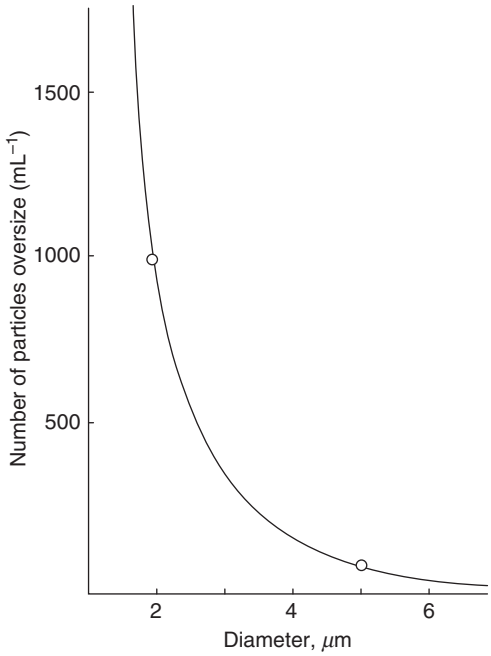


**FIGURE 16** Coordinate plot of log (number of particles oversize per milliliter) versus log (particle diameter) for BP specifications.

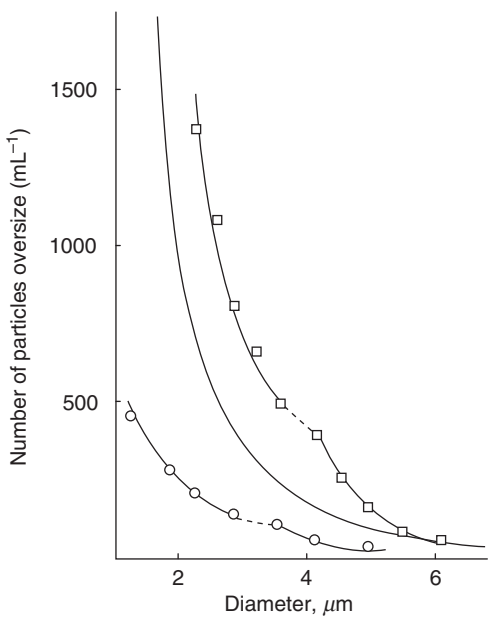


**FIGURE 17** Cumulative particle size distribution in 5% dextrose solutions for injection: (□) product 2AL packed in 500-mL glass containers with lacquer-coated rubber stopper; (○) product 2G packed in 500-mL plastic bags; (smooth curve), BP specification [80].

Techniques for counting particulate matter allow a discrimination of the particles by sizes even lower than the pharmacopoeial limits, and though not prescribed, it is also possible to characterize particulate matter by its composition. Table 41 displays the amount of particles found in infusion solutions classified according to size.

Since the results correspond to different studies, they are displayed in different fashions. Nevertheless, all solutions present particulate matter as contaminants, and it is possible to observe the exponential behavior of the relationship between the number and size of the particles.

Foroni et al. [81] measured and characterized inert particles both in individual solutions and in the final ternary mixture prepared by the sterile transfer technique. The results are presented in Table 40 and also in Figure 18. The difference between the number of particles greater than 2 and 5 μm is higher than 50. The histograms in Figure 18 show the number of particles of the individual components (Figure 18*a*) and the number of particles after transferring the components into an EVA bag (Figure 18*b*). The results show that the contamination of 10% KCl, 30% dextrose, amino acids, and lipid emulsion is significantly higher than of the other components (Figure 18*a*). The results also show that the concentration of particles in the components packaged in glass ampoules, KCl and disodium phosphate is also higher than in the other components. The identification of contaminants was also carried out, where particles composed of silica, aluminum, and sodium were identified in a KCl solution, and particles containing sulfur and silicate were identified in all bags. The authors consider the presence of sulfur to be related to rubber particles and attribute the presence of silicate to contact with talcum from gloves used by manipulators during the manufacturing of the bags.

Kamiya et al. [83] evaluated particulate contamination in 199 samples of admixed and un-admixed parenteral nutrition solution bags from 10 hospitals in Japan. Seven samples were used as controls since they had not been mixed with ampoules or vials (un-admixed samples). Size and number of particles were measured using a particle counter, and the identification of elements was carried out by scanning electron microscopy coupled to energy dispersion spectroscopy. The authors collected the residual volume of the samples (10–60 mL) after their usage. The results are presented in Table 40.

Figure 19 shows the scanning electron micrograph of two types of particles on the filter after filtration (0.22 μm membrane filter) of 50 mL of an admixture containing 1700 mL glucose/electrolytes/amino acids plus one vial of vitamin complex (Maltamin), two glass ampoules of 1 mL panthol (Pantol), two glass ampoules of 20 mL sodium chloride (Conclyte-Na), two glass ampoules of 2 mL metoclopramide (Primperan), and two glass ampoules of 10 mL potassium L-aspartate (Aspara K). Figure 19 also shows the identification of these particles. Their composition suggests that they are particles of glass (Figure 19*a*) and particles of rubber (Figure 19*b*).

A similar study was carried out by Ball et al. [82] in New Zealand and the United Kingdom. The authors analyzed 20 samples of adult and 20 samples of the pediatric PN admixtures, collecting the first and second fractions drawn from the infusion sets. The number of particles greater than 5 μm was 50 times higher than the number of particles greater than 40 μm in all solutions (Table 40). The analysis of the particulate matter allowed for a characterization of particles such as rubber and glass fragments (Table 41).

**TABLE 40 Mean (Range) of Particles Found in Mixed and Unadmixed Parenteral Nutrition Solutions Classified According to Their Size**

| | | | Particle Size | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | >1.3 μm | | >5 μm | | >10 μm | | >25 μm | | >50 μm | | |
| Sample | Volume (mL) | *n* | Mean | Range | Mean | Range | Mean | Range | Mean | Range | Mean | Range | Reference |
| 30 % Dextrose | 25 | 3 | 452 | ±127 | n.d.[a] | — | — | — | — | — | — | — | 81 |
| 50 % Dextrose | 25 | 3 | 2831 | ±278 | — | ±90 | — | — | — | — | — | — | 81 |
| Amino acids | 25 | 3 | 3715 | ±184 | — | ±58 | — | — | — | — | — | — | 81 |
| All-in-one solution (adult)[b] | 1 | 20 | 3.47 | ±1.24 | 1.40 | ±0.73 | 0.96 | ±0.45 | 1.06 | ±0.39 | — | — | 82 |
| Two-in-one (pediatric)[b] | 1 | 20 | 7.59 | ±2.56 | 11.77 | ±7.38 | 1.43 | ±1.09 | 0.38 | ±0.26 | — | — | 82 |
| Lipid emulsion (syringe packed) | 1 | 20 | 16.72 | ±10.85 | 8.98 | ±4.63 | 1.27 | ±1.23 | 0.76 | ±0.34 | — | — | 82 |
| Unadmixed samples | 10–60 | 7 | 62.7 | 8–146 | 1.70 | 0–4 | 0.4 | 0–2 | — | — | 0 | | 83 |
| Admixed samples[c] | 10–60 | 192 | 960.9 | 30–9539 | 42.8 | 0–587 | 6.4 | 1–146 | — | — | 0.09 | 0–1 | 83 |
| Solutions mixed using 1–3 glass ampoules[c] | 10–60 | 29 | 862.1 | 30–5707 | 31.3 | 0–176 | 4.4 | 0–24 | — | — | 0.1 | 0–1 | 83 |
| Solutions mixed during 4–13 glass ampoules[c] | 10–60 | 63 | 1163.4 | 142–9539 | 66.2 | 4–587 | 10.6 | 1–146 | — | — | 0.08 | 0–1 | 83 |

[a]n.d. = not detected.
[b]First milliliters collected.
[c]residual volume collected.

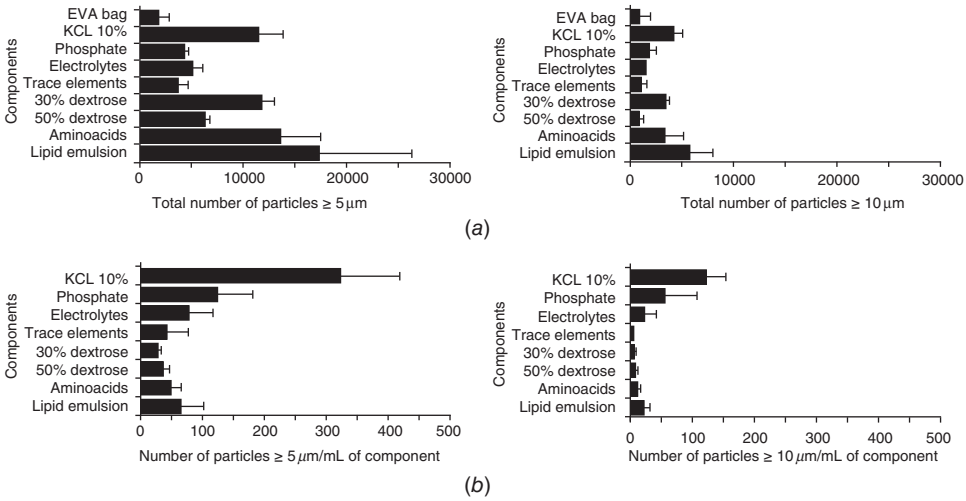


**FIGURE 18** (*a*) Total number of particles in each component of the admixture analyzed by filtration-observation method. (*b*) Number of particles per milliliter of each component analyzed by filtration-observation method.

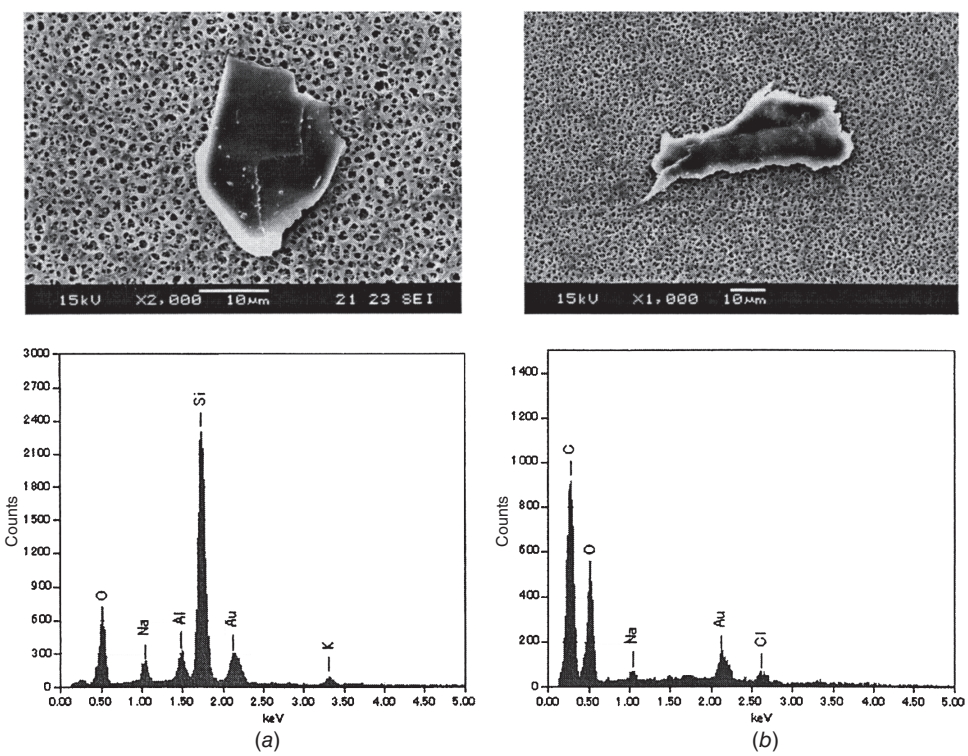


**FIGURE 19** Identification of two types of particles by scanning electron microscopy coupled to energy dispersion spectroscopy: (*a*) suggests glass particles; (*b*) suggests rubber particles [83].

Roseman et al. [84] used scanning electron microscopy to analyze flakes found in solutions stored in glass ampoules (type I glass ampoules). They observed that all flakes had similar characteristics (colorless, platelike, <1 μm in thickness), but their sizes ranged from a few micrometers to a 100 μm in length. The elemental analysis of four flakes gave the results presented in Table 42. Because the flakes presented the same elements of glass, the authors broke a glass ampoule and analyzed the fragments (results also presented in Table 42). Since the composition of the flakes and glass fragments were similar to each other, the authors believed that the flakes originated in the ampoule due to an attack on the glass surface by some chemical substance. It has been shown that, besides strong alkali or hydrofluoric acid, other substances such as EDTA and other organic acids such as citric acid, as well as phosphate salts are able to attack glass. The existence of such particulates in glass ampoules was attributed to a delamination process that could occur just above the bottom of the ampoule, a region of thermal stress because of the intense heat used to form the bottom of the container.

The major problem associated with particulate contamination in infusion fluids is not related to the composition itself (since they are mostly pieces of the container and therefore innocuous elements) but to the potential of each particle to cause

**TABLE 41 Elements in Particulate Matter Found in PN Solutions**

| Most Frequent Elements | Frequency of Appearance | Probable Origin | Reference |
|---|---|---|---|
| Oxygen | 1 | Glass; rubber | 82, 83 |
| Silicon | 1 | Glass; rubber | 81, 82, 83 |
| Carbon | 3 | Rubber | 82, 83 |
| Aluminum | 1 | Glass | 81, 82, 83 |
| Magnesium | 2 | Talc | 81, 82 |
| Sodium | 2 | Glass, rubber | 81, 83 |
| Chlorine | 3 | Rubber | 83 |
| Potassium | 3 | Glass | 83 |

*Note*: Frequency of appearance: 1 = high; 2 = medium; 3 = low.

**TABLE 42 Elemental Analysis of Flakes Found in Solution Stored Glass Ampoule and of Glass Fragments Resulting of a Broken Glass Ampoule**

| | Composition (%) | | | | |
|---|---|---|---|---|---|
| Element | Flake 1 | Flake 2 | Flake 3 | Flake 4 | Broken glass[a] |
| Silicon | 25–30 | 35–45 | 30–35 | 25–30 | 30–40 |
| Alumnum | 10–12 | 3–5 | 8–10 | 10–11 | 2–5 |
| Potassium | 4–6 | n.d.[b] | 3–5 | 3–4 | n.d. |
| Calcium | 1–2 | n.d. | n.d. | 1–2 | n.d. |
| Boron | 3–8 | 2–6 | 3–8 | Present | 4–10 |
| Sodium | 1–3 | 1–2 | 1–3 | 1–2 | 4–6 |

*Source*: From ref. 84.
*Note*: All ampoules were of type I glass.
[a]Mean of four samples; n.d. = not detected.

adverse reactions. The probability of an adverse reaction occurring is proportional to the number (and size) of particles introduced in the circulatory system. Turco et al. [85] suggest that responses resulting from the infusion of particles include physical occlusion, inflammatory responses, neoplastic responses, and antigenic responses. Animal studies have shown that the distribution of particles in tissues is related to their diameter. Particles larger than 8 μm are trapped in the lung capillaries, 3–6-μm particles are lodged in the spleen and hepatic lymph nodes, and 1-μm particles in the liver [86].

## 4 CONCLUDING REMARKS

The safety of drug therapy is closely related to the quality of drugs. All steps in drug processing may contribute to increasing the presence of foreign species in the final product. The requirements with respect to active ingredients, excipients, residual solvents, containers, and closures, although set up guidelines, are not a guarantee of products free of contaminants.

Although pharmacopeial biological reactivity tests for containers are good indicators of the toxicity of extractables, there are species that do not cause an acute toxic reaction but rather a chronic reaction, as is the case of phthalates (DEHP) and metallic species such as aluminum. Moreover, there are species that are extractable from packaging materials only by action of formulation constituents and therefore are not present in the extracts obtained by conventional pharmacopeial compendia tests.

The degradation of formulations due to lipid and vitamins peroxidation is a contamination problem not foreseen in pharmacopeial compendia since it is the exposure of the solutions to air and ambient light that induce peroxide generation.

Particulate contamination has been found in PN solutions and other intravenous drugs and fluids. Administration of particles through infusion solutions can result in adverse effects. The probability of these effects to occur increases proportionally with the amount of fluid administrated.

The presence of contaminants, though undesirable in all kind of drug formulations, is really critical in that intent for parenteral administration. Patients who require intensive or prolonged parenteral therapy, the immunocompromised, and neonates and infants might have increased susceptibility to the detrimental effect of contaminants.

## REFERENCES

1. U.S. Pharmacopeia (2005), USP 27, U.S. Pharmacopeial Convention, Rockville, MD.
2. World Health Organization (WHO) (2004), *Guidelines for Drinking-water Quality*, Vol. 1: *Recommendations*, WHO, Geneva.
3. Water programs, Environmental Protection Agency, National Interim Primary Drinking Water Regulations, *Code of Federal Regulations*, Part 141, (1985).
4. ASTM International (2005), *ASTM Book of Standards*, Vol. 11.02: *Water and Environmental* Technology: *Water (II)*, ASTM International, West Conshohocken, PA.

5. BP (2003), British Pharmacopoeia Commission, London.
6. Deutsche Arzneibuch (DAB) (2005), Deutschen Apotheker Verlag, Stuttgart.
7. European Pharmacopoeia (Ph.Eur.) (2004), 4th ed., European Directorate for the Quality of Medicines (EDQM), Ph.Eur., Strasbourg.
8. The International Pharmacopoeia (IP) (2003), 3rd ed., World Health Organization, Geneva.
9. Küster, F. W., and Thiel, A. (1985), *Rechentafeln für die Chemische Analytik*, 103 Auflage, Walter de Gruyter, Berlin, p. 258.
10. Koo, W. W., Kaplan, L. A., Horn, J., Tsang, R., and Steichen, J. (1986), Aluminum in parenteral nutrition solutions—sources and possible alternatives, *J. Parenteral Enteral Nutr.*, 10, 591–595.
11. Recknagel, S., Brätter, P., Chrissafidou, A., Gramm, H.-J., Kotwas, J., and Rösick, U. (1994), Parenteral aluminum loading in critical care medicine. Part I: Aluminum content of infusion solutions and solutions for parenteral nutrition, *Infusionsther. Transfusionsmed.,* 21, 266–273.
12. Bohrer, D., do Nascimento, P. C., Becker, E., de Carvalho L. M., and Dessuy, M. (2005), Arsenic species in solutions for parenteral nutrition, *J. Parenteral Enteral Nutr.*, 29, 1–7.
13. PluhatorMurton, M. M., Fedorak, R. N., Audette, P. J., Marriage, B. J., and Yatscoff, R. W. (1996), Extent of trace-element contamination from simulated compounding of total parenteral nutrient solutions, *Am. J. Health-Syst. Pharm.*, 53, 2299–2303.
14. Bohrer, D., Bortoluzzi, F., Nascimento, P. C., Carvalho, L. M., and Ramirez, A. G. (2008), Silicate release from glass for pharmaceutical preparations, *Int. J. Pharm.*, in press.
15. Bohrer, D., do Nascimento, P. C., Binotto, R., Becker, E., and Pomblum, S. G. (2002), Contribution of the raw material to the aluminum contamination in parenterals, *J. Parenteral Enteral Nutr.*, 26, 382–388.
16. Bohrer, D., Becker, E., do Nascimento, P. C., de Carvalho L. M., and Marques, M. S. (2006), Arsenic release from glass containers by action of intravenous nutrition formulation constituents, *Int. J. Pharm.*, 315, 24–29.
17. Thompson, J. E., and Davidow, L. (2003), *A Practical Guide to Contemporary Pharmacy Practice*, Lippincott Williams & Wilkins, Philadelphia.
18. Wade, A., and Weller, P. J. (1994), *Handbook of Excipients*, 2nd ed., Pharmaceutical Press, London.
19. Neuzil, J., Darlow, B. A., Inder, T. E., Sluis, K. B., Winterbourn, C. C., and Stocker, R. (1995), Oxidation of parenteral lipid emulsion by ambient and phototherapy lights: Potential toxicity of routine parenteral feeding, *J. Pediatr.*, 126, 785–790.
20. Steger, P. J. K., and Mühlebach, S. F. (1997), In vitro oxidation of IV lipid emulsions in different All-in-One admixture bags assessed by an iodometric assay and gas-liquid chromatography, *Nutrition*, 13, 133–140.
21. Murray, R. K., Granner, D. K., Mayes, P. A., and Rodwe, V. W. (2000), *Harper's Biochemistry*, 25th ed., McGraw-Hill, New York, p. 169.
22. Helbock, H. J., Motchnik, P. A., and Ames, B. N. (1993), Toxic hydroperoxides in intravenous lipid emulsions used in preterm infants, *Pediatrics*, 91, 83–87.
23. Picaud, J. C., Steghens, J. P., Auxenfans, C., Barbieux, A., Laborie, S., and Claris, O. (2004), Lipid peroxidation assessment by malondialdehyde measurement in parenteral nutrition solutions for newborn infants: A pilot study, *Acta Paediatr.*, 93, 241–245.
24. Pironi, L., Guidetti, M., Zolezzi, C., Fasano, M. C., Paganelli, F., Merli, C., Bersani, G., Pizzoferrato, A., and Miglioni, M. (2003), Peroxidation potential of lipid emulsions after compounding in all-in-one solutions, *Nutrition*, 19, 784–788.

25. Steger, P. J. K., and Mühlebach, S. F. (1998), Lipid peroxidatoin of IV lipid emulsions in TPN bags: The influence of tocopherols, *Nutrition*, 14, 179–185.
26. Burton, G. W., and Ingold, K. U. (1986), Vitamin E as an in vitro and in vivo antioxidant, *Ann. NY Acad. Sci.*, 570, 7–22.
27. Bowry, V. W., Ingold, K. U., and Stocker, R. (1992), Vitamin E in human low-density lipoprotein. When and how this antioxidant becomes a pro-oxidant, *Biochem. J.*, 288, 341–344.
28. Bowry, V. W., and Stocker, R. (1993), Tocopherol-mediated peroxidation. The pro-oxidant effect of Vitamin E on the radical-initiated oxidation of human low-density lipoprotein, *J. Am. Chem. Soc.*, 115, 6029–6040.
29. Steger, P. J. K., and Mühlebach, S. F. (2000), Lipid peroxidatoin of intravenous lipid emulsions and all-in-one admixtures in total parenteral nutrition bags: The influence of trace elements, *J. Parenteral Enteral Nutr.*, 24, 37–41.
30. Laborie, S., Lavoie, J-C., Rouleau, T., and Chessex, P. (2002), Multivitamin solutions for enteral supplementation: A source of peroxides, *Nutrition*, 18, 470–473.
31. Laborie, S., Lavoie, J-C., Pineaut, M., and Chessex, P. (2000), Contribution of multivitamins, air and light in the generation of peroxides in adult and neonatal parenteral nutrition solutions, *Ann. Pharmacother.*, 34, 440–445.
32. Helbock, H. J., Motchnick, P. A., and Ames, B. N. (1993), Toxic hydroperoxides in intravenous lipid emulsions used in preterm infants, *Pediatrics*, 91, 83–88.
33. Lavoie, J-C., Bélanger, S., Spalinger, M., and Chessex, P. (1997), Admixture of a multivitamin preparation to parenteral nutrition: The major contributor to in vitro generation of peroxides, *Pediatrics*, 99, 61–70.
34. Lavoie, J. C., Chessex, P., Rouleau, T., Migneault, D., and Comte, B. (2004), Light-induced byproducts of vitamin C in multivitamin solutions, *Clin. Chem.*, 50, 135–140.
35. Knafo, L., Chessex, P., Rouleau, T., and Lavoie, J-C. (2005), Association between hydrogen peroxide-dependent byproducts of ascorbic acid and increased hepatic acetyl-CoA carboxylase activity, *Clin. Chem.*, 51, 1462–1471.
36. Balet, A., Cardona, D., Jane, S., Molins-Pujol, A. M., Sanchez Quesada, J. L., Gich, I., and Mangues, M. A. (2004), Effects of multilayered bags vs ethylvinyl-acetate bags on oxidation of parenteral nutrition, *J. Parenteral Enteral Nutr.*, 28, 85–91.
37. Baker, M., Gregerson, M. S., Martin, S. M., and Buettner, G. R. (2003), Free radical and drug oxidation products in an intensive care unit sedative: Propofol with sulfite, *Crit. Care Med.*, 31, 787–792.
38. Lavoie, J-C., Lachance, C., and Chessex, P. (1994), Antiperoxide activity of sodium metabisulfite. A double-edged sword, *Biochem. Pharmacol.*, 47, 871–876.
39. Hayon, E., Treinin, A., and Wilf, J. (1972), Electronic spectra, photochemistry, and autoxidation: Mechanism of the sulfite-bisulfite-pyrosulfite systems. The $SO_2^-$, $SO_3^-$, $SO_4^-$ radicals, *J. Am. Chem. Soc.*, 94, 47–57.
40. International Conference on Harmonization (ICH) (1997), Guideline on residual solvents, ICH, Geneva.
41. Bohrer, D., do Nascimento, P. C., Becker, E., Bortoluzzi, F., depoi, F., and de Carvalho, L. M. (2004), Critical evaluation of the standard hydrolytic resistance test for glasses used for containers for blood and parenteral formulations, *PDA J. Pharm. Sci. Technol.*, 58, 96–105.
42. Scholze, H. (1988), *Glas, Natur, Struktur und Eigenschaften*, 3rd ed., Springer, Heidelberg, pp. 128, 324.

43. Bacon, F. R. (1986), Glass containers for parenterals, in Avis, K. E., Lachman, L., and Lieberman, H. A., Eds., *Pharmaceutical Dosages Forms: Parenteral Medications*, Vol 2, Marcel Dekker, New York, pp. 55–110.
44. Hak, E. B., Storm, M. C., and Helms, R. A. (1998), Chromium and zinc contamination of parenteral nutrient solution components commonly used in infants and children, *Am. J. Health-Syst. Pharm.*, 55, 150–154.
45. Borg, C., Constant, H., Fusselier, M., and Aulagner, G. (1994), Zinc, copper and iron in total parenteral nutrition mixtures: A contamination study, *Nutrition*, 13, 325–326.
46. Buchman, A. L., Neely, M., Grossie Jr, B., Truong, L., Lykissa, E., and Ahn, C. (2001), Organ heavy-metal accumulation during parenteral nutrition is associated with pathologic abnormalities in rats, *Nutrition*, 17, 600–606.
47. do Nascimento, P. C., Marques, M. S., Hilgemann, M., de Carvalho L. M., Bohrer, D., Pomblum, S. G., and Schirmer, S. (2006), Simultaneous determination of cadmium, copper, lead, and zinc in amino acid parenteral nutrition solutions by anodic stripping voltammetry and sample digestion by Uvirradiation, *Anal. Lett.*, 39, 1–14.
48. Alfrey, A. C., Le Gendre, G. R., and Kaenhy, W. D. (1976), The dialysis encefalophathy syndrome, possible aluminum intoxication, *N. Eng. J. Med.*, 294, 184–188.
49. Driscoll, W. R., Cummings, J. J., and Zorn, W. (1997), Aluminum toxicity in preterm infants, *N. Eng. J. Med.*, 337, 1090–1091.
50. Bohrer, D., do Nascimento, P. C., Binotto, R., and Pomblum, S. G. (2001), Influence of the glass packing on the contamination of pharmaceutical products by aluminium. Part I: Salts, glucose, heparin and albumin, *J. Trace Elem. Med. Biol.*, 15, 95–101.
51. Bohrer, D., do Nascimento, P. C., Martins, P., and Binotto, R. (2002), Availability of aluminum from glass and an Al-form exchanger in presence of complexing agents and amino acids, *Anal. Chim. Acta*, 459, 267–276.
52. Bohrer, D., do Nascimento, P. C., Binotto, R., and Becker, E. (2003), Influence of the glass packing on the contamination of pharmaceutical products by aluminium. Part III: Interaction container-chemicals during heating for sterilisation, *J. Trace Elements Med. Biol.*, 17, 107–115.
53. Bohrer, D., do Nascimento, P. C., Binotto, R., and Carlesso, R. (2001), Influence of the glass packing on the contamination of pharmaceutical products by aluminium. Part II: Amino acids for parenteral nutrition, *J. Trace Elements Med. Biol.*, 15, 103–108.
54. Popinska, K., Kierkus, J., Lyszkowska, M., Socha, J., Pietraszek, E., Kmiotek, W., and Ksiazky, J. (1999), Aluminum contamination of parenteral nutrition additives, amino acid solutions, and lipid emulsions, *Nutrition*, 15, 683–686.
55. Baydar, T., Aydin, A., Duru, S., Isimer, A., and Sahin, G. (1997), Aluminum in enteral nutrition formulas and parenteral solutions, *Clin. Toxicol.*, 35, 277–281.
56. Guidance for industry, container closure systems for packaging human drugs and biologics, chemistry, manufacturing and control documentation, Food and Drug Administration, 1999.
57. *Ullmann's Encyclopedia of Industrial Chemistry*, Vol. A5, 5th ed., VCH, Weinheim, 1993, p. 334.
58. Solomon, D. D., Jurgens, R. W., and Wong, K. L. (1986), Plastic containers for parenterals, in Avis, K. E., Lachman, L., and Lieberman, H. A., Eds., *Pharmaceutical Dosages Forms: Parenteral Medications*, Vol. 2, Marcel Dekker, New York, pp. 111–153.
59. U.S. Food and Drug Administration (FDA) (1999), Safety assessment of di-2-ethylhexylphthalate (DEHP) released from PVC medical devices, FDA, Center for Devices and Radiological Health, Rockville, MD.

60. Arabin, A., and Östelius, J. (1980), Determination by electron-capture gas chromatography of mono(2-ethylhexyl) phthalate and di(2-ethylhexyl) phthalate in intravenous solutions stored in poly(vinyl chloride) bags, *J. Chromatogra. B*, 193, 405–412.
61. Arabin, A., Jacobsson, S., Hagman, A., and Östelius, J. (1986), Studies on contamination of intravenous solutions from poly(vinyl chloride) bags with dynamic headspace gas chromatography–mass spectrometry and gradient liquid chromatography diode array techniques, *Int. J. Pharm.*, 28, 211–218.
62. Faouzi, M. A., Khalfi, F., Dine, T., Luyckx, M., Brunet, C., Gressier, B., Goudaliez, F., Cazin, M., Kablan, J., Belabed, A., and Cazin, J. C. (1999), Stability, compatibility and plasticizer extraction of quinine injection added to infusion solutions and stored in polyvinyl chloride (PVC) containers, *J. Pharm. Biomed. Anal.*, 21, 923–930.
63. Kambia, K., Dine, T., Gressier, B., Bah, S., Germe, A.-F., Luyckx, M., Brunet, C., Michaud, L., and Gottrand, F. (2003), Evaluation of childhood exposure to di(2-ethylhexyl) phthalate from perfusion kits during long-term parenteral nutrition, *Int. J. Pharm.*, 262, 83–91.
64. Allwood, M. C., and Martin, H. (1996), The extraction of diethylhexylphthalate (DEHP) from polyvinyl chloride components of intravenous infusion containers and administration sets by paclitaxel injection, *Int. J. Pharm.*, 127, 65–71.
65. Sautou-Miranda, V., Brigas, F., Vanheerswynghels, S., and Chopineau, J. (1999), Compatibility of paclitaxel in 5% glucose solutions with ECOFLAC low-density polyethylene containers-stability under different storage conditions, *Int. J. Pharm.*, 178, 77–82.
66. Marcato, B., Guerra, S., Vianello, M., and Scalia, S. (2003), Migration of antioxidants additives from various polyolefinic plastics into oleaginous vehicles, *Int. J. Pharm.*, 257, 217–225.
67. Smith, E. J., and Nash, R. J. (1986), Elastomeric closures for parenterals, in Avis, K. E., Lachman, L., and Lieberman, H. A., Eds., *Pharmaceutical Dosages Forms: Parenteral Medications*, Vol. 2, Marcel Dekker, New York, pp. 155–215.
68. Accardi-Dey, A., and Gschwend, P. M. (2003), Reinterpreting literature sorption data considering both absorption into organic carbon and adsorption onto black carbon, *Environ. Sci. Technol.*, 37, 99–106.
69. U.S. Food and Drug Administration (FDA) (1999), Guidance for industry, container closure systems for packaging human drugs and biologics, FDA, Rockville, MD.
70. Parenteral Drug Association (PDA) (1998), Pharmaceutical package integrity, Technical Report 27, PDA Bethesda, MD.
71. Mannermaa, J. P., Räisänen, J., Hyvönen-Dabek, M., Spring, E., and Yliruusi, J. (1994), Use of proton-induced X-ray emission (PIXE) analysis in the evaluation of large volume parenteral rubber stoppers, *Int. J. Pharm.*, 103, 125–129.
72. Delaunay-Bertoncini, N., van der Wielen, F. W. M., De Voogt, P., Erlandsson, B., and Schoenmakers, P. J. (2004), Analysis of low-molar-mass materials in commercial rubber samples by Soxhlet and headspace extractions followed by GC-MS analysis, *J. Pharm. Biomed. Anal.*, 35, 1059–1073.
73. Jenke, D. R. (1997), Utilization of extraction profiles to estimate the accumulation of extractables from polymeric materials, *J. Appl. Polym. Sci.*, 63, 843–848.
74. Loff, S., Kabs, F., Witt, K., Sartoris, J., Mandl, B., Niessen, K. H., and Waag, K. L. (2000), Polyvynilchloride infusion lines expose infants to large amounts of toxic plasticizers, *J. Pediatr. Surg.*, 35, 1775–1781.
75. Haishima, Y., Seshimo, F., Higuchi, T., Yamazaki, H., Hasegawa, C., Izumi, S., Makino, T., Nakahashi, K., Ito, R., Inoue, K., Yoshimura, Y., Saito, K., Yagami, T., Tsuchiya, T., and Nakasawa, H. (2005), Development of a simple method for predicting the levels of DEHP

migrated from PVC medical devices into pharmaceutical solutions, *Int. J. Pharm.*, 298, 126–142.

76. Kambia, K., Dine, T., Gressier, B., Germe, A.-F., Luyckx, M., Brunet, C., Michaud, L., and Gottrand, F. (2001), High-performance liquid chromatographic method for the determination of DEHP in total parenteral nutrition and in plasma, *J. Chromatogr. B*, 755, 297–303.
77. Boithias, S-B., Sautou-Miranda, V., Bourdeaux, D., Tramier, V., Boyer, A., and Chopineau, J. (2005), Leaching of DEHP from multilayer tubing into etoposide infusion solutions, *Am. J. Health-Syst. Pharm.*, 62, 182–188.
78. Jenke, D. R., Story, J., and Lalani, R. (2006), Estractables/leachables from plastic tubing used in product manufacturing, *Int. J. Pharm.*, 315, 75–92.
79. Groves, M. J. (1991), Particulate contamination in parenterals: Current issues, *Boll. Chim. Farma.*, 130, 347–354.
80. Bikhazi, A. B., Shiatis, J. A., and Haddad, A. F. (1977), Quantitative estimation of particulate matter in pharmaceutical preparations intended for intravenous administration, *J. Pharm. Sci.*, 66, 181–186.
81. Foroni, L. A., Rochat, M. H., Trouiller, P., and Calop, J. Y. (1993), Particle contamination in a ternary nutritional admixture, *J. Parenteral Sci. Technol.*, 47, 311–314.
82. Ball, P. A., Bethune, K., Fox, J., Ledger, R., and Barnett, M. (2001), Particulate contamination in parenteral nutrition solutions: Still a cause for concern? *Nutrition*, 17, 926–929.
83. Oie, S., and Kamiya, A. (2005), Particulate and microbial contamination in in-use admixed parenteral nutrition solutions, *Biol. Pharm. Bull.*, 28, 2268–2270.
84. Roseman, T. J., Brown, J. A., and Scothorn, W. W. (1976), Glass for parenteral products: A surface view using the scanning electron microscope, *J. Pharm. Sci.*, 65, 22–29.
85. Turco, S. J., and Davis, N. M. (1971), Detrimental effects of particulate matter on the pulmonary circulation, *J. Am. Med. Assoc.*, 217, 81–82.
86. Hearse, D. J., Sonmez, B., Saldanha, C., Braimbridge, M. V., Maxwell, M. P., and Erol, C. (1986), Particle-induced coronary vasoconstriction in the rat heart: Pharmacological investigation of underlying mechanisms, *Thorac. Cardiovasc. Surg.*, 34, 316–325.

# QUANTITATION OF MARKERS FOR GRAM-NEGATIVE AND GRAM-POSITIVE ENDOTOXINS IN WORK ENVIRONMENT AND AS CONTAMINANTS IN PHARMACEUTICAL PRODUCTS USING GAS CHROMATOGAPHY–TANDEM MASS SPECTROMETRY

ALVIN FOX
*University of South Carolina, Columbia, South Carolina*

**Contents**

1 Introduction
2 Analysis of GC–MS/MS Markers for LPS and PG and Applications
3 Operational Parameters
4 Concluding Remarks
References

## 1 INTRODUCTION

Endotoxins are bacterial cell envelope constituents that, when present in pharmaceutical products, cause pyrogenic reactions sometimes resulting in lethality. The toxicity of endotoxins is directly related to their chemical composition. However, the viability of the organism is irrelevant since endotoxin derived from dead or live microbes is equally active. The classical endotoxin is lipopolysaccharide (LPS). However, peptidoglycan (PG) also displays endotoxin-like activities. LPS is found only in gram-negative bacterial outer membranes, while PG is present in the cell

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

wall of both gram-positive and gram-negative bacteria. The *Limulus* amoebocyte lysate (LAL) test is widely used as a biological assay for LPS levels but has poor sensitivity for detection of PG. Substances are found in LPS and PG [3-hydroxy fatty acids (3-OH FAs) and muramic acid (Mur), respectively] that are rarely found elsewhere in nature and serve as chemical markers. Both markers for LPS and PG can be detected using gas chromatography–tandem mass spectrometry (GC–MS/MS) and this latter technology is the focus of this chapter.

A recently published book provides an excellent survey of issues that relate to contamination with endotoxins (present in both viable and nonviable bacteria), their released cell wall constituents, and also viable bacteria in the pharmaceutical industry [1]. It is important to know both the content of the work environment (e.g., indoor air) and the pharmaceutical products themselves. The former provides information on possible sources of microbial contamination and the latter the purity of the final commercial product (or precursors in various stages in its preparation). In some cases it is vital to know the actual bacterial species involved in contamination; culture-based methods are standard microbiological techniques which were the focus of Jimenez [1] and thus will not be discussed further. Any contamination (e.g., with endotoxins), regardless of the species of origin, is of utmost of importance (e.g., in determining the safety of a batch of antibiotics to be administered intravenously). This is determined optimally by non-culture-based methods.

Endotoxic reactions result from inflammatory responses to both culturable and nonculturable organisms. Thus, traditional culture-based methods often significantly underestimate the risk of endotoxin present. Culture- and non-culture-based monitoring methods differ greatly in their characteristics. Importantly, results from culture-based methods are strongly affected by variability in culture conditions (e.g., media selected, culture time, and temperature), making it difficult to standardize protocols for quantitation of contamination and only detecting those organisms that grow under the conditions selected. However, the use of non-culture-based procedures certainly does not eliminate the need for information obtained by culture.

Real-time polymerase chain reaction (PCR) methods have become the primary alternative to bacterial culture, measuring the levels of characteristic genes originating from particular species (whether culturable or not). PCR has proven to be highly amenable in the clinical microbiology laboratory, but its utility in more diverse environmental or pharmaceutical products can be more problematic. Of course, PCR will not detect an endotoxin. PCR enzymatically amplifies a genetic region, with a characteristic deoxyribonucleic acid (DNA) sequence; a flanking set of primers (short pieces of complementary DNA) focuses the enzyme on the DNA region of interest. Classical PCR involves detection of a PCR product by electrophoretic mobility on a gel, which is time consuming. Real-time PCR is distinct from classical PCR in that electrophoresis is avoided and the PCR product is detected simply by an increase in fluorescence; this can be performed with or without prior culture. This approach is of great utility, but there can be variation depending on the sample matrix (e.g., in some cases total inhibition of the reaction by metal ions or humic acids). Assessment of total bioload (e.g., by measuring the levels of chemical markers for bacterial endotoxins by GC–MS/MS) is much less subject to matrix-based variation and is more readily standardized.

An overview of the relative advantages (and disadvantages) of modern non-culture-based methods to each other and versus culture was also provided by

Jimenez [1], including discussing measurement of adenosine triphosphate (ATP) levels, microscopy/fluorescent-activated cell sorting, molecular biology (e.g., PCR and microarrays), and immunology-based methods. The endotoxin section focused on the LAL bioassay method, which is still the most common method for measuring such contamination; modern alternatives, including bioassays for cytokines and analytical chemical methods (GC–MS), were also discussed. The LAL assay measures biological activity which can vary with slight changes in structure among endotoxins affecting determination of levels; this is not the case for chemical analysis [2]. Furthermore biological assays, including the LAL assay, often suffer from false positives from cross-reactivity with other contaminants. As mentioned above, the LAL assay only detects the gram-negative endotoxin LPS, not the gram-positive endotoxin PG.

Chemical assays for endotoxin are considerably more selective than the LAL assay but until recently lacked the sensitivity required for routine detection of gram-positive and gram-negative endotoxins. This relates to the introduction of advanced but still user-friendly GC–MS/MS instrument which are rapidly replacing GC–MS for trace analysis in the pharmaceutical industry and elsewhere. This chapter will thus focus on GC–MS/MS methodology for detection of 3-OH FAs (markers for LPS) and Mur (a marker for PG), which are each well established [3–8]. Ergosterol is also used as a marker for fungal contamination using GC–MS/MS [9].

Endotoxicity results from the interaction of a bacterial cell envelope component (e.g., LPS or PG with a cell surface receptor constituting part of the nonspecific immune system, (i.e., a toll-like receptor on white blood cells). This results in the production of cytokines [e.g., interleukin 1 (IL-1) or tumor necrosis factor (TNF)] as part of an intracellular enzyme cascade which can cause severe tissue injury. Bioassays or immunoassays can be used to detect such reactions respectively. As noted above the most widely used bioassay is the LAL assay. A lysate of amoebocytes of the horseshoe crab (*Limulus*) contains an enzymatic clotting cascade which is activated by extremely low levels of LPS (nanogram levels or lower). There are variants of this assay that can detect PG, but they are not as widely used. As noted above, other bioassays employ cultured cell lines that respond to LPS or PG, respectively. Unfortunately bioassays are highly amenable to false positives (from the presence of cross-reactive substances) or false negatives from inhibition (by contaminants present in the sample) [10]. A detailed discussion of these assays is beyond the scope of this chapter and has been reviewed elsewhere [1].

## 2 ANALYSIS OF GC–MS/MS MARKERS FOR LPS AND PG AND APPLICATIONS

For groups experienced in trace analysis of chemical markers for bacteria, currently samples are analyzed almost exclusively by GC–MS/MS. GC–MS/MS assays are now well established and a wide range of clinical and environmental samples have been analyzed. However, application in the pharmaceutical industry requires further evaluation. For example, Mur is released by hydrolysis and analyzed as an alditol acetate; 3-OH FAs after methanolysis are converted to methyl trimethylsilyl derivative. Detailed analytical procedures have been described elsewhere for Mur [3, 4, 7, 8, 11] and for 3-OH FAs [2, 5, 12, 13]. The compounds of interest contain active

groups (e.g., OH or $CO_2H$) that interact with GC columns. Derivatization is necessary to convert the samples into a suitable form for analysis. Prederivatization and postderivatization clean-up yields simple chromatograms. Quantitation employs stable-isotope-labeled forms of the markers (e.g., $^{13}C$-labeled Mur derived from labeled blue-green algae).

The 3-OH FAs have had great utility in the determination of LPS levels in indoor air. However, in tissues and body fluids it has been determined that 3-OH FAs are naturally present at low levels as products of mammalian metabolism (mitochondrial fatty acid β oxidation). Due to this background GC–MS/MS for 3-OH FAs is not recommended as a general marker to determine trace LPS levels in clinical samples [14]. However, in certain situations the assessment of 3-OH FAs has been successfully used, for example, in the diagnosis of chronic peridontitis [15]. There is great potential for the utility of 3-OH FAs as markers for LPS contamination in pharmaceutical products, where often the background matrix would be anticipated to be much less complex.

Chemical analysis of Mur levels has proved effective in both clinical and environmental samples, since Mur is not synthesized by eukaryotic cells. For example, it is readily detected in infected human body fluids, for example, synovial fluids from patients with staphylococcal arthritis and spinal fluids from those with pneumococcal pneumonia [7, 16]. However, the most widely used method for its analysis, as an alditol acetate, is time consuming. A large number of derivatives have been tested in order to develop a simpler alternative. Unfortunately the limit of detection for these alternative approaches has not been optimal [17, 18].

The marker monomer is chemically converted into a volatile form suitable of passage through a gas chromatograph. The samples then pass into the tandem mass spectrometer where the ionized molecules are detected. GC–MS/MS employs GC separation coupled with the exquisite selectivity of MS/MS. In MS (monitoring/quantitation mode), background peaks are screened out. MS/MS screens out background a second time, thus dramatically lowering the detection limit. Alternatively, in the identification mode (MS/MS), a chemical fingerprint of the compound of interest allows definitive identification. The analysis of Mur serves as an example: natural $^{12}C$ muramic acid is first released from PG polymers (present as a minor component in a complex sample matrix) by hydrolysis. Conversion of $^{12}C$ muramic acid to a volatile derivative, muramicitol lactam pentaacetate molecular weight (MW) 445, is essential for GC–MS/MS analysis.

Maximal selectivity, in GC–MS analysis, requires selected ion monitoring (SIM). In SIM, one or more prominent ions characteristic of a given compound are monitored exclusively, thereby ignoring background or contaminant ions. Similarly, maximum selectivity in GC–MS/MS analysis involves twofold SIM, or multiple-reaction monitoring (MRM). In quadropole mass spectrometers, the first stage of MRM involves selective transmission of a molecular ion from the first mass spectrometer to the collision cell. This instrumental clean-up removes all nonidentical molecular weight contaminants produced from the initial ionization. The selected ion is then fragmented by collision with an inert gas (e.g., argon). In the second mass spectrometer the selected fragment ion is monitored. If either sensitivity or specificity is adversely affected, then the limit of detection is affected. GC–MS/MS provides much greater specificity than GC–MS. Using GC–MS it was impossible to reliably visually differentiate chromatograms that contained the lowest concentrations of

Mur from negative controls (including plants and fungi). GC–MS/MS chromatograms of dust were always readily differentiated from controls. However, these analyses were clearly performed at the current limits of sensitivity for GC–MS/MS. Current developments in MS technology may lead to dramatically improved sensitivity by GC–MS/MS.

Tandem mass spectrometry consists of two stages. For example, in Mur analysis, in the first stage, the molecule is isolated essentially intact with a MW of 403 due to the loss of a ketene (loss of 42). Coeluting molecules of different MW are largely but not totally eliminated; that is, only molecules with a MW of 403 are permitted to pass into the next stage. In the second stage, molecules with MW of 403 are broken into characteristic fragments including one that contains the original lactam with a MW of 198. Thus, in the second stage, only molecules with a MW of 198 are detected. Thus, for a second time, background molecules of different MW that coelute are essentially eliminated.

The use of a stable-isotope-labeled ($^{13}C$) analog of muramic acid, in each sample, verifies that there is no false-negative result. This assures that muramic acid is not lost during the sample preparation or hidden within the background in the instrumental analysis. Although $^{12}C$ muramic acid and $^{13}C$ muramic acid have the same retention time on GC analysis, they can be discriminated in the tandem mass spectrometer. GC–MS/MS analysis of $^{13}C$ muramic acid is identical to natural ($^{12}C$) muramic acid. However, the MWs in the first and second stages are correspondingly higher, 412 and 205, respectively. Thus in the tandem mass spectrometer, it is possible to monitor two separate windows simultaneously, one for $^{13}C$ muramic acid (top window) and one for natural muramic acid (bottom window). Accurate quantitation is readily accomplished by comparing the ratio of the area of $^{13}C$ versus $^{12}C$ muramic acid (see Figure 1).

A chemical fingerprint is generated consisting of scission products, in the MS/MS instrument, characteristic of the compound of interest. The parent molecule has a MW of 403. In each case the same major fragments are observed: masses 361, 301,

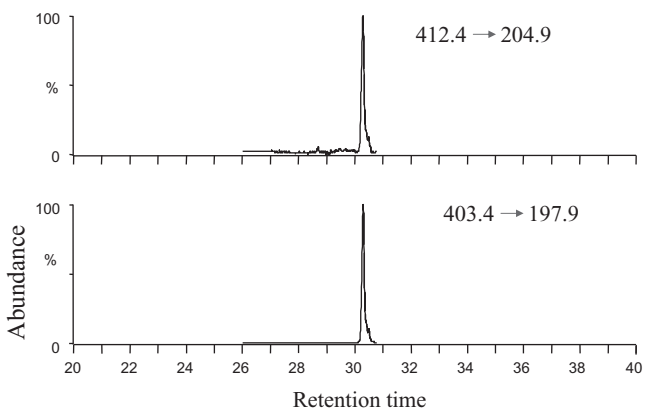


**FIGURE 1** GC–MS/MS monitoring chromatogram. The upper window depicts the internal standard ($^{13}C$ muramic acid) and the lower natural ($^{12}C$) muramic acid isolated from dust. The peak areas in the two separate windows are normalized relative to the highest peak in that window. Medical samples appear similar

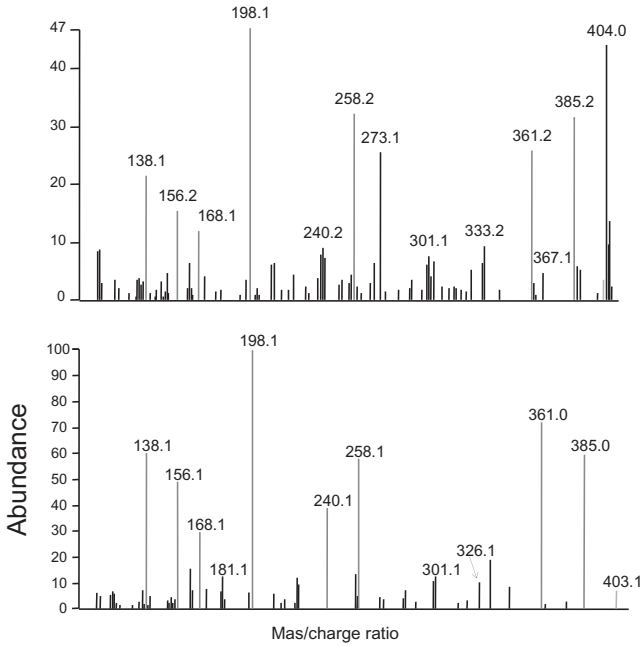


**FIGURE 2** GC-MS/MS chemical fingerprint (product ion spectrum) of (*a*) standard muramic acid (2 ng total in sample) and (*b*) muramic acid isolated from dust. Medical samples appear similar.

258, 240, 198, 156, and 138: 361 results from the loss of a ketene (loss of 42) and 301 from a subsequent loss of acetic acid (loss of 60). Breakage at C4–C5 (loss of 145) would generate 258 and further loss of acetic acid (loss of 60) would generate 198. Loss of ketene or acetic acid from 198 results in 156 and 138, respectively (see Figure 2).

## 3 OPERATIONAL PARAMETERS

Every marketed product has a level of endotoxin tolerated based on the minimum pyrogenic dose and amount of the drug to be administered as per Food and Drug Administration (FDA) guidelines [19]. However, there are none for the more advanced chemical assays described here. Indeed there are only a few highly specialized university laboratories that currently have experience in trace chemical analysis of LPS and PG. There are no commercial testing laboratories. Simplification and automation will allow more widespread availability of these methods.

At the current time it is recommended that 1–10 mg of solid sample be analyzed. In environments that are not heavily contaminated with airborne dust it takes several days for sample collection. In an unoccupied or minimally occupied room around 1 mg of dust needed for analysis can be collected in 72 h. In a heavily occupied room the levels of dust increase dramatically and collection takes around 6–8 h.

Chemically inert membranes (e.g., Teflon) are preferably used for dust collection since they are not affected by heating in sulfuric acid, the first stage of the chemical analysis.

Large (milligram-to-gram) quantities of dust can readily be collected from air conditioners or surfaces. The concentration of dust in air varies from micrograms to milligrams pero cubic meter. Low concentrations are more typical of office buildings and laboratories, while the high range is more typical of dusty environments such as barns or chicken houses. Since commonly used pumps collect a few liters per minute, except in grossly contaminated environments, only microgram amounts of dust can be collected without extended air-sampling periods. Chemical markers of interest are only one component of bacteria and the bacteria constitute only a small fraction of the dust. Mur is present at between 5 and 50 ng/mg (corresponding to approximately 50–500 ng PG/mg dust) in normal house and air conditioning dust. Somewhat higher levels have been found in airborne dust (over 100 ng Mur/mg). LPS is present at between 500 and 5000 ng/mg dust. Solids (e.g., pharmaceutical products) are more simply analyzed, generally without any sample pretreatment. Generally, for the optimal level of sensitivity, around 10 mg should be analyzed.

As noted above, sensitive and specific GC–MS/MS methods for the determination of 3-OH FAs and Mur have been developed. MS is an alternative to the classical LAL assay for determination of LPS, while no other regulated approach exists for PG assessment. These chemical methods are reproducible and provide quantitative, accurate determination of microbial biocontamination. At the present time mass spectrometric measurement of LPS and PG have matured sufficiently to be used for routine assessment of air quality. Numerous products of medical and environmental origin have been analyzed. However, use for assessment of pharmaceutical products remains limited.

The precision of MS assays is in the range typical of most clinical assays (i.e., under 5–15%). The best choice of internal standard is the stable-isotope-labeled form (preferably $^{13}C$) of the compound of interest (e.g., β-hydroxy myristic acid or muramic acid). Specific trace detection of chemical markers in complex matrices requires appropriate negative controls. Procedures are often described that do not employ the mass spectrometer and false positives are often reported. The mere analysis of blank filters or water blanks is not satisfactory since chemical noise contributed by the sample is much greater and is not accounted for with this form of control.

As noted above the predominant technique for measuring endotoxin levels is the LAL assay. This assay involves assessing activation of a clotting cascade in amoebocyte lysates (from the horseshoe crab). The LAL primarily detects LPS, and the sensitivity of detection of PG by LAL is extremely poor. The LAL bioassay and MS measurements of LPS sometimes correlate poorly. LAL measures biological activity while MS measures total quantity. Thus the two techniques are not strictly comparable. However, some differences (between results with LAL versus MS) may relate to the superior specificity of tandem mass spectrometry. It is worthy of note that GC–MS/MS can provide some information on the population of gram-negative bacteria present since distribution of hydroxy FAs varies among bacterial species. The 2- and 3-OH fatty acids with 10–18 carbon atoms are all common in organic dust.

## 4 CONCLUDING REMARKS

User-friendly commercial GC–MS/MS instruments have been available since the mid-1990s. Unfortunately there has been limited development in instruments that perform automated processing of samples for GC–MS/MS analysis, although a prototype automated derivatization has been built [20]. The pharmaceutical industry is well versed in the use of other types of mass spectrometry detection methods for high-throughput drug analysis [e.g., liquid chromatography–tandem mass spectrometry (LC-MS/MS) which eliminates the necessity for posthydrolysis derivatization by employing electrospray ionization mass spectrometry]. Unfortunately there has been limited interest in applying these methods for detection of LPS or PG markers [21]. Thus analysis of markers for endotoxins remains technically demanding and this has inhibited the widespread use of these techniques outside of a few specialist laboratories. However, there is great potential in the application of GC–MS/MS methods for assessing contamination with gram-positive and gram-negative bacterial endotoxins in the pharmaceutical industry. Tandem mass spectrometers are still expensive. However, such instruments are widely available in the pharmaceutical industry for general drug analysis. The most modern machines are run by Windows-based PC programs. Thus they can be readily operated by individuals, after appropriate training, with little prior experience of mass spectrometry or indeed analytical chemistry.

## ACKNOWLEDGMENT

The research described in this chapter was supported by Philip Morris USA, Inc., and Philip Morris International.

## REFERENCES

1. Jimenez, L. (2004), *Microbial Contamination Control in the Pharmaceutical Industry*, Marcel Dekker, New York.
2. Saraf, A., Larsson, L., Burge, H., and Milton, D. (1997), Quantification of ergosterol and 3-hydroxy fatty acids in settled house dust by gas chromatography–mass spectrometry: Comparison with fungal culture and determination of endotoxin by a *Limulus* amoebocyte lysate assay, *Appl. Environ. Microbiol.*, 63, 2554–2559.
3. Fox, A., Wright, L., and Fox, K. (1995), Gas chromatography tandem mass spectrometry for trace detection of muramic acid, a peptidoglycan chemical marker in organic dust, *J. Microbiol. Meth.*, 22, 11–26.
4. Fox, A., Krahmer M., and Harrleson, D. (1996), Monitoring muramic acid in air (after alditol acetate derivatization) using a gas chromatograph–ion trap tandem mass spectrometer, *J. Microbiol. Meth.*, 27, 129–138.
5. Saraf, A., and Larsson, L. (1996), Use of gas chromatography ion trap tandem mass spectrometer for the determination of microorganisms in organic dust, *J. Mass Spectrom.*, 31, 389–396.
6. Larsson, L., and Saraf, A. (1997), Use of gas chromatography–ion trap tandem mass spectrometry for the detection and characterization of microorganisms in complex samples, *Mol. Biotechnol.*, 7, 279–287.

7. Kozar, M., Krahmer, M. T., Fox, A., and Gray, B. M. (2000), Failure to detect muramic acid in normal rat tissues but detection in cerebrospinal fluid from patients with pneumococcal meningitis, *Infect. Immun.*, 68, 4688–4698.
8. Kozar, M., Laman, J. D., and Fox, A. (2002), Muramic acid is not generally present in human spleen as determined by gas chromatography–tandem mass spectrometry, *Infect. Immun.*, 70, 741–748.
9. Sebastian, A., and Larsson, L. (2003), Characterization of the microbial community in indoor environments: A chemical-analytical approach, *Appl. Environ. Microbiol.*, 69, 3103–3109.
10. Kaneko, T., Goldman, W. E., Mellroth, P., Steiner, H., Fucase K., Kusumoto, S., Harley, W., Fox, A., Golenbock, D., and Silverman, N. (2004), Monomeric and polymeric Gram-negative peptidoglycan but not purified LPS stimulate the *Drosophila* IMD pathway, *Immun.*, 20, 1–20.
11. Fox, A., Harley, W., Feigley, C., Salzberg, D., Toole, C., Sebastian, A., and Larsson, L. (2005), Large particles are responsible for elevated bacterial marker levels in school air upon occupation, *J. Environ. Monit.*, 7, 450–456.
12. Mielniczuk, Z., Mielniczuk, E., and Larsson, L. (1993), Gas chromatography–mass spectrometry methods for analysis of 2- and 3-hydroxylated fatty acids: Application for endotoxin measurement, *J. Microbiol. Meth.*, 17, 91–102.
13. Mielniczuk, Z., Mielniczuk, E., and Larsson, L. (1995), Determination of muramic acid in organic dust by gas chromatography–mass spectrometry, *J. Chromatogr. B*, 670, 167–172.
14. Sponazr, B., Norin, E., Midvedt, T., and Larrson, L. (2002), Limitations in the use of 3-hydroxy fatty acids to determine endotoxin in mammalian tissues, *J. Microbiol. Meth.*, 50, 283–289.
15. Ferrando, R., Szponar, B., Sánchez, A., Larsson, L., and Vaero-Guillén, P. L. (2005), 3-Hydroxy fatty acids in saliva as diagnostic markers in chronic peridontitis, *J. Microbiol. Meth.*, 62, 285–291.
16. Fox, A., Fox, K., Christensson, B., Krahmer, M., and Harrelson, D. (1995), Absolute identification of muramic acid at trace levels in human septic fluids *in vivo* and absence in aseptic fluids, *Infect. Immun.*, 64, 3911–3955.
17. Kozar, M., and Fox, A. (2002), Analysis of a stable halogenated derivative of muramic acid by gas chromatography–negative ion chemical ionization tandem mass spectrometry, *J. Chromatogr.*, 946, 229–238.
18. Sebastian, A., Harley, W., Fox, A., and Larsson, L. (2004), Evaluation of the methyl ester *O*-methyl acetate derivative of muramic acid for the determination of peptidoglycan in environmental samples by ion-trap GC-MS-MS, *J. Environ. Monit.*, 6, 1–6.
19. U.S. Department of Health and Human Services (DHHS) (1987), FDA guidelines on validation of the *Limulus* amoebocyte lysate test as an end product test for human and animal parenteral drugs, biological products and medical devices, DHHS, Rockville, MD.
20. Steinberg, P., and Fox, A. (1999), Automated derivatization instrument: Preparation of alditol acetates for analysis of bacterial carbohydrates using gas chromatography–mass spectrometry, *Anal. Chem.*, 71, 1914–1917.
21. Shahgholi, M., Ohorodnik, S., Callahan, J., and Fox, A. (1997), Trace detection of underivatized muramic acid in environmental dust samples by microcolumn liquid chromatography–electrospray tandem mass spectrometry, *Anal. Chem.*, 69, 1956–1960.

# MICROBIOLOGY OF NONSTERILE PHARMACEUTICAL MANUFACTURING

Ranga Velagaleti
*BASF Corporation, Florham Park*

**Contents**

1 Introduction
2 Global Regulations and Regulatory Guidances Relevant to Microbial Bioburden Control during Nonsterile Manufacturing
  2.1 GMPs for Finished Products and Components of Finished Products
  2.2 GMPs for Active Pharmaceutical Ingredients
  2.3 GMPs for Bulk Pharmaceutical Excipients
3 Pharmacopeial Guidelines Relevant to Microbiology of Nonsterile Manufacturing
  3.1 United States Pharmacopeia
  3.2 European Pharmacopoeia
  3.3 Japanese Pharmacopoeia
4 Current Regulatory Expectations for Microbial Bioburden Control during Nonsterile Manufacturing
5 Industry Perspective on Microbial Bioburden Control for Nonsterile Pharmaceutical Manufacturing
  5.1 Survey Results
  5.2 Recommendations
6 Microbial Bioburden Control during Shelf Life of Pharmaceutical Products
7 Summary and Conclusions
References

## 1 INTRODUCTION

In order to comply with general microbiological quality attributes for nonsterile pharmaceuticals and established monograph or drug application specifications for

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

raw materials and finished products, manufacturers need to follow the good manufacturing practice (GMP) regulations, with specific reference to contamination control stated in various sections of GMPs. For example, GMP regulations promulgated by the U.S. Food and Drug Administration (FDA) for finished products in the *Code of Federal Regulations* state (21 CFR Part 211.113): "Appropriate written procedures, designed to prevent objectionable microorganisms in drug products not required to be sterile, shall be established and followed" [1]. To meet this regulatory requirement, the microbial bioburden of nonsterile pharmaceutical raw materials and finished-product manufacturing environment (air, walls, and floors), equipment used for manufacturing, as well as the finished raw materials and drug products should be monitored and controlled to ensure that they have acceptable amount of total microbial bioburden and are free of objectionable microorganisms. Historically, the microbial contamination of nonsterile pharmaceuticals has been a concern, with a number of products recalled due to microbial contamination with undesirable microorganisms due to their suspected health hazards. As early as 1979, Dunnigan of the Bureau of Medicines at the FDA, commenting on the acceptable levels and contamination of microbiological contamination in drug products, expressed his concern that topical preparations contaminated with gram-negative organisms are a probable moderate to serious health hazard [2].

The U.S. Pharmacopeia (USP), in a recent revision of general chapter <1111> (585), "Microbiological Examination of Nonsterile Products: Acceptance Criteria for Pharmaceutical Preparations and Substances for Pharmaceutical Use," stated: "The presence of certain microorganisms in nonsterile preparations may have the potential to reduce or even inactivate the therapeutic activity of the product and has a potential to adversely affect the health of the patient. Manufacturers have therefore to ensure low bioburden of finished dosage forms by implementing current guidelines on GMPs during the manufacture, storage and distribution of pharmaceutical preparations" [3].

Pharmaceutical manufacturers and their trade associations have been proactive in addressing the concerns related to microbial contamination in nonsterile manufacturing environment. In an article in the March 1997 issue of *Pharmaceutical Technology*, the Pharmaceutical Research and Manufacturers Association of America (PhRMA) Environmental Monitoring Work Group provided a realistic assessment of the need for understanding and controlling microbial bioburden contributed by raw materials, primary packaging components, the manufacturing process, and the manufacturing environment. The work group emphasized in this article that, except for sterile products, most pharmaceutical products are neither intended to be nor represented as being sterile, but because they are administered to people who are ill and in a weakened state, the microbial content should be minimized [4]. The Australian Pharmaceutical Manufacturers' Association (APMA) published a guideline on nonsterile pharmaceuticals for human use emphasizing the need for control of microbial contamination and the importance of adhering to GMPs to control microbiological levels [5].

Besides GMP regulations for finished dosage forms (which also include regulatory requirements for inactive ingredients, active ingredients, containers, and closures used for finished products) published by the FDA [1], the International Conference on Harmonization (ICH) [6] (GMPs for active ingredients), authorities in Australia [7, 8] (pharmaceutical products), as well as the World Health Organization (WHO) [9] (pharmaceutical products), International Pharmaceutical

Excipients Council (IPEC) [10] (excipients), and the USP [11] have also published GMP guidelines. These guidelines require manufacture of pharmaceutical products with acceptable microbial bioburden and free of objectionable microorganisms.

This chapter discusses regulatory aspects of microbiology of nonsterile pharmaceutical manufacturing, including manufacturing environment, raw materials, and finished products. In addition, the microbiological bioburden control of pharmaceutical raw materials and products is reviewed, with particular reference to the type of finished dosage forms. Assessment of microbial bioburden requirements for nonsterile pharmaceutical manufacturing by the pharmaceutical industry is discussed. The microbiology of the pharmaceutical manufacturing environment for production of sterile pharmaceuticals or microbiological quality expectations for sterile products is not discussed in this chapter.

## 2 GLOBAL REGULATIONS AND REGULATORY GUIDANCES RELEVANT TO MICROBIAL BIOBURDEN CONTROL DURING NONSTERILE MANUFACTURING

GMP regulations for finished pharmaceutical drugs and components of finished drugs promulgated by the FDA [1], GMP guidelines for active pharmaceutical ingredients (Q7A) from the ICH [6], and GMP guidelines for bulk pharmaceutical excipients published by the USP [11] are discussed to illustrate the GMP requirements for pharmaceutical finished drug products, active (drug substances) and inactive (excipients) components of the drug products, respectively. In the context of this discussion on GMP regulations relevant to the microbiology of nonsterile manufacturing, *contamination* is defined as "the undesired introduction of impurities of a chemical or microbiological nature, or of foreign matter, into or on to a raw material, intermediate, or API during production, sampling, packaging, or repackaging, storage, or transport" [6]. A similar definition of contamination applies with reference to drug products. Microbiological bioburden refers to the level and type of microorganisms that may be present. The drug product, drug product components, manufacturing equipment, or environment can be considered contaminated when bioburden exceeds established specifications or acceptance criteria.

### 2.1 GMPs for Finished Products and Components of Finished Products

The FDA enforces the implementation of current good manufacturing practice (cGMP, GMP) regulations (21 CFR Parts 210 and 211) [1] for drug products manufactured in the United States (also applicable to drugs and drug components manufactured abroad and imported into the United Sates), under the Food, Drug, and Cosmetic Act (FD&C Act). Under FD&C Act 501(a)(2)(B), failure to comply with any GMP regulation shall render such drug to be adulterated and such drug as well as the person who is responsible for failure to comply shall be subjected to regulatory action [1].

It is important to understand the regulatory definitions of finished drug product, active ingredient, and inactive ingredient [1] in the context of the discussions on the subject presented here. Under GMP regulations 21 CFR Part 210.3(b)(4), drug product is defined as a finished dosage form, for example, tablets, capsule, and solution, that contains an active ingredient generally, but not necessarily, in association

with inactive ingredients. The term also includes a finished dosage form that does not contain an active ingredient but is intended to be used as a placebo. Placebo is used especially during safety and efficacy studies during drug development and in preclinical animal studies or human clinical trials. Active ingredient (also known as active drug substance) means any component that is intended to furnish pharmacological activity or other direct effect on the diagnosis, cure, mitigation, treatment, or prevention of disease or to affect the structure or any function of the body of humans or other animals [21 CFR Part 210.3(b)(7)]. Inactive ingredient (also known as excipient) means any component other than an active ingredient [21 CFR Part 210.3(b)(8)].

One of the important aspects of GMP regulations is the quality control during manufacturing to ensure that the drug products are free of contamination from objectionable microorganisms. GMPs [1] require that the personnel engaged in drug product manufacture wear protective apparel such as head, face, hand, and arm covering (garbing) as necessary to protect drug products from contamination [Part 211, Section 211.28(a)]. The regulations also require that the flow of components, drug product containers, closures, labeling, in-process materials, and drug products through the building or buildings be designed to prevent contamination [211.42(b)]. Contamination prevention also includes designation of separate or defined areas or such other control systems for the manufacturing operations as are necessary to prevent contamination [211.42(c)]. Appropriate equipment is required for adequate control over microorganisms [211.46(b)]. Adequate exhaust systems or other systems adequate to control contaminants are required where air contamination is likely to occur during production [211.46(c)]. Water is used in many areas of pharmaceutical manufacturing operations. GMPs require [211.48(a)] that incoming potable water into the manufacturing plant should comply with the U.S. Environmental Protection Agency's (EPA's) Primary Drinking Water Regulations (40 CFR Part 141), where control of microbiological contamination in potable water is defined. GMPs require written procedures be designed to prevent contamination of equipment, components, drug product containers, closures, packaging, and labeling materials or drug products and that such procedures be followed [21 CFR Part 211.56(c)]. Equipment cleaning and maintenance are also required to prevent contamination [21 CFR Part 211.67(a)]. Components and drug product containers and closures should at all times be handled and stored in a manner to prevent contamination [21 CFR Part 211.80(b)]. GMPs also require testing and release as described in 21 CFR Part 211.84(d)(6), which states that each lot of component, drug product container, or closure that is liable to microbiological contamination that is objectionable in view of its intended use should be subjected to microbiological tests before use. Appropriate written procedures, designed to prevent objectionable microorganisms in drug products not required to be sterile, should be established and followed [21 CFR Part 211.113(a)]. GMP regulations in 21 CFR Part 211.165(b) require appropriate laboratory testing be performed, as necessary, of each batch of drug product required to be free of objectionable microorganisms.

### 2.2 GMPs for Active Pharmaceutical Ingredients

The ICH Q7A guidance for GMPs for active pharmaceutical ingredients (APIs) [6] illustrates GMP expectations for microbial control during API production. The ICH

guideline specifies that where microbiological specifications have been established for intermediates during API production, or for an API, facilities should be designed to limit exposure to objectionable microbiological contaminants, as appropriate (ICH Q7A Section 4.1). Adequate ventilation, air filtration, and exhaust systems have been suggested to minimize risk of contamination due to microorganisms (Q7A Section 4.2). If tighter microbiological specifications for water used for API manufacture are called for, appropriate specifications for total microbial counts, objectionable microorganisms, and/or endotoxins should be established (Q7A Section 4.2). The guideline also states that where the manufacturer of a nonsterile API either intends or claims that it is suitable for use for further processing to produce a sterile drug product, water used in the final isolation and purification steps for such API should be monitored and controlled for total microbial counts, objectionable microorganisms, and endotoxins (Q7A Section 4.3). The guideline recommends cleaning of equipment assigned for continuous production or campaign production of successive batches of the same intermediate or API at appropriate intervals to prevent buildup and carryover of objectionable levels of microorganisms (Q7A Sections 5.2 and 8.5). Under the laboratory controls section of the guideline, ICH recommends that if the API has a specification for microbiological purity, appropriate action limits for total microbial counts, objectionable organisms, and endotoxins should be established and met (Q7A Section 11.1). Appropriate microbiological specification tests should be conducted on each batch of intermediate or API where microbial quality is specified (Q7A Section 11.2). The guidelines specify that equipment cleaning/sanitation studies should address microbiological and endotoxin contamination for those processes where there is a need to reduce total microbiological count or endotoxins in the API, where a nonsterile API is used to manufacture sterile products (Q7A Section 12.7). For APIs manufactured by cell culture, fermentation techniques, raw materials used such as media, and buffer components may have potential for microbiological contamination. For example, in such cases, depending on the intended use of the API or intermediate, control of bioburden, viral contamination, and/or endotoxins during manufacturing and monitoring of the process at appropriate stages may be necessary (Q7A Section 18.1). When the quality of the API can be affected by microbial contamination, manipulations using open vessels should be performed in a biosafety cabinet or similarly controlled environment (Q7A Section 18.3).

### 2.3 GMPs for Bulk Pharmaceutical Excipients

The USP general chapter <1078> outlines the GMP guidelines for bulk pharmaceutical excipients [11]. The guideline states that building and facilities should be designed so that operations performed within do not contribute to an actual or potential contamination of the excipient. An effective and regular cleaning program of equipment used is recommended to remove product residues and dirt, which may also contain microorganisms and act as a source of contamination. Further, the guideline states that all equipment that has been in contact with contaminated material should be thoroughly cleaned and disinfected before coming in contact with an excipient. A controlled environment may be necessary to avoid microbial contamination. Potable water used for production of excipients should be compliant with chemical and microbiological standards, including freedom from pathogenic

organisms. Purified water is also used for production of pharmaceutical excipients, and the systems used to produce purified water from potable water (deionizers, ultrafiltration, or reverse-osmosis systems) may have potential for microbial growth. Appropriate specifications for chemical and microbiological quality of water used for pharmaceutical production and periodic testing to demonstrate compliance with specifications are recommended. When excipient product specifications require it to be endotoxin or pyrogen-free, and water is used for production of that excipient, validation of the purified water systems to produce endotoxin and pyrogen-free water is required.

## 3 PHARMACOPEIAL GUIDELINES RELEVANT TO MICROBIOLOGY OF NONSTERILE MANUFACTURING

United States, European, and Japanese pharmacopeia have described general requirements, specifications, and tests for monitoring microbial bioburden in nonsterile pharmaceutical products. Although there are minor differences in the expectations of the pharmacopeia, the general principles for microbial bioburden monitoring remain similar as described below.

### 3.1 United States Pharmacopeia

In the general chapter "Microbiological Examination of Nonsterile Products" [3], the USP provides guidance for microbial examination of various nonsterile pharmaceutical dosage forms (Table 1). Enumeration of the total aerobic counts and total yeasts and molds in products are determined using procedures stated in USP general chapter <61> [12], and tests for specified microorganisms (*Staphylococcus aureus*, and *Pseudomonas aeruginosa, Escherichia coli, Salmonella* sp., and *Candida albicans*) are performed using procedures stated in USP general chapter <62> [13]. Acceptance criteria for microbiological quality of nonsterile dosage forms for various routes of administration are also provided. In addition, in pharmacopeial monographs, microbiological specifications are described where applicable in monographs for excipients, actives, and drug products. Microbiological best laboratory practices are described in general chapter <1117> [14], which provides guidance on implementing good laboratory practice standards when examining various sample matrixes for microbial bioburden in the laboratory. Alternative microbiological testing methods other than those specified in USP can be used provided such methods are appropriately validated. The procedures for validation of alternative microbiological methods are described in USP general chapter <1223> [15].

The USP also provides guidance on reduced microbial limits testing for product release and stability evaluation when drug products have reduced water activity [ratio of vapor pressure of $H_2O$ in product ($P$) to vapor pressure of pure $H_2O$ ($P_o$) at the same temperature] well below 0.75. Water is required for microbial growth in the pharmaceutical products [16]. However, the more resistant microorganisms, including spore-forming *Clostridium* sp., *Bacillus* sp., *Salmonella* sp., and filamentous fungi, which may not proliferate in the drug product with low water activity, may persist in the dormant state in the product. For example, a water activity of 0.61 is required for the growth of the fungus *Xeromuces bisporus*, a water activity

**TABLE 1 Acceptance Criteria for Microbial Bioburden Control of Nonsterile Dosage Forms**

| Route of Administration | Total Aerobic Count Limit (CFU/g, CFU/mL) | Total Combined Yeasts and Molds Limit (CFU/g, CFU/mL) | Specified Microorganism/s (in 1 g or 1 mL) |
|---|---|---|---|
| **Nonaqueous preparations for oral use** | 1000 | 100 | Absence of *E. coli* |
| **Aqueous preparations for oral use** | 100 | 10 | Absence of *E. coli* |
| **Rectal use** | 1000 | 100 | |
| **Oromucosal use (gingival, cutaneous, nasal, auricular)** | 100 | 10 | Absence of *S. aureus* and *P. aeruginosa* |
| **Vaginal use** | 100 | 10 | Absence of *S. aureus, P. aeruginosa*, and *C. albicans* |
| **Transdermal patches (one patch including adhesive layer and backing)** | 100 | 10 | Absence of *S. aureus* and *P. aeruginosa* |
| **Inhalation use (other than liquid preparations for nebulization)** | 100 | 10 | Absence of *S. aureus, P. aeruginosa*, and bile-tolerant gram-negative bacteria |

*Source*: From ref. 3.

of 0.62 is required for the growth of the yeast *Zygosaccharomyces rouxii*, while the majority of other fungi and yeasts require water activity higher than 0.75 for growth. Water activity of 0.75 is adequate for the growth of the bacterium *Halobacterium halobium*, but the majority of other bacteria require higher water activity for growth. The guidance also provides for a microbial limit testing strategy for pharmaceutical products for various routes of administration based on water activity. For example, for compressed tablets and liquid-filled capsules with representative water activities of 0.36 and 0.30, respectively, reduced testing is recommended, while for topical creams and nasal inhalants, with water activity of 0.97 and 0.99, respectively, testing for total aerobic counts, total yeasts and molds, *S. aureus*, and *P. aeruginosa* is recommended. Antimicrobial preservatives are added to nonsterile pharmaceutical dosage forms to protect them from microbial growth or from microorganisms that are introduced inadvertently during or subsequent to the manufacturing process. USP general chapter <51> describes how to perform antimicrobial effectiveness testing [17]. In this chapter, it is emphasized that antimicrobial preservatives should not be used as a substitute to GMPs or solely to reduce the viable microbial population of a nonsterile product. It is recommended that antimicrobial effectiveness of preservatives must be demonstrated for dosage forms such as multiple-dose topical and

oral dosage forms, ophthalmic, otic, nasal, and irrigation as described in the USP general chapter on pharmaceutical dosage forms [18].

The USP also provides guidance on water for pharmaceutical purposes [19]. The guidance points out that the major exogenous source of microbial contamination of bulk pharmaceuticals is the source water. Source water (potable water) should meet quality attributes of drinking water, in which coliform levels are controlled. Other microorganisms present in the incoming water, although not objectionable in nature, may compromise subsequent purification steps. Microorganisms present in source water may absorb to carbon beds, deionized resin beds, and filter membranes (used in the processing of potable water to purified water) and initiate the formation of a biofilm. Microorganisms in a biofilm are adapted to the prevailing low-nutrient environment. Microorganisms in the biofilm may also move downstream when they are shed from the existing biofilm and carried to other areas of the water system. A detailed account of biofilm formation in the pipelines and methods for detection and quantitation of biofilms is provided by Olson (1997) [20]. Microbial contamination may also come from unprotected vents, faulty air filters, ruptured disks in the water system, and backflow from contaminated outlets, among others. Several categories of microorganisms may be problematic to the pharmaceutical water systems and as a result to the pharmaceutical products, which are manufactured using water from these systems. The microorganisms may include opportunistic or overt pathogens, nonpathogenic indicators of potentially undetected pathogens, or microorganisms that could be resistant to a preservative in a drug product or which can degrade an active or inactive ingredient in a drug product. Defining alert and action levels of microorganisms in water systems and monitoring these levels by microbiological testing at scheduled intervals can serve as an early warning that will allow remedial actions to occur and prevent a system from producing water unfit for pharmaceutical use. For purified water, for example, maximum action levels are considered to be 100 colony-forming units of microorganisms per milliliter (CFUs/mL) [19].

### 3.2 European Pharmacopoeia

For pharmaceutical products, the European Pharmacopoeia describes sampling plans and tests for quantitative enumeration of bacteria and fungi that can grow under aerobic conditions [21]. Enterobacteria and certain other gram-negative bacteria, *E. coli, Salmonella* sp., *P. aeruginosa*, and *S. aureus* and Clostridia and counts for *Clostridium perfringens* [22] are described under tests for specified microorganisms. In the general chapter on microbiological quality of pharmaceutical preparations [23], pharmaceutical dosage forms are classified into three categories: Category 1 is for sterile pharmaceutical preparations, which is not the subject of this chapter, and categories 2 and 3 describe requirements for various nonsterile dosage forms. For example, for category 2, preparations for topical use, testing for total aerobic count and fungi, enterobacteria and other gram-negative bacteria, *P. aeruginosa*, and *S. aureus* is performed. Under category 3A, preparations for oral rectal administration, testing for total viable aerobic count and *E. coli* is performed, whereas under category 3B, where raw materials of natural origin are included in the drug products, testing for total viable aerobic count, nitrobacteria and other gram-negative bacteria, *Salmonella* sp., *E. coli*, and *S. aureus* is performed. Microbiological specifications

are detailed wherever applicable under each monograph. Antimicrobial preservative effectiveness tests are recommended, where antimicrobial preservatives are used, such as for multidose containers where there is a potential for microbiological contamination, or aqueous or topical preparations, which also have potential for microbial contamination with high moisture content in these dosage forms. Under the general chapter on efficacy of antimicrobial preservation [24], both the rationale for using antimicrobial preservatives and the tests for demonstrating effectiveness of antimicrobial preservatives are described.

### 3.3 Japanese Pharmacopoeia

In the general chapter on microbial attributes of nonsterile pharmaceutical products, the guidance suggests that the presence of microbial contaminants in nonsterile products [25] can reduce or inactivate the therapeutic activity of the product and has the potential to adversely effect the health of the patients and recommends manufacturers to ensure that contamination levels are as low as possible for finished dosage forms. Microbial enumeration limits for raw materials (total aerobic microbial count and total combined yeasts and molds count) and finished dosage forms are described. For inhalation, nasal, and topical routes of administration, tests for total aerobic microbial count and total combined and yeast and mold count, *P. aeruginosa*, and *S. aureus* are recommended. For vaginal preparations, testing for total aerobic microbial count and total combined yeast and mold count, *E. coli, S. aureus*, and *C. albicans* is recommended, while for oral liquids and solids, testing for total aerobic microbial count and total combined and yeast and mold count and *E. coli* is recommended. The microbiological assessment of preservatives is required when preservatives are used in a pharmaceutical product to control microbial bioburden. The test microorganisms and methods for evaluating the efficacy of the preservative in pharmaceutical products are described in the general chapter on preservative effectiveness tests [26].

## 4 CURRENT REGULATORY EXPECTATIONS FOR MICROBIAL BIOBURDEN CONTROL DURING NONSTERILE MANUFACTURING

Regulatory expectations for microbial bioburden for nonsterile pharmaceutical products are reviewed using the FDA guide to inspections of microbiological quality control laboratories [2], purified water systems [27], topical products [28], and oral solutions and suspensions [29].

In the guide to inspections of microbiological quality control laboratories [2], a number of problems associated with microbiological contamination of topical drug products, nasal solutions, and inhalation products have been emphasized. In that guidance, Dunnigan of the FDA is quoted as saying that topical preparations contaminated with gram-negative organisms are a probable moderate to serious health hazard.

In the guide to inspections of topical products [28], it is indicated that water deionizers are usually excellent breeding areas for microorganisms, where flow rates, temperature, surface area of resin beds, and microbial quality of the feed water all influence microbial growth. Since topical products (e.g., creams, ointments)

generally contain purified water, the microbial integrity of the product is contingent on the microbial bioburden of purified water. The guidance also points out that in assessing the significance of microbial contamination of the product, both the identification of isolated microorganisms and the number of organisms found are significant. When high numbers of nonpathogenic microorganisms are present, they may affect product efficacy and/or physical and chemical stability. Topical creams may tend to separate on standing. For topical products, data should be available to show continued effectiveness of preservatives throughout the product's shelf life.

Purified water is normally used for the manufacture of nonsterile pharmaceutical products. The guide to inspections of high-purity water systems [27] primarily discusses microbiological aspects and suggests that an action limit of over 100 CFU/mL for a purified water system is unacceptable. The purpose of establishing an action limit is to assure that the water system is under control. Purified water used in the manufacture of drug products should be free of "objectionable microorganisms," which are defined as organisms that when present can cause infections when the drug product is used as directed. The definition can also include organisms capable of growth in the drug product. Microorganisms may exist in water systems as either free floating or attached to the walls of pipes and tanks (biofilms). Microorganisms continuously slough off from biofilms and contribute to movement of microorganisms upstream or downstream in the water systems. The guidance emphasizes that establishing the level of contamination allowed in high-purity water systems used in the manufacture of a nonsterile product requires an understanding of the use of the product, the formulation (preservative system), and the manufacturing process. In the manufacture of antacids, which do not have an effective preservative system, an action limit below 100 CFU/mL is required. The inspection guide also points out that equipment used in purified water systems may be liable to contamination. Pumps used in water systems, if not continuously used, may have still (stagnant) water, which is the site of microbial growth. *Pseudomonas* sp. contamination reported in the water system was attributed to a pump which was operational only periodically. Deadlegs and fittings are a source of microbial bioburden in water systems. Threaded fittings should not be used in pharmaceutical water systems, as they are a source of microbial growth. Filters are another source of contamination, as microbes tend to accumulate and grow on their surface. While a 0.2-μm point-of-use filter can mask the level of microbiological contamination, it will not necessarily stop endotoxin contamination, since the toxins can pass through the filter. Filters should be frequently changed to prevent contamination.

In a water system, contamination by *Pseudomonas* sp. was found after FDA testing and the same species was also found in a topical steroid product after FDA testing [27]. The inspection resulted in product recall and issuance of a warning letter. The root cause for microbial bioburden was found to be a one-way water system that employ ultraviolet (UV) light to control microbiological contamination, where the UV light is turned on only when water is needed, and when the UV lights are not on, sterilization is ineffective and the standing water is a source of microbial growth. A flexible hose used in this water system that was difficult to sanitize may also have contributed to the observed microbial bioburden.

In the guide to inspection of oral solutions and suspensions [29], it is stated that in some oral liquids microbiological contamination can present significant health hazard. For instance, microbiological contamination with gram-negative

microorganisms is objectionable in some oral liquids, especially those used in infants and immunocompromised patients. In oral liquids such as antacids, contamination with *Pseudomonas* sp. is objectionable. In general, contamination of any preparations with gram-negative organisms is not desirable. Such contamination may suggest a deficient manufacturing process, inadequate preservative system, and potential use of contaminated raw materials.

## 5 INDUSTRY PERSPECTIVE ON MICROBIAL BIOBURDEN CONTROL FOR NONSTERILE PHARMACEUTICAL MANUFACTURING

The PhRMA Environmental Monitoring Work Group published an article in the March 1997 issue of *Pharmaceutical Technology* on microbiological monitoring of environmental conditions for nonsterile manufacturing [4]. This publication is a compilation of survey results of nonsterile manufacturing facilities within the United States and recommendations based on the survey results. In this section, a summary of the survey results and recommendations published in this article are reviewed with a few comments from this author based on experiences in this area.

### 5.1 Survey Results

Most manufacturers conducted the monitoring of nonsterile pharmaceutical manufacturing environments. Companies producing topicals, liquids, and aerosols tend to have more extensive programs than those companies manufacturing tablets or other solid oral dosage forms. Among the nonsterile manufacturing environments monitored were air, processing equipment, and water. Air quality was monitored using centrifugal air samplers, settling plates, and so on. Product contact surfaces of equipment were monitored using swabs or contact plates, with swabs being preferred because of better access of irregular surfaces. Since swabs used are presterilized, they are less likely to contaminate the surface being sampled, whereas contact plates which contain nutrient agar media may encourage growth of microorganisms after sampling, unless the sampled area is thoroughly cleaned with ethanol or other sterilizing agent. However, contact plates are easier to use on plain flat surfaces and also help provide an accurate assessment of in situ microbial status.

Companies with monitoring programs have established alert and action limits. Alert levels of microorganisms may indicate a change from normal operating procedures but may not require any corrective action, whereas action levels may suggest a significant change from normal operating procedures and require corrective actions to control further proliferation of microbes.

### 5.2 Recommendations

***Specific Recommendations*** The work group recommended that the routine monitoring of microbiological activity in nonsterile environments should not be mandatory but should be determined by the type of nonsterile dosage form manufactured. In general, cleaning and sanitization procedures, good maintenance schedules, well-defined process control programs, raw material quality, facility design and control, and training of personnel involved in various aspects of manufacturing can

contribute to microbial bioburden control. A defined monitoring program can provide data on historical trends and such data can help identify deviations, root-cause analysis, troubleshooting, and early and effective actions. Among nonsterile dosage forms, inhalation products are of particular concern, followed by liquids and topical formulations, with a lower priority given to solid oral dosage forms. When monitoring programs are instituted, sampling should include those areas that are most likely to be contaminated, such as product contact surfaces of processing equipment, ventilation systems, process gases and purified water, and water systems. Sampling frequency can be based on historical trends and types of dosage forms manufactured, focusing in particular on those products more likely to be susceptible to microbial contamination. Manufacturing process, cleaning, and utilities validations should include microbial sampling to ensure microbial bioburden control. Holding times for process steps such as coating solutions and wet granulation should be determined by sampling and testing these matrices for microbial bioburden during validation and subsequent manufacturing.

***General Recommendations*** Complying with GMP regulations specified in 21 CFR Part 211 [1] is the most effective way of controlling microbial bioburden in a nonsterile manufacturing environment as (as described under Section 2.1). Appropriate written procedures should be established to control microbial bioburden and prevent objectionable microorganisms. Microbiologists should be well trained in sampling and testing of the pharmaceutical manufacturing environment, so that microbial bioburden is not introduced into product or environmental samples during sampling and testing. Identification of isolates should be undertaken as applicable when action levels (or above acceptance criteria) are seen during quantitation of microbial bioburden.

## 6 MICROBIAL BIOBURDEN CONTROL DURING SHELF LIFE OF PHARMACEUTICAL PRODUCTS

GMPs require that the stability of pharmaceutical products be tested to evaluate the product integrity throughout the shelf life of the product [1, 6, 11]. The ICH guidance Q1A (R2), stability testing of drug substances and drug products [30], emphasizes that the stability testing should cover, as appropriate, the physical, chemical, biological, and microbiological attributes. In addition, for drug products, if antimicrobial preservatives are used, preservative content should also be investigated.

## 7 SUMMARY AND CONCLUSIONS

In nonsterile pharmaceutical manufacturing, control of microbial bioburden through appropriate design of manufacturing facilities and implementation of GMPs during manufacturing would help control microbial bioburden in pharmaceutical raw materials and finished products. The GMPs for finished product and its components (active and inactive ingredients) are discussed in this chapter to highlight expectations for contamination control during manufacturing. In support of GMP

expectations for contamination control and testing required to establish such control, pharmacopeia in the United States, Europe, and Japan provide test methods and procedures for determination of microbial bioburden, which are also summarized in this chapter. Manufacturers of pharmaceutical products through their trade associations have taken a proactive approach to microbial bioburden control in nonsterile manufacturing by publishing results from surveys of member companies and have made general recommendations for implementation of microbial bioburden control by member companies. A brief summary of the industry work group survey results on current industry practices and general recommendations for implementation is presented. The requirement for microbial bioburden monitoring throughout the shelf life of the products is also emphasized.

## REFERENCES


1. U.S. Food and Drug Administration (FDA) (2007), Current good manufacturing practice regulations, 21 CFR Parts 210 and 211, FDA, Rockville, MD.
2. U.S. Food and Drug Administration (FDA) (1993), Guide to inspections of microbiological pharmaceutical quality control laboratories, FDA, Rockville, MD.
3. U.S. Pharmacopeia (USP) (2007), <1111> Microbiological examination of nonsterile products: Acceptance criteria for pharmaceutical preparations and substances for pharmaceutical use, U.S. Pharmacopeial Convention, Rockville, MD.
4. Pharmaceutical Research and Manufacturers Association of America (PhRMA) (1997, March), Microbiological monitoring of environmental conditions for nonsterile manufacturing, *Pharm. Technol.*, 58–74.
5. Australian Pharmaceutical Manufacturers' Association (APMA) (1990), *The Control of Microbial Contamination in Nonsterile Pharmaceutical Products for Human Use*, APMA, Canberra, Australia.
6. International Conference on Harmonization (ICH) (2001), Guidance for industry. Q7A good manufacturing practice guidance for active pharmaceutical ingredients, ICH, Brussels, Belgium.
7. Therapeutic Goods Administration Laboratories of Australia (TGAL) (1990), *Guidelines for Assessing the Results of Microbiological Tests on Nonsterile Pharmaceuticals for Human Use*, TGAL, Canberra, Australia.
8. Tang, S. (1998), Microbial limits reviewed—Basis for unique Australian regulatory requirements for microbial quality of nonsterile pharmaceuticals, PDA, *J. Pharm. Sci. Technol.*, 52, 100–109.
9. World Health Organization (WHO) (2003), Good manufacturing practices for pharmaceutical products, main principles, WHO Technical Report Series No. 908, WHO, Geneva, Switzerland.
10. International Pharmaceutical Excipients Council (IPEC) (2006), *Good Manufacturing Practices*, IPEC, Washington DC.
11. U.S. Pharmacopeia (USP) (2006), Good manufacturing practices for bulk pharmaceutical excipients, general chapter <1078>, U.S. Pharmacopeial Convention, Rockville, MD.
12. U.S. Pharmacopeia (USP) (2007), Microbiological examination of nonsterile products: Microbial enumeration tests <61>, U.S. Pharmacopeial Convention, Rockville, MD.
13. U.S. Pharmacopeia (USP) (2007), Microbiological examination of nonsterile products: Tests for specified microorganisms <62>, U.S. Pharmacopeial Convention, Rockville, MD.

14. U.S. Pharmacopeia (USP) (2007), Microbiological best laboratory practices, general chapter <1117>, U.S. Pharmacopeial Convention, Rockville, MD.
15. U.S. Pharmacopeia (USP) (2007), Validation of alternative microbiological methods, general chapter <1223>, U.S. Pharmacopeial Convention, Rockville, MD.
16. U.S. Pharmacopeia (USP) (2007), Application of water activity determination to nonsterile pharmaceutical products, general chapter <1112>, U.S. Pharmacopeial Convention, Rockville, MD.
17. U.S. Pharmacopeia (USP) (2007), Antimicrobial effectiveness testing, general chapter <51>, U.S. Pharmacopeial Convention, Rockville, MD.
18. U.S. Pharmacopeia (USP) (2007), Pharmaceutical dosage forms, general chapter <1151>, U.S. Pharmacopeial Convention, Rockville, MD.
19. U.S. Pharmacopeia (USP) (2007), Water for pharmaceutical purposes, general chapter <1231>, U.S. Pharmacopeial Convention, Rockville, MD.
20. Olson, P. W. (1997), Biofilms in the pipeline and in the patient, *J. Pharm. Sci. Technol.*, 51(6), 252–261.
21. European Pharmacopoeia (Eur. Ph.) (2005), Microbiological examination of nonsterile products (total viable aerobic count), general chapter 2.6.12, Council of Europe, Strasbourg Cedex, France.
22. European Pharmacopoeia (Eur. Ph.) (2005), Microbiological examination of nonsterile products (tests for specified microorganisms), general chapter 2.6.13, Council of Europe, Strasbourg Cedex, France.
23. European Pharmacopoeia (Eur. Ph.) (2005), Microbiological quality of pharmaceutical preparations, general chapter 5.1.4, Council of Europe, Strasbourg Cedex, France.
24. European Pharmacopoeia (Eur. Ph.) (2005), Efficacy of antimicrobial preservation, general chapter 5.1.3, Council of Europe, Strasbourg Cedex, France.
25. Japanese Pharmacopoeia (JP) (2001), Microbiological attributes of nonsterile pharmaceutical products, general information chapter 7, Society of Japanese Pharmacopoeia, Tokyo.
26. Japanese Pharmacopoeia (JP) (2001), Preservatives—Effectiveness tests, general information chapter 12, Society of Japanese Pharmacopoeia, Tokyo.
27. U.S. Food and Drug Administration (FDA) (1993), Guide to inspections of high purity water systems, FDA, Rockville, MD.
28. U.S. Food and Drug Administration (FDA) (2000), Guide to inspections of topical drug products, FDA, Rockville, MD.
29. U.S. Food and Drug Administration (FDA) (1994), Guide to inspection of oral solutions and suspensions, FDA, Rockville, MD.
30. International Conference on Harmonization (ICH) (2003), Guidance for industry, Q1A (R2) stability testing of new drug substances and products, ICH, Brussels, Belgium.

# STABILITY AND SHELF LIFE OF PHARMACEUTICAL PRODUCTS

RANGA VELAGALETI
*BASF Corporation, Florham Park, New Jersey*

**Contents**

1 Introduction
2 Stability Requirements in GMP Regulations and Guidelines
  2.1 Finished Products
  2.2 Excipients
  2.3 Active Pharmaceutical Ingredients
3 Stability Requirements for Excipients
4 Stability Requirements for Drug Substances (APIs)
  4.1 Selection of Batches and Container Closure System
  4.2 Storage Conditions and Testing Frequency
  4.3 Stress Studies and Stability-Indicating Methods for Analysis of API Stability
  4.4 Evaluation of Stability Results
  4.5 Stability Commitment
  4.6 Storage Statement and Labeling
5 Stability Requirements for Drug Products
  5.1 Selection of Batches and Container Closure System
  5.2 Bracketing and Matrixing
  5.3 Storage Conditions and Testing Frequency
  5.4 Stress Studies and Stability-Indicating Methods for Analysis of Drug Product Stability
  5.5 Evaluation of Stability Results
  5.6 Stability Commitment
  5.7 Storage Statement and Labeling
6 Photostability Studies
  6.1 Photostability of APIs
  6.2 Photostability of Drug Products

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

7 Evaluation of Stability Data and Shelf Life Determination
7.1 Shelf Life Estimation for Drug Substances or Drug Products Intended for Room Temperature Storage
7.2 Shelf Life Estimation for Drug Substances or Drug Products Intended for Storage in a Refrigerator
7.3 Shelf Life Estimation for Drug Substances or Drug Products Intended for Storage in a Freezer
7.4 Shelf Life Estimation for Drug Substances or Drug Products Intended for Storage below –20°C
8 Assignment of Climatic Zones and Recommended Storage Conditions
9 Stability Testing Parameters for Different Dosage Forms
9.1 Stability Tests Common to All Dosage Forms
9.2 Stability Tests Specific to Dosage Forms
10 Summary and Conclusions
References

## 1 INTRODUCTION

Pharmaceutical drug products (e.g., tablets, capsules, creams, injectables, and inhalation products) and the components in a drug product, that is, the active pharmaceutical ingredient (API, drug substance), and the inactive ingredients (excipients) should be stable in the drug product for the proposed shelf life duration of the drug product and the proposed duration of the shelf life of the individual components. Shelf life (retest or expiration date) is the time period during which excipients, APIs, and drug products are expected to remain within approved shelf life specification, provided that they are stored under the conditions defined on the container label. Stability and storage conditions for excipients, APIs, and drug products are determined by evaluation of their quality parameters with time under the influence of a variety of environmental factors such as temperature, humidity, and light.

Stability applies to chemical, physical, microbiological, therapeutic, and toxicological properties. The U.S. Pharmacopeia (USP) [1] defined *stability* as the extent to which a product retains, within specified limits and throughout its period of storage and use (i.e., its shelf life), the same properties and characteristics that it possessed at the time of manufacture. For example, chemical stability implies that each active ingredient in a drug product retains its chemical integrity and label potency within the specified limits during the shelf life. Physical stability is assured if the original physical properties, including appearance, palatability, uniformity, dissolution, and suspendability, are retained during the shelf life. Microbiological stability is demonstrated if the sterility or resistance to microbial growth is retained according to the specified requirements and antimicrobial agents that are present retain effectiveness within the specified limits. The therapeutic effect of the drug should remain unchanged and no significant increase in toxicity should occur [1] during the shelf life if the product is stable.

Good manufacturing practice (GMP) regulations and guidelines allude to the importance of stability and shelf life of pharmaceutical products. The U.S. Food and

Drug Administration's (FDA's) GMP regulations for finished products require stability testing of drug products to assess the stability characteristics, the results of which are used for determining appropriate storage conditions and expiration dates [2]. The FDA's GMP guidance for APIs requires that a documented, on-going testing program should be established to monitor the stability characteristics of APIs and the results should be used to confirm the appropriate storage conditions and retest or expiry dates [3]. The USP guidance on GMPs for bulk pharmaceutical excipients suggests that there should be a documented testing program designed to assess the stability characteristics of excipients and the results of such stability testing should be used in determining appropriate storage conditions and their reevaluation (retest) or expiration date [4]. The GMP guideline for pharmaceutical products from the World Health Organization (WHO) [5] alludes to the stability of finished pharmaceutical products and starting materials.

In addition to GMP requirements, several guidelines are published specifically to address pharmaceutical stability. Organizations with international mandates such as the WHO [6] and International Conference on Harmonization [ICH, comprising principally the European Union (EU), Japan, and the United States] [7] have published stability guidelines. The stability guidance of the WHO [6] defines stability data package requirements for pharmaceutical products. In addition to addressing APIs and drug products, the WHO guidance also emphasizes that the stability of excipients that may contain or form reactive degradation products should be considered for stability testing. The ICH Q1A (R2) guidance published by the FDA [7] states that the purpose of stability testing is to provide evidence on how the quality of a drug substance or drug product varies with time under the influence of a variety of environmental factors such as temperature, humidity, and light and to establish a retest period for the drug substance or a shelf life for the drug product and recommend storage conditions. The ICH has published a comprehensive series of stability guidances covering different aspects of stability: for example, ICH Q1A (R2), Stability Testing of New Drug Substances and Products [7]; ICH Q1B, Photostability Testing of New Drug Substances and Products [8]; ICH Q1C, Stability Testing of New Dosage Forms [9]; ICH Q1D, Bracketing and Matrixing Designs for Stability Testing of New Drug Substances and Products [10]; and ICH Q1E, Evaluation of Stability Data [11]. Some of these guidelines have been adopted and published by the FDA (see References).

In this chapter, the stability and shelf life of pharmaceutical excipients, APIs, and finished drug products are reviewed, with information derived from GMP regulations and guidelines as well as technical and regulatory guidance documents on stability. The importance of photostability studies for stability-indicating method development and for appropriate container closure selection to assure stability of the marketed product through shelf life is discussed. The evaluation and assessment of stability results for the determination of shelf life are discussed. The use of validated stability-indicating analytical methods for assessment of the potency of drug substance and the amount of degradation products and impurities through the shelf life is emphasized. Recommended storage conditions for various climatic zones are discussed. The testing parameters/attributes for evaluation of stability of various drug dosage forms are outlined. (The mechanisms and pathways of degradation of pharmaceutical products are not discussed in this chapter.)

## 2 STABILITY REQUIREMENTS IN GMP REGULATIONS/GUIDELINES

### 2.1 Finished Products

GMPs require assessment of stability characteristics of the pharmaceutical drug product, and the principles of such assessment are outlined in 21 CFR Part 211, Section 211.166 (a)(b) [2], and summarized in this section. The drug product stability samples should be stored in the same container closure as that in which the drug product is marketed. Appropriate storage conditions for samples designated for stability evaluation and reliable, specific, and meaningful test methods should be used. If a drug product is reconstituted at the time of dispensing as directed on the label, testing of the material for compliance with stability specifications is required pre- and post-reconstitution. Results from stability testing should be used to determine the appropriate storage conditions and expiration dates. The product stability estimates should be valid based on statistical evaluation of each stability attribute. Shelf life, expiration date, and storage conditions assigned for a drug product should be supported by long-term stability storage and testing of an adequate number of batches. When drug applications are submitted to the FDA, long-term storage stability studies may be in progress and data may not be available for the full proposed shelf life duration to support shelf life, expiration dating, or storage conditions. In such cases, tentative expiration dates can be established based on the accelerated stability studies, which are designed to increase the rate of chemical degradation or physical change by using exaggerated temperature and humidity conditions. In addition, other basic stability information on known physical and chemical stability attributes of excipients and APIs and container closure systems is also relevant for shelf life determination. Tentative expiration dates thus established must be supported by long-term storage stability studies.

### 2.2 Excipients

The GMPs for bulk pharmaceutical excipients [4] emphasize that the stability of excipients is an important contributing factor to the stability of the finished dosage form. Changes in raw materials used or changes in manufacturing procedures of excipients may affect their stability. In addition, the container closure packaging for excipients may vary widely to include but not be limited to metal and plastic drums, plastic bottles, and tank cars, which can affect the stability of the product inside. The stability characteristics of the excipient should be assessed using appropriate storage conditions and storage intervals and the testing should be performed to determine compliance with the established stability specifications. Results of such testing should be used to determine appropriate storage conditions and reevaluation or retest dates. Samples should be examined at or before retest date to ensure that the material is still in compliance with the specification and thus is suitable for its intended use. The excipient stability testing program should be ongoing and should be documented through a stability protocol that will include number of lots tested per year, sample size, test intervals, storage conditions (e.g., temperature, humidity), and test methods that are stability indicating. The materials used for container closures for stability sample storage (if the containers and closures are smaller than

those used for the marketed product) should be similar to the packaging materials used for marketed products. Storage time and conditions should support the conditions that are likely to prevail during the designated shelf life of the product. A model approach can be used for excipients that are available in different grades such as various molecular weights of a polymer or different monomer ratios or mixtures or blends of other excipients. Stability data available on model products can be used to determine the theoretical stability for similar products. In addition, for excipients that have been on the market for a long time, historical data may be used to assign the shelf life, retest dates, and storage conditions [4].

### 2.3 Active Pharmaceutical Ingredients

GMPs for APIs [3] require a documented stability storage and testing program to monitor the stability characteristics of APIs against established stability specifications. The results from stability studies are used to establish storage conditions and retest or expiry dates. For APIs, retest rather than expiration date terminology is commonly used. The GMP guideline also emphasizes that the assay procedures (for drug substance and degradation products) in stability samples should be validated and it should be stability indicating. A stability-indicating assay is defined as "a validated quantitative analytical procedure that can detect the changes with time in the pertinent properties of the drug substance and measures the active ingredient accurately without interference from degradation products, process impurities or other potential impurities" [12; Section III C].

The GMP API guideline also emphasizes that stability samples should be stored in containers and closures (the sum of packaging components that together contain and protect API or drug product dosage forms) that simulate the market container and should be consistent with storage conditions specified in the FDA stability guidance [7]. The GMP guideline requires that the first three commercial production batches should be placed on the stability program to confirm the retest or expiry date. If the API is expected to remain stable for at least two years, fewer than three batches are allowed. The stability program should be ongoing with at least one batch of API manufactured each year tested annually to confirm stability, with testing done more frequently for APIs with shorter shelf lives.

The WHO GMPs for pharmaceutical products [5] require that quality control evaluate the stability of finished pharmaceutical products and, when necessary, of starting materials and intermediate products. The guideline requires that a written stability program be developed and implemented to generate the stability data. Expiration dates and shelf life specifications should be based on stability data generated under defined storage conditions. The WHO guideline also specifies that stability should be determined prior to marketing and following any significant change in, for example, process, equipment, and packaging materials.

## 3 STABILITY REQUIREMENTS FOR EXCIPIENTS

In the absence of specific stability guidance from the FDA and ICH (and limited guidance from the WHO) for excipients, the USP GMP guidance for bulk pharmaceutical excipients [4] is used to review stability requirements for excipients. A

description of stability requirements for excipients is provided in Section 2.2. The principles underlying the evaluation of stability data and determination of shelf life (retest dates) for excipients is similar to that described for APIs and drug products in Section 7.

## 4 STABILITY REQUIREMENTS FOR DRUG SUBSTANCES (APIs)

The guidance for industry Q1A (R2) [7] defines requirements for API stability data package submission with drug applications in the EU, Japan, and the United States.

### 4.1 Selection of Batches and Container Closure System

Stability data should be generated on at least three primary batches, which should be manufactured to a minimum of pilot scale by the same synthetic route and manufacturing process as the production batches. The quality of the API placed on a formal stability program should be similar to the quality of the material to be made on a commercial production scale. The container closure system must be the same or simulate the packaging proposed for storage and distribution of marketed product.

### 4.2 Storage Conditions and Testing Frequency

The guidance ICH Q1A (R2) [7] provides information on storage conditions and testing frequency for APIs under four intended storage conditions: (1) general cases, (2) Intended for storage in a refrigerator, (3) Intended for storage in a freezer, and (4) drug products intended for storage below –20°C (for this storage no specific guidance is provided except to be treated on case-by-case basis). The storage conditions are summarized in Table 1.

***Marketed API Intended for Room Temperature Storage Conditions*** The storage conditions defined in the guidance document [7] as a *general case* has no definition assigned in the guidance. In this chapter, the author is interpreting this term as requiring room temperature storage (vs. storage in refrigerator and freezer) conditions for the marketed API. The storage conditions (temperature and humidity) and lengths of storage for stability studies should be sufficient to cover the following stages after manufacture: storage, shipment, and subsequent use. For submission of drug application, the data for long-term storage should cover a minimum of 12 months on three primary batches. The testing should, however, continue for the duration of the proposed shelf life and retest period. The accelerated and, when necessary, intermediate (storage condition designed to moderately increase rate of chemical degradation or physical change) storage conditions should be carried out for 6 months. In this scenario, long-term storage can be conducted at either at 25°C ± 2°C; 60% relative humidity (RH) ± 5% RH or 30°C ± 2°C, 65% RH ± 5% RH. If the long-term stability study is conducted at 30°C ± 2°C, 65% RH ± 5% RH, no intermediate condition storage is required. If long-term storage studies are conducted at 25°C ± 2°C, 60% RH ± 5% RH and a significant change occurs in the test

**TABLE 1 Storage Conditions for Stability Evaluation of APIs**

| Stability Study Type | Stability Storage Conditions | Minimum Time Period Covered by Data at Submission (months) |
|---|---|---|
| *Marketed API Intended for Room Temperature (General Case) [7] Storage Conditions* | | |
| Long term | 25°C ± 2°C, 60% RH ± 5% RH or 30°C ± 2°C, 65% RH ± 5% RH | 12 |
| Intermediate | 30°C ± 2°C, 65% RH ± 5% RH | 6 |
| Accelerated | 40°C ± 2°C, 75% RH ± 5% RH | 6 |
| *Marketed API Intended for Storage in Refrigerator* | | |
| Long term | 5°C ± 3°C | 12 |
| Accelerated | 25°C ± 2°C, 60% RH ± 5% RH | 6 |
| *Marketed API Intended for Storage in Freezer* | | |
| Long term | –20°C ± 5°C | 12 |

*Source*: From ref. 7.

result at any time during the 6 months testing at the accelerated storage condition (40°C ± 2°C, 75% RH ± 5% RH), testing for established stability specifications at the intermediate storage condition (30°C ± 2°C, 65% RH ± 5% RH) should be conducted and evaluated against significant change criteria. The guidance document [7] defines *significant change* for an API as failure to meet its specification. If intermediate storage condition and testing become a necessity, a minimum of 6 months of data from this study should be submitted with the application.

The sampling frequency and testing in long-term stability studies should be targeted to generate data sufficient to establish a stability profile for the API. The guidance [7] recommends testing every 3 months over the first year, 6 months over the second year, and annually thereafter throughout the retest period under long-term storage condition. For a 6-month accelerated storage stability condition, sampling at 0, 3, and 6 months is recommended. When significant change to established test specification occurs under accelerated storage condition, sampling at time 0, 6, 9, and 12 is recommended for a 12-month intermediate storage condition.

***Marketed API Intended for Storage in Refrigerator*** For an API intended for storage in a refrigerator (5°C ± 3°C), if a significant change occurs between three and six months testing at the accelerated storage condition of 25°C ± 2°C, 60% RH ± 5% RH, the proposed retest period should be based on the real-time data available at the long-term storage condition of 5°C ± 3°C [7]. On the other hand, if a significant change occurs within three months at the accelerated storage condition, the effect of short-term excursions outside the label storage condition during shipping or handling should be discussed. When significant change occurs during the first three months storage at accelerated condition, consideration of further storage or testing at this condition is not necessary. However, testing on a single batch of API for a period shorter than three months, with more frequent sampling during

three months may be necessary to narrow the period in which the API can be demonstrated to be stable [7].

***Marketed API Intended for Storage in Freezer*** For an API with intended storage in a freezer (–20°C ± 5°C), the retest period should be based on the real-time data available at the long-term storage condition of –20°C ± 5°C [7]. Since no accelerated storage condition for a freezer-stored API is proposed [7], storage and testing of a single batch at elevated temperatures of 5°C ± 3°C or 25°C ± 2°C for an appropriate time period should be considered to understand the effect of short-term excursions outside the label storage condition during shipping or handling. Other storage conditions may be considered with appropriate justification for their selection.

### 4.3 Stress Studies and Stability-Indicating Methods for Analysis of API Stability

Appropriate physical, chemical, biological, and microbiological attributes of the API that are likely to be susceptible to change during storage and are likely to influence the quality, safety, and efficacy should be tested. The analytical testing performed to evaluate the stability of the API should be stability indicating as defined above under GMPs for APIs [3; Section 11.5]. The stability-indicating assay accurately measures the active ingredient without interference from degradation products, process impurities, excipients, or other potential impurities. In order to develop a stability-indicating method, it is necessary to conduct stress studies on an API. Stress testing is carried out on a single batch of the drug substance and the stresses can include acid (e.g., 0.2 *N* HCl) and base (e.g., 0.2 *N* NaOH) hydrolysis, temperature (10°C increments above the accelerated stability storage temperature of 40°C, e.g., 50°C, 60°C), humidity (75% RH or greater), photolysis (see Section 6), and oxidation (e.g., 10% $H_2O$) of the API [7]. Hydrolysis is a chemical transformation process whereby an organic molecule RX reacts with water ($H_2O$), resulting in direct replacement of X by OH. Among various stresses stated above, hydrolysis could be the most common stress leading to degradation of an API since amides, esters, and salts of weak acids and strong bases are well known to hydrolyze [13]. Photolysis is the process whereby chemicals are altered directly as a result of irradiation or indirectly through interaction with products of irradiation [13]. Carbonyl, nitroaromatic, N-oxide function, C—C double bond, weak C—H bond, suplfides, alkenes, polyenes, and phenols are functional groups that may react with light [14]. Oxidation reactions depend on several factors, including temperature, light, pH, oxygen concentration, impurities including metal ions (e.g., cupric, ferric), and the oxidizable component of the molecule [15].

The stress studies should demonstrate that impurities and degradants from the active ingredient do not interfere with the quantitation of the API [12]. Stress testing of the API, in addition to validating the stability-indicating power of the analytical method, can also help establish the degradation pathways and the intrinsic stability of the molecule [7].

### 4.4 Evaluation of API Stability Results

When the stability data for all three batches on stability at various sampling intervals show little variability from initiation of the stability study, the API can be considered

stable and no statistical analysis of the data are required. Where the data show API degradation with time, the time at which the 95%, one-sided confidence limit for the mean curve intersects the acceptance criteria should be defined. When the analysis shows small batch-to-batch variation, combining the data into an overall estimate is recommended by applying appropriate statistical tests to the slopes of regression lines and zero time intercepts for the individual batches. The overall retest period should be based on the minimum time a batch can be expected to remain within acceptance criteria. Appropriate statistical methods should be considered to test the goodness of the fit of the data on all batches and where applicable combined batches to the assumed degradation line or curve [7]. The stability evaluation should cover API content, levels of degradation products, and other appropriate attributes.

### 4.5 Stability Commitment

When the submission includes long-term stability studies on three batches of API covering the proposed retest period, a postapproval commitment is unnecessary. If, on the other hand, when at the time of approval long-term stability data for the primary batches do not cover the proposed retest period granted, a commitment should be made to continue the stability studies postapproval to firmly establish the retest period. If the submission includes data from stability studies on fewer than three production batches, a commitment should be made to continue these studies through the proposed retest period and to place additional production batches for a total of three and generate data through the proposed retest period. If the submission does not include stability data on the production batches, a commitment should be made to place the first three production batches on long-term stability studies through the proposed retest period.

### 4.6 Storage Statement and Labeling

A storage statement for the finished API should be prepared in compliance with the national and regional requirements. The established retest period, which is supported by the stability data, should be displayed on the certificate of analysis (COA) and as appropriate on the container label also.

## 5 STABILITY REQUIREMENTS FOR DRUG PRODUCTS

The guidance for industry Q1A (R2) [7] defines requirements for stability data package submission with drug applications in the EU, Japan, and the United States for drug products. Information on the chemistry of the API molecule and its degradation behavior from the stability studies of the API (see Section 4) should be useful in designing the stability program for the drug product, since one of the key measures of shelf life determination of the drug product is the stability of the API in the drug product formulation.

### 5.1 Selection of Batches and Container Closure System

Stability data should be generated on at least three primary batches of the drug product, with the manufacturing process simulating the production batches, with

formulation and quality specifications similar to those batches intended for marketing. The three batches of the drug product manufactured should use different batches of the drug substance where possible, at least two of three batches should be pilot scale (batch manufactured by a procedure fully representative of and simulating that applied to a full production-scale batch), and the third batch can be of the smaller size when justified [7].

The drug product for the stability studies should be packaged in the same container closure system as proposed for marketing of the drug product, and each individual strength and container size of the proposed packaging configuration should be placed on stability, unless bracketing and matrixing designs are used in compliance with ICH guidance for stability testing [10].

### 5.2 Bracketing and Matrixing

During the design of stability studies, bracketing and matrixing [10] may be used to achieve reduced testing while at the same time generating enough stability data for evaluation of shelf life.

In bracketing, the design may include reduction in storage and sampling of dosage strengths or container closure configuration. For example, in a three-batch stability study with dosage strengths of 50, 75, and 100 mg in 15-, 100-, and 150-mL high-density polyethylene (HDPE) containers, testing for 50- and 100-mg strengths in 15- and 150-mL container sizes may be adequate with no testing proposed for the 75-mg strength.

Matrixing design [10] may involve elimination of some stability sample pull time points to achieve reduced testing strategy. For example, a one-half reduction in time points eliminates one in every two time points from full study design, and one-third reduction eliminates one in every three time points. However, such a scenario must include full testing at initial, 12-month, and final time points under a 36-month shelf life study [10].

### 5.3 Storage Conditions and Testing Frequency

The drug product guidance Q1A (R2) [7] provides information on storage conditions and testing frequency for the drug products under six intended storage conditions: (1) general case (room temperature), (2) drug products packaged in impermeable containers, (3) drug products packaged in semipermeable containers, (4) drug products intended for storage in a refrigerator, (5) drug products intended for storage in a freezer, and (6) drug products intended for storage below –20°C (for this storage no specific guidance is provided except to be treated on a case-by-case basis). The storage conditions are summarized in Table 2.

***Marketed Drug Product Intended for Room Temperature Storage Conditions*** The storage condition defined in the guidance document [7] as a *general case* has no definition assigned in the guidance and in this chapter is interpreted as room temperature storage (other than refrigerator and freezer) condition for the marketed drug product. The storage conditions (temperature and humidity) and lengths of storage for stability studies should be sufficient to cover the following stages after manufacture: storage, shipment, and subsequent use.

**TABLE 2 Storage Conditions for Stability Evaluation of Drug Products**

| Stability Study Type | Stability Storage Conditions | Minimum Time Period Covered by Data at Submission (months) |
|---|---|---|
| *Marketed Drug Product Intended for Room Temperature Storage Conditions* | | |
| Long term | 25°C ± 2°C, 60% RH ± 5% RH or 30°C ± 2°C, 65% RH ± 5% RH | 12<br>12 |
| Intermediate | 30°C ± 2°C, 65% RH ± 5% RH | 6 |
| Accelerated | 40°C ± 2°C, 75% RH ± 5% RH | 6 |
| *Marketed Drug Product Packaged in Semipermeable Containers* | | |
| Long term | 25°C ± 2°C, 40% RH ± 5% RH or 30°C ± 2°C, 35% RH ± 5% RH | 12 |
| Intermediate | 30°C ± 2°C, 65% RH ± 5% RH | 6 |
| Accelerated | 40°C ± 2°C, no more than 25% RH | 6 |
| *Marketed Drug Product Intended for Storage in Refrigerator* | | |
| Long term | 5°C ± 3°C | 12 |
| Accelerated | 25°C ± 2°C, 60% RH ± 5% RH | 6 |
| *Marketed API Intended for Storage in Freezer* | | |
| Long term | –20°C ± 5°C | 12 |

*Source*: From ref. 7.

For submission of drug application, the data for long-term storage should cover a minimum of 12 months on at least three primary batches. The primary batches are those used in a formal stability study from which the stability data are derived and submitted in a registration application for the purpose of establishing a retest period or shelf life/expiration date. The testing should, however, continue for the duration of the proposed shelf life and retest period. The accelerated and, when necessary, intermediate storage conditions should be carried out for 6 months. In this scenario, long-term storage can be conducted at either 25°C ± 2°C, 60% RH ± 5% RH or 30°C ± 2°C, 65% RH ± 5% RH. If the long-term stability study is conducted at 30°C ± 2°C, 65% RH ± 5% RH, no intermediate condition storage is required. If long-term storage studies are conducted at 25°C ± 2°C, 60% RH ± 5% RH and a significant change occurs in the test result at any time during the 6 months testing at the accelerated storage condition (40°C ± 2°C, 75% RH ± 5% RH), testing for established stability specifications at the intermediate storage condition (30°C ± 2°C, 65% RH ± 5% RH) should be conducted and evaluated against significant change criteria (Table 2).

The guidance document [7] defines *significant change* for drug product as one or more of the following: (1) a 5% change in assay from initial value or failure to meet the acceptance criteria for potency when using biological or immunological procedures; (2) any degradation product exceeding its acceptance criterion; (3) failure to meet the acceptance criteria for appearance, physical attributes, and functionality test (e.g., color, phase separation, resuspendability, caking, hardness, and dose delivery per actuation; under accelerated storage conditions, some changes in physical attributes may be expected, e.g., softening of suppositories and melting and possible phase separation in of creams); (4) failure to meet acceptance criterion for pH; and (5) failure to meet acceptance criteria for dissolution for 12 dosage units.

If intermediate storage and testing become a necessity, a minimum of 6 months of data from this study should be submitted with the application.

The sampling frequency and testing in long-term stability studies should be targeted to generate data sufficient to establish a stability profile of the drug product. For long-term storage conditions, the guidance [7] recommends testing every 3 months over the first year, 6 months over the second year, and annually thereafter through the proposed shelf life for drug products when the proposed shelf life is at least 12 months.

For a 6-month accelerated storage stability condition, sampling at 0, 3, and 6 months is recommended. When significant changes to established test specifications are likely to occur under the accelerated storage condition, increased testing is required with inclusion of a fourth sampling point. When significant changes to established test specifications occur under the accelerated storage condition, testing at the intermediate storage condition for 12 months with sampling at time 0, 6, 9, and 12 is recommended.

***Drug Products Packaged in Impermeable Containers*** For drug products packaged in impermeable containers that provide a permanent barrier, moisture or solvent loss is not a concern and for such products stability studies can be conducted under any controlled or ambient humidity conditions.

***Drug Products Packaged in Semipermeable Containers*** Stability studies for aqueous-based drug products packaged in semipermeable containers (containers that allow the passage of solvent, usually water, while preventing solute loss) should be conducted under conditions of low relative humidity and temperatures specified in Table 2. Stability attributes such as potential water loss and physical, chemical, biological, and microbiological stability should be evaluated.

If long-term storage studies are conducted at 25°C ± 2°C, 40% RH ± 5% RH and a significant change other than water loss occurs during the six months testing at the accelerated storage condition (40°C ± 2°C, 75% RH ± 5% RH), testing for established stability specifications at the intermediate storage condition (30°C ± 2°C, 65% RH ± 5% RH) should be conducted to evaluate the effect of 30°C (Table 2). While a significant change in water loss (5% water loss from initial value) alone under accelerated storage conditions need not prompt testing of samples under the intermediate storage condition, water loss through the proposed shelf life should be monitored to ensure that the drug product has no significant water loss during long-term storage at 25°C ± 2°C, 40% RH ± 5% RH.

***Marketed Drug Product Intended for Storage in Refrigerator*** For a drug product intended for storage in a refrigerator, if a significant change occurs between three and six months testing at the accelerated storage condition of 25°C ± 2°C, 60% RH ± 5% RH, the proposed retest period should be based on the real-time data available at the long-term storage condition of 5°C ± 3°C. On the other hand, if a significant change occurs within three months at the accelerated storage condition, the effect of short-term excursions outside the label storage condition during shipping or handling should be discussed. When significant change occurs during the first three months storage at the accelerated condition, consideration of further storage or testing is not necessary. However, testing on a single batch of drug product for a period shorter than three months with more frequent sampling during the three months may be necessary to narrow the period in which the drug product can be demonstrated to be stable.

***Marketed Drug Products Intended for Storage in Freezer*** For a drug product with intended storage in a freezer, the retest period should be based on the real-time data available at the long-term storage condition of –20°C ± 5°C. Since no accelerated storage condition for a freezer-stored API is proposed, storage and testing of a single batch at elevated temperatures of 5°C ± 3°C or 25°C ± 2°C for an appropriate time period should be considered to understand the effect of short-term excursions outside the label storage condition during shipping or handling. Other storage conditions may be considered with appropriate justification for selection.

### 5.4 Stress Studies and Stability-Indicating Methods for Analysis of Drug Product Stability

Appropriate physical, chemical, biological, and microbiological attributes, antimicrobial preservative and antioxidant content, and dosage functionality test (e.g., dose delivery system) of the drug product that are likely to be susceptible to change during storage and are likely to influence quality, safety, and efficacy should be tested. The analytical testing performed to evaluate the stability of the drug product should be validated and stability indicating [12]. The stability-indicating assay accurately measures the active ingredient in drug product without interference from degradation products, process impurities, excipients, or other potential impurities. In order to develop a stability-indicating method, it is necessary to conduct stress studies. Stress testing (studies undertaken to elucidate the intrinsic stability of the drug substance) can be carried out on the drug product similar to that described for APIs in Section 4. Degradation information obtained from stress studies for the active ingredient in the drug product should demonstrate the specificity of the assay, that is, that impurities and degradants from the active ingredient and drug product excipients do not interfere with the quantitation of the active ingredient in the drug product [12].

### 5.5 Evaluation of Stability Results

Data on stability of the drug products should be presented for all testing intervals and evaluated with physical, chemical, and microbiological, microbial preservative effectiveness, antioxidant effectiveness, and functionality tests as appropriate to the

drug product dosage form and all established specifications for the attributes being considered.

When the stability data for all three batches on stability at various sampling intervals show little variability from initiation of the stability study, the drug product can be considered stable and no statistical analysis of the data is required. Where the data show change in quantitative attributes (API and degradation amounts, dissolution rates), the time at which the 95% one-sided confidence limit for the mean curve intersects the acceptance criteria should be defined. When the analysis shows small batch-to-batch variation, combining the data into an overall estimate is recommended by applying appropriate statistical tests to the slopes of regression lines and zero-time intercepts for the individual batches. The overall shelf life should be based on the minimum time during the stability study a drug product can be expected to remain within acceptance criteria. Appropriate statistical methods should be considered to test the goodness of fit of the data on all batches and where applicable combined batches to the assumed degradation line or curve [7]. The stability evaluation should cover the levels of degradation products and other appropriate attributes. The mass balance, that is, whether the addition of assay value and degradation products add up to 100% of the initial value, should be considered taking into account the margin of analytical error of the method used. Details on stability evaluation are discussed in Section 7.

### 5.6 Stability Commitment

When the submission includes long-term stability studies on three batches of drug product covering the proposed shelf life, a postapproval commitment is unnecessary. If, on the other hand, when at the time of approval long-term stability data for the primary batches of drug product do not cover the proposed shelf life granted, a commitment should be made to continue the long-term stability studies (and six-month accelerated stability study if not performed), postapproval, to firmly establish the shelf life. If the submission includes data from stability studies on fewer than three production batches (batches manufactured at production scale by using production equipment in a production facility as specified in the application), a commitment to continue long-term studies through proposed shelf life and accelerated studies for six months for a total of three batches should be made. If the submission does not include stability data on the production batches, a commitment should be made to place the first three production batches on long-term stability studies through the proposed shelf life and accelerated studies for six months. When significant change occurs in the accelerated stability study on the primary batches (requiring a study under intermediate conditions), stability of commitment batches can be conducted at either the accelerated or intermediate condition. On the other hand, if a significant change occurs for the commitment batches under the accelerated condition, testing at the intermediate storage condition is also required for commitment batches. A formal stability commitment protocol should be in place for primary and commitment batches [7].

### 5.7 Storage Statement and Labeling

A storage statement for the finished drug product should be established in compliance with the national and regional requirements, which should be based on the

conclusions derived from stability evaluation. For drug products that cannot tolerate freezing, specific instructions should be provided. The established expiration date (shelf life) should be displayed on the COA and as appropriate on the container label [7].

## 6 PHOTOSTABILITY STUDIES

The stability guidance for drug substances and drug products [7] suggests that photostability testing should be an integral part of stress testing. The guidance on photostability testing [8] suggests that intrinsic stability of new drug substances and drug products should be evaluated to demonstrate that light exposure does not result in unacceptable change. The guidance [8] also recommends a systematic approach to stability testing, including on drug substance by direct exposure, and tests on the exposed drug product outside the the immediate (primary) pack. In the immediate (primary) pack, the packaging material is in contact with the drug substance or drug product and also includes any label associated with the primary pack. If change after exposure of drug product outside the immediate pack is not acceptable, testing on the immediate pack should be conducted. If change in the immediate pack exposed to light is not acceptable, photostability testing on the marketing pack (combination of immediate pack and other secondary packaging such as carton) should be conducted. If the change in the marketed pack exposed to light is not acceptable, the packaging configuration should be redesigned or the product may require reformulation and photostability testing performed under the redesigned scenario.

The photostability guidance [8] provides recommendations for the light sources to which the API or drug product should be exposed. Two light exposure options are suggested. Option 1 includes light output similar to internationally recognized standards for outdoor light (D65) or indoor indirect light (ID 65), defined in the Internation Organization for Standardization (ISO) 10977. The light sources covered under option 1 are the artificial daylight fluorescent lamp with a combination of ultraviolet (UV) and visible outputs, xenon, or metal halide lamps. In option 2, it is recommended that the sample should be exposed to both the cool white fluorescent and near-UV lamp. The cool white fluorescent lamp should comply with the outputs specified in ISO 10977. It is recommended that the near-UV fluorescent lamp having a spectral distribution from 320 to 400 nm with a maximum energy emission between 350 and 370 nm be used with a significant portion of UV in both bands, that is, 320–360 and 360–400 nm. For confirmatory studies, the samples should be exposed to light providing an overall illumination of not less than 1.2 million lux hours and an integrated near-UV energy of not less than 200 $Wh/m^2$. Use of a validated chemical actinometric system is recommended to ensure that the samples are exposed to desired light exposure by exposing actinometer solutions to light side by side with the drug substance or drug product samples. Light exposure also should be measured using calibrated radiometers or lux meters. Dark controls (covered with an aluminum foil to protect from light) should also be placed side by side with the light-exposed samples. The guidance provides important references on chemical actinometers [16], interlaboratory studies by industry and FDA [17], and perspectives from FDA scientists [18] on photostability testing of pharmaceutical products. In addition, a reference on forced degradation testing (including but not limited to light stress) of ibuprofen bulk drug and tablets by Farmer et al. [19] illustrates the

differences in degradation of ibuprofen as neat API and ibuprofen in association with excipients in drug product. A reference is also provided to illustrate how different light sources and intensities effect the degradation of a light-sensitive compound valerophenone [20].

### 6.1 Photostability of APIs

Photostability studies of a drug substance can provide an overall evaluation of the sensitivity of the drug substance to light. Because the exposure levels under options 1 and 2 described above can facilitate worst-case exposure (forced degradation or light stress) of a drug substance, extensive degradation is likely to be seen, especially if the drug substance has functional groups that are sensitive to photooxidation. If degradation of APIs occurs during photostability study, that may present an opportunity for development of a stability-indicating method. Photostability of the drug substance may be carried out on its neat form or in solutions with samples placed in chemically inert and transparent containers for exposure (e.g., quartz glass). Depending on the known photosensitivity of the drug substance (i.e., based on the functional groups that are susceptible to oxidation or known absorbance since compounds in the 290–800-nm range are likely to undergo degradation [13]), the intensity of the light exposure and length exposed can be controlled to achieve the desired degradation in a forced degradation study. It is also important to partition the temperature effects on degradation of the drug substance from light effects since elevated temperatures seen under light exposure may influence degradation independent of light exposure. Appropriate temperature controls should be used. Changes in physical state such as melting, sublimation, and evaporation should be minimized. Use of quartz glass containers are recommended for photostability studies. An appropriate amount of solid APIs can be spread in a container with thickness of approximately 3 mm. Liquids can be exposed in appropriate containers that are closed tightly. Appropriate dark controls (containers covered with aluminum foil and placed side by side along the exposed samples) should be used. Analysis of samples should include both the exposed and dark controls and the analysis should be performed concomitantly. Parameters examined can include physical properties such as appearance and color of solids and, if a solution is exposed, the clarity and color of solution. The assay and degradation products should be evaluated using validated stability-indicating methods that clearly separate the degradation products and impurities from the drug substance in the sample chromatograms.

The confirmatory studies with drug substance should be designed to identify precautionary measures needed to protect the API from light during formulation or manufacturing of the drug product, especially if the drug substance is found to degrade under light exposure. These studies also should help in the design of primary and secondary packaging material for commercial packing of the API.

### 6.2 Photostability of Drug Products

Drug product should be exposed to light conditions as stated above. The testing should be done in a sequential manner with fully exposed drug product and as necessary followed by testing the product in the immediate pack and then in the

marketing pack [8]. Results of the photostability study should demonstrate that the drug product is adequately protected from light during the shelf life. The testing should include one batch during the developmental phase followed by confirmation of the photostability characteristics using one batch. Additional two batches should be conducted if the results in the confirmatory study with one batch are similar to the developmental study. Confirmatory studies should establish the photostability characteristics of the drug product under standardized conditions. Results from these studies should help in (1) identification of precautionary measures needed during manufacture and packaging, (2) container closure design for protection from light, and (3) storage conditions and light protection required during shelf life of the marketed product. While keeping the light exposure constant during the exposure period, care should be taken to control the temperature to ensure that the degradation products observed are light exposure related. Physical characteristics of the samples under test conditions should be monitored to ensure that changes in the physical state of the drug product are minimized. Dark controls (unexposed controls) should be tested side by side with the light-exposed samples. Maximum light exposure under direct exposure for solid oral dosage forms such as tablets and capsules can be achieved by spreading them in a single layer. Quartz glass containers are recommended as containers for direct exposure. Based on the results obtained during direct exposure of the drug product, if immediate and marketing pack exposure becomes necessary, samples should be placed in such a manner as to facilitate uniform exposure. Key physical parameters to be examined should include appearance, clarity and color of solutions, dissolution, and disintegration. Quantitation of drug substance, degradation products of drug substance (known and unknown), impurities, and excipients present in the samples of drug product from photostability studies should be performed using validated stability-indicating methods. It is important to sample representative exposed or dark control samples to make a realistic assessment of the photostability of the various dosage forms. Solutions, suspensions, and creams should be examined to ensure that there is no settling or partitioning of phases and that uniform and homogenous samples are analyzed.

## 7 EVALUATION OF STABILITY DATA AND SHELF LIFE DETERMINATION

Requirements for stability data evaluation and shelf life determination are described in the guidance for industry—ICH Q1E [11], which covers evaluation of stability data that should be submitted in registration applications for new molecular entities and associated drug products. It also provides recommendations for the establishment of retest periods and shelf lives for drug substances and drug products intended to be stored at room temperature, in refrigerator or freezer, and below –20°C. In addition to the guidance document, studies by Grimm [21, 22] should be reviewed which laid the foundation for testing and evaluation of stability data.

### 7.1 Shelf Life Estimation for Drug Substances or Drug Products Intended for Room Temperature Storage

The stability data evaluation for drug substance or drug product intended for storage at room temperature should include evaluation of any significant change at

the various time points during accelerated stability study and at intermediate storage conditions when they are used and evaluation of long-term stability data. Different scenarios for retest period and shelf life determination are presented based on the results observed under accelerated, intermediate, and long-term storage conditions [11], as described below.

***Long-Term and Accelerated Stability Data Show Little or No Change over Time and Little or No Variability*** The drug substance or drug product can be considered stable if the long-term and accelerated stability data show little or no change or little no variability over time. In this scenario, no statistical analysis is required but a justification should include a discussion of the pattern of change or variability or lack of change or variability, supportive accelerated stability data, mass balance of drug substance in samples, and/or other supporting data. The retest period or shelf life in this scenario can be up to twice as long as but not more than 12 months beyond the period covered by long-term storage stability data. However, extrapolation of the retest period or shelf life beyond the period covered by long-term data can be proposed.

***Long-Term and Accelerated Stability Data Show Change over Time and/or Variability*** In this scenario, statistical analysis of the long-term data can be useful in establishing a retest period or shelf life. When there are differences, for example, among batches or among dosage strengths, container sizes, and/or fill or any combination thereof that preclude combining the data, the proposed retest date or shelf life should not exceed the shortest period supported by long-term studies of any batch, other factors, or combinations of factors. However, when differences are attributed, for example, to strength of the dosage form, different shelf lives can be assigned to different strengths within the dosage form of the drug product, with appropriate discussion of the data generated. If the statistical analysis is performed on long-term data and the statistical and other relevant data are supportive, it can be appropriate to extrapolate and propose a retest period or shelf life up to twice as long but not more than 12 months beyond the period covered by long-term storage stability data. When long-term stability data are not amenable to statistical analysis, the proposed retest period or shelf life can be up to one and a half as long as but not more than 6 months beyond the period covered by long-term data.

***Accelerated Stability Data Show Significant Change*** When a significant change occurs at the accelerated storage condition, stability results at the intermediate condition and the long-tem storage condition will determine the retest period or shelf life. The guidance [11] suggests two exceptions to the observations at accelerated conditions: (1) physical change such as softening of a suppository designed to melt at 37°C if the melting point is clearly demonstrated and (2) failure to meet acceptance criteria for dissolution for 12 units of gelatin capsules or gel-coated tablets if the failure can be unequivocally attributed to gelatin cross-linking. These exceptions do not apply to physical change as exemplified by phase separation of a semisolid dosage form, creams, or other occurrences at accelerated conditions, and if such a change occurs, testing at the intermediate condition should be performed.

When no significant change is observed at the intermediate condition, extrapolation beyond the period covered by long-term data can be proposed when statistical analysis data and relevant supporting data back up such extrapolation. The proposed retest period or shelf life can be up to one and a half as long as but not more than six months beyond the period covered by long-term data. When long-term stability data are not amenable to statistical analysis but the relevant supportive data are available, the proposed retest period can be up to three months beyond the period covered by long-term data.

In a scenario where significant change occurs at the intermediate condition, the retest period or shelf life shorter than the period covered by long-term data is appropriate but should not exceed the period covered by long-term data.

### 7.2 Shelf Life Estimation for Drug Substances or Drug Products Intended for Storage in a Refrigerator

***Long-Term and Accelerated Stability Data Show Little Change over Time and/or Variability*** Under this scenario, the proposed retest period or shelf life can be up to one and a half times as long as but not more than six months beyond the period covered by long-term data without the support of statistical analysis.

***Long-Term and Accelerated Stability Data Show Change over Time and/or Variability*** Under this scenario, the proposed retest period and shelf life can be up to one and a half times as long as but not more than six months beyond the period covered by long-term stability data when backed by statistical analysis and supporting data. On the other hand, the proposed retest period and shelf life can only be up to three months beyond the period covered by long-term stability data if the long-term data are amenable to statistical analysis but the statistical analysis is not performed or the long-term data are not amenable to statistical analysis but data supporting the proposed retest date and shelf life are available.

***Accelerated Stability Data Show Significant Change or Variability*** Under this scenario, if significant change occurs during accelerated stability study between three and six months, the proposed retest period or shelf life should be based on the long-term data, and extrapolation is not considered appropriate, but the retest period or shelf life shorter than that supported by long-term stability data is considered appropriate. In the case of variability in the long-term stability data, verification of the proposed retest period or shelf life by statistical analysis is considered appropriate. On the other hand, if significant change occurs during accelerated stability study within the first three months, the proposed retest period or shelf life should be based on the long-term data, and extrapolation is not considered appropriate, but the retest period or shelf life shorter than that supported by long-term stability data is considered appropriate. Conducting a stability study at accelerated conditions for shorter than three months should be considered under this scenario at least for one batch with more frequent sampling and analysis.

### 7.3 Shelf Life Estimation for Drug Substances or Drug Products Intended for Storage in a Freezer

In this scenario, the retest period or shelf life should be based on long-term data.

### 7.4 Shelf Life Estimation for Drug Substances or Drug Products Intended for Storage below –20°C

In this scenario, the retest period or shelf life should be based on long-term data.

## 8 ASSIGNMENT OF CLIMATIC ZONES AND RECOMMENDED STORAGE CONDITIONS

In June 2004, the FDA issued a guidance document to the industry on stability data packaging for registration applications in climatic zones III and IV [23] (adopted from the ICH guidance Q1F). The guidance provided an update to the stability storage conditions for climatic zones I and II and provided new guidance for climatic zones III and IV based on recommendations provided in the WHO stability guidance [24]. The WHO guidance [24] described stability testing recommendations and storage conditions for all four climatic zones, considering Grimm's [25] recommendation for climatic zones III and IV. The QIF guidance [23] clarified that the long-term storage condition for climatic zones I and II is 25°C ± 2°C, 60% RH ± 5% RH, and the intermediate storage condition for climatic zones I and II is 30°C ± 2°C, 65% RH ± 5% RH. It also stated [23] that the storage condition of 30°C ± 2°C, 65% RH ± 5% can also be a suitable alternative for the long-term storage condition of 25°C ± 2°C, 60% RH ± 5% RH for climatic zones I and II, in which case no intermediate condition is required. For climatic zones III and IV, the Q1F guidance [23] suggested using 30°C ± 2°C, 65% RH ± 5% (for products intended to be stored at room temperature; general case stated in the guidance) as the long-term storage condition (12 months) with accelerated storage at 40°C ± 2°C, 75% RH ± 5% (6 months), and further no intermediate storage condition was recommended for climatic zones III and IV. For aqueous-based drug products packaged in semipermeable containers, the guidance [23] recommended 30°C ± 2°C, 35% RH ± 5% RH for long-term (12 months) and 40°C ± 2°C and not more than 25% RH ± 5% RH for accelerated storage (6 months). The WHO Expert Committee on Specifications for Pharmaceutical Preparations in its fortieth meeting, held at Geneva, Switzerland, in October 2005, recommended to split the climatic zone IV into zone IVA with long-term storage condition remaining at 30°C, 65% RH, whereas for climatic zone IVB, the long-term storage condition recommended was 30°C, 75% RH [26]. In Annex 1 of working document QAS/06.179 (2006) [26], the WHO defined the five climatic zones and proposed long-term testing conditions for each zone, which are summarized in Table 3. Following the WHO guidance [26] publication in 2006, the FDA Q1F [23] was withdrawn in 2006, with future revision of Q1F possible.

## 9 STABILITY TESTING PARAMETERS FOR DIFFERENT DOSAGE FORMS

The testing parameters published by the WHO as Annex 2 to working document QAS/06.179 in 2006 [27] is the basis for the information summarized in this section for various dosage forms. Although the parameters listed here provide guidance for various dosage forms, pharmacopeial and those approved by regulatory authorities

**TABLE 3 Definition of Climatic Zones and Recommended Long-Term Stability Conditions**

| Climatic Zone | Definition | Criteria [Mean Annual Temperatures Measured in Open Air (°C) and Mean Annual Partial Vapor Pressure (hPa)] | Long-Term Testing Condition [Temperature (°C) and RH] |
|---|---|---|---|
| I | Temperate climate | ≤15°C, ≤11 hPa | 21°C, 45% RH |
| II | Subtropical and mediterranean climate | >15–22°C, >11–18 hPa | 25°C, 60% RH |
| III | Hot and dry climate | >22°C, ≤15 hPa | 30°C, 35% RH |
| IVA | Hot and humid climate | >22°C, >15–27 hPa | 30°C, 65% RH |
| IVB | Hot and very humid climate | >22°C, >27 hPa | 30°C, 75% RH |

*Source*: From ref. 26.

in drug applications should be considered in applying the information described below for various dosage forms.

### 9.1 Stability Tests Common to All Dosage Forms

Appearance, assay, and degradation products, preservative, and antioxidant content as applicable should be considered for all dosage forms. The microbial bioburden of sterile dosage forms must be controlled and tested in compliance with pharmacopeial and/or internal specifications. The microbial bioburden of nonsterile dosage forms should be controlled, with appropriate sampling and testing.

### 9.2 Stability Tests Specific to Dosage Forms

*Tablets* Dissolution (or disintegration, if justified), water content, hardness/friability.

*Hard Gelatin Capsules* Brittleness, dissolution (or disintegration, if justified), water content, and microbial bioburden.

*Soft Gelatin Capsules* Dissolution (or disintegration, if justified), microbial bioburden, pH, leakage, and pellicle formation.

*Emulsions* Phase separation, pH, viscosity, microbial bioburden, mean size and distribution of dispersed globules.

*Oral Solutions and Suspensions* Formation of precipitate, clarity for solutions, pH, viscosity, microbial bioburden, extractables, and polymorphic conversion when applicable. Additional tests for suspensions include redispersability, rheological properties, mean size, and distribution of particles.

*Powders for Oral Solutions and Suspensions* Water content, reconstitution time, and reconstituted solutions and suspensions should be tested as above for oral solutions and suspensions.

*Topical, Ophthalmic, and Otic Preparations* Clarity, homogeneity, pH, resuspendability (for lotions), consistency, viscosity, microbial bioburden, and water loss should be tested. For ophthalmic and otic products additional attributes should include sterility, particulate matter, and extractables.

*Suppositories* Softening range, dissolution at 37°C.

*Transdermal Patches* In vitro release rates, leakage, microbial bioburden/sterility, and peel and adhesive forces.

*Metered-Dose Inhalers and Nasal Aerosols* Content uniformity, aerodynamic particle size distribution, microscopic evaluation, water content, leak rate, microbial bioburden, valve delivery, extractables, leachables from plastic and elastomeric components.

*Nasal Sprays* Clarity, microbial bioburden, pH, particulate matter, unit spray medication content uniformity, droplet and/or particle size distribution, weight loss, pump delivery, microscopic evaluation of suspensions, particulate matter, extractables, leachables from plastic and elastomeric components of container closure and pump.

*Small-Volume Parenterals* Color, clarity of solutions, particulate matter, pH, sterility, endotoxins. Powders for injection solutions include clarity, color, reconstitution time and water content, pH, sterility, endotoxins/pyrogens, and particulate matter. Suspensions for injection should include additional particle size distribution, redispersability, and rheological properties. Emulsion for injection should include phase separation, viscosity, mean size, and distribution of dispersed globules.

*Large-Volume Parenterals* Color, clarity, particulate matter, pH, sterility, endotoxin/pyrogen, and volume.

## 10 SUMMARY AND CONCLUSIONS

Studies on the stability of pharmaceutical products provide information on how the quality of excipients, APIs, and drug products varies with time under the influence of various environmental factors such as temperature, humidity, and light and help determine shelf life and recommended storage conditions for the life cycle of products. Good manufacturing practice regulations and GMP guidelines require that stability studies on pharmaceutical products be conducted and shelf life be determined based on the results from stability studies. Accelerated stability studies at elevated temperature and humidity and long-term stability studies at more moderate temperature and humidity conditions are performed during drug development and approval process to predict the shelf life of pharmaceutical products. The proposed shelf life is then confirmed by performing long-term studies for the duration of the shelf life or longer. Stress studies using acid, base, temperature, oxidation, and light stresses are conducted to predict the degradation products that may be formed during the accelerated and/or long-term studies and to develop stability-indicating methods required for stability evaluation. Photostability and temperature studies also help determine the packaging configuration as well as make recommendations for storage conditions during the shelf life. For a realistic assessment of shelf life in diverse climates that exist around the world, five climatic zones are identified and

the long-term storage condition for shelf life determination in these climatic zones is based on humidity and temperature conditions likely to prevail in those climatic zones. Quality parameters for evaluation of stability of pharmaceutical products depend on the chemical nature of the active ingredient being studied as well as the type of dosage form of the drug product.

## REFERENCES


1. U.S. Pharmacopeia (2006), Stability considerations in dispensing practices, general chapter ⟨1191⟩, U. S. Pharmacopeial Convention, Rockville, MD.
2. U.S. Food and Drug Administration (FDA) (2007) Current good manufacturing practice regulations, 21 CFR Parts 210 and 211, FDA, Rockville, MD.
3. U.S. Food and Drug Administration (FDA) (2001), Guidance for industry, Q7A Good manufacturing practice guidance for active pharmaceutical ingredients, FDA, Rockville, MD.
4. U.S. Pharmacopeia (2007), Good manufacturing practices for bulk pharmaceutical excipients, general chapter ⟨1078⟩, U.S. Pharmacopeial Convention, Rockville, MD.
5. World Health Organization (WHO) (2003), Good manufacturing practices for pharmaceutical products: Main principles, WHO Technical Report Series 908, WHO, Geneva.
6. World Health Organization (WHO) (2006), Stability testing of active substances and pharmaceutical products, Working Document QAS/06.179, WHO, Geneva.
7. U.S. Food and Drug Administration (FDA) (2003), Guidance for industry, ICH Q1A (R2), Stability testing of new drug substances and products, FDA, Rockville, MD.
8. U.S. Food and Drug Administration (FDA) (1997), Guidance for industry, ICH Q1B, Photostability testing of new drug substances and products, FDA, Rockville, MD.
9. International Conference on Harmonization (ICH) (1997), Guidance for industry, ICH Q1C, Stability testing of new dosage forms, ICH, Brussels.
10. International Conference on Harmonization (ICH) (2003), Guidance for industry, ICH Q1D, Bracketing and matrixing designs for stability of drug substances and drug products, ICH, Brussels.
11. U.S. Food and Drug Administration (FDA) (2004), Guidance for industry, ICH Q1E, Evaluation of stability data, FDA, Rockville, MD.
12. U.S. Food and Drug Administration (FDA) (2000), Guidance for industry, Analytical procedures and methods validation—Chemistry, manufacturing, and controls documentation, FDA, Rockville, MD.
13. Lymann, W. J., Reehl, W. F., and Rosenblatt, D. H. (1982), *Handbook of Chemical Property Estimation Methods*, McGraw-Hill, New York.
14. Albini, A., and Fasani, E. (1998), *Photochemistry of Drugs: An Overview of Practical Problems, Drugs, Photochemistry and Photostability*, The Royal Society of Chemistry, London.
15. Carstensen, J. T. (1990), *Drug Stability—Principles and Practices*, Marcel Dekker, New York.
16. Drew, H. D., Brower, J. F., Juhl, W. E., and Thornton, L. K. (1998), Quinine photochemistry: A proposed chemical actinometer system to monitor UV exposure in photostability studies of pharmaceutical drug substances and drug products, *Pharmacopeial Forum*, 24(3), 6334.

17. Drew, H. D., Thornton, L. K., Juhl, W. E., and Brower, J. F. (1998), An FDA/PhRMA interlaboratory study of the International Conference on Harmonization's proposed photostability testing procedures and guidelines, *Pharmacopeial Forum*, 24(3), 6317.
18. Sager, N., Baum, R. G., Wolters, R. J., and Layloff, T. (1998), Photostability studies of pharmaceutical products, *Pharmacopeial Forum*, 24(3), 6331.
19. Farmer, S., anderson, P., Burns, P. K., and Velagaleti, R. (2002, May), Forced degradation of ibuprofen in bulk drug and tablets and determination of specificity, selectivity, and the stability-indicating nature of the USP ibuprofen assay method, *Pharm. Technol.*, 28–42.
20. Farmer, S., McCauslin, L., Burns, P. K., and Velagaleti, R. (2004, Aug.), Photosensitivity of internal standard valerophenone used in USP ibuprofen bulk drug and tablet assay and its effect on quantitation of ibuprofen and its impurities, *Pharm. Technol.*, 68–74.
21. Grimm, W. (1985), Storage conditions for stability testing—Long term testing and stress tests (part I), *Drugs Made in Germany*, 28, 196–202.
22. Grimm W. (1986), Storage conditions for stability testing—Long term testing and stress tests (part II), *Drugs Made in Germany*, 29, 39–47.
23. U.S. Food and Drug Administration (FDA) (2004), Guidance for industry, ICH Q1F, Stability data package for registration applications in climatic zones III and IV, FDA, Rockville, MD.
24. World Health Organization (WHO) (2001), Stability testing of pharmaceutical products containing well established drug substances in conventional dosage form, WHO Technical Report Series 863, Annex 5, WHO, Geneva.
25. Grimm W. (1998), Extension of International Conference on Harmonization tripartite guideline for stability testing of new drug substances and products to countries of climatic zones III and IV, *Drug Dev. Ind. Pharm.*, 24, 313–325.
26. World Health Organization (WHO) (2006), Assignment of climatic zones and recommended storage conditions, Working Document QAS/06.179, Annex 1, WHO, Geneva.
27. World Health Organization (WHO) (2006), Testing parameters (for dosage forms), Working Document QAS/06.179, Annex 2, WHO, Geneva.

# DRUG STABILITY

Nazario D. Ramirez-Beltran,[1] Harry Rodriguez,[2] and L. Antonio Estévez[1]

[1]*University of Puerto Rico, Mayagüez, Puerto Rico*
[2]*Cordis LLC, a Johnson & Johnson Company, San German, Puerto Rico*

**Contents**

1 Overview of Stability
- 1.1 Introduction
- 1.2 Regulatory Requirements
- 1.3 Stability and Shelf Life
- 1.4 Short- and Long-Term Stability Studies
- 1.5 Statistical Considerations

2 Design of Stability Studies
- 2.1 Introduction
- 2.2 Basic Design Considerations
- 2.3 Design of Stability Studies

3 Long-Term Stability Analysis
- 3.1 Introduction
- 3.2 Drug Shelf Life for Single Batch
- 3.3 Drug Shelf Life for Multiple Batches
- 3.4 Shelf Life Estimation for Multiple Factors

4 Short-Term Stability Analysis
- 4.1 Introduction
- 4.2 Chemical Reaction Kinetics
- 4.3 Degradation Data Evaluation at Accelerated Conditions
- 4.4 Preliminary Shelf Life Calculation from Stressed Data

5 Concluding Remarks

Appendix Computer Programs

References

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

# 1 OVERVIEW OF STABILITY

## 1.1 Introduction

Every drug product on the market requires the expiration date to be specified on the immediate container label [1]. Regulations from the U.S. Food and Drugs Administration (FDA) and the International Conference on Harmonization (ICH) require that pharmaceutical companies present factual evidence supporting the shelf life for either existing or new products. The presence of the right amount of the active ingredients in a pharmaceutical formulation is extremely important for the drug product to be an effective medicament. A rigorous protocol is usually implemented to measure the amount of the active ingredient through time for a given drug product and the collected data are analyzed to estimate the shelf life, which leads to the calculation of the expiration date. Thus, the expiration date provides the consumer the confidence that the drug product will retain its identity, strength, quality, and purity throughout the expiration period.

The main purpose of this chapter is to describe a set of practical tools to design a stability study as well as to discuss and illustrate how to determine the shelf life for a given drug product. This chapter is organized as follows. This section presents a general overview of the stability studies and considers the current regulatory requirements. The stability study not only characterizes the degradation of the active ingredient for a drug product over time but also provides the basis to establish the shelf-life period. This section will also present a brief description of short- and long-term stability studies and the important statistical requirements and issues stated in the FDA and ICH guidelines. Section 2 presents the fundamentals for designing a stability study. The design of a stability study intends to establish a procedure to extract reliable data to determine the shelf life, which is based on testing a limited number of batches of a drug product and will be applicable to all future batches of the drug product that will be manufactured under similar conditions. Designing long-term stability studies usually involves the determination of the following factors: number of batches, strength, packaging configurations, and storage conditions. Essentially, the stability process design includes the application of matrixing and bracketing design. Section 3 presents a description of the conventional procedure to determine the shelf life, which is established based on long-term stability studies that are conducted under normal storage conditions. Calculation of a drug product shelf life for single and multiple batches will be discussed and illustrated by examples. A set of computer programs are included in the Appendix to carry out the stability calculations. Section 4 discusses the accelerated stability testing procedure, which is often used to provide a tentative shelf life. At the early stage of a drug product development, the primary goal of the stability studies is to determine the rates of chemical and physical reactions and their relationships with storage conditions such as temperature, moisture, light, and others. A short-term stability study is conducted under stressed conditions to increase the rate of chemical and physical degradation of the drug product. This section will discuss the chemical reaction kinetics and the statistical procedure to establish a preliminary shelf life. This section will cover the regression techniques to estimate the parameters of the Arrhenius equation.

### 1.2 Regulatory Requirements

The FDA requires pharmaceutical industries to establish a stability testing program for each drug product [1–3]. Matthews [4] reviews the regulatory situation in Europe. The purpose of the stability program is to design the appropriate procedure to extract the assay data and calculate the shelf life to determine the expiration date of a drug product. The program should be described in a written protocol including all the requirements established by the regulation. The FDA establishes that the requirements to assess stability are the following [1]:

1. A sample size and test intervals for each attribute to be tested should be specified based on statistical rationale.
2. The storage conditions for the samples should be specified. The storage conditions should be the same as those specified in the drug product labeling.
3. All test methods used for each of the drug product attributes need to be qualified to demonstrate adequacy and reliability.
4. The stability samples should be stored in the same packaging configuration used for the final drug product to be marketed.
5. If the drug product needs to be reconstituted, testing of the drug product before and after reconstitution should be provided.
6. The number of batches to be used for the stability program for a new drug product is at least three. Different batches of the drug substance should be used.
7. The pharmaceutical industry should maintain records for all data, protocols, and reports related to the stability program.

### 1.3 Stability and Shelf Life

The shelf life is the period of time for which the drug product is assured to maintain its identity, strength, quality, and purity when stored at the conditions specified on the labeling. The expiration date marks the end of the shelf life period. The shelf life and the expiration date for a drug product are determined by carrying out a stability study. The stability of the drug product is assessed by testing all the attributes required to release the product to the market. The tests are carried out at specified time intervals for a determined period of time. The expiration date is usually included on the label of the drug product.

### 1.4 Short- and Long-Term Stability Studies

The short-term stability study is conducted over extreme environmental conditions to increase the rate of chemical degradation. The data obtained from short-term stability studies are normally used to evaluate longer term chemical effects at non-extreme storage conditions, but they are also helpful to assess the effect of short-term excursions outside the recommended storage conditions that might occur during shipping. However, the results from short-term stability studies cannot always

be used to predict physical changes of the drug product. The recommended testing frequency should be a minimum of three time points, including the initial and final time points, for a six-month study. If significant changes are expected, additional or more frequent testing should be conducted as well as stability testing at intermediate conditions. The actual extreme and intermediate conditions depend on the recommended storage conditions for drug product, for example, room temperature or under refrigeration. A significant change occurs if one or more of the following events take place:

1. A 5% change in assay from its initial value or failure to meet the acceptance criteria for potency when using biological or immunological procedures.
2. Presence of any impurity at a level exceeding its acceptance criterion.
3. Failure to meet an acceptance criterion for appearance, physical attribute, or functionality test (e.g., color, phase separation, resuspendibility, caking, hardness, or dose delivery per actuation). However, some changes in physical attributes (e.g., softening of suppositories, melting of creams) may be expected under accelerated conditions.
4. Failure to meet the acceptance criterion for pH.
5. Failure to meet the acceptance criteria for dissolution of 12 dosage units.

The long-term stability study is performed under the recommended storage conditions and packaging configuration to be used for marketing the drug product and during the shelf life period, which is displayed on the product label. The recommended testing frequency should be every three months over the first year, every six months over the second year, and annually thereafter. The general storage conditions used for long-term, intermediate, and short-term stability studies for climatic regions I and II, which includes Europe, Japan, and the United States [5, 6], are shown in Table 1. Alternate storage conditions for long-term, intermediate, or short-term stability study could be used if justified.

**TABLE 1 Storage Conditions for Stability Studies**

| Study Condition | Storage Conditions | Minimum Time Period Covered at Submission (Months) |
|---|---|---|
| | *Product Stored at Room Temperature* | |
| Long Term | 25°C ± 2°C, 60% RH ± 5% RH | 12 |
| Intermediate | 30°C ± 2°C, 60% RH ± 5% RH | 6 |
| Accelerated | 40°C ± 2°C, 75% RH ± 5% RH | 6 |
| | *Product Stored in a Refrigerator* | |
| Long Term | 5°C ± 3°C | 12 |
| Accelerated | 25°C ± 2°C, 60% RH ± 5% RH | 6 |
| | *Product Stored in a Freezer* | |
| Long Term | –20°C ± 5°C | 12 |

*Note*: RH, relative humidity.

## 1.5 Statistical Considerations

It is recommended to use an appropriate statistical method to analyze the data generated during a stability study. The purpose of using statistics is to establish, with a high degree of confidence, the shelf life, that is, the period during which a quantitative attribute of the drug product remains within acceptance criteria for all future batches manufactured, packaged, and stored under similar conditions. Stability studies are expensive and time consuming and statistics can certainly help in this aspect. Statistical design principles can be applied to reduce the amount of testing required [7].

The shelf life for a single batch is usually computed based on regression techniques. An appropriate approach to shelf life estimation when using regression analysis is by calculating the earliest time at which the 95% confidence limit for the mean intersects the proposed acceptance criterion [8]. A detailed description of shelf life calculations is provided in Sections 3 and 4.

The analysis of covariance is the conventional statistical tool to determine whether or not the degradation lines from several batches belong to a single population. If the degradation lines are statistically different, the minimum criterion is used to determine the shelf life for the current and future batches. According to the minimum criterion, the shelf life of future batches corresponds to the shortest of all shelf lives from the tested batches. On the other hand, if the batches belong to a single population, regression analysis is used to develop a single degradation line based on pooled data of the tested batches.

A short-term stability study is conducted under stressed conditions to increase the rate of chemical and physical degradation of the drug product. The classical statistical procedure to establish a preliminary shelf life is based on regression techniques, which are used to estimate the parameters of the Arrhenius equation.

A systematic approach for stability data evaluation should be used to determine whether extrapolation beyond the time period covered in the stability study for the long-term data is appropriate to calculate the shelf life period. The approach consists of evaluating any significant change at the accelerated condition and, if applicable, at the intermediate condition in order to follow a guideline on how to establish the shelf life period by extrapolation. The required relevant supporting data and the commitment to place stability batches to be tested until the end of the extrapolated shelf life should be provided to the regulatory agency. The relevant supporting data include satisfactory long-term data from development batches manufactured with close related formulation in a small scale and stored in a packaging configuration similar to the primary stability batches. The data evaluation approach is executed as follows [8].

***Product to be Stored at Room Temperature*** For products to be stored at room temperatures the extent of extrapolation will depend on the evaluation of the results from the short-term, intermediate, and long-term data. If no significant change at the accelerated (short-term) condition is observed, then the long-term data are evaluated for change and variability:

- *Long-Term Data with Small or No Change over Time and Small or No Variability* In this case a statistical analysis may be considered unnecessary

but a justification should be provided. Extrapolation can be proposed and the resulting shelf life can be up to twice the period covered in the stability study for the long-term data but not more than 12 months beyond the period tested.

- *Long-Term Data with Change over Time and Variability* In this case the statistical analysis of the long-term data can be useful to determine the shelf life of the drug product. If there is statistical difference among batches, among factors, or among factor combinations, it is not possible to perform data pooling and the shelf life will correspond to the shortest of all shelf lives obtained from all batches, factors, or factor combinations. If the statistical difference is related to a particular factor (e.g., strength or packaging), different shelf lives can be assigned to each level of that factor. The extent of extrapolation depends on whether statistical analysis is applicable to the stability data:
  1. Data for which statistical analysis does not apply: The proposed shelf life can be up to 1.5 times but should not be more than 6 months beyond the period covered in the stability study for the long-term data. This will require relevant supporting data to show that the drug product will meet all attribute acceptance criteria by the end of the proposed shelf life period.
  2. Data for which statistical analysis does apply: When statistical analysis is applicable to the long-term data and it is not performed, a justification is required and the extent of extrapolation is the same as when statistical analysis does not apply. If the statistical analysis is performed, the extent of extrapolation can be up to twice but not more than 12 months beyond the period covered in the stability study for the long-term data.

If significant change at any time during the 6-month period of the accelerated (short-term) condition is observed, then stability testing at an intermediate condition is required and the extent of extrapolation will depend on the outcome of both the long-term and intermediate conditions:

- *Data with No Significant Change at Intermediate Conditions* In this case extrapolation of the long-term data can be proposed and the extent of extrapolation will depend on whether a statistical analysis is applicable.
  - Data for which statistical analysis does not apply: The proposed shelf life can be up to 3 months beyond the period covered in the stability study for the long-term data. This will require relevant supporting data to show that the drug product will meet all the attribute acceptance criteria by the end of the proposed shelf life period.
  - Data for which statistical analysis does apply: When statistical analysis is applicable to the long-term data and it is not performed, a justification is required and the extent of extrapolation is the same as when statistical analysis does not apply. If the statistical analysis is performed, the extent of extrapolation can be up to 6 months beyond the period covered in the stability study for the long-term data. This will require relevant supporting data to show that the drug product will meet all the attribute acceptance criteria by the end of the proposed shelf life period.

- *Data with Significant Change at Intermediate Conditions* In this case extrapolation of the long-term data is not allowed and the shelf life should not exceed the period covered by the long-term stability study.

***Product to be Stored In a Refrigerator*** For products to be stored in a refrigerator the same data evaluation approach but with more restrictive extent of extrapolation as for product to be stored at room temperature will be used unless otherwise specified below.

If no significant change at the the accelerated (short-term) condition is observed, then the long-term data is evaluated for change and variability:

- *Long-Term Data with Small or No Change over Time and Small or No Variability* Extrapolation can be proposed and the resulted shelf life can be up to 1.5 times but should not be more than six months beyond the period covered in the stability study for the long-term data.
- *Long-Term Data with Change over Time and Variability* The extent of extrapolation depends on whether statistical analysis is applicable to the stability data:
  - Data for which statistical analysis does not apply: The proposed shelf life can be up to three months beyond the period covered in the stability study for the long-term data.
  - Data for which statistical analysis does apply: When statistical analysis applies and is not performed, a justification is required and the extent of extrapolation is the same as when statistical analysis does not apply. If the statistical analysis is performed, the extent of extrapolation can be up to 1.5 times but not more than six months beyond the period covered in the stability study for the long-term data.

If significant change at any time during the six-month period of the accelerated (short-term) condition is observed, then extrapolation is not considered appropriate. If the long-term data show variability, the proposed shelf life can be verified by statistical analysis.

***Product to be Stored in a Freezer*** For a drug product to be stored in a freezer the shelf life is determined from the long-term stability data and no extrapolation is allowed.

## 2 DESIGN OF STABILITY STUDIES

### 2.1 Introduction

The FDA guidelines establish, first, that the stability study protocol must describe how the stability study is designed and carried out and, second, the statistical methods to be used to analyze the data. The design of a stability study is aimed at establishing an expiration dating period. The expiration dating period is based on testing on a

limited number of batches of a drug product, and the results of these tests are applied to future batches of the drug product manufactured under similar conditions. Therefore, the study design should reduce the bias and identify and control any expected or unexpected source of variations. An appropriate stability design should provide the highest accuracy and precision of the established shelf life.

In this section, the fundamentals for designing a stability study are described. A stability study will characterize the degradation of the active ingredient with respect to time and is required to determine the shelf life period used to calculate the expiration date of a drug product. The shelf life is the maximum allowable period of time for the drug product to be stored in its final packaging while maintaining the therapeutic amount of the active pharmaceutical ingredient (API). The expiration date marks the end of the shelf life period. The expiration date is calculated by adding the shelf life to the manufacturing end date of the drug product. The determined shelf life is applicable to all future batches of the drug product manufactured under similar conditions as the batches used in the stability study.

The design of a stability study should be based on knowledge of the behavior and properties of the drug substance obtained from stability studies on the drug substance and data generated during clinical formulation studies. The stability study is performed by establishing and executing a protocol. The stability study protocol should specify all the aspects to be considered (e.g., sample size, test methods, and acceptance criteria) while carrying out the stability study. It is important to comply with all FDA (or ICH) regulations when performing a stability study.

Designing a stability study is based on a factorial design of experiments where a systemic procedure is used to determine the effect on the response variable of various factors and factor combinations. A linear model is used to represent the relationship between the factors and factor combinations with the response variable. Once the experimental design is established, the assays are conducted and stability data are saved to finally estimate the shelf life period.

Typically, the following factors are involved in the design of stability studies: number of batches, strength, package configuration, and storage conditions. The typical response variable of the stability study is the API amount. However, any other response variables (drug product attributes) that are susceptible to change during storage and are likely to influence quality, safety, and efficacy of the drug product (e.g., physical appearance, sterility, drug release rate, impurities, and degradation products) should also be considered. The testing procedure should include, as appropriate, the physical, chemical, biological, and microbiological attributes, preservative contents (e.g., antioxidant or antimicrobial preservatives), and functionality tests (e.g., for a dose delivery system).

Section 2 describes practical guidelines to design stability studies that are aligned with the guidance for industries provided by the FDA [9]. A different approach can be used depending on the characteristics of the drug product and with an appropriate scientific justification.

## 2.2 Basic Design Considerations

When designing a stability study, the following aspects should be taken into account:

*Preliminary Stability Data* During the drug product design phase, stability data are generated to choose the final package configuration and storage conditions and characterize the product. Such data are helpful in the development of a good stability study design that can offer as much information as possible with as few data as possible. Also, the preliminary stability data may provide scientific justification to the selection of the type of stability study design.

*Drug Product Attributes* All attributes that may affect the quality of the drug product have to be included in design of a stability study. These attributes have to be tested at every time period and, thus, the number of samples needed at each time should be set accordingly.

*Drug Product Specifications* The specification for each attribute of the drug product is required to define the acceptance criteria of the stability study.

*Test Methods* All test methods used in the stability study should be previously qualified.

*Preliminary Testing Data* Any data generated during the developmental and clinical trial phases of the drug may be useful to determine the expected manufacturing process variability. The process variability is an important factor to define the sampling plan to be used in the stability study.

*Design Factors* Identification of the design factors is crucial when choosing the design of the stability study. It is important to identify the most relevant design factors that may affect the stability of the drug product during storage. Failure to identify a relevant design factor may cause a significant delay in the stability study completion and submission to the regulatory agency, for example, the FDA.

*Full Design versus Reduced Design* A full design requires testing the drug product for factor combinations and at all time periods. The full design provides the largest amount of information to determine the shelf life of a drug product. However, the number of required combinations increases exponentially with the number of factors. On the other hand, a reduced design requires drug product testing only at a fraction of factor combinations. The reduced design requires less testing effort but entails the potential risk to result in a shorter shelf life than the one obtained from a full design due to the reduced amount of data collected.

## 2.3 Design of Stability Studies

***Full Stability Study Design*** Assume that a stability study includes three factors: batch, strength (denoted S1, S2, and S3 in what follows), and packaging size (denoted P1, P2, and P3). Each value of a factor is normally referred to as *level*. Therefore, if there are four packaging sizes, the factor called packaging size has four levels. It should be noted that the FDA requires at least three batches and consequently the batch factor will have three levels. If the other two factors have three levels each, then the required number of experiments is 27. Table 2 shows the 27 experiments that should be performed at each point in time. The number of combinations, $C$, can be readily calculated by multiplying the number of levels of each factor:

**TABLE 2 Factor Combinations for Three Factors with Three Levels Each**

| | S1 | | | S2 | | | S3 | | |
|---|---|---|---|---|---|---|---|---|---|
| Batch | P1 | P2 | P3 | P1 | P2 | P3 | P1 | P2 | P3 |
| **1** | C1 | C2 | C3 | C10 | C11 | C12 | C19 | C20 | C21 |
| **2** | C4 | C5 | C6 | C13 | C14 | C15 | C22 | C23 | C24 |
| **3** | C7 | C8 | C9 | C16 | C17 | C18 | C25 | C26 | C27 |

$$C = \mathrm{LF}_1 \times \mathrm{LF}_2 \times \mathrm{LF}_3 \times \cdots \times \mathrm{LF}_n$$

where $\mathrm{LF}_1$ = number of levels of factor 1
$\mathrm{LF}_2$ = number of levels of factor 2
$\mathrm{LF}_3$ = number of levels of factor 3
$\mathrm{LF}_n$ = number of levels of factor $n$

Once the factor combinations are defined, the number of samples required for the study can be determined. Assuming that the only testing required for the stability study design specified in Table 2 is amount of drug and that the test requires one package of the drug product, the amount of samples required for the study is 27 packages per time period. A drug product package for each of the 27 combinations has to be included in the study for each time period.

For a proposed shelf life of 36 months, testing of the drug product has to be carried out at the following time periods ($t$): 0, 3, 6, 9, 12, 18, 24, and 36 months. Considering the full design shown in Table 2, a total of 216 samples are needed to perform all required testing. The first 27 samples are tested at the beginning of the study to obtain data for $t = 0$. All other samples are placed in an environmental chamber that will maintain the temperature and relative humidity as specified in the stability study protocol. At each testing time, 27 additional samples are pulled from the chamber for testing, as shown in Table 3. The data analysis for stability studies will be discussed in section 3.

***Reduced Stability Study Designs*** There are occasions where it is not possible to obtain the total number of required samples to perform a full stability study design or it is simply desired to reduce the sampling requirement; this is done by selecting a reduced stability study design. To perform a reduced stability study design, some assumptions are needed which should be fully justified; the justifications should be provided to the regulatory agency. Usually, a reduced design applies when a preliminary stability study is performed.

Bracketing and matrixing are the two reduced designs recommended by the FDA [9]. Each of these methods applies to different situations. Using both of them simultaneously may reduce the ability of the study to determine the shelf life since factor combinations can be confounded due to the aliasing effect [10].

***Bracketing*** The bracketing design consists in testing only two levels: the highest and lowest values of the factor. Bracketing requires showing that the selected levels are the extremes of the factor range. If the stability of the extreme levels is different,

**TABLE 3 Testing Schedule for Full Stability Study Design**

| | By Time in Months | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Factor Combination | 0 | 3 | 6 | 9 | 12 | 18 | 24 | 36 |
| C1 | T | T | T | T | T | T | T | T |
| C2 | T | T | T | T | T | T | T | T |
| C3 | T | T | T | T | T | T | T | T |
| C4 | T | T | T | T | T | T | T | T |
| C5 | T | T | T | T | T | T | T | T |
| C6 | T | T | T | T | T | T | T | T |
| C7 | T | T | T | T | T | T | T | T |
| C8 | T | T | T | T | T | T | T | T |
| C9 | T | T | T | T | T | T | T | T |
| C10 | T | T | T | T | T | T | T | T |
| C11 | T | T | T | T | T | T | T | T |
| C12 | T | T | T | T | T | T | T | T |
| C13 | T | T | T | T | T | T | T | T |
| C14 | T | T | T | T | T | T | T | T |
| C15 | T | T | T | T | T | T | T | T |
| C16 | T | T | T | T | T | T | T | T |
| C17 | T | T | T | T | T | T | T | T |
| C18 | T | T | T | T | T | T | T | T |
| C19 | T | T | T | T | T | T | T | T |
| C20 | T | T | T | T | T | T | T | T |
| C21 | T | T | T | T | T | T | T | T |
| C22 | T | T | T | T | T | T | T | T |
| C23 | T | T | T | T | T | T | T | T |
| C24 | T | T | T | T | T | T | T | T |
| C25 | T | T | T | T | T | T | T | T |
| C26 | T | T | T | T | T | T | T | T |
| C27 | T | T | T | T | T | T | T | T |

*Note*: T = sample tested.

**TABLE 4 Factor Combinations for Bracketing of Three Factors with Three Levels**

| | S1 | | | S2 | | | S3 | | |
|---|---|---|---|---|---|---|---|---|---|
| Batch | P1 | P2 | P3 | P1 | P2 | P3 | P1 | P2 | P3 |
| **1** | C1 | — | C3 | — | — | — | C19 | — | C21 |
| **2** | C4 | — | C6 | — | — | — | C22 | — | C24 |
| **3** | C7 | — | C9 | — | — | — | C15 | — | C27 |

it is expected that the stability of any intermediate level should be higher than the stability of the least stable extreme. Table 4 shows the bracketing design for the full design example presented in Table 2.

Note that bracketing was not applied to the *batch* factor because the FDA regulation requires testing at least three batches to determine a drug product shelf life. Even so, the sampling required for the bracketing design was reduced substantially. The sample size required per time period is 12, a small number when compared to

**TABLE 5 Testing Schedule for Bracketing Stability Study Design**

| | By Time in Months | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Factor Combination | 0 | 3 | 6 | 9 | 12 | 18 | 24 | 36 |
| C1 | T | T | T | T | T | T | T | T |
| C2 | — | — | — | — | — | — | — | — |
| C3 | T | T | T | T | T | T | T | T |
| C4 | T | T | T | T | T | T | T | T |
| C5 | — | — | — | — | — | — | — | — |
| C6 | T | T | T | T | T | T | T | T |
| C7 | T | T | T | T | T | T | T | T |
| C8 | — | — | — | — | — | — | — | — |
| C9 | T | T | T | T | T | T | T | T |
| C10 | — | — | — | — | — | — | — | — |
| C11 | — | — | — | — | — | — | — | — |
| C12 | — | — | — | — | — | — | — | — |
| C13 | — | — | — | — | — | — | — | — |
| C14 | — | — | — | — | — | — | — | — |
| C15 | — | — | — | — | — | — | — | — |
| C16 | — | — | — | — | — | — | — | — |
| C17 | — | — | — | — | — | — | — | — |
| C18 | — | — | — | — | — | — | — | — |
| C19 | T | T | T | T | T | T | T | T |
| C20 | — | — | — | — | — | — | — | — |
| C21 | T | T | T | T | T | T | T | T |
| C22 | T | T | T | T | T | T | T | T |
| C23 | — | — | — | — | — | — | — | — |
| C24 | T | T | T | T | T | T | T | T |
| C25 | T | T | T | T | T | T | T | T |
| C26 | — | — | — | — | — | — | — | — |
| C27 | T | T | T | T | T | T | T | T |

*Note*: T = sample tested.

the 27 samples required for the full design. As a result, the amount of samples required for the whole stability study is $8 \times 12$, or 96, whereas the full-design study required 216 samples. The execution procedure of stability testing for both a complete and a reduced stability study is the same, as shown in Table 5.

***Matrixing*** The matrixing design consists in selecting a fraction of the total number of possible combinations included in the full design. At each time, a different fraction will be tested. The main assumption is that the stability of each fraction represents the stability of all factor combinations at that particular time. The matrixing design can be applied across the factors or across the time points of the stability design. The degree of reduction (e.g., one-half or one-third) from a full design depends on the quantity of factors to be considered. The larger the number of factors and levels included in the full design, the larger the degree of reduction that can be implemented.

Applying a matrixing design on time points only, all factor combinations (full factorial design) should be tested at the initial and final points in time, while a

fraction of the full factorial design is tested at intermediate points in time. If the full long-term stability data required for the proposed shelf life will not be available for review before submission to the regulatory agency, the full factorial design should be tested at 12 months and at the last scheduled point in time prior to submission. Therefore, the FDA regulations require testing full factorial design at 0, 12, and 36 months for a matrixing design. In addition, data from at least three time points, including initial, should be available for each factor combination through the first 12 months of the study. For matrixing at accelerated storage conditions, testing should occur at a minimum of three time points, including initial and final, for each factor combination.

The matrixing design should be as balanced as possible so that each factor combination is tested to the same extent over the duration of the study. A matrixing design is applicable when the supporting data indicate predictable product stability and small variability. A statistical justification for using a matrixing design could be based on evaluating the proposed matrixing design with respect to its power to detect differences in the degradation rates among the factors or the level of accuracy on estimating the shelf life.

A matrixing design performed across factors other than time points generally has less precision in shelf-life estimation and yields a shorter shelf life than the corresponding full design due to the confounding and aliasing effects [10]. Matrixing design may have insufficient power to detect main factors or factor interaction effects. Thus an excessive reduction in the number of factor combinations may produce an unreliable estimation of the shelf life due to the missing factor combinations. On the other hand, a matrixing design on time points would often have similar ability as the full design to detect differences in rates of change among factors and to establish a reliable shelf life. This is because full testing of all factor combinations would still be performed at both the initial and the last time points. For illustration purposes, matrixing design will be implemented over time points, factors, and in both time points and factors. The reader should be wise when choosing a design for a particular stability study and should have in mind that supporting data and justifications have to be provided.

***Matrixing in Time Points*** When matrixing in time points design, a fraction of the combinations is selected using the fractional factorial design procedure. Note that the full design presented above is equivalent to a $3^k$ factorial design, where $k$ is the number of factors and 3 is the number of levels of each factor. The total number of combinations for three factors and three levels is $3^3$, or 27. Assuming that a one-third reduction for the matrixing design is desired, the required sample size per time period is 18 compared to 27 for the full design. As a result, the number of samples needed for the whole stability study is $5 \times 18 + 3 \times 27$, or 171, whereas the full design required 216 samples. Table 6 shows the schedule to implement a matrixing design over time points for 36 months of stability study. This table shows a full design on times 0, 12, and 36 and two-thirds on the remaining times. The sampling shown in Table 6 has been designed to maintain a proper balance. If bracketing was applied to one of the factors (either strength or packaging size) or the design has only two levels on one of those factors, the full experiment design will be reduced, as shown in Table 7. Table 8 shows a two-thirds matrixing design with three strength

**TABLE 6 Testing Schedule for Matrixing in Time Point Stability Study Design**

| | By Time in Months | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Factor Combination | 0 | 3 | 6 | 9 | 12 | 18 | 24 | 36 |
| C1 | T | — | T | T | T | — | T | T |
| C2 | T | T | — | T | T | T | — | T |
| C3 | T | T | T | — | T | T | T | T |
| C4 | T | T | — | T | T | T | — | T |
| C5 | T | T | T | — | T | T | T | T |
| C6 | T | — | T | T | T | — | T | T |
| C7 | T | T | T | — | T | T | T | T |
| C8 | T | — | T | T | T | — | T | T |
| C9 | T | T | — | T | T | T | — | T |
| C10 | T | T | T | — | T | T | T | T |
| C11 | T | — | T | T | T | — | T | T |
| C12 | T | T | — | T | T | T | — | T |
| C13 | T | — | T | T | T | — | T | T |
| C14 | T | T | — | T | T | T | — | T |
| C15 | T | T | T | — | T | T | T | T |
| C16 | T | T | — | T | T | T | — | T |
| C17 | T | T | T | — | T | T | T | T |
| C18 | T | — | T | T | T | — | T | T |
| C19 | T | T | — | T | T | T | — | T |
| C20 | T | T | T | — | T | T | T | T |
| C21 | T | — | T | T | T | — | T | T |
| C22 | T | T | T | — | T | T | T | T |
| C23 | T | — | T | T | T | — | T | T |
| C24 | T | T | — | T | T | T | — | T |
| C25 | T | — | T | T | T | — | T | T |
| C26 | T | T | — | T | T | T | — | T |
| C27 | T | T | T | — | T | T | T | T |

*Note*: T = sample tested.

**TABLE 7 Factor Combination Considering Two Levels in Packaging**

| | S1 | | S2 | | S3 | |
|---|---|---|---|---|---|---|
| Batch | P1 | P2 | P1 | P2 | P1 | P2 |
| **1** | C1 | C2 | C7 | C8 | C13 | C14 |
| **2** | C3 | C4 | C9 | C10 | C15 | C16 |
| **3** | C5 | C6 | C11 | C12 | C17 | C18 |

levels and two packaging size levels. Thus, 18 factor combinations are scheduled for time equal to 0, 12, and 36 months and 12 combinations for the remaining time points.

For this matrixing design, the total number of samples needed for the whole stability study is $3 \times 18 + 5 \times 12$, or 114, compared to 216 samples needed for the full $3^k$ design. Furthermore, for the case where both strength and packaging size have

**TABLE 8 Testing Schedule for Matrixing in Time Points (Two Levels for Packaging Size)**

| | By Time in Months | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Factor Combination | 0 | 3 | 6 | 9 | 12 | 18 | 24 | 36 |
| C1 | T | — | T | T | T | — | T | T |
| C2 | T | T | T | — | T | T | T | T |
| C3 | T | T | — | T | T | T | — | T |
| C4 | T | — | T | T | T | — | T | T |
| C5 | T | T | T | — | T | T | T | T |
| C6 | T | T | — | T | T | T | — | T |
| C7 | T | T | T | — | T | T | T | T |
| C8 | T | T | — | T | T | T | — | T |
| C9 | T | — | T | T | T | — | T | T |
| C10 | T | T | T | — | T | T | T | T |
| C11 | T | T | — | T | T | T | — | T |
| C12 | T | — | T | T | T | — | T | T |
| C13 | T | T | — | T | T | T | — | T |
| C14 | T | — | T | T | T | — | T | T |
| C15 | T | T | T | — | T | T | T | T |
| C16 | T | T | — | T | T | T | — | T |
| C17 | T | — | T | T | T | — | T | T |
| C18 | T | T | T | — | T | T | T | T |

*Note*: T = sample tested.

**TABLE 9 Factor Combination Considering Two Levels in Packaging Size and Strength**

| | S1 | | S2 | |
|---|---|---|---|---|
| Batch | P1 | P2 | P1 | P2 |
| **1** | C1 | C2 | C7 | C8 |
| **2** | C3 | C4 | C9 | C10 |
| **3** | C5 | C6 | C11 | C12 |

only two levels or bracketing is applied to both factors, the full experiment design is reduced, as shown in Table 9. In this case, the total number of factor combinations is 12 rather than 27. Table 10 shows the design when bracketing and matrixing are implemented. Bracketing is applied to factors while matrixing is applied to time points. Table 10 shows that there are 12 factor combinations for times 0, 12, and 36 and 8 factor combinations on the remaining time points. Thus, for this matrixing design, the number of samples required for the whole stability study is $3 \times 12 + 5 \times 8$, or 76, compared to 216 samples needed for the full $3^k$ design.

***Matrixing in Factors*** For the matrixing-in-factors design, the factor combinations are eliminated in a systematic way as shown in Table 11. As a result, not all factor combinations are tested during the stability study. Such a design can be used when the factor combination eliminated exhibits a similar behavior as the other factor

**TABLE 10 Testing Schedule for Matrixing in Time Points (Two Levels for Packaging Size and Strength)**

| | By Time in Months | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Factor Combination | 0 | 3 | 6 | 9 | 12 | 18 | 24 | 36 |
| C1 | T | — | T | T | T | — | T | T |
| C2 | T | T | T | — | T | T | T | T |
| C3 | T | T | — | T | T | T | — | T |
| C4 | T | — | T | T | T | — | T | T |
| C5 | T | T | T | — | T | T | T | T |
| C6 | T | T | — | T | T | T | — | T |
| C7 | T | T | — | T | T | T | — | T |
| C8 | T | — | T | T | T | — | T | T |
| C9 | T | T | T | — | T | T | T | T |
| C10 | T | T | — | T | T | T | — | T |
| C11 | T | — | T | T | T | — | T | T |
| C12 | T | T | T | — | T | T | T | T |

*Note*: T = sample tested.

**TABLE 11 Matrixing in Factor Combination Elimination**

| | S1 | | | S2 | | | S3 | | |
|---|---|---|---|---|---|---|---|---|---|
| Batch | P1 | P2 | P3 | P1 | P2 | P3 | P1 | P2 | P3 |
| **1** | — | C2 | C3 | C10 | C11 | — | C19 | — | C21 |
| **2** | C4 | — | C6 | — | C14 | C15 | C22 | C23 | — |
| **3** | C7 | C8 | — | C16 | — | C18 | — | C26 | C27 |

combinations within the study. The implementation of this matrixing design is shown in Table 12. All factor combinations are tested at times 0, 12, and 36 months, and just 18 factor combinations at the remaining times. For this matrixing design, the number of samples needed for the whole study is $3 \times 27 + 18 \times 5$, or 171, while 216 samples were needed for the full $3^k$ design.

***Matrixing in Factors and Time Points*** Matrixing in both factors and time points is a combination of the two previously mentioned matrixing designs and the schedule is shown in Table 13. In this matrixing design, all factor combinations are tested at times 0, 12, and 36 and fractional factorial at the remaining times. The total number of experiments to be conducted in this stability study is $3 \times 27 + 5 \times 12$, or 141, while 216 samples were required for the full $3^k$ design.

## 3 LONG-TERM STABILITY ANALYSIS

### 3.1 Introduction

This section considers the stability analysis of the long-term studies. The purpose of this section is to provide a set of fundamental statistical tools to calculate the shelf life for single and multiple packages and strengths. The statistical methods will be

**TABLE 12 Testing Schedule for Matrixing in Factor Stability Study Design**

| | By Time in Months | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Factor Combination | 0 | 3 | 6 | 9 | 12 | 18 | 24 | 36 |
| C1 | T | — | — | — | T | — | — | T |
| C2 | T | T | T | T | T | T | T | T |
| C3 | T | T | T | T | T | T | T | T |
| C4 | T | T | T | T | T | T | T | T |
| C5 | T | — | — | — | T | — | — | T |
| C6 | T | T | T | T | T | T | T | T |
| C7 | T | T | T | T | T | T | T | T |
| C8 | T | T | T | T | T | T | T | T |
| C9 | T | — | — | — | T | — | — | T |
| C10 | T | T | T | T | T | T | T | T |
| C11 | T | T | T | T | T | T | T | T |
| C12 | T | — | — | — | T | — | — | T |
| C13 | T | — | — | — | T | — | — | T |
| C14 | T | T | T | T | T | T | T | T |
| C15 | T | T | T | T | T | T | T | T |
| C16 | T | T | T | T | T | T | T | T |
| C17 | T | — | — | — | T | — | — | T |
| C18 | T | T | T | T | T | T | T | T |
| C19 | T | T | T | T | T | T | T | T |
| C20 | T | — | — | — | T | — | — | T |
| C21 | T | T | T | T | T | T | T | T |
| C22 | T | T | T | T | T | T | T | T |
| C23 | T | T | T | T | T | T | T | T |
| C24 | T | — | — | — | T | — | — | T |
| C25 | T | — | — | — | T | — | — | T |
| C26 | T | T | T | T | T | T | T | T |
| C27 | T | T | T | T | T | T | T | T |

*Note*: T = sample tested.

described first under the context of a general application and then numerical examples will be described step by step to illustrate particular applications. A set of computer programs are provided in the Appendix to facilitate the implementation of the stability analysis procedures. Chen et al. [11] pointed out that the stability analysis usually consists of three steps. The first step is to collect the assay results at several time intervals for the sample stored under appropriate conditions. The second step is to select the appropriate model to describe the relationship between assay results and sampling times, and the third step is to establish the expiration dating period. The first step was described in Section 2 and the second and third steps are described in Section 3.

### 3.2 Drug Shelf Life for Single Batch

Calculation of the shelf life for a drug product in a single package type will be illustrated assuming that assay results are obtained from a single batch. The FDA guideline establishes that the expiration dating period for a drug product consists

**TABLE 13 Testing Schedule for Matrixing in Factor and Time Point Stability Study Design**

| | By Time in Months | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Factor Combination | 0 | 3 | 6 | 9 | 12 | 18 | 24 | 36 |
| C1 | T | — | T | T | T | T | T | T |
| C2 | T | T | — | T | T | — | T | T |
| C3 | T | — | — | — | T | — | — | T |
| C4 | T | — | — | — | T | — | — | T |
| C5 | T | T | T | — | T | T | — | T |
| C6 | T | — | T | T | T | T | T | T |
| C7 | T | T | T | — | T | T | — | T |
| C8 | T | — | — | — | T | — | — | T |
| C9 | T | T | — | T | T | — | T | T |
| C10 | T | T | — | T | T | — | T | T |
| C11 | T | — | — | — | T | — | — | T |
| C12 | T | — | T | T | T | T | T | T |
| C13 | T | T | T | — | T | T | — | T |
| C14 | T | — | T | T | T | T | T | T |
| C15 | T | — | — | — | T | — | — | T |
| C16 | T | — | — | — | T | — | — | T |
| C17 | T | T | — | T | T | — | T | T |
| C18 | T | T | T | — | T | T | — | T |
| C19 | T | — | — | — | T | — | — | T |
| C20 | T | — | T | T | T | T | T | T |
| C21 | T | T | — | T | T | — | T | T |
| C22 | T | — | T | T | T | T | T | T |
| C23 | T | — | — | — | T | — | — | T |
| C24 | T | T | T | — | T | T | — | T |
| C25 | T | T | — | T | T | — | T | T |
| C26 | T | T | T | — | T | T | — | T |
| C27 | T | — | — | — | T | — | — | T |

*Note*: T = sample tested.

of determining the time at which the 95% one-sided lower confidence interval for the mean degradation curve intersects the lower acceptable specification limit, which is usually adopted by the FDA as 90% of the label claim (LC). Assuming that concentration of a drug product decreases linearly with time and it can be expressed as

$$y_i = \alpha + \beta x_i + \varepsilon_i \qquad i = 1, \ldots, n \tag{1}$$

where $y_i$ is the percentage of the label claim (assay result) at a given time $x_i$ for the $i$th sample, $x_i$ is the time at which the $i$th sample was analyzed, and $\alpha$ and $\beta$ are the regression parameters. The coefficient $\alpha$ represents the percentage label claim when $x_i = 0$ and is usually known as the batch effect, while coefficient $\beta$ is known as the degradation rate, and the product $\beta x_i$ is the stability loss over time. It is assumed that the random variable $\varepsilon_i$ follows a normal distribution with zero mean and a constant variance $\sigma^2$ and $n$ is the total number of samples. In Section 4, the various

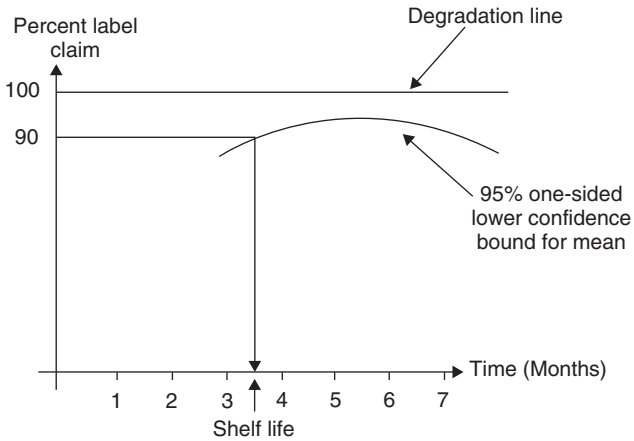


**FIGURE 1** Graphical representation of shelf life.

degradation kinetic models will be presented. In this context, Equation (1) describes a zero-order degradation.

The drug expiration date for a single batch exhibits a 95% confidence that the average drug characteristic of the dosage units in the batch is within specifications up to the end of the expiration date. The 95% one-sided lower confidence bounds for the mean degradation line is shown in Figure 1.

The 95% confidence interval can be expressed by the probability statement

$$P(t \le t_{0.05,n-2}) = 0.95 \tag{2}$$

where $t_{0.05,n-2}$ is the upper five percentile of the $t$ student distribution with $n - 2$ degrees of freedom. Assuming that the percentage of the label claim (%LC) follows a normal distribution, the statistic $t$ is defined as

$$t = \frac{z}{\sqrt{x_m^2/m}} = \frac{[\hat{y} - (\alpha + \beta x)] \Big/ \sqrt{\sigma^2\left[1/n + (x - \bar{x})^2 / S_{xx}\right]}}{\sqrt{\mathrm{SSE}/[\sigma^2(n-2)]}} \tag{3}$$

where

$$S_{xx} = \sum_{i=1}^{n} (x - \bar{x})^2 \qquad \mathrm{SSE} = \sum_{i=1}^{n} (y_i - \hat{y})^2 \tag{4}$$

where SSE is the sum-of-squares error, $z$ is a random variable that follows the standard normal distribution, $\chi_m^2$ is a random variable that follows the chi-square distribution with $m$ degrees of freedom, $\hat{y}$ is the estimated %LC at time $x$, $a$ and $b$ are the estimates of the parameters $\alpha$ and $\beta$, respectively, and $\bar{x}$ is the average sampling time.

The $t$ statistic was used to eliminate the unknown parameter $\sigma^2$. Thus, Equation (3) can be written as

$$t = \frac{\hat{y} - (\alpha + \beta x)}{S_{\hat{y}}} \tag{5}$$

$$S_{\hat{y}} = \sqrt{\mathrm{MSE}\left[\frac{1}{n} + \frac{(x - \bar{x})^2}{S_{xx}}\right]} \qquad \mathrm{MSE} = \frac{\mathrm{SSE}}{n-2} \tag{6}$$

where MSE is the mean-square error.

Thus, Equation (2) can be now written as

$$P\left(\frac{\hat{y} - (\alpha - \beta x)}{S_{\hat{y}}} \le t_{0.05,n-2}\right) = 0.95 \tag{7}$$

and after arranging terms, Equation (7) can be expressed as

$$P(\hat{y} - t_{0.05,n-2} S_y \le \alpha + \beta x) = 0.95 \tag{8}$$

Therefore, the 95% one-sided lower confidence bound is

$$L(x) = \hat{y} - t_{0.05,n-2} S_{\hat{y}} = a - bx - t_{0.05,n-2}\sqrt{\mathrm{MSE}\left[\frac{1}{n} + \frac{(x - \bar{x})^2}{S_{xx}}\right]} \tag{9}$$

The points where $L(x)$ intersects the acceptable lower specification limit $\delta$ can be obtained by finding the roots of the following equation: $\delta - L(x) = 0$. It should be pointed out that this equation can also be written in the form

$$f(x) = [\delta - (a + bx)]^2 - t^2_{0.05,n-2}\ \mathrm{MSE}\left[\frac{1}{n} + \frac{(x - \bar{x})^2}{S_{xx}}\right] = 0 \tag{10}$$

The quadratic equation (10) has two roots and the shelf life is obtained by computing the root of Equation (10) that is smaller than a reference point, which is defined as

$$x_{\mathrm{ref}} = \frac{b_0 - \delta}{-b_1} \tag{11}$$

A numerical example is presented below to illustrate, step by step, the shelf-life calculation procedure.

**Example 1 Shelf Life for Single Batch** The assay results for a batch of a drug product are given in Table 14. Based on these results, determine the shelf life for this batch.

**TABLE 14 Single-Batch Assay Results (%LC) for Example 1**

| Sampling time, months | 0 | 3 | 6 | 9 | 12 | 18 | 24 | 36 |
|---|---|---|---|---|---|---|---|---|
| Percent label claim | 103.5 | 97.3 | 97.2 | 94.2 | 94.8 | 94.1 | 91.3 | 88.7 |

***Solution*** First, regression techniques are used to estimate the expected degradation line [12]. This yields

$$\hat{y} = 99.64 - 0.3335x$$

It is well known that conventional statistical computer programs such as MINITAB, STATGRAPHICS, or SAS will conduct model fitting and provide the following additional information: $t_{0.05,n-2} = t_{0.05,6} = 1.943$, MSE = 4.1665, $n = 8$, $\bar{x} = 13.5$, and $S_{xx} = 1008$. For the reader's convenience, a computer program to perform model regression fitting and to compute all the statistics required to determine the shelf life is given in the Appendix.

Thus, the 95% one-sided lower confidence bound for the mean degradation rate is

$$L(x) = 99.64 - 0.3335x - 1.943\sqrt{4.1665\left[\frac{1}{8} + \frac{(x-13.5)^2}{1008}\right]}$$

Consequently, the shelf life can be obtained by using the equation

$$f(x) = [90 - (99.64 - 0.3335x)]^2 - 1.943^2(4.1665)\left[\frac{1}{8} + \frac{(x-13.5)^2}{1008}\right] = 0 \qquad (12)$$

To compute the shelf life, we compute first the reference point as

$$x_{\text{ref}} = \frac{b_0 - \delta}{-b_1} = \frac{99.64 - 90}{0.3335} = 28.90$$

Therefore, the shelf life is the root smaller than 28.90. A simple and practical tool to compute the roots of Equation (12) is perhaps solving the following equivalent problem. Find $x$ such that it minimizes the absolute value of $f(x)$. This root is obtained by using the quasi-Newton line search (QNLS) algorithm [13]. The computer program requires an initial point and we recommend using the value

$$x(0) = x_{\text{ref}} - d \qquad (13)$$

where $d$ is a positive value that should be explored by using a trial-and-error method until the program converges to a local minimum, usually the range of $d$ is between 0 and 10. For instance, when $d = 0$, the QNLS method converges to $x_R = 23.6989$. The shelf life $x_L$ is the integer part of the root $x_R$ and, in this case, the expiration dating period of this batch is $x_L = 23$ months since $f(x_R) = 0$, and $x_R < x_{\text{ref}}$. The QNLS method is implemented in MATLAB software and the computer program is also given in the Appendix.

### 3.3 Drug Shelf Life for Multiple Batches

***Test for Poolability of Batches*** The FDA guidelines establish that at least three batches must be used to determine the batch-to-batch variability. Thus, if the statistical procedure shows evidence that the three batches belong to the same population, a single shelf life for all batches will be determined by pooling data from all batches. The FDA guidelines also established that batch similarity of the degradation lines can be assessed by the equality of slopes and the equality of intercepts of individual batches.

Assuming that the degradation decreases linearly with time (zero-order degradation), it can be represented by the following model, which is analogous to Equation (1):

$$y_{ij} = \alpha_i + \beta_i x_{ij} + \varepsilon_{ij} \qquad i = 1, \ldots, I \qquad j = 1, \ldots, n_i \tag{14}$$

where $y_{ij}$ is the assay result (%LC) of the $i$th batch for a drug product sampled at time $x_{ij}$, $n_i$ is the number of sampling times for the $i$th batch, $I$ is the total number of batches, $\alpha_i$ and $\beta_i$ are the intercept and slope of the degradation line for the $i$th batch, respectively, and $\varepsilon_{ij}$ is assumed to be a random variable with zero mean and constant variance. It is worth mentioning that $\alpha_i$ is considered as the batch effect and $\beta_i$ as the degradation rate for the $i$th batch.

The FDA guidelines indicate that the tests for the equality slopes and the equality of intercepts should be performed at the 0.25 level of significance, as was suggested by Bancroft [14]. It would be desirable to test whether or not the batches belong to a single population and if that is the case to derive a model with a single intercept and slope for the entire population of batches. Thus, a poolability test is implemented to determine if a single population should be used. The poolability test is implemented in two steps: (1) testing for equality of slopes and (2) testing for equality of intercepts. The conventional procedure to test these hypotheses is accomplished by using the analysis of covariance for a completely randomized design [15, 16]. An alternative method that can easily be generalized for studying multifactor is using a regression model with indicator variables [17].

*Analysis of Covariance (ANCOVA) for Testing Similarity of Slopes* The first step is determining whether or not the degradation rates for all batches behave in a similar fashion. The following hypothesis will be tested:

$$H_0\text{:} \quad \beta_i = \beta_j \quad \text{for all } i \neq j, \quad i = 1, \ldots, I \qquad j = 1, \ldots, I \tag{15}$$

where $I$ is the number of batches.

To develop the $F$ statistics and be able to test the above hypothesis, it is required to compute different sum of squares and cross products for %LC and also for sampling times [15, 16, 18]. The aggregated sum of squares of the sampling times is defined as

$$Z_{xx} = \sum_{i=1}^{I} S_{xx}(i) \tag{16}$$

where

$$S_{xx}(i) = \sum_{j=1}^{n_i}(x_{ij} - \bar{x}_{i\bullet})^2 = \sum_{j=1}^{n_i} x_{ij}^2 - \frac{x_{i\bullet}^2}{n_i}$$
$$x_{i\bullet} = \sum_{j=1}^{n_i} x_{ij} \qquad \bar{x}_{i\bullet} = \frac{x_{i\bullet}}{n_i} \tag{17}$$

The aggregated sum of squares of %LC is defined as

$$Z_{yy} = \sum_{i=1}^{I} S_{yy}(i) \tag{18}$$

where

$$S_{yy}(i) = \sum_{j=1}^{n_i}(y_{ij} - \bar{y}_{i\bullet})^2 = \sum_{j=1}^{n_i} y_{ij}^2 - \frac{y_{i\bullet}^2}{n_i}$$
$$y_{i\bullet} = \sum_{j=1}^{n_i} y_{ij} \qquad \bar{y}_{i\bullet} = \frac{y_{i\bullet}}{n_i} \tag{19}$$

The aggregated sum of cross products between the %LC and sampling times is given as

$$Z_{xy} = \sum_{i=1}^{I} S_{xy}(i) \tag{20}$$

where

$$S_{xy}(i) = \sum_{j=1}^{n_i}(x_{ij} - \bar{x}_{i\bullet})(y_{ij} - \bar{y}_{i\bullet}) = \sum_{j=1}^{n_i} x_{ij} y_{ij} - \frac{x_{i\bullet} y_{i\bullet}}{n_i} \tag{21}$$

Thus, the sum-of-squares error based on the aggregated sums is computed as

$$\mathrm{SSE}_a = Z_{yy} - \frac{Z_{xy}^2}{Z_{xx}} \tag{22}$$

The aggregated sum-of-squares error for each batch, which is also needed, is calculated as

$$\mathrm{SSE} = \sum_{i=1}^{I} \mathrm{SSE}(i) \tag{23}$$

where

$$\mathrm{SSE}(i) = S_{yy}(i) - \frac{S_{xy}^2(i)}{S_{xx}(i)} \tag{24}$$

The mean-square errors due to slope contribution and the total mean-square errors are defined as

$$\mathrm{MS}_{\mathrm{slope}} = \frac{\mathrm{SSE}_a - \mathrm{SSE}}{I-1} \tag{25}$$

$$\mathrm{MSE} = \frac{\mathrm{SSE}}{N-2I} \qquad \text{where} \quad N = \sum_{i=1}^{I} n_i \tag{26}$$

Finally, the $F$ statistics to test hypothesis (15) can be written as

$$F(\mathrm{slope}) = \frac{\mathrm{MS}_{\mathrm{slope}}}{\mathrm{MSE}} \tag{27}$$

It should be noted that hypothesis (15) is rejected when $F(\mathrm{slope}) > F_{0.25,I-1,N-2I}$, where $F_{0.25,I-1,N-2I}$ is the upper percentile of the $F$ distribution with $I - 1$, and $N - 2I$ degrees of freedom. If hypothesis (15) is not rejected (i.e., the slopes are similar), it can be proceeded with the next step: to test whether or not the intercepts of the involved batches are similar. The computational procedure for this is described next.

*ANCOVA for Testing Intercept Similarities* The second step is to test whether or not the intercepts for the individual degradation lines are equal given that the degradation lines from the considered batches have similar slopes. The hypothesis of interest can be written as

$$H_0: \quad \alpha_i = \alpha_j \quad \text{for all } i \neq j \qquad i = 1, \ldots, I \qquad j = 1, \ldots, I \tag{28}$$

Since differences in slopes were not identified, model (14) reduces to a model with a single degradation rate as follows:

$$y_{ij} = \alpha_i + \beta x_{ij} + \varepsilon_{ij} \qquad i = 1, \ldots, I \qquad j = 1, \ldots, n_i \tag{29}$$

The intercepts can be decomposed into two elements: the common intercept and the batch effect. That is, $\alpha_i = \alpha + \tau_i$, where $\alpha = \mu - \beta\bar{x}_{..}$ is the common intercept and $\tau_i$ is the batch effect; $\mu$ is the expected value of $y_{ij}$, and $\bar{x}_{..}$ is the average of $x_{ij}$ and was defined by Equation (32). It should be noted that the deviation of each intercept from the common intercept is called the batch effect. Thus, model (29) can be written as

$$y_{ij} = \mu + \tau_i + \beta(x_{ij} - \bar{x}_{..}) + \varepsilon_{ij} \tag{30}$$

A model without a batch effect is compared with a model that includes the batch effect to be able to measure the intercept effect. Equation (30) includes the batch effect and will be called the complete model. The model with no batch effect will be called a reduced model and can be expressed as

$$y_{ij} = \mu + \beta(x_{ij} - \bar{x}_{\cdot\cdot}) + \varepsilon_{ij} \qquad i = 1, \ldots, I \qquad j = 1, \ldots, n_i \tag{31}$$

The sums of squares, cross products for totals, and errors for the reduced model are computed as [12, 18]

$$R_{xx} = \sum_{i=1}^{I}\sum_{j=1}^{n_i}(x_{ij} - \bar{x}_{\cdot\cdot})^2 = \sum_{i=1}^{I}\sum_{j=1}^{n_i}x_{ij}^2 - \frac{x_{\cdot\cdot}^2}{N} \qquad \bar{x}_{\cdot\cdot} = \frac{1}{N}\sum_{i=1}^{I}\sum_{j=1}^{n_i}x_{ij} \qquad x_{\cdot\cdot} = N\bar{x}_{\cdot\cdot} \tag{32}$$

$$R_{yy} = \sum_{i=1}^{I}\sum_{j=1}^{n_i}(y_{ij} - \bar{y}_{\cdot\cdot})^2 = \sum_{i=1}^{I}\sum_{j=1}^{n_i}y_{ij}^2 - \frac{y_{\cdot\cdot}^2}{N} \qquad \bar{y}_{\cdot\cdot} = \frac{1}{N}\sum_{i=1}^{I}\sum_{j=1}^{n_i}y_{ij} \qquad y_{\cdot\cdot} = N\bar{y}_{\cdot\cdot} \tag{33}$$

$$R_{xy} = \sum_{i=1}^{I}\sum_{j=1}^{n_i}(x_{ij} - \bar{x}_{\cdot\cdot})(y_{ij} - \bar{y}_{\cdot\cdot}) = \sum_{i=1}^{I}\sum_{j=1}^{n_i}x_{ij}y_{ij} - \frac{x_{\cdot\cdot}y_{\cdot\cdot}}{N} \tag{34}$$

where $N$ was defined in Equation (26).

The sum of squares for the complete model can be computed as

$$T_{xx} = \sum_{i=1}^{I}n_i(\bar{x}_{i\cdot} - \bar{x}_{\cdot\cdot})^2 = \sum_{i=1}^{I}\frac{x_{i\cdot}^2}{n_i} - \frac{x_{\cdot\cdot}^2}{N} \tag{35}$$

$$T_{yy} = \sum_{i=1}^{I}n_i(\bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot})^2 = \sum_{i=1}^{I}\frac{y_{i\cdot}^2}{n_i} - \frac{y_{\cdot\cdot}^2}{N} \tag{36}$$

$$T_{xy} = \sum_{i=1}^{I}n_i(\bar{x}_{i\cdot} - \bar{x}_{\cdot\cdot})(\bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot}) = \sum_{i=1}^{I}\frac{x_{i\cdot}y_{i\cdot}}{n_i} - \frac{x_{\cdot\cdot}y_{\cdot\cdot}}{N} \tag{37}$$

$$E_{xx} = \sum_{i=1}^{I}\sum_{j=1}^{n_i}(x_{ij} - \bar{x}_{i\cdot})^2 = R_{xx} - T_{xx} \tag{38}$$

$$E_{yy} = \sum_{i=1}^{I}\sum_{j=1}^{n_i}(y_{ij} - \bar{y}_{i\cdot})^2 = R_{yy} - T_{yy} \tag{39}$$

$$E_{xy} = \sum_{i=1}^{I}\sum_{j=1}^{n_i}(x_{ij} - \bar{x}_{i\cdot})(y_{ij} - \bar{y}_{i\cdot}) = R_{xy} - T_{xy} \tag{40}$$

The sum-of-squares errors for the reduced model can be computed as

$$\mathrm{SSE_r} = R_{yy} - \frac{R_{xy}^2}{R_{xx}} \tag{41}$$

The sum-of-squares errors for the complete model can be computed as

$$\mathrm{SSE_c} = E_{yy} - \frac{E_{xy}^2}{E_{xx}} \tag{42}$$

The mean-square errors for the intercept effect and the mean-square errors for the complete model can be written as

$$\mathrm{MSE_{int}} = \frac{\mathrm{SSE_r} - \mathrm{SSE_c}}{I-1} \qquad \mathrm{MSE_c} = \frac{\mathrm{SSE_c}}{N-I-1} \tag{43}$$

Finally, the $F$ statistics to measure whether or not the intercepts are different from batch to batch is given as

$$F(\mathrm{int}) = \frac{\mathrm{MSE_{int}}}{\mathrm{MSE_c}} \tag{44}$$

The hypothesis (28) is rejected if $F(\mathrm{int}) > F_{0.25,I-1,N-I-1}$, where $F_{0.25,I-1,N-I-1}$ is the upper percentile of the $F$ distribution with $I - 1$ and $N - I - 1$ degrees of freedom. Thus, if the null hypotheses (15) and (28) are not rejected at the 0.25 level of significance, the batches can be considered to come from a single population or pool and a single shelf life is computed based on the studied batches. If that is the case, model (14) reduces to the expression

$$y_{ij} = \alpha + \beta x_{ij} + \varepsilon_{ij} \qquad i = 1, \ldots, I \qquad j = 1, \ldots, n_i \tag{45}$$

where $\alpha$ and $\beta$ are the common intercept and slope, respectively, for model (45).

The approach to estimate the shelf life of a single batch can be applied to the pooled stability data from all batches. Thus, the shelf life is obtained by finding the minimum root of Equation (10), where $a$ and $b$ are the estimates of $\alpha$ and $\beta$ from model (45), respectively, $\bar{x}$ is the average sampling time, and $S_{xx}$ is defined by Equation (4) and considering for all sampling times.

If hypothesis (15) is not rejected and hypothesis (28) is rejected, model (14) reduces to the expression

$$y_{ij} = \alpha_i + \beta x_{ij} + \varepsilon_{ij} \qquad i = 1, \ldots, I \qquad j = 1, \ldots, n_i \tag{46}$$

Assuming a common slope with different intercepts for different batches, the expiration dating period is computed for each individual batch. The minimum of the expiration dating periods of the individual batches is the expiration dating period of the drug product.

A third alternative may occur if hypothesis (15) is rejected; it is concluded in this case that the batches do not belong to a single population and the shelf life should be computed for each individual batch. The minimum of the expiration periods of the individual batches is the shelf life of the drug product.

**Example 2 Slopes and Intercepts Are the Same** This example illustrates the application of the poolability test for batch similarity. Assay data from three batches, expressed as %LC, are shown in Table 15. Determine whether or not the three batches belong to a single population.

**TABLE 15 Assay Results (%LC) for Example 2**

| Sampling time, months | 0 | 3 | 6 | 9 | 12 | 18 | 24 | 36 |
|---|---|---|---|---|---|---|---|---|
| Batch 1 | 102.4 | 98.1 | 99.2 | 97.5 | 95.0 | 96.1 | — | — |
| Batch 2 | 101.1 | 101.2 | 99.0 | 97.2 | 96.4 | 95.5 | 94.3 | — |
| Batch 3 | 104.1 | 102.1 | 99.5 | 98.1 | 95.7 | 94.1 | 94.0 | 93.5 |

***Solution*** The poolability test is implemented in two steps.

**Step 1.** The first test will determine whether or not the slopes from the three batches are similar. From Table 15 the following information can be extracted: $I = 3$, $n_1 = 6$, $n_2 = 7$, $n_3 = 8$, and $N = 21$. The first hypothesis to be tested is

$$H_0: \quad \beta_1 = \beta_2 = \beta_3 = \beta \tag{47}$$

Using Equations (16)–(21), the sum of squares and cross products for %LC and sampling time are computed as follows:

$$S_{xx}(1) = \sum_{j=1}^{6} x_{1j}^2 - \frac{x_{1\cdot}^2}{6} = 3^2 + 6^2 + \ldots + 18^2 - \frac{48^2}{6} = 210$$

$$S_{xx}(2) = \sum_{j=1}^{7} x_{2j}^2 - \frac{x_{2\cdot}^2}{7} = 3^2 + 6^2 + \ldots + 24^2 - \frac{72^2}{7} = 429.43$$

$$S_{xx}(3) = \sum_{j=1}^{8} x_{3j}^2 - \frac{x_{3\cdot}^2}{8} = 3^2 + 6^2 + \ldots + 36^2 - \frac{108^2}{8} = 1008$$

$$S_{yy}(1) = \sum_{j=1}^{6} y_{1j}^2 - \frac{y_{1\cdot}^2}{6} = 102.4^2 + 98.1^2 + \ldots + 96.1^2 - \frac{588.3^2}{6} = 33.65$$

$$S_{yy}(2) = \sum_{j=1}^{7} y_{2j}^2 - \frac{y_{2\cdot}^2}{7} = 101.1^2 + 101.2^2 + \ldots + 94.3^2 - \frac{684.7^2}{7} = 43.75$$

$$S_{yy}(3) = \sum_{j=1}^{8} y_{3j}^2 - \frac{y_{3\cdot}^2}{8} = 104.1^2 + 102.1^2 + \ldots + 93.5^2 - \frac{781.1^2}{8} = 111.98$$

$$S_{xy}(1) = \sum_{j=1}^{6} x_{1j} y_{1j} - \frac{x_{1\cdot} y_{1\cdot}}{6} = 3(98.1) + 6(99.2) + \ldots + 18(96.1) - \frac{48(588.3)}{6} = -69.6$$

$$S_{xy}(2) = \sum_{j=1}^{7} x_{2j} y_{2j} - \frac{x_{2\cdot} y_{2\cdot}}{7} = 3(101.2) + 6(99) + \ldots + 24(94.3) - \frac{72(684.7)}{7} = -131.23$$

$$S_{xy}(3) = \sum_{j=1}^{8} x_{3j} y_{3j} - \frac{x_{3\cdot} y_{3\cdot}}{8} = 3(102.1) + 6(99.5) + \ldots + 36(93.5) - \frac{108(781.1)}{8} = -294.45$$

The total sums of squares are computed as

$$Z_{xx} = \sum_{i=1}^{3} S_{xx}(i) = 210 + 429.43 + 1008 = 1647.43$$

$$Z_{yy} = \sum_{i=1}^{3} S_{yy}(i) = 33.65 + 43.75 + 111.98 = 189.38$$

$$Z_{xy} = \sum_{i=1}^{3} S_{xy}(i) = -69.6 - 131.23 - 294.45 = -495.28$$

Using Equation (22), the aggregated sum-of-squares error is computed as

$$\mathrm{SSE_a} = Z_{yy} - \frac{Z_{xy}^2}{Z_{xx}} = 189.38 - \frac{495.28^2}{1647.43} = 40.48$$

The sum-of-squares errors for each batch is computed using Equation (24):

$$\mathrm{SSE}(1) = S_{yy}(1) - \frac{S_{xy}^2(1)}{S_{xx}(1)} = 33.65 - \frac{69.6^2}{210} = 10.58$$

$$\mathrm{SSE}(2) = S_{yy}(2) - \frac{S_{xy}^2(2)}{S_{xx}(2)} = 43.75 - \frac{131.23^2}{429.43} = 3.65$$

$$\mathrm{SSE}(3) = S_{yy}(3) - \frac{S_{xy}^2(3)}{S_{xx}(3)} = 111.98 - \frac{294.45^2}{1008} = 25.97$$

The aggregate sum-of-squares error is obtained using Equation (23):

$$\mathrm{SSE} = \sum_{i=1}^{3} \mathrm{SSE}(i) = 10.58 + 3.65 + 25.97 = 40.2$$

Thus, the mean-square error due to the slope is computed using Equation (25):

$$\mathrm{MS_{slope}} = \frac{\mathrm{SSE_a} - \mathrm{SSE}}{I - 1} = \frac{40.48 - 40.2}{3 - 1} = 0.14$$

The mean-square error for the regression model (14) is

$$\mathrm{MSE} = \frac{\mathrm{SSE}}{N - 2I} = \frac{40.2}{21 - 2(3)} = 2.68$$

Finally, the $F$ statistic for testing equal degradation rate [hypothesis (47)] is

$$F(\text{slope}) = \frac{\text{MS}_{\text{slope}}}{\text{MSE}} = \frac{0.14}{2.68} = 0.052$$

The critical value to test for equality on slopes is $F_{0.25,I-1,N-2I} = F_{0.25,2,15} = 1.52$. Since $F(\text{slope}) < F_{0.25,2,15}$, hypothesis (47) for equal degradation rate cannot be rejected at the 0.25 level of significance. It is concluded that there is no significant difference among the slopes of model (14), that is, the model represented by Equation (14) reduces to model (29).

**Step 2.** The second step consists in testing for equality of intercepts among the batches. To derive the appropriate statistic for testing the underlying hypothesis, it is required to compute the following sum and cross products for the %LC and the sampling time. The hypothesis to be tested is

$$H_0: \quad \alpha_1 = \alpha_2 = \alpha_3 = \alpha \tag{48}$$

The sum and cross products are computed by using equations (32)–(40):

$$R_{xx} = \sum_{i=1}^{3}\sum_{j=1}^{n_i} x_{ij}^2 - \frac{x_{\cdot\cdot}^2}{N} = 3^2 + 6^2 + \ldots + 36^2 - \frac{228^2}{21} = 1754.5$$

$$R_{yy} = \sum_{i=1}^{3}\sum_{j=1}^{n_i} y_{ij}^2 - \frac{y_{\cdot\cdot}^2}{N} = 102.4^2 + 98.1^2 + \ldots + 93.5^2 - \frac{2054.1^2}{21} = 189.97$$

$$R_{xy} = \sum_{i=1}^{3}\sum_{j=1}^{n_i} x_{ij} y_{ij} - \frac{x_{\cdot\cdot} y_{\cdot\cdot}}{N} = 3(98.1) + 6(99.2) + \ldots + 36(93.5) - \frac{228(2054.1)}{21} = -503.06$$

$$T_{xx} = \sum_{i=1}^{3} \frac{x_{i\cdot}^2}{n_i} - \frac{x_{\cdot\cdot}^2}{N} = \frac{48^2}{6} + \frac{72^2}{7} + \frac{108^2}{8} - \frac{228^2}{21} = 107.14$$

$$T_{yy} = \sum_{i=1}^{3} \frac{y_{i\cdot}^2}{n_i} - \frac{y_{\cdot\cdot}^2}{N} = \frac{588.3^2}{6} + \frac{684.7^2}{7} + \frac{781.1^2}{8} - \frac{2054.1^2}{21} = 0.59$$

$$T_{xy} = \sum_{i=1}^{3} \frac{x_{i\cdot} y_{i\cdot}}{n_i} - \frac{x_{\cdot\cdot} y_{\cdot\cdot}}{N} = \frac{48(588.3)}{6} + \frac{72(684.7)}{7} + \frac{108(781.1)}{8} - \frac{228(2054.1)}{21} = -7.78$$

$$E_{xx} = \sum_{i=1}^{3}\sum_{j=1}^{n_i} (x_{ij} - \bar{x}_{i\cdot}) = R_{xx} - T_{xx} = 1754.5 - 107.14 = 1647.36$$

$$E_{yy} = \sum_{i=1}^{3}\sum_{j=1}^{n_i} (y_{ij} - \bar{y}_{i\cdot})^2 = R_{yy} - T_{yy} = 189.97 - 0.59 = 189.38$$

$$E_{xy} = \sum_{i=1}^{3}\sum_{j=1}^{n_i}(x_{ij} - \bar{x}_{i\bullet})(y_{ij} - \bar{y}_{i\bullet}) = R_{xy} - T_{xy} = -503.06 + 7.78 = -495.28$$

The sum-of-squares errors for the reduced and complete models are given by Equations (41) and (42), respectively:

$$\mathrm{SSE_r} = R_{yy} - \frac{R_{xy}^2}{R_{xx}} = 189.97 - \frac{503.06^2}{1754.5} = 45.73$$

$$\mathrm{SSE_c} = E_{yy} - \frac{E_{xy}^2}{E_{xx}} = 189.38 - \frac{495.28^2}{1647.36} = 40.47$$

The mean-square errors for measuring intercept effects are calculated using Equation (43) as

$$\mathrm{MSE_{int}} = \frac{\mathrm{SSE_r} - \mathrm{SSE_c}}{I-1} = \frac{45.73 - 40.47}{3-1} = 2.63$$

$$\mathrm{MSE_c} = \frac{\mathrm{SSE_c}}{N-I-1} = \frac{40.47}{21-3-1} = 2.38$$

Finally the $F$ statistic for testing the equality of intercepts is given by Equation (44) and computed as

$$F(\mathrm{int}) = \frac{\mathrm{MSE_{int}}}{\mathrm{MSE_c}} = \frac{2.63}{2.38} = 1.105$$

The critical value to test the equality of the intercept is $F_{0.25,I-1,N-I-1} = F_{0.25,2,17} = 1.51$. Since $F(\mathrm{int}) < F_{0.25,2,17}$, there is not enough evidence to reject the null hypothesis expressed by Equation (47) at the 25% of level of significance. It can be concluded that the intercepts are not statistically different.

In summary, the null hypotheses for equality of slopes and intercepts are not rejected at the 0.25 significance level and, consequently, all batches are considered from the same population. Therefore, the expiration dating period can be computed by using model (45).

After pooling the data from the three batches, the regression line can be written as

$$\hat{y}_{ij} = a + bx_{ij} = 100.93 - 0.2867x_{ij}$$

The regression subroutine also provides the following calculations: MSE = 2.407, $\bar{x}$ = 10.8571, and $S_{xx}$ = 1.754.6. Thus, to determine the shelf life, the following equation can be used:

$$f(x) = [90 - (100.93 - 0.2867x)]^2 - 1.7291^2(2.407)\left[\frac{1}{21} + \frac{(x - 10.86)^2}{1754.6}\right] = 0$$

To find the shelf life, the root of the above equation has to be found that is smaller than the reference value; the reference value is given by

$$x_{\text{ref}} = \frac{a - 90}{-b} = \frac{100.93 - 90}{0.2867} = 38.11$$

The expiration dating period must be smaller than 38.11 months. Using the initial point $x(0) = x_{\text{ref}} - d = 38.11 - 8$, the QNLS algorithm converges to $x_{\text{R}} = 32.801$. Therefore, the expiration dating period for the underlying production batches is 32 months.

***Minimum Approach for Multiple Batches*** When the hypothesis for equality of slopes is rejected at the 0.25 significance level, the minimum approach should be implemented. This is because the degradation lines of individual batches cannot be considered the same since they have different degradation rates. In this situation the FDA guideline establishes that the overall expiration dating period has to ensure that the product will remain within acceptable limits regardless of the batch from which it comes. Thus, the shelf life for each batch is calculated and the expiration dating period is based on the lowest of all shelf lives. Mathematically, this can be expressed as

$$\min\{x_{\text{L}}(1), \ldots, K, x_{\text{L}}(k)\} \tag{49}$$

where the $x_{\text{L}}(i)$ is the shelf life of the $i$th batch, and $i$ is the total number of batches.

Since the minimum of all the expiration dating periods is the shortest shelf life among all batches, this estimate will provide a 95% confidence that the strength of the drug product will remain above the acceptable lower specification limit.

**Example 3 Slopes and Intercepts Are Different** A stability study provides the assay results shown in Table 16. Based on this information, determine the shelf life for this drug product.

**TABLE 16 Assay Results (%LC) for Example 3**

| Sampling time, months | 0 | 3 | 6 | 9 | 12 | 18 | 24 |
|---|---|---|---|---|---|---|---|
| Batch 1 | 99.2 | 97.1 | 96.1 | 95.2 | 93.8 | 93.1 | 92.4 |
| Batch 2 | 98.7 | 97 | 96.2 | 95.1 | 94.2 | 93.3 | — |
| Batch 3 | 102.5 | 98.9 | 97.1 | 95.6 | 94.1 | 93.1 | — |

***Solution*** The conventional approach to determine the shelf life will require to test whether or not the slopes and the intercepts of the straight lines associated with the degradation rate of the considered batches are the same. ANCOVA is also used to test equalities on both slopes and intercepts.

The strategy consists of testing first the null hypothesis of equalities of slopes, which is given by Equation (14). To determine whether or not this hypothesis is rejected, the statistics defined by Equation (27) is computed:

$$F(\text{slope}) = \frac{\text{MS(slope)}}{\text{MSE}} = \frac{4.0507}{0.8032} = 5.04$$

The null hypothesis, expressed by Equation (15), is rejected at the 0.25 level of significance because the critical value ($F_{0.25,2,13}$ = 1.55) of this test is smaller than $F$(slope). It is concluded that the slopes of the degradation lines of these batches are different and consequently the minimum approach applies. Thus, the shelf life for each batch is computed and the one that exhibits the minimum shelf life is applied to all manufactured batches.

The shelf life for the first batch is obtained by finding the root which is smaller than the reference point of the following equation:

$$[90-(98.0461-0.2898x)]^2 - 2.015^2(0.6689)\left[\frac{1}{7}+\frac{(x-10.28)^2}{429.43}\right]=0$$

The reference point of this equation is

$$x_{\text{ref}} = \frac{a-\delta}{-b} = \frac{98.0461-90}{0.2698} = 29.82$$

Using as an initial point $x(0) = x_{\text{ref}} - 8$, the root is $x_R(1) = 24.93$ and the shelf life for the first batch is $x_L(1) = 24$ months.

Using data for the second batch, the following equation is derived:

$$[90-(98.1157-0.2957x)]^2 - 2.1318^2(0.2328)\left[\frac{1}{6}+\frac{(x-8)^2}{210}\right]=0$$

The reference point for this equation is $x_{\text{ref}} = 27.44$ and the initial point to accomplish convergence is $x_{\text{ref}} - 5$. The corresponding root is $x_R(2) = 23.44$ and the shelf life for batch 2 is $x_L(2) = 23$ months.

The associated equation for the third batch is

$$[90-(100.91-0.5033x)]^2 - 2.1318^2(1.5415)\left[\frac{1}{6}+\frac{(x-8)^2}{210}\right]=0$$

The reference point for this equation is $x_{\text{ref}} = 21.67$ and the initial point to accomplish convergence is $x_{\text{ref}} - 5$. The root for this equation is $x_R(3) = 17.59$ and the shelf life for batch 3 is $x_L(3) = 17$ months.

Therefore, the shelf life that should be applied to future batches is

$$\min\{24 \quad 23 \quad 17\} = 17 \quad \text{months.}$$

**Example 4 Equality of Slopes and Different Intercepts** A stability study provides the assay results shown in Table 17. Based on this information, determine the shelf life for this drug product.

***Solution*** The ANCOVA method is used for testing whether or not the batches have a common slope. The statistics that determines whether or not this hypothesis is rejected is given by Equation (27) and yields

$$F(\text{slope}) = \frac{\text{MS(slope)}}{\text{MSE}} = \frac{0.9891}{2.7134} = 0.3645$$

The null hypothesis, expressed by Equation (15), cannot be rejected at the 0.25 significance level because the critical value ($F_{0.25,2,17} = 1.51$) of this test is larger than $F$(slope). It is concluded that the slopes of the degradation lines of these batches are similar.

It is necessary to test whether or not the intercepts of batches are similar. The statistic to test intercept similarities is given by Equations (43) and (44) and provides the following results:

$$\text{MSE}_{\text{int}} = \frac{\text{SSE}_{\text{r}} - \text{SSE}_{\text{c}}}{I-1} = \frac{172.19 - 48.10}{2} = 62.04 \qquad \text{MSE}_{\text{c}} = \frac{\text{SSE}_{\text{c}}}{N-I-1} = \frac{48.10}{19}$$

$$F(\text{int}) = \frac{\text{MSE}_{\text{int}}}{\text{MSE}_{\text{c}}} = \frac{62.04}{2.53} = 24.51$$

The null hypothesis, expressed by Equation (28), is rejected at the 0.25 significance level because the critical value ($F_{0.25,2,19} = 1.49$) of this test is larger than $F$(int). It is concluded that the intercepts are different at the 0.25 significance level and the shelf life is estimated for each batch and the lowest value is applied to all the manufactured batches.

**TABLE 17 Assay Results (%LC) for Example 4**

| Sampling time, months | 0 | 3 | 6 | 9 | 12 | 18 | 24 | 36 |
|---|---|---|---|---|---|---|---|---|
| Batch 1 | 98.4 | 96.1 | 94.2 | 93.5 | 90 | 89.1 | 89.2 | 87.3 |
| Batch 2 | 99.1 | 97.2 | 96.3 | 95.2 | 93.4 | 91.5 | 90.3 | — |
| Batch 3 | 104.1 | 102.1 | 99.5 | 98.1 | 95.7 | 94.1 | 94.0 | 93.5 |

The model that describes the degradation line is given in Equation (46), where $I = 3$, $n_1 = 8$, $n_2 = 7$, and $n_3 = 8$. To estimate the common slope and the three different intercepts of model (46), it has to be expanded as follows:

$$\begin{aligned}
y_{11} &= \alpha_1 + \beta x_{11} + \varepsilon_{11} \\
y_{12} &= \alpha_1 + \beta x_{12} + \varepsilon_{12} \\
&\vdots \\
y_{18} &= \alpha_1 + \beta x_{18} + \varepsilon_{18} \\
y_{21} &= \alpha_2 + \beta x_{21} + \varepsilon_{21} \\
y_{22} &= \alpha_2 + \beta x_{22} + \varepsilon_{22} \\
&\vdots \\
y_{27} &= \alpha_2 + \beta x_{27} + \varepsilon_{27} \\
y_{31} &= \alpha_3 + \beta x_{31} + \varepsilon_{31} \\
y_{32} &= \alpha_3 + \beta x_{32} + \varepsilon_{32} \\
&\vdots \\
y_{38} &= \alpha_3 + \beta x_{38} + \varepsilon_{38}
\end{aligned} \tag{50}$$

A matrix representation of model (50) is given by the expression

$$\mathbf{y} = \mathbf{Xb} + \mathbf{e} \tag{51}$$

where the elements of **y, X, b**, and **e** are given in Table 18.

Regression techniques are used to estimate the parameter of model (51). Thus the intercepts for batches 1, 2, and 3 are $a_1 = 96.369$, $a_2 = 97.871$ and $a_3 = 101.78$, respectively; the common slope is $b = -0.3069$. All $p$ values of the regression models are close to zero, indicating that the intercept of the batches are highly significant and the coefficient of multiple determination is very high, $R^2 = 0.9996$, indicating very good model fitting.

**TABLE 18 Description of the Matrix and Vectors of Model (50)**

| Vector **y** | Matrix **X** | | | | Vector **b** | Vector **e** |
|---|---|---|---|---|---|---|
| $y_{11}$ | $\alpha_1$ | $\varepsilon_{11}$ | 0 | $x_{11}$ | $\alpha_1$ | $\varepsilon_{11}$ |
| $y_{12}$ | $\alpha_2$ | $\varepsilon_{12}$ | 0 | $x_{12}$ | $\alpha_2$ | $\varepsilon_{12}$ |
| ⋮ | $\alpha_3$ | ⋮ | ⋮ | ⋮ | $\alpha_3$ | ⋮ |
| $y_{18}$ | $\beta$ | $\varepsilon_{18}$ | 0 | $x_{18}$ | $\beta$ | $\varepsilon_{18}$ |
| $y_{21}$ | 0 | $\varepsilon_{21}$ | 0 | $x_{21}$ | | $\varepsilon_{21}$ |
| $y_{22}$ | 0 | $\varepsilon_{22}$ | 0 | $x_{22}$ | | $\varepsilon_{22}$ |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | | ⋮ |
| $y_{27}$ | 0 | $\varepsilon_{27}$ | 0 | $x_{27}$ | | $\varepsilon_{27}$ |
| $y_{31}$ | 0 | $\varepsilon_{31}$ | 1 | $x_{31}$ | | $\varepsilon_{31}$ |
| $y_{32}$ | 0 | $\varepsilon_{32}$ | 1 | $x_{32}$ | | $\varepsilon_{32}$ |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | | ⋮ |
| $y_{38}$ | 0 | $\varepsilon_{38}$ | 1 | $x_{38}$ | | $\varepsilon_{38}$ |

**TABLE 19 Information Required for Quadratic Equation**

| Batch | SSE | MSE | $\bar{x}$ | $S_{xx}$ | $n_i$ |
|---|---|---|---|---|---|
| 1 | 18.8860 | 3.1476 | 13.5 | 1008 | 8 |
| 2 | 3.0331 | 0.6061 | 10.28 | 429.43 | 7 |
| 3 | 26.187 | 4.3646 | 13.5 | 1008 | 8 |

The expiration dating period will be estimated by solving the corresponding quadratic equation (10) for each batch. The major difficulty of using the quadratic equation is the estimation of the MSE for each batch. The MSE estimate for each batch may be computed by performing a regression analysis for each batch; however, this approach may not be correct because there is no guarantee that each regression provides the same slope. Thus, we recommend implementing the following approach. Use results from the regression analysis of model (51) and compute the residuals for the entire model. Extract the corresponding errors for each batch and compute the SSE for each batch using Equation (4). Use this result to calculate the MSE for each batch using Equation (6) and the sample size for the corresponding batch. In addition, extract the corresponding values of $x_{ij}$ for each batch and compute $\bar{x}$, and $S_{xx}$ using Equation (4). Results from the described procedure are shown in Table 19.

The quadratic equation for batch 1 can be written as

$$[90-(96.369-0.3069x)]^2-1.9432^2(3.1476)\left[\frac{1}{8}+\frac{(x-13.5)^2}{1008}\right]=0$$

The reference point for this equation is $x_{\text{ref}}$ = 20.7492 and the initial point to accomplish convergence is $x_{\text{ref}}$ − 6. The root for this equation is $x_{\text{R}}(1)$ = 16.62 and the shelf life for batch 1 is $x_{\text{L}}(1)$ = 16 months.

Following a similar procedure the quadratic equation associated with batch 2 is

$$[90-(97.871-0.3069x)]^2-2.015^2(0.6061)\left[\frac{1}{7}+\frac{(x-10.2857)^2}{429.43}\right]=0$$

The reference point for this equation is $x_{\text{ref}}$ = 25.6451 and the initial point to accomplish convergence is $x_{\text{ref}}$ − 6. The root for this equation is $x_{\text{R}}(2)$ = 22.1394 and the shelf life for batch 2 is $x_{\text{L}}(2)$ = 22 months.

The corresponding quadratic equation for batch 3 is

$$[90-(101.78-0.3069x)]^2-1.9432^2(4.3646)\left[\frac{1}{8}+\frac{(x-13.5)^2}{1008}\right]=0$$

The reference point for this equation is $x_{\text{ref}}$ = 38.38 and the initial point to accomplish convergence is $x_{\text{ref}}$ − 5. The root for this equation is $x_{\text{R}}(3)$ = 30.0519 and the shelf life for batch 3 is $x_{\text{L}}(3)$ = 30 months.

Therefore, the shelf life that should be applied to all manufactured batches is

$$\min\{16 \quad 22 \quad 30\}=16 \quad \text{months}$$

### 3.4 Shelf Life Estimation for Multiple Factors

Most drug products are manufactured with more than one strength and are marketed in more than one package and consequently the stability analyses must be carried out for every combination of package and strength. For instance, suppose that a drug product is available in two strengths and four containers sizes. Thus, eight sets of data from the $2 \times 4$ strength size combinations can be analyzed and eight separate shelf lives should be estimated to calculate the shelf life for the drug product.

A general analysis-of-covariance model for a stability design with several batches and packages can be expressed as

$$y_{ijk} = \alpha_{ij} + \beta_{ij} x_{ijk} + \varepsilon_{ijk} \qquad i = 1, \ldots, I \qquad j = 1, \ldots, J \qquad k = 1, \ldots, n \tag{52}$$

where $y_{ijk}$ is the response from the $k$th time point of the $i$th batch and the $j$th package of a drug product, $x_{ijk}$ is the sampling time at which the $y_{ijk}$ response was obtained, $\alpha_{ij}$ is the intercept for the $i$th batch and the $j$th package, $\beta_{ij}$ is the degradation rate of the $i$th batch and the $j$th package, and $\varepsilon_{ijk}$ is the random error assumed to be independently and normally distributed with zero mean and constant variance $\sigma^2$.

Chen et al. [11] show that there are 16 different models of Equation (52); however, these models reduce to 9. The general procedure consists in identifying the class of model that is associated to a given assay information. Once the model is determined, the appropriate procedure is implemented to determine the shelf life of the drug product. Chen et al. [11] established the procedure for estimating the shelf life.

***Identification of Analysis of Covariance Model*** A general procedure, based on regression analysis, to identify the analysis-of-covariance model that applies to a given set of assay results to determine the shelf life is introduced here. We call this procedure the regression model with indicator variables for testing poolability of batches and packages.

To introduce the fundamental concepts and facilitate the understanding, we consider first a simple case, second a general model, and finally the methodology as an example.

Suppose we have three batches and the corresponding indicator variables model is defined as

$$y_{ij} = \beta_0 + \delta_1 u_1 + \delta_2 u_2 + x_{ij}(\beta_1 + \Delta_1 u_2 + \Delta_2 u_2) + e_{ij} \tag{53}$$

where $y_{ij}$ is the assay result from the $j$th time point of the $i$th batch of a drug product, $x_{ij}$ is the time at which the assay sample $y_{ij}$ was obtained, $u_1$ and $u_2$ are indicator variables that assign a binary code to each factor combination, $\beta_0$ is part of the intercept of the regression line, and $\beta_1$ is part of the slope of the regression line. It should be noted that the intercept of model (53) is $\beta_0 + \delta_1 u_1 + \delta_2 u_2$ and the slope is given by $\beta_1 + \Delta_1 u_2 + \Delta_2 u_2$. The parameters $\delta$ and $\Delta$ are estimated from assay data; $e_{ij}$ is the random error assumed to be independently and normally distributed with zero mean and constant variance $\sigma^2$.

**TABLE 20 Binary Codification for Indicator Variables**

| Batch | $u_1$ | $u_2$ | Assay Result |
|---|---|---|---|
| B1 | 1 | 0 | $y_{1j}$ |
| B2 | 0 | 1 | $y_{2j}$ |
| B3 | 0 | 0 | $y_{3j}$ |

Let B1, B2, and B3 be codes that represent batches 1, 2, and 3, respectively. The indicator variables are used to indicate at which batch (level of factor) is assigned the response variable. For instance, a response variable associated with B1 is represented by $u_1 = 1$ and $u_2 = 0$. Similarly, a response variable from B2 is represented by $u_1 = 0$ and $u_2 = 1$. Also a response variable from B3 is represented by $u_1 = 0$ and $u_2 = 0$. Thus, the values of $u_1$ and $u_2$ define in a unique manner the factor combination. The indicator variables for this model are summarized in Table 20.

The response variable from batch 1 can be written as follows [after replacing the values of indicator variables in model (53)]:

$$\begin{aligned}
y_{11} &= \beta_0 + \delta_1 + x_{11}(\beta_1 + \Delta_1) + e_{11} \\
y_{12} &= \beta_0 + \delta_1 + x_{12}(\beta_1 + \Delta_1) + e_{12} \\
&\vdots \\
y_{1n_1} &= \beta_0 + \delta_1 + x_{1n_1}(\beta_1 + \Delta_1) + e_{1n_1} \\
y_{21} &= \beta_0 + \delta_2 + x_{21}(\beta_1 + \Delta_2) + e_{21} \\
y_{22} &= \beta_0 + \delta_2 + x_{22}(\beta_1 + \Delta_2) + e_{22} \\
&\vdots \\
y_{2n_2} &= \beta_0 + \delta_2 + x_{2n_2}(\beta_1 + \Delta_2) + e_{2n_2} \\
y_{31} &= \beta_0 + x_{31}\beta_1 + e_{31} \\
y_{32} &= \beta_0 + x_{32}\beta_1 + e_{32} \\
&\vdots \\
y_{3n_3} &= \beta_0 + x_{3n_3}\beta_1 + e_{3n_3}
\end{aligned} \tag{54}$$

where $n_i$ is the number of sampling times in the $i$th batch.

It should be noted that the system of linear equations expressed by (54) represents the response variable from the three batches. The required condition for the three batches to have the same intercept is that $\delta_1 = \delta_2 = 0$. The three batches will have the same slope if and only if $\Delta_1 = \Delta_2 = 0$. Thus, the problem for testing poolability reduces to fit the regression model (54) and test the following hypotheses: $h_1$: $\delta_1 = \delta_2 = 0$ and $h_2$: $\Delta_1 = \Delta_2 = 0$. Therefore, if the null hypothesis $h_2$ is not rejected at the 0.25 significance level, it implies that the slopes of the three batches are the same, that is, $\beta_1 = \beta_2 = \beta_3 = \beta$. Similarly, if the null hypothesis $h_1$ is not rejected, the intercepts of the three batches are the same, that is, $\alpha_1 = \alpha_2 = \alpha_3 = \alpha$. If that is the case, the shelf life is determined by a model with a single intercept and a single slope.

This procedure was applied to data from Examples 2 and 3 and the results from the regression analysis are given in Table 21.

**TABLE 21 Testing for Poolability of Batches**

| Example 2: Equality of Slopes and Intercepts[a] | | | | Example 3: Slopes and Intercepts Are Different[b] | | | |
|---|---|---|---|---|---|---|---|
| Parameter | Estimate | $t$ Statistic | $p$ Value | Parameter | Estimate | $t$ Statistic | $p$ Value |
| $\beta_0$ | 101.58 | 112.20 | 0.0000 | $\beta_0$ | 100.91 | 163.99 | 0.0000 |
| $\delta_1$ | –0.87 | –0.61 | 0.5513 | $\delta_1$ | –2.86 | –3.44 | 0.0044 |
| $\delta_2$ | –0.62 | –0.46 | 0.6543 | $\delta_2$ | –2.79 | –3.21 | 0.0068 |
| $\beta_1$ | –0.29 | –5.66 | 0.0000 | $\beta_1$ | –0.50 | –8.14 | 0.0000 |
| $\Delta_1$ | –0.04 | –0.32 | 0.7559 | $\Delta_1$ | 0.23 | 3.09 | 0.0085 |
| $\Delta_2$ | –0.01 | –0.14 | 0.8883 | $\Delta_2$ | 0.21 | 2.37 | 0.0337 |

Example 2:
[a]$h_1$: $\delta_1 = \delta_2 = 0$. $P$ Values of this table show that this hypothesis is not rejected at the 0.25 significance level and batches have common intercepts.
$h_2$: $\Delta_1 = \Delta_2 = 0$. This hypothesis is not rejected at the 0.25 significance level and batches have common slopes.
Example 3:
[b]$h_1$: $\delta_1 = \delta_2 = 0$. This hypothesis is rejected at the 0.25 significance level and batches have different intercepts.
$h_2$: $\Delta_1 = \Delta_2 = 0$. This hypothesis is rejected at the 0.25 significance level and batches have different slopes.

The indicator variables model to perform a poolability test for two factors (packages and batches) can be expressed as follows:

$$\begin{aligned} y_{ijk} = {} & \beta_0 + \beta_1 x_{ijk} + \delta_1 u_1 + \delta_2 u_2 + \ldots + \delta_r u_r + \phi_1 v_1 + \phi_2 v_2 \\ & + \ldots + \phi_s v_s + x_{ijk}(\Delta_1 u_1 + \Delta_2 u_2 + \ldots + \Delta_r u_r + \Phi_1 v_1 \\ & + \Phi_2 v_2 + \ldots + \Phi_s v_s + \omega_{11} u_1 v_1 + \omega_{12} u_1 v_2 + \ldots \\ & + \omega_{rs} u_r v_s) + e_{ijk} \end{aligned} \tag{55}$$

where $i = 1, \ldots, I$, $j = 1, \ldots, J$, and $k = 1, \ldots, n_{ij}$ and $I$ is the number of batches, $J$ is the number of packages, $n_{ij}$ is the sample size of the $i$th batch and the $j$th package, and $u_i$ and $v_j$ are indicator variables that are defined as shown in Table 22.

The first column of Table 22 shows the factor combinations. For instance, B1P1 represents the first batch and the first package, B1P2 indicates the first batch and the second package, and BIPJ represents the last batch and the last package. The subscripts $r$ and $s$ are defined as follows: $r = I - 1$, $s = J - 1$.

Testing for poolability requires fitting the regression model with indirect variables to the assay data and testing four possible hypotheses. Table 23 shows the hypotheses to be tested, the corresponding interpretation to whether or not the hypothesis is accepted or rejected, the code of the model, and the model established by the FDA [8, 11]. In Table 23 the letters A and R represent whether the considered hypotheses are accepted or rejected, respectively.

***Rules for Determining Shelf Life*** Once the analysis-of-covariance model has been identified, a set of rules for computing the shelf life must be implemented. This section describes the rules to follow to determine the expiration dating period for each of the nine representative models described in the previous section [8, 11]:

**TABLE 22 Binary Code for Factor Combinations**

| Factor Combinations | $u_1$ | $u_2$ | … | $u_r$ | $v_1$ | $v_2$ | … | $v_r$ | $y_{ijk}$ |
|---|---|---|---|---|---|---|---|---|---|
| B1P1 | 1 | 0 | … | 0 | 1 | 0 | … | 0 | $y_{11k}$ |
| B2P1 | 0 | 1 | … | 0 | 1 | 0 | … | 0 | $y_{21k}$ |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |
| BrP1 | 0 | 0 | … | 1 | 1 | 0 | … | 0 | $y_{r1k}$ |
| BIP1 | 0 | 0 | … | 0 | 1 | 0 | … | 0 | $y_{I1k}$ |
| B1P2 | 1 | 0 | … | 0 | 0 | 1 | … | 0 | $y_{12k}$ |
| B2P2 | 0 | 1 | … | 0 | 0 | 1 | … | 0 | $y_{22k}$ |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |
| BrP2 | 0 | 0 | … | 1 | 0 | 1 | … | 0 | $y_{r2k}$ |
| BIP2 | 0 | 0 | … | 0 | 0 | 1 | … | 0 | $y_{I2k}$ |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |
| B1PJ | 1 | 0 | … | 0 | 0 | 0 | … | 1 | $y_{1Jk}$ |
| B2PJ | 0 | 1 | … | 0 | 0 | 0 | … | 1 | $y_{2Jk}$ |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |
| BrPJ | 0 | 0 | … | 1 | 0 | 0 | … | 1 | $y_{rJk}$ |
| BIPJ | 0 | 0 | … | 0 | 0 | 0 | … | 1 | $y_{IJk}$ |

1. C0: The expiration dating period is computed separately for each batch and package combination. The minimum shelf life of the individual batches (at least three batches) from the same package is the shelf life of that package.
2. C1: The data from the individual batches are combined within each package. For each package:
   - M1: Assuming a common slope with different intercepts for different batches, the shelf lives are computed for individual batches. The lowest of all shelf lives of the individual batches is the expiration dating period of that package.
   - M3: Assuming a common slope and a common intercept for all batches, a single shelf life is computed as the expiration dating period of that package.
3. C2: The data from individual packages are combined within the batch:
   - M2: For each batch, assuming a common slope with different intercepts for different packages, the shelf lives are computed for individual packages. The lowest shelf life of the individual batches from the same package is used as the expiration dating period of that package.
   - M4: For each batch, assuming a common slope and a common intercept for all packages, a single shelf life is computed. The lowest expiration dating period of the individual batches is used as the shelf life for every package.
4. C3: The data for all packages and batches are combined:
   - M5: Assuming different intercepts but a common slope for all batches and package combinations. The lowest shelf life of the individual batches from the same package is used as the expiration dating period of that package.
   - M6: Assuming a common intercept for all batches with a common slope for all batch and package combinations, a single expiration dating period is computed for each package to be used as the expiration dating period of that package.

**TABLE 23 Procedure for Identifying Analysis of Covariance Model**

| Code | Hypotheses | A | R | Interpretation | ANCOVA Model |
|---|---|---|---|---|---|
| 1 C0: M0a | $H_1$: $\delta_1 = \delta_2 = \ldots = \delta_r = 0$ | X | | Intercepts of all batches are the same, $\alpha_{1j} = \alpha_{2j} = \ldots = \alpha_{Ij} = \alpha_j$ | $y_{ijk} = \alpha_j + \beta_{ij}x_{ijk} + e_{ijk}$ |
| 2 C1: M1 | $H_2$: $\Delta_1 = \Delta_2 = \ldots = \Delta_r = 0$, $\omega_{11} = \omega_{12} = \ldots = \omega_{rs} = 0$ | X | | Slopes of all batches are the same, $\beta_{1j} = \beta_{2j} = \ldots = \beta_{Ij} = \beta_j$ | $y_{ijk} = \alpha_{ij} + \beta_j x_{ijk} + e_{ijk}$ |
| 2 C1: M3 | $H_1$ and $H_2$ | X | | All batches have the same intercept and the same slope, $\alpha_{1j} = \alpha_{2j} = \ldots = \alpha_{Ij} = \alpha_j$ $\beta_{1j} = \beta_{2j} = \ldots = \beta_{Ij} = \beta_j$ | $y_{ijk} = \alpha_j + \beta_j x_{ijk} + e_{ijk}$ |
| 1 C0: M0b | $H_3$: $\phi_1 = \phi_2 = \ldots = \phi_s = 0$ | X | | Intercepts of all packages are the same, $\alpha_{i1} = \alpha_{i2} = \ldots = \alpha_{iJ} = \alpha_i$ | $y_{ijk} = \alpha_i + \beta_{ij}x_{ijk} + e_{ijk}$ |
| 3 C2: M2 | $H_4$: $\Phi_1 = \Phi_2 = \ldots = \Phi_s = 0$, $\omega_{11} = \omega_{12} = \ldots = \omega_{rs} = 0$ | X | | Slopes of all packages are the same, $\beta_{i1} = \beta_{i2} = \ldots = \beta_{iJ} = \beta_i$ | $y_{ijk} = \alpha_{ij} + \beta_i x_{ijk} + e_{ijk}$ |
| 3 C2: M4 | $H_3$ and $H_4$ | X | | Intercepts and slopes of all packages are the same, $\alpha_{i1} = \alpha_{i2} = \ldots = \alpha_{iJ} = \alpha_i$ $\beta_{i1} = \beta_{i2} = \ldots = \beta_{iJ} = \beta_i$ | $y_{ijk} = \alpha_j + \beta_i x_{ijk} + e_{ijk}$ |
| 3 C3: M5 | $H_2$ and $H_4$ | X | | All batches and all packages will have a common slope, $\beta_{1j} = \beta_{2j} = \ldots = \beta_{Ij} = \beta_j$ $\beta_{i1} = \beta_{i2} = \ldots = \beta_{iJ} = \beta_i$ $\beta_i = \beta_j = \beta$ | $y_{ijk} = \alpha_{ij} + \beta x_{ijk} + e_{ijk}$ |
| 4 C3: M6 | $H_1$, $H_2$, and $H_4$ | X | | All batches and packages have common slope and the intercepts of all batches are the same, $\alpha_{1j} = \alpha_{2j} = \ldots = \alpha_{Ij} = \alpha_j$ $\beta_{1j} = \beta_{2j} = \ldots = \beta_{Ij} = \beta_j$ $\beta_{i1} = \beta_{i2} = \ldots = \beta_{iJ} = \beta_i$ $\beta_i = \beta_j = \beta$ | $y_{ijk} = \alpha_j + \beta x_{ijk} + e_{ijk}$ |
| 4 C3: M7 | $H_2$, $H_3$, and $H_4$ | X | | All batches and packages have common slope and the intercepts of all packages are the same, $\alpha_{i1} = \alpha_{i2} = \ldots = \alpha_{iJ} = \alpha_i$ $\beta_{1j} = \beta_{2j} = \ldots = \beta_{Ij} = \beta_j$ $\beta_{i1} = \beta_{i2} = \ldots = \beta_{iJ} = \beta_i$ $\beta_i = \beta_j = \beta$ | $y_{ijk} = \alpha_i + \beta x_{ijk} + e_{ijk}$ |

**TABLE 23** ***Continued***

| Code | Hypotheses | A | R | Interpretation | ANCOVA Model |
|---|---|---|---|---|---|
| 4 C8: M8 | $H_1$, $H_2$, $H_3$, and $H_4$ | X | | Slopes and intercepts of all batches and packages are the same, $\alpha_{1j} = \alpha_{2j} = \ldots = \alpha_{Ij} = \alpha_j$ $\alpha_{i1} = \alpha_{i2} = \ldots = \alpha_{iJ} = \alpha_i$ $\alpha_i = \alpha_j = \alpha$ $\beta_{1j} = \beta_{2j} = \ldots = \beta_{Ij} = \beta_j$ $\beta_{i1} = \beta_{i2} = \ldots = \beta_{iJ} = \beta_i$ $\beta_i = \beta_j = \beta$ | $y_{ijk} = \alpha + \beta x_{ijk} + e_{ijk}$ |
| 1 C0 | $H_1$, $H_2$, $H_3$, and $H_4$ | | X | Intercepts and slopes of batches and packages are different | $y_{ijk} = \alpha_{ii} + \beta_{ij} x_{ijk} + e_{ijk}$ |

**TABLE 24 Assay Results (%LC)**

| | | By Sampling Time in Months | | | | | |
|---|---|---|---|---|---|---|---|
| Package | Batch | 0 | 3 | 6 | 9 | 12 | 18 |
| Bottle | 1 | 104.8 | 102.5 | 101.5 | 102.4 | 99.4 | 96.5 |
| | 2 | 103.9 | 101.9 | 103.2 | 99.6 | 100.2 | 98.8 |
| | 3 | 103.5 | 102.1 | 101.9 | 100.3 | 99.2 | 101.0 |
| | 4 | 101.5 | 100.3 | 101.1 | 100.6 | 100.7 | 98.4 |
| | 5 | 106.1 | 104.3 | 101.5 | 101.1 | 99.4 | 98.2 |
| Blister | 1 | 102.0 | 101.6 | 100.9 | 101.1 | 101.7 | 97.1 |
| | 2 | 104.7 | 101.3 | 103.8 | 99.8 | 98.9 | 97.1 |
| | 3 | 102.5 | 102.3 | 100.0 | 101.7 | 99.0 | 100.9 |
| | 4 | 100.1 | 101.8 | 101.4 | 99.9 | 99.2 | 97.4 |
| | 5 | 105.2 | 104.1 | 102.4 | 100.2 | 99.6 | 97.5 |

- M7: Assuming a common intercept for all packages and a common slope for all batches and package combinations, the expiration dating periods are computed for each batch. The minimum of the individual batches is used as the expiration dating period for every package.
- M8: Assuming a common slope and a common intercept for all packages and batch combinations, a single expiration dating period is computed to be used as the shelf life for every package.

**Example 5 Multiple-Factor Stability Study** A well-known example introduced by Shao and Chow [19] is used to illustrate the application of shelf life calculations for a multifactor case. A stability study was conducted on a 300-mg tablet of a drug product to establish the shelf life for each of the two types of packages used for this product: bottle and blister. The results are shown in Table 24. Each type of package includes five batches. The tablets were tested for potency at 0, 3, 6, 9, 12, and 18 months. Determine the shelf life based on these stability data.

***Solution*** Based on these data, we can extract the following information to build the regression model with indicator variables: $I = 5$ batches, $J = 2$ packages, $r = 4$, $s = 1$, and $n = 6$ sampling times for all batches: 0, 3, 6, 9, 12, and 18 months. The indicator variables are shown in Table 25 and the indicator variables model for this case is

$$\begin{aligned} y_{ijk} = {} & \beta_0 + \beta_1 x_{ijk} + \delta_1 u_1 + \delta_2 u_2 + \delta_3 u_3 + \delta_4 u_4 + \phi_1 v_1 \\ & + x_{ijk}(\Delta_1 u_1 + \Delta_2 u_2 + \Delta_3 u_3 + \Delta_4 u_4 + \Phi_1 v_1 + \omega_{11} u_1 v_1 \\ & + \omega_{21} u_2 v_1 + \omega_{31} u_3 v_1 + \omega_4 u_4 v_1) + e_{ijk} \end{aligned} \tag{56}$$

where subscript $i$ refers to the batch, subscript $j$ to the package type, and subscript $k$ to the sampling time.

A conventional computer package for regression analysis was used to estimate the parameters for the model described by Equation (56) and the results are summarized in Table 26.

**TABLE 25 Definition of Indicator Variables**

| Factor Combinations | $u_1$ | $u_2$ | $u_3$ | $u_4$ | $v_1$ | $y_{ijk}$ |
|---|---|---|---|---|---|---|
| B1P1 | 1 | 0 | 0 | 0 | 1 | $y_{11k}$ |
| B2P1 | 0 | 1 | 0 | 0 | 1 | $y_{21k}$ |
| B3P1 | 0 | 0 | 1 | 0 | 1 | $y_{31k}$ |
| B4P1 | 0 | 0 | 0 | 1 | 1 | $y_{41k}$ |
| B5P1 | 0 | 0 | 0 | 0 | 1 | $y_{51k}$ |
| B1P2 | 1 | 0 | 0 | 0 | 0 | $y_{12k}$ |
| B2P2 | 0 | 1 | 0 | 0 | 0 | $y_{22k}$ |
| B3P2 | 0 | 0 | 1 | 0 | 0 | $y_{32k}$ |
| B4P2 | 0 | 0 | 0 | 1 | 0 | $y_{42k}$ |
| B5P2 | 0 | 0 | 0 | 0 | 0 | $y_{52k}$ |

**TABLE 26 Parameter Estimation for Indicator Variables Model**

| Parameter | Estimate | $t$ Statistics | $p$ Value |
|---|---|---|---|
| $\beta_0$ | 104.95 | 185.76 | 0.0000 |
| $\delta_1$ | –1.63 | –2.23 | 0.0306 |
| $\delta_2$ | –1.33 | –1.83 | 0.0747 |
| $\delta_3$ | –2.83 | –3.88 | 0.0003 |
| $\delta_4$ | –3.64 | –4.99 | 0.0000 |
| $\phi_1$ | 0.44 | 0.96 | 0.3408 |
| $\beta_1$ | –0.43 | –6.91 | 0.0000 |
| $\Delta_1$ | 0.15 | 1.73 | 0.0901 |
| $\Delta_2$ | 0.07 | 0.88 | 0.3823 |
| $\Delta_3$ | 0.31 | 3.62 | 0.0007 |
| $\Delta_4$ | 0.26 | 3.01 | 0.0043 |
| $\Phi_1$ | –0.01 | –0.18 | 0.8557 |
| $\omega_{11}$ | –0.06 | –0.65 | 0.5161 |
| $\omega_{21}$ | 0.05 | 0.56 | 0.5770 |
| $\omega_{31}$ | –0.02 | –0.21 | 0.8313 |
| $\omega_{41}$ | 0.04 | 0.43 | 0.6660 |

*P Values* of Table 26 show the hypothesis $H_3$: $\phi = 0$ and $H_4$: $\Phi_1 = \omega_{11} = \omega_{21} = \omega_{31} = \omega_{41} = 0$ are not rejected at the 0.25 significance level and, according to Table 23, the intercepts and slopes of all packages are the same, that is, $\alpha_{i1} = \alpha_{i2} = \alpha_i$ and $\beta_{i1} = \beta_{i2} = \beta_i$. Therefore, the analysis-of-covariance model that should be applied to the assay results has the code 3 C2: M4 and has the form $y_{ijk} = \alpha_i + \beta_i x_{ijk} + e_{ijk}$.

Since there is no difference between the packages, a simple analysis-of-covariance model is suitable. The rules for computing the expiration dating period is given in Section 3.3 and indicate that a shelf life should be computed for each batch and the minimum criterion is used to determine the expiration dating period for the underlying batches of the drug product. Thus, regression model fitting should be conducted using model 3 C2: M4 and the results are used to develop the quadratic equation.

The corresponding quadratic equation to determine the shelf life for the first batch is

$$[90-(103.54-0.3233x)]^2 - 1.8125^2(1.48)\left[\frac{1}{12}+\frac{(x-8)^2}{420}\right]=0$$

The reference point for this equation is $x_{\text{ref}} = 41.89$ and the initial point to accomplish convergence is $x_{\text{ref}} - 7$. The root for this equation is $x_{\text{R}}(1) = 33.25$ and the shelf life for batch 1 is $x_{\text{L}}(1) = 33$ months. Since the sampling times are the same for all batches, the following values remain invariant for all batches: $n = 12$, $\bar{x} = 8$, $S_{xx} = 420$, and $t_{0.05,10} = 1.8125$. The values required by the quadratic equation and the associated shelf life for each batch are given in Table 27. The computer program to perform this calculation is given in the Appendix.

Therefore, the shelf life that should be applied to current and future batches is

$$\min\{33 \quad 32 \quad 54 \quad 49 \quad 31\} = 31 \quad \text{months}$$

This approach provides a conservative estimate of the overall shelf life because it provides more than 95% confidence for all batches except the batch from which it is estimated. It is worth mentioning that the minimum approach has been highly criticized by several researchers. For instance, Chow and Shao [20] pointed out that the minimum approach lacks statistical justifications. Ruberg and Stegeman [21] and Ruberg and Hsu [22] described the drawbacks of this methodology. However, Chen et al. [11] show that the FDA procedure performs reasonably well for data from a typical stability design with three batches. The FDA procedure is based on the assumption that batch effects are fixed. If the analysis shows that batch-to-batch differences are small, it is advantageous to combine all the data to obtain one overall estimate. If the analysis shows evidence of batch-to-batch differences, then the FDA

**TABLE 27 Required Parameters to Compute Shelf Life**

| batch | $a$ | $b$ | MSE | $x_{\text{ref}}$ | $x_{\text{R}}$ | Shelf Life |
|---|---|---|---|---|---|---|
| 2 | 103.84 | –0.3429 | 1.47 | 40.37 | 32.49 | 32 |
| 3 | 102.34 | –0.1429 | 1.24 | 86.4 | 54.26 | 54 |
| 4 | 101.54 | –0.1671 | 0.74 | 69.02 | 49.84 | 49 |
| 5 | 105.17 | –0.4426 | 0.45 | 34.28 | 31.08 | 31 |

uses the minimum of all estimates obtained from individual batches based on the premise that the overall expiration dating period may depend on the minimum time a batch may be expected to remain within acceptable limits. It also shows that the smaller is the experimental variance, the higher is the expiration period. Thus, the FDA exhibits conservatism when the number of batches is large with small batch variabilities. The FDA established that the 0.25 significance level is used to compensate for the expected low power of the design due to the relatively limited sample size in a typical formal stability study [8]. We supported the FDA arguments for the 0.25 significance level and this handbook outlines the procedure approved by the FDA.

## 4 SHORT-TERM STABILITY ANALYSIS

### 4.1 Introduction

Assuring acceptable stability of drugs remains a challenge to industry. As mentioned in Section 2.1 the stability concept involves the classical aspect of API degradation as well as the presence of degradation products at levels that represent a risk to the patient [23]. In general, two stability studies are conducted to ensure that the market drug product is under the required specifications: short-term and long-term studies. A short-term stability study is an accelerated stability testing study, that is, a study under stressed storage conditions. The main goal of accelerated stability testing is not only to determine the chemical reaction kinetics but also to establish a tentative expiration date under the environment of marketing storage conditions.

The stability of a drug product depends on the storage conditions, temperature and relative humidity, at which the product is exposed during its shelf life period. The effect of the environment on the degradation depends on the packaging configuration used and the chemical characteristics of the drug product. The degradation of a drug product is mainly caused by the chemical reaction of the API with the excipients or with species in the environment (atmospheric oxygen and humidity) causing a decrease of the assay results over time. During the whole shelf life period, the API of a drug product is degrading at a certain rate. Other quality attributes (e.g., physical appearance, sterility, drug release rate, impurities) may also be changing with time, affecting the functionality, efficacy, or purity of the drug product. The magnitude of these changes during the expiration dating period should in no way represent a risk to the patient.

The rate of a chemical reaction is a function of temperature and concentration of the reactants present. Therefore, to induce a change in the reaction rate, a change in temperature or reactant concentration must be caused. The temperature can be readily changed by varying the storage temperature of the drug product. The reactants in a drug product are the API, the excipients, and the environment air and humidity. Of these, only the environment air and humidity can be changed because the API and the excipients within the drug product are specified and cannot be changed.

It is of interest for the pharmaceutical industry to know the degradation of its drug products at accelerated conditions to assess degradation for longer term storage at nonextreme conditions and short excursions outside of the recommended storage

conditions that might occur during shipping. The short-term stability study is the method to evaluate degradation of a drug product at accelerated conditions. As an example of this, Gil-Alegre et al. [24] studied the degradation kinetics of mitonafide, an antineoplastic agent, at temperatures between 60°C and 90°C at 10°C intervals. This is a very stable drug and, thus, rather high temperatures had to be used to have detectable concentration changes. Normally, accelerated testing is carried out at constant temperature. However, ramping temperatures have been used in non-isothermal testing [25]. This approach may lead to better prediction of the low-temperature kinetic constant and thus expiration dating periods [23].

### 4.2 Chemical Reaction Kinetics

To understand how degradation data are treated, it is convenient to mention the basics of chemical reaction kinetics. The principles of chemical reaction engineering can be found in any reaction engineering or reactor design textbook [26]. A chemical reaction is the process whereby one or more components are transformed into one or more different components. The rate of reaction is the velocity at which the component(s) are being transformed in a chemical reaction. For the chemical reaction

$$\mathrm{A} + \mathrm{B} \rightarrow \mathrm{E} \tag{57}$$

the reaction rate can be expressed as

$$\frac{dC}{dt} = -k_{n+m}[\mathrm{A}]^n[\mathrm{B}]^m \tag{58}$$

where $C$ is the molar concentration of the component to be studied, that is, [A] or [B], the brackets represent the concentration of the reactants, and $k$ is the rate constant, also known as the specific reaction rate. The rate of reaction is positive if it refers to a product and negative if it refers to a reactant. The reaction order is the sum of the exponents ($n + m$) in Equation (58). According to this, a reaction of order zero is represented with the equation

$$\frac{dC}{dt} = -k_0 \tag{59}$$

which can be integrated to give

$$C = C_0 - k_0 t \tag{60}$$

where $C$ is the molar concentration at any time $t$, $C_0$ is the concentration at time $t = 0$, $k_0$ is the rate constant of the zero-order reaction, and $t$ is the time. For a first-order reaction, the rate can be expressed as

$$\frac{dC}{dt} = -k_1 C \tag{61}$$

which integrates to

$$\ln\left(\frac{C}{C_0}\right) = -k_1 t \qquad \text{or} \qquad C = C_0 - C_0 e^{-k_1 t} \tag{62}$$

The temperature has a strong effect on the rate constant; this effect is represented by the Arrhenius equation:

$$k = Ae^{-E/(RT)} \tag{63}$$

where $A$ is the frequency factor and has the same units of $k$, $E$ is the activation energy, $R$ is the gas constant, and $T$ is absolute temperature in kelvin. The logarithmic expression of the Arrhenius equation is

$$\ln k = -\frac{E}{RT} + \ln A \tag{64}$$

### 4.3 Degradation Data Evaluation at Accelerated Conditions

Chemical reaction kinetics can be used to evaluate degradation data at accelerated conditions and predict the drug product assay at normal conditions for periods longer than the proposed shelf life. This is applicable to limited cases because the reaction kinetics is often complex for drug products. The following example illustrates the procedure to follow to calculate the API concentration with time for a drug product stored at normal conditions when such data are not yet available.

**Example 6 Estimation of Degradation from Accelerated Data: First-Order Case** Consider the degradation data shown in Table 28, obtained at normal, intermediate, and accelerated storage conditions. All assay values are expressed as a percentage of the label claim (%LC). Ignoring the first two columns, that is, using only the data at higher temperatures (30°C and higher), estimate the assay values at 25°C as a function of time. [*Note:* The data in this example are artificial and will be used only for demonstration purposes. The normal-conditions data are displayed in the first two columns (25°C). Usually these data are not available when a new drug application (NDA) is submitted and that is why the procedure being presented here is important.]

**TABLE 28 Assay Degradation Data for Drug Product**

| $t$ (months) | Assay at 25°C | $t$ (months) | Assay at 30°C | Assay at 40°C | Assay at 50°C |
|---|---|---|---|---|---|
| 0 | 99.9 | 0 | 99.9 | 99.9 | 99.9 |
| 3 | 99.4 | 2 | 99.4 | 98.0 | 95.6 |
| 6 | 98.2 | 4 | 98.7 | 95.9 | 91.2 |
| 9 | 97.8 | 6 | 97.4 | 95.1 | 87.4 |
| 12 | 97.4 | | | | |
| 18 | 95.6 | | | | |
| 24 | 94.5 | | | | |
| 36 | 91.7 | | | | |

***Solution*** The analysis can be done assuming that the reaction is first order (this is usually the case in real life) or zero order (simpler analysis with no significant error for degradation of less than 10%). Both cases are presented here. For first-order kinetics [refer to Equation (61)], the first step is to obtain the natural logarithm of the concentrations, which is done in Table 29. Then, these values are graphed as a function of time where straight lines should be obtained as shown in Figure 2. The slope of each line corresponds to the value of $k_1$, which is shown in the bottom part of Table 29. Next, a plot of $\ln(k_1)$ versus $1/T$ is prepared with the three values of $k_1$, one for each temperature as shown in Figure 3. Note that absolute temperature should be used here. From Equation (64), the value of $k_1$ can be extrapolated to $T = 25°C$ ($1/T = 0.003354\ K^{-1}$) to then predict the drug product assay as a function of time. The values of slope and intercept are shown in the inset of Figure 3; the calculation is

$$\ln k_{25} = -8237.4 \times 0.003354 + 21.639 = -5.990$$

$$k_{25} = 0.00250\,\text{month}^{-1}$$

Once the specific reaction rate is calculated, the drug product assay can be predicted using Equation (62). The results are shown in Table 30 along with the actual data from Table 28.

**TABLE 29 Degradation Data Treated as First-Order Kinetics**

| *t* (months) | ln(assay) at 30°C | ln(assay) at 40°C | ln(assay) at 50°C |
|---|---|---|---|
| 0 | 4.60 | 4.60 | 4.60 |
| 2 | 4.59 | 4.58 | 4.55 |
| 4 | 4.58 | 4.55 | 4.50 |
| 6 | 4.57 | 4.55 | 4.46 |
| $k_1$ | 0.00415 | 0.00847 | 0.02224 |
| $\ln(k_1)$ | –5.483 | –4.771 | –3.798 |
| *1/T* | 0.00330 | 0.00319 | 0.00309 |

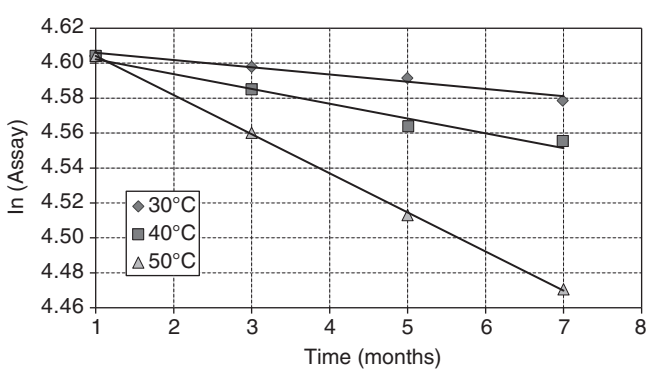


**FIGURE 2** Degradation data treated as first-order kinetics.

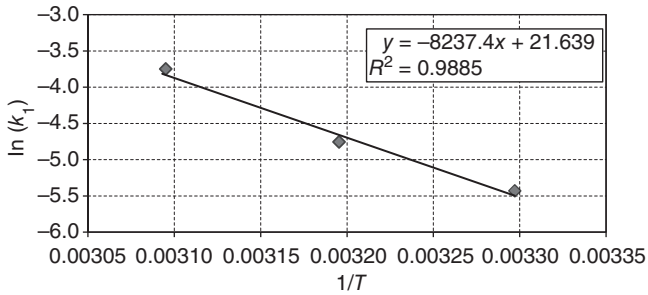


**FIGURE 3** Arrhenius plot for first-order degradation.

**TABLE 30 Predicted Assay Values Using Arrhenius and First-Order Kinetics**

| $t$ (months) | Assay | Predicted Assay |
|---|---|---|
| 0 | 99.0 | 100.0 |
| 3 | 98.5 | 98.3 |
| 6 | 97.3 | 97.5 |
| 9 | 96.9 | 96.8 |
| 12 | 96.5 | 96.1 |
| 18 | 94.7 | 94.6 |
| 24 | 93.6 | 93.2 |
| 36 | 90.8 | 90.5 |

**Example 7 Estimation of Degradation from Accelerated Data: Zero-Order Case** Repeat Example 6 assuming that the degradation is of zero order.

***Solution*** For zero-order kinetics [refer to Equation (60)], the original concentrations can be graphed as a function of time. Straight lines can be drawn through the points even if the actual degradation order is 1 because the degradation is small (less than 10%), as seen in Figure 4. The slope of each line corresponds to the value of $k_0$. These values are shown in Table 31. Next, a plot of $\ln(k_0)$ versus $1/T$ is prepared with the three values of $k_0$, one for each temperature, which is shown in Figure 5.

As in Example 6, the value of $k_0$ can be extrapolated to $T = 25°C$ ($1/T = 0.003354$ $K^{-1}$) using Equation (64) to then predict the drug product assay as a function of time. The values of slope and intercept are shown in the inset of Figure 5; the calculation is

$$\ln k_{25} = -7974.8 \times 0.003354 + 25.369 = -1.379$$

$$k_{25} = 0.2519 \text{ month}^{-1}$$

Once the specific reaction rate is calculated, the drug product assay can be predicted using Equation (60). The results are shown in Table 32 along with the actual data from Table 28.

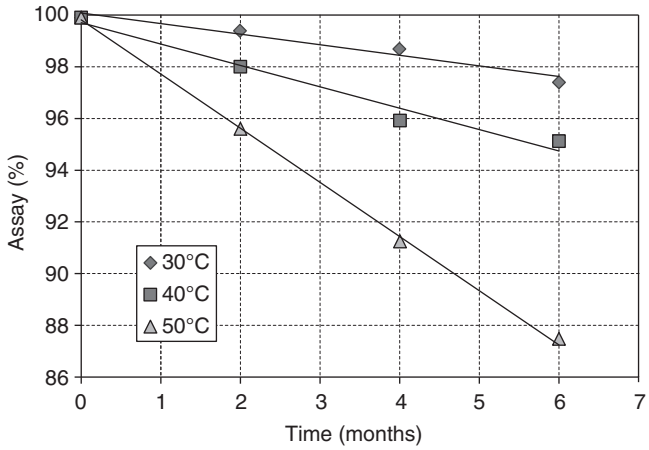


**FIGURE 4** Degradation data treated as zero-order kinetics.

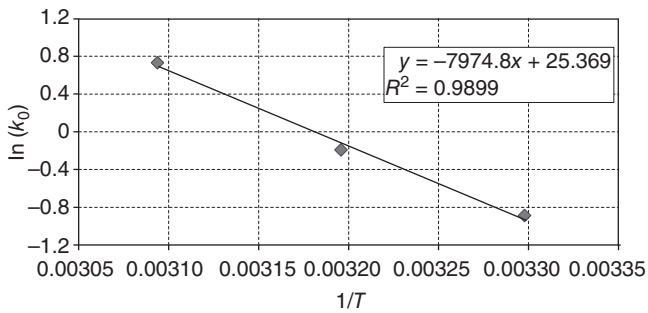


**FIGURE 5** Arrhenius plot for zero-order degradation.

**TABLE 31 Degradation Data Treated as Zero-Order Kinetics**

| | 30°C | 40°C | 50°C |
|---|---|---|---|
| $k_0$ | 0.410 | 0.825 | 2.095 |
| $\ln(k_0)$ | –0.892 | –0.192 | 0.740 |
| $1/T$ | 0.00330 | 0.00319 | 0.00309 |

**TABLE 32 Predicted Assay Values Using Arrhenius and Zero-Order Kinetics**

| $t$ (months) | Assay | Predicted Assay |
|---|---|---|
| 0 | 99.0 | 100.0 |
| 3 | 98.5 | 99.2 |
| 6 | 97.3 | 98.5 |
| 9 | 96.9 | 97.7 |
| 12 | 96.5 | 97.0 |
| 18 | 94.7 | 95.5 |
| 24 | 93.6 | 94.0 |
| 36 | 90.8 | 90.9 |
| Shelf life | 41 | 39 |

**TABLE 33 Actual and Predicted Assay Values**

| | | Predicted Assay | |
|---|---|---|---|
| *t* (months) | Assay | First Order | Zero Order |
| 0 | 99.9 | 100.0 | 100.0 |
| 3 | 99.4 | 99.3 | 99.2 |
| 6 | 98.2 | 98.5 | 98.5 |
| 9 | 97.8 | 97.8 | 97.7 |
| 12 | 97.4 | 97.0 | 97.0 |
| 18 | 95.6 | 95.6 | 95.5 |
| 24 | 94.5 | 94.2 | 94.0 |
| 36 | 91.7 | 91.4 | 90.9 |

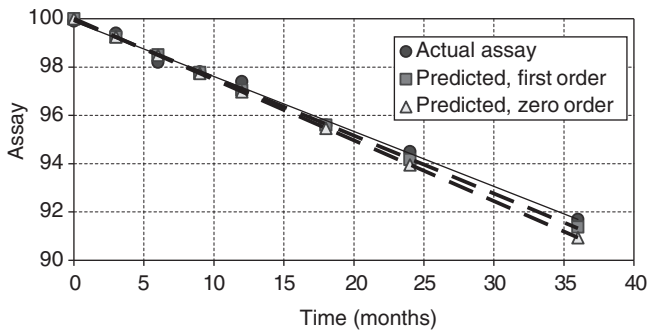


**FIGURE 6** Actual and predicted assay values.

***Comparison of Models*** To compare the predictions of both models (first- and zero-order kinetics), Table 33 and Figure 6 have been prepared. As expected, both predictions give quite similar results. The first-order model gives slightly better results, suggesting that the degradation is actually of first order. Both predictions yield similar results because degradation reactions are quite slow and therefore the time frame of these measurements is very small compared, say, to the half-life of the degradation (or the time it takes for the assay to reach the value of 50% LC). Although the discussion of half-life calculation is beyond the scope of this handbook, suffice it to say that based on the results of Examples 6 and 7, the half-life for first-order kinetics is about 23 years and for zero-order kinetics (probably less accurate) is 17 years. These values are about 40 times the time frame used in the accelerated (short-term) tests of 6 months (0.5 year). Thus, any kinetic model will work reasonably well.

### 4.4 Preliminary Shelf Life Calculation from Stressed Data

Preliminary shelf life can be calculated based on accelerated results. Regression techniques described in Section 3 applies to the estimation assays obtained from

accelerated data as illustrated in Examples 6 and 7. The procedure is the same as outlined in Section 3, that is, to develop the quadratic equation and determine the shelf life for each of the observed and predicted assay results.

The quadratic equation to determine the shelf life for the observed assay is

$$[90-(99.884-0.2275x)]^2-1.9432^2(0.0408)\left[\frac{1}{8}+\frac{(x-13.5)^2}{1008}\right]=0$$

The reference point for this equation is $x_{ref} = 43.44$ and the initial point to accomplish convergence is $x_{ref} - 1.5$. The root for this equation is $x_R(\text{obs}) = 41.78$ and the shelf life for observed assay results is $x_L(\text{obs}) = 41$ months. This value will be used for comparison purposes.

The quadratic equation for the predicted assay assuming a first-order chemical reaction applies is

$$[90-(99.95-0.2393x)]^2-1.9432^2(0.0032)\left[\frac{1}{8}+\frac{(x-13.5)^2}{1008}\right]=0$$

The reference point for this equation is $x_{ref} = 41.60$ and the initial point to accomplish convergence is $x_{ref} - 0.4$. The root for this equation is $x_R(\text{first}) = 41.17$ and the shelf life for the predicted assay assuming that the chemical reaction is first order is $x_L$ (first) = 41 months.

Now, assuming that the chemical reaction is of zero order, the quadratic equation for the predicted assay is

$$[90-(99.99-0.2515x)]^2-1.9432^2(0.0013)\left[\frac{1}{8}+\frac{(x-13.5)^2}{1008}\right]=0$$

The reference point for this equation is $x_{ref} = 39.74$ and the initial point to accomplish convergence is $x_{ref} - 0.3$. The root for this equation is $x_R(\text{zero}) = 39.49$ and the shelf life for the predicted assay assuming that the chemical reaction is zero order is $x_L(\text{zero}) = 39$ months.

In summary, the shelf life from the actual data is 41 months, that estimated from first-order kinetics is also 41 months, and the one predicted from zero-order kinetics is 39 months. All three values are similar. However, the results reveal that the actual degradation reaction of the drug product is more likely to be of first order.

## 5 Concluding Remarks

A succinct description of the basics of chemical reaction engineering has been presented and its application to the estimation of shelf life has been outlined through examples. These techniques are of crucial importance in NDAs to regulatory agencies such as the FDA. Normally, at the time a new drug application is submitted, not enough data at low temperature are available since long-term studies take years. The tools presented here are the alternative approved by the FDA and ICH.

## APPENDIX COMPUTER PROGRAMS

This Appendix contains four computer programs in MatLab that can be used by the reader to perform typical calculations related to shelf-life estimations.

This program computes the degradation line for single batch-and-save results to be input into the program called shelf life estimation:

```
% Regression Analysis for a single batch
clear, clc
close all
batch=3;
i=3;       % the number of the batch to be computed
n1(1)=7;
n1(2)=6;
n1(3)=6;
N=n1(i);
x=[0 3 6 9 12 18 24 36
   0 3 6 9 12 18 24 36
   0 3 6 9 12 18 24 36];
y=[99.2 97.1 96.1 95.2 93.8 93.1 92.4 0
   98.7 97 96.2 95.1 94.2 93.3 0 0
   102.5 98.9 97.1 95.6 94.1 93.1 0 0];
x1=[x(i,1:n1(i))];
y1=[y(i,1:n1(i))];
x1=x1';
X=[ones(N,1) x1];
Y=y1';
b=inv(X'*X)*X'*Y;
y_est=X*b;
e=Y-y_est;
SSE=e'*e;
SST=Y'*Y-N*(mean(Y)^2);
SSR=SST-SSE;
t=tinv(0.95,N-2)
R2=SSR/SST
MSE=SSE/(N-2)
av_x=mean(x1)
Sxx=x1'*x1-N*(av_x)^2
vec=[b' N MSE av_x Sxx]
save res vec
```

This program computes the degradation line after pooling data from several batches that have common slope and intercept:

```
.
% Regression Analysis for pooled data from several batches
clear, clc
close all
batch=3;
```

```
x=[0 3 6 9 12 18 24 36
   0 3 6 9 12 18 24 36
   0 3 6 9 12 18 24 36];
y=[102.4 98.1 99.2 97.5 95 96.1 95.2 94.3
   101.1 101.2 99 97.2 96.4 95.5 94.3 94.8
   104.1 102.1 99.5 98.1 95.7 94.1 94 93.5];
n(1)=6;
n(2)=7;
n(3)=8;
x1=[x(1,1:n(1)) x(2,1:n(2)) x(3,1:n(3))];
y1=[y(1,1:n(1)) y(2,1:n(2)) y(3,1:n(3))];
x1=x1';
N=sum(n);
X=[ones(N,1) x1];
Y=y1';
b=inv(X'*X)*X'*Y;
y_est=X*b;
e=Y-y_est;
SSE=e'*e;
SST=Y'*Y-N*(mean(Y)^2);
SSR=SST-SSE;
R2=SSR/SST
MSE=SSE/(N-2)
av_x=mean(x1)
Sxx=x1'*x1-N*(av_x)^2
vec=[b' N MSE av_x Sxx]
save res vec
```

This program computes the shelf life for data that have been input in either program 1 or 2:

```
clear, clc
close all
delete it* % para borrar archivos anteriores display('1:
  f = (90-(b0 - b1*x))^2-t^2*MSE*[1/n + (x- av_x)^2/Sxx] ')
flag_fun=input('Enter the number of the function ')
load res
b=vec(1:2);
n=vec(3);
MSE=vec(4);
av_x=vec(5);
Sxx=vec(6)
d=-5; % expiration date
reference=(b(1) - 90)/(-b(2))
if flag_fun==1
   x0=(b(1) - 90)/(-b(2))+d; % expiration date lower value
end
options = optimset('LargeScale','off');
```

```
cont=0;
save contador cont
[x,fval,exitflag,output] = fminunc(@(x)
obj_fun_expiration_date(x,flag_fun),x0,options)
```

This routine is required by program 3, shelf life estimation:

```
function f = obj_fun_expiration_date(x,flag_fun)
load contador
cont=cont+1
save contador cont
load res
b=vec(1:2);
n=vec(3);
MSE=vec(4);
av_x=vec(5);
Sxx=vec(6);
t=tinv(0.95,n-2);
if flag_fun==1
  f = abs((90-(b(1)+b(2)*x))^2-(t^2)*MSE*(1/n+(x-av_x)
    ^2/Sxx));
end
save(['it' num2str(cont)],'f','x','flag_fun')
```

This program performs a statistical test for determining whether or not the slopes and/or intercepts of the degradation lines from several batches are equal:

```
% ANCOVA
clear, clc
close all
batch=3;
x=[0 3 6 9 12 18 24 36
   0 3 6 9 12 18 24 36
   0 3 6 9 12 18 24 36];
y=[99.2 97.1 96.1 95.2 93.8 93.1 92.4 0
   98.7 97 96.2 95.1 94.2 93.3 0 0
   102.5 98.9 97.1 95.6 94.1 93.1 0 0];
n(1)=7;
n(2)=6;
n(3)=6;
N=sum(n);
%%%%%%%%%%%%%%%%%%
% testing slopes
%%%%%%%%%%%%%%%%%%
for i=1:batch
    sy=0;
    sx=0;
    sxy=0;
```

```
    for j=1:n(i)
        sy=sy+y(i,j)^2;
        sx=sx+x(i,j)^2;
        sxy=sxy+x(i,j)*y(i,j);
    end
    Syy(i)=sy-(sum(y(i,1:n(i))))^2/length(y(i,1:n(i)));
    Sxx(i)=sx-(sum(x(i,1:n(i))))^2/length(x(i,1:n(i)));
    Sxy(i)=sxy-(sum(y(i,1:n(i))))*(sum(x(i,1:n(i))))/
      length(y(i,1:n(i)));
end
SxxW=sum(Sxx);
SyyW=sum(Syy);
SxyW=sum(Sxy);
SSEW=SyyW-SxyW^2/SxxW;
for i=1:batch
    sse(i)=Syy(i)-Sxy(i)^2/Sxx(i);
end
SSE=sum(sse);
SS_slope=SSEW - SSE;
k=batch-1;
MS_slope=SS_slope/(batch-1);
MSE=SSE/(N-2*batch);
F_slope=MS_slope/MSE;
F_slope
df_num=batch-1
df_den=N-2*batch
F_cri=finv(0.75,df_num,df_den)
%%%%%%%%%%
% testing intercepts
%%%%%%%%%%
sy=0;
sy1=0;
sx=0;
sx1=0;
sxy=0;
for i=1:batch
    for j=1:n(i)
        sy=sy+y(i,j)^2;
        sx=sx+x(i,j)^2;
        sxy=sxy+x(i,j)*y(i,j);
    end
    sy1=sy1+sum(y(i,1:n(i)));
    sx1=sx1+sum(x(i,1:n(i)));
end
sy;
SY1=sy1^2/N;
SYY=sy-SY1;
sx;
```

```
SX1=sx1^2/N;
SXX=sx-SX1;
sxy;
SXY1=sx1*sy1/N;
SXY=sxy-SXY1;
tx=0;
ty=0;
txy=0;
for i=1:batch
    ty=ty+(sum(y(i,1:n(i))))^2/length(y(i,1:n(i)));
    tx=tx+(sum(x(i,1:n(i))))^2/length(x(i,1:n(i)));
    txy=txy+(sum(x(i,1:n(i))))*(sum(y(i,1:n(i))))/
      length(y(i,1:n(i)));
end
my=sy1^2/N;
mx=sx1^2/N;
mxy=sx1*sy1/N;
Tyy=ty - my
Txx=tx - mx
Txy=txy - mxy
Eyy=SYY-Tyy
Exx=SXX-Txx
Exy=SXY-Txy
SSEr=SYY-SXY^2/SXX
SSEc=Eyy-Exy^2/Exx
SSb0=SSEr - SSEc
MSb0 = SSb0/(batch-1)
MSEc=SSEc/(N-batch-1)
F_b0=MSb0/MSEc
df_num=batch-1
df_den=N-batch-1
F_cri_inter=finv(0.75,df_num,df_den)
```

## REFERENCES


1. Expiration Dating, Code of Federal Regulations, Title 21, Vol. 4, Part 211, 21CFR211.137.
2. American Medical Association (AMA) (2001), Report 1 of the Council on Scientific Affairs (A-01), Pharmaceutical expiration dates, available: http://www.ama-assn.org/ama/pub/category/print/13652.html, accessed on Feb. 19, 2007.
3. Bennett, B., and Cole, G., Eds. (2003), *Pharmaceutical Production. An Engineering Guide*, Institution of Chemical Engineers (IChemE), Rugby, Warwickshire, United Kingdom.
4. Matthews, B. R. (1999), Regulatory aspects of stability testing in Europe, *Drug. Dev. Ind. Pharm.*, 25, 831–856.
5. Dietz, R., Feilner, K., Gerst, F., and Grimm, W. (1993), Drug stability testing: Classification of countries according to climatic zone, *Drugs Made in Germany*, 36, 99–103.

6. U.S. Food and Drug Administration (FDA) (2001), Stability testing of new drug substances and products, ICH Q1A, FDA, Washington, DC.
7. Fairweather, W. R., Lin, T. Y., and Kelly, R. (1995), Regulatory, design, and analysis of complex stability studies, U.S. Food and Drug Administration, *J. Pharm. Sci.*, 84, 1322–1326.
8. U.S. Food and Drug Administration (FDA) (2004), Evaluation of stability data, ICH Q1E, FDA, Washington, DC.
9. U.S. Food and Drug Administration (FDA) (2003), Bracketing and matrixing designs for stability testing of new drug substances and products, ICH Q1D, FDA, Washington, DC.
10. Montgomery, D. G. (2004), *Design and Analysis of Experiments*, 6th ed., Wiley, New York.
11. Chen, J. J., Ahn, H., and Tsong, Y. (1997), Shelf-life estimation for multifactor stability studies, *Drug Inform. J.*, 31, 573–587.
12. Montgomery, D. C., Peck, E. A., and Vining, G. G., (2007), *Introduction to Linear Regression Analysis*, 3rd ed., Wiley, New York.
13. Reklaitis, G. V., Ravidran, A., and Ragsdell, K. M. (1983), *Engineering Optimization, Methods and Applications*, Wiley, New York.
14. Bancroft, T. A. (1964), Analysis of inference for incompletely specified models involving the use of preliminary test(s) of significance, *Biometrics*, 20, 427–442.
15. Snedecor, G. W., and Cochran, W. G. (1989), *Statistical Methods*, 8th ed., Iowa State University Press, Ames, IA.
16. Wang, S. G., and Chow, S. C. (1993), *Advanced Linear Model: Theory and Applications*, Marcel Dekker, New York.
17. Ramirez-Beltran, N. D., and Olivares, L. (1998), Drug Shelf-life Estimation Using Clustering Techniques. *Proceeding of the Computing Research Conference CRC'98*, edited by the University of Puerto Rico, Electrical and Computer Engineering Department. Mayaguez Puerto Rico, pp. 70–73.
18. Chow, S.-C., and Liu, J.-P. (1995), *Statistical Design and Analysis in Pharmaceutical Science*, Marcel Dekker, New York.
19. Shao, J., and Chow, S-Ch. (1994), Statistical inference in stability analysis, *Biometrics*, 50(3), 753–763.
20. Chow, S. C., and Shao, J. (1991), Estimating drug shelf-life with random batches, *Biometrics*, 47, 1071–1079.
21. Ruberg, S., and Stegeman, J. W. (1991), Pooling data for stability studies: Testing the equality of batch degradation slopes, *Biometrics*, 47, 1059–1069.
22. Ruberg, S., and Hsu, J. (1992), Multiple comparison procedures for pooling batches in stability studies, *Technometrics*, 34, 465–472.
23. Waterman, K. C., and Adami, R. C. (2005), Accelerated aging: Prediction of chemical stability of pharmaceuticals, *Int. J. Pharm.*, 293, 101–125.
24. Gil-Alegre, M. E., Bernabeu, J. A., Camacho, M. A., and Torres-Suárez, A. I. (2001), Statistical evaluation for stability studies under stress storage conditions, *Il Farm.*, 56, 877–883.
25. Oliva, A., Llabrés, M., and Fariña, B. (2006), Data analysis of kinetic modeling used in drug stability studies: Isothermal versus nonisothermal assays, *Pharm. Res.*, 23, 2595–2602.
26. Levenspiel, O. (1998), *Chemical Reaction Engineering*, 3rd ed., Wiley, Ney York.

# EFFECT OF PACKAGING ON STABILITY OF DRUGS AND DRUG PRODUCTS

EMMANUEL O. AKALA
*School of Pharmacy, Howard University, Washington, DC*

**Contents**

1 Introduction
  1.1 Rationale for Concern with Stability of Drugs and Drug Products
2 Factors Influencing Stability of Drugs and Drug Products
  2.1 Moisture, Hydrolysis, and pH
  2.2 Oxygen and Oxidation
  2.3 Light
  2.4 Temperature
  2.5 Microbes
  2.6 Active Pharmaceutical Ingredients and Excipients
3 Drug Packaging and Packaging Materials
  3.1 Introduction
  3.2 Materials of Fabrication of Packaging Components
4 Effect of Packaging on Drug Product Stability
  4.1 Introduction
  4.2 Solid Dosage Forms
  4.3 Nonsterile Liquid Dosage Forms
  4.4 Sterile Liquid Dosage Forms
  4.5 Container Extractables and Leachables and Drug Product Stability
  4.6 Biotechnological Products
  4.7 A Look into the Future of the Effects of Packaging on Stability of Drug Products
References

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

# 1 INTRODUCTION

Stability is an essential attribute of drug products as evidenced by continuing national and international governmental interventions in the form of regulations. It is required that stability testing should demonstrate physical and chemical stability of a drug and its product(s) at a variety of environmental conditions such as temperature, humidity, and light [1]. Moreover, International Conference on Harmonization (ICH) guidelines provide requirements for stability studies [2, 3]. Pharmaceutical and biopharmaceutical companies are also making frantic efforts to improve on the development and manufacture of drug products with adequate stability profiles. All these activities (by governments and pharmaceutical and biopharmaceutical companies) are for the following purposes [4]: to deliver drug products of high quality (efficacious and safe) to the final consumer (the patient), to meet the legal requirements stipulated by the regulatory agencies, to provide the assurance to the manufacturer that the drug products will perform as intended, and to serve as a baseline for future drug development efforts.

## 1.1 Rationale for Concern with Stability of Drugs and Drug Products

***Loss of Potency*** The effectiveness of any drug product (dosage form) is a function of its ability to deliver the required amount of the active therapeutic moiety to the biophase (site of action) for the intended length of time to achieve the purpose of therapy. Loss of the active ingredient in the dosage form, consequent upon chemical degradation, may result in poor performance of the drug product (i.e., the serum drug concentration may not reach the minimum effective concentration to elicit the desired therapeutic response). The potency of any drug product is not expected to remain constant ad *infinitum*; however it is expected that the quality of the drug product [the stability of the active pharmaceutical ingredients (APIs) among other attributes] is maintained until it either gets to the patients or reaches the expiration date, which is defined as the date placed on the immediate container label of the drug product that designates the date through which the product is expected to remain within specifications [5]. The specifications (the period within which the efficacy, safety, and esthetics of the drug product can be assured) are often defined in terms of the shelf life of the drug product. When the chemical degradation of the APIs is found to be the major contributor to the degradation process of the drug product, the shelf life is considered the time that elapses before the drug content falls below 90% of the label claim, with the assumption that the drug product is stored as directed on the label [4, 6].

***Toxic Decomposition Products*** In situations where the content of the API in a drug product is well above 90% as indicated above, the formation of toxic degradation products within the shelf life (which may cause untoward effects to the patients) may warrant the reassignment of a different expiration date or recall of the drug product in question. Consequently, the pharmaceutical industry is often concerned with both the amount as well as the nature of the degradation products. The formation of toxic products is particularly problematic with protein drugs which may maintain therapeutic activity after deliberate modification or pertubation of molecular structure in a domain removed from that associated with therapeutic activity

but can result in acquired immunogenicity [4]. A discussion of the examples of products of drug degradation which are more toxic than the original (parent) drugs has been given by Guillory and Poust [7]. It has been reported that the concerns for the potential increase in the amount of toxic degradation products informed the reluctance of the regulatory agencies to approve the design of new drug products with stability overages [4].

***Poor Bioavailability*** The bioavailability of a drug product is customarily defined in terms of the amount of API delivered to the blood (plasma concentration) and the rate at which it is delivered. For routes of administration other than intravascular, the extent of absorption is very important in that all the drug administered may not have a chance to get to the plasma and then get carried to the biophase. One of the most important problems of drug absorption and oral bioavailability is when the drug is not delivered from the drug product over an appropriate time frame (in solution form) to the sites in the gastrointestinal (GI) tract where it is well absorbed (i.e., problem of getting the drug into solution). This problem can be caused by changes in the quality attributes of the drug product on storage (especially the dissolution rate), though the potency is still acceptable and no toxic degradation product has been formed. The changes in quality attribute can be traced to changes in the physicochemical properties of the excipients with aging. The result is ineffectiveness of the drug product on administration.

***Microbial Contamination*** The microbiological quality of all drug products is receiving a lot of attention in today's drug product design, unlike in the past, when the concern was for drug products that must remain sterile: parenteral and ophthalmic products. The control of total bioburden and the exclusion of pathogenic microbes are considered desirable [4]. For proper maintenance of the microbiological quality of drug products the quality of the raw materials and the manufacturing facility must be controlled and assured. Further, in assessing the suitability for use of the container closure system, protection is considered to be one of the most important considerations. Thus the container closure system is expected to protect the drug product from causes of degradation such as light, temperature, loss of solvent, oxygen, water vapor, and microbial contamination [8]. Package integrity must not be compromised from the manufacturer through the distribution channels and to the final consumer: the patient.

***Changes in Physical Appearance of Drug Product*** Some of the physical changes may affect the efficacy of the drug products, while in others the potency may not change with the physical change but may affect pharmaceutical elegance such that patient acceptability suffers. Loss of an emulsifying agent may lead to breaking of an emulsion (phase separation) and caking of a suspension will lead to failure in resuspending the suspension on shaking. These changes will affect dose uniformity in the prescribed dosage regimen. Mottling of tablets may not affect the potency of the drug product, but such tablets will not be acceptable to patients and may affect compliance with dosage regimen. Moreover drug–excipient interactions may lead to changes in the physical appearance of the drug product as exemplified by the interaction between lactose and the amino functional group in drug which often

leads to the formation of yellow color on the surface of tablets with aging. Though the potency is still intact, such tablets will not appeal to patients.

***Loss of Functional Stability*** It is one thing to design a very effective drug formulation; it is another important thing for it to be delivered properly by the container closure system during the shelf life of the drug product. Thus the performance of the container closure system for drug products (defined as the ability to function in the manner for which it was designed, i.e. ability to deliver the dosage form in the amount or at the rate described in the package insert [8]) is an important part of drug product stability. Reports have shown that early transdermal patches had problems of skin adhesion with aging (they showed loss of adhesion with the tendency to fall off the patient's skin) [4]. Other drug products for which consistency in drug delivery should be monitored over the entire shelf life are as follows: prefilled syringes, a metered tube, a dropper, a spray bottle, a powder inhaler, and a metered-dose inhaler.

## 2 FACTORS INFLUENCING STABILITY OF DRUGS AND DRUG PRODUCTS

The preceding section has elaborated on the importance of the stability of drugs and drug products. The general principle in the design and manufacture of drug products is that quality (defined as the physical, chemical, microbiological, biological, bioavailability, and stability attributes that a drug product should maintain if it is to be deemed suitable for therapeutic and diagnostic use [8]) should be built into all the processes or stages of drug manufacture rather than be inspected into the final product ready for distribution. Consequently, knowledge of the factors that can promote all forms of drug and drug product instability such as chemical degradation (formation of new chemical entities), physical degradation (drug loss without the production of distinctly different chemical products), and biological degradation (the most important is microbial degradation, though cognizance should be taken of the fact that nonmicrobiological organisms such as ants and rats may be important) during the shelf life of the drug product is very important. The awareness of these factors will help not only in the preformulation and formulation stages of drug development but also in the selection and appraisal of the protective capability of the appropriate container closure system for the drug product in question.

### 2.1 Moisture, Hydrolysis, and pH

***Liquid Dosage Forms*** Most hydrolytic degradations take place in the presence of moisture. Thus hydrolysis together with the influence of pH on hydrolysis as a route of degradation of drugs is important for all types of dosage forms: liquid, solid, semisolid, and gases. Recognition of the chemical groups which are susceptible to hydrolysis will aid the drug formulation scientist to determine *a priori* the paths to take to achieve drug products of maximum stability. Examples of such chemical groups are as follows: ester linkages (acetyl salicylic acid, procaine, teracaine, and physostigmine), amides (cinchocaine, ergometrine, and chloramphenicol), lactams

(penicillins, cephalosporins, nitrazepam, and chlordiazepoxide), imides (glutethimide and ethosuximide), and lactones (pilocarpine and spironolactone) [7, 9]. The rate at which hydrolysis proceeds in each of the chemical groups highlighted above is a function of the chemical environment of the chemical group within the drug molecule. Substituent groups can exert electronic, steric, or hydrogen-bonding effects known to affect the susceptibility of the chemical groups to hydrolytic degradation. In fact, modification of drug chemical structure to control drug stability (though caution should be taken not to alter therapeutic efficacy) using appropriate substituents has been used to solve stability problems. The concept of Hammett linear free-energy relationship for the effects of substituents on the rates of aromatic side-chain reactions such as hydrolysis of esters has been reported and the principle has been used to produce and select the best substituents for allylbarbituric acids of optimum stability [9, 10].

While laboratory experiments together with the knowledge of chemical kinetics can be used to formulate a solution drug product at a pH of optimum stability (using appropriate buffers) for a drug whose hydrolytic degradation is catalyzed by hydrogen ion (specific acid catalysis) or hydroxyl ion (specific base catalysis), the known influence of buffer components (called general acid–base catalysis) in buffered drug products should be considered. Failure to evaluate complete stability of the drug by determining the catalytic coefficients for specific acid catalysis and base catalysis and the catalytic coefficients of buffer components will result in poor determination of the conditions for maximum stability of the drugs in solution. Furthermore, the ionization of the drug should be taken into account in the determination of the pH–rate profile for a complete determination of the conditions of maximum stability. Other strategies that have been used to stabilize drug products in solution involve the use nonaqueous solvents such as mono- and polyhydric alcohols (e.g., ethanol glycerin, and propyleneglycol) which can change the dielectric constant of the system. Micellar solubilization by surfactants and reduction in solubility (as exemplified by increase in the stability of penicillin in procaine penicillin in the presence of additives such as citrates, dextrose, sorbitol, and gluconate) are capable of reducing hydrolysis of drugs in solution [9].

The influence of packaging on the stability of drugs and drug products will be examined in greater detail later in this chapter, but suffice it to state here that the permeability of the packaging materials to gases and liquids should be considered for all types of dosage forms since it can lead to decomposition of the drug products due to oxidation and hydrolysis. In the case of solution, leachables from container closure systems can alter the predetermined stability of liquid dosage forms. For example, leaching of dioctyl phthalate, a plasticizer used in polyvinyl chloride (PVC) plastics, into intravenous solutions containing surfactants has been reported [11, 12]. The potential adverse effect of the leached compound is twofold: on the safety of the patient and on the stability of the drug administered intravenously. Under the consideration for the suitability for the intended use of any proposed packaging system, as indicated by the U.S. Food and Drug Administration (FDA), compatibility (packaging components are regarded compatible with the dosage form if there is no sufficient interaction to cause unacceptable changes in the quality of either the dosage form or the packaging component, i.e., leachable-induced degradation, precipitation, or changes in pH) and safety (packaging components should be constructed of materials that will not leach harmful or undesirable amounts of substances

to which a patient will be exposed when being treated with the drug product) are given prominent positions [8]. It is indicated clearly that for a drug product such as an injection, inhalation, ophthalmic, or transdermal, a comprehensive study should be carried out: extraction studies on the packaging component to determine which chemical species (and their concentrations) may migrate into the dosage form and a toxicological evaluation of those substances which are extracted to determine the safe level of exposure via the label-specified route of administration [8]. When the leached compound is an electrolyte, it can affect the ionic strength of the solution and hence the rate of degradation of the drug, which can be predicted by the Brönsted–Bjerrum model [9].

Liquid dosage forms which are disperse systems (colloidal, i.e., microspheres, nanoparticles, and micelles; suspensions; and emulsions) often contain preservatives which are methyl, ethyl, propyl, and butyl esters of *para*-hydroxybenzoic acid in various combinations. A typical example is the antacid suspensions which have high pH values which make the esters of the preservatives susceptible to hydrolysis. One way to circumvent this problem is to use several preservatives in combination with the hope that some quantities of the preservatives will remain to prevent the suspension from microbial attack. A report showing the assay of the four esters and the parent acid (one of the decomposition products) in drug products in which all the preservatives were used has been given [13].

***Semisolid Dosage Forms*** The nature of the base (vehicle) used for the fabrication of semisolid dosage forms affects their hydrolytic stability. Increased degradation of benzylpenicillin sodium in hydrogels of various natural and semisynthetic polymers has been reported [14]. Also at pH 6 in Carbopol hydrogels, the percentage of undecomposed pilocarpine at equilibrium is a function of the apparent viscosity of the medium [15].

***Solid Dosage Forms*** Moisture can affect the chemical stability as well as the physical stability of solid dosage forms. Reports have shown that the amount of moisture adsorbed by tablets in blister packages increased with increasing humidity and it resulted in decreased mechanical strength of the tablets [16]. Storage of prednisone and erythromycin tablets in moisture-permeable packaging changed drug release from the tablets [17, 18]. Drug release from enteric-coated and sugar-coated tablets was more susceptible to the effect of humidity than that from film-coated tablets. For example, storage of sugar-coated tablets changed the disintegration time, leading to increased or decreased dissolution rate [19]. Storage of two chloramphenicol capsules at high humidity prolonged the disintegration time and decreased the release rate [20]. Further, decrease in drug release from ampicillin capsules during storage at high humidity has been reported; it was attributed to the agglomeration of drug particles caused by moisture [21]. Gelatin-coated acetaminophen tablets exhibited a marked decrease in dissolution rate during storage at high humidity [30°C, 80% relative humidity (RH), and the effect was moderated by the addition of pancreatin to the dissolution medium. Pancreatin was believed to act by cleaving the cross-linked gelatin [22].

When solid dosage forms such as tablets adsorb moisture, drug present on the surface will be dissolved (if it is soluble). The drug in solution on the surface of the tablet will be subject to hydrolytic decomposition, and the process will be influenced

by the pH of the solution. It has been reported that an increase in the water vapor pressure significantly increases the decomposition of aminosalicylic acid [23]. A study on water permeation through rubber closures of injection vials indicates that rubber closures with a low permeability are often able to take significant amounts of water [24]. Thus freeze-dried products, which are highly hygroscopic and highly reactive with water, must be adequately protected against uptake of water in order to ensure the chemical stability of the active ingredient in the vial.

## 2.2 Oxygen and Oxidation

***Semisolid and Solid Dosage Forms*** Whether the oxidative degradation process is by auto-oxidation (an uncatalyzed reaction which proceeds slowly under the influence of molecular oxygen) or by chain processes which involve initiation, propagation, and termination reactions, molecular oxygen is very important. Consequently, every effort should be made to determine and control oxygen concentration in an aqueous solution. Florence and Attwood [9] have given an oxidation scheme involving a chain reaction. The description of the scheme is in order to provide an appreciation of the influence of the interaction of molecular oxygen, components of the solution and even leachables from the packaging materials on oxidation. It is believed that initiation can take place through the free radicals formed by organic compounds consequent upon the action of light, heat, or transition metals (e.g., copper and iron) present in trace amounts in the buffer. The free radicals readily combine with molecular oxygen in the propagation step to form a peroxy radical which then abstracts hydrogen from a molecule of organic compound to form a hydroperoxide and then create another free radical. The termination occurs when the free radicals are destroyed by inhibitors or by side reactions which break the chain.

The presence of certain functional groups is known to make some drugs highly susceptible to oxidative degradation [7, 9]. The phenol functional group present in steroids is sensitive to oxidation. The ether group is susceptible to oxidation as reported for econazole nitrate and miconazole nitrate. Catecholamines such as dopamine and isoproterenol are susceptible to oxidation. Phenothiaxines possess a thioether functional group which is oxidized to sulfoxide in the presence of water. Many drug molecules possess carbon–carbon double bonds which can be easily attacked by peroxyl radicals, leading to oxidative degradation. In fact, the double bond is highly susceptible to singlet oxygen (which is highly oxidizing); the singlet oxygen is believed to form from the ground state of oxygen called triplet oxygen when excited by light. Consequently, the general term oxidation is more than mere exposure of a susceptible drug molecule to oxygen but also exposure to conditions that favor oxidation such as photolysis [6]. Amphotericin is a polyene antibiotic with seven conjugated double bonds. It is oxidized by peroxy radicals with a loss of activity and aggregation [25]. Addition of peroxyl radicals to simvastatin can result in oxidation with the formation of polymeric peroxides [26]. Carboxylic acid is another group highly susceptible to oxidation: The rate of oxidation of ascorbic acid depends on oxygen concentration [27].

Literature on pharmaceutical sciences is replete with oxidative degradation of drugs in solution. 5-Aminosalicylic acid undergoes oxidation and the product of oxidation forms polymeric compounds [28]. Further, morphine has been reported to be oxidatively degraded in solution [29] as well as hydrocortisone [30].

***Semisolid and Solid Dosage Forms*** Since most oxidative reactions occur in solution, there is not much effort to investigate the effect of oxidative degradation in solid dosage forms. Auto-oxidation of tetrazepam tablets has been described [31].

## 2.3 Light

Photochemical decomposition is an important route of chemical degradation of drugs. Generalization on or a priori prediction of the effect of light on drug is very difficult because of the strong dependence of the decomposition on the spectral properties of drugs as well as the spectral distribution of the source of light [6]. Nevertheless, certain trends have been reported which might provide guidance on the effect of light on certain drug molecules so that efforts can be made at the early stage of drug development to circumvent the problems. Molecules with saturated bonds are not capable of interacting with visible or near-ultraviolet light. Those molecules with $\pi$ electrons [aromatic hydrocarbons (nitroaromatic and aryl halide functional groups), heterocyclic aromatic compounds, aldehydes, ketones, sulfides, alkenes, and polyenes] absorb light throughout the wavelength range of visible or near-ultraviolet light and are very susceptible to photochemical decomposition. Sunlight can potentially affect drugs capable of absorbing light at wavelengths below 280 nm, and those capable of absorbing light strongly at wavelengths greater than 400 nm can be potentially degraded by light in both sunlight and room light [7, 32].

The importance of the excipients used in the development of drug products vis-à-vis photochemical decomposition has been stressed [9]. Photochemical degradation of drugs may occur by the direct interaction of the drug with the light of a particular wavelength (primary photochemical reaction) or the excipeints may absorb light radiation and then transfer the energy to the drug molecule (which cannot absorb the incident light) to cause decomposition. Such excipients are called photosensitizers. In fact, the ICH Harmonized Tripartite Guideline on Stability Testing (Photostability Testing of New Drug Substances and Products) [33] recognized the role that excipients in finished drug products can play on photostability and recommends that photostability testing should be carried out as follows: tests on drug substance, tests on the exposed drug product outside of the immediate pack, and, if necessary, tests on the drug product in the immediate pack and tests on the drug product in the marketing pack. As indicated earlier, oxidative degradation may be initiated by light; thus all the factors affecting the stability of drugs and drug products do not occur in isolation. Stabilization of drugs against photochemical decomposition often involves the use of colored containers (amber glass excludes light of wavelength less than 470 nm and can give a good measure of protection to drugs sensitive to ultraviolet light) and storage in the dark. Incorporation of ultraviolet absorbers in polymer film for tablet coating is another method of protecting drugs from photochemical degradation [34].

***Liquid Dosage Forms*** Sodium nitropruside in aqueous solution for injection will remain stable for up to one year if protected from light; however, its shelf life is about 4 h when exposed to normal room light [35]. It has been reported that uric acid increases the photostability of sulfathiazole sodium in solutions [36]. Further *dl*-methionine increased the photostability of ascorbic acid in solution [37]. The

influence of light on the stability of molsidomine in infusion fluids has been reported [38]. It has been shown that the photochemical degradation of nifedipine is a function of light intensity (light source: high-pressure mercury lamp, sunlight, and fluorescent lamp) at room temperature: The amount of photodegraded nifedipine was proportional to the number of incident photons [39]. Stabilization of the photochemical degradation of daunorubicin in solution (at 290–700 nm) by the addition of various colorants has been reported: scarlet GN, amaranth, ponceau 6 R, and tartrazine [40].

***Solid Dosage Forms*** A complex relationship was found between the discoloration rate of sulfisomide in tablets irradiated with a mercury lamp in comparison with ultraviolet light intensity [41]. The functional attributes of capsules such as drug release rate can change following the interaction of dyes and gelatin in capsule shell under the influence of light. The changes were reported to be enhanced by the interaction of light and humidity [42]. The solid-state photodecomposition of the drug, tretinoin tocoferil, has been shown to be highly temperature dependent. It was found that the rate constant for the photodegradation of tretinoin tocoferil has an Arrhenius-type temperature dependence [43]. A report has shown the stabilization of indomethacin coloration by the addition of titanium dioxide to gelatin capsule shells [44]. Moreover, incorporation of synthetic iron oxides resulted in the stabilization of tablets of nifedipine and sorivudine against photodegradation [45].

## 2.4 Temperature

The temperature dependence of the rate constant for a chemical reaction is often expressed using the Arrhenius equation:

$$k = Ae^{-E_a/RT} \tag{1}$$

where $k$ is the kinetic rate constant, $A$ is the preexponential factor, $E_a$ is the activation energy, $R$ is the ideal gas constant, and $T$ is the temperature. As predicted by the equation, increasing the temperature increases the rate of reaction. Consequently, temperature is of considerable importance in consideration of the stability of drugs and drug products. Further, if the Arrhenius plot is linear, the degradation rates obtained at some temperature levels can be used to predict the rate of degradation at other temperatures. This equation is strictly valid for reactions that take place in solution, and it assumes there is no change in the degradation mechanism or the order of reaction; nevertheless, it has been used to predict the stability of dosage forms. For solid dosage forms, the effect of temperature on stability is complicated because of the effect of temperature on the excipients and the likely change in the properties of the excipients which may affect the stability of active ingredient. Though the Arrhenius-type equation has been used to model the dependence of reaction rate on temperature for solid dosage forms, the activation energy obtained could not be expected to have the same meaning as the one obtained with the solution of the drug [9].

Where the decomposition shows an approach to equilibrium, as found for vitamin E tablets [9, 46], the equilibrium concentrations of products of degradation and reactants are obtained at a series of temperatures. Then the logarithm equilibrium

constant $K$ is plotted against the reciprocal of temperature according to the van't Hoff equation to model the dependence of temperature on drug breakdown:

$$\ln k = -\frac{\Delta H}{RT} + \text{const} \tag{2}$$

***Liquid Dosage Forms*** The arrhenius equation has been used to model the discoloration of a liquid multisulfa preparation and the decomposition of liquid multivitamin preparations [47–49]. The degradation of codeine sulfate in solutions at several temperatures has been reported [50]. Further, the influence of pH on the Arrhenius plots for the hydrolysis of ciclosidomine has been studied [51]. The stability of a clofibride emulsion for oral administration was reduced (aggregation of the emulsion) when the storage temperature was increased from 25°C to 40°C. Moreover, storage at 4°C resulted in phase separation of the same emulsion [52]. It is believed that the stability of a vaccine can be predicted by estimating the loss of antigenicity during long periods of storage at different temperatures through accelerated stability studies. Though under certain conditions and for some vaccine products the Arrhenius equation can be used to make a prediction of real storage stability under accelerated stability data, at high temperatures many degradation processes of biomolecules are complicated by unfolding or other conformational or structural changes. Below the freezing point, the degradation rate can suddenly increase or decrease, showing no linear correlation with temperature [53]. Further the degradation rate constant $k$ is not the only factor determining the residual antigenicity $P$t of a vaccine. The time $t$ during which a vaccine is stored at a given temperature and the initial antigenicity of the vaccine $P_0$ also have an influence which can be expressed by the relationship [53, 54]

$$k = \frac{P_0 - P_t}{t} \tag{3}$$

***Semisolid Dosage Forms*** Heating easily brings about phase changes in semisolid drug products; consequently it is not feasible to use high temperatures to study the kinetics of degradation and hence the prediction of stability. Thus stability is often evaluated at the temperature in which the formulation is stored. This practice often takes a long time and stability problems may not be detected until the studies have been in progress for several months or years [55]. Storage temperature can bring about changes in the hardening of suppositories such that the time required for the suppository to melt is affected. It has been shown that the hardening effect increased with increased temperature up to 25°C but decreased at higher temperatures owing to partial melting of the suppository base [6].

***Solid Dosage Forms*** The decomposition of aspirin in tablets made in a microcrystalline cellulose base was modeled by first-order kinetics and rate constants obtained adhered to an Arrhenius equation [56]. A tablet formulation containing polyvinylpyrrolidone (a disintegrant) showed a change in drug release when the temperature of storage was increased from 23 to 65°C [57]. Changes in the dissolution rate of hydrochlorothiazide tablets at room temperature have been estimated from changes

observed at 37, 50, and 80°C [58]. A report of the interaction of certain furoic acids when tableted with microcrystalline cellulose to cause the formation of carbon monoxide has been given [59]. The interaction was fast at 55°C and caused the tablets to crumble but was less pronounced at room temperature.

### 2.5 Microbes

Microbial stability is one of the factors affecting drug product reliability. Under design and interpretation of stability studies, the Center for Drugs and Biologics [5] has elaborated on the microbial quality of drug products. It is clearly indicated that drug products containing preservatives to control microbial contamination should have the preservative content monitored at least at the beginning and end of the projected expiration dating period of the product. Also adequacy of the preservative system under conditions of use for multiuse containers should be considered. Further, nonsterile products that require control of microbial quality and that do not contain preservatives should be tested at specific intervals throughout the projected dating period according to the release specification for bioburden. Topical preparations should also be tested for the presence of pathogens that may be identified as potentially harmful. These statements are also reinforced under the consideration of the suitability of a packaging system: The container closure system should provide the dosage form with adequate protection from factors that can cause degradation; microbial contamination is one of the factors listed [8].

The U.S. Pharmacopeia/National Formulary USP 28/NF23 [1] refers to the stability of a drug product as the extent to which a product remains within specified limits and throughout its period of storage and uses the maintenance of the same properties and characteristics that it possessed at the time of manufacture. Microbial stability (defined as the retention of sterility or resistance to microbial growth according to the specified requirements; antimicrobial agents that are present retain effectiveness within the specified limits) is one of the main types of stability recognized by USP 28/NF23. Thus in addition to efforts to stabilize drug products against chemical and physical decompositions brought about by environmental factors already discussed, liquid and semisolid drug products must be protected against microbial attack where necessary. The antimicrobial preservatives added to pharmaceutical products should be evaluated during stability studies by either chemical methods or microbial challenge test. The microbial challenge test is recommended at the last time point during stability testing.

***Liquid Dosage Forms*** Both sterile products (consideration for the maintenance of sterility and preservative efficacy changes) and nonsterile products (consideration for the proliferation of microorganisms) are important. Ophthalmic and parenteral liquid products are often prepared in a sterile condition; nevertheless appropriative preservatives are often added so that they can remain sterile during the period of storage, distribution, and use. Further, other drug products that are not required to be sterile and therefore are not sterilized during their manufacture and may contain ingredients which can make then susceptible to microbial growth are also protected using antimicrobial preservatives [60]. Examples are aqueous products such as emulsions and suspensions. The use of methyl, ethyl, propyl, and butyl esters of *para*-hydroxybezoic acid in various combinations in drug products

has been mentioned previously [13]. The desire to allow certain amounts of the preservatives to remain in the suspension for preservation often informs the use of the preservatives in combination.

***Semisolid Dosage Forms*** The FDA requires that all ophthalmic ointments should be sterile. It is also required that a suitable preservative or mixture of preservatives to prevent the growth of microorganisms must be added to ophthalmic ointments that are packaged in multiple-use containers; chlorobutanol and methyl- and propyl-*para*-hydroxybezoic acid are often used as preservatives in ointments [61]. Since the sterilization of finished ophthalmic ointment is fraught with difficulty that centers on the stability of the components, aseptic methods of processing are normally used to achieve the sterility requirements [62]. Ointments for topical applications, though not required to be sterile, are expected to meet acceptable standards for microbial content [1]. Creams and gels are often formulated to contain appropriate antimicrobial preservatives, as the high water content favors the growth of microbes.

***Solid Dosage Forms*** At first sight one will expect solid dosage forms to be free of microbial contamination. However, reports have shown that tablet contamination could arise from the raw materials. Studies were carried out on the effects of the production, environment, method of production, and microbial quality of the starting materials on microbial loading during the various stages of tablet production [63]. High levels of microbial contamination were seen during the wet granulation process, which reduced significantly during the drying process.

## 2.6 Active Pharmaceutical Ingredients and Excipients

### *Physicochemical Properties of APIs*

*Hygroscopicity* Hygroscopicity is the amount of moisture absorbed by a powder when it is exposed to an atmosphere of a known relative humidity. It is part of the preformulation program and its importance lies in the use of the information gained to decide whether or not a particular salt of the drug in question could be used for a dosage form design. Report has shown that flurazepam is used as a monosulfate rather than the disulfate which however has many other desirable characteristics: It is so hygroscopic that it will remove water from a hard-shell capsule and then render it very brittle [64]. Knowledge of the hygroscopicity of the APIs can aid in the selection of packaging materials.

*Crystalline State or Amorphous State and Polymorphism* Solid drug particles occur as pure crystalline substances of definite indentifiable shape or as amorphous particles without definite structure. The energy required for a molecule of a drug to leave the lattice in a crystal is very much greater than the energy required in an amorphous powder. Drugs in the crystalline state have low reactivity. A linear relationship has been found between the solid-state degradation rate constant of various vitamin A derivatives and the inverse of the melting point [65]. Moreover, results of studies on the relationships between degradation rate and crystallinity of β-lactam antibiotics such as cefazolin indicate that a drug with low crystallinity tends to have decreased chemical stability [66].

Polymorphism is the phenomenon by which a solid particle may exist in different crystalline forms called polymorphs. Medicinal compounds form different types of crystals, depending on the conditions (temperature, solvent, time) under which crystallization is induced. The molecules of the drug exhibit different space–lattice arrangements in the crystalline form from one polymorph to another. It should be emphasized that polymorphism exists only in the solid state. Only one form of a pure drug substance is stable at a given temperature and pressure with the other forms, called metastable forms, converting in time to the stable crystalline form. It is common for a metastable form of a medicinal agent to change form even when present in a completed pharmaceutical preparation, although the time required for a complete change may exceed the normal shelf life of the product itself. Although the drug is chemically indistinguishable in each form, polymorphic forms differ significantly with respect to a number of physical properties such as density, melting point, solubility, stability, and dissolution characteristics, which are of prime importance to the proper development of the dosage forms containing the drug. Solid-state hydrolysis of carbamazepine from needle-shaped crystals with a high crystalline order was found faster than that of beam-shaped and prismatic forms and the reactivity of carbamazepine to light is strongly dependent on the crystalline form of the drug [67, 68], and a similar report has been given for the photodegradation of furosemide [69].

*Modification of Chemical Structure of Drug* The use of a Hammett linear free-energy relationship to investigate the effects of substituents on the rates of aromatic side-chain reactions such as hydrolysis of esters has been alluded to earlier vis-à-vis attainment of optimum stability [9, 10]. Degradation of erythromycin under acidic pH conditions is inhibited by substituting a methoxy group for the C-6 hydroxyl as found for the acid stability of clathromycin, which is 340 times greater than that of erythromycin [70].

*Vapor Pressure* Active pharmaceutical ingredients with sufficiently high vapor pressures can become lost by their volatilization through the containers such that the stability and content uniformity suffer. Further such compounds can interact with other drug molecules and packaging components [71]. Nitroglycerin is notorious for this behavior and special packaging materials are needed for dispensing sublingual nitroglycerin tablets. When unstabilized nitroglycerin sublingual tablets were stored in enclosed glass containers, the high volatility of the drug gave rise to redistribution of nitroglycerin among stored tablets with concomitant deterioration in the uniformity of the tablets on storage [72].

***Pharmaceutical Excipients*** Drugs are rarely administered to humans as pure chemical compounds. What is given is a preparation containing the drug and the materials of formulation called excipients. These excipients are added for various purposes to ensure adequate performance of the drug products. However, some of the excipients can exert deleterious effects on the stability of drug products. The degradation of codeine has been reported to be susceptible to the influence of buffer: Its hydrolytic rate constant in phosphate buffer at pH 7.0 is about 20 times faster than in buffered solution at the same pH [9]. Further, various phosphate species were found to enhance the degradation of bezypenicillin [73], cefadroxil

[74], carbenicillin [75], and spironolactone [76]. Some drugs may not be capable of absorbing light directly to undergo photolysis, but the excipients present in the formulation may absorb light radiation and transfer the absorbed energy to the drug to cause degradation [9]. Furthermore, transition metals such as copper and iron which are present in trace amounts in buffer are capable of initiating the chain process in oxidative degradation. Preservatives may affect the stability of dosage forms. Preservation of zinc insulin (in a multidose vial) with phenol can cause physical instability of the suspension. In the case ophthalmic pharmaceutical products, all raw materials used in the preparation should be of the highest quality available. Consequently, the practice by some drug manufacturers is to compound all ophthalmic drugs using water for injection.

Among other factors, the stability of APIs in the ointment base is of utmost importance in the selection of the appropriate base [60]. Studies have shown that the stability of hydrocortisone in ointment fabricated using polyethylene glycol base is very poor [77, 78]. Semisolid products often turn yellow or brown with age because of oxidative decompositions involving the base (especially natural fats and oils) used in the formulation of the products [55]. Extensive oxidation of natural fatty materials, termed rancidification, is associated with the development of a disagreeable odor. Various phase transitions, crystallizations, and transesterification reactions of suppository bases are believed to be responsible for the hardening of suppository on storage, which may adversely affect the quality of the suppository, such as poor availability of the active ingredients [6]. Moreover, aspirin was found to decompose in polyethylene glycols used as components of suppository bases; the decomposition was attributed to a transesterification reaction which gave rise to salicylic acid and acetylated polyethylene glycol [79]. Degradation of aspirin was also found when cocoa butter was used as the suppository base [80]. Polyethylene glycol esters of indomethacin were found in stored suppositories fabricated with polyethylene glycol bases [81].

Reports have shown that magnesium strearate (a lubricant) accelerated the discoloration of tablets containing amines and lactose [82]. Excipients such as binders (e.g., povidone) and disintegrants such as crosspovidone containing phenolic impurities can impact negatively the photostability of tablets by participating in the free-radical reaction [83]. Enhancement of the oxidation of phenylbutazone by dyes via the production of singlet oxygen that participates in chain reactions has been reported [84]. Excipients with high moisture content can increase the decomposition of drugs by increasing the amount of moisture associated with the drug. The percentage decomposition of aminosalicylic acid increased with increase in the water vapor pressure [85].

## 3 DRUG PACKAGING AND PACKAGING MATERIALS

### 3.1 Introduction

A packaging system, often called a container closure system, comprises packaging components that all together contain and protect the drug product [8]. It is customary to refer to two types of drug product packaging components: primary packaging component (the package components that are directly in contact with the drug

product) and the outer packaging components (usually a series of enclosures) often referred to as secondary packaging components, which include cartons, corrugated shippers, and pallets [86]. Contrary to the general belief that, with the possible exception of labeling, outer packaging components do not require any special consideration when applied to drug products, reports have shown that the outer package can extend the shelf life of drug products considerably: A blister pack exposed naked at 37°C, 90% RH had a shelf life of 21 days (suggesting that a blister might not be suitable for packaging the product). However, when the same blister pack was put in a carton and the carton in turn placed in a display outer overwrap and stored under the same conditions in the company's warehouse, the product was still in specification after 6 years [87].

The traditional definition of pharmaceutical packaging as a system capable of containing the drug product such that it remains safe and efficacious within its shelf life has shifted to a definition that recognizes the importance of packaging in drug product performance. Consequently, a pharmaceutical packaging has been defined as the combination of components necessary to contain, preserve, protect, and deliver a safe, efficacious drug product [87]. In fact, this definition echoes the statement in the guidance for industry Container Closure Systems for Packaging Human Drugs and Biologics [8] that the container closure system must be suitable for the intended use (it should adequately protect the dosage form; it should be compatible with the dosage form, and it should be composed of materials that are considered safe for use with the dosage form and the route of administration). It goes further to indicate that if the packaging system has a performance feature in addition to containing the product, the assembled container closure system should be shown to function properly. Thus drug products such as prefilled syringes, transdermal patches, metered-dose inhalers, and nasal sprays all contain formulations whose successful delivery to the final consumer (the patient) depends on the proper functioning of the packaging system. The requirements for protective, compatibility, safety, and performance characteristics of packaging systems vis-à-vis suitability for intended use vary from one type of dosage form and from one route of administration to another and a table that serves as a guide to the pharmaceutical industry on the suitability considerations for common classes of drug products has been provided elsewhere [8].

### 3.2 Materials of Fabrication of Packaging Components

The science of packaging materials vis-à-vis pharmaceutical and cosmetic products is very broad. The selection of a package often begins with a determination of the product's physical and chemical characteristics, its protective needs, and its marketing requirements. The materials selected must have the following characteristics: They must protect the preparation from various hazards (physical, climatic, chemical, and biological) [87]; they must not be reactive with the product (the current good manufacturing practice (CGMP) regulations [88]); they must not impart to the product tastes or odors; they must be nontoxic; they must be FDA approved (it should be indicated that the FDA approves only the materials used in the container and not the containers as such and an FDA publication lists substances generally recognized as safe (GRAS)]; they must meet applicable tamper resistance

requirements; and they must be adaptable to commonly employed high-speed packaging equipment [89].

***Glass Containers*** Glass is commonly used in pharmaceutical packaging because it possesses superior protective qualities, it is economical, and containers are readily available in a variety of sizes and shapes. It is essentially chemically inert, impermeable, strong, and rigid and has FDA clearance. Glass does not deteriorate with age, and with a proper closure system, it provides an excellent barrier against practically every element except light. Colored glass, especially amber, can give protection against light when it is required. The major disadvantages of glass as a packaging material are its fragility and its weight.

The USP/NF [1] recognizes four major types of glass containers: Type I borosilicate glass containers are used for preparations that are intended for parenteral administration. Type II glass containers which are made of soda-lime glass that is suitably dealkanized are usually employed for packaging acidic and neutral parenteral preparations (type I glass may also be used for this purpose, but it is used for packaging alkaline parenteral preparations). Type III soda–lime glass containers are not used for parenteral preparations except if stability data indicate otherwise. The containers of the fourth category (type NP glass) are intended for packaging nonparenteral articles (i.e., those intended for oral and topical use).

***Plastic Containers*** Plastics in packaging have proved useful for a number of reasons: The ease with which they can be formed, their high quality, and the freedom of design to which they lend themselves; plastic containers are extremely resistant to breakage and thus offer safety to consumers along with reduction of breakage losses at all levels of distribution and use [89]. Plastic-coated glass bottles are used in aerosol containers to protect from flying glass in the event of glass shattering; further, plastic coating around the neck of the container serves to absorb some of the shock from the crimping operation and decreases the danger of breaking around the neck.

Plastic containers for pharmaceutical products are primarily made from the following polymers:

(a) Polyethylene is a good barrier against moisture but a relatively poor one against oxygen and other gases; most solvents do not attack polyethylene, and it is unaffected by strong acids and alkalis; both high- and low-density polyethylene are used and they are long-chain polymers. The density of polyethylene directly determines the four basic physical characteristics of the blow-molded container—stiffness, moisture–vapor transmission, stress cracking, and clarity or translucency—and these characteristics determine the suitability of polyethylene used in containers for packaging drugs. The determinants of suitability of polyethylene have been identified by the USP/NF [1] as follows: oxygen and moisture permeability, modulus of elasticity, melt index, environmental stress crack resistance, and degree of crystallinity after molding).

(b) Polypropylene polymers are long-chain polymers synthesized from propylene or propylene and other olefins under controlled conditions of heat and pressure with the aid of catalysts. Polypropylene does not stresscrack under any conditions. Except for hot aromatic or halogenated solvents, which soften it, this polymer has

good resistance to almost all types of chemicals, including strong acids, alkalis, and most organic materials. Its high melting point makes it suitable for boilable packages and for sterilizable products. Lack of clarity is still a drawback, but improvement is possible with the construction of thinner walls. Polypropylene is an excellent gas and vapor barrier. Its resistance to permeation is equivalent to or slightly better than that of high-density or linear polyethylene, and it is superior to low-density or branched polyethylene [1, 89]. One of the biggest disadvantages of polypropylene is its brittleness at low temperatures.

(c) Polyethylene terephthalate (PET) is a condensation polymer typically formed by the reaction of terephthalic acid or dimethyl terephthalate with ethylene glycol in the presence of a catalyst. PETG resins are high-molecular-weight polymers prepared by the condensation of ethylene glycol with dimethyl terephthalate or terephthalic acid and 15–34 mole % of 1,4-cyclohexanedimethanol. PET and PETG resins and other ingredients used in the fabrication of these bottles conform to requirements in the applicable sections of the *Code of Federal Regulations*, Title 21, regarding use in contact with food and alcoholic beverages: PET and PETG bottles that are interchangeably suitable for packaging liquid oral dosage forms [1]. Other polymers that have been used or proposed for use in packaging of drug products include polycarbonate, polyvinyl chloride, and polystyrene and to a lesser extent polymethyl methacrylate, polyethylene terephthalate, polytrifluoroethylene, and polyamides [89].

Over the years efforts have been made to understand the various parameters to consider in the choice of plastics for packaging of drug products. Drug–plastic considerations have been divided into five separate categories: permeation, leaching, sorption, chemical reaction, and alteration in the physical properties of plastics or products [89].

***Metals*** Metals are used as collapsible tubes and in aerosol containers. The most common metals in use are tin, aluminum, and lead. Tin is the most expensive, while lead is the cheapest. Laminates of tin-coated lead provide the appearance and oxidation resistance of straight tin at lower prices [89]. Tin is the most chemically inert of all collapsible tube metals. It offers a good appearance and compatibility with a wide range of products. Aluminum tubes provide the attractiveness of tin at relatively lower cost. Lead has the lowest cost of all tube metals and is widely used for nonfood products such as adhesives. However, with internal linings, lead tubes are used for such products as fluoride toothpaste. If the product is not compatible with bare metal, the interior can be flushed with wax-type formulations or with resin solutions.

Aluminum is used as the material of construction for some aerosol drug products. Also a three-piece tin-plated steel container finds use in topical pharmaceutical aerosols, and to decrease the compatibility problems, an internal organic coating is often used [90].

***Rubber*** Stoppers, cap liners, and bulbs for dropper assemblies used in the pharmaceutical industry are made form rubber. The rubber stopper is used primarily for multiple-dose vials and disposable syringes. The rubber polymers most commonly used are natural, neoprene, and butyl rubber. As certain performance expectations

are expected from rubber, certain ingredients are commonly found in a rubber closure: for example, rubber vulcanizing agent, accelerator/activator, extended filler, reinforced filler, softener/plasticizer, antioxidant, pigment, and waxes. The complexity of rubber necessitates caution, especially when used in packaging parenteral products: When the rubber stopper comes in contact with parenteral solution, it may absorb active ingredient, antibacterial preservative, or other materials, and one or more ingredients of the rubber may be extracted into the liquid [89].

## 4 EFFECT OF PACKAGING ON DRUG PRODUCT STABILITY

### 4.1 Introduction

The purpose of stability testing is to investigate drug product (with and without container closure system) changes with time under the influence of various hazards such as temperature, humidity, and light and to establish a shelf life for the drug product and to recommend storage conditions. Generally, packaging suitability is based on four attributes: protection, safety, compatibility, and performance (function and/or drug delivery). Certain factors that must be considered in evaluating container closure systems are as follows: materials of construction of the container closure system, surface treatments and/or processing aids, dosage form active ingredients and excipients, sterilization and/or other related processing, and storage conditions. Guidelines are available in official compendia for the evaluation of container closure system. First guidelines are given for the evaluation of the container or the package materials: physicochemical and biological tests to evaluate glass and plastic bottles, metal closures, elastomeric closures, flexible and blister materials, syringe components, and aerosol packaging. Second, guidelines are available for stability studies (accelerated, short-term tests and long-term tests) to characterize the effects of the package component on the product. The general belief is that it is impossible to have a completely inert container closure system; thus the evaluation is often designed to identify, characterize, and monitor these interactions to achieve a safe, unadulterated, stable, and efficacious drug product [86]. At the end of the evaluation of a container closure system in the pharmaceutical industry, a technical report is issued which indicates the approach, results, and the conclusion.

### 4.2 Solid Dosage Forms

A stability study of losartan/hydrochlorothiazide tablets was carried out in three stages [91]. First, a stress test (forced-degradation study) was carried out without immediate packaging by exposing the tablets to severe storage conditions (separately 50 and 50°C, 80% RH). After four weeks it was concluded that the tablets are sensitive to moisture. Second, preliminary testing was carried out to select the packaging system. Two packaging systems with different barrier properties were studied: polyvinyl chloride 250 μm–polyethylene 25 μm–polyvinylidenechloride 60 g/m$^2$ foil (PVC–PE–PVdC–Al blisters capable of protecting the tablets partially from water vapor and gases) and oriented polyamide 25 μm–aluminum foil 45 μm–polyvinyl chloride 60 μm foil (OPA–AL–PVC–Al blisters capable of giving

absolute protection to the tablets). Drug assay, impurities, disintergration, and appearance were monitored. Following six months storage at 40°C, 75% RH, it was concluded that the protection against moisture offered by PVC–PE–PVdC–Al blisters was not sufficient and that losartan/hydrochlorothiazide tablets should be packaged in OPA–AL–PVC–Al blisters. Finally formal stability testing of the tablets packaged in OPA–AL–PVC–Al blisters was carried out. The parameters monitored were as follows: assay, impurities, dissolution, disintegration, hardness, water, appearance, and microbiological quality. The results obtained during 6 months of accelerated and 12 months of long-term stability testing showed that losartan/hydrochlorothiazide tablets packaged in OPA–AL–PVC–Al blisters were chemically, physically, and microbiologically stable and a shelf life of 24 h was proposed.

Studies were carried out to use the drug product and package characterization data (tablet equilibrium moisture content, degradation rate of unpacked product, and moisture barrier properties of the packages) to predict and subsequently confirm the package material that provided adequate stability for a moisture-sensitive compound (PGE-7762928) [92]. The physical and chemical stability [high-performance liquid chromatography (HPLC) assay] of the products were determined at 2, 4, 6, 8, 12, and 24 weeks at ICH conditions. At 6 months at 40°C, 75% RH, the percentage of active ingredient was 84% in polyvinyl chloride blisters, 91% in cyclic olefin blisters, 97% in aclar (polychlorotrifluoroethylene) blisters, 100% in cold-form aluminum blisters, and 99% in a high-density polyethylene bottle with a foil induction seal. The stability results for the packaged product were fairly consistent with the predictions based on the moisture sensitivity of the product and moisture barrier properties of the respective package. Based on the goal of 90% assay at 6 months, cold-form aluminum and Aclar blister packaging and high-density polyethylene bottles with foil induction seals provide acceptable PGE-7762928 tablet stability. Though inert atmosphere packaging is a common practice in the parenteral sector of the pharmaceutical industry, there are relatively few examples of solid dosage forms that are packaged under reduced oxygen levels. Model granule formulations which include a drug known to exhibit oxidative degradation were packaged in stoppered glass vials maintained at different head space oxygen concentrations and head space relative humidities and stored at 40°C. The oxidative degradation was quantified as a function of time and the data showed the dependence of oxidative degradation on head space oxygen concentration, relative humidity, drug loading, and time. It was recommended that the use of oxygen scavengers in bottles as well as inert atmospheric packaging foil–foil blister lines could be options for achieving pharmaceutical packages with low oxygen concentrations [93].

The effect of packaging and storage conditions on the in vitro performance of carbamazepine (CBZ) tablets was studied. Tablets used were Tegretol and the Egyptian generic Tegral, both presented in PVC–aluminum strip seals inside a carton and the German product Finlepsin dispensed in bottles of 50 tablets. In vitro performance was assessed through dissolution testing while chemical stability of CBZ was assessed via HPLC [94]. Tegral tablets were not affected by the tested stress conditions. Tegral tablets packaged in strips at 50 or 60°C and 75% RH showed fast disintegration and dissolution. The effect of 40°C, 97% RH for 6 months was similar to 1 month storage at 40°C, 97% RH; the tablets hardened and dissolved less than fresh Tegral tablets. Removal of Tegral tablets from their original strips

resulted in only 7% dissolution in 60 min. For Finlepsin, the effect of 97% RH at 40°C was more profound than 97% RH at 25°C, but both conditions caused a decrease in dissolution rate, the extent of which was found to be dependent on the tablet position in the bottle. However, all stressed tablets examined in this study showed no change in the chemical stability of CBZ tablets under all stressed conditions.

The model involving mass transfer characteristics of packaging materials and chemical stability has been used to characterize the diffusion of water through polyvinyl chloride blister packaging. It was found that the effect of moisture sorption on tablet crushing strength was dependent on the formulation and that it is necessary to consider the characteristics of the formulation when selecting a package [95]. Information on the tablet moisture sensitivity of compressed tablets together with package moisture permeability has been used to develop a physical model which predicted the changes in the crushing strength of tablet under a variety of storage and packaging conditions. The predictions are described as being useful in establishing long-term stability testing protocols. They are also useful in predicting what tablets stored in blister packages might reasonably be expected to experience in the marketplace where short- and long-term oscillating conditions are likely to occur. The theoretical predictions and experimental results emphasize the importance of matching the dosage formulation characteristics to the package material and testing conditions so that a more rational selection of packaging material and testing protocols can be made [96].

It is known that moisture transfer through packaging materials can affect hard gelatin capsules in two ways: changes in the properties of the shell and changes in the properties of the materials loaded in the capsule (contents of the capsules). Studies on capsules have recently examined the changes in both dissolution rates and potency of two brands of amoxicillin capsules packaged in PVC blister package (PVC blister having a thickness of 0.27 mm and aluminum base film of thickness 27 μm) and laminated–type package (the laminated package consisting of three layers: a nitrocellulose lacquer outer layer, a soft aluminum film intermediate layer, and a polyethylene inner layer). The two brands of amoxicillin stored outside the package at 76, 80, and 92% RH and inside the package at 92% RH showed no significant changes in dissolution profiles. Storage at 80 and 92% RH of the capsules of the two brands outside their packages resulted in a significant loss of amoxicillin potency. The laminated-type package afforded better protection compared with the PVC–aluminum blister package. Following 20 weeks storage at 92% RH at room temperature, only 6.4% loss in amoxicillin potency occurred for capsules packaged in the laminated package compared with 51.8% loss in the blister package [97].

The effects of two storage conditions (50°C, 50% RH and 40°C, 90% RH) on the properties (hardness, disintegration, tablet weight, dimensions, dissolution rate, and content of API) of two brands (A and B) of film-coated erythromycin stearate tablets and one brand (C) of enteric-coated tablets of erythromycin base were examined. The tablets were stored in paper bags, plastic dispensing bottles, and glass bottles. Aside from an increase in hardness by storage at 40°C, 90% RH for brand C tablets and a decrease in disintegration time for tablets of brand A under the same conditions, little or no changes were seen in most of the physical properties. However, dissolution rates significantly decreased for all the tablets by storage under the two conditions. Glass containers offered better protection for the tablets,

and consequently, the tablets retained higher dissolution rates as compared to tablets stored in plastic or paper containers [98].

The sorption of nitroglycerin by thermoplastic polymers and the stability of molded nitroglycerin tablets in strip packaging were studied. The polymers investigated varied greatly in their affinity for nitroglycerin, the order of decreasing affinity being vinyls > low-density polyethylene > ionomers > high-density polyethylene. With the proper choice of packaging, molded nitroglycerin tablets stabilized with povidone maintained acceptable potency for up to two years at 26°C when strip packaged in unit-dosage form. Chemical decomposition by hydrolysis of nitroglycerin was also investigated. Povidone accelerated the decomposition of nitroglycerin; at high temperature, decomposition was a significant factor in tablet stability for tablets containing povidone [99]. A liquid dispersion system of etodolac (20%) and Gelucire 44/14: D-α-tocopheryl polyethylene glycol 1000 succinate (TPGS) blend (80%) in different ratios was prepared. The capsule formulation was subjected to stability studies at different temperature humidity conditions based on ICH guidelines. Physical and chemical properties of the dispersion did not change during a period of storage at room temperature and at 4°C, 0% RH. However, the relative humidity and storage time exerted an effect on the dissolution behavior of etodolac. It was concluded that changes in dissolution behavior after storage under conditions of high humidity and temperature might be related to the formation of etodolac microcrystal and water absorption by the carrier during storage, thereby necessitating packaging in moisture-resistant packaging [100].

### 4.3 Nonsterile Liquid Dosage Forms

The effect of packaging materials on the stability of ultraviolet (UV) A, UVB, and infrared sunscreen emulsions after storage in different types of packaging materials (glass and plastic flasks; plastic and metallic tubes) has been reported [101]. The samples (emulsions containing benzophenone-3, octyl methoxycinnnamate, and Phycocorail) were stored at 10, 25, 35, and 45°C and representative samples were analyzed after 2, 7, 30, 60, and 90 days. Data showed that sample emulsions stored at different temperatures had similar rheological behaviors within 3 months and there were no significant changes in the physical and chemical stability of emulsions stored in different packaging materials. Glass and plastic packaging materials were found adequate for storing the products.

Reports have shown that chloroquine binds to glass but not to certain plastics. This information is very important in laboratory studies where significant reductions in chloroquine concentration may occur when the drug is prepared and stored in glass containers [102]. Further, buffered chloroquine diphosphate solutions of varying pH and concentration were stored in soda or borosilicate glass. Storage in soda glass showed a decrease in original drug concentration of up to 60 and 97% in test tubes and glass wools, respectively, Borosilicate glass did not show any binding. The highest binding recorded was at physiological pH. Thus borosilicate glass should be used when storing, assaying, or carrying out sensitivity tests for malaria in order to avoid loss of chloroquine from solution [103].

Oxygen is often used as a challenge to find out whether a particular drug is likely is to be affected by oxidative breakdown. The procedure involves storing the drug product in solution in ampoules purged with oxygen and then comparing their rate

of breakdown with similar solutions stored under nitrogen. Drug formulations shown to be susceptible to oxidation can be stabilized by replacing the oxygen in the storage containers with nitrogen or carbon dioxide; antioxidant can be added or the choice of container should aim at avoiding the presence of metal ions [9]. In the case of drug products (solution, semisolid, and solid) that are photolabile, the usual practice is to store them in containers which exclude UV light. It is believed that exposure to light in this wavelength range is the most frequent cause of photodecomposition. Amber glass is very effective in eliminating or reducing photodegradation because it is capable of excluding light of less than 470 nm [9]. Photolabile drugs can also be stored in the dark.

### 4.4 Sterile Liquid Dosage Forms

The use of prepackaged syringes in the hospital setting is increasing due to the advantages offered by their use and the current interest in unit-dose medication. It is known that all the various dosages and combinations prescribed in the hospital are not commercially available; consequently many hospital pharmacies are developing or operating a centralized parenteral admixture program. Packaging of the drug products requires consideration of the sterility and particulate matter content of single-syringe injections as well as compatibility between drugs and with the syringe components. Though plastic syringe packaging has a number of advantages, contact between plastic containers and drugs can present a number of problems, such as leaching, sorption, permeation, chemical reactivity of the plastic polymer, and alteration in physical properties of the plastic. Studies have been carried out on a commonly prescribed drug combination (hydroxyzine hydrochloride, meperidine hydrochloride, and atropine sulfate) used as preanesthetic medication to compare possible differences in stability between mixtures stored in glass and plastic containers. The combination drugs were stored in both glass and plastic syringes at 25 and 3°C for a 10-day period. Analyses were performed at intervals throughout the time period utilizing visual examination, pH determination, UV absorption spectra, and gas chromatography. No significant degradation of the syringe contents or appearance of additional constituents was detected in any of the admixture preparations. Thus the storage of such preparations in glass versus plastic syringes yielded no significant differences in product stability [104].

Report has shown the influence of the type of container on ceftazidime stability in intravenous solutions. Polypropylene (PP) and PVC (100 mL each) bags and 100-mL glass bottles were filled with 5% dextrose or 0.9% sodium chloride solutions containing ceftazidine (40 mg/mL) and stored at 20 and 35°C. One-milliliter samples were taken from each container at 0 and 20 h and assayed. Pyridine (the main degradation product) levels increased during storage and were higher in PVC and PP bags than in glass bottles in both diluents. Solutions stored in PP bags showed better stability than in PVC bags. Glass bottles seem to be the better container for storing ceftazidime solutions [105]. Studies have been carried out to evaluate the effect of diluent type, storage conditions, and nature of package on the stability of reconstituted Parecoxib sodium for injection (PSI) [106]. Results showed that the PSI reconstituted with normal saline, bacteriostatic normal saline, 5% dextrose injection, and 5% dextrose injection and half normal saline met visual acceptance criteria and showed almost no degradation under storage conditions, and no significant differ-

ences were seen between storage in glass vials or PP/glass syringes. However, PSI reconstituted with lactated Ringer's injection and lactated Ringer's and 5% dextrose injection showed visual precipitation in many vials as confirmed by HPLC assay values at all points.

The stability of an ophthalmic solution formulation of unoprostone isopropyl (UI) (a prostaglandin-like compound marketed as Rescula for treatment of elevated intraocular pressure in patients with primary open-angle glaucoma or ocular hypertension) was investigated in two types of packaging materials: PP and low-density polyethylene (LDPE). The concentrations of UI and its degradation products were monitored as a function of time. It was found that the rate of disappearance of drug was faster for the formulation stored in LDPE bottles than that stored in PP bottles. Further studies indicated that the inferior stability observed with LDPE packaging was primarily due to the sorption of UI to the packaging material and to a lesser degree chemical degradation. The sorption was found to be temperature dependent: Lowering the temperature reduced the sorption, thus improving the shelf life of the product [107].

The USP has proposed large-globule-size limits to ensure the physical stability of lipid injectable emulsions, expressed as the percent far greater than 5 μm, or $PFAT_5$, not exceeding 0.05%. Recent studies showed that injectable emulsions packaged in newly introduced plastic containers exceed the proposed USP $PFAT_5$ limits and subsequently become significantly less stable during simulated syringe-based infusion. Although modest growth in large-diameter fat globules was observed for the glass-based lipids, they remained within proposed USP globule size limits throughout the study. Thus glass-based lipids seem to be a more stable dosage form and potentially a safer way to deliver via syringe infusion to critically ill neonates [108].

Considerable interest has been shown, in recent years, in the clinical use of intravenous nitroglycerin for the treatment of myocardial infarction and in open heart surgery. In its intravenous use loss of the drug to PVC bags has been identified as a problem and has been characterized as a diffusionally controlled absorption process having a half-life for fractional absorption of 3.2 h at 30°C [109]. Moreover, in recent studies, administration set tubing has been implicated as the source for the observed loss of nitroglycerin from solution [110–112]. Quantitative studies have been reported on the loss of nitroglycerin to plastic intravenous tubing. It was found that in the short time periods the loss of nitroglycerin from normal saline solutions to PVC tubing could be treated as an adsorption process. The rate of adsorption was rapid and could be quantified as an apparent first-order process. The half-life of the loss was 2.6 min [113]. Recently a model describing the loss of nitroglycerin from the solution in the plastic bags as a rapid adsorption onto the plastic surface followed by partitioning into the plastic was proposed [114]. The mechanism of loss of nitroglycerin stored in plastic and glass containers was studied from an equilibrium and kinetic approach. Data showed that nitroglycerin was removed from aqueous solution by the plastic container material through an absorption process. Nitroglycerin loss from aqueous solution did not occur by hydrolysis since solutions stored in glass containers at pH 5.7 and 35°C retained potency for at least 48 h. Although experimental data supported the thesis that the uptake of nitroglycerin was migration of the drug into the plastic matrix, the finding did not exclude the possible adsorption of the drug on the surface: The rate of adsorption may be much

faster than the rate of absorption with the result that any adsorption may be obscured [115].

An investigation has been carried out to define the kinetics and mechanism of the interaction between various drugs and plastic infusion bags and to develop criteria that may be used in the prediction of such interactions. The sorption behavior of warfarin sodium, various benzodiazepines, and other drugs with PVC infusion bags was examined. The sorption of some compounds by PP infusion bags was also investigated. The sorption kinetics fro warfarin and diazepam could be accounted for by a diffusional model in which the loss of drug is determined primarily by the diffusivity of the compound in the plastic matrix. The rate and extent of sorption of warfarin showed a dependence on pH which could be interpreted in terms of ionization of the drug: Only the un-ionized form was sorbed. A rank-order relationship was found between the initial rate of uptake by the plastic bag and the hexane–water partition coefficients of the compounds. Thus hexane–water partition coefficient may be useful for the prediction of interactions between a drug substance and PVC infusion bags. Sorption of the compounds by infusion bags made from PP was insignificant, except for the highly lipophilic medazepam. Thus PP bags may be safer to use than PVC bags [116].

Drugs to be administered via the parenteral route which are unstable in water are often formulated as solids for reconstitution. Such drugs may be presented in vials or in two-compartment cartridges with powder in one side and diluent in the other. Rubber closures are used as primary packaging material because of their unique properties, such as elasticity for piercing and self-sealing which is maintained under freeze-drying conditions. Further, vial seals made from butyl or halogenated butyl rubber show low moisture vapor and gas transmission rate. During storage rubber closures are in intimate contact with vial contents and are a potential source of product contamination: shedding of particles, migration of benzothiazoles, leaching of metal ions, antioxidants and oligomers, silicone oil, sulfur, and paraffin wax. Reconstituted solutions in aqueous vehicles show haze formation which arises from precipitation in oversaturated solutions, polymerization of decomposed drugs, and release of nonpolar volatiles (saturated hydrocarbons, unchlorinated or chlorinated olefins, alkylbenzenes, and low-molecular-weight polydimethylsiloxanes) from rubber closures [117]. Studies have shown that high product surface area is a driving force for the adsorption of volatiles released from rubber stoppers (products with high surface areas show high levels of turbidity after reconstitution). Attempts to impede haze formation include the use of solubilizers and the use of closure materials such as bromobutyl closures instead of butyl and chlorobutyl rubber or Teflon-lined rubber stoppers. Reduction of migration of volatiles can be achieved by exposure of rubber stoppers to heat or vacuum within the sophisticated washing cycles [118].

Studies have been carried out on intravenous containers (plastic and glass) and plastic administration sets (with and without inline filters) on diazepam availability. No visual incompatibilities were observed and the pH remained constant. Solutions stored in glass bottles and infused through plastic sets retained greater than 90% of initial potency after 4 h. Solutions stored in plastic burette chambers and administered through plastic sets lost greater than 38% of potency after 2 h. The loss occurred principally in the plastic chamber and increased with increasing drug concentration and time. Drug availability was not affected by the type of intravenous

fluid, pH, flow rate, or filtration. It was recommended that if diazepam were to be infused, it should be diluted to at least 1:10 dilution in glass intravenous bottles and administered through plastic administration sets not having plastic burette chambers. Inline filters (0.45 μm) may be used without loss of potency [119]. Similar studies were carried out by Morris [120]. It was found that diazepam injection, in dilutions of 10 mg/dL or greater, was visually compatible and chemically stable in dextrose 5% in water, Ringer's injection, lactated Ringer's injection, and 0.9% sodium chloride in glass intravenous bottles when administered within 24 h. However, the same type of study carried out in plastic intravenous bags instead of glass bottles showed that the solutions had a potency loss of greater than 24% within 30 min after dilution. The percentage of diazepam lost from solution increased with increase in concentration and time [121].

Injectable lyophilized drug products are often presented in vials. The rubber stopper forms a critical barrier: Protection of the product against moisture or oxygen is strongly dependent upon the quality and functioning of this barrier. In fact, it has been reported that rubber not only is a barrier against moisture but also can be a source of water [122]. Studies have been carried out to investigate the speed of water uptake, the saturation, and finally the permeability of different types of rubber closures with a view to understanding the fundamental mechanisms which determine the barrier quality and obtaining the easy method to discriminate between different rubbers. Thirteen-millimeter closures of the same dimensions were used: the Helvoet FM 157 (grey) and FM 257 (grey); the Pharma Gummi PH 701/45 (red), PH 21/50 (red), and PH 4104/40 (grey); and the West Company W888 (grey). All rubbers were of the bromobutyl rubber type, with the exception of W888, which consists of a halobutyl/ polyisoprene blend. It was found that permeation through the closures is highly unpredictable owing to the concentration dependency of the diffusion coefficient $D$, which increases strongly with increasing RH. It was indicated that new types of rubbers could be evaluated with respect to their barrier properties by means of the absorption profiles under stress conditions such as 40°C 95% RH. Thus one does not require to make a selection on the basis of laborious permeation experiments [123].

### 4.5 Container Extractables and Leachables and Drug Product Stability

Containers and drug delivery devices for pharmaceutical and biological products may release low-molecular-weight compounds into the product. These compounds are called extractables and leachables. The toxicological concern is that an extractable or leachable compound from the container may affect the efficacy and biological safety of the drug-product. The drug-contacting packaging materials (synthetic polymeric formulations) may include antioxidants, colorants, slip agents, and plasticizers.

The subject of extractables and leachables in drug products is an area of active discussion in the pharmaceutical industry. Further the regulatory agencies have issued guidances on this subject in recent years. The FDA guidance on container closure systems defines extractables and leachables as follows [124]: *Extractables* are compounds that can be extracted from elastomeric or plastic components of the container closure system when in the presence of a solvent. *Leachables* are compounds that leach into the formulation from elastomeric or plastic components of

the drug product container closure system. Thus the term extractable refers to the removal of a compound from a container closure system under extractive conditions by exposure to solvents under nonstandard or atypical product handling conditions using a variety of solvents and/or stress conditions that are not used in the manufacturing process of the drug product (strong organic solvents or acid/alkaline solutions as well as elevated conditions of temperature and pressure). Leachable describes the removal of a compound from a container closure system under nonstressful solubilizing conditions (exposure of drug substance to surfaces and conditions defined specifically by the manufacturing process) [125]. The difference between the two terms is one of process and not related to a fundamental chemical property of the compound. An extractable or leachable is generated by partitioning of a given compound between two phases: the solid (container closure system) and liquid (drug product for leachables or a solvent for extractables).

It is necessary to test the extractables and/or leachables from each container or device followed by toxicological evaluation of the hazard potential. Toxicological risk assessment includes evaluating the toxicity of the packaging components and the specific extractable or leachable substance. Toxicological endpoints include single, repeated, and chronic dose toxicity, genotoxicity, carcinogenicity, immune sensitization, irritation, and blood compatibility [126]. The concern that the regulatory authorities (FDA) have regarding drug product leachables is directly related to the particular dosage form and the route of administration [8]. Of the highest concern are the inhalation drug products [orally inhaled and nasal drug products (OINDPs)] [124, 127, 128], which include metered-dose inhalers (MDIs), dry powder inhalers (DPIs), inhalation solutions and sprays, and nasal sprays. The FDA considers leachables to be a safety concern for OINDPs. Leachables are also of concern for injections and injectable suspensions, sterile powders and powders for injection, ophthalmic solutions and suspensions, and transdermal ointments and patches. Other dosage forms such as liquid and solid oral are of less concern [8].

There are many challenges (mainly scientific) for pharmaceutical development programs and teams regarding leachables and extractables in terms of analysis and questions such as: At what levels are drug product leachables of safety concern? What safety qualification processes can be applied to leachables? How do you develop control strategies (including specifications and acceptance criteria) for leachables? Can you control drug product leachables by controlling potential leachables (i.e., extratables)? [128]. A report has shown that The Product Quality Research Institute (PQRI) Leachables and Extractables Working Group proposed a complete pharmaceutical development process for OINDPs related to leachables and extractables [129] based on its own comprehensive scientific research and on some earlier proposals from the International Pharmaceutical Aerosol Consortium for Regulation and Science [130]. The PQRI recommendations also discuss selection criteria for OINDP container closure components which are designed to exclude potential leachables of known safety concern and minimize both the numbers and levels of other potential leachables. It is believed that these recommendations move beyond quality by testing paradigm and suggest a process by which quality is designed into OINDP container closure system components (quality-by-design paradigm) [128]. Extractable and leachable testing studies are recommended even if the containers or closures meet compendial suitability tests.

Aside from the toxicological risks posed by leachables and extractables, another source of drug product quality problems associated with leachables is stability because leachables could react chemically and may interact with drug product. Some of the more common chemically reactive leachables include transistion metals, radical initiators or propagators, organoperoxides, and reactive nucleophiles or electrophiles such as amines and aldehydes [131]. The principal sources of metal leachables include glass packages and product-contacting equipment in the manufacturing process, especially unpassivated stainless steel equipment. Metal leachables can often cause degradation via oxidation in which the metal catalyzes the formation of a short-lived peroxy radical, which further reacts with the drug substance [132]. Transition metals such as $Fe^{n+}$ and $Mn^{n+}$ are usually involved and the oxidation occurs via a Fenton catalytic cycle [133]. Degradation was observed for a developmental intravenous dosage form where metal ions leaching out of glass vials and stainless steel catalyzed single-electron oxidation. Interestingly, the antioxidants in the formulation (thioglycerol, ascorbic acid, or sodium bisulfite) were involved in the Fenton reactions leading to oxidation [134]. For a photosensitive product, the use of amber glass vials resulted in other stability problems: The higher levels of leachables from amber vials enhanced the metal-catalyzed drug oxidation [135]. Metal leachables in liquid formulations may result in the formation of insoluble complexes with the pharmaceutical active substance or other formulation ingredients. It has been reported that accelerated stability testing at higher temperatures might not be indicative of such problems since these precipitates tend to form and grow at lower temperatures in a nonlinear fashion [136]. Aluminum ions coming from USP type I glass and some plastic packages such as LDPE and rubber closures are the most prevalent source for the formation of particulates. The $Al^{3+}$ leaches out and accumulates in solution in levels from 45 ppb up to 6 ppm, depending on factors such as the presence of buffers, the solution pH, and autoclaving [137].

Organoperoxide radicals can be present in plastic packaging materials as a result of residual free radicals from dissociation of the peroxides formed during polymerization or melt processing of the plastic or from the long-lived radicals formed in the polymer as a result of gamma-induced radiolysis of the polymer chain. Studies on the migration of radiolysis products from plastics sterilized by gamma radiation or electron beam showed that smaller molecular weight fragments are formed by radiolysis of the plastic additives and residual oligomers [138, 139]. Very reactive species such as aldehydes have been reported to react directly with drug products possessing a nucleophilic site such as a primary or a secondary amine. Formaldehyde can be formed in minute amounts as a by-product of longer chain alcohol oxidation and from other sources such as acrylates. A sterile formulation containing an anti-stroke developmental product packaged in glass vials with rubber stoppers exhibited a formadehyde-adduct degradate in levels above 2% after 13 weeks at 30°C. The formaldehyde source was found to be in the rubber stopper (formed from a reinforcing agent used in the stopper) [140]. Strategies to mitigate the stability risk of leachables include a careful examination of mode of sterilization and packaging selection. Use of silica-coated glass vials, use of nitrogen head space in the package, use of plastics, and inclusion of a chelating agents will reduce the adverse effects of leachables on stability.

Protein aggregation has recently attracted a lot of attention from the FDA due to a number of incidents in which stability problems were caused by container

leachables/extractables or changes in excipients, as suggested in the case of Eprex (epoetin alfa, Johnson & Johnson, New Brunswisk, NJ). Since aggregation can induce immunogenicity reactions with potentially severe consequences, the FDA is requ iring manufactures to pay more attention [141]. Evaluation of packaging materials for potential extractables and leachables is critical to guarantee the integrity of the drug product and assure compliance with the *Code of Federal Regulations* (CFR), Title 21, Part 211.65, which states that equipment should be constructed so that contact surfaces that contact components, in-process materials, or drug products should not be reactive, additive, or absorptive so as to alter the safety, quality, or purity of the drug product beyond official or other established requirements [142]. The U.S. Food, Drug and Cosmetic (FD&C) Act also states that a drug device should be deemed to be adulterated if its container is composed in whole or in part of any poisonous or deleterious substances which may render the components injurious to health [143]. A review of regulatory and scientific considerations on testing for extractables and leachables has been provided [144]. To facilitate the evaluation of extractables (to detect, identify, and quantify organic extractables) mass spectometry was used recently. Using the Agilent G1888 Network Headspace/680N GC/5975 inert Mass Selective Detector (MSD) system, several primary packaging materials for pharmaceuticals were evaluated. Data on two of the materials (HDPE bottle and a soft elastomer liner from the inside of the screw cap) were presented. A multiple head space extraction technique was used and the highest attainable amount of extractables that could ever be concentrated in the drug product was calculated. It was found that the analytical results obtained from the inert 5975 head space gas chromatogrophy/mass spectrometry (GC/MS) provided excellent sensitivity [145].

The effect of ammonium sulfate treatment on cerium oxide glass vials was assessed following exposure to ionizing radiation. The bulk chemical composition of irradiated cerium oxide glass remains unchanged despite a temporary browning effect. Stability against alkali leachables of the internal silica matrix was enhanced with ammonium sulfate treatment. With the exception of alumina ($Al_2O_3$ and $Na_2O$), irradiation sterilization has a limited effect on altering the surface chemistry of ammonium sulfate–treated cerium oxide glass [146].

### 4.6 Biotechnological Products

Despite the efforts at delivering products of biotechnology (peptides and proteins) through novel drug delivery systems, injection still remains the main mode of delivery. The unit dose for many injectable biotechnological products is the single-dose vial and occasionally prefilled syringes. The product is often provided either as a solution or as a lypholized cake to be reconstituted and injected using syringes. Packaging represents the first line of defense for all formulated drug products, protecting the product from the outside world and vice versa. At the same time, the package must be fully compatible with the product [147]. Thus the requirements for product purity, activity, and shelf life dictate a high standard for injectable drug packaging.

Packaging conditions have effects on the stability of protein products. Peptide and protein drugs are high-molecular-weight compounds with unique physicochemical properties. They are extremely sensitive to their microenvironment: heat, light, pH, chemical contaminants, and so on. Trace amounts of metals, plasticizers, and

other materials from packaging may deactivate or denature therapeutic peptide and proteins. Further, peptides and proteins have a tendency to adsorb on to the surface of container closure systems, thereby removing virtually all active material from the drug formulation. Even when the drug desorbs back into the solution, the interaction could result in loss of potency. Lyophilized biopharmaceutical products can be affected by moisture if the seal is not adequate to prevent ingress of moisture into the container. Thus packaging can make and break final formulation of lyophilized product. Vials that are not designed specifically for lyophilization (with convex rather than flat bottoms) can make the lyophilization process less efficient. Rubber closures can also hinder freeze drying if they do not permit adequate venting during sublimation. Stopper rubbers adsorb and desorb moisture at different rates and under storage conditions stoppers that were not properly dehydrated can release water into lyophilized product.

Another source of instability of biopharmaceutical products traceable to packaging is silicone oil, which is commonly used to lubricate elastomeric stoppers during fill/finish to facilitate insertion of the stopper into the vial. Silicone oil is known to inactivate protein through nucleation of protein around oil droplets. The problem is being circumvented by using a fluoroelastomer coating on stoppers to provide the needed lubricity in addition to chemical innertness, barrier protection, and safety [148]. The fluoroelastomer films reduce adsorption of drug on to the stopper, provide lubricity for proper vial sealing, and also reduce the possibility of extractables migrating from the rubber stopper into the product. The prefilled syringes are becoming very popular in the injectable markets. Some of the challenges to overcome are the compatibility and stability issues that arise when dealing with biotechnology formulations. Biotechnological products can react with the oily form of silicone, which is used as a lubricant to coat the sliding components of the syringe. It is believed that the propensity for silicone to react with the formulation is a function of the concentration of the silicone in the syringe and its chemical activity. The chemical activity is determined by the number of terminal hydroxyl groups, which is greater the shorter the silicone polymer chain length. It has been reported that baking on the silicone (which involves heating the silicone-coated syringe to a specific temperature for an appropriate time) results in longer chains that are more closely adhered to the surfaces they coat. Thus the concentration of silicone in the syringe and its chemical reactivity are both reduced and the product's stability is increased [149]. Another benefit of baked-on silicone is that it reduces the frequency of the "break-loose" effect, which occurs during storage when the rubber closure inside the syringe barrel expands outward so that eventually it displaces the low-friction silicone coating and comes into direct contact with the inner glass surface. Another challenge is the prevention of undesirable pH change that sometimes occurs in liquids stored in prefilled syringes. The shift in pH occurs because the USP type I glass used in prefilled syringe manufacture is a borosilicate which must be subjected to various temperature changes during the glass tube production process. During storage, sodium ions are released into the product and increase the concentration of hydroxide ion. This problem is solved by spraying ammonium sulfate into the glass barrel before the tempering process in the formation of syringes begins.

Factor VIII (FVIII) is an essential coagulation factor in the blood which serves as a cofactor in the complex blood-clotting cascade. A deficiency in FVIII is the

cause of hemophilia (type A), a hereditary life-treatning bleeding disorder [150]. The effect of the container on FVIII stability was shown to be temperature dependent. At 4–8°C, 2 of the 15 constituted FVIII concentrates showed better stability in plastic containers and 2 concentrates showed better stability in glass containers; at 20–23°C most concentrates showed better stability in plastic contaners; and at 37°C all concentrates showed equal or better stability in plastic containers [151]. It was concluded that stability studies comparing different types of containers should be conducted at the product storage temperature. Due to limited diffusion of molecules in a solid state, containers would not be expected to play a critical role in storing lyophilized FVIII products unless the air permeation through the container and/or container stoppers is significantly different. This is because air has been shown to accelerate the inactivation of recombinant FVIII SQ not only in solution but also in a lyophilized state [152].

It is known that interaction of proteins with the surfaces of their containers is a potentially significant problem in biotechnology. The amphiphatic nature of protein molecules results in their adsorption to a wide variety of surfaces and can result in both their loss and destabilization. Reports have shown that when adsorption of a protein/peptide drug to container occurs, the drug molecule exchanges solution interactions for surface interactions where the free energy of exchange is negative [153]. Similarly, studies have shown that some surfactants are able to reduce/eliminate protein/peptide drug adsorption to glass and PP. In the presence of surfactants, where the surfactant–surface interaction is greater than the surface–protein/peptide interaction, drug adsorption is reduced or eliminated. For protein/peptide adsorption onto glass, where an electrostatic interaction (interaction between the positively charged peptide/protein and the negatively charged glass surface) predominates, only the most hydrophobic surfactants (polysorbate 20 and benzakonium chloride) were significantly effective to reduce adsorption to surfaces as demonstrated with salmon calcitonin and bovine serum albumin (BSA). The nonionic surfactant polysorbate 20 proved to be the most effective in eliminating adsorption to both plastic and glass surfaces. The predominat mechanism of protein/peptide adsorptrion to PP (plastic surface) was by a hydrophobic/dehydration mechanism [154]. The effect of Poloxamer 407 (Pluronic F-127), a nonionic surfactant, on the adsorption of granulocyte colony-stimulating factor (G-CSF) to PVC was assessed. It was found that Poloxamer 407 at a concentration of 0.05% w/w may show promise as a solvent additive with which to minimize G-CSF adsorption to PVC [155].

The amount of surface adsorption of a number of proteins ranging in molecular mass from 6.5 to 670 kDa and isoelectric point (pI) from 4.3 to 10.5 to several commonly used container surfaces (glass vials: either untreated, siliconized, sulfur treated or Purcoat treated; plastic vials: polyester + 0.3%, polyester 5 × 0, PP, and nylon). A 5-mL volume of protein solution was added to each vial, yielding a surface-to-volume ratio of 2.4 $cm^2$/mL. No correlation was found between the amount adsorbed and the molecular mass or isoelectric point, although glass surfaces appeared to bind more protein under the experimental conditions examined [156].

Patients receiving total parenteral nutrient solutions frequently require exogenous insulin to fully use the administered glucose. Consequently a study was conducted to determine the percentage of insulin adsorbed to the glass and PVC when added to a nutrient solution infusion system. The following parameters on insulin

availability from parenteral nutrient solutions were investigated: sample time for infusion, insulin concentration, amino acid or polypeptide source, electrolytes and vitamins, inline filters, glass and PVC infusion containers, and human albumin. Results showed that basic solutions of amino acids and protein hydrolysates in dextrose with 30 units of insulin failed to deliver approximately 44–47% of the added insulin. Varying the concentration of insulin had a small but statistically significant effect on the degree of insulin loss. The use of inline filters and PVC bags caused an even greater loss of insulin. The addition of albumin or electrolytes and vitamins decreased the insulin loss [157].

### 4.7 A Look into the Future of the Effects of Packaging on Stability of Drug Products

***New Drug Delivery Systems*** Reports have shown that the advent of new drug delivery systems for various routes of drug administration (e.g., oral, nasal, pulmonary, transdermal, needle free) and the development of new biotechnological drugs have resulted in the need for enhanced protection against such factors as moisture, light, oxygen, and mechanical forces as well as making packaging play a more integral role in drug delivery [158]. Novel drug delivery systems have necessitated the emergence of specialized packaging needs because not all of them can be packaged in bottles or standard blisters. More often than not, the packaging must be unit dose and must also become an integral part of the drug delivery technology, as drug product packaging and packaging design contribute significantly to the stability, shelf life, and performance of the drug and the drug delivery system. In fact, the old equation showing the relationship between a new chemical entity (NCE), drug delivery (DD), and drug product (DP) (i.e., NCE + DD = DP) is now being replaced by NCE + DD + packaging = DP. Table 1 gives the considerations needed when designing packaging for new drug delivery systems [158].

One of the stability problems associated with transdermal drug delivery devices is the degradation of the contents of the devices: drugs, permeation enhancers, matrix materials, and other components incorporated in the devices. Degradation undesirably breaks down the components as well as causes discoloration and formation of odors within the pouched system. Devices that are susceptible to degradation cannot be stored for a reasonable amount of time, thus causing practical problems in their distribution. A solution to this problem involves the incorporation of an antioxidant such as BHT into the drug formulation of the transdermal drug delivery device [159]. Also a desiccant has been incorporated within the sealed pouch of the transdermal drug delivery device. For example, the Climara transdermal estradiol system is packaged and sold within a sealed pouch containing a water scavenger to protect against hydrolysis of estradiol [160].

The effects of extreme temperatures on drug delivery of two albuterol sulfate hydrofluoroalkane (HFA) MDIs were evaluated. Three Proventil HFA and three Ventolin HFA MDIs were stored at room temperature and served as controls, while three of each product were placed in the trunk of a vehicle in Tuscon, Arizona. The temperature of the vehicle was monitored for six months. Product performance for each of the MDIs was evaluated at room temperature. An additional study was performed to investigate the performance of the two products when actuated at 4, 22, 47, and 60°C. Within one week of being placed in the test vehicle, all MDI

**TABLE 1 Packaging Needs and Considerations for New Drug Delivery Technologies**

| Delivery Route | Company Names and Systems | Characteristics | Nature | Packaging Needs and Description |
|---|---|---|---|---|
| Oral | CIMA, OralSolv<br>RP Scherer, Zydis<br>Biovail, Flashdose<br>EthyPharm, Flashtab<br>Yamanouchi, wowTab<br>Elan, FASTMelt<br>Eurand, Ziplets | Fast-dissolving tablets or orally disintegrating dosage Forms | Hygroscopic (polysaccharide or protein-based, taste masked, fragile | Protection from mechanical forces: stiff packaging<br>Peelable opening<br>Ultrahigh moisture barrier, e.g., Aclar or foil to maintain shelf life stability for 2–3 years |
| Pulmonary | Inhale/Inhance™<br>Aradigm/AERx®<br>Alkermes/AIR®<br>BatellePharma Therapeutics/ EHD pulmonary delivery | Inhalable delivery deep into lungs, either powder or liquid | Can be fine powder with high surface area or aqueous/ nonaqueous formulation | Medium–high barrier protection<br>Mechanical needs depend on delivery device design<br>Sterile barrier<br>Chemical inertness, especially for high liquid doses<br>Clear barrier offers QA check for both producer and end user |
| Transdermal | 3M/Latitude™<br>Noven/DOT Matrix™<br>Alza/D-Trans® | Sustained delivery/ absorption through skin | Depending on drug/formulation | Protection of active from environmental aggrevates<br>Clear barrier offers aesthetics<br>Pouching as primary/ secondary package<br>Barrier property on carrier web critical<br>Chemical inertness |
| Transmucosal delivery | Cephalon, Actiq<br>Atrix Labs, BEMA<br>CIMA Labs, Dravescent | Sustained delivery via mucosal layer | Moisture senstivity | Ultrahigh barrier moisture protection<br>Clear material offers QA check<br>Aesthetics |

canisters developed physical distortions. The Proventil HFA MDIs bulged at the base of the canister, while the Ventolin HFA MDIs bulged around the valve. These distortions were thought to be the result of an increase in vapor pressure brought about by the initial high temperatures. After exposure to extreme environmental temperatures, an increase in propellant-leak rate was observed with Proventil HFA and Ventolin HFA MDIs, but little to no change was found in particle size, dose per

actuation, respirable mass, and nonrespirable mass. Despite the two formulations' tolerance to extreme temperatures, drug delivery was affected when the MDIs were tested at specific temperatures outside of the recommended usage [161]. One of the alternatives to the pressurized metered-dose inhaler (pMDI) is the breath-actuated DPI. Stable and predictable therapeutic responses require a consistent dose delivery from an inhaler throughout its life and consistency of doses from one inhaler to another. Recognizing this, specifications for inhaler dose uniformity have been defined by regulatory agencies, including the European Pharmacopoeia (EP) [162] and the FDA [163]. DPI inhaler design, especially the geometry of the mouth piece, is critical for patients to produce an airflow sufficient to lift the drug from the dose chamber, break up the agglomerates in a turbulent air stream, and deliver a dose to the lungs as therapeutically effective fine particles [164, 165]). The airflow generated by inhalation directly determines particle velocity and hence the ease with which particles are deagglomerated. The materials used in the construction of DPIs and the characteristics of the formulation affect electrostatic charge accumulation. Some formulations as well as inhaler materials accumulate and retain electrostatic charge more strongly than others, and this will affect both drug retention within these inhalers as well as delivered aerosol behavior [166].

Pulmozyme®, recombinant deoxyrucleic acid (DNA)–derived human DNase I (rhDNase) has been formulated for local delivery to the lung by inhalation. rhDNase was filled originally in glass vials for clinical studies. A direct stability comparison study was made between rhDNase in glass vials and rhDNase filled into plastic ampoules using the blow-fill seal technology of Alp, which uses low-density polyethylene resins which are gas permeable. The problem of gas permeation was addressed by packing the plastic ampoules in a gas-impermeable foil pouch which can be filled with nitrogen. The rate of deamindation was similar for rhDNase stored at 2–8°C in foiled and unfoiled ampoules but substantially different from protein stored in glass vials. This difference in deamidation rate was attributed to the 0.5-unit difference in pH of rhDNase in glass (due to leaching of ions, possibly sodium, from the glass surface) versus plastic ampoules. This result provided the rationale for the choice of container closure component for rhDNase aerosols [167].

Improvements in the design of pMDIs has facilitated the extension of pMDIs to administering macromolecules such as peptides and proteins. The belief is that the HFA environment within the pMDI is inert and essentially moisture free, thereby providing good stability for macromolecules. Further, advances in valve and actuator technology in the pMDI help ensure a consistent and efficient delivery, an important consideration based on the cost and potency of biotechnological drug substances [168]. The 3M company has made improvement in container closure systems (CCSs) for biopharmaceuticals. One improvement reduces the potentially problematic deposition of the active drug within the CCS by introducing can coating. Further 3M is using a novel semipermeable membrane system component to improve the dosing reproducibility of suspension formulations [169]. The company is also developing new fast-fill, fast-empty valve designs to improve dose-to-dose uniformity as well as to avoid the need for patients to prime the system if only used intermittently [170–172].

***Methods for Prediction of Effects of Packaging on Stability of Drug Products***

Drug stability studies often involve multiple batches to ensure that a product will

consistently remain within specifications for its entire expiration date. The studies usually involve the same drug products in similar packages or in multiple strengths. The belief now is that application of sound statistical design principles can reduce the amount of testing required. The principles stated in the FDA 1987 publication Guideline for Submitting Documentation for the Stability of Human Drugs and Biologics have been extended to setting expiration dating periods for more complex situations [173]. PP may be used as a substitute for glass in primary packaging for various drug products. However, the diffusion of water through the PP plastic wall often builds up the water content with time during the long-term storage. It is often the shelf life limiting parameter for chemically stable aqueous solutions in PP bottles. The water diffusion rates of the widely distributed X-ray contrast agent VisipaquesTM in PP bottles and magnetic resonance contrast agent OmniscanTM in PP syringes were evaluated with the objective of developing a mathematical method of estimating the rate constant for water diffusion through the PP wall as a function of the following variables: temperature, humidity, surface area, wall thickness, concentration of the active ingredient, and fill volume. The effect of the variables was estimated by partial least-squares regression. The predictive ability of the cross-validated models was good for the two active agents. The models were used to predict the shelf life for the relevant combinations of temperature and humidity for the four climatic zones. The method presents an opportunity to measure the effect of the variables influencing drug stability [174].

A multivariate model has been developed to identify factors influencing stability, to estimate shelf life, to select new batches for further stability testing, and to evaluate changes in new batches. The model was capable of predicting the degradation rate constant as a function of storage temperature, pH, product concentration, and container volume [175]. Other investigators have reported on the applications of statistical analysis in stability studies of drug products. The time dependency of product change has been examined by ordinary least-squares regression and variance analysis for testing the possibility of pooling batches to increase the precision in shelf life estimates [176]. The Bayesian approach has been used to quantify the uncertainty in predicting shelf life from the Arrhenius equation as a function of stability consisting of various error distributions [177]. Monte Carlo simulation has been used to test the power of the analysis of variance (ANOVA) on matrixing data and single data to estimate shelf life. The simulation showed that the large amount of data from a matrixing design gave more precise shelf life estimates [178].

Desiccant is frequently included in the packaging of moisture-sensitive products in order to maintain low relative humidity inside the package and hence protect the product from moisture. Sorption–desorption moisture transfer (SDMT) models have been successfully applied to predict moisture transfer between a solid product and a desiccant inside a closed package. Further, theoretical simulations extended the use of the SDMT model to take into account the moisture permeation properties of the package [179]. The SDMT model was also used to predict the effect of desiccant quantity, tablet quantity, and tablet initial moisture content on the relative humidity inside high-density polyethylene (HDPE) bottles containing a moisture-sensitive drug product, roxifiban tablets. Desiccant quantity, tablet weight, and initial moisture content before packaging were found to have an effect on stability. The use of theoretical calculations utilizing the SDMT model was shown to be useful in the understanding of packaging requirements for the product. Calculated relative

humidity data corroborated experimental findings regarding the effect of the variables studied on tablet stability [180]. It is known that hard gelatin capsule brittleness is a function of moisture content. A study has been carried out to indicate that the brittleness of empty capsules occurred at humidities below 40%. Mexitil-loaded capsules exhibited a similar profile. The SDMT model was employed to estimate the final relative humidity for the Mexitil/gelatin capsule system and results were presented to demonstrate the general applicability of the SDMT model for predicting the incidence of brittleness problems and for the formulation to ensure the absence of brittleness [181].

The sorption of two weak acids (warfarin and thiopentone) and two weak bases (chlorpromazine and diltiazem) into PVC infusion bags was described by a constant partition model. PVC–water partition coefficients were obtained using three different methods: equilibrium values for sorption into PVC bags, the sorption versus pH relationship, and partition into PVC strips. The data were compared with similar values derived from a liquid–liquid partition system and different organic solvents (octanol, dichloromethane, carbon tetrachloride, and hexane). Octanol is the preferred solvent, and it is suggested that octanol–water partition data can be used to predict sorption behavior [182].

Reports have shown that changes in the hardness of tablets composed of lactose and cornstarch were due to variation of the moisture content. Further it was shown that the hardness of the tablets in moisture-semipermeable packages (such as strip packs and press-through packs with or without an overwrap film) could be predicted by an iterative calculation procedure through a mathematical model based on the physicochemical properties of the tablets and the moisture permeabilities of the packaging materials [183]. In another study by the same research group, moisture and temperature were used to predict the shelf-life of packaged tablets. Cognizance was taken of the flunctuations of temperature and relative humidity during prolonged storage. A sugar-coated tablet with a core containing ascorbic acid whose color is affected by both moisture content and ambient temperature was investigated. It was found that the color change of ascorbic acid core of the sugar-coated tablet was dependent on its moisture content and the ambient temperature and that the color change of the tablet in moisture-semipermeable packages could be predicted by an iterative calculation procedure using a mathematical model based on the kinetics of the color change and the moisture permeabilities of the packaging materials [184].

It is believed that a container/formulation couple is compatible if the magnitude of ingredient loss is within acceptable limits over the entire shelf life of the product. In this connection an approach for assessing the interaction of drug with PVC plastic infusion bags was developed. The approach correlates the partition coefficients and dissociation constant (when appropriate) of the solute, the physical dimensions of the container, and the solution pH with single parameters that dictate the shape of the sorption profile. To determine the equilibrium sorption level of PVC containers, the fractional binding of a solute was correlated with its hexane–water and octanol–water partition coefficients. Calculations based on single partition coefficients are believed to be less effective in terms of mimicking the behavior of PVC. To determine the sorption profile (fractional binding with time), the partition coefficients are related to the fractional binding at a particular time through a single parameter referred to as the sorption number. Equilibrium fractional binding and

sorption profiles for various drugs stored in PVC containers are generated with the models and agree well with reported behavior [185]. The relationship between sorption number (a parameter defining initial solute uptake by PVC infusion bags) and solute octanol–water partition coefficient was investigated by following the time course of the sorption of drugs by PVC infusion bags which was approximated using a diffusion model in which the plastic was assumed to act as an infinite sink. The model was found to be suitable for estimation of storage times relevant to clinical usage and enabled the magnitude of the uptake in a specific time to be described by a single parameter referred to as the sorption number. This parameter was defined by the plastic infusion solution partition coefficient, the diffusion coefficient in the plastic, the fraction un-ionized in the solution, the volume of the infusion solution, and the surface area of the plastic. The sorption number can be extrapolated to allow prediction of the effects of time, plastic surface area, solution volume, and solution pH on fractional solute loss. A reasonable correlation was established between the logarithm of this parameter and the logarithm of the octanol–water partition coefficient of various solutes. The model allows the fraction of a solute remaining in a plastic infusion bag at a given storage time to be estimated from the octanol–water partition coefficient of the solute and other readily available data [186].

## REFERENCES


1. U.S. Pharmacopeia/National Formulary (2005), *The Official Compendia of Standards*, USP 28/NF23, U.S. Pharmacopeial Convention, Rockuille, Maryland.
2. International Conference on Harmonization (ICH) Steering Committee (1994), ICH harmonized tripartite guideline, Stability testing of new drug substances and products, ICH, Geneva.
3. International Conference on Harmonization (ICH) Steering Committee (1998), Draft ICH harmonized tripartite guideline, Stability testing of new drug substances and products, ICH, Geneva.
4. Rhodes, C. T. (2000), Reasons for stability testing, in Carstensen, J. T., and Rhodes, C.T. Eds., *Drug Stability, Principles and Practices*, Marcel Dekker, New York, p. 11.
5. Center for Drugs and Biologics, U.S. Food and Drug Administration (FDA), Department of Health and Human Services (1987), Guideline for submitting documentation for the stability of human drugs and biologics, CDB, Washington, DC.
6. Yoshioka, S., and Stella, V. J. (2000), *Stability of Drugs and Dosage Forms*, Kluwer Academic, New York.
7. Guillory, J. K., and Poust, R. I. (1996), Chemical kinetics and drug stability, in Banker, G. S., and Rhodes, C. T., Eds., *Modern Pharmaceutics*, 3rd ed., Marcel Dekker, New York, p. 179.
8. Center for Drug Evaluation and Research and Center for Biologics Evaluation and Research, U.S. Food and Drug Administration (FDA), Department of Health and Human Services (1999), Guidance for industry: Container closure systems for packaging human drugs and biologics, FDA, Washington, DC.
9. Florence, A. T., and Attwood, D. (2006), *Physicochemical Principles of Pharmacy*, 4th ed., Pharmaceutical Press, London, p. 94.
10. Cartensen, J. T., Serenson, E. G., and Vance, J. J. (1964), Use of Hammett graphs in stability programs, *J. Pharm. Sci.*, 53, 1547–1548.

11. Moorhatch, P., and Ciou, W. L. (1974), Interactions between drugs and plastic intravenous fluid bags. II. Leaching of chemicals from bags containing various solvent media, *Am. J. Hosp. Pharm.*, 31, 149–152.
12. Venkataramanan, R., Burckart, G. J., Ptachcinski, R. J., Blaha, R., Logue, L. W., Bahnson, A., Giam, C., and Brady, J. E. (1986), Leaching of diethyl phthalate from polyvinyl chloride bags into intravenous cyclosporine solution, *Am. J. Hosp. Pharm.*, 43, 2800–2802.
13. Scheiffer, G. W., Palermo, P. J., and Pollard-Walker, S. (1984), Simultaneous determination of methyl, ethyl, propyl, and butyl 4-hydroxybenzoates and 4-hydroxybenzoic acid in liquid antacid formulations by gas chromatography, *J. Pharm. Sci.*, 73, 128.
14. Ullmann, E., Thoma, K., and Zelfel, G. (1963), The stability of sodium penicillin G in the presence of ionic surfactants, organic gel formers, and preservatives, *Pharm. Acta. Helv.*, 38, 577–586.
15. Testa, B., and Etter, J. C. (1975), Hydrolysis of pilocarpine in carbopol hydrogels, *Can. J. Pharm. Sci.*, 10, 16–20.
16. Amidon, G. L., and Middleton, K. R., (1988), Accelerated physical stability testing and long-term predictions of changes in the crushing strength of tablets stored in blister packages, *Int. J. Pharm.*, 45, 79–89.
17. Taborsky-Urdinola, C. J., Gray, V. A., and Grady, L. T. (1981), Effects of packaging and storage on the dissolution of model prednisone tablets, *Am. J. Hosp. Pharm.*, 38, 1322–1327.
18. Gouda, M. W., Moustafa, M. A., and Molokhia, A. M. (1980), Effect of storage conditions on erythromycin tablets marketed in Saudi Arabia, *Int. J. Pharm.*, 5, 345–347.
19. Barrett, D., and Fell, J. T. (1975), Effect of aging on physical properties of phenylbutazone tablets, *J. Pharm. Sci.*, 64, 335–337.
20. Khalil, S. A., Ali, L. M., and Abdel-Khalek, M. M. (1974), Effects of aging and relative humidity on drug release. *Pharmazie*, 29, 36–37.
21. Georgarakis, M., Htzipantou, P., and Kountourelis, J. E. (1988), Effect of particle size, content in lubricant, mixing time, and storage relative humidity on drug release from hard gelatin ampicillin capsules, *Drug Dev. Ind. Pharm.*, 14, 915–923.
22. Dahl, T. C., Sue, I. T., and Yum, A. (1991), The effect of pancreatin on the dissolution performance of gelatin-coated tablets exposed to high-humidity conditions, *Pharm. Res.*, 8, 412–414.
23. Kornblum, S. S., and Sciarrone, B. J. (1964), Decarboxylation of p-aminosalicylic acid in the solid state, *J. Pharm. Sci.*, 53, 935.
24. Vromans, H., and van Laarhoven, J. A. H. (1992), A study on water permeation through rubber closures of injection vials, *Int. J. Pharm.*, 79(1–3), 301–308.
25. Lamy-Freud, M. T., Ferreira, V. F. N., Faljoni-Alario, A., and Schreiter, S. (1993), Effect of aggregation on the kinetic of autoxidation of the polyene antibiotic amphotericin B, *J. Pharm. Sci.*, 82, 162–166.
26. Smith, G. B., DiMichele, L., Colwell, L. F., et al. (1993), Autooxidation of simvastin, *Tetrahedron*, 49, 4447–4462.
27. Khan, M. M. T., and Martel, A. E. (1967), Metal ion and metal chelate catalyzed oxidation of ascorbic acid by molecular oxygen. I. Cupric and ferric ion catalyzed oxidation, *J. Am. Chem. Soc.*, 89, 4176–4185.
28. Jensen, J., Cornett, C., Olsen, C. E., Tjornelund, J., and Hansen, S. H. (1992), Identification of major degradation products of 5-aminosalicylic acid form in aqueous solutions and in pharmaceuticals, *Int. J. Pharm.*, 88, 177–187.
29. Yeh, S., and Lach, J. L. (1961), Stability of morphine in aqueous solution III. Kinetics of morphine degradation in aqueous solution, *J. Pharm Sci.*, 50, 35–42.

30. Pitman, I. H., Higuchi, T., Alton, M., and Wiley, R. (1972), Deutrium isotope effects on degradation of hydrocortisone in aqueous solution. *J. Pharm. Sci.*, 61, 818–820.
31. Boccardi, G., Deleuze, C., Gachon, M., Palmisano, G., and Vergnaud, J. P. (1992), Autoxidation of tetrazepam in tablets: Prediction of degradation impurities from oxidative behavior in solution, *J. Pharm. Sci.*, 81, 183–185.
32. Greehill, J. V., and McLelland, M. A. (1990), Photodecomposition of drugs, *Prog. Med. Chem.*, 27, 51–121.
33. International Conference on Harmonization (ICH) (1996), Harmonized tripartite guideline: on stability testing: Photostability testing of new drug substances and products, Q1B, 1996 ICH, Geneva.
34. Matsuda, Y., Inouye, H., and Nakanishi, R. (1978), Stabilization of sulfisomidine tablets by use of film coating containing UV absorber: Protection of coloration and photolytic degradation from exaggerated light, *J. Pharm. Sci.*, 67, 196–201.
35. Frank, M. J., Johnson, J. B., and Rubin, S. H. (1976), Spectrophotometric determination for sodium nitroprusside and its photodegradation products, *J. Pharm. Sci.*, 65, 44.
36. Asker, A. F., and Larose, M. (1987), Influence of uric acid on photostability of sulfathiazole sodium solutions, *Drug Dev. Ind. Pharm.*, 13, 2239.
37. Asker, A. F., Canady, D., and Cobb, C. (1985), Influence of DL-methionine on the photostability of ascorbic acid solutions, *Drug Dev. Ind. Pharm.*, 11, 2109.
38. Vandenbossche, G. M., deMuynk, C., Colardyn, F., and Remon, J. P. (1993), Light stability of molsidomine in infusion fluids, *J. Pharm. Pharmacol.*, 45, 486.
39. Akimoto, K., Kurosaka, K., Nakagawa, H., and Sugimoto, I. (1988), A new approach to evaluating photo-stability of nifedipine and its derivatives in solution by actinometry, *Chem Pharm. Bull.*, 36, 1483–1490.
40. Thoma, K., and Klimek, R. (1991), Photostabilization of drugs in dosage forms without protection from packaging materials, *Int. J. Pharm.*, 67, 169–175.
41. Matsuda, Y., and Minamida, Y. (1976), Stability of solid dosage forms: II. Coloration and photolytic degradation of sulfisomide tablets by exaggerated ultraviolet irradiation, *Chem. Pharm. Bull.*, 24, 2229–2236.
42. Murthy, K. S., Reisch, R. G., and Fawzi, M. B. (1989), Dissolution stability of hard-shell capsule products: Part I: The effects of exaggerated storage conditions, *Pharm. Technol.*, 13(3), 72–86.
43. Teraoka, R., Konishi, Y., and Matsuda, Y. (2001), Photochemical and oxidative degradation of the solid-state tretinoin tocoferil, *Chem. Pharm. Bull.*, 49(4), 368–372.
44. Matsuda, Y., Itooka, T., and Mitsuhashi, Y. (1980), Photostability of indomethacin in model gelatin capsules: effects of film thickness and concentration of titanium dioxide on the coloration and photolytic degradation, *Chem. Pharm. Bull.*, 28, 2665–2671.
45. Desai, D. S., Abdelnasser, M. A., Rubitski, B. A., and Varia, S. A. (1994), Photostabilization of uncoated tablets of sorivudine and nifedipine by incorporation of synthetic iron oxides, *Int. J. Pharm.*, 103, 69–76.
46. Carstensen, J. T., Johnson, J. B., Spera, D. C., and Frank, M. J. (1968), Equilibrium phenomena in solid dosage forms, *J. Pharm. Sci.*, 57, 23
47. Garrett, E. R., and Carper, R. F. (1955), Prediction of stability in pharmaceutical preparations I. Color stability in a liquid multisulfa preparations. *J. Am. Pharm. Assoc. Sci. Ed.*, 44, 515–518.
48. Garrett. E. R. (1956), Prediction of stability in pharmaceutical preparations II. Vitamin stability in liquid multivitamin preparations. *J. Am. Pharm. Assoc. Sci. Ed.*, 45, 171–178.

49. Garrett, E. R. (1956), Prediction of stability in pharmaceutical preparations III. Comparison of vitamin stabilities in different multivitamin preparations. *J. Am. Pharm. Assoc. Sci. Ed.*, 45, 470–473.
50. Powell, M. F. (1986), Enhanced stability of codeine sulfate: Effect of pH, buffer, and temperature on the degradation of codeine in aqueous solution, *J. Pharm. Sci.*, 75, 901.
51. Carney, C. F. (1987), Solution stability of ciclosidomine, *J. Pharm. Sci.*, 76, 393.
52. Magalhales, N. S. S., Cave, G., Seiler, M., and Benita, S. (1991), The Stability and in vitro release kinetics of clofibride emulsion, *Int. J. Pharm.*, 76, 225–237.
53. Shi, L., DeHaven, P. A., and Burke, C. J. (2005), Biopharmaceutical stability studies: stable vaccine dosage form development for commerical use, *Am. Pharm. Rev.*, 8(6), 86–92.
54. Tydeman, M. S., and Kirkwood, T. B. L. (1984), Design and analysis of accelerated degradation tests for the stability of biological standards. I. Properties of maximum likelihood estimators, *J. Biol. Stand.*, 12, 195–206.
55. Flynn, G. L. (1996), Cutaneous and transdermal delivery: Processes and systems of delivery, in Banker, G. S., and Rhodes, C. T., Eds., *Modern Pharmaceutics*, 3rd ed., Marcel Dekker, New York, p. 239.
56. Carstensen, J. T. (1974), Stability of solids and solid dosage forms, *J. Pharm. Sci.*, 63(1), 1–14.
57. Horhota, S. T., Burgio, J., Lonski, L., and Rhodes, C. T. (1976), Effect of storage at specified temperature and humidity on properties of three directly compressible tablet formulations, *J. Pharm. Sci.*, 65, 1746–1749.
58. Alam, A. S., and Parrott. E. L. (1971), Effect of aging on some physical properties of hydrochlorothiazide tablets, *J. Pharm. Sci.*, 60, 263–266.
59. Carstensen, J. T., and Kothari, R. (1983), Solid-state decomposition of alkoxyfuroix acids in the presence of microcrystalline cellulose, *J. Pharm. Sci.*, 72, 1149.
60. Ansel, H. C., Allen, L. V., and Popovich, N. G. (1999), *Pharmaceutical Dosage Forms and Drug Delivery Systems*. 7th ed., Lippincott Williams and Wilkins, Baltimore, MD, p. 98.
61. Hecht, G., Roehs, R. R., Lang, J. C., Rodheaver, D. P., and Chowhan, M. A. (1996), Design and evaluation of ophthalmic pharmaceutical products, in Banker, G. S., and Rhodes, C. T., Eds., *Modern Pharmaceutics*, 3rd ed., Marcel Dekker, New York, p. 489.
62. Lee, J. Y. (1992), Sterilization control and validation for topical ointments, *Pharm. Tech.*, 16, 104–110.
63. Ibrahim, Y. K. E., and Olurinola, P. R. (1991), Comparitive microbiological contamination levels in wet granulation and direct compression methods of tablet production, *Pharm. Acta Helv.*, 66, 298.
64. Carstensen, J. T. (1996), Preformulation, in Banker, G. S., and Rhodes, C. T., Eds., *Modern Pharmaceutics*, 3rd ed., Marcel Dekker, New York, p. 213.
65. Guillory, J. K., and Higuchi, T. (1962), Solid state stability of some vitamin a compounds, *J. Pharm. Sci.*, 51, 100–105.
66. Pikal, M. J., Lukes, A. L., Lang, J. E., and Gaines, K. (1978), Quantitative crystalline determinations for b–lactam antibiotics by solution calorimetry: Correlation with stability, *J. Pharm. Sci.*, 67, 767–773.
67. Krahn, F. U., and Mieck, J. B. (1989), Effect of type and extent of crystalline order on chemical and physical stability of carbamazepine, *Int. J. Pharm.*, 53, 25–34.
68. Matsuda, Y., Akazawa, R., Teraoka, R., and Totsuka, M. (1994), Pharmaceutical evaluation of carbamazepine modifications: Comparative study of photostability of carbamazepine polymorphs by using fourier-transformed reflection-absorption infrared spectrocopy and colorimetric measurement, *J. Pharm. Pharmacol.*, 46, 162–167.

69. DeVilliers, M. M., van der Watt, J. G., and Lotter, A. P. (1992), Kinetic study of the solid-state photolytic degradation of two polymorphic forms of furosemide, *Int. J. Pharm.*, 88, 275–283.
70. Nakagawa, Y., Itai, S., Yoshida, T., and Nagai, T. (1992), Physicochemical properties and stability in the acidic solution of a new macrolide antibiotic, clarithromycin, In comparison with erythromycin, *Chem. Pharm. Bull.*, 40, 725–728.
71. Pikal, M., and Lukes, A. L. (1976), Kundsen vapor pressure measurements on pure materials and solutions dispersed in porous media: Molded nitroglycerin tablets, *J. Pharm. Sci.*, 65, 1269.
72. Fusari, S. A. (1973), Nitroglycerin sublingual tablets II: preparation and stability of a new stabilized sublingual molded nitroglycerin tablet, *J. Pharm. Sci.*, 62, 2012.
73. Finholt, P., Jurgensen, G., and Kristiansen, H. (1965), Catalytic effect of buffers on degradation of penicillin G in aqueous solution, *J. Pharm. Sci.*, 54, 387–393.
74. Tsuji, A., Nakashima, E., Deguchi, Y., Nishide, K., Shimizu, T., Horiuchi, S., Ishikawa, K, and Yamana, T. (1981), Degradation kinetics and mechanism of aminocephalosporins in aqueous solution: Cefadroxil, *J. Pharm. Sci.*, 70, 1120–1128.
75. Zia, H., Teharan, M., and Zargarbashi. R. (1974), Kinetics of carbenicillin degradation in aqueous solutions, *Can. J. Pharm. Sci.*, 9, 112–117.
76. Pramar, Y., and Gupta, V. D. (1991), Preformulation studies of spironolactone: Effect of pH, two buffer species, ionic strength, and temperature on stability, *J. Pharm. Sci.*, 80, 551–553.
77. Allen, A. E., and Das Gupta, V. (1974), Stability of hydrocortisone in polyethyleneglycol ontiment base, *J. Pharm. Sci.*, 63, 107–109.
78. Das Gupta, V. (1978), Effect of vehicles and other active ingredients on stability of hydrocrtisone, *J. Pharm. Sci.*, 67, 299–302.
79. Jun, H. W., Whitworth, C. W., and Luzzi, L. A. (1972), Decomposition of aspirin in polyethyleneglycols, *J. Pharm. Sci.*, 61, 1160–1162.
80. Whitworth, C. W., Luzzi, L. A., Thompson, B. B., and Jun, H. W. (1973), Stability of aspirin in liquid and semi-solid bases. II. Effect of fatty additives on stability in a polyethyleneglycol base, *J. Pharm. Sci.*, 62, 1372–1374.
81. Ekman, R., Liponkoski, L., and Kahela, P. (1982), Formation of indomethacin esters in polyethylene glycol suppositories, *Acta Pharm. Suec.*, 19, 241–246.
82. Castello, R. A., and Mattocks, A. M. (1962), Discoloration of tablets containing amines and lactose, *J. Pharm. Sci.*, 51, 106–108.
83. Tonnesen, H. H. (2001), Formulation and stability testing of photolabile drugs, *Int. J. Pharm.*, 225, 1–14.
84. Baugh, R., Calvert, R. T., and Fell, J. T. (1977), Stability of phenylbutazone in the presence of pharmaceutical colors, *J. Pharm. Sci.*, 66, 733–735.
85. Kornblum, S. S., and Sciarrone, B. J. (1964), Decarboxylation of p-aminosalicylic acid in the solid state, *J. Pharm. Sci.*, 53, 935.
86. Liebe, D. C. (1996), Packaging of pharmaceutical dosage forms, in Banker, G. S., and Rhodes, C. T., Eds., *Modern Pharmaceutics*, 3rd ed., Marcel Dekker, New York, p. 681.
87. Dean, D. A. (2000), Packaging, package evaluation, stability and shelf-life, in Carstensen, J. T., and Rhodes, C. T., Eds., *Drug Stability, Principles and Practices*, Marcel Dekker, New York, p. 483.
88. *Code of Federal Regulations*, Food and drugs, Title 21, Part 211, Current good manufacturing practice for finished pharmaceuticals, U.S. Government Printing Office, Washington, DC, rev. Apr. 1, 2001.

89. Croce, C. P., Fischer, A., and Thomas, R. H. (1986), Packaging materials science, in, Lachman, L., Lieberman, H. A., and King, J. L., Eds., *The Theory and Practice of Industrial Pharmacy*, 3rd ed., Lea & Febiger, Philadelphia, p.711.
90. Sciarra, J. J. (1996), Pharmaceutical aerosols, in Banker, G. S., and Rhodes, C. T. Eds., *Modern Pharmaceutics*, 3rd ed., Marcel Dekker, New York, p. 547.
91. Lusina, M., Cindric, T., Tomaic, J., Peko, M., Pozaic, L., and Musulin, N. (2005), Stability study of losartan/hydrochlorothiazide tablets, *Int. J. Pharm.*, 291, 127–137.
92. Allinson, J. G., Dansereau, R. J., and Sakr. A. (2001), The effects of packaging on the stability of a moisture sensitive compound, *Int. J. Pharm.*, 221, 49–56.
93. Mahajan, R., Templeton, A., Harman, A., Reed, R. A., and Chern, R. T. (2005), The effect of innert atmospheric packaging on oxidative degradation in formulated granules, *Pharm. Res.*, 22(1), 128–140.
94. Al-Zein, H., Riad, L. E., and Abd-Elbary, A. (1999), Effect of packaging and storage on the stability of carbamazepine tablets. *Drug Dev. Ind. Pharm.*, 25(2), 223–227.
95. Veillard, M., Bentejac, R., Duchene, D., and Carstensen, J. T. (1979), Moisture transfer tests in blister package testing, *Drug Dev. Ind. Pharm*. 5, 227–244.
96. Amidon, G. E., and Middleton, K. R. (1988), Accelerated physical stability testing and long-term predictions of changes in the crushing strength of tablets stored in blister packages, *Int. J. Pharm.*, 45,79–89.
97. Khalil, S. A. H., Barakat, N. S., and Boraie, N. A. (1991), Effects of package type on in vitro release and chemical stability of amoxycillin in capsules, *Pharm. Ind.*, 53(7), 698–701.
98. Gouda, M. W., Moustafa, M. A., and Molokhia, A. M. (1980), Effect of storage conditions on erythromycin tablets marketed in saudi arabia, *Int. J. Pharm.*, 5, 345–347.
99. Pikal, M. J., Bibler, D. A., and Rutherford, B. (1977), Polymer sorption of nitroglycerin and stability of molded nitroglycerin tablets in unit-dose packaging, *J. Pharm. Sci.*, 66(9), 1293–1297.
100. Barakat, N. S. (2006), Etodolac-liquid-filled dispersion into hard gelatin capsules: An approach to improve dissolution and stability of etodolac formulation, *Drug Dev. Ind. Pharm.*, 32, 865–876.
101. Santoro, M. I. R. M., Oliveira, D. A. G. D. C. E., Kedor-Hackmann, E. R., and Singh, A. K. (2005), The effect of packaging materials on the stability of sunscreen emulsions, *Int. J. Pharm.*, 297, 197–203.
102. Geary, T. G., Akkod, M. A., and Jensen, J. B. (1983), Characteristics of cheloroquine binding to glass and plastic, *Am. J. Trop. Med. Hyg*. 32, 19–23.
103. Yahya, M. A., McElnay, J. C., and D'Arcy, P. F. (1985), Binding of chloroquine to glass, *Int. J. Pharm.*, 25, 217–223.
104. Stanaszek, W. F., and Pan, I. H. (1978), Comparison of drug stability in glass versus plastic containers: Analysis of prefilled syringe admixtures, *Proc. Okla. Acad. Sci.*, 58, 102–105.
105. Arsene, M., Favetta, P., Favier, B., and Bureau, J. (2002), Comparison of ceftazidime degradation in glass bottles and plastic bags, *J. Clin. Pharm. Ther.*, 27(3), 202–209.
106. Crane, I. M., Mulhern, M. G., and Nema, S. (2003), Stability of reconstituted parecoxib for injection with commonly used diluents. *J. of Clin. Pharm. Thera.*, 28, 363–369.
107. Wong, M., Marion, R., Reed, K., and Wang, Y. (2006), Sorption of unoprostone isopropyl to packaging materials, *Int J. Pharm*. 307, 163–167.
108. Driscoll, D. F., Ling, P. R., and Bistrian, B. R. (2007), Physical stability of 20% lipid injectable emulsions via simulated syringe infusion: effects of glass versus plastic product packaging, *J. Parenteral Enteral Nutr.*, 31(2), 148–153.

109. Yuen, P. H., Denman, S. L., Sokoloski, T. D., and Burkman, A. M. (1979), Loss of nitroglycerin from aqueous solution into plastic intravenous delivery systems. *J. Pharm. Sci.*, 68, 1163–1166.
110. Baaske, D. M., Amann, A. H., Wagenknecht, D. M., Carter, J. E., Hoyt, H. J., and Stoll, R. G. (1980), Nitroglycerin compatibility with intravenous fluid filters, containers and administration sets, *Am. J. Hosp. Pharm.*, 37, 201–205.
111. Cossum, P. A., Roberts, M. S., Galbraith, A. J., and Boyd, G. W. (1978), Loss of nitroglycerin from intravenous infusion sets, *Lancet.*, 2, 34–350.
112. Crouthamel, W. G., Dorsch, B., and Shangraw, R. (1978), Loss of nitroglycerin from plastic intravenous bags, *New Engl. J. Med.*, 299, 262.
113. Sokoloski, T. D., Wu, C. C., and Burkman, A. M. (1980), Rapid adsorptive loss of nitroglycerin from aqueous solution to plastic, *Int. J. Pharm.*, 6, 63–76.
114. Malick, A. W., Amann, A. H., Baaske, D. M., and Stoll, R. G. (1981), Loss of nitroglycerin from solutions to intravenous plastic containers: A theoretical treatment. *J. Pharm. Sci.*, 70, 798–800.
115. Yuen, P. H., Denman, S. L., Sokoloski, T. D., and Burkman, A. M. (1979), Loss of nitroglycerin from aqueous solution into plastic intravenous delivery systems, *J. Pharm. Sci.*, 68(9), 1193–1166.
116. Illum, L., and Bubdgaard, H. (1982), Sorption of drugs by plastic infusion bags, *Int. J. Pharm.*, 10, 339–351.
117. Pikal, M. J., and Lang, J. E. (1978), Rubber closures as a source of haze in freeze-dried parenterals: Test methodology for closure evaluation, *J. Parent. Drug. Assoc.*, 32, 162–173.
118. Jahnke, R. W. O., Kreuter, J., and Ross. G. (1991), Content/container interactions: The phenomenon of haze formation on reconstitution of solids for parenteral use, *Int. J. Pharm.*, 77, 47–55.
119. Parker, W. A., and MacCara, M. E. (1980), Compatibility of diazepam with intravenous fluid containers and administration sets, *Am. J. Hosp. Pharm.*, 37, 496–500.
120. Morris, M. E. (1978), Comaptibility and stability of diazepam injection following dilution with intravenous fluids, *Am. J. Hosp. Pharm.*, 35, 669–672.
121. Parker, W. A., Morris, M. E., and Shearer, C. A. (1979), Incompatibility of diazepam injection in plastic intravenous bags, *J. Am. Hosp. Pharm.*, 36, 505–507.
122. Van Amerongen, G. J. (1964), Diffusion in elastomers, *Rubber Chem. Technol.*, 37, 1065–1152.
123. Vromans, H. and van Laarhoven, J. A. H. (1992), A study on water permeation through rubber closures of injection vials, *Int. J. Pharm.*, 79, 301–308.
124. U.S. Food and Drug Administration (FDA) (1998, Oct.), Guidance for industry: Metered dose inhaler and dry powder inhaler drug products, FDA Washington, DC.
125. Castner, J., Anderson, J., and Benites, P. (2007), Strategy for development and characterization of HPLC methods to investigate extractables and leachables, *Am. Pharm. Rev.*, 10(3), 10.
126. Northup, S. J. (2005), Assesing the biological safety of extractable and leachable chemicals in pharmaceutical and medicial products, *Am. Pharm. Rev.*, 8(4), 38–43.
127. U.S. Food and Drug Administration (FDA) (2002, July), Nasal spray and inhalation solution, suspension, and spray drug products.: Chemistry, manufacturing, and controls documentation, Guidance for industry. FDA, Washington, DC, 1–45.
128. Norwood. D. L. (2007), Understanding the challenges of extractables and leachables for the pharmaceutical industry—Safety and requlatory enviroment for pharmaceuticals, *Am. Pharm. Rev.*, 10(2), 32.

129. U.S. Food and Drug Administration (FDA), Extra 6: Safety thresholds and best practices for extractables and leachables in orally inhaled and nasal drug products, PQRI Leachables and Extratables Working Group, available: http://www.pqri.org/pdfs/LE_Recommendations_to_FDA_09-26-06.pdf.
130. IPAC-RS (1991, Mar.) Extra 7: Leachables and extractables testing: points to consider, available: http://www.ipacrs.com/leachables.html.
131. Allan, L., and Wang, Q. (2007), Impact of package on the stability of pharmaceutical products, *Am. Pharm. Rev*., 10(4), 38–44.
132. Waterman, K. C., Adami, R. C., Alsante, K. M., Hong, J., Landis, M. S., Lombardo, F., and Roberts, C. J. (2002), Stabilization of pharmaceuticals to oxidative degradation, *Pharm. Dev. Tech*., 7(1), 1–32.
133. Reed, R. A., Harmon, P., Manas, D., Wasylaschuk, W., Galli, C., Biddell, R., Bergquist, P. A., Hunke, W., Templeton, A. C., and Ip, D. (2003), The role of excipients and package components in the photostability of liquid formulations, *PDA J. Pharm. Sci. Technol*., 57(5), 351–368.
134. Hong, J., Lee, E., Carter, J. C., Masse, J. A., and Oksanen, D. A. (2004), Antioxidant-accelerated oxidative degradation: A case study of transistion metal ion catalyzed oxidation in formulation, *Pharm. Dev. Tech*., 9(2), 171–179.
135. Quarry, M. A., Sebastian, D. S., and Diana, F. (2002), Investigation of 4,5-epoxymorphian degradation during analysis by HPLC, *J. Pharm. Biomed. Anal*. 30, 99–104.
136. Nakamura, K., Yokohama, S., Sawada, M. T., and Sonobe, T. (2003), A new stressed test to predict the foreign matter formation of minodronic acid in solution, *Int. J. Pharm*., 251, 99–106.
137. Bohere, D., Nascimento, P. C., Binotto, R., and Becker, E. (2003), Influence of the glass packing on the contamination of pharmaceutical products by aluminium. Part III. Interaction of the container-chemicals during heating for sterilization, *J. Trace Elem. Med. Biol*., 17(2), 107–115.
138. Shen, F. W., and McKellop, H. A. (2002), Interaction of oxidation and crosslinking in gamma-irradiated ultrahigh molecular-weight polyethylene, *J. Biomed. Mater. Res*., 61(3), 430–439.
139. Welle, F. (2005), Migration of radiolysis products from radiation-sterilized plastics, *Pharm. Ind*., 67(8), 970–972.
140. Nassar, M. N., Nesarker, V. V., Lozano, R., Yande, H., and Palaniswany, V. (2005), Degradation of a lyophilized formulation of BMS-204352: Identification of degradants and the role of elastomeric closure. *Pharm. Dev. Tech*., 10(2), 227–232.
141. Glaser, V. (2005), Addressing stability of biological drugs: forced degradation studies predict effects on bioproducts in drug development and manufacture, *GEN Genetic eng. news*, 25(6), 1–5.
142. *Code of Federal Regulations*, Food and drugs, Title 21, Part 211.65, U.S. Government Printing Office, Washington, DC, rev. Apr. 1, 2004.
143. Federal Food, Drug and Cosmetic Act as amended through Dec. 21, 2000, Chapter V, Drugs and devices, Section 501(a)(3), U.S. Government Printing Office, Washington DC, 2001.
144. Taborsky, C. J., and Sheinin, E. B. (2006), A critical approach to the evaluation of packaging components and the regulatory and scientific considerations in developing a testing strategy, *Am. Pharm. Rev*., 9(4), 10–15.
145. Gudat, A. E., and Firor. R. L. (2006, Sept.), The determination of extractables and leachables in pharmaceutical packaging materials using headspace GC/MS, Application Note, Agilent Technology.

146. Janimak, J. J., and Marteleur, M. (2004), On the stability of cerium oxide glass for terminal radiation sterilization, *Rad. Phys. Chem*. 71, 195–198.
147. DeGrazio, F. L. (2005), Parenteral packaging concerns for drugs: Early understanding of drug to package compatibility lowers the cost in the long run. *GEN: Genetic Eng. News (Bioprocess. Channel)*, 25(21), 2.
148. DeGrazio, F. L. (2005), Parenteral packaging concerns for drugs: Early understanding of drug to package compatibility lowers the cost in the long run. *GEN: Genetic Eng. News (Bioprocess. Channel)*, 25(21), 2.
149. Romacker, M. (2005), Prefilled syringes: Why new developments are important in injectable delivery today, in Roessling, G., Ed., *Prefilled Sringes: Innovations That Meet the Growing Demand*, ONdrugDelivery, Sussex, United Kingdom.
150. Wang, W., Wang, Y. J., and Kelner, D. N. (2003), Coagulation factor VIII: Strucuture and stability, *Int. J. Pharm*., 259, 1–15.
151. Schulman, S., Gitel, S., and Martinowitz, U. (1994), Stability of factor VIII concentrates after reconstitution, *Am. J. Hematol*., 45, 217–223.
152. Osterberg, T., Fatouros, A., Neidhardt, E., Warne, N., and Mikaelsson, M. (2001), B-domain-deleted recombinant factor VIII formulation and stability, *Semin. Hematol*., 38, 40–43.
153. Andrade, J. D., and Hlady, V. (1986), Protein adsorption and materials biocompatibility: A tutorial review and suggested hypothesis, *Adv. Polym. Sci*., 79, 1–63.
154. Duncan, M. R., Lee, J. M., and Warchol, M. P. (1995), Influence of surfactants upon protein/peptide adsorption to glass and polypropylene, *Int. J. Pharm*., 120, 179–188.
155. Wang, P. L., Udeani, G. O. and Johnston, T. P. (1995), Inhibition of granulocyte colony stimulating factor (c-csf) adsorption to polyvinyl chloride using a nonionic surfactant, *Int. J. Pharm*., 114, 177–184.
156. Burke, C. J., Steadman, B. L., Volkin, D. B., Tsai, P. K., Bruner, M. W., and Middaugh, C. R. (1992), The adsorption of proteins to pharmaceutical container surfaces, *Int. J. Pharm*., 86, 89–93.
157. Weber, S. S., Wood, W. A., and Jackson, E. A. (1977), Availability of insulin from parenteral nutrient solutions, *Am. J. Hosp. Pharm*., 34, 353–357.
158. Weeren, R. V., and Gibboni, D. J., Barrier packaging as an integral part of drug delivery, available: http://www.drugdeliverytech.com/cgi-bin/articles=52 (accessed June 1, 2007).
159. U. S. Patent 5,028,431: Franz, T. J., Shah, K. R., and Kydonieus, A. Article for the Delivery to Animal Tissue of a Pharmacological Active. July 2, 1991; U. Patent 5,242,433: Smith, J. A., and Murphy, B., Packaging System with In-Tandem Applicator Pads for Topical Drug Delivery. Decemeber 7, 1992.
160. U. S. Patent 5,698,217: Wilking, S. L. Transdermal Drug Delivery Device Containing a Dessicant. December 16, 1997.
161. Hoye, W. L., Mogalian, E. M., and Myrdal, P. B. (2005), The effects of extreme temperatures on drug delivery of albuterol sulfate hydrofluoroalkane inhalation products, *Am. J. Health-Syst. Pharm*., 62(21), 2271–2277.
162. European Pharmacopoeia (EP) (2001), Preparations for inhalation, EP supplement 2001: 2.9.18, EP, Strasbourg, France.
163. U.S. Food and Drug Administration (FDA) (1998), Guidance for industry: Metered dose inhaler (MDI) and dry powder inhaler (DPI) drug products, FDA CDER, Washington, DC.
164. Petersson, G., and Wiren, J. E. (1980), The brochodilator response from inhaled terbutaline is influenced by the mass of small particles: A study on a dry powder inhaler (Turbuhaler), *Eur. Respir*. 2, 253–256.

165. Newman, S.P., and Busse, W. W. (2002), Evolution of dry powder inhaler design, formulation and performance, *Respir. Med*., 96, 293–304.
166. Carter, P. A., Rowley, G., Fletcher, E. J., and Sylianopoulos, V. (1998), Measurement of electrostatic charge decay in pharmaceutical powdersand polymer materials used in dry powder inhaler devices, *Drug Dev. Ind. Pharm*., 24, 1083–1088.
167. Shire, S. J. (1996), Stability, characterization and formulation of recombinant human deoxyribonuclease (Pulmozyme®, Dornase Alpha), in Pearlman, R., and Wang, Y. J. Eds., *Formulation, Characterization and Stability of Protein Drugs*, Plenum Press, New York, pp. 393–426.
168. Arnum, P. V. (2007), 3M drug delivery systems advances inhalation and transdermal drug delivery, Pharm. Technol. Online, available http://www.pharmtech.com, Jan. 31.
169. Jinks, P., and Marsden, S. (1999), The development and performance of a fluoropolymer lined can for suspension metered dose inhaler product, in Fradley, G., Ed., Proceedings of Drug Delivery to the Lungsa X, The Aerosol Society, Portshead, United Kingdom, pp. 177–180.
170. Wilby, M. (2005), Increasing dose consistency of pMDis, *Drug Deliv. Technol*., 5(9), 59–65.
171. Wilby, M. (2006), Novel valve designs to eliminate loss of prime, in Dalby, X. R. N. et al., Eds., Proceedings of Respiratory Drug Delviery, Vol. 2, Davis Healthcare, Boca Raton, FL, pp. 373–376.
172. Jinks, P., and Hunt, K. (2006), Improving suspension MDI dose consistency in patient use by incorporation of a novel semi-permemable system component, in Van Arnum, P., Ed., Proceedings of Drug Delivery to the Lungs XVII, The Aerosol Society, Portshead, United Kingdom, pp. 172–175.
173. Fairweather, W. R., Lin, T. Y. D., and Roswitha, K. (1995), Regulatory, design, and analysis aspects of complex stability studies, *J. Pharm. Sci*., 84(11), 1322–1326.
174. Dyrstad, K., Veggeland, J., and Thomassen, C. (1999), A multivariate method to predict the water vapour diffusion rate through polypropylene packaging, *Int. J. Pharm*., 188, 105–109.
175. Dyrstad, K., Thomassen, C., and Eivindvik, K. (1999), An opportunistic stability strategy: simulation with real data, *Int. J. Pharm*., 188, 97–104.
176. Ruberg, S. J., and Hsu, J. C. (1991), Multiple comparison procedures for pooling batches in stability studies, *Technometrics*, 34, 465–472.
177. Su, X. Y., Po, A. L. W., and Yoshioka, S. (1994), A bayesian approach to arrhenius prediction of shelf life, *Pharm. Res*., 11, 1462–1466.
178. Yoshika, S., Aso, T., and Kojima, K. (1996), Statistical evaluation of shelf-life of pharmaceutical products estimated matrixing, *Drug Stab*. 1, 147–151.
179. Kontny, M. J., Koppenol, S., and Graham, E. T. (1992), Use of the sorption-desorption moisture transfer model to assess the utility of a desiccant in a solid product, *Int. J. Pharm*., 84, 261–271.
180. Badway, S. I. F., Gawronski, A. J., and Alvarez, F. J. (2001), Application of sorption-desorption moisture transfer modeling to the study of chemical stability of a moisture sensitive drug product in different packaging configurations, *Int. J. Pharm*., 223(1–2), 1–13.
181. Kontny, M. J., and Mulski, C. A. (1998), Gelatin capsule brittleness as a function of relative humidity at room temperature, *Int. J. Pharm*., 54(1), 79–85.
182. Illum, L., Bundgaard, H., and Davis, S. S. (1983), A constant partition model for examining the sorption of drugs by plastic infusion bags, *Int. J. Pharm*., 17, 183–192.

183. Nakabayashi, K., Shimamoto, T., and Mima, H. (1980), Stability of packaged solid dosage forms. I. shelf-life prediction for packaged tablets liable to moisutre damage, *Pharm. Chem. Bull.*, 28, 1090–1098.
184. Nakabayashi, K., Shimamoto, T., and Mima, H. (1980), Stability of packaged solid dosage forms. II. Shelf-life prediction for packaged sugar-coated tablets liable to moisture and heat damage, *Pharm. Chem. Bull.*, 28, 1099–1106.
185. Jenke, D. R. (1993), Modeling of solute sorption by polyvinyl chloride plastic infusion bags, *J. Pharm. Sci.*, 82(11), 1134–1139.
186. Roberts, M. S., Kowaluk, E. A., and Polack, A. E. (1991), Prediction of solute sorption by polyvinyl chloride plastic infusion bags, *J. Pharm. Sci.*, 80(5), 449–455.

# PHARMACEUTICAL PRODUCT STABILITY

Andrew A. Webster
*McWhorter School of Pharmacy, Birmingham, Alabama*

**Contents**

1 Introduction
2 Kinetic Equations and Half-Life
  2.1 Rate Expression
  2.2 Order Determination
  2.3 Predicting Shelf Life
  2.4 Arrhenius Equation and Accelerated Stability Testing
3 Degradation Pathways of Pharmaceuticals
4 Chemical Degradation
  4.1 Solvolysis
  4.2 Oxidation
  4.3 Photolysis
  4.4 Dehydration
  4.5 Racemization
5 Physical Degradation
  5.1 Polymorphism
  5.2 Vaporization
6 Microbial Degradation
7 Stability Guidelines and Regulations
8 ICH Quality Guidelines
  References

## 1 INTRODUCTION

Pharmaceutical product stability may be defined as the capability of a particular formulation to remain within its physical, chemical, microbiological, therapeutic,

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

and toxicological specifications while in a specific container closure system. A collection of valid data on the drug in its specific container leads to assurances that the product will be stable for the assigned shelf life.

Often, stability of a drug has been referred to as the time from the date of manufacture and packaging until its chemical or biological activity is not less than a predetermined level of labeled potency without the physical characteristics changing appreciably. For most drugs, 90% of labeled potency is generally recognized as the minimum acceptable potency.

There are many factors that can affect the stability of a pharmaceutical product. These include the stability of the active drug(s), interactions between active and inactive ingredients, the dosage form, manufacturing process, the container system, and environment for shipping, handling, and storage.

The U.S. Pharmacopeia 29/National Formulary 24 (USP29/NF24) defines stability as the extent to which a product retains within specified limits and throughout its period of storage and use (i.e., shelf life) the same properties and characteristics that it possessed at the time of its manufacture [1]. The USP29/NF24 further identifies five generally recognized types of stability:

- Chemical degradation—Each active ingredient retains its chemical integrity and labeled potency within specified limits.
- Physical—The original physical properties, including appearance, palatability, uniformity, dissolution, and suspendability, may be affected.
- Microbiological—Sterility or resistance to microbial growth is retained according to the specified requirements. Antimicrobial agents that are present retain effectiveness within the specified limits.
- Therapeutic—The therapeutic effect remains unchanged.
- Toxicological—No significant increase in toxicity occurs.

Credible pharmaceutical product expiration dates are obtained by rigorous, scientifically designed studies using reliable, meaningful, and specific stability-indicating assays, appropriate statistical concepts, and computers to analyze the resulting data [2]. A comprehensive review of all aspects of pharmaceutical product stability has been published by Lintner [3] and more recently by Connors et al. [4].

## 2 KINETIC EQUATIONS AND SHELF LIFE

Consider the reaction

$$a\mathrm{A} + b\mathrm{B} \rightarrow p\mathrm{P} + \mathrm{q}Q \tag{1}$$

where A and B are the reactants, P and Q are the products, and *a, b, p, q* are the stoichiometric coefficients describing the reaction. The rate of change of the concentration *C* of any of the species can be expressed by

$$-\frac{dC_{\mathrm{A}}}{dt},\quad -\frac{dC_{\mathrm{B}}}{dt},\quad \frac{dC_{\mathrm{P}}}{dt},\quad \frac{dC_{\mathrm{Q}}}{dt}$$

The reactants decrease in concentration relative to time—hence the negative sign. On the other hand, the products increase over time and are preceded by a positive sign. The rates of disappearance of A and B and the rates of appearance of P and Q are interrelated by equations that take into account the stoichiometry of the reaction:

$$-\frac{1}{a}\frac{dC_{\mathrm{A}}}{dt} = -\frac{1}{b}\frac{dC_{\mathrm{B}}}{dt} = \frac{1}{p}\frac{dC_{\mathrm{P}}}{dt} = \frac{1}{q}\frac{dC_{\mathrm{Q}}}{dt} \tag{2}$$

## 2.1 Rate Expression

The rate expression is a mathematical description of the rate of the reaction at any time $t$ in terms of the concentrations(s) of the molecular species present at that time. By simplifying Equation (1) to

$$\mathrm{A} + \mathrm{B} \rightarrow \text{products} \tag{3}$$

the rate expression can be written as

$$-\frac{dC_{\mathrm{A}}}{dt} = -\frac{dC_{\mathrm{B}}}{dt} \propto C^{a}_{\mathrm{A}(t)} C^{b}_{\mathrm{B}(t)} \tag{4}$$

Equation (4) in essence states that the rate of change of the concentration of A at time $t$ is equal to that of B and that each of these changes at time $t$ is proportional to the product of the concentrations of the reactants raised to the respective powers. Note that $C_{\mathrm{A}(t)}$ and $C_{\mathrm{B}(t)}$ are time-dependent variables. As the reaction proceeds, both $C_{\mathrm{A}(t)}$ and $C_{\mathrm{B}(t)}$ will decrease. For simplicity, these concentrations can be denoted by $C_{\mathrm{A}}$ and $C_{\mathrm{B}}$, respectively:

$$-\frac{dC_{\mathrm{A}}}{dt} = -\frac{dC_{\mathrm{B}}}{dt} = kC^{a}_{\mathrm{A}} C^{b}_{\mathrm{B}} \tag{5}$$

where $k$ is a proportionality constant, commonly referred to as the reaction rate constant or the specific rate constant. The format for rate expressions generally involves concentration terms of only the reactants and very rarely those of the products. The latter occurs only when the products participate in the reaction once it has been initiated.

The order of the reaction, $n$, can be defined as $n = a + b$. Extended to the general case, the order of a reaction is the numerical sum of the exponents of the concentration terms in the rate expression. Thus, if $a = b = 1$, the reaction described above is said to be second order overall but first order relative to A and first order relative to B. In principle, the numerical value of $a$ or $b$ can be integral or fractional.

At times the rate of reaction is apparently independent of the concentration of one of the reactants, even though this reactant is consumed during the reaction. For example, in the reaction between an ester and water (hydrolysis) in a predominantly aqueous environment, the theoretical rate expression for the ester can be written in terms of the concentrations of the ester ($C_{\mathrm{E}}$) and water ($C_{\mathrm{W}}$):

$$-\frac{dC_E}{dt} = kC_E C_W \tag{6}$$

If the initial concentration of the ester is less than or equal to $0.5\,M$, complete hydrolysis of the ester will bring about a corresponding decrease in the concentration of water of $0.5\,M$ or less. With the initial concentration of water being about $55\,M$ for an aqueous solution, the relative loss of water through a reaction is insignificantly small, enabling $C_W$ to be considered a constant throughout the entire course of the reaction. Therefore,

$$-\frac{dC_E}{dt} = k\pi C_E \tag{7}$$

where $k_\pi = kC_w$. The reaction appears to be first order relative to the ester and zero order relative to water. The overall reaction is known as a pseudo-first-order reaction with $k_\pi$ the pseudo-first-order constant.

Pseudo-first-order kinetics are observed whenever the concentration of one of the reactants is maintained constant, either by a substantial excess in initial concentration or by rapid replenishment of one of the reactants. If one of the reactants is the hydrogen ion or the hydroxide ion, its concentration, though probably small when compared with that of the drug, can be kept constant throughout the reaction by using buffers in the solution. The concentration of an unstable drug in solution can be maintained invariant by utilization of a suspension, which provides excess solid in equilibrium with the drug in solution.

### 2.2 Order Determination

Reaction orders can be determined by several methods:

*Substitution Method* The accumulated data from a kinetic study can be substituted in the integrated form of the various equations that describe reaction orders. When the $k$ values of one of the iterations remain constant, the reaction is considered to be of that order.

*Graphic Method* A plot of the data can be used to ascertain the order. If a plot of concentration versus time yields a straight line, the reaction is zero order. A straight line from the plot of $\log(a - x)$ versus time is first order and second order if the plot of $1/(a - x)^2$ versus time is a straight line (where the initial concentrations are equal).

*Half-Life Method* For a zero-order reaction the half-life ($t_{1/2}$) is proportional to the initial concentration. The half-life for a first-order reaction is independent of the initial concentration while a second-order reaction is proportional to 1/initial concentration.

### 2.3 Predicting Shelf Life

The shelf life of a formulation is currently based on rigorous physical–chemical laws and statistical concepts to obtain reliable estimates. McMinn and Lintner have

developed an information processing system for handling product stability data [5]. This system saves the time of formulators in analyzing and interpreting their product stability data. For products such as those of vitamins, for example, where large overages are required, the statistical portions of this advanced technique aid the manufacturer to tailor the formula composition to obtain the desired and most economical expiration dating.

This system stores both physical and chemical data and retrieves the information in three different formats (one of which was designed specifically for submitting to regulatory agencies). It analyzes single-temperature data through analysis of covariance and regression. Multiple-temperature data are analyzed either weighted or unweighted using the Arrhenius relationship. This method provides estimates of the shelf life of the preparation with appropriate confidence intervals, preprints the assay request cards that are used to record the results of the respective assay procedures and to enter the data into the system, and produces a 5-year master stability schedule as well as periodic 14-day schedules of upcoming assays.

Analysis of stability data obtained at a single temperature is based on a linear (zero-order) model

$$Y_{mn} = \beta_m X_{mn} + \alpha_m + \varepsilon_{mn} \tag{8}$$

where $Y_{mn}$ is the percentage of label of the $n$th stability assay of the $m$th lot, $X_{mn}$ is the time in months at which $Y_{mn}$ was observed, $\beta_m$ and $\alpha_m$ are the slope and intercept, respectively, of the regression line of the $m$th lot, and $\varepsilon_{mn}$ is a random error associated with $Y_{mn}$. The random errors are assumed to be identically and independently distributed normal variables with a zero mean and a common variance $\sigma^2$.

A summary of the regression analysis for each individual lot and for the combination of these lots plus a summary of the analyses of covariance and deviation from regression is prepared by the computer.

Since the stability data from individual lots are pooled, these data are examined for validity by the $F$ test. The mean square of the regression coefficient (slope) is divided by the mean square of the deviation within lots, and similarly, the adjusted mean ($y$ intercept) is divided by the common mean square to give the respective $F$ ratios. The latter values then are compared with the critical 5% $F$ values. When the calculated $F$ values are smaller than the critical $F$ values, the data may be combined and the pooled data analyzed.

### 2.4 Arrhenius Equation and Accelerated Stability Testing

The purpose of stability testing is to assess the effects of temperature, humidity, light, and other environmental factors on the quality of a drug substance or product. These data sets are used to establish storage conditions, retest periods, and shelf loss and to justify overages included in products for stability reasons. The most useful equation relating temperature and reaction rate is the Arrhenius equation

$$\frac{d \ln k}{dT} = \frac{E_a}{RT^2} \tag{9}$$

This can be rewritten as

$$k = Ae^{-E_a/(RT)} \tag{10}$$

$$\ln\left(\frac{k_1}{k_2}\right) = \frac{E_a}{R}\left(\frac{1}{T_2} - \frac{1}{T_1}\right) \tag{11}$$

where $E_a$ is a constant and the subscripts 1 and 2 denote the two different temperature conditions. A plot of $\ln k$ as a function of $1/T$, referred to as the Arrhenius plot, is linear according to Equation (10) if $E_a$ is independent of temperature. This observation makes it possible to conduct kinetic experiments at elevated temperatures and obtain estimates of rate constants at lower temperatures by extrapolation of the Arrhenius plot. This procedure, commonly known as accelerated stability testing, is most useful when the reaction at ambient temperature is too slow to be monitored conveniently and when $E_a$ is relatively high. As an example, in a reaction with an $E_s$ of 25 kcal/mol with an increase from 25 to 45°C, there is about a 14-fold increase in the reaction rate constant. In contrast, a rate increase of just threefold is obtained for the same elevation in temperature when $E_a$ is 10 kcal/mol. The slope of the Arrhenius plot for a reaction yields the magnitude of $E_a$. Hydrolysis reactions typically have an $E_a$ of 10–30 kcal/mol while oxidation and photolysis reactions have smaller energies of activation [6].

The elevated temperatures most commonly used are 40, 50, and 60°C in conjunction with ambient humidity. Occasionally, higher temperatures are used. The samples stored at the highest temperatures are examined weekly for physical and chemical changes. If a substantial change is seen, samples stored at lower temperatures are examined. If there is no change after 30 days at 60°C, the stability prognosis is excellent. Corroborative evidence must be obtained by monitoring the samples stored at lower temperatures for longer durations.

An underlying assumption of the Arrhenius equation is that the reaction mechanism does not change as a function of temperature. Since accelerated stability testing of pharmaceutical products normally employs a narrow range of temperature, it is often difficult to detect nonlinearity in the Arrhenius plot from experimental data, even though such nonlinearity is expected from the reaction mechanism [7].

Non-Arrhenius behavior has been observed in pharmaceutical systems [8]. This may be attributed to the possible evaporation of solvent, multiple reaction pathways, and the change in physical form of the formulation when the temperature of the reaction is changed [9]. Nonlinearity in Arrhenius plots frequently is observed in following the temperature dependence of protein degradation. Degradation mechanisms in proteins often change with temperature [10]. The aggregation of interleukin 1β (IL-1β) in aqueous solution follows apparent first-order behavior at 60°C to 30% of drug remaining. At or below 55°C the aggregation deviates from apparent first order and becomes biphasic [11].

Considerable interest has been generated in the use of accelerated stability testing based on a singe condition of elevated temperature and humidity. For abbreviated new drug applications (ANDAs) U.S. Food and Drug Administration (FDA) stability guidelines suggest that a tentative expiration date of 24 months may be granted for a drug product if satisfactory stability results can be documented under a stressed condition of 40°C and 75% relative humidity [12]. The simplicity of such

a guideline is attractive because a substantial saving in time can be obtained in advancing a drug product to the marketplace [7].

## 3 DEGRADATION PATHWAYS OF PHARMACEUTICALS

Many degradation pathways are similar to reactions described for organic compounds in organic chemistry texts. The decomposition of a drug is likely to be mediated by reaction with water, oxygen, or light. The routes most pharmaceuticals degrade include hydrolysis, oxidation, photolysis, and racemization. Chemical degradation occurs when a new chemical entity is formed as a process of the degradation process. Physical degradation occurs when drug loss does not produce distinctly different chemical products [13]. Degradation reactions of a variety of mechanisms can often be avoided by decreasing storage temperature and/or buffering pH to a stability optimum.

## 4 CHEMICAL DEGREDATION

### 4.1 Solvolysis

Solvolysis is the decomposition of the active drug with the solvent that is present. When water is the solvent, the process is called hydrolysis. The most common solvolysis reactions involve liable carbonyl compounds such as esters and β-lactams (see Table 1). Aspirin is hydrolyzed to acetic acid and salicylic acid in the presence of moisture, but in a dry environment this hydrolysis is negligible. The most significant catalysts of hydrolysis are adverse pH and specific compounds (e.g., dextrose and copper or other metal ions).The rate of hydrolysis is pH and temperature dependent. There is a generalized rule of thumb that with each 10°C increase in temperature, the rate of the reaction increases exponentially. The actual factor of rate increase depends on the activation energy of the particular reaction, with the activation energy being a function of the specific reactive bond and the drug formulation.

Hydrolysis leads to a decrease in active drug and an increase of decomposition products. The effect of this change on reaction rate depends on the order of the

**TABLE 1 Liable Functional Groups to Hydrolysis**

| Functional Group | Examples |
|---|---|
| Esters | Aspirin, procaine, alkaloids, estrone sulfate, dexmethasone sodium phosphate |
| Lactams | Penicillins, cephalosporins |
| Amides | Chloramphenicol, thiacinamide |
| Lactones | Pilocarpine |
| Oximes | Steroid oximes |
| Imides | Gluthethimide |
| Malonic ureas | Barbiturates |
| Nitrogen mustards | Melphalan |
| Azo methines | Benzodiazepines |

reaction. For zero-order reactions, the decomposition rate is independent of the drug concentration. Zero-order hydrolysis shows a lesser percentage of decomposition for higher concentration solutions that those of lower concentration. Unlike zero-order reactions, hydrolysis following first-order kinetics shows a rate of change which is directly proportional to the concentration of the reacting ingredient. Therefore, changes in active ingredient concentration do not have an influence on the percentage concentration.

Structural modifications of the drug may be employed to retard hydrolysis. The most frequently encountered hydrolysis reaction involves the ester. Substituents can have an effect on these reaction rates. Hansch and Taft [14] and Hammett [15] provide excellent reviews of the topic. Additionally, a compound may be stabilized by reducing its solubility. This can be accomplished by the addition of lipophilic substituents to side chains or aromatic rings. Often less soluble salts or esters have been employed to aid in product stability.

The rate of hydrolysis is impacted by the ionic strength of the total concentration of dissolved electrolytes. In general, the rate constant of hydrolysis is directly proportional to the ionic strength of ions of like charge (drug cation and excipient cation) and inversely proportional to the ionic strength with oppositely charged ions.

### 4.2 Oxidation

Oxidation is a primary cause of instability and usually involves the addition of oxygen or the removal of hydrogen. Auto-oxidations are oxidative degradations usually mediated through reactions with atmospheric oxygen. Most auto-oxidations are free-radical reactions. A drug's susceptibility to auto-oxidation can be determined by investigating its stability in a high-oxygen-tension atmosphere, usually 40% oxygen. Results are compared against those under inert or ambient temperature [16]. A listing of several functional groups that are subject to oxidation can be found in Table 2.

Oxidations are catalyzed by acids, bases, pH values that are higher than the optimum, polyvalent metal ions, peroxides, hydroperoxides, and exposure to oxygen and ultraviolet (UV) illumination. These reactions may necessitate the use of antioxidant chemicals, inert atmospheres, and opaque packaging. Chelating agents

**TABLE 2 Functional Groups Subject to Oxidation**

| Functional Group | Examples |
|---|---|
| Phenols | Catecholamines, morphine |
| Conjugated dienes | Vitamin A |
| Thioethers | Chlorpromazine |
| Nitrites | Amyl nitrite |
| Aldehydes | Paraldehyde, flavors |
| Amines | Clozapine |
| Carboxylic acids | Fatty acids |
| Thiols | Dimerceprol |
| Ethers | Diethylether |

**TABLE 3 Common Antioxidants**

| Aqueous Systems | Oil-Based Systems | Chelating Agents |
|---|---|---|
| Sodium sulfite | Ascorbyl palmitate | Ethylenediamine tetraaceticacid (EDTA) |
| Sodium metabisulfite | Hydroquinone | Dihydroethylglycine |
| Sodium bisulfite | Propyl gallate | Citric acid |
| Sodium thiosulfate | Nordihydroguaiaretic acid | Tartaric acid |
| Ascorbic acid | Butylated hydroxytoluene | Gluconic acid |
| | Butylated hydroxyanisole | Saccharic acids |
| | α-Tocopherol | |

added to water sequester heavy metals. Parenteral formulations should not come into contact with heavy metal ions during their manufacture, packaging, or storage [17].

Antioxidants are very effective in stabilizing products undergoing a free-radical mediated chain reaction. These products possess lower oxidation potentials than the active drug. Ideally, antioxidants are stable over a wide pH range and remain soluble in the oxidized form, colorless, and nontoxic. A listing of commonly used antioxidants can be found in Table 3.

Visual identification of oxidation products is frequently attainable due to the introduction of conjugation, but these changes may not be visible in certain concentrations.

### 4.3 Photolysis

When molecules absorb energy, they proceed to a higher energy state where they release the energy in a chemical reaction to attain their ground energy state. When the energy of activation that is absorbed by the compound comes from a light source, the decomposition reaction is called photolytic. The activated species can release the energy either as light of a different frequency (fluorescence) or by decomposition (photolysis). Exposure to room light or sunlight can lead to drug degradation by causing photo-oxidation and photolysis of covalent bonds. In susceptible compounds, usually those with π electrons, photochemical energy creates free-radical intermediates, which can perpetuate chain reactions.

In general, drugs that absorb UV light below 280 nm undergo decomposition in sunlight while compounds that absorb above 400 nm have the potential to degrade in both sunlight and room light. Photo-induced reactions are common in steroids [18]. The photodegradation of sodium nitroprusside in aqueous solution remains a classic example of photolytic decomposition [19].

### 4.4 Dehydration

Dehydrations are chemical reactions that involve the loss of water. The acid-catalyzed dehydration of tetracycline yields the toxic epianhydrotetracycline [20]. The physical dehydration of theophylline hydrate and ampicillin trihydrate leads to a change of the crystalline structure of the drug [21].

### 4.5 Racemization

Racemization is the process of changing from an optically active compound into a racemic mixture. Pfeiffer provided one of the earliest discussions on the importance of stereospecificity in drug action [22].

The most widely known drugs that undergo racemization are tetracycline, epinephrine [23], pilocarpine [24], and ergotamine. In tetracycline, the reaction occurs rapidly when the dissolved drug is exposed to a pH greater than 3, resulting in a steric rearrangement of the dimethylamino group [25].

Generally, racemization follows a first-order reaction rate and is dependent on temperature, solvent system, catalysts, and the presence or lack of light. Resonance stabilization through substituents adjacent to the asymmetric center tends to accelerate racemization.

## 5 PHYSICAL DEGRADATION

### 5.1 Polymorphism

Many pharmaceutical solids can exist in different physical forms. Polymorphism is often characterized as the ability of a drug substance to exist as two or more crystalline phases that have different arrangements and/or conformations of the molecules in the crystal lattice [26]. Polymorphic forms of a drug substance can have different chemical and physical properties, including melting point, chemical reactivity, apparent solubility, dissolution rate, optical and mechanical properties, vapor pressure, and density. These properties can have a direct effect on the ability to process and/or manufacture the drug substance and the drug product as well as on drug product stability, dissolution, and bioavailability. Thus, polymorphism can affect the quality, safety, and efficacy of the drug product [27]. Amorphous solids consist of disordered arrangements of molecules and do not possess a distinguishable crystal lattice. Solvates are crystalline solid adducts containing either stoichiometric or nonstoichiometric amounts of a solvent incorporated within the crystal structure. If the incorporated solvent is water, the solvates are also commonly known as hydrates. Drug exposure to changes in temperature, pressure, humidity, and pulverization during granulation, milling, and compression may lead to polymorphic phase conversion. The extent of conversion generally depends on the relative stability of the polymorphs, kinetic barriers for phase conversion, and applied stress [28]. Sulfonamides, barbiturates, and steroids are known for their propensity to form polymorphs [29].

### 5.2 Vaporization

Several drugs and adjuvants possess high vapor pressures at room temperature that their vaporization exists as a route of drug loss. Low-molecular-weight alcohols and "aromatics" used for flavoring and aroma may be lost through vaporization. Nitroglycerine is the most frequently cited example of loss on vaporization. The FDA issued a special regulation governing the types of containers that may be used for

dispensing nitroglycerine tablets [30]. The addition of macromolecules such as microcrystalline cellulose allows for the preparation of a stabilized nitroglycerine sublingual tablet [31].

## 6 MICROBIAL DEGRADATION

Microorganisms could present a risk of infection or degradation of medicinal products. These organisms may proliferate during normal storage conditions or during patient use, especially in multidose preparations. Preparations containing water bear the greatest risk of contamination. These products include solutions, suspensions, and emulsions as well as sterile multidose injections and ophthalmic preparations.

Antimicrobial preservatives are used to prevent or inhibit the growth of microorganisms. Factors influencing the level of observable efficacy include the chemical structure of the preservative, the physical and chemical characteristics of the pharmaceutical product, the concentration of the preservative, and the type and load of initial contamination. At no time should preservatives be used as an alternative to good manufacturing practice.

## 7 STABILITY GUIDELINES AND REGULATIONS

FDA guidelines for stability testing are outlined in the *Code of Federal Regulations* (CFR), Part 21, Section 211.166, under Current Good Manufacturing Practice for Finished Pharmaceuticals. The guidelines state that there must be a testing program designed to assess the stability characteristics of drug products. The results of such stability testing are to be used in determining appropriate storage conditions and expiration dates. This section of the *Federal Register* provides guidelines for sample size and test intervals, storage conditions for samples, and specific test methods [32].

Section 211.166 guidelines also require an adequate number of batches of each drug product tested for expiration date assignment. Accelerated studies, combined with basic stability information on the components, drug products, and container closure system may be used to support tentative expiration dates provided full shelf life studies are not available and are being conducted. Where data from accelerated studies are used to project a tentative expiration date that is beyond a date supported by actual shelf life studies, there must be stability studies conducted, including drug product testing at appropriate intervals, until the tentative expiration date is verified or the appropriate expiration date determined.

Additional guidelines outline the need for reserve samples [33], expiration dating [34], and laboratory recordkeeping [35].

## 8 ICH QUALITY GUIDELINES

The International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH) is a project that brings together the regulatory authorities of Europe, Japan, and the United States and

**TABLE 4 ICH Document Codes and Guidelines for Stability**

| ICH Code | Title | Description |
|---|---|---|
| Q1A(R2) | Stability Testing of New Drug Substances and Products | Stability testing protocols including temperature, humidity, and trial duration |
| Q1B | Stability Testing: Photostability Testing of New Drug Substances and Products | Basic testing protocol required to evaluate light sensitivity and stability of new drugs and products |
| Q1C | Stability Testing for New Dosage Forms | Extends main stability guideline for new formulations of already approved medicines and defines circumstances under which reduced stability data can be accepted |
| Q1D | Bracketing and Matrixing Designs for Stability Testing of New Drug Substances and Products | General principles for reduced stability testing and provides examples of bracketing and matrixing designs |
| Q1E | Evaluation of Stability Data | Explains possible situations where extrapolation of retest periods/shelf lives beyond real-time data may be appropriate; provides examples of statistical approaches to stability data analysis |
| Q1F | Stability Data Package for Registration Applications in Climatic Zones III and IV | Besides proposing acceptable storage conditions for long-term and accelerated studies, gives guidance on data to cover situations of elevated temperature and/or extremes of humidity; referenced literature provides information on classification of countries according to climatic zones |

experts from the pharmaceutical industry in the three regions to discuss scientific and technical aspects of product registration. The purpose is to make recommendations on ways to achieve agreement in the interpretation and application of technical guidelines and requirements for product registration to reduce the need to duplicate the testing carried out during the research and development of new medicines. The objective of such harmonization is a more economical use of human, animal, and material resources and the elimination of unnecessary delay in the global development and availability of new medicines while maintaining safeguards on quality, safety and efficacy, and regulatory obligations to protect public health [36].

The guidelines provide a breath of recommendations from stability testing protocols, including temperature, humidity, and trial duration [Q1A(R2)]; basic testing protocols required to evaluate the light sensitivity and stability of new drugs and products (Q1B); and stability testing for new formulations of already approved medicines (Q1C). A description of the various ICH guidelines for stability can be found in Table 4.

The ICH also provides guidelines in analytical validation [Q2(R1), impurities (Q3 series), pharmacopeias (Q4 series), quality of biotechnological products (Q5

series), specifications (Q6 series), good manufacturing practice (Q7), pharmaceutical development (Q8), and risk management (Q9)].

## REFERENCES


1. U.S. Pharmacopeia (USP), Stability considerations in dispensing practice, USP 29/NF 24, USP, Rockville, MD, p. 3029.
2. Guillory, P., and Poust, R. (2002), Chemical kinetics and drug stability, in Banker and Rhodes, Eds., *Drugs in the Pharmaceutical Sciences*, Vol. 121, Marcel Dekker, New York, pp. 139–166.
3. Lintner, C. C. (1973), *Quality Control in the Pharmaceutical Industry*, Vol. 2, Academic, New York, p. 141.
4. Connors, K. A., Amidon, G. L., and Stella, J. V. (1986), *Chemical Stability of Pharmaceuticals*, 2nd ed., Wiley, New York.
5. McMinn, C. S., and Lintner, C. J. (1973, May), paper presented at the American Pharmaceutical Association Academy of Pharmaceutical Sciences Meeting, Pharmaceutical Technical Section, Chicago, IL.
6. Lachman, L., DeLuca, P., and Akers, J. J. (1986), Kinetic principles and stability testing, in Lachman, L., Lieberman, H. A., and Kanig, J. L., Eds., *The Theory and Practice of Industrial Pharmacy*, 3rd ed., Lea & Febiger, Philadelphia, p. 766.
7. Fung, H.-L., and King, S.-Y. (1983), in Pharmaceutical Technology Conference '83 Proceedings, Aster Publishing, Springfield, OR.
8. Pikal, M. J., Lukes, A. L., and Lang, J. E. (1977), *J. Pharm. Sci.*, 66, 1312.
9. Woolfe, J., and Worthington, H. E. C. (1974), *Drug Dev. Commun.*, 1, 185.
10. Wang, W. (1999), Instability, stabilization, and dormulation of liquid protein pharmaceuticals, *International Journal of Pharmaceutics*, 185, 129–188.
11. Gu, L.C., Erdos, E. A., Chang, H.-S. (1991), *Pharm. Res*. 8, 485.
12. (a) Center for Drugs and Biologics, U.S. Food and Drug Administration (FDA) (1987, Feb.), Guideline for submitting documentation for the stability of human drugs and Biologics, FDA, Rockville, MD. (b) Center for Drug Evaluation and Research and Center for Biologics Evaluation and Research, U.S. Department of Health and Human Services, FDA (1998, June), Guidance for industry: Stability testing of drug substances and drug products: Draft guidance, FDA, Rockville, MD.
13. U.S. Pharmacopeia (USP), Pharmaceutical stability, USP 29/ NF 24, USP, Rockville, MD. p. 2994.
14. Hansch, A. L., and Taft, R. E. (1991), A survey of Hammett substituent constants and resonance and field parameters, *Chem. Rev.*, 91, 165.
15. Hammett, L. P. (1970), *Physical Organic Chemistry*, 2nd ed., McGraw-Hill, New York.
16. Johnson, D. M., and Gu, L. C. (1998), Autooxidation and antioxidants, in Swarbrick, J., and Boylan, J. C. Eds., *Encyclopedia of Pharmaceutical Technology*, Marcel Dekker, New York, pp. 415–449.
17. Vadas, E. B. (2000), Stability of pharmaceutical products, in *Remington: The Science and Practice of Pharmacy*, 20th ed, Lippincott Williams and Wilkins, Baltimore, MD, p. 989.
18. Ogata, M., Noro, Y., Yamada, M., Tahara, T., and Nishimura, T. (2000), Photo-degradation products of methylprednisolone sulphanate in aqueous solution—Evidence of a bicyclo[3.1.0]hex-3-en-2-one intermediate, *J. Pharm. Sci.*, 87(1), 91–95.
19. Hauser, U., Oestreich, V., and Rohrweck, H. D. (1977), On optical dispersion in transparent molecular systems, *Zeits. Phys. A*, 280, pp. 17–25.

20. Yuen, P. H., and Sokoloski, T. D., (1977), Kinetics of concomitant degradation of tetracycline to epitetracycline, anhydrotetracycline, and epianhydrotetracycline in acid phosphate solution, *J. Pharm. Sci.*, 66(11), 1648–1650.
21. Shefter, E., Fung, H.-L., and Mok, O. (1973), Dehydration of crystalline theophylline monohydrate and ampicillin trihydrate, *J. Pharm. Sci.*, 62, 791.
22. Pfeiffer, C. C. (1956), Optical isomerism and pharmacological action, a generalization, *Sci.*, 123, 3210, 29–31.
23. Schroeter, L. C., and Higuchi, T. (1958), Racemization of epinephrine, *J. Am. Pharm. Assoc. (Baltimore)*, 47, 6, 426, 30.
24. Nunes, M. A., and Brochmann- Hanssen, E. (1974), Hydrolysis and epimerization kinetics of pilocarpine in aqueous solution, *J. Pharm. Sci.*, 63, 716.
25. Sheberstova, N. V., Perel'son, M. E., and Kuzovkov, A. D. (1975), Study of the epimerization of tetracycline by the NMR method, *Chem. Nat. Comp.*, 10(1), 61–65.
26. Haleblian, J., and McCrone, W. (1969), Pharmaceutical applications of polymorphism, *J. Pharm. Sci.*, 58, 911.
27. U.S. Department of Health and Human Services Food and Drug Administration Center for Drug Evaluation and Research (CDER), U.S. Food and Drug Administration (FDA) (2004, Dec.), Guidance for industry ANDAs: Pharmaceutical solid polymorphism chemistry, manufacturing, and controls information, FDA, Rockville, MD.
28. Vippagunta, S. R., Brittain, H. G., and Grant, D. J. W. (2001), Crystalline solids, *Adv. Drug Deliv. Rev.*, 48, 3–26.
29. Kuhnert-Brandstatter, M. (1971), Thermomicroscopy, in *The Analysis of Pharmaceuticals*, Pergamon, Oxford, pp. 37–42.
30. *Federal Register*, 37, 15959 (1972).
31. Fung, H.-L., Yap, S. K., and Rhodes, C. T. (1974), Development of a stable sublingual nitroglycerin tablet. I. Interaction of nitroglycerin with selected macro-molecules, *J. Pharm. Sci.*, 63, 1810.
32. *Code of Federal Regulations*, Title 21, Food and drugs, Part 211, Current good manufacturing practice for finished pharmaceuticals, Subpart I, Laboratory controls, Section 211.166, Stability testing.
33. *Code of Federal Regulations*, Title 21, Food and drugs, Part 211, Current good manufacturing practice for finished pharmaceuticals, Subpart I, Laboratory controls, Section 211.170, Reserve samples.
34. *Code of Federal Regulations*, Title 21, Food and drugs, Part 211, Current good manufacturing practice for finished pharmaceuticals, Subpart G, Packaging and labeling control, Section 211.137, Expiration dating.
35. *Code of Federal Regulations*, Title 21, Food and drugs, Part 211, Current good manufacturing practice for finished pharmaceuticals, Subpart J, Records and reports, Section 211.194, Laboratory records.
36. International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use, available: http://www.ich.org/cache/compo/363-272-1.html.

# ALTERNATIVE ACCELERATED METHODS FOR STUDYING DRUG STABILITY: VARIABLE-PARAMETER KINETICS

Giuseppe Alibrandi
*Università di Messina, Messina, Italy*

**Contents**

1 Introduction
2 Theory
3 Experimental
  3.1 Computer Simulation
  3.2 Devices to Obtain Variable-Parameter Conditions
  3.3 Analytical Instruments
  3.4 Software for Processing Experimental Data
4 Examples of Variable-Parameter Kinetic Experiments
  4.1 Variable-Temperature Kinetic Experiments
  4.2 Variable-Concentration Kinetic Experiments
  4.3 Variable-Ionic-Strength Kinetic Experiments
5 Conclusions
  References

## 1 INTRODUCTION

Studies on drug stability have a central role in physicochemical profiling [1–6]. It is important to know (i) how long the substrate maintains its chemical identity in various environments carrying out its therapeutic action and (ii) the pathway followed when it degrades. Both sets of information are vitally important issues in the pharmaceutical industry. The first is necessary to decide whether to continue to investigate the molecule of interest or not: An unstable drug can reduce its efficiency

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

to a little percentage and can generate products with undesired effects; the second is necessary to understand whether it could be possible to find a remedy or not, for example synthesizing a new drug with a similar structure and reduced reactivity or devising a suitable carrier able to stabilize the molecule and deliver it to the desired target.

Temperature, pH, ionic strength, concentration of a metal ion, and other environmental parameters influence a chemical reaction and varying their values can signify drastic changes, depending on the case, on the rate, mechanism, or direction of the reaction. For this reason quantitative studies on the effects of physical parameters on reactivity often take a very long time in this kind of research.

The usual way consists of isolating and characterizing the reaction of interest and then determining the rate constant under first- or pseudo-first-order conditions for different values of the involved parameters. The result is a polydimensional representation of the specific rate as a function of these parameters that can be of great use for practical applications and, particularly, constitutes the experimental structure on which to base the investigation directed to clarify the reaction mechanism. In fact, the dependence of $k_{obs}$ on parameters is a consequence of their influence on the rate-determining step and fitting kinetic data to suitable mathematic models gives information that enables one to formulate a reaction scheme, discriminate between kinetically equivalent mechanisms, and determine the values of elementary kinetic constants [1–4, 7, 8].

Many examples are present in the scientific literature underlining the effort in producing kinetic data [9–11]. The Edwards historical study that started the investigation on the mechanism of the hydrolysis of aspirin required hundreds of kinetic experiments [12, 13]. Several examples are reported by Carstensen [1] in his review on the subject where, beside the large space dedicated to the determination of the pH–rate profile, the effect of temperature, ionic strength, buffer concentration, and dielectic constant on the stability of drugs was treated.

Physicochemical profiling concerns the determination of some basic molecular properties of pharmaceutical interest that can support the successive clinical studies and can help a rapid identification and elimination of compounds with unsuitable physicochemical and pharmacokinetic properties. New accelerated methods for physicochemical profiling have assumed particular importance in recent years because of the great amount of new chemical entities coming from modern synthetic strategies. Thousands of compounds have to be profiled each year and traditional technologies cause a bottleneck in drug development. Methods have been developed for high-throughput physicochemical profiling for solubility, permeability, $pK_a$, lipophilicity, stability, and integrity, and a great interest from pharmaceutical industries and instrument maker groups is reserved for this subject [5, 6].

Among physicochemical properties, the stability of drug candidates is receiving increasing attention. Unfortunately, its evaluation very often requires a lot of experimental time and profiling of many molecules would be nearly impossible without new, computer-aided methods. Any effort useful to facilitate this first part of pharmaceutical investigation is appreciated because it can be converted to a considerable lowering of the research cost.

Here a panoramic picture of the effort directed to accelerating the characterization of the reactivity in solution of molecular candidates is shown. In particular,

variable-parameter kinetics (VPaK) is considered, a potential new way to save time and chemicals while studying drug stability.

## 2 THEORY

Consider a molecule A degrading in products,

$$\mathrm{A} \rightarrow \mathrm{P} \tag{1}$$

the rate law, operating under first- or pseudo-first-order conditions, is given by Equation (2), where $C$ is the molar concentration of the substrate and $k_{\mathrm{obs}}(\mathrm{Par}_1, \mathrm{Par}_2, \ldots)$ is the observed rate constant function of various parameters:

$$-\frac{dC}{dt} = k_{\mathrm{obs}}(\mathrm{Par}_1, \mathrm{Par}_2, \ldots)C \tag{2}$$

In the usual kinetic experiments all these parameters must be rigorously constant so that the experimental $C$–$t$ data, that is, the kinetic profile obtained following the change in concentration with time, can be easily fitted to the differential form of the first-order equation (2) or to its integrated, exponential [Equation (3)] or logarithmic [Equation (4)], forms to obtain the optimized values of $k_{\mathrm{obs}}$:

$$C = C_0 \exp(-k_{\mathrm{obs}}t) \tag{3}$$

$$\ln C = -k_{\mathrm{obs}}t + \ln C_0 \tag{4}$$

The dependence of $k_{\mathrm{obs}}$ on one parameter (temperature $T$, pressure $P$, pH, ionic strength $I$, etc.), that is, the $k_{\mathrm{obs}}(\mathrm{Par}_i)$ profile, can be obtained by carrying out several experiments for different values of that parameter. This yields a set of $k_{\mathrm{obs}}$–$\mathrm{Par}_i$ data and, having an analytical model describing this dependence, by a second step fitting treatment, the terms regulating such dependence can be obtained. For example, for the parameter temperature ($T$) the dependence function can be the Eyring equation (5) and the terms regulating the dependence, the entropy of activation $\Delta S^{\ddagger}$ and the enthalpy of activation $\Delta \mathrm{H}^{\ddagger}$ ($k$ = Boltzmann's constant, $h$ = Planck's constant, $R$ = gas constant) [1, 7, 8]:

$$k_{\mathrm{obs}} = \frac{kT}{h}\exp\left(\frac{\Delta S^{\ddagger}}{R}\right)\exp\left(-\frac{\Delta H^{\ddagger}}{RT}\right) \tag{5}$$

The Eyring equation is usually used in the logarithmic form (6) to have a linear plot of kinetic data. The intercept gives the $\Delta S^{\ddagger}$ value and the slope gives the $\Delta H^{\ddagger}$ value:

$$\ln\frac{k_{\mathrm{obs}}}{T} = \ln\frac{k}{h} + \frac{\Delta S^{\ddagger}}{R} - \frac{\Delta H^{\ddagger}}{RT} \tag{6}$$

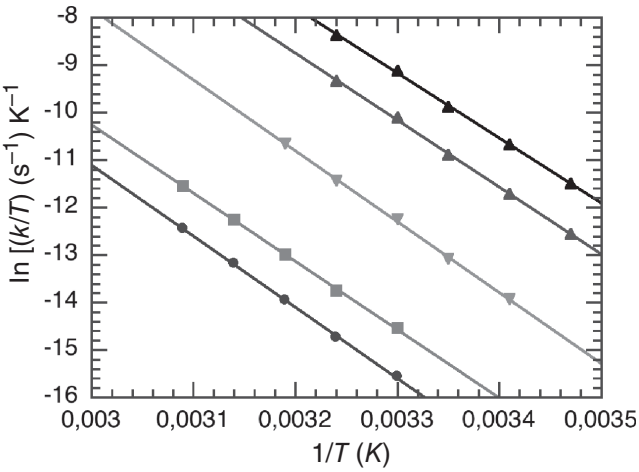


**FIGURE 1** Eyring plots (solid lines) for reactions of homologous series of five molecules obtained with kinetic experiments carried out at constant temperature (plain markers) (simulated data from ref. 14).

Figure 1 shows an example of this kind of study where 25 constant-temperature kinetic (CTK) experiments were necessary to find the activation parameters for the decomposition of five molecules. When the parameter is a concentration of a species present in the reaction environment ($H^+$, $OH^-$, $Me^{n+}$, a nucleophile, a catalyst, etc.), the dependence function can be, for example, of the type shown in the equation

$$k_{obs} = k_{OH}[OH^-] \tag{7}$$

and the fit of $k_{obs}$–$[OH^-]$ data, obtained in several experiments carried out at constant $[OH^-]$, gives the term $k_{OH}$ (Figure 2). These dependence functions, where concentrations are concerned, can be of great complexity but they are very important information for the formulation of the reaction mechanism [1, 3, 7, 8].

When the parameter is the ionic strength the dependence function is given by the Brönsted–Bjerrum equation (8) [1, 7, 8, 15], where $k_0$ is the rate constant at zero ionic strength, $a$ is a constant (for water at 25°C, $a$ = 0.50925 $L^{1/2}/mol^{1/2}$ [16]) and $Z_A Z_B$ is the product of charges of the reacting species taking part in the rate-determining step:

$$k_{obs} = k_0 \times 10^{2aZ_A Z_B \frac{\sqrt{I}}{1+\sqrt{I}}} \tag{8}$$

Even in this case, traditionally, a logarithmic form is used [Equation (9)] to treat kinetic data. The intercept of the linear plot (Figure 3) gives the value of $k_0$ and the slope the value of $Z_A Z_B$, and this is very useful information for ionic reactions to discriminate between kinetically equivalent mechanisms:

$$\log k_{obs} = \log k_0 + 2aZ_A Z_B \frac{\sqrt{I}}{1+\sqrt{I}} \tag{9}$$

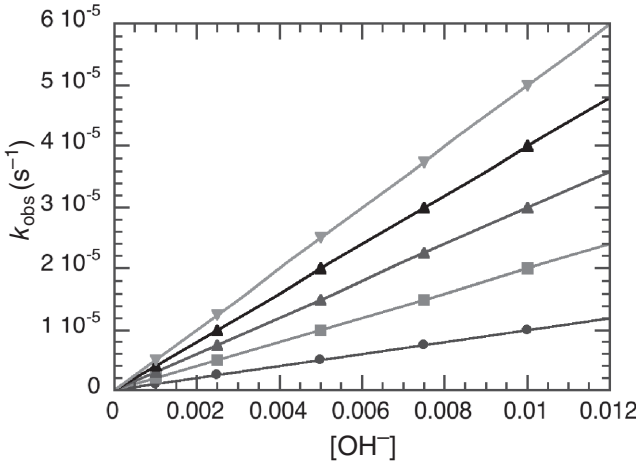


**FIGURE 2** $k_{obs}([OH^-])$ profiles (solid lines) obtained for homologous series of five molecules by 25 kinetic experiments (plain markers) carried out at constant concentration of $OH^-$ (simulated data).

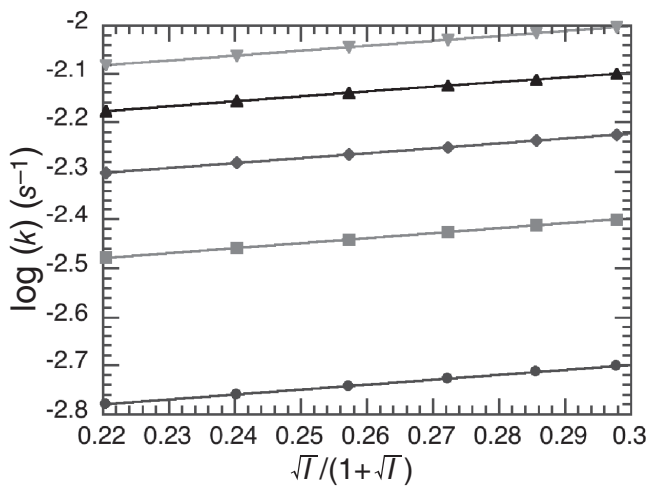


**FIGURE 3** Brönsted–Bjerrum plots (solid lines) obtained by constant-ionic-strength kinetic experiments (plain markers) for homologous series of five compounds having the same mechanism (same values of $Z_A Z_B = +1$) and different intrinsic reactivity (different values of $k_0$) (simulated data).

Constant-parameter kinetic (CPaK) experiments are very time consuming. Nowadays, the use of personal computers and new-generation analytical instruments enables a step forward to be made in collecting kinetic data.

VPaK [17] can be defined as that part of chemical kinetics concerning experiments carried out while varying the value of an environmental parameter in a controlled way. The aim is to obtain, in a single run, the dependence of the observed rate constant on that parameter in a time 10–100 times less compared to that usually

spent using traditional methods. This is possible by fitting the kinetic profile obtained during the reaction to a suitable mathematic model analytically describing it.

VPaK can be considered a generalization of nonisothermal kinetics [18, 19], largely applied in thermal analysis [20–25] for solid-state systems. Several examples of variable-temperature experiments concerning solution chemistry were reported in the past. Different approaches were used depending on the different scientific background of the authors (inorganic, organic, organometallic, and pharmaceutical) and on the instruments available at the time [14, 26–51]. Few examples of variable-concentration kinetic (VCK) experiments were reported [31, 35, 52, 53]. Just one example of variable-ionic-strength kinetic experiment has so far been reported [54]. Here a simple approach is described making a reference to important similar studies.

The mathematical form describing a VPaK experiment is given by the equation,

$$-\frac{dC}{dt} = \{k_{\text{obs}}[\text{Par}_i(t)]\}_{\text{Par}\neq\text{Par}_i} C \tag{10}$$

where $C$ is the concentration of the monitored reacting species. In this equation $k_{\text{obs}}$ is not a constant but depends on a parameter ($\text{Par}_i$) varying with time. Then $k_{\text{obs}}[\text{Par}_i(t)]$ is a function of a function: $k_{\text{obs}}(\text{Par}_i)$ is the dependence function ($D$) describing the dependence of the rate constant on the parameter $i$ and $\text{Par}_i(t)$ is the modulating function ($M$) showing the way the parameter changes with time while the other parameters are maintained constant. The experimental kinetic profile contains in each point information about the value of the rate constant at that time, as can be seen writing Equation (10) in the form of Equation (11). The ratio of the derivative of the concentration on the concentration itself gives $k_{\text{obs}}$. Then, the $k_{\text{obs}}(\text{Par}_i)$ profile can be obtained in a single kinetic run:

$$-\frac{1}{C}\frac{dC}{dt} = \{k_{\text{obs}}[\text{Par}_i(t)]\}_{\text{Par}\neq\text{Par}_i} \tag{11}$$

Furthermore, knowing the mathematical form of the dependence function, a single step fitting to Equation (10), or its integrated form, gives the terms regulating such dependence. For example, in variable-temperature kinetic (VTK) experiments the dependence function, as stated above, is the Eyring equation and the terms are the activation parameters $\Delta S^{\ddagger}$ and $\Delta H^{\ddagger}$. If the temperature changes in a linear way, that is, $M$: $T = T_0 + \alpha t$, where $T_0$ is the initial temperature and $\alpha$ is the temperature gradient, the model changes to the equation

$$-\frac{dC}{dt} = \frac{k(T_0+\alpha t)}{h}\exp\left(\frac{\Delta S^{\ddagger}}{R}\right)\exp\left(-\frac{\Delta H^{\ddagger}}{R(T_0+\alpha t)}\right)C \tag{12}$$

or to the relative integral form

$$C = C_0 \exp\left[-\int_0^t \frac{k(T_0+\alpha t)}{h}\exp\left(\frac{\Delta S^{\ddagger}}{R}\right)\exp\left(-\frac{\Delta H^{\ddagger}}{R(T_0+\alpha t)}\right)dt\right] \tag{13}$$

The fit of *C–t* experimental data, that is, the kinetic profile obtained in a single VTK run, to one of these equations gives the optimized values of $\Delta S^{\ddagger}$ and $\Delta H^{\ddagger}$.

For an experiment concerning the investigation of the dependence on the concentration of a catalyst Y, [Y], the molar concentration of the catalyst could be varied in a linear way, that is, *M*: [Y] = $\beta t$, where $\beta$ is the concentration gradient. If the dependence function is known, for example, *D:* $k_{obs} = k_Y[Y]$, the mathematical model will assume the forms

$$-\frac{dC}{dt} = k_Y \beta t C \tag{14}$$

$$C = C_0 \exp\left(-\frac{1}{2} k_Y \beta t^2\right) \tag{15}$$

A single fit of the *C–t* experimental data to one of these equations gives the optimized value of the term $k_Y$.

When the dependence function is not known, Equation (16) has to be used:

$$-\frac{1}{C}\frac{dC}{dt} = k_{obs}\{[Y](t)\} \tag{16}$$

and the $k_{obs}([Y])$ profile can be obtained by dividing the derivative of the kinetic profile to the profile itself. This method is particularly important when no kind of information is available on the dependence of the rate constant on concentrations of species present in the reaction environment and taking part in the rate-determining step. The VCK experiment, in this case, gives the empirical relation that enables a hypothesis on the dependence function to be formulated.

## 3 EXPERIMENTAL

Some basic points have to be discussed to conveniently carry out VPaK experiments:

1. Computer simulation
2. Devices to obtain variable-parameter conditions
3. Analytical instruments
4. Software for processing experimental data

All these points involve in some way the use of a personal computer. It is important to underline that it is the availability of fast and inexpensive computers that enables, nowadays, VPaK to be applied, providing a real saving of time and acceptable confidence in the obtained data.

### 3.1 Computer Simulation

Computer simulation can be very useful in the phase of projecting a VPaK experiment. VPaK profiles are often difficult to imagine compared to CPaK ones where

the rate always decreases exponentially with time [Equation (3)]. In VPaK the reaction rate is given at each time $t$ by the product of two factors: $k_{obs}[Par_i(t)]$ and $C$. The concentration always decreases with time but $k_{obs}[Par_i(t)]$ can vary in many ways depending on the form of both the dependence function and the modulating function. The search for good experimental parameters could signify a waste of precious time. Using a personal computer it is possible in a few minutes to have an acceptable profile, just obtaining some preliminary data from tests that are always carried out before a kinetic study [55]. Any software able to plot a function and, when necessary, to evaluate numerically a derivative or an integral can be used. We found the MicroMath SCIENTIST program [56] to be very versatile, very easy to use, and enabling us to use the differential form of the rate equation (10) without the necessity of integrating it. With this tool, giving as input the dependence function and the modulating function, the kinetic profile can be obtained immediately for every value of the experimental parameters. Figure 4 shows a typical list of a SCIENTIST program where the modulating function, dependence function, and first-order differential equation are indicated by arrows for a VTK simulation.

Figure 5 shows a profile obtained imposing as experimental data the values $T_0$ = 298 K and $\alpha$ = 0.0033 K/s. It is also possible to obtain in a single simulation several profiles relative to different values of terms conditioning them. Figure 6, for example, shows a multiparametric simulation of VCK profiles for a reaction of basic hydrolysis with dependence function $k_{obs} = k_{OH}[OH^-]$ and a modulating function $[OH^-] = \alpha t$, with $10^{-3} \times \alpha$M s$^{-1}$ = 1, 2.5, 4, 5.5, 7.

### 3.2 Devices to Obtain Variable-Parameter Conditions

A device is necessary to obtain VPaK conditions inside the reaction vessel. The parameter must change homogeneously and coherently with the modulating function and without altering any other factors important for the reaction or its analytical monitoring.

The parameter can change in a vessel being part of the analytical instrument, for example, an ultraviolet–visible (UV–Vis) spectrophotometric cell [39, 41, 45, 14, 47, 48], an infrared (IR) cell [42, 46], or a fluorometer cell [45, 51], or a polarimetric tube [27, 49]. It can change in a reactor vessel where the analytical signal can be read in some way, for example using an optical fiber cell for spectrophotometry [52–54] or a conductometric cell [16, 34, 40]. Another possibility is to transport the solution from the reaction vessel to the analytical instrument by a peristaltic pump [38]. When altenative ways are not practicable, samples can be taken at suitable time intervals and analyzed apart [29, 31, 35, 39, 43, 50].

Any system satisfying the conditions cited above can be useful as VPaK devices. For VTK programmable thermostatic baths (with the possibility of external circulation) and temperature programmers using the Peltier effect are very convenient. It is always necessary to homogenize the temperature and the composition of the solution with good stirring, monitoring the temperature by measuring it inside the reaction vessel, possibly memorizing it in a computer. A linear increasing temperature is sufficient and it is enough easy to realize, but other modulating functions, for various reasons, have been tested [33, 35, 47]. However, for future particular applications, the use of a computer with suitable software can easily generate $T(t)$ profiles having various shapes.

```
// VTK_SIM
IndVars: t
DepVars: C, V, K,TE
Params: ΔH, ΔS, C0, kb, h, R, TEo, α
→ TE = TEo + α*t
→ K = (kb*TE/h)*exp(ΔS /R)*exp(-ΔH /(R*TE)
→ C' = -K*C
V = -C'
// Parameter values
ΔH = 118000
ΔS = 61
→ TEo = 298
→ α = 0.0033
C0 = 1e-3
kb = 1.3806e-16
h = 6.6262e-27
R = 8.314
// Initial conditions
t = 0
C = C0
```

**FIGURE 4** List of SCIENTIST model used to simulate VTK experiment. The first three arrows show, respectively, the modulating function, dependence function, and first-order kinetic equation. The last two arrows show experimental parameters $T_0$ and α.

The thermal expansion of the reaction solution and vessel, because of their influence on the concentration of the reacting species, can be a problem for coherence to the kinetic model. The best way to solve it is to limit the range of temperature to 20–25°C; otherwise corrections are required that make the procedure less simple. Furthermore, a limited thermal excursion assures one about the constancy of the activation parameters during the experiments.

For VCK autoburettes releasing concentrated solution of species having some effect on the reaction are usually used. The concentration gradient α (in molars per second) inside the vessel is given by the relation

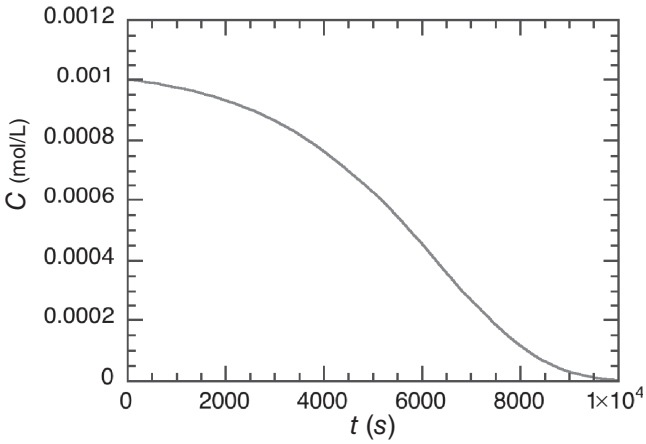


**FIGURE 5** Simulated kinetic profile for VTK experiment obtained by SCIENTIST model in Figure 4.

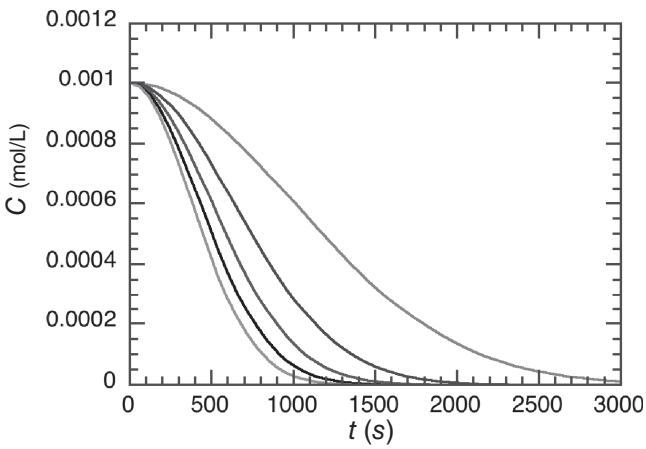


**FIGURE 6** Simulated VCK profiles obtained by SCIENTIST multiparametric model. *D:* $k_{obs} = k_{OH}[OH^-]$, *M*: $[OH^-] = \alpha t.\ 10^{-3} \times \alpha M\ s^{-1} = 1, 2.5, 4, 5.5, 7$.

$$\alpha = \frac{gM}{V_0} \tag{17}$$

where $g$ (in liters per second) is the releasing rate, $M$ (moles per liter) the molar concentration of the added solution, and $V_0$ (liters) the initial reaction volume. The chemical homogeneity is assured by good stirring. A problem can arise when the added percentage volume is large and cannot be neglected in the mathematical model. Computer simulation can help to find suitable conditions [55]. Some attempts have been made to introduce a correction in the kinetic equation [31]. The species released into the reaction environment must influence the reaction exclusively in the way described by the dependence function.

Equation (10) is generally valid for any parameter provided that a suitable apparatus is available to generate the relative experimental conditions. Some cases are very difficult to manage. For example, to our knowledge, variable-pressure kinetic experiments have so far not been reported.

### 3.3 Analytical Instruments

Theoretically, any analytical instrument useful to follow a reaction in a traditional constant-parameter kinetic experiment can be used for VPaK experiments. Obviously, some practical applications can be problematic. For example, while it is very easy to change the temperature inside a UV–Vis spectrophotometric cell, changing the concentration of a species inside a nuclear magnetic resonance (NMR) tube can be very complicated.

Following a physical property of the reacting species instead of their concentration leads to a series of consequences. First, while Equations (2), (3), and (4) used in CPaK experiments change respectively to the equations

$$-\frac{d(\lambda-\lambda_\infty)}{dt} = k_{\text{obs}}(\text{Par}_1, \text{Par}_2, \ldots)(\lambda-\lambda_\infty) \tag{18}$$

$$\lambda-\lambda_\infty = (\lambda_0-\lambda_\infty)\exp(-k_{\text{obs}}t) \tag{19}$$

$$\ln(\lambda-\lambda_\infty) = -k_{\text{obs}}t + \ln(\lambda_0-\lambda_\infty) \tag{20}$$

Equation (10) and its integral form change respectively to the equations

$$-\frac{d(\lambda-\lambda_\infty)}{dt} = \{k_{\text{obs}}[\text{Par}_i(t)]\}_{\text{Par}\neq\text{Par}_1}(\lambda-\lambda_\infty) \tag{21}$$

$$\lambda = (\lambda_0-\lambda_\infty)\exp\left(-\int_0^t k_{\text{obs}}[\text{Par}_i(t)]dt\right)+\lambda_\infty \tag{22}$$

where $\lambda$ is a physical quantity related to the compounds involved in the reaction at time $t$, $\lambda_0$ and $\lambda_\infty$ its values, respectively, at the start and at the end of the reaction. The parameter $\lambda$ can be the absorbance, the conductance, the optical rotation, the area of an NMR peak, and so on.

The more the precision of the instrument and the more the points for the time unit in the acquired profile, the better the result of the fitting of experimental data. For this reason instruments with a low measure error and connectable to a computer for the automatic and continous aquisition of data are very much prefered. The UV–Vis spectrophotometer is by far the most used instrument in chemical kinetics. It has a good sensitivity and a good control of the temperature. It is connected or easily connectable to a computer and is available nearly everywhere. The absorbance has a very low dependence on the temperature so that, in the used temperature range, its variation can be neglected during the VTK experiments.

For VCK experiments the usual 10-mm cuvette may be not suitable as a reaction vessel so an external reactor can be used where the absorbance can be read by an

UV–Vis sensor or using a flux cell. A problem can arise by adding a reagent absorbing in the same range where the reacting species or products absorb. When this is unavoidable, a correction of the kinetic profile can be made by a blank scan [53].

Unfortunately not all the molecules absorb so that other physical properties and other analytical instruments have to be used.

For reactions involving change in optical rotation, a professional polarimeter is a good solution. It has high sensitivity and can be connected to a computer for the automatic acquisition of the analytical data. The changing and monitoring of temperature can be easily done [49]. The effect of temperture on optical rotation is not large. Examples of VCK experiments using this instrument have not been reported, but they could be realized, for example, using an external reactor connected to a flux cell.

Infrared [42, 46], fluorimeters [45, 51], and conductometers [16, 34, 40] have been used with success. Gas–liquid chromatography (GLC) and high-performance liquid chromatography (HPLC) are very much used in pharmaceutical studies, for both analytical investigations and kinetic studies. Unfortunately they are not the ideal for VPaK experiments. The kinetic profile obtained is made by a few points (each one requiring a time-consuming cromatographic analysis) and the fit to the model can lead to a large error in the evaluation of the terms. Nevertheless, examples are reported and the results are acceptable [38, 39, 43, 50]. If large numbers of routine measurments would require it, suitable automation in the analysis [38] and software for the fast processing of the computer acquired chromatograms could be devised.

### 3.4 Software for Processing Experimental Data

When a VPaK experimental profile has been produced and stored on a computer, suitable software is needed for its quick processing. While in the past the treatment of constant-temperature kinetic data was easy to do, even without computers, in VPaK the use of a computer is of fundamental importance for both the complexity of the mathematical model, which does not require time-consuming calculations, and confidence in the obtained results. Algorithms are necessary to fit the experimental data to Equation (21) or (22) in their particular form depending on the investigated parameter and the used analytical instrument. Algorithms are also necessary for the evaluation of the derivative and sometimes the integral [e.g., in Equation (11) it is not possible to solve the integral in terms of elementay functions so it must be evaluated numerically]. Any language can be used for a personalized, home-made program, but application programs are commercially available to facilitate the work. The MicroMath SCIENTIST, cited for simulation for the search of the experimental conditions, is an example. The same models used to simulate VPaK profiles are used to fit the obtained data with the difference that the terms given for the simulation are in this case the input values of the terms to be optimized. SCIENTIST uses a Powell modified Marquadt for the fitting and the (default) Episode method to solve the differential equation [57].

The modulating function is always known because the value of the parameter is imposed by a program and/or monitored by a sensor. The dependence function is sometimes known and sometimes not. For example, for the parameter temperature the dependence function is always known (Arrhenius equation, Eyring equation). For the parameter concentration the dependence function can be known in studies

where a mechanism has already been proposed or guessed and a series of similar substrates have to be quantitatively investigated in detail. It can be unknown for new studies. In these cases Equation (21) can be written in the form

$$-\frac{1}{\lambda-\lambda_\infty}\frac{d\lambda}{dt}=\{k_{\text{obs}}[\text{Par}_i(t)]\}_{\text{Par}\neq\text{Par}_i} \tag{23}$$

to underline that the main calculation required is for the evaluation of the derivative of the kinetic profile. The ratio of the derivative on $\lambda - \lambda_\infty$, that is, the profile minus its value at the end of the reaction, gives at each point the value of $k_{\text{obs}}$.

Some examples concerning pharmaceutical systems will be discussed.

## 4 EXAMPLES OF VARIABLE-PARAMETER KINETIC EXPERIMENTS

### 4.1 Variable-Temperature Kinetic Experiments

Figure 7 shows a kinetic profile relative to a VTK experiment concerning the racemization of (–)-adrenaline in acidic aqueous solution [49]. It was obtained polarimetrically following the optical rotation $\alpha$. The modulating function used was $T = T_0 + \gamma t$, where $T_0 = 309.6$ K and $\gamma = 0.001694$ K/s, realized by a circulation of water coming from a programmed thermostatted bath. The temperature was read inside the polarimetric cell by a platinum resistor. Temperature and absorbance were automatically acquired by a computer connected to the instruments. The typical sigmoidal shape comes from the relation $r = k_{\text{obs}}[T(t)]C$, that is, the reaction rate is given by the product of two terms: $k_{\text{obs}}[T(t)]$, which always increases for the increasing temperature, and $C$, which decreases continously during the reaction. For this reason the reaction is accelerated in its first part but after the inflection point the low concentration of $C$ is overwhelming and the reaction decelerates until the rate

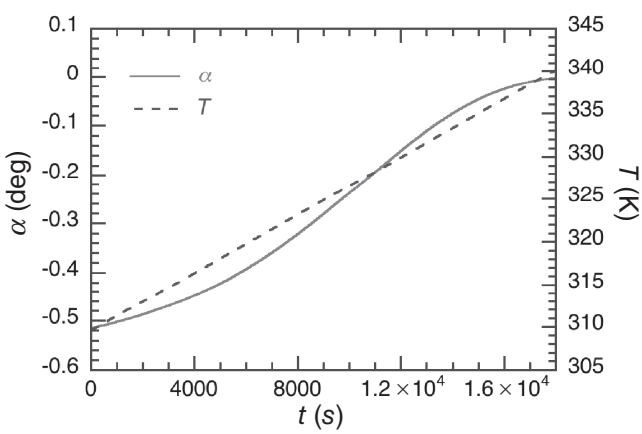


**FIGURE 7** Change in optical rotation (solid line) during racemization of (–)-adrenaline in aqueous solution (1 *M* HCl) carried out under variable-temperature conditions. *M:* $T$(K) = 309.6 + 0.001694$t$ (dashed line).

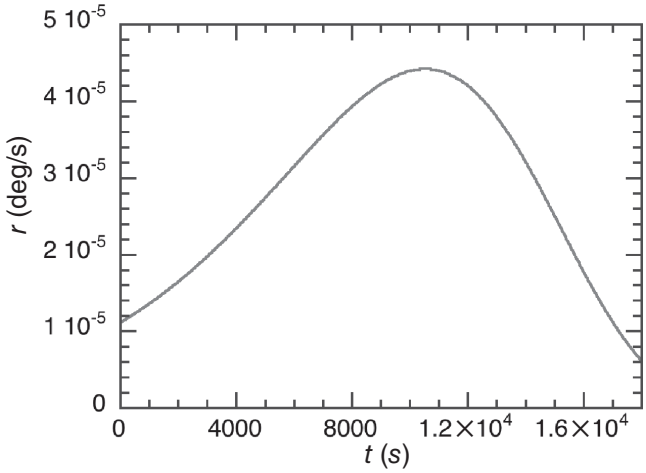


**FIGURE 8** Change in reaction rate during racemization of (−)-adrenaline as obtained using Savitzky–Golay method [58] by derivative of VTK profile reported in Figure 7.

gets to zero at the end of the reaction (Figure 8). A direct fitting of this profile to Equation (24), with $\alpha_0$, $\alpha_\infty$, $\Delta S^\ddagger$, and $\Delta H^\ddagger$ the parameters to be optimized, gave the activation parameters ($\Delta S^\ddagger = -27 \pm 1$ J/K·mol, $\Delta H^\ddagger = 95 \pm 1$ kJ/mol, $R^2 = 0.99999$). Their values were identical to those obtained by five traditional VTK experiments:

$$\alpha = (\alpha_0 - \alpha_\infty)\exp\left\{-\frac{k}{h}\exp\left[\frac{\Delta S^\ddagger}{R}\right]\int_0^t (T_0 + \gamma t)\right.$$
$$\left.\exp\left[-\frac{\Delta H^\ddagger}{R(T_0 + \gamma t)}\right]dt\right\} + \alpha_\infty \tag{24}$$

Figure 9 shows the profile relative to another VTK experiment carried out spectrophotometrically [45]. The reaction followed was the hydrolysis of aspirin [12, 13, 59, 60] at pH 4.50. The temperature was controlled by a cell compartment thermostatted by a Peltier temperature programmer and measured by a platinum resistor inserted into the spectrophotometric cell. Magnetic stirring was assured by a suitable device. Lots of data points were memorized by a computer connected to the analytical instrument and easy processings, both differential and integral, were carried out. Values of activation parameters were in agreement with each other and in agreement with those obtained by comparative constant-temperature kinetic experiments carried out in the same conditions ($\Delta S^\ddagger = -115 \pm 1$ J/K · mol, $\Delta H^\ddagger = 69 \pm 1$ kJ/mol, $R^2 = 0.99999$).

The sigmoidal profile in Figure 10 is relative to the hydrolysis of aspirin carried out at pH 7.00 under variable-temperature conditions. In this case a fluorometer was used as an analytical instrument irradiating the salicylic acid at 310 nm and recording the fluorescence signal at 404 nm [45, 61]. A programmable thermostatted bath was used to apply a linear increasing temperature ($T_0 = 323.05$ K, $\alpha = 1.631 \times 10^{-3}$ K/s) in the thermostatted fluorometric cell compartment. The temperature was

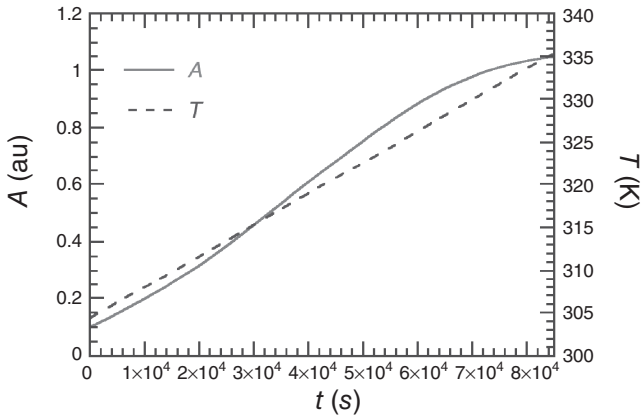


**FIGURE 9** VTK profile (solid line) obtained spectrophotometrically (λ = 298.5 nm) for hydrolysis of aspirin in water (pH = 4.50). *M:* $T$(K) = 304.36 + 3.647 × $10^{-4}t$ (dashed line).

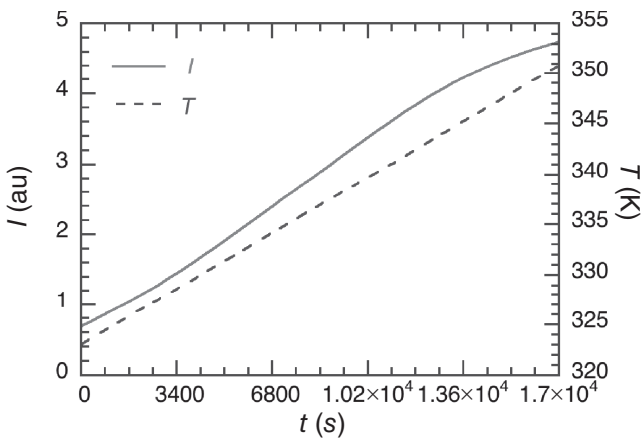


**FIGURE 10** Change in fluorescence intensity (solid line; 404 nm) during hydrolysis of aspirin (pH 7.00) in variable-temperature kinetic experiment. *M:* $T$(K) = 323.05 + 1.631 × $10^{-3}t$ (dashed line).

read inside the cell and monitored and stored by a computer together with the fluorescence intensity.

A fluorometer is more sensitive than a spectrophotometer and enables a very low concentration of substrate to be used. This can be convenient in the first stage of pharmaceutical studies. Unfortunately the dependence of fluorescence intensity on temperature cannot be avoided. A blank scan at the end of the reaction enables a temperature-independent signal [45] useful for VTK processing to be obtained. The results were consistent ($\Delta S^{\ddagger}$ = –109 ± 1 J/K · mol, $\Delta H^{\ddagger}$ = 71 ± 1 kJ/mol, $R^2$ = 0.99999) with those obtained spectrophotometrically under both CTK and VTK conditions.

Polarimeters, spectrophotometers, and fluorimeters connected to the computer can store, during a VPaK experiment, hundreds or, when necessary, even thousands

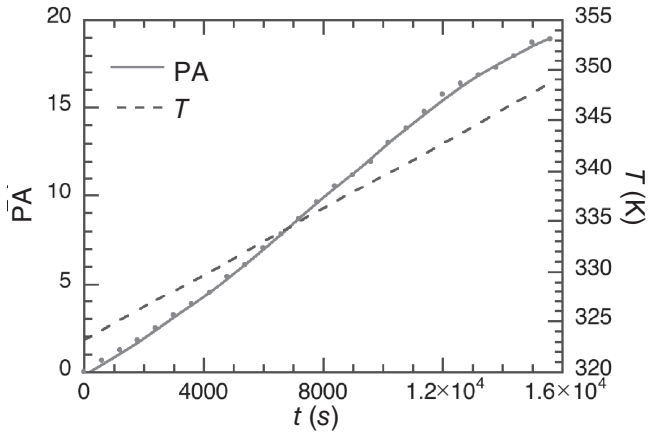


**FIGURE 11** VTK profile obtained by HPLC peak area (PA) of salicylic acid (plain circles) during hydrolysis of aspirin at pH 7.00. *M:* $T(\mathrm{K}) = 323.2 + 1.631 \times 10^{-3}t$ (dashed line). Solid line is relative theoretical curve.

of analytical points. This increases the fitting of experimental data to the mathematical model or the evaluation of the derivative of the kinetic profile.

Figure 11 shows a VTK profile obtained by HPLC analysis concerning the hydrolysis of aspirin at pH 7.00 and at a linear increasing temperature, with $T_0 = 323.2\,\mathrm{K}$ and $\alpha = 1.631 \times 10^{-3}\,\mathrm{K/s}$ [50]. The experiment was carried out in a reaction vessel immersed in a programmable thermostatic bath. Samples of the reaction mixture were taken at suitable time intervals and injected into the liquid chromatography (LC) column. Few points were collected (27) to build it up because of the time required for the chromatographic separation. Nevertheless, the fit to the relative mathematical model gave values of activation parameters similar to those obtained spectrophotometrically, although with greater statistical error ($\Delta S^{\ddagger} = -102 \pm 8\,\mathrm{J/K \cdot mol}$, $\Delta H^{\ddagger} = 73 \pm 2\,\mathrm{kJ/mol}$, $R^2 = 0.9997$).

### 4.2 Variable-Concentration Kinetic Experiments

The dependence on the concentration of species present in the reaction environment is certainly the most important in the context of mechanistic studies. A species can influence the reaction by taking part in it as a reagent or as a catalyst or simply by altering in some way the physicochemical character of the reaction environment: for example, a nucleophile in a nucleophilic substitution, an ionic metal in a reaction catalyzed by it, or a salt altering the ionic strength or molecules of solvent altering the dielectric constant. In the first case the dependence function is the heart of the rate law. It indicates the species taking part in the rate-determining step and gives an idea of the way they interact with each other [1, 3, 7, 8]. For this reason the search for its form and for the values of the terms inside it requires a great deal of attention.

A study on the nucleophilic substitution on the square planar complex *trans*-$[Pt(PEt_3)_2Cl_2]$, a substrate similar to cisplatinum and other compounds largely used as antitumoral agents [53, 62–65], has been carried out spectrophotometrically while changing the concentration of the nucleophile with time. The reaction vessel was

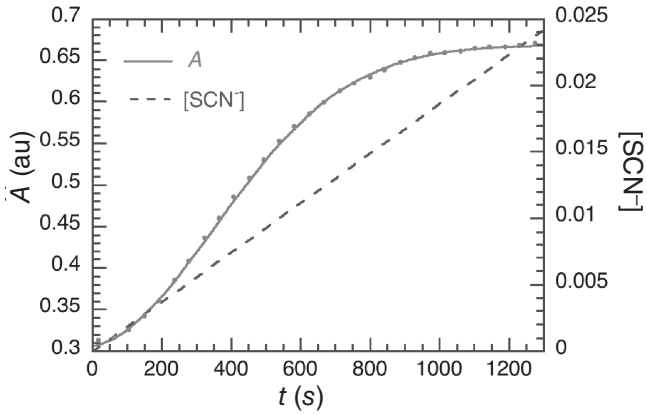


**FIGURE 12** VCK profile obtained spectrophotometrically (plain cirlce, selected points) and theoretical curve (solid line) (280 nm) for reaction *trans*-$[Pt(PEt_3)_2Cl_2]$ + 2 $SCN^-$ → *trans*-$[Pt(PEt_3)_2(SCN)_2]$ + 2 $Cl^-$ in methanol. *M*: $[SCN^-] = 0.048t$ *M* (dashed line); $T = 303.2$ K.

immersed in a thermostatted bath, an autoburette released a suitable concentrated solution of nucleophile (thiourea, iodide, bromide, thiocianide), the absorbance was read by an optical fiber cell, and good stirring was assured by an immersion stirrer.

Pseudo-first-order conditions were assured. In classic constant-concentration kinetic (CCK) experiments this means the use of a reactant in a high enough concentration, compared to that of the substrate, to neglect its consumption during the reaction. In this way the simple first-order kinetic equation can be used for the data treatment. In VCK experiments these conditions have to be considered in a different way: The concentration of the species must vary with time in a way that deviation from the modulating function, caused by its consumption, can be neglected. Concentrated solutions able to create a suitable excess of reagent in the reaction environment can be easily prepared. In particular cases, when this is not possible, the consumption of the species can be inserted into the modulating function [53].

Figure 12 shows the variable-concentration kinetic profile obtained for the reaction of the platinum complex with $SCN^-$. The modulating function was $[SCN^-] = \alpha t$, with $\alpha = 0.048$ M/s. Even in this case the reaction is accelerated in the first part of the kinetics, but this is caused by the increasing concentration of nucleophile. The dependence function for this reaction is given as

$$k_{obs} = k_S + k_Y[Y] \tag{25}$$

where $k_S$ is a solvolytic constant and $k_Y$ is the direct attack constant for a generic nucleophile Y [62, 63]. Fitting the profile to the model

$$-\frac{dA}{dt} = (k_S + k_Y \alpha t)(A - A_\infty) \tag{26}$$

gave the optimized values of $k_S$ and $k_Y$ (respectively, $0.92 \times 10^{-4}\,s^{-1}$ and $0.351\,M^{-1} \cdot s^{-1}$)

```
  // VCK_SIM
  IndVars: t
  DepVars: A, V, K,Y
  Params: KS, KY, A0, AI
  α = g*M/V0
→ Y =  α*t
→ K = KS + KY*Y
→ A' = -K*(A-AI)
  V = -A'
  // Parameter values
  KS = 0.0001
  KY = 0.3
→ g = 0.59
→ M = 0.63
→ V0 = 20000
  A0 = 0.3
  AI = 0.7
  // Initial conditions
  t = 0
  A = A0
```

**FIGURE 13** SCIENTIST list of model used for processing of VCK profile in Figure 11. The first three arrows show, respectively, the modulating function, dependence function, and first-order kinetic equation. The last three arrows show the experimental parameters *g, M*, and $V_0$.

Figure 13 shows the SCIENTIST list of the model used for the fitting procedure of the kinetic profile, where KS, KY, A0, and AI are input values to be optimized.

The reaction of the same substrate with thiourea in variable-concentration conditions was also followed conductometrically with good results [16]. The reaction was carried out in a thermostatted reaction vessel. An autoburette was used to add a concentrated solution of thiourea. The conductance was read by means of a conductometric cell and acquired by a computer.

Among the studies on the dependence of $k_{obs}$ on the concentration of species present in the reaction environment, in the pharmaceutical field a particular space is dedicated to pH–rate profiles, that is, the dependence on $[H^+]$ and/or $[OH^-]$. They can give a lot of information on the reaction mechanism and on the way to confront the instability [1, 3, 4]. Very often these studies require lots of kinetic experiments because the hydrogen ion concentration varies over 14 orders of magnitude. To

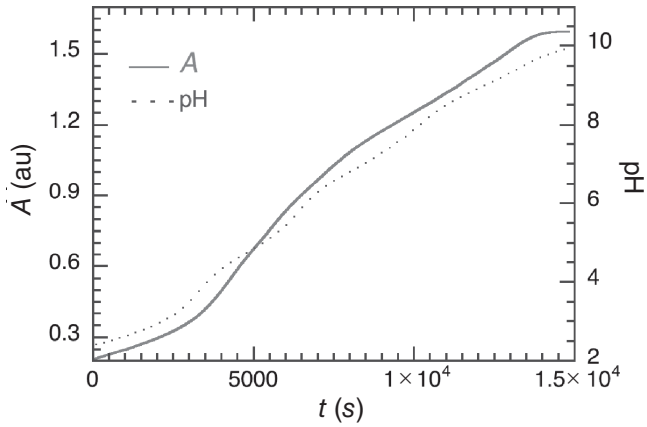


**FIGURE 14** Change in absorbance (solid line) during a VpHK experiment concerning hydrolysis of aspirin at $T = 342.5$ K, $\lambda = 298.5$ nm. The dotted line shows the variation of pH. *M*: not known analytically.

delineate the pH–rate profile of aspirin, for example, about 50 kinetic experiments are necessary. Considering different temperatures, comparative molecules, and other various effects, a study on homologous series can be very hard to carry out.

With VCK experiments, a single run can be enough to obtain the pH–rate profile of a reacting substrate.

Figure 14 shows the variable-pH kinetic (VpHK) profile obtained spectrophotometrically for the reaction of hydrolysis of aspirin with pH varying in the range 2–10 at $T = 342.5$ K. The variable-concentration conditions were realized by adding a concentrated solution of NaOH (0.6 *M*) to the thermostatted reaction vessel containing the aqueous solution of acetylsalicylic acid and a buffer composed of acetic acid (0.01 *M*), fosforic acid (0.01 *M*), and boric acid (0.01 *M*). In this way an almost linear increase of pH was generated. The absorbance was read by an optical fiber cell and stored in a computer. The pH was monitored by a pH sensor connected to a computer.

This profile is more complex than the others seen before. This is caused by the fact that, in the reaction rate equation, given by the product of $k_{obs}$ and the concentration of substrate, $k_{obs}[\mathrm{pH}(t)]$ varies with the pH in an irregular way because of the particular shape of the pH–rate profile. Figure 15 shows the derivative of the pH–rate profile as obtained by the Savitzky–Golay method [58]. It gives an idea of the variation of the reaction rate during the VpHK experiment. There is an acceleration in the first part, caused by the increase of $k_{obs}$ with the pH, followed by a deceleration for the stabilization of $k_{obs}$ and the continuous decrease of the substrate concentration. Then, the rate increases again for the increase of $k_{obs}$ for higher values of pH and after decreases again for concentration approaching zero.

The ratio of the derivative of the profile to $A - A_\infty$ gave, according to Equation (27), the entire pH–rate profile in a single scan (Figure 15).

For such large ranges of pH, the dependence function can be very complex. For example, in the case of aspirin, changing the pH, four different mechanisms of reaction operate and the global rate constant requires several terms [3]. Nevertheless, once the pH–rate profile has been obtained and a reaction mechanism formulated,

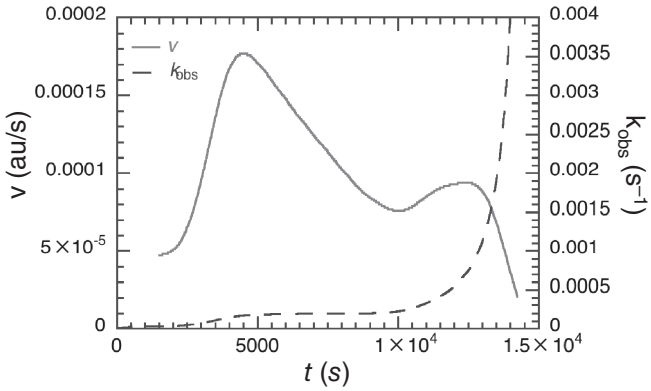


**FIGURE 15** Change in derivative of VpHK profile reported in Figure 14 (solid line) and $k_{obs}$ [pH($t$)] profile as obtained by Equation (27) (dashed line).

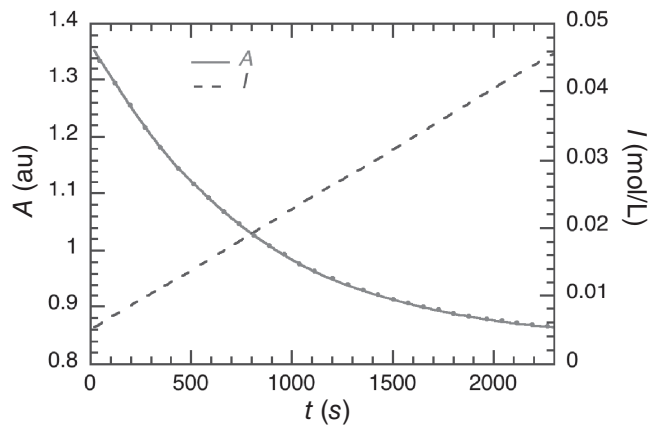


**FIGURE 16** Change in absorbance (plain circle, selected points) during VIK experiment concerning reaction of indomethacine with NaOH 0.005 $M$ and $T = 299.0 \pm 0.1$ K. The dashed line shows the increase in ionic strength. The solid line is relative to the theoretical model.

a dependence function can be inserted into Equation (27) and a global fit can be done for the evaluation of the values of the elementary constants:

$$-\frac{d(A - A_\infty)}{dt} = \{k_{obs}[[H^+](t)]\}(A - A_\infty) \tag{27}$$

### 4.3 Variable-Ionic-Strength Kinetic Experiments

Figure 16 shows the kinetic profile for a variable-ionic-strength kinetic (VIK) experiment concerning the hydrolysis of indomethacin [54, 66, 67] at NaOH 0.005 $M$ and $T = 299.0 \pm 0.1$ K at a varying ionic strength $I = I_0 + \alpha t$ with $I_0 = 0.005$ $M$, due to the concentration of NaOH, and $\alpha = 1.77 \times 10^{-5}$ M/s, calculated by the equation $\alpha = gM/V_0$. The experimental apparatus was similar to that used for the VpHK experiment. In this case a concentrated solution of LiCl (3 $M$) was released by the auto-

burette into the reaction vessel. Hundreds of absorbance data points were acquired. A fit of the experimental data to the mathematical model (28) was performed using the MicroMath SCIENTIST program with $A_0$, $A_\infty$, $k_0$, and $Z_A Z_B$ the parameters to be optimized. The fitting was excellent ($R^2 = 0.99999$) and the results were in good agreement with those obtained in the traditional way [$k_0 = (1.02 \pm 0.04)\ 10^{-4}\ s^{-1}$, $Z_A Z_B = 1.01 \pm 0.04$]:

$$-\frac{dA}{dt} = k_0 \times 10^{1.04 Z_A Z_B \frac{\sqrt{I}}{1+\sqrt{I}}} (A - A_\infty) \tag{28}$$

The differential method was also applied for processing the experimental data. The result was in good agreement.

## 5 CONCLUSIONS

Stability is an essential property of the drug product. A fast screening of new molecular entities and low-time-consuming detailed studies on selected candidates can avoid delay and consequent high cost in the first part of pharmaceutical investigation.

Potentially, VPaK provides a powerful new method for collecting kinetic data. It is a new way to look at kinetic experiments. Instead of obtaining a single value of the observed rate constant, with a single run, it is possible to obtain the entire dependence of the rate constant on a physical parameter. Usually, in mechanistic studies, what one looks for is the intimate way of interaction of the reagents in the rate-determining step. The nature of the reagent and products are well known because of the very good analytical instuments available. Kinetic experiments are just routine operations but, so far, they are of fundamental importance because kinetics is the only way to look at this aspect of the reactivity of the substrate.

Without fear of spending too much time it is possible to obtain a full panoramic picture of the chemical behavior of a long series of homologous compounds. Plain mechanistic studies can be carried out using easily accessible instruments and software. Furthermore, kinetic data are obtained using a single sample, avoiding the inhomogeity characterizing traditional CPaK experimental data, loss of time in preparing more samples, and often, but not of secondary importance, using a lower quantity of compound.

## REFERENCES

1. Carstensen, J. T. (2000), Solution Kinetics; Kinetic pH profiles; Oxidation in solution; Catalysis, Complexation, and Photolysis, in Carstensen, J. T. and Rhodes, C. T., Eds., *Drug Stability, Principle and Practice*, 3rd ed., Marcel Dekker, New York, Chapters 2–5, pp. 19–143.
2. Connors, K. A., Amidon, G. L., and Stella, V. J. (1986), *Chemical Stability of Pharmaceuticals: A Handbook for Pharmacists*, 2nd ed., Wiley-Interscience, New York.
3. Loudon, G. M. (1991), Mechanistic interpretation of pH–rate profiles, *J. Chem. Ed.*, 68, 973–984.
4. Jenks, W. P. (1969), *Catalysis in Chemistry and Enzymology*, McGraw Hill, New York.

5. Kerns, E. H. (2001), High throughput physicochemical profiling for drug discovery, *J. Pharm. Sci.*, 90, 1838–1858.
6. Avdeef, A., and Testa, B. (2002), Physicochemical profiling in drug reserch: A brief survey of the state-of-the-art of experimental techniques, *Cell. Mol. Life Sci.*, 59, 1681–1689.
7. Moore J. W., and Pearson R. G. (1981), *Kinetics and Mechanism*, Wiley, New York.
8. Wilkins R. G. (1991), *Kinetics and Mechanism of Reactions of Transition Metal Complexes*, VCH, Weinheim.
9. Kresge, A. J. (1987), Unusual reactivity of prostacyclin: Rational drug design through physical organic chemistry, *Acc. Chem. Res.*, 20, 364–370.
10. Page, M. I., and Webster, P. (1990), The hydrolysis of azetidinyl amidinium salts. Part 1. The unimportance of strain release in the four-membered ring, *J. Chem. Soc. Perkin Trans.*, 2, 805–811.
11. Lajis, N. H., Noor, H. M., and Khan, M. N. (1995), Kinetic and mechanism of the alkaline hydrolysis of securinine, *J. Pharm. Sci.*, 84, 126–130.
12. Edwards, L. J. (1950), The hydrolysis of aspirin. A determination of the thermodynamic dissociation constant and a study of the reaction kinetics by ultraviolet spectrophotometry, *Trans. Faraday Soc.*, 46, 723–735.
13. Edwards, L. J. (1952), The hydrolysis of aspirin. Part 2, *Trans. Faraday Soc.*, 48, 696–699.
14. Romeo, R., and Alibrandi, G. (1997), Structure-reactivity correlations for the dissociative uncatalyzed isomerization of monoalkylbis(phosphine)platinum(II) solvento complexes, *Inorg. Chem.*, 36, 4822–4830.
15. Carstensen, J. T. (1970), Kinetic salt effect in pharmaceutical investigation, *J. Pharm. Sci.*, 59, 1140–1143.
16. Koryta, J., Dvorak J., and Kavan, L. (1993), *Principles of Electrochemistry*, Wiley, Chichester.
17. Alibrandi, G. (1994), Variable-concentration kinetics, *J. Chem. Soc. Chem. Commun.*, 23, 2709–2710.
18. Koch, E. (1977), *Non-Isothermal Reaction Analysis*, Academic, London.
19. Brown, M. E., and Phillpotts, C. A. R. (1978), Non-isothermal kinetics, *J. Chem. Ed.*, 55, 556–560.
20. Kissinger, H. E. (1957), Reaction kinetics in differential thermal analysis, *Anal. Chem.*, 29, 1702–1706.
21. Flynn, J. H. (1969), in Schwenker, R. V., and Garn, P. D., Eds., *Thermal Analysis*, Vol. 2, Academic, New York, p. 1111.
22. Daniels, T. (1973), *Thermal Analysis*, Kogan Page, London.
23. Dollimore, D. (1996), Thermal analysis, *Anal. Chem.*, 68, 63R–71R.
24. Ozawa, T. (2000), Thermal analysis—Review and prospect, *Thermoch. Acta*, 335, 35–42.
25. Glass, B. D., Vovàk, Cs., and Brown M. E. (2004), The thermal and photostability of solid pharmaceuticals. A review, *J. Therm. Anal. Calorim.*, 77, 1013–1036.
26. Rogers, A. R. (1963), An accelerated stage test with programmed temperature rise, *J. Pharm. Pharmacol.*, 15, 101T–105T.
27. Ahlberg, P., and Wold, S. (1970), Evaluation of activation parameters for a first order reaction from one kinetic experiment. Theory, numerical methods and computer program, *Acta Chem. Scand.*, 24, 618–632.
28. Ahlberg, P. (1970), Determination of activation parameters in one kinetic experiment. *Acta Chem. Scand.*, 24, 1883–1893.
29. Madsen, B. W., and Anderson, R. A. (1974), Integral approach to nonisothermal estimation of activation energy, *J. Pharm. Sci.*, 63, 777–781.

30. Edel, B., and Baltzer, M. O. (1980), Nonisothermal kinetics with programmed temperature steps, *J. Pharm. Sci.*, 69, 287–290.
31. Tucker, I. G., and Owen, W. R. (1982), High information kinetic studies: Non-isothermal programmed acid concentration kinetics, *Int. J. Pharm.*, 10, 323–337.
32. Gonzalez, J. L., and Salvador, F. (1982), Kinetics of reactions in solution: Method for the treatment of data from non-isothermal chemical kinetic experiments, *React. Kinet. Catal. Lett.*, 21(1–2), 167–171.
33. Gonzalez, J. L., and Salvador, F. (1984), Comparative studies of non-isothermal methods in linear and non-linear temperature variation, *React. Kinet. Catal. Lett.*, 25(1–2), 125–130.
34. Mason, T. J., and Lorimer, J. P. (1983), A method for the determination of the activation energy for a reaction from a single kinetic run, *Comp. Chem.*, 7(4), 159–163.
35. Li Wan Po, A., Elias, A. N., and Irwin, W. J. (1983), Non-isothermal and non-isopH kinetics in formulation studies, *Acta Pharm. Suec.*, 20, 277–286.
36. Ortiz Uribe, M. I., Romero Salvador, A., and Irabien Gulias, A. (1985), Kinetic analysis for liquid-phase reactions from programmed temperature data. I. Simple analysis of potential kinetic laws, *Thermochem. Acta*, 94, 323–331.
37. Kipp, J. E. (1985), Non-isothermal kinetics—Comparison of two methods of data treatment, *Int. J. Pharm.*, 26, 339–354.
38. Kipp, J. E., Jensen, M. M., Kronholm, K., and McHalsky, M. (1986), Automated liquid chromatography for non-isothermal kinetic studies, *Int. J. Pharm.*, 34, 1–8.
39. Bunce, N. J., Forber, C. L., and McInnes, C. (1988), Single-step methods for calculating activation parameters from raw kinetic data, *J. Chem. Soc. Perkin Trans.*, II, 363–368.
40. Schoenemann, E., Hahn, H., and Bracht, A. (1991), Determination of kinetic parameters from non-isothermal conductivity measurements by an integral method, *Thermochim. Acta*, 185(1), 171–176.
41. Alibrandi, G. (1994), Non-isothermal spectrophotometric kinetics applied to inorganic reactions, *Inorg. Chim. Acta*, 221, 31–34.
42. Zhang, S., and Brown, T. L. (1995), Application of non-isothermal approach to the kinetics of organometallic reactions: The substitution of ($\eta^5$-pentamethylcyclopentadienyl) dicarbonylrhodium(I), *Inorg. Chim. Acta*, 240, 427–433.
43. Junnarkar, G. H., and Stavchansky, S. (1995), Isothermal and nonisothermal decomposition of famotidine in aqueous-solution, *Pharm. Res.*, 12(4), 599–604.
44. Lee, M. L., and Stavchansky, S. (1995), Isothermal and nonisothermal decomposition of thymopentin and its analogs in aqueous-solution, *Pharm. Res.*, 15(11), 599–604.
45. Alibrandi, G., Micali, M., Trusso, S., and Villari, A. (1996), Hydrolysis of aspirin studied by spectrophotometric and fluorometric variable-temperature kinetics, *J. Pharm. Sci.*, 85, 1105–1108.
46. Maeder, M., Molloy, K. J., and Schumacher, M. M. (1997), Analysis of non-isothermal kinetic measurements, *Anal. Chim. Acta*, 337, 73–81.
47. Hodgson, S. C., Ngeh, L. N., Orbell, J. D., and Bigger, S. W. (1998), A student experiment in non-isothermal chemical kinetics, *J. Chem. Ed.*, 75(9), 1150–1153.
48. Ficarra, R., Villari, A., Micali, N., Tommasini, S., Calabrò, M. L., Di Bella, M. R., Melardi, S., Agresta, M. F., Coppolino, S., and Stancanelli, R. (1999), Stability study of piroxicam and cinnoxicam in solid pharmaceuticals, *J. Pharm. Biomed. Anal.*, 20, 283–288.
49. Alibrandi, G., Coppolino, S., D'Aliberti, S., Ficarra, P., Micali, N., and Villari, A. (2002), Temperature-rate profiles by polarimetric variable-temperature kinetic experiments to study racemization reactions, *J. Pharm. Biomed. Anal.*, 29, 1025–1029.

50. Alibrandi, G., Coppolino, S., D'Aliberti, S., Ficarra, R., Micali, N., and Villari, A. (2003), Fast drug stability determination by LC variable-parameter kinetic experiments, *J. Pharm. Biomed. Anal.*, 32, 1073–1079.
51. Mood, A. R. H., Haghighi, S., and Gholami, M. R. (2004), Fluorometric variable-temperature kinetic investigations of the transesterification reaction of procaine with aliphatic alcohols, *J. Pharm. Pharm. Sci.*, 7(1), 88–91.
52. Alibrandi, G., Coppolino, S., Micali, N., and Villari, A. (2001), Variable-pH kinetics: An easy determination of pH-rate profile, *J. Pharm. Sci.*, 90, 270–274.
53. Alibrandi, G., D'Aliberti, S., and Tresoldi, G. (2003), Spectrophotometric variable-concentration kinetic experiments applied to inorganic reactions, *Int. J. Chem. Kinet.*, 35, 497–502.
54. Alibrandi, G., Coppolino, S., D'Aliberti, S., Ficarra, P., Micali, N., and Villari, A. (2003), Variable-ionic strength kinetic experiments to study drug stability, *J. Pharm. Sci.*, 92, 1730–1733.
55. Alibrandi, G., D'Aliberti, S., and Pedicini, R. (2001), Computer simulation of variable-parameter kinetic experiments, *Chem. Ed.*, 6, 185–191.
56. MicroMath Scientific Software, Salt Lake City, UT.
57. Press, W. H., Flannery, B. P., Teukolsky, S. A., and Vetterling, W. T. (1986), *Numerical Recipes*, Cambridge University Press, Cambridge.
58. Savitzky, A., and Golay, M. J. E. (1964), Smoothing and differentiation of data by simplified least squares procedures, *Anal. Chem.*, 36, 1627–1639.
59. Garrett, E. R. (1957), The kinetics of solvolysis of acyl esters of salycilic acid, *J. Am. Chem. Soc.*, 79, 3401–3408.
60. Fersht, A. R., and Kirby, A. J. (1967), The hydrolysis of aspirin. Intramolecular general base catalysis of ester hydrolysis, *J. Am. Chem. Soc.*, 89, 4857–4863.
61. Miles, C. I., and Schenk, G. H. (1970), Fluorescence of acetylsalicylic acid in solution and its measurement in presence of salicylic acid, *Anal. Chem.*, 42, 656–659.
62. Tobe, M. L. (1987), Substitution reactions, in Wilkinson, G., Gillard, R. D., and McCleverty, J. A., Eds., *Comprehensive Coordination Chemistry*, Vol. 1, Pergamon, Oxford, Chapter 7.1, pp. 281–329.
63. Belluco, U., Cattalini, L., Basolo, F., Pearson, R. G., and Turco, A. (1965), Nucleophilic constants and substrate discrimination factors for substitution reactions of platinum(II) complexes, *J. Am. Chem. Soc.*, 87, 241–246.
64. Sherman, S. E., and Lippard, S. J. (1987), Structure aspect of platinum anticancer drug interactions with DNA, *Chem. Rev.*, 87, 1153–1181.
65. Nicolini, M., Ed. (1988), *Platinum and Other Metal Coordination Compounds in Cancer Chemiotherapy*, Martinus Nijoff Publishing, Boston, MA.
66. Hajratwala, B. R., and Dawson, J. E. (1977), Kinetics of indomethacin degradation I: Presence of alkali, *J. Pharm. Sci.*, 66, 27–29.
67. Cipiciani, A., Ebert, C., Linda, P., Rubessa, F., and Savelli, G. (1983), Kinetics and mechanism of the basic hydrolysis of indomethacin and related compounds: A reevaluation, *J. Pharm. Sci.*, 72, 1075–1076.

# ANALYTICAL METHOD VALIDATION: PRINCIPLES AND PRACTICES

Chung Chow Chan
*Azopharma Contract Pharmaceutical Services, Miramar, Florida*

**Contents**

1 Introduction
2 Why Validate Analytical Procedures
3 Current Good Manufacturing Practices in Twenty-First Century
4 Cycle of Analytical Methods
5 Analytical Method Validation Characteristics
  5.1 Accuracy
  5.2 Method Precision
  5.3 Specificity
  5.4 Detection Limit
  5.5 Quantitation Limit
  5.6 Linearity
  5.7 Range
  5.8 Robustness
6 Process of Analytical Method Validation
7 Information Required in Analytical Procedure
8 Phase-Appropriate Method Validation
9 Method Verification
10 Method Revalidation
11 Conclusion
References

## 1 INTRODUCTION

Validation of an analytical procedure is the process by which it is established, by laboratory studies, that the performance characteristics of the procedure meet the

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

requirements for its intended use. All analytical methods that are intended to be used for analyzing any clinical samples will need to be validated. Validation of analytical methods is an essential but time-consuming activity for most analytical development laboratories. It is therefore important to understand the requirements of method validation in more detail and the options that are available to allow for optimal utilization of analytical resources in a development laboratory.

## 2 WHY VALIDATE ANALYTICAL PROCEDURES

There are many reasons for the need to validate analytical procedures. Among them are regulatory requirements, good science, and quality control requirements. The *Code of Federal Regulations* (CFR) 311.165c explicitly states that "the accuracy, sensitivity, specificity, and reproducibility of test methods employed by the firm shall be established and documented." Of course, as scientists, we would want to apply good science to demonstrate that the analytical method used had demonstrated accuracy, sensitivity, specificity, and reproducibility. Finally management of the quality control unit would definitely want to ensure that the analytical methods that the department uses to release its products are properly validated for its intended use so the product will be safe for human use.

## 3 CURRENT GOOD MANUFACTURING PRACTICES IN TWENTY-FIRST CENTURY

The overarching philosophy in current good manufacturing practices (cGMPs) of the twenty-first century and in robust modern quality systems is *quality should be built into the product, and testing alone cannot be relied on to ensure product quality.* From the analytical perspective, this will mean that analytical methods used to test these products should have quality attributes built into them. To have quality attributes built into the analytical method will require that fundamental quality attributes be applied by the bench-level scientist. This is a paradigm shift that requires the bench-level scientist to have the scientific and technical understanding, product knowledge, process knowledge, and/or risk assessment abilities to appropriately execute the quality functions of analytical method validation. It will require (1) the appropriate training of the bench-level scientist to understand the principles involved with method validation and able to validate an analytical method and understand the principles involved with the method validation, (2) proper documentation and understanding and interpreting data, and (3) cross-functional understanding of the effect of their activities on the product and the customer (the patient). It is the responsibility of management to verify that skills gained from the training are implemented in day-to-day performance.

## 4 CYCLE OF ANALYTICAL METHODS

The analytical method validation activity is not a one-time study. This is illustrated and summarized in the life cycle of an analytical procedure in Figure 1. An analytical

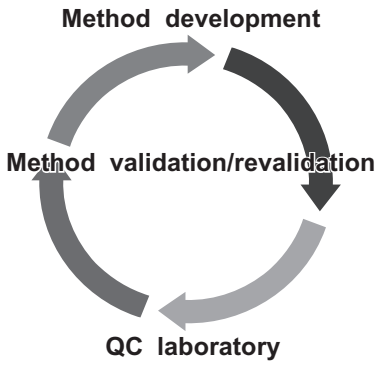


**FIGURE 1** Life cycle of analytical method.

method will be developed and validated for use to analyze samples during the early development of an active pharmaceutical ingredient (API) or drug product. As drug development progresses from phase 1 to commercialization, the analytical method will follow a similar progression. The final method will be validated for its intended use for the market image drug product and transferred to the quality control laboratory for the launch of the drug product. However, if there are any changes in the manufacturing process that have the potential to change the analytical profile of the drug substance and drug product, this validated method may need to be revalidated to ensure that it is still suitable to analyze the API or drug product for its intended purpose.

## 5 ANALYTICAL METHOD VALIDATION CHARACTERISTICS

Typical analytical performance characteristics that should be considered in the validation of the types of procedures described in this chapter are listed below. Each validation characteristic is defined to ensure consistency in usage of terminology and interpretation:

- Accuracy
- Precision
  - Repeatability
  - Intermediate precision
- Specificity
- Detection limit
- Quantitation limit
- Linearity
- Range
- Robustness

### 5.1 Accuracy

The International Convention on Harmonization (ICH) defines the accuracy of an analytical procedure as the closeness of agreement between the values that are accepted either as conventional true values or an accepted reference value and the value found. For drug substance, accuracy may be defined by the application of the analytical procedure to an analyte of known purity (e.g., a reference standard). For the drug product, accuracy will be determined by application of the analytical procedure to synthetic mixtures of the drug product components to which known amounts of analyte have been added within the range of the procedure. The ICH document also recommends assessing a minimum of nine determinations over a minimum of three concentration levels covering the specified range (e.g., three concentrations/three replicates).

Accuracy is usually reported as percent recovery by the assay (using the proposed analytical procedure) of known added amount of analyte in the sample or as the difference between the mean and the accepted true value together with the confidence intervals. The range for the accuracy limit should be within the linear range. Typical accuracy of the recovery of the drug substance is expected to be about 99–101%. Typical accuracy of the recovery of the drug product is expected to be about 98–102%. Values of accuracy of recovery data beyond this range need to be investigated as appropriate.

### 5.2 Method Precision

The precision of an analytical procedure expresses the closeness of agreement (degree of scatter) between a series of measurements obtained from multiple samples of the same homogeneous sample under prescribed conditions. Precision is usually investigated at three levels: repeatability, intermediate precision, and reproducibility. For simple formulation it is important that precision be determined using authentic homogeneous samples. A justification will be required if a homogeneous sample is not possible and artificially prepared samples or sample solutions are used.

***Repeatability*** Repeatability is a measure of the precision under the same operating conditions over a short interval of time, that is, under normal operating conditions of the analytical method with the same equipment. It is sometimes referred to as intra-assay precision.

The ICH recommends that repeatability be assessed using a minimum of nine determinations covering the specified range for the procedure (e.g., three concentrations/three replicates as in the accuracy experiment) or using a minimum of six determinations at 100% of the test concentration. Reporting of the standard deviation, relative standard deviation (coefficient of variation), and confidence interval is required. The assay values are independent analyses of samples that have been carried through the complete analytical procedure from sample preparation to final test result. Table 1 provides an example set of repeatability data.

***Intermediate Precision*** Intermediate precision is defined as the variation within the same laboratory. The extent to which intermediate precision needs to be estab-

**TABLE 1 Repeatability Data**

| Replicate | Percentage of Labeled Claim |
|---|---|
| 1 | 100.6 |
| 2 | 102.1 |
| 3 | 100.5 |
| 4 | 99.4 |
| 5 | 101.4 |
| 6 | 101.1 |
| Mean | 100.9 |
| Percentage relative standard deviation (%RSD) | 0.90 |

lished depends on the circumstances under which the procedure is intended to be used. Typical parameters that are investigated include day-to-day variation, analyst variation, and equipment variation. Depending on the extent of the study, the use of experimental design is encouraged. Experimental design will minimize the number of experiments that need to be performed. It is important to note that ICH allows exemption from doing intermediate precision when reproducibility is proven. It is expected that the intermediate precision should show variability that is in the same range or less than repeatability variation. ICH recommends the reporting of standard deviation, relative standard deviation (coefficient of variation), and confidence interval of the data.

***Reproducibility*** Reproducibility measures the precision between laboratories. This parameter is considered in the standardization of an analytical procedure (e.g., inclusion of procedures in pharmacopeias and method transfer between different laboratories).

To validate this characteristic, similar studies need to be performed at different laboratories using the same homogeneous sample lot and the same experimental design. In the case of method transfer between two laboratories, different approaches may be taken to achieve the successful transfer of the procedure. The most common approach is the direct-method transfer from the originating laboratory to the receiving laboratory. The originating laboratory is defined as the laboratory that has developed and validated the analytical method or a laboratory that has previously been certified to perform the procedure and will participate in the method transfer studies. The receiving laboratory is defined as the laboratory to which the analytical procedure will be transferred and that will participate in the method transfer studies. In the direct-method transfer, it is recommended that a protocol be initiated with details of the experiments to be performed and acceptance criteria (in terms of the difference between the means of the two laboratories) for passing the method transfer. Table 2 provides examples of a set of method transfer data between two laboratories.

### 5.3 Specificity

The ICH defines specificity as the ability to assess unequivocally an analyte in the presence of components that may be expected to be present. In many publications,

**TABLE 2 Results from Method Transfer between Two Laboratories**

| | Runs | Average Percent |
|---|---|---|
| Originating laboratory | 12 | 100.7 |
| Receiving laboratory | 4 | 100.2 |

selectivity and specificity are often used interchangeably. However, there are debates over the use of specificity over selectivity and some authorities, for example, the International Union of Pure and Applied Chemistry (IUPAC), have preferred the term *selectivity*, reserving *specificity* for those procedures that are completely selective. For pharmaceutical application, the above definition of ICH will be used.

For identity test, compounds of closely related structures which are likely to be present should be discriminated from each other. This could be confirmed by obtaining positive results (by comparison with a known reference material) from samples containing the analyte, coupled with negative results from samples which do not contain the analyte. Furthermore, the identification test may be applied to material structurally similar or closely related to the analyte to confirm that a positive response is not obtained. The choice of such potentially interfering materials should be based on sound scientific judgment with a consideration of the interferences that could occur.

The specificity for an assay and impurity tests should be approached from two angles:

1. *When Impurities Are Available* The specificity of an assay method is determined by comparing test results from an analysis of sample containing the impurities, degradation products, or placebo ingredients with those obtained from an analysis of samples without the impurities, degradation products, or placebo ingredients. For a stability-indicating assay method, degradation peaks need to be resolved from the drug substance. However, these impurities do not need to be resolved from each other.

   For the impurity test, the determination should be established by spiking drug substance or drug product with the appropriate levels of impurities and demonstrating the separation of these impurities individually and/or from other components in the sample matrix. Representative chromatograms should be used.
2. *If Impurities Are Not Available.* Specificity may be demonstrated by comparing the test results of samples containing impurities or degradation products to a second well-characterized procedure or other validated analytical procedure (orthogonal method). This should include samples stored under relevant stress conditions (light, heat, humidity, acid/base hydrolysis and oxidation). For the assay method, the two results should be compared; for impurity tests, the impurity profiles should be compared. Peak homogeneity tests should be performed using PDA or mass spectrometry to show that the analyte chromatographic peak is not attributable to more than one component. Figure 2 illustrates the selectivity of a method to resolve known degradation peaks from the parent peak.

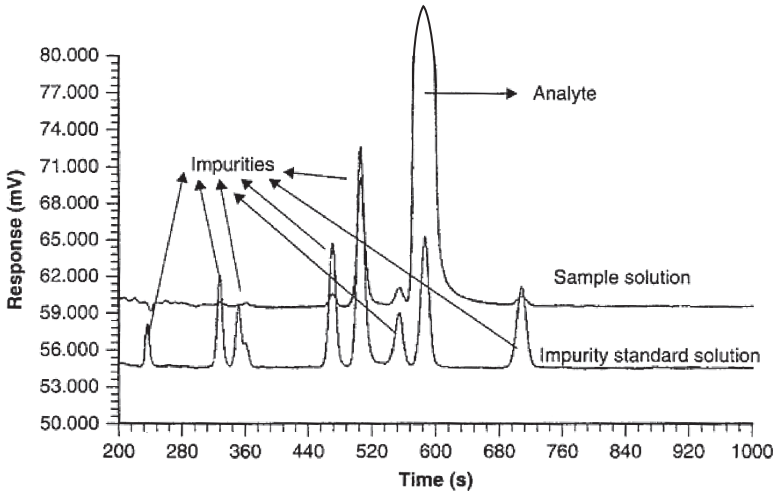


**FIGURE 2** Overlay chromatogram of impurity solution with sample solution.

### 5.4 Detection Limit

The detection limit (DL) is a characteristic for the limit test only. It is the lowest amount of analyte in a sample that can be detected but not necessarily quantitated under the stated experimental conditions. The detection is usually expressed as the concentration of the analyte in the sample, for example, percentage, parts per million (ppm), or parts per billion (ppb).

There are several approaches to establish the DL. Visual evaluation may be used for noninstrumental (e.g., solution color) and instrumental methods. In this case, the DL is determined by the analysis of a series of samples with known concentrations and establishing the minimum level at which the analyte can be reliably detected. Presentation of relevant chromatograms or other relevant data is sufficient for justification of the DL.

For instrumental procedures that exhibit background noise, it is common to compare measured signals from samples with known low concentrations of analyte with those of the blank samples. The minimum concentration at which the analyte can reliably be detected is established using an acceptable signal-to-noise ratio of 2:1 or 3:1. Presentation of relevant chromatograms is sufficient for justification of the DL.

Another approach estimates the DL from the standard deviation of the response and the slope of the calibration curve. The standard deviation can be determined either from the standard deviation of multiple blank samples or from the standard deviation of the $y$ intercepts of the regression lines done in the range of the DL. This estimate will need to be subsequently validated by the independent analysis of a suitable number of samples near or at the DL:

$$\mathrm{DL} = \frac{3\sigma}{S}$$

where σ is the standard deviation of the response and $S$ is the slope of the calibration curve.

### 5.5 Quantitation Limit

The quantitation Limit (QL) is a characteristic of quantitative assays for low levels of compounds in sample matrices, such as impurities in bulk drug substances and degradation products in finished pharmaceuticals. QL is defined as the concentration of related substance in the sample that will give a signal-to-noise ratio of 10:1. The QL of a method is affected by both the detector sensitivity and the accuracy of sample preparation at the low concentration of the impurities. In practice, QL should be lower than the corresponding ICH report limit.

ICH recommends three approaches to the estimation of QL. The first approach is to evaluate it by visual evaluation and may be used for noninstrumental methods and instrumental methods. QL is determined by the analysis of samples with known concentrations of analyte and by establishing the minimum level at which the analyte can be quantitated with acceptable accuracy and precision.

The second approach determines the signal-to-noise ratio by comparing measured signals from samples with known low concentrations of anlayte with those of blank samples. QL is the minimum concentration at which the analyte can be reliably quantified at the signal-to-noise ratio of 10:1.

The third approach estimates QL by the equation

$$\mathrm{QL} = \frac{10\sigma}{S}$$

The slope $S$ may be estimated from the calibration curve of the analyte. The value of σ may be estimated by (1) calculating the standard deviation of the responses obtained from the measurement of the analytical background response of an appropriate number of blank samples or (2) calculating the residual standard deviation of the regression line from the calibration curve using samples containing the analyte in the range of the QL.

Whatever approach is applied, the QL should be subsequently validated by the analysis of a suitable number of samples prepared at the QL and determining the precision and accuracy at this level.

### 5.6 Linearity

ICH defines linearity of an analytical procedure as the ability (within a given range) to obtain test results of variable data (e.g., absorbance and area under the curve) which are directly proportional to the concentration (amount of analyte) in the sample. The data variables that can be used for quantitation of the analyte are the peak areas, peak heights, or the ratio of peak areas (heights) of analyte to the internal standard peak. Quantitation of the analyte depends on it obeying Beer's law for the spectroscopic method over a concentration range. Therefore, the working sample concentration and samples tested for accuracy should be in the linear range.

There are two general approaches for determining the linearity of the method. The first approach is to weigh different amounts of standard directly to prepare

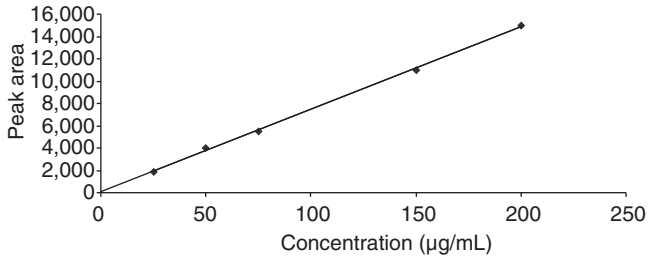


**FIGURE 3** Linearity plot of peak-area response versus concentration.

linearity solutions at different concentrations. However, it is not suitable to prepare solution at very low concentration, as the weighing error will be relatively high.

Another approach is to prepare a stock solution of high concentration. Linearity is then demonstrated directly by dilution of the standard stock solution. This is more popular and the recommended approach. Linearity is best evaluated by visual inspection of a plot of the signals as a function of analyte concentration. Subsequently, the variable data are generally used to calculate a regression line by the least-squares method. At least five concentration levels should be used. Under normal circumstances, linearity is acceptable with a coefficient of determination ($r^2$) of ≥0.997. The slope, residual sum of squares, and $y$ intercept should also be reported as required by ICH.

The slope of the regression line will provide an idea of the sensitivity of the regression, and hence the method that is being validated. The $y$ intercept will provide an estimate of the variability of the method. For example, the ratios percent of the $y$ intercept with the variable data at nominal concentration are sometimes used to estimate the method variability.

For the determination of potency assay of a drug substance or a drug product, the usual range of linearity should be ±20% of the target or nominal concentration. For the determination of content uniformity, it should be ±30% of the target or nominal concentration. Figure 3 illustrates the linearity of a sample set of data.

### 5.7 Range

The range of an analytical procedure is the interval between the upper and lower concentration of analyte in the sample for which it has been demonstrated that the analytical procedure has a suitable level of precision, accuracy, and linearity. The range is normally expressed in the same units as test results (e.g., percent, parts per million) obtained by the analytical procedure.

For the assay of drug substance or finished drug product, it is normally recommended to have a range of 80–120% of the nominal concentration.

For content uniformity, a normal range would cover 70–130% of the nominal concentration, unless a wider and more appropriate range (e.g., metered-dose inhalers) is justified.

For dissolution testing, a normal range is ±20% over the specified range. If the acceptance criterion for a controlled-release product covers a region from 20% after

1 h, and up to 90% after 24 h, the validated range would be 0–110% of the label claim. In this case, the lowest appropriate quantifiable concentration of analyte will be used as the lowest limit as 0% is not appropriate.

### 5.8 Robustness

Robustness of an analytical procedure is a measure of the analytical method to remain unaffected by small but deliberate variations in method parameters and provides an indication of its reliability during normal usage.

The evaluation of robustness is normally considered during the development phase and depends on the type of procedure under study. Experimental design (e.g., fractional factorial design or Plackett–Burman design) is common and useful to investigate multiple parameters simultaneously. The result will help to identify critical parameters that will affect the performance of the method. Common method parameters that can affect the analytical procedure should be considered based on the analytical technique and properties of the samples:

1. Sample preparation
   a. Extraction time
   b. Sample solvent (pH ± 0.05 unit, percent organic ± 2% absolute)
   c. Membrane filters
   d. Sample and standard stability
2. High-performance liquid chromatography (HPLC) conditions
   a. Mobile-phase composition (pH ± 0.05 unit, percent organic ± 2% absolute)
   b. Column used (equivalent columns, lots and/or suppliers, age)
   c. Temperature
   d. Flow rate
3. Gas chromatography (GC) conditions
   a. Column used (lots and/or suppliers, age)
   b. Temperature
   c. Flow rate

When the results are affected by some critical experimental parameters, a precautionary statement should be included in the analytical procedure to ensure that this parameter is tightly controlled between experiments. For example, if percent ionpairing of mobile phase affects the results significantly, the analytical procedure should explicitly be written with a precautionary statement for aqueous component, for example, 40% aqueous 20 m*M* octanesulfonic acid ± 2% absolute.

Other robustness considerations for ruggedness of the analytical procedure during validation include the following:

(a) *Sample Extraction* Mechanical shaking is preferred over sonication as the latter is affected by a number of factors, for example, water level in bath and position of sample.

(b) *Dilution of Sample and Solvent.* Minimize the number of dilution steps to reduce introduction of error. Dilution solvent should be as similar to mobile phase as possible.

## 6 PROCESS OF ANALYTICAL METHOD VALIDATION

The typical process that is followed in an analytical method validation is chronologically listed below:

1. Planning and deciding on the method validation experiments
2. Writing and approval of method validation protocol
3. Execution of the method validation protocol
4. Analysis of the method validation data
5. Reporting the analytical method validation
6. Finalizing the analytical method procedure

The method validation experiments should be well planned and laid out to ensure efficient use of time and resources during execution of the method validation. The best way to ensure a well-planned validation study is to write a method validation protocol that will be reviewed and signed by the appropriate person (e.g., laboratory management and quality assurance).

The validation parameters that will be evaluated will depend on the type of method to be validated. Analytical methods that are commonly validated can be classified into three main categories: identification, testing for impurities, and assay. Table 3 lists the ICH recommendations for each of these methods.

Execution of the method validation protocol should be carefully planned to optimize the resources and time required to complete the full validation study. For example, in the validation of an assay method, linearity and accuracy may be validated at the same time as both experiments can use the same standard solutions. A normal validation protocol should contain the following contents at a minimum:

(a) Objective of the protocol
(b) Validation parameters that will be evaluated
(c) Acceptance criteria for all the validation parameters evaluated
(d) Details of the experiments to be performed
(e) Draft analytical procedure

The data from the method validation data should be analyzed as the data are obtained and processed to ensure a smooth flow of information. If an experimental error is detected, it should be resolved as soon as possible to reduce any impact it may have on later experiments. Analysis of the data includes visual examination of the numerical values of the data and chromatograms followed by statistical treatment of the data if required.

**TABLE 3 Validation Parameters**

| Type of Analytical Procedure Characteristics | Identification | Testing for Impurities | | Assay – Dissolution (Measurement Only) – Content/ Potency |
|---|---|---|---|---|
| | | Quantitation | Limit | |
| Accuracy | – | + | – | + |
| Precision | – | | | |
| Repeatability | – | + | – | + |
| Intermediate precision | | +[a] | – | +[a] |
| Specificity[b] | + | + | + | + |
| Detection limit | – | –[c] | + | – |
| Quantitation limit | – | + | – | – |
| Linearity | – | + | – | + |
| Range | – | + | – | + |

*Note*: –, characteristic not normally evaluated; +, characteristic normally evaluated.[a]In cases where reproducibility has been performed, intermediate precision is not needed.[b]Lack of specificity of one analytical procedure could be compensated by other supporting analytical procedure(s).[c]May be needed in some cases.

Upon completion of all the experiments, all the data will be compiled into a detailed validation report that will conclude the success or failure of the validation exercise. Depending on the company's strategy a summary of the validation data may also be generated. Successful execution of the validation will lead to a final analytical procedure that can be used by the laboratory to support future analytical work for the drug substance or drug product.

## 7 INFORMATION REQUIRED IN ANALYTICAL PROCEDURE

The minimal information that should be included in a final analytical procedure are as follows:

(a) Rationale of the analytical procedure and description of the capability of the method. Revision of analytical procedure should include the advantages offered by the new revision.
(b) Proposed analytical procedure. This section should contain a complete description of the analytical procedure in sufficient detail to enable another analytical scientist to replicate it. The write-up should include all important operational parameters and specific instructions, for example, preparation of reagents, system suitability tests, precautions, and explicit formulas for calculation of the test results.
(c) List of permitted impurities and its levels in an impurity assay.
(d) Validation data. Either a detailed set or summary set of validation data is included

(e) Revision history.
(f) Signature of author, reviewer, management, and quality assurance.

## 8 PHASE-APPROPRIATE METHOD VALIDATION

The original intent of the cGMPs was to describe standards and activities designed to ensure the strength, identity, safety, purity, and quality of pharmaceutical products introduced into commerce. However, the GMPs are silent on explicit guidances for the development phase of pharmaceuticals in several areas.

Regulatory bodies recognize that knowledge of the drug product and its analytical methods will evolve through the course of development. This is stated explicitly in ICH Q7A: Changes are expected during development, and every change in product, specifications, or test procedures should be recorded adequately. It is therefore reasonable to expect that changes in testing, processing, packaging, and so on, will occur as more is learned about the molecule. However, even with the changes, the need for ensuring the safety of subjects in clinical testing should not be compromised.

According to the ICH guidance, the objective of method validation is to demonstrate that analytical procedures "are suitable for their intended purpose." Therefore the method's purpose should be linked to the clinical studies and the pharmaceutical purpose of the product being studied.

The purpose in the early phase of drug development is to deliver a known dose that is bioavailable for clinical studies. As product development continues, increasing emphasis is placed on identifying a stable, robust formulation from which multiple, bioequivalent lots can be manufactured and ultimately scaled up, transferred, and controlled for commercial manufacture.

The development and validation of analytical methods should follow a similar progression. The purpose of analytical methods in early stages of development is to ensure potency, to understand the impurity and degradation product profile, and to help understand key drug characteristics. As development continues, the method should be stability indicating and capable of measuring the effect of key manufacturing parameters to ensure consistency of the drug substance and drug product.

Analytical methods used to determine purity and potency of an experimental API that is very early in development will need a less rigorous method validation exercise than would be required for a quality control laboratory method at the manufacturing site. An early phase project may have only a limited number of lots to be tested and the testing may be performed in only one laboratory by a limited number of analysts. The ability of the laboratory to "control" the method and its use is relatively high, particularly if laboratory leadership is clear in its expectations for the performance of the work.

The environment in which a method is used changes significantly when the method is transferred to a quality control laboratory at the manufacturing site. The method may be replicated in several laboratories, multiple analysts may use it, and the method may be one of many methods used in the laboratory daily. The developing laboratory must therefore be aware of the needs of the receiving

**TABLE 4 Assay Method Validation in Early Phase for Drug Substance and Drug Product**

| | Drug Substance | Drug Product |
|---|---|---|
| Accuracy | Inferred from precision, linearity, and specificity | Recovery at 100% for each strength (bracket for multiple strengths) |
| Repeatability | Three sample preparation at 100% nominal | Three sample preparations at 100% nominal |
| Intermediate precision | To be completed at later stages of development | To be completed at later stages of development |
| Specificity | Resolution from most likely impurities | Resolution from impurities and excipients |
| Quantitation/detection limit | Not required | Not required |
| Linearity | Minimum three levels from 80 to 120% | Minimum three levels from 80 to 120% |
| Range | Defined in linearity | Defined in linearity |
| Robustness | Solution stability | Solution stability |

laboratories, for example, quality control laboratory, and regulatory expectations for the successful validation of a method to be used in support of a commercial product.

An example of the minimum requirement for potency assay of the drug substance and drug product is tabulated in Table 4. Note that the postponement of intermediate precision is aligned with previous discussion that the use of early phase analytical method resides mainly in one laboratory and is used only by a very limited number of analysts. Each individual company's phased method validation procedures and processes will vary, but the overall philosophy is the same. The extent of and expectations from early phase method validation are lower than the requirements in the later stages of development. The validation exercise becomes larger and more detailed and collects a larger body of data to ensure that the method is robust and appropriate for use at the commercial site.

However, certain fundamental concepts of cGMPs must be applied regardless of the details of the phased appropriate method validation strategy used. Examples are (1) proper documentation, (2) change control, (3) deviations, (4) equipment and utilities qualification, and (5) proper training.

A detailed method validation report may not be necessary until submission of the final market application. However, summary reports should be available to facilitate efficient data retrieval and fulfill requests from regulatory agencies for the information when required.

## 9 METHOD VERIFICATION

The U.S. Food and Drug Administration (FDA) regulation 21 CFR 211.194(a)(2) specifically states that users of analytical methods in the U.S. Pharmacopeia/National Formulary (USP/NF) are not required to validate the accuracy and reliability of these methods but merely verify their suitability under actual conditions of use. USP

**TABLE 5 Validation and Verification Requirements for HPLC Assay of Final Dosage Forms**

| Performance Characteristics | Validation | Verification |
|---|---|---|
| Accuracy | Yes | Maybe |
| Precision | Yes | Yes |
| Specificity | Yes | Yes |
| Limit of detection (LOD) | No | No |
| Limit of quantification (LOQ) | No | No |
| Linearity | Yes | No |
| Range | Yes | No |

has issued a guidance for verification in general chapter ⟨1226⟩. This proposal provides general information to laboratories on the verification of compendial procedures that are being performed for the first time to yield acceptable results utilizing the laboratories' personnel, equipment, and reagents.

Verification consists of assessing selected analytical validation characteristics described earlier to generate appropriate, relevant data rather than repeating the validation process for commercial products. The guidance in this general chapter is applicable to applications such as titrations, chromatographic procedures (related compounds, assay, and limit tests), and spectroscopic tests. However, general tests (e.g., water, heavy metals, residue on ignition) do not typically require verification.

Table 5 summarizes the comparison of the validation requirements with the verification requirements of the HPLC assay of an example final dosage form. ICH requires the validation of accuracy, precision, specificity, linearity, and range. Generally, verification will only require a minimal of precision and specificity validation. The accuracy requirements will be dependent on the specific situation of the final dosage form.

## 10 METHOD REVALIDATION

There are various circumstances under which a method needs to be revalidated. Some of the common situations are described below:

1. During the optimization of the drug substance synthetic process, significant changes were introduced into the process. To ensure that the analytical method will still be able to analyze the potentially different profile of the API, revalidation may be necessary.
2. If a new impurity is found that makes the method deficient in its specificity, this method will need to be modified or redeveloped and revalidated to ensure that it will be able to perform its intended function.
3. A change in the excipient composition may change the product impurity profile. This change may make the method deficient in its specificity for the assay or impurity tests and may require redevelopment and revalidation.

4. Changes in equipment or suppliers of critical supplies of the API or final drug product will have the potential to change their degradation profile and may require the method to be redeveloped and revalidated.

## 11 CONCLUSION

This chapter summarizes the validation parameters that are required according to the requirements of ICH Q2R(1). The paradigm shift under cGMP in the twenty-first century that requires the bench-level scientist to have the scientific and technical understanding, product knowledge, process knowledge, and/or risk assessment abilities to appropriately execute the quality functions of analytical method validation is presented in detail. The method validation process and the minimum requirements to be included in a regulatory method are also discussed. An overview of phase-appropriate method validation, method verification, and method revalidation are presented to stimulate ideas and the thought process to follow when such situations are encountered.

## FURTHER READING

1. International Conference on Harmonization (ICH) (2005, Nov.), Harmonised tripartite guideline Q2(R1), Validation of analytical procedures: Text and methodology.
2. International Conference on Harmonization (ICH) (1999, Oct.), Harmonised tripartite guideline Q6A, Specifications: Test procedures and acceptance criteria for new drug substances and new drug products: Chemical substances.
3. International Conference on Harmonization (ICH) (2000, Nov.), Harmonised tripartite guideline Q7A GMP for active pharmaceutical ingredient.
4. International Conference on Harmonization (ICH) (2006, Sept.), Guidance for industry: Quality systems approach to pharmaceutical cGMP.
5. Code of Federal Regulations (CFR), Part 211, Current good manufacturing practice for finished pharmaceuticals.
6. Chan, C. C., et al., (2004), *Analytical Method Validation and Instrument Performance Verification*, J Wiley, Hoboken, NJ.
7. U.S. Pharmacopeia (USP), General Chapter ⟨1225⟩, Validation of compendial procedures, USP, Rockville, MD.
8. U.S. Pharmacopeia (USP), General Chapter ⟨1226⟩, Verification of compendial procedures, USP, Rockville, MD.

# ANALYTICAL METHOD VALIDATION AND QUALITY ASSURANCE

Isabel Taverniers, Erik Van Bockstaele, and Marc De Loose
*Institute for Agricultural and Fisheries Research (ILVO), Scientific Institute of the Flemish Community, Merelbeke, Belgium*

**Contents**

1 Introduction
2 Traceability and Measurement Uncertainty
  2.1 Introduction: Quality of Analytical Results
  2.2 Role of Method Validation in Traceability and MU
  2.3 Guidelines on Traceability and Uncertainty of Results
  2.4 Concept of Traceability
  2.5 Concept of MU
  2.6 Different Operational Definitions of MU
  2.7 Approaches to Establish MU
  2.8 Importance of Traceability and MU
  2.9 Summary
3 Method Validation and Quality Assurance
  3.1 Role of Method Validation in AQA
  3.2 Guidelines and Guidance on AQA
  3.3 Approaches for Evaluating Acceptable Methods of Analysis
  3.4 Method Performance Characteristics and Criteria Approach
  3.5 Analytical Quality Assurance
4 Summary
  References

## 1 INTRODUCTION

Credibility of analytical data has never caught the public's eye more than today. Rather than on the used techniques and methodologies themselves, attention is nowadays paid to the quality and reliability of the final results. This is influenced by

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

a higher demand for regulatory compliance, a higher consciousness of the customer—the client wants to know the level of confidence of the reported result—and under impulse of new, more exigent European and international standards such as the International Organization for Standardization/International Electrotechnical Commission (ISO/IEC) 17025 norm for laboratory accreditation. The underlying key principle is comparability of results between laboratories and on a wider, international basis. In order for results to be comparable, they must be reported with a statement of measurement uncertainty (MU) and they must be traceable to common primary references. Methods must be validated to show that they actually measure what they are intended to measure—that they are fit for their specific purpose.

Because validation and quality assurance (QA) apply for a specific analytical method, it is important to approach each method on a case-by-case basis. An analytical method is a complex, multistep process starting with sampling and ending with the generation of a result. Although every method has its specific scope, application, and analytical requirement, the basic principles of validation and QA are the same regardless the type of method or sector of application. The information in this chapter is mainly taken from analytical chemistry, but it applies to other sectors as well. The validation of analytical methods, the establishment of traceability of results, and the assessment of MU should be done in a uniform, harmonized way, conforming with internationally recognized standards from institutions such as Eurachem, the International Union of Pure and Applied Chemistry (IUPAC), or ISO.

It is important to issue a common understanding on the topics of method validation, traceability, and uncertainty of measurements. Here, the interrelationships between method validation, traceability, and MU of results will be elucidated. Throughout the landscape of guidelines and standards, the most relevant information is selected, compiled, and summarized. Great importance is attached to the different method performance parameters and their definitions, ways of expression, and approaches for practical assessment. We discuss the role of method validation within QA as well as the topics of standardization, internal and external quality control (IQC and EQC, respectively), and accreditation and the links between these different aspects.

This chapter provides a general, complete, and up-to-date overview of the topic of quality of analytical measurements in the wide sense of the word. It is useful for the completely inexperienced scientist as well as those involved in this topic for a long time who have lost their way in the labyrinth and are looking for more explanation on a particular aspect or deeper insight and knowledge.

## 2 TRACEABILITY AND MEASUREMENT UNCERTAINTY

### 2.1 Introduction: Quality of Analytical Results

Innumerable types of analytical methods exist in the fields of analytical and bioanalytical chemistry, biochemistry, biology, clinical biology, and pharmacology and related application domains such as forensic, toxicological, environmental, agricultural, and food analyses. Regardless of the type of method, scope, and application, laboratories must be able to produce reliable data when performing analytical tests for a client or for regulatory purposes. Together with the fast development of ana-

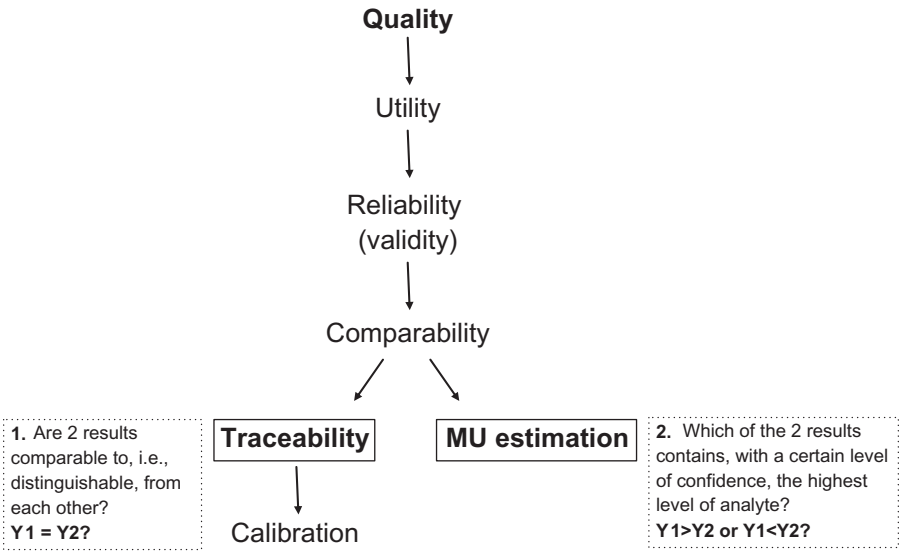


**FIGURE 1** Relationship between quality, traceability, and MU of results [3].

lytical methodologies, great importance is nowadays attached to the "quality" of the measurement data. *Quality* of analytical measurement data encompasses two essential criteria: *utility* and *reliability* (Figure 1) [1]. Utility means that analytical results must allow reliable decision making. A key aspect of reliability or *validity* of results is that they are *comparable*, whatever their origin. Comparability between results in the strict sense is provided by *traceability* to appropriate standards. Traceability to common reference standards underlies the possibility of making a comparison (i.e., a distinction) between different results. If results are to be compared also in terms of their quantities or levels of analyte, additional information on the analytical result is needed: *measurement uncertainty*. Uncertainty of results arises from the combination of all uncertainties of the reference values (to which the results are traceable) and all additional uncertainties associated with the measurement procedure. Measurement uncertainty and traceability are related concepts both defining the quality of analytical data (Figure 1) [2, 3].

Quality of results reflects *adequacy* (or *inadequacy*) of a method in terms of the extent to which the method fulfills its requirements or is fit for its particular analytical purpose (see below). Quality is always a relative notion, referring to the requirements fixed beforehand on the basis of national or international regulations or customer needs [1, 4]. The need for reliability of analytical data is stressed by the fact that measurement results will be used and may form the basis for decision making. Unreliable results bring along a high risk for incorrect decisions and may lead to higher costs, health risks, illegal practices and so on. Imagine, for instance, the consequences if results are false positives or if the uncertainty is much larger than reported [1, 5, 6].

### 2.2 Role of Method Validation in Traceability and MU

An analysis is a complex multistage investigation of the property values of materials, being the identity and the concentration of a specific component in a specific sample

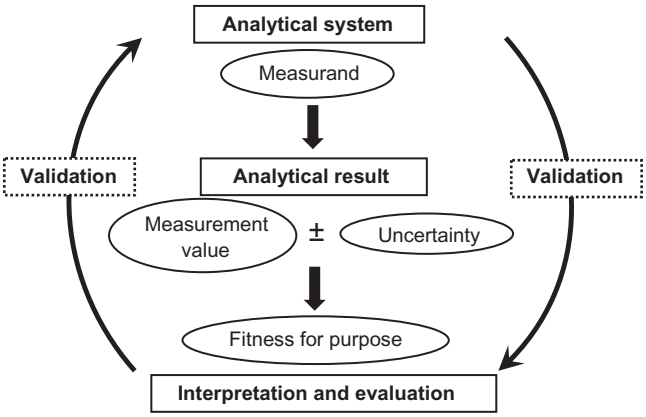


**FIGURE 2** Role of method validation in quality of analytical measurements. Validation is the process to demonstrate the fitness for purpose of the analytical system [4, 8, 14, 15].

material [2, 7]. Van Zoonen et al. [1] presented a chemical analysis as a cyclic process in which the final objective is the generation of chemical information. This integral process starts with defining the basic analytical problem (specifying the analytical requirement) and ends with evaluating and reporting the analytical result. Ideally, the last step provides an answer to the starting problem, stated by a client or based on regulatory requirements.

The process of providing an answer to a particular analytical problem is presented in Figure 2. The analytical system—which is "a defined method protocol, applicable to a specified type of test material and to a defined concentration rate of the analyte"—must be "fit for a particular analytical purpose" [4]. This analytical purpose reflects the achievement of analytical results with an acceptable standard of accuracy. Without a statement of uncertainty, a result cannot be interpreted and, as such, has no value [8]. A result must be expressed with its expanded uncertainty, which in general represents a 95% confidence interval around the result. The probability that the mean measurement value is included in the expanded uncertainty is 95%, provided that it is an unbiased value which is made traceable to an internationally recognized reference or standard. In this way, the establishment of traceability and the calculation of MU are linked to each other. Before MU is estimated, it must be demonstrated that the result is traceable to a reference or standard which is assumed to represent the truth [9, 10].

Traceability and MU both form parts of the purpose of an analytical method. Validation plays an important role here, in the sense that it "confirms the fitness-for-purpose of a particular analytical method" [4]. The ISO definition of validation is "confirmation by examination and provision of objective evidence that the particular requirements of a specified intended use are fulfilled" [7]. Validation is the tool used to demonstrate that a specific analytical method actually measures what it is intended to measure and thus is suitable for its intended purpose [2, 11]. In Section 3, the classical method validation approach is described based on the evaluation of a number of method performance parameters. Summarized, the criteria-based validation process consists of precision and bias studies, a check for

specificity/selectivity, a linearity check, robustness studies, and eventually, based on the practical requirements of the method, an assessment of the limits of detection and/or quantification. The objective of validation is to verify that the measurement conditions and the equation used to calculate the final result include all the influences that will affect the final result. Validation measures the different effects throughout the whole analytical system which have an influence on the result and ensures that no other effects have to be taken into account. A specificity test ensures that the method responds to the specific analyte of interest only, and not to other interferents or contaminants. A linearity check verifies that the supposed relationship between the signal and units used for the analyte may actually be used. A bias study is a certified reference material check and as such demonstrates that the method is not significantly biased, and precision and robustness studies cover the effects of variability in conditions, operators, equipment, and time.

The roles of method validation in the achievement of reliable results are (1) to include all possible effects or factors of influence on the final result, (2) to make them traceable to stated references [reference methods, reference materials, or International System of Units (SI)], and (3) to know the uncertainties associated with each of these effects and with the references. Validation is thus a tool to establish traceability to these references [2, 4]. In this context, it is important to see the difference between traceability and accuracy. A method which is accurate, in terms of "true" (i.e., approximating the true value), is always traceable to what is considered to be the true value. The opposite however is not correct. A method that is traceable to a stated reference is not necessarely true (accurate). Errors can still occur in this method, depending on the reference [12].

Analytical method validation forms the first level of QA in the laboratory. Analytical quality assurance (AQA) is the complete set of measures a laboratory must undertake to ensure that it is able to achieve high-quality data continuously. Besides the use of validation and/or standardized methods, these measures are effective IQC procedures (use of reference materials, control charts, etc.), with participation in proficiency testing schemes and accreditation to an international standard, normally ISO/IEC 17025 [4]. Method validation and the different aspects of QA form the subject of Section 3.

### 2.3 Guidelines on Traceability and Uncertainty of Results

Table 1 shows an overview of prominent guiding institutions and their guidelines on traceability, MU, and related topics. In Europe, a leading role is played by Eurachem, a working group on analytical chemistry centralized at and originating from the United Kingdom's LGC (Laboratory of the Government's Chemists). Basic references are CITAC/Eurachem guides on quality in analytical chemistry [2] and traceability in chemical measurement [3] and a Eurachem guide on MU [13, 14]. Eurachem also published guides on related topics such as reference materials [7] and method validation [15].

On the international level, relevant standards are available from the IUPAC, ISO, and AOAC International [4, 8, 16, 17] and from the Codex Alimentarius' working group CCMAS [18–21]. Other helping guides have been published by the EAL [22] and ILAC [23] (see Table 1 for explanation of abbreviations).

**TABLE 1 Overview of European and International Guiding Institutions and Regulatory Bodies with Guidelines and Standards on Traceability, MU and Related Topics**

| Body | Full Name | Guidance On | References |
|---|---|---|---|
| Eurachem/CITAC | A Focus for Analytical Chemistry in Europe/ Cooperation on International Traceability in Analytical Chemistry | Traceability, MU, reference materials, validation | 2, 3 13, 14, 7, 15 |
| IUPAC, ISO, and AOAC International | International Union of Pure and Applied Chemistry, International Organisation for Standardization, and Association of Official Analytical Chemists | MU | 4, 8, 16, 17 |
| FAO/WHO: Codex/CCMAS | Food and Agricultural Organization /World Health Organisation: Codex Committee on Methods of Analysis and Sampling | MU | 18–21 |
| EAL | European Cooperation for Accreditation | MU | 22 |
| ILAC | International Laboratory Accreditation Cooperation | MU | 23 |

## 2.4 Concept of Traceability

***Definitions*** Traceability is a relatively new term, gaining more and more attention in analytical measurement sciences. Traceability can be assigned to different aspects related to a measurement: traceability of a result, method, procedure, laboratory, product, material, equipment, and son on. As such, there is no single definition of traceability. Before exploring the different concepts of traceability, we can look to a more general, extended meaning. According to Valcarcél and Rios [24], the basic meaning of traceability integrates (1) the establishment of one or more relationships to well-stated references or standards and (2) the documented "history" of a product or a system. These two parts of the basic meaning can be found again when defining traceability as a property or a characteristic of different analytical facets. The different traceability concepts are shown in Figure 3.

The most obvious approach for traceability is a "property of the result of a measurement or the value of a standard whereby it can be related to stated references, usually national or international standards, through an unbroken chain of comparisons all having stated uncertainties." [2] The different elements and the practical use of this basic definition will be explained further below under Traceability in Practice.

Traceability of a result is related to traceability of a method, which in turn is linked to traceability of standards and traceability of the equipment used in the analytical procedure (Figure 3). A method is called traceable when it produces results (with their uncertainties) which are characterized by a defined traceability to well-stated references [24]. Walsh [25] defines traceable methods as "validated official or standard methods or validated methods which contain uncertainty state-

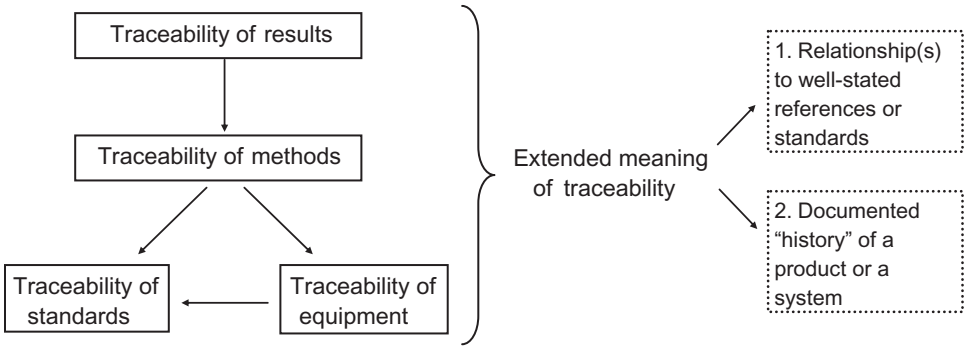


**FIGURE 3** Different concepts and extended meaning of traceability [24].

ments and which are embedded into a quality system and anchored to a common reference point." Traceability among standards is considered the most relevant basis for traceability of results [26], as shown in Figure 3. Traceability of equipment is defined as "the detailed, timely, and customised recording of installation, malfunctioning and repairs, periodic calibration and corrections (if needed), hours of use, samples processed, standard used, etc., in such a way that all questions (what?, how?, who?, etc.) should have a detailed answer in the pertinent documents" [24]. Calibration is the set of operations used to establish the relationship between values shown by a measuring instrument and the values of measurement standards. By doing so, the results of measurements are related to and thus made traceable to values of standards or references. In practice, calibration is performed by measuring samples with known amounts of analyte, such as certified reference materials, and monitoring the measurement response [2, 3]. These definitions confirm the links between traceability of equipment, standards, and results (Figure 3).

***Traceability in Practice*** The practical establishment of traceability is based on a step-by-step implementation of the definition. The ISO definition of traceability, originating from metrology (see above), can be translated into three basic steps:

1. Establish one or more links to well-stated references,
2. through an unbroken chain of comparisons and
3. estimate all uncertainties associated with those comparisons [12, 24, 27].

This definition is very much in line with the more practical definition described by Eurachem/CITAC [3]. Their procedure for traceability consists of the following steps [3]:

1. Specifying the measurand, scope of measurement, and required uncertainty
2. Choosing the method of measurement
3. Validating the method of measurement
4. Identifying/quantifying all influences which will affect the result
5. Choosing appropriate references
6. Estimating the uncertainty components associated with all influences and references.

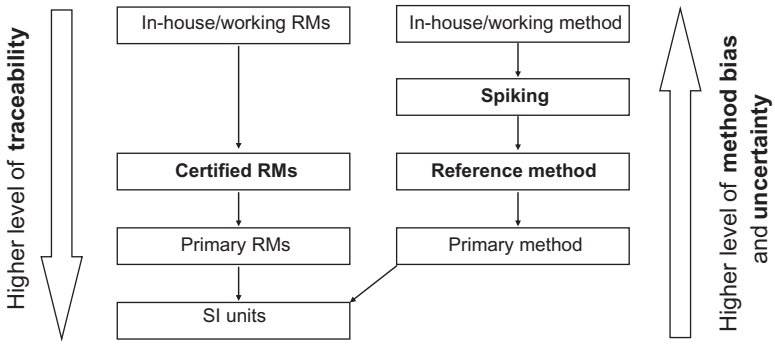


**FIGURE 4** Traceability chain and relationship between traceability and uncertainty of measurements. The three possibilities for establishing traceability referred to in Figure 5 are indicated in bold [25, 28].

The key principle in both approaches is that relationships to stated references are established and that this is done through an unbroken chain of comparisons. Practically, this means that the analytical procedure is first described as a chain or a flow diagram (step 2 in ISO definition; steps 1 and 2 in Eurachem/CITAC definition). The word *unbroken* means that there is no loss of information when considering the different steps in the analytical procedure leading to the measurement result. Each step in the procedure then needs to be linked to a reference method, a reference material, or a SI unit (step 1 in ISO definition; step 5 in Eurachem/CITAC definition) [12, 24, 27]. Figure 4 depicts the successive classes of stated references (materials or methods) in a so-called traceability chain. Establishing traceability through a traceability chain brings along a certain level of uncertainty, called *calibration uncertainty* or *traceability uncertainty* (see also below) [3, 7, 28]. This brings us then to the third key element when applying the traceability definitions: the stated uncertainties. Each step in the traceability chain, with the uncertainty components of all the stated references, will contribute to the measurement result and thus to the uncertainty associated with it. Uncertainty components must thus be estimated at each step in the analytical process (step 3 in ISO definition; step 6 in Eurachem/CITAC definition). As described above, validation is a tool to identify all possible effects or factors within the analytical procedure which can influence the final result. As such, steps 3 and 4 in the Eurachem/CITAC definition are additional steps which can be very helpful in the establishment of traceability [3].

Examples of how traceability is established in practice can be found in the literature. In 2004, Trends in Analytical Chemistry published a special issue on challenges for achieving traceability of environmental measurements (volume 23, 2004). This issue contains a lot of up-to-date information and practical examples in the particular domain of environmental analysis. Many authors have reported on the most important and most difficult step in establishing traceability: the selection of stated references or standards. For different stages in the traceability chain shown in Figure 4, descriptions and examples are given by Quevauviller and Donard [27], Charlet and Marschal [29], and Segura et al. [30]. Pan [28], Förstner [31], and Theocharopoulos et al. [32] applied the ISO definition for establishing traceability in different types of environmental methods of analysis. A similar approach was followed by

Sabé and Rauret [33] and Drolc et al. [34], however based upon the Eurachem/CITAC guide on traceability [3]. In their examples, all authors take into account specific steps or influences in the analytical procedure which can lead to a "broken" chain of comparisons, such as sampling and sample treatment or preparation steps. Some authors have reported on uncertainties associated with sampling in particular [35, 36].

### 2.5 Concept of MU

Measurement uncertainty is the most important criterium in both method validation and IQC. It is defined as "a parameter, associated with the result of a measurement, that characterizes the dispersion of the values that could reasonably be attributed to the measurand" [14]. The measurand refers to the particular quantity or the concentration of the analyte being measured. The parameter can be a standard deviation or the width of a confidence interval [14, 37]. This confidence interval represents the interval on the measurement scale within which the true value lies with a specified probability, given that all sources of error are taken into account [37]. Within this interval, the result is regarded as being accurate, that is, precise and true [11].

It cannot be overemphasized that MU is different from error. The error of an individual analytical result, the difference between the result and the true value of the measurand, is always a single value [38]. Part of the value of a known error, the systematic error, can be used to correct a result. This means that after correction the result of an analysis may be very close to the true value. However, the uncertainty of the measurement may still be very large because there is doubt or limited knowledge about how close the result is to the value. Uncertainty is expressed as a range and applies to an analytical procedure and a specific sample type but to different determinations and thus measurement results. The value of the uncertainty cannot be used to correct a measurement result.

The error of an analytical result is related to the (in)accuracy of an analytical method and consists of a systematic component and a random component [14]. Precision and bias studies form the basis for evaluation of the accuracy of an analytical method [18]. The accuracy of results only relates to the fitness for purpose of an analytical system assessed by method validation. Reliability of results however has to do with more than method validation alone. MU is more than just a single-figure expression of accuracy. It covers all sources of errors which are relevant for all analyte concentration levels. MU is a key indicator of both fitness for purpose and reliability of results, binding together the ideas of fitness for purpose and quality control (QC) and thus covering the whole QA system [4, 37].

The MU of an analytical procedure is thus derived from but is different from the error of a single analytical result. The deviation of the measurement result from the true value comprises a number of systematic and random errors, as shown in Figure 5. Each of these error components adds its own uncertainty to the total uncertainty budget of the analytical procedure. Therefore, the different error components are referred to as *sources of uncertainty*. Depending on the sources of uncertainty taken into account and thus on the conditions of the measurement, the overall MU will be different and another definition of MU will apply. This means that there is no single, straighforward definition of MU. It is rather a concept the interpretation of

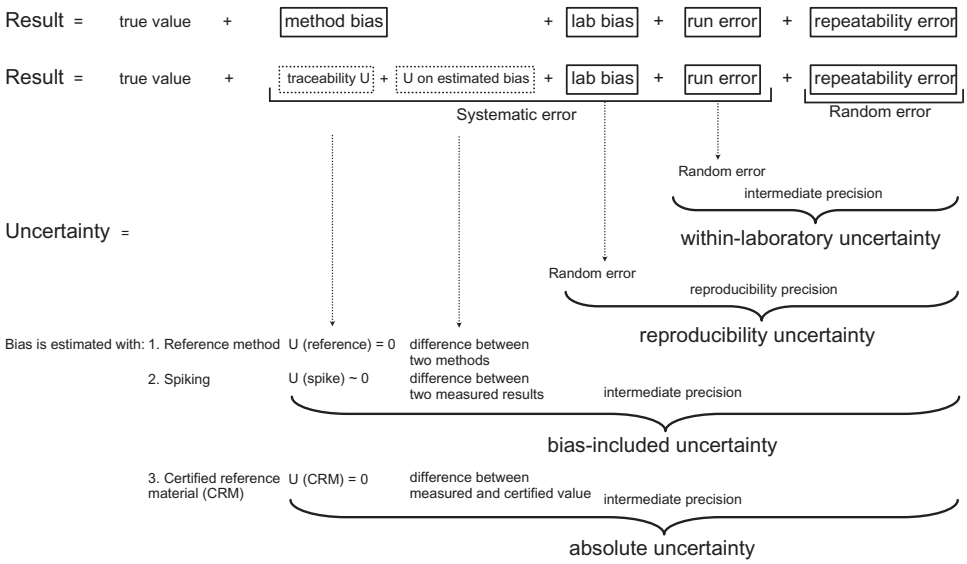


**FIGURE 5** Composition of error of analytical result related to MU. Different operational definitions of MU [10]. At the bottom on the left are the three possibilities for establishing traceability (see also Figure 4).

which changes according to the measurement conditions and to the reference to which the result is traceable [10]. The different definitions of MU form the subject of the next topic.

## 2.6 Different Operational Definitions of MU

As illustrated in Figure 5, the error of an analytical result for a specified analyte concentration is composed of different error components, forming together the "ladder of errors":

- The *method bias* is a systematic error associated with the method as such.
- The *laboratory bias* is either a systematic error—if the laboratory is considered on its own—or a random error—if the laboratory is considered as one of a group, as is the case in interlaboratory studies.
- The *run error* is seen as a systematic error for one run and as a random variation over several runs performed intralaboratory.
- The *repeatability error* is a random error due to the replicate measurements performed within a single run [10].

As the considered error only applies for a specified concentration of analyte isolated from a specified type of sample or matrix, *sampling errors* and *matrix variation effects* are not included here [4].

The more traditional distinction of error components is between *random errors* and *systematic errors*. In this classical approach, random errors are generally referred to as "precision" (repeatability, intermediate precision, and reproducibility), while systematic errors are typically attributed to the uncertainty on the bias estimate and

calibration uncertainty. To this classification, other uncertainty contributions are added, such as sampling effects, matrix effects, and uncertainties associated with certain assumptions which underlie the measurement method and/or the calculation equation [2].

As mentioned above, each of these error components is a potential source of uncertainty. Depending on the conditions under which the analysis is performed, different sources of uncertainty contribute to the overall uncertainty. Hund et al. [10] introduced different operational definitions of uncertainty, according to the number and type of uncertainty sources considered (Figure 5):

- *Within-laboratory uncertainty* is derived from intermediate precision and includes only the repeatability error and the run error.
- *Reproducibility uncertainty* is derived from reproducibility precision (interlaboratory tests) and counts for the repeatability error, run, and laboratory effects.
- *Bias-included uncertainty* and *absolute uncertainty* additionally take into account the method bias. Method bias is the most important source of uncertainty because it refers to a reference or a standard to which the method is considered to be traceable. If the working method is not a primary method—which is traceable to SI units (see Figure 4)—the method is always compared to another, reference method or is applied on appropriate certified reference materials. This reference or standard is to be considered when the uncertainty associated with the method bias is estimated. In addition to the uncertainty associated with this reference or standard, there is the uncertainty on the estimated bias (Figure 5). The different possibilities of bias estimation—and thus of traceability—are depicted in Figure 5. If the method is compared to a reference method, the uncertainty associated with this reference method is considered negligible and the bias is estimated as the difference between the two methods (case 1 in Figure 5). If there is no method to compare with, bias can be estimated by spiking samples and assessing the difference between the spiked sample and the measured sample. In this case also, the uncertainty on the spike will approximate zero (case 2) and the only method bias is the difference between the measured sample and the spiked sample. Absolute uncertainty can only be estimated if certified reference materials are used (case 3). Only in this case, full traceability to SI units can be guaranteed [10].

### 2.7 Approaches to Establish MU

In general, to estimate the overall uncertainty on a particular result, it is necessary to know (1) all uncertainties arising from the measurement procedure itself and (2) all uncertainties associated with the references or standards to which the analytical results are made traceable [3]. Different approaches exist for the estimation of overall MU, as reviewed by several authors [9, 10, 39] and as summarized in Table 2.

The most well-known and traditional approach is based on identifying, quantifying, and combining all individual contributions to uncertainty. In this "bottom-up approach," the overall uncertainty is derived from the uncertainties of the individual components. The component-by-component assessment of MU was originally

**TABLE 2 Approaches for Estimating MU**

| Name of Approach | Basic Principle | Strengths | Weaknesses | Reference |
|---|---|---|---|---|
| Bottom-up, error budget, error propagation or component-by-component approach | Identification, quantification, and combination of all sources of uncertainly | Holistic: all important sources of error should be included | Complex, expensive, time consuming | Eurachem [13], ISO [16] |
| Fitness-for-purpose approach | Establishment of a fitness function $u = f(c)$ based mainly on precision and bias studies | Simple, MU can be assessed for different concentrations | Some sources of uncertainly may be overlooked | Codex Alimentarius/ CCMAS [18–21] |
| Top-down approach | Based on data obtained from interlaboratory studies (precision) | MU can be assessed for different concentrations | Some sources of uncertainly may be overlooked; only if data on collaborative studies are available | Analytical Methods Committee [37] |
| Validation-based approach | Based on inter- or intralaboratory validation studies (precision, trueness, robustness) | An extension of validation work, no extra work needed | Some sources of uncertainly may be overlooked | Eurachem [14], Barwick and Ellison [47] |
| Robustness-based approach | Based on robustness tests as intralaboratory simulations of interlaboratory studies | Simple, time-efficient approach | Some sources of uncertainly may be overlooked; method must first show to be robust | Hund et al. [39] |

developed for physical measurements and adopted by Eurachem for chemical measurements [13]. Because of its complexity, however, this methodology requires significant costs in time and effort and has never found widespread applications.

A simplified approach to assess MU is the *fitness-for-purpose approach*, defining a single parameter called the fitness function. This fitness function has the form of an algebraic expression $u = f(c)$ and describes the relationship between the MU and the concentration of the analyte. For example, $u = 0.05c$ means that the MU is 5% of the concentration. Calculation of the MU will hereby rely on data obtained by evaluating individual method performance characteristics, mainly on repeatability and reproducibility precision, and preferably also on bias [21, 40, 41]. This approach can more or less be seen as a simplification of the step-by-step protocol for testing the MU, as described by Eurachem [14].

Although MU comprises more than systematic and random errors, it can be estimated from method validation data. Data from method performance studies can give all or nearly all information required to evaluate the uncertainty [2, 4, 18, 37]. This includes the use of data from in-house and collaborative validation studies (typically precision data), proficiency testing schemes (typically bias data), or whichever QA data relevant for uncertainty. If such data are available and used to estimate the uncertainty, it is not necessary to estimate MU using the component-by-component approach [18, 19]. In particular, validation studies and QC measures are considered as highly relevant sources for MU estimation [42, 43]. Table 2 describes three methodologies for MU assessment which are based on validation data. In the Analytical Methods Committees top-down approach [37], the laboratory is seen from a higher level, as a member of a population of groups. As a consequence, systematic errors within one laboratory become random errors and the estimated uncertainty is the reproducibility uncertainty (Figure 5). Examples of MU estimation studies using data from collaborative ring trials are worked out by Dehouck et al. [44] and Maroto et al. [45, 46]. The other two approaches mentioned in Table 2 [14, 39, 47] make use of different method performance parameters. All three validation-based methodologies can be seen as more simple, time- and cost-efficient extensions of validation. However, important to note is that not all sources of uncertainty are covered by method performance data. Some sources that may need particular consideration in addition to the available data are sampling, pretreatment, method bias, variation in conditions, and changes in the sample matrix [14, 18, 41].

Many of those principles for MU estimation have been applied in a case study for toluene in groundwater performed by Armishaw [48]. His idea is that for a routine method which has been validated previously and which is performed in a laboratory where QC measures are in place, it is possible to estimate MU in a working afternoon. The key is to extract all relevant information from already available data: validation studies (bias/recovery and precision data), instrument calibration data [reference material (RM) uncertainties], information from regularly performed QC measures (replicate analyses and control samples), and the method procedure itself (sampling, homogeneity of samples, etc.). After identification of the components contributing to the uncertainty budget and assignment of the relevant sources of information, standard uncertainties were quantified and combined to form the combined standard uncertainty. The results for toluene in water were reported as $x \pm U$, where $U$ is the expanded uncertainty, obtained by multiplying

the combined standard uncertainty $u_c$ with a coverage factor of 2. In addition to this bottom-up MU estimation (approach 1), Armishaw reported the expanded uncertainty as a function of the concentration of toluene (approach 2). Finally, the authors compared this experimentally assessed MU with calculated MU values, based on (1) a within-laboratory reproducibility estimate, (2) proficiency test data, and (3) the models of Horwitz [49] and Thompson and Lowthian [50]. These models allow to calculate the percent relative standard deviation (RSD) values as a function of the analyte concentration. To obtain expanded uncertainty values, Armishaw multiplied the predicted standard deviation (SD) values from the models by 2 [48]. All three MU calculations are variants of validation-based approaches (approaches 3 and 4, Table 2).

### 2.8 Importance of Traceability and MU

The underlying motivation to establish traceability and MU is the need to make decisions based on the obtained analytical results or to be in compliance with regulatory limits (for quantitative determinations) [51]. MU is an essential feature of analytical results for different reasons. First, customers want to have an idea about the dispersion of results and about the comparability of results between different laboratories [43]. Any result must be accompanied by a statement of MU, so that the user of the result knows what level of confidence is associated with it [52]. The concepts of *comparability* and *reliability* of results have been discussed briefly in Section 1 and are presented in Figure 1. Second, MU is necessary because it demonstrates *traceability*. Before MU can be evaluated, traceability to stated references or standards must be established. Moser et al. [43] claim that traceability is proven by the appropriate use of reference materials or standards and by a full uncertainty budget. This brings us then to a third reason for MU estimation, which is the requirement to *know* the method, to understand the underlying principles and mechanisms of the measurement procedure. Lack of profound knowledge on the method itself will lead to certain unknown uncertainty contributions not being taken into account and thus to gaps in the uncertainty budget. MU can only be estimated if the method is well understood [43, 53].

Measurement uncertainty is increasingly gaining attention, in particular in the framework of accreditation. The new accreditation standard ISO/IEC 17025 [17], which has been in force from December 2002 on, contains clear requirements on the estimation of MU and when and how it should be stated in test reports. ISO/IEC 17025 requires MU to be reported when required by the client and when relevant to the application and interpretation of the measurement results in the framework of certain specifications or decision limits. The MU should be readily available and reported together with the result as $X \pm U$, where $U$ is the expanded uncertainty [17, 47, 51, 54]. Also Eurachem and CCMAS within the Codex Alimentarius deal with MU as a separate issue [14, 18–20]. Some even claim that MU will become the main unifying principle of analytical data quality [37].

### 2.9 Summary

Rather than on the used techniques and methodologies themselves, attention is nowadays paid to the quality and reliability of the final results. This is influenced by

a higher demand for regulatory compliance, a higher consciousness of the customer—the client wants to know the level of confidence of the reported result—and under impulse of new, more exigent European and international standards such as the ISO/IEC 17025 norm for laboratory accreditation. On the basis of quality and reliability of analytical data lies comparability of results between laboratories. In order for results to be comparable, they must be reported with a statement of MU and they must be traceable to common primary references. Methods must be validated to show that they actually measure what they are intended to measure—that they are fit for their specific purpose.

An analytical method is a complex, multistep process, starting with sampling and ending with the generation of a result. Although every method has its specific scope, application, and analytical requirement, the basic principles of QA are the same regardless of the type of method or sector of application. The information in this chapter is mainly taken from the analytical chemistry, but it applies to other sectors as well. The validation of analytical methods, the establishment of traceability of results, and the assessment of MU should be done in a uniform, harmonized way, conforming with internationally recognized standards from institutions such as Eurachem, IUPAC, or ISO.

This update on analytical quality issues a common understanding on the topics of method validation, traceability, and MU of measurements. The interrelationships between method validation and traceability and MU of results have been elucidated. Throughout the landscape of guidelines and standards, the most relevant information was selected, compiled, and summarized. Different approaches are discussed for establishing traceability and assessing MU of analytical methods in general. The importance of both concepts and the link with method validation and analytical quality assurance are highlighted.

## 3 METHOD VALIDATION AND QUALITY ASSURANCE

### 3.1 Role of Method Validation in AQA

The terms validation and QA are widely used. However, a lot of analysts and laboratories do not know the exact meaning—neither the difference nor the relationship between the two terms. Validating a method is investigating whether the analytical purpose of the method is achieved, which is obtaining analytical results with an acceptable uncertainty level [4]. Analytical method validation forms the first level of QA in the laboratory (Figure 6). AQA is the complete set of measures a laboratory must undertake to ensure that it is able to achieve high-quality data continuously. Besides the use of validation and/or standardized methods, these measures are effective IQC procedures (use of reference materials, control charts, etc.), participation in proficiency-testing schemes, and accreditation to an international standard, normally ISO/IEC 17025 [2, 4, 6].

The different levels in Figure 6 represent the different measures a laboratory must undertake in order to ensure that it is qualified and competent to perform analytical measurements which satisfy their agreed requirements. A laboratory must be capable of providing analytical data of the required quality. The "agreed requirement" of an analytical method and the "required quality" of an analytical result refer to the fitness for purpose of the method [4, 8, 15].

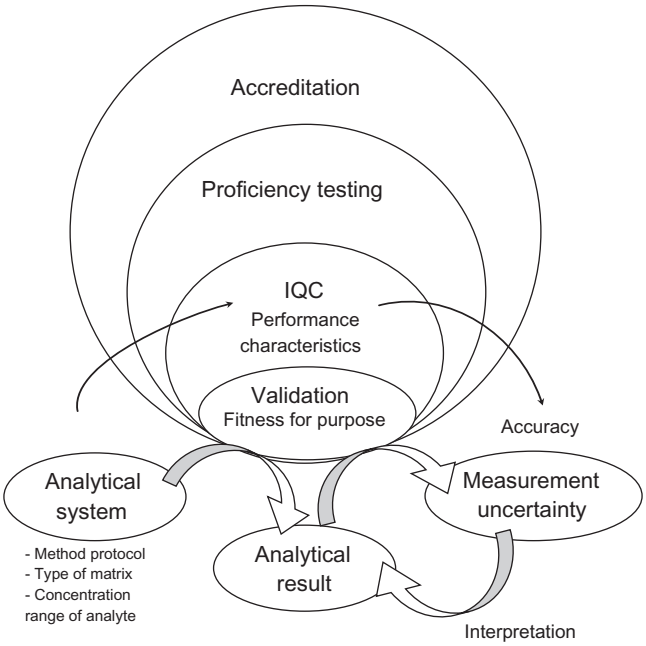


**FIGURE 6** Different levels of QA measurements for analytical chemistry and food laboratories [4, 8, 15].

The ISO definition of validation is "confirmation by examination and provision of objective evidence that the particular requirements of a specified intended use are fulfilled" [15]. Method validation is needed to "confirm the fitness for purpose of a particular analytical method," that is, to demonstrate that "a defined method protocol, applicable to a specified type of test material and to a defined concentration rate of the analyte"—the whole is called the "analytical system"—"is fit for a particular analytical purpose" [4]. This analytical purpose reflects the achievement of analytical results with an acceptable standard of accuracy. An analytical result must always be accompanied by an uncertainty statement, which determines the interpretation of the result (Figure 6). In other words, the interpretation and use of any measurement fully depend on the uncertainty (at a stated level of confidence) associated with it [8]. Validation is thus the tool used to demonstrate that a specific analytical method actually measures what it is intended to measure and thus is suitable for its intended purpose [11, 55, 56].

Validation is required in the first place for any new method. As the definition says, validation always concerns a particular analytical system. This means that for a particular type of material and a particular operating range of concentrations, the method must be able to solve a particular analytical problem [4]. As a consequence, *revalidation* is needed whenever any component of the analytical system is changed or if there are indications that the established method does not perform adequately anymore [15, 56, 57].

Method validation is closely related to method development. When a new method is being developed, some parameters are being evaluated already during the *development stage* while in fact this forms a part of the *validation stage* [15]. On the other hand, a validation study may indicate that a change in the method protocol is necessary, which may then require revalidation [58].

Before any method validation is started, the scope of validation must be fixed, comprising both the *analytical system* and the *analytical requirement*. A description of the analytical system includes the purpose and type of method, the type and concentration range of analyte(s) being measured, the types of material or matrices for which the method is applied, and a method protocol. On the basis of a good analysis lies a clear specification of the analytical requirement. The latter reflects the minimum fitness-for-purpose criteria or the different performance criteria the method must meet in order to solve the particular problem. For example, a minimum precision (RSD, see below) of 5% may be required or a limit of detection (LOD) of 0.1% (w/w) [2, 4, 15, 58]. The established criteria for performance characteristics form the basis of the final acceptability of analytical data and of the validated method [58].

Validation of a new analytical method is typically done at two levels. The first is the level of prevalidation, aiming at fixing the scope of the validation. The second level is an extensive, "full" validation performed through a collaborative trial or interlaboratory study. The objective of full validation, involving a minimum number of laboratories, is to demonstrate that the method performs as was stated after the prevalidation.

### 3.2 Guidelines and Guidance on AQA

As shown in Figure 6, using validated methods is the first level of QA required within a system of IQC. The latter is needed for participance in proficiency-testing schemes (PTSs), which in turn form a prerequisite for accreditation [4].

For the different levels of QA presented in Figure 6, guidelines and requirements are well described in detail by several regulatory bodies, standardization agencies, and working groups or committees. Just as for traceability and MU (Section 2), relevant guidelines are given in Table 3. Eurachem guides are published on quality in the laboratory in general [2], method validation [15], and proficiency testing [59]. A guideline on proficiency testing from the joint Eurachem–Eurolab–EA group is in addition available [60]. On the European level there is also the CEN, working through different technical committees and working groups on standardization of analytical methods for all sectors [61].

On the international level, we distinguish IUPAC, ISO, and the AOAC International. All three bodies develop validation and standardization frameworks for analytical chemistry. AOAC International introduced the AOAC Peer Verified Methods Program [62]. Different harmonized guidelines and protocols are developed by the IUPAC, ISO, and AOAC International together [4, 8, 63–67], in addition to a number of ISO standards [68–71]. The FDA, USP, and ICH developed guidelines specific for pharmaceutical and biotechnological methods [55, 72–74].

The international Codex Alimentarius Commission within the United Nations FAO/WHO Food Standards Program has a CCMAS. CCMAS works out criteria

**TABLE 3 Overview of European and International Regulatory Bodies and Their Guidelines and Standards on Different Aspects of Analytical QA**

| Body | Full Name | Guidance On | References |
|---|---|---|---|
| Eurachem/CITAC and EA | A Focus for Analytical Chemistry in Europe/Cooperation of International Traceability in Analytical Chemistry and European Cooperation for Accreditation | Method validation, Proficiency testing, QA, accreditation | 2, 15, 59–60 |
| CEN | European Committee for Normalization | Standardization | 61 |
| IUPAC, ISO, and AOAC International | International Union of Pure and Applied Chemistry, International Organisation for Standardization, Association of Official Chemists | Method validation, Standardization, internal quality control, proficiency testing, accreditation | 4, 8, 62–71 |
| FDA, USP, and ICH | U.S. Food and Drug Administration, U.S. Pharmacopeia, International Conferences on Harmonization | Method validation | 55, 72–74 |
| FAO/WHO: Codex/CCMAS | Food and Agricultural Organization/World Health Organisation: Codex Committee on Methods of Analysis and Sampling | Method validation | 21, 75–78 |
| ILAC | International Laboratory Accreditation Cooperation | Proficiency testing, accreditation | 79–81 |

for evaluating the acceptability of methods of analysis as well as guidelines on single-laboratory and interlaboratory validation of methods [75–78]. For single-laboratory validation, CCMAS defends the harmonized IUPAC guidelines [4]. On the international level, also ILAC provides guidelines on proficiency testing [79] and accreditation [80, 81] (Table 3).

### 3.3 Approaches for Evaluating Acceptable Methods of Analysis

The purpose of an analytical method is the deliverance of a qualitative and/or quantitative result with an acceptable uncertainty level. Therefore, theoretically, "validation" boils down to "measuring uncertainty". In practice, method validation is done by evaluating a series of method performance characteristics, such as precision, trueness, selectivity/specificity, linearity, operating range, recovery, LOD, limit of quantification (LOQ), sensitivity, ruggedness/robustness, and applicability. Calibration and traceability have been mentioned also as performance characteristics of a method [2, 4]. To these performance parameters, MU can be added, although MU is a key indicator for both fitness for purpose of a method and constant reliability of analytical results achieved in a laboratory (IQC). MU is a comprehensive parameter covering all sources of error and thus more than method validation alone.

In practice, data from method validation and collaborative studies form the basis for but are only a part of MU estimation. MU is thus more than just a "method performance parameter," as described extensively in Section 2. Over the years, the concept of MU has won attention in all analytical areas and this has led to two different approaches currently accepted and used for analytical method validation.

The traditional *criteria approach* is to identify specific performance parameters and to assign numeric values to these. These numeric values represent cutoff or threshold values the method parameters must meet in order for the method to be acceptable. The alternative approach is focused on fitness for purpose and MU. In this *fitness-for-purpose approach*, the overall MU is estimated as a function of the analyte concentration (see Section 2).

Generally, the criteria approach is used for rational methods, that is, methods where the measurement result can be obtained independently of the method used. Opposite to rational methods, there are empirical methods in which the measured value depends on the method used. For empirical methods, the criteria approach cannot simply be applied. Instead, precision data from collaborative studies are normally used as a basis for MU estimation and validation [75, 76].

Validation is needed to demonstrate that the analytical method complies with established criteria for different performance characteristics [82]. When these different characteristics are being evaluated individually, this is generally done for the analytical method as such—where the input is the purified or isolated analyte and the output is the analytical result. However, MU covers the whole analytical procedure, starting from the original sample lot. The assessment of MU (see Section 2) is in line with the so-called modular validation approach. Modular validation refers to the "modularity" of an analytical procedure divided up into several sequential steps needed to analyze the material. These may be sample preparation, analyte extraction, and analyte determination (Figure 7). Each step in the procedure can be seen as an analytical system and can thus be validated separately and combined

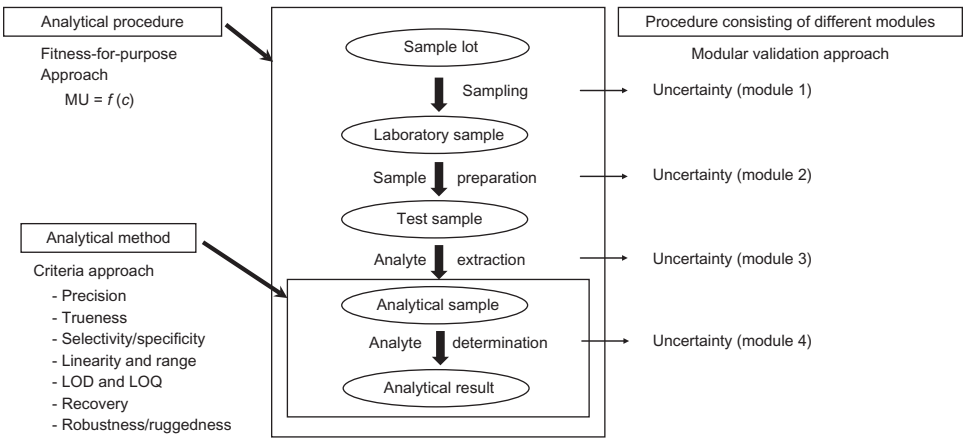


**FIGURE 7** Schematic representation of "analytical method" in "analytical procedure" and different approaches for validation: $f(c)$ = function of concentration.

later on with other "modules" in a flexible way. Modular validation is thus a stepwise validation of a whole procedure, taking into consideration all possible difficulties or uncertainty factors at each level in the procedure. The concept of modular validation originates from the domain of predictive microbiology and is now being proposed for methods of analysis for genetically modified organisms (GMOs) [83]. The relationship between the three validation approaches described above is shown in Figure 7.

### 3.4 Method Performance Characteristics and Criteria Approach

***Extent of Validation Depends on Type of Method*** On the one hand, the extent of validation and the choice of performance parameters to be evaluated depend on the status and experience of the analytical method. On the other hand, the validation plan is determined by the analytical requirement(s) as defined on the basis of customer needs or as laid down in regulations. When the method has been fully validated according to an international protocol [63, 68] before, the laboratory does not need to conduct extensive in-house validation studies. It must only verify that it can achieve the same performance characteristics as outlined in the collaborative study. As a minimum, precision, bias, linearity, and ruggedness studies should be undertaken. Similar limited validation is required in cases where it concerns a fully validated method which is applied to a new matrix, a well-established but noncollaboratively studied method, and a scientifically published method with characteristics given. More profound validation is needed for methods published as such in the literature, without any characteristic given, and for methods developed in-house [84].

Which performance criteria have to be evaluated depends also on the purpose of the method. Different ICH/USP guidelines are set up for (1) identification tests, (2) impurity tests, and (3) assay tests. An identification test ensures the identity of an analyte in a sample by comparison to a known reference material. An impurity test is intended to confirm the identity of (limit impurity test) or to accurately quantify (quantitative impurity test) an impurity, defined as an entity which may normally not be present. An assay test finally implies the major component or active ingredient in a sample and quantifies the drug substance as such as a whole or the drug substance in a drug product.

For an assay test, where the major component or active ingredient is supposed to be present at high levels, other criteria should be investigated than for an impurity test. The same is valid for quantitative tests versus identification and limit impurity tests (Table 4) [55, 56, 72, 85].

The literature gives a wide range of practical guidelines for the evaluation of method performance characteristics [58]. Besides the diversity of approaches, also the terminology and way of reporting results vary widely. Differences may occur depending on the purpose and the application field of the method, and validation studies may become more difficult as the complexity of the analysis increases [86]. In what follows, terms and formulas are taken from the accepted IUPAC nomenclature for the presentation of results of chemical analysis [66]. For each validation parameter, definitions, ways of expression, determination guidelines, and acceptance criteria are reported in Table 5.

**TABLE 4 Criteria to Establish for Different Categories of Methods of Analysis**

| Method Performance Parameter | Identification Test | Impurity Test: Limit Impurity Test | Impurity Test: Quantitative Impurity Test | Assay Test |
|---|---|---|---|---|
| Precision | –[a] | – | + | + |
| Trueness | – | –[a] | + | + |
| Specificity | + | + | + | + |
| LOD | –[a] | + | – | – |
| LOQ | –[a] | – | + | – |
| Linearity | –[a] | – | + | + |
| Range | –[a] | –[a] | + | + |
| Ruggedness | + | + | + | + |

*Source*: From ref. 56.

[a]May be performed.

***Accuracy*** Precision and bias studies, which form a part of the MU estimate, are the most important validation criteria.

*Precision* measures are divided into (1) repeatability precision measures $s$ or SD ($s_r$ or $SD_r$) and RSD ($RSD_r$), (2) intralaboratory reproducibility precision or intermediate-precision measures SD and RSD, and (3) interlaboratory reproducibility precision $s$ or SD ($s_R$ or $SD_R$) and RSD ($RSD_R$) [66].

Besides standard deviations and coefficients of variation, repeatability/reproducibility values or limits ($r$, $R$) are additional parameters of high value in the assessment of precision (see formulas in Table 5). These criteria mean that the absolute variation between two independent results—obtained within the same laboratory respectively between different laboratories—may only exceed the values of $r$ and $R$ in a maximum 5% of the cases [2]. Another measure for the precision is the confidence interval, in which all measurements fall with a certain propability or confidence level $1 - \alpha$ ($\alpha$ is often 0.05 giving a probability here of 95%) [66].

Calculated repeatability, intermediate precision, and reproducibility values can be compared with those of existing methods. If there are no methods with which to compare the precision parameters, theoretical relative reproducibility and repeatability standard deviations can be calculated from the Horwitz equation and the Horrat value (Table 5). Horwitz RSD values are reported in Table 6. Higher variability is expected as the analyte levels approach the detection limit (see below). Next to the Horwitz equation, the AOAC's Peer Verified Program proposes its own levels of acceptability of %RSD as a function of analyte concentration level [56, 72].

Precision data can be documented in bar charts or control charts such as Shewhart control charts (see the discussion of internal quality control in Section 3.5). Bar charts plot %RSD values with their corresponding confidence interval. Control charts plot the individual measurement results and the means of sets of measurements with their confidence level (or with horizontal lines representing "limits," see below) as a function of the measurement number and the run number, respectively [15, 55, 56, 58, 72, 85].

**TABLE 5 Summary of Method Performance Parameters**

| Parameter | Definition | Expression | Requirements | Practical Assessment |
|---|---|---|---|---|
| Accuracy | Closeness of agreement between test result and accepted reference value | | | |
| 1. Precision | Closeness of agreement between independent test results obtained under stipulated conditions | • Standard deviation $s$ or SD<br>• Relative standard deviation $s_{rel}$ or RSD<br>• % coefficient of variation %CV or %<br>• RSD = 100 × SD/$\bar{x}$<br>• repeatability limit $r$ = 2.83 × $SD_r$ and reproducibility limit $R$ = 2.83 × $SD_R$<br>• confidence interval CI = $\bar{x}$ ± $C$, where $C = \frac{s \times t_{p,v}}{\sqrt{n}}$ | • $RSD_r$ = 0.5 at 0.6 times theoretical values determined by Horwitz function = 2 exp (1–0.5 log/$C$) with $C$ = analyte concentration in decimal fraction<br>• $RSD_R$ = 0.5 at 2 times theoretical values determined by Horwitz function<br>• Horrat value = $RSD_R$ (trial)/$RSD_R$(Horwitz) ≤ 2<br>• $RSD_r$ and $RSD_R$ values according to AOAC peer-verified program | • Minimum of three concentration levels over full range of analytical method (~at LOD, in middle, high)<br>• Minimum of three repeats per concentration level<br>• Calculate repeatability precision $SD_r$, $RSD_r$, $r$ = 2.8 × $SD_r$, $C$, CI<br>• Calculate intermediate precision $SD_{int}$, $RSD_{int}$, $r$ = 2.8 × $SD_{int}$, $C$, CI<br>• Calculate reproducibility precision $SD_R$, $RSD_R$, $r$ = 2.8 × $SD_R$, $C$, CI<br>• Document in bar chart or control chart |
| 1.1. Repeatability precision (intrarun precision) | *Precision under conditions where independent test results are obtained with same method on identical test items in same laboratory by same operator using same equipment within short intervals of time* | | | |
| 1.2. Intermediate precision (interrun precision) | *Precision under conditions where independent test results are obtained with same method on identical test items in same laboratory but by different analysts using different equipment over extended period of time* | | | |
| 1.3. Reproducibility precision (interlaboratory precision) | *Precision under conditions where test results are obtained with same method on identical test items in different laboratories with different operators using different equipment* | | | |

| | | | | |
|---|---|---|---|---|
| 2. Trueness | Closeness of agreement between expectation of test result (expected mean value) and accepted reference value (true value) | *If CRMs are used*:<br>Bias = % error: difference between measured value and true value<br>$Z$ score: difference between measured value and certified reference value %<br>$$z = \frac{X_{\text{found}} - X_{\text{certified}}}{\sqrt{\frac{\text{SD}_{\text{found}}}{n_{\text{found}}} + \left(\frac{\text{SD}_{\text{certified}}}{n_{\text{certified}}}\right)^2}} = \frac{X_{\text{found}} - X_{\text{certified}}}{\sqrt{\frac{\text{SD}_{\text{found}}}{n_{\text{found}}} + \left(\frac{\text{CI}}{2}\right)^2}}$$<br>*If no CRMs are available*:<br>Percent recovery of known, spiked amount of analyte | *General*: compare results with second, validated reference method<br>*If CRMs are used*:<br>$Z$ score ≤ \|2\|<br>*If spiking method is used*:<br>Percent recovery according to USP/ICH guidelines (depending on analyte concentration level) | • Minimum of three concentration levels over full range of analytical method (~at LOD, in middle, high)<br>• Minimum of three at 10 repeats per concentration level<br>• *Use of CRMs*: calculate bias (% error) and/or $Z$ score<br>• *No CRMs*: spike typical matrix and blank sample with known amount of analyte; calculate %recovery = 100 × [$\text{spike}_{\text{matrix sample}}/\text{spike}_{\text{blanco}}$] |
| Recovery | Fraction of analyte added to test sample (fortified or spiked) prior to analysis, which is measured by method | Percent recovery of known, spiked amount of analyte | Percent recovery according to USP/ICH guidelines (depending on analyte concentration level) | • Minimum of six repeats of matrix blanks or samples unfortified and fortified with analyte at different concentrations<br>• Calculate %recovery = 100 × [$\text{spike}_{\text{matrix sample}}/\text{spike}_{\text{blanco}}$] |

**TABLE 5** ***Continued***

| Parameter | Definition | Expression | Requirements | Practical Assessment |
|---|---|---|---|---|
| Specificity | Ability of method to determine accurately and specifically analyte of interest in presence of other components in sample matrix (that may be expected to be present in sample matrix) under stated conditions of test (specificity = 100% selectivity) | *Note*: cannot be expressed, must be demonstrated; depends on type and purpose of method<br>*For identification tests*:<br>Percent correct classification of<br>• Non-analyte-containing samples as negatives<br>• Analyte-containing samples as positives<br>Ability to discriminate with compounds of closely related structures (negative results)<br>*For quantitative tests*:<br>Percent recovery of samples, spiked with possible interferants | *For identification tests*:<br>Percent correct classification~100<br>*For quantitative tests*:<br>Percent recovery according to USP/ICH guidelines (depending on analyte concentration level) | *Identification tests*:<br>• Determine percentage of false positives for minimum number of blank samples<br>• Determine percentage of false negatives for minimum number of positive samples (RMs)<br>• Test minimum number of structurally similar or closely related samples which must be negative<br>*Impurity and assay tests*:<br>• Spike different samples with possible interferents and calculate percentage of recovery<br>• If no interfering compounds are available, compare with second method |

| | | | |
|---|---|---|---|
| Limit of detection (LOD) | Lowest concentration or amount of analyte:<br>• that can be reliably distinguished from zero or<br>• that can be detected/measured with reasonable statistical certainty | • *Note*: LOD is expressed as concentration or amount of analyte, which is derived from measured signal<br>• Lowest signal: $x_L = x_{bl} + 3s_{bl}$<br>• Lowest concentration/amount: LOD = $q_L(c_L) = x_L/S = 3s_{bl}/S$<br>• *Note*: signal of three times $SD_{blanco}$ corresponds to %RSD of 33% | • Measure, each once, a minimum of either 10 independent sample blanks ($x_L = x_{bl} + 3s_{bl}$) or 10 independent sample blanks fortified at lowest acceptable concentration ($x_L = 3s_x$)<br>• Choose number of low concentration levels near the expected LOD<br>• Set up calibration curve: signal $x = A$ concentration $+ B$ and thus $x_L = A$LOD<br>• Determine slope $A$ and calculate LOD = $x_L$/A = $[x_{bl} + 3s_{x,bl}]$/A |
| Limit of quantitation (LOQ) | Lowest concentration or amount of analyte that can be determined quantitatively with acceptable level of repeatability, precision, and trueness | • *Note*: LOQ is expressed as concentration or amount of analyte, which is derived from measured signal<br>• Lowest signal: $x_L = x_{bl} + 10s_{bl}$<br>• Lowest concentration/amount: LOQ = $q_L(c_L) = x_L/S = 10s_{bl}/S$<br>• *Note*: a signal of 10 times $SD_{blanco}$ corresponds to % RSD of 10% | • Measure minimum of 10 independent sample blanks each once ($x_L = x_{bl} + 10s_{bl}$)<br>• Choose number of low concentration levels near calculated LOD<br>• Set up calibration curve: signal $x = A$ concentration $+ B$ and thus $x_L = A$LOQ<br>• Determine slope $A$ and calculate LOQ = $x_L$/A = $[x_{bl} + 10s_{bl}]$/A |

**TABLE 5** ***Continued***

| Parameter | Definition | Expression | Requirements | Practical Assessment |
|---|---|---|---|---|
| Linearity<br>Linear range = working range = linearity limits | Ability of method to obtain test results proportional to concentration of analyte (within given range)<br>Range of concentrations/amounts of analyte over which<br>• method gives test results proportional to concentration of analyte or<br>• linear calibration model can be applied with known confidence level | • *Note*: *linearity* cannot be expressed, must be demonstrated<br>• *Range*: a range of concentrations between lower and upper linearity limits | *For assay tests*:<br>range = 80–120% of analyte level<br>*For impurity tests*:<br>range = LOQ 50–150% or 0–150% according to USP/ICH and IUPAC guidelines<br>Lower linearity limit = LOQ | • First rough estimation of linear range: measure blank + minimum of six independent concentration levels, set up calibration curve and estimate linear range visually<br>• Choose CRMs or spiked samples of minimum of six different concentrations within estimated linear range<br>• Calculate residual *Y* values from calibration curve and plot them as funcion of concentration<br>• If randomly distributed → linearity; if *systematic* trends → non-linearity<br>• Calculate relative signals = signal/concentration and plot them as function of concentration<br>• If horizontal line → linearity; linearity limits correspond to 95% and 105% relative signal values →*range* |

| | | | |
|---|---|---|---|
| Ruggedness | The (intralaboratory tested) behavior of analytical process when small changes in environmental and/or operating conditions are made (generally used term) | *Ruggedness*: measure for variability (reproducibility of results obtained under variety of conditions); expressed as %RSD (interlaboratory) | • Evaluate effect of different small changes in paramters (in days, instruments, analysts, reagents, material, amount of sample material used, etc.) individually<br>• Calculate precision data |
| Robustness | Measure of capacity of analytical procedure to remain unaffected by small but deliberate variations in method performance parameters, which provides an indication of its reliability during normal usage (term used by USP/ICH only) | | |
| Sensitivity | Change in response of measuring instrument divided by corresponding change in stimulus | Slope of calibration curve (arbitrary) | |

*Source:* From refs. 4, 15, 21, 55–56, 72, 75, and 76.

*Note:* Definitions, ways of expression, requirements or acceptance criteria, and guidelines for practical assessment (for more details, see text).

$t_{p,v}$ is the Student factor corresponding to the confidence level $1 - \alpha$ and $v$ degrees of freedom. The symbol $p$ represents the percentile or percentage point of the $t$ distribution. For one-sided intervals, $p = 1 - \alpha$; for two-sided intervals, $p = 1 - \alpha/2$. Values of $t$ can be found in the IUPAC nomenclature ($t = 2.776$ for $n = 5$ and $t = 3.182$ for $n = 4$ at $p = 0.95$) [67].

$X$ is the mean determined value and $n$ is the number of measurements for which the SD was calculated. If SD data of the certified reference materials are not available, 95% confidence limits may be used as an estimate of CRM SD (see second form of formula for $z$ score) [21].

$x_{bl}$ is the mean of the blank measurements, $s_{bl}$ is the SD on the blank measurements and $S$ is the sensitivity of the method or the slope of the calibration function. The calibration function is the relationship between the measured response $x_L$ and the concentration $c_L$ or amount $q_L$[56, 72, 95–96].

**TABLE 6 Horwitz Function**

| Analyte Percent | Analyte Ratio | Unit | Horwitz %RSD | AOAC PVM |
|---|---|---|---|---|
| 100 | 1 | 100% | 2 | 1.3 |
| 10 | $1 \times 10^{-1}$ | 10% | 2.8 | 2.8 |
| 1 | $1 \times 10^{-2}$ | 1% | 4 | 2.7 |
| 0.1 | $1 \times 10^{-3}$ | 0.10% | 5.7 | 3.7 |
| 0.01 | $1 \times 10^{-4}$ | 100 ppm | 8 | 5.3 |
| 0.001 | $1 \times 10^{-5}$ | 10 ppm | 11.3 | 7.3 |
| 0.0001 | $1 \times 10^{-6}$ | 1 ppm | 16 | 11 |
| 0.00001 | $1 \times 10^{-7}$ | 100 ppb | 22.6 | 15 |
| 0.000001 | $1 \times 10^{-8}$ | 10 ppb | 32 | 21 |
| 0.0000001 | $1 \times 10^{-9}$ | 1 ppb | 45.3 | 30 |

*Note:* This function if given as an empirical relationship between the precision of an analytical method and the concentration of the analyte regardless of the nature of the analyte, matrix, and method used. Acceptable $RSD_R$ and $RSD_r$ values according to [75] and to AOAC International [56, 62]; % RSD = percentage relative standard deviation, PVM = peer-verified methods program.

Precision relates to the random error of a measurement system (see Figure 8) and is a component of MU (see also Section 2 and Figure) [2].

*Trueness* is expressed in terms of *bias* or percentages of error. Bias is the difference between the mean value determined for the analyte of interest and the accepted true value or known level actually present [87]. It represents the systematic deviation of the measured result from the true result. Method trueness is an indicator also for utility and applicability of that method with real samples [88].

Different sources of systematic errors contribute to the overall bias (Figure 8). Thompson and Wood [8] describe *persistant bias* as the bias affecting all data of the analytical system over longer periods of time and being relatively small but continuously present. Different components contribute to the persistant bias, such as laboratory bias, method bias, and the matrix variation effect. Next to persistant bias, the larger *run effect* is the bias of the analytical system during a particular run [4, 8, 15].

One or more of these bias components are encountered when analyzing RMs. In general, RMs are divided into certified RMs (CRMs, either pure substances/solutions or matrix CRMs) and (noncertified) laboratory RMs (LRMs), also called QC samples [89]. CRMs can address all aspects of bias (method, laboratory, and run bias); they are defined with a statement of uncertainty and traceable to international standards. Therefore, CRMs are considered useful tools to achieve traceability in analytical measurements, to calibrat equipment and methods (in certain cases), to monitor laboratory performance, to validate methods, and to allow comparison of methods [4, 15, 30]. However, the use of CRMs does not necessarely guarantee trueness of the results. The best way to assess bias practically is by replicate analysis of samples with known concentrations such as reference materials (see also Section 2). The ideal reference material is a matrix CRM, as this is very similar to the samples of interest (the latter is called *matrix matching*). A correct result obtained with a matrix CRM, however, does not guarantee that the results of unknown samples with other matrix compositions will be correct [4, 89].

The usefulness of CRMs for validation (in particular for trueness assessment) and traceability purposes has been debated for years. This is illustrated by the enor-

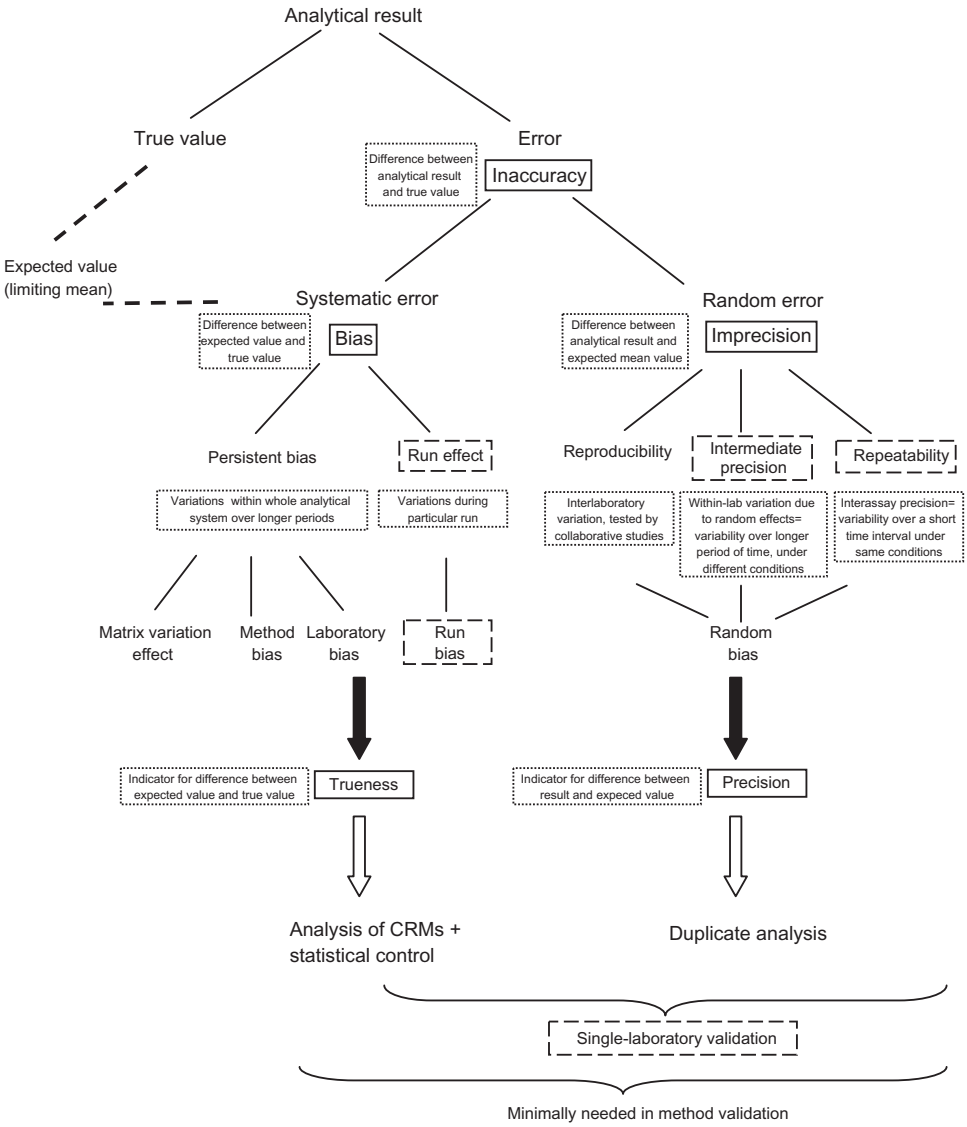


**FIGURE 8** Composition of error of analytical result related to accuracy of analytical method [4, 8].

mous number of papers published on this topic. We mention here only some interesting references [89–91]. Examples of the use of pure substance RMs, matrix CRMs, or LRMs can be found in the special issues of *Accreditation and Quality Assurance* (volume 9, 2004) and *Analytical and Bioanalytical Chemistry* (volume 278, 2004) on biological and environmental reference materials and the special issue of *Trends in Analytical Chemistry* (volume 23, 2004) on challenges for achieving traceability of environmental measurements.

If no such (certified) reference materials are available, a blank sample matrix of interest can be "spiked" with a known amount of a pure and stable in-house material, called the *spike* or *surrogate*. Recovery is then calculated as the percentage of

**TABLE 7 Acceptable recovery percentages as a function of the analyte concentration [56]**

| Analyte Percent | Analyte Ratio | Unit | Mean Recovery (%) |
|---|---|---|---|
| 100 | 1 | 100% | 98–102 |
| 10 | $1 \times 10^{-1}$ | 10% | 98–102 |
| 1 | $1 \times 10^{-2}$ | 1% | 97–103 |
| 0.1 | $1 \times 10^{-3}$ | 0.10% | 95–105 |
| 0.01 | $1 \times 10^{-4}$ | 100 ppm | 90–107 |
| 0.001 | $1 \times 10^{-5}$ | 10 ppm | 80–110 |
| 0.0001 | $1 \times 10^{-6}$ | 1 ppm | 80–110 |
| 0.00001 | $1 \times 10^{-7}$ | 100 ppb | 80–110 |
| 0.000001 | $1 \times 10^{-8}$ | 10 ppb | 60–115 |
| 0.0000001 | $1 \times 10^{-9}$ | 1 ppb | 40–120 |

the measured spike of the matrix sample relative to the measured spike of the blank control or the amount of spike added to the sample. The smaller the recovery percent, the larger the bias which is affecting the method and thus the lower the trueness [4, 56, 72, 92, 93]. An indication of trueness can also be obtained by comparing the method with a second, well-characterized reference method under the condition that the precision of the established reference method is known. Results from the two methods, performed on the same sample or set of samples, are compared. The samples may be CRMs, in-house standards, or just typical samples [15]. For a comparison between the three ways of establishing bias, see also Section 2 on analytical quality. It should be clear here that the use of recovery estimates and comparing methods are alternative ways which encompass serious limitations. They can give an idea about data comparability, but trueness cannot be assured [89].

Trueness or exactness of an analytical method can be documented in a control chart. Either the difference between the mean and true value of an analyzed (C)RM together with confidence limits or the percentage recovery of the known, added amount can be plotted [56, 62]. Here, again, special caution should be taken concerning the used reference. Control charts may be useful to achieve trueness only if a CRM, which is in principle traceable to SI units, is used. All other types of references only allow traceability to a "consensus" value, which however is assumed not to be necessarely equal to the "true" value [89]. The expected trueness or recovery percent values depend on the analyte concentration. Therefore, trueness should be estimated for at least three different concentrations. If recovery is measured, values should be compared to acceptable recovery rates as outlined by the AOAC Peer Verified Methods Program (Table 7) [56, 62]. Besides bias and percent recovery, another measure for the trueness is the $z$ score (Table 5). It is important to note that a considerable component of the overall MU will be attributed to MU on the bias of a system, including uncertainties on reference materials (Figures 5 and 8) [2].

*Recovery* is often treated as a separate validation parameter (Table 5). Analytical methods intend to estimate the true value of the analyte concentration with an uncertainty that is fit for the intended purpose. However, in such analytical methods, the analyte is transferred from the complex matrix to a simpler solution whereby there is a loss of analyte. As a consequence, the measured value will be lower than the true concentration present in the original matrix. Therefore, assessing the effi-

ciency of the method in detecting all of the analyte present is a part of the validation process. Eurachem, IUPAC, ISO, and AOAC International state that recovery values should always be established as a part of method validation. Recovery or spiking studies should be performed for different types of matrices, several examples of each matrix type, and each matrix type at different levels of analyte concentration [2, 4, 15].

***Specificity and Selectivity*** Specificity and selectivity both give an idea of the reliability of the analytical method (for definitions see Table 5). Some authors give different definitions for both terms while for others they are identical. The term *specific* generally refers to a method that produces a response for a single analyte only, while the term *selective* is used for a method producing responses for different chemical entities or analytes which can be distinguished from each other. A method is called selective if the response is distinguished from all other responses. In this case, the method is perfectly able to accurately measure an analyte in the presence of interferences [56, 94]. According to Eurachem, specificity and selectivity principally reflect the same characteristic and are related very closely to each other in such a way that specificity means 100% selectivity. In other words, a method can only be specific if it is 100% selective. Another related term is *confirmation of identity*, which is the proof that "the measurement signal, which has been attributed to the analyte, is only due to the analyte and not to the presence of something chemically or physically similar or arising as coincidence" [15]. A method must first show high specificity before true quantification can be performed [87].

There is no single expression form for specificity. It is rather something which must be demonstrated. The way in which this is done depends on the objective and the type of analytical method (see also below). For identification tests, the goal is to ensure the identity of an analyte. Specificity is here the ability to discriminate between compounds of closely related structures which can be present.

For impurity tests (limit impurity test, quantitative impurity test) and assay tests, the accent lies on the ability to determine or discriminate for the analyte in the presence of other interferants. Selectivity can be assessed by spiking samples with possible interferants (e.g., degradation products) [55, 56, 72].

***Limit of Detection*** There is no analytical term or parameter for which there exists a larger variety in terminology and formulations than for the limits of detection and quantification. *Limit of detection* or *detection limit* is the terminology most widely used as accepted by Eurachem. ISO uses *minimum detectable net concentration* while IUPAC prefers *minimum detectable (true) value* [15]. All official organizations however refer to the same definition: the lowest amount of an analyte in a sample which can be detected but not necessarily quantified as an exact value. In general, the LOD is expressed as a concentration $c_L$ or a quantity $q_L$, derived from the smallest signal $x_L$ which can be detected with reasonable certainty for a given analytical procedure. The lowest signal $x_L$ is the signal which lies $k$ times $SD_{blank}$ above the mean blank value, whereby $k$ is a numerical factor chosen according to the level of confidence required [56, 72, 95–97].

The larger the value of $k$, the larger the confidence level. Eurachem and IUPAC recommend a value of 3 for $k$, meaning that the chance that a signal more than $3s$ above the sample blank value is originating from the blank is less than 1%. The

detection limit is thus the concentration or amount corresponding to a measurement level (response, signal) three $s_{bl}$ units above the value for zero analyte (Table 5). At the concentration or amount level of 3 times the $s_{bl}$, the RSD or coefficient of variation (CV) on the measured signal is 33% (measure for uncertainty) [2, 4, 15, 75, 95, 98]. According to USP/ICH, the detection limit corresponds to that signal where the *signal-to-noise ratio* is 2:1 or 3:1 [72, 85].

It is not true—as is often thought—that detection or quantification is impossible below the determination limit, but at these lower levels, the uncertainty of the detection/quantification measurement is higher than the actual value itself [21]. In this context, Huber [56] defines the limit of detection also as the point at which a measured value is larger than the uncertainty associated with it. According to Krull and Swartz [88], the LOD is a concentration point where only the qualitative identification is possible but no accurate and precise quantification.

For qualitative methods, the LOD is defined as the threshold concentration at which the test becomes unreliable. A series of blank samples, spiked with different concentrations of the analyte, are each analyzed at least 10 times. The threshold, or *cutoff*, concentration is determined visually based on a response curve, plotting the percentage of positive results versus the concentration. In this respect, the LOD is also defined as the concentration at which 95% of the experiments give a clearly positive signal [15].

The LOD may not be confused with the sensitivity of the method. The latter is the capability of the method to discriminate small differences in concentration or mass of the test analyte and is equal to the slope of the calibration curve (see below) [56].

***Limit of Quantification*** For the *limit of quantification*, or *limit of determination*, definitions and formulas are very similar to those of LOD, except that for LOQ, $k$ is taken to be 5, 6, or even 10 [2, 4, 15, 56, 72, 96]. A value of 10 for $k$ means that the %RSD at the limit of quantification is 10%. The LOQ thus corresponds to that concentration or amount of analyte quantifiable with a variation coefficient not higher than 10% [98]. The LOQ is always higher than the LOD and is often taken as a fixed multiple (typically 2) of the detection limit [4]. Also, the determination limit is referred to as the signal 10 times above the noise or background signal, corresponding to a signal-to-noise ratio of 10:1 [72, 85].

In practice, the LOQ can be calculated in an analogous way as for the LOD, as indicated in Table 5. An alternative way of practically assessing the LOD and LOQ is the following. In a first step, 10 independent sample blanks are measured each once, the blank standard deviation $s_{bl}$ is calculated, and the lowest signals corresponding to both the LOD and the LOQ are calculated as $x_{LOD} = x_{bl} + 3s_{bl}$ and $x_{LOQ} = x_{bl} + 10s_{bl}$, respectively. In a second step, sample blanks are spiked with various analyte concentrations (e.g., 6) close to the LOD. Per concentration, 10 independent replicates are measured and the standard deviation of the measured signals calculated. These standard deviations $s$ (or the %RSD) are then plotted against the concentration. LOD and LOQ values are those concentrations of analyte corresponding to %RSD values of 33 and 10%, respectively [15, 21].

As is said for the LOD, neither is it true that at and below the LOQ quantification becomes impossible. Quantification is possible; however it becomes unreliable as the uncertainty associated with it at these lower levels is higher than the measure-

ment value itself. Quantification becomes reliable as soon as the MU is lower than the value measured [21].

***Decision Limit and Detection Capability: For Specific Sectors Only*** In the context of analytical method validation, the terms *decision limit* ($CC_\alpha$) and *detection capability* ($CC_\beta$) as well as *minimum required performance limits* (MRPLs) are often used and need some clarification. These terms are applicable for the measurement of organic residues, contaminants, and chemical elements in live animals and animal products, as regulated within the European Union (EU) by directives 96/23/EC [99], 2002/657/EC [82], and 2003/181/EC [100]. The Commission distinguishes group A substances for which no *permitted limit* (*maximum residue level*, MRL) has been established and group B substances having a fixed permitted limit.

The decision limit $CC_\alpha$ is the limit at and above which it can be concluded with an error probability of $\alpha$ that a sample is *noncompliant*. If a permitted limit (PL) has been established for a substance (group B or the regulated compounds), the result of a sample is noncompliant if the decision limit is exceeded ($CC_\alpha = x_{PL} + 1.64s_{MRL}$). If no permitted limit has been established (group A), the decision limit is the lowest concentration level at which the method can discriminate with a statistical certainty of 1-$\alpha$ that the particular analyte is present ($CC_\alpha = x_{bl} + 2.33s_{sample}$). The detection capability $CC_\beta$ is the smallest content of the substance that may be detected, identified, and/or quantified in a sample with an error probability of $\beta$ ($CC_\beta = CC_\alpha + 1.65s_{sample}$).

Minimum required performance limits have been established for substances for which no permitted limit has been fixed and in particular for those substances whose use is not authorized or even prohibited within the EU (group A). A MRPL is the minimum content of an analyte in a sample which at least has to be detected and confirmed. A few MRPLs for residues of certain veterinary drugs have been published so far in directive 2003/181/EC.

For group A substances (no PL established), $CC_\alpha$ and $CC_\beta$ are comparable with LOD and LOQ, respectively, as their concentrations correspond to measured signals laying $y$ times above the blank signal. For substances having a PL (group B), $CC_\alpha$ and $CC_\beta$ are not related to LOD and LOQ but are expressed in relation to this PL. It is important to note that these terms apply specifically for inspection of animals and fresh meat for the presence of residues of veterinary drugs and specific contaminants and are therefore different from LOD and LOQ [82, 99–102].

***Linearity and Range*** For assessment of the linearity of an analytical method, linear regression calculations do not suffice. In addition, residual values should be calculated (Table 5). The latter represent the differences between the actual $y$ value and the $y$ value predicted from the regression curve for each $x$ value. If residual values, calculated by simple linear regression, are randomly distributed about the regression line, linearity is confirmed, while systematic trends indicate nonlinearity. If such a trend or pattern is observed, this suggests that the data are best treated by weighted regression. For either simple or weighted linear regression, linearity supposes that the intercept is not significantly different from zero [4, 15, 21, 75].

An alternative approach to establish linearity is to divide the response by the respective concentrations and to plot these "relative responses" as a function of the concentration on a log scale. The obtained line should be horizontal over the full

linear range, with a positive deviation at low concentrations and a negative deviation at higher concentrations. By drawing parallel horizontal lines, corresponding to, for example, 95 and 105% of the horizontal relative response line, the intersection points can be derived at which the method becomes nonlinear [56].

It is important that a linear curve is repeatable from day to day. However, linear ranges may be different for different matrices. The reason for this is a possible effect of interferences inherent to the matrix. A test for general matrix effects can be performed by means of "standard additions" or the method of analyte additions. For a set of samples, obtained by adding different concentrations of analyte to a certain matrix, the slope of the calibration curve is compared with the slope of the usual calibration function. A lack of significance (curves are parallel) means that there is no matrix effect [21, 75].

***Ruggedness and Robustness*** Although the terms *ruggedness* and *robustness* are often treated as the same and used interchangeably, separate definitions exist for both terms, as indicated in Table 5.

To have an idea about the ruggedness, Eurachem recommends to introduce deliberate variations to the method, such as different days, analysts, instruments, and reagents and variations in sample preparation or sample material used. Changes should be made separately and the effect evaluated of each set of experimental conditions on the precision and trueness [4, 15, 85]. To examine the effects of different factors, a "factorial design" methodology can be applied, as described by von Holst et al. [103]. By combining changes in conditions and performing a set of experiments, one can determine which factors have a significant or even critical influence on the analytical results. In ICH/USP guidelines, ruggedness is not defined separately but treated under the same denominator as reproducibility precision: It is "the degree of reproducibility of the results obtained under a variety of conditions, expressed as %RSD" [56].

Robustness is a term introduced by USP/ICH [88]. Although Eurachem has included the term robustness in its official list of definitions, the term is not used by official organizations other than USP/ICH. According to Eurachem, both parameters do present the same and are thus synonyms [15, 72, 85].

***Sensitivity*** The sensitivity of a method is the gradient of the response curve. In practical terms, sensitivity refers to the slope of the calibration curve. Sensitivity is often used together with detection and quantification limits. Indeed, the slope of the calibration curve is used for the calculation of limits of detection and quantification. A method is called sensitive if a small change in concentration or amount of analyte causes a large change in the measured signal [4, 15, 21]. Sensitivity is not always mentioned as a validation parameter in official guidelines. According to Thompson et al. [4], it is not useful in validation because it is usually arbitrary, depending on instrument settings. USP/ICH does not mention sensitivity at all.

### 3.5 Analytical Quality Assurance

Quality assurance is the complete organizational infrastructure that forms the basis for all reliable analytical measurements [8]. It stands for all the planned and systematic activities and measures implemented within the quality system [2, 104]. A quality system has a quality plan, which emphasizes the implementation of good

laboratory practices (GLPs). GLPs are comparable to the good manufacturing practices (GMPs) and the larger HACCPs (hazard analysis critical control points) quality systems of food production factories. Attention goes to all aspects of quality management in the laboratory organization, including staff training, the maintenance and calibration of all equipment used, the laboratory environment, safety measures, the system of sample identification, recordkeeping, and storage—the latter may be simplified by the use of a laboratory information management system (LIMS), the use of validated and standardized methods, and the documentation of these methods and of all information concerning the followed procedures (standard operating procedures, or SOPs).

Quality assurance embraces both QC and "quality assessment." QC is defined as the mechanism or the practical activities undertaken to control errors, while quality assessment is the mechanism to verify that the system is operating within acceptable limits. Quality assessment and control measures are in place to ensure that the measurement process is stable and under control [2, 8].

Within QC, internal and external quality control are distinguished. In general, QA comprises the following topics as also schematized in Figure 6.

***Use of Validated Methods: In-House Versus Interlaboratory Validation*** Wherever possible or practically achievable, a laboratory should use methods which have been "fully validated" through a collaborative trial, also called interlaboratory study or method performance study. Validation in collaborative studies is required for any new analytical method before it can be published as a standard method (see below). However, single-laboratory validation is a valuable source of data usable to demonstrate the fitness for purpose of an analytical method. In-house validation is of particular interest in cases where it is inconvenient or impossible for a laboratory to enter into or to organize itself a collaborative study [4, 5].

On the one hand, even if an in-house validated method shows good performance and reliable accuracy, such a method cannot be adopted as a standard method. In-house validated methods need to be compared between at least eight laboratories in a collaborative trial. On the other hand, a collaborative study should not be conducted with an unoptimized method [58]. Interlaboratory studies are restricted to precision and trueness while other important performance characteristics such as specificity and LOD are not addressed [105]. For these reasons, single-laboratory validation and interlaboratory validation studies do not exclude each other but must be seen as two necessary and complementary stages in a process, presented in Figure 9. The added value of single-laboratory validation is that it simplifies the next step—interlaboratory validation—and thereby minimizes the gap between internally (validated or not) developed methods and the status of interlaboratory validation. By optimizing the method first within the laboratory, as a kind of preliminary work, an enormous amount of collaborators' time and money is saved [58].

The importance of conducting such a single-laboratory preliminary validation step is increasingly highlighted by international standardization agencies. The IUPAC and AOAC International include a preliminary work paragraph in their guidelines for collaborative studies, stating that within-laboratory testing data are required on precision, bias, recovery, and applicability. Additionally, a clear description of the method, including statements on the purpose of the method, the type of method, and the probable use of the method, is required within this preliminary work [62, 63]. It is however not only in the harmonized guidelines for collaborative

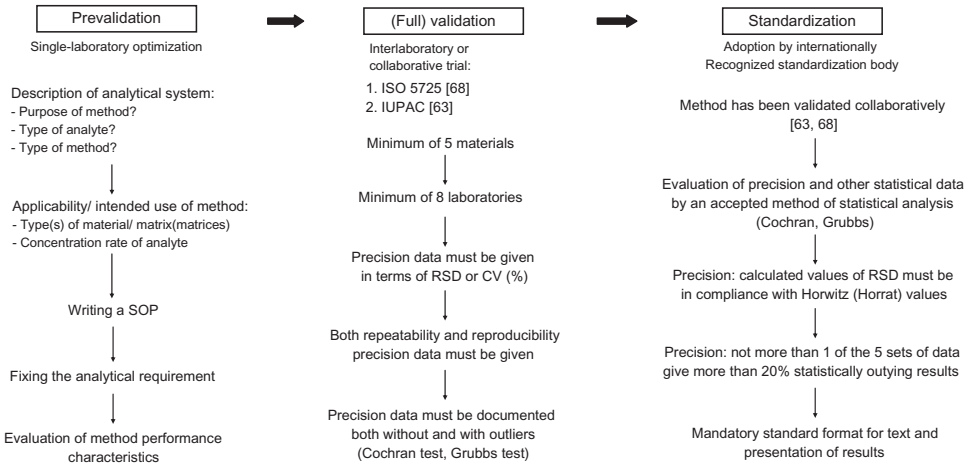


**FIGURE 9** Hierarchy of, relationship between, and objectives and requirements for prevalidation [106], validation [62, 63, 68], and standardization of analytical methods [62, 63, 67, 68, 75, 84]: RSD = relative standard deviation, CV = coefficient of variation, SOP = standard operating procedure.

trials (see below) that the link between a single-laboratory prevalidation step and the collaborative trial is emphasized. Separate guidelines for single-laboratory validation of methods of analysis have recently been published by the IUPAC, ISO, and AOAC International [4]. The IUPAC guidelines have also been considered and accepted by the CCMAS [77]. In addition, specific, individual working groups or scientists are presenting their own framework for pre-, single-laboratory validation of methods of analysis. The latter do not concern official, published guidelines but can be found on the Internet [see, e.g., 58, 85] or are distributed through national or international specific working groups. The objectives of a single-laboratory or in-house validation process are depicted in Figure 9. Depending on the type of method (Table 4), data can be obtained for all criteria except for the reproducibility (interlaboratory) precision. However, it is this *among-laboratories variability* or reproducibility which is the dominating error component in analytical measurement and which underlies the need for interlaboratory validation [106].

Interlaboratory or collaborative validation studies can be organized by any laboratory, institute, or organization but should preferably be conducted according to one of the following recognized protocols: (1) ISO 5725 on Accuracy (Trueness and Precision) of Measurement Methods and Results [68] or (2) IUPAC Protocol for the Design, Conduct and Interpretation of Method-Performance Studies [63]. The latter revised, harmonized guidelines have been adopted also by AOAC International as the guidelines for the AOAC Official Methods Program [62]. The main requirements for collaborative studies outlined in these guidelines are shown in Figure 9.

Precision plays a central role in collaborative studies. Wood [84] defines a collaborative trial as "a procedure whereby the precision of a method of analysis may be assessed and quantified." Precision is the objective of interlaboratory validation studies, and not trueness or whichever other method performance parameter. Evalu-

ation of the acceptability of precision data is important for the standardization of methods (see below).

***Use of Standardized Methods*** The first level of AQA is the use of validated or standardized methods. The terms *validated* and *standardized* here refer to the fact that the method performance characteristics have been evaluated and have proven to meet certain requirements. At least, precision data are documented, giving an idea of the uncertainty and thus of the error of the analytical result. In both validated and standardized methods, the performance of the method is known.

Validated methods can be developed by the laboratory itself or by a standardization organization after interlaboratory studies. Standardized methods are developed by organizations such as the AOAC International, ISO, USP (see Table 3), U.S. Environmental Protection Agency (EPA), American Society for Testing and Materials (ASTM), or Food Standards Association (FSA) [56]. Here exactly lies the difference between a validated and a standardized method: An analytical method can only be standardized after it has been validated through interlaboratory comparisons. The main prerequisite for a standards organization is that a method has been adequately studied and its precision shown to meet a required standard, as summarized in Figure 9. The format of a standard method as outlined in ISO Guide 78/2 [58] is shown in Table 8 [2]. A specific IUPAC protocol [67] describes in detail how to present AQA data such as the performance characteristics.

**TABLE 8 Mandatory Text Format for Standardized Methods**

| | |
|---|---|
| 1. Scope | States briefly what method determines |
| 2. Definitions | Precise definition of analyte or parameter determination by method |
| 3. Fields of application | Type of materal(s)/matrix(ces) to which method is applicable |
| 4. Principle | Basic steps involved in procedure |
| 5. Apparatus | Specific apparatuses required for determination are listed |
| 6. Reagents | Analytical reagent-grade reagents needed for determination are listed |
| 7. Sampling | |
| 8. Procedure | Divided into numbered paragraphs or subclauses, includes a preparation-of-test-sample step and reference to QA procedures |
| 9. Calculation and expression of results | Indication of how final results are calculated and units in which results are to be expressed |
| 10. Notes | Additional information as to procedure; may be in form of notes, placed here or in body of text |
| Annex | Includes all information on analytical quality control, such a precision clauses (repeatability and reproducibility data), table of statistical data outlining accuracy (trueness and precision) of method |
| References | Reference to published report on collaborative study which was carried out prior to standardization of method |

*Source*: According to ISO Guide 78/2 [2, 58].

***Effective Internal Quality Control (IQC)*** In the IUPAC harmonized guidelines for IQC, Thompson and Wood [8] define IQC as a "set of procedures undertaken by the laboratory staff for the continuous monitoring of operation and the results of measurements in order to decide whether results are reliable enough to be released". IQC guarantees that methods of analysis are fit for their intended purpose, meaning the continuous achievement of analytical results with the required standard of accuracy. The objective of IQC is in fact the elongation of method validation: to continuously check the accuracy of analytical data obtained from day to day in the laboratory. In this respect, both systematic errors, leading to bias, as well as random errors, leading to imprecision, are monitored. In order to be able to monitor these errors, they should remain constant. Within the laboratory, such constant conditions are typically achieved in one analytical run. The word *internal* in IQC implicates that repeatability conditions are achieved. Thus monitoring the precision as the objective of IQC concerns not reproducibility or interlaboratory precision but only repeatability or intralaboratory precision. In fact, the monitoring of accuracy of an analytical method in IQC can be translated into the monitoring of the analytical system [8].

Two aspects are important for IQC: (1) the analysis of *control materials* such as reference materials or spiked samples to monitor trueness and (2) replication of analysis to monitor precision. Of high value in IQC are also blank samples and blind samples. Both IQC aspects form a part of statistical control, a tool for monitoring the accuracy of an analytical system. In a control chart, such as a *Shewhart control chart*, measured values of repeated analyses of a reference material are plotted against the run number. Based on the data in a control chart, a method is defined either as an *analytical system under control* or as an *analytical system out of control*. This interpretation is possible by drawing horizontal lines on the chart: $\bar{x}$(mean value), $\bar{x} + s$ (SD) and $\bar{x} - s$, $\bar{x} + 2s$ (upper warning limit) and $\bar{x} - 2s$ (lower warning limit), and $\bar{x} + 3s$ (upper action or control limit) and $\bar{x} - 3s$ (lower action or control limit). An analytical system is under control if no more than 5% of the measured values exceed the warning limits [2, 6, 85].

***Participance in Proficiency Testing Schemes*** Proficiency testing (PT) is the periodic assessment of the competency or the analytical performance of individual participating laboratories [23]. An independent coordinator distributes individual test portions of a typical uniform test material. The participating laboratories analyze the materials by their method of choice and return the results to the coordinator. Test results obtained by different laboratories are subsequently compared with each other and the performance of each participant evaluated based on a single competency score [64, 107]. International harmonized protocols exist for the organization of PT schemes [59, 60, 64, 69, 79].

Participation in proficiency tests is not a prerequisite or an absolute substitute for IQC measures or vice versa. However, participance in proficiency tests is meaningless without a well-developed IQC system. IQC underlies participance in PT schemes, while IQC and participance in PT schemes are both important substitutes of AQA (Figure 6). It is shown that laboratories with the strongest QC procedures score significantly better in PT schemes [8, 50]. Participance in PT can to a certain extent improve the laboratory's performance; however unsatisfactory performance in schemes (up to 30% of all participants) has been reported. This means that there

**TABLE 9 Differences between Method Performance Studies and Proficiency-Testing Schemes**

| Characteristic | Collaborative/Interlaboratory (Method Performance) studies | Proficiency-Testing Schemes |
|---|---|---|
| Main Objective | *Validation* of new methods | *Competency check* of analytical laboratories |
| Application | New methods, required for full validation and standardization, first prerequisite for IQC and QA | Routinely used (validation and/or standardization methods), recommended within IQC and QA system |
| Results aimed at | *Precision*: multiple results, both repeatability and reproducibility; >%RSD is compared to theoretical Horwitz Horrat values | *Trueness*: Single result per test material, > calculation of $Z$ score as measure for bias |
| Method and protocol used | Single prescribed method for which SOP must strictly be followed | Multiplicity of methods; participants have free choice of (validated and/or standardized) method |
| Test materials | Minimum of five different materials, no stipulations for homogeneity and stability of test samples | No minimum, per round often less than five test samples; homogeneity and stability of materials must be assured |
| Participating laboratories | Minimum of eight, assumed to be equally competent | No minimum, variety in participants is possible throughout one scheme (different rounds); not assumed to have equal competency (will be tested) |

*Source*: From refs. 8, 78, and 107.

is no correlation between good analytical performance and participation in PT [108]. However, PT has a significant educational function as it helps a laboratory to demonstrate competency to an accreditation body or another third party [60].

The terms *PT schemes* and *collaborative trials* are often confused with each other, as in both QA measures, a number of different laboratories are involved. However, there is a clear distinction between both. The mean differences with respect to the objective and application, the results, used method, test materials, and participating laboratories are summarized in Table 9. It is important to note also that the results obtained from PT schemes, as well as those from collaborative performance studies, can be used for assessing the MU (see Section 2).

***External Quality Control and Accreditation*** Participation in PT schemes is an objective means of evaluating the reliability of the data produced by a laboratory. Another form of external assessment of the laboratory performance is the physical inspection of the laboratory to ensure that it complies with externally imposed standards. Accreditation of the laboratory indicates that it is applying the required

QA principles. The "golden standard" ISO/IEC 17025 [17], which is the revised version of ISO Guide 25 [70], describes the general requirements for the competence of calibration and testing laboratories. In Europe, the accreditation criteria have been formalized in European standard EN45001 [109]. Participation in PT schemes forms the basis for accreditation, because PT is a powerful tool for a laboratory to demonstrate its competency. Accreditation guides use the information obtained by PT schemes [6, 17, 60, 64].

Accreditation is a formal recognition that a laboratory is competent to carry out specific (types of) calibrations or tests [2]. After the use of validated and standardized methods, the introduction and use of appropriate IQC procedures and the participation in PT schemes, accreditation to ISO/IEC 17025 is the fourth basic principle related to laboratory QA in general [4]. Guidelines on the implementation of ISO/IEC 17025, including the estimation of MU (see also Section 2), are published in the literature and by official accreditation bodies such as Eurachem, CITAC, EA, Eurolab, and ILAC (see Table 1) [2, 60, 80, 81, 110]. It is worthwile to mention that accreditation, just like participance in PT schemes, does not necessarily indicate good performance of the laboratory [108].

## 4 SUMMARY

Together with the fast development of analytical methodologies, great importance is nowadays attached to the quality of the measurement data. Besides the necessary reporting of any result with its MU and traceability of the results to stated standards or references (Section 2), a third crucial aspect of analytical methods of whichever type is their status of validation. It is internationally recognized that validation is necessary in analytical laboratories. However, less is known about *what* is validation and *what* should be validated, *why* validation is important, *when* and by *whom* validation is performed, and finally, *how* it is carried out practically. This Chapter has tried to answer these questions.

Method validation is defined in detail and different approaches to evaluate the acceptability of analytical methods are described. Great importance is attached to the different method performance parameters, their definitions, ways of expression, and approaches for practical assessment. Validation of analytical methodologies is placed in the broader context of QA. The topics of standardization, internal and external quality control, and accreditation are discussed as well as the links between these different aspects. Because validation and QA apply for a specific analytical method, it is important to approach each method on a case-by-case basis. An analytical method is a complex, multistep process, starting with sampling and ending with the generation of a result. Although every method has its specific scope, application, and analytical requirement, the basic principles of validation and QA are the same, regardless of the type of method or sector of application. The information in this work is mainly taken from analytical chemistry, but it applies to other sectors as well.

Section 3 on quality in the analytical laboratory provides a good, complete, and up-to-date collation of relevant information in the fields of analytical method validation and QA. It is useful for the completely inexperienced scientist as well as for those involved in this topic for a long time but somewhere having lost the way in

the labyrinth, looking for more explanation on a particular aspect, or longing for deeper insight and knowledge.

## ACKNOWLEDGMENTS

The authors wish to thank Simon Kay for preliminary discussions on the topic, Janna Puumalainen for reading and commenting on early versions of the chapter, Andrew Damant for giving suggestions, Arne Holst-Jensen for refreshing ideas, and Friedle Vanhee for many helpful discussions, reading, and assistance throughout the writing process.

## REFERENCES

1. van Zoonen, P., Hoogerbrugge, R., Gort, S. M., Van de Wiel, H. J., and Van't Klooster, H. A. (1999), Some practical examples of method validation in the analytical laboratory, *Trends Anal. Chem.*, 18, 584–593.
2. CITAC/Eurachem Guide (2002), *Guide to Quality in Analytical Chemistry—An Aid to Accreditation*, available: http://www.eurachem.org.
3. Eurachem/CITAC Guide (2003), *Traceability in Chemical Measurement. A Guide to Achieving Comparable Results in Chemical Measurement*, Joint Eurachem/CITAC Working Group on Measurement Uncertainty and Traceability, available: http://www.eurachem.org.
4. Thompson, M., Ellison, S., and Wood, R. (2002), Harmonised guidelines for single-laboratory validation of methods of analysis, IUPAC Technical Report, *Pure Appl. Chem.*, 74, 835–855.
5. Battaglia, R. (1996), Quality assurance in a food analytical laboratory—The introduction of EN 45001 in the food analytical laboratories of a retail company, *Accred. Qual. Assur.*, 1, 256–261.
6. Mesley, R. J., Pocklington, W. D., and Walker, R. F. (1991), Analytical quality assurance: A review, *Analyst*, 116, 975–990.
7. Eurachem Guide EEE/RM/062rev3 (2002), *The Selection and Use of Reference Materials.* A Basic Guide for Laboratories and Accreditation Bodies, available: http://www.eurachem.org.
8. Thompson, M., and Wood, R. (1995), Harmonised guidelines for internal quality control in analytical chemistry laboratories, *Pure Appl. Chem.*, 67, 649–656.
9. Maroto, A., Boqué, R., Riu, J., and Rius, F. X. (1999), Evaluating uncertainty in routine analysis, *Trends Anal. Chem.*, 18, 577–584.
10. Hund, E., Massar, D. L., and Smeyers-Verbeke, J. (2001), Operational definitions of uncertainty, *Trends Anal. Chem.*, 20, 394–406.
11. Fleming, J., Albus, H., Neidhart, B., and Wegschieder, W. (1996), Glossary of analytical terms (II), *Accred. Qual. Assur.*, 1, 87–88.
12. Quevauviller, Ph. (2004), Traceability of environmental chemical measurements, *Trends Anal. Chem.*, 23, 171–176.
13. Eurachem/CITAC Guide (1995), in Ellison, S. L. R., Rosslein, M., and Williams, A., Eds., *Quantifying Uncertainty in Analytical Measurement*, 1st ed., available: http://www.eurachem.org.

14. Eurachem/CITAC Guide (2000), in Ellison, S. L. R., Rosslein, M., and Williams, A., Eds., *Quantifying Uncertainty in Analytical Measurement*, 2nd ed., available: http://www.eurachem.org.
15. Eurachem Guide (1998), *The Fitness for Purpose of Analytical Methods. A Laboratory Guide to Method Validation and Related Topics*, LGC, Teddington, available: http://www.eurachem.org.
16. International Organization for Standardization (ISO) GUM (1995), *Guide to the Expression of Uncertainty in Measurement*, ISO, Geneva.
17. International Organization for Standardization (ISO)/IEC 17025 (1999), *General Requirements for the Competence of Calibration and Testing Laboratories*, ISO, Geneva.
18. CX/MAS 01/8 (2001), Codex Alimentarius Commission, Codex Committee on Methods of Analysis and Sampling (FAO/WHO), Measurement uncertainty. Relationship between the analytical result, the measurement uncertainty and the specification in Codex standards, agenda item 4a of the 23rd session, Budapest, Hungary, Feb. 26–Mar. 2, 2001.
19. CX/MAS 02/6 (2002), Codex Alimentarius Commission, Codex Committee on Methods of Analysis and Sampling (FAO/WHO), Proposed draft guidelines on measurement uncertainty, agenda item 5 of the 24th session, Budapest, Hungary, Nov. 18–22, 2002.
20. CX/MAS 02/13 (2002), Codex Alimentarius Commission, Codex Committee on Methods of Analysis and Sampling (FAO/WHO), The use of analytical results: Sampling, relationship between the analytical results, the measurement uncertainty, recovery factors and the provisions in Codex standards, agenda item 9 of the 24th session, Budapest, Hungary, Nov. 18–22, 2002.
21. CX/MAS 02/4 (2002), Codex Alimentarius Commission, Codex Committee on Methods of Analysis and Sampling (FAO/WHO), Proposed draft guidelines for evaluating acceptable methods of analysis, agenda item 4a of the 24th session, Budapest, Hungary, Nov. 18–22, 2002.
22. European Cooperation for Accreditation of Laboratories (EAL) (1996), *The Expression of Uncertainty in Quantitative Testing*, EAL-G23, The Netherlands.
23. International Laboratory Accreditation Cooperation (ILAC) (2002), *Introducing the Concept of Uncertainty of Measurement in Testing in Association with the Application of the Standard ISO/IEC 17025*, ILAC-G17, ILAC Technical Accreditation Issues Committee, available: www.ilac.org.
24. Valcarcél, M., and Rios, A. (1999), Traceability in chemical measurements for the end users, *Trends Anal. Chem.*, 18, 570–576.
25. Walsh, M. C. (1999), Moving from official to traceable methods, *Trends Anal. Chem.*, 18, 616–623.
26. Valcarcél, M., and Rios, A. (1997), Is traceability an exclusive property of analytical results? An extended approach to traceability in chemical measurement, *Fresen. J. Anal. Chem.*, 359, 473–475.
27. Quevauviller, Ph., and Donard, O. F. X. (2001), Stated references for ensuring traceability of chemical measurements for long-term environmental monitoring, *Trends Anal. Chem.*, 20, 600–613.
28. Pan, X. R. (1996), The traceability scheme in chemical measurement, *Accred. Qual. Assur.*, 1, 181–185.
29. Charlet, P., and Marschal, A. (2004), Improvement in the traceability of environmental analysis by the relevant use of certified pure solutions and a matrix certified reference material, *Trends Anal. Chem.*, 23, 178–184.

30. Segura, M., Camara, C., Madrid, C., Rebollo, C., Azcarate, J., Kramer, G. N., Gawlik, B. M., Lamberty, A., and Quevauviller, Ph. (2004), Certified reference materials (CRMs) for quality control of trace-element determinations in wastewaters, *Trends Anal. Chem.*, 23, 194–202.
31. Förstner, U. (2004), Traceability of sediment analysis, *Trends Anal. Chem.*, 23, 217–236.
32. Theocharopoulos, S. P., Mitsios, I. K., and Arvanitoyannis, J. (2004), Traceability of environmental soil measurements, *Trends Anal. Chem.*, 23, 237–251.
33. Sabé, R., and Rauret, G. (2004), Challenges for achieving traceability of analytical measurements of heavy metals in environmental samples by isotopic dilution mass spectrometry, *Trends Anal. Chem.*, 23, 273–280.
34. Drolc, A., Ros, M., and Cotman, M. (2004), Establishment of traceability of ammonium nitrogen determination in wastewater, *Anal. Bioanal. Chem.*, 378, 1243–1250.
35. Thompson, M. (1998), Uncertainty of sampling in chemical analysis, *Accred. Qual. Assur.*, 3, 117–121.
36. Roy, S., and Fouillac, A.-M. (2004), Uncertainties related to sampling and their impact on the chemical analysis of groundwater, *Trends Anal. Chem.*, 23, 185–193.
37. Analytical Methods Committee (1995), Uncertainty of measurement: Implications of its use in analytical science, *Analyst*, 120, 2303–2308.
38. Fleming, J., Albus, H., Neidhart, B., and Wegschieder, W. (1997), Glossary of analytical terms (VIII), *Accred. Qual. Assur.*, 2, 160–161.
39. Hund, E., Massart, D. L., and Smeyers-Verbeke, J. (2003), Comparison of different approaches to estimate the uncertainty of a liquid chromatographic assay, *Anal. Chim. Acta.*, 480, 39–52.
40. Eurachem/EA Guide 04/10 (2002), *Accreditation for Microbiological Laboratories*, available: http://www.eurachem.org.
41. Küppers, S. (1998), Is the estimation of measurement uncertainty a viable alternative to validation? *Accred. Qual. Assur.*, 3, 412–415
42. Ellison, S. L. R., and Barwick, V. J. (1998), Using validation data for ISO measurement uncertainty estimation. Part 1. Principles of an approach using cause and effect analysis, *Analyst*, 123, 1387–1392.
43. Moser, J., Wegscheider, W., and Sperka-Gottlieb, C. (2001), Quantifying the measurement uncertainty of results from environmental analytical methods, *Fresen. J. Anal. Chem.*, 370, 679–689.
44. Dehouck, P., Vander Heyden, Y., Smeyers-Verbeke, J., Massart, D. L., Crommen, P. H., Marini, R. D., Smeets, O. S., Decristoforo, G., Van de Wauw, W., De Beer, J., Quaglia, M. G., Stella, C., Veuthey, J.-L., Estevenon, O., Van Schepdael, A., Roets, E., and Hoogmartens, J. (2003), Determination of uncertainty in analytical measurements from collaborative study results on the analysis of a phenoxymethylpenicillin sample, *Anal. Chim. Acta*, 481, 261–272.
45. Maroto, A., Riu, J., Boqué, R., and Rius, F. X. (1999). Estimating uncertainties of analytical results using information from the validation process, *Anal. Chim. Acta*, 391, 173–185.
46. Maroto, A., Boqué, R., Riu, J., and Rius, F. X. (2001), Estimation of measurement uncertainty by using regression techniques and spiked samples, *Anal. Chim. Acta*, 446, 133–145.
47. Barwick, V. J., and Ellison, S. L. R. (2000, Jan.), *VAM Project 3.2.1*: *Development and Harmonisation of Measurement Uncertainty Principles. Part d. Protocol for Uncertainty Evaluation from Validation Data*, LGC, UK, available: http://www.vam.org.uk.

48. Armishaw, P. (2003), Estimating measurement uncertainty in an afternoon. A case study in the practical application of measurement uncertainty, *Accred. Qual. Assur.*, 8, 218–242.
49. Horwitz, W. (1982), Evaluation of analytical methods used for regulation of food and drugs, *Anal. Chem.*, 54, 67A–76A.
50. Thompson, M., and Lowthian, P. J. (1997), The Horwitz function revisited, *J. AOAC Int.*, 80, 676–679.
51. King, B. (2001), Meeting the measurement uncertainty and traceability requirements of ISO/IEC standard 17025 in chemical analysis, *Fresen. J. Anal. Chem.*, 371, 714.
52. Mueller-Harvey, I. (2003), Do we need quality assurance and quality control of analytical measurements in R&D laboratories? *Food Agric. Environ.*, 1, 9–11.
53. Rösslein, M. (2000), Evaluation of uncertainty utilising the component by component approach, *Accred. Qual. Assur.*, 5, 88–94.
54. Mueller, N. (2002), Introducing the concept of uncertainty of measurement in testing in association with the application of the standard ISO/IEC 17025, *Accred. Qual. Assur.*, 7, 79–80.
55. International Organization on Harmonization (1995), Guideline for industry: Text on validation of analytical procedures, ICH-Q2A, available: http://www.fda.gov/cder/guidance/index.htm.
56. Huber, L. (1998), Validation of analytical methods. Part 9, in Huber, L., Ed., Validation and Qualification in Analytical Laboratories, Interpharm Press, Agilent Technologies, available: http://www.labcompliance.com.
57. Wells, R. J. (1998), Validation requirements for chemical methods in quantitative analysis—Horses for courses? *Accred. Qual. Assur.*, 3, 189–193.
58. Green, M. (1996), A practical guide to analytical method validation, *Anal. Chem.*, 68, 305A–309A.
59. Eurachem Guide (2000), *Selection, Use and Interpretation of Proficiency Testing (PT) Schemes by Laboratories*, available: http://www.eurachem.org.
60. Eurachem/Eurolab/EA Guide EA-03/04 (2001, Aug.), *Use of Proficiency Testing as a Tool for Accreditation in Testing*, available: http://www.eurachem.org.
61. European Committee for Normalisation (CEN), (2006), Standardization and related activities—General Vocabulary (from ISO/IEC Guide 2:2004), EN45020:2006, www.cenorm.be.
62. Association of Official Analytical Chemists (AOAC) International (2000), Method validation programs (OMA/PVM Department), including Appendix D: Guidelines for collaborative study procedures to validate characteristics of a method of analysis, available: http://www.aoac.org/vmeth/devmethno.htm.
63. Horwitz, W. (1995), Protocol for the design, conduct and interpretation of method performance studies, *Pure Appl. Chem.*, 67, 331–343.
64. Thompson, M., and Wood, R. (1993), The international harmonized protocol for the proficiency testing of (chemical) analytical laboratories, *Pure Appl. Chem.*, 65, 2123–2144: also *J. AOAC Int.*, 76, 26–940 (1993).
65. Currie, L. A. (1995), Nomenclature in evaluation of analytical methods including detection and quantification capabilities, *Pure Appl. Chem.*, 67, 1699–1723.
66. Currie, L. A., and Svehla, G. (1994), Nomenclature for the presentation of results of chemical analysis, International Union of Pure and Applied Chemistry (IUPAC), *Pure Appl. Chem.*, 66, 595–608.
67. Pocklington, W. D. (1990), Harmonized protocols for the adoption of standardized analytical methods and for the presentation of their performance characteristics,

International Union of Pure and Applied Chemistry (IUPAC), *Pure Appl. Chem.*, 62, 149–162.

68. International Organization for Standardization (ISO) Guide 5725 (1994), *Accuracy (Trueness and Precision) of Measurement Methods and Results*, ISO, Geneva.
69. International Organization for Standardization (ISO) Guide 43 (1984), *Development and Operation of Laboratory Proficiency Testing*, ISO, Geneva.
70. International Organization for Standardization (ISO) Guide 25 (1990), *General Requirements for the Competence of Calibration and Testing Laboratories*, ISO, Geneva.
71. International Organization for Standardization (ISO)/IEC 17025 (1999), *General Requirements for the Competence of Calibration and Testing Laboratories*, ISO, Geneva.
72. International Conference on Harmonization (1996), *Guidance for Industry: Validation of Analytical Procedures: Methodology*, ICH-Q2B, available: http://www.fda.gov/cder/guidance/index.htm.
73. International Conference on Harmonization (1999), *Harmonised Tripartite Guideline: Specifications: Test Procedures and Acceptance Criteria for Biotechnological/Biological Products*, ICH-Q6B, available: http://www.fda.gov/cder/guidance/index.htm.
74. U.S. Food and Drug Administration (FDA)/CDER/CVM (2001), Guidance for industry—Bioanalytical method validation, available: http://www.fda.gov/cder/guidance/index.htm.
75. CX/MAS 01/4 (2001), Codex Alimentarius Commission, Codex Committee on Methods of Analysis and Sampling, Criteria for evaluating acceptable methods of analysis for Codex purposes, agenda item 4a of the 23rd session, Budapest, Hungary, Feb. 26–Mar. 2, 2001.
76. CX/MAS 02/5 (2002), Codex Alimentarius Commission, Codex Committee on Methods of Analysis and Sampling (FAO/WHO), Criteria for evaluating acceptable methods of analysis for Codex purposes, agenda item 4b of the 24th session, Budapest, Hungary, Nov. 18–22, 2002.
77. CX/MAS 02/11 (2002), Codex Alimentarius Commission, Codex Committee on Methods of Analysis and Sampling (FAO/WHO), Requirements for single-laboratory validation for Codex purposes, agenda item 8b of the 24th session, Budapest, Hungary, Nov. 18–22, 2002.
78. CX/MAS 02/12 (2002), Codex Alimentarius Commission, Codex Committee on Methods of Analysis and Sampling (FAO/WHO), Validation of methods through the use of results from proficiency testing schemes, agenda item 8c of the 24th session, Budapest, Hungary, Nov. 18–22, 2002.
79. International Laboratory Accredidation Cooperation (ILAC) (2000), Guidelines for the requirements for the competence of providers of proficiency testing schemes, ILAC-G13 ILAC Technical Accreditation Issues Committee, available: www.ilac.org.
80. International Laboratory Accredidation Cooperation (ILAC) (2001), Guidance for accreditation to ISO/IEC 17025, ILAC-G15, ILAC Technical Accreditation Issues Committee, available: www.ilac.org.
81. International Laboratory Accredidation Cooperation (ILAC) (2002), The scope of accreditation and consideration of methods and criteria for the assessment of the scope in testing, ILAC-G18, ILAC Technical Accreditation Issues Committee, available: www.ilac.org.
82. European Commission (EC) (2002), Commission decision 2002/657/EC implementing council directive 96/23/EC concerning the performance of analytical methods and the interpretation of results, *Off. J. Eur. Commun.*, L 221/8, 17.8.2002.
83. Holst-Jensen, A., and Berdal, K. G. (2004), The modular analytical procedure and validation approach and the units of measurement for genetically modified materials in foods and feeds, *J. AOAC Int.*, 87, 1–9.

84. Wood, R. (1999), How to validate analytical methods, *Trends Anal. Chem.*, 18, 624–632.
85. Waters Corporation (2004), Validation Guidelines: Terminology and Definitions, available: http://www.waters.com/WatersDivision.
86. Hibbert, D. B. (1999), Method validation of modern analytical techniques, *Accred. Qual. Assur.*, 4, 352–356.
87. Fleming, J., Neidhart, B., Albus, H., and Wegscheider, W. (1996), Glossary of analytical terms (III), *Accred. Qual. Assur.*, 1, 135.
88. Krull, I. S., and Swartz, M. (1999), Analytical method development and validation for the academic researcher, *Anal. Lett.*, 32, 1067–1080.
89. Quevauviller, Ph. (2004), Traceability of environmental chemical measurements, *Trends Anal. Chem.*, 23, 171–176.
90. Holcombe, G., Lawn, R., and Sargent, M. (2004), Improvements in efficiency of production and traceability for certification of reference materials, *Accred. Qual. Assur.*, 9, 198–204.
91. Lauwaars, M., and Anklam, E. (2004), Method validation and reference materials, *Accred. Qual. Assur.*, 9, 253–258.
92. Fleming, J., Albus, H., Neidhart, B., and Wegschieder, W. (1996), Glossary of analytical terms (IV), *Accred. Qual. Assur.*, 1, 191.
93. Thompson, M., Ellison, S. L. R., Fajgelj, A., Willets, P., and Wood, R. (1999), Harmonised guidelines for the use of recovery information in analytical measurement, *Pure Appl. Chem.*, 71, 337–348.
94. Vessman, J. (1996), Selectivity or specificity? Validation of analytical methods from the perspective of an analytical chemist in the pharmaceutical industry, *J. Pharm. Biomed. Anal.*, 14, 867–869.
95. Analytical Methods Committee (1987), Recommendations for the definition, estimation and use of the detection limit, *Analyst*, 112, 199–204.
96. Fleming, J., Albus, H., Neidhart, B., and Wegschieder, W. (1997) Glossary of analytical terms (VII), *Accred. Qual. Assur.*, 2, 51–52.
97. Huber, W. (2003), Basic calculations about the limit of detection and its optimal determination, *Accred. Qual. Assur.*, 8, 213–217.
98. Kuselman, I., and Sherman, F. (1999), Assessment of limits of detection and quantification using calculation of uncertainty in a new method for water determination, *Accred. Qual. Assur.*, 4, 124–128.
99. European Commission (EC), Council directive 96/23/EC of April, 29 1996, on measures to monitor certain substances and residues thereof in live animals and animal products and repealing directives 85/258/EEC and 86/469/EEC and decisions 89/187/EEC and 91/664/EEC, *Off. J. Eur. Commun.*, L125, 0010–0032, 23.05.1996.
100. European Commission (EC), Commission decision 2003/181/EC amending decision 2002/657/EC as regards the setting of minimum required performance limits (MRPLs) for certain residues in food of animal origin, *Off. J. Eur. Commun.*, L 71/17, 0017–0018, 15.3.2003.
101. Antignac, J.-P., Le Bizec, B., Monteau, F., and Andre, F. (2003), Validation of analytical methods based on mass spectrometric detection according to the "2002/657/EC" European decision: Guideline and application, *Anal. Chim. Acta*, 483, 325–334.
102. De Wasch, K., De Brabander, H. F., Courtheyn, D., Van Hoof, N., Poelmans, S., and Noppe, H. (2003), The commission decision 2002/657/EC: Disscussion on some new analytical aspects. EURO FOOD CHEM XII, in Strategies on Safe Food, Proceedings, Vol. 1, Brugge, Belgium, Sept. 24–26, 2003, pp. 45–48.

103. von Holst, C., Müller, A., Björklund, E., and Anklam, E. (2001), In-house validation of a simplified method for the determination of PCB's in food and feedingstuffs, *Eur. Food Res. Technol.*, 213, 154–160.
104. Prichard, E., Albus, H., Neidhart, B., and Wegscheider, W. (1997), Glossary of analytical terms (IX), *Accred. Qual. Assur.*, 2, 348–349.
105. van der Voet, H., Van Rhijn, J. A., Van de Wiel, H. J. (1999), Inter-laboratory, time, and fitness-for-purpose aspects of effective validation, *Anal. Chim. Acta*, 391, 159–171.
106. Horwitz, W., and Albert, R. (1996), Reliability of the determinations of polychlorinated contaminants (bifenyls, dioxins, furans), *J. AOAC Int.*, 79, 589–621.
107. Thompson, M., and Lowthian, P. J. (1993), Effectiveness of analytical quality control is related to the subsequent performance of laboratories in proficiency tests, *Analyst*, 118, 1495–1500.
108. King, B., Boley, N., Kannan, G. (1999), The correlation of laboratory performance in proficiency testing with other QA characteristics, *Accred. Qual. Assur.*, 4, 280–291.
109. The Joint European Standard Institution (1989), *General Criteria for the Operation of Testing Laboratories*, EN45001, CEN/CENELEC, Brussels.
110. Pritzkow, J. (2003), Practical experience of the laboratories in implementing the ISO/IEC 17025, *Accred. Qual. Assur.*, 8, 25–26.

# VALIDATION OF LABORATORY INSTRUMENTS

Herman Lam
*Wild Crane Horizon, Inc., Scarborough, Ontario, Canada*

**Contents**

1 Introduction
2 Scope
3 Laboratory Instrument Classifications
4 Validation Phases
  4.1 Planning and Requirements Phase
  4.2 Qualification (Testing) Phase
  4.3 Operational Phase
  4.4 End of Life
5 Summary
  References

## 1 INTRODUCTION

The reliability of chemical and physical measurements is critically dependent on the suitability and performance of the instruments from which the measurements are obtained. It is a challenge for any laboratory to develop a pragmatically structured validation program for laboratory instruments of varying complexity. However, implementation of such a program is highly valuable as it provides assurances that instruments meet performance requirements and are suitable for their intended use. These assurances are required to comply with the good manufacturing practices (GMPs) and good laboratory practices (GLPs).

In recent years, tremendous efforts have been put forth by many organizations to address the validation of a wide variety of laboratory instruments of varying complexity. Instrument validation–related topics have been discussed at great length

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

in meetings and conferences. Many guidance documents and reference literatures on the topic of the validation of laboratory instruments are available [1–11]. Several key concepts and a common framework for the validation of laboratory instruments have been consolidated into the form presented herein:

1. The fundamental purpose of laboratory instrument validation is to provide assurances that the instrument is suitable for its intended use. The assurance is supported by documented evidence that the system consistently performs according to predetermined specifications for its intended applications.
2. The validation of laboratory instruments is the responsibility of the user's organization and not the suppliers of the instruments. The users can partner with the suppliers to facilitate installation and validation testing.
3. Most of the instruments used in the laboratory are commercial off-the-shelf (COTS) instruments, and consequently the users have little or no input into their design. A full system development life-cycle (SDLC) approach [8], which is used to develop complex computerized systems such as Laboratory Information Management System (LIMS) or Chromatographic Data System (CDS) or custom design laboratory equipment, is not appropriate for COTS instruments. Some laboratory instruments such as a pH meter or centrifuge are fairly simple and therefore do not warrant the SDLC approach.
4. Scalable validation for laboratory instruments is being developed to provide appropriate levels of coverage for laboratory instruments of various complexities for their intended use.
5. Laboratory instruments are classified into different categories according to their complexity to enable a scalable validation approach. The expected validation deliverables and documentation requirements for each class of laboratory instrument can be defined in the validation guide for the laboratory.
6. Validation efforts should be commensurate with the complexity of the different classes of instruments as defined in the validation guide developed for the laboratory.

## 2 SCOPE

This chapter provides an overview of laboratory instrument validation for GMPs and GLPs. The term GxP is used as a general referral to GMP and GLP. The focus of this chapter is on the validation of COTS instruments using a scalable approach.

The validation of large-scale computerized systems such as LIMS, chromatographic data systems, and custom-designed systems (bespoke systems) are beyond the scope of this chapter. Readers are encouraged to consult *GAMP (Good Automated Manufacturing Practice) Guide for Validation of Automated System in Pharmaceutical Manufacturing*, Version 4[8] for the validation for those systems.

## 3 LABORATORY INSTRUMENT CLASSIFICATIONS

A practical way to provide an appropriate level of validation for a wide range of laboratory instruments is to develop a classification system to categorize the laboratory instruments according to their functional complexity, usage, and risk to data integrity. For each category or class of instruments, a set of predetermined scalable deliverables and testing is defined. These definitions will in turn be used to determine the extent and type of testing required to validate various instruments. This instrument classification system and its associated deliverables provide simple and structured guidance on deciding the validation effort required for a whole spectrum of laboratory instruments from a simple pH meter to a fully automated dissolution workstation.

An instrument classification system which is based on the complexity of the instrument is given in Table 1. There are similar laboratory instrument classifications outlined in *GAMP: Good Practice Guide: Validation of Laboratory Computerized Systems* [10], and the U.S. Pharmacopeia (USP) general chapter on analytical instrument qualification [11]. As there can be overlap between categories of instrument classifications, individual organizations should make modifications to the definitions and the types of instruments in the classification system as required, thereby customizing the classification system to their respective operation. After the laboratory instrument classification has been established, the deliverables required for the validation for each class of instruments will be defined to provide the framework for the validation process. Table 2 lists the deliverables commonly encountered in the laboratories.

**TABLE 1 Instrument Classification**

| Category[a] | Definition | Examples[b] |
|---|---|---|
| A | Equipment/tools used for sample preparations | Hotplates, stirrers, shakers |
| B | Basic firmware instruments used to measure fundamental physical parameters such as weight, dimension, temperature, and pH | Balances, pH meters, digital thermometers, centrifuges, sonicators |
| C | Advanced firmware instruments that utilize spectrometry, chromatography, dissolution, etc. | Liquid chromatographs, gas chromatographs, dissolution baths, Karl Fischer titrators |
| D | Commercial COTS controlled by external computer<br>Hybrid systems such as automated dissolution workstation with high-performance liquid chromatography (HPLC) or ultraviolet–visible (UV–Vis) interface | Liquid chromatographs, gas chromatographs, UV/Vis spectrophotometers, Fourier transform infrared (FTIR) spectrophotometers, near-infrared (NIR) spectrophotometers, mass spectrometers, atomic absorption spectrometers, thermal gravimetric analyzers, COTS automation workstations |

[a]Categories may vary. Consult the user's organization laboratory instrument classification guide or references in the GAMP guide and USP general chapter ⟨1058⟩ [11, 12].
[b]Examples are not all inclusive.

**TABLE 2 Key Instrument Validation Deliverables**

| | By Category | | | |
|---|---|---|---|---|
| Deliverable | A | B | C | D |
| Business requirements[a] | ✓ [b] | ✓ [b,c] | ✓ | ✓ |
| Compliance assessment (GxP criticality assessment) | ✓ [b] | ✓ [b,c] | ✓ | ✓ |
| Validation plan | ✓ [b] | ✓ [b] | ✓ [c] | ✓ |
| Electronic records/electronic signatures (ER/ES) assessment | ✗ | ✗ | ✗ | ✓ |
| User and functional requirements (design qualification) | ✓ [b] | ✓ [a] | ✓ [c] | ✓ |
| Installation qualification | ✓ [b] | ✓ [c] | ✓ [c] | ✓ |
| Operational qualification | ✓ [b] | ✓ [c] | ✓ [c] | ✓ |
| Performance qualification | ✓ [b] | ✓ [c] | ✓ [c] | ✓ |
| Validation report | ✗ | ✓ [c] | ✓ [c] | ✓ |
| Traceability matrix and configuration documents | ✗ | ✗ | ✗ | ✓ |
| Standard operating procedures | ✓ [b] | ✓ | ✓ | ✓ |
| Performance verification | ✗ | ✓ | ✓ | ✓ |
| Periodic review | ✗ | ✓ [b] | ✓ | ✓ |
| Decommissioning | ✓ [b] | ✓ | ✓ | ✓ |

*Note*: ✓: required; ✗: not required.[a]Supplier audit may be required for some category D instruments.
[b]Optional.
[c]Simplified standard forms can be used.

## 4 VALIDATION PHASES

Conceptually, a life-cycle approach to laboratory instrument validation can be divided into four phases, as shown in Figure 1.

### 4.1 Planning and Requirements Phase

The activities in the planning phase are commenced after a business need to implement an instrument in the laboratory has been identified. Planning-phase activities include concisely documenting the user requirements and functional requirements of the instrument. Once established, the requirements are then used to evaluate the instrument candidates available from suppliers. A validation plan with deliverables that are commensurate with the class of the instrument being implemented should be developed during the planning phase.

***Validation Plan*** A validation plan is prepared to highlight the activities and deliverables required in the implementation of the laboratory instrument. The details of the validation plan should be scaled according to the complexity of the system, its intended use, and the impact on the business. A generic validation plan can be used for similar types of laboratory instruments with similar applications. For example, one validation plan can be prepared for the HPLC. A validation plan typically includes the following sections:

- *Objectives and Scope of Project* Rationale for implementing the instrument with its key applications and to state the assumptions, exclusions, and limitations of the project.

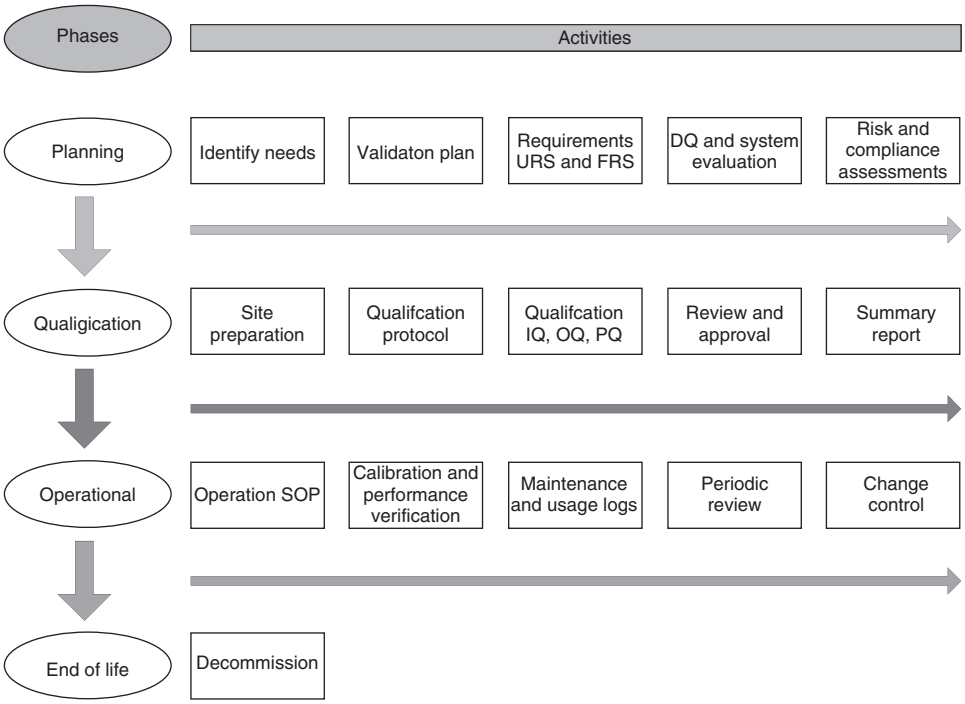


**FIGURE 1** Validation phases. URS: user requirement specification; FRS: functional requirement specification; DQ, IQ, OQ, PQ: design, installation, operational, and performance qualifications; SOP: standard operating procedure.

- *System Description* Hardware, software, and system configuration (if the system is comprised of different modules or components).
- *Compliance Assessment* State whether the use of the analytical instrument is subject to GxP and regulatory requirements. Instruments subjected to GxP will need validation. A rationale should be provided to support the decision of the assessment.
- *Validation Approach and Deliverables* Outline the required documents, qualification testing, and reports to be included in the validation project. The deliverables should be based upon the complexity and the risk associated with the intended use of the system. The deliverables will highlight the acceptance criteria to demonstrate that the system has met the requirements for its intended use.
- *Roles and Responsibilities* List the group(s) or personnel that are responsible for the validation of the instrument.
- *Project Timeline and Milestones* List the projected completion date of major tasks in the validation. This information is useful in monitoring the progress of the project and making sure proper resources are available at various stages of the validation.
- *Record Management* Describe the format, numbering sequence, revision, version control, and storage of the validation documentation.

- *References of Relevant Information* List the internal and external instrument validation guidance documents, compendial reference methods, qualification testing, and related SOPs for validation.

***User Requirements*** The user requirements outline the applications that the user wants the analytical instrument system to run and the major tasks that the instrument is required to execute in order to deliver the applications. The requirements should be clear, unambiguous, and verifiable or testable. The user requirements dictate the amount of validation that will be required. A successful validation should ultimately demonstrate that the user requirements can be met by the instrument and that the instrument can be used reliably for its intended applications. It is very important to state clearly what the instrument is supposed to do. Only the mandatory requirements should be validated. Including the "nice to have" or optional features in the user requirements will unnecessarily add more work to the validation.

For example, a simple HPLC system with an isocratic mobile-phase delivery system and a multiple-wavelength UV detector may be sufficient for simple routine product release testing. However, an HPLC system for running impurity assays or stability-indicating methods is likely to require a gradient mobile-phase delivery system and a diode array UV detector. If only the isocratic applications will be required for an HPLC system, there is no need to validate the gradient feature of the HPLC system even though the system is capable of running gradient applications. Another example is that of a dissolution bath that is solely utilized for running paddle dissolution methods. Although the bath may have the capability to run both basket and paddle methods, it is only the validation of the paddle method that should be considered necessary.

***Functional Requirements*** The high-level user requirements establish the framework for the functional specifications which identify the system functions, mode of operation, and operation environment necessary to meet the user requirements. Again, the requirements should be verifiable or testable to demonstrate that the requirements can be met during the qualification testing.

Functional requirements usually include the following considerations:

- Functions of each system hardware component or modules: The expected operational range of each function and its performance attributes such as the accuracy, precision, and linearity required for its intended use should be defined.
- Site requirements to support the operations of the instrument: Power supply (electrical voltage and current), ventilation, gas supply, water supply, and drainage.
- Health and safety requirements: Mechanical safety for robotic movement, radiation safety for systems that involve the use of radioactive sources, and laser safety for systems that use high-power lasers.
- Computer operation system and network requirements.
- Data type (analog and/or digital output/input) and capacity (size of the data files).

**TABLE 3 Typical Functional Requirements for HPLC with UV–Vis Detector**

| Modules | Functional Requirements |
|---|---|
| Pump | • The pump should be capable of a flow rate between 0.50 and 5.00 mL/min.<br>• The pump should be a quaternary gradient pump and have a compositional accuracy of ±1.5% of the theoretical values for the four channels.<br>• The pump should have relative standard deviation (RSD) of ≤2.0% for six successive readings from a calibrated flowmeter. |
| Autosampler | • The autosampler should have an injector capable of injecting sample volumes from 1 to 100 μL.<br>• The injector should have a precision of ≤1.5% RSD.<br>• The injector carryover should be ≤0.5% of peak area.<br>• The autosampler should pick the correct vial.<br>• The autosampler should use relays and/or contacts to communicate with the laboratory CDS.<br>• Temperature-controlled sample racks should be capable of maintaining samples in the temperature range of 4–15 °C (±3 °C). |
| UV–Vis detector | • The detectors should operate with a wavelength range from 200 to 800 nm.<br>• The detector should have a wavelength accuracy of ±2 nm.<br>• The detector response should be linear with a correlation coefficient $r^2$ not less than 0.999 over the full dynamic range.<br>• The linear range should be up to 2.0 Absorbance Unit Full Scale (AUFS). |
| Column compartment | • The column oven should be capable of maintaining a temperature range of 5 °C above ambient to 60 °C.<br>• The oven should be within ±3 °C of the set temperature and the temperature precision should be ≤2.0%. |

- Requirements for electronic records and electronic signatures [13–15]: System security, data integrity, traceability, and data archive/retrieval.
- Communication and control of other systems.
- System recovery from a major failure or disaster.

Typical hardware functional requirements for a gradient HPLC system with UV–Vis detection are listed in Table 3 as an example.

***Design Qualification*** For COTS analytical instruments, the user has little or no input into the design. The design information is usually not available to the users. In this case, the design qualification demonstrates the user requirements and the functional requirements can be fulfilled by the analytical instrument being considered by comparing the requirements against the technical specifications from the suppliers. For analytical systems which comprise of multiple COTS components (e.g., a dissolution system with online HPLC analysis capability), the configuration and compatible components from different suppliers should be demonstrated in the DQ.

***System Evaluation and Supplier Assessment*** Usually there are many choices and suppliers for common COTS laboratory instruments. The user requirements and the operational requirements will provide the basic criteria for the selection. Obviously the chosen instrument must be able to fulfill the key requirements for its intended use. Other factors concerning the instrument such as its ease of use, maintenance, and reputation of the suppliers in terms of quality, reliability, and support should be considered. From a practical point of view, a supplier audit may not be viable or necessary for commonly used COTS instruments. A supplier assessment is sometimes used to evaluate whether the supplier has a good-quality system in place to support the development and manufacturing of the instrument of interest. The need for a supplier assessment depends on the criticality and complexity of the system to be obtained.

### 4.2 Qualification (Testing) Phase

The terms *instrument qualification* and *instrument validation* are sometimes used indiscriminately. In this chapter, the term *qualification* refers to the site preparation and the testing employed to demonstrate that the instrument is properly installed in a suitable environment and the performance meets the predetermined specifications for its intended use. Qualification is a part of the whole validation life cycle. Validation refers to the process to provide assurance that the instrument is suitable for the intended application throughout the lifetime of the instrument. Installation qualification (IQ), operation qualification (OQ), and performance qualification (PQ) are performed to provide evidence that the user requirement specifications (URSs), functional requirement specifications (FRSs), and design qualification (DQs) have been met. The sequence of requirements setting and qualification events as well as the relationships between IQ, OQ, PQ and URS, FRS, and DQ are generally illustrated by the "V" diagram shown in Figure 2. Installation qualification demonstrates the fulfillment of the DQ. Similarly, OQ demonstrates the fulfillment of the functional requirements and PQ demonstrates the fulfillment of the user requirements.

***Site Preparation*** Prior to the installation of the instrument in the laboratory, all the preparations that support the operation of the instrument must be ready. The

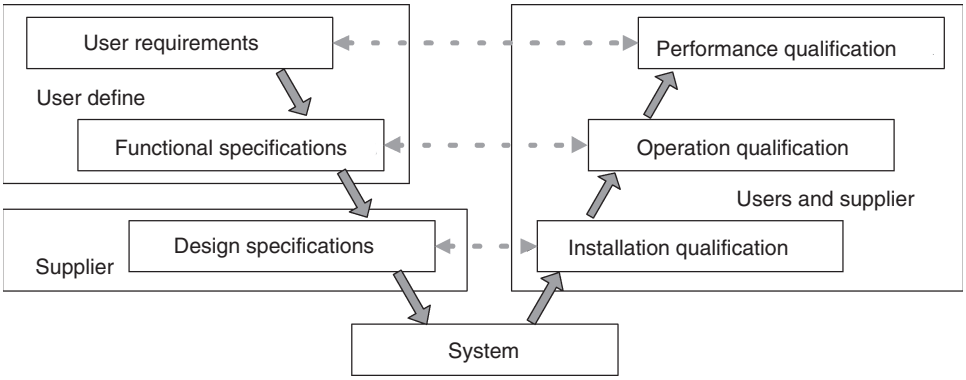


**FIGURE 2** Qualification "V" diagram.

supplier usually provides a site preparation document which outlines the information of the necessary facilities required to support the operations of the instrument. The site preparation document usually includes the following information:

- Physical dimensions and weight of the instrument
- Environmental conditions for proper operations: temperature, humidity, and vibration control
- Utilities: power supply, water, gas, drainage, ventilation, network connection
- Health and safety requirements

It is a common mistake to underestimate the effort and time required for site preparation. The users should carefully study the site preparation guide to ensure that the necessary preparations to house the new instrument in the laboratory are completed prior to installation. Inadequate site preparation can cause major inconveniences and long delays in the installation process. It is a waste of time and resources to have the service engineer show up in the laboratory but not able to do anything due to incomplete site preparation.

***Qualification Approaches*** If an analytical instrument is comprised of different functionally discrete modules, a modular approach to qualification testing that focuses on the specific operations of the individual module can be suitable for certain aspects of some operational testing (such as the flow rate precision and accuracy testing of a HPLC pump and the temperature accuracy column compartment).

When it comes to performance qualification, a holistic approach must be taken to test the analytical system with all the necessary modules working together to deliver the intended applications as specified in the user requirement (Figure 3). The proper functioning of each individual module of the analytical system does not

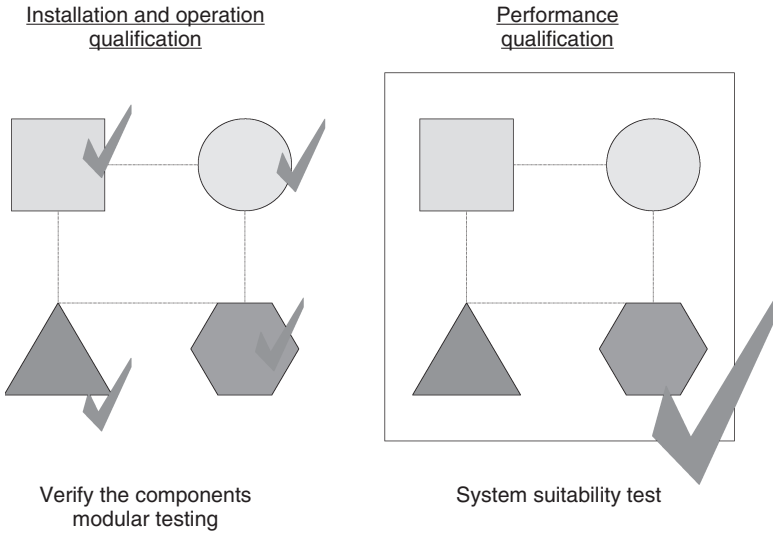


**FIGURE 3** Modular versus holistic approach for qualification.

infer proper functioning of the whole system with all the modules working in concert.

***Qualification Protocols*** For COTS instruments, the supplier often provides test protocols for installation and operation qualification. It is a common practice to purchase the qualification services from the supplier to execute the testing. The supplier should have a good knowledge of the instrument being installed and qualified, and therefore utilizing their services to perform these steps can provide significant time and resource savings. However, it is the responsibility of the user to review the test protocol for meaningful testing, that proper procedures will be used, and that reasonable acceptance criteria for the testing to ensure the objective of the testing are met. The supplier procedures and/or testing should be compatible with the procedures and practices of the user company. Acceptance criteria which are too loose do not provide enough assurance to demonstrate instrument performances. In contrast, acceptance criteria which are too tight can lead to unnecessary failures that cause a lot of effort and time to investigate and justify. The acceptance criteria should reflect both the operation requirements and the user requirements. The protocols must be approved before execution.

The test protocol typically has a general description of the system, the configurations, and the intended use. The test scripts in the test protocol provide detailed information of the testing procedures. In each test script, the following information should be provided:

- Description of the module to be tested
- Purpose and objective of the tests
- Scope and limitations
- Test procedure
- Acceptance criteria for the testing
- Sections to capture the test results
- A section to document deviation and exceptions encountered during the testing

***Installation Qualification*** Installation qualification provides documented evidence that the instrument was received and successfully installed in accordance with the approved design requirements and properly installed in an environment suitable for its operation. Proper installation is the first step to ensure that the instrument will function properly. An improperly installed instrument is likely to cause problems during the operational qualification and performance qualification. The following are some checks for the IQ process:

IQ checks before instrument is installed:

- Verify hardware and software against shipping list.
- Check for visible damage.
- Complete site preparation check list.
- Health and safety precautions.
- Add the instrument to the instrument inventory list.

IQ checks during installation:

- Document the location of the instrument.
- Document all the hardware components, including computer, printer, analytical instrument firmware, interfaces, and network connection..
- Document of all software applications, the operating system, and the storage location of actual software package.
- Document the system configuration.

IQ checks after installation:

- System powers up properly.
- Proper initialization and homing position.
- Proper communication between modules.
- Calibration of modules if necessary.
- Software version verification.
- Proper launching of the software and compatibility with hardware.
- Back up critical files for the system settings.
- Set up log book.

Availability of reference documents:

- Site preparation guide.
- System operation manuals.
- Sale and shipping documents.
- Factory testing records if available.
- Certificate of quality compliance from the vendor.

During system installation and qualification testing, use screen capture to print the information for evidence whenever possible as it is a more efficient and complete way to document the results and observations than writing down the information on paper.

***Operational Qualification*** Operational qualification provides documented evidence that the instrument will perform in accordance with its functional requirements throughout the representative operation range in a suitable environment. The OQ for simple instruments such as pH meters, balances, stirrers, water baths, and thermometers can simply achieved by execution of a calibration for the instrument. For moderately complex systems, OQ verifies the correct operation of the hardware and software for the instrument.

The testing of the hardware should cover the functionality of the instrument expected during normal operation. For example, the testing for an HPLC system would include the operation of the pump, the injector, and the detector [15].

Typical OQ tests for HPLC modules include:

- Pump: flow rate accuracy and gradient accuracy
- Detector: linearity of response, noise, drift, and wavelength accuracy
- Injector: precision, linearity, and carryover
- Column heater: temperature accuracy

Typical OQ tests for the UV–Vis spectrophotometers [16] include:

- Wavelength accuracy and reproducibility
- Stray light
- Resolution
- Photometric accuracy, reproducibility, and linearity
- Noise
- Baseline flatness and stability

In addition to testing the system components, a test of software functionality would be performed to test the system software operation and electronic records and electronic signatures (ERES) compliance (security, data integrity, data backup, and archive). In order to test the software functionality, a predetermined set of instructions can be entered step by step into the system. The system responses are then compared to the expected outcomes of the instruction and any problems with the execution are determined. Some vendors will provide a standard set of data which can be processed by the system to verify the data-handling capability of the system.

***Electronic Records and Electronic Signatures Considerations*** ERES compliance testing for computerized (personal-computer-controlled) instruments is required to demonstrate the functional requirements in the following three key areas [12–14]:

1. *Security* To prevent unauthorized access.
   - Control unauthorized access to the instrument through physical, logical, and procedural controls. Physical control is normally achieved by controlling access to the site, building, laboratory, and instrument. Logical security is provided by setting up different levels of access accounts such as user, supervisor, and system administrator and distinct passwords for each account user. Dual logical access controls are usually available from the applications software or the operating system. A procedural control using SOPs can be used to assign users access.
2. *Data Integrity* To demonstrate that records are valid and trustworthy. A record is a combination of raw data and the metadata (processing parameters plus other related information necessary to reconstruct the records):
   - Traceability: Helps to reconstruct the events.
   - Use of audit trails to track the activities for the creation, modification, or deletion of a record. It provides a record of who did what, wrote what, when, and why.
   - Use of secure date and time stamp for the creation, modification, and deletion of data.
3. *Data Preservation* To preserve a complete and accurate set of records by copying electronic data to a removable or remote storage medium and to manage the data retention. The data integrity must not be compromised during the preservation process:

- Data backup and restore: Active data are periodically copied from the hard drive or the personal computer controlling the instrument to a suitable medium such as CD-ROM or DVD or to a separate location such as a controlled file server.
- Data archive and retrieval: Data that are no longer required in the day-to-day operation of the instrument are archived in a retrievable manner.

***Performance Qualification*** Performance qualification is the process that provides documented evidence to demonstrate that the instrument has fulfilled the user requirements. Holistic testing which involves all the functional components in the system is required for the PQ testing. Performance qualification can be demonstrated by running a typical application that requires all the modules to function together as a whole system to deliver the intended application and the expected results.

If applicable, it is advantageous to execute a highly reliable test that is frequently performed by users on the particular type of instrument for the PQ. For example, the PQ test for a gradient HPLC system can be a gradient HPLC method with which the user has in-depth experience. In this case, the test results will mainly reflect the performance of the instrument and will not be affected by the uncertainty of the method used.

***Test Exceptions and Data Review*** In case the test results fail to meet the acceptance criteria, an investigation is required to determine the cause of the failure. The failures may be caused by execution errors of the test procedure or by instrument-related problems. Based on the outcome of the investigation, corrective actions can be taken to rectify the problem. Once the problem has been resolved, retesting can be executed to confirm that the instrument operations meet the requirements. The failure investigation, impact assessment, corrective actions, and retesting must be documented in the exception log of the qualification testing. After the qualification testing has been completed, the testing process, data generated, and results must be reviewed for accuracy, completeness, and justifiable deviations. All major issues with regard to compliance have to be addressed prior to releasing the instrument for production use.

***Summary Report*** After completion of the qualification activities, a summary report should be prepared to summarize all the qualification testing and results and deviations from the validation plan as well as include a conclusion to state whether the instrument is ready for its intended uses. Any deviations from planned activities including testing failures and requirements not met by the system should be addressed in the report. The summary report typically includes the following sections:

- Introduction
- System description
- Reference documents
- Validation activities, testing, acceptance criteria, and results
- Summary of deviations, problems, and mitigations

- Restrictions on the system (if any)
- Deliverables
- Conclusion

For simple analytical instruments, a simple table to summarize the qualification testing, acceptance criteria, results, and pass/fail decision of the tests will be sufficient since there are fewer tests that are required and the tests are usually relatively simple. For complex analytical systems, a more complex table often referred to as a traceability matrix which traces the requirements, testing, acceptance criteria, test results, and storage locations of the validation documents, test data, and other supporting documents is usually included in the summary report for easy reviewing and quick references.

After the actual qualification testing and regular performance verification testing, the documents and the related test data are the only proof that the instrument has gone through such testing and has been properly installed and maintained to support its intended applications. The document should be stored systematically in a centralized location and maintained with care in order to prevent any losses. A good recordkeeping system can be extremely useful in audit preparation and help to speed up the turnaround time for documents during an inspection.

### 4.3 Operational Phase

After an instrument has been qualified, it is ready for production use. The activities in the operational phase support the day-to-day use and maintain the instrument in a validated state.

***Standard Operating Procedure*** A SOP has to be written to provide instructions for the operation, maintenance, and calibration of the new instrument. A typical SOP should include:

- A general system description
- Operation instructions
- Responsibilities of the system users and system administrators
- Calibration or performance verification requirements, acceptance criteria, frequency of testing, and the actions required if the instrument does not meet the performance verification requirements.
- Maintenance requirements
- Service, major and minor repairs, and parts replacement that will necessitate a requalification of the instrument. For example, the replacement of a UV lamp in a UV detector does not require a full requalification, whereas a replacement of circuitry board will warrant full requalification.

***Maintenance*** Normal wear and tear as well as aging of various components may compromise the performance of the instrument or lead to operation failure. The instrument needs to be maintained in order to function consistently and reliably. A preventive maintenance program which identifies and replaces the consumable parts will likely save time and money in the long run. The usage and service records

are kept with the instrument to provide a performance history of the system. These records can provide clues to simplify troubleshooting in case of instrument failure and in addition are used to help develop a meaningful preventive maintenance schedule.

***Performance Verification and Calibration*** In order to maintain the instrument in a validated state, regular performance verification and calibration are required to demonstrate the instrument is functioning reliably according to a predetermined set of criteria to support the applications required in the user requirements. The current GMP (cGMP) requirements also dictate the calibration of instruments at suitable intervals in accordance with an established written program. The terms *calibration* and *performance verification* are very often used interchangeably. Calibration involves measuring and adjusting the instrument response using known standards. Performance verification verifies the operation and performance characteristics of an instrument against a predetermined set of requirements. Calibration can be considered as a part of performance verification.

The frequency of performance verification testing should be based on the knowledge of the operational reliability for the type of instrument and the type of operations that the instrument will be supporting. Initial frequency can be based on the recommendation from the supplier. A good performance record can justify less frequent verification. One potential drawback for overextending the period between performance verification can be the increased amount of impact assessment on the data generated since the last performance verification in case of system failure. Running system suitability testing before the analysis cannot replace the need for regular instrument calibration. System suitability testing is method specific whereas system calibration verifies the general performance of the instrument. The system suitability test only demonstrates that the instrument is suitable for a particular method at the time of analysis. It cannot reveal marginal performance of the system. For example, the system suitability test for an HPLC assay using UV detection is not likely to detect a wavelength accuracy problem since both the standards and the samples are analyzed at the same wavelength.

***Change Control*** Change control provides a structured mechanism for requesting, authorizing, evaluating, testing, implementing, and releasing changes to validated systems. It should be performed in accordance with approved and documented procedures. These procedures should include the following elements:

Prechange:

- Description of the proposed change
- Assessment of impact of a proposed change
- Reviewing and approval prior to execution
- Communicating changes to system users
- Management approval to implement the change

Postchange:

- Implementing the change
- Testing after the change is implemented

- Training and/or retraining system users
- Management approval to complete the change process

Typical change control forms to are shown in Figures 4 and 5.

The version of firmware may be inadvertently updated by the service engineer during service or routine maintenance without making the changes known to the system owners. The change may not have any impact on the operation of the instrument and may not be detected. However, the new firmware version is different from

| | | |
|---|---|---|
| Equipment make and type | Instrument identifier | |
| Current software | New software | |
| Operating system | Validation plan reference | |
| Description of change | | |
| Reason for change | | |
| Impact of change | | |
| Validation document references | | |
| Documentation that requires updating | | |
| Requalification necessary (attach protocols) | | |
| Request initiator | | Date |
| Mangement approval | | Date |
| QA approval | | Date |

**FIGURE 4** Prechange control form.

| | | |
|---|---|---|
| Equipment (make and type) | Instrument identifier | |
| Test results summary | | |
| Requalification reference | Reference | Date |
| Documentation reviewed | Reference | Date |
| Equipment records updated | Reference | Date |
| SOP reviewed | Reference | Date |
| System register updated | Reference | Date |
| Program validation data archived | Reference | Date |
| System ready for use | | Date |
| Request initiator | | Date |
| Management approval | | Date |
| QA approval | | Date |

**FIGURE 5** Postchange control form.

the version documented in the validation documents. The system owners should work with the service engineers to prevent this potential violation.

***Business Continuity Planning (Disaster Recovery)*** A disaster recovery plan should be in place to ensure the continued operation of the laboratory in case of an adverse event that renders the instrument out of commission and hence causes interruption to the business processes which the system supports. Adverse events like the failure of the critical hardware components of the instrument and the failure of the application software do happen in the day-to-day operation of a laboratory. The disaster recovery plan should provide the necessary steps to restore the systems back to a functional state. The steps typically include instructions to reinstall the application software to the personal computer controlling the instrument, to reconfigure the instrument, and to restore the backup data to the instrument.

***Periodic Review*** The performance of the instrument should be reviewed on a regular basis, typically once every two to three years, to ensure the instrument is still reliable and that it continues to comply with the user requirements. The review should include:

- *Operation Environment* Changes such as temperature, humidity, and vibration that may impact the operation of the instrument.
- *Change Control* System configuration changes; hardware, firmware, and software change and its impacts.
- *Records* Usage, maintenance, services, performance verification testing, and location of the records.
- *Documentation* Validation documents are current and the requirements are being met by the instrument, SOPs, operation manuals, business continuity plan, and location of the documents.
- *User Training* Adequate training for a new version of software operation.

The outcome of the review will determine whether the instrument is maintained in a validated state. In the case that the records indicated the instrument is more prone to certain types of failure, a preventive maintenance program may be desirable to avoid system failure during operation.

### 4.4 End of Life

The decommissioning of an instrument is the last step in the validation life cycle. When an instrument is no longer required in the laboratory and is ready for retirement, the following activities and related recordkeeping are required:

- Document in the instrument binder or logbook the reason for decommissioning and the effective date.
- Archive all related records such as the instrument binder or logbook, software, and manuals if no longer required in the laboratory.
- Ensure that all electronic records are archived following established procedures.

- Disconnect all building services from the instrument.
- Make arrangement for the removal of the instrument.
- Withdraw or amend any affected SOPs.
- Update the calibration program and the inventory list.

## 5 SUMMARY

The fundamental purpose of validating laboratory instruments is to provide assurance that the instrument is suitable for its intended use. The validation effort associated with a laboratory instrument should be commensurate with the complexity of the instrument, its intended use, and the impact of the data generated. A scalable validation approach to manage the whole life cycle of the instrumentation in the laboratory, from the planning stage to decommissioning, is a good business practice for smooth laboratory operation and will avoid preventable instrument failures. A systematic approach to instrument validation based on sound scientific rationales balancing the potential risks and business affordability can convey confidence to the auditors during laboratory inspections.

## REFERENCES

1. Freeman, M., Leng, M., Morrison, D., and Munden, R. (1995), Position paper on the qualification of analytical equipment, *Pharm. Technol. Eur.*, November 1995, 40.
2. Huber, L. (1995), *Validation of Computerized Analytical Systems*, Interpharm Press.
3. Huber, L. (1996), Quality assurance and instrumentation, *Accred. Qual. Assur.*, 1, 24.
4. Huber, L. (1999), *Validation and Qualification in Analytical Laboratories*, Interpharm Press.
5. International Organization for Standardization (ISO)/IEC 17025 (1999), General requirements for the compliance of testing and calibration laboratories, ISO, Geneva.
6. Miller, J. M., and Crowther, J. B. (2000), *Analytical Chemistry in a GMP Environment—A Practical Guide*, Wiley Interscience, Hoboken, NJ.
7. International Conference on Harmonization (ICH) (2000), Quality guideline, Q7A, ICH, Geneva.
8. International Society for Pharmaceutical Engineering (ISPE) (2002), *GAMP (Good Automated Manufacturing Practice) Guide for Validation of Automated System in Pharmaceutical Manufacturing*, Version 4, ISPE.
9. Lam, H. (2004), Procurement qualification and calibration of laboratory instrument: An overview, in Chan et al., Eds., *Analytical Method Validation and Instrument Performance Verification*, Wiley Interscience, Hoboken, NJ, Chapter 9.
10. International Society for Pharmaceutical Engineering (ISPE) (2005), *GAMP (Good Automated Manufacturing Practice) Good Practice Guide: Validation of Laboratory Computerized Systems*, ISPE.
11. U.S. Pharmacopoeia (USP) (2006), general chapter ⟨1058⟩, Analytical instrument qualification, USP, Rockville, MD.
12. U.S. Food and Drug Administration (FDA) (2005), Guidance for industry, 21 CFR Part 11: Electronic records and electronic signatures, FDA, Rockville, MD.

13. International Society for Pharmaceutical Engineering (ISPE) (2002), *Good Practice and Compliance for Electronic Records and Signatures, Part 1, Good Electronic Records Management*, ISPE and PDA.
14. International Society for Pharmaceutical Engineering (ISPE) (2005), *GAMP (Good Automated Manufacturing Practice) Good Practice Guide: A Risk-Based Approach to Compliant Electronic Records and Signatures*, ISPE.
15. Lam, H. (2004), Performance verification of HPLC, in Chan et al., Eds., *Analytical Method Validation and Instrument Performance Verification*, Wiley Interscience, Hoboken, NJ, Chapter 11.
16. Lam, H. (2004), Performance verification of UV—vis spectrophotometers, in Chan et al., Eds., *Analytical Method Validation and Instrument Performance Verification*, Wiley Interscience, Hoboken, NJ, Chapter 10.

# PHARMACEUTICAL MANUFACTURING VALIDATION PRINCIPLES

E. B. Souto,[1,3] T. Vasconcelos,[2,3] D. C. Ferreira,[3] and B. Sarmento[3]

[1]*Free University of Berlin, Berlin, Germany*
[2]*Laboratory of Pharmaceutical Development, BIAL, S. Mamede do Coronado, Portugal*
[3]*Faculty of Pharmacy, University of Porto, Porto, Portugal*

**Contents**

1 Introduction
2 Scope of Validation Processes
3 Validation Master Plan
4 Validation Protocols and Reports
  4.1 Validation Protocols
  4.2 Validation Reports
5 Facilities Validation
  5.1 Generalities
  5.2 Design of Facilities
6 Manufacturing Process Validation
7 Analytical Methods
8 Equipment and Computer Systems
  8.1 Equipment Systems
  8.2 Computer Systems
9 Cleaning Validation
10 Conclusions
  References

## 1 INTRODUCTION

The pharmaceutical industry has been a pioneer in the development of quality and safety procedures assuring that the risk of its work is reduced to a minimum. The

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

establishment in the last century of good manufacturing practices (GMPs) was a small step in the pharmaceutical work but a giant step in the reduction of risk to patients and operators and financial losses in pharmaceutical industry. Furthermore, to assure a GMP, one of the most important means is the validation procedure. Validation aims to confirm that quality is assured step by step in the process and not just at the end of the pharmaceutical procedures. Generally, an entire process is validated and every step is verified. Validation procedures are intended to ensure that all the different parts, steps, and components of a process are well defined and controlled in order to guarantee that the final product will not change over time. In general, validation is the process of checking if something satisfies a certain criterion and providing accurate and documented evidence that a process or system, when operated within established parameters, can effectively and reproducibly be performed while meeting its predetermined specifications and quality attributes. Therefore, validation is an integral part of quality assurance, involving the methodical study of systems, facilities, and processes aimed at determining whether they perform their intended functions adequately and consistently as specified. When validated the process will provide a high degree of uniform batch assurance, meeting the required specifications and therefore being formally approved. Validation does not improve processes but confirms that these have been properly developed and are under control. It is not only a requirement by the regulatory authorities but also an improvement to the pharmaceutical industry. The pharmaceutical industry benefits with the validation procedures since decreases the risk of problems and thus assure the smooth running of the processes. Also, validation procedures contribute to decreasing the risk of defect costs, the risk of regulatory noncompliance, and in process controls and end-product testing.

Validation can also be defined as documenting that any procedure, process, and activity consistently lead to the expected results, including the qualification of systems and equipment. The processes to be validated must be defined in a validation master plan (VMP) that consists of an approved written plan of objectives and actions stating how and when a company will achieve compliance with the GMP requirements regarding validation. The validation procedure must be well defined in a validation protocol, which is a written plan of actions defining how the process validation will be conducted, specifying as well who will conduct the various tasks and describing the testing parameters. The validation protocol must also contain sampling plans, testing methods, specifications of product characteristics, and equipment to be used. Moreover, it must specify the acceptance criteria and who will sign/approve/disapprove the conclusions derived from such a scientific study. The validation report is the final product of a validation procedure and must contain the results and interpretation of the tasks defined in the validation protocol.

Validation procedures must be applicable to computer systems; cleaning processes; manufacturing processes heating, ventilation, and air-conditioning systems; water systems; and analytical methods and equipment.

The process validation provides a high degree of assurance that a specific process will consistently result in a product that congregates its predetermined specifications and quality characteristics. Process validation provides documented evidence that a process is capable of reliably and repeatedly rendering a product of the required quality. Water and air systems validations establish that the system is under control

over a long period of time. Analytical validation proves that the analytical procedure is suitable for its intended purpose. Equipment validation is applicable to critical equipment whose performance may have an impact on the quality of the product guaranteeing that it works in a suitable manner. Computer validation is intended to provide a high degree of assurance that a computerized system analyzes, controls, and records data correctly and that data processing complies with predetermined specifications. Cleaning validation establishes that cleaning procedures are removing residues to predetermined levels of acceptability, taking into consideration factors such as batch size, dosing, toxicology, and equipment size.

To summarize, validation procedures are a tool to develop a product with good and reproducible characteristics, assuring quality throughout the lifetime and improving security for patients and the pharmaceutical industry.

## 2 SCOPE OF VALIDATION PROCESSES

The scope of validation has been widely stated and attempts to clearly define the competences of validation has been described [1]. It is now perfectly understood in both regulatory and compliance points that all laboratory processes, including facilities, equipment, analytical methods, and computer programs used in the analytical testing of pharmaceutical products, must be validated [2]. In the implementation of validation, it is fundamental to outline from the beginning the objectives throughout the validation process, which means that there should be proper preparation and planning before validation is performed as well as a specific program for validation activities.

There should be an appropriate and sufficient structured system, including organizational structure and documentation infrastructure, sufficient sufficient personnel and financial resources to perform validation tasks in a timely manner. Management and the personnel responsible for quality assurance should be involved in this discussion. Validation performance must be under the responsibility of appropriate and experience personnel that should represent different departments depending on the validation work to be performed.

Validation should be performed for new premises or equipment, utilities, or even systems, processes, and procedures at periodic intervals and when major changes have been made in accordance to written protocols. Validation can be prospective, concurrent, or retrospective, depending on when the validation is performed.

There should be a clear distinction between in-process controls and validation. In-process tests are performed during the manufacture of each batch using specifications and methods devised during the development phase. The aim is to monitor the process continuously and not exactly validate it. When a new manufacturing formula or method is adopted, steps should be taken to demonstrate its suitability for routine processing prior to the validation. The defined process using the materials and equipment specified should be shown to yield a product consistently of the required quality.

Essential validation work must therefore be identified to prove control of the critical aspects of the operations. A risk assessment approach should be used to determine the scope and extent of validation.

## 3 VALIDATION MASTER PLAN

Validation is intended to establish documented evidence with a high degree of probability that a process previously established will constantly perform according to the intended specified products.

According to the World Health Organization (WHO), the VMP is a high-level document that establishes an umbrella validation plan for the entire project and summarizes the manufacturer's overall philosophy and approach to be used for establishing performance adequacy. It provides information on the manufacturer's validation work program, including the manufacturer's intentions and methods used to establish the adequacy of performance of equipment, systems, and controls and the processes to be validated. The VMP is an approved document that provides the details and time scales for the validation work to be performed, including the responsibilities relating to the plan. VMPs may be defined as structured, detailed plans of work providing information about the overall manufacturer's philosophy, intentions, and approaches to be used to establish performance adequacy and how all of the validation work on a project will be controlled.

Although only a recommendation, the VMP must be, undeniably, the most significant document in any validation program. Throughout this document must clearly postulate the validation policy of each manufacturer, referring the organizational structure of all validation activities. The VMP supplies the main guidelines for the validation program, identifies the responsibilities of the personnel involved in the validation activities, identifies all aspects subject to validation, and expresses the nature and extent of testing on each point [3]. Other issues that must be mentioned in the VMP are a summary of existent facilities, systems, and equipment; a list of processes already validated and those intended to be; the planning and scheduling; the changing control; and references of existing documents prior to elaboration of the VMP. In this way, it is intended to be a retrospective, current, and prospective validation plan. All validation activities relating to critical technical operations relevant to product and process controls within a firm should be included in a VMP as well. This includes qualification of critical manufacturing and control equipment.

The VMP should be a summary document and should therefore be brief, concise, and clear. It should not repeat information documented elsewhere but should refer to existing documents such as policy documents, standard operating procedures (SOPs), and validation protocols/reports. The documentation format is illustrated in the VMP.

A VMP should be divided into chapters covering different subjects. First, an introduction should state the manufacture's validation policy, general description of the scope of those validation activities covered by the VMP, and their objectives, derivation, location, and schedule. Then, it must declare all validation activities and their organizational structure in terms of personnel responsibility for the VMP, validation protocols, validation work, report and document preparation projects, approval of the same validation protocols, reports in all stages of validation processes, and the training needs in support of validation. Other requirements of the VMP are cross references to other documents and to specific characteristics of the processes that are critical for yielding a quality product. Next, all validation activities comprised in the VMP should be summarized and compiled in a matrix format. Such a matrix should provide an overview and contain all items covered by the VMP that

are subject to validation describing the extent of validation required. It should include validation of analytical techniques which are to be used in determining the validation status of other processes or systems, validation approaches, revalidation activities, actual status, and future planning. Finally, the VMP must possess a general statement on key acceptance criteria for all validation activities. A planning section with detailed planning of subprojects must be included with an estimate of staffing, equipment, validation schedule of activities, and other specific requirements to complete the validation effort described. This time plan could be included in the above-mentioned matrix. A VMP requires regular updating. It must end with a statement of the enterprise's commitment to controlling critical changes to materials, facilities, equipment, or processes (including analytical techniques) as well as references and a glossary.

The VMP should contain references to the SOP's and information relevant to all aspects of the validation process. These actions are carried out in the GMP manufacturing process of a product and executed in accordance with written instructions if the manufacture of the product is intended to be GMP compliant. The list of relevant SOPs must be included in the VMP and should define how they must be validated. By planning and scheduling the validation procedure, the VMP defines the periodicity required to assure the validation statements. Validation of the different activities, facilities, and systems will occur in predetermined areas as defined by the VMP. The VMP should state who is responsible for preparing the VMP, preparing the protocols and SOPs, the validation work, the report and documentation preparation and control, approval of validation protocols and reports, tracking systems, and defining training needs.

The VMP helps not only the manufacturer and team members but also the inspection. It helps all members of the validation team to know their tasks, responsibilities of various groups during the validation of the equipment, and utility and validation program with respect to time, people, and money. It also helps auditors to understand the enterprise approach to validation and the setup and organization of all validation activities. Furthermore, the VMP is a living document updated and adjusted during the course of the project. Therefore, there are special changes that require revalidation, namely software changes, site and operational changes, and changes in the source of materials, the process, the equipment, production areas, and support systems.

## 4 VALIDATION PROTOCOLS AND REPORTS

### 4.1 Validation Protocols

A protocol is a written set of instructions broader in scope than a SOP. SOPs are the detailed written instructions for procedures routinely performed in the course of any of the activities associated with pharmaceutical manufacturing. A protocol describes the details of a comprehensive planned study to investigate the consistent operation of new system/equipment, a new procedure, or the acceptability of a new process before it is implemented.

Protocols include significant background information, explain the rationale behind and the aim of the study, and give a full description of the procedures to be followed. This means as well description of the site of the study; the responsible

personnel; the equipment to be used; the standards and criteria for the relevant products and processes; the type of validation, sampling, testing, and monitoring requirements; a description of how the results will be analyzed; and predetermined acceptance criteria for conclusive purposes. Validation, stability, and clinical studies are examples of written protocols for pharmaceutical manufacturers.

A validation protocol is a document that describes the item to be qualified, the tests and checks to be performed, as well as the results that are expected to be obtained. It is a file in which the records, results, and evaluation of a completed validation program are assembled. It may also contain proposals for the improvement of processes and/or equipment [1]. Validation protocols are important in ensuring that documented evidence is taken which demonstrates that an equipment item, a system, a process, or a method consistently performs at a specified level [4].

Validation protocols are required to describe the objective, methodology, and acceptance criteria for installation, operational, and performance qualifications. They are written to ensure test methods, and acceptance criteria are reviewed and approved before qualification of protocols. In practical terms, there are several stages for the production of protocols. First, an acceptable format needs to be agreed. No universal format exists for protocols, but to some extent, the type of equipment, the size of the project, and the personal preferences will dictate the protocol style. However, some norms have been established. Like other controlled documents, protocols are assigned unique reference numbers and revision numbers. They are titled and numbered on every page and have a particular place for approval signatures. Other common elements in protocols tend to be brief descriptions of the item being qualified and a clear statement of responsibilities.

Often the protocol will incorporate test sheets or sections for recording data. In this way, once the protocol has been executed, the document constitutes a record of the results and conclusion [3]. It must describe the activities to be performed in a validation, including the acceptance criteria for the approval of a manufacturing process or a part thereof for routine use [1].

A validation protocol is necessary to define the specific items and activities that will constitute any validation study. It is advisable for companies to draw up a VMP indicating the overall validation strategy for either the product range or equipment type or the entire site. The protocol must be prepared before initiation of the study and must either include or refer to the documentation required to provide information about a specific process, the parameters involved in that specific process, the personnel responsibilities, and the acceptance criteria.

The list of items to be qualified in the validation protocol must be produced and the approach agreed upon. There are several procedures to perform the validation protocols. Nevertheless, it is important that the approach agreed upon is executed to ensure internal consistency and prevent items being inadvertently forgotten. Thus, for instance, it is possible to group identical items together under one protocol or to have a protocol per item. A single or multiple protocols can be written covering installation qualification (IQ), operational qualification (OQ), and design qualification (DQ). For complex items such as large utilities, is it acceptable to address the generating system and the distribution system in separate protocols. Computer systems validation or control systems for process equipment can be documented within the mechanical validation protocol or in a separate protocol.

The protocol should be approved prior to its use. Furthermore, any changes to a protocol should be approved prior to implementation of the change. Once the approach and format are agreed upon, the protocol preparation can start. Preparing protocols requires individuals who are trained to extract information from a variety of sources and synthesize it into a coherent whole. Typically, sources of information include material requisition, specifications, data sheets, piping and instrument diagrams, manufacturer's literature, and equipment operating manuals. Significant protocols cannot be written until sufficient sources of information are available. Once the first draft is written, protocols require review. Usually, individuals from manufacturing, engineering, and quality assurance will perform the review. It is fairly normal for first-draft protocols to require considerable modification as this will be the first time that other interested parties have seen them. Rather than check everything, it is preferable that the review disciplines concentrate on their specialist areas. For example, manufacturing should evaluate if the proposed acceptance criteria are consistent with the process requirements; engineering should ensure that the list of instruments, major components, and utilities requirements are accurate; and quality assurance should ensure that compendial requirements have been met and that the protocol meets normal quality expectations.

In a large project, the coordination of protocol review, keeping track of the revision status, and the storage and retrieval of protocols are tasks that require careful planning. At some points it is probable that a number of protocols at different draft stages are circulating for review, some are being modified, new protocols are being prepared, and approved protocols are stored prior to execution. A tracking system should be implemented that can very quickly identify where a protocol is, providing a picture of the current protocol production status. The method and time allowed for review must also be agreed upon. Review methods range from traditional circulation of a single document to each individual for comment to assembling a team for a joint review to electrically accessing the document and revising or commenting on it. The method chosen must reflect the available time, resource, and technology. In some cases, the resources for protocol review are a limiting factor that can critically affect the schedule [3].

### 4.2 Validation Reports

A validation report is a written document that cross-references the validation protocol, summarizes the results obtained, describes any deviations observed, and draws the necessary conclusions, including recommending changes required to correct deficiencies for the qualification and validation performed [5]. In this report it is required to present both the results and conclusions and the secure approval of the study. The report should include a summary of the procedures used to clean, sample, and test as well as the physical and analytical test results or references for the same. The conclusions regarding the acceptability of the results should also be included. Other information would be the status of the procedures being validated, any recommendations based on the results, or any relevant information obtained during the study. These include, revalidation practices (if applicable), the approved conclusions, and any deviations of the protocol that might have occurred. In cases where it is unlikely that further batches of the product will be manufactured for a period

of time, it is advisable to generate temporary reports on a batch-by-batch basis until such time as the cleaning validation study has been completed [1].

Finally, the departments responsible for the qualification and validation work should approve the completed report and the conclusion of the report should state if the outcome of the qualification and/or validation was considered successful. The final review is performed by the quality assurance department, which gives the approval of the report according to the company's quality assurance system [1].

## 5 FACILITIES VALIDATION

### 5.1 Generalities

Facilities validation is related to the location, design, and construction of the plant to facilitate cleaning, maintenance, and operations in order to be appropriate for the type and stage of manufacture [6].

All facilities in a pharmaceutical industry must be designed and validated in order to assure the minimum risk of cross-contamination. Facilities validation must include the products and personnel flow, rooms design and cleaning, air and humidity systems, and water pipelines.

### 5.2 Design of Facilities

With regard to the design, the overall facilities should always be developed according to the most simplistic route of material flow and control of cross-contamination. Several layouts have therefore been described [7], with the aim of separating released materials from quarantined or rejected ones.

One of the most popular layouts consists of a center, or core, that has been conceived as a storage or warehouse area for raw materials, packaging components, and bulk stocks. Following this warehouse area immediately in the outer perimeter the manufacturing and packaging operations should be located to allow the flow of raw materials and components from the receiving and quarantine areas to approved storage. After materials are weighed into batch quantities, they are moved into the manufacturing area. When the manufacturing process is finished, the obtained products are placed in quarantine and then moved to bulk stock upon release. The packaging run follows when scheduled, and then the product and packaging components are delivered from the bulk stock and approved storage areas. An advantage of this layout is space conservation by virtue of having the supply areas close to the areas being supplied. Nonetheless, a significant disadvantage is the crossover of materials with the potential risk of contamination or mix-up.

An alternative layout design could be having the receiving, approved raw materials, components storage, and dispensing on one side and the manufacturing, quarantine, bulk stock, and packaging areas across a central corridor. Material flow from one area to another occurs in the same way as in the previous layout. However, in this case, flow is circular, eliminating much of the crossover described above.

To minimize contamination or mix-up, a third layout could be basic straight-line flow moving the materials along a critical path. The main advantage over the above-mentioned layouts is minimal crossover of materials, thus minimizing the potential

for contamination or mix-up. The additional space required to accommodate this configuration can be pointed to as the main disadvantage.

***Cross-Contamination Control*** Two parameters particularly useful in controlling the cross-contamination are the air-handling systems and dust collection.

Regarding the GMPs, air-handling systems are designed with the supply of outside air, moving on to the filtration systems that will be used, determining where positive and negative air pressures are required and whether to recirculate or exhaust spent 100% air, and finally to dust collection and exhaust systems.

Air filtration systems, including prefilters and particulate matter air filters, should be used when appropriate on air supplies to production areas. If air is recirculated of dust from production in areas where air contamination occurs during production, appropriate exhaust systems or other systems adequate to control contaminants should be requested.

A typical design involves one or more bag or cartridge filters located close to the area of dust generation. These coarse filtration devices should remove 95% of the dust generated from normal pharmaceutical manufacturing operations [8]. The prefiltered air is then mixed with 10–15% outside make-up air and passed through a high-efficiency particulate air (HEPA) filter and reenters the rooms through the supply plenum diffuser.

Regarding dust collection, the first area that must be addressed is the raw material sampling rooms and dispensing area. This area should be an enclosed facility with separate booths or hoods where the individual weighing or sampling can take place. These areas may be designed using horizontal laminar flow or appropriate hoods and other dust pickup devices. The supply air to these stations will therefore be HEPA filters either at the pickup stations or after the dust collector prior to returning to the general area or supply air.

The last area of dust collection must be the packaging. Some machines are designed with a self-contained vacuum system that returns the air, filtered through an absolute filter, back to the packaging area. There should also be some provisions made at the cottoning stations.

***Humidity and Temperature Control*** Humidity and temperature control systems are also important considerations as these impair both product protection and working environment comfort. Unless otherwise stated, 45% room humidity and 21°C are usually appropriate for critical manufacturing areas. Comfortable conditions should be provided for all operations. However, temperature controls should be such that little or no variation will be caused by external ambient temperatures. Thus, comfortable working conditions are achieved and there should be no impact on the characteristics of in-process materials. Warehousing operations should have adequate ventilation, particularly in areas of high storage, either pallet racks or pallet-to-pallet storage. The ventilation could be provided by large roof fans to circulate air. In addition, some form of supplemental air heating, such as hot-air blowers, should be provided for cold areas, such as shipping or receiving docks [7].

***Water Systems Control*** The supply of potable water in a plumbing system must be free of defects that could contribute contamination to pharmaceutical products.

Therefore, an effective water system is required. Nowadays, several techniques can be used to obtain water of high pharmaceutical quality. These include ionexchange treatment, reverse osmosis, distillation, electrodialysis, and ultrafiltration. However, there is no single optimum system for producing high-purity water, and selection of the final system is dependent on factors such as the quality of raw water, intent of its use, flow rate, and costs. In the pharmaceutical industry, the different water classes normally encountered are well water, potable water, purified water, and specially purified grades of water, such as water for injection (e.g., MilliQ water).

Water drawn directly from a well is called well water. The water may not be either chemically or microbiologically pure because it is untreated [7]. Therefore, the use of well water should be restricted to nonmanufacturing operations, such as lawn sprinklers, fire protection systems, and utilities.

Potable water is city water or private well water that has usually been subjected to some form of microbiological treatment, such as chlorine addition [7]. Potable water can be used in processing operations for cleaning and sanitation purposes. Periodic monitoring of use points should be conducted to ensure adequate residual chlorine levels and the absence of microbial contamination. Purified water is treated to attain specified levels of chemical purity and it is the type of water used in most pharmaceutical processing operations and final equipment cleaning. Purified water generally is produced by deionization or distillation, although reverse osmosis or ultrafiltration systems might be utilized if the required chemical purity could be achieved. Water softening or activated carbon filtration is frequently employed as a pretreatment process to remove calcium and magnesium ions or chlorine and organic materials. Ion exchange and demineralization through deionization is a very common method to obtain the purified water used in the pharmaceutical industry.

Deionization equipment should have proper size to allow frequent regeneration. A recirculation system should also be installed in the unit that approaches the rated flow of the deionization unit. Procedures should be written to ensure that all water treatment equipment is properly operated, monitored, maintained, and sanitized on a regular basis.

Regarding water filtration procedures first, prefilters should be provided to prevent large particulates from entering the system and microfiltration to remove bacteria. Prefilters are usually the replaceable cartridge type with porosities ranging as high as 25 μm. Microfiltration follows and is generally accomplished with 0.2-μm absolute filters, which will remove most bacteria.

After definition of the water type required, the water pipeline must be validated in order to ensure an adequate flow and purity (chemical and microbiological) of water. After the validation process, periodic verification of the pipeline and water collection points is required. This verification must be based on well-defined SOPs.

Sanitizing is best accomplished through several methods. After periods of low water usage, the system should be flushed with a supply of water that has residual chlorine. Periodic hyperchlorination and microbial control are also recommended. Microbial control can be achieved by storing the water at 80 °C. Alternatively, ultraviolet radiation can be applied.

***Pest Control*** Each manufacture, processing, packing, or holding area of a pharmaceutical product should be maintained clean and sanitary. This means that these

areas must be free of infestation by rodents, birds, insects, and others (other than laboratory animals). Trash and organic waste matter should be held and disposed of in a timely and sanitary manner.

A pest control program should be developed in order to assure the integrity and quality of products produced and comply with existing guidelines. The program should be written to include a general statement of the purpose and the company position. Effectiveness of the program should be assured by defining the plant individual with overall responsibility for the program and how the responsibility will be carried out. In addition, the extermination staff, whether they be in-house or subcontracted personnel, should have their training and experience requirements well defined. Assistance in supporting the program may be gained from other plant personnel by their pointing out problem areas.

This program should be treated to line management personnel. Furthermore, its content regarding the list of approved pesticides to be used in the plant should also be verified periodically. Basic information should be controlled, such as the trade name of the pesticide; classification; type of action; chemical name and concentration of active ingredient; effectiveness, usefulness, area of usage, mode, and frequency of application; toxicities and any specific toxic symptoms; status of government approval; and specific restrictions and cautions [7]. The development of sheets depicting such information will serve a twofold purpose. First, these sheets are subject to approval by the plant safety organization to determine if the materials comply with the Occupational Safety and Health Administration (OSHA) requirements and the requirements of other state or local agencies. Additionally, these sheets would also facilitate compliance with the GMP regulations. Written procedures for use of suitable rodenticides, insecticides, fungicides, fumigating agents, and cleaning and sanitizing agents should be provided in order to prevent the contamination of equipment, components, drug product containers, closures, packaging, labeling materials, or drug.

All manufacturing areas should be constructed using nonporous materials on the walls and floors. Any protrusions, such as pipes and electrical boxes, should be minimized. Space should be allocated carefully to provide sufficient rooms for all operations. There should be adequate lighting and the areas should be remote from any openings to the outside. Adequate training in understanding GMPs should be given to all personnel, including outside contractors.

Scheduled inspections and preventive treatments should be documented specifying the problem encountered as well as the special service that has shown effectiveness to the treatment.

***Manufacturing Rooms*** Manufacturing rooms must be well designed in order to ensure adequate cleaning and reduction of cross-contamination. Points of dust accumulation like 90° angles areas must be avoided in the room, dust collectors or air lines must be presented restrictly at the wall surface and only the minimum equipments must be presented in a room.

Air pressure inside the manufacturing rooms must be positive or negative depending on whether the product is a liquid or a powder, respectively, in order to avoid cross-contamination. It must be monitored by validated pressure systems.

***Packaging and Labeling Control*** Type of construction in the packaging area is very important because it is particularly sensitive to mix-ups in either product or labeling. Individual packaging lines should have minimum separation between each line. Depending on the equipment used, more space might need to be used. Consideration should be given to the separation of lines by partitions. The partitions should be closed at floor level to prevent migration of product and be high enough to prevent any crossover of product.

One possibility might be to start packaging operations with a prestaging area large enough to contain all required components for the packaging operation. This area is separate but adjacent to or in front of the filling area. The filling area should be equipped, if possible, with air filtered through absolute filters. Dust collector from the central dust collector system should be located in the filling area to remove dust from the filling area. The air circulation in the room must enter in the filling area and then be moved to the area where cottoning, capping, or labeling takes place. Thus, maximum product protection is exercised in the filling area while the product is exposed, with lesser degrees of control being required in the other areas because the product is in containers.

SOPs should be developed for all packaging operations. These include specific procedures for line setup, approval of line before start of operations, periodic line check during operation, close out of line, line clearance at end of operation, and reconciliation of product and components.

Labeling operations must also receive special attention. The use of cut labels in the pharmaceutical industry has already disappeared. Electronic label counting and verification are the norm, with bar codes, universal product codes (UPCs) or health industry bar codes (HIBCs) being used more frequently as label identifiers. The storage of labels is very important for both security and preservation reasons. The verification of labeling at the final label application point is also becoming more popular, since the the U.S. Food and Drug Administration (FDA) has identified errors.

SOPs must be developed for label accountability with specific tolerances spelled out. Provision must be made for label security, such as locked cabinets on operating lines. Accountability sheets for other components should be provided so that reconciliation between used and finished packages can be developed. If possible, a cleaning area and shop should be set up for the cleaning of filters and parts and the disassembly and assembly of filling machines.

## 6 MANUFACTURING PROCESS VALIDATION

The purpose of manufacturing process validation is to ensure that the manufacturing process for the drug product is fully controlled and capable of providing a product that complies with established quality specifications consistently and repetitively. This validation process covers manufacturing operations from formulation through packaging. Process validation is being conducted in support of launch of the drug product.

Process validation also establishes documented evidence that the manufacturing process, when executed according to the VMP and pertinent process SOPs, consis-

tently manufactures a drug product that meets all of the critical specifications required for its purpose. It is designed to ensure that all of the prerequisites are in place as well as to allow technical analysis of all applicable performance tests.

Nowadays the manufacturing validation tendencies tend to focus on pharmaceutical development. Therefore, the primary responsibility of the pilot plant is to ensure that the developed product is efficiently, economically, and consistently reproducible on a large scale. Since low production costs are of higher competitive advantage, attention must be paid to the production costs. Each operation unit should therefore be optimized. To avoid excessive production times, the manufacturing instructions transferred to the production department should be clearly written, readily understood, and unambiguous. Unless the economic advantage of purchasing new equipment has been proven to be necessary, in-house equipment should be preferentially used. Nonetheless, if international companies are intended to manufacture products at several sites, alternate manufacturing equipment and procedures might also be required.

The physical properties and specifications for the manufactured formulations should match those established earlier by the product formulator, the pilot plant staff, and the quality control department. Therefore, the product manufactured on a large scale should possess the appropriate quality properties.

Additional responsibilities for pilot plant staff are those related to the evaluation of new processing equipment, which aims to find causes and solutions for problems that occasionally might arise during the production. Since the department of pharmaceutical research and its development division are responsible for developing the formulations and the manufacturing processes for the dosage forms, the experience gained in the development of the manufacturing process will be of major importance during the validation process. Such validation conducted at a pilot scale should simplify large-scale validation.

It is clear that the aim of pharmaceutical research is to achieve zero defects and zero batch rejections, and this can be verified by process validation. One must bear in mind that exhaustive finished testing of product is not a substitute for in-process controls and process validation.

The validation process will consistently deliver product that has a direct relationship to a dosage form on which clinical efficiency and safety were determined. Thus, comparisons between the manufacturing process, the raw material used, in-process control, and finished-process test results are in order. During scale up, consideration of these factors at the development stage must be carefully performed to produce quality batches. Matching equipment from laboratory to production helps to eliminate possible validation problems afterward. The physicochemical characteristics of the active drug substances should be controlled. Special attention should be paid to products in which the drug substance comprises a very small percentage of the pharmaceutical dosage forms.

The physicochemical characteristics of raw materials play an important role in content uniformity and bioavailability. Therefore, bioavailability of the drug over time must be thoroughly investigated before any significant changes are made. Once the physicochemical properties of drug substances (e.g., particle size of raw materials) can influence the availability and clinical effect of a product, the main characteristics of raw materials should be considered in a validation program. The

characteristics of raw materials can vary among manufacturers, but also from lot to lot of the same manufacturer. Control of the physicochemical characteristics of excipients is also important and must be stated in the specifications.

In order to develop a reproducible manufacturing process, attention must be given to particular instructions and screening procedures. For instance, excipients should be free of lumps and proper screening will aid raw material dispersion. Additionally, one should specify the size and design of containers and all equipment to be used.

Changes in process decisions regarding the suitability of new manufacturing equipment that is intended to be employed should be performed by management and by production personnel. Equipment changes for some products might guarantee additional dissolution and/or content uniformity studies. The degree and depth of a study are largely dependent upon the specific product. For some products with very good historical data coming from a reproducible and controlled process, particular tests can be obviated. However, if potent dosage forms are to be developed, those are usually performed (e.g., sampling the mix in tablets technology) [9]. If products with lack of historical data are to be handled, additional finished product testing should be performed. For instance, with respect to tablet production dissolution testing and content uniformity testing are usually implied in the finished product tests.

The batches utilized in this type of approach for manufacturing process validation should be presented on different scales as the manufacturing process is developed. Laboratory-scale batches are of very small size (e.g., 100–1000 times less than production scale) and are produced at the research and early development laboratory stages, which are used to support the formulation and packaging development and clinical and/or preclinical studies. In the development pharmaceutics, data validation of laboratory-scale batches can contribute to the choice of the appropriate manufacturing process (evolution and definition of critical product performance characteristics). Pilot batches correspond to at least 10% or 100,000 units for tablets (the biggest) of the production batches and are used in the development or optimization stage to support stability studies. Pilot batches are used to provide data predictive of the production batches and therefore provide the link between process development and industrial manufacture of the pharmaceutical product. Finally, production-scale batches will be produced during routine marketing of the drug product.

Traditionally a proper process validation performance must be established based on the following considerations:

- All raw materials used for the validation batches must be approved by quality control (QC).
- Drug products are validated concurrently. Samples from three consecutive batches should meet all of the predetermined specifications without unexplained failures. If any of the three batches does not comply, the validation will be repeated up to three times to try to obtain three consecutive acceptable results. If not obtained, the tests will be suspended until the manufacturing process has been reviewed.
- The validation batches must use the regular production equipment and personnel.

- Critical in-process control parameters must be specified by ranges, and full-range monitoring is performed. The limit determined to be the "worst case" will be challenge at least twice and the other limit challenged at least once to complete the full-range monitoring.
- For critical process control parameters specified by fixed points, the value ought to be challenged within an acceptable tolerance, typically ±1 unit.
- All tests must be conducted by training and experienced technical personnel and must be documented in a scientific manner using the established format of the protocol.
- Any test function that does not have results which support the parameters defined in the approved protocol must be conclusively rationalized for their nonconformance and approved; otherwise, the qualification will be considered invalid.

Samples are to be taken during and/or after each critical manufacturing step. All control parameters for the manufacturing process have to be monitored and recorded. Each sample analysis will be performed in duplicate using validated or accepted pharmacopeia methods. The sample results will be used to confirm in-process and final product quality attributes as defined by the preestablished specifications. Conformance with specifications will justify the appropriateness of the critical parameters used during the process validation.

Validation data should be generated for all products to demonstrate the adequacy of the manufacturing process. The process validation data may not always be available, however, where the manufacturing process uses a nonstandard method of manufacture. Data demonstrating the validity of that method should be submitted in the marketing authorization file.

## 7 ANALYTICAL METHODS

The aim of validation of an analytical procedure is to demonstrate that the method employed in any product testing, such as the identification, control of impurities, assay, dissolution, particle size, water content, or residual solvents, is validated in the most important characteristics. Identification tests, quantitative tests for impurities content, limit tests for control of impurities, and quantitative tests of the active moiety in samples of pharmaceutical product are the most common types of analytical procedures that validation addresses [1].

However, other analytical procedures, such as dissolution testing for dosage form or particle size determination for drug substance, are required for validation of analytical procedures. The revalidation of an analytical procedure is possible when, in particular circumstances, it could show changes in the synthesis of the drug substance, the composition of the finished product, or the analytical procedure. However, certain other changes may require validation as well.

Method validation should confirm that the analytical procedure employed for a specific test is suitable for its intended use. The validation of an analytical method

is the process by which it is established by laboratory studies that the performance characteristics of the method meet the requirement for the intended application. This implies that validity of a method can be demonstrated only through laboratory studies. Methods should be validated or revalidated before their introduction and routine use whenever the conditions change for which the method has been validated (e.g., an instrument with different characteristics) and the method change is outside the original scope of the method.

Depending on the use of the assay, different parameters will have to be measured during the assay validation. Validation of analytical assays is the process of establishing one or more of the following as appropriate to the type of assay: accuracy precision (repeatability, intermediate precision), linearity, range, limit of detection, limit of quantification, specificity, and robustness [1]. For physicochemical methods there are accepted defined limits for these test parameters:

1. Specificity is the ability to assess unequivocally the analyte in the presence of components, which may be expected to be present and include impurities, degradants, matrices, and so on.
2. Accuracy is the replica of agreement between the value which is accepted as either a conventional true value or an accepted reference value and the value found.
3. Precision is the degree of agreement between a series of measurements obtained from multiple sampling of the same homogeneous sample under prescribed conditions. Precision may be considered at three levels: repeatability (precision under the same operating conditions over a short interval of time), intermediate precision (precision within-laboratory variations: different days, different analysis, different equipment, etc.), and reproducibility (precision between laboratories, collaborative studies, usually applied to standardization of methodology).
4. Detection limit is the lowest amount of analyte in a sample which can be detected but not necessarily quantified as an exact value.
5. Quantitation limit is the lowest amount of analyte in a sample which can be quantitatively determined with suitable precision and accuracy. The quantification limit is a parameter of quantitative assays for low levels of compounds in sample matrices and is used particularly for the determination of impurities and/or degradation products.
6. Linearity is the ability, within a given range, to obtain test results which are directly proportional to the concentration of analyte in the sample.
7. Range is the interval between the upper and lower concentration of analyte in the sample for which it has been demonstrated that the analytical procedure has a suitable level of precision, accuracy, and linearity.
8. Robustness is a measure of the capacity to remain unaffected by small but deliberate variations in the method parameters and provides an indication of its reliability during normal usage.

## 8 EQUIPMENT AND COMPUTER SYSTEMS

### 8.1 Equipment Systems

***Generalities*** With the advent of the industrial revolution in the eighteenth century, machinery became essential for our lifestyle. Machinery is one of most important key factors of our society and the same occurs in the pharmaceutical industry. In the middle of the twentieth century, many procedures in the pharmaceutical industry were hand manufactured, requiring a lot of time and manpower. Packaging was one of the most rudimentary tasks performed; an image of hundreds of persons seated at a table packaging pharmaceutical products was a normal view in the middle of the last century. In the second half of the twentieth century manpower was replaced by machinery. Blister machines, automatic capsule-filling machines, and bigger compression and drying equipment became common in industry, dramatically increasing output and profits. Since then, the equipment used in the manufacture of medications has shown great improvement. According to GMPs all equipment must be located, designed, constructed, adapted, and maintained to suit the operations to be carried out. Their layout and design must minimize the risk of errors, allowing effective cleaning and maintenance in order to avoid cross-contamination, buildup of dust or dirt, and, in general, any adverse effects on the quality of products [10]. Equipment and premises must improve the product quality and safety and should never create any risk to the product. Adequate design must be taken into account when new equipment is created in order to improve its efficiency and cleanness and reduce errors or breakdowns.

Equipment should be placed under adequate environmental conditions in order to function accurately and in such a way as to prevent any risk of error or contamination. The environment must show minimal risk of causing contamination of materials or products when considered together with measures to protect the manufacture [10].

Manufacturing equipment should be designed, located, and maintained to suit its intended purpose. Repair and maintenance operations should not present any hazard to the quality of the products [10]. It should be designed so that it can be easily and thoroughly cleaned. It should be cleaned according to detailed and written procedures and stored only in a clean and dry condition. Washing and cleaning equipment should be chosen and used in order not to be a source of contamination.

Production equipment should not present any hazard to the products. The parts of the production equipment that come in contact with the product must not be reactive, additive, or absorptive to such an extent that it will affect the quality of the product and thus present any hazard [10]. Equipment should not be made of materials that contaminate the final product; therefore manufacturing equipment is usually composed of stainless steel and polymeric materials that are easily cleaned. According to this, natural materials must be avoided.

Balances and measuring equipment of an appropriate range and precision should be available for production and control operations. Measuring, weighing, recording, and control equipment should be calibrated and checked at predetermined intervals by appropriate methods. Adequate records of such tests should be maintained [10].

All equipment for which calibration is applicable must be periodically calibrated. Periodicity of calibration must take into account the type of the equipment, risk assessment, and previous results. An extremely short calibration periodicity becomes extremely expensive, whereas an extremely long calibration periodicity could result in poor verification results. Poor verification results mean that the results generated are wrong and therefore all results obtained since the previous calibration must be reviewed. In conclusion, adequate evaluation of the calibration periodicity is essential.

All equipment must be well identified with name and code and must have a single operation instruction. A good operation instruction must contain at least the following points: (i) equipment description, containing the function of the equipment, name, code, serial number, model, manufacturer, accessories, dimensions, power source, connections, and component descriptions; (ii) function, containing a description of a procedure of how to handle the equipment, and all the steps required and options; (iii) calibration procedure, described in detail and providing the calibration periodicity; (iv) maintenance, a simple and easily maintenance procedures should be available; (v) cleaning described in detail; and (vi) security procedures.

***Equipment System Validation*** The GMP has special features regarding the equipment, setting several points for the validation, such as design qualification, installation qualification, operational qualification, and performance qualification. Each point must be given in a different document, so four documents must be generated for each equipment. Without any of these documents, the equipment cannot be considered adequate for GMP purposes. GMPs do not refer equipment validations; nonetheless, they refer these four items, and therefore, equipment can only be considered validated after the approval of these four documents.

Equipment validation is not reliable, and several special features are required instead. GMP is concerned with several procedures before the equipment is placed into the service. After entrance into service, maintenance and calibration must be made periodically. The following focuses on the procedures required for new equipment to enter into service.

*Design Qualification* When a pharmaceutical industry intends to purchase new equipment, it must first know and define what type of equipment and which specifications it must have. Specifications must be set taking into account specific requirements of the company or facility. All the specifications of the equipment must be reviewed. This is called design qualification. The requirements of the equipment must be defined in a specific document previous to the purchase by any pharmaceutical company. The document generated is used to justify the equipment selection from the various proposals.

*Installation Qualification* After equipment selection, it is necessary to assure that the equipment is installed well. The IQ document describes and validates the procedure of the equipment installation. It establishes confidence that the process equipment and ancillary systems are capable of consistently operating within established limits and tolerances [10]. The equipment manufacturer and pharmaceutical company must agree and check the IQ, which must be approved by the pharmaceutical company at the end. This document certifies that equipment was installed as specified by the manufacturer and the purchaser.

*Operational Qualification* The OQ document certifies that the equipment works as desired and defined by the manufacturer and the purchaser. An example is the acquisition of a new high-shear mixer/granulator where the paddle is put on rotation with a real calibrated rotation speed tester and, if the value obtained meets the specifications, the mixer paddle rotation passes the OQ test. If not, additional requalification must be performed. All the test results must be introduced and verified in the OQ report that is approved by the company at the end. The OQ document must describe several tests and related specifications to perform on the equipment in order to evaluate if it is working well, and the test to be performed must be described and approved by the manufacturer and the purchaser. Therefore, tests must be performed on the equipment, and for each one a description and signature of who performed and verified the test are required. Usually the tests are performed by the manufacturer and verified by the purchaser. These tests usually consist of evaluating if the mechanical and electric components of the equipment are working as desired.

*Performance Qualification* After evaluation of the way the equipment is working, it is required to perform some tests applying real situations to evaluate if the results are in agreement with those specified. It is called performance qualification (PQ) and establishes confidence through appropriate testing that the finished product obtained by a specified process meets all release requirements for functionality and safety [11].

Commonly when a pharmaceutical industry purchases a new compression machine, a PQ is conducted. First, the company must select some of its well-known products and prepare them until the compression phase as usual. The product must be compressed by the new equipment using the same compression conditions, such as compression force and tablet output. The obtained tablets must meet all the specifications of that product. The parameters specified could be aspect, hardness, thickness, diameter, average mass and uniformity of mass, friability, and tablet output. If the tablets obtained comply with specifications, the compression machine is considered reliable to obtain a product with good quality and in a reproducible manner and the equipment is considered performance qualified.

Also in the PQ a document must be created with the description of the tests to be performed and the related specifications. This document must be verified and approved before the test performance. After the tests, the results must be introduced in the document and the pharmaceutical company must approve the final version.

All the equipment qualification documents must define the equipment designation, tests to be performed, test specifications, materials, operators, reviewers, and responsibility for approval.

After approval of design qualification, installation qualification, operational qualification, and performance qualification, the equipment is considered adequate for GMP proposes and can be placed in service.

### 8.2 Computer Systems

***Generalities*** Technological advances are increasing extraordinarily, with the advent of computer system processes that were unthinkable a few years ago but nowadays are easy and simple. These advances are available every day in almost

everything. Particularly in the pharmaceutical industry the last decade has seen the development of new analytical equipment, such as near Infrared, Fourier transform infrared (FTIR), or Raman spectroscopy using complex software to perform modulation or high-performance liquid chromatography (HPLC) and mass spectrophotometry, which have easier and simpler software. All this evolution in analytical equipment allows the industry to produce better and have higher profits. The computational systems revolution has reached not only analytical equipment used in pharmaceutical industry but also manufacturing equipment. Nowadays in the pharmaceutical industry almost every operation uses a computer system. Therefore, all computer systems in the pharmaceutical industry are supposed to be validated in order to assure that the results they generate are accurate and precise. In addition, wherever a computerized system replaces a manual operation, there should be no resultant decrease in product quality or quality assurance. Consideration should be given to the risk of losing aspects of the previous system which could result from reducing the involvement of operators.

A computer system is composed of software and hardware, equipment, a processor, and a user, and it is used to execute a specific procedure. Regardless of whether the computer system is developed in-house or by a contractor or purchased off the shelf, establishing documented end-user requirements is extremely important for computer systems validation. Without first establishing end-user needs and intended use, it is virtually impossible to confirm that the system can consistently meet them. Once established, it should obtain evidence that the computer system implements those needs correctly and that they are traceable to system design requirements and specifications. It is important that the end-user requirements specifications take into account predicate rules [12].

***Computer System Validation*** All computer systems presently used in the pharmaceutical industry need to be evaluated for risk assessment. The first task in the validation of a computer system must be to determine if a potential failure in the computer system can cause a risk in product security, efficacy, and quality. This must be applicable in all pharmaceutical industry areas, related with good clinical practices (GCPs), good distribution practices (GDPs), good laboratory practices (GLPs), or GMPs. Validation of a computer system also provides evidence that all of its components are working perfectly and generating adequate results. Thus, all of its components must be validated, such as applications, processes, users, and facilities.

Risk assessment is the first critical step in the validation of a computer system. After risk assessment, the validation protocol must be created, including all the points referred in the VMP. The procedures to be executed in a computer system should be defined, in addition to the specifications and tests to be performed. These tests could be trials for IQ, OQ, and PQ. After the test procedure and data evaluation, a validation report must be available. Therefore, a very well written and periodically reviewed (e.g., every year) VMP, validation protocol, and validation report are always necessary.

The extent of validation necessary will depend on a number of factors, including the use to which the system is to be put, whether the validation is to be prospective or retrospective, and whether novel elements are incorporated. Validation should be considered as part of the complete life cycle of a computer system. This cycle

includes the stages of planning, specification, programming, testing, commissioning, documentation, operation, monitoring, and modifying [13].

A retrospective validation is applicable to computer systems in use when the VMP is elaborated. In this case, the real use of the system must be pointed out in order to allow a correct test plan.

To produce an adequate validation close cooperation between key personnel and those involved with the computer systems is essential. People in responsible positions should have the appropriate training for the management and use of systems within their field of responsibility which utilizes computers. This should include ensuring that appropriate expertise is available to provide advice on aspects of design, validation, installation, and operation of computerized systems [12].

After the validation procedure and personal characterization, it is necessary to describe how to handle the computer systems to validate them. As previously referred, a computer system is used to execute a specific procedure to give a defined result. Therefore, attention should be paid to siting equipment in suitable conditions where inappropriate factors cannot interfere with the system [13]. Therefore, the physical place of the computer system must be described in the validation protocol, and every time its place is changed a revalidation plan is required.

A written detailed description of the system should be provided (including diagrams) and kept up to date. It should describe the principles, objectives, security measures and scope of the system, the main features of the way in which the computer is used, and how it interacts with other systems and procedures [13]. A computer system is a particular type of equipment, and, as for any equipment, it is required to have a SOP available and, if possible, near the equipment.

The software—a program enabling computer to perform a specific task—is a critical component of a computerized system. Even a simple calculus sheet could be considered software, as it allows a computer to perform a specific task. The user of such software should take all reasonable steps to ensure that it has been produced in accordance with a system of quality assurance [13].

The system should include built-in checks of the correct entry and processing of data. In order to verify the validation data, some computer systems may periodically be submitted to a defined group of inputs for which the result is known and the result must be kept and filed. If the results are acceptable, the computer system is operating well, but if the results do not match the expected ones, the computer system is not working properly and maintenance is required. In this situation, all the results obtained from the referred computer system since the last validation verification are consided questionable and must be reevaluated.

Before a system using a computer is brought into use, it should be thoroughly tested and confirmed as being capable of achieving the desired results. If a manual system is being replaced, the two should be run in parallel for a time as part of this testing and validation [13]. This means that every time maintenance has made an intervention in the equipment and any piece has been adjusted or replaced, a revalidation is required. This maintenance intervention could be in both hardware and software.

Data should only be entered or amended by authorized people. Suitable methods of deterring unauthorized entry of data include the use of keys, pass cards, personal codes, and restricted access to computer terminals. There should be a defined

procedure for the issue, cancellation, and alteration of authorization to enter and amend data, including changing personal passwords. Consideration should be given to systems allowing for recording of attempts to access by unauthorized people [13]. A computer system must have different user levels and the access to each application should be protected by a user name and password. Critical operations, such as method modifications, calibration procedures, and adjustments, must only be possible by specific personnel. Operators must only have access to procedures defined and without any possibility of changes. All the operations must be automatically recorded in the equipment, including operator user name, time, and date. These records may be available for a quality assurance audit.

When critical data are being entered manually (e.g., the weight and batch number of an ingredient during dispensing), there should be an additional check on the accuracy of the record which is made. This check may be done by a second operator or by validated electronic means [13]. Frequently, computer systems require external data to realize the specified task. The introduction of an external data is a critical step in the computer system procedure, because an error in data introduction may produce an erroneous result. To avoid errors one can proceed by double verification followed by the automatic registry prove or, alternatively, double verification consisting of the introduction of data by one operator while another verifies the data introduced and both sign the register. An automatic registry proves that it is available when the values introduced in a computer system came from another computer system registry. For example, in the dissolution test for tablets, tablet weights introduced in the computer system could be available from a balance associated with an automatic registry; that registry is the proof that the data introduced are correct.

The system should record the identity of operators entering or confirming critical data. Authority to amend entered data should be restricted to nominated people. Any change to an entry of critical data should be authorized and recorded with the reason for that change. Consideration should be given regarding the creation of a complete record of all entries and amendments (an "audit trail") [13].

As previously mentioned, changes to a system or to a computer program should only be made in accordance with a defined procedure and should include provision for validating, checking, approving, and implementing the change. Such a change should only be implemented and recorded with the agreement of the person responsible for the part of the system concerned. Every significant modification should be validated.

For quality auditing purposes, it should be possible to obtain clear printed copies of electronically stored data. Data should be secured by physical or electronic means against wilful or accidental damage. Stored data should be checked for accessibility, durability, and accuracy. If changes are proposed to the computer equipment or its programs, the above-mentioned checks should be performed at a frequency appropriate to the storage medium being used. Data should be protected by backing up at regular intervals. Back-up data should be stored as long as necessary at a separate and secure location [13].

All data generated in a computer system must be available, secure, and safely stored. Unauthorized people cannot have access to data files. It must be impossible to overwrite any data, but all recalculation required must generate a new data and not substitute the previous one. Safety procedures must be available. Data should be backed up periodically following specific SOPs, and backup copies must be iden-

tified and stored in a defined and safe place. All backup data must be evaluated for the backup process, ensuring that the data backup has been well performed. Periodically all data must be checked for integrity in order to ensure that they are well preserved and no data have been lost.

Adequate alternative arrangements should be available for systems which need to be operated in the event of a breakdown. The time required to bring the alternative arrangements into use should be related to the possible urgency of the need. If the system fails or breaks down, the procedures to be followed should be defined and validated. Any failures and corrective actions taken should be recorded [13]. Special cases of breakdowns are power breakdowns. It is required that some computer systems work for 24 h and should not be interrupted after starting data acquisition. Consequently, an extra uninterruptible power supply (UPS) is required to allow the operation of these computerized systems during power breakdowns.

A procedure should be established to record and analyze errors and to enable corrective action to be taken. Every computer system can originate errors that must be documented. Every time an error occurs, it must be analyzed and, if applicable, some adjustments in the computer system may be required. Adjustments have to be well defined and documented; in some cases a revalidation of the system may be required.

When outside agencies are used to provide a computer service, there should be a formal agreement including a clear statement of the responsibilities of that outside agency [13]. When the release of batches for sale or supply is carried out using a computerized system, the system should allow only a qualified person (QP) to release the batches and it should clearly identify and record the person releasing the batches. This is possible by giving the QP an operational level in the computer system that allows it to release batches. No other person should have the same operational level authorization as the QP.

*Software Validation* Authorities are demanding rules in order to outline the software validation principles used in medical device software or the validation of software used to design, develop, or manufacture medical devices.

Guidances recommend an integration of software life-cycle management and risk management activities. Software validation and verification activities must be conducted throughout the software life cycle [12, 14]. Software verification and validation are terms frequently confused. Software verification is defined as the process that provides objective evidence that the design outputs of a particular phase of the software development life cycle meet all of the specified requirements for that phase [14]. Software verification consists of tasks performed to evaluate if the software is performing desired tasks and providing desirable results. Software verification is part of software validation. Software testing is one of the many verification activities intended to confirm that software development output meets its input requirements. Other verification activities include various static and dynamic analyses, code and document inspections, walkthroughs, and other techniques [14].

Although actual guidelines are only applicable in software used as a component part or accessory of a medical device (e.g., blood establishment software, programmable logic controllers in manufacturing equipment, software that records and maintains the device history record), the validation principles presented in these guidelines could be applicable to any software validation.

Software is not a physical entity and, unlike some hardware failures, software failures occur without advanced warning. One of the most common software failures is branching, that is, the ability to execute alternative series of commands based on differing inputs. The software branching capacity makes the commands extremely complex and difficult to validate once errors occur as an answer of a specific input, and until the introduction of that specific input error has not been detected. Software input can be almost any data and, and since it is impossible to introduce all data into a software, validation of data is extremely difficult. Thus, results are considered to be of high confidence level. The majority of software problems occur as a consequence of errors in the software design and development and are not directly related to the software manufacture. It is simple to manufacture several software copies that work perfectly and as the original one.

Software validation is not separately defined in the quality system regulation. The FDA considers software validation to be the "confirmation by examination and provision of objective evidence that software specifications conform to user needs and intended uses, and that the particular requirements implemented through software can be consistently fulfilled" [14]. Software validations have special concerns on software installation, implementation, and utilization. A software validation consists in several tests, inspections, and verifications performed to assure the adequate installation and use of software and that the tasks performed meet all the specifications defined. Software validations must be performed under the environmental conditions to which software will be submitted. This is particularly important in medical devices that are used under special conditions, such as close to or inside the human body.

A software validation plan must take into account the risk analyses for the software; therefore a critical software must have a high level of confidence and must be submitted to deep validation processes, whereas noncritical software may be submitted to less extensive validation processes.

Seemingly insignificant changes in software code can create unexpected and very significant problems elsewhere in the software program. The software development process should be sufficiently well planned, controlled, and documented to detect and correct unexpected results from software changes [14]. Maintenance of a software must be carefully performed because even a few small changes could develop dramatic software results. Accuracy and thorough documentation are essential in order to assure the software validation.

Software validation is not a simple and easy task; therefore an adequate schedule and task plan for software validation and verification are required to avoid unnecessary time and money expenses. One day of planning with no experiments performed is better than one day of experiments without planning.

*Electronic Documents* Several special cases of electronic documents have been targeted by specific official regulations. These are electronic records and electronic signatures. For instance, the FDA issued regulations that provide criteria for acceptance, under certain circumstances, of electronic records, electronic signatures, and handwritten signatures executed to electronic records as equivalent to paper records and handwritten signatures executed on paper. These regulations, which apply to all FDA program areas, were intended to allow the widest possible use of electronic technology compatible with the FDA's responsibility to protect public health [12].

Some interpretations for this guideline included almost every record generated in a computer as the final target, but another was very simple, considering only a few documents under the scope of this guideline. The records required to be maintained under predicate rules or submitted to the FDA would apply whenever records in electronic format replace paper hardcopies. On the other hand, when the computer is used to generate paper printouts of electronic records and those paper records meet all the requirements of the applicable predicate rules, the FDA would generally not consider people to be "using electronic records in lieu of paper records." The same is considered when people rely on paper records to perform their regulated activities. In these cases, the use of computer systems in the generation of paper records would not trigger the official requirements [12]. Computer systems that generate a document that is printed out and signed by hand are excluded in this guideline. The guideline only regulates documents generated in a computer system that are filed, signed, and modified electronically. Since the documents are not printed and signed, special considerations must be taken in order to assure that an adequate file is performed, undesirable modifications are avoided, and the signature is safe and personalized.

This guideline is only applicable to documents for which it is required to be maintained. All documents created and signed electronically that are not required to be maintained are not under the scope of this guideline. The decision of which documents are under the scope of this guideline is taken by the pharmaceutical industry and must be well documented and justified.

Electronic signatures are intended to be the equivalent of handwritten signatures, initials, and other general signings required by predicate rules. They include electronic signatures that are used, for example, to document the fact that certain events or actions occurred in accordance with the predicate rule (e.g., approved, reviewed, and verified).

*Validation Team* A well-defined validation team with a well-written description of responsibilities is required and assures the adequate realization of the validation tasks. A validation team should be composed by different responsibilities: responsible-of-validation team, team leader, archive manager, test coordinator, quality assurance member, tester, and witness. The responsible-for-validation team elaborates and approves the VMP, protocols, and reports. The team leader should be responsible for the computer system validation and utilization. An archive manager is responsible for the management of all computer system validation documents. The test coordinator is responsible for the computer system test and coordinates the elaboration and operation of tests for evaluating the performance of the computer system. A quality assurance member is required to periodically inspect and train the personnel and review all the validation documents. The tester is responsible for the execution of the tests required to perform the validation protocol. The witness is responsible for observing and reviewing the operations of the tester.

## 9 CLEANING VALIDATION

Cleaning validation is defined as the "the process of providing documented evidence that the cleaning methods employed within a facility consistently controls potential

carryover of product (including intermediates and impurities), cleaning agents and extraneous material into subsequent product to a level, which is below predetermined levels" [15].

The reasons behind the validation of cleaning procedures are the assurance of the safety and purity of the product (customer requirement), it is a regulatory requirement in active pharmaceutical ingredient product manufacture, and it assures the quality of the process from an internal control and compliance point of view [15].

Cleaning should be carried out with relative ease and with the use of standard cleaning materials [7]. Vacuum facilities should be available for cleaning and contact parts fouls be wiped down and sanitized utilizing a sanitizing agent. The equipment should be washed, dried, covered, and stored in an equipment storage area.

As in manufacturing, the packaging operation starts with the generation of the packaging order. This consists of the approved packaging components, the batch number of the product to be packaged, and quantities of each. A supervisor verifies the accuracy and completeness of the packaging order, including expiration date, line being used, and any other special equipment being used for that operation. These steps are accomplished prior to bringing components to the line. The complete line area, including all equipment, is verified as being properly disassembled and cleaned of all product and components from the previous packaging operation. After the QA department has verified that the area has been cleared and cleaned, the components must be brought to the line for mechanical setup. Once the setup mechanic has completed all the adjustments required, a supervisor oversees the prestart procedure. This consists of cleaning all products used during the setup; counting labels used in the label setup; rechecking all components and lot numbers; verifying the bottle count, all stamps, and lot numbers; and signing the packaging order that all is in readiness to start packaging operations.

Nonappropriate cleaning procedures will develop batches of poor quality due to the risk of presence of a number of contaminants, such as precursors of the drug, degradation products, solvents and other materials employed during the manufacturing process, microorganisms, cleaning agents, and lubricants [15].

## 10 CONCLUSIONS

Since the middle of the twentieth century the pharmaceutical industry has been a leader in terms of quality and security of manufacturing, associating higher production efficiency with higher profits. Therefore, validation became essential, that is, the confirmation and evidence that all facilities, equipment and processes work as desired, generating quality products.

In the present chapter, the pharmaceutical industry validation system has been reviewed. To have an appropriate validation system it is first required to define which equipment, facilities, and processes will be validated, when they will be validated, and by whom this must be performed. This definition is based on a risk assessment priority and is written in a specific document, the so-called MVP. In order to generate an adequate validation report, all the validation activities should be described in the validation protocols, SOPs, and specific procedures.

Facilities validation is a critical process in a pharmaceutical industry and the types of pharmaceutical forms produced must be considered. Facilities that produce different pharmaceutical forms have different specific requirements and different critical parameters based on risk assessment. All facilities must have an adequate flow of people, raw materials, bulk products, and finished products. These flows must be created in order to avoid cross-contamination. Additionally, pressurized rooms and adequate SOPs should be supplied to minimize the risk of cross-contamination. Controlled air temperature and humidity are also required and should be validated to ensure adequate stability of the product.

MVP is a crucial procedure to confirm that the product manufacture is adequate and is generated in a consistent manner with the same quality. Definitions of equipment, rooms, quality of raw materials, and process parameters are necessary to validate the manufacturing process. Maintaining these definitions ensures that the product has the same quality after the process validation. Changes in equipment, raw materials, or process parameters require a manufacturing process revalidation.

Analytical methods validation is one of the most regulated validation processes in the pharmaceutical industry. Analytical validations are required to demonstrate that the methods employed are the most indicated for each product and that the results obtained are reliably correct. All methods employed in raw and finished product materials analysis are required to be validated.

Equipment validation is comprised of four critical operations: design, installation, operational, and performance qualifications. These operations will confirm that the equipment has adequate specifications, installation, and functions, manufacturing a product with adequate properties. After these procedures, whenever equipment is installed, it must be periodically verified and calibrated in order to ensure adequate performance.

Computer systems are a specific type of equipment that must be validated as well. These systems have specific requirements if they are used to collect and process data. Associated with computer hardware is the software, carefully validated in order to prove that the data generated by them are correct.

In order to avoid cross-contamination, another concern with respect to equipment is the cleaning process, which must comprise cleaning SOPs to ensure adequate cleanliness. Cleaning validation must be performed based on risk assessment and worst-case scenarios.

## REFERENCES

1. World Health Organization (WHO) (2006), WHO expert committee on specifications for pharmaceutical preparations (fortieth report)—Supplementary guidelines on good manufacturing practices (GMP): Validation, WHO, Geneva.
2. Thomas, E., Grochulski, A., Patel, R., George, S. M., and Zhang L. (2006, Sept.), The laboratory control system: Fulfilling cGMP requirements, *BioPharm Int.*, 26–32.
3. Slater, S. (1999), *Biopharmaceutical Validation: An Overview in Biopharmaceuticals, an Industrial Perspective*, Springer-Verlag.
4. Chaloner-Larsson, G., Anderson, R., and Egan, A. *A WHO Guide to Good Manufacturing Practice (GMP) Requirements*, 1997, World Health Organization, Geneva.

5. International Conference on Harmonization (ICH) (2000), Good manufacturing practice guide for active pharmaceutical ingredients, ICH, Geneva.
6. European Medicines Evaluation Agency (EMEA) (2003), *Good Manufacturing Practices*, Chapter 4: Building and facilities, EMEA.
7. Nash, R. A. (1990), The essential of process validation, in *Pharmaceutical Dosage Forms, Tablets*, Marcel Dekker Inc, New York.
8. Hanna, S. A. (1990), Quality assurance, in *Pharmaceutical Dosage Forms, Tablets*, Marcel Dekker Inc, New York.
9. Connolly, R. J., Berstler, F. A., and Coffin-Beach, D. (1990), Tablet production, in *Pharmaceutical Dosage Forms, Tablets*, Marcel Dekker Inc, New York.
10. European Medicines Evaluation Agency (EMEA) (2003), *Good Manufacturing Practices*, Chapter 3: Premise and equipment, EMEA.
11. U.S. Food and Drug Administration (FDA) (1987), Guideline on general principles of process validation, FDA, Rockville, MD.
12. U.S. Food and Drug Administration (FDA) (2003), Guidance for industry, CFR Part 11, Electronic records; electronic signatures, good manufacturing practices, FDA, Rockville, MD.
13. European Medicines Evaluation Agency (EMEA) (2003), *Good Manufacturing Practices*, Annex 11: Computerised systems, EMEA.
14. U.S. Food and Drug Administration (FDA) (2002), general principles of software validation, good manufacturing practices, Rockville, MD.
15. Active Pharmaceutical Ingredients Committee (APIC) (1999), Cleaning validation in active pharmaceutical ingredient manufacturing plants, APIC.

# *Applications of High Performance Ion-Exchange Chromatography in the Pharmaceutical Industry*

**Matthew L. Workentine, Vy A. Tran and Raymond J. Turner**
*Department of Biological Sciences, Faculty of Science, University of Calgary, Alberta, Canada,*

## 1 ESSENTIALS OF ION-EXCHANGE CHROMATOGRAPHY

### 1.1 Introduction to Ion-Exchange Chromatography

Ion-exchange chromatography (IEC) is based on the interaction of ionic species in the mobile phase with charged functional groups immobilized on a stationary phase. IEC is divided into anion-exchange chromatography and cation-exchange chromatography depending on the type of ion it is separating. Anion exchangers contain positively charged functional groups in the stationary phase and cation exchangers have negatively charged functional groups. High performance IEC (HPIC) uses mechanically stable stationary phase supports that allow the high pressures needed for rapid and high resolution separations. HPIC can be considered a mature analytical technique [1] and thorough descriptions can be found in various handbooks [2, 3].

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing,*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

The purpose here is to understand HPIC as a tool for use in pharmaceutical separations. In that context the following descriptions of column selection, elution parameters, sample preparation, and detection methods will focus on current technologies and their application in the pharmaceutical industry. The theory and essentials of IEC in the following sections are adapted primarily from the two handbooks that the reader is directed to for more in-depth descriptions [2, 3].

## 1.2 Basic Theory

Separation of ions in an ion-exchange column is based on the principle that different ions will have differing affinities for the stationary phase. The stationary phase contains a fixed charge with a mobile counter ion. As solute ions in the mobile phase are passed down the column they can exchange with the mobile counter ion and are temporarily retained by the fixed charge of the stationary phase. Based on the strength of interaction with the stationary phase different ions will spend different amounts of time in the column and hence will be separated. Let us consider the example of a strong anion exchange column containing a quaternary ammonium functional group. If a solution of NaOH is passed down this column the hydroxide anion exchanges with the counter ion of the quaternary ammonium group and the quaternary ammonium group is now completely in its $OH^-$ form. A solution of $Br^-$ and $Cl^-$ ions are then injected onto the column and the NaOH is continually pumped on. The $Br^-$ and $Cl^-$ begin to move down the column as they do so they compete with $OH^-$ and with each other for the positive charge on the stationary phase. This exchange reaction between an analyte ion and $OH^-$ is controlled by a chemical equilibrium and a corresponding equilibrium constant, termed the *selectivity coefficient* (see Fig. 1). In our example, the $Br^-$ and $Cl^-$ ions will have differing equilibrium constants (and different selectivities) when competing with $OH^-$ for charge interaction sites and will be retained to differing extents before passing out the bottom of the column to the detector. Choosing an eluant with an appropriate selectivity is very important for IEC to ensure that the ions are retained on the column long enough to be separated, retained with a reasonable column residence time.

## 1.3 Resins and Functional Groups

Ion-exchange columns can have either a silica-based resin, as in typical HPLC columns, or, more commonly an organic polymer resin. The most common resins are polystyrene, made up of styrene cross-linked with different amounts of divinylbenzene (structure shown in Fig. 2), which brings the necessary mechanical stability to the resin [2, 3]. Other common polymers for resins include polymethacrylates and polyvinylalcohol [1]. Organic polymer resins, although sometimes exhibiting secondary, non-polar interactions with the analyte [1], are stable at the extremes of pH often utilized in IEC [3]. Resins can be classified as micro or macroporus depending on how the polymerization takes place, although microporus is more common for ion-exchange. Macroporus resins are considerably more rigid and are often utilized when

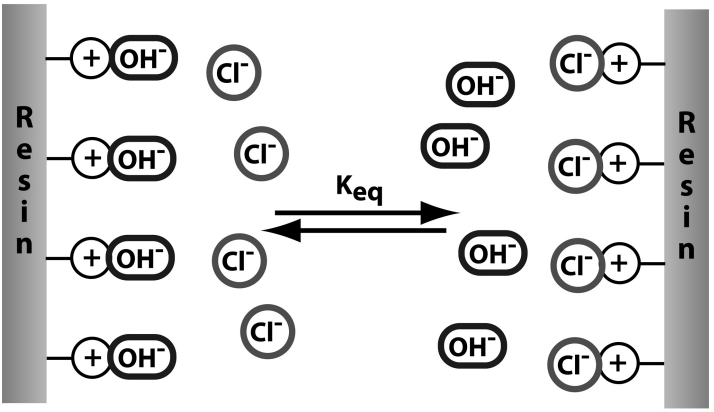


$$K = \frac{[Cl^-]_{bound}\,[OH^-]_{free}}{[OH^-]_{bound}[Cl^-]_{free}}$$

***FIGURE 1*** *The selectivity of an ion-exchange resin is determined by the equilibrium between free and bound mobile ions. In this example, the positive charge is on the stationary phase but the negative counter-ion ($OH^-$) is mobile and can exchange with other anions in the mobile phase. As the sample ion, $Cl^-$ is passed down the column it will exchange with $OH^-$ and be retained temporarily on the stationary phase. The equilibrium of this exchange will determine just how long the anion is retained on the column. It should also be noted that the sample counter-ion ($Na^+$ for example, not shown on the figure) will not be retained at all, but will travel through the column.*

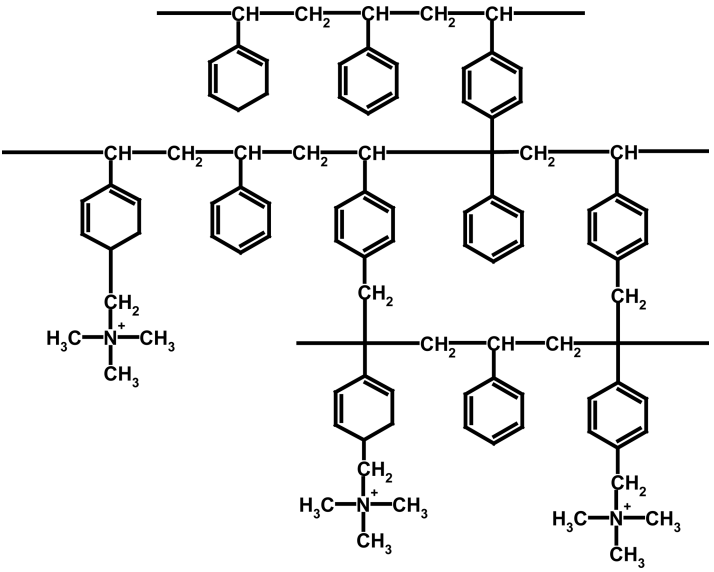

***FIGURE 2*** *A schematic illustrating the chemical structure of a styrene/divinylbenzene copolymer, a common stationary phase for HPIC. The diagram also shows some of the polymer functionalized with a quaternary amino anion-exchanger, used on many anion-exchange resins.*

organic solvents are needed because they do not exhibit the swelling and shrinking that may occur with microporus resins.

Resins are turned into ion-exchangers by the addition of functional groups to the polymeric beads. Depending on the type and amount of functional groups added, the selectivity and strength of the ion-exchanger will vary. Functional groups are typically divided into strong, intermediate, or weak anion and cation exchangers. The most common strong anion exchanger is quaternary ammonium functional group (Fig. 2) and the most common strong cation exchanger is sulfonic acid [2].

Another important concept with regard to ion-exchange resins is the ion-exchange capacity. Ion-exchange capacity is defined as the number of ion-exchange sites per weight equivalent of the column packing [3]. This is usually expressed in milli-equivalents per gram of resin (mequiv/g) and in general low-capacity ion-exchange resins are preferred. This is because as the capacity increases it requires a higher ionic strength eluant to completely elute the analyte. High ionic strength buffers are usually avoided, as this leads to interference with the conductivity detector which is the most common method of detection in IEC (see below). However, alternative detection methods do allow the use of high ionic strength buffers and higher capacity resins.

## 1.4 Detectors

There are many approaches used to detect the ions as they come off the column, but the most common by far is suppressed conductivity. Spectrophotometric and pulsed amperometric detection are also quite common and useful with regard to the application of HPIC to pharmaceutical studies.

***1.4.1 Suppressed Conductivity*** Measuring the conductivity of a solution as it passes out of the column is the most widely used detection method for ion chromatography [3]. Conductivity is the ability of a solution to conduct electricity across two electrodes [2]. Solutions with higher ionic strength (more ions) will have a higher conductivity. As the analyte ions come off the column they increase the conductivity of the solution thus increase current which is interpreted as a peak on the chromatogram. The problem with ion-exchange and conductivity is the relatively high ionic strength of the eluant resulting in high background noise. Suppressed conductivity is the solution for this. In suppressed conductivity, the sample passes through another ion-exchange column which replaces the counterion of the eluant [2]. For example in an anion exchange experiment where NaOH is the eluant, a suppressor would exchange the $Na^+$ ions for $H^+$ resulting in the production of water and almost no background conductivity [2].

***1.4.2 Spectrophotometric Detection*** Compared to suppressed conductivity, spectrophotometric methods of detection have not seen as much prevalence in HPIC for pharmaceuticals [3]; however, absorbance or fluorescence can be very useful for the separation of biological ions. For example, phosphorylated nucleotides where the nucleotides, strongly absorb light at 254 nm and this makes them very amendable to detection by UV absorbance. UV detection is the method of choice for the detection of

nucleotides from bacterial samples [4]. Used in combination with conductivity, UV absorbance can be a very useful tool for detecting several different types of metabolites [5] (see example in Fig. 3). Fluorescence, on the other hand, is more useful in combination with a post-column derivitization method where select ions are derivitized by reaction or complexation with a reagent that renders them fluorescent [3]. Although potentially useful, particularly as a highly selective method (i.e. only your specific analyte is detected), fluorescence detection is rarely used.

***1.4.3 Pulsed Amperometric Detection*** Electrochemical detection detects compounds based on the catalytic oxidation or reduction at the surface of an electrode [2]. The potential between a reference and a working electrode is maintained and the current is sampled over time. As the compound emerges from the column it is oxidized or reduced on the surface of the working electrode,

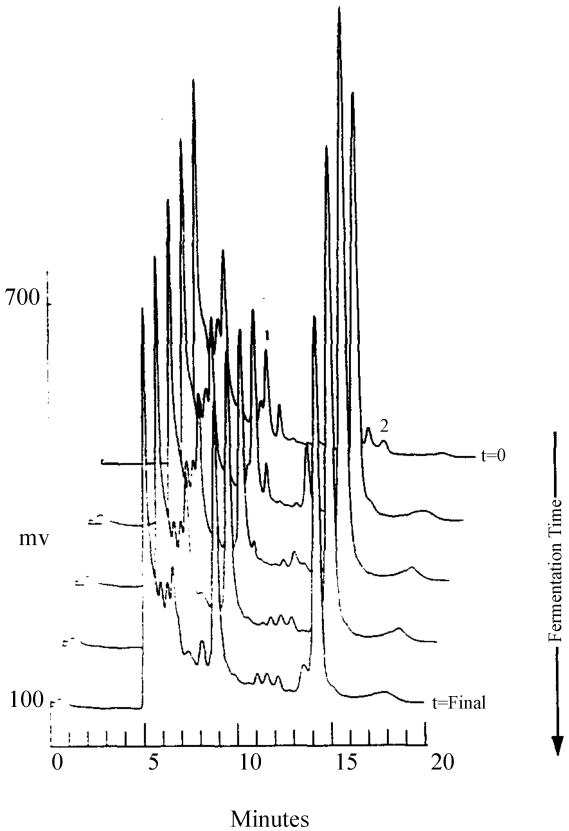


***FIGURE 3*** *Separation of glycerol and ethanol in yeast fermentation broth by isocratic elution. Shown is a time course of representative chromatograms of fermentation broth of Sacchoromyces cerevisiae grown in complex media. Detection was done using PAD with a platinum electrode. Peak 1 is glycerol and peak 2 is ethanol. Reproduced from Ref. [35] with permission.*

generating a detectable current [6]. Pulsed electrochemical detection uses a series of potential pulses to oxidatively clean and then reduce the working electrode surface [2]. This method works very well for carbohydrates because of their tendency to adsorb on to the surface of the working electrode rendering it inactive [7]. An oxidative pulse completely oxidizes the surface of the electrode, thereby removing the adsorbed contaminant. A reductive pulse reduces the oxidized surface to pristine metal ready for the next measurement [7]. Carbohydrates can be accurately and sensitively detected using either a gold or platinum working electrode, although gold seems to be the best choice [8]. This method of pulsed amperometry has been used extensively for the detection of carbohydrates from a variety of complex samples [8, 9].

## 1.5 Considerations for the Separation by IEC

The first choice one should consider when doing IEC is the type of column. This, of course will be dependent on the compound being analyzed. The resin ion-exchange capacity, functional group, and solvent compatibility must be considered when choosing a column. The IonPac AS 11 from Dionex is quite often used for high-performance anion exchange [5, 10-13], particularly for organic acid separation. The CarboPac A1 also from Dionex has been used for the separation of sugar phosphates and polysaccharides [12, 14, 15]. Both of these function at a broad pH range and are compatible with organic solvents. There are a large number of columns available, many for specific purposes, and a recent review highlights some of the newer stationary phases for HPIC and their uses [16].

The type of column and chromatography will help determine the type and concentration of eluant. Compatibility with the detection method must also be taken into consideration when choosing an eluant. HPIC coupled to a mass spectrometer (MS) can be problematic because of the high ionic strength of the eluants. Online desalters are one solution to this [17]. Modern ion-exchange resins are of much lower capacity than earlier ones [3] and as such analytes can be eluted with lower ionic strength buffers which is important for detection methods like conductivity and MS. The pH of the eluant is extremely important in IEC because the presence of a charge on the analyte will be dependent on the pH of the solvent. Carbohydrates in particular need to be in a high pH solvent in order to become anions and be separated [9].

Once the eluant is chosen one needs to determine whether an isocratic or gradient elution will be used and if a gradient is chosen, the type of gradient will need to be determined. At one time isocratic elution (a single concentration of eluant) was the primary choice as gradients proved difficult to implement [1]. However, isocratic elutions tend not to give optimal peak resolution and gradient elutions have become the dominant choice as they provide much greater flexibility and peak resolution. Old problems with gradients, especially hydroxide gradients were essentially eliminated by the introduction of electrolytic eluant generators which produce hydroxide (for anion separations) and methanesulfonic acid (for cation separation) using only water. Gradient elutions are used routinely and can be quite complex with multiple steps and tend to be difficult to predict [1]. Previous work separating the compounds of interest

can be a useful starting place and there have been recent developments in computational tools to optimize separations [1].

Finally the detection method will have to be considered. This again will depend on the ion(s) being analyzed, their concentration, as well as the type of eluant need to elute them. As will be discussed in Section 2.3.2 carbohydrates can be detected quite well with pulsed amperometry [9]. Detection of phospho-nucleotides is typically done by UV absorbance at 254 nm, the wavelength at which the nucleotide bases strongly absorb. Multiple detection methods can also be useful in expanding the range of analytes that can be separated on a single column [5]. Multidimensional chromatography separations linked to mass spectrometry are also very powerful as these can greatly expand the number of compounds that can be separated. Some good examples of HPIC linked to MS for the analysis of proteins is given in Section 2.4.

## 2 EXAMPLES OF HPIC IN THE PHARMACEUTICAL INDUSTRY

### 2.1 Quality Analysis

HPIC has developed into a very useful method with a variety of applications in the analytical field for the separation and identification of various organic, biochemical, and inorganic species. It is therefore also an attractive and powerful tool for pharmaceutical analysis in order to meet the requirements by regulatory authorities. Ion chromatography can be used to identify and quantitate pharmaceutically relevant compounds, including the active pharmaceutical ingredients, trace amounts of impurities such as intermediates that have developed during formulation, degradation products and various other undesired side products [18]. The amounts of inactive materials such as preservatives, anti-oxidants, and counter ions that are also included with the drug or samples used for pharmaceutical research are also of particular interest. In this section, a few key examples of HPIC use are provided.

***2.1.1 Analysis of Aliphatic Amines*** The presence of amines in drug production is quite broad, including reagents for synthesis of emulsifying agents and medicines, counter-ions, and reaction intermediates [19-21]. Therefore, it is necessary to not only separate but also quantitate these species in the presence of other impurities. The classes of aliphatic amines include alkanolamines as well as alkyl amines. Analysis of aliphatic amines is problematic for pharmaceutical analysis because these compounds are non-chromophoric only allowing detection at wavelengths below 200 nm. However, at this wavelength, there is a risk of interference from species in the eluent and matrices used. While a wide range of methods in literature are available for the analysis of aliphatic amines, they are time consuming, requiring derivitization of complex mixtures that are typical of pharmaceutical samples. High performance liquid chromatography (HPLC) and gas chromatography (GC) are normally used in conjunction with derivitization, but the traditional HPLC silica-based columns can interact with the amines causing

poor resolution and significant tailing of peaks [20-22]. Alternatively, the application of HPIC for the analysis of aliphatic amines is a more reliable technique; as a consequence, over the past decade many methods have been described using simple ion chromatography [19-24].

Krol *et al.* [20] have reported that ion chromatography could be used for the separation of alkyl amines and alkanolamines using a poly(butadiene-maleic acid)-coated silica-based cation exchanger (Waters IC-Pak Cation M/D). Detection was achieved by indirect conductivity of an EDTA-nitric acid containing acetonitrile or methanol mobile phase. The separation of standard mixtures of alkanolamines, low-molecular mass amines, and quaternary amines which could also contain sodium, ammonium, lithium, calcium, and magnesium were achieved with good resolution in one single isocratic run, by varying the amounts of nitric acid or organic modifier in the mobile phase. Application of this method on amine analysis was carried out on a pharmaceutical formulation of bethanechol, a synthetic analog of choline. Sodium, ammonium, potassium, magnesium, and calcium were successfully separated from bethanechol. Furthermore, the method claims to be able to separate biogenic quaternary amines such as choline and acetylcholine from bethanechol using the same conditions.

Another compound that can be analyzed by ion chromatography is 1-(butylamino)-1-deoxy-D-glucitol (BDG) which is a polyhydroxy aliphatic amine. BDG is an important intermediate and it is necessary to isolate and test the purity of the compound before continuing with synthesis. The impurities of BDG are 1-(amino)-1-deoxy-D-glucitol (ADG), 1-(butylamino)-1-deoxy-D-mannitol (BDM), and *n*-butylamine (BA). Kersten [24] reported a method where BDG could be separated from the related amine impurities using a styrene-divinyl-benzene polymeric cation exchanger (Dionex IonPac CS3) with suppressed conductivity detection. The mobile phase of dilute HCl and DL-2,3-diaminopropionic acid (DAP) was used. As the concentration of HCl increased, retention of the amines decreased, and in the presence of DAP which has two positive charges, the analytes were effectively eluted. This mobile phase is compatible with suppressed conductivity detection since in the presence of the alkaline conditions of the suppressor, DAP is converted to the anionic form which is then removed from the eluent by the cation suppressor membrane regeneration system.

Although the examples given above pertain to an assortment of alkyl amines, those studies did not focus on quantitation of the compounds and more importantly have not been applied to trace amounts in pharmaceutical analysis. A study of the separation of amylamine and *tert*-butylamine (TBA) in pharmaceuticals by ion chromatography was developed by Jagota *et al.* [19] to address the limitations of previous studies. Amylamine is used in the synthesis of BMS-181 866-02 and thus may be present in trace quantities. BMS-181 866-02 is further used in the synthesis of thromboxane antagonist. TBA is used as a counter ion for the prodrug ester of an oral hypocholesterolemic agent and thus the amount of TBA is necessary in order to assign a purity value to the drug. Separation of amylamine and TBA were achieved with isocratic conditions using an acetonitrile–methanesulfonic acid mobile phase on a carboxylated poly(ethylvinylbenzene/divinylbenzene) column (Dionex CS14) with conduc-

tivity detection. The detection limit for amylamine using the method developed by Jagota *et al.* was 0.01% (w/w) with a signal-to-noise ratio of 3. This method is attractive because it provides rapid ion chromatography determination of amylamine and TBA in a wide range of concentrations (0.01–10%) with direct injection and easy clean-up.

***2.1.2 Detection of Azide*** Protein samples and diagnostic reagents used for pharmaceutical research typically contain sodium azide as a preservative or an antimicrobial agent [25-27]. Azide is also used in the synthesis of drugs such as antihypertensives, anti-HIV drugs, and some antibiotics. However, a consequence to the presence of azide is its high level of toxicity with a very low lethal dose of less than 50 mg/kg in humans [25]. As such, it is necessary to quantitate azide levels in samples used for pharmaceutical research. Despite the numerous methods described for azide detection, including derivitization which is time consuming, most lack the sensitivity required for determining trace amounts. Alternatively, ion chromatographic approaches have been designed for determining azide in pharmaceutically relevant samples. With a p*K*a of 4.8, azide will be in its anionic form under alkaline conditions, making ion chromatography ideal for its analysis. One challenge faced when separating azide from other ionic members in the sample is the tendency for similarly sized monovalent anions such as nitrite, nitrate, and bromide to co-elute with the azide peak. Nonetheless, a few groups have successfully applied simple, straightforward, and interference free approaches of ion chromatography to quantify azide in various proteins samples.

In 1991, Annable and Sly [26] developed an analytical method for the identification and separation of azide from common anions found in protein samples. Optimization of azide analysis was done using an aminated latex bead sulfonated polymer [poly(styrene-divinylbenzene)] column (Dionex IonPacAS4A) with a sodium tetraborate mobile phase and chemically suppressed conductivity detection. Samples tested by the method described contained 10 ppm amounts of fluoride, bromate, chloride, nitrite, bromide, azide, nitrate, phosphate, and sulfate (in the order they elute). Of interest is the resolution of azide from nitrate and bromide. The group initially tried a simple approach with dilute carbonate–bicarbonate eluent previously described in literature. However, resolution of azide from nitrate and bromide could not be achieved on the various columns that were tested, even with varying the carbonate–bircarbonate eluent concentration. Tetraborate was then chosen as the eluent despite its weak elution strength (as sulfate is retained strongly), it has greater resolving power, and furthermore it is capable of being chemically suppressed. On the IonPac AS4A column, the typical retention time of azide, based on a 10 ppm standard is 8–12 min with a detection limit of about 30 ppb using the tetraborate eluent system. Additionally, four proteins: bovine serum albumin (BSA), zinc metallo enzyme, collagen, and a growth factor were also analyzed for the quantitation of azide using the described method. The amount of azide in BSA was successfully determined. The samples of zinc metallo enzyme and growth factor were expected to be azide free, and using this approach it was confirmed that azide was not present. Overall, this method allows for quality

assurance experiments and the crucial detection of trace amounts of azide in pharmaceutically relevant samples.

Vinkovic and Drevenkar [25] were concerned with azide detection in the presence of high concentrations of chloride. They addressed a limitation with the tetraborate eluent system described [26], as it is not possible to apply a gradient elution when using tetraborate and thus some anions are not eluted within an acceptable time and resolution. Instead, a sodium hydroxide gradient was applied for the detection of azide in a conservation buffer for immunoaffinity column which is used for the separation of a pharmaceutically active protein from its reaction mixture. Here the goal was to ensure that the protein is not contaminated with azide. The IonPac AS11 column functionalized with quaternary ammonium compounds on poly(ethylvinylbenzene/divinylbenzene) was used in the separation of azide and chloride. The two peaks could be successfully resolved even when samples contained trace amounts of azide in the presence of chloride concentration of more than 190,000 times the azide. Using this system, the detection limit of azide was determined to range from 0.007 to 0.02 mg/L, which is reportedly lower than those previously reported in the literature. Though azide was efficiently separated from other anions with increasing concentrations of the eluent, the baseline is disturbed when sodium hydroxide concentration is increased, due to an effect on the conductivity detector response. When the concentration was reduced to initial values, the baseline began to minimize once again. The solution proposed to minimize this baseline drift was to use an eluent generator and carbonate removal device provided by Dionex, which is now common and considered the best approach when using NaOH eluents. Care also needs to be taken when dealing with very high concentrations of chloride as the chromatographic peak would overlap the azide peak, thus the only sample preparation necessary would be to dilute the sample prior to loading it. Overall, this method developed by Vincovic and Drevenkar is a suitable and robust tool for the analysis of trace amounts of azide in a variety of sample types. Along with simple sample preparation, the method selectivity, linearity, accuracy, precision, and sensitivity meets the criteria for application in the pharmaceutical industry.

***2.1.3 Determination of Sodium Sulfite*** Sodium metabisulfite is a common anti-oxidant that is used in pharmaceuticals. It decomposes into sulfite ions with $-2$ charge in aqueous solution thus making it ideal for analysis by HPIC. However, sulfite readily self-oxidizes into sulfate, so additional sample preparation is necessary to stabilize the sulfite ions. The application of HPIC on the analysis of sulfite is not a new endeavour. Herbranson *et al.* [28] developed a separation technique using suppressed conductivity detection in 1987. To overcome the oxidation problem, the group prepared the sulfite solution in 0.01% (v/v) formaldehyde which forms a detectable complex with the sulfite ions. The described technique allows for the quantitation of both sulfite and its reduced form sulfate in order to determine the true sulfite amounts in pharmaceutical samples. Separation was performed on a Dionex IonPac AS4 column using a carbonate eluent which provided efficient resolution of the sulfite and sulfate peaks in a quick 10 min run.

***2.1.4 Analysis of Clodronate*** Clodronate or disodium clodronic acid (disodium dichloromethylene bisphosphonic acid) is a member of the bisphosphonate class of drugs which are is used in the treatment of bone diseases such as hypercalcaemia, bone pain in patients suffering from multiple myeloma, or carcinoma of the breast, and is also a targeted treatment for Paget's disease and osteoporosis [29, 30]. Synthesis of clodronate is achieved by the two-step reaction of tetraisopropylmethylenebisphosphonate with sodium hypochlorite, followed by de-esterification with hydrochloric acid [29]. Clodronic acid is also a starting reagent for the synthesis of nucleotide analogs by neutralizing it with one equivalent of tributylamine to form monotributylammonium salt [29].

The impurities that may result from the synthesis of clodronate are chloride, partial esters, methylenebisphosphonic acid, monochloromethylenebisphosphonic acid as well as degradation products under basic conditions including orthophosphate and carbonylbisphosphonate. Identification of bisphosphonates is particularly difficult because of the lack of a chromophore for detection by UV or fluorescence. Although several methods in literature pertain to chromatographic detection of bisphosphonates, these studies utilize pre-column or post-column derivitizations. The sample preparation process for derivitization is time consuming, and most routine pharmaceutical laboratories are limited to the availability of specialized equipment that is necessary for the detection of these derivatives. However, two distinct approaches from Taylor [29] and Tsai et al. [30] have been identified and provide an alternative to derivitization of the non-chromophore bisphosphonate compounds.

Tsai et al. [30] developed a method in which the measurement of the bisphosphonate drugs could be achieved through the use of single column (Waters IC—PAK HR anion exchange column) ion chromatography with indirect UV detection. The UV detection measures the decrease in the UV absorption of the nitric acid eluent which absorbs within the range of 230–245 nm. As bisphosphonate ions are eluted, these anions replace the nitrate anions in the detector cell and what is seen in the chromatograph is a negative signal owing to the decrease in the mobile phase peak. To visualize these peaks on the chromatograph, the polarity of the detector output was reversed.

Chloride ions, one of the aforementioned impurities, tend to co-elute with the bisphosphonates, but in the presence of lower pH, the resolution of the two peaks is increased. Thus based on this, Tsai et al. chose nitric acid as a means of preserving the anionic charge of the compounds of interests. Rather than a gradient, 12 mM $HNO_3$ was used and detected at 245 nm. The group also quantitatively analyzed two other bisphosphonates, alendronate, and etidronate using different concentrations of the mobile phase, than had been used for clodronate. However, Taylor [29] recognized that at low pH conditions such as those carried out by Tsai et al. the structurally similar compounds to clodronate, such as monochloromethylenebisphosphonic acid and methylenebisphosphonic acid and most other impurities, would be in the monovalent $-1$ state. In addition to this, the isocratic mobile phase used would make separation of these impurities rather challenging. Although greater concentrations of acidic eluent may be used, the baseline tends to drift upwards, making integration of peaks difficult.

Alternatively Taylor [29] used a sodium hydroxide gradient as the eluent. Under alkaline conditions all analytes will be fully ionized thus producing a sample that is multivalent allowing the analytes to interact with the anion exchange column to varying degrees. Utilizing NaOH as the eluent also makes it possible to use suppressed conductivity detection and thus direct detection of analytes are possible. This study also recognized the importance of hydrophobic interactions of bisphosphonates with the column matrix for more adequate separation and thus tested five different anion exchange columns. The columns were used with a sodium hydroxide gradient and suppressed conductivity with a self-regenerating anion suppressor in the recycle mode. The IonPac AS5 column functionalized with alkanol quaternary ammonium ions gave the best selectivity for all the impurities. Other columns with stronger hydrophobic interactions such as the IonPac AS4 column required a high sodium hydroxide concentration in order to elute the strongly retained clodronate, resulting in the less charged impurities eluting in the solvent front. The OmniPac PAX-100 (quaternary ammonium functionalized ethylvinylbenzene/divinylbenzene polymer) gave the best resolution for the bisphosphonates when using acetonitrile as an organic modifier; however, the HPLC-grade acetonitrile gave background conductivity likely due to ionic impurities in the acetonitrile, and thus this column was not chosen. Taylor also explored the effect of column temperature on column selectivity and found that with the IonPac AS5, an increase in temperature caused the anions to have an increased retention time. The optimal temperature was thus chosen to be 45°C.

## 2.2 Fermentation Analysis

Traditionally, fermentation techniques have been employed in the food production industry such as dairy and bakery products, as well as wines and beers. In addition to these methods, recombinant genetic techniques have become more prevalent generating microorganisms to produce antibiotics, enzymes, amino acids many other pharmaceutically active agents and more recently, biofuels. Cultivation of these microorganisms requires a complex mixture of nutrients and minerals present in the correct concentration so as not to hinder or inhibit growth and production. In order to optimize yield and to control the fermentation process, a number of components need to be monitored and quantified. Analysis of the components can be problematic because of the high concentrations and simply because of the large amounts of different analytes in the media. HPIC is a user-friendly method with simple sample preparation that can analyze essential components of fermentation in a short time allowing for real time monitoring and optimization. Ion chromatography can be implemented for the measurement of a number of anions and cations including amino acids, carbohydrates, alcohols and other metabolic by-products, transition metals, vitamins, minerals, substrates, and products [31-36].

***2.2.1 Determination of Organic and Inorganic Anions*** Anion-exchange chromatography coupled with integrated pulsed amperometric detection (HPAEC-IPAD) is a powerful tool for monitoring the fermentation process. It can be applied to

the simultaneous analysis of carbohydrates, amino acids, and alcohols, which exist in conjunction with one another [33, 34, 36]. Carbohydrates are the primary carbon source for energy production and are thus essential substrates in media used for cell growth and synthesis. These include glucose, lactose, galactose, and ribose, to name a few. However, resolution is limited since the amount of carbohydrates present can result in peak tailing and co-elution. Further complications arise from the presence of amino acids and alcohols which may also co-elute with the carbohydrate peaks. Most microorganisms require amino acids to be supplemented in the growth media because they cannot synthesize those amino acids from other nutrients. When carbohydrates are not available, amino acids are used as an alternative energy source. Thus, it is important to monitor the substrate levels to ensure that deficiencies do not occur which can limit the production of the desired products. To ensure there is an adequate supply, the amino acid levels in the fermentation broth need to be monitored, but most amino acids are poor chromophores so require an alternative detection method to UV absorbance. Additionally, other analytes in the culture media which are chromophoric can interfere with the absorbances by amino acids [33]. As an alternative carbohydrates and amino acids can be oxidized and therefore can be directly measured by amperometry.

To determine the conditions necessary for optimal elution of a mixture of carbohydrates and amino acids in fermentation media, a retention study of a mixture of 42 different carbohydrates was analyzed by Hanko and Rohrer [33] in the presence of 30 amino acids. Anion exchange coupled with integrated PAD using a gold working electrode was used for separation on a Dionex AminoPac PA10 column with various initial NaOH eluent concentrations and isocratic durations, followed by a NaOH gradient. The group also designed a method for the separation of carbohydrates and amino acids in YPD (yeast extract peptone dextrose) broth, LB (Luria-Bertani) broth, serum-free protein-free hybridoma medium, and MEM (minimal essential medium). The analysis of these media did not require any sample preparation or carbohydrate removal unless the samples contained over 0.2% (10 mM) glucose.

Robinett and Herber [36] also described a method for carbohydrate analysis using PAD with a gold electrode. Separation was accomplished using a Dionex CarboPac PA1 column. Two elution systems were used: an isocratic elution system consisting of a 150 mM NaOH mobile phase and a gradient system using a combination of (A) water and (B) 50 mM NaOH-3 mM $CH_3COOH$. For the gradient analysis, the mobile composition was held constant at 94%A-6%B and varied linearly to 0%A-100%B. Separation of a mixture of carbohydrates in the order of alcohols, pentoses, hexoses, and disaccharides was achieved in 20 min.

Alcohol separation has also been described by Robinett and Herber [36]. With isocratic conditions, the analysis of alcohols was performed on sulfonated column (a Dionex IonPac ICE-AS1) with PAD detection (platinum electrode) and a 100 mM $HClO_4$ mobile phase. While carbohydrate analysis does detect ethanol and glycerol, co-elution of the two peaks often occurred. To resolve this, ion-exclusion methods were applied, where retention of the lower molecular mass alcohols is based on inclusion while retention of higher molecular mass alcohols is based on adsorptive properties with the resin. Analysis time is also only 20 min. The chromatogram seen in

Fig. 3 represents the separation of a mixture of alcohols with very efficient resolution between glycerol and ethanol.

Joergensen *et al.* [34] investigated the use of ion chromatography for the detection and quantitation of anions in media. Separation was done on two different columns, the IonPac AS5A and the OmniPac PAX-500. Both columns, equipped with an anion exchange latex produced good separation of acetate, formate, chloride, nitrate, sulfate, and phosphate with conductivity detection and an aqueous gradient of sodium hydroxide and methanol. The method was then applied to fermentation experiments of methanotrophic bacteria before and after a switch of substrate from methane to natural gas. It was found that before the switch to natural gas, acetate levels were very low, but steadily increased after the switch. Since acetate has been shown to inhibit the growth of methanotrophic bacteria as they are unable to utilize it, the group cultured the acetate utilizing bacteria *Comamonas acidovorans*, and followed the decline of acetate in the fermentation media by ion chromatography.

***2.2.2 Determination of Inorganic Ions and Trace Transition Metals*** Inorganic cations are important for cellular functions, such as membrane maintenance and function. However, depending on the organism, the optimal concentration for ions such as $Na^+$, $K^+$, $H^+$, $NH_4^+$, $Mg^{2+}$, as well as transition metals ($Fe^{2+}$, $Zn^{2+}$, $Mn^{2+}$, $Co^{2+}$, $Cu^{2+}$) will vary. Requirement of these cations also vary depending on the metabolic state of the organism—simply growing or producing antibiotics. The detection of these ions can be difficult due to the presence of chelating agents as well as interference from other components in the media. Moreover, the concentration of some ions will vary throughout the course of the fermentation process. Since some concentrations can be inhibiting, while insufficient amounts can also hinder growth, it is very desirable to apply a simple process to both detect and monitor inorganic cation concentrations over time. Various literature sources have investigated the role of ion chromatography in the detection of inorganic cations, which include alkali- and alkaline-earth metals and trace amounts of transition metals [31, 32, 34–36]. The methods that are described are not only important for monitoring and optimizing growth conditions, but also for quality control analysis prior to inoculation. Metal contamination from corrosion of steel equipment used in fermenters can occur during the fermentation process, and thus ion chromatographic methods can be used as a means to ensure the quality of the media and the machinery used.

Robinett *et al.* [35] improved on a method previously developed by the research group to simultaneously quantitate eight inorganic cations that are commonly present in culture media used for fermentation. The method offers a simple sample preparation, requiring only dilution and filtration of the broth and an analysis time of only 20 min. A Dionex CS12 column functionalized with carboxylic acids was used with a methanesulfonic acid isocratic eluent system and suppressed conductivity detection. This method can allow for the detection of up to 11 cations, if baseline resolution is not required; however, the group focused on five key ions (sodium, potassium, ammonium, magnesium, and calcium). A limit of detection of 0.5–1 μg/mL was reported, so

dilution is the only sample preparation necessary prior to analysis. Furthermore, conductivity detection removes the possibility of interference from organic cations such as alcohols and carbohydrates since their presence is not recorded by the conductivity detector. However, some cations such as amines and transition metals are recorded, but their peaks do not interfere with the analytes of interest. Application of the method was performed on fermentation broths for *Saccharomyces cerevisiae* and *Escherichia coli*, as well as mammalian and an insect cell cultures over a period of time. In these cases, the concentration of cations could be monitored over time. A chromatogram of the separation of eight cations in a standard solution is seen in Fig. 4.

A procedure developed by Bell [32] offers two solutions to the removal of interferences that could pose difficulties for metal ion determination. Two methods: ashing of the sample or solid-phase extraction were used; however, these processes require large amounts of media. Ashing also takes several hours to complete, thus this method is not as desirable. Nonetheless, Bell was able to separate calcium, copper, lead, zinc, nickel, cobalt, cadmium, iron (III), magnesium, and manganese close to 1 ppm concentrations, in less than 35 min once sample preparation was completed. Here, ion-pair chromatography is used, where octanesulfonate is used as the counter ion in the eluent consisting of tartaric acid, sodium octanesulfonate, and methanol on an adsorbosphere HS $C_{18}$ column with conductivity detection. Further sensitivity could be obtained from the application of post-column derivitization reactions prior to detection.

Alternatively, Joergensen *et al.* [34] used acid digestion to remove interferences from the media, which provides a faster means of sample preparation than had been described by Bell [32]. The group investigated fermentation broth used to culture methanotrophic bacteria which have a special requirement for copper. At high concentrations, copper is poisonous to the organism, and more challenging is that

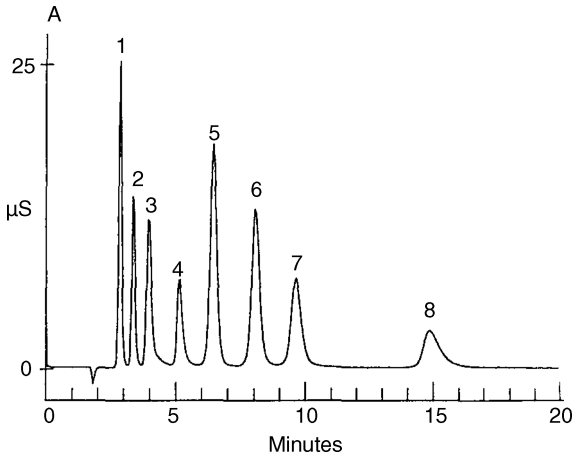


***FIGURE 4*** *Separation of eight inorganic cation standards by isocratic elution using a Dionex IonPac CS12, taken. Baseline resolution could be achieved for all eight cations, which were detected using suppressed conductivity. Peaks: 1 = Li, 2 = Na, 3 = $NH_4$, 4 = K, 5 = $Mg^{2+}$, 6 = $Ca^{2+}$, 7 = Sr, 8 = $Ba^{2+}$. Reproduced from Ref. [34] with permission.*

the range between optimal and detrimental concentrations is very narrow. Difficulties arise for the measurement of copper like other transition metals, because organic components present in the media can form complexes with the metals. In order to remove the organic materials, it is necessary to degrade them before sample analysis. A simple method for digestion was the use of nitric acid followed by heating in a microwave oven. This sample preparation could be applied to fermentation media samples that were filtered prior to acid digestion or samples containing bacterial cells. For the analysis of cations such as sodium, ammonium, potassium, magnesium, and calcium, only filtration prior to sample analysis is necessary. A polystyrene/divinylbenzene column functionalized with sulfonic acid (IonPac CS1) was used with an isocratic mobile phase of hydrochloric acid and DL-2,3-diaminopropionic acid (DAP-HCl) and conductivity detection. The detection limit of these alkali- and alkaline earth metals was 10 ppb. On the other hand, measurement of transition metals was achieved by the isocratic chelation ion chromatographic method following nitric acid digestion producing a detection limit of 0.01–0.05 ppb using UV detection. An application of the method described here was on a magnesium washout experiment. A culture was changed from medium containing magnesium to one without. The magnesium concentration in the fermenter was followed by ion chromatography over time, and the limiting concentration for growth was found, thus allowing for the identification of the relationship between magnesium concentration and growth of the bacteria.

The use of ion-pair chromatography is described by Schuster and Hampel [31] to analyze trace levels of transition metals. The metal ions were separated on a Waters μBondapak $C_{18}$ column using a tartaric acid–sodium octanesulfonate eluent. However, the downside to this method is the use of post-column derivitization with sodium 4-(2-pyridylazo) resorcinol (PAR) for the formation of colored-complexes which can then be monitored by spectrophotometric detection. Regardless of the extra step that is required, detection of the metal ion complexes in the visible region allows for determination in the μg/L range. The fermentation process that Schuster and Hampel are concerned with is for the production of citric acid by *Aspergillus ruger*. Trace levels of $Fe^{2+}$, $Mn^{2+}$, and $Zn^{2+}$ are monitored in the fermentation broth, in the presence of high levels of citric acid and sucrose without interference. However, high concentrations of $Mg^{2+}$, $K^{+}$, and $Ca^{2+}$ can overload the column, thus affecting the retention times of the transition metals. To compensate for the high concentrations of mono- and divalent ions, dilution of the sample is necessary, but this increases the detection limit.

From these examples it is shown that HPIC offers a capability of analyzing complex media consisting of a variety of organic anions (carbohydrates and alcohols) as well as inorganic anions in order to determine optimal growth conditions for fermentations. The robustness of this chromatography is amenable to simple dilution and filtration sample preparation, and short analysis times.

## 2.3 Metabolomics Applications

The preceding sections have illustrated the usefulness of HPIC in both the production (Section 2.2) and the analysis (Section 2.1) of pharmaceuticals. Another important aspect is the analysis of the *effect* of drug compounds on cellular physiology, particularly in terms of proteins and metabolites. This section will address the use of HPIC for the analysis of metabolites and the next section will deal with the analysis of proteins. The metabolome is defined as the complement of low molecular weight metabolites in a cell under a particular physiological condition [37]. In terms of physiology, metabolites could be considered the final output and therefore serve as an indication of the "true" state of the cell [37]. Metabolites are more sensitive to perturbations than either mRNA transcripts (transcriptome) or protein level (proteome) and the metabolome is downstream of both and arguably more useful [37]. Therefore, determining the concentrations of all the metabolites in the cell is important in obtaining a true picture of cellular physiology. However, the chemical composition of cellular metabolites is extremely diverse [38], even in a "simple" bacteria such as *E. coli*. As such it is very difficult, and perhaps impossible to utilize a single analytical method to quantify all the cellular metabolites. In this regard the best approach seems to be to use a variety of analytical techniques, such as NMR [39, 40], GC-MS [39, 41, 42], and liquid chromatography separations combined with MS [39, 42]. In this section, we will present HPIC as one of many tools in the toolbox of the metabolomics researcher. Metabolomics platforms are being developed that utilize a variety of techniques that can each separate different types of compounds thereby gaining as much coverage of the metabolome as possible [42]. HPIC has already been used in combination with GC-MS and NMR to study metabolic flux on a large scale in *S. cerevisiae* [43]. In this context, then, HPIC can be used as a metabolomics tool to study the effect of various drug compounds on the metabolite pools in the cell. For the sake of simplicity the focus in this section will be on microbial cells, starting with a discussion on metabolite extraction considerations yet the same considerations are valid for higher life forms.

***2.3.1 Metabolite Extraction Considerations*** Metabolite turnover in microbial cells can be on the order of less then a second [44] and so it is desirable that when extracting compounds for analysis there is some sort of procedure in place that rapidly quenches metabolism [45]. Centrifugation to harvest cells is typically a no–no as it is expected that cellular physiology would change drastically, with respect to low-molecular weight compounds. For example, even a 2 minute centrifugation of *E. coli* cells can have a drastic effect on intracellular nucleotide pools [4]. Rapid filtration (<45 s) or rapid quenching with −40°C, 60% methanol followed by centrifugation are typical protocols [4, 45], although with the later leakage of metabolites from the cell following quenching can be a problem that one needs to be aware of [45].

The other consideration for metabolite extraction is the method used for extraction. The method of choice typically depends on the type of compound that is being extracted. Cold methanol is a popular choice [45] and of the current extraction

protocols seems to be able to extract the widest range of compounds [46]. Other protocols include acidic acetonitrile [47], boiling buffered ethanol [12, 48], formic acid [49–51], solid phase extraction [14], potassium hydroxide [52], and perchloric acid [46, 53]. One must consider the type of compound being extracted as well as the composition of the buffer for the IEC separation. For example an extraction in methanol will require the sample to be dried in either a vacuum centrifuge or lyophilized, then re-suspended in the appropriate buffer (typically a hydroxide or carbonate buffer for anion-exchange). Acid or base extractions on the other hand may not need this step and can be injected directly on the column. Thought should also be put into the volume of sample to be extracted in order to detect the compound(s) under investigation. This may require some trial and error, particularly if approximate cellular concentrations are unknown or if the procedure has never been done before. The volume of sample to be extracted will depend on the sensitivity of the detector, the intracellular concentration of the analyte, and the growth state of the cells.

***2.3.2 Carbohydrate and Organic Acid Analysis*** As described in Section 2.2.1, the combination of HPAEC with PAD is a very powerful tool for the separation of a wide range of carbohydrates such as reducing and non-reducing sugars, oligosaccharides, amino sugars, and sugar phosphates, to name a few [8]. In strong alkaline conditions, the hydroxyl groups of sugars are transformed into oxyanions, thereby allowing them to be separated by anion-exchange chromatography [9]. Detection by pulsed-amperometry is accomplished with the use of a gold working electrode which can catalyzed the oxidation of the –COH groups in the sugar under the alkaline conditions used for the chromatography [8]. The use of a triple-pulse allows the regeneration of the oxidized electrode surface after each measurement [9]. Smits et al. [14] utilized HPAEC-PAD for the analysis of glycolytic intermediates in yeast. This study, however, was utilized primarily to analyze the effectiveness of a solid-phase extraction technique on the recovery of seven sugar phosphate metabolites from yeast cells. Groussac and Ortiz [12] developed a HPAEC-PAD protocol to quantify 12 glycolytic and pentose phosphate metabolites in *S. cerevisia* [12]. They used this protocol to then study intracellular metabolite concentrations after a glucose pulse. Another study used HPAEC in combination with several detection methods, including PAD, suppressed conductivity, and UV absorbance [48]. The authors used different columns for different types of metabolites and were able to analyze 14 sugar phosphates from yeast cell extracts with detection limits using PAD down to 0.5 μM [48]. Another group has coupled HPAEC-PAD to electrospray ionization mass spectroscopy (ESI-MS) in order to study the structures of capsular polysaccharides from *Sinorhizobia* spp. [15], demonstrating this technique as an alternative to NMR and GC/MS, which are usually used for polysaccharide analysis.

Although HPAEC-PAD has proven quite useful for the analysis of microbial carbohydrates, there are also several examples where HPAEC is coupled directly to mass spectrometry as a detection method. The high ionic strength of the eluant required for HPAEC provides a technical challenge but lowering the ionic strength of the buffer and using an online desalter prior to mass spectrometry seems to solve this

problem quite well [17]. ESI-MS coupled to HPAEC is the method of choice and has been used to quantify glycolytic intermediates (mainly sugar phosphates) and extensive metabolic flux analysis [10, 11, 13, 43].

Organic acids are commonly separated by HPAEC for analytical purposes [2,3] and there are some examples of HPAEC for the separation of microbial organic acids, like TCA cycle intermediates. A study in 1995 developed a method to separate 27 anionic metabolite standards from several different *Escherichia coli* metabolic pathways in a single run [5]. HPAEC was used in combination with suppressed conductivity and UV detection to quantify 16 of these metabolites in *E. coli* cell extracts, including several organic acids, sugar phosphates, and nucleotides [5]. Groussac and Ortiz [12] used HPAEC with suppressed conductivity detection to separate organic acids as well as some inorganic ions. The method was also applied to glucose pulse experiments in the yeast *S. cerevisiae*.

***2.3.3 Phosphorylated Nucleotides*** Traditionally, separation of cellular nucleotide pools in microbiological studies is achieved with two-dimensional thin layer chromatography (2D-TLC). This involved growing the cells on radioactive phosphate, followed by extraction, separation by 2D-TLC, and comparison to standards [50]. However, HPLC methods provide an attractive alternative to this approach, due to the fact that HPLC is more quantitative, gives better separations and does away with the need for radioactive labeling [4]. In light of this, several ion chromatography techniques have been used to study cellular nucleotides. In 1996, HPAEC was used with suppressed conductivity and UV detection to separate a variety of phosphorylated compounds from *Mycobacterium*, mainly mono-, di-, and triphosphonucleotides [54]. The authors utilized this method primarily to analyze phosphoantigens extracted from *Mycobacterium tuberculosis*. The chromatogram of

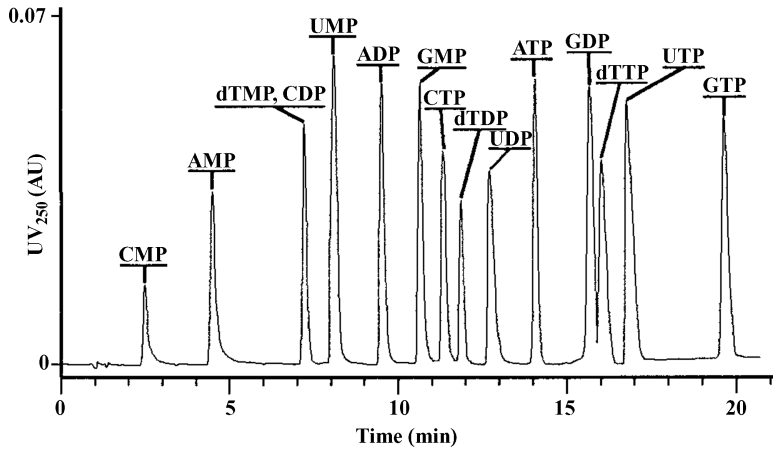


***FIGURE 5*** *Representative chromatogram of HPAEC separation of a standard mixture of phosphorylated nucleotides. Separation was performed on a Dionex IonPac AS11 with a gradient elution of NaOH. Detection was by absorbance at 250 nm. Reproduced from Ref. [54] with permission.*

the standard mixture of phosphorylated nucleotides from this study is shown in Fig. 5, which illustrates the effectiveness of HPAEC to separate these compounds.

Ion-pair reverse phase HPLC (IP-RPLC) is an alternative form of ion chromatography that has also seen use in separation phosphorylated nucleotides. IP-RPLC involves the addition of a lipophilic ion to the mobile phase in a typical RPLC separation [3]. For the case of nucleotides, tetrabutyl ammonium phosphate is typically used as an ion pair reagent [49, 52], which allows the separation of nucleotides in an RPLC column based on a combination of interactions with the non-polar stationary phase as well as potential anion-exchange interactions [3]. Buckstein *et al.* have developed a protocol for the analysis of nucleotide pools in *E. coli* based on a combination of IP-RPLC and HPAEC [4]. This method allowed the detailed analysis of cellular nucleotide pools at different points of the growth phase.

## 2.4 Multi-Dimensional Proteomic Separations

The proteome is the set of proteins in a biological system at a specific physiological time and condition. Proteomics can be applied to a cell, a microorganism, or body fluids in order to analyze the global changes in response to specific stimuli. Thus over the years, the discipline of proteomics has become an area of interest for the pharmaceutical field for drug development and action. Samples used in proteomics contain a very complex mixture of proteins and thus require more than one separation dimension to achieve the resolution necessary for analysis. Furthermore, protein or peptide mixtures require separation prior to analysis by mass spectrometry, which is a key analytical tool for proteome analysis in protein identification. Classically, the separation of proteomic samples was achieved by two-dimensional polyacrylamide gel electrophoresis (PAGE) where isoelectric focusing (IEF) is the first dimension followed by sodium dodecyl sulfate (SDS–PAGE) in the second dimension. The protein bands are then cut out of the gel and identified by matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) or electrospray ionization mass-spectrometry (ESI-MS) [55, 56]. Although this method is still widely used, 2D-PAGE also has its limitations. Difficulty arises for the analysis of low abundance proteins, proteins with extreme pIs (i.e., extremely basic or acidic), membrane proteins, and very large or small proteins [55–57]. Alternatively, the development of shotgun proteomics emerged as a solution to the limitations of 2D-PAGE analysis. This technique is a high throughput automatable system that couples HPLC with tandem mass spectrometry (MS/MS). Furthermore, multi-dimensional separation using two or more independent chromatography techniques prior to the MS allows for high resolution and sensitivity. Multi-dimensional proteomics can be applied to intact protein samples or for peptide mixtures.

***2.4.1 The role of Strong Cation Exchange (SCX) as a First Dimension Separation*** An evolving approach to proteomics is multi-dimensional separation method via online or offline 2D-LC-MS/MS, using SCX as the first dimension followed by reversed phase (RP) chromatography as the second dimension. This 2D approach thus separates on the basis of charge then hydrophobicity. Together, the

two dimensions simplify the mixture before it is analyzed in the mass spectrometer. Online 2D-LC-MS/MS uses a SCX column in series with an RP column to separate tryptic digests by a stepwise gradient where each peak is directly eluted onto the RP column. On the other hand, with the offline approach, the 1st dimension (SCX) is not directly coupled to the 2nd dimension [55, 56]. Rather, fractions from the SCX column are collected and later loaded onto the RP column. The online approach is much faster and sample loss is minimized. Although both methods are suitable for resolving large proteomes, the offline method offers significantly greater resolution of complex mixtures as well as greater peak capacity, leading to a higher number of identified proteins [56]. The use of SCX-RP-MS/MS can be advantageous for the analysis of complex whole-cell lysates from a variety of human extracts, such as human serum [58, 59], human liver [60], and human cortex [61].

Despite the resolution offered from the two-dimensional separation of SCX and RP, sample pre-fractionation prior to protein digestion is becoming necessary [58]. This approach addresses the problem of low abundance proteins being overpowered by high abundance proteins. Pre-fractionation increases the effective concentration of each protein in different fractions while reducing mixture complexity. As a result less interference between peptides occur and more proteins can be identified. In this respect, SCX can also be used for pre-fractionation of the protein mixture prior to digestion. Barnea *et al.* [58] investigated the differences between various pre-fractionation methods. The group found that SCX pre-fractionation followed by MudPIT (multi-dimensional protein identification technology) resulted in the identification of the largest number of proteins as well as the best coverage of those proteins.

## 2.5 General Considerations on Sample Preparation

When using IEC in metabolomics, proteomics, or examination of other complex mixtures from crude cell or organism lysates, one should consider the variety of biomolecules aside from the targets that will interact with the column. Generally one should think not only of the target molecules but the characteristics of the contaminants. The majority of biomolecules contain a charge. Other charged biomolecules will decrease the capacity of the column and as a result optimal sample separation is substantially decreased. Whole cell lysates will include molecules such as lipids, nucleic acids, carbohydrates, small molecule metabolites, and electrolites which are in various levels of abundance, but overall can be at a significant concentration relative to one's targets. One of the most common mistakes in ion exchange experiments is not removing the DNA and RNA, which will interfere with protein binding to the column. Another mistake is the wrong choice of buffering compound used during the cell or organism lysate extraction, such as the mistake of using a phosphate buffer with an anion exchanger.

Processing of the cell lysate needs to be done in order to ensure the removal of such unwanted cellular compounds. If such charged biomolecule contaminants are a problem, a pre-fractionation method is easily added including a different charged matrix HPIC, size exclusion chromatography, differential precipitation, and centrifugation or digestions. For example, treatment with DNases and RNases to remove nucleic

acids or centrifugation to pellet cellular debris, which include nucleic acids, larger carbohydrate polymers and lipids. Furthermore, it is desirable to remove high abundance proteins or metabolites so that the detection of low abundance targets is possible.

## 3 CLOSING REMARKS

As evidenced by the studies presented in this chapter, HPIC sees a wide range of uses across many areas in the pharmaceutical industry. Many powerful and advanced analytical chemistry methods are available to today's researcher and yet a more simple and "older" technique can accomplish many of the same tasks. A large number of physiological compounds are charged and easily lend themselves to be separated by IEC. With some thought to sample preparation, proper choice of column, eluant and detector, many important compounds can be rapidly separated and quantified. Everything from quality analysis to fermentation analysis to metabolomics and proteomic applications can use HPIC as a useful tool.

The goal of this chapter was to present an overview of some of the diverse uses of HPIC in the pharmaceutical industry. HPIC was presented as an analytical tool for diverse areas of pharmaceutical research, from analyzing drug production and purity to analyzing physiological effects of drug use. The power of HPIC lies in its simplicity of use, inexpensive instrumentation, and diversity of compounds it can separate and quantify.

## REFERENCES

1. Haddad PR, Nesterenko PN, Buchberger W. Recent developments and emerging directions in ion chromatography. J Chromatogr A 2008; 1184: 456–473.
2. Fritz JS, Gjerde DT. Ion chromatography. Weinheim: Wiley-VCH; 2000.
3. Weiss J. Handbook of ion chromatography. Weinheim: Wiley-VCH; 2004.
4. Buckstein MH, He J, Rubin H. Characterization of nucleotide pools as a function of physiological state in Escherichia coli. J Bacteriol 2008; 190: 718–726.
5. Bhattacharya M, Fuhrman L, Ingram A, Nickerson KW, Conway T. Single-run separation and detection of multiple metabolic intermediates by anion-exchange high-performance liquid chromatography and application to cell pool extracts prepared from Escherichia coli. Anal Biochem 1995; 232: 98–106.
6. Harris DC. Quntitative chemical analysis. New York: W. H. Freeman and Company; 2003.
7. Johnson DC, LaCourse WR. Liquid chromatography with pulsed electrochemical detection at gold and platinum electrodes. Anal Chem 1990; 62: 589A–597A.
8. Cataldi TRI, Campa C, De Benedetto GE. Carbohydrate analysis by high-performance anion-exchange chromatography with pulsed amperometric detection: the potential is still growing. Fresenius' J Anal Chem 2000; 368: 739–758.
9. Lee YC. High-performance anion-exchange chromatography for carbohydrate analysis. Anal Biochem 1990; 189: 151–162.

10. Kiefer P, Nicolas C, Letisse F, Portais JC. Determination of carbon labeling distribution of intracellular metabolites from single fragment ions by ion chromatography tandem mass spectrometry. Anal Biochem 2007; 360: 182–188.
11. Van Winden WA, Van Dam JC, Ras C, Kleijn RJ, Vinke JL, Van Gulik WM, Heijnen JJ. Metabolic-flux analysis of Saccharomyces cerevisiae CEN.PK113-7D based on mass isotopomer measurements of 13C-labeled primary metabolites. FEMS Yeast Res 2005; 5: 559–568.
12. Groussac E, Ortiz M, Francois J. Improved protocols for quantitative determination of metabolites from biological samples using high performance ionic-exchange chromatography with conductimetric and pulsed amperometric detection. Enzyme Microb Technol 2000; 26: 715–723.
13. Van Dam JC, Eman MR, Frank J, Lange HC, Van Dedem GWK, Heijnen SJ. Analysis of glycolytic intermediates in Saccharomyces cerevisiae using anion exchange chromatography and electrospray ionization with tandem mass spectrometric detection. Anal Chim Acta 2002; 460: 209–218.
14. Smits HP, Cohen A, Buttler T, Nielsen J, Olsson L. Cleanup and analysis of sugar phosphates in biological extracts by using solid-phase extraction and anion-exchange chromatography with pulsed amperometric detection. Anal Biochem 1998; 261: 36–42.
15. Chataigne G, Couderc F, Poinsot V. Polysaccharides analysis of sinorhizobial capside by on-line anion exchange chromatography with pulsed amperometric detection and mass spectrometry coupling. J Chromatogr A 2008; 1185: 241–250.
16. Paull B, Nesterenko PN. Novel ion chromatographic stationary phases for the analysis of complex matrices. Analyst 2005; 130: 134–146.
17. Walters JJ, Grayson MA, Gross ML, Hughes M, Shearer G, Kohl DH, Bashkin J. Ion-exchange chromatography followed by ESI-MS for quantitative analysis of sugar monophosphates from glucose catabolism. J Am Soc Mass Spectrom 2006; 17: 104–107.
18. Roy J. Pharmaceutical impurities—a mini-review. AAPS PharmSciTech 2002; 3: E6.
19. Jagota NK, Chetram AJ, Nair JB. Ion chromatography of amylamine and tert-butylamine in pharmaceuticals. J Chromatogr A 1996; 739: 343–349.
20. Krol J, Alden PG, Morawski J. Ion chromatography of alkylamines and alkanolamines using conductivity detection. J Chromatogr 1992; 626: 165–170.
21. LaCourse WR, Jackson WA, Johnson DC. Pulsed amperometric detection of alkanolamines following ion pair chromatography. Anal Chem 1989; 61: 2466–2471.
22. Vialle J, Navarro P, Nguyet TT, Lanteri P, Longeray R. Experimental design for the optimization of the separation of aliphatic amines by ion chromatography. J Chromatogr 1991; 549: 159–174.
23. Daigle R, Mehra MC, Mallet V. Single column ion chromatography of ammonium ion, alkali metals and alkyl amines. Chromatographia 1991; 32: 143–147.
24. Kersten B. Ion chromatography of polyhydroxy aliphatic amines of pharmaceutical interest. Chromatographia 1992; 34: 607–612.
25. Vinkovic K, Drevenkar V. Ion chromatography of azide in pharmaceutical protein samples with high chloride concentration using suppressed conductivity detection. J Chromatogr B 2008; 864: 102–108.
26. Annable P, Sly LA. Azide determination in protein samples by ion chromatography. J Chromatogr 1991; 546: 325–334.

27. Perati P, Rohrer J, Thomas D. Determination of azide in human plasma using a Reagent-Free ion chromatography system with suppressed conductivity detection. LC-GC North Am 2006; 24: 17.
28. Herbranson DE, Eliason MS, Karnatz NN. Development of a high performance ion chromatographic (HPIC) method for the determination of sodium metabisulfite in parenteral formulations. J Liq Chromatogr 1987; 10: 3441–3450.
29. Taylor GE. Determination of impurities in clodronic acid by anion-exchange chromatography. J Chromatogr A 1997; 770: 261–271.
30. Tsai EW, Chamberlin SD, Forsyth RJ, Dominic CB, Brooks MA. Determination of bisphosphonate drugs in pharmaceutical dosage formulations by ion chromatography with indirect UV detection. J Pharm Biomed Anal 1994; 12: 983–991.
31. Schuster G, Hampel WA. Determination of trace levels of transition metal ions in fermentation broth by ion-pair chromatography. Anal Chim Acta 1992; 258: 275–280.
32. Bell RG. Cation analysis of fermentaion broth by high performance liquid chromatography utilizing ion interaction. J Chromatogr 1991; 546: 251–257.
33. Hanko VP, Rohrer S. Determination of amino acids in cell culture and fermentation broth media using anion-exchange chromatography with integrated pulsed amperometric detection. Anal Biochem 2004; 324: 29–38.
34. Joergensen L, Weimann A, Botte HF. Ion chromatography as a tool for optimization and control of fermentation processes. J Chromatogr 1992; 602: 179–188.
35. Robinett RSR, George HA, Herber WK. Determination of inorganic cations in fermentation and cell culture media using cation-exchange liquid chromatography and conductivity detection. J Chromatogr A 1995; 718: 319–327.
36. Robinett RSR, Herber WK. Analysis of substrates and metabolites in fermentation broth by ion chromatography. J Chromatogr A 1994; 671: 315–322.
37. Kell DB, Brown M, Davey HM, Dunn WB, Spasic I, Oliver SG. Metabolic footprinting and systems biology: the medium is the message. Nat Rev Microbiol 2005; 3: 557–565.
38. Nobeli I, Ponstingl H, Krissinel EB, Thornton JM. A structure-based anatomy of the E. coli metabolome. J Mol Biol 2003; 334: 697–719.
39. Dunn WB, Ellis DI. Metabolomics: current analytical platforms and methodologies. Trends Analyt Chem 2005; 24: 285–294.
40. Grivet JP, Delort AM, Portais JC. NMR and microbiology: from physiology to metabolomics. Biochimie 2003; 85: 823–840.
41. Koek MM, Muilwijk B, vanderWerf MJ, Hankemeier T. Microbial metabolomics with gas chromatography/mass spectrometry. Anal Chem 2006; 78: 1272–1281.
42. van der Werf MJ, Overkamp KM, Muilwijk B, Coulier L, Hankemeier T. Microbial metabolomics: toward a platform with full metabolome coverage. Anal Biochem 2007; 370: 17–25.
43. Kleijn RJ, Geertman JMA, Nfor BK, Ras C, Schipper D, Pronk JT, Heijnen JJ, Van Maris AJA, Van Winden WA. Metabolic flux analysis of a glycerol-overproducing Saccharomyces cerevisiae strain based on GC-MS, LC-MS and NMR-derived 13C-labelling data. FEMS Yeast Res 2007; 7: 216–231.
44. De Koning W, Van Dam K. A method for the determination of changes of glycolytic metabolites in yeast on a subsecond time scale using extraction at neutral pH. Anal Biochem 1992; 204: 118–123.

45. Mashego MR, Rumbold K, De Mey M, Vandamme E, Soetaert W, Heijnen JJ. Microbial metabolomics: past, present and future methodologies. Biotechnol Lett 2007; 29: 1–16.
46. Maharjan PR, Ferenci T. Global metabolite analysis: the influence of extraction methodology on metabolome profiles of Escherichia coli. Anal Biochem 2003; 313: 145–154.
47. Rabinowitz JD, Kimball E. Acidic acetonitrile for cellular metabolome extraction from Escherichia coli. Anal Chem 2007; 79: 6167–6173.
48. Loret MO, Pedersen L, Francois J. Revised procedures for yeast metabolites extraction: application to a glucose pulse to carbon-limited yeast cultures, which reveals a transient activation of the purine salvage pathway. Yeast 2007; 24: 47–60.
49. Payne SH, Ames BN. A procedure for rapid extraction and high-pressure liquid chromatographic separation of the nucleotides and other small molecules from bacterial cells. Anal Biochem 1982; 123: 151–161.
50. Bochner BR, Ames BN. Complete analysis of cellular nucleotides by two-dimensional thin layer chromatography. J Biol Chem 1982; 257: 9759–9769.
51. Fischer M, Zimmerman TP, Short SA. A rapid method for the determination of guanosine 5′-diphosphate-3′-diphosphate and guanosine 5′-triphosphate-3′-diphosphate by high-performance liquid chromatography. Anal Biochem 1982; 121: 135–139.
52. Little R, Bremer H. Quantitation of guanosine 5′,3′-bisdiphosphate in extracts from bacterial cells by ion-pair reverse-phase high-performance liquid chromatography. Anal Biochem 1982; 126: 381–388.
53. Winder CL, Dunn WB, Schuler S, Broadhurst D, Jarvis R, Stephens GM, Goodacre R. Global metabolic profiling of Escherichia coli cultures: an evaluation of methods for quenching and extraction of intracellular metabolites. Anal Chem 2008; 80: 2939–2948.
54. Poquet Y, Constant P, Peyrat MA, Poupot R, Halary F, Bonneville M, Fournie JJ. High-pH anion-exchange chromatographic analysis of phosphorylated compounds: application to isolation and characterization of nonpeptide mycobacterial antigens. Anal Biochem 1996; 243: 119–126.
55. Fournier MJ, Gilmore JM, Martin-Brown SA, Washburn MP. Multidimensional separations-based shotgun proteomics. Chem Rev 2007; 107: 3654–3686.
56. Nagele E, Vollmer M, Horth P, Vad C. 2D-LC/MS techniques for the identification of proteins in highly complex mixtures. Expert Rev Proteomics 2004; 1: 37–46.
57. Jiang X, Ye M, Zou H. Technologies and methods for sample pretreatment in efficient proteome and peptidome analysis. Proteomics 2008; 8: 686–705.
58. Barnea E, Sorkin R, Ziv T, Beer I, Admon A. Evaluation of prefractionation methods as a preparatory step for multidimensional based chromatography of serum proteins. Proteomics 2005; 5: 3367–3375.
59. Liu X, Valentine SJ, Plasencia MD, Trimpin S, Naylor S, Clemmer DE. Mapping the human plasma proteome by SCX-LC-IMS-MS. J Am Soc Mass Spectrom 2007; 18: 1249–1264.
60. Gao M, Deng C, Yu W, Zhang Y, Yang P, Zhang X. Large scale depletion of the high abundance proteins and analysis of middle- and low-abundance proteins in human liver proteome by multidimensional liquid chromatography. Proteomics 2008; 8: 939–947.
61. Pan S, Shi M, Jin J, Albin RL, Lieberman A, Gearing M, Lin B, Pan C, Yan X, Kashima DT, Zhang J. Proteomics identification of proteins in human cortex using multidimensional separations and MALDI tandem mass spectrometer. Mol Cell Proteomics 2007; 6: 1818–1823.

# *Automation in Quality*

**J. Alford**
*Purdue University, West Lafayette, Indiana*

## 1 INTRODUCTION

The noble pursuit of pharmaceutical product quality typically involves the identification and measurement of critical product attributes, some of which have historically been difficult to determine automatically in real time (e.g., dissolution rate) or to correlate with measurements that are available on-line. Pharmaceutical manufacturing has also historically involved a large number of manual operations. Such operations can be prone to operator error and to variability due to differences between operators, despite the existence of training programs and documented SOPs.

Pursuing product quality has also involved drawing conclusions about product quality based on sampling and testing of small amounts of active pharmaceutical ingredient (API) and finished products. While such testing is still required, it is insufficient by itself to conclude that all product meets specifications. Validated processes are required. This is in keeping with the well-known axiom:Quality cannot be tested into a product; it must be built into the processes that make the product

Examples of the limitations of testing product samples include the following:

- A small sample, taken from a lightly contaminated sterile operation, may not contain any contaminating microorganisms, leading to a false conclusion about the sterility of a lot.
- Certain product quality tests conducted on individual vials, cartridges, capsules, and/or tablets may destroy the sample, so obviously cannot be performed on all product.

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing,*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

- Periodic sampling may not catch unacceptable product resulting from a process deviation that occurred in-between sampling events.

Automation is a technology that can improve product quality from a number of perspectives. It can:

- replace manual operations for operations that are well defined and measurable, thus reducing the frequency of deviations due to operator errors,
- reduce operator contact with product and/or hazardous environments,
- provide a more consistent manufacturing process, thereby reducing variability, as operations are performed exactly as configured or programmed to do so, with alarms generated and/or pre-programmed corrective action taken immediately when process measurements or logic indicates that a problem exists.

Automation is also an enabling technology, permitting functionality not possible or practical to implement with manual operations: Examples include:

- implementation of virtual sensors (i.e., on-line model estimates of parameters not easily measured directly),
- implementation of advanced control algorithms (e.g., model predictive control), permitting more consistent (i.e., less variable) and optimal process operations,
- rapid collection, consolidation, sorting, and analysis of data needed in generating batch reports and in investigating process deviations.

The pharmaceutical industry typically defines product quality in terms of product safety, strength, identity, and purity, as well as use of quality (i.e., validated) processes. Validated processes include those used in manufacturing, laboratory, and plant utility operations that play a role in producing a quality product. Automation can be an important contributor regarding all of these five quality attributes. For example, automation can:

- generate and read bar codes (e.g., on raw materials, final product containers), which helps with material identity,
- monitor and control utility operations (e.g., generation of water-for-injection – WFI),
- assist with laboratory operations (e.g., assay instruments, LIMS),
- monitor and consistently control most manufacturing unit operations (e.g., bioreactors, crystallizers, separation columns, filling machines, packaging operations), which helps to produce products with consistent purity and strength. Automating unit operations also helps to reduce manual operations and avoid contamination possibilities, thereby contributing to product and personnel safety,
- help collect, display, analyze, and report data (e.g., for use in investigating potential product quality deviations).

## 1.1 Brief History of FDA Sponsorship and/or Endorsement of Initiatives Involving Automation

One of the FDA's most well-known initiatives regarding automation was the publication of its *Guide to Inspection of Computerized Systems in Drug Processing* [1]. This

was primarily intended as an educational tool for the FDA's field inspectors, most of whom, at the time, were not highly knowledgeable about computer systems.

The publication of FDA's *Guide to Inspection of Computerized Systems in Drug Processing* prompted the Pharmaceutical Manufacturer's Association (PMA) to form two computer system validation committees consisting of representatives of many pharmaceutical companies. One of the committees focused on GMP systems and the other on GLP systems. These committees developed and published several technical articles in subsequent years, including an overview article on a computer system's life cycle [2], and follow-up articles expanding on various key computer system life cycle activities (e.g., installation qualification, [3]).

In the mid-1990s, the International Society of Pharmaceutical Engineering (ISPE) published a leading reference book titled *Good Automation Manufacturing Practices (GAMP)*, which has since been updated several times [4].

By the dawn of the 21st century, the pharmaceutical industry had acquired a reputation as operating in a stifling regulatory environment, characterized, e.g., by highly bureaucratic and time-consuming procedures required to make changes to improve and optimize existing manufacturing processes. One result was an industry lagging others by several years in adopting new state-of-the-art technology (e.g., on-line analytical systems, model predictive control). Another result was long manufacturing cycle times due to numerous hold points during the process when off-line testing was pursued in order to get the "green light" for forward processing of lots.

In recognizing this dilemma, the FDA began working with industry and other groups in supporting and/or endorsing various initiatives to cultivate more efficient and cost-effective manufacturing operations while maintaining a strong emphasis on product quality. These initiatives included:

- Quality by Design (QbD),
- Risk-Based Decision Making,
- Process Analytical Technology (PAT).

***1.1.1 Quality by Design (QbD)*** Quality by Design (QbD) is a systematic approach to product and process design and development [5]. The concept was initially developed in the early 2000s with the expectation that it would be an evolving process for several years. QbD was part of a FDA Chemistry, Manufacturing, and Controls (CMC) pilot program, launched in July 2005, that contained features documented in the International Conference on Harmonization's guidelines, Q8 Pharmaceutical Development and Q9 Quality Risk Management and Process Analytical Technology (PAT).

Quality by Design includes the following key elements:

- Target the product profile.
- Determine critical quality attributes (CQAs).
- Link raw material attributes and process parameters to CQAs and perform risk assessment.
- Develop a design space.

- Design and implement a control strategy.
- Manage product life cycle, including continual improvement.

Automation can clearly play a role in some of the above elements (e.g., designing and implementing a control strategy).

***1.1.2 Risk-Based Decision Making*** A pharmaceutical cGMP initiative, Pharmaceutical cGMPs for the 21st Century: A Risk-Based Approach, was launched in August 2002. It applies to human and biological drug products and veterinary drugs and has several objectives, some of which are:

- to encourage the early adoption of new technological advances by the pharmaceutical industry,
- to facilitate industry application of modern quality management techniques, including implementation of quality systems approaches, to all aspects of pharmaceutical production and quality assurance,
- to encourage implementation of risk-based approaches that focus both industry and agency attention on critical areas.

Selected guidance documents resulting from this initiative include the following:

- Part 11, Electronic Records, Electronic Signatures – Scope and Application
  This guideline clarifies the scope and application of the 21CFR Part 11 regulation and provides for enforcement discretion in certain areas. The guideline explains the goals of this initiative and removes barriers to scientific and technological advances and encourages the use of risk-based approaches.
- Sterile Drug Products Produced by Aseptic Processing: Current Good Manufacturing Practices
- PAT – A Framework for Innovative Pharmaceutical Manufacturing and Quality Assurance

An example of the use of risk-based decision making is the use of Failure Modes and Effects Analysis (FMEA). In this technique, often used during system or product design and in manufacturing operations, the various failure modes of a process/product are evaluated. Each failure mode is given a score for:

- severity of consequences of the failure,
- anticipated frequency of occurrence of the failure,
- ability to measure/detect the failure.

The three scores are multiplied together to provide the RPN (Risk Prioritization Number), which is then compared to a pre-determined value (threshold). Usually, design/operation changes will only be considered for those failure modes that have a RPN above the threshold. This means that some failure modes (e.g., with low RPN)

will remain with the product or system and be a reality of life. Note: some flexibility exists in how companies administer FMEAs. For example, some applications only consider severity and frequency in determining RPNs.

Another example of risk-based decision making is in reliability engineering. In this discipline, system testing occurs iteratively such that the results of each iteration drive action plans to improve the process/product. Once the improvements are made, another round of testing is pursued. The objective metric is often the estimated "mean time between failure (MTBF)" with projects often having a defined MTBF "end in mind" when the process/product is in development. Each iteration of testing and follow-up improvements should result in an increase in MTBF. Several different published reliability growth models can help in quantifying such improvement.

***1.1.3 Process Analytical Technologies (PAT)*** Process Analytical Technology (PAT) is a FDA/industry-promoted methodology based on designing, analyzing, and controlling manufacturing processes based on:

- an understanding of the scientific and engineering principles involved,
- identification of the variables that affect product quality.

Regarding scientific and engineering principles, PAT supports the need that chemical, physical, and biopharmaceutical characteristics of the drug and other components of the drug product be determined. Some of the science involves relatively mature disciplines (e.g., determining a compound's identity and purity) while other aspects are not as mature (e.g., measuring and controlling physical attributes such as solid form, particle size, and particle shape).

The PAT initiative is consistent with the current FDA belief that quality cannot be tested into products but should be built in by design.

From the FDA's perspective, the desired state of pharmaceutical manufacturing includes objectives such as:

- product quality and performance ensured through the design of effective and efficient manufacturing processes,
- product and process specifications based on a mechanistic understanding of how formulation and process factors affect product performance,
- quality assurance being continuous and real time,
- relevant regulatory policies and procedures being tailored to accommodate the most current level of scientific knowledge,
- risk-based regulatory approaches recognizing both the level of scientific understanding and the capability of process control related to product quality and performance.

The above objectives make it clear how PAT ties into QbD and risk management.

Defining PAT in more detail, Process Analytical Technologies are "systems for analysis and control of manufacturing processes based on timely measurements,

during processing, of critical quality parameters and performance attributes of raw and in-process materials and processes to assure acceptable end product quality at the completion of the process." PAT is actually a broad-based continuous improvement initiative, encouraging greater use of new technologies, risk-based management, increased automation, and data analysis to achieve objectives such as greater understanding of processes, increased on-line decision making, reduced cycle time, improved process control, and reduced process variation, while maintaining product quality [6]. The paradigm shift to PAT can, in part, be illustrated schematically as shown in Figure 1.

PAT measurements are a subset of the overall PAT vision. They range from traditional on-line measurement sensors [monitoring critical process parameters (CPPs) or monitoring parameters that are highly correlated to CPPs] to the moving of more complex off-line or laboratory-based assays to the production floor and interfacing them on-line or at-line to the process. In part, the PAT initiative captures the evolution from the historical use of "univariate" process measurements (e.g., temperature, pressure) to ones that better capture the biological, physical, and/or chemical attributes or state of materials being processed.

However, PAT involves more than appropriate measurements. It includes use of multivariate tools for design, data acquisition and analysis, process control, knowledge management, and continuous process improvement. The objective of PAT is that the information generated will be timely (avoiding unnecessary process delays and holds) and can be used for appropriate real-time or near-real time decision making and control of existing process lots; i.e., the goal is to enhance understanding and control of the manufacturing process. Reference 7 captures this aspect of PAT with the term "continuous real-time quality assurance." PAT is believed to be one of several contributors to the ultimate goal of operating a capable, compliant, robust, in-control, and reproducible process. That is, PAT can be an important part of building quality into a process, as opposed to testing for quality "after the fact."

Some of the expected benefits from implementing PAT include:

- real-time forward processing and product release,
- reduction of cycle times via use of on-, in-, or at-line measurements and controls,
- reduction of rejected product and wastes
- reduced manual operations
- better understanding of processes (e.g., via developing of CQA–CPP correlations)
- facilitating more "continuous" (vs. batch) processing, resulting in increased capacity

While the topic of PAT has received much recent attention in the literature and at technical conferences, primarily due to the FDA's recent endorsement, PAT has actually been a successful part of many commercial pharmaceutical processes since the 1980s. For example, use of on-line mass spectrometry and high performance liquid chromatography (HPLC) have a long track record for improved understanding and control of many pharmaceutical API processes.

***1.1.4 PAT and Process Control*** Once the important properties of API and drug product formulations are understood, CPPs are defined and their proven acceptable ranges determined. Process variables that affect CPPs need to be prospectively identified and integrated into automation strategies and process control recipes. A multivariate approach is often needed. Process control often involves integrating several direct sensor and analytical technologies, interfaced to plant automation systems. For example, the thickness of some coating layers (including the enteric and active product coating layers) on a bead substrate can be CPPs, which, in turn, are influenced by the controlled temperature and humidity of the air stream and liquid feed rate of coating solutions/suspensions in a fluidized bed spray dryer.

***1.1.5 Other Quality Perspectives*** While this chapter focuses primarily on pharmaceutical product quality, it is noted that quality can also be defined in other ways, such as in the "quality" of the environment in which employees work. The environment for employees is usually associated with safety (e.g., OSHA, corporate) and environmental (e.g., MACT rule) regulations and best practices that are implemented/managed in plants with the help of automation systems.

As examples, automation systems are known to exist in pharmaceutical manufacturing that, as part of their mission, help reduce employee exposure to toxic chemicals and/or reduce the chances of a chemical explosion. They are also used, e.g., in controlling the pH of plant waste streams to city utilities and in monitoring the performance of gas scrubbers used to prevent release of unacceptable amounts of toxic vapors into the atmosphere. Automation systems are also used in monitoring building environments, e.g., with regard to fire detection and security, and in generating associated alarms.

## 2 DEFINITIONS

### 2.1 Automation

The discipline of executing process control, data collection, and/or data analysis automatically and, usually, in real time.

### 2.2 Clean-in-Place (CIP) [8]

The act of cleaning process equipment and/or systems (tanks, vessels, filling machines, distribution lines, etc.) without dismantling them.

### 2.3 Commissioning

A well-planned, documented, and managed engineering approach to the start-up and turnover of facilities, systems, and equipment to the end user that results in a safe and functional environment that meets established design requirements and stakeholder expectations.

## 2.4 Computer Program [8]

A collection of logically interrelated statements or instructions that, when executed by a computer, makes possible the performance of a pre-defined task.

## 2.5 Computer System [8]

A group of hardware components assembled to perform in conjunction with a set of software programs. Collectively, they are designed to perform a specific function or group of functions.

## 2.6 Control Algorithms

The use of logic and/or mathematical equations, usually coded or configured into software, that operates on one or more input signals (often an error signal representing the difference between a measurement's current value and its desired value) and determines the value of one or more output signals that are then directed to final control element(s), such as valves.

## 2.7 Critical Process Parameter [8]

A control parameter that has a direct relationship to the quality, safety, effectiveness, or performance of the intermediate or final product.

## 2.8 Distributed Control System

The most comprehensive, functional, and sophisticated of the various options for process control. Such systems include the commonly expected components and functionality for process control such as input/output (I/O) modules, human machine interface, control algorithms, alarm functions, and continuous historian. They are also able to communicate with a wide range of peripherals and field devices using a variety of analog and digital communication protocols. Such systems can accommodate plants containing many thousands of I/O points.

## 2.9 Final Control Elements

A variety of hardware devices, normally installed in the field, that receive a signal originating from the automation system and converts it to useful work. Examples of final control elements are continuous range control valves, on-off valves, multiplexing valves, and robotic arms.

## 2.10 Fo [8]

The time, in minutes, to produce the equivalent steam sterilization effect as at 121.1°C (= 250°F). A measure of the lethality of a sterilization process.

$F_o$ can also be defined as the logireductions in living microorganisms/spores due to a sterilization operation.

## 2.11 Ground Loop [8]

The grounding of a circuit at more than one point.

## 2.12 Hardware [8]

Physical apparatus or equipment that makes up a computer and its related peripheral devices (e.g., tape drives, disk drives, printers) or other system. This term is also used to describe various pieces of equipment in the system.

## 2.13 Historian

A database, connected to or an integral part of an automation system, which contains the values of process measurements, alarms, and other relevant plant data. There are two types of historians (1) continuous historians, embedded, e.g., in DCSs, and (2) batch historians, appropriate for batch processes, which may be supplied either by a DCS/PLC vendor or a third party. The data contained in historians are typically used for trend plots for operators, batch reports, deviation investigations, statistical process control charts, model building, and other data mining applications.

## 2.14 IQ [8]

1. The performance of documented verification that an equipment/system installation adheres to approved contract specification and achieves design criteria. The IQ is developed from P&IDs, electrical drawings, piping drawings, purchase specifications, purchase orders, instrument lists, engineering specifications, equipment operating manuals, and other necessary documentation. All developmental documentation will be included in an IQ. The manufacturer's recommendations, local and state codes, and the cGMP should also be considered. The IQ precedes the Operational Qualification (OQ).
2. Establishing the documentary evidence that a sub-system or equipment is installed in compliance with the technical specifications, standards, codes, and regulations.
3. Documented verification that all key aspects of hardware installation adhere to appropriate codes and approved design intentions and that the recommendations of the manufacturer have been considered.

## 2.15 Life Cycle [8]

An approach to computer system development and operation that begins with identification of a user's requirements and continues through design, integration, qualification, validation, control, and maintenance, ending only when commercial use of the system is discontinued.

## 2.16 Measurement Uncertainty [8]

An estimate of the range of values around the measurement value in which the accepted value is believed to fall. The total uncertainty represents the sum of a measurement of the random error and the estimated bounds of the systemic error.

## 2.17 Operating Parameter [8]

The process variables that are measured to monitor and maintain the normal sate of a process or system. All factors, including control parameters, that may potentially affect the process state of control and/or fitness for use of the final product.

## 2.18 Operating System [8]

A set of programs provided with a computer that function as the interface between the hardware and the applications programs.

## 2.19 OQ [8]

1. The documented verification that the equipment/system performs per design criteria over all defined operating ranges. The OQ includes qualification of operating and maintenance records. The OQ precedes the Performance Qualification (PQ).
2. Establishing the documentary evidence that a sub-system or set of equipment is capable of repeated operation within the limits defined in the specifications.

## 2.20 Programmable Logic Controller

One of the most popular options for process control. These systems were originally developed as an electronic replacement for mechanical relay racks and were programmed in a language called ladder logic. In recent years, their functionality has expanded such that the distinction between DCSs and PLCs has become fuzzy. Most PLC systems handle up to a few hundred I/O, are often associated with individual process unit operations, and are easily interfaced to plant DCSs, if needed.

## 2.21 Process Analytical Technologies (PAT)

An FDA industry initiative, encouraging greater use of new technologies, risk-based management, increased automation, and data analysis to achieve greater understanding of processes, increased on-line decision making, reduced cycle time, improved process control, and reduced process variation, while maintaining product quality.

## 2.22 Process Control

The automated and/or manual manipulation of final control elements (e.g., valve position, pump speed) in order to keep desired measured parameters/attributes within pre-defined ranges.

## 2.23 Process Measurements

The measurement or estimates of product or process attributes via the use of direct measurement hardware sensors, virtual (model-based) sensors, and/or analytical systems.

## 2.24 Product Quality

The adherence of manufactured product meeting specifications regarding strength, identity, safety, and purity, and having been manufactured by a validated process.

## 2.25 Qualification [8]

1. Used to describe the testing and review of a piece of equipment, system, or sub-system of a process to assure its fitness for use. Qualification deals with components or elements of a process while validation deals with the entire manufacturing process for a product. Qualification is also the operation aimed at proving, with regard to materials, equipment, or personnel, that the required conditions are met and that they actually provide the expected results.
2. Action of proving that any equipment works correctly and actually leads to the expected results.

## 2.26 Quality by Design (QbD)

A systematic approach to product and process design and development that focuses on critical quality attributes.

## 2.27 Retrospective Validation [8]

Establishing documented evidence that a system does what it purports to do based on a review and analysis of historical data and information obtained during the production of clinical or marketable product.

## 2.28 Software [8]

A collection of programs, routines, and subroutines that control the operation of a computer or computerized system.

### 2.29 Steam-in-Place (SIP) [8]

The process for steam sterilizing (or sanitizing) process systems (tanks, vessels, lyophilizers, filling machines, distribution lines, etc.) without dismantling and autoclaving.

### 2.30 Worst Case [8]

A set of conditions encompassing the upper and lower processing limits and circumstances that pose the greatest chance of process or product failure when compared to ideal conditions. Such conditions do not necessarily induce product or process failure.

## 3 DESCRIPTION OF AUTOMATION SYSTEMS

Automation covers a broad range of systems and technology, most being microprocessor based. Applications include process monitoring and control, data collection and analysis, generation of electronic documents, laboratory operations, and supporting utility operations.

A key assumption in the use of automation systems in helping to manage a manufacturing process is the existence of electronic measurements. This is in keeping with the well-known axiom:

If you can't measure it, you can't manage it

A generic diagram of a process automation system is shown in Fig. 2.

The system can include:

- field elements such as sensors and transmitters that generate electronic signals representing process measurements (e.g., temperature, pH) and then send them to a process control computer,
- field elements such as valves and transducers that can receive electrical signals from a process control computer and convert them into useful mechanical work,
- on-line or at-line analytical systems (e.g., mass spectrometers, HPLC, IR/NIR),
- operator and engineering workstations, containing the human machine interface (HMI),
- process control computer (e.g., DCS, PLCs, PCs), including functionality enabling the configuration of common applications (e.g., PID algorithms, alarms) and the ability to program other applications (e.g., coding on-line models),
- continuous and batch historians, including applications (e.g., statistical control charts, alarm record Pareto charts) and the ability to program other applications,
- connectivity to other computers (e.g., LIMS, MES, alarm loggers, paging systems), including specialized sub-systems (e.g., real-time expert systems; 9, 10),
- appropriate Control and LAN networks (e.g., Fieldbus, DeviceNet, Profibus, Ethernet).

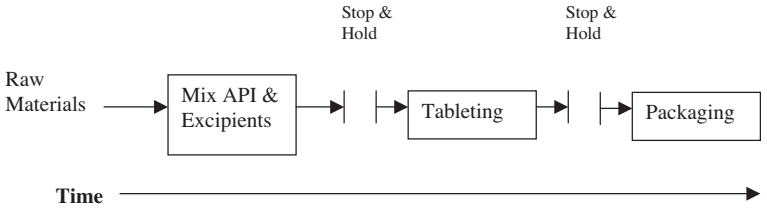


**"End of Phase" testing of quality (to reduce risk in moving to the next stage) is based on off-line laboratory tests / assays.**

**PAT based process**

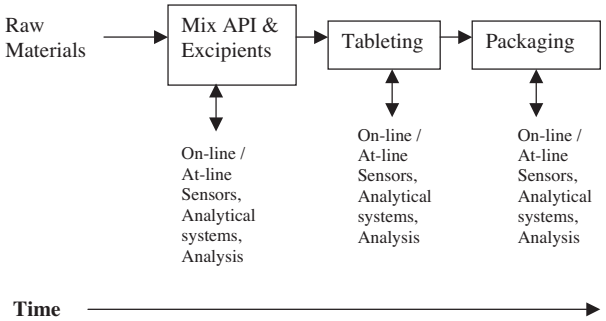


**Forward Processing (to reduce risk in moving to the next stage) is based on on-line and/or at-line measurements and analysis**

***FIGURE 1*** *A paradigm shift from lab-based testing to PAT.*

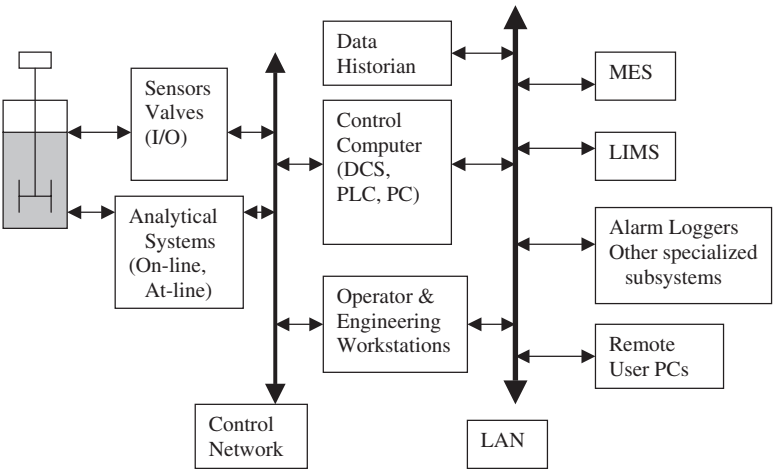


***FIGURE 2*** *Generic diagram of a process automation system.*

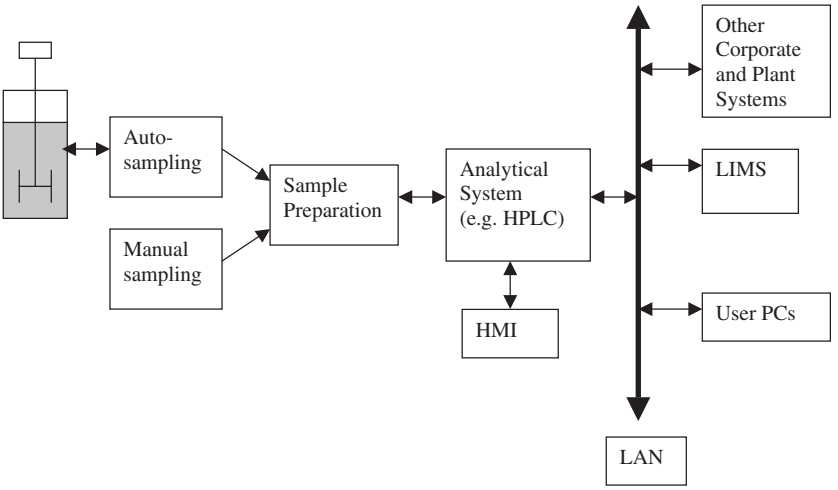


***FIGURE 3*** *Generic diagram of a laboratory automation system.*

Laboratory systems represent another major use of automation. However, a laboratory operation is seldom totally automated. Manual operations can include acquiring a sample, certain steps of sample preparation, and a lab supervisor's review and release of assay results. A generic diagram of a laboratory automation system is shown in Fig. 3.

Equipment vendor-supplied embedded PLCs represent a large subset of automation systems used in the pharmaceutical industry. A generic diagram of a product vial/cartridge-filling machine is shown in Fig. 4.

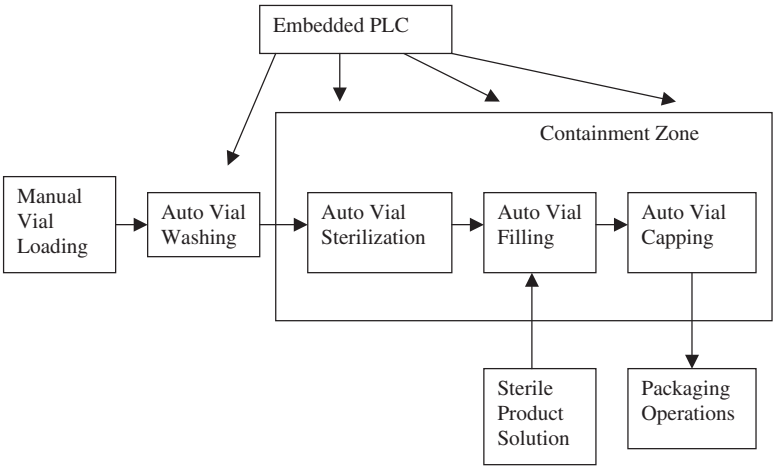


***FIGURE 4*** *Generic diagram of an automated product filling line with embedded PLC.*

## 3.1 Safety Instrumented Systems (SISs)

For highly hazardous processing operations, the addition of a separate safety instrumented system (SIS) may be required if the installed plant DCS or PLC system dose not have sufficient overall reliability to satisfy the requirements of the hazardous operation. Most SISs involve independently operating hardware and software but may, in some cases, utilize a plant's DCS HMI. The requirements associated with a SIS are documented in the standard ANSI/ISA S-84.

## 3.2 Automation Hierarchy

An alternative way to view automation operations is with respect to ANSI/ISA S–95, which is the international standard for the integration of enterprise and control systems. This standard defines several levels of plant computer systems, as follows:

- levels 0–2, focusing on real-time discrete, continuous, and batch process control, including monitoring, control, diagnostics, data archiving, HMI, process control recipes, and alarming,
- level 3, focusing on manufacturing execution systems (e.g., detailed production scheduling, dispatching, maintenance management),
- level 4, focusing on enterprise resource planning systems (e.g., plant production scheduling, operational management, and other business systems).

It is important to define and understand the different levels of computer systems as industry is progressively moving toward integrating (i.e., linking) the different levels together, despite their use of different data bases and operating paradigms. In support of this integration, many companies are merging together their IT and plant automation groups.

## 3.3 Process Monitoring

***3.3.1 Process Sensors*** Many different electronic analog sensors are used in automating pharmaceutical processes. Among the most common sensors are those for measuring:

- Temperature,
- gage pressure,
- gas flow rates,
- liquid flow rates,
- agitator mixing speed,
- weight of tank contents (e.g., load cells, differential pressure),
- pH,
- dissolved oxygen,
- conductivity,
- tank level [e.g., microwave (i.e., radar), ultrasonic, differential pressure].

Some of the less commonly used analog sensors include:

- oxidation reduction potential (redox),
- dissolved carbon dioxide,
- turbidity (optical density),
- ion probes,
- paramagnetic oxygen and infrared carbon dioxide analyzers.

Digital sensors are also in common use, such as:

- conductance or capacitance probes (e.g., for detecting when foam in a vessel reaches a pre-determined level),
- limit switches on valves (verifying their fully open or closed position),
- proximity sensors (e.g., noting the location of a bottle, cartridge, vial, robotic arm).

Most analog process sensors/transmitters have historically been designed to have a 4- to 20-mA output. Dedicated wires (e.g., twisted pair) for each transmitter are installed from the field to computer I/O cabinets where the signals are converted to digital form (e.g., via 12- or 16-bit A/D converters) and then read by the computer. Plants incorporating large numbers of analog sensors, then, have historically needed to install a large amount of field wiring. Also, the analog transmission of signals, especially low current or low voltage signals, can be susceptible to radio frequency interference (RFI) and electromagnetic interference (EMI) unless special care is taken during installation, such as:

- using shield materials (e.g., foil) on signal wiring, grounded at only one end,
- avoiding installation of signal and power wiring in parallel and in close proximity to each other.

An alternative to 4- to 20-mA analog transmission has been use of RS 232, a digital communication protocol, often used in computers communicating with peripherals (e.g., printers). This protocol has transmission length constraints of only a few hundred feet and, as with 4- to 20-mA signals, only allows communication between two devices.

In recent years, more sophisticated and functional digital alternatives to analog wiring and RS 232 have been developed. Formats/protocols include:

- HART (Highway Addressable Remote Transducer). This protocol superimposes/overlays digital communications on top of the standard 4- to 20-mA analog signal,
- Ethernet [IEEE 802.3 – a technology for local area networks (LANs), which defines wiring and signaling standards for the physical layer of the network protocol],
- Foundation Fieldbus [IEC 61158 – a family of industrial computer network digital protocols used for real-time distributed control; permits daisy-chain, star, ring, branch, and tree network topologies),

- Profibus (Process Field Bus; a standard for field bus communications in automation systems),
- DeviceNet (a communication protocol used in automation to interconnect control devices for data exchange. It is based on controller area network (CAN) technology. It defines the media, physical, data link, and application layers of the ISO/OSI 7 layer model,
- Modbus (a serial communication protocol uses to establish master-slave/client-server communication between intelligent industrial electronic devices).

Most of these digital communication options allow for major reductions in analog wiring as well as reduced susceptibility to RFI and EMI. Also, many field instruments are now "smart." That is, they provide on-line diagnostic information as well as the measurement itself. For example, smart sensors now contain embedded microprocessors and an A/D converter such that the output of the sensor is converted to engineering units in digital form within the sensor device. Also, the sensor output includes additional information useful for diagnostics (e.g., a sensor involving a photo eye can communicate if the photo eye is dirty). Smart valves contain sensors (e.g., that measure valve stem position, supply pressure, and actuator pressure) that can enable an automation system to diagnose on-line a problem with the valve. Such smart field devices are allowing manufacturing plants to move from a "preventive maintenance" paradigm, involving expensive plant shutdowns and manual checks of field devices, most of which are still in good working order, to a more cost-effective and efficient "predictive maintenance" paradigm.

It is important for sensors located in processing equipment or in piping associated with such equipment to be compatible with all cleaning and sterilization requirements. For example, liquid mass flow meters operating via the Coreolis principle are popular since the fluid, whose flow rate is to be measured, travels only through a vibrating "U" tube. No pockets or cavities exist in which material can accumulate and be a source of contamination.

Also, many sensors installed through the walls of manufacturing vessels, such as temperature RTDs and pH electrochemical sensors, are smooth cylindrical probes that contain no surface cavities or pockets and, therefore, can be easily and reliably cleaned and sterilized.

In some cases, instruments and/or their transducers/transmitters must be interfaced to computers via intrinsic safety barriers, or housed in explosion-proof boxes, which minimizes the chances of an explosion due to generation of electrical sparks.

General considerations for instrumentation installation include the following [11]:

- Avoid sample tubing that allows the accumulation of fluids.
- Avoid dead legs in instrument taps and process piping; ensure the design allows for proper drainability.
- Use hygienic fittings; avoid threaded fittings.
- Ensure materials and liquid-filled elements (e.g., electrochemical probes) are not harmful to the product.

Critical process instruments must be properly calibrated. This requires that [11]:

- a calibration procedure is followed,
- the calibration is documented,
- the calibration is performed by qualified personnel,
- the calibration (if in the United States) is traceable to a NIST standard.

Other instrumentation and associated gasket requirements may exist regarding materials of construction, especially when direct contact with product is possible. These include the following [11]:

- The material must be non-reactive and non-absorptive.
- The material must be cleanable by CIP fluids.
- The material must be capable of being exposed to live steam during SIP operations.
- The material must be smooth and free of scratches, pits, and other surface blemishes.

Most validation protocols require verification/certification of the product contact materials [12, 13].

***3.3.2 On-Line/At-line Analytical Systems*** Process measurements can include more sophisticated analytical systems that typically contain their own computer which processes a high volume of raw data into meaningful results that can be presented to its own HMI or to an external automation system for display, control, and/or logging purposes.

One example of an on-line or at-line analytical system is a process mass spectrometer, multiplexed to a plant's bioreactors [14] Figure 5. These systems can measure with high accuracy the composition of gas streams entering and leaving the bioreactor, enabling the near real-time monitoring of oxygen, carbon dioxide, nitrogen, argon, and other volatiles (e.g., alcohols). The systems can detect compound concentrations from 10 ppm to 100% with traditional Faraday collectors, and concentrations lower than 10 ppm with multiplier tube detectors. Interfacing process mass spectrometers to bioreactors is relatively straightforward, taking advantage of the positive pressure inside bioreactors to drive sample gas to the mass spectrometer, which operates under vacuum. Sample conditioning is focused on filtering out any particulates and avoiding water vapor in the sample gas from condensing. Many mass spectrometer systems contain their own multi-port sampling valves and include automated calibration. A gas chromatography front end is normally not needed, as most of the compounds in a fermentation gas sample have at least one mass/charge fragment that is different from those in other compounds present. This feature allows analysis of a gas sample to occur in only a few seconds. Also, since bioprocesses are relatively slow changing, one mass spectrometer can be used to serve many

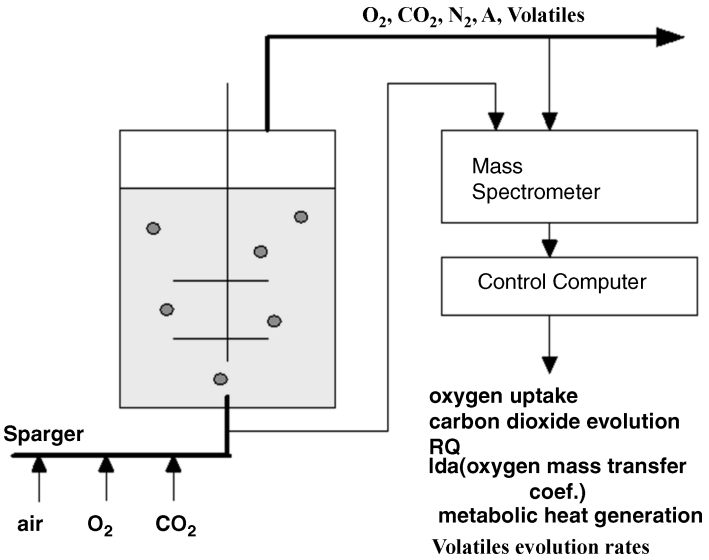


***FIGURE 5*** *Adding a mass spectrometer to a bioreactor.*

bioreactors, thereby resulting in a major reduction in analytical system cost per bioreactor.

Another analytical system example is on-line HPLC. HPLC is a very common laboratory instrument in the pharmaceutical industry, used for assaying the concentration of compounds in a mixture. A sample to be assayed is introduced to a tubular column containing a solid material (e.g., a resin) and operated using a carrier gas. Sometimes the column is subjected to a programmed temperature ramp to help separate compounds in the sample mixture and/or to shorten total analysis time, which is usually several minutes per sample. The different compounds in the sample absorb and desorb from the column resin at different rates and so arrive at the detector at different times. Assuming ideal development of the application, each peak on a chromatogram (detector intensity vs. time) will be well shaped, be unique to a specific compound, and have a stable well-defined baseline to facilitate assay computations. In cases when (1) sample collection and preparation are amenable to automation and (2) potential benefits are substantial, HPLC has been moved to the plant floor and interfaced directly to processing equipment [15].

Analytical systems similar to HPLC, such as ion chromatography and gas chromatography, are also used in the pharmaceutical industry. Basic principles of operation are the same, although items such as sample introduction and choice of detectors can be different.

Another example of on-line analytical systems is use of IR/NIR for monitoring the concentration of specific chemicals in mixtures, such as in complex raw materials, in bioreactor broth, or in blending and drying operations [16].

IR/NIR technology takes advantage of radiation absorption properties of chemicals at various wavelengths. Features of some commercial NIR systems with diode array detectors include:

- no moving parts,
- a non-invasive technique,
- results in less than a second,
- ability to operate from battery power and send results to a receiver via radio frequency (useful for containment areas where wire cables are to be minimized).

In some applications involving complex mixtures, calibration of IR/NIR systems has been a significant challenge – due to other chemicals that may be present in the mixture being analyzed that may absorb radiation at the same wavelength(s) as the chemical of interest.

***3.3.3 Virtual Sensors*** Virtual sensors are on-line model predictions of important process parameters that are impractical to measure directly in the time frame desired. For example, there are relatively few hardware sensors that are viable for use in sterile bioreactors since many sensors cannot survive the sterilization phase, may fail in the presence of a gritty highly agitated viscous broth, and/or cannot maintain calibration for the days or weeks that may be required to complete the batch operation. Examples of bioprocess parameters that are desired for on-line monitoring and/or control purposes include cell mass and substrate concentrations [17]. These process parameters are generally available as off-line assays (and may even be on-line for a few specialized bioprocesses), but the time delay involved for most bioprocesses is too long for this information to be useful for on-line control and real-time troubleshooting purposes.

Studies have shown that for many cases, the use of process mass spectrometry-derived information, such as culture oxygen uptake and/or carbon dioxide evolution, plus a simple model can accurately estimate, on-line, important parameters such as cell mass and substrate concentration [18, 19]. Models include use of empirical relationships, stoichiometric equations, and/or neural nets.

## 3.4 Process Control Loops

One of the most important functions of a process automation system is the ability to control process parameters.

The basic control of a process parameter usually involves three key elements:

- One or more process measurements
- Control algorithm
- One or more final control elements

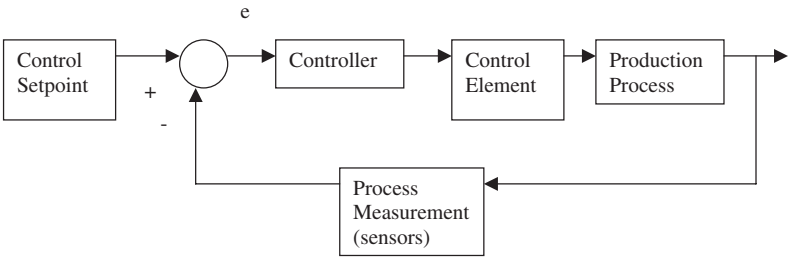


***FIGURE 6*** *Typical process control feedback loop.*

These are often configured as a feedback control loop, such as shown in Fig. 6.

In Fig. 6, the control setpoint is the desired value of the process parameter. The setpoint is typically compared in real time to the incoming measurement (which could be a direct field measurement, a virtual sensor, or calculated variable) in order to create an "error" (e) signal. The controller, often using a proportional + integral + derivative algorithm, determines what corrective action, if any, is needed and calculates an appropriate output value. Other portions of the automation system take the result (i.e., output) of the control algorithm and convert it to an appropriate signal to drive a final control element (e.g., valve). The following text provides additional detail on the measurement, control algorithm, and final control element portions of a control loop.

***3.4.1 Control Algorithms*** The objective of a control algorithm is to receive an indication of how the current value (measurement or calculated value) of a process parameter compares to its desired value (setpoint) and then to determine the corrective action needed, if any, to drive the current value to the setpoint. Control is often complicated in that time delays and process dynamics are usually involved in a process responding to a specific corrective control action (e.g., a valve changing position), so control algorithm tuning parameters need to accommodate such phenomena. For example, the pH in a solution in a vessel, as measured by a pH probe, may be different than the desired value. An automation system may then be programmed to add an acid or base feed to the solution. The dynamics of the feed rate control signal will need to account for the fact that the pH response of the solution in the vessel will not be immediate, nor will it reach a final steady state value quickly, due to factors such as:

- the location of the pH probe relative to the entry point of acid/base,
- the vessel's mixing time.

Of course, there are other complicating artifacts of pH control as well, since the pH response to the addition of a finite amount of acid or base will depend on where the current pH is on the 0–14 scale (due to the highly non-linear titration curve) and on the current buffering capacity of the solution being controlled.

Continuous control algorithms can take one of several forms:

- feedback (single input-single output),
- feed forward,
- adaptive,
- multi-variable (multiple inputs and/or outputs),
- ratio,
- cascade,
- fuzzy logic,
- model reference control,
- model predictive control.

As a few general comments on the above list:

- The most common feedback control algorithm is known as PID (proportional + integral + derivative). Mathematically, one version of it is:

$$\text{Controller output} = K^*e + K/T_i^* \int e\,dt + K^*T_d^*de/dt$$

  where "$e$" is the error signal (difference between setpoint and current measurement), and $K$, $T_i$, and $T_d$ are tuning constants that are determined by process or automation personnel using one of several available tuning procedures. All commercial process automation systems have one or more versions of the above algorithm available for use. Since many versions of the PID algorithm exist, a tuned PID loop from one vendor usually cannot be transferred to another vendor's system without retuning.Commercial products (e.g., from Expertune, DCS vendors) are available to automatically tune a process controller. These products usually require that the process be perturbed (i.e., during tuning) from normal setpoint values in order to determine how the process responds to control element output changes.
- Feed forward is a means of proactively controlling a process by measuring process disturbances and adjusting control action before the disturbance affects the controlled value of the desired process parameter.
- Adaptive control can have several meanings, but most often refers to the automated adjustment of tuning parameters (i.e., $K$, $T_i$, or $T_d$) based on factors such as batch time or process conditions. Adaptive control is sometimes needed in batch processes when the state of the process and dynamics involved change significantly as the process progresses. As an example, a fermentation starts out with very low cell density in aqueous solution with a broth viscosity about 1 cp, with very low oxygen and agitation requirements. As the fermentation proceeds, cell mass increases by several orders of magnitude, sometimes resulting in a

non-Newtonian broth viscosity of 200 apparent cp or more, with high oxygen and agitation requirements. Therefore, a specipic finite increase in air sparge rate to correct a given error in dissolved oxygen could have a dramatic effect early in the batch but a much smaller effect later on. The integral term of a PID controller can help deal with much of this process difference, but adaptive tuning can also help.

- Ratio control is used, when it is desired to control the ratio of two parameters (e.g., the ratio of dry air to liquid feed in a coating operation).
- Cascade control is used, e.g., when using certain physical control loops to control another measurable parameter. For example, it is usually desired to control dissolved oxygen in a bioreactor. While a defined $DO_2$ control valve does not formally exist, $DO_2$ can still be well controlled by controlling the agitation rate, airflow, oxygen feed, or other process parameters in a cascade control loop configuration. In this case, there are actually two control loops:
  1. the $DO_2$ control loop, with its own PID controller, known as the Master Loop,
  2. the e.g., agitation loop, with its own PID controller, which gets its setpoint from the output of the Master Loop controller, and which is known as the Slave Loop.
- Model Predictive Control (MPC) is an advanced multi-variable control algorithm currently supported by some commercial systems (e.g., Delta V). It is more commonly implemented for continuous steady-state chemical processes (e.g., petrochemicals) than it is for non-steady-state batch processes, but a few batch process applications have been reported.

Readers are referred to process control references [20–22] for further information of the various forms of continuous control algorithms, types of control, and controller tuning techniques. Specific to model predictive control, readers are referred to Refs. 23 and 24.

***3.4.2 Final Control Elements (Valves, Pumps, Robotic Devices)*** As a result of process measurements and logic programmed or configured into an automation system, decisions can be automatically made and implemented to manipulate mechanical devices in the field to do useful work.

While many types of valves exist in the process industries, a few are well suited to sanitary operations, as they are easily cleaned and sterilized, and contain few, if any, cavities in which material can accumulate and be a source of contamination. Valves used most often for sanitary operations include [11]:

- diaphragm valves (especially for on-off control of fluids),
- solenoid valves (for on-off control),
- ball valves (for process isolation),
- other regulating valves (for continuous control operations).

Some of these valves (e.g., solenoid valves) operate directly from electrical signals originating from the automation system's I/O. Traditional continuous control valves typically operate via a 3- to 15-psi pneumatic pressure force and therefore require a current to pressure (*I/P*) transducer in the control loop, leading to the final control element.

Inappropriate specification of continuous range control valves and their actuators is the source of many control loop problems in industry. For example, it is theoretically desired for control valves to operate within the 20–80% values of their range, avoiding operating near their fully open or fully closed positions. However, many engineers, in the spirit of providing safety margins, specify oversized control valves, resulting in their operation near the end of their range.

Other valve challenges include:

- non-linearity for several types of valves (e.g., equal percentage valves). The most common way to analyze a control loop is to assume all portions of the loop are linear elements, connected together. Having some elements as non-linear (e.g., control valves, or the process itself) significantly complicates the theoretical analysis of the loop and in determining appropriate tuning constants.
- stiction (a term combining the words "sticking" and "friction") – which is the minimum amount of controller output required to start a valve moving. Significant valve stiction can cause cyclic oscillation of the controlled variable.
- hysteresis ("lost motion" of the control valve). Significant hysteresis can also cause cycle oscillation of the controlled variable.

Some final control elements exist as variable speed pumps. Peristaltic pumps, e.g., are a popular choice for pilot and small production scale liquid feed applications, as they contain no cavities to promote contamination and are easily calibrated and cleaned.

Some final control elements exist as digital valves (e.g., solenoid valves or multi-port multiplexing valves), activated by electrical signals from the automation system and capable of providing an electronic feedback signal as to their position. Such valves can be found on automated process sampling systems in which samples from different vessels, calibration fluids, and/or test standard fluids are directed, as programmed, to an on-line or at-line analyzer. Digital valves are also used for on-off control of certain fluids fed to processes, such as an ammonia, caustic, or acid addition to a fermentor for pH control.

Robotic applications are often found in product-filling machines where a mechanical arm/hand device is used to place bottles, vials, cartridges, caps, filling apparatus, etc. in a desired two- or three-dimensional physical location. While the software involved with such applications is normally stable, the same is not necessarily true of the mechanical moving parts of the system, which can fail or drift over time. Automated procedures are often needed to, e.g., periodically bring the robotic arm back to a known, verifiable "home" position – to confirm the robotic element is still operating correctly.

## 3.5 Identification Techniques/Devices

***3.5.1 Bar Codes*** A bar code is a machine-readable representation of information, usually implemented as dark ink on a light background to create high and low reflectance. Light from a scanner is absorbed by the dark bars and reflected by the light spaces. Ultimately, the reflected light signal is represented in a bar code reader as 1s and 0s. Originally, bar codes stored data based on the widths and spacings of printed parallel lines, but printed codes today also come in patterns of dots, concentric circles, and text codes hidden within images. Bar codes can be read by optical scanners (containing a photocell detector) called bar code readers or scanned from an image by special software. Bar codes are widely used to implement Auto ID Data Capture (AIDC) systems that improve the speed and accuracy of computer data entry. An advantage of bar coding over other methods is its cost, costing less than $0.01 per bar code to implement (compared, e.g., to over $0.10 for some forms of RFID). Bar codes are usually reference (i.e., identification) numbers. The reference numbers are then used by a computer to look up all the pertinent detailed data associated with the reference (i.e., ID) data. For example, a bar code on a grocery store item will identify the food item but not other related information, such as the cost and final store destination. It is the database in the computer that will contain information such as cost, manufacturer, and location, and that will then automatically update, e.g., a merchandise inventory.

Uses of bar coding include:

- tracking most items sold in grocery and department stores,
- tracking of item movement, e.g., rental cars, airline luggage, nuclear waste, mail, and parcels,
- use on special event tickets (e.g., sporting events) that require validation before allowing the holder to enter the special event venue.

***3.5.2 RFID (Radio Frequency Identification)*** RFID [25] is a device/material identification data collection technology that uses electronic tags for storing data Fig. 7. It includes use of an RFID microchip attached to a radio antenna. Tags can either be battery powered (called active tags) or derive their power from the RF waves coming from a reader (called passive tags).

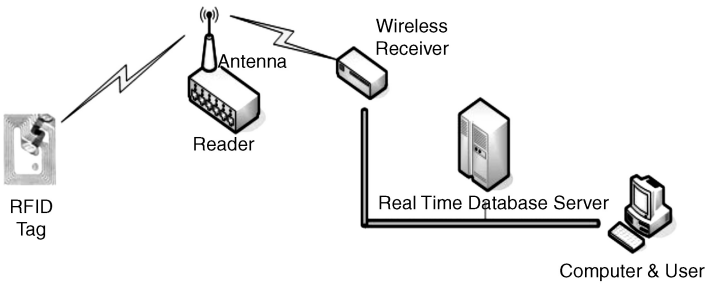


***FIGURE 7*** *RFID system diagram.*

Active tags, which cost up to a few dollars, can be used to transmit up to 300 ft indoors and more than a thousand feet outdoors. Passive tags, which cost about $ 0.10, can be read up to about 10 ft from the microchip tag.

Hybrid RFID tags also exist, known as "semi-passive" tags, which combine passive characteristics with a battery that allows the device to beep, blink, or perform some operation. For example, a semi-passive tag on refrigerated items can include a sensor that, when interrogated, reports the temperature range during shipment.

Unlike bar codes, which must be in close proximity and line of sight to the scanner for reading, RFID tags do not require line of sight and can be embedded within packages. In addition, RFID-tagged items can be read many times faster than bar codes. RFID tags hold more data than bar codes. For example, a bar code might indicate a product's model number while an RFID tag might include a unique serial number for an item in addition to the model number.

Known RFID applications include:

- tracking livestock,
- vehicle, container, and people tracking, e.g., putting RFID tags on children in amusement parks to help locate them if lost,
- fast checkout of books in libraries,
- retrieving pictures of people from a database, based on the RFID embedded in ID cards they present at security gates.

In the pharmaceutical industry, several pilot projects exist for products most vulnerable to theft and/or counterfeiting (e.g., OxyContin at Purdue Pharma and Viagra at Pfizer). The pilot projects are evaluating the use of tags both on pallets and on individual bottles containing product.

## 3.6 Process Automation Computer Links to Other Computers

While process automation computers (e.g., DCSs, PLCs) are highly functional with regard to monitoring, alarming, and controlling a manufacturing process, they often do not provide all the functionality that a plant site may need regarding the management of their manufacturing operations. Other computer systems may exist to provide other manufacturing functionality, and it may be desirable to provide electronic communication to these computers. Some of the computers that a process automation system may link to include:

- on-line or at-line analytical systems,
- batch historian,
- real-time expert systems,
- alarm loggers,
- LIMS (Laboratory Information Management System),
- manufacturing execution systems.

A few general comments on the above list are as follows:

- Many analytical systems (e.g., IR, NIR, HPLC, mass spectrometry) deal with large amounts of raw data in order to calculate a final analytical result. They normally contain their own computer to do the raw data processing and then send the result to a LIMS or process automation system.
- Most DCS or PLC systems contain a "continuous" historian as an embedded portion of the system. The DCS or PLC then retains data for a relatively short period of time (e.g., a few weeks) and uses the data, e.g., to generate real-time trend plots for operators. However, "continuous" historians have significant constraints regarding batch processes. Therefore, separate batch historians are often interfaced to the automation system in such cases. Batch historians have greater capability than continuous historians in (1) managing data by lot number and by batch step/phase, (2) displaying data in different time formats (e.g., batch relative time rather than calendar time), and (3) data analysis utilities available. Batch historians also store data for longer periods of time (e.g., years).
- Expert systems (sometimes called rule-based systems) are a branch of artificial intelligence. They are especially proficient at managing large numbers of user-defined "if-then-else" rules and other types of linguistic paradigms, all contained in a "knowledge base" that represents how process experts (operators, engineers, scientists, etc.) think about and describe their operations. The "inference engine" portion of the operating system takes care of appropriately linking the information in the knowledge base together. Therefore, users do not have to worry about precisely locating "knowledge" information within a computer program such as they would when using a procedural language such as Fortan or C++ [9].
- Commercial alarm loggers are sometimes interfaced to automation systems to provide expanded capability regarding alarm data recording and analysis. For example, alarm logger functionality usually includes Pareto charting showing the most frequently occurring alarms.
- LIMS is typically the repository where final laboratory assay results are stored, generated from individual microprocessor-controlled lab instruments (e.g., HPLC). The LIMS system may then generate reports, provide means of supervisor approval and release of lab results to users, and communicate results to other computer systems. Laboratory SOPs, assay protocols, validation documents, and other laboratory supporting information may also be stored in LIMS.

## 4 APPLICATIONS/BENEFITS OF AUTOMATION

### 4.1 General Uses

In general, operations that are well defined, electronically measurable, and frequently conducted are good candidates for automation. Automation enables such operations to occur accurately, consistently, on-time, and with minimal manual labor.

Consider the historical use of manual operations such as:

- reading and recording process measurements (e.g., in manufacturing tickets) at periodic intervals,
- changing control loop (and/or alarm) setpoints,
- calculations,
- collection and preparation of process samples for analysis,
- opening and closing valves.

Such manual operations were subject to operator errors, operator-to-operator variability, and the possibility of time-sensitive tasks not being done on time. Automation can usually perform the same operations, avoiding the limitations of manual operations, as well as conduct operations at higher frequency. For example, it is relatively easy for most process control systems to monitor standard field measurements (e.g., temperature, pressure, flow rates, pH) every second. This is desirable for continuous control loop operations and for the immediate detection of process excursions.

Automation can also do much more. Value-adding examples include:

- continuous real-time monitoring, display, and recording of most process and utility measurements,
- starting up and shutting down process unit operations,
- verifying material makeup additions (e.g., via reading bar codes, NIR),
- conducting sterilization and "clean-in-place" operations,
- controlling plant utility operations (e.g., WFI, clean steam, compressed air),
- sequencing from step to step for batch operations,
- assisting with model development (searching for correlations, fitting model parameters),
- implementing on-line virtual sensors (a model application),
- implementing advanced control (e.g., feed forward, cascade, ratio, adaptive, multi-variable, model predictive),
- calculating compound concentrations based on information from on-line, at-line, or laboratory analytical systems,
- generating batch reports,
- implementing intelligent remote alarming,
- implementing and graphing the results of advanced data analysis (e.g., principal component analysis, partial least squares, statistical process control),
- prompting when preventive maintenance (PM) activities are due,
- checking final product formulations,
- verifying amount of product in vials, cartridges, and bottles,
- printing and/or putting labels on bottles,
- sending data, reports, and other information to other plant computers,
- creating a plant's operations schedule.

## 4.2 Process Example: Bioreactors

The raw data measured by process mass spectrometers (i.e., oxygen, carbon dioxide, nitrogen, and alcohol analysis of bioreactor exhaust gas) enables the subsequent near-real-time calculations of culture oxygen uptake, carbon dioxide evolution, respiration quotient, oxygen mass transfer coefficient, and other derived variables (e.g., cell mass, substrate concentration) [14]. These calculated parameters greatly increase the real-time information available to scientists and technical service personnel regarding a culture and the bioreactor, contributing to improved control, reduced variability, and real-time process troubleshooting (e.g., early indication of contamination). Such information is also helpful in determining the equivalency of bioprocesses run at different scales.

## 4.3 Process Example: Chromatography Separation Columns

A high performance liquid chromatography (HPLC) analytical system can be interfaced to a production scale chromatography column that provides separation of desired product (i.e., eluate fraction containing the desired product) from undesired compounds (i.e., undesired eluate fraction). The objective is to monitor the eluate output of the chromatography column during the phase when the desired product is desorbing (unbinding) from column resin and eluting from the column [15] Fig. 8. The HPLC system can be used to continuously monitor the output of an HPLC detector; compute corrected compound concentration values from raw data, baseline, and calibration information; and then make decisions as to the appropriate time to direct electronic signals to valves in the field to separate column eluate fractions. Because of the time needed for the system to complete a chemical analysis (typically several minutes), multiplexing a HPLC to multiple pieces of equipment is uncommon; a dedicated instrument per chromatography column is typical.

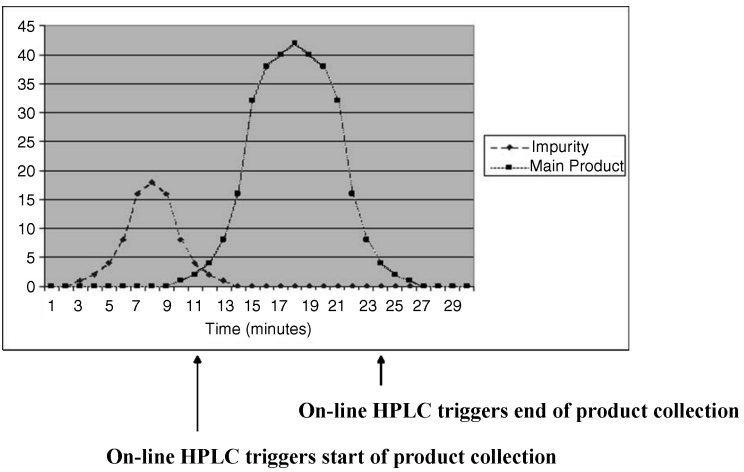


***FIGURE 8*** *Use of on-line HPLC to automate collection of eluate fractions from a chromatography column.*

## 4.4 Process Examples: Powder Blending and Drying

Near infrared (IR) instrumentation has been increasingly used in pharmaceutical manufacturing applications such as in determining the end point for powder blending and drying operations.

Powder blending is a key step in, e.g., solid dosage form preparation. One problem with historical manual operations is that it often brings an operator in contact with the mixture and any associated hazards. Another issue is that of most manual operations, i.e., sampling errors and the time required for sample preparation and analysis. Pharmaceutical company Pfizer [16] has reported on their use of NIR in monitoring the progress of and determining the end point of powder blending and has indicated the following benefits:

- No operator contact with mixture
- No sampling errors
- Real-time information
- Multi-ingredient uniformity
- Improved process understanding
- Process fingerprinting, useful for scale-up
- Getting it right the first time
- Fast release of blend to next step, reduced cycle time

NIR (or IR) can also help in confirming "identity" of chemical compounds in a manufacturing process due to the absorbance by compounds of directed radiation energy at specific wavelengths.

Regarding drying operations, the absorbance of light by water or other solvents at known NIR wavelengths is utilized. As with blending, such applications are near continuous, near real time, and can involve non-invasive measurement and measurements involving no moving parts.

## 4.5 Process Example: Final Product Check Weighing

A common need in the pharmaceutical industry is to "check weigh" product in vials/cartridges such that all samples are weighed in a high throughput, non-contact, high accuracy inspection. Historically, check weighing has often been accomplished by removing samples from production lines and manually weighing them on precision balances, both before and after filling. This has been time consuming and also resulted in inspection rates of only a few percent.

With nuclear magnetic resonance (NMR), the technique only measures materials with a magnetic moment, which includes, e.g., compounds containing $^{1}H$ and $^{13}C$. Most pharmaceutical liquid (and powder) compounds have a measurable amount of, e.g., $^{1}H$, and can therefore be monitored with NMR (i.e., the NMR signal is proportional to the number of $^{1}H$ atoms in the product being weighed and, therefore, its mass). The glass vial or cartridge contributes no signal, thereby enabling a single-step, tare-less mass measurement of drug product. This technique has been commercialized (BOC

Edwards) and used on a pharmaceutical filling line with a throughput of up to 400 vials per minute, and performs fill level verification on every vial.

## 4.6 Process Example: steam sterilization

[Note: sterilization can also be accomplished in other ways, e.g., gamma radiation, ethylene oxide (EtO), hydrogen peroxide ($H_2O_2$).]

Consider the challenge in steam sterilizing a bioreactor. The objectives are to:

- kill any microorganisms, including spores, that may be present in the bioreactor and its directly connected piping,
- achieve a consistent broth nutrient concentration at the end of the sterilization phase.

Steam sterilization operations are pursued manually in some plants, using a procedure such as introducing steam to appropriate portions of the bioreactor system, taking the system up to 121°C for 20 minutes, and then cooling the system down. In practice, such manual procedures can result in significant process variability. For example, an operator may be busy doing other things when it is time to start the cool down. Also it may take a long time to achieve 121°C in a plant's bioreactors, especially for very large vessels or when trying to sterilize multiple vessels at the same time. This can extend the time that the vessel is at high temperatures when nutrients may be reacting, changing the concentration mix of the broth inside the vessel. For example, glucose is a common nutrient (i.e., the principal carbon source) in many bioprocesses. It is known that at sterilization temperatures, glucose:

- can form polymers,
- reacts with inorganic phosphates,
- can react with certain amino acids to form Browning reaction products.

Therefore, the post-sterilization concentration of glucose, minerals, and other nutrients can vary significantly as a result of variation in the total heat introduced to the bioreactor during sterilization.

One part of an automated sterilization strategy can be to calculate $F_o$, a well-known algorithm for determining the end point of sterilization. $F_o$ is easily coded into an automation system, i.e., [26]:

$$\begin{aligned} F_o &= \text{log reductions of viable organisms} \\ &= \Delta T * \sum 10^{((\text{Temp.}-121)/Z)} \end{aligned}$$

where $T$ = calculation iteration time interval (min)

Temp. = bioreactor temperature (°C)

$Z$ = the temperature change (°C) required for a survivor decrease of 1 log unit, e.g., 10°C

When $F_o$ reaches a predetermined value (e.g., 12), this means that 12 log reductions in viable organisms have theoretically occurred. The idea is to choose a $F_o$ target such that less than one chance in a million exists that a viable microorganism remains in the bioreactor (which could grow, multiply, and contaminate the bioprocess). Twelve is a typical minimum value of $F_o$ in industry. Most plants use a higher number to provide additional assurance of a contamination free outcome. The $F_o$ algorithm allows for sterilization to take place regardless of whether the temperature reaches 121°C. If the temperature stays below 121°C, the sterilization will take longer; if the temperature exceeds 121°C, the sterilization cycle will be shorter.

## 4.7 Process Example: Drug Product Coating (E.g., Coating Pans, Fluidized Beds)

A "drug product" example, which combines the principles of QbD and PAT, is the addition of coating layers (e.g., product, enteric, color) to substrates, as can occur in coating pan and fluidized bed operations. Scientists have determined that the "quality" of a film coating is highly correlated to the drying rate – which is then suggested as a critical quality attribute; i.e., processes that result in the same drying rate will also produce the same film coating quality.

A calculated parameter, which can be determined on-line, known as EEF (environmental equivalency factor), has subsequently been developed [27, 28]. EEF is a dimensionless number that is proportional to the ratio of the dry area of a substrate being coated to the wetted area and, as such, is indicative of the drying rate of the film being applied. Therefore, coating conditions that yield identical EEFs will result in equivalent film coatings. Parameters affecting drying rate, needed for the calculation of EEF, can be continuously measured on-line (i.e., inlet airflow rate, inlet airflow temperature, dew point, liquid spray rate, and outlet air temperature). The availability of these measurements then leads to an appropriate control strategy, whereby the above listed measurements are continuously monitored, EEFs then calculated on-line, and then liquid spray rate, inlet airflow rate, inlet airflow temperature, and/or dew point adjusted in order to keep EEF in a desired target range (typically between 2 and 5). Sometimes, outlet air temperature is also included in the control strategy as a constraint; e.g., it may be important to avoid exceeding a particular temperature due to coating stability or product physical property considerations.

## 4.8 Process Example: Agglomeration

Agglomeration is one of several methods of producing granules. It typically involves the combining of, e.g., API powder, binder(s) and diluents, and equipment such as mixers, dryers, and machinery for physical transfer of material. There are several key physical attributes of agglomerates that are essential for product performance, such as granule size, size distribution, density, flowability, mechanical integrity, compressibility, and dispersion. While it is not always possible to physically measure the key physical attributes in a way that is convenient for on-line control, good progress has

been made in monitoring and controlling important aspects of such processes. For example, it is known that the growth in the size of many commercially important granules is strongly influenced by the addition rate of liquid binder.

Regarding process control, agglomerate size distribution can be measured in-line at any number of points in the process, using a commercially available high speed camera and appropriate lighting. Image analysis (using direct camera imaging, light diffraction, or reflectance method) is used to obtain measures of particle size from the camera images. This information can then be used to manipulate the level of binder dilution, which, in turn, changes granule size and/or granule size distribution [29].

## 4.9 Utilities Example: Water-For-Injection

Automation can play a vital role in various manufacturing supporting processes such as in monitoring and control of process utilities [e.g., sterile air, water-for-injection (WFI), clean steam]. For example, WFI must meet strict requirements for purity. The system (see Fig. 9) that produces and distributes the WFI must be able to maintain the required level of purity. This is frequently accomplished with the help of automation by sanitizing the system at high temperatures, followed by maintaining continuous flow at elevated temperature through the distribution system. Sensors normally included in such automated systems include temperature (RTDs or thermocouples), flow rate, and, sometimes, conductivity and pH [11].

## 4.10 Example: Product Storage

Opportunities for automation do not end when a drug product is packaged. For example, a finished product may have a temperature CPP – such that the temperature environment (often monitored via automation) must stay within a proven acceptable range (PAR) in order to ensure that product stability is maintained and published "shelf life" and/or the "expiration date" remains valid. This temperature range must not only be adhered to during product filling, finishing, and packaging operations, but also during storage and transportation activities. Techniques include monitoring and recording temperatures in warehouses as well as recording temperature history in RFID tags.

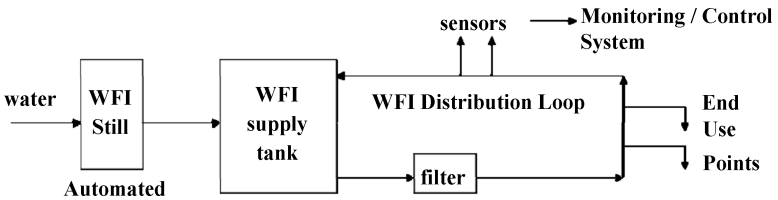


***FIGURE 9*** *Generic schematic of a plant WFI system.*

## Examples: Automated Data Analysis

***4.11.1 Current vs. Historical Trend Plots*** A key objective of on-line monitoring of a process is to determine if the process run is operating normally or if the run is abnormal. This can serve as, e.g., a warning activity that the process may be drifting toward a "deviation" condition.

While typically configured alarms, based on a comparison of current values vs. setpoints, may accomplish this for some process operations, it is insufficient for others. For example, most of the parameters normally controlled to a setpoint in a fermentation (air sparge rate, agitation rate, broth temperature, back pressure, and pH) indicate very little about the health of the culture that is growing and producing product inside the vessel. However, other time-varying information is often available that can be used as indicators of culture health (i.e., mass and viability).

One way of determining if a bioprocess is running normally is by calling up individual trend plots of the running process and comparing the current trends to those associated with normal historical runs. This is time consuming to do manually for those unit operations with a high number of process parameters, such as a bioreactor (process parameters $>20$), and also challenging for batch processes, where the expected value of some process parameters (e.g., culture oxygen uptake) is a function of time. Further challenge occurs due to the natural variability (COV $>10\%$) of some bioprocesses. A way to assist operators and technical service personnel in this situation is to automate the presentation of current and historical trend results. An example is shown in Fig. 10.

In this trend plot, the current value might be a fermentation culture oxygen uptake, carbon dioxide evolution, or perhaps the total amount of sodium hydroxide added in order to maintain pH control. The deviation plots represent the result of having averaged several satisfactory historical runs and computing the time varying average $\pm 2$ std. deviations. The $\pm 2$ std. deviation calculated plots are then overlaid in the HMI on the current trend plot. Therefore, operators and technical service personnel can quickly see if a process parameter is outside the range determined from successful

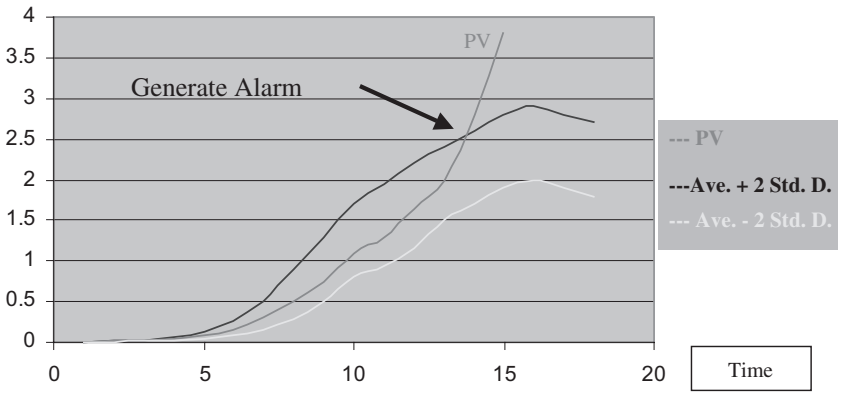


***FIGURE 10*** *Fermentation process variable (PV) vs. historical range of values.*

historical lots. If desired, an automated alarm can be configured to occur when this situation occurs. In Fig. 10, the plot displayed (at about 13 hours) is typical of a bioprocess becoming contaminated with bacteria, so a culture respiration-related PV exceeding the historical average PV plus 2 std. deviation is an excellent early warning sign of process contamination. Note: the basis of this contamination application is that many contaminating microorganisms, such as several common bacteria, grow and multiply much faster than the desired pure culture (e.g., mammalian cells, streptomycetes, or fungal microorganisms). Therefore, once introduced into the pure culture, it is only a matter of time until the contaminating microorganisms have sufficient mass to change the normal pure culture plots showing the uptake of oxygen and other nutrients.

***4.11.2 Principal Component Analysis (PCA)*** In this method [23], the large number of process measurements are reduced to a much smaller number of mutually uncorrelated variables, called principal components (PCs). The idea is that it may be easier to draw conclusions about the progress of a production run by analyzing two or three independent variables (even though PC variables are not directly observable independent measurements with well-known labels that operators/users can understand) than to call up and analyze 15–20 process trend plots.

A nominal set of PCs can be determined by analyzing historical successful runs. These PCs (or derived statistics) can then be compared to a real-time operation to determine if the run is progressing normally or not.

A related use of PCA is to separately determine the time-varying PCs for historically good batch lots vs. historically poor batch lots. Assuming good pattern differentiation exists between the good vs. poor batch run PCs when plotting the PCs against one another, users can pursue subsequent analysis, iteratively deleting parameters via trial and error from the data sets used to compute PCs, to sometimes obtain meaningful insights as to the cause of poor batch runs.

***4.11.3 Partial Least Squares (PLS)*** PLS is a tool suitable when plant variables can be partitioned into cause-and-effect values [23]. The method may be used for regression or, similarly to PCA, with resulting reduction of the effective dimensionality of data. Regarding regression, PLS has been used as a prediction tool to estimate the end concentration of product in a batch. In this application, the PLS model acts as a classification system and indicates if the current batch is more consistent with high or low yield batches. If used as a classifier, it is important that the model be developed using historical data from both high and low yield batches.

## 4.12 Automation in Research/Development

Manufacturing is not the only venue for automation. Pharmaceutical companies also employ automation in many of their research and development operations. Examples include:

- automated retrieval of cultures from their storage domain – then testing them for the existence of compounds that may be active against, e.g., a pathogen,
- automated filling of shake flasks with media + inoculation with different strains of a culture, followed by incubation on automated shakers in controlled temperature and humidity rooms – for strain evaluation studies,
- automated filling of shake flasks with media of different compositions, for media optimization studies,
- automated programming of HPLC analytical systems to try a variety of different methods (e.g., different temperature programs) to maximize the separation of compounds, used in analytical method development,
- operation of bench-top bioreactors in scientist laboratories,
- pursuit of informatics in, e.g., new drug design, analyzing genetic codes.

## 4.13 Automation's Relevance to Factory Physics

***4.13.1 Plant Capacity (Throughput)*** A major benefit of using automation systems is in more timely decision making. This, of course, is one of the main objectives of PAT. Putting more instrumentation on processing equipment, especially instruments monitoring CPPs, and moving traditional laboratory analytical systems to an on-line or at-line configuration on the plant floor can provide more immediate information to technical service personnel. This can shorten the time a lot is "on hold" before being forward processed, which, in turn, shortens overall cycle time. Shorter cycle times, in turn, increases a plant's capacity (i.e., throughput) in accordance with Little's law:

$$\text{Throughput} = \frac{\text{Work in progress (WIP)}}{\text{Cycle time}}$$

***4.13.2 Process Variability*** Low process variability is highly desired in manufacturing, as it can result in a more consistent product, reduces frequency of queues, reduces average cycle time, reduces the need to change the production schedule, and reduces the number of production runs needed to evaluate proposed improvements to a process (i.e., via plant trials) or to see a significant unplanned change in plant operation (via monitoring statistical control and/or CUSUM charts). The relationship between variability and task time, queue time, and resource utilization is captured in Kingman's equation [31]:

$$\text{Average Queue Time} = \text{Average Task Time}*((\text{Ca}+\text{Ct})/2)*((\text{Ur}/(1-\text{Ur}))$$

where $C_a$ = coefficient of variation in the arrival rate of things to be worked on
$C_t$ = coefficient of variation in the time required to complete a work task activity
$U_r$ = fraction of time that required resources are utilized

An examination of Kingman's equation shows that increases in the variability of incoming project work and/or in executing tasks can dramatically increase average queue (idle wait time) associated with an activity, especially when resources (i.e., equipment and/or people) are utilized a high fraction of their available time.

## 4.14 Automation's Relevance to Process Deviations

The addition of automation to a plant generally implies that more process parameters will be monitored and controlled, which, if implemented well, should result in a more consistent batch-to-batch process. The resulting reduced process variability, combined with reduced operator errors due to fewer manual operations, can be expected to reduce the frequency of process deviations.

Use of automation can also provide earlier alerts to abnormal situations (i.e., warning alarms), providing the opportunity to reduce actual deviations, which, in turn, require precious time and resources to investigate. One commercial fermentation plant has reported that their use of real-time expert systems, used primarily for intelligent remote alarming and alarm response guidance for operators, resulted in more than a 10% reduction in process variability – and also a 4% yield increase. This was due primarily to earlier detection and better troubleshooting of poorly performing fermentations [10].

Automation can also shorten the time and reduce manual resources needed to prepare batch reports and summarize data needed in investigating process deviations. For example, an automation system is capable of reviewing collected electronic data in a historian at the end of a batch run and generating a report of all instances in which a CPP exceeded PAR values. These represent deviations (a key subset of all deviations) that must be investigated. Given key words associated with a deviation, an automation system can also do a search of all other deviations previously recorded in a deviation database to look for similarities. That is, most deviations are not unique but similar to one or more incidents that have occurred previously. Locating a similar incident previously investigated and analyzed can avoid duplicating past work when a similar incident occurs in the future. This kind of automation system functionality, then, can reduce manual labor and speed the decision regarding forward processing and/or rejection of lots.

## 4.15 Other Automation Perspectives

Automation encompasses many other applications and benefits (in addition to those affecting product quality), such as in cost reduction, in increasing yield, and in process optimization. These will not be further discussed here, due to this chapter's focus on product quality.

Finally, automation does not need to be perceived as a way to reduce operations staff. Rather, it can be perceived as a way to change the role of operators, i.e., to reduce their use for routine repetitive tasks and increase their time available for more intelligent uses, such as helping in analyzing data, process troubleshooting, and working on process improvements.

# 5 VALIDATING AUTOMATION SYSTEMS

## 5.1 General Principles

If automation systems are used in the monitoring, control, support, and/or data recording of pharmaceutical manufacturing processes, particularly when CPPs are involved, then such systems must usually be validated.

Validation means that a process or system has documented evidence that it does what it purports to do, has done so in the past, and can be expected to continue doing so in the future.

Examining this definition in more detail,

- What a system purports to do implies the existence of a Functional Requirements (i.e., System Requirements Specification) document. Some high level requirements and work process expectations may exist in other documents, e.g., a plant automation philosophy or validation master plan.
- A system operating correctly in the past implies the existence of a validated system. If an older system was not formally validated when initially installed, a documented retrospective validation is expected. At a minimum, historical maintenance, change control, and periodic review records of the system should have been maintained, as well as training records of operators and system support personnel. Batch records of previous lots should also have been maintained, often for several years per regulatory requirements.
- A system that can be expected to continue operating correctly in the future implies the existence of a preventive (and/or predictive) maintenance program, formal change control, records verifying that operators and system support personnel are qualified to do their job, business continuity (i.e., disaster recovery) plan, and periodic system reviews.

Computer system validation is not a guarantee of successful operation. Rather, it represents the formal implementation and practice of computer system management best practices that, in turn, maximize the probability that a system will operate correctly.

A life cycle diagram for new automation systems is shown in Fig. 11.

The following comments expand on the activities in an automation system's life cycle from a validation perspective:

## 5.2 Validation Master Plan

A validation master plan (or equivalent) is expected that describes:

- required work processes (e.g., change control), including group and individual responsibilities,
- testing (e.g., IQ, OQ, PQ) expectations,
- documentation expectations,

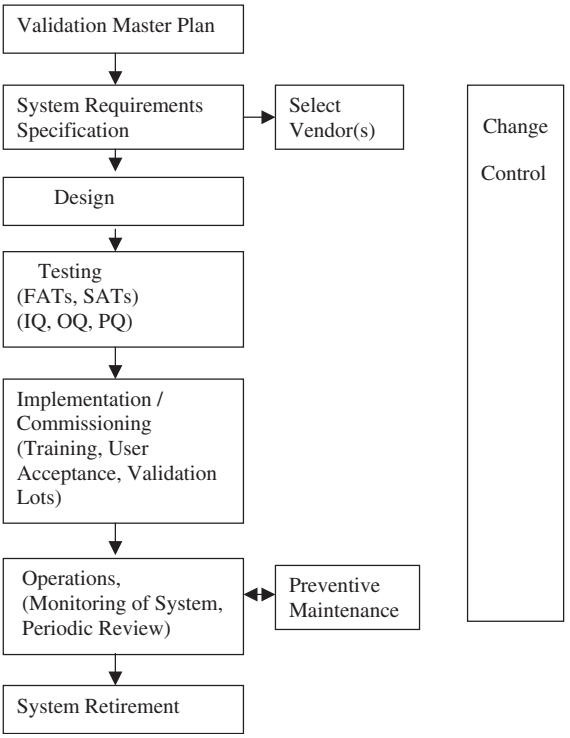


***FIGURE 11*** *Life cycle of a new automation system.*

- corporate requirements (e.g., security, ER/ES compliance, adherence to applicable regulatory safety and environmental regulations, adherence to ISO and technical standards),
- other company-mandated validation expectations.

## 5.3 System Requirements Specification

The System Requirements Specification (sometimes called Functional Requirements) is typically generated near the beginning of a new automation system project or major retrofit to an existing system. The requirements specification describes, in detail, WHAT it is that the system is expected to do. Requirements are usually a combination of user requirements, system support, and plant infrastructure requirements.

The requirements specification can be used for several purposes, including:

- obtaining quotes from vendors,
- evaluating vendor system capabilities/vendor selection,
- source document for system design,
- system testing.

While generating the System Requirements Specification is sometimes a one-time activity, the requirements specification is theoretically a living document up to the implementation phase of the life cycle. Sometimes changes are made to system requirements due to:

- technical issues that surface during design,
- project scope creep,
- vendor product limitations,
- project time and budget constraints.

Keeping the requirements specification up to date is important since it is a key source document in driving system design and in prompting specific system tests.

## 5.4 Design

After the requirements specification document has been approved, the automation system vendor can be selected (unless already known from other considerations, e.g., system replication) and system design can begin. The design establishes the "HOWS" in meeting the system requirement "WHATS." Many documents are typically associated with design, including e.g.,:

- plant electrical power/grounding drawings,
- layout of the automation system,
- piping and instrument drawings (P&IDs),
- operator console information displays (HMI).

For large projects, design documentation is often structured as high level design documents, with each driving one or more detailed design documents.

Frequently a design qualification exercise is pursued near the end of the design phase to verify that the system design will meet requirements.

An important aspect of system design is to consider potential system failure modes. This is often pursued using a Failure Modes and Effects Analysis (FMEA). Design modifications (e.g., addition of an uninterruptible power supply) can then be considered to help reduce the chance of failures occurring or mitigate the consequences of failures should they occur.

## 5.5 Testing

Significant testing occurs during system development – prior to system implementation/commissioning. For example, during development, reusable software modules might be developed or statistical algorithms coded that can be individually tested via the creation of software "scripts." Late-phase testing that occurs at the vendor factory, before system delivery to the customer, is called Factory Acceptance Testing (FAT). Testing that occurs after delivery is called Site Acceptance Testing (SAT).

A major objective in testing is to prove that Functional Requirements have been met. A common tool used to document this linkage is known as a traceability matrix, which links specific requirements to design features that, in turn, are linked to specific tests. Using this tool, one can:

- verify that all requirements have been tested,
- identify the specific test(s) run that verifies that a certain requirement has been met. If a test fails, the matrix can help identify the design feature to be investigated.

***5.5.1 General Testing*** Formal testing in pharmaceutical manufacturing and clinical trial pilot plants is usually organized into sequential phases known as Installation Qualification (IQ), Operational Qualification (OQ), and Performance Qualification (PQ). IQ is usually specific to hardware and ensures that components of the automation system have been received, are consistent with what was ordered (based on system design), and are installed in accordance with the manufacturer's recommendations and recognized best practices. OQ verifies that properly installed system components operate correctly. For example, OQ might verify that a temperature control loop properly controls a tank temperature to a specified setpoint (which, in turn, requires that all hardware and software components of the loop are properly configured and work together properly). PQ tests that the overall integrated system works correctly, but usually stopping short of making actual product. Water batching (running of the process using aqueous solutions and perhaps other inexpensive raw materials) is often pursued during PQ.

***5.5.2 Worst-Case Testing*** A special type of testing, especially applicable in validating automation systems, is known as worst-case testing. The idea is to ensure that tests are run at the limits of expected (i.e., realistic) operational boundaries. Failures are more likely to occur at such boundaries vs. nominal operating conditions. Examples of worst-case testing include:

- testing the ability to interact with a database (e.g., historian) while several users are simultaneously accessing the same database,
- testing applications requiring precise timing operations (e.g., totalizations) while the system is fully loaded with running applications,
- testing the accurate receipt by the computer system of low voltage/low current measurements in the presence walkie-talkies operating near the sensors. This will check that the system is immune to extraneous RFI.

***5.5.3 Pre-implementation Change Control*** During project phases prior to system implementation, a pre-implementation version of change control is appropriate. Change control is needed during these phases since projects rarely run exactly according to the original project management plan. Also, testing activities

invariably result in the identification of defects that need to be corrected. Descriptions and challenges of managing projects in general, and change control in particular, are included in Ref. 30. Pre-implementation change control is usually a less bureaucratic and more streamlined process than is post-implementation change control.

## 5.6 Implementation/Commissioning

Once successful tests are completed, the implementation (sometimes called commissioning) phase begins [32]. This phase typically includes operator training, completing any remaining project minor bucket list items, and user acceptance of the system. cGMPs require that operators be qualified to do their jobs and that training supporting operator qualification be documented. Training activities often include vendor courses, in-house classes, and/or on-the-job training.

Implementation/commissioning may also include the execution of validation lots (e.g., if the mission of the system is to manufacture a new product). Validation lots represent the running of the process multiple times (usually three or more batch lots) in which actual product is manufactured. The results of validation lots are normally included, e.g., in the New Drug Application (NDA) that a company submits to the FDA in requesting approval to manufacture a new medicine. The end of implementation is a key benchmark, marking the point in time where the project team turns ownership of the automation system over to plant operations or their automation/IT support group.

## 5.7 Operations

During plant operations, several activities occur that help ensure that the automation system stays in a state of validation. These include:

- continuous monitoring by users and support staff,
- preventive/predictive maintenance of hardware components,
- system periodic reviews,
- formal "post-implementation" change control,
- development of a contingency plan (i.e., disaster recovery plan).

Monitoring can help identify concerns such as gaps in data records, existence of nuisance alarms, and degradation of the system. System degradation could be due to, e.g., the accumulation of changes, weakening electrical components from voltage spikes, and overheating incidents in areas containing system hardware.

A major focus of system validation is formal change control. Systems can fall out of a state of validation due to a single change or to the accumulation of many changes over time. Root causes can include insufficient testing in changing application software, blind acceptance of vendor upgrades, increased loading on the system, introducing sources of RFI and/or EMI, and authorizing unqualified personnel to operate or change the system.

Therefore, changes to an automation system must be formally managed. Most companies use a SOP to govern changes, which includes a structured form that prompts information as to what the change is, why it is needed, testing performed, names of people involved with making, approving, and accepting the change, date of change, and the new version identifier (if the change is to software). Therefore, two versions of change control often exist, a formal version used after system implementation, and a less formal version used prior to implementation.

A highly desirable "change control" feature of a computer system, if available from a vendor, is automated audit rails. Audit trail functionality triggers automatic records to a database when certain manual changes (e.g., to a control loop setpoint or alarm limit) are made to the application software of an operating computer system. Such records may include what the change was and when it was made. If possible, the name, initial, or other identifier of the person making the change should be included in the record. Such records can be useful in investigating deviations and/or in understanding any upset that may have occurred to the process. Audit trail functionality mimics what would be required manually if an automation system did not exist; i.e., an operator making a control loop setpoint change would manually record this action in the manufacturing ticket or, alternatively, initial and date the line item if it were already an instruction in the ticket. Thus, audit trail records provide formal documentation of software changes that, at a plant's discretion, can become part of official batch records.

A contingency plan, also known as a disaster recovery plan or a business continuity plan, is expected to exist for an automation system if, e.g., loss of the system could potentially impact the supply of product to the market possibly, impact product quality if the process had to be run manually, or have other significant business impact to a company. Process control systems can potentially take a long time to replace should a contingency (e.g., fire) occur due to:

- the large amount of wiring between the computer and field elements,
- integration of several different technologies and devices,
- time to complete relevant verification and validation activities.

A typical contingency plan:

- describes a particular system (including a list of major components),
- describes the impact to the company and/or customers should the system be lost,
- defines procedures to minimize the chances or consequences of a contingency (e.g., use of appropriate physical and software security, remotely stored backup copy of software, maintaining up to date wiring diagrams, obtaining agreements with vendors regarding expedited replacement of hardware),
- describes the action plan, including personnel responsibilities, to be implemented in the event of a contingency.

## 5.8 System Retirement

The final activity in the life cycle diagram is system retirement. This is not an inconsequential activity, as key logic may exist in utility and application software that may be needed in porting to a new system. Also, regulatory retention requirements may exist for portions of the system and its output, such as process recipes and historical lot data. Therefore, a significant overlapping time line may exist when migrating from one system to another. Even if a manufacturing process is to be terminated, much process data and other information must typically be retained and be recoverable, often for several years. This means that:

- sub-systems remain available to read, display, and/or print historical information that is stored on whatever electronic storage devices a plant has used, or
- historical data are migrated to new databases (possibly requiring reformatting) or to other repositories, with the migration process being formally validated.

## 5.9 Other Computer System Validation Topics

***5.9.1 Electronic Records/Electronic Signatures*** An important cGMP (21 CFR Part 11), which has evolved in recent years, pertains to electronic records and electronic signatures. The need for this cGMP was based on the paradigm shift from the historical use of hard copy manufacturing tickets, manual collection of data, and manual signatures on documents to today's environment where much of the information is in electronic form. For example, manufacturing tickets are now frequently implemented in electronic form. Also, significant manufacturing instruction content previously contained in manufacturing tickets, now exist in other electronic documents such as process control recipes. Further, data are collected and stored electronically and automatically, and signatures can be obtained electronically. It is important that electronic forms of data and manufacturing ticket information, and any changes to these documents and data, adhere to similar or equivalent controls that have historically existed for paper/hard copy versions of this information.

Reference 33 contains more information on this topic.

***5.9.2 Vendor Assessments/Audits*** Potential vendors supplying computer systems for use in making pharmaceuticals are usually assessed/audited by the pharmaceutical company customer to make sure they are likely to be in business for the long term and that they utilize recognized system development and support best practices (e.g., formal change control, adequate technical support, use of alpha and beta testing, remote storage of current and historical versions of software). The results of vendor assessments/audits can be used to:

- assist with automation system vendor selection,
- encourage a vendor to consider work process or system functionality changes to help the customer meet regulatory/validation expectations.

***5.9.3 Security*** Security begins with plant security, allowing only authorized (or chaperoned) personnel access to a manufacturing plant, computer equipment rooms, and/or plant operations control rooms. Security also applies to accessing running process control software and to data records. Different levels of access should exist, depending on what information an individual is qualified and authorized to monitor and/or change.

If cipher locks or passwords are used to help with security, it is important that they do not use easily guessed alphanumeric character sequences and that they be periodically changed. Details regarding security expectations are typically identified in a Standard Operating Procedure.

Security considerations may contribute to an automation system's design, such as the common practice of implementing a "local area network," independent of a company's intranet system, for the process control portion of a plant automation system.

***5.9.4 Source Code Review*** It is expected that custom source code, such as unique software utilities and applications, developed by a company, its suppliers, and/or system integrators for validated computer systems, be reviewed by a qualified second person.

The objectives of source code reviews include checking that:

- code will meet its intended purpose,
- code meets requirements as specified in a plant's software development SOP,
- code conforms to programming best practices (e.g., absence of dead code, liberal commenting of code),
- code is sufficiently well organized and documented that a person other than the original programmer could be expected to understand and support the code in the future.

Source code reviews sometimes include on-line tests as well as off-line reviews. For example, an on-line test might include running a script to check that execution time is reasonable. This, in turn, would verify the absence of infinite loops and/or confirm fast convergence of iterative calculations. An on-line test could also confirm that a calculated result agrees with that determined manually.

It is not expected that vendor-supplied operating system and utility source code be reviewed by a customer. This would be impractical to do, given that such software is typically sophisticated and proprietary.

## 6 CHALLENGES IN USING AUTOMATION

While use of automation can be cost effective and provide many benefits to a company, it is not without its challenges and potential pitfalls. Areas of concern may include:

## 6.1 Inappropriate applications

Microsoft's founder, Bill Gates, has stated: "The first rule of any technology used in a business is that automation applied to an efficient operation will magnify the efficiency. The second is that automation applied to an inefficient operation will magnify the inefficiency." This suggests that automation should not be viewed as a "cure" for a poorly run and inadequately understood process. As noted earlier, automation is best applied to repetitive operations which are well understood and for which on-line (or near real-time) measurements exist.

Remember, if you can't measure it, you can't manage it.

## 6.2 System Degradation and Failures

Automation systems, while generally reliable, can degrade or fail for a variety of reasons, including:

- plant utility failures (e.g., electrical power, pneumatic),
- improper grounding,
- lack of redundancy for critical components, including data highways,
- software bugs (operating system, utilities, and applications),
- increased loading (e.g., number of I/O points), which results in the system's inability to read inputs, process data, communicate over networks, execute control algorithms, and/or update operator consoles within desired/required time frames,
- insufficient adherence to validation best practices – including testing, training, security (i.e., authorized access to hardware and software), business continuity planning, and change control,
- insufficient prospective evaluation of failure modes (e.g., via FMEA),
- incompatibilities in interfacing computers together from multiple vendors,
- contingencies (e.g., fire, flood, lightning, tornado, sabotage).

## 6.3 Invalid Data/Parameter Values

Various field operating phenomena and/or nuances of automation design can result in values of software parameters or collected data that are not appropriate for their intended use. Examples include the following:

- RFI/EMI
  The existence of radio frequency interferences (RFI, e.g., use of walkie-talkies) and/or electromagnetic interferences (EMI, e.g., close proximity to large induction motors), can induce voltages/currents in improperly protected nearby signal wiring, thereby changing the apparent value of a process measurement. This effect is sometimes observed, e.g., when analog measurements originate in

low current or voltage (e.g., millivolt) field sensors that are some distance from their associated transmitters and/or analog/digital converter – such that the connecting wiring can act as an antenna, especially if wiring is not protected with metallic shielding that is grounded at only one end. The effect can also be observed when transmitter housings are not made of appropriate materials and/or properly protected from electronic interferences and when signal wiring is installed in parallel too close to power wiring.

- Measurement uncertainty when setting alarm limits
  A temptation exists to set alarm limits at proven acceptable range (PAR) values for CPPs. However, alarm limits may need to be set inside the PAR values to avoid missing deviations due to measurement uncertainty. For example, if a temperature CPP's upper PAR value is 30°C, then a deviation would be missed if the real temperature were slightly above 30°C, but the measurement (sensor + transmitter + computer A/D converter) indicated a value slightly under 30°C, due to measurement uncertainty. The FDA has increased its expectation that measurement uncertainty be considered for CPP operational alarms. The discipline and algorithms regarding the determination of measurement uncertainty are well established as practiced, e.g., in the nuclear industry. Information exists in several International Society of Automation (ISA) publications (e.g., [34]).
- Data filtering
  Filtering of data in automation systems is often pursued to smooth signals (e.g., to reduce high frequency noise) for use by control algorithms or in trend plots. Filtering has the effect of reducing the displayed magnitude of measurement perturbations (i.e., spikes) and lengthening their perceived duration. Therefore, as illustrated in Fig. 12, a real measurement spike, shown as a solid line trend, which exceeds a PAR value, might not be recorded as such in a data historian due to filtering that occurs as the signal is being processed by, e.g., a smart sensor or the process control system. That is, use of data filtering has the potential of causing missed awareness of real process deviations.

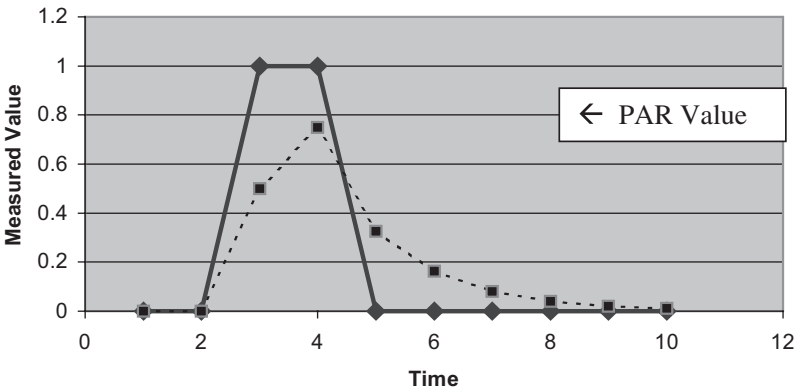


***FIGURE 12*** *Raw data (solid line) vs. filtered data (dashed line) plot of a process parameter.*

- Data compression/decompression
  Automation systems operate in real time, with many systems processing thousands of measurements per second. Historically, disc storage capacity was too limited and expensive to save all the measurements, so data compression algorithms were employed. For example, a temperature sensor might be read by the automation system every second, but perhaps only one or two values per hour were recorded in a historian while the temperature remained under good control. If needed, a data decompression software program could then be used for data analysis/mining applications to create estimates of the temperature at periodic time points. Depending on the compression algorithm used and the choice of compression constants, it is possible for stored compressed data to miss reporting some of the important details of raw signals (measured or calculated parameters) needed to evaluate process deviations. Thus, use of compression algorithms need to be validated, especially when the measurements involve CCPs and/or when the stored data could be used for process deviation investigations. The need for data compression has been reduced in recent years due to large increases in the capacity of inexpensive data storage. However, it is still sometimes utilized due, in part, to the long-term regulatory storage requirement (several years) for critical data.

## 6.4 HMI Design

The primary purpose of a human machine interface (HMI) is to provide operators with well-organized and timely information to help them do their job. This includes providing information and alerts (e.g., schematics, numerical values, plots, lights, horns) to help operators:

- monitor a process,
- easily navigate between computer screens in a logical way,
- become aware of and appropriately respond to abnormal situations requiring a response.

Unless well designed, automation system HMIs can be a frustration for operators. Examples include:

- difficulty in navigating between screens to locate process information for large plants containing thousands of I/O points,
- mixing routine messages into alarm alerts, contributing to “information overload” during crisis situations (note: this has a contributing factor to the severity of some plant disasters),
- generation of multiple alarms for an individual abnormal situation – which can contribute to information overload,
- alarms existing for situations not representing an abnormal situation requiring a response (these are usually referred to as “nuisance alarms”),

- not including the desired operator response (or list of options) to an alarm in the HMI, a problem particularly when a plant has thousands of I/O points,
- not prioritizing alarms for operators, resulting in confusion when multiple alarms exist during crisis situations,
- ineffective use of colors in helping operators interpret presented information/alarms.

Many of the above items refer to alarms. Regarding HMI design, two of the most important guidelines in configuring alarms are [35, 36]:

- an alarm should represent an abnormal situation requiring a response,
- an effective alarm alerts, informs, and guides.

Unless well designed, plant automation systems can sometimes generate alarm floods (i.e., generating alarms faster than operators can be expected to respond).

References 35–36 contain metrics as to the suggested maximum frequency of alarms that operators are typically capable of handling.

Plant automation systems can generate large numbers of nuisance alarms, which clearly frustrates operators. Anecdotal information suggests that some operators end up losing respect for the automation system in such cases and, sometimes, then start ignoring alarm alerts, therefore missing real plant abnormal situations. This lack of attentiveness to alarm alerts then potentially contributes to increased product quality deviations.

Unless the above noted alarm principles and published best practices [35, 36] are adhered to, a manufacturing plant's nuisance alarm problem may become progressively worse with the evolution to smart sensors and valves – which is significantly increasing the amount of information available to automation systems. Unless strict alarm management discipline is adhered to, the temptation is to configure alarms to much of the increased information available with smart field devices.

## 6.5 Meeting User Expectations

A fundamental activity in planning for new automation systems, or major upgrades to existing systems, is the generation of Functional Requirements (sometimes called a requirements specification) for the system. This document is typically a combination of user requirements and corporate, support, and infrastructure requirements. Projected users of the system are typically interviewed in order to document their expectations (i.e., to generate user requirements).

It is important for users to avoid making assumptions that a system will provide needed functionality just because it is from a well-known and respected automation vendor. This is especially true regarding regulated batch processes, such as are common in the pharmaceutical industry, when using systems (e.g., DCSs) that were originally developed targeting continuous steady-state processes (e.g., petrochem-

icals). While some of these automation systems have added more "batch" functionality in recent years, many still have significant limitations in automating non-steady-state batch processes. For example, some commercial products cannot easily (1) configure alarm attributes, such as limits and priority, as a function of process step or phase, or (2) sort, report, and display process data as a function of lot number and/or relative time. Also, many commercial automation systems are not 21 CFR Part 11 (i.e., electronic record/electronic signature) compliant. The point here is that users and support personnel need to list all key requirements when generating FRs and avoid making assumptions. Note that FRs are a primary resource document when designing and testing the system, so the absence of key FRs capturing user expectations could mean the system will be missing key functionality or will not be sufficiently tested.

It is also important for users to look beyond the "given" process control capabilities of an automation system and think about how collected data will be utilized. For example, if it is important to search, sort, and display data and alarm records by batch lot number, then this should be identified as a functional requirement (note: many systems do not include lot number as a defined field in data records).

In addition, linkages and compatibility with other data bases used in the plant should be considered. Perhaps the most common barrier to more productive data mining and analysis in plants is the difficulty and significant manual effort required to prepare data sets for analysis. While challenging enough when data are in a single database, it is especially challenging when data come from different sources where they exist in different formats with different header/tag identifiers. Some of the challenges in readying datasets for useful analysis include:

- converting calendar time stamps to relative time (needed for batch processes),
- adding additional labels within record tags (e.g., lot number, process step) or in generating software dictionaries to provide needed additional data identifiers,
- removing data outliers,
- decompressing compressed data (i.e., interpolate to fill in missing values),
- reformatting data, if needed, especially when combining data from different sources,
- reducing the thousands of data points representing a continuous process measurement to a much smaller number of data records representing key attributes of a trend,
- combining discrete (e.g., product analytical assays) with continuous process data,
- normalizing data.

Data analysis is useful functionality to include in an automation system, as it is a means of extracting "information" and "knowledge" from data sets. The concept is illustrated in Fig. 13.

The good news is that automation systems have done a great job in enabling industrial companies to collect large amounts of data on the processes running in their

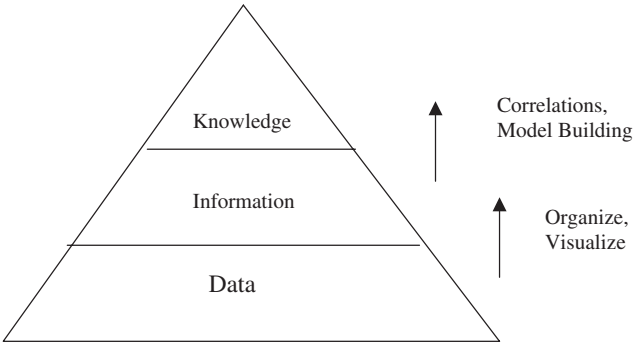


***FIGURE 13*** *Moving from data to information and knowledge.*

plants. Unfortunately, all too often, the wealth of data continues to remain as just data, with little usage other than displaying it in trend graphs. A continuing challenge then is to mine the data for the "nuggets of gold" representing its information and knowledge content. While good statistical, PCA, PLS, 3-D graphics, and model building tools are available for analysis, the effort to prepare data sets for use of such tools often limits their use. Specifying requirements up front in an automation project regarding data to be collected – including attributes, formats, and record tag information – as well as prospectively thinking through what sorting, query, model building, visualization, and analysis tools will be needed, will go a long way toward facilitating efficient analysis and use of collected data once an automation system is implemented.

Companies often face a dilemma when following up to user requirements as some requirements may exceed the capability of standard commercial off-the-shelf (COTS) products, resulting in the prospect of customizing the system and/or integrating in third-party products, often at considerable extra project expense as well as higher ongoing support costs. As a result, compromises in user requirements may sometimes be needed due to time, cost, resourcing, and/or vendor capability constraints.

A second source of user frustration can be a company's "blind acceptance" of vendor software, i.e., not performing adequate testing. Almost all software programs, including vendor operating systems, contain "bugs," both known and unknown. Note that one of the major purposes of periodic vendor software upgrades is to provide customers with fixes to some previously identified "bugs." Also, the number of different software configurations that are possible with commercial software is usually far too great for vendors to test them all, so users need to test the system with the specific configuration and mix of applications they intend to use. Failure to do so not only violates validation expectations but can result in negative operational surprises.

A third source of user frustration can be a plant's proliferation of different vendor computer systems, each with their own HMI, alarm system, database, and data record formats. Operators would certainly prefer to deal with a single HMI and alarm system,

and process engineers/technical support personnel would certainly prefer to have all their manufacturing data (process, lab assay, on-line analytical system) in the same format in one database – which would then greatly facilitate data analysis. Utilizing a DCS for overall plant control, interfaced to (1) any PLCs associated with individual unit operation equipment and (2) any on-line or at-line analytical systems, is one way to work toward the goal of a common HMI, alarm system, and database for users.

Another user expectation that is often a source of frustration is the need for flexible computer systems that can be easily and quickly adapted to new conditions, requirements, and/or measurements, and that can be integrated with other plant computers. Such flexibility is often needed, e.g., in the pursuit of continuous improvement and process optimization initiatives. Significant flexibility can often be realized when using general purpose automation systems (e.g., DCSs) but is traditionally more difficult when dealing with unit operation equipment control systems (e.g., using embedded PLCs) for which the proprietary application software was developed and is supported by the equipment manufacturer. In such cases, a customer's ability to add desired functionality to an equipment vendor's system depends on vendor cooperation. Even if a vendor is willing to make a change, the time and the customer expense required are often considerable.

## 6.6 Support Issues

Sophisticated applications (e.g., advanced control, use of models incorporating parameters that need to be periodically "re-fit") often remain viable only as long as the developer (i.e., champion) is available to support the application. The same can sometimes be said of using highly customized proprietary automation systems developed by the customer. Given the reality in manufacturing plants of personnel losses and job rotations, there is much to be said, from a support perspective, in using standard automation platforms from well-known vendors. The same can be said for using highly configurable software for which a significant number of personnel (plant and vendor) are knowledgeable. A strong economic justification should exist before pursuing highly proprietary, customized, and/or hard coded automation systems and in pursuing process control strategies that few plant personnel will understand.

## 6.7 Money Pit

The cost of automation systems can range from thousands to millions of dollars, depending largely on the sophistication of the system, the amount of customization pursued, and the number of I/O points (i.e., sensors, instruments, valves). Cost categories include:

- capital cost of system,
- annual vendor support costs, e.g., to cover the cost of troubleshooting assistance and in developing the next versions of software (Support costs typically range from 10% to 20% of the system capital cost per year.),
- cost to develop (code and/or configure) applications,

- cost for any system customization,
- cost to test/validate, document, and implement the system,
- cost to maintain the ongoing automation support staff.

Rules of thumb (e.g., $x,000 capital cost per control loop) are often used to help provide a rough estimate of an automation project's cost before deciding to pursue an automation project.

In summary, automation does not represent a "free lunch." Automation, properly applied to appropriate applications, can provide high returns on investment. Payback, in many cases, has been reported to be less than three years.

## REFERENCES


1. Food and Drug Administration. Guide to inspection of computerized systems in drug processing. Washington, D.C.: Government Printing Office; Feb. 1983.
2. Validation concepts for computer systems used in the manufacture of drug products, the PMA's GMP Computer Systems Validation Committee. Pharmaceutical Technology, May 1986.
3. Alford JS, Cline FL(principal authors), and the PMA Validation Committee, Computer system validation – staying current: installation qualification, published in Pharmaceutical Technology, Sept. 1990 and also in Biopharm, Sep. 1990.
4. GAMP (Good Automated Manufacturing Practice) Guide for validation of automated systems, an ISPE publication, 5th ed. 2008.
5. A FDA perspective on quality by design, PharmTech, Dec. 2007.
6. U.S. Department of Health and Human Services, Food and Drug Administration. Guidance for industry PAT – a framework for innovative pharmaceutical manufacturing and quality assurance, Aug. 2003.
7. ASTM Standard E2363-04: Standard terminology related to PAT.
8. The validation dictionary, The Institute of Validation Technology.
9. Alford J, Cairney C, Higgs R, Honsowetz M, Huynh V, Jines A, Keates D, Skelton C. Real rewards from artificial intelligence. Intech; Apr. 1999.
10. Alford J, Cairney C, Higgs R, Honsowetz M, Huynh V, Jines A, Keates D, Skelton C. On-line expert-system applications; use in fermentation plants. Intech; Jul. 1999.
11. Buckbee G, Alford, JS. Automation applications in bio-pharmaceuticals, The Instrumentation, Systems, and Automation Society, 2008.
12. ISPE baseline guide for bio-pharmaceutical manufacturing facilities, ISPE, Tampa, FL, 2004.
13. ASME BPE-2002: Bioprocessing equipment, The American Society of Mechanical Engineers, New York, 2002.
14. Alford JS. Process mass spectrometry, a PAT application. J Proc Analyt Technol May–June, 2006. vol. 3. Issue 3, pp 6–10.
15. Cooley RE. Automating the control of process-scale purification columns using on-line liquid chromatography, BioPharm International Supplement, Feb. 2007.

16. Sekulic S, Wakeman JH, Doherty P, Hailey P, Automated system for the on-line monitoring of powder blending processes using near-infrared spectroscopy. Part II. Qualitative approaches to blend evaluation. J Pharm Biomed Anal 1998; 17, 8.
17. Alford JS. Bioprocess control, advances and challenges, presented at CPC VII, Jan. 2006 and published in a special issue of Computers and Chemical Engineering, Sep. 2006.
18. Alford JS. Modeling cell concentration in complex media. Biotechnol Bioeng 1978; 20, 12.
19. Gbewonyo K, Jain D, Hunt G, Drew S, Buckland B. On-line analysis of avermectin fermentation cell growth kinetics in an industrial pilot plant. Biotechnol Bioeng 1989; 34, 2.
20. Shinskey FG. Process control systems, application, design, and tuning. 3rd ed. McGraw Hill; 1988, New York, NY.
21. Riggs JB, Karim MN. Chemical and bio-process control. 3rd ed. Ferret Publishing; Lubbock, Texas 2006.
22. Trevathan V. A guide to the automation body of knowledge. 2nd ed. ISA; Research Triangle Park, NC 2006.
23. Mohan P., Glassey J. Montague GA. Pharmaceutical operations management: manufacturing for competitive advantage. McGraw Hill; New York, NY 2006.
24. Boudreau MA, McMillan GK. New directions in bioprocess modeling and control: maximizing process analytical technology benefits. ISA; Research Triangle Park, NC 2007.
25. Sweeny PJ. RFID for dummies. Wiley Publishing; Hoboken, NJ 2005.
26. Boeck LD, Wetzel RW, Burt SC, Huber FM, Fowler GL, Alford JS. Sterilization of bioreactor media on the basis of computer calculated thermal input, designated as $F_0$. J Ind Microbiol 1988; 3.
27. Ebey, Glenn C. A thermodynamic model for aqueous film coating, Pharmaceutical Technology, Apr. 1987.
28. U.S. Patent 6770141, Systems for controlling evaporative drying processes using environmental equivalency, issued to SmithKline Beecham Corp., Aug. 3, 2004.
29. 1Mort PR, Capeci SW, Holder JW. Control of agglomerate attributes in a continuous binder-agglomeration process, Powder Technol 2001; vol. 117, pp. 173–176.
30. A guide to the project management body of knowledge, 3rd ed., American National Standard, ANSI/PMI 99-001-2004.
31. Hopp W, Spearman M. Factory physics: foundations of manufacturing management. 3rd ed. McGraw Hill; New York, NY 2008.
32. ISPE ISPE commissioning and qualification guide, 5, International Society for Pharmaceutical Engineering, Tampa, FL, 2001.
33. Guidance for industry. Part 11: Electronic records; electronic signatures – scope and application, division of drug information, HFD-240, 2003. http://www.fda.gov/cder/guidance/index.htm.
34. ANSI/ISA Standard S-67.04.01-2000, Setpoints for nuclear safety-related instrumentation.
35. EEMUA Publication 191, Alarms systems: a guide to design, management and procurement, 1999.
36. ANSI/ISA Standard 18.02-2009, Management of Alarm Systems for the process industries.

# *Capillary Electrophoresis of Dendrimer-Related Medical Nanodevices*

**Mingwu Shen and Xiangyang Shi**

*College of Chemistry, Chemical Engineering and Biotechnology, Donghua University, Shanghai, P. R. China*

## 1 INTRODUCTION

Dendrimers are a new class of highly branched, monodispersed, and synthetic macromolecules with well-defined composition and architecture (1). The tailored core, interior structure, surface groups, and generation-dependent geometric properties of dendrimers make them a quite unique material for a range of applications in catalysis (2), sensors (3–5), optics (6), electronics (7,8), environmental remediation (9–11), and drug delivery (12–14). Since the pioneering work of the Tomalia, Newkome, and Vogtle groups in the 1980s (15–18), a surge of interest has been experienced in the synthesis and applications of dendrimers (1,19). Various types of dendrimers have been synthesized, including poly(amidoamine) dendrimers (Starburst PAMAMs), poly(propyleneimine) dendrimers (Astramol, PPI), the Newkome type cascade (ZCascade: methane (4):(3-oxo-6-oxa-2-azaheptylidyne):(propanoic acids)) dendrimers, the Frechet-type polyether dendrimers (20), the melamine

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing,*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

dendrimers (21), and the polyester dendrimers (22), to name a few. Among all of the synthetic dendrimers, PAMAM and PPI dendrimers have received a great deal of scientific and technological interest in medical applications due to their commercial availability.

Recent advances in dendrimer-based nanomedicine show that dendrimers have been used in many different ways in medical applications. For instance, dendrimers can be used as a platform to covalently conjugate targeting moieties, imaging dyes, and drug molecules for subsequent cancer cell targeting, imaging, and therapy (23–28). Dendrimers or functionalized dendrimers can also be used to physically encapsulate drug molecules inside their interior to improve the water solubility and bioavailability of the drugs (13,29–34). Multiple gadolinium chelator-modified dendrimer conjugates can complex much more gadolinium ions than the chelator, which significantly enhances the magnetic resonance (MR) imaging sensitivity *in vivo* (35,36). In addition, dendrimers can be used to entrap or stabilize metal nanoparticles for subsequent transmission electron microscopic imaging and therapy of cancer cells (37–42). Furthermore, functionalized dendrimers can be assembled onto magnetic iron oxide nanoparticles for MR imaging of tumor cells *in vitro* and *in vivo* (43–45).

As the synthesis and application of dendrimers go farther and farther, the analysis and quality control of dendrimer products are becoming more and more important in order to ensure their successful applications. Practical dendrimers and dendrimer derivatives always contain generational, skeletal, and substitutional diversities, which gives rise to great challenges for comprehensively characterizing their structural properties. Various types of analytical techniques have been utilized to characterize dendrimeric nanoparticles, including but not limited to, potentiometric titrations (46–48), dynamic light scattering (49), size exclusion chromatography (SEC) (24,25), high performance liquid chromatography (HPLC) (50–53), mass spectrometric methods (e.g., matrix-assisted laser desorption ionization time of flight, MALDI-TOF, electrospray ionization mass spectrometry, ESI-MS) (54,55), polyacrylamide gel electrophoresis (PAGE) (56–58), and capillary electrophoresis (CE) (47,48,58–63).

CE is a powerful chromatographic method utilized in the analysis of biologic macromolecules, such as DNA (64), proteins (65), and peptides (66), etc. CE has high efficiency, high sensitivity, short run time, and high automation capability, and it is suitable for the routine analysis of diverse dendrimer architectures, especially those that are carrying multiple charges. This chapter gives an extensive overview of the recent advances of CE analysis and characterization of various dendrimer nanomaterials from simple dendrimer derivatives to complex multifunctional dendrimeric medical nanodevices. The CE analysis technique varies depending on the structure, charge, and properties of dendrimers. In the mean time, similar dendrimer structures may display different migration behaviors under the electric field when different CE methods were adopted. This greatly helps understanding of the dendrimeric nanostructures from different aspects, further directing the successful application of dendrimers in the area of nanomedicine and pharmaceutical sciences.

## 2 CAPILLARY ELECTROPHORESIS OF DENDRIMERS AND DENDRIMER DERIVATIVES

### 2.1 Bare-Fused Silica Capillary

Bare-fused silica capillary has been typically used for analyzing polyanionic dendrimers because the negatively charged silica surface does not absorb the polyanionic dendrimers. This significantly helps improve the analytical reproducibility. Dubin and coworkers (67) studied the effect of pH on the mobility of carboxylic acid-terminated cascade dendrimers (ZCascade:methane (4):(3-oxo-6-oxa-2-azaheptylidyne):(propanoic acids)) (Newkome dendrimer) of generations 2–5 using capillary zone electrophoresis (CZE). It was found that even though the net charge for G5 (−972) is much higher than that of G2 (−36), its electrophoretic mobility (EM) is lower. The effective surface charge density of G5 is lower than the geometric surface charge density, which is attributed to counterion binding effect. It was suggested that a critical condition for non-specific counterion binding to spheres is analogous but not identical to the well-known Manning condition for counterion condensation on polyelectrolytes. In their further work (68), effect of electric field on the mobility of carboxyl-terminated Newkome dendrimers was investigated using CZE. It was shown that the mobility of the studied dendrimers increases with the field strength at moderate ionic strength, but varies inversely with $V$ at low ionic strength. Other carboxyl-terminated dendrimers (e.g., phenylacetylene dendrimers) (69) have also been analyzed using bare silica capillary. Results indicate that some form of cross-linking between or within higher generation dendrimer molecules occurs during the thermolysis reaction to produce the carboxyl termini. Therefore, CE analysis provides a useful tool for monitoring the homogeneity of the dendrimers.

Shi et al. (48,58) analyzed poly(amidoamine) succinamic acid (PAMAM-SAH) dendrimers of different generations using CZE and found that the EMs for all generations (1–7) are similar (Fig. 1), indicating that the polyanionic PAMAMs are not adsorbed on the surface of the quartz capillary, thus only the charge/mass ratio and the electroosmotic flow (EOF) influence the separation. The movement of ions mainly depends on the electric field and minor differences in charge/mass ratios are insufficient to achieve effective separations between generations. A G5(G3·SAH)$_n$ core-shell tectodendrimer with charge/mass ratio lower than that of single generation PAMAM-SAH dendrimers displayed a lower EM. Electropherogram of lower generations (1 and 2) showed that there are multiple peaks present before the main peak. This may be attributable to the presence of trailing generations and other imperfect dendrimer species converted into succinamic acid derivatives from the starting primary amine-terminated dendrimers. The charge/mass ratio of these dendrimer species can be differentiated at low generation level (1 and 2) when the respective charge/mass ratios are different from each other. With the increasing number of dendrimer generations, the charge/mass ratio differences of the trailing generations and imperfect structures become smaller and smaller. Accordingly, for G3·SAH through G7·SAH, only a single symmetric dendrimer peak was present in the electropherograms. Another possible reason is related to the manufacture of the

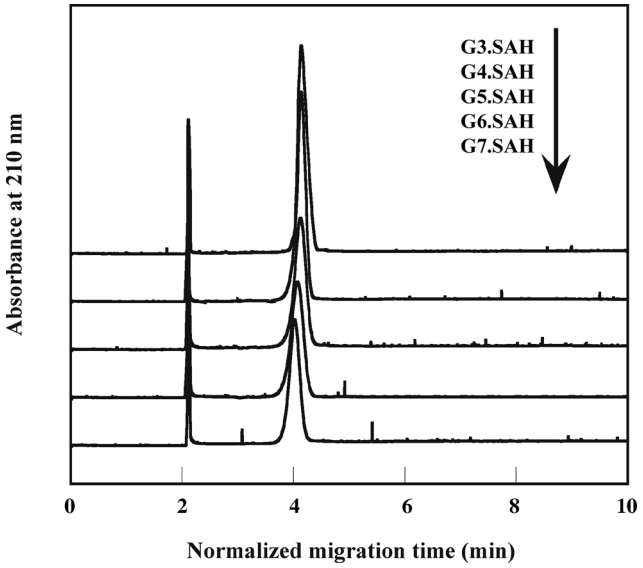


***FIGURE 1*** *Normalized electropherograms of PAMAM-SAHs of generation 3 through 7 analyzed using a bare fused-silica capillary (ID 75 $\mu$m, effective length 40 cm, and total length 48.5 cm). Injection time: 3 s. The first peak in all electropherograms corresponds to mesityl oxide, which was used as a neutral marker. The normalized CE electropherogram was based on the average migration time of the neutral maker (2.12 min, run time $n = 23$).* Reprinted with permission from Ref. (62). Copyright 2008 American Chemical Society.

starting primary amine-terminated dendrimers, during which lower generations (generation 0 through 2) are not purified by ultrafiltration process, therefore the purity of their succinamic acid derivatives is relatively low.

In another recent work, Shi et al analyzed G5 PAMAM dendrimer succinamic acids with defined charge density using a bare-fused silica capillary (59). Results showed that the migration time of these PAMAMs increases with the increase of carboxylation degree, which is ascribed to the increased charge/mass ratio, resulting in their increased EMs (Fig. 2). It was found that the peak of partially carboxylated G5 PAMAMs is much broader than 100% carboxylated G5 dendrimer. In addition, the practical charge/mass ratio versus carboxylation degree is linear, while their EMs as a function of both carboxylation degree and charge/mass ratio are nonlinear, in an "S" shape, indicating that beyond the charge/mass ratio, additional mechanisms are also involved in their separation. It was suggested that counter-ion binding onto charged dendrimer spheres and dendrimer hydrodynamic size may play a major role in influencing the nonlinear increase of EM of anionic-surfaced dendrimers with increasing charge/mass ratio. The molecular distribution and EM studies of similar generation dendrimers with defined charge density provides basic understanding of the molecular heterogeneity of multifunctional dendrimer nanodevices formed through multistep dendrimer surface functionalization.

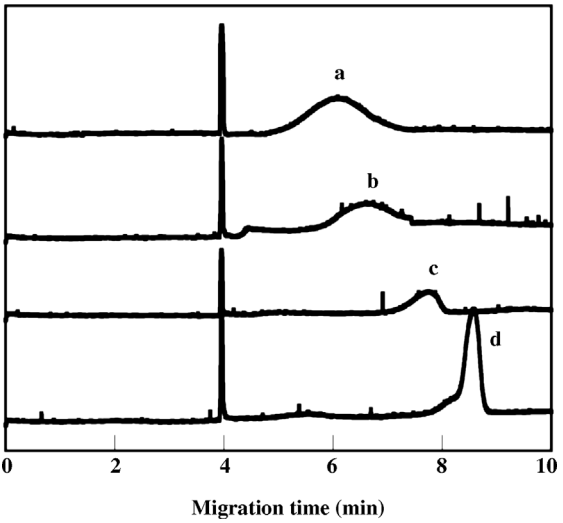


***FIGURE 2*** *Normalized electropherograms of carboxylated G5 PAMAMs having carboxylic acid termini of 25% (a), 50% (b), 75% (c), and 100% (d) of the available end groups collected using a bare silica capillary (ID 75 $\mu$m, total length 64.5 cm, and effective length 56 cm) under borate buffer (pH 8.3, 20 mM) condition. Injection time: 3 s. First peak corresponds to 4-methoxybenzyl alcohol, the neutral marker.* Reprinted with permission from Ref. (62). Copyright 2008 American Chemical Society.

Half-generation carboxylic acid-terminated PAMAM dendrimers have also been analyzed using CZE and showed similar EM behavior to that of PAMAM-SAH dendrimers (70,71). Compared with half-generation carboxyl-terminated PAMAM dendrimers made by hydrolysis of methyl esters (70), PAMAM-SAHs have the following advantages:modification step is simple and more controllable,polydispersity of PAMAM-SAHs is much lower than that of the half generation PAMAM carboxylates (70), and the number of terminal carboxyl groups remains similar to that of primary amine groups of the starting full-generation dendrimers, which makes comparisons between parent dendrimers and derivatives easier.

Bare silica capillary has also been used to separate polycationic primary amine-terminated PAMAM dendrimers. Brother et al. (70) and Ebber et al. (72) utilized bare-fused silica capillary to analyze ethylenediamine (EDA) core PAMAM dendrimers of different generations. Although the generational separation can be achieved, the reproducibility between runs is a problematic issue, which may affect the effective identification of dendrimers for the purpose of quality control. The irreproducibility of CE runs using a bare-fused silica capillary is due to the strong electrostatic interaction between the polycationic dendrimers and the negatively charged silica capillary surface.

In the case of G1 PAMAMs with small molecular weight, huge adsorption of PAMAMs onto bare silica capillary surface is not expected. Shi et al. (48) analyzed amine-, hydroxyl-, and acetyl-terminated G1 PAMAM dendrimers using a bare-fused

silica capillary. In their study, 2,3-diaminopyridine (2,3-DAP) was used as an internal standard in order to normalize the migration times of dendrimers with different terminal groups. This study provides new insights on the structural deviations of PAMAM dendrimers. Commercial full generation PAMAM dendrimers always contain "trailing generation(s)" as well as dimers, and real PAMAM molecules almost always display missing arms and molecular loops (56,57,60,70,73). These defects result from the four main types of side reactions during the synthesis: (1) a retro-Michael reaction gives rise to asymmetrical structures due to missing arms; (2) dimers may form in the amidation step by intermolecular coupling and (3) intramolecular cyclization (intramolecular coupling) may also occur during the same step as EDA is a bifunctional reactant; and (4) the residual EDA acts as a new core to initiate a trailing $(g-1)$ generation. Understanding the structural deviations of low generations is crucial for further characterization of higher generation PAMAMs. In Shi et al.'s study (47), they synthesized G1 PAMAMs with acetamide (G1·NHAc) and hydroxyl (G1·N(Gly)OH) termini from the same starting material, an amine-terminated PAMAM (G1·$NH_2$), and analyzed the structural deviations of these PAMAM derivatives by CE as well as other techniques for comparison (47). Comparison of the analytical results of these derivatives to those of the parent dendrimer is necessary to determine the real composition. The purity and homogeneity can be evaluated using CE based on the EM of the components including charge distributions between the molecules. Figure 3 shows normalized CE electropherograms of G1·$NH_2$, G1·NHAc,

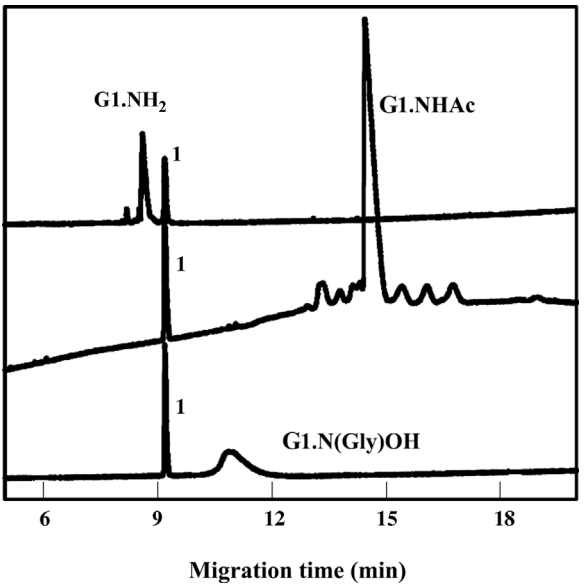


***FIGURE 3*** *Normalized CE electropherograms of G1·$NH_2$, G1·NHAc, and G1·N(Gly)OH (normalized migration time of the internal standard, $t_{IS}$ = 9.20 min, the average value of seven runs, detection wavelength at 210 nm). Peak 1: 2,3-diaminopyridine. CE was performed using a bare-fused silica capillary (ID 75 μm, total length 64.5 cm, and effective length 56 cm). Injection time: t = 3 s.* Reprinted with permission from Ref. (47). Copyright 2008 Elsevier Science.

and G1·N(Gly)OH. EM of G1 PAMAM decreases after surface modifications, as expected. At pH 2.5, the observed EM is in the order of G1·$NH_2$ > G1·N(Gly)OH > G1·NHAc, i.e., in the same order as their theoretical net charge/mass ratio. The peak shapes were also compared in Fig. 3, where the main peaks of G1·$NH_2$ and G1·NHAc are much narrower than the G1·N(Gly)OH peak. It can be clearly seen that the intensity of the main peak of G1·NHAc is much higher than that of G1·$NH_2$. After acetylation, there are more small visible peaks appearing in the G1·NHAc electropherogram (Fig. 3), and the differences between various components of G1·NHAc seem to be magnified compared to the starting material G1·$NH_2$. G1·N(Gly)OH (Fig. 3) displays only one broad peak, as all the individual peaks are hidden under an envelope. Further MALDI-TOF spectrometry verified that incomplete hydroxylation occurred even if 10 times molar excess of glycidol was added in the reaction mixture (47). CE results indicate that, after acetylation, the differences between substituted PAMAM molecules are amplified, whereas the differences diminish after hydroxylation.

Although bare-fused silica capillary has been used for analysis of smaller generation PAMAM dendrimers, the between-run irreproducibility always happens for the CE analysis of higher generation PAMAM dendrimers because of their strong interaction with capillary surface. Therefore, utilization of coated capillary for the analysis of polycationic dendrimers is essential.

## 2.2 Silanized Capillary

Since separation of polycationic PAMAM dendrimers using a bare fused-silica capillary lacks reproducibility, silanization of the capillary surface to decrease the interaction of the polycationic dendrimers with the capillary surface is necessary in order to improve the separation reproducibility (47,60–62). Shi et al. (60) compared the reproducibility of bare-fused silica capillary and chlorotrimethylsilane-modified silica capillary in the separation of G2·$NH_2$, G3·$NH_2$, G4·$NH_2$, and G5·$NH_2$ dendrimers (Figs. 4 and 5). The silanization reaction renders the capillary surface with hydrophobic methyl groups, which significantly reduce the adsorption of PAMAM dendrimers onto a silica substrate (74). It is clear that migration times of dendrimers are decreasing in the consecutive runs approaching a plateau. The long migration times (30–35 min) (see Fig. 4a) of the first two runs indicate strong binding of PAMAM dendrimers to the silica wall. After an initial "conditioning" of the capillary, a dynamic equilibrium develops and the PAMAM migration times become very similar to the ones in the previous runs. Figure 4b shows the achievable separation of the primary amine terminated PAMAM generations G2·$NH_2$, G3·$NH_2$, G4·$NH_2$, and G5·$NH_2$ using a bare-fused silica capillary in the fifth run. Silanization of the capillary surface may decrease the number of active centers and not only leads to faster equilibration but also clearly makes the separation more reproducible. When silanized capillary was used for the separation of the same mixture of PAMAM dendrimers (Fig. 5a), the migration times rapidly decreased in subsequent injections and stabilized at steady values. A typical separation at run 6 is shown in Fig. 5b, for G2·$NH_2$, G3·$NH_2$, G4·$NH_2$, and G5·$NH_2$ separated from their mixture. Although the

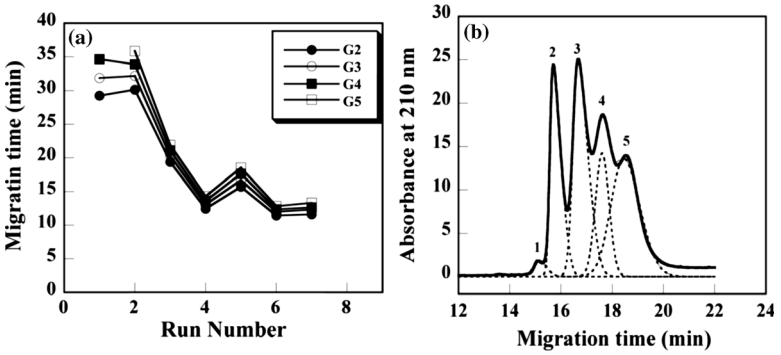


***FIGURE 4*** *(a) Relationship between the migration times of G2·NH$_2$, G3·NH$_2$, G4·NH$_2$, and G5·NH$_2$ separated using a bare fused-silica capillary as the function of consecutive run numbers. (b) A typical capillary electropherogram of the mixture of EDA core amine-terminated PAMAM dendrimers of generations 2–5 using a bare fused-silica capillary (ID 100 $\mu$m, total length 78.5 cm, and effective length 70 cm). Injection time: 3 s. Peaks 1, 2, 3, 4, 5 correspond to G1·NH$_2$ trace (in G2·NH$_2$ dendrimer), G2·NH$_2$, G3·NH$_2$, G4·NH$_2$, and G5·NH$_2$.* Reprinted with permission from Ref. (62). Copyright 2008 American Chemical Society.

run-to-run reproducibility using silanized silica capillary is significantly increased, day-to-day runs of similar dendrimer materials still experiences some irreproducibility issues, therefore utilization of internal standard to normalize the migration time of similar dendrimer species is necessary (60).

Using silanized silica capillary, Shi et al. (48) were able to analyze amine-, hydroxyl-, and acetyl-teminated G4 PAMAM dendrimers. These studies provide new

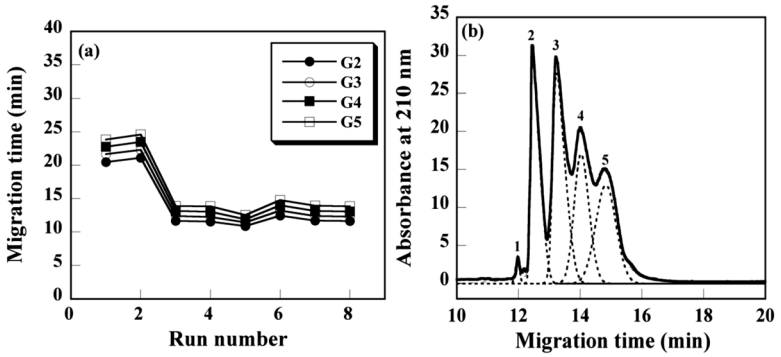


***FIGURE 5*** *(a) Relationship between the migration times and the consecutive run numbers for G2·NH$_2$, G3·NH$_2$, G4·NH$_2$, and G5·NH$_2$ dendrimers separated on silanized silica capillary. (b) A typical capillary electropherogram of the mixture of EDA core amine-terminated PAMAM dendrimers of generations 2–5 using a silanized silica capillary (ID 100 $\mu$m, total length 78.5 cm, and effective length 70 cm). Injection time: 3 s. Peaks 1, 2, 3, 4, 5 correspond to G1·NH$_2$ trace (in G2·NH$_2$ dendrimer), G2·NH$_2$, G3·NH$_2$, G4·NH$_2$, and G5·NH$_2$.* Reprinted with permission from Ref. (62). Copyright 2008 American Chemical Society.

insights on the structural deviations of PAMAM dendrimers. In a recent study, Shi and coworkers utilized amine-terminated PAMAM dendrimers of G4 as starting materials to synthesize acetyl, hydroxyl-terminated PAMAM derivatives. It is expected that high generation PAMAMs and derivatives will exhibit different properties compared to G1 dendrimers. Characteristically, density of the PAMAM molecules increases with generation and higher generation (G4 and above) PAMAM molecules have a close-to-spherical shape because of the exponentially growing number of terminal groups. With the growing number of terminal groups the relative differences become smaller, and the substitutional dispersity of higher generation PAMAMs may be masked in certain cases. Both quantitative analysis and charge distribution evaluation of the PAMAM derivatives were performed by CE. Figure 6 shows the CE electropherograms of G4·$NH_2$, G4·NHAc, and G4·N(Gly)OH dendrimers. The migration peaks of polycationic G4 PAMAM derivatives were normalized using the average migration time of 2,3-DAP (11.8 min). By comparing peak positions in the normalized electropherograms (Fig. 6), the EM of G4 PAMAM derivatives can be calculated. The EM of G4 PAMAMs follows the order of G4·$NH_2$ > G4·N(Gly)OH > G4·NHAc. Compared to CE studies of G1 PAMAMs, the substitutional dispersity of G4 PAMAM upon acetylation is not magnified as expected. The structural dispersity of hydroxyl-terminated G4·N(Gly)OH remains masked similar to G1 PAMAMs. The hidden substitutional dispersity of higher generation PAMAMs as well as their smaller generation counterparts mainly results from the incomplete hydroxylation reaction in nature. The incomplete hydroxylation produces mixtures of hydroxyl-terminated PAMAMs with very close molecular structures. Even at the G1 level, CE cannot

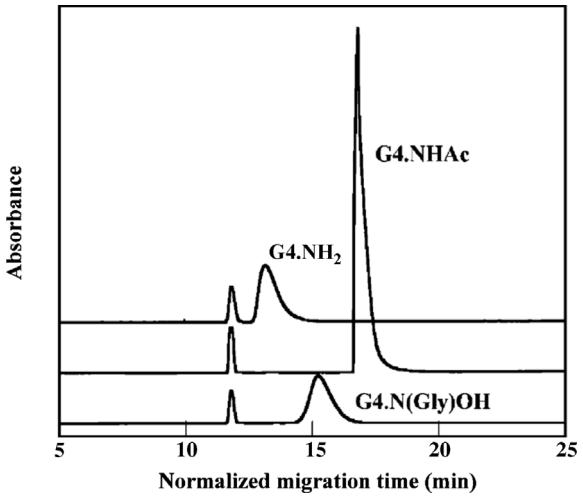


***FIGURE 6*** *CE electropherograms of polycationic G4 and derivatives (G4·$NH_2$, G4·NHAc, G4·N (Gly)OH). The first peak corresponds to the internal standard 2,3-DAP. The electropherograms are normalized using the average migration time of 2,3-DAP (11.8 min).* Reprinted with permission from Ref. (62). Copyright 2008 American Chemical Society.

differentiate the mixture of hydroxyl-terminated PAMAM derivatives with close molecular structures.

Silanized silica capillary has also been used to analyze PAMAM dendrimers of similar generation but with defined charge densities. PAMAM dendrimers with systematically varied surface charges are expected to serve as a nearly ideal model for the investigation of EM. The EM of the same generation PAMAM dendrimers with different functionalities (which is the case during the synthesis of multifunctional dendrimer nanodevices) has been investigated using CE. This provides new insights on further characterization of PAMAM-based multifunctional nanodevices. It is well established that the EM of linear polyelectrolytes is linearly proportional to their charge density at different zones (75). Charged PAMAM dendrimers, because of their nearly spherical shape, persistent size, and high concentration of nitrogen ligands in the interior and on the surface, are thought to exhibit behavior different from linear polyelectrolytes. As highly charged macromolecules are often cytotoxic, the dendrimer effective charges have to be reduced and dendrimer-based multifunctional nanodevices are usually prepared using partially modified PAMAM dendrimers as intermediate materials. Charge and charge distribution of partially modified dendrimers play an important role in determining physicochemical properties of the final products and influencing their interactions with biologic entities.

PAMAM dendrimer derivatives with defined substitutions were synthesized from generation 5 amine-terminated dendrimers ($G5{\cdot}NH_2$) with a systematic change of their acetylation degrees using standard methods (59). CE was employed to investigate the effect of charge on the EM of the polycationic dendrimer molecules as well as the molecular distribution changes upon defined substitutions (59). Figure 7 shows the normalized electropherograms of the polycationic acetylated G5 PAMAMs. Clearly, the migration time increases with the increase of acetylation degree, which is attributed to the decreased charge/mass ratio. It was found that the migration peaks of partially acetylated G5 PAMAMs are much broader than those of non-acetylated ($G5{\cdot}NH_2$) or of the fully acetylated G5·100Ac. It was also found that although the practical charge/mass ratio of these polycationic dendrimers (determined by $^1$H NMR and titration) decreases linearly with the increase of acetylation degree, their EMs decrease nonlinearly, in a concave downward shape, which indicates that the difference in their charge/mass ratios is not solely responsible for the migration rate differences of the PAMAM dendrimer derivatives. Partially modified G5 PAMAMs have much broader overall distribution than those of fully modified or non-modified G5 dendrimers due to the stochastic nature of this substitution reaction. Consequently, the highest charge distribution ensues when 50% of the surface functionalities are substituted. EM is usually linearly dependent on the charge/mass ratio of the studied materials. It often changes when different mechanisms are involved in the separation process. Regarding the nonlinear increase of EM (concave downward shape) of G5 PAMAM acetamides (59), there may be several factors influencing their separation. (1) Adsorption–desorption of dendrimers onto the capillary surface: silanization coating onto the capillary internal surface cannot exclude the complete desorption of polycationic PAMAM dendrimers onto the quartz surface (58,60). G5 PAMAM acetamides with lower acetylation degree (higher charge/mass ratio) display much

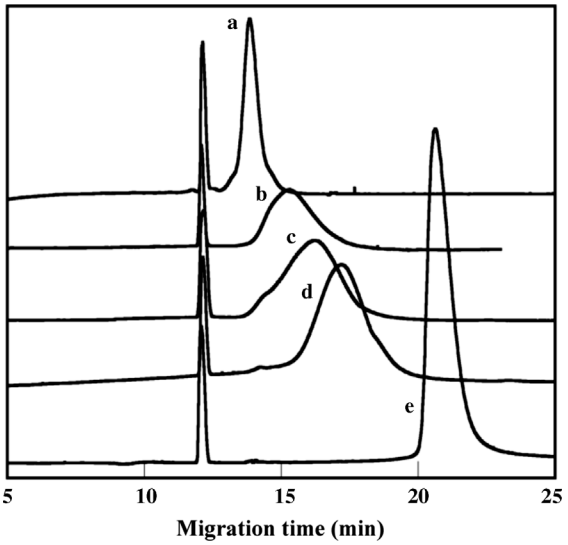


***FIGURE 7*** *Normalized electropherograms of acetylated G5 PAMAMs with acetylation percentage of (a) 0%, (b) 25%, (c) 50%, (d) 75%, and (e) 100% analyzed using a silanized silica capillary (ID 100 μm, total length 78.5 cm, and effective length 70 cm) under phosphate buffer (pH 2.5, 50 mM) condition. Injection time: 3 s. First peak corresponds to the internal standard 2,3-DAP.* Reprinted with permission from Ref. (62). Copyright 2008 American Chemical Society.

stronger adsorption on the capillary surface than that of G5 PAMAM acetamides with higher acetylation degree. Consequently, the EMs of G5 PAMAM acetamides with higher charge/mass ratio are slightly reduced. (2) Counterion binding onto charged dendrimer spheres: as demonstrated by Manning's theory (76) and confirmed later by Dubin and co-workers' studies (67,68,75) on linear and dendrimeric polyelectrolytes using CE, counterion condensation around polyelectrolytes strongly influences their electrokinetic mobility. Acetylated G5 PAMAMs bearing more terminal amine groups may experience higher counterion binding force during electrophoresis. (3) Dendrimer hydrodynamic size: on the basis of previous work (46), an increasing number of acetyl termini leads to the shrinkage of the dendrimer molecule. Larger dendrimer spheres may experience a larger inhibition force during the electrophoresis, and the acetylated G5 PAMAMs having more acetamide groups experience less hindrance than those bearing more terminal amine groups.

## 2.3 Dynamic Coating Capillary

In order to decrease the adsorption of amine-terminated PAMAM dendrimers onto capillary surface, Sedlakova et al. (77) developed a suitable method for the analysis of highly positively charged amine-terminated PAMAM dendrimers. The interaction between amino groups of PAMAM dendrimers and silanol groups was prevented by the dynamic coating of a fused-silica capillary with an amino compound (polyethylenimine) present in the background electrolyte. This new method based on the

dynamic coating of the capillary was developed for the separation of dendrimers at pH 7.4, instead at extremely acidic pH (pH = 2.5) used by other groups (60,70,72). The method comprised separation in a fused-silica capillary (57 cm total length, 50 cm to the detector, ID 75 μm) and a background electrolyte consisting of a Tris-phosphate buffer (50 mmol/L, pH 7.4) and 0.05% (w/v) polyethyleneimine. This system is suitable for the separation of seven generations of dendrimers (generations 0–6). The dynamic coating agent (polyethyleneimine) also improves the separation at acid pH.

## 2.4 Poly(Vinyl Alcohol)-Coated Capillary

Poly(vinyl alcohol) (PVA) has been permanently coated onto the surface of capillary in order to improve the reproducibility of dendrimers. Sharma and coworkers (78) demonstrated that the use of a PVA-coated capillary significantly decreases the migration times of the nanomolecules without compromising resolution. Dendrimer mixtures containing generations 0–5 are separated as discrete, non-overlapping peaks in about 15 min. In addition, the between-day precision of retention times was dramatically improved without the need for internal standards or data normalization. Dendrimers of various generations and cores run on different days showed an RSD of retention times of less than 4%. The PVA coating was very stable as shown by the excellent precision of migration times obtained on a capillary used for a month with more than 100 injections. Welch and Hoagland (79) utilized PVA-coated capillary to investigate the EM of PPI dendrimers and concluded that charged PPI dendrimers behave as charged spheres.

Using a PVA-coated capillary, Desai et al. analyzed the PAMAM succinamic acid dendrimers of different generations (63). Due to the excellent stability and osmotic flow-shielding effect of the PVA-coated capillary, highly reproducible migration times were achieved for all generations of dendrimer (e.g., RSD for the migration times of G5 dendrimer was 0.6%). It was also observed that a reverse trend in migration times exists for the PAMAM-SAH dendrimers (i.e., higher generations migrate faster than lower generation dendrimers) compared to amine-terminated PAMAM dendrimers reported in the literature (60,70,72,78). This reversal in migration times was attributed to the difference in counterion binding around these negatively charged dendrimers. The counterion binding effect, in certain circumstances, could play an important role in the separation of dendrimers. When compared with a bare silica capillary, a PVA-coated capillary with a permanent coating of hydrophilic polymer on its surface can completely shield the effects of EOF. Therefore, the migration of the PAMAM-SAH dendrimer samples is dependent on both their charge/mass ratios and the effects of counterion binding. Smaller generations (G1–G3) exhibit a relatively open molecular structure and behave as a linear polymer, thereby receiving a higher inhibition force because of the counterion binding effect; while high generations (G4–G10) exhibit a densely close globular structure and behave as a colloidal sphere, thus receiving a smaller inhibition force. The differences in the counterion binding effect for the CE analysis of PAMAM-SAH dendrimers with different generations allows the separation of mixed generations for small generations of PAMAM-SAH dendrimers (Fig. 8), which are impossible to be

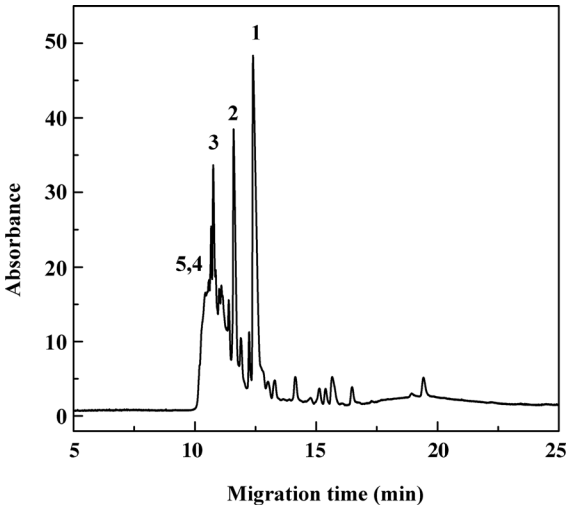


***FIGURE 8*** *A typical capillary electropherogram of the mixture of PAMAM-SAH dendrimers of generations 1–5 using a PVA-coated capillary (ID 50 μm, total length 64.5 cm, and effective length 56 cm). Injection time: 4 s. The concentration of each generation is 1 mg/mL.* Reprinted with permission from Ref. (63). Copyright 2008 Wiley-VCH.

separated using a bare silica capillary (58). Because of their dense spherical molecular conformation, close migration times, and broadness of the eluting peak, it is hard for higher generation dendrimers (G4 and G5) to be separated under these conditions (Fig. 8). It should also be noted that under the separation buffer condition (pH 7.0), it is likely that lower generation dendrimers could have a different p*I*, compared to higher generation dendrimers, thus the percentage of ionization for various generations at pH 7.0 could be slightly different. This could affect the overall charge/mass ratio of the dendrimer from generation to generation. This change in charge/mass ratio is larger between lower generations, compared to higher generations (G4 and G5), resulting in a better resolution (generational separation) for lower generation dendrimers.

## 2.5 Other Coated Capillaries

Other coated capillaries (e.g., surfactant, polyethylene glycol (PEG)) have also been used to analyze dendrimers. Welch and Hoagland (79) tried to use positively charged cetyltrimethylammonium bromide (CTAB) to reverse the bare capillary wall charge for the CE analysis of PPI dendrimers. CTAB appears to associate with PPI dendrimers at higher pH (pH > 9.0), raising their charge. Therefore, only PVA-coated capillary were used at elevated pH for their work. Stockigt et al. (80) utilized PEG-coated capillary to analyze polydisperse dendrimeric diaminobutane (DAB)-based polynitriles through the use of the online coupling of CE with a sector mass spectrometer via an electrospray ionization (ESI) source. The CE/MS method provided fast and detailed information concerning the heterogeneity of the dendrimer samples.

### 2.6 Capillary Gel Electrophoresis

In capillary gel electrophoresis (CGE) analysis, the separation of dendrimers is based on both size and charge/mass ratio. Literature reports do not show many applications of CGE analysis of dendrimers. In a recent report, Tomalia and coworkers (81) demonstrated that a novel, star-shaped copolymers composed of G3 or G4 PAMAM dendrimer core and polyacetylene shell can be analyzed and identified through the use of CGE. In the CGE method, 0.5% hydroxypropylmethylcellulose (HPCE) was filled in silica capillary as a gel sieving matrix before analysis. Other type of CGE method could include the use of preformed poly(acrylamide) gel in the capillary which has been used for DNA sequencing. However, this type of CGE has not been used for dendrimer analysis in literature.

## 3 CAPILLARY ELECTROPHORESIS OF DENDRIMERIC MEDICAL NANODEVICES

### 3.1 Mono- and Multifunctional Complex Dendrimer Medical Nanodevices

PAMAM dendrimers are ideal nanoplatforms to conjugate a defined number of targeting moieties, dye tags, and drugs for multifunctional targeting, imaging, and treatment of biological systems, especially for targeted cancer therapy. A novel PAMAM dendrimer-based cancer therapy agent carrying biospecific targeting moieties (e.g., folic acid (FA)), fluorescent dyes, and drugs have been developed at the Michigan Nanotechnology Institute for Medicine and Biological Sciences (MNIMBS), University of Michigan (23–25,27,82,83). These dendrimer nanodevices are demonstrated to be able to effectively transverse vascular pores and directly infuse tumor cells. Although the research on medical applications of dendrimer-based nanodevices continues progressively, the characterization of these multifunctional dendrimer nanodevices still remains a great challenge. The comprehensive understanding and analytical characterization of the structural heterogeneity of these novel synthetic nanodevices have not been widely investigated.

For mono- and multifunctional dendrimer-based nanodevices, beyond the generational and skeletal dispersities that are inherently related to the synthetic technology, substitutional dispersity is often expected due to the multistep surface modifications with different functional moieties. As opposed to other techniques, CE allows simultaneous evaluation of both purity and charge distribution of the analyzed materials.

Using silanized silica capillary, Shi et al. demonstrated that monofunctional dendrimer-dye conjugates could be analyzed and identified (62). G5-6T (6-carboxytetramethyl-rhodamine) conjugates (28,84) with both amino and acetamide termini were synthesized for biosensing purposes and analyzed using CE. The detection wavelengths were selected at 210, 250, 280, and 520 nm. The 520 nm wavelength is the maximum absorbance of 6 T dye molecules. Figure 9 shows the normalized electropherograms of G5·$NH_2$-6T and G5·Ac-6T conjugates at both 210 nm (a) and 520 nm (b). It is clear that both G5·$NH_2$-6T and G5·Ac-6T materials exhibit similar

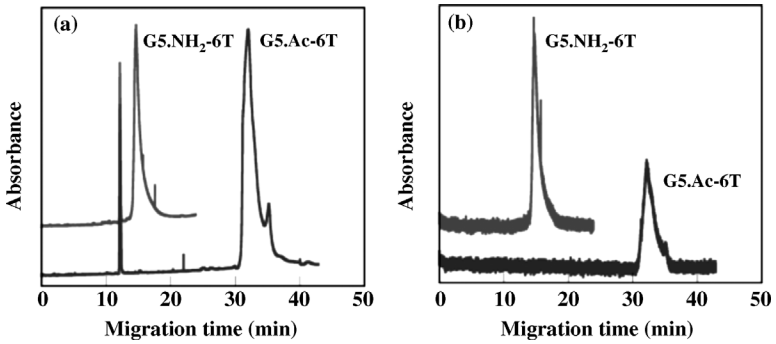


***FIGURE 9*** *Normalized electropherograms of G5·NH$_2$-6T and G5·Ac-6T at 210 nm (a) and 520 nm (b).* Reprinted with permission from Ref. (62). Copyright 2008 American Chemical Society.

peak shapes at two different wavelengths, indicating that the materials are quite homogeneous. The CE peak at 520 nm (see Fig. 9b) further supports the presence and distribution of 6 T moieties, because at 520 nm only 6 T moieties have absorbance. In another study, Shi and coworkers demonstrated that the purity and homogeneity of carboxyl-terminated G5·SAH-6T dendrimer conjugates can be analyzed and identified using a PVA-coated capillary (63).

For multifunctional complex dendrimer medical nanodevices, evaluation of the molecular heterogeneity is very important. Shi et al. demonstrated that the molecular heterogeneity is largely dependent on the synthetic approaches (61). In a recent work, Shi and coworkers have synthesized a variety of G5 PAMAM dendrimer-based functional nanodevices using two synthetic approaches (see Fig. 10) and analyzed by

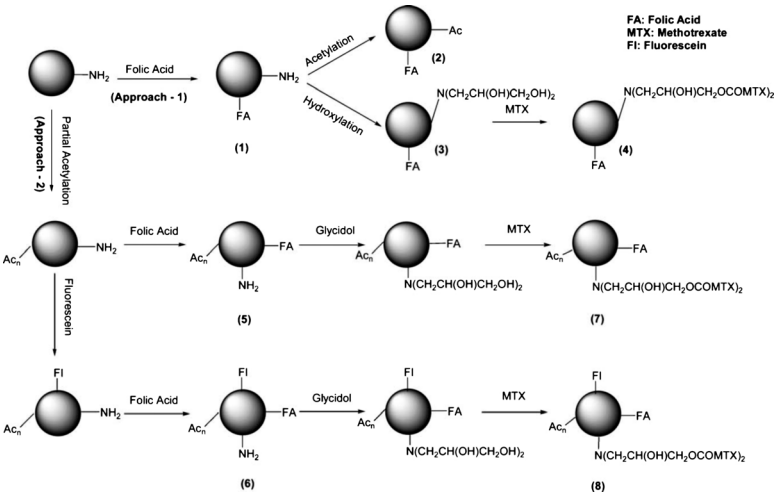


***FIGURE 10*** *Schematic representation of the synthetic approaches used to synthesize mono- and multifunctional dendrimer nanodevices.* Reprinted with permission from Ref. (61). Copyright 2008 Royal Society of Chemistry.

CE using a silanized capillary. Approach 1: synthesis from amine-terminated G5·$NH_2$ dendrimers; the synthesized materials include monofunctional G5–FA conjugates (terminated with amine, acetamide, and hydroxyl groups), and bifunctional G5·FA–OH–MTX (MTX denotes methotrexate) conjugates terminated with hydroxyl groups. Approach 2: synthesis from partially acetylated G5 dendrimers; the synthesized materials include a set of mono- and multifunctional G5·$Ac_n$–FA, G5·$Ac_n$–FI–FA, G5·$Ac_n$–FA–OH–MTX, and G5.$Ac_n$–FI–FA–OH–MTX nanodevices (FI denotes fluorescein).

For the analysis of G5–FA conjugates terminated with amino, acetamide, and hydroxyl groups synthesized using Approach 1, it is apparent that conjugate 3 exhibits a relatively homogeneous migration peak, while 1 and 2 display at least bimodal distribution. Analysis of G5–FA–MTX conjugate (4) synthesized using Approach 1 showed that G5·FA–OH–MTX bifunctional nanodevice (4) exhibits a monomodal symmetrical peak. The material was confirmed to be pure, as no impurity peaks were present in the electropherogram even at a long migration time up to 50 min.

In another experiment, a set of mono- and multifunctional G5 nanodevices synthesized using Approach 2 were analyzed using CE. Figure 11 shows the CE electropherograms of G5·$Ac_{75}$, G5·–$Ac_{75}$–FA (5), G5·$Ac_{75}$–FI–FA (6), G5·$Ac_{82}$–FA–OH–MTX (7), and G5–$Ac_{82}$–FI–FA–OH–MTX (8) conjugates. It is clear that all dendrimer conjugates display a single peak, indicating their relatively good monodispersity.

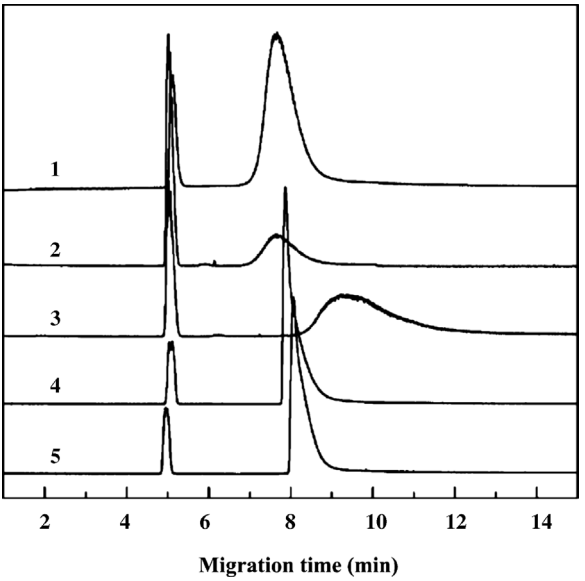


***FIGURE 11*** *CE electropherograms of G5 dendrimer conjugates synthesized using Approach 2 detected at 210 nm. The first peak is related to the internal standard 2,3-DAP. The CE was run using a silanized silica capillary (ID 100 μm) with effective length of 40 cm and total length of 48.5 cm. Curve 1: G5·Ac75; curve 2: G5·Ac75–FA (5); curve 3: G5·Ac75–FI–FA (6); curve 4: G5·Ac82–FA–OH–MTX (7); curve 5: G5–Ac82–FI–FA–OH–MTX (8).* Reprinted with permission from Ref. (61). Copyright 2008 Royal Society of Chemistry.

Conjugate 6 shows a longer migration time than both G5·$Ac_{75}$ and conjugate 5, which is probably due to its lower charge/mass ratio and structural complexity after conjugation with the FI moiety. G5·$Ac_{82}$–FA–OH–MTX (7) and G5–$Ac_{82}$–FI–FA–OH–MTX (8) conjugates were synthesized from a partially acetylated G5·$Ac_{82}$ dendrimer. After conjugation with FA and/or FI, the intermediate products were hydroxylated by reacting with glycidol, followed by conjugation with MTX to give the final MTX-conjugated nanodevices. Both conjugates (7 and 8) show narrower peaks than those of G5·$Ac_{75}$, 5, and 6 (please note that the comparison is reasonable because the difference between G5·$Ac_{75}$ and G5·$Ac_{82}$ is small enough and can be neglected). The narrow peaks of 7 and 8 might result from the hydroxylation reaction, which has been shown to mask the polydispersity of dendrimers in previous work (47,48). In a sense of molecular heterogeneity, hydroxylation actually improves the monodispersity of the multifunctional dendrimer nanodevices.

As opposed to other analytical techniques (i.e., SEC, UV–Vis spectrometry, and NMR) that can only detect the average molecular physicochemical parameters, CE provides a unique way to detect the distribution of molecular species based on the differences in their EMs resulting from the change of their charge/mass ratios. Possible factors affecting the molecular distribution of functional dendrimer conjugates are briefly discussed as follows: (1) The classification of dendrimer terminal groups influences the surface moiety distribution of dendrimer conjugates. We previously demonstrated that the dispersities of dendrimers could be magnified by acetylation and masked by hydroxylation for amine-terminated PAMAM dendrimers (47). For all materials synthesized using Approach 1 (starting from amine-terminated G5·$NH_2$), more prominent FA moiety distribution can be observed in G5·FA–Ac (2) than that of G5·$NH_2$–FA (1), while G5·FA–OH (3) and G5·FA–OH–MTX (4) only show a monomodal distribution. In the case of hydroxyl-terminated dendrimer conjugates, because of the incomplete hydroxylation and the existence of the resulting secondary terminal amines, the very close number of terminal hydroxyl groups masks the molecular differences between individual dendrimer conjugate molecules. Similarly, multifunctional G5·$Ac_{82}$–FA–OH–MTX (7) and G5–$Ac_{82}$–FI–FA–OH–MTX (8) nanodevices synthesized using Approach 2 display narrower CE peaks than those of G5·$Ac_{75}$, 5, and 6 that do not possess hydroxyl terminal groups (Fig. 11). This further suggests the dispersity masking effect of hydroxylation reaction for dendrimer conjugation. (2) The synthetic approaches. Two synthetic approaches were employed to synthesize mono- and multifunctional G5 dendrimer nanodevices. Compared with Approach 1 which the conjugation reaction started from amine-terminated G5·$NH_2$ dendrimer, in Approach 2 partially acetylated G5 dendrimer was instead used as a starting material. The high local concentration of amine groups of G5·$NH_2$ makes it highly reactive, thereby significantly distributing the conjugated moiety molecules. Whereas, after partial acetylation reaction, the small portion of amine groups remaining on the G5 dendrimer surface can preserve the homodispersity of the dendrimer for subsequent bioconjugation (25). Consequently, mono- and multifunctional dendrimer nanodevices synthesized using Approach 2 always display a monomodal distribution.

## 3.2 Complex Dendrimeric Contrast Agents

Complex dendrimeric contrast agents offer medical imaging applications. Dendrimer-based gadolinium (Gd)-containing MR imaging contrast agents have been prepared from various dendrimeric nanoparticles, e.g., PAMAM dendrimers (36), PPI dendrimers (85), and DAB dendrimers (35). A promising clinical candidate is Gadomer, which is currently developed as a blood pool agent for MR imaging (86,87). Although research on dendrimer-based Gd contrast agents have been developed for pre-clinical uses, the CE analysis of the related Gd-dendrimer complex contrast agents has just appeared since 2006 (88,89).

In a recent report, Vetterlein et al. (89) developed a CE method for the characterization of the complex dendrimeric contrast agent Gadomer. The method was capable of separating the target substance Gadomer 24 from related dendrimers containing amino or carboxyl functions and from impurities of lower molecular weight (Fig. 12). The compounds were separated in a fused-silica capillary. The optimized background electrolyte (BGE) consisted of 15 mM sodium phosphate, pH 6.3, containing 0.5 mM CTAB. The addition of CTAB led to a positively charged capillary wall, reversed the EOF, and prevented interaction of Gadomer with negatively charged silanol groups. The method allows the determination of the purity and stability of the drug substance Gadomer as well as its injectable formulation. On the basis of the study, a strategy for the quality assurance and quality control of the complex dendrimeric drug candidate Gadomer may be devised. By using CE–MS, the complex dendrimeric contrast agent Gadomer has been comprehensively characterized. The analytical challenges arising from the structure of this analyte, e.g., the presence of a complex isotopic pattern originating from multiple gadolinium chelates could be overcome using this combined approach. In order to make the CE method compatible with the MS detection, neither the use of sodium phosphate nor of CTAB as constituents of the buffer system could be used. In the CE method, polybrene/dextran sulfate multilayer-coated capillary surface proved to be suitable for CE–MS analysis of Gadomer. The profile of related dendrimers and impurities of lower molecular weight present in Gadomer could be elucidated and variances in the pattern of related dendrimers present in batches obtained via different synthetic routes could be detected. The qualitative analysis of Gadomer using MS techniques in combination with CE provides the fundamental basis for a framework of analytical methodologies for the characterization of Gadomer. By designing such a framework of methodologies, the challenges associated with the introduction of dendrimers into the clinical stage may be overcome and the reproducible quality and safety of these innovative products can be ensured.

## 3.3 Interaction of Biomolecules with Dendrimers

Understanding the interaction between dendrimers and biomolecules (e.g., nucleic acids, protein, and peptides) is important for the development of dendrimers as drug and gene delivery carriers. CE has been utilized for the analysis of binding behavior between dendrimers and biomolecules (90–92).

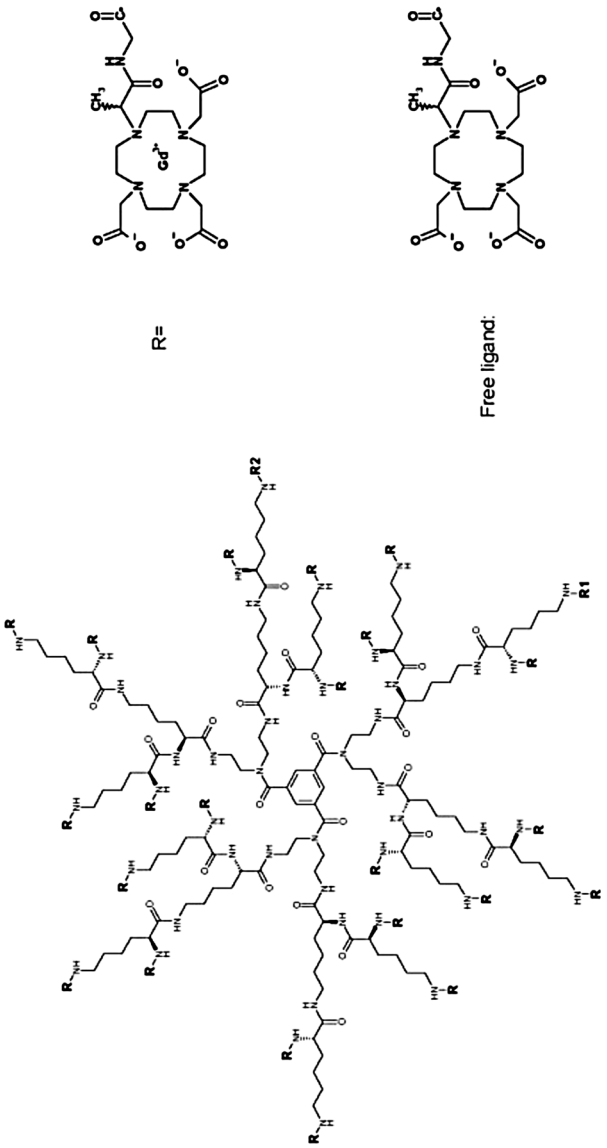


| Compound name | Number | R1 | R2 |
|---|---|---|---|
| Gadomer 24 | 1 | R1 = R | R2 = R |
| Gadomer 23 (3COOH) | 2 | R1 = free ligand | R2 = R |
| Gadomer 22 (6COOH) | 3 | R1 = free ligand | R2 = free ligand |
| Gadomer 16 (1COOH) | 4 | complete branch detached from core of molecule | |
| Gadomer 23 (1$NH_2$) | 5 | R1 = H | R2 = R |

***FIGURE 12*** *Structures of Gadomer 24 and selected related dendrimers.* Reprinted with permission from Ref. (89). Copyright 2008 Wiley-VCH.

Shcharbin et al. analyzed the interaction between PAMAM G4 dendrimer and serum albumins using CE (91). It was shown that serum albumins and PAMAM G4 dendrimer form the complex with stoichiometry of 4–6:1 for G4:HSA and 4–5:1 for G4:BSA molar ratio. In the CE analysis, it was found that the addition of G4 dendrimer to protein leads to (1) the decrease of protein area and (2) the decrease of peak elution times. At dendrimer/protein molar ratio $n = 10$ the peak of pure protein disappeared. From circular dichroism studies, it has been proposed that dendrimer does not significantly affect the protein secondary structure.

In another recent study, using a bare-fused silica capillary, Jiang and coworkers (92) investigated the influence of 2–5 generations of PAMAM dendrimers on the inhibition of Tat protein/TAR RNA binding behavior by CE. Because of the perfect structure and charge distribution, PAMAM dendrimers can be used as a good model for studying the interaction of ligand and acceptor, and parameters that cause the different affinity of drug. In that study, the different binding abilities of PAMAM dendrimers/or Tat peptide to TAR RNA was investigated using CE to gain a deeper insight into the possible mechanism. Mixture of TAR/PAMAM dendrimers/Tat having a molar ratio of 1:5:5 was used to examine the competition between TAR-PAMAM and TAR-Tat. It shows that the complex of each generation of PAMAM dendrimers with TAR RNA exhibits their own migration time. The migrating time of G2-TAR is the closest one to that of TAR RNA, and less than TAT-Tar, implying that the interaction between G2 and RNA is weak in solution. The sequence of mobility of PAMAM dendrimers after complexing with RNA changed completely, i.e., the mobility is G3-RNA > G5-RNA > G4-RNA > Tat-RNA > G2-RNA. It is intriguing that this sequence is inconsistent with the binding coefficient ($K_D^{-1}$)s data, implying that the interaction between RNA and PAMAM dendrimers is not only a simple electrical attraction or weight effect as the theoretical charge/mass ratios are very close to each other. This result shows that the other factors, which are more essential for the ligand–receptor complexation, must exist. If it is not a question of electrical attraction, then perhaps it is the size of PAMAM that could fit the loop in TAR more properly. The CE data combined with quartz crystal microbalance (QCM) results gave a quantitative comparison of the complexation ability of 2–5 generations of PAMAM with TAR RNA. The binding ability of G3 is the strongest one which should be the best candidate for the future drug. The results could be used to screen the numerous candidates of HIV drugs *in vitro* simply and rapidly, providing a new method for the pharmacology study of HIV-1.

For drug delivery applications, it is important to understand the binding or encapsulation capability of the target drug molecules with or within dendrimers. PAMAM dendrimers have been shown to be an excellent carrier to encapsulate hydrophobic drug molecules (12,30–32). However, little is known for the binding or encapsulation of peptide drugs with or within dendrimers. Vasoactive intestinal peptide (VIP), a 28-aa neuropeptide widely distributed in the central and peripheral nerve system, exhibits many interesting anti-inflammatory and immunomodulatory *in vitro* and *in vivo* properties (93–96). However, the half-life of exogenous VIP in biological fluids is very short, most likely due to degradation and inactivation of the

peptide (97,98). The rapid degradation of VIP is a major factor limiting the potential clinical applications of this peptide. To overcome this problem, strategies have been proposed to improve VIP efficacy and/or stability in biological fluids by means of new delivery systems, using for instance, the potential of liposomes. Nevertheless, due to their poor stability, liposomes are not suitable to retain efficiently the VIP peptide for a long time. In order to develop the use of PAMAM dendrimers as a possible drug delivery vehicle for VIP, Dribek et al. (90) reported the use of affinity capillary electrophoresis (ACE) to study the molecular interactions between VIP and PAMAM dendrimers. In their work, poly(ethylene oxide) (PEO)-, poly(diallylmethylammonium chloride (PDDA)-, and poly(acrylamide) (PAA)-coated capillaries were used to analyze the affinity between VIP and four PAMAMs: two amine-terminated PAMAM dendrimers (G2·$NH_2$ and G5·$NH_2$) and two carboxyl-terminated PAMAM derivatives (G2·COOH and G5·COOH). The potential of PDDA and PEO coatings to prevent the adsorption of the strongly basic peptide was demonstrated. For the interaction between VIP and amine-terminated dendrimers, PDDA-coated capillary was used; while for the interaction between VIP and carboxyl-terminated PAMAM dendrimers, PEO-coated capillary was used. The ACE method developed showed that among the dendrimers studied, only the cationic dendrimer of generation 5 G5·$NH_2$ could establish non-covalent interactions with VIP (Fig. 13), while no affinity was evidenced between VIP and all other dendrimers investigated. It was demonstrated that assuming a 1:1 stoichiometry, the binding constant between VIP and G5·$NH_2$, determined from nonlinear and linear regressions, was around $(6.7 \pm 1.1) \times 10^{-4}\ M^{-1}$. Thus, the ACE approach developed could be useful for further screening of a set of dendrimers bearing various physicochemical properties (i.e., neutral terminal

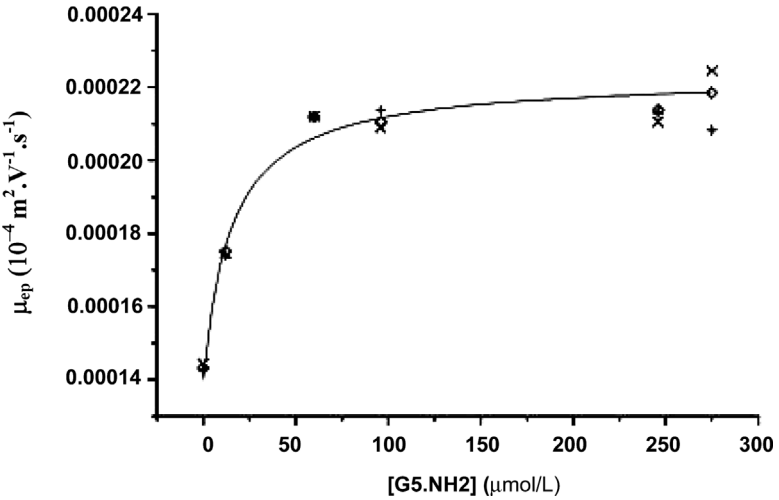


***FIGURE 13*** *Plot of VIP corrected electrophoretic mobility as a function of G5·$NH_2$ dendrimer concentration added to the BGE.* Reprinted with permission from Ref. (90). Copyright 2008 Wiley-VCH.

group, higher size, etc.) in order to gain understanding on the parameters influencing the VIP–dendrimer interactions that could be helpful for the development of novel VIP formulations based on dendrimer encapsulation strategy.

## 4 SUMMARY AND OUTLOOKS

This chapter has described the application of CE for the analysis of diverse dendrimer nanoparticles ranging from amine-terminated dendrimers, carboxyl-terminated dendrimers, fully modified dendrimers of different generation and functionalities, dendrimers of defined substitutions, and multifunctional dendrimer-based medical nanodevices and contrast agents, as well as the interaction between dendrimers and various types of biomolecules (e.g., protein, peptides, and nucleic acids). These dendrimer-related various derivatives and nanodevices have been analyzed using various types of capillaries, including, but not limited to bare-fused silica capillary, silanized capillary, PVA-coated capillary, dynamic coating capillary, surfactant-coated capillary, and other polymer-coated capillary. In addition, CGE mode has also been used for the analysis of dendrimer architectures. As one of the routine analytical techniques, CE has been demonstrated to be a vital tool to analyze PAMAM dendrimer nanoparticles for medical applications. Various dendrimer nanoparticles have been evaluated by CE in terms of their purity, EM, molecular charge distribution, and dendrimer/biomolecule interaction.

The tailorable dendrimer surface chemistry and various bioconjugation techniques allow one to design various complex dendrimer medical nanodevices for various biomedical applications. The adjustable CE techniques should be amenable for the analysis of various novel dendrimer nanodevices. As dendrimers are continuously recognized as powerful nanoplatforms for biological nanotechnologies used for treatment, diagnosis, monitoring, and control of biological systems, we envision that CE will continuously play an important role to evaluate and analyze various novel multifunctional dendrimer nanodevices for detection, targeting, drug delivery, treatment, and therapeutics in various biological systems.

## REFERENCES

1. Tomalia DA, Frechet JMJ. Dendrimers and other dendritic polymers. New York: John Wiley & Sons Ltd; 2001.
2. van Manen HJ, van Veggel FCJM, Reinhoudt DN. Non-covalent synthesis of metallodendrimers. In: Vogtle F, Schalley CA, editors. Dendrimers IV: metal coordination, self-assembly, catalysis. Berlin: Springer-Verlag; 2001; p 163–199.
3. Kim C, Park E, Song CK, Koo BW. Ferrocene end-capped dendrimer: synthesis and application to CO gas sensor. Synth Met 2001; 123: 493–496.
4. Koo BW, Song CK, Kim C. CO gas sensor based on a conducting dendrimer. Sens Actuators B-Chem 2001; 77: 432–436.

5. Miller LL, Kunugi Y, Canavesi A, Rigaut S, Moorefield CN, Newkome GR. "Vapoconductivity." Sorption of Organic Vapors Causes large Increases in the Conductivity of a Dendrimer. Chem Mater 1998; 10: 1751–1754.
6. Busson P, Ortegren J, Ihre H, Gedde UW, Hult A, Andersson G, Eriksson A, Lindgren M. Preparation of mesogen-functionalized dendrimers for second-order nonlinear optics. Macromolecules 2002; 35: 1663–1671.
7. Ma D, Lupton JM, Beavington R, Burn PL, Samuel IDW. Improvement of luminescence efficiency by electrical annealing in single-layer organic light-emitting diodes based on a conjugated dendrimer. J Phys D: Appl Phys 2002; 35: 520–523.
8. Ma H, Chen B, Sassa T, Dalton LR, Jen AK. Highly efficient and thermally stable nonlinear optical dendrimer for electrooptics. J Am Chem Soc 2001; 123: 986–987.
9. Diallo MS, Balogh L, Shafagati A, Johnson JH Jr, Goddard WA III, Tomalia DA. Poly (amidoamine) dendrimers: a new class of high capacity chelating agents for Cu(II) ions. Environ Sci Technol 1999; 33: 820–824.
10. Diallo MS, Christie S, Swaminathan P, Balogh L, Shi X, Um W, Papelis C, Goddard WA III, Johnson J. H. Jr. Dendritic chelating agents. 1. Cu(II) binding to ethylene diamine core poly(amidoamine) dendrimers in aqueous solutions. Langmuir 2004; 20: 2640–2651.
11. Diallo MS, Christie S, Swaminathan P, Johnson JH Jr, Goddard WA III. Dendrimer enhanced ultrafiltration. 1. Recovery of Cu(II) from aqueous solutions using PAMAM dendrimers with ethylene diamine core and terminal NH2 groups. Environ Sci Technol 2005; 39: 1366–1377.
12. Esfand R, Tomalia DA. Poly(amidoamine) (PAMAM) dendrimers: from biomimicry to drug delivery and biomedical applications. Drug Discov Today 2001; 6: 427–436.
13. Patri AK, Kukowska-Latallo JF, Baker JR. Targeted drug delivery with dendrimers: comparison of the release kinetics of covalently conjugated drug and non-covalent drug inclusion complex. Adv Drug Deliv Rev 2005; 57: 2203–2214.
14. Patri AK, Majoros I, Baker JR Jr. Dendritic polymer macromolecular carriers for drug delivery. Curr Opin Chem Biol 2002; 6: 466–471.
15. Buhleier E, Wehner W, Vogtle F. "Cascade" and "Nonskid-Chain-like" syntheses of molecular cavity topologies. Synthesis 1978; 155–158.
16. Newkome GR, Yao ZQ, Baker GR, Gupta VK. Cascade molecules: a new approach to micelles. A [27]-Arborol. J Org Chem 1985; 50: 2003.
17. Tomalia DA, Baker H, Dewald JR, Hall M, Kallos G, Martin S, Roeck J, Ryder J, Smith P. A new class of polymers: starburst-dendritic macromolecules. Polym J 1985; 17: 117.
18. Tomalia DA, Baker H, Dewald JR, Hall M, Kallos G, Martin S, Roeck J, Ryder J, Smith P. Dendritic macromolecules: synthesis of starburst dendrimers. Macromolecules 1986; 19: 2466–2468.
19. Newkome GR, Moorefield CN, Vogtle F. Dendritic molecules: concepts, syntheses, perspectives. New York: Wiley-VCH; 1997.
20. Fréchet JMJ. Functional polymers and dendrimers: reactivity, molecular architecture, and interfacial energy. Science 1994; 263: 1710–1715.
21. Zhang W, Simanek EE. Dendrimers based on melamine. divergent and orthogonal, convergent syntheses of a G3 dendrimer. Org Lett 2000; 2: 843–845.
22. Carnahan MA, Grinstaff MW. Synthesis and characterization of polyether-ester dendrimers composed of glycerol and lactic acid. J Am Chem Soc 2001; 123: 2905–2906.

23. Kukowska-Latallo JF, Candido KA, Cao ZY, Nigavekar SS, Majoros IJ, Thomas TP, Balogh LP, Khan MK, Baker JR. Nanoparticle targeting of anticancer drug improves therapeutic response in animal model of human epithelial cancer. Cancer Res 2005; 65: 5317–5324.
24. Majoros IJ, Myc A, Thomas T, Mehta CB, Baker JR. PAMAM dendrimer-based multifunctional conjugate for cancer therapy: synthesis, characterization, and functionality. Biomacromolecules 2006; 7: 572–579.
25. Majoros IJ, Thomas TP, Mehta CB, Baker JR. Poly(amidoamine) dendrimer-based multifunctional engineered nanodevice for cancer therapy. J Med Chem 2005; 48: 5892–5899.
26. Shukla R, Thomas TP, Peters JL, Desai AM, Kukowska-Latallo J, Patri AK, Kotlyar A, Baker JR. HER2 specific tumor targeting with dendrimer conjugated anti-HER2 mAb. Bioconjug Chem 2006; 17: 1109–1115.
27. Thomas TP, Majoros IJ, Kotlyar A, Kukowska-Latallo JF, Bielinska A, Myc A, Baker JR Jr. Targeting and inhibition of cell growth by an engineered dendritic nanodevice. J Med Chem 2005; 48: 3729–3735.
28. Thomas TP, Patri AK, Myc A, Myaing MT, Ye JY, Norris TB, Baker JR Jr. *In vitro* targeting of synthesized antibody-conjugated dendrimer nanoparticles. Biomacromolecules 2004; 5: 2269–2274.
29. Man N, Cheng Y, Xu T, Ding Y, Wang X, Li Z, Chen Z, Huang G, Shi Y, Wen L. Dendrimers as potential drug carriers. Part II. Prolonged delivery of ketoprofen by *in vitro* and *in vivo* studies. Eur J Med Chem 2006; 41: 670–674.
30. Morgan MT, Carnahan MA, Immoos CE, Ribeiro AA, Finkelstein S, Lee SJ, Grinstaff MW. Dendritic molecular capsules for hydrophobic compounds. J Am Chem Soc 2003; 125: 15485–15489.
31. Morgan MT, Nakanishi Y, Kroll DJ, Griset AP, Carnahan MA, Wathier M, Oberlies NH, Manikumar G, Wani MC, Grinstaff MW. Dendrimer-encapsulated camptothecins: increased solubility, cellular uptake, and cellular retention affords enhanced anticancer activity *in vitro*. Cancer Res 2006; 66: 11913–11921.
32. Jansen JFGA, de Brabander van den Berg EMM, Meijer EW. Encapsulation of guest molecules into a dendritic box. Science 1994; 266: 1226–1229.
33. Kolhe P, Misra E, Kannana RM, Kannanb S, Lieh-Lai M. Drug complexation, *in vitro* release and cellular entry of dendrimers and hyperbranched polymers. Int J Pharm 2003; 259: 143–160.
34. Dhanikula RS, Argaw A, Bouchard J-F, Hildgen P. Methotrexate loaded polyether-copolyester dendrimers for the treatment of gliomas: enhanced efficacy and intratumoral transport capability. Mol Pharm 2008; 5: 105–116.
35. Kobayashi H, Kawamoto S, Choyke PL, Sato N, Knopp MV, Star RA, Waldmann TA, Tagaya Y, Brechbiel MW. Comparison of dendrimerbased macromolecular contrast agents for dynamic micro-magnetic resonance lymphangiography. Magn Reson Med 2003; 50: 758–766.
36. Wiener EC, Brechbiel MW, Brothers H, Magin RL, Gansow OA, Tomalia DA, Lauterbur PC. Dendrimer-based metal chelates: a new class of magnetic resonance imaging contrast agents. Magn Reson Med 1994; 31: 1–8.
37. Shi X, Ganser TR, Sun K, Balogh LP, Baker JR Jr. Characterization of crystalline dendrimer-stabilized gold nanoparticles. Nanotechnology 2006; 17: 1072–1078.

38. Shi X, Wang S, Meshinchi S, Van Antwerp M, Bi X, Lee I, Baker JR Jr. Dendrimer-entrapped gold nanoparticles as a platform for cancer-cell targeting and imaging. Small 2007; 3: 1245–1252.
39. Shi X, Wang S, Sun H, Baker JR Jr. Improved biocompatibility of surface functionalized dendrimer-entrapped gold nanoparticles. Soft Matter 2007; 3: 71–74.
40. Bielinska A, Eichman JD, Lee I, Baker JR Jr, Balogh L. Imaging {Au0-PAMAM} gold-dendrimer nanocomposites in cells. J Nanopart Res 2002; 4: 395–403.
41. Lesniak W, Bielinska AU, Sun K, Janczak KW, Shi X, Baker JR, Balogh LP. Silver/dendrimer nanocomposites as biomarkers: fabrication, characterization, *in vitro* toxicity, and intracellular detection. Nano Lett 2005; 5: 2123–2130.
42. Shi X, Lee I, Baker JR Jr. Acetylation of dendrimer-entrapped gold and silver nanoparticles. J Mater Chem 2008; 18: 586–593.
43. Shi X, Thomas TP, Myc LA, Kotlyar A, Baker JR Jr. Synthesis, characterization, and intracellular uptake of carboxylterminated poly(amidoamine) dendrimer-stabilized iron oxide nanoparticles. Phys Chem Chem Phys 2007; 9: 5712–5720.
44. Wang S, Shi X, Van Antwerp M, Cao Z, Swanson SD, Bi X, Baker JR Jr. Dendrimer-functionalized iron oxide nanoparticles for specific targeting and imaging of cancer cells. Adv Funct Mater 2007; 17: 3043–3050.
45. Shi X, Wang SH, Swanson SD, Ge S, Cao Z, Van Antwerp ME, Landmark KJ, Baker JR Jr. Dendrimer-functionalized shell-crosslinked iron oxide nanoparticles for *in vivo* magnetic resonance imaging of tumors. Adv Mater 2008; 20: 1671–1678.
46. Majoros IJ, Keszler B, Woehler S, Bull T, Baker JR. Acetylation of poly(amidoamine) dendrimers. Macromolecules 2003; 36: 5526–5529.
47. Shi X, Banyai I, Islam MT, Lesniak W, Davis DZ, Baker JR Jr, Balogh L. Generational, skeletal and substitutional diversities in generation one poly(amidoamine) dendrimers. Polymer 2005; 46: 3022–3034.
48. Shi X, Lesniak W, Islam MT, MuÑiz MC, Balogh L, Baker JR Jr. Comprehensive characterization of surface-functionalized poly(amidoamine) dendrimers with acetamide, hydroxyl, and carboxyl groups. Colloids Surf A 2006; 272: 139–150.
49. Tan NCB, Balogh L, Trevino SF, Tomalia DA, Lin JS. A small angle scattering study of dendrimer-copper sulfide nanocomposites. Polymer 1999; 40: 2537–2545.
50. Choi Y, Thomas T, Kotlyar A, Islam MT, Baker JR. Synthesis and functional evaluation of DNA-assembled polyamidoamine dendrimer clusters for cancer cell-specific targeting. Chem Biol 2005; 12: 35–43.
51. Islam MT, Majoros IJ, Baker JR. HPLC analysis of PAMAM dendrimer based multi-functional devices. J Chromatogr B Analyt Technol Biomed Life Sci 2005; 822: 21–26.
52. Islam MT, Shi XY, Balogh L, Baker JR. HPLC separation of different generations of poly (amidoamine) dendrimers modified with various terminal groups. Analyt Chem 2005; 77: 2063–2070.
53. Shi X, Bi X, Ganser TR, Hong S, Myc LA, Desai A, Banaszak Holl MM, Baker JR Jr. HPLC analysis of functionalized poly(amidoamine) dendrimers and the interaction between a folate-dendrimer conjugate and folate binding protein. Analyst 2006; 131: 842–848.
54. Kallos GJ, Tomalia DA, Hedstrand DM, Lewis S, Zhou J. Molecular-weight determination of a polyamidoamine starburst polymer by electrospray ionization mass-spectrometry. Rapid Commun Mass Spectrom 1991; 5: 383–386.

55. Schwartz BL, Rockwood AL, Smith RD, Tomalia DA, Spindler R. Detection of high molecular weight starburst dendrimers by electrospray ionization mass spectrometry. Rapid Commun Mass Spectrom 1995; 9: 1552–1555.
56. Sharma A, Desai A, Ali R, Tomalia DA. Polyacrylamide gel electrophoresis separation and detection of polyamidoamine dendrimers possessing various cores and terminal groups. J Chromatogr A 2005; 1081: 238–244.
57. Sharma A, Mohanty DK, Desai A, Ali R. A simple polyacrylamide gel electrophoresis procedure for separation of polyamidoamine dendrimers. Electrophoresis 2003; 24: 2733–2739.
58. Shi XY, Patri AK, Lesniak W, Islam MT, Zhang CX, Baker JR, Balogh LP. Analysis of poly (amidoamine)-succinamic acid dendrimers by slab-gel electrophoresis and capillary zone electrophoresis. Electrophoresis 2005; 26: 2960–2967.
59. Shi X, Banyai I, Rodriguez K, Islam MT, Lesniak W, Balogh P, Balogh LP, Baker JR. Electrophoretic mobility and molecular distribution studies of poly(amidoamine) dendrimers of defined charges. Electrophoresis 2006; 27: 1758–1767.
60. Shi XY, Banyai I, Lesniak WG, Islam MT, Orszagh I, Balogh P, Baker JR, Balogh LP. Capillary electrophoresis of polycationic poly(amidoamine) dendrimers. Electrophoresis 2005; 26: 2949–2959.
61. Shi XY, Majoros IJ, Patri AK, Bi XD, Islam MT, Desai A, Ganser TR, Baker JR. Molecular heterogeneity analysis of poly(amidoamine) dendrimer-based mono- and multifunctional nanodevices by capillary electrophoresis. Analyst 2006; 131: 374–381.
62. Shi X, Majoros IJ, Baker JR Jr. Capillary electrophoresis of poly(amidoamine) dendrimers: from simple derivatives to complex multifunctional medical nanodevices. Mol Pharmaceutics 2005; 2: 278–294.
63. Desai A, Shi X, Baker JR Jr. CE of poly(amidoamine) succinamic acid dendrimers using a poly(vinyl alcohol)-coated capillary. Electrophoresis 2008; 29: 510–515.
64. Mitnik L, Novotny M, Felten C, Buonocore S, Koutny L, Schmalzing D. Recent advances in DNA sequencing by capillary and microdevice electrophoresis. Electrophoresis 2001; 22: 4104–4117.
65. Patrick JS, Lagu AL. Review applications of capillary electrophoresis to the analysis of biotechnology-derived therapeutic proteins. Electrophoresis 2001; 22: 4179–4196.
66. Righetti PG. Capillary electrophoretic analysis of proteins and peptides of biomedical and pharmacological interest. Biopharm Drug Dispos 2001; 22: 337–351.
67. Huang QR, Dubin PL, Moorefield CN, Newkome GR. Counterion binding on charged spheres: effect of pH and ionic strength on the mobility of carboxyl-terminated dendrimers. J Phys Chem B 2000; 104: 898–904.
68. Seyrek E, Dubin PL, Newkome GR. Effect of electric field on the mobility of carboxyl-terminated dendrimers. J Phys Chem B 2004; 108: 10168–10171.
69. Pesak DJ, Moore JS. Synthesis and characterization of water-soluble dendritic macromolecules with a stiff, hydrocarbon interior. Macromolecules 1997; 30: 6467–6482.
70. Brothers HM, Piehler LT, Tomalia DA. Slab-gel and capillary electrophoretic characterization of polyamidoamine dendrimers. J Chromatogr A 1998; 814: 233–246.
71. Castagnola M, Zuppi C, Rossetti DV, Vincenzoni F, Lupi A, Vitali A, Meucci E, Messana I. Characterization of dendrimer properties by capillary electrophoresis and their use as pseudostationary phases. Electrophoresis 2002; 23: 1769–1778.

72. Ebber A, Vaher M, Peterson J, Lopp M. Application of capillary zone electrophoresis to the separation and characterization of poly(amidoamine) dendrimers with an ethylenediamine core. J Chromatogr A 2002; 949: 351–358.
73. Zhang C, Tomalia DA. Gel electrophoretic characterization of dendritic polymers In: Frechet JMJ, Tomalia DA, editors. Dendrimers and other dendritic polymers. New York: John Wiley & Sons Ltd; 2001; p 239–253.
74. Betley TA, Holl MMB, Orr BG, Swanson DR, Tomalia DA, Baker JR Jr. Tapping mode atomic force microscopy investigation of poly(amidoamine) dendrimers: effects of substrate and pH on dendrimer deformation. Langmuir 2001; 17: 2768–2773.
75. Gao JY, Dubin PL, Sato T, Morishima Y. Separation of polyelectrolytes of variable compositions by free-zone capillary electrophoresis. J Chromatogr A 1997; 766: 233–236.
76. Manning GS. Limiting laws and counterion condensation in polyelectrolyte solutions.I. Colligative properties. J Chem Phys 1969; 51: 924–933.
77. Sedlakova P, Svobodova J, Miksik I, Tomas H. Separation of poly(amidoamine) (PAMAM) dendrimer generations by dynamic coating capillary electrophoresis. J Chromatogr B Analyt Technol Biomed Life Sci 2006; 841: 135–139.
78. Carter B, Desai A, Sharma A. Use of poly(vinyl alcohol)-coated capillaries for separation of amino-terminated polyamidoamine dendrimers. Electrophoresis 2007; 28: 335–340.
79. Welch CF, Hoagland DA. The electrophoretic mobility of PPI dendrimers: do charged dendrimers behave as linear polyelectrolytes or charged spheres? Langmuir 2003; 19: 1082–1088.
80. Stockigt D, Lohmer G, Belder D. Separation and identification of basic dendrimers using capillary electrophoresis on-line coupled to a sector mass spectrometer. Rapid Commun Mass Spectrom 1996; 10: 521–526.
81. Balogh L, de Leuze-Jallouli A, Dvornic P, Kunugi Y, Blumstein A, Tomalia DA. Architectural copolymers of PAMAM dendrimers and ionic polyacetylenes. Macromolecules 1999; 32: 1036–1042.
82. Baker JR Jr, Quintana A, Piehler LT, Banazak-Holl M, Tomalia D, Raczka E. The synthesis and testing of anti-cancer therapeutic nanodevices. Biomed Microdevices 2001; 3: 61–69.
83. Quintana A, Raczka E, Piehler L, Lee I, Myc A, Majoros I, Patri AK, Thomas T, Mule J, Baker JR. Design and function of a dendrimer-based therapeutic nanodevice targeted to tumor cells through the folate receptor. Pharm Res 2002; 19: 1310–1316.
84. Patri AK, Myc A, Beals J, Thomas TP, Bander NH, Baker JR Jr. Synthesis and *in vitro* testing of J591 antibody-dendrimer conjugates for targeted prostate cancer therapy. Bioconjug Chem 2004; 15: 1174–1181.
85. Langereis S, deLussanet QG, vanGenderen MHP, Backes WH, Meijer EW. Multivalent contrast agents based on gadolinium-diethylenetriaminepentaacetic acid-terminated poly (propylene imine) dendrimers for magnetic resonance imaging. Macromolecules 2004; 37: 3084–3091.
86. Herborn CU, Barkhausen J, Paetsch I, Hunold P, Mahler M, Shamsi K, Nagel E. Coronary arteries: contrast-enhanced MR imaging with SH L 643A–experience in 12 volunteers. Radiology 2003; 229: 217–223.
87. Misselwitz B, Schmitt-Willich H, Ebert W, Frenzel T, Weinmann HJ. Pharmacokinetics of Gadomer-17, a new dendritic magnetic resonance contrast agent. Magma 2001; 12: 128–134.

88. Vetterlein K, Bergmann U, Bueche K, Walker M, Lehmann J, Linscheid MW, Scriba GKE, Hildebrand M. Comprehensive profiling of the complex dendrimeric contrast agent Gadomer using a combined approach of CE, MS, and CE–MS. Electrophoresis 2007; 28: 3088–3099.
89. Vetterlein K, Bueche K, Hildebrand M, Scriba GKE, Lehmann J. Capillary electrophoresis for the characterization of the complex dendrimeric contrast agent Gadomer. Electrophoresis 2006; 27: 2400–2412.
90. Dribek M, Le Potier I, Rodrigues A, Pallandre A, Fattal E, Taverna M. Determination of binding constants of vasoactive intestinal peptide to poly(amidoamine) dendrimers designed for drug delivery using ACE. Electrophoresis 2007; 28: 2191–2200.
91. Shcharbin D, Janicka M, Wasiak M, Palecz B, Przybyszewska M, Zaborski M, Bryszewska M. Serum albumins have five sites for binding of cationic dendrimers. Biochimica et Biophysica Acta 2007; 1774: 946–951.
92. Wang W, Guo Z, Chen Y, Liu T, Jiang L. Influence of generation 2-5 of PAMAM dendrimer on the inhibition of Tat peptide/TAR RNA binding in HIV-1 transcription. Chem Biol Drug Des 2006; 68: 314–318.
93. Delgado M, Ganea D. Inhibition of endotoxin-induced macrophage chemokine production by vasoactive intestinal peptide and pituitary adenylate cyclase-activating polypeptide *in vitro* and *in vivo*. J Immunol 2001; 167: 966–975.
94. Delgado M, Munoz-Elias EJ, Gomariz RP, Ganea D. VIP and PACAP inhibit IL-12 production in LPS-stimulated macrophages. Subsequent effect on IFNgamma synthesis by T cells. J Neuroimmunol 1999; 96: 167–181.
95. Delgado M, Munoz-Elias EJ, Gomariz RP, Ganea D. Vasoactive intestinal peptide and pituitary adenylate cyclase-activating polypeptide enhance IL-10 production by murine macrophages: *in vitro* and *in vivo* studies. J Immunol 1999; 162: 1707–1716.
96. Delgado M, Pozo D, Martinez C, Leceta J, Calvo JR, Ganea D, Gomariz RP. Vasoactive intestinal peptide and pituitary adenylate cyclase-activating polypeptide inhibit endotoxin-induced TNF-alpha production by macrophages: *in vitro* and *in vivo* studies. J Immunol 1999; 162: 2358–2367.
97. Gao XP, Noda Y, Rubinstein I, Paul S. Vasoactive intestinal peptide encapsulated in liposomes: effects on systemic arterial blood pressure. Life Sci 1994; 54: 247–252.
98. Ikezaki H, Paul S, Alkan-Onyuksel H, Patel M, Gao XP, Rubinstein I. Vasodilation elicited by liposomal VIP is unimpeded by anti-VIP antibody in hamster cheek pouch. Am J Physiol 1998; 275: 56–62.

# *From Concept to the Clinics: Development of Novel Large Molecule Cancer Therapeutics*

**Balveen Kaur[1], Gregory B. Lesinski[2], and Abhik Ray Chaudhury[3]**

[1]*Dardinger Laboratory for Neuro-oncology and Neurosciences, Department of Neurological Surgery; College of Medicine, The Ohio State University Medical Center, Columbus, Ohio*

[2]*Department of Internal medicine, Division of Hematology and Oncology, The Ohio State University Medical Center, Columbus, Ohio*

[3]*Department of Pathology, The Ohio State University Medical Center, Columbus, Ohio*

## 1 INTRODUCTION

Development of a therapeutic agent is a long and laborious process commencing with the discovery of a "druggable target" underlying the pathology of a disease. This finding is then exploited to identify molecules/compounds which can activate/antagonize this target. Rigorous testing for the specificity and potency of the identified agents then leads to the identification of a lead molecule/compound. The pharmacological formulation of the compound is then optimized for efficient and effective

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing,*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

drug delivery. Initial laboratory testing is subsequently followed by detailed animal studies confirming its *in vivo* absorption, distribution, metabolism, elimination, efficacy, and toxicology. Following these initial laboratory observations, the drug then undergoes thorough pharmacological testing to evaluate its pharmacokinetic and pharmacodynamic properties. During this time, it may return back to the laboratory for additional modifications to modulate its half life, solubility, bioavailability, and so on. The modified agent then goes through rigorous pre-clinical development in GMP certified laboratories followed by studies geared to demonstrate its bioequivalency, and efficacy relative to the initial molecule. The candidate agent is then carefully tested for toxicity in appropriate animal models in multiple species, to define a dose which can be safely administered to human patients without adverse effects. After the successful completion of all these pre-clinical studies the developer of the experimental drug typically completes an IND (investigational new drug) application to be approved for testing in human patients. After obtaining IND approval, the drug is then tested for safety and efficacy in clinical trials. Phase I clinical trials are typically dose escalation studies meant to evaluate the safety, adverse events, and maximum tolerated dose in human patients. Agents which appear to be reasonably safe then progress to Phase II and eventually Phase III studies to evaluate its efficacy prior to government approval and marketing. In this chapter, we will discuss three recently approved agents that have been through this process and are now being tested in combination with other therapies, and one experimental agent that is actively engaged in the developmental process.

## 2 BEVACIZUMAB/AVASTIN

Bevacizumab is a humanized anti-vascular endothelial growth factor (VEGF) monoclonal antibody which, in combination with chemotherapy, was the first anti-angiogenic agent to be FDA approved as first-line treatment for metastatic colorectal cancer. In this section, we will describe the discovery of VEGF, its function and role played in tumor angiogenesis, followed by a discussion of the pre-clinical and clinical studies leading to the development of Bevacizumab.

### 2.1 Tumor Angiogenesis

Angiogenesis/vasculogenesis is a highly regulated process controlling the generation of new blood vessels during development, wound healing and ovulation in adults. This process commences with increased production of angiogenic factors, and a concomitant reduction in angiostatic factors. The secreted angiogenic factors then bind to and activate specific receptors on endothelial cell receptors, initiating a signaling cascade leading to increased proliferation, migration, and invasion of endothelial and/or vascular progenitor cells. The process that transforms this nascent vasculature into mature vasculature is exquisitely regulated, turning “on” and “off” discrete sets of molecular events resulting in branching, remodeling, pruning, and recruitment of

perivascular supporting cells [1]. Dysregulation of this process manifests itself as new vessel development in several pathological states such as diabetic retinopathy, chronic inflammatory disorders, and tumor growth and metastasis. The development of a functional vascular network is considered a critical prerequisite for the progressive growth and metastasis of neoplasms. The significance of angiogenesis in tumor growth is made apparent by observations of small avascular tumors found upon, histological examination of autopsies from patients dying of unrelated events. The frequency of detecting these small dormant lesions is much higher compared to the frequency of people who present with pathological disease in similar age groups [2]. This suggests that these microscopic tumors are mostly dormant and need additional signals to grow and become lethal tumors. These observations could be interpreted to imply that dormant microscopic tumors are kept in check by the body's ability to prevent the tumors from establishing a new blood supply essential to provide them with oxygen and nutrients [2–4].

The hypothesis that anti-angiogenic strategies to attack the tumor's blood supply could be exploited as an effective anti-cancer therapy was first postulated by Dr. Folkman in 1971 [3]. Since then several angiogenic and anti-angiogenic factors involved in tumoral angiogenesis have been described, and various strategies to counter the pathways they activate are currently being investigated as potential anti-cancer therapies.

## 2.2 Vascular Endothelial Growth Factor

Vascular endothelial growth factor is one member of a family of growth factors which includes VEGF-A, VEGF-B, VEGF-C, VEGF-D, VEGF-E, and Placenta like growth factor (Fig. 1). VEGF-A was the first growth factor found to have endothelial cell specific mitogenic properties. It was purified to homogeneity from media conditioned by bovine pituitary follicular cells as a growth factor with remarkable specificity for endothelial cells [5]. It had also been described as a vascular permeability factor (VPF) which facilitated vascular leakage into the skin [6]. Subsequent isolation and cloning of VEGF and VPF revealed them to be identical [7,8]. VEGF is a homodimeric heparin binding secreted protein which functions as a potent stimulator of growth and migration of vascular endothelial cells derived from arteries, veins and even lymphatics (reviewed in Reference [9]) (Fig. 1). VEGF-A binds and activates VEGF receptor-1 (VEGFR-1/Flt-1) and VEGF receptor 2 (VEGFR-2/KDR) expressed primarily on endothelial cells. Both VEGFR-1 and VEGFR-2 are membrane bound tyrosine kinase receptors. VEGF binding to VEGFR-2 initiates receptor dimerization and activates a signaling cascade resulting in increased endothelial cell proliferation, while its binding to VEGFR-1 results in endothelial cell migration and differentiation (reviewed in Reference [10]). VEGF-A plays a very significant role in embryonic vessel development wherein targeted genetic disruption of even a single allele of VEGF results in abnormal blood vessel development and lethality observed in mice [11,12]. Decades of research have uncovered the extremely significant role played by VEGF in the regulation of both normal and pathological angiogenesis.

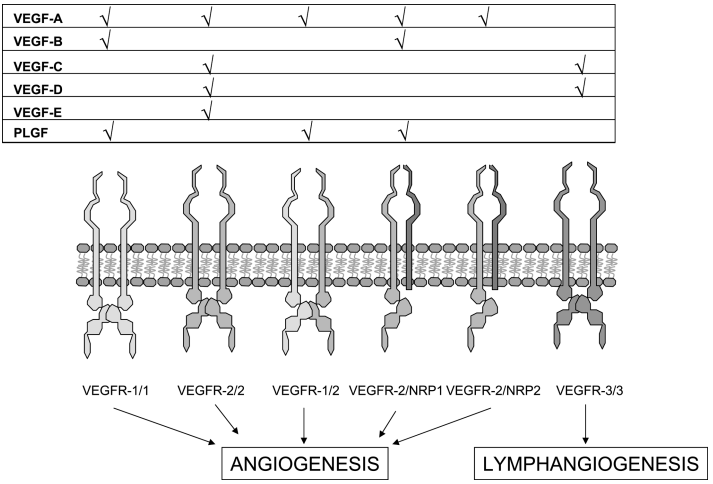


***FIGURE 1*** *VEGF family members (VEGFA, VEGFB, VEGFC, VEGFD, VEGFE, and PLGF) bind differentially to the VEGF receptors VEGR1, VEGFR2, VEGFR3, NRP1, or NRP2 resulting in receptor dimerization and activation. The non-signaling NRP1 and NRP2 function by hetrodimerizing with VEGFR-2 leading to its activation. Lymphangiogenesis is initiated by VEGFC and VEGFD binding to VEGFR3, leading to its homodimerization and subsequent activation.*

## 2.3 VEGF Regulation

VEGF expression in human cancers has been shown to correlate with increased tumor vascularity and in many instances a worse prognosis. Over-expression of VEGF in human tumors promotes neovascularization and malignant growth. VEGF gene expression is rapidly induced by hypoxia and its increased expression in solid tumors is primarily a consequence of poor tissue oxygenation due to aberrant tumor vasculature. This is perhaps most apparent in pathological specimens derived from glioblastoma multiforme (GBM) grade IV malignant astrocytomas, in which areas of hypoxia/necrotic tissue are thought to be the result of poor tissue oxygenation [13] (Fig. 2). The resulting tissue hypoxia results in stabilization and activation of hypoxia inducible factor (HIF1). HIF1 is a hetrodimeric transcription factor which activates transcription of numerous hypoxia responsive genes including VEGF [10]. Apart from increased gene expression, hypoxia responsive 3′UTR in VEGF mRNA can also function to increase its stability. Consistent with these observations, increased expression of VEGF is often observed adjacent to hypoxic necrotic areas in tissue specimens (Fig. 3). Increased VEGF in GBM often results in the appearance of hyperplastic or rapidly dividing endothelial and pericyte cell aggregates called ‘glomeruloid bodies’. These are often located in proximity to hypoxic necrotic centers within the tissue [14]. Figure 3 shows an example of a GBM specimen displaying hypoxic necrotic tissue cores and glomeruloid bodies. Hypoxia independent activation of signaling pathways resulting from mutations in oncogene and/or tumor suppressor genes can also trigger increased levels of VEGF and subsequent

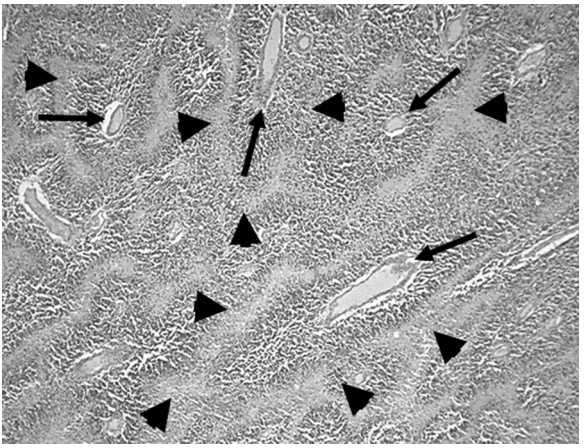

***FIGURE 2*** *Glioblastoma multiforme with geographic necrosis (pseudopalisading of tumor cells surrounding linear necrotic zones). Hematoxylene and Eosin images of glioblastoma specimen at 40X magnification. Areas distant from tumor vasculature (arrows) appear to be devoid of tumor cells and are referred to as hypoxic necrotic areas (arrow heads). These hypoxic necrotic areas appear to form a ring around the tumor vasculature and represent areas with poor perfusion.*

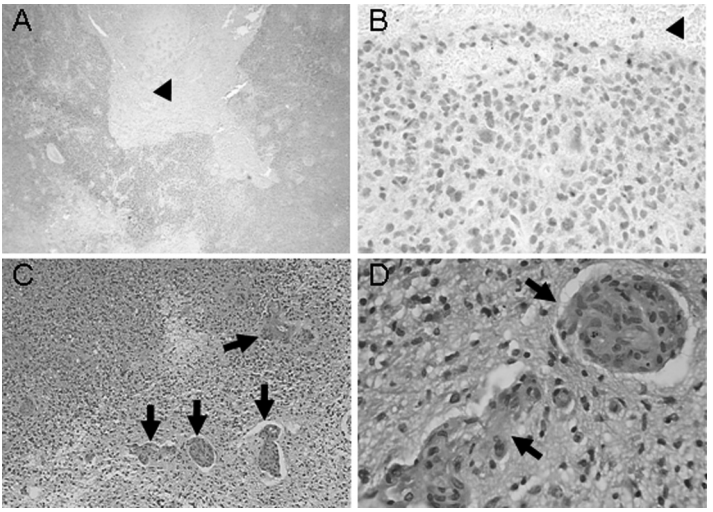


***FIGURE 3*** *Increased VEGF in hypoxic tumor areas. (A and B) Sections derived from human glioblastoma tissue showing tumor cells adjacent to a necrotic zone (arrows) showing positive reaction with VEGF immunostain (brown staining) at low (A) and high magnification (B). The positive cells display brown cytoplasmic staining. Note the increased expression of VEGF adjacent to hypoxic necrotic areas (arrow head) in the tumor. (C and D) Hematoxin and Eosin staining of glioblastoma multiforme tissue showing microvascular proliferation at low (C) and high magnification (D). Presence of tufted aggregates of proliferating endothelial and pericytes called Glomeuloid bodies is indicated by arrows.*

angiogenesis [15]. Hence VEGF dysregulation appears to be a common factor in many different malignancies.

## 2.4 Pre-Clinical Studies Targeting Tumor VEGF as a Therapeutic Target

Research over the last decade has established VEGF as a key endothelial cell specific mitogen which is essential for pathological and normal vascularization. Anti-VEGF murine monoclonal antibodies were initially generated to identify biologically relevant VEGFR binding regions of human VEGF [16]. One of these VEGF binding antibodies (A.4.6.1) abrogated the ability of VEGF to bind to and activate both VEGFR-1 and VEGFR-2 on endothelial cells. This murine monoclonal antibody was then evaluated for its ability to suppress angiogenesis and tumor growth in athymic mice bearing subcutaneous rhabdomyosarcoma, and glioblastoma cells. The results indicated that treatment of tumor bearing mice with A.4.6.1 inhibited tumor growth with a concurrent decrease in blood vessel density [17]. Ensuing studies using numerous human cancer cell lines confirmed the anti-tumorigenic potential of targeting VEGF using A.4.6.1, and other VEGF antibodies, soluble dominant negative VEGFR-2 receptors, small molecule inhibitors and anti-VEGFR-2 antibodies *in vivo* (Table 1) [18].

## 2.5 Pharmacological Formulation of Bevacizumab

Interfering with VEGF signaling by delivery of the murine monoclonal antibody A.4.6.1 was found to have a therapeutic advantage for anti-cancer therapy in many different animal models (Table 1). However a major limitation in the use of murine monoclonal antibodies in human patients is the anti-immunoglobulin response initiated against the antibody. Humanization of monoclonal antibodies is a powerful technique used to develop antibodies that closely resemble the human IgG sequence. This technique is an effective means to avoid an antiglobulin response against itself upon administration to human patients. Antibodies are composed of two identical heavy (H) polypeptide chains and two identical light (L) chains, that are covalently bonded to each other by interchain cysteine–cysteine disulfide (S-S) linkages (Fig. 4). Humanization of the murine A.4.6.1 was achieved by transferring six complementary determining regions (CDR) from the murine A.4.6.1 into a consensus human framework, consisting of the human heavy chain subgroup III and the light chain subgroup $\kappa$I [19]. However, simple interchange of the CDRs from murine to human antibody framework reduced the VEGF binding affinity of the humanized antibody by more than 1350-fold as compared to the murine A.4.6.1. Seven non-CDR framework residues were then mutagenized to achieve the recombinant humanized monoclonal antibody: Bevacizumab/Avastin (RhuMAb). Bevacizumab consisted of a sequence that was 93% homologous to human immunoglobulin (the entire framework) and 7% homologous to murine immunoglobulin (the region in A.4.6.1 that binds specifically to human VEGF). Analysis of the binding affinity kinetics of Bevacizumab and A.4.6.1 revealed that both antibodies had very similar binding affinity to human

**TABLE 1 Preclinical Studies Demonstrating Anti-Tumor Efficacy by Targeting Tumor VEGF**

| | Preclinical Tumor Model Tested | Mechanism of VEGF Inhibition | Noteworthy Observations (Ref.) |
|---|---|---|---|
| 1 | Human rhabdomyosarcoma (A673), glioblastoma multiforme (G55), and leiomyosarcom (SK-LMS-1) cells implanted as subcutaneous xenograft tumors in athymic nude mice | Anti-human VEGF mouse monoclonal antibody (A.4.6.1) | Inhibition of tumor growth and vascularity *in vivo* [17] |
| 2 | Colon carcinoma (HM7) cells grown as experimental hepatic metastases in athymic nude mice | Anti-human VEGF mouse monoclonal antibody (A.4.6.1) | Reduction in the number and size of hepatic metstases [160] |
| 3 | Human prostate cancer (DU145) cells grown as subcutaneous xenografts in athymic nude mice | Anti-human VEGF mouse monoclonal antibody (A.4.6.1) | Tumor growth and formation of distant metastases in lungs was suppressed [161] |
| 4 | Human fibrosarcom (HT1080) cells | Anti-human VEGF mouse monoclonal antibody (MV303) | Inhibition of primary and metastatic tumor growth in nude mice [162] |
| 5 | Human glioblastoma (C6) cells grown as subcutaneous tumors in athymic nude mice | Retrovirus encoding dominant negative mutant of VEGFR/Flk-1 (Flk-1TM) | Suppression of tumor xenograft growth [163] |
| 6 | Epidermal carcinoma (A431), colon carcinoma (Ls174T, and HT-29), prostate carcinoma (PC-3, DU-145, and CWR-22) cells grown as subcutaneous tumors in mice | PTK787, a small molecule inhibitor of VEGF receptor tyrosine kinase | Suppression of tumor growth and development of distant metastases [164] |
| 7 | Breast (MDA-MB-231) lung (Calu-3) prostate (PC-3), and ovarian (SKOV-3) tumors grown as xenografts in nude mice | ZD4190, a small molecule inhibitor of VEGF receptor tyrosine kinase | Significant anti-tumor efficacy observed in prostate and lung cancer cell line derived tumors in mice [165] |
| 8 | Human colon (SW620), lung (Calu-6), prostate (PC3), breast (MDA-MB-231), and ovary (SKOV-3) cells grown as subcutaneous tumors in mice | AZD2171, a small molecule inhibitor of VEGF receptor tyrosine kinase | Siginificant inhibition of tumor growth observed in all tumor models [166] |
| 9 | Human epidermoid (A431), renal cell carcinoma (SK-RC-29), pancreatic (BxPC-3), or glioblastoma (GBM-18) tumors grown as subcutaneous tumors in nude mice, and Lewis lung, B16 melanoma, and 4T1 breast tumors established in C57BL/6 or BALB/c mice | Anti-VEGFR/Flt-1 mouse monoclonal antibody (DC101) | Tumor suppression observed in several mouse and human tumor models [167] |

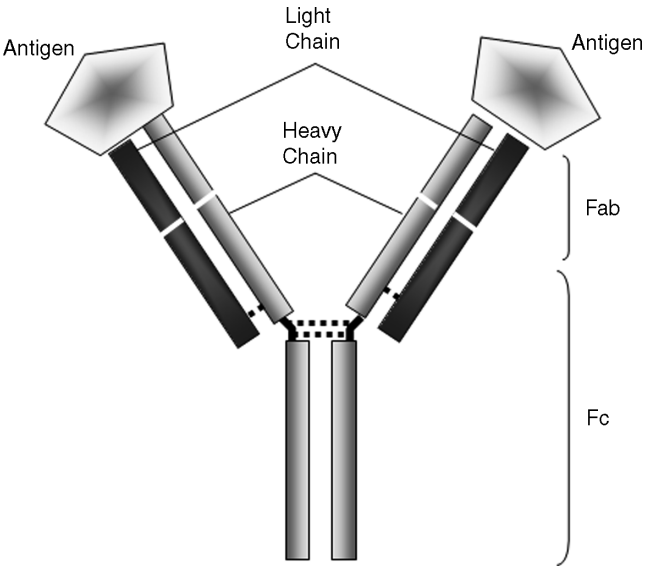


***FIGURE 4*** *Structure of an antibody: each antibody contains two light chains (red) and two heavy chains (green), held together by disulfide linkages (dotted lines). Each chain has a variable region and a constant region. The variable regions together constitute the antigen recognition domain within the antibody. Cleavage by Papain at pH 7.0, results in the fragmentation of the antibody into two Fab fragments and one Fc fragment.*

VEGF. Additionally the two antibodies were indistinguishable in their ability to antagonize VEGF bioactivity both in endothelial cell cultures (ED50s against endothelial cell proliferation were $50 \pm 5$ and $48 \pm 8$ ng/mL respectively) *in vitro* and in their ability to antagonize tumor growth *in vivo* in mice [19]. Numerous subsequent studies further confirmed the anti-tumor efficacy of the humanized anti-VEGF antibody alone and in combination with other agents in mouse models of tumorigenesis (Table 2). This engineering increased the half-life of Bevacizumab while maintaining its high specificity for VEGF.

For clinical studies, 100 mg of RhuMAb is provided in glass vials containing histidine, trehalose, polysorbate 20, and sterile water ready for injection. The vials contain no preservative and are intended for single use. The vial is diluted into 5% dextrose (D5W) prior to infusion.

## 2.6 Safety Studies

Detailed pre-clinical safety evaluation of Bevacizumab consisted of examining tissue reactivity, investigation of its hemolytic potential and blood compatibility studies, and multiple dose toxicity studies in adult cynomologus monkeys. Absence of non-target tissue binding was confirmed by immunohistochemistry of tissue derived from human donors, cynomologous monkey and rabbit tissue. No tissue cross reactivity of

**TABLE 2 Summary of Preclinical Studies for Anti-Tumor Efficacy of the Humanized Anti-VEGF Antibody (rhuαVEGF)/BEVACIZUMAB**

| Single Agent | | | | |
|---|---|---|---|---|
| Agent | Tumor Type | Cell Line (Graft Location/Mouse Strain) | Dosage | Noteworthy Observation (Ref.) |
| rhuαVEGF | Androgen independent prostate cancer | CWR223 (s.c/nude athymic BALB/c mice) | 5 mg/Kg, given i.p. twice weekly for 4 weeks | >85% growth inhibition of established tumors with reduced MVD [168] |
| rhuαVEGF | Neuroblastoma | SK-N-AS<br>IMR-32<br>SH-SY5Y (s.c immunodeficient mice) | 5 mg/Kg, given i.p. twice weekly once tumor volume was >0.3 mL | Significant reduction in NB tumor growth *in vivo* with 30–63% reduction of angiogenesis [169] |
| Bevacizumab | Non-small cell lung cancer cells | SCH1299 (athymic nude BALB/c mice) | 100 ug/mouse i.p. twice weekly | Significant reduction in tumor growth [170] |

| Combination with Chemotherapy | | | | |
|---|---|---|---|---|
| Agent | | Tumor Type | Cell Line | Noteworthy Observation (Ref.) |
| rhuαVEGF + palcitaxel | Androgen independent prostate cancer | CWR223 (s.c/nude athymic BALB/c mice) | rhuαVEGF: 5 mg/Kg, i.p. twice weekly for 4 weeks, and palcitaxel: 6.25 mg/Kg s.c 5 times/week for 3 weeks | 98% growth inhibition of established tumors [168] |
| Bevacizumab + ad-mda7 | Non-small cell lung cancer cells | SCH1299 (athymic nude BALB/c mice) | Bevacizumab: 100 ug/mouse i.p. twice weekly for 4 weeks<br>Ad-mda7: $5 \times 10^9$ virus particles/dose intratumorally twice weekly for 4 weeks | Significant reduction in tumor growth, compared to ad-mda7 treated alone or Bevacizumab treated alone animals [170] |

(*Continued*)

**TABLE 2 (*Continued*)**

| Combination with Chemotherapy | | | | |
|---|---|---|---|---|
| Agent | | Tumor Type | Cell Line | Noteworthy Observation (Ref.) |
| Bevacizumab + Ad5/3-Δ24 | Metastatic renal cancer | ACHN cells injected i.p. (SCID mice) | Bevacizumab: 50 ug/mouse once weekly for 5 weeks<br>Ad5/3-Δ24: 108 virus particles once weekly for 3 weeks | No toxicity but no enhancement observed with co-treatment compared to treatment with the oncolytic adenovirus alone [171] |
| Bevacizumab + yeast derived beta-glucan | Ovarian carcinoma | SKOV-3 injected s.c (SCID mice) | Bevacizumab: 0.2 mg every third day, Beta-glucan: 1.2 mg every third day, | Mice treated with combination had significantly smaller tumors that Bevacizumab treated alone mice, and 86% of mice treated with combination were long term survivors compared to 43% Bevacizumab treated alone mice [172] |

non-target tissue was observed with low (10 μg/mL) or high (400 μg/mL) doses of Bevacizumab [20]. The compatibility of Bevacizumab with human serum was further tested by incubating human serum with the antibody. No hemolysis, precipitation or coagulation of human or monkey serum was observed indicating its serum/plasma compatibility [20]. Bevacizumab (0, 2, 10, and 50 mg/Kg) was administered twice weekly intravenously to cynomologus monkeys for 4 or 13 weeks to determine its toxicity. Animals were monitored for weight loss, appetite, blood pressure measurements, electrocardiograms, physical activity, ophthalmic exams, clinical pathology (hematology, clinical chemistry, and urinalysis). No Bevacizumab related clinical signs of toxicity were observed upon administration of the recombinant antibody. Bevacizumab treatment also led to inhibition of physiological angiogenesis into the growth plate and ovary and was reversible as shown by the resumption of angiogenesis when the treatment was stopped. This is consistent with its known function and was not considered an adverse event [20].

## 2.7 Clinical Studies with Bevacizumab

Based on the safety profile of Bevacizumab, Genetech initiated phase I clinical trials with Bevacizumab as a single agent and in combination with chemotherapy. In the first phase I study, 25 patients with metastatic cancer, were given 0.1–10 mg/Kg Bevacizumab on days 0, 28, 35, and 42 by a 90 min. intravenous (i.v.) infusion. Overall the infusions were well tolerated with no grade III or grade IV treatment related adverse events [21]. Following these results several phase II studies with Bevacizumab given as a single agent alone or in combination with chemotherapy were initiated [18]. A randomized phase II study comparing Bevacizumab and fluorouracil (F)/leucovorin (L) with F/L treatment alone for patients with previously untreated metastatic colorectal cancer was initiated. One hundred and four patients were randomly assigned to be treated with either F/L alone (36 patients) or F/L + high dose Bevacizumab (10 mg/Kg) (33 patients) or F/L + low dose Bevacizumab (5 mg/Kg) [22]. Treatment with Bevacizumab in combination with FU/LV showed an improvement in time to progression and overall survival for both the doses. Interestingly patients treated with low dose Bevacizumab showed a better response than patients treated with the higher dose [22]. Fatal thrombosis was observed in one patient, and the lower dose was recommended for further studies in colorectal cancer patients. In a subsequent phase III trial, 813 patients were randomly segregated to receive irrinotecan, fluorouracil and leucovorin (IFL) + Bevacizumab (IFL + Bevacizumab: 402 patients) or IFL + placebo (411 patients). The primary end point was to evaluate changes in overall survival and secondary end points was to determine Bevacizumab treatment induced progression free survival, response rate, duration of response, safety and quality of life. The addition of Bevacizumab to IFL resulted in a statistically significant and clinically meaningful improvement in survival of patients with metastatic colorectal cancer. In 2004, administration of Bevacizumab in combination with F.L based therapies was approved by the US Food and drug Administration (FDA) for the treatment of patients with metastatic colorectal cancer. Administration of Bevacizumab in combination with other chemotherapeutic

drugs is being tested for efficacy as a front line treatment for a variety of other tumor types. A recent randomized phase III study investigating Bevacizumab in combination with interferon-alpha (IFN-α) for patients with metastatic clear cell renal cell carcinoma showed improvements in progression free survival and overall response rates [23].

### 2.8 Adverse Events Related to Bevacizumab Treatment

Hypertension was the most frequent Bevacizumab-related adverse event that was noted in the studies testing combined therapy with Bevacizumab + IFL in colorectal cancer patients. The underlying mechanism of hypertension is thought to be related to reduction in the levels of nitric oxide production as a consequence of VEGF blockade [24]. Other complications associated with Bevacizumab treatment include GI perforation, complications in wound healing, hemorrhage, thromboembolic events, proteinurea, and congestive heart failure. Careful monitoring and discontinuation of the drug is advised in such cases. More serious and even fatal pulmonary hemorrhage was noted in patients with NSCLC of squamous cell histology who were treated with Bevacizumab [25]. Due to its impact on wound healing, Bevacizumab therapy also cannot be initiated for at least 28 days before and after a surgery. Since angiogenesis during embryogenesis is essential for normal fetal development Bevacizumab is rated pregnancy category C. Due to its long half life, patients are advised to avoid getting pregnant soon after discontinuing the drug.

## 3 PROTEASOME INHIBITORS

### 3.1 The Proteasome as a Therapeutic Target

The ubiquitin proteasome pathway regulates intracellular protein turnover and its function is critical to cellular homeostasis [26]. The 26s proteasome is a large multienzyme complex that functions to degrade or process intracellular proteins, including those involved in cell cycle progression and apoptosis [27,28]. The proteasome is appreciated as playing a critical role in degrading intracellular proteins through both ubiquitin-dependent and ubiquitin-independent proteolysis [26,29]. The proteasome consists of a 20S catalytic cylindrical complex capped at both ends by 19S regulatory subunits and is typically present within both the nucleus and cytoplasm of eukaryotic cells [30,31]. Although other intracellular proteolytic processes are mediated by the immunoproteasome, the 26S proteasome mediates a substantial degree of protein turnover. The proteasome is central to the ubiquitin-proteasome pathway for ordered, temporal degradation of cyclins, CDK inhibitors (p21 and p27) and other proteins required for cell cycle progression in malignant cells as well as normal cells [32]. Because of its actions on cell cycle proteins, the proteasome represents a rational target for targeted therapy against neoplastic cells. Furthermore, the proteasome is known to play a role in regulation of apoptosis and/or cell survival through its effects on key mediators such as caspases, the Bcl2-family of proteins and nuclear factor $\kappa$B (NF-$\kappa$B) [32–34]. Based on these observations, a class of anti-tumor agents, termed

'proteasome inhibitors' has emerged as a novel therapeutic option for the treatment of malignancies.

## 3.2 The Discovery of Bortezomib as a Proteasome Inhibitor

To date, proteasome inhibitors have been separated into five classes based on their chemical structure: peptide aldehydes, peptide vinyl sulfones, peptide boronates, peptide epoxiketones, and β-lactones. The first agents to be identified as inhibitors of the proteasome were the natural inhibitor 'lactacystin' and synthetic peptide aldehydes [27]. However, only two compounds (bortezomib/VELCADE/PS-341 and PS-519/MLN519) are currently in clinical development [27]. The latter compound, PS-519/MLN519, has been tested in healthy volunteers to determine safety, tolerability, and pharmacodynamic profiles [35]. Based on its protective ability in animal models of cerebral ischemia, it is thought that this compound may be an effective neuroprotective agent [36]. The former compound, bortezomib has been primarily employed in the treatment of malignancy and will be the focus of the present discussion. Bortezomib is a small molecule proteasome inhibitor (Fig. 5) developed by Millennium Pharmaceuticals, Inc., as a novel agent to treat malignancy. This drug is currently approved by the US FDA for the treatment of patients with multiple myeloma [37]. The rationale for developing bortezomib was that substitution of the aldehyde with boronic acid would create a compound that could form covalent and reversible complexes with the proteasome, resulting in increased potency and selectivity compared with their corresponding aldehydes [27]. As a result, a series of dipeptidyl boronic-acid proteasome inhibitors was developed and screened for activity against the standard National Cancer Institute (NCI) panel of 60 human tumor cell lines [38]. In these assays, bortezomib demonstrated effective *in vitro* cytotoxic activity against the tumor cell lines and was subsequently pursued in further studies for clinical development. Through its ability to inhibit the proteasome, bortezomib affects multiple signaling pathways.

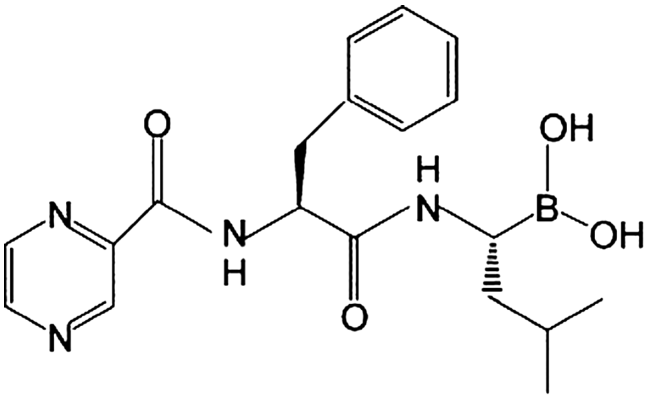


***FIGURE 5*** *Structure of Bortezomib (Reproduced from Ref. [173] with copy right Permission to GL).*

### 3.3 Pharmacological Formulation of Bortezomib

Bortezomib is a modified dipeptidyl boronic acid formulated as an anti-neoplastic agent for intravenous injection. The chemical name for bortezomib, the monomeric boronic acid, is [(1R)-3-methyl-1-[(2S)-1-oxo-3-phenyl-2-[(pyrazinylcarbonyl) amino]propyl]amino]butyl] boronic acid. The product is provided as a mannitol boronic ester which, in reconstituted form, consists of the mannitol ester in equilibrium with its hydrolysis product, the monomeric boronic acid. The drug substance exists in its cyclic anhydride from as a trimeric boroxine. Typically the compound is provided as a sterile lyophilized powder (Velcade [bortezomib] for injection (from the package insert)).

### 3.4 Bortezomib Pre-Clinical Studies and Mechanism of Action

Bortezomib-mediated reductions in the degradation of proteins involved in cell cycle progression and transcription factor activation leads to arrested growth of malignant cells via multiple mechanisms [33]. The first evidence that proteasome inhibition was a rational approach to cancer treatment arose from *in vitro* studies showing that proteasome inhibitors induced apoptosis in leukemic cell lines which were resistant to chemotherapy and radiation and *in vivo* studies using an animal model of Non-Hodgkin's Lymphoma [39–42]. Importantly, the apoptosis observed in these initial experiments appeared to be limited to transformed cells.

The anti-neoplastic effect of bortezomib likely involves several distinct mechanisms, including inhibition of cell growth and survival pathways, induction of apoptosis, and inhibited expression of genes that control cellular adhesion, migration and angiogenesis. Thus, the mechanisms by which bortezomib elicits its anti-tumor activity may vary among tumor types, and the extent to which each affected pathway is critical to the inhibition of tumor growth could also differ. Bortezomib has a novel pattern of cytotoxicity against the National Cancer Institute (NCI) panel of 60 human tumor cell lines [33]. In addition, bortezomib has cytotoxic activity that is preferential to neoplastic cells in a variety of xenograft tumor models [32,43,44].

One particular pathway by which bortezomib is thought to exert is anti-tumor activity is via its effects on NF-$\kappa$B [45]. NF-$\kappa$B activity maintains cell viability through the transcription of factors that inhibit the apoptotic process induced by environmental stress and cytotoxic agents [46]. The proteasome promotes the activation of the NF-$\kappa$B signaling pathway via degradation of its inhibitory protein I$\kappa$B. Stabilization of I$\kappa$B (e.g., by proteasome inhibition with bortezomib) and blockade of NF-$\kappa$B activity renders cells more susceptible to apoptosis [38]. Reduced NF-$\kappa$B activity also inhibits the expression of cell adhesion molecules such as ICAM-1 and VCAM-1 that are important for tumor metastasis and angiogenesis *in vivo* [27].

Despite the broad effects downstream of NF-$\kappa$B, bortezomib and other proteasome inhibitors work through a number of mechanisms separate from its inhibition of NF-$\kappa$B. For example, in the presence of bortezomib, cyclin, and CDK inhibitor

proteins required for cell cycle progression are stabilized and lead to cell cycle arrest [32]. Proteasome inhibition with bortezomib may also inhibit tumor progression via the inhibited release of pro-angiogenic factors such as VEGF and basic fibroblast growth factor (bFGF) [47]. Notably, bortezomib induces apoptosis in cells that over express bcl-2, a genetic trait that confers unregulated growth and resistance to conventional chemotherapeutics [34] and is believed to induce apoptosis by a p53-independent mechanism [28,32,38,48]. Furthermore, bortezomib has been reported to activate both caspase-8 and caspase-9 suggesting involvement of both intrinsic and extrinsic apoptotic pathways. Notably, the mechanism of bortezomib-induced apoptosis in transformed cells appears to be highly variable, across tumor cells of different origin. For example, the differential induction of the pro-apoptotic Noxa protein has been shown to occur in melanoma cells upon exposure to bortezomib [49,50]. Most recently, proteasome inhibition has been reported to induce endoplasmic reticulum stress and apoptosis through the unfolded protein response in human pancreatic cancer cell lines [51–53]. Together these data highlight the wide spectrum of pro-apoptotic events that are induced in response to proteasome inhibition.

### 3.5 Clinical Experience with Bortezomib

As an anti-cancer agent, bortezomib has shown particular promise in the setting of multiple myeloma. It has recently been approved by the FDA for patients with multiple myeloma who have received at least two prior therapies and have demonstrated disease progression on their last therapy [54]. Bortezomib is thought to be efficacious in multiple myeloma via its inhibition of NF-$\kappa$B activation, its attenuation of interleukin-6 (IL-6)-mediated cell growth, a direct apoptotic effect, and possibly anti-angiogenic and other effects [32]. To date, only one trial has tested bortezomib as a single agent in the treatment of metastatic melanoma and no complete or partial responses were observed [55]. However, bortezomib has shown clinical activity in other solid tumors, including renal cell carcinoma, non-small cell lung carcinoma, androgen-independent prostate carcinoma, mantle cell lymphoma, and follicular non-Hodgkin's lymphoma [54,56,57].

An acceptable overall therapeutic index has been documented with bortezomib however, there remains some risk of toxicity from this agent [37]. The most common dose-limiting side effects of bortezomib observed in phase I studies were fatigue, malaise, gastrointestinal effects such as diarrhea, nausea and loss of appetite, fluid retention, hypokalemia, hyponatremia, orthostasis, thrombocytopenia and sensory neuropathy [58]. Subsequent clinical studies in patients with multiple myeloma have demonstrated that toxicity was observed in approximately 10% of patients receiving single agent bortezomib and that a majority of the dose-limiting toxicity associated with this drug are likely due to thrombocytopenia and neuropathy [58]. More detailed information on the pharmacokinetics and pharmacodynamics of bortezomib are beyond the scope of this review, but is readily available from the product insert or the United States Food and Drug Administration [37,58].

**TABLE 3 Clinical Trials Employing Bortezomib in Combination with Other Agents**

| Other Agents Combined with Bortezomib | Tumor Type | Ref. |
|---|---|---|
| Carboplatin | Ovarian or primary peritoneal | [59] |
| Thalidomide and dexamethasone | Multiple myeloma | [173] |
| Melphalan | Multiple myeloma | [62] |
| Pegylated liposomal doxorubicin | Multiple myeloma | [63] |
| Dexamethasone | Multiple myeloma | [65,66] |
| Paclitaxel and carboplatin | Advanced malignancies | [67] |
| Melphalan and prednisone | Multiple myeloma | [68] |
| Docetaxel | Advanced solid tumors | [69] |
| Pegylated liposomal doxorubicin | Multiple myeloma | [70,71] |
| Gemcitabine | Advanced solid tumors | [73] |
| Irinotecan | Advanced solid tumors | [73] |
| Gemcitabine and cisplatin | Advanced solid tumors | [74] |
| Thalidomide and dexamethasone | Multiple myeloma | [75] |

## 3.6 Future Studies: Combination Therapy with Bortezomib

The wide range of pro-apoptotic properties and acceptable toxicity profile of bortezomib in patients suggests that this agent may be best utilized in combination with other anti-cancer therapeutics. Indeed bortezomib has been administered clinically in combination with a variety of other therapeutic agents (Table 3). A number of pre-clinical and clinical studies have demonstrated that bortezomib is an effective chemo-sensitizing agent [59–76]. Furthermore, pre-clinical work has demonstrated that combined treatment of malignant cells with bortezomib and receptor tyrosine kinase inhibitors, such as sorafenib, leads to enhanced apoptosis *in vitro* [77]. Proteasome inhibition has been recognized as an effective means for sensitizing tumor cells to TRAIL-mediated apoptosis [78] and also appears to be a rational means of enhancing the anti-tumor activity of immune enhancing cytokines. Khan *et al.* demonstrated enhanced anti-tumor activity of bortezomib when administered together with interleukin-12 or interleukin-2 in murine models of neuroblastoma [79]. Other studies have demonstrated that combined administration of bortezomib and interferon-alpha can promote apoptosis and inhibit angiogenesis in human bladder cancer cell lines [78] and induce synergistic apoptosis in human melanoma cells [80]. Recent studies in B16 melanoma and multiple myeloma models also indicated that bortezomib may be an effective means to sensitize tumor cells to the lytic effects of dendritic cell (DC)-activated immune effector cells [81,82]. The effects of proteasome inhibition on immune effector cells therefore remains an important area of investigation, as other reports have shown that proteasome inhibition can also lead to apoptosis in DC and inhibit their stimulatory capacity [83–85] and suppress $CD4^+$ T cell function [86]. Finally, second and third generation proteasome inhibitors are currently under evaluation and the available data suggests

that these compounds will likely have enhanced specificity and anti-tumor properties [87,88]. It will therefore be critical to evaluate the toxicity profile of these new generation inhibitors as single agents and when given in combination with other anti-cancer and immune modulating therapies.

## 4 INTERFERON-ALPHA

### 4.1 History of Interferon-Alpha

The biological phenomenon known as 'interference' was first described by Isaacs and Lindenmann in 1957 [89]. This phenomenon was defined as protection against viral infection in cells that had previously been exposed to another virus. This concept was examined because of its public health implications during a time in which there was an ongoing search for 'anti-viral penicillin'. In their research examining the molecular mechanisms of this phenomenon, Isaacs and Lindenmann discovered interferon (IFN) [89,90]. Crude protein fractions, consisting of less than one percent of IFN and unsuccessful attempts to purify and produce IFN in large quantities was a major limitation in the field [91]. It was not until the 1960s when IFN was recognized as a family of distinct structurally related proteins rather than a single molecule [92]. In this same decade, additional interest in the biological properties of IFN led to the discovery of its cytostatic, anti-tumor and immunomodulatory properties [92,93]. IFN research was later advanced by the work leading to its purification to homogeneity in solution. This facilitated production of the cytokine in sufficient amounts to study its physicochemical properties [94] and later for the cloning of the first IFN gene [95]. Competition between the pharmaceutical industry and academic research eventually led to the production of IFN for use in patients, and this substance was one of the first natural proteins to become available for the clinical treatment of malignancies [96]. Human IFNs are a heterogenous family of proteins which function endogenously to protect the host from viral infection. Type I IFN genes are clustered on human chromosome 9p21 [97] and mediate their biological effects through a common receptor, IFN-αβR, which is ubiquitously expressed throughout the body. The IFN gene family consists of 13 IFN-α members, a single member each of IFN-β, IFN-$\kappa$, IFN-ω and IFN-ε [98–100]. Despite the large number of endogenously produced IFNs, IFN-α is the subtype that is used most frequently in clinical medicine and therefore will remain the focus of the present chapter.

### 4.2 Pharmacologic Formulation of IFN-α

Interferon-alpha is available from two commercial sources. Either Roferon-A (Interferon-alfa-2a, recombinant; Roche, Inc.) or INTRON-A (Interferon-alpha-2b; Schering-Plough, Inc.). Both formulations are manufactured by recombinant DNA technology that employs a genetically engineered *Escherichia coli* bacterium containing DNA that encodes for either the human Inteterferon-alpha-2a or interferon-alpha-2b protein (165 amino acids; molecular weight = 19,000 Daltons). Fermentation is carried out in a defined nutrient medium containing the antibiotic tetracycline

hydrochloride, however the presence of the antibiotic is undetectable in the final, purified product. Roferon-A is supplied both as an injectable solution in a vial or a pre-filled syringe. Roferon-A formulations typically contain Interferon-alpha-2a ranging from 3–36 million International Units (IU), and contain polysorbate 80, benzyl alcohol, and ammonium acetate as a preservative. Intron-A is supplied as a sterile lyophilized powder for reconstitution and injection, which is formulated with glycine, sodium phosphate dibasic, sodium phosphate monobasic, and human albumin. Intron-A is also supplied in solution for injection or as a multi-dose pen for injection. These formulations contain sodium chloride, sodium phosphate dibasic, sodium phosphate monobasic, edentate disodium, polysorbate 80, and *m*-cresol as a preservative.

### 4.3 Clinical Experience with IFN-α as a Treatment for Malignancy

Cytokine therapy with interferon-alpha (IFN-α) is regularly used as a treatment for metastatic melanoma and renal cell carcinoma. IFN-α is also FDA-approved for melanoma patients as an adjuvant following the surgical resection of thick primary tumors and dissection of regional lymph nodes [101–103]. Historically this cytokine has also been used to treat chronic myelogenous leukemia and Kaposi's sarcoma [104,105]. The recombinant form of this cytokine became available for use in human clinical trials during the 1980s [106]. In the setting of metastatic melanoma, the response rate to IFN-α typically ranges from 10 to 20%, where in general, complete responses represent one-third of total responses [107–110]. Likewise, a 15% objective response rate and significantly prolonged overall survival of patients with metastatic renal cell carcinoma (RCC) was observed in response to single agent IFN-α in randomized clinical trials [111,112]. Subsequently, a series of clinical trials evaluated the use of IFN-α in the adjuvant setting for melanoma patients who underwent complete surgical resection of their tumor, but were at high-risk for recurrent disease. An initial phase III, trial was conducted in which patients were randomized to receive either high-dose IFN-α (20 MU/m$^2$/d i.v.) for 1 month followed by 10 MU/m$^2$/thrice weekly subcutaneously or observation [113]. The results of this trial indicated an increase in both the overall and relapse-free survival of IFN-α-treated patients [113]. Subsequent trials of high-dose and low-dose IFN-α regimens (3 MU thrice weekly $\times$ 2 years) also demonstrated an increase in relapse free survival of high-dose IFN-α-treated patients. However, these studies did not report a difference in overall survival in a similar population of surgically treated patients. Despite these conflicting results, an adjuvant trial comparing an anti-ganglioside vaccine to high-dose IFN-α was stopped as the result of an early analysis which showed a clear survival benefit to patients in the IFN-α arm [103]. Overall, IFN-α remains the only FDA-approved adjuvant therapy for patients with melanoma who have undergone surgical resection of high-risk lesions. There remains no alternative therapy that has shown efficacy in randomized trials in this setting and IFN-α will likely continue as the standard of care in the adjuvant setting and will continue to be given in combination with other therapeutic agents. Accordingly, there remains a need to identify relevant biomarkers which might be useful for predicting clinical response or toxicity in response to this therapy.

## 4.4 Dosage and Toxicity Considerations

As with many other cytokine therapies, the use of IFN-α is tolerated quite well by many patients, but results in a marked degree of toxicity in others. Particularly when given at high-doses, IFN-α is associated with a unique spectrum of toxicities that often limit patient compliance and have tempered enthusiasm for using this therapy as a first-line agent for metastatic disease [114]. The most common adverse reactions to high-dose IFN-α are flu-like symptoms including fever, fatigue, headache, myalgias chills and arthralgias. In some cases, IFN-α treatment may result in depression, therefore IFN-α is not recommended in patients with a pre-existing psychiatric condition. Gastrointestinal effects, mild proteinuria, intermittent or transient skin rashes and myelosuppression are also commonly seen. Hepatic toxicity is manifested primarily as liver enzyme disturbances which are generally mild to moderate and transient. Pulmonary complications are relatively unusual and may include dyspnea, cough or more rarely pneumonitis or pneumonia. Edema, hypotension, arrhythmia, and tachycardia have been observed and may be related in some cases to the febrile reaction that occurs frequently during initial therapy. IFN-α could potentially lead to a depression of the hepatic clearance of certain drugs that use the hepatic microsomal cytochrome P-450 pathway [107,114].

In contrast, lower or moderate doses of this cytokine are often well tolerated and in rare cases can be administered in a sustained manner for several months or even years [101,115]. The dosages employed in early Phase I and Phase II trials ranged from 3 to 100 million units (MU) per square meter ($m^2$) and would be considered very high dose regimens according to the current guidelines [107]. The route of IFN-α administration does not appear to be a critical factor influencing clinical response. Indeed, investigators have reported complete clinical responses following both intravenous (i.v.) or intramuscular (i.m.) routes of administration in sequential clinical trials conducted at the same institution [109] Despite this variability in dosage and in subject populations of the early trials, the response rates ranged between 5 and 30%. Although complete clinical responses have been observed most frequently upon the administration of higher doses (10–50 MU/$m^2$), complete responses have been shown to occur in response to very low doses of IFN-α (1–3 MU/$m^2$) [109,114]. Recent data have also demonstrated that escalating clinical doses of IFN-α are no more effective at initiating signal transduction and gene expression in immune effector cells [116,117]. Indeed, recent experience with tumor-targeted antibodies or other biologic agents demonstrate that higher doses do not necessarily exert the optimal biologic effect. These data are intriguing in that they suggest elevated doses of cytokines associated with greater toxicity may not necessarily be producing a greater clinical benefit. Therefore, it remains of value to explore strategies to individualize dosing of IFN-α and other immunotherapeutic cytokines on a patient-by-patient basis.

## 4.5 Mechanism of Action

The precise mechanism by which IFN-α mediates its anti-tumor activity remains undefined. Because IFN-α can bind to its receptor present upon both tumor cells and

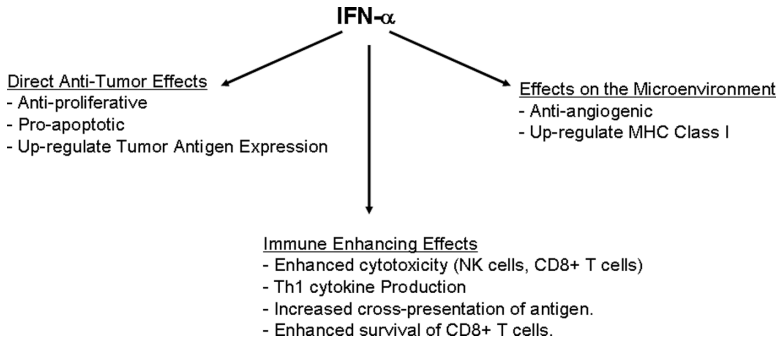


***FIGURE 6*** *Mechanism of interferon-alpha anti-tumor action. The molecular targets critical for the anti-tumor properties of IFN-( have not been precisely defined. However, IFN-( is known to act directly on the tumor cell, on the microenvironment and via its ability to stimulate the host immune system.*

host cells, this cytokine can mediate pleiotropic effects in a biological system (Fig. 6). Indeed, *in vitro* studies have demonstrated that treatment of tumor cells with IFN-α leads to decreased cell proliferation and in some cases can promote apoptosis, or programmed cell death [118–120]. Thus the direct actions of IFN-α on the tumor cell were thought to constitute a major mechanism of action. Furthermore, other studies have confirmed a role for IFN-α in inhibiting the secretion of pro-angiogenic molecules from tumor cells (e.g., basic fibroblast growth factor) [121,122] and up-regulating the expression of major histocompatibility complex class I (MHC Class I) proteins on tumor cells [123,124].

Despite the apparent direct effects of IFN-α on the tumor cell, Belardelli, and Gresser have conducted numerous eloquent studies investigating the mechanisms of anti-tumor effects of Type I IFN in mice transplanted with Friend leukemia cells (FLC) [92,125–130]. These tumor cells express Friend leukemia virus antigens on the cell surface and are metastatic to the liver and spleen following injection into syngeneic DBA/2 mice. Belardelli and Gresser demonstrated that IFN-α induced anti-tumor activity in mice transplanted with an IFN-resistant FLC clone. From these and subsequent studies [126,131,132], it was demonstrated that IFN-α acted to elicit anti-tumor activity through host-dependent mechanisms.

A series of more recent studies have further dissected the mechanism of host-dependent, IFN-α-mediated anti-tumor activity in murine models of melanoma. We found that the anti-tumor effects of IFN-α were dependent upon STAT1-mediated signal transduction in the host immune cells, rather than the tumor cell [133,134]. Subsequent studies have further demonstrated that exogenous IFN-α has profound anti-tumor activity when administered to mice lacking the negative regulatory suppressor of cytokine signaling-1 (SOCS1) protein (an inhibitor of STAT1 signal transduction) [135]. This finding is particularly intriguing as it suggests that modulating the negative regulators of IFN-α induced signal transduction might be a means of sensitizing a cell to enhanced effects of IFN-α.

Clinical studies have shown that infiltration of $CD8^+$ T cells in cutaneous lesions from melanoma patients was associated with a favorable clinical response to IFN-α

therapy [136]. Finally, studies indicate that melanoma patients who develop autoantibodies and clinical manifestations of autoimmunity during the course of adjuvant IFN-α therapy are less likely to experience disease recurrence. In a report by Gogas *et al.*, autoantibodies were detected in 26% of melanoma patients following adjuvant IFN-α therapy ($n = 52$ patients total). Univariate and multi-variate regression analyses indicated that autoimmunity was an independent prognostic marker for improved relapse-free survival and overall survival [137]. The development of autoimmunity may signify that exogenous IFN-α can break the tolerance of host T lymphocytes to host antigens as well as melanoma associated antigens and further suggests that a principal mechanism of its anti-tumor action is immunologic in nature.

### 4.6 Future Studies with IFN-α

The extensive clinical experience with IFN-α and the fact that it can bind to receptors on both host immune effectors and tumor cells makes it an attractive candidate for combination therapy. Over the last several years, IFN-α has been administered in combination with multiple chemotherapeutic agents, cytokines or tumor targeting antibodies [23,138–140]. In addition, more stable, pegylated forms (PEG-INTRON) of IFN-α have become available for clinical use [141]. It is thought that this formulation might potentially allow for prolonged stability of the drug *in vivo* and fewer toxic side effects. As with most other cytokines, there remains a critical need to identify the molecular determinants that differentiate optimal dosing, patient toxicity, and clinical responsiveness to this drug. Advances in our mechanistic understanding in these areas could lead to important improvements in choosing which patients may be most likely to gain benefit from this potentially toxic therapy and to identify rational combinations for clinical use.

## 5 ONCOLYTIC VIRUS

Oncolytic virus (OV) is a biological cancer-therapy based on the natural or engineered propensity of some viruses to replicate in and destroy cancer cells. While the concept of using live viruses to infect and destroy tumors dates back nearly/over a century ago, advances in molecular biology and virology have accelerated the development of OV in the last two decades [142]. Therapeutic viruses used for the treatment of cancer can broadly be classified as replication deficient and conditionally replication competent viruses. Replication deficient viruses are mostly used as gene therapy vehicles to deliver a therapeutic, or cytotoxic gene to the cancer cells. Conditionally replication competent viruses are viruses whose replication potential is compromised in normal cells, but can manipulate the aberrant molecular/genetic pathways in tumors to replicate efficiently in cancer cells. This leads to the lytic destruction of the infected malignant cell. Oncolytic viruses currently being tested for safety and efficacy include herpes simplex virus (HSV), and adenovirus, and RNA viruses such as Newcastle disease virus (NDV), measles virus, Vaccinia virus (VV), and Vesicular Stomatitis virus (VSV).

In this section, we will describe the conception and testing of oncolytic viruses derived by genetically manipulating wild type HSV-1 derived viruses.

## 5.1 Herpes Simplex Virus-1 Derived Oncolytic Viruses

HSV-1 is an enveloped, double stranded DNA virus containing a large, well characterized, fully sequenced genome of about 152 kb of DNA that encodes for more than 80 genes (Fig. 7). While the large size makes genetic manipulations cumbersome, it also provides ample opportunities to remove genes that are not essential for replication (estimated to be about 30 kb). These genes can be replaced to allow for the insertion of therapeutic transgenes within the viral backbone. Furthermore, the ability of HSV-1 to remain as an episome avoids the possibility of any insertional mutagenesis of the infected cell. These properties make HSV-1 a very desirable vector for therapeutic applications. In addition, HSV is sensitive to several anti-herpetic drugs such as acyclovir (ACV) and gancyclovir (GCV) which are readily

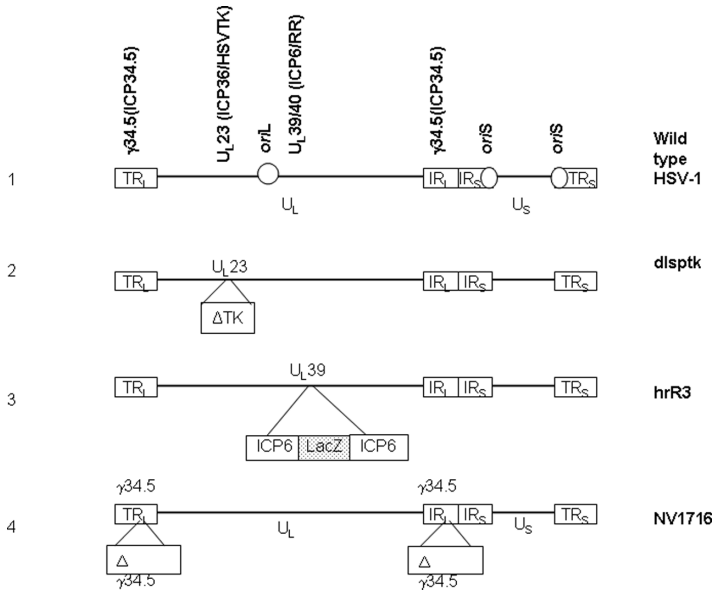


***FIGURE 7*** *Structure of the HSV-1 wild type genome and some of the oncolytic viruses generated from it. (1) The HSV-1 genome consists of a linear double stranded DNA of about 152 kb, encoding for more than 80 genes. Four inverted repeat sequences: Terminal repeat long ($TR_L$), Terminal repeat small ($TR_S$), and Internal repeat long ($IR_L$) and internal repeat small ($IR_S$) divide the genome into unique long ($U_L$) and a unique small ($U_S$) segment. The three origins of replication within the HSV genome are indicated by circles. Locations of genes that are disrupted/deleted for the creation of tumor specific replication competent viruses are indicated. (2) dlsptk is generated by a 360 bp deletion within the locus UL23, encoding for viral HSVTK. (3) hrR3 is generated by an in frame gene disrupting insertion of the E. Coli beta galactosidase gene into the UL39 locus encoding for large subunit of viral Ribonucleotide Reductase. (4) NV1716 is generated by deletion of the both the copies of the gene encoding for ICP34.5 within the terminal repeat sequences of the HSV genome.*

available to keep viral replication in check if needed. Together these factors make HSV an excellent candidate for oncolytic therapy.

## 5.2 HSV Thymidine Kinase Deleted Oncolytic Virus

HSV thymidine kinase (HSVTK) is a homidimeric protein encoded by $U_L23$ (ICP36) gene of the HSV genome. Unlike its mammalian counterpart, the viral enzyme has a very broad substrate specificity permitting it to phosphorylate pyrimidine and purine nucleosides along with several nucleoside analog molecules. Viral HSVTK is required for nucleotide metabolism essential for DNA replication in non-dividing cells which do not express the mammalian counterpart of this enzyme. Hence an HSV virus deficient in HSVTK activity would be restricted for its replication in dividing cells, where the cellular enzyme could compensate for this defect, and could be exploited for the generation of an oncolytic HSV. Subsequently the first genetically altered HSV was created by a 360 bp deletion within the viral gene UL39/ICP36 encoding for the enzyme thymidine kinase (HSVTK) [143] (Fig. 7). This mutant HSV (dlsptk) was tested for oncolytic efficacy in human glioblastoma cell lines and cultures *in vitro*. Further dlsptk treatment of human gliomas implanted into immune deficient mice showed that treated tumors were smaller and mice survived longer than untreated mice with gliomas [143]. Further analysis of the pathogenesis of corneal infection of SCID mice by dlsptk revealed persistent infection which resulted in death of inoculated mice within $80.2 \pm 21.3$ days. The lack of sensitivity of dlsptk to commonly used anti-herpetic drugs such as ACV and GCV (which exploit the virus expressed HSVTK), combined with its neurovirulence precluded the use of this virus in human patients. However, the efficacy seen in pre-clinical animal models underscored the need for the development of safer oncolytic HSV vectors which retain the anti-herpetic drug sensitizing HSVTK, gene and are also less virulent.

## 5.3 Ribonucleotide Reductase Deficient OV

Ribonucleotide reductase (RR) is required for the *de novo* synthesis of deoxyribonucleotides, and is essential for viral DNA synthesis and replication. HSV-1 encodes for its own RR, composed of a large and small subunit encoded by $U_L39$ and $U_L40$, respectively. The proteins encoded by $U_L39$ and $U_L40$ together make up an $\alpha 2\beta 2$ tetramer constituting the functional RR. Like the HSVTK, viral RR is essential for viral replication in non-dividing cells. However the cellular enzyme compensates for this defect in cycling cells, but not in non-dividing cells which do not express the cellular RR. Hence an HSV lacking its own RR would be replication competent only in mitotic cancer cells but not in non-replicating normal cells. To test this hypothesis a mutant virus, deficient for the large subunit of RR (ICP6/RR) was created by the insertion of an *Eschericia coli* LacZ gene within the UL39 gene locus [144] (Fig. 7). Cytopathic and drug sensitivity assays using this attenuated virus (hrR3) revealed efficient cancer cell killing *in vitro* and also efficient tumor killing in animal models of tumorigenesis [145,146]. Additionally drug sensitivity assays revealed that the attenuated hrR3 was 10-fold more sensitive to anti-herpetic drug GCV. HSVTK results in phosphorylation of GCV to a toxic nucleotide precursor which results in

death of infected cells and also uninfected bystander cells. This suggested that hrR3 infected glioma cells would be sensitive to killing by GCV. Consistent with this treatment of rats bearing intracranial tumors with GCV in combination with hrR3 significantly enhanced the anti-tumor effect of hrR3[147].

More recent evidence also indicates that RR mutant HSV may be "molecularly targeted" to replicate more efficiently in cells with specific oncogenic mutations such as homozygous deletion of the p16 gene [148].

## 5.4 ICP34.5 Deficient Oncolytic HSV

Interferon-induced, double-stranded RNA-activated protein kinase (PKR) is one of the principal cellular defenses against viral infection. Cellular PKR is maintained at low levels in uninfected cells but is induced by interferon. Activated PKR phosphorylates $\alpha$ subunit of eIF-2$\alpha$ leading to a total shutoff of host protein synthesis. HSV gene $R_L$-1 encodes for ICP34.5 protein which binds to protein phosphatase 1$\alpha$ and leads to the dephosphorylation of $\alpha$ subunit of eIF-2$\alpha$. HSV-1 ICP34.5 was identified to be the neurovirulence gene and attenuation of both the copies of this gene led to reduction of neurovirulence associated with HSV-1 *in vitro* and *in vivo* [149,150]. More significantly this mutant virus retained the ability to replicate in actively dividing mouse embryo cells but not in confluent mouse embryo cultures [150]. Consistent with these results an RL-1 null HSV-1 mutant (1716), which fails to produce ICP34.5 was able to efficiently replicate in and cause subsequent cell death in a majority of glioma cell lines and primary tumor derived cells [151]. The *in vivo* anti-tumor efficacy of 1716 was tested in mice bearing experimental brain tumors, and treatment resulted in a statistically significant increase in survival [152,153]. Pathological examination of mice injected intracranially with 1716, revealed a finite, self-limiting host response to 1716, highlighting the potential of using it for therapeutic purposes.

## 5.5 Pharmacological Formulation

The clinical grade HSV 1716 used in human trials was prepared by amplifying low passage laboratory stocks in BHK-21/C13 cells from a master seed cell bank under cGMP conditions7 by Q-One Biotech Ltd, Glasgow, UK. The amplified clinical grade viral stocks were tested for contamination, and evaluated for titers and stored at $-70°C$. The virus was stored at the appropriate dilution in 1 mL of human serum albumin, (0.5–1.0%) in isotonic PBS. The titers of the virus were reconfirmed a week prior to the injection date. All studies were approved by the UK gene therapy advisory committee.

## 5.6 Clinical Testing in Human Patients

***5.6.1 HSV Mutant 1716*** Based on its safety profile in animals, the mutated HSV 1716 was tested in UK for safety in human patients with malignant brain tumors [154]. This clinical study represented the first foray of inoculating mutated oncolytic HSV

vectors into the brains of human patients with malignant brain tumors, and for concerns of safety, only a very moderate dosage of the virus was permitted to be tested. In this study, nine patients with malignant glioma were treated with doses of $10^3$, $10^4$, and $10^5$ plaque forming units (p.f.u.) of HSV1716 by direct intra-tumoral injection. All the patients were diagnosed with recurrent malignant glioma (eight glioblastoma and one anaplastic astrocytoma) that had relapsed following radical treatment, including surgery and radiation therapy. Patients were closely monitored and no adverse clinical events which could be attributed to the administration of 1716 were observed. No evidence of encephalitis, or ongoing lytic replication was found in any treated patient. Buccal sawbs and serum samples from patients were examined weekly for presence of HSV and no evidence of viral shedding or reactivation of endogenous latent virus was seen in any patient [154]. Five of these nine patients underwent a subsequent tumor progression and biopsy samples from these patients showed no signs of 1716 or wild type HSV by PCR [154]. No evidence of efficacy was observed with inoculation of this dosage in patients, nevertheless the absence of any significant toxicity which could be attributed the administered OV was encouraging and subsequently two clinical trials to test therapeutic efficacy of HSV mutant 1716 were initiated. In a subsequent report by the investigators it was noted that two of these nine patients were alive and stable at 4 years and 4 months and 3 years and 7 months after treatment [155].

The very low dosage given in the first clinical study was based on the hypothesis that the oncolytic virus should be able to replicate and hence produce a higher dose within the tumor. In the second clinical study evidence of viral replication in human high grade glioma was examined. Twelve patients with high grade malignant glioma received $10^5$ p.f.u. of HSV mutant 1716 by direct intratumoral stereotactic injection [156]. Tumor resection was carried out 4–9 days after treatment and the tissue was examined for presence of HSV DNA by PCR, southern blotting immuno-histochemistry and the presence of infectious viral particles by plaque assay. Results from plaque assays revealed the presence of infectious viral particles in resected tumors from two out of twelve patients. Southern blot analysis confirmed that the recovered virus was HSV mutant 1716. HSV DNA was detected by PCR at the site of injection in 10 out of the 12 patients while specimens from distal tumor sites of only four patients were found to be positive for HSV DNA [156]. Immunohistochemistry of tissue samples revealed positive staining for HSV proteins in tumor tissues from two of the twelve patients. The absence of detectable viral DNA protein or infectious particles in some specimens could be attributed to tissue sampling errors owing to non-uniform spread of the injected virus throughout the tumor tissue, or simply reflect the need to improve current methods used for detection. Nevertheless the study demonstrated evidence of HSV presence in tissue without any obvious signs of toxicity [156].

Based on these studies another clinical trial in the UK was initiated to test the therapeutic efficacy of the HSV mutant 1716 in human glioma patients. Despite the lack of toxicity seen in those trials, due to paramount safety concerns the therapeutic dose given to the patients was not increased as would have been done for clinical trials with a chemotherapeutic agent. A total of 12 patients who had been diagnosed with high grade recurrent (six patients) or newly diagnosed (six patients) glioma

underwent gross tumor resection. After the initial tumor debulking, $10^5$ p.f.u of HSV mutant 1716 was injected into 8–10 sites within the tissue surrounding tumor cavity. The patients were carefully monitored with a comprehensive clinical work up including immunological, hematological (full blood count with white cell count differential) and biochemical (urea, electrolytes, liver function tests, and bone profile), and neurological examinations. The lack of any virus related toxicity in this study considerably strengthened the safety profile of this approach even though the virus had been injected into tissue involving functioning brain. The expected median survival for patients with GBM is 1 year from the time of initial diagnosis and only 3–6 months following diagnosis of a recurrent tumor. One patient with a recurrent tumor and two patients diagnosed with primary GBM survived for more than 22, 18, and 15 months, respectively. The patient who survived for more than 22 months was diagnosed with a recurrent tumor but was disease free at 22 months. However, a causal relationship between response and viral treatment could not be established. This study demonstrated that HSV1716 can be injected into the normal brain of individuals without any ensuing toxicity.

The therapeutic potential of HSV1716 has also been assessed in five patients with stage 4 metastatic melanoma. In each patient paired distinct nodules were injected with either dilutent or $10^3$ p.f.u of HSV 1716 [157]. Flattening of previously palpable tumor nodules was seen after direct injections of HSV1716 in one patient, and in three patients there was microscopic evidence of tumor necrosis. These findings suggest that HSV1716 is non-toxic and could be of therapeutic benefit in patients with cancer.

## 5.7 Future of HSV 1716

HSV1716 is currently being pursued by Crusade Laboratories, Glasgow, UK. The company has been granted orphan drug status in Europe for "the use of HSV mutant 1716 in the treatment of glioma". Crusade laboratories have received regulatory approvals to initiate a Phase III trial, for glioma patients. Glioma patients with a first recurrence after surgery will be randomized into one of two treatment arms: HSV1716 or conventional chemotherapy. Satisfactory results in this trial could lead to a license and marketing authorization in glioma. The company is also pursuing Phase I/II safety and biological activity studies in patients with ovarian carcinoma and unresectable hepatocellular carcinoma [158].

## 5.8 Second Generation Multi-Mutated Oncolytic Viruses: G207

In parallel with the HSV mutant 1716 clinical studies in UK, a doubly attenuated oncolytic HSV, has also been investigated in human patients in the USA. Owing to the overriding concerns of safety arising from the use of singly mutated viruses, oncolytic viruses with multiple genetic defects were created for the treatment of cancer. G207, is one such virus that has deletions at both gamma 34.5 (RL1) loci and an in-frame and gene disrupting insertion of *E. Coli* beta galactosidase gene within the ICP6 gene (UL39), making it replication-competent in dividing cells. This minimizes the

chances of G207 reverting to wild type, and maintains an attenuated neurovirulence, temperature sensitivity, as well as ganciclovir hypersensitivity [159]. The insertion of β-galactosidase permits easy detection of cells harboring the virus in infected tissue. Pre-clinical testing of this virus was performed in nude mice harboring subcutaneous or intracerebral gliomas. Intraneoplastic administration of the virus resulted in slower tumor growth and/or prolonged survival [159].

Pre-clinical toxicity studies of G207 also found it to be avirulent upon intracerebral inoculation of mice and HSV-sensitive non-human primates. Shedding and biodistribution of G207 was determined in new world owl monkeys (*Aotus nancymae*) after intracerebral inoculation of $3 \times 107$ p.f.u. of clinical grade virus. Even after 2 years of inoculation G207 DNA was found to be restricted to the brain, although no infectious viral particles were obtained. Histopathology of the necrospied tissue revealed normal brain even at the site of injection. Absence of virus DNA in tears saliva and vaginal secretions indicated lack of viral shedding and suggested that strict biohazard management of G207 patients may not be required. Together these results implicate the possibility of considering mutimutated oncolytic viruses such as G207 for clinical evaluation for the treatment of cancer.

Safety of intracerebral inoculation of G207 in humans was determined in a dose escalation study in patients suffering from recurrent malignant glioma. The protocol was approved by the Recombinat DNA Advisory Committee of the National Institute of Health and Food and Drug Administration. The trial was initiated with the administration of $1 \times 10^6$ p.f.u. of G207, into the brains of three patients post-tumor resection. After 28 days of observation to confirm lack of acute toxicity three patients were then recruited into the next cohort and were treated with the higher dosage. In the last cohort the patients were treated with $1 \times 10^9$ p.f.u of G207 inoculated into five sites after surgical resection of the brain tumor. No toxicity which could be directly attributed to the infectious virus was noted in this study. This study suggested that G207 can be safely inoculated into human brain tumors upto doses of $1 \times 10^9$ p.f.u. without any adverse events. This agent is currently being investigated for efficacy in Clinical trials. Future studies will reveal the efficacy of combining oncolytic viruses with radiation and chemotherapy in patients.

## REFERENCES

1. Jain RK. Molecular regulation of vessel maturation. Nat Med 2003; 9: 685–693.
2. Folkman J, Kalluri R. Cancer without disease. Nature 2004; 427: 787.
3. Folkman J. Tumor angiogenesis: therapeutic implications. N Engl J Med 1971; 285: 1182–1186.
4. Folkman J. What is the evidence that tumors are angiogenesis dependent? J Natl Cancer Inst 1990; 82: 4–6.
5. Ferrara N, Henzel WJ. Pituitary follicular cells secrete a novel heparin-binding growth factor specific for vascular endothelial cells. Biochem Biophys Res Commun 1989; 161: 851–858.

6. Senger DR, Galli SJ, Dvorak AM, Perruzzi CA, Harvey VS, Dvorak HF. Tumor cells secrete a vascular permeability factor that promotes accumulation of ascites fluid. Science 1983; 219: 983–985.
7. Keck PJ, Hauser SD, Krivi G, Sanzo K, Warren T, Feder J, Connolly DT. Vascular permeability factor, an endothelial cell mitogen related to pdgf. Science 1989; 246: 1309–1312.
8. Leung DW, Cachianes G, Kuang WJ, Goeddel DV, Ferrara N. Vascular endothelial growth factor is a secreted angiogenic mitogen. Science 1989; 246: 1306–1309.
9. Ferrara N. Vascular endothelial growth factor: basic science and clinical progress. Endocr Rev 2004; 25: 581–611.
10. Kaur B, Khwaja FW, Severson EA, Matheny SL, Brat DJ, Van Meir EG. Hypoxia and the hypoxia-inducible-factor pathway in glioma growth and angiogenesis. Neuro-oncol 2005; 7: 134–153.
11. Carmeliet P, Ferreira V, Breier G, Pollefeyt S, Kieckens L, Gertsenstein M, Fahrig M, Vandenhoeck A, Harpal K, Eberhardt C, Declercq C, Pawling J, Moons L, Collen D, Risau W, Nagy A. Abnormal blood vessel development and lethality in embryos lacking a single vegf allele. Nature 1996; 380: 435–439.
12. Ferrara N, Carver-Moore K, Chen H, Dowd M, Lu L, O'shea KS, Powell-Braxton L, Hillan KJ, Moore MW. Heterozygous embryonic lethality induced by targeted inactivation of the vegf gene. Nature 1996; 380: 439–442.
13. Brat DJ, Van Meir EG. Vaso-occlusive and prothrombotic mechanisms associated with tumor hypoxia, necrosis, and accelerated growth in glioblastoma. Lab Invest 2004; 84: 397–405.
14. Brat DJ, Castellano-Sanchez A, Kaur B, Van Meir EG. Genetic and biologic progression in astrocytomas and their relation to angiogenic dysregulation. Adv Anat Pathol 2002; 9: 24–36.
15. Mizukami Y, Kohgo Y, Chung DC. Hypoxia inducible factor-1 independent pathways in tumor angiogenesis. Clin Cancer Res 2007; 13: 5670–5674.
16. Kim KJ, Li B, Houck K, Winer J, Ferrara N. The vascular endothelial growth factor proteins: identification of biologically relevant regions by neutralizing monoclonal antibodies. Growth Factors 1992; 7: 53–64.
17. Kim KJ, Li B, Winer J, Armanini M, Gillett N, Phillips HS, Ferrara N. Inhibition of vascular endothelial growth factor-induced angiogenesis suppresses tumour growth *in vivo*. Nature 1993; 362: 841–844.
18. Ferrara N, Hillan KJ, Novotny W. Bevacizumab (avastin), a humanized anti-vegf monoclonal antibody for cancer therapy. Biochem Biophys Res Commun 2005; 333: 328–335.
19. Presta LG, Chen H, O'connor SJ, Chisholm V, Meng YG, Krummen L, Winkler M, Ferrara N. Humanization of an anti-vascular endothelial growth factor monoclonal antibody for the therapy of solid tumors and other disorders. Cancer Res 1997; 57: 4593–4599.
20. Ryan AM, Eppler DB, Hagler KE, Bruner RH, Thomford PJ, Hall RL, Shopp GM, O'neill CA. Preclinical safety evaluation of rhumabvegf, an antiangiogenic humanized monoclonal antibody. Toxicol Pathol 1999; 27: 78–86.
21. Gordon MS, Margolin K, Talpaz M, Sledge GW Jr, Holmgren E, Benjamin R, Stalter S, Shak S, Adelman D. Phase i safety and pharmacokinetic study of recombinant human

anti-vascular endothelial growth factor in patients with advanced cancer. J Clin Oncol 2001; 19: 843–850.

22. Kabbinavar F, Hurwitz HI, Fehrenbacher L, Meropol NJ, Novotny WF, Lieberman G, Griffing S, Bergsland E. Phase ii, randomized trial comparing bevacizumab plus fluorouracil (fu)/leucovorin (lv) with fu/lv alone in patients with metastatic colorectal cancer. J Clin Oncol 2003; 21: 60–65.
23. Escudier B, Pluzanska A, Koralewski P, Ravaud A, Bracarda S, Szczylik C, Chevreau C, Filipek M, Melichar B, Bajetta E, Gorbunova V, Bay JO, Bodrogi I, Jagiello-Gruszfeld A, Moore N. Bevacizumab plus interferon alfa-2a for treatment of metastatic renal cell carcinoma: a randomised, double-blind phase iii trial. Lancet 2007; 370: 2103–2111.
24. Shih T, Lindley C. Bevacizumab: an angiogenesis inhibitor for the treatment of solid malignancies. Clin Ther 2006; 28: 1779–1802.
25. Bergsland E, Dickler MN. Maximizing the potential of bevacizumab in cancer treatment. Oncologist 2004; 9 (Suppl 1): 36–42.
26. Orlowski RZ, Kuhn DJ. Proteasome inhibitors in cancer therapy: lessons from the first decade. Clin Cancer Res 2008; 14: 1649–1657.
27. Adams J. The proteasome: a suitable antineoplastic target. Nat Rev Cancer 2004; 4: 349–360.
28. Adams J, Kauffman M. Development of the proteasome inhibitor velcade (bortezomib). Cancer Invest 2004; 22: 304–311.
29. Demartino GN, Gillette TG. Proteasomes: machines for all reasons. Cell 2007; 129: 659–662.
30. Brooks P, Fuertes G, Murray RZ, Bose S, Knecht E, Rechsteiner MC, Hendil KB, Tanaka K, Dyson J, Rivett J. Subcellular localization of proteasomes and their regulatory complexes in mammalian cells. Biochem J 2000; 346 (Pt 1): 155–161.
31. Peters JM, Cejka Z, Harris JR, Kleinschmidt JA, Baumeister W. Structural features of the 26 s proteasome complex. J Mol Biol 1993; 234: 932–937.
32. Hideshima T, Richardson P, Chauhan D, Palombella VJ, Elliott PJ, Adams J, Anderson KC. The proteasome inhibitor ps-341 inhibits growth, induces apoptosis, and overcomes drug resistance in human multiple myeloma cells. Cancer Res 2001; 61: 3071–3076.
33. Adams J, Palombella VJ, Elliott PJ. Proteasome inhibition: a new strategy in cancer treatment. Invest New Drugs 2000; 18: 109–121.
34. Bold RJ, Virudachalam S, Mcconkey DJ. Chemosensitization of pancreatic cancer by inhibition of the 26s proteasome. J Surg Res 2001; 100: 11–17.
35. Shah IM, Lees KR, Pien CP, Elliott PJ. Early clinical experience with the novel proteasome inhibitor ps-519. Br J Clin Pharmacol 2002; 54: 269–276.
36. Zhang L, Zhang ZG, Zhang RL, Lu M, Adams J, Elliott PJ, Chopp M. Postischemic (6-hour) treatment with recombinant human tissue plasminogen activator and proteasome inhibitor ps-519 reduces infarction in a rat model of embolic focal cerebral ischemia. Stroke 2001; 32: 2926–2931.
37. Kane RC, Farrell AT, Sridhara R, Pazdur R. United states food and drug administration approval summary: bortezomib for the treatment of progressive multiple myeloma after one prior therapy. Clin Cancer Res 2006; 12: 2955–2960.
38. Adams J, Palombella VJ, Sausville EA, Johnson J, Destree A, Lazarus DD, Maas J, Pien CS, Prakash S, Elliott PJ. Proteasome inhibitors: a novel class of potent and effective antitumor agents. Cancer Res 1999; 59: 2615–2622.

39. Delic J, Masdehors P, Omura S, Cosset JM, Dumont J, Binet JL, Magdelenat H. The proteasome inhibitor lactacystin induces apoptosis and sensitizes chemo- and radio-resistant human chronic lymphocytic leukaemia lymphocytes to tnf-alpha-initiated apoptosis. Br J Cancer 1998; 77: 1103–1107.
40. Imajoh-Ohmi S, Kawaguchi T, Sugiyama S, Tanaka K, Omura S, Kikuchi H. Lactacystin, a specific inhibitor of the proteasome, induces apoptosis in human monoblast u937 cells. Biochem Biophys Res Commun 1995; 217: 1070–1077.
41. Orlowski RZ, Eswara JR, Lafond-Walker A, Grever MR, Orlowski M, Dang CV. Tumor growth inhibition induced in a murine model of human burkitt's lymphoma by a proteasome inhibitor. Cancer Res 1998; 58: 4342–4348.
42. Shinohara K, Tomioka M, Nakano H, Tone S, Ito H, Kawashima S. Apoptosis induction resulting from proteasome inhibition. Biochem J 1996; 317 (Pt 2): 385–388.
43. Leblanc R, Catley LP, Hideshima T, Lentzsch S, Mitsiades CS, Mitsiades N, Neuberg D, Goloubeva O, Pien CS, Adams J, Gupta D, Richardson PG, Munshi NC, Anderson KC. Proteasome inhibitor ps-341 inhibits human myeloma cell growth *in vivo* and prolongs survival in a murine model. Cancer Res 2002; 62: 4996–5000.
44. Teicher BA, Ara G, Herbst R, Palombella VJ, Adams J. The proteasome inhibitor ps-341 in cancer therapy. Clin Cancer Res 1999; 5: 2638–2645.
45. Berenson JR, Ma HM, Vescio R. The role of nuclear factor-kappab in the biology and treatment of multiple myeloma. Semin Oncol 2001; 28: 626–633.
46. Hayden MS, Ghosh S. Shared principles in nf-kappab signaling. Cell 2008; 132: 344–362.
47. Nawrocki ST, Bruns CJ, Harbison MT, Bold RJ, Gotsch BS, Abbruzzese JL, Elliott P, Adams J, Mcconkey DJ. Effects of the proteasome inhibitor ps-341 on apoptosis and angiogenesis in orthotopic human pancreatic tumor xenografts. Mol Cancer Ther 2002; 1: 1243–1253.
48. Chauhan D, Li G, Shringarpure R, Podar K, Ohtake Y, Hideshima T, Anderson KC. Blockade of hsp27 overcomes bortezomib/proteasome inhibitor ps-341 resistance in lymphoma cells. Cancer Res 2003; 63: 6174–6177.
49. Fernandez Y, Verhaegen M, Miller TP, Rush JL, Steiner P, Opipari AW Jr, Lowe SW, Soengas MS. Differential regulation of noxa in normal melanocytes and melanoma cells by proteasome inhibition: therapeutic implications. Cancer Res 2005; 65: 6294–6304.
50. Qin JZ, Ziffra J, Stennett L, Bodner B, Bonish BK, Chaturvedi V, Bennett F, Pollock PM, Trent JM, Hendrix MJ, Rizzo P, Miele L, Nickoloff BJ. Proteasome inhibitors trigger noxa-mediated apoptosis in melanoma and myeloma cells. Cancer Res 2005; 65: 6282–6293.
51. Nawrocki ST, Carew JS, Dunner K Jr, Boise LH, Chiao PJ, Huang P, Abbruzzese JL, Mcconkey DJ. Bortezomib inhibits pkr-like endoplasmic reticulum (er) kinase and induces apoptosis via er stress in human pancreatic cancer cells. Cancer Res 2005; 65: 11510–11519.
52. Nawrocki ST, Carew JS, Pino MS, Highshaw RA, Andtbacka RH, Dunner K Jr, Pal A, Bornmann WG, Chiao PJ, Huang P, Xiong H, Abbruzzese JL, Mcconkey DJ. Aggresome disruption: a novel strategy to enhance bortezomib-induced apoptosis in pancreatic cancer cells. Cancer Res 2006; 66: 3773–3781.

53. Nawrocki ST, Carew JS, Pino MS, Highshaw RA, Dunner K Jr, Huang P, Abbruzzese JL, Mcconkey DJ. Bortezomib sensitizes pancreatic cancer cells to endoplasmic reticulum stress-mediated apoptosis. Cancer Res 2005; 65: 11658–11666.
54. Papandreou CN, Daliani DD, Nix D, Yang H, Madden T, Wang X, Pien CS, Millikan RE, Tu SM, Pagliaro L, Kim J, Adams J, Elliott P, Esseltine D, Petrusich A, Dieringer P, Perez C, Logothetis CJ. Phase i trial of the proteasome inhibitor bortezomib in patients with advanced solid tumors with observations in androgen-independent prostate cancer. J Clin Oncol 2004; 22: 2108–2121.
55. Geyer, SM, Markovic SM, Fitch TR, Albertini MR, Maples WJ, Fracasso PM, Erlichman C. Phase 2 study of ps-341 (bortezomib) in the treatment of metastatic malignant melanoma (mmm). Proceedings of ASCO; 2004. p 7526.
56. Aghajanian C, Soignet S, Dizon DS, Pien CS, Adams J, Elliott PJ, Sabbatini P, Miller V, Hensley ML, Pezzulli S, Canales C, Daud A, Spriggs DR. A phase i trial of the novel proteasome inhibitor ps341 in advanced solid tumor malignancies. Clin Cancer Res 2002; 8: 2505–2511.
57. Orlowski RZ, Stinchcombe TE, Mitchell BS, Shea TC, Baldwin AS, Stahl S, Adams J, Esseltine DL, Elliott PJ, Pien CS, Guerciolini R, Anderson JK, Depcik-Smith ND, Bhagat R, Lehman MJ, Novick SC, O'connor OA, Soignet SL. Phase i trial of the proteasome inhibitor ps-341 in patients with refractory hematologic malignancies. J Clin Oncol 2002; 20: 4420–4427.
58. Bross PF, Kane R, Farrell AT, Abraham S, Benson K, Brower ME, Bradley S, Gobburu JV, Goheer A, Lee SL, Leighton J, Liang CY, Lostritto RT, Mcguinn WD, Morse DE, Rahman A, Rosario LA, Verbois SL, Williams G, Wang YC, Pazdur R. Approval summary for bortezomib for injection in the treatment of multiple myeloma. Clin Cancer Res 2004; 10: 3954–3964.
59. Aghajanian C, Dizon DS, Sabbatini P, Raizer JJ, Dupont J, Spriggs DR. Phase i trial of bortezomib and carboplatin in recurrent ovarian or primary peritoneal cancer. J Clin Oncol 2005; 23: 5943–5949.
60. Amiri KI, Horton LW, Lafleur BJ, Sosman JA, Richmond A. Augmenting chemosensitivity of malignant melanoma tumors via proteasome inhibition: implication for bortezomib (velcade, ps-341) as a therapeutic agent for malignant melanoma. Cancer Res 2004; 64: 4912–4918.
61. Barlogie B, Anaissie E, Van Rhee F, Haessler J, Hollmig K, Pineda-Roman M, Cottler-Fox M, Mohiuddin A, Alsayed Y, Tricot G, Bolejack V, Zangari M, Epstein J, Petty N, Steward D, Jenkins B, Gurley J, Sullivan E, Crowley J, Shaughnessy JD Jr. Incorporating bortezomib into upfront treatment for multiple myeloma: early results of total therapy 3. Br J Haematol 2007; 138: 176–185.
62. Berenson JR, Yang HH, Sadler K, Jarutirasarn SG, Vescio RA, Mapes R, Purner M, Lee SP, Wilson J, Morrison B, Adams J, Schenkein D, Swift R. Phase i/ii trial assessing bortezomib and melphalan combination therapy for the treatment of patients with relapsed or refractory multiple myeloma. J Clin Oncol 2006; 24: 937–944.
63. Biehn SE, Moore DT, Voorhees PM, Garcia RA, Lehman MJ, Dees EC, Orlowski RZ. Extended follow-up of outcome measures in multiple myeloma patients treated on a phase i study with bortezomib and pegylated liposomal doxorubicin. Ann Hematol 2007; 86: 211–216.

64. Canfield SE, Zhu K, Williams SA, Mcconkey DJ. Bortezomib inhibits docetaxel-induced apoptosis via a p21-dependent mechanism in human prostate cancer cells. Mol Cancer Ther 2006; 5: 2043–2050.
65. Harousseau JL, Attal M, Leleu X, Troncy J, Pegourie B, Stoppa AM, Hulin C, Benboubker L, Fuzibet JG, Renaud M, Moreau P, Avet-Loiseau H. Bortezomib plus dexamethasone as induction treatment prior to autologous stem cell transplantation in patients with newly diagnosed multiple myeloma: results of an ifm phase ii study. Haematologica 2006; 91: 1498–1505.
66. Jagannath S, Durie BG, Wolf J, Camacho E, Irwin D, Lutzky J, Mckinley M, Gabayan E, Mazumder A, Schenkein D, Crowley J. Bortezomib therapy alone and in combination with dexamethasone for previously untreated symptomatic multiple myeloma. Br J Haematol 2005; 129: 776–783.
67. Ma C, Mandrekar SJ, Alberts SR, Croghan GA, Jatoi A, Reid JM, Hanson LJ, Bruzek L, Tan AD, Pitot HC, Erlichman C, Wright JJ, Adjei AA. A phase i and pharmacologic study of sequences of the proteasome inhibitor, bortezomib (ps-341, velcade), in combination with paclitaxel and carboplatin in patients with advanced malignancies. Cancer Chemother Pharmacol 2007; 59: 207–215.
68. Mateos MV, Hernandez JM, Hernandez MT, Gutierrez NC, Palomera L, Fuertes M, Diaz-Mediavilla J, Lahuerta JJ, De La Rubia J, Terol MJ, Sureda A, Bargay J, Ribas P, De Arriba F, Alegre A, Oriol A, Carrera D, Garcia-Larana J, Garcia-Sanz R, Blade J, Prosper F, Mateo G, Esseltine DL, Van De Velde H, San Miguel JF. Bortezomib plus melphalan and prednisone in elderly untreated patients with multiple myeloma: results of a multicenter phase 1/2 study. Blood 2006; 108: 2165–2172.
69. Messersmith WA, Baker SD, Lassiter L, Sullivan RA, Dinh K, Almuete VI, Wright JJ, Donehower RC, Carducci MA, Armstrong DK. Phase i trial of bortezomib in combination with docetaxel in patients with advanced solid tumors. Clin Cancer Res 2006; 12: 1270–1275.
70. Orlowski RZ, Nagler A, Sonneveld P, Blade J, Hajek R, Spencer A, San Miguel J, Robak T, Dmoszynska A, Horvath N, Spicka I, Sutherland HJ, Suvorov AN, Zhuang SH, Parekh T, Xiu L, Yuan Z, Rackoff W, Harousseau JL. Randomized phase iii study of pegylated liposomal doxorubicin plus bortezomib compared with bortezomib alone in relapsed or refractory multiple myeloma: combination therapy improves time to progression. J Clin Oncol 2007; 25: 3892–3901.
71. Orlowski RZ, Voorhees PM, Garcia RA, Hall MD, Kudrik FJ, Allred T, Johri AR, Jones PE, Ivanova A, Van Deventer HW, Gabriel DA, Shea TC, Mitchell BS, Adams J, Esseltine DL, Trehu EG, Green M, Lehman MJ, Natoli S, Collins JM, Lindley CM, Dees EC. Phase 1 trial of the proteasome inhibitor bortezomib and pegylated liposomal doxorubicin in patients with advanced hematologic malignancies. Blood 2005; 105: 3058–3065.
72. Ryan DP, Appleman LJ, Lynch T, Supko JG, Fidias P, Clark JW, Fishman M, Zhu AX, Enzinger PC, Kashala O, Cusack J Jr, Eder JP. Phase i clinical trial of bortezomib in combination with gemcitabine in patients with advanced solid tumors. Cancer 2006; 107: 2482–2489.
73. Ryan DP, O'neil BH, Supko JG, Rocha Lima CM, Dees EC, Appleman LJ, Clark J, Fidias P, Orlowski RZ, Kashala O, Eder JP, Cusack JC Jr. A phase i study of bortezomib plus irinotecan in patients with advanced solid tumors. Cancer 2006; 107: 2688–2697.
74. Voortman J, Smit EF, Honeywell R, Kuenen BC, Peters GJ, Van De Velde H, Giaccone G. A parallel dose-escalation study of weekly and twice-weekly bortezomib in combination

with gemcitabine and cisplatin in the first-line treatment of patients with advanced solid tumors. Clin Cancer Res 2007; 13: 3642–3651.

75. Wang M, Giralt S, Delasalle K, Handy B, Alexanian R. Bortezomib in combination with thalidomide-dexamethasone for previously untreated multiple myeloma. Hematology 2007; 12: 235–239.
76. Weigert O, Pastore A, Rieken M, Lang N, Hiddemann W, Dreyling M. Sequence-dependent synergy of the proteasome inhibitor bortezomib and cytarabine in mantle cell lymphoma. Leukemia 2007; 21: 524–528.
77. Yu C, Friday BB, Lai JP, Yang L, Sarkaria J, Kay NE, Carter CA, Roberts LR, Kaufmann SH, Adjei AA. Cytotoxic synergy between the multikinase inhibitor sorafenib and the proteasome inhibitor bortezomib *in vitro*: induction of apoptosis through akt and c-jun nh2-terminal kinase pathways. Mol Cancer Ther 2006; 5: 2378–2387.
78. Papageorgiou A, Kamat A, Benedict WF, Dinney C, Mcconkey DJ. Combination therapy with ifn-alpha plus bortezomib induces apoptosis and inhibits angiogenesis in human bladder cancer cells. Mol Cancer Ther 2006; 5: 3032–3041.
79. Khan T, Stauffer JK, Williams R, Hixon JA, Salcedo R, Lincoln E, Back TC, Powell D, Lockett S, Arnold AC, Sayers TJ, Wigginton JM. Proteasome inhibition to maximize the apoptotic potential of cytokine therapy for murine neuroblastoma tumors. J Immunol 2006; 176: 6302–6312.
80. Lesinski GB, Shah N, Go M, Bouchard P, Carson WE. Velcade potentiates the anti-tumor activity of interferon-alpha *in vitro* and in a murine model of malignant melanoma. Proceedings of AACR-NCI-EORTC International Conference on Molecular Targets and Cancer Therapeutics; Boston, MA; 2003.
81. Schumacher LY, Vo DD, Garban HJ, Comin-Anduix B, Owens SK, Dissette VB, Glaspy JA, Mcbride WH, Bonavida B, Economou JS, Ribas A. Immunosensitization of tumor cells to dendritic cell-activated immune responses with the proteasome inhibitor bortezomib (ps-341, velcade). J Immunol 2006; 176: 4757–4765.
82. Spisek R, Charalambous A, Mazumder A, Vesole DH, Jagannath S, Dhodapkar MV. Bortezomib enhances dendritic cell (dc)-mediated induction of immunity to human myeloma via exposure of cell surface heat shock protein 90 on dying tumor cells: therapeutic implications. Blood 2007; 109: 4839–4845.
83. Nencioni A, Garuti A, Schwarzenberg K, Cirmena G, Dal Bello G, Rocco I, Barbieri E, Brossart P, Patrone F, Ballestrero A. Proteasome inhibitor-induced apoptosis in human monocyte-derived dendritic cells. Eur J Immunol 2006; 36: 681–689.
84. Nencioni A, Schwarzenberg K, Brauer KM, Schmidt SM, Ballestrero A, Grunebach F, Brossart P. Proteasome inhibitor bortezomib modulates tlr4-induced dendritic cell activation. Blood 2006; 108: 551–558.
85. Straube C, Wehner R, Wendisch M, Bornhauser M, Bachmann M, Rieber EP, Schmitz M. Bortezomib significantly impairs the immunostimulatory capacity of human myeloid blood dendritic cells. Leukemia 2007; 21: 1464–1471.
86. Berges C, Haberstock H, Fuchs D, Miltz M, Sadeghi M, Opelz G, Daniel V, Naujokat C. Proteasome inhibition suppresses essential immune functions of human cd4(+) t cells. Immunology 2008; 124(4): 234–246.
87. Chauhan D, Catley L, Li G, Podar K, Hideshima T, Velankar M, Mitsiades C, Mitsiades N, Yasui H, Letai A, Ovaa H, Berkers C, Nicholson B, Chao TH, Neuteboom ST, Richardson P, Palladino MA, Anderson KC. A novel orally active proteasome inhibitor induces

apoptosis in multiple myeloma cells with mechanisms distinct from bortezomib. Cancer Cell 2005; 8: 407–419.

88. Kuhn DJ, Chen Q, Voorhees PM, Strader JS, Shenk KD, Sun CM, Demo SD, Bennett MK, Van Leeuwen FW, Chanan-Khan AA, Orlowski RZ. Potent activity of carfilzomib, a novel, irreversible inhibitor of the ubiquitin-proteasome pathway, against preclinical models of multiple myeloma. Blood 2007; 110: 3281–3290.
89. Isaacs A, Lindenmann J. Virus interference i: the interferon. Proc R Soc Lond Ser B 1957; 147: 258–267.
90. Isaacs A, Lindenmann J, Valentine RC. Virus interference, ii: some properties of interferon. Proc R Soc Lond 1957; 147B: 268–273.
91. Burke DC, Isaacs A. Some factors affecting the production of interferon. Br J Exp Pathol 1958; 39: 452–458.
92. Gresser I, Bourali C. Exogenous interferon and inducers of interferon in the treatment balb-c mice inoculated with rc19 tumour cells. Nature 1969; 223: 844–845.
93. Paucker K, Cantell K, Henle W. Quantitative studies on viral interference in suspended l cells, iii: effect of interfering viruses and interferon on the growth rate of cells. Virology 1962; 17: 324–334.
94. Rubinstein M, Rubinstein S, Familletti PC, Gross MS, Miller RS, Waldman AA, Pestka S. Human leukocyte interferon purified to homogeneity. Science 1978; 202: 1289–1290.
95. Taniguchi T, Fujii-Kuriyama Y, Muramatsu M. Construction and identification of a bacterial plasmid containing the human fibroblast interferon gene sequences. Proc Japan Acad 1979; 55B: 464–469.
96. Hall SS. The cloning of interferon and other mistakes. In: A commotion in the blood. New York, NY: Henry Holt and Company, Inc.; 1997. p 178–208.
97. Owerbach D, Rutter WJ, Shows TB, Gray P, Goeddel DV, Lawn RM. Leukocyte and fibroblast interferon genes are located on human chromosome 9. Proc Natl Acad Sci USA 1981; 78: 3123–3127.
98. Allen G, Diaz MO. Nomenclature of the human interferon proteins. J Interferon Cytokine Res 1996; 16: 181–184.
99. Conklin DC, Grant FJ, Rixon MW, Kindsvogel W.Interferon-e. US Patent 6329175. 2002.
100. Lafleur DW, Nardelli B, Tsareva T, Mather D, Feng P, Semenuk M, Taylor K, Buergin M, Chinchilla D, Roshke V, Chen G, Ruben SM, Pitha PM, Coleman TA, Moore PA. Interferon-kappa, a novel type i interferon expressed in human keratinocytes. J Biol Chem 2001; 276: 39765–39771.
101. Hancock BW, Wheatley K, Harris S, Ives N, Harrison G, Horsman JM, Middleton MR, Thatcher N, Lorigan PC, Marsden JR, Burrows L, Gore M. Adjuvant interferon in high-risk melanoma: the aim high study--united kingdom coordinating committee on cancer research randomized study of adjuvant low-dose extended-duration interferon alfa-2a in high-risk resected malignant melanoma. J Clin Oncol 2004; 22: 53–61.
102. Kirkwood JM, Ibrahim JG, Sondak VK, Richards J, Flaherty LE, Ernstoff MS, Smith TJ, Rao U, Steele M, Blum RH. High- and low-dose interferon alfa-2b in high-risk melanoma: first analysis of intergroup trial e1690/s9111/c9190. J Clin Oncol 2000; 18: 2444–2458.
103. Kirkwood JM, Ibrahim JG, Sosman JA, Sondak VK, Agarwala SS, Ernstoff MS, Rao U. High-dose interferon alfa-2b significantly prolongs relapse-free and overall survival compared with the gm2-klh/qs-21 vaccine in patients with resected stage iib-iii mela-

noma: results of intergroup trial e1694/s9512/c509801. J Clin Oncol 2001; 19: 2370–2380.

104. Burchert A, Neubauer A. Interferon alpha and t-cell responses in chronic myeloid leukemia. Leuk Lymphoma 2005; 46: 167–175.
105. Krown SE. Interferon-alpha: evolving therapy for aids-associated kaposi's sarcoma. J Interferon Cytokine Res 1998; 18: 209–214.
106. Dorr RT, Salmon SE, Robertone A, Bonnem E. Phase i-ii trial of interferon-alpha 2b by continuous subcutaneous infusion over 28 days. J Interferon Res 1988; 8: 717–725.
107. Balch C, Houghton A, Sober A, Soong S-J. Biologic therapy. In: Balch C, Houghton A, Sober A, Soong S-J, editors. Cutaneous melanoma. St. Louis: Quality Medical Publishing; 1998. p 419–436.
108. Agarwala S. Improving survival in patients with high-risk and metastatic melanoma: immunotherapy leads the way. Am J Clin Dermatol 2003; 4: 333–346.
109. Kirkwood JM, Ernstoff MS, Davis CA, Reiss M, Ferraresi R, Rudnick SA. Comparison of intramuscular and intravenous recombinant alpha-2 interferon in melanoma and other cancers. Ann Intern Med 1985; 103: 32–36.
110. Legha SS. Interferons in the treatment of malignant melanoma. A review of recent trials. Cancer 1986; 57: 1675–1677.
111. Pyrhonen S, Salminen E, Ruutu M, Lehtonen T, Nurmi M, Tammela T, Juusela H, Rintala E, Hietanen P, Kellokumpu-Lehtinen PL. Prospective randomized trial of interferon alfa-2a plus vinblastine versus vinblastine alone in patients with advanced renal cell cancer. J Clin Oncol 1999; 17: 2859–2867.
112. Steineck G, Strander H, Carbin BE, Borgström E, Wallin L, Achtnich U, Arvidsson A, Söderlund V, Näslund I, Esposti PL, Norell SE. Recombinant leukocyte interferon alpha-2a and medroxyprogesterone in advanced renal cell carcinoma. A randomized trial. Acta Oncol 1990; 29: 155–162.
113. Kirkwood JM, Strawderman MH, Ernstoff MS, Smith TJ, Borden EC, Blum RH. Interferon alfa-2b adjuvant therapy of high-risk resected cutaneous melanoma: the eastern cooperative oncology group trial est 1684. J Clin Oncol 1996; 14: 7–17.
114. Kirkwood J. Cancer immunotherapy: the interferon-alpha experience. Semin Oncol 2002; 29: 18–26.
115. Schuchter LM. Adjuvant interferon therapy for melanoma: high-dose, low-dose, no dose, which dose? J Clin Oncol 2004; 22: 7–10.
116. Lesinski GB, Kondadasula SV, Crespin T, Shen L, Kendra K, Walker MJ, Carson WE. Multiparametric flow cytometric analysis of inter-patient variation in stat1 phosphorylation following interferon alfa immunotherapy. J Natl Cancer Inst 2004; 96: 1331–1342.
117. Zimmerer JM, Lehman AM, Ruppert AS, Noble CW, Olencki T, Walker MJ, Kendra K, Carson WE. 3rd. Ifn-alpha-2b-induced signal transduction and gene regulation in patient peripheral blood mononuclear cells is not enhanced by a dose increase from 5 to 10 megaunits/m2. Clin Cancer Res 2008; 14: 1438–1445.
118. Levy DE, Gilliland DG. Divergent roles of stat1 and stat5 in malignancy as revealed by gene disruptions in mice. Oncogene 2000; 19: 2505–2510.
119. Thyrell L, Erickson S, Zhivotovsky B, Pokrovskaja K, Sangfelt O, Castro J, Einhorn S, Grander D. Mechanisms of interferon-alpha induced apoptosis in malignant cells. Oncogene 2002; 21: 1251–1262.

120. Thyrell L, Hjortsberg L, Arulampalam V, Panaretakis T, Uhles S, Dagnell M, Zhivotovsky B, Leibiger I, Grander D, Pokrovskaja K. Interferon alpha-induced apoptosis in tumor cells is mediated through the phosphoinositide 3-kinase/mammalian target of rapamycin signaling pathway. J Biol Chem 2004; 279: 24152–24162.

121. Dinney CP, Bielenberg DR, Perrotte P, Reich R, Eve BY, Bucana CD, Fidler IJ. Inhibition of basic fibroblast growth factor expression, angiogenesis, and growth of human bladder carcinoma in mice by systemic interferon-alpha administration. Cancer Res 1998; 58: 808–814.

122. Ezekowitz RA, Mulliken JB, Folkman J. Interferon alfa-2a therapy for life-threatening hemangiomas of infancy. N Engl J Med 1992; 326: 1456–1463.

123. Martin-Henao GA, Quiroga R, Sureda A, Gonzalez JR, Moreno V, Garcia J. L-selectin expression is low on cd34+ cells from patients with chronic myeloid leukemia and interferon-a up-regulates this expression. Haematologica 2000; 85: 139–146.

124. Von Stamm U, Brocker EB, Von Depka Prondzinski M, Ruiter DJ, Rumke P, Broding C, Carrel S, Lejeune FJ. Effects of systemic interferon-alpha (ifn-alpha) on the antigenic phenotype of melanoma metastases. Eortc melanoma group cooperative study no. 18852. Melanoma Res 1993; 3: 173–180.

125. Belardelli F, Gresser I, Maury C, Duvillard P, Prade M, Maunoury MT. Antitumor effects of interferon in mice injected with interferon-sensitive and interferon-resistant friend leukemia cells. III. Inhibition of growth and necrosis of tumors implanted subcutaneously. Int J Cancer 1983; 31: 649–653.

126. Belardelli F, Gresser I, Maury C, Maunoury MT. Antitumor effects of interferon in mice injected with interferon-sensitive and interferon-resistant friend leukemia cells. II. Role of host mechanisms. Int J Cancer 1982; 30: 821–825.

127. Belardelli F, Gresser I, Maury C, Maunoury MT. Antitumor effects of interferon in mice injected with interferon-sensitive and interferon-resistant friend leukemia cells. Int J Cancer 1982; 30: 813–820.

128. Gresser I. Production of inteferon by suspensions of human leukocytes. Proc Soc Exp Biol Med 1961; 108: 799–803.

129. Gresser I, Belardelli F, Maury C, Tovey MG, Maunoury MT. Anti-tumor effects of interferon in mice injected with interferon-sensitive and interferon-resistant friend leukemia cells. IV. Definition of optimal treatment regimens. Int J Cancer 1986; 38: 251–257.

130. Gresser I, Carnaud C, Maury C, Sala A, Eid P, Woodrow D, Maunoury MT, Belardelli F. Host humoral and cellular immune mechanisms in the continued suppression of friend erythroleukemia metastases after interferon alpha/beta treatment in mice. J Exp Med 1991; 173: 1193–1203.

131. Gresser I, Maury C, Belardelli F, Maunoury MT, Machover D. Effectiveness of mouse interferon alpha/beta compared to single-agent chemotherapy in increasing survival time of mice after intravenous inoculation of friend erythroleukemia cells. J Natl Cancer Inst 1988; 80: 126–131.

132. Gresser I, Maury C, Carnaud C, De Maeyer E, Maunoury MT, Belardelli F. Anti-tumor effects of interferon in mice injected with interferon-sensitive and interferon-resistant friend erythroleukemia cells. VIII. Role of the immune system in the inhibition of visceral metastases. Int J Cancer 1990; 46: 468–474.

133. Badgwell B, Lesinski GB, Magro C, Abood G, Skaf A, Carson WE. The anti-tumor effects of interferon-alpha are maintained in mice challenged with a stat1-deficient murine melanoma cell line. J Surg Res 2003; 116: 129–136.

134. Lesinski GB, Anghelina M, Zimmerer J, Bakalakos T, Badgwell B, Parihar R, Hu Y, Becknell B, Abood G, Raychaudhury A, Magro C, Durbin J, Carson WE. The anti-tumor effects of interferon-alpha are abrogated in a stat1-deficient mouse. J Clin Invest 2003; 112: 170–180.

135. Zimmerer JM, Lesinski GB, Kondadasula SV, Karpa VI, Lehman A, Raychaudhury A, Becknell B, Carson WE. 3rd. Ifn-alpha-induced signal transduction, gene expression, and antitumor activity of immune effector cells are negatively regulated by suppressor of cytokine signaling proteins. J Immunol 2007; 178: 4832–4845.

136. Moschos SJ, Edington HD, Land SR, Rao UN, Jukic D, Shipe-Spotloe J, Kirkwood JM. Neoadjuvant treatment of regional stage iiib melanoma with high-dose interferon alfa-2b induces objective tumor regression in association with modulation of tumor infiltrating host cellular immune responses. J Clin Oncol 2006; 24: 3164–3171.

137. Gogas H, Ioannovich J, Dafni U, Stavropoulou-Giokas C, Frangia K, Tsoutsos D, Panagiotou P, Polyzos A, Papadopoulos O, Stratigos A, Markopoulos C, Bafaloukos D, Pectasides D, Fountzilas G, Kirkwood JM. Prognostic significance of autoimmunity during treatment of melanoma with interferon. N Engl J Med 2006; 354: 709–718.

138. Atkins MB. Cytokine-based therapy and biochemotherapy for advanced melanoma. Clin Cancer Res 2006; 12: 2353s–2358s.

139. Melichar B, Koralewski P, Ravaud A, Pluzanska A, Bracarda S, Szczylik C, Chevreau C, Filipek M, Delva R, Sevin E, Negrier S, Mckendrick J, Santoro A, Pisa P, Escudier B. First-line bevacizumab combined with reduced dose interferon-{alpha}2a is active in patients with metastatic renal cell carcinoma. Ann Oncol 2008; 19(8): 1470–1476.

140. Varker KA, Biber JE, Kefauver C, Jensen R, Lehman A, Young D, Wu H, Lesinski GB, Kendra K, Chen HX, Walker MJ, Carson WE. 3rd. A randomized phase 2 trial of bevacizumab with or without daily low-dose interferon alfa-2b in metastatic malignant melanoma. Ann Surg Oncol 2007; 14: 2367–2376.

141. Bukowski RM, Tendler C, Cutler D, Rose E, Laughlin MM, Statkevich P. Treating cancer with peg intron: pharmacokinetic profile and dosing guidelines for an improved interferon-alpha-2b formulation. Cancer 2002; 95: 389–396.

142. Liu TC, Galanis E, Kirn D. Clinical trial results with oncolytic virotherapy: a century of promise, a decade of progress. Nat Clin Pract Oncol 2007; 4: 101–117.

143. Martuza RL, Malick A, Markert JM, Ruffner KL, Coen DM. Experimental therapy of human glioma by means of a genetically engineered virus mutant. Science 1991; 252: 854–856.

144. Goldstein DJ, Weller SK. Herpes simplex virus type 1-induced ribonucleotide reductase activity is dispensable for virus growth and DNA synthesis: isolation and characterization of an icp6 lacz insertion mutant. J Virol 1988; 62: 196–205.

145. Kasuya H, Nishiyama Y, Nomoto S, Hosono J, Takeda S, Nakao A. Intraperitoneal delivery of hrr3 and ganciclovir prolongs survival in mice with disseminated pancreatic cancer. J Surg Oncol 1999; 72: 136–141.

146. Mineta T, Rabkin SD, Martuza RL. Treatment of malignant gliomas using ganciclovir-hypersensitive, ribonucleotide reductase-deficient herpes simplex viral mutant. Cancer Res 1994; 54: 3963–3966.

147. Boviatsis EJ, Park JS, Sena-Esteves M, Kramm CM, Chase M, Efird JT, Wei MX, Breakefield XO, Chiocca EA. Long-term survival of rats harboring brain neoplasms treated with ganciclovir and a herpes simplex virus vector that retains an intact thymidine kinase gene. Cancer Res 1994; 54: 5745–5751.

148. Aghi M, Visted T, Depinho RA, Chiocca EA. Oncolytic herpes virus with defective icp6 specifically replicates in quiescent cells with homozygous genetic mutations in p16. Oncogene 2008.

149. Chou J, Kern ER, Whitley RJ, Roizman B. Mapping of herpes simplex virus-1 neurovirulence to gamma 134.5, a gene nonessential for growth in culture. Science 1990; 250: 1262–1266.

150. Bolovan CA, Sawtell NM, Thompson RL. Icp34.5 mutants of herpes simplex virus type 1 strain 17syn + are attenuated for neurovirulence in mice and for replication in confluent primary mouse embryo cell cultures. J Virol 1994; 68: 48–55.

151. Mckie EA, Maclean AR, Lewis AD, Cruickshank G, Rampling R, Barnett SC, Kennedy PG, Brown SM. Selective *in vitro* replication of herpes simplex virus type 1 (hsv-1) icp34.5 null mutants in primary human cns tumours--evaluation of a potentially effective clinical therapy. Br J Cancer 1996; 74: 745–752.

152. Lasner TM, Kesari S, Brown SM, Lee VM, Fraser NW, Trojanowski JQ. Therapy of a murine model of pediatric brain tumors using a herpes simplex virus type-1 icp34.5 mutant and demonstration of viral replication within the cns. J Neuropathol Exp Neurol 1996; 55: 1259–1269.

153. Kesari S, Randazzo BP, Valyi-Nagy T, Huang QS, Brown SM, Maclean AR, Lee VM, Trojanowski JQ, Fraser NW. Therapy of experimental human brain tumors using a neuroattenuated herpes simplex virus mutant. Lab Invest 1995; 73: 636–648.

154. Rampling R, Cruickshank G, Papanastassiou V, Nicoll J, Hadley D, Brennan D, Petty R, Maclean A, Harland J, Mckie E, Mabbs R, Brown M. Toxicity evaluation of replication-competent herpes simplex virus (icp 34.5 null mutant 1716) in patients with recurrent malignant glioma. Gene Ther 2000; 7: 859–866.

155. Harland J, Papanastassiou V, Brown SM. Hsv1716 persistence in primary human glioma cells *in vitro*. Gene Ther 2002; 9: 1194–1198.

156. Papanastassiou V, Rampling R, Fraser M, Petty R, Hadley D, Nicoll J, Harland J, Mabbs R, Brown M. The potential for efficacy of the modified (icp 34.5(-)) herpes simplex virus hsv1716 following intratumoural injection into human malignant glioma: a proof of principle study. Gene Ther 2002; 9: 398–406.

157. Mackie RM, Stewart B, Brown SM. Intralesional injection of herpes simplex virus 1716 in metastatic melanoma. Lancet 2001; 357: 525–526.

158. Crusade-Laboratories.

159. Mineta T, Rabkin SD, Yazaki T, Hunter WD, Martuza RL. Attenuated multi-mutated herpes simplex virus-1 for the treatment of malignant gliomas. Nat Med 1995; 1: 938–943.

160. Warren RS, Yuan H, Matli MR, Gillett NA, Ferrara N. Regulation by vascular endothelial growth factor of human colon cancer tumorigenesis in a mouse model of experimental liver metastasis. J Clin Invest 1995; 95: 1789–1797.

161. Melnyk O, Zimmerman M, Kim KJ, Shuman M. Neutralizing anti-vascular endothelial growth factor antibody inhibits further growth of established prostate cancer and metastases in a pre-clinical model. J Urol 1999; 161: 960–963.

162. Asano M, Yukita A, Matsumoto T, Kondo S, Suzuki H. Inhibition of tumor growth and metastasis by an immunoneutralizing monoclonal antibody to human vascular endothelial growth factor/vascular permeability factor121. Cancer Res 1995; 55: 5296–5301.

163. Millauer B, Shawver LK, Plate KH, Risau W, Ullrich A. Glioblastoma growth inhibited *in vivo* by a dominant-negative flk-1 mutant. Nature 1994; 367: 576–579.

164. Wood JM, Bold G, Buchdunger E, Cozens R, Ferrari S, Frei J, Hofmann F, Mestan J, Mett H, O'reilly T, Persohn E, Rosel J, Schnell C, Stover D, Theuer A, Towbin H, Wenger F, Woods-Cook K, Menrad A, Siemeister G, Schirner M, Thierauch KH, Schneider MR, Drevs J, Martiny-Baron G, Totzke F. Ptk787/zk 222584, a novel and potent inhibitor of vascular endothelial growth factor receptor tyrosine kinases, impairs vascular endothelial growth factor-induced responses and tumor growth after oral administration. Cancer Res 2000; 60: 2178–2189.

165. Wedge SR, Ogilvie DJ, Dukes M, Kendrew J, Curwen JO, Hennequin LF, Thomas AP, Stokes ES, Curry B, Richmond GH, Wadsworth PF. Zd4190: an orally active inhibitor of vascular endothelial growth factor signaling with broad-spectrum antitumor efficacy. Cancer Res 2000; 60: 970–975.

166. Wedge SR, Kendrew J, Hennequin LF, Valentine PJ, Barry ST, Brave SR, Smith NR, James NH, Dukes M, Curwen JO, Chester R, Jackson JA, Boffey SJ, Kilburn LL, Barnett S, Richmond GH, Wadsworth PF, Walker M, Bigley AL, Taylor ST, Cooper L, Beck S, Jurgensmeier JM, Ogilvie DJ. Azd2171: a highly potent, orally bioavailable, vascular endothelial growth factor receptor-2 tyrosine kinase inhibitor for the treatment of cancer. Cancer Res 2005; 65: 4389–4400.

167. Prewett M, Huber J, Li Y, Santiago A, O'connor W, King K, Overholser J, Hooper A, Pytowski B, Witte L, Bohlen P, Hicklin DJ. Antivascular endothelial growth factor receptor (fetal liver kinase 1) monoclonal antibody inhibits tumor angiogenesis and growth of several mouse and human tumors. Cancer Res 1999; 59: 5209–5218.

168. Fox WD, Higgins B, Maiese KM, Drobnjak M, Cordon-Cardo C, Scher HI, Agus DB. Antibody to vascular endothelial growth factor slows growth of an androgen-independent xenograft model of prostate cancer. Clin Cancer Res 2002; 8: 3226–3231.

169. Segerstrom L, Fuchs D, Backman U, Holmquist K, Christofferson R, Azarbayjani F. The anti-vegf antibody bevacizumab potently reduces the growth rate of high-risk neuroblastoma xenografts. Pediatr Res 2006; 60: 576–581.

170. Inoue S, Hartman A, Branch CD, Bucana CD, Bekele BN, Stephens LC, Chada S, Ramesh R. Mda-7 in combination with bevacizumab treatment produces a synergistic and complete inhibitory effect on lung tumor xenograft. Mol Ther 2007; 15: 287–294.

171. Guse K, Ranki T, Ala-Opas M, Bono P, Sarkioja M, Rajecki M, Kanerva A, Hakkarainen T, Hemminki A. Treatment of metastatic renal cancer with capsid-modified oncolytic adenoviruses. Mol Cancer Ther 2007; 6: 2728–2736.

172. Salvador C, Li B, Hansen R, Cramer DE, Kong M, Yan J. Yeast-derived beta-glucan augments the therapeutic efficacy mediated by anti-vascular endothelial growth factor monoclonal antibody in human carcinoma xenograft models. Clin Cancer Res 2008; 14: 1239–1247.

173. Pineda-Roman M, Zangari M, Van Rhee F, Anaissie E, Szymonifka J, Hoering A, Petty N, Crowley J, Shaughnessy J, Epstein J, Barlogie B. Vtd combination therapy with bortezomib-thalidomide-dexamethasone is highly effective in advanced and refractory multiple myeloma. Leukemia 2008.

# *Karl Fischer Titration: Determination of Water Content in Pharmaceuticals*

**Dr. Katrin Schoeffski and Helga Hoffmann**
*Sigma-Aldrich Laborchemikalien GmbH, Seelze, Germany*

## 1 BASICS

The Karl Fischer titration is based on the reaction from iodine with sulfur dioxide in alcohol in the presence of water.

During the determination water reacts stoichiometric with iodine in presence of an excess of sulfur dioxide. The reaction investigated thoroughly by Scholz (1) proceeds in two steps:

$$\text{1st step:} \quad SO_2 + \text{R-OH} + B \Leftrightarrow \text{R-}SO_3^- + BH^+ \tag{1a}$$

$$\text{2nd step:} \quad RSO_3^- + I_2 + H_2O + 2B \Leftrightarrow RSO_4^- + 2BH^+ + 2I^- \tag{1b}$$

B = Base as imidazole

One mole iodine reacts with one mole water.

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing,*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

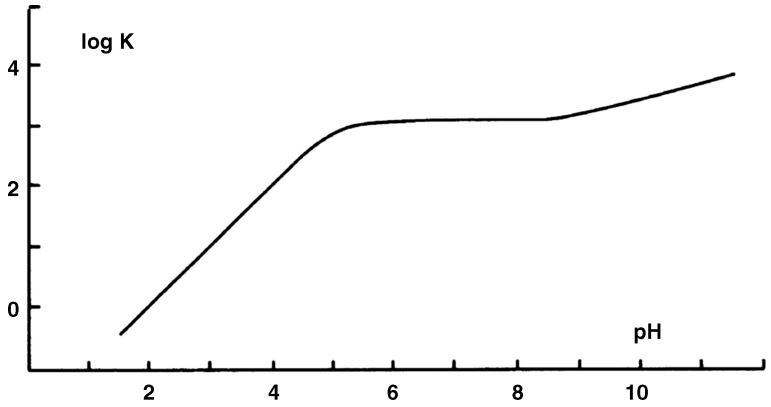


***FIGURE 1*** *pH dependence of the reaction rate constant of the Karl Fischer titration (2).*

## 2 IMPORTANCE OF THE pH

As shown in the equation mentioned above, protons were generated and have to be buffered during the Karl Fischer reaction. So the reaction rate of the Karl Fischer reaction is dependent of the pH. In Fig. 1 this dependence is shown.

A high reaction speed is necessary for the Karl Fischer titration since a long titration time enlarges the error caused by invading humidity from outside. In Fig. 1, it can be seen that a pH lower than 6 should be avoided since the titration is too slow. On the other hand at pH values higher than 8 a side reaction occurs. Therefore, the base used in Karl Fischer reagents is very important. It must be strong enough to push the equilibrium of the reaction Eq. 1b to the right side. But the base must also not too strong, since beginning at pH 8 the iodine starts to decompose and pretends a higher water content or hinders the system to find a stable end point. Imidazole and its substitutes have been found as ideal Karl Fischer bases. They keep the pH value of Karl Fischer reagents in the ideal range of 5–7. Strong acid or basic samples can also influence the pH in the titration vessel. To hold the pH value constant a buffer is necessary. For acids the use of imidazole as buffer has been successful, salicylic acid or benzoic acid are used for strong basic samples to keep the pH on an ideal level (3).

## 3 SIDE-REACTIONS OF SAMPLES WITH KARL FISCHER REAGENTS

Due to the nature of the Karl Fischer reagent, a variety of side-reactions of the reagent with the sample are possible. The Karl Fischer reagent consists of iodine, an organic base and sulfur dioxide in an alcoholic solvent. The first group of side-reactions occurs of substances, which undergo a redox reaction with the iodine/iodide complex as peroxides or unsaturated hydrocarbons. Several methods can hinder the oxidation. The change of the pH may be useful, masking of double bonds another possibility.

Whether a sample reacts with iodine can be found out easily when a methanolic iodine solution is dropped into the sample. In case of decoloring a reaction takes place and a special sample preparation is necessary.

Another possible side-reaction is the reaction with sulfur dioxide. The bisulfite addition of aldehydes is a common chemical reaction, which belongs to this class. This is one of the rare cases, where the water content found can be too low. The aldehyde, ingredients of the reagent and some of the water of the sample form a "bisulfite adduct". In this case dedicated reagents, a suitable titration agent and a dedicated medium for the titration vessel as well have to be used. The most frequent side-reaction of a Karl Fischer reagent during Karl Fischer titration is the formation of acetals and ketals respectively, from carbonyl compounds. The reaction forms water, which is also titrated, resulting in vanishing end points, and an erroneously high water content occurs. In some cases no end point can be found. This side-reaction can be suppressed when methanol is replaced by other suitable alcohols. Depending on the reactivity of the carbonyl compounds in the sample also a titration reagent developed especially for the titration of ketones and aldehydes is necessary. Coulometric titration is NOT recommended for water determination in these samples.

## 4 KARL FISCHER TITRATORS

Several manufacturers of laboratory equipment offer special Karl Fischer titrators. The end point of the titration is detected potentiometrically. On a double-platinum electrode a defined current is laid, the required potential is measured. When the measured potential goes rapidly down to a very low voltage the end point is reached. Commercial available volumetric Karl Fischer titrators work with one and two component reagents as well. Pre-titration starts automatically, and the result is calculated and shown on a display. Also Karl Fischer coulometers are offered. They automatically dry the coulometric cell after addition of the reagent and keep it dry during the stand-by. In both types of titrators standard methods and calculations are already implemented. For automation sample changers can be used. Karl Fischer ovens can be connected to many of the commercial available volumetric and coulometric Karl Fischer titrators. And at last, documentation systems are also available which enables the user to keep all records according to quality and legislative requirements.

## 5 KARL FISCHER REAGENTS

Decades ago, laboratory personnel were expected to prepare Karl Fischer reagents of their own. The unpleasant smell of pyridine, the rapid loss of titer, and the great ineffectively made this duty very unpopular so most laboratories have changed to commercial reagents in the meantime. These reagents have surpassed the popularity of home-made Karl Fischer reagents in many application areas. They contain more effective bases and are characterized by a rapid course of titration. In addition a more

stable end point is caused by the higher reaction speed. A better accuracy is also reached with modern high-speed reagents. An additional benefit for the user is caused since the reagent suppliers improve the safety for the laboratory personnel by eliminating toxic and unpleasant ingredients.

The following reagent depictions give also an overview about the typical "Karl Fischer nomenclature" for the various types of Karl Fischer titration techniques.

The volumetric one-component titration reagent is called "Composite". It contains all active ingredients iodine, sulfur dioxide, and a suitable mixture of bases in an appropriate solvent. The number 5, 2, and 1, respectively, behind the "Composite" gives the titer in mg/mL water. In the titration vessel a "Working Medium" is filled. This can be pure methanol or a ready-made mixture of various solvents, e.g., to improve the solubility of the sample or to suppress side-reactions. Buffer can also be included.

For the two-component titration a solution of iodine in a suitable alcohol called "Titrant" is filled into the buret, the titration vessel is equipped with the "Solvent", a mixture of sulfur dioxide and the base in a suitable alcohol.

For coulometric titration the anolyte or "Coulomat A" is filled into the anodic compartment of the coulometric cell. It consists of iodide, sulfur dioxide, bases, and appropriate solvents. The catholyte or "Coulomat C" is filled into the cathodic compartment of the diaphragm cell and consists of a special mixture of salts in methanol, which can be reduced in a cathodic electrochemical reaction.

Special reagents for certain groups of samples can also be identified by the name. A "K" indicates reagents for ketones and aldehydes. "Oil" displays the suitability for non-polar samples. The suffix "E" indicates a reagent based on ethanol, which is less hazardous and more environmentally friendly. "Standards" are substances or solvent mixtures with a dedicated water content for the calibration of Karl Fischer reagents, for the control of instruments according to ISO 9000, for the validation of procedures and for the suitability tests according to Ph.Eur. as well.

The latest developments are reagents based on ethanol. This solvent is significantly less toxic than methanol and has better solubility properties for non-polar samples. (4)

## 6 STANDARD PROCEDURES FOR KARL FISCHER TITRATIONS

As already mentioned the titration can be performed volumetrically and coulometrically. Typical for the Karl Fischer titration is the so-called pre-titration. Since humidity is everywhere in the atmosphere, the equipment and also reagents can never be absolute water-free. When a Karl Fischer instrument is filled with the reagent or working medium, in the first step the cell must be titrated to dryness. In the second step the sample is added and its water content is titrated.

In the volumetric one-component titration the reagent must be filled into a buret. In the titration vessel the working medium is added. It must contain at least 25–50% methanol, since the alcohol is not only a solvent but takes also part in the reaction mentioned above. The other 50–75% can be a large variety of solubilizers as formamide, halogenated solvents or long-chained alcohols. The used quantity

depends on the nature of the solubilizer. The working medium is titrated to dryness (pre-titration) in the first step, then the sample is added and the water content is determined. The one-component titration has the advantage that the working medium in the titration vessel can be varied according to the requirements of the sample.

In the volumetric two-component titration the buret is filled with the "Titrant". In the titration vessel the Karl Fischer Solvent is added. The solvent in the vessel is titrated to dryness, then the sample is added and the water content is determined. Two-component titrations have a rapid titration rate and highly accurate results.

For both types of volumetric Karl Fischer titration, the titration with a one-component reagent and a two-component reagent, respectively, a titer determination is necessary.

The Karl Fischer anolyte consists of iodide as active ingredient. The iodine necessary for the Karl Fischer reaction is generated by anodic oxidation of iodide to iodine. The iodine reacts at once with the water according to Eq. 1. The amount of water titrated is proportional to the total current consumption and can be calculated using Faraday's law. For this type of water determination two coulometric cell types are in use, a cell with diaphragm, consisting of an anodic and a cathodic compartment, and a cell without diaphragm, respectively. The so-called diaphragmless cell consists of one compartment, the anodic compartment only. For a coulometric titration the anodic Karl Fischer reagent is filled in the large anodic compartment of the coulometric cell. The cell with diaphragm requires the catholyte, which has to be filled into the small cathodic compartment. The compartment is separated from the anodic compartment with a ceramic diaphragm. In the diapragmless cell, a special cathodic reagent is not necessary. In the first step the reagent is automatically titrated to dryness. Then the sample is added and the water content is determined. Since the coulometric titration is much more sensitive than the volumetric titration the sample has to be added using a syringe. A rubber seal is pricked by the tip of the syringe. Using this procedure the sample can be added without transfer of air humidity into the cell. In coulometric titration a titer determination is not necessary since the current is measured directly.

A third form of Karl Fischer titration is the non-direct titration using Karl Fischer oven technique. There are samples, which should not have direct contact with a Karl Fischer reagent. In some cases, they react with the ingredients of the reagent, sometimes they are not soluble in the typical Karl Fischer solubilizers. Those samples can be analyzed using the Karl Fischer oven. Depending on type of oven, the sample is weighted in a vial or small bottle, then it is heated at a suitable temperature until the water is evaporated and transferred by a carrier gas into the volumetric or coulometric Karl Fischer cell. There the water is determined according to the standard techniques.

## 7 THE KARL FISCHER TITRATION IN THE NATIONAL PHARMACOPOEIAS

The national pharmacopoeias describe the Karl Fischer titration as a method for water determination. Traditionally a one-component reagent is preferred in the

pharmaceutical quality control. But in the latest editions, other modern techniques have also found their way into the pharmacopoeias. Here we will give a short overview of the recommendations.

## 7.1 USP (United States Pharmacopoeia), 921 Water Determination (5)

It describes a titrimetric, gravimetric and destillation method for water determination. The titrimetric Karl Fischer titration is divided into direct volumetric, volumetric residue and coulometric titration. A composition instruction for a pyridine-containing reagent is given but the reagents can be purchased as well. Bases and solvents in the commercially available reagents can be varied in a broad rage so that pyridine and methanol are not necessary. A detailed standardization procedure is given which prescribes the use of sodium tartrate dihydrate as water reference substance. For the coulometric Karl Fischer titration commercially available apparatus and reagents are recommended and not specified in detail.

## 7.2 European Pharmacopoeia (6)

***7.2.1 Water: Semi-Micro Determination*** First in the European Pharmacopoeia a volumetric titration method is prescribed. Direct titration and residue titration is possible. The reagent (so-called "titrant") is not given in detail, each commercially available reagent is possible which fulfils the requirements of the standardization procedure. Water or a certified reference material must be used for standardization, the titer of the reagent must be not less than 80% of that is indicated by the supplier. The accuracy of the determination with the chosen titrant must be verified for each substance to be examined. First time a detailed suitability test is prescribed in the monograph 5.7 (2007).

***7.2.2 Water: Micro Determination*** On this page the coulometric Karl Fischer titration is listed. The apparatus consists of a cell with diaphragm, the reagent composition is not given in detail, and commercially available reagents are indicated to be suitable.

## 7.3 Japanese Pharmacopoeia, 2.48 Water Determination (Karl Fischer Method) (7)

***7.3.1 Volumetric Titration*** In the Japanese Pharmacopoeia detailed preparation methods for three various "Karl Fischer Titrating Solutions" (TS) are prescribed. Similar reagents with the mentioned ingredients can be purchased by reagent suppliers. Also a standardization procedure for the reagent is given which uses a methanol–water solution as standardization reagent. The text gives detailed procedures for direct titration, residue titration, direct extraction of insoluble samples and, at least gas extraction using the Karl Fischer oven technique.

***7.3.2 Coulometric Titration*** For coulometric Karl Fischer titration the use of a cell with diaphragm is recommended. The preparation of three various anolyte and catholyte solutions is given in detail. Similar solutions are offered by common reagent suppliers so the work of preparation of own reagents can be saved. The technique of direct coulometric titration is prescribed, for insoluble samples the Karl Fischer oven technique is recommended.

## 8 PRACTICE OF THE KARL FISCHER TITRATION

In this chapter, standard titration techniques and their variations for particular samples will be shown. For each technique a typical example out of the field of pharmaceutical chemistry is given.

### 8.1 Direct Titration with a One-Component Reagent: Example Ipatropium Bromide

Thirty milliliter of dry methanol or another appropriate solvent are placed in the titration vessel and titrated to dryness with Karl Fischer reagent Composite 5. Around 500 mg of the sample, exactly weighed by difference, is weighed in and titrated with Composite 5.

The analysis can also be done with the two component reagent.

The same procedure is usable for Cefalexine, Polysorbate 80, Povidone, Oxytetracycline hydrochloride, Sodium acetate trihydrate, Atropinsulfate hydrate, Dexpanthenol, Promethazine hydrochloride, Salbutamol hemisulfate, or Poloxamere.

### 8.2 Direct Titration with a One-Component Reagent in a Buffered Working Medium: Example Hexetidine

Thirty milliliter of methanol and 5 g of salicylic acid are placed in the titration vessel and titrated to dryness with Composite 5. Around 500 mg of the sample, exactly weighed by difference, is weighed in and titrated.

The analysis cannot be done with a two-component reagent.

The same procedure is usable for Penicilline, Dobutamine hydrochloride, or Calcium nitrate tetrahydrate.

### 8.3 Direct Titration Using a Solubilizer for Non-Polar Samples: Example Sorbitan Monostearate

Thirty milliliter of a mixture of methanol and chloroform are placed in the titration vessel and titrated to dryness with Composite 5. A sample of 1 g is weighed in exactly using differential weighing. It dissolves quickly and the water content is titrated with Composite 5.

If chloroform should be avoided for environmental reasons the working medium could also made of a mixture of methanol and hexanol or of an ethanolic working medium. In both cases the titration vessel must be heated to dissolve the sample

completely. The same procedure can be used for Peanut oil, Almond oil, Cellulose acetate phtalate, Fluticasone proprionate and Hexane.

## 8.4 Direct Titration Using a Solubilizer for Hydrophilic Samples: Example l-Leucovorin

The sample does not dissolve in usual Karl Fischer working media. A mixture of methanol and formamide and the addition of salicylic acid give appropriate results.

The titration vessel is filled with 20 mL methanol, 20 mL formamide and 7 g salicylic acid and titrated to dryness with Composite 5.

Approximately 0.1 g sample is accurately weighed in using the "weighing-by-difference" technique and titrated with Composite 5. The titration vessel can be heated to achieve acceleration. The same procedure can be used for Caciumfolinate and Cloracilline sodium salt hydrate.

## 8.5 Direct Titration Using Karl Fischer Solvent in the Titration Vessel: Example Salbutamol Hemisulfate

Salbutamol hemisulfate does not dissolve in methanol. When the more polar Karl Fischer Solvent (containing sulfur dioxide and a base in methanol, see Chapter 5) is used, the substance dissolves rapidly.

Thirty milliliter Karl Fischer Solvent are placed in the titration vessel and titrated to dryness with Titrant 5 or Composite 5. Approximately 0.3 g sample, determined by differential weighing, is precisely weighed in and the water content is titrated with Titrant or Composite.

A titration in methanol or ethanolic working media is also possible, the sample does not dissolve completely but the water is extracted.

## 8.6 Direct Titration of Acids Using a Buffered Working Medium: Example Folic Acid

Folic acid does not dissolve in the usual Karl Fischer working media. The titration times are very long. When a mixture of sulfur dioxide and imidazol in methanol is used in the titration vessel, the extraction of the water is highly improved.

Ten milliliter methanol and 20 mL Karl Fischer Buffer Acid are filled into the titration vessel and titrated to dryness with Composite 5. Approximately 0.2 g sample, exactly weighed by difference, is added into the titration vessel. The water content is determined with Composite 5.

The same procedure can be used for adenosine triphosphate.

## 8.7 Direct Titration with Suppressed Side-Reaction: Example Penicillin

Pure penicillin does not undergo a side-reaction with Karl Fischer reagents. But in respect to its production process free penicillin acid is found as impurity in many

commercial penicillins. The penicillin acid reacts with iodine and therefore must be neutralized.

Thirty milliliter methanol and 5 g of salicylic acid are placed in the titration vessel and titrated to dryness with Composite 5. Then approximately 0.5 g sample is weighed in accurately by back-weighing and its water content titrated with Composite 5. The same procedure can be used for Beserazid hydrochloride and Trolamine.

## 8.8 Direct Titration of Ketones and Aldehydes: Example Acetone

As mentioned in Chapter 3 ketones and aldehydes need particular reagents since they tend to undergo side-reactions with Karl Fischer reagents. Methanol should be avoided. Suitable solvents for the Karl Fischer titration in ketones are chloroethanol (caution: highly toxic!), less toxic are mixtures of chloroform and methoxypropanol.

Thirty milliliter Karl Fischer Medium K is filled in the titration vessel and titrated to dryness with a special Karl Fischer Composite For Ketones. 5 mL acetone is added using a glass syringe and weighed be difference. The water content is titrated with Composite 5 For Ketones. If more than one sample should be investigated, the working medium has to be replaced after 10 mL acetone. The same procedure can be used for Methyl-iso-butylketone.

## 8.9 Titration in a Heated Titration Vessel: Example D-(+)Glucose (Dextrose)

Thirty milliliter dry methanol are placed in the titration vessel, heated to 50°C and titrated to dryness with Karl Fischer reagent Composite 5. Approximately 500 mg of the sample is then precisely weighed in by difference. It should dissolve quickly in the heated working medium. The water content is titrated with the titrating reagent.

The titration can also be done at room temperature. In this case the working medium has to be replaced by a 1:1 mixture of dry formamide and methanol. Further examples to be used with this technique are Lactose Monohydrate, Fructose or other mono saccharides.

Samples as Corn starch, Glycerol dibehenate, or Sorbitan monostearate require definitive the titration at 50°C.

## 8.10 Titration in a Cooled Titration Vessel

This variation is also possible but, according to our knowledge, not required in the various pharmacopoeias. Just for the sake of completeness of Karl Fischer techniques it should be mentioned. The titration vessel of a volumetric Karl Fischer titrator can be cooled down to 0°C or, in some particular cases to −40°C. Then the titration should be performed with a two-component-reagent. This technique is used for the water determination in organic peroxides as peresters and percarbonates and requires dedicated end-point parameters.

## 8.11 Indirect Titration Using the Karl Fischer Oven

The use of the Karl Fischer oven is an elegant way to avoid the contact of the sample and the Karl Fischer reagents. It is recommended for insoluble and reactive substances. A stream of dried air is conducted over the sample, which is heated to the necessary temperature. The water evaporates and is conducted into the Karl Fischer cell. There, the titration takes place. Usually, the Karl Fischer oven is used in connection with a Karl Fischer coulometer. An additional benefit of this technique is the possibility for automation.

We do not describe a method for a special product since for each product a suitable temperature has to be found. If a lot of identical samples has to be analyzed the Karl Fischer oven can be a reliable alternative to manual Karl Fischer titration.

## 8.12 External Extraction in Dry Methanol

Another unusual but advantageous way of Karl Fischer titration is the external extraction in dry methanol. The sample is added in a vial which contains a defined amount of dry methanol, for example 10 mL. The vial is closed with a septum cap. After a suitable extraction time (overnight has proved itself), an aliquot is taken with a microliter glass syringe and added into the titration cell. There the water content is determined without problems of solubility or side-reactions. It is important to take the blank value into consideration, since the vial, the methanol and the syringe bring a minimal water amount into the analysis. It must be also considered that 1 mL dry methanol can extract maximal 1 mg water. In the last 20 years Herold (8) has made comprehensive investigations on this technique. He developed mutual standard procedures for a wide range of unrelated products as herbal drugs, gelatine capsules, organic solvents, powders and tablets.

## 8.13 The External Extraction Using Dried Karl-Fischer Anolyte: Example Lyophilized Material

Pharmaceutical preparations for injection are often supplied as lyophilized substrate in a vial. These substances are highly hygroscopic so the vial should not be opened for sample preparation. The water content is typically in the range about 100 μg per vial. For the determination of these small water contents the coulometric titration and the extraction of the water with dried Karl Fischer anolyte is recommended.

The coulometric titration cell is filled and dried as prescribed in Chapter 6. A syringe with a long needle is used the take approximately 5 mL anolyte out of the cell and re-inject it back at once. This procedure is repeated several times until the syringe is dry. Then 5 mL anolyte was taken again and this amount is injected into the vial with the lyophilized material. The substance is dissolved by shaking. Then the solution is drawn into the syringe again and injected into the anodic compartment of the coulometric cell. The water content is determined as usual. A blank of the whole

method can be investigated when the whole procedure is carried out with an empty vial.

In rare cases lyophilisates have a water content of 1% and more. For these materials the vials can be opened and the extraction with anolyte is not necessary.

## 9 TITRATION CURVES

In this chapter, we want to show several typical titration curves of Karl Fischer titrations. A variety of Karl Fischer titrators offer the possibility to print out the titration curve of the actual titration. It can be used to check the stability of the end point, the occurrence of a side-reaction or an extraction problem.

The curve above shows the reagent consumption versus titration time (Figure 2). A quick Karl Fischer titration of approximately 3 minutes can be seen. In the beginning of the titration the reagent is added fast, nearer to the end point the reagent dosage becomes slower until it stops.

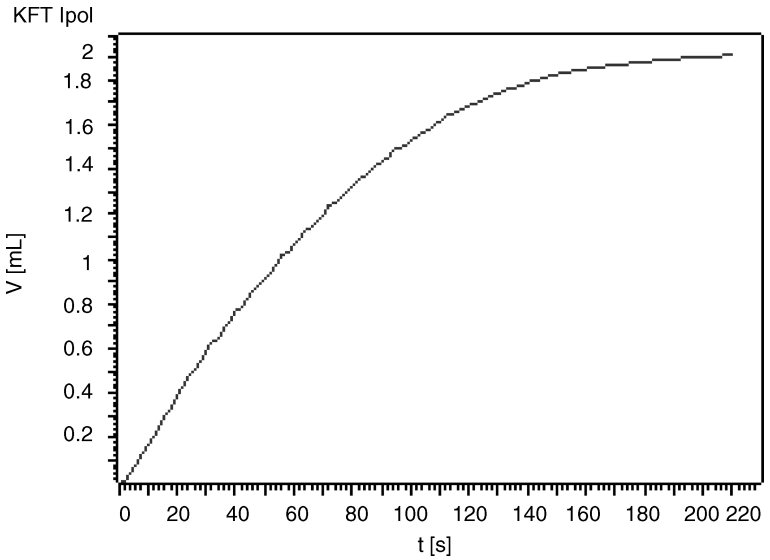


***FIGURE 2*** *Karl Fischer titration without problems: glycerol in methanol by volumetric titration.*

In Fig. 3a the typical side-reaction of ketones in methanol can be seen. The acetone forms an acetal with methanol and water is released during this slow reaction. The water is titrated and no end point can be found. After 15 minutes the titration is stopped manually. In Fig. 3b the titration takes place in Medium K, the methanol-free working medium for Karl Fischer titration. The side reaction is suppressed, not water is formed and a stable end point is achieved after 2 minutes.

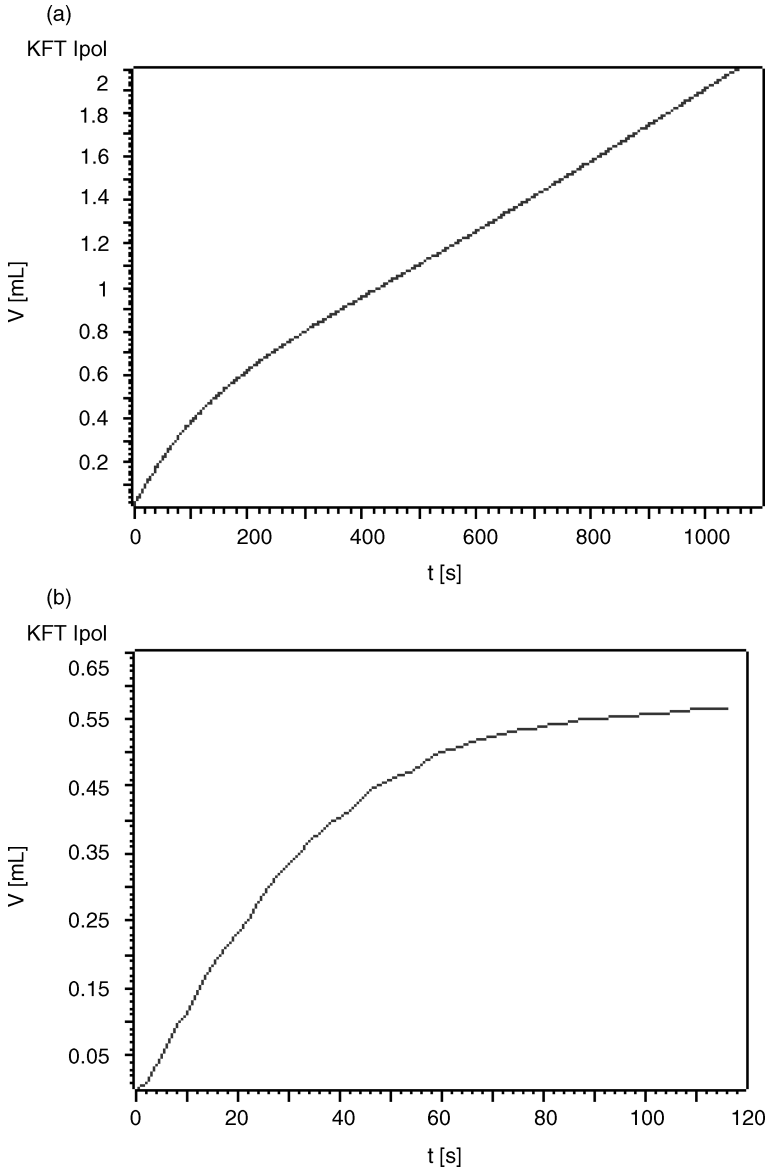


***FIGURE 3*** *Aceton in (a) methanol and (b) Medium K.*

In Fig. 4 the advantage of a solubilizer is shown. Curve 2 shows the titration behavior in pure methanol as working medium. The salt is dissolved incompletely and therefore the water is released slowly. The titration does not find an end point and is manually stopped after 15 minutes. Curve 1 shows the titration in a mixture of formamide and methanol. This solution dissolves the sample quickly and a stable end point is found after 3 minutes.

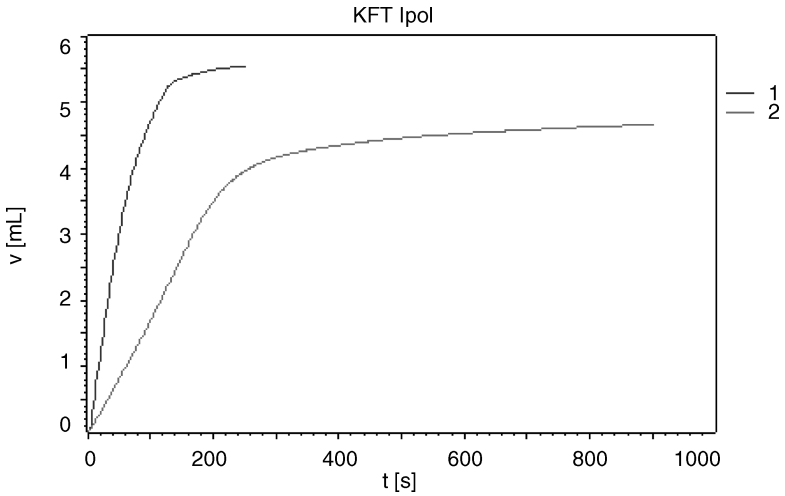


***FIGURE 4*** *Potassium citrate in methanol (2) and in methanol/formamide (1).*

## 10 QUALITY ASSURANCE

Karl Fischer titrators are quality relevant instruments in respect to ISO 9000 and GLP and therefore have to fulfill whose requirements. In practice three main issues have to be considered for these guidelines:

- Working instructions for water determinations and for instrument control must be defined in writing.
- Results must be documented to assure the retrieval of the analytical data at a later time.
- The analytical instrumentation must be inspected periodically to assure delivery of exact and precise measurement.

Validating analytical instruments is necessary and ensures proper functioning of the titrators. The documentation of such tests suggests that the results obtained from the inspected unit are correct. We recommend regular check of the whole system instrument AND reagents with defined commercially available water standards. These standards are measured against NIST SRM 2890, easy to apply and include a certificate of analysis showing the individual water content and its uncertainty.

## REFERENCES

1. Scholz E. Karl-Fischer-titration. *Berlin, Heidelberg, New York, Tokyo*: Springer-Verlag; 1984.
2. Verhoef JC, Barendrecht E. *Mechanism and reaction rate of the Karl-Fischer titration reaction*. J Electroanal Chem 1976; 71: 305–315.

3. Scholz E. *HYDRANAL®-Manual*. D-30926 Seelze, Germany: Sigma-Aldrich Laborchemikalien GmbH; 2006.
4. Hoffmann H. *Karl Fischer reagent containing ethanol*. European Patent 0, 933, 634. 1997.
5. United States Pharmacopeia; USP 31, NF 26, Volume 1. *921 Water Determination*, 2008; p 370–371.
6. European Pharmacopoeia, 5.7. *2.5.12.Water: Semi-micro Determination*, 2008; p 4785.
7. Japanese Pharmacopoeia; XV, 2.48. *Water Determination (Karl Fischer Method)*, 2007; p 47–50.
8. Herold G. *Presentation at 25th Karl–Fischer-Seminar*, Seelze, Germany, 25th + 26th November 2008.

# *EDTA Titrations*

**Dr. Zheng Chang**
*Department of Biological and Physical Sciences, South Carolina State University, Orangeburg, SC*
*Department of Physics, Virginia Polytechnic Institute and State University, Blacksburg, VA*

## 1 INTRODUCTION

The EDTA titration is based on the stoichiometric formation of water-soluble metal-EDTA complex to determine the concentration of the metal ion in aqueous solution. "EDTA" is the abbreviation of "ethylenediaminetetraacetic acid" or its anion, "ethylenediaminetetraacetate." Owing to its chemical structure, EDTA forms stable, water-soluble, 1:1 complexes with most polyvalent metal ions. This distinctive property sets the basis of the EDTA titration, and is also the reason that EDTA titrations are categorized as complexation titrations. In fact, a vast majority of complexation titrations employ EDTA as the complexing agent, and are thus EDTA titrations.

Fundamental studies on EDTA and related compounds can be traced back to 1930s. Schwarzenbach and his coworkers have studied the titrations with EDTA since the mid-1940s. In 1948, they reported that the endpoint of an EDTA titration might be detected by metal-sensitive indicators such as Murexide and Eriochrome Black T. Since then, the application of EDTA in titrations has been widely studied. A number of authors have presented the topic in full-length and well-organized styles [1–4].

Nowadays, with the development of analytical technology, various techniques are available for the analysis of metal ions. In addition to the conventional methods, advanced instrumental techniques such as atomic absorption spectroscopy (AAS),

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing,*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

inductively coupled plasma mass spectrometry (ICP-MS), and inductively coupled plasma atomic emission spectroscopy (ICP-AES) have proved great advantages for metal analysis. However, the EDTA titration is still widely used for quantitative analysis because of its unique advantages such as ease of operation, low cost of equipment, and minimal maintanence requirement. This is because the EDTA titration is simple to operate, can produce accurate results for samples with different complexities, and does not require costly instrumentation and specialized personnel. Moreover, with well-designed masking/de-masking schemes, EDTA titrations are capable for sequential analysis of multiple metal ions in a single sample without separation and extraction operations.

In this chapter, the EDTA titration is discussed from basic principles, titration curves, endpoint detection, titration types, and practical considerations. Some practical examples on the determination of various ions are also given at the end of the chapter.

## 2 IONIZATION OF EDTA

The structural formula of ethylenediaminetetraacetic acid (molecular weight 292.25) can be written as:

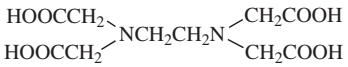


It can be seen that EDTA is a quadrivalent acid. The four hydrogen ions on the carboxyl groups can be ionized successively with the increase of pH. Schwarzenbach et al. have measured the ionization constants at the ionic strength of 0.1 M KCl and at 20°C. Based on their data [5], the ionization of the four hydrogen ions on EDTA can be expressed as:

$$H_4Y \rightleftharpoons H^+ + H_3Y^- \qquad K_1 = \frac{[H^+][H_3Y^-]}{[H_4Y]} = 10^{-2.0} \tag{1}$$

$$H_3Y^- \rightleftharpoons H^+ + H_2Y^{2-} \qquad K_2 = \frac{[H^+][H_2Y^{2-}]}{[H_3Y^-]} = 10^{-2.67} \tag{2}$$

$$H_2Y^{2-} \rightleftharpoons H^+ + HY^{3-} \qquad K_3 = \frac{[H^+][HY^{3-}]}{[H_2Y^-]} = 10^{-6.16} \tag{3}$$

$$HY^{3-} \rightleftharpoons H^+ + Y^{4-} \qquad K_4 = \frac{[H^+][Y^{4-}]}{[HY^{3-}]} = 10^{-10.26} \tag{4}$$

where $H_4Y$, $H_3Y^-$, $H_2Y^{2-}$, $HY^{3-}$, and $Y^{4-}$ refer to the molecular and ionized species of EDTA at different ionization levels, respectively. $K_1$, $K_2$, $K_3$, and $K_4$ are the corresponding ionization constants. It can be seen from the constants that EDTA is a mid-strong acid and the first two hydrogen ions are much easier to ionize than the rest of the hydrogen ions.

The mole fraction of each of the EDTA species is a function of pH if the ionic strength and the temperature are constant. From Eqs. 1–4, the mole fractions of the EDTA species can be obtained:

$$a_0 = \frac{[H_4Y]}{C_{EDTA}} = \frac{[H^+]^4}{[H^+]^4 + K_1[H^+]^3 + K_1K_2[H^+]^2 + K_1K_2K_3[H^+] + K_1K_2K_3K_4} \quad (5)$$

$$a_1 = \frac{[H_3Y^-]}{C_{EDTA}} = \frac{K_1[H^+]^3}{[H^+]^4 + K_1[H^+]^3 + K_1K_2[H^+]^2 + K_1K_2K_3[H^+] + K_1K_2K_3K_4} \quad (6)$$

$$a_2 = \frac{[H_2Y^{2-}]}{C_{EDTA}} = \frac{K_1K_2[H^+]^2}{[H^+]^4 + K_1[H^+]^3 + K_1K_2[H^+]^2 + K_1K_2K_3[H^+] + K_1K_2K_3K_4} \quad (7)$$

$$a_3 = \frac{[HY^{3-}]}{C_{EDTA}} = \frac{K_1K_2K_3[H^+]}{[H^+]^4 + K_1[H^+]^3 + K_1K_2[H^+]^2 + K_1K_2K_3[H^+] + K_1K_2K_3K_4} \quad (8)$$

$$a_4 = \frac{[Y^{4-}]}{C_{EDTA}} = \frac{K_1K_2K_3K_4}{[H^+]^4 + K_1[H^+]^3 + K_1K_2[H^+]^2 + K_1K_2K_3[H^+] + K_1K_2K_3K_4} \quad (9)$$

$$C_{EDTA} = [H_4Y] + [H_3Y^-] + [H_2Y^{2-}] + [HY^{3-}] + [Y^{4-}] \quad (10)$$

where $\alpha_0$, $\alpha_1$, $\alpha_2$, $\alpha_3$, and $\alpha_4$ are the mole fractions for $H_4Y$, $H_3Y^-$, $H_2Y^{2-}$, $HY^{3-}$, and $Y^{4-}$, respectively. $C_{EDTA}$ refers to the sum of the concentrations of the molecular EDTA and its four ionic species. Figure 1 shows the mole fractions of these EDTA species as a function of pH. It can be seen that $H_4Y$ and $H_3Y^-$ are the major species when pH < 3. $H_2Y^{2-}$ and $HY^{3-}$ mainly exist as pH ranges between 3 and 10, in which

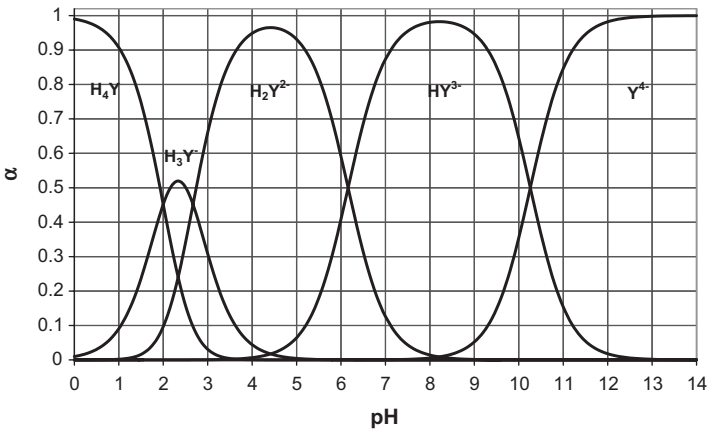


***FIGURE 1*** *The ionization degree of EDTA as a function of pH.*

most EDTA titrations are carried out. $Y^{4-}$ exists only in strong alkaline medium (pH > 10). Only a few of EDTA titrations are carried out in strong alkaline medium. This is because most of the metal ions of interest hydrolyze when pH > 9. Therefore, $H_2Y^{2-}$ and $HY^{3-}$ are the ionic species that exist and react with the metal ions in most EDTA titrations.

## 3 BASIC EDTA AGENTS FOR TITRATION

*MolecularEDTA ($H_4Y$):* The molecular EDTA is a free acid, which is a white, odorless, anhydrous powder. It is nearly insoluble in water (0.02 g per 100 g water at 22°C). The acid can be freed of absorbed water at 130–145°C for several hours.

*Sodium Dihydrogen Salt ($Na_2H_2Y_2{\cdot}2H_2O$):* The dihydrate of sodium dihydrogen salt is fairly soluble in water (11.1 g per 100 g at 21°C). This salt can slowly absorb up to 0.3% of moisture in the air. The moisture can be removed by drying at 80°C for at least 24 hours, so that the composition is brought to exactly that of the dehydrate formula.

## 4 FORMATION OF METAL-EDTA COMPLEXES

The complexation between EDTA and metal ions is characterized by three points: (*i*) EDTA forms stable complexes with most metal ions; (*ii*) EDTA complexes with metal ions at a ratio of 1:1 regardless of the charge of the metal ion; and (*iii*) all the hydrogen ions on the EDTA carboxyl groups are released once the metal-EDTA complex is formed.

The stability constant ($K_{MY}$) of the metal-EDTA complex is defined as:

$$M^{n+} + Y^{4-} \rightleftharpoons MY^{n-4} \tag{11}$$

$$K_{MY} = \frac{[MY^{n-4}]}{[M^{n+}][Y^{-4}]} \tag{12}$$

where $M^{n+}$ refers to the "free" metal ion with positive charge n. $MY^{n-4}$ refers to the metal-EDTA complex. Table 1 lists the stability constants for common metal ions at certain ionic strengths and temperatures. It can be seen from the table that most of the metals including alkaline earths, transition metals, rare earths, and actinides have a stability constant larger than $10^8$, which is large enough for the metal ion to be determined by the EDTA titration. Only a few metals, such as the alkali metals, Ba, Ra, Ag, and Tl(I), have a stability constant less than $10^8$. These metal ions form weak metal-EDTA complexes and are not suitable to be directly determined by EDTA titration, but can be measured by indirect EDTA titration methods.

Because $H_2Y^{2-}$ and $HY^{3-}$ are the major EDTA species in most titrations, the formation of the metal-EDTA complex in the EDTA titration may be expressed as:

$$M^{n+} + H_2Y^{2-} \rightleftharpoons MY^{n-4} + 2H^+ \quad \text{(pH 3–6)} \tag{13}$$

$$M^{n+} + HY^{3-} \rightleftharpoons MY^{n-4} + H^+ \quad \text{(pH 6–10)} \tag{14}$$

**TABLE 1 Stability Constants of Metal-EDTA Complexes* [6]**

| Metal Ion | Log $K_{MY}$ | Metal Ion | Log $K_{MY}$ | Metal Ion | Log $K_{MY}$ |
|---|---|---|---|---|---|
| Co(III) | 41.4 | Er(III) | 18.89 | $VO_2$(I) | 15.5 |
| Tl(III) | 35.3 | Cu(II) | 18.78 | La(III) | 15.36 |
| Hf(IV) | 29.5 | VO(II) | 18.7 | Fe(II) | 14.30 |
| Zr(IV) | 29.3 | Ho(III) | 18.56 | Mn(II) | 13.89 |
| Bi(III) | 27.8[a] | Ni(II) | 18.4 | Cr(II) | 13.6[a] |
| V(III) | 25.9[a] | Sn(II) | 18.3[b] | V(II) | 12.7[a] |
| U(IV) | 25.7 | Dy(III) | 18.3 | Ca(II) | 10.65 |
| Pd(II) | 25.6[a] | Y(III) | 18.08 | Be(II) | 9.7 |
| Fe(III) | 25.1 | Pb(II) | 18.0 | Mg(II) | 8.79 |
| In(III) | 24.9 | Tb(III) | 17.87 | Sr(II) | 8.72 |
| Cr(III) | 23.4[a] | Sm(III) | 17.06 | | |
| Th(IV) | 23.2 | Pm(III) | 16.9 | Ba(II) | 7.88 |
| Sc(III) | 23.1[a] | Nd(III) | 16.51 | Ra(II) | 7.4 |
| Ga(III) | 21.7 | Zn(II) | 16.5 | Ag(I) | 7.20 |
| Hg(II) | 21.5 | Cd(II) | 16.5 | Tl(I) | 6.4 |
| Ti(III) | 21.3 | Co(II) | 16.45 | Li(I) | 2.95 |
| Lu(III) | 19.74 | Al(III) | 16.4 | Na(I) | 1.86 |
| Yb(III) | 19.49 | Pr(III) | 16.30 | K(I) | 0.8 |
| Tm(III) | 19.32 | Ce(III) | 15.93 | | |

* Most of the stability constants are obtained at 25°C and ionic strength 0.1 M except:
[a] 20°C and ionic strength 0.1 M,
[b] 20°C and ionic strength 1 M.

Notice that all carboxyl hydrogen ions on the EDTA are released once the complex is formed.

The fact that EDTA forms $My^{n-4}$ complex with metal ions can be explained from its chemical structure. EDTA is an aminopolycarboxylic acid with six ligand groups (i.e., four single-bonded oxygen atoms in the ionized carboxyl groups and two nitrogen atoms in the amino groups). As most metal ions show a coordination number of 6 or 8 in the formation of complexes, all the six ligand atoms on EDTA will take part in the complexation. Note that EDTA has a structure with any of the two neighbored ligand atoms being separated by two carbon atoms. This leads to a five-numbered ring structure when the two neighbored ligand atoms chelate with the metal ion. Thus, a metal-EDTA complex can form as many as 5 five-numbered rings if all the four carboxyl hydrogen ions are ionized. According to coordination chemistry, a five-numbered ring structure is the most stable chelate structure. Therefore, when complexing with a metal ion, EDTA will release all the hydrogen ions on the carboxyl groups and form 1:1 metal-EDTA complex with five chelate rings, so that a structure with high stability is achieved (Fig. 2).

## 5 CONDITIONAL STABILITY CONSTANT

In the EDTA titration, the stability constant of the metal-EDTA complex is mainly affected by two important factors: (*i*) the medium pH and (*ii*) the concentration of the auxiliary complexing agent.

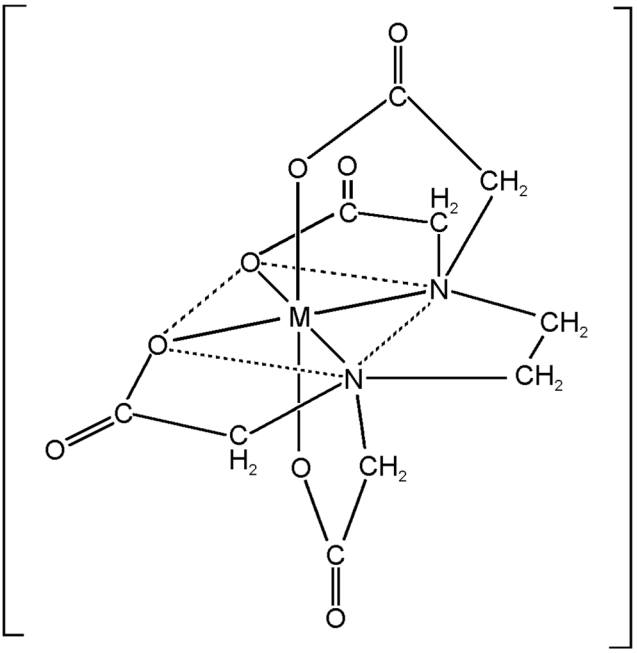


***FIGURE 2*** *The structure of a metal-EDTA complex.*

Metal ions with different stability constants can be titrated by adjusting these two factors. The conditional stability constant of the metal-EDTA complex is the stability constant, considering the effects of these two factors, which are quantitatively discussed in the following.

***pH Effect:*** The stability of the metal-EDTA complex is very sensitive to the medium pH. At low pH, the ionic EDTA species tend to combine with hydrogen ion. The complexing equilibrium in Eq. 13 or 14 shifts toward the left. The metal-EDTA complex becomes less stable. On the other hand, at high pH, more hydrogen ions ionize from the EDTA. The complexing equilibrium in Eq. 13 or 14 will shift toward the right. The metal-EDTA complex becomes more stable. However, as the metal ion hydrolyzes to form various hydroxyl species or precipitates in alkaline condition, complicated species other than 1:1 metal-EDTA complex will be produced when pH is too high, which is generally avoided in the EDTA titration.

Due to the strong dependence on pH, the EDTA titration is normally carried out in a buffered solution where the medium pH is controlled at a constant predetermined value for the titration. Under such a condition, the stability of a metal-EDTA complex can be readily evaluated.

Combining Eqs. 9 and 12, a conditional stability constant ($K_{\text{MY-pH}}$) can be obtained to evaluate the pH effect on the metal-EDTA complexation:

$$K_{\text{MY-pH}} = \frac{[\text{MY}^{n-4}]}{[\text{M}^{n+}]C_{\text{EDTA}}} = \frac{[\text{MY}^{n-4}]}{[\text{M}^{n+}]([\text{Y}^{-4}]/\alpha_4)} = K_{\text{MY}}\alpha_4 \tag{15}$$

where $C_{EDTA}$ refers to the sum of the concentrations of all the five EDTA species ($H_4Y$, $H_3Y^-$, $H_2Y^{2-}$, $HY^{3-}$, and $Y^{4-}$). From Eq. 15, it can be seen that $K_{MY\text{-}pH}$ increases with the increase of pH and decreases with the decrease of pH. Note that Eq. 15 only considers the effect of pH on the ionization of EDTA; the hydrolysis of the metal ion is not included. This is because the hydrolysis of metal ion will significantly change the mechanism of the metal-EDTA reaction and complicate the basic complex reaction on which the EDTA titration is carried out. Practically, the medium pH is controlled in the range where the hydrolysis of metal ion does not affect the formation of the metal-EDTA complex. Under the circumstances that the EDTA titration has to be conducted in the pH range where the metal-ion hydrolysis is significant, an auxiliary complexing agent is usually added. The effect of the auxiliary complexing agent on the EDTA titration is discussed in the following section.

***The Effect of the Auxiliary Complexing Agent:*** In many EDTA titrations, an auxiliary complexing agent is added to prevent the hydrolysis of metal ion. Normally, a monodentate ligand is used as the auxiliary complexing agent, which complexes with the metal ion and changes the stability of the metal-EDTA complex. The effect of the auxiliary agent can be readily evaluated because it is often added in excess amount and its concentration is considered constant throughout the EDTA titration. Suppose the auxiliary complexing agent, L, combines with the metal ion by forming a series of complex species with the corresponding stability constants ($K_{ML1}$, $K_{ML2}$, $K_{ML3}$, $K_{ML4}$, ...):

$$M^{n+} + L^{m-} \rightleftharpoons ML^{n-m} \qquad K_{ML1} = \frac{[ML^{n-m}]}{[M^{n+}][L^{m-}]} \tag{16}$$

$$ML^{n-m} + L^{m-} \rightleftharpoons ML_2^{n-2m} \qquad K_{ML2} = \frac{[ML_2^{n-2m}]}{[ML^{n-m}][L^{m-}]} \tag{17}$$

$$ML_2^{n-2m} + L^{m-} \rightleftharpoons ML_3^{n-3m} \qquad K_{ML3} = \frac{[ML_3^{n-3m}]}{[ML_2^{n-2m}][L^{m-}]} \tag{18}$$

$$\cdots \quad \cdots \quad \cdots \quad \cdots \quad \cdots \quad \cdots$$

$$\beta_{ML} = \frac{[M^{n+}]}{C_M} = \frac{1}{1 + K_{ML1}[L] + K_{ML1}K_{ML2}[L]^2 + K_{ML1}K_{ML2}K_{ML3}[L]^3 + \cdots} \tag{19}$$

where $m$ is the charge of the auxiliary complexing agent. $C_M$ is the sum of the concentrations of the metal ion and its complex species with L. $\beta_{ML}$ is the mole fraction of the metal ion with respect to $C_M$. It is noticed that the protonation of the auxiliary complexing agent is not considered here. This approximation is reasonable because the auxiliary complexing agent is normally used at high pH to prevent the metal hydrolysis. The protonation of the auxiliary complexing agent is negligible compared with its metal-complexation reaction.

Therefore, the conditional stability constant of the metal-EDTA complex considering the effect of the complexation with the auxiliary agent can be written as:

$$K_{\text{MY-L}} = \frac{[\text{MY}^{n-4}]}{C_{\text{M}}[\text{Y}^{-4}]} = \frac{[\text{MY}^{n-4}]}{([\text{M}^{n+}]/\beta_{\text{ML}})[\text{Y}^{-4}]} = K_{\text{MY}}\beta_{\text{ML}} \tag{20}$$

Combining Eqs. 15 and 20 gives a conditional stability constant ($K'_{\text{MY}}$) considering the effects of both EDTA ionization and the complexation with the auxiliary agent:

$$K'_{\text{MY}} = \frac{[\text{MY}^{n-4}]}{C_{\text{M}}C_{\text{EDTA}}} = \frac{[\text{MY}^{n-4}]}{([\text{M}^{n+}]/\beta_{\text{ML}})([\text{Y}^{-4}]/\alpha_4)} = K_{\text{MY}}\alpha_4\beta_{\text{ML}} \tag{21}$$

$$\log K'_{\text{MY}} = \log K_{\text{MY}} + \log a_4 + \log \beta_{\text{ML}} \tag{22}$$

## 6 TITRATION CURVES

In analogy to the acid-base titration, the EDTA titration curve can be drawn by plotting pM ($= -\log\,[\text{M}^{n+}]$) versus the volume of the EDTA titrant or the percent of the stoichiometric amount of EDTA added. The titration curve is not only for understanding the titration process, but also for evaluating the key factors in developing new EDTA titrations.

Suppose the sample solution to be titrated has a volume of $V_0$ and a metal concentration of $C^0_{\text{M}}$. The EDTA titrant has a concentration of $C^0_{\text{Y}}$. If the volume of the EDTA titrant added at any point of the titration is $V_{\text{t}}$, then the percent of the stoichiometric amount of EDTA added can be expressed as $\phi = C^0_{\text{Y}}V_{\text{t}}/C^0_{\text{M}}V_0$. The titration curve can thus be obtained by plotting pM versus $\phi$.

The pH buffer and auxiliary complexing agent in a titration are normally in excess amount. So Eq. 21 can be directly used for the calculation of pM at any given point of titration because both $\alpha_4$ and $\beta_{\text{ML}}$ basically remain constant throughout the titration process. Combining Eqs. 9,19, and 21, a titration curve can be calculated in three steps:

1. *Before the equivalence point* ($\phi < 1$):

$$\text{pM} = \log\left[\frac{1+\phi\cdot\gamma}{C^0_{\text{M}}\beta_{\text{ML}}(1-\phi)}\right] \tag{23}$$

where $\gamma = C^0_{\text{Y}}/C^0_{\text{M}}$.

2. *At the equivalence point* ($\phi = 1$):

$$\mathrm{pM} = \frac{1}{2}\log\left[\frac{\alpha_4 K_{\mathrm{MY}}(1+\phi\cdot\gamma)}{C_{\mathrm{M}}^0\beta_{\mathrm{ML}}}\right] \tag{24}$$

3. *After the equivalence point* ($\phi > 1$):

$$\mathrm{pM} = \log[\alpha_4 K_{\mathrm{MY}}(\phi - 1)] \tag{25}$$

Figure 3 shows a group of titration curves with different stability constants ($K_{\mathrm{MY}}$). It can be seen that the titration curves have an "S" shape, with the inflexion point at the equivalence point of the curve. Among all the factors that affect the shape of the titration curve, the stability constant of the metal-EDTA complex is the most important. As shown in Fig. 3, the larger the $K_{\mathrm{MY}}$ value, the greater the pM change at the equivalence point. It is also noticed that $K_{\mathrm{MY}}$ does not affect the titration curve before the equivalence point. But as $K_{\mathrm{MY}}$ increases, the titration curve after the equivalence point is shifted up, and the pM change at the equivalence point is increased consequently. This can be understood in Eqs. 23–25, where $K_{\mathrm{MY}}$ appears only in the last two equations.

Since any titration is dependent on the quick pM change around the equivalence point, a titration curve with a sharp and large pM jump is crucial for the successful detection of the endpoint of the titration. Generally, it needs at least 2 units of pM change for a detection method to sense the endpoint. From Fig. 3, it can be seen that the titration curves with $K_{\mathrm{MY}} > 10^8$ have more than 2 units of pM change at the

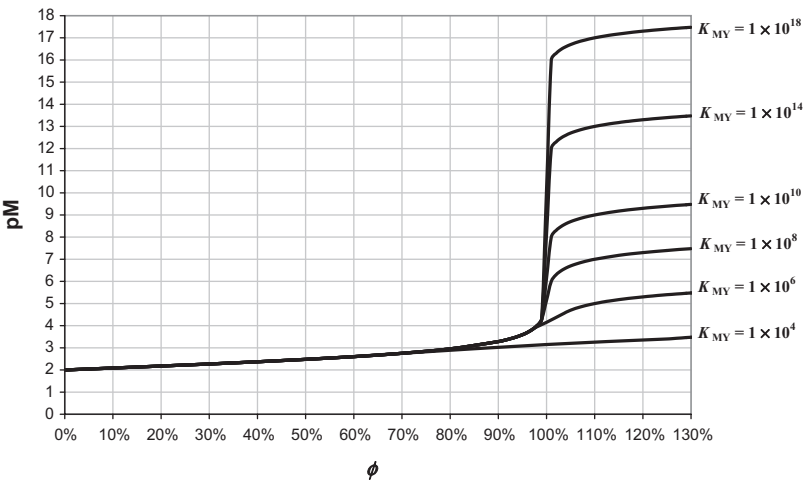


***FIGURE 3*** *Titration curves with different stability constants ($K_{\mathrm{MY}}$).* Suppose the titrations are carried out at the condition: $C_{\mathrm{M}}^0 = 0.01\,\mathrm{M}$, $C_{\mathrm{Y}}^0/C_{\mathrm{M}}^0 = 1$, $\alpha_4$=1, and $[\mathrm{L}^{m-}] = 0\,\mathrm{M}$.

equivalence point. While the curves with $K_{MY} < 10^8$ have less than 2 units of pM change. This is the reason why it is difficult or impossible to conduct an EDTA titration with $K_{MY} < 10^8$. It can be seen from Table 1 that most of the metals have a stability constant greater than $10^8$ and thus can be determined by the EDTA titration.

***pH Effect on the Titration Curve:*** The medium pH can affect the conditional stability constant ($K'_{MY}$) of the metal-EDTA complex by changing the ionization of EDTA. From Eqs. 23–25, it can be seen that the titration curve before the equivalence point will not be affected by the medium pH. But the curve after the equivalence point will shift up with the increase of pH and shift down as the pH decreases. For this reason, a higher pH is generally favorable, provided that the hydrolysis of the metal ion does not affect the complexation reaction.

On the other hand, it is interesting to find out the minimum pH that causes a given $K'_{MY}$ decrease to $10^8$, which is just big enough to produce a 2-unit pM jump at the equivalence point of the titration curve. The minimum pH can be calculated with Eq. 15 by supposing $K'_{MY} = 10^8$. These values are important for the design of the EDTA titration under various conditions. Reilley and Schmid [7] have plotted the stability constant of the metal-EDTA complex versus the minimum pH for a number of metals. Similar diagram (Fig. 4) was drawn in this chapter with the data in Table 1. It can be seen from the figure that many metal ions can be titrated in the range between pH 3 and 10, while tetravalent and trivalent metal ions such as Zr (IV), Th(IV), Bi(III), Fe(III), In(III), and Sc(III) can be titrated in acidic media (pH < 3).

***Effect of the Auxiliary Complexing Agent on the Titration Curve:*** Under certain circumstances, the metal ion may hydrolyze at the pH where the EDTA titration is designed to conduct. The hydrolysis of metal ion produces hydroxides and precipitates that can significantly influence the EDTA titration by slowing down the EDTA complexation reaction or changing the complexation mechanism. Therefore, significant metal hydrolysis should always be avoided in an EDTA titration.

The addition of an auxiliary complexing agent can help prevent the metal hydrolysis. However, the stability of the metal-auxiliary agent complex should be much less than that of the metal-EDTA complex, so that the conditional stability constant ($K'_{MY}$) will not be less than $10^8$. In this way, as the titration proceeds, the auxiliary agent will be gradually replaced when the metal-EDTA complex is formed.

Unlike the pH effect, the addition of an auxiliary complexing agent will not change the titration curve after the equivalence point but move up the curve before the equivalence point. This can be understood from the equations of the titration curve. $\beta_L$ is contained in Eqs. 23 and 24, but not in Eq. 25.

Figure 5 shows the titration curves for Zn(II) at pH 10, with different concentrations of an auxiliary complexing agent, $NH_3$. The stepwise stability constants of Zn-$NH_3$ complexes [8] are $\log K_1 = 2.32$, $\log K_2 = 2.29$, $\log K_3 = 2.36$, and $\log K_4 = 2.39$, respectively. Together with the stability constant of Zn-EDTA complex ($\log K_{Zn\text{-}EDTA} = 16.5$ from Table 1) and the ionization constants of EDTA, the titration curves in Fig. 5 can be obtained. It can be seen that the curve before the equivalence point is shifted up with the increase of $[NH_3]$. The higher the $[NH_3]$, the smaller the pM jump

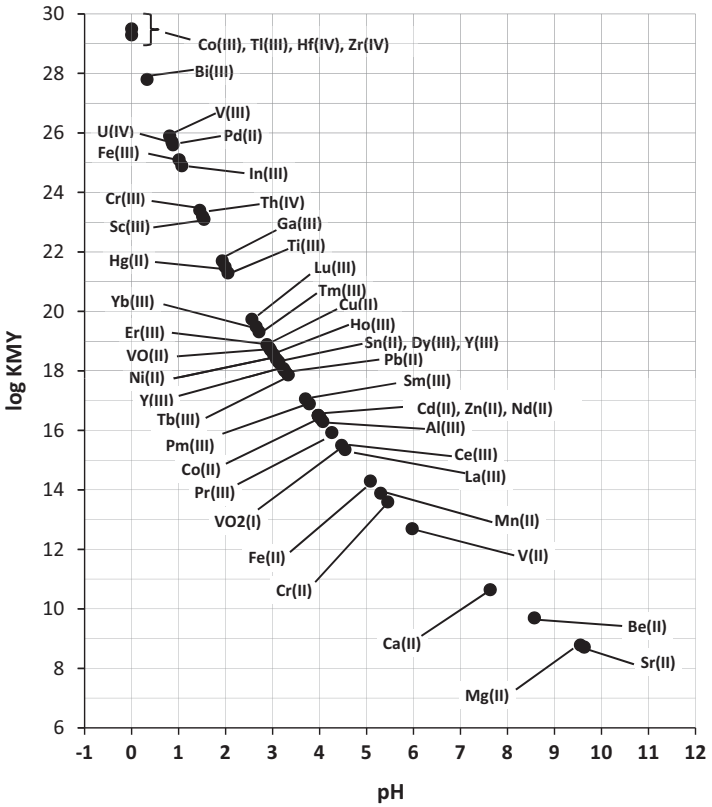


***FIGURE 4*** *The pH effect on the stability constants of the metal-EDTA complexes.*

at the equivalence point. Therefore, as long as that the metal hydrolysis is masked; the addition of the auxiliary complexing agent should be limited to the lowest concentration.

***The Concentration Effect:*** The concentration of the metal ion ($C_M^0$) may also affect the titration curve. It can be seen from Eqs. 23 and 24 that the titration curve before the equivalence point shifts up with the decrease of $C_M^0$. As a rule of thumb, every 10-fold decrease in $C_M^0$ will lead to the curve before the equivalence point shift up by 1 unit. Practically, $C_M^0$ is often adjusted in the range between 0.01 and 0.001 M for the EDTA titration. So the titration curve before the equivalence point is positioned between 2 and 4 on the pM-axis. If much lower $C_M^0$ has to be used, it is suggested that before the analysis the titration curve is plotted, the pM change at the equivalence point is checked, and the endpoint detection method is evaluated.

The concentration ratio of EDTA over the metal ion ($\gamma = C_Y^0/C_M^0$) affects the titration curve before the equivalence point. Figure 6 shows the titration curves for Ca with different $\gamma$ values. It can be seen that if the EDTA concentration is larger than that of the metal ion ($\gamma > 1$), the titration curve before the equivalence point is moved up,

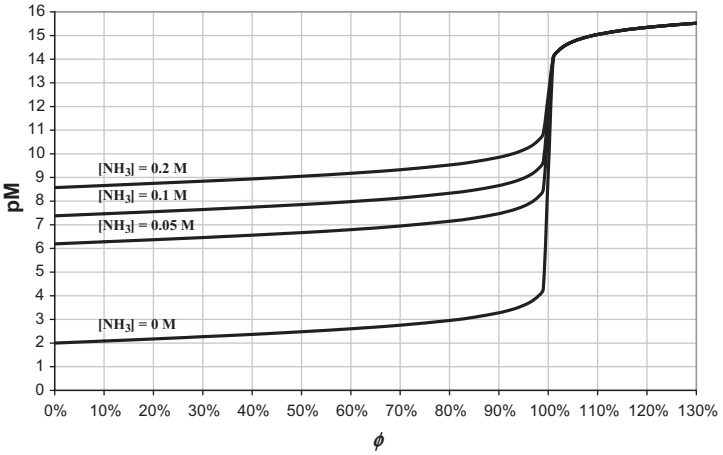


***FIGURE 5*** *Titration curves for Zn(II) in the medium with different free [$NH_3$]. Suppose the titrations are carried out at the condition: $K'_{MY} = 10^{16.5}$, $C^0_M = 0.01$ M, pH = 10, and $C^0_Y/C^0_M = 1$.*

and the pM jump at the equivalence point is decreased. Furthermore, a larger analytical error will be introduced at the endpoint when the EDTA concentration is significantly larger than that of the metal ion. On the other hand, if the EDTA concentration is smaller than that of the metal ion ($\gamma < 1$), the curve before the equivalence point can shift down as much as 1 unit of pM. Further dilution of the EDTA concentration will not make any significant change to the titration curve, but will cause difficulty in the titration because it is impractical to accurately add a large

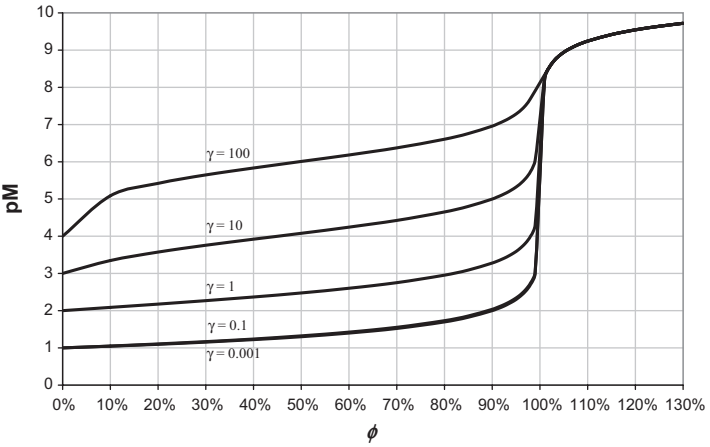


***FIGURE 6*** *Titration curves for $Ca^{2+}$ with different $\gamma$ ($= C^0_Y/C^0_M$). Suppose that $K'_{MY} = 5 \times 10^{10}$, $C^0_M = 0.01$ M, pH = 10, and $[L^{m-}] = 0$.*

volume of the titrant into the sample solution. Therefore, the proper concentration ratio ($\gamma$) for the titration should be in the range between 0.1 and 1.

## 7 ENDPOINT DETECTION METHODS

A successful titration depends on the accurate detection of the equivalence point, at which the amount of the titrant added is equivalent to the amount of the analyte presenting in the sample. The method used to detect the equivalence point is called "endpoint detection method." The endpoint is not necessarily to be exactly equal to the equivalence point. The precision related to the endpoint detection is affected not only by the endpoint method but also by the shape of the titration curve. Normally, an endpoint located within the pM jump at the equivalence point will produce a result with an error less than 1%, which is higher than the precision (5%) required for a titration analysis.

Many detection methods have been developed for the EDTA titration. These methods can be categorized into two groups: visual and instrumental detection methods.

A visual detection method usually employs a special agent (indicator) that shows the endpoint by a visual effect (change of color or production/disappearance of precipitation) that is discernable to the eye of the experimenter. Examples for visual indicators include metal-ion indicators, redox indicators, and precipitation indicators. Visual detection methods are normally easy to operate. Skillful experimenters can complete an EDTA titration with a visual detection within several minutes.

Instrumental detection methods include photometric, potentiometric, amperometric, conductometric, and thermometric methods. Instrumental methods are especially advantageous when combined into an automated titration system that is suitable for handling large numbers of samples with similar properties. Instrumental methods are also competitive in micro-titrations where very small amounts of sample are involved and it is difficult for human to discern a visual change. However, instrumental methods require more technical maintenance and are not as flexible as visual methods. It usually takes more time to complete a titration if an instrumental detection is used.

In the following two sections, we will discuss metal-ion indicators and potentiometric methods as examples of visual and instrumental methods.

***Metal-Ion Indicators:*** These visual indicators are organic dyestuffs characterized with functional groups that show different colors when complexing or not complexing with the metal ion. The functional groups of the metal-ion indicators are often composed of phenol and naphthol groups with many conjugated $\pi$-bonds. The delocalized electrons in these $\pi$-bonds are easy to be excited by visible lights so that different ionization states of these functional groups display various colors. A sharp color contrast between the dyestuff itself and the metal-dyestuff complex is essential for a dyestuff to be used as endpoint indicator in the EDTA titration.

When added as the metal-ion indicator, the dyestuff will complex with the metal ion and show a certain color. As the titration proceeds, the dyestuff is progressively replaced from the metal-dyestuff complex by the addition of EDTA. The liberated dyestuff displays a color different from that of the metal-dyestuff complex. The hue of the solution color changes slowly in the course of titration. At the endpoint, a color transition occurs when about half of the dyestuff is released from the metal-dyestuff complex. If it occurs within the pM jump of the titration curve, this color transition will be sharp enough for an experimenter to discern because the metal concentration changes most quickly in the pM change around the equivalence point. Consequently, the color transition provides a visual signal for the arrival of the endpoint.

Actually, the metal-ion indicators are often organic acids and have multiple ionic species with different colors. Taking a widely used metal-ion indicator, Eriochrome Black T, as an example, its ionic species at different pH can be shown as:

$$\underset{\text{red}}{H_2In^-} \xleftrightarrow{\text{pH } 5.3-7.3} \underset{\text{blue}}{HIn^{2-}} \xleftrightarrow{\text{pH } 10.5-12.5} \underset{\text{yellow/orange}}{In^{3-}} \tag{26}$$

It can be seen that Eriochrome Black T alone can show red, blue, and yellow/orange color in the ranges of pH $< 7$, pH 7–11, and pH $> 11$, respectively. As it has to give a sharp color change at the endpoint of titration, Eriochrome Black T can only work in certain pH range. The complexes of Eriochrome Black T with many metals (Ca, Mg, Cd, Zn, Ba, Mn, Hg, Pb, Cu, Al, Fe, Ti, Co, Ni, etc.) have a red color. This suggests that the indicator can only be used when pH is larger than 7. When pH $< 7$, the indicator itself is red, and no color transition will occur at the endpoint. Therefore, Eriochrome Black T is not suitable for the EDTA titration carried out in the media with pH $< 7$.

Another important condition for a dyestuff to be a metal-ion indicator is that the stability constant of the metal-dyestuff complex should match the titration so that the color transition (i.e., endpoint) occurs within the pM jump around the equivalence point of the titration. If the metal-dyestuff complex is not stable enough, the endpoint will occur before the pM jump. If the metal-dyestuff complex is too stable, a sluggish endpoint, a late endpoint, or no endpoint at all will be observed. Therefore, it is important to understand the relationship between the stability of the metal-dyestuff complex and the equivalence point of the titration.

Analogous to the EDTA complexation shown in Eq. 21, the conditional stability constant, $K'_{MIn}$, for the metal-dyestuff complex can be expressed as:

$$K'_{MIn} = \frac{[MIn]}{[M^{n+}]C_{In}} \tag{27}$$

where [MIn] refers to the concentration of the metal-indicator complex. The charge of the metal-dyestuff complex is omitted for simplicity. $C_{In}$ is the concentration of all the indicator species not combined with the metal ion. As to Eriochrome Black T, $C_{In}$ is equal to $[H_3In] + [H_2In^-] + [HIn^{2-}] + [In^{3-}]$ according to Eq. 26.

The relationship between $K'_{\mathrm{MIn}}$ and the color transition point can be evaluated by rearranging Eq. 27:

$$\log K'_{\mathrm{MIn}} = \mathrm{pM} + \log \frac{[\mathrm{MIn}]}{C_{\mathrm{In}}} \tag{28}$$

Suppose the color transition occurs at the point where half amount of the indicator is released from the metal-dyestuff complex; then Eq. 28 at the endpoint can be written as:

$$\log K'_{\mathrm{MIn}} = \mathrm{pM}, \quad \text{when}[\mathrm{MIn}] = C_{\mathrm{In}} \tag{29}$$

From Eq. 29, it is possible to tell whether a metal-ion indicator functions properly for a given EDTA titration. If log $K'_{\mathrm{MIn}}$ falls in the region of the pM jump at the equivalence point of the titration curve, the indicator is good for the titration. If log $K'_{\mathrm{MIn}}$ falls out of the region of the pM jump, the indicator is not suitable for the titration.

Flaschka [9] further studied the suitability of the metal-ion indicator to EDTA titration. According to his study, when common concentrations of the meal and the auxiliary complexing agent are used, a good metal-ion indicator should be the following: (*i*) The value of log $K'_{\mathrm{MIn}}$ must be 4 units less than that of log $K'_{\mathrm{MIn}}$ of the metal-EDTA complex; (*ii*) the value of log $K'_{\mathrm{MIn}}$ itself should be larger than 5; and (*iii*) the concentration ratio of the indicator to the metal ion should be less than $10^{-2}$.

It is noticed that similar to log $K'_{\mathrm{MIn}}$, log $K'_{\mathrm{MIn}}$ for the metal-dyestuff complex is also dependent on the medium pH because metal-ion indicators are usually organic acids. Therefore, the suitability of a metal indicator to a given EDTA titration is related to the medium pH of the titration. For example, the log $K'_{\mathrm{MIn}}$ of Eriochrome Black T complexes with Ca and Mg were found varying at different pH, as shown in Table 2. If a 0.01 M Ca sample solution is to be titrated under the condition: $[NH_4{}^+] = 0.1$ M and pH = 10, one can see from Fig. 7 that Eriochrome Black T is not a satisfactory indicator because log $K'_{\mathrm{CaIn}}$ (3.8) is small and falls out of the pM jump (4–8) of Ca titration curve. This leads to a vague color transition and early endpoint detection. A method called "replacement titration" is often used to overcome such a problem. As to Ca titration, several drops of Mg-EDTA are added in the sample solution before the titration. Because Ca-EDTA complex is more stable than Mg-EDTA complex, EDTA will complex with Ca before complexing with Mg. Therefore, it is Mg, not Ca, to show the endpoint with the dyestuff (Eriochrome Black T). Because log $K'_{\mathrm{MgIn}} = 5.4$ at pH 10, the color transition is within the pM jump of the Mg titration curve. A sharp color change from red to blue occurs to give the correct endpoint. Note that no titrant is consumed by Mg in this method because Mg is added as its EDTA complex.

**TABLE 2 The Conditional Stability Constants of Eriochrome Black T with Ca and Mg at Different pH**

| pH | 7 | 8 | 9 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|
| Log $K'_{\mathrm{MgIn}}$ | 2.4 | 3.4 | 4.4 | 5.4 | 6.3 | 6.8 |
| Log $K'_{\mathrm{CaIn}}$ | 0.8 | 1.9 | 2.8 | 3.8 | 4.7 | 5.3 |

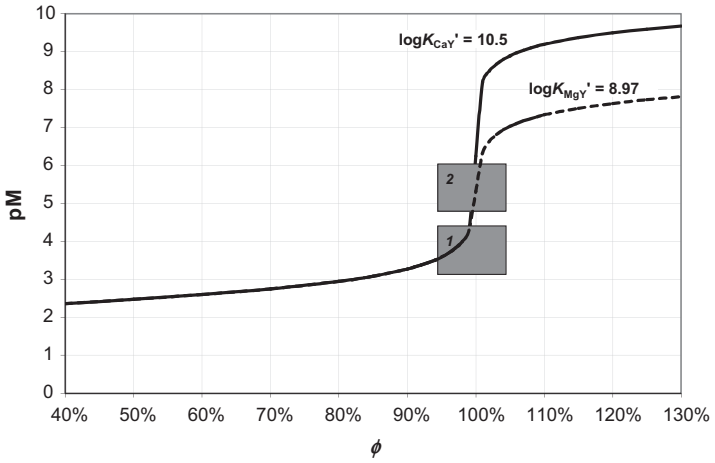


***FIGURE 7*** *The relationship between the metal-ion indicator and the endpoint. Sample solution: $[Ca^{2+}] = 0.01$ M; buffer $NH_3/NH_4^+$; pH = 10. Solid line: titration curve of Ca. Dotted line: titration curve of Mg. Box 1: Endpoint position without Mg-EDTA in the sample. Box 2: Endpoint position if several drops of Mg-EDTA are added.*

Many metal-ion indicators have been developed since G. Schwarzenbach and his coworkers reported that the endpoint of the EDTA titration can be marked by Murexide [10] and Eriochrome Black T [11] in the 1940s. Although the basic principle of indicating the endpoint of the EDTA titration is very similar, these dyestuffs possess different acid-base and metal complexing properties. Table 3 summarizes some common metal-ion indicators with their basic properties and applications for various EDTA titrations.

***Potentiometric Methods:*** The potentiometric method is one type of instrumental methods that is widely used in the EDTA titration. Like other instrumental methods, the potentiometric detection is superior when higher precision is required or micro-titration needs to be conducted. Potentiometric methods can be readily combined into an automated titration system, in which the endpoint can be determined by a graphical extrapolation technique that will yield better precision when the pM change around the equivalence point is not large enough for a visual method to determine.

Two types of electrodes are often used in the potentiometric detection: redox electrodes and ion-selective electrodes (ISEs). These two types of electrodes are discussed in the following.

*The Redox Electrode:* The electrical potential of a half-cell redox reaction is dependent on the concentration ratio of the reduction and oxidation states of the chemical species according to Nernst equation. At room temperature (25°C), Nernst equation is written as:

$$E = E_0 + \frac{0.0591}{z} \log \frac{[\text{Ox}]}{[\text{Red}]} \tag{30}$$

**TABLE 3 Properties and Applications of Some Common Metal-Ion Indicators**

| Indicator and Its Preparation | Chemical Name | Colors of Free Indicator Species at Different pH | Color of Metal-Indicator Complexes | Applications |
|---|---|---|---|---|
| Calmagite (0.05 g in 100 mL water. Stable for 1 year in polyethylene bottle away from sunlight) | 1-(1-Hydroxl-4-methyl-2-phenylazo)-2-naphthol-4-sulphonic acid | pH < 7, red ($H_2In^-$); | Wine red | A good substitute for Eriochrome Black T, with much better stability in aqueous solution |
| | | 8 < pH < 12, blue ($HIn^{2-}$); pH > 13, yellow orange ($In^{3-}$) | | |
| Eriochrome Black T (Solochrome Black) (0.4 g in 100 ml methanol. Stable for 1 month) | 1-(1-Hydroxy-2-naphthylazo)-6-nitro-2-naphthol-4-sulphonic acid | pH < 6, red ($H_2In^-$); | Wine red | *Direct titration* Mg, Ba, Cd, In(III), Mn, Pb, Zn and rare earths |
| | | 7 < pH < 11, blue ($HIn^{2-}$); | | *Back titration* Al, Bi(III), Ca, Co, Cr(II), Cu(II), Ni, Fe(II), Ti(IV) |
| | | pH > 12, yellow orange ($In^{3-}$) | | *Replacement titration* Au (III), Ba, Ca, Cu(II), Hg(II), Pb, Pd(II), Sr |
| Murexide | 5-[(Hexahydro-2,4,6-trioxo-5-pyrimidinyl)imino]-2,4,6 (1H,3H,5H)-pyrimidine-trione monoammonium salt | pH < 9, red-violet ($H_4In^-$); | Red with $Ca^{2+}$ | *Direct titration* Ca, Ni, Co, Cu(II) |
| | | 10 < pH < 11, violet ($H_3In^{2-}$); pH > 11, blue ($H_2In^{3-}$) | Yellow with $Co^{2+}$, $Ni^{2+}$ and $Cu^{2+}$ | *Back titration* Ca, Cr(III), Ga (III) *Replacement titration* Ag and halides |
| PAN | 1-(2-Pyridylazo)-2-naphthol | pH < 12, yellow-orange (HIn); | Red | *Direct titration* Cu(II), Zn, Cd, Sc, In(III) |

(*Continued*)

**TABLE 3** (*Continued*)

| Indicator and Its Preparation | Chemical Name | Colors of Free Indicator Species at Different pH | Color of Metal-Indicator Complexes | Applications |
|---|---|---|---|---|
| | | pH > 13, pink ($In^-$) | | *Back titration* Zn, Cd, Ni, Pb, Fe(III), Ga(III), Sc(III) *Replacement titration* Ca, Mg, Mn, Zn, Cd, Pb, In(III), Ga(III), Hg(II), V, Ni, Co |
| Pyrocatechol Violet (0.1 g in 100 mL water. Stable for many months) | 3,3′-4-trihydroxyfuchsone -2″-sulfonic acid | pH < 1.5, red ($H_4In$); 2 < pH < 6, yellow ($H_3In^-$); 7 < pH < 9, violet ($H_2In^{2-}$); pH > 10, red-purple ($HIn^{3-}$) | Blue, except red with Th (IV) | *Direct titration* Al, Bi, Ca, Cd, Co, Fe(III), In(III), Mn, Ni, Zn, Th(IV) *Back titration* Al, Bi, Ca, Cd, Co, Fe(III), In(III), Mn, Ni, Zn, Th(IV) |
| Xylenol Orange (0.5 g in 100 mL ethanol. Stable indefinitely) | 3,3′-Bis[*N*, *N*-di(carboxyl-methyl)aminomethyl]-*o*-cresolsulfonephthanlein | pH < 6.3, lemon yellow ($In^{4-}$); pH > 6.5, red/pink ($In^{4-}$) | Red/violet | *Direct titration* Bi(III), Th (IV), Sc(III), La(III), Zn, Pb, Hg(II), Ca, Co, Ta(V). *Back titration* Sn, In, Ga(III), Fe(III), Ni, Co, Cu, U(IV), V, Cr(III), Ti(IV), Zr(IV), Al, Lanthanides |

where $E_0$ is the standard potential of the redox couple. $E$ is the potential when the concentrations of the redox couple are [Ox] and [Red], respectively. $z$ refers to the number of electrons transferred in the half-cell reaction.

For metal ions such as Cu(II) and Fe(III), a reversible redox equilibrium can quickly establish between the reduced and oxidized couple of the metal (i.e., Cu(II)/Cu and Fe(III)/Fe(II)) on the surface of a chemically inert electrode (platinum or graphite). The electrode potential can be measured if it is coupled with a reference electrode (e.g., calomel electrode). The titration curve for the metal ion can be thus detected by the potential of the redox electrode. The EDTA titration of Hg(II) can be also detected by such a potentiometric method because a reversible redox equilibrium between the couple, Hg(II)/Hg, can be established on a dropping mercury electrode or an amalgamated gold electrode. Nevertheless, most of the metals do not have a reversible redox couple suitable for such a detection. The method using the redox couple of the metal to be titrated is thus limited to only a few metal ions (Cu(II), Fe (III), or Hg(II)).

However, Fe(III)/Fe(II) and Hg(II)/Hg redox couples can be used as the indicator electrode for the titration of other metal ions. This can be understood from the complexation equilibria and the Nernst equation. Take Hg(II)/Hg couple as an example, if several drops of Hg-EDTA are added in the sample for the titration of a metal, $M^{n+}$, the corresponding complexation equilbria will be:

$$K_{MY} = \frac{[MY^{n-4}]}{[M^{n+}][Y^{4-}]}, \qquad K_{HgY} = \frac{[HgY^{2-}]}{[Hg^{2+}][Y^{4-}]} \tag{31}$$

where $K_{MY}$ and $K_{HgY}$ refer to the stability constants of the EDTA complexes of $M^{n+}$ and Hg(II), respectively. Combining these equations with the Nernst equation on a dropping mercury electrode at 25°C will give

$$E_{Hg} = E^0_{Hg} + \frac{0.0591}{2}\log\left[Hg^{2+}\right] = E^0_{Hg} + \frac{0.0591}{2}\log\left(\frac{[M^{n+}]}{[MY^{n-4}]} \cdot \frac{[HgY^{2-}]K_{MY}}{K_{HgY}}\right) \tag{32}$$

If Hg-EDTA is not as stable as the EDTA complex of $M^{n+}$, Hg(II) ion will be released from its complex as soon as Hg-EDTA is added into the sample. Then, the potential of the dropping mercury electrode ($E_{Hg}$) is decided by $[Hg^{2+}]$, as shown in the first half of Eq. 32. The endpoint will occur when all of the $M^{n+}$ ion is complexed with EDTA and Hg(II) starts to combine with EDTA.

If Hg-EDTA complex is more stable than the EDTA complex of the metal ion to be titrated, the electrode potential ($E_{Hg}$) is dependent on the concentration ratio, $[M^{n+}]/[MY^{n-4}]$, as shown in the second half of Eq. 32. $[HgY^{2-}]$ is approximately constant during the titration process because only a small amount of Hg-EDTA is added. As the value of $[M^{n+}]/[MY^{n-4}]$ changes dramatically around the equivalence point, the endpoint of the titration can be easily detected by the potential change of the dropping mercury electrode.

Generally, three requirements must be met in order to use Fe(III)/Fe(II) or Hg(II)/Hg redox couple as the electrode indicator for an EDTA titration:

i. The reduction potential of the metal ion should be less than that of the redox couple. No interfering reaction occurs on the electrode. This is true to many metal ions because the reduction potentials of Fe(III)/Fe(II) (0.771 V) and Hg (II)/Hg (0.85 V) are relatively high.
ii. The medium pH should be in the range in which the complex reaction of Fe(III) or Hg(II) with EDTA is not hindered by other processes. When pH is too low (pH $< 1$ for Fe(III); pH $< 2$ for Hg(II)), the EDTA complex of these metals dissociates to such an extent that only a poorly defined titration curve can be obtained. When pH is too high (pH $> 11$), oxygen and hydroxide ions will participate into the reaction, and the titration curve is deteriorated.
iii. Interfering species such as halides, cyanides, and sulfides will form rather stable species with Hg(II) and Fe(III). The presence of these species should be avoided when the redox electrode is used for the endpoint detection.

In spite of the above limitations, the redox pairs, Fe(III)/Fe(II) and Hg(II)/Hg, have been successfully used in the EDTA titrations of many metals. Practically, a dropping mercury electrode is more useful than an Fe(III)/Fe(II) inert electrode because the latter has a problem of being easily poisoned in the titration process. The metals that are often titrated with a dropping mercury electrode include Al, Bi, Ca, Cd, Cu, Co, Pb, Zn, and Ni.

For example, a dropping mercury electrode can be used for the EDTA titration of Cd at pH 5. Before the titration, several drops of Hg-EDTA are added and an electrical cell composing of a dropping mercury electrode and a reference electrode (e.g., a calomel electrode) is inserted in the sample solution. The potential of the dropping mercury electrode ($E_{Hg}$) is recorded when standard EDTA is added in the sample under stirring. As the reduction potential of Cd(II)/Cd couple ($-0.403$ V) is much less than that of the Hg(II)/Hg couple, the potential on the dropping mercury electrode is dependent on the concentration ratio of Cd(II) and Cd-EDTA as shown in Eq. 32. When close to the equivalence point, the complexation of Cd(II) with EDTA is about to complete and the concentration of the "free" EDTA increases quickly. This causes a fast decrease in $[Cd^{2+}]$/[Cd-EDTA] ratio and a sharp decrease in the potential of the dropping mercury electrode. The endpoint can be located graphically or numerically by plotting the electrode potential ($V$) versus the percent of titration ($\phi$).

*The Ion-Selective Electrode (ISE):* In the last two decades, a great deal of efforts has been made in search of ISEs for various metal ions. This is because ISEs can be used to quickly and automatically estimate the concentration of the metal ion of interest in a solution where other metal ions and species are present. Nowadays, ISEs are available for most of the metal ions. As to common metal ions, such as Ca, Cd, Cu (II), and Pb, their ISEs are available in the market with high quality. For other metals,

including transition metals, lanthanides, and actinides, the ISEs can be readily prepared according to the procedures offered in literatures.

The key part of an ISE is the membrane that separates the sample solution from a chamber in which a reference electrode (e.g., AgCl/Ag electrode) is immersed in a solution (sometimes solid powders) of the metal ion to be detected. Although glasses or inorganic crystals are used (e.g., pH electrode), the ISE membrane is normally made of a plastic material containing a polymer (e.g., poly(vinyl chloride)), a plasticizer, an anion excluder (e.g., sodium tetraphenyl borate, NaTBP), and a complexing agent (ionophore) that has high selectivity to the metal ion to be detected. After solidified on the top of the electrode chamber, the membrane is conditioned with a solution of the metal ion to be measured. The potential of an ISE is dependent on the Gibbs energy change in the mass transfer process of the metal ion across the membrane. According to the structure and composition of the membrane, the mass transfer may have a mechanism of ion exchange, solvent extraction, etc. In a certain concentration range, the potential of an ISE is found to have a linear relationship (Nernstian response) with the logarithm of the activity of the metal ion in the sample solution. Especially, a good ISE shows high selectivity to the metal ion to be detected and much lower response to other ionic species. This is because the ionophore used in the membrane is highly selective to the target metal ion. This property makes the ISEs to be able to estimate the concentration of the target metal ion in the solution where multiple ionic species are present.

The dynamic response time of an ISE ranges from 5 to 30 s. The voltage response to the metal concentration is about 20 mV/decade. An ISE normally shows the Nernst ion response in the range from $10^{-1}$ to $10^{-6}$ M of the metal concentration. When coupled with a reference electrode, an ISE can be readily used as the indicator electrode for the EDTA titration. With the fast development of the complexing agents and ISE technology, high-quality ISEs for most metal ions have become available. Particularly, ISEs are very useful for automated titration systems, where a simple and generalized detection method is desired. Table 4 lists common ISEs with their basic properties.

Some disadvantages about the ISE technique are noticed:

i. The pH range is relatively narrow. The pH for most of the ISE membranes to work properly is between pH 3 and 10.
ii. The dynamic response time is often too long for manual titration.
iii. The ISE membranes are not durable. For many ISEs, the membrane has a life of several months. New membranes have to be made when the old ones are poisoned or physically damaged.
iv. Some ISEs are not very selective to alkali metal ions. This will cause problems because the EDTA agents are normally made of sodium or potassium dihydrogen salts of EDTA.
v. As the potential response of most of the ISEs is not lower than $10^{-6}$ M, a titration with the metal concentration less than $10^{-4}$ M may be problematic.

**TABLE 4 A List of Some Common ISEs WITH BASIC SPECIFICATIONS**

| ISE | Membrane/ Ionophore– | Conc. Range (M) | pH Range | Slope (mV/decade) | Resp. Time (s) | Interference |
|---|---|---|---|---|---|---|
| Ca-ISE [12] | Polymer | $1–5 \times 10^{-6}$ | 3–10 | $27 \pm 2$ | 30 | Pb(II), Hg(II), Cu(II), and Ni. |
| Cd-ISE [12] | Solid state | $0.1–5 \times 10^{-7}$ | 2–12 | $57 \pm 2$ | 20 | Ag, Hg(II), Cu(II). High conc. of Pb(II) and Fe(II). |
| Cu-ISE [12] | Solid state | $0.1–5 \times 10^{-8}$ | 2–12 | $27 \pm 2$ | 20 | Ag, Hg(II), high conc. of $Cl^-$, $Br^-$, Cd(II) and Fe(II). |
| Pb-ISE [12] | Solid state | $0.1–5 \times 10^{-6}$ | 3–8 | $25 \pm 2$ | 20 | Ag, Hg(II), Cu(II). High conc. of Pb(II) and Fe(II) |
| Al-ISE [13] | PVC/Morin | $0.1–5 \times 10^{-7}$ | 3.5–5 | $19.7 \pm 0.1$ | 5 | High conc. of Fe(III), Cr(III), K, Na. |
| Co-ISE [14] | PVC/calixarene | $0.1–5.3 \times 10^{-6}$ | 2.5–6.0 | $30 \pm 0.1$ | 10 | High conc. of Zn, Hg(II), Pb(II), Ag, Ni, Cr(III). |
| Cr-ISE [15] | PVC/AHMTO | $0.1–1.0 \times 10^{-6}$ | 2.7–6.6 | $19.7 \pm 0.3$ | $<10$ | High conc. of Al, La, VO(II), Cu(II), Ce(III). |
| Cu-ISE [16] | PVC/Porphyrin | $0.1–4.4 \times 10^{-6}$ | 2.8–7.9 | $29.3 \pm 0.2$ | 8 | Ni, High conc. of Cd, Sr, K. |
| Dy-ISE [17] | PVC/BBH | $0.01–1.0 \times 10^{-6}$ | 3.0–8.0 | $20.1 \pm 0.8$ | $<20$ | Na, Yb(III), Tm, Ho. |
| Gd-ISE [18] | PVC/PyTu4NO2 | $0.1–1.0 \times 10^{-7}$ | 4.0–9.0 | $20.0 \pm 0.3$ | 10 | High conc. of Sm, Eu, Yb. |
| La-ISE [19] | PVC/NPYFH | $0.1–1 \times 10^{-6}$ | 3.5–10 | $19.2 \pm 0.6$ | 10 | High conc. of Sm, Nd, Eu, Yb(III). |
| Lu-ISE [20] | PVC/TPD | $0.01–1 \times 10^{-6}$ | 4.5–8 | $20.5 \pm 0.4$ | $<10$ | Nd, high conc. of Dy, Cu(II), Ho. |
| Ni-ISE [21] | PVC/NED | $0.1–5 \times 10^{-6}$ | 3.6–7.4 | $29.9 \pm 1$ | 20 | Na, K, Ba, Zn, Ag, Hg(II), Pb(II). |
| Pr-ISE [22] | PVC/PBH | $10^{-3}–10^{-8}$ | 3.0–8.5 | 21.1 | 10 | No interference. |
| Sm-ISE [23] | PVC/SnBP | $0.1–1 \times 10^{-5}$ | 4.0–10 | $40 \pm 0.9$ | $<10$ | No interference. |
| Sr-ISE [24] | PVC/Calixarenes | $0.1–3.2 \times 10^{-5}$ | 3–10 | 30 | 10 | Pb(II), Cd(II), Zn, Cs, high conc. of other metal ions. |
| Tb-ISE [25] | PVC/ATO | $0.1–1 \times 10^{-6}$ | 3.8–8.2 | $19.4 \pm 0.5$ | 15 | High conc. of Ga(III), Ni, Na. |
| Th-ISE [26] | PVC/DPTD | $0.01–1 \times 10^{-6}$ | 3.0–9.0 | $15.2 \pm 0.5$ | 30 | No interference. |
| Tm-ISE [27] | PVC/DADS | $0.01–1 \times 10^{-6}$ | 4.8–8.5 | $19.5 \pm 0.3$ | $<15$ | Er, Lu, Pb(II), Yb(III), Pr |
| V-ISE [28] | Graphite/NTHA | $2 \times 10^{-3}–1 \times 10^{-6}$ | 6 | – | – | Cr(III), Mo(III), Co(II). |
| Yb-ISE [29] | PVC/PMTO | $0.1–1 \times 10^{-6}$ | 3.3–8.0 | $19.5 \pm 0.3$ | 10 | High conc. of Sm, La, Cd, Na |

## 8 TYPES OF EDTA TITRATIONS

As discussed in previous sections, all EDTA titrations are based on the same principle – formation of stable 1:1metal-EDTA complexes and detection of the equivalence point at which an equal amount of the metal and EDTA are combined. However, such an ideal procedure is often complicated under practical circumstances by the factors including slow kinetics of metal-EDTA complexation, "blocking" effect of trace metals on the endpoint detection, hydrolysis of the metal ion, and existence of multiple metal ions and other chemical species. Accordingly, various procedures have been developed to deal with the analytical tasks in the real world. These procedures include direct titration, back titration, replacement titration, selective titration, and indirect titration. The selective titration can be further divided into pH-adjustment, masking with agent, and kinetic masking. All these approaches will be discussed in the following.

***Direct Titration:*** The simplest procedure is to directly titrate the sample with standardized EDTA. Except at very low or very high pH (pH $< 2$ or pH $> 12$), the addition of a pH buffer ($HAc/Ac^-$ for ~pH 5, Hexamine for ~pH 6, $NH_3/NH_4NO_3$ for ~pH 10) is generally necessary to control the pH in a narrow range. If the titration needs to be conducted at a pH where the hydrolysis of metal ion is significant, an auxiliary complexing agent (e.g., ammonia water, citrate, tartrate, and triethanolamine) is added to prevent the hydrolysis. It is noted that when the titration is needed to be conducted in the pH range between 9 and 10, concentrated ammoniacal solution is often used as the pH buffer as well as the auxiliary complexing agent. For example, Zn, Cd, Ni, and Cu(II) ions can form soluble ammine complexes at pH 10 and be successfully titrated with EDTA.

It is not always straightforward to find a suitable endpoint detection method for a direct titration. For some metals, one may run into a problem that the metal-ion indicator does not produce a sharp color change at the endpoint. A valuable way to deal with this problem is to add several drops of the EDTA complex of another metal that can give a sharp color change at the endpoint. Similarly, if the metal to be titrated does not have a reversible redox couple for the potentiometric detection, Fe(III)/Fe(II) or Hg(II)/Hg reversible redox couple may be used to potentiometrically detect the endpoint. It is noted that the metal added for the endpoint detection is normally in its EDTA complex form, which makes it unnecessary to correct the titration result because the second metal added for the endpoint detection will not consume any EDTA in the titration. Such method is called "partial replacement titration," and will be discussed in detail in the following section.

Some trace metals contained in a sample can "block" certain metal-ion indicators. For example, trace amount of copper combines irreversibly with Eriochrome Black T and hinders the use of this indicator in the direct titration of Ca, Mg, Ba, and Sr. Under such situations, an agent complexing more strongly with the trace metal is often added to prevent the trace metals from interfering with the indicator. Such a complexing agent is called "masking agent." For example, trace metals such as Ag, Cd, Co, Cu(II), Hg(II), Ni, Pd(II), Pt, and Zn can be masked by potassium cyanide. A reducing agent such as ascorbic acid or hydroxylamine can change the oxidation states of some

oxidative metal ions (Cr(III), Fe(III), Cu(II), etc.) and protect the indicators from being "blocked" by these metal ions. When a masking agent is used, enough time should be permitted for the masking reactions to occur. After that, the indicator can be added for the direct titration.

Slow kinetics often occur to the reactions between the metal ion and EDTA or between the metal ion and the metal-ion indicator. A direct titration may be influenced by such slow kinetic processes. An easy solution to this problem is to heat the sample solution and keep it at an elevated temperature during the titration. For example, in the direct titration of Mg, a sharp color change will appear at the endpoint only when the titration is conducted at an elevated temperature (>50°C). This is because the Mg-EDTA complex forms at a slow rate and causes the titration sluggish at room temperature.

Table 5 lists typical direct EDTA titrations for some metal ions. The titration procedures in Table 5 are arranged in the ascending order of the medium pH at which the titration is conducted. Common masking agents used to remove the interference under various situations are also listed in the table. Further discussion on utilizing masking agents is included in the section "Use of Masking Agents." In Table 5, only metal-ion indicators are listed as the endpoint detection method. However, other methods such as the potentiometric methods with redox electrodes or ISE electrodes can also be used in many titrations listed in the table. The basic properties of common ISEs can be found in Table 4.

***Back Titration:*** Back titration is a procedure designed to overcome most of the difficulties encountered in direct titration. A back titration is composed of two basic steps: i) add an excess amount of EDTA to the sample and ii) "back-titrate" unreacted EDTA with a standard metal solution under a suitable condition. Both of the EDTA agent and the back-titration metal solution employed in the back titration should be standardized. It is worthwhile to note that the EDTA complex of the metal ion used to back-titrate the excess EDTA should be less stable than that of the metal to be determined. Otherwise, the latter may be replaced from its EDTA complex in the process of titration.

As the metal to be determined is complexed with EDTA before the titration, problems such as the formation of insoluble hydroxides, the existence of trace metals that block the endpoint indicator, and the slow kinetic processes of complexing reactions no longer affect the titration. This is the reason why the back titration can overcome most of the difficulties encountered in the direct titration.

For example, Al and Zr(IV) are readily hydrolyzed in pH 9–10, which is the proper range for a direct titration with Eriochrome Black T. The metal hydrolysis leads to a sluggish endpoint color transition and large analytical error. If a back titration is employed, excess amount of EDTA is added in the sample at lower pH to ensure a complete complexation (with boiling); then the unreacted EDTA is titrated with a standard Mg solution at pH 10. For example, metal ions, including Cu(II), Ni, and Co, form complexes with Eriochrome Black T that are more stable than the complexes with EDTA. This makes the indicator unsuitable for the direct titration of the above metal ions. With a back titration, excess amount of EDTA is added before the addition of the indicator. Then, the unreacted EDTA can be titrated by Pb, Zn, or Mg standard solution, using Eriochrome Black T as the indicator.

**TABLE 5 Typical Direct Titrations for Some Metal Ions**

| Metal Ion | pH | Buffer | Endpoint Detection Method | Remarks |
|---|---|---|---|---|
| Zr(IV) | <1 | HCl | Eriochrome Black T (b/r) [30] | Boil the acidic solution before titration. Ascorbic acid reduces Hg(II) and Fe(III). Sulfate masks Th and Ti(IV). |
| | | $HNO_3$<br>$H_2SO_4$ | Xylenol Orange (r/y) [30] | |
| Bi(III) | 1–3 | $HNO_3/NO^-_3$ | Xylenol Orange (r/y) [31] | Ascorbic acid reduces Hg(II) and Fe(III). Sulfate masks Th (IV) and Ti(IV). Fluoride masks Al, Ba, Be, Fe(III) (in small amount), Mg, Nb, rare earths, Sr, Ta(V), Ti(IV), and Zr(IV). |
| | | | Pyrocatechol Violet (b/y) [32]<br>PAN (r/y) | |
| VO(II) | 2 | $HClO_4/ClO_4^-$ | Xylenol Orange (r/y) [30] | Phosphate masks W(IV). |
| Sc(III) | 2–3 | $HAc/Ac^-$ | Cu-PAN (r/y) [30]<br>Murexide (y/v) [30] | |
| Th(IV) | 2–3 | $HAc/Ac^-$ | Xylenol Orange (r/y) [31] | Ascorbic acid reduces Hg(II) and Fe(III). $H_2O_2$ masks W(IV). 2,3-Mercaptopropanol-1 masks Bi and Pb at pH 3. |
| | | | Methylthymol blue (b/y) [31]<br>Pyrocatechol Violet (r/y) [30] | |
| Tl(III) | 2–5 | $HAc/Ac^-$ | PAN (R/Y) [30]<br>Xylenol Orange (r/y) [30, 31] | Heat the solution to 80°C. |
| Fe(III) | 2–5 | $HAc/Ac^-$ | Xylenol Orange (r/y) [32]<br>Pyrocatechol Violet (r/b) [32]<br>Sulphosalicylic acid (v/c) [33]<br>Cu-PAN (r/y) [32] | $H_2O_2$ masks W(IV). Thiourea masks Cu(II). |
| In(III) | 2–4 | $HAc/Ac^-$ | Cu-PAN (r/y) [34, 35] | Ascorbic acid reducees Hg(II) and Fe(II). Cyanide masks Ag, Cd, Co, Cu(II), Hg(II), Ni, Pd(II), Pt, and Zn. Fluoride masks Al, Ba, Be, Ca, Fe(III) (in small amount), Mg, Nb, rare earths, Sr, Ta, Ti(IV), and Zr(IV). Formic acid or formaldehyde reduces Tl(III). |

(*Continued*)

**TABLE 5** (*Continued*)

| Metal Ion | pH | Buffer | Endpoint Detection Method | Remarks |
|---|---|---|---|---|
| Al | 2–3 | HAc/Ac$^-$ | Cu-PAN (r/y) [30, 35, 36] | Titrate in hot HAc solution. 1,10-phenanthroine masks Cd, Co, Cu(II), Mn, Ni, $UO_2$(II), and Zn. |
| Cu(II) | 4 | HAc/Ac$^-$ | Murexide (y/r) [30] | Fluoride masks Al, Ba, Be, Ca, Fe(III) (in small amount), Mg, Nb, rare earths, Sr, Ta, Ti(IV), and Zr. Citrate masks Fe(III) and Mo(VI). Tartrate masks Sb(III), Sn(IV), and Zr(IV). |
| Cd | 5 | HAc/Ac$^-$ | Xylenol Orange (r/y) [31] | Fluoride masks Al, Ba, Be, Ca, Fe(III) (in small amount), Mg, Nb, rare earths, Sr, Ta, Ti(IV), and Zr. Triethanolamine masks Al, Fe(III), Sn(IV), and Ti(IV), and small amount of Mn. |
| | | | PAN (r/y) [30, 37]<br>Cu-PAN (r/y) [35] | |
| Fe(II) | 5–6 | Hexamine | Xylenol Orange (r/y) [31] | Add ascorbic acid to prevent oxidation. |
| Co | 6 | Hexamine | Xylenol Orange (r/y) [30, 31] | Citrate masks Sn(II), Th(IV), and Zr(IV). Pyrophosphate masks Cr(III) and Fe(III). Tartrate masks Sb(III), Sn(IV), and Zr(IV). Thosulfate masks Cu(II). Triethanolamine masks Al, Fe(III), Sn(IV), and Ti(IV). |
| | | | Cu-PAN (r/y) [9] | |
| Pb | 6 | Hexamine | Xylenol Orange (r/y) [30, 31] | Cyanide masks Cd, Co, Cu(II), Hg(II), Ni, Pd(II), Pt, and Zn. Tartrate masks Sb(III), Sn(IV), and Zr(IV). 1,10-phenanthroine masks Cd, Co, Cu(II), Mn, Ni, $UO_2$(II), and Zn. |
| | | | Cu-PAN (r/y) [35] | |
| Hg(II) | 6 | Hexamine | Xylenol Orange (r/y) [31]<br>Cu-PAN (r/y) [30, 35] | |
| Zn | 6 | Hexamine | Xylenol Orange (r/y) [30, 31] | Fluoride masks Al, Ba, Be, Ca, Fe(III) (in small amount), Mg, Nb, rare earths, Sr, Ta, Ti(IV), and Zr. Ascorbic acid reduces Cu(II). Citrate masks Al, Sn(II), Th, and Zr. |
| | | | Cu-PAN (r/y) [30, 35] | |

| | | | | |
|---|---|---|---|---|
| Sn(II) | 6 | Hexamine | Xylenol Orange (r/y) [31] | |
| Ni | 7–10 | $NH_3/NH_4^+$ | Murexide (y/v) [30, 31, 38], Bromophyrogallol red (b/r) [31]. | Triethanolamine masks Al, Fe(III), Sn(IV), and Ti(IV), and small amount of Mn. $H_2O_2$ masks W(IV). Citrate masks Al, Sn(II), Th, and Zr. Tartrate masks Sb(III), Sn(IV), and Zr(IV). |
| Zn | 8-9 | $NH_3/NH_4^+$ | Murexide (y/v). | Fluoride masks Al, Ba, Be, Ca, Fe(III) (in small amount), Mg, Nb, rare earths, Sr, Ta, Ti(IV), and Zr(IV). Ascorbic acid reduces Cu(IV). Citrate masks Al, Sn(IV), Th(IV), and Zr (IV). Carbonate masks U(IV). |
| Ga(III) | 9 | $NH_3/NH_4^+$ | Eriochrome Black T (r/b) [30]. | Fluoride masks Al, Ba, Be, Fe(III) (in small amount), Mg, Nb, rare earths, Sr, Ta, Ti(IV), and Zr(IV). |
| Mn(II) | 10 | $NH_3/NH_4^+$ | Eriochrome Black T (r/b) [31, 39] | Add some ascorbic acid or hydroxylamine to prevent oxidation and precipitation. |
| In(III) | 8–10 | $NH_3/NH_4^+$ | Eriochrome Black T (r/b) [30, 40] Pyrocatechol Violet (b/y) [30] | Add tartrate to prevent precipitation. Boil the solution before titration. Ascorbic acid reduces Hg(II) at pH 2. Cyanide masks Cd, Co, Cu(II), Hg(II), Ni, Pd(II), Pt, and Zn. Fluoride masks Al, Ba, Be, Ca, Fe(III) (in small amount), Mg, Nb, rare earths, Sr, Ta, Ti(IV), and Zr(IV). |
| Zn | 10 | $NH_3/NH_4^+$ | Eriochrome Black T (r/b) [30, 31] | Fluoride masks Al, Ba, Be, Ca, Fe(III) (in small amount), Mg, Nb, rare earths, Sr, Ta, Ti(IV), and Zr(IV). Ascorbic acid reduces Cu(II). Citrate masks Al. Carbonate masks U(IV). |
| Mg | 10 | $NH_3/NH_4^+$ | Eriochrome Black T (r/b) | |
| Ba | 12 | $NH_3/NH_4^+$ | Methylthymol blue (r/gr) [31] | Cyanide masks Cd, Co, Cu(II), Hg(II), Ni, Pd(II), Pt, and Zn. Citrate masks Al, Sn(II), Th(IV), and Zr(IV). Carbonate masks U(IV). Tartrate masks Sb(III), Sn(IV), and Zr(IV). |
| Ca | $>12$ | NaOH | Murexide (r/v) [30,38] Methylthymol blue (r/gr) [31] | Cyanide masks Cd, Co, Cu, Hg, Ni, Pd(II), Pt, and Zn. Citrate masks Al, Sn(II), Th(IV), and Zr(IV). Carbonate masks U(IV). Tartrate masks Sb(III), Sn(IV), and Zr(IV). |

Note: The color change of each metal-ion indicator is shown by the characters with the meanings: r = red; b = blue; y = yellow; gr = gray; v = violet.

Furthermore, many metal ions such as Cr(III), Ti(IV), and U(IV) complex with EDTA at slow rates, making the direct titration of these metal ions very difficult to conduct. If excess amount of EDTA is added and the sample solution is boiled for enough time, the metal to be titrated will be completely complexed with EDTA. Then, the unreacted EDTA can be readily titrated with a standard metal solution under a suitable condition.

In a back titration, the procedure to titrate the excess EDTA should be easy to operate and produce a sharp signal for the endpoint. The titration procedures frequently used at different pH are listed in the following:

Proc. (i): Back titration at pH 2–3 with Xylenol Orange. Adjust to pH 2–3 with a strong acid. Titrate the excess EDTA with a standard Th(IV) solution using Xylenol Orange. The color changes from yellow to red at the endpoint.

Proc. (ii): Back titration at pH 4–6 with PAN or Xylenol Orange. Adjust to pH 4–6 with $HAc/Ac^-$ buffer or hexamine powders. Titrate the excess EDTA with a standard Cu(II) or Zn solution using PAN or Xylenol Orange. The color changes from yellow to red at the endpoint.

Proc. (iii): Back titration at pH 5–7 with Pyrocatechol Violet. Adjust to pH 5–7 with $HAc/Ac^-$ buffer. Titrate the excess EDTA with a standard Zn, Cu(II), or Pb solution using Pyrocatechol Violet. The color changes from yellow to blue at the endpoint.

Proc. (iv): Back titration at pH 7–9 with Murexide. Adjust the solution to pH 7–9 with hexamine or buffer. Titrate the excess EDTA with a standard Ni, Co, Zn, or Cu(II) solution using Murexide. The solution changes from red to yellow.

Proc. (v): Back titration at pH 9–10 with Eriochrome Black T. Adjust the solution to pH 9–10 with buffer. Boil the solution and titrate the excess EDTA with a standard Mg, Zn, or Pb solution using Eriochrome Black T. The solution changes from blue to red.

Table 6 lists the back titrations for some metal ions. In the table, the titration procedures used to determine the unreacted EDTA are listed as Proc. (i), Proc. (ii), etc., corresponding to the procedures summarized above. More examples of back titrations can be found in literatures [30, 31].

***Replacement Titration:*** This procedure is also called substitution titration or displacement titration. In a replacement titration, an excess amount of the EDTA complex of the replacement metal is added to the sample. Equivalent amount of the replacement metal will be liberated from its EDTA complex by the metal ion to be determined, and the metal liberated is titrated with EDTA under a proper condition. The replacement metal should meet the following conditions: (*i*) its EDTA complex is less stable than that of the metal to be determined and (*ii*) a good endpoint detection method is available for the direct titration of the replacement metal.

The replacement titration is similar as the back titration in that both procedures employ the complexation between EDTA and a second metal as the titration reaction, and both procedures can overcome the problems in the direct titration. However, in a

**TABLE 6 Typical Back Titrations for Some Metal Ions**

| Metal Ion | Addition of EDTA | Back-Titrated by | Remarks |
|---|---|---|---|
| Co(III) | Add $HNO_3$ adjust to pH 2. Add $H_2O_2$ and excess EDTA. | Proc. (i) at pH 2 | $H_2O_2$ prevents Co(III) from being reduced to Co(II) |
| Zr(IV) | Adjust to pH 2–3 with $HNO_3$. Add excess EDTA and heat the solution. | Proc. (i) | |
| U(IV) | Adjust to pH 2–3. Add excess EDTA and boil the solution. | Proc. (i) | |
| Sc(III) | Add excess EDTA. | Proc. ii) at pH 4. | |
| Al | Add excess EDTA and HAc to adjust to pH 3. Boil for $>2$ min. | Proc. (ii) at pH 4–6. | 1,10-phenanthroline masks Cd, Co, Cu(II), Mn, Ni, $UO_2$(II), and Zn |
| Cr(III) | Add excess EDTA at pH 4–4.5. Boil for 15 min. | Proc. (ii) at pH 4–5<br>Proc. (iv) at pH 9<br>Proc. (v) at pH 9–10 | |
| Sb(III) | Adjust to pH 4. Add excess EDTA. | Proc. (ii) at pH 4 | |
| Sn(IV) | Add excess EDTA at pH 5. Heat to 70–80°C. | Proc. (iii) at pH 5 | |
| Ga(III) | Add excess EDTA.<br>Fluoride masks Al, Ba, Be, Ca, Fe(III) (in small amount), Mg, Nb, rare earths, Sr, Ta, Ti(IV), and Zr(IV). Formic acid or formaldehyde reduces Tl(III). | Proc. (iv) at pH 9<br>Proc. (v) at pH 9 | |
| Hg(II) | Add excess EDTA. | Proc. (iv) at pH 9<br>Proc. (v) at pH 9 | |

(*Continued*)

**TABLE 6** (*Continued*)

| Metal Ion | Addition of EDTA | Back-Titrated by | Remarks |
|---|---|---|---|
| V(IV) | Adjust to pH 6. Add EDTA and heat the solution. | Proc. (ii) at pH 6 | |
| Mo(V) | Add excess EDTA. Heat to boiling. Cool to 40–60°C | Proc. (v) at pH 9–10 | |
| Pd(II) | Add excess EDTA. | Proc. (i) at pH 3 | |
| Mn | Adjust to pH 4.5 with NH3/$NH_4^-$. Heat to 60°C and add excess EDTA. | Proc. (ii) | Cyanide masks Cd, Co, Cu, Hg(II), Ni, Pd(II), Pt, and Zn. Fluoride masks Al, Ba, Be, Ca, Fe(III) (in small amount), Mg, Nb, rare earths, Sr, Ta, Ti(IV), and Zr(IV). |
| In(III) | Add tartaric acid. Adjust pH to 8–10. Add excess EDTA and heat to boiling. | Proc. (v) | Ascorbic acid reduces Hg(II) and Fe(III). Cyanide masks Ag, Cd, Co, Cu, Hg(II), Ni, Pd(II), Pt, and Zn. Fluoride masks Al, Ba, Be, Ca, Fe(III) (in small amount), Mg, Nb, rare earths, Sr, Ta, Ti(IV), and Zr(IV). Formic acid or formaldehyde reduces Tl(III). |
| Co | Add excess EDTA and adjust to pH 9–10. | Proc. (v) | Citrate masks Sn(II), Th(IV), and Zr(IV). Pyrophosphate masks Cr(III) and Fe(III). Tartrate masks Sb(III), Sn(IV), and Zr(IV). Thosulfate masks Cu(II). Triethanolamine masks Al, Fe(III), Sn(IV), and Ti(IV). |
| Ni | Add excess EDTA. | Proc. (v) | Triethanolamine masks Al, Fe(III), Sn(IV) and Ti(IV), and small amount of Mn. $H_2O_2$ masks W(IV). Citrate masks Al, Sn(II), Th(IV), and Zr(IV). Tartrate masks Sb(III), Sn (IV), and Zr(IV). |
| Ca | Adjust to pH 10 and add excess EDTA. | Proc. (v) | Cyanide masks Cd, Co, Cu, Hg(II), Ni, Pd(II), Pt, and Zn. |

replacement titration, there is no need to accurately quantify the amount of the metal-EDTA agent added in the sample. This simplifies the procedure. Compared with the back titration, replacement titration may yield smaller analytical error because only the EDTA agent for the titration needs to be accurately quantified.

Mg-EDTA solution is the most common replacement agent because Mg-EDTA complex is less stable than the EDTA complexes of most metal ions. Mg is readily titrated at pH 10 at an elevated temperature (>50°C), using Eriochrome Black T as the metal-ion indicator. Mn-EDTA and Zn-EDTA solutions are also useful for the replacement titration of the metals that form more stable EDTA complexes.

Sometimes a complex other than metal-EDTA is used in the replacement titration. Such an arrangement can be used to titrate the species that does not form stable complexes with EDTA. For example, a nickel complex, $Ni(CN)_4^{2-}$, is used as the metal complex in the replacement titration of silver ion. As $Ag(CN)_2^-$ is much more stable than $Ni(CN)_4^{2-}$ (at 25°C, $K(Ag(CN)_2^-) = 10^{20.4}$, $K(Ni(CN)_4^{2-}) = 10^{31.3}$), the following replacement reaction will occur quantitatively:

$$Ni(CN)_4^{2-} + 2Ag^+ \rightleftharpoons 2Ag(CN)_2^- + Ni^{2+} \tag{33}$$

The nickel liberated from its cyanide complex can be directly titrated with EDTA, and the silver can be determined from the stoichiometry of the above reaction. Since the cyanide complexes of both Ni and Ag are much more stable than their EDTA complexes, the cyanide complexes will not interfere with the titration. Therefore, the requirements for the second complexing agent used in the replacement titration can be summarized as follows: (*i*) its complexes with both of the metal ions are more stable than the corresponding metal-EDTA complexes and (*ii*) its complex with the metal ion to be determined is more stable than that with the replacement metal ion.

If poor endpoint detection is the only reason for using the replacement titration, a complete replacement is not necessary. A partial replacement titration that uses a small amount of the EDTA complex of the displacement metal is quite enough to give a sharp endpoint signal. For example, Cu-PAN, a widely used metal-ion indicator, is actually an agent using the principle of partial replacement titration. Cu-PAN is a solution containing Cu-EDTA complex and PAN indicator. As Cu-PAN complex is rather stable, many metal ions can replace Cu from its EDTA complex to lead to the formation of Cu-PAN complex, which has a bright-red color. At the equivalence point, when the first drop of excess EDTA is added, EDTA will replace PAN from Cu-PAN complex. The solution color is turned from red into yellow, which is the color of PAN indicator itself.

***Selective Titration:*** The EDTA titrations of a single metal ion are relatively simple and easy to operate. Difficulties may be encountered if the sample contains multiple metal ions and if multiple metal ions existing in one sample are to be titrated. Theoretically, a metal ($M_I$) can be selectively titrated in the presence of the second metal ($M_{II}$) only when

$$\frac{K'_{M_I Y}}{K'_{M_{II} Y}} > 10^8 [M_{II}] \tag{34}$$

where $K'_{M_IY}$ and $K'_{M_{II}Y}$ are the conditional stability constants for $M_I$ and $M_{II}$, respectively. $[M_{II}]$ is the molarity of the second metal ion. Such a condition is often unsatisfied as to the samples in real world. Extra measures have to be taken to selectively titrate the metal ion coexisting with other metal ions.

Most of the selective titration procedures are based on the addition of a chemical agent to alter the selectivity of EDTA over the metals in question. These procedures, including pH-adjustment, masking with agent, and kinetic masking, will be discussed in the following. Other procedures, mainly based on the separation of metals into different phases before titration, are principally not selective titrations and will not be discussed in this section.

*pH-Adjustment:* It can be seen from Table 1 the stability constants of EDTA complexes of the metal ions can be roughly divided into three groups:

| | |
|---|---|
| Tri- and tetravalent metal ions, including Co(III), V(III), In(III), Th(IV), Fe(III), Sc, Hg(II), Ga, Zr(IV), etc. | Log $K > 20$ |
| Divalent transition metals, rare earths, Pb, Al, U(IV), Ti(IV), etc. | $19 > \log K > 12$ |
| Alkaline earths | $11 > \log K$ |

From Fig. 4, one can see that the metals in the first group can be titrated at pH as low as pH 1–3. The second group can be titrated at pH as low as pH 4–6, and the third group can be titrated only in the range of pH 8–11. This suggests that a sample containing metals from different groups may be successively titrated by controlling the medium pH. However, it is noticed that difficulties may occur discerning several endpoints when more than one metal-ion indicators are used. The colors of multiple indicators often interfere with each other to make the visual detection impossible. Under such circumstances, detection methods such as potentiometric methods become important complementary methods for a sequential titration of multiple metal ions.

For example, a mixture of Bi(III), Pb, and Mg can be sequentially titrated by controlling the pH. At acidic medium (pH 1–2), Bi(III) can be directly titrated with a Fe (III)/Fe electrode and several drops of Fe-EDTA complex for the detection of the endpoint. Then, the solution is adjusted to pH 5 with $HAc/Ac^-$ buffer and Pb can be directly titrated with the Fe(III)/Fe electrode or a Pb-ISE. Finally, the solution is brought up to pH 10 with ammoniacal buffer and heated to $>50°C$. The Mg is directly titrated with Eriochrome Black T, which changes color from red to blue at the endpoint.

In a few cases, one metal-ion indicator is effective in a wide pH range and can be used to detect several endpoints. This makes it possible to conduct some sequential EDTA titrations with a single metal-ion indicator, without running into the trouble of color interference due to multiple metal-ion indicators. For example, a sample composed of a trivalent/tetravalent metal ion (Th(IV), Fe(III), or Bi(III)) and a bivalent metal ion (Zn, Pb, Cd, or Co) can be successively titrated using Xylenol Orange as indicator. As to such a sample, the solution is first adjusted to pH 1–3 and the trivalent/tetravalent metal ion is titrated. At the endpoint, Xylenol Orange is replaced by EDTA from the metal complex and it changes from red to yellow. Then,

the solution is neutralized, and hexamine powders are added to control the solution at pH 6. The solution turns into red again as Xylenol Orange complexes with the divalent metal ion. The divalent metal ion can be directly titrated, and the color changes to yellow at the endpoint.

*Masking with Agent:* A masking agent is a substance that can selectively "mask" certain metals from complexing with EDTA and make it possible for the titration of some other metals. According to the mechanism of the process, the masking processes can be sorted as complexation masking, redox masking, and precipitation masking. Since the pH-adjustment is an approach to alter the EDTA selectivity over metals by changing the proton concentration, it can be considered as one of the masking processes as well.

The agents used for complexation masking are fluoride, cyanide, citrate, tartrate, 2-mercaptoethanol, triethanolamine, etc. For example, tartrate can mask Al, Fe(III), Nb (V), Ta(V), Ti(IV), W(VI), Sb(III), and Sn(IV), enabling the titration of Zn, Bi, Cd, Co (III), Cu(II), Ni, rare earths, and alkaline earths. Ammonium fluoride can mask Al, Ca, and Mg to enable the determination of Bi(III), Cd, Cu(II), Ga(III), In(III), Mn, Ni, Pb, and Zn.

Cyanide ion is one of the effective complexation masking agents. At $pH > 8$, cyanide ion forms stable complexes with the metal ions such as Cd, Zn, Hg(II), Cu(II), Co, Fe(II), Fe(III), Ni, Hg(II), Ag, Pd(II), Tl(III), and Pt(II), but do not interact with alkaline earths, Mn(II), and Pb. Therefore, the metal ions in the latter group can be selectively titrated at alkaline medium where the metals in the former group are masked by cyanide.

Moreover, Cd and Zn can be de-masked from their cyanide complexes by the addition of formaldehyde at room temperature. Formaldehyde liberates Cd and Zn from their cyanide complexes by forming glycolonitril, $HOCH_2CN$, with cyanide ion. Other metals will not be affected unless large excess amount of formaldehyde is added. This property can be used to determine Cd or Zn in the sample in which cyanide has to be added as masking agent.

Two points should be noticed when cyanide is used as masking agent: First, cyanide can only be used in alkaline medium because of its formation of deadly gas HCN in acidic medium. Before the addition of the cyanide agent, the sample solution should be brought to alkaline pH. This may lead to the formation of the metal hydroxides that are difficult to decompose. To avoid such a problem, tartaric acid or other complexing agent is added before raising the medium pH. Then, cyanide is added to combine with the metal ions to form stable cyanide complexes. Second, Fe-cyanide complex can oxidize some metal-ion indicators including Eriochrome Black T. If Fe(III) exists, the alkaline solution is usually boiled with ascorbic acid to reduce Fe(III)-cyanide complex to Fe(II)-cyanide complex before titration.

The agents used for redox masking include ascorbic acid, hydrazine, formaldehyde, hydrogen peroxide, etc. For example, ascorbic acid can reduce Fe(III) to Fe(II) and Tl(III) to Tl(I), masking these metals from interfering with the titration of In(III), Bi(III), Th(IV), or Zr(IV). Cu(II) may be reduced in acidic solution using hydrazine so that Ni or Co can be titrated without interference from the Cu(II).

A typical example of precipitation masking is to titrate Ca at high pH while Mg precipitates as $Mg(OH)_2$. Masking by precipitation is only used in a few of favorable

cases. This is because analytical error is likely to be introduced by co-precipitation, post-precipitation, and other processes related with precipitates. The precipitation can also obscure the color change at the endpoint. In many titrations involving precipitation, separation of the precipitates from the solution before the titration is proven to be valuable. As chemical separation is not considered as selective titration, such techniques are not discussed in this section.

Table 7 lists common masking agents and their applications. More detailed discussions can be found in literatures [30, 31].

*Kinetic Masking:* Sometimes, a selective titration can be devised by taking the advantage of the difference in the kinetic rates of the complexation reactions. For example, Cr(III), in spite of its high stability constant with EDTA, has a very slow rate reacting with EDTA at room temperature. One has to boil the solution with excess EDTA for >15 minutes to transfer Cr(III) completely into Cr-EDTA complex. This property allows Fe(III) or Bi(III) to be titrated in the presence of Cr(III) at pH 1–3 and room temperature. Cr(III) can be determined by a back titration after the sample is boiled with excess EDTA at pH 4 [41].

The slow rate of replacement of Ni from its EDTA complex can be used in kinetic masking as well. In ice-cold solution, Co, Zn, Cd, and Pb can be rapidly replaced from their EDTA complexes by Bi(III). But Ni is only replaced at a very slow rate. This property permits the determination of Ni in the presence of other metal ions [42].

***Indirect Titration:*** The EDTA titration can be used not only to measure the metal ions that form stable 1:1 metal-EDTA complexes but also to measure many other species, including anions, alkali metals, and organic molecules. Theoretically, any species can be determined by an EDTA titration if this species can quantitatively form a compound with a metal ion that can be determined by an EDTA titration and if this compound can be quantitatively separated from the sample. By titrating the metal containing in the compound or the excess metal after the formation of the compound, the species of interest can be determined. For quantitative analysis, the compound containing the metal and the species is often a solid and is separated from the solution before the EDTA titration.

For example, sulfate ion can be determined by first precipitating with a known amount of barium and then titrating the excess Ba with EDTA in the supernatant after the precipitates are separated. Halides or thiocyanate anions can be determined using such an approach: the species is first precipitated with silver ion. Then, the precipitates are isolated and dissolved in a $K_2Ni(CN)_4$ solution at pH 10. Each Ni(II) ion will be liberated by two silver ions in the precipitates. Nickel is titrated with EDTA using Murexide as indicator. The amount of the species can be determined by the stoichiometric relationship between Ni and the species. As to phosphate ion, it can form crystalline precipitates ($MgNH_4PO_4 \cdot 6H_2O$) with Mg and $NH_3$ in a hot solution. After being separated from the sample solution and washed with $NH_3$ water, the precipitates can be dissolved in hydrochloric acid. The Mg from the precipitates can be titrated by direct titration in a warm solution (>50°C) at pH 10 with Eriochrome Black T as indicator. Detailed discussions on the indirect titrations of many other chemical species can be found in literatures (30, 31).

**TABLE 7 Common Masking Agents and Their Behaviors in EDTA Titrations**

| Masking Agent (Type of Masking) | Metal Ions That Can Be Masked | Metal Ions That Can Be Titrated with the Masking Agent |
|---|---|---|
| Ascorbic acid (redox masking) | Cr(III) (in boiling acidic solution), Tl(III), Ti (IV), Cu(II) → Cu(I) (in alkaline medium), Fe(III) → Fe(II) (at pH 2), Hg(II) → Hg (at pH 2) | Ca, Mg, Mn, Ni, Zn, Co, Pb, Al, V(IV), Bi (III), Th(IV), $UO_2$(II), Zr(IV), Ga(III), In(III), and Pd(II) |
| Citrate (complexation masking) | Al (in small amount), Fe(III), Mo(VI), Sn (II), Th(IV), and Zr(IV) at pH 8.5 ~ 9.5 | Zn, Cd, Cu(II), Pb, Co, and Ni |
| Cyanide (complexation masking) | Cu(II), Ag, Cd, Co, Fe(III), Fe(II), Hg(II), Ni, Pd(II), Pt(II), and Zn in alkaline solution. To prevent the precipitation of the metal ions, add tartaric acid before increasing the pH. Cd and Zn can be demasked by adding formaldehyde at room temperature. Ferricyanide ion needs to be reduced with ascorbic acid because the former can oxidize many azo dye indicators. | Al, Ca, Mg, Mn(II), Pb(II), Sr, In(III), and Zn. |
| Fluoride (complexation masking) | Al, alkaline earths, rare earths, Nb, Ta, Ti (IV), Au(III), Sn(IV), and Zr in neutral and alkaline media | Zn, Ga, In, Bi(III), Cd(II), Cu(II), Mn(II), Ni (II), and Pb(II). Add tartrate to prevent these metal ions from forming insoluble hydroxides at higher pH. |
| Hydrazine (redox masking) | Fe(III) → Fe(II), Cu(II) → Cu(I), Mo(VI) → Mo(V) in acidic and neutral media | Bi(III), Ni(II), Th(IV), Mo(V) |
| Hydrogen peroxide (redox masking) | Hg(II) when pH > 2 | Cu(II), Zn |
| Iodide (complexation masking) | | |

(*Continued*)

**TABLE 7** (*Continued*)

| Masking Agent (Type of Masking) | Metal Ions That Can Be Masked | Metal Ions That Can Be Titrated with the Masking Agent |
|---|---|---|
| 2-Mercaptoethanol (complexation masking) | Hg(II), Cu(II), Pd(II), Pb(II), Cd(II), Tl(III) | Co, Ni, Zn, Al, Ce, La, In(III), Zr(IV), Pt(IV), V(IV), $UO_2$(II) |
| 2,4-Pentanedione (complexation masking) | Al, Fe(III), and Pb (by precipitation) | Pb(II) and Zn |
| 1,10-Phenanthroline (complexation masking) | Cd(II), Co(II), Cu(II), Mn(II), Ni(II), Pd(II), Hg(II), $UO_2^{2+}$, and Zn | Al, In, and Pb(II) |
| Tartrate (complexation masking) | Al, Fe(III), Nb(V), Ta(V), Ti(IV), W(VI), Sb (III), Sn(IV), Th(IV), and Zr(IV) | Mn, Mo(V), Bi(III), Cd(II), Co(II), Cu(II), Ni, Zn, and rare earths |
| Thiourea | Cu(II) | Fe(III), Ni(II), Pb(II), Sn(II), and Zn |
| Thiosulfate (redox masking) | Cu(II) | Cd(II), Ni(II), Pb(II), and Zn |
| Triethanolamine (complexation masking) | Al, Fe(III), Fe(II), Ti(IV), and Sn(IV) in alkaline solution | Ni(II), Zn, Mn(II) (add ascorbic acid), Cd (II), Pb(II), and rare earths |

## 9 SOME PRACTICAL CONSIDERATIONS

Although EDTA titration procedures are generally easy to carry out, some operational mistakes frequently happen and produce unpredictable errors in the analytical results. It is worthwhile to pay attention to the following points in the practical operation of EDTA titrations:

i. Many metal-ion indicators are very sensitive to even trace amount of heavy metals. De-ionized water should be always used in the preparation of all solutions for the titration.
ii. The medium pH is critical for the EDTA titration. Many titrations are required to be operated in a narrow pH range. It is suggested to use a calibrated pH meter for the pH adjustment. If a titration is carried out in a strong acidic or strong basic medium where no pH buffer is used, the pH should be monitored as the titration is close to the endpoint. When adjusting pH from acidic to alkaline, a good approach is to carefully bring the pH to neutrality with NaOH or KOH before adding the pH buffer. By doing this, the pH buffer can effectively control the solution within the correct pH range.
iii. The concentration of the metal ion to be titrated is generally controlled between 0.01 and 0.001 M. If the sample concentration is too high, the titration is often deteriorated by complicated chemical processes and blunt endpoint. On the other hand, the metal ion at very low concentration could be titrated if the stability constant of the metal-EDTA complex is sufficiently large and proper endpoint detection methods are available. The concentration effect on the titration curve is discussed in the section "Concentration Effect."
iv. The amount of metal-ion indicator to be added should be the least amount that gives discernable color changes for the titration operator. Generally, this amount is about 2–4 drops of a liquid indicator, or 20–50 mg of a solid indicator. It is noted that too much indicator gives a sharper color change at the endpoint but introduces larger analytical error due to the consumption of the metal ion by the indicator itself.
v. The standard agents, metal-EDTA solutions, and pH buffers for the EDTA titration should have the exact concentration and composition as they are meant to. Good preparation procedures should be followed to make these solutions so that unnecessary errors will not be introduced. Table 8 lists the standard procedures of preparing some agents for the EDTA titration.

## 10 SOME EXAMPLES OF EDTA TITRATIONS

### 10.1 Determination of Bi(III): DIRECT TITRATION AT pH 1-3

*Procedure:* Dilute the sample into a 250-mL volumetric flask. Pipette a 25.00-mL aliquot into a 250-mL conical flask. Monitor pH with a pH meter. If necessary adjust to

**TABLE 8 Preparation Procedures of Some Agents for the EDTA Titration**

| Agent | Preparation |
|---|---|
| 0.01 M EDTA disodium solution | Dissolve and dilute 3.722 g of sodium dihydrogen ethylenediaminetetraacetate dehydrate into a 1000-mL volumetric flask with de-ionized water. Transfer the solution to a polyethylene bottle for storage. |
| Standard EDTA disodium solution | An EDTA solution can be standardized by titrating with a standard Zn, Cu, or Pb solution. For example, when a standard Zn solution is used for standardization, 25.00 mL Zn solution is neutralized with NaOH and mixed with 5 mL of pH 10 ammoniacal buffer solution. Add 2 drops of Eriochrome Black T solution and titrate with the EDTA solution. |
| 0.0100 M standard $ZnCl_2$ solution | Treat 10 g of reagent grade Zn granules with 0.1 M HCl to remove any oxide coating. Clean the treated Zn granules with water and acetone. After dried up, 6.541 g of the Zn granules is weighed and dissolved in minimum amount of 5 M HCl by heating the solution to boil. Dilute the solution into a 1000-mL flask with de-ionized water. |
| 0.0100 M standard $Cu(NO_3)_2$ solution | Rinse 10 g of pure electrolytic Cu metal with diluted $HNO_3$ solution. Dissolve 6.355 g of the cleansed Cu into 120 mL 0.1 M $HNO_3$. Dilute the solution into a 1000-mL flask with de-ionized water. |
| Mg-EDTA solution | Prepare approximately 0.05 M $MgCl_2$ and 0.05 M EDTA solutions. Pipette 25.00 mL of the $MgCl_2$ solution into a conical flask. Add 15 mL of water and 2 drops of Eriochrome Black T solution. Warm to 60°C and titrate with the EDTA titration until the solution turns from red to blue. Mix equivalent amounts of the $MgCl_2$ and EDTA solutions and store in a polyethylene bottle. |
| Cu-EDTA solution | Prepare approximately 0.05 M $CuSO_4$ and EDTA solutions. Pipette 25.00 mL of the $CuSO_4$ solution into a 250-mL conical flask. Add pH 10 ammoniacal buffer dropwise until the solution turns to blue and the precipitate has re-dissolved. Add 100 mL of de-ionized water. Add more pH buffer if the turbidity is present. Add a pinch of Murexide indicator powder and titrate with the EDTA solution. |
| pH 5 acetate buffer | Dissolve 27.3 g of reagent grade sodium acetate trihydrate ($NaC_2H_3O_2 \cdot 3H_2O$) in 60 mL of 1 M HCl solution. Dilute with de-ionized water into a 1-L volumetric flask. |
| pH 10 ammoniacal buffer | Dissolve 70 g of reagent grade $NH_4Cl$ in 570 mL of reagent grade concentrated ammonia water (28 to 30% $NH_3$). Dilute with de-ionized water into a 1-L volumetric flask. Store in a polyethylene bottle. |

pH 1–3 by adding dilute nitric acid or ammonia water under stirring. Add 15 mL of de-ionized water and add 2 drops of Xylenol Orange indicator. Titrate with standard EDTA to the endpoint (from red to yellow). The weight of bismuth ($W_{Bi}$) in the 250-mL volumetric flask can be calculated from the volume ($V_{EDTA}$) and concentration ($C_{EDTA}$) of the standard EDTA used in the titration:

$$W_{Bi}(\text{mg}) = C_{EDTA}(\text{M}) \times V_{EDTA}(\text{mL}) \times \text{Molar mass of Bi} \times 10 \qquad (35)$$

*Interferences:* Alkaline earth metals and most divalent transition metals are masked in this pH range and do not interfere with the titration. The major interference metal ions are Fe(III), Th(IV), Cu(II), Sb(III), Hg(II), and Ti(IV). Aluminum ion in large amount also interferes with the titration. Ascorbic acid can mask Fe(III) and Hg(II) by reducing them to Fe(II) and Hg, respectively. Hydrazine can reduce Cu(II) to Cu(I). Sulfate masks Th(IV) and Ti(IV). Tartrate masks Sb(III). Aluminum can be masked by fluoride.

*Remarks:* Fe(III), Cu(II), Hg(II), Th(IV), Tl(III), and Sc(III) can be also titrated by this procedure. Other metal-ion indicators, including PAN, Pyrocatechol Violet, Cu-PAN, and Murexide, are used for the procedure as well.

## 10.2 Determination of Zn: DIRECT TITRATION AT pH 10

*Procedure:* Prepare the sample into a 250-mL volumetric flask. Pipette a 25.00-mL aliquot into a 250-mL conical flask. If the solution is acidic, add 1 M of NaOH dropwise under stirring until the solution approaches neutrality. Monitor pH with a pH meter. Add 15 mL of water, 5 mL of pH 10 ammoniacal buffer, and 2 drops of Eriochrome Black T immediately prior to the titration. Titrate with standardized EDTA to the endpoint (from red to blue). The weight of zinc ($W_{Zn}$) in the 250-mL volumetric flask can be calculated from the volume ($V_{EDTA}$) and concentration ($C_{EDTA}$) of the standard EDTA used in the titration:

$$W_{Zn}(\text{mg}) = C_{EDTA}(\text{M}) \times V_{EDTA}(\text{mL}) \times \text{Molar mass of Zn} \times 10 \qquad (36)$$

*Interferences:* Many metal ions can interfere with the titration. Masking is an effective method to screen the interference metals. For example, tartrate masks Fe(III), Al, Sb (III), and Sn(IV); 2-mercaptoethanol masks Hg(II), Cu(II), Pd(II), Pb, Cd, and Tl(III). Fluoride masks Al, Ba, Be, Ca, Sr, Fe(III) (in small amount), Mg, rare earths, Ti(IV), and Zr(IV). Ascorbic acid reduces Cu(II). Citrate masks Al, Sn(II), Th(IV), and Zr (IV). Carbonate masks U(IV).

*Remarks:* Many metal ions can be determined by this procedure with little or no modifications. Several metal-ion indicators are available for this procedure. If Eriochrome Black T (red to blue), Calmagite (red to blue), or Pyrocatechol Violet (violet to blue) is used as the indicator, cadmium can be determined by the procedure. Magnesium can be determined when the sample is heated to >50°C to increase the formation rate of Mg-EDTA complex. Calcium, barium, and even strontium can be determined if Mg-EDTA is added. Lead and indium(III) can be titrated if sufficient

tartrate is added to prevent hydroxide formation. Rare earths can be titrated by adding sufficient tartrate or citrate and boiling the solution. Manganese(II) can be determined if some ascorbic acid or hydroxylamine is added to prevent its oxidation and hydroxide formation.

As Cu(II), Ni, and Co form with Eriochrome Black T strong complexes that are too stable to be decomposed by EDTA, Murexide is a suitable indicator for the titration of these metal ions. Nickel can be directly titrated with the procedure. Cobalt can be determined when sufficient acetate is added. Copper can also be titrated by the procedure, but too much ammoniacal buffer is avoided because Cu-$NH_3$ complex is rather stable. If alkaline earths and Pb are determined by this procedure, cyanide is an effective agent to mask Cd, Co, Cu(II), Hg(II), Ni, Pd(II), Pt(II), Zn, and other transition metals.

## 10.3 Determination of Co: BACK TITRATION AT pH 5–6

*Procedure:* Dilute the sample into a 250-mL volumetric flask. Pipette a 25.00-mL aliquot into a 250-mL conical flask. Add excess amount of standard EDTA. Neutralize the solution with NaOH under stirring; use a pH meter to monitor pH. Add 5 mL of acetate buffer and heat the solution to boiling. Add 2 drops of PAN solution and titrate with standard $CuSO_4$ until the solution turns from blue to red-violet. The weight of Co can be calculated from the volumes and concentrations of the standard EDTA and $CuSO_4$ solutions.

$$W_{\mathrm{Co}}(\mathrm{mg}) = [C_{\mathrm{EDTA}}(\mathrm{M}) \times V_{\mathrm{EDTA}}(\mathrm{mL}) - C_{\mathrm{CuSO4}}(\mathrm{M}) \times V_{\mathrm{CuSO4}}(\mathrm{mL})] \times \text{Molar mass of Co} \times 10 \tag{37}$$

*Interferences:* Interferences from various metals can be eliminated by the addition of masking agents before neutralizing the solution. Ascorbic acid can mask Cr(III), Tl (III), Ti(IV), Fe(III), and Hg(II). Fluoride can mask Al, rare earths, and alkaline earths. Citrate masks Al, Sn(II), Th(IV), and Zr(IV). Pyrophosphate masks Cr(III) and Fe (III). Tartrate masks Sb(III), Sn(IV), and Zr(IV). Triethanolamine masks Al, Fe(III), Sn(IV), and Ti(IV).

*Remarks:* Metal ions such as Co, Ni, Fe(III), V(IV), Cu(II), Hg(II), Pb, Zn, Mn, rare earths, Al, Ga(III), and In(III) can be determined by this procedure. It is necessary to boil the solution because many metal ions form EDTA complexes at a slow rate, and Cu-PAN indicator functions properly at elevated temperature. Vanadium(IV) can be determined in the presence of ascorbic acid. As to Al, the sample solution should be boiled for 15 minutes to make sure that Al is completely transferred into its EDTA complex.

## 10.4 Determination of Ca AND Mg: SEQUENTIAL TITRATION

*Procedure:* Dilute the sample into a 250-mL volumetric flask. Pipette a 25.00-mL portion into a 250-mL conical flask. Add 3 drops of a 1% polyvinyl alcohol solution

and adjust the solution over pH 12.5 with KOH. Add 15 mL of de-ionized water, and add 5% of KCN and heat the solution to 60°C. Add 2 drops of Calcon indicator (1-(2-hydroxy-1-naphthylazo-2-naphthol-4-suphonic acid in 4 w/v% methanol). Titrate with standard EDTA to the endpoint (from red to blue). The calcium content is determined.

$$W_{\mathrm{Ca}}(\mathrm{mg}) = C_{\mathrm{EDTA}}(\mathrm{M}) \times V_{\mathrm{EDTA}}(\mathrm{mL}) \times \text{Molar mass of Ca} \times 10 \tag{38}$$

Add 5 mL of pH 10 ammoniacal buffer. Boil the solution until magnesium hydroxide dissolves. The solution turns back to red again because of the Mg-Calcon complex. Add 1 drop of Eriochrome Black T and titrate the solution with standard EDTA to the endpoint (from red to blue). This determines the content of magnesium.

$$W_{\mathrm{Mg}}(\mathrm{mg}) = C_{\mathrm{EDTA}}(\mathrm{M}) \times V_{\mathrm{EDTA}}(\mathrm{mL}) \times \text{Molar mass of Mg} \times 10 \tag{39}$$

*Remarks:* The polyvinyl alcohol reduces the adsorption of Calcon on the surface of $Mg(OH)_2$ precipitate[43]. Cyanide is used to mask Ag, Cd, Co, Cu(II), Fe(II), Hg(II), Ni, Pd(II), Pt(II), and Zn. Interferences from Cu(II), Fe(II), Au(III), Co, Ni, Cd, and Zn can be also masked by fluoride. Fe(III), Al, Ti(IV), and Sn(IV) can be masked by triethanolamine.

## 10.5 Determination of Phosphate Anion: Indirect Titration

*Procedure:* Dilute the phosphate sample into a 250-mL volumetric flask. Pipette a 25.00-mL aliquot into a 250-mL beaker, and add 50-mL of de-ionized water. Add 1 mL of concentrated HCl solution, a few drops of methyl red, and excess amount of $MgCl_2$ solution. Heat the solution to boiling. Add concentrated ammonia water dropwise under vigorous stirring until methyl red turns from red to yellow. Further add a few more milliliters and allow standing for 2 hours. $MgNH_4PO_4{\cdot}6H_2O$ crystalline precipitate forms. Filter through a sintered glass crucible (porosity G4) and wash thoroughly with 1 M ammoniacal solution.

Rinse the 250-mL beaker with 25 mL of hot 1 M HCl solution. Pour the solution into the glass crucible to dissolve the precipitate and wash into a 500-mL conical flask. Wash the beaker and the crucible with further 10 mL of HCl solution and 50 mL of de-ionized water. Add a measured excess of standard EDTA. Neutralize the solution with 1 M NaOH. Add 5 mL of pH 10 ammoniacal buffer and a few drops of Eriochrome Black T. Warm the solution to 60°C and back-titrate the excess EDTA with a standard $MgCl_2$ solution. The weight of phosphate in the 250-mL sample can be determined by the difference of the amount of standard EDTA and the amount of the standard $MgCl_2$ used in the back titration:

$$\begin{aligned} W_{\mathrm{Phosphate}}(\mathrm{mg}) = {} & [C_{\mathrm{EDTA}}(\mathrm{M}) \times V_{\mathrm{EDTA}}(\mathrm{mL}) - C_{\mathrm{Mg}}(\mathrm{M}) \times V_{\mathrm{Mg}}(\mathrm{mL})] \\ & \times \text{Molar mass of } PO_4^{3-} \times 10 \end{aligned} \tag{40}$$

*Remarks:* Trace amount of heavy-metal ions can be co-precipitated with $MgNH_4PO_4 \cdot 6H_2O$. As many of these metals can block the indicator, they should be masked with proper agents. For example, tartrate can be used to mask Fe, Mn, Al, and Cu. Cyanide can be added to mask Cu(II), Ag, Cd, Co, Fe(III), Fe(II), Hg(II), Ni, Pd(II), Pt(II), and Zn. Tartrate should be added when the solution is acidic. Cyanide should be added after the solution is neutralized. Safety tools such as masks and gloves should be prepared when cyanide is used.

## 10.6 Determination of Al, Pb(II), AND Zn: SEQUENTIAL TITRATION

*Procedure:* Dilute the sample into a 250-mL volumetric flask. Pipette a 25.00-mL aliquot into a 300-mL conical flask, and dilute to 100 mL with de-ionized water. Add excess amount of standard EDTA and 0.1 g of ascorbic acid. Boil the solution for 1 minute. To this solution 25 mL of 20% KCN solution was added, and immediately, 10 mL of concentrated ammonium hydroxide was added to prevent the formation of HCN and adjust the medium at pH 10. Gloves and masks must be used as a precaution when working with KCN. Add a few drops of Eriochrome Black T solution. The solution is cooled to 10°C in an ice bath for the rest of the procedure. Titrate the excess EDTA with standard $MnSO_4$ to the endpoint (from blue to red).

To determine Al, add 20 mL of triethanolamine and 5 mL of hydroxyl amine 10% (w/v) to the above solution and boil the solution for 1 minute. Add a few drops of Eriochrome Black T solution and titrate the liberated EDTA with standard $MnSO_4$ to the endpoint (from blue to red). The amount of Al in the sample can be determined by:

$$W_{\mathrm{Al}}(\mathrm{mg}) = C_{\mathrm{EDTA}}(\mathrm{M}) \times V_{\mathrm{EDTA}}(\mathrm{mL}) \times \text{Molar mass of Al} \times 10 \qquad (41)$$

After the titration of Al, 5 mL of 20% alcoholic solution of 2,3-dimercaptopropanol is added with slow stirring. Lead(II) is masked from the EDTA, and the liberated EDTA is titrated with $MnSO_4$ using Eriochrome Black T to the endpoint (from blue to red). The amount of Pb(II) can be determined by:

$$W_{\mathrm{Pb}}(\mathrm{mg}) = C_{\mathrm{EDTA}}(\mathrm{M}) \times V_{\mathrm{EDTA}}(\mathrm{mL}) \times \text{Molar mass of Pb} \times 10 \qquad (42)$$

Finally, zinc is quantitatively de-masked from its cyanide complex by the addition of (3:1) 20 g of 4% formaldehyde: acetone solution. The liberated Zn is titrated with standard EDTA at pH 10 using Eriochrome Black T to the endpoint (from red to blue). After titration, the addition of the formaldehyde:acetone mixture is repeated, followed by a second titration, until the solution no longer turned red. The total amount of EDTA used can be used to determine the amount of Zn in the sample:

$$W_{\mathrm{Zn}}(\mathrm{mg}) = C_{\mathrm{EDTA}}(\mathrm{M}) \times V_{\mathrm{EDTA}}(\mathrm{mL}) \times \text{Molar mass of Zn} \times 10 \qquad (43)$$

*Remarks:* This process is designed to sequentially determine Al, Pb(II), and Zn in glass and glass frit samples, which contain small amounts of Ca and Mg, and trace amounts of Fe, Cu, Mn, Co, Ni, Ti, and Zr. Ascorbic acid and cyanide are added to reduce Fe(III), Cu(II), and Mn(VII)/(IV) and mask Fe(II), Mn(II), Cr, Co, Ni, and Ti. EDTA is added to mask Al, Pb(II), and Zn together with Zr, Ca, and Mg. Then, the Al is de-masked with triethanolamine and the released EDTA is titrated at pH 10. Lead is de-masked with 2,3-dimercaptopropanol and the liberated EDTA is titrated. Then, zinc is de-masked by adding formaldehyde/acetone solution and titrated by standard EDTA. As all the titrations are carried out at pH with $MnSO_4$, the solution is cooled to 10°C, at which Mg and Ca will not interfere. More detailed information about this process can be found in the literature [44].

## 10.7 Determination of Fe(III), Cd, AND Mg: SEQUENTIAL TITRATION

*Procedure:* Dilute the sample into a 250-mL volumetric flask. Pipette a 25.00-mL aliquot into a 500-mL conical flask. Add excess amount of standard EDTA. Neutralize with NaOH while stirring with a magnetic bar. Monitor pH with a pH meter. Add 5 mL of pH 5 acetate buffer and add 2 drops of PAN indicator solution. Back-titrate the solution with standard $CuSO_4$ to the endpoint (from yellow to red). The amount of Fe (III) and Cd is obtained. To the same solution, add 5.00 mL of pH 10 ammoniacal buffer. Titrate with EDTA until the solution turns back to yellow. This determines the amount of Mg.

$$W_{\mathrm{Mg}}(\mathrm{mg}) = C_{\mathrm{EDTA}}(\mathrm{M}) \times V_{\mathrm{EDTA}}(\mathrm{mL}) \times \text{Molar mass of Mg} \times 10 \tag{44}$$

Then, add 10 mL of 5% KCN solution, following with 5.00 mL of concentrated ammonium hydroxide solution. Conduct this operation in a fume hood. Prepare gloves and masks in case of accidents. Cyanide will replace EDTA from its complex with cadmium. Add 2 drops of Eriochrome Black T and titrate with standard $MgNO_3$ until the solution turns from blue to red. The result determines the amount of Cd.

$$W_{\mathrm{Cd}}(\mathrm{mg}) = C_{\mathrm{Mg}}(\mathrm{M}) \times V_{\mathrm{Mg}}(\mathrm{mL}) \times \text{Molar mass of Cd} \times 10 \tag{45}$$

The amount of Fe can be calculated from the difference between the results of the back titration and replacement titration.

$$\begin{aligned} W_{\mathrm{Fe}}(\mathrm{mg}) = {} & [C_{\mathrm{EDTA}}(\mathrm{M}) \times V_{\mathrm{EDTA}}(\mathrm{mL}) - C_{\mathrm{CuSO4}}(\mathrm{M}) \\ & \times V_{\mathrm{CuSO4}}(\mathrm{mL}) - C_{\mathrm{Mg}}(\mathrm{M}) \times V_{\mathrm{Mg}}(\mathrm{mL})] \times \text{Molar mass of Fe} \times 10 \end{aligned} \tag{46}$$

*Remarks:* Many sequential EDTA titrations have been developed and are available in literature. It is noticed there are numerous possibilities of combining pH-adjustment,

masking/de-masking schemes, titration procedures, and endpoint detection methods to achieve the goal of the determination of multiple metal ions in a sample solution. If a new titration approach is to be designed for an analytical task, steps for achieving a successful procedure may be generalized as the following:

1. Inspect the stability constants of the EDTA complexes for the metal ions to be determined. List the titration procedures and conditions (pH range and endpoint detection methods) for each single metal ion to be determined. The tables included in this chapter are helpful for obtaining such information. Table 1 lists the stability constants of the EDTA complexes for common metal ions. Table 3 lists the properties and applications of some metal-ion indicators. Tables 5 and 6 list the direct and back titrations for a number of metal ions. More data and information can be found in the literature as well.
2. pH-adjustment should always be the first to be considered as an effective means of sequential titration if the stability constants of the metal-EDTA complexes are significantly different from each other. If the stability constants of the metal-EDTA complexes are close to each other, masking/de-masking approaches will be considered. Furthermore, elegant procedures for the determination of multiple metal ions with close properties can be achieved by combining different types of titration procedures (direct, back-, and indirect titrations) with masking schemes. Table 7 lists common masking agents and their applications in EDTA titrations.
3. It is noted that a sequential titration with one aliquot of sample is not always successful if multiple metal-ion indicators or masking agents are necessary. A satisfactory sequential titration normally uses one or two metal-ion indicators. If more than two indicators are necessary for the titrations, the titrations may be conducted in more than one aliquot. Bear in mind that metal-ion indicators are not always good choices for EDTA titrations. Sometimes, other detection methods can dramatically simplify the process. A good example is the use of a potentiometric titration to determine Bi(III), Cu(II), and Ca in a single portion of the sample. A mercury electrode permits a sequential titration of these three metals by adjusting the medium at pH 1, 5, and 10. Ion-selective electrodes (ISEs) can also provide good endpoint detections for many metals if the titration is conducted in the proper pH range for ISE. Table 4 lists some common ISEs with basic specifications. In some cases, when the sample is so complicated that no masking/de-masking scheme works properly, chemical separation procedures have to be considered.

## REFERENCES

1. Schwarzenbach G, Flaschka H. Complexometric titrations. Methuen: London; 1969.
2. Flaschka HA. EDTA titrations – an introduction to the theory and practice. Pergamon Press: New York; 1959.

3. Reilley CN, Bernard AJ Jr., Puschel R.In: Meites L,editor. Handbook of analytical chemistry. McGraw-Hill: New York, 1963; pp. 3–76 to 3–234.
4. Ringbom A, Wanninen E, Chapter 20, Complexation reactions. in Treatise on analytical chemistry, Part I. Theory and practice. Vol. 2. John Wiley & Sons Inc.: New York, USA; 1979.
5. Schwarzenbach G, Ackermann H. Complexones V. Ethylenediaminetetraacetic acid. Helv Chim Acta 1947; 30: 1798–1804.
6. Martell AE, Smith RM, Moteckaitis RJ,NIST critically selected stability constants of metal complexes. In: NIST Standard Reference Database 46, Gaithersburg, MD, 2001.
7. Relley CN, Schmid RW. Chelometric titrations with potentiometric end point detection. Anal Chem 1958; 30: 947–953.
8. Supplement No 1, Part 1, Compiled by Sillen LG, Martell AE. Stability constants of metal-ion complexes. The Chemical Society: Burlington House, London; 1971.
9. Flaschka H. Theory of visual indication and selectivity for complexometric titrations. Talanta 1958; 1: 60–75.
10. Schwarzenbach G, Biedermann W, Bangerter F. Complexons VI. New simple, titrimetric method for determining water Hardness. Helv Chim Acta 1946; 29: 811–818.
11. Biedermann W, Schwarzenbach G. Complexons XI. Complexometric titration of alkaline earths and some other metals with Eriochrome Black T. Chimica 1948; 2: 56–59.
12. Made by Van London – pHoenix Company, 6103 Glenmont Dr. Houston, TX 77081, USA.
13. Gupta VK, Jain AK, Maheshwari G. Aluminum(III) selective potentiometric sensor based on morin in poly(vinyl chloride) matrix. Talanta 2007; 72: 1469–1473.
14. Gupta VK, Jain AK, Khayat MAl, Bhargava SK, Raisoni JR. Electroanalytical studies on cobalt(II) selective potentiometric sensor based on bridge modified calixarene in poly (vinyl chloride). Electrochim Acta 2008; 53: 5409–5414.
15. Zamani HA, Rajabazadeh G, Ganjali MR. Highly selective and sensitive chromium(III) membrane sensors based on 4-amino-3-hydrazino-6-methyl-1,2,4-trianzin-5-one as a new neutral ionophore. Sens Actuators 2006; B 119: 41–46.
16. Gupta VK, Jain AK, Maheshwari G, Lang H, Ishtaiwi Z. Copper(II)-selective potentiometric sensors based on porphyrins in PVC matrix. Sens Actuators 2006; B 117: 99–106.
17. Ganjali MR, Dodangeh M, Ghorbani H, Norouzi P, Adib M. Dy(III) sensor based on a new asymmetrical Schiff's base. Anal Lett 2006; 39: 295–506.
18. Ganjali MR, Norouzi P, Alizadah T, Tajarodi A, Hanifahpour Y. Fabrication of a highly selective and sensitive Gd(III)-PVC membrane sensor based on *N*-(2-pyridyl)-*N*-(4-nitrophenyl) thiourea. Sens Actuators 2007; B 120: 487–493.
19. Zamani HA, Ganjali MR, Norouzi P, Adib M, Aceedy M. Synthesis of N′-(1-pyridin-2-ylmethylene)-2-furohydrazide and its application in construction of a highly selective PVC-based membrane sensor for La(III) ions. Anal Sci 2006; 22: 943–948.
20. Ganjali MR, Tamaddon A, Norouzi P, Adib M. Novel Lu(III) membrane sensor based on a new asymmetrically S-N Schiff's base. Sens Actuators 2006; B 120: 194–199.
21. Kumar KG, Poduval R, Augustine P, John S, Saraswathyamma B. A PVC plasticized sensor for Ni(II) ion based on a simple ethylenediamine derivative. Anal Sci 2006; 22: 1333–1337.
22. Ganjali MR, Norouzi P, Mirnaghi FS, Riahi S, Faridbod F. Lanthanide recognition: monitoring of praseodymium(III) by a novel praseodymium(III) microsensor based on N′-(pyridin-2-ylmethylene) benzohydrazide. IEEE Sens J 2007; 7: 1138–1144.

23. Mittal SK, Sharma HK, Kumar ASK. Samarium(III) selective membrane sensor based on tin(IV) boratophosphate. Sensors 2004; 4: 125–135.
24. Jain AK, Gupta VK, Jitendra RR. Strontium(II)-selective protentiometric sensor based on ester derivative of 4-tert-butylcalix(8) arene in PVC matrix. Sensors 2004; 4: 115–124.
25. Zamani HA, Rajabzadeh G, Ganjali MR. Construction of a novel potentiometric terbium (III) membrane sensor and its application for the determination of terbium ion in binary mixture and fluoride ion in a mouth wash preparation. J Braz Chem Soc 2006; 17: 1297–1303.
26. Ganjali MR, Norouzi P, Faridbod F, Riahi S, Yaftian MR, Zamani A, Matt D. Highly selective and sensitive $Th^{4+}$-PVC-based membrane sensor based on 2-(diphenylphosphorothioyl)-N′,N′-diphenylacetamide. J Appl Electrochem 2007; 37: 827–833.
27. Ganjali MR, Norouzi P, Tamaddon A, Husain SW. Novel Tm(III) membrane sensor based on 2, 2′-dianiline disulfide and its application for the fluoride monitoring of mouth wash preparations. Bull Korean Chem Soc 2006; 27: 1418–1422.
28. Quintar SE, Santagata JP, Cortinez VA. Determination of vanadium(V) by direct automatic potentiometric titration with EDTA using a chemically modified electrode as a potentiometric sensor. Talanta 2005; 67: 843–847.
29. Zamani HA, Rajabzadeh G, Ganjali MR. A new ytterbium(III) PVC membrane electrode based on 6-methy-4-{[1-(1H-pyrrol-2-yl)methylidene]amino}-3-thioxo-3,4-dihydro-1,2,4-triazin-5(2H)-one. Talanta 2007; 72: 1093–1099.
30. Dean JA. Analytical chemistry handbook. McGraw-Hill: New York, USA; 1995. P 3–104–3–109.
31. Jeffery GH, Bassett J, Mendham J, Denney RC. Vogel's textbook of quantitative chemical analysis. Longman Scientific & Technical, John Wiley & Sons: New York, 2000. p. 384.
32. Flaschka HA, Barnard AJ Jr, Titrations with EDTA and related compounds. In: Wilson CL, Wilson DW,editors. Comphrehensive analytical chemistry. Elsevier Publishing Co: New York; 1960. P. 350.
33. Flaschka H. Microtitrations with ethylenediaminetetraacetic acid, I. Mikrochemie, ver Mikrochim Acta 1952; 39: 38–50.
34. Cheng KL. Complexometric titration of indium. Anal Chem 1955; 27: 1582–1583.
35. Flaschka H, Abdine HZ. EDTA titrations using copper-PAN complex as indicator. Chem Anal 1956; 45: 58–61.
36. Flaschka H, Abdine HZ. Complexometric titration of aluminum and iron and the sum of both. Anal Chem 1956; 152: 77–85.
37. Flaschka H, Abdine HZ. EDTA titrations using PAN complex as indicator. Chem Anal 1956; 45: 2–3.
38. Biedermann W, Schwarzenbach G. Complexones XI. Complexometric titration of alkaline earths and some other metals with Eriochrome Black T. Chimica (Switz.) 1948; 2: 56–59.
39. Flaschka H. Direct volumetric determination of divalent manganese with EDTA in presence of other metals. Chem Anal 1953; 42: 56–58.
40. Flaschka H, Amin AMZ. The titrimetric determination of indium with ethylenediaminetetraacetic acid. Anal Chem 1953; 140: 6–9.
41. Patzak P, Doppler G. Determination of chromium, iron and aluminum with ethylenediaminetetraacetate when all three cations are present simultaneously. Z Anal Chem 1957; 156: 248–257.

42. Flaschka F, Puschel R. Complexometric titration of nickel in the presence of cobalt and of some other metals. Z Anal Chem 1955; 147: 354–360.
43. Lott P, Cheng KL. Improved end point by addition of polyvinyl alcohol in the EDTA titration of calcium with Calcon as indicator. Chem Anal 1959; 48: 13.
44. Kayal N, Singh N. Selective masking and demasking for the stepwise complexometric determination of aluminium, lead and zinc from the same solution. Chem Cent J 2008; 2: 4,

# Electronic Circular Dichroism Spectroscopy

**Ferenc Zsila M.D., Ph.D.**
*Chemical Research Center, Institute of Biomolecular Chemistry, Budapest, Hungary*

## 1 INTRODUCTION

### 1.1 Definitions and Abbreviations Used in the Chapter

***1.1.1 Definitions*** *Anisotropy factor (termed also as Kuhn's dissymmetry ratio or g factor)*: It is a dimensionless parameter expressing a ratio of the strength of CD to strength of light absorption of the chiral sample at specific wavelengths. $g = \Delta A/A = \Delta\varepsilon/\varepsilon$, where $\Delta A$ is the difference of the absorption of the sample against left- and right-circularly polarized light beams ($A_L - A_R$). It is a useful parameter of how difficult it is to measure the CD in the region of a given absorption band, because $\Delta A/A$ is proportional to the signal/noise ratio where $\Delta A$ determines the signal and $A$ determines the noise. The anisotropy factor is rarely larger than $10^{-2}$ and more commonly is around $10^{-4}$. $g$ factor can be used as analytical tool to characterize the diastereomeric and enantiomeric purities of optically active materials.

*Chiral exciton coupling*: It is based on the through space exciton interaction between two or more chirally oriented chromophores which are not conjugated to each other and show strong ($\varepsilon \geq 10^4$) $\pi \rightarrow \pi^*$ type absorption. Coupling of the $\pi \rightarrow \pi^*$ transition moments in a symmetric and anti-symmetric fashion gives rise to a

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing,*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

CD band pair of mutually opposite signs in the region of the absorption peak (bisignate CD spectrum). The exciton coupling may occur intra- or intermolecularly, it may be degenerate (between identical chromophores), or non-degenerate (between different chromophores).

*Chiroptical ("chiral optical") spectroscopy*: It covers various spectroscopy techniques which measure differences in the interaction of chiral molecules with left- and right-circularly polarized light. These techniques include: electronic circular dichroism (CD or ECD), synchrotron radiation circular dichroism, vibrational circular dichroism, fluorescence-detected circular dichroism, circularly polarized luminescence, and vibrational Raman optical activity.

*Chromophore*: A molecule (e.g., benzoic acid) or a molecular subgroup (e.g., the polyene chain of amphotericine B) that gives rise to absorption of electromagnetic radiation of specified wavelengths. The term is equally used in the context of electronic transitions in the ultraviolet ($\lambda < 350$ nm) and in the visible spectral region.

*Circular dichroism*: CD is the difference in absorption of left- and right-circularly polarized light beams passed through an optically active (chiral) medium: $\Delta A = A_L - A_R$. In molar units: $\Delta\varepsilon = \varepsilon_L - \varepsilon_R$, where $\Delta\varepsilon$ is the molar circular dichroic absorption coefficient in $M^{-1}\,cm^{-1}$.

*Circularly polarized light (cpl)*: In cpl the electric field vector remains constant in magnitude, but traces out a helix as a function of time. It rotates in a clockwise (right) or in an anti-clockwise (left) sense (rcpl, lcpl). The vector summation of in-phase rcpl and lcpl results in plane polarized light.

*Cotton effect (CE)*: Positive or negative bands in the CD spectrum associated with light absorption bands of chromophoric groups of the sample compound. Stereochemical information of the analyte are encoded in the number, sign, magnitude, width, and spectral position of CEs.

*Ellipticity ($\Theta$)*: Due to the interaction with a chiral compound, the electric field vector of the light traces out an elliptical path after passing through the sample. The angle of the ellipsoidally polarized transmitted light is defined as the ellipticity angle ($\Theta$). The ellipticity of a chiral molecule is directly proportional to $\Delta A$. Due to historical reasons, most CD spectropolarimeters produce CD spectra in units of in millidegrees (mdeg) versus wavelength (nm) although they directly measure differential absorbance values ($\Delta A$). The relationship between $\Theta$ and $\Delta A$ is $\Delta A = 4\pi\Theta$ (degrees)/180ln10 $= \Theta$(mdeg)/32982.

*Enantiomeric purity (e.p.) or enantiomeric excess (e.e.)*: It means the excess (in per cent) of one enantiomer over the other: if $\%R > \%S$ then e.e.$\% = \%R - \%S$ and e.e.$\% = \%S - \%R$ if $\%R < \%S$. For instance, a 75%:25% mixture of *R*- and *S*-enantiomers has a $75\% - 25\% = 50\%$ e.e. Enantiomeric purity relates directly to the CD signal because if there is no enantiomeric excess (50%*R*:50%*S*) then there is no CD signal.

*Extrinsic circular dichroism: see Induced circular dichroism*

*g Factor: see Anisotropy factor*

*Induced (extrinsic) circular dichroism (ICD)*: Achiral molecules can acquire CD activity through binding to asymmetric hosts such as proteins, nucleic acids, cyclodextrines. The ICD bands are shown in the light absorption region of the guest

and their sign and magnitude are crucially dependent on the binding geometry as well as on the chromophore interaction energy with the host matrix.

*Intrinsic circular dichroism*: CD activity caused by genuine disymmetry of chiral entities. Most frequently, it is associated with the presence of asymmetric centers in the molecule.

*Mean residue (or residual) ellipticity ($[\Theta]_{MRW}$)*: For far-UV CD spectra (190–240 nm) of proteins and peptides, CD data are routinely normalized by scaling to the repeating unit of the polymer. Because essentially every residue in a protein is associated with a peptide bond, spectra in this region can be directly compared by using the peptide bond as the repeating unit. CD data are expressed in terms of the mean residue weight (MRW), which is defined by $\mathrm{MRW} = M/(N-1)$, where $M$ is the molecular mass of the polypeptide chain in daltons and $N$ is the number of amino acids, so the value of $N-1$ gives the number of peptide bonds. The mean residue ellipticity is given by the equation of

$$[\Theta]_{\mathrm{MRW}} = \mathrm{MRW} \times \Theta/10 \times c \times l$$

where $\Theta$ is the observed ellipticity (in mdeg), $c$ is the concentration (mg/mL), and $l$ is the optical path length (cm). The dimensions of $[\Theta]_{\mathrm{MRW}}$ is deg cm$^2$/dmol, and the quantity is independent of molecular weight.

*Molar circular dichroic absorption coefficient ($\Delta\varepsilon$ in $M^{-1}\,cm^{-1}$)*: It is the difference of the molar absorption coefficients ($\varepsilon_{\mathrm{L}}$ and $\varepsilon_{\mathrm{R}}$) of the chiral solute: $\Delta A = A_{\mathrm{L}} - A_{\mathrm{R}} = \varepsilon_{\mathrm{L}} \times c \times l - \varepsilon_{\mathrm{R}} \times c \times l = \Delta\varepsilon \times c \times l \Leftrightarrow \Delta\varepsilon = \varepsilon_{\mathrm{L}} - \varepsilon_{\mathrm{R}}$, where $c$ is the molar concentration (M) and $l$ is the optical path length (cm).

*Molar ellipticity ($[\Theta]$)*: While $\Delta\varepsilon$ (or $\Delta A$) is the preferred way to express CD values, authors frequently report data as molar ellipticity $[\Theta]$ in deg cm$^2$/dmol: $[\Theta] = \Theta/10 \times c \times l$ where $\Theta$ is the observed ellipticity (in mdeg), $c$ is the molar concentration (M), and $l$ is the optical path length (cm). Molar ellipticity and molar circular dichroic absorption are related by the expression $[\Theta] = 3298 \times \Delta\varepsilon$.

*Optical activity*: The ability of chiral chemical compounds to rotate the plane of polarization of plane polarized light, an effect that results from the difference in refractive indices lcpl and rcpl components of the plane polarized light.

*Plane polarized light*: A transverse light wave of which electric and magnetic field vectors oscillate in single planes perpendicular to each other. Its polarization is defined by the direction of the electric field. Plane polarized light can be decomposed into two, left- and right-circularly polarized components of opposite helicity but identical frequency, speed, and intensity.

### 1.1.2 Abbreviations

| | |
|---|---|
| (+) | positive or clockwise rotation of plane polarized light |
| (−) | negative of counterclockwise rotation of plane polarized light |
| (±) | no rotation of plane polarized light (racemic sample) |
| AAG | human serum $\alpha_1$-acid glycoprotein |
| BSA | bovine serum albumin |
| CD | circular dichroism |

| | |
|---|---|
| CE | Cotton effect |
| cpl | circularly polarized light |
| CyD | cyclodextrin |
| e.e. | enantiomeric excess |
| e.p. | enantiomeric purity |
| HPLC | high-performance liquid chromatography |
| HSA | human serum albumin |
| ICD | induced (extrinsic) circular dichroism |
| lcpl | left-circularly polarized light |
| LOD | limit of detection |
| LOQ | limit of quantitation |
| M | molar concentration (mol/L) |
| mdeg | millidegree |
| MRW | mean residue (or residual) ellipticity |
| NMR | nuclear magnetic resonance |
| NSAID | non-steroidal anti-inflammatory drug |
| ORD | optical rotatory dispersion |
| rcpl | right-circularly polarized light |
| RI | refractive index |
| UV/VIS | ultraviolet/visible |

## 1.2 Basic Features of Electronic Circular Dichroism Spectroscopy

ECD or CD is the manifestation of the important property of chiral compounds to absorb lcpl and rcpl to different extents. CD spectroscopy is closely related to the conventional UV/VIS absorption spectrophotometry, a well-known and routinely applied technique in the broad fields of pharmaceutical sciences [1–3]. Mutual comparison of principal features of the two spectroscopic techniques can be useful for understanding the basic nature of the CD method (Table 1).

The CD instrument generates lcpl and rcpl which are passed alternately through the sample. There is no absorption difference for achiral molecules against lcpl and rcpl ($A_L = A_R$, $\Delta A = 0$). Conversely, unpolarized light used in UV/VIS spectrophotometers is unsuitable for discrimination of enantiomers due to their identical absorption. However, lcpl and rcpl do interact differently with optically active materials resulting in a small but measurable difference in the light absorption of the sample. The absorption difference is expressed conventionally as $\Delta A = A_L - A_R$ or $\Delta\varepsilon = \varepsilon_L - \varepsilon_R$ and is plotted in the function of the wavelength. Accordingly, CD bands can be positive ($A_L > A_R$) or negative ($A_L < A_R$). Importantly, CD bands (called also Cotton effects, CE) can only be measured where absorption bands are present. If chiral molecules do not show optical absorbance (e.g., carbohydrates, ethambutol above 200 nm), their CD signal will be zero in spite of the presence of asymmetric centers.

Sign, shape, number, intensity, and wavelength position of CEs in the CD spectrum and their changes under different experimental conditions carry a wealth of valuable structural information on the compound under study.

**TABLE 1 Comparison of Principal Features of UV/VIS Absorption and CD Spectroscopy**

| | UV/VIS Absorption Spectroscopy | CD Spectroscopy |
|---|---|---|
| Type of compounds can be studied | Chiral and achiral | Only chiral |
| Type of the detector used | Photomultiplier | Photomultiplier |
| Polarization of the electromagnetic radiation | Unpolarized | Circularly polarized |
| Wavelength range of the electromagnetic radiation | 180–800 nm | 180–800 nm |
| Type of transitions detected | Electronic excitations ($\pi \rightarrow \pi^*$, $n \rightarrow \pi^*$) | Chirally perturbed electronic excitations ($\pi \rightarrow \pi^*$, $n \rightarrow \pi^*$) |
| Relative strength of the signals measured | 1 $A$ unit | $10^{-3} - 10^{-5}$ $\Delta A$ unit |
| Mode of data acquisition | $A$ values in the function of wavelength | $\Delta A$ (and $A$) values in the function of wavelength |
| Sign of the bands in the spectrum | Always positive ($A > 0$) | Positive or negative ($0 > \Delta A$ or $\Delta A > 0$) |
| Spectra of pure enantiomers | Identical curves | Mirror-image curves |
| Validity of the Beer's law | Valid | Valid |
| Molar units used | $\varepsilon$ ($M^{-1}\,cm^{-1}$) | $\Delta\varepsilon$ ($M^{-1}\,cm^{-1}$) |

Discussion of the theoretical basis of CD spectroscopy is beyond the scope of this chapter. For in-depth treatments, readers are referred to several excellent books [1,4–15] and review papers [16–28].

## 1.3 CD Instrumentation

CD instruments are available from several manufacturers including Jasco Inc. (http://www.jascoinc.com), Aviv Biomedical Inc. (http://www.avivbiomedical.com), and Applied Photophysics (http://www.photophysics.com).

CD bands are most commonly measured in the UV and VIS regions of the spectrum from 170–180 to 700–800 nm.

The CD signal ($\Delta A$) is of the order of $\pm 10^{-5}$ to $\pm 10^{-3}$ for usual samples corresponding to an ellipticity ($\Theta$) of $\approx \pm 0.3$ to $\pm 30$ mdeg, so that highly sensitive instrumentation is required for its accurate measurement. Due to historical reasons, most CD spectropolarimeters display CD spectra in units of ellipticity ($\Theta$), in millidegrees against wavelength (nm) although they directly measure $\Delta A$ values. The relationship between $\Theta$ and $\Delta A$ is $\Delta A = \Theta(\text{millidegrees})/32982$. Chiroptical data for small organic/inorganic molecules are usually expressed in $\Delta\varepsilon$, units while CD data of biopolymers are often given in ellipticity values ($\Theta$ or $[\Theta]$).

Most commercial CD equipments are single beam and use the modulation technique [15,17,29,30]. First, unpolarized monochromatic light is linearly polarized and then passes through a photoelastic modulator (PEM), a piezoelectric device which produces lcpl and rcpl alternately, with a switching frequency of typically 50 kHz. For optically inactive samples absorption of the two beams is equal and the output current

generated in the detector is constant. In chiral samples, the absorption of the lcpl and rcpl will be different so the light intensity seen by the photomultiplier will also be modulated at the same frequency. The larger direct current component of the detector is characteristic to the total light absorption of the sample, whereas the small alternating current component relates to the CD of the sample and is directly proportional to $\Delta A$. Thus, beside CD spectra CD instruments record simultaneously UV/VIS absorption curves as well.

Single beam CD spectropolarimeters should be calibrated regularly, at least monthly if the machine is in regular use, using a standard sample with known absorbance and CD value [11,17]. The most frequently used standard for CD calibration is the D-( + )-10-camphorsulfonic acid.

A new double beam instrument from OLIS Inc. (On-Line Instrument Systems, Inc., http://www.olisweb.com) has recently been introduced on the market, which uses direct substraction of lcpl and rcpl for obtaining the CD signal [15,30]. The instrument utilizes two modulated cpl beams out of phase by 180° which pass separately through different parts of one cuvette and detected simultaneously by two detectors. So, absorption of lcpl and rcpl is measured directly. Since each beam is a reference for the other, this technique has some advantages. For instance, energy fluctuations of the light source are eliminated since they affect both beams equally. In addition, this system needs no calibration and results in zero drift over time and nearly flat baselines. However, the two beams pass through different parts of the sample. So, it is particularly important to prepare homogenous and particle-free sample solution, to use high-quality quartz cuvettes with no intrinsic strain, and to clean cuvettes regularly to avoid buildup of hard-to-remove organic and inorganic deposits.

The high-energy light sources used in CD machines (150–450 W xenon arc lamp, water, or air cooled) generate ozone during operation. Thus, a supply of high-purity, dry nitrogen gas is essential to purge oxygen in order to avoid degradation of the optics by ozone, as well as to decrease background optical absorbance of $O_2$ in the sample chamber for performing measurements below 200 nm [17].

A Peltier system for temperature control and thermal ramping is an invaluable accessory, particularly for studies of the thermal unfolding of proteins/nucleic acids, and for thermodynamic and kinetic investigations of ligand–biopolymer interactions.

## 1.4 Conditions for Employing CD Spectroscopy

***1.4.1 Light Absorption and Stereochemical Requirements of CD Spectroscopic Investigations*** Sample must contain one or more chromophore showing light in the wavelength range where CD equipments operate.

Chromophores must be associated with a chiral environment. Here, "association" means

- connection of the chromophore to chiral center(s) via one or more covalent bonds;

OR

- non-covalent association of the chromophoric moiety to an asymmetric (chiral) host (e.g., proteins, DNA/RNA, cyclodextrins, etc.).

In spite of the lack of any stereocenter, a small group of chromophores are themselves chiral (inherent chirality). Their mirror images are not superimposable with each other, since their structures devoid of either a center or a plane of symmetry. Such type of chromophores (hexahelicene, twisted biphenyls) exhibit pronounced CD activity [16].

It is important the total absorbance (cell and sample solution combined) to be kept in the analytical range between 0.2 and 1.2 AU [11,30]. The maximum absorbance of the sample tolerated in CD experiments can be about 2 AU. An absorbance of 0.4–1.0 AU gives the optimal signal-to-noise (S/N) ratio.

***1.4.2 Chemical Nature of Materials Suitable for CD Spectroscopic Studies*** Chiral organic/inorganic compounds, organometallics [9,12,13]; peptides, polypeptides, proteins [9,10]; natural and synthetic oligo- and polynucleotides [9,10]; mono-, oligo-, and polysaccharides [10]; synthetic chiral polymers [31,32].

***1.4.3 Physical Forms of the Sample Material Suitable for CD Measurements*** Dissolved in organic or in aqueous solvents (most frequent); liquids; vapor phase [33]; crystalline/semicrystalline phases (solid state CD) [34]; liquid crystals [20]; thin films [35].

***1.4.4 Solvents for CD Spectroscopic Measurements*** Organic and aqueous solvents used in CD spectroscopic investigations must be optically inactive and have the lowest possible absorbance in the spectral region scanned. Organic solvents should be of spectroscopic purity and should be checked for the absence of absorbing impurities.

Distilled water is particularly transparent solvent but buffer additives (phosphate, chloride, nitrate, borate, MOPS, EDTA, SDS, etc.) can significantly increase light absorption below 230 nm [11,36,37].

Since oxygen absorbs light below 200 nm, solvents should be degassed before measuring in that spectral region.

Due to the single-beam construction of most CD instruments, sample CD spectra must always be corrected by subtracting a baseline spectrum of the cell plus blank solvent obtained under identical acquisition conditions.

## 1.5 Advantageous and Disadvantageous Features of Circular Dichroism Spectroscopy

Advantages

- CD spectroscopy provides instant answer for structural problems qualitative in hand: discrimination between enantiomers and diastereomers (Fig. 1) by simple comparison of their CD spectra, detection of chirality of biopolymers (e.g. B or Z conformations of DNA), demonstration of drug–biopolymer (e.g., drug–protein, drug–nucleic acid, drug–cyclodextrin) and biopolymer–biopolymer (e.g., protein–protein, protein–DNA) binding interactions [38].
- Molecules can be studied under simulative physiological conditions, i.e., pH 7.4 aqueous buffer systems, 37°C.

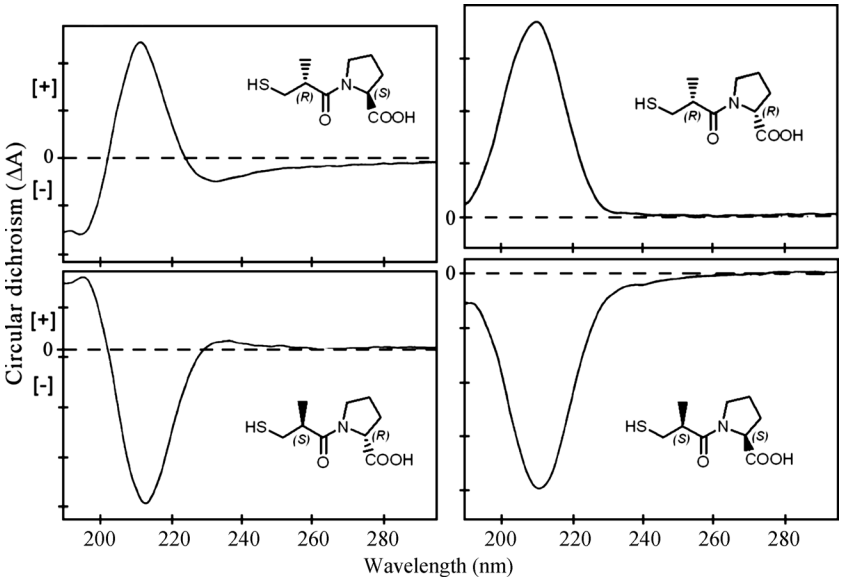


***FIGURE 1*** *Comparison of CD spectra of the angiotensin converting enzyme inhibitor captopril (1-[(2S) 3-mercapto-2-methylpropionyl]-(S)-proline) and its stereoisomers measured in deionized water. Note the mirror-image CD curves of the enantiomers and the spectral differences caused by inverted chirality of the side-chain or the ring stereocenter. Adapted and modified from Ref. [28].*

- CD spectra can be taken under a wide range of experimental conditions (concentration, pH, temperature, ionic strength, detergents, etc.) [39,40].
- CD spectroscopy is a non-destructive technique. If necessary, sample materials can be restored after measurements.
- CD instruments record two kinds of experimental data simultaneously; the CD and the UV/VIS absorption spectrum of the sample. The absorption curve is essential for proper interpretation of the CD data and it often provides further information on the compound under study [21,41,42]. In addition, the ratio of the dichroic/absorbance signal ($\Delta A/A$, $g$ factor) is a valuable analytical tool for determination of enantiomeric composition of chiral pharmaceuticals (see Sections 2.1.2 and 2.1.3).
- CD spectra of multi-component samples can be recorded and studied without previous analytical separation [14,43–45]. Since optically inactive compounds give no CD contribution, they do not disturb CD signals of the chiral component. Chiral but optically transparent molecules (e.g., mono-, oligo-, and polysaccharides) do not interfere by CD bands of chiral substances above 210 nm. CD investigation of mixtures of chiral substances showing separate CD bands above 200 nm is also possible.
- In contrast to other physical methods such as NMR and mass-spectrometry, CD spectroscopic characterization of glycoproteins and membrane proteins can be

performed without rigorous pre-treatment of the samples since neither carbohydrate moieties nor residual lipids disturb CD signals of the protein component.

- Recording of CD spectra is quick and can be completed within 15–30 min [37]. In practice, a single protein far-UV CD spectrum over the 190–240 nm range can take as little as ≈3 min. In addition, only small quantities of protein are required for measuring far-UV CD spectra (200 μl of a 1 to 10 μM solution in a rectangular cell with 0.1 cm pathlength). Titration experiments (e.g., ligand–protein, DNA–protein binding), temperature-dependent measurements need more time, usually 1–2 h [39,46,47]. CD and UV absorbance detection at a fixed wavelength have been used in a high throughput HPLC flow through assay to distinguish between enantiomers of 1-phenylethanol with a throughput of 700–900 assays per day [48].
- Since the electronic transitions between the ground and excited states occurs on a very short time scale ($\sim 10^{-15}$ s), a wide range of biologically important processes can be time resolved by CD spectroscopy.
- Using stopped-flow CD equipment makes possible to study structural changes in proteins which occur in a few tens of milliseconds timescale [49,50].

Limitations and pitfalls

- Samples must be optically active. Achiral compounds cannot be studied by CD spectroscopy.
- Chiral centers and the chromophores can mutually be located in a manner (separated from each other by unhindered single bonds, e.g., thyroxine, dihydrofolate) that produces only a very weak or no measurable CD activity [43,51,52].
- Conventional CD spectroscopic investigation of chiral molecules showing light absorption only below 200 nm is difficult (e.g., mono- and disaccharides). At wavelengths less than 200 nm such compounds often have incompletely developed CD spectra, not reaching a maximum signal. In addition, most organic solvents and buffers show limited optical transparency below 200 nm even by using shorther path length cuvettes (1 mm and below).
- Compared to NMR and X-ray crystallography methods, CD spectroscopy provides only low-resolution structural information both for small molecules and biopolymers [17,53,54]. For instance, secondary structure of proteins (λ-helix, β-sheet content) can be estimated from CD data but CD spectra are not able to provide detailed information such as which regions of the protein have what structure, what are the shape, and amino acid composition of ligand binding pockets and clefts.
- Relatively large amounts of proteins (typically ≈1–2 mg) are required for measuring near-UV CD spectra (240–320 nm) and for conducting ligand–protein CD titration studies [17,40].
- Limits of direct CD measurements made on bulk samples using direct transmission detection are similar to those for absorbance, approximately 1 μM for a 1 cm path length.

- Compounds having absorption bands below cut-off point of the solvent applied cannot be studied by CD spectroscopy due to the low intensity of the light reaching the detector. This problem arises when the sample compound is insoluble in other, optically transparent solvents.
- Materials with strong fluorescence or phosphorescence might cause artifacts in the CD spectrum due to the significant number of emitted photons reaching the detector [55].
- Samples containing colloidal particles (liposomes, molecular aggregates, membrane fragments) might cause differential light absorption and light-scattering effects resulting in optical artifacts in CD and absorption spectra [5,56,57]. These problems can be offset to some extent by adjustment of cuvette path length, but they remain a significant limitation.
- Membrane proteins should be solubilized for CD measurements to eliminate optical artifacts. However, different detergents may lead to the alteration of the native protein structure which will be seen in the CD spectra [58–60].

## 1.6 Main Application Fields of CD Spectroscopy

### *1.6.1 Stereochemical Characterization of Organic and Inorganic Compounds*

- Determination of the absolute configuration of molecules by using empirical and non-empirical methods. Conformational analysis of chiral molecules with known absolute configuration [7,9,12–14,16,18,19,26,28,61,62].

There are numerous reports on CD spectroscopic investigation of stereochemistry of chiral pharmaceutical agents. Since the main purpose of this chapter to summarize the practical aspects of CD spectroscopy most relevant for the pharmaceutical industry, its application in configurational, and conformational analysis of organic/inorganic compounds is not covered here. Readers who are interested in specific details are referred to Table 2 where selected examples of therapeutic drugs studied by CD spectroscopy are shown [25,41,51,63–95].

### *1.6.2 Structural Characterization of Biopolymers*

- Estimation of secondary (proteins/peptides) [10,54,96–98] and characterization of tertiary structure of proteins [49,96] under various experimental conditions.
- Determination of thermodynamics and kinetics of folding/unfolding of proteins and peptides.
- Characterization of protein/peptide aggregations [10,30].
- Determination of conformation and conformational transitions of nucleic acids [10,30,99].
- Evaluation of conformation of carbohydrates [10].

**TABLE 2 Examples for CD Spectroscopic Studies of Stereochemistry of Various Pharmaceutical Agents**

| Drug | Ref. |
|---|---|
| Anthracycline antibiotics | [63] |
| Antibiotics | [64–68] |
| Barbiturates | [69] |
| Benzodiazepines | [70,71] |
| Bradykinin | [72] |
| $Ca^{2+}$ antagonists | [73] |
| Captopril | [25] |
| Cinchona alkaloids | [74] |
| Codeine derivatives | [75] |
| Colchicine | [76,77] |
| Ketamine | [41] |
| Methotrexate | [78] |
| NSAIDs | [51,79,80] |
| Nicotine | [81] |
| Ouabain | [82] |
| Prostacyclins | [83] |
| Riboflavin | [84] |
| Steroid hormones | [85,86] |
| Strychnine alkaloids | [87] |
| Taxol | [88] |
| Tolperisone | [89] |
| Vasopressin | [90] |
| *Vinca* alkaloids | [91–93] |
| Vitamin $B_{12}$ | [94] |
| Warfarin | [95] |

### 1.6.3 Investigation of Biopolymer–Biopolymer Interactions

- Detection and characterization of protein–protein [10,100], protein–peptide, protein/peptide–nucleic acid [10,47], and protein/peptide–carbohydrate [101] binding interactions.
- Estimation of the binding constant, stoichiometry, thermodynamics, and kinetics of binding [21,27,40].

### 1.6.4 Investigation of Ligand–Biopolymer Interactions

- Detection and characterization of ligand–protein binding interactions (e.g., drug–receptor [102], substrate/inhibitor–enzyme [103,104], co-enzyme–enzyme [27,105], drug–plasma protein [21,106]).
- Determination of binding constant/stoichiometry [107,108], mechanism, thermodynamics, and kinetics of binding.
- Characterization of the binding environment. Detection of protein conformational changes upon ligand binding [9,30,109].

- Detection and characterization of ligand–nucleic acid binding interactions [27,109]. Determination of binding mode (intercalation, groove binding), estimation of binding constant/stoichiometry [110], thermodynamics, and kinetics of binding.
- Detection and characterization of ligand–polysaccharide binding interactions [27,111].

#### *1.6.5 Analytical Applications*

- Determination of concentration and enantiomeric purity of chiral drugs in pharmaceutical dosage forms by direct CD spectroscopic measurements or by a CD detector coupled with HPLC system [14,44,112].

## 2 PRACTICAL APPLICATIONS OF CD SPECTROSCOPY IN PHARMACEUTICAL SCIENCES

### 2.1 Analytical Applications

#### *2.1.1 Drug Determination by CD Spectroscopic Method in Pharmaceutical Preparations and in Biological Samples*

##### *2.1.1.1 Direct CD Spectroscopic Determination of Chiral Drugs in Pharmaceutical Formulations*

*Summary* Due to the high-selectivity of CD spectroscopy for molecular chirality, this technique can efficiently be used for the identification and direct quantitative determination of chiral drugs in pharmaceutical preparations, even in the simultaneous presence of several kinds of additives. Such assays can be performed without prolonged workup including pre-concentration, derivatization, or various separation steps.

*Selected Examples* A quick and simple CD method has been elaborated for chiroptical determination of 15 cephalosporins in commercial oral suspensions, injections, and capsules [65]. Based on the discriminative CD spectral features of pure forms, the antibiotics were classified into five groups. Using CD spectroscopic data measured at different concentrations, calibration graphs have been constructed for each drug for the purpose of quantitative analysis. The limit of quantitation (LOQ) was found to be 5 μg/mL. Subsequent measurement of CD spectra of commercial samples dissolved in distilled water without performing any derivatization, filtration, or separation steps allowed correct determination of antibiotic levels of these preparations.

Penicillin V and cephalothin contents of laboratory mixtures containing lactose, starch, and caffeine have successfully been determined by exploiting distinctive features of CD spectra of that antibiotics [113].

Based on the large difference between CD spectra of tetracycline and its epimer 4-epitetracycline, a simple and sensitive CD assay has been presented for quantitative determination of the epimers in commercial tetracycline capsules [66].

Feasibility of the CD detection method has been demonstrated by estimation of water-soluble vitamin contents of pharmaceutical and food preparations [43]. Quantified amounts of commercial vitamin tablets and juice products were directly dissolved in or diluted with buffer + EDTA solution and then CD spectra were taken and compared to the reference curves of pure ascorbic acid and riboflavin.

Quick and easily reproducible CD method for quantitative determination of heroin in illicit materials has been reported [114]. Among the procedures tested, best results were obtained for direct CD determination of heroin at 253 nm as a morphine formed in the sample after basic hydrolysis. This step significantly enhanced the limit of detection (LOD) since the positive UV CD maximum of morphine is seven times higher than that of heroin.

CD spectroscopy might be useful in direct determination of two or more chiral drugs being present simultaneously in the same preparation. A simple protocol, which requires no separation or sample pre-treatment steps has been developed for direct quantification of cocaine and morphine [115]. Though both drugs are chiral and display partly overlapping CEs, their simultaneous CD investigation is still possible by measuring and analyzing CD data pairs of the mixture and of either components at pre-selected wavelength pairs.

Subtle differences in CD spectra of structurally related compounds can be used for diagnostic purposes, as it has been demonstrated by identification and quantitative determination of various opiates in binary and ternary mixtures including morphine, codeine, thebaine, and noscapine [116].

*Comments*

- Depending on the stereochemistry, CD signal of some chiral drugs may be so weak that makes possible only qualitative analysis.
- Achiral molecules showing UV/VIS absorption bands interfere to the extent that their absorption lowers the S/N ratio and limits the level of detection. Natural dyes and coloring agents added to pharmaceuticals are among the worst interferences.

#### *2.1.1.2 Direct CD Spectroscopic Determination of Chiral Drugs in Biological Fluids*

*Summary and Comments* Direct determination of CD active chiral drugs in serum or plasma is possible only if the compounds show CD bands above 300–330 nm, where intrinsic CD contribution of major plasma proteins (albumin, immunoglobulins) is negligible. For chiroptical study of drugs in serum showing CEs only between 240 and 300 nm, protein removal is needed for the elimination of strong masking effect of near-UV protein CD bands.

*Selected Examples* Protein precipitation had to be performed prior to the quantitative CD determination of β-lactam antibiotics in serum samples [117]. Positive and negative CD bands of ceftriaxone, cephalothin, and cephalexin were

measured between 240 and 290 nm in protein-free serum at different drug concentrations. Validity and accuracy of the method was confirmed by using the model drug cefoxitin, cefuroxime, and cefotaxime.

Direct estimation of optically active antibiotics such as ampicillin, cefoxitin, and cephalexin has been performed by CD method in voided urine of patients, excreting different amounts of proteins, under multiple drug administration [118]. To simulate protein CD contribution of the samples obtained from subjects showing proteinuria, human serum albumin (HSA) was used as the analytical standard for the correction of CD spectra. The validation procedure demonstrated that CD values followed the Beer's law and were linear relation to drug concentration in the range of $\approx 10^{-6} - 10^{-4}$ M.

Direct analysis of tetracycline in human urine has also been described using CD spectroscopy [119]. By measuring amplitude of the strong positive CE at 295 nm, the LOD was found to be $4 \times 10^{-6}$ M.

#### *2.1.1.3 Indirect CD Spectroscopic Identification and Determination of Drugs by Chirality Induction*

*Summary* Achiral drugs and racemic mixtures cannot be studied directly by CD spectroscopy. It is conceivable that such compounds can be made "visible" for the CD detector and can be quantitated if an extrinsic/induced CD signal is generated upon inclusion of the drug chromophore into asymmetric environment of a chiral host. Cyclodextrines (CyDs) are well-known chiral complexation agents showing no CD activity above 210–220 nm. Inclusion of drug molecules into CyD cavity can result in solute-induced CD band(s) which can be used for drug identification and determination [120,121]. The procedure is simple: an aliquot of standardized CyD stock solution has to be added to the sample solution prior to CD measurement. It can be applied for achiral drugs [120], racemic mixtures, and even for chiral compounds. In the latter case, intrinsic CD curve of the enantiomerically pure guest is altered upon CyD complexation [122].

Drug content of pharmaceutical preparations might be assayed without prior mixture separation.

*Selected Examples* Distinctive features of CD spectra of 10 therapeutic benzodiazepines induced by β-CyD have been described [123]. Total drug concentration of commercial tablets and capsules containing diazepam and flurazepam has been determined directly by measuring ICD bands after addition of β-CyD to the sample solution. No time-consuming separation procedure was necessary for the CD assays of which reproducibility were better than 0.2%.

Meperidine (achiral) content of commercially available tablets was directly assayed using ICD data ($\Delta\varepsilon_{max}$ at 271 nm) obtained for a simple solution of tablets prepared in alkaline β-CyD ($10^{-2}$ M) stock reagent [121]. Correspondence with the prescribed amount was better than 98%.

ICD data have been used to estimate the association constants of eight barbiturate-CyD complexes [124]. The association constant obtained for seconal then was used

for quantitative assays of the drug in commercial seconal sodium suppositories without prior mixture separation.

*Comments*

- In order to determine drug concentration from ICD values, formation constant of drug-CyD complex must be known. This can be estimated from series of spectroscopic data obtained by measuring ICD values at different host/guest molar ratios.
- For working with racemic drugs, pure enantiomers are needed for determining stability constants.
- Achiral additives of pharmaceutical preparations bearing chromophoric groups (e.g., procaine, lidocaine, benzoic acid, caffeine) might also be included in the CyD cavity generating additional extrinsic CD bands which can be mixed to ICD bands of the drug under study.
- See also the Comments in Section 2.2.1 (Drug–Cyclodextrin Complexes).

### 2.1.2 Purity Investigation of Chiral Pharmaceutical Compounds by CD Spectroscopy

*2.1.2.1 Summary* Unfortunately, the literature of direct CD spectrometric purity investigations of drug substances and preparations is very short. Between the reasons could be found the high cost of a CD spectropolarimeter, the instruments, and methods with highest performance and sensitivity than the chiroptical methods and the limited possibility of use. On the other hand, the chiroptical methods have intrinsic selectivity in the analysis of chiral compounds. The most widely spread techniques are the different kinds of separation methods. The chiroptical aspects of separation techniques are detailed in Section 2.1.3, the unique and rare examples, which do not apply separation, are collected here.

Chiral pharmaceutical compounds could contain achiral and chiral impurities. In the mixtures containing achiral substances the chiroptical methods are selective and specific for the chiral components. The chiral impurities could be sorted in two groups:

1. Molecules with similar structure (structure isomers, homologs)
2. Stereoisomers (diastereomers, enantiomers)

*2.1.2.2 Selected Examples in Group 1* In this group the steroids have largest pharmaceutical impact. For instance, ethynodiol is manufactured from norethisterone, so norethisterone acetate might always remain as impurity in the product ethynodiol diacetate. The norethisterone has selective wavelengths over 331 nm due to the chiraly perturbed $\Delta^4$-3-keto chromophore, and in this way 0.5–5% of impurities could be selectively determined [125]. Similar problem is the $\Delta^5$-ethisterone determination in ethisterone [126]. Quick and accurate direct and indirect CD spectroscopic methods

were developed for the simultaneous determination of ethisterone (17α-ethinyl-17-hydroxy-4-androstene-3-one) and its $\Delta^5$-isomer ($\Delta^5$-ethisterone). The direct method is based on the selective negative CE of the $\Delta^4$-3-oxo group in ethisterone (negative maximum at 348 nm in dioxane) and measurement of the ellipticity at 296 nm (positive maximum of $\Delta^5$-ethisterone), where the measured ellipticity is the sum of those of the two isomers. In the indirect procedure $\Delta^5$-ethisterone was transformed to ethisterone by base catalyzed isomerization and the ellipticities were measured at 339 nm in ethanol before and after isomerization.

Drug molecules, like 6-oxo morphinane derivatives and 3-keto-steroids due to the $n \rightarrow \pi^*$ electronic transition (270–350 nm) can be measured directly with high precisity and selectivity [127]. The oxim formation reaction of keton group with hydroxil-amine is a short and quantitative reaction and causes dramatic change in the CD spectrum. The difference CD spectroscopy gives further sensitivity and selectivity to the method eliminating interference from other chiral compounds in pharmaceutical products. In addition to the improved selectivity, a further advantage of the method suggested is the high sensitivity since the 6-oxo-morphinans studied can be determined at a concentration of 1 mg/mL this procedure can certainly be applied to the analysis of biological fluids. The method based on the use of difference CD spectra might have an important role in the field of illicit drug analysis.

The quantitative determination of $\Delta^4$-3-ketosteroids is based on the considerable difference between the ellipticities before and after oxime formation around 320 nm [128]. The difference ellipticity for the six ketosteroids examined varies linearly with the concentration in the interval $6 \times 10^{-6} - 3 \times 10^{-3}$ mol/L. The advantages of the determination of $\Delta^4$-3- ketosteroids by difference CD spectroscopy are improved selectivity and high sensitivity. This method can be applied to the determination of $\Delta^4$-3-ketosteroids in the presence of achiral compounds and those chiral substances, which do not react with hydroxylamine hydrochloride. The progestogen norgestimate can be contaminated with levonorgestrel acetate. The reference [128] demonstrates that 0.02% $\Delta^4$-3-ketosteroid contamination in norgestimate can be determined with required accuracy by the difference CD spectroscopic method.

*2.1.2.3 Selected Examples in Group 2* Erskine *et al.* [129] elaborated a method for the determination of e.p. using CD and UV spectroscopy for ephedrine and mandelic acid enantiomers and ephedrine–pseudo-ephedrine mixtures. The CD and UV spectra overlap each other in the investigated region (256–280 nm). The CD and UV data were collected at 0.1 nm interval and the results were calculated by the partial least squares method. The UV spectral data give information about the total concentration of the sample, while the sign and intensity of the chiroptical signal provide information about enantiomeric composition and the enantiomeric excess.

The enantiomeric composition of phenylglycine and mandelic acid was successfully determined by the application of the anisotropy factor ($g$ factor) derivable from the ratio of the CD and the absorbance signals ($g = \Delta A/A$) [130]. This is an intensive, concentration-independent physico-chemical parameter, which eliminates concen-

tration-related errors caused by sample preparation, dilution, and so on. For the two enantiomers, it has the same numerical value with opposite signs, and is therefore a function of the enantiomeric composition. The procedure based on the determination of the $g$ factor is rapid, accurate, and eminently suitable for routine quality control purposes enabling the determination to be carried out with an accuracy of 0.1%. Only one restriction factor must be considered: if any kind of contamination (chiral or achiral) distorts the CD or UV signal, it must be separated.

Complex formation can be used as an aid to improve the sensitivity of the determination of e.p. of compounds; for instance, in the case of ethambutol complexation with copper(II) was used for this purpose [131]. A method based also on complex formation has been developed for determination of the binary mixtures of pseudoephedrine and ephedrine using chiral Cu(II)–tartrate complexes as derivatizing agents [132]. The reaction is a simple ligand exchange between tartrate and the chiral drug and results in a shift of the wavelength of CD detection to the visible range. The CD spectral change is proportional to the concentration of D-pseudoephedrine and at 527 nm the L-pseudoephedrine complexes have an isobestic point. At this wavelength, the D-pseudoephedrine content can be determined. The effects are entirely opposite if L-tartrate is changed to D-tartrate in which case the D-pseudoephedrine complexes will produce the isobestic point and the L-pseudoephedrine content can be determined. The detection limit for the enantiomeric impurity was in the order of 2%.

A specific complex formation of the anti-inflammatory drug (±)-ketoprofen with bovine serum albumin (BSA) was investigated [79]. The CE arising from the $n \rightarrow \pi^*$ transition of ketoprofen shows magnification due to the protein binding and it can be used to measure the e.p. of (−)-ketoprofen with excellent precision.

Separation and purification of drug enantiomers (thiopental, pentobarbital, methadone) by chiral capillary electrophoresis was supported by CD spectroscopic characterization of the enantiomeric composition of the collected fractions [133,134]. The samples were prepared by chiral recycling isotachophoresis of a racemic mixture.

A method has been elaborated for the determination of enantiomeric purity of amphetamine samples [135]. The procedure is based on the derivatization of amphetamine with 2,4-dinitrofluorobenzene and subsequent measurement of the anisotropy factor by dual (CD/UV) spectroscopy. The advantage of the derivatization is that due to the bathochromic shift, more selective and sensitive analysis can be performed than it would be possible in the underivatized form. Both ellipticity and absorbance values were in linear relation to the concentration in the range of $5 \times 10^{-6} - 6 \times 10^{-4}$ mol/L. By investigating binary mixtures of amphetamine enantiomers containing 63.3–98.8% of (+)-amphetamine the accuracy of the determination of e.p. was found to be 0.2–0.3%.

### 2.1.3 Coupled HPLC-CD Techniques for Solving Pharmaceutical Problems

*2.1.3.1 Advantages and Limitations of CD Detectors* The detectors (UV, fluorescence, refractive index (RI), electrochemical) generally used in the

analytical practice cannot detect chiral properties of investigated molecules and cannot differentiate between the enantiomers directly. The identification of enantiomers and the detection of chiral properties could be carried out with chiroptical detectors only.

There are two types of chiroptical detectors basically: the polarimetric (optical rotatory dispersion (ORD)) and CD equipments. The ORD detectors which is based upon RI differences between rcpl and lcpl can detect all kind of chiral compounds, but the sensitivity and selectivity of the method is not high enough as required in pharmaceutical industry. The sensitivity of CD detectors exceed the sensitivity of ORD, but the presence of chiral or chirally perturbed cromophore is required in the sample. This restriction limits the range of investigated materials, but gives inherent selectivity to the method. The sensitivity of CD detectors exceeds the sensitivity of ORD and RI devices, but equal or less sensitive as the UV detectors. Further problem is the relatively high cost and the limited wavelength scale. The shortest wavelength is 220 nm, typically, which could be used in the case of highly transparent eluent. The noise increases below 240 nm significantly, if the eluent contains slightly transparent component.

The most important aspect in method development for CD detection is the $g$ factor. The CD detector measures $\Delta A$, but measures the sum of absorptivity ($A$) of the cpl components also. The low value of $\Delta A/A$ ($\Delta\varepsilon/\varepsilon$) ratio limits the possibilities. If $\Delta A$ is too low, it manifests in low CD signal while if $A$ is too high, the compound cannot be detected with the precisity required by pharmaceutical industry, due to the poor S/N ratio. Additional limitation could be arisen from low $\Delta A$ values: to increase the measured CD signal, concentration of the sample must be increased, but the injected quantity is limited by the column loadability.

In spite of these limitations, the CD detection could be used at several fields of pharmaceutical analysis. The representative applications are as follows:

1. Characterization of optical and geometrical isomers after separation and quantitative determination of chiral compounds (Section 2.1.3.1)
2. Determination of e.p. after chiral or achiral separations (Section 2.1.3.2)

### *2.1.3.2 HPLC-CD Detection of Chiral Pharmaceuticals*

***Summary and Comments*** The analytical scale separation of isomers or different chiral substances allows on-line chiroptical characterization with CD detector and there is no need to purchase preparative columns and collect the fractions for posterior process of the collected samples. This method is especially important in cases when a dynamic equilibrium persists between the isomers and the time scale is relatively short, but exceed the separation time.

In time mode when CD detection occurs at a fixed wavelength, the elution order of isomers could be determined while the stop-flow method or the fast spectrum registration mode enables to register CD spectra of eluted components. If HPLC-CD systems are operated in fixed wavelength mode, stereochemical information carried by the complete CD spectrum cannot not be utilized. The sign of the measured

CD signal is variable, depending upon the wavelength selected, and cannot be correlated with the absolute stereochemistry of the analyte.

The construction of CD equipments allow the parallel detection of the CD and UV signals. This technical feature bears specific application possibilities. In practice, the CD spectropolarimeter itself can be directly used as a CD detector by coupling it with an HPLC system. A novel, and much more economic solution is the use of a CD detector, designed specifically for HPLC measurements [136]. This detector is suitable to determine 99.8–98% e.e. values. Basic and general information about the HPLC-CD detection could be achieved from the literature [14,137–142].

*Selected Examples* The neuromuscular blocking agent atracurium besylate contains 10 isomers in various proportions as four racemic pairs and two meso compound and their *cis*- or *trans*-isomers. Directly coupled HPLC-NMR and HPLC-CD detectors were applied to the identification of the isomers. The HPLC-NMR method was used for identifying the isomeric configuration and the enantiomeric pairs. The HPLC-CD experiments were complementary in the determination of the absolute stereochemistry [143].

Takahashi et al. [144] investigated the warfarin and 7-hydroxywarfarin binding process to proteins and measured the enantiomers from plasma and urine. The serially coupled CD detection combined with stop-flow method supported the identification of isomers and allowed to register on-line CD spectra of eluted components.

A diazepam analog was synthetized, characterized and its stereospecific HSA binding properties were investigated by NMR, CD spectroscopy, and HPLC-CD methods. The on-line CD detection clearly demonstrated that the stereoselectivity of the compound on immobilized HSA stationary phase is opposite to all the other studied benzodiazepines [145].

The metabolism of lorazepam enantiomers was also studied from plasma by on-line CD detection [146].

The 2,3-benzodiazepine derivative tofisopam has two types of chirality: the C(5) atom generates central chirality, while the diazepine ring affects the existence of two helical conformers. The four isomers were separated on cellulose-based stationary phase and were detected by the coupled CD equipment. The on-line CD spectra of the peaks obtained by stop-flow method definitely showed different chiral character [147].

A non-nucleoside reverse transcriptase inhibitor chiral molecule (DPC 961) which is used in the treatment of HIV infection was investigated. The plasma extract was analyzed with coupled HPLC-CD method to determine, if racemization occurred during metabolism. The tandem detection method (CD/UV) confirmed that there was no racemization of the chiral center of DPC 961 during metabolism. The LOD of the CD detector for DPC 961 at 240 nm was approximately 5 ng injected on column while the LOD for the UV detector at 250 nm was approximately 0.2 ng under the same conditions [148].

Ethisterone is one of the classical synthetic progestogenes and it is an important intermediate in the synthesis of the diuretic and cardiac drug spironolactone. In the course of the synthesis, a large amount of the $\Delta^5$-isomer is also formed. The ratio

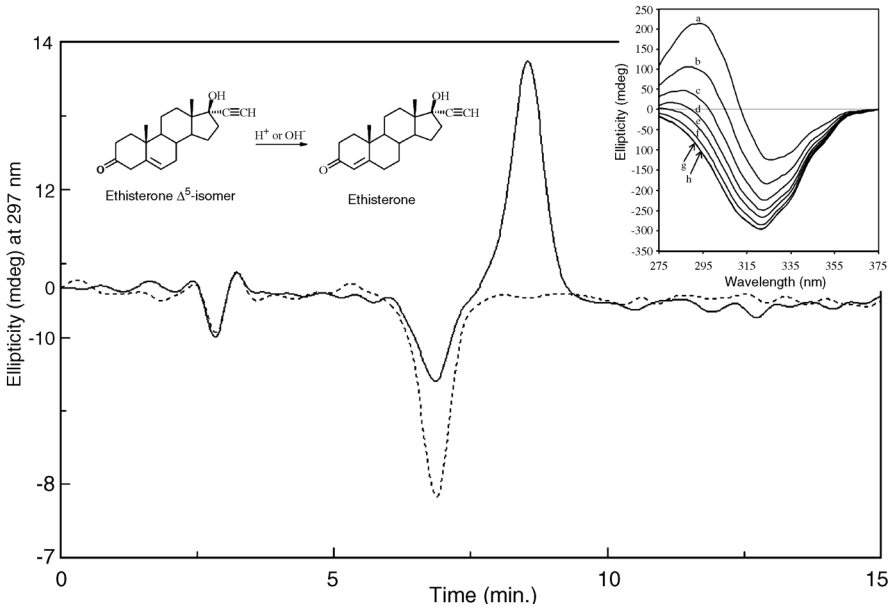


***FIGURE 2*** *HPLC-CD chromatogram of a sample containing 41.4% of ethisterone and 58.6% of $\Delta^5$-isomer before (continuous line) and after (dashed line) base-catalyzed isomerization. Inset: CD spectra of the reaction mixture of the isomerization reaction at different time intervals from start of the reaction up to the completion. (a) 0 min, (b) 2 min, (c) 4 min, (d) 6 min, (e) 8 min, (f) 12 min, (g) 25 min, (h) 35 min. The (g) and (h) curves correspond to the CD spectrum of pure ethisterone. Note the CD sign inversion at 297 nm. Adapted and modified from Ref. [126].*

of $\Delta^4$- and $\Delta^5$ isomers should be measured in the course of the in-process control during the spironolactone synthesis. Using HPLC-CD method (Fig. 2) it is possible to detect not only ethisterone with negative CE, but also the $\Delta^5$-isomer with low UV activity, and above all this, the base-catalyzed isomerization may be examined as well [126].

The application of HPLC-CD for identification of pharmacologically active compounds, such as dexamphetamine, 5-hydroxytryptophan, (−)-huperizine A, interferon as standards, in registered drugs, in falsifications, and in food supplements was described [149]. The UV, fluorescence, and CD detection were compared and validated for chromatographic monitoring of conformational properties of human interferon α2b [150].

The biologically important dehydroepiandrosterone (DHEA) is a precursor in the biological steroid synthesis, and it can be converted into other hormones. Gergely *et al.* [151] has developed an HPLC-CD method, where DHEA and three related biosynthetic steroids were separated. The use of CD detector provided selective detection because of its inherent selectivity for chiral compounds and it allowed the application of high detection wavelength (295 nm), presenting more selective detection with minimal background noise.

A specific scope of the dual CD and UV detection is the peak homogeneity monitoring. A general method is described in [152] and the feasibility of the method is exemplified by the analysis of partially overlapping codeine-hydrocodone and completely overlapping codeine-oxycodone samples. Another possibility for the peak homogeneity control is the fast CD spectrum recording during the separation process. Apparently homogeneous peaks of very similar analytes were analyzed and resolution was carried out [153].

CD detection combined with HPLC was used to demonstrate the chiral separation of perbenzoyl derivatives of the anti-tuberculosis agent ethambutol stereoisomers: the enantiomeric (−)-(*R*,*R*) and (+)-(*S*,*S*)-ethambutols together with the (*meso*)-(*R*,*S*)-form [154]. Comparison of the UV-HPLC and CD-HPLC traces showed that the first two UV peaks with equal and opposite CD effects correspond to the therapeutically active (+)-(*S*,*S*)-isomer (eutomer), and the inactive (−)-(*R*,*R*)-isomer (distomer) while the third peak could be seen only in the UV trace corresponds to the CD inactive *meso* form (Fig. 3). Derivatization was necessary since ethambutol lacks chromophoric groups and thus cannot be studied by CD spectroscopy. Upon derivatization, exciton type CD signals can be measured above 200 nm in the strong $\pi \rightarrow \pi^*$ UV band of the benzoyl moieties.

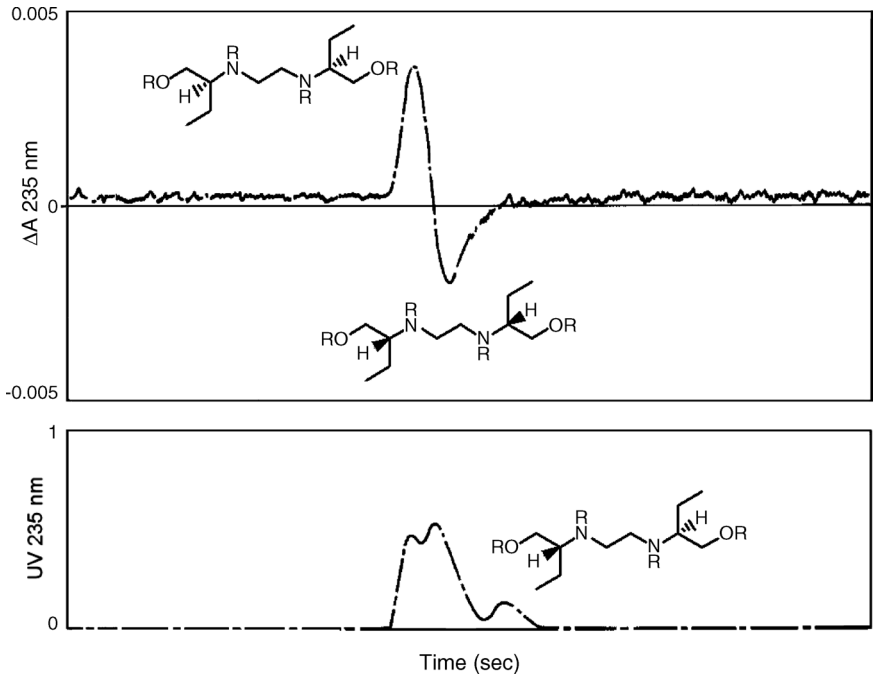


***FIGURE 3*** *HPLC-CD/UV chromatogram of perbenzoyl (R = COPh) derivatives of ethambutol (R = H) stereoisomers. The first and second peaks (CD and UV) correspond to the (+)-(S,S)- and the (−)-(R,R)-enantiomers. The third peak (UV only) represents the meso-form. Adapted and modified from Ref. [154].*

The absolute configurations of two novel and selective cyclooxygenase-2 inhibitors were assigned by the comparison of computed and experimental CD spectra measured on-line with HPLC-CD detection [155].

#### 2.1.3.3 Enantiomeric Purity Determination Using HPLC-CD System

*Summary and Comments* Resolution of enantiomers on chiral stationary phases is a commonplace recently. The chiroptical detection after chiral separation is preferred if the elution order of isomers is not known and if the sensitivity is acceptable. The opposite CD sign of enantiomers identifies them unambiguously. The restrictive features of the CD detection are the LOD and LOQ whose parameters vigorously depend on the $g$ factor. The $g$ value is not merely a restrictive factor, but it may be a useful analytical tool, also. As was previously mentioned, CD instruments detect and register simultaneously the CD ($\Delta A$) and UV ($A$) signal. The ratio of the $\Delta A/A$ for a single enantiomer is a constant value, the sign of the $g$ factor is identical as the sign of the CD signal and independent from the concentration. For details see Section 2.1.2.

By using the $g$ factor no chiral separation is needed and the e.e. can be determined from the chromatographic results of a non-chiral separation. The separation of other components showing own CD signal or UV absorbance is required, since these components distort $g$ values and cause erroneous purity determination. Additionally, the method is valid only if the $g$ factor is independent of concentration and if it is linear with respect to e.e. These conditions may not hold in the case of dimerization or self-association of the chiral analyte because own particular CD contribution of such enantiomeric or diastereomeric species would modify the CD profile of the monomeric form. Thus, it is mandatory to check this possibility in each new system under study.

Chiral analytes (e.g., sugars) showing no optical absorbance above 200 nm have to be made visible for the CD detector by previous derivatization step using chromophoric reactants.

*Selected Examples* A sensitive, selective, and versatile CD-based HPLC detection system was used for validation of the e.p. assay in the quality control of chiral drugs upon non-chiral stationary phases [156]. The precision and accuracy of the method were checked for selected samples showing values of the $g$ factor in the order of $10^{-1}$–$10^{-4}$. Very high accuracy was obtained in the case of extreme e.p. values (less than 1% or higher than 99% e.p.) and also for a low $g$ value ($2 \times 10^{-4}$) compound. Plots of $g$ values as a function of e.p. showed good linearity for all analytes, with correlation coefficients ($r^2$) between 0.970 and 0.999, even in the range of extreme e.p. values. The high selectivity of this detection system allows a selective monitoring of analytes in complex mixtures and makes the baseline stable. The minimum detectable e.p. value is mainly related to the maximum $g$ value, while the minimum amount of the detectable analyte depends on the values of the CD and absorption signals.

A simple and rapid non-enantioselective HPLC method using a CD detector for the determination of the e.p. of dexchlorpheniramine maleate was developed [157]. Such

a method, which does not require the enantiomer separation, seems to ensure adequate performance; indeed the results obtained and the reported data on precision, linearity and sensitivity show that the described HPLC-CD system is suitable for the determination of e.p. of compounds having an unfavorable CD signal.

The HPLC determination of efaroxan e.p. was successfully performed by using a $C_8$ silica column and a CD detector [136]. This method presents the advantages of being fast and precise and does not require the use of an expensive chiral column. Moreover, due to UV and CD detections, it was possible to simultaneously determine the optical and chemical purities of drugs. This last feature could be applied to study racemization and metabolization of the eutomer in biological fluids, too. For the first time, this HPLC method—with achiral support and CD detection—has been fully validated according to the ICH (International Conference on Harmonization) criteria and can replace the classical HPLC approach—chiral support and UV detection. Nevertheless, the LOQ in terms of e.p. is for the moment too high (98%) compared with results generally obtained with UV detection. Some technical improvements concerning the sensitivity and the linearity of the CD detector are required in the near future to achieve quantification of e.p. up to 99.8% level.

A specific, inexpensive optics were inserted into a conventional photo-diode array detector and this modification resulted in a multi-beam CD detector. With this equipment the e.p. determination of malic acid in beverages was performed [158]. The samples whose e.p. were lower than 97.3% were classified as adulterated and it was confirmed by capillary electrophoresis. Some of these samples were labeled as 100% pure natural product and no additives was listed in their ingredients. In this context, the HPLC-CD detection is a feasible method to explore falsifications in natural products.

Using non-chiral HPLC coupled with a CD detector, a new method was developed to determine the e.p. of glucose [159]. Due to the lack of chromophores absorbing light above 200 nm, the sugar molecules were derivatized with *p*-bromobenzoyl chloride that made CD detection possible in the UV region. In order to achieve more accurate e.p. determination the wavelength of detection was set at 254 nm, near to the maximum difference of *g* factors of D- and L-glucose derivatives. The method is suitable for microscale analysis of glycoside and saccharide. As an application, the component glucose of a glucan was assayed to be pure D-isomer.

The HPLC-CD approach can be used for the development of fast enantiopurity assays in support of high throughput pharmaceutical process research investigations, e.g., for monitoring e.p. of products of asymmetric catalysis [48,55]. The HPLC flow injection analysis (FIA) combined with UV and CD detection results in FIA chromatograms (Fig. 4) which can be visually interpreted, allowing the user to rapidly assess which samples are showing conversion to product, which are forming the desired enantiomer, and allowing an immediate rough estimate of e.p. [55]

Holik and Mannschreck [160] discussed the possible errors determining e.p. from overlapping (non-resolved) HPLC peaks, while Gergely *et al.* [161] reported further practical information, how unbiased *g* factor can be calculated from the chromatographic data.

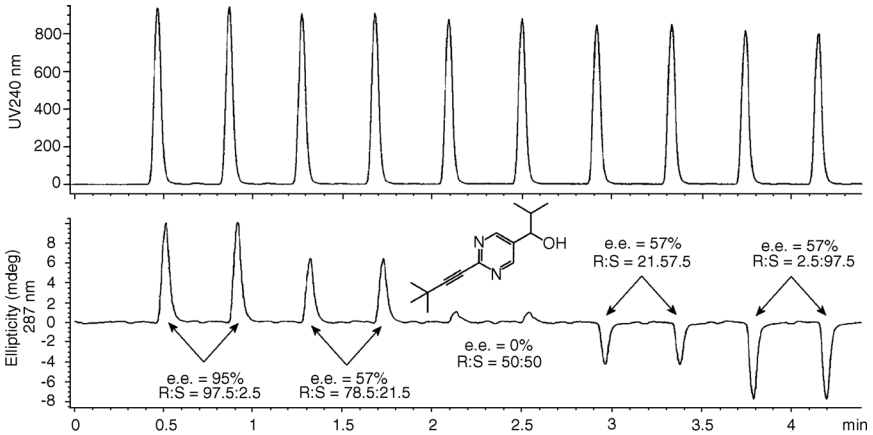


***FIGURE 4*** *Flow injection HPLC-CD analysis of mixtures of a compound of varying degrees of enantiopurity. Duplicate injections of samples prepared by mixing solutions of the individual enantiomers, nominally 95% e.e. (R), 57% e.e. (R); 0% e.e., 57% e.e. (S), 95% e.e (S). Note that 10 measurements were taken within 4.5 min. Adapted and modified from Ref. [55].*

## 2.2 Using CD Spectroscopy for Characterization of Pharmaceutical Preparations

### *2.2.1 Drug–Cyclodextrin Complexes*

*2.2.1.1 Summary* α-, β-, and γ-cyclodextrins (CyD) and their derivatives have been used extensively by the pharmaceutical industry, especially for enhancing the solubility, stability, and bioavailability of drug molecules [120,162]. These cyclic sugar hosts formed by the degradation of starch possess a hydrophilic surface and a hydrophobic internal cavity suitable for complexation of a wide variety of pharmaceutical agents. Due to the presence of numerous stereogenic centers (e.g., 35 in β-CyD) inclusion of achiral guest compounds in the asymmetric cavity might induce CEs in the CD spectrum of guest compounds [121,162,163] (Fig. 5A). For enantiomers of racemic mixtures, the resulted complexes are in diastereomeric relation of which non-equivalent chiroptical contributions result in non-zero CD activity [164,165] (Fig. 5B). CyD complexation of pure enantiomers might also be detected due to the modification of their intrinsic CD profile in relation to that measured in CyD-free solution [166–169] (Fig. 5C). It should be noted that simultaneous inclusion of two guest molecules is also possible especially for CyDs with large cavity (γ-CyD). In such cases, mutual chiral disposition of the entrapped guest chromophores dictated by the asymmetric host might result in guest–guest chiral intermolecular exciton coupling and characteristic ICD band splitting [170–173].

Chiroptical investigation of drug–CyD complexes can be useful to resolve several kinds of pharmaceutically relevant issues.

- CyD complexation of achiral/racemic or chiral drugs can be quickly demonstrated by measuring extrinsic or modified intrinsic CD bands of the guest compounds [162,167,168,174,175].

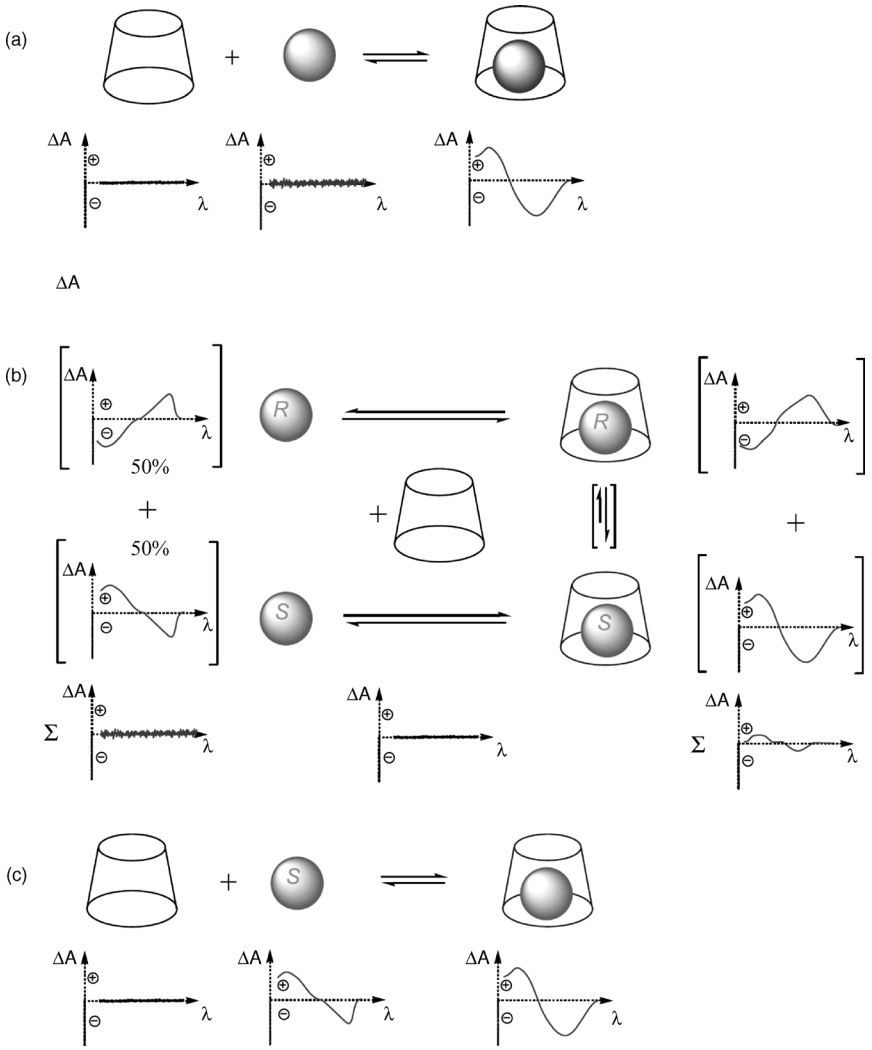


***FIGURE 5*** *(a) Representation of CD spectra ($\lambda > 210$ nm) of a CyD and an achiral guest before and after the formation of the inclusion complex. (b) CD spectra of a CyD and a racemic mixture before and after the formation of inclusion complexes being in diastereomeric relation. (c) CD spectra of a CyD and a chiral guest showing intrinsic CD bands before and after the formation of the inclusion complex. Adapted and modified from Ref. [162].*

- Concentration dependence of the CD signals allows to estimate binding parameters of the inclusion complexes. From series of CD spectroscopic data obtained by increasing guest concentration in sample solution of CyD kept at fixed concentration (or vice versa), the equilibrium constant and the stoichiometry of the complexes can be calculated [162,169,176].

- The presence/absence of ICD bands of achiral/racemic compounds or the alteration of intrinsic CD signals of chiral agents can serve as spectroscopic tool for screening of different CyDs to pre-select hosts with promising inclusion capability for selected drug molecules [177,178].
- Monitoring the change of ICD spectra of drug–CyD complexes under various experimental conditions (pH, temperature, time, ionic strength, additives, etc.) allows the characterization of physico-chemical stability and optimal conditions of formation of the inclusion complexes [162,168,177,179,180].
- Sign, shape, and intensity of extrinsic CEs of drug–CyD complexes are structure-dependent and can thus be used to assess the geometrical arrangement of guest compounds inside the CyD cavity [162,163,175].

*2.2.1.2 Selected Examples* Formation of CyD inclusion complexes of numerous pharmaceutical substances were demonstrated by measuring extrinsic CD spectra. In Table 3, some representative examples are shown with binding parameters estimated from analysis of ICD data [123,124,165,167,174,175,177,181–185].

The antipsychotic agent ziprasidone was formulated by using β-CyD sulfobutyl ether for solubilization of the drug. ICD spectra of various salt forms of ziprasidone suggested that the counterion of each salt form participates in the complex formation [166]. The ICD values at 318 nm measured at different guest:host molar ratios indicated 1:1 stoichiometric relationship.

Far-UV CD spectra of the bradykinin potentiating pentapeptide complexed with β-CyD qualitatively demonstrated the host:guest interaction [166]. CD curve of the peptide characteristic for the random coil structure changed drastically in the presence of β-CyD suggesting greater conformational rigidity within the complex.

Alteration of the intrinsic CD spectrum of luteinizing hormone-releasing hormone agonist buserelin acetate indicated that the aromatic side chains of the peptide are incorporated into the hydrophobic cavity of dimethyl-β-CyD [186].

**TABLE 3 Examples for Drug–CyD Interactions Studied by CD Spectroscopy**

| Drug | CyD | Stoichiometry [drug]:[CyD] | $K_a$ ($M^{-1}$) | Ref. |
|---|---|---|---|---|
| Barbiturates | β | 1:1 | $0.2–5.5 \times 10^2$ | [124] |
| Benzodiazepines | β | 1:1 | $1.1–9.3 \times 10^2$ | [123] |
| Diflunisal | β | 1:1 | $1.7 \times 10^3$ | [181] |
| Diclofenac | β | 1:1 | $3.5 \times 10^2$ | [182] |
| Dicoumarol | β | 1:1 | $4.9 \times 10^3$ | [175] |
| Doxorubicin | γ | 1:1 | $6.7 \times 10^2$ | [177] |
| Ephedrine | β | 1:1 | $1.8 \times 10^2$ | [183] |
| Flufenamic acid | β | 1:1 | $1.3 \times 10^3$ | [167] |
| Hydrocortisone | α, β, γ, dimethyl-β | 1:1 | $1.7–6.0 \times 10^3$ | [184] |
| Licochalcone-A | β | 1:1 | $3.2 \times 10^2$ | [174] |
| (±)-Metoprolol | α, β, HP-β | 1:1 | $0.6–3.5 \times 10^2$ | [165] |
| Zileuton enantiomers | β, γ | 1:1 | $0.7–5.1 \times 10^3$ | [185] |

Association constant and stoichiometry values were estimated from ICD titration data.

#### 2.2.1.3 Comments

- CyDs are optically transparent above 210–220 nm and they have no contribution either to the CD or the absorption spectra. For CD spectral analysis below 220 nm where intrinsic CEs of CyDs can be observed, chiroptical contribution of CyDs must always be taken into consideration by arithmetic subtraction of CD curve of free CyD from that of the guest–CyD complexes.
- Small molecule–CyD complexation does not necessarily result in CD spectral changes [184,187]. This might be due to various reasons, such as inclusion of the non-chromophoric part of the guest, and sterically rigid chiral molecules of which intrinsic CD signals are not or hardly affected upon the inclusion [163].
- There is no correlation between the value of the binding constant of guest–CyD complexes and the intensity of the ICD bands. In spite of a high constant, ICD amplitudes may be small and vice versa [163,184].
- Interpretation of ICD spectra may be difficult when the guest molecule contains more than one chromophore. The CD bands induced by different chiral environments of the chromophores might overlap resulting unusual CD profile which can only be resolved by using additional information obtained with monochromophoric guests [188].
- In some cases, especially when guest molecules have high conformational flexibility, measure of the CD spectral changes induced by CyDs might not correlate with the stability constants estimated by other methods [189].
- Depending on the concentration used, CyDs can form aggregates [190]. In most cases, size of these aggregates remains small (<20 nm) and does not affect optical properties of the sample but in some cases they can grow causing significant light scattering and thus distorted CD spectra.
- CyD associates can also act as solubilizers binding of drugs through non-inclusion manner [184,191]. In such cases, care must be taken in interpreting and using ICD data since they can reflect both inclusion and non-inclusion interactions, respectively.

### 2.2.2 Protein and Peptide Pharmaceuticals

*2.2.2.1 Summary* CD spectroscopy is a sensitive and straightforward tool for elucidation solution structure of peptides and proteins under variety of experimental conditions. In the far-UV region of the spectrum (190–240 nm), CD bands of proteins arise from the $\pi \rightarrow \pi^*$ and $n \rightarrow \pi^*$ transitions of peptide bonds and aromatic chromophores and report on the basic elements of secondary structure (e.g., α-helix, β-sheet) [10,192,193]. The near-UV CD spectrum (240–320 nm) is determined by asymmetric environment of the aromatic side chains of Trp, Tyr, Phe, and by the chirality of disulfide bonds [194,195]. CD signals of this region are highly dependent on the tertiary structure of the protein. Evaluation of the far-UV and near-UV CD curves enables

- to verify the correct folding and final conformation of a protein;
- quantitative assessment of the secondary structure elements (λ-helix, β-sheet content);

- to test changes of biologically active conformation of proteins under different conditions (e.g., effect of temperature, pH, freeze-thawing cycles, additives, derivatization) [196–199];
- to follow structural changes during formulation, storage, release, and administration (degradation, folding/unfolding, self-association/dissociation) [200–202];
- to investigate the structural equivalency of mutant/native [203], recombinant/native [204], and recombinant/recombinant therapeutic proteins;
- to check impurities of protein samples (e.g., DNA).

*2.2.2.2 Selected Examples* Far-UV and near-UV CD curves of a monoclonal antibody were measured and compared before and after its covalent conjugation with a lipid anchor for incorporation into immunoliposomes [205]. The CD data displayed no major alterations either of the secondary or tertiary structures suggesting the retained antigen binding activity of the derivatized antibody.

By measuring ICD bands generated upon binding of haptens (di- and trinitrophenyl derivatives), CD spectroscopy was shown to offer a new and powerful approach for studying and comparing the fine structure of the combining sites of antibodies of similar specificity and to probe the reformed site resulting from the reassociation of heavy and light chains [206].

Hydrophilic β-CyDs added to insulin solution caused CD spectral changes indicative for dissociation of the insulin aggregates to monomer or lower-order aggregates [207].

For developing a carrier for oral delivery, insulin was incorporated into chitosan/alginate nanoparticles. Far-UV CD spectra demonstrated the preservation of insulin structure during the nanoparticle formulations as well as after its release from the particles under gastrointestinal simulated conditions [208]. Despite of the presence of asymmetric centers of chitosan and alginate, the high optical transparency of these compounds allowed to measure the protein CD curve down to 190 nm.

Far-UV CD spectra indicated profound, irreversible structural changes of interleukin 2 caused by transient surface association with the infusion pump/catheter tubing system used for long-term delivery purposes [202]. The biological activity of the delivered protein was dramatically reduced by ≈90% after a 24 h infusion program.

pH and temperature dependency of CD spectra have been utilized for characterization of the recombinant anthrax antigen, showing highest thermal stability within the pH range of 6–8 [209]. CD spectroscopic screening of various stabilizers required to prevent the aggregation of the antigen helped to select compounds showing the most effective stabilization on secondary and tertiary structure of the anthrax protein.

Far-UV and near-UV CD spectra of diphtheria toxin and toxoids were taken to compare their secondary and tertiary structure. Near-UV CD spectra have shown that among the inactivation methods applied for preparing toxoids, acetylation and methylation of lysine residues hardly but glutaraldehyde and formaldehyde treatments drastically altered the protein tertiary structure [196,197]. In contrast, the far-UV CD spectra were essentially the same, indicating that secondary structure was unaffected by the detoxification process.

The anti-microbial preservative benzyl alcohol is widely used in the formulation of multi-dose parenteral products. Such agents, however, can greatly accelerate formation of non-native protein aggregates. In addition to the potential loss of biological activity due to aggregation, the presence of aggregates can cause adverse responses in patients. The addition of benzyl alcohol at 0.9% (w/v) to liquid formulations of human interferon-γ caused gradual decrease in the near-UV ellipticity values indicating the loosening of protein tertiary structure. This interaction pre-disposed the formation of high molecular weights protein aggregates as demonstrated by dynamic light scattering. Effects of the protein concentration, buffer components, buffer concentration on the aggregation were also studied by recording changes in the near-UV CD spectrum [201].

Analysis of ICD spectra of dye–HSA and dye–enzyme complexes served as an aid for the selection of dyes suitable to prepare sorbents for protein purification [210]. In addition, measuring the extinction of extrinsic CD signals of the complexes upon addition of various agents helped to establish the materials necessary for the elution of proteins bound on the sorbent.

#### *2.2.2.3 Comments*

- There are a range of various computer programs available for calculating the percentage of different structural motifs from the far-UV molar CD data of proteins, thus precise determination of the protein concentration is important [17]. In general, a basis set of reference proteins of known secondary structure is used to calibrate these methods by providing information on CD contributions of the various types of secondary structure existing in real proteins [37,211,212] (Fig. 6). These methods should be applied with caution for peptides, because globular proteins are used in the reference sets.
- It is a common practice to use the mean residue concentration (MRW) for displaying far-UV CD spectra of proteins and peptides as mean residue ellipticity units ($[\Theta]_{MRW}$). In contrast, near-UV and visible region of the CD spectra are reported in molar ellipticity ($[\Theta]$) or in $\Delta\varepsilon$ units in terms of protein or protein subunit molarity. It is important to differentiate between these kinds of data expression and to clarify the basis on which the molar concentration is calculated.
- Distilled water is the optimal choice for recording far-UV CD spectra of proteins and peptides. Buffers used for such purpose must have a low absorbance. Maximum absorbance of the samples should do not exceed 1.0 AU (including solvent and cell) [17,37].
- Particles of nanometer size (aggregates, impurities) in the sample solution may cause significant light scattering and thus overestimation of the protein concentration determined by UV spectroscopic method. Consequently, molar CD values will be lower which might be problematic for quantitative comparison of CD spectra of the same protein isolated from different biopharmaceutical formulations [213].
- Far-UV CD spectra of proteins might be insensitive for protein aggregation if it does not alter the secondary structure [201].

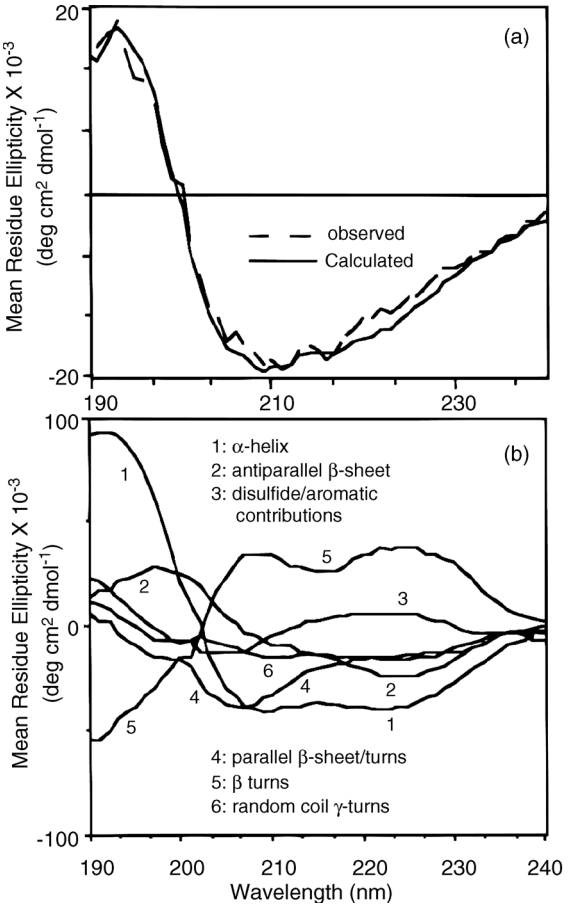


***FIGURE 6*** *Secondary structure analysis of an 81 amino acid yeast killer toxin from its far-UV CD data. (a) CD spectrum of the toxin in water and a spectrum calculated by reconvolution of the discrete structural contributions determined by the Convex Constraint Analysis method. (b) Results of deconvolution of the CD spectrum of the toxin into CD component curves of six pure secondary structural motifs. Adapted and modified from Ref. [212].*

### *2.2.3 Liposomes*

*2.2.3.1 Summary* Liposomes are the most clinically established nanometer-scale drug-carrier systems which provide a means to alter the biodisposition of hydrophobic drugs, genes, vaccines, as well as imaging agents. Among the biophysical techniques applied in the field of liposome research, CD spectroscopy can be used for

- detection of successful incorporation of drugs into liposomes [214–216];
- evaluation of the molecular state (e.g., monomer-aggregate, metal-complex) of drugs entrapped in the lipid environment [214,216,217];

- following the stability of liposomal preparations in the function of various environmental parameters (pH, temperature, time, etc.) [218,219];
- assessing structural integrity/conformational changes of therapeutic agents formulated in liposomes [220–222].

*2.2.3.2 Selected Examples* CD spectra of the chiral anthracycline antibiotics doxorubicin, daunorubicin, idarubicin and idarubicinol were compared in aqueous solution and in the presence of large unilamellar vesicles. Changes of CD profiles of the drugs indicated that both doxorubicin and daunorubicin bind to membranes in monomeric form. In contrast to this, the highly hydrophobic idarubicin and idarubicinol form small molecular associates within the bilayer involving 2–3 drug molecules, membrane cholesterol, and phosphatidic acid [216].

Taxol (paclitaxel) and its derivatives (docetaxel) are novel chemotherapeutic agents introduced to the clinical practice for treatment of malignant diseases. Due to their poor water-solubility, both paclitaxel and docetaxel require a complex solvent system for commercial formulations. Regarding their better toxicology profile, liposomal-encapsulated form of taxanes is a promising avenue for parenteral administration of these drugs [223]. CD spectroscopic evaluation of formulation characteristics can provide the information necessary to develop stable taxane-liposome preparations [214]. High similarity of the CD spectra of taxol measured in methanol and in liposome-incorporated state at low drug/lipid ratios indicated molecular solubilization of the drug in the lipid membrane. Upon increase the drug/lipid ratio, prominent changes of the taxol CD bands were observed which signify the drug destabilization and aggregation. By recording CD value of liposome-encapsulated taxol around 295 nm in the function of time enabled comparison of temporal stability of various formulations. In addition, quantitative analysis of CD spectral data permitted estimation of the fraction of taxane incorporated in the liposome membrane.

When protein or peptide pharmaceuticals are formulated in liposome delivery systems it is important to check their structural integrity on association with and after dissociation from membrane bilayers. Far-UV and near-UV CD spectra of human interferon-$\gamma$ recorded in liposome associated state and after desorption from liposomes showed no substantial difference compared to the native protein indicating full retention of both the secondary and tertiary structure [220].

CD spectroscopic investigation of the secondary structure of liposomal vasoactive intestinal peptide (VIP) revealed significant increase in the $\alpha$-helical content upon interaction of the peptide with phospholipids compared with the predominantly random coil form in aqueous solution [221]. The enhanced *in vitro* activity of the liposomal VIP was correlated with this structural transition since the $\alpha$-helix is the preferred conformation for VIP–receptor interactions.

CD spectroscopy is particularly useful for probing aggregation state of the strongly hydrophobic chiral polyene antibiotics like amphotericin B (AmB), filipin, and nystatin in liposome formulations [224]. The asymmetric centers of the macrolide ring render the molecular associates chiral (supramolecular chirality). Positive–negative CD peaks at 325–330 and 343–348 nm are assigned to the

chiral associates of AmB, while three positive peaks at higher wavelengths (378, 397, and 412 nm) are indicative to the monomeric state of the drug. Relative intensities, shape, time- and temperature [218,219] dependence of these CEs give valuable data on the monomeric–oligomeric drug equilibrium, localization of AmB molecules inside liposomes [217], interaction of AmB with plasma components [225] and help to determine the optimal conditions when the drug self-association is minimal [226].

#### *2.2.3.3 Comments*

- Intense light scattering of large liposomes distorts the CD spectrum making difficult to record CD signals of drugs incorporated. Accordingly, the light scattering interference must be reduced to a minimal level. This can be achieved by the reduction of the concentration of liposomes, using a short path length cuvette, and by a brief sonication of the sample solutions prior to the CD measurement to decrease the size of liposomes.

### *2.2.4 Gene Delivery Systems*

*2.2.4.1 Summary* Gene transfer represents an important treatment option for several genetic and acquired diseases with concrete perspectives of clinical application. Gene delivery systems are generally divided into two categories in terms of gene transfection carrier materials: viral and non-viral systems. For non-viral systems, cationic/anionic/neutral lipids and polymers complexed with chromosomal/linear/plasmid DNA or with synthetic oligonucleotides are frequently used due to their advantageous biological and formulation properties [57]. These systems are capable of condensing DNAs and improving their *in vivo* stability and cellular uptake.

Due to presence of chiral sugar units both single- and double-strand nucleic acids acquire helical secondary structure in which chiral exciton-coupling arises between the $\pi$–$\pi^*$ transitions of the stacked nucleobases resulting strong CD signals below 320 nm [10,30,57]. For nucleic acids, peak position, shape, sign, and intensity of the CD signals are functions of the nature of base-stacking interactions and the helicity of DNA. Thus, CD spectra present excellent indicators of various structural forms (A, B, C, D, T, Z) of nucleic acids. CD spectral modifications of nucleic acids induced by their interactions with gene delivery components can be applied to

- demonstrate complex formation between nucleic acids and the carrier [57];
- evaluate and compare the conformation and organization of nucleic acids in different formulations [227–231];
- study the organizational state of nucleic acids within the gene delivery vehicles under variety of experimental conditions (pH, temperature, ionic strength, etc.).

*2.2.4.2 Selected Examples* Amine-substituted gemini surfactant-based nanoparticles containing plasmid DNA were examined by using various methods including CD spectroscopy [227]. The red shift and intensity changes of the CD bands suggested distortion of the B-DNA conformation as a result of compaction of the DNA molecule by the surfactant.

Conformation and condensation of DNA complexed with cationic polymers were studied by CD spectroscopy. Depending on the charge ratio applied, B → C conformational transition of the double helix [228] and the appearance of polymer-and-salt-induced $\Psi^{-}$-type DNA [229], a highly condensed chiral tertiary DNA structure were observed.

CD studies indicated the right-handed B-DNA → left-handed Z-DNA conformational transition during anionic lipoplex formation [232]. Comparison of chiroptical profiles of the anionic and cationic lipoplexes suggested significant differences between the conformational transitions of DNA in the two formulations. It was noted that the formation of Z-DNA may contribute to the high transfection efficiency of anionic lipoplexes.

#### *2.2.4.3 Comments*

- Most gene delivery vehicles are heterogeneous, non-covalent complexes of DNA, and the delivery agent. Therefore, consistency of the procedures used to prepare the complexes is essential for the reproducibility of CD spectroscopic data.
- In general, chiral compounds (e.g., cholesterol and its derivatives) used to synthetize carrier components of gene-delivery formulations have negligible CD contribution above 220 nm and thus allow CD signals of DNA to be clearly resolved with little or no interference from the lipid component. In contrast, using some chiral cationic polymers such as polylysine, polyarginine, polyhistidine which display strong intrinsic CD signals below 230 nm may limit the spectral region suitable for CD scanning.
- Due to inadequate purification, chromosomal DNA can be associated with histones and other nuclear proteins, which will potentially contribute to the CD spectra.
- Since diameter of the gene-delivery particles are typically greater than 120 nm and display a heterogeneous size distribution, effects of light scattering to CD as well as UV absorption spectra (flattening effects) have to be taken into consideration [57,230].

## 2.3 CD Spectroscopic Investigation of Drug–Biopolymer Interactions

### *2.3.1 Drug–Protein Binding*

*2.3.1.1 Summary* There is a close analogy between chiroptical phenomena described for drug–CyD and drug–protein interactions. Similarly to CyDs, proteins are also asymmetric hosts, thus protein-bound ligands might display ICD bands, too. It is to be noted that most CyDs show no intrinsic CD activity above 220 nm but due to the presence of several asymmetrically perturbed chromophores proteins display intrinsic CD activity in the near-UV region (240–320 nm) which can disturb measuring ICD bands of ligands in this region. In addition, aromatic amino acids, especially tryptophan side chains might participate in generating ICD bands of protein-bound ligands via non-degenerate chiral exciton coupling mechanism [21,105,233,234]. Further mechanisms

which can also be involved in CD spectroscopic manifestation of ligand–protein interactions are

(a) the chiral perturbation of electronic transitions of the guest molecule by the asymmetric protein matrix [21,79,235];

(b) the helical (twisted) conformation of the chromophoric moiety of the ligand molecule adopted at the binding site [21,235–238];

(c) intramolecular chiral exciton coupling between the chromophoric parts of a protein-bound ligand molecule [21,42,235];

(d) intermolecular exciton coupling between two (or more) guest molecules bound close to each other (<10 Å) to the protein matrix [21,42,235].

Practical applications of ICD spectra in studying drug–protein interactions are as follows:

- Measuring ICD bands is a valuable diagnostic tool for demonstration of drug–protein binding interactions (Table 4).
- Quantitative analysis of ICD spectroscopic data measured at different [drug]/[protein] molar ratios (CD titration) allows determination of the number of binding sites per a protein molecule and the affinity constant [105,235,237,239–242].
- By monitoring ICD signals of marker ligands known to bind to specific binding areas of the protein upon addition of a drug, binding site of the latter can be identified (CD displacement method) [42,240,243]. Alternatively, displacement/competition studies can be carried out by measuring the change of the ICD signal of a protein-bound drug with unknown binding site upon the addition of site markers [242].
- Drug-binding properties of native, recombinant, chemically modified forms [233,235,244] or genetic variants [42,237,240] of the same protein can be assessed by comparative evaluation of ICD spectra.
- The species differences of drug binding parameters of pharmaceutically important proteins can be evaluated by ICD methodology [245,246].
- Comparison of extrinsic CD spectra of drug-native protein and drug-mutant protein complexes may help for identification of amino acid residues involved in the drug-binding mechanism [233,247].
- Allosteric, cooperative, and competitive interactions between different ligands bound to the protein host can be investigated by measuring ICD spectra [42,248–250].
- Conformational changes of the protein provoked by drug binding can be studied by measuring far-UV CD spectra [102,251].

*2.3.1.2 Selected Examples* Numerous papers are available reporting CD spectroscopic studies of binding of various therapeutic drugs to the principal transporter proteins of human blood, HSA, and $\alpha_1$-acid glycoprotein (AAG). The binding was detected and binding parameters were determined by measuring and analyzing ICD spectra of the protein-bound drugs (Table 4).

**TABLE 4 Selected Examples for Drug–Human Plasma Protein Binding Interactions Studied by ICD Spectra**

| Drug | HSA | AAG |
|---|---|---|
| Antimalarial agents | | [240] |
| Benzodiazepines | [252] | [252,253] |
| Coumarin anti-coagulants | [248] | [253,254] |
| Dipyridamole | | [255] |
| Imatinib | | [237] |
| NSAIDs | [256–258] | [253] |
| Phenothiazine neuroleptics | | [259] |
| Prazosin | | [260] |
| Sulfonamides | [261] | |
| Thyroxine | [262] | |

The site II specific marker (±)-ibuprofen decreased the ICD signal of leukotriene $B_4$ bound to HSA indicating the binding of leukotriene to the sub-domain IIIA [243].

Changes of the ICD spectrum of HSA-bound diclofenac and carprofen proved the site-II-to-site-I displacement of the drugs upon addition of ibuprofen [263].

Comparison of ICD spectra of the tyrosine kinase inhibitor imatinib obtained with the main genetic variants of AAG helped to assign the high-affinity binding site of the drug on the F1/S variant [237].

Interaction of several substrates (e.g., dihydrofolate), coenzymes (NADPH), and inhibitors with dihydrofolate reductase was demonstrated by measuring extrinsic CD bands [105,264] and the binding parameters were estimated from ICD data of the enzyme-bound ligands [105]. ICD spectra showed the formation of reductase–NADPH–methotrexate ternary complex [105].

The strong negative CD band of colchicine at 340 nm vanished or was greatly reduced upon its specific binding to tubulin. Decrease of the CD values enabled to estimate the colchicin–tubulin association constant [241].

The potent immunosuppressive drug rapamycin (Sirolimus) mediates its effects via binding to an intracellular protein termed FKBP (immunophilin). CD spectroscopic investigation of rapamycin/FKBP complexes disclosed very subtle conformational changes of the ligand undetectable either by NMR or by X-ray crystallographic methods [265].

ICD spectra of acetylsalicylic acid indicated its binding interaction with human myeloma immunoglobulin [266].

The binding of rifampin to human tear lipocalin was monitored by measuring alteration of intrinsic CD spectrum of the chiral drug above 300 nm [242]. Using the change in CD activity at 359 nm, the binding parameters were calculated. The binding site of rifampin was identified by CD displacement experiment performed with palmitic acid known to bind in the central cavity of the protein. pH dependence of the CD spectrum of the rifampin–tear lipocalin complex showed that the drug may be released from the protein in acidic environment. The results suggested the possibility of using tear lipocalin as a specific carrier for the delivery of rifampin to tuberculous lesions that are acidic and rich in fatty acids.

Specific changes in the far-UV CD spectrum of the rat β-adrenoceptor protein obtained in the presence of (−)-epinephrine, but not with its (+)-antipod, indicated the stereoselective binding of the pharmacologically active (−)-enantiomer [102] (Fig. 7).

Far-UV CD spectra demonstrated that upon binding of the anti-depressant benzodiazepine agent alprazolam the α-helicity of hemoglobin subunits are gradually lost [251].

#### *2.3.1.3 Comments*

- Strong intrinsic CD activity of proteins masks ICD signals below 240 nm. Accordingly, protein binding of drugs showing no light absorption above this region cannot be studied by ICD method.
- The main disadvantages to using far-UV CD spectroscopy to detect protein structural changes upon ligand binding are that *1)* a fairly substantial conformational change is required to detect increase or decrease of CD band intensities below 240 nm, *2)* the CD spectrum provides no information on which sequences of the protein are involved in the conformational rearrangement.
- Depending on the intensity of the CD signals, relatively high protein concentration is required to perform CD titration or CD displacement experiments ($\approx 10^{-5}$–$10^{-6}$ M in 1 or in 2 mL sample volume).
- Drug-protein association constants much higher than $10^7$ $M^{-1}$ cannot be determined by regression analysis of ICD titration data since the binding equilibrium is completely shifted toward the protein-bound species.
- Similarly to drug–CyD complexes, drug–protein binding interactions might also be "CD silent" resulting in no ICD signal despite of the formation of the complex [258,267,268].
- If a drug possesses two (or more), a high affinity and a low affinity CD inducer protein binding sites, it should be considered that due to the different local asymmetry, these sites might induce overlapping but different CD band patterns. In such cases, sign inversion, band shift or appearance of additional CEs can be observed during the titration procedure [269,270].
- For ICD measurements aimed to study ligand–ligand competition or allosteric interactions, the one participant must be optically transparent (e.g., valproate, fatty acids) [242] or its intrinsic/extrinsic CD band(s) must be separated from CE of the other ligand [258,271].
- Intrinsic CD profile of sterically restricted chiral molecules (e.g., steroids) might be little affected upon protein binding.

### *2.3.2 Drug–Nucleic Acid Interactions*

*2.3.2.1 Summary* Nucleic acids are obvious targets for drug intervention since modification of their structures clearly will have significant biological implications. The principal non-covalent binding modes of drugs to double-stranded nucleic acids

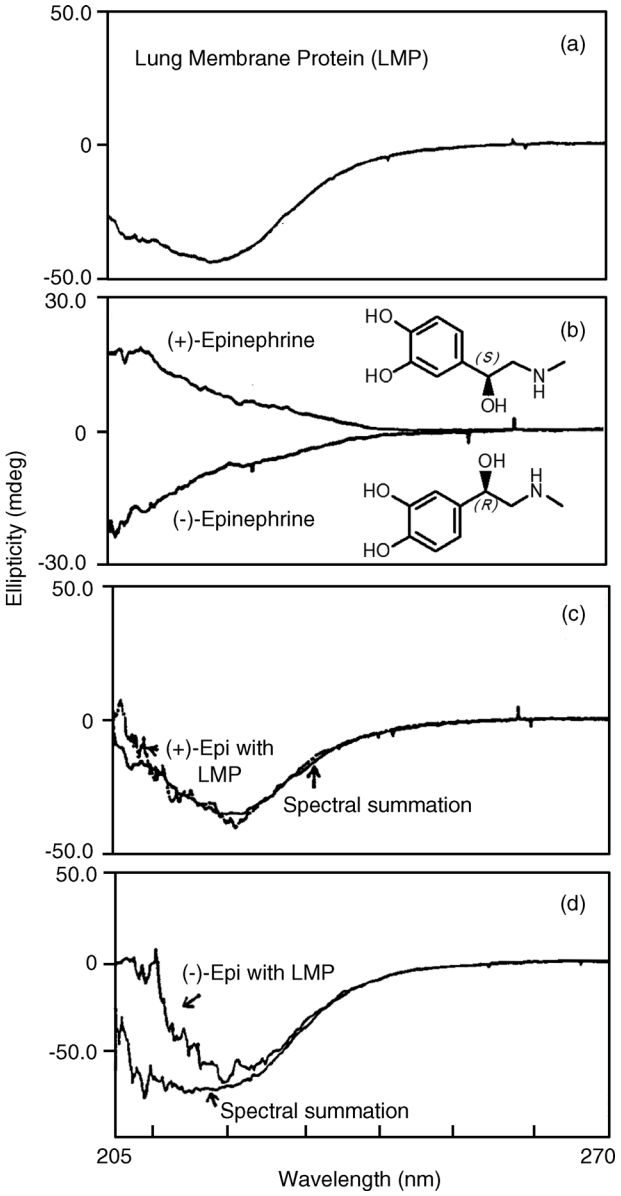


***FIGURE 7*** *CD spectra of (a) rat lung membrane protein (50 μg/mL) in phosphate buffer, (b) enantiomers of epinephrine (200 μM) in phosphate buffer, (c) rat lung membrane protein (50 μg/mL) in combination with (+)-epinephrine (200 μM). The spectral summation of data obtained in (a) and (b) is also provided for comparison. (d) Rat lung membrane protein (50 μg/mL) in combination with (−)-epinephrine (200 μM), along with spectral summation. Adapted and modified from Ref. [102].*

include groove binding (e.g., distamycin A, netropsin, Hoechst 33258) or intercalation (e.g., daunorubicin, doxorubicin) between the base pairs [272,273]. In single-stranded polynucleotides drugs can bind only by intercalation. Other agents, such as cisplatin, mitomycin C, and ecteinascidins, form covalent linkages with DNA, while duocarmycin, bleomycin/pepleomycin, and enediyne antibiotics cause DNA backbone cleavages [273]. Nucleic acid association of drugs may give rise to various kinds of ICD spectra depending on the mode of the binding and the value of the [drug]/[base pair] molar ratio. Intercalation (typically planar, aromatic molecules) and groove binding at low [drug]/[base pair] values result in a single, positive, or negative CE [274–277]. Upon increasing the [drug]/[base pair] ratio, the ICD pattern may become bisignate [275,278] displaying both positive and negative CEs within the principal absorption band of the drug. As more and more drug molecules bind and stack together in chiral orientation relative to each other in the minor or major groove of the duplex, the exciton interaction between their electronic transitions will be more and more dominant resulting characteristic biphasic ICD peaks. It is to be noted that compared to the intercalative mode, ICD signal of ligands bound in the minor groove is stronger by one order of magnitude [279,280].

Applications of drug–nucleic acid binding induced CD spectroscopic changes are as follows:

- Measuring ICD activity verifies the occurence of drug–nucleic acid binding interaction.
- Intercalative and non-intercalative binding modes can be distinguished by comparison of ICD spectra obtained under different experimental conditions (e.g., low/high [drug]/[base pair] ratios, various temperatures, ionic strength) [274,278].
- Nucleic acid binding mode of a ligand can also be studied by competition experiments: monitoring the effects of addition of a known minor/major groove binder or intercalator compound on ICD band of the ligand–nucleic acid sample helps to distinguish between the binding modes [274].
- From ICD data obtained by titrating a drug into a solution containing fixed concentration of a polynucleotide (or vice versa), the binding constant and the number of base pairs covered by a single drug molecule can be determined [110,278,281].
- Comparison of CD spectra of drugs obtained with synthetic polynucleotides with alternating (e.g., poly[d(A-T)]·poly[d(A-T)]) and homologous sequences (e.g., poly(dA) ·poly(dT)) informs on the sequence preference of the drug binding [282,283].
- Changes of the intrinsic CD pattern of polynucleotides elicited upon drug association might also be utilized for characterization of the binding interaction [283,284].

*2.3.2.2 Selected Examples* Comparison of changes in the intrinsic CD spectra of ofloxacin enantiomers measured in the presence of natural and synthetic

polynucleotides revealed enantioselective binding of the enantiomers in the minor groove of B-DNA [285,286].

Binding of the anti-cancer agent 9-hydroxyellipticine to calf thymus DNA and synthetic polynucleotides was studied by various biophysical techniques including CD spectroscopy [278]. The CD data showed that the drug adopts two main binding modes, intercalative and groove binding, which involves the formation of drug oligomers in the major groove. ICD values proportional to the concentration of the DNA-bound drug enabled to calculate the equilibrium binding constant.

CD spectroscopy was employed to study the dimerization of the topoisomerase I poison chiral topotecan and the increase of the dimerization process upon topotecan–DNA binding [287].

Alteration of the intrinsic CD spectra of a DNA dodecamer indicated the binding of the enediyne antibiotic calicheamicins which provokes marked conformational change of the B-form DNA. Binding constants were estimated based on analysis of CD spectral changes [284].

Addition of calf-thymus DNA to the solution of the optically inactive anti-neoplastic mitoxantrone gave rise to ICD bands above 450 nm [288]. Behavior of the CEs recorded at various [drug]/[base pair] ratios suggested intercalative binding mode. No ICD activity was seen with the Pd (II) complex of mitoxantrone indicating that the presence of the metal ion prevents intercalation of the drug between the base pairs.

Due to the presence of 11 stereocenters and several chromophores, the anti-cancer agent paclitaxel displays intrinsic CEs below 300 nm. Intensity changes and shift of these CEs observed in the presence of DNA suggested that paclitaxel molecules go from an aggregated state to a non-aggregated form upon binding with the double helix [289]. Comparison of CD spectra of mixtures of alternating and homologous synthetic polynucleotides with paclitaxel indicated AT sequence specific binding of the drug [282].

Nucleic acid association of ligands showing no light absorption above 220 nm can also be followed by measuring binding generated changes in the intrinsic CD profile of DNA. Though the lowest-energy UV transition of the natural lipid, oleic acid is below 220 nm, its binding to double but not to triple stranded oligo- and polydeoxyribonucleotides was clearly demonstrated by recording dramatic changes of DNA CD bands [283]. It was deduced from the spectroscopic results that oleic acid recognizes AT base pair motifs by minor groove binding.

#### *2.3.2.3 Comments*

- The ICD approach is unsuitable to identify specific nucleotide units or base pairs with which a given drug interacts since ICD data provide averaged structural information only.
- Modification of intrinsic CD spectra of chiral drugs can also be used for verifying drug–polynucleotide binding interactions [285,286]. However, it can occur that the strong intrinsic CD profile of sterically restricted chiral drugs remain unaltered despite of nucleic acid binding [290].

- For drugs showing light absorption in the 220–310 nm region, the measured CD spectrum will be the sum of intrinsic (nucleic acid) and extrinsic (bound drug) CD contributions. In such cases, it might be difficult to resolve the dominant nature of spectral changes since drug binding may alter the secondary structure of polynucleotides and this modified CD pattern can be mixed with the extrinsic CD signal of the drug [280,291–293]. This phenomenon is less important for proteins showing low-intensity near-UV CD activity compared to the much stronger CD bands of polynucleotides.

## 3 APPENDIX

Taking into consideration the time consuming compilation process of the Encyclopedia, update of the manuscript finalized in 2008 seemed to be necessary. The Appendix overviews the most relevant literature references published between 2008 and 2010 together with some previously overlooked works which are related to the applications of CD spectroscopy in the field of pharmaceutical sciences.

### 3.1 Stereochemical Characterization of Organic and Inorganic Compounds

Theoretical bases of CD spectroscopy and its applications in stereochemical studies of bioactive natural products with special emphasis on the CD exciton chirality method have been summarized [294].

### 3.2 Structural Characterization of Biopolymers

The most important applications of ECD and related CD techniques in biomolecular (proteins, peptides, nucleic acids) and nanostructural analysis (nanoparticles, nanocarriers) have been summarized [295]. Extended by a brief theoretical description, Schweitzer-Stenner *et al.* have provided an overview on how absorption and CD spectroscopy can be used to identify and characterize different conformational states of cytochrome *c* in solution [296].

### 3.3 Investigation of Biopolymer–Biopolymer Interactions

Applications of CD spectroscopy in structural studies of glycosaminoglycans and glycosaminoglycan-protein binding interactions have recently been reviewed [297].

***3.3.1 Enantiomeric Purity Determination Using HPLC-CD System*** Applying simultaneous UV and CD detections and *g* factor calculation, a simple, rapid and enantioselective method for the separation and determination of carnitine enantiomers has been developed for their quantitative analysis in dietary supplements and pharmaceutical formulations [298]. The method lacks of derivatization step and of a chiral support to separate the enantiomers and enables to calculate an e.e. of the potentially harmful D-carnitine.

The separation of benzoxathiepin diastereoisomers and their chemical impurities on an achiral sorbent has been followed by CD/UV detection. The determination of chemical purity was achieved with UV detection whereas the diastereoisomeric purity was determined by plotting the *g* factors versus the e.e. of each couple of enantiomers [299]. A mathematic model was developed to express the diastereoisomeric excess vs. the e.e. of each pair of enantiomers. Practical advantages and limitations of analytical assays based on HPLC-CD and CD/UV methods for screening of enantioselective enzyme-catalyzed reactions have been discussed [300].

## 3.4 Drug–Cyclodextrin Complexes

Alterations of far-UV CD spectrum of the organoantimonial drug, meglumine antimonate have been measured to investigate its association with β-CyD and the effects of heating and freeze-drying on the complex [301]. Modification of the intrinsic CD band of simvastatin upon its interaction with α-, β- and 2-HP-β-CyD has been used to study the inclusion geometry of the complexes [302].

## 3.5 Protein and Peptide Pharmaceuticals

The book entitled "Modern Techniques for Circular Dichroism and Synchrotron Radiation Circular Dichroism Spectroscopy" provides a comprehensive and up-to-date account of CD spectroscopy and its application to structural biology. A separate chapter discusses in detail the experimental aspects of CD spectroscopic characterization of biopharmaceuticals [303].

The conformational stability of the therapeutic fusion protein abatacept has been probed at various pH and temperature values by monitoring far-UV CD spectra [304].

Phenol used as a preservative in pharmaceutical insulin preparations binds to specific sites on the insulin hexamer and act as an allosteric effector. The binding isotherm constructed using CD data which were obtained by titration of the insulin hexamer with phenol allowed to calculate the dissociation constant [305].

In the course of developing a process for obtaining stable dry protein formulation of interferon α2a and immunoglobulin G, CD spectroscopy has been used to compare the secondary and tertiary structure of the reconstituted proteins with the structure of the native form [306, 307]. Intranasal formulations of the therapeutic peptide salmon calcitonin have been assessed by various methods. Far- and near-UV CD data supported structural similarity in secondary and tertiary structures of the peptides formulated in chlorobutanol and in benzalkonium chloride as preservatives [308].

'Solid-state' CD spectroscopy method has been proposed to study the surface-adsorbed conformation of therapeutic proteins following microencapsulation [309].

As has been determined by CD spectroscopy, secondary structures of chemically synthesized immunogenic anti-malarial peptide vaccines show certain characteristic structural configurations which can aid the easier design of additional immunogenic and protection-inducing peptides [310].

### 3.6 Gene Delivery Systems

A novel book chapter has been published focusing on the practical applications of CD, UV absorption, IR, and fluorescence spectroscopies for the structural analysis of the oligonucleotide component of nonviral gene delivery vectors [311].

### 3.7 Drug–Protein Binding

AAG binding interactions of several kinase inhibitors evaluated by using ICD spectra of the complexes have showed high-affinity protein association of most compounds [312]. Changes of the ICD bands of dicoumarol bound to the 'A' variant of AAG provoked by the anxiolytic agent deramciclane have proved that AAG can simultaneously accommodate both drug molecules [313]. Analysis of the ICD data showed that the presence of deramciclane enhanced by five-times of the AAG binding affinity of dicoumarol. AAG variant selective binding of nine basic drugs in clinical use has been examined by measuring their effects on extrinsic CEs of the 'F1/S' and 'A' variant selective CD label dicoumarol and acridine orange, respectively [314]. An interesting application of CD spectroscopy is its use to study mechanisms by which anesthesia occurs, i.e. the examination of the effects inhaled anesthetics (e.g., halothane, isoflurane) on protein CD spectra and to study the thermodynamic stability of proteins in the absence and presence of anesthetics to estimate the binding energetics [315–318].

### 3.8 Drug–Nucleic Acid Interactions

Theoretical and practical aspects of how circular and linear dichroism spectroscopy can be used to provide information about drug-DNA systems have been outlined [319]. Photogenotoxic potential of several pharmaceutical substances (e.g., chlorpromazine, furosemide, quinine) has been studied employing CD spectroscopy to analyse their DNA binding properties [320].

## 4 ACKNOWLEDGEMENT

This work was supported by research grant of OTKA K69213.

## REFERENCES

1. Poole RK, Kalnenieks U. Introduction to light absorption: visible and ultraviolet spectra. In: Gore MG, editor. Spectrophotometry and spectrofluorimetry: a practical approach. Oxford: Oxford University Press; 2000. p 1–32.
2. Sun SF. Physical chemistry of macromolecules: basic principles and issues. Hoboken: John Wiley & Sons, Inc.; 2004. p 399–412.
3. Görög S. Ultraviolet-visible spectrophotometry in pharmaceutical analysis. Boca Raton: CRC Press; 1995.

4. Gordon GH. Spectroscopy for the biological sciences. New Jersey: John Wiley & Sons Inc.; 2005. p 63–87.
5. Serdyuk IN, Zaccai NR, Zaccai J. Methods in molecular biophysics: structure, dynamics, function. Cambridge: Cambridge University Press; 2007. p 601–624.
6. Salvadori P, Di Bari L, Rosini C. Electronic circular dichroism – fundamentals, methods and applications. In: Lough WJ, Wainer IW, editors. Chirality in natural and applied science. Oxford: Blackwell Science Ltd.; 2002. p 203–239.
7. Gawronski J, Skowronek P. Electronic circular dichroism for chiral analysis. In: Busch KW, Busch MA, editors. Chiral analysis. Amsterdam: Elsevier Science; 2006. p 397–459.
8. Urbanová M, Maloň P. Circular dichroism spectroscopy. In: Schalley CA, editor. Analytical methods in supramolecular chemistry. Weinheim: Wiley-VCH Verlag GmbH & Co.; 2007. p 265–304.
9. Berova N, Nakanishi K, Woody RW. Circular dichroism: principles and applications. New York: Wiley-VCH, Inc.; 2000.
10. Fasman GD. Circular dichroism and the conformational analysis of biomolecules. Plenum Press: New York; 1996.
11. Martin SR, Schilstra MJ. Circular dichroism and its application to the study of biomolecules. In: Correia JJ, Detrich HW, editors. Biophysical tools for biologists, volume 1: *in vitro* techniques. San Diego: Academic Press; 2008. p 263–293.
12. Harada N, Nakanishi K. Circular dichroic spectroscopy: exciton coupling in organic stereochemistry. Mill Valley: University Science Books; 1983.
13. Lightner DA, Gurst JE. Organic conformational analysis and stereochemistry from circular dichroism. New York: John Wiley and Sons, Inc.; 2000.
14. Purdie N, Brittain HG. Analytical application of circular dichroism. Amsterdam: Elsevier; 1994.
15. Pančoška P. Circular dichroism in analysis of biomolecules. In: Meyers RA, editor. Encyclopedia of analytical chemistry. Chichester: John Wiley & Sons Ltd.; 2000. http://dx.doi.org/10.1002/9780470027318.a0203.
16. Berova N, Di Bari L, Pescitelli G. Application of electronic circular dichroism in configurational and conformational analysis of organic compounds. Chem Soc Rev 2007; 36: 914–931.
17. Kelly SM, Jess TJ, Price NC. How to study proteins by circular dichroism. Biochim Biophys Acta 2005; 1751: 119–139.
18. Superchi S, Giorgio E, Rosini C. Structural determinations by circular dichroism spectra analysis using coupled oscillator methods: an update of the applications of the DeVoe polarizability model. Chirality 2004; 16: 422–451.
19. Gawronski J, Skowronek P. New chromophores for organic stereochemical analysis by exciton-coupled circular dichroism. Curr Org Chem 2004; 8: 65–82.
20. Allenmark S. Induced circular dichroism by chiral molecular interaction. Chirality 2003; 15: 409–422.
21. Zsila F, Bikádi Z, Fitos I, Simonyi M. Probing protein binding sites by circular dichroism spectroscopy. Curr Drug Discov Technol 2004; 1: 133–153.
22. Schellman JA. Circular dichroism and optical rotation. Chem Rev 1975; 75: 323–331.
23. Norden B, Kurucsev T. Analysing DNA complexes by circular and linear dichroism. J Mol Recognit 1994; 7: 141–155.

24. Snatzke G. Circular dichroism and absolute conformation—application of qualitative MO theory to chiroptical phenomena. Angew Chem Int Ed Engl 1979; 18: 363–377.
25. Brittain HG. Chiroptical spectroscopic studies of pharmaceutical compounds. Microchem J 1997; 57: 137–148.
26. Purdie N, Swallows KA. Circular dichroism—I. Theory and practice. Trends Anal Chem 1990; 9: 94–97.
27. Perrin JH, Hart PA. Small molecule-macromolecular interactions as studied by optical rotatory dispersion—circular dichroism. J Pharm Sci 1970; 59: 431–448.
28. Brittain HG. Applications of chiroptical spectroscopy for the characterization of pharmaceutical compounds. J Pharm Biomed Anal 1998; 17: 933–940.
29. Johnson WC Jr. Circular dichroism instrumentation. In: Fasman GD, editor. Circular dichroism and the conformational analysis of biomolecules. New York: Plenum Press; 1996. p 635–652.
30. Gore MG. Spectrophotometry and spectrofluorimetry: a practical approach. Oxford: Oxford University Press; 2000. p 99–139.
31. Harada A, Hashidzume A, Takashima Y. Cyclodextrin-based supramolecular polymers. Adv Polym Sci 2006; 201: 1–43.
32. Fujiki M, Nakashima H, Toyoda S, Koe JR. Chirality in the polysilanes. In: Green MM, Nolte RJM, Meijer EW, editors. Materials-chirality: volume 24 of Topics in stereochemistry. Hoboken: John Wiley & Sons, Inc.; 2003. p 209–280.
33. Al-Basheer W, Pagni RM, Compton RN. Spectroscopic and theoretical investigation of (*R*)-3-methylcyclopentanone. The effect of solvent and temperature on the distribution of conformers. J Phys Chem A 2007; 111: 2293–2298.
34. Kuroda R, Honma T. CD spectra of solid-state samples. Chirality 2000; 12: 269–277.
35. McPhie P. Circular dichroism studies on proteins in films and in solution: estimation of secondary structure by g-factor analysis. Anal Biochem 2001; 293: 109–119.
36. Calero M, Gasset M. Fourier transform infrared and circular dichroism spectroscopies for amyloid studies. In: Sigurdsson EM, editor. Methods in molecular biology, volume 299, Amyloid proteins: methods and protocols. Totowa: Humana Press Inc.; 2005. p 129–151.
37. Greenfield NJ. Using circular dichroism spectra to estimate protein secondary structure. Nat Protoc 2006; 1: 2876–2890.
38. Sauer K. Why spectroscopy? Which spectroscopy? Methods Enzymol 1995; 246: 1–10.
39. Greenfield NJ. Using circular dichroism collected as a function of temperature to determine the thermodynamics of protein unfolding and binding interactions. Nat Protoc 2006; 1: 2527–2535.
40. Kelly SM, Price NC. Circular dichroism to study protein interactions. In: Coligan JE, Dunn BM, Speicher DW, Wingfield PT, editors. Current protocols in protein science, supplement 46. New York: John Wiley and Sons, Inc.; 2006. Unit 20.10 http://dx.doi.org/10.1002/0471140864.ps2010s4.
41. Zsila F, Gergely A, Szász G. Conformational study on ketamine by circular dichroism and ultraviolet spectroscopy. Chirality 1999; 11: 280–285.
42. Fitos I, Visy J, Zsila F, Bikádi Z, Mády G, Simonyi M. Specific ligand binding on genetic variants of human alpha1-acid glycoprotein studied by circular dichroism spectroscopy. Biochem Pharmacol 2004; 67: 679–688.

43. Purdie N, Swallows KA. Direct determination of water soluble vitamins by circular dichroism. J Agric Food Chem 1991; 39: 2171–2175.
44. Purdie N, Swallows KA, Murphy LH, Purdie RB. Analytical applications of circular dichroism. J Pharm Biomed Anal 1989; 7: 1519–1526.
45. Purdie N, Swallows KA, Murphy LH, Purdie RB. Circular dichroism—II. Analytical applications. Trends Anal Chem 1990; 9: 136–142.
46. Rodger A, Marrington R, Roper D, Windsor S. Circular dichroism spectroscopy for the study of protein–ligand interactions. In: Nienhaus GU editor. Methods in molecular biology, volume 305, Protein–ligand interactions: methods and applications. Totowa: Humana Press Inc.; 2005. p 343–363.
47. Cary PD, Kneale GG. Circular dichroism for the analysis of protein–DNA interactions. In: Moss T, Leblanc B, editors. Methods in molecular biology, volume 543, DNA-protein interactions: principles and protocols. 3rd ed. Totowa: Humana Press Inc.; 2009. p 613–624.
48. Reetz MT, Kuhling KM, Hinrichs H, Deege A. Circular dichroism as a detection method in the screening of enantioselective catalysts. Chirality 2000; 12: 479–482.
49. Greenfield NJ. Analysis of the kinetics of folding of proteins and peptides using circular dichroism. Nat Protoc 2006; 1: 2891–2899.
50. Kuwajima K. Stopped-flow circular dichroism. In: Fasman GD, editor. Circular dichroism and the conformational analysis of biomolecules. New York: Plenum Press; 1996. p 159–182.
51. Wenzel S, Buss V. Circular dichroism and electronic structure calculations on naproxen. J Phys Org Chem 1992; 5: 748–754.
52. Craig JC, Labelle B, Ohnsorge U. The absolute configuration of the enantiomers of 6-chloro-9-(4′-diethylamino-1′-methylbutyl)-amino-2-methoxyacridine (quinacrine). Chirality 1991; 3: 436–437.
53. Verdino P. Structural characterization of pollen allergens. Clin Rev Allergy Immunol 2006; 30: 73–95.
54. Kelly SM, Price NC. The use of circular dichroism in the investigation of protein structure and function. Curr Protein Pept Sci 2000; 1: 349–384.
55. Welch CJ, Biba M, Sajonz P. Fast methods of enantiopurity determination for the Soai reaction: towards a general enantioenrichment detector? Chirality 2007; 19: 34–43.
56. Wallace BA, Teeters CL. Differential absorption flattening optical effects are significant in the circular dichroism spectra of large membrane fragments. Biochemistry 1987; 26: 65–70.
57. Braun CS, Kueltzo LA, Middaugh CR. Ultraviolet absorption and circular dichroism spectroscopy of nonviral gene delivery complexes. In: Findeis MA, editor. Methods in molecular medicine, volume 65, Nonviral vectors for gene therapy: methods and protocols. Totowa: Humana Press Inc.; 2001. p 253–284.
58. Kuroda Y, Maeda Y, Sawa S, Shibata K, Miyamoto K, Nakagawa T. Effects of detergents on the secondary structures of prion protein peptides as studied by CD spectroscopy. J Pept Sci 2003; 9: 212–220.
59. Noordermeer MA, Veldink GA, Vliegenthart JF. Spectroscopic studies on the active site of hydroperoxide lyase; the influence of detergents on its conformation. FEBS Lett 2001; 489: 229–232.

60. Mandal A, Das S, Chakraborti T, Kar P, Ghosh B, Chakraborti S. Solubilization, purification and reconstitution of $Ca^{2+}$-ATPase from bovine pulmonary artery smooth muscle microsomes by different detergents: preservation of native structure and function of the enzyme by DHPC. Biochim Biophys Acta 2006; 1760: 20–31.
61. Sandström J. Chiroptical methods in conformational analysis. Chirality 1995; 7: 181–192.
62. Berova N, Borhan B, Dong JG, Guo JS, Huang XF, Karnaukhova E, Kawamura A, Leu JH, Matile S, Nakanishi K, Rickman B, Su J, Tan Q, Zanze I. Solving challenging bioorganic problems by exciton coupled CD. Pure Appl Chem 1998; 70: 377–383.
63. Fiallo MML, Tayeb H, Suarato A, Garnier-Suillerot A. Circular dichroism studies on anthracycline antitumor compounds. Relationship between the molecular structure and the spectroscopic data. J Pharm Sci 1998; 87: 967–975.
64. Mitscher LA, Bathala MS, Sokoloski TS. Spectropolarimetry of antibiotics. Methods Enzymol 1975; 43: 346–373.
65. Gortazar P, Vazquez JT. Discrimination and direct determination of cephalosporins by circular dichroism. J Pharm Sci 1994; 83: 1204–1208.
66. Miller RF, Sokoloski TD, Mitscher LA, Bonacci AC, Hoener BA. Use of circular dichroism in analysis of mixtures of tetracycline and 4-epitetracycline and its application to assay of commercial products. J Pharm Sci 1973; 62: 1143–1147.
67. Tongaree S, Flanagan DR, Poust RI. The interaction between oxytetracycline and divalent metal ions in aqueous and mixed solvent systems. Pharm Develop Technol 1999; 4: 581–591.
68. Franceschi G, Bedeschi A, Rizzo V, Vigevani A, Oberti R. Structural studies of ritipenem acoxil (FCE 22891)—X-ray crystal structure and chiroptical properties. Bioorg Med Chem Lett 1993; 3: 2333–2336.
69. Knabe J, Wunn W. Chiroptical investigations on optically active barbiturates, hydantoins and pyrazolidinediones. Arch Pharm 1982; 315: 977–989.
70. Konowal A, Snatzke G, Alebic-Kolbah T, Kajfez F, Rendic S, Sunjic V. General approach to chiroptical characterization of binding of prochiral and chiral 1,4-benzodiazepin-2-ones to human serum albumin. Biochem Pharmacol 1979; 28: 3109–3113.
71. Snatzke G, Konowal A, Sabljic A, Blazevic N, Sunjic V. Circular dichroism of optically active 1,4-benzodiazepines. Croat Chem Acta 1982; 55: 435–455.
72. Kotovych G, Cann JR, Stewart JM, Yamamoto H. NMR and CD conformational studies of bradykinin and its agonists and antagonists: application to receptor binding. Biochem Cell Biol 1998; 76: 257–266.
73. Ota A, Kawashima Y, Ohishi H, Ishida T. Conformational studies of semotiadil (SD-3211), a novel $Ca^{2+}$ antagonist. Chem Pharm Bull (Tokyo) 1993; 41: 1681–1685.
74. Gawronski J, Brzostowska M, Koput J. A circular dichroism study of molecular association of cinchona alkaloids and carboxylic acids. Croat Chem Acta 1989; 62: 97–107.
75. Bognár R, Gaál G, Kerekes P, Lévai A, Makleit S, Snatzke F, Snatzke G. Chiroptical properties of morphinan derivatives. Collect Czech Chem Commun 1975; 40: 670–680.
76. Hrbek J, Hruban L, Simanek V, Santavy F, Snatzke G, Yemul SS. Circular dichroism of alkaloids of colchicine type and their derivatives. Collect Czech Chem Commun 1982; 47: 2258–2279.
77. Pyles EA, Rava RP, Hastie SB. Effect of B-ring substituents on absorption and circular dichroic spectra of colchicine analogs. Biochemistry 1992; 31: 2034–2039.

78. Faupel P, Buss V. Evidence for a folded conformation of methotrexate in solution. Angew Chem Int Ed Engl 1988; 27: 423–424.
79. Zandomeneghi M. Circular dichroism of ketoprofen complexed to serum albumins: conformational selection by the protein: a novel optical purity determination technique. Chirality 1995; 7: 446–451.
80. Tan SC, Baker JA, Stevens N, DeBiasi V, Salter C, Chalaux M, Afarinkia K, Hutt AJ. Synthesis, chromatographic resolution and chiroptical properties of carboxyibuprofen stereoisomers: Major metabolites of ibuprofen in man. Chirality 1997; 9: 75–87.
81. Testa B, Jenner P. Circular dichroic determination of the preferred conformation of nicotine and related chiral alkaloids in aqueous solution. Mol Pharmacol 1973; 9: 10–16.
82. Dong JG, Guo J, Akritopoulou-Zanze I, Kawamura A, Nakanishi K, Berova N. Theoretical and experimental CD of conformationally flexible complex molecules-application to ouabain pentanaphthoate and analogs. Chirality 1999; 11: 707–721.
83. De B, Andersen NH, Ippolito RM, Wilson CH, Johnson WD. Synthesis and chiroptical characterization of prostacyclin diastereomers. Prostaglandins 1980; 19: 221–247.
84. Harders H, Forster S, Voelter W, Bacher A. Problems in electronic state assignment based on circular dichroism. Optical activity of flavines and 8-substituted lumazines. Biochemistry 1974; 13: 3360–3364.
85. Li MP, Birmingham MK, Chan TH. 18-Hydroxy-11-deoxycorticosterone. Chemical synthesis, structure, and circular dichroism. J Org Chem 1976; 41: 2552–2556.
86. Wellman KM, Briggs WS, Djerassi C. Optical rotatory dispersion studies. C. Variable-temperature circular dichroism studies of ring-conformational and rotational equilibria in cyclohexanones. J Am Chem Soc 1965; 87: 73–81.
87. Snow JW, Hooker TM. The chiroptical properties of strychnine alkaloids: strychnine, beta-colubrine, brucine, and their dihydro derivatives. Can J Chem 1978; 56: 1222–1230.
88. Balasubramanian SV, Alderfer JL, Straubinger RM. Solvent- and concentration-dependent molecular interactions of taxol (Paclitaxel). J Pharm Sci 1994; 83: 1470–1476.
89. Zsila F, Hollósi M, Gergely A. Stereochemical study of tolperisone, a muscle relaxant agent, by circular dichroism and ultraviolet spectroscopy. Chirality 2000; 12: 720–726.
90. Cort JH, Fric I, Carlsson L, Gillessen D, Bystricky S, Skopkova J, Gut V, Studer RO, Mulder JL, Blaha K. Biological and chiroptical sequelae of graded alkyl substitutions in vasopressin ring. Mol Pharmacol 1976; 12: 313–321.
91. Dong JG, Bornmann W, Nakanishi K, Berova N. Structural studies of vinblastine alkaloids by exciton coupled circular dichroism. Phytochemistry 1995; 40: 1821–1824.
92. Szabó L, Kolonits P, Kalaus G, Szántay C, Kálmán A, Párkányi L, Kajtár M, Ángyán, J. Synthesis of Vinca alkaloids and related compounds. Part 18. Stereochemical investigations on some intermediates leading to (+)-vincamine. J Chem Soc-Perkin Trans 1984; 2: 1629–1636.
93. Parish CA, Dong JG, Bornmann WG, Chang J, Nakanishi K, Berova N. Circular dichroism studies of bisindole *Vinca* alkaloids. Tetrahedron 1998; 54: 15739–15758.
94. Firth RA, Hill HAO, Pratt JM, Williams RJ, Jackson WR. The circular dichroism and absorption spectra of some vitamin $B_{12}$ derivatives. Biochemistry 1967; 6: 2178–2189.
95. Heimark LD, Trager WF. The preferred solution conformation of warfarin at the active site of cytochrome P-450 based on the CD spectra in octanol/water model system. J Med Chem 1984; 27: 1092–1095.

96. Woody RW. Circular dichroism of protein-folding intermediates. Methods Enzymol 2004; 380: 242–285.
97. Adler AJ, Greenfield NJ, Fasman GD. Circular dichroism and optical rotatory dispersion of proteins and polypeptides. Methods Enzymol 1973; 27: 675–735.
98. Pain R. Determining the CD spectrum of a protein. In: Coligan JE, Dunn BM, Speicher DW, Wingfield PT, editors. Current protocols in protein science, supplement 38. New York: John Wiley and Sons, Inc.; 2004. Unit 7.6 http://dx.doi.org/10.1002/0471140864.ps0706s38.
99. Gray DM, Ratliff RL, Vaughan MR. Circular dichroism spectroscopy of DNA. Methods Enzymol 1992; 211: 389–406.
100. Greenfield NJ. Circular dichroism analysis for protein–protein interactions. In: Fu H, editor. Methods in molecular biology, volume 261, Protein–protein interactions: methods and protocols. Totowa: Humana Press Inc.; 2004. p 55–77.
101. Stone AL, Beeler D, Oosta G, Rosenberg RD. Circular dichroism spectroscopy of heparin-antithrombin interactions. Proc Natl Acad Sci U S A 1982; 79: 7190–7194.
102. Patil PN, Eraundorfer P, Dutta PK. Stereoselective modification of circular dichroism spectra of rat lung beta-adrenoceptor protein preparation by enantiomers of epinephrine. Chirality 1996; 8: 463–465.
103. Ulmer DD, Vallee BL. Extrinsic Cotton effects and the mechanism of enzyme action. Adv Enzymol Relat Areas Mol Biol 1965; 27: 37–104.
104. Edwards RA, Woody RW. Molecular orbital calculations of the optical properties of Congo Red and Cibacron Blue and their complexes with proteins. J Phys Chem 1983; 87: 1329–1337.
105. Reddy AV, Behnke WD, Freisheim JH. Binary and ternary complexes of dihydrofolate reductase with substrates, coenzymes and inhibitors. Circular dichroic and magnetic circular dichroic studies. Biochim Biophys Acta 1978; 533: 415–427.
106. Ascoli GA, Domenici E, Bertucci C. Drug binding to human serum albumin: abridged review of results obtained with high-performance liquid chromatography and circular dichroism. Chirality 2006; 18: 667–679.
107. Roosdorp N, Sjoholm I. Determination of binding constants from continuous circular dichroism titration data by numerical analysis. Biochem Pharmacol 1976; 25: 2141–2145.
108. Vandev D, Prodanova K, Russeva V. New mathematical procedure for analysis of data obtained by means of circular dichroism titration method. Arzneimittelforschung 1998; 48: 1190–1193.
109. Eriksson M, Norden B. Linear and circular dichroism of drug-nucleic acid complexes. Methods Enzymol 2001; 340: 68–98.
110. Stootman FH, Fisher DM, Rodger A, Aldrich-Wright JR. Improved curve fitting procedures to determine equilibrium binding constants. Analyst 2006; 131: 1145–1151.
111. Zsila F, Gedeon G. Binding of anti-prion agents to glycosaminoglycans: evidence from electronic absorption and circular dichroism spectroscopy. Biochem Biophys Res Commun 2006; 346: 1267–1274.
112. Gergely A. Polarimetry, ORD and CD spectroscopy. In: Görög S, editor. Progress in pharmaceutical and biomedical analysis, vol. 4. Amsterdam: Elsevier Science; 2000. p 553–561.

113. Purdie N, Swallows KA. Direct determination of beta-lactam antibiotics by circular dichroism. Anal Chem 1987; 59: 1349–1351.
114. Bowen JM, Crone TA, Kennedy RK, Purdie N. Determination of heroin by circular dichroism. Anal Chem 1982; 54: 66–68.
115. Atkinson WM, Bowen JM, Purdie N. Determination of "Brompton's cocktails" by circular dichroism. J Pharm Sci 1984; 73: 1827–1828.
116. Han SM, Purdie N. Simultaneous determination of opiates by circular dichroism. Anal Chem 1986; 58: 113–116.
117. Gortazar P, Roen A, Vazquez JT. Determination of drug levels in human serum by circular dichroism. Chirality 1998; 10: 507–512.
118. Gortazar P, Ravina M, Vazquez JT. Direct quantitative determination of optically active absorbing drugs in human urine by circular dichroism—simultaneous direct determination of beta-lactam antibiotics and Proteins. J Pharm Sci 1995; 84: 1316–1321.
119. Bowen JM, Purdie N. The direct analysis of tetracycline in urine by circular dichroism spectropolarimetry. J Pharm Sci 1982; 71: 836–837.
120. Han SM, Purdie N. Solute-induced circular dichroism: complexation of achiral drugs with cyclodextrin. Anal Chem 1984; 56: 2822–2825.
121. Han SM, Atkinson WM, Purdie N. Solute-induced circular dichroism: drug discrimination by cyclodextrin. Anal Chem 1984; 56: 2827–2830.
122. Bowen JM, Purdie N. Identification of cocaine and phencyclidine by solute-induced circular dichroism. Anal Chem 1981; 53: 2239–2242.
123. Han SM, Purdie N, Swallows KA. Determination of the benzodiazepin-2-ones by circular dichroism. Anal Chim Acta 1987; 197: 57–64.
124. Han SM, Purdie N. Cyclodextrin complexation of barbiturates in aqueous solution. Anal. Chem. 1984; 56: 2825–2827.
125. Gergely A, Szász G, Soós J. Determination of delta-4-3-ketosteroid contamination in drug base substances by CD spectroscopy. 9. Application of chiroptical methods in drug analysis. Fresenius J Anal Chem 1984; 318: 528–530.
126. Szegvári D, Horváth P, Gergely A, Németh S, Görög S. CD spectrometric methods for the simultaneous determination of ethisterone and its delta(5)-isomer. Anal Bioanal Chem 2003; 375: 713–717.
127. Szentesi A, Gergely A, Horváth P, Hosztafi S, Szász G. Determination of 6-oxomorphinans, as the oximes, by difference circular dichroism spectroscopy. Anal Bioanal Chem 2002; 374: 427–431.
128. Szentesi A, Gergely A, Horváth P, Szász G. Determination of Δ(4)-3-ketosteroids based on oxime formation by difference circular dichroism spectroscopy. Fresenius J Anal Chem 2000; 368: 384–388.
129. Erskine SR, Quencer BM, Beebe KR. Rapid optical purity determination using chiral spectroscopy, achiral spectroscopy, and multivariate analysis. Appl Spectrosc 1995; 49: 1682–1691.
130. Horváth P, Gergely A, Noszál B. Determination of enantiomeric purity by simultaneous dual circular dichroism and ultraviolet spectroscopy. Talanta 1997; 44: 1479–1485.
131. Kitagawa T, Oie S, Taniyama H. Analysis of *d*-ethambutol by circular dichroism of its copper-chelates. Chem Pharm Bull (Tokyo) 1976; 24: 3019–3024.

132. Engle AR, Purdie N. Determination of enantiomeric purities using CD/CD detection. Anal Chim Acta 1994; 298: 175–182.

133. Zaugg S, Caslavska J, Theurillat R, Thormann W. Characterization of the stereoselective metabolism of thiopental and its metabolite pentobarbital via analysis of their enantiomers in human plasma by capillary electrophoresis. J Chromatogr A 1999; 838: 237–249.

134. Hoffmann P, Wagner H, Weber G, Lanz M, Caslavska J, Thormann W. Separation and purification of methadone enantiomers by continuous- and interval-flow electrophoresis. Anal Chem 1999; 71: 1840–1850.

135. Hegedüs H, Gergely A, Veress T, Zsila F. Determination of the enantiomeric purity of amphetamine after derivatization with Sanger's reagent [2,4-dinitrofluorobenzene [DNFB]) by simultaneous dual circular dichroism and ultraviolet spectroscopy. Fresenius J Anal Chem 1999; 364: 749–752.

136. Lorin M, Delepee R, Ribet JP, Morin P. Validation of a method using an achiral liquid chromatography sorbent and a circular dichroism detector—analysis of the efaroxan enantiomers. J Chromatogr A 2007; 1141: 244–250.

137. Salvadori P, Bertucci C, Rosini C. Circular dichroism detection in HPLC. Chirality 1991; 3: 376–385.

138. Zukowski J, Tang YB, Berthod A, Armstrong DW. Investigation of a circular dichroism spectrophotometer as a liquid chromatography detector for enantiomers: sensitivity, advantages and limitations. Anal Chim Acta 1992; 258: 83–92.

139. Mannschreck A. On-line measurement of circular dichroism spectra during enantioselective liquid chromatography. Trends Anal Chem 1993; 12: 220–225.

140. Yamamoto A, Kodama S, Matsunaga A, Hayakawa K, Kitaoka M. Multi-beam circular dichroism detector for HPLC. Analyst 1999; 124: 483–485.

141. Bobbitt DR, Linder SW. Recent advances in chiral detection for high performance liquid chromatography. Trends Anal Chem 2001; 20: 111–123.

142. Kott L, Holzheuer WB, Wong MM, Webster GK. An evaluation of four commercial HPLC chiral detectors: a comparison of three polarimeters and a circular dichroism detector. J Pharm Biomed Anal 2007; 43: 57–65.

143. Mistry N, Roberts AD, Tranter GE, Francis P, Barylski I, Ismail IM, Nicholson JK, Lindon JC. Directly coupled chiral HPLC-NMR and HPLC-CD spectroscopy as complementary methods for structural and enantiomeric isomer identification: application to atracurium besylate. Anal Chem 1999; 71: 2838–2843.

144. Takahashi H, Kashima T, Kimura S, Muramoto N, Nakahata H, Kubo S, Shimoyama Y, Kajiwara M, Echizen H. Determination of unbound warfarin enantiomers in human plasma and 7-hydroxywarfarin in human urine by chiral stationary-phase liquid chromatography with ultraviolet or fluorescence and on-line circular dichroism detection. J Chromatogr B 1997; 701: 71–80.

145. Salvadori P, Bertucci C, Ascoli G, UccelloBarretta G, Rossi E. Direct resolution, characterization, and stereospecific binding properties of an atropisomeric 1,4-benzodiazepine. Chirality 1997; 9: 495–505.

146. Kanazawa H, Kunito Y, Matsushima Y, Okubo S, Mashige F. Stereospecific analysis of lorazepam in plasma by chiral column chromatography with a circular dichroism-based detector. J Chromatogr A 2000; 871: 181–188.

147. Zsila F, Gergely A, Horvath P, Szász G. Separation and identification of tofisopam stereoisomers by hyphenated HPLC-CD technique. J Liq Chromatogr Relat Technol 1999; 22: 713–719.

148. Williams RC, Edwards JF, Joshi AS, Aubry AF. Chiral analysis of drug substance in clinical plasma extracts using achiral HPLC with circular dichroism detection. J Pharm Biomed Anal 2001; 25: 501–509.

149. Slijkhuis C, Hartog KD, van Alphen C, Blok-Tip L, Jongen PM, de Kaste D. Analysis of optically active compounds using conventional chromatography with a circular dichroism detector. J Pharm Biomed Anal 2003; 32: 905–912.

150. Luykx DM, Goerdayal SS, Dingemanse PJ, Jiskoot W, Jongen PM. HPLC and tandem detection to monitor conformational properties of biopharmaceuticals. J Chromatogr B 2005; 821: 45–52.

151. Gergely A, Szász G, Szentesi A, Gyimesi-Forras K, Kökösi J, Szegvári D, Veress G. Evaluation of CD detection in an HPLC system for analysis of DHEA and related steroids. Anal Bioanal Chem 2006; 384: 1506–1510.

152. Gergely A, Horváth P, Noszál B. Determination of peak homogeneity by dual detection. Anal Chem 1999; 71: 1500–1503.

153. Horváth P, Gergely A, Noszál B. Characterisation of peak purity by fast multiscan circular dichroism detection. Chromatographia 2000; 48: 584–588.

154. Blessington B, Beiraghi A, Lo TW, Drake A, Jonas G. Chiral HPLC-CD studies of the antituberculosis drug (+)-ethambutol. Chirality 1992; 4: 227–229.

155. Di Bari L, Pescitelli G, Salvadori P, Rovini M, Anzini M, Cappelli A, Vomero S. Synthesis, resolution, and absolute configuration of two novel and selective cyclooxygenase-2 inhibitors based on the 1,5-diarylpyrrole structure. Tetrahedron Asymmetry 2006; 17: 3430–3436.

156. Bertucci C, Andrisano V, Cavrini V, Castiglioni E. Reliable assay of extreme enantiomeric purity values by a new circular dichroism based HPLC detection system. Chirality 2000; 12: 84–92.

157. Bossu E, Cotichini V, Gostoli G, Farina A. Determination of optical purity by nonenantioselective LC using CD detection. J Pharm Biomed Anal 2001; 26: 837–848.

158. Yamamoto A, Akiba N, Kodama S, Matsunaga A, Kato K, Nakazawa H. Enantiomeric purity determination of malic acid in apple juices by multi-beam circular dichroism detection. J Chromatogr A 2001; 928: 139–144.

159. Song S, Sun L, Yuan L, Sun T, Zhao Y, Zuo W, Cong Y, Li X, Wang J. Method to determine enantiomeric excess of glucose by nonchiral high-performance liquid chromatography using circular dichroism detection. J Chromatogr A 2008; 1179: 125–130.

160. Holik M, Mannschreck A. Determining enantiomeric excess from overlapping HPLC peaks: discussion of errors in the methods. Chemom Intell Lab Syst 2004; 72: 153–160.

161. Gergely A, Horváth P, Noszál B. Deconvolution of composite chromatographic peaks by simultaneous dual detections. J Chromatogr Sci 2000; 38: 425–429.

162. Krois D, Brinker UH. Circular dichroism of cyclodextrin complexes. In: Dodziuk H, editor. Cyclodextrins and their complexes: chemistry, analytical methods, applications. Weinheim: WILEY-VCH Verlag GmbH & Co. KGaA; 2006. p 289–298.

163. Zhdanov YA, Alekseev YE, Kompantseva EV, Vergeychik EN. Induced optical activity in cyclodextrin complexes. Russian Chem Rev 1992; 61: 563–575.

164. Blanco M, Coello J, Iturriaga H, Maspoch S, Perez-Maseda C. Circular dichroism spectra of cyclodextrins-ketoprofen inclusion complexes—determination of enantiomeric purity. Anal Chim Acta 2000; 407: 233–245.

165. Ikeda Y, Hirayama F, Arima H, Uekama K, Yoshitake Y, Harano K. NMR spectroscopic characterization of metoprolol/cyclodextrin complexes in aqueous solution: cavity size dependency. J Pharm Sci 2004; 93: 1659–1671.

166. Denadai AM, Ianzer D, Alcantara AF, Santoro MM, Santos CF, Lula IS, de Camargo AC, Faljoni-Alario A, dos Santos RA, Sinisterra RD. Novel pharmaceutical composition of bradykinin potentiating penta peptide with beta-cyclodextrin: physical-chemical characterization and anti-hypertensive evaluation. Int J Pharm 2007; 336: 90–98.

167. Ikeda K, Uekama K, Otagiri M. Inclusion complexes of beta-cyclodextrin with antiinflammatory drugs fenamates in aqueous solution. Chem Pharm Bull (Tokyo) 1975; 23: 201–208.

168. Uekama K, Hirayama F, Otagiri M, Otagiri Y, Ikeda K. Inclusion complexation of beta-cyclodextrin with some sulfonamides in aqueous solution. Chem Pharm Bull (Tokyo) 1978; 26: 1162–1167.

169. Mura P, Furlanetto S, Cirri M, Maestrelli F, Corti G, Pinzauti S. Interaction of naproxen with ionic cyclodextrins in aqueous solution and in the solid state. J Pharm Biomed Anal 2005; 37: 987–994.

170. Suzuki M, Kajtár M, Szejtli J, Vikmon M, Fenyvesi E. Induced circular dichroism spectra of complexes of cyclomaltooligosaccharides and azo dyes containing naphthalene nuclei. Carbohydr Res 1992; 223: 71–80.

171. Suzuki M, Kajtár M, Szejtli J, Vikmon M, Fenyvesi E, Szente L. Induced circular dichroism spectra of complexes of cyclomaltooligosaccharides and azobenzene derivatives. Carbohydr Res 1991; 214: 25–33.

172. Kobayashi N, Hino Y, Ueno A, Osa T. Induced circular-dichroism of 2 chromophores included in gamma-cyclodextrin. Bull Chem Soc Jan 1983; 56: 1849–1850.

173. Buss V. Circular dichroism of a chiral cyanine dye dimer trapped in gamma-cyclodextrin. Angew Chem Int Ed Engl 1991; 30: 869–870.

174. Marconi G, Monti S, Manoli F, Ottani S. Circular dichroism and theoretical studies on the inclusion of the antimalarial drug licochalcone-A in beta-cyclodextrin. J Incl Phenom Macrocycl Chem 2007; 57: 279–282.

175. Otagiri M, Fokkens JG, Hardee GE, Perrin JH. Interaction of some coumarin anticoagulants with beta-cyclodextrin in phosphate buffers. Pharm Acta Helv 1978; 53: 241–247.

176. Ishizuka Y, Nagawa Y, Nakanishi H, Kuboyama A. Circular dichroism spectra of the inclusion complexes of phlorizin in cyclodextrins. J Incl Phenom Macrocycl Chem 1990; 9: 219–225.

177. Bekers O, Beijnen JH, Otagiri M, Bult A, Underberg WJ. Inclusion complexation of doxorubicin and daunorubicin with cyclodextrins. J Pharm Biomed Anal 1990; 8: 671–674.

178. Ciaran FP, Russell NR, McNamara M. Spectroscopic characterisation of metallo-cyclodextrins for potential chiral separation of amino acids and L/D-DOPA. J Incl Phenom Macrocycl Chem 2006; 56: 395–403.

179. Ventura CA, Giannone I, Musumeci T, Pignatello R, Ragni L, Landolfi C, Milanese C, Paolino D, Puglisi G. Physico-chemical characterization of disoxaril-dimethyl-beta-cyclodextrin inclusion complex and *in vitro* permeation studies. Eur J Med Chem 2006; 41: 233–240.

180. Hirayama F, Utsuki T, Uekama K, Yamasaki M, Harata K. Prominent inclusion effect of dimethyl-beta-cyclodextrin on photoisomerization of the thromboxane synthetase inhibitor (E)-4-(1-imidazoylmethyl)cinnamic acid. J Pharm Sci 1992; 81: 817–822.
181. Sortino S, Giuffrida S, Fazio S, Monti S. Spectroscopic characterization and photochemical behavior of host-guest complexes between beta-cyclodextrin and drugs containing a biphenyl-like chromophore. New J Chem 2001; 25: 707–713.
182. Barbato F, Cappello B, La Rotonda MI, Miro A, Quaglia F. Diclofenac/beta-cyclodextrin binary systems: a study in solution and in the solid state. J Incl Phenom Macrocycl Chem 2003; 46: 179–185.
183. Ndou T, Mukundan S, Warner IM. Complexation of ephedrine with beta-cyclodextrin: an NMR spectroscopy study. J Incl Phenom Macrocycl Chem 1993; 15: 9–25.
184. Chun IK, Yun DS. Inclusion complexation of hydrocortisone butyrate with cyclodextrins and dimethyl-beta-cyclodextrin in aqueous solution and in solid state. Int J Pharm 1993; 96: 91–103.
185. Ramusino MC, Bartolomei M, Gallinella B. H-1 NMR, UV and circular dichroism study of inclusion complex formation between the 5-lipoxygenase inhibitor zileuton and beta- and gamma-cyclodextrins. J Incl Phenom Macrocycl Chem 1998; 32: 485–498.
186. Matsubara K, Irie T, Uekama K. Spectroscopic characterization of the inclusion complex of a luteinizing hormone-releasing hormone agonist, buserelin acetate, with dimethyl-beta-cyclodextrin. Chem Pharm Bull (Tokyo) 1997; 45: 378–383.
187. Arima H, Yunomae K, Miyake K, Irie T, Hirayama F, Uekama K. Comparative studies of the enhancing effects of cyclodextrins on the solubility and oral bioavailability of tacrolimus in rats. J Pharm Sci 2001; 90: 690–701.
188. Bobek MM, Krois D, Brinker UH. Induced circular dichroism of cyclodextrin inclusion complexes: examining the cavity with a bilateral probe. Org Lett 2000; 2: 1999–2002.
189. Brewster ME, Loftsson T. Cyclodextrins as pharmaceutical solubilizers. Adv Drug Deliv Rev 2007; 59: 645–666.
190. Loftsson T, Masson M, Brewster ME. Self-association of cyclodextrins and cyclodextrin complexes. J Pharm Sci 2004; 93: 1091–1099.
191. Loftsson T, Hreinsdottir D, Masson M. The complexation efficiency. J Incl Phenom Macrocycl Chem 2007; 57: 545–552.
192. Woody RW. Circular dichroism. Methods Enzymol 1995; 246: 34–71.
193. Grishina IB, Woody RW. Contributions of tryptophan side chains to the circular dichroism of globular proteins: exciton couplets and coupled oscillators. Faraday Discuss 1994; 99: 245–262.
194. Kahn PC. The interpretation of near-ultraviolet circular dichroism. Methods Enzymol 1979; 61: 339–378.
195. Sreerama N, Manning MC, Powers ME, Zhang JX, Goldenberg DP, Woody RW. Tyrosine, phenylalanine, and disulfide contributions to the circular dichroism of proteins: circular dichroism spectra of wild-type and mutant bovine pancreatic trypsin inhibitor. Biochemistry 1999; 38: 10814–10822.
196. Metz B, Jiskoot W, Hennink WE, Crommelin DJ, Kersten GF. Physicochemical and immunochemical techniques predict the quality of diphtheria toxoid vaccines. Vaccine 2003; 22: 156–167.

197. Metz B, Jiskoot W, Mekkes D, Kingma R, Hennink WE, Crommelin DJ, Kersten GF. Quality control of routine, experimental and real-time aged diphtheria toxoids by *in vitro* analytical techniques. Vaccine 2007; 25: 6863–6871.

198. Schule S, Schultz-Fademrecht T, Bassarab S, Bechtold-Peters K, Friess W, Garidel P. Spray-drying of IgG1 formulations: a CD and FTIR investigation with respect to protein secondary structure. Respir Drug Deliv 2005; 1: 185–188.

199. Jovanovic N, Bouchard A, Hofland GW, Witkamp GJ, Crommelin DJ, Jiskoot W. Stabilization of IgG by supercritical fluid drying: optimization of formulation and process parameters. Eur J Pharm Biopharm 2007; 68: 183–190.

200. Gilg D, Riedl B, Zier A, Zimmermann MF. Analytical methods for the characterization and quality control of pharmaceutical peptides and proteins, using erythropoietin as an example. Pharm Acta Helv 1996; 71: 383–394.

201. Lam XM, Patapoff TW, Nguyen TH. The effect of benzyl alcohol on recombinant human interferon-gamma. Pharm Res 1997; 14: 725–729.

202. Tzannis ST, Hrushesky WJM, Wood PA, Przybycien TM. Irreversible inactivation of interleukin 2 in a pump-based delivery environment. Proc Natl Acad Sci U S A 1996; 93: 5460–5465.

203. Uversky VN, Garriques LN, Millett IS, Frokjaer S, Brange J, Doniach S, Fink AL. Prediction of the association state of insulin using spectral parameters. J Pharm Sci 2003; 92: 847–858.

204. Deechongkit S, Aoki KH, Park SS, Kerwin BA. Biophysical comparability of the same protein from different manufacturers: a case study using Epoetin alfa from Epogen and Eprex. J Pharm Sci 2006; 95: 1931–1943.

205. Ng K, Zhao L, Meyer JD, Rittmann-Grauer L, Manning MC. Use of circular dichroism spectroscopy in determining the conformation of a monoclonal antibody prior to its incorporation in an immunoliposome. J Pharm Biomed Anal 1997; 16: 507–513.

206. Rockey JH, Freed RM. Circular dichroism of hapten–antibody complexes: characterization of the combining sites of native and reformed MOPC-315 protein, its isolated subunits, and its Fv fragment. Scand J Immunol 1976; 5: 655–666.

207. Irie T, Uekama K. Cyclodextrins in peptide and protein delivery. Adv Drug Deliv Rev 1999; 36: 101–123.

208. Sarmento B, Ferreira DC, Jorgensen L., van de Weert, M.. Probing insulin's secondary structure after entrapment into alginate/chitosan nanoparticles. Eur J Pharm Biopharm 2007; 65: 10–17.

209. Jiang G, Joshi SB, Peek LJ, Brandau DT, Huang J, Ferriter MS, Woodley WD, Ford BM, Mar KD, Mikszta JA, Hwang CR, Ulrich R, Harvey NG, Middaugh CR, Sullivan VJ. Anthrax vaccine powder formulations for nasal mucosal delivery. J Pharm Sci 2006; 95: 80–96.

210. Glemza AA. Investigation of dye/protein interaction and its application to enzyme purification. J Mol Recognit 1990; 3: 137–141.

211. Sreerama N, Woody RW. Estimation of protein secondary structure from circular dichroism spectra: comparison of CONTIN, SELCON, and CDSSTR methods with an expanded reference set. Anal Biochem 2000; 287: 252–260.

212. Balasubramanian SV, Bruenn J, Straubinger RM. Liposomes as formulation excipients for protein pharmaceuticals: a model protein study. Pharm Res 2000; 17: 344–350.

213. Heavner GA, Arakawa T, Philo JS, Calmann MA, Labrenz S. Protein isolated from biopharmaceutical formulations cannot be used for comparative studies: Follow-up to "a case study using Epoetin Alfa from Epogen and EPREX". J Pharm Sci 2007; 96: 3214–3225.
214. Straubinger RM, Balasubramanian SV. Preparation and characterization of taxane-containing liposomes. Methods Enzymol 2005; 391: 97–117.
215. Abraham SA, Edwards K, Karlsson G, MacIntosh S, Mayer LD, McKenzie C, Bally MB. Formation of transition metal-doxorubicin complexes inside liposomes. Biochim Biophys Acta 2002; 1565: 41–54.
216. Gallois L, Fiallo M, Garnier-Suillerot A. Comparison of the interaction of doxorubicin, daunorubicin, idarubicin and idarubicinol with large unilamellar vesicles. Circular dichroism study. Biochim Biophys Acta 1998; 1370: 31–40.
217. Moribe K, Maruyama K, Iwatsuru M. Molecular localization and state of amphotericin B in PEG liposomes. Int J Pharm 1999; 193: 97–106.
218. Gaboriau F, Cheron M, Leroy L, Bolard J. Physico-chemical properties of the heat-induced "superaggregates" of amphotericin B. Biophys Chem 1997; 66: 1–12.
219. Szponarski W, Bolard J. Temperature-dependent modes for the binding of the polyene antibiotic amphotericin B to human erythrocyte membranes. A circular dichroism study. Biochim Biophys Acta 1987; 897: 229–237.
220. van Slooten ML, Visser AJ, van Hoek A, Storm G, Crommelin DJ, Jiskoot W. Conformational stability of human interferon-gamma on association with and dissociation from liposomes. J Pharm Sci 2000; 89: 1605–1619.
221. Sethi V, Onyuksel H, Rubinstein I. Liposomal vasoactive intestinal peptide. Methods Enzymol 2005; 391: 377–395.
222. Koppenhagen FJ, Visser AJ, Herron JN, Storm G, Crommelin DJ. Interaction of recombinant interleukin-2 with liposomal bilayers. J Pharm Sci 1998; 87: 707–714.
223. Hennenfent KL, Govindan R. Novel formulations of taxanes: a review. Old wine in a new bottle? Ann Oncol 2006; 17: 735–749.
224. Jullien S, Vertut-Croquin A, Brajtburg J, Bolard J. Circular dichroism for the determination of amphotericin B binding to liposomes. Anal Biochem 1988; 172: 197–202.
225. Barwicz J, Beauregard M, Tancrede P. Circular dichroism study of interactions of Fungizone or AmBisome forms of amphotericin B with human low density lipoproteins. Biopolymers 2002; 67: 49–55.
226. Larabi M, Gulik A, Dedieu JP, Legrand P, Barratt G, Cheron M. New lipid formulation of amphotericin B: spectral and microscopic analysis. Biochim Biophys Acta 2004; 1664: 172–181.
227. Wettig SD, Badea I, Donkuru M, Verrall RE, Foldvari M. Structural and transfection properties of amine-substituted gemini surfactant-based nanoparticles. J Gene Med 2007; 9: 649–658.
228. Zuidam NJ, Barenholz Y, Minsky A. Chiral DNA packaging in DNA-cationic liposome assemblies. FEBS Lett 1999; 457: 419–422.
229. Simberg D, Danino D, Talmon Y, Minsky A, Ferrari ME, Wheeler CJ, Barenholz Y. Phase behavior, DNA ordering, and size instability of cationic lipoplexes. Relevance to optimal transfection activity. J Biol Chem 2001; 276: 47453–47459.
230. Braun CS, Jas GS, Choosakoonkriang S, Koe GS, Smith JG, Middaugh CR. The structure of DNA within cationic lipid/DNA complexes. Biophys J 2003; 84: 1114–1123.

231. Wang DA, Narang AS, Kotb M, Gaber AO, Miller DD, Kim SW, Mahato RI. Novel branched poly(ethylenimine)-cholesterol water-soluble lipopolymers for gene delivery. Biomacromolecules 2002; 3: 1197–1207.

232. Patil SD, Rhodes DG, Burgess DJ. Biophysical characterization of anionic lipoplexes. Biochim Biophys Acta 2005; 1711: 1–11.

233. Zsila F, Iwao Y. The drug binding site of human alpha1-acid glycoprotein: insight from induced circular dichroism and electronic absorption spectra. Biochim Biophys Acta 2007; 1770: 797–809.

234. Rockey JH, Montgomery PC, Underdown BJ, Dorrington KJ. Circular dichroism studies on the interactions of haptens with MOPC-315 and MOPC-460 mouse myeloma proteins and specific antibodies. Biochemistry 1972; 11: 3172–3181.

235. Zsila F, Matsunaga H, Bikádi Z, Haginaka J. Multiple ligand-binding properties of the lipocalin member chicken alpha1-acid glycoprotein studied by circular dichroism and electronic absorption spectroscopy: the essential role of the conserved tryptophan residue. Biochim Biophys Acta 2006; 1760: 1248–1273.

236. Fitos I, Visy J, Zsila F, Mády G, Simonyi M. Conformation selectivity in the binding of diazepam and analogues to alpha1-acid glycoprotein. Bioorg Med Chem 2007; 15: 4857–4862.

237. Fitos I, Visy J, Zsila F, Mády G, Simonyi M. Selective binding of imatinib to the genetic variants of human alpha1-acid glycoprotein. Biochim Biophys Acta 2006; 1760: 1704–1712.

238. Alebic-Kolbah T, Kajfez F, Rendic S, Sunjic V, Konowal A, Snatzke G. Circular dichroism and gel filtration study of binding of prochiral and chiral 1,4-benzodiazepin-2-ones to human serum albumin. Biochem Pharmacol 1979; 28: 2457–2464.

239. Bertucci C, Domenici E. Reversible and covalent binding of drugs to human serum albumin: methodological approaches and physiological relevance. Curr Med Chem 2002; 9: 1463–1481.

240. Zsila F, Visy J, Mády G, Fitos I. Selective plasma protein binding of antimalarial drugs to alpha1-acid glycoprotein. Bioorg Med Chem 2008; 16: 3759–3772.

241. Detrich HW, 3rd, Williams RC, Jr, Macdonald TL, Wilson L, Puett D. Changes in the circular dichroic spectrum of colchicine associated with its binding to tubulin. Biochemistry 1981; 20: 5999–6005.

242. Gasymov OK, Abduragimov AR, Gasimov EO, Yusifov TN, Dooley AN, Glasgow BJ. Tear lipocalin: potential for selective delivery of rifampin. Biochim Biophys Acta 2004; 1688: 102–111.

243. Zsila F, Bikádi Z, Lockwood SF. *In vitro* binding of leukotriene $B_4$ ($LTB_4$) to human serum albumin: evidence from spectroscopic, molecular modeling, and competitive displacement studies. Bioorg Med Chem Lett 2005; 15: 3725–3731.

244. Dockal M, Carter DC, Ruker F. The three recombinant domains of human serum albumin. Structural characterization and ligand binding properties. J Biol Chem 1999; 274: 29303–29310.

245. Pistolozzi M, Bertucci C. Species-dependent stereoselective drug binding to albumin: a circular dichroism study. Chirality 2008; 20: 552–558.

246. Brown NA, Jahnchen E, Muller WE, Schmidt W, Wollert U. Species differences of the stereoselective binding of phenprocoumon and warfarin enantiomers to serum albumins. Comp Biochem Physiol C 1979; 62C: 101–105.

247. Petersen CE, Ha CE, Harohalli K, Feix JB, Bhagavan NV. A dynamic model for bilirubin binding to human serum albumin. J Biol Chem 2000; 275: 20985–20995.
248. Perrin JH, Vallner JJ, Nelson DA. Some quantitative investigations of the binding to and the displacement of bishydroxycoumarin from human serum albumin. Biochem Pharmacol 1975; 24: 769–774.
249. Otagiri M, Miyoshi T, Yamamichi R, Maruyama T, Perrin JH. Effects of tricyclic drug on induced circular dichroism spectra of dicumarol bound to alpha1 acid glycoprotein. Biochem Pharmacol 1991; 42: 729–733.
250. Ascoli G, Bertucci C, Salvadori P. Stereospecific and competitive binding of drugs to human serum albumin: a difference circular dichroism approach. J Pharm Sci 1995; 84: 737–741.
251. Maitra S, Saha B, Santra CR, Mukherjee A, Goswami S, Chanda PK, Karmakar P. Alprazolam induced conformational change in hemoglobin. Int J Biol Macromol 2007; 41: 23–29.
252. Maruyama T, Furuie MA, Hibino S, Otagiri M. Comparative study of interaction mode of diazepines with human serum albumin and alpha1 acid glycoprotein. J Pharm Sci 1992; 81: 16–20.
253. Otagiri M, Yamamichi R, Maruyama T, Imai T, Suenaga A, Imamura Y, Kimachi K. Drug binding to alpha1 acid glycoprotein studied by circular dichroism. Pharm Res 1989; 6: 156–159.
254. Otagiri M, Yamamichi R, Imai T, Imamura Y, Takadate A. Study on the binding of dicumarol to alpha1 acid glycoprotein using circular dichroism spectroscopy. Chem Pharm Bull (Tokyo) 1988; 36: 4958–4962.
255. El-Gamel S, Wollert U, Müller WE. Optical studies on the specific interaction of dipyridamole with alpha1 acid glycoprotein (orosomucoid). J Pharm Pharmacol 1982; 34: 152–157.
256. Rosen A. The measurement of binding constants using circular dichroism, Binding of phenylbutazone and oxyphenbutazone. Biochem Pharmacol 1970; 19: 2075–2081.
257. Russeva V, Zhivkova Z, Prodanova K, Rakovska R. Protein binding of piroxicam studied by means of affinity chromatography and circular dichroism. J Pharm Pharmacol 1999; 51: 49–52.
258. Cheruvallath VK, Riley CM, Narayanan SR, Lindenbaum S, Perrin JH. A quantitative circular dichroic investigation of the binding of the enantiomers of ibuprofen and naproxen to human serum albumin. J Pharma Biomed Anal 1997; 15: 1719–1724.
259. El-Gamal S, Wollert U, Müller WE. Binding of several phenothiazine neuroleptics to a common binding site of alpha1 acid glycoprotein, orosomucoid. J Pharm Sci 1983; 72: 202–205.
260. Brunner F, Müller WE. Prazosin binding to human alpha1 acid glycoprotein (orosomucoid), human serum albumin, and human serum. Further characterization of the "single drug binding site" of orosomucoid. J Pharm Pharmacol 1985; 37: 305–309.
261. Otagiri M, Nakamura H, Imamura Y, Matsumoto U. Circular dichroic investigations into binding of sulfonamides to serum albumin. Pharm Weekbl Sci 1989; 11: 207–212.
262. Tokuoka R, Okabe N, Tomita K. Circular dichroism studies on the interaction between human serum albumin and thyroxine. J Biochem 1980; 87: 1729–1734.
263. Rahman MH, Maruyama T, Okada T, Imai T, Otagiri M. Study of interaction of carprofen and its enantiomers with human serum albumin - II. Stereoselective site-to-site displacement of carprofen by ibuprofen. Biochem Pharmacol 1993; 46: 1733–1740.

264. Seng G, Bolard J. Circular dichroism studies in the near UV of ligand binding to chicken liver dihydrofolate reductase. Biochimie 1983; 65: 169–175.
265. Chen YQ, Zhou P, Berova N, Zhang HZ, Nakanishi K, Failli A, Steffan RJ, Molnarkimber K, Caggiano TJ. Conformational changes of rapamycin and analogs upon complexing with FKBP associated with activity: An application of second derivative CD spectroscopy. J Am Chem Soc 1994; 116: 2683–2684.
266. Su YY, Jirgensons B. Optical activity studies of drug-protein complexes. The interaction of acetylsalicylic acid with human serum albumin and myeloma immunoglobulin. Biochem Pharmacol 1978; 27: 1043–1047.
267. Lim K, Owens SM, Arnold L, Sacchettini JC, Linthicum DS. Crystal structure of monoclonal 6B5 Fab complexed with phencyclidine. J Biol Chem 1998; 273: 28576–28582.
268. Galat A. Interaction of riboflavin binding protein with riboflavin, quinacrine, chlorpromazine and daunomycin. Int J Biochem 1988; 20: 1021–1029.
269. Rahman MH, Maruyama T, Okada T, Yamasaki K, Otagiri M. Study of interaction of carprofen and its enantiomers with human serum albumin—I. Mechanism of binding studied by dialysis and spectroscopic methods. Biochem Pharmacol 1993; 46: 1721–1731.
270. Anderson SR. The circular dichroism spectra of the complexes of 1-anilino-8-naphthalenesulfonate with bovine serum albumin. Evidence for heterogeneity of binding. Biochemistry 1969; 8: 4838–4842.
271. Bertucci C, Andrisano V, Gotti R, Cavrini V. Use of an immobilised human serum albumin HPLC column as a probe of drug-protein interactions: the reversible binding of valproate. J Chromatogr B 2002; 768: 147–155.
272. Bischoff G, Hoffmann S. DNA-binding of drugs used in medicinal therapies. Curr Med Chem 2002; 9: 321–348.
273. Yang XL, Wang AHJ. Structural studies of atom-specific anticancer drugs acting on DNA. Pharmacol Ther 1999; 83: 181–215.
274. Nitta Y, Kuroda R. Quantitative analysis of DNA–porphyrin interactions. Biopolymers 2006; 81: 376–391.
275. Karlsson HJ, Eriksson M, Perzon E, Akerman B, Lincoln P, Westman G. Groove-binding unsymmetrical cyanine dyes for staining of DNA: syntheses and characterization of the DNA-binding. Nucleic Acids Res 2003; 31: 6227–6234.
276. Wilson WD, Tanious FA, Ding DY, Kumar A, Boykin DW, Colson P, Houssier C, Bailly C. Nucleic acid interactions of unfused aromatic cations: evaluation of proposed minor-groove, major-groove, and intercalation binding modes. J Am Chem Soc 1998; 120: 10310–10321.
277. Reddy BS, Sondhi SM, Lown JW. Synthetic DNA minor groove-binding drugs. Pharmacol Ther 1999; 84: 1–111.
278. Ismail MA, Sanders KJ, Fennell GC, Latham HC, Wormell P, Rodger A. Spectroscopic studies of 9-hydroxyellipticine binding to DNA. Biopolymers 1998; 46: 127–143.
279. Lyng R, Rodger A, Norden B. The CD of ligand-DNA systems. 2. Poly(dA-dT) B-DNA. Biopolymers 1992; 32: 1201–1214.
280. Lee EJ, Yeo JA, Cho CB, Lee GJ, Han SW, Kim SK. Amine group of guanine enhances the binding of norfloxacin antibiotics to DNA. Eur J Biochem 2000; 267: 6018–6024.

281. Strat D, Missailidis S, Drake AF. A novel methodological approach for the analysis of host-ligand interactions. Chemphyschem 2007; 8: 270–278.
282. Bischoff G, Gromann U, Lindau S, Meister WV, Hoffmann S. A spectroscopic and thermodynamic study of taxol nucleic acid complexes. Nucleosides Nucleotides 1999; 18: 2201–2217.
283. Zhdanov RI, Strazhevskaya NB, Jdanov AR, Bischoff G. A spectroscopic and surface plasmon resonance study of oleic acid/DNA complexes. J Biomol Struct Dyn 2002; 20: 231–242.
284. Krishnamurthy G, Ding W, Obrien L, Ellestad GA. Circular dichroism studies of calicheamicin-DNA interaction: evidence for calicheamicin-induced DNA conformational change. Tetrahedron 1994; 50: 1341–1349.
285. Hwangbo HJ, Yun BH, Cha JS, Kwon DY, Kim SK. Enantioselective binding of *S*- and *R*-ofloxacin to various synthetic polynucleotides. Eur J Pharm Sci 2003; 18: 197–203.
286. Lee EJ, Yeo JA, Jung K, Hwangbo HJ, Lee GJ, Kim SK. Enantioselective binding of ofloxacin to B form DNA. Arch Biochem Biophys 2001; 395: 21–24.
287. Streltsov S, Oleinikov V, Ermishov M, Mochalov K, Sukhanova A, Nechipurenko Y, Grokhovsky S, Zhuze A, Pluot M, Nabiev I. Interaction of clinically important human DNA topoisomerase I poison, topotecan, with double-stranded DNA. Biopolymers 2003; 72: 442–454.
288. Kolodziejczyk P, Garnier-Suillerot A. Circular dichroism study of the interaction of mitoxantrone, ametantrone and their Pd(II) complexes with deoxyribonucleic acid. Biochim Biophys Acta 1987; 926: 249–257.
289. Krishna AG, Kumar DV, Khan BM, Rawal SK, Ganesh KN. Taxol-DNA interactions: fluorescence and CD studies of DNA groove binding properties of taxol. Biochim Biophys Acta 1998; 1381: 104–112.
290. Bush CA, Pesce A, Gaizutis M, Nair V. Binding and circular dichroism studies on interactions of lysergic acid diethylamide with deoxyribonucleic acid. Mol Pharmacol 1972; 8: 104–109.
291. Wang H, Yang P, Tian Y, Zhang Z, Zhao C. Experimental antitumor activity of the Ce(IV)-mitoxantrone complex and its interaction with deoxyribonucleic acid. J Inorg Biochem 1997; 68: 117–121.
292. Hossain M, Giri P, Suresh Kumar G. DNA intercalation by quinacrine and methylene blue: a comparative binding and thermodynamic characterization study. DNA Cell Biol 2008; 27: 81–90.
293. Lopez MC, Ruiz LM, Craciunescu D, Doadrio A, Osuna A, Alonso C. Studies on the interaction between antitrypanosome *cis*-DDP analogs with DNA. Chem Biol Interact 1986; 59: 99–111.
294. Berova N, Ellestad G, Harada N. Characterization by circular dichroism spectroscopy. In: Mander L, Liu HW, editors. Comprehensive natural products II, Chemistry and biology, volume 9, Modern methods in natural products chemistry. Oxford: Elsevier; 2010. p 91–146.
295. Ranjbar B, Gill P. Circular dichroism techniques: biomolecular and nanostructural analyses-a review. Chem Biol Drug Des 2009; 74: 101–120.
296. Schweitzer-Stenner R, Hagarman A, Verbaro D, Soffer JB. Conformational stability of cytochrome c probed by optical spectroscopy. Methods Enzymol 2009; 466: 109–153.

297. Rudd TR, Yates EA, Hricovíni M. Spectroscopic and theoretical approaches for the determination of heparin saccharide structure and the study of protein-glycosaminoglycan complexes in solution. Curr Med Chem 2009; 16: 4750–4766.

298. de Andrés F, Castañeda G, Ríos A. Achiral liquid chromatography with circular dichroism detection for the determination of carnitine enantiomers in dietary supplements and pharmaceutical formulations. J Pharm Biomed Anal 2010; 51: 478–483.

299. Lecoeur-Lorin M, Delepee R, Adamczyk M, Morin P. Simultaneous determination of optical and chemical purities of a drug with two chiral centers by liquid chromatography-circular dichroism detection on a non-chiral stationary phase. J Chromatogr A 2008; O 1206: 123–130.

300. Reetz M. Screening for enantioselective enzymes. In: Svendsen A, editor. Enzyme functionality. Boca Raton: CRC Press; 2003. p 559–598.

301. Frezard F, Martins PS, Bahia AP, Le Moyec L, de Melo AL, Pimenta AM, Salerno M, da Silva JB, Demicheli C. Enhanced oral delivery of antimony from meglumine antimoniate/ β-cyclodextrin nanoassemblies. Int J Pharm 2008; 347: 102–108.

302. Matei I, Soare L, Tablet C, Hillebrand M. Characterization of simvastatin and its cyclodextrin inclusion complexes by absorption and circular dichroism spectroscopies and molecular mechanics calculations. Rev Roum Chim 2009; 54: 133–141.

303. Ravi J, Hills AE, Knight AE. Reproducible circular dichroism measurements for biopharmaceutical applications. In: Wallace BA, Janes RW, editors. Modern techniques for circular dichroism and synchrotron radiation circular dichroism spectroscopy. Amsterdam: *IOS* Press; 2009. p 125–140.

304. Fast JL, Cordes AA, Carpenter JF, Randolph TW. Physical instability of a therapeutic Fc fusion protein: domain contributions to conformational and colloidal stability. Biochemistry 2009; 48: 11724–11736.

305. Maltesen MJ, Bjerregaard S, Hovgaard L, Havelund S, van de Weert M. Analysis of insulin allostery in solution and solid state with FTIR. J Pharm Sci 2009; 98: 3265–3277.

306. Kumar V, Sharma VK, Kalonia DS. In situ precipitation and vacuum drying of interferon alpha-2a: Development of a single-step process for obtaining dry, stable protein formulation. Int J Pharm 2009; 366: 88–98.

307. Jovanovic N, Bouchard A, Hofland GW, Witkamp GJ, Crommelin DJ, Jiskoot W. Stabilization of IgG by supercritical fluid drying: optimization of formulation and process parameters. 2008; 68: 183–190.

308. Costantino HR, Culley H, Chen L, Morris D, Houston M, Roth S, Phoenix MJ, Foerder C, Philo JS, Arakawa T, Eidenschink L, Andersen NH, Brandt G, Quay SC. Development of calcitonin salmon nasal spray: similarity of peptide formulated in chlorobutanol compared to benzalkonium chloride as preservative. J Pharm Sci 2009; 98: 3691–3706.

309. van der Walle CF, Sharma G, Ravi Kumar M. Current approaches to stabilising and analysing proteins during microencapsulation in PLGA. Expert Opin Drug Deliv 2009; 6: 177–186.

310. Patarroyo ME, Cifuentes G, Bermudez A, Patarroyo MA. Strategies for developing multi-epitope, subunit-based, chemically synthesized anti-malarial vaccines. J Cell Mol Med 2008; 12: 1915–1935.

311. Ausar SF, Joshi SB, Middaugh CR. Spectroscopic methods for the physical characterization and formulation of nonviral gene delivery systems. In: Le Doux JM, editor. Methods in molecular biology, volume 434, Gene therapy protocols, Design and

characterization of gene transfer vectors. 3rd ed. Totowa: Humana Press Inc.; 2008. p 55–80.

312. Zsila F, Fitos I, Bencze G, Kéri G, Örfi L. Determination of human serum alpha1-acid glycoprotein and albumin binding of various marketed and preclinical kinase inhibitors. Curr Med Chem 2009; 16: 1964–1977.
313. Fitos I, Visy J, Simonyi M, Mády G, Zsila F. Selective binding interactions of deramciclane to the genetic variants of human alpha1-acid glycoprotein. Biochim Biophys Acta 2010; 1800: 367–372.
314. Ishizaki J, Fukaishi A, Fukuwa C, Yamazaki S, Tabata M, Ishida T, Suga Y, Arai K, Yokogawa K, Miyamoto K. Evaluation of selective competitive binding of basic drugs to alpha1-acid glycoprotein variants. Biol Pharm Bull 2010; 33: 95–99.
315. Nagele P, Mendel JB, Placzek WJ, Scott BA, D'Avignon DA, Crowder CM. Volatile anesthetics bind rat synaptic snare proteins. Anesthesiology 2005; 103: 768–778.
316. Pidikiti R, Zhang T, Mallela KM, Shamim M, Reddy KS, Johansson JS. Sevoflurane-induced structural changes in a four-α-helix bundle protein. Biochemistry 2005; 44: 12128–12135.
317. Eckenhoff RG, Pidikiti R, Reddy KS. Anesthetic stabilization of protein intermediates: myoglobin and halothane. Biochemistry 2001; 40: 10819–10824.
318. Johansson JS, Eckenhoff RG. Experimental approaches to the study of volatile anesthetic-protein interactions. In: Moody E, Skolnick P, editors. Molecular bases of anesthesia. Boca Raton: CRC Press; 2001. p 37–68.
319. Rodger A. Circular and linear dichroism of drug-DNA systems. In: Fox KR, editor. Methods in molecular biology, volume 613, Drug-DNA interaction protocols. 2nd ed. New York: Humana Press, a part of Springer Science + Business Media, LLC; 2010. p 37–54.
320. Onoue S, Seto Y, Oishi A, Yamada S. Novel methodology for predicting photogenotoxic risk of pharmaceutical substances based on reactive oxygen species (ROS) and DNA-binding assay. J Pharm Sci 2009; 98: 3647–3658.

# *Good Manufacturing Practices—Steps to Improve Quality*

**John E. Snyder**

## 1 GOOD MANUFACTURING PRACTICES: GOOD BUSINESS PRACTICES

With the advent of Current Good Manufacturing Practices (CGMP) in 1978 (1) came a major shift in the way the Food and Drug Administration (FDA) enforces drug quality in the pharmaceutical industry. Simply stated, the quality of a product is no longer determined by the results of end-product tests alone. It must also be documented that product has been manufactured, packaged, and held according to the written procedures that comply with CGMP regulations.

In fact, the Food, Drug, and Cosmetic Act goes so far as to say that a drug product shall be deemed to be adulterated if the methods used in, or the facilities or controls used for, its manufacture, processing, or holding do not conform to or are not operated or administered in conformity with current good manufacturing practices (2). Therefore, regardless whether a drug product is materially defective, deviations from CGMP regulations cause drug products to be adulterated within the meaning of the Act, and violative articles are subject to seizure and condemnation in district courts.

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing,*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

Likewise, the United States Pharmacopoeia (USP), the compendium of U.S. official analytical methods, agrees that although every unit of product must conform to requirements, this cannot be established by product testing alone. Rather "data derived from manufacturing process validation studies and from in-process controls may provide greater assurance that a batch meets a particular monograph requirement than analytical data derived from an examination of finished units drawn from that batch (3)."

The assurance of product quality down to the individual unit level is important because that is what patients take—individual tablets, capsules, injections—often daily and often lifelong. With this in mind the premise of CGMP is to drive quality by enforcing regulations that require manufacturers to document their capability to operate in a state of continuous control. However, fear of enforcement action is not the best inspiration for sustaining a quality operation. CGMP regulations must be viewed as an enabler of good business practices that achieve the shared goals between the FDA and the manufacturer of product safety, efficacy, potency, purity and other quality attributes. Thus the recognition of CGMP as a catalyst for quality, rather than a business antagonist, is the first step to improve quality.

When examined from a business perspective, CGMP regulations cover aspects of daily manufacturing operations that are the very foundation upon which businesses compete, which is to have reliable product supply that meets customer requirement with economic control of quality (4). We can make this connection between complying with CGMP and enabling good business practice by broadly considering the business implications of six major themes in the regulation.

## 1.1 There must be Product and Process Knowledge to Achieve Consistency (5)

This means that we will understand the science and technology of the product development from beginning to end, which led to the routine operation of converting raw materials and packaging components into the final packaged dosage form.

A range of knowledge is necessary such as supplier capabilities to meet requirements; equipment and instruments selection and operating ranges; and test methods and specifications used to evaluate acceptability. We know the process control points that are critical and those that are not. We know which upstream factors cause downstream affects and just how far to the edge of the range we could go without affecting important outcomes. The reason we are knowledgeable of our process is because the variables have been studied during the product development phase before being transferred and accepted into the routine manufacturing environment.

We also have conducted studies and documented that when we operate according to our approved manufacturing and packaging instructions, we have a high degree of statistical confidence that our product will meet requirements and fall in a normal random distribution comfortably within the limits of our specifications. Since these specifications are based upon outcomes in clinical studies, we have proven and validated our capability to manufacturer product that delivers intended results.

For the business, product and process knowledge translates into a reliable and efficient operation. A very low probability of rework or reprocessing, and minimal process-related down time is expected that enables a predictable, right-first-time manufacturing schedule and a reliable supply of products to customers. Also every batch of product has a pedigree linking it to the manufacturing and testing methodology used for clinical studies, which was the basis for the approval of the drug product application. When errors or unintended drift occurs within the operating parameters and specifications, a data-driven conclusion on the product effect can be readily determined by referring to the collective body of process knowledge.

## 1.2 Manufacturing Process Resides within a Controlled Facility and Environment (6)

The facility is designed to preclude mix-ups and contamination by controlling the paths of materials, components, in-process and final drug product, and people. We design-in the ability to clean manufacturing, packaging, and holding areas by selecting appropriate materials for construction of floors, walls, and ceilings, and that the cleaning agents and methods achieve the degree of cleanliness or sanitization required. We ensure that the supply of critical utilities such as air, water, gases meet standards for intended use. Environmental temperature and humidity meet the requirements to maintain the stability of raw material, in-process, and final product. We carefully consider material storage space and layout to support the in-process product flow and warehousing of raw materials and components to prevent mix-ups.

Protecting operators is critical in terms of what they touch and breathe, as well as how operator movement, behavior, and traffic flow through the facility could potentially contaminate product. So proper HVAC systems, air flow patterns, and defined spaces are engineered and monitored. We provide for the safety and ergonomic needs of our operators. We consider the security of the facility and to control access of employees and visitors and prohibit the access by rodents and other pests.

For the business, this means efficient and predictable movement of materials from dock to stock. Product rejection and re-inspection due to contamination by pharmaceutical dust, operator handling, or foreign product are engineered out of the process. Public relations issues from costly recalls and customer complaints are prevented due to mix-ups in product and labeling due to inadvertent commingling of materials. When a facility works with the process and people, there is a positive work environment and reduced human resource issues.

## 1.3 Sources of Unwanted Variation must be Prevented from Impacting the Process (7)

The facility is kept in good repair and good housekeeping is practiced. Equipment and instruments are well maintained by placing them on preventive maintenance and calibration schedules based on manufacturer recommendations and extent of use so that we have confidence in the accuracy and precision of in-process readings from equipment and test instruments. We ensure that everyone is trained for their specific

job duties and has a working understanding of their role in complying with CGMP. When new employees are hired, they demonstrate to responsible management their ability to perform tasks as trained and established in written procedures in order to control the human variable.

We understand that we must operate in a state of control that enables us to predict a successful product outcome. Therefore, we guard against anything that has a potential effect on the state of control of our operation. Proposed changes to anything we do are carefully, if not skeptically, considered on a technical level to ensure that the impact is understood and acceptable—even if the proposed change is viewed as an improvement. The effectiveness of changes is confirmed against the pre-established intended result.

For the business, this means that a check and balance exists for when innovative process technology and business practices are introduced to an established product manufacturing process. Innovation should always be explored, especially if it translates into increased productivity and the ability to better compete in the marketplace. The intent of CGMP is never to deter technological advances, rather to combine process knowledge with regulatory intent to achieve "mindful compliance (8)." However, caution is exercised by demanding data-driven conclusions regarding product impact and equivalence as filed in the regulatory application in order that product integrity is maintained.

## 1.4 There must be a Record of the Manufacture and Controls used to Make a Product Batch (9)

This includes written instruction such as procedures, protocols, test methods, master batch records that describe the manner in which work activities must be performed. We understand that these instructions form much of the content of specific job training and that these instructions must be written clearly at the level of the person performing the operation. Informal instructions are not permitted and only those that are written, current and approved are available in the work area.

Documentation also includes the recorded output from written instructions. We recognize that activities are recorded at the time they were performed and results obtained at the time they were observed. The person actually doing the work, and likewise the actual witness for critical steps sign and date the record. We establish documentation conventions to preclude perception of fraud or misunderstanding as to the intention of the entry such as significant figures, correcting entry errors, and handling unused fields in records. Taken collectively, batch-specific documentation establishes an audit trail of the complete history of any given batch of product. The importance of documentation is underscored by the fact that FDA inspections routinely involve a documentation review as evidence of following CGMP, more than product testing.

For the business, documentation is legal evidence of product manufacture. It is the record of meeting regulatory compliance and product quality requirements and the basis upon which product is released into the market. It is an essential source of information for data mining for technical studies, process improvement as well as

for an investigation into a process deviation or a customer complaint. Documentation that supports a validated and stable process in general, as well as specific batches of product is invaluable for product liability defense cases.

## 1.5 Problems must be Detected and Handled Appropriately when they Occur (10)

This requires that we establish feedback mechanisms that report negative trends and failures of products, processes, and systems. Such feedback is obtained from established processes that handle customer complaints, returned goods, annual product reviews (APRs), performance metrics, deviation and failure reports from the manufacturing floor, and quality control laboratories.

Feedback detection systems are described in procedures and include the type of issues, metrics, and trends that are routinely reported and reviewed, as well as those that are quickly escalated to management responsible for governance.

We respond to reports from detection systems by determining the impact on product, whether or not the product has been distributed (11). This involves a thorough investigation into the facts of the situation, determining the root cause or most probable cause, the degree of risk, the necessity of interim controls while the issue is corrected and prevented from recurring (12). Depending upon the degree of risk, the product may need to be voluntarily recalled from commerce. Responsible management is always informed of regulatory inspections, recalls, significant issues being investigated, and key performance metrics.

For the business, these requirements provide the processes to continuously measure the effectiveness of the quality system, the state of control of the operation, as well as highlight anomalies for immediate action and improvement. This means that a quality improvement program is embedded within the regulation, which is focused on error cause removal, and that many of the classic tools of quality are directly applicable. When compiled into periodic reports, management with executive responsibility is provided with an important input for corporate governance and an essential input into the planning process of the business.

## 1.6 The Quality Unit is Responsible for Testing, and Implementing and Measuring the Effectiveness of the Quality System (13)

An independent quality unit is traditionally separated into two functional units: quality control and quality assurance. Quality control is a laboratory or measuring function such as analytical chemistry, microbiology, immunochemistry, biological, attributes inspection, or physical testing that examines raw materials, components, in-process material, and final packaged product against pre-determined specifications. We ensure that quality control has adequate laboratory facilities to perform their function. We typically operate a quality control unit as a work center with planning values and resources for testing cycle times that are balanced to meet the volume demand and production schedule. Laboratory operations may be performed by

a contract company that has been approved by a qualification protocol and operate under a quality agreement.

Quality assurance has several responsibilities that are required by CGMP. They are responsible for establishing the systems that make possible all the other themes described above, as well as measuring the effectiveness through audits and quality metrics. They have the responsibility of reviewing and approving CGMP documents as well as proposed changes before they are implemented. Quality assurance is responsible for reviewing and approving the adequacy of investigations related to manufacturing deviations, failures to meet specification, and customer complaints. They review production batch records and other manufacturing records and are empowered to make the decision to release product into the marketplace. We are careful that responsibilities are assigned to quality assurance are not a conflict of interest, for example, not creating situations where quality assurance performs work for which they are expected by regulation to review and approve.

For the business, the quality unit is the objective function within the operation that assists management to exercise its governance responsibility. The quality unit is uniquely positioned to provide responsible management with performance measures that not only report on nonconformance, but also enable management to anticipate and identify potential product quality and regulatory compliance risks. The management challenge is to allow the quality assurance function to operate independently, yet integrate the function into the business. This often is achieved by balancing independence for assessing and reporting with accountably for establishing systems that are both compliant and efficient and their promoting ownership for product quality and compliance outcomes across all functional areas.

This broad survey of themes illustrates how interrelated the concerns are between CGMP regulations and business. It is interesting to note, however, that CGMP regulations do not prescribe in all cases exactly how a business is to comply. Although there may be FDA guidance on how to interpret and apply regulations, FDA is thankfully silent on how exactly to design a system, step-by-step for any specific business application.

It is tempting sometimes for someone to transplant procedures from their previous company experience—just changing the letterhead. It may come with an endorsement that says, "The FDA had no problem with this at my previous company." Here is where it must be emphasized that one size truly does not fit all, and that there is no such thing as a universal template for what a company should specify in its procedures.

In fact, a lot of harm can be done to a business in the name of experience when there is a failure to recognize that a quality system is more than a set of written procedures. A procedure is just words on a page unless it takes into consideration the environment it is in. For example, it is quite possible, say, for a document management procedure to require that a simple, clarifying wording change in a test method to go through the same document review and change control systems as major change in equipment with little regard for the rigor of review being commensurate with the risk being managed. Although, compliance is achieved if the procedure is followed, such an approach could negatively impact a business that needs to be nimble in order to respond quickly, yet effectively to market changes to be competitive.

A pharmaceutical business needs to be both CGMP compliant and competitive. Since there are many ways to implement the CGMP regulations, the best one is the one that works to serve both ends. This fact opens a door of opportunity to enhance both compliance effectiveness and business efficiency by improving the design and implementation of the CGMP quality system.

The first step to improving quality is to recognize that the CGMP regulation is an enabler of good business practices as well as product and process quality. The quality system implemented for CGMP compliance is a worthy target for improvement, since enhancement will be realized for both quality and the business.

However, if a quality system is more than a set of procedures—more than words on a page, then what is a system? Understanding the answer to this question is the second step to improving quality.

## 2 GOOD MANUFACTURING PRACTICES: GOOD SYSTEM PRACTICES

The concept of a system can be illustrated by a radio. The antenna is not a radio; neither are the diodes, resistors, or speaker. However, each of these parts has a purpose and each part works together so that collectively they are the radio and operate predictably regardless of who turns it on. Over time the radio has evolved in its reliability to transmit and replicate sound with greater fidelity so that now we can enjoy music as if we were in the very midst of the orchestra. Over time improvement has been directed toward each part. As one part was improved, others followed. Each worked with the others to enhance the overall quality of the radio. Improved quality was an output of improved radio parts that work together as a system.

So it is with the CGMP quality system. Quality is an output of all the individual parts of the system working together. The degree to which each part of the quality system is designed to meet requirements—and, interacts, or relates well with each other and with the business environment is the degree to which a quality output is achieved.

Thus, the second step to improving quality is to understand that a system is comprised of many parts that must interact well together. Like the illustration of the radio, the component parts of the CGMP quality system must work well together to be effective.

When evaluating how well the parts of a CGMP quality system work together, it may be useful to consider the three levels of features that promote effective interaction: those that are *within*, *between*, and *around* the parts of the system.

Let us examine the first level of features that enhance interaction that are *within* a given part of the system.

### 2.1 Written Procedures

This might seem obvious, since throughout the CGMP regulations we read that instructions for meeting the requirements are to be in writing—and followed. The term

"established" is typically used in the regulatory and pharmaceutical fields to mean that instructions on how requirements are met are in writing, approved by the quality unit, and implemented into the workplace by training all affected personnel. Without procedures it is difficult, if not impossible to control the human variable, or defend how a company intentionally expects work to be performed. Procedures are at the core of the training curriculum and a company defense; therefore, considerable attention must be given to the adequacy of written procedures and competency-based training programs to ensure procedures are translated into performance in the work area.

It is common for a company to have an architecture of policies and procedures that go from the general requirements at the corporate level to the more specific site standard operating procedures. Architecture provides a balanced outline and structure to the library of interlinking procedures at a site, or network of sites throughout the corporation. Guidance is commonly provided to standardize the format, style, and level of writing so that the entire body of written instructions fits together. Even with guidance, not all are gifted with the ability to write clearly and preserve the balance of the body of procedures as a whole. Unfortunately, it is far too common for a procedure to be written to plug a gap, rather than to consider where the instructions would more appropriately fit in the process. For this reason, authoring procedures should be vested with the best of writers, and the review and approval process should ensure the highest of writing standards are preserved. In short, if one were to view the table of contents for the body of policies and procedures, it should resemble the logic and orderliness of the table of contents of a college textbook.

## 2.2 Process Flow

The process flow follows and facilitates the order in which work is actually performed. Flowcharting is the tool most often used by owners and stakeholder teams to visualize the process when either creating a new process or improving an existing process. A process flow diagram can promote conversations and enhance understanding between functions that have a stake in how a current process operates, or how a new process will operate before the procedure becomes effective. Independent persons can also add value to the mapping process by watching for redundancies, re-cycling of activities, or other inefficiencies that may be deeply held conventions by those close to the process. Process flow diagrams also enhance procedure training by showing trainees the big picture, key interactions, and decision points, as well as their role in the overall effectiveness of the process.

Incomplete, inefficient, and disjointed processes open the door to informal practices to develop over time. "Failure to follow the procedure" is a familiar regulatory inspection citation. However, it does not always reflect an unwillingness or inability of the user. It may reflect the lack of a fluid, practical or understandable way of doing the job. The term "work around system" is often used to describe an unauthorized, sometimes easier, way to get the job done. Although a "work around system" cannot be condoned in the pharmaceutical industry, no one says that the right way has to be confusing. It is important for a process to be complete, but not difficult to follow.

## 2.3 Established Responsibilities

Clearly defined responsibilities, including designees—or back-ups are important to ensure that the appropriate individuals are qualified and authorized to take on the specified roles. There are some instances where councils convene regularly for a specific purpose, such as an environmental review board or material review board. In these cases, the responsibility of the individuals who participate, as well as the group as a whole should be described in the procedure. If a responsibility includes review and approval, the procedure should specify what the reviewer and approver are signifying by his or her signature to be clear and emphasize the importance of this responsibility.

Aside from the various responsibilities for the process, it is important for there to be an owner who has overall responsibility for the performance of a given part of the system. The owner is responsible for ensuring that the process keeps up with evolving compliance requirements and current industry practice. The owner is also responsible for being the spokesperson for the process, knowing how well the process is working, and continually seeking way to improve it. Without clear ownership and defined responsibilities, there can be organizational confusion and inefficiency.

## 2.4 Qualified Decision-Making Authority

All processes have decision points. While some decisions may be very straightforward, others may require critical thinking skills to weigh the available information, assess the risks and implications in order to render the best decision. Therefore, care must be taken where decision-making is vested to ensure that the appropriate qualified persons make decisions. Consideration should be given whether an individual, or a group makes a decision. Decisions by groups tend to be slower in order to reach consensus, but there is more buy-in. Decisions by individuals are quicker, but more difficult to implement. There are some situations that may require the escalation of a decision. Escalation may be invoked when an individual or team cannot reach a decision or when the risk involved warrants participation in the decision-making process by more senior management.

As discussed earlier, the unique role of the quality unit in decision-making must be preserved as defined in the regulation. However, the quality unit is well served by consulting other operating functions to make informed decisions.

## 2.5 Records Document Work Performed

A process will always include pathways for records to be completed to document process inputs and outputs, as well as signatures and dates of those who perform and verify steps. Records include forms, batch records, logbooks, recording charts, printouts, and other means of permanent documentation of process information and results. It takes skill to design simple and user-friendly records that require all the essential, value-added information to be captured. Instructions on how to complete the record must be given in sufficient detail in a related procedure to achieve consistent

documentation across multiple users. Procedures and records should be reviewed together when changes are made to avoid inconsistencies.

An emphasis must be placed on careful documentation. It is common that general training is given on good documentation practices to promote clear and consistent entries, as well as documentation practices to prevent perception of fraudulent records.

Specific training on each record is also essential for users to understand how to properly document specific work, and to understand the reason behind each entry. Only through complete, accurate, and timely records can investigations be conducted and the manufacturing history be told. In fact, so essential are records that without them there is no product. For this reason, original records are controlled through an established document management process to keep them secure during use and catalogued in a storage area, or archive that prevents loss or damage.

## 2.6 Measurement of Performance

Each part of the CGMP quality system should establish the metrics that are to be reported and reviewed. The procedure should provide the frequency, distribution and review and approval responsibility. The metrics selected should indicate how well the system is operating, and what the system is detecting. These two dimensions are necessary to give a complete picture of the effectiveness of the quality system. For example, a Pareto analysis of customer complaints will show what the system is detecting. However, a report on the number of new complaint files opened, the number complaint files closed during the period and the cycle time of pending complaint files (0–30 days; 31–60 days; 61–90 days, etc.) will tell how well the system is operating.

Operating metrics are often viewed as "efficiency" or "business" metrics, not CGMP. However, nothing could be farther from reality. As discussed above, CGMP and the business are really inseparable. In the earlier example of complaint handling, operational metrics may be the only indication that a backlog liability is mounting, which is a CGMP issue. Backlogs of any kind within the quality system such as complaint handling and manufacturing deviation investigations, preventive maintenance, training, annual product reviews, validation and requalification are signals of inattention to activities that can identify and prevent future failure. Backlogs of any kind are most always associated with a reactionary work environment (14).

In some cases, a diagnostic metric may be necessary to go a level deeper to pinpoint where specific remedial action is required. In the complaint handling example, if a significant backlog were revealed by the operating metric, a life cycle metric would breakdown the number of days a complaint file dwells at the various stages of the handling process. This diagnostic metric could reveal, for instance, if the bottleneck was due to repeated review cycles between manufacturing and quality to rework a better-written investigation. This fact would lead to targeted improvement initiatives. The use and interaction with performance measures is part of the responsibility of the system owner to manage the process; and it is the responsibility of the owner to escalate issues when the performance metrics show unfavorable trends that are beyond his or her control to quickly resolve.

System failures that result in a serious business loss and regulatory enforcement action are often traced to the inability to self-detect and self-correct failures or trends toward failure in sufficient time for management to intervene at the most economical time to resolve the problem. When there is a product, process or system failure, or a field alert, recall or regulatory action, there should always be an examination into the performance metric that was missing or unheeded that could have alerted the right persons at the best possible time to take appropriate action. Performance metrics enable an organization to move from a reactionary to an anticipating environment. More about performance metrics follows later.

As an example of the above first level of interactions *within* a part of the CGMP system, let us consider features of a supplier qualification program to assure control over suppliers.

One would expect that a qualification process begins with the selection of the candidate supplier for the specific material to be procured. The process would further include how the material and the manufacturer would be evaluated and approved for use.

Responsibilities would be defined for all who participate in the process such as product development who identifies the grade and specifications for the material; purchasing identifies a reliable supplier that is capable of meeting the specifications at the right price; and quality assurance conducts the audit of the supplier against the appropriate standard and assesses inherent risk in the ability of the supplier to meet the specifications. The responsibility for preparing the supply agreement will be with Purchasing, as well as the quality agreement by the Quality unit.

The decision to use a supplier may be made jointly between purchasing and quality assurance, or perhaps a material management team. The quality unit gives the final approval by placing the supplier on the approved supplier list and determines the audit frequency and audit technique to be used based on risk to the process, product, and patient. If there is a change in supplier, the quality unit approves the technical justification through the change control process.

One or more files or databases are defined in the procedure as the formal records that document the supplier qualification process. This would include the supplier audit and responses; copies or references to supplier performance at in-coming inspection; deviations that point to a material as the root cause; as well as corrective and preventive action (CAPA) plans.

A quarterly supplier performance report will be issued to the functional heads of product development, manufacturing, purchasing, and quality assurance. The report includes the approval status of suppliers, noting particularly those that are associated with laboratory or manufacturing failures along with the status of corrective and preventive action (CAPA). Additionally, the status of supplier audits, planned versus conducted, is provided.

All these features reflect interactions *within* the supplier qualification component of the CGMP quality system. Each of these features is important for the system to operate effectively and efficiently.

In our continuing list of system features, let us now examine the second level interactions, which are *between* parts of the CGMP quality system.

## 2.7 Defined Relationship between Parts of the System

The CGMP quality system is a web of interdependent relationships. The effectiveness of one part is dependent on the effectiveness of other parts. Each part of the system has inputs and outputs that link one to the other, and deficiencies in one place can have far-reaching effects elsewhere in the system.

Take CAPA for instance. A CAPA system is only as effective as the ability to detect failure or trend toward failure throughout the operation. One cannot correct or prevent what one does not detect. Thus, all inputs must be identified where a deviation in product, process, or system could occur, and processes must be in place to link these as inputs to the CAPA process in order to effectively investigate, correct, and prevent recurrence. Likewise, the output of the CAPA process is a link back to the originating area of the deviation to verify that the measures taken to prevent recurrence were indeed effective, as well as to keep the area informed where the deviation originated. Therefore, CAPA must be linked to other parts of the CGMP quality system, which must be carefully identified as inputs and outputs. Without these defined relationships, the process can become an island of activity creating a false sense security—a lot of motion and resources, but no real benefits.

It is a common practice for a procedure to have a list references. The most effective use of a list of references is to help identify the interrelated systems. This is particularly useful when a change is made to the procedure. The reference list tells you where else to look that may be impacted by the change. Much like an ecosystem, one cannot make a change in one place without examining the effect on the others.

It is also worthwhile to map the interrelationship. Much like the process flow mapping in the first feature, process mapping between parts of the CGMP quality shows the practical ways the parts interrelate. This also promotes a better understanding between all persons involved in executing the work.

## 2.8 Ownership of Dependent Processes

Much like having an owner for each component part of the CGMP quality system, an owner of interrelated processes operates at a higher level where the dependent processes come together. There is no better way to break down the silos within an organization than to create a network of owners that are responsible for the performance of the dependent processes that invariably operate cross functionally.

Continuing with the supplier example, an owner of dependent processes would include all the processes that insure materials used for CGMP purposes from the point of identification of supplier to its use on the manufacturing floor. The overall owner, business process owner if you will, is identified by name as well as the individual process owners. When such a team is identified and challenged with the responsibility and authority to make data-driven decisions and measurable improvement to materials management, a powerful interdisciplinary resource is created to manage major segments of the business. Rather than creating *ad hoc* teams to react to the problem-of-the-day, ongoing process teams as described here should be communicating regularly and interacting with performance metrics to prevent failure and continuously

improving the operation. This approach creates a natural structure to manage and improve the business, as opposed to popular "quality improvement teams" that are sometimes directed to "spotty" problems and not well integrated with the operation on a continuing basis. Continuous improvement requires continuous integration.

As an example of this second level of interaction *within* a part of the CGMP system, let us consider the dependent processes of transferring technology from a sponsoring operation to a contract manufacturing operation.

The business process owner at the contract manufacturer may be the Technology Transfer function, which typically understands the technical capability of the operation and the functional and regulatory requirements to communicate well with the business development function, internal operating areas, and customer.

A procedure is in place and in use to direct the transfer and timing to ensure all requirements are met. The content of a quality agreement is also procedurally established to identify those who are responsible, accountable, consulted, and informed between each entity for the myriad of activities involving shared responsibilities. The procedure guards against practices that affect variability and reliability. For example, effort is made not to introduce differing ways of working at the contractor site, and to safeguard essential steps from the transfer process affect reliability.

Team members include process owners from facility and process engineering, validation, purchasing, operations, training, and quality who understand what it takes to establish a reliable manufacturing process. This would include requirements for the facility, utilities, equipment, and operational capabilities; the implication of the customer bill of materials (BOM) for identifying approved suppliers; validation to ensure that quality attributes, critical control parameters are defined and translated into manufacturing processes and procedures; operators are trained in the associated procedures and batch records. Team membership may also extent to the sponsor company.

The component part process owners, as led by the overall business process owner, continually examine the effectiveness of the interrelated process to ensure that the requirements are met in the most effective and efficient way by reviewing the performance measures that each brings to the table; making data-driven decisions and setting measurable goals to enhance business performance.

Finally, let us conclude our list of features as we examine the third level interactions, which are *around* parts of the CGMP quality system.

## 2.9 Supporting Organization

Having the right organization structure and number of people with the right skills is not only essential for executing the work, but is essential to operate in compliance and in a state of control. However, this is often a management challenge in a competitive marketplace where headcount and cost of goods sold (COGS) is always scrutinized to improve profitability. However, there are opportunities to streamline organizations that support the CGMP quality system and reduce costs without wandering into compliance problems.

There is an erroneous view that all things quality and CGMP belong in the quality unit. This notion duplicates functionality, enables lack of ownership for quality and compliance in the operation areas, and drives up costs. For example, nowhere is it required that the quality unit should perform training, validation, and investigations. Many organizations place these functions within the quality unit, for the sole reason that this function is thought to be more knowledgeable about the regulatory requirements to perform these activities. While the quality unit has responsibility to ensure a system is established for these functions, and it has a review an approval authority, the activities do not necessarily belong within the quality unit. In fact, this could be viewed as a conflict of interest. The education, skills, and experience should more appropriately reside in the disciplines of human resources, engineering, and manufacturing supervision respectively. If the skills do not exist in these functional areas, then the strategy should be to build the capabilities within these parts of the organization, which is more cost effective than adding headcount to the quality unit to double-team the work. The more a quality unit moves toward its basic regulatory commission, the smaller the quality organization needs to be, while the knowledge in the operating areas is increased.

However, those who control operating budgets must understand that a CGMP quality system must be in place to comply fully with the regulations whether one batch or a thousand batches are produced. For example, regardless of production volume documents are still managed; equipment and instruments are still calibrated and maintained; changes are still made and must be technically justified; the environment is still monitored; annual product reviews are still written, analyzed, and reviewed; complaints and manufacturing deviations are still investigated; metrics are still performed, reviewed and actions taken—the list goes on. Therefore, there must be a sufficient staff level to establish, maintain, and operate the system as the price of admission into the industry.

Nonetheless, there are aspects of most systems that do relate directly to volume such as the number of laboratory tests for in-coming materials, in process controls, and final product; the number of batched records reviewed; the number of deviations investigated; and so forth. Thus, the CGMP quality system should not be above review for cost reduction opportunities, but making a distinction between those aspects that are variable based on volume from those that are fixed is important to prevent the quality system from collapsing.

## 2.10 Supporting Infrastructure

This is the assortment of bricks and mortar, computer systems, and software programs—even workspace that either helps, or gets in the way of effectively implementing the CGMP quality system to support the operation.

For example, inadequate warehouse space can prohibit segregation and adequate cleaning; manual record keeping can inhibit the ability for "look backs" to go through voluminous historical records to identify similar product complaint and trends; and insufficient number of computer terminals can make computer-aided training modules difficult to deploy, schedule, and complete on time; or outdated hardware and

software versions of engineering and maintenance systems can make these difficult to service.

Although infrastructure is often less tangible, it nonetheless can affect productivity and the ability to adequately comply with CGMP regulations.

Automating a quality system may not always be the answer to improving the infrastructure, particularly for small operations and low volumes, since computer system validation regrettably introduces another level of complexity to maintain from a regulatory perspective. Bricks and mortar issues can sometimes be resolved creatively, and employees often hold the answers. However, when examining the infrastructure issues, it is important to have as the super-ordinate goal to facilitate information and process flows and to eliminate non-valued-added steps, redundant files, and time delays.

These features of interrelationships *within*, *between*, and *around* a CGMP quality system help us to take a broader view of the nature of a system. If one were to have the misfortune to be the recipient of an FDA483 or a Warning Letter, one might pause before taking the typical approach in the name of system thinking of simply making a revision to a standard operating procedure and attaching it to the response. Indeed, a CGMP quality system is indeed more than words on a page.

Thus, the second step to improving quality is to recognize that a system involves the myriad of interrelationships that work together to provide predictable and sustainable results. An examination into the effectiveness of the relationships within, between, and around the CGMP quality system is another worthy target to identify opportunities for improving quality.

But real-world experience says that things do not go perfectly all the time. In fact, CGMP regulations acknowledge this by a provision for handling deviations through a process of technical review, technical justification, and documentation. However, the question is not whether a pharmaceutical manufacturing operation should expect only normal, random variation to be operating all the time. Rather, the question is whether the operation has the capability to detect deviation from the norm and respond appropriately with good decisions to bring the operation back into control. This brings us to the third step to improve quality: performance metrics.

## 3 GOOD MANUFACTURING PRACTICES: GOOD METRICS PRACTICES

There is no CGMP regulation that specifies the metrics to be reviewed to evaluate the performance of the quality system. Drug regulation 21CFR211.180e requires a rather belated review of records at least annually to determine whether or not changes are needed for specifications, manufacturing process, or controls. In essence, this is a retrospective analysis of process capability to determine whether changes may be required to ensure a manufacturing process is stable and operating well within specifications and operating control limits.

Closer to the intent is found on the medical device side where 21CFR820.20 provides for a management review of the suitability and effectiveness of the quality

system at defined intervals. Also, the review of "quality data" is mentioned as a source of information for identifying existing and potential causes for quality problems in 21CFR820.100 as part of the CAPA system. Thus, it is presumed that quality data is being collected that would reflect the adequacy of the quality system.

The International Conference on Harmonization (ICH) Pharmaceutical Quality System Q10 guidance document offers some detail to "manage and continually improve the pharmaceutical quality system." The first is a formal periodic review of key performance indicators such as complaints, deviations, CAPA, and change management processes, as well as self-assessments, regulatory inspections, and customer audits. The second is a review of factors that impact the quality system such as new regulations, guidance, and industry trends, as well as changes in the business that could result from innovation and new strategies. Finally, the guidance suggests that the outcome of the review should promote quality system improvement, evaluation of resources, and enhanced communication or escalation of issues to senior management.

Appropriately, the regulations and guidance have left it to individual companies to determine their own approach to applying specific performance metrics to manage the quality system, and the range of industry practices widely varies.

On the rather simplistic side, some companies search for a single index to know how well the quality system is working so that a 9.5 provides more assurance than a 4.0 that the quality system is working well and the operation is in a state of control. Some have argued that the stock price of the company is the single most important numerical indicator—though rather lagging in its ability to point to an emerging quality problem. However, after a problem manifests itself in a supply interruption, or after a regulatory enforcement action becomes public information, then this single indicator will certainly come into play, though in an unfavorable direction.

There are some who believe that the true measure of the adequacy of the quality system is the outcome of regulatory inspections where fewer observations are better than many, and the absence of a negative is an implied positive. Some go as far as to compare the number of observations for the current year to a running average over the history of the company to serve as a barometer of their system as though some statistical treatment can be applied to these events. Such believers are eventually caught off guard and typically surprised when an inspection does not go well—or when regulators rebuke them for having to be inspected into compliance.

On the other end of the spectrum, there are companies that have an exhaustive set of metrics that roll up into a dashboard that summarizes massive amounts of information about the entire operation at a glance. Embedded somewhere are elements of the quality system such as corporate and internal audits. Line items are often rolled up to an overarching category and reduced to a color code that flashes the attention of the reader to the problem areas. Backup data is typically displayed in a spreadsheet, which may be compressed onto as few pages as possible. However, rarely are trends toward failure detected in this manner, or connections made between the dots for sporadic, but significant emerging issues.

Still other approaches major in graphic displays as though the objective is to use every quality tool in the pocket guide. Though appealing to those who prefer pictures

to columns of numbers, the point can be missed amongst the hackneyed pie charts and graphs—particularly in the absence of well-written analysis to direct the eye of the reader.

These scenarios are not uncommon and reflect differing views and attempts to satisfy a range of management interest and attention spans. The challenge is not only to prepare meaningful performance metrics, but also to make the business case for the value of management being engaged with quality system metrics with the same interest as sales figures from the various territories and regions it serves. In both cases, relevant data are reviewed against expectations, and when expectations are not met, adjustments are made and the effects monitored. When expectations are met, there are rewards and celebrations. Whether its sales or CGMP quality system performance metrics, it is unreasonable to expect that all numbers come in on target all the time. However, it is a reasonable expectation that management engages with performance metrics in order to manage the performance of the company (15). For both sales and the CGMP quality system, the effects on the business are significant in order to sustain the operation and to compete long-term in the marketplace.

Thus, the third step to improving quality is to manage the effectiveness and the efficiency of the quality system through meaningful performance metrics and action-oriented management reviews.

There are many kinds of tools and techniques for reporting, trending, and summarizing data. There are many textbooks and pocket guides to choose from that deal with metrics. In the end, the right metrics are the ones that work; and those that work answer three basic questions: Is the system being managed or functioning as expected? What is the system detecting? What is causing the problem?

The first question is of fundamental importance, since a poorly managed or poorly functioning system may give a false impression—such as no news is good news. Operational metrics are designed to give us information about the system itself.

Let us take the self-quality audit system as an example. An operational metric could be the number of audits performed versus planned. Another could be the number of overdue audit report responses versus number of audit reports issued. So for instance, if the gap between the number of audits planned and performed begins to increase over time, then an inquiry should be made into the reason for the backlog. Reasons could point to performance, absenteeism, competing priorities, or even unwillingness for the auditee to allow audits to be scheduled. Likewise, a backlog in overdue audit responses may signal such things as the lack of attention to the task or an inability to prepare a response to the satisfaction to the auditor. Another example is the analytical laboratory. An operational metric could be the cycle time from the time a sample is collected to the time the result is officially reported versus the planned time. An increase in the cycle time could be a function of increased testing volume, a problem with the method or instrumentation, or perhaps lack of a proper test schedule.

The second question is only of complete value after the first question is answered. The issues or the magnitude of the issues detected by the system presume that the system is operating as intended to be able to detect problems. For example, given the

complaint handling system is operating well, what complaints are we seeing? Given the laboratory operations systems are working well, what do the results tell us? Given the deviation system is operating well, what non-conformances are we experiencing? And so forth.

So it presumes, for instance, that the Pareto analysis used as a detection metric to report the top compliance issues detected by the system were based on all the planned audits, not just the two out of the twelve that were actually performed; or that the favorable cycle time for reporting lab results was based on all tests required for release, not just finished product.

Operational metrics and detection metrics each tell part of the story. Reporting only one or the other is a common flaw in the application of metrics, and both are necessary for proper interpretation and decision-making.

The third question is answered by diagnostic metrics, which attempt to direct focused action toward the underlying reason for problems detected by properly functioning systems. Continuing with the self-quality audit example, with the operational metrics indicating that all planned audits were performed and response reports closed, the issues identified in rank order by frequency detected that the most frequent audit finding was failure to follow the company policy on good documentation practices. Upon reviewing this fact at the regularly scheduled management review, it was decided to look into this matter further to diagnose the problem in order to direct a targeted initiative toward specific improvement. The quality group was asked to further break down the findings to determine the type of records and the work center where the documentation errors were found.

This exercise resulted in the discovery that most of the documentation errors occurred in product batch records related to the sterile manufacturing operation. Armed with this information, the manufacturing process owner along with owners of related systems such as technical services and quality assurance examined these specific records and interviewed employees in the specific area. It was discovered that since this particular work group and process technology was recently added to the capability of the site, and that the timetable for the technology transfer process was compressed, the training requirements were truncated, which included training on good documentation practices. It was also discovered that the master batch records transferred with the process were error prone, since data fields in many instances were inconsistent with documentation requirements.

By using diagnostic metrics, management was able to direct specific action. In this example, the sterile manufacturing work group was given the requisite training. Additionally, management directed a review of required standards and training for technology and personnel transferred to the site—while they also reflected upon the effect of their previous decision to cut the training time.

Although practices may vary across companies for how quality system performance metrics are selected, reported, and reviewed, developing a common approach within an organization enhances communication and the overall effectiveness as an essential business tool. Here are seven attributes of good metrics practices that will help performance metrics for the quality system to be clear, concise, concrete, comparative, and cogent.

### 3.1 Metrics and their Calculations are Defined

This is particularly important for common metrics that are reported across functions within a site, or metrics that are reported between sites. The source of the data and how the metric is calculated is important to ensure they are derived the same way. Definition and standardization ensures that there is no misunderstanding about what exactly is being reported, so that true comparisons can be made and pooling of data can be done. Let us take customer complaints as an example. It is important to know whether both prescription and OTC products are included, and whether the complaints include both medical and product attribute complaints. Also the denominator such as per number of batches distributed per time period must be established.

### 3.2 Metrics are Comparative and Time-Related

For metrics to answer, "Are we getting better, worse, or the same?" metrics need to be presented in a way that enables us to compare performance over a relevant time period and set of conditions. For example, the variability of a critical dimension of a packaging component from a supplier before and after implementation of corrective measures requires the need to know dates of the turn key or phase-in implementation, as well as supplier lot number information to be able to measure the improvement at incoming inspection. Another example is the need to know the production shift, manufacturing line, and employees where the improved production batch record was directed in order to show if there was a reduction in errors within the same group compared to the previous version.

### 3.3 Metrics have Direction and Scale

Decide on the direction on the graph that will be the universal direction that indicates improvement. It can be very confusing when up means better one time, and down is better the next within or between reports. The most common practice is for "down" to signify "better." This practice is most likely associated with the "Zero Defects" quality school, where zero defects are the quality standard. Also, be sure that the scales on the axis are chosen appropriately for sensitivity of the measurement. Changing the scale on the graph between reporting periods and using overly large units of scale are the oldest tricks in the book to obscure a negative trend.

### 3.4 Measure the Opportunity, Not the Success

This guidance is perhaps the most controversial between optimists and pessimists. The optimist might rather focus on the number of investigation completed, rather than the number still open. The pessimist might rather see the number of batches with errors, rather that those that are correct. It is true that "Right First Time (RFT)" sounds a lot better than "Wrong First Time (WFT)." However, measuring the opportunity is not intended to convey pessimism, rather to focus on the exposure and risk.

## 3.5 Metrics are Actionable

Metrics should answer the questions previously discussed: Is the system being managed or functioning as expected? What is the system detecting? What is causing the problem? Anyone presenting metrics for management review should be prepared to interpret the data and make recommendations. When action is warranted, then it should be clear who is accountable for achieving favorable results, and all the participants that affect the outcome should be identified to partner in the initiative. Goals and target dates should be set, as well as the frequency to receive updates on the metrics to ensure that the initiative is supported and getting results.

## 3.6 Metrics Enable an Organization to Predict Performance and be Proactive

Perhaps the most important benefit of quality system metrics is the capability it gives to the organization to predict a problem and proactively address it at the most economical time, before it becomes a more costly quality problem or a regulatory enforcement issue. Therefore, systems should define a trend, and to specify when special action is required when a trend is identified. Some systems may follow the simple rule of "One is a point, two is a line, and three is a trend." When traditional statistical process control (SPC) techniques are used, then the Westinghouse Rules may apply whereby a trend or an outlier is flagged based on some practical rules of thumb-based statistical theory. The customer complaint handling is an example of a system that needs to err on the side of a false positive trend to detect an early product liability issue. For certain, trends need to be applied to the systems that manage the most significant regulatory and financial risks in order to ensure escalation of issues to higher levels of management as quickly as possible.

## 3.7 Align Metrics with the Business Plan

Quality system metrics are very well suited to provide a component to the business process of setting company goals and personal performance objectives. They add the "specific and measurable" dimension. Nothing could be more powerful as a serious driver for change than to align quality system metrics with the business plan. In fact, recommendations for targeted improvement should be the natural outcome of the ongoing management review of quality system metrics. One of the most significant contributions by the quality function is to bring data-driven targets to the business planning sessions, as well as to objectively report on the results from the previous year targets.

The question often arises as to who should be responsible in an organization for establishing metrics practices and reporting quality system performance for management review. Although there could be a number of suitable answers to this question, a case can be made for this to be the responsibility of the quality unit.

The quality unit is the only functional organization with a job description in the Code of Federal Regulations. Among other duties prescribed, it has the responsibility

to independently ensure an effective quality system is established, and that the effectiveness of the quality system is evaluated. Metrics are probably the single most effective means to make that evaluation, far more effective that audits. Audits provide feedback for a point in time on whatever the focus of the audit may be at the time. However, metrics provide an ongoing intelligence across the entire quality system—and the business. Because of the power of metrics, the quality function has a significant vested interest that metrics are solidly in place procedurally and in use behaviorally.

The fact that the quality function is responsible for the quality system is underlined by the fact that more warning letters are beginning with, "The Quality Unit has failed to ensure an adequate system is in place to"—then fill in the blank. The focus is turning to whether there is a strong and independent quality organization that is able to establish an effective quality system, and whether the quality system is capable of detecting a failure and righting itself in a timely manner through corrective and preventive action. Metrics is the quintessential tool of the quality function, which gives it a diagnostic capability.

For organizations that have a well-implemented set of quality system metrics, the way quality audits can be performed is revolutionary. In an effective metrics environment, routine internal quality audits are able to evolve and add value in ways that are unheard of within the pharmaceutical industry today. Audits could examine the validity and integrity of databases from which metrics are derived; confirm that calculations and metrics as reported for high risk areas are accurate; re-evaluate the appropriateness of metrics described in procedures; and examine whether quality systems should be enhanced based on regulatory and benchmarking surveillance. This is in contrast to the prevalent mode of simply conducting audits of departments against their existing procedure—regardless of how adequate they may be. An evolving audit function should be part of the vision for an enlightened, futuristic quality organization that is thoroughly embedded in a quality system metrics environment.

Reporting on quality system metrics should be the responsibility of the quality function. First, the weight of the regulatory responsibility for the quality system falls squarely on the quality function. Secondly, the independence and objectivity of the quality function puts them in the unique position to deliver the difficult messages often revealed by metrics. Company executives should nurture and welcome quality function "straight talk," which may be difficult for less independent functions to deliver.

Metrics also provide the quality function with a means to significantly contribute to the business planning process. Too often the quality function is not quite up to par where business planning is concerned. Unfortunately, business planning is often thought by quality professional to be the domain of the business units, or associates with business degrees. Nothing could be farther from the truth. Keep in mind that embedded in the wealth of quality system metrics is data about waste, errors, non-conformance, complaints, and all manner of other sources for cost improvement. Armed with quality system metrics, the quality function should be well prepared to offer each planning period to the executive team the three to five most significant issues that, when addressed, will result in annualized savings to the business, as well

as improve quality and reduce regulatory compliance risk. The quality function should also be able to provide data-driven evidence of improvement and savings for the initiatives identified the previous year.

Metrics reveal three possible states of a given element of the quality system: stable, event failure, and system failure.

The term stable is used to signify that the metrics are usual and expected; and comparable to a relevant period. For SPC applications, this description signifies "in statistical control" where only "common cause" or normal, random variation is operating. There is no statistically significant difference in means and variability between data sets; and that the data points are randomly distributed about the mean.

For example, metrics indicate that the number of investigations closed equals the number opened; and that the average time to closure is 20 days with few investigation exceeding 30 days; those that exceed 30 days are justifiable and approved; and that the average number of investigations per batch is stable at a low targeted level. These metrics have been maintained over the last 6 months. The conclusion is that the quality system for investigations is stable.

The term event failure is used to signify that the quality system detected an event or a non-conformance, either in the operation of the system or in what the system detected. For SPC applications, this description signifies an instance or instances of being "out of statistical control" due to an assignable cause. When the problem is corrected, the metrics indicate a return to the stable state.

For example, metrics indicate that there had been a delay in meeting the procedural deadline for APR reports for the second consecutive month. Further investigation into this operational metric showed that the delays corresponded to a vacancy in the position responsible for these reports. Although the issue might have been better tended to when the first overdue report was noted, the metrics did bring the issue front and center. The failure related to the vacancy event, and the system was brought back into control through a temporary assignment, as an interim control then was fully resolved when the position was permanently filled.

The term system failure is used to signify a breakdown in the operation of the system, or breakdown of whatever the system is detecting. For SPC applications this description signifies "out of statistical control" due to causes that generally affect the system from operating as established. To remedy a system failure requires an assessment against the 10 attributes of a system as described above.

For example, the FDA visited to follow-up on numerous field complaints on a critical care medical device due to the melting of a plastic connector used in a critical care device. The failure caused a malfunction of the device during patient use. An investigation revealed that a change from a brass to a plastic connector had been implemented as a cost improvement initiative. The change was expedited in order to capture the savings in the current quarter. Since the supplier of the plastic component was ISO certified and the connector was a standard catalogue item, the decision was made by the quality unit to approve the change. When the FDA asked to see customer complaint records, it was discovered that there were four additional complaints unnoted for this malfunction waiting processing due to a backlog.

There were multiple system failures in this example. The change control system failed to technically evaluate the effect of the change; the supplier qualification system failed to accept the plastic connector based on fit for purpose; the complaint handling system failed to detect associated complaints in a timely manner to be able to address the problem; the quality unit failed to ensure there was an adequate system in place, and further it enabled an improper decision.

Thus, an ongoing review of performance metrics is an essential means for corporate and site governance in order to know the state of the CGMP quality system, and to direct a data-driven response toward improvement.

## 4 CONCLUSION: GOOD MANAGEMENT PRACTICES

An exasperated company executive once declared that he was tired of their firm being a "learning organization." Puzzled by his comment, he further explained that he meant he was tired of all the "lessons learned" meetings that followed when there were quality problems, supply interruptions, and surprise findings from FDA inspections. He wanted the organization to become an "anticipating organization" that was capable of seeing further down the road and around the corner to be able to make adjustments before it was too late.

His comment led to an examination of the information that was collected and reported throughout the organization on their quality system and their state of CGMP compliance.

Reports centered on quality projects from their continuous improvement office. Regulatory compliance activity was limited to timeliness of closing out audit responses, and an indication that FDA had not visited during the reporting period. The report was dotted with reassurance of efforts to benchmark and their recent internal award for best-demonstrated practices across the company network.

It was clear from an analysis of the information and reports he routinely received that he had been flying blind. He had not been receiving the type of intelligence from the CGMP quality system that he needed to truly manage the regulatory compliance aspect of his business. In the final analysis, there was nothing much to base his confidence on the effectiveness of the quality system, or state of control of the operation. Yet, the operation was busy doing "quality."

Over time his paradigm shifted about the CGMP quality system. No longer was it the onerous obligation to satisfy hard-nosed FDA inspectors, rather a strategic imperative that supported his concern for patients and customers, and the ability to compete. The quality system was not just a set of procedures, but the strategic intent to align the organization and operate in a state of control—yet flexible in how the intent of regulations can be carried out. Most of all, "management review" was no longer empty, mail-drop report, rather an objective and data-driven cross-functional session that gave him confidence that the CGMP quality system was operating well, and directed attention to identified problems, unacceptable trends, and risks. This gave him the edge he was looking for to stay ahead of the problems.

In the end, he recognized that the CGMP quality system itself is a worthy target of continuous improvement to ensure that good business practices, good system practices, and good metric practices were applied—or one could sum up by saying good management practices—in order to compete and be worthy of patient trust.

## REFERENCES


1. FR 45077, September 29, 1978.
2. FD&C Act, Section 501 (a)(2)(B).
3. United States Pharmacopoeia, Volume XXIV, p. 7, Rockville, MD.
4. Shewhart WA. Economic control of quality of manufactured product. New York: Van Nostrand; 1931. p 26–34.
5. Code of Federal Regulations, Title 21, Volume 4, Revised as of April 1, 2007, Subparts D, E.
6. Code of Federal Regulations, Title 21, Volume 4, Revised as of April 1, 2007, Subparts C, H.
7. Code of Federal Regulations, Title 21, Volume 4, Revised as of April 1, 2007, Subparts B, D, F.
8. Snyder J.Mindful compliance: where knowledge and regulations meet. BioPharm International Supplement, Guide to Good Manufacturing Practices 2004; November: 26–34.
9. Code of Federal Regulations, Title 21, Volume 4, Revised as of April 1, 2007, Subparts D, E, J.
10. Code of Federal Regulations, Title 21, Volume 4, Revised as of April 1, 2007, Subparts F, J.
11. Snyder J. Deemed and redeemed: principles of justification for deviation reviewers. J GXP Compliance 2004; 8 (3): 6–14.
12. Snyder J. Corrective and preventive action: planning to achieve sustainable GMP compliance. J GXP Compliance 2002; 6 (3): 29–39.
13. Code of Federal Regulations, Title 21, Volume 4, Revised as of April 1, 2007, Subparts B, I.
14. Hayes RH, Wheelwright SC, Clark KB. Dynamic manufacturing: creating the learning organization. New York: Free Press; 1988. p 225.
15. Snyder J. 1999. Management responsibility for the quality system: a practical understanding for the CEO in FDA-regulated industries. J cGMP Compliance 3 (3): 55–59.

# *High Performance Liquid Chromatography (HPLC) in the Pharmaceutical Analysis*

**Shulamit Levin**
*Medtechnica, Petach Tivka, Israel*

## 1 INTRODUCTION

Drug manufacturing control requires high level and intensive analytical and chemical support of all stages to ensure the drug's quality and safety [1]. The pharmacopeia constitutes a collection of recommended procedures for analysis and specifications for the determination of pharmaceutical substances, excipients, and dosage forms that is intended to serve as source material for reference or adaptation by anyone wishing to fulfill pharmaceutical requirements. The most important analytical technique used during the various steps of drug development and manufacturing is the separation technique: high performance liquid chromatography (HPLC).

The key to a proper HPLC system operation is knowledge of the principles of the chromatographic process, as well as understanding the reasons behind the choice of the components of the chromatographic systems such as column, mobile phase, and detectors. A scheme of an HPLC system is shown in Fig. 1. A high pressure pump is

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing,*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

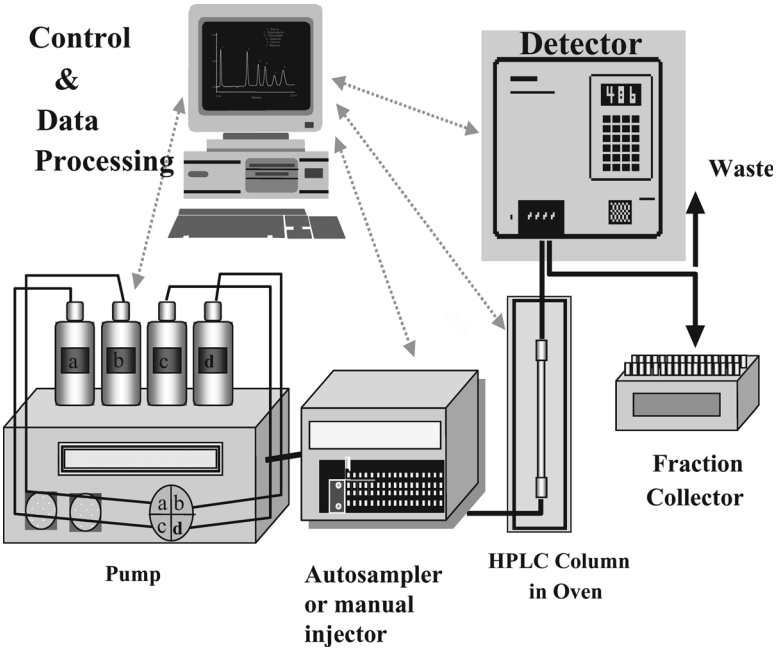


***FIGURE 1*** *A scheme of an HPLC system.*

required to force the mobile phase through the column at typical flow rates of 0.5–2 mL/min. The sample to be separated is introduced into the mobile phase by injection device, manual or automatic, prior to the column. The detector usually contains low volume cell through which the mobile phase passes carrying the sample components eluting from the column. There are books describing the practicality of HPLC operation [2–11]. It is expected of any proper HPLC system that is used in the pharmaceutical laboratories to produce highly accurate and precise results, due to health-related issues of improper measurements. Every HPLC system must be qualified to comply with the strict demands from health authorities for high quantitative performance.

Quality standards in pharmaceutics require that all instruments should be adequately designed, maintained, calibrated, and tested. The approach that has been adopted in the environment of the analytical instrument has become known as the "Four Qs": design qualification (DQ), installation qualification (IQ), operational qualification (OQ), and performance qualification (PQ). Design qualification is performed at the vendor's site, and it is representative of the way an instrument is developed and produced, usually governed by International Organization for Standardization (ISO) criteria.

The IQ process can be divided into two steps: *pre-installation* and *physical installation*. During pre-installation, all information relevant to the proper installation, operation, and maintenance of the instrument is checked. Workers confirm the site requirements and the receipt of all of the parts, pieces, and manuals necessary to

perform the installation of the specific HPLC unit. During physical installation, serial numbers are recorded and all fluidic, electrical, and communication connections are made for system components. Documentation describing how the instrument was installed, who performed the installation, and other various details are archived.

The operational qualification process ensures that the separate modules of a system (pump, injector, and detector) are operating according to the defined specifications such as accuracy, linearity, and precision. Specific tests are performed to verify parameters such as detector wavelength accuracy, flow rate, or injector precision.

The PQ step verifies system performance as a whole. Performance qualification testing is conducted under real operating conditions in the analytical laboratory that is going to be using the instrument. In practice, sometimes operational and performance qualification blend together, particularly for linearity and precision (repeatability) tests, which can be conducted more easily at the system level.

The performance qualification test of the HPLC system uses a method with a well-characterized analyte mixture, column, and mobile phase. It incorporates type of measurements from the system suitability section of the general chromatography Chapter ⟨621⟩ in the *U.S. Pharmacopeia* [12]. In the end of the process proper documentation is archived.

## 2 MODES OF HPLC

There are various modes of operation of HPLC. The mechanism of interaction of the solutes with the stationary phases determines the classification of the mode of liquid chromatography. Table 1 summarizes the variety of modes of liquid chromatography, of which reversed phase stands out as the most widely used mode in HPLC, therefore, the discussion will elaborate on this mode.

### 2.1 Reversed Phase [13]

Reversed phase liquid chromatography (RPLC) is considered as the method of choice for the analysis of pharmaceutical compounds for several reasons, such as its compatibility with aqueous and organic solutions as well as with different detection systems and its high consistency and repeatability. Sensitive and accurate RPLC analysis, whether in the pharmaceutical or bioanalytical field, necessitates the use of stationary phases which give symmetrical and efficient peaks. Therefore, manufacturers of stationary phases are continuously improving and introducing new RPLC products, and the selection of various types of reversed phase stationary phases is high. The needs for consistency as well as the globalization of the pharmaceutical companies require that the methods will be transferred from site to site, using either the same column brands or their equivalents. Therefore, an extensive categorization or characterization of the rich selection of stationary phases has been done in recent years [14–21].

The stationary phase in the reversed phase chromatographic columns is a hydrophobic support that is consisted mainly of porous particles of silica gel in various

**TABLE 1 Summarizes the Most Abundant Modes of Chromatography in Pharmaceutical Analysis**

| Mode | Normal phase | Reversed phase | Ion exchange | Chiral | Affinity | Size exclusion |
|---|---|---|---|---|---|---|
| Stationary phases chemistry | Polar-hydrophilic | Non-polar-lipophilic | Ion-bonding | Chiral recognition | Bioaffinity | Sieving by size |
| Typical stationary phases | Silica, alumina | Alkylated silica, mostly C18 | Ionic functional groups on silica or polymer | Chiral groups on silica surfaces | Either substrates or biomolecules, | Gel type polymers |
| Typical mobile phase | Hexane, isopropa-nol. methylene chloride | Water, methanol, acetonitrile, buffers, ion pairing agents | Water, buffers, acid, base | Two modes: aqueous and non-aqueous | Water, buffers | Two modes: aqueous and non-aqueous |
| Typical solutes | Fatty and oily | Almost all organic compounds | Any ion-charged compounds | Enantiomers small and large molecules | Biomolecules or their substrates | Polymers: synthetic or biological |

Reversed phase chromatography is by far the most widely used mode of chromatography.

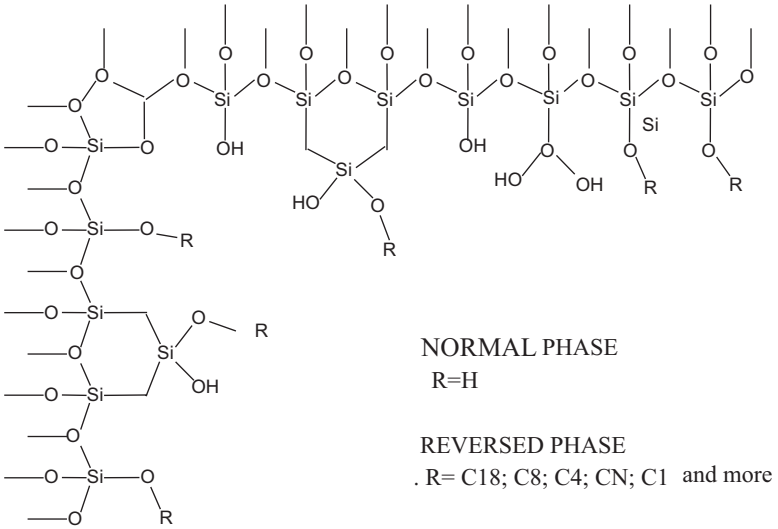


***FIGURE 2*** *The surface of reversed phase stationary phases.*

shapes (spheric or irregular) at various diameters (1.8, 3, 5, 7, 10 μM, etc.) at various pore sizes (such as 60, 100, 120, 300). The surface of these particles is covered with various chemical entities, such as various hydrocarbons (C1, C6, C4, C8, C18, etc.) as can be seen in Fig. 2. There are also hydrophobic polymeric supports that are used as stationary phases when there is an extreme pH in the mobile phase. In most of the methods used currently to separate medicinal materials, C18 columns are used, which sometimes are called ODS (octedecylsilane) or RP-18.

The more hydrophobic are the sample components the longer they stay in the column thus they are separated. The mobile phases are mixtures of water and organic polar solvents mostly methanol and acetonitrile. These mixtures contain frequently additives such as acetate, phosphate, citrate, and/or ion-pairing substances, which are surface active substances such as alkylamines as ion-pairing of anions or alkylsulfonates, ion-pairing of cations. The purpose of using such additives is to enhance efficiency and/or selectivity of the separation, mostly due to control of their retention.

The parameters that govern the retention in reversed phase systems are the following:

A. The chemical nature of the stationary phase surface [22–27].
B. The type of solvents that compose the mobile phase and their ratio [28].
C. The pH and ionic strength and additives of the mobile phase [23,29].

When the effect of these parameters on the retention of the solutes is understood it is possible to manipulate them to enhance selectivity.

***2.1.1 The Chemical Nature of the Stationary Phase [16]*** The surface of the stationary phase is described in Fig. 2. The chemical nature is determined by

the size and chemistry of hydrocarbon bonded on the silica gel surface, its bonding density (units of μmole/$m^2$), and the purity and quality of the silica gel support. As a rule, the more carbons in a bonded hydrocarbon the more it retains organic solutes (as long as similar % coverage is compared). The higher the bonding density the longer the organic solutes are retained. A column is considered relatively hydrophobic if its bonding density exceeds 3 μmole/$m^2$.

Very important modifiers of the stationary phase's surface are surface-active substances used as mobile phase's additives, acting as ion-pair reagents. These are substances such as tri-ethylamine or tetrabutylamine or hexyl, heptyl, octyl sulfonate. They are distributed between the mobile phase and the hydrophobic surface and cover it with either positive (alkylamines) or negative (alkylsulfonates) charges. This change of the surface into charged surface affects the retention significantly, especially on charged species in the sample [30].

***2.1.2 Composition of the Mobile Phase*** As a rule, the weakest solvent in reversed phase is the most polar one, water. The other polar organic solvents are considered stronger solvents, where the order of solvent strength follows more or less their dielectric properties or polarity. The less polar the solvent added to the mobile phase, the stronger it gets, shortening the retention times.

***2.1.3 PH and Ionic Strength of the Mobile Phase*** When the samples contain solutes of ionizable functional groups, such as amines, carboxyls, phosphates, phosphonates, sulfates, and sulfonates, it is possible to control their ionization degree with the help of buffers in the mobile phase. As a rule, the change of an ionizable molecule to an ion makes it more polar and less available to the stationary phase. For example, increasing the pH of the mobile phase above 4–5, which is the typical p$K_a$ of carboxyl groups, reduces the retention of carboxyl containing compounds. On the other hand, substances that contain amines whose p$K_a$ is around 8 will retain longer when the pH will be above 8. In most of the traditional silica-gel based stationary phases it is not possible to increase the mobile phase's pH above 8 due to hydrolysis of the silica gel. During the 2000s there have been developed extended pH stationary phases [31,32]

## 3 HPLC THEORY: SYSTEM SUITABILITY PARAMETERS

High performance liquid chromatography is defined as a separation of mixtures of compounds due to differences in their distribution equilibrium between two phases, the stationary phase packed inside columns and the mobile phase, delivered through the columns by high pressure pumps. Components whose distribution into the stationary phase is higher, are retained longer, and get separated from those with lower distribution into the stationary phase. The theoretical and practical foundations of this method were laid down at the end of 1960s and at the beginning of 1970s. The theory of chromatography has been used as the basis for *System Suitability* tests, which are set of quantitative criteria that test the suitability of the chromatographic

system to identify and quantify drug-related samples by HPLC at any step of the pharmaceutical analysis.

## 3.1 Retention Time ($t_R$), CAPACITY FACTOR k′, AND RELATIVE RETENTION TIME (RRT)

The time elapsed between the injection of the sample components into the column and their detection is known as the retention time ($t_R$). The retention time is longer when the solute has higher affinity to the stationary phase due to its chemical nature. For example, in reverse phase chromatography, the more lypophilic compounds are retained longer. Therefore, the retention time is a property of the analyte that can be used for its identification.

A non-retained substance passes through the column at a time $t_0$, called the void time. The *Retention Factor* or *Capacity Factor* $k'$ of an analyte is measured experimentally as shown in Fig. 3 and Eq. 1:

$$k' = \frac{t_R - t_0}{t_0} \tag{1a}$$

The *Capacity Factor* describes the thermodynamic basis of the separation and its definition is the ratio of the amounts of the solute at the stationary and mobile phases within the analyte band inside the chromatographic column:

$$k' = \frac{C_s}{C_m}\Phi \tag{1b}$$

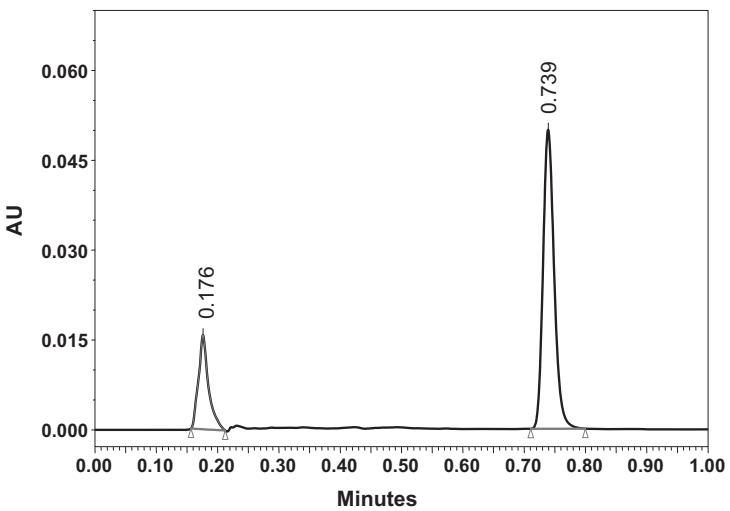


***FIGURE 3*** *Example of capacity factor calculation in LC. In this case* $t_R = 0.739$, $t_0 = 0.17$, therefore $k' = (0.739 - 0.176)/0.176 = 3.20$.

where $C_s$ is the concentration of the solute at the stationary phase and $C_m$ is its concentration at the mobile phase and $\Phi$ is the ratio of the stationary and mobile phase volumes all within the chromatographic band.

The *Retention Factor* (Eq. 1a) is used to compare the retention of a solute between two chromatographic systems, normalizing it to the column's geometry and system flow rate. The need to determine the void time can be tricky sometimes, due to the instability of the elution time of the void time marker, $t_0$, therefore, when the chromatogram is complex in nature, and one known component is always present at a certain retention time, it can be used as a retention marker for other peaks. In such cases, the ratio between the retention time of any peak in the chromatogram and the retention time of the marker is used ($t_{R(Peak)}/t_{R(Marker)}$) and referred to as the *relative retention time* (RRT). RRT is also used instead of the capacity ratio for the identification of the analyte as well as to compare its extent of retention in two different chromatographic systems.

## 3.2 Efficiency: Plate Count N and Peak Capacity $P_c$

Figure 4 describes a chromatogram with four peaks and a detected void peak. The parameters of the system suitability are displayed in the inserted table. The sharpness of a peak relative to its retention time is a measure of the system's efficiency, calculated as $N$, plate count. Band-broadening phenomena in the column such as eddy diffusion, molecular diffusion, mass-transfer kinetics, and extra-column effects reduce the efficiency of the separation [33]. The sharpness of a peak is relevant to the limit of detection and limit of quantification of the chromatographic system. The

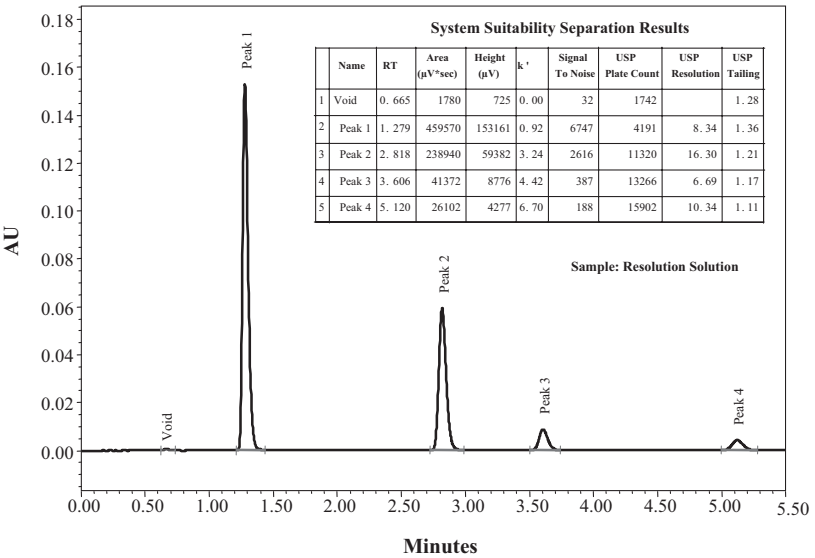


| | Name | RT | Area (µV*sec) | Height (µV) | k ' | Signal To Noise | USP Plate Count | USP Resolution | USP Tailing |
|---|---|---|---|---|---|---|---|---|---|
| 1 | Void | 0. 665 | 1780 | 725 | 0. 00 | 32 | 1742 | | 1. 28 |
| 2 | Peak 1 | 1. 279 | 459570 | 153161 | 0. 92 | 6747 | 4191 | 8. 34 | 1. 36 |
| 3 | Peak 2 | 2. 818 | 238940 | 59382 | 3. 24 | 2616 | 11320 | 16. 30 | 1. 21 |
| 4 | Peak 3 | 3. 606 | 41372 | 8776 | 4. 42 | 387 | 13266 | 6. 69 | 1. 17 |
| 5 | Peak 4 | 5. 120 | 26102 | 4277 | 6. 70 | 188 | 15902 | 10. 34 | 1. 11 |

***FIGURE 4*** *Example of a system suitability run's result.*

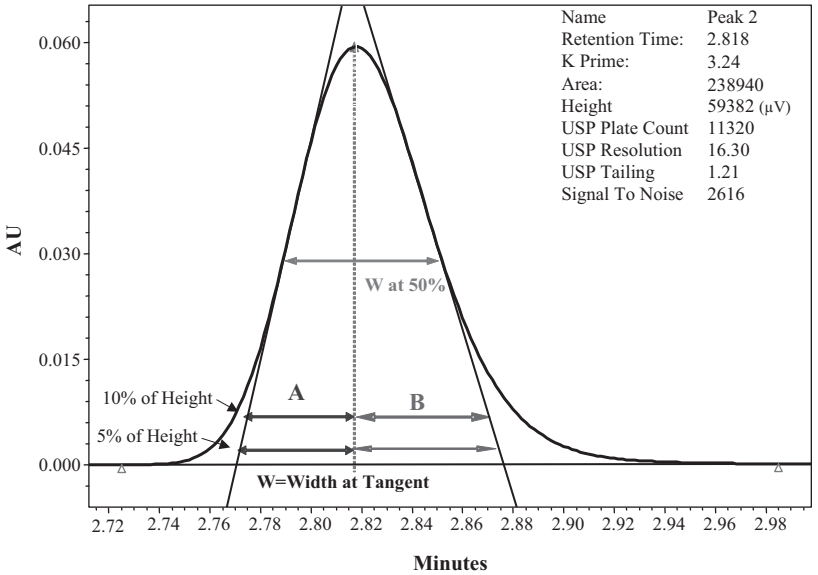


***FIGURE 5*** *System suitability measurements on a single peak.*

sharper the peak for a specific area, the better is its signal-to-noise; hence the system is capable of detecting lower concentrations. Therefore, the efficiency of the chromatographic system must be established by the system suitability test before the analysis of low concentrations that requires high sensitivity of the system, such as the analysis of drug impurities and degradation products.

The efficiency of the separation is determined by the plate count $N$ when working at isocratic conditions, whereas it is usually measured by peak capacity $P_c$ when working at gradient conditions [34]. The following equation for the plate count is used by the United States Pharmacopoeia (USP) to calculate $N$ (see Fig. 5):

$$N = 16 \times \left( \frac{t_R}{w_{(\text{Baseline})}} \right)^2 \tag{2}$$

where $w$ is measured from the baseline peak width calculated using lines tangent to the peak width at 50% height (see Fig. 5). European and Japanese Pharmacopeias use the peak width at 50% of the peak height (see Fig. 5), hence the equation becomes

$$N = 5.54 \times \left( \frac{t_R}{w_{(50\%)}} \right)^2 \tag{3}$$

*Peak capacity* $P_c$ is defined as number of peaks that can be separated within a retention window for a specific pre-determined resolution. In other words, it is the runtime measured in peak width units [34]. It is assumed that peaks occur over the

gradient chromatogram. Therefore, *peak capacity* can be calculated from the peak widths $w$ in the chromatogram as follows:

$$1 + \frac{t_g}{(1/n)\sum_{1}^{n} w} \tag{4}$$

where $n$ is the number of peaks at the segment of the gradient selected for the calculation, $t_g$. Thus peak capacity can be simply the gradient run time divided by the average peak width. The sharper the peaks the higher is the peak capacity, hence the system should be able to resolve more peaks at the selected run time as well as detect lower concentrations.

Another measure of the column's chromatographic efficiency is the *height equivalent to theoretical plate* (HETP) which is calculated from the following equation:

$$\text{HETP} = (L/N) \tag{5}$$

where $L$ is column length and $N$ is the plate count. HETP is measured in micrometer. For example, in the peak of Fig. 5 the column length was 100 mm (0.1 m) therefore HETP was 8.8 μm at the flow rate of the experiment.

The behavior of HETP as function of linear velocity has been described by various equations [35]. It is frequently called "The Van-Deemter curve," and it is frequently used to describe and characterize various chromatographic stationary phases' performance and compare them to each other [36–39]. The lower are the values of HETP, the more efficient is the chromatographic system, enabling the detection of lower concentrations due to the enhanced signal-to-noise ratio of all the peaks in the chromatogram.

### 3.3 Peak Asymmetry Factor, $A_f$ AND TAILING FACTOR T

The chromatographic peak is assumed to have a Gaussian shape under ideal conditions, describing normal distribution of the velocity of the molecules populating the peak zone migrating through the stationary phase inside the column. Any deviation from the normal distribution indicates non-ideality of the distribution and the migration process, therefore might jeopardize the integrity of the peak's integration, reducing the accuracy of the quantitation. This is the reason why USP Tailing is a peak's parameter almost always measured in the system suitability step of the analysis.

The deviation from symmetry is measured by the *Asymmetry Factor*, $A_f$ or *Tailing Factor T*. The calculation of *Asymmetry Factor*, $A_f$, is described by the following equation:

$$A_f = \frac{A_{(10\%h)}}{B_{(10\%h)}} \tag{6}$$

where $A$ and $B$ are sections in the horizontal line parallel to the baseline, drawn at 10% of the peak height as shown in Fig. 5.

The calculation of *Tailing Factor*, $T$, which is more widely used in the pharmaceutical industry, as suggested by the pharmacopeias, is described by the following equation:

$$T = \frac{A_{(5\%\mathrm{h})} + B_{(5\%\mathrm{h})}}{2A_{(5\%\mathrm{h})}} \tag{7}$$

where $A$ and $B$ are sections in the horizontal line parallel to the baseline, drawn at 5% of the peak height, as also shown in Fig. 5. The USP suggests that *Tailing Factor* should be in the range of 0.5 up to 2 to assure a precise and accurate quantitative measurement. The peaks in Figs. 4 and 5 have a Tailing Factor within the USP requirements.

### 3.4 Selectivity Factor α and Resolution Factor Rs

The separation is a function of the thermodynamics of the system. Substances are separated in a chromatographic column when their rate of migration differs, due to their different distribution between the stationary and mobile phases. The *Selectivity Factor, a*, and *Resolution Factor, Rs*, measure the extent of separation between two adjacent peaks. The *Selectivity Factor* accounts only for the ratio of the *Retention Factors, k′*, of the two peaks ($k'_2/k'_1$), whereas the *Resolution Factor, Rs*, accounts for the difference between the retention times of the two peaks relative to their width.

The equation that describes the experimental measurement of the *Resolution Factor, Rs*, is as follows:

$$Rs = \frac{t_{\mathrm{R}(2)} - t_{\mathrm{R}(1)}}{0.5(w_2 + w_1)} \tag{8}$$

where $t_{\mathrm{R}}$ is the retention time of peaks 1 and 2 respectively and $w$ is their respective peak width at the tangents' baseline (see Fig. 5). According to the pharmacopeias *Rs* should be above 1.5 for an accurate quantitative measurement. Figure 4 shows that the resolution measured between every two adjacent peaks in the chromatogram was above 1.5, therefore, it was above the minimum required.

The resolution is a critical value when working with complex samples such as drug impurities and degradation products, or when the formulation is complex and excipients might interfere with the quantitative measurements. Therefore, it is an essential part of the system suitability measurement stage before the quantitative work of these types of samples.

The sample used for the measurements of *Rs* during the system suitability runs is sometimes called *Resolution Solution*, as can be seen in Fig. 4 It usually contains the components that are the most difficult to resolve.

The theoretical description of the *Resolution Factor Rs* equation is shown in Eq. 9. It includes some of the above parameters, the plate count $N$, the selectivity $a$, and the average of the two peaks' capacity factors $k'$:

$$Rs = \frac{\sqrt{N}}{4}\left(\frac{\alpha-1}{\alpha}\right)\left(\frac{k'_{\text{ave}}}{k'_{\text{ave}}+1}\right) \tag{9}$$

It can be clearly seen from this equation that the plate count is the most effecting parameter in the increase of the chromatographic resolution. Since the plate count increases with the reduction in particle diameter, it explains the reduction in particle diameter of the stationary phase material during the last three decades of HPLC. This is also the rational behind the recent trend in HPLC, the use of sub-2 μm particle columns and the development of a specially design of ultra performance HPLC systems to accommodate such columns [36,40–42].

## 4 THE ROLE OF HPLC IN DRUG ANALYSIS

The most characteristic feature of the development in the methodology of pharmaceutical and biomedical analysis during the past 25 years is that HPLC became undoubtedly the most important analytical method for the identification and quantification of drugs, either in their active pharmaceutical ingredient (API) or in their formulations during the process of their discovery, development, and manufacturing [43–50].

Figure 6 summarizes the various stages of the pharmaceutical streamline of drug discovery, development, and manufacturing. Drug development starts with the discovery of a molecule with a therapeutic value [51,52]. This can be done by high throughput screening during which separations by fast or ultra-fast HPLC are performed [53]. At the discovery stage there can be also characterizing synthetic or natural products [54–57]. Drug metabolism and pharmacokinetics (DMPK) is the step where the candidate compounds for drug are tested for their metabolism and pharmacokinetics. The studies involve use of LC–MS or LC–MS/MS [58–62].

The next stage is the development stage, where HPLC is used to characterize products of the chemical synthesis, by analyzing the API, their impurities [63,64], and/or degradation products generated by accelerated aging [65,66]. The development of formulation requires also studies of the dissolution properties of solid dosage forms as well as assays of the pharmaceutical formulations [67]. Method for the verification of system's cleanliness during the manufacturing process is developed and used at this stage [68]. All the HPLC methods that have been finalized at the developmental stage are validated and transferred to the manufacturing laboratories for a quality control analysis [69–76].

### 4.1 The Discovery Stage

The goal in the discovery stage of drug development is to discover a new, safe and active chemical entity (NCE) that will become medication for diseases. During the

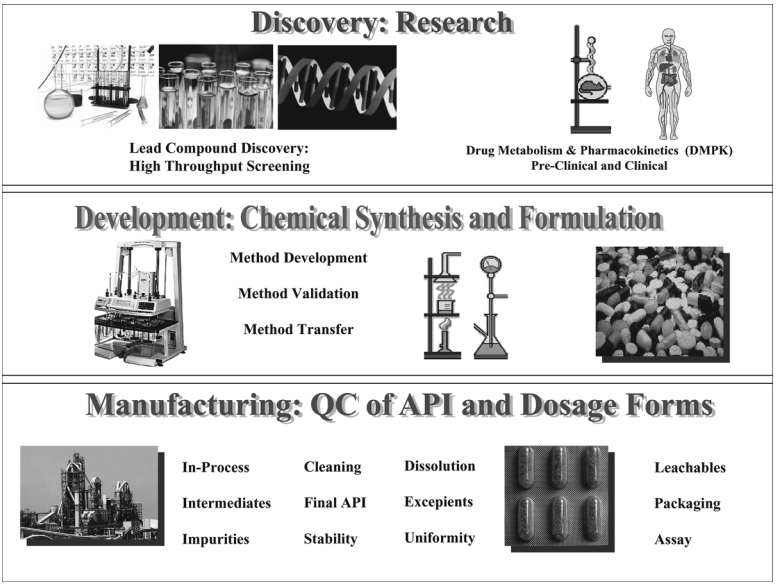


***FIGURE 6*** *Steps in a drug development and manufacturing in the pharmaceutical industry.*

last decade parallel synthesis of potential lead compounds, using combinatorial chemistry, has been done [51,77–79]. The large numbers of products created by the combinatorial chemistry are then identified by fast LC–MS methods and screened by *in vitro* bioassays and/or pharmacological or chemical tests to allow a selection of a few chosen drug candidates [53,59,61,80,81].

When the lead compounds are selected, the next steps involve pharmacological studies. Due to its high sensitivity and selectivity, HPLC coupled with tandem mass spectrometry (MS), HPLC–MS/MS, has become the predominant method in bioassays, and pharmacokinetic and metabolic studies [58–62,82]. In addition, another new development in this field has been the introduction of column's packing with ultra-fine particles ($<2\,\mu m$) enabling short columns to be used (5 cm or less) and rapid analyses (e.g., 5 min or even less than 1 min) to be carried out by UPLC (ultra performance liquid chromatography) [36,40–42,53,83–88].

## 4.2 Bioanalytical Studies

The support of bioanalytical testing includes quantitation of a target compound in various biological matrices as part of pre-clinical lead optimization by measurement its adsorption, distribution, metabolism, and excretion (ADME) [53,58–62,84,89–91]. The samples are physiological fluids such as plasma, serum, tissue extracts, and urine, typically from experimental animals or human subjects [70,84,92]. During the last decade conventional HPLC methods using fluorescence or electrochemical detection, following an elaborate sample clean-up, have been replaced by HPLC–MS/MS or U-HPLC–MS/MS methodologies. Sample clean-up is simpler and more automated

and usually involves protein precipitation and/or solid phase extraction techniques [92].

The chromatographic method is typically very rapid, gradient, or isocratic, using small columns' dimensions, such as $2.1 \times 50\,mm^2$ or less. A trace of a typically stable daughter ion signal, obtained from the fragmentation of a parent analyte is used for the detection and quantitation. Such a mode is called MRM (multiple reaction monitoring) or SRM (selective reaction monitoring), which is more stable, sensitive, and selective than either the fluorescence or electrochemical signal. Samples are usually ordered in 96 microplates for higher throughput and minimum variability of the sample vials [93]. An internal standard is used typically to minimize assay variability, due to the clean-up stage and MS/MS response, which is still less stable than UV response [94].

## 4.3 The Development Stage

The development stage is the part during which some of the HPLC methods that will be used during the subsequent manufacturing stages are developed, validated, and then transferred. The methods involve identification of the drug substance, its assay, impurities, stability, the dosage form content uniformity, dissolution, and so on. The methods are validated first, and then transferred to the quality control laboratories. Details are given in Section 3.3.

## 4.4 Method Validation

All HPLC methods used for the development of pharmaceuticals and for the determination of their quality have to be validated. In cases whereby methods from the Pharmacopeia's are used, it is not necessary to evaluate their suitability, provided that the analyses are conducted strictly according to the methods' intended use. In most other cases, especially in cases of modification of the drug composition, the scheme of synthesis or the analytical procedure, it is necessary to re-evaluate the suitability of the HPLC method to its new intended use [95].

The parameters tested throughout the method validation as defined by the ICH, USP and FDA, and other health organizations [95–97] are the following: specificity or selectivity, precision (repeatability, intermediate precision, reproducibility, or ruggedness), accuracy or trueness or bias, linearity range, limit of detection, limit of quantitation, and robustness.

The terms *selectivity* and *specificity* are often used interchangeably. The USP monograph defines selectivity of an analytical method as its ability to measure accurately an analyte in the presence of interference, such as synthetic precursors, excipients, enantiomers, and known (or likely) degradation products that might be present in the sample matrix. A method whose selectivity is verified is a "Stability Indicating Method (SIM)," for details please see Section 3.3.

*Precision* of a method is measured by injecting a series of standards and measuring the variability of the quantitative results. The measured standard deviation can be subdivided into three categories: repeatability, intermediate precision, and reproducibility (or ruggedness):

- Repeatability is obtained when one operator using one system over a relatively short time-span carries out the analysis in one laboratory. At least five or six determinations of three different matrices at two or three different concentrations should be done and the relative standard deviation calculated.
- Intermediate precision is a term that has been defined by ICH [96] as the long-term variability of the measurement process and is determined by comparing the results of a method run within a single laboratory over a number of weeks. A method's intermediate precision may reflect discrepancies in results obtained by different operators, from different instruments, with standards and reagents from different suppliers, with columns from different batches or a combination of these. Objective of intermediate precision validation is to verify that in the same laboratory the method will provide the same results once the development phase is over.
- Reproducibility (or reggedness), as defined by ICH represents the precision obtained between laboratories. The objective is to verify that the method will provide the same results in different laboratories, preparing it for the transfer to other sites.

Typical variations affecting a method's reproducibility are:

- Differences in room temperature and humidity.
- Operators with different experience and thoroughness.
- Equipment with different characteristics, such as delay volume of an HPLC system or injection modes.
- Variations in material and instrument conditions, for example different protocols of the mobile phases preparation; changes in composition, pH, flow rate of mobile phase.
- Equipment and consumables of different ages.
- Columns from different suppliers or different batches.
- Solvents, reagents, and other material with different quality.

*Accuracy* of an analytical method is the extent to which test results are close to their true value. It is measured from the result of a quantitative determination of a well-characterized known sample. The amount measured is compared to the known amount.

*Linearity* of an analytical method is determined by a series of three to six injections of five or more standards whose concentration's span is 80–120% of the expected concentration range. The response should be proportional to the concentrations of the analytes, directly or by means of a well-defined mathematical calculation. A linear regression equation, applied to the results, should have an intercept not significantly different from zero. If a significant non-zero intercept is obtained, it should be demonstrated that there is no effect on the accuracy of the method. The range of concentrations that an analytical method can be implemented on is the interval between the upper and lower levels (including these levels) that have

been demonstrated to have the appropriate precision, accuracy, and linearity. The range is normally expressed in the same units of the test results (e.g., percentage, parts per million) obtained by the analytical method.

*Limit of detection*: It is the lowest concentration of analyte in a sample that can be detected but not necessarily quantified. In chromatography the detection limit is the injected amount that results in a peak height of at least two or three times as high as the baseline noise level.

*Limit of quantitation*: It is the minimum injected amount that gives precise measurements. In chromatography it typically requires peak heights of 10–20 times higher than baseline noise at precision of <10–15% RSD between results.

*Robustness* of analytical method is a measure of its capacity to remain unaffected by small but deliberate variations in method parameters and provides an indication of its reliability during normal usage.

Once validated, the methods are ready for transfer to the manufacturing quality control laboratories. Method transfer is the last stage of the validation, whereby results are tested on both the development and the manufacturing sites. Technology transfer is especially important in the era of increasing globalization of the pharmaceutical companies [69,74,98].

## 4.5 The Manufacturing Stage

***4.5.1 Identification*** The identification test by HPLC is aimed to confirm the identity of the active pharmaceutical ingredient inside the samples of either drug substance or drug product. Two independent tests are needed, for example, one chromatographic and one spectroscopic. Therefore, a chromatographic run with a diode-array or MS detector will provide both parameters, the retention time in the chromatogram and the spectrum UV or MS of the eluting peak, matched against a known standard [99–103].

***4.5.2 Assay and Content Uniformity*** An assay of a sample containing either drug substance or drug product quantitatively measures the actual amount of the active ingredient compared to that expected in the drug substance assay, whereas in drug products the actual amount measured is verified against the label claim. The official assays are described in the pharmacopeias, such as United State's (USP) [12].

When a solid dosage form such as tablet or capsule is tested, it is usually requested to sample a composite of 10–20 units to avoid tablet-to-tablet variations. The API is extracted from the dosage form in the simplest most effective way to optimize recovery. A portion of the composite group, equivalent to an average unit dosage form weight is taken for the analysis [104,105]. Typical specifications for most drug products are 90–110% of the label claim.

A content uniformity test is similar to the drug product assay but an individual solid dosage form is tested instead of a pool of the tablets/capsules as in the assay tests. Ten tablets are taken for the assay, and if the specifications of uniformity are not met, more tablets are taken. The typical calculations are made according to a scheme given by the USP [106].

Most assays and content uniformity methods involve the use of UV detection, no extensive sample preparation is needed, the standards are usually external and highly qualified [107], typical specifications can be 98–102% purity on dried basis. Specifications for system suitability were determined by the pharmacopeia [108].

***4.5.3 Dissolution*** Drug absorption from a solid dosage form after oral administration depends on its release from the drug product, its dissolution, or solubilization under physiological conditions, and the permeability across the gastrointestinal tract. Because of the critical nature of the first two of these steps, *in vitro* dissolution may be relevant to the prediction of *in vivo* performance. Based on this general consideration, *in vitro* dissolution tests for solid oral dosage forms, such as tablets and capsules, are used to (1) assess the lot-to-lot quality of a drug product; (2) guide the developers of new formulations; and (3) ensure continuing product quality and performance after certain changes, such as changes in the formulation, the manufacturing process, the site of manufacture, and the scale-up of the manufacturing process.

The dissolution test measures the release of the drug substance from its dosage form into a dissolution bath under standardized conditions, specified by the pharmacopeias [67,109–116]. The US Food and Drug Administration issued a guidance for the industry in 1997 [95]. The tests are performed usually on two types of bath apparatus, Type I is a basket method and Type II is a paddle method. Since most drugs have absorption in the UV–VIS range, their dissolution has been traditionally measured by UV–VIS spectrometry and HPLC–UV. The advantage of HPLC over UV spectrometry is its separation capabilities, providing higher specificity and sensitivity as well as its applicability in formulations with multiple API's or very low dose.

The method is typically isocratic and very short to accommodate for the high throughput needed, therefore the columns are typically very short. The sample is collected from the dissolution apparatus at certain time intervals, designed according to its release pattern, and the released substance must follow strict release specifications. A dissolution profile of Prednisone tablets, used to calibrate a dissolution bath, is shown in Fig. 7.

***4.5.4 Drug Impurities*** Many potential impurities result from the API manufacturing process, including starting materials, isomers, intermediates, reagents, solvents, catalysts, and reaction by-products [117]; therefore, impurity control is essential during the process of drug development and manufacturing. HPLC has become the major technique in this process [63,64,118–126].

The quality of the API starting material is tested by HPLC to qualify it for the production of API that meets its specifications, as its impurities can be carried through directly or can participate in the reaction chemistry to produce significant impurities in the final product, the API. It is likely that different synthetic routes that produce different impurities may be used by different suppliers. In some cases, the synthetic route is not disclosed to the API producer; therefore, a thorough investigation of impurities in starting materials must then be conducted. This investigation includes HPLC method development, a search for potential impurities, as well as identification

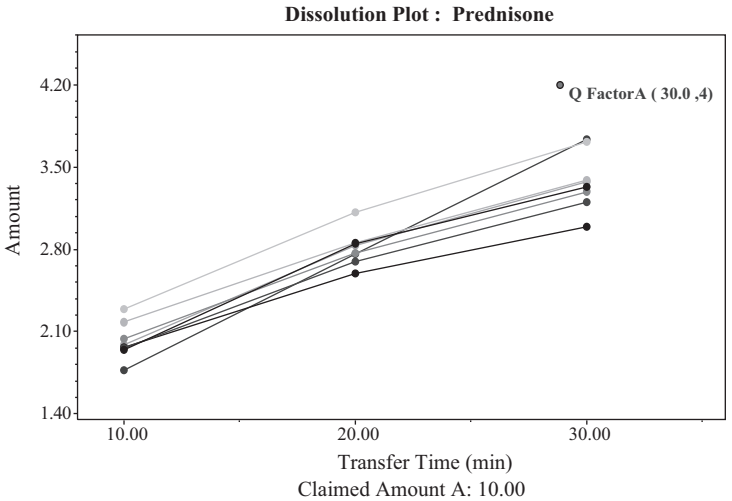


***FIGURE 7*** *Dissolution profile of Prednisone tablets.*

of significant unknown impurities and assessment of their impact on the API quality. Results of such an investigation are used to set specifications for the starting material that will assure its suitability for use in API production [117]. Similar considerations apply to the drug product. Strict international regulatory requirements have been enforced for several years as outlined in the International Conference on Harmonization (ICH) Guidelines Q3A(R), Q3B(R), and Q3C [127–129].

A typical method for the measurement of impurities can be rather complicated and difficult, as it requires high resolution to distinguish between closely related compounds, high sensitivity to detect low concentrations of the impurities, and wide dynamic range of the system to be able to detect the major components as well as their impurities and degradation products in one run. The method should have the following important attributes:

- First of all the sample preparation should be as simple as possible, to stay true to the original composition of the sample, either a drug substance or a drug product, not to lose or create an impurity during this process.
- The chromatographic separation should be optimal, to resolve all the impurities from each other and from the API's.
- The quantitative protocol involves frequently a calibration, using well-characterized impurities standards, and/or calibration against a diluted standard of the main API, taking into account their relative response factor. Sometimes the impurities are just normalized against the main peak using % area.
- The main peak is frequently at overloaded concentrations and if the normalization method of % area is used, its inclusion within the linear range should be established.

- According to the ICH guidelines and depending on the daily dose, the quantification limits of unspecified impurities can get as low as 0.05–0.1% of the main peaks. Specified known impurities can be reported at higher levels, provided their toxicology was determined.
- Method validation must include the establishment of limit of detection (to be used for rejection of small peaks) and limit of quantitation (which is used as the minimum value for linearity range specifications).
- There is a substantial effort to ensure that the main peak does not hide a co-eluted impurity by measuring peak purity (see next section).

Most methods for impurities involve a gradient of the mobile phase, to be able to resolve a complex mixture of all compounds in one run, polar as well as non-polar. In recent years, the mobile phase is programmed to be MS compatible, i.e., the buffers used are volatile, so that impurities can be identified by LC–MS at the developmental and verification stages of the analysis.

***4.5.5 Drug Stability*** Stability testing of drug substance or product is performed to ensure that their quality does not vary with time under the influence of a variety of environmental factors such as temperature, humidity, and light [130]. In addition, stability studies are typically done using HPLC methods, to determine the re-test period of a drug substance and/or the shelf life of the drug product as well as its recommended storage conditions. Guidelines to the industry are given by the ICH [129].

An HPLC method used for stability studies is termed SIM [65,66,131–137]. According to the FDA guidelines, an SIM is defined as a validated analytical procedure that accurately and precisely measures active ingredients (drug substance or drug product) free from potential interferences like degradation products, process impurities, excipients, or other potential impurities. FDA recommends that all assay procedures for stability studies will be stability indicating. A typical chromatogram of a drug substance in a stability program is shown in Fig. 8. The results are shown in Table 2.

The HPLC system used for stability indicating methods development and validation must contain 3D data capabilities such as diode-array detectors and/or mass spectrometers, to be able to detect spectral non-homogeneity within the chromatographic peaks [99–101,138]. Figures 9 and 10 present 3D data obtained by photo-diode array and mass detectors respectively. The modern UV–VIS diode-array technology is capable of evaluating specificity of the method to the analyte in question by collecting spectrum at each acquisition data point across a peak, and through software manipulations each such spectrum is compared to a reference spectrum, to determine spectral homogeneity, hence determine peak purity. Figures 11 and 12 demonstrate graphically how the software collects spectra from various points across the peaks to determine their similarity. We can see in this example that there is a minor interference only in Peak 2, as its UV spectra do not perfectly overlap. The mass spectrometer's spectra in Fig. 12 were collected by combining spectra across the peak.

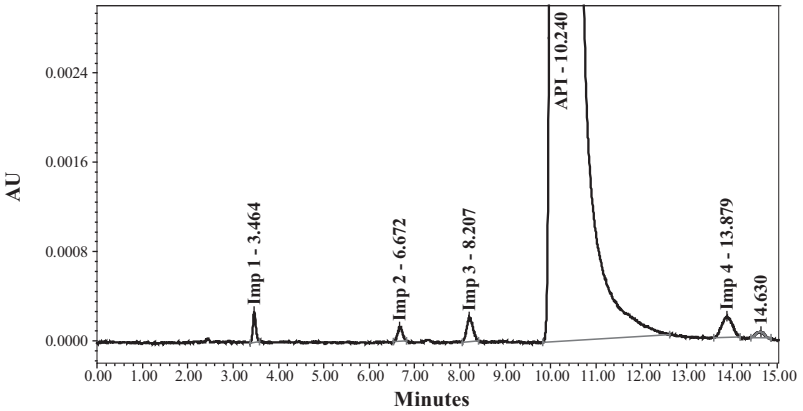


***FIGURE 8*** *A chromatogram of an API sample at a stability program. Detection: UV at 286 nm, isocratic run (70:30 water:acetonitrile with 0.1% TFA); flow rate = 1 mL/min. Results are shown in Table 2.*

The spectra confirm the suspicion and indicate the co-eluting interference in Peak 2; however, they also indicate that there is an interference in Peak 1 as well, although the UV spectra could not detect it!

Since 2006 it has been recommended that a peak purity test, based upon photodiode-array (PDA) detection, will also be based on MS whenever possible, to be more definite [99,101,138]. As already was noticed in Figs. and discussed in the previous section, the spectroscopic data obtained by the MS detector is much more reliable in evaluating the interferences co-eluting with the major chromatogram's peaks.

As noticed in Figs. 11 and 12, the UV–VIS diode-array detectors might be limited in evaluating peak purity in cases whereby the UV–VIS spectra of the main peak and its impurities are similar and/or the noise in the system is relatively high, masking the co-eluting impurities and degradants.

Three conditions must be fulfilled for the major components and their related compounds to be effective in the peak purity determination by diode-array detectors:

- They must have a UV chromophore, or some absorbance in the wavelength range selected.

**TABLE 2 Results from the HPLC Run Shown in Fig. 8 of an API After 3 Months Stability Test in Heat**

| Name | RT | Area | % Area | RRT | Signal to noise | $k'$ | USP resolution | USP plate count |
|---|---|---|---|---|---|---|---|---|
| Imp 1 | 3.46 | 1319 | 0.01 | 0.34 | 18.2 | 0.41 | | 13539 |
| Imp 2 | 6.67 | 1163 | 0.01 | 0.65 | 8.9 | 1.72 | 18.68 | 14124 |
| Imp 3 | 8.21 | 2119 | 0.01 | 0.80 | 14.1 | 2.35 | 6.40 | 17169 |
| API | 10.24 | 14505569 | 99.94 | 1.00 | 70348.3 | 3.18 | 6.65 | 14214 |
| Imp 4 | 13.88 | 3243 | 0.02 | 1.36 | 12.4 | 4.66 | 9.52 | 13251 |
| | 14.63 | 861 | 0.01 | 1.43 | 4.1 | 4.97 | 1.84 | 21690 |

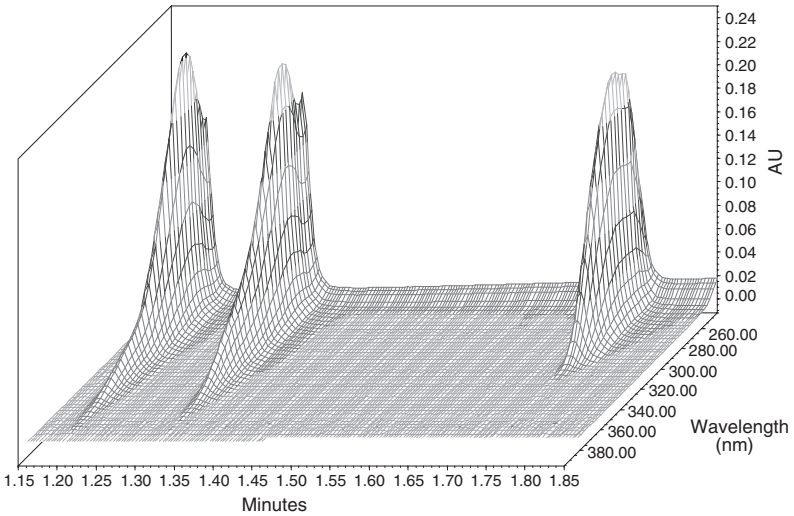


***FIGURE 9*** *3-Dimensional plot of a chromatographic run of three sulfa drugs, using a diode-array detector.*

- There should be a minimal degree of chromatographic resolution between the main peak and its co-eluting impurities (chromatographic resolution between 0.3 and 0.7).
- There must be some degree of spectral difference between the analyte peak and the co-eluting impurity.

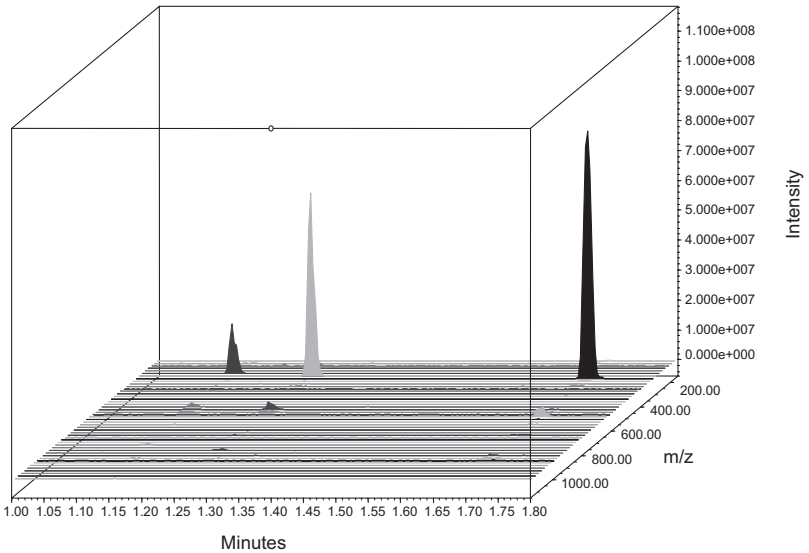


***FIGURE 10*** *3-Dimentional plot of the same three sulfa drugs as in Fig. 9, using a mass spectrometer detector.*

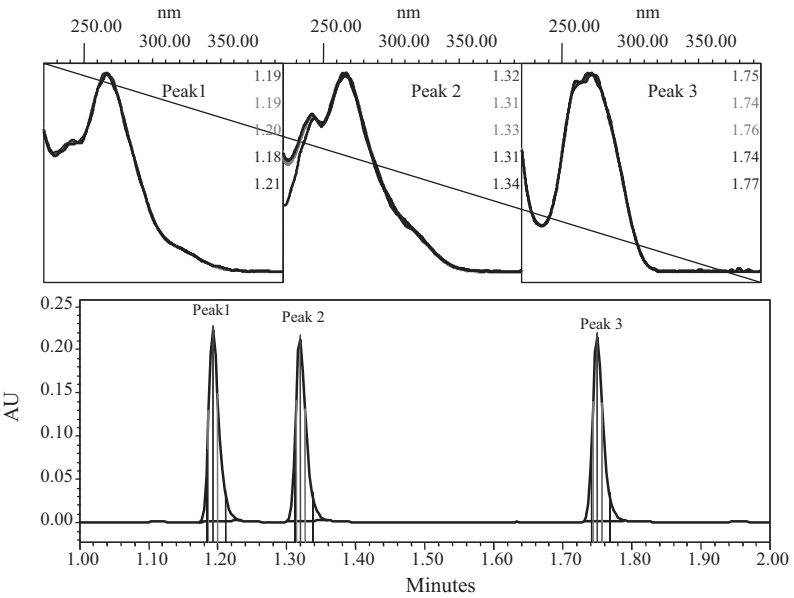


***FIGURE 11*** *UV chromatogram at 270 nm of three sulfa drugs shown in Fig. 9, and a plot of UV spectra collected from each peak's apex, inflection, and two offset points, to demonstrate comparison of spectra collected at various points across the peak.*

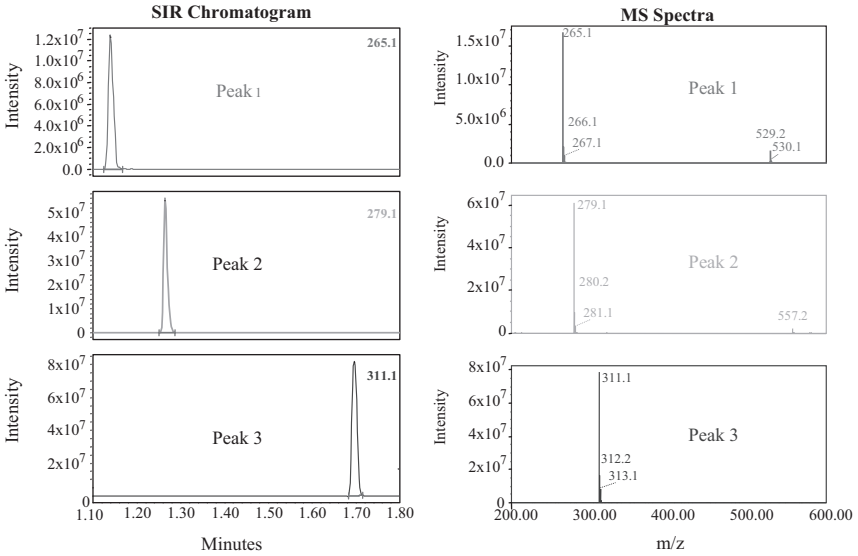


***FIGURE 12*** *SIR chromatograms of the same sulfa drugs as in Fig. 10 and their corresponding mass spectra, obtained by combining spectra across the entire peak.*

Stress testing is carried out on a single batch of the drug substance and/or the drug product, using forced degradation of the drug substance [65]. The results of such studies can help identify the likely degradation products, which can in turn help establish the degradation pathways and the chemical stability of the molecule and to validate the stability indicating capabilities of the analytical procedures used.

The nature of the stress testing will depend on the individual drug substance and the type of drug product involved. It usually includes the effect of temperatures, humidity, oxidation, and photolysis on the drug substance. The testing also evaluates the vulnerability of the drug substance to hydrolysis across a wide range of pH values when in solution or suspension. Photo-stability testing is also an integral part of a stress testing [65,66,124,129,131,135,137].

***4.5.6 Cleaning Validation*** Cleaning validation tests are performed to assure the cleanliness of the pharmaceutical manufacturing equipment, such as blender, tablet press, and so on. There are strict specifications for the maximum allowable amounts of the residues, according to their toxicology and daily dosage [139]. The analytical methods must be extremely sensitive, to be able to track the residues of the analytes, their degradants, or impurities, other contaminations or the cleaning reagents themselves [85,140–147]. The methods must be very sensitive and short, as the entire manufacturing process depends on the approval of the cleanliness of apparatus. Due to the enhanced sensitivity as well as speed of analysis, the use of ultra performance liquid chromatography has immediately adopted in the cleaning validation tests [85].

***4.5.7 In-Process Control*** In-process control method monitors the progress in the manufacturing of an active pharmaceutical ingredient or its formulation. HPLC is frequently employed for the in-process assay of active pharmaceutical ingredient synthesis [68,72,148,149]. The results signal the production chemist/pharmacist whether to proceed with a subsequent unit operation. The decision about whether to use an in-process control test in a manufacturing is established during process development and is based on scientific judgment. In recent years, the FDA issued a recommendations for process analytical technology (PAT) [150]. The guidance intended to describe a regulatory framework that will encourage the voluntary development and implementation of innovative pharmaceutical development, manufacturing, and quality assurance. The new LC ultra performance/rapid technologies are emerging as the technology of choice for the in-process control assay due to their extended efficiency, sensitivity, and speed of the analysis [68,151].

## 5 SPECIAL TOPICS

### 5.1 Ultra High Performance Liquid Chromatography

In order to enhance chromatographic performances in terms of efficiency and speed, LC has recently evolved thanks to the development of short columns packed with

small particles (sub-2 μm), working at high pressures (Above 1000 bar) [40,152–155]. This approach has been described 30 years ago according to the fundamental chromatographic equations [156]. However, systems and columns compatible with such high pressures have been introduced into the market since 2004 only. Advantages of small particles working at high pressure have been vividly discussed in the literature, in terms of sensitivity, efficiency, resolution, and analysis time, showing how systems working at a maximum pressure of 1000 bar give reliable and reproducible results. The implementation of the new technology in the pharmaceutical and life sciences is soaring and it becomes the next benchmark technology for the near future [41,42,84,85,152,157].

There are important experimental parameters one should be aware of while reducing the particle size [158]. First and foremost is the column pressure. Equation 10 shows the dependence of column head pressure on a number of experimental parameters including the particle size. As can be seen in Eq. 10, the pressure is inversely proportional to the square of the particle size.

$$P = \frac{\Phi L \eta \mu}{100 \times d_p^2} \tag{10}$$

where $P$ is the pressure drop, $\Phi$ is the flow resistance parameter, $\eta$ is viscosity (mPa/s), $L$ is column length (mm), $\mu$ is the linear velocity (mm/s), and $d_p$ particle size (μm). So when the particle size is halved, the pressure goes up by a factor of 4. However, often for fast separations, the column length is also reduced, so the pressure increase is not nearly as high as one would expect, because pressure is proportional to column length.

If longer columns such as 100 or 150 mm are required to achieve higher resolution and sensitivity, then higher pressure pumps are required. Currently, there are commercial HPLC systems with upper pressure limits as high as 15 000 psi. It should be noted that the total pressure that the HPLC system experiences is the sum of the column and the instrument backpressures. The latter results when small internal diameter capillaries are used in the flow paths to reduce extra-column effects and minimize the gradient delay volume [33]. As the flow rate increases, the back pressure due to these capillaries increases proportionally.

An experimental parameter that requires the most careful consideration when reducing the column length and especially internal diameter is the extra-column band broadening. Some of the modern ultra-fast LC columns are only 1.5-cm long with an internal diameter of 2.1 mm. Such a column has a total void volume of around 33 μL. Many conventional HPLC instruments were developed for typical 15 and 25 cm × 4.6 mm analytical columns, whose void volumes are ~1.6 and ~2.6 mL, respectively. A 1.5 cm × 2.1 mm column packed with 1.8-μm particles often produces a few microliters peaks, which requires that extra-column band broadening will be minimized to accommodate for the true advantages of these small columns. Some modern liquid chromatographs have been designed or modified to minimize the extra-column

effects [157]. For others, upgrade kits are available to modify some older HPLC instruments to work satisfactorily with sub-2-μm columns.

Some other experimental parameters that must be taken into account when running fast and ultrafast separations in LC are as follows [86,159]:

- *Detector time constant*: Peaks can be only a second or two wide when short narrow-bore columns run at high flow rates. A time constant that is too slow will make the peaks artificially broad because the detector cannot keep up with the rapidly changing signal.
- *Data sampling (acquisition) rate*: Data system needs enough data points, minimum of 20 across a peak to define the peak as proper for integration of area, determine retention time, and so on.
- *Autosampler cycle time*: Separations whose run time is less than a minute or two, throughput can be slowed down by a slow autosampler.
- *Gradient delay volume*: If the volume between the points, where the gradient forms (mixer), to the column head is too large, gradient profiles will be compromised because gradient has not reached column in time.

When transferring methods from HPLC to UPLC there is a scale-down process of the sample load, flow-rate, and gradient times [86,159]. Sample load must be reduced proportionally to the columns volumes, a typical factor can be from ~1.5 mL (4.6 × 150 mm column) to ~0.15 mL (2.1 × 100 mm column). Equation 11 describes the calculations required for such a scale-down.

$$M_{\mathrm{UPLC}} = M_{\mathrm{HPLC}} \frac{L_{\mathrm{UPLC}}}{L_{\mathrm{HPLC}}} \frac{\mathrm{ID}_{\mathrm{UPLC}}}{\mathrm{ID}_{\mathrm{HPLC}}} \tag{11}$$

where $M$ is the sample load in mg, $L$ is column's length, and ID is internal column's diameter.

It should be noted that the separation mechanisms in UPLC compared to HPLC are still the same; chromatographic principles are maintained while speed, sensitivity, and resolution is improved. This all supports easier method transfer from HPLC to UPLC and its revalidation. The main advantages are significant reduction of analysis time and flow-rates (which means also reduction in solvent consumption!), enhanced peak capacity and sensitivity, which means that complex analytical determination of pharmaceutical preparations can be accomplished much easier within a reasonable time. Moreover, time spent with new method development, optimization, and validation is significantly saved.

A typical chromatogram obtained in the UPLC system is shown in Fig. 13. A mixture of alkylphonone is chromatographed on a 2.1 × 50 $mm^2$ Acquity BEH C18 column packed with 1.7 μM particles at room temperature using a gradient of 5% MeCN to 95% Acetonitrile within 3 min at a flow rate of 0.5 mL/min. Re-equilibration time was 1.5 min and solvent consumption was 2.5 mL/run. The spectacular resolution and peak widths in such a short time are noticeable.

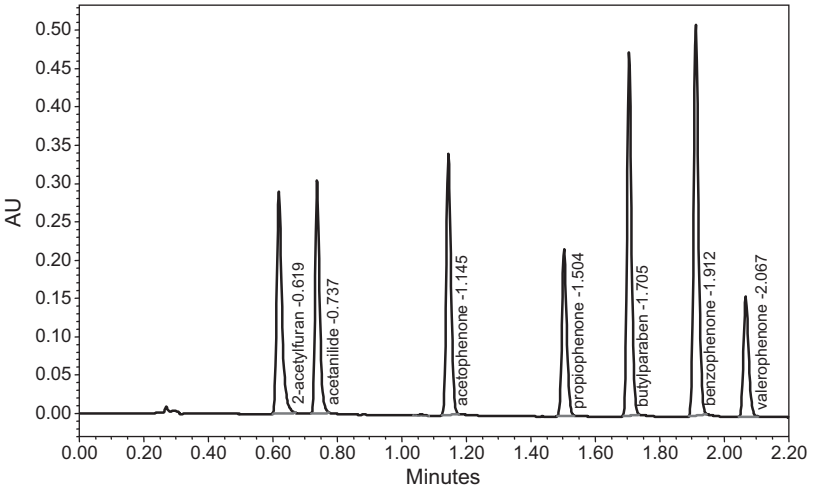


***FIGURE 13*** *Mixture of alkylphenones injected into waters' acquity UPLC using acquity BEH C18 2.1 × 50 mm² 1.7 μM.*

## 5.2 Preparative Chromatography

Preparative chromatography is used to isolate and purify compounds [6,73,160–166]. Typically this type of separations requires larger columns and the chromatograms generally display broader peaks, due to the increased column load. The separation is optimized for higher throughput (weight/time), then sample load is maximized on small analytical columns, and finally quantities are scaled up to the preparative column according to the desired purity and recovery.

The parameters that will govern efficiency of the preparative separation are the plate count, the capacity factor $k'$, and the selectivity between the solutes (see Section .2). As selectivity, $a$, between two adjacent peaks of interest is improved, a better separation between them at higher sample loads is obtained, getting better product purity. Mobile phase changes cause this change in selectivity, as well as selecting the appropriate stationary phase.

Peak shape is no longer a concern, only the collected product's purity. The scaling up of the mass load can be done according to the column's volume ratio as is shown in Eq. 11, but instead of ratio between UPLC and HPLC, the ratio is between the analytical column and the preparative one.

A rational approach to scaling up HPLC runs, using adsorption isotherms measurements and calculations has been described [167]. In the pharmaceutical work, preparative chromatography is used mostly to isolate drug impurities for the purpose of their identification and/or using them as working standards for impurities determination [168]. Purification of compounds during drug development and manufacturing has been done in recent years using UV in line with mass spectrometer, to achieve a mass-directed purification, whereby the mass spectrometer is used to determine the cutoff of the purified compound and direct the pure compound into its collection vessel [169].

### 5.3 Chiral Chromatography

The biological activity of chiral substances often depends upon their stereochemistry, since the living body is a highly chiral environment [170–172]. A large percentage of commercial and investigational pharmaceutical compounds are enantiomers, and many of them show significant enantioselective differences in their pharmacokinetics and pharmacodynamics. The importance of chirality of drugs has been long recognized, and the consequence of using them as racemates or as enantiomers has been frequently discussed in the pharmaceutical literature [173–175]. With increasing evidence of problems related to stereoselectivity in drug action, chiral chromatography has become the major tool in their analysis [171,173–184]. Great efforts have been devoted to the development of better methodology for enantioselective chromatography during the past two decades, and have resulted in variety of chiral stationary phases. Chiral agents were derivatized and immobilized on the surface of the support (silica gel mostly), and served as the *in situ* chiral discriminators during the chromatographic process.

The chiral stationary phases can be categorized into the following classes:

A. Chiral affinity by proteins (serum albumin, a1-acid glycoprotein, ovomucoid, and chymotrypsin) [185].
B. Stereoselective access to helical chiral polymers (derivatized or free polysaccharides) [186,187].
C. Steric interactions between p-donor p-acceptor type of chiral aromatic amide groups (Pirkle) [188].
D. Host–guest interactions inside chiral cavities (cyclodextrins, crown ethers, macrocyclic antibiotics, and imprinted polymers) [189–192].

Major developments in chiral stationary phases have led to dramatic advances in enantioselective chromatography [175]. Chromatographic enantioseparation on chiral stationary phases results from energy differences between transient diastereomeric complexes formed by the solute with the chiral agent on the surface of the chiral stationary phase interactions. These differential interactions are directly influenced by the mobile phase environment. The solvents in the mobile phase can be either normal phase, or aqueous reversed phase's [193] or polar organic [187].

During recent years the trend in the pharmaceutical manufacturing of drugs containing chiral centers is to develop them as single enantiomers rather than racemates due to the complexity of the tests and research involved in the use of racemates as API's [64].

## 6 CONCLUSION

HPLC technology has matured to the extent that almost any existing organic compound can be analyzed by an existing method that can be found in the analytical literature, such as professional journals, protocol books such as Pharmacopeia's or

AOAC manuals. The most remarkable change in the pharmacopeias in the past 25 years has been the increasing importance of HPLC technology in the analysis of all aspects of drug development and manufacturing.

## REFERENCES


1. Velagaleti R, Burns P, Gill M. Drug Inf J 2003; 37: 407–438.
2. Dong MW. Modern HPLC for practicing scientists. New Jersey: John Wiley & Sons; 2006.
3. Rochet JC. Am J Pharm Educ 2006; 70.
4. Katz E. Handbook of HPLC. CRC Press; 1998.
5. Farb D, Luttrell A, Kirsch R. Pharmaceutical Quality Control Lab Guide book. University Of Health Care; 2005.
6. Wellings D. A practical handbook of preparative HPLC. Elsevier Science; 2006.
7. Winslow P, Meyer R. Compliance handbook for pharmaceuticals, medical devices, and biologics. 2004.
8. Hage DS. Handbook of affinity chromatography. Taylor & Francis Group; 2006.
9. Walker JM. The protein protocols handbook. Humana Pr Inc; 2002.
10. Ahuja S, Dong MW. Handbook of pharmaceutical analysis by HPLC. Elsevier Academic Press; 2005.
11. Ahuja S, Scypinski S. Handbook of modern pharmaceutical analysis. Academic Press; 2001.
12. USP-NF. *United State Pharmacopeia National Formulary*. United State Pharmacopeial Convention, Inc; 2005.
13. Levin S. J Liq Chromatogr Relat Technol 2004; 27: 1353–1376.
14. Stella C, Rudaz S, Gauvrit JY, Lant'ri P, Huteau A, Tchapla A, Veuthey JL. J Pharm Biomed Anal 2007; 43: 89–98.
15. Claessens HA, van Straten MA. J Chromatogr A 2004; 1060: 23–41.
16. Marchand DH, Snyder LR, Dolan JW. J Chromatogr A 2008; 1191: 2–20.
17. Luo H, Ma L, Zhang Y, Carr PW. J Chromatogr A 2007.
18. Euerby MR, Petersson P. J Chromatogr A 2003; 994: 13–36.
19. Haghedooren E, Diana J, Noszal B, Hoogmartens J, Adams E. Talanta 2007; 71: 31–37.
20. Van Gyseghem E, Jimidar M, Sneyers R, Redlich D, Verhoeven E, Massart DL, Vander Heyden Y. J Chromatogr A 2004; 1042: 69–80.
21. Dehouck P, Visky D, Heyden YV, Adams E, Kovacs Z, Noszl B, Massart DL, Hoogmartens J. J Chromatogr A 2004; 1025: 189–200.
22. O'Gara JE, Alden BA, Walter TH, Petersen JS, Niederlaender CL, Neue UD. Anal Chem 1995; 67: 3809–3813.
23. Neue UD, Walter TH, Alden BA, Jiang Z, Fisk RP, Cook JT, Glose KH, Carmody JL, Grassi JM, Cheng Y. Am Lab 1999; 31: 36–39.
24. Neue UD, Tran KV, M'ndez A, Carr PW. J Chromatogr A 2005; 1063: 35–45.
25. Neue UD. J Sep Sci 2007; 30: 1611.

26. Neue UD, O'Gara JE, M'ndez A. J Chromatogr A 2006; 1127: 161–174.
27. Cheng YF, Walter TH, Lu Z, Iraneta P, Alden BA, Gendreau C, Neue UD, Grassi JM, Carmody JL, O Gara JE. LC GC 2000; 18: 1162–1173.
28. Neue UD, Phoebe CH, Tran K, Cheng YF, Lu Z. J Chromatogr A 2001; 925: 49–67.
29. Neue UD, Mendez A, Iran KV, Diehl DM. HPLC made to measure: a practical handbook for optimization. 2006.
30. Gilar M, Fountain KJ, Budman Y, Neue UD, Yardley KR, Rainville PD, Russell Ii RJ, Gebler JC. J Chromatogr A 2002; 958: 167–182.
31. Jenkins K, Diehl D, Morrison D, Mazzeo J. LC GC North Am 2005; 92–93.
32. Fountain KJ, Morrison D, Diehl DM, Martin J. LC GC North Am 2007; 64–65.
33. Kirkland JJ, Yau WW, Stoklosa HJ, Dilks CH, Jr. J Chromatogr Sci 1977; 15: 303–316.
34. Neue UD. J Chromatogr A 2005; 1079: 153–161.
35. Usher KM, Simmons CR, Dorsey JG. J Chromatogr A 2008.
36. Wren SAC, Tchelitcheff P. J Chromatogr A 2006; 1119: 140–146.
37. Neue UD, Kele M. J Chromatogr A 2007; 1149: 236–244.
38. Jones MD, Plumb RS. J Sep Sci 2006; 29: 2409–2420.
39. Gritti F, Guiochon G. J Chromatogr A 2006; 1128: 45–60.
40. Swartz M. J Liq Chromatogr Relat Technol 2005; 28: 1253–1263.
41. Novakova L, Matysova L, Solich P. Talanta 2006; 68: 908–918.
42. King S, Peter J, Stoffolano ER, Eichhold TE, Ii SHH, Baker TR, Richardson EC, Wehmeyer KR. LC GC North Am 2005; 36–39.
43. Schirmer R. Modern methods of pharmaceutical analysis. CRC Press; 1991.
44. McGill J, Holly D. Am J Pharm Educ 1996; 60: 370–374.
45. Tang L, Fitch W, Alexander M, Dolan J. Anal Chem 2000; 72: 5211–5218.
46. Lee D, Webb M. Pharmaceutical analysis. Blackwell Science Ltd; 2003.
47. Lunn G, Wiley J. HPLC methods for recently approved pharmaceuticals. Wiley-Interscience; 2005.
48. Gorog S. Trends Analyt Chem 2006; 25: 755–757.
49. Gorog S. Trends Analyt Chem 2007; 26: 12–17.
50. Kazakevich YV, LoBrutto R. HPLC for pharmaceutical scientists. Wiley-Interscience; 2007.
51. Bleicher K, Bohm H, Muller K, Alanine A. Nat Rev Drug Discov 2003; 2: 369–378.
52. Daniel B. HPLC for pharmaceutical scientists. 2007.
53. Castro-Perez J, Plumb R, Granger JH, Beattie I, Joncour K, Wright A. Rapid Commun Mass Spectrom 2005; 19: 843–848.
54. Butler MS. J Nat Prod 2004; 67: 2141–2153.
55. Smyth WF, McClean S, Hack CJ, Ramachandran VN, Doherty B, Joyce C, O'Donnell F, Smyth TJ, O'Kane E, Brooks P. Trends Analyt Chem 2006; 25: 572–582.
56. Butler MS. Nat Prod Rep 2005; 22: 162–195.
57. Lam KS. Trends Microbiol 2007; 15: 279–289.
58. Chu I, Anomeir A. Curr Drug Metab 2006; 7: 467–477.
59. Kassel DB. Using mass spectrometry for drug metabolism studies. 2005.
60. Lappin G, Garner RC. Expert Opin Drug Metab Toxicol 2005; 1: 23–31.

61. Kassel DB. Curr Opin Chem Biol 2004; 8: 339–345.
62. Wang J, Urban L. Drug Discov 2004; 73.
63. Ng L, Lunn G, Faustino P. Analysis of drug impurities. 2007.
64. Argentine M, Owens P, Olsen B. Adv Drug Deliv Rev 2007; 59: 12–28.
65. Bakshi M, Singh S. J Pharm Biomed Anal 2002; 28: 1011–1040.
66. Consultant P, Raleigh N. Drug stability: principles and practices. 2000.
67. Rosanske, T. W., Brown, C. K., Christopher, M. R. a. T. W. R., Progress in Pharmaceutical and Biomedical Analysis. Vol. 3. Elsevier; 1996. p. 169.
68. Gomis DB, Nunez NS, Enguita PB, Abrodo PA, Alvarez MDG. J Liq Chromatogr Relat Technol 2006; 29: 931–948.
69. Rozet E, Dew W, Morello R, Chiap P, Lecomte F, Ziemons E, Boos KS, Boulanger B, Crommen J, Hubert P. J Chromatogr A 2008; 1189: 32–41.
70. Kataoka H. Trends Analyt Chem 2003; 22: 232–244.
71. Koh HL, Yau WP, Ong PS, Hegde A. Drug Discov Today 2003; 8: 889–897.
72. Wrezel PW, Chion I, Hussain MS. LC GC North Am 2004; 22: 1006–1009.
73. Brandt, A, Schering, AG, Berlin, G, Kueppers, S. LC GC Eur 2002.
74. Varma S. Qual Assur J 2001; 5: 85–92.
75. Jimidar, MI, De Smet, M. HPLC method development for pharmaceuticals. 2007.
76. Gibson, M. Technology transfer introduction and objectives, PDA. 2005.
77. Gardner CR, Almarsson O, Chen H, Morissette S, Peterson M, Zhang Z, Wang S, Lemmo A, Gonzalez-Zugasti J, Monagle J. Comput Chem Eng 2004; 28: 943–953.
78. Maurer H. Comb Chem High Throughput Screen 2000; 3: 467–480.
79. Schultz L, Garr CD, Cameron LM, Bukowski J. Bioorg Med Chem Lett 1998; 8: 2409.
80. Corti G, Maestrelli F, Cirri M, Zerrouk N, Mura P. Eur J Pharm Sci 2006; 27: 354–362.
81. Alsenz J, Kansy M. Adv Drug Deliv Rev 2007; 59: 546–567.
82. Korfmacher WA. Drug Discov Today 2005; 10: 1357–1367.
83. Plumb R, Jones M, Rainville P, Nicholson J. J Chromatogr Sci 2008; 46: 193–198.
84. Pedraglio S, Rozio MG, Misiano P, Reali V, Dondio G, Bigogno C. J Pharm Biomed Anal 2007; 44: 665–673.
85. Fountain KJ, Wingerden M, Diehl DM. LC GC Mag North Am Solut Sep Sci 2007; 37: 66–67.
86. de Villiers A, Lestremau F, Szucs R, G'l'bart S, David F, Sandra P. J Chromatogr A 2006; 1127: 60–69.
87. Nardo C. 2005.
88. Johnson KA, Plumb R. J Pharm Biomed Anal 2005; 39: 805–810.
89. Korfmacher WA, Cox KA, Bryant MS, Veals J, Ng K, Lin C, Watkins R. Drug Discov Today 1997; 2: 532–537.
90. Wickremsinhe ER, Singh G, Ackermann BL, Gillespie TA, Chaudhary AK. Curr Drug Metab 2006; 7: 913–928.
91. Kamel A, Prakash C. Curr Drug Metab 2006; 7: 837–852.
92. Alnouti Y, Csanaky IL, Klaassen CD. J Chromatogr B 2008; 873: 209.
93. Mallet CR, Lu Z, Fisk R, Mazzeo JR, Neue UD. Rapid commun Mass Spectrom 2003; 17: 163–170.

94. Li R, Dong L, Huang J. Anal Chim Acta 2005; 546: 167–173.
95. CDER: Center for Drug Evaluation and Research. U.S. Department of Health and Human Services, Food and Drug Administration; 1997.
96. ICH-Q2B. Geneva; 1996.
97. USP-26-NF21. 2003.
98. Gibson DV, Williams F. Technology transfer. Newbury Park, CA: Sage Publications; 1990.
99. Li W, Hu C. J Chromatogr A 2008.
100. Gilliard JA, Ritter C. J Chromatogr A 1997; 786: 1–11.
101. Lindholm J, Westerlund D, Karlsson K, Caldwell K, Fornstedt T. J Chromatogr A 2003; 992: 85–100.
102. Swartz ME. Chirality 1996; 867: 76.
103. Rivier L. Applications of LC-MS in toxicology 2006.
104. Lambropoulos J, Spanos GA, Lazaridis NV. J Pharm Biomed Anal 1999; 19: 793–802.
105. Berman J, Planchard JA. Drug Dev Ind Pharm 1995; 21: 1257–1283.
106. http://www.usp.org/USPNF/compendialTools.html
107. Williams RL. J Pharm Biomed Anal 2006; 40: 3.
108. Renger B. J Chromatogr B: Biomed Sci Appl 2000; 745: 167.
109. Aiache JM. J Pharm Biomed Anal 1990; 8: 499.
110. Zahirul M, Khan I. Int J Pharm 1996; 140: 131.
111. Qureshi SA. Drug Inf J 1996; 30: 1055–1061.
112. Shiu GK. Drug Inf J 1996; 30: 1045–1054.
113. Castellano P, Vignaduzzo S, Maggio R, Kaufman T. Anal Bioanal Chem 2005; 382: 1711–1714.
114. Bateman S, Paul W, Alan T, Colin P. Encyclopedia of analytical science. Oxford: Elsevier; 2005. p. 101.
115. Azarmi S, Roa W, Lybenberg R. Int J Pharm 2007; 328: 12.
116. Deng G, Ashley A, Brown W, Eaton J, Hauck W, Kikwai L, Liddell M, Manning R, Munoz J, Nithyanandan P. Pharm Res 2008; 25: 1100–1109.
117. Gavin PF, Olsen BA, Wirth DD, Lorenz KT. J Pharm Biomed Anal 2006; 41: 1251–1259.
118. Smith RJ, Webb ML. Analysis of drug impurities. Blackwell Publishing; 2007.
119. Ahuja S. Adv Drug Deliv Rev 2007; 59: 3–11.
120. Kovaleski J, Kraut B, Mattiuz A, Giangiulio M, Brobst G, Cagno W, Kulkarni P, Rauch T. Adv Drug Deliv Rev 2007; 59: 56–63.
121. Basak A, Raw A, Al Hakim A, Furness S, Samaan N, Gill D, Patel H, Powers R, Yu L. Adv Drug Deliv Rev 2007; 59: 64–72.
122. Qiu F. J Liq Chromatogr Relat Technol 2007; 30: 877–935.
123. Olsen B, Castle B, Myers D. Trends Analyt Chem 2006; 25: 796–805.
124. Baertschi S. Trends Analyt Chem 2006; 25: 758–767.
125. Nageswara Rao R, Nagaraju V. J Pharm Biomed Anal 2003; 33: 335–377.
126. Roy J. AAPS PharmSciTech 2002; 3: 1–8.
127. ICH. International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use; 2006.

128. ICH. International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use; 2005.
129. ICH. International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use; 2003.
130. Carstensen J, Rhodes C. Drug stability: principles and practices. Informa Healthcare; 2000.
131. Kumar V, Bhutani H, Singh S. J Pharm Biomed Anal 2007; 43: 769–773.
132. Mohammadi A, Mehramizi A, Moghaddam F, jabarian L, pourfarzib M, Kashani H. J Chromatogr B 2007; 854: 152–157.
133. Shah R, Bryant A, Collier J, Habib M, Khan M. Int J Pharm 2008; 360: 77–82.
134. Mohammadi A, Rezanour N, Ansari Dogaheh M, Ghorbani Bidkorbeh F, Hashem M, Walker R. J Chromatogr B 2007; 846: 215–221.
135. Xu QA, Trissel LA,American Pharmacists Association. Stability-indicating HPLC methods for drug analysis. American Pharmaceutical Association; 1999.
136. Chan ECY, Wee PY, Ho PC. Clin Chim Acta 1999; 288: 47–53.
137. Swartz M, Krull I. *Chromatography Online*; 2005.
138. Fourneron JD. J Chromatogr Sci 2001; 39.
139. APICActive Pharmaceutical Ingredients Committee. 1999.
140. Yang P, Burson K, Feder D, Macdonald F. Pharm Technol 2005.
141. Schmidt AH. J Liq Chromatogr Relat Technol 2006; 29: 1663.
142. Zayas J, Colnn H, Garced O, Ramos LM. J Pharm Biomed Anal 2006; 41: 589–593.
143. Forsatz B, Snow N. LC GC North Am 2007; 25: 960.
144. Kathiresan K, Kannan K, Valliappan K, Karar PK, Manavalan R. Asian J Chem 2007; 19: 5034.
145. Baran PA, Bielejewska A, Glice MM, Beczkowicz H, Maruszak W, Kosmacinska B, Golebiewski P. Przemysl Chemiczny 2007; 86: 747.
146. Kathiresan K, Kannan K, Kannappan N, Venkatesan P, Shenoy KRP, Manavalan R. Asian J Chem 2007; 19: 4438.
147. Resto W, Hernndez D, Rey R, Colnn H, Zayas J. J Pharm Biomed Anal 2007; 44: 265–269.
148. Laviana L, Llorente I, Bayod M, Blanco D. J Pharm Biomed Anal 2003; 32: 167–174.
149. Moes JJ, Ruijken MM, Gout E, Frijlink HW, Ugwoke MI. Int J Pharm 2008; 357: 108.
150. Guidance F.CDER: Center for Drug Evaluation and Research, editor. U.S. Department of Health and Human Services, Food and Drug Administration; 2004.
151. Anurag S, et al. Biotechnol Bioeng 2008; 100: 306–316.
152. Swartz ME. LC-GC 2005; 23: 8–14.
153. Neue UD, Smith B, Johnson KA. J Sep Sci 2007; 30: 1158–1166.
154. Mazzeo JR, Neue UD, Kele M, Plumb RS. Anal Chem (Washington, DC) 2005; 77: 460–467.
155. Diehl DM, Tran K, Grumbach ES, Neue UD, Mazzeo JR. LC GC North Am 2003; 67–68.
156. Giddings JC,P. 1965.
157. Jerkovich AD, LoBrutto R, Vivilecchia RV. LC/GC North Am 2005; 15–21.
158. Majors RE.

159. Guillarme D, Nguyen DTT, Rudaz S, Veuthey JL. Eur J Pharm Biopharm 2008; 68: 430–440.
160. Mazza CB, Lu Z, Tran K, Sirard T, Mazzeo J, Neue UD. LC GC North Am 2002; 20: 1052–1061.
161. Guiochon G. J Chromatogr A 2002; 965: 129–161.
162. Rathore AS, Dekker M. Scale-up and optimization in preparative chromatography: principles and biopharmaceutical applications. CRC Press; 2003.
163. Neue UD, Wheat TE, Mazzeo JR, Mazza CB, Cavanaugh JY, Xia F, Diehl DM. J Chromatogr A 2004; 1030: 123–134.
164. Schmidt-Traub H. Preparative chromatography of fine chemicals and pharmaceutical agents. Wiley-VCH; 2005.
165. Scott C. BioProcess Int 2005.
166. Kuesters E. HPLC for pharmaceutical scientists. 2007.
167. Guiochon G, Shirazi DG, Felinger A, Katti AM. Fundamentals of preparative and nonlinear chromatography. Academic Press; 2006.
168. Terfloth G. Analysis of drug impurities. 2007.
169. Kagan M, Chlenov M, Bach A, McConnell O. J Liq Chromatogr Relat Technol 2004; 27: 1817–1834.
170. Ariens EJ. Trends Pharmacol Sci 1986; 7: 200–205.
171. Caldwell J. Chem Ind 1995; 176–179.
172. Lien EJ. J Drug Target 1995; 2: 527–532.
173. Krzek J, Bojarski J, Podolak I, Leciejewicz-Ziemecka E. Thin layer chromatography in chiral separations and analysis. 2007.
174. Agrawal YK, Bhatt HG, Raval HG, Oza PM, Gogoi PJ. Mini Rev Med Chem 2007; 7: 451–460.
175. Murakami H. Novel optical resolution technologies. 2007. p. 273.
176. Ahuja S. Chiral separations: applications and technology. An American Chemical Society Publication; 1997.
177. Caner H, Groner E, Levy L, Agranat I. Drug Discov Today 2004; 9: 105–110.
178. Welch CJ. Preparative enantioselective chromatography. 2005.
179. Maier NM, Lindner W. Chirality in drug research. 2006.
180. Patil PA, Kothekar MA, Statistics A, Comments R, Downloaded PDF. Development 2006; 60: 427–437.
181. Beesley TE, Lee JT. LC GC North Am 2006; 24: 28.
182. Wang X. Am Pharm Rev 2007; 10: 60.
183. Subramanian G. Chiral separation techniques: a practical approach. Wiley-VCH; 2007.
184. Shimazawa R, Nagai N, Toyoshima S, Okuda H. J Health Sci 2008; 54: 23–29.
185. Domenici E, Bertucci C, Salvadori P, F'lix G, Cahagne I, Motellier S, Wainer IW. Chromatographia 1990; 29: 170–176.
186. Tang Y. Chirality 1996; 8: 136–142.
187. Miller L, Orihuela C, Fronek R, Murphy J. J Chromatogr A 1999; 865: 211–226.
188. Peter A, Torok G, Armstrong DW. J Chromatogr A 1998; 793: 283–296.

189. Chang SC, Reid GL, Chen S, Chang CD, Armstrong DW. Trends Analyt chem (Regular ed.) 1993; 12: 144–153.
190. Ekborg-Ott KH, Liu Y, Armstrong DW. Chirality 1998; 10: 434–483.
191. Stationary C, Soc P. Bull Korean Chem Soc 2005; 26: 1153.
192. Li X. J Liq Chromatogr Relat Technol 2007; 30: 937–964.
193. Krause K, Girod M, Chankvetadze B, Blaschke G. J Chromatogr A 1999; 837: 51–63.

# *Microdialysis*

**Jamie Carroll,[1] Bethany R. Brookshire,[2] and Tiffany A. Mathews[1]**

[1]*Department of Chemistry, Wayne State University, Detroit, Michigan*
[2]*Department of Physiology and Pharmacology, Wake Forest University School of Medicine, Winston-Salem, North Carolina*

## OVERVIEW

With over 14,000 papers currently in the literature, and more appearing daily, microdialysis is one of the most popular techniques for sampling biological molecules from living tissue. Dialysis, the process of sampling fluid in the *in vivo* environment, utilizes a semi-permeable membrane, allowing endogenous biological molecules to diffuse into a sample via a concentration gradient. The main focus of this chapter will be microdialysis in brain. *In vivo* brain microdialysis sampling has been critical in enhancing our understanding of the native biochemistry of important central nervous system (CNS) proteins. Additionally, brain microdialysis has been essential in revealing many physiological and molecular aspects of human brain disease. There are a multitude of other techniques available for the studying brain neurochemistry, such as positron and single photon emission tomography, biosensors (including carbon fiber microelectrodes), push–pull perfusion, and ultrafiltration. The advantages of microdialysis lie in its ability to simultaneously sample a variety of biologically relevant signaling molecules, measure basal or stimulated levels of endogenous neurotransmitters, and monitor the pharmacokinetics of exogenously administered compounds. *In vivo* microdialysis has historically been performed in rats, but with the advent of genetically altered mice in the mid-1990s, manufacturers have miniaturized microdialysis probes to take advantage of these important models,

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing,*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

allowing neuroscientists to study the neurochemistry of knockout, knock-in and transgenic mice. Although the majority of microdialysis research has been preformed in rodents, the technique now is being employed in primates and humans.

Currently, there are over 100 compounds that have been identified as putative neurotransmitters in the mammalian brain, including small organic molecules (amino acids and monoamines), neuropeptides, and gaseous molecules (nitric oxide). Coupling microdialysis to analytical techniques with extremely low sensitivities is critical for detecting the nanomolar to picomolar levels of neurotransmitters present in discrete brain regions. Typically, microdialysis is coupled to separation techniques such as high performance liquid chromatography (HPLC) or capillary electrophoresis (CE) and electrochemical, fluorescence, or mass spectrometry to detect the neurotransmitter(s) of interest. Coupled to sensitive detection techniques, microdialysis is ideal for studying low levels of neurotransmitters as they function in the vast extracellular milieu.

As stated in the opening sentence, there are currently about 14,000 journal articles on the topic of microdialysis and an exhaustive review of the microdialysis literature is beyond the scope of this review chapter. However, we do invite the reader to use our list of references as a starting point if they are at all interested in this technique.

## 1 INTRODUCTION

### 1.1 Historical Perspective: Microdialysis

Earlier techniques before microdialysis included a "push–pull" method, first developed by Gaddum et al. in 1961. In this design, liquid enters the brain through one tube while another tube adjacent removes the fluid that entered the tissue at the same time [1,2]. The "push–pull" technique has several disadvantages; possible tissue damage from pressure due to a high flow rate, inefficient recovery of injected fluid, or the removal of tissue if the flow rates are not correctly balanced [1,2]. Bito et al. in 1966 first engineered enclosed dialysis membranes bags with a 6% dextran saline solution, which allowed analytes to be separated and collected within the tissue. The sac was placed in various tissues and allowed to equilibrate within the surrounding area [2,3]. After several days, the sac was then removed and the analytes of interest, mainly amino acids from the neck and brain tissue, were measured [2,3].

This first attempt using a dialysis membrane overcame some of the problems associated with the "push–pull" method by using the membrane bags, but a second surgery was required to remove the dialysis membrane to collect the sac containing the analytes [2,3]. In 1972, Delgado et al. developed a modification of the push–pull cannula technique termed a "dialytrode" which was able to provide continuous monitoring of analytes. The general construction of the dialytrode consists of two tubes that are connected to a dialysis membrane bag composed of polysulfone [1].

Ungerstedt and Pycock [4] later refined the dialytrode technique using several different designs. One design includes a dialysis membrane made of cellulosic tubing

with an outer diameter of approximately 250 μm and a molecular cutoff weight of less than 5 kDa [4]. The dialysis tubing is attached to steel tubes and bent into a loop to allow perfusion [4]. Using this innovative technique, Ungerstedt and Pycock examined the neurotransmitter dopamine (DA) following an injection of amphetamine in the brains of awake and freely moving rats [4]. The development of this novel microdialysis probe has provided great insight into the *in vivo* functions of a wide variety of neurotransmitters, and is still the technique of choice for understanding neurochemical transmission *in vivo* [4–9].

## 1.2 Basics Principles of Microdialysis

The semi-permeable membrane of the microdialysis probe has a typical molecular weight cut off between 5 and 30 kDa, which allows the probe to exclude macromolecules such as proteins and enzymes [2]. As shown in Fig. 1, the semi-permeable membrane permits the free diffusion of hydrophilic, low molecular weight

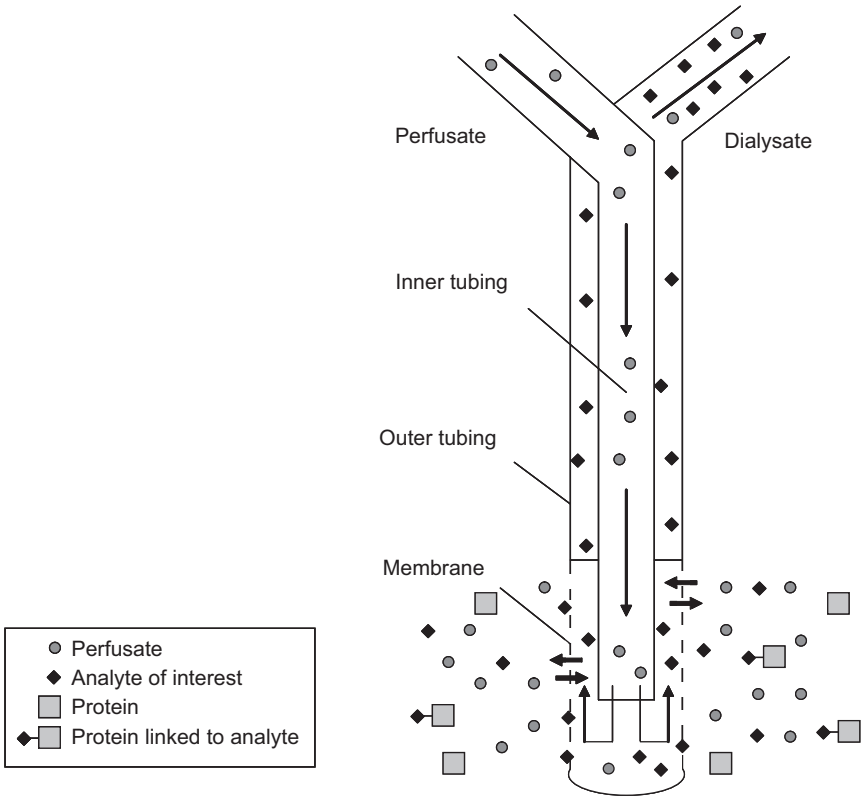


***FIGURE 1*** *Concentric microdialysis probe used for in vivo analysis of neurotransmitters. Small hydrophilic molecules are able to pass through the semipermeable membrane, while larger molecules such as proteins including enzymes are excluded.*

compounds. The probe can be implanted in almost any kind of tissue or organ including the brain, blood vessels, eyes, tumors, or the liver [10]. Syringe pumps are used to push a perfusate, usually a fluid mimicking the extracellular fluid of the tissue, into the inlet tubing of the probe at low flow rates; approximately 0.5–5.0 μL/min [10]. Once the perfusate reaches the semi-permeable membrane, dialysis sampling occurs based on the principle of diffusion. A concentration gradient exists between the perfusate, which typically is missing the analyte of interest (neurotransmitter) allowing for diffusion of the neurotransmitter from the sample matrix to the microdialysis probe. Mass transport of the analytes of interest flows down their concentration gradient, which in most cases is toward the probe. The analytes then diffuse through the semi-permeable membrane, to be carried away by the perfusate and into the dialysate sample [2,11,12]. Dialysis is defined as the exchange of analytes occurring at the membrane interface between the lumen and the tissue [12–14]. Analytes in the perfusate may also diffuse from the lumen of the probe into the sample matrix, a reverse process known as retrodialysis [13]. Typical microdialysis analytes include neurotransmitters such as biogenic monoamines, amino acids, and neuropeptides, metabolites such as glucose and hormones, and pharmaceutical or illicit drugs such as fluoxetine, methylphenidate, cocaine, or amphetamine [15]. After a fraction is collected, it is separated and coupled to the most appropriate analytical technique to determine and measure its components [16]. As shown in Fig. 1, microdialysis can be viewed as a crude separation technique that allows only hydrophilic, low molecular weight compounds to diffuse through the membrane. The power of microdialysis is that this technique can be coupled to the most appropriate analytical technique for the analytes of interest to be detected and measured [16].

## 1.3 Common Probe Characteristics and Basic Setup

Several kinds of probe designs exist, though all must have common features necessary to ensure successful dialysis. These features include a dialysis membrane that is hydrophilic and can consist of compounds such as polycarbonate, regenerated cellulose, or polyacrylonitrile (see Fig. 2) [10,14]. The dialysis membrane usually has an outer diameter of approximately 250–350 μm and a length between 1 and 10 mm [10,13]. The basic microdialysis setup is shown in Fig. 3. The animal is connected to a swivel by a collar that allows free movement within the cage. The swivel is attached to a balanced arm to prevent tubing from tangling or twisting. The syringe pump sits on top of the table and pumps the perfusate into the implanted microdialysis probe, which is situated inside a guide cannula which has been surgically implanted into the brain. The dialysate is pumped slowly through the membrane and out the other side, allowing time for analytes to diffuse into the fluid. The sample is then collected at regular intervals and analyzed for compounds of interest [1] [17].

[1] If the reader is interested a detail description for setting up the monoamine microdialysis measurements they are referred to the Acworth reference.

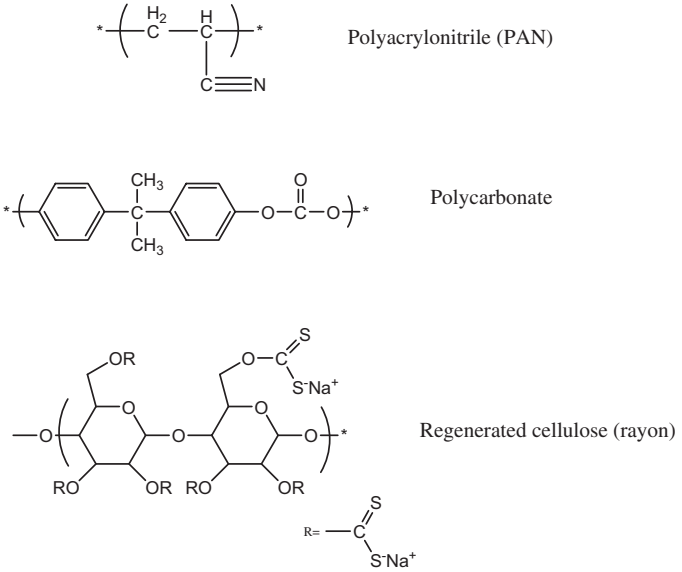


***FIGURE 2*** *Chemical structures of the different types of hydrophillic membranes used.*

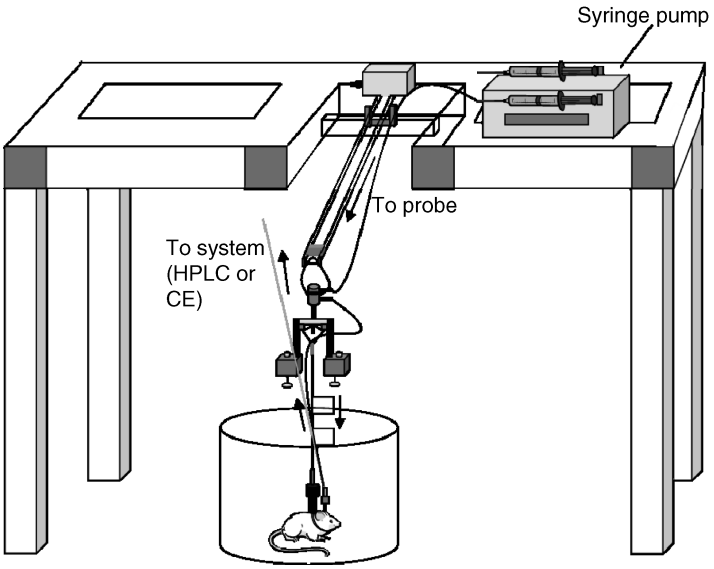


***FIGURE 3*** *In vivo microdialysis setup. If setup were for online analysis the microdialysate sample would be directly connected to a HPLC, CE or microfluidic system. If off-line analysis was being used the outlet tubing would be inserted into a microcentrifuged tube.*

## 1.4 Advantages and Potential Pitfalls

Microdialysis as a sampling technique has four important advantages: (1) the construction of the probe allows for discrimination of molecules based on molecular weight, (2) the ability to continuously monitor analytes *in vivo* over a long period of time, (3) different probe designs allowing sampling flexibility for numerous types of tissue or organs, and (4) the ability to couple microdialysis to sensitive analytical techniques allows the detection and measurement of low nanomolar to picomolar concentrations of analytes (see Section 2) [2,12]. Since the microdialysis probe is small (diameter ~200 μm, length 1–4 mm), it does not severely damage the tissue (but see Section 1.7 for a complete discussion of tissue damage related to the microdialysis probe), and is capable of separating out discrete brain regions such as the caudate–putamen, nucleus accumbens, ventral tegmental area, hippocampus, and frontal cortex. In a few studies, multiple microdialysis probes have been implanted into different brain areas such as the caudate–putamen and ventral tegmental area of the same animal to examine different release profiles of neurotransmitters to the same exogenous drug [18,19]. Microdialysis as a sampling technique is both sensitive and discriminating. Though small analytes are allowed to diffuse into the probe, proteins and enzymes are excluded, which allows the collection of analytes without fear of degradation [2,12] The technique is used in pharmacokinetic studies to examine absorption of the drug of interest, effects of the drug on the system, metabolism, and elimination, and can be used for studies of novel therapeutics, hormones, steroids, or other compounds [12,20,21]. Samples collected can be analyzed immediately without further purification by several analytical techniques (see Section 2). Finally, because microdialysis employs continuous sampling, multiple measurements containing many analytes can be taken over an extended period of time, reducing the number of animals used [2].

Microdialysis is a very common, sensitive, and useful technique for *in vivo* sampling. However, as with any technique, there are pitfalls associated with its use. Although the dialysis membrane can sample low molecular weight compounds (~<7000 Da) and provides a fairly clean sample, but this semi-permeable membrane excludes hydrophobic molecules, analytes bound to proteins, or high molecular weight compounds [2,12]. Due to the limitations of the dialysis membrane, some classes of molecules have been overlooked, such as proteins and small peptides, which, though they are not "fast" neurotransmitters, play a critical role in neuromodulation [16,22–25]. Further, microdialysis sampling does not take place at equilibrium, with only a limited time for diffusion, and thus only a percentage of the analyte is actually recovered (see Section 1.6.1). Additionally, analyte recovery is dependent upon the path of diffusion through the extracellular fluid. Molecules must first come into the vicinity of the probe, and then overcome tortuosity, in which molecules do not take direct paths to the probe membrane, and must overcome barriers to diffusion [5,15]. The path to the microdialysis membrane is complex, involving potential interactions between receptors, transporters, and degrading enzymes of the analyte of interest [5,15]. As shown in Fig. 4, it is easier for the analyte to diffuse down its concentration gradient toward the membrane in *in vitro*

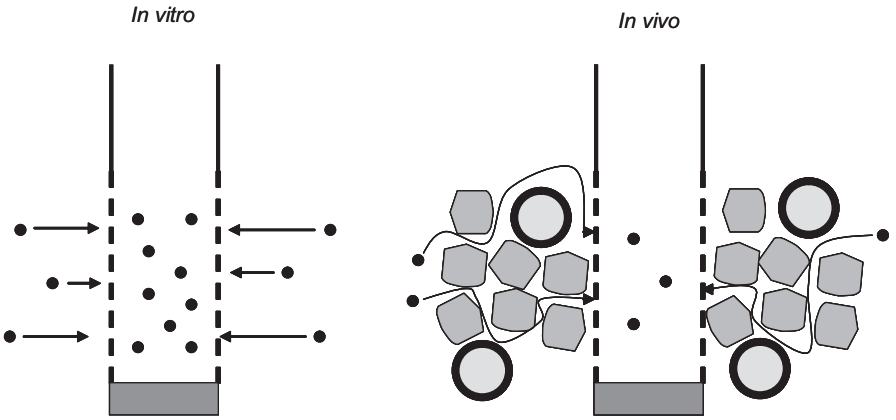


***FIGURE 4*** *The diffusion paths of analytes in an in vitro compared to an in vivo system. In in vitro studies, analytes have a direct path to microdialysis membrane. In in vivo, analytes do not have direct path but are forced to travel around cells, but they also may be taken up by cells or enzymatically degraded in the extracellular fluid.*

studies compared to *in vivo,* where the numbers of complex biological interactions is dramatically increased.

Another potential pitfall is the flow rate used. The perfusion flow rate can determine the amount of analyte recovered from the sample matrix. The slower the flow rate, the more time is allowed to approach equilibrium, and the higher the percent recovery of the analyte [12]. However, lowering the flow rate to sub-1 μL/min results in longer collection times, smaller sample volumes recovered, and poor temporal resolution [12].

Another point that has drawn considerable criticisms is the tissue damage associated with microdialysis after insertion of the probe [26–30]. Microdialysis was originally thought as a 'minimally invasive' technique, but recent studies have shown that the area around the probe and up to 1.5 μm away might be compromised; leading to decreased accuracy of analyte measurements (see Section 1.7) [26,28].

## 1.5 Probe Designs

There are several different probe designs available, though all contain dialysis membranes that are hydrophilic, and probes between 1 and 10 mm long with an outer diameter of approximately 250–350 μm [10,13]. Different probes have been developed over a number of years that are specific for various types of target tissue, a development that has increased the versatility of the microdialysis technique. The next few sections describe these different designs as well as their applicability.

***1.5.1 Side-By-Side and Loop Probe*** First introduced by Sandberg et al. in 1986, the side-by-side probe consists of two flexible tubes next to each other, with one slightly longer than the other (see Fig. 5). The longer tube is inserted into the dialysis membrane, while the shorter tube is positioned parallel to the long tube outside the

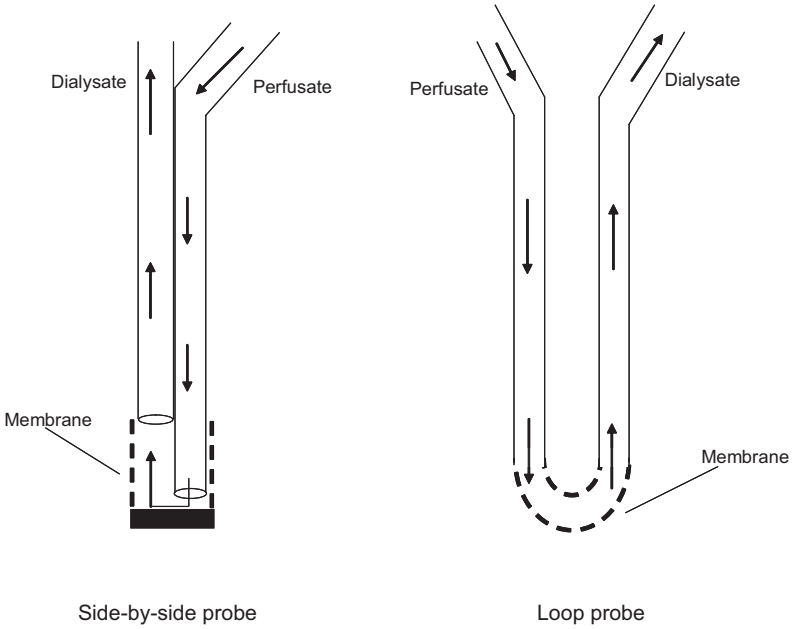


***FIGURE 5*** *Schematic of side by side and loop probe designs.*

membrane [5]. As the perfusate enters the longest tubing, it flows to the bottom of a round dialysis membrane and flows up through the shorter tube. The molecular cut-off weight of the dialysis membrane is approximately 3 kDa and the diameter of the probe is approximately 300 μm [5].

Loop probes can be made using a short piece of dialysis membrane tubing attached to two steel cannulae. There is approximately 12 mm of dialysis membrane exposed between the two cannulae, but the exposed membrane can be as large as 2–4 cm [11] The dialysis membrane is then folded into a tight loop to create the loop design. One end of the steel cannulae is connected to a pump to push the perfusate through the probe while the other end is the outlet. These membranes have an approximately cut-off weight of 5 kDa and are suitable for subcutaneous tissues or the peritoneal cavity (shown in Fig. 5) [7,11].

***1.5.2 Concentric Cannula Probe*** The concentric cannula probe was first developed by Zetterstrom et al. and later refined by Hernandez et al. and Ungerstedt [9,11]. The concentric cannula probe consists of a round dialysis membrane attached to a metal or non-metal cannula with two pieces of tubing. As shown in Fig. 1, one tubing (inner) is placed inside the other tubing (outer) to create a concentric design. [5,12,13]. The inner tube extends past the outer tube down to the tip of the dialysis membrane. As the perfusate solution travels down the inner tube, it flows to the end of the dialysis membrane and diffuses from the inner tube to the membrane. Concentric probes are approximately 250–350 μm in diameter and

varying length ranging anywhere from 1 to 4 mm in rodents to 6–10 mm in primates [13]. The major function of this type of probe is to measure neurotransmitters from the brains of either mice, rats, monkeys, or humans [11].

The concentric probe has an inflexible structure that makes it stable and sturdy enough to be implanted into the brain. First, a guide cannula containing a "dummy" filler is inserted into the brain, directly above the region of interest, and is fixed into place in the skull using glue or dental cement with bone screws [11]. After the guide cannula surgery, the animals are given a post-operational recovery varying from a few hours to a few days. After recovery, the "dummy" is removed and the microdialysis probe is inserted in its place [11,12]. An advantage of the concentric microdialysis probe is that analytes can be measured *in vivo* in freely moving awake animals without major damage to brain tissue. Though the majority of research with concentric probes has measured neurotransmitters within the brain, the concentric probe has also been used to measure drugs in the eye cavity of rabbits [13]. Although the concentric probe design is convenient for sampling in the brain, it is not intended for use in blood vessels or internal organs because of its inflexible design, and the high probability that tearing or piercing of the tissue or vessel may occur [12,13].

***1.5.3 Flexible Probe*** Developed by Wages et al. in 1986, and further refined by Telting-Diaz et al. in 1992, the flexible probe is designed to overcome the inflexible nature of the concentric probe [31,32]. The flexible probe is composed of two pieces of fused silica tubing having internal diameters of 75 μm placed in 5-mm long polyethylene tubing. For a side-by-side design, one of the tubing pieces is introduced into a polyethylene tube and extended out approximately 5 mm, while the other fused silica tubing is inserted into the polyethylene, but does not extend beyond the tubing [32]. The dialysis membrane is placed on the fused silica tubing extending beyond the polyethylene tubing and the end is sealed with epoxy (shown in Fig. 6) [32]. Because the flexible probe eliminates the structural support of the

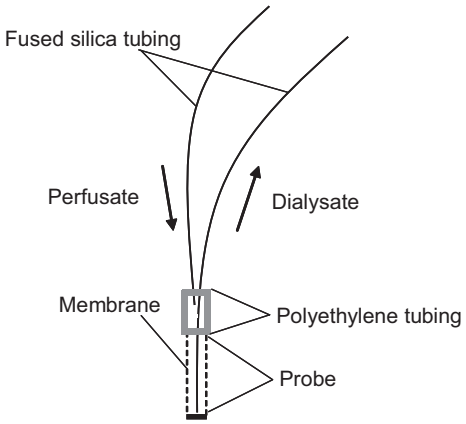


***FIGURE 6*** *Flexible probe, which has a similar design to concentric probe, but has flexible tubing for intravenous use.*

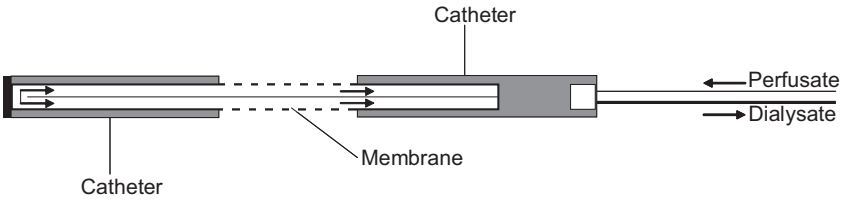


***FIGURE 7*** *Modification of a flexible probe that has catheter design.*

cannula design, damage due to tearing or piercing of tissues or vessels is reduced and allows for this style probe to be used intravenously, most commonly in the jugular vein [11–13]. Although the flexible probe overcomes the structural disadvantages of a concentric cannula probe, it is difficult to attach the probe to target organs such as the liver or muscle without disrupting the integrity of the tissue. Yang et al. reported on a flexible concentric probe using a catheter design for insertion into the inferior vena cava of rats (see Fig. 7) [11,33]. This new type of design allows for easy placement in the vessel and prevents the membrane from being easily damaged since it is placed in the middle and not at the end [11].

***1.5.4 Linear Probe*** The linear probe, designed in 1974 by Ungerstedt and Pycock, is a flexible design suitable for implantation in peripheral tissues or organs such as the liver, muscle, and skin (dermis), as well as sampling from tumors and fluid from the spine [11–13]. The probe is made by taking a hollow fiber membrane and attaching it to flexible tubing (Fig. 8). A skeleton fiber is then inserted into the flexible tubing and dialysis membrane to provide support, as well as to assist with implantation of the probe by threading it through the desired tissue [13,34]. The skeleton fiber also protects the membrane from being damaged. After the inlet and outlet ends of the flexible tubing are glued in place, the skeleton fiber can be removed if necessary. There are several advantages to using the flexible probe, including making direct contact with the sample of interest, reducing tissue damage due to probe flexibility, and increased structural stability allowing the probe to be used

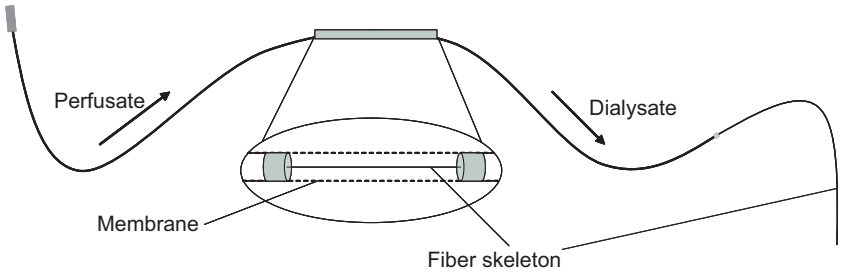


***FIGURE 8*** *Linear probe design. Completely flexible and is used in peripheral tissues such as the muscle and dermis.*

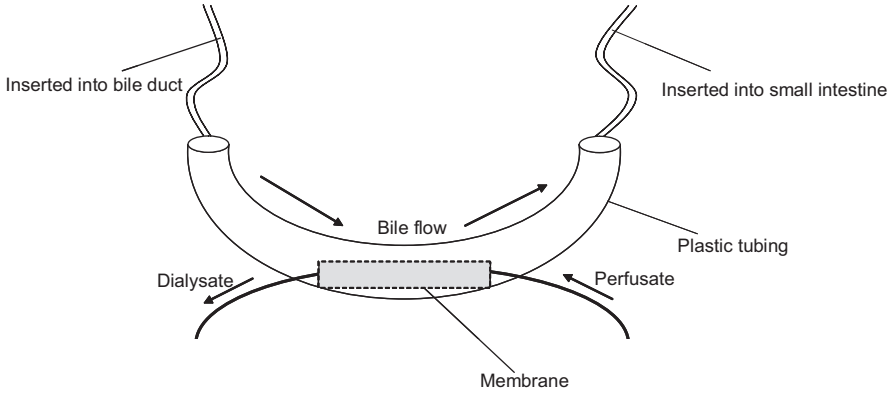


***FIGURE 9*** *Shunt probe design used in the liver to analyze bile.*

*in vivo* [13,34]. Linear probes are most often used to examine the effects of drugs in tumor tissue and in the extracellular fluid of skin tissue, studying red blood cell binding, and for kinetic studies involving enzymes *in vitro* [13]. Unlike the concentric probe, the flexible probe is unable to be used in the brain because of its large size, making it difficult to distinguish between specific brain regions [13,34].

***1.5.5 Shunt Probe*** The first shunt probe was designed in 1993 by Scott and Lunte, to examine the contents of the liver (Fig. 9) [35]. The design of the shunt probe consists of a membrane approximately 3–6 cm long (attached to flexible tubing), which is inserted into a piece of polyethylene tubing [11] The inlet of the polyethylene tubing is inserted and secured with sutures into the bile duct toward the liver and the outlet is inserted into the small intestine. The sample can then flow past the dialysis membrane and interact with the perfusate that is pumped in the opposite direction through the flexible tubing [11–13,36,37]. The advantage of the shunt probe design is that it allows for continuous sampling of the bile (generally low molecular weight compounds) from the liver [35]. Previous designs have collected bile over an extensive period of time (>48 h), which can lead to the break down of bile or alter the effects of drugs. The shunt probe offers a major advantage in that samples from bile can be tested continuously without degradation over long periods of time [35]. Therefore, several experiments testing the metabolism of bile or drug studies involving the analyte of interest can be determined in rats [35]. Additionally, the shunt probe has been used for desalinating samples for analysis in mass spectrometry (MS). However, the shunt probe must flow continuously, while the flow of bile is sometimes sporadic, due to influences of body temperature, bile concentration, and anesthesia or drugs administered into the system [37].

While there are currently several types of probes in use and more in development for sampling of tissues, it remains clear that the principles and measurements obtained

from microdialysis remain the same. We shall now shift to discussion of the technical points of microdialysis, which apply across probe types.

## 1.6 Calibration

One of the critiques of microdialysis is that concentration of neurotransmitters, when uncorrected for probe recovery, vastly underestimates actual extracellular concentrations in the brain [26,38]. The extracellular value typically reported in microdialysis experiments does not correct for *in vivo* recovery [26,38]. Several mathematical theories of mass transport in microdialysis experiments have evolved over the past 30 years which have led to methods to correct for *in vivo* recovery [5,39,40]. One issue that should be considered when determining quantitative determination of extracellular neurotransmitter levels is the recovery of the dialysis probe. Recovery of neurotransmitters from the brain by microdialysis is low (<50 %) and varies between individual probes and animals [38] However, the microdialysis probe is unable to recovery a 100% of the analyte from the brain matrix due to barriers arising from probe material, length, and area of the membrane, diffusion coefficient, flow rate, matrix tortuosity, and temperature. Firstly, we will discuss calibration of microdialysis probes *in vitro*, and then discuss three methods for estimating basal extracellular levels *in vivo* that have evolved to correct for probe recovery: zero flow, zero net flux, and retrodialysis.

The barriers to probe recover can be compensated for in a variety of ways. For example, it has been shown that if the material of the dialysis membrane is composed of polyacrylnitrile (PAN), which is a thicker (30 μm), more negatively charged membrane compared to polycarbonate or regenerated cellulose membranes (9 μm), anionic compounds will have a lower recovery with PAN membranes versus other membranes compared [11]. Increased length of the microdialysis membrane will increase percent recovery because a larger surface area is available to sample from the matrix, though the length of membrane used depends heavily on the tissues being sampled [2,10]. In brain tissues, shorter membranes (<2 mm) are typically used, while longer membranes are used in other tissues such as the muscle, allowing for increased recovery of analytes compared to brain [11]. The diffusion coefficient of the analyte is also affecting probe recovery. The diffusion coefficient is defined by the amount of analyte that travels through a concentration gradient over a certain area and time [11]. Based on the equation by the Stokes–Einstein relation, the diffusion coefficient is directly related to the viscosity, temperature, and the size of the analyte. Therefore, higher molecular weight compounds (or molecules with a larger radius) have a lower diffusion coefficient than lower molecular weight compounds and will hinder probe recovery. Finally, to optimize probe recovery, the temperature must be consistent with the sample matrix. For this reason, the perfusion liquid is generally kept at 37°C, or equalized to the temperature of the *in vitro* tissue being sampled [2].

As can be seen from above, probe recovery is dependent on a number or factors, many of which the experimenter can manipulate to optimize *in vivo* recovery. To further increase the amount of analyte recovered from the sample matrix, other factors must be taken into additional consideration. Flow rate of the perfusate can be

decreased, which increases in the amount of analyte recovered, by increasing the amount of time available for diffusion [12]. However, decreasing the flow rate also results in poor temporal resolution (how precise measurements are over a period of time) due to longer collection times and smaller sample volumes [12]. Additionally, the concentration of the analyte in the dialysate sample can be increased by adding inhibitors that prevent degradation or uptake [2]. For example, using a selective serotonin reuptake inhibitor will increase the amount of extracellular serotonin and increases the concentration in the sample matrix [41].

***1.6.1 In Vitro Microdialysis Calibration*** In the past, one of the most common methods for evaluating the recovery of the microdialysis probe was to perform *in vitro* recovery, which was used as an index of *in vivo* probe recovery [11]. *In vitro* recovery of the probe is a fairly straightforward process. The probe is placed inside a beaker (typically well stirred and at 37°C, although room temperature has been used) containing a known concentration of the analyte of interest in extracellular fluid. The inlet side of the microdialysis probe is perfused with extracellular fluid, and the sample is collected and measured by an analytical method (see Section 2). The *in vitro* relative recovery ($R$) of an analyte is defined by the following equation:

$$R = C_{\mathrm{p,f}}/C_{\mathrm{e,f}}$$

where $C_{\mathrm{p,f}}$ is the concentration of the analyte obtained from the microdialysis probe and $C_{\mathrm{e,f}}$ is the concentration of analyte in the beaker. Although *in vitro* recovery used to be a popular technique to estimate *in vivo* recovery, it is not considered as reliable as more recent *in vivo* methods. An *in vitro* solution is not a complex matrix where active biological processes are taking place, and where the path length of analytes to the microdialysis probe is tortuous, limiting the mass transport of the analyte. Instead, *in vitro* recovery allows for diffusion without all these complicating factors, which can drastically alter the recovery of a probe *in vivo*. Thus, the *in vitro* model is not capable of mimicking the *in vivo* system accurately, and in recent years has given way to the recent *in vivo* methods of zero flow, zero net flux, and retrodialysis.

***1.6.2 Flow Rate Method: Extrapolation to Zero Flow and Zero Flow Rate*** The zero flow method is based on the principle that recovery of the probe is inversely proportional to the flow rate, meaning the slower the flow rate, the greater the recovery [11]. Typical flow rates for rodent microdialysis experiments are in the order of 1–2 μL/min, allowing a collection of 10–20 μL in a 20 min sample. These flow rates permit a respectable sample size for analysis, while staying within a biologically relevant timescale. Because recovery is inversely proportional to flow rate, the flow rate and the relatively recovery can be used to extrapolate to the point of zero flow. In zero flow microdialysis, various flow rates ranging from 0.3 to 5 μL/min are chosen and the analyte's concentration is determined at each flow as a function of the flow rate. The data are fitted and one can extrapolate (estimate the value) to a steady state, defined as the point when the probe is in equilibrium with the extracellular fluid.

Justice and co-workers took this analysis one step further by deducing that, if one measured the concentration of analytes at very slow flow rates (<100 nL/min), the dialysate and extracellular space would be near equilibrium [42]. At very slow flow rates, the dialysate has the maximum time to equilibrate with the extracellular space, allowing for the greatest possible recovery. The advantage of this slow flow technique over the extrapolation method is that only one flow rate is evaluated, a less time consuming design than changing flow rates and waiting for equilibrium to be re-established for each measurement. The drawback with both methods, however, is poor temporal resolution, as flow rates less than 1 μL/min involves longer sampling times to obtain the needed quantity of sample. Additionally, the method of slow perfusion rate may be hindered with sample evaporation during long collection times as well as the stability of the slow perfusion rate [42]. This means either (a) longer collection time is necessary to obtain a sample, or (b) a microbore or CE system will be required coupled with a normal sample time and low flow rate, to analyze a small sample volume adequately.

***1.6.3 Zero Net Flux*** Zero net flux is a method that is commonly used to estimate basal neurotransmitter levels *in vivo*. In zero net flux experiments, the concentration of the analyte of interest is perfused into the brain region of interest at various concentrations ($C_{in}$; concentration of the analyte going in). Diffusion will then take place *in vivo* across the dialysis membrane. The amount of analyte exiting the microdialysis probe is defined as $C_{out}$. When $C_{out}$ is greater than the amount of analyte being perfused, there is a net gain of the analyte in the microdialysis probe as neurotransmitter diffuses into the sample (represented by the positive $y$-axis values; see Fig. 10). However, if $C_{out}$ is less than the amount of analyte being perfused, there is less extant neurotransmitter in the brain than in the perfusate. Thus, there is a net loss of analyte in the sample as it diffuses from the microdialysis probe to the brain. Determination of zero net flux is plotted as a linear regression plot, where $C_{in}$ is plotted on the $x$-axis and the difference between $C_{in}$ and $C_{out}$ appears on the $y$-axis. The point where the line crosses the $x$-axis is referred to as $C_{ext}$, where the Concentration perfused in is the same as that in brain, resulting in no diffusion across the membrane (the point of zero net flux). This point represents an estimate of the concentration of the extracellular neurotransmitter of interest that corrects for sub-maximal recovery across the dialysis membrane. Previous studies by Justice and co-workers have shown that changes in transporter function can alter microdialysis probe recovery [43–47]. Thus, it has been hypothesized that the only biologically significant process affecting probe recovery is the active uptake of neurotransmitters by their plasma membrane transporters.

***1.6.4 Retrodialysis*** An advantage of using the microdialysis technique is that it allows for the perfusion of an exogenous drug (antagonist or agonist) or analyte directly into the target tissue to measure the local response. Retrodialysis takes advantage of this ability by applying an internal standard to the perfusate. Retrodialysis is based on the principle that the relative loss (RL) of the internal

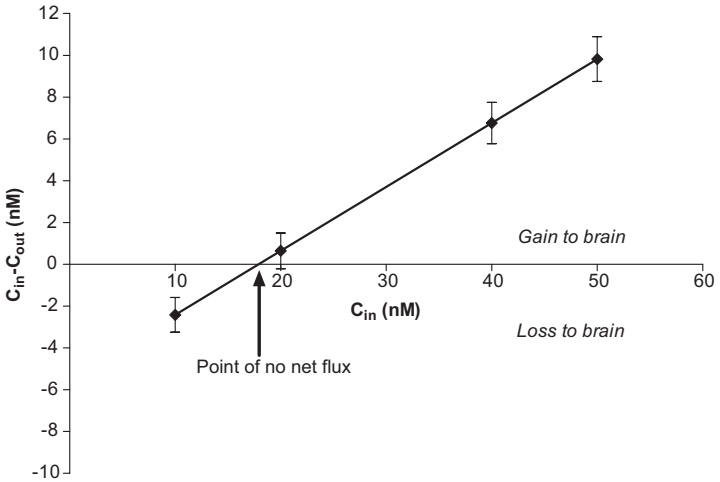


***FIGURE 10*** *A hypothetical graph of no net flux. The x-axis represents the amount of neurotransmitter being perfused into the brain ($C_{in}$), while the y-axis represents the difference between the amount of neurotransmitter being perfused in and the amount that is measured ($C_{out}$). Where the line crosses the x-axis ($y = 0$) is the point of no net flux and represents the corrected neurotransmitter level.*

standard is related to the relative recovery (RR) of the analyte of interest. Relative recovery as defined in *in vitro* experiments is the concentration of the dialysate ($C_d$) divided by the concentration of the extracellular levels ($C_{ext}$), while relative recovery is defined as the loss of the internal standard to the tissue ($C_{perfusate} - C_{dialysate}$) (divided by the concentration in the perfusate. Using retrodialysis as a calibration model entails perfusion of a non-native neurotransmitter with a similar chemical structure, for example *N*-methyl-serotonin for the neurotransmitter serotonin [48]. However, the relative recovery being measured is not from the analyte of interest, but rather the recovery of the non-native neurotransmitter. The recovery of the non-native neurotransmitter is used to make assumptions about the recovery and diffusion of the native neurotransmitter, assuming the non-native neurotransmitter has values equivalent to the analyte of interest [49]. Retrodialysis overcomes some of the disadvantages associated with *in vitro* probe recovery by placing the probe in the complex biological matrix. However, the non-native neurotransmitter used may display different properties compared to a native neurotransmitter with respect to metabolism and uptake *in vivo* and these issues are often overlooked. Although this method appears to have more advantages compared to *in vitro* microdialysis, it is still not a widely used method for calculating *in vivo* recovery.

## 1.7 Tissue Damage

One of the major hurdles associated with *in vivo* monitoring is the tissue damage which occurs when inserting a probe into the brain [50]. Although, relatively speaking, the microdialysis rodent probe is small ($\sim$200 μm in diameter, length of

1–4 mm), and the damage is often described as minimally invasive [16,50], local tissue responses must be taken into consideration. Following surgical implantation of the probe, a series of responses develop near the implantation site. Firstly, within 2 days of implantation, edema and leukocytes surround the probe. There may also be hemorrhaging and accumulation of polymorphonuclear leukocytes (a type of white blood cell signifying an immune response). Macrophages (another characteristic response to foreign invasion) are also observed in the implantation area between 2 and 3 days following surgery [13]. Ungerstedt et al. has observed an increase in g (28) lucose metabolism, decrease in blood flow, and a decrease in transmitter release taking place after probe insertion, which could last from minutes to hours depending on the type of tissue being analyzed [2]. Beneviste demonstrated similar effects, showing an increase in glucose metabolism with decreased blood flow after approximately 3 h, but recovering after 24 h [5]. Beneviste also observed astrocyte growth 3 days after implantation, and after 14 days, reticulin-positive fibers were present and had encapsulated the probe [5]. Previous studies have also shown edema, hemorrhages, monocytes, and an increase in the number of vacuoles (vacuolization) present in the probe implantation area [5].

Clapp-Lilly et al. have also studied the effects of dialysis probe tissue damage, measuring intracellular disruption and damage in rat striatal tissue. Following implantation of a microdialysis probe, there was evidence of decreased neuronal density in the striatum ~400 um away from the probe tract, and intracellular disruption as far as 1.4 mm away from the probe [28]. The group then performed a more extensive study of tissue damage, finding evidence of neuronal damage, compromised synapses, and disruptions in neurotransmitter release up to 1.4 mm away from the probe [28]. Using light microscopy, they observed neuronal damage, including myelinated axon damage, at 0–100 um distance from the probe site. At 350 μm, axons containing fiber bundles were swollen, and at 500 μm there was general visible neuronal damage [28]. Between 500 and 1000 μm from the probe site, some neurons appeared normal, while others showed evidence of degeneration with the presence of vacuoles. Astrocytes present within this range also showed signs of swelling [28]. With the use of electron microscopy, Clapp-Lilly et al. further observed the effects of probe implantation on gliosis (the accumulation of large numbers of astrocytes, occurring at sites of tissue damage or injury) at distances of 0–300 μm [28]. At 300–400 μm, neuron damage was less severe, though there remained evidence of gliosis and darken neurons. Even at the distance of 1.2–1.4 mm, there was still observation of deteriorating mitochondria within neurons, synapse damage, and damage to myelinated axons [28]. Given these observations, it is clear that substantial tissue damage takes place following microdialysis probe implantation in to the brain tissue.

Given the evidence of tissue damage at some distance from the site of probe implantation, there has been some debate over the effectiveness of microdialysis measures. If synapses and neurons are severely damages, how can microdialysis correctly estimate the concentration of biological molecules in the area of interest [27,30]? In a study to assay the effects of microdialysis probe implantation, Bungay et al. examined the function of the neurotransmitter dopamine. Two common

techniques, microdialysis and voltammetry, were observed for the effects of tissue damage [27]. After observing the clearance rates of dopamine from the extracellular fluid of brain tissue, Bungay et al. found that uptake rates for dopamine were higher when measured via voltammetry. This suggests that tissue damage may be less using voltammetry techniques, possibly due to the smaller size of the voltammetry electrode [27]. The authors concluded that microdialysis was not an accurate depiction of extracellular dopamine or dopamine clearance rates, presumably due to the extensive amount of tissue damage following probe implantation [27]. In an unrelated study, Grabb et al. examined the effects of acute (3 h) versus chronic (24 h) implantation of probes in the brain [51]. Both acutely and chronically implanted probes showed no difference in the amount of microvascular hemorrhaging [51]. Though at the 24 h time point there were significant increases in inflammation, fibrin (a blood clotting factor), edema, and the presence of microglia. Probes implanted for only 3 h showed no increase in the number of microglia present compared to controls, and though macrophages where present, there were low levels compared to the 24 h times point [51]. Finally, chronic probe implantation showed no difference in the amount of astrocytes present, while increased astrocytes were seen in the acute probe implantation. Results by Grabb et al. thus suggest that chronically implanted probes may be less accurate in ascertaining extracellular analytes levels, due to increases in inflammation, edema, and fibrin, and that new probe designs or implantation methods should be developed to overcome these issues [51].

## 2 ANALYTICAL METHODS

### 2.1 Introduction

Microdialysis itself is not a sensitive technique, but coupling microdialysis to a wide range of separation methods and sensitive detectors can improve its versatility and power. Although it is possible to couple microdialysis to techniques without a separation step such as immunoassays or mass spectrometric analysis, the majority of microdialysis studies employ a separation component, such as liquid chromatography or capillary electrophoresis (CE), to separate small molecular weight analytes from one another [52,53]. Once the analytes are separated, a sensitive detector is required to measure the low concentrations of endogenous compounds. A variety of analytical instruments are used to detect the analytes of interest, including ultraviolet-visible (UV–Vis) detector, electrochemical detection, fluorescence, and mass spectrometry.

### 2.2 Online vs. Off-Line Collection

After the analytes of interest in the extracellular space perfuse down their concentration gradients and into the probe, the samples can be collected in one of two ways; “online” or “off-line”. Both online and off-line systems have essentially the same setup; a syringe pump, a syringe selector if more than one perfusion liquid is required,

and a microfraction collector to collect the sample (off-line). For online collection, the microdialysis tubing feeds directly into the injection valve for automated collection, taking the place of the microfraction collector. When using the off-line collection method, samples are collected into microcentrifuge tubes and placed on ice during collection or immediately frozen after collection has finished [45,54]. The advantage of the off-line method is that many animals can be dialyzed at the same time, without being limited by the number of analytical instruments available. Samples are collected manually by changing vials or automatically using a fraction collector [2]. Fraction collectors automate off-line collection, which minimizes the handling of small sample volumes, providing an easy interface to collect from multiple (2 or 3) probes. Samples collected off-line are stored at −80°C and analyzed at a later time point [55]. It is important to ensure that the microcentrifuge tubes are immediately sealed and stored properly to prevent evaporation of the sample [2]. Some fraction collectors have a sophisticated system that keeps the microcentrifuge tubes refrigerated and sealed. Temporal resolution for off-line collection is governed by the dialysate flow rate and the amount of sample required for analysis [10]. If the analyte of interest is below the detector's limit of detection, then dialysate samples must be collected over a longer period of time, which results in poor temporal resolution [56].

The online collection method is used when the outlet tube of the probe is connected directly to a sample loop with a valve. The sample can then be directly injected into a high performance liquid chromatography (HPLC) or CE system. There are several advantages to using the online method: (1) samples can be analyzed in real time without evaporation [2,20,56], (2) automated sample collecting is easier for sub- to microliter sample size, (3) longer experiments can be performed without the presence of the experimenter (>12 h), and (4) near real-time monitoring of the resulting chromatograms can be performed. Real-time monitoring allows the experimenter to immediately identify problems with the experimental subject, microdialysis probe, flow, or co-eluting peaks, and thus to trouble-shoot them before the experiment can be compromised. When using the online collection method, both the microdialysis and the chromatogram systems have to be optimized simultaneously [2]. In online microdialysis, the temporal resolution is dependent on the run time of the chromatography (or other analysis system). For example, if it takes 30 min to resolve the analytes, then the next dialysate sample cannot be injected, separated, and analyzed until the end of the previous sample. Furthermore, the sample must be flushed properly during online collection, or there will be carry-over into the next collection. Samples can also be lost to waste if there is a time separation between injections. Despite this, online collection is useful for long, optimized experiments with small sample sizes [13].

## 2.3 Separation Techniques (Chromatography)

Microdialysis is itself a separation technique, capable of discriminating sample molecules by molecular weight, but chromatography is often utilized to further separate structurally similar molecules collected by dialysis from one another.

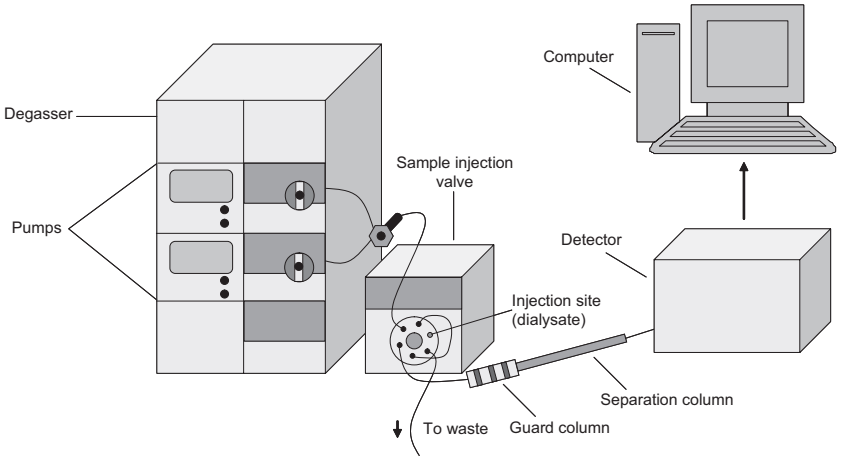


***FIGURE 11*** *Schematic of HPLC setup with major components which include the HPLC pumps (gradient example shown), degasser, injector, guard column, column, and detector.*

Although there are many types of chromatography including size-exclusion, ion exchange, and affinity chromatography, the chromatography techniques generally used for microdialysis include HPLC or CE [57]. Separation techniques like HPLC and CE are popular for several reasons: (1) ability to analyze a wide variety of biological materials including blood, tissue, urine, and cerebral spinal fluid, (2) ability to accommodate small sample volumes, and (3) reasonable temporal resolution of analytes of interest [13,57,58].

***2.3.1 High Performance Liquid Chromatography*** The objective of HPLC is to separate chemically related species present in a single mixture [57]. In HPLC, the sample is transported via a liquid mobile phase, and forced through a solid stationary phase packed into a rigid column [57]. As shown in Fig. 11, the HPLC instrumentation setup contains the following components: a degasser, HPLC pump, guard, and analytical column, and detector. The degasser is used to remove any bubbles that may occur due to oxygen and nitrogen present in the mobile phase. The bubbles are removed by purging the system with helium gas [59]. The mobile phase must run through the degasser prior to pumping through the HPLC system. If air is present within the system, these air bubbles could result in significant noise and loss of efficiency [59].

Depending on the analytes of interest, an isocratic mobile phase may be used, but in those cases where more than one mobile phase is required, gradient mobile phases are needed. Gradient mobile phases are produced by using more than one pump to mix different solvents with differing polarities, ionic strength, and concentration together [57,60]. Gradient elutions can be fashioned in a stepwise or continuous gradient, and are used to separate the sample more efficiently due to the difficulty of separating analytes with similar chemical characteristics [57,60]. The HPLC pump pushes the mobile phase and sample through the column at high pressures ($\leq$6,000 psi) [57].

Although there are several different HPLC pumps available, the most common is the reciprocating pump, which contains dual pump heads to push the solvent through. This ensures a constant flow of mobile phase through the system.

The column comprises the stationary phase of the system, over which the mobile phase containing the sample is run. The analytes can then separate based on the length of interaction with the stationary phase [60]. The time it takes for an analyte to elute from the column is known as the compound's retention time (RT). Columns used for microdialysis typically have an internal diameter of <5 mm containing packing particles between 3 and 10 μm, though particles as small as <2.2 um have recently been developed [57,60]. Since most biological molecules are polar, reverse phase liquid chromatography (RPLC) is used. A RPLC stationary phase column contains silica that has been modified with non-polar materials such as $CH_3$, $C_4H_9$, $C_{18}H_{37}$, or $C_8H_{17}$. The mobile phase for RPLC consists of polar solvents such as water with lesser amounts of acetonitrile or methanol [60]. Therefore, the more an analyte is retained within the stationary phase, the longer it takes for the analyte to elute or leave the column [60]. For example, the neurotransmitter dopamine is not only a small polar compound but also contains a non-polar aromatic ring moiety. This duality of structure allows this chemical species to interact with both the stationary phase of the column ($C_{18}$, non-polar) and the polar mobile phase. Molecules spending more time in the mobile phase versus the stationary phase have shorter retention times, while molecules interacting more closely with the stationary phase elute from the column later. The column stationary phase is normally stable, but it should be noted that extremes in pH can decrease the stability of the silica material and decrease the lifetime of the column. It is therefore recommended that the mobile phase pH is between 2 and 8.

To extend the lifetime of an HPLC and reduce maintenance time, guard columns and in-line filters can be used. Researchers use a guard column, which is placed before the analytical column to protect it from impurities and analytes that may bind irreversibly to the stationary phase. The guard column contains the same stationary phase as the column, with an increased particle size to account for the pressure difference between the two columns. Some analytical detectors are very sensitive, and the manufacturers may recommend an in-line filter to protect the detector from contaminants, not only from the mobile phase but also from the sample injected. This is especially true of microdialysis, as the dialysis membrane will allow very small peptides (<6,000 Daltons) to collect in the same. These peptides can be retained very easily within detectors or on columns, reducing column lifetime. Therefore, the use of guard columns or filters can prevent these materials from adhering to critical components of the system.

Ultimately, when analyzing low volumes and concentrations (low nanomolar) of monoamine neurotransmitters, amino acids, and other compounds, it is imperative that the HPLC system be optimized. One of the first concerns with separation techniques such as HPLC is that injection of the samples onto the column further dilutes a low volume and low concentration sample when it is combined with the mobile phase. This issue can be resolved by using a shorter column. There are several other factors that can affect the resolution, selectivity, and sensitivity of the sample

analysis [60]. First, the amount of dead volume within the HPLC system should be minimized to decrease the width of the peak (or band). The peak width is dependent on extra-column effects such as the volume required for the detector, length of tubing and connectors, or filters [58]. Column lengths and diameters should also be evaluated to determine if the correct size column is being used to elute the analytes of interest. A simple change to improve the sensitivity of the HPLC system is to decrease the column diameter (example from 4.6 to 3.0 mm). The smaller internal diameter of the column decreases the column volume dilution factor and band broadening for the analytes of interest, resulting in taller and sharper peaks. Failure to maximize the column efficiency results in band broadening and decreased resolution, which ultimately affects the quality of the sample being analyzed. Therefore, all steps to improve resolution, column efficiency, selectivity, and sensitivity should be careful considered.

***2.3.2 Microbore Liquid Chromatography (Microbore-LC)*** Types of liquid chromatography are often classified according to the column which is used. Use of columns with internal diameters less than 75 μm is referred to as nano-chromatography. An internal diameter from 25 to 100 μm is usually defined as capillary liquid chromatography (cLC), while 0.5–1 mm columns are classified as microbore. Microbore columns have been shown to have a 20-fold smaller cross-section compared to conventional columns (4.6 mm i.d.), which can lead to enhancement in sample concentration and decrease in band broadening [11,13]. Finally, narrow bore chromatography columns have internal diameters of 1–2 mm, and the conventional column width for microdialysis is around 4.6 mm. The internal particle size can also be changed to increase sensitivity. Ultra-performance liquid chromatography involves particles below 2 μm and results in detection at picomolar concentration, though retention time is sacrificed.

Traditional HPLC columns with an internal diameter of 3.0–4.6 mm are the most commonly used in *in vivo* microdialysis, though microbore columns, with their increased sensitivity to analytes, have recently enjoyed increased popularity. One report showed a 2000% increase in sensitivity using a microbore column compared to a traditional HPLC column [58]. Microbore columns also allow for the use of smaller sample sizes (1–10 μL), and can reduce the limits of detection. Finally, Justice and co-workers demonstrated that by using a microbore system, dialysate sampling time can be reduced to 1 min for the neurotransmitter dopamine, providing high temporal resolution for microdialysis [61].

Generally, conventional columns are used in microdialysis studies compared to microbore-LC or capillary-LC. Conventional columns do not generally require special instruments such as reduced flow rate pumps, low dead volume injectors, or a low volume detector, products that are often required on other systems. Also, extracolumn effects are not as much of an issue in a conventional system [58]. There have been many advances in column technology to optimize resolution, band broadening, selectivity, and sensitivity when using traditional columns. These areas are particularly important to microdialysis where small sample volumes with low concentrations of analytes are common.

***2.3.3 Capillary Liquid Chromatography (cLC)*** An emerging method to analyze microdialysis samples is cLC, which uses columns with internal diameters of 25–100 μm, compared to traditional columns with 3–5 mm internal diameters [13]. This decrease in internal diameter enhances the mass sensitivity to allow detection of attomolar analyte concentrations, a major advantage for microdialysis. In addition, small diameter columns lead to a reduction in flow rate, allowing this method of chromatography to be coupled very easily to mass spectrometer detectors [50,62]. Coupling with mass spectrometer detectors can prove useful in the measurement of compounds extant in the brain in picomolar concentrations. Recently, a number of microdialysis reports have coupled capillary liquid chromatography and mass spectrometry together for the analysis of neuropeptides [50]. cLC has several advantages over traditional microdialysis columns: (1) low band broadening (dispersion), (2) increased sensitivity to attomolar detection limits, (3) reduced mobile phase consumption, and (4) separation of up to 16 different amino acids from a sample volume of 200 nL [50]. However, the cLC method has several drawbacks. The system itself is unstable due to high sensitivity and the small size of the column diameter. The smaller internal diameter can result in frequent clogs and column exchanges. Analysis times are generally much longer than that of conventional columns. Finally, due to the sensitivity of the system, extracolumn effects are common and difficult to minimize [13,50]. Despite these drawbacks, cLC is a valuable tool for detection of rare compounds and small sample sizes.

***2.3.4 Capillary Electrophoresis*** Within the last 10 years, CE has grown in popularity in the microdialysis field, in part due to work from the Kennedy and Lunte laboratories [50]. CE can be performed with several different modifications, depending on the analytes of interest and the sensitivity required. These include capillary zone electrophoresis, capillary gel electrophoresis (gel matrix), capillary isoelectric focusing (separates amphiprotic species), capillary isotachophoresis (analytes migrate at the same velocity), and micellar electrokinetic chromatography (separate neutral or chiral species) [57,60].

CE is a separation technique wherein analytes are separated in open fused silica capillary tubing having an internal diameter of approximately 25–75 μm, as opposed to the column use required by chromatography [63]. The capillary is filled with a running buffer composed of an aqueous buffer solution [57,60]. Two electrodes, a positive and a negative (termed anode and cathode, respectively) are placed on either end of the setup (in the buffer reservoirs) and a voltage power is applied (shown in Fig. 12). An electric current is then generated which allows ions to migrate from the anode toward the cathode [57,60]. Therefore, the analytes of interest separate based on their charge (or electrophoretic mobility) and mass. The electrophoretic mobility of the analytes can be affected by the charge of the ion, viscosity or other frictional forces, the size of the ion, and its shape [60]. Highly positive ions will migrate faster toward the cathode compared to negatively charged ions, while larger ions will move more slowly due to their mass.

Samples are injected through hydrostatic forces such as gravity, pressure, or a vacuum. These methods can be achieved by raising the sample above the destination

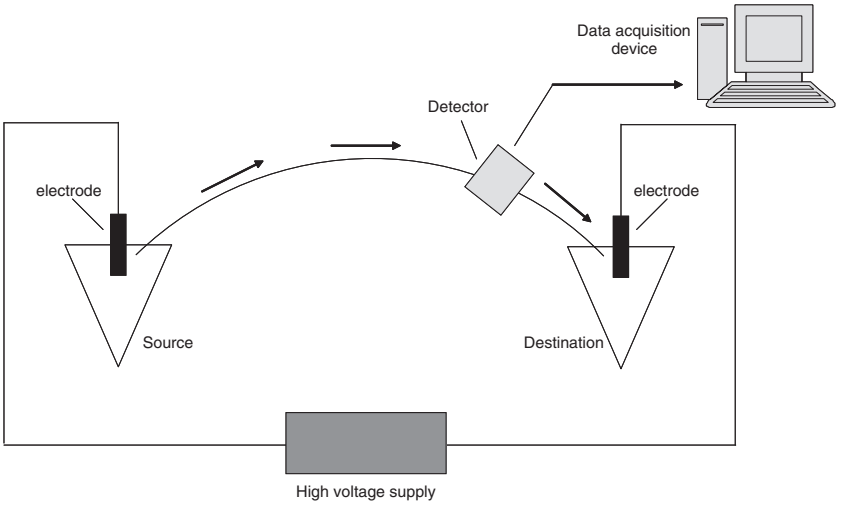


***FIGURE 12*** *Schematic of CE set up. Capillary is exposed (gray box) to provide a detection window for UV–Vis or fluorescent detection.*

vial, applying pressure, or applying a vacuum at the destination end of the setup. Samples can also be introduced by electromigration, where the sample vial replaces the buffer reservoir and a voltage is applied for a period of time [57,60]. Electromigration is the preferred injection method, as it is the most efficient for drawing the sample into the fused silica for separation.

Capillary electrophoresis is advantageous in that band broadening and interferences are minimal. Flow velocity remains constant through the capillary, regardless of the position of the analyte in the tubing. This creates a "plug-like" flow and produces narrow bands [64]. The most important features of CE include rapid separation of

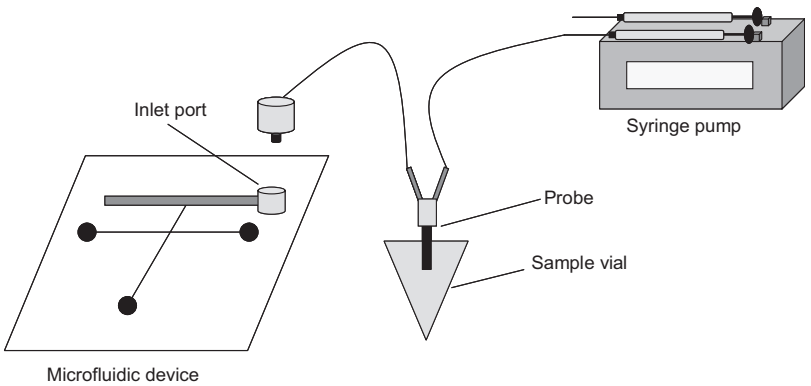


***FIGURE 13*** *The coupling of microdialysis to a microfluidic device. The probe can be placed in a vial for in vitro* or an animal for *in vivo* collection.

analytes, the option of small sample volumes (<10 nL), high mass sensitivity, high resolution, and low cost [16,63,65,66]. An additional feature of using CE is the ability to couple it to the wide variety of detectors that are commercially available as well as custom built detectors. CE can be coupled to laser-induced fluorescence, mass spectrometry, ultraviolet (UV) absorbance, and electrochemistry. The detector coupled to CE will vary depending on the analyte of interest [57]. By far the biggest advantage of coupling CE with microdialysis is the enhancement in temporal resolution. Samples can be collected and analyzed in 10-s intervals (>3000 samples for an hour of collection). For example, Qain et al. was able to determine a detection limit of 5 nM with a 1-min temporal resolution using CE coupled to electrochemistry [13]. CE also has the advantage of a low sample volume requirement (approximately 10–20 nL) Separation usually occurs after 70 s with a limit of detection of 100 nM [67]. CE has been used to identify and measure a wide variety of neurotransmitters, including amino acids, dopamine, substance P, serotonin, and noradrenaline as well as peptides such as luteinizing hormone-releasing hormone [16,63,66,68].

For off-line analysis using CE, sample volumes of approximately 1–7 μL are necessary. This avoids problems associated with the handling of sub-microliter volumes, evaporation, and loss in separation efficiency [13,50]. Initial attempts using off-line analysis involved collecting the sample in a vial, adding a derivatization agent if necessary, and injecting the sample onto the capillary. However, recent modifications of off-line collection have improved temporal resolution to 1–30 s. These modifications include mixing derivatization agents with the dialysate sample online prior to collection by using polyethylene tubing chambers, a method called "crossing" [50]. The outlet of the microdialysis probe is directly coupled to capillary tubing containing all of the necessary reagents for derivatization. After derivatization, a collection tube sends the derivatized samples to the vial. Therefore, sample collection, derivatization, and the actual separation are all online with no manual handling of samples. Lunte and co-workers was the first group to use NDA/CN to measure glutamate and aspartate online in the hippocampus of awake and freely moving animals [67]. In this CE method, two crosses were used. One cross was used to mix the derivatization agents, and the other cross to mix the derivatization agents with the dialysate sample [67]. The mixed sample is injected by a microinjection valve onto the CE [67].

Separations of 3–100 s are possible using capillary tubing of approximately 10 μm (length or diameter) [50]. However, there are several factors that can effect CE efficiency: injection length (including the volume of sample), diffusion of the analyte, electromigration dispersion (conductivity differences), joule heating, and solute interactions with the capillary wall [60].

***2.3.5 Microfluidics*** Within the last few years the area of micro-total analysis (μ-TAS) systems, or "lab-on-a-chip" devices has become a rapidly developing field, where sample introduction, separation, and detection occurs within a miniaturized or microfluidic analysis system [69,70]. These devices are engineered from various polymers or glass coated with a substrate of interest (Fig. 13). However, polymer chips

such as those made of poly(dimethylsiloxane) (PDMS) are more popular, being easy to make and less expensive compared to glass chips [69,70]. Microfluidic channels are approximately 10–100 micrometers in diameter or length in size, allowing very small sample volumes ranging from $10^{-9}$ to $10^{-18}$ [70]. Besides the small sample volumes that can be used with microfluidic devices, they are capable of faster separation time [70]. With ultra-small sample volumes being used with microfluidic devices, the challenge is that they must be coupled to very sensitive detectors. As with most traditional separation techniques, microfluidic devices have been coupled to conventional detectors such as laser-induced fluorescence, electrochemical detection, and mass spectrometry, but also non-traditional detectors including infrared, Raman, nuclear magnetic resonance (NMR), surface plasmon resonance, absorbance, and chemiluminescence detectors [70]. However, many of the detectors being used with microfluidic devices are the bulk detectors (emphasizing only the separation component is small), but recent work by Mecker and Martin integrated an electrochemical detector on a chip for dopamine analysis [71].

Microdialysis is a technique that generates samples with low volumes, and as a result there is great interest in coupling microdialysate samples to microfluidic devices. Microfluidic systems or "lab-on-a-chip" refers to the miniaturization of the analytical system for separation and measurement of analytes of interest. Although still in the beginning stages, it is anticipated that, microfluidic devices could successfully and quickly separate many analytes with low solvent requirements and minimal sample volume [69]. Current studies employing microdialysis and microfluidics monitor classical neurotransmitters such as amino acids and peptides. It is hoped that the future format for microfluidic devices will be able to analyze a wider variety of analytes or perform various assays in parallel [72–75]. Sandlin et al. of the Kennedy laboratory have already produced one of the first papers demonstrating the successful coupling of microdialysis and microfluidics [75]. Their online technique using a microfluidics electrophoresis chip with detection via laser-induced fluorescence allowed for the detection of various amino acid neurotransmitters in the rat brain [75]. Future work with microfluidics will hopefully build upon the promising potential of this form of electronic separation.

## 2.4 Detectors

Following collection of the dialysate sample, the molecules of interest are separated via HPLC or CE, and then the analytes of interest are detected and quantified. In general, detectors for microdialysis should possess the following characteristics: sensitivity (concentration of low nanomolar or less), stability, reproducibility of signal response, linear response to the analytes of interest, short response time, high reliability, and ease of use [57]. The three most commonly used detectors are electrochemical, fluorescence, and MS, detectors with low detection limits and selectivity of compounds. Interestingly, each one is used predominantly for a particular class of neurotransmitters. For example, electrochemistry is used for biogenic monoamines (including norepinephrine, dopamine, and serotonin),

fluorescence is used for the amino acids (GABA and glutamate), and mass spectrometry is used for neuropeptides.

***2.4.1 Electrochemical Detection of Analytes*** Electrochemical detection is favored for biogenic amine neurotransmitters because it can achieve exceptionally low detection limits (high picomolar) and it is selective for electroactive species. Monoamine neurotransmitters are found in very low concentrations in the brain extracellular space (typically <5 nM, not corrected for probe recovery). These species are oxidized (lose electrons) easily (at $\sim +600$ mV versus Ag/AgCl reference electrode), which makes them suitable analytes for electrochemical detection, as opposed to other neurotransmitters such as GABA and glutamate. The most widely used methods for electroactive biological compounds, like neurotransmitters, are amperometric and coulometric methods [76]. Amperometry and coulometry measure the change in the current generated during the oxidation or reduction of an analyte. In an amperometric system, the analyte of interest is allowed to flow either parallel or perpendicular to the working electrode in a thin layer configuration. The parallel cell design results in 5–10 % oxidation of the analyte to efficiencies approaching 100 % in the end-on perpendicular design, particularly when the latter is coupled to narrow-bore HPLC columns (i.d. <1 mm). Amperometric and coulometric detectors are roughly equivalent with respect to signal to noise (S/N) ratio. The coulometric detector utilizes an electrochemical cell which consists of three electrodes; a working electrode (oxidation), a counter electrode (reduction) and a reference electrode. In this design, the working and counter electrodes are composed of porous graphite which allows the mobile phase to flow through, oxidizing approximately 100% of the analyte. The potentials at each of the electrodes are held constant and as the analyte of interest is oxidized at the working electrode, a change in current is generated. The integrated current is the charge and this is proportional to the amount of analyte as explained by Faraday's law.

The applied potential for analysis of a multicomponent sample is a balance between maximizing the oxidation of the analytes of interest, while minimizing the oxidation of species that may interfere with analyte detection [57]. The identity of each neurotransmitter and metabolite is confirmed using known standard solutions to determine the retention times of each analyte. Standard curves bracketing the expected concentration for each analyte should be established before collecting dialysate samples. Calibration curves are based on peak areas of the known standard concentrations calculated by integrating peaks using current software. Sample peaks are then integrated and the peak areas are regressed to standard curves to determine their concentrations.

***2.4.2 Fluorescence*** Fluorescence is based upon a molecule's ability to absorb a photon and be promoted to a higher energy state, known as excitation. The fluorescence signal is observed when the molecule relaxes back to the ground state, releasing the photon which is then detected, a process is known as emission. Molecules that fluoresce easily are fused ring structures with a diffuse $\pi$ bonding system, a property easily applied to amino acid neurotransmitters, though the majority

of animo acids do not naturally fluoresce. The three amino acids that do have fluorescence properties are tryptophan, tyrosine, and phenylalanine, but to make the other amino acids fluoresce, a process known as derivitization is used. In derivatization, a fluorophore is covalently bound to the primary amine, allowing for the detection and quantification of amino acids by fluorescence. There are a number of pre-column derivitization compounds including 1-dimethyl aninoaphthaldenesulfonyl chloride (DANSYL), phenylisothiocynanate (PITC), 6-aminoquinolyl-*N*-hydroxysuccinimidyl carbamate (ACQ), orthophthaldialdehyde/β-mercaptoethnaol (OPA-β-MCE), and naphthalene-2,3-dicarboxaldehyde/cyanide (NDA/CN), of which the two most popular derivitization compounds for the amino acids are OPA-β-MCE and NDA/CN. This is due to their stability and their native ability not to fluoresce. NDA is the slightly more favored derivitization agent due to its ability to provide higher quantum efficiency and enhanced stability [77].

Fluorescence can be easily coupled to HPLC or a CE system to detect analytes of interest from dialysate samples. When using a HPLC system, fluorescence is measured using a commercially available fluorescence detector. However, when using CE, a laser must be used as the source for measuring fluorescence [16,25,65,66]. In both cases a light source (lamp or laser) are used to promote the analyte of interest from the ground state to the excited state. When the molecule of interest falls from the excited state to the ground state, the resulting emission is measured. Fluorescence is a popular detection scheme for neurotransmitters due to its low concentration detection limits (~100 fM), mass detection limits (~10,000 molecules), high selectivity, broad detection range, high sensitivity, and high selectivity [25]. Additionally, fluorescence is being employed as a detector in online capillaries, off-line capillaries, and microchips, which provides an additional element of versatility.

The routine separation scheme for dialysate samples in the neurosciences includes use of HPLCs. However, there are drawbacks associated with HPLC separation, including poor mass sensitivity, large sample volumes required (leading to lengthy dialysis sampling times), and ultimately poor temporal resolution. As a result of these drawbacks, researchers have been developing CE separation techniques with fluorescence to detect amino acid neurotransmitters. Examples include coupling CE-laser induce fluorescence (LIF) with microdialysis, to make measuring small dialysate fractions (nL) and detecting low concentrations feasible. The Kennedy group has demonstrated rapid sampling of amino acid neurotransmitters in less than 10 s using CE-LIF, which greatly enhances the temporal resolution of the dialysate sample [78]. If microdialysate samples are collected off-line, this would require handling of very small fractions (nanoliters), which are susceptible to evaporation. However, if these samples were to be measured online, successful derivatization and separation must occur on a compatible time scale to accomplish the high throughput required (sampling ~6 s, would generate ~600 samples) [50].

***2.4.3 Mass Spectrometry*** MS is a powerful analytical technique that is capable of providing the element composition of samples and the structure of compounds, as well as qualitative and quantitative composition [57]. Unlike

electrochemical or fluorescent detectors, MS can in principle be applied to analyze any analyte of interest. MS is capable of precisely measuring the molecular mass of a compound, and can also be used to determine the sequence of a biomolecule [50]. MS is a destructive technique which ionize components of a sample, which in turn are separated using electric and/or magnetic fields. The mass spectrum demonstrates the resulting abundance of *mass-to-charge ratio* (*m*/*z*). Although MS destroys the sample in the process, sensitivity is high and little sample volume is required [79,80]. During the 1990s MS underwent an explosion of growth, providing key advances in biological research by providing high precision of structure and molecular mass [79,80]. During this time, MS became more amendable to measuring proteins and peptides due to improved ionization methods such as electrospray ionization (ESI) and matrix-absorbed laser desorption/ionization (MALDI) [79]. Although both methods are critical for evaluating and identifying key biopolymers, the focus of this mass spectrometry section will be on ESI. This is because MALDI is used for tissue analysis of proteins and is unable to discriminate between intracellular and extracellular tissue levels, and thus is not used in *in vivo* microdialysis sampling.

ESI is an ionization method which produces multiply charged ions. These ions are generated in a strong electric field, which induces charge accumulation, producing multiply charged droplets. Typically, ESI is coupled to a separation technique such as HPLC, capillary LC, nano-LC, or CE, and the sensitivity of ESI is dependent on the separation mode being used. Separation techniques with flow rates in the microliter to nanoliter per minute range offer enhanced sensitivity by increasing the sample's concentration in the elution solvent [79]. ESI is a very powerful method which permits the molecular weight determination of biomolecules, with very low detection limits reported to be in the attomole range [25,79].

Once the analytes have been ionized, they are ready for structural analysis or selective detection of a specific class of compounds by tandem MS (or abbreviated as MS/MS or $MS^2$) [79]. For example, the first mass spectrometer is used to isolate the precursor ion, while the second mass spectrometer analyzes the products [79]. There are many advantages of using tandem mass spectrometry on complex biological samples like microdialysis. These include the speed of isolation, sequence specific detection, and quantification of unknowns. In theory, tandem MS could be used for any peptide [16]. The most common mass analyzer is the quadruple analyzer, composed of four parallel rods that are opposite frequency or charge.

Currently, over a hundred known neurotransmitters have been identified, but not all of these neurochemicals can be detected by electrochemical or fluorescent detection. Of the neurochemicals available for analysis, mass spectrometry is typically used to evaluate *in vivo* neuropeptide levels. These samples are particularly challenging; they are present in low concentrations, sample size is small, and numerous similar compounds can contaminate the sample. Only within the last 20 years has mass spectrometry been directly coupled to microdialysis to determine the extracellular concentration of a small peptides [24,25]. The first few experiments with microdialysis coupled to mass spectrometry were analytically challenging, but much progress has been made. By coupling MS and microdialysis together, this can provide

an important first step for *in vivo* monitoring of neuropeptides, an underrepresented area of research at this time.

Neuropeptides are probably the most important biological molecules in the central nervous system (CNS). However, their low levels *in vivo* and their complex chemical structure make them difficult to be detected and analyzed with electrochemical or fluorescent detection. Although MS can be used to evaluate traditional neurotransmitters such as the monoamines and amino acids, its primary role with regards to microdialysis is to measure more challenging neurotransmitters such as acetylcholine, purine and pyrimidine bases, anandamide, and prostaglandins [24]. MS has been particularly useful in identifying extracellular concentrations of endocannobinoids, substance P, and neuropeptides such as neuroensin and met-enkephalin. Although the study of these neurochemicals is a developing field, understanding neuropeptides and other non-traditional neurotransmitters may prove important for future work with both traditional and non-traditional neurotransmission.

With the increasing popularity of microbore and capillary LC in recent years, there has been an increase desire to couple these separation components to MS. Sensitivity of the analytes detected by MS is currently in the picomolar range. With increased improvement of the HPLCs, such as ultra flow and nanoflow LC, this will add another dimension and enhancement to mass spectrometry analysis, enhancing its use in the microdialysis field.

## 3 MICRODIALYSIS SAMPLING APPLICATIONS

### 3.1 Introduction

With over 14,000 publications in the area of microdialysis ranging from neurochemical alterations following drug administration to pharmacokinetics of exogenous compounds, it is clear that microdialysis measurements can be made in virtually any tissue or organ. Although a comprehensive in depth coverage of the applications of microdialysis is beyond the scope of this chapter, we do highlight a few key neurochemical applications in this section.

### 3.2 Biogenic Amines and Amino Acids

Biogenic amines such as histamine, tyramine, dopamine, epinephrine, serotonin, and tryptamine are a class of compounds which contain an amine group. Dopamine is one of the most well studied neurotransmitter systems because of its involvement in the brain reward pathway (drug abuse/addiction), neuropathological conditions such as Parkinson's Disease, schizophrenia, attention-deficit hyperactivity disorder (ADHD), and Tourette's syndrome [81,82]. Dopamine is an ideal neurotransmitter for analysis as it is easily detectable in many brain areas and oxidizes easily. Under normal physiological conditions, DA is known to regulate locomotor activity, cognition, affect, and neuroendocrine hormone secretion [83]. Extracellular dopamine levels, which are easily measured with microdialysis, are a balance of release, uptake, and

metabolism. Release is controlled by many other neurotransmitter systems which terminate on to DA neurons, as well as feedback from the DA system itself. Termination of the DA signal is accomplished by reuptake by a 12 transmembrane protein known as the dopamine transporter (DAT) [83]. The primary function of the DAT is to inactivate extracellular dopamine by transporting it back into the cytosol. DA that remains in the synapse is broken down quickly by monoamine oxidase catechol-o-methyltransferase. All of these components of the DA system are the targets of diverse pharmacotherapies and areas of intense study.

Within the neuroscience community one of the most studied areas is the effect of illicit drugs on neurotransmitter systems. These include psychostimulants such as cocaine and amphetamines, CNS depressants such as opioids and alcohol, cholinergics such as nicotine, and inhalants such as toluene [84–87]. Some of the first microdialysis studies by Ungerstedt and colleagues examined extracellular dopamine levels in the striatum (caudate–putamen; CPu) in response to drugs of abuse. The authors were able to monitor behavioral responses to drug as well as monitoring neurotransmitter responses in real time in freely moving animals. For example, one study examined the acute response to amphetamine, and the subsequent microdialysis demonstrated that this effect was due to a potentiation of dopamine in the striatum [4,5]. Microdialysis experiments have also demonstrated that an intraperitoneal (i.p.) injection of cocaine elevates extracellular DA levels by $\sim$ 200% in the caudate–putamen and nucleus accumbens. Microdialysis continues to provide valuable information on the mechanism of action of drugs of abuse such as cocaine, ethanol, and amphetamines, and will continue to be important in understanding how the brain reacts to acute and chronic drug administration.

One of the most well characterized genetically altered mouse models is the dopamine transporter knockout mouse (DAT-KO) developed in the Caron laboratory in the mid-1990s [88]. Without the DAT, DA cannot be recycled back into the intracellular storage pool, leading to a deficiency of intracellular DA. An excess of extracellular DA leads to a hyperactive phenotype. The DAT-KO mouse, coupled to *in vivo* microdialysis, has been particularly useful in evaluating the DA system with respect to the molecular mechanisms of drugs of abuse (cocaine, amphetamine, and ethanol), dopaminergic neurotoxins such as 6-hydroxydopamine (6-OHDA), and 1,2,3,6-tetrahydro-1methyl-4-phenylpyridine (MPTP), and neurological disorders such as ADHD and Parkinson's disease.

DAT-KO mice treated with $\alpha$-methyl-*p*-tyrosine, a potent irreversible inhibitor of tyrosine hydroxylase, acquire Parkinson's disease like symptoms including tremors and rigidity [89]. Parkinson's disease is a neurological disorder that is characterized by a severe deficit in striatal DA [90]. Extensive degeneration of DA neurons and changes in other neuronal subpopulations leads to bradykinesia, rigidity, resting tremor, and postural instability. When DAT-KO mice were administered $\alpha$-methyl-*p*-tyrosine, their striatal DA levels showed significant decreases 20 minutes after treatment and a behavioral phenotype of Parkinson's followed [89]. These Parkinson-like symptoms were reversed by administering amphetamine derivatives, specifically 3,4-methylenedidoxymethampphetamine (MDMA) [89], which releases DA into the extracellular matrix. However, when DA dialysate levels were evaluated, there was no

recovery of extracellular DA levels following the administration of MDMA. The MDMA microdialysis data indicate that pharmacological activation of non-DA neuronal pathways may offer another option in restoring movement in Parkinson's patients [89]. This is only one example of the many microdialysis studies on DA systems, which have illuminated mechanisms of drug abuse, as well as mechanisms of movement disorders such as Parkinson's.

## 3.3 Neuropeptides

Small peptides are an important sub-group of neurotransmitters (neuroactive peptides). These are distinguishable from animo acid and monoamine transmitters, and tend to co-localize with these more prevalent neurotransmitters [24]. Currently, there are over 100 known neuropeptides such as endorphins, enkephalin, growth-hormone, releasing hormone, neurotensin, and somatostatin. Typically, neuropeptides are as fast acting as classical neurotransmitters and play a more modulatory role in the central nervous system [24]. Neuropeptides are also larger in size and present in lower levels than classical neurotransmitters, making electrochemical detection a challenge.

The number of microdialysis studies evaluating neuropeptides levels or identifying novel neuropeptides has increased over the last 5 years. For example, Haskins et al. successfully measured non-opioid peptides derived from proenkephalin A [23]. Microdialysis samples are taken directly from the extracellular space and discriminate based on molecular weight cutoff. Thus, the dialysate sample may have a complex array of small neuropeptides, and this can be particularly challenging when concentrations of neuropeptides are low and a finite amount of sample is available. To overcome these challenges, various database programs such as SEQUEST, Mascot, and Lutefisk are used to analyze the samples and provide a true representation of the analytes in the sample [23]. All together, Haskins et al. found 29 peptides from six different proteins including fragments from the brain acidic membrane protein NAP-22, neurogranin, fibrinogen, proenkephalin A, and peptide I 1-12 [23]. Much more work remains in the area of neuropeptides, and improvements in microdialysis technique will aid in discovery and analysis of these modulatory neurotransmitters.

## 3.4 Pharmacokinetics

As discussed so far in this chapter, microdialysis is a technique which can allow neuroscientists to monitor neurotransmitter release, uptake, and metabolism *in vivo*. An equally important area for microdialysis is monitoring the pharmacokinetics of compounds *in vivo*. Pharmacokinetics is the study of absorption, distribution, metabolism, and elimination of biologically active compounds in an organism. Typically, pharmacokinetic data are obtained by administering exogenous compounds and evaluating absorption, distribution, and metabolism in blood and/or tissue at specific time points. However, these experiments involve large numbers of animals, often have poor temporal resolution relative to the physiological or pharmacological events, and involve significant sample preparation time [10]. Finally, the

results obtained from tissue content pharmacokinetic studies do not discriminate between the intra- and extracellular concentrations of exogenous analyte, and thus cannot differentiate the amount of drug taken up from cells versus the amount in the extracellular space.

Using *in vivo* microdialysis to evaluate pharmacokinetics of biologically active compounds has many advantages over typical blood and tissue analyses. As previously discussed in this chapter, microdialysis discriminates molecules by molecular weight, and by this virtue provides fairly clean samples with no large macromolecular contaminants such as proteins or enzymes [10]. Once the dialysate is collected there is no further breakdown of analytes, since enzymes are excluded by molecular weight [10]. Microdialysis involves the constant perfusion at the region of interest, where there is no net loss of fluid from the organism, and analytes are monitored continuously at the site of interest [49,91]. Microdialysate samples contain only the free fraction of the biological compound of interest; hence microdialysis monitors only the extracellular levels of the compound of interest [49]. Although the topic of this chapter has focused mainly on the brain, virtually any tissue or organ can be monitored with the correct microdialysis probe [49]. Thus, one can imagine simultaneously sampling from a multitude of sites within the body to monitor drug pharmacokinetics. Another unique advantage of microdialysis is different parts of the same tissue can be monitored simultaneously. For example, many studies have focused on evaluating both the blood and the brain to examine transport kinetics of a biologically active compound across the blood brain barrier [92]. Using microdialysis to monitor pharmacokinetics allows a variety of experimental designs to be employed. For example, a within subject design can be used, where baseline is monitored prior to drug exposure. This allows each animal to serve as its own control, decreasing the number of animals required and minimizing the variation due to individual differences [20,21,49]. Pharmacokinetic microdialysis experiments using a within subject design include (1) acute or chronic drug administration, (2) dose-ranging studies, and (3) simultaneous measurements from the same tissue or organ [20]. With continuous monitoring, multiple sites (tissue/organs), or multiple analytes (endogenous or exogenous) generates a large amount of data. This large amount of data provides a more complete picture of absorption, distribution, metabolism, and excretion compared to single time point studies. Thus, microdialysis provides one of the best methods for studying pharmacokinetics in the living organism.

## 3.5 Blood and Tissue

Microdialysis was originally developed to analyze biological endogenous and exogenous compounds in the brain, and so many advances in the technique have been developed for brain-specific studies. But the brain is not the only tissue that can be sampled using microdialysis, and with the development of non-concentric probes many other organs and tissues have been evaluated using *in vivo* microdialysis, including blood, heart, liver, bile, kidney, skin, ocular, adipose, tumor, and muscle [49,93]. A comprehensive review of all the available microdialysis methods in organs and tissue is beyond the scope of this chapter. However, we have highlighted

some of the most recent studies of microdialysis in blood, liver, myocardium, bile, and stomach, to provide examples of recent work which highlights the multiple uses of microdialysis.

When microdialysis is used as the sampling technique to monitor analytes in the blood, the microdialysis probe is commonly implanted in the right atrial junction, providing a larger blood pool around the probe [94,95]. Microdialysis sampling of blood has the advantages of providing continuous sampling over time, as opposed to blood draws, which are time point specific. To better understand the pharmacokinetics of the antibiotic cefsulodin, Tsai et al. monitored the blood and brain levels of cefsulodin [96]. A recent study by Zolkowska et al. used *in vivo* microdialysis to examine serotonin (5-HT) plasma levels in non-hypoxic rats, following chronic treatment with the anorectic drug fenfluramine and the anti-depressant fluoxetine [97]. The goal was to determine if chronic administration of these drugs elevates plasma 5-HT levels. Zolkowska et al. demonstrated elevated plasma 5-HT levels in Sprague-Dawley rats following chronic administration of fenfluramine, although the levels were much below the micromolar levels that produce adverse cardiovascular functions [97].

The liver is one of the most commonly transplanted tissues. This is partly due to the fact that part of the liver can be transplanted to restore liver function, allowing for live donation. However, there is a risk that donated liver tissue may be dysfunctional, and that the recipient may reject the organ. Richards et al. employed *in vivo* microdialysis to monitor liver grafts through post-reperfusion to give indication of poor or non-functioning grafts [98]. Richards et al. found that none of the patients in the population exhibited poor or non-functional grafts, but the study demonstrated that microdialysis is important for characterizing metabolic changes from pre- to post-transplant grafts [98]. Bergström et al. has also developed a novel microdialysis method to follow known and unknown compounds in a model of liver disease, the first time that unknown analytes were followed in evaluation of a disease state [99]. By employing mass spectrometry and chemometric data evaluation, Bergström et al. discovered new biomarkers and biomarker patterns in a model of induced liver damage [99].

In 2004, Bahlman et al. used *in vivo* microdialysis to evaluate myocardial lactate and pyruvate concentrations in patients undergoing coronary artery bypass grafting and valve surgery [100]. This represented the first time that microdialysis was performed in humans with commercially available microdialysis catheters. Bahlmann et al. measured the lactate and pyruvate levels before, during, and after the cardiopulmonary bypass, and these compounds were shown to be sensitive markers of the myocardial redox status [100]. Thus, microdialysis has provided another tool to evaluate biologically relevant molecules during coronary artery bypass, and may prove to be a vital component of patient monitoring.

Tumors are another area where microdialysis has made inroads in patient monitoring. Most often patients must undergo tumor biopsy to determine whether a mass is benign or malignant. However, this does not provide details regarding the extracellular matrix or the pharmacokinetics effects of anti-tumor drugs. Implantation of a microdialysis probe directly into the tumor mass allows for both drug concentration and metabolism to be monitored, though this procedure is more invasive and

time consuming for the patient. A recent study by Apparaju et al. evaluated the anticancer agent gemcitabine in tumors and tumor-free regions of the brain to better understand the pharmacokinetic behavior of the drug in the CNS [101,102]. Their studies demonstrated that extracellular levels of gemcitabine were lower in tumor-free tissue compared to tumor tissue, suggesting that the tumor exhibits higher drug uptake. Microdialysis is thus an underutilized area of study in cancer research, though more work will surely be done as inpatient methods improve.

## 4 COMPARISON TO OTHER *IN VIVO* METHODS

Microdialysis is not the only *in vivo* method available to quantitate extracellular neurotransmitter levels in the brain. Positron emission tomography (PET) and voltammetry are two other neurochemical methods typically used. PET, a minimally invasive technique, and voltammetry, a more invasive technique will be discussed below. Each of these techniques provides advantages and disadvantages over microdialysis, but all of these methods offer an important glimpse into the neurochemistry of the brain. Work in all of these techniques will hopefully lead to better understanding of neurological diseases and disorders, with a view to creating improved treatments and therapeutics.

### 4.1 PET Imaging

PET imaging is a non-invasive *in vivo* imaging technique which uses radioactive ligands to measure *in vivo* metabolism. In a PET experiment, a small amount of radioactive analyte is administered to the subject. The subject is then placed in a ring of detectors for the $\gamma$-rays that are emitted when the positrons of the radioactive isotope collide with electrons *in vivo*. The information obtained through PET imaging is similar to the *in vitro* technique of autoradiography a spatial map of radioactivity distribution within the tissue [102]. PET has advantages over autoradiography in that it is non-invasive allowing for it to be used in clinical research, while autoradiography is an end point, measuring receptor or transporter levels *in vitro*. PET imaging has a spatial resolution of 3–12 mm compared to a few 100 μm for autoradiography [102]. Although autoradiography has improved spatial resolution, PET imaging done *in vivo* allows for longitudinal studies across the same individual [102]. One major hurdle for PET imaging is the synthesis of the radioactive ligand. In PET imaging, radioisotopes are substituted for one of the atoms making up a common drug, which can then be administered to a patient. Some examples include oxygen-15 ($^{15}O$), nitrogen-13 ($^{13}N$), carbon-11 ($^{11}C$), and fluorine-18 ($^{18}F$). The typical half life of these isotopes is between 2 and 109 min [102]. The short half life of the isotopes used means that temporal resolution is fairly high, usually a few minutes, though continuous monitoring of the patient must take place for as long as the radioactive ligand is available [103].

Synthesis of the positron-emitting radioisotopes used in PET imaging takes place in a cyclotron (a particle accelerator which promotes charged particles to higher

energies) [102,103]. The radioactive ligands are administered to the patient in very small doses, to prevent possible side effects associated with radioactive compounds. When the isotopes reach the tissue of interest, positrons are emitted from the radioactive ligand as it binds with receptors in the area. The positrons do not travel far (~ a few millimeters) before they collide with electrons present *in vivo* (annihilation), and the result is two high intensity γ-rays (photons) traveling in opposite directions. Detectors for the γ-rays are situated in a ring around the subject, and the γ-rays that arrive at near-simultaneously at completely opposite detectors are measured ("coincidence detection"). As a result complex algorithms are used to determine the quantification of the ligand binding that is taking place *in vivo*, which result in images with a high signal to noise ratio [102,103]. Although PET is very sensitive and non-invasive, it is a very expensive and time intensive technique with relatively poor spatial resolution, and as a result it is difficult to justify such huge expenses for rats or mice *in vivo*.

## 4.2 Voltammetry

Within the last 30 years there has been a surge in the development of biosensors to measure biologically relevant molecules *in vivo* (or *in vitro*). Biosensors offer fast temporal resolution (typically a measurement is made in a matter of seconds) and improved spatial resolution where a biosensor can be placed into small compartments or discrete brain regions. The biosensor field originated in 1956, with the development of the oxygen sensor patented by Leland Clark and has been used in medical and environmental research fields for measuring dissolved oxygen [104]. The oxygen sensor has served as the foundation for many of the biosensors in use today, such as the glucose meter [105]. Clark expanded on the oxygen electrode and developed a glucose biosensor by placing the enzyme glucose oxidase in between the membrane and the electrode [105]. Microdialysis has been in competition with biosensor techniques in recent years. Biosensors such as carbon fiber electrodes offer a few advantages over microdialysis enhanced/improved temporal resolution and a less invasive implantation [106]. Voltammetry is considered a biosensor technique because of its ability to detect only one neurotransmitter, as opposed to the complex mixture of analytes typically found in microdialysis.

Although there are a variety of electrochemical techniques to measure biomolecules such as amperometry, chronoamperometry, differential pulse voltammetry, and fast scan cyclic voltammetry (FSCV) the focus of this section will be FSCV. Electrochemical measurements provide another means for measuring biomolecules such as neurotransmitters. Monoamine neurotransmitters such as DA, serotonin and norepinephrine are easily oxidizable, which makes them ideal analytes for electrochemical detection. FSCV uses a microelectrode, which is a carbon fiber (length ~200 μm, diameter ~5 μm) enclosed in a pulled glass capillary to detect electroactive species. A triangle waveform is applied to the microelectrode in FSCV, initially the microelectrode is held at a constant potential versus a reference electrode (Ag/AgCl), then a rapid increase in potential is applied (forward scan) followed by an immediate

return (reverse scan) to the holding potential [107,108]. The potential of the working electrode is scanned between −400 and +1200 mV (versus Ag/AgCl), and the detection of the monoamines occurs at approximately +600 mV. Typically, fast scan rates (greater than 100 V/s) are used and the triangle waveform is completed in less than 10 ms allowing for electrochemical measurements to be made every 100 ms. An advantage of FSCV is that analytes are differentiated from one another based on their peak oxidation and reduction position.

In the work pioneered by Wightman and co-workers, electrodes are lowered into a brain region of interest and the release of DA and DA uptake are measured via oxidation and reduction of a carbon fiber electrode [109,110]. These measurements can be preformed in tissue slices (*in vitro*), anesthetized or freely moving animals to probe the dynamics of neurotransmitter signaling, which are known to occur on the time scale of milliseconds [111–116]. Voltammetry is the only bioanalytical technique that has the temporal resolution to measures changes in neurotransmitter concentrations on this time scale. Therefore, voltammetry is particularly useful for measuring changes in neurotransmission as a result of pharmacological or behavioral manipulations [107,117].

*In vivo* microdialysis work is most consistently compared with voltammetry work performed with carbon fiber electrodes, known as FSCV, a technique that has been in competition with microdialysis for the last 15 years. The carbon fiber electrode used in FSCV has a diameter of approximately 7 μm (20 times smaller than a concentric microdialysis probe), which allows the electrode to discriminate between the core and shell of the nucleus accumbens, a distinction that cannot be made when using microdialysis probes in a mouse model. Measurements are made within 15–30 s, a vast improvement over the minutes long analysis time of microdialysis. DA is the most common analyte measured using carbon fiber electrodes, as it is easily oxidizable and occurs in detectable levels in many brain areas. The limit of detection for carbon fiber electrodes is around 100 nM for DA, approximately 10 times the basal level observed using *in vivo* microdialysis. As a result of the poor limits of detection, an electrical stimulation must be applied to release enough DA in the general vicinity of the carbon fiber (*in vivo* anesthetized animals, see References [118–121]). However, recent evidence has indicated that fast scan cyclic voltammetry can measure burst firing *in vivo* with no electrical stimulation. Fast measurement and small size have helped FSCV gain in popularity, but there are limitations with the carbon fiber biosensor. Unlike microdialysis probes which discriminate based on molecular weight, carbon fiber electrodes are only able to measure one analyte based on its oxidation potential; there are no samples or separation components allowing for detection of multiple analytes. Currently, however, voltammetric methods are not sensitive enough detect basal neurotransmitter levels and chemical selectivity is poor. It is well understood that many neurochemical events are occurring simultaneously *in vitro* and *in vivo*, and voltammetry is limited to examining one neurochemical event at a time. Microdialysis offers the benefit of examining a number of chemically related molecules at the same time, at the sacrifice of temporal resolution (due to the length of time required to obtain a sizeable sample and the length of time necessary for analysis).

## 5 CONCLUSIONS

This chapter highlights some of the benefits and limitations of *in vivo* microdialysis in measuring biologically significant changes in various tissues, with a focus on the brain. *In vivo* microdialysis has been an established neurochemical technique for approximately 35 years, and many improvements can still be made with regards to technical aspects and appropriate data analysis and interpretation. Microdialysis has seen a surge in popularity over the last 15 years, partially due to the advent of genetically altered mice. However, there are two important drawbacks to *in vivo* microdialysis. Tissue damage cause by the probe is still an important limitation, as well as the inability to detect very small, but biologically relevant changes in neurotransmitter levels. These pitfalls have spurred new interest in developing "smaller" probes, such as capillaries that directly sample the extracellular fluid, to both reduce tissue damage and increase sensitivity [16].

The two most common *in vivo* bioanalytical methods for measuring neurotransmitter levels in the extraneuronal signaling space are *in vivo* microdialysis and *in vivo* or *in vitro* voltammetry. Microdialysis can be coupled to either HPLC or CE with electrochemical, mass spectrometry, or fluorescence detection, while voltammetry utilizes a carbon fiber microelectrode to stimulate release, oxidation, and reduction of analytes. Microdialysis has many advantages over voltammetry. It is sensitive enough to detect extracelleular neurotransmitter levels in the basal state. Further, analytes can be separated and identified by retention time or mass, allowing multiple analytes to be analyzed at the simultaneously. However, microdialysis sampling times are long (minutes) and semi-permeable microdialysis membranes are relatively large ($\sim$200 μm) with respect to sampling regions, especially in the mouse brain. The relative size of the probe can lead to significant damage in the region of interest [27,30]. Furthermore, determination of the recovery of neurotransmitters from the extracellular space is complex. Finally, in order to collect sufficient analytes to detect basal neurotransmitter levels, dialysate samples on traditional HPLC columns usually require 10–30 min of collection time. These long sampling times decrease the temporal resolution of *in vivo* microdialysis.

Voltammetry solves many problems of probe damage and temporal resolution. The dynamics of neurotransmitter signaling are known to occur on the time scale of milliseconds, and currently voltammetry is the only bioanalytical technique with the temporal resolution to measure changes on that time scale. However, only one chemical species can be monitored at a time. It is well understood that many neurochemical events are occurring simultaneously, and microdialysis offers the benefit of examining a number of chemically related molecules at the same time. Thus, when using microdialysis temporal resolution is sacrificed for chemical selectivity, multiple analyte detection, and analysis of basal neurotransmitter homeostasis. Recently, microdialysis coupled to CE separations has shown promise for reducing sampling times to seconds [122]. CE is highly advantageous for measuring neuroactive amino acids due to their high concentration in the extracellular space. However, measuring low nanomolar concentrations of the monoamines is much more difficult, and thus not all temporal resolution problems can be avoided [13]. However,

in cases where detections limits are adequate, coupling microdialysis to CE may allow *in vivo* microdialysis to compete temporally with voltammetry. Advances in analytical chemistry have also made it possible to measure low nano- to picomolar concentrations of biologically significant molecules using a variety of detection techniques, such as electrochemical detection, mass spectrometry, and fluorescence detection. If analytical chemistry continues to advance under the current philosophy of "smaller and faster", then real improvements for detecting temporally significant changes in neurotransmission will be achieved [16,78,123]. The future of the microdialysis probe is bright. It is the only technique currently available to measure basal neurotransmitter levels in the brain, and recent work has expanded microdialysis research to many other areas. Future experiments will only expand and improve the current techniques, resulting in new data and insights into brain and body, and elucidating the mechanisms responsible for disease and function.

## REFERENCES


1. Delgado JM, DeFeudis FV, Roth RH, Ryugo DK, Mitruka BM. Dialytrode for long term intracerebral perfusion in awake monkeys. Arch Int Pharmacodyn Ther 1972; 198 (1): 9–21.
2. Ungerstedt U. Microdialysis—principles and applications for studies in animals and man. J Intern Med 1991; 230 (4): 365–373.
3. Bito L, Davson H, Levin E, Murray M, Snider N. The concentrations of free amino acids and other electrolytes in cerebrospinal fluid, *in vivo* dialysate of brain, and blood plasma of the dog. J Neurochem 1966; 13 (11): 1057–1067.
4. Ungerstedt U, Pycock C. Functional correlates of dopamine neurotransmission. Bull Schweiz Akad Med Wiss 1974; 30 (1–3): 44–55.
5. Benveniste H. Brain microdialysis. J Neurochem 1989; 52 (6): 1667–1679.
6. Westerink BHC, De Vries JB. Characterization of *in vivo* dopamine release as determined by brain microdialysis after acute and subchronic implantations: methodological aspects. J Neurochem 1988; 51 (3): 683–687.
7. Zetterstroem T, Sharp T, Ungerstedt U. Effect of neuroleptic drugs on striatal dopamine release and metabolism in the awake rat studied by intracerebral dialysis. Eur J Pharmacol 1984; 106 (1): 27–37.
8. Zetterstrom T, Sharp T, Marsden CA, Ungerstedt U. *In vivo* measurement of dopamine and its metabolites by intracerebral dialysis: changes after d-amphetamine. J Neurochem 1983; 41 (6): 1769–1773.
9. Zetterstrom T, Vernet L, Ungerstedt U, Tossman U, Jonzon B, Fredholm BB. Purine levels in the intact rat brain. Studies with an implanted perfused hollow fibre. Neurosci Lett 1982; 29 (2): 111–115.
10. Weiss DJ, Lunte CE, Lunte SM. *In vivo* microdialysis as a tool for monitoring pharmacokinetics. Trends Analytical Chemistry 2000; 19 (10): 606–616.
11. Chaurasia CS. *In vivo* microdialysis sampling: theory and applications. Biomed Chromatogr 1999; 13 (5): 317–332.

12. Davies MI. A review of microdialysis sampling for pharmacokinetic applications. Anal Chim Acta 1999; 379 (3): 227–249.
13. Davies MI, Cooper JD, Desmond SS, Lunte CE, Lunte SM. Analytical considerations for microdialysis sampling. Adv Drug Deliv Rev 2000; 45 (2–3): 169–188.
14. Lunte CE, Scott DO, Kissinger PT. Sampling living systems using microdialysis probes. Anal Chem 1991; 63 (15): 773A–780A, 776A–778A, 7780A.
15. Kehr J. A survey on quantitative microdialysis: theoretical models and practical implications. J Neurosci Methods 1993; 48 (3): 251–261.
16. Kennedy RT, Watson CJ, Haskins WE, Powell DH, Strecker RE. *In vivo* neurochemical monitoring by microdialysis and capillary separations. Curr Opin Chem Biol 2002; 6 (5): 659–665.
17. Acworth I, Cunningham ML. The measurement of monoamine neurotransmitters in microdialysis perfusates using HPLC-ECD. Methods Mol Med 1999; 22: 219–236.
18. Mateo Y, Budygin EA, John CE, Jones SR. Role of serotonin in cocaine effects in mice with reduced dopamine transporter function. Proc Natl Acad Sci USA 2004; 101 (1): 372–377.
19. O'Dell LE, Parsons LH. Serotonin1B receptors in the ventral tegmental area modulate cocaine-induced increases in nucleus accumbens dopamine levels. J Pharmacol Exp Ther 2004; 311 (2): 711–719.
20. Elmquist WF, Sawchuk RJ. Application of microdialysis in pharmacokinetic studies. Pharm Res 1997; 14 (3): 267–88.
21. Garrison KE, Pasas SA, Cooper JD, Davies MI. A review of membrane sampling from biological tissues with applications in pharmacokinetics, metabolism and pharmacodynamics. Eur J Pharm Sci 2002; 17 (1–2): 1–12.
22. Haskins WE, Wang Z, Watson CJ, Rostand RR, Witowski SR, Powell DH, Kennedy, RT. Capillary LC-MS2 at the attomole level for monitoring and discovering endogenous peptides in microdialysis samples collected *in vivo*. Anal Chem 2001; 73 (21): 5005–5014.
23. Haskins WE, Watson CJ, Cellar NA, Powell DH, Kennedy RT. Discovery neurochemical screening of peptides in brain extracellular fluid by chemical analysis of *in vivo* microdialysis samples. Anal Chem 2004; 76 (18): 5523–5533.
24. Lanckmans K, Sarre S, Smolders I, Michotte Y. Quantitative liquid chromatography/mass spectrometry for the analysis of microdialysates. Talanta 2008; 74 (4): 458–469.
25. Lapainis T, Sweedler JV. Contributions of capillary electrophoresis to neuroscience. J Chromatogr A 2008; 1184 (1–2): 144–158.
26. Borland LM, Shi G, Yang H, Michael AC. Voltammetric study of extracellular dopamine near microdialysis probes acutely implanted in the striatum of the anesthetized rat. J Neurosci Methods 2005; 146 (2): 149–158.
27. Bungay PM, Newton-Vinson P, Isele W, Garris PA, Justice JB Jr. Microdialysis of dopamine interpreted with quantitative model incorporating probe implantation trauma. J Neurochem 2003; 86 (4): 932–946.
28. Clapp-Lilly KL, Roberts RC, Duffy LK, Irons KP, Hu Y, Drew KL. An ultrastructural analysis of tissue surrounding a microdialysis probe. J Neurosci Methods 1999; 90 (2): 129–142.
29. Mitala CM, Wang Y, Borland LM, Jung M, Shand S, Watkins S, et al. Impact of microdialysis probes on vasculature and dopamine in the rat striatum: a combined

fluorescence and voltammetric study. J Neurosci Methods 2008; 174 (2): 177–185 (PMCID: 2713687).

30. Peters JL, Michael AC. Modeling voltammetry and microdialysis of striatal extracellular dopamine: the impact of dopamine uptake on extraction and recovery ratios. J Neurochem 1998; 70 (2): 594–603.
31. Telting-Diaz M, Scott DO, Lunte CE. Intravenous microdialysis sampling in awake, freely-moving rats. Anal Chem 1992; 64 (7): 806–810.
32. Wages SA, Church WH, Justice JB Jr. Sampling considerations for on-line microbore liquid chromatography of brain dialysate. Anal Chem 1986; 58 (8): 1649–1656.
33. Yang H, Wang Q, Elmquist WF. The design and validation of a novel intravenous microdialysis probe: application to fluconazole pharmacokinetics in the freely-moving rat model. Pharm Res 1997; 14 (10): 1455–1460.
34. Zuo H, Ye M, Davies MI. The linear probe: a flexible choice for *in vivo* microdialysis sampling in soft tissues. Curr Sep 1995; 14 (2): 54–57.
35. Scott DO, Lunte CE. *In vivo* microdialysis sampling in the bile, blood, and liver of rats to study the disposition of phenol. Pharm Res 1993; 10 (3): 335–342.
36. Heppert KE, Davies MI. Using a microdialysis shunt probe to monitor phenolphthalein glucuronide in rats with intact and diverted bile flow. Anal Chim Acta 1999; 379 (3): 359–366.
37. Huff JK, Heppert KE, Davies MI. The microdialysis shunt probe: profile of analytes in rats with erratic bile flow or rapid changes in analyte concentration in the bile. Curr Sep 1999; 18 (3): 85–90.
38. Tang A, Bungay PM, Gonzales RA. Characterization of probe and tissue factors that influence interpretation of quantitative microdialysis experiments for dopamine. J Neurosci Methods 2003; 126 (1): 1–11.
39. Parsons LH, Justice JB Jr. Perfusate serotonin increases extracellular dopamine in the nucleus accumbens as measured by *in vivo* microdialysis. Brain Res 1993; 606 (2): 195–199.
40. Lonnroth P, Jansson PA, Smith U. A microdialysis method allowing characterization of intercellular water space in humans. Am J Physiol 1987; 253 (2 Pt 1): E228–E231.
41. Tao R, Ma Z, Auerbach SB. Differential effect of local infusion of serotonin reuptake inhibitors in the raphe versus forebrain and the role of depolarization-induced release in increased extracellular serotonin. J Pharmacol Exp Ther 2000; 294 (2): 571–579.
42. Menacherry S, Hubert W, Justice JB Jr. *In vivo* calibration of microdialysis probes for exogenous compounds. Anal Chem 1992; 64 (6): 577–583.
43. Cosford RJ, Parsons LH, Justice JB Jr. Effect of tetrodotoxin and potassium infusion on microdialysis extraction fraction and extracellular dopamine in the nucleus accumbens. Neurosci Lett 1994; 178 (2): 175–178.
44. Cosford RJ, Vinson AP, Kukoyi S, Justice JB Jr. Quantitative microdialysis of serotonin and norepinephrine: pharmacological influences on *in vivo* extraction fraction. J Neurosci Methods 1996; 68 (1): 39–47.
45. Olson Cosford RJ, Vinson AP, Kukoyi S, Justice JB Jr. Quantitative microdialysis of serotonin and norepinephrine: pharmacological influences on *in vivo* extraction fraction. J Neurosci Methods 1996; 68: 39–47.
46. Parsons LH, Justice JB Jr. Extracellular concentration and *in vivo* recovery of dopamine in the nucleus accumbens using microdialysis. J Neurochem 1992; 58 (1): 212–218.

47. Parsons LH, Justice JB Jr. Quantitative approaches to *in vivo* brain microdialysis. Crit Rev Neurobiol 1994; 8 (3): 189–220.
48. Clement R, Malinovsky JM, Dollo G, Le Corre P, Chevanne F, Le Verge R. *In vitro* and *in vivo* microdialysis calibration using retrodialysis for the study of the cerebrospinal distribution of bupivacaine. J Pharm Biomed Anal 1998; 17 (4–5): 665–670.
49. Hansen DK, Davies MI, Lunte SM, Lunte CE. Pharmacokinetic and metabolism studies using microdialysis sampling. J Pharm Sci 1999; 88 (1): 14–27 (PMCID: 2519818).
50. Schultz KNaK Robert T. Time-Resolved Microdialysis for *In Vivo* Neurochemical Measurements and Other Applications. Annu Rev Anal Chem 2008; 1: 21.1–21.35.
51. Grabb MC, Sciotti VM, Gidday JM, Cohen SA, Van Wylen DGL. Neurochemical and morphological responses to acutely and chronically implanted brain microdialysis probes. J Neurosci Methods 1998; 82 (1): 25–34.
52. Ao X, Stenken JA. Microdialysis sampling of cytokines. Methods 2006; 38 (4): 331–341.
53. Hansen DK, Lunte SM. Determination of kynuernic acid by capillary electrophoresis with laser-induced fluorescence detection. J Chromatogr A 1997; 781 (1–2): 81–89.
54. Chefer VI, Zakharova I, Shippenberg TS. Enhanced responsiveness to novelty and cocaine is associated with decreased basal dopamine uptake and release in the nucleus accumbens: quantitative microdialysis in rats under transient conditions. J Neurosci 2003; 23 (7): 3076–3084.
55. Olson RJ. Quantitative microdialysis under transient conditions. Anal Chem 1993; 65: 1017–1022.
56. Davies MI, Lunte CE. Microdialysis sampling coupled on-line to microseparation techniques. Chem Soc Rev 1997; 26: 215–222.
57. Skoog DA, Leary, JJ. Principles of Instrumental Analysis. Fort Worth: Saunders College Publishing; 1992.
58. Robinson TE, Justice JB Jr. Microdialysis in the Neurosciences. New York: Elsevier Science Publishing Co. Inc.; 1991.
59. Scott RPW. An introduction to chromatography. Chromatogr Sci Ser 1994; 67: 1–13.
60. Weston A, Brown PR. HPLC and CE: Principles and Practices. Academic Press; 1997.
61. Newton AP, Justice JB Jr. Temporal response of microdialysis probes to local perfusion of dopamine and cocaine followed with one-minute sampling. Anal Chem 1994; 66 (9): 1468–1472.
62. Shackman HM, Shou M, Cellar NA, Watson CJ, Kennedy RT. Microdialysis coupled on-line to capillary liquid chromatography with tandem mass spectrometry for monitoring acetylcholine *in vivo*. J Neurosci Methods 2007; 159 (1): 86–92.
63. Dawson LA. Capillary electrophoresis and microdialysis: current technology and applications. J Chromatogr B Biomed Sci Appl 1997; 697 (1 + 2): 89–99.
64. Jorgenson JW, Lukacs KD. Capillary zone electrophoresis. Science 1983; 222 (4621): 266–272.
65. Kennedy RT. Bioanalytical applications of fast capillary electrophoresis. Anal Chim Acta 400: 1999; 163–180.
66. Kennedy RT, Thompson JE, Vickroy TW. *In vivo* monitoring of amino acids by direct sampling of brain extracellular fluid at ultralow flow rates and capillary electrophoresis. J Neurosci Methods 2002; 114 (1): 39–49.

67. Zhou SY, Zuo H, Stobaugh JF, Lunte CE, Lunte SM. Continuous *in vivo* monitoring of amino acid neurotransmitters by microdialysis sampling with on-line derivatization and capillary electrophoresis separation. Anal Chem 1995; 67 (3): 594–599.
68. Ciriacks CMBM. Monitoring d-serine dynamics in the rat brain using online microdialysis-capillary electrophoresis. Anal Chem 2004; 76: 6582–6587.
69. Chen G, Lin Y. Microfabricated devices for sample extraction, concentrations, and related sample processing technologies. In: Wang W, Soper SA, editors. Bio-MEMS: technologies and applications; CRC Press, Taylor & Francis Group, Boca Raton, FL 2007.
70. Crevillen AG, Hervas M, Lopez MA, Gonzalez MC, Escarpa A. Real sample analysis on microfluidic devices. Talanta 2007; 74 (3): 342–357.
71. Mecker LC, Martin RS. Integration of microdialysis sampling and microchip electrophoresis with electrochemical detection. Anal Chem 2008; 80 (23): 9257–964 (PMCID: 2771943).
72. Hsieh Y-C, Zahn JD. Glucose recovery in a microfluidic microdialysis biochip. Sens Actuators B Chem 2005; B107 (2): 649–656.
73. Huynh BH, Fogarty BA, Nandi P, Lunte SM. A microchip electrophoresis device with on-line microdialysis sampling and on-chip sample derivatization by naphthalene 2,3-dicarboxaldehyde/2-mercaptoethanol for amino acid and peptide analysis. J Pharm Biomed Anal 2006; 42 (5): 529–534.
74. Kurita R, Yabumoto N, Niwa O. Miniaturized one-chip electrochemical sensing device integrated with a dialysis membrane and double thin-layer flow channels for measuring blood samples. Biosens Bioelectron 2006; 21 (8): 1649–1653.
75. Sandlin ZD, Shou M, Shackman JG, Kennedy RT. Microfluidic electrophoresis chip coupled to microdialysis for *in vivo* monitoring of amino acid neurotransmitters. Anal Chem 2005; 77 (23): 7702–7708.
76. Erickson BE. Electrochemical detectors for liquid chromatography. Anal Chem 2000; 72 (9): 353A–357A.
77. Shah AJ, de Biasi V, Taylor SG, Roberts C, Hemmati P, Munton R, et al. Development of a protocol for the automated analysis of amino acids in brain tissue samples and microdialysates. J Chromatogr B Biomed Sci Appl 1999; 735 (2): 133–140.
78. Boyd BW, Witowski SR, Kennedy RT. Trace-level amino acid analysis by capillary liquid chromatography and application to *in vivo* microdialysis sampling with 10-s temporal resolution. Anal Chem 2000; 72 (4): 865–871.
79. de Hoffmann E, Stroobant V. Mass Spectrometry: Principles and Applications. 3rd ed. West Sussex: John Wiley & Sons Ltd; 2007.
80. Rouessac F, Rouessac A. Chemical Analysis. 2nd ed. West Sussex, England: John Wiley & Sons Ltd; 2007.
81. Missale C, Nash SR, Robinson SW, Jaber M, Caron MG. Dopamine receptors: from structure to function. Physiol Rev 1998; 78 (1): 189–225.
82. Wise RA, Rompre PP. Brain dopamine and reward. Annu Rev Psychol 1989; 40: 191–225.
83. Cooper JR, Bloom FE, Roth RH. The Biochemical Basis of Neuropharmacology. 8th ed. New York, NY: Oxford University Press; 2002.
84. Daws LC, Callaghan PD, Moron JA, Kahlig KM, Shippenberg TS, Javitch JA, et al. Cocaine Increases Dopamine Uptake and Cell Surface Expression of Dopamine Transporters. Biochemical and Biophysical Research Communications. 2002; 290 (5): 1545–1550.

85. Doyon WM, Howard EC, Shippenberg TS, Gonzales RA. k-Opioid receptor modulation of accumbal dopamine concentration during operant ethanol self-administration. Neuropharmacology 2006; 51 (3): 487–496.
86. Riegel AC, Zapata A, Shippenberg TS, French ED. The abused inhalant toluene increases dopamine release in the nucleus accumbens by directly stimulating ventral tegmental area neurons. Neuropsychopharmacology 2007; 32 (7): 1558–1569.
87. Zapata A, Gonzales RA, Shippenberg TS. Repeated ethanol intoxication induces behavioral sensitization in the absence of a sensitized accumbens dopamine response in C57BL/6J and DBA/2J mice. Neuropsychopharmacology 2006; 31 (2): 396–405.
88. Giros B, Jaber M, Jones SR, Wightman RM, Caron MG. Hyperlocomotion and indifference to cocaine and amphetamine in mice lacking the dopamine transporter. Nature 1996; 379 (6566): 606–612.
89. Sotnikova TD, Beaulieu J-M, Barak LS, Wetsel WC, Caron MG, Gainetdinov RR. Dopamine-independent locomotor actions of amphetamines in a novel acute mouse model of Parkinson disease. PLoS Biol 2005; 3 (8): 1488–1500.
90. Hornykiewicz O. Biochemical aspects of Parkinson's disease. Neurology 1998; 51 (2 Suppl 2): S2–S9.
91. Schmidt S, Barbour A, Sahre M, Rand KH, Derendorf H. PK/PD: new insights for antibacterial and antiviral applications. Curr Opin Pharmacol 2008; 8 (5): 549–556.
92. Davies MI, Lunte CE. Simultaneous microdialysis sampling from multiple sites in the liver for the study of phenol metabolism. Life Sci 1996; 59 (12): 1001–1013.
93. Plock N, Kloft C. Microdialysis—theoretical background and recent implementation in applied life-sciences. Eur J Pharm Sci 2005; 25 (1): 1–24.
94. Tsai TH. Assaying protein unbound drugs using microdialysis techniques. J Chromatogr B Analyt Technol Biomed Life Sci 2003; 797 (1–2): 161–173.
95. Tsai TH, Chen YF, Chen IF, Chen CF. Measurement of unbound caffeic acid in rat blood by on-line microdialysis coupled with liquid chromatography and its application to pharmacokinetic study. J Chromatogr B Biomed Sci Appl 1999; 729 (1–2): 119–125.
96. Tsai TH, Cheng FC, Chen YF, Chen CF. On-line microdialysis coupled with microbore liquid chromatography with ultraviolet detection for continuous monitoring of free cefsulodin in rat blood. J Chromatogr A 2001; 914 (1–2): 83–87.
97. Zolkowska D, Baumann MH, Rothman RB. Chronic fenfluramine administration increases plasma serotonin (5-hydroxytryptamine) to nontoxic levels. J Pharmacol Exp Ther 2008; 324 (2): 791–797.
98. Richards DA, Silva MA, Murphy N, Wigmore SJ, Mirza DF. Extracellular amino acid levels in the human liver during transplantation: a microdialysis study from donor to recipient. Amino Acids 2007; 33 (3): 429–437.
99. Bergstrom SK, Goiny M, Danielsson R, Ungerstedt U, Andersson M, Markides KE. Screening of microdialysates taken before and after induced liver damage; on-line solid phase extraction-electrospray ionization-mass spectrometry. J Chromatogr A 2006; 1120 (1–2): 21–26.
100. Bahlmann L, Misfeld M, Klaus S, Leptien A, Heringlake M, Schmucker P, et al. Myocardial redox state during coronary artery bypass grafting assessed with microdialysis. Intensive Care Med 2004; 30 (5): 889–894.

101. Apparaju SK, Gudelsky GA, Desai PB. Pharmacokinetics of gemcitabine in tumor and non-tumor extracellular fluid of brain: an *in vivo* assessment in rats employing intracerebral microdialysis. Cancer Chemother Pharmacol 2008; 61 (2): 223–229.
102. Frey KA, Minoshima S, Kuhl DE. Neurochemical imaging of Alzheimer's disease and other degenerative dementias. Q J Nucl Med 1998; 42 (3): 166–178.
103. Brunner M, Langer O. Microdialysis versus other techniques for the clinical assessment of *in vivo* tissue drug distribution. AAPS J 2006; 8 (2): E263–E271.
104. Clark LC Jr, Kaplan S, Matthews EC, Schwab L. Oxygen availability to the brain during inflow occlusion of the heart in normothermia and hypothermia. J Thorac Surg 1956; 32 (5): 576–582.
105. Magner E. Trends in electrochemical biosensors. Analyst 1998; 123 (10): 1967–1970.
106. Peters JL, Miner LH, Michael AC, Sesack SR. Ultrastructure at carbon fiber microelectrode implantation sites after acute voltammetric measurements in the striatum of anesthetized rats. J Neurosci Methods 2004; 137 (1): 9–23.
107. Kita JM, Wightman RM. Microelectrodes for studying neurobiology. Curr Opin Chem Biol 2008; 12 (5): 491–496.
108. Robinson DL, Hermans A, Seipel AT, Wightman RM. Monitoring rapid chemical communication in the brain. Chem Rev 2008; 108 (7): 2554–2584.
109. Dayton MA, Ewing AG, Wightman RM. Evaluation of amphetamine-induced *in vivo* electrochemical response. Eur J Pharmacol 1981; 75 (2–3): 141–144.
110. Ewing AG, Wightman RM, Dayton MA. *In vivo* voltammetry with electrodes that discriminate between dopamine and ascorbate. Brain Res 1982; 249 (2): 361–370.
111. Day JJ, Carelli RM. The nucleus accumbens and Pavlovian reward learning. Neuroscientist 2007; 13 (2): 148–159.
112. Day JJ, Roitman MF, Wightman RM, Carelli RM. Associative learning mediates dynamic shifts in dopamine signaling in the nucleus accumbens. Nat Neurosci 2007; 10 (8): 1020–1028.
113. John CE, Jones SR. Voltammetric characterization of the effect of monoamine uptake inhibitors and releasers on dopamine and serotonin uptake in mouse caudate-putamen and substantia nigra slices. Neuropharmacology 2007; 52 (8): 1596–1605 (PMCID: 2041899).
114. Montague PR, McClure SM, Baldwin PR, Phillips PE, Budygin EA, Stuber GD, et al. Dynamic gain control of dopamine delivery in freely moving animals. J Neurosci 2004; 24 (7): 1754–1759.
115. Phillips PE, Robinson DL, Stuber GD, Carelli RM, Wightman RM. Real-time measurements of phasic changes in extracellular dopamine concentration in freely moving rats by fast-scan cyclic voltammetry. Methods Mol Med 2003; 79: 443–464.
116. Phillips PE, Stuber GD, Heien ML, Wightman RM, Carelli RM. Subsecond dopamine release promotes cocaine seeking. Nature 2003; 422 (6932): 614–618.
117. Robinson DL, Hermans A, Seipel AT, Wightman RM. Monitoring rapid chemical communication in the brain. Chem Rev 2008; 108 (7): 2554–2584.
118. Garris PA, Budygin EA, Phillips PE, Venton BJ, Robinson DL, Bergstrom BP, et al. A role for presynaptic mechanisms in the actions of nomifensine and haloperidol. Neuroscience 2003; 118 (3): 819–829.
119. Garris PA, Walker QD, Wightman RM. Dopamine release and uptake rates both decrease in the partially denervated striatum in proportion to the loss of dopamine terminals. Brain Res 1997; 753 (2): 225–234.

120. Kilpatrick MR, Rooney MB, Michael DJ, Wightman RM. Extracellular dopamine dynamics in rat caudate-putamen during experimenter-delivered and intracranial self-stimulation. Neuroscience 2000; 96 (4): 697–706.

121. Walker QD, Rooney MB, Wightman RM, Kuhn CM. Dopamine release and uptake are greater in female than male rat striatum as measured by fast cyclic voltammetry. Neuroscience 2000; 95 (4): 1061–1070.

122. McKenzie JA, Watson CJ, Rostand RD, German I, Witowski SR, Kennedy RT. Automated capillary liquid chromatography for simultaneous determination of neuroactive amines and amino acids. J Chromatogr A 2002; 962 (1–2): 105–115.

123. Bowser MT, Kennedy RT. *In vivo* monitoring of amine neurotransmitters using microdialysis with on-line capillary electrophoresis. Electrophoresis 2001; 22 (17): 3668–3676.

# *Handbook of Analysis and Pharmaceutical Quality: Nano-Methods*

**Katherine M. Tyner**
*Food and Drug Administration, CDER/OPS/OTR/DAPR, Silver Spring, Maryland*

The views presented in this chapter do not necessarily reflect policy of the Food and Drug Administration. The mention of commercial products, their sources, or their use in connection with material reported herein is not to be construed as either an actual or implied endorsement of such products by the Department of Health and Human Services.

## 1 INTRODUCTION

Due to the unique features that arise from materials at the nanometer length scale, it is anticipated that nanotechnology will play a large role in emerging technologies, including those found in the pharmaceutical industry. The impact of drug delivery systems, nano-therapeutics, or combination products could affect treatment of disease by enabling targeted delivery or controlled delivery of therapeutics, which could minimize drug usage or decrease the frequency of drug use, while enhancing the safety profile (1,2). In addition, the ability to design multifunctional particles that can

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing,*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

target, image and treat disease (3), protect fragile therapeutics from enzymatic and chemical degradation (4), and enhance bioavailability due to larger surface areas associated with smaller particles (5) can decrease the number of times a patient needs to be treated thereby boosting patient compliance and lowering treatment costs.

Nanotechnology and its many derivatives have numerous definitions from a wide range of sources (6), though a growing consensus, including the National Nanotechnology Initiative, is defining "nano" as containing features on the dimensions of 1–100 nm (7). The majority of US regulatory agencies including the Food and Drug Administration (FDA) have not, as yet, adopted a specific definition for nanotechnology and related materials[a]. Despite the lack of specific definitions, the FDA has in the past and is currently in the process of reviewing products that contain controlled features on the nanometer length scale.

With the expansion of nanotechnology into the pharmaceutical field, there has been increased pressure for adequate regulation and safety assessment of any resulting materials. Pre-clinical characterization as well as quality control during the manufacturing of nano-therapeutics is fundamental for understanding resulting safety issues stemming from product use. All regulated products are required to submit basic physio-chemical characterization as part of Investigational New Drug (IND) Applications or New Drug Applications (NDA) submitted to the FDA for review (8). Included in the characterization is proper identification of the therapeutic, including structure, composition, crystal structure, quality, stability, and purity, as well as the synthetic route of the therapeutic. Products containing nano-therapeutics are currently subject to all of the same regulatory review steps and characterization requirements that must be completed as a part of IND/NDA submissions for non-nano-material containing products.

Although nano-materials have always been present, the conscious directed assembly and enhancements available for nano-therapeutics may present novel safety and quality control concerns. In addition to the standard physio-chemical information, it has been proposed that products containing nano-materials should be additionally characterized to include information about particle morphology (size and shape), particle size distribution, surface effects (including surface area, reactivity, and coatings), and aggregation/agglomeration effects in relevant systems (9,10).

The unique features of nano-therapeutics can arise from multiple sources. Not only are surface features enhanced over bulk properties, but also size alone can affect biodistribution (11). Dissolution profiles can be altered (5,12), substantially increasing the bioavailability of traditionally insoluble drugs. Surface coatings can mask particles from the immune system or target hotspots in the body (2,3). These novel properties, if applicable to the nano-therapeutic, should be identified and noted for specific characterization.

In conjunction with the above determinations, particle state (aggregation/agglomeration effects) as well as particle morphology will be a key feature to characterizing the materials. Due to the high surface to bulk ratio for nano-materials and the

[a] For the purposes of this chapter, the prefix "nano" will refer to materials that contain at least one dimension ranging between 1 and 100 nm.

subsequent increased importance of surface effects, the raw nano-material may differ significantly from the formulated product. As a result, nano-therapeutics should be characterized in the final formulation in addition to any characterization of the raw material. Specific areas to consider are the aggregation/agglomeration potential of the therapeutic, potential interactions of the nano-material with the rest of the formulation, and changes in the nano-material over the course of its use. In the case of multifunctional materials, it may be desired that the nano-material changes over time (i.e., release of the therapeutic, targeting, degradation, etc.). In terms of multifunctional therapeutics, all components, their interactions with the material and the final formulation, and any potential release, intentional or not, should be examined in detail.

Table 1 details some of the physio-chemical analysis needed for therapeutics during the NDA/IND submission process as well as validation of methodology. Since some nano-therapeutics may be considered combination products, there may be additional considerations for these materials during characterization.

Determining how the nano-therapeutic can change, or if it can change other components of the formulation will play a major role in both characterization and quality control of the product. Although this type of characterization is required for all regulated materials, the characterization of nano-materials within a formulated product can be challenging in terms of appropriate instrumentation. The size and

**TABLE 1 Guidelines for CMC Characterization of Various Therapeutics and Methodologies and Suggested Additional Information for Nanoproducts**

| Therapeutics[a] | Combination Products[b] | Methods Validation[c] | Nanoproducts (Suggested)[d] |
|---|---|---|---|
| Identity | Leachables/ extractables | Method principle | Size |
| Potency | Stability of drug | Sampling procedure | Size distribution |
| Quality | Stability of device | Equipment | Shape |
| Purity | Drug adhesion/ absorbtion | Reagents | Aggregation/ agglomeration |
| Structure | Breakdown of inactive parts | Preparation of materials | Charge |
| Crystal struture | | Procedure | Surface moeity analysis |
| Synthetic/ manufacturing route | | Calculations | Stability |
| Stability | | Reporting of results | |
| Bioavailability | | | |

a) Guidance for Industry, Content and format of investigational new drugs (INDs) application for Phase 1 studies of drugs, including well-characterized, therapeutic, biotechnology-derived products, November 1995. Guidance for Industry, INDs for Phase 2 and Phase 3 Studies, chemistry, manufacturing, and controls information, May 2003.

b) Guidance for Industry and FDA Staff, Early Development Considerations for Innovative Combination Products, September 2006.

c) Draft Guidance Guidance for Industry, Analytical Procedures and Methods Validation, Chemistry, Manufacturing, and Controls Documentation, August 2000.

d) Warheit D. How meaningful are results of nanotoxicity studies in the absence of adequate material characterization. Toxicological Sciences 2008;101: 183–185.

complexity of nano-therapeutics can sometimes cause the resulting product to fall between the range and sensitivities of traditional instrumentation used for characterizing small molecules and/or bulk materials. Recent advances as well as application/modification of older technologies have emerged to address this issue, but there still can be shortcomings in terms of adequately characterizing different aspects of nano-therapeutics. In terms of quality control during manufacturing, equipment that has traditionally been available for nano-materials characterization may not be amenable to in-line processes, or may have long characterization times. There may, however, be methods that are adaptable from small molecule pharmaceuticals to nano-pharmaceuticals, and these methods need to be evaluated and validated for generalized use. In addition, if the instrumentation is not commonly used in an industrial setting, some of the concepts found in FDA guidances concerning method validation may need to be considered (13,14).

While the possibilities for novel nano-therapeutics are vast, the choices for characterization can also be quite extensive. Several key areas to focus on when choosing instrumentation include accuracy, precision, and detection limit of the instrument, as well as what type of information is being provided (bulk, statistical, image, etc.). There have been several review articles on basic nanoparticle characterization (15–18). The following chapter will provide an overview of methods typically used to characterize nano-materials specifically relating to the pharmaceutical industry, their relevance to the field, and what type of information is provided by the method. Figure 1 lists the acronyms used for this chapter.

## 2 INSTRUMENTATION OVERVIEW

Morphology has been, by far, the most common consideration when it comes to characterizing many nano-materials. While other properties may be as or more important than morphology, discussions about the size, shape, and related features of a product will factor into any subsequent characterization of nano-therapeutics. Morphological characterization can encompass not only the pertinent size and shape of a material, but also the size distribution of the features, and the aggregation/agglomeration status. It is important to remember that "nano" does not necessarily imply particle. The prefix "nano" has been used for particles, platelets, tubes, wires, or features on an extended material, and so on (19). Nano-materials can be individual moieties or embedded within a matrix, with each new feature able to impart different characteristics to the overall material.

In terms of determining size, it can be important to discriminate between primary particles, aggregates, and agglomerates, which can be defined as the following (20,21):

*Primary particle*: Individual crystals or discreet particles of a material.

*Aggregate*: Primary particles that associate during synthesis or post-synthesis conditions such that the overall surface area of the new moiety is reduced as compared to the individual primary particles. Aggregates will not revert back to primary particles under normal conditions.

| | |
|---|---|
| AAS | Acoustic Attenuation Spectroscopy |
| ADFI | Annular Dark-field Imaging |
| AES | Auger Electron Spectroscopy |
| AFM | Atomic Force Microscopy |
| APFIM | Atom-Probe Field-Ion Microscopy |
| APT | Atom-Probe Tomography |
| ATOFMS | Aerosol Time-of-Flight Mass Spectrometry |
| AUC | Analytical Ultracentrifugation |
| BET | Brunauer, Emmett, Teller |
| BSE | Backscattered Electrons |
| CE | Capillary Electrophoresis |
| CHDF | Capillary Hydrodynamic Fractionation |
| CMC | Chemistry, Manufacturing, and Control |
| DLS | Dynamic Light Scattering |
| DMA | Differential Mobility Analysis |
| EBSD | Electron Back-scatter Diffraction |
| ED | Electron Diffraction |
| EDAX | See EDS |
| EDS | Energy Dispersive X-ray Analysis |
| EDX | See EDS |
| EELS | Electron Energy Loss Spectroscopy |
| EM | Electron Microscopy |
| ESA | Electroacoustic Spectroscopy |
| ESEM | Environmental Scanning Electron Microscopy |
| FDA | Food and Drug Administration |
| FFF | Field Flow Fractionation |
| FIB | Focused Ion Beam |
| FIM | Field-Ion Microscopy |
| FWHM | Full Width Half Maximum |
| GE | Gel Electrophoresis |
| GPC | Gel Permeation Chromatography see SEC |
| HPLC | High Performance Liquid Chromatography |
| IAP | Imaging Atom Probe |
| ICP-MS | Inductively Coupled Plasma Mass Spectrometry |
| IND | Investigational New Drug |
| LSCM | Laser Scanning Confocal Microscopy |
| MRI | Magnetic Resonance Imaging |
| NAA | Neutron Activation Analysis |
| NDA | New Drug Application |
| NNI | National Nanotechnology Initiative |
| NSOM | Near-field Scanning Optical Microscopy |
| OM | Optical Microscopy |
| PCCS | Photon Cross Correlation Spectroscopy |
| PCS | Photon Correlation Spectroscopy (see DLS) |
| PDI | Polydispersity Index |
| PGSE NMR | Pulsed Gradient Spin-Echo Nuclear Magnetic Resonance |
| PHA-UPC | Pulse Height Analyzer Ultrafine Condensation Particle Counter |
| PP | Primary Particle |
| PSD | Particle Size Distribution |

| | |
|---|---|
| RI | Refractive Index |
| RS | Raman Spectroscopy |
| SAM | Scanning Auger Microscopy |
| SANS | Small Angle Neutron Scattering |
| SAXS | Small Angle X-Ray Scattering |
| SE | Secondary Electrons |
| SEC | Size Exclusion Chromatography |
| SEM | Scanning Electron Microscopy |
| SIMS | Secondary Ion Mass Spectrometry |
| SLS | Static Light Scattering |
| SMPS | Scanning Mobility Particle Sizer |
| SNMS | Secondary Neutral Mass Spectrometry |
| SPM | Scanning Probe Microscopy |
| SPR | Surface Plasmon Resonance |
| STED | Stimulated Emission Depletion Microscopy |
| STEM | Scanning Transmission Electron Microscopy |
| STM | Scanning Tunneling Microscopy |
| TEM | Transmission Electron Microscopy |
| UC | Ultracentrifugation |
| VPSEM | Variable Pressure Scanning Electron Microscopy |
| WDS | Wavelength Dispersive Spectroscopy |
| XDC | X-Ray Disk Centrifuge |
| XRD | X-Ray Diffraction |
| XRF | X-Ray Fluorescence |

***FIGURE 1*** *Common acronyms for nano-methods.*

*Agglomerate*: Loosely formed features made up of primary particles, aggregates, or a combination of the two. The overall surface area is not reduced, and the original particles can be recovered, depending on conditions.

Each of these features can contribute in different ways to the overall performance of a product (e.g., viscosity for agglomerates, surface reactivity for aggregates and primary particles). Depending on the product and application, therefore, it can become important to identify and characterize each size feature present in a formulation. Figure 2 shows various nano-entities.

For many methods, there will be a distinction between what type of information is provided by a particular instrument. Results can be a bulk average of the entire sample population or can measure individual, single, moieties. Both types of analysis have their advantages and disadvantages. Bulk measurements, on average, are rapid and more amenable to in-line analysis. They do not always, however, identify small populations of outliers that may be relevant for safety or quality control profiles. Instruments providing population profiles on single particles, while able to provide specific information for individual and small populations of a product, can have difficulty obtaining enough data to make statistically significant conclusions and can sometimes require lengthy analysis times. Figure 3 shows the results of a bulk method (DLS) and a single particle method (SEM) for 30 nm gold nanoparticles.

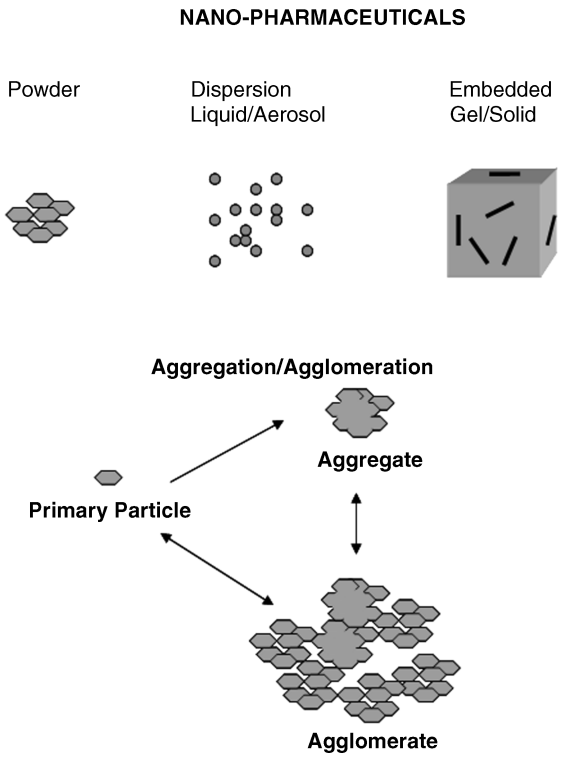


***FIGURE 2*** *Nano-materials may be individual entities or contained in a much larger material. Depending on conditions, primary particles can aggregate or agglomerate into larger entities.*

In general, compositional determination of the overall formulation, including the chemical composition, purity, and crystal structure of a material, will be possible using methods available for small molecule therapeutics (e.g., if crystal structure is needed, then X-ray diffraction would be an appropriate method to use). Characterization techniques such as ICP, IR, NMR, X-ray diffraction (XRD), and so on have already been discussed in previous chapters and will not be discussed further in this chapter.

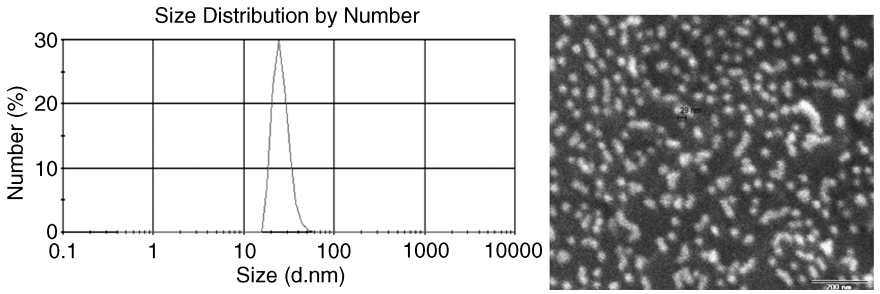


***FIGURE 3*** *Size evaluation of gold nanoparticles by a bulk (DLS) method versus a single particle (SEM) method.*

**TABLE 2 Information Obtained from Nano-Methods**

| | Morphology | | | | | | Composition | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Instrument | Resolution[a] | Image | Shape | Bulk Dispersion | Average/ Single | Environmental | Detected | Limit of Detection[a] | Lateral Resolution[a] | Bulk Comp | Surface Info | Crystal | Comments |
| **TEM** | Sub nm | Y | Y | N | S | Y, but not common | | | | | | | Extensive sample preparation |
| ED | Sub nm | | | | S | | | | | Y | N | Y | |
| EELS | Sub nm | | | | S | | All elements | 10 ppm | Sub nm | Y | Y | | |
| EDS | 1 nm | Y | Y | N | A | | B and heavier | 50 ppm | Sub nm | Y | Y | | |
| **SEM** | 5 nm | Y | Y | N | S | Y | | | | | | | |
| BSE | 5 nm | Y | Y | N | S | | | | | Y | | | Brightness increases with atomic number |
| EDS | Over 100 nm | Y | Y | Y | A | | B and heavier | 10 ppm | 50 nm | Y | Y | | |
| WDS | Over 100 nm | Y | Y | | | | Be and heavier | 10 ppm | 50 nm | Y | Y | | |
| FIB | 5 nm | Y | Y | Y | S | N | | | | | | | Can etch into a sample to determine 3D distribution |
| EBSD | 10 nm | | | | | | | | | | | Y | |
| AES | | | | | A | N | All except H and He | 0.1 atomic % | | Y | Y | N | Chemical state and reactivity can be determined |
| SAM | 1 nm | Y | Y | Y, with sputtering | S | N | All except H and He | 0.1 atomic % | 1 nm | Y | Y | N | Chemical state and reactivity can be determined |
| **STEM** | Sub nm | Y | Y | N | S | Y, but not common | | | | | | | |

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| EDS | 1 nm | Y | Y | N | Both | | B and heavier | 50 ppm | Sub nm | Y | Y | | Can characterize individual particles with good resolution |
| ADFI | Sub nm | Y | Y | N | S | | | | | Y | | | |
| **AFM** | Sub nm | Y | Y | N | S | Y | | | Sub nm | N | Y | | Information depends on tip and detector. Various surface information including forces and bonding |
| **STM** | Sub nm | Y | Y | N | S | Y | | | | | | | Metallic or semi-conducting surface needed |
| **NSOM** | 25–100 nm | Y | Y | N | S | Y | | | 25–100 nm | N | Y | | RI and chemical structure |
| **Cytoviva** | 90 nm | Y | Y | Y | S | Y | | | | | | | Information similar to typical OM |
| **STED** | 50 nm | Y | Y | Y | S | Y | | | | | | | Information similar to typical OM |
| **FIM** | Sub nm | Y | Y | Y | S | N | All | Atomic | Atomic | Y | Y | Y | Can etch away surface layers to get a (small) bulk profile. Material must conduct electricity. |

a) Resolution is best possible. Sample and sample conditions may lower the achievable resolution.

In some cases, however, traditional small molecule analytical techniques may have special applications for nanotechnology (e.g., XRD used to determine the primary particle size of a nano-material). In other cases, techniques can characterize the composition of specific particles to allow for individual analysis of particles, which can become important for impurity analysis. Finally, other techniques are able to provide specific information about the surface of nano-materials. Techniques that provide specific nano-characterization will be discussed below. In-depth fundamentals of the methods are not provided and instead, the reader is referred to other reference books and papers, a partial list of which may be found in the following references (22–39, 49–51).

*A note on instrument resolution*: This chapter and the tables contained within it report the best achievable instrumental resolution at the time the chapter was written. Often sample or measurement conditions will lower the achievable resolution, sometimes quite dramatically. For example, SEM resolution can be as small as 5 nm. When introducing wet samples into an environmental chamber, however, the instrument may only be able to resolve 100 nm features. Thus, if resolution is important to a particular measurement, the effect that the sample and the sample conditions may have on the instrument resolution should be kept in mind when choosing analytical techniques. Often the introduction of moisture or pressure (i.e., environmental conditions) will lower an instrument's resolution capabilities. Table 2 lists the different microscopy methods and the type of information provided.

## 2.1 Microscopy

Microscopic methods are popular nano-methods as they often allow one to directly visualize individual moieties of the nano-material. Such visualization allows for extensive morphological evaluation. As individual particles are analyzed, the techniques usually provide statistics of individual size and shape populations of a material. The type of microscopy used will determine both lateral and depth resolution. In addition, many microscopies may be coupled with other techniques to provide compositional information.

***2.1.1 Electron Microscopy (EM)*** EM methods are prevalent in nano-material characterization due to their high resolution imaging capabilities. In each method, an electron beam interacts with a sample to produce an image. By utilizing results from different outcomes of the electron beam interacting with the sample, different types of information, including compositional, may be obtained. Scanning electron microscopy (SEM) and transmission electron microscopy (TEM) are the most common methods used, although modifications to each have produced a wide sub-set of instrumentation.

*TEM* (Transmission electron microscopy): A high resolution technique that provides an image of the sample. An electron beam is transmitted through a thin sample specimen. Regions where the electrons are scattered appear dark on the viewing screen. Samples are very thin ($< 100$ nm) and are often analyzed in high vacuum.

Attachments to TEM:

ED (Electron diffraction): Electrons from the TEM beam are scattered elastically to produce a diffraction pattern. By analyzing the pattern, crystal morphological information may be obtained. Small regions within the sample can be analyzed with this technique.

EELS (Electron energy loss spectroscopy): The energy loss of low-energy electrons hitting the sample is examined through an attached detector. The energy spectrum of the electrons provides information about the composition and structure of materials. Sub-nanometer resolution is possible with this method.

EDS (Energy dispersive spectroscopy): The electron beam interacts with the sample, causing electrons to be ejected from the inner shell of a sample's atoms, forming an electron hole. An electron from an outer shell fills this hole, causing energy release in the form of an X-ray. The X-ray spectral lines are characteristic for individual elements, allowing for compositional determination.

*SEM* (Scanning electron microscopy): A high resolution imaging technique whereby an electron beam is scanned over the surface of a sample. The interaction of the beam with the sample surface causes electrons to be scattered or absorbed, resulting in the emission of electrons and electromagnetic radiation that can be detected to produce images of the sample surface as well as used to identify elements within the sample. As the electron beam is interacting with the surface of the material, it is an inherently surface method.

Attachments to SEM:

BSE (Back scattered electrons): BSE are high energy electrons from the scanning beam that are reflected back out of the sample through elastic collisions with the sample and can be collected with a specific detector. An attribute of BSE is that the brightness of the image increases with atomic number, and thus high contrast may be achieved between components of different elemental compositions.

EDS: See TEM.

WDS (Wavelength dispersive spectroscopy): As with EDS, characteristic X-rays are produced from the electron beam interacting with the sample. How the X-rays are separated differentiates the two techniques. Typically, sensitivity is improved with WDS.

FIB (Focused ion beam): A sample is bombarded with an ion beam that is used to etch away the sample. As the sample is etched away, structural and compositional data can be obtained with SEM analysis from the removed layers or from the walls of an etched hole, providing morphological and compositional data in three dimensions.

EBSD (Electron back-scatter diffraction): This technique analyzes diffraction patterns of backscattered electrons. Patterns are analyzed to provide information on crystal phase and morphology of the material.

AES (Auger electron spectroscopy): The electron beam interacts with the sample, causing electrons to be ejected from the inner shell of a sample's atoms, forming an electron hole. An electron from an outer shell fills this hole, causing energy release in the form of an energetic electron. The energy of the electron is characteristic of elements, allowing for compositional information about a sample surface. Elemental identification, quantification, chemical states, and chemical reactivity are possible with this method.

SAM (Scanning Auger microscopy): An electron microscope specifically designed to analyze Auger electrons is used in this technique. High resolution compositional images may be produced. The method can be combined with sputtering to get in-depth profiles.

*STEM*: (Scanning transmission electron microscopy). This method combines the high resolution capabilities of a TEM with the scanning raster beam of a SEM.

Attachments to STEM:

EDS: SEE TEM.

EELS SEE TEM.

ADFI (Annular dark-field imaging): In this technique, a detector collects electrons that are diffracted, rather than transmitted. As with BSE, this technique provides contrast that is highly sensitive to atomic number.

***2.1.2 Scanning Probe Microscopy (SPM)*** SPM microcopies are also popular techniques used to analyze nano-materials due to their high resolution capabilities. In SPM, a cantilever is used to scan a specimen surface. Depending on the cantilever, sample conditions, and type of SPM, different properties of the sample may be measured under environmental conditions on a wide range of substrates. Atomic force microscopy is the most common SPM method used due to its high versatility in analyzing many different sample characteristics.

*AFM* (Atomic force microscopy): This method is a high resolution probe microscopy that can achieve atomic resolution by scanning or tapping a cantilever on or close to the sample surface. In addition to height imaging, forces, chemical bonding, and a wide variety of sample-probe interactions may be measured. Three-dimensional surface profiles may be obtained with this technique in contact, non-contact and dynamic modes. The samples can be imaged in a variety of environmental (including wet) conditions.

*STM* (Scanning tunneling microscopy): This method can achieve atomic resolution using quantum tunneling to construct three-dimensional images of a sample. Samples must be very flat and free from contamination. Samples must be either conductive or semi-conductive.

*NSOM* (Near-field scanning optical microscopy): A probe method that breaks the optical resolution of light by placing the detector close to the specimen surface in order to obtain the high frequency spatial information found in the evanescent fields close to the sample surface. Properties such as refractive index and chemical structure can be studied with this technique. Contrast techniques available for use with optical microscopy can be used with NSOM.

**2.1.3 Optical Microscopy (OM)** In general, OM methods cannot accurately size nano-scale materials due to optical resolution limits. Advances in some optics and illumination systems, however, can now resolve particles smaller than 100 nm (40,41). Even without the spatial resolution, if the nano-material has significant contrast over the background material/formulation, the particles can often be detected, if not sized. In some cases, this type of detection will allow for sufficient characterization. Such detection, coupled with some optical methods being able to examine materials in depth, may allow for bulk distribution analysis of the material.

*High Resolution Illuminator* (*Cytoviva®*): The standard microscope condenser is replaced by a high resolution illuminator with a fixed-geometry. The result is highly collimated light that is focused onto the surface, allowing for a high signal to noise ratio and a corresponding increase in resolution. This method maintains OM features, such as imaging in environmental conditions, while resolving individual nano-sized features.

*STED* (Stimulated emission depletion microscopy): The optical diffraction barrier is broken in far-field fluorescence confocal microscopy. Spontaneous fluorescent emission is inhibited at the periphery of the diffraction limited spot of a confocal microscope. This results in a fluorescent spot (effective point-spread-function) that is smaller than the diffraction limited dimensions, allowing for individual nanoparticle detection while maintaining confocal capabilities.

**2.1.4 Other Microscopy** *FIM* (Field-ion microscopy): A sharp metal tip is placed in an ultra high vacuum environment, which is then filled with helium or neon. The tip is cooled and a positive voltage is applied across the tip. Repelled ions are detected and form an image.

Derivatives of FIM:

- APFIM (Atom-probe field-ion microscopy): Time of flight spectroscopy is couple with FIM to allow for 3D reconstructions of samples with chemical profiles at the atomic level.
- IAP (Imaging atom probe): Ions emitted from the surface are mass analyzed by "time-gating" the detector. Crystal information and nano-material distribution can be determined.
- APT (Atom-probe tomography): A three dimensional atomic reconstruction of a surface can be generated by using a position-sensitive detector.

## 2.2 Scattering Methods

Scattering methods are also very common in the analysis of nano-materials. For the most part, these methods provide a bulk measurement. Depending on the method, assumptions about the sample (i.e., spherical shape for light scattering) are inherent in the measurement calculations. Many of the methods are easily automated. Table 3 lists the different methods and the type of information provided.

**TABLE 3 Information Obtained from Nano-Methods**

| | MORPHOLOGY | | | | | | Composition | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Instrument | Resolution[a] (nm) | Image | Shape[b] | Bulk Dispersion | Average/ Single | Environmental | Detected | Limit of Detection[a] | Lateral Resolution[a] | Bulk Comp | Surface Info | Crystal | Comments |
| **DLS** | 0.5 nm | N | N | Y | A | If solution based | | | | | | | Most calcs assume spherical shape. |
| **SLS** | 20 nm | N | N | Y | A | If solution based | | | | | | | Absolute molecular weight and concentration may be determined. |
| **PCCS** | 1 nm | N | N | Y | A | If solution based | | | | | | | Absolute molecular weight and concentration may be determined. Calcs assume spherical shape |
| **XRD** | <200 nm | N | Y, C | N | A | Y | | Depends on count time | | Crystal phase | | Y | Primary particles sized. |
| **SAXS** | 1–100 nm | N | Y, C | Y | A | Y | | | | | | | |
| **SANS** | 1 nm | N | Y, C | Y | A | Y | | | | | | | |
| **SIMS** | 20 nm | Y | Y | Y | Both | N | All elements | ppb | 20 nm | Y | Y | N | Inherently destructive. Needs special detector for imaging. |
| **SNS** | 50 nm | Y | Y | Y | Both | N | All elements | ppb | 50 nm | Y | Y | N | See SIMS |
| **FIB** | 5 nm | Y | Y | Y | S | N | | | | | | | Inherently destructive |
| **FFF** | 1–100 nm | N | Y, C | Y | A | If solution based | Surface funct | | | N | Y | N | Size limit and surface information dependent on sub-technique |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **AUC** | 1 nm | N | Y, C | Y | A | If solution based | | | | | | |
| **CHDF** | 10 nm | N | N | Y | A | If solution based | | | | | | |
| **XDC** | 5 nm | N | Y, C | Y | A | If solution based | | | | | | |
| **HPLC** | <1 nm | N | Y, C | Y | A | If solution based | Surface funct | | | Y | N | May not handle very concentrated systems. Surface info dependent on column |
| SEC | <1 nm | N | N | Y | A | If solution based | | | | | | |
| **AAS** | 5 nm | N | N | Y | A | If solution based | Surface charge | | | Y | N | Concentrated solutions are possible |
| **ATOFMS** | 30–300 nm | N | N | N | S | N | Charged species | 1–800 Da | N | N | N | Single particle analysis |
| **RS** | 1–100 nm | N | N | N | A | Y | molecular species | ppm | Y | Chemical bonding | Y | Primary particles detected. Lateral resolution of imaging is not in the nano range |
| **PGSE NMR** | 1 nm | N | Y, C | N | A | N | | | Y | Y | N | Dilue systems. Bulk compositional information with other NMR techniques |
| **DMA** | 1 nm | N | N | Y | A | N | | | | | | Aerosol based method. Dilute samples needed |
| **PHA UCPC** | 3–10 nm | N | N | N | A | N | | | | | | |

a) Resolution is best possible. Sample and sample conditions may lower the achievable resolution.
b) Calculations (C) used to determine shape effect on measurements.

**2.2.1 Light Scattering** Light scattering techniques, coupled with EM analysis, are probably the most common methods used to measure nanoparticles in a solution. The results are presented as an average size distribution, and for most methods, the calculations assume that the particle is spherical.

*DLS* (Dynamic light scattering): This is also known as photon correlation spectroscopy. This method uses scattered laser light to measure the diffusion of particles dispersed in a liquid. By utilizing the Stokes-Einstein equation (refer to Section .3), the hydrodynamic radius of the particle may be calculated.

*PCCS* (Photon cross correlation spectroscopy): This method is a modification of DLS and uses a three-dimensional cross correlation technique that suppresses multiple scattering of the laser light, allowing for high sample concentrations to be measured. Particle size and stability of suspensions may be obtained using this technique.

*SLS* (Static light scattering): This method measures light scattered by a particles in solution, often at a number of angles. The time averaged scattering intensities are detected at multiple angles, with the scattering pattern depending on the shape and size of the particles. The average radius of gyration is generated with this method. In addition, the intensity of the scattered light is proportional to the weight-average molecular weight and concentration of the particles.

**2.2.2 X-Ray Methods** *XRD* (X-ray diffraction): X-rays are diffracted from randomly oriented crystals in a sample, providing information on the sample's crystal structure. Polymorph identification and quantification may be obtained. In addition, if the primary crystallites are below ~200 nm, an estimate of the particle size and particle size distribution may be calculated through line broadening analysis.

*SAXS* (Small angle X-ray scattering): Elastic scattering of X-rays is recorded at very low angles (0.1–10°). Information about bulk particle sizes, shapes, distribution and surface to volume ratio may be obtained. Unlike XRD, a crystalline sample is not needed.

**2.2.3 Electron Scattering** See Electron Microscopy.

**2.2.4 Neutron Scattering** *SANS* (A complementary technique to SAXS): A beam of neutrons is elastically scattered off of a sample to determine structures between 1 and 100 nm. Particle size may be determined as well as size distributions. Size data are reported in the form of radius of gyration. A feature of SANS is its sensitivity to light elements such as hydrogen and carbon.

**2.2.5 Ion Scattering** *SIMS* (Secondary ion mass spectrometry): A sample surface is bombarded with an ion beam. The emission of entities from the surface (electrons, charged and neutral atomic and molecular fragments) are measured and analyzed by mass spectrometry.

*SNMS* (Secondary neutral mass spectrometry): Similar to SIMS, but ejected neutral atoms are analyzed after post-ionization.

*FIB* (Focused ion beam): A technique similar to SEM, but using an ion beam (usually $Ga^+$) rather than electrons. Sputtered ions or secondary electrons are collected to form an image. In addition, FIB can be used to flatten and mill samples for TEM analysis.

## 2.3 Separation Methods

Not surprisingly, separation techniques typically provide better resolution between different size populations of a sample. Different detection methods and detectors may impact the resolution and applicability of the measurement to nano-therapeutics. Table 3 lists the different methods and the type of information provided.

*AUC* (Analytical ultracentrifugation): In this method, a sample is spun and the concentration distribution in a centrifuge cell is monitored in real time through an optical detection system. Sample purity, shape, size, size distribution, and molecular weight may all be determined with this method.

*XDC* (X-ray disc centrifuge particle size analyzer): Gravitation and centrifugal sedimentation are combined in this method to allow for a narrow separation of particle population distributions. Particles are detected by X-ray attenuation.

*FFF* (Field flow fractionation): This method is solution based separation method where a solvent stream is introduced into a thin cell, with different forces applied perpendicular to the sample flow causing separation and sizing of a wide range of samples. Many varieties of this method exist, and some are listed below.

*Sedimentation*—The channel is spun resulting in a separation of the materials at a right angle to the sample.

*Thermal*—A temperature gradient is produced across the channel with particles driven to the colder side.

*Hydraulic*—A cross flow over the sample pushes particles to the channel wall, which is permeable. The pore size determines the lower size limit.

*Electric*—A potential gradient is placed across the channel, separating out species of different charges.

*Steric*—The opposite channel wall provides a steric barrier, with the smaller particles able to get closer to the wall than larger particles.

*CHDF* (Capillary hydrodynamic fractionation): This method flows a sample through a capillary tube. The particles elute out of the capillary tube based on their hydrodynamic radius, and are detected by a UV detector.

*HPLC* (High performance liquid chromatography): This is a form of column chromatography that pushes a liquid mobile phase through a stationary column. Depending on the column packing material, entities may be separated by polarity, size, surface charge, and so on. A detector monitors the retention time of the entities as they elute from the column.

*SEC* (Size exclusion chromatography): This is also known as gel permeation chromatography and may be part of an HPLC system. Particles are separated by their hydrodynamic volume.

## 2.4 Miscellaneous Methods

Table 3 lists the different methods and the type of information provided.

*AAS* (Acoustic attenuation spectroscopy): Also know as acoustic spectroscopy. The attenuation and sound speed of ultrasound pulses are measured over a wide range of frequencies. The generated spectra can be used to calculate particle size distributions.

*ATOFMS* (Aerosol time-of-flight mass spectrometry): This method uses a single particle mass spectrometry system that analyzes the size and chemical composition of single particles in aerosol form.

*RS* (Raman spectroscopy): A spectroscopic technique analyzing vibrational and rotational modes in a sample. Analyzing the line broadening and peak asymmetry in a spectrum can provide information about the particle size.

PGSE NMR (Pulsed gradient spin-echo nuclear magnetic resonance): The hydrodynamic radius of small particles in a dilute solution may be measured with this NMR method by measuring the self-diffusion coefficient. The concentration dependence of diffusion may also be determined with this method.

*DMA* (Differential mobility analysis): An aerosol based system that analyzes particles based upon their electric mobility, which is based on both size and charge.

*SMA* (Scanning mobility particle sizer): An aerosol based system used in conjunction with DMA to obtain a complete particle size distribution by scanning the voltage in the DMA system.

*PHA-UPC* (Pulse height analyzer ultrafine condensation particle counter): This is an aerosol method where supersaturated vapor condenses onto small nanoparticles and are detected and counted with an optical detector. Small aerosol concentrations may be determined.

***2.4.1 Bulk Surface Methods*** There are also several bulk methods that are commonly used to provide information about nano-materials. Included in these methods are various titrations for surface charge and surface functionalization.

*BET* (Brunauer, Emmett, Teller): A method named after the authors of the journal article describing the rules of gas molecule adsorption onto solid surfaces. Through application of their theory, specific surface area of materials may be measured.

*Zeta potential*: The electrostatic potential from ions on the surface of a colloidal particle (electrical double layer) is measured by a laser when a known E field is applied across a sample. Electrophoretic mobility is related to the velocity of particle movement, causing a frequency shift in the incident laser beam and allowing the zeta potential to be calculated.

*ESA* (Electroacoustic spectroscopy): This method is related to AAS. The electric and acoustic fields are measured in a sample to determine the zeta potential of the particles.

*GE* (Gel electrophoresis): An electric field is applied across a gel matrix. Compounds are separated out by electric charge and frictional forces.

*CE* (Capillary electrophoresis): Also capillary zone electrophoresis. An electric field is applied across a small capillary filled with an electrolyte. All ionic species are

pulled through the tube in the same direction by electro-osmotic flow. Separation is due to electrophoretic mobility, and moieties are typically detected by UV.

*SPR* (Surface plasmon resonance): An optical technique used to measure changes in the RI of a surface when different moieties bind to it. Binding interactions between different moieties may be examined with this technique.

## 3 NANO-METHODS: NOTES AND EQUATIONS

All analytical methods will have their benefits and limitations. Especially at the beginning of the characterization process, it is beneficial to compare two or more techniques to account for each instrument's suitability for analysis.

### 3.1 Microscopy

Microscopy methods allow for individual particles to be analyzed. While this data can provide specific details for the sample population, in order to gain statistical significance of the size values, a sufficient number of particles need to be analyzed. The number of particles needed to be analyzed depends on, among other things, the particle size distribution and the desired measurement error. There are specific algorithms that have been produced to determine the number of particles needed to be analyzed for statistical significance (42).

***3.1.1 Electron Microscopy*** EM methods are often surface techniques, with the analysis confined to the top layer of a material, or in the case of TEM, to a thin section of a sample. In order to analyze a sample in bulk, additional sample preparation is necessary. Such preparation could include FIB with analysis of removed layers, or cryogenic methods, followed by fractionation and cross-section analysis.

*3.1.1.1 Transmission Electron Microscopy* TEM is typically run under a high vacuum, so samples must be dry when analyzed. Although environmental TEM set-ups do exist, they are not yet common. Samples must be between 100 and 200 nm in thickness. For nano-materials that are embedded within a matrix or are part of a larger material, this restriction means that the sample must be cut or thinned. Techniques such as microtome procedures or FIB preparation can produce these thin samples. If the nano-material is in a liquid suspension, then a portion of the sample may simply be dropped onto a TEM grid and allowed to dry. Although this is a facile sample preparation, artifacts from the drying process, including a misrepresentation of the aggregation/agglomeration state may occur with this sample preparation. Cryogenically freezing the sample, followed by analysis with cryo-TEM will prevent such artifacts, but requires a cryogenic attachment to the TEM.

Less electron-dense materials (such as carbon based materials) do not deflect the electron beam sufficiently for a signal and often cannot be imaged well with TEM. Heavy metal staining procedures, such as using phosphotungstic acid can be used to

increase contrast in these materials. If surface groups attached to a core nano-material are carbon based (such as PEG, protein, or peptide surface coatings), these coatings may not appear in the TEM analysis, and the particle size will appear smaller than it actually is. Coupling TEM with another type of analysis, such as DLS, will help alleviate some of these errors as well as confirm that the portion of material analyzed (a small sample size inherent with the method) is representative of the entire sample.

*3.1.1.2 Scanning Electron Microscopy* Although this method has less resolution than TEM, the operation of the instrument is often easier and more amenable to high throughput analysis. Traditional SEMs run under a full vacuum. Environmental chambers (ESEM) allow for both humidity and pressure in the sample chamber to be controlled, thus allowing for formulated materials to be directly analyzed without having to dry the sample. Alternatively, cryogenic methods may also be used with this technique, if a cryogenic stage is available.

SEM samples are typically coated with a conductive coating to prevent charging of the samples (which reduces the resolution of the instrument). While environmental set-ups do not require a conductive coating, resolution will often be decreased if it is not present on the sample due to charging effects. In addition, soft samples may be damaged in the electron beam, resulting in measurement errors.

When coupled to EDS, a comprehensive picture of the formulation may be determined. A special detector is needed to provide elemental mapping images. Although EDS does not have the lateral resolution of SEM to make precise sizing measurements, often individual features and contamination can still be determined with these image maps.

***3.1.2 SPM*** SPM methods can be extremely versatile, with the limitations of analysis being the measurement modes, the attachments to the instrument, and the probes used. As with TEM, however, only a small area is typically examined, which can lead to a non-representative portion of a sample being analyzed. In addition, artifacts from tip damage or the scanning method may complicate analysis. Lengthy and complex analysis as well as what can sometimes be a long optimization process to obtain images does not always lend SPM methods to be the first line of analysis. SPM methods, however, can provide otherwise inaccessible sample information, and analysis and imaging parameters, once optimized, can even become automated (43).

***3.1.3 Optical Microscopy*** The achievable resolution of most optical microscopes follow:

$$d = \frac{\lambda}{2A_n}$$

where $d$ is the resolution, $\lambda$ the wavelength of light, and $A_n$ is the numerical aperture of objective lens.

With the exception of some notable examples (refer to Section 2.2.3), most optical scopes will not be able to resolve nano-scale features. Depending on the conditions, materials, and microscope set-up, nano-scale features might be able to be detected, if not accurately sized. This type of information may be sufficient for some types of analyses. Most OM methods are either environmental or may be easily set-up to allow for a wide variety of environmental settings. In addition, OM may encompass a wide range of methods including fluorescent microscopy and all its related techniques and confocal microscopy. Confocal microscopy makes use of pinhole geometry to selectively focus on a single optical plane in a sample. By collecting stacks of optical planes in the $Z$ direction, a three-dimensional representation can be constructed with little or no modification made to the sample, allowing for in-depth, three-dimensional analysis.

## 3.2 Scattering Methods

Unlike microscopy evaluations, particles are not directly visualized when using scattering methods. Instead of measuring feature sizes, particle characteristics are often determined through well-established equations.

### *3.2.1 Light Scattering*

*3.2.1.1 Dynamic Light Scattering* If particles interacting with light are sufficiently smaller than the wavelength of the light they are scattering, the light is scattered by the particles uniformly in all directions. This type of scattering is called Rayleigh scattering (for larger particles roughly above 250 nm, Mie scattering dominates, which is angle dependent). Due to Brownian motion, the particles are in constant motion, and the intensity of the fluctuation of the laser light due to the particle interactions can provide the diffusion coefficient, which can then be placed into the Stokes–Einstein equation:

$$D = \frac{k_b T}{6\pi\eta a}$$

where $D$ is diffusion constant, $k_b$ the Boltzmann constant, $T$ the temperature (in Kelvin), $\eta$ the viscosity of the solvent, and $a$ is the hydrodynamic radius.

Historically, DLS systems could only measure particles in highly dilute systems. This limitation was mainly due to the multiple scattering of laser light in concentrated systems. Advances in the technology, including the development of backscatter detectors, have extended the utility of this method to highly concentrated samples. Thus, most samples do not have to be diluted prior to evaluation.

Typical instruments will provide different numerical weightings of the sample size distribution as well as the polydispersity of the sample.

The polydispersity index of a sample can be expressed as

$$\varepsilon = \sigma/R$$

where σ is the standard deviation and $R$ is the average particle radius.

Depending on the application, the polydispersity index may become important. In many cases, a low PDI (more uniform) product is desired, although there are notable cases where a wide distribution of particle sizes is preferable (e.g., inorganic sunscreens).

In addition, the weighting applied to the particle size distribution can greatly influence interpretation of results. The most common weightings are number, volume, and intensity and scale as $n = 0$, 3, 6, respectively, for the following equation:

$$P_n(R) = \frac{R^n N(R)}{\int_0^\infty R^n N(R)\mathrm{d}R}$$

where $P_n(R)$ is the distribution, $N(R)$ the particle number density, $N(R)\mathrm{d}R$ the number of particles per united volume with radii between $R$ and $R + \mathrm{d}R$ and $R$ is the particle radius.

Intensity weighting tends to favor larger particles, thus if the sample has aggregates, these may appear as a larger fraction of the sample. The type of weighting used in DLS analysis should always be reported with the value.

Deviations from a sphere, misrepresenting the viscosity and refractive index of the dispersing medium, or having a sample that absorbs the laser light, cause the most common errors in DLS. In addition, agglomeration and settling of the material can cause measurement inaccuracies. As the method does not directly visualize the particle, however, the entire particle diameter is measured, including organic coatings that may not be taken into account with EM methods.

DLS results are presented as a bulk average of the tested solution, with polydispersity typically being calculated along with the weighted average size. Separate size populations that have similar sizes can sometimes appear as one large, broad, size distribution rather than two separate, narrow, populations. Modifications may be performed to standard DLS technique to increase the resolution between separate size populations, including first separating the particles through methods such as FFF or chromatography.

*3.2.1.2 Static Light Scattering* SLS measures the scattered light intensity as a function of scattering angle and concentration. Most instruments use the theoretical foundation set out by Zimm to obtain particle information from the intensity of the scattered laser light.

$$\frac{Kc}{R(\theta)} = \frac{1}{M_\mathrm{w} P(\theta)} + 2A_2 c$$

where $K$ is a constant, $c$ the concentration (g/mL), $R(\theta)$ the excess Rayleigh ratio (the

excess of light scattered in a solution per unit solid angle θ over the pure solvent, divided by the incident light intensity), $M_w$ the weight average molar mass, $A_2$ the second virial coefficient, and $P(\theta)$ is the Form factor of the material.

Since the form factor depends on the structure of the material, the radius of gyration as well as other structural information about the sample may be obtained.

### 3.2.2 X-Ray Scattering

*3.2.2.1 X-Ray Diffraction* X-ray diffraction patterns obey Bragg's Law:

$$n\lambda = 2d\sin\theta$$

where $n$ is the integer of the order of reflection, λ the wavelength of incident X-ray beam, $d$ the distance between atomic layers in a crystal, and θ is the angle of incidence.

Whereas the diffraction pattern is indicative of a certain crystal state, analysis of the individual peaks may provide additional information about the material. The Scherrer equation uses peak broadening to obtain approximate crystallite diameters. Calculation errors may be as high as 20%, and will increase if other factors, such as crystallite strain, are not accounted for.

The Scherrer equation analyzes the peak broadening at FWHM of a material's characteristic peaks. Before sample peak broadening can be analyzed, instrumental broadening must first be accounted for:

$$\mathrm{FWHM}^2_{\mathrm{observed}} = \mathrm{FWHM}^2_{\mathrm{instrument}} + \mathrm{FWHM}^2_{\mathrm{size+strain}}$$

Then the Scherrer equation may be applied:

$$D \sim \frac{K\lambda}{\mathrm{FWHM}_{\mathrm{size}}\mathrm{Cos}\,\theta}$$

where $K$ is the a unit cell geometry dependent constant (typically between 0.85 and 0.99), $D$ the particle diameter (volume weighted), λ the wavelength of incident X-ray beam, θ is the angle of incidence, and FWHM is in radians.

If strain is present in the system ($\mathrm{FWHM}_{\mathrm{strain}} = \eta\mathrm{Tan}\,\theta$, $\eta$ = microstrain), then the equation may be written as

$$\mathrm{FWHM}_{\mathrm{size+strain}}\mathrm{Cos}\,\theta \sim \frac{K\lambda}{\mathrm{D}} + \eta\sin\theta$$

The particle diameter and microstrain can be determined from a linear regression of $\mathrm{FWHM}_{\mathrm{size\,+\,strain}}\mathrm{Cos}\,\theta$ versus $\mathrm{Sin}\,\theta$.

By analyzing peak shape, height, and relative intensities, other properties of the crystallite may be calculated including shape and thickness (44).

*3.2.2.2 Small Angle X-Ray Scattering* Structural information of both particles and particle–particle interactions is contained in the SAXS scattering pattern. It can be difficult to interpret SAXS data, especially for complex systems. For a monodisperse spherical system, the scattering intensity is related to the form factor by the following equation:

$$I(q) = \Delta\rho^2 N_\mathrm{p} V_\mathrm{p}^2 P(q)$$

where $I(q)$ is the scattering intensity, $P(q)$ the form factor, $q$ the scattering vector, $4\pi\sin(\theta/2)/\lambda$, $\Delta\rho$ the difference in the scattering length density between the solvent and the colloid, $N_\mathrm{p}$ the number of particles in a sample, and $V_\mathrm{p}$ is the volume of a particle.

Particle dimensions may be calculated by determining a good fit of the scattering pattern with established models or by using a Guinier plot to determine radius of gyration.

$$I(q) = \Delta\rho^2 N_p V_\mathrm{p}^2 \exp(-q^2 R^2/3)$$

where $I(q)$ is the intensity, $\Delta\rho$ the difference in the scattering length density between the solvent and the colloid, $N_\mathrm{p}$ the number of particles in a sample, $V_\mathrm{p}$ the volume of a particle, $R$ the radius of gyration, and $q$ is the scattering vector, $4\pi\sin(\theta/2)/\lambda$.

Plots may be constructed by plotting the natural log of intensity versus $q^2$, with a slope of $-R^2/3$. This is only valid for $q < 1/R$ (45).

***3.2.3 Neutron Scattering*** Neutrons are generally scattered by the nuclei of atoms, and the scattering is not dependent upon the atomic number. Atomics with close atomic numbers (including hydrogen) may be distinguished from each other. The basic scattering equations are similar to SAXS. Additional structural information may be gained by SANS by taking advantage of "contrast matching" out portions of the sample that are not being analyzed (e.g., the solvent) by using the correct ratio of $D_2O$ to $H_2O$.

***3.2.4 Ion Scattering*** Unlike the previous methods, ion scattering methods tend to be inherently destructive to a sample. Not only are sample layers etched away in the ion beam bombardment, but also ions can become imbedded within the sample. In addition, a full vacuum is usually required for sample analysis.

## 3.3 Separation

***3.3.1 Analytical Ultracentrifugation*** AUC can provide structural and conformational information about particles. Two main modes are used in AUC, sedimentation velocity and sedimentation equilibrium and they can be used sequentially to provide information about the particle and particle interactions in a particular system.

Sedimentation velocity experiments use high rotor speeds and detectors to determine the time course of sedimentation. Size, size distribution, and shape may be calculated with this method.

Sedimentation equilibrium experiments use lower rotor speeds and can probe interactions between particles and other moieties in the system such as binding information, strength of binding, and self-association as each non-interacting species is separated into a unique boundary. The size of the complexes may also be determined. As the chemical equilibrium of the system is not perturbed with this method, AUC can be an ideal technique for looking at formulated products (provided that they are in a form amenable to this type of procedure).

Three forces will act on the suspended particle when subject to a gravitational field, the gravitational force, the buoyant force, and the frictional force. These forces come into balance and provide the following equation:

$$\frac{M(1-\bar{\nu}\rho)}{Nf} = \frac{u}{\omega^2 r}$$

where $M$ is the molar weight of the solute (in g/mol), $N$ the Avogadro's number, $\nu$ the partial specific volume of thee particle (in mL), $\rho$ the density of the solvent (g/mL), $\omega$ is the angular velocity (in radians per second), $r$ the distance of the particle from the axis of rotation, $f$ is the frictional coefficient, $u$ the particle velocity, and $u/\omega 2r$ is the sedimentation coefficient, which depends on the particle properties.

## 3.4 Miscellaneous Methods

***3.4.1 Acoustic Attenuation Spectroscopy*** As no optics are involved, AAS can be used to analyze particle size and particle size distribution in dilute to opaque samples. Thus, opaque and/or concentrated particle suspension that cannot be dried and analyzed may be characterized using this technique.

***3.4.2 Raman Spectroscopy*** As with XRD, reducing the crystallite size to nano-dimensions results in broadening the peaks of the Raman spectrum. In addition to peak broadening, the following effects may also occur: small downshifts in transverse optic mode wave numbers, decreases in the separation of longitudinal and transverse modes, and changes in band intensity. These effects occur after decreasing the crystallite size below 1 μm. Different models may be applied to calculate crystallite size from the Raman spectrum, which are beyond the scope of this chapter (46,47).

***3.4.3 Differential Mobility Analysis*** Not all pharmaceuticals are solvent or solid based. DMA is a traditional instrument for measuring aerosol particles. Particles are injected into a clean gas and exposed to an electric field. Separation of the particles

is based upon electrical mobility, which is based upon the size and charge of the particle. By varying the applied field, a size distribution of the particles may be obtained.

## 3.5 Surface Methods

Zeta potential measurements are frequently used to evaluate a sample's surface charge and stability. The particle surface itself is not measured. A charged particle's surface is surrounded by a layer of ions called the "fixed layer." This layer is, in turn, surrounded by a more diffuse region comprised of oppositely charged counter ions. This area is termed the "diffuse double layer" and the entire region is electrically neutral. During a zeta potential measurement, a voltage is applied across the solution, and particles are attracted to the electrode of opposite polarity. The fixed layer goes along for the ride with the particle, as does some of the diffuse double layer. The boundary in the diffuse double layer between going with the particle and staying put is what is measured. Stability in a colloidal system occurs when columbic repulsions between the particles is greater than the Van der Waals forces. Typically a zeta potential of ±30 will result in a stable system and is measured by electrophoretic mobility by the Henry equation:

$$U_{\mathrm{E}} = 2\varepsilon\zeta f(\kappa a)/3\eta$$

where $U_{\mathrm{E}}$ is the electrophoretic mobility, $\varepsilon$ the dielectric constant, $\zeta$ the zeta potential, $\eta$ the viscosity, and $f(\kappa a)$ is the Henry's function, with "$1/\kappa$" the "thickness of the electric double layer (Debye length), and "$a$" the particle radius. In polar media, the maximum value of Henry's function is approximated as 1.5 (Smoluchowski approximation). The minimum value is 1 for non-polar media (Hückel approximation).

Usually, measurement systems are limited to currents lower than 200 μA (to avoid electrolytic reactions) and conductivities of 10 μS/cm (ionic strengths of 0.1–100 mM).

***3.5.1 Brunauer, Emmett, Teller (48)*** This method calculates the total and external surface area of a material based upon gas (usually nitrogen) adsorption. The measurements are based upon BET theory:

$$\frac{P}{V(P^{\mathrm{o}}-P)} = \frac{1}{V_{\mathrm{m}}c} + \frac{(c-1)}{V_{\mathrm{m}}c} \times \frac{P}{P^{\mathrm{o}}}$$

where $V$ is the volume of gas adsorbed at pressure $P$, $P^{\mathrm{o}}$ the saturated pressure of adsorbing gas at a designated temperature, $c$ the BET constant ($\exp[(H_{\mathrm{a}} - H_{\mathrm{l}})/\mathrm{RT}]$) with $H_{\mathrm{a}}$ the enthalpy of adsorption of the gas on the solid and $H_{\mathrm{l}}$ the heat produced on forming the second and succeeding layers, $T$ is temperature and $R$ is the gas constant, and $V_{\mathrm{m}}$ is the volume of gas adsorbed to form a monolayer on a surface.

The surface area of the material may be determined by plotting $P/V(P^{\mathrm{o}} - P)$ versus $P/P^{\mathrm{o}}$ (Intercept $= 1/V_{\mathrm{m}}c$, Slope $= (c-1)/V_{\mathrm{m}}c$), $V_{\mathrm{m}} = 1/(\text{slope} + \text{intercept})$.

$$\text{The number of molecules adsorbed onto the surface is} = V_m N / V_R$$

where $N$ is the Avogadro's number and $V_R$ is the molar volume of and ideal gas.

Finally, the surface area is the number of molecules multiplied by the area per molecule (0.162 $nm^2$ for $N_2$).

## 3.6 Pharmaceutical Specific Considerations

In some cases, it is necessary to distinguish between primary particles, aggregates, and agglomerates in a nano-therapeutic. In addition, if the particles are changing form during the life time of the material, the different entities will need to be characterized. In general, techniques such as EM and SPM can often identify the different moieties for visual analysis of the material. Scattering and separation methods, while able to size the different size populations, cannot always identify which specific type of moiety is being sized. Methods such as DLS, however, are able to evaluate how different conditions (time, temperature, pH, solvent, etc.) affect the aggregation/agglomerate on particles using real-time time course experiments.

***3.6.1 In vitro Considerations*** In many cases, small molecule techniques used to determine characteristics such as dissolution, bioavailability, and so on, may be sufficient to analyze nano-materials. However, nanoparticle interference has been observed for some assays, especially chemical-based toxicity assays (51). As this inference can produce strong false-positive or false-negative results, it is important to run the appropriate controls with each nanoparticle and assay to determine how, if at all, the nanoparticles are interfering with a particular test. Along with nanoparticle controls, effects of any buffers or chemicals upon the nano-pharmaceutical should be noted (e.g., strong agglomeration in a buffer).

***3.6.2 In vivo Considerations*** After adequate characterization of the nano-material, it will be necessary to characterize how it behaves in animal models. Nano-materials are large enough that they can be directly observed in tissues through various methods described earlier. This visualization may allow for a more direct determination of therapeutic localization (in contrast to small molecules where methods such as histopathology may provide indirect information about localization of the therapeutic). In the case of non-carbon based nano-systems, the identification of the materials may be straightforward. Carbon based materials present more challenges, although they too can be positively identified.

As with small molecule therapeutics, radio labeling can provide accurate localization of materials. In addition, if the material is not carbon based (e.g., a metal oxide) bulk measurement techniques, such as ICP-MS, XRF, or NAA, can provide tissue concentrations in the ppb range. If a nano-material incorporates imaging moieties, then imaging techniques such as MRI may be employed. Such versatility in localization and imaging is an advantage over radiolabeling. However, it can often

be difficult, especially with complex nano-materials, to determine the state of the particle once inside the tissue (e.g., are ligands still attached, etc.).

In terms of locating the particles within tissues, TEM is often the gold standard. When coupled with one of the compositional attachments such as EDS or EELS, particles may be accurately identified not only in tissues but also in sub-cellular locations. Depending on the size and concentration of the materials, other EM methods could be used. If the material is fluorescently labeled, or has a substantially different RI than the surrounding tissue, OM methods can also be used to locate particles. Both optical and SEM methods can be made environmental, but additional staining procedures may still be needed to enhance contrast.

For TEM procedures, tissues are usually fixed, embedded, stained, and sectioned prior to analysis (24). Cryo-techniques may also be used to avoid interactions with staining dyes or fixatives and the nano-materials. As most TEMs require high vacuum and thin samples, sample preparation is more involved. Introducing stains and fixative procedures may also introduce artifacts, and it is extremely important to ensure that these artifacts are not mistaken for the nano-material as they are often similar in size.

In a typical fixation procedure, tissues are first fixed with an aldehyde (glutaraldehyde or formaldehyde), post-fixed with osmium tetroxide, dehydrated through ethanol grades, infiltrated/embedded with resin, sectioned and stained (uranyl acetate and lead citrate are common) (24). As some stains or fixatives may interact with the nanoparticles, the staining protocol should be planned in its entirety prior to working with the tissues. This planning will prevent unfortunate accidents such as a fixative dissolving the nanoparticles.

## 4 METHODS

In all cases, proper sample handling and preparation should be observed for characterization procedures. In all cases, it is imperative to obtain a representative sample of the therapeutic, and methods are available to help with sample handling (25). Nano-materials can be formulated into a wide range of materials, and these materials may lend themselves to certain analytical methods. Table 4 highlights techniques that may be used for the characterization of different materials.

### 4.1 Deciding on Appropriate Instrumentation

When beginning the characterization procedure, several factors need to be taken into account including the material to be tested, the conditions desired for the measurement, and the type of data needed for the characterization. Some of these factors are more limiting than others and should be considered earlier in the decision process. Figure 4 shows an example of a decision tree used to determine appropriate nano-methods for sizing a powder sample under environmental conditions.

Many techniques are complementary to each other, while others can verify information from another method. For this reason, it is recommended that more

**TABLE 4 Methods for Various Nano-Materials**

| Powder | Liquid Dispersion | Aerosol | Gel/Complex | Solid | Environmental |
|---|---|---|---|---|---|
| **Size** | | | | | |
| EM[a] | DLS | DMA | ESEM | EM | ESEM |
| SPM | SLS | SMA | AFM | SPM | AFM |
| OM | PCSS | PHA-UCPC | OM | OM | NSOM |
| XRD | FFF | ATOMS | NSOM | FIM | OM |
| SAXS | AUC | | OM | XRD | XRD |
| SANS | CHDF | | SAXS | SAXS | SAXS |
| SIMS | XDC | | SANS | SANS | SANS |
| SNMS | HPLC-SEC | | RS | SIMS | RS |
| RS | AAS | | XRD | SNMS | |
| | PGSE-NMR | | | FIB | |
| | XRD | | | RS | |
| | SAXS | | | | |
| | SANS | | | | |
| | OM | | | | |
| | ESEM | | | | |
| | AFM | | | | |
| **Composition** | | | | | |
| EM | FFF | | ESEM | EM | ESEM |
| SPM | HPLC | | AFM | SPM | AFM |
| XRD | AAS | | NSOM | OM | NSOM |
| OM | NMR | | OM | FIM | OM |
| SIMS | XRD | | RS | XRD | XRD |
| SNMS | OM | | XRD | SAXS | SAXS |
| RS | AFM | | | SIMS | SANS |
| | ESEM | | | SNMS | RS |

a) If a general class, rather than a specific method is listed, then all sub-instruments and attachments in that class may be used.

than one technique initially be used to characterize nano-materials. Once a technique has been established and validated, then it may become acceptable for one method to be used, but initially, every reasonable effort should be made to test a material with multiple nano-methods.

## 4.2 Tools for Specific Materials

***4.2.1 Powders*** If the particle must remain in the powder state for analysis, most of the microscopy methods, including ones requiring a high vacuum and their various attachments can be utilized. The range of instrumentation increases if the particles are able to be suspended in a liquid medium.

***4.2.2 Liquid Dispersions*** Liquid dispersions are a very versatile class of materials to analyze. Not only are many of the characterization techniques solvent based but also if the material may be cryo-frozen or dried, then most of the characterization techniques are available for use.

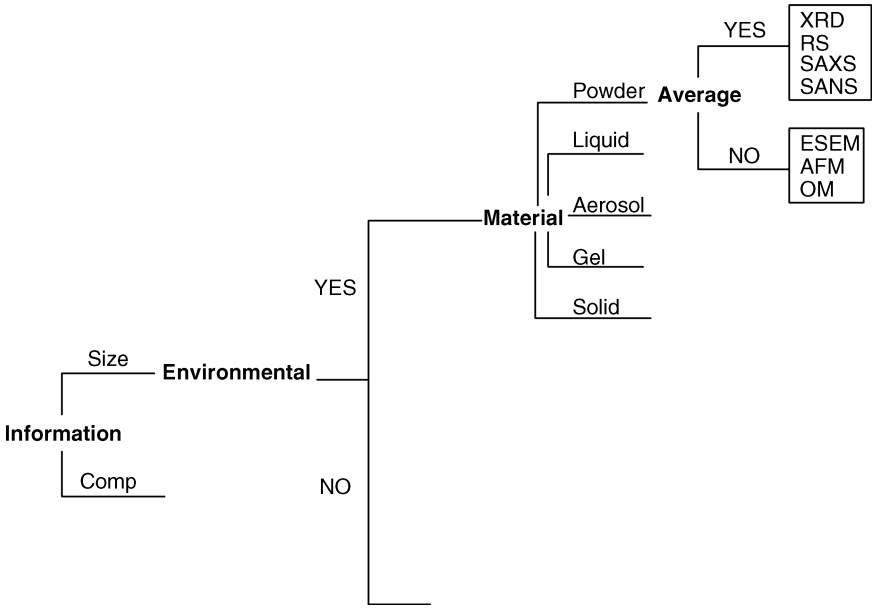


***FIGURE 4*** *Hypothetical example of a decision tree used to determine appropriate nano-methods. In this example, the goal is to obtain sizing information for powders under environmental conditions.*

***4.2.3 Aerosols*** If formulations are delivered in aerosol form ATOMS, PHA-UCPC, or DMA would be appropriate instruments to use for characterization. If the final material is placed onto sample grids, then techniques used to analyze powders may also be applied for both sizing and compositional information.

***4.2.4 Gel/Paste/Emulsion/Complex Formulation*** Although analytical nano-methods should be considered for each material, complex formulations require additional thought. If the formulation is viscous and will be fundamentally changed by dilution (such as many creams or lotions) then many of the liquid based methods will be unavailable. On the other hand, if drying the material will change the formulation and/or the nano-material and its interaction with the formulation, then all methods requiring a vacuum will be inappropriate as well. In some cases, only environmental techniques will be possible, and although this will limit the type and resolution of the data obtained, many environmental techniques are now providing data at increasingly high resolution. If the material may be dried or diluted, then solid and liquid dispersion techniques can be used.

***4.2.5 Solid Formulations*** In some cases, the nano-material will be embedded into another material (such as nanocomposites for tissue scaffolds or drug eluting stents). In these cases, the entire material may need to be evaluated and, as a result, many solvent-based techniques will be eliminated. In order to access some instruments, the material may need to be sectioned (TEM) or damaged (FIB, SIMS) for analysis.

### 4.3 Non-Perturbing/Environmental Methods

In some cases, the material must be tested in environmental conditions. This requirement is especially true if modifications to the sample during preparation or analysis (e.g., drying or dilution) is expected to change the nano-content or how it interacts with the rest of a formulation. Examples would include some nano-emulsions, or topical products such as creams that contain nano-materials. In such cases, the choice of instrumentation may be severely limited. Depending on the material tested, standard testing procedures may not be possible. If the material is analyzed with a destructive technique, the additional changes to the material should be noted. Although some instruments may be environmental for certain samples (e.g., liquid dispersions and light scattering techniques), truly environmental methods should be able to handle all materials.

## 5 CONCLUSIONS

Nano-therapeutics have been often viewed similarly to small molecule therapeutics. However, differences may arise during some of the characterization steps where nano-materials may interfere with common assays and tests, or non-traditional analytical instrumentation may be needed in order to fully characterize the material. In order to ensure that the material is correctly and completely identified, several analytical methods may be initially needed to evaluate the material. Finally, the final formulation should be tested to characterize any changes to the formulation or nano-material during the manufacturing process.

## REFERENCES

1. Dobson J. Magnetic nanoparticles for drug delivery. Drug Dev Res 2006; 67: 55–60.
2. Nishiyama N, Kataoka K. Current state, achievements, and future prospects of polymeric micelles as nanocarriers for drug and gene delivery. Pharm Ther 2006; 112: 630–648.
3. LaVan D, McGuire T, Langer R. Small-scale systems for in vivo drug delivery. Nat Biotechnol 2003; 21: 1184–1191.
4. Choy JH. Intercalative route to heterostructured nanohybrid. J Phys Chem Solids 2004; 65: 373–383.
5. Hu J, Johnston K, Williams R. Nanoparticle engineering processes for enhancing the dissolution rates of poorly water soluble drugs. Drug Dev Ind Pharm 2004; 30: 233–245.
6. Siegel R, Hu E, Roco M (STEC, Loyola College in Maryland), editors. *Nanostructure science and technology. A worldwide study*. National Science and Technology Council Committee on Technology, The interagency working group on nanoscience, engineering and technology. Final Report September 1999. International Technology Research Institute. World Technology Division, Loyola College in Maryland. Available from: http://itri.loyola.edu/nano/final/

7. Vogel V(University of Washington, Seattle and Swiss Federal Institute of Technology, Zurich), Baird B (Cornell University), editors. Nanobiotechnology. *Report of the National Nanotechnology Initiative Workshop*, Nanoscale Science, Engineering, and Technology Subcommittee; October 9–11, 2003. Available from: www.nano.gov/nni_nanobiotechnology_rpt.pdf
8. *Guidance for Industry, Content and format of investigational new drug applications for phase 1 studies of drugs, including well-characterized, therapeutic, biotechnology-derived products*. US Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research, and Center for Biologics Evaluation and Research; 1995. Available from: http://www.fda.gov/cder/guidance/phase1.pdf.
9. Warheit D. How meaningful are results of nanotoxicity studies in the absence of adequate material characterization. Toxicol Sci 2008; 101: 183–185.
10. Oberdorster G, Oberdorster E, Oberdorster J. Nanotoxicology: an emerging discipline evolving from studies of ultrafine particles. Environ Health Perspect 2005; 113: 823–839.
11. Gaur U, Sahoo S, De T, Ghosh P, Maitra A, Ghosh P. Biodistribution of fluoresceinated dextran using novel nanoparticles evading reticuloendothelial system. Int J Pharm 2000; 202: 1–10.
12. Liversidge G, Cundy K. Particle size reduction for improvement of oral bioavailability of hydrophobic drugs: I. Absolute oral bioavailability of nanocrystalline danazol in beagle dogs. Int J Pharm 1995; 125: 91–97.
13. *Guidance for Industry, INDs for Phase 2 and Phase 3 Studies, Chemistry, Manufacturing, and Controls Information*. US Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research; May 2003. Available from: http://www.fda.gov/cder/guidance./index.htm
14. *Draft Guidance, Guidance for Industry, Analytical Procedures and Methods Validation, Chemistry, Manufacturing, and Controls Documentation*. US Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research, and Center for Biologics Evaluation and Research; August 2000. Available from: http://www.fda.gov/cder/guidance./index.htm
15. Magenheim B, Benita S. Nanoparticle characterization: a comprehensive physiochemical approach. STP Pharma Sci 1991; 1: 221–241.
16. Burleson DJ, Driessen MD, Penn RL. On the characterization of environmental nanoparticles. J Environ Sci Health 2004; A39 (10): 2707–2753.
17. Hall JB, Dobrovolskaia MA, Patri AK, McNeil SE. Characterization of nanoparticles for therapeutics. Nanomedicine 2007; 2 (6): 789–803.
18. Tinke AP, Govoreanu R, Vanhoutte K. Particle size and shape characterization of nano-and submicron liquid dispersions. Am Pharm Rev 2006; September/October: 1–5.
19. Zuin S, Pojana G, Marcomini A. Effect-oriented physiochemical characterization of nanomaterials. In: Nancy A, Riviere M, Tran CL, editors. Nanotoxicology: characterization, dosing, and health effects. New York: Informa Healthcare; 2007. p 19–57.
20. Hartley PA, Parfitt GD. Agglomeration of powders containing submicron particles. In: International powder and bulk solids handling and processing. Int Powder 1984; 9: 171–181 C (Proceedings of the Technical Program).

21. Boyle J, Manas-Zloczower I, Feke DL. Influence of particle morphology and flow conditions on the dispersion behavior of fumed silica in silicone polymers. Part Part Syst Charact 2004; 21 (3): 205–212.
22. Clarke AR, Eberhardt CN. Microscopy techniques for materials science. Abington: Woodhead Publishing; 2002.
23. Holmberg K, editor. Handbook of applied surface and colloid chemistry. Vol. 1–2. West Sussex, England: JohnWiley and Sons; 2002.
24. Mathiowitz E, editor. Encyclopedia of controlled drug delivery. Vols 1 and 2. New York: John Wiley & Sons; 1999.
25. Fayed ME, Otten L, editors. Handbook of powder science and technology. 2nd ed. New York: Chapman and Hall; 1997.
26. Vaughan D, editor. Energy-dispersive X-ray microanalysis. An introduction. Wisconsin: Thermo NORAN; 2001.
27. Postek MT,Chair. *NNI Instrumentation and Metrology for Nanotechnology Report of the National Nanotechnology Initiative Workshop*, Gaithersburg, MD; January 27–29 2004. Available from: www.nano.gov/NNI_Instrumentation_Metrology_rpt.pdf.
28. Witt W, Geers H, Aberle L. Measurement of particle size and stability of nanoparticles in opaque suspensions and emulsion with photon cross correlation spectroscopy (PCCS). *PARTEC, Proceedings, P107*, Nurnburg; 2004.
29. Feldman LC, Mayer JW. Fundamentals of surface and thin film analysis. Englewood Cliffs: Prentice Hall; 1986.
30. Moffet RC, Prather KA. Extending ATOFMS measurements to include refractive index and density. Anal Chem 2005; 77: 6535–6541.
31. Valentini M, Vaccaro A, Rehor A, Napoli A, Hubbell JA, Tirelli N. Diffusion NMR spectroscopy for the characterization of the size and interactions of colloidal matter: the case of vesicles and nanoparticles. J Am Chem Soc 2004; 126: 2142–2147.
32. de la Mora JF, de Juan L, Eichler T, Rosell J. Differential mobility analysis of molecular ions and nanometer particles. Trends Analyt Chem 1998; 17 (6): 328–339.
33. Song Y, Jimenez V, McKinney C, Donkers R, Murray RW. Estimation of size for 1-2 nm nanoparticles using and HPLC electrochemical detector of double layer charging. Anal Chem 2003; 75: 5088–5096.
34. Price WS. Applications of pulsed gradient spin-echo NMR diffusion measurements to solution dynamics and organizations. Diffus Fundam 2008; 2: 112.1–112.19.
35. White WB. The structure of particles and the structure of crystals: information from vibrational spectroscopy. J Ceram Proc Res 2005; 6 (1): 1–9.
36. Bussche EV, Deene YD, Dubruel P, Vergote K, Schacht E, Wagter CD. The use of static light scattering for the structure analysis of radiosensitive polymer gels: a literature survey. J Phys Conf Ser 2004; 3: 180–183.
37. Ralston G. Introduction to analytical ultracentrifugation. Fullerton, CA: Beckman Instruments; 1993.
38. Zeta Potential: An introduction in 30 minutes. Zetasizer Nano series technical note. Malvern Instruments. www.malvern.com. accessed March 25, 2010.
39. Hammouda B. Probing nanoscale structures—the SANS toolbox. Gaithersburg, MD: National Institute of Standards and Technology; 2008. Available from http://www.ncnr.nist.gov/staff/hammouda/the_SANS_toolbox.pdf.

40. Vainrub A, Pustovyy O, Vodyanoy V. Resolution of 90 nm (lamda/5) in an optical transmission microscope with an annular condenser. Opt Lett 2006; 31: 2855–2857.
41. Klar TA, Engle E, Hell SW. Breaking Abbe's diffraction resolution limit in fluorescence microscopy with stimulated emission depletion beams of various shapes. Phys Rev E 2001; 64: 1–9.
42. Jillavenkatesa A, Dapkunas SJ. Lum LSH. NIST recommended practice guide special publication 960-1 particle size characterization. Gaithersburg, MD: National Institute of Standards and Technology; 2001. Available from: http://www.msel.nist.gov/practice-guides/SP960_1.pdf
43. Adams N, De Gans BJ, Kozodaev D, Sanchez C, Bastiaansen C, Broer D, Schubert U. High-throughput screening and optimization of photoembossed relief structures. J Combin Chem 2006; 8: 184–191.
44. Sherwood D, Emmanuel B. Computing shapes of nanocrystals from X-ray diffraction data. Crystal Growth Des 2006; 6 (6): 1415–1419.
45. Richardson R. Scattering and reflection techniques. In: Cosgrove T, editor. Colloid science—principles, methods, and applications. Oxford UK: Blackwell Publishing; 2005. p 228–254.
46. White WB. The structure of particles and the structure of crystals: information from vibrational spectroscopy. J Ceram Proc Res 2005; 6 (1): 1–9.
47. Colomban C. Raman analyses and "smart" imaging of nanophases and nanosized materials. Raman Spectrosc Spectrosc Eur 2003; 15 (6): 8–16.
48. Brunauer S, Emmett PH, Teller E. Adsorption of gases in multimolecular layers. J Am Chem Soc 1938; 60 (2): 309–319.
49. Yogamalar R, Srinivasan R, Vinu A, Ariga K, Bose AC. X-ray peak broadening analysis in ZnO nanoparticles. Solid State Commun 2009; 149: 1919–1923.
50. Fairhurst D, Dukhin AS. Acoustic attenuation spectroscopy: a new technique to characterize the stability and structure of semi-solid topical delivery systems for cosmetic and pharmaceutical applications. In: Wiechers J, editor. Science and applications of skin delivery. Illinois, USA: Allured Publishing Corporation; 2008. p. 185–205
51. Monteiro-Riviere NA, Inman AO, Zhang LW. Limitations and relative utility of screening assays to assess engineered nanoparticle toxicity in a human cell line. Toxicol Appl Pharmacol 2009; 234: 222–235

# *Quality by Design for the Pharmaceutical Industry*

**Richard N. McGrath[1] and Dennis K.J. Lin[2]**

[1]*Applied Statistics and Operations Research, Bowling Green State University, Bowling Green, Ohio*

[2]*Department of Statistics, Penn State University, University Park, Pennsylvania*

## 1 INTRODUCTION: WHY DESIGN AND RUN EXPERIMENTS IN THE PHARMACEUTICAL INDUSTRY?

While some discoveries and revolutionary improvements have come about by ad hoc methods, tinkering, or even accidents, a disciplined approach to discovering new products, developing them into salable form, and improving the processes by which they are produced is more efficient and effective. With such a disciplined approach, the quality of products and processes are embedded in their designs. This "quality by design" may be achieved through effective use of experimental methods. In this brief overview, we discuss many different methods that may be used to explore how changes to inputs (factors, explanatory or independent variables) affect outputs (responses or dependent variables). As opposed to passively observing changes and reacting to them, we describe methods in which we purposively change the values of factors in order to study their impact on one or more responses.

When attempting to optimize a product, process, or system, researchers often vary one factor at a time (OFAT), holding others fixed, in order to assess the impact of the factor on the response or responses. Design of experiments (DoE) methods allow

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing,*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

researchers to vary multiple factors at a time in a systematic fashion, in order to obtain more information for the same amount of resources. DoE is by no means new, having been applied in the agricultural sciences since the early 20th century. Varying amounts and types of fertilizers, as well as irrigation levels, allow one to study the impact of these factors on crop yield. Of course, there are many differences between performing experiments in a field and performing them in a laboratory or a manufacturing setting. For example, agricultural experiments may require months to obtain data while laboratory results may be available in minutes or hours. Agricultural experiments are often victims of much variation between fields, clouding differences between fertilizer and irrigation treatments. In a chemical experiment, labortory conditions and measurement error may be the main sources of background noise in the data.

Yet, many industries have incorporated DoE as an important tool for product/process improvement. Many of these firms, General Electric and Ford Motor Company, for example, use DoE as part of a "Six Sigma" effort for process improvement and product development. The types of experiments and their analyses that are applicable in the auto industry may differ somewhat from those in agriculture, which may differ from those in the pharmaceutical industry. We will focus on those of main interest in pharmaceuticals.

So where are these techniques used in the pharmaceutical industry? According to Laird (1), "Combinatorial chemistry and high-throughput screening have dramatically influenced the way compound discovery is carried out. . . ." He later states ". . . automated techniques lend themselves to a combination with statistically designed approaches (design of experiments, DOE) such as factorial designs, simplex optimisation, and response surface designs." Gooding et al. (2) applied DoE in conjunction with an automated synthesizer to optimize the synthesis of amides. Owen et al. (3) provide a step-by-step description of a case study involving the desilylation of a silyl ether to produce alcohol, a key step in the synthesis of a trinem antibiotic. Lutz et al. (4) descibe the use of DoE in high-thoughput screening. Tye (5) provides encouragement for DoE usage in the pharmaceutical industry and cites many example applications. Throughout this chapter, we will introduce other examples.

The rest of the chapter is organized as follows. The next section introduces basic experiments and their analyses where only one or two factors are studied. (We assume the reader has some familiarity with multiple linear regression and common statistical distributions including the normal and $t$. For more details on regression, see Kutner et al. (6).) Section 3 introduces factorial experiments and experiments useful for screening. Screening designs are useful for separating a large number of factors into those that are the most important and those that have less of an impact on the response. They are very useful when there are numerous variables resulting in many combinations to be tested as in, for example, drug discovery. Having reduced the number of factors of interest, in Section 4 we then discuss experiments that are useful for optimization. In these experimental designs such as response surface and simplex designs, a small number of factors are studied in order to find a combination of settings that will provide the most desirable response or responses.

Our discussion will focus on the design and analysis of an experiment once the factors have been identified. Due to page limitations, we will not discuss important issues such as what factors to include, how to determine the levels (settings) of a quantitative factor, and others. For a detailed discussion of these and other experimental matters, see Coleman and Montgomery (7).

## 2 BASIC EXPERIMENTAL DESIGNS AND ANALYSES

### 2.1 Single Factor Studies

Suppose all process variables are set to their nominal values in a chemical process and two batches of the same chemical are produced. If the yield of each is measured, will these yields be identical? Generally, the answer is no. If there is enough precision in the measuring system, the yields will not be identical, differing due to small changes in the process variables that were presumably fixed and other variables that were not controlled (ambient humidity and temperature for example). We can describe this with a statistical model as follows: Let $y_i$ be the response (yield in our example) for batch $i$. Then if we process multiple batches at the same settings of the process variables, we expect the yields to vary somewhat randomly around some average, say $\beta_0$. The variation around this average is a result of "noise" from minor deviations from nominal in the process variables and variation in other uncontrolled and possibly unobserved variables. This leads to the model in equation (1):

$$y_i = \beta_0 + \varepsilon_i \tag{1}$$

Here, $\beta_0$ is the mean yield and $\varepsilon_i$ is a random variable representing all of the noise or *error*. We assume that $\varepsilon_i$ has expected value 0 and variance $\sigma^2$. In order to perform statistical inference, we will also assume the $\varepsilon_i$ follows a normal distribution, or more succinctly, $\varepsilon_i \sim N(0, \sigma^2)$.

Now suppose the reaction temperature was 225°C for the first batch and was changed to 250°C for the second batch. We need to adapt model (1) to include the value and effect of reaction temperature. Let $x_{i1}$ be the reaction temperature for batch $i$ ($x_{11} = 225$ and $x_{21} = 250$). Now we may use the model in equation (2) below.

$$y_i = \beta_0 + \beta_1 x_{i1} + \varepsilon_i \tag{2}$$

Viewing this as a regression model, $\beta_1$ is the slope, which quantifies how much the response changes, on average, for a unit change in the factor. Note that if $\beta_1 = 0$, then the mean yield is the same regardless of reaction temperature and (2) reduces to (1). In this case, we refer to reaction temperature as being inactive. For $\beta_1 \neq 0$, reaction temperature is considered to be active, as increasing the temperature increases (decreases) the yield for $\beta_1 > 0 (\beta_1 < 0)$.

However, the parameters $\beta_0$, $\beta_1$, and $\sigma^2$ are unknown and must be estimated from the data from the experiment. If we only observe two batches (one at each of two temperatures), then we can estimate $\beta_0$ and $\beta_1$ but cannot estimate $\sigma^2$. Without knowing (or being able to estimate) $\sigma^2$, we cannot perform statistical inference to

decide if $\beta_1 \neq 0$ or not. So, in a simple experiment such as this, it is desirable to replicate the experiment by observing more than one batch at each level of the factor. The order that the experimental units are processed should be randomized. For example, if we were to observe three batches at each temperature, the order that the six batches are processed should be determined randomly.

## 2.2 Two-Factor Studies

***2.2.1 Unreplicated Two-Factor Studies*** In most designed experiments, multiple factors are systematically varied in order to assess their effect on one or more responses. Extending our chemical processing example, suppose we study catalyst amount as a factor in addition to reaction temperature. Temperature is studied at two levels as before (225 and 250) and catalyst at levels of "low" and "high." When each factor has only two levels, it is convenient to label them $-1$ and $+1$, or simply $-$ and $+$. Note that there are four temperature–catalyst combinations or *treatments*: (225, low), (250, low), (225, high), and (250, high) or $(-, -)$, $(+, -)$, $(-, +)$, and $(+, +)$. A single batch was produced for each treatment. In an unreplicated experiment such as this, each treatment is only observed once. Table 1 shows the levels of temperature ($x_1$) and catalyst ($x_2$) as well as the yield ($y$). We will discuss the other columns shortly.

The four responses are $y_{--} = 73$, $y_{+-} = 80$, $y_{-+} = 77$, and $y_{++} = 77$. Based on these four observations, can we quantify the effect of temperature ($x_1$) on the response? A simple and effective way to do so is to compare the average response at high temperature ($x_1 = 1$) to the average response at low temperature ($x_1 = -1$). The quantities we need are $(y_{+-} + y_{++})/2 = (80+77)/2 = 78.5$ and $(y_{--} + y_{-+})/2 = (73+77)/2 = 75$. So our estimate of the effect of temperature is $78.5-75 = 3.5$, meaning we estimate that the yield is 3.5% greater at the higher temperature. With our $+1$ and $-1$ coding, the "high" level of a factor is two units above the "low" level of the factor. So a two-unit increase in $x_1$ is associated with an increase of 3.5 in $y$, or equivalently, a one unit increase in $x_1$ is associated with an increase of 1.75 in $y$. This, of course, is the estimated slope of $x_1$. In a similar manner, the effect of catalyst is $(y_{-+} + y_{++})/2 - (y_{--} + y_{+-})/2 = (77+77)/2 - (73+80)/2 = 0.5$ and the estimated slope is 0.25.

***2.2.2 Two-Factor Model*** For an experiment with two factors ($x_1$ and $x_2$), our model may be written as

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_{12} x_{i1} x_{i2} + \varepsilon_i, \qquad \varepsilon_i \sim N(0, \sigma^2) \tag{3}$$

**TABLE 1 Unreplicated Two-Factor Example**

| $x_0$ | $x_1$ | $x_2$ | $x_1x_2$ | $y$ |
|---|---|---|---|---|
| 1 | −1 | −1 | 1 | 73 |
| 1 | 1 | −1 | −1 | 80 |
| 1 | −1 | 1 | −1 | 77 |
| 1 | 1 | 1 | 1 | 77 |

$\beta_0$ is a constant representing the mean response and may be estimated by $\hat{\beta}_0 = (y_{--} + y_{+-} + y_{-+} + y_{++})/4 = (73 + 80 + 77 + 77)/4 = 76.75$. $\beta_1$ and $\beta_2$ are the slopes quantifying the increase in the response for unit changes in factors $x_1$ and $x_2$, respectively. Above we estimated these quantities with $\hat{\beta}_1 = 1.75$ and $\hat{\beta}_2 = 0.25$.

The product $x_{i1}x_{i2}$ (which, notationally, could be labeled $x_{i3}$) represents the interaction between factors $x_1$ and $x_2$, and $\beta_{12}$ represents the magnitude of this interaction. Two factors are said to interact if the effect of one factor depends on the level of the other. Note that if we multiply the elements of $x_1$ by those of $x_2$, we get $x_1x_2 = (+, -, -, +)'$. So as we did for the effects of $x_1$ and $x_2$, we estimate the interaction effect as $(y_{--} + y_{++})/2 - (y_{+-} + y_{-+})/2 = (73 + 77)/2 - (80 + 77)/2 = -3.5$ and $\hat{\beta}_{12} = -1.75$. One way to interpret this quantity is the effect of $x_1$ when $x_2 = +1$ minus the effect of $x_1$ when $x_2 = -1$. So here, the effect of increasing temperature by one unit at the high level of catalyst is 1.75 less than at the low level of catalyst. Putting this together we can get a model for predicting the yield, $\hat{y}_i = 76.75 + 1.75x_{i1} + 0.25x_{i2} - 1.75x_{i1}x_{i2}$. However, without replication we cannot estimate $\sigma^2$ and, therefore, cannot test hypotheses. In other words, we cannot judge whether any of the estimated coefficients are significantly different from 0.

Eqs. (1–3) may all be written in matrix form as

$$\mathbf{y} = \mathbf{X}\beta + \varepsilon, \qquad \mathrm{Var}(\varepsilon) = \sigma^2\mathbf{I} \tag{4}$$

The $n \times p$ matrix $\mathbf{X}$ consists of columns $\mathbf{x}_0, \ldots, \mathbf{x}_{p-1}$. $\mathbf{x}_0$ is a column of $n$ 1s representing the intercept. $\beta = (\beta_0, \ldots, \beta_{p-1})$ is a $p$-dimensional vector of unknown parameters that need to be estimated. To estimate these $p$ quantities, we need $p$ degrees of freedom and thus, at least, $p$ observations. With a total of $n$ observations, we have $n-p$ degrees of freedom left to estimate $\sigma^2$. The first four columns of Table 1 show the $\mathbf{X}$ matrix for our unreplicated example. Note that here $n = p = 4$, so $n-p = 0$, and we do not have any degrees of freedom to estimate $\sigma^2$ as previously stated.

Using this matrix notation, formulas can be given for many quantities of interest. More details can be found in any regression text (see, for example, Kutner et al. (6)). We list the ones that will be most useful for our purposes. As statistical software is readily available for this, we will then focus on interpreting the output from software.

The ordinary least squares estimator of $\beta$ is

$$\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \tag{5}$$

and the variance–covariance matrix of $\hat{\beta}$ is

$$\mathrm{Var}(\hat{\beta}) = \mathbf{V} = (\mathbf{X}'\mathbf{X})^{-1}\sigma^2 \tag{6}$$

The fitted values of the response are

$$\hat{\mathbf{y}} = \mathbf{X}\hat{\beta} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \tag{7}$$

and we can estimate $\sigma^2$ with

$$s^2 = (\mathbf{y}-\hat{\mathbf{y}})'(\mathbf{y}-\hat{\mathbf{y}})/(n-p) \tag{8}$$

Putting it all together, we can estimate $\mathbf{V}$ with

$$\widehat{\mathrm{Var}(\hat{\beta})} = \hat{\mathbf{V}} = (\mathbf{X}'\mathbf{X}')^{-1}s^2 \tag{9}$$

The $j$th diagonal term of $\hat{\mathbf{V}}$, denoted $v_{jj}, j = 0, \ldots, p-1$, is the estimate of $\mathrm{Var}(\hat{\beta}_j)$, and the square root of this quantity is known as the standard error of $\hat{\beta}_j$, or $\sqrt{v_{jj}} = \mathrm{s.e.}(\hat{\beta}_j)$. We will see $\sqrt{v_{jj}} = \mathrm{s.e.}(\hat{\beta}_j)$ in much of our computer output.

***2.2.3 Replicated Two-Factor Studies*** Once again extending our example, now suppose that 12 batches were processed as follows. Each of the four treatments were applied to three different batches (resulting in $r = 3$ replications) with the order of experimentation randomized. In Table 2, column $i$ shows the order of experimentation, or batch number, and the $x_1$ and $x_2$ columns define the experimental design. Collectively, the columns $x_0$, $x_1$, $x_2$, and $x_1x_2$ form the $\mathbf{X}$ matrix. Note we have $n-p = 12-4 = 8$, so we have eight degrees of freedom to estimate $\sigma^2$ using Eq. (8). We (or more likely a software package) can then test to see if the coefficients differ significantly from 0.

Running least squares regression in any statistics package will result in output similar to that given in the Table 3. The column labled "estimate" gives the estimated slopes or coefficients ($\hat{\beta}_j$s). For our $-1$ and $+1$ coding, the slope is half of the difference between the mean response at the $+$ and $-$ levels of a column. From Table 2 we can calculate the mean response for each of the four unique treatments, namely, $\bar{y}_{--} = 69$, $\bar{y}_{+-} = 79$, $\bar{y}_{-+} = 74$, and $\bar{y}_{++} = 81$, where the first subscript refers to the level of $x_1$ and the second to that of $x_2$. We may then perform calculations as we did in the unreplicated case using means in place of the individual observations. So $\hat{\beta}_1$ is $\frac{1}{2}\left(\frac{79+81}{2} - \frac{69+74}{2}\right) = 4.25$, and the mean yield is $2 \times 4.25 = 8.5\%$ greater at

**TABLE 2 Replicated Two-Factor Example**

| $i$ | $x_0$ | $x_1$ | $x_2$ | $x_1x_2$ | Yield (y) |
|---|---|---|---|---|---|
| 1 | + | − | − | + | 73 |
| 2 | + | + | − | − | 80 |
| 3 | + | − | + | − | 77 |
| 4 | + | + | + | + | 77 |
| 5 | + | − | + | − | 70 |
| 6 | + | − | + | − | 75 |
| 7 | + | + | − | − | 81 |
| 8 | + | − | − | + | 69 |
| 9 | + | − | − | + | 65 |
| 10 | + | + | + | + | 82 |
| 11 | + | + | + | + | 84 |
| 12 | + | + | − | − | 76 |

**TABLE 3 Two-Factor Experiment Regression Output**

| | Estimate | Std. Error | *t*-value | Pr(> \|t\|) |
|---|---|---|---|---|
| (Intercept) | 75.750 | 1.010 | 74.973 | 1.12e-12 |
| $x_1$ | 4.250 | 1.010 | 4.206 | 0.00297 |
| $x_2$ | 1.750 | 1.010 | 1.732 | 0.12150 |
| $x_1x_2$ | −0.750 | 1.010 | −0.742 | 0.47913 |

the higher temperature than at the low. The other columns in the display are used to perform inference about the coefficients. Recall that a factor (or interaction) is active if its slope (coefficient) differs from 0. The output allows us to perform the hypothesis test $H_0 : \beta_j = 0$, $H_1 : \beta_j \neq 0$. Under the assumptions of our model, $t_j = \text{estimate}/\text{std. error} = \hat{\beta}_j/\text{s.e.}(\hat{\beta}_j)$, called a test statistic, follows a $t$ distribution with $n-p$ degrees of freedom.

The $p$-value for a hypothesis test is the probability of observing the test statistic we did ($t_j$), or something more in favor of $H_1$, assuming that $H_0$ is true. The larger the $t_j$-value is, the smaller the $p$-value is and the more evidence there is against $H_0$. So for us, a small $p$-value (0.05 is often used to define small) suggests that $H_1 : \beta_j \neq 0$ is true, i.e., that the factor has an effect on the mean response. A common cutoff value for deciding to reject $H_0$ in favor of $H_1$ is 0.05.

From Table 3 we can see the intercept and $\beta_1$ (the $x_1$ term) differ significantly from 0. So we can conclude that the higher reaction temperature leads to a higher yield, but we do not have sufficient evidence to state that the yield is affected by changing the amount of catalyst. It is important to note that this does not mean that the amount of catalyst has no effect on the yield. It does mean that based on the levels studied for this factor, and the amount of data collected, we do not have enough evidence to see a statistically significant difference from 0 in $\beta_2$.

Similarly, the interaction between temperature and catalyst is not significant. The slope for the interaction is $-0.75$ equivalent to the interaction effect being $-1.50$. The interaction effect is the difference in the temperature effect at the high level of catalyst minus the temperature effect at the low level of catalyst. If there is no dfference between these two temperature effects, then the temperature effect does not depend on the catalyst level, and there is no interaction. This effect is actually $(1/2) \times [(81-74)-(79-69)] = -1.5$.

Figure 1, known as an interaction plot, shows the effect of temperature at the low level of catalyst separately from the effect of temperature at the high level of catalyst. Theoretically, the two lines in the plot will be parallel if there is no interaction. Substantial departures from parallelism indicate an interaction between the factors. We can see that in this case, the lines are roughly parallel, agreeing with our previous conclusion of no interaction.

## 2.3 Multi-Factor Studies

The example just discussed was quite simple in that only two factors were studied, each with two levels. Increasing the number of factors or the number of levels of factors increases the number of treatments. Continuing the chemical example,

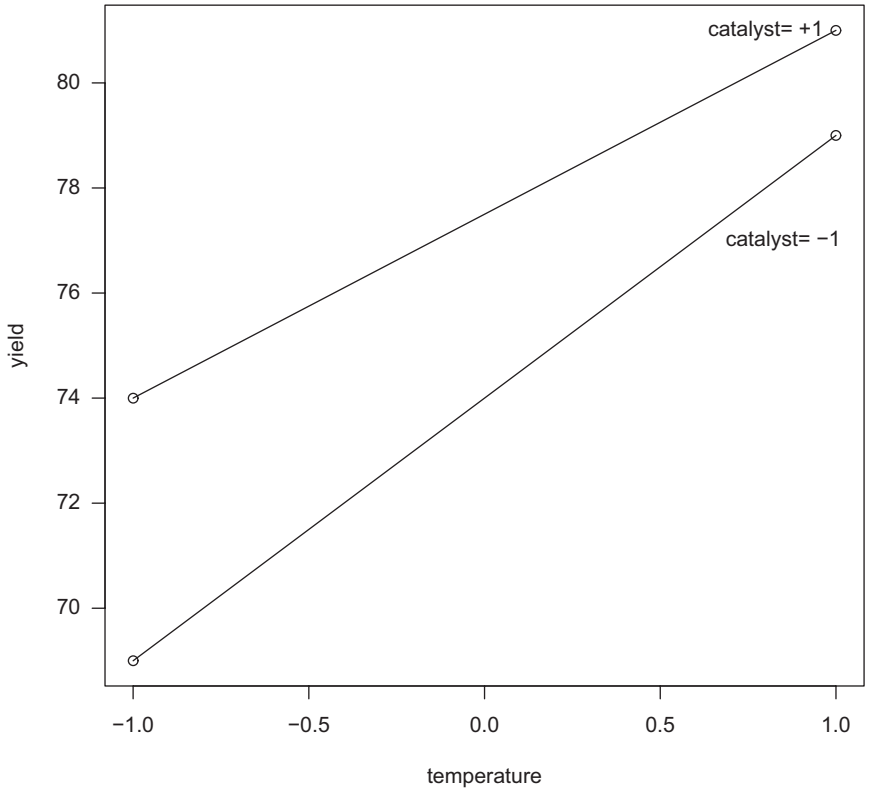


***FIGURE 1*** *Interaction plot – temperature vs. catalyst.*

suppose three temperatures, two catalyst amounts, and four reaction times were to be studied. If all combinations are included in the experiment, there will be $3 \times 2 \times 4 = 24$ unique treatments requiring 24 batches even without any replication. When all possible treatments are studied as just discussed, we have a *factorial design*.

The number of factors to include and the number of levels of each depends on the goal of the study. There may be many goals when performing an experiment including:

- sorting factors into those that affect the response (active) and those that do not appear to affect the response (inactive);
- assessing the magnitude of the effect of factors on the response;
- determining which factors interact and assessing the magnitude of such interactions;
- finding settings of the factors that produce desirable responses.

Often, a series of experiments are used, each with different goals. An initial study may be used to do screening, i.e., find the factors that appear to be the most important.

These screening experiments generally have several to many factors, but each is studied at only two or perhaps three levels. Then, after identifying a small number of key factors, additional experiments may be performed in order to better quantify the effect of each factor or to find optimal settings. In the next section we discuss screening designs and then introduce designs for optimization in the section thereafter.

## 3 DESIGNS FOR SCREENING

Two factors at two levels each leads to $2 \times 2 = 2^2 = 4$ treatments. If each factor had three levels, we would have $3^2 = 9$ treatments, while three factors each at three levels leads to $3^3 = 27$ treatments. So increasing the number of factors or the number of levels of the factors rapidly increases the number of treatments. If the intent of the experiment is to separate the wheat from the chaff (find the important few factors among the less important many), then we may wish to study a relatively large number of factors, each with only two or three levels. In this section we will briefly study designs where all factors are studied at two levels. For more discussion of these designs and their analysis, see Box et al. (8), Wu and Hamada (9), and Montgomery (10).

### 3.1 Factorial Designs

Factorial designs consist of all combinations of factors; i.e., each treatment is observed. If each of $k$ factors has only two levels, then there are $2^k$ treatments. In Table 1 we have a $2^3$ tablet experiment from Khanvilkar et al. (11) in which three factors (A = hardness, B = viscosity, and C = polymer composition) were studied and multiple responses were measured. (To show another common notation, we have used A, B, and C instead of $x_1$, $x_2$, and $x_3$ to label the factors.) We focus on $t_{\text{lag}}$, which is the time (min) taken by the matrix tablet to get hydrated and achieve a state of quasi-equilibrium before erosion and the advance of solvent front through the matrix occur. The columns AB, AC, and BC represent interactions among two factors, and ABC represents the three-factor interaction (Table 4). (A three-factor interaction occurs when a two-factor interaction effect depends on the level of a third factor.)

**TABLE 4 Unreplicated $2^3$ Experiment from Khanvilkar et al. (11)**

| A | B | C | AB | AC | BC | ABC | $t_{\text{lag}}$ |
|---|---|---|---|---|---|---|---|
| − | − | − | + | + | + | − | 10.08 |
| + | − | − | − | − | + | + | 10.84 |
| − | + | − | − | + | − | + | 10.65 |
| + | + | − | + | − | − | − | 10.92 |
| − | − | + | + | − | − | + | 5.14 |
| + | − | + | − | + | − | − | 6.06 |
| − | + | + | − | − | + | − | 5.46 |
| + | + | + | + | + | + | + | 5.65 |

**TABLE 5 Output from Unreplicated $2^3$ Experiment from Khanvilkar et al. (11)**

| | Estimate | Std. Error | *t*-value | Pr(> \|*t*\|) |
|---|---|---|---|---|
| (Intercept) | 8.1000 | 0.0300 | 270.000 | 0.00236 |
| A | 0.2675 | 0.0300 | 8.917 | 0.07110 |
| B | 0.0700 | 0.0300 | 2.333 | 0.25776 |
| C | −2.5225 | 0.0300 | −84.083 | 0.00757 |
| AB | −0.1525 | 0.0300 | −5.083 | 0.12366 |
| AC | 0.0100 | 0.0300 | 0.333 | 0.79517 |
| BC | −0.0925 | 0.0300 | −3.083 | 0.19966 |

Note that in this experiment, only one observation is recorded for each treatment. Recall that this is known as an *unreplicated design*. The model may be written as

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_{12} x_{i1} x_{i2} + \beta_{13} x_{i1} x_{i3} + \beta_{23} x_{i2} x_{i3} +$$

$$\beta_{123} x_{i1} x_{i2} x_{i3} + \varepsilon_i, \quad \varepsilon_i \sim N(0, \sigma^2)$$

There are eight observations ($n = 8$) and eight $\beta$s ($p = 8$), each requiring one degree of freedom for estimation. Therefore, all degrees of freedom are used for parameter estimation ($n-p = 0$). In this case, $\hat{y}_i = y_i$ and $s^2 = 0$, so we cannot estimate $\sigma^2$ and cannot perform hypothesis tests.

One way to tackle this issue, the one used in Khanvilkar et al., is to assume one or more of the higher order interactions are inactive and leave them out of the model. Doing so pools that term or terms into error, providing at least one degree of freedom for testing. Khanvilkar et al. assumed the three-factor interaction was inactive ($\beta_{123} = 0$) and performed regression resulting in the output in Table 5. Clearly, factor C is significant, and no other terms are with the possible exception of A.

An alternative analysis method is to use the fact that if no effects are active, the estimates (ignoring the intercept) all follow a common normal distribution with mean 0 and unknown but constant variance. As such, they should behave as a random sample from a common normal distribution. Active effects have greater expected absolute values and should stick out from the noise of the inactive effects. Daniel (12) suggested using a half-normal plot of the effects or regression coefficients. Those that deviate substantially above a linear pattern emanating from 0 through the first few points are deemed to be active. Figure 2 shows a half-normal plot of the seven estimates from the tablet example. $\hat{\beta}_C$ clearly deviates from the roughly linear pattern of the others. An advantage of this approach is that all effects of interest can be assessed. A disadvantage is that it assumes effect sparcity, that only a few of the effects are actually active, and that the rest are trivial.

## 3.2 Fractional Factorial Designs

The previous example was an unreplicated $2^3$ design. By not replicating, experimental resources were saved. However, situations occur where resources are quite limited and even one observation for every treatment combination is not feasible. We saw that

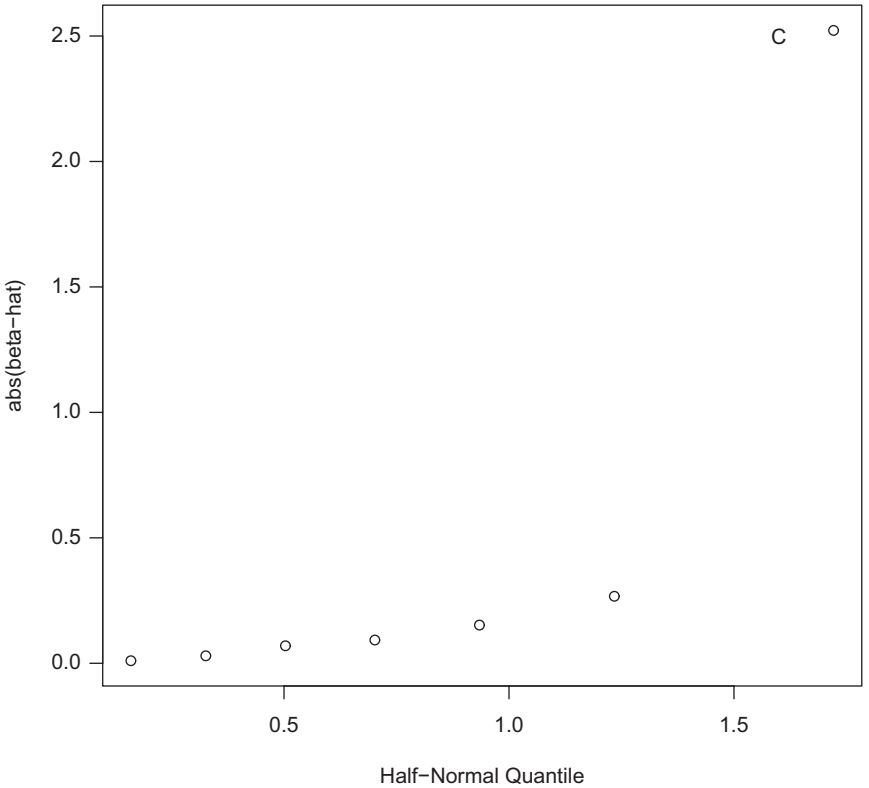


***FIGURE 2*** *Half-normal plot – Khanvilkar et al., unreplicated $2^3$ with response $= t_{lag}$.*

from an analysis perspective, we can assume one or more effects are inactive and essentially use these to form an error term. But if we are willing to entertain the idea that some effects are inactive, then we may not need to observe all treatments. Consider a study reported by Menon et al. (13). Five factors (percent povidone ($P$), liquid/dry binder addition for povidone ($B$), atomization pressure ($A$), spray rate ($S$), and inlet air temperature ($I$)) were studied with responses of $y_1 =$ particle size (microns) and $y_2 =$ crushing strength (kP).

Suppose only the first four factors were to be studied in a $2^4$ design. Then once again using the $+1, -1$ coding, the design is displayed in columns $1-5$ in Table 6. Columns 5–15 represent the six two-factor interactions, the four three-factor interactions, and the four-factor interaction for these factors. If we want to study a fifth factor without additional runs, we assign the fifth factor to one of these columns, equating it with that effect. Column 15 is the PBA$S$ interaction, and we see that $I$ has been assigned to that column. So now column 15 represents a pair of effects, PBA$S$ and $I$. If we multiply columns 1 and 15 ($P$ and $I$) together, we get column 14 ($PI$), which is also BA$S$. In other words, PI $=$ BA$S$. Each column represents a pair of effects resulting in the pattern shown in Table 7. The effects in each pair are said to be confounded.

**TABLE 6 $2^{5-1}$ Example**

| $i/j$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | $y_1$ | $y_2$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | *P* | *B* | *A* | *S* | *PB* | *PA* | *PS* | *BA* | *BS* | *AS* | *SI* | *AI* | *BI* | *PI* | I | | |
| 1 | 1 | −1 | −1 | −1 | −1 | 1 | 1 | 1 | 1 | 1 | 1 | −1 | −1 | −1 | −1 | 1 | 189.00 | 11.28 |
| 2 | 1 | 1 | −1 | −1 | −1 | −1 | −1 | −1 | 1 | 1 | 1 | 1 | 1 | 1 | −1 | −1 | 56.00 | 3.00 |
| 3 | 1 | −1 | 1 | −1 | −1 | −1 | 1 | 1 | −1 | −1 | 1 | 1 | 1 | −1 | 1 | −1 | 93.75 | 5.70 |
| 4 | 1 | 1 | 1 | −1 | −1 | 1 | −1 | −1 | −1 | −1 | 1 | −1 | −1 | 1 | 1 | 1 | 80.00 | 5.93 |
| 5 | 1 | −1 | −1 | 1 | −1 | 1 | −1 | 1 | −1 | 1 | −1 | 1 | −1 | 1 | 1 | −1 | 212.50 | 10.02 |
| 6 | 1 | 1 | −1 | 1 | −1 | −1 | 1 | −1 | −1 | 1 | −1 | −1 | 1 | −1 | 1 | 1 | 212.50 | 11.34 |
| 7 | 1 | −1 | 1 | 1 | −1 | −1 | −1 | 1 | 1 | −1 | −1 | −1 | 1 | 1 | −1 | 1 | 76.00 | 6.50 |
| 8 | 1 | 1 | 1 | 1 | −1 | 1 | 1 | −1 | 1 | −1 | −1 | 1 | −1 | −1 | −1 | −1 | 125.00 | 5.20 |
| 9 | 1 | −1 | −1 | −1 | 1 | 1 | 1 | −1 | 1 | −1 | −1 | −1 | 1 | 1 | 1 | −1 | 351.25 | 7.73 |
| 10 | 1 | 1 | −1 | −1 | 1 | −1 | −1 | 1 | 1 | −1 | −1 | 1 | −1 | −1 | 1 | 1 | 534.00 | 11.89 |
| 11 | 1 | −1 | 1 | −1 | 1 | −1 | 1 | −1 | −1 | 1 | −1 | 1 | −1 | 1 | −1 | 1 | 275.00 | 11.60 |
| 12 | 1 | 1 | 1 | −1 | 1 | 1 | −1 | 1 | −1 | 1 | −1 | −1 | 1 | −1 | −1 | −1 | 219.00 | 5.90 |
| 13 | 1 | −1 | −1 | 1 | 1 | 1 | −1 | −1 | −1 | −1 | 1 | 1 | 1 | −1 | −1 | 1 | 153.50 | 9.52 |
| 14 | 1 | 1 | −1 | 1 | 1 | −1 | 1 | 1 | −1 | −1 | 1 | −1 | −1 | 1 | −1 | −1 | 752.00 | 10.30 |
| 15 | 1 | −1 | 1 | 1 | 1 | −1 | −1 | −1 | 1 | 1 | 1 | −1 | −1 | −1 | 1 | −1 | 374.00 | 4.87 |
| 16 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 478.00 | 8.61 |

**TABLE 7 Confounding Pattern for the $2^{5-1}$ Example**

| Intercept = PBASI | PS = BAI |
|---|---|
| P = BASI | PI = BAS |
| B = PASI | BA = PSI |
| A = PBSI | BS = PAI |
| S = PBAI | BI = PAS |
| I = PBAS | AS = PBI |
| PB = ASI | AI = BAS |
| PA = BSI | SI = PBA |

So in Table 6, each column represents a pair of effects, and we have labeled each column with its lower order effect from the pairings listed above. The coefficients associated with $\log_{10}(y_1)$ and $y_2$ are listed in Table 8. Figure 3 is a half-normal plot, based on the $\log_{10}$ of response $y_1$ = particle size. From this plot it is clear that $S$ has the largest impact on the response. However, Menon et al. used the procedure of Lenth (14) as adapted by Haaland and O'Connell (15). First, we calculate $s_0 = 1.5 \times \text{median}(|\beta_j|)$. Next, a pseudo-standard error is calculated as $\text{PSE} = 1.5 \times \text{median}_{|\beta_j|<2.5s_0}|\beta_j|)$. Then each effect is tested by comparing $|\beta_j|$ to $t_{(1-\alpha/2,(p-1)/3)} \times \text{PSE}$. (Haaland and O'Connell suggest minor adjustments to these quantities.) For $\log_{10}(y_1)$, $s_0 = 0.05119085$, $\text{PSE} = 0.05011927$, $t_{(.975,15/3)} = 2.570582$, and the critical value is $t_{(.975,15/3)} \times \text{PSE} = 0.1288357$. Thus we see that $S$ is the only significant effect. (We note that Menon et al. showed a different (apparently incorrect) value for PSE and found many effects to be significant.)

Using the coefficients for $y_2$ = strength, Lenth's procedure results in PSE = 1.028438 and a critical value of 2.643683. Based on this procedure, none of the effects are significant. Half-normal plotting and Lenth's procedure are just two of many methods that have been developed to analyze unreplicated fractional factorial designs. See Hamada and Balakrishnan (16) for an extensive discussion of other methods that may be used to analyze unreplicated two-level experiments.

Of the $2^4$ treatments in the factorial design of this example, only $1/2 = 2^{-1}$ or $(2^{-1})(2^4) = 2^{4-1}$ were studied. This is a half-fraction of the full factorial design. In general, if only a $2^{-f}$ fraction of a $2^k$ factorial is run, the design is often called a $2^{k-f}$. Selection of the specific fraction to use is very important and is discussed in texts such as Montgomery (10).

**TABLE 8 Regression Coefficients for the $2^{5-1}$ Example**

| Term | P | B | A | S | PB | PA | PD | BA |
|---|---|---|---|---|---|---|---|---|
| $y_1$ estimate | 0.031 | −0.071 | 0.062 | 0.240 | −0.012 | 0.095 | 0.079 | 0.010 |
| $y_2$ estimate | −0.316 | −1.298 | 0.208 | 0.716 | −0.063 | 0.883 | 0.688 | −0.702 |
| Term | BS | AS | SI | AI | BI | PI | I | |
| $y_1$ estimate | 0.033 | −0.030 | −0.034 | −0.096 | −0.006 | 0.076 | −0.003 | |
| $y_2$ estimate | 0.241 | −0.689 | 0.106 | −0.800 | −0.126 | 0.174 | 1.497 | |

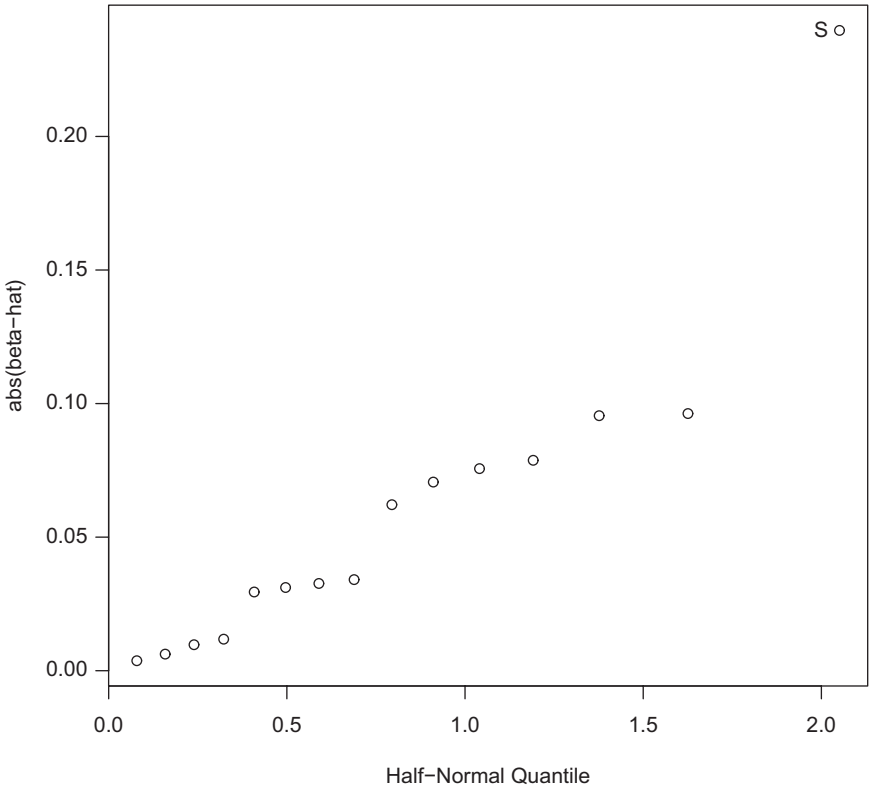


***FIGURE 3*** *Half-normal plot – Menon et al., unreplicated* $2^{5-1}$ *with response = particle size.*

## 3.3 Plackett and Burman Design

Plackett and Burman (17) provided a series of two-level fractional factorial designs for examining $n-1$ factors in $n$ runs, where $n$ is a multiple of 4 and $n \leq 100$. They omitted the case $n = 92$, which was later provided by Baumert et al. (18). When only main effects exist, these designs estimate all of them. Moreover, they are available for all run sizes that are a multiple of 4 (vs. a power of two as in a $2^{k-f}$ design). PB designs are thus extremely useful in screening situations. Note that the main effect of each factor can be evaluated by $\bar{y}_+ - \bar{y}_-$ as described in Section 2.2. Moreover, each column is orthogonal to all other columns. In other words, the inner product of any two columns is zero. Designs with such a property are called *orthogonal main-effect designs*.

*Construction of Plackett and Burman Designs*

A 12-run Plackett and Burman design can be obtained as follows:

1. Write down the set of signs $++-+++---+-$, provided by Plackett and Burman (17).

2. Permute the signs in 11 rows total, by taking the sign from the right hand side and moving it to the left hand side.
3. Add a 12th row of all minus signs.

For $n \leq 24$, all of the Plackett and Burman designs can be obtained by such a cyclic permutation. The signs for the first rows are:

$$
\begin{aligned}
n &= 8: \; + + + - + - - \\
n &= 12: \; + + - + + + - - - + - \\
n &= 16: \; + + + + - + - + + - - + - - - \\
n &= 20: \; + + - - + + + + - + - + - - - - + + - \\
n &= 24: \; + + + + + - + - + + - - + + - - + - + - - - -
\end{aligned}
$$

For $n$ a power of 2 (e.g., $n = 8$ and 16), we obtain a standard $2^{k-f}$ design. Note that Plackett and Burman designs can be viewed as a special class of orthogonal arrays (OAs). A complete list of OAs is given in Hedayat et al. (19) and can be found at the Web site http://www.research.att.com/~njas/oadir/.

## 3.4 Optimal Design

Recall from (4) that $\mathbf{X}$ represents the design matrix. From Eqs. (5)–(9) we see that $\mathbf{X}'\mathbf{X}$ plays an important role in inference. Specifically, an "optimal" design should be the one which minimizes $\mathrm{Var}(\hat{\beta})$ as in Eq. (6). In other words, one seeks the design whose $(\mathbf{X}'\mathbf{X})^{-1}$ matrix is minimized in some manner. Many design criteria have been proposed for choosing the design matrix $\mathbf{X}$ or for comparing design matrices. Essentially, most of them deal with the eigenvalues of the matrix $\mathbf{X}'\mathbf{X}$, $\lambda_1, \lambda_2, \ldots, \lambda_k$. The major alphabetical optimality criteria are as follows:

- *D-optimality*: maximize $\|\mathbf{X}'\mathbf{X}\| = \lambda_1 \times \lambda_2 \times \cdots \times \lambda_k$.
- *A-optimality*: minimize trace $(\mathbf{X}'\mathbf{X})^{-1} = \sum_{\mathrm{i}=1}^{\mathrm{k}} \lambda_{\mathrm{i}}^{-1}$.
- *E-optimality*: maximize the smallest eigenvalue of the $\mathbf{X}'\mathbf{X}$ matrix.
- *G-optimality*: minimize the maximum prediction variance over the operation region.
- *V-optimality*: minimize the average prediction variance over the operation region.

More (generalized) design optimalities can be found in Keifer (20), or more recently Pukelsheim (21). In essence, given the underlying model and the optimality criterion, any modern optimization technique can be used to generate an optimal design for the practitioner. However, the resulting design may not have desirable design properties, such as orthogonality, few levels in each factor, and symmetry, as in a $2^{k-f}$ design.

The problem of finding a D-optimal design has been very thoroughly explored mathematically; two early articles are by Hotelling (22) and by Mood (23). For two-level saturated D-optimal designs, the problem becomes finding an $n \times n$ matrix with

all its elements being either 1 or $-1$ such that its determinant is maximized. Such a "det-max" matrix problem has received a great deal of attention in the literature. Many of the resulting designs are available in software. If the underlying first-order design is believed to be true and if the optimality criterion is appropriate, these optimal designs can be very useful.

The practical value of saturated D-optimal designs, apart from Plackett and Burman (17) designs, is worthy of further investigation. First, these designs do not contain an equal number of high-level and low-level values. This can be quite disturbing to experimenters. The non-equal occurrence property implies that the factor is being partially confounded with the constant term (the column with all $+$'s), which is usually significantly different from zero. Second, unlike orthogonal designs, these designs lack similarity relationships among all the columns. For example, the correlations between every pair of columns are not necessarily the same. This raises questions about which factors to assign to which columns and whether it matters.

## 3.5 Supersaturated Designs

In order to obtain an unbiased estimate of the main effect of each factor, the number of experiments must exceed (or at least be equal to) the number of factors plus one (for estimating the grand average). When these two are equal, the design is called saturated and represents the minimum effort required to estimate all main effects. The standard advice given to users in such a screening process is to use the saturated design, which is "optimal" based upon certain theoretical optimality criteria. However, the non-significant effects are not of interest. Estimating all main effects may be wasteful if the goal is simply to detect those few active factors. If the number of active factors is indeed small, the use of a slightly biased estimate will still allow one to accomplish the identification of the active factors but significantly reduce the amount of experimental work. Developing such screening designs has long been a well-recognized problem, certainly since Satterthwaite (24).

When all factors can be reasonably arranged into several groups, group screening designs can be used (see, e.g., Reference 25). Only those factors in groups that are found to have large effects are studied further. The basic assumptions (such as the directions of possible effects are known), in fact, depend heavily upon the grouping scheme. While such methods may be appropriate in certain situations (e.g., blood tests), we are interested in systematic supersaturated designs for two-level factorial designs that can examine $k$ factors in $N < k + 1$ experiments in which no grouping scheme needs to be made.

***3.5.1 Supersaturated Designs Using Hadamard Matrices*** Lin (26) proposed a class of special supersaturated designs that can be easily constructed via half-fractions of Hadamard matrices. These designs can examine $k = N-2$ factors with $n = N/2$ runs, where $N$ is the order of the Hadamard matrix used. The Plackett and Burman (17) designs, which can be viewed as a special class of Hadamard matrices, are used to illustrate the basic construction method.

**TABLE 9 Plackett–Burman 12-Run Design**

| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 |
|---|---|---|---|---|---|---|---|---|---|---|
| + | + | − | + | + | + | − | − | − | + | − |
| + | − | + | + | + | − | − | − | + | − | + |
| − | + | + | + | − | − | − | + | − | + | + |
| + | + | + | − | − | − | + | − | + | + | − |
| + | + | − | − | − | + | − | + | + | − | + |
| + | − | − | − | + | − | + | + | − | + | + |
| − | − | − | + | − | + | + | − | + | + | + |
| − | − | + | − | + | + | − | + | + | + | − |
| − | + | − | + | + | − | + | + | + | − | − |
| + | − | + | + | − | + | + | + | − | − | − |
| − | + | + | − | + | + | + | − | − | − | + |
| − | − | − | − | − | − | − | − | − | − | − |

Table 9 shows the original 12-run Plackett and Burman design. If we take column 11 as the branching column, then the runs (rows) can be split into two groups: group I with the sign of $+1$ in column 11 (rows 2, 3, 5, 6, 7, and 11), and group II with the sign of $-1$ in column 11 (rows 1, 4, 8, 9, 10, and 12). Deleting column 11 from group I causes columns 1–10 to form a supersaturated design to examine $N-2 = 10$ factors in $N/2 = 6$ runs (runs 1–6, as indicated in Table 10). It can be shown that if group II is used, the resulting supersaturated design is an equivalent one. In general, a Plackett and Burman (17) design matrix can be split into two half-fractions according to a specific branching column whose signs equal $+1$ or $-1$. Specifically, take only the rows that have $+1$ in the branching column. Then, the $N-2$ columns other than the branching column will form a supersaturated design for $N-2$ factors in $N/2$ runs. Judged by a criterion proposed by Booth and Cox (27), these designs have been shown to be superior to other existing supersaturated designs. For recent advances in supersaturated design, see Lin (28).

## 4 DESIGNS FOR OPTIMIZATION

### 4.1 Response Surface Designs

We have focused on two-level designs to this point, limiting our model to linear and cross-product (interaction) terms. If we are seeking to optimize a response by

**TABLE 10 The Resulting Supersaturated Design for ($n$, $k$) = (6, 10)**

| No. | $x_1$ | $x_2$ | $x_3$ | $x_4$ | $x_5$ | $x_6$ | $x_7$ | $x_8$ | $x_9$ | $x_{10}$ |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | + | − | + | + | + | − | − | − | + | − |
| 2 | − | + | + | + | − | − | − | + | − | + |
| 3 | + | + | − | − | − | + | − | + | + | − |
| 4 | + | − | − | − | + | − | + | + | − | + |
| 5 | − | − | − | + | − | + | + | − | + | + |
| 6 | − | + | + | − | + | + | + | − | − | − |

maximizing (or minimizing) it, then we need to fit second-order models with quadratic terms. To do so, we must study at least three levels of a factor. One option is to study each of the $k$ factors at three levels and run all $3^k$ treatments. This experimental design grows in size very rapidly even without replication requiring 27 runs for three factors and 81 for four. Response surface designs generally allow some replication as well as the capability of fitting second-order models while requiring fewer runs than a $3^k$ design. We will discuss central composite designs, although there are other simlar designs such as Box–Behnken (29). These designs are used with quantitative factors.

Recall that we use $-1$ and $+1$ to denote the low and high levels in a two-level design. Then for a quantitative factor, 0 is obviously the level in the middle of this range. A central composite design for $k$ factors consists of a a two-level factorial or fractional factorial ($2^k$ or $2^{k-f}$) design, c centerpoints, and $2k$ axial points. The centerpoints have each factor set to its central value of 0. It is common to use $c > 1$ replications at this location in order to estimate experimental error. There are two axial or star points for each of the $k$ factors in the design. An axial point has one factor set at $\pm\alpha$ and all other factors set at 0. Setting each factor to $-\alpha$ once and $+\alpha$ once leads to the $2k$ axial points. The value of $\alpha$ may be chosen to give the design different properties. If $\alpha = 1$, then we have a face-centered design, as the axial points are located in the middle of each face of the cuboidal design region. Setting $\alpha = \sqrt{k}$ yields a spherical design, as the factorial and axial points lie on a sphere of radius $\sqrt{k}$. Figure 4 is an example of a two-factor central composite design with $\alpha = \sqrt{2}$. If $\alpha = n_f^{1/4}$, where $n_f$ is the number of factorial points, the design is said to be rotatable. In a rotatable design, the variance of the predicted response is the same for all points that are equidistant from the center. See Montgomery (10) for more discussion.

Vega et al. (30) describe a central compostie design used to study the effect of three factors (pH, flurbiprofen concentration, and P188 (a surface coating agent) concentration) on the physicochemical properties of nanoparticles and the drug entrapment efficiency in an ocular delivery system. They used a $2^3$ with $c = 2$ centerpoints, and a rotatable design with $\alpha = 8^{1/4} \approx 1.68$. From Table 11 we see that factor $x_2$ does not appear to affect the mean particle size. Factors $x_1$ and $x_3$ have main effects, but do not appear to interact or have much curvature over the regions studied.

## 4.2 Mixture Designs

In drug development, it is crucial to find the appropriate proportions of active ingredients, buffers, etc. For example, when producing a simple cough syrup, a% of the contents are from component A, b% is from b, and so on. The sum of the components constitutes the whole, so the factor levels cannot be set independently. If there are three factors, $x_1$, $x_2$, and $x_3$ for example, then $x_{i1} + x_{i2} + x_{i3} = 1$ for all $i$. With three factors, the design space is a triangular region. In general, it is known as a simplex. Additional restrictions may occur, such as a requirement of $x_{i1} > 2x_{i2}$, which remove areas from the triangular space.

We discuss an example from Marti-Mestres et al. (31). Three components of a sunscreen were studied, $X_1$ = coconut oil, $X_2$ = Primol, and $X_3$ = isopropyl myristate

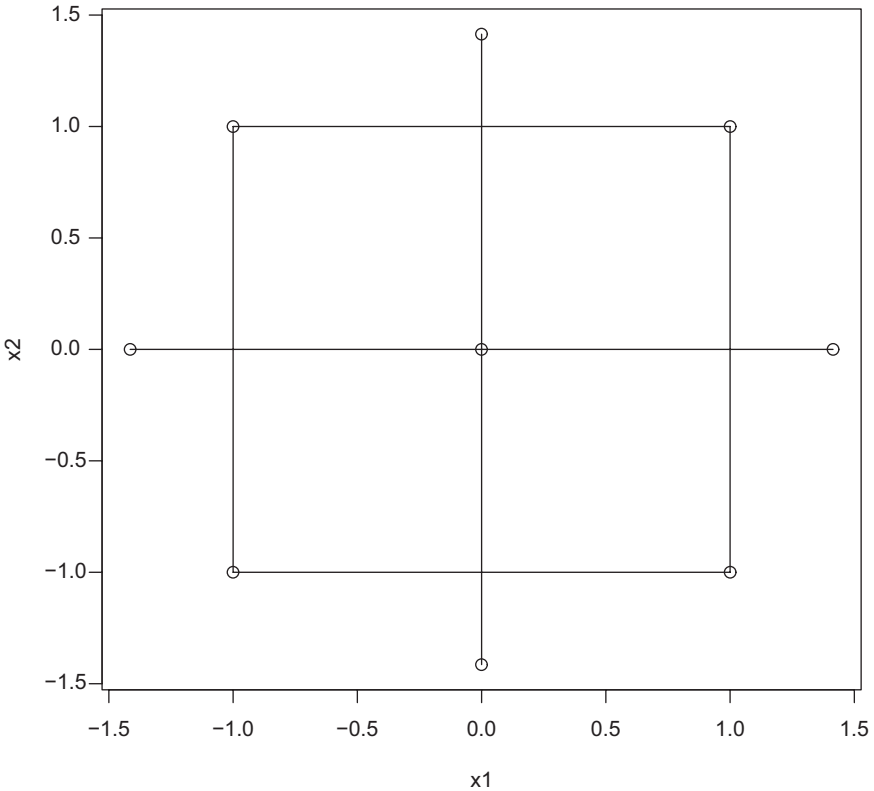


***FIGURE 4*** *Central composite design.*

(MIP), and the response was $y =$ oxybenzone solubility. Figure 5 shows the design. Note that each "corner" is included, which is 100% of a single component. The points on the edge represent absence of one component and an equal mixture of the remaining two. The interior points labeled as boxes comprise of two-thirds of one component and one-sixth of the other two. Finally, we have an equal mixture with each

**TABLE 11 Central Composite Design Output**

| | Estimate | Std. Error | *t*-value | Pr($>\|t\|$) |
|---|---|---|---|---|
| (Intercept) | 201.968 | 14.018 | 14.407 | 7e-06 |
| $x_1$ | −21.437 | 5.384 | −3.982 | 0.00727 |
| $x_2$ | 5.432 | 5.384 | 1.009 | 0.35196 |
| $x_3$ | 20.566 | 5.384 | 3.820 | 0.00876 |
| $x_1x_2$ | 13.162 | 7.031 | 1.872 | 0.11037 |
| $x_1x_3$ | 6.013 | 7.031 | 0.855 | 0.42531 |
| $x_2x_3$ | 4.063 | 7.031 | 0.578 | 0.58443 |
| $x_1sq$ | 11.302 | 6.542 | 1.728 | 0.13481 |
| $x_2sq$ | 11.355 | 6.542 | 1.736 | 0.13330 |
| $x_3sq$ | 3.046 | 6.542 | 0.466 | 0.65790 |

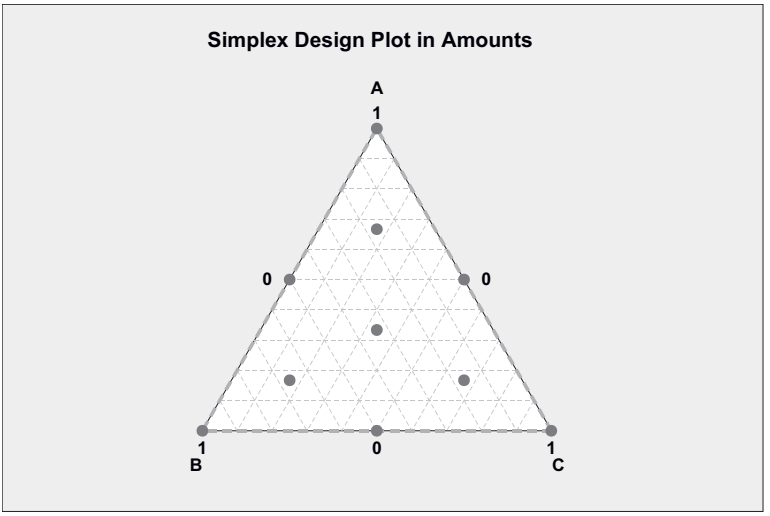


***FIGURE 5*** *Simplex design.*

component comprising one-third of the mixture. (For more information on designing and analyzing mixtures, see Cornell (32).)

Due to the constraint that $x_1 + x_2 + x_3 = 1$, the analysis is performed a little bit differently. We generally fit regression models without an intercept term. For this experiment, Marti-Mestres et al. found the following model to fit the data well:

$$y = 11.69x_1 - 0.23x_2 + 13.48x_3 - 7.4x_1x_2 + 4.35x_1x_3 + 7.49x_2x_3 + 54.41x_1x_2x_3$$

Figure 6 shows the response over the mixture region. Using close to 100% Primol ($X_2$) results in $< 0.5\%$ oxybenzone while the percentage of oxybenzone is $> 12.5\%$ for small amounts of Primol. The curvature shown in the plot is indicative of the interactions among the three factors.

## 5 FINAL REMARKS

In this chapter, we have outlined several different DoE techniques that are very useful for achieving quality by design in the pharmaceutical industry. These techniques can be used in conjunction with others such as combinatorial chemistry and high-throughput screening. While high-throughput screening and other methods allow a large quantity of data to be obtained relatively quickly, combining these methods with experimental design techniques allows generation of large volumes of high-quality (easily analyzed) data. Screening designs, including fractional factorials, Plackett–Burman designs, and supersaturated designs, make efficient use of experimental resources to cull a large number of potential factors down to a manageable few. These vital few factors may then be explored in a response surface (e.g., central

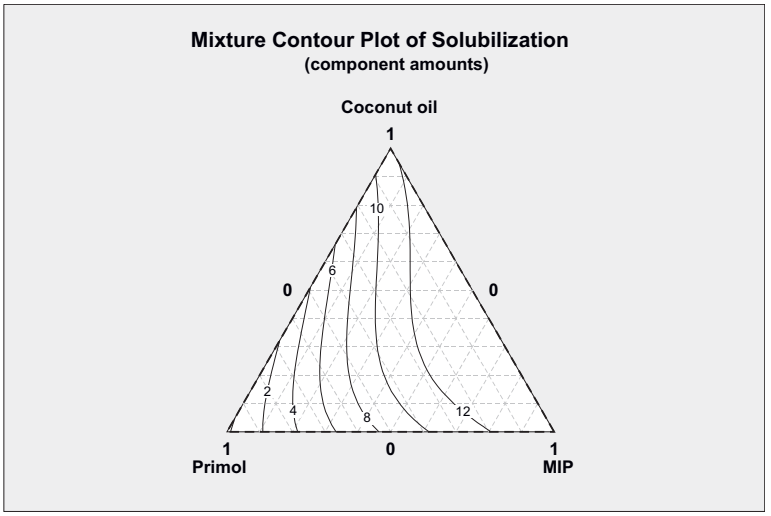


***FIGURE 6*** *Contour plot from mixture experiment.*

composite) design or a mixture (simplex) design to find the best combination of factor levels. An example of this sequence is given by Rosso et al. (33). Their studies included a screening DoE to study a broad range of factors and a central composite design to optimize those found to be important. By applying high-throughput methods, they were able to optimize the yield of a reaction by performing a large number of experiments in a short period of time.

There are many other DoE techinques that may be useful in pharmaceuticals that we could not discuss in this brief introduction. For example, the designs (and analyses) discussed in this chapter assume that the treatments were randomly assigned to the experimental units and that the order the experimental units were processed was randomized. This may be difficult or prohibitively expensive in practice. For example, if it is very costly to change reactor temperature, we may set it at one value and vary the other factors. Then we change reactor temperature and vary the other factors in a different random order, repeating for each reactor temperature to be studied. This is an example of a split plot design that requires a different analysis method than described herein.

The experimenter may also be aware that ambient laboratory conditions may vary from day to day. In this case, the results may be more favorable on Tuesday than Monday for reasons unrelated to those of interest in the study. Randomized complete block designs may be used to "block out" the variation due to day-to-day differences, thereby reducing the noise in the experiment.

In addition to finding settings of factors to optimize the mean response, the experimenter may also be interested in finding settings that minimize the variance in the response. Settings that provide desirable mean values of the response with low variation are known as robust settings. Again, different designs and analyses may be used in these robust-design studies.

So while there are many other DoE topics to be explored, we have introduced several designs and analyses that can help the pharmaceutical researcher become a more efficient experimenter. Classical experimental designs, including those popularized in the agricultural sciences, such as factorials and fractional factorials, can play an important role. Others that are more applicable in a laboratory setting, such as mixture designs, have been developed and implemented in food sciences and chemical development. Collectively, these methods can be used to achieve quality by design in the pharmaceutical industry.

## REFERENCES


1. Laird T. Special feature section: laboratory automation in process R&D, editorial. Org Process Res Dev 2001; 5: 272.
2. Gooding OW, Vo L, Bhattacharyya S, Labadie J. Use of statistical design of experiments in the optimization of amide synthesis utilizing polystyrene-supported N-hydroxybenzotrizole resin. J Comb Chem 2002; 4: 576–583.
3. Owen MR, Luscombe C, Lai L-W, Godbert S, Crookes DL, Emiabata-Smith D. Efficiency by design: optimisation in process research. Org Process Res Dev 2001; 5: 308–323.
4. Lutz MW, Menius A, Choi TD, Gooding Laskody R, Domanico PL, Goetz AS, Saussy DL. Experimental design for high-throughput srceening. Drug Discov Today 1996; 1: 277–286.
5. Tye H. Application of statistical design of experiments methods in drug discovery. Drug Discov Today 2004; 9: 485–491.
6. Kutner MH, Nachtsheim CJ, Neter J, Li W. Applied linear statistical models. 5th ed.; New York: McGraw-Hill; 2005.
7. Coleman DE, Montgomery DC. A systematic approach to planning a designed industrial experiment (with discussion). Technometrics 1993; 35: 1–27.
8. Box GEP, Hunter JS, Hunter WG. Statistics for experimenters. 2nd ed.; New York: Wiley; 2005.
9. Wu CFJ, Hamada M. Experiments: planning, analysis and parameter design optimization. New York: Wiley; 2000.
10. Montgomery DC. Design and analysis of experiments. 7th ed.; New York: Wiley; 2009.
11. Khanvilkar KH, Huang Y, Moore AD. Influence of hydroxypropyl methylcellulose mixture, apparent viscosity, and tablet hardness on drug release using a $2^3$ full factorial design. Drug Dev Ind Pharm 2002; 28: 601–608.
12. Daniel C. Use of half-normal plots in interpreting two-level factorial experiments". Technometrics 1959; 1: 311–349.
13. Menon A, Dhodi N, Mandella W, Chakrabarti S. Identifying fluid-bed parameters affecting product variability. Int J Pharm 1996; 140: 207–218.
14. Lenth RV. Quick and easy analysis of unreplicated factorials. Technometrics 1989; 31: 469–473.
15. Haaland PD, O'Connell MA. Inference for effect-saturated fractional factorials. Technometrics 1995; 37: 82–93.
16. Hamada M, Balakrishnan N. Analyzing unreplicated factorial experiments: a review with some new proposals (Disc: p29-42). Stat Sin 1998; 8: 1–23.

17. Plackett RL, Burman JP. The design of optimum multifactorial experiments. Biometrika 1946; 33: 305–325.
18. Baumert L, Golomb SW, Hall J. Discovery of a Hadamard matrix of order 92. Bull Am Math Soc 1962; 68: 237–238.
19. Hedayat A, Sloane NJA, Stufken J. Orthogonal arrays: theory and application. New York: Springer; 1999.
20. Keifer J. Optimum experimental designs. J R Stat Soc B 1959; 21: 272–304.
21. Pukelsheim F. Optimal design of experiments. New York: Wiley; 1993.
22. Hotelling H. Some improvements in weighing and other experimental techniques. Ann Math Stat 1944; 15: 297–306.
23. Mood AM. On Hotelling's weighing problem. Ann Math Stat 1946; 17: 432–446.
24. Satterthwaite FE. Random balance experimentation (with discussion). Technometrics 1959; 1: 111–137.
25. Watson GS. A study of the group screening method. Technometrics 1961; 3: 371–388.
26. Lin DKJ. A new class of supersaturated designs. Technometrics 1993; 37: 213–223.
27. Booth KH, Cox DR. Some systematic supersaturated designs. Technometrics 1962; 4: 489–495.
28. Lin DKJ. Industrial experimentation for screening. Handbook of statistics in industry Chapter 2 Volume 22 Rao CR, Khattree R, editors. New York: North Holland; 2003. p. 33–73.
29. Box GEP, Behnken D. Some new three level designs for the study of quantitative variables. Technometrics 1960; 2: 455–475.
30. Vega E, Egea MA, Valls O, Espina M, García ML. Flurbiprofen loaded biodegradable nanoparticles for ophtalmic administration. J Pharm Sci 2006; 95: 2393–2405.
31. Marti-Mestres G, Nielloud F, Fortune R, Fernandez C, Maillois H. Effect of the addition of oxybenzone or octyl-methoxycinnamate on particle size of submicron emulsions. Drug Dev Ind Pharm 2000; 26: 349–356.
32. Cornell JA. Experiments with mixtures. New York: Wiley; 1990.
33. Rosso VW, Pazdan JL, Venit JJ. Rapid optimization of the hydrolysis of N-trifluoroacetyl-S-tert-leucine-N-methylamide using high-throughput chemical development techniques. Org Process Res Dev 2002; 5: 294–298.

# *X-Ray Fluorescence Spectrometry in Medicine*

**Linda Nie**

*School of Health Sciences, Purdue University, 550 Stadium, Mall Drive, West Lafayette, Indiana*

## 1 INTRODUCTION

Some elements have toxic effects on the human body, and some are essential but have adverse effects at high levels. Monitoring these elements is critical for scientists to assess the association between exposures and health outcomes, and for clinicians to diagnose disease related to exposure (1,2). X-ray fluorescence (XRF) and neutron activation analysis (NAA) are the two main non-invasive methods for *in vivo* and *in vitro* element analysis (3–6). The XRF technology for *in vivo* element measurement will be described in detail in this chapter. In an *in vivo* element measurement, the subject/patient is positioned in front of an XRF device by a trained researcher or technician for a fixed time, and the concentration of the element in a target tissue is determined from an XRF spectrum.

The first *in vivo* XRF measurement was performed in the 1960s to measure iodine in the thyroid (7). Later in the 1980s, a technique to measure iodine in human tissue *in vivo* from X-ray contrast agents was reported (8). Ahlgren et al. (9) made the first bone lead measurement in the 1970s, and since then a number of XRF techniques have been developed, among which XRF bone lead measurement is the most intensively studied and applied. Bone lead has been used to investigate the association between

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing,*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

lead exposure and many health outcomes, such as hypertension, kidney impairment, and cognitive loss (10–15). The *in vivo* XRF technology was also explored to assess occupational or high level environmental exposure to other elements. These elements include but are not limited to Cd, As, Hg, and U.

Another rapidly rising technique, XRF microprobe imaging, will also be briefly described. XRF microprobe imaging maps an element's distribution in human or animal tissue/cells. The imaging process is mostly performed *in vitro* using necropsy or biopsy tissues. Synchrotron radiation XRF (SRXRF) provides excellent trace element sensitivity with its highly intensive beam, and is a valuable microprobe imaging technology (5,16,17). The advanced technologies in synchrotrons and X-ray optics open the door to nano-scale trace element probing, which brings the technology down to the molecular or cellular level (18,19).

Both *in vivo* and synchrotron XRF techniques have made significant contributions to our understanding of the role of trace elements in terms of health effects and biological consequences. There is great potential for both techniques to be improved and broadly applied in the near future.

## 2 DEFINITIONS

### 2.1 X-Ray Fluorescence

XRF is the emission of characteristic "secondary" (or fluorescent) X-rays from a material that has been bombarded with X-rays or $\gamma$-rays. The energies of these fluorescent X-rays are specific to the elements being assessed, and hence the concentration of the element can be calculated from the spectrum collected by a detection system.

### 2.2 Monte Carlo (MC) Simulation

Monte Carlo methods are a class of computational algorithms that rely on repeated random sampling and are widely employed to simulate radiation transportation in materials. For example, to simulate the transportation of photons in a material, the MC program generates random numbers to determine the direction in which the photon travels, the type of interaction between the photons and the material, the energy transition during an interaction, and all the parameters that are required to track the photon. The program tracks the photon until it is absorbed or leaves the relevant area. Normally, millions of photons are tracked to obtain a statistically significant result. The result can be the energy deposited in the material, a signal detected by a detector, or anything else related to the photons that can be theoretically predicted.

### 2.3 Minimum Detection Limit (MDL)

The minimum detection limit is the smallest concentration of an element that can be identified by an XRF system. It is defined as three times the standard deviation of the calculated concentrations for repeat measurements.

## 2.4 Phantom

A phantom is a tissue-equivalent standard doped with a known concentration of the element of interest. Phantoms are manufactured to simulate the *in vivo* measurement or to calibrate the XRF system.

## 2.5 KXRF and LXRF

An orbit electron is knocked out of the shell during the XRF process. If the electron is from the K shell, then the characteristic X-rays produced are called K-X-rays, and the system that detects an element by counting its K-X-ray is called a KXRF system (an LXRF system counts L-X-rays, etc.).

## 2.6 X-Ray Tube

X-ray tubes normally consist of an electron generator, a vacuum tube with a metal target, and a cooling system. The X-rays are produced through the interaction of the electrons with the target, either with the whole atom to produce bremsstrahlung X-rays or with the orbit electron to produce characteristic X-rays.

## 2.7 Polarized X-Ray Beam

Radioisotope- and X-ray tube-induced X-ray sources are isotropic, producing symmetrically distributed polarized X-rays. If a polarizer is placed in front of an X-ray beam, it can select the X-rays polarized on a certain plane and produce a partially or totally polarized X-ray beam. While the X-rays interact with a material, the coherent scattering and incoherent scattering cross-sections are strongly dependent on the angle between the incident beam and the scattered beam. Hence applying a polarized X-ray beam to induce X-ray fluorescence and to detect the characteristic X-rays at the axis of X-ray polarization can significantly reduce the scattered background radiation. One disadvantage of polarized X-ray beam is that the intensity of the beam is greatly reduced after polarization.

## 2.8 Radiation Detector

A radiation detector is used to detect, track, and/or identify radiation particles, such as those produced by nuclear decay, cosmic radiation, or radiation interactions. It works by interacting with the radiation particle and collecting and recording the charges produced. A radiation detector can work as a dosimeter to measure the deposited energy, or as a spectrometer to measure the pulse height distribution of the particles. The detectors described in this chapter are spectrometers that record the signals produced during the interaction of the X-rays and a material or tissue, which include the characteristic X-ray signals.

### 2.9 Radiation Dose

Radiation dose is a measure of the energy deposited in a medium by ionizing radiation. It is expressed as energy deposited per unit mass of medium. Two types of doses are used in this chapter: the organ dose, or the energy deposited in the organ divided by the mass of the irradiated organ, and then multiplied by the weighting factor of the radiation; and the whole body effective dose, which is the organ dose multiplied by the fraction of the organ being irradiated, multiplied by the risk factor of the organ, which is defined by the ICRP (20).

### 2.10 Synchrotron Radiation

Synchrotron radiation is electromagnetic radiation generated by the acceleration of ultrarelativistic (i.e., moving near the speed of light) charged particles through magnetic fields. The synchrotron radiation-induced XRF is called synchrotron XRF and is several orders of magnitude more intense than conventional X-rays as well as being highly focused and polarized.

## 3 THE PRINCIPLE OF XRF

When a beam of X-rays falls onto a given specimen, four basic phenomena result: photoelectric absorption, Compton scattering (also called incoherent scattering), pair production, and Rayleigh scattering (also called coherent scattering) (21). Photoelectric absorption by the elements in the specimen results in the emission of monochromatic X-rays that are uniquely characteristic of the elements.

In the photoelectric absorption process, an incident X-ray undergoes an interaction with an absorber atom in which the X-ray completely disappears and an orbital electron is ejected by the atom from one of its bound shells (K, L, M, etc.). The whole process is subject to the laws of energy conservation as well as momentum conservation, so it cannot take place with free electrons. For X-rays of sufficient energy, the most probable origin of the photoelectron is the most tightly bound shell (K) of the atom. Part of the incident X-ray energy is used to overcome the binding energy of the photoelectron, and the rest is converted to the energy of the photoelectron.

In addition to the photoelectron, the interaction also creates an ionized atom with a vacancy in one of its bound shells. This vacancy is filled through rearrangement of electrons from other shells of the atom, and eventually a free electron is captured from the medium. As a consequence, one or more characteristic X-rays are generated. In some instances, the emission of an electron (called an Auger electron) from another shell may substitute for the characteristic X-ray in carrying away the atomic excitation energy. The incident X-ray could knock out an electron from the K, L, or M shells, and so on, and the vacancy presents in these shells with certain proportions. If the vacancy is in the K shell, then the characteristic X-ray ejected while the electron jumps from other shells to the vacancy is called a K-series X-ray; a vacancy in the L shell produces an L-series X-ray, and so on. Subscript Greek letters (e.g., $K_\alpha$s and $K_\beta$s) indicate the

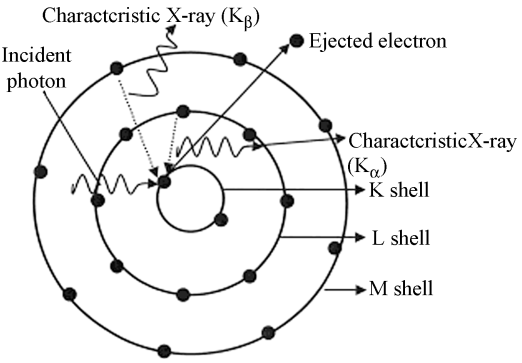


***FIGURE 1*** *Process of X-ray fluorescence.*

shell and layer from which the electron originates. The energy of the X-ray is determined by the charge, or atomic number, of the nucleus. By measuring the energy and intensity of the characteristic X-rays, we can determine the type and concentration of the element. This is the basic principle underlying XRF measurement of a sample. Figure 1 illustrates the process of X-ray fluorescence.

The other three interactions do not contribute to characteristic X-rays, but Compton scattering and Rayleigh scattering sometimes provide useful information for spectrum and data analysis or produce large background in the energy range of the X-rays. In Compton scattering, an X-ray interacts with an absorber atom and transfers a portion of its energy to an electron. In Rayleigh scattering, the X-ray interacts with all the electrons of an absorber elastically, keeps all its energy, and changes its direction.

## 4 XRF SPECTROMETRY

### 4.1 X-Ray Source, Detection System, and Electronic System

An XRF spectroscopy system consists of three main components: the X-ray source, the detector system, and the electronic system. The X-ray source interacts with the sample and produces characteristic X-rays for target elements; the detector collects the signals; and the electronics process the signals. A spectrum is then displayed on a computer screen and analyzed by a specific software program. The primary goal in developing an *in vivo* system is to achieve the highest sensitivity (lowest detection limit or DL) with an acceptable radiation dose. The selection of radiation source, detection system, and electronics is based on this rule. There are two types of X-ray sources for *in vivo* XRF: radioisotope and X-ray tube. Radioisotope sources have fixed energies and intensities, and X-ray tube sources have flexible energy spectra based on the tube setup and the use of filters. The energy of the source has to be higher than the binding energy of the shell electrons to produce characteristic X-rays for the corresponding shell. The closer the source energy and the binding energy, the greater

the photoelectric cross-section and hence the yield of the characteristic X-rays. While it is important to choose a source with energy close to the binding energy, one needs to be aware that both the X-ray signal count and the background count under the X-ray signal determine the minimal detection limit. Simply put,

$$\mathrm{MDL} \propto \frac{\sqrt{N_{\mathrm{bg}}}}{N_{\mathrm{net}}}$$

where $N_{\mathrm{net}}$ is the net count of X-ray signals, and $N_{\mathrm{bg}}$ is the background count under the X-ray energy range. Improvement of the system means a lower ratio of $\sqrt{N_{\mathrm{bg}}}/N_{\mathrm{net}}$. The ultimate goal of system design is to maximize the signal and to reduce the background.

HPGe detector and Si(Li) detector are two of the most frequently applied detectors for XRF spectroscopy; both have excellent energy resolution (21). The relevant energy range for XRF spectroscopy is 10–100 keV. Most HPGe detectors have high efficiency at energy above 30 keV. Si(Li) detectors are better for energies between 10 and 30 keV. HPGe detectors can be designed for high efficiency in an extra-low-energy range; however, it is difficult to design a Si detector with high efficiency at energy above 80 keV. One disadvantage of these two types of detectors is that they must be cooled by liquid nitrogen to reduce noise. Currently, scientists are working on other types of detectors that can be operated at room temperature or can be cooled by thermoelectronics (22–24). Selection of the detectors is primarily determined by the relevant energy range, resolution of the detector, and price.

There are two types of electronics for XRF spectroscopy (25,26): the conventional analog system and the modern digital system. The analog system digitizes the signals at the end of the signal processing chain, and modern digital signal processing (DSP) digitizes the signals at the beginning of the signal processing chain (after the preamplifier). The DSP approach minimizes the amount of analog circuitry necessary, resulting in increased stability and reproducibility, enhancing energy resolution and increasing signal throughput. The compact digital systems are gradually replacing conventional analog systems. Selection of the electronics is largely based on the energy resolution, signal throughput, and price.

## 4.2 Spectrum Analysis

Once a spectrum is collected, the next step is to analyze and acquire useful information from it. The net count of characteristic X-ray signals is the usual information we need, and there are commercial programs to extract this information. *In vivo* XRF element measurement is a specialized field within X-ray spectrometry, and spectrum analysis programs are generally developed by the investigators in this field. There are many ways to fit a spectrum. The easiest way is to define a function and fit the spectrum with an available algorithm in commercial software. One can also write a program with a theoretically proven algorithm, for example, least square or maximum likelihood, to fit any arbitrary functions (27). Genetic algorithms are also applied to fit the spectrum (28).

### 4.3 System Calibration and Validation

A set of phantoms is manufactured for system calibration. The phantoms are made of material equivalent to the target organ and doped with a known amount of the element of interest. By comparing the net counts of characteristic X-rays of the interested element for the phantom and the unknown samples (human organ/tissue), one can calculate the element concentration in the unknown. If the organ/tissue is buried under another organ/tissue, the radiation attenuation of the overlaying organ/tissue needs to be corrected for. At least one phantom is required to calibrate the system; generally a set of phantoms is measured to construct a calibration line.

Before doing any measurement with a live human being, a system needs to be validated. Other standardized analytical procedures, for example, NAA and inductively coupled plasma-mass spectroscopy (ICP-MS), are used to validate the system (29,30). Normally postmortem organ/tissue samples are used for system validation.

## 5 *IN VIVO* XRF MEASUREMENT OF LEAD IN BONE—INTRODUCTION AND SYSTEM DEVELOPMENT

When an XRF system is designed, two methods can be applied to test its sensitivity and feasibility before a real system is assembled. One can perform a theoretical calculation to obtain the most important values of the system (such as MDL and radiation dose). The calculation of the intensity of characteristic X-rays is illustrated in Appendix A. The calculation of background (coherent and incoherent scattering) follows similar procedure. The second method is to run some Monte Carlo simulations to test and optimize the system. The optimized system is then examined with experiments. In the following sections, different XRF bone lead measurement systems are described, and the development of a KXRF bone lead measurement system is illustrated.

### 5.1 Introduction

Both K- and L-X-rays are suitable for the analysis of lead in bone. To induce K-X-ray fluorescence, the incident radiation has to be above the 88.001 keV K-absorption edge (31). The sources used in K-XRF bone lead measurement system include $^{57}$Co and $^{109}$Cd. The maximum binding energy of the L electrons of lead is 15.870 keV, and the sources used in L-XRF bone lead measurement system include $^{109}$Cd and $^{125}$I, as well as X-ray tube sources.

***5.1.1 In vivo LXRF Bone Lead Measurement Systems*** Feasibility study of the *in vivo* measurement of lead in bone using L-X-ray fluorescence was first investigated by Wielopolski et al. (32). The researchers concluded that $^{109}$Cd and $^{125}$I could be used to detect bone lead by inducing lead L X-rays. Later, Wielopolski (33) conducted a postmortem study to measure bone lead with a $^{125}$I source. The first *in vivo* LXRF study estimated bone Pb values in 45 workers (34).

Later, an X-ray tube source based LXRF system was designed for potential use in children (35,36). This system utilizes polarized X-rays and hence greatly reduced the intensity of the scattering background. Another LXRF system with a secondary target was investigated by Todd (37–39). The polarization of a $^{109}$Cd source to induce lead L X-rays was discussed by Ao et al. (40). Recent research related to LXRF *in vivo* bone lead measurement includes a study to combine K and LXRF methods for the tibia lead concentration (41) and a preliminary study on K and L coincidence spectroscopy (42). The development of a portable LXRF bone lead measurement system is under way (43).

The LXRF method, which is limited to the measurement of lead in the near surface of the bone, has potential for making measurements of short duration with portable equipment. However, the LXRF method has major disadvantages in that its detection limit depends strongly on the thickness of the overlying tissue, and the technique is unlikely to be practical for overlying tissue thicknesses much greater than 5 mm.

***5.1.2 In vivo KXRF Bone Lead Measurement Systems*** The first *in vivo* measurement of bone lead was reported by Ahlgren et al. (9). In this study, researchers utilized γ-rays from $^{57}$Co to excite the K series X-rays of lead in finger bone with a 90° geometry. Later, $^{109}$Cd was found to have significant advantages (44,45). The initial $^{109}$Cd XRF bone lead measurement system consists of a 16 mm diameter HPGe detector with an annular $^{109}$Cd source mounted in front of the detector. An improved system developed by Gordon et al. (46) is based on a larger-area HPGe detector with a diameter of 51 mm and a point $^{109}$Cd source and is the system installed in most of KXRF labs worldwide. Recently a new KXRF bone lead measurement system was developed that consists of a clover-leaf-shaped four-detector array, four digital signal processing units, and a point $^{109}$Cd source (47,48). The high sensitivity of the new system opens the door to the investigation of low-level lead toxicity in children. The development of this new system is summarized in section 411.5.2.

***5.1.3 $^{109}$Cd γ-Ray Induced KXRF Bone Lead Measurement*** In $^{109}$Cd γ-ray induced KXRF bone lead measurement, human tibia, calcaneus, or other bone sites are irradiated to obtain lead K X-rays and then to calculate bone lead concentration. The lead K edge is 88.001 keV. The $^{109}$Cd source emits γ-rays of 88.035 keV with 3.6% intensity (31), just above the energy threshold for the K shell absorption edge in lead, thus maximizing the photoelectric cross-section and hence the X-ray fluorescence yield. The γ-rays interact with a K shell electron in lead, and the electron is ejected. This produces a vacancy in the K shell, which an electron from an outer shell will fill. The energy is released as $K_\alpha$ or $K_\beta$ X-rays with 96% intensity or Auger electrons with around 4% intensity (31). There are also coherent (Rayleigh) scattered γ-rays with energy 88.035 keV and Compton scattered γ-rays with energy range 65.47 keV (180° scattering) to 88 keV for X-rays scattered through very small angles in the forward direction. Figure 2 plots the fluoresced spectrum using a $^{109}$Cd source for K XRF lead measurement of a lead-doped phantom with a concentration of 200 μg/g. The Compton-scattered photons have peak energy of 66 keV, as expected for the 180° back-scattered geometry used. The individual characteristic K X-rays of

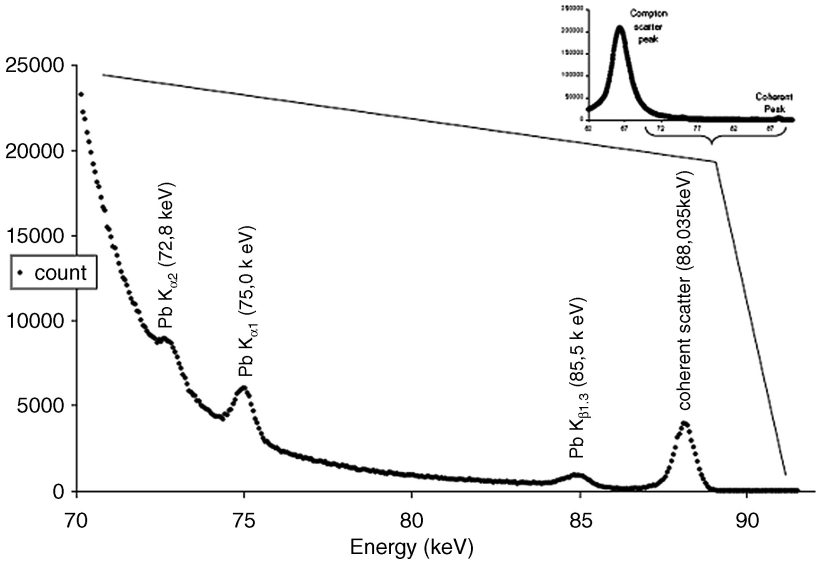


***FIGURE 2*** *Spectrum for $^{109}$Cd KXRF lead measurement.*

lead and the coherent peak at 88 keV, coming primarily from scattering from the calcium in the bone, are identified in the spectrum.

A unique feature of using a $^{109}$Cd source and backscattering geometry is that both α peaks and β peaks can be normalized to the coherent peak, making it unnecessary to correct for tissue attenuation and geometry (45) to obtain the lead concentration as a fraction of the calcium concentration.

The system is calibrated by 10 phantoms made of plaster of Paris (a bone equivalent material) with known lead concentrations (μg Pb/g plaster of Paris). The ratios of both ($K_{\alpha 1}$ peak)/(coherent peak) and ($K_{\beta 1}$ peak)/(coherent peak) are calculated from the spectra by a peak-fitting program. The two calibration lines (α calibration line and β calibration line) are the lines of the ratios versus the Pb concentrations. The concentration of Pb in bone for *in vivo* measurement is calculated from these two calibration lines. The final bone lead concentration is an inverse-variance-weighted value from α calibration and β calibration (49,50).

## 5.2 Development of a New KXRF Bone Lead Measurement System

As mentioned above, the ultimate goal in developing an *in vivo* XRF element measurement system is to decrease the MDL while minimizing the radiation dose. In the following section, the development of a new $^{109}$Cd induced KXRF bone lead measurement system is summarized (48). The development of other systems follows a similar procedure.

***5.2.1 Theoretical Calculations*** It is possible to estimate the MDL by manual calculation before a system is assembled. The count of characteristic X-ray signals and the background signals that will be detected by a specific detector are calculated by theoretical formula using interaction cross-sections, detector efficiency, and geometry factors at a fixed geometry, and the MDL is calculated as three times the uncertainty derived from the signal and the background. Manual calculation is difficult for the $^{109}$Cd-induced bone lead measurement system, because there is a large Compton background under the lead K-X-ray peaks, and the counts of the background depend heavily on the Compton broadening effect and the detector reaction function. Because this new system is an improvement of the previous systems, relative MDL (or MDL ratio between the new system and the conventional system) is calculated instead. The new system design is based on an array HPGe detector, with four 16 mm diameter detectors, compared to one 51 mm diameter large detector for the conventional system. We know that $\mathrm{MDL} \propto \sqrt{N_{\mathrm{bg}}}/N_{\mathrm{net}}$. Four detectors produce four times the net signal and four times the background (given a more active $^{109}$Cd source), which gives rise to an MDL ratio of 0.5. The smaller detectors have a better resolution (about 550 ev at the 88.035 keV coherent peak), compared to a resolution of about 750 ev for the conventional large detector. The better resolution accounts for an MDL improvement factor of $\sqrt{550}/\sqrt{750} = 0.86$ resulting from the reduced background. The predicted MDL improvement factor is $0.5 \times 0.86 = 0.43$.

***5.2.2 Monte Carlo Simulations*** MC programs are powerful tools to simulate radiation transportation. They have long been applied to simulate the transportation of neutrons and relatively high energy photons (51–53). In recent years, MC simulation has been successfully applied to the transportation of low energy X-rays. Two of the most frequently utilized MC programs are MCNP and EGS. The old versions of MCNP (MCNP4 and previous versions) and EGS (EGS4 and previous versions) have some apparent defects for the simulation of low-energy X-ray transportation, particularly in the application of XRF system development. For example, the characteristic X-rays are not very well simulated in the programs, and the Doppler broadening of Compton scattering is not simulated at all. The physics of the low-energy X-rays are greatly improved in the latest version of these two programs. To obtain a program that is tailored for one's specific application, the best way is to write one's own program if time and budget permit. The program used to explore the MDL of the new $^{109}$Cd KXRF bone lead measurement system is an in-house program written for radioisotope-induced *in vivo* XRF element measurement, modified to incorporate bone lead measurement (54). MC simulations were performed for both the conventional system and the new array system. Different array designs were simulated, as well as four geometries with different detector sample distance (DSD) and source sample distance (SSD). The simulations test two types of phantoms: the bare phantom made of plaster of Paris to simulate bare bone measurement; and the bare phantom in a soft tissue equivalent plastic phantom to simulate the *in vivo* leg measurement. The absolute MDLs are lower than expected when compared to the MDLs calculated from experiments with conventional systems, possibly because the Compton scattering was not very well simulated. However, the

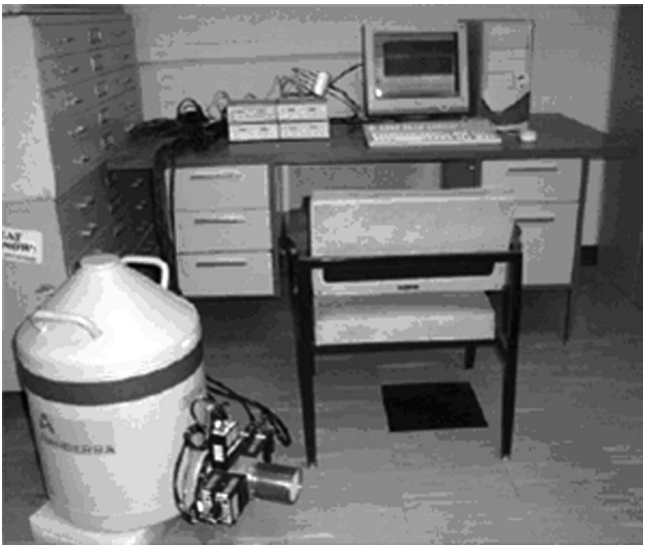

***FIGURE 3*** *The advanced* $^{109}$Cd *bone lead measurement system.*

MDL ratios provide valuable information on which design and geometry are the best choices.

***5.2.3 Experiments and System Calibration*** The optimized design and setting of the system was chosen through the MC simulations and manufactured by a company producing radiation detectors. The performance of three types of electronic systems was also compared to choose the best one to process the signals from the detectors. The affiliated devices, such as the device to position a patient and the device to hold the $^{109}$Cd source, were manufactured. Figure 3 is a picture of the detection system, which includes an array detector covered under the iron sleeve, a liquid nitrogen Dewar, four preamplifiers, four digital signal processing devices, and a computer. The $^{109}$Cd source is mounted in front of the detector. Experiments were performed on this system to verify the MDL improvement. The average MDLs calculated from the experiments were 6 μg/g for the conventional system and 2 μg/g for the new system. Two calibration lines, one for $K_{\alpha 1}$ peak and one for $K_{\beta 1}$ peak, were produced with the lead-doped phantoms.

***5.2.4 Dosimetry Study*** A small amount of radiation is involved in the *in vivo* XRF element measurement. The researchers or the hospitals containing this type of facility are required to understand the minimal radiation risk to the patient and apply for permission from the human subject committee before using this system. The dose can be estimated from the dose calculated for the conventional system (55), which is equivalent to 5 min of natural background radiation. The new system requires a more active $^{109}$Cd source, about 10 times that of the old source, hence the dose delivered by

the new system is about 50 min of the natural background radiation. Because the geometry of the new system is different from that of the conventional system, and the new system is intended for bone lead measurement in children, a dosimetry study was conducted to accurately estimate the dose to children (56). The result shows that the doses delivered to adults and children are negligible. After the dosimetry study, the researchers applied for and were granted a license to use this system for *in vivo* bone lead measurement.

***5.2.5 In vivo Investigation*** An *in vivo* investigation was conducted with the new system to verify the experimental result. Twenty volunteers were recruited, and bone lead was measured at the tibia site by both the new system and the conventional system. The study confirmed the improvement of the new system over the conventional system, that is, the average MDL of the new system is about 2 μg/g, compared with about 6 μg/g for the conventional system (48).

***5.2.6 Epidemiological and Clinical Significance of the Bone Lead Measurement System, and the Implication of the New System*** There are about 20 KXRF bone lead measurement laboratories around the world, three of which installed the newly developed KXRF system. The KXRF systems in these laboratories made significant contributions to our understanding of the impact of cumulative lead exposure in adult toxicity. The significance of bone lead as a biological marker of lead dose for adults has been reviewed in some previous articles (57–59). Lead exposure has largely declined over the last two decades; nevertheless, lead remains a highly significant environmental toxicant with regard to exposure to children as well as a persistent occupational health problem for hundreds of thousands of workers. It continues to be the focus of intensive screening in pediatric and occupational medicine clinics. In the past, the relatively high MDL of the system and the low overall bone density of children prevented the measurement of bone lead in patients younger than 12–14 years. The new system has a MDL of 2 μg/g and is sufficiently sensitive to measure bone lead levels in children. Using the new system, at least two studies are under way to investigate the association of lead exposure and health effects (particularly neurological effects) among children with bone lead as a biomarker. Historically, clinical research on lead poisoning has focused on prevention of exposure. While such work is important, efforts to mitigate lead's toxicity among children already exposed has received scant attention. By better quantifying the cumulative dose of lead with this new system, we will be able to better identify those children who (i) have the highest exposure; (ii) are most likely to suffer the adverse health effects of lead toxicity; and (iii) are most likely to benefit from medical and social interventions to reduce lead exposure and toxicity. The improved sensitivity of the new system has significance in epidemiological studies, in that it will greatly enhance the significance of the results of analysis with the same population or greatly reduce the size of the required survey population without reducing the statistical significance.

## 6 *IN VIVO* XRF MEASUREMENT SYSTEMS FOR OTHER ELEMENTS

Most of the elements with atomic number $Z$ greater than 15 in the periodic table can be detected at very low levels *in vitro* by XRF. Several factors restrict the *in vivo* measurement of elements. Firstly, the energies of the characteristic X-rays for moderately high $Z$ elements are low, and hence limit detection of these elements to superficial organs such as skin. Secondly, the human body consists mainly of low-$Z$ elements; the relatively high incoherent scattering cross-section in low-$Z$ matrices leads to a large background in measured spectrum and eventually an increased MDL. Thirdly, correction of the geometry and attenuation is difficult. Fourthly, the X-ray source intensity is limited by radiation dose. Despite these limitations, *in vivo* measurements of many elements have been investigated, and such studies are expected to accelerate with each improvement in *in vivo* techniques. In the following section, applications of *in vivo* measurement of elements other than lead are summarized, in the order of their atomic number.

### 6.1 Iron, Copper, Zinc, and Arsenic in Skin

Iron and zinc are essential elements in the human body; copper is neutral; arsenic is toxic. These are moderately high $Z$ elements, with energies of characteristic X-rays ranging from 6 to 12 keV. The intensive attenuation of these low energies in soft tissue limits the *in vivo* measurement of these elements to skin. One application of measuring iron, copper, and zinc in skin *in vivo* is to investigate the association of the concentration of these elements with skin lesions and disorders (60). *In vivo* measurement of iron can also be applied to monitor the efficacy of chelation of Fe, a therapy used to alleviate the tendency of Fe to build up as a result of treatment of the hereditary blood disorder β-thalassemia (61).

The first *in vivo* XRF system to measure iron, copper, and zinc in skin was developed by an Israeli research group (62,63), who used it in normal skin and in various non-malignant and malignant lesions (60). The system is based on a fine-focused X-ray tube with a gold target, and produces 11.4 keV X-rays emanated from a bent crystal monochromator (64). The spectrum is collected by a Si(Li) detector and analyzed by a conventional multichannel system. This technique was improved recently by Farquharson et al. (65–67), who placed different filters in front of an X-ray tube operating at 15 kVp and 23 mA, and optimized the X-ray tube output. The estimated MDLs for iron, zinc, and copper are 20, 10, and 5 μg/g skin tissue, respectively. The reported skin iron and zinc levels among the general population are comparable to MDLs, while the MDL for copper is much higher than the normal physiological skin level (67).

There are concerns about arsenic for both occupational and environmental exposures. In Bangladesh and West Bengal, thousands of people were exposed to water contaminated with high levels of arsenic and developed many types of skin diseases, including cancers. The biomarkers used for arsenic exposure assessment for epidemiology studies include hair and toenails, which are flawed with contamination. In addition, these are not target tissues for arsenic exposure. *In vivo* XRF measurement

of arsenic in skin is a promising method in that skin is a target organ for arsenic. Arsenic binds to the sulfhydryl groups through keratin in skin, and accumulates in the skin. Only one group of researchers studied the feasibility of measuring arsenic in skin *in vivo* by XRF (68–70). Different radioisotope sources and X-ray tube sources were explored with measurement of arsenic-doped skin-equivalent phantoms. MDL of 3.5 and 2.6 μg/g were obtained for $^{109}$Cd source-based and $^{125}$I source-based systems, respectively. The MDL was decreased to 0.7–5 μg/g (based on different skin arsenic distribution models) with an X-ray tube, which is sufficiently sensitive for *in vivo* measurement of arsenic in skin for a moderately or heavily exposed population, with a reported concentration of 0.3–24 μg/g in Bangladesh (71). The i*n vivo*$^{125}$I source-based system was used to measure arsenic in skin in a population with no known arsenic exposure (69). Other studies have used arsenic-doped phantoms. The estimated radiation doses delivered to the subjects are minimal for all these systems. Current studies indicate that XRF could be a valuable tool to assess arsenic in skin among exposed populations.

Currently, our group is in the process of developing a portable XRF device for skin and bone measurement of several elements (43). By combining the proper X-ray tube setting, filter selection and arrangement, and appropriate measurement geometry, this device could be adapted to the measurement of all these elements.

## 6.2 Strontium in Bone

Strontium is a bone-seeking element; over 99% of the body strontium burden is in bone (72). Administration of a small amount of strontium has been reported to have beneficial effects on bone (73,74). However, large amounts of strontium have been found to be detrimental to human health (75,76). The health effects of strontium are still controversial. Strontium has a relatively high concentration in bone, with an estimated level of 32 μg/g bone in Reference Man (77), and hence it is relatively easy to detect by XRF. Three studies have been performed on *in vivo* bone strontium measurement, the first by Wielopolski et al. (78). The XRF system involved was a $^{109}$Cd/$^{125}$I source-based system in a 90° geometry, and the strontium was measured in the tibia. The MDL achieved in this study was about 15 μg strontium/g bone. Later, Pejovic-Milic et al. reported an *in vivo* study on 10 environmentally exposed subjects with a similar system (79). The average MDL for this $^{109}$Cd-induced measurement system was 52 μg strontium/g bone. Five of the ten subjects had strontium concentrations lower than the MDL. Zamburlini et al. (80,81) improved the system with a $^{125}$I source and an optimized geometry, and obtained a detection limit of 5 μg strontium/g bone. The Zamburlini system was sensitive enough to detect the individual bone strontium in all 22 environmentally exposed subjects. The portable XRF system mentioned above is sensitive enough to obtain the strontium peak within 30 s (as opposed to 30 min for most of the *in vivo* measurement systems) with an acceptable radiation dose (43). The MDL of this portable system is still under investigation. The available data indicate that *in vivo* XRF bone strontium measurement is sufficiently sensitive for exposure assessment among the general population.

## 6.3 Silver in Skin

Health effects of occupational silver exposure include gastrointestinal irritation, renal and pulmonary lesions, and argyria and argyrosis, in which discoloration occurs in skin and in the eye (82). Argyria and argyrosis are of greatest concern. *In vivo* measurement of silver in skin, if proven feasible, will provide valuable information on the association between silver exposure and argyria, and possibly identify individuals at risk of the disease. There is only one study of *in vivo* measurement of silver in skin (83). This system is based on a $^{125}$I radioisotope source, and the measurements were taken on skin-mimic phantoms. The estimated MDL is 3–4 μg/g with 10–20 min measurement. There are little data to date regarding typical concentrations of silver in skin among exposed or non-exposed populations. Based on the available data, which report skin silver concentrations of 72 μg/g for an argyria patient (84) and a derived skin concentration of 1.4–43 μg/g after an acute inhalation of silver to induce argyria (83,85), the MDL with the current XRF system is encouraging for use among an exposed population.

## 6.4 Cd in Kidney and Bone

Kidney and liver are target organs with relatively high concentrations of cadmium, and contain over 50% of the body burden (86). Cadmium exposure is related to renal disease, bone disease, and can affect pulmonary function and neurological function (87). Over the past 30 years, both *in vivo* NAA and *in vivo* XRF techniques have been developed to non-invasively measure cadmium levels in the liver and kidney (88,89). Two systems were developed by the same group of scientists to measure cadmium in kidney with XRF. The first system generates Cd characteristic X-rays ($K_{\alpha 1,2} = 23.1$ keV) with an $^{241}$Am radioisotope source (90). The source is set up at the side of the person measured, and a Ge(Li) detector is located at the back of the subject. The MDL for the system is strongly influenced by the dorsal depth from the subject's back to the kidney surface and ranges from 20 to 40 μg/g for depths between 30 and 40 mm. A more sensitive system was developed a couple of years later (91) based on excitation of Cd by partially polarized X-ray beams from an X-ray therapy unit. The MDL varies between 3 and 20 μg/g for distances between the skin and the kidney surface of 30 to 70 mm. Borjesson et al. (92) proposed a system with an expected detection limit of 1 μg/g by optimizing the X-ray spectrum from a novel X-ray tube. A few studies have used XRF to examine the health effects of cadmium exposure (93). All studies focused on occupationally exposed workers except one (94), in which significant differences in cadmium concentrations in the kidneys were detected between non-occupationally exposed smokers and non-smokers. These results suggest that *in vivo* XRF measurement of cadmium in kidney is useful tool for cumulative dose assessment in occupationally exposed workers, but the technique still requires substantial improvement and validation for cumulative exposure assessment in the general population.

Bone is another target organ for Cd. Skeletal damage caused by chronic Cd exposure has been noted in humans (95) and in experimental animals (96), the

mechanism for which has been reviewed in several papers (97–99). The concentration of Cd in bone is relatively low even among exposed populations, with reported values of 0–1.5 μg/g (100,101). Two studies have been conducted to investigate the feasibility of measuring Cd in bone *in vivo* with XRF. Carew et al. (102) conducted a pilot study on the radioisotope source-based system and concluded that MDL with a $^{109}$Cd source in backscatter geometry was superior to several other possible geometries and excitation sources. Chettle (103) analyzed the capabilities of two HPGe detection systems and a lithium-drifted silicon (Si(Li)) system to detect bone Cd X-rays emitted through $^{109}$Cd sources and obtained an optimized MDL of 2.1 μg/g for 3 mm of overlaying soft tissue. Cadmium-doped standards were used in both studies, and no *in vivo* measurements were made. Current XRF techniques require improvement to assess bone cadmium in occupationally exposed populations.

## 6.5 Iodine in Thyroid

Iodine is as easy to detect as strontium due to its high concentration in thyroid. Knowing the amount of iodine stored in the thyroid may contribute to our understanding of thyroid diseases, the uptake and release of iodine from the thyroid, and possibly to the prediction of treatment outcome. The first *in vivo* iodine measurement in thyroid was reported in the 1960s (7). Several studies performed later used similar X-ray tube-based systems (104–106). One recent study reinforced the contention that XRF is a valuable method to determine the iodine content in thyroid (107). However, to improve accuracy, care must be taken to obtain precise knowledge of the measurement geometry, that is, the exact location of the thyroid and the measurement (108,109). All the described systems are sensitive enough to detect iodine in thyroid in the general population, though correction for geometry and attenuation still requires further investigation.

## 6.6 Platinum and Mercury in Kidney

People are exposed to platinum through mining and cytotoxic drugs (cisplatin and its transplatin analogues). Mercury exposure occurs in the general population from diet and dental amalgams. Occupational exposure is generally much higher, from sources such as mercury and gold mining, chloralkali and thermometer factories, and dental amalgam handling. Both platinum and mercury are toxic. Even low-level exposure can cause kidney disease and malfunction of other organs (110,111). The XRF measurement system for *in vivo* studies of platinum and mercury in the kidney is essentially that used for the measurement of cadmium in kidney (112). The average MDL of this system is 8 μg/g for platinum (113) and 26 μg/g for mercury (114,115). In Börjesson's study, half of the exposed workers had kidney mercury higher than the MDL, and none of the referents' values exceeded the individual MDL (114). The current technique for mercury is restricted to moderately and heavily exposed populations. It is a potentially valuable tool to assess platinum in tumors of patients treated with platinum-related cytotoxic drugs.

### 6.7 Uranium in Bone

Uranium, which accumulates in bone and kidneys, is a potential health hazard. Occupational exposure is the primary concern. The rise of the nuclear industry, which is increasing the demand for uranium mining, is expected to make the health effects of uranium a more significant issue. A $^{57}$Co-source-based XRF bone uranium measurement system was proposed by O'Meara et al. (116,117), whose system has an MDL of 20 μg uranium/g bone mineral. That MDL is more than three orders of magnitude greater that the 0.01 μg/g bone uranium level reported for the general population and more than one order of magnitude greater than the maximum of 2 μg/g for occupationally exposed populations (116). Current XRF techniques are not sensitive enough for measurements of even highly exposed subjects.

## 7 SUMMARY AND FUTURE DEVELOPMENT OF *IN VIVO* XRF ELEMENT EXPOSURE ASSESSMENT

Table 1 presents the characteristics of most XRF systems for *in vivo* element measurement. Among these techniques, measurements of iron and zinc in skin, strontium and lead in bone, and iodine in thyroid are applicable to general population; measurements of arsenic and silver in skin and cadmium, platinum, and mercury in kidney are applicable to moderately and highly exposed population; measurements of copper in skin and cadmium and uranium in bone demand further improvement.

Many techniques and methods can be applied to improve the *in vivo* XRF element measurement system. Advanced detectors and electronics can be adopted from other applications. Capillary optics can be used to focus the X-rays to a small beam with

**TABLE 1 Applications of *In vivo* Element Measurement**

| Element | Site | MDL (μg/g)* | Current Status |
|---|---|---|---|
| Iron | Skin | 20 | Applicable to general population |
| Copper | Skin | 5 | Demands further improvement |
| Zinc | Skin | 10 | Applicable to general population |
| Arsenic | Skin | 0.7–5 | Applicable to highly exposed population |
| Strontium | Bone | 5 | Applicable to general population |
| Silver | Skin | 3–4 | Applicable to highly exposed argyria patients |
| Cadmium | Kidney | 3–20 | Applicable to occupationally exposed population |
| | Bone | 2.1 | Demands further improvement |
| Iodine | Thyroid | 50 μg/mL | Applicable to general population |
| Platinum | Kidney | 8 | Applicable to patients treated with platinum related drugs |
| | Tumor | | |
| Mercury | Kidney | 26 | Applicable to moderately and heavily exposed population |
| Lead | Bone | 2 | Applicable to general population |
| Uranium | Bone | 20 | Demands further improvement |

*Latest MDL.

a relatively high intensity. For the X-ray tube source, the use of a different target or a secondary target can be explored, and different filters can be applied to optimize the X-ray source. Polarized X-rays can also be considered to reduce the background. Monte Carlo simulation can always be employed to determine the best set-up.

While technical improvement is an important issue, it is also critical to establish an effective connection between the technique developed and the end user. The advance of the technique and the direction of the improvement will be largely driven by the demand of the technique.

## 8 SYNCHROTRON RADIATION-INDUCED X-RAY FLUORESCENCE IN MEDICINE AND BIOLOGY

Another rapidly growing application of XRF in medicine and biology is microprobe imaging, also referred to as synchrotron microprobe imaging because the fluorescing source is produced by a synchrotron. The use of synchrotron radiation for XRF has several advantages over an X-ray tube or a radioisotope source (118), including: a much higher intensity, a polarized X-ray source, and a sub micro-scale source beam. The disadvantage is that the access to a synchrotron is limited.

### 8.1 Principles of Synchrotron XRF

The principles of synchrotron XRF are similar to those of XRF induced by conventional radioisotopes or X-ray tube sources. The main difference is that the X-ray source is generated by the acceleration of charged particles. Conventional synchrotron uses a ring magnet to guide the particles in an orbit, with a radio-frequency cavity to accelerate the beam in a periodic manner. Modern synchrotron utilizes a storage ring to accelerate the particles. The straight sections in the storage ring are used for inserting periodic magnetic devices (wigglers) that force the particles into a sinusoidal path. The particles are then further accelerated in a coherent fashion to produce high energy, high intensity X-rays. The size of the X-ray source is determined by the size of the particle bunch that is in orbit around the ring, and it can be minimized to a sub-micro-scale size using X-ray optics. The synchrotron is described in detail elsewhere (119). The synchrotron X-ray beam can be tailored to an intensive monoenergy bean by placing a monochromator in front of the source. The intensity of the X-ray source produced by the synchrotron is several orders higher than that of the highest-powered X-ray tubes, and hence synchrotron XRF is several orders more sensitive than conventional XRF. The X-ray is concentrated in the plane of the electron orbit and hence it is highly polarized in the same plane. The polarization of the X-ray beam further increases the sensitivity of the synchrotron XRF because of the reduced background from scattered coherent and incoherent radiation.

### 8.2 Element Mapping with Synchrotron XRF

The characteristics of the synchrotron beam make it a perfect choice to explore metal or trace element distributions. Its small beam size allows elements to be scanned at

a superior spatial resolution; its high intensity enable the mapping of some thousand points, collected within a reasonable time period. Two samples with synchrotron XRF are discussed below, representing application at micro- and cellular scales.

#### *8.2.1 Distribution of Metals and Trace Elements in Tissues*

Distribution of metals and trace elements in tissues can be explored with synchrotron XRF (120–123). In *in vivo* XRF measurement, the concentration of the element is calculated in a relatively large volume of tissue; synchrotron XRF tells you how the element is distributed. During the scanning process, an extra-thin slice of tissue sample (on the order of μm) is positioned in front of the source, and the source scans the slice in a 2D manner. The 2D composition of the elemental distribution is reconstructed with a specific software package. Depending on the source size and the demands of the application, the spatial resolution ranges from tens of μm to 1 mm. The tissues of interest include breast, bone, heart, brain, and other sites. With this technique, researchers have found elevated metal levels in breast cancer cells (124). Zoeger et al. (125) reported elevated lead and zinc concentrations at the tidemark of bone, where bone remodeling takes place. Szczerbowska (126) discovered differences in elemental composition of brain tissues between Parkinson's disease cases and a control group. The information obtained by synchrotron XRF element mapping in these areas provides valuable information in understanding the significant role metals and trace elements play in human health, whether beneficial or detrimental.

#### *8.2.2 Distribution of Metals and Trace Elements in Cells*

Synchrotron XRF microprobe imaging is now able to map the distribution of metals and trace elements in cells (127). Palmer et al. (128) studied zinc and calcium distribution in heart muscle cells. McRae et al. (129) investigated elemental maps for mouse fibroblast cells. Metal concentrations in metal-binding proteins were also explored (130). Though the number of studies in this area is small, this technique has great potential to become a valuable biotechnology tool, as summarized in Paunesku et al.'s (131) comprehensive review.

### 8.3 Future Development

Improvement of synchrotron XRF systems includes enhancement of the source, advancement of detectors and electronics, and extensive application of MC simulations. Establishment of a productive connection between the technique developed and the medical and biological community is also very important.

## ACKNOWLEDGEMENTS

The author is grateful to Dr. Lee Grodzins for correcting some of the physics concepts, providing valuable comments on the content, and contributing in Fig. 4 and formula in Appendix A. The author would also like to thank Paul Guttry for her comprehensive editing.

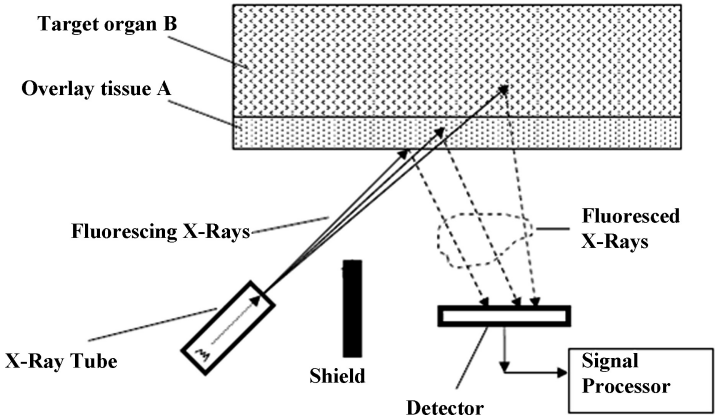


***FIGURE 4*** *A schematic diagram of an in vivo element measurement setup with X-ray tube.*

## APPENDIX A: CALCULATION OF INTENSITY OF CHARACTERISTIC X-RAYS

Figure 4 illustrates a setup for *in-vivo* element measurement with an X-ray tube. The target organ is B, and there is an overlaying tissue A on top of B. The following equations calculate the intensity of characteristic X-rays of element E from target organ B. A monoenergetic, incident X-ray or γ-ray beam is assumed; the fluorescing and fluoresced radiations enter and exit perpendicular to a uniform overlay tissue surface; geometrical and detection efficiencies are subsumed in the constant K. At thickness t of target organ B, the characteristic X-rays produced over thickness of d$t$ and collected by the detector is given by

$$\begin{aligned} \mathrm{d}(I_{\mathrm{E,X}}) = K \times I_0 \times \mathrm{e}^{-\mu_{\mathrm{in,A}} \times t_{\mathrm{A}}} \times \mathrm{e}^{-\mu_{\mathrm{in,B}} \times t} \times \mu_{\mathrm{PE}} \times Y \times f_{\mathrm{E}} \times b \times \mathrm{d}t \\ \times \rho_{\mathrm{B}} \times \mathrm{e}^{-\mu_{\mathrm{X,B}} \times t} \times \mathrm{e}^{-\mu_{\mathrm{x,A}} \times t_{\mathrm{A}}} \end{aligned} \tag{1}$$

Where $I_{\mathrm{E,X}}$ is the number of characteristic X-rays measured by a detector; $I_0$ is the photon flux ($\mathrm{cm}^{-2}$); $t_{\mathrm{A}}$ is the thickness of overlaying tissue A; $\mu_{\mathrm{in,A}}$ is linear attenuation for incident photons in tissue A; $\mu_{\mathrm{X,A}}$ is linear attenuation for characteristic X-rays in tissue A; $\mu_{\mathrm{in,B}}$ is linear attenuation for incident photons in target organ B; $\mu_{\mathrm{X,B}}$ is linear attenuation for characteristic X-rays in target organ B; $\mu_{\mathrm{PE}}$ is the cross-section for photoelectric interaction ($\mathrm{cm}^2$/g); Y is the yield of characteristic X-ray; $f_{\mathrm{E}}$ is the concentration of element E given in g/(g organ B); b is the irradiated area ($\mathrm{cm}^2$); $\rho_{\mathrm{B}}$ is the density of organ B. Integrate the above equation over the thickness of target organ B, $t_{\mathrm{B}}$

$$\int_0^{t_{\mathrm{B}}} \mathrm{d}(I_{\mathrm{E,X}}) = K \times I_0 \times \mathrm{e}^{-(\mu_{\mathrm{in,A}} + \mu_{\mathrm{X,A}}) \times t_{\mathrm{A}}} \times \mu_{\mathrm{PE}} \times f_{\mathrm{E}} \times b \times \rho_{\mathrm{B}} \times \int_0^{t_{\mathrm{B}}} \mathrm{e}^{-(\mu_{\mathrm{in,B}} + \mu_{\mathrm{X,B}}) \times t} \mathrm{d}t \tag{2}$$

Calculate the integration,

$$I_{E,X} = K \times I_0 \times f_E \times b \times \frac{\mu_{PE} \times \rho_B}{\mu_{in,B} + \mu_{X,B}} \times \left[1 - e^{-(\mu_{in,B} + \mu_{X,B}) \times t_B}\right] \times e^{-(\mu_{in,A} + \mu_{X,A}) \times t_A} \tag{3}$$

When the thickness of target organ B is sufficiently large, that is, the target organ B is "infinitely" thick, the number of the X-rays that can be measured by a detector is then given by

$$I_{E,X} = K \times I_0 \times f_E \times b \times \frac{\mu_{PE} \times \rho_B}{\mu_{in,B} + \mu_{X,B}} \times e^{-(\mu_{in,A} + \mu_{X,A}) \times t_A} \tag{4}$$

## REFERENCES


1. Hu H, Aro A, Rotnitzky A. Bone lead measured by X-ray fluorescence: epidemiologic methods. Environ Health Perspect 1995; 103 (suppl 1): 105–110 (Review).
2. Börjesson J, Mattsson S. Toxicology: *in vivo* X-ray fluorescence for the assessment of heavy metal concentrations in man. Appl Radiat Isot 1995; 46: 571–576 (Review).
3. Mattsson S, Thomas BJ. Development of methods for body composition studies. Phys Med Biol 2006; 51: R203–228 (Review).
4. Wedeen RP. Environmental renal disease: lead, cadmium and Balkan endemic nephropathy. Kidney Int Suppl 1991; 34: S4–S8 (Review).
5. Theodorakou C, Farquharson MJ. Human soft tissue analysis using X-ray or gamma-ray techniques. Phys Med Biol 2008; 53: R111–R149 (Review).
6. Börjesson J, Isaksson M, Mattsson S. X-ray fluorescence analysis in medical sciences: a review. Acta Diabetol 2003; 40 (suppl 1): S39–S44.
7. Hoffer PB, Jones WB, Crawford RB, Beck R, Gottschalk A. Fluorescent thyroid scanning: a new method of imaging the thyroid. Radiology 1968; 90: 342–344.
8. Grönberg T, Sjöberg S, Almen T, Golman K, Mattsson S. Noninvasive estimation of kidney function by X-ray fluorescence analysis: elimination rate and clearance of contrast media injected for urography in man. Invest Radiol 1983; 18: 445–452.
9. Ahlgren L, Liden K, Mattsson S, Tejning S. X-ray fluorescence analysis of lead in human skeleton *in vivo*. Scand J Work Environ Health 1976; 2: 82–86.
10. Hu H, Aro A, Payton M, Korrick S, Sparrow D, Weiss ST, Rotnitzky A. The relationship of bone and blood lead to hypertension: the Normative Aging Study. JAAA 1996; 275: 1171–1176.
11. Korrick SA, Hunter DJ, Rotnitzky A, Hu H, Speizer FE. Lead and hypertension in a sample of middle-aged women. Am J Public Health 1999; 89: 330–335.
12. Cheng Y, Schwartz J, Sparrow D, Aro A, Weiss ST, Hu H. Bone lead and blood lead levels in relation to baseline blood pressure and the prospective development of hypertension: the Normative Aging Study. Am J Epidemiol 2001; 152: 164–171.

13. Glenn BS, Bandeen-Roche K, Lee BK, Weaver VM, Todd AC, Schwartz BS. Changes in systolic blood pressure associated with lead in blood and bone. Epidemiology 2006; 17: 538–544.
14. Shih RA, Glass TA, Bandeen-Roche K, Carlson MC, Bolla KI, Todd AC, Schwartz BS. Environmental lead exposure and cognitive function in community-dwelling older adults. Neurology 2006; 67: 1556–1562.
15. Wu MT, Kelsey K, Schwartz J, Sparrow D, Weiss S, Hu H. A delta-aminolevulinic acid dehydratase (ALAD) polymorphism may modify the relationship of low-level lead exposure to uricemia and renal function: the Normative Aging Study. Environ Health Perspect 2003; 111: 335–341.
16. Paunesku T, Vogt S, Maser J, Lai B, Woloschak G. X-ray fluorescence microprobe imaging in biology and medicine. J Cell Biochem 2006; 99: 1489–1502 (Review).
17. Lobinski R, Moulin C, Ortega R. Imaging and speciation of trace elements in biological environment. Biochimie 2006; 88: 1591–1604 (Review).
18. Fahrni CJ. Biological application of X-ray fluorescence microscopy: exploring the subcellular tomography and speciation of transition metals. Curr Opin Chem Biol 2007; 11: 121–127 (Review).
19. Zanini F, Lausi A, Savoia A. The beam line of ELETTRA and their application to structural biology. Gnetica 1999; 106: 171–180.
20. ICRP 60. Recommendations of international commission on radiological protection. Oxford: Pergamon Press; 1990.
21. Knoll GF. Radiation detection and measurement. 3rd ed. New Jersey: John Wiley and Sons Inc; 1999. p 48–53.
22. Mori I, Takayama T, Motomura N. The CdTe detector module and its imaging performance. Ann Nucl Med 2001; 15: 487–494.
23. Guerra P, Santos A, Darambara DG. Development of a simplified simulation model for performance characterization of a pixellated CdZnTe multimodality imaging system. Phys Med Biol 2008; 53: 1099–1113.
24. Leclair RJ, Wang Y, Zhao P, Boileau M, Wang L, Fleurot F. An analytic model for the response of a CZT detector in diagnostic energy dispersive X-ray spectroscopy. Med Phys 2006; 33: 1329–1337.
25. Mitra S, Wielopolski L, Hendrey G. Coomparison of a digital and an analog signal processing system for neutron inelastic gamma-ray spectrometry. Appl Radiat Isot 2004; 61: 1463–1468.
26. Reguigui N, Morel J, Kraiem HB, Mahjoub A. Characterization of three digital signal processor systems used in gamma ray spectrometry. Appl Radiat Isot 2002; 56: 93–97.
27. Bevington PR, Robinson DK. Data reduction and error analysis. 3rd ed. New York: McGraw-Hill Companies, Inc; 2003.
28. Luo LQ, Chettle DR, Nie H, McNeill FE, Popovic M. Curve fitting using a genetic algorithm for the X-ray fluorescence measurement of lead in bone. J Radioanal Nucl Chem 2006; 269: 325–329.
29. Aro A, Amarasiriwardena C, Lee ML, Kim R, Hu H. Validation of K X-ray fluorescence bone lead measurements by inductively plasma mass spectrometry in cadaver legs. Med Phys 2000; 27: 119–123.

30. Todd AC, Moshier EL, Carroll S, Casteel SW. Validation of X-ray fluorescence-measured swine femur lead against atomic absorption spectrometry. Environ Health Perspect 2001; 109: 1115–1119.
31. Browne E, Doebler RE, Dairiki JM, Shihab-Eldin A, Jardine LJ, Tuli JK, Buyrn AB. In: Lederer CM, Shirley VS,editors. Tables of isotopes. 7th ed. New York: John Wiley and Sons, Inc.; 1978.
32. Wielopolski L, Slatkin DN, Vartsky D, Ellis KJ, Cohn SH. Feasibility study for the *in vivo* measurement of lead in bone using L-X-ray fluorescence. ISEE Trans Nucl Sci NS28: 1981; 114–116.
33. Wielopolski L, Rosen JF, Slatkin DN, Vartsky D, Ellis KJ, Cohn SH. Feasibility of noninvasive analysis of lead in the human tibia by soft X-ray fluorescence. Med Phys 1983; 10: 248–251.
34. Wielopolski L, Ellis KJ, Vaswani AN, Cohn SH, Greenberg A, Puschett JB, Parkinson DK, Fetterolf DE, Landrigan PJ. *In vivo* bone lead measurements: a rapid monitoring method for cumulative lead exposure. Am J Ind Med 1986; 9: 221–226.
35. Wielopolski L, Slatkin DN, Kalef-ezra JA, Rosen JF. Non-invasive detection of tibia bone lead in intact legs by L X-ray fluorescence. The Toxicologist Abstract 1987; 77.
36. Weiloposki L, Rosen JF, Slatkin DN, Zhang R, Kalef-ezra JA, Rothman JC, Maryanski M, Jenks ST. *In vivo* measurement of cortical bone using polarized radiation. Med Phys 1989; 16: 521–528.
37. Todd AC. L-shell X-ray fluorescence measurement of lead in bone: system development. Phys Med Biol 2002; 47: 507–522.
38. Todd AC. L-shell X-ray fluorescence measurement of lead in bone: theoretical consideration. Phys Med Biol 2002; 47: 491–505.
39. Todd AC. L-shell X-ray fluorescence measurement of lead in bone: accuracy and precision. Phys Med Biol 2002; 47: 1399–1419.
40. Ao Q, Lee SH, Gardner RP. Optimization of *in vivo* X-ray fluorescence analysis methods for bone lead by simulation with the Monte-Carlo code CEARXRF. Appl Radiat Isot 1997; 48: 1413–1423.
41. Lee SH, Gardner RP, Todd AC. Preliminary studies on combining the K and L XRF methods for *in vivo* bone lead measurement. Appl Radiat Isot 2001; 54: 893–904.
42. Guo W, Gardner RP, Metwally WA. Preliminary studies on K and L coincidence spectroscopy for optimizing the *in vivo* XRF measurement of lead in bone. Nucl Instrum Methods Phys Res B 2004; 213: 574–578.
43. Nie H, Sanchez S, Grodzins L, Cleveland R, Weisskopf M. Feasibility and Methodology of the Development of a Portable *In Vivo* X-ray Fluorescence (XRF) Bone Lead Measurement System. Abstract. Conference Proceeding for the 49th Annual Meeting of Society of Toxicology 2010; p 271.
44. Somervaille LJ, Laird EE, Chettle DR, Scott MC. Lead body stores assessed *in vivo* by X-ray fluorescence. International Conference on Heavy Metals in the Environment 1, CEP Consultants, Edinburgh; 1983. p 521–524.
45. Somervaille LJ, Chettle DR, Scott MC. *In vivo* measurement of lead in bone using X-ray fluorescence. Phys Med Biol 1985; 30: 929–943.
46. Gordon CL, Chettle DR, Webber CE. An improved instrument for the *in vivo* detection of lead in bone. Br J Ind Med 1993; 50: 637–641.

47. Nie H, Chettle DR, Stronach IM, Arnold ML, Huang SB, McNeill FE, O'Meara J. A study of MDL improvement for the *in vivo* measurement of lead in bone. Nucl Instrum Methods Phys Res B 2004; 213: 579–583.
48. Nie H, Chettle DR, Luo LQ, O'Meara JM. *In vivo* investigation of a new $^{109}$Cd gamma-ray induced K-XRF bone lead measurement system. Phys Med Biol 2006; 51: 351–360.
49. Todd AC, Landrigan PJ. X-ray fluorescence analysis of lead in bone. Environ Health Perspect 1993; 101: 494–495.
50. Chettle DR, Arnold ML, Aro ACA, Fleming DEB, Kondrashov VS, McNeill FE, Moshier EL, Nie H, Rothenberg SJ, Stonach IM, Todd AC. An agreed statement on calculating lead concentration and uncertainty in XRF *in vivo* bone lead analysis. Appl Radiat Isot 2003; 58: 603–605.
51. Zaidi H, Xu XG. Computational anthropomorphic models of the human anatomy: the path to realistic Monte Carlo modeling in radiological sciences. Annu Rev Biomed Eng 2007; 9: 471–500 (Review).
52. Rogers DW. Fifty years of Monte Carlo simulations for medical physics. Phys Med Biol 2006; 51: 287–301 (Review).
53. Verhaegen F, Seuntjens J. Monte Carlo modeling of external radiotherapy photon beams. Phys Med Biol 2003; 48: 107–164 (Review).
54. O'Meara JM, Chettle DR, McNeill FE, Prestwich WV, Svensson CE. Monte Carlo simulation of source-excited *in vivo* X-ray fluorescence measurement of heavy metals. Phys Med Biol 1998; 43: 1413–1428.
55. Todd AC, McNeill FE, Palethorpe JE, Peach DE, Chettle DR, Tobin MJ, Strosko SJ, Rosen JC. *In vivo* X-ray fluorescence of lead in bone using K X-ray excitation with $^{109}$Cd sources: radiation dosimetry studies. Environ Res 1992; 57: 117–132.
56. Nie H, Chettle DR, Luo LQ, O'Meara JM. Dosimetry study for a new in-vivo X-ray fluorescence (XRF) bone lead measurement system. Nucl Instrum Methods Phys Res B 2007; 263: 225–230.
57. Schwartz BS, Hu H. Adult lead exposure: time for change. Environ Health Perspect 2007; 115: 451–454 (Review).
58. Hu H, Shih R, Rothenberg S, Schwartz BS. The epidemiology of lead toxicity in adults: measuring dose and consideration of other methodologic issues. Environ Health Perspect 2007; 115: 455–462.
59. Navas-Acien A, Guallar E, Silkergeld EK, Rothenberg SJ. Lead exposure and cardiovascular disease—a systematic review. Environ Health Perspect 2007; 115: 472–482.
60. Gorodetsky R, Sheskin J, Weinreb A. Iron, copper, and zinc concentration in normal skin and in various non malignant and malignant lesions. Int J Dermatol 1986; 25: 440–445.
61. Gorodetsky R, Goldfarb A, Dagan I, Rachmilewitz EA. Noninvasive analysis of skin iron and zinc levels in beta-thalassemia major and intermedia. J Lab Clin Med 1985; 105: 44–51.
62. Sheskin J, Zeimer R. *In vivo* Study of trace elements in leprous skin. Int J Dermatol 1977; 16: 745–747.
63. Zeimer R, Belkin M, Leitersdorff E, Rachmilewitz EA. A noninvasive method for the evaluation of tissue iron deposition in beta-thalassemia major. J Lab Clin Med 1978; 91: 24–31.
64. Sheskin J, Gorodetsky R, Loewinger E, Weinreb A. In-vivo measurement of iron, copper and zinc in the skin of prurigo nodularis patients treated with thalidomide. Dematologica 1981; 162: 86–90.

65. Farquharson MJ, Bradley DA. The feasibility of a sensitive low-dose method for the *in vivo* evaluation of Fe in skin using K-shell X-ray fluorescence (XRF). Phys Med Biol 1999; 44: 955–965.
66. Bradley DA, Farquharson JJ. X-ray fluorescence and the *in vivo* evaluation of Fe, Cu, and Zn in skin. J Radioanal Nucl Chem 1999; 224: 213–217.
67. Farquharson MJ, Bagshaw AP, Porter JB, Abeysinghe RD. The use of skin Fe levels as a surrogate marker for organ Fe levels, to monitor treatment in cases of iron overload. Phys Med Biol 2000; 45: 1387–1396.
68. Studinski RC, McNeill FE, O'Meara JM, Chettle DR. A method detectionlimit for potential *in vivo* arsenic measurements with a 50W X-ray tube. Phys Med Biol 2006; 51: N381–N387.
69. Studinski RC, McNeill FE, Chettle DR, O'Meara JM. Estimation of a method detection limit for an *in vivo* XRF arsenic detection system. Phys Med Biol 2005; 50: 521–530.
70. Studinski RC, O'Meara JM, McNeill FE. The feasibility of *in vivo* measurement of arsenic and silver by X-ray fluorescence. X-ray Spectrometry 2008; 37: 51–57.
71. Karim MM. Arsenic in groundwater and health problems in Bangladesh. Wat Res 2000; 34: 304–310.
72. Schroeder HA, Tipton IH, Nason AP. Trace metals in man: strontium and barium. J Chronic Dis 1972; 25: 491–517.
73. Marie PJ, Garba MT, Hott M, Miravet L. Effect of low doses of stable strontium on bone metabolism in rats. Miner Electrolyte Metab 1985; 11: 5–13.
74. Grynpas MD, Hamilton E, Cheung R, Tsouderos Y, Deloffre P, Hott M, Marie PJ. Strontium increases vertebral bone volume in rats at a low dose that does not include detectable mineralization defect. Bone 1996; 18: 253–259.
75. Oxgür S, Sümer H, Koçoğlu G. Rickets and soil strontium. Arch Dis Child 1996; 75: 524–526.
76. Cohen-Solal M. Strontium overload and toxicity: impact on rneal osteodystrophy. Nephrol Dial Transplant 2002; 17 (suppl l2): 30–34 (Review).
77. ICRP23. Report of the task group on reference man. Oxford: Pergamon Press; 1975.
78. Wielopolsk L, Vartsky D, Yasumura S, Cohn SH. Application of XRF to measure strontium in human-bone. Adv X-ray Anal 1983; 26: 415–421.
79. Pejović-Milić A, Stronach IM, Gyorffy J, Webber CE, Chettle DR. Quantification of bone strontium levels in humans by *in vivo* X-ray fluorescence. Med Phys 2004; 31: 528–538.
80. Zamburlini M, Pejović-Milić A, Chetle DR, Webber CE, Gyorffy Y. *In vivo* study of an X-ray fluorescence system to detect bone strontium non-invasively. Phys Med Biol 2007; 52: 2107–2122.
81. Zamburlini M, Pejović-Milić A, Chettle DR. Spectrometry methods for *in vivo* bone strontium measurements. X-ray spectrometry 2008; 37: 42–50.
82. World Health Organization. Summary of toxicological data of certain food additives, No 12: silver. Geneva: WHO; 1977.
83. Graham SA, O'Meara JM. The feasibility of measuring silver concentrations *in vivo* with X-ray fluorescence. Phys Med Biol 2004; 49: N259–266.
84. East BW, Boddy K, Williams ED, Macintyre D, Mclay ALC. Silver retention, total body silver and tissue silver concentrations in argyria associated with exposure to an anti-smoking remedy containing silver acetate. Clin Exp Dermatol 1980; 5: 305–311.

85. Armitage SA, White MA, Wilson HK. The determination of silver in whole blood and its application to biological monitoring of occupationally exposed groups. Ann Occup Hyg 1996; 40: 331–338.
86. World Health Organization, Cadmium. Environmental health criteria 134. Genera: WHO; 1992.
87. Nordberg GF, Herber RFM, Alessio L, editors. Cadmium in the human environment: toxicity and carcinogenicity. New York: Oxford University Press; 1992.
88. Börjesson J, Mattsson S. Toxicology: *in vivo* X-ray fluorescence for the assessment of heavy metals concentrations in man. Appl Radiat Isot 1995; 46: 571–576.
89. Nilsson U, Skerfving S. *In vivo* X-ray fluorescence measurement of cadmium and lead. Scand J Work Environ Health 1993; 19 (suppl 1): 54–58.
90. Ahlgren L, Mattsson S. Cadmium in man measured *in vivo* by X-ray fluorescence analysis. Phys Med Biol 1981; 26: 19–26.
91. Christoffersson JO, Mattson S. Polarized X-rays in XRF-analysis for improved *in vivo* detectability of cadmium in man. Phys Med Biol 1983; 28: 1135–1144.
92. Börjesson J, Olsson M, Mattson S. Feasibility of a fluorescent X-ray source for *in vivo* X-ray fluorescence measurements of kidney and liver cadmium. Ann N Y Acad Sci 2000; 904: 255–258.
93. Gerhardsson L, Börjesson J, Grubb A, Hultberg B, Mattson S, Schutz A, Skerfving S. *In vivo* XRF as a means to evaluate the risk of kidney effects in lead and cadmium exposed smelter workers. Appl Radiat Isot 1998; 49: 711–712.
94. Nilsson U, Schütz A, Skerfving S, Mattson S. Cadmium in kidneys in swedes measured *in vivo* using X-ray fluorescence analysis. Int Arch Occup Environ Health 1995; 67: 405–411.
95. Alfven T, Elinder CG, Carlsson MD, Grubb A, Hellstrom L, Persson B, Pettersson C, Spang G, Schütz A, Jarup L. Low-level cadmium exposure and oseoporosis. J Bone Miner Res 2000; 15: 1579–1586.
96. Brzoska MM, Moniuszko-Jakoniuk J. Disorders in bone metabolism of female rats chronically exposed to cadmium. Toxic Appl Pharmacol 2005; 202: 68–83.
97. Wilson AK, Bhattacharyya MH. Effects of cadmium on bone: an *in vivo* model of the early response. Toxic Appl Toxicol 1997; 145: 68–73.
98. Bertin G, Averberk D. Cadmium: cellular effects, modifications of biomoldecules, modulation of DNA repair, and genotoxic consequences. Biochimie 2006; 88: 1549–1559 (Review).
99. Kazantzis G. Cadmium, osteoporosis, and cadmium metabolism. Biometals 2004; 17: 493–498.
100. Baranowska I, Czernicki K, Aleksandrowicz R. The analysis of lead, cadmium, zinc, copper and nickel content in human bones from the upper Silesian industrial district. Sci Total Environ 1995; 159: 155–162.
101. D'Haese PC, Couttenve MM, Lamberts LV, Elseviers MM, Goodman WG, Schrooten I, Cabrera WE, De Broe ME. Aluminum, iron, lead, cadmium, copper, zinc, chromium, magnesium, strontium, and calcium content in bone of end-stage renal failure patients. Clin Chem 1999; 45: 1548–1556.
102. Carew SE, Gastaldo J, Roels HA, O'Meara JM, Chettle DR. Development of a K-shell X-ray fluorescence measurement of cadmium in bone. X-ray Spectrom 2005; 34: 498–501.
103. Chettle D. Private Communication.

104. Aubert B, Fragu P, Di Paola M, Rougier P, Tubiana M. Application of X-ray fluorescence to the study of iodine distribution and content in the thyroid. Eur J Nucl Med 1981; 6: 407–410.

105. Jonckheer MH. Clinical usefulness of thyroid imaging by means of X-ray fluorescence. Horm Res 1987; 26: 42–47 (Review).

106. Jonckheer MH, Deconinck F. X-ray-fluorescence determination of stable iodine in the thyroid gland. A review Acta Clin Belg 1982; 37: 30–40.

107. Milakovic M, Berg G, Eggertsen R, Nystrom E, Olsson A, Larsson A, Hansson M. Determination of intra thyroidal iodine by X-ray fluorescence analysis in 60- to 65- year olds living in an iodine-sufficient area. J Intern Med 2006; 260: 69–75.

108. Hansson M, Isaksson M. A Monte Carlo (MC) based individual calibration method for *in vivo* X-ray fluorescence analysis (XRF). Phys Med Biol 2007; 52: 2009–2019.

109. Hansson M, Berg G, Isaksson M. *In vivo* X-ray fluorescence analysis (XRF) of the thyroid iodine content—influence of measurement geometries on the iodine $K_\alpha$ signal. X-ray Spectrom 2008; 37: 37–41.

110. World Health Organization. Environmental health criteria 125, platinum. Geneva: WHO; 1991.

111. Clement Associates and U.S. Environmental Protection Agency. Toxicological profile for mercury. Agency for toxic substances and disease registry. Georgia: U.S. Public Health Service; 1989.

112. Börjesson J, Mattsson S. *In vivo* X-ray fluorescence measurements of lead, cadmium and mercury in occupational and environmental studies: a review of work conducted in Sweden 1970–2005. X-ray Spectrom 2008; 37: 58–68.

113. Jonson R, Mattsson S, Unsgaard B. A method for *in vivo* analysis of platinum after chemotherapy with cisplatin. Phys Med Biol 1988; 33: 847–857.

114. Börjesson J, Barregård Sallsten G, Schütz A, Jonson R, Alpsten M, Mattsson S. *In vivo* analysis of mercury: the relation between concentrations in the kidney and the urine. Phys Med Biol 1995; 40: 413–426.

115. O'Meara JM, Börjesson J, Chettle DR. Improving the *in vivo* X-ray fluorescence (XRF) measurement of renal mercury. Appl Raiat Isot 2000; 53: 639–646.

116. O'Meara JM, Chettle DR, McNeill FE, Webber CE. The feasibility of measuring bone uranium concentrations *in vivo* using source excited K X-ray fluorescence. Phys Med Biol 1997; 42: 1109–1120.

117. O'Meara JM, Chettle DR, McNeill FE, Webber CE. *In vivo* X-ray fluorescence (XRF) measurement of uranium in bone. Appl Radiat Isot 1998; 49: 713–715.

118. Jones KW, Gordon BM, Hanson AL, Kwiatek WM, Pounds JG. X-ray fluorescence with synchrotron radiation. Ultramicroscopy 1988; 24: 313–328.

119. Hofmann A. The physics of synchrotron radiation. New York: Cambridge University Press; 2004.

120. Zoeger N, Roschger P, Hofstaetter JG, Jokubonis C, Pepponi G, Falkenberg G, Fratzl P, Berzlanovich A, Osterode W, Streli C, Wobrauschek P. Lead accumulation I tidemark of articular cartilage. Osteoarthr Cartil 2006; 14: 906–913.

121. Vincze L, Vekemans B, Brenker FE, Falkenberg G, Rickers K, Somogyi A, Kersten M, Adams F. Three-dimensional trace element analysis by confocal X-ray microfluorescence imaging. Anal Chem 2004; 76: 6786–6791.

122. Geraki K, Farquharson MJ, Bradley DA. Concentrations of Fe, Cu and Zn in breast tissue: a synchrotron XRF study. Phys Med Biol 2002; 47: 2327–2339.

123. Geraki K, Farquharson MJ, Bradley DA. X-ray fluorescence and energy dispersive X-ray diffraction for the quantification of elemental concentrations in breast tissue. Phys Med Biol 2004; 49: 99–110.

124. Geraki K, Farquharson MJ, Bradley DA, Gundogdu O, Falkenberg G. The localization of biologically important metals in soft and calcified tissues using a synchrotron X-ray fluorescence technique. X-ray Spectrom 2008; 37: 12–20.

125. Zoeger N, Streli C, Wobrauschek P, Jokubonis C, Pepponi G, Roschger P, Hofstaetter J, Berzlanovich A, Wegrzynek D, Chinea-Cano E, Markowicz A, Simon R, Falkenberg G. Determination of the elemental distribution in human joint bones by SR micro XRF. X-ray Spectrom 2008; 37: 3–11.

126. Szczerbowska-Boruchowska M. X-ray fluorescence spectrometry, an analytical tool in neurochemical research. X-ray Spectrom 2008; 37: 21–31.

127. Miller L, Wang Q, Smith RJ, Zhong H, Elliott D, Warren J. A new sample substrate for imaging and correlating organic and trace metal composition in biological cells and tissues. Anal Bioanal Chem 2007; 387: 1705–1715.

128. Palmer BM, Vogt S, Chen Z, Lachapelle RR, Lewinter MM. Intracellular distributions of essential elements in cardiomyocytes. J Struct Biol 2006; 155: 12–21.

129. McRae R, Lai B, Vogt S, Fahrni CJ. Correlative microXRF and optical immunofluorescence microscopy of adherent cells labeled with ultrasmall gold particles. J Struct Biol 2006; 155: 22–29.

130. Verbi FM, Arruda SC, Rodriguez AP, Pérez CA, Arruda MA. Metal-binding proteins scanning and determination by combining gel electrophoresis, synchrotron radiation X-ray fluorescence and atomic spectrometry. J Biochem Biophys Methods 2005; 62: 97–109.

131. Paunesku T, Vogt S, Maser J, Lai B, Woloschak G. X-ray fluorescence microprobe imaging in biology and medicine. J Cell Biochem 2006; 99: 1489–1502.

# *Statistical Tools*

**Antje Christensen and Sille Esbjerg**
*Novo Nordisk A/S, Bagsværd, Denmark*

## 1 INTRODUCTION

### 1.1 Examples

*Example 1-1:* During a fermentation process, a sensor registers temperature measurements at regular time intervals, resulting in a large dataset for each fermentation batch. A process chemist who is new to the factory is given data from a couple of batches to familiarize himself with the way the process behaves with regard to temperature. Looking down the long list of data is not helpful, he needs some tools.

*Example 1-2:* A pharmaceutical company is considering two types of tanks for their fermentation process. It is believed that the yield from one tank is higher than from the other, but the company would like to investigate this before deciding which tank to use.

*Example 1-3*: Measurements of assay from two stability studies, each consisting of three batches, are analyzed together. The analysis is supposed to show how an assay develops over time, if the development is the same for batches in the same study or, if possible, the same in all batches.

*Example 1-4:* In a production process, one wants to raise the yield. There are a number of factors that may influence the yield, such as holding time, feeding speed, and various conditions of the raw materials. Which of these actually has an influence, and how large the influence is, has not been established. While a lot

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing,*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

of personal opinions exist among the process operators, there are few data outside the current set-point.

*Common background for the examples:* All of these examples call for a tool from the statistical toolbox, be it to summarize and graphically depict data, or be it to test a hypothesis, build a model, or design an experiment.

### 1.2 Introduction of the Tool

To the academic statistician, statistics is a branch of mathematics. To the practitioner, statistics is a tool to better understand the numerical data that are abundant in a modern production environment. This chapter is written for the practitioner. It is an introduction to a useful toolbox, lining up the different tools side by side. If you have an overview of the toolbox, you avoid resembling the man with the hammer to whom everything looks like a nail. Instead, there are a number of different tools for different situations. Sometimes, different tools can be used for the same situation. There is no absolute right and wrong in the choice. In general, statistics is not about absolute rights and wrongs. It is a set of tools to help you distinguish the probable from the improbable and find the most compelling explanation for your observations.

### 1.3 Examples Continued

*Example 1-1:* Descriptive statistics, like mean and standard deviation, and graphs are helpful in this situation. See the paragraphs on these two subjects.

*Example 1-2:* See the paragraph on hypothesis testing.

*Example 1-3:* This example calls for a model of the data. See the paragraph on mixed models.

*Example 1-4:* Statistics does not start with the data. If you plan to collect data systematically through experiments, it makes sense to use a statistical tool to set up your experiments in the way that gives you the most information for the data collection effort. See the paragraph on design of experiments.

## 2 TYPES OF DATA

### 2.1 Examples

*Example 2-1:* Measurements of strength of insulin in insulin vials are made as part of the quality control before release of the batch.

*Example 2-2:* Switches used in the production of a medical device are examined by quality control and sorted into good items and scrap.

*Example 2-3:* Subjects entering a clinical trial have their body mass index (BMI) measured. However, the exact BMI is not registered, only the categories obese ($\text{BMI} > 30$), overweight ($25 < \text{BMI} \leq 30$), normal ($17.5 < \text{BMI} \leq 25$), and underweight ($\text{BMI} \leq 17.5$). In the study the ethnicity of the subject is also registered. It can be "Caucasian," "African," "Asian," or "Other."

## 2.2 Introduction of the Tool

There are three types of data: continuous data, discrete data, and categorical data. Continuous data are data measured on a continuous scale. The strength measurement in Example 2-1 is an example of continuous data. Discrete data can assume all integer values ($\ldots, -1, 0, 1, 2, \ldots$). Both continuous and discrete data can be placed in an infinite number of categories. Categorical data, on the other hand, can only be placed in a finite number of categories. Categorical data can be ordered or unordered. The BMI groups in Example 2-3 are ordered data since a person in the overweight group is known to have a larger BMI than those in the normal group and a smaller BMI than those in the obese group. The grouping of the subjects by their ethnicity is obviously unordered. When there are only two categories, as in Example 2.2, the notion of order is meaningless. Sometimes continuous data are rounded so much that they appear to be categorical. If that is the case one should try and get some more decimals; otherwise, the data have to be analyzed as categorical. It is important do determine which type of data one is working with since the type is key to the distribution that could describe data and which kind of analysis one should make. For instance if one has height measurements for children as the variable of interest and age measurements as the explanatory variable then dependent on the age measurements, one could make different kinds of analyses. If the age measurements were only in two or three categories, a one-sided ANOVA (ANalysis Of VAriance) model could be fitted, and if the age measurements varied more, e.g., if they were continuous, a linear regression could be made.

# 3 GRAPHS

## 3.1 Examples

*Example 3-1:* From a fermentation process, data on oxygen and sugar consumption are recorded at regular time intervals. One wants to detect patterns in the data over time and understand how the consumption data relate to process parameters such as temperature and pH, which can be adjusted.

*Example 3-2:* In a clinical study, a number of data are obtained from patients, including HbA1c, average fasting glucose, and average post-prandial glucose before and after treatment. The patients have received different doses of the medication. For each subject, a number of background variables have been recorded, e.g., gender, age, BMI, and time since diagnosis.

*Example 3-3:* A screening test for cancer is offered countrywide, but the number of people accepting the offer differs widely between different counties. In a survey, the number of tests conducted per year is determined and related to the number of eligible inhabitants for each county. One wants to understand the patterns of the variation and hopes to discover a reason for the different acceptance of the test.

*Common background for the examples:* In all examples, one approaches the data with a desire to explore and understand whatever is in the data. This can be a situation

where one does not have a specific question to the data that warrants a test. But also when one has planned a test, it makes sense to start with a graph to get a feeling for the data. Once knowledge of the structure in the data has been obtained, it needs to be conveyed to other people who have not participated in the analysis and wish to quickly and easily gain an understanding of the results.

## 3.2 Introduction of the Tool

As in other fields of exploration, one of the most important things one can do is to get to the scene and see for oneself. To see your data, you need to arrange them graphically. There are a wide range of graphs that will highlight different aspects of the data. You can use them to scan groups of data, rather than single observations at a time. Graphs are also a strong tool for communicating your findings, as they typically do not require any familiarity with statistical procedures.

## 3.3 Types of Graphs

***3.3.1 Time Plots*** Patterns occur often over time. If your data have an inherent time order, plot them in this order. If the time points of data collection are unequally spaced, reflect this spacing on your time axis. For continuous data, use dots, possibly connected (see Fig. 1). For categorical data, you can use stacks of colored bars that reflect either the number of observations or the percentage of observations in each category (see Fig. 2). You can include several series of data, e.g., the oxygen and sugar consumption from Example 3-1, in the same time plot to see whether the variables' variation over time is related.

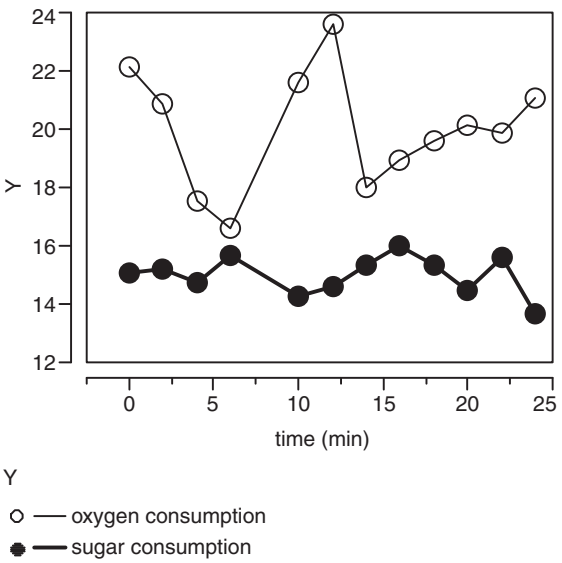


***FIGURE 1*** *Time plot for numerical data.*

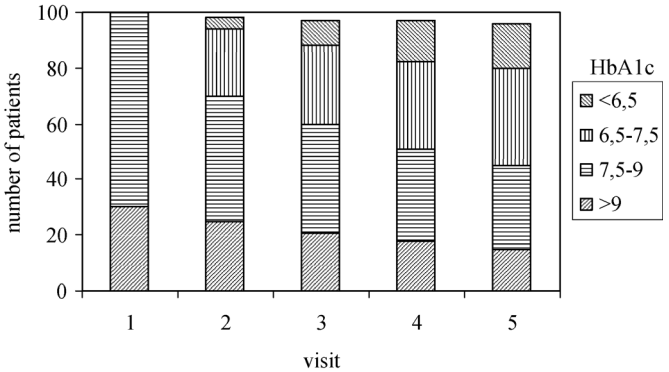


***FIGURE 2*** *Time plot for percentages.*

***3.3.2 Scatter Plots*** To explore the relationship between pairs of continuous data, plot them as dots in a coordinate system with one variable on the x-axis and the other on the y-axis (see Fig. 3). You can then nicely see correlations. If time may play a role, you can connect the dots in time order to see the process wander through the plane that the variables span. You can also construct scatter plots with three axes for three variables, but these can be difficult to interpret on a two-dimensional computer screen or piece of paper.

***3.3.3 Group Plots*** If your data belong to different categories, e.g., male-female, or fermentor A, B, and C, and are on a continuous scale, draw them side by side (see Fig. 4). You can include lines indicating the averages, ranges, standard deviations, etc. You can then easily see whether, e.g., one category stands out from the others, or the categories fall into groups that are similar.

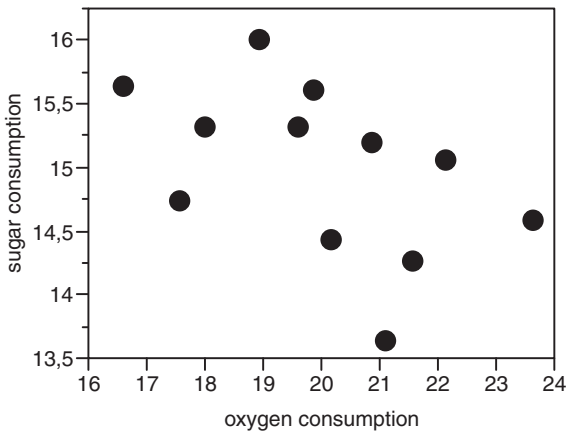


***FIGURE 3*** *Scatter plot.*

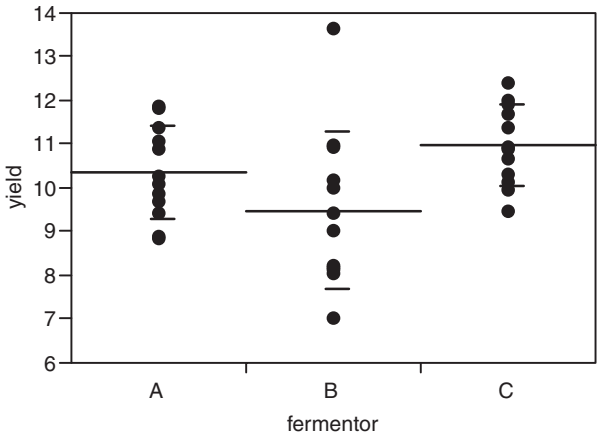


***FIGURE 4*** *Group plot.*

***3.3.4 Bar Charts*** For categorical data belonging to different categories, proceed as for a group plot, but use bars instead of dots (see Fig. 5).

***3.3.5 Pie Charts*** These are useful to show percentages, especially if there are few categories, or few of the categories account for a large part of the data. Wedges of a circle are colored such that the area for each category corresponds to the percentage of that category (see Fig. 6).

***3.3.6 Histograms*** If you want to check the distribution of a single continuous variable, divide the range of data into a suitable number of categories or bins. Then count how many observations there are in each bin. Set off the categories on a horizontal axis and draw columns whose height corresponds to the number of observations (see Fig. 7). Draw the columns without spaces between them. If, e.g.,

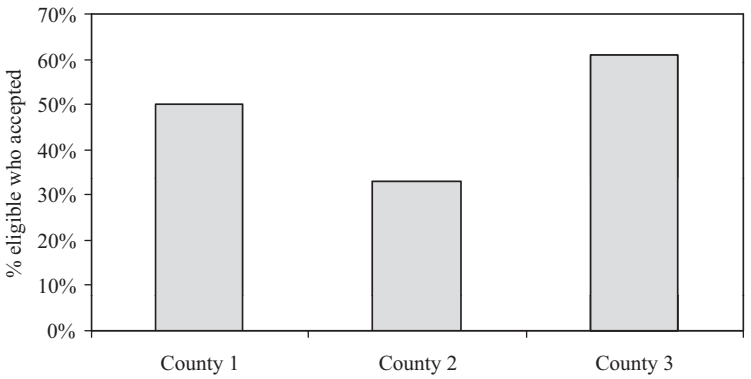


***FIGURE 5*** *Bar chart.*

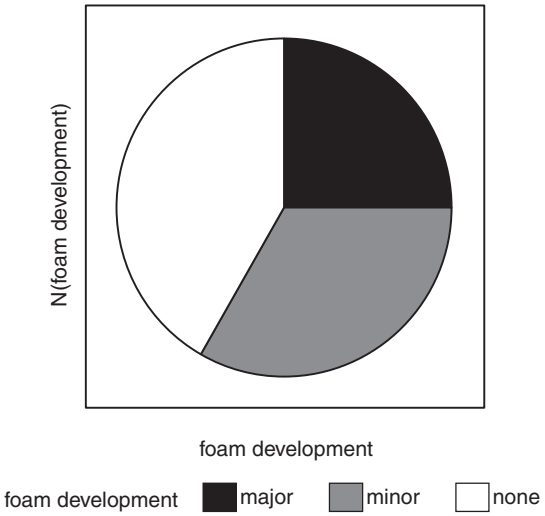


***FIGURE 6*** *Pie chart.*

your data are close to normally distributed, the columns will form a bell shape. A histogram will show you whether, e.g., you have a skewed distribution or heavy tails.

***3.3.7 Other Charts*** While most of the charts mentioned above are widely applicable and easily accessible in statistical software, you need not let yourself be restricted by this list. Special datasets may be best represented in a different way that takes some specific quality of the data into account. For example, if the dataset involves data representing some position in space, you may want to reflect this by arranging representations of the data in space (see Fig. 8).

## 3.4 Traps

***3.4.1 First Glance Conclusions*** Graphs typically highlight one or few specific aspects of the data. This is their strength, as it helps you focus on this aspect. But it is also their weakness, if it removes an important aspect of the data from focus. It is therefore recommended to represent data in various graphs during

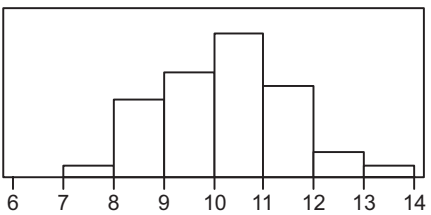


***FIGURE 7*** *Histogram.*

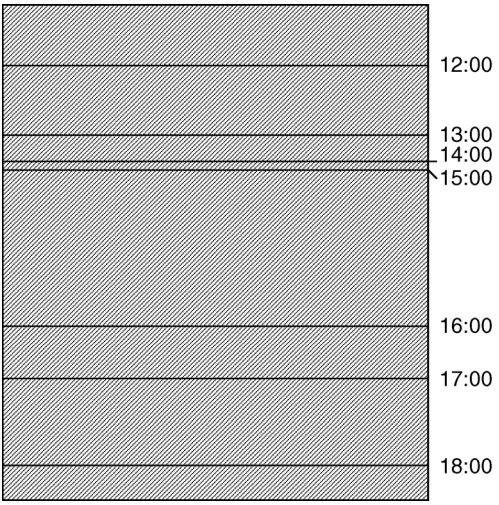


***FIGURE 8*** *Graph for spatial data: Filling height in a tank depending on time.*

exploration, just as you would turn an unfamiliar object in your hands to see it from different angles.

***3.4.2 Scaling the y-Axis*** A widespread misuse of graphs consists in constructing bar charts, or time plots for categorical data, where the y-axis does not start at 0, but at a random point below the lowest data point. Small differences can be blown up to look enormous.

***3.4.3 Depicting Only Part of the Data*** You can also create a misleading impression by choosing the data for a graph from a narrow range. For example, if showing a time plot of some continuous data from a short time period, you can create the impression of large fluctuations, while in fact the fluctuations are small as compared to the normal process range.

***3.4.4 Setting Graphs with Corresponding Axes, but Different Scales Side by Side*** If two graphs with similar data are placed close to each other, you can prompt comparisons that do not reflect the actual correspondence between the two if the axes use different scales. People will typically look at the data and not check the surrounding information unless prompted to do so. So you should ensure that they need not check to understand the data.

***3.4.5 Overloading a Graph with Graphical Elements*** Graphs in news media typically contain a large number of graphical elements, part of which convey information and part of which only have decorative purposes. This is fine as long as the view to the data is not obscured by the graphics. In general, the fewer elements a graph contains, the easier it is to grasp. You can therefore make your graph easier accessible as a communication tool if you restrict yourself to elements that

convey information. These considerations start already at the choice between two- and three-dimensional bars for a bar chart: Both contain the same information; the third dimension is a purely decorative element.

### 3.5 Examples Continued

*Example 3-1:* Oxygen and sugar consumption can be shown in time plots, possibly in the same plot with different axis. The relation between the two variables can be shown in a scatter plot. Also, scatter plots can be used to show oxygen consumption vs. temperature, oxygen consumption vs. pH, and so on. The four graphs can be arranged in a matrix, with corresponding axes having the same scales.

*Example 3-2:* A typical graph to consider shows the difference between ending and starting HbA1c plottet against dose. One may want to subdivide each dose group into the two gender groups, or into groups relating to BMI. One can also draw a scatter plot for the difference between ending and starting HbA1c vs. the non-categorized BMI or vs. the starting HbA1c. The same can be done for average fasting glucose and average post-prandial glucose.

*Example 3-3:* One can show the survey results nicely by coloring a map of the country. Each county is shown in a color corresponding to, e.g., the number of yearly tests per 1000 inhabitants. Colors are assigned to ranges of test numbers such that the ordering becomes clear; e.g., one can use different intensities of the same color. One may also be suggestive in coloring: Use red for the category of figures that is least desired by the customer to the analysis, and green for the category that is most desired. One can then add other relevant spatial information, e.g., the location of hospitals that may have an influence.

## 4 DESCRIPTIVE STATISTICS

### 4.1 Examples

*Example 4-1:* A pharmaceutical company is considering two different suppliers of glass vials for their pharmaceutical product. They get a sample of 1000 vials from each of the two vial producers and want to compare the two samples of vials. The vials are described by their length, their diameter, the thickness of the glass, etc.

*Example 4-2:* In a clinical trial, one compares two treatments for obesity. Some of the important characteristics of the subjects participating in the trial are gender, height, weight (before and after treatment), and BMI (before and after treatment).

*Common background for the examples:* In the above examples, one wants to describe a sample or a population by some well-defined characteristics that are of interest; e.g., in the clinical trial, the weights are of interest. However, the description should be in a summarized form, not a list of the weights for all subjects, since it is mostly impossible to get an overview from a raw list of data.

## 4.2 Measures of Location

The level of a specific characteristic in a sample, e.g., vial length in Example 4-1, can be described by different measures.

***4.2.1 Average*** This is the sum of the measurements divided by the number of them:

$$\bar{x} = \frac{1}{n}(x_1 + \cdots + x_n) = \frac{1}{n}\sum_{i=1}^{n} x_i$$

where $n$ is the number of observations and $x_i$ is measurement $i$.

***4.2.2 Median*** This is the value where half of the observations have a value that is smaller than or equal to this value. To determine the median, one should order the observations in ascending order. If the number of observations is equal, the median is the mid-point value between the two middle observations; e.g., if there are 10 observations, the median is midway between the values of observation 5 and observation 6. If the number of observations is unequal, the median is exactly the observation in the middle.

## 4.3 Measures of Variation

When the level/location of a characteristic is known for a sample, it is also desirable to know how much variation there is in the sample. For example, in Example 4-1, the two producers of vials may have the same average vial length, but the variation in the two samples may be quite different and very important to the pharmaceutical company.

***4.3.1 Variance*** The variance is calculated as $\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$, where $n$ is the number of observations and $x_i$ is measurement $i$, and $\bar{x}$ is the average of the $n$ observations. The variance is a sort of average of the squared distance between the observations and their average. Using the above definition, the variance is only well defined if there is more than one observation; otherwise, one will divide with zero, which is impossible. However, in some statistical textbooks, the divisor is $n$; in that definition the variance is zero if there is only one observation. Normally, the variance is always larger than zero. The variance is often used in statistical calculations, but as a descriptive measure, it is not that useful since the scale is the squared natural scale; e.g., in Example 4-2, the measurements of weights could be in kilograms (kg) and then the variance would be in kilograms squared ($kg^2$). As a descriptive measure of the variation, the following is used.

***4.3.2 Standard Deviation*** This measure is given as $\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2}$, which is the square root of the variance. The standard deviation is measured on the same scale as the average – the scale that the observations are made on.

***4.3.3 Range*** The range is calculated as the difference between the highest and the lowest value of the observations. It is a very crude measure of the variation in the sample but very easy to calculate.

There are other measures such as skewness and mode that can be of interest. For a description of these, refer to [1].

## 4.4 Traps

***4.4.1 Outliers*** When one describes the sample of interest, one should be aware of outliers (see 6.4.3) and their influence on the different descriptive statistics. The average is sensitive to outliers. For example, if there is an observation with a very high value, the average will be quite high compared to what it is if the outlier is removed. The median is far more robust since the removal of a single observation (the outlier) will not change the median very much. It can therefore be advisable to state both the median and the average. If they are close, the distribution the observations stem from is symmetric, and there are no outliers. Both measures of variation are sensitive to outliers.

***4.4.2 Small Samples*** If the number of observations in a sample is small, the descriptive statistics may not be representative for the entire population that the sample is drawn from. Also, it will be hard to determine if there are any outliers (see 6.4.3)

## 4.5 Examples Continued

*Example 4-1:* Descriptive statistics should be calculated for each of the measurements (vial length, glass thickness, etc.) for each of the two samples.

*Example 4-2:* For each of the two treatment groups, descriptive statistics should be found for height, weight, etc. The two groups might further be split in groups defined by there sex. It would make sense to investigate that treatment differences are not due to differences in, e.g., height or weight in the two treatment groups.

# 5 DISTRIBUTIONS

## 5.1 Examples

*Example 5-1:* A supplier of components used in a medical device plans to deliver a description of his components. He would like to state an interval within which the length of the components will be with a given certainty.

*Example 5-2:* As a part of a quality control program, a sample is taken from every produced batch of a medical device. Every device in the sample is inspected and either accepted or rejected. It is desired to have a description of the sample attributes.

*Example 5-3:* In the factory of a pharmaceutical company, areas are classified according to their state of cleanliness. To ensure that the areas are as clean as intended, samples are taken on different surfaces in the rooms and the numbers of germs counted. It is desired to use the samples for a monitoring program for the state of cleanliness.

*Common background for the examples:* In all the examples, one wants to describe some data and, based on the data, set some kind of limits.

## 5.2 Distributions

Distributions are mathematical functions that fulfill specific criteria. Some of the distributions are very well described mathematically and have some nice properties. Practically, distributions are useful for describing different kinds of data, for calculating confidence intervals, setting limits, and modeling dependencies. They can also be used in hypothesis testing. Some distributions are only defined for non-negative integer numbers (0, 1, 2, ...) and may only be defined for a finite number of values. Such distributions are known as *discrete distributions.* Other distributions are defined for all non-negative real values or for all real values. These distributions are often referred to as *continuous distributions.* The name of a distribution, e.g., the normal distribution, the binomial distribution, is really not a name for *one* specific distribution but a whole class of distributions that all have some special characteristics. For example, there are many binomial distributions, but they are all defined on a finite number of integer values; they are all described by their probability parameter $p$, etc. Generally, the family of distributions has some common characteristics and then some parameters that defines exactly which distribution one is dealing with. For instance, the binomial distribution has the probability parameter and the count parameter. Similarly, the normal distribution has the mean and the variance parameters.

***5.2.1 Normal Distribution*** The normal distribution is one of the most well-known distributions. It is characterized by its mean $\mu$ and its variance $\sigma^2$. The normal distribution is sometimes referred to as the bell curve since its shape resembles a bell. The shape, however, depends on the values of the mean and variance (see Fig. 9). The normal distribution is a continuous distribution defined on the entire real axis. It is often used for describing measurements such as length, height, and weight. There are many nice assets to the normal distribution; e.g., if two normal distributions are "added" (folded), the new distribution is again a normal distribution. One of the most famous and useful assets for the normal distribution is the central limit theorem, which states that if $X_1, X_2, X_3, X_4, X_5, \ldots$ are all independent stochastic variables following a distribution with finite mean $\mu$ and finite variance $\sigma^2 > 0$, then the average $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ will approximately follow a normal distribution with mean $\mu$ and variance $\sigma^2/n$. For practical purposes, this means that if a sample is taken and the average calculated, then the average will behave as if it came from a normal

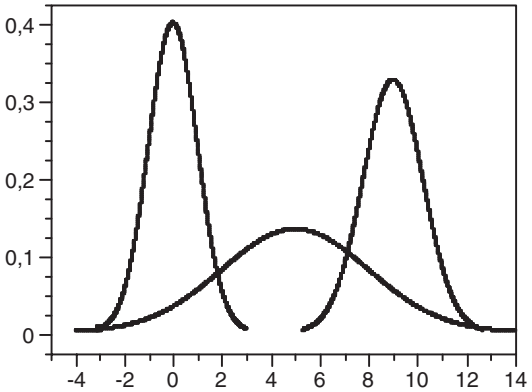


***FIGURE 9*** *Normal distributions with mean 0 and variance 1, mean 9 and variance 1.5, and mean 5 and variance 9.*

distribution, even though the distribution that the different observations in the sample stem from is not a normal distribution. For example, if "samples" each of 6 tosses with a fair dice are gathered, then the average of the samples will follow a normal distribution with mean 3.5. This property is very useful for setting limits or making control charts.

***5.2.2 Log-Normal Distribution*** The log-normal distribution is right-skewed – it has a long tail to the right – and is defined for all non-negative values (Fig. 10). If data follow a log-normal distribution, the natural logarithm of the data follows a normal distribution. So if one has data that are continuous and positive but skewed, then taking the natural logarithm of them might make them follow a normal distribution. It is then possible to use all the properties of and tools for the normal distribution and then transform any estimates back to the natural scale.

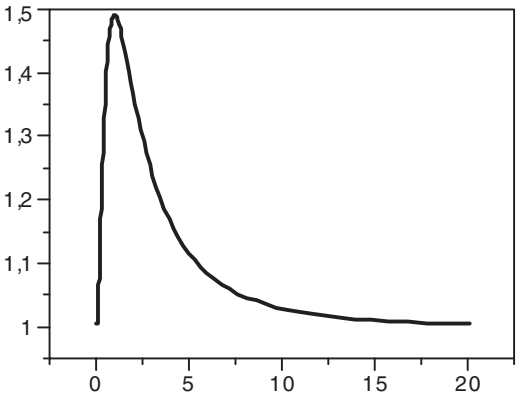


***FIGURE 10*** *Log-normal distribution with mean 0 and variance 1.*

***5.2.3 Binomial Distribution*** The binomial distribution is defined for two values, e.g., 0 and 1, heads and tails, and go or no-go. It can be used for describing the number of heads in $n$ coin tosses or the number of rejected items in a batch. It is described by the probability parameter $p$ and the "count" parameter $n$. $n$ is the number of coin tosses or the number of items in a batch, and $p$ can be thought of as the probability of heads in coin toss. Then the probability of tails is of course $1 - p$. The mean in a binomial distribution is $np$ and the variance is $np(1 - p)$.

***5.2.4 Poisson Distribution*** The Poisson distribution is defined for all non-negative integer values, i.e., 0, 1, 2, . . . . It can be used for describing rare events in a large population, e.g., the number of complaints on a specific product in the population of sold items of that product.

***5.2.5 Other Distributions*** Some distributions are used in hypothesis testing, e.g., the $t$-distribution, the $F$-distribution, and the $\chi^2$-distribution. All these distributions have a scale parameter denoted as the "degrees of freedom," which determines their shape. The $F$-distribution actually has two sets of degrees of freedom. There are many other distributions that are of interest. For a thorough introduction to these, refer to [2].

## 5.3 Examples Continued

*Example 5-1:* The average and standard deviation of the lengths is calculated and a histogram made to see if it is reasonable to assume that the measurements follow a normal distribution. If that is the case, a tolerance interval for the lengths can be calculated and stated in the description. This will give the receiver of the components an idea of how much the components vary in lengths.

*Example 5-2:* The number of rejects in the samples will follow a binomial distribution with count parameter, the size of the sample, and an unknown probability of rejects. However, an estimate of the probability of rejects could be made from the sample.

*Example 5-3:* The count of germs in a room will be a non-negative integer value, and for the very clean areas, there will probably be a lot of times where there are no germs at all. Therefore, a Poisson distribution might be suitable to describe data, but other distributions could be useful as well (log-normal or negative binomial). To assess fit, a histogram of data should be made along with fits of the different distributions. Once a distribution has been chosen, alert limits for the number of germs could be established based on the percentiles in that distribution.

# 6 SAMPLING

## 6.1 Examples

*Example 6-1:* From a batch of a pharmaceutical finished product, samples are taken during the production process and analyzed for the content of the active

ingredient. From the samples, both the average content and the variation of the content throughout the batch are estimated. Both need to comply with specification limits.

*Example 6-2:* A supplier delivers devices for inclusion in a manufacturer's product. The manufacturer tests the quality of the delivered batches by conducting a destructive test on a number of samples from each batch. The manufacturer and the supplier have agreed on a specific maximum percentage of defective devices. When delivery is stable, the manufacturer wishes to spend as few devices as possible on testing. However, if problems are seen, the manufacturer wishes to intensify the tests. To ensure agreement, both parties wish to agree beforehand on the amount of testing and acceptance criteria.

*Example 6-3:* Following a change in a production process, a process validation is conducted. The change was done in order to raise a parameter from its former to a new level. The process validation has the goal to show that the new level is obtained, within specified limits.

*Common background for the examples:* In all examples, a sample from a batch is supposed to represent the whole batch. From test results for the sample, it is estimated what is probable as test results for the whole batch, had it been tested in its entirety. The choice of the sample size is a compromise between obtaining sufficient knowledge to ensure a minimum quality level and minimizing testing costs.

## 6.2 Introduction of the Tool

When a sample is drawn from a batch, statistics such as the mean, variance, or percentage of defects will not be exactly the same in the sample as in the entire batch. But from the sample, one can estimate the corresponding statistics for the entire batch. The estimates for the entire batch come with an uncertainty that depends on the type and distribution of the data and on the sample size. Typically, one can reduce uncertainty by drawing larger samples.

## 6.3 Practical Details

***6.3.1 Point Estimates*** The best estimates for the mean and standard deviation in a batch are the average and standard deviation in a sample drawn from the batch. Typically, specification limits are directly defined as limits for the average and standard deviation of a sample of a certain size, allowing for simple comparison once the results for the sample are known.

***6.3.2 Confidence intervals*** If one draws a new sample from a batch, the mean and the standard deviation will very likely not be exactly the same as in the first sample. However, the larger the sample, the closer to each other will the estimates typically be (percentage estimates from small samples in large populations are an exception). Depending on sample size, one can calculate a confidence interval for the point estimate. A 95% confidence interval contains the true value of the parameter of interest parameter with 95% probability. That is, if you calculate a 95% confidence interval for an average from a sample, the true average for the entire batch is with 95%

probability within the confidence interval. Thus, the interval tells you how precise your estimate is. You can compare the interval to pre-defined limits that are based on requirements to the batch and thus conclude whether the quality of the batch is sufficient for the requirements. If your interval is very broad, the information you obtained from the sample may not be useful. You may need to increase the sample size to obtain more precise information.

***6.3.3 Sample Size Calculations*** Typically, one starts out with requirements to the batch, and an expectation as to what the quality will be. In this situation, one can calculate the appropriate sample size that will ensure a confidence interval that stays clear of the limit for acceptable quality for the batch, if indeed the batch quality is as expected (see, e.g., [3]). The closer the expected quality is to the worst acceptable quality, the larger a sample one will need in order to discriminate between the two. The methods for sample size calculation in this case are the same as those used by biostatisticians to determine the sample size in clinical trials (see, e.g., [4]).

***6.3.4 Acceptance Sampling*** There are ISO standards describing procedures for inspection sampling that can be helpful to contractually define the amount of testing and acceptance criteria [5,6]. The parties agree to an acceptable quality level (AQL), i.e., an average of a continuous measurement, or a percentage of non-conforming items per batch, which is the worst acceptable level. The supplier is asked to provide an average quality better than AQL. Depending on the batch size, the standard then prescribes a sample size and an acceptance number, i.e., for go/no-go measurements a maximum number of non-conforming items in the sample necessary to accept the batch, and for continuous measurements a limit for an acceptance number that is calculated from average and standard deviation in the sample, and the specification limits. The standards also prescribe procedures for relaxing and tightening the inspection, depending on the history of quality in earlier deliveries, and schemes for actually skipping batches for inspection if one is confident in the quality.

## 6.4 Traps

***6.4.1 Representativeness*** For the described procedures to work, it is necessary that the sample is representative. That is, all variation that occurs in the batch should also occur in the sample. It is therefore not advisable to collect all samples from one time spot during production. Sampling at fixed time intervals can also upset representativeness, if systematic variation occurs in the periods between samples – e.g., if high values of content occur just before a container of supply is changed, but samples are drawn systematically just after the change.

***6.4.2 Systematic Variation in a Limited Part of the Batch*** An underlying assumption for sampling is that the batches are homogenous. If, e.g., a disturbance causes highly diverging values in a limited part of the batch, a sample is likely not to pick up on it. Either none of the affected items ends up in the sample, or if it does, the

sample values may only be slightly affected, as the problematic part of the batch is small as compared to the rest.

***6.4.3 Outliers*** When drawing a sample, you may end up with one or few data points that look inconsistent with the remainder of the sample. Such data can have a strong influence on the point estimates. If there is actually something wrong with these data points, e.g., a measurement error or recording error, they can render the point estimates useless. In this case it may be wise not to include the outliers in the estimate calculation. There are outlier tests that help you decide whether one or several strange-looking data points actually are likely to stem from the same distribution as the remainder of the sample (see, e.g., [7]). But beware: When excluding outliers, you risk excluding data that actually do represent your process, and thereby heavily underestimating variation. Furthermore, often it is the different-looking data points that can put you on track of a process property that you did not know about, and thereby can be highly valuable sources of knowledge. Don't toss them away too easily.

## 6.5 Examples Continued

*Example 6-1:* Sample size and limits for the average and standard deviation in the sample are chosen such that the confidence intervals for the mean and standard deviation in the entire batch are within reasonable limits. Sample size and limits for the results from the samples are documented in the specification.

*Example 6-2:* In this example, the ISO standards for sampling procedures for attributes can be employed with advantage.

*Example 6-3:* Sample size calculations are conducted to ensure proper documentation of the level change, assuming that the level indeed has changed to the desired level.

# 7 STATISTICAL PROCESS CONTROL (SPC)

## 7.1 Examples

*Example 7-1:* In every analysis run of an HPLC-analysis, a control sample is measured. The purpose of the control sample is to verify that the analysis is performing as intended; i.e., there are no abrupt changes in the level, drift over time, or increase in variation. Furthermore, it is desired to use the control sample to follow the analysis over time.

*Example 7-2:* During production of a medical device, one of the parts for the device is tested in a strength test. From every produced batch of the part, a sample of six parts is taken and tested using the strength test. Both the level of the strength and the variation should be controlled.

*Example 7-3:* During production of switches for a medical device, samples of switches are collected on a regular basis, and the number of non-conforming

switches is determined. The purpose of the sampling is to check whether the production of switches is in control, which means the number of non-conforming switches must not increase.

*Common background for the examples:* In all the above examples, it is desired to follow a process (analysis, production line) over time and control that it is performing as intended, to react if the process is out of control, and not to react to random fluctuations as long as the process is in control. In order to achieve that, samples are periodically drawn from the process and measured. In Examples 7-1 and 7-2 the control measurements are continuous, whereas in Example 7-3 the control measurement is a go/no-go determination.

## 7.2 Introduction of the Tool

Statistical process control is a tool to understand and control processes. Generally, all work can be seen as a system of interconnected processes and all processes are subject to variation. The latter can be separated into two parts: variation due to special causes and variation due to common causes, or random variation. Common causes are causes that cannot be removed, e.g., day-to-day variation or operator-to-operator variation – they are inherent to the system. Special causes have a large effect on the system and are "abnormal" or exceptions. One specific operator or machine can cause them, or they can happen on one specific weekday. Sometimes, it can be hard to determine if a cause is special or common; e.g., operator-to-operator variation is a common cause, but if one operator is especially good or bad, that operator will be a special cause. To obtain control and stability of a process, the variation should be reduced as much as possible. In order to do that, one must know and understand the process and its variation. Ishikawa diagrams, process flow charts, and Pareto analyses can all be useful in the process understanding phase. Once knowledge of the process is obtained, the special causes of variation should be removed. Then the common causes should be reduced as much as possible. A process that is only subject to variation due to common causes is said to be in statistical control. A process that is in control has no drift and no sudden shifts in level. In Fig. 11 a process is shown, which is *in control*. Figure 12 shows a process that is *out of control*.

When the state of statistical control is reached, control charts can be constructed and used to maintain the state of control. There are many different kinds of control

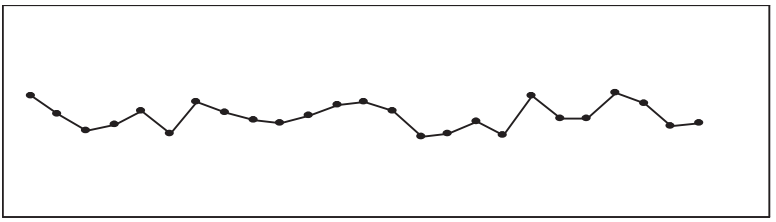

***FIGURE 11*** *Process in statistical control.*

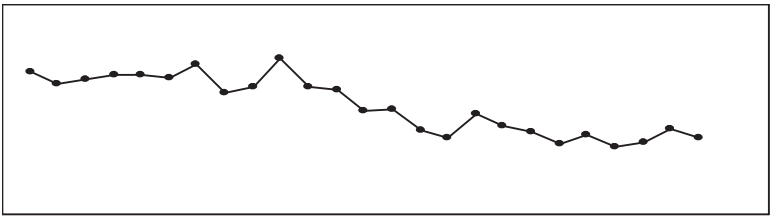

***FIGURE 12*** *Process which has drifted and thus is out of control.*

charts; in the following a few of those will be briefly described. For a more thorough description, see [1].

## 7.3 Types of Control Charts

***7.3.1 Shewhart Charts*** Shewhart charts are charts for normally distributed data. If, as in Example 7.2, samples of several items are drawn, both charts for the level (mean) and variation (standard deviation/range) can be constructed. The mean chart ($\bar{x}$-chart) is constructed using the average ($\bar{x}$) as the mid-level and $\bar{x} \pm 3\times$(standard deviation of the mean estimate) as the lower and upper limits of the chart. If data are truly normally distributed, 99.73% of the observations will be within the upper and lower limits. Thus it is very unlikely that an observation falls outside the limits by chance – unless the process has changed. The chart for variation – the $s$-chart – is constructed in a similar way with $\bar{s}/c_4$ as the centerline and $\bar{s} \times B_4$ and $\bar{s} \times B_3$ as the upper and lower limits. $c_4$, $B_3$, and $B_4$ are all constants that depend on the size of the sample [1]. An example of a $\bar{x}$- and $s$-chart is shown in Fig. 13. In order to determine limits, data from at least 20 samples should be available. Based on these, the limits can be calculated and the chart drawn. The 20 samples must all be from a period where the process is in control. When the charts have been constructed every time a sample is taken, $\bar{x}$ and s should be calculated based on the sample and plotted in the chart. If both of them are within the limits, the process is in control. Besides the limits, seven run rules exist to warn if the process is drifting, variation is increasing, etc. The run rules

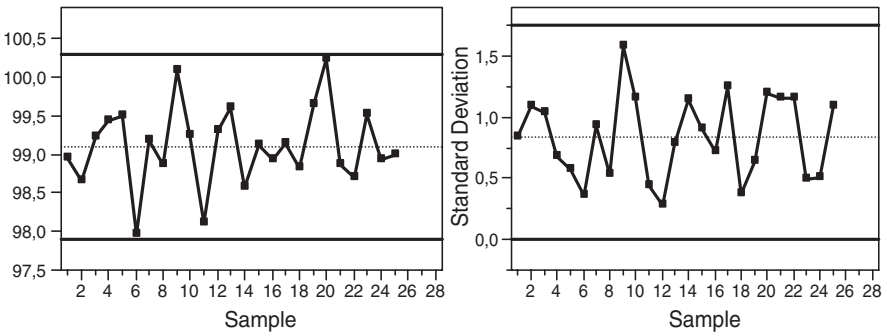


***FIGURE 13*** *x-bar chart and s-chart for a process in control.*

are designed to increase the charts' sensitivity toward specific kinds of expected changes. It is not advisable to apply all the run rules to the charts since that will sharply increase the number of false alarms.

***7.3.2 p-Charts*** When control measurements are in the form of go/no-go, the control chart needs to reflect this. If the measurements are recorded as fraction non-conforming or the number of non-conformities in the sample, $p$-charts or $np$-charts are used respectively. Both of these are based on the binomial distribution. The $p$-chart monitors the fraction of non-conforming items. It consist of a centerline $\bar{p}$ and a lower and upper limit often calculated as $3\times$(standard deviation of $\bar{p}$). $\bar{p}$ is the average fraction of non-conforming items. The process is in control as long as $\bar{p}$ for the samples are within the limits. $np$-charts are constructed in the same way, except that the centerline is given as $n\times\bar{p}$, and the limits are $3\times$(standard deviation of $n\times\bar{p}$). The advantage of the $np$-chart compared to the $p$-chart is that if the sample size is always the same, one plots the number of non-conforming items directly without further calculations.

***7.3.3 Other Charts*** The charts discussed above can be modified in the sense that the limits can be changed to cover another percentage of the distribution. If it is critical that the process is stable, one could use $2\times$(standard deviation of $\bar{x}$) as limits. This would cause more alarms but make the system more sensitive to a change in the process. In the same way if a process change is not critical and the workload caused by alarms is high, one could use wider limits than $3\times$(standard deviation of $\bar{x}$).

If data are not very well described by a normal distribution and there are plenty of data, or knowledge from similar processes exists, one could use the empirical distribution – the exact distribution created by the actual data – as the basis for the chart. The limits could then be chosen as percentiles from the distribution. For example, the 1 and 99 percentiles could be used as limits.

## 7.4 Traps

***7.4.1 Non-Normality*** If Shewhart charts are used, it must be investigated if the measurements follow a normal distribution. If the distribution heavily differs from a normal distribution, the limits could be erroneous and cause too many or too few alarms depending on the real distribution. In this case, one could try to transform data into something which is normally distributed. The natural logarithm is often useful as a function for transformation.

***7.4.2 Overinterpreting Alarms*** If 99% limits are used, then in theory one out of hundred observations should fall outside the limits. So alarms can appear without the process having changed.

***7.4.3 Changing Process*** The limits should be evaluated once in a while. If the process has not changed, new limits could be calculated on the basis of all data, reducing the uncertainty on the limits. If there have been no alarms at all for a long

time of operations, it could be a sign that the process has actually improved. If a shift in the process happens, the limits should be recalculated based on the new process level.

### 7.5 Outlook

Many other control charts exist besides the Shewhart and the $p$-charts (see, e.g., [1]). CUSUM charts model sums of binomial variables describing some characteristic of the process. There are also charts for moving averages, exponentially moving averages, etc. It is also possible to do multi-dimensional SPC to monitor several correlated process variables in the same chart, especially if data are normally distributed. Finally, if other limits exist than those set in the control chart, e.g., tolerance limits, process performance can be assessed by relating these external limits to the process-inherent limits derived for the charts. Such fractions are called capability indexes. A capability index is a measure of the amount of space the process has to "move within" without exceeding the external limits. If there is a lot of room for the process within the limits (the capability index is large), the process can be out of statistical control and still be within the limits. If the capability index is small, the process is close to at least one of the external limits, and scrap is produced as soon as a slight change in process level or variation occurs.

### 7.6 Examples Continued

*Example 7-1*: A control chart for individual measurements could be constructed for the measurements of the control sample. It is also possible to construct a chart following the variation of the measurements, but that has to be based on observation from more than one analysis run. If a change in the process appears, one needs to consider if it is the control sample which has changed (a drift could be due to degradation of the sample) or if it is the HPLC-system which has changed.

*Example 7-2:* If the strength measurements follow a normal distribution, Shewhart control charts can be made for both $\bar{x}$ and $s$. The $\bar{x}$-chart controls the level of the process and the $s$-chart controls the variation. It would be advisable to apply some of the run rules. These should be chosen to reflect what is most likely to happen to the process or what is most critical to the process.

*Example 7-3:* An $np$-chart could be constructed monitoring the number of non-conforming switches in the sample. If the sample is large and the number of non-conforming items is very small, a control chart based on the Poisson distribution may be more suitable.

## 8 HYPOTHESIS TESTING

### 8.1 Examples

*Example 8-1:* A pharmaceutical laboratory wants to compare two analytical methods. It is believed that one of the methods is better than the other, and

it is desired to verify that belief. A number of samples are analyzed using both methods. Data are continuous and can be described by normal distributions.

*Example 8-2:* A pharmaceutical company is considering two types of tanks for their fermentation process. It is believed that the yield from one tank is higher than from the other, but the company would like to investigate this before deciding which tank to use.

*Common background for the examples:* In both examples one wants to compare two things, and one has a belief that one of the things is better than the other.

## 8.2 Introduction to the Tool

Statistical testing is reminiscent of solving an advanced sudoku puzzle. If you cannot determine directly which of two numbers goes into a specific field, you choose one as hypothesis and follow the puzzle through from there. If a contradiction occurs, the other number must have been the correct one. If no contradiction occurs, the number first chosen is possible – but you do not know whether the second number would have led to a solution as well. If the sudoku was constructed according to the rules, there should only be one possible solution, but in real life that is not always the case.

Also in statistical testing, you choose a hypothesis and see whether it leads to a contradiction. You choose a null hypothesis that is actually the opposite of what you want to show. The hypothesis you are really interested in is referred to as the alternative hypothesis. In the case of the two analysis methods, the null hypothesis is that the two analysis methods are equally good. Hypotheses must be formulated in mathematical terms, and they concern the population parameters of the distribution (s) describing the actual data. For example, it may make sense to translate "method A is better than B" into "method A has a smaller variance than B," or "the measured level of some reference standard is closer to its certified level when measured with method A than with B." The population parameters are unknown, but they can be estimated from the actual data (see 6). Using the distribution of data and the hypothesis, a test statistic can be calculated. When the null hypothesis is true, the distribution of the test statistics is known, and it can be determined how likely it is to observe something at least as extreme as the test statistic. The notion "extreme" means further out in the tail(s) of the distribution of the test-statistic (see Fig. 14). If the chance of observing something at least as extreme is small, the null hypothesis is rejected. If the chance of observing something at least as extreme as the test statistic is large, the null hypothesis is accepted. The chance of observing something at least as extreme as the test statistic is often referred to as the *p-value* or the test probability. The idea behind this setup is that the distribution of the test statistic is known when the null hypothesis is true. If the observed test statistic – based on the actual data – is extreme (far out in tail of its distribution), there are two possibilities: Either one has by chance observed a very unlikely test statistic, *or* the null hypothesis is wrong. As the first possibility by definition is very unlikely, the conclusion is that the null hypothesis is wrong.

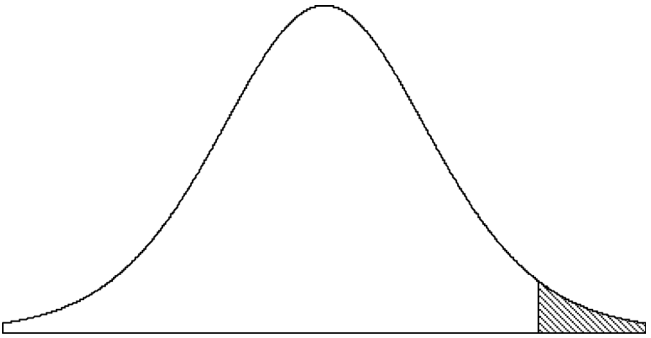

***FIGURE 14*** *The hatched area is in the right tail of the distribution.*

## 8.3 Practical Details

***8.3.1 Significance Level*** How small should the chance of observing something at least as extreme as the test statistic be in order to reject the null hypothesis? That is determined by the significance level. Often a significance level of 5% or 1% is used. If the chance of observing something at least as extreme as the test statistic is less than the significance level, the hypothesis is rejected.

***8.3.2 Type I Error*** If the null hypothesis is really true and a significance level of 5% is used, then in 1 out of 20 times, one will get a $p$-value that is less than 5% and thus reject the null hypothesis even though it is true. This type of error is known as the type I error. This is similar to a criminal court where the accused is considered innocent until the opposite is proved. In the statistical case, the null hypothesis is considered true until proven wrong. If it is very important that the null hypothesis is not rejected when it could be true, a low significance level should be chosen. If the null hypothesis is accepted, it is not the same as having shown that it is true. Again this is similar to the court case: Even if an accused is not convicted that is not the same has having proved him or her innocent. What one can conclude is that there is no evidence in the data to reject the null hypothesis.

***8.3.3 Type II Error*** The type II error represents the risk of accepting the null hypothesis when it is in fact wrong. The type II error is often denoted $\beta$ ($0 \leq \beta \leq 1$). $1 - \beta$ is known as the power, which is the chance of actually rejecting the null hypothesis when the alternative hypothesis is true. The power cannot be controlled for a given sample size but depends on the true parameters. The power is often used when one wants to determine the sample size in an experiment or clinical trial [1]. Often $\beta$ is chosen to be 10% or 20%.

***8.3.4 Sample Size*** It can be very important to assess the right sample size before an experiment is run. If the sample size is too small, the experiment will be inconclusive. On the other hand, it is often quite costly to do an experiment; therefore, it is desirable to keep the sample size as small as possible (see 6).

***8.3.5 One- and Two-Sided Tests*** In a two-sided test, both tails of the distribution of the test statistic are used. If the null hypothesis is $a = b$, and the alternative hypothesis is $a \neq b$, then this is a two-sided test since one is not assuming whether $a$ is larger than $b$ or smaller than $b$. In the corresponding one-sided test, the null hypothesis is still $a = b$, but the alternative hypothesis is $a \leq b$ or $a \geq b$. In either of these cases one would only be using one tail of the distribution of the test statistic. In the case of the two fermentation tanks that would mean testing the hypothesis "The two tanks give the same yield" against the alternative "Tank B gives higher yield than tank A." In this case, if B gave a lower yield than A, that would not be considered extreme.

***8.3.6 Equivalence Test*** In Example 8-1 if the laboratory really wanted to show that the two methods were equal, they should use an equivalence test. The idea in an equivalence test is to show that the difference between the parameters describing the two groups of interest, e.g., the level of the two methods, is smaller than some predefined value $K$. The value $K$ is chosen in advance of the test and is the largest acceptable difference between the two parameters. One finds the estimated difference between the two and, e.g., the 90% confidence interval for the difference. If the confidence interval is within $[-K, K]$, one has shown that the nominal difference between the two parameters of interest is less than $K$.

## 8.4 Different Tests

Statistical tests are often referred to by the name of the distribution of the test statistic. So a $t$-test is a test where the test statistic follows a $t$-distribution. Furthermore, some distributions are also described by a parameter called the "degrees of freedom" [2].

***8.4.1 t-test:*** The $t$-test is a test comparing the means of two groups, where data can be described by a normal distribution. It is possible to make a $t$-test either assuming that the variance is the same in the two groups, or assuming it differs. The difference between the two situations is the variance estimate and the degrees of freedom. If a $t$-distribution has a high number of degrees of freedom, it is very similar to a normal distribution with mean 0 and variance 1.

***8.4.2 F-test:*** The $F$-test is typically for comparing two variances. In an adaptation, it can also be used for an overall assessment of whether the means in several groups of normally distributed data are different or all equal. The $F$-test has two degrees of freedom, relating to the two variances.

***8.4.3 Chi-square test:*** The chi-square distribution is used in many different kinds of tests. One of the most well-known applications is the likelihood ratio test, which compares two statistical models, where one is nested in the other. The chi-square distribution is also referred to by its degrees of freedom.

***8.4.4 Non-parametric tests:*** If data cannot be described by any known distribution, a number of "distribution-free" tests exist [2]. These tests may not

yield as much knowledge as a parametric test, but if the assumptions for a parametric test are not fulfilled, a non-parametric test may be more adequate.

### 8.5 Examples Continued

*Example 8-1:* The two methods could be compared using a *t*-test. The null-hypothesis would be that there is no difference between the methods.

*Example 8-2:* An experiment should be designed to investigate if there is difference in yields between the two types of tanks and then the null hypothesis of no difference should be tested.

## 9 COMPARING GROUPS: ANOVA

### 9.1 Examples

*Example 9-1:* Two machines are producing components for a pharmaceutical device. It is suspected that the lengths of the components differ for the two machines, and an investigation is started to determine if there is a difference between components from the two machines. Data from both machines are collected for 10 days of production.

*Example 9-2:* A step in a pharmaceutical process has been optimized. Management would like to know if the optimized process has an effect on the strength of the product.

*Example 9-3:* In a QC laboratory an HPLC-analysis is routinely performed as a quality control of a given product. The analysis is performed on three HPLC-systems. One wants to know if differences in results stem from the use of different HPLC-systems, different technicians, or possibly both.

*Common background for the examples:* In the examples given above, several features are the same. It is desired to compare results from at least two groups (two machines, pre- and post-optimizations, three HPLC-systems). All results are measured on a continuous scale, and finally, it is desired to find out if there is a difference between the groups.

### 9.2 Introduction of the Tool

ANOVA is short for ANalysis Of VAriance. The term is used for statistical analyses where data are assumed to follow a normal distribution, observations are independent, and explanatory variables are categorical.

### 9.3 Practical Details

***9.3.1 Two Groups of Normally Distributed Data*** In the following it is assumed that data are normally distributed. We want to compare two groups of measurements to determine if they are different. The two groups of data stem from two normal distributions. Two normal distributions can differ in several ways: They can

have both different means and different variances, they can have the same variance but different means, they can have equal means and different variances, and finally both means and variances can be equal, meaning that the distributions are the same.

***9.3.2 Comparing Means*** Assume that the variances in the two groups are the same but unknown and that the means are unknown as well. We want to know if the means are different. Therefore, we want to test the null hypothesis that the means are equal against the alternative hypothesis that the means are different. Under the null hypothesis, all data stem from the same normal distribution. Therefore, a way to test the null hypothesis is to measure how far the difference in group means is from zero, taking into account the variance of the difference. This is done by constructing the t-statistic:

$$T_0 = \frac{\bar{x}_1 - \bar{x}_2}{\text{Std. dev.}\,(\bar{x}_1 - \bar{x}_2)}$$

Note that the numerator is the estimated difference between the two group means and the denominator is the estimated standard deviation of the numerator. When the variances are the same in the two groups, this statistic is known to follow a $t$-distribution with $n-2$ degrees of freedom where $n$ is the total number of observations. The hypothesis is rejected, if the chance of getting something more extreme (further away from zero) than $T_0$ is less than 5%, when the significance level is 5%. It is also possible to compare the two means by a t-statistic when the variances in the two groups are not the same, but then the degrees of freedom are different [8].

***9.3.3 Comparing Two Variances*** It is also possible to test if the two variances are the same. The test is performed under the assumption that the means are different. Here the null hypothesis is that the ratio between the two variances is 1. A test statistic is constructed, which follows an $F$-distribution with $(n_1 - 1,\ n_2 - 1)$ degrees of freedom where $n_1$ is the number of observations in group 1 and $n_2$ is the number of observations in group 2. If the test probability ($p$-value) is larger then the chosen significance level (often 5%) the null hypothesis cannot be rejected. This is NOT the same as having shown that the hypothesis is true; e.g., the variances are the same. If one really wants to show that two parameters are the same, one should use an equivalence test [9]. However, for modeling purposes, when the null hypothesis cannot be rejected, the non-significant effect is removed. In Example 9-1, that means one would usually start by performing the test for equal variances, and if the hypothesis is not rejected, it is assumed that the variances are equal, and one proceeds with the test for equal means.

***9.3.4 Comparing Several Groups*** In the same way as for two groups, it is possible to test if several group means are the same. This is done using an F-statistic (but different than the one used for comparing two variances). The test statistic is a measure of how big the variance is within the groups compared to how big the variance is between the groups as shown in Figs. 15 and 16.

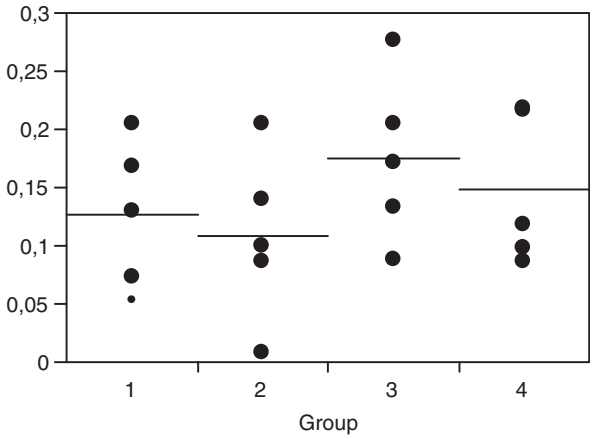


***FIGURE 15*** *There is small variation between the groups compared to the variation within the groups.*

Figure 15 shows the case where there is large variance within the groups compared to the variance between the groups. In this case it is not possible to reject the null hypothesis. In other words, there is no evidence in the data to show that the groups are different. In Fig. 16 there is little variance within the groups compared to the variance between the groups. In this case the null hypothesis of equal means in the groups would be rejected, and it is concluded that the groups are different in the sense that their means are different. Again it is assumed that the variances are the same in all the groups. A number of tests exist for testing this assumption ([10], p.227). When there are several groups, it might also be relevant to make pairwise comparisons. This is done as described under comparing two groups. However, if many pairwise comparisons are made, one should consider using a correction for multiple testing [11].

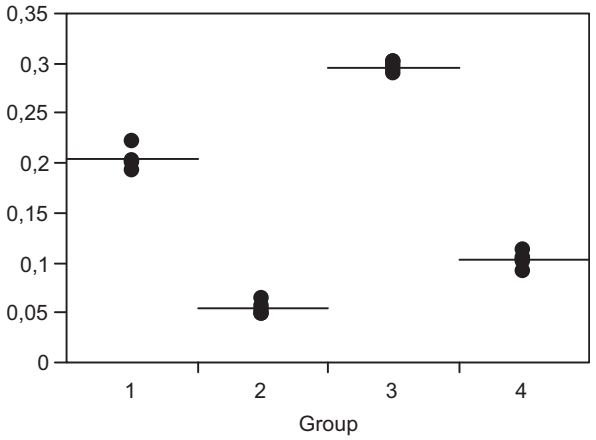


***FIGURE 16*** *There is large variation between the groups compared to the variation within the groups.*

***9.3.5 Group as Random Effect*** Consider machine 1 from Example 9-1. One might want to model the variation between days. In this case it is not the levels for the different days that are of interest since these are just a random sample of all the days, where data might have been collected. Thus we would like to model the differences between days as variation and determine how large the day-to-day variation is and how large the within-day variation is. It is thus assumed that all measurements are normally distributed with the same mean and with a variance that consists of a day-to-day component and a within-day component. The estimation of the components is very similar to the one under the section "Comparing Several Groups." When the number of observations on each day is not the same, the calculation of degrees of freedom is not straightforward, but it can be handled by most statistical software programs.

***9.3.6 Two-Way ANOVA*** In Example 9-3, there are two explanatory grouped variables, and it is desired to investigate their effect on the measurements. This can be analyzed in a two-way ANOVA. Data are again assumed to follow a normal distribution with the same variance. In the interaction effect model, the mean is assumed to be different for each combination of the two factors (technician and HPLC-system) in an "unsystematic" way (see Table 1).

One can start by making an overall $F$-test of whether the mean depends on any factors at all. If the test is rejected, one can proceed in the following way. First, the interaction effect is tested.

If the interaction effect is non-significant, the model can be reduced to a two-way additive model. In this model the mean still depends on the two factors but now in a systematic way, in the sense that a shift from one level in factor 1 to another level influences the mean in the same way for all levels of factor 2. From this model the effect of either factor 1 or factor 2 can be tested, and if that hypothesis is accepted, the hypothesis of no effect of the remaining factor can be tested (see Fig. 17). If one of these hypotheses is accepted, then the other one can be tested afterwards. If all the hypotheses are accepted, it means that there is no evidence in data to support the hypothesis of any differences between the two factors. This conclusion could also have been reached if one had started out with the overall $F$-test. A little care should be exercised though in using this approach since, sometimes, testing the two different ways might not give the same result. In order to test the hypothesis of no interaction between the two factors, there should be more than one observation for each of the

**TABLE 1 Interaction Effects**

| No Interaction Effect | | | Interaction Effect | | |
|---|---|---|---|---|---|
| | HPLC 1 | HPLC 2 | | HPLC 1 | HPLC 2 |
| Tech. 1 | 10 | 20 | Tech. 1 | ***30*** | 20 |
| Tech. 2 | 15 | 25 | Tech. 2 | 15 | 25 |
| Tech. 3 | 8 | 18 | Tech. 3 | 8 | 18 |
| Tech. 4 | 12 | 22 | Tech. 4 | 12 | 22 |

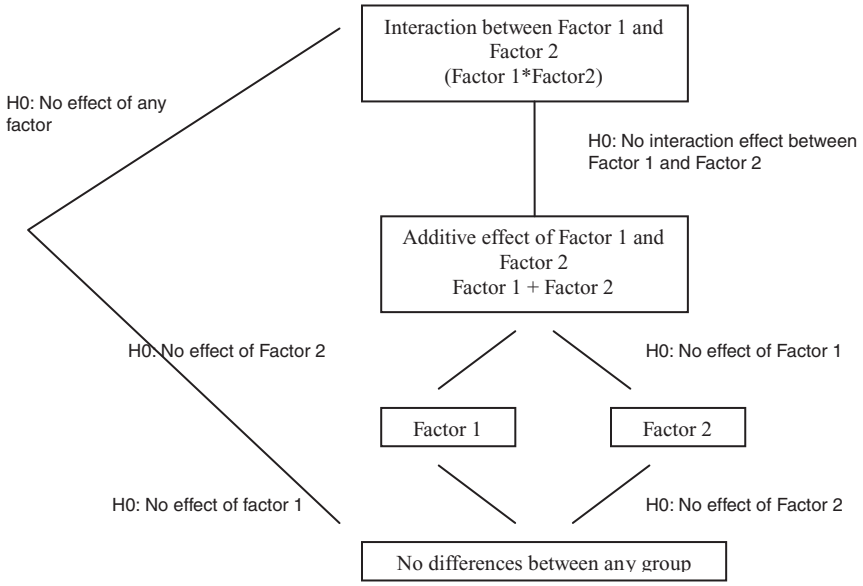


***FIGURE 17*** *Possible ways to test in a two-factor ANOVA.*

combinations of the levels of factor 1 and factor 2. If there is only one observation for each combination, the starting model is the additive model.

***9.3.7 Several Factors*** ANOVA with three or more factors can be handled similarly to the one described for a two-factor ANOVA. It is not always relevant to test the highest order interaction or it might not be possible due to the amount of data needed.

***9.3.8 Model Assumptions*** In all the ANOVA models, it is assumed that data follow a normal distribution and that the observations are independent. These assumptions should be verified by model checking, e.g., histograms and residual plots.

## 9.4 Examples Continued

*Example 9-1:* One can start out by comparing the variances from the two machines using an $F$-test. If the $F$-test is accepted, one can proceed comparing the two means as described above. If the difference is statistically significant, one can evaluate if the difference is also practically significant. This is not always the case. If there are more than one observation from the machines on each day, one could also determine the day-to-day variation and the within-day variation for each machine.

*Example 9-2:* Measurements of the strength before and after the process optimization could be compared. Both the variation and the mean of the strength should be compared.

*Example 9-3:* The effect of HPLC-system and technician could be assessed using a two-way ANOVA. If each technician on each HPLC-system makes several measurements, it is possible to investigate if there is an interaction effect between system and technician. If there is no interaction, it means that a shift from one HPLC-system to another influences the mean in the same way for all technicians. From that model the hypothesis of no difference between technicians can be tested, or the hypothesis of no difference between HPLC-systems.

# 10 MODELING LINEAR DEPENDENCE: LINEAR REGRESSION

## 10.1 Examples

*Example 10-1:* In a stability study, the concentration of a degradation product raises over time at a constant rate. First, one needs to determine the rate. Next, one needs to make sure that product released at the release limit will, with a certain probability, not exceed the shelf life limit at the end of shelf life.

*Example 10-2:* In an assay, one measures concentration of a certain molecule. One cannot determine concentration directly, but the degree of opacity which in this case is known to depend linearly on the concentration. One then analyzes the test samples together with a number of reference samples of known concentration. On the basis of the degrees of opacity of all samples, and the known concentrations of the reference samples, one can estimate the concentrations in the test samples.

*Example 10-3:* A yield depends on pressure and temperature in the process. If the pressure is kept constant, yield depends linearly on temperature (within a certain temperature range) and vice versa. However, changing both at the same time can have a different effect on yield than the sum of the effect of changing each of them from the starting point.

*Common background for the examples:* A dependent variable $y$ depends linearly on a continuous independent variable, or regressor, $x$. In all cases, one wants to model the dependence. The model is then used for different purposes: In the stability example, the analytical response $y$ is predicted for a given time point $x$. In the assay example, the concentration $x$ is estimated for a given response $y$. In the yield example, the dependent variable depends on two different regressors at the same time.

## 10.2 Introduction of the Tool

***10.2.1 Simple Regression Model*** For pairs of independent variables $y_i$ and regressors $x_i$, the model equation is $y_i = a + bx_i + e_i$, where the error term $e_i$ describes the difference between the predicted value $\hat{y}_i = a + bx_i$ and the data value $y_i$ for each $x_i$ (see Fig. 18). The parameters $a$ and $b$ describe the intercept

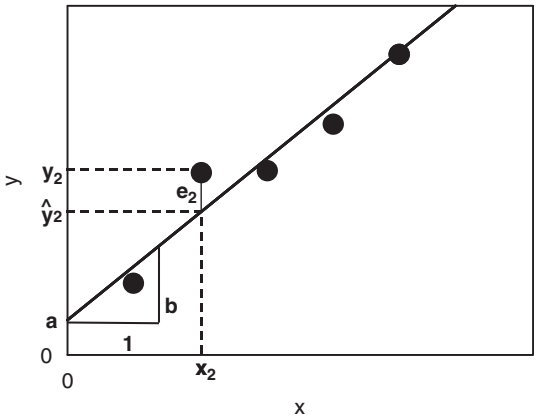


***FIGURE 18*** *Simple regression.*

and the slope of the regression line, respectively. Collectively, they are termed regression coefficients. The error terms $e_i$ are assumed to be independent and to follow a normal distribution.

***10.2.2 Multiple Regression Model*** A dependent variable can depend on several regressors, as in Example 10-3. The full model equation for two regressors $x$ and $z$ is

$$y_i = a + b \cdot x_i + c \cdot z_i + d \cdot x_i \cdot z_i + e_i$$

where $d$ is the parameter of the interaction term. One can test by means of a $t$-test whether $d$ is significantly different from 0. In this case, one says that the variables $x$ and $z$ interact. If two continuous variables do not interact, the graph for the model shows a plane. If they do interact, the plane is twisted (see Fig. 19). Computationally, any number of variables can be included in a model.

## 10.3 Practical Details

***10.3.1 Estimating the Regression Coefficients*** There are different methods to estimate the regression coefficients for a given dataset. The most common method is the method of least squares, which minimizes the sum of squared error terms. Least square estimates for $a$ and $b$ can be calculated using a closed formula (see, e.g., [2], p. 475). Other methods of estimation exist, e.g., the maximum likelihood method ([2], p. 293), which in many cases results in the same estimates as least squares, or Bayesian methods ([2], p. B-28). For the latter, there are usually no closed formulas for the estimates of the regression coefficients, but they are obtained iteratively.

***10.3.2 Checking the Assumption of Independent, Normally Distributed Error Terms*** Graphical methods are very helpful here. A histogram of the error

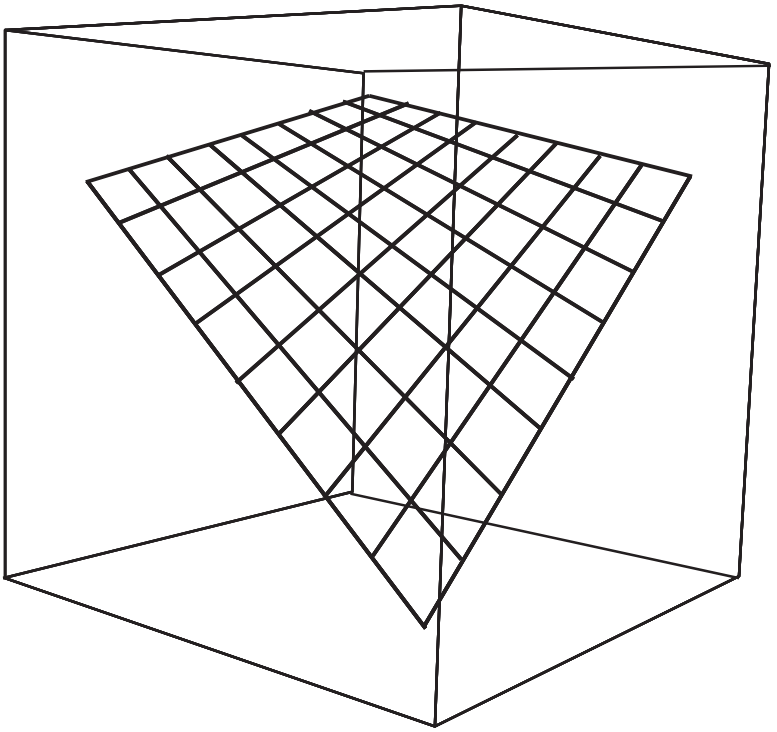

***FIGURE 19*** *Multiple regression with interaction.*

terms and a scatter plot of the error terms vs. the regressor(s) can show whether any systematic variation is left in the error terms or whether the error terms follow, e.g., a highly skewed distribution. In this case, tests in the model will not be valid.

***10.3.3 Testing for Overall Goodness of Fit*** *R*-square is a commonly used, but controversial ([2], p. 510), method to assess overall goodness of fit. It is a measure of how much of the variation in the data is explained by the model. It is computed as 1 minus the ratio of the error sum of squares and the total sum of squares. It is a number between 0 and 1, with 1 indicating that the model perfectly fits the data. *R*-square can be misleading when used to compare different models for the same data. For example, it increases with increasing number of regressors in a multiple regression model, also if the newly included variables do not contribute to explaining the variation. The adjusted *R*-square is an attempt to remedy this weakness. It includes a penalty for the number of regressors.

***10.3.4 Testing for Individual Parameters*** In multiple regressions, models can easily become rather complicated. To keep it simple, one needs a method to point out those regressors and interactions in the model that explain a significant part of the variation in the dependent variable. All other possible regressors can be left out of the

model. To decide which regressors and interactions to include, one can use $t$-tests. The tests are conducted by constructing an ANOVA table. Typically, one starts with the full model including all regressors and all their interactions, or with all interactions up to a certain degree. As for the ANOVA, one conducts an $F$-test for the overall significance of the model, and $t$-tests for the significance of individual parameters, starting with the higher order interactions.

***10.3.5 Confidence and Prediction Intervals*** Confidence intervals for the parameters result in an area around the regression line where each possible line within the area represents a model that would be consistent with the data. In simple regression, the area has a double trumpet shape; i.e., the uncertainty is greatest at the extremes of the data range (see Fig. 20). Prediction intervals show for each $x$ the range of $y$'s that would be consistent with the model. The prediction intervals are wider than the area defined by the parameter confidence intervals because they are based on both the model uncertainty, which is the basis for the confidence intervals, and the uncertainty associated with a single new observation.

## 10.4 Traps

***10.4.1 Extrapolation*** Extrapolating means using the regression outside the range of the observed data points. One assumes that the model derived from the observed data will also be useful in a different range. Caution should be exercised here, as many relations between measurable variables are far from linear when considering a wide range of values for the regressors. As no data are available to indicate the model fit in the extrapolated range, predictions can be highly misleading. See Fig. 21 for a dataset where a regression line estimated from the first seven observations would be very far from predicting the 20th observation correctly.

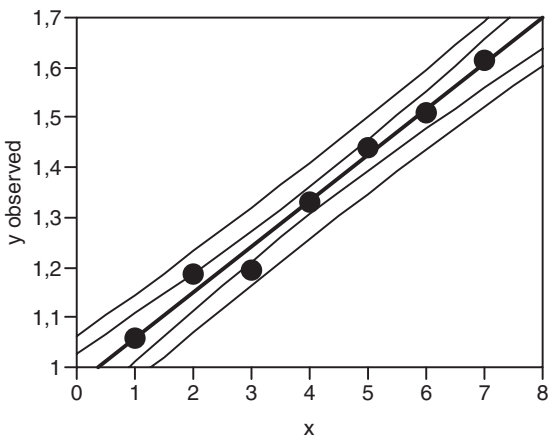


***FIGURE 20*** *Regression line with confidence limits for the line (closest to the regression line) and individual prediction limits.*

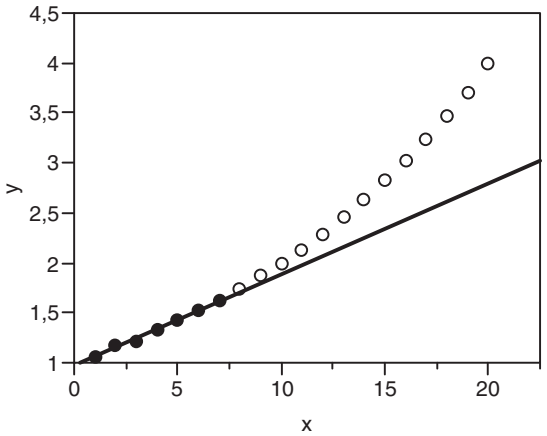


***FIGURE 21*** *Regression line which is not suitable for extrapolation.*

***10.4.2 Single Points with High Impact*** The further away from the average of $x$'s the $x$ value of a data point is, the larger is its impact on the slope of the regression line. A single data point with high uncertainty, if it has a very high or low $x$ value, can therefore seriously jeopardize the predictive value of a regression model (see Fig. 22). If you can plan where data are to be collected, consider placing many data points at the extremes of the range you want to cover, in order to minimize the uncertainty of the slope estimate.

***10.4.3 Non-Uniform Error Variance*** Regression models assume that the error variance is the same over the entire model. In many real-life situations, this is not the case. For example, the variance is often larger for higher dependent variable values. In

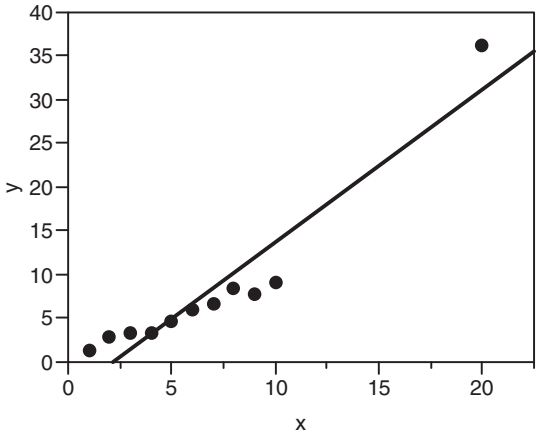


***FIGURE 22*** *Regression line highly influenced by one observation.*

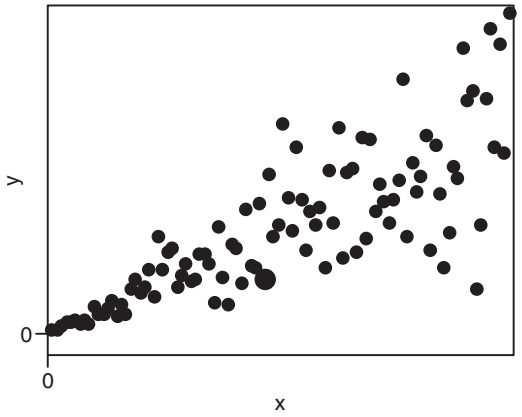


***FIGURE 23*** *Linear relationship with non-uniform error variance.*

simple regression, such non-uniformity of the error variance can already be seen in a scatter plot of the data (see Fig. 23). For multiple regression models, one can detect it by plotting the error terms against the regressors, one by one. In the case where the variance is higher for higher values of a regressor, you can sometimes solve the issue by modeling the logarithms of the regressors and the dependent variable, rather than modeling the regressor and the dependent variable directly. When using this solution, we recommend you check whether indeed the residuals expose a more uniform error variance than in the original model.

***10.4.4 Multi-Collinear Regressors*** Multi-collinearity arises in multiple regression models if the regressors are not independent of each other; i.e., one of them is close to a linear combination of the others. In this case, the regression coefficients are very sensitive to the specific data and the model chosen, and can vary widely during the model building process. Worse, the prediction of values outside the dataset on which the model is built can be extremely poor (see Fig. 24). The most effective way to deal with multi-collinearity is to collect additional data that break the linear dependence between the regressors. Another option is to respecify the model such that the regressors become independent, e.g., by removing one regressor from the model or by using a suitable function of the original regressors instead. Also, specialized statistical procedures have been proposed, e.g., ridge regression (see [12]) and principal component regression (see [20], p. 353).

***10.4.5 Interpretation: Significant Relationship Does Not Imply Causality*** As with correlations, a significant relationship does not imply that the value of the regressor is actually the cause for the value of the dependent variable. Unlike for correlation, the regressor and dependent variables have different roles. If you exchange the roles in simple regression, you end up with a different estimated regression line (see Fig. 25). This effect occurs because you minimize the variance of the residuals only for the dependent variable; i.e., you do not assume that the regressor

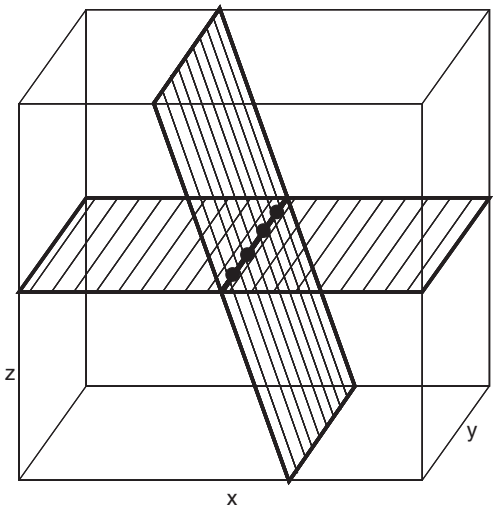


***FIGURE 24*** *Several surfaces consistent with a set of multicollinear data.*

values carry uncertainty. If you need to estimate a line where both variables are treated equally, you can use orthogonal regression (see [20]), which is closely related to correlation.

## 10.5 Examples Continued

*Example 10-1:* Conduct linear regression. The slope $b$ is the rate of degradation. You can use a $t$-test to determine whether $b$ is significantly different from 0, i.e.,

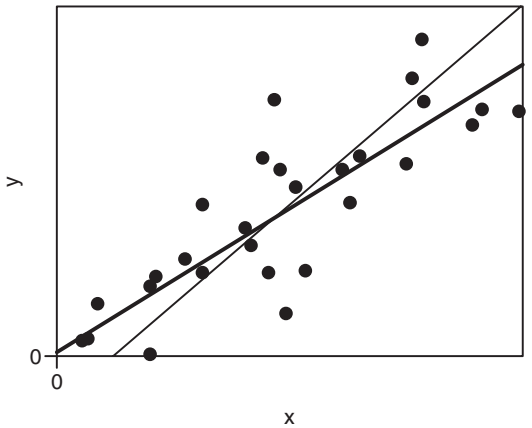


***FIGURE 25*** *Regression lines where x and y have the usual (solid line) and reversed roles (dashed line).*

whether the product degrades over time. If yes, make sure there is room for this degradation between the release and shelf life limits. If these limits are given, you can predict the content at the end of shelf life from the regression line. Compare not only the predicted value to the shelf life limits but also the prediction interval.

*Example 10-2:* Construct your model based on the data from the reference samples, with concentration as regressor $x$ and degree of opacity as dependent variable $y$. You can determine the concentration for the test sample as the concentration that, when plotted into the model, results in the observed degree of opacity for this sample.

*Example 10-3:* Construct a model with two regressors: time and pressure. You can then estimate the parameters in the model, and predict the effect that changes in time and pressure will have on the yield. Beware not to extrapolate, as time and pressure very seldom yield linear responses over a large range.

# 11 NON-LINEAR REGRESSION

## 11.1 Examples

*Example 11-1:* In a stability study, the amount of a degradation product is followed. Over time, the rate of degradation becomes smaller. The concentration of the degradation product does not surpass a certain level, which is dependent on the batch. One wants to predict at an early stage whether a batch is likely to stay within the shelf life limit for the degradation product throughout its shelf life.

*Example 11-2:* In a production process, a protein is activated. The activated molecule promotes the activation of additional molecules; i.e., the rate of activation accelerates over time. One wants to determine the time point at which at least a certain percentage of the molecule is activated, as the process can be stopped at this time.

*Common background for the examples:* In both examples, one measures a dependent variable over time and models the relationship, which is non-linear. One then uses the model in much the same way as a linear regression model: For intra- or extrapolating a response for a time point where one does not have a measurement, and for determining a time point where a certain response is most likely to occur. One could also be interested in testing the fit of the model. While time is the independent variable $x$ in both examples, one could imagine any other independent variable $x$.

## 11.2 Introduction of the Tool

In general, non-linear regression is about fitting a model $y_i = f(x_i, (a, b, \ldots)) + e_i$, where $f$ is any non-linear function in $x$ with parameters $a, b, \ldots$. Below are some examples of non-linear functions that are useful as models in many situations:

*Exponential growth or decay:* The growth or decay rate, which is proportional to the current size, $f(x) = ab^x$ (see Fig. 26).

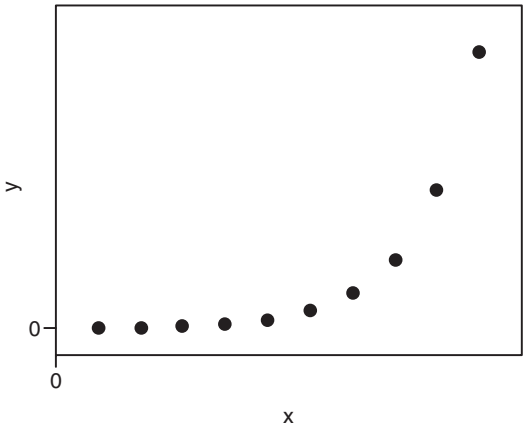


***FIGURE 26*** *Exponential growth.*

*Saturation curve*: Slowing growth, which approaches a saturation level $K$ without surpassing it, $f(x) = Kx/(b + x)$ (see Fig. 27)

*Polynomial curve:* $f(x) = ax^k + bx^{k-1} + \cdots$ (see Fig. 28)

## 11.3 Practical Details

***11.3.1 Choosing Between Non-Linear Functions*** If sufficiently many data with sufficiently high precision are available, a graph of the data may suggest a specific non-linear function that will work well for intrapolation without applying further knowledge of the mechanistic background of the data. On the other hand, if the mechanistics of the data generation process are well understood, they can suggest a specific non-linear function (see, e.g., [13]).

***11.3.2 Estimation*** Typically, there are no closed expressions for the best fitting parameters. Methods of numerical optimization, starting from a guessed value, are

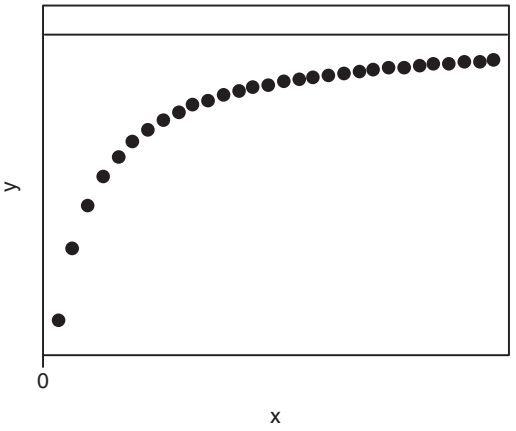


***FIGURE 27*** *Saturation curve.*

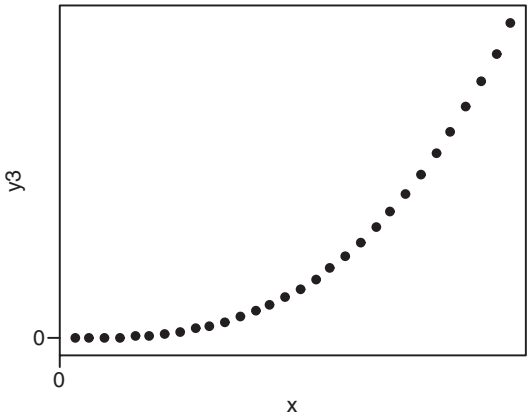


***FIGURE 28*** *Polynomial curve.*

typically used (see, e.g., Reference [14]). By the best fitting parameters, one often understands those that minimize the sum of the squared residuals; i.e., one uses least square estimation. Other methods are also feasible and can be warranted in cases of non-homogenous error variances.

***11.3.3 Linearization*** In some cases, one can transform the data such that they follow a linear relationship. For example, if $y = ae^{bx}$, you can transform data by applying the natural logarithm to both sides of the equation. This results in the linear curve $\ln(y) = \ln(a) + bx$. Transformation changes the influence of the data values and the error structure. You need therefore to consider these aspects before choosing whether to linearize.

## 11.4 Traps

In general, the same traps as with regression analysis are relevant. One trap that is often ignored is extrapolation.

***11.4.1 Extrapolation*** If you use a non-linear model to predict values for the dependent variable for values for the independent variables outside the observed data range, you can be even more off the right track than when extrapolating from a regression model, because non-linear models leave you more choices. For example, if you model the population of the earth for the years 1900–2000 by an exponential growth curve, and then use the model to extrapolate to the year 2100, you end up with a population figure that, in the current understanding, is rather unrealistic.

## 11.5 Examples Continued

*Example 11-1:* The initial data can be modeled by a saturation curve with time as independent variable $x$ and the concentration of the degradation product as dependent variable $y$. If the estimate of $K$ is below the shelf life limit for the

degradation product, the entire estimated curve will be below the shelf life limit. If $K$ is larger than the limit, one can determine from the curve what the concentration of the degradation product is estimated to be at the end of shelf life. It can, however, be difficult to obtain sufficiently certain estimates of the parameters as long as data are only available from the first half of the s-curve.

*Example 11-2:* The data can be modeled in an exponential growth model with time as independent variable $x$ and concentration of the activated molecule as dependent variable $y$. From the model, the time point can be determined where the curve surpasses the desired percentage. The time point where the curve surpasses the desired percentage can be determined from the model.

In both models, neither the uncertainty of the parameter estimates nor the prediction uncertainty for a specific value is taken into account, as both can be difficult to assess. This limitation should be kept in mind when making decisions based on the model predictions.

## 12 MIXED MODELS

### 12.1 Examples

*Example 12-1:* Measurements of assay from two stability studies, each consisting of three batches, are analyzed together. The analysis is supposed to show how assay develops over time, if the development is the same for batches in the same study or, if possible, the development is the same in all batches. The response variable is assay, and variables such as batch, time, and study are explanatory variables. Measurements are made at time 0 and after 1, 3, 6, 9, 12, 18, 24, and 36 months. At each time point, a double determination of assay is made for each of the batches in the study.

*Example 12-2:* A new reference material is being calibrated against an old one. Activity is measured in an assay for both the old material and the new one. The activity is determined using a number of independent series (runs). In each series (run), several vials from each of the two reference materials are used. From each vial a dilution series of five levels is made. It is desired to model the activity as a function of the explanatory variables reference material, series, and concentration (dilution level).

*Example 12-3:* Two laboratories are using the same analysis. A study is made where samples from the same batch are analyzed at each of the two laboratories on eight days. On each day three measurements are made in each laboratory. The purpose of the analysis is to investigate if there is a difference between the two laboratories. The difference can both be in terms of level and variation.

*Common background for the examples*: In all the three examples, the variable of interest is continuous and some kind of comparison between two groups is desired. Furthermore, in Examples 12-1 and 12-2, both continuous and categorical explanatory variables are present. In Example 12-3 there are only categorical variables. It is reasonable to assume that data from all the examples can be modeled using models for normally distributed data.

## 12.2 Introduction of the Tool

There is a very general framework of models for normal data that is denoted "mixed models." Both ANOVA and regression models are part of this framework. Within this framework it is possible to model mean structures, variance components, and structures for correlations in data.

The mean function is constructed as a sum of main effects and interaction effects. The effects can both be categorical as in an ANOVA and continuous as in regression models. An interaction effect between a continuous linear and a categorical variable is a linear function for each level of the categorical variable.

In a similar fashion, variance components can be used in the model. Variance components are typically categorical factors where one is not interested in the specific levels of the factor. The levels are instead a sample of all possible levels. In Example 12-3 the day factor could very well be seen as a random factor since the specific days are not interesting – they are just a sample of all possible days, and the variation between days is just a part of the variation that the analysis is subject to.

Finally, it is also possible to model correlation structures within the framework of mixed models. Examples of correlated data could be results from the same HPLC-run as opposed to results from different HPLC-runs or repeated measurements on the same subject. It is also possible to specify different residual errors for different groups.

Within the framework of mixed models, quite complex models can be built. It is possible to model data using models that are not sub-models of each other – this is often the case when one is trying to model correlation structures. Since the models are not naturally ordered, one has to choose which model is the best. There are several approaches to determine which model fits data better [15].

## 12.3 Traps

Since there are rarely finite solutions for the estimates in the models, estimates in very complex models may depend on the choice of algorithm, number of iterations, starting values, etc. When working with a very complex model, it is wise to check that the conclusions do not depend on starting values or the choice of algorithm. Also one should consider the risk of over-fitting the model. The fewer observations one has to base an estimate on, the more sensitive the estimate becomes to outliers. The robustness of such a model should therefore be examined. Generally, model assumptions should be checked using graphical methods for the distribution and structure of the residuals.

## 12.4 Outlook: Non-Normal Data

The mixed models framework can be extended to handle other types of data than normally distributed. For a discussion of these models, see [16].

## 12.5 Examples Continued

*Example 12-1:* One way to model these data is a model where each batch has its own straight line describing the development in assay over time. It would then

be possible to test if the three batches within each study have the same slope and in case that is accepted, if the slope from the two studies are different. Another way to model the same data would be to model a common slope for each study and then model the batch deviations from the common slope as random. Again it would then be possible to test if there is a difference between the slopes from each study.

The double determinations for each batch at each time point could very well be modeled as correlated since they are performed close together in time and possibly in the same analysis run. They are handled under the same conditions, which may be quite different for another double determination performed at another time point. The difference in conditions could be different technicians, different equipment, different buffers, different room temperature, different humidity, etc. This correlation could be modeled together with the two models described above.

*Example 12-2:* One way to model the activity as a function of concentration is by a straight line for each reference material and each series. Furthermore, the residual variation could differ for each reference material. Another way could be to model the effect of series as random variations of a common straight line for each reference material. It would also be possible to model the dependency of activity on concentration as a polynomium.

*Example 12-3:* The measurements could be modeled with different means for each laboratory and day as a random effect – different for each laboratory – and finally different residual errors (within-day variance) for each of the two laboratories. From this model it is possible to test the difference in mean between the two labs and also to test if the residual errors are the same and if the effects of day are the same.

## 13 DESIGN OF EXPERIMENTS

### 13.1 Examples

*Example 13-1:* In a production process, one wants to raise the yield. There are a number of factors that may influence the yield, such as holding time, feeding speed, and various conditions of the raw materials. Which of these actually has an influence, and how large the influence is, has not been established. While a lot of personal opinions exist among the process operators, there are few data outside the current set-point.

*Example 13-2:* For an analytical method, one wants to quantify the uncertainty and learn which factors, such as analyst, equipment, reagent batches, batch of reference material, and weather conditions, have the most important influence on the measurement results.

*Example 13-3:* The yield of a chemical production process is known to be mainly influenced by temperature and pressure. One wants to determine the optimal settings of these two factors to achieve maximum yield while maintaining acceptable product quality.

*Common background for the examples:* In all examples, known variables, called factors, cause variation in some well-defined outcome. One wants to determine which factors cause the most variation in the outcome, and how much. Therefore, one sets up an experiment prospectively. One varies the factors systematically among different levels. For example, if temperature is a factor, it can be changed systematically to the levels 5°C, 25°C, and 30°C. The resulting data are then collected systematically. Thanks to the design of the experiment, they are particularly suited for the intended analysis. Furthermore, the risk that the observed variation is caused by other influences than the systematic changes in the factors is minimized. This is obtained through the prospective setup and also through specific techniques to reduce noise, such as blocking and randomization. The factors are varied simultaneously to reveal any interactions among them. While the amount of possible combinations may greatly exceed the number of experimental runs that are feasible, systematic choice of combinations makes it possible to get as much information as possible out of a limited number of experimental runs.

## 13.2 Overview of Designs

***13.2.1 Factorial Designs*** The simplest factorial design consists of two factors. Each factor is tested at two levels, i.e., two different settings for each factor. This setup results in $2 \times 2 = 4$ different combinations of factor levels or treatments. Restricting oneself to two levels is useful when examining whether a factor has an influence on the response at all. When adding a third factor, the number of treatments is doubled, as each combination of the other factor levels is examined together with both the "high" and the "low" level of the third factor. For a general number $k$ of factors, one can construct a $2^k$ factorial design.

Factorial experiments are analyzed using ANOVA if the factors are categorical, regression analysis if the factors are continuous, or mixed models if there are both categorical and continuous factors. Not only main effects between the high and low levels of a factor can be tested and estimated, but interactions between the factors can also be tested and estimated. This is one of the most important properties of designs where several factors at a time are varied.

When specifying a full model with all interaction effects, down to the $k$-way interaction of all factors, testing is not possible because the data yield no estimate of error variance with which the effects can be compared. There are several ways to address this issue:

*Replicate the full design.* This may be warranted if the variation of the outcomes for the same treatment is high compared to the differences between treatments, and the experiment is small enough to make full replication feasible.

*Replicate only one treatment.* For the sake of balance, if your factors are continuous, the replicated treatment can be a center point where all factors are set at a "middle" level.

*Do not include the highest interaction in the model, assuming it is zero.* This solution is based on the "sparsity of effects principle," which states that non-random

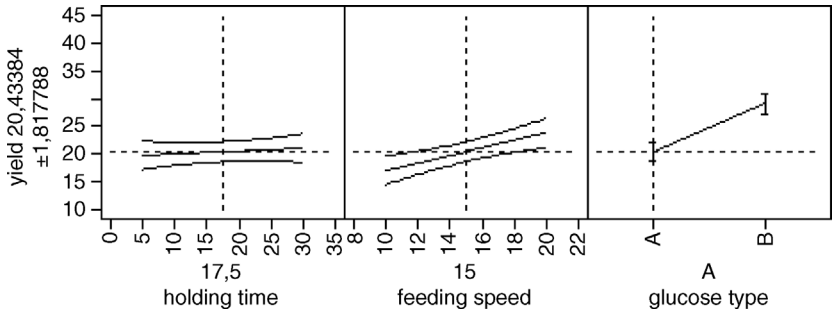


***FIGURE 29*** *Main effect plots: Feeding speed and glucose type have a significant effect.*

effects are sparse among all possible effects in an experiment. You can then use the variance of the highest interaction as initial estimate of error variance. When removing further interactions and main effects from the model, the error estimate improves. This is mainly useful for experiments with many factors, all of which are not expected to contribute to the variation of the outcome.

Useful tools for exploratory analysis are plots of the main effects (see Fig. 29) and the interactions that involve only few factors (see Fig. 30).

***13.2.2 Fractional Factorial Designs*** If many factors are considered in the same experiments, the number of runs necessary to try out all combinations quickly becomes very high. For example, for five factors, you need $2^5 = 32$ runs. Depending

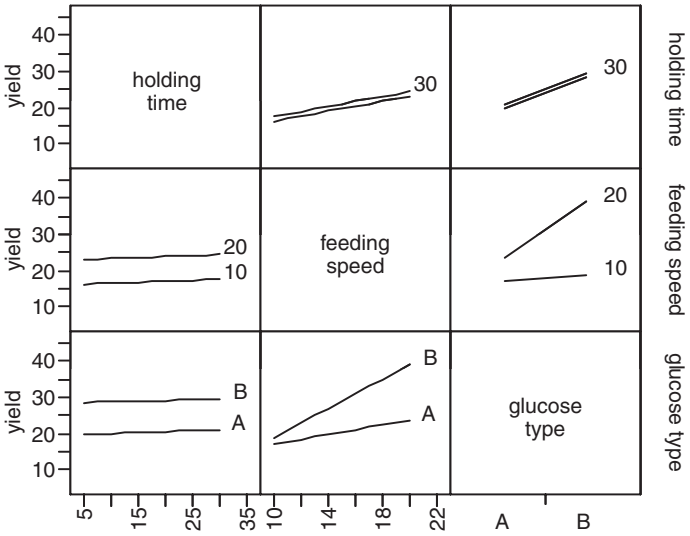


***FIGURE 30*** *Interaction plots for two-way interactions: Feeding speed and glucose type interact.*

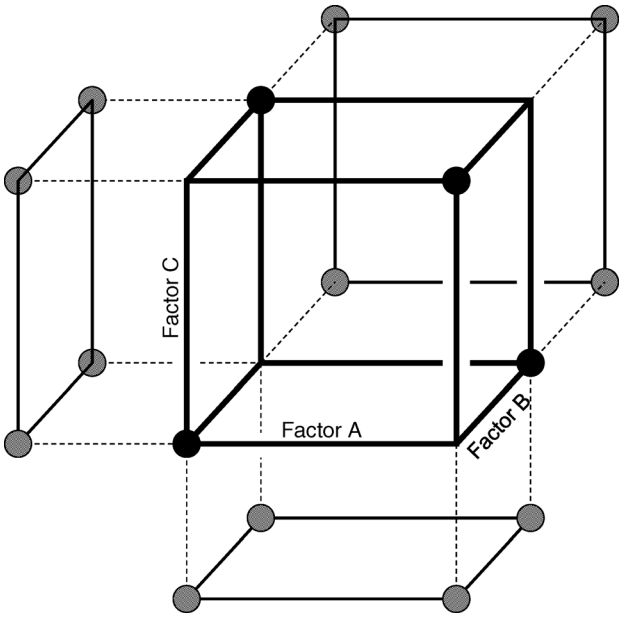


***FIGURE 31*** *Fractional factorial design.*

on the experiment, this may be much higher than what is technically or economically feasible. You can reduce the number of runs and still consider all factors, if you assume the higher order interactions to be zero. For example, in an experiment with three factors at two levels, you can only run one half of the $2^3 = 8$ experiments and still be able to estimate the main effects. This is so because, when ignoring any of the three factors, you still have a full factorial design in the two remaining factors (see Fig. 31). However, you cannot distinguish the main effects from the two-factor interactions. Each main effect is confounded with the two-factor interaction of the other two factors. This design is therefore only helpful if you expect the interaction effects to be negligible. It's your choice which of the two halves of the experiment you run. You may want to choose the one that includes your current standard operating conditions or avoid a combination that is particularly risky or stressful for the equipment. The two halves of the $2^3$ design are called $2^{3-1}$ designs.

With higher numbers of factors, you can choose to further reduce your designs at the cost of additional interactions, generating $2^{n-k}$ designs for any $k$ that is smaller than $n$. The figure $n - k + 1$ is called the resolution of the design. It tells you how much information you can extract from the data. The higher the resolution, the more information you get. The alias structure of a $2^{n-k}$ design tells you which main effects and interactions are confounded. It helps you to choose the right resolution and to choose between different $2^{n-k}$ designs. You should take care not to confound two effects with each other, both of which you are interested in estimating. For more details, see, e.g., [8].

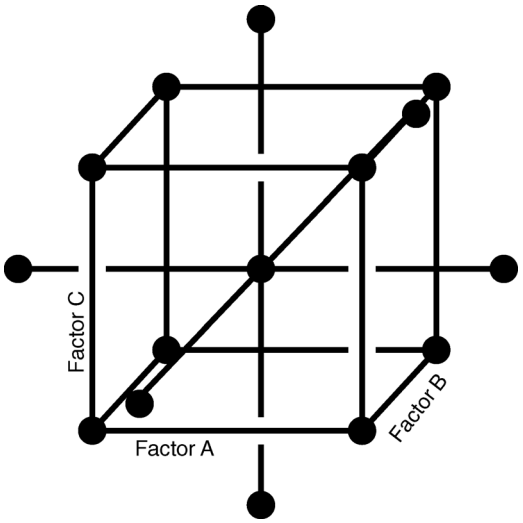


***FIGURE 32*** *Central composite design.*

Factorial and fractional factorial designs rest on the assumption that the outcome depends linearly on the factors. They are therefore not suitable in cases where the outcome follows a curve, and where you want to determine the optimum in order to turn the corresponding combination of factor levels into the future standard operation condition.

***13.2.3 Central Composite Designs*** Central composite designs are constructed from factorial designs with center points by adding "star points" on or outside the faces of the factorial design (see Fig. 32). Most commonly, the star points are located at the same distance in space from the center point as the corners of the factorial design, thus distributing information evenly in the space of conditions. Each factor appears at five levels (or three, if the star points are placed on the faces of the cube represented by the factorial design), which allows estimation of curvature.

***13.2.4 Box–Behnken Designs*** Unlike factorial designs, these designs avoid treatments where all factors are set at one of their extremes. Instead, either one or all factors are set at the middle of their range (see Fig. 33). In this way, all factors are used at three levels, which allows estimation of curvature.

***13.2.5 Plackett–Burman Designs*** As fractional factorials, Plackett– Burman designs are efficient as screening designs for the detection of few large main effects among many factors. Their main advantage is their economy of runs. For example, the smallest Plackett–Burman design for 11 factors has just 12 runs. The number of runs in these designs is a multiple of four, rather than a power of two as for factorials. So it is easier to take advantage of a budget for a few more experimental runs. Main effects are, in general, heavily confounded with two-factor interactions, which makes the

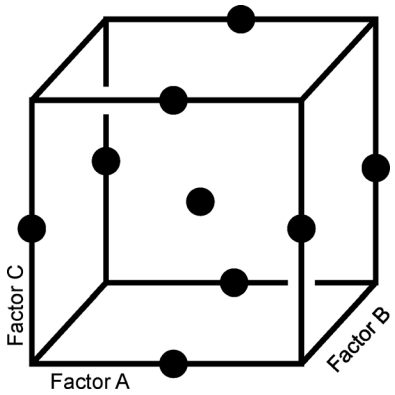


***FIGURE 33*** *Box-Behnken design.*

designs less suitable for understanding interactions. The designs do not have simple graphical representations like the factorials. They are generated using algebraic tools and are available in tables and software (see, e.g., [17]).

## 13.3 Techniques for Sequential Experimenting

***13.3.1 Augmentation*** When a need for additional data arises, an experiment can be augmented by adding new runs. For example, a fractional factorial design can be augmented by another fraction of the same design, allowing some of the confounding to be resolved. A factorial design can be augmented into a central composite design, allowing estimation of curvature.

***13.3.2 Response Surface Methodology*** This methodology is used when looking for the optimum of an outcome that depends on several factors, e.g., in order to turn the optimal combination of factor levels into future standard operating conditions. It follows a sequential approach, with each experiment leading to a new experiment in an area of conditions closer to the optimum. Typically, one starts with a factorial or fractional factorial design. First, factors with significant effects are separated from those that have no significant effect. Next, a line of steepest ascent is calculated from the results of the initial experiment. Along this line, in a suitable distance from the original conditions, a second experiment is set up, typically only involving the factors that turned out significant in the first experiment. The second experiment can still be factorial and lead to a new line of steepest ascent. Contour plots (see Fig. 34) can be helpful to visualize the journey through the space of conditions. When approaching the optimum, methods can be refined by using more complicated designs, such as the central composite, by replacing the linear line of steepest ascent by a second degree polynomial and by setting a shorter distance between two consecutive experiments (see [18]).

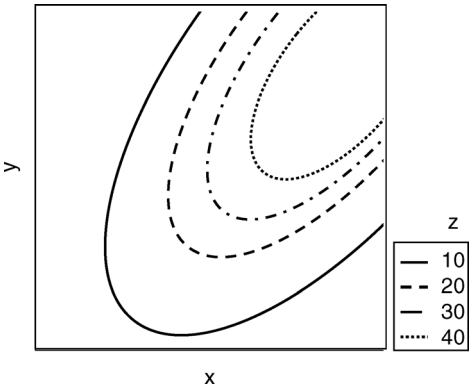


***FIGURE 34*** *Contour plot.*

Alternative approaches exist, e.g., a technique for high-cost experiments where one set of conditions is added at a time, effectively triangulating the response surface (see, e.g., [8]).

All approaches based on finding the shortest way to the optimum assume that there is only one optimum, and if it is high, you get there by going up all the time. While this assumption often holds in real-life situations, there are more complicated situations where response surface methods can be misleading. You can, e.g., be led to a local optimum, which is far inferior to the global optimum that can only be reached by crossing a deep valley of inferior conditions. Where local optima exist in isolated places, it is easy to miss them altogether.

## 13.4 Practical Details

***13.4.1 Blocking*** Differences in outcome between experimental units, e.g., batches, analytical setups, or patients, can be large as compared to the differences between treatments. In this case, it can be difficult to detect the treatment effects. One can reduce such known, but irrelevant, sources of variation by using similar experimental units. Where this is either not feasible, or one wants to spread the experiment over a group of units that is representative for a population, one can still reduce the variation by arranging the units into similar groups and assigning treatments systematically across the groups. The technique is similar to introducing a new factor without raising the number of runs. For example, in a factorial design, one can assign one half-fraction to one block and the other half-fraction to another block.

***13.4.2 Randomization*** Not all relevant sources of variation are known. To prevent unknown sources of variation from upsetting experimental results, one uses randomization to remove any predictability from the assignment of treatments to experimental units. For example, runs are conducted in random time order, and patients are randomly assigned to the treatments in a clinical trial. Various

randomization techniques, e.g., random number tables or dice, can be used to generate a randomization. Randomization of an experiment can be complete or only partial. The latter can be relevant when blocking is used or when some factors are hard to change.

***13.4.3 Testing for Effects and Effect Estimation*** Outcomes of designed experiments are analyzed by ANOVA for categorical variables, by regression analysis for continuous variables or, if both continuous and categorical variables are involved, by estimating a mixed model.

## 13.5 Traps

***13.5.1 Is the Measurement System Adequate?*** If the uncertainty in the measurement system is large compared to the size of the effects, it can be difficult to detect and estimate the effects. One can reduce variation by repeating the experiment or by working on the measurement system to reduce its inherent variation.

***13.5.2 Are All Relevant Factors Included?*** If a relevant factor is not included in the design and varies uncontrolled and unobserved during the experiment, it can seriously upset the data. Observing conditions that are suspected to have an influence, but cannot be changed, e.g., the weather, can be relevant to understanding the outcomes. If a relevant factor is not included in the design and is constant during the experiment, the outcomes may not be predictive for a different situation where the omitted factor has changed.

***13.5.3 Are the Factors Varied Sufficiently?*** When screening for effects, it is important to vary the levels of a factor as widely as possible within the relevant space of conditions in order to raise chances that the effect is large enough to be detected.

***13.5.4 Non-Linear Responses in Factorial Designs*** When running factorial designs with their inherent assumption of linearity in a situation where an outcome does not depend linearly on the factors, one may well be lucky that the approximation is good enough for the purpose. But it is possible to run into a situation where small changes in conditions yield widely different effect estimates or no significant effect at all (see Fig. 35).

## 13.6 Examples Continued

*Example 13-1:* In this example, one would screen the factors using a factorial, fractional factorial, or Plackett–Burman design, depending on the feasible number of runs and any suspicions about interactions. One would vary the levels as widely as possible. One would consider if there are any relevant sources of variation that cannot be kept constant during the experiment. For example, if several, but not all, sets of conditions can be tried out within the same batch, batch might be a relevant blocking variable. Others might include personnel,

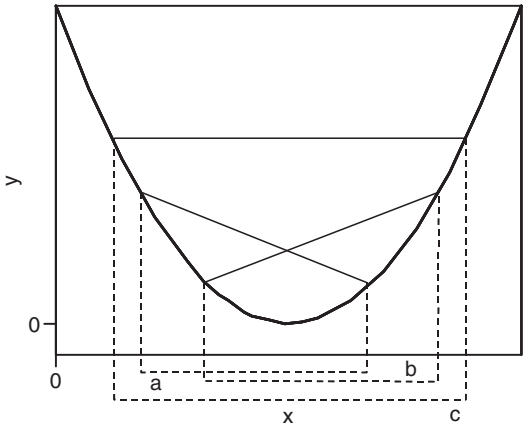


***FIGURE 35*** *Pairs of observations from a non-linear relationship suggesting y to decrease with x (a), increase with x (b), or be independent of x (c).*

equipment, or methods. The adequacy of the measurement system should be established prior to the experiment. Runs should be done in random order. Data are collected systematically and analyzed using ANOVA. With the learnings from the results, one might need to follow up with a second experiment.

*Example 13-2:* One lists the relevant sources of variation and sets up a factorial experiment with the sources of variation as factors. Typically, one uses samples that are as similar to each other as possible, e.g., from the same batch, or repeats the experiment for two types of samples that are known to be at the low and high ends of the analytical method's range. Days are a typical blocking factor. If weather conditions are suspected to play a role, one notes the weather conditions on the respective days. Analysis is done by ANOVA using random effects; i.e., a significant source of variation is quantified as a variance. For example, there will be a variance describing the variation between technicians in general, not a numeric difference between two specific technicians. The residual variance is the repeatability of the method, while the sum of all the variances shows the intra-laboratory reproducibility.

*Example 13-3:* A series of experiments with temperature and pressure as factors is set up according to the response surface methodology. The response variable is yield, while the measure of product quality serves as a boundary to the outcome space: Once quality is unacceptable, you backtrack.

## 14 TIME SERIES ANALYSIS

### 14.1 Examples

*Example 14-1:* A pharmaceutical company would like to forecast how much it is going to sell of a certain product in different countries. The company knows

how much it has sold previously in the countries and how many resources in terms of sales force and marketing budget it has planned to spend in the different countries.

*Example 14-2:* In a pharmaceutical production using a fermentation process, the responsible chemist would like to be able to predict how much yield he is going to get each day from the process. He knows that the yield depends on the amount of nutrition the cells have, and he believes the yield also depends on how much was harvested the previous day.

*Common background for the examples:* In both examples it is desired to forecast something (yield, amount to be sold) for specific time periods. The forecast will depend not only on some explanatory variables but also on how large the variable of interest was in the previous time period.

### 14.2 Introduction of the Tool

In time series analysis, models are built for processes where observations are made at given time intervals, e.g., daily or at each harvest from a fermentation process. Data are ordered in time sequence. The idea is to model the next observation in the process as a function of the previous observations, and sometimes also of some explanatory variables that may depend on time as well. Also the model will usually reflect the facts that observations that are close together in time are more correlated than observations further apart. Time series analysis can be used for understanding what drives a certain process and for forecasting the variable of interest. A classical example of that is in econometrics, where one forecasts the price of a share based on the price the day before. For a thorough introduction to time series analysis, see [19].

### 14.3 Examples Continued

*Example 14-1:* A time series model could be built forecasting the sale in the coming month as a function of sales in the previous months in the relevant country, the size of the sales force, and the marketing budget. It is possible to build in some kind of seasonal variation if that is meaningful (e.g., for hay fever products, or where hospital budget rounds have a major impact on sales).

*Example 14-2:* The yield from the fermentation process on a given day could be modeled as a function of the yield in the previous days and the amount of nutrition that was added the previous days. Additional explanatory variables could be included in the model.

## 15 MULTIVARIATE ANALYSIS

### 15.1 Examples

*Example 15-1:* A fermentation process generates a wealth of data. At regular intervals throughout the process, measurements are recorded for process parameters such as feeding speed for various ingredients and paddle speed,

as well as for process outcomes such as temperature, pH, and cell counts. Many of these measurements are correlated. A collective analysis of data from a range of batches is intended to yield a better understanding of how the many different parameters and outcomes are related to each other.

*Example 15-2:* Near Infrared Spectroscopy (NIR) method is used to determine water content in freeze-dried samples. The primary result of measurement is a spectrum, consisting of many different data points for each measurement. When calibrating the method, samples are analyzed first using NIR and afterward by Karl Fischer titration, which results in a reference measurement. The aim of the calibration is to link the many data of the spectrum for each sample to the single Karl Fischer result for this sample. The link allows then to read a water content figure from the spectrum, which means to predict the Karl Fischer result.

*Example 15-3:* From a fermentation process, a number of harvests are drawn throughout the process. The content of one degradation product, as determined by chemical analysis, follows a characteristic curve throughout the full length of the process. One wants to determine at various time points during the process whether the profile follows the usual curve or develops into a different profile.

*Common background for the examples:* In all examples, data consist of a large number of variables that are correlated. Multivariate modeling allows for using the full range of data while preventing correlation from introducing ambiguity to the model.

## 15.2 Introduction of the Tool

***15.2.1 Principal Component Analysis (PCA)*** The correlation structure in a multi-dimensional dataset is analyzed and used for compressing the dataset into a set of lower dimension. Instead of a large number of original variables that are correlated, the compressed dataset consists of a smaller number of principal components that capture the majority of the variation in the original dataset and are not correlated, see Fig. 36. The number of components indicates how much independent information there is in the

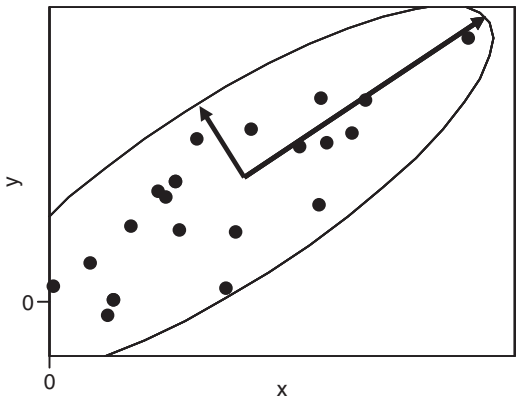


***FIGURE 36*** *Small multivariate dataset with principal components.*

dataset. The principal components are linear combinations of the original variables. The contribution of each variable to a principal component is called its loading. The loadings can give an understanding of the correlation structure and possible interpretations of the principal components. For a thorough introduction, see, e.g., [21].

***15.2.2 Factor Analysis*** In this analysis, the observed variables are assumed to be linear combinations of underlying factors. The method was developed in psychometrics, where, e.g., the observed variables were test scores, and an underlying factor could be intelligence. The factors are estimated from the data. If the error terms of the linear combinations are assumed to be equal, this analysis is closely related to PCA.

## 15.3 Examples Continued

*Example 15-1:* PCA can be used to reduce the wealth of variables to a smaller number of independent principal components, by which the batches can be characterized. By examining the loadings, one can get an understanding of how the variables group and contribute to the principal components. Graphical methods are recommended to support this process towards understanding the correlation structure.

*Example 15-2:* PCA can be used to go from the many data per spectrum, one for each recorded wavelength, to one principal component. Then the relation between the spectrum's principal component score and the content determined by Karl Fischer can be modeled by a relevant method, e.g., linear regression. With the model in place, the component score can be determined for each new spectrum, and the content, as it would have been determined by Karl Fischer, can be estimated.

*Example 15-3:* PCA or factor analysis allows to reduce the profiles into one or a few components. For these, ordinary or multivariate SPC can be established. One can do this not only for the full profiles but also for profiles up to a certain time point. During the production of a batch, one can determine the principal component score of the batch and enter it into the process control chart for the relevant time point. If a batch wanders beyond the control limits, this is a sign that the profile is clearly distinct from the usual profiles. A simpler and possibly comparably effective approach to detecting aberrant profiles may be to plot the profiles directly against a representative collection of profiles from earlier batches.

## REFERENCES

1. Montgomery DC. Introduction to statistical quality control. 5th ed. New York: John Wiley & Sons Inc.; 2005.
2. Montgomery DC, Runger GC. Applied statistics and probability for engineers. New York: John Wiley & Sons Inc.; 1994.
3. Guenther WC. Sample size formulas for normal theory T tests. Am Stat 1981; 34: 243–244.

4. Lachin JM. Introduction to sample size determination and power analysis for clinical trials. Control Clin Trials 1981; 2: 93–113.
5. ISO, 2859 series 1992–2007: Sampling procedures for inspection by attributes.
6. ISO, 3951 series (2007): Sampling procedures for inspection by variables.
7. Kuwahara S. Outlier testing: its history and application. Biopharm 1997; 10 (4): 64–67.
8. Montgomery DC. Design and analysis of experiments. 6th ed. New York: Wiley; 2004.
9. Wellek S. Testing statistical hypotheses of equivalence. Chapman & Hall/CRC; 2003.
10. Sall J, Creighton L, Lehman A. JMP® start statistics. 4th ed. SRS Institute 2007.
11. Bland JM, Altman DG. Multiple significance tests: the Bonferroni method. Br Med J 1995; 310: 170.
12. Hoerl AE, Kennard RW. Ridge regression: biased estimation for non-orthogonal problems Technometrics 1970; 12: 55–67.
13. Box GE, Hunter WG, Hunter JS. Statistics for experimenters: design, innovation, and discovery. New York: Wiley; 2005.
14. Bates DM, Watts DG. Nonlinear regression analysis and its applications. 2nd ed. 2007. New York: Wiley; 1988.
15. Verbeke G, Molenberghs G. Linear mixed models for longitudinal data. Springer; 2000.
16. Myers RH, Montgomery DC, Vining GG. Generalized linear models. New York: Wiley; 2002.
17. Plackett RL, Burman JP. The design of optimum multifactorial experiments. Biometrika 1946; 33 (4): 305–25.
18. Box GEP, Wilson KB. On the experimental attainment of optimum conditions (with discussion) J Royal Stat Soc Series B 1951; 13 (1): 1–45.
19. Chattfield C. The analysis of time series: an introduction 6th ed. Chapman & Hall/CRC; 2004.
20. Montgomery DC, Peck E. Introduction to linear regression analysis. New York: Wiley; 1992.
21. Sharma S. Applied multivariate techniques. New York: Wiley; 1996.

# *Systems-Based Approach to Laboratory Compliance with the Current Good Manufacturing Practice (CGMP) Regulations*

**David M. Bliesner**
*Delphi Analytical Services, Inc., Indian Rocks Beach, Florida*

## 1 INTRODUCTION

In January 2001 the FDA issued the document Compliance Program Guidance Manual (CPGM) 7356.002 titled "Drug Manufacturing Inspections (Pilot Program)." This manual provided guidance to FDA personnel for conducting inspections of pharmaceutical manufacturing sites in an efficient and structured fashion (1). The objective of this manual was to implement a quality systems-based approach to enforcing 21 CFR parts 210 and 211, the Current Good Manufacturing Practice (CGMP) regulations (2). In February 2002 the FDA formally adopted this guidance

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing,*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

manual and in August of that same year announced the Pharmaceutical CGMPs for the 21st Century Initiative. This initiative explained the agency's intent to integrate quality systems and risk management approaches into all its existing programs, the goal of which was to encourage industry to adopt more modern and innovative technologies. Subsequently, in September 2006 the FDA issued a guidance to industry entitled "Quality Systems Approach to Pharmaceutical CGMP Regulations" (3). This guidance document built upon the principles delineated in CPGM 7356.002 and bridged the gap between CGMPs and the current industry understanding of quality systems.

The guidance describes in a straightforward manner a comprehensive quality systems model for manufacturing drugs that, if properly implemented, not only ensures compliance with the CGMPs but also ensures manufacturers provide high-quality drug product to patients and prescribers. The bottom line is that a quality systems-based approach to drug product manufacturing is just good business. As in CPGM 7356.002, the guidance document describes an inspectional model that divides the operations of pharmaceutical manufacturing into six systems. These include (1) the quality system, (2) the packaging and labeling system, (3) the production system, (4) the facilities and equipment system, (5) the materials system, and (6) the laboratory control system. The diagram below shows the relationship among these six systems. Note that the quality system provides the basis for the other five manufacturing systems that are linked and function within it.

Of these five "inner" systems, the laboratory control system remains perhaps one of the most difficult to understand and master and leads to numerous quality failures.

A cursory review of Warning Letters listed on the FDA Web site http://www.fda.gov/foi/warning.htm shows that the laboratory is *rarely not* part of the regulatory action.

The root causes for lack of compliance with the CGMPs within the laboratory control system are perhaps most directly related to the multifunctional aspects of laboratory operations in general. In addition, many of the responsibilities within the laboratory are truly unique to that environment and vary somewhat from laboratory to laboratory depending upon the product line and size of the company. For example, a small generic solid dosage form manufacturing may have only a few established products and a few test methods. Compare this to a large innovator company that is constantly developing, validating, and testing new and different products. This means that each of these laboratories may require specific instrumentation, methods, and uniquely qualified personnel to execute the job functions. Because of this the laboratory control system is actually a collection of "systems within the system," each requiring special consideration and care.

However, for the purposes of most discussions, the laboratory control system can be divided into at least seven different sub-elements that are required to ensure overall compliance with the CGMPs. These include (1) Laboratory Managerial and Administrative Systems, (2) Laboratory Documentation Practices and Standard Operating Procedures, (3) Laboratory Equipment Qualification and Calibration, (4) Laboratory Facilities, (5) Methods Validation, (6) Laboratory Computer Systems, and (7) Laboratory Investigations.

The intent of this chapter is to demystify the laboratory control system and to help people understand each of the sub-elements that are simply "systems within the system" and are required for smooth and compliant laboratory operations of a typical analytical laboratory. To accomplish this, each of the seven laboratory control system sub-elements is analyzed in more detail. Specifically, this analysis takes the form of a series of questions an FDA inspector may ask during a pre-approval inspection (PAI) or a manufacturing site biennial inspection.

## 2 LABORATORY MANAGERIAL AND ADMINISTRATIVE SYSTEMS

Laboratory managerial and administrative systems are those sub-elements that provide the infrastructure for efficient and compliant operations of an analytical laboratory. It is sometimes difficult to directly relate items within this sub-element to the CGMPs. However, deficiencies within this sub-element that cannot be directly attributed to the CGMPs can lead to compliance failures. For example, 21 CFR parts 210 and 211 do not specifically require that someone be assigned the duties of training manager. However, failing to have an individual responsible for training and consequently not have an effective training program can degrade the overall quality of data that are generated by laboratory personnel.

The Laboratory Managerial and Administrative Systems sub-element includes at least eight individual topics. A laboratory that is in compliance with CGMPs should at least address each of these topics (as applicable). These topics include (1) Organizational Structure and Roles and Responsibilities, (2) Training, (3) Tracking

and Trending-Statistical Quality Control, (4) Complaints, (5) Laboratory Purchasing and Requisition, (6) Laboratory Administration, (7) Laboratory Chemicals, Solutions, Reagents, and Supplies, and (8) Laboratory Reference Standards and Solutions. These eight topics are listed below in the form of possible questions an FDA inspector may ask when conducting an audit of the laboratory. The list of items under each topic is fairly comprehensive but may not be totally inclusive of all those components that may constitute the sub-element.

---

**LABORATORY MANAGERIAL AND ADMINISTRATIVE SYSTEMS SUB-ELEMENT**

*Organizational Structure and Roles and Responsibilities*

Are current organization charts available and accurate?
Is there a policy/procedure that defines the responsibility and authority of the Quality Control Unit?
Have the responsibilities of each functional group and supervisor been clearly defined, including testing and operational requirements, SOPs, and all other critical functions?
Is an organizational structure in place that is properly staffed to assure that all required testing/monitoring and support activities are performed?
Is the span of control and authority assigned to the Quality Control adequate to allow proper execution of these activities?
What is the ratio of supervision to analyst (8:1 recommended)?
Are the roles and responsibilities for each position clearly defined?
Are job descriptions available?
Are signature authority, responsibility, and accountabilities appropriate and clearly defined?
Do systems exist to enhance communications, understanding, and working relationships between laboratory and Quality Assurance personnel?
Does a personnel performance evaluation system exist that tracks laboratory personnel strengths and weaknesses and establishes corrective action procedures to mitigate any weaknesses?
Does a master testing schedule or similar document(s) exist to ensure smooth workflow and minimize laboratory personnel over commitment?
Do current CVs and resumes exist for all personnel including consultants?

*Training*

Have the educational, training, and work experience requirements for each laboratory position been clearly defined and do they reflect current standards in the industry?
Are training requirements clearly documented in an SOP or similar guidance document (including managers, supervisors, analysts, and temporary staff)?
Has a training curriculum been developed for each position that clearly identifies training requirements for all required SOPs and policies, safety procedures, CGMPs, as well as all other internal and external courses or programs? (Database or hard copy?)
Is job-specific training identified and executed (e.g., laboratory analyst qualification system)?
Does each employee have a training file or similar record?
Are the training histories for each individual employee kept current? (Database or hard copy?)
Have all laboratory personnel been properly trained?
Is this training documented and are the documents available for review?
Is SOP training conducted in a fashion other than “read and understand”?
Are metrics used to determine effectiveness of SOP training?
Is there an analyst qualification program in place?
Has an individual been designated as the training coordinator or manager?
Is there evidence of management support for training and training programs?
Does a formal training budget exist?
Do vendor training records exist?
Is there evidence that newly hired employees are evaluated for skill level and competency?

**TABLE** (*Continued*)

Is there a formal training schedule in place and is it being executed?
Is there evidence of employee re-training and re-qualification?
Is there a positive attitude with respect to training on the part of management and laboratory personnel?
*Tracking and Trending-Statistical Quality Control*
Is tracking and trending (e.g., control charting) of data performed with the following points to be considered:

- For finished product testing?
- For raw material testing?
- Procedures delineated in an SOP?
- Appropriate periodic evaluation of control chart data included annual product reviews?
- Appropriateness of response to evaluations (e.g., excursions or trends are formally addressed as per SOP)?
- For finished product stability data (release and retain samples)?

*Complaints*
Is there a formal product complaint handling system in place, including some system of complaint monitoring, and statistical review with the following points to be considered:

- Procedures delineated in an SOP?
- Site procedures consistent with standard industry practice?
- Quality Control laboratories are involved and provide testing as appropriate?

*Laboratory Purchasing and Requisition*
Are levels of approval defined and appropriate for purchasing and requisitions for the laboratory?
Are items purchased from qualified vendors (e.g., raw materials, reagents, standards, and instruments)?
Does the laboratory have a defined yearly budget? Is the budget derived from existing data and projected workloads?
Does the laboratory have a defined capital budgeting process? Is the budget derived from existing data and projected needs?
Does the laboratory have a defined process for requesting and receiving approval to hire new personnel?
Does the laboratory have specific personnel assigned to purchasing and budgeting activities?
Do these personnel have the proper skills and training to perform planning, budgeting, and purchasing activities?
*Laboratory Administration*
Does the laboratory have a general administrator, office manager, or similar administrative support person?
Is someone assigned the responsibility for:

- Managing work orders for office and laboratory repairs?
- Budget preparation, oversight, and review?
- Overseeing stockroom operations?
- Supervising requisition of administrative and operational supplies?
- Managing external contractors and contracts?
- Managing/enhancing communication between the laboratory and internal/external customers?
- Supervising laboratory safety training and compliance?
- Monitoring overtime billing?
- Capital equipment purchases?
- Supervising administrative service personnel such as maintenance and calibration, laboratory computers, documentation section, and training section?

Are there laboratory corrective and preventive action committees, or similar groups in place?
Are the laboratory corrective and preventive action committees' meeting minutes published and proposed actions communicated to all appropriate personnel?
Are data reported in all types of reports easily traceable to raw data?
*Laboratory Chemicals, Solutions, Reagents, and Supplies*

**TABLE** (*Continued*)

| |
|---|
| Is there a written procedure for receipt and storage of chemicals and reagents? |
| Are chemicals and reagents properly labeled with date of receipt, date opened, and expiration and re-test dates? |
| Are laboratory-prepared reagents and solutions properly identified (e.g., chemical name or symbol, concentration, date of preparation, initials of the analyst who prepared it, and expiration date)? |
| Are records available to document preparation and standardization of volumetric solutions? |
| Is the frequency of standardization of various reagents described? Is it consistent with the current USP? |
| Are standardized reagents properly stored to assure integrity? |
| Are there procedures that describe the preparation of laboratory reagents and cultures? |
| Are there procedures that describe the maintenance of laboratory reagents and cultures? |
| *Laboratory Reference Standards and Solutions* |
| Is there a written procedure for ordering and receipt of compendial reference standards and non-compendial reference standards? |
| Are the primary standards of the current lot listed in the USP, EP, or JP? |
| Is the receipt of the standards logged? |
| Are all standards labeled with name, source, lot number, and expiration date? |
| Do written procedures include provisions for prevention of contamination of primary standards? |
| Are standards stored in a secured area under environmentally controlled and monitored conditions? |
| Are procedures for assuring standard integrity available? |
| Are working or "house" standards checked against primary standards at appropriate intervals? |
| Are stock solutions appropriately identified, and is their expiration based on their true stability? |
| Do procedures exist for the certification and use of non-commercially available reference standards? |
| Do certificates of analysis exist for all reference standards and are these certificates stored as controlled documents? |
| Have provisions been made for handling controlled substance reference standards? |
| Does the reference standard SOP address (map) proper handling of controlled substance reference standards? |

## 3 LABORATORY DOCUMENTATION PRACTICES AND STANDARD OPERATING PROCEDURES

Laboratory Documentation Practices and Standard Operating Procedures form a critical component of the overall laboratory control system. Much of the essence of CGMPs involves the generation, review, approval, revision, and control of procedures and data. In compliance parlance, "If it isn't written down, it didn't happen."

Laboratory Documentation Practices and Standard Operating Procedures subelement includes at least nine individual topics. A laboratory that is in compliance with CGMPs should address each of these topics (as applicable). These topics include (1) SOPs-General, (2) SOP-Specific Procedures, (3) Laboratory Test Procedures, (4) Laboratory Data and Results, (5) Security of Data, (6) Distribution of Results, (7) Chromatography, (8) In-Process Testing, and (9) Assignment of Re-Test/Expiry Dates.

These nine topics are listed below in the form of possible questions an FDA inspector may ask when conducting an audit of the laboratory. The list of items under

each topic is fairly comprehensive but may not be totally inclusive of all those components which may constitute the sub-element.

---

**LABORATORY DOCUMENTATION PRACTICES AND STANDARD OPERATING PROCEDURES SUB-ELEMENT**

*SOPs-General*

Is there a comprehensive and current laboratory document control system in place?
Is there a list of all approved SOPs?
Are the SOPs current, clearly written, and accessible to all appropriate personnel?
Is there a system for periodic review of all SOPs to ensure that they are consistent with current company and industry practices?
Is there an SOP governing the SOP or document control system (e.g., an SOP on SOPs)?
Is there a system for controlling the issuance and revision of all SOPs?
Do policies and manuals exist that supplement the SOPs (e.g., Level I and Level II guidance documents)?
Are the guidance documents readily accessible to all laboratory personnel?
Are all SOPs reviewed and updated at least every year?

*SOP-Specific Procedures*

Are there specific standard operating procedures covering:

- Change control?
- Record review and approval requirements?
- Description of the requirements for performing data review?
- Laboratory records (e.g. data recorded in bound notebooks or controlled worksheets)
- Dissolution testing?
- Sample receipt, documentation, handling, storage, and control?
- Laboratory investigations, deviations, and handling of out-of-specification (OOS) results?
- Policy for identification and reporting of new impurities and/or higher levels of previously known impurities?
- Validation and transfer of analytical methods?
- Laboratory computer validation?
- Preparation, labeling, identification, expiration-dating, and storage of chemicals, reagents, and solutions?
- Equipment/instrument use, preventive maintenance (PM), calibration, and qualification?
- Appropriate labeling of out-of-use lab equipment?
- Glassware washing (mechanical and manual)?
- Practices regarding housekeeping, safety glasses, eating in labs, smoking, etc.?
- Hazardous waste disposal?
- Handling and disposing laboratory samples?

Is there an SOP for defining appropriate sampling plans for all Quality Control laboratory testing?
Are samples randomly chosen and are they representative of all portions of the lot(s)?
Is there an SOP for the receipt, documentation, handling, storage, and distribution of laboratory samples?
Is a sample logbook maintained?
Are samples tracked? Is sample disposition included in tracking?
Are storage areas maintained for sample disposition (e.g., retention and destruction)?
Is there a master signature list, including person names, initials, and responsibilities?
Are samples labeled appropriately to include sample description, source, quantity, date sampled, date sample received for testing?
Are sample storage areas properly identified and maintained?
Is there a backlog with respect to sample analysis, and if so, is the backlog great enough to have an impact on testing results and product release?

*Laboratory Test Procedures*

Is there an index listing the testing documents and specifications?
Is this list updated and properly controlled?

**TABLE** (*Continued*)

Do the test procedures include sufficient instructions on how to conduct the testing and operate the specific lab instruments?
Is there a system for controlling the issuance and revision of all testing-related records?
Are laboratory test procedures and specifications approved by the Quality Assurance Unit?
Is there a written change control procedure that describes how analytical method changes are made?
Do specifications for compendial products meet compendial requirements?
Is there a procedure for determining "in-house" or tighter release specifications?
Are raw materials released by using validated analytical methods?
*Laboratory Data and Results*
Are data documented in bound pre-numbered logbooks, notebooks, or other data storage and acquisition systems?
Is the use of scrap paper, "post-its," or similar uncontrollable paper specifically prohibited in the laboratory?
Are corrections to data entry errors covered in an SOP?
Are all handwritten documents/data recorded in permanent ink?
If bound notebooks are used, is there a system for control, issuance, use, and return?
Is each notebook page dated and signed by the analyst and second reviewing authorized individual and performed in a timely manner?
Are all entries checked and approved for completeness of sample identity, proper use of reagents, standards, and execution of experimental procedures?
Are all outputs such as chromatograms and spectra appropriately identified, stored, referenced to notebooks, and readily retrievable?
Are all data, calculations, and results verified by second qualified individual for accuracy, completeness, and compliance with specifications?
Are the qualifications of persons conducting data review documented and acceptable?
Is data review documented by full signature and date?
Are data review and approval activities captured in LIMS?
Are data reviewed by the laboratory and approved by the Quality Assurance Unit before distribution? Is this process defined by an SOP?
Is there a system in place to assure that the method being used in the laboratory is the current approved version method for testing?
Is the method used for analysis recorded onto the laboratory testing results sheet or recorded into a LIMS?
Is there a written SOP for handling of re-integration of HPLC/GC data?
If a computer software package is used to calculate results, is there an example calculation included with the reported results verifying proper algorithm execution?
Are all correction factors used in calculations expressed with the proper units?
Are raw data defined in an SOP?
*Security of Data*
Is access to stored data limited?
Are data protected from fire, water, and other environmental hazards?
Are data readily retrievable?
Is the length of time that documents need to be retained defined in writing and does this time frame meet current industry practice?
Are backup copies of data stored at an off-site location?
Is there a disaster recovery plan, and has it been rehearsed?
*Distribution of Results*
Are results entered into a LIMS or other electronic management system?
Is LIMS used to compare testing results to the specifications for determining "pass/fail" decisions?
Is LIMS used to provide data for use in annual product reviews?
If data are transcribed from notebooks and reports, is transcription accuracy being verified?
*Chromatography*
Is system suitability performed routinely (e.g., HPLC methods, GC methods)?

**TABLE** (*Continued*)

| |
|---|
| Are the system suitability parameters determined from validation data? |
| Must system suitability parameters be calculated and met prior to executing a quality control sample analysis? |
| Is there an SOP that delineates a standard sample queue (i.e., interspersing of standards and samples during the course of the chromatographic run)? |
| *In-Process Testing* |
| Are validated methods used for in-process testing? |
| Do formal sampling plans exist for all in-process testing? |
| Are there established acceptance criteria/ranges/specifications for the in-process tests? |
| Is sample integrity preserved during handling and storage of in-process samples? |
| Are all the necessary industry standard in-process tests conducted (e.g., blend uniformity, weight variation, hardness, thickness)? |
| *Assignment of Re-Test/Expiry Dates* |
| Are expiration/re-test dates based on available in-house storage (stability) data or otherwise justified? |
| Does an SOP exist that differentiates between re-test and expiry dating? |
| Is there a limit for extending the re-test life of the substance evaluated? |
| Are there procedures for handling/storage of reference standards and reference cultures? |

## 4 LABORATORY EQUIPMENT QUALIFICATION AND CALIBRATION

Since the vast majority of data in a modern analytical laboratory are acquired via instrumentation, it is essential that these instruments have been properly installed and qualified, and are kept in proper working order. The Laboratory Equipment Qualification and Calibration sub-element addresses these issues.

Laboratory Equipment Qualification and Calibration includes at least three individual topics. A laboratory that is in compliance with CGMPs should address each of these topics (as applicable). These topics include (1) Laboratory Equipment Procedures-General, (2) Laboratory Equipment Procedures-Specific, and (3) Laboratory Equipment Procedures-Computer Controlled.

These three topics are listed below in the form of possible questions an FDA inspector may ask when conducting an audit of the laboratory. The list of items under each topic is fairly comprehensive but may not be totally inclusive of all those components that may constitute the sub-element.

| **LABORATORY EQUIPMENT QUALIFICATION AND CALIBRATION SUB-ELEMENT** |
|---|
| *Laboratory Equipment Procedures-General* |
| Are the laboratories equipped with all of the necessary instruments for the analytical testing to be performed? |
| Is there a written qualification, calibration, and PM program in place that is described in an SOP? |
| Is there a master equipment list available and is it properly maintained? |
| Is there a master equipment maintenance and calibration schedule? |
| Are specific personnel assigned to maintain the master equipment list and equipment maintenance and calibration schedule? |
| Are calibration/service vendors qualified and are their training records available? |

**TABLE** (*Continued*)

| |
|---|
| Is there a validation master plan, which indicates timing and responsibility for equipment qualification and re-qualification? |
| Are there "triggers" (e.g., specific criteria and/or circumstances) for re-qualification, and are they integrated into the master plan? |
| Does each piece of equipment have a logbook or file that documents instrument maintenance, calibration, and repair histories? |
| Does each piece of equipment have a usage log? |
| Do labels on the instruments identify the person who performed the calibration, date of calibration, and due date of next calibration? |
| Are IQ/OQ/PQ requirements, which clearly define performance testing acceptance criteria for each piece of equipment, clearly defined in an approved protocol? |
| Has IQ/OQ/PQ been performed and properly documented for all equipment? |
| Is IQ/OQ/PQ performed during initial system installation? |
| Are calibration and PM, including tolerances, based on the manufacturer's recommendations? |
| Are cleaning procedures included in the SOPs for instrument qualification or calibration? |
| Is there an SOP that requires that instruments failing calibration are removed from service? |
| Does the SOP provide directions for investigations when instrumentation is found to be out of calibration, yet still is being used to generate CGMP data? |
| Are the qualification specifications for equipment based on sound scientific principals? |
| Are calibration/metrology records secure (locked up)? |
| *Laboratory Equipment Procedures-Specific* |
| Are balances checked for corner loading? |
| Are pH meters standardized and calibrated? |
| Are balances calibrated at both upper and lower weighing capability using NIST traceable standards? |
| Are calibration correction factors attached to thermometers (if appropriate)? |
| Are pH meters calibrated at two points, not differing by more than 4 pH units? |
| Are UV spectrophotometers calibrated for wavelength and photometric accuracy? |
| Is there a written procedure for setup, calibration, and operation of dissolution baths? |
| Are there written procedures for calibration and maintenance of HPLCs, including pumps, autosampler reproducibility, and wavelength accuracy? |
| Do HPLC PMs include seal and valve changes? |
| *Laboratory Equipment Procedures-Computer Controlled* |
| Is there a written policy or procedures for validation of computer-interfaced instrumentation? |
| Have the computer systems/software, which control instruments, been validated? |
| Is there a system in place to install new software versions and to decide what scope of re-validation must be done during software upgrades? |
| Is computer software validation a part of the OQ process? |
| Is the computer software validation documented? |
| Are there systems in place to prevent installation of unauthorized software? |
| Are integrators validated using a validated signal generator? |

## 5 LABORATORY FACILITIES

As with a manufacturing area, it is essential that the laboratory be properly designed and equipped. This ensures that the data generated accurately reflect the true composition of the sample from which they came. In addition, it ensures a safe working environment for laboratory personnel.

The Laboratory Facilities sub-element includes three individual topics. A laboratory that is in compliance with CGMPs should at least address each of these topics (as applicable). These topics include (1) Laboratory Facilities-General, (2) Safety and Environmental Concerns, and (3) Laboratory Glassware.

These three topics are listed below in the form of possible questions an FDA inspector may ask when conducting an audit of the laboratory. The list of items under each topic is fairly comprehensive but may not be totally inclusive of all those components that may constitute the sub-element.

---

**LABORATORY FACILITIES SUB-ELEMENT**

*Laboratory Facilities-General*

Are the physical construction of the laboratory areas adequate for testing and all routine activities with respect to:

- Size?
- Layout and design for optimum workflow and efficiency?
- Sample receipt?
- Appropriate tables for balances/instruments, etc.?
- Location of SOPs, location of methods, specification sheets, and safety data sheets?
- Location of reagents, standards, solutions, and solvents?
- Data entry, recording, writing areas?
- Sample storage and retention?
- Refrigeration?

Are proper systems in place to minimize cross-contamination during sample preparation and laboratory testing?

Are all controlled temperature/humidity storage areas, incubators, etc. monitored to assure that proper conditions are maintained?

Have temperature monitoring systems and equipment been properly validated?

Have humidity monitoring systems and equipment been properly validated?

Are controlled substances properly segregated and controlled according to DEA requirements?

Is the environment/temperature control correct for the areas for different types of equipment?

Are radioisotopes (if used) in proper storage areas?

*Are uninterruptible power supply (UPS) units used to minimize the impact of power surges and outages?*

If an UPS exists, has it been qualified and has an SOP describing its use been written?

Are piping systems specified at the correct pressure, and have they been cleaned, flushed, leak tested passivated (sanitary use), and certified?

Have appropriate hood air filters been installed and smoke tests performed?

Have facility monitoring systems been validated?

Have purified water systems been validated?

Is there an SOP for lab or engineering maintenance of the air system?

*Are there appropriate controls/specifications for handling particulate matter and microbes available?*

Is there control of HVAC in instrument rooms, balance area, etc.?

Does the hood exhaust into an appropriate location (e.g., a sufficient distance from the air intake system as to not interfere with lab operations)?

Is the HVAC for stability/retention samples adequately monitored?

Has the laboratory water system been properly qualified?

Is there routine monitoring to assure adequate quality of the water?

Are the types of water required for the lab identified, and are specifications properly set?

Is lab water part of the site water monitoring program, and is testing frequency established and appropriate?

Are there action plans for following water related out of specification results or deviations?

Is there a schedule for servicing the laboratory water system?

**TABLE** (*Continued*)

| |
|---|
| *Safety and Environmental Concerns* |
| Is there a safety and environmental SOP? |
| Is there testing of hoods? |
| Are hazardous materials identified and segregated? |
| Are safety data sheets available, and are they kept current? |
| Is there a program that describes handling and disposal of biohazardous waste? |
| Are toxic/dangerous chemicals handled and disposed of according to local, state, and federal regulations? |
| Are employees trained to handle toxic/dangerous materials? |
| *Laboratory Glassware* |
| Has glassware washing/cleaning been validated to include demonstrating the removal of chemical residue? |
| Is there a PM program for mechanical glassware washers? |
| Is there a program in place to check for defective and worn glassware? |

## 6 METHODS VALIDATION AND TECHNOLOGY TRANSFER

Methods validation is the process of demonstrating that analytical procedures are suitable for their intended use. The methods validation process for analytical procedures begins with the planned and systematic collection validation data to support analytical procedures. In addition, the methods validation process is considered to be complete once the methods have been successfully transferred from the originating to the receiving or end user laboratory.

The Methods Validation and Technology Transfer sub-element includes at least three individual topics. A laboratory that is in compliance with CGMPs should at least address each of these topics. These topics include (1) Validation of Analytical Methods-General, (2) Cleaning Methods Validation, and (3) Procedures for Methods Transfer.

These three topics are listed below in the form of possible questions an FDA inspector may ask when conducting an audit of the laboratory. The list of items under each topic is fairly comprehensive but may not be totally inclusive of all those components that may constitute the sub-element.

**METHODS VALIDATION AND TECHNOLOGY TRANSFER SUB-ELEMENT**

*Validation of Analytical Methods-General*

Is there a general SOP for methods validation?

Are the requirements in the SOP consistent with FDA/ICH/USP guidelines?

Are methods validation protocols used for each validation?

Are all of the appropriate elements of validation included in the methods validation protocol such as:

- Accuracy?
- Precision (repeatability and intermediate precision)?
- Selectivity/specificity?
- Robustness?
- Linearity?
- Range?
- Limit of detection?

**TABLE** (*Continued*)

- Limit of quantitation?
- System suitability?
- Sample solution stability?
- Completeness of extraction (from dosage forms)?
- Filter retention studies?
- Forced degradation studies?

Is there an SOP or section within an SOP, which discusses the requirements for the verification of compendial methods?
Have compendial methods been shown to work under the actual conditions of use?
Have impurity methods been properly validated? Are they stability indicating and selective or specific?
Are methods validation reports published following validation of an analytical method?
Can results presented in the methods validation reports be tracked back to raw data?
*Cleaning Methods Validation*
Is there an ongoing cleaning validation program in place?
Is cleaning validation conducted according to a master plan and schedule?
Are there validated sampling procedures developed to support cleaning validation studies?
Are there written acceptance criteria?
Were recovery studies conducted?
Are there validated analytical methods developed to support cleaning validation studies?
Does the cleaning validation testing include analysis of area swap samples?
*Procedures for Methods Transfer*
Are there site policies and procedures governing transfer of methods, and do these documents include defining acceptance criteria?
Are the responsibilities for who is in charge clearly defined (e.g., receiving laboratory or originating laboratory)?
Is there a requirement to form an analytical technology transfer team or similar organization to transfer methods?
Is a formal receiving laboratory readiness assessment performed to include evaluation of equipment, facilities, and personnel?
Is the receiving laboratory given the opportunity to review the analytical methodology prior to the transfer process?
Is an analytical technology transfer qualification strategy document generated and used to guide the transfer of methods?
Are personnel in the receiving laboratory given the opportunity to use the method under the supervision of the originating personnel subject matter experts prior to formal transfer?
Is an analytical technology transfer protocol, with pre-determined acceptance criteria, used to formally transfer the methodology?
Is a reviewed and approved technology transfer report generated upon successful execution of the transfer protocol?
Can results presented in the cleaning methods validation reports be tracked back to raw data?

## 7 LABORATORY COMPUTER SYSTEMS

In a modern analytical laboratory, nearly every instrument either is controlled by or has data residing and processed by a computer. Subsequently, it is critical that reliability of these systems and their associated software be evaluated.

The Laboratory Computer Systems sub-element includes at least seven individual topics. A laboratory that is in compliance with CGMPs should address each of these

topics (as applicable). These topics include (1) Laboratory Computer Systems-General, (2) Centralized and Network-Attached Data Systems, (3) Stand-Alone Data Systems, (4) SOPs and Records, (5) Laboratory Information Management System (LIMS), (6) Spreadsheets, and (7) Other Systems.

These seven topics are listed below in the form of possible questions an FDA inspector may ask when conducting an audit of the laboratory. The list of items under each topic is fairly comprehensive but may not be totally inclusive of all those components that may constitute the sub-element.

---

**LABORATORY COMPUTER SYSTEMS SUB-ELEMENT**

*Laboratory Computer Systems-General*

Is there an inventory of laboratory data systems (both centralized and benchtop systems) including any associated servers?
Is a disaster recovery plan in place addressing laboratory data systems?
Has it been tested?
Is documentation for the plan and any tests available for review?
Have all GMP laboratory data systems and associated file servers, etc. been validated in accordance with applicable standards and procedures?
Have all GMP laboratory data systems and associated file servers, etc. been reviewed for compliance with 21 CFR Part 11?
Is there a remediation/replacement plan for those that were found to be deficient?
Are GMP workstations clearly labeled as such?
Is access to laboratory workstations provided through operating systems that have individual user accounts?
Are passwords properly controlled and protected by the end users?
Are passwords properly managed with assignment of permissions and privileges?
Are workstations configured with password-controlled screensavers or other automatic locking mechanism?
Are data structures for laboratory systems documented?
Does this include defining folders/directories on local hard drives and file servers where data are stored?
Do system SOPs specify these folders/directories for data storage as appropriate?
Are procedures in place to govern data naming conventions for projects, analyses, products, etc.?
Is data “archiving” performed on one or more laboratory data systems?
Is the distinction between archiving and backing up data understood by laboratory staff?
Is archived data subject to formalized retention standards?
Has the de-archival process ever been tested?

*Centralized and Network-Attached Data Systems*

Are backups of data files made regularly?
Is it governed by SOP?
Is there an individual who is assigned this responsibility?
Is the backup media securely stored, with inventory data that are current and accurate?
Are there documented service-level agreements with the central support organization defining mutual expectations and responsibilities?
If backups are performed by an automatic procedure provided by a central support organization, is clear documentation in place to define expected data structures and problem-alerting/resolution procedures?
Are data restores governed by an SOP?
Are the restores documented as required by SOP?

**TABLE** (*Continued*)

Does the central support organization have server/application SOPs, which address data management, incident management, performance monitoring, etc., specific to the laboratory?

*Stand-Alone Data Systems*

Are backups performed either automatically (scheduled through a system procedure) or manually as defined by SOPs?

Is there a procedure in place to determine whether backup media have been modified?

Are backup media identified and securely stored?

Are the restores documented as required by SOP?

*SOPs and Records*

Do SOPs exist for validation of laboratory data systems?

Have appropriate laboratory staff been trained on the laboratory data system SOPs?

Are training records for personnel trained on the SOPs available for review?

Do SOPs exist for operation of CGMP laboratory data systems?

Have appropriate laboratory staff been trained on the SOPs for CGMP laboratory data system operation?

Are training records for personnel trained on the SOPs available for review?

*Laboratory Information Management System (LIMS)*

If a LIMS is used for data management (either custom or commercial system), has it validated?

Has the validation been properly documented, and are data available for review?

Is the system operated and maintained by laboratory staff or a systems support organization?

Is the system maintained under change control that is delineated in an SOP?

Have lab personnel been trained on LIMS data entry, and are training records available for review?

Are there automated data entry and extraction points in the LIMS? Have the links been validated for:

- Bar code readers?
- GC data stations?
- HPLC data stations?
- Balances?
- Spreadsheets?
- Other components?

Is there a definition of what are considered raw data?

Can data entries be made by only authorized, password-protected individuals?

Are electronic signatures used in the LIMS, and are they in compliance with 21 CFR Part 11?

*Spreadsheets*

How are spreadsheets used in the laboratory, and if so, do they reside on the following:

- Networks?
- Workstation hard drives?

Are procedures in place governing the development/modification of spreadsheets for GMP use?

Are the spreadsheets in use unalterable and qualified/validated?

Are procedures in place to ensure that correct versions of spreadsheets are being used?

Are there procedures in place for archiving and retrieving retired versions of spreadsheets?

*Other Systems*

Is an electronic document management system used, and if so, is it validated?

Is an electronic training-record database used, and if so, is it validated?

Are laboratory periodic maintenance and calibration activities scheduled and recorded by a maintenance management system, and if so, is it validated?

## 8 LABORATORY INVESTIGATIONS

CGMPs exist to not only prevent problems from occurring but also mitigate and address deficiencies when they do occur. The Laboratory Investigations sub-element is designed to systematically identify the root cause of failures, to determine how to correct the problems, and to prevent them from re-occurring in the future.

The Laboratory Investigations sub-element includes at least three individual topics. A laboratory that is in compliance with CGMPs should address each of these topics (as applicable). These topics should include (1) Laboratory Investigations-General, (2) Laboratory Investigations-Execution, and (3) Laboratory Investigations-Documentation.

These three topics are listed below in the form of possible questions an FDA inspector may ask when conducting an audit of the laboratory. The list of items under each topic is fairly comprehensive but may not be totally inclusive of all those components that may constitute the sub-element.

---

**LABORATORY INVESTIGATIONS SUB-ELEMENT**

*Laboratory Investigations-General*

Is there an SOP that defines in detail when and how laboratory investigations are conducted?
Does the SOP distinguish between OOS and deviations or atypical events?
*Does the SOP have a flowchart and/or decision tree to assist in performing laboratory investigations?*
Does the SOP have a checklist to assist in determining the root cause of the deviation or failure?
Does the checklist include review of specifics such as:

- Use of proper test methods and proper execution of the test method?
- Performance of proper calculations without mathematical error?
- Use of proper reagents, standards, and glassware?
- Execution of proper sample preparation?
- Execution of proper standard preparation?
- Analysis of proper samples?
- Execution of proper analytical technique such as mixing, shaking, centrifuging, filtering weighing, and dilution?
- Use of proper equipment?
- Use of properly qualified and calibrated equipment?
- Special considerations for proper execution of HPLC methods?
- Special considerations for proper execution of GC methods?
- Special considerations for proper execution of dissolution testing?
- Special considerations for proper execution of IR analysis?
- Special considerations for proper execution of UV/Vis spectroscopy analysis?
- Special considerations for proper execution of titrations including Karl Fischer?

Are OOS investigations handled in a manner consistent with the Barr decision and FDA guidance documents?
Does the SOP prevent release of product by continuous testing, re-testing, or re-sampling (e.g., testing into compliance)?
Does the SOP clearly define situations when re-analysis or re-testing is appropriate, and provide guidelines on how it will be executed?
Does the SOP address special circumstances such as unidentified chromatographic peaks or shifting retention times of known peaks?
Are roles and responsibilities for executing laboratory investigations clearly defined in the SOP?
Are all laboratory personnel trained in the proper execution of laboratory investigations?

*Laboratory Investigations-Execution*

**TABLE** (*Continued*)

| |
|---|
| Is supervisor immediately notified of an OOS, deviation, or atypical event? |
| Are test material (glassware, test solutions, standard solutions, etc.) retained? |
| Is equipment and instrumentation left in the same conditions as when the event that cause the investigation occurred? |
| Are investigations performed in a timely manner? |
| Are the individuals responsible for generating the OOS, deviation, or atypical event involved in the investigation? |
| Is a concerted effort made to find a root or assignable cause of all laboratory investigations? |
| Are investigations closed in a timely manner? |
| If the OOS is discovered after product release, is there a system in place to evaluate the impact of product that is in the market? |
| Is the Quality Assurance Unit integrated into the laboratory investigations process? |
| If laboratory investigations cannot be completed in their allotted time frame, is there a formal mechanism in place to extend their due date? |
| *Laboratory Investigations-Documentation* |
| Is software (e.g., tracking, trending, and reporting) used to complete laboratory investigations? |
| Has the software been validated? |
| Is there a standard report or forms that are used for documenting laboratory investigations? |
| Are investigations supported with sufficient data to include chromatograms, spectra, raw data, etc.? |
| Is there a process for implementing corrective and preventive actions upon closure of laboratory investigations? |
| Are laboratory investigations tracked and trended? |
| Are the trending reports used during annual product reviews? |
| If the root cause of the investigation is not determined, is a conscious and rational decision on the impact of continued testing made? |
| Can final investigation reports be understood from someone outside the organization and stand on its own with limited verbal explanation? |

## 9 CONCLUSION

As shown in the sections above, the laboratory control system is actually a collection of "systems within a system." This chapter has outlined seven potential sub-elements within the laboratory control system. Depending upon your individual laboratory, there may be additional sub-elements that you may identify. The intent of the questions posed for each sub-element is to encourage the reader to begin thinking of the laboratory control system as a series of interrelated sub-elements, each of which is needed for the entire laboratory to operate in a state of control. In the end, if you address the questions highlighted in the seven sub-elements listed, your laboratory will be in control and comply with the requirements of 21 CFR parts 210 and 211, the Current Good Manufacturing Practice Regulations.

## 10 GLOSSARY OF TERMS

The following terms are commonly encountered while working in a CGMP laboratory:

*21 CFR Part 11* Title 21, Code of Federal Regulations, Part 11. Part 11 established the criteria under which electronic records and signatures will be considered equivalent to paper records and handwritten signatures in manufacturing processes regulated by the FDA.

*483 (Form 483)* The designation of the FDA form which is used to capture and report observations of CGMP deficiencies related to FDA on-site audits. These observations are compiled and become part of the Facilities Inspection Report. Form 483 observations become part of the public record.

*Abbreviated New Drug Application (ANDA)* Application made to the FDA for a generic version of a branded drug. The duplicate drug must be bioequivalent to the branded version of the drug. The first company to receive FDA approval of an ANDA receives 180 days of market exclusivity for the generic drug.

*Absorption, Distribution, Metabolism, and Excretion/Toxicology (ADME/T)* The objective of ADME/T testing is to measure what happens to a compound in the human physiology. ADME/T tests are done during the preclinical stage of the drug discovery process and are a necessary part of any clinical trial, prior to filing an IND.

*Acceptance Criteria* Numerical limits, ranges, or other suitable measures used to determine the acceptability of the results of analytical procedures.

*Accuracy* Expresses the closeness of agreement between the value found and the value that is accepted as either a conventional true value or an accepted reference value. It may often be expressed as the recovery by the assay of known, added amounts of analyte.

*Action Level/Alert Level* Alert level is used to identify the point at which a parameter has drifted toward the extreme of the specified operating range. Action level is when the parameter has drifted outside of the specified operating range. Alert and action levels must be tighter than registration specifications. Alerts are reported to management and evaluated. If an action level is reached, it is reported to management and investigated, and a corrective action initiated.

*Active Pharmaceutical Ingredient (API)* Also known as drug substance. Any component that is intended to furnish pharmacological activity or other direct effect in the diagnosis, cure, mitigation, treatment, or prevention of disease, or to affect the structure of any function of the body of humans or other animals.

*Adverse Drug Reaction (ADR)* An adverse drug reaction is an undesirable effect associated with the use of a drug. It may occur as a standard part of the drug's action, or it may be unpredictable.

*Agonist* An agonist is a substance or drug that can interact with a receptor and initiate a physiological or pharmacological response characteristic of that receptor.

*Analytical Performance Characteristics* A term used by the USP, analytical performance characteristics refer to those characteristics of an analytical method that define its performance as an analytical technique. These performance characteristics include accuracy, precision, specificity, detection limit, quantitation limit, linearity, and range. These characteristics need to be considered when validating anyone of the USP method categories.

*Antagonist* An antagonist is a drug or compound that opposes the physiological effects of another. At the receptor level, it is a chemical entity that opposes the receptor-associated responses normally induced by another bioactive agent.

*Aseptic Processing* A method of producing a sterile (devoid of living organisms) product in which sterile bulk drug or raw materials are compounded and assembled with sterile packaging components.

*Assay* A laboratory test or technique to identify and/or measure the amount of a particular substance in a sample, or for determining characteristics such as composition, purity, activity, and weight. Used to determine whether compounds (drugs, chemicals, etc.) have the desired effect in a living organism, outside the organism, or in an artificial environment.

*Atypical Result* Results generated on a test article that are within specifications but are inconsistent with previous data, established trends, or other results for the same sample on test.

*Audit* An effort to determine through investigation the adequacy of and adherence to established procedures, instructions, specifications, codes, and standards or other applicable requirements as well as the effectiveness of their implementation.

*Audit Trail* An automatic feature of some computer programs or operating systems that creates a running record of all transactions.

*Batch* A specific quantity of a drug or other material that is intended to have a uniform character and quality within specified limits, and is produced according to a single manufacturing order during the same one cycle of manufacture.

*Batch Record* A record prepared for each batch of drug product or API produced that includes complete information relating to the production and control of each batch.

*Bioanalysis (Bioanalytical)* Analytical procedure(s) to quantify results of target analyte(s) in a defined biological matrix.

*Bioassay* The determination of the biological activity of a drug by observing its effect on an organism (or organ) compared to a standard preparation.

*Bioavailability* A measure of the uptake of an ingested substance by the body as assessed by its concentration in the blood. The rate and the extent of its appearance in the blood are important determinants of bioavailability.

*Bioequivalence* A scientific basis on which generic and brand name drugs are compared with one another. Drugs are bioequivalent if they enter circulation at the same rate when given in similar doses under similar conditions. Proof of bioequivalence is crucial for generic drugs and must be demonstrated in ANDAs.

*Biological* According to the FDA's CBER, a biological product is any virus, therapeutic serum, toxin, antitoxin, vaccine, blood, blood component or derivative, allergenic product, or analogous product applicable to the prevention, treatment, or cure of diseases or injuries to humans. Biological products include, but are not limited to, bacterial and viral vaccines, human blood plasma and their derivatives, and certain products produced by biotechnology, such as interferons and erythropoietin.

*Biologics License Application (BLA)* An application to the FDA for a license to market new biological products in the United States.

*Biotechnology* The industrial use of living organisms (or parts of living organisms) to produce foods, drugs, or other products. The oldest biotechnologies include fermentation and plant and animal hybridization.

*Blank* A sample or standard of a particular matrix or composition without analyte.

*British Pharmacopoeia (BP)* The authoritative, current collection of standards for UK medicinal substances and the official source of all UK quality standards. It is used by individuals and organizations involved in pharmaceutical research, development, manufacturing, and testing.

*Calibration* The demonstration that a particular instrument or device produces results within specified limits by comparison with those produced by a reference or traceable standard over an appropriate range of measurements.

*Change Control Procedure* A procedure describing measures to be taken for controlling and maintaining an audit trail when changes are made to any part of a system (e.g., standard operating procedure, test method, or specification).

*Calibration Curve* A calibration curve is a plot of standard solution concentration, on the x-axis, vs. instrument response, on the y-axis. In chromatographic analyses, calibration curves are generated by analyzing standard analyte solutions of known concentration and measuring the resulting chromatographic peak area. The resulting plot is then used to determine the concentration of unknown sample solutions containing the same analyte. This is done by measuring the unknown peak area ($y$) and using the equation for the line to solve for the concentration of the unknown ($x$). Although referred to as a "curve," it is usually a linear plot with a well-defined slope and y-intercept.

*Capacity Factor $k'$* A dimensionless quantity used to describe the retention of a compound. It is calculated by the following formula: $(t_r - t_0)/t_0$, where $t_r$ is the measured retention time of the component of interest and $t_0$ is the retention time of an un-retained component. $t_0$ is most measured at the first disturbance of the baseline in HPLC analyses. $t_r$ is measured at the peak apex. $k'$ is a normalized value for retention. Values range between 2 and 10 for acceptable chromatography.

*CBE 30 Change by Effect, 30 Days* Supplemental changes to applications that do not require prior approval by the FDA. These changes may be implemented within 30 days following submission to the FDA if the agency has no comments.

*Center for Biologics Evaluation and Research (CBER)* The FDA successor to the Bureau of Biologics concerned with biologic drugs and with the new protein and peptide drugs arising from biotechnology. According to its charter, the CBER is responsible for ensuring the following:

- The safety of the U.S. entire blood supply and the products derived from it
- The production and approval of safe and effective childhood vaccines, including any future AIDS vaccines
- The proper oversight of human tissue for transplantation

- An adequate and safe supply of allergenic materials anti-toxins
- The safety and efficacy of biological therapeutics, including an exciting new array of biotechnology-derived products used to treat diseases such as cancer and AIDS

*Center for Drug Evaluation and Research (CDER)* The successor to the Bureau of Drugs of the FDA concerned with all SVPs (Small-Volume Parenterals), LVPs (Large-Volume Parenterals), and non-biological or small-molecule drugs.

*CGMPs* Current Good Manufacturing Practices. Section 21 Code of Federal Regulations parts 210 and 211. Federal regulations that describe the minimum CGMPs for preparation of drug products for administration to humans and animals. They include methods to be used in and the facilities or controls to be used for the manufacturing, processing, packing, or holding of a drug to assure that such drug meets the requirements of the act as to safety, has the identity and strength, and meets the quality and purity characteristics that it purports or is represented to possess.

*Check Standard* A second preparation of the working standard that is analyzed as part of the system suitability run. The check standard is prepared at the same concentration as the working standard. Prior to continuing the chromatographic run, the ratios of the response factors (response factor = area/concentration) for the working standard and the check standard are calculated. RF check standard/RF working standard should normally be within $\pm 2.0\%$. This provides assurance that the working standard was prepared correctly.

*Chemistry, Manufacturing, and Controls* Known as CMC, it is the section on a BLA, IND, NDA, or ANDA describing the composition, manufacture, and specifications for drug product and its ingredients.

*Chiral Compound* Molecules that have the same chemical formula and the same structural formula but are mirror images to each other. Stereochemically speaking, they exhibit "handedness." Molecules with the same empirical formula that are not superimposable on their mirror images are chiral.

*Chromatography* Separation of a mixture of substances by size, charge, or other characteristic by allowing the mixture to partition between a moving phase and a stationary phase.

*Clinical Trial* Carefully designed investigations of the effects of investigational new drugs, medical treatments, or devices on a group of patients. There are typically three phases of clinical trials prior to the filing of an NDA, sometimes followed ongoing post-approval studies. See Phase I, Phase II, Phase III, and Phase IV.

*Compendial Test Methods* Test methods that appear in official compendia such as the United States Pharmacopoeia (USP/NF).

*Complaint* Any verbal, written, or electronic report that alleges deficiencies related to the identity, strength, quality, purity, or effectiveness of a product after it has been released for distribution.

*Compliance* A state of operations that ensures activities is following documented procedures.

*Component* Any ingredient intended for use in the manufacture of a drug product, including those that may not appear in such drug product and primary packaging components.

*Compounding* Bringing together into homogenous mix of active ingredients, excipients, and, as applicable, solvent components.

*Consent Decree* A voluntary legal agreement a drug firm enters into with the FDA for the expressed purpose of correcting deficiencies related to CGMPs.

*Computer Systems Validation* Confirmation by examining and providing objective evidence that computer system specifications conform to user needs and intended uses, and that all requirements can be consistently fulfilled.

*Contract Manufacturing Organization (CMO)* A company that offers manufacturing services, with volume capabilities ranging from small amounts for preclinical R&D to larger volumes necessary for clinical trial purposes and commercialization.

*Contract Research Organization (CRO)* A company involved in performing clinical research on a contract basis for a pharmaceutical company, research organization, or other health organization. CROs are contracted to perform some or all of the duties by the sponsor for a clinical trial; examples include monitoring the trial, enrolling patients, performing statistical analysis, and writing the protocols. By some definitions, CROs may also perform CMC development work.

*Degradation Product* A molecule resulting from a chemical change in the drug molecule brought about over time and/or the action of light, temperature, pH, water, etc. or by reaction with excipient and/or the immediate container–closure system.

*Detection Limit* The detection limit (DL) or limit of detection (LOD) of an individual procedure is the lowest amount of analyte in a sample that can be detected but not necessarily quantitated as an exact value. The LOD is a parameter of limit tests (tests that only determine if the analyte concentration is above or below a specification limit). In analytical procedures that exhibit baseline noise, the LOD can be based on a signal-to-noise ratio (3:1), which is usually expressed as the concentration (e.g., percentage, parts per billion) of analyte in the sample.

*Development Report* A report that summarizes the major stages of drug product or API development from early stages through large-scale manufacturing.

*Documentation* Any combination of text, graphics, data, audio, or video information that can be used to clearly and completely recreate an activity, event, or process. Documentation includes life cycle documents as well as records.

*Document Control System* A system for managing preparation, review, approval, issuance, distribution, revision, retention, archival, obsolescence, and destruction of life cycle documents.

*Dosage Form* The form in which the drug is delivered to the patient. Dosage forms include parenteral, topical, tablet, oral (liquid or dissolved powder) suppository, inhalational, and transdermal.

*Drug Substance* See Active Pharmaceutical Ingredient (API).

*Drug Product* The combination of API and excipients that are processed into a dosage form and marketed to the public. Common examples include tablets, capsules, oral solutions, etc. Also referred to as finished product or dosage form.

*Efficacy* Measures the power to produce, in a controlled setting such as a clinical trial, a stated effect typically attributable to a known physiologic phenomenon. It is important to derive measures of efficacy because, with appropriate statistical achievements, these data can approximate real-world effectiveness.

*Effective Date* Date by which the approved standard or procedure shall be implemented and in use. All required training must be completed prior to this date.

*Endpoint* In clinical trials, a parameter used to compare the results in different arms of the trial. Endpoints may be directly related to the condition (such as progression of the disease) or may be measurements of surrogate markers.

*European Pharmacopoeia* Official compendium of the member states of the Council of Europe, including all EC and European Free Trade Association (EFTA) countries.

*Excipient(s)* An excipient is a raw material that may perform a variety of roles in a drug product (e.g., tablet press lubricant, filler, diluent, disintegration accelerator, and colorant). However, unlike the API, which is pharmacologically active, the excipient has no intrinsic pharmacological activity.

*Extraction Efficiency* Is a measure of the effectiveness of extraction of the drug substance from the sample matrix. Studies are conducted during methods validation to determine that the sample preparation scheme is sufficient to ensure complete extraction without being unnecessarily excessive. This is normally investigated by varying the shaking or sonication times (and/or temperature) as appropriate.

*Expiration Date* The date placed on the container/labels of a drug product or API designating the time during which a batch of product is expected to remain within the approved shelf life specification if stored under defined conditions, and after which it must not be used.

*Food and Drug Administration (FDA)* The U.S. Federal Agency charged with promoting and protecting the public health by helping safe and effective products reach the market in a timely way and monitoring products for continued safety after they are in use. Founded in 1938.

*Filter Study* A comparison of filter to unfiltered solutions in a methods validation to determine whether the filter being used retains any active compounds or contributes unknown compounds to the analysis.

*Forced Degradation Studies* Studies undertaken to degrade the sample (e.g., drug product or API) deliberately. These studies, which may be undertaken during method development and/or validation, are used to evaluate an analytical method's ability to measure an active ingredient and its degradation products without interference and are an integral part of the validation of a method as being specific and stability indicating.

*Formulation* The recipe that describes the quantity and identity of API and excipients that make up a drug product. For example, a 100-mg tablet of Advil® may actually weigh 165 mg total. The formulation may include 100 mg of ibuprofen

(the active ingredient), 50 mg of starch (filler), 5 mg of talc (lubricant for the tablet press), and 10 mg of iron oxide (colorant).

*Good Laboratory Practice (GLP)* A set of rules and criteria for a quality system concerned with organizational process and the conditions under which non-clinical health and environmental safety studies are planned, performed, monitored, recorded, archived, and reported. The GLP principles have been developed to promote the quality and validity of data generated in the testing of chemicals in order to facilitate their recognition for the purposes of assessment and other uses relating to the protection of human health and the environment. U.S. GLPs are found in 21 CFR Part 58.

*Good Documentation Practices* The handling of written or pictorial information describing, defining, specifying, and/or reporting of certifying activities, requirements, procedures, or results in such a way as to ensure data integrity.

*GxP* Good [*x*] Practice. Commonly refers to GLP, GMP, and GCP (good clinical practices).

*High Performance Liquid Chromatography (HPLC)* A separation technique based on a solid stationary phase and a liquid mobile phase. Separations (into distinct bands) are achieved by partition, adsorption, or ion-exchange processes depending upon the type of the stationary phase used. Each band is then profiled, as the solvent flows through an electronic detector, typically UV, fluorescence, or refractive index. Sometimes called high pressure liquid chromatography.

*ICH* The Tripartite International Conference on Harmonization, an international organization formed to establish uniform guidance within the pharmaceutical industry. For drug development and manufacture, the ICH issues guidance such as ICH Q2A *Text on Validation of Analytical Procedures* (March 1995), which is designed to instill uniformity within the industry with respect to various drug development and manufacturing issues. The ICH is composed of industry experts who work jointly and with the FDA to develop these guidance documents.

*International Organization for Standardization (ISO)* Non-governmental organization responsible for developing international standards founded in 1947. The ISO's 13,700 standards are voluntary and market driven. The ISO 9000 series has become an international reference for quality requirements in business-to-business dealings, and the ISO 14000 standards address environmental issues.

*Investigational New Drug Application (IND)* An application to the FDA to begin clinical trials of a new drug or biologic on humans. The IND gives the plan for the study and contains formulation, manufacturing, and animal test results information.

*Identification Test* An analytical method that is capable of determining the presence of the analyte and can discriminate between closely related compounds.

*In-Process Control/Test* Checks performed during production in order to monitor and, if appropriate, to adjust the process and/or to ensure that the product meets its specifications. The control of the environment or equipment may also be regarded as part of in-process control.

*Intermediate* Any substance, whether isolated or not, that is produced by chemical, physical, or biological action at some stage in the production of a drug product or API and that is subsequently used at another stage of production.

*Internal Audit* A systematic examination conducted by an internal organizational unit to determine whether quality activities and related results comply with policies, standards, and requirements and whether practices have been implemented effectively and are suitable to achieve objectives and ensure compliance with regulatory requirements.

*IQ/OQ/PQ* Installation, Operation, and Performance Qualification. The process by which laboratory equipment is properly installed and determined to be operating within specifications for its intended use. IQ/OQ/PQ is executed via a protocol with predetermined acceptance criteria.

*Label Claim* Theoretical strength of the product as given on the marketed product label.

*Laboratory Qualification* A process where a laboratory that has demonstrated that it has the systems in place necessary to properly perform the tests being conducted.

*LIMS* Laboratory Information Management System. Any computer system that is used to collect, compile, organize, and report laboratory data. LIMS systems may also be used to calculate results as well.

*Linearity* Evaluates the analytical procedure's ability (within a given range) to obtain a response that is directly proportional to the concentration (amount) of analyte in the sample. If the method is linear, the test results are directly, or by well-defined mathematical transformation, proportional to the concentration of analyte in samples within a given range. Linearity is usually expressed as the confidence limit around the slope of the regression line. The line is calculated according to an established mathematical relationship from the test response obtained by the analysis of samples with varying concentrations of analyte. Linearity may be established for all active substances, preservatives, and expected impurities. Evaluation is performed on standards. Note that this is different than *Range* (sometimes referred to as *Linearity of Method*), which is evaluated using samples and must encompass the specification range of the component assayed in the drug product.

*LIR* Laboratory Investigation Report. An investigation of any laboratory results or observation that does not meet acceptance criteria or falls outside the expected operational parameters. Similar to an OOS investigation, but does not have to do with meeting a release specification.

*Lot* A batch or any portion of a batch having uniform character and quality within specified limits; or, in case of a drug product, a continuous process, it is a specific identified amount produced in a unit of time or quantity in a manner that assures its having uniform character and quality within specified limits.

*Matrix (Sample Matrix)* The components and physical form with which the analytes of interest are intimately associated. In the case of drug product, the matrix is the combination of excipients in which the active ingredient is diluted and formed within. For example, the matrix of a transdermal patch is an adhesive, in which the

drug substance is dissolved, fixed to a plastic backing covered by a release liner. The chemical composition and physical structure of the matrix can have a substantial effect on sample preparation and extraction of the active moiety.

*Master Production Record* A document that specifies complete manufacturing and control instructions, sampling, and testing procedures, specifications, special notations, and precautions to be followed in production of the product and assures uniformity from batch to batch by specifying the batch size, components, and quantities and theoretical yields.

*Maximum Allowable Residue* The maximum allowable residue is used to calculate the acceptance criteria for cleaning verification methods. Residue limits should be practical, achievable, verifiable, and based on the most deleterious residue. Limits may be established based on the minimum known pharmacological, toxicological, or physiological activity of the API or its most deleterious component.

*Method Development* The process by which methods are developed and evaluated for suitability of use as test methods, and as precursor to validation.

*Method Qualification* Preliminary methods validation conducted during phases 1, 2, and 3 to support the drug development process and the associated release of clinical trial material.

*Method Transfer* Moving any analytical technique (chemical or microbiological) from one site/area to another.

*Methods Validation* The process by which it is established, by systematic laboratory studies, that the performance characteristics of the analytical method meet the requirements for intended analytical applications. The FDA states in their guidance document, "Methods validation is the process of demonstrating that analytical procedures are suitable for their intended use. The methods validation process for analytical procedures begins with the planned and systematic collection by the applicant of the validation data to support analytical procedures."

*Method Verification* The process by which compendial methods are determined to be suitable for analysis of a given test article by a given laboratory. As cited in 21 CFR Part 211.194(a) (2), method verification is used for USP/NF compendial methods or those compendia (and other recognized references such as the AOAC book of methods) that have documented evidence that the methods have been validated. At a minimum, this involves sample solution stability, specificity, and intermediate precision.

*New Drug Application (NDA)* An application to the FDA for a license to market new drug in the United States that sponsor companies submit NDAs after completing clinical trials on a new drug.

*OOS Investigation* The systematic and planned search for the root cause that generated an out-of-specification result. An OOS investigation should include formal reporting and a description of corrective and preventative actions taken.

*OOS Result* Out-of-Specification Result. An examination, measurement, or test outcome that does not comply with the specification or predetermined acceptance criteria.

*Percent RSD (% RSD)* Percent Relative Standard Deviation. A common expression and measure of the relative precision of an analytical method for a given set of measurements. Percent RSD is calculated by dividing the standard deviation for a series of measurements by the mean of the same sets of measurements and multiplying by 100, i.e., % RSD = (σn − 1/mean) × 100. Large percent RSDs for a series of measurements indicated significant scatter and lack of precision in the technique.

*Personnel Qualification* The combination of education, training, and/or experience that enables an individual to perform assigned tasks.

*Pharmacokinetics (PK)* The movement of drugs within biological systems as affected by uptake, distribution, elimination, and biotransformation.

*Pharmacology* The study of how drugs produce their effects. Pharmacology relies on knowledge of physiology, biochemistry, molecular biology, and other scientific disciplines.

*Phase I* Initial safety studies in humans. May include as few as 10 subjects, often in healthy volunteers, and includes PK, ADME/T, and dose escalation studies to determine some side effects. Usually open label.

*Phase II* Following initial safety (Phase I) testing, a drug is tested for efficacy, typically blind, randomized trials, in which a control group receives a placebo. Phase II testing may last from several months to two years. Phase II trials involve 100–300 subjects with the disease or condition of interest. Includes PK, dose ranging, safety, and efficacy.

*Phase III* Following Phase II testing, a drug is tested in a large-scale setting (several hundred to several thousand patients) to determine effectiveness, benefits, and the range of possible adverse reactions. Most Phase III studies are randomized and double blinded and typically take several years. Usually, two well-controlled studies are necessary to establish efficacy. Once the Phase III trials are successfully completed, a pharmaceutical company can request FDA approval for marketing the drug, by filing a New Drug Application (NDA).

*Phase IV* Following FDA approval and marketing, drug companies may conduct further studies on their products. These post-approval studies have several objectives, including comparing the drug with other drugs already in the market, monitoring the drug's long-term effectiveness, and determining additional potential uses for the drug.

*Pre-Approval Inspection (PAI)* An inspection of an establishment that performs manufacturing steps in an approved BLA or NDA and for which the applicant has submitted a supplement for significant manufacturing change or other change that ordinarily requires on-site review of the change.

*Process Impurity* Any component of the drug product resulting from the manufacturing process that is not the chemical entity defined as the drug substance or an excipient in the drug product.

*Placebo* A formulation containing all ingredients of a drug product, except the active ingredient for which the method is being developed.

*Precision* Expresses the closeness of agreement (degree of scatter) between a series of measurements obtained from multiple aliquots of a homogenous sample under the

prescribed conditions. The precision of an analytical procedure is usually expressed as the percent relative standard deviation. According to the FDA and the ICH, precision may be considered at three levels, namely:

*Repeatability* Refers to the use of the analytical procedure within a laboratory by a single analyst, on a single instrument, under the same operation conditions, and over a short interval of time. This is sometimes referred to as method precision.

*Intermediate Precision* Refers to variations within a laboratory as with different days, with different instruments, by different analysts, and so forth. Formally known has ruggedness.

*Reproducibility* The measure of the capacity of the method to remain unaffected by a variety of conditions such as different laboratories, analysts, instruments, reagent lots, elapsed assay times, and days – more succinctly, the use of the method in different laboratories. Reproducibility is not part of the expected methods validation process. It is addressed during technical transfer of the method to different sites.

*Prospective Validation* Validation conducted prior to the distribution of either a new product, or product made under a revised manufacturing process, where the revisions may affect the product characteristics.

*Protocol* An approved documented experimental design that, when executed, will demonstrate the ability of the subject method to perform as intended. Formal methods validation studies require a protocol. The protocol must have pre-assigned and approved acceptance criteria for each stage of the validation. The protocol may be a stand-alone document or may reference the methods validation SOP for specific details.

*Q* The amount of dissolved active ingredient specified in the monograph, expressed as a percentage of the labeled content of the dosage form, obtained during dissolution testing. For example, a tablet may have a specification stating that 85% of the active must be dissolved in 30 minutes of dissolution testing using USP apparatus II (paddles). In this case $Q = 85\%$. Also, a nefarious character in the Star Trek: The Next Generation TV series.

*Quality Agreement* The procedures established to ensure that a product is manufactured or a clinical trial is performed in compliance with the appropriate standards and regulatory requirements, and that the process or results are properly documented.

*Quality Assurance Unit* The Quality Assurance Unit serves the role of the Quality Control Unit as defined in 21 CFR 211.11. In this document, only the compliance function of the Quality Unit is addressed, not the testing functions. In the past, there has been some confusion with respect to Quality Assurance and Quality Control. CGMPs now generally recognize QA = compliance, and QC = testing.

*Qualification* Action of proving and documenting that equipment or ancillary systems are properly installed, work correctly, and actually lead to the expected results. Qualification is part of validation, but the individual steps alone do not constitute process validation.

*Quality System* A group of interrelated activities representing an integrated approach to philosophy and practices of manufacturing APIs and drug products to assure safety, identity, strength, purity, and quality.

*Quality Unit* An organizational unit independent of production which ensures that the manufacture, testing, storage, and distribution of drug products, active pharmaceutical ingredients, and components are performed in compliance with regulatory requirements and conformance to company policies and industry practices. See Quality Assurance Unit.

*Quantitation Limit* The quantitation limit (QL) or limit of quantitation (LOQ) of an individual analytical procedure is the lowest amount of analyte in a sample that can be quantitatively determined with suitable precision and accuracy. The quantitation limit is a parameter of quantitative assays for low concentrations of compounds in sample matrices and is used particularly for the determination of impurities and/or degradation products. It is usually expressed as the concentration (e.g., percentage and parts per million) of analyte in the sample. For analytical procedures that exhibit baseline noise, the LOQ is generally estimated from a determination of signal-to-noise ratio (10:1) and may be confirmed by experiments.

*Quarantine* The status of materials isolated physically or by other effective means to preclude their use pending a decision on their subsequent disposition.

*Range* The interval between the upper and lower concentrations (amounts) of analyte in the sample (including these concentrations) for which it has been demonstrated that the analytical procedure has a suitable level of precision, accuracy, and linearity. Range is normally expressed in the same units as test results (e.g., percent and parts per million) obtained by the analytical method. Range (sometimes referred to as *Linearity of Method*) is evaluated using samples and must encompass the specification range of the component assayed in the drug product.

*Raw Data* Raw data are defined as the original record of measurement or observation. Raw data may include, but are not limited to, printed instrument output, electronic signal output, computer output, hand-recorded numbers, digital images, hand-drawn diagrams, etc. Raw data are proof of the original measurement or observation and, by definition, cannot be regenerated once collected.

*Raw Material* Any ingredient intended for use in the production of intermediates, APIs, or drug products.

*Reference Standard* A reference standard is a highly purified compound that is well characterized. It is used as a reference material to confirm the presence and/or amount of the analyte in samples.

*Related Compounds* Related compounds are categorized as process impurities, degradants, or contaminants that may be found in finished drug products.

*Repeatability* The variation experienced by a single analyst on a single instrument. Repeatability does not distinguish between variation from the instrument or system alone and from the sample preparation process.

*Reporting Limit* The level, at or above the LOQ, below which values are not reported (e.g., reported as "<0.05% for a reporting limit = 0.05%). The reporting limit may be defined by ICH thresholds.

*Reprocessing* Introducing a previously processed material that does not conform to standards or specifications, back into the process, and repeating a step or steps that are part of the established manufacturing process.

*Resolution* A measure of the efficiency of the separation of two component mixtures. In chromatographic analyses, a resolution of >1.5, which means two peaks are separated from each other all the way to the baseline, is desirable.

*Response Factor* A measure of the signal generated by a detector normalized to the amount of analyte present. In HPLC it is usually the peak area for a given component (the response) divided by the concentration or mass that generated that response.

*Retain Sample* Reserve samples. An appropriately identified sample that is representative of each lot or batch of drug product or API that is stored under conditions consistent with the product labeling, in the same container–closure system in which the product is marketed or in one that is essentially of the same characteristics. Used for testing that may be involved in analyzing complaint samples.

*Relative Response Factor (RRF)* The ratio of the response factor of a major component to the response factor of a minor peak. This allows the accurate determination of a minor component without the need for actual standards.

*Relative Retention Time (RRT)* The normalization of minor peaks to the parent peak in a chromatogram. The RRT of the parent will be 1.0. Peaks eluting before the parent will have RRTs < 1.0. Peaks eluting after the parent will have RRTs > 1.0.

*Retrospective Validation* Validation of a process for a product already in distribution based upon accumulated production, testing, and control data.

*Revalidation* The process of partially or completely validating a method or process after changes or modifications have been made to the manufacturing process, analytical methodology, equipment, instrumentation, or other parameter(s) that may affect the quality and composition of the finished product. The USP specifically sites changes in synthesis of drug substance; changes in the composition of the drug product; and changes in the analytical procedure.

*Robustness* The measure of the ability of an analytical method to remain unaffected by small but deliberate variations in method parameters (e.g., pH, mobile phase composition, temperature, instrument settings); provides indication of its reliability during normal usage. Robustness testing is a systematic process of varying a parameter and measuring the effect on the method by monitoring system suitability and/or the analysis of samples. It is part of the formal methods validation process.

*Ruggedness* A date term, now commonly accepted as intermediate precision. See Intermediate Precision.

*Selectivity* Selectivity is the ability of the method to separate the analyte from other components that may be present in the sample, including impurities. Determination of selectivity normally includes analyzing placebo, blank, media for dissolution, dilution solvent, and mobile phase injections. Also, no chromatographic peaks, such as related compounds, should interfere with the analyte peak, or internal standard peak if applicable. Selective = separate and shows every component in the sample.

*Signal-to-Noise* In chromatography the measure of average baseline noise (e.g., peak-to-peak) to the signal given by an analyte peak. S/N calculations are performed when determining the LOD and the LOQ.

*Specification* The quality standards (e.g., tests, analytical procedures, and acceptance criteria) provided in an approved application to confirm the quality of the drug substances, drug products, intermediates, raw materials, reagents, and other components including in-process materials.

*Spiking* The addition of known amounts of a known compound to a standard, sample, or placebo, typically for the purpose of confirming the performance of an analytical procedure or the calibration of an instrument.

*Spiked Placebo* Preparation of a sample to which known quantities of analyte are added to placebo material. Performed during validation to generate accurate and reproducible samples that are used to demonstrate recovery from the sample matrix.

*Specificity* The ability to assess unequivocally the analyte in the presence of components that may be expected to be present, such as impurities, degradation products, and excipients. There must be inarguable data for a method to be specific. Specific = measure only the desired component without interference from other species that might be present; separation is not necessarily required.

*Stability* Refers to the physicochemical condition of a parenteral, biologic, or shelf life of other drugs. Certain drugs must pass USP or CFR stability tests. Manufacturers must have documentation of the potency of products under labeled storage conditions. Product stability is determined through execution of a formal stability protocol, which requires storage of the drug product, under controlled environmental conditions, in its planned marketed product containers for periods up to three years.

*Stability Indicating Methodology* A validated quantitative analytical procedure or set of procedures that can detect the changes with time in the pertinent properties (e.g., active ingredient, preservative level, or appearance of degradation products) of the drug substance and drug product.

*Stability-Indicating Assay* An assay that accurately measures the component of interest (the active ingredient(s) or degradation products) without interference from other degradation products, process impurities, excipients, or other potential interfering substances.

*Stability Indication Profile* A set of procedures or assays that, collectively, detect changes with time, although may not do so individually.

*Standard Operating Procedures (SOPs)* The description of activities necessary to respond to normal and abnormal situations in the operation system. SOPs may include troubleshooting checklists, lists of personnel to contact, etc. They should also describe normal operation, maintenance, and cleaning of the system, as well as normal operating parameters. In the laboratory, all analytical methods are required to be documented in SOPs that are commonly referred to as standard test methods (STMs).

*Standard and Sample Solution Stability* The stability of standards and samples is established under normal benchtop conditions, under normal storage conditions, and sometimes in the instrument (e.g., an HPLC autosampler) to determine if special

storage conditions are necessary, for instance, refrigeration or protection from light. Stability is determined by comparing the response and impurity profile from aged standards or samples to that of a freshly prepared standard and to its own response from earlier time points. Note that these are short-term studies and are not intended to be part of the stability indication assessment or product stability program.

*Stressed Studies* See Forced Degradation Studies.

*Supplemental New Drug Application (sNDA)* Application submitted for an already-approved NDA for any changes in packaging, labeling, dosages, ingredients, or new indications.

*System Suitability* System suitability is the evaluation of the components of an analytical system to show that the performance of a system meets the standards required by a method. A system suitability evaluation usually contains its own set of parameters; for chromatographic assays, these may include tailing factors, resolution, and precision of standard peak areas, and comparison to a confirmation standard, capacity factors, retention times, theoretical plates, and calibration curve linearity.

*Tailing Factor* A measure of peak asymmetry. Peaks with a tailing factor of $>2$ are usually considered to be unacceptable due to difficulties in determining peak start and stop points, which complicates integration. Tailing peaks are an indication that the chromatographic conditions for a separation have not been properly optimized.

*Technology Transfer* The transfer of a process or method including demonstration of equivalence according to predetermined criteria between the receiving and transferring sites.

*Test Method* An approved, detailed procedure describing how to test a sample for a specified attribute (e.g., assay), including the amount required, instrumentation, reagents, sample preparation steps, data generation steps, and calculations used for evaluation.

*Theoretical Plates* A dimensionless quantity used to express the efficiency or performance of a column under specific conditions. A decrease in theoretical plates can be an indication of HPLC column deterioration.

*Transcription Accuracy Verification (TAV)* The process where the transcription of data from one location to another is confirmed by a second party. Important with methods validation reports which summarize results in tables derived from raw data.

*United States Pharmacopoeia (USP)* A non-governmental, not-for-profit organization that promotes the public health by establishing standards to ensure the quality of medicines and other health care technologies. These standards are developed by a publicly involved process and are accepted worldwide. The USP provides standards for more than 3400 prescription and non-prescription drugs, nutritional and dietary supplements, veterinary drug standards, and health care products. These standards are published in the United States Pharmacopoeia and National Formulary (USP-NF).

*USP <1225> Category I* One of four method categories for which validation data should be required. Category I methods include analytical methods for quantitation of major components of bulk drug substances or active ingredients (including preser-

vatives) in finished pharmaceutical products. Category I methods are typically referred to as assays.

*USP <1225> Category II* One of four method categories for which validation data should be required. Category II methods include analytical methods for determination of impurities in bulk drug substances or degradation compounds in finished pharmaceutical products. These methods include quantitative assays and limit tests.

*USP <1225> Category III* One of four method categories for which validation data should be required. Category III methods include analytical methods for determination of performance characteristics (e.g., dissolution, drug release).

*USP <1225> Category IV* One of four method categories for which validation data should be required. Category IV methods include analytical methods used as identification tests.

*Validation* A documented program that provides a high degree of assurance that a specific process, method or system will consistently produce a result meeting predetermined acceptance criteria.

*Validation Characteristics* See Analytical Performance Characteristics.

*Validation Parameters* See Analytical Performance Characteristics.

*Validation Protocol* A prospective plan that, when executed as intended, produces documented evidence that a process, method, or system has been properly validated.

*Validation Report* A summary of experiments and results that demonstrate the method is suitable for its intended use, approved by responsible parties.

*Verification* The act of reviewing, inspecting, testing, checking, or otherwise establishing and documenting whether items, processes, services, or documents conform to specified requirements.

## 11 READER RESOURCES

The following selected references are good sources of information and will help the reader stay current with respect to laboratory CGMP compliance:

1. Food and Drug Administration Web site: www.fda.gov
   a. Warning Letters
   b. Guidance Documents
   c. CPGM Documents
   d. Monthly E-mail Services
2. FOI Services, Inc.: www.foiservices.com
   a. Actual Form 483 Observations (online purchase)
   b. Establishment Inspection Reports (EIRs) (online purchase)
3. International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH): www.ich.org
4. Analytical chemistry in a GMP environment. J. Miller, J. Crowther, editors. Wiley InterScience, New York, 2000.

5. "The quality audit handbook. 2nd Edition. J.P. Russell, editor. ASQ Quality Press, Milwaukee, WI, 2000.
6. A laboratory quality handbook of best practices. D. Singer, editor. ASQ Quality Press, Milwaukee, WI, 2000.
7. FDC Reports "The Gold Sheet": www.fdcreports.com
8. FDC Reports "The Pink Sheet": www.fdcreports.com
9. Contract Pharma Magazine
10. C&E News Magazine
11. Pharmaceutical Formulation and Quality Magazine
12. Lab Manager Magazine
13. AAPS NEWSMAGAZINE Magazine
14. Pharmaceutical Review Magazine
15. "Establishing a CGMP laboratory audit system: a practical guide. David M. Bliesner. John Wiley & Sons, Inc. Hoboken, New Jersey, 2006.
16. Validating chromatographic methods: a practical guide. David M. Bliesner. John Wiley & Sons, Inc. Hoboken, New Jersey, 2006.
17. The course "Introduction to Laboratory Current Good Manufacturing Practices (CGMPs): A Quality Systems Based Approach." Online purchase at www.delphianalytical.com provided in the Better-Than-Live™ electronic learning format.

## REFERENCES

1. Abc;Compliance Program Guidance Manual for FDA Staff: "Drug Manufacturing Inspections Program" 7356.002, February 1, 2001. http://www.fda.gov/AboutFDA/Centers Offices/CDER/ucm095598.htm
2. Code of Federal Regulations, Food and Drugs, Title 21 Parts 210 and 211: "Current Good Manufacturing Practice in Manufacturing, Processing, Packing, or Holding of Drugs: General" and "Current Good Manufacturing Practice for Finished Pharmaceutical." Revised April 1, 2005.
3. FDA Guidance for Industry: *Quality Systems Approach to Pharmaceutical Current Good Manufacturing Practice Regulations*, September 2006.

# *Uses of X-Ray Powder Diffraction In the Pharmaceutical Industry*

**Igor Ivanisevic, Richard B. McClurg, and Paul J. Schields**
*SSCI, a Division of Aptuit, West Lafayette, IN*

## 1 INTRODUCTION

Among the many experimental techniques available for the identification of solid forms, including polymorphs, solvates, salts, cocrystals and amorphous forms, X-ray powder diffraction (XRPD) stands out as a generally accepted "gold standard." While this does not mean that XRPD should be used to the exclusion of other experimental techniques when studying solid forms, X-ray diffraction (XRD) has applications throughout the drug development and manufacturing process, ranging from discovery studies to lot release. The utility of X-ray diffraction becomes evident when one considers the direct relationship between the measured X-ray diffraction pattern and the structural order and/or disorder of the solid. XRPD provides information about the structure of the underlying material, whether it exhibits long-range order as in crystalline materials, or short-range order as in glassy or amorphous materials. This information is unique to each structure—whether crystalline or amorphous—and encoded in the uniqueness of the XRPD pattern collected on a well-prepared sample of the material being analyzed.

One must draw a distinction between crystalline materials, which give rise to XRPD patterns with numerous well-defined sharp diffraction peaks, and glassy or amorphous

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing,*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

**TABLE 1 Types of Solid Forms Described by the Wunderlich (1) Classification System**

| Solid Form | Translation | Orientation | Conformation |
|---|---|---|---|
| Crystal | Long range | Long range | Long range |
| Condis crystal (glass) | Long range | Long range | Short range |
| Plastic crystal (glass) | Long range | Short range | Short range |
| Liquid crystal (glass) | Short range | Long range | Short range |
| Amorphous (glass) | Short range | Short range | Short range |

materials whose XRPD patterns contain typically three or less broad maxima (X-ray amorphous halos). In practice, using XRPD, one can usually measure a sequence of progressively more disordered crystalline materials that ultimately result in glass. A classification system has been proposed by Wunderlich (1), Table 1, to describe the type of structural order and molecular packing present in molecular organic solid forms using three order parameter classes: translation, orientation, and conformation.

Solid forms of a given molecule exhibiting long-range crystalline order (e.g., polymorphs, solvates, co-crystals, and salts) can be identified and characterized using XRPD by their unique combination of order parameters. Amorphous solid forms do not exhibit any long-range order but are identifiable and characterized by their unique local molecular order, apparent in the X-ray amorphous diffraction pattern (2).

Knowing that X-ray powder diffraction is sensitive to structural order, some of its typical applications in the analysis of solid-state properties of a drug substance or product include:

- Identification of existing forms of the active pharmaceutical ingredient (API).
- Characterization of the type of order present in the API (crystalline and/or amorphous).
- Determination of physical and chemical stability.
- Identification of the solid form of the API in the drug product.
- Identification of excipients present in a drug product.
- Monitoring for solid form conversion upon manufacturing.
- Detection of impurities in a drug product.
- Quantitative analysis of a drug product.

Where appropriate data are available, XRPD analysis can determine the solid-form structure and crystal-packing relationship among individual molecules in the solid. This information is essential to the understanding of solid-state chemistry of drugs and important from the regulatory perspective.

## 2 X-RAY DIFFRACTION THEORY

When dealing with organic samples, a common simplification called the first Born approximation (3,4) is useful in explaining the X-ray diffraction process. Solid forms

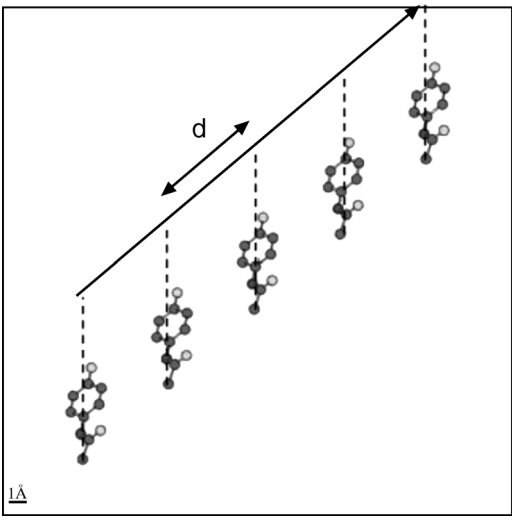


***FIGURE 1*** *A simple periodic array of molecules with a single orientation and conformation and constant spacing, d.*

of organic materials are expected to interact weakly with incident X-rays (generally true provided the crystallite size is not too large). This means that the amplitude of doubly and multiply scattered radiation will be very small and negligible when compared to the singly scattered radiation (4). In the presence of crystal defects, grain boundaries or disordered systems (i.e., typical laboratory samples and not large, perfect single crystals), the multiply scattered radiation becomes even less significant.

Respecting the limits of these assumptions, we can model the diffraction process as a Fourier transform of the electron density within the sample. Figure 1 shows a simple example containing a periodic array of molecules separated by a constant spacing, $d$, and with a single orientation and conformation. While each atom is considered a point source of scattering (5), the molecules themselves can be reduced to point sources of scattering under the assumption that the electron density distribution of a collection of atoms is the sum of the electron density distributions attributed to individual centered atoms (4). Note that atoms in a molecule are not necessarily the same as isolated, free atoms, though they are often approximated as such. The latter approximation allows us to substitute values for the Fourier transform of electron density of each individual atom using the tables of atomic scattering factors (6).

During the diffraction process, the ordered arrangement of point scattering centers in real space produces a set of diffraction events in reciprocal space corresponding to sharp peaks. A spacing of $d$ between the point centers (molecules) in real space will correspond to a peak spacing of $2\pi/d$ in reciprocal space (also called $Q$-space). This simple Fourier identity, illustrated in Fig. 2, will apply to any molecular-translational order existing within a solid form (4).

Therefore, diffracted peak positions can be expressed in terms of $d$-space, $Q$-space or, most commonly, $2\theta$, the angle between diffracted and undeviated X-ray waves (5).

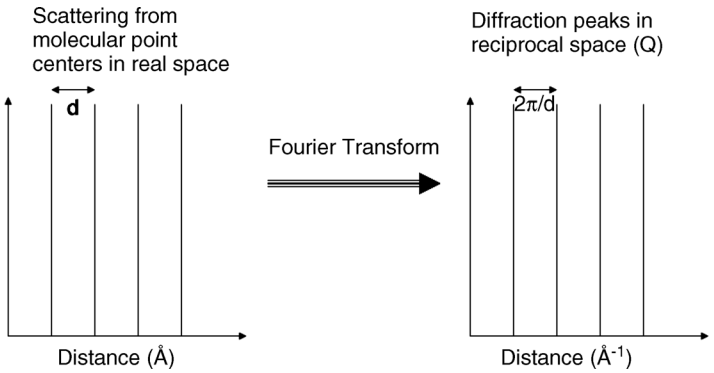


***FIGURE 2*** *A Fourier transform relates a periodic array of molecular point scattering centers in real space to a family of diffraction peaks in reciprocal space.*

Bragg's law can be used to relate the X-ray half-scattering angle $\theta$ to the aforementioned spacing parameter, $d$, as seen in Eq. 1.

$$n\lambda = 2d\sin\theta \tag{1}$$

The parameter $\lambda$ is the wavelength of the incident X-ray radiation, with a typical value of approximately 1.54 Å for a Cu X-ray radiation source. $n$ is an integer and can best be understood through the explanation of why Bragg's law holds. Consider a case of many parallel scattering planes—each of which is a collection of previously introduced point-scattering centers—reflecting incoming X-ray radiation of wavelength $\lambda$ at the angle $\theta$, as seen in Fig. 3 (5). The two diffracted waves in Fig. 3 will interfere with each other either constructively (adding together to produce stronger peaks) or destructively (subtracting from each other to some extent), depending on whether they are in phase or out of phase, respectively (4). Since the wave diffracted by the bottom plane in Fig. 3 travels a greater distance (greater by exactly $a + b$) than the wave diffracted by the top plane, that extra distance must be equal to an integer multiple of $\lambda$ for points to remain in phase and total constructive reinforcement to

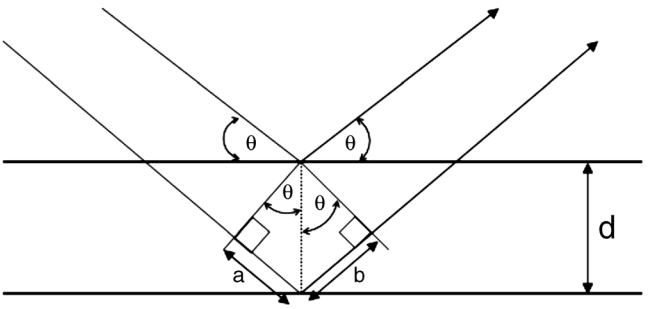


***FIGURE 3*** *Bragg's law for parallel planes.*

occur between the scattering from these planes (Eq. 2).

$$n\lambda = a + b \tag{2}$$

$a$ and $b$ can be expanded using the law of sines as $\sin\theta = a/d = b/d$. Simple substitution then transforms Eq. 2 into Bragg's law (1). The integer $n$ refers to the order of the diffraction. For additional readings on X-ray diffraction theory, see for example References (4–8).

## 3 X-RAY DIFFRACTION EXPERIMENTAL PROCEDURES

XRPD patterns display diffracted intensity as a function of the experimental parameter $2\theta$ (angle between diffracted and undeviated X-ray waves). Intensity is typically expressed in counts or counts per second while peaks are listed as positions in $°2\theta$ or $d$-spacings (usually measured in Å or nm).

### 3.1 Crystalline Materials

For materials exhibiting long-range (crystalline) order, XRPD patterns will contain sharp peaks, Fig. 4, whose shape and width will depend on the type of instrument on which the data were collected. Measurement ranges for crystalline materials depend on the type of study. For example, in a common application of XRPD—small-molecule polymorphism study—1–40°$2\theta$ is a typical measurement range used, since

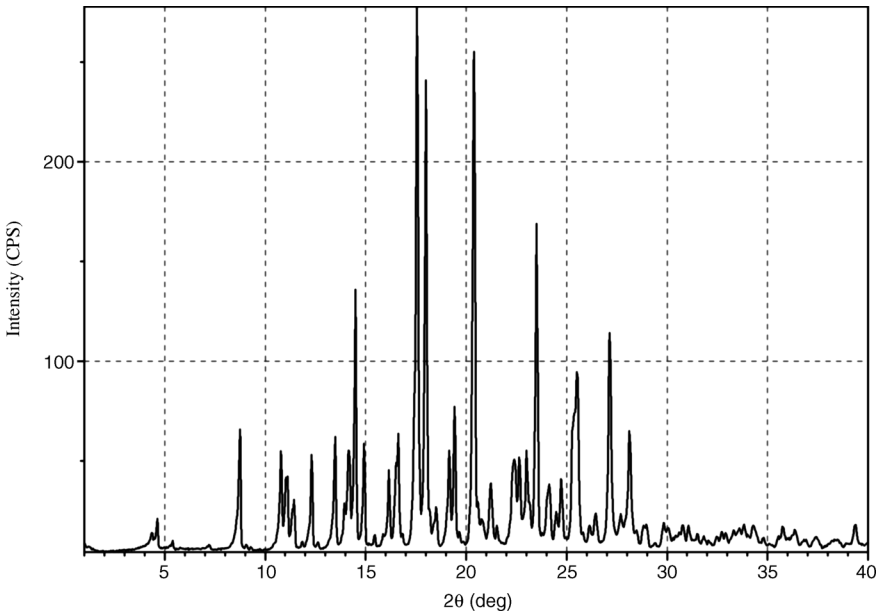


***FIGURE 4*** *XRPD pattern of crystalline itraconazole.*

peaks above approximately 30°2$\theta$ become too numerous and overlapped to be useful in differentiating between polymorphs. When working with large molecules (e.g., biologicals), it is advantageous to measure to as low an angle as allowed by the geometry of the instrument (approximately 0.5°2$\theta$ is achievable in typical laboratory settings with modern instruments, sub 0.5°2$\theta$ using a synchrotron or with a dedicated small-angle scattering instrument). Certain computational methods may require measurements to significantly higher 2$\theta$ angles, up to 100°2$\theta$.

Collection time varies per application but, for polymorphism studies, good patterns of crystalline material on modern instruments can be obtained in 2–10 min. In high throughput configurations for well plates, collection times of less than a minute per pattern can readily be achieved. Depending on the instrument geometry and application, 2–20 mg of crystalline sample may be required, and the sample can be reused afterwards as XRPD is typically non-destructive.

XRPD suffers from two common sample-related effects that can play a significant role when characterizing or identifying crystalline material. Ideal XRPD samples have large numbers of randomly oriented crystallites. The reproducibility of an XRPD pattern is dependent on particle-orientation statistics, while preferred orientation limits the degree to which a pattern accurately represents the structure, as opposed to a particular sample preparation. For those reasons, one must assess the particle statistics and degree of preferred orientation before one can be confident the peak identifications made from a particular pattern are representative of the material.

The effect of non-random (preferred) orientation of crystallites in a sample is to increase the relative intensities of some peaks and decrease the relative intensities of others. The variation in peak intensity is proportional to the degree of preferred orientation. In extreme cases, it can cause peaks to disappear completely while exaggerating otherwise faint peaks. Even in mild cases, a different set of relative intensities would result from diffractometers with different sample geometries.

Poor particle statistics are displayed by samples where a relatively small number of crystallites contribute to the integrated XRPD patterns. Since the small population of large crystallites cannot represent all possible orientations, the measured relative intensities are not reproducible. Irreproducible peak intensities lead to the expectation that the pattern may differ dramatically if the same experiment was repeated on a different sample of the same material even using the same diffractometer.

Both preferred orientation and poor particle statistics can be reduced to some extent in laboratory samples by spinning the sample in a holder (9, 10). Diffractometers allow the sample to be spun around one or two axes. In addition, a simple way of assessing the degree of preferred orientation or particle statistics in a sample is to analyze it on two different diffractometers with different spinning geometries. A summary of methods to minimize preferred orientation can be found in References (11–14). Bish (5) includes a detailed discussion on sample preparation and related problems.

## 3.2 X-Ray Amorphous Materials

In the case of X-ray amorphous (disordered, glassy or amorphous) materials, there will be no sharp peaks observed in the XRPD pattern, only broad halos (Fig. 5).

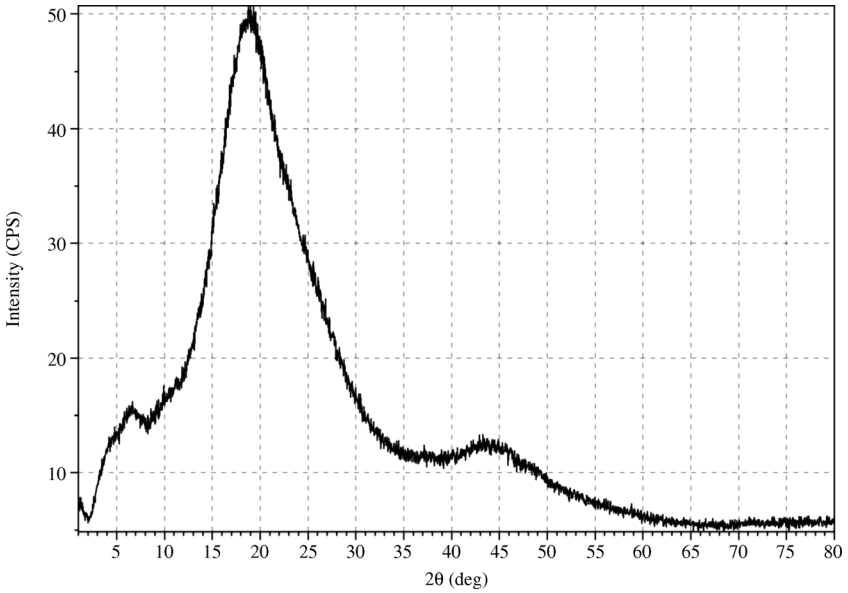


***FIGURE 5*** *XRPD pattern of amorphous itraconazole.*

Nevertheless, as we shall see later on, it is possible to extract structural information from these types of patterns using computational methods. Many of the computational methods used to analyze disordered materials require that the data are collected over a broad range, typically from 1 to 100°$2\theta$. Furthermore, because the signal-to-noise ratio in X-ray amorphous patterns is typically poor, longer collection times are often used. 30–60 min is not uncommon for a single pattern. Finally, a greater amount (50–100 mg) of amorphous sample is typically warranted to obtain a good XRPD pattern.

XRPD patterns of both crystalline and X-ray amorphous materials contain some experimental artifacts, for example, instrument-background functions, sample-holder fingerprints, incoherent (Compton) scattering, polarization and Lorentz effects, and air scatter (5). The relatively low signal generated from X-ray amorphous samples means those artifacts will represent a much greater portion of the overall diffracted intensity. Therefore, computational methods used to analyze X-ray amorphous materials (15–17) are very sensitive to experimental artifacts and care must be taken to minimize their presence. Instrument intensity-correction functions can be determined using known standards and computationally modeled thereafter. Sample holders that produce significant X-ray scattering of their own in the relevant regions (e.g., glass capillaries) should be avoided when working with amorphous materials. Instead, samples can be sandwiched between thin polymer films with low X-ray background (e.g., Etnom®). Given the chemical formula of the material, Compton scattering can be calculated from atom-scattering tables (6) and subtracted from the overall signal. Air scatter can be largely eliminated using custom-built enclosures for a helium atmosphere.

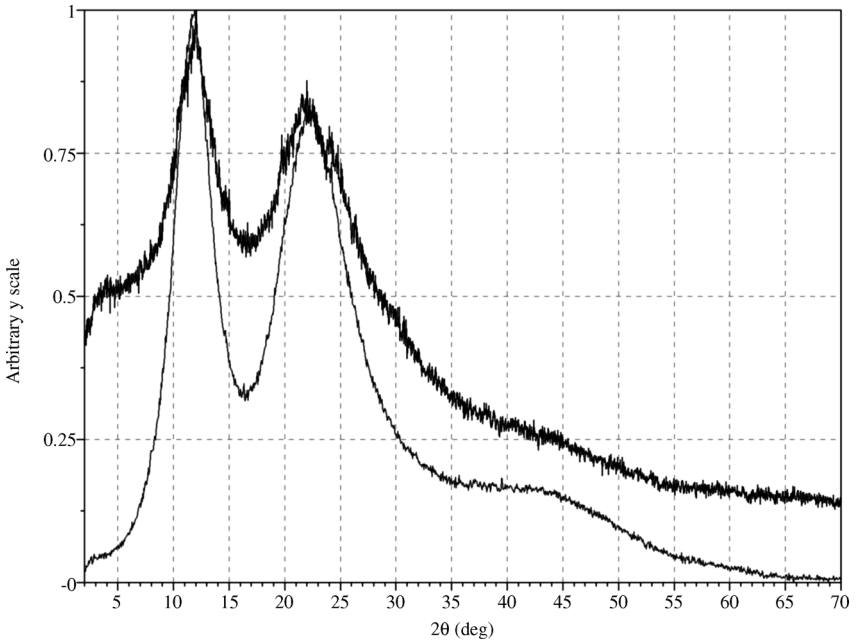


***FIGURE 6*** *The effects of experimental artifacts on amorphous X-ray patterns. Both patterns were collected from the same sample of amorphous nifedipine. The top pattern was collected from approximately 10 mg of material packed in a glass capillary under ambient conditions. The collection time was 5 min. The bottom pattern was collected from approximately 100 mg of material sandwiched between two thin sheets of Etnom® film in a He-purged enclosure. The collection time was 1 h.*

Figure 6 shows an X-ray amorphous pattern of nifedipine collected on a standard instrument with no special considerations and the sample packed in a glass capillary, compared to the same sample collected on a specially configured instrument using a low background sample holder, He atmosphere, and longer collection time. Note the patterns in this figure have not been offset but normalized to top intensity. As seen in the bottom pattern of Fig. 6, the low angle air scatter contribution is largely eliminated through the use of He. Furthermore, the broad glass scattering around $25°2\theta$ which causes a shift in the position of the second amorphous halo in the top pattern is not evident in the bottom pattern. Typically, in screening applications, no more than a few milligrams of material are available, resulting in patterns where the glass/air scatter signal overwhelms the signal from the material itself and little structure is evident.

Certain types of analyses involving crystalline materials (e.g., quantitative studies or indexing) may also benefit from careful experimental setups but the most common applications (i.e., polymorph detection or identification (18)) generally do not require the specialized setups or longer collection times used for amorphous materials.

## 3.3 Instrumentation

Laboratory X-ray diffractometers typically consist of an X-ray source, sample chamber and detector. The most commonly used X-ray source in laboratory experiments with organics is Cu, though instruments typically allow different sources to be used with some configuration. Slits and optics are used to focus the incident and diffracted radiation on the sample and detector, respectively. Specimens usually can be rotated to alleviate some of the intensity artifacts discussed earlier. Detectors can be point, line or area, with the latter offering advantages in both speed of acquisition and ability to assess the particle statistics and preferred orientation in a sample through examination of Debye rings (19). Synchrotron sources are sometimes used for specialized measurements to collect high quality data. Detailed descriptions of X-ray instrumentation can be found, for example in Reference (5).

The alignment of a diffractometer is maintained through the use of calibration standards, such as silicon, to verify the position calibration (5). Regular calibrations are accepted industry practice (cGMP), with the frequency of calibration ranging from every $X$ number of samples to set time periods (e.g., daily), depending on how much data one is willing to risk invalidating due to a failed calibration. When a silicon standard is used, the Si 111 peak position is typically verified using a short scan in the appropriate region (around 28.4°$2\theta$). Other standards may be used, for example, to verify low angle alignment.

Diffractometers can typically be operated in either reflection or transmission configurations. Reflection is by far the more common and also referred to as Bragg-Brentano geometry (Fig. 7). In reflection measurements, incoming X-rays are "reflected" off the sample surface and focused by the instrument optics onto the detector. Errors due to sample transparency to X-rays are common for organic samples (5,14). The X-rays penetrate many atomic layers below the surface meaning the average diffracting surface lies somewhat beneath the surface of the sample. This type of error can lead to peak position errors in measured patterns of as much as a tenth of a degree. Therefore, low absorbing samples are often prepared as thin films using

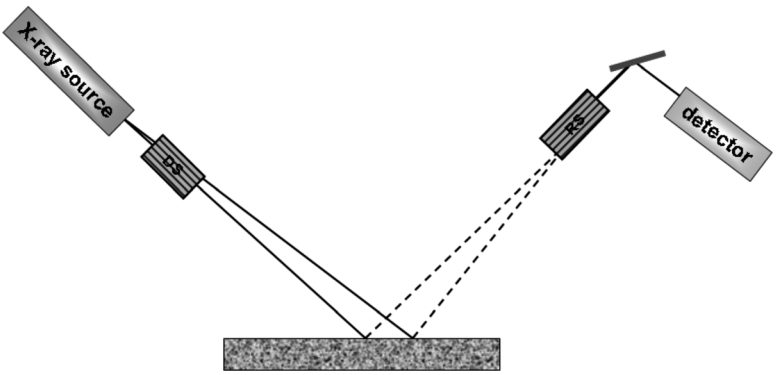


***FIGURE 7*** *Bragg-Brentano diffractometer geometry.*

a low background holder, to reduce the penetration effect. Displacement errors arise from the difficulty of preparing samples such that the surface of the sample is level with the surface of the holder (where the instrument is focused). Sample-displacement errors can result in significant, systematic shifts in measured peak positions, on the order of several tenths of a degree in particularly bad cases. Experience in preparing samples such that they are level with the surface of the holder is the only solution to this problem, although computational methods can be used to shift the pattern and correct peak positions, where the error is recognized. A further limitation of reflection measurements is the inability to measure to very low angles—typically $2.5°2\theta$ is a practical limit—making this type of measurement less useful for the analysis of large molecules which are expected to have peaks in the $0–2.5°2\theta$ range.

Transmission mode analysis overcomes many of the limitations of reflection mode when carefully configured. In transmission measurements, the incident X-rays pass through the sample and are diffracted not only on the surface but throughout the sample. This type of analysis is possible for organics due to their relative transparency to X-rays. The sample no longer needs to be level with the holder but thickness of the sample is important and can cause peak displacement errors (when the sample is too thick). Transmission configurations generally allow lower angle measurements than reflection, making them useful for large molecules. However, it becomes essential that the holder containing the sample is as X-ray transparent as possible, because its "fingerprint" will be part of the X-ray pattern collected on each sample. Disposable, low background thin polymer films are commonly used to hold the sample in transmission measurements. In addition, it is common practice to regularly collect blanks (X-rays of the holder material itself) to ensure its signature does not change between different lots. Nevertheless, the extra effort required to operate instruments in transmission mode is well worth the improvement in data quality, especially when working with X-ray amorphous materials.

## 4 APPLICATIONS OF X-RAY DIFFRACTION IN DRUG DEVELOPMENT AND MANUFACTURING

XRD has a broad range of applications in various stages of drug development and manufacturing. This section will address many of the common XRPD uses from a practical standpoint. In the broadest terms, these applications can be divided between API characterization and identification. While there is some overlap in both categories, the former is more commonly applied during drug development (before the drug is on the market) while the latter is directed more toward manufacturing, regulatory aspects and intellectual property.

### 4.1 API Characterization

Guidelines from regulatory authorities regarding the need for characterization of a drug substance under development have been clearly stated. Below is an example relating to the issue of polymorphism (20):

"Polymorphic forms of a drug substance can have different chemical and physical properties, including melting point, chemical reactivity, apparent solubility, dissolution rate, optical and mechanical properties, vapor pressure, and density. These properties can have a direct effect on the ability to process and/or manufacture the drug substance and the drug product, as well as on drug product stability, dissolution, and bioavailability. Thus, polymorphism can affect the quality, safety, and efficacy of the drug product."

While there are a number of methods to characterize polymorphs of a drug substance (21), the two broadly accepted methods of providing unequivocal proof of polymorphism recognized by the above source are single crystal X-ray diffraction and X-ray powder diffraction (20). Other techniques (e.g., thermal or spectroscopic methods) can be helpful in further characterizing drug products but only X-ray provides the necessary structural information to uniquely identify different polymorphs. Therefore, in early drug development, X-ray powder diffraction is often used as a primary experimental technique and a means of differentiating between the experimentally generated materials. Fully characterizing any material requires the use of complementary techniques (thermal or spectroscopic) but X-ray is typically done first because it is fast, non-destructive, requires little material, and provides the necessary structural information.

Synchrotron X-ray diffraction has frequently been used to characterize pharmaceutical materials in applications where additional sensitivity not provided by laboratory X-ray diffractometers may be required (e.g., crystallization monitoring (22,23)). The tradeoff is the greater expense and time investment typically associated with such measurements. Since such applications tend to be specialized, this section will focus primarily on laboratory XRPD methods.

***4.1.1 Qualitative Analysis of Materials (Phase Identification)*** Every structurally different crystalline material will exhibit a unique XRPD pattern upon analysis (14). Therefore, the use of XRPD for phase identification was recognized early and remains the most common application of XRPD to pharmaceuticals. This so-called qualitative analysis typically refers either to the initial characterization of material previously not analyzed by XRPD or to the identification of a phase or phases in a sample of material by comparison to reference patterns. Reference patterns are previously collected XRPD patterns of the same material. Where available, XRPD patterns calculated from, for example, single crystal structures can be substituted but one should remember that the temperature at which the pattern is calculated can have a significant effect on the calculated XRPD profile. When dealing with mixtures of phases, qualitative analysis can provide an estimate of the relative proportions of different phases in the sample, usually based on the comparison of peak intensities for characteristic peaks of the different phases. Due to sample artifacts such as preferred orientation and poor particle statistics, this type of analysis should never be confused with quantitative analysis of mixtures, addressed later in this chapter. Databases of known XRPD patterns for various pharmaceutical materials are published annually by the Centre for Diffraction Data (ICDD) and the Cambridge Crystallographic Data Centre (CSD).

XRPD patterns are typically compared by overlaying and aligning the data from different samples. This procedure is typically done electronically, either using the software provided by the XRPD instrument vendor or using custom-developed software. The primary assessment might include determining whether each sample is X-ray amorphous or crystalline based on the absence or presence of crystalline peaks, respectively. When comparing XRPD data of crystalline samples, one notes any differences in peak positions (to within a certain precision, e.g., 0.1 °$2\theta$ (24)) which correspond to structural differences between the samples. Intensities are generally not relied upon for qualitative analysis due to previously mentioned instrument and sample artifacts, although they have to be used to some degree to allocate the peak positions (based on local maxima).

It is not uncommon for two patterns to share some but not all of the peak positions. This can be a coincidence or it can be due to one of the samples being a mixture of multiple phases, including the phase in the other sample. Experience and data from complementary experimental techniques are needed to resolve such ambiguous cases. It should also be noted that, at higher $2\theta$ values, peaks of most organic materials become considerably overlapped and determining their exact positions becomes difficult. Therefore, free-standing peaks at low angles are the primary means of differentiating structures and XRPD data above approximately 30 °$2\theta$ are rarely useful for qualitative analysis.

Figure 8 shows XRPD patterns of two crystalline polymorphs of sulfamerazine. The patterns in Fig. 8 were collected from material crystallized in glass capillaries during a polymorphism screen. A polymorphism screen is typically run early in the

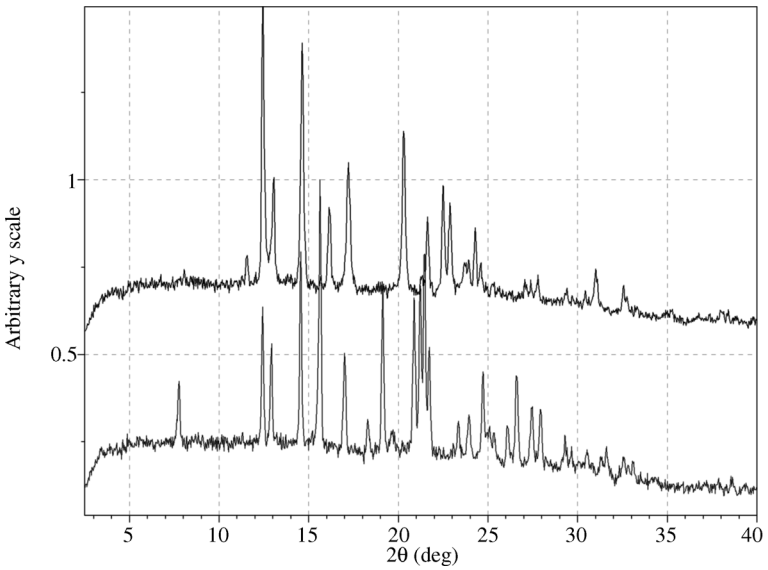


***FIGURE 8*** *Two polymorphs of sulfamerazine. Patterns offset for clarity.*

drug development process to identify and (partially) characterize the different polymorphs of a drug substance. Assuming XRPD was the first analytical technique used on the samples, the data in Fig. 8 could be used to make a qualitative assessment regarding the probable nature of the material generated during crystallization experiments. Therefore, one could designate the first of the patterns crystalline "Pattern A" and the other crystalline "Pattern B," noting the sharp peaks and lack of diffuse halos as a sign of crystallinity and the structural differences, as evidenced by the different peak positions in each pattern. There is insufficient information at this stage to designate either pattern as a polymorph of the material (e.g., they could be a solvate, hydrate, or a mixture of two or more polymorphs). However, it is clear that both materials are crystalline and structurally different. Further characterization using for example, thermal methods (TGA, DSC) would confirm these materials are not solvates or mixtures but actual polymorphs and aid in determining the thermodynamically stable polymorph. XRPD provides information about the structure of materials, not thermodynamics, although variable-temperature XRPD has been used to study changes in structure at different temperatures.

One can envision a large number of different crystallization experiments (using different solvents or conditions) performed on the API, some possibly in automated fashion, with the resulting material characterized initially by XRPD. This is in fact a common approach to polymorphism, salt, and cocrystal screening and perhaps the most common application of XRPD in the drug development process. The latter two screens are usually performed when the polymorph(s) of the drug candidate itself are not sufficiently bioavailable, in an effort to produce a formulation that addresses the bioavailability problem. An XRPD pattern is taken of the API and the guest material (e.g. acid) and the mixture of the two. If a salt or cocrystal was formed, the XRPD pattern of the mixture should be more than just a sum of the reference patterns of the API and the guest.

Therefore, the first application for XRPD during drug development is typically to identify the materials generated using different experimental methodologies, often in automated high throughput screening environments (25–27). To simplify this pattern recognition problem that often involves hundreds or thousands of experimental data sets per screen, people have developed various computational approaches to recognize, sort, and classify unknown XRPD patterns, either through comparison to a known database of materials (28) or simply within the experimental set of unknown patterns (18, 29, 30). The latter often uses an approach called hierarchical clustering (31, 32).

XRPD data are often catalogued in databases using the so called Hanawalt system (33–34). In this system, the data are stored as $d$ versus $I/I_{max}$ pairs. The use of $d$-space eliminates the need to specify the radiation source wavelength and allows comparison between laboratories using different instrumentation. A similar system is often used for intellectual property filings, as discussed later in this chapter.

However, as we shall see in following sections, there is considerable structural information available in a typical XRPD pattern that can be used to characterize the material. Making use of this information usually requires high quality laboratory data and the use of advanced computational methods.

***4.1.2 Structural Analysis of Crystalline Forms*** XRPD patterns of crystalline materials have sharp peaks due to constructive interference of diffracted radiation. As mentioned previously, XRPD patterns can be used to identify particular substances by comparison with reference patterns, analogous to the use of fingerprints to identify people. For example, the experimental XRPD pattern of Mannitol in Fig. 9 (bottom pattern) compares best with the simulated pattern of the beta form (second pattern) indicating that the sample is likely form beta, and not form alpha or delta. The positions of the reflections in XRPD patterns are functions of the size and shape of the crystallographic unit cell. Reflection intensities are functions of the atomic positions within the unit cell. A variety of instrumental and sample preparation artifacts influence the intensities, and to a lesser extent, the positions of the reflections. A first step in the further characterization of an XRPD pattern of a crystalline sample is to rationalize the reflection positions using a process called indexing.

*4.1.2.1 Indexing* Indexing is the process of determining the size and shape of the unit cell given the peak positions in a diffraction pattern. The term gets its name from the assignment of Miller index labels to individual peaks. For most applications, the index labels are less important than are the unit cell length and angle parameters that provide the link between crystal properties and the diffraction pattern.

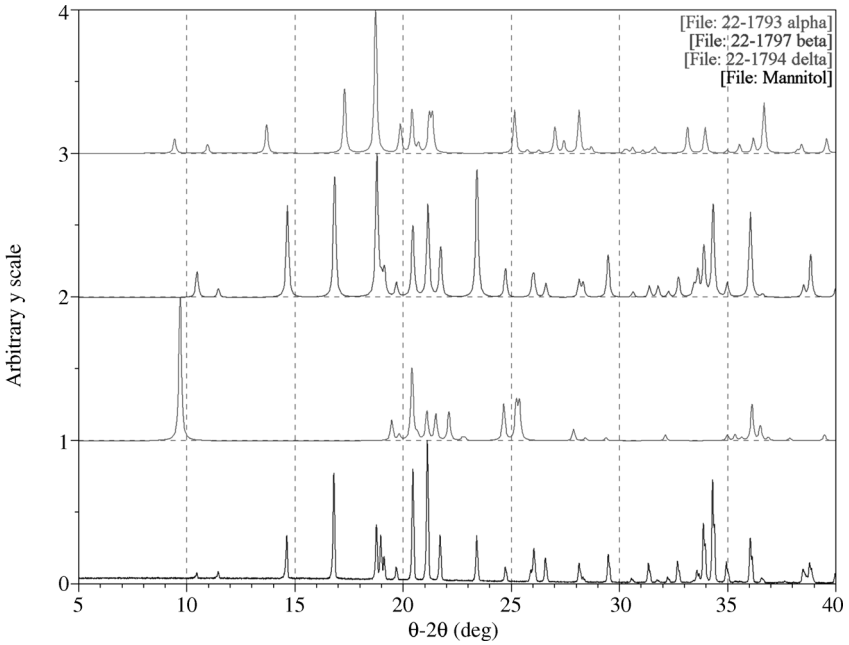


***FIGURE 9*** *Comparison of simulated (top three) and measured (bottom) XRPD patterns of Mannitol. The legend provides ICDD reference codes and form designations for the simulated patterns.*

Indexing plays a role in single crystal and powder diffraction pattern analysis, but there are qualitative differences between indexing single crystal and powder diffraction data. For single crystal diffraction, reflections are recorded as a function of the three-dimensional orientation of a crystal relative to the incident radiation. Each reflection in single crystal diffraction corresponds to a solution of the Laue equations, or equivalently the vector form of the Bragg equation. For powder diffraction, the (ideally) isotropic distribution of particle orientations leads to cylindrical symmetry in the diffracted radiation. Therefore, each reflection in powder diffraction corresponds to a solution of the scalar form of the Bragg equation (35). As a result of the loss of information regarding the relative orientations of diffracting crystallites, indexing of XRPD data is more challenging than single crystal diffraction data.

There are a variety of methods available for indexing XRPD patterns, some of which are in the public domain and others that are sold commercially. Examples include ITO (36), TREOR (37), N-TREOR (38), singular value decomposition (39), X-CELL (40), and DICVOL (41). The methods differ in their methods for identifying and refining potential solutions, ability to determine low symmetry solutions, efficiency, and sensitivity to extraneous and/or missing peaks. Development of novel methods and refinement of existing methods continues to be an area of active research. When applying these automated methods, it is important to be aware that multiple solutions can appear to adequately index a XRPD pattern. These degenerate solutions are called lattice metric singularities (42). In such cases, it is common practice to accept the higher symmetry solution unless there is other evidence supporting the lower symmetry solution.

Successful XRPD indexing serves several purposes. If all of the peaks in a pattern are indexed using a single unit cell, this is strong evidence that the sample contains a single crystalline phase. Given the indexing solution, the unit cell volume may be calculated directly. The difference in molar volume between an anhydrous crystal form and the molar volume of hydrates, solvates, and co-crystals can be useful for determining their stoichiometries. If the chemical composition of a crystal is known, then indexing provides the means of determining very accurate true densities. Indexing is also a robust description of a crystalline form. Common practice is to characterize forms using a list of XRPD peak positions for selected intense peaks, but the intensities of individual peaks are sensitive to changes in temperature, defect density, and preferred orientation effects as a result of sample preparation. Also, the positions of peaks are sensitive to temperature, strain, and changes in hydration state. Although unit cell parameters change as a result of these effects, the changes are less pronounced when comparing indexing solutions than when comparing peak lists based on particular selection criteria. Indexing can be used to show that a XRPD pattern is consistent with the structure of a crystal determined via single crystal diffraction. Also, indexing is a preliminary step for other analyses including structure determination from XRPD data and Rietveld refinement.

*Data* Not all XRPD patterns are amenable to indexing. XRPD patterns must have sufficient signal-to-noise and peak resolution to provide enough peak positions to uniquely define the indexing solution. Signal-to-noise can be improved by eliminating sources of diffuse scattering such as glass sample holders and by collecting the pattern

for a longer time. Peak width is a convolution of instrumental and sample contributions. Therefore, a diffractometer with adequate monochromatization is necessary, but not sufficient to ensure good peak resolution. Samples containing nanoscale crystalline regions or defective crystals display broadened peaks in their XRPD patterns. In the extreme, these broadened peaks can overlap and produce a pattern that appears to contain only amorphous halos as discussed elsewhere in this chapter.

Wilson showed that centrosymmetric crystal structures (those containing an inversion center) are more likely to have very weak and very strong reflections than are non-centrosymmetric crystal structures (lacking an inversion center) which tend to have a narrower distribution of peak intensities (43). Ordered crystals containing only one enantiomer (or diastereomer) of a chiral molecule cannot crystallize in space groups with mirror planes, glide planes, or inversion centers since those operations change the chirality of the molecule. Therefore, crystals containing non-chiral molecules or racemic mixtures of chiral molecules may adopt a centrosymmetric structure whose XRPD pattern has a large fraction of very weak reflections. Since these weak reflections are often important for successful indexing of the XRPD pattern, care should be taken to optimize peak sensitivity when collecting XRPD data for indexing, particularly for non-chiral molecules or racemic mixtures. This can be accomplished by reducing sources of background scattering and by increased collection times, for example.

The bottom XRPD pattern in Fig. 9 is an experimental XRPD pattern of Mannitol, form beta. Since Mannitol is chiral, it is expected to crystallize in a non-centrosymmetric structure and the corresponding XRPD pattern should have relatively few extremely weak or strong reflections. There is relatively little diffuse background scattering and the signal-to-noise ratio appears to be very good in the pattern. Also, the peaks are quite sharp leading to good peak resolution. All of these observations suggest that the XRPD pattern is amenable to indexing under the assumption that the pattern represents a single crystalline phase.

*Results* A successful indexing solution for the Mannitol XRPD pattern from Fig. 9 is illustrated in Fig. 10. The bars in Fig. 10 correspond to positions of allowed reflections based on Bragg's Law. All of the observed peaks in the XRPD pattern are indexed by one or more of the allowed reflections. Also, there are relatively few allowed reflections that are not observed. In some cases, the reflections are too weak to be evident at the scale shown in Fig. 10, but are evident using an expanded scale. If the allowed reflections did not account for all of the observed reflections, then the solution

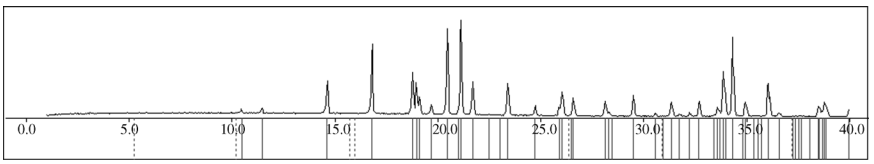


***FIGURE 10*** *Indexed XRPD pattern of Mannitol form beta. Green bars indicate allowed reflections based on the unit cell dimensions and assigned space group $P2_12_12_1$ (no. 19). Dashed bars indicate extinct reflections.*

would have been rejected as unable to fully describe the observed pattern. If there were a large number of allowed but unobserved peaks, then the solution would be rejected in favour of a smaller volume and/or higher symmetry solution. Since the illustrated solution falls into neither category, it is accepted as a good description of the observed pattern. This confirms that the XRPD pattern is of a single crystalline phase.

There are some allowed reflections, marked with dashed bars in Fig. 10, that are not observed. These systematic extinctions are the result of space group symmetry elements of the contents of the unit cell. Systematic extinctions are the result of centered unit cells, screw axes, and/or glide planes. Rules indicating which reflections are extinct based on their Miller indices are tabulated (35). Given the observed systematic extinctions in a pattern, an extinction symbol may be assigned (44). In the illustrated case, the extinction symbol ($P2_12_12_1$) corresponds to only one space group, $P2_12_12_1$ (no. 19). In many other cases, there are two or more space groups corresponding to the observed extinction symbol and additional analysis is needed to fully determine the space group.

The solution illustrated in Fig. 10 is orthorhombic with length parameters given in Table 2. Given the length parameters, the unit cell volume is readily calculated. The number of asymmetric units in the unit cell for a given space group is tabulated in Reference (44). For space group $P2_12_12_1$ there are four asymmetric units in the unit cell ($Z = 4$). Assuming that there is one Mannitol molecule per asymmetric unit ($Z' = 1$), then the molar volume and density are readily calculated. The density is a bit higher than typical organic molecules, but is as expected for sugars such as Mannitol. A reasonable density confirms that the correct number of molecules per asymmetric unit was assumed.

Having successfully indexed the XRPD pattern, determined the extinction symbol, and calculated the density, the analysis of peak positions in the Mannitol XRPD pattern is complete. Additional information regarding the Mannitol sample can be derived from the peak shapes and intensities using the Rietveld method.

**TABLE 2 Indexing Results and Derived Quantities**

| Substance and Form | Mannitol Form Beta |
|---|---|
| Composition | $C_6H_{14}O_6$ |
| MW (amu/molecule) | 182.172 |
| Family and space group | Orthorhombic $P2_12_12_1$ (no. 19) |
| $a$ (Å) | 8.6790 |
| $b$ (Å) | 16.8980 |
| $c$ (Å) | 5.5502 |
| $\alpha$ (deg) | 90 |
| $\beta$ (deg) | 90 |
| $\gamma$ (deg) | 90 |
| $V$ (Å$^3$) | 813.65 |
| $Z'$ (molecules/unit) | 1 |
| $Z$ (units/cell) | 4 |
| $V/Z$ (Å$^3$/unit) | 203.41 |
| $\rho$ (g/cm$^3$) | 1.487 |

*4.1.2.2 Rietveld Analysis* The Rietveld method is a computational tool for extracting structural and microstructural information about a crystalline solid from its powder-diffraction pattern. This computational method was pioneered by Hugo Rietveld in the late 1960s (45, 46). The ability to obtain structural information from polycrystalline materials from which a crystal suitable for single-crystal analysis was either unavailable or practically impossible to obtain was a significant achievement. Initially the method was mostly used to refine the atomic positions and lattice parameters of crystal structures. Use of the Rietveld method proliferated with the increased accessibility of digital data in the 1980s and is now routinely used for quantitative analysis of phase mixtures for a wide range of materials and applications from cement and minerals to pharmaceuticals. This method is also used to characterize crystal defects and the microstructural properties of polycrystalline solids, such as preferred orientation, microstrain, and crystal size.

The Rietveld method computes a powder pattern using information describing the crystal structure and the instrumental technique used to collect the diffraction pattern. An iterative algorithm refines the structural and instrumental parameters used in the computation to minimize the difference (residual) between the observed and calculated diffraction intensity at each scattering angle ($2\theta$). The success of a Rietveld analysis depends on building an appropriate model for the algorithm so the best fit (global minimum) to the experimental pattern can be efficiently discovered while avoiding false minima. Many freeware computer programs available for Rietveld analysis are listed and described on the website (47); many programs are also commercially available. To set up a model for a system of interest and obtain reliable information, each program requires a skilled user with understanding of diffraction theory and experiment. A recommended way to learn and assess the reliability of a Rietveld program is by refining patterns of well-characterized powders. $\alpha$ $Al_2O_3$ NIST SRM 676 is a very good standard because it has no preferred orientation, two refinable atomic positions determined by single-crystal analysis (48), and two lattice parameters. The accuracy of the measured intensity and $2\theta$ combined with the skill of the Rietveld analyst and the accuracy of the Rietveld refinement algorithm determines the accuracy of the refined parameters. The applicable theoretical corrections used to compute the diffraction pattern for a particular data-collection technique require the user to judiciously construct the appropriate models (49, 50).

The minimum information needed to calculate a diffraction pattern is the atomic positions, space group, and lattice parameters of the crystal structure. The Rietveld program calculates the position and intensity of the diffraction peaks and assigns them a $2\theta$-dependent profile determined by the instrumental model. The parameters describing the profile are obtained from the results of a Rietveld analysis of a pattern of NIST SRM 660a LaB6 profile standard. This standard can also be used to check the accuracy of the lattice-parameter refinement and to refine the wavelength intensity ratio if more than one X-ray wavelength was used for data collection. For example, the ratio of Cu $K\alpha 2$ to $K\alpha 1$ is $\sim 0.5$ and can vary several hundredths depending on the type and alignment of the monochromator of the instrument used to collect the powder pattern. Standardizing the profile function allows microstrain and size (51) to be more

accurately refined as well as improving the fit of the calculated pattern to the experimental pattern.

Several quantitative and qualitative ways of assessing the "goodness of fit" between the experimental and calculated patterns are available. The qualitative assessment is a careful visual examination of the fit and the residual intensity. Visual assessment is essential and compliments the quantitative measures of the fit. The quantitative measures of the fit include the weighted pattern residual ($R_w$) and the ratio of $R_w$ to the expected residual ($R_w/R_{exp}$). This ratio indicates how close the fit is to the theoretically best residual calculated using the number of data points in the observed pattern, the number of refined variables, and the uncertainty of the observed intensity intrinsic to counting photons.

The refinement is usually done in several stages (52) with only some of the refinable parameters being refined in each stage. The scale factor for the incident-beam intensity is always refined, and the background is usually refined in the first stage. The background is best minimized in the experimental pattern by use of antiscatter slits and helium; reducing background also improves the detection limit. Background is usually modeled with a polynomial and, if the background cannot be adequately fit, it can be artificially minimized by subtracting an appropriate blank pattern collected with identical collection parameters as the specimen of interest. Subsequent stages depend on the goal of interest. None of the refinable parameters will usually be refined well the first time; therefore, several iterations are used to find the global minimum.

A successful Rietveld refinement requires the user to know if the pattern of interest displays artifacts from inadequate orientation statistics (OS) and preferred orientation. These artifacts affect relative peak intensities and can generate significantly inaccurate structure refinements. If preferred orientation is well understood, it can be corrected by a variety of models available in many Rietveld programs. OS artifacts are manifest in random fluctuations of relative intensity for each preparation of the sample and cannot be corrected. The spiky peaks generated by OS artifacts can be difficult to detect by visual examination of the one-dimensional XRPD pattern. Fortunately, they can be readily detected by collecting a two-dimensional pattern with an area detector. If the Debye rings in the two-dimensional pattern are spotty, then OS is a potential problem depending on how the pattern of interest is collected. Both PO and OS artifacts can be practically eliminated by collecting the pattern in transmission geometry using a Gandolfi spinner and by comminution of the sample. Unfortunately comminution is often not a viable option because many samples, especially organics, are susceptible to stress-induced phase transformation.

We provide an example of the limitations of refining a pattern displaying a slight amount of PO. Rietveld refinement using Maud (version 1.999) was applied to a pattern of β mannitol collected with a laboratory diffractometer in transmission geometry. To correct PO we used a March-Dollase model on the 130 hkl (Fig. 11). The atomic positions were compared to those from a single-crystal structure determination by Kaminsky et al. (53). The thermal parameters were bound to be the same for all atoms and set to be isotropic; hydrogen was not included. Without using PO corrections, the fit to the measured pattern was poor (Fig. 12).

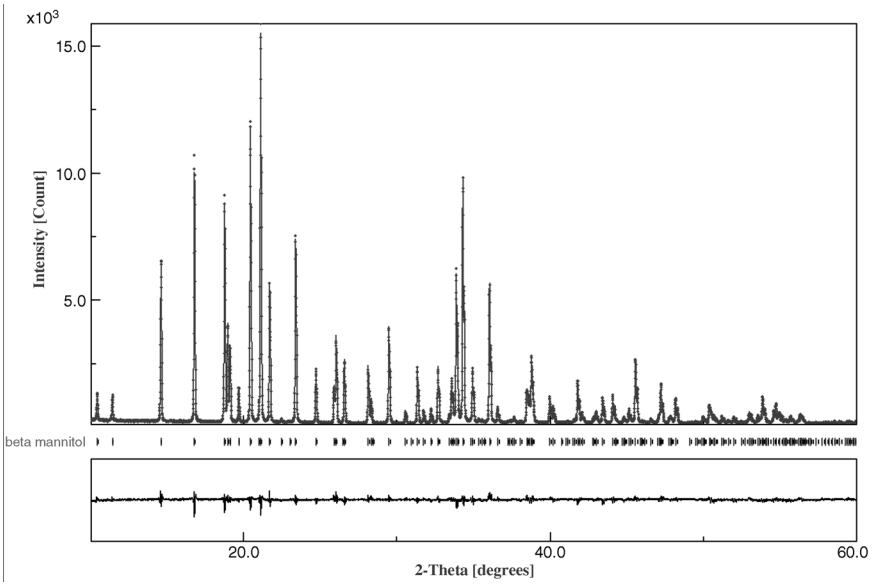


***FIGURE 11*** *Overlay of a Rietveld refinement of $\beta$ mannitol with $R_w = 7.19\%$* and *$R_w/R_e = 1.6$* used to investigate the use of spherical harmonics to facilitate the refinement of the lattice parameters and thermal factor.

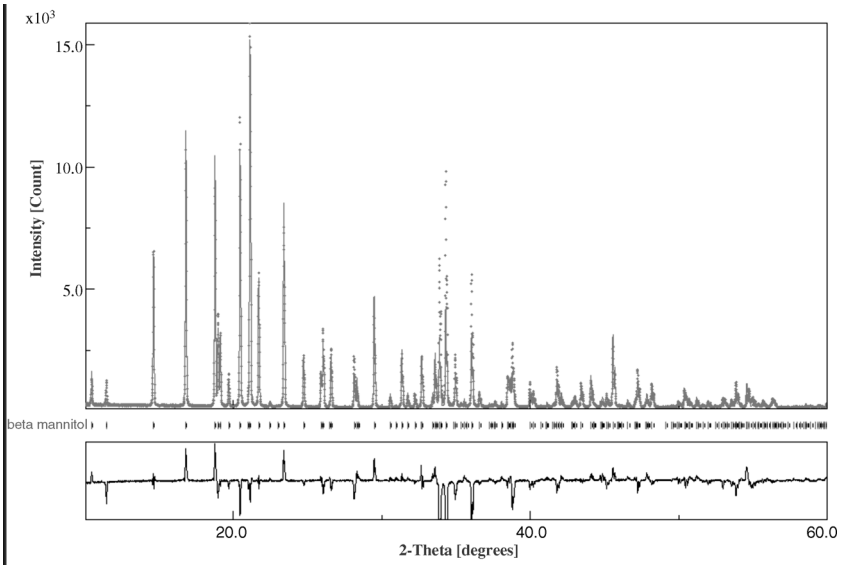


***FIGURE 12*** *Rietveld refinement of $\beta$ mannitol without spherical harmonics, $R_w = 26.3\%$* and *$R_w/R_e = 5.9$.*

**TABLE 3 Bond Distances for β-Mannitol from a Rietveld Refinement of a Pattern with Orientation-Statistics Artifacts Compared to Those from a Single-Crystal Analysis by Kaminsky et al. (53) and a Rietveld Refinement by Botez et al. (54)**

| Bond Distance | Kaminsky et al. 1997 (Å) | Botez et al. 2003 (Å) | Difference (Å) | This Study (Å) | Difference (Å) |
|---|---|---|---|---|---|
| C1-C2 | 1.517 (2) | 1.545 (17) | 0.028 | 1.585 | 0.068 |
| C2-C3 | 1.539(2) | 1.546 (15) | 0.007 | 1.543 | 0.004 |
| C3-C4 | 1.523 (2) | 1.515 (16) | −0.008 | 1.555 | 0.032 |
| C4-C5 | 1.539 (2) | 1.592 (14) | 0.053* | 1.618 | 0.079 |
| C5-C6 | 1.521 (2) | 1.552 (16) | 0.031 | 1.569 | 0.048 |
| C1-O1 | 1.425(2) | 1.416 (17) | −0.016 | 1.485 | 0.053 |
| C2-O2 | 1.440 (2) | 1.443 (16) | 0.016 | 1.444 | 0.017 |
| C3-O3 | 1.433(2) | 1.390 (17) | −0.042 | 1.431 | −0.001 |
| C4-O4 | 1.448(2) | 1.434 (15) | −0.004 | 1.356 | −0.082* |
| C5-O5 | 1.445(2) | 1.439 (16) | 0.011 | 1.461 | 0.033 |
| C6-O6 | 1.436(2) | 1.431 (18) | 0.004 | 1.460 | 0.033 |
| Standard deviation | | | 0.026 | n/a | 0.040 |
| Sum | | | 0.080 | n/a | 0.284 |

*Maximum deviation.

The bond distances, Table 3, from our refinement had a 0.04 Å standard deviation relative to the single-crystal values and a maximum absolute error as high as 0.08 Å. This uncertainty is ten times larger than the uncertainty of the single-crystal bond distances and slightly higher than those refined by Botez et al. (54) using a pattern collected with a synchrotron. Our bond distances also have a positive bias. These relatively high uncertainties reflect the intensity inaccuracies from not completely correcting the PO. The inaccuracies of atomic positions will increase as the magnitude of PO and OS artifacts increase. The calculated bond angles also will be more significantly more uncertain than the reliable single-crystal values.

The magnitude of thermal displacements above those commonly measured by single-crystal analysis indicates the amount of static-displacement defects in the crystal. The refined thermal parameter (B) was 2.4 $Å^2$, and the single-crystal values for carbon and oxygen ranged from 1.64 to 2.81 $Å^2$. This result indicates the $2\theta$-dependent intensity corrections provided by the Debye-Scherrer geometry and the curved position-sensitive detector models are reasonably accurate. The intensity fluctuations from slight PO are not expected to significantly affect the thermal parameter if enough of the pattern was refined, but, depending on the magnitude of the PO artifacts, a reasonably accurate B value might not always be obtained unless an accurate PO correction is applied.

The orthorhombic lattice parameters were refined to $a = 8.6790$ Å, $b = 16.8980$ Å, and $c = 5.5502$ Å, and the zero offset was 0.0004°. These lattice parameters are within 0.01% of those refined by Botez et al. (54). The negligible offset indicates the diffractometer was very well aligned.

The beginning Rietveld analyst must have a good understanding of experimental and theoretical diffraction physics and the Rietveld model to assess the goodness of fit. The β mannitol example described here demonstrates a fairly good fit to the observed

pattern using PO corrections does not necessarily mean the refined parameters are reliable and accurate; the bond distance and angles must be compared to known ranges from reliable single-crystal determinations to see if the results are reasonable. Values of bond distance and angle for moieties on organic molecules can be found in the Cambridge Crystal Structure database. Although patterns containing intensity artifacts can be analyzed using Rietveld analysis to provide useful structural information, the use of appropriate data-collection techniques is always the best way to obtain the most reliable results.

***4.1.3 Structural Analysis of Disordered Crystalline Forms*** Highly crystalline forms of APIs are preferred in the drug development process because of their high level of purity and resistance to physical and chemical instabilities under ambient conditions (34). Unfortunately, most organic crystals are generally imperfect, containing different types of defects in the structure of the crystal lattice (4). The defective regions in a crystal introduce disorder into what is otherwise an ordered system, and correspond to sites with higher levels of energy in the solid (55). In many cases, defects can eventually lead to the formation of glass or amorphous (supercooled liquid) materials. Defects are typically formed through kinetic processes, for example, milling or dehydration, as opposed to thermodynamic processes. Whereas defects can offten be reversed, the formation of amorphous material creates a different solid state that can agglomerate or recrystallize into a different crystalline form. Pharmaceutical interest in the study of defects arises from the demonstrated role that these high energy sites can play in affecting a number of important physical and chemical phenomena. The presence of defects can, for example, produce higher dissolution rates (56, 57), greater chemical instability (58, 59), altered mechanical properties (60–62), and enhanced hygroscopicity (63, 64).

XRPD combined with computational methods can be used to study crystalline defects (4, 15), even quantitate kinetic disorder parameters (e.g., crystal strain). Random defects within a crystal structure will generate diffuse X-ray scattering similar to X-ray amorphous scattering, though with halo positions and widths that are typically different from those measured from the amorphous (supercooled liquid) phase (15). This rule can be broken when defects in the material form a kinetic glass but it is possible to monitor the agglomeration of defects starting with perfect crystals all the way through glass, using XRPD. Such a study was reported in Reference (15) and is summarized below as an example application of XRPD to the study of crystalline defects.

In this study, defects were introduced into Raffinose pentahydrate through dehydration over time. XRPD measurements were taken at regular time intervals, Fig. 13, to monitor per cent crystallinity (65) and total diffraction methodology was used to model both the coherent long range crystalline contributions and incoherent short range disorder components together as a single system (66, 67). Rietveld methods were used for structure modeling (68, 69) and to track the structural phase changes in crystalline material. In addition, methods built around the use of the Pair Distribution Function (PDF) (8, 11, 70–75), $G(r)$ in Eq. 3, were employed to confirm the Rietveld results. Finally, it was shown that the introduction of defects through

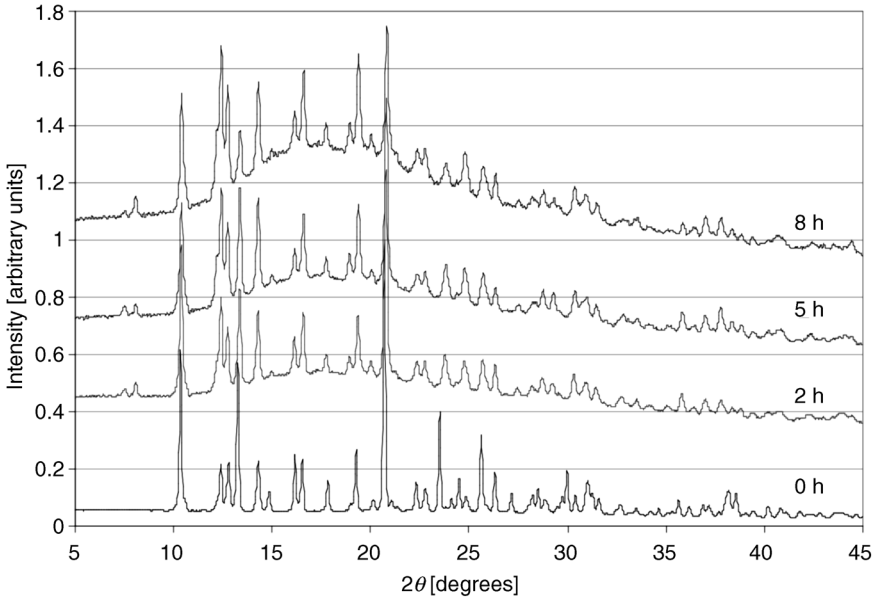


***FIGURE 13*** *XRPD patterns collected in transmission mode for raffinose pentahydrate dried at 60° C under vacuum for (from bottom to top): 0, 2, 5, and 8 h, respectively.*

dehydration of an organic crystal hydrate eventually leads to the formation of amorphous material.

$$G(r) = 4\pi r[\rho(r) - \rho_0] \tag{3}$$

Where $\rho_0$ is the average number density of the structure and $\rho(r)$ is the atom pair density for X-ray scattering (76) (Eq. 4).

$$\rho(r) = \left(\frac{1}{4}\pi r^2\right) \sum_{p,q} \left(\frac{f_p f_q}{\langle f \rangle^2}\right) \delta(r - r_{pq}) \tag{4}$$

The terms in Eq. 4 include $f_{p,q}$, the individual atomic form factors; $\langle f \rangle$, the mean atomic form factor of the structure; and the distance $r_{pq}$, corresponding to the atom pair separation. $\rho(r)$ represents the probability of finding an atom pair separated by the distance $r$, weighted by the atomic form factors and averaged over all atom pairs in the structure. The calculation needs to be evaluated over a crystal structure large enough to give the atom pair relationships for a distance of interest.

The PDF is a generally useful computational tool when analyzing X-ray amorphous or disordered systems. PDF analysis unlocks information present (but not obvious) in XRPD patterns, providing a measure of inter-atomic distances that define a solid form (70). From these distances, one can draw conclusions regarding the relationship between ordered and disordered forms of a material. By convention (71),

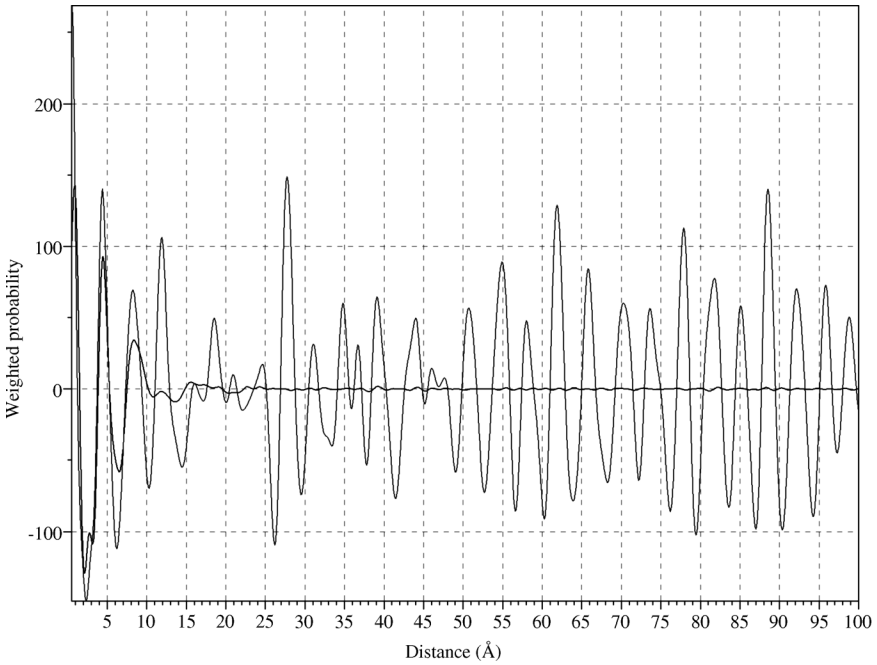


***FIGURE 14*** *PDF traces of crystalline (lighter, with many peaks) and amorphous (darker, with few initial peaks) felodipine. The crystalline sample shows long-range order as evidenced by peaks in the PDF up to and exceeding 100 Å. The amorphous sample shows only short-range order with PDF peaks disappearing after approximately 20 Å. Note that both samples have similar short-range order with peaks in the PDF overlapping below 10 Å.*

the PDF trace is typically plotted as distance (in Å) vs. (weighted) probability of finding two atoms with that separation. Peaks in the PDF correspond to inter-atomic distances that are common to the material in question, with the product of the peak area and distance giving the number of atom units with that specific separation. Figure 14 illustrates a PDF trace of amorphous felodipine superimposed against a PDF trace of crystalline felodipine. Both traces show similar short-range order (below 10 Å) but whereas the amorphous trace has no peaks in the PDF trace above approximately 20 Å (and therefore no long-range order), the crystalline trace contains peaks up to and exceeding 100 Å (indicating long-range order). Differences between defective crystalline (X-ray amorphous) and amorphous (supercooled liquid) materials can be seen in the PDFs calculated from XRPD data (15).

***4.1.4 Structural Analysis of Organic Amorphous Materials*** Amorphous organic solids, also referred to as disordered systems (77, 78), are characterized by a lack of long-range order that is observed in crystalline structures (2). They are often described as supercooled liquids, turned solid by the removal of thermal energy or a solvent without inducing crystallization (65, 78). Many of the early stage drug candidates currently under development exhibit very poor aqueous solubility (2, 79).

Higher molecular weights than were common in the past and greater hydrophobicity are thought to contribute to the decrease in aqueous solubility of organic compounds. With data to suggest that the dissolution rate is the rate-limiting step in oral absorption, it is expected that reduced solubility will lead to reduced oral bioavailability (80).

Amorphous organic compounds tend to have significantly higher initial aqueous solubilities (non-equilibrium) and dissolution rates as well as greater compressibility and different hygroscopicity, as compared to crystalline forms of the same compound (65, 81). Therefore, in compounds where the crystalline forms are poorly soluble, the amorphous form often presents an attractive formulation option (82). Dozens of APIs and excipients currently on the market are listed as being amorphous (65). Because crystalline forms are more thermodynamically stable than amorphous forms, there is a driving force for crystallizing the amorphous state. Therefore, amorphous materials are inherently unstable (both physically and chemically) and, once crystallized, the material loses the solubility advantage provided by the amorphous form. The inherent instability coupled with a lack of developed characterization tools has limited the commercial potential of working with amorphous materials at present (65, 83).

A common method used by formulation scientists to overcome the instabilities associated with amorphous materials is to prepare a composition of the amorphous API with pharmaceutically acceptable excipients that provide a barrier to crystallization of an amorphous drug product upon storage. Such compositions are often referred to as "dispersions" or "solid dispersions." Stabilizers are often selected from a variety of polymers, a commonly used one being polyvinylpyrrollidone (PVP). Intimate mixing (mixing at the molecular level) between such excipients and the amorphous API is an important factor in compositions resistant to crystallization (84). To provide a barrier to structural changes, the excipient needs to be intimately mixed (e.g., in a solution) with the API, which is often difficult to accomplish or even identify.

As noted earlier, XRPD is a useful technique in the analysis of amorphous materials. In its simplest application, XRPD is used to identify X-ray amorphous materials and classify them separately from crystalline materials (18). It can also be used to further differentiate between (supercooled liquid) amorphous material and kinetic glassy material (15, 85) or to monitor the conversion of crystalline material into amorphous material upon processing (e.g., grinding) (85) (Fig. 15). In the latter application, note the steady peak broadening and loss of intensity at higher angles in the XRPD patterns, Fig. 15, as the material becomes progressively disordered. However, some materials may immediately collapse into a binary amorphous-crystalline mixture and eventually pure amorphous material, without any broadening of the crystalline peaks that would be indicative of disorder in the crystalline phase (85).

Finally, in the analysis of multi-component amorphous mixtures, XRPD along with computational methods (16) have emerged as techniques that can be used to identify the so-called "miscible" dispersions, as compared to the phase separated physical mixtures where intimate mixing was not achieved. XRPD is often accom-

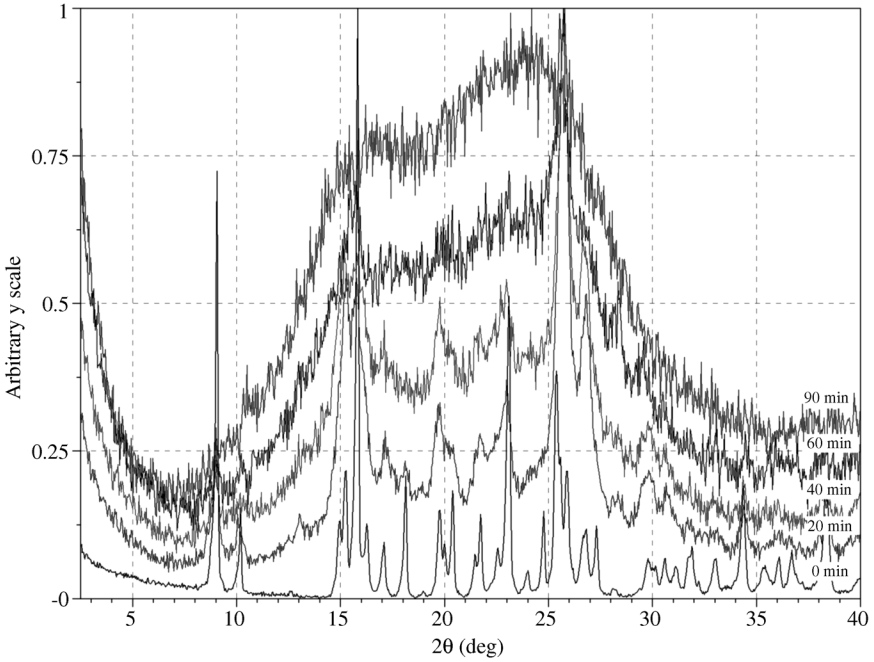


***FIGURE 15*** *Monitoring of cryo-grinding experiments on piroxicam form II using XRPD. Bottom-to-top, starting pattern, after 20, 40, 60, 90 min of grinding, respectively.*

panied by the use of complementary experimental techniques in this type of amorphous mixture analysis. Traditionally, thermal methods such as DSC have been used to measure glass transition temperatures when evaluating miscibility (86) and, while those methods have been shown to have limitations (16), their use is still recommended when analyzing amorphous mixtures (in addition to XRPD).

An example of this type of XRPD and computational analysis was reported in Reference (16). Solid dispersions of three different systems were analyzed initially using thermal measurements. Glass transition temperatures were identified from thermal data. For the system where two glass transition temperatures were identified, phase separation was indicated. The other two systems had a single glass transition temperature, suggesting miscibility and therefore stability to crystallization. XRPD data were collected on all three systems and the patterns were then analyzed using computational methods. Linear combinations of measured XRPD patterns and calculated PDFs were used to model the systems to determine the level of interaction between the two components in each system. The conclusions derived from thermal data were confirmed by XRPD for two of the three systems. However, the third system was identified as a so-called nano-suspension with domain sizes smaller than the resolution of thermal methods. Therefore, a prediction was made that this system is in fact phase separated despite the single glass transition temperature, and would re-crystallize. This prediction was confirmed elsewhere in the literature (87).

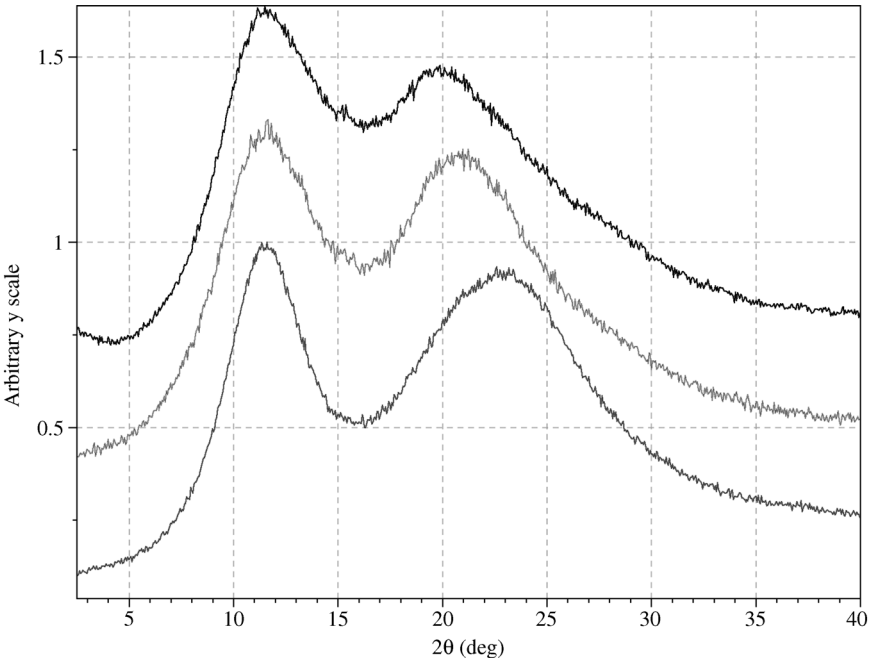


***FIGURE 16*** *XRPD patterns of, top-to-bottom, amorphous PVP, a solid dispersion of PVP and felodipine (70% PVP, 30% felodipine by weight), and amorphous felodipine.*

Figure 16 shows example XRPD patterns of a pharmaceutically acceptable excipient (PVP, top pattern), amorphous drug product (felodipine, bottom pattern), and a solid dispersion of the two (70% PVP and 30% felodipine by weight, middle pattern). Despite the lack of long-range order (and therefore crystalline peaks), one can observe clear differences in the structure of these amorphous materials from these XRPD patterns, using the position and width of the two broad halos in each pattern. For example, the second halo in the dispersion is located in between the second halos for the excipient and the drug product, suggesting it is a mixture of the two materials. Additional computational analysis could be applied (16) to determine the level of miscibility of the two components in the dispersion.

***4.1.5 Characterization of Multi-Component Mixtures*** A common application for X-ray powder diffraction is in the characterization of multi-component mixtures. The technique is highly effective when used for this application provided the materials are structurally different (88) and fall within the limit of detection of XRPD. With good laboratory instrumentation and sample preparation methods and typical counting times, the limit of detection for XRPD can be on the order of 2–3% by weight, or even lower when detecting the presence of crystalline materials in an amorphous-crystalline mixture. In multi-component crystalline mixtures the limit of detection can become significantly worse, especially if the component being detected has no unique peaks as compared to

the other components in the mixture. The detection of low levels of amorphous material mixed with a predominately crystalline sample can be difficult and is a subject of ongoing research (65). XRPD can also be used to quantify the amount of (crystalline or amorphous) component in the mixture as the diffracted intensity in the pattern is related to the weight percentage of the component in the mixture.

*4.1.5.1 Characterization of Crystalline Mixtures (Quantitative Analysis)* Quantitative analysis of crystalline mixtures typically refers to the determination of the relative amounts of different phases in a sample containing multiple phases (i.e., in a mixture), using experimental XRPD data. In broad terms, quantitative methods can be separated into those requiring the use of internal or external standards and ones based around the use of the full diffraction pattern (5,11, 14, 89, 90). The application of standards to this problem is not new, dating back to the 1930s (91). When using standards, it is essential to work with a randomly oriented fine powder specimen to avoid problems with preferred orientation or poor particle statistics and to be confident that the sample is representative of the material being analyzed (5, 14). In addition, there are a number of structure, instrument, and measurement-sensitive factors (14) that affect quantitative analysis.

The internal standard method (92, 93) (also called the RIR method, for reference intensity ratio) is the most general and the most commonly used of any XRPD quantitative phase analysis methods in the past. It requires that a known weight percentage of a standard material be homogeneously mixed with the sample whose phase composition one is trying to determine. The method is based on dividing the intensity of a line (peak) of the phase whose ratio in the composition is being determined by the intensity of the line for the internal standard used. As such, it is important to choose an internal standard that has peaks that do not overlap with any found in the mixture itself. Examples of standards include corundum, silicon, and SRM 676 ($Al_2O_3$) (5). The latter is particularly good at minimizing preferred orientation problems. For a detailed discussion on the RIR method see for example, (5, 14), in recent years it has been increasingly replaced by full pattern methods because of difficulties in ensuring random orientation of all components in the mixture.

The full pattern methods can be traced to quantitative work performed on cements in the 1960s (94–98). The same principles apply to pharmaceuticals with some modifications, the most important of which is the need to normalize the total diffracted intensity across different patterns before using the full pattern method. The simplest of such full pattern methods rely on fitting entire observed or calculated patterns of reference materials to an observed pattern of the mixture (5, 99). If not previously available, XRPD patterns of the reference (standard) materials must be first collected, meaning one must know the composition of the mixture beforehand. The references must be analyzed under the same conditions and instrumental settings as the mixture. When possible, the instrumental background and noise should be removed by data processing (100). Provided the same sample preparation techniques were used to

make the reference materials and the mixture(s), sample artifacts should not present as big a problem as for internal standard methods as the full pattern methods do not rely on single peaks. The weighted sum of the selected reference patterns is fit using least-squares minimization (e.g., (101)) to the unknown mixture. In this procedure, each point in the pattern is treated as an independent observation and a single weight is applied to each pattern, giving its ratio in the mixture. The best calculated fit is reported along with the weights used to obtain this pattern (5, 98, 99).

A second commonly used full pattern quantitative method is based on the use of the Rietveld method (5, 14, 102). In this method, the refinement is done by minimizing the sum of weighted, squared differences between measured and calculated intensities at every $2\theta$ position in the pattern. The Rietveld method does require the knowledge of the crystal structure for every component (of interest) in the mixture. This method is applicable even in the presence of overlapping peak lines and is less sensitive to preferred orientation problems. However, the requirement to have at least approximate crystal structures of every component can be problematic for APIs under development. Some examples of Rietveld applications to quantitative analysis include (103–105). Freely available software, for example, TOPAS (106) or GSAS (107), can be used for Rietveld quantitative analysis.

Finally, novel quantitative approaches using full pattern fitting are still an active area of research. Recent example include References (89, 108).

*4.1.5.2 Characterization of Crystalline-Amorphous Mixtures (Per cent Crystallinity)* Degree (or per cent) of crystallinity is perhaps the most common quantitative application of XRPD in the drug development process. When the amorphous phase is present or suspected, XRPD can be used to characterize the material and determine the ratio of crystalline to amorphous material in the sample. Most X-ray methods used for this purpose are based on total sample scattering (i.e., scattering from both the amorphous and crystalline phases). In addition, the total diffracted area in each measurement needs to be normalized between samples to account for example, weight or absorption differences across samples. Once the patterns are normalized, per cent crystallinity can be easily calculated as a fraction of crystalline content in the total diffracted intensity measured from a sample (109). Several methods exist to estimate the crystalline content in an XRPD pattern (65, 110):

i. Measuring the integrated peak area under the crystalline peaks and dividing by the X-ray intensity scattered by the entire sample (65). This is a common approach that works well except when many overlapping crystalline peaks are present. Examples of this type of analysis can be found in, References (110–122).
ii. Measuring the maximum intensity of a subset of characteristic peaks (109). This method can only be applied if all the samples are analyzed under the same conditions, there is no peak broadening due to disorder or particle size issues, and a reproducible ratio of intensity can be demonstrated for the chosen peaks across multiple samples. Typically, the use of internal standards is recommended for this method (65,109) to ensure the above conditions are met. Such

internal standards can include lithium fluoride, silicon metal powder, or zinc oxide. A variation on this method uses peak area rather than height (123). Examples of these types of studies include References (123–129).

iii. Quantification of the amorphous content using a region where no crystalline peaks appear in the mixture pattern (130). Since the diffraction in such regions is caused by short-range ordering of amorphous materials, it can be used to estimate the amorphous content present in the sample. The method may be more robust and sensitive than peak intensity measurements (110) though it is limited to mixtures where crystalline peaks are not present in the entire range of amorphous diffraction.

Other methods for the determination of per cent crystallinity have been reported in literature, for example, using specialized instrument optics (131). If both amorphous and crystalline reference patterns are available, full pattern quantitative or semi-quantitative methods used in crystalline mixture analysis can be applied here as well. Such an example is shown in Fig. 17. Physical mixtures of amorphous and crystalline sucrose were prepared with known weight ratios (0%, 20%, 40%, 60%, 80%, and 100% crystalline). The pure amorphous and pure crystalline patterns were then linearly fitted using Brent minimization (101) to each mixture. A single weight was used per reference pattern (applied to all intensity points) and the sum of squared differences at each intensity point was used as the error metric. All patterns were first

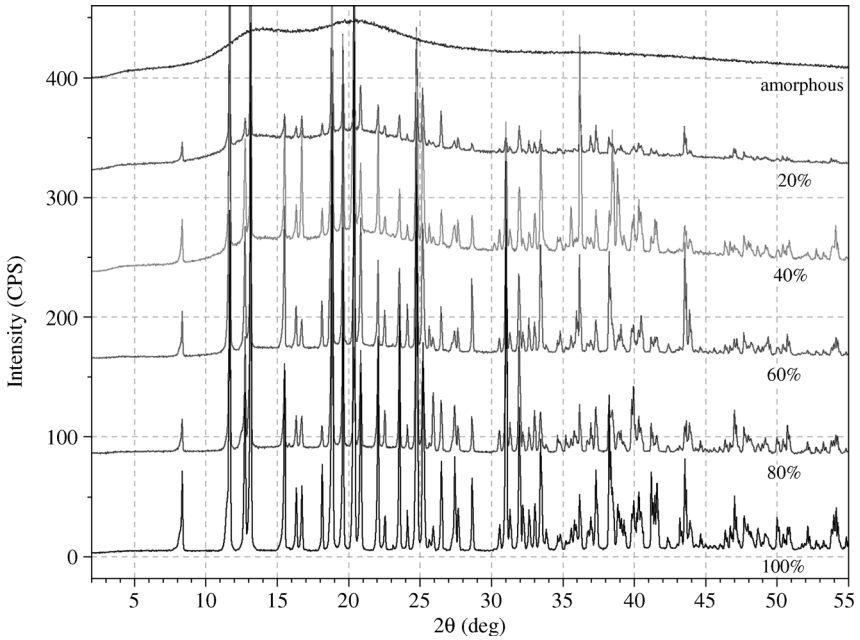


***FIGURE 17*** *XRPD quantitative analysis of physical mixtures of amorphous and crystalline sucrose. Top-to-bottom, amorphous sucrose, 20, 40, 60, 80 and 100% crystalline sucrose.*

normalized to same total diffracted intensity. The calculated per cent crystallinity values for the 20/40/60/80 mixtures were 17.2%, 40.7%, 64.4%, 84.3%, respectively. In practice, this approach coupled with good XRPD data consistently yields results within 5% of actual values for mixtures ranging from 10 to 90% crystalline.

*4.1.5.3 Re-Crystallization of Amorphous Materials* XRPD can also be used to monitor re-crystallization of amorphous materials. A typical application of this kind would be to characterize the extent of crystallinity during processing steps (e.g., scale-up of bulk material, formulation, manufacturing) or simply upon storage over the intended shelf life of the drug product, to ensure safety and efficacy (132). As noted earlier, amorphous materials will have different performance than crystalline counterparts (65,133), including for example sensitivity to moisture (134–136), reduced chemical stability (113,126,137), postcompression hardness (112), and enhanced dissolution rate (81,138).

Figure 18 shows the monitoring of recrystallization of amorphous felodipine by XRPD over a 7 day period. The top pattern is pure amorphous felodipine, the remaining patterns were collected on the same sample stored under ambient conditions (25°C and 50% RH) after a period of 2, 3, 5, and 7 days, respectively. The aforementioned computational methods were used to determine per cent crystallinity in each pattern, with the first and last patterns used as reference amorphous and crystalline material, respectively. Given those assumptions, the pattern collected after

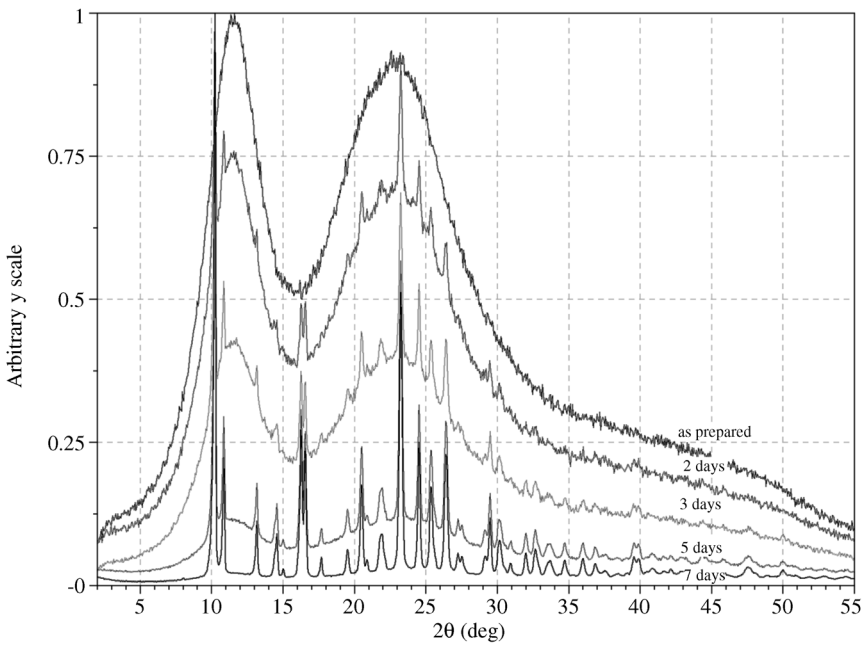


***FIGURE 18*** *Recrystallization of amorphous felodipine monitored by XRPD. Top-to-bottom, amorphous felodipine, same sample after 2, 3, 5, and 7 days in ambient conditions, respectivelly.*

2 days is approximately 5% crystalline, followed by 11% after 3 days and 45% after 5 days. After 7 days the sample is almost completely crystalline.

***4.1.6 Regulatory Considerations*** The Food and Drug Administration's (FDA) New Drug Application (NDA) guideline (139) states that "appropriate" analytical procedures should be used to detect polymorphic, solvated (including hydrated), or amorphous forms of a drug substance. Likewise, the guideline states that it is the applicant's responsibility to control the crystal form of the drug substance. If bioavailabilty is affected by the change in crystal form, the applicant must demonstrate the suitability of the control methods. This highlights the importance of controlling the crystal form of the drug substance and, as shown in previous sections, the use of XRPD can provide essential structural information on crystalline and amorphous forms of a drug substance.

The ICH Q6A document (139) provides clear guidelines on how X-ray information should be used in the NDA. As noted in the section on API characterization, XRPD is one of two broadly accepted methods of providing unequivocal proof of polymorphism by the regulatory authorities (20), the other being single crystal X-ray diffraction. Characterization of the API by X-ray and/or some other method is required by the NDA guidelines as different solid forms may have different properties (20,140). An often-cited example is ritonavir (141), whose amorphous form is up to 20 times more bioavailable than the crystalline form, and which crystallized into a new less soluble form after 2 years on the market. XRPD is directly mentioned in the ICH decision trees as part of the drug application process, Fig. 19 (140,142). Some NDA submissions include XRPD methods on the drug substance.

It is often necessary to establish a specification or test (e.g., XRPD or IR) to ensure that the proper drug form is manufactured. Failure to do so may result in the manufacture of different polymorphs with different properties, potentially altering the dosage form performance (as in the case of ritonavir). Regulatory authorities are aware of such challenges and expect the manufacturer to demonstrate control over the

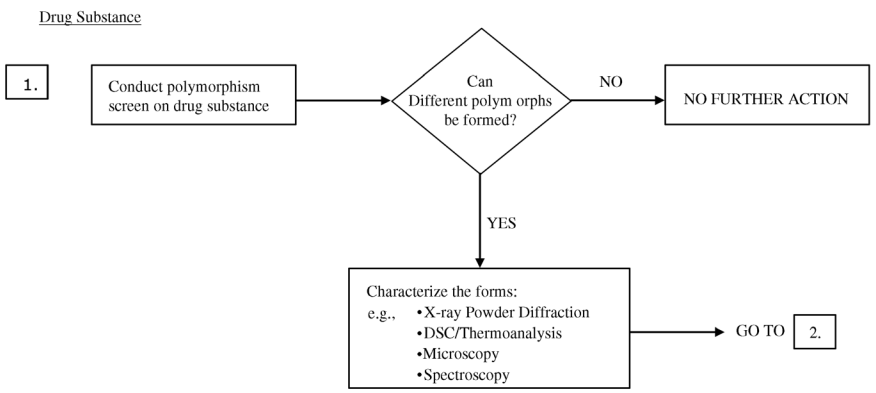


***FIGURE 19*** *ICH Q6A decision tree directing characterization of the forms by X-ray powder diffraction (adapted by author from Reference (140)).*

crystalline or amorphous form in the manufacturing process. Production scale-up requires a process be developed that can reproducibly make the desired polymorph. Where mixtures of forms cannot be avoided, quantitative control (e.g., using XRPD as outlined earlier) is typically required to ensure that the ratios between forms are maintained. Furthermore, this requirement may extend beyond the manufacturing process and through the retest date of the drug substance and potentially throughout the shelf life of the product, which can be a difficult requirement if the forms interconvert. Quantitative XRPD techniques are typically validated by following the ICH Q2A and Q2B guidance (139). It is therefore advantageous to select the most stable solid polymorph for production, determined using a process called polymorph screening in which XRPD plays a major role as outlined earlier. (The use of the most stable form would ensure that there would be no conversion into other forms).

Certain USP monographs require the use of XRPD on the drug substance. One example is the carbamazepine monograph which states that the X-ray diffraction pattern conforms to that of USP carbamazepine RS, similarly determined (140). The exact method of certification used by the USP in the reference standard is not known. In addition, USP recommends the precision to use when comparing peak positions between different XRPD patterns, that is, a peak in one XRPD pattern that falls within the USP-specified range of a peak in the other pattern is to be considered the same. That precision has varied from 0.1 to 0.2 $^{\circ}2\theta$, depending on the year of USP issue (24).

XRPD has been identified as a unique tool for drug substance and/or product analysis that can directly contribute to FDA's "quality by design" initiative. In addition, this tool can help improve the time to market by providing information regarding the structure of the molecule early on in the development process. Finally, XRPD methods can be used to ensure the drug product and/or substance is consistent and has the same identity, purity, and potency (140).

## 4.2 API Identification

***4.2.1 Analysis of a Drug Product*** XRD is frequently applied to the analysis of a drug product. A simple XRD method used for this purpose might involve the following steps (140):

1. Determination of the XRD patterns of all solid components in the drug product.
2. Selection of a region or regions in $2\theta$ where only the polymorph of interest (analyte) diffracts.
3. Preparation of reference physical mixtures containing all API forms as well as excipients.
4. Development of computational/mathematical methods.
5. Validation of the method following the ICH Q2A and Q2B methods.
6. Application of the method.

Chemometric (PLS) methods (143–146) may also be used to increase sensitivity and lower the detection limit. Such analytical XRD methods can have a variety of

applications ranging from online process monitoring to support in litigation of drug product counter-fitting.

One example of the above approach that also includes quantitative analysis of the active ingredient is the XRD method developed by Suryanarayanan (147). The method was applied to intact tablets and required preparation of mixtures containing various weight fractions of the drug and excipient. The quantitative analysis relied on identifying integrated intensities of several diffraction lines for each compound and calculating their ratios as a function of the weight loading of the drug in the tablet. The reported error for the method was less than 10% for a drug loading of 40% weight-in-tablet.

***4.2.2 Monitoring Effects of Processing and Manufacturing on the API*** XRD has been used to study the effects of processing and manufacturing on the solid form of an API. Many of the previously mentioned XRPD methods used for API characterization can also be applied for this purpose.

For example, the antiviral drug MCC-478, known to exist in several anhydrous polymorphic forms as well as a hydrate, was studied in tablet formulation using XRD (148). The XRD patterns of the tablets containing the various forms of the MCC-478 revealed at least one peak unique to each form—generally a prerequisite for this type of work. A semi-quantitative method was developed to characterize the physical form of the API in intact film-coated tablets despite the relatively low weight loading of the API (<20%) in the tablet that also contained a highly crystalline excipient, mannitol, in much greater proportion (60%). The use of this method confirmed that the aqueous film-coating process did not change the form of the API. Furthermore, the same method was used to verify that the API form remained stable (unchanged) over 6 months under accelerated aging conditions (40°C/75% RH). Therefore, the XRD method found application not only for process control during manufacturing, but also for the quality control of the final product.

With increased emphasis by regulatory authorities on understanding and controlling production processes, the implementation of online XRD methods as part of Process Analytical Technology (PAT) is gaining momentum (140). Production processes that are well understood and controlled are considered to be lower risk than those that have not been subject to PAT. It should be noted that spectroscopic methods (NIR, IR, Raman) are also frequently used to monitor polymorphic conversion during processing. Each method has its advantages and disadvantages but typically XRD methods are capable of analyzing the largest sample volume simultaneously and have been used extensively for this purpose in the cement, gypsum and catalyst industries.

An example application of XRD online monitoring was a study of polymorphic conversion during wet granulation (140,149). Wet granulation is a size enlargement process that uses a binder and water to form larger agglomerates, potentially causing phase transformations during the water addition step. Two enantiotropically related forms of flufenamic acid were monitored with a transition temperature of 42°C. XRD was able to identify the form present at various steps of the process.

Online XRD methods used to monitor crystallization of a drug substance have been the subject of extensive research (e.g., Reference (150)). Online processing cells have

been developed for the examination of temperature-dependant polymorphic changes in pharmaceutical materials using *in situ* XRD. Such cells can recirculate the crystallizing solution and monitor the changes by XRD.

***4.2.3 Intellectual Property Considerations*** XRPD is frequently used to identify and/or characterize new materials in patent application filings. Typically, a list of XRPD peak positions is provided, listing $2\theta$ and/or $d$-space position and relative (to maximum) intensity of crystalline peaks. Estimated experimental errors in position ($2\theta$ and/or $d$-space) may be provided, often adopted from the USP (24). It is important to remember that aforementioned instrumental and sample artifacts can and do affect peak positions and intensities. Therefore, providing strict limits on peak positions or relying on intensity information in patent claims yields data that are too narrowly defined. A skilled patent practitioner should therefore be consulted on how best to incorporate solid-state data into a patent application.

The XRPD peak listings may include a list of all observed peaks that can be visually identified above the noise level in the pattern (typically determined through an intensity threshold). However, since it is unlikely that all such peaks will be reproduced in every XRPD pattern of the material, a subset of low angle, non-overlapping peaks with strong intensity (10% or more of maximum observed peak intensity) may be included as peaks "representative" of the material. In this case, it is helpful to have access to XRPD data collected on more than one type of diffractometer geometry in order to assess the reproducibility of peak intensities, which can frequently be affected by for example, preferred orientation and poor particle statistics in a sample.

Where materials exhibit polymorphism and multiple forms are found, it is typical to provide a listing of characteristic (unique) peaks that differentiate one form of a material from all the other known forms of the same material. For different salts or co-crystals where the material can be differentiated chemically, this consideration may be less important but it still applies to polymorphs of a given salt/cocrystal.

## REFERENCES

1. Wunderlich B. A classification of molecules and transitions as recognized by thermal analysis. Thermochim Acta 1999; 340/41: 37–52.
2. Yu L. Amorphous pharmaceutical solids: preparation, characterization and stabilization. Adv Drug Deliv Rev 2001; 48: 27–42.
3. Sakurai JJ. Modern quantum mechanics. Reading, MA: Addison Wesley; 1994.
4. Cowley JM. Diffraction physics. 2nd revised edition. Amsterdam: Elsevier Science; 1990.
5. Bish DL, Post JE.In: Ribbe PH, editor. Modern powder diffraction. Reviews in mineralogy. Vol. 20. Washington, DC: Mineralogical Society of America; 1989.
6. Cowley, JM. International Tables for X-ray Crystallography, Vol. C, Part 4.3., pp. 259–429, International Union of Crystallography, ISBN 978-1-4020-4969-9.

7. James RW. The optical principles of the diffraction X-rays. Ithaca, NY: Cornell University Press; 1965.
8. Warren BE. X-ray diffraction. New York: Dover Publications; 1990.
9. de Wolff PM, Taylor JM, Parrish W. Experimental study of effects of crystallite size statistics on X-ray diffractometer intensities. J Appl Phys 1959; 30: 63–69.
10. Parrish W, Huang TC. Accuracy and precision of intensities in X-ray polycrystalline diffraction. Adv X-ray Anal 1983; 26: 35–44.
11. Klug HP, Alexander LE. X-ray diffraction procedures for polycrystalline and amorphous materials. New York: Wiley; 1974.
12. Smith DK, Barett CS. Special handling problems in X-ray diffractometry. Adv X-ray Anal 1979; 22: 1–12.
13. Bunge HJ. Texture analysis in material science. 2nd ed. Gottingen: Cuvillier Verlag; 1993.
14. Jenkins R, Snyder RL. Introduction to X-ray powder diffractometry. In: Winefordner JD, editor. Chemical analysis. Vol. 138. New York: John Wiley & Sons; 1996.
15. Bates S, Kelly R, Ivanisevic I, Shields P, Zografi G, Newman A. Assessment of defects and amorphous structure produced in raffinose pentahydrate upon dehydration. J Pharm Sci 2007; 96 (5): 1418–1433.
16. Newman A, Engers D, Bates S, Ivanisevic I, Kelly R, Zografi G. Characterization of API: polymer mixtures using X-ray powder diffraction. J Pharm Sci 2008; 97 (11): 4840.
17. Billinge SJL, Kanatzidis MG. Beyond crystallography: the study of disorder, nanocrystallinity and crystallographically challenged materials using pairwise distribution functions. Chem Commun 2004; 749–760.
18. Ivanisevic I, Bugay DE, Bates S. On pattern matching of X-ray powder diffraction data. J Phys Chem B 2005; 109 (16): 7781–7787.
19. Bergese P, Bontempi E., Colombo I, Depero LE. Micro X-ray diffraction on capillary powder samples: a novel and effective technique for overcoming preferred orientation. J Appl Cryst 2001; 34: 663–665.
20. *Guidance for Industry: ANDAs: Pharmaceutical Solid Polymorphism Chemistry, Manufacturing, and Controls Information*. U.S. Department of Health and Human Services Food and Drug Administration Center for Drug Evaluation and Research (CDER); 2007.
21. Brittain H. Polymorphism in pharmaceutical solids. New York: Marcel Dekker, Inc; 1999.
22. Varshney DB, Kumar S, Shalaev EY, Kang S-W, Gatlin LA, Suryanarayanan R. Solute crystallization in frozen systems–use of synchrotron radiation to improve sensitivity. Pharm Res 2006; 23 (10): 2368–2374.
23. Blagden N, Davey R, Song M, Quayle M, Clark S, Taylor D, Nield A. A novel batch cooling crystallizer for in situ monitoring of solution crystallization using energy dispersive X-ray diffraction. Cryst. Growth Des 2002; 3 (2): 197–201.
24. United States Pharmacopeia. X-ray diffraction. Natl Formulary 2008; 31 (26): 374.
25. Hertzberg RP, Pope AJ. High-throughput screening: new technology for the 21st century. Curr Opin Chem Biol 2000; 4: 445–451.
26. Barberis A. Cell-based high-throughput screens for drug discovery. Eur Biopharm Rev 2002; (Winter).
27. Johnston PA. Cellular platforms for HTS: three case studies. Drug Discov Today 2002; 7: 353–363.

28. Marquart RG. Katsnelson I, Milne GW A, Heller SR, Jnr GGJ, Jenkins R. A search-match system for X-ray powder diffraction data. J. Appl. Cryst. 1979; 12 (6): 629–634.
29. Gurley K, Kijewski T, Kareem A. First- and higher-order correlation detection using wavelet transforms. J Engr Mech 2003; 129 (2): 188–201.
30. Gilmore CJ, Barr G, Paisley J. High-throughput powder diffraction. I. A new approach to qualitative and quantitative powder diffraction pattern analysis using full pattern profiles. J Appl Cryst 2004; 37: 231–242.
31. Johnson SC. Hierarchical clustering schemes. Psychometrika 1967; 2: 241–254.
32. Borgatti SP. How to explain hierarchical clustering. Connections 1994; 17 (2): 78–80.
33. Hanawalt JD, Rinn HW, Frevel LK. Chemical analysis by X-ray diffraction-classification and use of X-ray diffraction patterns. Ind Eng Chem, Anal Ed 1938; 10: 457–512.
34. Byrn SR, Pfeiffer RR, Stowell JG. Solid-state chemistry of drugs. 2nd ed. West Lafayette, IN: SSCI Inc; 1999.
35. Massa W. Crystal structure determination. 2nd ed. Berlin: Springer; 2004.
36. Visser JW. A fully automatic program for finding the unit cell from powder data. J Appl Cryst 1969; 2: 89–95.
37. Werner P-E, Eriksson L, Wetdahl M. TREOR, a semi-exhaustive trial-and-error powder indexing program for all symmetries. J Appl Cryst 1985; 18: 367–370.
38. Altomare A, Giacovazzo C, Guagliardi A, Moliterni AGG, Rizzi R, Werner P-E. New techniques for indexing: N-TREOR in EXPO. J Appl Cryst 2000; 33: 1180–1186.
39. Coelho AA. Indexing of powder diffraction patterns by iterative use of singular value decomposition. J Appl Cryst 2003; 36: 86–95.
40. Neumann MA. X-cell: a novel indexing algorithm for routine tasks and difficult cases. J Appl Cryst 2003; 36: 356–365.
41. Boultif A, Louer D. Powder pattern indexing with the dichotomy method. J Appl Cryst 2004; 37: 724–731.
42. Mighell AD. Lattice metric singularities and their impact on the indexing of powder patterns. Powder Diffr 2000; 15 (2): 82–85.
43. Wilson AJC. Determination of absolute from relative X-ray intensity data. Nature 1942; 150: 151–152.
44. Looijenga-Vos A, Buerger MJ. International tables for crystallography. Vol. A. 5th ed. International Union of Crystallography; 2005. ISBN 978-1-4020-4969-9 (Online tables).
45. Rietveld HM. Line profiles of neutron powder-diffraction peaks for structure refinement. Acta Crystallogr 1967; 22: 151–152.
46. Rietveld HM. The early days: a retrospective view (Chapter 2). In: Young RA,editor. The rietveld method. Oxford, UK: Oxford University Press; 1993.
47. http://www.ccp14.ac.uk/solution/rietveld_software/index.html.
48. Lewis J, Schwarzenbach D, Flack HD. Electric field gradients and charge density in corundum, alpha-Al2O3. Acta Crystallogr 1982; A38: 733–739.
49. Taylor JC. Rietveld made easy: a practical guide to the understanding of the method and successful phase quantifications. Canberra: Sietronics Pty Ltd; 2001.
50. McCusker LB, Von Dreele RB, Cox DE, Louër D, Scardi P. Rietveld refinement guidelines. J Appl Cryst 1999; 32: 36–50.

51. Schields PJ, Bates S, McClurg RB, Gendron C. Comparison of simulated and experimental XRPD patterns of silver with twin faults using MAUD and DIFFaX. Adv X-ray anal 2008; 51: D076.
52. Will G. Powder diffraction: the Rietveld method and the two stage method to determine and refine crystal structures from powder diffraction data. Berlin, Heidelberg: Springer; 2006.
53. Berman HM, Jeffrey GA, Rosenstein RD. The crystal structures of the alpha and beta forms of D-mannitol. Acta Crystallogr 1968; B24 (3): 442–449.
54. Botez CE. Stephens PW, Nunes C, Suryanarayan R. Crystal structure of anhydrous delta-D-mannitol. Powder Diffr 2003; 18 (3): 214–218.
55. Zhang J. Ebbens S, Chen X, Jin Z, Luk S, Madden C, Patel N. Roberts C. Determination of surface free energy of crystalline and amorphous lactose by atomic force microscopy adhesion measurement. Pharm Res 2006; 23 (2): 401–407.
56. Burt HM, Mitchell AG. Crystal defects and dissolution. Int J Pharm 1981; 9: 137–152.
57. Prasad KVR, Ristic RI, Sheen DB, Sherwood JN. Dissolution kinetics of paracetamol single crystals. Int J Pharm 2002; 238 (1–2): 29–41.
58. Huttenrauch R, Fricke S, Zielke P. Mechanical activation of pharmaceutical systems. Pharm Res 1985; 2 (6): 302–306.
59. Shalaev E, Shalaeva M, Zografi G. The effect of disorder on the chemical reactivity of an organic solid, tetraglycine methyl ester: change of the reaction mechanism. J Pharm Sci 2002; 91 (2): 584–593.
60. Hiestand E, Smith DP. Indices of tableting performance. Powder Tech 1984; 38: 145–159.
61. Wildfong PLD, Hancock BC, Moore MD, Morris KR. Towards an understanding of the structurally based potential for mechanically activated disordering of small molecule organic crystals. J Pharm Sci 2006; 95 (12): 2645–2656.
62. Huttenrauch R, Keiner I. Rekristallisation aktivierter feststoffe beim verpressen. Pharmazie 1977; 32: 129–130.
63. Huttenrauch R. Abhangigkeit der hygroscopizitat von der kristallinitat. Pharmazie 1977; 32: 240–244.
64. Ahlneck C, Zografi G. The molecular basis of moisture effects on the physical and chemical stability of drugs in the solid state. Int J Pharm 1990; 62: 87–95.
65. Shah B, Kakumanu VK, Bansal AK. Analytical techniques for quantification of amorphous/crystalline phases in pharmaceutical solids. J Pharm Sci 2006; 95 (8): 1641–1665.
66. Wellberry TR, Butler BD. Interpretation of diffuse X-ray scattering via models of disorder. J Appl Cryst 1994; 27 (3): 205–231.
67. Keen DA, Dove MT. Total scattering studies of silica polymorphs: similarities in glass and disordered crystalline local structures. Mineral Mag 2000; 64 (3): 447–455.
68. Ferrari M, Lutterotti L. Method for the simultaneous determination of anisotropic residual stresses and texture by X-ray diffraction. J Appl Phys 1994; 76 (11): 7246–7255.
69. Matthies S, Lutterotti L, Wenk HR. Advances in texture analysis from diffraction spectra. J Appl Cryst 1997; 30: 31–42.
70. Billinge SJL, Thorpe MF. Local structure from diffraction. New York: Plenum Press; 1998.
71. Debye P. Scattering from noncrystalline substances. Ann Phys 1915; 46: 809–823.
72. Proffen T. Analysis of occupational and displacive disorder using atomic pair distribution function: a systematic investigation. Z Kristallogr 2000; 215: 1–8.

73. Proffen T, Billinge SJL, Egami T, Louca D. Structural analysis of complex materials using the atomic pair distribution function—a practical guide. Z Kristallogr 2003; 218: 132–143.
74. Bishop M, Bruin C. The pair correlation function: a probe of molecular order. Am J Phys 1984; 52 (12): 1106–1108.
75. Guinier A. X-ray diffraction in crystals, imperfect crystals, and amorphous bodies. New York: Dover Publications; 1994.
76. Peterson PF. Bozin ES, Proffen T, Billinge SJL. Improved measures of quality for the atomic pair distribution function. J Appl Cryst 2003; 36: 53–64.
77. Kauzmann W. The nature of the glassy state and the behavior of liquids at low temperatures. Chem Rev 1948; 43: 219–256.
78. Hancock BC. Disordered drug delivery: destiny, dynamics and the deborah number. J Pharm Pharmacol 2002; 54: 737–746.
79. Angell CA. The old problems of glass and the glass transition, and the many new twists. Proc Natl Acad Sci USA 1995; 92: 6675–6682.
80. Yalkowsky SH. Solubility and solubilization in aqueous media. Oxford: Oxford University Press; 1999.
81. Hancock BC, Parks M. What is the true solubility advantage for amorphous solids? Pharm Res 2000; 17 (4): 397–404.
82. Pouton CW. Formulation of poorly water-soluble drugs for oral administration: physicochemical and physiological issues and the lipid formulation classification system. Eur J Pharm Sci 2006; 29 (3–4): 278–287.
83. Kaushal AM, Gupta P, Bansal AK. Amorphous drug delivery systems: molecular aspects, design and performance. Crit Rev Ther Drug Carrier Syst 2004; 21: 133–193.
84. Serajuddin ATM. Solid dispersion of poorly water soluble drugs: early promises, subsequent problems, and recent breakthroughs. J Pharm Sci 1999; 88 (10): 1058–1066.
85. Bates S, Zografi G, Engers D, Morris K, Crowley K, Newman A. Analysis of amorphous and nanocrystalline solids from their X-Ray diffraction patterns. Pharm Res 2006; 23 (10): 2333–2348.
86. Taylor LS, Zografi G. Spectroscopic characterization of interactions between PVP and indomethacin in amorphous molecular dispersions. Pharm Res 1997; 14 (12): 1691–1698.
87. Vasanthavada M, Tong W.-Q, Joshi Y, Kislalioglu MSS. Phase behavior of amorphous molecular dispersions I: determination of the degree and mechanism of solid solubility. Pharm Res 2004; 21 (9): 1598–1606.
88. Cincic D, Friscic T, Jones W. Isostructural materials achieved by using structurally equivalent donors and acceptors in halogen-bonded cocrystals. Chem Eur J 2008; 14: 747–753.
89. Dong W, Gilmore CJ, Barr G, Dallman C, Feeder N, Terry S. A quick method for the quantitative analysis of mixtures. I. Powder X-ray diffraction. J Pharm Sci 2008; 97 (6): 2260–2276.
90. Zevin LS, Kimmel G. Quantitative X-ray diffracometry. New York: Springer-Verlag; 1995.
91. Clark GL, Reynolds DH. Quantitative analysis of mine dusts. Ind Eng Chem, Anal Ed 1936; 8: 36–42.
92. Chung FH. Quantitative interpretation of X-ray diffraction patterns. I. Matrix-flushing method of quantitative multicomponent analysis. J Appl Cryst 1974; 7: 519–525.

93. Chung FH. Quantitative interpretation of X-ray diffraction patterns. II. Adiabatic principle of X-ray diffraction analysis of mixtures. J Appl Cryst 1974; 7: 526–531.
94. Berger RL, Frohnsdorff GJC, Harris PH, Johnson PD. Application of X-Ray diffraction to routine mineralogical analysis of portland cement. *Highway Research Board Special Report*, Vol. 90; 1966. p 234–253.
95. Ballantyne C.M. Quantitative automatic determination of cement compound composition using X-ray diffraction techniques. *Cement and Concrete Association Report*, 1968.
96. Gutteridge WA. Quantitative X-ray powder diffraction in the study of some cementitive materials. Br Ceram Proc 1984; 35: 11–23.
97. Smith DK, Johnson GG, Wims AM. Use of full diffraction spectra. Both experimental and calculated in quantitative powder diffraction analysis. Aust J Phys 1988; 41: 311–321.
98. Stutzman PE, Pattern fitting for quantitative X-ray powder diffraction analysis of portland cement and clinker. In: 18th international conference on cement microscopy. Houston, TX: International Cement Microscopy Association; 1996.
99. Smith DK, Scheible A, Wims AM, Johnson JL, Ullmann G. Quantitative X-ray powder diffraction method using the full diffraction pattern. Powder Diffr 1987; 2: 73–77.
100. Savitzky A, Golay MJE. Smoothing and differentiation of data by simplified least squares procedures. Anal Chem 1964; 36: 1627–1639.
101. Brent RP. Algorithms for minimization without derivatives. Englewood Cliffs, NJ: Prentice-Hall; 1973.
102. Rietveld HM. A profile refinement method for nuclear and magnetic structures. J Appl Cryst 1969; 2: 65–71.
103. Hill RJ, Howard CJ. Quantitative phase analysis from neutron powder diffraction data using the Rietveld method. J Appl Cryst 1987; 20: 467–474.
104. Bish DL, Chipera SJ. Problems and solutions in quantitative analysis of complex mixtures by X-ray powder diffraction. Adv X-ray Anal 1988; 31: 295–308.
105. Bish DL, Howard SA. Quantitative phase analysis using the Rietveld method. J Appl Cryst 1988; 21: 86–91.
106. Coelho AA. TOPAS user manual. Karlsruhe, Germany: Bruker AXS GmbH; 2003.
107. Larson AC, Von Dreele RB.General Structure Analysis System (GSAS), Los Alamos National Laboratory LAUR, Vol. 86; 2000. p 748.
108. Chipera SJ, Bish DL. FULLPAT: a full-pattern quantitative analysis program for X-ray powder diffraction using measured and calculated patterns. J Appl Cryst 2002; 35 (6): 744–749.
109. Suryanarayan R. X-ray powder diffractometry. In: Brittain HG,editor. Physical characterization of pharmaceutical solids. New York: Marcel Dekker Inc; 1995. p 187–221.
110. Byard SJ. Jackson SL, Smail A, Bauer M, Apperly DC. Studies on the crystallinity of a pharmaceutical development drug substance. J Pharm Sci 2005; 94 (6): 1321–1335.
111. Matsunaga Y, Bando N, Yuasa H, Kanaya Y. Effects of grinding and tabletting on physicochemical stability of an anticancer drug, TAT-59. Chem Pharm Bull 1996; 44: 1931–1934.
112. Nakai Y, Yamamoto K, Terada K, Kajiyama A. Relationships between crystallinity of β-cyclodextrin and tablet characteristics. Chem Pharm Bull 1985; 33: 5110–5112.

113. Pikal MJ, Lukes AL, Lang JE, Gaines K. Quantitative crystallinity determinations for β-lactam antibiotics by solution calorimetry: correlation with stability. J Pharm Sci 1978; 67: 767–772.
114. Black DB, Lowering EG. Estimation of the degree of crystallinity in digoxin by X ray and infrared methods. J Pharm Pharmacol 1977; 29: 684–687.
115. Kamat MS, Osawa T, Deangelis RJ, Koyama Y, DeLuca PP. Estimation of the degree of crystallinity of cefazolin sodium by X-ray and infrared methods. Pharm Res 1988; 5 (7): 426–429.
116. Saleki-Gerhardt A, Stowell JG, Byrn SR, Zografi G. Hydration and dehydration of crystalline and amorphous forms of raffinose. J Pharm Sci 1995; 84 (3): 318–323.
117. Thompson KC, Draper JP, Kaufman MJ, Brenner GS. Characterization of the crystallinity of drugs: B02669, a case study. Pharm Res 1994; 11 (9): 1362–1365.
118. Ohta M, Buckton G. Determination of the changes in surface energetics of cefditoren pivoxil as a consequence of processing induced disorder and equilibration to different relative humidities. Int J Pharm 2004; 269: 81–88.
119. Nakai Y, Fukuoka E, Nakajima S, Hasegawa J. Crystallinity and physical characteristics of microcrystalline cellulose. Chem Pharm Bull 1977; 25: 96–101.
120. Otsuka M, Kaneniwa N. Effects of grinding on the crystallinity and chemical stability in the solid state of cephalothin sodium. Int J Pharm 1990; 62: 65–73.
121. Gubskaya AV, Lisnyak YV, Blagoy YP. Effect of cryogrinding on physico-chemical properties of drugs. I. Theophylline evaluation of particle sizes and the degree of crystallinity, relation to dissolution parameters. Drug Dev Ind Pharm 1995; 21: 1953–1964.
122. Nakai Y, Fukuoka E, Nakajima S, Morita M. Physico-chemical properties of crystalline lactose. I. Estimation of the degree of crystallinity and the disorder parameter by an X-ray diffraction method. Chem Pharm Bull 1982; 30: 1811–1818.
123. Saleki-Gerhardt A, Ahlneck C, Zografi G. Assessment of disorder in crystalline solids. Int J Pharm 1994; 101: 237–247.
124. Suryanarayan R. Evaluation of two concepts of crystallinity using calcium gluceptate as a model compound. Int J Pharm 1985; 24: 1–17.
125. Clas SD, Faizer R, O'Connor RE, Vadas EB. Quantification of crystallinity in blends of lyophilized and crystalline MK-0591 using X-ray powder diffraction. Int J Pharm 1994; 121 (1): 73–79.
126. Otsuka M, Kaneniwa N. Effect of grinding on the degree of crystallinity of cephalexin powder. Chem Pharm Bull 1983; 31: 4489–4495.
127. Matsuda Y. Otsuka M, Onoe M, Tatsumi E. Amorphism and physicochemical stability of spray-dried frusemide. J Pharm Pharmacol 1992; 44: 627–633.
128. Imaizumi H, Nambu N, Nagai T. Stability and several physical properties of amorphous and crystalline indomethacin. Chem Pharm Bull 1980; 28: 2565–2569.
129. Livingston RA, Stutzman PE, Schumann I. Quantitative X-ray diffraction analysis of handmolded brick. In: Baer NS, Fritz S, Livingston RA,editors. Foam conservation of historic brick structures. Shafesbury, UK: Donhead Publishing Ltd; 1998. p 105–116.
130. Crocker LS, McCauley JA. Comparison of the crystallinity of imipenem samples by X-ray diffraction of amorphous material. J Pharm Sci 1995; 84 (2): 226–227.

131. Chen C, Bates S, Morris K. Quantifying amorphous content of lactose using parallel beam X-ray powder diffraction and whole pattern fitting. J Pharm Biomed Anal 2001; 26: 63–72.

132. Seyer JJ, Luner PE, Kemper MS. Application of diffuse reflectance near-infrared spectroscopy for the determination of crystallinity. J Pharm Sci 2000; 89 (10): 1305–1316.

133. Craig DQM, Royall PG, Kett VL, Hopton ML. The relevance of the amorphous state to pharmaceutical dosage forms: glassy drugs and freeze dried systems. Int J Pharm 1999; 179: 179–207.

134. Fukuoka E, Makita M, Yamamura S. Some physicochemical properties of glassy indomethacin. Chem Pharm Bull 1986; 34: 4314–4321.

135. Yoshioka M, Hancock BC, Zografi G. Crystallization of indomethacin from the amorphous state below and above its glass transition temperature. J Pharm Sci 1994; 83 (12): 1700–1705.

136. Hancock BC, Shamblin SL, Zografi G. Molecular mobility of amorphous pharmaceutical solids below their glass transition temperature. Pharm Res 1995; 12: 799–806.

137. Oberholzer ER, Brenner GS. Cefoxitin sodium: solution and solid state chemical stability studies. J Pharm Sci 1979; 68 (7): 836–866.

138. Mullins J, Macek T. Some pharmaceutical properties of novobiocin. J Am Pharm Assoc Sci Ed 1960; 49: 245–248.

139. http://www.fda.gov/cder

140. Byrn SR, Bates S, Ivanisevic I. Regulatory aspects of X-ray powder diffraction part II. Am Pharm Rev 2005; 8 (5): 137–141.

141. Law D, Schmitt EA, Marsh KC, Everitt EA, Wang W, Fort JJ, Krill SL, Qiu Y. Ritonavir-PEG 8000 amorphous solid dispersions: in vitro and in vivo evaluations. J Pharm Sci 2004; 93 (3): 563–567.

142. Byrn SR, Bates S, Ivanisevic I. Regulatory aspects of X-ray powder diffraction, part I. Am Pharm Rev 2005; 8 (3): 55–59.

143. Höskuldsson A. Prediction methods in science and technology. Vol. 1. Copenhagen: Thor Publishing; 1996.

144. Höskuldsson A. PLS regression methods. J Chemom 1988; 2: 211–228.

145. Wold S, Johansson E, Cocchi M. PLS—partial least squares projections to latent structures. In: Kubinyi H,editor. 3D QSAR in drug design, theory, methods and applications. Leiden: ESCOM Science Publishers; 1993.

146. Eriksson L, Johansson E, Kettaneh-Wold N, Wold S. Multi- and megavariate data analysis—principles and apllications. Umeå, Sweden: Umetrics; 2001.

147. Suryanarayanan R, Herman CS. Quantitative analysis of the active tablet ingredient by powder X-ray diffractometry. Pharm Res 1991; 8 (3): 393–399.

148. Yamada H, Suryanarayanan R. X-ray powder diffractometry of intact film coated tablets—an approach to monitor the physical form of the active pharmaceutical ingredient during processing and storage. J Pharm Sci 2007; 96 (8): 2029–2036.

149. Davis T, *Ph.D. Thesis*, Purdue University, West Lafayette; 2003.

150. MacCalman ML, Roberts KJ, Kerr C, Hendriksen B. On-line processing of pharmaceutical materials using in situ X-ray diffraction. J Appl Cryst 1995; 28: 620–622.

# *Water Determination By Karl Fischer Titration*

**Michael Margreth, Regina Schlink, and Alfred Steinbach**
*Metrohm International Headquarters, Switzerland*

## 1 INTRODUCTION

Besides pH measurement, weighing and acid–base titration, the determination of the water content is one of the most frequently used methods in laboratories around the world. In addition to methods requiring complex apparatus, such as infrared reflection spectrometry, gas chromatography or microwave spectroscopy, two methods in particular have been able to establish themselves:

### 1.1 Drying Methods (Drying Ovens, Infrared Lamps, and Infrared Balances)

This method is commonly found in various standards, but suffers from the following disadvantages:

- In principle the loss on drying (LOD) is determined, and not necessarily the water content. Apart from water, other volatile components of the sample and/or decomposition products are also determined. In order to obtain comparable results, work must be carried out under strictly controlled conditions (temperature and duration of the drying process).

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*,
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

- It takes a long time to obtain the analytical results (several hours in a drying oven). This can lead to bottlenecks, for example, in the monitoring of manufacturing processes.

### 1.2 Titration Methods

In contrast to drying, this is a specific and selective method. If no side reactions occur, only water will be determined. The method is rapid (determination normally takes only a few minutes), it can be validated and therefore fully documented. Since its introduction more than 60 years ago (1), it has become indispensable in many laboratories. The name **Karl Fischer** (1901–1958), **KF** for short, is therefore well known in most laboratories.

KF titration allows to determine both free and bound water, e.g., surface water on crystals or the water contained inside them. The method works over a wide concentration range from ppm up to 100% and supplies reproducible and correct results.

## 2 REAGENTS

When developing his new analytical method, Karl Fischer took into account the well-known Bunsen reaction, which is used for the determination of sulfur dioxide in aqueous solutions:

$$SO_2 + I_2 + 2H_2O \rightarrow H_2SO_4 + 2HI$$

This reaction can also be used for the determination of water if sulfur dioxide is present in excess and the acids so produced are neutralized by a base. The selection of pyridine, the base used by Karl Fischer, was actually due to chance: "pyridine was just standing in the shelf". This led to the formulation of the classical KF reagent (KFR), a solution of iodine and sulfur dioxide in a mixture of pyridine and methanol.

### 2.1 Original Formulation According to Karl Fischer

Dissolve 254 g iodine in 5 L of anhydrous methanol. Add 790 g pyridine, mix thoroughly and then add 192 g liquid $SO_2$.

1 mL KFR corresponds to approximately 3 mg $H_2O$.

The titer of this combined reagent was not very stable—titer determinations had to be carried out virtually every day. Help was provided by the use of two-component reagents. The titration solution consisted of 30 g iodine in 1 L methanol. For the reaction solution 60 g $SO_2$ was introduced into a mixture of 300 mL each of methanol and pyridine. The titer of the titration solution prepared in this way (approximately 1.7 mg $H_2O$/mL) only decreased by approximately 1.3% over a 2-month period and could be regarded as being relatively stable.

However, for most applications it is advisable to work with a one-component reagent with a relatively stable titer. In order to obtain such a reagent, the methanol was replaced by ethylene glycol monomethyl ether (Cellosolve).

### 2.2 Further Developments

The noxious odor and toxicity of pyridine was a source of annoyance for many users of the KF method. Work had to be carried out in the fume cupboard, which was helpful to the user but not to the environment. Research was carried out into the use of other bases in order to obtain pyridine-free reagents. During this work it was also found that the KF reaction only proceeded rapidly and stoichiometrically at a pH between 5 and 7.

Among the bases tested with varying success were diethanolamine, sodium acetate, and sodium salicylate. The ideal substance was finally found: imidazole ($C_3H_4N_2$), which is similar to pyridine but odorless and has a higher basicity (2).

The latest trend is to replace the toxic methanol by ethanol. Today there is general differentiation between one-component and two-component reagents; the two compounds are called titrant and solvent in volumetry while in coulometry one has the anode and cathode solutions.

A wide range of KF reagents and auxiliary solutions are offered by various manufacturers. They are intended for particular applications, that is, custom-made:

- One-component and two-component reagents for volumetry and coulometry.
- Pyridine-containing and pyridine-free reagents.
- Special reagents for water determination in aldehydes and ketones.
- Stable volumetric KF reagents with a titer of 1 or 2 mg $H_2O$/mL for the determination of low water contents.
- Solvent additives for fats and oils, with halogenated hydrocarbons (usually undesirable) or without these (e.g., with decanol).
- Solvent additives for polar substances (e.g., salts, sugars).
- Buffer substances or buffer solutions for strongly alkaline or strongly acidic samples.
- Certified water standards from 0.1 to 10 mg/mL $H_2O$ with a density of 1 g/mL.

This wide range of possibilities allows the user to determine the water content in practically any matrix (3). It is certainly no longer worth preparing the reagents yourself!

## 3 CHEMICAL REACTIONS

Karl Fischer presented the following equation for water determination with his reagent:

$$\begin{aligned} &2H_2O + SO_2 \times (C_5H_5N)_2 + I_2 + 2C_5H_5N \\ &\quad \rightarrow (C_5H_5N)_2 \times H_2SO_4 + 2C_5H_5N \times HI \end{aligned}$$

This would result in molar ratios of

$$H_2O : I_2 : SO_2 : \text{pyridine} = 2 : 1 : 1 : 4$$

under the assumption that methanol only functions as a solvent and pyridine forms additive compounds with the acids.

However, this molar ratio was incorrect, as established by the studies of Smith, Bryant, and Mitchell (4). These authors recognized the participation of methanol and determined the following molar ratios:

$$H_2O : I_2 : SO_2 : \text{pyridine} : CH_3OH = 1 : 1 : 1 : 3 : 1$$

In a first partial reaction water forms the (hypothetical) pyridine sulfur trioxide, which then further reacts with methanol:

$$H_2O + I_2 + SO_2 + 3C_5H_5N \rightarrow 2C_5H_5N \times HI + C_5H_5N \times SO_3$$

$$C_5H_5N \times SO_3 + CH_3OH \rightarrow C_5H_5N \times HSO_4CH_3$$

The authors subsequently revised this equation by formulating iodine and $SO_2$ as pyridine adducts.

Based on the pH dependency of the reaction constants Verhoef and Barendrecht (5) later realized that it is not the $SO_2$ in the KF reagent that forms the reactive component, but rather the monomethyl sulfite ion that is formed from $SO_2$ and methanol:

$$2CH_3OH + SO_2 \rightleftarrows CH_3OH_2^+ + SO_3CH_3^-$$

They also established that pyridine does not take part in the reaction but only acts as a buffer substance (such that at the same pH the same reaction rate is achieved with sodium salicylate). Pyridine can therefore be replaced by another suitable base (RN). The added base results in a considerable displacement of the equilibrium to the right:

$$CH_3OH + SO_2 + RN \Rightarrow [RNH]SO_3CH_3$$

This means that the KF reaction in methanolic solution can now be formulated as follows:

$$H_2O + I_2 + [RNH]^+SO_3CH_3^- + 2RN \rightarrow [RNH]^+SO_4CH_3^- + 2[RNH]^+I^-$$

### 3.1 Influence of the Water Content

The water content of the working medium influences the stoichiometry of the KF reaction. If it exceeds 1 mol/L (18 g/L) then the reaction behavior changes in favor of the Bunsen reaction for aqueous solutions. This means that two $H_2O$ are

consumed for one $I_2$ or one $SO_2$. It is uneconomic to titrate such a high water content in a large volume of sample. In this case it is important to dilute the sample and/or to use a small amount of sample. If two-component KF reagents are used then the water capacity of the solvent must be taken into consideration, particularly for subsequent titrations. The water capacity of these solvents is normally about 5–7 mg/mL, for a solvent volume of 25 mL this means maximum 125–175 mg $H_2O$.

### 3.2 Influence of the Organic Solvent

Theoretically non-alcoholic KF reagents can also be used, but in this case the stoichiometry also changes. In protic solvents, for example, methanol or ethanol, the ratio is $H_2O{:}I_2 = 1{:}1$, in aprotic solvents, for example, DMF, the ratio is $H_2O{:}I_2 = 2{:}1$. Intermediate values (uneven ratios) have been observed, for example, with propanol, butanol, and other long-chain alcohols. This has to do with the reactivity of the alcohol used. In methanol about 100%, in propanol approximately 80% and in butanol only approximately 50% is present as sulfite ester. With the KF reagents normally used an alteration in the stoichiometry is not to be expected because:

- the titer determination and water determination are carried out in the same titration medium and
- the alcohol fraction in the sample solution continuously increases during the course of the titration.

### 3.3 Kinetics of the KF Reaction

Cedergren [6,7] was presumably the first to thoroughly investigate the KF reaction sequence and establish that the reaction rate increases as the concentrations of $SO_2$, $I_2$, and $H_2O$ increase. The reaction is first order referring to each individual component. For the rate constant K he used the equation:

$$-\mathrm{d}[\mathrm{I_2}]/\mathrm{dt} = \mathrm{K}[\mathrm{I_2}] \times [\mathrm{SO_2}] \times [\mathrm{H_2O}]$$

to obtain the value:

$$K = (1.2 \pm 0.2) \times 10^3\ \mathrm{L^2/mol^2/s} \qquad \log K = 3.08 \pm 0.08$$

Verhoef and Barendrecht confirmed the results of Cedergren, but at the same time found that the reaction constant $K$ depends on the pH of the solution:

- Up to pH = 5, log $K$ increases linearly with the pH.
- Between pH = 5.5 and 8 the reaction rate is constant (there is a plateau).
- Above pH 8, the reaction rate again increases slightly (probably due to side reactions).

From the increases in $K$ under acidic conditions the authors concluded that not $SO_2$, but the sulfite base is the reactive component.

The reaction rate is also strongly influenced by the concentration of iodide ions. The iodine in the KF solution reacts with iodide to form triiodide; this strongly displaces the equilibrium to the right:

$$I_2 + I^- \Rightarrow I_3^-$$

However, during the oxidation of the methyl sulfite the free iodine reacts far more quickly than the triiodide ($\Delta K$ approximately $10^4$). This reaction is also pH-dependent.

## 4 VOLUMETRY AND COULOMETRY

The following are important for both methods (8):

- the KF reagent must be added accurately, reproducibly, and with the highest possible resolution and
- the titration cell must be as impervious as possible. It should be penetrated by as little atmospheric water as possible;
- the water adhering to the walls of the titration cell (inner water film) must be removed by swirling about the conditioned contents of the titration vessel.

### 4.1 Volumetry

It is only since the introduction of piston burets by Metrohm in the 1950s that it has been possible to carry out volumetric Karl Fischer titrations without requiring a cumbersome setup. The introduction of the Exchange Units (also by Metrohm) in the 1960s brought further improvements.

By way of example, Metrohm Titrinos titrate with a resolution of 10 000 pulses per buret volume, while Metrohm Titrandos have a resolution of 20 000 pulses per buret volume. Normally piston burets with a volume of 5, 10, or 20 mL are used for the titration; this means that the Titrinos have a resolution of 0.5–2 μL and the Titrandos' resolution is even 0.25–1 μL.

Although this resolution allows a high precision, this does not mean that titrations can be carried out down to the resolution limit. Even with impervious titration cells the introduction rate of atmospheric humidity can amount to a reagent consumption of 20 μL per min. Modern titrators determine this value as the blank drift and automatically include it in the final calculation.

Higher (absolute) water contents are preferably determined by volumetric titration. Volumetry also has the advantages that solid or pasty samples can be introduced directly into the titration vessel and work can be carried out with a variety of suitable organic solvents specially adapted to particular sample.

Of course it is also possible to determine small amounts of water volumetrically, for example, in solvents. If an appropriate amount of sample and diluted KF reagents are used then the lower determination limit is approximately 50–100 ppm $H_2O$.

Despite great progress in the manufacture of KF reagents, these do not have a stable titer. One of the minor disadvantages of volumetry is therefore that the titer has to be determined at regular intervals.

## 4.2 Coulometry

Instead of a buret the electric current is used for generating a reagent—an "electronic buret", so to speak. The current releases the stoichiometrically corresponding amount of iodine from the iodide-containing KF reagent by electrolysis.

Faraday's law applies: $m = \frac{M \times Q}{z \times F}$

where m is the mass of converted substance in g, M is the molar mass in g/mol, Q is the measured amount of charge in ampere-seconds, z is the number of exchanged electrons (equivalence number, charge number), $F$ is the electrochemical equivalent (1 F = 96.485 coulomb/mol, 1 coulomb = 1 C = 1 ampere-second = 1 A s).

Example for iodine: $2I^- - 2e^- \rightarrow I_2$

$$z = 2,\ M = 253.8(I_2)$$

126.9 g iodine is released by 96,485 A in 1 s—or 1.315 mg iodine is generated by 100 mA in 10 s.

Requirements for coulometric titration are

- The process must take place with 100% current efficiency, that is, there must be no side reactions.
- Oxidation or reduction must lead to a defined oxidation stage.

Current KF Coulometers meet these requirements with modern coulometric reagents. They work according to the galvanostatic coulometry principle, that is, with a constant current.

The same chemical processes take place as in a volumetric KF titration, that is, $H_2O$ consumes one $I_2$. As the iodine is generated electrolytically from the iodide in the KF solution this means that the coulometric determination of water is an absolute method—a titer does not need to be determined!

In order to generate iodine at the anode, Metrohm KF Coulometers work with variable current strengths and pulse lengths (100, 200, and 400 mA—with diaphragm-less cells with a constant current strength of 400 mA). Higher current strengths have not been able to establish themselves—side reactions occur and heat is produced. Working with variable pulse lengths allows precise "iodine addition" even in the region of the titration endpoint.

The instrument measures the time and current flow that is required to reach the titration endpoint. The product of "time × current" is directly proportional to the amount of iodine generated and therefore to the amount of water determined.

Coulometric water determination is primarily used for the determination of small amounts of water. Larger amounts of water require a lot of time and/or may exceed the water capacity of the KF reagent, which could lead to incorrect results. Modern

KF Coulometers typically have a determination range of 10 μg–200 mg $H_2O$ with a resolution of 0.1 μg $H_2O$.

## 5 DETECTION METHODS

In order to determine the end of the KF reaction—that is, the titration endpoint at which all the $H_2O$ has been converted—a detection or indication method is required. Titrimetric volumetric analysis stands and falls with correct endpoint recognition. This means that the indication method is crucial as it determines to a great extent the reproducibility and, above all, the correctness of the results.

### 5.1 Visual or Photometric Indication

Karl Fischer himself only had the possibility of determining the titration endpoint visually. This method requires a high degree of routine and can normally only be used for uncolored sample solutions. When excess iodine is present the solution turns increasingly yellow to brown (the latter with a large excess of iodine). This already indicates the difficulty of deciding the coloration point to which the titration is to be continued. In addition, the coloration in polar solvents (e.g., methanol) differs from that in nonpolar solvents (e.g., chloroform, DMF). Further disadvantages of these methods are that they can hardly be automated or validated. They fell out of use rapidly as electrometric indication methods became available toward the end of the 1940s.

### 5.2 Electrometric Indication Methods

In the meantime indication methods using two polarized electrodes have established themselves. Double Pt-wire or double Pt-ring electrodes are commonly used.

***5.2.1 Biamperometric Indication ($U_{pol}$ or Dead Stop)*** In this case a constant voltage of max. 500 mV is applied to the electrodes and the resulting current is measured. This method was first used mainly for aqueous iodometric titrations. The following reactions occur at the electrodes:

$$\text{Cathode}: \quad I_2 + 2e^- \rightarrow 2I^- (\text{reduction})$$

$$\text{Anode}: \quad 2I^- - 2e^- \rightarrow I_2 (\text{oxidation})$$

The relationship between the voltage, current, and concentration of the depolarizer (iodine or triiodide) is similar to that found in polarography, that is, initially only low voltages are required to produce a flow of current. Then the voltage increases steeply—as a function of the iodine concentration—and, when the limiting diffusion current has been achieved, the current increases only slightly. If the voltage is then further increased decomposition of either the solvent or conductive salt could occur.

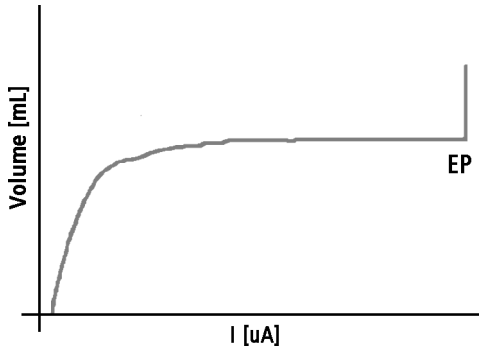


***FIGURE 1*** *Biamperometrically indicated titration curve.*

This means that a current flows for as long as iodine is present. When the iodine has been consumed (e.g., by titration with thiosulfate) the current drops abruptly to 0 (which is where the name "dead stop" comes from), the resistance between the electrodes increases strongly and the electrodes are again polarized. In KF titration the reverse occurs. If excess $H_2O$ is present only a minimal current flows; this increases to a few μA when excess iodine is present. The "water loss" resulting from electrochemical conversion can be ignored. At 50 μA for 30 s it is theoretically approximately 1 μg $H_2O$. Figure 1 shows an example of a biamperometrically indicated titration curve:

***5.2.2 Bivoltametric Indication ($I_{pol}$)*** In this case, a small direct or alternating current is applied between the electrodes and the resulting voltage is measured. AC is much more sensitive than DC, which is why AC is usually only used for coulometric titrations. The shape of the titration curve is similar to that obtained by biamperometry; however, the jumps are usually more marked and larger. If the sample solution and titrant are electrochemically active then peak-shaped or V-shaped titration curves (Fig. 2a) are obtained. If only one partner is electrochemically active (the KFR with excess iodine) then L-shaped curves are obtained (Fig. 2b).

The resulting voltages for excess iodine or water depend, of course, on the applied current ($I_{pol}$). Three examples (double Pt-wire electrode):

$I_{pol} = 50$ μA, $U$ approximately 100 mV for excess iodine, approximately 680 mV for excess $H_2O$;
$I_{pol} = 10$ μA, $U$ approximately 5 mV for excess iodine, approximately 580 mV for excess $H_2O$;
$I_{pol} = 1$ μA, $U$ approximately 0 mV for excess iodine, approximately 350 mV for excess $H_2O$.

The titration endpoint is determined by entering a switch-off voltage; reagent addition ceases when this limit is undercut. This "ideal endpoint" is best found

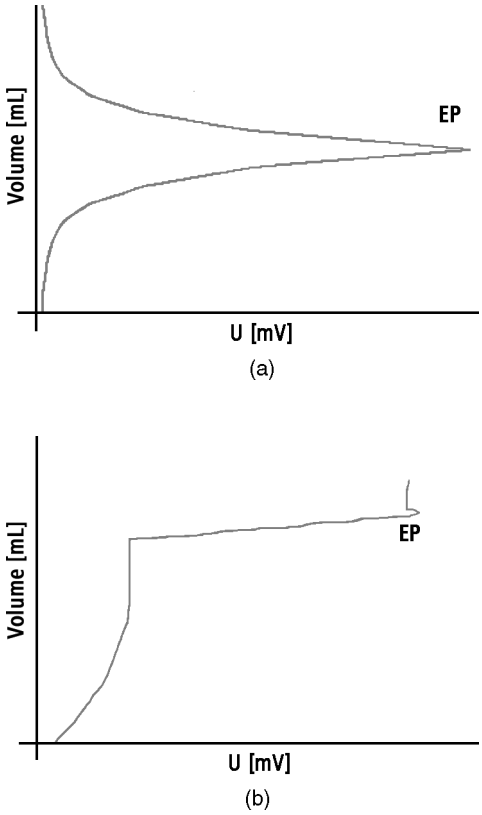


***FIGURE 2*** *Examples of bivoltametrically indicated titration curves.*

experimentally. If a too low voltage is selected then the amount of excess iodine required to achieve the switch-off point will be too large. If it is too high then the titration will not start because no free iodine is required for this voltage to be achieved (decomposition of the solvent or electrolyte).

Once all the water has reacted, the indicator electrode detects the very slight excess of iodine and the Karl Fischer titration stops. In comparison to the visual methods, the electrometric methods have the advantage that the titration is always carried out to the same slight excess of iodine and in this way a better reproducibility and accuracy can be achieved.

## 6 INSTRUMENTS

### 6.1 Volumetric Karl Fischer Titration

Different manufactures offer instruments for volumetric KF titration, ranging from simple instruments for routine determinations up to high end equipment with

***FIGURE 3*** *901 Titrando with 900 Touch Control and 803 KF Titration Stand (Metrohm AG).*

numerous advantages. What all instruments have in common is that they are easy to operate and produce rapid, accurate, and reproducible results.

The Karl Fischer titrators, one example is shown in Fig. 3, are automated titration systems that carry out the titration automatically, that is, the titration rate and volume increments are controlled according to the signal measured by the indication system. In addition, the sample weight can be transmitted directly to the titration system via a connected balance. Results, curves, parameters, and the configuration data can either be printed out or transferred to a PC database.

The titrators have both bivoltametric ($I_{pol}$) and biamperometric ($U_{pol}$) endpoint detection. For both types of detection, the control parameters can be freely selected. This has the advantage that the control parameters can be optimized for difficult samples and special reagents, for example, for the addition of solubility promoters. Good results are normally obtained with the default settings.

Slowly reacting reagents (e.g., those containing pyridine) require an additional stop delay when the endpoint is reached, that is, the endpoint voltage or the endpoint current must be maintained for a few seconds when the endpoint has been reached. In the course of the development of new reagents and the improvement of the instruments, this stop delay has been replaced by the drift-stop criterion. In addition to a pre-defined voltage (or current), a certain drift value must also be achieved. The drift is the amount of consumed KF reagent per unit time to keep the titration cell dry. In general, the drift value for titration cells titrated to dryness is in the order of a few μL/min. The aim of the absolute drift-stop correction is to achieve the drift value measured before the titration. In practice it has been shown that the

achievement of an identical drift can considerably increase the titration time, which is why the use of a relative drift-stop criterion is recommended. If a relative drift stop is used then the titration will be stopped at a slightly higher drift than the initial drift.

If very small amounts of water are to be determined, or if a long titration time is necessary, then work should be carried out using automatic drift correction. In this case the drift value is multiplied with the titration time and the value obtained is subtracted from the added volume.

In addition to the indication system, the resolution of the buret makes a contribution to the accuracy and reproducibility of the results. Metrohm standard burets, for example, add the reagent with a resolution of 10,000 steps per cylinder volume. For a 5 mL buret, this means that in the region of the endpoint, addition will be carried out in 0.5 μL steps. If a Titrando is used in combination with an intelligent Exchange Unit, you can store not only the current titer on the data chip of the Exchange Unit but also carry out reagent addition with the unique resolution of 20,000 steps per cylinder volume.

In principle all commonly available KF reagents can be used with the current KF titrators. Since these reagents do not have a stable titer, it must be determined regularly. The interval between the titer determinations depends on the choice of reagent as well as on how impervious the reagent bottle is.

Leading reagent manufacturers offer certified water standards for determining the titer. It is also possible to determine the titer with ultrapure water. In this case work has to be carried out with very small volumes and handling takes some getting used to. This is why the use of certified standards is recommended; it is possible to back-weigh the added amount and volume dosing is no longer necessary. Solid substances such as di-sodium tartrate dihydrate can also be used, but care must always be taken that the substance is completely dissolved. Standard solutions are available with water contents of $10.0 \pm 0.1$ mg/g (for volumetric water determination), $1.0 \pm 0.03$ mg/g, or $0.1 \pm 0.005$ mg/g (for coulometric determinations) and are offered in ampoules. When handling the standards proceed as follows:

Immediately after the ampoule is opened a syringe is approximately half-filled with the standard solution. The syringe is rinsed by inverting it and ejecting the complete contents (air and standard solution). This is to ensure that the inner walls of the syringe are completely wetted with the solution. The syringe is then filled with the whole of the remaining ampoule contents without entraining any air. After the needle has been wiped off the syringe is tared on a balance and the KF titrator is started. Part of the standard solution is then injected at one go. There are two ways of doing this:

- The standard solution is injected directly beneath the surface of the KF solution.
- It is injected without immersing the needle. The last drop is drawn back into the syringe.

The syringe is then pulled out and back-weighed. The sample weight can be transmitted directly from the balance to the KF titrator or entered manually.

Particular attention should also be given to the titration cell. Water is present everywhere: at a relative humidity of 60%, 0.5 mL air at 25°C contains approximately 7 μg $H_2O$. This is why it is essential that the titration cell is impervious; it should be checked at regular intervals. Special attention should be paid to the sealing rings of the titration cell and the molecular sieve—check the drift! The titrator automatically conditions the titration cell before and between determinations. It is advisable to swirl the cell about from time to time in order to remove any water adhering to the walls. It is important that a constant drift value that is as low as possible is achieved.

## 6.2 Coulometry

In contrast to volumetric Karl Fischer titration, in KF coulometry no titrant is added but the required iodine is generated directly in the iodide-containing electrolyte by electrochemical means ("electronic buret"). There is a strictly quantitative relationship between the amount of electric charge and the amount of iodine generated. This relationship allows the extremely precise addition of the iodine.

KF Coulometers such as the 851 Titrando (Fig. 4) work with variable current strengths and pulse lengths (100, 200, and 400 mA) for generating iodine at the anode.

***FIGURE 4*** *851 Titrando with 900 Touch Control, 801 Stirrer, and 800 Dosino (Metrohm AG).*

PSE415 03/26/2013 8:54:23 Page 14



Higher current strengths have not proven themselves in practice—side reactions can occur and excess heat is produced. Working with variable pulse lengths allows precise "iodine dosification" even in the region of the titration endpoint.

The instrument measures the time and current flow that is required to reach the titration endpoint. The product of time and current is directly proportional to the amount of iodine generated and therefore the amount of water determined.

Coulometric water determination is primarily used for the determination of small amounts of water. Larger amounts of water require a lot of time and/or may exceed the water capacity of the KF reagent, which could lead to incorrect results. Metrohm KF Coulometers, for example, have a determination range of 10 μg–200 mg $H_2O$ with a resolution of 0.1 μg $H_2O$.

The iodine is produced by a generator electrode. There are two different types of generator electrodes:

A. *Generator electrode without diaphragm [9–11]:* The generator electrode without diaphragm (Fig. 5) is easy to handle and easy to clean. Only a single reagent is required. However, care must be taken that only those reagents that are intended for generator electrodes without diaphragm are used. As no moisture can adhere to the diaphragm there is a further advantage—the electrode is rapidly ready for use.
B. *Generator electrode with diaphragm:* A generator electrode with diaphragm (Fig. 5) should be used whenever your samples contain ketones and aldehydes, as the special reagents used for them are only available for generator electrodes with diaphragms. An electrode with diaphragm should also be used for reagents with a low conductivity (the addition of chloroform, for

***FIGURE 5*** *Generator cells for KF Coulometers: on the left without, on the right with diaphragm (Metrohm AG).*

example, reduces the conductivity) and when measurements are to be made in the lower trace range. Reagents for coulometric water determination consist of an anolyte that is filled into the titration vessel and a catholyte that is filled into the cathode compartment. Both liquids should have about the same level in order to prevent pressure differences from forcing the anolyte into the cathode compartment and vice versa.

The endpoint is determined by a double platinum electrode and is largely similar to that in volumetric Karl Fischer titration.

As the coulometric KF method is an absolute method this means that no titer has to be determined. It must only be ensured that the reaction that generates the iodine takes place to 100%. Although this is the case with the reagents that are available, the whole system should be checked from time to time with the aid of certified water standards with a water content of $1.0 \pm 0.03$ mg/g or $0.1 \pm 0.005$ mg/g. Handling the standards is described in 6.1 "Volumetric KF titration".

In coulometry, cells made completely of glass are used as titration cells, as cells with a plastic cover have a certain permeability to water vapor. This does not have much influence on volumetric KF titrations, as these are normally used to determine higher water contents than in KF coulometry. Care must also be taken that the cell is as impervious as possible.

### 6.3 Karl Fischer Oven

Many substances only release their water slowly or at elevated temperatures. This means that they are not suitable for direct Karl Fischer titration. A further problem is the low solubility of certain samples in alcohols. In such cases traditional methods recommend complicated sample preparation procedures or the use of solubility promoters, which represent a health hazard. Other substances react with the KF reagents to release water or consume iodine; this falsifies the results.

These problems can be avoided by using a heating method. The substance under investigation is heated in an oven located upstream from the cell and the released water is transferred by a flow of dry carrier gas to the titration cell where it is determined by Karl Fischer titration. As only the water enters the titration cell and the sample itself does not come into contact with the KF reagent, this means that side reactions and matrix effects are ruled out. Work is carried out either with a conventional tube oven or with an oven using the vial technique:

The 860 KF Thermoprep uses the vial technique. The samples are weighed out directly into vials that are then hermetically sealed and placed in the oven. A double hollow needle pierces the vial septum. A stream of dry carrier gas (air or inert gas) is introduced through the inlet needle and passes through the heated sample. The carrier gas containing the released moisture then passes through the outlet needle and via a heated transfer tube directly into the titration cell, in which the Karl Fischer water determination takes place. Depending on the water content of the sample the determination can be carried out volumetrically or, in the trace range, coulometrically. Working with vials means that the oven cannot be contaminated

by the sample and, as a result, no carryover or memory effects can occur. In addition, the hermetic vial seals with PTFE-coated septa reliably prevent the water content from being falsified by atmospheric moisture.

## 6.4 Automation in Karl Fischer Titration

The use of a sample changer is well worthwhile for large numbers of samples. The advantages are not only the amount of time saved by the laboratory personnel: automated systems control the operating procedures and improve both reproducibility and accuracy. In automated KF titration there is an additional difficulty; normal KF sample changers do not work in closed systems and penetration by atmospheric moisture cannot be prevented. For a water content $>1\%$ this interference can be corrected by the subtraction of a previously determined blank value. In principle, the use of such sample changers can only be recommended for volumetric determinations.

The MATi 10 system in Fig. 6 allows the automation of volumetric Karl Fischer titrations. The samples are weighed out into beakers, sealed to protect them against atmospheric moisture, and placed on the rack. Before titration the samples are

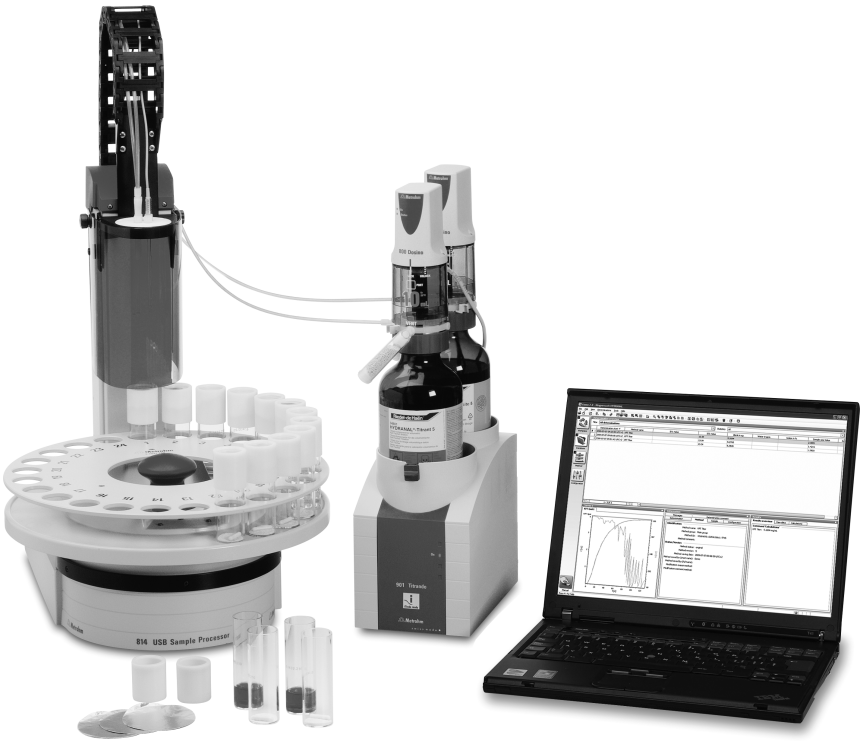


***FIGURE 6*** *The MATi 10 system for the automation of volumetric Karl Fischer titrations (Metrohm AG).*

***FIGURE 7*** *The MATi 4 system for automated coulometric titrations (Metrohm AG).*

dissolved in a known amount of Karl Fischer working medium and the water content is then titrated in a second step. This method is particularly suitable for samples that dissolve quickly in the KF working medium. The solubility can also be improved by the use of a suitable solubility promoter.

It is also possible, however, to automate the analysis of samples with a water content $<1\%$.

The 815 Robotic USB Sample Processor Xl (Fig. 7) is a fully automated system for coulometric titrations with high sample throughput. As in the oven technique, the samples are weighed out into vials, hermetically sealed with septa and placed on the sample rack. With the aid of a pipetting system adapted to the special conditions of the KF titration (minimization of the penetration of atmospheric humidity), a defined volume of the sample is transported to the coulometer cell where the water content is determined.

By using the 874 USB Oven Sample Processor (Figs 8 and 9), automation of the heating method is also possible. The principle corresponds to that of the 860 KF Thermoprep, with the exception that the sample vials are placed on the rack of the 874 USB Oven Sample Processor. During the analysis, the turntable positions the sample vial above the oven. The vial is then lowered automatically into the heating block.

In addition, the 874 USB Oven Sample Processor offers the possibility of using temperature gradients. The recorded water release curve can be used to determine

PSE415 03/26/2013 8:54:26 Page 18

***FIGURE 8*** *874 Oven Sample Processor with 851 Titrando (Metrohm AG).*

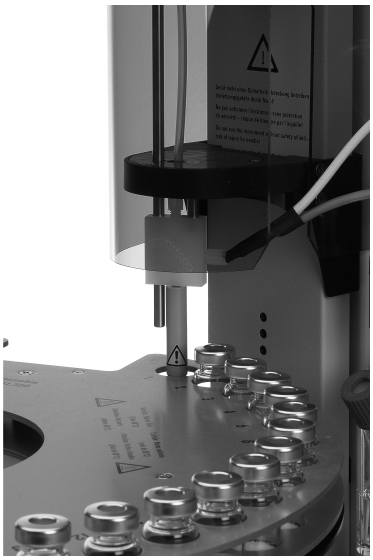

***FIGURE 9*** *Side view of the 874 USB Oven Sample Processor (Metrohm AG).*

PSE415 03/26/2013 8:54:28 Page 19



the optimum heating temperature for the particular sample. The curve also allows statements to be made about the kinetics of water release as a function of the temperature. Depending on the water content of the sample, the titration can be carried out volumetrically or, in the trace range, coulometrically.

## 6.5 Accessories for Karl Fischer Titration

A wide range of accessories makes it easier to carry out Karl Fischer titrations in the laboratory.

Use the 803 KF Titration Stand (Fig. 10) for stirring and manual exchange of the used titration solution. Solvents can be added and aspirated with the built-in membrane pump without having to open the cell. This considerably shortens the conditioning time and minimizes reagent handling by the user.

With KF Coulometers or a Titrando, reagent exchange is even easier to carry out by using a Dosino. All you have to do is define the time and the instrument will carry out the exchange automatically.

***FIGURE 10*** *803 KF Titration Stand.*

***FIGURE 11*** *Weighing spoon for solid samples.*

If special additional steps are necessary, such as titration of a heated solution or homogenization of the sample in the titration cell, then special accessories are available to suit your particular application. Some examples are shown in Figs 11 and 12.

## 7 SIDE REACTIONS

KF water determination is only specific if no side reactions with the KF reagents take place. This means that no water should be released in side reactions, nor should the sample consume or release iodine. Most side reactions can be suppressed by suitable measures. The most important interfering substances are as follows.

### 7.1 Carbonates, Hydroxides, and Oxides

During the KF titration, methyl sulfuric acid and hydriodic acid are produced. Although these are attached to the base, they still form weak acids that react with the above-mentioned compounds to form salts and release water:

$$Na_2CO_3 + 2HI \rightarrow 2NaI + CO_2 + H_2O$$

$$Ca(OH)_2 + H_2SO_4 \rightarrow CaSO_4 + 2H_2O$$

$$MgO + 2HI \rightarrow MgI_2 + H_2O$$

***FIGURE 12*** *Heatable cell for titrations at elevated temperatures.*

Not all oxides take part in an acidic reaction. In addition to alkali oxides and alkaline earth oxides, the presence of the following oxides leads to the formation of water: $Ag_2O$, HgO, $MnO_2$, PbO, $PbO_2$, and ZnO.

We recommend to determine the $H_2O$ in these compounds by the oven method.

## 7.2 Aldehydes and Ketones

These two classes of compounds combine with the methanol in the KFR to form acetals and ketals respectively with the release of water.

$$\text{Aldehyde; } RHC = O + 2CH_3OH \rightarrow \text{Acetal; } RHC(OCH_3)_2 + H_2O$$

$$\text{Ketone; } R_2C = O + 2CH_3OH \rightarrow \text{Ketal; } R_2C(OCH_3)_2 + H_2O$$

With aldehydes the reactivity decreases as the chain length increases. Aromatic aldehydes react more slowly than aliphatic aldehydes. Acetaldehyde reacts the quickest and causes the most problems. However, there are always exceptions to any rule, formaldehyde and chloral do not form acetals and can be titrated without any problems.

Ketones are less reactive than aldehydes. The reactivity also decreases as the chain length increases. Ketal formation is most rapid with acetone and cyclohexanone. The following are regarded as being stable (no ketal formation): diisopropyl

ketone, benzophenone, deoxybenzoin, benzil, benzoin, camphor, alizarin, dibenzalacetone, and dichlorotetrafluoroacetone.

Acetal and ketal formation is also influenced by the alcohol present. As the chain length increases the reactivity decreases. This means that methanol is the quickest to react.

There are several ways of preventing acetal or ketal formation:

a. By using KF reagents specially labeled for this purpose.
b. By using methanol-free KF reagents. Methanol is replaced by a higher alcohol, for example, 2-methoxyethanol.
c. By making use of the different reaction rates and slowing down acetal or ketal formation so much that they do not interfere with the titration. This can be achieved by carrying out the titration at $-10°C$. This also freezes hydrogen sulfite addition (see following section).

With aldehydes a second reaction—hydrogen sulfite addition—can also interfere. In this case $H_2O$ is consumed, so that the water content determined could be too low.

$$RHC = O + SO_2 + H_2O + NR' \rightarrow RHC(OH)SO_3HNR'$$

## 7.3 Strong Acids

Strong acids can react with the methanol contained in the KF reagent and release water:

$$2CH_3OH + H_2SO_4 \rightarrow (CH_3O)_2SO_2 + 2H_2O$$

$$CH_3OH + HOOCH \rightarrow CH_3O\text{-}OCH + H_2O$$

Among the organic acids the strongest esterification is shown by formic acid; acetic acid only undergoes partial esterification. Methanol esterifies the quickest. Slower reaction rates are achieved with propanol or 2-methoxyethanol or both.

However, strong acids also lower the pH value of the reaction solution so much that the KF reaction either stops or proceeds unsatisfactorily. The reaction mixture can be buffered by the addition of a base (e.g., imidazole) and the pH value brought back to the optimal range again. If this measure is taken then no esterification occurs even with strong mineral acids (e.g., $H_2SO_4$ or $HNO_3$).

## 7.4 Silanols and Siloxanes

The terminal silanol groups esterify with the methanol contained in the KF reagents and release water. This means that methanol-free reagents must be used to determine the water content of these substances.

$$R_3SiOH + CH_3OH \rightarrow R_3SiOCH_3 + H_2O$$

## 7.5 Boron Compounds

Boric acid reacts with methanol with esterification and the release of water:

$$H_3BO_3 + 3CH_3OH \rightarrow B(OCH_3)_3 + 3H_2O$$

Boron trioxide ($B_2O_3$), metaboric acid ($HBO_2$), and borax/sodium tetraborate ($Na_2B_4O_7$) all react in the same way.

The esterification rate decreases as the chain length of the alcohol increases (methanol esterifies the quickest).

## 7.6 Metal Peroxides

These form water with KF solutions and cannot be titrated directly. In some cases, the oven method can be used (provided that the substances release their $H_2O$ at $<250°C$).

## 7.7 Reducing Agents

Reducing or oxidizing agents can react with the iodine in the KF reagents and simulate a water content that is too high or too low, respectively.

0.5 mol $H_2O$/mol substance in the sample is simulated by:

$$\text{Iron(III) salts} \quad FeCl_3 \times 6H_2O + 5.5I_2 + 6SO_2 + 6CH_3OH \rightarrow FeI_2 + 9HI + 3HCl + 6CH_3SO_4H$$

$$\text{Copper(II) salts} \quad CuCl_2 \times 2H_2O + 1.5I_2 + 2SO_2 + 2CH_3OH \rightarrow CuI + 2HI + 2HCl + 2CH_3SO_4H$$

$$\text{Nitrites} \quad KNO_2 + 0.5I_2 + SO_2 + CH_3OH \rightarrow KI + NO + CH_3SO_4H$$

$$\text{Thiosulfates} \quad 2Na_2S_2O_3 + I_2 \rightarrow 2NaI + Na_2S_4O_6$$

1 mol $H_2O$/mol substance in the sample is simulated by:

$$\text{Ascorbic acid (Vitamin C)} \quad C_6H_8O_6 + I_2 \rightarrow C_6H_6O_6 + 2HI$$

$$\text{Disulfites} \quad Na_2S_2O_5 + I_2 + CH_3OH \rightarrow 2NaI + SO_2 + CH_3SO_4H$$

$$\text{Hydrazine and its derivatives} \quad 2N_2H_4 + 2I_2 \rightarrow N_2H_6I_2 + N_2 + 2HI$$

$$\text{Selenites} \quad Na_2SeO_3 + I_2 + 3SO_2 + 3CH_3OH \rightarrow 2NaI + Se + 3CH_3SO_4H$$

$$\text{Sulfites} \quad Na_2SO_3 + I_2 + CH_3OH \rightarrow 2NaI + CH_3SO_4H$$

$$\text{Tellurites} \quad Na_2TeO_3 + I_2 + 3SO_2 + 3CH_3OH \rightarrow 2NaI + Te + 3CH_3SO_4H$$

$$\text{Tin(II) salts} \quad SnCl_2 \times 2H_2O + 3I_2 + 2SO_2 + 2CH_3OH \rightarrow SnI_4 + 2HI + 2HCl + 2CH_3SO_4H$$

2 mol $H_2O$/mol substance in the sample is simulated by:

Arsenites $NaAsO_2 + 2I_2 + 2SO_2 + 2CH_3OH \rightarrow NaI + AsI_3 + 2CH_3SO_4H$

3 mol $H_2O$/mol substance in the sample is simulated by:

Arsenates $Na_2HAsO_4 + 3I_2 + 4SO_2 + 4CH_3OH \rightarrow 2NaI + AsI_3 + HI + 4CH_3SO_4H$

Mercaptans simulate a higher water content and are oxidized to disulfides:

$$2RSH + I_2 \rightarrow RSSR + 2HI$$

The water content of salts can usually be determined by the oven method. For interfering hydrazines, the following procedure provided good results:

The sample the hydrazine dissolved in methanol and titrated with methanolic iodine solution. This oxidizes the hydrazine and the KF water determination is no longer interfered with. In the same solution, a titration can then be carried out with a combined KF solution. The $H_2O$ content of the methanol used must be determined separately and subtracted from the result.

### 7.8 Oxidizing Agents

These release elemental iodine from the iodides in the KF reaction and simulate a water content that is too low.

a. Dialkyl peroxides, diacyl peroxides, and peresters form iodine. However, the reaction takes place slowly and can be "frozen" at low temperatures. The titration temperature to be used is product-specific and can be as low as −60°C. Hydroperoxides—for example $H_2O_2$—do form iodine, but also an equivalent amount of $H_2O$. The KF titration is free from interference.
b. Elemental chlorine and bromine also release iodine. In this case, excess $SO_2$ is passed through the sample. This reduces these substances to chloride and bromide respectively, which no longer interfere.
c. The water content of dichromates, chromates, and permanganates is determined by the oven method.

## 8 SAMPLE PREPARATION TECHNIQUES

Each sample preparation is preceded by sampling. Particularly with inhomogeneous samples—and there are a lot of them, even if it is not always obvious—this initial step is important for the correctness and accuracy of the results. Correct sampling is therefore an essential requirement for the subsequent determination!

Many standardized methods provide detailed instructions concerning this subject. This may also apply for in-house instructions. If no instructions are available then the following points must be observed:

- The water content of the sample taken must be the same as the average water content of the original sample—that is, the sample must be representative.
- Atmospheric moisture is an ever-present source of contamination, and not only in tropical countries.
- Hygroscopic samples have a large amount of water adhering to their surfaces.
- Samples that easily release water or become weathered have only a small amount of water on their surfaces.
- Liquids can release water if the water is not present in dissolved form but as a dispersion (precipitation or flotation, adherence to vessel walls—for example, ointments, creams, and lotions). The same applies for samples in which the solubility limit for water is undercut on cooling down.
- The more heterogeneous the water distribution in the sample, the larger the sample to be taken.
- Sampling must take place as quickly as possible and the water content of the sample must not change during transport and storage.

## 8.1 General

Many samples can be weighed out and/or injected directly into the titration vessel or titration cell, provided that they are soluble in the solvent used and do not undergo side reactions with the KFR.

In contrast, other samples require one or more preparation steps. The most important sample preparation techniques are described below. Further information can be found under the individual classes of substances.

It is very important that the sample preparation technique used does not introduce any additional water into the sample and that no water losses occur by heating the sample too strongly.

## 8.2 Dilution

For reasons of accuracy (sample size), samples with a high water content (>40%) are diluted with an inert dry solvent before determination. Methanol can normally be used for this purpose. If the sample contains emulsified fats oils or both then a dry mixture of 60% methanol and 40% decanol (or, if permitted, chloroform) should be used. The procedure is as follows:

Approximately 1 g sample is weighed out exactly into a dry Erlenmeyer flask fitted with a septum stopper. This is treated with approximately 20–25 g dry solvent (mixture), the flask is then sealed, and the contents are mixed. Part of the mixture is drawn off in a dry syringe and the syringe and its contents are tared. Some of the mixture is injected into the titration vessel and titrated. The amount of sample injected is determined by back-weighing the syringe. If a dry solvent is used then it is not absolutely necessary to take its blank value into account. Here is an example:

| | |
|---|---|
| Sample weight | 1.575 g |
| Methanol weight | 22.725 g |
| Injected mixture | 1.500 g |
| KFR titer | 4.855 mg $H_2O$/mL |
| KFR consumption | 6.255 mL |
| *Calculation* | |
| Diluted sample | 22.725 g + 1.575 g = 24.300 g (100%) |
| Sample fraction | 1.575 g: 24.3 g = 0.0648 |
| Injected sample | 0.0648 × 1500 mg = 97.22 mg |
| $H_2O$ found | 6.255 mL × 4.855 mg/mL = 30.37 mg |
| $H_2O$ in sample | 30.37 × 100/97.22 = **31.24%** |

The following samples can be treated in this way:

- Ointments
- Cosmetic lotions and emulsions
- Alcohol-containing samples.

## 8.3 Solvent Extraction

This can take place in two different ways—external or during the titration in the titration vessel. It is usually carried out either warm or hot, less frequently at room temperature. If an external extraction is carried out then please remember that all the water can never be extracted, an average water content between the sample and the solvent is obtained. Example: 10 g sample with a water content of 10% is extracted with 50 mL methanol (density approximately 0.8 g/cm$^3$) with a water content of 100 ppm. The $H_2O$ content of the mixture is 1004 mg/50 g, corresponding to 20.08 mg $H_2O$/g or 2.008% $H_2O$.

This also means that the larger the ratio between the sample and dry solvent, the less water will be extracted from the sample. The blank value of the solvent must, of course, be taken into consideration when calculating the water content.

Better values are obtained if the extraction is carried out directly in the titration vessel and the water is titrated immediately. A titration vessel equipped with a heating jacket and a reflux condenser is used and the titration is carried out, for example, at 50°C. Work can be carried out with or without an extraction time (set at the titrator) as required. If an extraction time is used then please note that a sufficiently long-stop delay must be observed for the endpoint.

## 8.4 Homogenization

In liquids water can float, adhere to the inner walls of the sample bottle, or be deposited on the bottom. This occurs particularly when the sample is taken at a high temperature and allowed to cool down to room temperature before the analysis. The water can again be distributed homogeneously by immersing the bottle in an ultrasonic bath.

The separated water can also be dissolved again by the addition of an exactly weighed-out amount of a solubility promoter (e.g., isopropanol), whose water

***FIGURE 13*** *The 841 Titrando with Polytron PT 1300 D high-frequency homogenizer (Metrohm AG).*

content has been accurately determined (blank value). This method is suitable for pastes and viscous liquids.

Solids are usually inhomogeneous and must be thoroughly comminuted and homogenized. The procedure depends on the constitution of the sample:

Coarse particles, hard and tough: Ball mill or analytical mill (closed) with water cooling. This method is especially suitable for pills.

Coarse, fatty: Grating and high-frequency comminution.

Coarse, soft, inhomogeneous: Comminute with a meat grinder or knife, then high-frequency comminution.

Pasty, inhomogeneous: Homogenization with a blender, then possibly high-frequency comminution.

High-frequency comminution: High-frequency comminution can be directly carried out in the sealed titration vessel. Any alteration to the water content caused by atmospheric humidity is eliminated. With the Titrando/Polytron system shown in Fig. 13 tablets are dissolved in the KF reagent in a short time regardless of their coating.

## 8.5 The Karl Fischer Oven Method

A dry stream of gas is passed through the heated sample; in each case the released water is transferred to the titration vessel and immediately titrated. The heating temperature should be set as high as possible—just high enough so that the sample does not decompose and no interfering secondary constituents distill over. The heating temperature and duration must be optimized in preliminary tests. The appropriate blank values must be subtracted.

This method is used for samples that cannot be extracted (or cannot be completely extracted) with a suitable solvent or that undergo side reactions with the KFR, such as some starting materials for pharmaceutical products.

## 9 KARL FISCHER WATER DETERMINATION IN PHARMACEUTICALS

### 9.1 Water Content in Pharmaceuticals

Quality, hardness, compaction, and shelf life of pharmaceuticals depend to a large extent on the content of adsorbed (surface water) or bound water (water of crystallization). Accordingly, straightforward and accurate methods for water determination are required. Several methods exist for the determination of the water content: loss on drying, Fourier Transform Infrared (FTIR), and Raman Spectroscopy as well as dielectric measurements. Among these, Karl-Fischer titration is certainly the method of choice when trace amounts of water have to be accurately determined in a reasonable time. Hence, for many years now KF titration has been stipulated as the standard method for determining the water content of pharmaceuticals. The most important pharmacopoeias are:

- European Pharmacopoeia (Ph. Eur., (12)) and the
- United States Pharmacopeia (USP, (13)).

Both of these contain detailed information about sample size/sample weight, titration method (>95% direct volumetric titration), range of the water content permitted, and reagents to be used. The European Pharmacopoeia describes more than 400 substances for which the water content has been determined by the Karl Fischer method, the USP describes more than 700 substances (14).

### 9.2 Direct Volumetric Titration

With methanol in the titration cell, the water content of the following substances has been determined by direct volumetric titration:

Pyrazolone, cipofloxacin hydrochloride, effortil, folic acid, nicotinic acid, and scopolamine hydrobromide.

And in methanol at 50°C:

Glucose, hyaluronic acid sodium salt, lysine acetylsalicylate, acetylsalicylic acid, calcium folinate, and lactose monohydrate.

The water content of the following substances has been determined in a solvent mixture, in some cases after heating to 50°C:

Mannitol and sorbitol (methanol/formamide 2:1), erythromycin (pyridine), isomalt, glucosylmannitol/glucosylsorbitol (methanol/formamide 1:1, ground-up sample), magnesium aspartate, magnesium lactate, calcium glycerophosphate, caffeine (methanol/formamide 2:1, 50°C), glycerol monostearate, and suppositories (methanol/chloroform 2:1, titrant with a titer of 2 mg $H_2O$/mL), lanolin (methanol/

chloroform 1:1), serine phosphoric acid (methanol/formamide 2:1, 50°C), and herb extracts (methanol/formamide 2:1).

## 9.3 Karl Fischer Oven Method for Pharmaceuticals

With some products or substances side reactions can occur if the titration is carried out directly in methanol (pH alteration of the KF system, oxidation by iodine). Moreover, many substances release their water only slowly or at high temperatures. They are therefore not suitable for direct Karl Fischer titration. An additional problem is the low solubility of certain samples in alcohols. In these cases, traditional methods recommend laborious sample preparation procedures or the use of toxic solvents to promote dissolution. The European Pharmacopoeia recommends the quantification via the LOD in a drying cabinet or desiccator. However, besides the water to be determined, this method also includes all volatile compounds (e.g., impurities, decomposition products) released at higher temperatures.

The KF Oven Method allows to resolve these problems (see section 8.4). Depending on the water content of the sample, KF titration can be performed either volumetrically for higher or coulometrically for lower water contents. Considering the relatively small amount of water in most pharmaceuticals, samples are mainly analyzed coulometrically.

***9.3.1 Procedure*** The calculation of the sample's water content involves also the determination of the humidity in the sample vessel and the moisture adhering to the vessel surface, vial cap, and septum. Three empty but sealed sample vials are "analyzed" under the same conditions as the particular sample. The resulting mean value is stored as a blank value and automatically taken into account in the calculation (15).

Subsequently a three-fold determination for each compound is performed. The accuracy of the KF titrations is checked at regular intervals with a certified KF Oven standard.

***9.3.2 Water Release Curve for Metamizole Sodium*** The optimum oven temperature for driving off the water should ensure complete water extraction in a reasonable time excluding decomposition of the sample. This means that the oven temperature should be chosen as high as possible to ensure short determination times, but still be 20–30°C below the decomposition temperature. The oven temperatures for unknown samples are determined on the basis of the water-release curves that were recorded for all investigated pharmaceuticals in the temperature range of 50–250°C (16).

The determination of the adequate oven temperature is illustrated by metamizole sodium, a pharmaceutical that is not accessible to direct KF titration. The dashed curve in Fig. 14 represents the absolute water released as a function of total time while the continuous curve corresponds to the associated drift as a function of oven temperature.

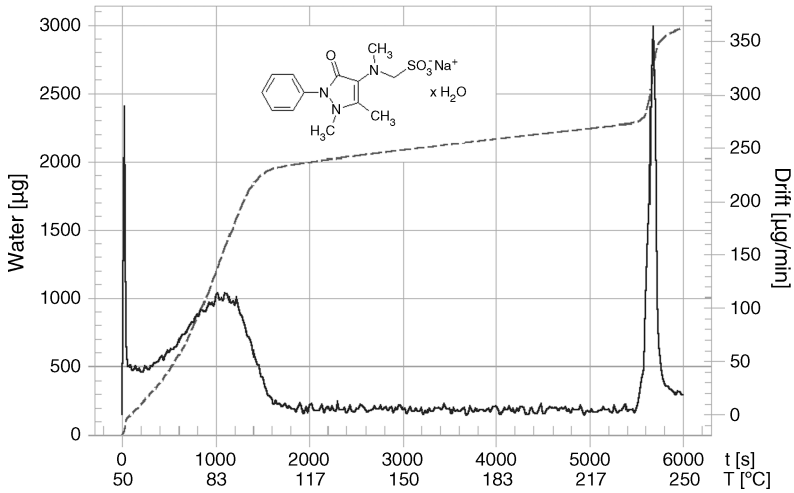


***FIGURE 14*** *Water-release curve for metamizole sodium in the temperature range 50–250°C.*

The water release curve was recorded using a heating rate of 2°C/min. Both surface moisture and water of crystallization are released within the first 1600 s (50–103°C). Afterwards the drift (continuous curve) drops to its original value of approximately 10 µg/min and remains almost constant for 3800 s, increasing again at 5400 s. At 5400 s (230°C), both curves indicate by their increase that water is released by beginning decomposition. Hence, for the rapid determination of the water content of metamizole sodium, a temperature of approximately 210°C assures a fast water release without decomposition.

***9.3.3 Titration Curve for Morphine Hydrochloride*** Figure 15 shows the absolute amount of water (dashed curve) and the drift (continuous curve) as a function of time. The three drift peaks show that morphine hydrochloride is present in the form of its trihydrate. After 180 s no more water is released. The slight increase of the dashed curve can be attributed to the background drift.

***9.3.4 Comparison Karl Fischer Oven Method/Loss on Drying*** In contrast to the wide range of the recovery rates for the LOD specified by the European Pharmacopoeia (65–117%), the KF oven method provides recovery rates between 96 and 101% (Table 1). Additionally, the water content is determined with an excellent repeatability in less than 12 min.

***9.3.5 Advantages of the Karl Fischer Oven Method*** The Oven Sample Processor:

- enables accurate and precise analysis of pharmaceuticals that formerly were not amenable to KF titration,

PSE415 03/26/2013 8:54:30 Page 31



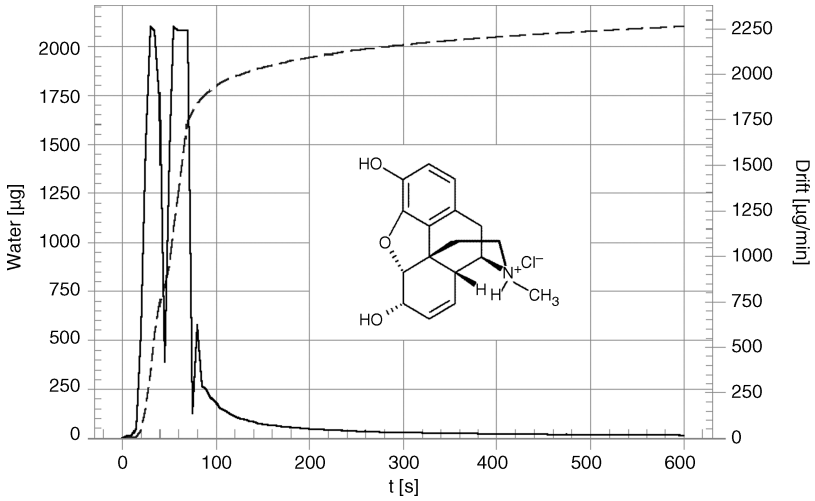


***FIGURE 15*** *Titration curve for morphine hydrochloride at an oven temperature of 180°C. The absolute amount of water (dashed curve) and the drift (continuous curve) are shown as a function of time.*

- avoids contamination of the titration cell; consequently there are no carryover and memory effects,
- works non-destructively and requires only 15–30 mg of the potentially expensive pharmaceutical,
- allows recording of water release curves providing information about the kinetics of the water release,
- significantly reduces manual sample preparation steps,
- enhances sample throughput (high efficiency),
- analyzes all samples under identical and reproducible conditions, providing results of the highest precision.

## 9.4 Alternative Procedures

For some products, the water content cannot be accurately determined neither in methanol nor by the heating method. For these samples the following procedures are suggested (17):

a. Salves and creams are first diluted with methanol/chloroform. Approximately 1.5 g sample is weighed out into a dry Erlenmeyer flask fitted with a septum stopper. Approximately 50 g methanol/chloroform 1:1 is added together with a magnetic stirrer bar, the flask is sealed and vigorously shaken. It is then placed on the magnetic stirrer for 5 min. 25 mL methanol/chloroform 2:1 is placed in the titration cell and titrated to dryness. Sample solution is removed through the septum stopper with a dry syringe and approximately 1 mL is injected into the titration vessel. The sample weight is obtained by differential

**TABLE 1 Results of the Karl Fischer Water Determination for Selected Pharmaceuticals from the European Pharmacopoeia**

| Substance and Empirical Formula | Water Content: Calcul. (%) | Water Content: Loss on Drying[a] (LOD) (%) | Water Content: KF Oven Method (%) | Water Content: KF Oven Method RSD[b] (%) | Recovery: LOD (%) | Recovery: KF | Melting Point (°C) | Temp. 874 Oven (°C) |
|---|---|---|---|---|---|---|---|---|
| Carbidopa $C_{10}H_{14}N_2O_4 \times H_2O$ | 7.37 | 6.9–7.9 | 7.17 | 0.35 | 93.6–107.2 | 97.3 | 203–205 | 150 |
| Cysteinehydrochloride monohydrate $C_3H_8ClNO_2S \times H_2O$ | 10.25 | 8.0–12.0 | 10.33 | 0.30 | 78.0–117.1 | 100.8 | ≥170[c] | 150 |
| Ethacridine lactate monohydrate $C_{18}H_{21}N_3O_4 \times H_2O$ | 4.98 | 4.5–5.5 | 5.05 | 0.59 | 78.0–117.1 | 101.4 | ≥180[c] | 180 |
| Metamizole sodium $C_{13}H_{16}N_3NaO_4S \times H_2O$ | 5.12 | 4.9–5.3 | 5.14 | 0.30 | 95.7–103.5 | 100.4 | 220–210[c] | 150 |
| Morphinehydrochloride $C_{17}H_{20}ClNO_3 \times 3H_2O$ | 14.37 | 12.0–15.0 | 13.83 | 0.44 | 83.5–104.4 | 96.2 | ≥200[c] | 180 |
| Quinine sulfate $C_{40}H_{50}N_4O_8S \times 2H_2O$ | 4.6 | 3.0–5.0 | 4.63 | 0.33 | 65.2–108.7 | 100.7 | 144–155 | 135 |

[a]Given in the European Pharmacopoeia.
[b]Mean value of three-fold determinations.
[c]Melting with decomposition.

Calculated water contents as well as information about the loss on drying given in the Pharmacopoeia.

weighing of the syringe. A blank value for the solvent mixture is determined in the same way.

b. In lozenges the water is determined in the same way as in salves. Three grams comminuted sample is dissolved in 20 g formamide under gentle heating and approximately 1 g is titrated with the separate KF reagents for aldehydes and ketones.
c. Bismuth subnitrate is treated in the same way (5 g sample in approximately 50 g methanol, 2–3 g of the separated out supernatant solution is titrated).
d. Of the highly viscous pentanol approximately 1 g is diluted with approximately 10 g methanol and approximately 2 g of the mixture is titrated with the separate KF reagents for aldehydes and ketones.
e. For amoxicillin, penicillin G (18), dobutamine hydrochloride monohydrate, arginine aspartate, and norepinephrine hydrochloride, 30 mL methanol and 5 g salicylic acid is placed in the titration cell and titrated to dryness; the sample is then added and its water content is determined.
f. For S-carboxymethyl-L-cysteine 20 mL of a 1:1 mixture of methanol/formamide is placed in the titration vessel with 5 mL KF buffer solution and titrated to dryness. Approximately 1 g sample is added and its water content is determined.
g. With adenosine-5′-triphosphoric acid disodium salt imidazole (approximately 3 g) is added to the methanol in the titration cell in order to stabilize the pH of the KF system.
h. Vitamin C (ascorbic acid) and ascorbates are oxidized by the iodine in the KF reagents and an increased water content is simulated. The following procedure—with two-component reagents—is proposed for these substances:

   25 mL methanol is placed in the titration vessel and the sample is dissolved in it. The vitamin C or ascorbate is then titrated with methanolic iodine solution (the titrant consumption does not need to be noted down). 10 mL working medium ($SO_2$ plus base in methanol) is added and the water content is titrated with the same titrant. The blank value of 25 mL methanol + 10 mL working medium (without sample) is determined and taken into account in the calculation.
i. Lindane contains free chlorine and its water content must also be determined with separate KF reagents:

   25 mL working medium ($SO_2$ and base in methanol) is placed in the titration vessel and approximately 2 g sample is added. The solution is stirred for 5 min and the water content is then determined by titration with iodine solution.

   The blank value for 25 mL working medium (without sample) must then be determined separately and taken into account in the calculation.
j. For N-acetylcysteine 20 mL methanol and 5 mL KF buffer solution and 1 g *N*-methylmaleinimide is placed in the titration vessel and titrated to dryness with KF solution. Approximately 1 g sample is added, the solution is stirred for 5 min and the water content is then titrated.

## REFERENCES


1. Fischer K. Neues Verfahren zur massanalytischen Bestimmung des Wassergehaltes von Flüssigkeiten und festen Körpern. *Angew Chem* 1935; 48: 394–396.
2. Scholz E. *Karl Fischer titration: determination of water*. Berlin, Heidelberg, New York: Springer Verlag; 1984. P 138.
3. Jones AG. A review of some developments in the use of the Karl Fischer reagent. *Analyst* 1951; 76: 5–12.
4. Smith DM, Bryant WMD, Mitchell J Jr. Analytical procedures employing Karl Fischer reagent, I. Nature of the reagent. *J Am Chem Soc* 1939; 61: 2407–2412.
5. Verhoef JC, Barendrecht E. Mechanism and reaction rate of the Karl Fischer titration reaction. *J Electroanal Chem* 1977; 75: 705–717.
6. Cedergren A. Reaction rates between water and the Karl Fischer reagent. *Talanta* 1974; 21: 265–271.
7. Cedergen A. Determination of the kinetics of the Karl Fischer reaction based on coulometry and true potentiometry. *Anal Chem* 1996; 68: 784–791.
8. Margolis SA. Source of the difference between the measurement of water in hydrocarbons as determined by the volumetric and coulometric Karl Fischer methods. *Anal Chem* 1999; 71: 1728–1732.
9. Larsson W, Cedergren A. Coulometric Karl Fischer titration of trace water in diaphragma-free cells. *Talanta* 2005; 65: 1349–1354.
10. Cedergen A, Jonsson S. Progress in Karl Fischer coulometry using diaphragm-free cells. *Anal Chem* 2001; 73: 5611–5615.
11. Nordmark U, Cedergren A. Conditions for accurate Karl Fischer coulometry using diaphragm-free cells. *Anal Chem* 2000; 72: 172–179.
12. European Pharmacopoeia 6th Edition — Supplements 6.3–6.5; 2009.
13. United States Pharmacopeia USP 32-NF27; 2010.
14. Metrohm AG. Pharmaceutical potentiometric analysis collection (Pharm PAC, Article Number: 6.6042.001), Herisau; 2003.
15. Metrohm Application Bulletins AB-109, AB-193, AB-255, AB-273, AB-280 and AB-282 (downloadable under http://products.metrohm.com).
16. Schlink R, Dengler S, Kirschenbühler P. Automatic Karl Fischer water determination in pharmaceuticals. *Lab Asia* 2003; 10: 22–24.
17. Bruttel P, Schlink R. *Water determination by Karl Fischer titration, Metrohm Monograph* (Article Number: 8.026.5003), Metrohm AG in Herisau, Switzerland, 2003, P 80.
18. Metrohm Application Notes AN-K-004, AN-K-006, and AN-K-039 (Available at: http://products.metrohm.com).

# Chapter 1


## Abstract

A variety of analytical techniques have been used for water determination. These include drying methods and titration methods. In comparison of drying, titration is rapid and can be validated. This article focuses on water determination by Karl Fischer titration.



## Keywords

drying, karl fischer titration, water, chemical reactions, reaction kinetics.

# CURRENT NEEDS FOR NEW THERAPEUTIC AGENTS AND DISCOVERY STRATEGIES—A SYSTEMS PHARMACOLOGY APPROACH

Sophie C. Desbiens and David H. Farb

*Boston University School of Medicine, Laboratory of Molecular Neurobiology, Department of Pharmacology and Experimental Therapeutics, Boston, Massachusetts*

## 1 INTRODUCTION: A BRIEF HISTORY OF DRUG DISCOVERY

The use of remedies to treat or alleviate symptoms of a medical condition can be traced as far back as ancient Egypt. The Ebers papyrus, dating from 1555 BC, was found to contain 876 concoctions to treat a wide variety of disorders [1]. Early medicinal efforts were also used by the Greeks, most notably Hippocrates, and by several Asian cultures including the Chinese [2]. However, the identification of active ingredients and the development of the interdisciplinary science of pharmacology that bridged organic chemistry, zoology, and pharmacology did not emerge until the late 1800's. These advances were made possible by progress in chemistry, including theories on acids and bases and on the structure of aromatic molecules such as dyes [3]. In the 1870's, Paul Ehrlich proposed the existence of "chemoreceptors", which differed between microorganisms and the host tissue, based on his studies of dyes in biological tissues [3]. He suggested these could be used therapeutically, which eventually gave rise to the development of a class of drug treatments known as chemotherapy.

At the beginning of the twentieth century, pharmacology progressed quickly. In 1905, J. N. Langley introduced the concept of a "receptive substance", the modern basis for the study of receptor agonists and antagonists [4, 5]. In 1933, Meldrum and Roughton identified the enzyme carbonic anhydrase while studying the effects of sulfanilamide, the active metabolite of the antibiotic sulfamidochrysoidine [6]. This discovery led to the concept of enzymes as a good target for drug discovery and gave further importance to the biochemical characterization of cellular functions. Following Alexander Fleming's discovery of penicillin as a product from the *Penicilium* mold that killed *Staphylococcus* bacteria [7], many drug companies invested in microbiology, resulting in the discovery of more antibiotic and other therapeutic agents [3].

Over the next decades, drug discovery progressed along with our understanding of the basic sciences that underlie the discipline of pharmacology. What is currently referred to as interdisciplinary and translational basic science became known as pharmacology and experimental therapeutics. The large explosion of chemical libraries in the 1980's by the large scale implementation of combinatorial chemistry required the simultaneous development of high-throughput screening (HTS) methodologies. With these new technologies, many hundreds of thousands of compounds could be synthesized and then screened against a target of interest. However, this approach was often largely detached from a physiologically relevant screening signal. Another major breakthrough that changed the way society as

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

well as pharmacologists thought about therapeutics coincided with the development of the human genome project in 1986. The hope that a large number of drug targets would soon be identified and, in combination with the HTS technology, that a rapid increase in the discovery of therapeutic agents would result [8] effused through the pharmacology and biomedical sciences community as well as Wall Street aficionados. However, this original optimism met with a declining number of drugs approved by the U.S. Food and Drug Administration (FDA) since 1996, seeming to suggest that a target-based random screening approach was not the panacea originally envisaged for drug discovery.

In this chapter, we will provide a brief overview of some current therapeutic agents, including their weaknesses, and highlight selected opportunities for continued drug discovery efforts in certain disease areas. We will then focus our attention on different strategies used to identify new therapeutic agents, comparing target-based discovery (TBD) and what we refer to as systems-based discovery (SBD). Although pharmaceutical companies have favored TBD during the past thirty years, emerging evidence indicates that SBD may help in solving some of the issues that have plagued the modern approach to drug discovery. Finally, we will explore the challenges associated with the application of SBD to neurological disorders. This chapter is intended neither to be exhaustive or fair and balanced but presents a selective viewpoint. We apologize to those pioneers whose work we may have failed to cite and for simplifications we introduced in the interest of streamlining this chapter.

## 2 CURRENT STATE OF THERAPEUTIC AGENTS AND THE NEED FOR NEW AGENTS

### 2.1 Overview of the Current State of Therapeutic Agents

Therapeutic agents currently on the market can be placed into either one of four categories: small molecules, which are chemically synthesized compounds such as aspirin; natural products, which are small, naturally occurring molecules that are generally isolated from plants, fungi and mold (such as penicillin); biotherapeutics, which are macromolecules that occur naturally (such as insulin and tissue plasminogen activator) or are engineered based upon a biological template (such as trastuzumab (Herceptin, Genentech), a monoclonal antibody designed to treat HER2-positive breats cancer); and nucleic-acid-based therapeutics, which are deoxyribonucleic acid (DNA) or ribonucleic acid (RNA) molecules designed to interfere with the stability or translation of a messenger RNA (mRNA) (such as fomivirsen (Vitravene, ISIS Pharmaceutical and Novartis), an antisense molecule designed to treat cytomegalovirus infections in the retina of immunocompromised patients). The decision to use a small molecule or a biotherapeutic agent is generally dictated by the desired target and, as discussed below, these two classes of therapeutic agents complement each other in terms of their respective capabilities and weaknesses.

***Small Molecules and Natural Products*** For drug discovery, small molecules are typically less than 500 molecular weight (MW) units, although this is not an absolute cut-off as molecules of higher molecular weight could be acceptable if sufficient bioavailability could be achieved. Ultimately, most small molecules are chemically synthesized, but their discovery may be the result of different strategies. Historically, most small molecules were initially isolated from natural products, which is now in disfavored due to the unfavorable logistics of developing commercially viable large-scale chemical syntheses of complex natural products. For example, the compound paclitaxel (Taxol, Bristol-Myers-Squib) was isolated from the bark of a North American yew in the 1960's (Fig 1) [9]. It was later shown to have novel anticancer properties by binding to tubulin, resulting in the stabilization of microtubules and disruption of mitosis [10]. However, the treatment of one cancer patient would have required the harvesting of six yew trees. To address this issue, a synthesis scheme was eventually developed using an analog of paclitaxel,

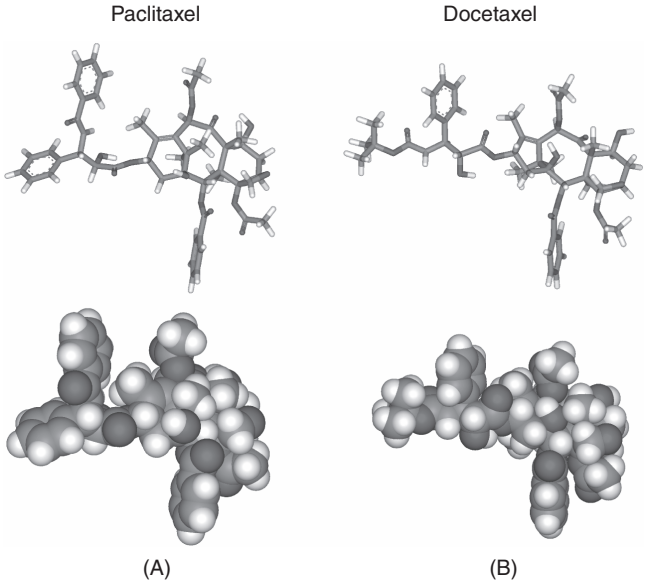


**Figure 1** Chemical structure of (A) paclitaxel and (B) docetaxel.

10-deacetylbaccatin III (DAB), as starting material (reviewed in ref. [11]). DAB was obtained from the needles of European and Himalayan yew trees, avoiding the harvesting of trees. As shown by this example, isolation of a natural substance may require large amounts of starting material to identify and isolate sufficient quantities of the active ingredient. However, the effort and development costs were clearly justified in this case by the discovery of this life-saving anti-cancer agent. A second and important added value is that the advances in science resulting from the study of natural products may also lead to a drug that acts via a new mechanism and thereby initiate a new area of research on the disease biology.

Iterative modification is a strategy in which novel small molecules are synthesized based upon a well-known template until a novel structure, with improved properties and/or efficacy is obtained. Such an approach can lead to the production of a substantiative patent estate. This is the case with docetaxel (Taxotere), an analog of paclitaxel developed by Potier and colleagues at the Centre national de la recherche scientifique (CNRS) [12] and commercialized by Sanofi Aventis (Fig 1). Docetaxel was identified in a structure–activity relationship campaign where paclitaxel derivatives were tested in a microtubule depolarization assay [12]. In this in vitr*o* assay, docetaxel was about twice as potent as paclitaxel. When compared to paclitaxel in different cell line models of tumor, docetaxel was 1.3- to 12-fold more potent and this was found to be due to docetaxel's higher affinity for microtubules [13]. Further studies using in vivo xenograft models revealed that docetaxel was as effective as or more effective than paclitaxel [14]. In addition, docetaxel treated tumors, such as the C38 colon adenocarcinoma, in which paclitaxel was ineffective. Aside from its higher potency and broader efficacy spectrum, docetaxel is also an improvement over paclitaxel in that it is more soluble [12] and is considered schedule independent [14]. This example demonstrates how structure–activity relationship studies based on variants of a known drug can give rise to novel, improved small molecules.

With advances in the technology for characterization of the three-dimensional structure of proteins, the concept that the design of small molecules could be guided by the structure of the target recognition site(s) emerged. Information required to deduce the structure of the active site and other binding pockets can be obtained using X-ray crystallography and nuclear magnetic resonance (NMR) spectroscopy. Candidate molecules may then be designed to interact with specific amino acids at the desired site of action using molecular modeling, bioinformatics, and computational chemistry. In the final steps of the lead identification process, compounds can be modified iteratively to improve upon or optimized the initial lead's chemical properties, efficacy, and/or potency. This process of "rational drug design" resulted in much enthusiasm during the 1980's and early 1990's but seemed to become incorporated with the platform technology of medicinal chemisty rather than the driving force behind new pharmacological discovery.

The Bcr-Abl kinase inhibitor imatinib mesylate (Gleevec, Novartis) is considered to be the first drug designed rationally [15, 16]. Chronic myeloid leukemia is the result of a reciprocal translocation of chromosomes 9 and 22 which creates a gene encoding for the Bcr-Abl kinase, a highly active tyrosine kinase [17, 18]. The design of a specific Bcr-Abl kinase inhibitor takes advantage of the fact that, although several kinases display sequence homology and similar active conformations, the kinases differ in the conformation of their inactive state [19]. Hence, a specific inhibitor could be designed to stabilize the enzyme's inactive conformation, which is different from that of other kinases, and act by preventing phosphorylation of the substrate(s). The crystal structure of Bcr-Abl kinase bound to an analog of imatinib mesylate appeared to reveal an atomic basis for the specificity of interaction [19]. The compound and Bcr-Abl kinase form a number of hydrogen bonds and van der Waals interactions, which are not possible with other kinases. Interestingly, imatinib mesylate interacts specifically with the inactive conformation of the Bcr-Abl kinase as activation causes a conformational change in the kinase that prevents the drug from binding.

During the development of small molecules therapeutics, optimization is geared toward obtaining a desirable set of physico-chemical properties. At the molecular level, these generally include potency and efficacy and, in most cases, specificity for a particular target. At the level of the whole organism, low toxicity (at an efficacious dose) and adequate pharmacodynamic and pharmacokinetic properties must be achieved. For most therapeutic candidates, the ability of the compound to cross the blood–brain barrier (BBB) and blood–nerve fiber barrier (BNFB) must also be taken into consideration. Nervous tissues are very effectively shielded from endogenous and exogenous molecules, whether centrally or peripherally, through the use of physical, biochemical, and cell biological mechanisms (reviewed in ref. [20]). The endothelial cells lining the intracerebral capillary walls are unusual in that they are connected by tight junctions forming patent zonula occludens and lack fenestrations. In addition, these endothelial cells possess metabolic enzymes and transmembrane transporters that further decrease the penetration of most drugs. Small molecules may cross the BBB with both passive or active mechanisms, as well as vesicular

trancellular mechanisms. For passive transport, the size of the molecule, its lipophilicity and charge ($pK_a$) are the main determinants. Drug candidates with specific functional groups may also be actively taken into the brain via transporters. In some cases, however, BBB integrity may be disrupted, facilitating the penetration of small and large molecules. This is seen most often following pathological conditions, such as stroke, trauma, infections, and neuroinflammatory diseases.

Over the course of the small-molecule development process, BBB permeability may be assessed at different stages by using several methods [20–22]. Computational models have been developed based on known drugs that act in the central nervous system (CNS) (called a "training set"). However, these models are limited by the small size of the training set and by the relatively narrow chemical space it covers [21]. In vivo assays can be conducted in both humans and animals using analysis of pure venous blood, microdialysis, and/or positron emission tomography (PET) imaging [20, 22]. However, these methods are generally low-throughput, labor intensive and costly. The use of in vitro assays, based on cell lines transfected with efflux transporters, was found to be a good complement to previously described methods [23]. It is also important to keep in mind that while BBB permeability is critical for drugs acting in the CNS, it is also an advantage to reduce the BBB permeability of drugs targeted outside of the nervous system to decrease potential CNS-mediated side effects. This is the case for loperamide, a therapeutic agent used to treat diarrhea [24]. Loperamide is a peripherally acting μ-opioid receptor agonist targeting receptors located in the large intestine. Because it does not cross the BBB, loperamide is devoid of analgesic properties usually observed with BBB-permeable μ-opioid agonists such as morphine.

Unfortunately, many drugs are marketed without providing doctors and patients information about BBB permeability and possible neurological side effects. An example of this situation is provided by the aromatase inhibitors (AIs), used to reduce the recurrence of estogen-responsive breast cancer, either in conjunction with or following treatment with tamoxifen [a selective estrogen receptor modulator (SERM) acting as an antagonist in breast tissue]. AIs work by inhibiting the aromatase (also called CYP19), the enzyme responsible for producing estrogen. Three AIs are currently the market: the non-steroidal agents anatrozole (Arimidex, AstraZeneca) and letrozole (Femara, Novartis), and the steroidal compound exemestane (Aromasin, Pfizer). Aromatase is present in several tissues including adipose tissue, uterus, bones and brain. While a reduction of estrogen levels in adipose tissue is critical to prevent the resurgence of breast cancer, studies also suggest that estrogen may be important for maintaining cognitive functions [25]. Hence inhibiting estrogen production in the brain may lead to cognitive decline. Strikingly, no information regarding cognitive function following chronic AI treatment is provided to the physicians or patients, even though these drugs were approved over 8 years ago. Studies regarding the possible effects of AIs on cognition in breast cancer patients have only been recently initiated, and, although preliminary results are beginning to come out, the numbers are still too small to determine whether there is an effect (reviewed in ref. [26]). If one or more of the AIs were known to be BBB impermeable, doctors and patients may be able to make a better informed decision regarding the treatment choices and quality of life.

Small molecules may be active at intracellular, extracellular, or both locations of targets and, as such, have traditionally been the platform of choice to act on enzymes and receptors. Moreover, depending on their physico-chemical properties, small-molecule drugs have been designed to be orally bioavailable, which greatly facilitates administration and patient compliance.

***Biotherapeutics*** A biotherapeutic has thus far been considered to be a protein that works by mimicking the action of a naturally occurring substance. In many cases, the therapeutic was no more than the endogenous protein that was produced by a different method, for example, insulin produced synthetically using recombinant DNA technology as compared with purified insulin. Historically, protein therapeutics were isolated from human or animal sources, and their use carried risks such as variable efficacy, contamination by resident infectious agents, and immunological reactions. Most biological therapeutics are now manufactured using recombinant DNA technologies, allowing for a reliable and consistent product [27]. Immunogenic reactions to biotherapeutics are still a concern, but the added level of safety that is derived from a more chemically defined preparation justifies the approval as adverse reactions can be managed in most cases [28].

Recombinant protein therapeutics developed on the heels of the revolution in the area of molecular biological technologies. In 1972, the first article describing the use of restriction enzymes to cut two viral DNAs was published [29], and, by 1978, Genentech, the first company founded to develop biotechnologies, had expressed human recombinant insulin in *Escherichia coli* (*E. coli*) and produced about 20 ng of purified protein [30]. This joint venture with the pharmaceutical giant Lilly resulted in clinical trials beginning in 1980, and in 1982 human insulin manufactured by fermentation of *E. coli* became the first recombinant biotherapeutic to be approved by the FDA.

Over the years, recombinant therapeutic design has evolved and most protein therapeutics that are on the market are not exact duplicates of their endogenous human counterpart. For example, amino acid substitutions are sometimes included to improve pharmacokinetic properties [28]. In other cases, mutations may be induced to change a protein's function. Pegvisomant (Somavert, Pfizer) is a growth hormone receptor antagonist used to treat acromegaly (reviewed in ref. [31]). Substitution of 1 amino acid in the human growth hormone's 191 amino-acid-sequence is enough to turn this agonist into an antagonist. To create pegvisomant, eight other residues were mutated, increasing binding affinity to the receptor, and polyethylene glycol molecules were added to promote stability of the antagonist. Interestingly, even with modifications recombinant protein therapeutics are generally well tolerated and, for the most part, do not trigger immunological complications [28].

The use of antibodies as therapeutic agents has also emerged as a new technology. The success of immunization as a method to protect against infectious agents such as the measles and rabies virus stimulated thought that polyclonal antibodies isolated from human or animal serum could be used to combat toxins and venoms as well as to protect against infections. When this approach was tried, hypersensitive reactions were often triggered because the immune system recognized these polyclonal antibodies as foreign agents that needed to be removed. The advent of monoclonal antibody (mAb) production techniques in the 1970's, along with molecular biology, paved the way for the development and production of chimeric mAbs (murine Fv fragments linked to human IgG Fc fragment), humanized mAbs (human antibody except for the antigen-binding region, which is derived from mouse), and fully human mAbs (derived either from human B cells or from transgenic mice expressing human IgG). Although the immunogenicity of these mAbs is reduced, adverse reactions are still observed in some patients even with fully human antibodies, requiring coadministration of an immunosuppressive agent such as methotrexate (reviewed in ref. [32]). At the molecular level, antibodies may neutralize or antagonize the target, act as agonists, or deplete the target protein from the blood supply. The specific therapeutic action against a target is determined in part by the location of the epitope and by the constant region of the antibody. The ultimate pharmacological success of the therapeutic will require not only biological efficacy but a constant region that minimizes the induction of clearance, complement-dependent cytotoxicity, and antibody-dependent cellular cytotoxicity or apoptosis.

The development of a biotherapeutic for rheumatoid arthritis (RA) treatment provides a good example of the challenges associated with mAb therapeutics [32]. RA is a chronic disease caused by a complex pathology that involves the release of pro-inflammatory cytokines, such as tumor necrosis factor-α (TNF-α), and leads to the destruction of cartilage and bones of non-weight-bearing joints. A major difficulty associated with the use of mAbs is that animal models and ex vivo assays (such as whole-blood assays) offer limited predictability for human efficacy and adverse reactions. For example, at least eleven different antibodies against the T-cell surface antigen CD4 have been developed that are well tolerated and reverse autoimmunity in animal models. However, these same mAbs lacked clinical efficacy when tested in humans. The lack of equivalence between animal models, including non-human primates, and human subjects can sometimes have unexpectedly serious consequences. The activating, anti-CD28 mAb TGN1412 (CD28 SuperMAB, TeGenero Immuno) was shown to be extremely effective at reducing autoimmune reactions without apparent side-effects in both rodent and non-human primate models [33]. However, when TGN1412 was administered to healthy human subjects, all 6 volunteers developed multi-organ failure and extreme cytokine activation. It is unclear whether this reaction could have been otherwise predicted, but it emphasizes the differences between animals and humans when it comes to the pharmacological effects of antibody-based therapeutics.

In contrast to T-cell-directed agents, biological therapeutics targeting TNF-α have been very successful. FDA-approved agents include the anti-TNF-α chimeric mAb infliximab (Remicade, Centocor) and the humanized mAb adalimunab (Humira, Abbott). Both are well tolerated and produce immunogenic reactions in a very small fraction of patients, especially when taken with methotrexate. Another interesting agent is the fusion protein etanercept (Enbrel, Amgen/Wyeth), which is composed of the extracellular domain of the p75 TNF-α receptor attached to the hinge and constant domains 2 and 3 of the human IgG1. This biological therapeutic acts by binding TNF-α, reducing its concentration in blood and preventing its binding to the endogenous TNF-α receptors.

In addition to their traditional roles as immune modulators, antibodies can also be engineered to act as carriers designed to deliver a toxic "warhead" to specific target cells, such as in cancerous tissue. For this application, a specific mAb is conjugated to a toxin, small molecule, or radioisotope that is designed to selectively kill the targeted neoplastic cells [34]. This approach, in principle, allows the the accumulation of a high local concentration of active toxic compounds while minimizing side effects. For example, $^{131}$I tositumomab (Bexxar, GlaxoSmithKline), which binds to CD20, is used for the

treatment of refractory non-Hodgkins lymphoma [35]. The CD20 antigen is present on normal and malignant B cells, including on 90% of B-cell non-Hodgkins lymphomas. Following a single infusion of $^{131}$I tositumomab, 68% of patients saw improvement in their pathological and clinical profiles, including 33% who showed a complete response (Bexxar product information, GlaxoSmithKline).

The large-scale development of biotherapeutics is hindered by the limited options for drug delivery. Due to their large size and susceptibility to degradation in the gut, these agents must be administered via injections or by implantable control release devices. It is also currently impossible to use biotherapeutics for the treatment of disorders of the nervous system due to the fact they lack BBB and BNFB permeability. However this issue may be resolved as pre-clinical studies have been conducted using molecular Trojan horses; that is, fusion proteins between molecules that cross the BBB via receptor-mediated transport and BBB-impermeable agents (reviewed in ref. [36]). In those experiments, peptides such as the vasoactive intestinal peptide, and ribonucleic acid (RNA) interference (RNAi) against the epidermal growth factor (EGF) receptor were successfully delivered to the brain parenchyma. However, further studies will have to be conducted to determine the applicability of this method of biotherapeutic drug delivery to human nervous tissue.

***Nucleic-Acid-Based Therapeutics*** The latest new category of therapeutic agents to emerge, in the 1990's, was based on nucleic acids. The early drug candidates in this group took advantage of antisense sequences as a method to specifically inhibit translation of a given gene product. Although the exact mechanism of antisense inhibition is not fully understood, it is believed that hybridization of a short, single-stranded DNA or RNA sequence complementary to an mRNA targets that mRNA for degradation and/or prevents the ribosome from binding to the mRNA, resulting in decreased levels of protein product. Administration of antisense could be used to treat a variety of conditions such as viral infections, cancers and disorders caused by a mutant protein exhibiting a gain-of-function. However, there are also numerous challenges associated with these therapeutic agents including the short half-life of oligonuclotides in biological fluids, their cellular uptake, and the toxicity they may induce [37]. Chemically modified bases can be used to increase stability of the antisense oligonucleotides in vivo. The most common is the phosphorothioate deoxynucleotide, which increases serum half-life from 1 hour for standard oligonucleotide to 9–10 hours for phosphorothioate-containing oligonucleotides. Cellular uptake can be improved by using delivery systems composed of lipids, polymers or nanoparticles. An alternative strategy is to conjugate the antisense oligonucleotide to an antibody or ligand recognizing a receptor on the surface of the cell, achieving targeted delivery. Toxicity is often the result of an off-target effect(s) and needs to be assessed on a case-by-case basis [39].

The first and only FDA-approved antisense oligonucleotide to date is fomivirsen (Vitravene, ISIS Pharmaceutical and Novartis). This therapeutic agent received approval in 1998 to treat cytomegalovirus-induced retinitis in AIDS (acquired immunodeficiency syndrome) patients (reviewed in ref. [38]). Fomivirsen is composed of phosphorothioate deoxynucleotides and is administered intravitreally. However, although this medication answered a specific medical need, the small number of patients and low sales lead Novartis to discontinue its production. Several other antisense therapies have entered clinical trials, but none have demonstrated sufficient efficacy to obtain FDA approval [37].

RNA interference (RNAi) is a method endogenously used by a wide variety of organisms to target mRNA molecules for degradation [40–42]. Unlike antisense-mediated inhibition, the mechanism by which RNAi works is well understood, and several putative therapeutic agents at different stages of development make use of this technology [43, 44]. RNAi can be achieved using micro-RNA (miRNA), small interfering RNA (siRNA) or short hairpin RNA (shRNA). miRNAs are endogenously encoded single-stranded RNAs that interact with the 3′ untranslated region of mRNAs. Since they do not require perfect sequence homology, miRNAs can interact with 100–200 genes, some of which may be involved in the same signaling pathway [43]. Inhibition of a given mRNA by miRNAs is not complete, but additive effects may be observed if several transcripts in the same pathway are targeted by one miRNA. Endogenous miRNAs, such as miR-146, miR-155 and miR181a, are involved in the development and regulation of the immune system and are altered in chronic inflammatory diseases [45]. In addition, miRNA changes have been reported to correlate with the progression and prognosis of certain cancers [43]. These observations suggest that miRNAs may be an interesting therapeutic target.

The other two types of RNAi tools, siRNAs and shRNAs, are exogenously applied to modify gene expression [46]. siRNAs are short (19–23 nucleotide) double stranded RNAs with a perfect complementarity to their target mRNA. Once they enter the cell, they are taken-up by the RNA-induced silencing complex (RISC) which unwinds the siRNA and degrades the sense strand. The antisense strand is then used as a template to identify the target mRNA, which is then degraded by Argonaute-2. A good siRNA can mediate >90% inhibition of its target. A number of siRNAs are

currently in clinical trials [43]. While naked siRNAs may be appropriate for targeting easily accessible tissues such as the eye or the lung, they are not stable enough to be delivered via systemic circulation. For deeper tissues such as the kidney, chemical modifications and/or delivery systems are used [43]. The most advanced siRNA, currently in phase III clinical trials, is bevasiranib (Opko) against vascular endothelial growth factor. It is a naked siRNA administered by intravitreal injection for wet age-related macular degeneration. AKIi-5 (Quark/Silence) is a chemically modified siRNA that targets p53 and that is delivered intravenously. It recently began phase I trials for acute renal failure.

As their name indicates, shRNAs are short, single stranded RNA sequences that form a hairpin. In the cytoplasm, this hairpin is cleaved by Dicer, leading to an siRNA that can interact with the RISC as described above. The advantage of shRNAs is that they can be encoded on a viral vector, which can be engineered for tissue-specific and/or controlled delivery (reviewed in ref. [47]). A variety of vectors have been used successfully in animal models, including adenoviruses, adeno-associated viruses (AAVs), and lentiviruses. Two shRNAs recently entered clinical trials, targeting hepatitis B infections (Nucleonics/Novosom) and AIDS-related lymphoma (Benitec).

RNAi methods are not without risks and challenges. The most common issue observed in animal models has been the risk of saturation of the endogenous RNAi machinery. For example, in mice receiving large doses of one of 49 different AAVs encoding for shRNAs against 6 different genes, 36 constructs led to liver toxicity, including 23 which resulted in death [40]. This was shown to be due to saturation of Exportin-5, a transporter for precursors of siRNAs and miRNAs. This serious problem can be avoided by using naked siRNAs (mature siRNA enter the RNAi machinery downstream from Exportin-5) or by using a lower titer of virus. Viruses engineered to offer controlled expression of their transcript may also be useful, representing a significant opportunity for the development of new therapeutics.

Using RNAi may also cause off-target effects and/or trigger an interferon-mediated response. Both of these can be reduced by chemically modifying the siRNA and by avoiding specific sequences, which are known to be immunogenic. Finally, resistance to a particular RNAi may develop as a single base-pair mutation is enough to disrupt the interaction between an siRNA and its target. This may be avoided by engineering a vector with a few different shRNAs against the same target.

***Portrait of Drug Approval in Recent Years*** Most new drugs authorized by the United-States Food and Drug Administration (FDA) are classified as small molecules, whereas biotherapeutics account for 5–29% of the yearly approvals (Fig 2). However, despite technical advances and increased spending in research and development by pharmaceutical companies [48], the number of drugs approved has been declining since 1996. This effect is particularly noticeable for small molecule

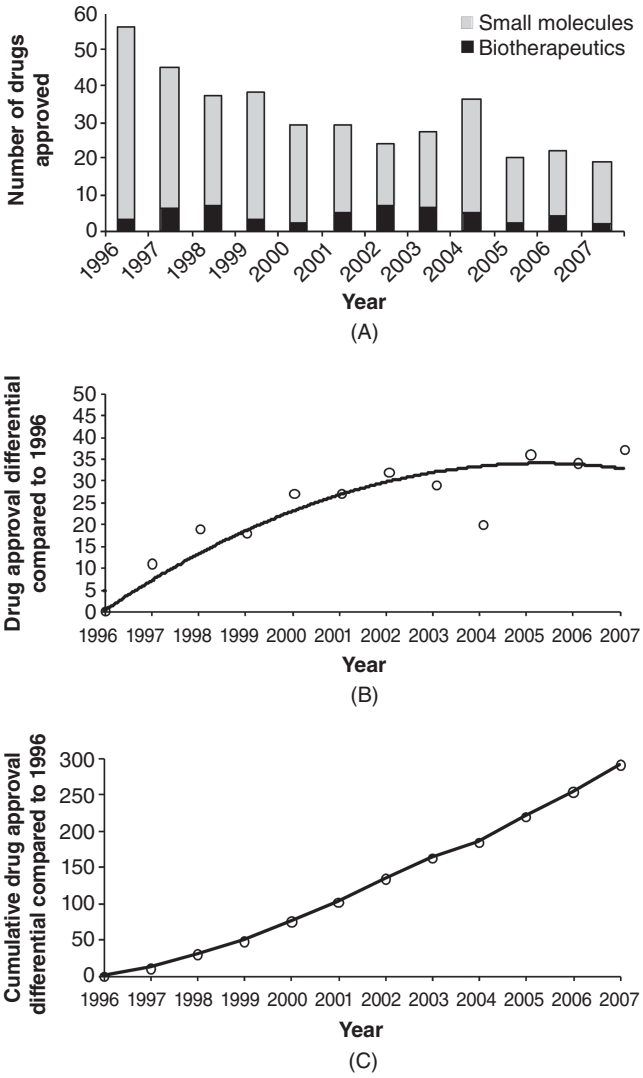


**Figure 2** Drug approval by the Food and Drug Administration (FDA) since 1996. (A) Number of new drugs approved per year by category. Small molecules (gray bars) require the submission of a new drug application (NDA), while biotherapeutics (black bars) are reviewed under the form of a biological licence application (BLA) (Data source: www.fda.org.) (B) The drug approval differential, as compared to the number of drugs approved in 1996, reveals that there has been a progressive decline in approvals in the past ten years. In the last few years however, rate of decline seems to be leveling off. (C) Despite stabilization in the number of new drug approvals, the cumulative drug approval differential suggest that if approvals had been maintained to the level of 1996, we would have nearly 300 additional new drugs on the market today.

therapeutics. There are many reasons for this trend, including increased requirements for safety, reduced side-effects, and a litiginous system of law [39, 49]. Given the current risks and elevated cost of developing a new drug, pharmaceutical and biotech companies often turn to repurposing, the identification of new indications for existing drugs, or to reformulation, such as developing continuous release or combination therapies, to increase their market share [49, 50].

If the outlook for new drug discovery at large is declining, the glacier trail leading to breakthrough neurotherapeutic agents is even more daunting and treacherous. According to the World Health Organization [51], 40 million people worldwide are affected by epilepsy, 24 million suffer from Alzheimer's disease or other forms of dementia, 62 million are diagnosed with cerebrovascular diseases, and 326 million suffer from migraine. However, despite the high prevalence of these disorders, only a few drugs with neurological indications are approved each year (Fig. 3). Clearly, therapeutic agents targeting neurological conditions must cross the BBB, but it is estimated that 98% of putative neurotherapeutics fail due to lack of BBB permeability [52]. In addition, candidates often fail during clinical trials due to poor efficacy or safety concerns. As expected, CNS-directed agents tend to cause CNS-mediated side effects such as seizures, dizziness and nausea. Moreover, the lack of validated biomarkers makes it difficult to assess whether the drug reaches concentrations that are sufficiently high at the target to be efficacious [53]. In 2006 alone, 11 drug programs targeting 7 neurological disorders were halted during clinical testing [54]. In the current situation, there is a dire need for new therapeutics directed towards even the most common of neurological conditions. It is sobering to note that with all of its undesirable side effects and adverse reactions morphine, and its derivative opiate receptor agonists, still provide the mainstay of analgesics.

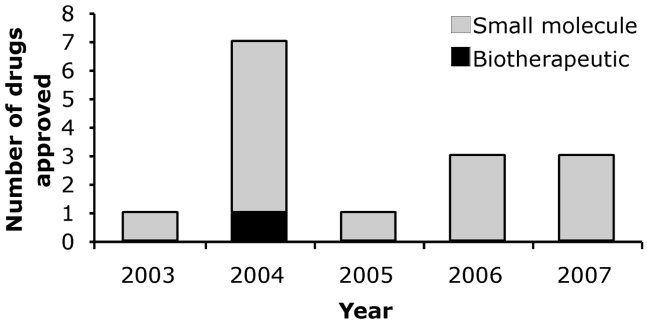


**Figure 3** Number of new drugs approved by the FDA from 2003 to 2007 with a primary indication for a neurological disorder or neuropathic pain. Small molecules are indicated by gray bars and biotherapeutics by by black bars. The only biotherapeutic recently approved is natalizumab (Tysabri, Biogen/Elan), an anti-α4 integrin monoclonal antibody used to treat relapsing multiple sclerosis (Data source: www.fda.org.)

## 2.2 Need for New Therapeutic Agents

Over the last century, there have been major advances in our understanding of pharmacology, biology of disease, and in methods for the discovery of therapeutic agents toward a wide range of maladies. Indeed, the discovery of the "wonder drugs", the tricyclic antidepressants, phenothiazine antipsychotics, and benzodiazepine anxiolytics, completely changed the way society thought about mentally ill: from the notion of insanity and incarceration to the concept of neurological disorder and outpatient treatment. However, there is a great need for continued research and development to improve existing therapeutics, develop new classes of pharmacotherapies for diseases where the current treatments may not work in some or even most patients, and to establish new treatments for conditions where none are available. Given the decrease in new drug approvals in the last several years, it is important to emphasize that there is a crucial need for more new therapeutics to be developed and submitted for review. In order to improve its productivity, not only must the pharmaceutical industry rethink its drug discovery strategies but the federal government and international world health agencies must work to provide an optimal setting for the high risk high stakes business of drug discovery and development to thrive.

***Improving on Currently Existing Therapeutics*** It is easy to imagine the properties of the ideal drug: orally bioavailable, once a day treatment, with no side effects, and completely effective in all patients. Improving on existing therapeutics is an important and valuable method to generate new therapeutics. This strategy mitigates the risks associated with first-in-class therapeutic agents as the chemical genres and target are often well-validated, and certain aspects associated with modulation of the target in humans have been explored. Among the improvements that are generally feasible are enhancements in efficacy, and/or potency, and/or specificity, all aimed at enhancing desired as compared with undesired pharmacological effects via altered pharmacodynamic and/or pharmacokinetic properties. Incremental improvements on the method or frequency of delivery can also provide significant benefits to patients and thus justify the allocution of limited resources to such objectives.

However, this is not a safe haven for the pharmacologist as any modification can cause deleterious side effects even if the discovery effort focused on the creation of a new therapeutic agent acting on the same

target as the approved drug. Novelty is also a key aspect of new drug design, whether small molecule or biotherapeutic, as patent protection is an essential element to commercialization. While this driver is often bemoaned, it serves a useful purpose in rewarding pharmaceutical companies that bring breakthrough science to the benefit of medicine via the market place.

***Finding New Targets for Conditions Where Therapeutic Agents Already Exist*** An exciting and challenging direction is to discover a new therapeutic agent that acts at a novel target that improves the treatment relative to a medication that already exists. The target can be selected from within a pathway that has already been exploited or in a different pathway that is also important for the disease biology. In addition to resolving intellectual property issues identified with the previous strategy, this approach offers a number of advantages. For example, a drug acting at a different target may be effective in non-responders to existing therapeutics due to a number of factors, such as tolerance, heterogeneity of disease pathology, or variations in genetics among populations. A therapeutic agent with a new mechanism of action is also likely to have a different side effect profile, which may be better tolerated. However, working on a new target brings risks that are usually not a factor when working with a known target. For example, choosing a target which is involved in multiple biological pathways or whose biological role(s) is(are) poorly understood may lead to unanticipated adverse events, such as previously described for the super-agonist anti-CD28 mAb TGN1412 [33], or the breakthrough discovery of the selective cyclooxygenase-2 (COX-2) enzyme inhibitor rofecoxib (Vioxx, Merck). It is, therefore, sobering to realize that the particular target must be chosen carefully and thoroughly validated to ensure that it is both implicated in the therapeutic pathway and unlikely to produce serious adverse effects when its activity is modulated. In this regard nature is unforgiving and while patent law rewards novelty, the best science may not always be rewarded. As an example, while Merck scientists discovered a selective COX-2 inhibitor rofecoxib, a number of adverse events that would result from selective inhibition of this enzyme were not anticipated. Inhibition of the COX-1 enzyme was deemed responsible for the gastro-intestinal bleeds observed following chronic treatment with non-steroidal anti-inflammatory drugs (NSAIDs). Pfizer developed the less selective and therefore seemingly less innovative drug celecoxib (Celebrex), which inhibits mainly COX-2 at a therapeutic dose with minimal effects on COX-1 [55]. Although mechanistically similar, rofecoxib was voluntarily withdrawn yet celecoxib has remained a specialty pharmaceutical with a black box warning. This teaches us that the objective of developing a highly potent selective inhibitor of one of two enzymes in a pathway firmly implicated in a physiological process is sound but the side effects that result from involvement of a particular target in other patho-physiological processes can surface as well.

Combination therapy using two or more classes of therapeutic agents to treat a disorder can also provide substantial benefits with reduced risks. This approach has helped with the management of human immunodeficiency virus 1 (HIV-1) infections, for which there are currently six different classes of drugs available. According to the U.S. Department of Health and Human Services, initial therapy with two or three antiviral agents is most effective at reducing HIV-1 load since infected patients often present with a small proportion of viruses harboring drug-resistant mutations [56]. Upon resurgence of the viral load, a new drug combination can be used to suppress infection by a resistant virus strain.

In addition to delaying the appearance of drug-resistant viruses or cancer cells, combination therapy may also be useful to increase the efficacy of a treatment by co-administering two drugs which are not completely effective when given alone. This strategy is often use for the treatment of hypertension, where single agents may not sufficiently lower the patient's blood pressure [57]. Another possibility is to combine two therapeutics at a sub-optimal dose to reduce side effects and achieve increased efficacy. One example is the combined use of acetaminophen with an opioid receptor agonist such as oxycodone. Acetaminophen reduces the dose of the narcotic analgesic required and thus reduces the adverse side effects, limitations, and liabilities for both. However, de novo side effects or undesirable events may arise. These may result from interactions of the different drugs with liver metabolizing enzymes, competition for transport mechanisms, or biological interactions between the different targets and pathways modulated. For example, in prescription drug narcotic addiction, addicts learn that pulverizing controlled-release oxycontin tablets provides a rush by circumventing the controlled-release formulation. Consumption of large amounts of acetaminophen/oxycodone can also result in chronic overdose of acetaminophen, possibly resulting in severe kidney damage.

***Finding New Therapies for Untreated Conditions*** Perhaps the greatest challenge, and the greatest need, in drug discovery resides in finding therapeutic agents against disorders for which no treatment exists. Factors influencing the decision to launch a new drug discovery program include unmet medical need, market size, the ability to diversify the company's intellectual property

portfolio, the availability of required technical expertise, the prospects for upcoming competitors, and the anticipated level of difficulty with the new drug approval process [27]. Although pharmaceutical companies generally focus on a disease that affects a significant portion of the population, there has been a trend in the United States for smaller biotech companies to work on orphan drugs, designed to treat rare disorders affecting less than 200,000 Americans [49]. In addition, for therapeutic agents that treat serious medical conditions where there is an unmet medical need, a priority application can be filed with the FDA, which accelerates the approval process and reduces the total time to market. In recent years, the proportion of priority applications approved by the FDA has risen in comparison with standard approvals (see www.fda.gov), suggesting that this avenue may be advantageous for both medical and commercial reasons.

The identification and validation of a target remains a major challenge with developing a drug for an untreated condition. This is especially true if the disease biology is poorly understood. While the recent emphasis by the pharmaceutical industry has been on identifying a single target and creating a highly potent and specific drug, examples taken from older "dirty" drugs and combination therapies show that drugs acting on multiple subtypes of a single target group or on multiple targets can be highly beneficial. One example is the non-selective benzodiazepine positive modulator diazepam which acts on at least twelve different γ-aminobutyric acid ($GABA_A$) receptor subtypes containing the $\gamma_2$ subunit. Another example is the double reuptake inhibitor venlafaxine (Effexor, Wyeth), which inhibits both norepinephrine and serotonine transporters.

## 3 APPROACHES FOR DRUG DISCOVERY

Historically, drug discovery began with the empirical observation of a therapeutic effect, often by random testing in animals, which was then confirmed and further refined using bioassays and preclinical models (Fig 4).

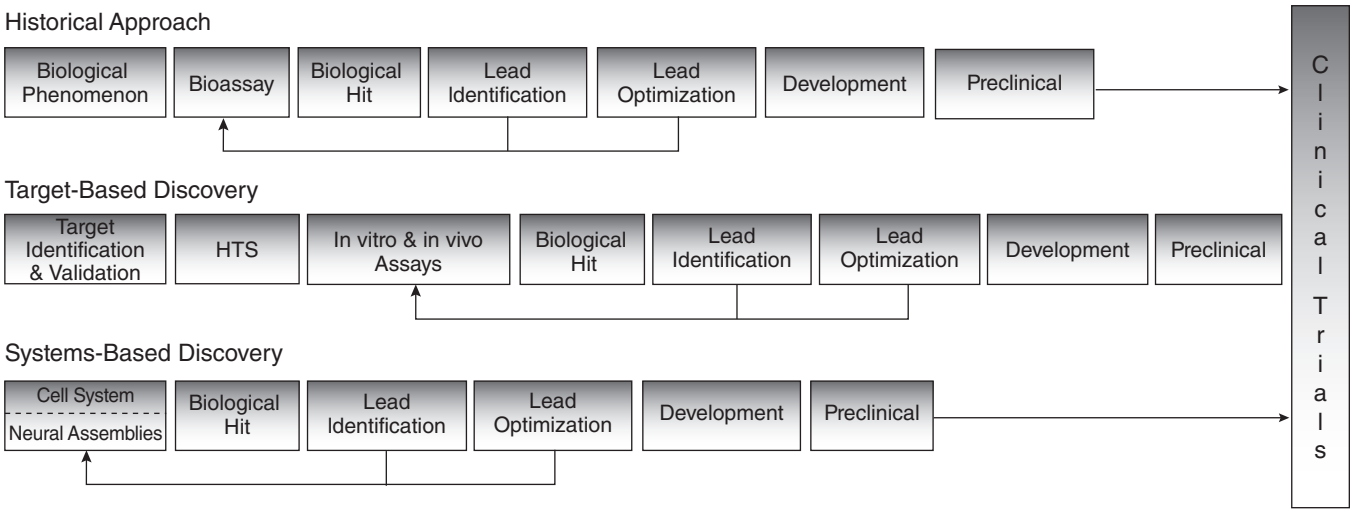


**Figure 4** Drug discovery pipeline. The discovery of new a therapeutic can be initiated in a number of ways as illustrated here by the historical approach, target-based discovery (TBD) and systems-based discovery (SBD). Although the first step (first box for each method) differs, the lasts steps of the process leading to optimization and development of the compound, pre-clinical, and clinical trials are fairly similar. In the case of SBD, the nature of the initial screening system dependent on the disease studied. For example, while cancer and inflammation can be modeled accurately using a collection of human cells, neurological disorders may require a more complex system such as an *in vivo* recording model in order to obtain data from neural assemblies. Neural assemblies are units of neurons whose activity, when taken together, contributes to the observed behavior. Interestingly, although the approach proposed for systems-based drug discovery seems to be similar to the historical approach, our improved understanding of the disease biology today facilitates the selection of an appropriate model for screening compounds, eliminating the need for a trial-and-error approach and increasing our chances for success. It is assumed tha for SBD, the input of compounds would not necessarily involve HTS but could begin with high value focused chemical libraries.

Today, in the postgenomic era, there has been a change in paradigm. Drug discovery is now usually initiated by the identification and validation of a target, even before any putative therapeutic is tested. While this was an effective strategy in the late 1980's and early 1990's, the trend towards decreasing new drug approvals during the past decade raises questions about the long-term effectiveness of this approach. In addition, to overcome challenges associated with the increased scrutiny from regulatory authorities and the ever-rising cost of development for new therapeutic agents, the industry may consider revising its functional model for drug discovery, by transitioning from a target-based approach to a systems-based or combined approach.

In the sections below, we review current drug discovery strategies and suggest a number of opportunities for advances based upon existing needs for improvement.

### 3.1 Target-Based Discovery (TBD)

Target-based discovery is the most widely used strategy by pharmaceutical and biotech companies. Although specific details may vary, the general framework of TBD is usually similar across companies (Fig 4). The first step is to identify a target that is important to the onset or progression of disease pathology [58]. As part of this process, the "druggability" of a putative target, that is, whether and how such a target could be modulated by a therapeutic agent, is of key strategic importance [59]. In addition, financial considerations, public perception of a disease, and governmental health policies may also affect the target selection process [60]. Previously validated targets are an obvious choice, as discussed above, but information regarding novel targets may be obtained from various sources. In some cases, human genetic associations identify a chromosomal area or specific gene that is linked to the disease and further associated based upon a valid animal model [58]. If no suitable target can be identified from in vivo studies, in vitro methods including activity assays and gene or protein expression studies may provide useful information. Alternatively, in silico methods such as bioinformatics (genomics, proteomics, transcriptomics) and pathway analysis may allow to rapidly identify potential targets, but these will require additional validation [61]. It is important to note that results from all the above-mentioned methods may also be found in the literature. Although targets identified from the literature are not technically novel, because the knowledge is publicly available, they can present a lower risk profile if they emerge from high-quality data.

The next step in TBD is to validate the targets identified. This is a crucial step in that it provides the foundation for the entire drug discovery program. Failure to properly validate a target may cause the program to fail at later stages when much time and financial ressources have already been invested. To establish a validated target, several criteria must be fulfilled [58]. When comparing patients affected by the disease and healthy controls, a good target may have a different expression level or activity level. However, interpretation of data such as a change in protein levels or activity may reflect either the cause of the disease or the consequence of the disease. Changes could also result as an adaptive self curative process. In the later case, inhibition of a molecule involved in a negative feedback pathway may worsen the disease. As part of the validation process, the expression pattern of the target molecule must be assessed and localized in a tissue that is relevant to the pathology. Clearly, the ultimate step in target validation is to demonstrate that modulating the target in human patients improves their condition. This is a demanding task but of great importance. New technologies and/or strategic approaches that advance the validation of the target function in human disease will be on the cutting edge for years to come.

As its name suggests, TBD is based on finding drug candidates that act specifically at the chosen target. In the pre-exploratory phase of drug discovery, the assays that are going to be used to test drug candidates are developed, and the specific assays chosen depend upon the type of target. For example, enzyme targets can be tested conveniently in isolated protein assays [62]. However, to find a drug modulating a G-protein-coupled receptor (GPCR), a functional screen in a cell-based assay may be more appropriate. This approach could employ a focused chemical library of a small number of compounds and lower throughput high value screens, such as high throughput electrophysiology [63], or high throughput screening (HTS) of large scale compound libraries.

For the HTS approach, the first step precludes washes and requires an output that can be measured easily [64]. Secondary assays of increasing order of complexity are conducted following the HTS screening. These assays utilize automated benchtop technologies, can be more detailed, but should still have a throughput that is high enough to screen thousands of hits from the HTS screen. The secondary assay phase can employ a cell-based assay, if not already carried-out in the primary assay, and includes selectivity screens. Assessing selectivity early-on is critical as off-target effects may induce side effects or adverse events in pre-clinical studies.

Following in vitro assays, short-term in vivo assays are performed. These assays can usually be carried out on the time-scale of hours to 1 or 2 days and require less compound. Although they may not be a good predictor of efficacy in the disease model, they allow

a determination as to whether the compounds can have some effect when introduced into an animal. Pharmacokinetic studies establish the dosage and route of administration to be employed in the long-term in vivo studies using animal models of disease. At the end of the pre-exploratory phase, an established screening tree beginning with target-based screening and ending with selective pre-clinical animal models of disease is fully implemented.

Once the assays to test the activity of compounds at the target have been established, the TBD exploratory phase can be initiated. The objective of this phase is to identify hits from a library of compounds, and then from those hits to generate one or a few lead families of compounds. Following HTS, thousands of potential hits may be identified [64]. If structural information is available on the target or on a known ligand of the target, pre-screening of the library using structure-based computer modeling may help decrease the total number of hits and increase hit quality [61, 64]. Compounds identified by HTS then have to be re-screened and/or counter-screened to confirm their activity and establish a basic dose–response curve. After confirmation, hits are subjected to the secondary assays devised during the pre-exploratory phase. Following each assay, compounds are ranked and structure–activity relationships (SAR) and structure–property relationships (SPR) emerge. These relationships help to better understand how the compounds work and how they can be improved using chemical modifications [65]. A few hits are generally selected to undergo iterative rounds of chemical modification and testing in secondary assays. At this point, one or a few lead series are typically chosen to move on to the next step.

For each lead series in the discovery phase, the SAR and SPR would be further characterized, pharmacokinetic/pharmacodynamics (PK/PD) studies performed, and candidates tested in in vivo models. It is interesting to note that for TBD this is the first phase where chemical entities are actually introduced in animals. As a result, compound production must be scaled-up to achieve a medium scale [66]. PK studies are important to establish the time-course for drug absorption, distribution, metabolism and excretion [67, 68]. As part of the metabolism studies, possible interactions between the drug and the hepatic cytochrome P450 (CYP) family of enzymes are tested. Drug-mediated inhibition of a CYP enzyme that is responsible for the metabolism of the compound studied or of other commonly used therapeutic agents may lead to severe side effects or to drug–drug interactions [68]. Given the possible severity of the consequences of CYP inhibition, such occurrence would be carefully evaluated. PD studies, on the other hand, relate the pharmacological and toxic effects observed upon administration of a drug to the concentration of its active metabolite in the plasma (or other tissue of interest).

Throughout this process, chemical modifications are still being performed to improve the compounds' properties and to address any potential issues that could result in failure to identify a lead compound. The information obtained from the PK/PD studies is critical for the subsequent in vivo studies as the dosage and dosing frequency selected need to provide an adequate drug concentration for the appropriate duration to obtain a therapeutic effect with minimal toxicity [69]. Animal studies begin with the short-term models devised during the pre-exploratory phase, which have a faster readout and require less compound. Then, the more promising candidates are tested in the long-term disease models. While in vivo testing is present throughout the discovery phase, determination of the efficacy of a compound in a disease model is not done until the later part of this stage and, although chemical optimization is ongoing, there is a risk that the lead(s) obtained may fail and be unsalvageable.

When a lead compound or compounds is/are identified, further testing in non-human primate models including PK, PD, proof-of-concept, and efficacy in a disease model are carried out. This last testing phase allows to obtain information which may be more indicative of the drug's effect in humans, although some unpredicted events may still occur later on in the clinical trials. If a good TBD lead candidate is obtained, drug development can begin.

As described, TBD is a lengthy process with a general progression from the simplest, most reductionist models towards the more complex systems. In parallel, iterative rounds of chemical modifications are performed to establish SAR and SPR in an attempt to improve the compound properties and resolve problems that may arise. While TBD has been successful in identifying new drugs against previously validated targets (i.e. modulated by a known drug), generation of therapeutic agents acting at novel targets has been disappointing [70]. This can be explained in part by the fact that although TBD leads to compounds whose mechanism of action and contact sites with the target are reasonably well-understood, it is also prone to late-stage failures as testing in more relevant in vivo models is not done until the end of the discovery process. This creates an unfortunate situation where, by the time an advanced compound is rejected, several years of discovery and huge opportunity costs have already been invested in the project without any return.

### 3.2 Systems-Based Discovery (SBD)

As the limitations of TBD described above become evident, an alternative view of drug discovery is emerg-

ing. Systems-based discovery (SBD) methods are based on the use of complex systems that model the disease to look for potential therapeutic compounds, instead of focusing on simpler assays that test drug interactions with a single target (Fig 4). While SBD might in some way seem to be a return to the traditional approach to drug discovery, recent technologies allow for the automation of high content assays, with a concomitant increase in throughput. However, advances in computational modeling facilitate the identification of pathways and mechanisms important to the disease biology [70, 71].

A range of systems could be suitable to use as the starting point for SBD. Human cell systems offer many advantages and provide a practical approach to SBD [70]. The complexity of the pathways involved in a disease may be modeled better using a combination of human cell types instead of a cell line. Rather simple modifications can also be made to a human cell system to reflect different physiological or pathological states, further increasing the chances of uncovering therapeutic properties for the drugs tested. For example, a model system for studying inflammatory diseases could be composed of endothelial cells and blood mononuclear cells, stimulated using a combination of cytokines [70]. A disadvantage of such a system, however, is that it may be difficult to obtain the correct combination of primary human cells that will meaningfully reflect the biology of the disease. Some conditions, such as psychiatric disorders, may not be amenable to cellular-level modeling while others, such as these regulated by organism-wide hormonal signaling, may only be modeled crudely by a human cell system. However, new stem cell technologies may remedy this deficiency by providing an endless supply of undifferentiated tumor cells and tissue components.

Another advantage of SBD compared to TBD is that drug screening is performed using a marker of disease progression/regression so there is no procedural requirement to identify and validate a target which, in principle, could reduce discovery costs by allowing only higher value molecules to be identified as leads. Moreover, although target validation in animal models of disease is widely accepted, a given target overexpresed in a cellular or non-cellular system may select for molecules that work in an animal model but are not as relevant to the disease pathology in humans or target modulation may cause unexpected adverse effects. SBD screens compounds against an unknown number of targets but would theoretically generate relatively advanced hits.

Following system-based screening, further testing is conducted and the promising compounds are optimized. As previously described for TBD, these assays include PK and PD characterization as well as short- and long-term animal models. At this stage, a lead compound may be obtained. If desired, target deconvolution strategies can then be applied to identify the target or targets that is/are modulated by the lead compound (specific examples of deconvolution techniques are reviewed in ref. [72]). However, these methods depend upon a physical interaction between the lead compound and one or more targets, with varying affinities and efficacies. Target deconvolution can be straightforward to intractable and should be considered optional as clinical trials can be initiated if safety and efficacy have been demonstrated in cell-based and animal models of the disease [70].

A more complex alternative to the use of human cell systems is to generate system response profiles [73]. These can be obtained by subjecting samples such as body fluids or biopsy samples to various analyses (gene array, mass spectrometry, NMR spectrometry, immunoassays, etc) to obtain levels of gene transcripts, proteins, or metabolites. System response profiles can then be compared across healthy and sick individuals or before and after administration of a therapeutic compound, for example. By collecting samples at different time points, an analysis of the system's dynamics can also be performed. This extracts information about disease progression and/or provides a better understanding of the transient effects of a therapeutic agent. System response profiles may help identify sub-groups of patients with a given diagnosis that are expressing different biochemical markers.

For a systems pharmacology assessment, such a classification may also provide biomarkers for drug discovery and help predict which patients are likely to respond to a given treatment. This application alone could help increase the success of clinical trials, especially for neurological disorders where the lack of validated biomarkers has been cited as a cause for the high rate of failures, due to possible improper patient classification and difficulties in assessing the suitable amount of drug to use [53]. In a different application, system response profiles may be useful in assessing the validity of an animal or cell-based model of the disease [74]. By comparing the system response profiles of the disease model to that of human patients, it may be possible to determine which pathways hold predictive validity and are more likely yield a successful therapeutic agent. Furthermore, different models may be better predictors for different subgroups of the patient population.

Golub and colleagues developed a similar system, called the Connectivity Map, using data collected from treating human cells with a variety of biologically active compounds [75]. This publicly accessible database can be searched for information regarding the activation of certain genes following the treatment of cultured human cells with drugs. By comparing different profiles a new biological function for a known compound can

be identified. Gene expression-based high-throughput screening (GE-HTS) has been used to screen for inhibitors of platelet-derived growth factor receptor (PDGFR) signaling [76]. PDGFR signaling leads to activation of the extracellular regulated kinase (ERK) pathway, which is known to be up-regulated in a number of cancers. In human TIP5 fibroblast cells treated with PGDF with or without an ERK inhibitor and subjected to microarray, a number of regulated genes correlated with ERK activation were selected as markers for ERK inhibitor activity. Application of GE-HTS to a chemical library led to the identification of aurintricarboxylic acid, which inhibits PGDFR phosphorylation, preventing downstream activation of ERK. GE-HTS does not require a detailed knowledge of the signaling transduction cascade and captures molecules that act either upstream or downstream of the known kinase.

When applied during clinical trials, system response profiles may also generate useful information and potentially rescue drug candidates that would otherwise fail [73]. For example, the system response profiles of drugs with limited efficacy could be used to predict useful therapeutic combinations, either with a previously marketed agent or with another low-efficacy drug. System response profiles may also help to better understand mechanisms by which drugs work, including potential off-target effects. These unknown mechanisms can either enhance therapeutic effects or cause side effects. Knowing which disease pathways are targeted by a given treatment may also expose unexploited areas of the disease profile, including potential new drug targets.

Despite these limitations and given their potential benefits, system response profiles could fulfill the need for the discovery of new therapeutic agents by delivering higher value lead candidates earlier in the discovery process.

### 3.3 Balancing TBD and SBD

TBD and SBD offer very different strategies, as illustrated by the dissimilarities in their workflow pipelines (Fig 4). In TBD, the first discovery milestone identifies and validates a drug target. By contrast, SBD begins with the development of a systems-level assay representative of the disease and then proceeds with testing a library of compounds to identify hits. With TBD, assay development and hit generation occurs later in the process, following target identification and validation. SBD generates hits at an earlier stage and, at least theoretically, selects compounds that have better druglike properties. Whereas TBD uses simple and usually cell-free assays for HTS, SBD is generally carried-out using an intact cell assay on a system ideally composed of human cells. The quality of the hits obtained plays a major role in determining how much effort will have to be devoted to lead identification/optimization. Another major difference between the two strategies resides in the fact that in TBD, compounds are tested in cell-based systems during the discovery stage, which occurs much later in the process as compared with SBD, which begins with a cell-based system or tissue. In concept, compounds identified by HTS in TBD fail at a much later stage than those from SBD. Given these differences, one may think that, on average, SBD would generate more higher quality leads in a shorter time span. However, this may be counter balanced by the fact that additional efforts will have to be devoted to identify the target(s) modulated by the SBD lead compound. This last step is useful from the point of view of understanding how a given compound works, but it is not required to fully understand the mechanism of action of a drug in order to obtain permission to test it in humans. The prerequisites for a compound to enter clinical trials are demonstrated efficacy and safety in animal models of disease. As can be seen from this discussion, the development of a systems-based approach to drug discovery is in its infancy and represents an emerging need for new technological approaches.

Once an optimized lead compound is obtained, the remainder of the process to obtain a new therapeutic agent is similar between TBD and SBD, even though outcomes may be different. In both cases, drug candidates need to be tested in animal models of the disease. In addition to being ready for the preclinical phase faster, leads obtained via SBD may act on several targets. In contrast, leads from TBD are, in principle, selected to be effective at and selective for a single molecular target, although side effects exerted via actions at multiple unanticipated mechanisms is common. While modulation of a single target may be sufficient to treat a disease, cellular pathways are often resistant to perturbations [77], and drugs acting at several targets may be more successful. For the development phase, TBD and SBD converge. At this post-discovery stage, there are major differences between the preparation of small molecules and biotherapeutics. The development phase is the last stage before the drug is tested in humans. It consists in establishing manufacturing conditions to produce the compound and conducting any additional pre-clinical studies that may be needed. It is also in this phase that the clinical trials are planned. This requires determining the population to be tested, establishing protocols and planning for centers to conduct the trials. When this is done, an Investigational New Drug Application can be filled with the regulatory agency. This document details the findings obtained in in vitro and in vivo studies, with particular emphasis on

safety and efficacy, and the plans for clinical trials. If there are no concerns on the part of the regulatory agency, clinical trials are allowed to begin.

Administration of a test substance into humans is a highly regulated process whose details are beyond the scope of this chapter. However, to summarize the process, clinical trials are divided into three phases. Participants in each phase are required to sign an informed consent form detailing the potential risks and benefits associated with their participation in the trial. In phase I clinical trials, a small number of healthy subject are recruited to test the safety of the investigational new drug. The first PK/PD studies are also conducted during this stage. If no adverse events are reported, the study then proceeds with phase II. At this stage, varying doses of the drug are tested in an intermediate number of patients afflicted by the condition studied. These studies may also be conducted with a group of control patients receiving a placebo (with or without the standard approved treatment). The objective of phase II trials is to establish an efficacious dose and dosing regimen for the drug. In phase III, the drug is tested in a large number of patients across different geographical areas in a double-blind, placebo-controlled trial. The objective of this phase is to establish efficacy in a demographically diverse patient population. The large number of subject recruited may also help in identifying unforseen events occurring in only a small fraction of the population, which may have been missed in phases I and II due to the small sample size. Following the completion of the phase III studies and if the drug is deemed safe and efficacious, a New Drug Application (for a small molecule) or a Biological License Application (for a biotherapeutic) is filed with the regulatory agency to obtain the permission to commercialize the new therapeutic agent. As compounds enter clinical trials, it is thought that the SBD pipeline would be more successful on average due to early selection for activity in complex animal and/or human cell systems leading to the optimization of low toxicity hits targeting several members of a pathway or even different pathways involved in the disease.

### 3.4 Brief Look at SBD in the Context of Neurological Disorders

***Inherent Difficulties Related to the Study of the Nervous System*** For a successful application of SBD, a suitable animal model with objective and appropriate biomarkers must be established. When considering the molecular basis of a disorder, this may seem at first glance to be a fairly straightforward task as many molecular signaling cascades are conserved across the animal kingdom. For example, the components of the janus kinase (JAK)-signal transducer and activator of transcription (STAT) signaling pathway signaling pathway, which has been involved in a variety human diseases including brain cancer, neuronal injury following stroke, and depression in rheumatoid arthritis patients, can be found in the nematode *Caenorhabditis elegans*, the fruit fly *Drosophila melanogaster*, and in the fish *Danio rerio*, as well as in mammals such as mice, rats and monkeys [78–80].

However, the structure and/or functional domains of proteins may also diverge during evolution with important consequences for disease biology. As an example, mutations in the human leucine-rich repeat kinase 2 (LRRK2) gene have been associated with an autosomal-dominant form of Parkinson's disease (reviewed in ref. [81]). Although *Drosophila* and *C. elegans* both express LRRK, their gene lacks an N-terminal repeat compared to the human LRRK2, which occurred later in evolution [82]. Studies the endogenous LRRK protein in these animal models, therefore, may not be appropriate to further our understanding of the role of LRRK2 in healthy humans and in those afflicted with Parkinson's disease. This is a good example of a protein that acquired additional functional domains during evolution, and these new functional domains may be important for the biological basis of a disease. It follows that the development of new therapeutics for the treatment of Parkinson's disease would reasonably focus on LRRK2 rather than LRRK if TBD were used. With SBD, the function of LRRK in the context of systems biology could correlate with the function of LRRK2. If true, and if a valid system approach could be conceived of, results from a lower organism could be used, perhaps in combination with TBD. Alternatively, it is possible that addition of new functional domains on a protein would affect its signaling and/or protein–protein interactions, in which case the pathway in which this specific protein is involved would need to be studied for SBD.

One example of an SBD strategy to streamline discovery for neurological disorders would be to pre-screen the initial chemical library to reduce its size and to begin the drug discovery process by using a hybrid approach that incorporates elements of TBD. Since 98% of compounds do not cross the BBB [52], performing an initial BBB screen on the entire chemical library would result in a focused library for neurological applications. Then, using knowledge of the molecular basis of a particular disorder, to the extent that it exists, some target-based screening might also be valuable. For example, positive modulation of a selected group of $GABA_A$ receptor subtypes by certain benzodiazepines and related therapeutics reduces anxiety in a safe and effective way. But all produce side effects such as drowsiness and sedation that limit their usefulness for day time anxiety. Whereas

the search for a selective anxiolytic free of sedative-hypnotic side effects has been driven by a TBD approach based upon $GABA_A$ receptor α2 subunit selectivity (reviewed in ref. [83]), ocinaplon was found to be a non-sedating anxiolytic in human subjects but surprisingly, relatively non-selective towards the α1, α2 and α3 subunit-containing $GABA_A$ receptors [84].

In the search for new anxiolytics, a hybrid approach in which BBB-permeable compounds are screened for activity against the benzodiazepine-responsive $GABA_A$ receptor subtypes could be effective. Instead of focusing on subtype selectivity as in TBD, compounds identified through HTS could be tested using the SBD approach. Following animal toxicity studies and phase I clinical trials, compounds that are likely to be effective in humans could be identified using system response profiling. It would seem that the search for new treatments for anxiety disorders and other neurological disorders would benefit greatly from future improvements in technologies for high-resolution analysis of nervous system function that are applicable to the intact human brain and spinal cord. Using these technologies nervous system responses in patients and normal controls could be generated. Drug candidates could then be evaluated in a small number of subjects using focused clinical trials to identify lead compounds that bring the patient's nervous system response profile toward that of a normal control. This approach would likely yield compounds acting on multiple targets but may improve future success rates.

Expanding the scale to a specific group of cells or an anatomical region containing a physiological pathway that is known to correlate with or be critical in the disease biology could be used. For example, the development of new therapeutic agents in the area of cognitive enhancement would fulfill an important unmet medical need. The hippocampus and parahippocampal region of the medial temporal lobe are well known to be important for declarative memory, the type of memory involved in remembering places, time, and events, as well as the emotions associated with these situations. One of the most famous examples for the association of the medial temporal lobe with declarative memory is provided by H.M., a patient with intractable epilepsy, who had his parietal temporal lobe surgically removed bilaterally in 1953 [85]. While H.M.'s seizures became manageable, he suffered severe anterograde amnesia, with partial retrograde amnesia, making him unable to complete episodic or semantic memory tasks, both of which are part of declarative memory (reviewed in ref. [86]). However, his short-term memory and ability to acquire new motor skills remained intact. Subsequent lesion studies in monkeys and rodents have confirmed the importance of the hippocampus and parahippocampal region for declarative memory, although the specific role of each individual anatomical region within the medial temporal lobe remains controversial (reviewed in ref. [87] and [88]).

Divergences in studying declarative memory across species may be explained in part by evolutionary divergences in the anatomical structures involved. Looking across mammalian species, the hippocampus appears to be well conserved both in its anatomical organization and in the intrinsic connections between its cornu ammonis (CA) field and dentate gyrus [89]. However, small variations can be observed in the parahippocampal region, and further disparities are found in the neocortex, which provides input to the hippocampus via the parahippocampal region. These differences may arise from connectivity patterns, cortical size and/or laminar stratification. Hence, if sensory information coming to the hippocampus from the neocortex differs across species, it is a strong possibility that the behavior of the animal, influenced by the hippocampal output, will also differ.

While some behaviors, such as recollection of previous events, may differ somewhat between species, other demeanors are strictly human and even more difficult to model in animals. An example of such behaviors found in schizophrenia, which includes positive (hallucinations, delusions, thought disorders), negative (anhedonia, lack of emotion and motivation, catatonia), and cognitive (attention and memory impairments) symptoms [90] that have yet to be reproduced in a comprehensive animal model. Models currently used can be classified as either predictive if they can be used to predict the therapeutic ability of a drug; or partially homologous if they can recapitulate some of the symptoms observed in humans [91]. One criticism associated with using predictive models is that they may miss drug candidates acting via a novel mechanism because these animal models are established and validated using currently available therapeutics. In addition, predictive models may be using behavioral assessment that represents only a minor problem in schizophrenic patients [91]. In the case of partially homologous models, in the absence of a better understanding of the disease biology of schizophrenia, it is difficult to assess whether treating the symptom(s) or objective responses represented in the animal model will truly yield an effective therapeutic in human patients.

Identification of new therapeutics for neurological disorders poses a problem in that there is often a lack of evolutionary conservation in the biological processes involved. While this can be solved at the molecular level by studying the human components involved in a pathway, it is difficult to establish a convincing parallel between information obtained from animal models in rodents and non-human primates and the results obtained in clinical trials. Furthermore, as shown in

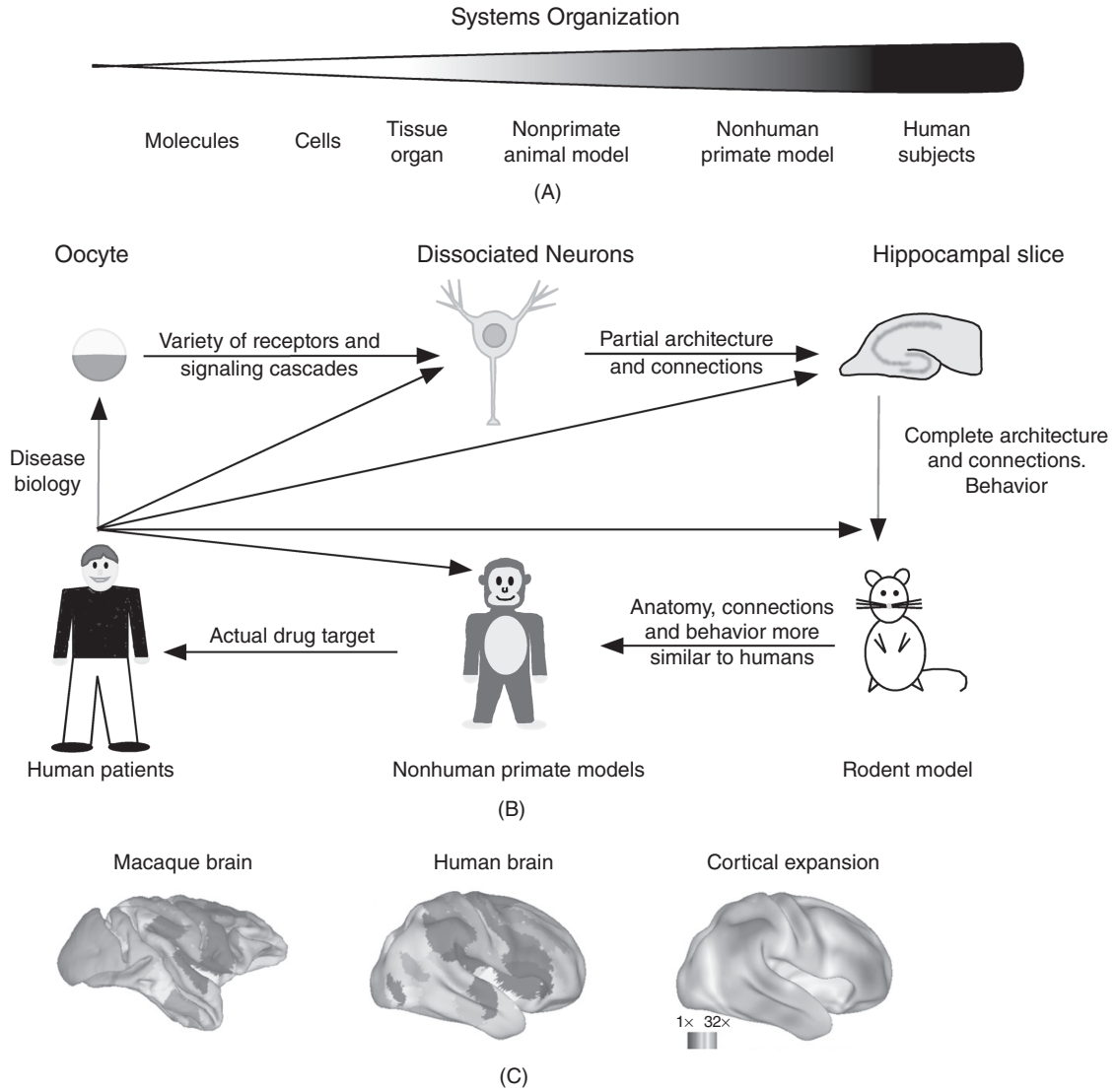


**Figure 5** A systems pharmacology view of a platform for drug discovery in the nervous system. (A) The systems used for drug discovery can be arranged by increasing organization level. (B) Example of an adaptation of the organization level scale for drug discovery oriented towards neurological disorders. Although drug discovery programs may start at different steps of the process outlined here, a number of adaptations must be made when going from one model to the next. (C) There are a number of differences even between the human and macaque brain. PALS (population-average, Landmark, and Surface-based) rendering of the macaque and human brains are shown along with functional areas defined by Lewis and Van Essen [94]. Based on the localization and size of those functional areas, a map of cortical expansion can be designed. (Reproduced from ref. [92].)

Figure 5, discrepancies can arise at multiple levels, making it even more difficult to find an accurate animal model, even in species such as the macaque which is closely related to the human.

***On the use of Integrative Databases as a Tool to Facilitate SBD for Neurological Disorders*** Since the use of animal models is essential to drug discovery and development, one SBD approach would be to track and record differences and similarities between subregions of the nervous system of different species. This includes their wiring diagram at strategic developmental stages depending upon the disorder to be treated. This neural systems-level information on circuitry could then be linked to anatomical information with functional data such as electrophysiology, behavior, functional magnetic resonance imaging (fMRI), gene expression and neurochemistry. Since there is a need to represent a

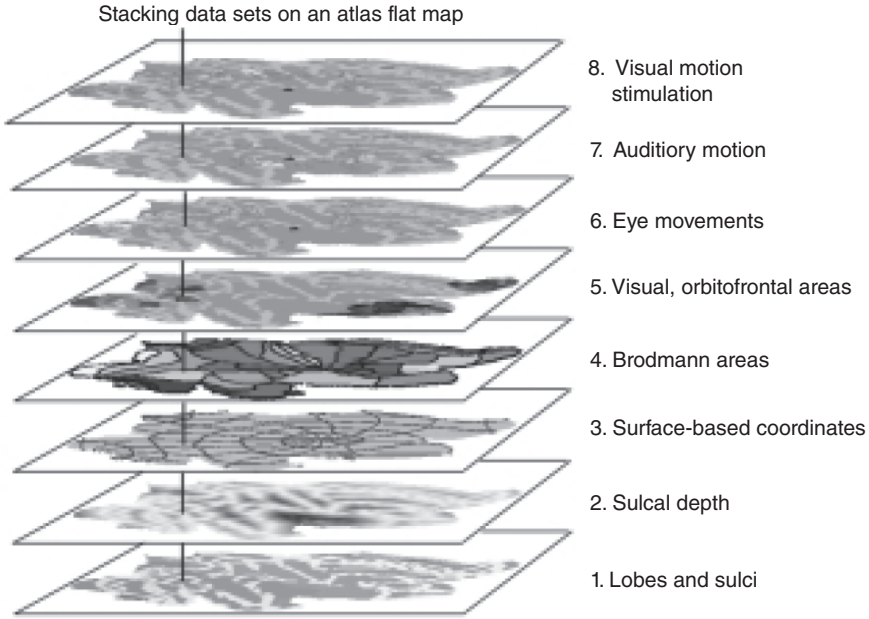


**Figure 6** Example of a brain atlas. Anatomical and functional information may be displayed on a flat map (reproduced from ref. [97].)

large amount of annotated anatomical information, databases referred to as multimodal brain atlases have emerged (Fig 6) [92]. Electronic brain atlases offer several advantages over print atlases in that they provide a dynamic link between structure and function, which is easily accessible, searchable, updated and linked to other databases and resources [93]. In addition, whereas traditional atlases were based on post-mortem studies of one or a few brains, multimodal atlases incorporate information from MRI and allow the creation of population averages and probabilistic analysis.

There are a number of databases currently available which may help in the application of SBD to neurological disorders. Human and macaque surface-based brain atlases using flat images and/or 3-dimentional reconstructions (reviewed in ref. [93]) are derived mainly from imaging studies, while other databases include different types of information. The Allen Brain Atlas provides a comprehensive analysis for the expression of 20,000 genes in the brain of adult (post-natal day 56, P56) male mice [95]. Searches in this freely available online atlas provide images from in-situ hybridization, annotated for major brain structures, and links for the gene or protein of interest in other databases. The GenePaint Atlas also provides information on gene expression in the mouse brain but includes brain from different developmental stages [96]. Although the number of genes covered is smaller than that of the Allen Brain Atlas, information on spatio-temporal expression of genes can be especially important for developmental disorders such as autism and schizophrenia. In addition, the database is updated weekly, and researchers can submit requests for candidate genes [98].

Another issue frequently encountered when studying a developmentally regulated process is the "conversion factor" between the brain of two species. To address this question, Clancy and colleagues used neuroinformatics to translate brain development data across 10 species including mice, rats, cats, macaques and humans [99]. Since different brain regions mature at different times in different species, this program further allows to compare two animals for cortical events, limbic events, and non-cortical/limbic events. In the context of SBD, this would help establishing a valid animal model for developmentally regulated processes and diseases.

## 4 CONCLUDING REMARKS

The science of pharmacology has progressed tremendously since its inception. Microbial infections that were once fatal can now be treated with antibiotics and other conditions for which there is still no cure, such as type I diabetes and certain breast cancers can be managed in many patients using a number of approaches including the appropriate medication. However the promise of a horn of plenty filled with new biology "wonder drugs" based on modern discoveries of pharmacology com-

bined with advances such as the sequencing of the human genome remains at the end of the rainbow. One could even wonder whether the outlook for the development of breakthrough therapeutics does not look grim with the continued decrease in yearly new drug approvals over the past decade combned with increased legal, social and political pressure. To resolve this issue, the pharmaceutical industry may need to bring a shift in its drug discovery and development strategies, whether it is for the generation of new small molecules, biotherapeutics, or nucleic-acid-based therapeutics and/or more productive focused clinical trials. Although TBD has seemed, in theory, the way to increase efficacy while reducing side effects of new therapeutics, a look at the current rates of success and failure for new drug candidates, as well as their associated costs, seems to suggest otherwise. An alternative approach, as suggested here, would be to consider enhancing SBD strategies within the drug discovery scheme. This may reduce the fraction of drug candidates that have inadequate efficacy and/or induce significant side effects, decreasing the average cost of discovery and development that accrues as drugs emerging from company pipelines fail pre- and postmaketing. At the same time it is possible to generate compounds that act through novel mechanisms or via combined actions at several targets. Similarly, a polypharmaceutical approach in which a novel combination of approved therapeutics may be identified by SBD provides practical advantages if the screening process could be shown to yield substantially higher value lead therapeutic combinations. However, the application of SBD may be particularly challenging in certain diseases where it is difficult to accurately represent the spectrum of symptoms observed in humans within an animal model. To remedy this, an important objective for R&D is to identify better biomarkers for disease progression, develop a better understanding of cross-species differences and, based on this information, to develop more valid animal models that can be monitored using the same biomarkers carried through to clinical trials.

Nowhere is the disconnect between human disease and animal models for disease more apparent than for those disorders based upon dysfunction of the nervous system. As just one example, the nervous system challenges biomedical science to break the barrier of disconnect between the anxiety perceived by the patient with generalized anxiety disorder and the animal models for it, using biomarkers in order to guide discovery. Regardless of the approach taken, efforts must be made to diversify the treatment options available and to provide new therapeutics for conditions where there is a medical need. For many neurological disorders, such as schizophrenia and depression, treatment options are associated with severe side effects and are effective in only a fraction of the population, whereas in other conditions, such as Alzheimer's and Parkinson's diseases, therapeutics only reduce somewhat the appearance of symptoms in some patients. Whereas therapeutics exist for the treatment of anxiety disorders, convulsive disorders and sleep disorders, the existing treatments are palliative and burdened with untoward side effects. In addition, there is no treatment available for a number of neurological diseases including spinal muscular atrophy, a hereditary motor neuron disorder. While the nervous system is complex and presents with additional challenges, such as the BBB and the BNFB, which are absent in other organs, it is critical that efforts be invested in this area to overcome this lag in the development of new therapeutic agents. It is our hope that advances in SBD, combined with biomarker development, validated animal models for diseases, and enhancements in the design, surveillance and interpretation of clinical trials will lead to the discovery of therapeutic cures rather than palliative treatments for disorders and diseases that afflict us.

## ACKNOWLEDGMENTS

This work was supported by the National Institute of Mental Health (MH49469), National Institute of General Medical Sciences (GM008541), National Institute on Drug Abuse (DA219682) and by the PhRMA Foundation (pre-doctoral fellowship).

## REFERENCES

1. Sullivan R (1995). A brief journey into medical care and disease in ancient Egypt. *J R Soc Med 88*, 141–145.
2. Subbarayappa BV (2001). The roots of ancient medicine: An historical outline. *J Biosci 26*, 135–143.
3. Drews J (2000). Drug discovery: A historical perspective. *Science 287*, 1960–1964.
4. Langley JN (1905). On the reaction of cells and of nerve-endings to certain poisons, chiefly as regards the reaction of striated muscle to nicotine and to curari. *J Physiol 33*, 374–413.
5. Maehle AH (2004). "Receptive substances": John Newport Langley (1852–1925) and his path to a receptor theory of drug action. *Med Hist 48*, 153–174.
6. Meldrum U, Roughton JW (1933). Carbonic anhydrase. Its preparation and properties. *J Physiol 80*, 131.
7. Fleming A (1929). On the antibacterial action of cultures of a penicillium, with special reference to their use in the isolation of B. influenzae. *Br J Exp Pathol 10*, 226–236.
8. Barnhart J (1989). DOE Human Genome Program. *U.S. Department of Energy, Human Genome News (v1n1)*.

9. Wani MC, Taylor HL, Wall ME, Coggon P, McPhail AT (1971). Plant antitumor agents. VI. The isolation and structure of taxol, a novel antileukemic and antitumor agent from Taxus brevifolia. *J Am Chem Soc 93*, 2325–2327.
10. Schiff PB, Fant J, Horwitz SB (1979). Promotion of microtubule assembly *in vitro* by taxol. *Nature 277*, 665–667.
11. Guénard D, Guéritte-Voegelein F, Dubois J, Potier P (1993). Structure-activity relationships of Taxol and Taxotere analogues. *J Natl Cancer Inst Monogr 15*, 79–82.
12. Guéritte-Voegelein F, Guénard D, Lavelle F, Le Goff MT, Mangatal L, Potier P (1991). Relationships between the structure of taxol analogues and their antimitotic activity. *J Med Chem 34*, 992–998.
13. Riou JF, Naudin A, Lavelle F (1992). Effects of Taxotere on murine and human tumor cell lines. *Biochem Biophys Res Commun 187*, 164–170.
14. Bissery MC, Guénard D, Guéritte-Voegelein F, Lavelle F (1991). Experimental antitumor activity of taxotere (RP 56976, NSC 628503), a taxol analogue. *Cancer Res 51*, 4845–4852.
15. Zimmermann J, Buchdunger E, Mett H, Meyer T, Lydon NB (1997). Potent and selective inhibitors of the Abl-kinase: Phenylamino-pyrimidine (PAP) derivatives. *Bioorg Med Chem Lett 7*, 187–192.
16. Capdeville R, Buchdunger E, Zimmermann J, Matter A (2002). Glivec (STI571, imatinib), a rationally developed, targeted anticancer drug. *Nat Rev Drug Discov 1*, 493–502.
17. Rowley JD (1973). Letter: A new consistent chromosomal abnormality in chronic myelogenous leukaemia identified by quinacrine fluorescence and Giemsa staining. *Nature 243*, 290–293.
18. Chan LC, Karhi KK, Rayter SI, Heisterkamp N, Eridani S, Powles R, Lawler SD, Groffen J, Foulkes JG, Greaves MF (1987). A novel abl protein expressed in Philadelphia chromosome positive acute lymphoblastic leukaemia. *Nature 325*, 635–637.
19. Schindler T, Bornmann W, Pellicena P, Miller WT, Clarkson B, Kuriyan J (2000). Structural mechanism for STI-571 inhibition of abelson tyrosine kinase. *Science 289*, 1938–1942.
20. Cecchelli R, Berezowski V, Lundquist S, Culot M, Renftel M, Dehouck MP, Fenart L (2007). Modelling of the blood–brain barrier in drug discovery and development. *Nat Rev Drug Discov 6*, 650–661.
21. Goodwin JT, Clark DE (2005). In silico predictions of blood–brain barrier penetration: Considerations to "keep in mind". *J Pharmacol Exp Ther 315*, 477–483.
22. Upton RN (2007). Cerebral uptake of drugs in humans. *Clin Exp Pharmacol Physiol 34*, 695–701.
23. Terasaki T, Ohtsuki S, Hori S, Takanaga H, Nakashima E, Hosoya K (2003). New approaches to in vitro models of blood-brain barrier drug transport. *Drug Discov Today 8*, 944–954.
24. Linnet K, Ejsing TB (2008). A review on the impact of P-glycoprotein on the penetration of drugs into the brain. Focus on psychotropic drugs. *Eur Neuropsychopharmacol 18*, 157–169.
25. Sherwin BB, Henry JF (2008). Brain aging modulates the neuroprotective effects of estrogen on selective aspects of cognition in women: A critical review. *Front Neuroendocrinol 29*, 88–113.
26. Hadji P (2008). Menopausal symptoms and adjuvant therapy-associated adverse events. *Endocr Relat Cancer 15*, 73–90.
27. Rick NG (2004). *Drugs: From Discovery to Approval*. Hoboken, NJ: Wiley, p. 355.
28. Dingermann T (2008). Recombinant therapeutic proteins: Production platforms and challenges. *Biotechnol J 3*, 90–97.
29. Morrow JF, Berg P (1972). Cleavage of Simian virus 40 DNA at a unique site by a bacterial restriction enzyme. *Proc Natl Acad Sci USA 69*, 3365–3369.
30. Johnson IS (2003). The trials and tribulations of producing the first genetically engineered drug. *Nat Rev Drug Discov 2*, 747–751.
31. Muller AF, Kopchick JJ, Flyvbjerg A, van der Lely AJ (2004). Clinical review 166: Growth hormone receptor antagonists. *J Clin Endocrinol Metab 89*, 1503–1511.
32. Strand V, Kimberly R, Isaacs JD (2007). Biologic therapies in rheumatology: Lessons learned, future directions. *Nat Rev Drug Discov 6*, 75–92.
33. Stebbings R, Findlay L, Edwards C, Eastwood D, Bird C, North D, Mistry Y, Dilger P, Liefooghe E, Cludts I, Fox B, Tarrant G, Robinson J, Meager T, Dolman C, Thorpe SJ, Bristow A, Wadhwa M, Thorpe R, Poole S (2007). "Cytokine storm" in the phase I trial of monoclonal antibody TGN1412: Better understanding the causes to improve preclinical testing of immunotherapeutics. *J Immunol 179*, 3325–31.
34. Jain M, Kamal N, Batra SK (2007). Engineering antibodies for clinical applications. *Trends Biotechnol 25*, 307–316.
35. Park SI, Press OW (2007). Radioimmunotherapy for treatment of B-cell lymphomas and other hematologic malignancies. *Curr Opin Hematol 14*, 632–638.
36. Pardridge WM (2006). Molecular Trojan horses for blood–brain barrier drug delivery. *Curr Opin Pharmacol 6*, 494–500.
37. Kurreck J (2003). Antisense technologies. Improvement through novel chemical modifications. *Eur J Biochem 270*, 1628–1644.
38. Nichols WG, Boeckh M (2000). Recent advances in the therapy and prevention of CMV infections. *J Clin Virol 16*, 25–40.
39. Kramer JA, Sagartz JE, Morris DL. (2007). The application of discovery toxicology and pathology towards the design of safer pharmaceutical lead candidates. *Nat Rev Drug Discov 6*, 636–649.
40. Grimm D, Streetz KL, Jopling CL, Storm TA, Pandey K, Davis CR, Marion P, Salazar F, Kay MA (2006). Fatality in mice due to oversaturation of cellular microRNA/short hairpin RNA pathways. *Nature 441*, 537–541.

41. Fire A, Xu S, Montgomery MK, Kostas SA, Driver SE, Mello CC (1998). Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. *Nature 391*, 806–811.
42. Davidson BL, Boudreau RL (2007). RNA interference: A tool for querying nervous system function and an emerging therapy. *Neuron 53*, 781–788.
43. Love TM, Moffett HF, Novina CD (2008). Not miR-ly small RNAs: Big potential for microRNAs in therapy. *J Allergy Clin Immunol 121*, 309–319.
44. Gewirtz AM (2007). On future's doorstep: RNA interference and the pharmacopeia of tomorrow. *J Clin Invest 117*, 3612–3614.
45. Sonkoly E, Ståhle M, Pivarcsi A (2008). MicroRNAs and immunity: novel players in the regulation of normal immune function and inflammation. *Semin Cancer Biol 18*, 131–140.
46. Grimm D, Kay MA (2007). Therapeutic application of RNAi: Is mRNA targeting finally ready for prime time? *J Clin Invest 117*, 3633–3641.
47. Raoul C, Barker SD, Aebischer P (2006). Viral-based modelling and correction of neurodegenerative diseases by RNA interference. *Gene Ther 13*, 487–495.
48. Pharmaceutical Research and Manufacturers of America (PhRMA) (2007). Annual Report PhRMA. Washington, DC.
49. Owens J (2007). 2006 drug approvals: Finding the niche. *Nat Rev Drug Discov 6*, 99–101.
50. Hughes B (2008). 2007 FDA drug approvals: A year of flux. *Nat Rev Drug Discov 7*, 107–109.
51. World Health Organization (WHO) (2006). *Neurological Disorders: Public Health Challenges*. WHO, Geneva, p. 232.
52. Pardridge WM (2002). Drug and gene targeting to the brain with molecular Trojan horses. *Nat Rev Drug Discov 1*, 131–139.
53. Pangalos MN, Schechter LE, Hurko O (2007). Drug development for CNS disorders: strategies for balancing risk and reducing attrition. *Nat Rev Drug Discov 6*, 521–532.
54. Collins SD (2007). Discontinued drugs in 2006: Central and peripheral nervous system drugs. *Expert Opin Investig Drugs 16*, 1743–1751.
55. Schwartz JI, Dallob AL, Larson PJ, Laterza OF, Miller J, Royalty J, Snyder KM, Chappell DL, Hilliard DA, Flynn ME, Cavanaugh PF, Wagner JA (2008). Comparative Inhibitory Activity of Etoricoxib, Celecoxib, and Diclofenac on COX-2 Versus COX-1 in Healthy Subjects. *J Clin Pharmacol 48*, 745–754.
56. Panel on Antiretroviral Guidelines for Adults and Adolescents (2008). *Guidelines for the use of antiretroviral agents for HIV-1-infected adults and adolescents*. Developed by the DHHS Panel on Antiretroviral Guidelines for Adults and Adolescents—A Working Group of the Office of AIDS Research Advisory Council (OARAC). Bethesda, MD. p. 128.
57. National High Blood Pressure Education Program Coordinating Committee (2003). Paper presented at *The Seventh Report of the Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure (JNC 7 Express)*. Bethesda, MD. p. 34.
58. Lindsay MA (2003). Target discovery. *Nat Rev Drug Discov 2*, 831–838.
59. Keller TH, Pichota A, Yin Z (2006). A practical view of "druggability". *Curr Opin Chem Biol 10*, 357–361.
60. Knowles J, Gromo G (2003). A guide to drug discovery: Target selection in drug discovery. *Nat Rev Drug Discov 2*, 63–69.
61. Ekins S, Mestres J, Testa B (2007). In silico pharmacology for drug discovery: Applications to targets and beyond. *Br J Pharmacol 152*, 21–37.
62. Kumar RA, Clark DS (2006). High-throughput screening of biocatalytic activity: Applications in drug discovery. *Curr Opin Chem Biol 10*, 162–168.
63. Yaghoubi N, Malayev A, Russek SJ, Gibbs TT, Farb DH (1998). Neurosteroid modulation of recombinant ionotropic glutamate receptors. *Brain Res 803*, 153–160.
64. Walters WP, Namchuk M (2003). Designing screens: How to make your hits a hit. *Nat Rev Drug Discov 2*, 259–266.
65. Huber W (2005). A new strategy for improved secondary screening and lead optimization using high-resolution SPR characterization of compound–target interactions. *J Mol Recognit 18*, 273–281.
66. Nikitenko AA (2006). Compound scale-up at the discovery–development interface. *Curr Opin Drug Discov Devel 9*, 729–740.
67. Di L, Kerns EH (2003). Profiling drug-like properties in discovery research. *Curr Opin Chem Biol 7*, 402–408.
68. Yengi LG, Leung L, Kao J (2007). The evolving role of drug metabolism in drug discovery and development. *Pharm Res 24*, 842–858.
69. Agoram BM, Martin SW, van der Graaf PH (2007). The role of mechanism-based pharmacokinetic-pharmacodynamic (PK-PD) modelling in translational research of biologics. *Drug Discov Today 12*, 1018–1024.
70. Butcher EC (2005). Can cell systems biology rescue drug discovery? *Nat Rev Drug Discov 4*, 461–467.
71. Haney SA, LaPan P, Pan J, Zhang J (2006). High-content screening moves to the front of the line. *Drug Discov Today 11*, 889–894.
72. Terstappen GC, Schlüpen C, Raggiaschi R, Gaviraghi G (2007). Target deconvolution strategies in drug discovery. *Nat Rev Drug Discov 6*, 891–903.
73. van der Greef J, McBurney RN (2005). Innovation: Rescuing drug discovery: In vivo systems pathology and systems pharmacology. *Nat Rev Drug Discov 4*, 961–967.
74. Ideker T, Galitski T, Hood L (2001). A new approach to decoding life: Systems biology. *Annu Rev Genomics Hum Genet 2*, 343–372.
75. Lamb J, Crawford ED, Peck D, Modell JW, Blat IC, Wrobel MJ, Lerner J, Brunet JP, Subramanian A, Ross KN, Reich

M, Hieronymus H, Wei G, Armstrong SA, Haggarty SJ, Clemons PA, Wei R, Carr SA, Lander ES, Golub TR (2006). The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease. *Science 313*, 1929–1935.

76. Antipova AA, Stockwell BR, Golub TR (2008). Gene expression-based screening for inhibitors of PDGFR signaling. *Genome Biol 9*, R47.
77. Araujo RP, Liotta LA, Petricoin EF (2007). Proteins, drug targets and the mechanisms they control: The simple truth about complex networks. *Nat Rev Drug Discov 6*, 871–880.
78. Arbouzova NI, Zeidler MP (2006). JAK/STAT signalling in Drosophila: insights into conserved regulatory and cellular functions. *Development 133*, 2605–2616.
79. Malemud CJ, Miller AH (2008). Pro-inflammatory cytokine-induced SAPK/MAPK and JAK/STAT in rheumatoid arthritis and the new anti-depression drugs. *Expert Opin Ther Targets 12*, 171–183.
80. Planas AM, Gorina R, Chamorro A (2006). Signalling pathways mediating inflammatory responses in brain ischaemia. *Biochem Soc Trans 34*, 1267–1270.
81. Morris HR (2007). Autosomal dominant Parkinson's disease and the route to new therapies. *Expert Rev Neurother 7*, 649–656.
82. Marín I (2006). The Parkinson disease gene LRRK2: evolutionary and structural insights. *Mol Biol Evol 23*, 2423–2433.
83. Berezhnoy, Gravielle C, Farb H (2007). Pharmacology of the GABA-A receptor In Silbey D, Kuhar M, Hanin I, Skolnick P, Eds. *Handbook of Contemporary Neuropharmacology*. Hoboken, NJ: Wiley, pp. 465–568.
84. Lippa A, Czobor P, Stark J, Beer B, Kostakis E, Gravielle M, Bandyopadhyay S, Russek SJ, Gibbs TT, Farb DH, Skolnick P (2005). Selective anxiolysis produced by ocinaplon, a $GABA_A$ receptor modulator. *Proc Natl Acad Sci USA 102*, 7380–7385.
85. Scoville WB, Milner B (1957). Loss of recent memory after bilateral hippocampal lesions. *J Neurol Neurosurg Psychiatry 20*, 11–21.
86. Corkin S (2002). What's new with the amnesic patient H.M.? *Nat Rev Neurosci 3*, 153–160.
87. Buckley MJ (2005). The role of the perirhinal cortex and hippocampus in learning, memory, and perception. *Q J Exp Psychol B 58*, 246–268.
88. Mumby DG (2001). Perspectives on object-recognition memory following hippocampal damage: lessons from studies in rats. *Behav Brain Res 127*, 159–181.
89. Manns JR, Eichenbaum H (2006). Evolution of declarative memory. *Hippocampus 16*, 795–808.
90. National Institute of Mental Health (2007). *Schizophrenia* Bethesda, MD: NIMH, p. 25.
91. Tordjman S, Drapier D, Bonnot O, Graignic R, Fortes S, Cohen D, Millet B, Laurent C, Roubertoux PL (2007). Animal models relevant to schizophrenia and autism: Validity and limitations. *Behav Genet 37*, 61–78.
92. Toga AW, Thompson PM, Mori S, Amunts K, Zilles K (2006). Towards multimodal atlases of the human brain. *Nat Rev Neurosci 7*, 952–966.
93. Van Essen DC, Dierker DL (2007). Surface-based and probabilistic atlases of primate cerebral cortex. *Neuron 56*, 209–225.
94. Lewis JW, Van Essen DC (2000). Mapping of architectonic subdivisions in the macaque monkey, with emphasis on parieto-occipital cortex. *J Comp Neurol 428*, 79–111.
95. Lein ES, et al. (2007). Genome-wide atlas of gene expression in the adult mouse brain. *Nature 445*, 168–176.
96. Visel A, Thaller C, Eichele G (2004). GenePaint.org: An atlas of gene expression patterns in the mouse embryo. *Nucleic Acids Res 32*, D552–D556.
97. Van Essen DC (2002). Windows on the brain: the emerging role of atlases and databases in neuroscience. *Curr Opin Neurobiol 12*, 574–579.
98. Koester SE, Insel TR (2007). Mouse maps of gene expression in the brain. *Genome Biol 8*, 212.
99. Clancy B, Finlay BL, Darlington RB, Anand KJ (2007). Extrapolating brain development from experimental species to humans. *Neurotoxicology 28*, 931–937.

# *Vibrational Circular Dichroism: Ensuring Quality of Pharmaceutical Products*

**Dimiter Tsankov[1], Alexander Polyanichko[2], and Helmut Wieser[3]**

[1] *Institute of Organic Chemistry, Bulgarian Academy of Sciences, Sofia, Bulgaria*
[2] *Faculty of Physics and Institute of Cytology, RAS, St. Petersburg State University, St. Petersburg, Russian Federation*
[3] *Department of Chemistry, University of Calgary, Calgary, Alberta, Canada*

## 1 BACKGROUND

Aside from very early interests in optical rotation as well as rotatory dispersion and circular dichroism (CD), "optical activity" of chiral molecules has motivated chemists and physicists for some time (1). Pertinent for the purpose of this article is that optical activity can arise from *electronic* transitions, which normally occur in the *visible* and *ultraviolet* spectral regions, or from *molecular vibrations* observed in the *infrared* (IR) region. Of specific interest in the text to follow are the differences in absorption between left and right circularly polarized light. Known respectively as *electronic* circular dichroism (ECD) and *vibrational* circular dichroism (VCD), both

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing,*
Edited by Shayne C. Gad
Copyright © 2011 John Wiley & Sons, Inc.

techniques are useful and unique spectroscopic tools. The interest in VCD in particular has mushroomed during the 1970s resulting largely from endeavors of researchers at the Universities of Minnesota (G. Holzwarth), Southern California (P.J. Stephens), Chicago (T.A. Keiderling), and Syracuse (L.A. Nafie) from which a truly remarkable and relatively rapid development followed. Discussions in the context of the essay to follow will be restricted to discoveries and applications since the turn of this century assuming that the earlier information is reasonably well documented, known, and understood.

Upon perusing pertinent literature citations, it will be clear that VCD represents the difference in absorption intensity of left versus right circularly polarized radiation most commonly in the infrared region and particularly between 800 and 2000 $cm^{-1}$. Although this limitation is not absolute, the VCD signals of molecular vibrations in this region are those most pertinent and prominent. The vibrational modes within this region are largely skeletal distortions, be they stretching or bending modes. Cyclic molecules therefore are the most interesting targets but not exclusively so. For this reason as well, favorite molecular systems for investigation are a plethora of biological molecules that perform myriads of important functions. Among these are such highly interesting and pertinent molecular components as DNA, RNA, and proteins in addition to individual drug molecules all of which usually contain one or more cyclic components. Fortunately, and although some of the pioneering investigators managed successfully often with home-built instruments, VCD spectrometers are now available from several manufacturers.[a] One can therefore expect the use of VCD to grow at a goodly pace.

In support of this prediction, one needs to mention only a few examples of applications, including relatively simple measurements of VCD data for several synthetic polyribonucleotides (2), VCD studies of the A-to-B conformation transition in DNA (3), B–Z conformational transitions of DNA including *ab initio* interpretations of VCD spectra (4), determining absolute configuration and solution conformation of materials such as gossypol (5), vibrational circular dichroism spectroscopy and the effects of metal ions on DNA structure (6), determining solution conformations of cyclosporin complexes (7), DNA–oligonucleotide–cisplatin binding (8), and many more. Pertinent aspects of these and other systems will be discussed in greater detail below. Of particular interest, impact, and utility for this article is a recent publication in which the authors provided a detailed description of progress in vibrational optical activity for determining structural properties of pharmaceuticals (9). Last but not the least, the general term "optical activity" also includes Raman optical activity (ROA) for which Barron et al. (10) have provided interesting and pertinent examples; however, this topic is not included in the text to follow.

[a] PMA 50 accessory in conjunction with the Tensor™ series FT-IR (Bruker Optics), Chiral*IR*™ (BioTools), Nicolet™ FT-IR coupled with PEM module (Nicolet Thermo), FVS-4000 vibrational CD spectrometer (JASCO), and Varian 7000 FT-IR spectrometer with Hinds PEM.

# 2 GENERAL DESCRIPTION

## 2.1 Vibrational Circular Dichroism

VCD is a chiroptical spectral method, which while similar to its analogs, optical rotatory dispersion (ORD) and ECD, discriminates between enantiomers because of their differential response to left versus right circularly polarized light. While ORD is due to the difference in the refractive indices that cause different propagation velocities of oppositely polarized radiation components through chiral media (circular birefringence), CD, either electronic or vibrational, originates from different absorption of polarized light by chiral molecules. Primarily, circular dichroism was measured in the ultraviolet and visible spectral range where the presence of a chromophore is required to detect electronic transitions. Vibrational circular dichroism is caused by $3N - 6$ vibrations of the atoms in a molecule and is related to transitions between distinct vibrational levels.

## 2.2 Instrumentation Methods

The VCD spectra are measured as the difference in absorbances for left and right circularly polarized radiation by a chiral molecule as defined by Eq. 1:

$$\Delta A = A_L - A_R, \tag{1}$$

where $A_L$ and $A_R$ are the corresponding absorbances. The values of $A_L$ and $A_R$ are to a great extent alike, which makes their difference quite small and usually three to four orders of magnitude smaller than either one of them. The detection of such weak signals poses substantial challenges for the VCD experiment. The experimental measurement requires decomposition of the electric vector of incident linearly polarized radiation into left and right circularly polarized components that are further allowed to irradiate the sample. This can be achieved by introducing phase retardation of 90° ($\frac{1}{4}\lambda$) between these components. Three distinct physical approaches have been employed so far for this purpose, namely, the photoelastic modulator (PEM), polarizing Michelson interferometer, and total internal reflection. At present, photoelastic modulation remains to be the only practically usable method. Phase retardation is attained by periodical mechanical stress applied to an isotropic IR transparent crystal. As a result, the crystal becomes birefringent; for example, the polarized radiation components will travel with different speeds through the crystal. An additional yet substantial benefit that follows is the retardation, which is modulated between two states with phase difference of 90° (left and right circular polarization) in accordance with the applied mechanical stress. Thus, the sample is irradiated by polarized light alternating between left and right circular polarization and resulting in a differential interferogram registered by the detector. The light modulation ensures the desired *ac advantage* for measuring very small differences between absorption signals of nearly even amplitudes. A drawback of the method is that the optimal circular polarization is obtained only twice during the sinusoidal cycle of the applied

stress and for one wavenumber only. Apart from that, the radiation is elliptically polarized, which reduces the efficiency of the modulator.

In the polarizing Michelson interferometer (11), the conventional beam splitter is replaced by an IR polarizer, which serves to separate the components of the polarized beam. The phase modulation is attained at Fourier frequencies and is limited by the speed of the scanning mirror. Although this method circumvents the spectral limitations of the photoelastic modulation, it is unsuitable for routine applications because of long data acquisition times and accompanying instrumental drift.

Phase retardation could also be attained by means of total internal reflection. In order to reach 90° retardation at a certain azimuth of the incident polarization and definite refractive index of the totally reflecting element, it is only necessary to switch between $+45°$ and $-45°$ angles of incidence. Such a retarder, *Fresnel rhomb*, which is highly achromatic in the entire IR range, has already been used in dispersive (12) and FT spectrometers (13). Without modulation, the spectra are of low signal-to-noise (S/N) ratio and show persistent artifacts. This method still has unexploited potential and would allow for VCD measurement in the whole mid-IR range provided that proper modulation between the opposite polarization states is ensured.

The photoelastic modulation remains the most advantageous method when selected spectral intervals are to be scanned; all commercially available instruments offered by different manufacturers are based on this principle. Some Fourier transform spectrometers operate with double modulation, the interferogram being modulated once at Fourier frequencies and thereafter at PEM frequency. A typical scheme and optical design of a FT-VCD spectrometer used in essentially all modifications is shown in Fig. 1.

Among the commercially available spectrometers, only BioTools Inc. currently offers completely stand-alone instruments, where the VCD optics is incorporated and prealigned into the spectrometer. Other manufacturers, such as Bruker Optics, Nicolet Thermo, Varian, and JASCO, provide VCD modules that can be attached to standard FT-IR spectrometers. The IR beam from the interferometer is first linearly polarized

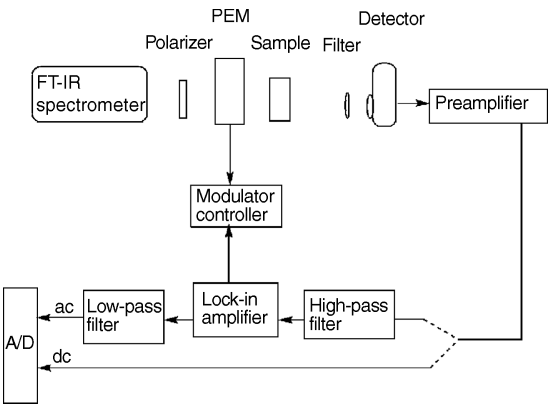


***FIGURE 1*** *Optical scheme of VCD setup.*

by a wire-grid polarizer and then enters the PEM. The polarizer is set at 45° with respect to the modular axis of the crystal to allow equal amplitudes of linearly polarized light to reach the orthogonal stress axes of the crystal (ZnSe) driven at 37 kHz. The modulated circularly polarized light passes through the sample positioned adjacent to the PEM and focused with a ZnSe lens onto an HgCdTe detector. The proper choice of a narrowband and low-noise detector is essential for the measurement. A long-wave pass optical filter placed either in front of the polarizer or in front of the detector stops the radiation above 1800 $cm^{-1}$. The low spectral range is limited by the ZnSe optics and a detector whose cutoff is about 800 $cm^{-1}$. These precautions preclude detector saturation in the wavenumber region where the PEM has low efficiency. The detector signal is preamplified and then divided into two paths. While the first path is directly fed to the analog-to-digital (A/D) converter (Fig. 1), the signal passes through several electronic devices in the second path. The signal is first filtered by a high-pass filter to remove the Fourier frequencies, thereafter demodulated by a lock-in amplifier tuned to the modulation frequency, passed through a second filter to remove any residual "dc" signal, and then fed to the A/D converter. The first path measures the ordinary sample interferogram usually designated as "dc". The second path yields the differential interferogram of the sample denoted as "ac". Fourier transformation generates the corresponding single beam spectra $I_{ac}$ and $I_{dc}$, and their ratio defines the raw VCD spectrum (14):

$$\frac{I_{ac}(\tilde{\nu})}{I_{dc}(\tilde{\nu})} = 2J_1[\alpha_0(\tilde{\nu})]\exp[-(2V\tilde{\nu})\tau](1.1513)\Delta A_{LR}(\tilde{\nu}). \tag{2}$$

The raw VCD spectrum is directly related to the difference in the absorbances, $\Delta A_{LR}(\tilde{\nu})$, and is thus expressed in arbitrary units. $J_1[\alpha_0(\tilde{\nu})]$ is the first-order Bessel function and represents the efficiency of the PEM at a definite wavenumber $\tilde{\nu}$ and maximum retardation $\alpha_0(\tilde{\nu})$. The exponential term characterizes the attenuation factor of the lock-in amplifier, where $V$ stands for the speed of the scanning mirror and $\tau$ is the time constant of the lock-in. The coefficients and hence the true absorbance values can be determined by a suitable calibration procedure in which the sample is replaced by a quarter-wave plate with a second polarizer behind it. The quarter-wave plate, whose fast axis is set parallel to the stress axis of the PEM, transforms the circularly polarized light back to linear, which is analyzed by the second polarizer. The calibration procedure involves measurement of two VCD spectra, one in which the analyzer is parallel to the first polarizer and the second in which it is perpendicular. Both VCD spectra are displayed in Fig. 2.

The connection of the intersection points with straight lines produces the calibration plot, which visualizes the efficiency of the circularly polarized light. Dividing the raw VCD by this curve and substituting the time constant of the lock-in yields the true VCD spectrum in absorbance units. One more adjustment is necessary for routine application of VCD. The phase correction algorithm of Mertz implemented in standard IR software for Fourier transformation can be troublesome for VCD spectroscopy since it accounts for positive peaks only. The negative peaks are often folded into the positive domain, which leads to erroneous results. This uncertainty can

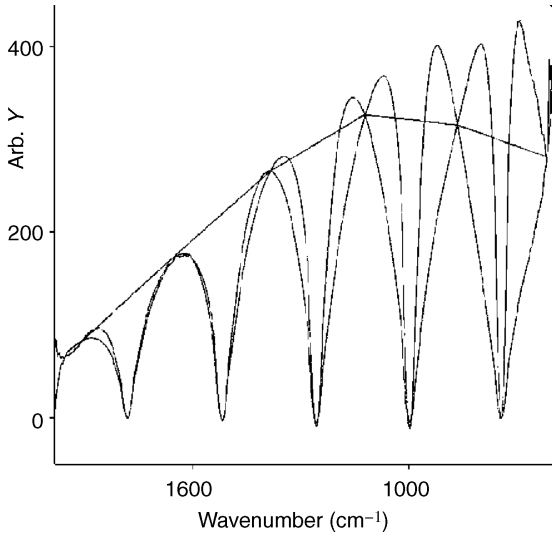


***FIGURE 2*** *Calibration spectrum. The straight lines connecting the intersecting points show the efficiency of the circularly polarized light.*

be avoided either by employing a modified phase correction algorithm (15) or by transferring the phase correction from a stressed optical plate, which produces a monosignate VCD spectrum (16).

Further advance in VCD instrumentation has been achieved by implementing the dual polarization modulation method (17). It involves introducing a second PEM directly after the sample and operating at slightly different frequency. The VCD spectrum is measured separately and in parallel at each frequency of the two PEMs. The modulated beam intensities reaching the detector are demodulated simultaneously by two lock-in amplifiers and then subtracted electronically in real time. The subtraction eliminates artifacts that are caused by birefringence since they appear to be nearly equal for both PEMs, while VCD remains unaffected. The proposed method needs precise electronic adjustment but helps to eliminate the influence of linear birefringence, which distorts the baseline of the VCD spectrum. In this way, the VCD spectra can be measured without the need for background measurement.

## 2.3 Uniqueness and Applicability of Results

All chiral molecules whose IR absorption spectra have sufficient intensity in solution should show measurable VCD. Nonetheless, there is no direct correlation between IR absorbance and VCD intensities. As VCD is sensitive to molecular conformation in solution, the molecules that possess multiple conformations of nearly even populations might show much weaker VCD signal than those with one prevailing conformation. To the last category belong bicyclic molecules similar to most terpenes, which have rigid structures due to their fused rings. α-Pinene and camphor have generally been adopted as the qualitative measure for the performance of a VCD spectrometer.

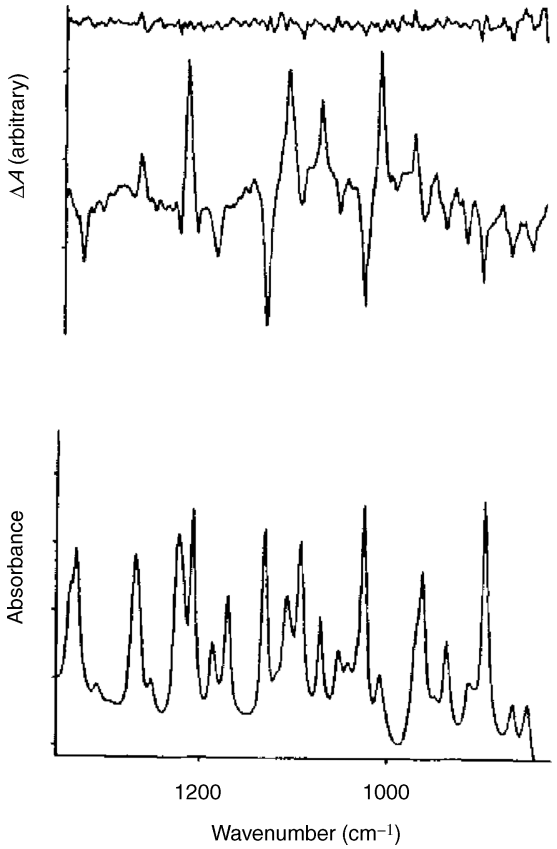


***FIGURE 3*** *Absorption (bottom) and raw VCD (top) spectra of neat (−)-α-pinene, 40% (v/v) in $CCl_4$ in 0.15 mm NaCl cell. The VCD spectrum is a result of 500 ac and 25 dc scans collected on a modified Bomem MB100 spectrometer, 4 $cm^{-1}$ resolution, and 0.4 mm/s mirror speed (18).*

Nonetheless, α-pinene is more frequently used because it also enables sampling as a neat liquid. Both molecules show strong VCD signals and yield good S/N ratios at relatively short measurement times. The VCD spectrum of (−)-α-pinene together with the corresponding absorption spectrum is presented in Fig. 3 (18).

The raw VCD spectrum is the result of 500 ac scans and 25 dc scans, which comprises about 10 min data acquisition time. Since the VCD band intensities are much smaller than those of absorptions, and consequently the VCD intensity scale is much expanded, the subtle baseline distortions due to optical imperfections may become visible (Fig. 3). As a result, the baseline of the VCD spectrum deviates from its zero value and needs to be corrected. The usual way to make this correction is to measure the VCD spectrum of the opposite enantiomer taken at precisely the same experimental conditions. The subtraction of the two spectra and the division of the result by 2 yields the enantiomer corrected spectrum. This subtraction effectively eliminates the noise, which has a random character, while the VCD bands remain

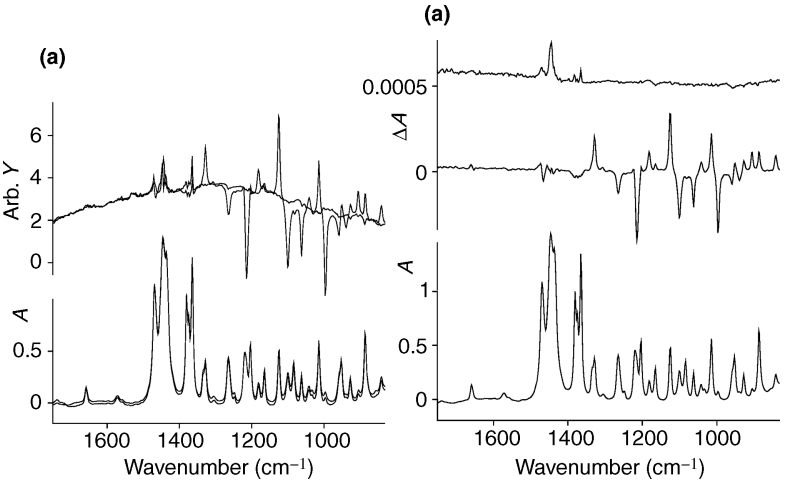


***FIGURE 4*** *(a) IR absorption and VCD spectra of (+)-α-pinene and racemic (±)-α-pinene. (b) IR absorption (bottom), enantiomer corrected VCD (middle), and noise estimate (top) of neat (+)-α-pinene. Both VCD spectra were acquired with 20 min data acquisition time (18).*

unchanged. The opposite mathematical operation on the two spectra determines the VCD zero line and is referred to as noise estimate. It should represent the "VCD" spectrum without the sample in the instrument's optical train. The same result has to emerge when the sample is a racemic compound or simply the spectral solvent. Examples of raw VCD spectra of (+)-α-pinene and racemic (±)-α-pinene are displayed in Fig. 4 (18).

Some deviations persist in the baseline of racemic α-pinene, where the strongest absorption bands occur. These deviations cannot be compensated for also in the noise estimate of (+)-(α-pinene (Fig. 4). These distortions occur noticeably in the wavenumber range where the photoelastic modulator is less efficient and consequently the VCD bands are less intense. The main sources for the baseline distortions are polarization and absorption artifacts. The first of these two reasons is inherent for each particular spectrometer and the optics used in the VCD setup. The polarization artifacts appear as weak "VCD" bands typically in the spectral range where the sample does not absorb. The artifacts arise from optical arrangements and some imperfections in the optical elements. Particularly with our instrument, the imperfections were due to reflections from the internal ledges of the sample cell holder as well as residual stress and birefringence left in the detector window. It is known that reflections from metal surfaces as well as passing through birefringent material may cause phase shifts of the incident polarized radiation. While the reflections from the sample holder can be avoided either by inserting an aperture in front of the cell or by using mild focusing of the IR beam, the second reason is difficult to circumvent. The residual stress left in the crystal can depend on the manufacturing process and on the type of the crystal. Baseline distortions due to

residual birefringence can be partially compensated for by repositioning the optical filter in front of the detector or by rotation of the ZnSe lens.

## 3 APPLICATIONS

### 3.1 Determination of Enantiomeric Purity and Enantiomeric Excess

The abundance of one enantiomer over another in a chemical substance is known as enantiomeric excess (ee). Enantiomeric excess is defined by the respective concentrations of both enantiomers according to Eq. 3:

$$\mathrm{ee} = \frac{c_R - c_S}{c_R + c_S} \times 100\%, \tag{3}$$

where $c_R$ is the concentration of the $R$ enantiomer and $c_S$ is the concentration of the $S$ enantiomer. This equation can be used for ee assay provided the amount of each enantiomer produced is known. Mixing two enantiomers will show net optical rotation unless it is a racemate. Since the optical rotation is proportional to the concentrations of each enantiomer, it can be used for assaying enantiomeric purity according to Eq. 4:

$$\mathrm{ee} = \frac{[\alpha]_\lambda^T}{[\alpha]_{max}} \times 100\%, \tag{4}$$

where

$$[\alpha]_\lambda^T = \alpha / lc \times 100 \tag{5}$$

is the specific optical rotation of the sample at a given wavelength $\lambda$ and temperature $T$. $\alpha$ is the measured angle of optical rotation, $l$ is the thickness of the polarimetric cell, and $c$ is the solute concentration; $[\alpha]_{max}$ is the specific optical rotation of the pure enantiomer measured at exactly the same conditions as $[\alpha]_{max}$ of the sample.

Historically, this is the oldest method for enantiomeric purity assay and is routinely applied for ee evaluations. Some shortcomings are typical for this method. They are related to difficulties with reproducibility because optical rotations are sensitive to concentration, solvent, temperature, and traces of impurities with high optical rotation. Furthermore, polarimetry cannot accurately and reliably establish the enantiomeric purity of a sample unless the optical rotation values are calibrated by an independent method.

NMR spectroscopy with chiral shift reagents is one of the simplest and most accessible techniques available for determining enantiomeric purity. The method consists of adding to the sample a chiral lanthanide reagent, which shifts some of the NMR bands. If the corresponding peaks of each enantiomer can be resolved, the

percentage ratio of the enantiomers can be measured as relative peak heights or areas. Although routinely applied, NMR with shift reagents has limited applicability for analyzing enantiomeric purity. Among the drawbacks are finding suitable shift reagents. In addition, difficulties may arise when substances with several chiral centers are to be investigated.

The most frequently applied technique for enantiomeric excess is high-performance liquid chromatography (HPLC) using chiral columns. Enantiomers are affected differently by the chiral substrate in the column and appear at different retention times. The ratio of the band areas of the separated enantiomers is the basis for the ee evaluation with accuracy reaching 0.1% ee when complete separation is possible (19). The chiral columns are, however, expensive and sometimes inefficient when complex mixtures are to be separated. The chiral stationary phases are specific to chemical functionalities that require maintenance of several types of columns for relatively small ranges of optically active compounds. Combining HPLC with UV and/or chiroptical detection can improve the quantitation and detection limits in unfavorable cases when complete separation is not possible (19). Such a combination is particularly useful for online monitoring of asymmetric syntheses when the reaction steps can result in racemization of the product. Analyses of reaction mixtures are usually maintained by reverse-phase HPLC. The combined use of UV and CD detectors can help identify impurities that are not chiral or detectable by CD detectors, but the extent of racemization cannot be determined (19). In addition, chiroptical detection raises the detection limits substantially for mixed chiral samples when full separation is not possible.

In comparison, VCD is a simple and nondestructive method for assaying enantiomeric purity. There is no need to physically separate the enantiomers as is required for chiral chromatography, nor are any foreign additives needed such as chiral shift reagents. According to Beer's law, VCD intensities of a single enantiomer are linearly concentration related. As the VCD bands of both enantiomers are but mirror images, they have the same peak positions and intensity but opposite sign. When both enantiomers are present in a mixture, the VCD band intensities decrease in direct proportion with the relative concentration of each enantiomer. Thus, ee assays can be performed when an appropriate calibration plot using single enantiomers is created previously. The availability of multiple transitions in VCD spectra ensures more accurate calibration than the use of a single chromophore band. Because of lower signal-to-noise ratio, the accuracy of the method is limited to about 1% ee obtained to date. Further improvement of accuracy can be expected after development of appropriate chemometric software. The applicability of VCD for determining ee is illustrated for (+)-α-pinene in Fig. 5 (19) and represents the overlaid plot of 11 VCD spectra of (+)-α-pinene with the ee percentage ranging between 100% and 0% in 11 steps. The relationship between the real and predicted ee values is plotted linearly in Fig. 6 (19). Partial least squares (PLS) have been used to account for the entire array of points across the spectrum. This approach results in better correlation between the actual and predicted ee values.

The capabilities of VCD and optical rotation for determining enantiomeric purity were also discussed by Zhao and Polavarapu (20). The advantages of VCD

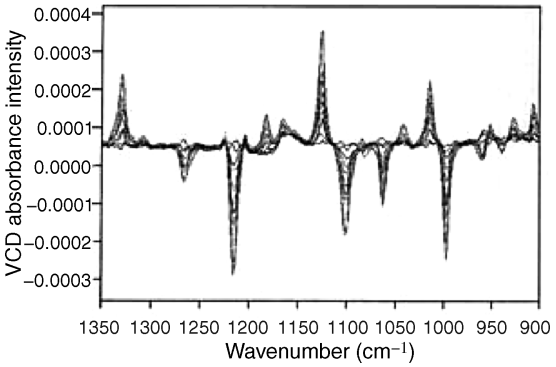


***FIGURE 5*** *Superposition of the VCD spectra of 11 samples of (+)-α-pinene in the mid-IR region (19)* (copyright PharmaPress Ltd.).

are indisputable when chiral mixtures are to be studied because the individual components can be monitored separately. Further advance in VCD for determining % ee was achieved by constructing a flow cell sampling accessory (21) with which one can monitor simultaneously and with time any changes in fractional composition of a mixture of two different chiral molecules. The VCD set of data was processed by the PLS plus/IQ module in Grams/32 AI (6.0) (Galactic Industries Inc., Salem, NH) because it makes it possible to apply the partial least squares chemometric analysis to determine the percentage ee from VCD spectral data. The proposed sampling accessory can be applied to either single species or a chiral mixture. The accuracy of % ee determination is about 1% for single species and about 2% when two components in a mixture have to be monitored simultaneously (21).

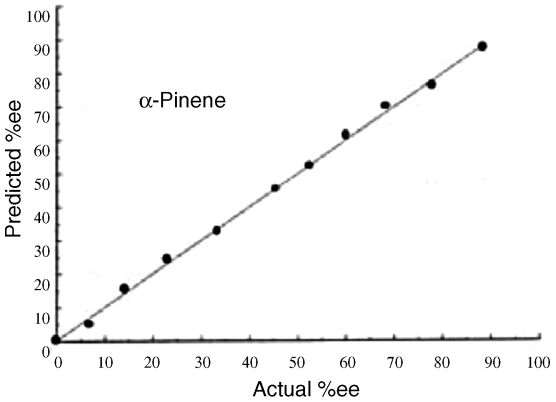


***FIGURE 6*** *Linear plot of prepared VCD % ee versus predicted VCD % ee value (19)* (copyright PharmaPress Ltd.).

## 3.2 Absolute Configuration

The three-dimensional arrangement of atoms around the chiral center of a molecule is known as absolute configuration. Direct access to determine absolute configuration can be assured only by Bijvoet analysis of single-crystal X-ray diffraction or chiroptical methods such as ECD and VCD. X-ray analysis requires a single crystal of a compound of sufficient size and the presence of a "heavy" atom beyond the first row, such as S or Cl. The atom absorbs X-rays to an appreciable extent so that a phase change occurs for the X-ray that is scattered by that atom relative to the other atoms in the structure. This anomalous scattering causes different intensities of reflection that, compared to the calculated intensities for the accepted model, indicate whether the model represents the correct configuration.

NMR spectroscopy is an alternative method for assaying absolute configuration in solution (22). Useful as the method may be in many instances, it is limited for cases in which derivatizing substrates with a chiral substitute are needed and known configurations are required (22). Determining directly the absolute configuration of underivatized chiral compounds is still a challenge. Some progress has been made by using NMR shift reagents; however, these combine the coordination ability of lanthanide ions with their paramagnetic properties (23). In general, routine applicability still remains elusive because finding suitable shift reagents can be difficult.

Other chiroptical methods, such as electronic CD and quite recently optical rotation, could also serve in principle as direct means for determining absolute configuration. For electronic CD, these assessments are based either on one or on a few bands available in the spectral range of interest. The theoretical interpretation depends on the development of computational quantum mechanical methods for both the ground and the excited electronic states. Recent developments of *ab initio* time-dependent density functional theory (TDDFT) for calculating optical rotation in absorption-free spectral regions of chiral molecules made it possible to use optical rotation measurements alone or in combination with other chiroptical methods for determining absolute configurations (24,25).

The other method capable of solving absolute configurations in solution and that requires no prior knowledge for a precursory structure is VCD. It is based on comparing experimental VCD spectra with those computed theoretically either by *ab initio* means or by the more frequently used DFT methods. The theoretical spectrum is generated based on a definitive spatial arrangement of the atoms around the chiral center and an assumed conformation. When an agreement between the band signs of the theoretical and the experimental VCD spectra is reached, the absolute configuration chosen as an input for the calculations is the same as that of the sample. An opposite sign of a calculated band compared to that determined experimentally testifies unambiguously that the opposite enantiomer was selected for calculation. Without calculations, experimentally measured VCD spectra can predict empirically the absolute configuration of chiral molecules provided reference to a structure close to a known absolute configuration exists or a correlation between VCD sign and absolute configuration obeys established chiral rules. Some limitations currently exist. They are imposed either by the VCD experiment or by theoretical treatment.

Sample molecules must show measurable VCD spectra with satisfactory signal-to-noise ratio. Molecules must be reasonably soluble in the usual spectral solvents and show distinct VCD bands. The rigidity of the molecular skeleton, which reduces the number of conformations, usually contributes to intense VCD bands. Conversely, higher molecular flexibility resulting from multiple conformations often leads to weak VCD and poor signal-to-noise ratio. Conformations may also raise the complexity of calculations since every single conformation, which contributes to the overall conformational equilibrium, must be accounted for. The main limitation related to the computations is usually the size of the targeted molecule. Since most computations today are performed on personal computers, there is a practical limit of up to 30 heavy atoms such as carbon, nitrogen, and oxygen. As already mentioned, another reason is the existence of an equilibrium among evenly populated multiple conformations. A relatively small number of conformations could be easily handled, but multiple conformation equilibria effectively increase CPU time. These limitations, however, become less important as the power of personal computers increases.

Solving molecular stereochemistry is straightforward when a chiral structure has a single predominant conformation. For instance, the absolute configuration of 1-azabicyclo[3.1.0]hexane, which is the principal structural fragment of the antitumor antibiotics azinomycins A and B, has been determined by VCD spectroscopy (26). This bridgehead molecule has two chiral centers in its aziridine ring. Geometry optimization and VCD calculations were carried out using RHF/6-31G*(0.3), Becke3LYP/6-31G*, and MP2/6-31G* methods. The optimization by each of these methods yielded only the boat conformation (Fig. 7). The experimental and simulated IR and VCD spectra in the 1500–800 $cm^{-1}$ range are displayed in Fig. 8.

The level of agreement among virtually all VCD bands is very high. There is but one obvious discrepancy observed for the band at 1240 $cm^{-1}$ for which only RHF/6-31G*(0.3) predicted the correct sign. The band cluster at 1300–1150 $cm^{-1}$ is a readily observable signature of unambiguously identifiable enantiomers. The predictions of the three theoretical methods agree well with the experimental features yielding a definitive assignment of the absolute configuration (1*R*,5*S*)-(–)-1-azabicyclo[3.1.0] hexane.

## 3.3 Solution Conformations

Both IR and VCD spectroscopy originate from $3N - 6$ independent vibrations of the molecule, where the frequencies of the molecular transitions range between $10^{12}$ and $10^{14}$ Hz. The timescale of vibrational transitions makes it possible to resolve molecular motions reaching femtosecond limits. This rapid spectral response makes both IR and VCD spectroscopy incisive tools for studying molecular conformations in solution.

The experimental VCD spectrum represents the conformational equilibrium under the conditions chosen for the experiment. When multiple conformations exist, each conformer contributes to the VCD spectrum in accordance with its percentage population. In this case, a conformational search must be performed first leading to structures, relative free energies, and equilibrium populations of the conformers. The theoretical treatment requires that a separate calculation for every single

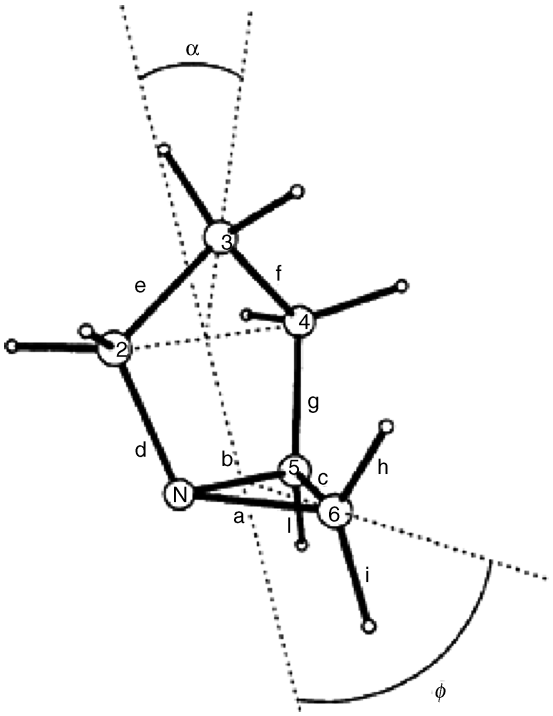


***FIGURE 7*** *Optimized structure of (1S,5R)-azabicyclo[3.1.0]hexane (26).*

conformer be carried out. Then, the simulated VCD spectrum can be constructed by summing the contributions of the separate conformers weighted by the fractional populations to yield the conformationally averaged composite spectrum. A close match of the experimental and the simulated spectra yields not only the absolute configuration but also the distribution of the solution-phase conformation.

As an illustration, it is useful to demonstrate how (+)- and (−)-enantiomers of *cis*, *cis*-spiro[4,4]-nonane-1,6-diol can be identified as 1*S*,5*S*,6*S*-(+) and 1*R*,5*R*,6*R*-(−), and how their conformation distribution is determined (27). The VCD spectra were obtained with a Bomem MB100 interferometer equipped with VCD optics and described in detail elsewhere (18). The *ab initio* calculations were carried out at the SCF-RHF/6-31G*(0.3) level using the *Gaussian 90* program package. The conformational analysis disclosed six conformers with close energies differing in the ring conformations and the orientation of the hydroxyl groups. The first two conformers separated by 0.3 kJ/mol constituted 86% of the whole population. They were identified to differ in ring conformations and were seemingly intramolecularly hydrogen bonded. The next two conformers with about 5 kJ/mol greater energy contributed another 7% and the last two highest energy conformers displaced by 3 kJ/mol gave the remaining 7% of the conformation distribution. The simulated VCD spectra generated with vibronic coupling theory at the distributed origin gauge are presented in Fig. 9. The first two spectra (a and b) represent the first two conformers of

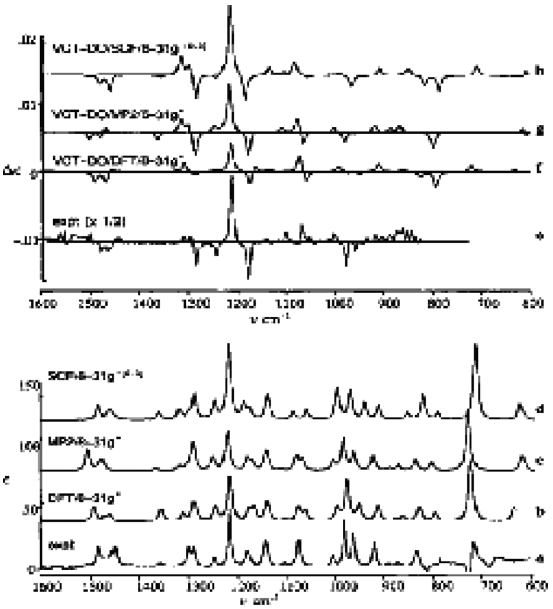

***FIGURE 8*** *Experimental (a, e) and theoretical (b–d, f–h) IR (a–d) and VCD (e–h) spectra of (1R,5S)-1-azabicyclo[3.1.0]hexane: (f) geometry, force field, and atomic polar tensors at Becke3LYP level and atomic axial tensors at SCF level; (g) geometry, force field, atomic polar tensors, and atomic axial tensors at MP2 level; (h) geometry, force field, atomic polar tensors, and atomic axial tensors at SCF level (26).*

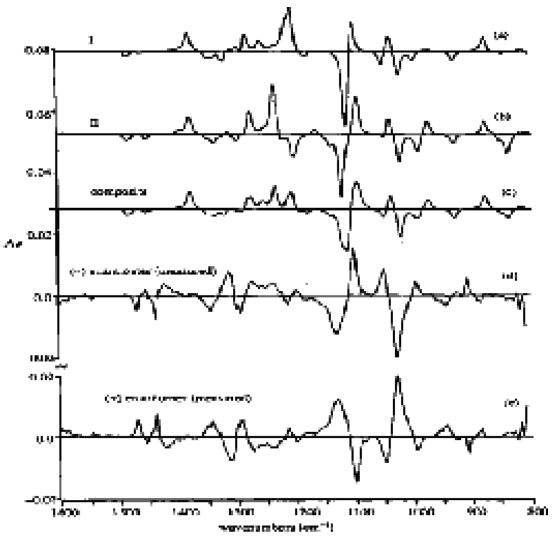

***FIGURE 9*** *Calculated VCD spectra of (−)-(1R,5R,6R)-spiro[4,4]nonane-1,6-diol: 6-31G*(0.3) geometry, force field, atomic polar tensors, and 6-31G atomic axial tensors (a–c). Experimental VCD spectra of both enantiomers: (d) (−)-isomer,* $[\alpha]_D^{22} = -101.4$*, 100% ee; (e) (+)-isomer,* $[\alpha]_D^{22} = +97.1$*, 100% ee (27).*

(−)-(1*R*,5*R*,6*R*)-spiro[4,4]nonane-1,6-diol, whereas the next simulated VCD spectrum (c) is a composite of the six conformers. These are compared with the experimental VCD spectra of the two enantiomers presented in the same figure (d and e). The most intense and characteristic band combination is located between 1150 and 1000 $cm^{-1}$ and appears as negative–positive and positive–negative combinations of bands unambiguously associated with the (−)-enantiomer. The theoretical composite spectrum closely matches the experimental VCD spectrum in almost all features suggesting a close relationship between the actual and assumed conformation distribution. Unlike the absorption where hydrogen bonding complicates the spectrum to such an extent that it seems poorly reproduced, hydrogen bonding in VCD seems not to interfere with these identifications.

For cyclic molecules for which free rotation around simple bonds cannot occur and only a limited number of dihedral angles for internal rotation are available, the most dependable way to find the conformation distribution is to perform potential energy surface (PES) scans by the DFT-PES method. Further optimization of the lowest energy conformers leads to the equilibrium structures of these conformations. This approach was demonstrated by Devlin et al. on a series of the chiral cyclic sulfoxides including 1-thiochroman-4-one *S*-oxide, 1-thiaindan *S*-oxide, and 1-thiochroman *S*-oxide (28). Nearly perfect agreement of the predicted and the experimental spectra was obtained leading to unequivocal assignments of the absolute configurations of these molecules.

The DFT-PES calculation is ineffective for molecules that possess extra flexibility. Finding the distribution of their conformation requires another approach. It is usually carried out in two steps. First, a conformational search using a molecular mechanics force field finds the lowest energy conformers. Their geometries are further optimized by DFT and the VCD spectrum of each conformer is computed. The conformationally averaged VCD spectrum is then compared with the experimental one to provide the assignment of the absolute configuration. The application of this approach is exemplified by solving the absolute configuration of an oxadiazol-3-one *calcium channel blocker* (29). The authors took advantage of a relatively new theoretical development for which the results relied on the simultaneous use of vibrational circular dichroism, electronic circular dichroism, and optical rotation.

The foregoing examples already demonstrate the predictive power of VCD spectroscopy in conjunction with *ab initio* calculations, and more recently in connection with other chiroptical mechanisms. The comparison of the observed and calculated spectra is an effective means for determining the absolute stereochemistry and the conformational distribution. The close agreement of experiment and theory is gratifying and confirms that VCD is a reliable tool for stereochemical analyses.

## 3.4 Pharmaceutical Molecules

The determination of absolute stereochemistry of pharmaceutical molecules is central for their use as drugs given the different biological activities of opposite stereoisomers. The increasing demand posed by the growing interest of pharmaceutical industry requires stereochemical characterizations of relatively large molecules

whereas the level of theoretical models and computational capabilities still remain reasonably modest, although considerable progress is evident. However, and as noted in the preceding section, a different approach emerged in which the interpretation of VCD spectra is based on calculating single fragments of chiral molecules. Since the achiral portion of the molecule in general does not contribute significantly to VCD intensities, it does practically not interfere in particular spectral ranges, such as C–H stretching region for instance. As an example, this approach has been tested on a number of antiarrhythmic and anti-inflammatory drugs depicted in Fig. 10 (30). Diverse VCD markers were established for the molecules studied, which included NH and OH vibrations for free and intramolecularly bonded structures. The appropriate choice of the molecular models allowed the authors to find the most abundant conformers, determine the conformational distribution, and solve the absolute configuration.

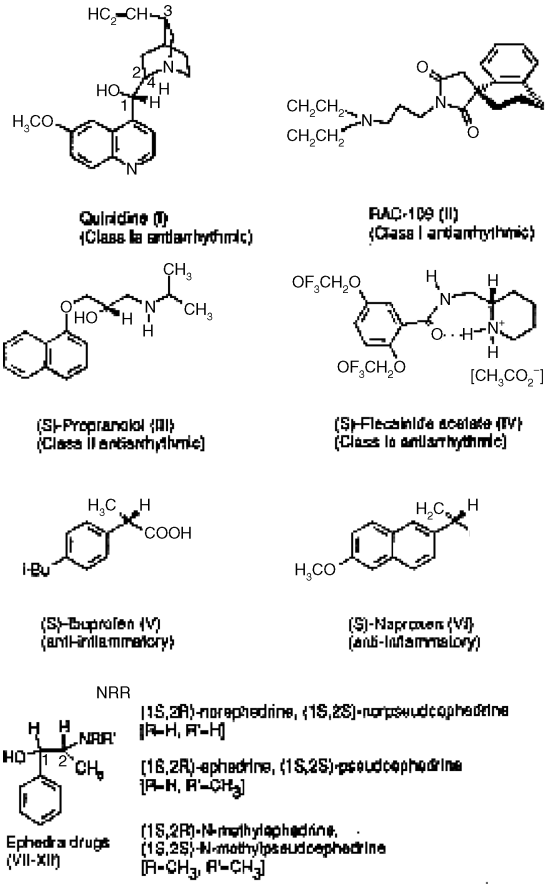


***FIGURE 10*** *Molecular structure of antiarrhythmic, anti-inflammatory, and Ephedra drugs (30)* (copyright Marcel Dekker, Inc.).

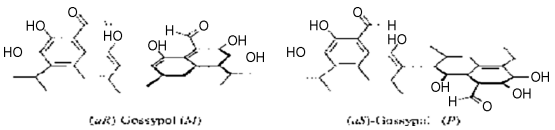


***FIGURE 11*** *Structures of (M)- and (P)-gossypol (5)* (copyright Wiley-Liss, Inc.).

Historically, it is interesting to note that an absolute configuration was determined of a plant product known as gossypol (5). This compound is a polyphenolic binaphthyl dialdehyde (Fig. 11) for which the hindered rotation around the binaphthyl bond results in axial chirality and consequently in helical arrangement of different handedness M or P. Gossypol itself has attracted considerable interest because, besides its contraceptive activity for males (31), it was also tested as an agent against different human malignancies (32). The absolute configurations of gossypol enantiomers were first established by Huang et al. using electronic circular dichroism when assigning the M configuration to the (−)-enantiomer (33). From among many and varied investigations, one learns that (−)-gossypol's chirality can overcome resistance to apoptosis and may therefore turn out to be a "promising new agent to treat malignancies that are resistant to conventional therapies" (34).

Freedman et al. achieved the first definitive assignment for gossypol by calculating optimized geometries, force fields, and infrared and VCD intensities using *Gaussian 98* at the DFT/B3LYP level and 6-31G* basis set (Fig. 12) (5). Although the binaphthyl skeleton is rather rigid, a molecule can evidently adopt several conformations including different intramolecular hydrogen bonding and various orientations of the aldehyde and isopropyl groups. A conformational search discriminated among three lowest energy conformations that exist at one and the same type of intramolecular hydrogen bonding and differ in relative orientations of the isopropyl group. While the single conformers were separated by about 1 kcal/mol from each other, the relative population was found to be 70:28:2 in favor of the most stable conformer. Precise geometry optimization established a value of 79.1° for the dihedral angle between the naphthalene rings. The conformationally averaged infrared and VCD spectra of (+)- and (−)-gossypol were constructed by assuming Boltzmann distribution of the separate populations together with the experimental spectra. The very close agreement between virtually all spectral features testifies that the P configuration previously assumed for computation can be associated unambiguously with (+)-gossypol. Therefore, it is *the* physiologically active substance and will have M configuration as suggested 20 years ago by applying coupled oscillator theory to its CD (33). This seminal example further demonstrates the appreciable potential of VCD for assigning absolute configurations of relatively large molecules. Without doubt, the combination of VCD with DFT calculations turned out to be eminently capable of predicting the hydrogen bonding pattern and conformation distribution in solution (5).

Although X-ray crystallography has been remarkably useful for resolving absolute configurations, VCD has also become an indispensable method for stereochemical analysis providing additional information when liquid materials are involved. For

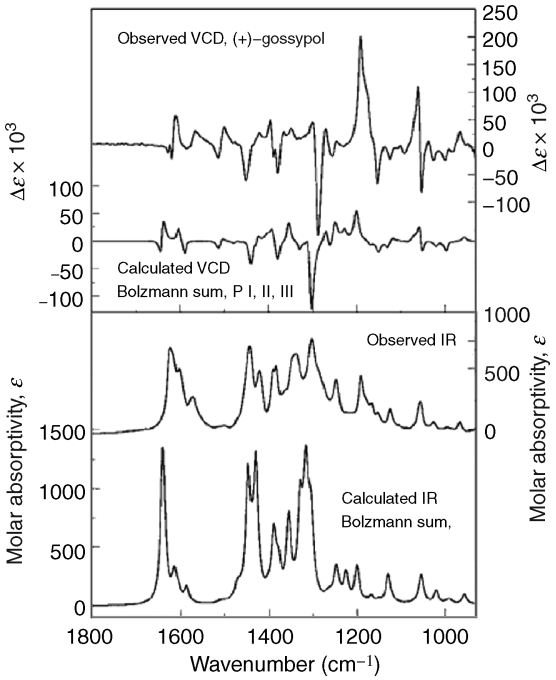


***FIGURE 12*** *Infrared and vibrational circular dichroism spectra of gossypol (5)* (copyright Wiley-Liss, Inc.).

example, when attempting to identify new nonpeptide vasopressin receptor antagonists and after the X-ray methods failed, the absolute configuration of a key physiologically active chiral component, such as benzodiazepine, was determined by VCD spectroscopy (Fig. 13) (35). Infrared absorption and VCD spectra were recorded in $CDCl_3$ in the spectral range 1700–800 $cm^{-1}$ on a Chiral*IR* instrument (ABB Bomem/BioTools). Theoretical calculations were carried out at the DFT B3LYP/6-31G* level using *Gaussian 98*. Three distinct conformers were identified by geometry optimization differing in the conformations of the six- and seven-membered rings. Separate calculations of absorption and VCD intensities of every

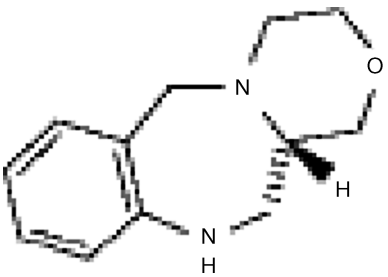


***FIGURE 13*** *Structure of (+)-chiral benzodiazepine (35)* (copyright Wiley-Liss, Inc.).

single conformer have been necessary in order to associate the observed VCD spectrum of the (+)-enantiomer with the true absolute *S* configuration of the asymmetric carbon (35).

Full agreement was achieved for (+)-McN-5652-X in the absolute configuration between prior single-crystal X-ray diffraction and recent VCD measurements (36). This compound belongs to a novel series of pyrrolo[2,1-*a*]isoquinolines considered to be selective inhibitors of serotonin reuptake *in vivo* and therefore possessing potential antidepressant activity. The single-crystal X-ray diffraction of (+)-McN-5652-X salts with $HClO_4$ and (+)-(2*R*,3*R*)-tartaric acid provided evidence for the (6*S*,10b*R*) absolute configuration and relative *trans* configuration between $C_6$ and $C_{10b}$ (Fig. 14). A VCD study in $CDCl_3$ confirmed the previous X-ray crystallographic assignment and identified the highest populated solution conformation as a transfused 5,6 ring junction in which the nitrogen configuration is *R* (36).

Dunmire et al. examined the VCD spectra of ketoconazole, itraconazole, and miconazole, three chiral drugs that are used as antifungal agents (37). These VCD studies confirmed the previous assignment of the absolute configurations, which were inferred solely on the basis of syntheses. In order to avoid DFT calculations of (+)-ketoconazole and (+)-itraconazole, which would have been necessary for

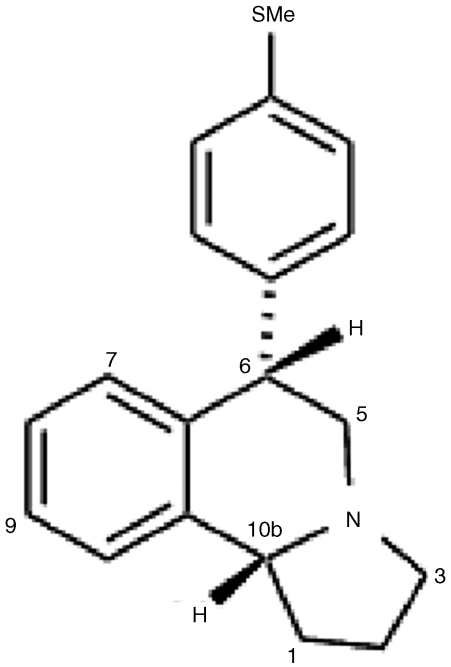


*trans*-(6*S*, 10b*R*)

***FIGURE 14*** *Structure of (+)-McN-5652-X (36)* (copyright Elsevier Science Ltd.).

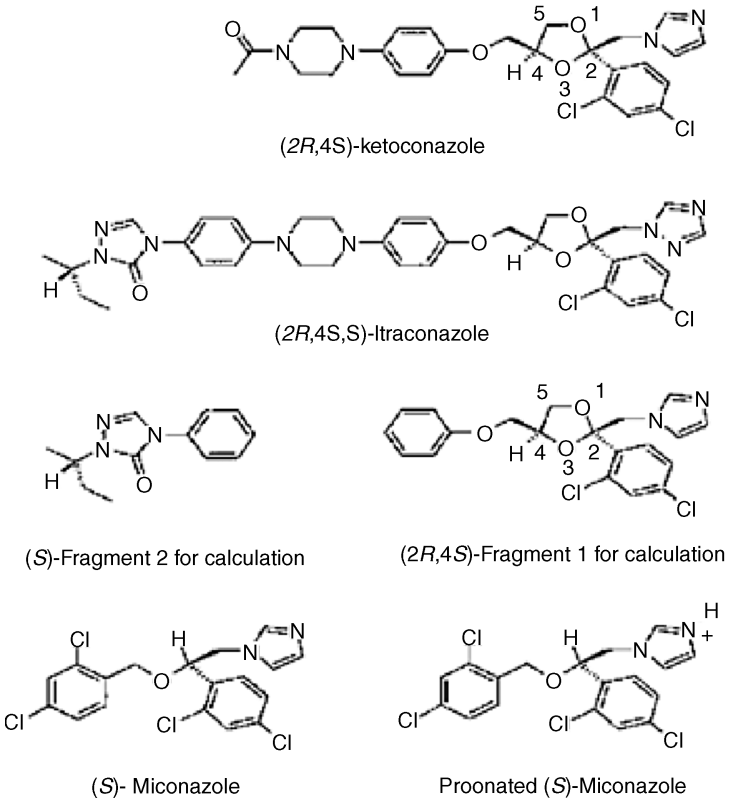


***FIGURE 15*** *Structures of (2R,4S)-ketoconazole, (2R,4S)-itraconazole, and (S)-miconazole and fragments 1 and 2 used for calculations (37)* (copyright Wiley-Liss, Inc.).

identifying solution conformations and establishing the absolute configurations, a fragment common for the two molecules was used (Fig. 15). This fragment contains two chiral centers, which should have been sufficient for identifying features to compare with the VCD experiment. The contribution of the third chiral center of (+)-itraconazole was accounted for by calculating the second fragment. The VCD spectrum of (*S*)-miconazole was simulated by computing the geometry, force field, and VCD intensity of the whole molecule and its protonated form. By comparing the composite spectra constructed from the most abundant conformation populations and the experimental VCD spectra, the absolute configuration was confirmed for (+)-(2*R*,4*S*)-ketoconazole, (+)-(2*R*,4*S*)-itraconazole, and (+)-(*S*)-miconazole.

Most of the chiral molecules studied conveniently have isolated vibrational modes in the VCD spectrum in which the chiral center is involved. In some instances, these modes are either hidden in the fingerprint region or obscured by solvent absorption rendering them unsuitable for stereochemical assignments. To determine by VCD the absolute configuration of the cruciferous phytoalexin metabolites (−)-dioxybrassinin and (−)-3-cyanomethyl-3-hydroxyoxindole, Monde et al. had to use DMSO-$d_6$,

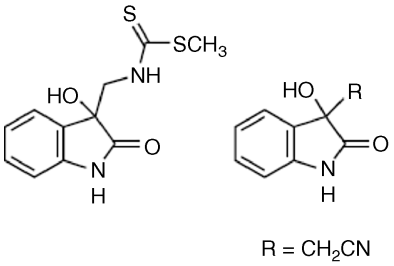


***FIGURE 16*** *Structures of cruciferous phytoalexin-related metabolites (−)-dioxybrassinin and (−)-3-cyanomethyl-3-hydroxyoxindole (38)* (copyright Elsevier Ltd.).

which required a very thin (6 μm) $CaF_2$ cell, thereby substantially reducing the VCD window. Only a few characteristic bands located at 1730, 1620, and 1470 $cm^{-1}$ were available of which the first two were readily assigned to C=O stretch and aromatic ring stretch. As indicated earlier, this *concerted* mechanism combined the calculations for VCD with ECD followed by optical rotation (38). In this manner, the absolute configurations of (−)-dioxybrassinin and (−)-3-cyanomethyl-3-hydroxyoxindole were unequivocally determined. This was the first attempt to determine absolute configurations of tertiary alcohols by VCD. The approach could well be generalized for other stereochemical assignments of tertiary alcohols (Fig. 16). Noteworthy in this particular case was that their absolute configurations were determined by involving vibrational CD together with the more conventional electronic CD.

Earlier in the foregoing text, oxadiazol-3-one was referred to as a "calcium channel blocker" (29). This designation was originally proposed by B.J. Materson in 1995 to avoid confusing designations with related drugs (39). Since then, a variety of calcium channel blockers were scrutinized in order to test a series of substances, explicitly [1,4]thiazino[3,4-*c*][1,2,4]oxadiazol-1-one derivatives (40) and a new class of calcium channel modulators (41). In a more recent publication, these authors speak of a multidisciplinary project that "has led to the discovery of novel, structurally diverse, L-type calcium entry blockers (CEBs)." The absolute configuration of a recently reported CEB has been determined by vibrational circular dichroism spectroscopy to assign the stereospecificity of the ligand–channel interaction (42). From a "purchasable" database, they screened 340,000 molecules and tested 20 compounds *in vitro* obtaining several potentially promising calcium entry blockers. As mentioned earlier, the absolute configuration of the related compound 8-(4-Br-phenyl)-8-ethoxy-5-methyl-8*H*-[1,4]thiazino[3,4-*c*][1,2,4]oxadiazol-3-one (Fig. 17) (29) also possesses the proven activity of a calcium channel blocker as determined with VCD, ECD, optical rotation, and accompanying DFT calculations. Oxadiazol-3-one is rather flexible with multiple conformations arising from the nonplanar thiazino ring as well as from free rotations of the ethoxy and phenyl groups. A conformational search by MMFF94 force field detected a large number of conformations, which when reoptimized by B3LYP/6-31G* led to 13 stable conformers within 8 kcal/mol. More accurate DFT calculations using the larger basis set TZ2P rejected most of the conformations but kept the three lowest energy conformers a, b, and c (Fig. 18). The

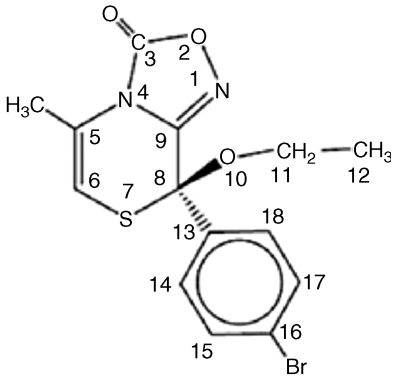


***FIGURE 17*** *Structure of 8-(4-Br-phenyl)-8-ethoxy-5-methyl-8H-[1,4]thiazino[3,4-c][1,2,4]oxadiazol-3-one (29)* (copyright American Chemical Society).

conformers a and b are separated by only 0.78 kcal/mol but consist of 97.5% of the whole population at room temperature. Conformer a alone consists of more than 80%. The very thorough analysis of VCD spectra of the oxadiazol-3-one established definitively that the absolute configuration of the (+)-enantiomer is *S*, and the (−)-enantiomer having higher physiological activity has *R* configuration (29).

## 3.5 Natural Products

The absolute stereochemistry of a series of natural products was investigated extensively by the group of Philip Stephens who reported the "first concerted application of density functional theory calculations," which included VCD, ECD,

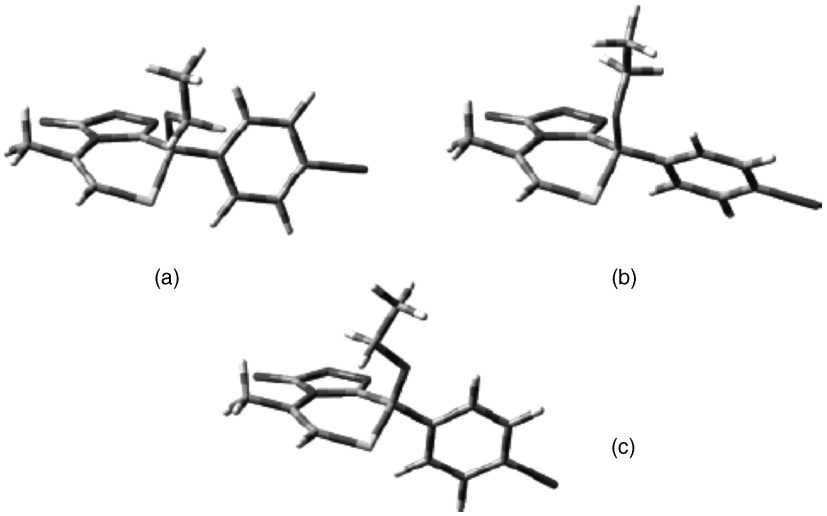


***FIGURE 18*** *B3LYP/6-31G* structures of conformations a, b, and c of (S)-oxadiazol-3-one (29)* (copyright American Chemical Society).

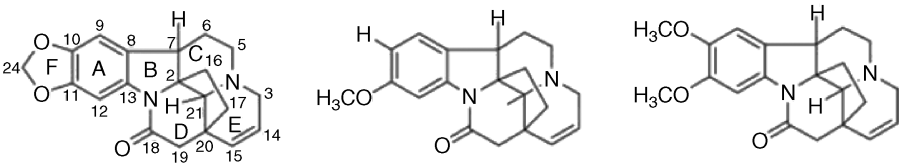


***FIGURE 19*** *Structures and absolute configurations of the schizozygane alkaloid schizozygine (43) and the isoschizozygane alkaloids isoschizogaline and isoschizogamine (44)* (copyright American Chemical Society).

and optical rotation spectra. Among the compounds, they first investigated the schizozygane alkaloid schizozygine (43) and the isoschizozygane alkaloids isoschizogaline and isoschizogamine (Fig. 19) (44). All three of these alkaloids were isolated from *Schizozygia caffaeoides*, a plant traditionally used in Kenyan folk medicine. Although the structures of these alkaloids were determined by NMR, their absolute configurations were still unclear until very recently. By combining chiroptical methods and high-level DFT calculations, Stephens et al. were able to assign unambiguously the absolute configurations of the alkaloid (+)-schizozygine, which has (2*R*,7*S*,20*S*,21*S*) configuration (43), as do both (−)-isoschizogaline and (−)-isoschizogamine, which have one and the same configuration, namely, (2*R*,7*R*,20*S*,21*S*) (44).

Equally compelling are the assignments for plumericin and isoplumericin iridoids. Plumericin and its isomer isoplumericin, first isolated from *Plumeria multiflora* and *Plumeria rubra* var. *alba*, were found to have antifungal and antibacterial activity. The structures were established originally by Albers-Schönberg and Schmid in 1961 (45), although some doubt remained whether these assignments were correct. Consequently, a comprehensive study of the above-mentioned iridoids was performed very recently bringing to bear VCD, commensurate density functional calculations, electronic CD and optical rotation (46) (Fig. 20). A conformational search revealed four stable conformations within $\Delta E = 2$ kcal/mol. The refined geometry of the lowest

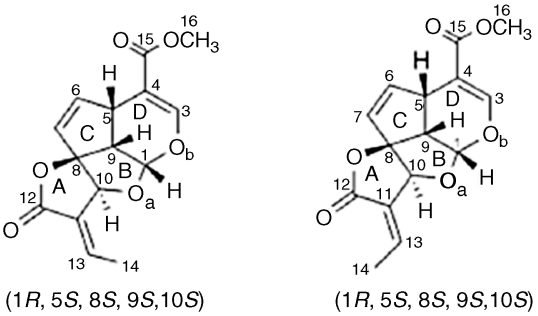


***FIGURE 20*** *Structures of plumericin and its isomer isoplumericin (46)* (copyright American Chemical Society).

energy conformers at the B3PW91/TZ2P level of theory led to conformationally averaged infrared and VCD spectra, which were found to be equivalent to those observed. It was therefore concluded definitively that the absolute configuration of both (+)-plumericin and (+)-isoplumericin is (1*R*,5*S*,8*S*,9*S*,10*S*), confirming the previous assignments of Albers-Schönberg and Schmid (45). Two years earlier, a different absolute configuration was proposed (45) suggesting that the absolute configuration of plumericin and isoplumericin should be (1*S*,5*R*,8*R*,9*R*,10*R*), that is, the reverse of that identified by density functional theory and experimental spectroscopic evidence. Stephens et al. (44) evidently provided the correct interpretation and also demonstrated that the method is more reliable based partially on the combination of VCD, TDDFT, electronic CD, and optical rotation (46).

The structure of a *new* and highly cytotoxic iridoid, isolated from prismatomerin (*Prismatomeris tetrandra*), was found in which a methyl group is substituted by a *para*-phenol group (Fig. 21) (47) and was very similar to the previously studied plumericin. In order to avoid intermolecular hydrogen bonding, the absolute configuration of prismatomerin was inferred from its acetate derivative, which has the same absolute configuration. Based on the VCD spectra in $CDCl_3$ and B3PW91/TZ2P calculations, it was concluded that the naturally occurring (−)-prismatomerin has (1*R*,5*S*,8*S*,9*S*,10*S*) configuration (47,48), identical to that of (+)-plumericin (46) even though it has positive $[\alpha]_D$.

Other natural products with closely related chemical structures and nearly equal $[\alpha]_D$ revealed that they have opposite absolute configurations. For example, the iridoid oruwacin, isolated from *Morinda lucida*, was identified by NMR to have a structure very close to that of plumericin (49). The $[\alpha]_D$ values measured in $CHCl_3$ are +193 for oruwacin and +197 for plumericin. Seemingly, the replacement of the methyl group at $C_{14}$ by a phenyl ring bearing $OCH_3$ and OH groups had not affected the

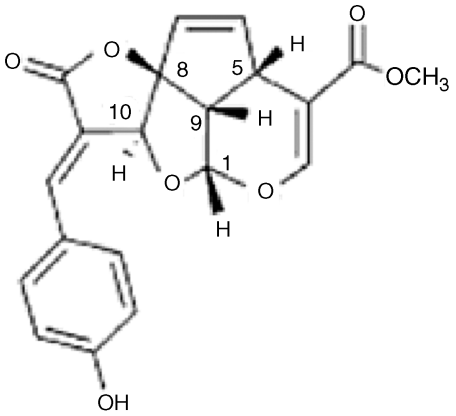


(–)-(1*R*, 5*S*, 8*S*, 9*S*, 10*S*)

***FIGURE 21*** *Structure of prismatomerin (47,48).*

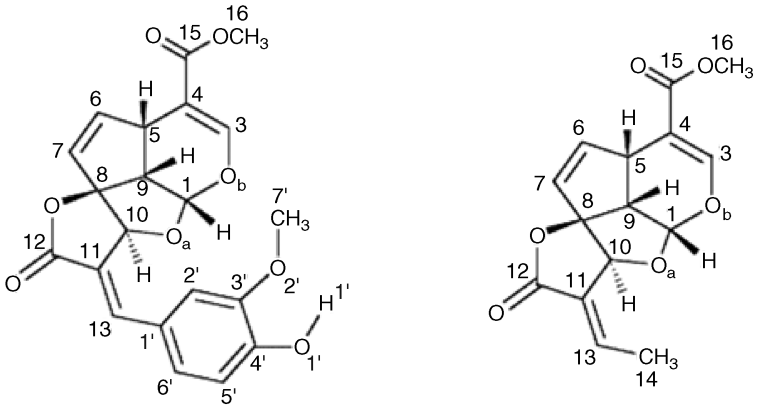


***FIGURE 22*** *Structures of (+)-oruwacin and (+)-plumericin (50)* (copyright American Chemical Society and American Society of Pharmacognosy).

optical rotation. On this basis, it was concluded that both (+)-enantiomers have (1*R*,5*S*,8*S*,9*S*,10*S*) configuration.

However, the calculation of the optical rotation for the two enantiomers of oruwacin using TDDFT has shown that the absolute configuration (1*R*,5*S*,8*S*,9*S*,10*S*) has $[\alpha]_D$ −193, while (1*S*,5*R*,8*R*,9*R*,10*R*) has + 193 (Fig. 22) (50). This result clearly demonstrates that assignments of absolute configurations solely made on the basis of closely related structures and almost equal $[\alpha]_D$ may be erratic.

## 4 METAL IONS, DRUGS, AND PROTEINS

It is well known that DNA and RNA structures respond to interactions with metal ions, drugs, and proteins. Examples of these are selected and articulated below to demonstrate the variety of outcomes of some investigations documented in the available scientific literature.

### 4.1 Metal Ions

Metal ions have distinctive functions in biological systems where they participate in a number of biological processes. Being a major component in life's systems either for the benefit or to the detriment of physical health, one can hardly ignore the need to better understand details of why and how vibrational circular dichroism is able to assist in determining the role it can bring to bear on attempting to ensure *quality of pharmaceutical products*. A realization of the involvement of metal ions in nucleic acids (51,52) followed only a few years after the discoveries of Watson and Crick. The first particularly extensive and comprehensive publication appeared in 1976 (53). Of further interest are several publications that deal with metal ions interacting with

nucleotides, nucleic acids, and their constituents (54–60). By the same token, it is also important to be aware of less desirable aspects of metal ions particularly for human life cycles (61–64).

Compared to conventional infrared absorption spectroscopy as applied to interactions of DNA (or oligomers) with metal ions (65–68), relatively few applications of VCD spectroscopy on these topics can be found in the literature. Considering that VCD signals are generally less intense than those of absorption, simpler oligonucleotides rather than whole DNA have originally provided the desired insights. In one specific case, VCD measurements were used to examine the interaction of $Mn^{2+}$ with $d(GC)_{20}$ and $d(ATGCATGCAT)_2$ (69). The VCD and absorption spectra of the two duplexes were compared without $Mn^{2+}$, with 0.64 [Mn]/[P], and with 0.50 [Mn]/[P]. Essentially no differences in absorption were apparent with only a minor change in VCD in the phosphate region. However, more substantive differences emerged when the spectra of the two oligomers were compared *without* and *with* the manganese ions. The appreciably different spectra suggested unambiguously that VCD provides better definition of structural changes in helical conformations (69).

A theoretical interpretation of the B–Z transition was particularly informative (4). The self-complimentary $d(GC)_{20}$ duplex was measured in the B- and Z-forms and compared to that observed, as well as computed B-conformation without and with explicit water molecules. The results of these aspects are at least twofold: first, the confirmation of the experimental observations with respect to the B–Z transition (69), and second, the computational results that correspond to the observed spectra. The absorption and VCD spectra for the two conformations are portrayed not only in the nitrogen base region but also in the phosphate regions.

The foregoing text dealt mainly with DNA *oligomers* and the interaction with metal ions, specifically $Mn^{2+}$ as a test case. The principal importance of this example is the extent to which VCD can provide structural information in this context (70). The metal concentration was varied from 0 to 0.73 M $Mn^{2+}$ (0–8.5 moles of $Mn^{2+}$) per mole of oligonucleotide phosphate, [Mn]/[P]. In general, the spectra displayed considerable changes as the $Mn^{2+}$ concentrations were increased and the differences were frequently complementary in expression and extent. The main phase of the transition occurred between 0.8 and 1.1 [Mn]/[P]. The spectra transformed gradually in terms of the B–Z transformation binding $Mn^{2+}$ to phosphates and the nitrogen bases as well as partial denaturation. Below 0.6 [Mn]/[P], only a small portion of the oligonucleotide adopted the Z conformation within 3 h; conversion was completed in the same time interval for concentrations between 0.9 and 1.2 [Mn]/[P].

Whereas the absorption spectrum does not change significantly upon the addition of $Cr^{3+}$, changes in the intensity of certain VCD peaks (either positive or negative) are very distinct (Fig. 23) (71). There is no evidence of DNA denaturation even at the highest $Cr^{3+}$ concentration, which was one of the principal purposes for this particular exercise. The main differences between 1750 and 1600 $cm^{-1}$ upon metal ion binding are connected to N7 of guanine. When the $Cr^{3+}$ concentration is raised to 0.6 [Cr]/[P] (DNA + 0.06 M $Cr^{3+}$), the VCD amplitudes increase considerably at several wavenumber positions. At higher concentration, for example, 3 [Cr]/[P], the VCD and absorption features are again similar to those at the start. Within existing

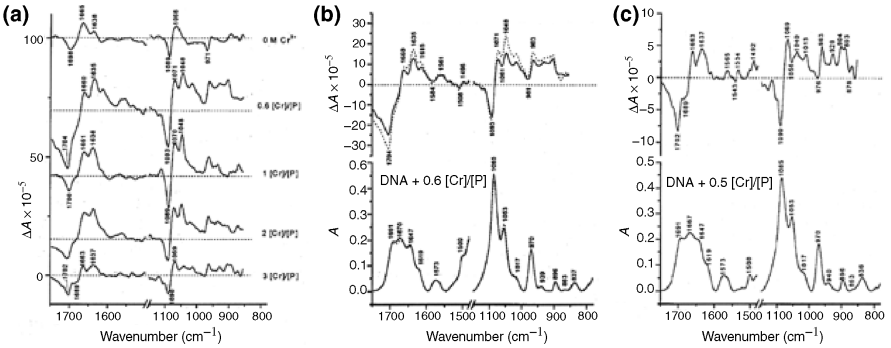


***FIGURE 23*** *Absorption and VCD spectra of DNA with $Cr^{3+}$: (a) DNA + 0, 0.6, 1, 2, and 3 [Cr]/[P]; (b) DNA + 0.6 [Cr]/[P], solid and dotted contours correspond to spectra taken 100 and 220 min after preparing the sample; (c) DNA + 3 [Cr]/[P], 220 min after preparing the sample.*

theory (72,73), the greatly increased amplitudes are referred to as "ψ-type" spectra. Cheng and Mohr (74) referred to this phenomenon as causing DNA to "collapse into rapidly sedimenting form"; Shin and Eichhorn (75) spoke of "structural polymorphism of DNA involved in determining its biological activities"; and Kim et al. (76), who developed the theory of "interaction of light with large inhomogeneous molecular aggregates," referred to it as "psi-type circular dichroism," which is the currently accepted chemical term. In this context, VCD can detect DNA condensation as "particles" of a size comparable to the wavelength of the probing light; in other words and together with suitable metal ions, it can tell something about DNA packing in a living cell. The principal motivation for including trivalent chromium $Cr^{3+}$ ion into this discussion was to demonstrate the capacity and versatility of VCD for detecting DNA *condensation*. The phenomenon itself is well understood and has been known for some time. But it was an important question whether VCD spectroscopy would join the investigative biophysical arsenal (69–76).

The literature that deals with complexes of $Cu^{2+}$ with DNA and oligomers is rather captivating and extensive. Many of the publications dating back to 1966 are captured in a recent treatment dealing explicitly with VCD in combination with infrared absorption spectroscopy (77). In order to be able to distinguish the changing shape of the spectra upon increasing the amount of [Cu]/[P], the solid lines delineate DNA with progressively increasing amounts of [Cu]/[P], whereas the dotted lines represent DNA together with preceding amounts of [Cu]/[P] (Fig. 24). The first addition of $Cu^{2+}$ ions ([Cu]/[P] = 0.2) led only to very minor changes in the DNA spectrum. For subsequently increasing quantities of [Cu]/[P], that is, 0.4, 0.5, and 0.7, each spectrum was compared to that of the previous combination. The most prominent changes are those that occurred for DNA + 0.5 [Cu]/[P] and DNA + 0.7 [Cu]/[P]. Although interpretation is a challenging task, the emerging details with regard to the effects of $Cu^{2+}$ ions on VCD spectra are unmistakable. For example, a barely detectable VCD peak at 1642(+)/1617(−) $cm^{-1}$ for DNA itself emerges as a distinct feature already with 0.2 [Cu]/[P] and thereafter develops as a prominent VCD couplet

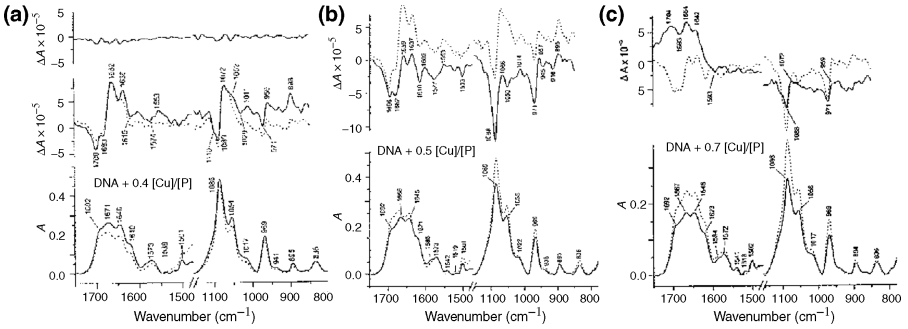


***FIGURE 24*** *Absorption (bottom) and VCD (top) spectra of calf thymus DNA with [Cu]/[P]: 0.4 (a), 0.5 (b), and 0.7 (c).*

at 1636(+)/1619(−) cm$^{-1}$ for DNA + 0.4 [Cu]/[P] and at 1637(+)/1616(−) cm$^{-1}$ for DNA + 0.5 [Cu]/[P] (Fig. 24a and b), but becomes indistinguishable for DNA + 0.7 [Cu]/[P] (Fig. 24c). Prominent also are the VCD signals in the sugar–phosphate region. Emerging details via VCD suggest that little if any significant distortion of Watson–Crick hydrogen bonding or denaturation occurs at 0.2 [Cu]/[P]; distinct changes in VCD indicate $Cu^{2+}$ binding to N7 of guanine, but $Cu^{2+}$ does not interact with $O_2$ of thymine; rather, $Cu^{2+}$ interacts with N3 of cytosine. No significant DNA denaturation is evident. Base stacking is reduced appreciably above 0.4 [Cu]/[P]. The sugar–phosphate backbone remains highly ordered. Finally, the VCD spectrum at 0.7 [Cu]/[P], which has undergone major changes, indicates denaturation of DNA, although quite weak couplets in the sugar–phosphate region signal a partially ordered structure. As one final observation, it is of interest to emphasize that in this case the VCD spectra enabled a cogent summary of why $Cu^{2+}$ ions behave differently with DNA from other metal ions.

If $Mn^{2+}$ ions can induce *condensation* in DNA, either at room temperature or at elevated temperature, can it also lead to *aggregation*? For clarity, the word *condensation* is defined as "formation of dense and relatively large particles in which DNA molecules are packed with a regular arrangement," whereas *aggregation* is understood as "a significantly decreasing volume occupied by the DNA molecule forming large but relatively loose molecular aggregates with an unordered arrangement" (78). The particles of the former are capable of creating ψ-type spectra, whereas those of the latter associate in unordered arrangements. Keeping in mind that theoretically as well as experimentally condensation and aggregation have long been understood, the principal objective of this investigation was to determine whether VCD would be able to determine the presence or absence of DNA condensation and aggregation and also distinguish between these two phenomena. As the first step, the absorption and VCD spectra were measured for calf thymus DNA at different temperatures. Thereafter, the spectra were obtained for four different sets of [Mn]/[P] ratios and two different temperatures (23 and 55°C).

The set of VCD experiments involving $Mn^{2+}$ ions revealed several insights. The secondary structure remained in the B-form during the whole ion concentration range

at room temperature. No significant denaturation of DNA occurred at room temperature even at the highest metal ion concentration. At elevated temperatures, the manganese ions caused decreasing stability, and melting temperature of DNA accompanied denaturation. Evidently, VCD is sensitive to DNA condensation or aggregation, and by the same token it is also able to distinguish between these processes. Although condensation of DNA became evident in a narrow range of experimental conditions near 2.4 [Mn]/[P] and 55°C, aggregation took place in the presence of $Mn^{2+}$ ions at these elevated temperatures, during or after denaturation.

From a relatively simple experiment by today's standards, Rosenberg, Van Camp, and Krigas provided inadvertently a possibility to combat cancer: "In an investigation of the possible effects of an electric field on growth processes in bacteria, we have discovered a new and interesting effect." (79). Of course, the chemical substance pertaining to this event is cisplatin, as is well known. Much has been learned and published since then about the benefits and shortcomings of this relatively simple material as well as its structure and mode of action. Some attempts have been made to determine what contributions this physical–chemical method could offer from the perspective of vibrational circular dichroism, although the nature of the subject and its tools have so far limited the scope of these endeavors. In two relatively recent publications, the authors have probed the feasibility, accuracy, and practical value of VCD for this purpose. In the case of the octamer d(CCTG*G*TCC)·d(GGACCAGG) (80), the multiple positive and negative peaks in the VCD spectra looked enticingly interesting (Fig. 25). Reasonable wavenumber positions are relatively easy to ascertain particularly for this octamer and without

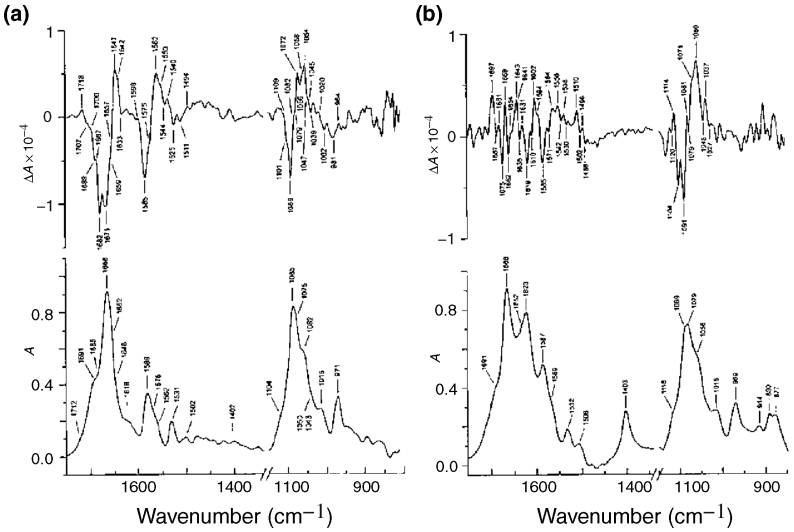


***FIGURE 25*** *Absorption and VCD spectra of cisplatin–d(CCTG* G* TCC)·d(GGACCAGG) adduct: (a) octamer without cisplatin; (b) octamer with cisplatin.*

cisplatin; what appear to be complicated spectra especially for the cisplatin complex can also be readily assigned (Table 1). Although the VCD spectra may look somewhat confusing, their interpretation too is relatively easy. Notable, for instance, are significant intensity decreases in the VCD spectrum of the platinated octamer (1750–1500 $cm^{-1}$), or specifically a new couplet at 1631(+)/1619(−) $cm^{-1}$, which is not apparent in the unplatinated sample. Upon being exposed to cisplatin, the octamer isomerizes from an *intra*strand to an *inter*strand cross-link as demonstrated in an earlier NMR analysis (81). Figure 24a and b clearly confirms that while the absorption spectra of the complex are less affected, the VCD spectra differ considerably. It is not entirely clear from this particular experiment what structure emerges, but it is clear that the spectra are affected to a considerable extent in both the 1750–1500 and 1100–900 $cm^{-1}$ ranges. Specifically, the fourth base step between G4pG5 is connected to platinum via the N7 atoms of guanine, where most prominent changes occur in the spectra. Whereas absorptions do not change with time, the VCD spectra are progressively altered due to a slow isomerization from *intra*strand to *inter*strand cross-links (80).

The absorption and VCD spectra of the same octamer together with cisplatin were modeled by *ab initio* simulation using density functional computations and a Cartesian coordinate-based transfer of molecular property tensors from smaller DNA fragments (8). If successful in this endeavor, situations are likely to emerge when this combination of theoretical and physical methods will provide compelling support for interpreting VCD spectra alongside infrared absorption and other physical–chemical methods. Although a computed *absorption* spectrum, which used a scaled BPW91/6-31G** force field, might be considered a reasonable reproduction, the VCD spectra leave something yet to be desired.

Compared to the infrared absorption spectrum of d(CCTG*G*TCC)·d(GGACCAGG), its VCD trace appears greatly more interesting and challenging. Although perhaps difficult, the multitude of absorption signals without cisplatin is likely to be readily interpreted, but those of VCD are more difficult from a different perspective (Table 1). The correlation between the spectra of the free and complexed oligomers is truly remarkable, although it can be expected that the spectral details of an entire DNA are likely even more complicated. The strong absorptions stand out clearly, whereas the weaker ones are often only suggestive. In contrast, the VCD spectra reveal a rich variety of signals that are even more prominent in their details especially in the presence of cisplatin.

While informative, the results of cisplatin when complexed with the octamer did not indicate whether spontaneous conversion from an *intra*strand to an *inter*-strand oligomer takes place independent of the chain length. The dodecamer d(CCTCTG*G*TCTCC)·d(GGAGACCAGAGG) (Fig. 26) was therefore complexed with cisplatin and the absorption as well as VCD spectra recorded (Fig. 26a and b, Table 2) (82). Both the absorption and VCD spectra differ appreciably between those of the octamer and dodecamer, but differ more prominently between the free and complexed dodecamer. VCD features of the dodecamer are significantly overlapped particularly in the 1750–1500 $cm^{-1}$ range and then even more upon complexation with cisplatin. Although changes in absorption are also significant, they are not as

**TABLE 1 Absorption and VCD Data of d(CCTG*G*TCC)·d(GGACCAGG) Free and Complexed with Cisplatin (Fig. 25)**

| Assignment | | Octamer Free | | Octamer–Pt Complex | |
|---|---|---|---|---|---|
| | | Absorption | VCD ($cm^{-1}$) | Absorption | VCD ($cm^{-1}$) |
| C2=O stretch thymine | { | 1712 | 1718(+)/1707(−) | | |
| | | 1691 | 1700(+)/1689(−) | 1691 | 1697(+)/1687(−) |
| C6=O stretch guanine | { | 1685 | 1687(+)/1682(−) | 1666 | 1681(+)/1675(−) |
| | | 1666 | 1671(−)/~1659(+) | | 1668(+)/1662(−) |
| C=O stretch cytosine | | 1652 | ~1657(−)/1647(+) | 1652 | 1654(−)/1643(+) |
| C4=O stretch thymine | | 1648 | 1642(+)/1633(−) | | 1641(+)/1635(−) |
| Ring in-plane mode of adenine | | 1618 | | 1623 | |
| | { | | | | 1610(−)/1602(+) |
| C–N(D)$_2$ stretch guanine | | 1589 | 1598(+)/1585(−) | 1587 | 1594(+)/1585(−) |
| | | 1575 | 1575(−)/1562(+) | 1569 | 1571(−)/1564(+) |
| C=N stretch guanine | | 1562 | 1553(+)/1544(−) | | 1550(+)/1542(−) |
| Ring in-plane cytosine | | 1531 | 1540(+)/1525(−) | 1532 | 1538(+)/1530(−) |
| C=N stretch cytosine | { | 1502 | 1511(−)/1494(+) | 1506 | 1510(+)/1502(−) |
| | | | | | 1496(+)/1488(−)[a] |
| | | | | 1403 | |
| Sym PO$_2$ plus C–O stretch | { | | | 1118 | 1120(−)[a] |
| | | 1104 | 1109(+)/1101(−) | 1110 | 1114(+)/1104(−) |
| Sym PO$_2$ stretch | | 1085 | 1089(−)/1082(+) | 1089 | 1091(−)/1081(+) |
| Phosphate diester backbone coupled to sugar modes | { | 1075 | 1079(−)/1072(+) | 1079 | 1079(−)/1071(+) |
| | | 1062[a] | 1066(−)/1058(+)[a] | | |
| | | 1050[a] | 1054(+)/1047(−)[a] | 1056 | 1060(+)/1045(−) |
| | | 1043[a] | 1045(+)/1039(−)[a] | | 1037(+)/1027(−) |
| Sugar modes | { | 1016 | | 1016 | |
| | | 971 | 981(−)/964(+)[a] | 969 | |
| | | | | 914 | |
| | | 890 | | 890 | |
| | | | | 877 | |

[a]Assignments/designations uncertain.

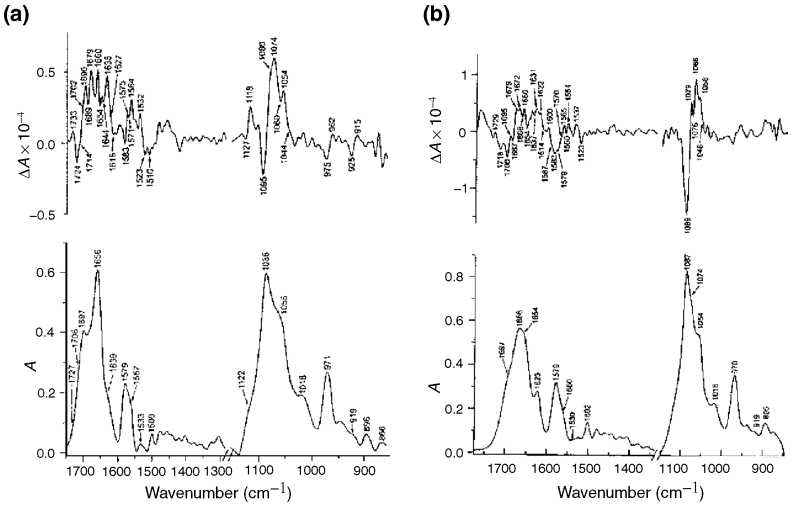


***FIGURE 26*** *Absorption and VCD of cisplatin–d(CCTCTG*G*TCTCC)·d(GGAGACCAGAGG) adduct: (a) dodecamer without cisplatin; (b) dodecamer with cisplatin.*

prominent as in VCD. Both the absorption and VCD spectra indicate a typical B configuration for the dodecamer and the octamer. Pertinent is also the fact that both the absorption and VCD spectra are changed significantly upon complex formation. Of all the differences that were noted, the most pertinent observation is that the effects of platination on the dodecamer duplex were changes in VCD expressed specifically by the vibrations of the guanines to which the platinum atoms become attached, namely, G*6 and G*7. Further, the spectra did not indicate that any isomerization of the complex had taken place contrary to what was observed for the octamer. The major thrust that emerged from the foregoing text is that VCD is capable of making significant contributions to generating new knowledge about specific drugs perhaps aided in combination with appropriately sophisticated *ab initio* simulations.

## 4.2 From Simple to Greater Challenges

Various aspects of metal ions were considered in Section 4.1. The components include a range of publications all of which focused mainly on interactions of oligomers or DNA with metal ions (69–71,77,78). Examples of infrared absorption and vibrational circular dichroism combining DNA with metal ions *and* proteins are articulated below (80–89). Included are combinations of DNA with $Mn^{2+}$ (69,70,78,83), $Cu^{2+}$ (77), $Pt^{2+}$ (80,82), and the proteins HMGB1 and H1 in addition with DNA forming large supramolecular complexes using UV, infrared absorption, and vibrational circular dichroism (84–88), HMGB1 in the presence of $Mn^{2+}$ using UV and CD (85) and DNA/HMGB1/H1 complex with $Mn^{2+}$ (86). In all these instances, the effects on DNA and the selected proteins in the presence of the metal ions are clearly evident. In another investigation, which used infrared absorption and vibrational circular dichroism as well as electronic circular dichroism spectroscopy, the effect of $Ca^{2+}$ ions on DNA was considered

**TABLE 2 Absorption and VCD Data of d(CCTCTG*G*TCTCC)·d(GGAGACCAGAGG) Free and Complexed with Cisplatin (Fig. 26)**

| Assignment | Dodecamer Free | | Dodecamer–Pt Complex | |
|---|---|---|---|---|
| | Absorption | VCD ($cm^{-1}$) | Absorption | VCD ($cm^{-1}$) |
| C=C base stretching modes | 1727 | 1733(+)/1724(−) | | 1729(+)/1718(−) |
| C=O stretch of thymine | 1706 | 1714(−)/~1702(+) | | |
| | 1697 | ~1702(−)/1695(+) | 1697 | 1700(−)/1695(+) |
| C6=O stretch of guanine | | 1689(−)/1679(+) | 1666 | 1687(−)/1679(+) |
| | | | | 1672(+)/1666(−) |
| C=O stretch of cytosine | 1656 | ~1660(+)/1654(−) | 1654 | 1660(+)/1654(−) |
| C4=O stretch of thymine | 1639 | 1644(−)/1634(+) | | 1637(−)/1631(+) |
| Ring in-plane mode of adenine | | 1627(+)/1616(−) | 1623 | 1622(+)/1614(−) |
| C–N$(D)_2$ stretch guanine | 1579 | 1583(−)/1575(+) | 1579 | 1600(+)/1587(−) |
| | | | | 1583(−)/1570+) |
| C=N stretch guanine | 1567 | 1571(−)/1564(+) | ~1560 | 1579(−)/1565(+) |
| | | | | 1560(−)/1554(+) |
| Ring in-plane mode cytosine | 1533 | 1532(+)/1523(−) | 1530 | 1537(+)/1523(−) |
| C=N stretch cytosine | 1500 | 1510(−)/?(+) | 1502 | |
| Sym $PO_2^-$ plus C–O stretch | 1122 | 1127(−)/1118(+) | | |
| Sym $PO_2^-$ stretch | 1086 | 1095(−)/1080(+) | 1087 | 1089(−)/1079(+) |
| Phosphate diester backbone coupled to sugar modes | | 1074(+)/1060(−) | 1074 | 1076(−)/1068(+) |
| | ~1056 sh | 1054(+)/1044(−) | 1054 | 1058(+)/~1048(−) |
| Sugar modes { | 1018 | | 1016 | |
| | 971 | 975(−)/962(+) | 970 | |
| | 919 | 924(−)/915(+) | 919 | |
| | 896 | | 896 | |
| | 866 | | 866 | |

leading to compaction when complexed with non-histone HMGB1 (87). Intricate details emerged from the IR/VCD perspective via a study of DNA-$Ca^{2+}$ interaction in the infrared regions from 1750 to 1300 and 1150 to 750 $cm^{-1}$. The molar ratio of $Ca^{2+}$ to DNA phosphate in solution, $R = [Ca^{2+}]/[P]$, ranged from 0.015 to 1.3; the absorption and VCD features resembled very closely those of $Mn^{2+}$ and DNA (70). The regions between 1750 to 1500 and 1150 to 800 $cm^{-1.}$ differed only in lesser nuances. The forerunner to this work was an earlier analysis, which compared the behavior of DNA on interaction with HMGB1 and H1 at different ionic strengths and protein content (88). The applications of IR/VCD spectroscopy to study proteins and their complexes with DNA were described recently elsewhere (89).

## 5 SUMMARY AND CONCLUDING REMARKS

The principal purpose of this article was to provide a relatively brief overview of today's utility of *vibrational circular dichroism*. This article focuses on instrumental methods, unique applicabilities and capabilities, review of scientific applications such as enantiomeric purity and absolute configuration, solution conformations, pharmaceuticals, and natural products, and concludes with an assessment of the instrumentation's utility when metal ions play significant roles in biological and medicinal processes, and the observation that the same metal ions significantly affect large molecular biological complexes that arise from the interactions of DNA, proteins, and medicinal drugs.

Section 2 deals with the principles of vibrational circular dichroism and required instrumentation methods, and thereafter discusses the uniqueness and applicability of results. These components are intended as quick recitals for someone who is not familiar with the instrumentation. In Section 3, one is reminded of several aspects of vibrational circular dichroism spectroscopy. These turn out to be significant, important, and useful. Among them are topics that deal with enantiomeric purity, enantiomeric excess, and absolute configuration. All three of these are generally useful. Conformation of solutions is usually a desirable element, which can assist in identifying details of absolute configuration. The subsection that deals with pharmaceutical molecules focuses on materials of medicinal purpose such as *Ephedra* drugs, gossypol, and oxydiazol-3-one, to mention but a few. The last subsection highlights natural products, usually extracts of various plant materials, such as the family of alkaloids for instance. It is in this subsection that the density functional theory calculations have proven to be very successful.

Finally in Section 4, metal ions in combination with medicinal drugs and proteins demonstrate a functional capability for expressing stereochemical effects on nucleic acid oligomers, whole DNA, and potential drugs of various kinds.

## REFERENCES

1. Stephens PJ, Clark R. Vibrational circular dichroism: the experimental viewpoint. In: Mason SF,editor. Optical Activity and Chiral Discrimination. Proceedings of the NATO

Advanced Study Institute, University of Sussex, UK, September 10–22, 1978. Vol. 48. London: D. Reidel Publishing; 1978. Chapter X, p 263–287.

2. Anamalai A, Keiderling T. Vibrational circular dichroism of poly(ribonucleic acids). J Am Chem Soc 1987; 109: 3125–3132.
3. Wang L, Keiderling TA. Vibrational circular dichroism studies of the A-to-B conformational transition in DNA. Biochemistry 1992; 31: 10265–10271.
4. Andrushchenko V, Wieser H, Bour P. B–Z conformational transition of DNA monitored by vibrational circular dichroism. *Ab initio* interpretation of the experiment. J Phys Chem B 2002; 106: 12623–12634.
5. Freedman TB, Cao X, Oliveira RV, Cass QB, Nafie LA. Determination of the absolute configuration and solution conformation of gossypol by vibrational circular dichroism. Chirality 2003; 15: 196–200.
6. Andrushchenko V, Tsankov D, Wieser H. Vibrational circular dichroism spectroscopy and the effects of metal ions on DNA structure. J Mol Struct 2003; 661–662: 541–560.
7. Bodack LA, Freedman TB, Chowduhry BZ, Nafie LA. Solution conformations of cyclosporins and magnesium–cyclosporin complexes determined by vibrational circular dichroism. Biopolymers 2004; 73: 163–177.
8. Andrushchenko V, Wieser H, Bour P. DNA oligonucleotide–*cis*-platin binding: *ab initio* interpretation of the vibrational spectra. J Phys Chem A 2007; 111: 9714–9723.
9. Nafie LA, Dukor RK. Applications of Vibrational Optical Activity in the Pharmaceutical Industry. Handbook of Vibrational Spectroscopy. Applications in Life, Pharmaceutical and Natural Sciences. Wiley; 2007. p 662–676.
10. Barron LD, Bogaard MP, Bukingham AD. Solutions structure and dynamics of biomolecules from Raman optical activity. J Am Chem Soc 1973; 95: 603–605.
11. Ragunathan N, Lee N-S, Freedman TB, Nafie LA, Tripp C, Buijs H. Measurement of vibrational circular dichroism using a polarizing Michelson interferometer. Appl Spectrosc 1990; 44: 5–7.
12. Dudley RJ, Mason SF, Peacock RD. Infrared vibrational circular dichroism. J Chem Soc Chem Commun 1972; 1084–1085.
13. Wang W, Griffiths P, Kemeny G. Vibrational circular dichroism measurements by optical subtraction FT-IR spectroscopy. Appl Spectrosc 1984; 38: 337–343.
14. Nafie LA, Vidrine DW. Double modulation Fourier transform spectroscopy. In: Ferraro JR, Basile LJ, editors. Fourier Transform Infrared Spectroscopy. Vol. 3. New York: Academic Press; 1982. p 83–123.
15. McCoy C, de Haseth JA. Phase correction of vibrational circular dichroic features. Appl Spectrosc 1988; 42: 336–341.
16. Nafie LA, Diem M. Theory of high frequency differential interferometry: application to the measurement of infrared circular and linear dichroism via Fourier transform spectroscopy. Appl Spectrosc 1979; 33: 130–135.
17. Nafie LA. Dual polarization modulation: real-time, spectral multiplex separation of circular dichroism from linear birefringence spectral intensities. Appl Spectrosc 2000; 54: 1634–1645.
18. Tsankov D, Eggimann T, Wieser H. Alternative design for improved FT-IR/VCD capabilities. Appl Spectrosc 1995; 49: 132–138.

19. Shah R, Nafie LA. Spectroscopic methods for determining enantiomeric purity and absolute configuration in chiral pharmaceutical molecules. Curr Opin Drug Discov Dev 2001; 4: 764–775.
20. Zhao C, Polavarapu P. Comparative evaluation of vibrational circular dichroism and optical rotation for determination of enantiomeric purity. Appl Spectrosc 2001; 55: 913–918.
21. Guo C, Shah RD, Dukor RK, Cao X, Freedman TB, Nafie LA. Determination of enantiomeric excess in samples of chiral molecules using Fourier transform vibrational dichroism spectroscopy: simulation of real-time reaction monitoring. Anal Chem 2004; 76: 6956–6966.
22. Seco JM, Quiñoá E, Riguera R. The assignment of absolute configuration by NMR. Chem Rev 2004; 104: 17–117.
23. Omata K, Fujiwara T, Kabuto K. Use of diamagnetic lanthanide complex for extending the scope of NMR determination of absolute configuration by the modified Mosher's method. Tetrahedron: Asymmetry 2002; 13: 1655–1662.
24. Cheeseman JR, Frisch MJ, Devlin FJ, Stephens PJ. Hartree–Fock and density functional theory *ab initio* calculation of optical rotation using GIAOs: basis set dependence. J Phys Chem 2000; 104: 1039–1046.
25. Stephens PJ, Devlin FJ, Cheeseman JR, Frisch MJ. Calculation of optical rotation using density functional theory. J Phys Chem 2001; 105: 5356–5371.
26. Rauk A, Yang D, Tsankov D, Wieser H, Koltypin Yu, Gedanken A, Shustov G. Chiroptical properties of 1-azabicyclo[3.1.0]hexane in the vacuum UV and IR regions. J Am Chem Soc 1995; 117: 4160–4166.
27. Nieman JA, Keay BA, Kubicki M, Yang D, Rauk A, Tsankov D, Wieser H. Determining absolute configuration by vibrational circular dichroism: (+)-(1*S*,5*S*,6*S*)-and (−)-(1*R*,5*R*,6*R*)-spiro[4,4]nonane-1,6-diol. J Org Chem 1995; 60: 1918–1919.
28. Devlin FJ, Stephens PJ, Scafato P, Superchi S, Rosini C. Conformational analysis using infrared and vibrational circular dichroism spectroscopies: the chiral cyclic sulfoxides 1-thiochroman-4-on *S*-oxide, 1-thiaindan *S*-oxide and 1-thiochroman *S*-oxide. J Phys Chem A 2002; 106: 10510–10524.
29. Stephens PJ, Devlin FJ, Gasparrini F, Ciogli A, Spinelli D, Cosimelli B. Determination of the absolute configuration of a chiral oxadiazol-3-one calcium channel blocker, resolved using chiral chromatography, via concerted density functional theory calculations of its vibrational circular dichroism, electronic circular dichroism and optical rotation. J Org Chem 2007; 72: 4707–4715.
30. Freedman TB, Ragunathan N, Alexander S. Vibrational circular dichroism in *Ephedra* molecules. Experimental measurement and *ab initio* calculation. Faraday Discuss 1994; 99: 131–149.
31. Ye WS, Liang JC, Hsu TC. Toxicity of a male contraceptive, gossypol, in mammalian cell cultures. In Vitro 1983; 19: 53–57.
32. Tsui Y-C, Creasy MR, Hultén MA. The effect of the male contraceptive agent gossypol on human lymphocytes *in vitro*: traditional chromosome breakage, micronuclei, sister chromatid exchange, and cell kinetics. J Med Genet 1983; 20: 81–85.
33. Huang L, Si YK, Snatzke G, Zheng DK, Zhou J. Absolute configuration of gossypol. Collect Czech Chem Commun 1988; 53: 2664–2666.

34. Oliver CL, Miranda MB, Shangary S, Land S, Wang S, Johnson DE. (−)-Gossypol acts directly on the mitochondria to overcome Bcl-2- and Bcl-$X_L$-mediated apoptosis resistance. Mol Cancer Ther 2005; 4: 23–31.
35. Dyatkin A, Freedman T, Cao X, Dukor R, Maryanoff B, Maryanoff C, Matthews JM, Shah R, Nafie L. Determination of the absolute configuration of a key tricyclic component of a novel vasopressin receptor antagonist by use of vibrational circular dichroism. Chirality 2002; 14: 215–219.
36. Maryanoff BE, McComsey DF, Dukor R, Nafie LA, Freedman TB, Cao X, Day VW. Structural studies on McN-5652-X, a high-affinity ligand for the serotonin transporter in mammalian brain. Bioorg Med Chem 2003; 11: 2453–2470.
37. Dunmire D, Freedman TB, Nafie LA, Aeschlimann C, Gerber JG, Gal J. Determination of the absolute configuration and solution conformation of the antifungal agents ketoconazole, itraconazole, and miconazole with vibrational circular dichroism. Chirality 2005; 17: S101–S108.
38. Monde K, Taniguchi T, Miura N, Nishimura S-I, Harada N, Dukor R, Nafie L. Preparation of cruciferous phytoalexin related metabolites, (−)-dioxybrassinin and (−)-3-cyanomethyl-3-hydroxyoxindole, and determination of their absolute configuration by vibrational circular dichroism (VCD). Tetrahedron Lett 2003; 44: 6017–6020.
39. Materson BJ. Calcium channel blockers: is it time to split the lump? Am J Hypertens 1995; 8: 325–329.
40. Budriesi R, Cosimelli B, Ioan P, Lanza CZ, Spinelli D, Chiarini A. Cardiovascular characterization of [1,4]thiazino[3,4-*c*][1,2,4]oxadiazol-1-one derivatives: selective myocardial calcium channel modulators. J Med Chem 2002; 45: 3475–3481.
41. Budriesi R, Carosati E, Chiarini A, Cosimelli GC, Ioan P, Spinelli D, Spisani R. A new class of selective myocardial calcium channel modulators. 2. Role of the acetal chain in oxadiazol-3-one derivatives. J Med Chem 2005; 48: 2445–2456.
42. Carosati E, Cruciani G, Chiarini A, Budriesi R, Ioan P, Spisani R, Spinelli D, Cosimelli B, Fusi F, Frosini M, Matucci R, Gasparrini F, Ciogli A, Stephens PJ, Devlin FJ. Calcium channel antagonists discovered by a multidisciplinary approach. J Med Chem 2006; 49: 5206–5216.
43. Stephens PJ, Pan JJ, Devlin FJ, Urbanová M, Hájiček J. Determination of the absolute configurations of natural products via density functional theory calculations of vibrational circular dichroism, electronic circular dichroism and optical rotation: the schizozygane alkaloid schizozygine. J Org Chem 2007; 72: 2508–2524.
44. Stephens PJ, Pan JJ, Devlin FJ, Urbanová M, Julínek O, Hájiček J. Determination of the absolute configurations of natural products via density functional theory calculations of vibrational circular dichroism, electronic circular dichroism, and optical rotation: the *iso*-schizozygane alkaloids *iso*-schizogaline and *iso*-schizogamine. Chirality 2007; 20: 454–470.
45. Albers-Schönberg G, Schmid H. Über die Struktur von Plumericin, Isoplumericin, β-Dihidroplumericin und der β-Dihidroplumericinsäure. Helv Chim Acta 1961; 44: 1447–1473.
46. Stephens PJ, Pan JJ, Devlin FJ, Krohn K, Kurtán T. Determination of the absolute configurations of natural products via density functional theory calculations of vibrational circular dichroism, electronic circular dichroism and optical rotation: the iridoids plumericin and isoplumericin. J Org Chem 2007; 72: 3521–3536.

47. Stephens PJ, Pan JJ, Krohn K. Determination of the absolute configurations of pharmacological natural products via density functional theory calculations of vibrational circular dichroism: the new cytotoxic iridoid prismatomerin. J Org Chem 2007; 72: 7641–7649.
48. Krohn K, Gehle D, Sohrab MH, Dey SK, Nahar N, Mosihuzzaman M, Sultana M, Stephens PJ, Pan JJ, Sasse F. Prismatomerin, a new iridoid from *Prismatomeris tetrandra*. Structure elucidation, determination of the absolute configuration, and cytotoxicity. J Nat Prod 2007; 70: 1339–1343.
49. Adesogan EK. Oruwacin, a new iridoid ferulate from *Morinda lucida*. Phytochemistry 1979; 18: 175–176.
50. Stephens PJ, Pan JJ, Devlin FJ, Cheeseman JR. Determination of the absolute configurations of natural products using TDDFT optical rotation calculations: the iridoid oruwacin. J Nat Prod 2008; 71: 285–288.
51. Wacker WEC, Vallee BL. Nucleic acids and metals. I. Chromium, manganese, nickel, iron, and other metals in ribonucleic acids from diverse biological sources. J Biol Chem 1959; 234: 3257–3262.
52. Eichhorn GL. Metal ions as stabilizers or destabilizers of the deoxyribonucleic acid structure. Nature 1962; 194: 474–475.
53. Sissoëff I, Grisvard J, Guillé E. Studies on metal ions–DNA interactions: specific behaviour of reiterative DNA sequences. Prog Biophys Mol Biol 1976; 31: 165–199.
54. Spiro TG. Metal Ions in Biology. Vol. 1. Nucleic Acid–Metal Ion Interactions. Wiley; 1980.
55. Duguid JG, Bloomfield VA, Benevides JM, Thomas GJ Jr. Raman spectroscopy of DNA–metal complexes. I. Interactions and conformational effects of divalent cations: Mg, Ca, Sr, Ba, Mn, Co, Ni, Cu, Pd, and Cd. Biophys J 1993; 65: 1916–1928.
56. Duguid JG, Bloomfield VA. Aggregation of melted DNA by divalent metal ion-mediated cross-linking. Biophys J 1995; 69: 2642–2648.
57. Sigel A. Interactions of metal ions with nucleotides and nucleic acids and their constituents. Chem Soc Rev 1993; 22: 213–292.
58. Kazakov S, Hecht SM. Nucleic acid–metal ion interactions. Encyclopedia of Inorganic Chemistry. Wiley–Interscience; 2006. p 1–34.
59. Sigel A, Sigel H, Sigel RKO. Metal Ions in Life Sciences. Vol. 1. Neurodegenerative Diseases and Metal Ions. Wiley; 2006.
60. Sigel A, Sigel H, Sigel RKO. Metal Ions in Life Sciences. Vol. 2. Nickel and Its Surprising Impact in Nature. Wiley; 2007.
61. Eichhorn GL, Butzow JJ, Shin YA. Some effects of metal ions on DNA structure and genetic information transfer. Proc Int Symp Biomol Struct Interact Suppl J Biosci 1985; 8: 527–535.
62. Hartwig A. Recent advances in metal carcinogenicity. Pure Appl Chem 2000; 72: 1007–1014.
63. Valko M, Morris H, Cronin MTD. Metals, toxicity and oxidative stress. Curr Med Chem 2005; 12: 1161–1209.
64. Adlard PA, Bush AI. Metals and Alzheimer's disease. J Alzheimers Dis 2006; 10: 145–163.
65. Theophanides T. Fourier transform infrared spectra of calf thymus DNA and its reactions with the anticancer drug cisplatin. Appl Spectrosc 1981; 35: 461–465.

66. Strangret J, Savoie R. Vibrational spectroscopic study of the interaction of metal ions with diethyl phosphate, a model for biological systems. Can J Chem 1992; 70: 2875–2883.
67. Arakawa H, Ahmad R, Naoui M, Tajmir-Riahi H-A. A comparative study of calf thymus DNA binding to Cr(III) and Cr(IV). J Biol Chem 2000; 275: 10150–10153.
68. Ahmad R, Arakawa H, Tajmir-Riahi HA. A comparative study of DNA complexation with Mg(II) and Ca(II) in aqueous solution. Biophys J 2003; 84: 2460–2466.
69. Andrushchenko VV, van de Sande JH, Wieser H. Interaction of deoxyribo-oligonucleotides with divalent manganese ions: comparison of vibrational circular dichroism and absorption spectroscopy. Vibr Spectrosc 1999; 19: 341–345.
70. Andrushchenko VV, van de Sande JH, Wieser H, Kornilova SV, Blagoi YP. Complexes of $(dG\text{-}dG)_{20}$ with $Mn^{2+}$ ions: a study by vibrational circular dichroism and infrared spectroscopy. J Biomol Struct Dyn 1999; 17: 545–560.
71. Andrushchenko V, Leonenko Z, Cramb D, van de Sande H, Wieser H. Vibrational CD (VCD) and atomic force microscopy (AFM) study of DNA interaction with $Cr^{3+}$ ions: VCD and AFM evidence of DNA condensation. Biopolymers 2002; 61: 243–260.
72. Keller D, Bustamante C. Theory of the interaction of light with large inhomogeneous molecular aggregates. I. Absorption. J Chem Phys 1986; 84: 2961–2971.
73. Keller D, Bustamante C. Theory of the interaction of light with large inhomogeneous molecular aggregates. II. Psi-type circular dichroism. J Chem Phys 1986; 84: 2972–2980.
74. Cheng S-M, Mohr SC. The thermal transition of 'psi' DNA monitored by circular dichroism. FEBS Lett 1974; 49: 37–42.
75. Shin YA, Eichhorn GL. Formation of $\psi(+)$ and $\psi(-)$ DNA. Biopolymers 1984; 23: 325–335.
76. Kim M-H, Ulibarri L, Keller D, Maestre MF, Bustamante C. The psi-type circular dichroism of large molecular aggregates. III. Calculations. J Chem Phys 1986; 84: 2981–2989.
77. Andrushchenko V, van de Sande JH, Wieser H. Vibrational circular dichroism and IR absorption of DNA complexes with $Cu^{2+}$ ions. Biopolymers 2003; 72: 374–390.
78. Andrushchenko V, van de Sande H, Wieser H. DNA interaction with $Mn^{2+}$ ions at elevated temperatures: VCD evidence of DNA aggregation. Biopolymers 2003; 69: 529–545.
79. Rosenberg B, Van Camp L, Krigas T. Inhibition of cell division in *Escherichia coli* by electrolysis products from a platinum electrode. Nature 1965; 205: 698–699.
80. Tsankov D, Kalisch B, van de Sande H, Wieser H. Cisplatin–DNA adducts by vibrational circular dichroism spectroscopy: structure and isomerization of d(CCTG*G*TCC)·d(GGACCAGG). J Phys Chem B 2003; 107: 6479–6485.
81. Yang D, van Boom SSGE, Reedijk J, van Boom JH, Wang AH-J. Structure and isomerization of an intrastrand cisplatin cross-linked octamer DNA duplex by NMR analysis. Biochemistry 1995; 34: 12912–12920.
82. Tsankov D, Kalisch B, van de Sande JH, Wieser H. Cisplatin adducts of d(CCTCTG*G*TCTCCC)·d(GGAGACCAGAGG) in aqueous solution by vibrational circular dichroism spectroscopy. Biopolymers 2003; 72: 490–499.
83. Polyanichko AM, Andrushchenko VV, Chikhirzhina EV, Vorob'ev VI, Wieser H. The effect of manganese(II) on DNA structure: electronic and vibrational circular dichroism studies. Nucleic Acids Res 2004; 32: 989–996.

84. Polyanichko A, Wieser H. Fourier transform infrared/vibrational circular dichroism spectroscopy as an informative tool for the investigation of large supramolecular complexes of biological macromolecules. Biopolymers 2005; 78: 329–339.
85. Polyanichko AM, Chikhirzhina EV, Kostyleva EI, Vorob'ev VI. Structure of DNA complexes with nonhistone chromosomal protein HMGB1 in the presence of manganese ions. Mol Biol 2004; 38: 891–898.
86. Polyanichko AM, Chikhirzhina EV, Andrushchenko VV, Wieser H. The effect of manganese(II) on the structure of DNA/HMGB1/H1 complexes: electronic and vibrational circular dichroism studies. Biopolymers 2006; 83: 182–192.
87. Polyanichko AM, Chikhirzhina EV, Andrushchenko VV, Wieser H, Vorob'ev VI. The effect of $Ca^{2+}$ ions on DNA compaction in the complex with HMGB1 nonhistone chromosomal protein. Mol Biol 2004; 38: 701–712.
88. Chikhirzhina EV, Polyanichko AM, Skvortsov AN, Kostyleva EI, Houssier C, Vorob'ev VI. HMG1 domains: the victims of circumstances. Mol Biol 2002; 36: 412–418.
89. Polyanichko AM, Wieser H. (2007) Vibrational circular dichroism and its applications to protein studies. In "Methods in Protein Structure and Stability Analysis: Vibrational Spectroscopy", E. Permyakov and V. Uversky (Eds.), Nova Science Publishers, Inc., New York, USA, 2007, pp 267–302.

# *New Approaches to the Characterization of Drug Candidates by Solid-State NMR*

**Frederick G. Vogt,[1] Jacalyn S. Clawson,[1] Mark Strohmeier,[1] Tran N. Pham,[2] Simon A. Watson[2] and Andrew J. Edwards[2]**

[1]*Chemical Development, GlaxoSmithKline plc., King of Prussia, PA*
[2]*Chemical Development, GlaxoSmithKline plc., Stevenage, Hertfordshire SG1 2NY, UK*

## 1 INTRODUCTION

Solid-state NMR (SSNMR) spectroscopy is a commonly applied analytical technique in pharmaceutical development. Reviews of the pharmaceutical applications of SSNMR have recently appeared (1–4); texts in the field of solid-state pharmaceutical chemistry also include detailed discussions and applications of SSNMR (5–8). This chapter attempts to summarize the current status and latest developments in the field of applied pharmaceutical SSNMR from an industrial perspective, focusing on the utility of the technique during the development of a drug from the candidate stage through to a marketed, approved product. Particular attention is given to the latest approaches to the

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing,*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

use of SSNMR in pharmaceutical analysis, including detailed structural analysis of active pharmaceutical ingredients (APIs) and analysis of formulations.

The solid-state structure of a pharmaceutical product can influence its properties, manufacturability, and stability. Since these characteristics can impact product performance and patient safety, regulatory agencies require control over solid-state structure and subsequent properties (9–11). Solid-state analysis of pharmaceuticals is a challenging field and a number of complementary techniques are frequently needed to understand the complete system. Typically, knowledge of the API polymorphic forms, salts, and solvates, as well as the physical characteristics of excipients, granulations, tablets, parenteral products, and other dosage forms, is built during the drug development process. SSNMR can be a critical part of an integrated analytical strategy for addressing the questions faced in the development process because of its high information content and flexible experimental design (12). Subsequently, analytical methods are developed to enable rapid testing of many batches and ensure control. Although SSNMR methods can be quite sensitive and have many other desirable properties, the complexity and expense associated with current instrumentation often lead to the use of X-ray powder diffraction (XRPD) and the various forms of vibrational spectroscopy when routine analysis is required (12). However, in some cases, SSNMR can be applicable to the testing of a wide range of batches, particularly to detect and understand subtle variations in solid-state structure that are not detected by lower resolution techniques, or when high sensitivity drug product analysis is needed.

SSNMR is closely related to the widely used technique of solution-state NMR and shares many of the same features (13). The theory used to describe SSNMR experiments and their results is identical to that of solution-state NMR, but includes special consideration of three NMR interactions that have particular manifestations in the solid state: the anisotropic chemical shift, magnetic dipolar coupling, and quadrupolar coupling (14,15). Because of these interactions, specialized experimental methods and pulse sequences, such as magic-angle spinning (MAS) and cross-polarization (CP), are widely used and thoroughly permeate any discussion of SSNMR (16).

This chapter reviews the current status of applied pharmaceutical SSNMR, first focusing on the latest innovations in technology and methodology of relevance to pharmaceutical analysis and then providing an overview of the present and future applications of this technology and methodology to API, formulated products, and excipients, illustrated with examples from current practice. SSNMR has advanced considerably in the past decade, with multinuclear and multidimensional experiments now commonly encountered in pharmaceutical applications (17).

## 2 METHODS

### 2.1 Accessible NMR Nuclei in Pharmaceuticals

Experimental methods for SSNMR spectroscopy of pharmaceuticals are closely related to the properties of the nuclei of interest. The isotopes generally used in

pharmaceutical applications are summarized in Table 1. Although the $^{13}C$ isotope is the dominant nucleus used in pharmaceutical applications of SSNMR, other nuclei have begun to play a wider role. Pharmaceutical SSNMR analyses generally rely on the natural abundance of the isotopes given in Table 1. Isotopic spin labeling, although it is predominant in other applications of SSNMR, is not commonly used in pharmaceutical SSNMR analysis. However, spin labeling has been used in certain special applications requiring high sensitivity, for example, to detect the form of an API or an excipient at levels of $<1\%$ w/w (12,18). The pharmaceutical relevance, key NMR properties, and experimental NMR parameters for each nucleus in Table 1 are discussed below; applications of each nucleus to various analytical issues in pharmaceutical development are discussed in Section 3.

***2.1.1 Carbon-13*** Carbon forms the backbone of most pharmaceutical molecules and has reasonable sensitivity and natural abundance. The $^{13}C$ nucleus is well positioned to report on changes in the 3D structure of a drug molecule or excipient, including such effects as conformation, hydrogen bonding, packing interactions, and ionization. As a result, $^{13}C$ is by far the most commonly observed NMR isotope in pharmaceutical SSNMR applications. Extensive literature covering the empirical chemical shift parameters for $^{13}C$ is available from solution-state studies (19). The $^{13}C$ SSNMR spectra of pharmaceutical solids are generally referenced to TMS via an external reference such as glycine or hexamethylbenzene (20) and can serve as the base reference for other nuclei studied by a SSNMR system using ghost referencing (21). The $^{13}C$ chemical shift tensor (CST), which offers information beyond that available from isotropic shifts alone and is discussed in detail below, has also been extensively studied and related to structural trends (22). The $^{13}C$ isotope often shows large, interpretable effects in the principal values of its CST; this is often referred to as strong chemical shift anisotropy (CSA) (22).

***2.1.2 Hydrogen Isotopes*** Hydrogen, the other backbone element of organic pharmaceuticals, contains two isotopes of particular interest in pharmaceutical SSNMR. The $^{1}H$ isotope is a sensitive, highly abundant spin 1/2 nucleus, as shown in Table 1. However, its relatively narrow chemical shift range and ability to engage in strong dipolar coupling cause most $^{1}H$ SSNMR spectra of pharmaceuticals to be broad and featureless unless special methods are used (23). Dipolar coupling arises because of the through-space interaction of the magnetic moments of one nucleus with other nuclei that are spatially proximate (14–16). The magnitude of this interaction between $^{1}H$ nuclei can reach up to 50 kHz and is significantly stronger than the weak and difficult to measure $^{1}H$ CSA. The dipolar interaction provides an avenue to obtain a wealth of structural information when it can be removed and selectively reintroduced by specialized SSNMR experiments (23). Hydrogen also exists as a spin 1 isotope $^{2}H$ (deuterium), a quadrupolar nucleus that can also be used in the study of pharmaceuticals. Although not generally useful in natural abundance, $^{2}H$ exchange in readily available deuterated solvents and reagents can be achieved, so that partial incorporation of deuterium into pharmaceutical solids

**TABLE 1 Properties of Pharmaceutically Relevant NMR Active Nuclei[a]**

| Isotope | Spin ($I$) | Frequency in MHz at 2.344 T | Natural Abundance | Receptivity [b] | Notes |
|---|---|---|---|---|---|
| $^{1}H$ | 1/2 | 100.000 | 99.985% | 1.0000 | Narrow chemical shift range, very sensitive. Historically used infrequently due to large solid-state linewidths, but becoming more prevalent because of higher fields and increased resolution from dipolar decoupling and fast MAS |
| $^{2}H$ | 1 | 15.350 | 0.015% | 0.00000145 | Too insensitive for natural abundance, but introduced through isotope labeling. Can be used to study hydrates and water exchange |
| $^{11}B$ | 3/2 | 32.084 | 80.1% | 0.132 | Useful for characterization of boron-containing pharmaceuticals |
| $^{13}C$ | 1/2 | 25.145 | 1.10% | 0.000170 | The most widely used and informative nucleus for pharmaceutical analysis |
| $^{15}N$ | 1/2 | 10.137 | 0.366% | 0.00000384 | Low abundance, insensitive, and useful for determining ionization state |
| $^{17}O$ | 5/2 | 13.556 | 0.038% | 0.0000111 | Only accessible via isotope labeling. Can be used to assess the role of water in pharmaceutical hydrates and water exchange |
| $^{19}F$ | 1/2 | 94.094 | 100.0% | 0.834 | Sensitive, present in many pharmaceuticals but not in excipients, and large chemical shift range |
| $^{23}Na$ | 3/2 | 26.451 | 100.0% | 0.0927 | Analysis of salts of API acids, also present in excipients |
| $^{27}Al$ | 5/2 | 26.056 | 100.0% | 0.207 | Present in some pharmaceutical excipients and in vaccine adjuvants |
| $^{29}Si$ | 1/2 | 19.867 | 4.67% | 0.000367 | Present in some pharmaceutical excipients |
| $^{31}P$ | 1/2 | 40.480 | 100.0% | 0.0665 | Sensitive, useful for phosphate salts and phosphorus-containing pharmaceuticals, and also present in some excipients |
| $^{77}Se$ | 1/2 | 19.071 | 7.63% | 0.000537 | Useful for characterization of selenium-containing pharmaceuticals |

[a]Values are from Ref. (21).
[b]Relative to $^{1}H$ at 1.0000.

is straightforward. In many cases, the deuterated solvent itself is of interest, if it can be incorporated into drug hydrates and solvates, polymeric excipients, and many other pharmaceutical materials as bound or surface solvent for studies that track moisture or examine its physical state (24–26). Deuterium may also receive increased attention for solid form analysis as deuterated drug molecules advance in clinical development (27).

***2.1.3 Nitrogen-15*** Nitrogen is frequently present in pharmaceutical molecules, in the guise of a wide variety of amino, amido, imino, and heterocyclic functional groups (28), and new applications of nitrogen-containing functional groups in medicinal chemistry are constantly expanding (29). Amines are commonly used to add ionizable groups and affect solubility and membrane permeability, while also allowing the formation of salts (30). The spin 1/2 nucleus $^{15}N$ is commonly studied despite its low natural abundance; extensive tabulations of experimental $^{15}N$ NMR data are available from solution-state NMR studies (31). Typically, $^{15}N$ SSNMR data are externally referenced to either ammonia or nitromethane via solid glycine or ammonium chloride samples (21). Because it has a spin of 1, has a strong quadrupolar broadening, and lacks a central transition, the quadrupolar $^{14}N$ nucleus is not commonly studied by SSNMR; however, new developments in $^{14}N$ SSNMR may ultimately make it accessible for pharmaceutical studies (32). The closely related technique of nuclear quadrupole resonance (NQR) is also emerging as a potential method for analysis of this isotope in pharmaceuticals (33). The effects of the $^{14}N$ isotope are commonly seen in the $^{13}C$ spectra of pharmaceuticals, as are the effects of other quadrupolar nuclei, through residual dipolar coupling effects (34).

***2.1.4 Fluorine-19*** Fluorine is encountered in a growing number of marketed and research pharmaceuticals, particularly because of its ability to improve metabolic stability to common liver enzymes; one analysis from 2002 indicates that 30% of the new chemical entities approved in the United States were fluorinated molecules (35,36). Aryl fluorines and $CF_3$ groups are commonly found in pharmaceuticals. Fluorinated steroid molecules containing aliphatic CF groups are also frequently encountered. With a natural abundance of 100% and a high gyromagnetic ratio, plus high selectivity and freedom from interferences and background (37), $^{19}F$ is an ideal nucleus for high-sensitivity SSNMR analysis of fluorinated API. $^{19}F$ has a wide chemical shift range, covering 1000 ppm, and a relatively high CSA (31,37). Although most pharmaceuticals contain only a few fluorine atoms, there is the potential for strong homonuclear dipolar couplings (37). $^{19}F$ spectra are referenced to liquid $CFCl_3$ via a number of procedures, including ghost referencing, as there is no accepted external solid reference standard (21).

***2.1.5 Other Nuclei of Interest*** $^{31}P$ is another nucleus with a natural abundance of 100% and other generally favorable NMR properties that are useful in pharmaceutical analysis. $^{31}P$ has a chemical shift range covering $\sim$75 ppm for pharmaceutical samples and can have a relatively high CSA. Phosphorus is most

commonly observed in phosphate salts (30) and in some pharmaceutical excipients, although a number of pharmaceutical compounds contain organic phosphorus, including the ACE inhibitor fosinopril (38), fosamprenavir, auranofin, and the osteoporosis medication alendronic acid. Selenium has recently started to appear in drug molecules (39); it has a favorable spin 1/2 nucleus ($^{77}$Se) as shown in Table 1. Colloidal silicon dioxide and other silicate excipients can be probed using $^{29}$Si SSNMR, which has the added benefit of being applicable to studies of stationary phases widely employed in pharmaceutical chromatographic analysis.

Other NMR active nuclei that appear in pharmaceuticals are often quadrupolar isotopes. Sodium is frequently encountered in pharmaceutical materials, both in excipients and in sodium salts of drug molecules, and is amenable to study using $^{23}$Na SSNMR methods (12,40). Boron-containing drugs can benefit from SSNMR studies of the quadrupolar $^{11}$B nucleus, which can readily distinguish trigonal and tetrahedral coordination among other useful parameters (41). Boron-containing drugs have recently been the subject of renewed interest through a growing range of medicinal chemistry approaches, culminating in the recent marketing of bortezomib, an organoboronic acid proteasome inhibitor (42–44). With the advent of higher static magnetic fields, other nuclei commonly found in pharmaceutical salts, such as $^{35}$Cl and $^{79}$Br/$^{81}$Br, have also become accessible (45). Aluminum, accessible via the $^{27}$Al nucleus, also appears in excipients and particularly in the adjuvants used in numerous vaccines (46). Like $^{2}$H, $^{17}$O is a useful probe of water incorporation through solution- or vapor-phase exchange into crystalline and amorphous drugs and excipients and can also be labeled directly into molecules of interest, albeit at a greater expense (24,47,48). Finally, other nuclei of pharmaceutical interest, such as $^{33}$S, may become increasingly accessible as fields continue to trend higher (49).

## 2.2 Experimental NMR Methods

***2.2.1 Instrumentation*** The current design of SSNMR instrumentation closely parallels the technology used in solution-state NMR. As a consequence, most modern SSNMR systems utilize digital filtering and digital signal generation to produce radiofrequency (RF) pulses and detect NMR signals (50–52). Digital receivers have greatly improved spectral baseline while adding dynamic range and sensitivity to modern SSNMR experiments. Direct digital synthesis of RF signals allows very rapid frequency and phase shifts and has made a number of specialized SSNMR experiments of relevance to pharmaceutical analysis far easier to implement, such as fast frequency shifts on the order of $\pm 100$ kHz as used for homonuclear decoupling (see below).

The primary difference between solution-state NMR and SSNMR instrumentation is the probe. Standard MAS probe designs have been available for many years; most are designed to operate in wide-bore (89 mm) NMR magnets. Unlike solution-state NMR probes, SSNMR probes usually employ a single solenoidal inductor for excitation and detection of multiple nuclei (53). The design of the tuning circuit, and in particular its power handling and isolation, is critical to the eventual performance of SSNMR experiments. The engineering of the stator and rotor for

stable, high-speed MAS is also a critical consideration, as is the ability of the probe to change temperatures over a wide range. Sizes of outer diameters of common MAS rotor vary from 7.5 to 2.5 mm and can hold sample masses of 400 to 5 mg depending on the internal volume and the use of restrictor plugs. Static probes are also utilized for specialized applications, including $^{2}$H and $^{17}$O wideline SSNMR (53) and high-pressure and wide-temperature range NMR studies (54). Triple-tuned single-solenoid probes and filters that can simultaneously access $^{1}$H, $^{19}$F, and an X-nucleus with good isolation and excellent power handling, using transmission-line designs, are a recent development of special importance, given in the large number of drug molecules that contain fluorine (55).

A number of innovative rotor and stator designs have been developed recently that have seen immediate applications to pharmaceuticals. MAS stators and rotors that are capable of spinning to speeds between 35 and 70 kHz have been developed for $^{1}$H NMR applications, albeit using very small sample quantities (23,56). Because of the commercial availability of ultrafast MAS probes, $^{1}$H MAS experiments at 70 kHz on pharmaceuticals are now possible. Stator designs that flip up to allow easy sample ingress and egress, for narrow bore (54 mm) NMR magnets, are also commercially available and widely used. Probes containing several spinning modules that can be cycled through the most homogeneous central area of the magnet have also been developed; these can be beneficially applied to samples with long $^{1}$H $T_1$ relaxation times so that sequential experiments can be conducted on the stator at the center point while the others relax, allowing for optimal use of the expensive spectrometer (57).

***2.2.2 Routine Experiments*** The basic experiments used to acquire 1D SSNMR spectra and obtain initial chemical shift information are direct-polarization (DP) and CP experiments (16). In pharmaceutical applications, both are commonly combined with $^{1}$H decoupling and MAS. DP–MAS experiments, also referred to as "single pulse" or "pulse and acquire," are commonly used for high-sensitivity nuclei, including $^{1}$H, $^{19}$F, and $^{31}$P, and have the advantage of being both simple and quantitative (if a sufficient relaxation delay is used); these experiments can also be used to study solution fractions of wet solid samples, as discussed in detail below. However, in general, the relaxation rate and the natural abundance of the $^{13}$C and $^{15}$N nuclei prohibit the routine use of DP–MAS. CP–MAS experiments take advantage of the dipolar coupling of the X-nuclei to abundant spins such as $^{1}$H and enhance the sensitivity and reduce relaxation delay by transferring magnetization from those spins to the less-abundant and sensitive spins, the latter benefit being possible because the relaxation rate is determined by the typically efficient relaxation rate of the $^{1}$H spin bath (16). Magnetization transfer, or contact times, generally range from 1 to 10 ms and can lead to semiquantitative signal intensities (58). CP is also the basis of many 2D NMR correlation experiments.

Because of CSA, and the potential overlap of sidebands with signals of interest at typical field strengths and spinning rates particularly with $^{13}$C SSNMR (such as carbonyl groups with aliphatic groups), sideband suppression methods such as the total sideband suppression (TOSS) method are often applied (59). After the application of CP and TOSS, for example, in $^{13}$C SSNMR, the resulting spectrum contains

signals that can be directly assigned without the potential confusion that occurs with small sidebands. Because the Hartmann–Hahn condition is split into a series of sidebands under the influence of MAS (16), the matching condition can be sensitive to sampling conditions; this effect is ameliorated by the use of ramped-amplitude CP methods (60). Thus, an experiment that is well suited for $^{13}C$ SSNMR in pharmaceutical analysis is the CP–TOSS experiment with a ramped $^{1}H$ or $^{13}C$ contact time pulse.

Decoupling plays a key role in many SSNMR experiments and is particularly relevant to hydrogen-rich pharmaceutical compounds. Currently, the most popular heteronuclear decoupling schemes for pharmaceutical solids (useful for decoupling, for example, $^{1}H$ when observing $^{13}C$) are the two-pulse phase modulation (TPPM) scheme and its derivatives, such as SPINAL-64 (61). In pharmaceutical SSNMR of the $^{13}C$ nucleus, these decoupling schemes perform most efficiently at $^{1}H$ decoupling strengths of $\sim$100 kHz; as a result, smaller diameter MAS probes in the 4–5 mm range are widely used because of their ability to produce RF fields in this range. Homonuclear decoupling schemes that decouple $^{1}H$ nuclei from each other, such as the frequency-switched Lee–Goldburg (FSLG) method, are also in widespread use for observation of $^{1}H$ in both 1D and 2D experiments (62). Windowed sequences are available for direct observation of $^{1}H$ spectra if a full 2D measurement is not desired, particularly using the phase-modulated Lee–Goldburg (PMLG) method (63,64). The performance of homonuclear decoupling schemes at higher MAS spinning rates is becoming more important with the trend toward faster speeds; sequences such as PMLG are known to perform well at speeds approaching 65 kHz (65). Newer $^{1}H$ homonuclear decoupling techniques, such as the DUMBO method, also perform well at high speeds (66).

The forces involved in MAS are known to occasionally interconvert solid forms of certain pharmaceutically related compounds. For example, solvates and hydrates can be interconverted by the centrifugal force or frictional heating of MAS (67,68). The use of cooling in the sample area can help ameliorate these effects, if they occur.

***2.2.3 Spectral Editing and Assignment*** Several spectral editing experiments are available, primarily for identifying $^{13}C$ and $^{15}N$ positions by their number of attached protons. The oldest method for spectral editing in organic solids, and still the most widely used, is the dipolar dephasing experiment (69). This experiment, also known as nonquaternary suppression (NQS), utilizes $^{1}H$–$^{13}C$ dipolar couplings to dephase transverse magnetization during an echo period in the experiment; $^{13}C$ sites with stronger $^{1}H$–$^{13}C$ dipolar couplings dephase more rapidly and are removed from the spectrum. In practice, this experiment eliminates signals from methine and methylene carbons and is useful for isolating quaternary carbon aromatic signals in pharmaceuticals for detailed analysis. An $^{15}N$ version of the NQS experiment is also easily implemented, an example of which is shown in Section 3. A variety of more sophisticated SSNMR spectral editing experiments have been developed over the past 15 years; these are based on polarization inversion (PI) (70) or on multiplicity assignment via *J* couplings (71,72). These experiments allow separation of CH, $CH_2$, and $CH_3$ signals, enabling assignments to be made on

overlapped positions. The primary limitation of these experiments is the need to carefully calibrate the PI period or the need to test a range of *J* delays because of the variable performance of homonuclear decoupling sequences. The main application of these experiments in pharmaceutical SSNMR is to refine $^{13}C$ assignments, particularly in molecules with overlapped aliphatic groups, such as steroids and macrolides, where overlap between CH and $CH_2$ signals is significant.

### 2.2.4 2D SSNMR Experiments for Structural Characterization

Numerous 2D SSNMR experiments are available for the detailed structural characterization of solid materials. A list of the experiments that are both robust enough for routine use and have general relevance to pharmaceutical analysis is given in Table 2. In general, these experiments fall into two primary categories: those that probe the dipolar coupling and those that report on the CST.

Experiments that study the dipolar coupling make use of either the inverse cubic dependence of the direct magnetic dipolar coupling between nuclei or longer range dipolar spin diffusion effects (73). The direct dipolar coupling can be divided into heteronuclear and homonuclear couplings, both of which are applicable to pharmaceutical analysis. Most of these experiments are performed as 2D experiments, which provide a connectivity map between nuclei (74,75). Homonuclear and heteronuclear correlation (HETCOR) experiments can correlate a variety of nuclei, including $^{1}H$, $^{13}C$, and $^{19}F$, in pharmaceutical systems that include polymorphs and cocrystals (76,77). Several examples are given in Table 2. Many dipolar experiments involve the $^{1}H$ nucleus, so that the use of homonuclear dipolar decoupling sequences can be a critical feature of their performance. Spin diffusion effects can also appear during HETCOR experiments and can be useful in detecting molecular association (77), and they can also be suppressed when necessary to focus on direct dipolar transfers (78).

The second subclass of experiment that utilizes dipolar coupling effects, those that rely on spin diffusion (73), is also useful for structural analysis of organic and pharmaceutical molecules (79,80). Long-range distance information can be extracted by modeling spin diffusion rates, such that the proximity of multiple proton positions can be determined from the same experiment. Currently, spin diffusion approaches to crystal structure solution have been confined to molecules that are relatively small in size compared to most pharmaceutical compounds, although they hold promise for larger molecules (81). Because of the challenging nature of the analysis, 2D SSNMR experimental methods used for crystal structure determination are often used in conjunction with computational methods and other experimental approaches (such as XRPD powder solution methods) (76,81,82). The computational methods used for this purpose are discussed in more detail in a later section.

The other major class of experiments involves measurement of the three principal components (PCs) of the CST. The CST provides a measure of the spatial dependence of the chemical shift interaction (83). In powdered samples, as is usually the case for pharmaceutical samples, the three principal components of the CST are accessible; these are designated by $\delta_{11}$, $\delta_{22}$, and $\delta_{33}$. A large number of

**TABLE 2 Summary of 1D, 2D, and Pseudo-2D Experiments Commonly Used in Pharmaceutical SSNMR Analyses**

| Experiment | Short Description | Typical Duration | References |
|---|---|---|---|
| $^{1}$H DQ-BABA | 2D dipolar correlation between hydrogen bonding $^{1}$H and neighboring $^{1}$H nuclei, useful for polymorph interrelationships and cocrystals | 1–2 h | (74,76,77) |
| $^{1}$H spin diffusion with CRAMPS | 2D spin diffusion between $^{1}$H nuclei, with effects ranging up to 30 Å, for structural analysis | 1–2 h | (23,79,80) |
| $^{19}$F (or other X-nucleus) spin diffusion | 2D heteronuclear DARR spin diffusion, for pharmaceutical phase purity assessment | 8–12 h | (146) |
| $^{13}$C-detected $^{1}$H $T_1$ saturation recovery | Determination of the $^{1}$H spin lattice relaxation rates ($T_1$) by indirect detection via the $^{13}$C spectrum, for pharmaceutical phase purity assessment and phase mixing experiments | 12 h | (93) |
| $^{1}$H–$^{13}$C CP-HETCOR | 2D detection of dipolar connectivity between $^{1}$H and $^{13}$C, shows spin diffusion when used with longer ramp-CP mixing (800 μs–2 ms), and useful for polymorph interrelationships and cocrystals | 12 h | (75–77,170) |
| $^{1}$H–$^{13}$C LGCP-HETCOR | Detects dipolar connectivity between $^{1}$H and $^{13}$C, minimizes spin diffusion, and useful for solid form interrelationships and assignments | 12 h | (78) |
| $^{1}$H–$^{19}$F LGCP-HETCOR | Detects dipolar connectivity between $^{1}$H and $^{19}$F for fluorinated pharmaceuticals, minimizes spin diffusion | 1 h | (77,78,155) |
| $^{1}$H –$^{19}$F–$^{13}$C DCP-TOSS | 1D experiment for fluorinated pharmaceuticals, detects dipolar connectivity between $^{13}$C and $^{19}$F. | 8 h | (76,77) |
| $^{13}$C CP-PASS | Chemical shift anisotropy ($^{13}$C) for studies of hydrogen bonding and ionization, and for comparison with calculations | 18–24 h | (59,76,84) |
| $^{13}$C CP-FIREMAT | Chemical shift anisotropy ($^{13}$C) for studies of hydrogen bonding and ionization, and for comparison with calculations | 18–24 h | (77,85) |
| $^{13}$C CP-SUPER | Chemical shift anisotropy ($^{13}$C) for studies of hydrogen bonding and ionization, and for comparison with calculations. | 18–24 h | (86) |

The references given include both the original technical implementations of the pulse sequence and subsequent applications to pharmaceutical or similar organic molecules with spectral interpretation.

experiments have been developed for measurement of the principal components in complex samples, but only a few are suited for pharmaceutical applications. The cross-polarization phase-adjusted spinning sidebands (CP-PASS) method was one of the earlier robust methods to be developed and applied to pharmaceuticals (59,84). CP-PASS has been used for the determination of a pharmaceutical crystal structure from a powdered sample (76). The CP-FIREMAT experiment (85)

also provides a highly accurate method for CST analysis and has been applied to a pharmaceutical cocystal (77). The robust CP-SUPER method has also been demonstrated on organic molecules (86).

Besides these two major classes of structural analysis experiments, other 2D SSNMR experiments are useful for pharmaceutical analysis. In particular, the multiple-quantum MAS (MQMAS) experiment is useful for studies of quadrupolar nuclei with moderate quadrupolar interaction strengths, such as $^{17}O$, $^{23}Na$, and $^{27}Al$ (87,88). The MQMAS experiment produces high-resolution isotropic spectra while also allowing access to quadrupolar parameters and has the added benefit of being easily implemented on most modern SSNMR hardware using fast-spinning probes, including systems optimized for pharmaceutical analysis.

***2.2.5 Dynamics and Relaxation Experiments*** Like solution-state NMR, SSNMR is uniquely capable of the study of motion in pharmaceutical systems. Variable temperature NMR plays a critical role in these experiments, as it readily distinguishes static disorder from apparent disorder actually caused by molecular dynamics. Various NMR relaxation measurements can probe motions of different timescales in pharmaceutical molecules and are useful when dynamic motion in the solid state is of interest (5,16).

Samples with longer $T_1$ times may benefit from the use of paramagnetic relaxation enhancing compounds, such as copper sulfate. However, this generally requires recrystallization of the solid form of interest with the relaxation agent and is thus primarily applicable to readily prepared API forms. The benefits of these approaches are much faster data acquisition; such a benefit has recently been demonstrated for protein samples studied with 2D SSNMR (89).

## 2.3 Data Analysis and Computational Methods

***2.3.1 NMR Data Processing*** SSNMR data processing is similar in many respects to that employed in solution-state NMR, given the requirements of a Fourier transform spectroscopic method. In particular, apodization and linear prediction are both widely used in pharmaceutical SSNMR analyses (90). Linear prediction is extremely useful in 2D SSNMR analyses as it allows enhanced resolution in the $F_1$ dimension, helping to optimize signal-to-noise ratio on longer experiments (90). Similarly, although Hadamard transform-based methods are only in limited use, they allow more efficient 2D data collection in SSNMR (91). More advanced processing methods can be applied when needed to handle data produced by specialized experiments, including those that specifically probe dipolar and quadrupolar interactions (92). For example, experiments that produce exponential decays (such as $^1H$ $T_1$ relaxation measurements) can be analyzed using nonlinear fitting methods or the more sophisticated DECRA method; the latter approach is particularly useful for analysis of pharmaceutical phase mixtures (93).

Lineshape fitting and deconvolution are widely used methods for fitting SSNMR spectra of pharmaceuticals. There are a variety of reasons to fit lineshapes, most notably to extract chemical shift and quadrupolar parameters (quantitative aspects of

lineshape fitting are also of great importance, and they are discussed in Section 2.3.2). The principles of lineshape fitting and general applications to SSNMR have been well documented in the literature (94,95). One of the most common uses of lineshape fitting is to extract CST values from MAS spinning sideband patterns. The major disadvantage of lineshape fitting is that a prior knowledge of the spin system (e.g., the number of nuclei) must be known. Software programs are widely available for this task, including both commercial packages and freely available packages from academic laboratories (96).

***2.3.2 Quantitative Analysis*** Quantitative analysis using SSNMR is readily accomplished but requires special care, as is also the case with XRPD, DSC, and vibrational spectroscopy. SSNMR has an inherent advantage over XRPD, DSC, and vibrational spectroscopy in that the signals observed even in CP experiments are approximately quantitative in nature, and the specificity and resolution is such that the need for multivariate analysis (MVA) is minimized. When quantitative analysis with SSNMR is being performed, relaxation parameters, spectral resolution, and spectral phasing must be considered. Spin lattice relaxation ($T_1$) is one of the most crucial relaxation parameters to consider in a quantitative experiment. A recycle delay of at least five times the longest $T_1$ of all components of a sample will ensure full relaxation, although a delay of three times $T_1$ is sufficient to achieve 95% relaxation. For optimum sensitivity of a particular component, a recycle delay of $1.3 \times T_1$ can be used for a 90° flip angle (13). Quantitative results can still be calculated from the integral of a non-fully relaxed spectrum by considering the Bloch equations of steady-state and equilibrium magnetization (13,97). Quantitative results, without the requirement of a calibration curve, can be obtained from the CP–MAS (or CP–TOSS) spectrum if the $^1$H spin lattice relaxation in the rotating frame ($T_{1\varrho}$) and the CP rate constant ($T_{CH}$) are known (98,99). These parameters can be obtained by fitting the CP intensity at various contact times to a kinetic model, from which an extrapolation to "zero contact time" is obtained (99,100).

A detection limit of about 1% w/w using $^{13}$C CP experiments is achievable in favorable cases. For example, a detection limit of 2–3% w/w was reported for polymorph mixtures of delavirdine mesylate (98), whereas a detection limit of 0.5% w/w was achievable for crystalline forms of neotame (99). The detection limit can be experimentally determined from a sample of known concentration as three times the concentration divided by the signal-to-noise ratio measured on the sample (101). Similarly, the limit of quantitation can be estimated as 10 times the signal-to-noise ratio.

When a sample of interest shows poor spectral resolution, because of a lack of crystallinity or a short transverse relaxation time, deconvolution can be used to retrieve overlapped lineshapes for quantitative purposes. However, excessive line broadening can cause a deconvolution analysis to converge to a false minimum, leading to inaccurate results. In such cases, using the fitted linewidth from another resolved peak in the spectrum and then constraining the linewidth of the unresolved component to the same values may be sufficient to allow recovery of the correct lineshape and improve the accuracy of results. Another difficulty arising with

overlapped peaks is spectral phasing, which is subjective and can affect the quantitative results.

Provided that all influence factors to data quality have been carefully considered, the integral of NMR signal is inherently quantitative and can be used directly for analysis without the requirement of a gravimetric calibration curve. Though linewidths of SSNMR signals are substantially broader than that in solution state, the resolution usually suffices for simple integration methods. In cases where signals of different components are overlapped, for example, in amorphous or disordered materials, deconvolution or other methods may be required to obtain more accurate integrals. An example of theoretical lineshape deconvolution for the quantitative analysis of a mixture of an amorphous and an overlapped crystalline phase is shown in Figure 1. Using this $^{19}$F SSNMR method, it is possible to rapidly quantify levels of amorphous content to approximately 5% w/w using a short experiment time (of typically less than 10 min). This process can also be applied to mixtures of crystalline phases, and additional parameters such as CSA and quadrupolar coupling may also be included in fits as needed.

Finally, two other methods are useful for quantitative analysis. In solution-state NMR, principal component analysis (PCA) (102) and independent component

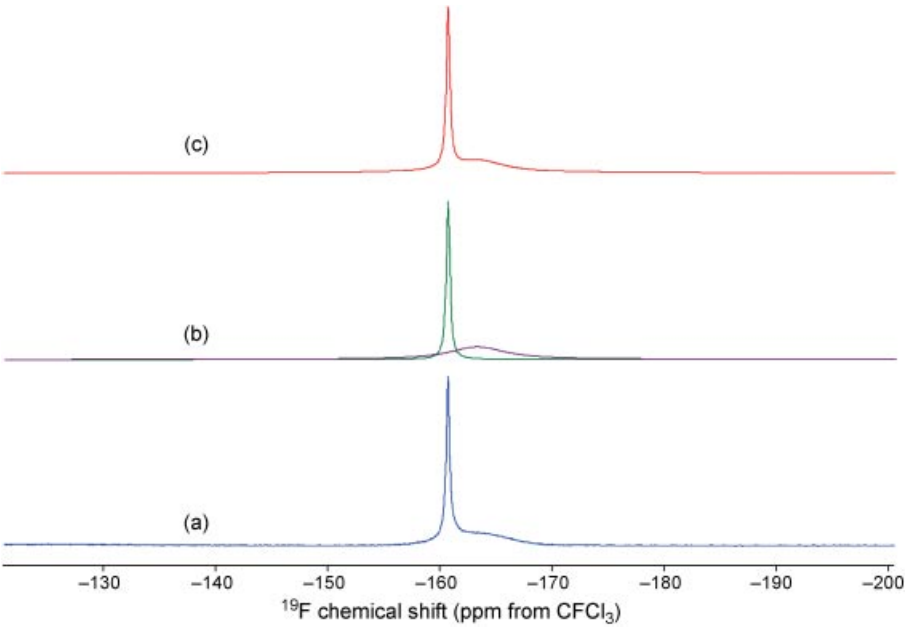


***FIGURE 1*** *Lorentzian/Gaussian lineshape deconvolution of the* $^{19}$F DP–MAS spectrum of a mixture of 50/50 amorphous and crystalline phases. (a) Experimental spectrum, (b) individual subspectra fitted to the two components, and (c) addition spectrum of the two subspectra. The experimental spectrum was obtained at 8.4 T and 273 K.

analysis (ICA) (103) have been commonly used for quantifying overcrowded spectra of chemical mixtures (104,105) and metabolites (106,107). PCA is a statistical method that analyses a large set of experimental spectra to pull out the PCs, which for SSNMR are generally the pure phase spectra. The magnitude of a PC in an individual experimental spectrum provides a quantitative value for that component. Availability of pure phase reference materials can reduce PCA to ICA, where only a small set of reference spectra are required. The ICA method involves the least-squares regression of a reference spectrum to an experimental data set. PCA has been applied to SSNMR in other areas outside of pharmaceutical analysis (108). An example of the use of ICA for amorphous quantification in a crystalline API is shown in Figure 2. The advantage of ICA is exactly the opposite of theoretical deconvolution; that is, a prior knowledge of the spin system is not required for quantitative analysis, although the

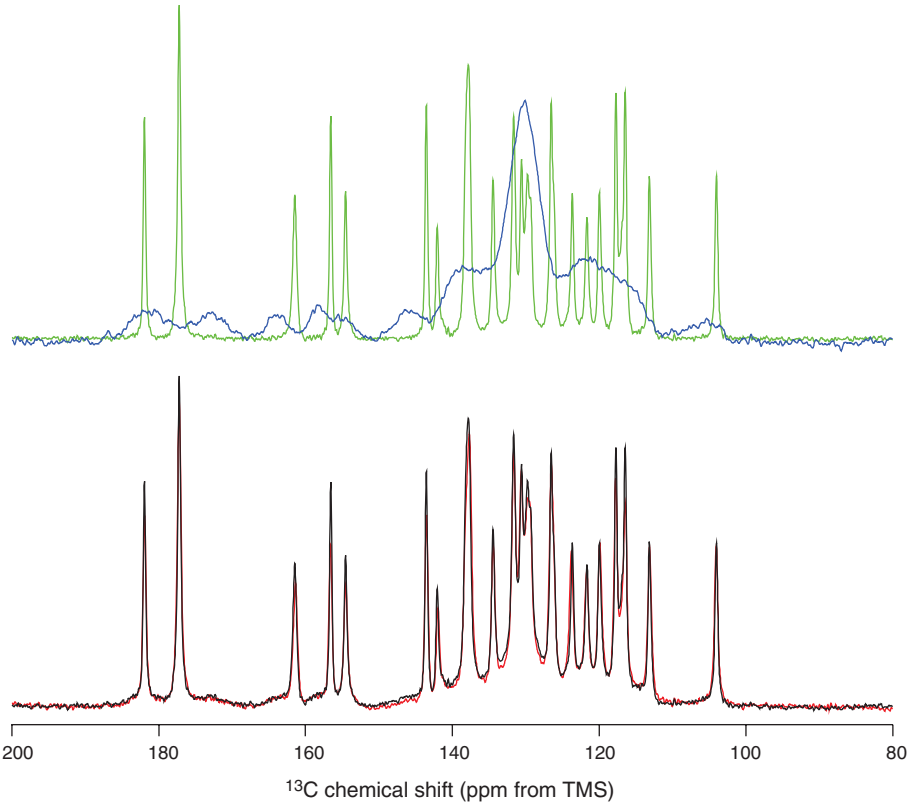


***FIGURE 2*** *Quantification of amorphous content in a crystalline API from a* $^{13}$C CP–TOSS spectrum by least-squares regression of reference spectra (using the ICA method). The top traces show normalized spectra of crystalline (green) and amorphous (blue) reference samples. The bottom trace is an experimental spectrum of a 20% amorphous mixture (black) overlaid with an ICA calculated spectrum (red). Spectra were obtained at 8.4 T and 273 K.

method requires phase-pure reference samples. In addition, caveats for the use of ICA include ensuring that the linewidth and phasing of the references, particularly the crystalline reference, are as similar to the experimental spectrum as possible.

***2.3.3 Prediction of NMR Properties*** Computational methods can reliably reproduce SSNMR parameters from molecular models, such as the chemical shift, the spin–spin coupling, and the electric field gradient (EFG) (109), and greatly facilitate the interpretation of experimental CST and EFG differences between solid phases (110). Computational methods aid in the understanding of molecular conformation, the assessment of hydrogen bonding and π-stacking (76), the assignment of ionization states, the assignment of resonances (111,112), and the elucidation of powder structures (76,113,114), and they can often provide a link between low-temperature crystal structures and room temperature bulk material. Although many studies have focused on the prediction of $^{13}C$ and $^{15}N$ chemical shifts, studies on other nuclei are also useful (109). A brief overview of the computational methods useful for pharmaceutical SSNMR is provided here; more details can be found in recent reviews (109).

Recent improvements in chemical shift prediction in particular have been driven by the availability of computational resources and the further development of theoretical methods. This in turn has enhanced the applicability of these methods to pharmaceutical SSNMR. Although simple Hartree–Fock (HF) calculations are sufficient to reproduce chemical shifts when electron correlation effects are small, such as in the $^{13}C$ shift prediction of aliphatic carbons, post-HF methods are required to accurately predict chemical shifts when correlation effects become significant, as is the case for most molecules of pharmaceutical interest (109). Although expensive perturbation and coupled cluster methods can account for these effects, they are often not applicable to larger molecular structures as encountered in pharmaceutical research and development. As a result, density functional theory (DFT), which also accounts for electron correlation, has been shown to accurately predict chemical shifts and field gradients, especially using hybrid functionals such as B3LYP (109). The computational requirements of DFT methods scale more favorably with the number of electrons ($N^4$), so that computations on typical pharmaceutical molecules are feasible on desktop computers. For Gaussian basis set methods, full double zeta and triple zeta basis sets including polarization and diffuse functions are recommended for reliable results (109).

In addition to improvements in molecular orbital theory, various methods have been developed to include solid-state effects such as steric interactions, long-range electrostatic interactions, collective intermolecular hydrogen bonding, ring current effects, and band structures. These include simple cluster methods (109), as well as methods based on periodic boundary conditions, Ewald summation, Madelung potentials, plane wave basis sets (115), and ONIOM methods with localized basis sets (116). These methods are often needed, since solid-state effects in pharmaceutical compounds can alter the computed wavefunctions and consequently the computed NMR parameters (76,77). The cluster approach includes neighboring molecules in the calculations to describe the intermolecular effects. This approach has been shown to

yield satisfactory $^{13}C$ and $^{15}N$ chemical shifts when only short-range interactions are present, such as steric interactions or intermolecular hydrogen bonding (76,110). Large charge separations within the molecule of interest can result in significant long-range Coulomb interactions. Further improvement over the simple cluster approach may be achieved by including these interactions with the ONIOM method that includes a shell of point charges beyond the molecule of interest. However, the finite size of the point charge array can still lead to significant truncation effects because of the poor convergence of Coulomb lattice sums. The electrostatic interactions in molecular crystals may be included by fitting point charges to reproduce the Madelung potential at an area of interest in which the quantum mechanical calculations are performed, as implemented in the embedded ion method (117). These pure point charge methods are computationally less demanding and can be performed on large unit cells and large molecules with limited computational resources, and they also may be combined with cluster approaches; however, the method again requires a crystal structure and is not commercially available (118).

A full quantum mechanical description of all intermolecular crystal effects can be achieved by periodic boundary conditions and plane wave basis sets (115). Although this approach allows the inclusion of all intermolecular effects that can be modeled by DFT, the method is computationally intensive and requires considerable computational resources, except in cases of small molecules and unit cells. When a crystal structure is available, CST and EFG calculations are possible with these methods, the most notable of which is the gauge-including projector-augmented wave (GIPAW) method, as implemented in the commercially available CASTEP software package (119). The GIPAW method has been applied to pharmaceutical molecules such as flurbiprofen and testosterone to enhance spectral assignments (111,120). As computational resources continue to increase in performance and become more accessible, the application of these methods to pharmaceutical compounds is likely to increase.

## 3 APPLICATIONS

### 3.1 Active Pharmaceutical Ingredient

SSNMR studies of drug substance still represent the predominant use of the technique in the pharmaceutical industry. Although it can be used in the study of chemical structure in the solid state, in cases where a molecule of interest is insoluble or where a particular effect (such as keto-enol tautomerism) must be studied directly in the solid (123), the primary application of SSNMR is in the detailed study of the physical form of molecules for which the chemical structure is already known. SSNMR is thus widely used to support the development of crystallization processes to control the solid form of a drug substance and to understand the solid form itself, including applications to polymorphs, salts, solvates, cocrystals, and amorphous phases. The resolution, specificity, and information content of SSNMR offer many advantages compared to other techniques for API solid-state analysis, including infrared (IR) and Raman spectroscopy, XPRD, thermal methods, and microscopic methods. In this role,

SSNMR is often applied in conjunction with other methods to gain complementary information.

One of the key applications of SSNMR to drug substance is in the determination of phase purity. Polymorphs, salts, solvates, cocrystals, and amorphous forms can all be collectively referred to as phases, where the term phase refers to a single crystalline or noncrystalline entity. A phase impurity is any other physically separated phase that is present besides the desired phase. For example, another undesired polymorph present at the 10% w/w level in a batch of a desired polymorph can be called a phase impurity (particularly if its specific identity as a polymorph is not yet known). In a salt, the presence of a free base form can be called a phase impurity. SSNMR is one of the most straightforward methods for identifying phase impurities when reference standards for the impurities are not available, and it works well in tandem with the other dominant method of assessing phase purity, XRPD indexing (12). The ease of assignment of SSNMR spectra is the reason for this utility. When determining polymorphic purity by SSNMR, the chemical purity of the sample should first be quickly assessed using other analytical techniques such as HPLC or solution-state NMR. Routine $^{13}$C SSNMR spectra can then be obtained, and, if present in the molecule, $^{19}$F and $^{31}$P spectra may also be recorded. A long relaxation delay should be used to ensure that polymorphs of unknown relaxation rates are observed (e.g., 10–20 s). In the resulting spectra, all resonance lines should be assigned to the solid-state structure, bearing in mind that the asymmetric unit may have more than one molecule ($Z' > 1$), that magnetic equivalencies observed in solution-state NMR are generally not observed in the solid state, and that potentially complicated peak splitting may appear for nuclei coupled to quadrupolar nuclei (such as $^{14}$N and $^{35/37}$Cl) (34) and to nondecoupled spin 1/2 nuclei (such as $^{19}$F) (121,122). If all peaks can be assigned, then it is likely that the sample is phase pure. If additional resonances are observed, particularly of low intensity, the resonances need to be assigned to another form or to defects in the crystal lattice. The simplest method to distinguish these two possibilities is to compare SSNMR spectra of multiple batches of API, since the level of a polymorphic impurity is likely to change from batch to batch, whereas the proportion of a lattice defect often remains constant. If several batches are not available, additional SSNMR experiments may be applied to distinguish these situations. A common approach involves measuring $^{1}$H $T_1$ relaxation rates via the $^{13}$C-detected saturation- or inversion-recovery experiment to take advantage of the resolution offered by $^{13}$C SSNMR (93). Since protons in same crystal lattice will have the same relaxation rate, disordered sites can be distinguished from a phase-separated polymorphic impurity. The experiment, however, does not differentiate between two polymorphs with the same relaxation rate. When unassigned resonances are observed in $^{19}$F NMR spectra, either $^{19}$F-detected $^{1}$H $T_1$ relaxation experiments or spin diffusion methods can be exploited. Spin diffusion is not typically applied to $^{13}$C or $^{15}$N due to the low natural abundance of the nuclei, and the resolution in $^{1}$H SSNMR is often not sufficient. When peaks of unknown origin are present in $^{1}$H, $^{19}$F, or $^{31}$P data, a phase impurity can be distinguished from disorder by use of the 2D dipolar-assisted rotational resonance (DARR) experiment (146) or other 2D spin diffusion experiments that utilized RF-driven or proton-driven approaches; if

correlations between the API and a low-intensity resonance of interest are detected, the molecules that cause these shifts exist in the same crystal lattice and structural disorder is the likely cause of the extra signals.

***3.1.1 Polymorphism*** Polymorphism is one of the principal applications of SSNMR in the pharmaceutical industry. Polymorphs are defined as different crystalline structures of the same molecule; they can differ in molecular conformation, packing arrangements, and close contact interactions such as hydrogen bonding. Strictly speaking, polymorphs do not contain any other species, including water, in the crystal structure. In the pharmaceutical industry, knowledge of the polymorphic behavior of a molecule and control of the desired polymorph of an API is needed for regulatory and intellectual property reasons. Therefore, analytical methods that can distinguish and quantify the polymorphic forms are widely used. The chemical shifts and relaxation times measured in SSNMR experiments are sensitive to differences in the local environment of nuclei, and as a result, the SSNMR spectra of polymorphs are usually unique. Numerous reviews of the field of polymorphism in small-molecule pharmaceuticals and the application of SSNMR to its study have appeared, and only a brief summary is given in this chapter to highlight the current trends in the use of SSNMR in polymorphism studies (5,6,124–126). SSNMR has been applied to the analysis of polymorphism in many classes of pharmaceutical compounds, covering virtually every class of compound commonly encountered in modern drug development (5,6,125,126,132,136,137). In all cases, SSNMR has offered high-resolution spectra with analytical specificity similar to that of high-quality XRPD analysis, which is widely recognized as the premier technique in distinguishing polymorphs.

Figure 3 shows a typical example of the $^{13}C$ SSNMR spectra of two polymorphs of a pharmaceutical drug candidate molecule obtained using the CP–TOSS experiment. The spectra highlight the general nature of SSNMR analyses of polymorphs; they are clearly distinguishable from each other and would allow quantification of one form to the other if appropriate criteria are met. Although SSNMR spectra can be used as fingerprints, the chemical shifts are influenced in a known fashion by the solid-state conformation of the molecule and interactions with the neighboring molecules; these effects can be used for more detailed structural examinations, as described later. In principle, any NMR active nucleus in the molecule can be used to distinguish polymorphs, although $^{13}C$ is by far the most widely employed given the organic nature of most pharmaceuticals, and the large number of distinct carbon sites, which enhances specificity (as many sites are available that might show a different chemical shift between polymorphs). Although $^{13}C$ SSNMR has adequate sensitivity for practical data acquisition time (i.e., $<8$ h in most cases), $^{19}F$ SSNMR is considerably more sensitive and can be utilized whenever fluorine is present in the molecule (37,127). Despite this advantage, literature examples showing the use of $^{19}F$ SSNMR in the detection and quantitative analysis of crystalline solids are limited, as probes capable of simultaneously exciting and detecting $^{19}F$, $^{1}H$, and an X-nucleus have only become commercially available in the past 10 years (55). Since the advent

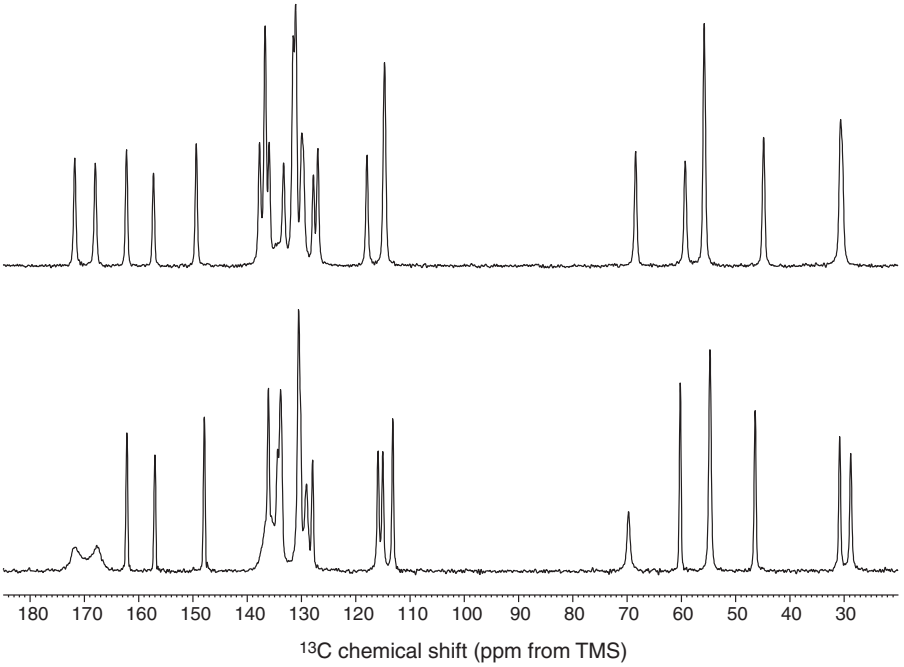


***FIGURE 3*** $^{13}$C CP–TOSS spectra of two polymorphs of an organic molecule illustrating the typical resolution, sensitivity, and information content of the technique in polymorphism studies. Each spectrum required approximately 8 h of acquisition time. Spectra were obtained at 8.4 T and 273 K.

of this hardware, a number of examples illustrating quantitative applications of $^{19}$F SSNMR to polymorphism or discrimination of polymorphs in API have been presented (128,129). An example is shown in Figure 4, where the $^{19}$F SSNMR spectra of four polymorphs of a trifluorinated steroid are depicted. Even though the molecule contains only three chemically distinct fluorine atoms, the polymorphs can be readily distinguished, and the presence of amorphous content is also easily discerned.

Beyond the basic ability to distinguish polymorphs, SSNMR offers a wealth of information about the crystal structure of each form that can be used in the absence of single-crystal X-ray (SCXRD) solutions to study structural relationships between polymorphs. This is particularly useful early in the development before a form has been selected and permits the study of all of the polymorphs including metastable or transient forms that are difficult to crystallize. A simple example demonstrating the sensitivity of SSNMR to crystal structure parameters is shown in Figure 5. The resonances at about 190 ppm are assigned to a single carbonyl carbon in the molecule, yet two NMR chemical shifts are observed in the spectrum of polymorph II (bottom). These resonances arise from structurally unique molecules in the crystal lattice and are referred to as the asymmetric unit ($Z'$) by crystallographers, or the smallest structural unit that, when symmetry is applied, forms the crystal lattice. The two polymorphs shown in Figure 5 have $Z' = 1$ (top) and $Z' = 2$ (bottom). The $Z'$ value is of

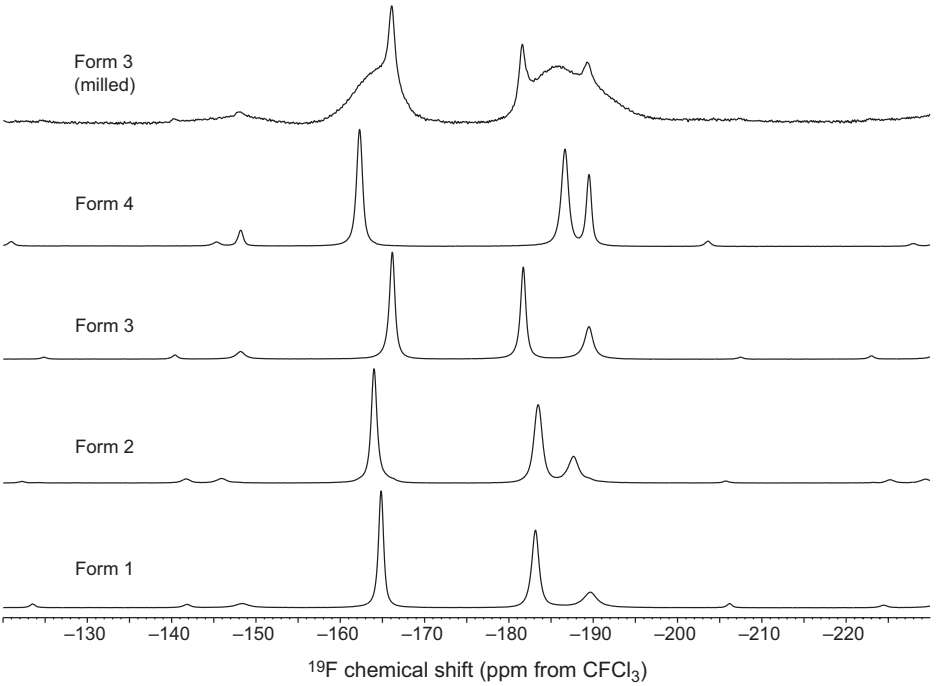


***FIGURE 4*** $^{19}$F DP–MAS spectra of four polymorphs and an amorphous form of a trifluorinated steroid. The topmost spectrum shows the spectrum of Form 3 after grinding and contains a large amount of amorphous material. Spectra were obtained at 8.4 T and 273 K.

great utility in indexing an XRPD pattern to obtain crystallographic information (134,135). Typical organic molecules have a $Z'$ of 0.5–4; however, larger values are observed. For instance, the presence of a $Z' = 6$, superlattice was identified by $^{1}$H relaxation measurements, whereas the solution to the X-ray structure was ambiguous (138,139).

The NMR chemical shift is sensitive to small changes in conformation and hydrogen bonding. The latter effect has been particularly well characterized for $^{1}$H, $^{2}$H, $^{13}$C, and $^{15}$N environments and can be interpreted empirically in many cases (140). For example, for carbonyl hydrogen bond acceptors, shorter hydrogen bonds lead to more deshielded $^{13}$C carbonyl resonances (141,142). For the $^{15}$N chemical shift of nitrogen hydrogen bond acceptors, such as pyridine and aliphatic nitrogen groups, shorter hydrogen bonds lead to more shielded resonances (143–145). Conversely, the $^{15}$N chemical shift of nitrogen hydrogen bond donors (e.g., amines and amides) is observed to be more deshielded with shorter hydrogen bonds. These effects have been used in numerous studies to probe hydrogen bonding in pharmaceutical polymorphs, solvates, and hydrates, especially where only limited crystal structure data are available (5,131–133).

The polymorph isolated from a given process may not be the most thermodynamically stable phase, but may be favored kinetically under the conditions of that

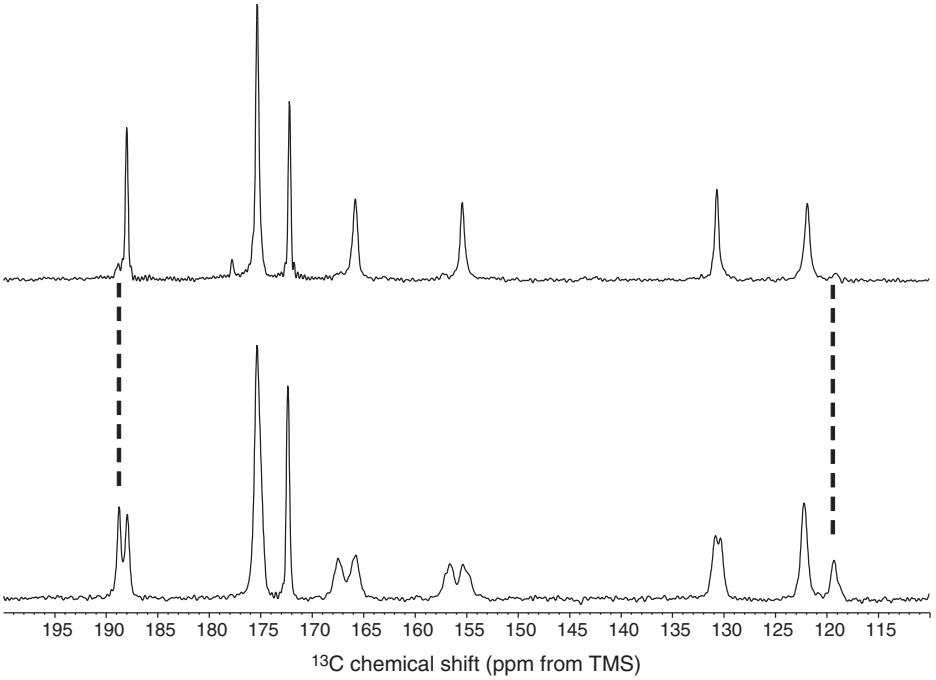


***FIGURE 5*** $^{13}$C CP–TOSS spectra of steroid drug before (bottom) and after (top) particle size reduction by micronization (top), illustrating the conversion from a metastable phase to a phase of greater stability. The dotted line highlights a small amount of the residual metastable phase remaining in the final material. Spectra were obtained at 8.4 T and 273 K.

process. Conversion between polymorphs can be induced by application of energy and can be achieved mechanically by particle size reduction or grinding, by heating, or by slurrying. The effect of micronization, a particle size reduction process (130), on a steroid molecule is shown in Figure 5. In this example, the polymorph isolated from the process is metastable, and during the particle size reduction, is almost completely converted to a more stable one.

***3.1.2 Disordered Solid Systems*** The term "disorder" has several meanings in modern pharmaceutical development. One common usage of this term arises from the processing of crystalline pharmaceutical solids, which can frequently lead to crystal defects and amorphous solids (147). Disorder can also be used to refer isolated molecular sites or moieties being in several positions in otherwise well-crystallized lattices. Since SSNMR is highly sensitive to local structural perturbations, it is able to detect all these types of disorder. Disorder in the solid state is of great concern to the pharmaceutical industry because it can lead to chemical stability issues (152,153).

A solid devoid of translational, rotational, or conformational order, and thus has no long-range order, is considered to be an amorphous solid. While some compounds

naturally crystallize in this state, the amorphous state can be induced by poor crystallization conditions or postprocessing of the bulk, such as drying, milling, compression, and micronization. It is common, however, to induce partial conversion, in which regions of purely amorphous and purely crystalline forms coexist in the same batch. In these cases, a convoluted SSNMR spectrum of sharp resonances of the crystalline form and the broad resonances of the amorphous form is observed (5). It should be noted, however, that unlike XRPD where a broad "halo" is observed for the amorphous phase, highly distinct (albeit broadened) resonances are still observed in SSNMR spectra. Further details of SSNMR analyses and the amorphous state in general can be found in numerous studies devoted to the amorphous state (5,148,149) and are also discussed in the context of amorphous dispersions with excipients later.

Disorder can be limited to a functional group such as a phenyl (150) or *tert*-butyl (151) group, or to any included solvate molecules. In these cases, only a few broadened low-intensity resonances of the particular structural moiety are observed in the SSNMR spectra, while the remaining peaks remain sharp. Site disorder can be caused by molecular mobility, and decreasing the analysis temperature can slow down this motion, reducing the broadening of the NMR resonance lines. When the disorder is static and the groups exist in distinct locations, temperature changes will not influence the SSNMR spectrum. In SCXRD, site disorder can be modeled and the atoms are given multiple positions with a corresponding site occupancy factor. However, the nature of the disorder, static or dynamic, cannot be measured directly from diffraction data. In addition, site disorder is not easily discernable by XRPD.

A special example of disorder is a solid solution or a solid whose crystal structure incorporated a structurally similar molecule into the lattice (154). SSNMR is particularly useful in studies of solid solutions, often in a complementary fashion with XRPD and thermal methods (155). Solid solutions involving an API molecule and a structurally related chemical impurity are of interest in pharmaceutical crystallization development, as are disordered crystalline API structures containing inclusions (or occlusions) of solvents or other small molecules; SSNMR analysis can be beneficial in these cases because of its ability to detect disordered regions of an otherwise crystalline material.

***3.1.3 Solvates*** The term solvate is used here to refer to a crystalline solid that has incorporated solvent molecules, including water, into the crystalline lattice. Similar to nonsolvated forms, solvates can be polymorphic (5,6). Solvates and hydrates have been classified in several different ways (156–161); in many cases, it is most useful to consider a continuum of solvation states based on the strength of binding of the solvent to the crystalline lattice. At one extreme, the solvent molecule is strongly hydrogen bound to a defined position in the lattice and typically has a stoichiometric relationship to the parent drug molecule. An illustrative example of the analysis of stoichiometric hydrates using $^{13}C$ and $^{15}N$ SSNMR has been given for three dihydrate polymorphs of olanzapine (162). Channel solvates, in which the solvent is located in channels that run through the lattice, may also have a highly ordered arrangement, but can also be disordered and exhibit variable stoichiometry. The presence of channels provides a mechanism for solvent to be lost and gained

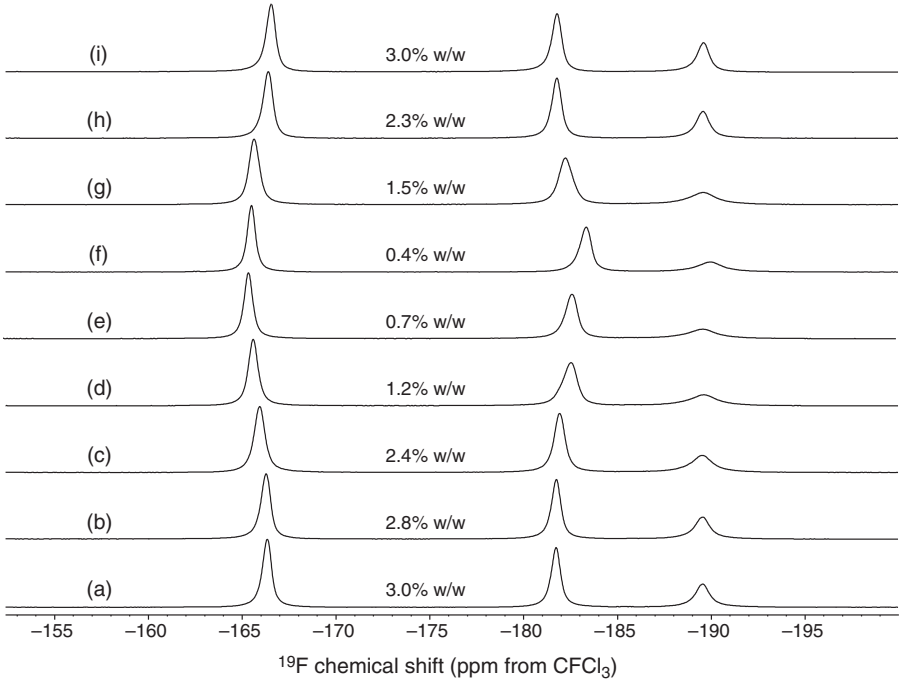


***FIGURE 6*** $^{19}$F DP–MAS spectra of a channel hydrate of a trifluorinated drug with increasing drying time (a–f) and then subsequently exposed to ambient humidity (g–i). The water content for each sample was measured separately and is shown for each point. Spectra were obtained at 8.4 T and 273 K.

according to ambient conditions as the lattice expands and contracts to accommodate more or less solvent. As an example of this type of solvate, Figure 6 shows the $^{19}$F SSNMR spectra of a channel hydrate of a steroid API as a function of its water content. As water is lost and gained through the process of drying and rehydration, the hydrogen bonding in the host lattice changes, as does the lattice spacing, and a continuous change in the appearance of the $^{19}$F resonances is observed. The effect is reversible and as no other resonances appear during the experiment, the lattice is expanding and contracting without collapsing irreversibly to a desolvated phase. A number of SSNMR studies of pharmaceuticals exhibiting nonstoichiometric hydration have been published, illustrating the different approaches to tackling this problem and exemplifying how SSNMR can provide insight into the nature of the solvation processes (24,25,133,163,164). In many cases, SSNMR has the unique ability to directly study the solvent, for example, by observation of $^{1}$H, $^{2}$H, and $^{17}$O nuclei (24,25).

At the other extreme, solvents can be present in larger channels or merely physisorbed to the surface of excipient/drug particles and are very weakly bound. Weak solvates may only exist in the presence of excess solvent (e.g., slurries and wet cakes) and can completely desolvate upon isolation or drying. An example of this is

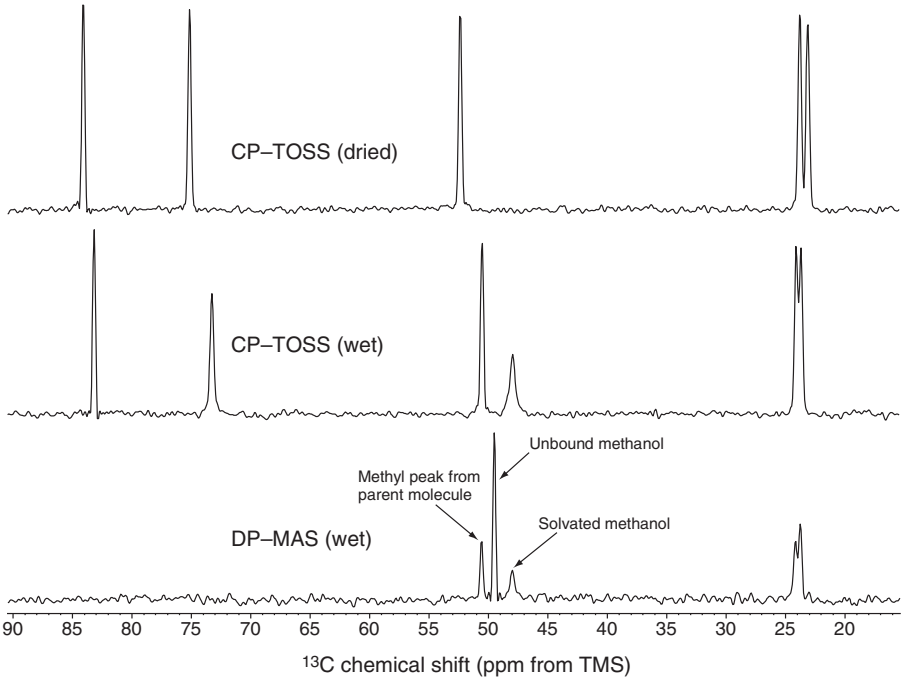


***FIGURE 7*** $^{13}$C DP–MAS spectrum of a weak methanol solvate (bottom), showing peaks from all methyl groups in the sample. $^{13}$C CP–TOSS spectra of same sample (middle) showing only peaks that are part of the solid lattice-including solvated methanol. $^{13}$C CP–TOSS spectra of this sample dried (top) showing loss of solvated methanol. Spectra were obtained at 9.4 T and 273 K.

shown in Figure 7, where a methanol solvate present in a wet cake sample loses solvent once dried to convert to a desolvated form. $^{13}$C SSNMR can distinguish between solvent bound to the lattice and unbound solvent using DP and CP–MAS experiments. In the DP–MAS spectrum, resonances from all species are observed and the intensity is based on the relaxation rate of the nucleus. Fast relaxing groups such as methyl groups or unbound liquids/solutions with unrestricted mobility are typically observed when using short relaxation delays (30 s or less). Conversely, the signal intensity in a $^{13}$C CP spectrum arises from the transfer of magnetization from $^{1}$H to $^{13}$C, which is disrupted by molecular motion, so that highly mobile solvent resonances are not observed. SSNMR is particularly well suited to studies of weak solvates because it can easily handle wet samples. The detection of weak solvates helps in the interpretation of competitive ripening results often used to establish thermodynamic stability between polymorphs (as the presence of weak solvates can confound the results in multisolvent studies).

***3.1.4 Salts*** It is often desirable to develop a salt of the active molecule either to obtain a crystalline product or to improve the characteristics of the drug such as

solubility or hygroscopicity (30). As such, it is important to establish salt formation or that protonation or deprotonation has occurred. When the salt former contains NMR active nuclei, it can be observed directly by SSNMR, while the occurrence of proton transfer can be established indirectly by the effect protonation or deprotonation has on the NMR chemical shifts of the parent molecule or directly by spectral editing experiments based on attached proton multiplicity. For example, eltrombopag olamine forms a salt with one or two moles of monoethanolamine (MEA), referred to as the mono- and bis-MEA salts, as shown in Figure 8. The compound has two acidic sites, with $pK_a$ values estimated at 4 for the carboxylic acid (C13) and 8.5 for the phenol (C8). The $^{13}C$ CP–TOSS spectra of crystalline MEA salts and free acid are compared in Figure 8. The MEA resonances are observed in the 60–40 ppm range, and the salt formation changes the carboxylic acid chemical shift from 167 to 173 ppm between the free acid and mono-MEA and the phenol (C8) from 142 to 156 ppm upon conversion to the bis-MEA salt.

Since a large number of pharmaceutical compounds contain basic nitrogen functional groups, the formation of an acidic salt can be confirmed by observing the $^{15}N$ chemical shift in the solid state. Unlike other spectroscopic methods, such as vibrational spectroscopy, SSNMR is able to readily locate the site(s) of protonation in a molecule containing various basic groups with suitable $pK_a$ values (165,166). Comparison of the chemical shifts of the nitrogen atoms of free base and putative salt can be informative, but the relative shifts can depend on the chemical environment of the nitrogen. Figure 9a compares the $^{15}N$ CP–MAS NMR spectrum of a free base and a tosylate salt of a drug candidate. In this case, salt formation is suggested by the 16 ppm shift of an amine resonance from −305 to −289 ppm. The minor shifts observed for the other nitrogen sites are related to differences between the crystal structures of the free base and the tosylate; computational methods would be required to resolve this ambiguity. Also, in some instances, it may not be possible to obtain a SSNMR spectrum of the parent molecule for comparison. In those cases, it is possible to establish the protonation state of the nitrogen atom by utilizing the spectral editing pulse sequences described in Section 2, such as dipolar dephasing. For example, the $^{15}N$ dipolar dephasing spectrum in Figure 9b shows a decrease in the intensity of the signal of those nitrogen atoms that are protonated.

When the pharmaceutical salt is a carboxylate, the $^{13}C$ CST can be measured to determine whether protonation has occurred (166–170). Deprontated carboxyl groups will generally yield a $\delta_{11}$ value of less than 250 ppm, while $\delta_{11}$ for protonated groups will generally lie above 250 ppm. These effects must be carefully interpreted, as hydrogen bonding involving the carboxylate group will also influence the CST; theoretical calculations may be necessary to fully resolve carbonyl ionization state questions. In addition to the more typical cases of carboxylic acid and nitrogen base salts, SSNMR is applicable to more complex studies of ionization and salt formation, as shown by its ability to resolve zwitterionic forms of piroxicam even in an amorphous phase (171). Other complex examples of mixed salt and cocrystal phases involving a composite crystal have also been studied (170,172). Recently, SSNMR of the quadrupolar nucleus $^{35}Cl$ (45) has been used to analyze and characterize polymorphs of hydrochloride salts. Although it requires very high magnetic fields,

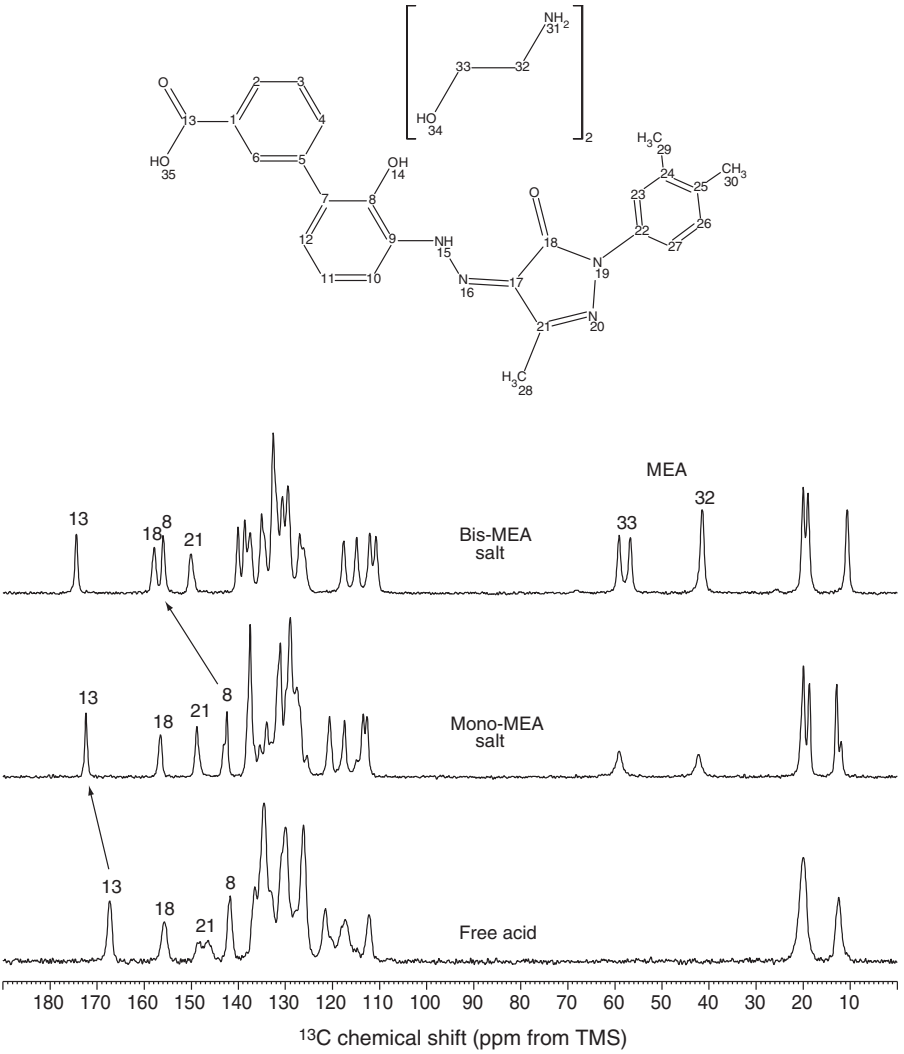


***FIGURE 8*** *Comparison of the* $^{13}$C CP–TOSS spectra of eltrombopag olamine (bis-monoethanolamine) salt to crystalline forms of the mono-MEA salt and the free acid, showing the chemical shift changed on salt formation. The MEA resonances are observed in the 60–40 ppm range. Spectra were obtained at 9.4 T and 273 K.

the technique is able to identify the number of chlorine sites in the lattice and provides an indication of the hydrogen bonding environment of this nucleus.

***3.1.5 Cocrystals*** Cocrystals represent an alternative to salt formation for the development of APIs and are a key element of the growing field of crystal engineering (173). A generally accepted definition of a pharmaceutical cocrystal

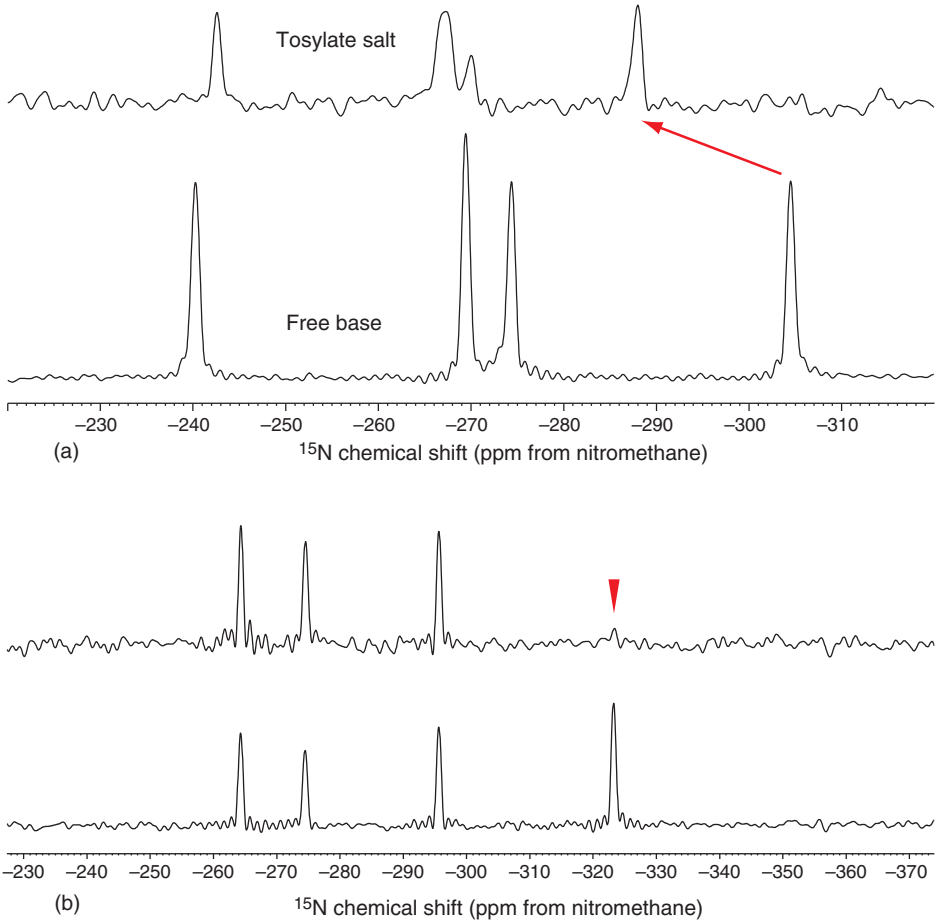


***FIGURE 9*** Demonstration of the analysis of nitrogen ionization states using $^{15}$N SSNMR. (a) $^{15}$N CP–MAS spectra demonstrating typical chemical shift change (red arrow) of a secondary amine ionized by protonation (top) compared to the free base (bottom). (b) $^{15}$N CP–MAS spectrum of a maleate salt (bottom) compared with the dipolar-dephased $^{15}$N CP–MAS spectrum (top) acquired on the same sample. A nitrogen atom is dephased (red arrow) by the strong dipolar interaction with its attached proton, showing the basic site where protonation has occurred. This methodology has the advantage that the information on protonation is available from a single sample, when comparison between a salt and a free base reference sample is not possible. Spectra were obtained at 8.4 T and 273 K.

is a multicomponent crystal made of an API and one or more small organic molecules, such that the components remain in their neutral (non-ionized) forms and interact with each other by hydrogen or other noncovalent interactions; a stricter definition also in use requires the pure components to be solids at room temperature (174). SSNMR is particularly useful for studying the association of the molecules within a cocrystal via many of the approaches already mentioned, particularly through dipolar correlation,

spin diffusion, and hydrogen bonding and ionization effects. Detailed illustrations of the applications of SSNMR methodology to the identification and characterization of pharmaceutical cocrystals have been presented (77,170). Application of $^{13}C$ and $^{15}N$ SSNMR to an unusual case of a nonstoichiometric pharmaceutical L-tartaric acid cocrystal was useful in identifying protonation states and intermolecular interactions (175).

## 3.2 Excipients

A wide variety of excipients are employed as binders, fillers, diluents, disintegrants, lubricants, or preservatives in modern pharmaceutical products, and SSNMR is a useful and popular method for their characterization (176). It is often desirable to closely study the physical and chemical nature of these inactive ingredients in bulk and in formulation. Many small-molecule excipients can exhibit polymorphism and solvation. For example, D-mannitol is often used as a stabilizer for lyophilized pharmaceutical protein-based products. The three known polymorphs of D-mannitol (177) can undergo form conversion depending on the lyophilization conditions (178). The $^{13}C$ CP–TOSS spectra of the polymorphs are shown in Figure 10 and can be used to identify the polymorph present. Excipients can also exist as hydrates; for example, magnesium stearate has four hydration states, a mono-, di-, and trihydrate, and an anhydrous form, although it is typically sold as a mixture of these states. Understanding the phase content of magnesium stearate can be critical in drug development, as studies have shown that the hydration state of magnesium stearate can affect its performance as a lubricant (179). Trehalose is another excipient that is amenable to characterization by $^{13}C$ SSNMR, particularly since it forms a dihydrate and an amorphous form; the latter is not easily detected by diffraction methods. $^{13}C$ SSNMR provides a high-performance method for quantifying amorphous content in trehalose dehydrate (180). A wide variety of semisolids and gels, including lipids, are used as excipients and can display complex polymorphism and mesomorphism (181). For example, $^{13}C$ SSNMR studies of lipids using CP, dipolar dephasing, and $^{13}C\,T_1$ relaxation measurements provided insight into mesomorphism, ordering, and mobility in α- and γ-oleic acid (182). Information of this nature is useful in pharmaceutical development because the complex polymorphism and crystallization behavior of fats and lipids can impact key product performance attributes in drug products, such as softening temperatures in capsules.

Numerous polymeric excipients are used in pharmaceutical products. SSNMR is a widely used technique in polymer science (95), and pharmaceutical polymers can benefit from much of this experience. An example is provided by poly(methyl methacrylate)-based polymers and copolymers, commonly used in tablet film coating preparations, which have been extensively studied by SSNMR methods including relaxation and spin diffusion studies (183–185). Detailed SSNMR studies of a wide range of other polymeric pharmaceutical materials, such as celluloses, amyloses, and cyclodextrins, have also been reported (186–188). In general, SSNMR can provide information on structure, dynamics, miscibility, and phase separation and help relate these to the macroscopic performance of the polymer.

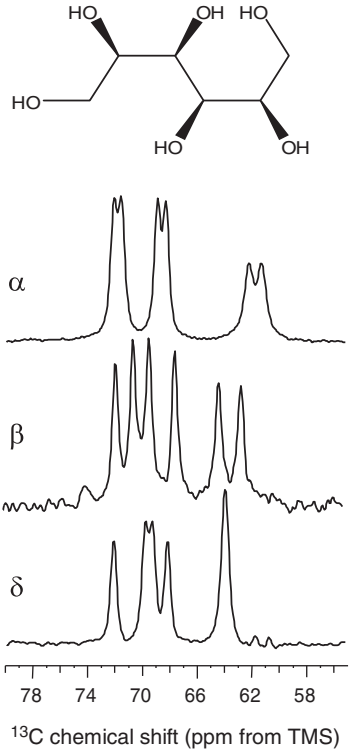


***FIGURE 10*** $^{13}$C CP–TOSS spectra of three polymorphic forms of D-mannitol, illustrating the ability of SSNMR to distinguish the forms of this excipient. Spectra were obtained at 9.4 T and 273 K.

It is well known that many excipients will adsorb water and that the absorbed water dramatically influences the physical properties or stability of excipients or drug products. For example, sodium starch glycolate, a disintegrant, can adsorb ~25% w/w of moisture when exposed to high humidity; this can be directly and selectively monitored by chemical shift changes in the $^{23}$Na SSNMR spectrum (12). SSNMR is relatively unique in its ability to selectively detect the impact of moisture on excipients, whether through its effects on $^{13}$C, $^{15}$N, or other spectra or through direct observation of protic water by $^{1}$H SSNMR or isotopically labeled water by $^{2}$H and $^{17}$O SSNMR (26), and can thus enhance moisture tracking studies involving vibrational and thermal methods.

## 3.3 Drug Product

Analysis of solid drug products can be broken down into three broadly defined categories: determination of the characteristics of the API in formulation, investigations into the performance of the product *in vitro* and *in vivo*, and assessment of physical solid-state stability. The complexity of a modern drug product demands a

multidisciplinary approach, and the versatility of SSNMR makes it a valuable contributor, particularly given the large number of potential transformations that can occur during the formulation process (189).

***3.3.1 Analysis of API Solid Phases in Formulations*** $^{13}$C SSNMR has been used to identify and quantify the API in a variety of drug products (8,191–198), although $^{13}$C SSNMR is intrinsically a low-sensitivity technique; drug loads of less than 5% w/w are difficult to analyze in a reasonable amount of time except in favorable cases. It should be noted, however, that the sensitivity of SSNMR for low-dose formulations has been enhanced using $^{13}$C and $^{15}$N labeling (12,18,198). Figure 11 shows a typical $^{13}$C SSNMR spectrum of a small-molecule API in a tablet dosage form. Drug molecules often contain various aromatic, carboxylic, and heterocyclic functional groups that are absent in excipients, making it straightforward to both assign all the components of an SSNMR spectrum and find regions devoid of excipient resonances. Other solid-state analysis techniques often require detailed analysis or MVA to separate these components; furthermore. Typical XRPD instruments cannot readily detect amorphous phases. There are many areas of drug product analysis where SSNMR has been underutilized, particularly in studies of the role of

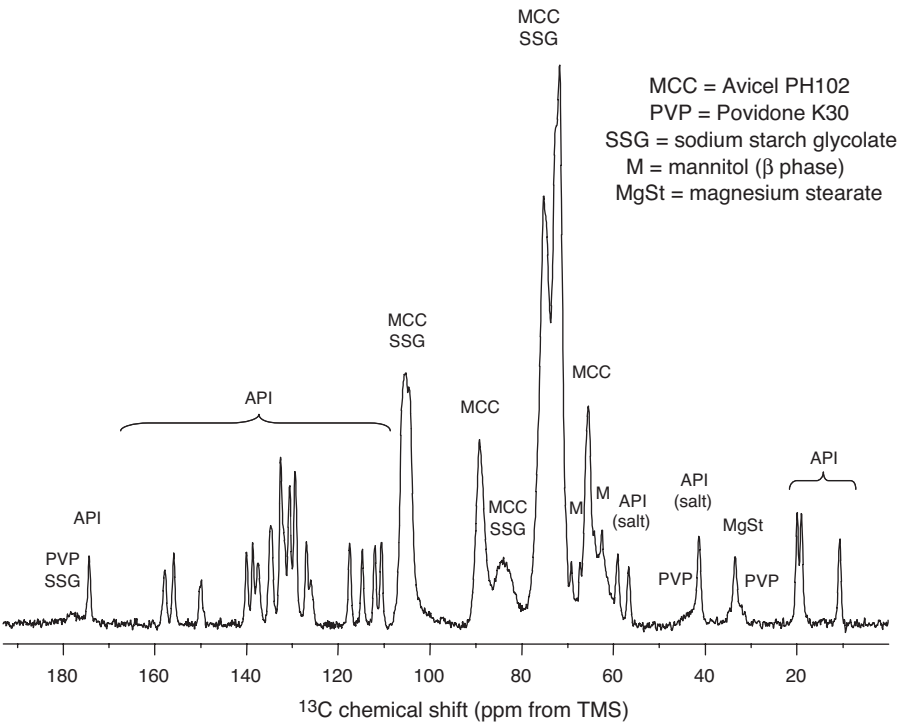


***FIGURE 11*** $^{13}$C CP–TOSS spectrum of eltrombopag olamine in an oral tablet dosage form. The excipients and API chemical shifts are identified. The spectrum was obtained at 9.4 T and 273 K.

moisture in API and excipients. The following examples illustrate the potential of SSNMR in these and other drug product applications.

Although $^{13}C$ SSNMR analysis of drug products is routinely used, more sensitive nuclei such as $^{1}H$, $^{19}F$, $^{23}Na$, or $^{31}P$ can be used as alternatives (199) to extend the limit of detection or obtain the necessary specificity. $^{1}H$ SSNMR analysis is not typically utilized in drug product analysis because of dipolar broadening discussed earlier, as well as spectral overlap with excipients. However, Griffin et al. (200) have demonstrated identification of an API in formulation by exploiting the resolution increase offered in certain two-dimensional NMR techniques, specifically $^{1}H$double-quantum (DQ) MAS experiments. The utility of $^{19}F$ SSNMR is particularly relevant to formulations, since excipients do not typically contain fluorine groups; however, there are only limited literature examples of this approach. In one case, $^{19}F$ SSNMR was used to identify and detect amorphous content in 2% w/w tablets of a selective $M_3$ neuronal receptor antagonist (128). In another example, $^{19}F$ SSNMR was used as a limit test in tablets containing 8.3% w/w of API; a limit of 1.5% w/w for undesired forms was achieved from high signal-to-noise spectra (129). Taking advantage of the previously discussed ability of SSNMR to analyze liquid and wet samples, $^{19}F$ CP–MAS was used to investigate the crystalline form of a drug inside a suspension formulation (201).

About 80% of pharmaceuticals are developed as tablets or capsules, but there are many other formulation classes in which SSNMR can be used including ointments, inhaled powders, and solid dispersions. Figure 12 depicts the $^{13}C$ SSNMR spectrum of a 1% retapamulin ointment; the semisolid excipient petrolatum is clearly observed along with the desired polymorph of the API used in this product. In the past few years, numerous improvements in inhaled dry powder formulations have been made. The complexity of the delivery system makes product formulation a crucial attribute and great efforts are made to fully understand the system. Recently, SSNMR was utilized to study the encapsulation of rifampicin with liposomes for an inhalation product. SSNMR provided insight into the order of the lipic acyl chains and location of the rifampicin in the matrix (202).

Formulations that involve solid dispersions of API in a polymer can also benefit from study by SSNMR. SSNMR can help study the degree of mixing of the API and the polymers present (or any other additives, such as surfactants). Because of its ability to detect the association of different molecules over a range of distances via dipolar correlation and spin diffusion methods, SSNMR is particularly useful for understanding the physical state of a solid dispersion, namely, whether it is a solid solution, nanosuspension, or microsuspension (203). Given the amorphous or mixed crystalline–amorphous nature of these materials, SSNMR has an inherent advantage over XRPD analysis. Several examples are available that illustrate the use of different SSNMR approaches in the study of solid dispersions, including 2D CP-HETCOR experiments (Table 2) for studying API molecular conformation in the amorphous dispersion, and static $^{19}F$ studies of mobility (204–207).

In addition to evaluating the final drug product, SSNMR can be used to understand and potentially direct formulation processes in a variety of systems, including controlled release formulations (208). For example, the crystal structure of a

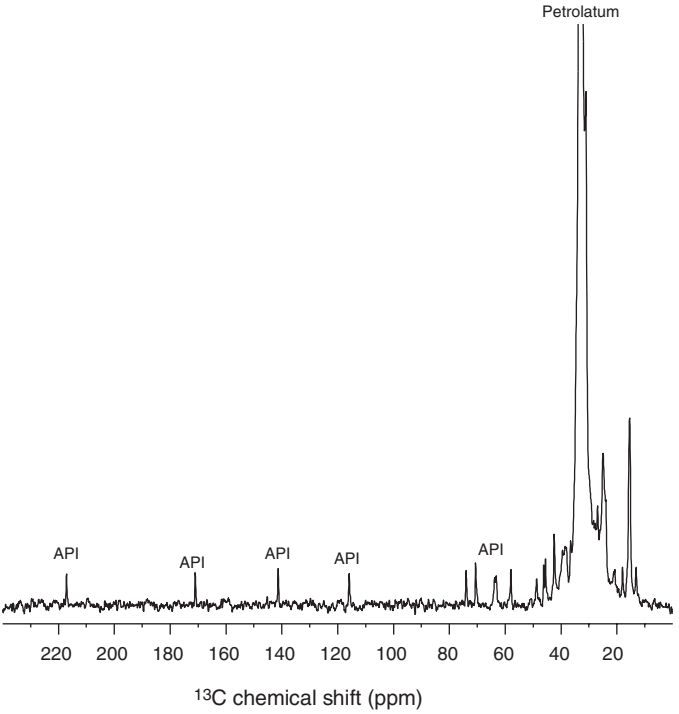


***FIGURE 12*** $^{13}$C CP–TOSS spectrum of a 1% w/w retapamulin petrolatum ointment, illustrating the ability of $^{13}$C SSNMR to detect the polymorph present in a lose-dose topical formulation. The spectrum was obtained at 9.4 T and 273 K.

derivatized pleuromutilin (*trans*-4-aminocyclohexyl (1*S*,2*R*,3*S*,4*S*,6*R*,7*R*,8*R*,14*R*)-4-ethenyl-3-hydroxy-2,4,7,14-tetramethyl -9-oxotricyclo[5.4.3.01,8]tetradec-6-yl imidodicarbonate) possess large-diameter channels that are capable of filling with nonstoichiometric amounts of succinic acid, ranging from 0.5 to 2.0 mol (170). There was concern that the succinic acid could be washed out of the channels during a top spray fluid bed granulation process. In conjunction with XRPD and chromatographic analysis, SSNMR was used to evaluate granules exposed to different in-process moisture levels, drying rates, and final moisture contents. While no changes in the SSNMR were observed with in-process moisture or drying rate, Figure 13 shows the significant change that final moisture content of the granules had on the succinic acid resonances at ~180 ppm. Chromatographic analysis showed that the total succinic acid content was not changed by the granulation process; therefore, the apparent loss of signal intensity in the CP–TOSS spectrum is caused by an interruption of the cross-polarization. Similar behavior was observed in the pure API and is caused by water entering the channels and triggering molecular motion of the cocrystallized succinic acid molecule. On the basis of these data, fluid bed granulation was used for the process, and a specification for the final moisture content was developed.

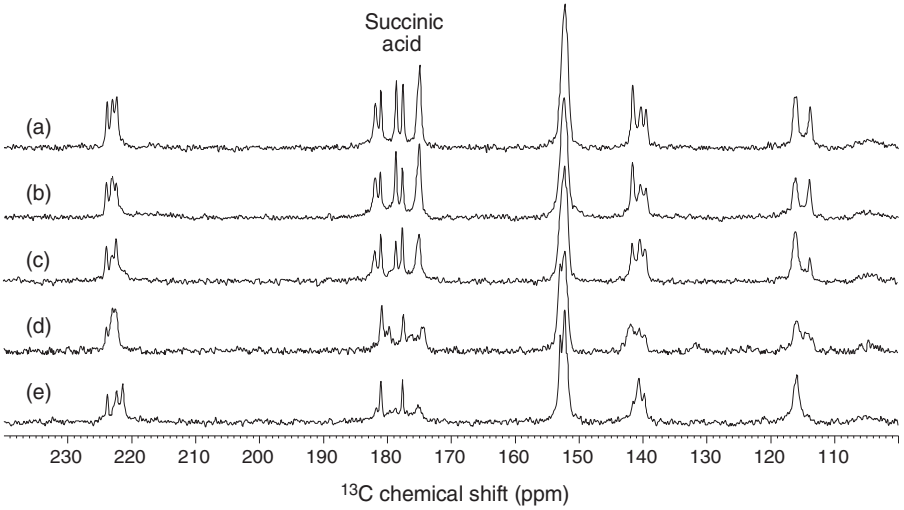


***FIGURE 13*** *Expanded region of the* $^{13}$C CP–TOSS spectra of derivatized pleuromutilin granules at various levels of final moisture content: (a) 0.79, (b) 1.09, (c) 1.66, (d) 2.24, and (e) 2.69 mol relative to the API. The succinic acid signals appear at approximately 180 ppm. Spectra were obtained at 9.4 T and 273 K.

***3.3.2 Investigation of Drug Product Performance Issues*** Failure of tablets to fully dissolve, or slow dissolution, is a common problem during drug product development, and in some cases, SSNMR can provide insight into the problem. For example, using $^{13}$C SSNMR, a dissolution failure of delavirdine mesylate tablets exposed to high humidity was linked to free base formation (209). The formation of crystalline drug fractions in capsules of SCH 48461, a cholesterol absorption inhibitor being developed as an amorphous form, was linked to a decrease in the dissolution rate (210). Sometimes, the dissolution medium can interact with the tablet, causing apparent dissolution problems. Partially disintegrated tablets can be removed from the dissolution media and evaluated without drying by SSNMR. Figure 14 shows the $^{13}$C SSNMR spectra of a tablet of an API sodium salt pulled from a 0.1 N HCl dissolution medium after 15 min compared to the tablet prior to disintegration. The lactose is dissolved and a resonance at 170 ppm, concordant with the parent molecule, is observed. Since the parent molecule is insoluble, a new dissolution medium was developed for this API. Sometimes, surfactants such as sodium lauryl sulfate (SLS) are added to the dissolution media of poorly water-soluble drugs. Figure 15 shows a partially dissolved API pulled from a dissolution medium containing 50 mM acetate buffer with 0.75% SDS. The resonances at 71 and 32 ppm are related to the SDS; however, two sets of API-like resonances are observed. For example, the peak at 97.9 ppm is consistent with the initial form of the API, but a second peak at 99.1 ppm is also observed. The chemical shifts of these additional peaks were not consistent with the known forms of the API and changes in the chemical shift of peaks assigned to SDS

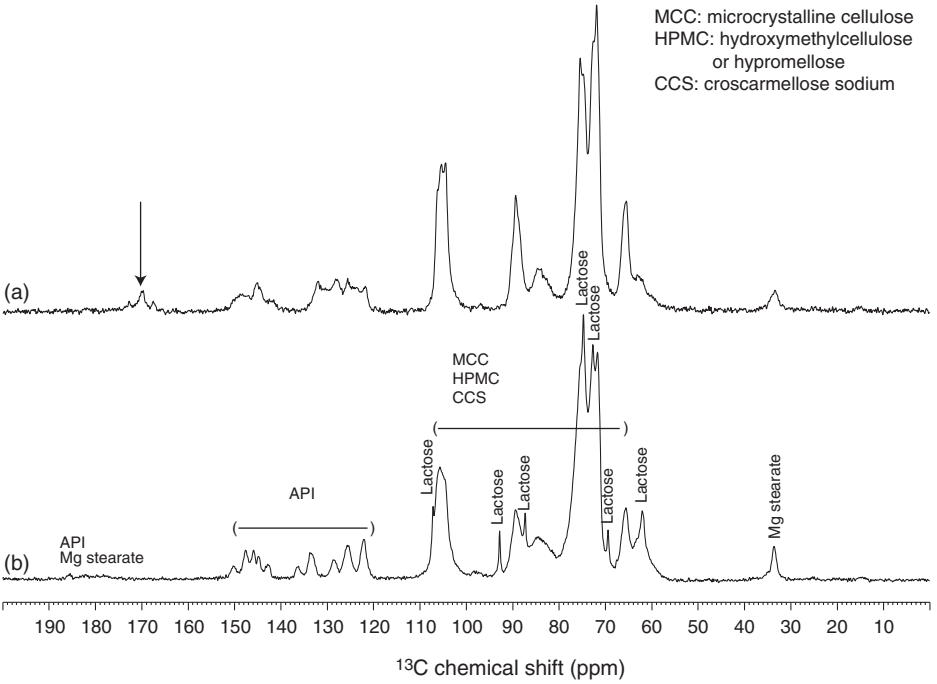


***FIGURE 14*** $^{13}$C CP–TOSS spectra of (a) a 50 mg tablet pulled from a 0.1 N HCl dissolution media after 15 min and (b) a tablet prior to testing. The arrow shows peaks associated with a phase change of the API (see text). Spectra were obtained at 9.4 T and 273 K.

suggested that a salt or cocrystal may have formed. This was later confirmed by preparation of an authentic SDS salt reference.

Although dissolution failures are not always related to API physical transformations, the ability of SSNMR to analyze complex mixtures (including wet samples) allows it to quickly confirm or deny a potential physical transformation as the cause of a product dissolution failure, representing an important and growing area of use for SSNMR.

***3.3.3 Physical Stability Analysis with SSNMR*** SSNMR analysis of drug products during stability studies can be relevant in several situations. The API may be known to undergo a physical change, and SSNMR may be the best method for observing API in drug product. Even if the API is known to be stable, it may react with excipients in the formulation, or the excipients may induce noninteraction degradation causing stability issues. Although SSNMR does not directly test chemical stability, it can detect many of the reasons behind instability observed during stability testing, particularly if moisture is involved with the underlying process.

Figure 16 shows an example of a desolvation phase transition involving a low-dose API in drug product monitored by $^{19}$F CP–MAS. Two sets of 1 mg tablets were stored

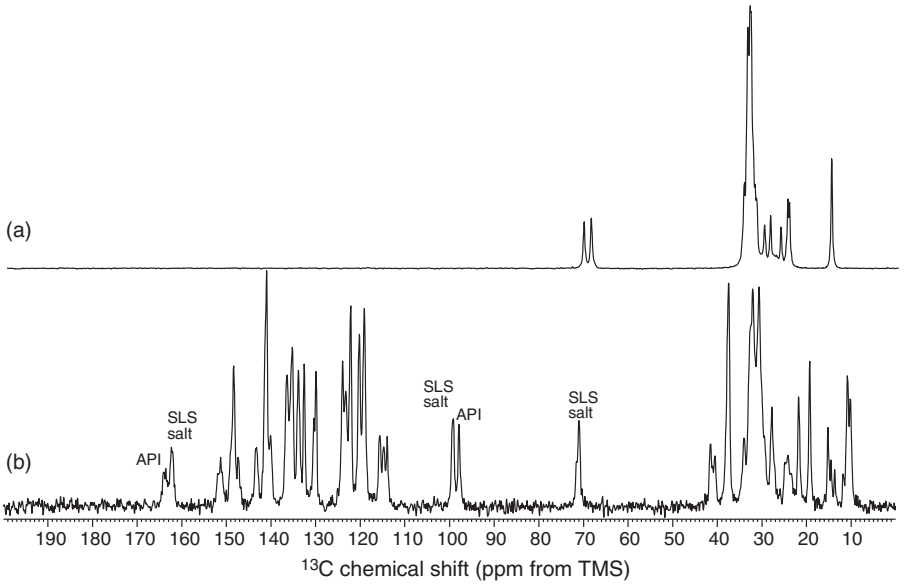


***FIGURE 15*** $^{13}$C CP–TOSS spectra of (a) sodium lauryl sulfate and (b) an API dissolved in a 50 mM acetate buffer with 0.75% SLS for 10 min. In addition to the resonances of the API (starting material), a second set of API-related peaks and SLS resonances are observed in (b); several isolated site resonances are labeled. These extra peaks were later confirmed as the SLS salt of the API. Spectra were obtained at 9.4 T and 273 K.

at 40°C, 75% RH for 6 months with both as-received and dried, low water content excipients. The reduction of the moisture content in the excipients drastically reduced the desolvation of the API. Further efforts to reduce the ingress of moisture into tablets by moisture blocking packaging and desiccation did not prevent the minor desolvation of the API upon stability. Since desolvation was related to the *in situ* release of moisture from the excipients, it became more significant at lower drug loads where the API to excipient ratio is greater. This demonstrates the utility of $^{19}$F SSNMR in the detection of low levels of unwanted forms in low-dose formulations without interference from excipients, and it is an example of excipient moisture uptake causing API physical instability (211).

In addition to stability studies, SSNMR can aid in gaining an understanding of stability phenomena via forced degradation studies and formulation compatibility studies. For example, excipients can engage in ionic interactions with an API. Byrn et al. monitored the acid–base interaction of magnesium oxide with (+) and (±) ibuprofen, showing that the carboxylic acid of (+)-ibuprofen deprotonates to form a magnesium salt at a much faster rate than the racemic mixture (211). A second example involves the use of tartaric acid as an acidifying agent in a formulation for solubility enhancement. During the formulation development, formulations that involved addition of tartaric acid during granulation (intragranular) and postgranulation (extragranular) to a fluorinated free base API were examined. The $^{13}$C CP–TOSS

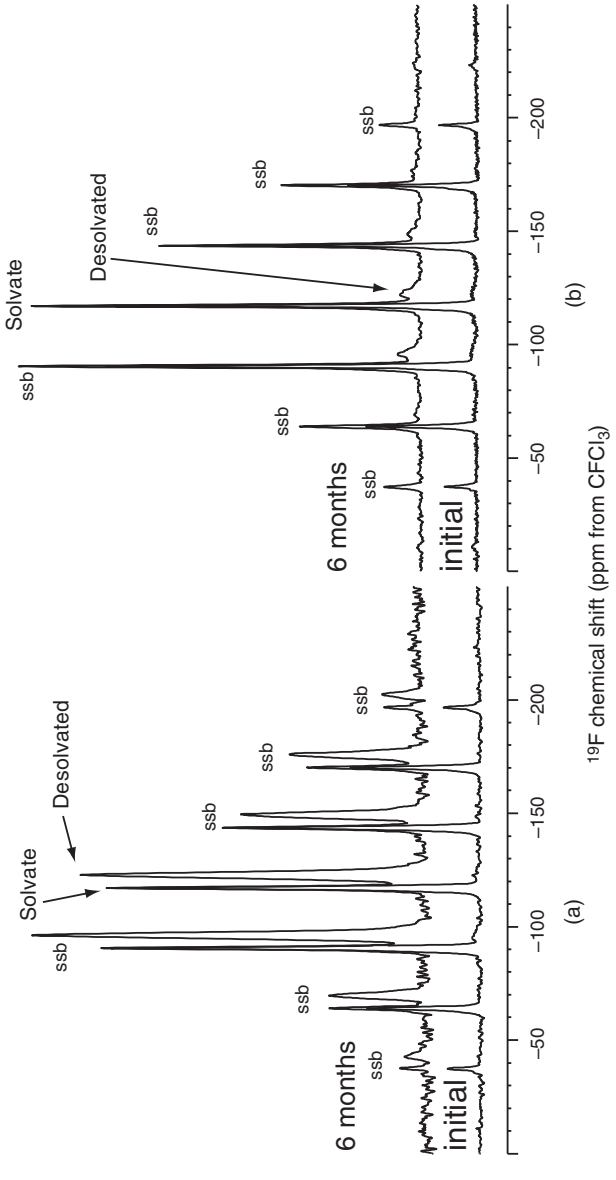


***FIGURE 16*** $^{19}$F CP–MAS spectra of 1 mg tablets (a) initial and 6 months at 40°C/75% RH stability sample with excipients as received, (b) initial and 6 months at 40°C/75% RH stability sample with low water content excipients. Spinning sidebands are indicated with "ssb." Spectra were obtained at 11.7 T and 273 K.

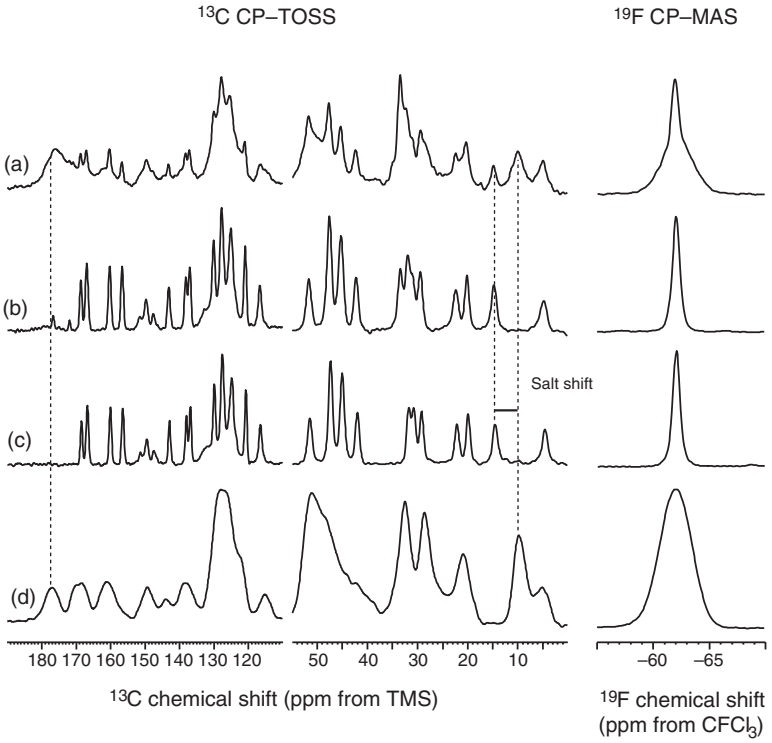


***FIGURE 17*** $^{13}$C CP–TOSS and $^{19}$F CP–MAS spectra of (a) tablets with 10% w/w intragranular tartaric acid, (b) tablets with 10% w/w extragranular tartaric acid, (c) API free base, and (d) a separately prepared amorphous API tartrate salt. Spectra were obtained at 11.7 T and 273 K.

and $^{19}$F CP–MAS spectra of the resulting tablets with 10% w/w intragranular and extragranular tartaric acid are shown in Figure 17. A significant shift is observed in the methyl signals of the protonated ethylamine group upon salt formation. The observed chemical shift of this methyl group in tablets reveals the formation of an amorphous tartrate salt in the intragranular formulation, whereas tablets with extragranular tartaric acid are still consistent with free base. The $^{19}$F signal provides further evidence for the salt formation through a broad peak for the amorphous tartrate salt. In tablets with intragranular tartaric acid, this broad peak is clearly evident.

## 4 FUTURE DIRECTIONS AND DEVELOPMENTS

The field of pharmaceutical SSNMR continues to advance at a rapid rate. Many of the areas of expected growth have already been demonstrated to some extent, as previously discussed, even if they are not yet fully applied across the industry. However, several new developments in the wider field of SSNMR, not yet applied to pharmaceutical analysis, deserve special mention because of the high likelihood that

they will eventually be utilized to an appreciable extent. These new developments, and several new applications, are described here.

A number of instrumental and technological improvements in the field of NMR continue to offer new possibilities. A general trend toward higher static field strengths offers increasing access to many quadrupolar nuclei of pharmaceutical interest that are currently not in widespread use. Improvements in probe design, such as the high RF fields generated by microcoil designs, can enhance wideline SSNMR of quadrupolar nuclei. Finally, improvements in automation including automatic tuning probes (similar to those available for solution-state NMR) and improved autosampler designs will likely enhance the usage of SSNMR by simply allowing more analyses to be performed for more drug development projects using less instrumentation. Recently, sensitivity enhancements have been obtained by embedding a microcoil inside a specially constructed MAS rotor and spinning the rotor while relying on RF coupling to an outer coil (212). This technique, known as magic-angle coil spinning (MACS), offers sensitivity enhancements for most common experiments. Because of its ease of implementation, it may soon appear as a routine method in many pharmaceutical SSNMR analyses. Cryogenic probe cooling technology, though it is complex, may also ultimately be adapted for routine SSNMR use; experimental designs have already been demonstrated wherein the coil is maintained at 12 K while thermally isolated from the spinning rotor (213). A variety of polarization enhancement techniques are also available, including those that utilize hyperpolarized xenon gas and parahydrogen (214); the main limitation of these techniques is that the enhancement is generally selective, for example, for surfaces, although this can also be considered an advantage for certain applications.

In the realm of applications, several areas warrant mention. SSNMR may be applied to pharmaceutical molecules of ever increasing size and complexity, ranging from large macrolides to oligonucleotides and peptides to proteins and larger molecules. Polymorphism in crystalline proteins has been studied by SSNMR (215); however; the majority of biopharmaceutical products are amorphous in nature. SSNMR has numerous applications and a great potential in studies of amorphous biopharmaceutical products, particularly those manufactured by lyophilization. Gross information about the structure of a large molecule is available from basic $^{13}C$ and $^{15}N$ SSNMR spectra, allowing at least the detection of batch-to-batch variations in tertiary protein structure. SSNMR is sensitive to dynamics through both spectral variation with temperature and various relaxation measurements and thus has potential applications to the study of glass transition temperatures impacting drug product stability (190).

## 5 CONCLUSIONS

Over the past decade, SSNMR has been increasingly utilized for pharmaceutical analysis. It is widely applied in studies of pharmaceutical solids, ranging from detailed structural studies of API forms to analysis of complex drug products and formulations. It is expected that SSNMR will be increasingly applied to studies of

complex drug products as both the sensitivity of the basic experiments and the performance of more advanced experiments continue to increase. The applications of SSNMR in pharmaceuticals closely parallel those in other areas of materials science, and cross-fertilization will likely continue. Applications to more complex APIs, such as peptides, oligonucleotides, and biopharmaceuticals, will also develop in the future as new molecular classes are developed and new formulations and delivery modes are attempted. As drug molecules, formulations, and products evolve, SSNMR techniques will likely evolve with them.

## REFERENCES


1. Geppi M, Mollica G, Borsacchi S, Veracini CA. Solid-state NMR studies of pharmaceutical systems. Appl Spectros Rev 2008; 43: 202–302.
2. Tishmack PA, Bugay DE, Byrn SR. Solid-state nuclear magnetic resonance spectroscopy—pharmaceutical applications. J Pharm Sci 2003; 92: 441–474.
3. Harris RK. NMR studies of organic polymorphs and solvates. Analyst 2006; 131: 351–373.
4. Reutzel-Edens SM, Bush JK. Solid-state NMR spectroscopy of small molecules: from NMR crystallography to the characterization of solid oral dosage forms. Am Pharm Rev 2002; 5: 112–115.
5. Byrn SR, Pfeiffer RR, Stowell JG. Solid-state chemistry of drugs. West Lafayette (IN): SSCI; 1999.
6. Bernstein J. Polymorphism in molecular crystals. New York: Oxford University Press; 2002.
7. Hilfiker R. Polymorphism in the pharmaceutical industry. Weinheim: Wiley-VCH; 2006.
8. Munson EJ, Lubach JW. Solid-state NMR in the characterization of pharmaceutical formulations. In: Swarbrick J, editor. Encyclopedia of pharmaceutical technology. 3rd ed. Informa Healthcare USA; 2007. p 3297–3309.
9. Raw AS, Furness MS, Gill DS, Adams RC, Holcombe FO, Yu LX. Regulatory considerations of pharmaceutical solid polymorphism in Abbreviated New Drug Applications (ANDAs). Adv Drug Deliv Rev 2004; 56: 397–414.
10. Lee SL, Raw AS, Yu L. Significance of drug substance physiochemical properties in regulatory quality by design. Preformulation in Solid Dosage Form Development, Drugs and the Pharmaceutical Sciences. Informa Healthcare; 2008. p 571–583.
11. ICH. International Conference on Harmonization Q6A Guideline: Specifications for New Drug Substances and Products. ICH, Geneva, Switzerland, 1999.
12. Vogt FG. A multi-disciplinary approach to the solid-state analysis of pharmaceuticals. Am Pharm Rev 2008; 11: 50–57.
13. Ernst RR, Bodenhausen G, Wokaun A. Principles of nuclear magnetic resonance in one and two dimensions. New York: Oxford University Press; 1987.
14. Slichter CP. Principles of magnetic resonance. 3rd ed. Heidelburg: Springer-Verlag; 1996.
15. Gerstein BC, Dybowski CR. Transient techniques in the NMR of solids. San Diego (CA): Academic Press; 1985.
16. Stejskal EO, Memory JD. High resolution NMR in the solid state. New York: Oxford University Press; 1994.

17. Vogt FG. The evolution of solid-state NMR in pharmaceutical analysis. Future Med Chem 2010; 2: 915–921.
18. Remenar JF, Wenslow RM, Ostovic D, Peresypkin A. Solid-state nuclear magnetic resonance determination of the physical form of BHA on common pharmaceutical excipients. Pharm Res 2004; 21: 185–188.
19. Kalinowski HO, Berger S, Braun S. Carbon-13 NMR spectroscopy. New York: Wiley; 1988.
20. Earl WL, Vanderhart DL. Measurement of $^{13}C$ chemical shifts in solids. J Magn Reson 1982; 48: 35–54.
21. Harris RK, Becker ED, Cabral de Menezes SM, Goodfellow R, Granger P. NMR nomenclature. Nuclear spin properties and conventions for chemical shifts (IUPAC recommendations 2001). Solid State Nucl Magn Reson 2002; 22: 458–483.
22. Duncan TM. A compilation of chemical shift anisotropies. Chicago: Farragut Press; 1990.
23. Brown SP. Probing proton–proton proximities in the solid state. Prog NMR Spectrosc 2007; 50: 199–251.
24. Vogt FG, Brum J, Katrincic LM, Flach A, Socha JM, Goodman RM, Haltiwanger RC. Physical, crystallographic, and spectroscopic characterization of a crystalline pharmaceutical hydrate: understanding the role of water. Cryst Growth Des 2006; 6: 2333–2354.
25. Vogt FG, Dell'Orco PC, Diederich AM, Su Q, Wood JL, Zuber GE, Katrincic LM, Mueller RL, Busby DJ, DeBrosse CW. A study of variable hydration states in topotecan hydrochloride. J Pharm Biomed Anal 2006; 40: 1080–1088.
26. Aso Y, Yoshioka S, Terao T. Effect of the binding of water to excipients as measured by $^{2}H$-NMR relaxation time on cephalothin decomposition rate. Chem Pharm Bull 1994; 42: 398–401.
27. Yarnell A. Heavy-hydrogen drugs turn heads, again. C&E News, June 22, 2009, p 36–39.
28. Delgado JN, Remers WA, editors. Wilson and Gisvold's textbook of organic medicinal and pharmaceutical chemistry. 9th ed. Philadelphia (PA): J.P. Lippincott; 1991.
29. Braese S, Gil C, Knepper K, Zimmermann V. Organic azides. An exploding diversity of a unique class of compounds. Angew Chem Int Ed Engl 2005; 44: 5188–5240.
30. Stahl PH, Wermuth CG, editors. Handbook of pharmaceutical salts: properties, selection, and use. New York: Wiley-VCH; 2002.
31. Berger S, Braun S, Kalinowski HO. NMR spectroscopy of the non-metallic elements. New York: Wiley; 1997.
32. O'Dell LA, Schurko RW. Fast and simple acquisition of solid-state $^{14}N$ NMR spectra with signal enhancement via population transfer. J Am Chem Soc 2009; 131: 6658–6659.
33. Balchin E, Malcolme-Lawes DJ, Poplett IJF, Rowe MD, Smith JAS, Pearce GES, Wren SAC. Potential of nuclear quadrupole resonance in pharmaceutical analysis. Anal Chem 2005; 77: 3925–3930.
34. Harris RK, Olivieri AC. Quadrupolar effects transferred to spin-$^{1}/_{2}$ magic-angle spinning spectra of solids. Prog NMR Spectrosc 1992; 24: 435–456.
35. Böhm H-J, Banner D, Bendels S, Kansy M, Kuhn B, Müller K, Obst-Sander U, Stahl M. Fluorine in medicinal chemistry. ChemBioChem 2004; 5: 637–643.
36. Müller K, Faeh C, Diederich F. Fluorine in pharmaceuticals: looking beyond intuition. Science 2007; 318: 1881–1886.
37. Ando S, Harris RK, Scheler U. Fluorine-19 NMR of solids containing both fluorine and hydrogen. In: Grant DM, Harris RK,editors. Encyclopedia of nuclear magnetic resonance. Vol. 9. New York: Wiley; 2002.

38. Brittain HG, Morris KR, Bugay DE, Thakur AB, Serajuddin ATM. Solid-state NMR and IR for the analysis of pharmaceutical solids: polymorphs of fosinopril sodium. J Pharm Biomed Anal 1993; 11: 1063–1069.
39. Saulnier MG, Velaparthi U, Zimmermann K. Five-membered ring systems: with N and S (Se) atoms. Prog Heterocycl Chem 2004; 16: 228–271.
40. Edge S, Belu AM, Potter UJ, Steele DF, Young PM, Price R, Staniforth JN. Chemical characterization of sodium starch glycolate particles. Int J Pharm 2002; 240: 67–78.
41. Siedle AR. $^{11}$B NMR spectroscopy. Annu Rep NMR Spectrosc 1988; 20: 205–314.
42. Groziak MP. Boron heterocycles as platforms for building new bioactive agents. Prog Heterocycl Chem 2000; 12: 1–21.
43. Kliegel W. Boron compounds of pharmaceutical chemical interest. Pharmazie 1972; 27: 1–14.
44. Richardson PG. A review of the proteasome inhibitor bortezomib in multiple myeloma. Exp Opin Pharmacother 2004; 5: 1321–1331.
45. Hamaed H, Pawlowski JM, Cooper BFT, Fu RQ, Eichhorn SH, Schurko RW. Application of solid-state $^{35}$Cl NMR to structural characterization of hydrochloride pharmaceuticals and their polymorphs. J Am Chem Soc 2008; 130: 11056–11065.
46. Klein J, Ushio M, Burrell LS, Wenslow B, Hem SL. Analysis of aluminum hydroxyphosphate vaccine adjuvants by $^{27}$Al MAS NMR. J Pharm Sci 2000; 89: 311–321.
47. Zhang QW, Zhang HM, Usha MG, Wittebort RJ. $^{17}$O NMR and crystalline hydrates. Solid State NMR 1996; 7: 147–154.
48. Lemaitre V, Smith ME, Watts A. A review of oxygen-17 solid-state NMR of organic materials—towards biological applications. Solid State NMR 2004; 26: 215–235.
49. Wagler TA, Daunch WA, Panzner M, Youngs WJ, Rinaldi PL. Solid-state $^{33}$S MAS NMR of inorganic sulfates. J Magn Reson 2004; 170: 336–344.
50. Villa M, Tian F, Confrancesco P, Halamek J, Kasal M. High-resolution digital quadrature detection. Rev Sci Instrum 1996; 67: 2123–2129.
51. Momo F, Sotgiu A, Testa L, Zanin A. Digital frequency synthesizers for nuclear magnetic resonance spectroscopy. Rev Sci Instrum 1994: 65: 3291–3292.
52. Yun J, Yu J, Hongyan T, Gengying L. A complete digital radio-frequency source for nuclear magnetic resonance spectroscopy. Rev Sci Instrum 2002; 73: 3329–3331.
53. Doty D. Solid-state probe design. The encyclopedia of nuclear magnetic resonance. New York: Wiley; 1995. p 4475–4485.
54. Wasicki J, Kolenzo DP, Pnakov SE, Bilsi P, Pajzderska A, Hanckock BC, Medek A, Nawrcik W, Savenko BN. NMR search for polymorphic phase transformations in chlorpropamide form-A at high pressures. J Pharm Sci 2009; 98: 1426–1437.
55. Stringer JA, Drobny GP. Methods for the analysis and design of a solid-state nuclear magnetic resonance probe. J Chem Phys 1998; 69: 3384–3391.
56. Samoson A, Tuherm T, Past J, Reinhold A, Anupold T, Heinmaa I. New horizons for magic-angle spinning NMR. Top Curr Chem 2005; 246: 15–31.
57. Nelson BN, Schieber LJ, Barich DH, Lubach JW, Offerdahl TJ, Lewis DH, Heinrich JP, Munson EJ. Multiple-sample probe for solid-state NMR studies of pharmaceuticals. Solid State NMR 2006; 29: 204–213.
58. Alemany LB, Grant DM, Pugmire RJ, Alger TD, Zilm KW. Cross polarization and magic angle sample spinning NMR spectra of model organic compounds. 1. Highly protonated molecules. J Am Chem Soc 1983; 105: 2133–2141.

59. Antzutkin ON. Sideband manipulation in magic-angle spinning NMR. Prog NMR Spectros 1999; 35: 203–266.
60. Metz G, Wu X, Smith SO. Ramped-amplitude cross-polarization in magic-angle-spinning NMR. J Magn Reson A 1994; 110: 219–227.
61. Hodgkinson P. Heteronuclear decoupling in the NMR of solids. Prog NMR Spectrosc 2005; 46: 197–222.
62. Levitt MH, Bielecki A, Kolbert AC, Ruben DJ. High-resolution $^{1}$H NMR in solids with frequency-switched multiple-pulse sequences Solid State NMR 1993; 2: 151–163.
63. Vinogradov E, Madhu PK, Vega S. High-resolution proton solid-state NMR spectroscopy by phase-modulated Lee–Goldburg experiment. Chem Phys Lett 1999; 314: 443–450.
64. Sakellariou D, Lesage A, Hodgkinson P, Emsley L. Homonuclear dipolar decoupling in solid-state NMR using continuous phase modulation. Chem Phys Lett 2000; 319: 253–260.
65. Leskes M, Madhu PK, Vega S. Why does PMLG proton decoupling work at 65 kHz MAS? J Magn Reson 2009; 199: 208–213.
66. Salager E, Stein R, Steuernagel S, Lesage A, Elena B, Emsley L. Enhanced sensitivity in high-resolution $^{1}$H solid-state NMR spectroscopy with DUMBO dipolar decoupling under ultra-fast MAS. Chem Phys Lett 2009; 469: 336–341.
67. Xu M, Harris KDM. Altering the polymorphic product distribution in a solid-state dehydration process by rapid sample rotation in a solid-state NMR probe. J Am Chem Soc 2005; 127: 10832–10833.
68. Zell MT, Padden BE, Grant DJW, Schroeder SA, Wachholder KL, Prakash I, Munson EJ. Investigation of polymorphism in aspartame and neotame using solid-state NMR spectroscopy. Tetrahedron 2000; 56: 6603–6616.
69. Opella SJ, Frey MH. Selection of non-protonated carbon resonances in solid-state NMR. J Am Chem Soc 1979; 101: 5854–5856.
70. Wu X, Zilm KW. Complete spectral editing in CPMAS NMR. J Magn Reson A 1993; 102: 205–213.
71. Sakellariou D, Lesage A, Emsley L. Spectral editing in solid-state NMR using scalar multiple quantum filters. J Magn Reson 2001; 151: 40–47.
72. Lesage A, Steuernagel S, Emsley L. Carbon-13 spectral editing in solid-state NMR using heteronuclear scalar couplings. J Am Chem Soc 1998; 120: 7095–7100.
73. Cheung TTP. Spin diffusion in solids. The encyclopedia of nuclear magnetic resonance. New York: Wiley; 1996. p 4518–4524.
74. Feike M, Demco DE, Graf R, Gottwald J, Hafner S, Spiess HW. Broadband multiple-quantum NMR spectroscopy. J Magn Reson A 1996; 122: 214–221.
75. van Rossum BJ, Förster H, de Groot HJM. High-field and high-speed CP-MAS $^{13}$C NMR heteronuclear dipolar-correlation spectroscopy of solids with frequency-switched Lee–-Goldburg homonuclear decoupling. J Magn Reson 1997; 124: 516–519.
76. Vogt FG, Katrincic LM, Long ST, Mueller RL, Carlton RA, Sun YT, Johnson MN, Copley RCB, Light ME. Enantiotropically-related polymorphs of {4-(4-chloro-3-fluorophenyl)-2-[4-(methyloxy)phenyl]-1, 3-thiazol-5-yl} acetic acid. Crystal structures and multinuclear solid-state NMR. J Pharm Sci 2008; 97: 4756–4782.
77. Vogt FG, Clawson JS, Strohmeier M, Edwards AJ, Pham TN, Watson SA. Solid-state NMR analysis of organic cocrystals and complexes. Cryst Growth Des 2008; 9: 921–937.

78. van Rossum BJ, de Groot CP, Ladizhansky V, Vega S, de Groot HJM. A method for measuring heteronuclear ($^{1}$H-$^{13}$C) distances in high speed MAS NMR. J Am Chem Soc 2000; 122: 3465–3472.
79. Elena B, Emsley L. Powder crystallography by proton solid-state NMR. J Am Chem Soc 2005; 127: 9140–9146.
80. Elena B, Pintacuda G, Mifsud N, Emsley L. Molecular structure determination in powders by NMR crystallography from proton spin diffusion. J Am Chem Soc 2006; 128: 9555–9560.
81. Salager E, Stein RS, Pickard CJ, Elena B, Emsley L. Powder NMR crystallography of thymol. Phys Chem Chem Phys 2009; 11: 2610–2621.
82. Pickard CJ, Salager E, Pintacuda G, Elena B, Emsley L. Resolving structures from powders by NMR crystallography using combined proton spin diffusion and plane wave DFT calculations. J Am Chem Soc 2007; 129: 8932–8933.
83. Grant DM. Chemical shift tensors. The encyclopedia of nuclear magnetic resonance. New York: Wiley; 1995. p 1298–1321.
84. Antzutkin ON, Young YK, Levitt MH. $^{13}$C and $^{15}$N-chemical shift anisotropy of ampicillin and penicillin-V studied by 2D-PASS and CP/MAS NMR. J Magn Reson 1998; 135: 144–155.
85. Alderman DW, McGeorge G, Hu JZ, Pugmire RJ, Grant DM. A sensitive, high resolution magic angle turning experiment for measuring chemical shift tensor principal values. Mol Phys 1998; 95: 1113–1126.
86. Liu S-F, Mao J-D, Schmidt-Rohr K. A robust technique for two-dimensional separation of undistorted chemical-shift anisotropy powder patterns in magic-angle-spinning NMR. J Magn Reson 2002; 155: 15–28.
87. Amoureux J-P, Pruski M. Advances in MQMAS NMR. In: Grant DM, Harris RK,editors. The encyclopedia of nuclear magnetic resonance. New York: Wiley; 2002. p 226–251.
88. Rocha J, Morais CM, Fernandez C. Progress in multiple-quantum magic-angle spinning NMR spectroscopy. Top Curr Chem 2005; 246: 141–194.
89. Wickramasinghe NP, Parthasarathy S, Jones CR, Bhardwaj C, Long F, Kotecha M, Mehboob S, Fung LWM, Past J, Samoson A, Ishii Y. Nanomole-scale protein solid-state NMR by breaking intrinsic $^{1}$H$T_1$ boundaries. Nat Methods 2009; 6: 215–218.
90. Hoch JC, Stern AS. NMR data processing. New York: Wiley-Liss; 1996.
91. Kupce E, Nisshida T, Freeman R. Hadamard NMR spectroscopy. Prog NMR Spectrosc 2003; 42: 95–122.
92. Vogt FG, Mueller KT. Orientationally broadened NMR spectra: analysis by special processing methods. In: Grant DM, Harris RK, editors. The encyclopedia of nuclear magnetic resonance. New York: Wiley; 2002. p 112–125.
93. Zumbulyadis N, Antalek B, Windig W, Scaringe RP, Lanzafame AM, Blanton T, Helber M. Elucidation of polymorph mixtures using solid-state $^{13}$C CP/MAS NMR spectroscopy and direct exponential curve resolution algorithm. J Am Chem Soc 1999; 121: 11554–11557.
94. Duer M, editor. Solid-state NMR spectroscopy: principles and applications. Oxford: Blackwell Science; 2002.
95. Schmidt-Rohr K, Spiess HW. Multidimensional solid-state NMR and polymers. London: Academic Press; 1994.

96. Massiot D, Fayon F, Capron M, King I, Le Calvé S, Alonso B, Durand J, Bujoli B, Gan Z, Hoatson G. Modelling one- and two-dimensional solid state NMR spectra. Magn Reson Chem 2002; 40: 70–76.
97. Fukushima E, Roeder SBW. Experimental pulse NMR: a nuts and bolts approach. Massachusetts: Addison-Wesley; 1981.
98. Gao P. Determination of the composition of delavirdine mesylate polymorph and pseudopolymorph mixtures using $^{13}$C CP/MAS. Pharm Res 1996; 13: 1095–1104.
99. Offerdahl TJ, Salsbury JS, Dong Z, Grant DJW, Schroeder SA, Prakash I, Gorman EM, Barich DH, Munson EJ. Quantitation of crystalline and amorphous forms of anhydrous neotame using $^{13}$C CPMAS NMR spectroscopy. J Pharm Sci 2005; 94: 2591–2605.
100. Mehring M. High resolution NMR spectroscopy in solids. Berlin: Springer-Verlag; 1976.
101. Boumans PWJM. Detection limits and spectral interferences in atomic emission spectroscopy. Anal Chem 1994; 66: 459A–467A.
102. Stoyanova R, Kuesel AC, Brown TR. Application of principal-component analysis for NMR spectral quantitation. J Magn Reson 1995; 115: 265–269.
103. Ladroue C, Howe FA, Griffiths JR, Tate AR. Independent component analysis for automated decomposition of *in vivo* magnetic resonance spectra. Magn Reson Med 2003; 50: 697–703.
104. Zhang F, Bruschweiler R. Spectral deconvolution of chemical mixtures by covariance NMR. Chem Phys Chem 2004; 5: 794–796.
105. Nord LI, Vaag P, Duus JØ. Quantification of organic and amino acids in beer by $^{1}$H NMR spectroscopy. Anal Chem 2004; 76: 4790–4798.
106. Pelczer I. High-resolution NMR for metabomics. Cur Opin Drug Discov Dev 2005; 8: 127–133.
107. Pulkkinen J, Häkkinen A-M, Lundbom N, Paetau A, Kauppinen RA, Hiltunen Y. Independent component analysis to proton spectroscopic imaging data of human brain tumours. Eur J Radiol 2005; 56: 160–164.
108. Alam TM, Alam KM. Chemometric analysis of NMR spectroscopy data: a review. Annu Rep NMR Spectrosc. 2005; 5: 41–80.
109. Orendt AM, Facelli JC. Solid state effects on NMR chemical shifts. Annu Rep NMR Spectrosc 2007; 62: 115–178.
110. Strohmeier M, Orendt AM, Alderman D, Grant DM. Investigation of the polymorphs of dimethyl-3,6-dichloro-2, 5-dihydroxyterephthalate by $^{13}$C solid-state NMR spectroscopy. J Am Chem Soc 2001; 123: 1713–1722.
111. Harris RK, Joyce S, Pickard CJ, Cadars S, Emsley L. Assigning carbon-13 NMR spectra to crystal structures by the INADEQUATE pulse sequence and first principles computation: a case study of two forms of testosterone. Phys Chem Chem Phys 2006; 6: 137–143.
112. Barich DH, Clawson JS, Stueber D, Strohmeier M, Pugmire RJ, Grant DM. Determination of $^{13}$C chemical shift tensors in the presence of hydrogen bonding and $^{14}$N quadrupolar coupling: *p*-aminosalicylic acid, isoniazid, and pyrazinamide. J Phys Chem A 2002; 106: 11375–11379.
113. Harper JK, Barich DH, Heider EM, Grant DM. A combined solid-state NMR and X-ray powder diffraction study of a stable polymorph of paclitaxel. Cryst Growth Des 2005; 5: 1737–1742.

114. Harper JK, Grant DM, Zhang Y, Lee PL, Von Dreele RJ. Characterizing challenging microcrystalline solids with solid-state NMR shift tensor and synchrotron X-ray powder diffraction data: structural analysis of ambuic acid. J Am Chem Soc 2006; 128: 1547–1552.

115. Winkler B. An introduction to computational crystallography. Z Kristallogr 1999; 214: 506–527.

116. Maseras F, Morokuma K. IMOMM: A new integrated *ab initio* + molecular mechanics geometry optimization scheme of equilibrium structures and transition states. J Comput Chem 1995; 16: 1170–1179.

117. Stueber D, Guenneau F, Grant DM. The calculation of $^{13}$C chemical shielding tensors in ionic compounds utilizing point charge arrays obtained from Ewald lattice sums. J Chem Phys 2001; 114: 9236–9243.

118. Strohmeier M, Stueber D, Grant DM. Accurate $^{13}$C and $^{15}$N chemical shift and $^{14}$N quadrupolar coupling constant calculations in amino acid crystals: zwitterionic, hydrogen-bonded systems. J Phys Chem A 2003; 107: 7629–7642.

119. Pickard CJ, Mauri F. Calculations of magnetic resonance parameters in solids and liquids using periodic boundary conditions. In: Kaupp M, Bühl M, Malkin VG, editors. Calculation of NMR and EPR parameters. Weinheim: Wiley-VCH; 2004. p 265–278.

120. Yates JR, Dobbins SE, Pickard CJ, Mauri F, Ghic PY, Harris RK. A combined first principles computational and solid-state NMR study of a molecular crystal: flurbiprofen. Phys Chem Chem Phys 2005; 7: 1402–1407.

121. Carss SA, Harris RK, Fletton RA. Carbon-13 NMR investigations of three fluorinated steroids. Magn Reson Chem 1995; 33: 501–505.

122. Harris RK, Packer KJ, Thayer AM. Slow magic-angle rotation carbon-13 NMR studies of solid phosphonium iodides. The interplay of dipolar, shielding, and indirect coupling tensors. J Magn Reson 1985; 62: 284–297.

123. Likar MD, Taylor RJ, Fagerness PE, Hiyama Y, Robins RH. The 3′-keto-diol equilibrium of trospectomycin sulfate bulk drug and freeze-dried formulation: solid-state carbon-13 cross-polarization magic angle spinning (CP/MAS) and high-resolution carbon-13 nuclear magnetic resonance (NMR) spectroscopy studies. Pharm Res 1993; 10: 75–79.

124. Lu J, Rohani S. Polymorphism and crystallization of active pharmaceutical Ingredients (APIs). Curr Med Chem 2009; 16: 884–905.

125. Harris RK. Applications of solid-state NMR to pharmaceutical polymorphism and related matters. J Pharm Pharmacol 2007; 59: 225–239.

126. Brittain H. Polymorphism and solvatomorphism 2006. J Pharm Sci 2008; 97: 3611–3636.

127. Carss SA, Harris RK, Holstein P, Fletton RA. $^{19}$F NMR of proton-containing solids. Magn Res Chem 1996; 34: 63–70.

128. Wenslow RM. $^{19}$F solid-state NMR spectroscopic investigation of crystalline and amorphous forms of a selective muscarinic $M_3$ receptor antagonist, in both bulk and pharmaceutical dosage form samples. Drug Dev Ind Pharm 2002; 28: 555–561.

129. Katrincic LM, Sun YT, Carlton RA, Diederich AM, Mueller RL, Vogt FG. Characterization, selection, and development of an orally dosed drug polymorph from an enantiotropically related system. Int J Pharm 2009; 366: 1–13.

130. Thibert R, Tawashi R. Micronization of pharmaceutical solids. Microspheres, microcapsules and liposomes. Vol. 1. New York: Citus Books; 1999. p 327–347.

131. Vogt FG, Cohen DE, Bowman JD, Spoors GP, Zuber GE, Trescher GA, Dell'Orco PC, Katrincic LM, DeBrosse CW, Haltiwanger RC. Structural analysis of polymorphism and solvation in tranilast. J Pharm Sci 2005; 94: 651–665.

132. Wenslow RM, Baum MW, Ball RG, McCauley JA, Varsolona RJ. A spectroscopic and crystallographic study of polymorphism in an aza-steroid. J Pharm Sci 2000; 89: 1271–1285.

133. Reutzel SM, Russell VA. Origins of the unusual hygroscopicity observed in LY297802 tartrate. J Pharm Sci 1998; 87: 1568–1571.

134. David WIF, Shankland K. Structure determination from powder diffraction data. Acta Cryst 2008; A64: 52–64.

135. David WIF, Shankland K, McCusker LB, Baerlocher Ch, editors. Structure determination from powder diffraction data. Oxford: Oxford University Press; 2006.

136. Othman A, Harris RK, Hodgkinson P, Christopher EA, Lancaster RW. Structural characterization of two pharmaceutically important steroids by solid-state NMR. New J Chem 2008; 32: 1796–1806.

137. Yang J, Ho Y, Tzou DM. A $^{13}$C solid-state NMR analysis of steroid compounds. Magn Res Chem 2008; 46: 718–725.

138. Rheingold AL, Figueroa JS, Dybowski C, Beckmann PA. Superlattices, polymorphs, and solid-state NMR spin-lattice relaxation ($T_1$) measurements of 2,6-di-*tert*-butylnaphthalene. Chem Commun 2000; 651–652.

139. Beckmann PA, Burbank KS, Clemo KM, Slonaker EN, Averill K, Dybowski C, Figueroa JS, Glatfatter A, Koch S, Liable-Sands LM, Rheingold AL. $^{1}$H nuclear magnetic spin-lattice relaxation, $^{13}$C magic-angle-spinning nuclear magnetic resonance spectroscopy, differential scanning calorimetry, and X-ray diffraction of two polymorphs of 2,6-di-*tert*-butylnaphthalene. J Chem Phys 2000; 113: 1958–1965.

140. Jeffrey GA. An introduction to hydrogen bonding. New York: Oxford University Press; 1997.

141. Asakawa N, Kuroki S, Kurosu H, Ando I, Shoji A, Ozaki T. Hydrogen-bonding effect on carbon-13 NMR chemical shifts of L-alanine residue carbonyl carbons of peptides in the solid state. J Am Chem Soc 1992; 114: 3261–3265.

142. Kalsbeek N, Shaumburg K, Larsen S. Short hydrogen bonds in salts of dicarboxylic acids; structural correlations from solid-state $^{13}$C and $^{2}$H NMR spectroscopy. J Mol Struct 1993; 299: 155–170.

143. Naito A, Tuzi S, Saito H. A high-resolution $^{15}$N solid-state NMR study of collagen and related polypeptides. The effect of hydration on formation of interchain hydrogen bonds as the primary source of stability of the collagen-type triple helix. Eur J Biochem 1994; 224: 729–734.

144. Saito H, Tanaka Y, Nukada K. Hydrogen bond studies by nitrogen-14 nuclear magnetic resonance. III. Electron redistribution associated with hydrogen bonding and metal-complex formation of formamide and *N*-methylacetamide. J Am Chem Soc 1971; 93: 1077–1081.

145. Saito H, Nukada K. Hydrogen bond studied by nitrogen-14 nuclear magnetic resonance. II. Heteronuclear magnetic double resonance study of nitrogen-14 hydrogen-bond shifts of pyrroles and indole. J Am Chem Soc 1971; 93: 1072–1076.

146. Takegoshi K, Nakamura S, Terao T. $^{13}$C-$^{1}$H dipolar-assisted rotational resonance in magic-angle spinning NMR. Chem Phys Lett 2001; 334: 631–637.

147. Saleki-Gerhardt A, Ahlneck C, Zografi G. Assessment of disorder in crystalline solids. Int J Pharm 1994; 191: 237–247.
148. Hilden LR, Morris KR. Physics of amorphous solids. J Pharm Sci 2004; 93: 3–12.
149. Yu L. Amorphous pharmaceutical solids: preparation, characterization and stabilization. Adv Drug Deliv Rev 2001; 48: 27–42.
150. Cosgrove SD, Steele G, Auton TK, Plumb AP, Stensland B, Ferrari E, Harris KJ. Understanding the polymorphic behavior of sibenadet hydrochloride through detailed studies integrating structural and dynamical assessment. J Pharm Sci 2005; 94 (11): 2403–2415.
151. Othman A, Evans JSO, Evans IR, Harris RK, Hodgkinson P. Structural study of polymorphs and solvates of finasteride. J Pharm Sci 2007; 96(5): 1380–1397.
152. Luthra SA, Hodge IM, Utz M, Pikal MJ. Correlation of annealing with chemical stability in lyophilized chemical glasses. J Pharm Sci 2008; 97: 5240–5251.
153. Lubach JW, Xu D, Segmuller BE, Munson EJ. Investigation of the effects of pharmaceutical processing upon solid-state NMR relaxation times and implications to solid-state formulation stability. J Pharm Sci 2007; 96: 777–787.
154. Datta S, Grant DJW. Crystal structures of drugs: advances in determination, prediction and engineering Nat Rev Drug Discov 2004; 3: 42–47.
155. Vogt FG, Vena JA, Chavda M, Clawson JS, Strohmeier M, Barnett ME. Structural analysis of 5-fluorouracil and thymine solid solutions. J Mol Struct 2009; 932: 16–30.
156. Morris KR, Rodriguez-Hornedo N. Pharmaceutical hydrates. In: Swarbrick J, Boylan JC, editors. Encyclopedia of pharmaceutical technology. Vol. 7. New York: Marcel Dekker; 1993. p 393–440.
157. Perrier PR, Byrn SR. Influence of crystal packing on the solid-state desolvation of purine and pyrimidine hydrates: loss of water of crystallization from thymine monohydrate, cytosine monohydrate, 5-nitrouracil monohydrate, and 2′-deoxyadenosine monohydrate. J Org Chem 1982; 47: 4671–4676.
158. Khankari RK, Law D, Grant DJW. Determination of water content in pharmaceutical hydrates by differential scanning calorimetry. Int J Pharm 1992; 82: 117–127.
159. Stephenson GA, Groleau EG, Kleeman RL, Xu W, Rigsbee DR. Formation of isomorphic desolvates: creating a molecular vacuum. J Pharm Sci 1998; 87: 536–542.
160. Chen LR, Young VG, Lechuga-Ballesteros D, Grant DJW. Solid-state behavior of cromolyn sodium hydrates. J Pharm Sci 1999; 88: 1191–1200.
161. Ahlqvist MUA, Taylor LS. Water dynamics in channel hydrates investigated using H/D exchange. Int J Pharm 2002; 241: 253–261.
162. Reutzel-Edens SM, Bush JK, Magee PA, Stephenson GA, Byrn SR. Anhydrates and hydrates of olanzapine: crystallization, solid-state characterization, and structural relationships. Cryst Growth Des 2003; 3: 897–907.
163. Te RL, Griesser UJ, Morris KR, Byrn SR, Stowell JG. X-ray diffraction and solid-state NMR investigation of the single-crystal to single-crystal dehydration of thiamine hydrochloride monohydrate. Cryst Growth Des 2003; 3: 997–1004.
164. Apperly DC, Basford PA, Dallman CI, Harris RK, Kinns M, Marshall PV, Swanson AG. Nuclear magnetic resonance investigation of the interaction of water vapor with sildenafil citrate in the solid state. J Pharm Sci 2005; 94: 516–523.

165. Li ZJ, Abramov Y, Bordner J, Leonard J, Medek A, Trask AV. Solid-state acid-base interactions in complexes of heterocyclic bases with dicarboxylic acids: crystallography, hydrogen bond analysis, and $^{15}N$ NMR spectroscopy. J Am Chem Soc 2006; 128: 8199–8210.

166. Gardiennet-Doucet C, Henry B, Tekely P. Probing ionization state of functional groups by chemical shift tensor fingerprints. Prog NMR Spectrosc 2006; 49: 129–149.

167. Gu Z. Drueckhammer DG, Kurz L, Liu K, Martin DP, McDermott A. Solid state NMR studies of hydrogen bonding in citrate synthase inhibitor complexes. Biochemistry 1999; 38: 8022–8031.

168. Gu Z, Zambrano R, McDermott A. Hydrogen bonding of carboxyl groups in solid-state amino acids and peptides: comparison of carbon chemical shielding, infrared frequencies, and structures. J Am Chem Soc 1994; 116: 6368–6372.

169. Facelli JC, Gu Z, McDermott A. Carbon-13 chemical shift tensors of carboxylic acids. GIAO calculations in acetic acid + methylamine dimer. Mol Phys 1995; 86: 865–872.

170. Clawson JS, Vogt FG, Brum J, Sisko J, Patience DB, Dai W, Sharpe S, Jones AD, Pham TN, Johnson MN, Copley RCB. Formation and characterization of crystals containing a pleuromutilin derivative, succinic acid and water. Cryst Growth Des 2008; 8: 4120–4131.

171. Sheth AR, Lubach JW, Munson EJ, Muller FX, Grant DJW. Mechanochromism of piroxicam accompanied by intermolecular proton transfer probed by spectroscopic methods and solid-phase changes. J Am Chem Soc 2005; 127: 6641–6651.

172. Coppens P, Maly K, Petricek V. Composite crystals: what are they and why are they so common in the organic solid state? Mol Cryst Liq Cryst 1990; 181: 81–90.

173. Alamrsson Ö, Zaworotko MJ. Crystal engineering of the composition of pharmaceutical phases. Do pharmaceutical co-crystals represent a new path to improved medicines? Chem Commun 2004; 1889–1896.

174. Vishweshwar P, McMahon JA, Bis JA, Zaworotko MJ. Pharmaceutical co-crystals. J Pharm Sci 2006; 95: 499–516.

175. Variankaval N, Wenslow R, Murry J, Hartman R, Helmy R, Kwong E, Clas S-D, Dalton C, Santos I. Preparation and solid-state characterization of nonstoichiometric cocrystals of a phosphodiesterase-IV inhibitor and L-tartaric acid. Cryst Growth Des 2006; 6: 690–700.

176. Bugay DE, Findlay WP. Pharmaceutical excipients: characterization by IR, Raman, and NMR spectroscopy. New York: Marcel Dekker; 1999.

177. Fronczek FR, Kammel HN, Slattery M. Three polymorphs ($\alpha$, $\beta$, and $\delta$) of D-mannitol at 100 K. Acta Crystallogr C 2003; 59: O557–O570.

178. Nakagawa K, Murakami W, Andrieu J, Vessot S. Freezing step controls the mannitol phase composition heterogeneity. Chem Eng Res Des 2009; 87: 1017–1027.

179. Müller BW. Polymorphism of magnesium stearate and the influence of the crystal structure on the lubricating behavior of excipients. Acta Pharm Suecica 1981; 2: 74–75.

180. Lefort R, De Gussemem A, Willart J-F, Danede F, Decamps M. Solid state NMR and DSC methods for quantifying amorphous content in solid dosage forms: an application to ball milling of trehalose. Int J Pharm 2004; 280: 209–219.

181. Sato K. Crystallization behavior of fats and lipids—a review. Chem Eng Sci 2001; 56: 2255–2265.

182. Crowley KJ, Forbes RT, York P, Apperly DC, Nyqvist H, Camber O. Characterization of oleic acid and propranolol oleate mesomorphism using $^{13}C$ solid-state nuclear magnetic resonance spectroscopy (SSNMR). J Pharm Sci 2000; 89: 1286–1295.

183. Diniz TMFF, Tavares MIB. A high-resolution solid-state NMR investigation of molecular mobility of poly(methyl methacrylate)/poly(vinyl pyrrolidone)/poly(ethylene oxide) ternary blends. J Appl Poly Sci 2006; 100: 1492–1495.

184. Ishidaa M, Oshimaa J, Yoshinagaa K, Horiib F. Structural analysis of core-shell type polymer particles composed of poly(butyl acrylate) and poly(methyl methacrylate) by high-resolution solid-state $^{13}C$ n.m.r. spectroscopy. Polymer 1999; 40: 3323–3329.

185. Earnshaw RG, Price CA, O'Donnell JH, Whittaker AK. Determination of residual unsaturation in highly crosslinked, dough-moulded poly(methyl methacrylate) dental polymers by solid-state $^{13}C$ NMR. J Appl Polym Sci 1986; 32: 5337–5344.

186. Larsson PT, Wickholm K, Iverson I. A CP/MAS $^{13}C$ NMR investigation of molecular ordering in celluloses. Carbohydr Res 1997; 302: 19–25.

187. Gidley MJ, Bociek SM. $^{13}C$ CP/MAS NMR studies of amylose inclusion complexes, cyclodextrins, and the amorphous phase of starch granules: relationships between glycosidic linkage conformation and solid-state $^{13}C$ chemical shifts. J Am Chem Soc 1988; 110: 3820–3829.

188. Morgan KR, Furneaux RH, Larsen NG. Solid-state NMR studies on the structure of starch granules. Carbohydr Res 1995; 276: 387–399.

189. Morris KR, Griesser UJ, Eckhardt CJ, Stowell JG. Theoretical approaches to physical transformation of active pharmaceutical ingredients during manufacturing processes. Adv Drug Deliv Rev 2001; 48: 91–114.

190. Pikal MJ, Rigsbee D, Roy ML, Galreath D, Kovach KJ, Wang BQ, Carpenter JF, Cicerone MT. Storage stability of freeze-dried human growth hormone (hGH) with structure and dynamics in the glassy solid. J Pharm Sci 2008; 97: 5106–5121.

191. Chang C, Diaz LE, Morin F, Grant DM. Solid-state $^{13}C$ NMR study of drugs: aspirin. Magn Reson Chem 1986; 24: 768–771.

192. Jagannathan NR. High-resolution solid-state carbon-13 nuclear magnetic resonance study of acetaminophen: a common analgesic drug. Curr Sci 1987; 56: 827–830.

193. Saindon PJ, Cauchon NS, Sutton PA, Chang CJ, Peck GE, Byrn SR. Solid-state nuclear magnetic resonance (NMR) spectra of pharmaceutical dosage forms. Pharm Res 1993; 10: 197–203.

194. Brittain HG, Fiese EF. Effects of pharmaceutical processing on drug polymorphs and solvates. In: Brittain HG, editor. Polymorphs in pharmaceutical solids. Vol. 95. New York: Marcel Dekker; 1999. p 331–361.

195. Nobuyki S, Takao K. Evaluation of solid state form of troglitazone by solid-state NMR spectroscopy. J Pharm Biomed Anal 2005; 37: 177–181.

196. Stephanie S, Fabio Z, Stephanie V, Corinne D, Stafano C. Improved solid-state NMR quantification of active principles in pharmaceutical formulations. J Pharm Biomed Anal 2008; 47: 683–687.

197. Harris RK, Hodgkinson P, Larsson T, Muruganantham A. Quantification of bambuterol hydrochloride in a formulated product using solid-state NMR. J Pharm Biomed Anal 2005; 38: 858–864.

198. Booy K-J, Wiegerinck P, Vader J, Kaspersen F, Lambregts D, Vromans H, Kellenbach E. The use of $^{13}C$ labeling to enhance the sensitivity of $^{13}C$ solid-state CPMAS NMR to study polymorphism in low dose solid formulations. J Pharm Sci 2005; 94: 458–463.

199. Liu J, Nagapudi K, Kiang Y-H, Martinez E, Jona J. Quantification of compaction-induced crystallinity reduction of a pharmaceutical solid using $^{19}F$ solid-state NMR and powder X-ray diffraction. Drug Dev Ind Pharm 2009; 35: 969–975.

200. Griffin JM, Martin DR, Brown SP. Distinguishing anhydrous and hydrous forms of an active pharmaceutical ingredient in a tablet formulation using solid-state NMR spectroscopy. Angew Chem Int Ed Engl 2007; 119: 8182–8184.

201. Farrer BT, Peresypkin A, Wenslow RM. Quantitation of crystalline material within a liquid vehicle using $^{1}H/^{19}F$ CP/MAS NMR. J Pharm Sci 2007; 96: 264–267.

202. Changsan N, Chan H-K Separovic F, Srichana T. Physiochemical characterization and stability of rifampicin liposome dry powder formulation for inhalation. J Pharm Sci 2009; 98: 628–639.

203. Pham TN, Watson SA, Edwards AJ, Chavda M, Clawson JS, Strohmeier M, Vogt FG. Analysis of amorphous solid dispersions using 2D solid-state NMR and $^{1}H\,T_1$ relaxation measurements. Mol Pharm 2010; 7: 1667–1691.

204. Schachter DM, Xiong J, Tirol GC. Solid state NMR perspective of drug-polymer solid solutions: a model system based on poly(ethylene oxide). Int J Pharm 2004; 281: 89–101.

205. Geppi M, Guccione S, Mollica G, Pignatello R, Veracini CA. Molecular properties of ibuprofen and its solid dispersions with Eudragit RL100 studied by solid-state nuclear magnetic resonance. Pharm Res 2005; 22: 1544–1555.

206. Mollica G, Geppi M, Pignatello R, Veracini CA. Molecular properties of flurbiprofen and its solid dispersions with Eudragit RL100 studied by high- and low-resolution solid-state nuclear magnetic resonance. Pharm Res 2006; 23: 2129–2140.

207. Aso Y, Yoshioka S, Miyazaki T, Kawanishi T. Feasibility of $^{19}F$-NMR for assessing the molecular mobility of flufenamic acid in solid dispersions. Chem Pharm Bull 2009; 57: 61–64.

208. Sotthivirat S, Lubach JW, Haslam JL, Munson EJ, Stella VJ. Characterization of prednisolone in controlled porosity osmotic pump pellets using solid-state NMR spectroscopy. J Pharm Sci 2007; 96: 1008–1017.

209. Rohrs BR, Thamann TJ, Gao R, Stelzer DJ, Bergren MS, Chao RS. Tablet dissolution affected by a moisture mediated solid-state interaction between drug and disintegrant. Pharm Res 1999; 16: 1850–1856.

210. Markovich RJ, Anderson EC, Coscolluela CB, Zibas SA, Rosen J. Spectroscopic identification of an amorphous-to-crystalline drug transition in a solid dispersion SCH 48461 capsule formulation. J Pharm Biomed Anal 1997; 16: 661–673.

211. Byrn SR, Xu W, Newman AW. Chemical reactivity in solid-state pharmaceuticals: formulation implications. Adv Drug Deliv Rev 2001; 48: 115–136.

212. Sakellariou D, le Goff G, Jacquinot J-F. High-resolution, high-sensitivity NMR of nanolitre anisotropic samples by coil spinning. Nature 2007; 447: 694–698.

213. Mizuno T, Hioka K, Katsuya F, Fujioka K, Takegoshi K. Development of a magic-angle spinning nuclear magnetic resonance probe with a cryogenic detection system for sensitivity enhancement. Rev Sci Instrum 2008; 79: 1–6.

214. Harel E, Schröder L, Xu S. Novel detection schemes of nuclear magnetic resonance and magnetic resonance imaging: applications from analytical chemistry to molecular sensors. Ann Rev Anal Chem 2008; 1: 133–164.

215. Frericks Schmidt HL, Sperling LJ, Gao YG, Wylie BJ, Boettcher JM, Wilson SR, Rienstra CM. Crystal polymorphism of protein GB1 examined by solid-state NMR spectroscopy and X-ray diffraction. J Phys Chem B 2007; 111: 14362–14369.

# TARGET SELECTION IN DRUG DISCOVERY

Charles F. Albright
*Bristol-Myers Squibb, Neuroscience Discovery, Research and Development, Wallingford, Connecticut*

## 1 INTRODUCTION

The annual output of the pharmaceutical industry has been relatively constant from 1993 to 2003 with an average of 30 new molecular enties (NME) per year [1, 2]. Only about 10% of these NMEs were directed at novel targets as identified by the Food and Drug Administration (FDA) [1]. This relatively constant output and low innovation has occurred despite a twofold increase in investment by both the National Institutes of Health (NIH) and the pharmaceutical industry [2] as well as dramatic advancements in technology, such as sequencing the human genome, transcriptomics, and expanded small-molecule collections. The combination of the constant output plus increased investment has driven the cost per successful NME from $1.1 billion for the period 1995–2000 to $1.7 billion for the period 2000–2002 [2]. These trends clearly threaten the long-term health of the pharmaceutical industry.

The reasons for attrition of therapeutic compounds have been studied to help understand how to increase the number of NMEs and innovative therapies. Based on an analysis of investigational new drugs (IND) that were abandoned in the period 1981–1992, one study found that compound attrition was due to a lack of efficacy (35%), economics (32%), safety (20%), and other reasons (13%) [3]. In a separate study, data gathered directly from 10 large pharmaceutical companies for the period 1991–2000 showed that attrition was primarily due to lack of efficacy (about 30%) and safety (about 30%) [4]. Taken together, these studies indicate that efficacy, safety, and economics are the major reasons for attrition. Clearly, identifying targets that are well linked to the disease, likely to have an acceptable therapeutic index with respect to mechanism-based toxicity, and appear economically attractive will decrease attrition. This review will focus on these criteria and an additional criterion, chemical tractability, that assesses the likelihood of identifying a compound with the desired pharmacology and pharmacokinetics. While every target in drug discovery brings unexpected challenges, an accurate assessment of targets with respect to these four criteria will help increase the likelihood of developing successful therapeutics. Other reviews with additional perspectives on target selection for drug discovery are available [5–7].

## 2 CRITERION 1: TARGET LINKAGE TO DISEASE

Connecting a target to the disease is the most important criterion for selecting a target for drug discovery. There are several ways to make this connection, including

1. Human clinical results
2. Human genetics
3. Pathway linked to the disease by human clinical results or human genetics
4. Preclinical studies in model systems
5. Target expression perturbion in the disease

In practice, good targets are supported by several of these connections.

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

### 2.1 Target Linked to Disease by Human Clinical Results

Human clinical results provide the strongest support for a compound and, consequently, its target in a disease. These clinical results can be with an approved compound or a compound that was efficacious but failed in clinical trials for reasons that could be addressed with an improved compound. For instance, a compound that was abandoned due to off-target toxicity or poor pharmaceutical properties may nonetheless strongly link a target to the disease.

Recent results with a calcitonin gene-related peptide (CGRP) antagonist provide an excellent example of target validation with a compound that has pharmaceutical issues. In a clinical study, intravenous administration of BIBN 4096 BS was effective in treating acute migraine attacks [8]. Since intravenous administration is not an ideal delivery method for a migraine therapy, CGRP antagonists that are delivered using a more convenient method, such as oral administration, would be a preferred therapy relative to BIBN 4096 BS.

Positive clinical results with a biological therapy can link a target to a disease that could be targeted with small molecules. For instance, the efficacy of trastuzumab and cetuximab, biologics that bind the ErbB2 and ErbB1 receptor tyrosine kinases, respectively, validated these targets for certain cancers. Based on this validation, small-molecule kinase inhibitors of ErbB1, gefitinib and erlotinib, and ErbB1 plus ErbB2, lapatinib, were developed [9]. In addition, a biologic that prevents the dimerization of ErbB2 with its partners including ErbB1, pertuzumab, has also been developed [9].

When pursuing targets that are linked to the disease by approved therapies, the ability to to differentiate the new therapy from the existing therapies is an important consideration. A lack of differentiation can create economic issues (criterion 4), which will be discussed further in a later section.

### 2.2 Target Linked to Disease by Human Genetics

Familial diseases, where a mutation in a gene causes or significantly contributes to the disease, provide a strong link between the target and the disease. An analysis of the 100 top-selling drugs in 2003 supports this notion [1]. In particular, these 100 drugs were directed toward 43 distinct proteins. Of these 43 targets, 12 targets had mutations in the corresponding gene that caused a monogenic disorder, a disease where a mutation in a single gene was sufficient to cause the disease. Importantly, these therapies are used for patients with sporadic forms of the disease, not just patients with familial forms of the disease. The development of imatinib, an inhibitor of the Abl (V-abl Abelson murine leukemia viral oncogene homolog 1) tyrosine kinase, exemplifies a target where human genetics played a large role in target selection. In particular, 95% of patients with chronic myelogenous leukemia (CML) contain a chromosomal translocation that fuses the breakpoint cluster region (*Bcr*) gene and *Abl* gene, thereby creating a gene that encodes the Bcr-Abl fusion protein [10]. Studies showed that the Bcr-Abl protein functions as a constitutively active tyrosine kinase that increases proliferation and decreases cell death; both effects required the kinase activity of the Bcr-Abl protein. Based on these results, a small-molecule inhibitor of the Bcr-Abl kinase, imatinib, was discovered and developed.

There are over 1800 human monogenic diseases that identify potential targets for therapeutic intervention [11]. While many of these targets are not tractable (criterion 2) or economically viable (criterion 4), attractive targets for monogenic diseases remain. For example, loss-of-function mutations in the gene encoding PCSK9 decrease plasma low-density lipoprotein cholesterol (LDL-C) levels suggesting that PCSK9 inhibitors would be therapeutically beneficial for atherosclerosis [12]. While there are relatively few targets for major unmet medical need that have corresponding monogenic diseases, monogenic diseases will continue to provide an important source of targets for less prevalent indications, such as those eligible for orphan drug status, which will be discussed in criterion 4.

In addition to monogenic diseases, identifying genes that increase disease susceptibility can identify targets. For example, apolipoprotein E (*ApoE*) alleles affect susceptibility to Alzheimer's disease (AD). Of the three common *ApoE* alleles, which include *ApoE2*, *ApoE3*, and *ApoE4*, studies showed that the *ApoE4* gene dosage increased the risk and onset age for AD while *ApoE2* decreased the risk and onset age for AD [13, 14]. Based on these findings, therapeutic strategies aimed at *ApoE4* are being pursued for AD [15]. A large effort to identify new targets at GlaxoSmithKline by gene association studies was recently described [16]. In addition, the HapMap project identified a collection of single nucleotide polymorphisms to detect linkage disequilibrium throughout the genome in multiple populations. The successful completion of this effort [17] should aid the detection of susceptibility loci for all diseases including more complex, polygenic diseases such as schizophrenia. Targets identified due to disease association studies raise issues not encountered with targets identified by monogenic diseases. In particular, association of a target with a disease could indicate that this target only influences the disease in a subset of patients or affects only limited aspects of the disease. Despite these issues, gene association studies offer the potential to greatly expand

the number of high-quality targets for more complex diseases.

### 2.3 Target on Pathway Linked to Disease by Human Clinical Data or Human Genetics

Understanding the pathways affected by targets linked to the disease by human clinical data or human genetics can lead to new targets. The development of HMG-CoA inhibitors, or statins, to reduce cholesterol levels exemplifies this approach. The identification of a mutation in the LDL receptor that greatly increased both plasma LDL-C levels and the risk of myocardial infarction strongly implicated plasma LDL-C levels in atherosclerosis [18]. This finding stimulated the search for methods to decrease plasma LDL levels that subsequently led to statins. In particular, investigators found that the inhibition by compactin of HMG-CoA reductase, the rate-limiting step in cholesterol biosynthesis, increased LDL receptor expression in the liver and thereby decreased plasma LDL-C levels. Positive clinical results with compactin began the development of HMG(3-hydroxy-3-methyl-glutaryl)-CoA reductase inhibitors for humans.

A similar argument is driving much of the current drug discovery effort for AD. In particular, mutations in the genes encoding the amyloid precursor protein, presenilin-1, and presenilin-2 lead to inherited, early-onset forms of AD [19]. An analysis of these mutants has led to the hypothesis that amyloid beta (Aβ42), a peptide formed by the cleavage of the amyloid precursor protein by BACE (beta-site APP cleaving enzyme) and γ-secretase, is essential for the initiation and progression of AD. Presenilin-1, one therapeutic target for reducing Aβ42, forms part of the γ-secretase enzyme that produces Aβ42 and is therefore linked to the disease by a monogenic disease. Other approaches to inhibit Aβ42, including BACE inhibitors, anti-Aβ antibodies, and Aβ42 aggregation inhibitors, are linked to AD due to their role in the Aβ pathway [20].

### 2.4 Target Linked to Disease by Preclinical Studies

Screening for compounds that produce a desired physiologic effect in a disease model was the typical method for drug discovery before the advent of target-based drug discovery [21]. For example, thiazolidinedione derivatives were known to increase insulin sensitivity in rodents [22] for over a decade before it was recognized that they were ligands for peroxisome proliferators-activated receptor γ [23]. The poor success rate of target-based drug discovery and technologic advancements have prompted a resurgence of interest in the discovery of therapeutics and targets using disease models. For purposes of this review, the identification of therapeutics based on their effects in disease models without requiring prior knowledge of the therapeutic target will be termed forward pharmacology. Forward pharmacology can employ both standard cellular assays, such as cytotoxicity or signaling, more complicated cellular assays, such as highly engineered systems [24] or parallel assays for multiple pathways [25], and in vivo disease models. While these forward pharmacology approaches occur without knowing the molecular target, such identification is usually possible using proteomics, genetics, or other methods.

The use of forward pharmacology is advantageous since this approach may identify targets that are either novel or of suspect chemical tractability (criterion 3) based on precedent. The reliance of this method on a disease model is a potential limitation since most human diseases are often poorly modeled. For example, diseases that take decades to develop are typically compressed into months in a rodent. Further, validation of these models is frequently based on known therapies and, therefore, the predictive value of these models for novel targets is unknown. Additionally, assessment of mechanism-based toxicity (criterion 2) requires target identification, which can require a significant effort.

The tractability of genetic studies in model systems, including *Saccharomyces cerevisiae*, *Caenorhabditis elegans, Drosophila melanogaster*, *Danio rerio*, *Mus musculus*, and cultured human cells has stimulated the search for new targets using these systems [26]. These searches can use several methods such as forward genetic screens, forward pharmacology, and RNA interference (RNAi) screens. A retrospective analysis of targets for known drugs using mouse knockouts provides a strong case for the validity of this approach [1]. In particular, the 100 top-selling drugs in 2003 targeted 43 distinct proteins. The mouse gene that corresponded to these human targets was knocked out for 34 of the 43 targets. Encouragingly, the knockout phenotypes for 29 of these 34 genes were judged informative with respect to identifying therapeutic targets. While a similar analysis has not been conducted for other genetic systems, the frequent conservation of protein function in these model systems suggests that results from these systems will usually be applicable to human targets [27].

Genetic approaches to target identification are advantageous since they can rapidly identify novel targets, including an assessment of both efficacy and the potential for mechanism-based toxicity. As for the forward pharmacology approach, the genetic approach depends critically on the assay to identify mutants. An additional problem is that humans may have multiple genes that could perform the function of the gene identified in the model organism. Understanding the role of

these human homologs in the pathogenesis of the human disease can require significant effort.

### 2.5 Target Linked to Disease by Altered Expression in Disease Tissue

The explosion of methods to profile gene and protein expression has stimulated the search for new targets based on expression differences between normal and disease tissue. In particular, transcriptomics, proteomics, and metabonomics have led to a large number of genes, proteins, and metabolites, respectively, that are differentially expressed in disease tissues. While this data can be the basis for further investigation of a target, altered expression by itself only weakly links a target to the disease since it is not possible to determine whether the altered expression is a cause or an effect of the disease.

## 3 CRITERION 2: POTENTIAL THERAPEUTIC INDEX

The potential therapeutic index of a new therapy is the next question that should be addressed once the target is linked to the disease. The therapeutic index is the ratio of the compound concentration/dose needed for efficacy divided by the concentration/dose that causes toxicity. An effective therapeutic requires a therapeutic index significantly greater than unity. While the therapeutic index may be limited by either compound-specific or mechanism-specific toxicity, efforts should be made to understand mechanism-based toxicity at the time of target selection. In parallel, efforts should be undertaken to understand the amount of target perturbation that will be required for a therapeutic effect. A knowledge of likely mechanism-based toxicity can then be used with tool compounds in disease models to estimate the potential therapeutic index.

A phenotypic analysis of mouse knockouts for the target gene is an excellent starting point to determine potential mechanism-based toxicities. Since a retrospective analysis of current therapeutic targets shows that most knockouts provide evidence of the therapeutic effect [1], it follows that most knockouts will also provide evidence of mechanism-based toxicity. The applicability of mouse knockout phenotypes to human therapeutics can, however, be confounded by several factors, including the effect of knockouts on developmental processes, compensation for the knockout phenotype by altered expression of other genes, and differences in gene function between mice and humans. Despite these potential issues, the phenotype of a knockout of the gene encoding the therapeutic target protein should be determined to identify potential mechanism-based toxicities. Tool molecules can also be helpful in identifying mechanism-based toxicities. Multiple tool molecules from divergent chemical series are usually required to differentiate compound-specific and mechanism-specific toxicity.

In most cases, tool molecules are required to determine the amount of target perturbation required for a therapeutic effect. Such tool molecules will usually be available when the target is identified by clinical compounds. For most novel targets, tool compounds of sufficient quality for these studies will be lacking until well into the drug discovery process. In these cases, a genetic analysis may be helpful. For example, heterozygous knockouts can be crossed into disease-relevant models to determine whether a therapeutic effect is observed. A positive result in these studies combined with an understanding of the effect of the heterozygous knockout on activity of the target provides an estimate of the amount of target perturbation required for efficacy. In rare cases, alleles that cause a range of perturbation of target activity may be available for study in either model systems or from human samples.

## 4 CRITERION 3: CHEMICAL TRACTABILITY

The potential to identify an acceptable clinical compound for either small-molecule targets or biologic targets is an important consideration when selecting a new target. The criteria for small-molecule targets and biologic targets differ significantly and will be considered separately.

### 4.1 Small-Molecule Target Tractability

Small-molecule therapeutics must have acceptable potency, selectivity, bioavailability, pharmacokinetics, and pharmaceutical properties. These constraints have limited the number of targets for which successful therapeutics have been developed. The biochemical classes of small-molecule targets have been collated in three recent reviews. In the first review, the targets of all 483 approved therapeutics as of 1996 contained primarily receptors (45%), which were largely G-protein coupled receptors (GPCR), and enzymes (28%) (Table 1) [21, 28]. In a second review, Hopkins and Groom identified 399 targets in the Investigational Drugs Database and Pharmaprojects Database that had compounds with binding affinities below $10\,\mu M$ that complied to Lipinski's rules [29]. Lipinski's rules, which were developed by evaluating successful therapeutics, state that small-molecule therapeutics should contain fewer than 5 hydrogen bond donors, molecular mass less than 500 Da, low lipophilicity ($c \log P < 5$), and less than 10 hydrogen bond acceptors [30]. These 399 targets contained 120 marketed therapeutics that were comprised

**TABLE 1 Small-Molecule Target Classes for Marketed Drugs**

| Target Class | Percentage of Therapeutics in Each Target Class | | |
|---|---|---|---|
| | First Review[a] | Second Review[b] | Third Review[c] |
| Receptor and GPCR | 45 | 34 | 23 |
| Enzymes | 28 | 47 | 50 |
| Hormones and factors | 11 | — | 3 |
| Nuclear receptors | 2 | 4 | 6 |
| Channels and transporters | 5 | 11 | 12 |
| Other | 9 | 8 | 6 |

[a]Based on data from [21, 28].
[b]Based on data from [29].
[c]Based on data from [31].

primarily of receptors (34%) and enzymes (47%) (Table 1). In a third review, authors collated the targets of the 268 successful drugs [31] contained in the Therapeutic Target Database [32] and also found that small-molecule targets were primarily receptors (23%) and enzymes (50%) (Table 1).

While classifying small-molecule targets as receptors and enzymes is helpful, these broad categories mask much of the complexity of small-molecule target classification. To better understand this complexity, Hopkins and Groom [29] took the 399 targets identified above and grouped them into families based on InterPro domains [33]. The InterPro database is an annotated collection of protein domains based on several algorithms for domain identification [33]. These 399 targets represented 130 protein families where nearly half of the targets were represented by 6 families that included GPCRs, serine-threonine and tyrosine kinases, zinc metallopeptidases, serine proteases, nuclear hormone receptors, and phosphodiesterases. A compilation of InterPro domains that are targeted by more than 3 compounds is shown in Table 2. Interestingly, the majority of proteins that contained these 130 InterPro domains did not have small-molecule compounds in the

**TABLE 2 InterPro Domains Targeted by Small-Molecule Therapeutics**

| InterPro Domain Name | Targets per InterPro Domain | Human Proteins with InterPro Domain |
|---|---|---|
| Serine-threonine protein kinase | 22 | 402 |
| Rhodopsin-like GPCR | 92 | 401 |
| Tyrosine protein kinase | 14 | 264 |
| Cation channel | 7 | 137 |
| Serine protease | 12 | 115 |
| Tyrosine protein phosphatase | 3 | 112 |
| Neutral zinc metallopeptidase | 14 | 65 |
| GPCR family 2 | 5 | 64 |
| Cytochrome P450 | 7 | 64 |
| Ligand binding domain of nuclear hormone receptor | 13 | 62 |
| Short-chain dehydrogenase | 3 | 47 |
| Neurotransmitter-gated ion channel | 6 | 43 |
| 3′-5′-Cyclic nucleotide phosphodiesterase | 12 | 36 |
| Sodium-neutrotransmitter symporter | 4 | 23 |
| Intergrins, α-chain | 5 | 22 |
| Eukaryotic cysteine protease | 6 | 21 |
| Ionotrophic glutamate receptor | 4 | 20 |
| GPCR family 3 | 5 | 16 |
| FAD and NAD oxidoreductase | 3 | 14 |
| Eukaryotic and viral aspartic protease | 3 | 11 |
| Animal haem peroxidase | 3 | 10 |
| Metallopeptidase, family M24 | 3 | 9 |
| Integrins, β-chain | 3 | 9 |
| MAPEG | 3 | 6 |
| **Subtotal** | **252** | **1973** |
| **Other** | **124** | **1078** |
| **Total** | **376** | **3051** |

*Source*: Based on data from [29].

databases used for this analysis. For instance, only 22 of 402 serine-threonine kinases and 12 of 115 serine proteases contained small-molecule targets. This analysis also revealed a diverse array of targets in relatively small families. For example, HMG-CoA reductase, the target of statin therapeutics, is part of an InterPro family that contains only a single member.

In a separate analysis, Zheng and co-workers identified 268 targets for successful small-molecule therapies and 1267 targets for investigational therapies in the Therapeutic Target Database [31]. Within these targets, they found identifiable drug-binding domains in 190 successful targets and 1035 research targets. These drug-binding domains were then grouped into families using the Pfam database to identify and define protein families [34]. This grouping showed that successful targets contained 88 Pfam domains and research targets contained 357 Pfam domains. About half of the successful and research targets were contained in 11 Pfam families and 24 Pfam families, respectively (Table 3). There was a significant overlap in the Pfam families that were frequently contained in successful and research targets. Not surprisingly, these frequently identified Pfam families, which included GPCRs, protein kinases, and nuclear hormone receptors, were similar to the frequently identified InterPro domains identified by Hopkins and Groom [29] (Table 2).

Taken together, these studies show that about half of small-molecule therapeutics are targeted at relatively large protein families, such as protein kinases and GPCRs, where successful therapeutics were previously identified. Identifying and optimizing leads for these therapeutic classes will frequently be easier since inhibitors of related proteins will often provide useful leads and insight into compound optimization. Challenges within these classes will include identifying selective molecules and ensuring an acceptable intellectual property position. These studies also suggest that about half

**TABLE 3 Protein Families Targeted by Small-Molecule Therapeutics**

| | Targets per Pfam domain | |
|---|---|---|
| Pfam Family Name | Successful Targets | Research Targets |
| 7-Transmembrane receptor rhodopsin family | 32 | 94 |
| Nuclear hormone receptor | 11 | 16 |
| Zinc finger | 11 | 15 |
| Ion transport protein | 7 | 18 |
| Protein kinase | 5 | 61 |
| Short-chain dehydrogenase | 4 | 6 |
| Amino acid permease | 4 | |
| Cytochrome P450 | 4 | |
| Neurotransmitter-gated ion channel 1 | 4 | 6 |
| Fibronectin type III domain | 4 | 15 |
| Sodium/neurotransmitter symporter | 3 | |
| Immunoglobulin | | 25 |
| Trypsin | | 21 |
| SH2 domain | | 17 |
| Receptor family ligand binding region | | 12 |
| Phorbol ester/diacyl glycerol binding domain | | 12 |
| Leucine-rich repeat | | 12 |
| Ankyrin repeat | | 11 |
| Papain family cysteine protease | | 10 |
| Lectin C-type domain | | 10 |
| Matrixin | | 9 |
| Small cytokines | | 9 |
| 3′-5′-Cyclic nucleotide phosphodiesterase | | 8 |
| Hemopexin | | 8 |
| ATP-binding cassette transporter | | 7 |
| Hormone receptor | | 7 |
| Eukaryotic-type carbonic anhydrase | | 7 |
| **Subtotal** | **89** | **416** |
| **Other Pfam domains** | **101** | **619** |
| **Total** | **190** | **1035** |

*Source*: Based on data from [31].

of small-molecule therapies will come from protein families that contain relatively few members. Identifying and optimizing leads for these targets is likely to be the largest challenge.

### 4.2 Biologic Target Tractability

Biologic therapeutics can be broadly grouped into three categories: factors, enzymes, and binding proteins. Factors include peptidic hormones, such as insulin, erythropoietin, and growth hormone. Factors are typically agonists, although dominant-negative factors may be possible. Enzymes are used either for enzyme replacement, such as acid β-glucosidase or factor VIIa, or to catalyze the removal of a toxic species formed in a disease state, such as tPA (tissue plasminogen activator) to dissolve a thrombus. Binding proteins include primarily antibodies that neutralize factors, such as anti-TNF (tumor necrosis factor) and anti-VEGF (vascular endothelial growth factor) antibodies, or bind to cell surface proteins, such as anti-ErbB1 and anti-ErbB2 antibodies, to interfere with protein function or recruit antibody effectors. While antibodies represent the bulk of binding proteins, non-antibody-binding proteins have also been developed. A review of the current biological targets that are contained within the DrugBank [35] shows 56 unique targets for biologics that are divided between factors (38%), enzymes (29%), and binding proteins (34%).

The assessment of the chemical tractability of biological targets should include information about target location, required specificity, target amount, and duration of exposure needed. Most biologic targets are extracellular proteins or extracellular regions of plasma membrane-spanning proteins in peripheral tissues since biologics have limited ability to cross membranes. The number of binding sites, required duration of effect, and turnover of the ligand are also important factors. It will be increasingly important to develop models that incorporate these variables to set the design criteria for a biologic [36].

## 5 CRITERION 4: ECONOMIC CONSIDERATIONS

Economic considerations are an important factor in target selection. When selecting a target for drug discovery one must consider whether the resulting compound is likely to generate revenues that will yield an acceptable rate of return on the investment required to discover, develop, and market the therapy. While a detailed financial analysis is neither practical nor needed at the time of target selection, some estimate of the potential revenues and expenses should be incorporated into evaluation of the target. There are scenarios where an acceptable rate of return is not the proper criterion. One such example is the development of therapeutics for developing countries that are funded by nonprofit organizations. In such cases, the magnitude of the unmet medical need and expenses should be considered when evaluating targets.

The expected revenues from a therapeutic agent will be affected by the total market size, market share, price, and exclusivity period. It is usually straightforward to estimate the total market size and exclusivity period for a new chemical entity. Estimating the market share and price are more complicated issues that depend on the magnitude of the unmet medical need that will be addressed by the new therapeutic. In particular, if a new therapeutic addresses a need that is clearly recognized by patients and payors, then a new therapeutic has the potential to capture a significant market share at a premium price even if existing therapies are available. Conversely, if a new therapeutic performs similarly to available generic therapies, then a large market share and premium pricing are unlikely. Recent experience with acetylcholinesterase inhibitors (AChEI) for the treatment of patients with mild and moderate AD illustrates the growing role payors will play in economics. In particular, a meta-analysis of randomized double-blind placebo-controlled clinical trials with AChEIs showed that only 10% of patients responded to these therapies [37]. Based on this analysis, the United Kingdom's National Institute for Health and Clinical Excellence recommended that AChEIs be used only for patients with moderate AD even though they were approved for both mild and moderate AD patients. Clearly, this recommendation may impact the economics of AChEIs since mild AD patients may not use these therapies if they are not reimbursed.

The expected costs for a new therapeutic agent will be affected by development and marketing costs. For each successful therapeutic compound, clinical development is estimated to cost $1.7 billion [2]. Furthermore, only 11% of compounds that enter human clinical studies are successfully registered [4]. Even these typical values vary between indications. For example, oncology drugs have a 5% success rate while cardiovascular drugs have a 20% success rate. In addition to these success rates, development costs will be significantly affected by the clinical stage at which unsuccessful compounds are terminated. Hence, indications where efficacy can be rapidly assessed in the clinic will reduce the cost of attrition due to lack of efficacy. For example, the effect of antiviral therapy on viral load can often be determined after a single administration. In other cases, it may be possible to accelerate registration based on surrogate

endpoints. In particular, drugs may be approved for serious or life-threatening diseases based on surrogate endpoints if the therapy provides significant benefits relative to existing therapy. Several drugs for cancer have received accelerated approval in the past decade based on the endpoints of objective tumor response rates and time to progression, which serve as surrogates for survival [38]. Finally, development costs can be reduced if the therapy targets an orphan indication. While the definition of orphan indication varies between countries, a rare disease in the United States is defined as affecting less than 200,000 individuals [39]. A variety of incentives, potentially including 7 years market exclusivity, tax credits, and accelerated approval, may be available for orphan drugs depending upon the regulatory agency. Interestingly, several widely used drugs, including erythropoietin for anemia and imatinib for CML, were approved as orphan drugs. In the United States, 282 therapies have been approved as orphan drugs in the last 24 years [40].

### 5.1 Selecting Targets for Drug Discovery

In this review, four criteria were proposed to evaluate new targets: linkage to the disease, potential therapeutic index, chemical tractability, and economics. The following questions are proposed as an additional guide to the application of these criteria for new targets:

- How will the new therapy be differentiated from any existing therapies?
- If the target was identified by a monogenic disease, will the new therapy only be applicable for patients with a sporadic form of the disease?
- If the target was identified using genetics in a model organism, how will the corresponding human target(s) be identified?
- How much target perturbation is required for efficacy?
- What are the potential mechanism-based toxicities?
- Would this therapeutic index and the likely mechanism-based toxicities be acceptable for the target indication?
- Is the target part of a protein family for which therapeutics have been previously identified?
- Does the market size justify the development costs?
- At what point in clinical testing will the key clinical issues be addressed?
- Would the therapeutic be eligible for an accelerated approval or orphan indication?

Even the best targets have issues. The ability to identify these issues, develop a plan to address these issues, and apply judgment as to the likely outcome of these issues will have a significant impact on the success of any drug discovery and development effort. Viewed from this perspective, target selection is a process not an event. In particular, as internal and external data become available, targets should be reevaluated with respect to these criteria to determine the appropriateness of continued work.

## ACKNOWLEDGMENTS

I would like to thank Richard Olson, Jeremy Toyn, and Donna Barten for helpful discussions and comments on the manuscript.

## REFERENCES

1. Zambrowicz BP, Sands AT (2003). Knockouts model the 100 best-selling drugs–Will they model the next 100? *Nat Rev Drug Discov 2*, 38–51.
2. (FDA) (2004). *Challenge and opportunity on the critical path to new medical products*. Washington, DC: FDA.
3. Dimasi JA (2001). Risks in new drug development: Approval success rates for investigational drugs. *Clin Pharmacol Ther 69*, 297–307.
4. Kola I, Landis J (2004). Can the pharmaceutical industry reduce attrition rates? *Nat Rev Drug Discov 3*, 711–715.
5. Lindsay MA (2003). Target discovery. *Nat Rev Drug Discov 2*, 831–838.
6. Kramer R, Cohen D (2004). Functional genomics to new drug targets. *Nat Rev Drug Discov 3*, 965–972.
7. Hardy LW, Peet NP (2004). The multiple orthogonal tools approach to define molecular causation in the validation of druggable targets. *Drug Discov Today 9*, 117–126.
8. Olesen J, Diener HC, Husstedt IW, Goadsby PJ, Hall D, Meier U, Pollentier S, Lesko LM (2004). Calcitonin gene-related peptide receptor antagonist BIBN 4096 BS for the acute treatment of migraine. *N Engl J Med 350*, 1104–1110.
9. Citri A, Yarden Y (2006). Egf-ErbB signalling: Towards the systems level. *Nat Rev Mol Cell Biol 7*, 505–516.
10. Druker BJ (2002). Perspectives on the development of a molecularly targeted agent. *Cancer Cell 1*, 31–36.
11. Antonarakis SE, Beckmann JS (2006). Mendelian disorders deserve more attention. *Nat Rev Genet 7*, 277–282.
12. Lambert G, Krempf M, Costet P (2006). PCSK9: A promising therapeutic target for dyslipidemias? *Trends Endocrinol Metab 17*, 79–81.
13. Corder EH, Saunders AM, Strittmatter WJ, Schmechel DE, Gaskell PC, Small GW, Roses AD, Haines JL, Pericak-Vance MA (1993). Gene dose of apolipoprotein E type 4

allele and the risk of Alzheimer's disease in late onset families. *Science 261*, 921–923.

14. Corder EH, Saunders AM, Risch NJ, Strittmatter WJ, Schmechel DE, Gaskell PC Jr, Rimmler JB, Locke PA, Conneally PM, Schmader KE, et al. (1994). Protective effect of apolipoprotein E type 2 allele for late onset Alzheimer disease. *Nat Genet 7*, 180–184.
15. Mahley RW, Weisgraber KH, Huang Y (2006). Apolipoprotein E4: A causative factor and therapeutic target in neuropathology, including Alzheimer's disease. *Proc Natl Acad Sci USA 103*, 5644–5651.
16. Roses AD, Burns DK, Chissoe S, Middleton L, St Jean P (2005). Disease-specific target selection: A critical first step down the right road. *Drug Discov Today 10*, 177–189.
17. International HapMap Consortium. (2005). A haplotype map of the human genome. *Nature 437*, 1299–1320.
18. Brown MS, Goldstein JL (1981). Lowering plasma cholesterol by raising LDL receptors. *N Engl J Med 305*, 515–517.
19. Selkoe DJ (2001). Alzheimer's disease: Genes, proteins, and therapy. *Physiol Rev 81*, 741–766.
20. Siemers ER, Dean RA, Demattos R, May PC (2006). New pathways in drug discovery for Alzheimer's disease. *Curr Neurol Neurosci Rep 6*, 372–378.
21. Drews J (2000). Drug discovery: A historical perspective. *Science 287*, 1960–1964.
22. Lebovitz HE (1993). Insulin-mimetic and insulin-sensitizing drugs. *Diabetes Res Clin Pract 20*, 89–91.
23. Lehmann JM, Moore LB, Smith-Oliver TA, Wilkison WO, Willson TM, Kliewer SA (1995). An antidiabetic thiazolidinedione is a high affinity ligand for peroxisome proliferator-activated receptor gamma (PPAR gamma). *J Biol Chem 270*, 12953–12956.
24. Vogt A, Lazo JS (2005). Chemical complementation: A definitive phenotypic strategy for identifying small molecule inhibitors of elusive cellular targets. *Pharmacol Ther 107*, 212–221.
25. Butcher EC (2005). Can cell systems biology rescue drug discovery? *Nat Rev Drug Discov 4*, 461–467.
26. Austen M, Dohrmann C (2005). Phenotype-first screening for the identification of novel drug targets. *Drug Discov Today 10*, 275–282.
27. Matthews DJ, Kopczynski J (2001). Using model-system genetics for drug-based target discovery. *Drug Discov Today 6*, 141–149.
28. Drews J, Ryser S (1997). The role of innovation in drug development. *Nat Biotechnol 15*, 1318–1319.
29. Hopkins AL, Groom CR (2002). The druggable genome. *Nat Rev Drug Discov 1*, 727–730.
30. Lipinski CA, Lombardo F, Dominy BW, Feeney PJ (2001). Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv Drug Deliv Rev 46*, 3–26.
31. Zheng CJ, Han LY, Yap CW, Ji ZL, Cao ZW, Chen YZ (2006). Therapeutic targets: Progress of their exploration and investigation of their characteristics. *Pharmacol Rev 58*, 259–279.
32. Chen X, Ji ZL, Chen YZ (2002). TTD: Therapeutic target database. *Nucleic Acids Res 30*, 412–415.
33. Apweiler R, Attwood TK, Bairoch A, Bateman A, Birney E, Biswas M, Bucher P, Cerutti L, Corpet F, Croning MD, Durbin R, Falquet L, Fleischmann W, Gouzy J, Hermjakob H, Hulo N, Jonassen I, Kahn D, Kanapin A, Karavidopoulou Y, Lopez R, Marx B, Mulder NJ, Oinn TM, Pagni M, Servant F, Sigrist CJ, Zdobnov EM (2001). The InterPro database, an integrated documentation resource for protein families, domains and functional sites. *Nucleic Acids Res 29*, 37–40.
34. Bateman A, Birney E, Cerruti L, Durbin R, Etwiller L, Eddy SR, Griffiths-Jones S, Howe KL, Marshall M, Sonnhammer EL (2002). The Pfam protein families database. *Nucleic Acids Res 30*, 276–280.
35. Wishart DS, Knox C, Guo AC, Shrivastava S, Hassanali M, Stothard P, Chang Z, Woolsey J (2006). Drugbank: A comprehensive resource for in silico drug discovery and exploration. *Nucleic Acids Res 34*, D668–672.
36. Rao BM, Lauffenburger DA, Wittrup KD (2005). Integrating cell-level kinetic modeling into the design of engineered protein therapeutics. *Nat Biotechnol 23*, 191–194.
37. Maggini M, Vanacore N, Raschetti R (2006). Cholinesterase inhibitors: Drugs looking for a disease? *PLoS Med 3*, e140.
38. Dagher R, Johnson J, Williams G, Keegan P, Pazdur R (2004). Accelerated approval of oncology products: A decade of experience. *J Natl Cancer Inst 96*, 1500–1509.
39. Wastfelt M, Fadeel B, Henter JI (2006). A journey of hope: Lessons learned from studies on rare diseases and orphan drugs. *J Intern Med 260*, 1–10.
40. Haffner ME (2006). Adopting orphan drugs—Two dozen years of treating rare diseases. *N Engl J Med 354*, 445–447.

# NEW DRUG APPLICATIONS

John S. McInnes
*Center for Medicare, CMS, Baltimore, Maryland**

## 1 INTRODUCTION TO THE U.S. DRUG APPROVAL PROCESS

A new drug must be approved by the U.S. Food and Drug Administration (FDA) prior to marketing in the United States. "New drug" is defined by Federal Food, Drug, and Cosmetic Act (FDCA) § 201(p) as a human drug that is "not generally recognized . . . as safe and effective for use under the conditions prescribed, recommended, or suggested in the labeling" or one that, although recognized as safe and effective, has not been "used to a material extent or for a material time." The Supreme Court has concluded that a drug must have "adequate and well-controlled investigations"† of the drug's effectiveness to not be considered a new drug [1]. Because almost all drugs (unless already FDA approved) do not meet these criteria, they must go through the FDA drug approval process.

There are different approval processes for the different types of products that are regulated by the FDA. This chapter focuses on the application requirements for new human drugs, both through the new drug approval (innovator drugs) and abbreviated new drug approval (generic drugs) routes. The statutory provisions governing these requirements are found in the FDCA. Other drug application processes that will not be discussed include the over-the-counter review process, new animal drug applications, abbreviated new animal drug applications, and biologics license applications. The new drug application (NDA) process begins with evidence development to satisfy the conditions for approval. Clinical studies that are part of that evidence, if performed in the United States (and optionally if performed outside of the United States), require submission of an Investigational New Drug (IND) application.

## 2 INVESTIGATIONAL NEW DRUG APPLICATIONS

Prior to human clinical trials, preclinical investigation of a drug is performed in the laboratory and in animals to determine the drug's potential effects and to determine if the drug is safe for human testing. This testing typically includes both in vitro (in the laboratory outside of the body) and in vivo (in the body of a living organism) studies evaluating a drug's toxicity and pharmacology. FDA notification is not required prior to preclinical testing, but these investigations are subject to FDA regulations known as good laboratory practices. (See 21 CFR Part 58.)*

Commencement of human clinical trials in the United States require the submission of an IND application.

*This chapter was prepared by John McInnes in his personal capacity. The opinions expressed in this chapter are the author's own and do not reflect the view of the Centers for Medicare & Medicaid Services, the Department of Health and Human Services, or the United States government.

†Such investigations are also required for approval of a new drug application [FDCA § 505(d)].

*These regulations apply to the laboratory work and facilities that perform nonclinical studies intended to support a marketing application for FDA-regulated products, including drugs and biologics.

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

(See 21 CFR Part 312.) The ultimate goal of most human clinical trials is to gather sufficient information about the efficacy and safety of a drug to support an NDA. This is the beginning of the FDA's role in the clinical development of the new drug. The IND provides an exemption from the requirement that a drug be the subject of an approved marketing application prior to shipment of the drug across state lines. Once an IND is submitted, the FDA has 30 days to object to the IND or it automatically becomes effective and clinical trials may begin. (See 21 CFR Part 312.40). However, if the FDA detects a problem with the IND, it may place a clinical hold on the IND, prohibiting the start of the clinical studies until the problem is resolved. Clinical holds are discussed below.

## 2.1 Types of INDs

INDs can be categorized as either *commercial*, meaning that they are filed by companies intent on obtaining marketing approval for a new drug [2] or *noncommercial*, meaning that they are generally for research purposes or to make an unapproved drug available in certain situations and will not support an application for marketing approval. Types of INDs other than the standard commercial IND are described below.

***Research or Investigator INDs*** Research (noncommercial) INDs are submitted by a physician who both initiates and conducts an investigation. The investigator might want either to study an unapproved drug or to study an approved product for a new indication or in a new patient population. The investigator assumes all responsibilities of the sponsor under an investigator IND (21 CFR Part 312.3). In most cases the purpose of a research IND is not to create data to support marketing approval of either a new indication or a new product. Instead, the physician-researcher may want to publish the study or develop data to support use of the drug in a particular protocol.

***Emergency Use INDs*** *Emergency use* INDs are sometimes called either compassionate use or single-patient INDs. These INDs are for emergency use of an unapproved drug when the clinical situation does not allow sufficient time to submit an IND in accordance with 21 CFR §§ 312.23, 312.24. These are most commonly used for life-threatening conditions for which there is no standard treatment. Other situations in which an emergency use IND can be used are for patients who do not meet existing study criteria or when an approved study protocol does not exist.

***Treatment INDs*** *Treatment* INDs make widely available promising drugs for serious or immediately life-threatening conditions during the completion of clinical trials and during FDA review prior to approval. FDA provided examples of "serious" conditions in the IND final rule:

> [The] "serious disease or condition" requirement refers to conditions that have an important effect on functioning (e.g., stroke, schizophrenia, rheumatoid arthritis, osteoarthritis) or on other aspects of quality of life (e.g., chronic depression, seizures). Alzheimer's dementia, Amyotrophic Lateral Sclerosis (ALS), and narcolepsy are specific examples of serious conditions for which FDA has granted expanded access to investigational drugs in the past. [3]

***Screening INDs*** *Screening* INDs allow a sponsor to test several closely related chemical entities in initial clinical trials. FDA defines a screening IND as "[a] single IND submitted for the sole purpose of comparing the properties of closely related active moieties to screen for the preferred compounds or formulations. These compounds or formulations can then become the subject of additional clinical development, each under a separate IND" [4, p. 2]. Section 6030.4 of MaPPs also describes the properties of screening INDs as follows [4, p. 3]:

> Screening INDs are appropriate when single-dose or short-term, repeat-dose clinical trials (3 days of dosing) are proposed using multiple, closely related compounds. The compounds include different salts or esters and active moieties that are slightly different chemically, but appear to be similar in pharmacodynamic properties. The proposed studies could be a single trial with multiple compounds or similar trials involving only one compound (e.g., several PK studies, one for each compound). The number of compounds tested should usually be 5. Normally, the intent of the study is to compare the properties of the closely related active moieties to screen for further development.

## 2.2 IND Application Content

An IND depends upon a successful preclinical development program demonstrating that the drug has pharmacologic activity that justifies further development and is safe enough to commence human clinical trials. The content of the application may be divided into the following categories.*

1. *Preclinical Testing* Animal pharmacology and toxicology studies comprise the essential preclinical data to permit an assessment as to whether the product is reasonably safe for initial testing in

*The content requirements for an IND are discussed at 21 CFR § 312.23.

humans. Also included are any previous experience with the drug in humans (often foreign use).

2. *Manufacturing Information* Manufacturing information pertains to the composition, manufacturer, and stability of, and the controls used for, manufacturing the drug substance and the drug product. This information is assessed to ensure that the company can adequately produce and supply consistent batches of the drug.
3. *Investigator Information* The IND must include information on the qualifications of clinical investigators, that is, the professionals (generally physicians) who oversee the administration of the experimental drug to the study subjects, so that FDA can assess whether the investigators are qualified to fulfill their clinical trial duties.
4. *Clinical Trial Protocols* The clinical trial protocols are the centerpiece of the IND. FDA evaluates the detailed protocols for proposed clinical studies to assess whether the initial-phase trials will expose subjects to unnecessary risks. For later stage studies, FDA also evaluates whether the trials will, if successful, provide evidence for approval of the drug.
5. *Other Commitments* Finally, commitments to obtain informed consent from the research subjects, to obtain review of the study by an institutional review board (IRB), and to adhere to the investigational new drug regulations.

### 2.3 Clinical Trial Phases

During all phases of investigation, both the sponsor and investigators have responsibilities to ensure patient safety and data integrity. Clinical investigations are usually classified as one of four temporal phases. In practice, these phases frequently overlap.

***Phase I Studies*** Phase I studies are designed to evaluate the safety and the side effects of the drug under study and are the initial introduction of an investigational new drug into humans [5]. Phase I studies are usually conducted in healthy volunteers. Their purpose is to determine the metabolic and pharmacologic actions of the drug in humans, the side effects associated with increasing doses, and, sometimes, to gain early evidence on effectiveness. The data obtained in phase I studies about the investigational drug's pharmacokinetics and pharmacodynamics is used in the design of well-controlled, scientifically valid, phase II studies.

Additional questions pursued during phase I studies include drug metabolism, structure–activity relationships, and the mechanism of action of the drug in humans. The total number of subjects included in phase I studies varies with the drug but is generally in the range of 20–80. During phase I studies, Center for Drug Evaluation and Research (CDER) can impose a clinical hold (i.e., prohibit the study from proceeding or stop a trial that has started) for reasons of safety or because of a sponsor's failure to accurately disclose the risk of study to investigators.

***Phase II Studies*** Phase II is the beginning of evaluations of therapeutic efficacy [21 CFR § 312.21(b)]. Phase II studies are designed to evaluate efficacy and the dose range for the drug, including early controlled clinical studies conducted to obtain some preliminary data on the effectiveness of the drug for a particular indication or indications in patients with the disease or condition. In addition, more information is obtained on short-term side effects and risks associated with the drug. Phase II studies are typically controlled, closely monitored, and conducted in a relatively small number of patients, usually involving several hundred people. An important purpose of phase II trials is to determine the regimen and drug dose or doses for phase III trials.

***Phase III Studies*** Phase III studies are expanded controlled and uncontrolled trials relative to phase II. These trials build upon initial evidence of effectiveness of the drug obtained in phase II and are intended to gather the additional information about effectiveness and safety that is needed to evaluate the overall risk–benefit relationship of the drug. Phase III studies also provide an adequate basis for extrapolating the results to the general population and including that information in the physician labeling. Phase III studies commonly include several hundred to several thousand people.

In both phases II and III, CDER can impose a clinical hold if a study is unsafe (as in phase I) or if the protocol is clearly deficient in design in meeting its stated objectives. Great care is taken to ensure that this determination is not made in isolation but reflects current scientific knowledge, agency experience with the design of clinical trials, and experience with the class of drugs under investigation.

***Phase IV Studies*** Phase IV studies are done postapproval. Although FDA can seek phase IV commitments from a sponsor prior to or after approval, it has until recently been able to require postapproval studies only for drugs approved under the accelerated approval program, which is referred to as the Subpart H Program because the governing regulations are located at 21 CFR Part 314 subpart H.

More recently, the FDA Amendment Acts of 2007 [6] provide the FDA with new authority to require

postapproval studies or clinical trials, as well to require labeling changes for approved drug products. The FDA will now have the authority to require a postapproval study or clinical trial of a drug to assess a known serious risk, or signals of a serious risk, and to identify an unexpected serious risk. Studies or trials are only to be required if the Secretary of Health and Human Services (HHS) determines that current reporting and the postmarket risk identification analysis system are not sufficient to address the concern. For each required study or trial, the Secretary will require a timetable for its completion and periodic reports on status. A responsible person may appeal a requirement under FDA's normal dispute resolution procedures.

### 2.4 Institutional Review Boards

One of the most significant clinical trial subject protections is the Institutional Review Board (IRB), which is required for every clinical trial to ensure the rights and welfare of the trial subjects both before and during their trial participation. IRBs at hospitals and research institutions throughout the United States make sure that participants are fully informed of the risks associated with the experimental treatment and have given their written informed consent prior to commencement of the study.

An IRB must be composed of no less than five experts and laypeople with varying backgrounds to ensure a complete and adequate review of activities commonly conducted by research institutions. In addition to possessing the professional competence needed to review specific activities, an IRB must be able to ascertain the acceptability of applications and proposals in terms of institutional commitments and regulations, applicable law, standards of professional conduct and practice, and community attitudes. Therefore, IRBs must be composed of people whose concerns are in relevant areas. IRBs are monitored by the FDA, and the regulations governing IRBs are at 21 CFR Part 56.

### 2.5 Clinical Holds

FDA may halt or forestall a clinical investigation through a "clinical hold," which FDA uses when it does not believe, or cannot confirm, that the study can be conducted without unreasonable risk to the subjects [7]. If such a determination is made, FDA will contact the sponsor to order it not to start, or to stop, the clinical trial. Clinical holds may either delay the start of an early-phase trial on the basis of information submitted in the IND or stop an ongoing study based on a review of newly submitted clinical protocols, safety reports, protocol amendments, or other information. When a clinical hold is issued, a sponsor must address the issue that is the basis of the hold before the order is removed.

## 3 FULL NEW DRUG APPLICATIONS

Section 505 of the FDCA describes three types of new drug applications: (1) an application that contains full reports of investigations of safety and effectiveness under § 505(b)(1), (2) an application that contains full reports on safety and effectiveness but where at least some of the information required for approval comes from studies not conducted by or for the applicant and for which the applicant has not obtained a right of reference under § 505(b)(2), and (3) an application that contains information to show that the proposed product is identical in active ingredient, dosage form, strength, route of administration, labeling, quality, performance characteristics, and intended use, among other requirements, to a previously approved product under § 505(j).

Right of reference or use [the absence of which may make an NDA a 505(b)(2) application] is defined as "the authority to rely upon, and otherwise use, an investigation for the purpose of obtaining approval of an application, including the ability to make available the underlying raw data from the investigation for FDA audit, if necessary" [21 CFR 314.3(b)]. Sponsors have the right of reference to any studies: (1) they conduct, (2) that are conducted for them, or (3) for which they formally obtain a documented right of reference. An applicant is not considered to have a right of reference to published studies because the applicant does not have access to the raw data. However, if the raw data are in the public domain, a right of reference is unnecessary [8].

The subject of this section is the first type of application that contains full reports of investigations of safety and effectiveness, which are known as full NDAs. The two other types of applications, 505(b)(2) applications and 505(j) applications, also known as abbreviated new drug applications (ANDAs), are discussed in subsequent sections of this chapter.

### 3.1 Process for Full NDAs

***Background—Application User Fees*** In 1992, Congress enacted the Prescription Drug User Fee Act of 1992 (PDUFA). One significant purpose of PDUFA was to decrease review times for new drugs through new FDA staffing and other infrastructure funded by industry-paid user fees. This legislation was reauthorized in 1997 (PDUFA II) as a part of the Food and Drug Administration Modernization Act (FDAMA) and again in 2002 (PDUFA III).

PDUFA was most recently reauthorized in 2007 as title I of the Food and Drug Administration Amendments Act of 2007 (FDAAA). User fees collected under PDUFA IV are to be dedicated to expediting drug development and human drug application review, including postmarket drug safety activities.

PDUFA IV's scope is greater than the original PDUFA mandate of providing the FDA with additional resources for the review of new drug applications. PDUFA IV expands the FDA's use of user fee funding to include collecting, developing, and reviewing safety information (including adverse-event reports), developing and using adverse-event data collection systems, developing and using improved analytical tools to assess potential safety problems, and complying with other requirements related to postapproval studies, clinical trials, and labeling changes. PDUFA IV raises fee revenue amounts to $392.8 million for fiscal year 2008 and includes adjustment methods for fiscal years 2009–2012. It also provides for an additional $225 million in fees for fiscal years 2008–2012 for postmarket drug safety activities.

***Time Constraints on FDA Review*** The FDCA states that FDA must either approve an NDA or issue a notice opportunity for hearing on a proposed denial of approval within 180 days after the "filing" of the application [FDCA § 505(e)]. However, this 180-day limit does not mean that final approval is granted by the FDA within 180 days of submission of the NDA. Although the PDUFA legislation, discussed above, has generally decreased review times for new drugs, the process from start to completion generally takes longer.

There are two overlapping 180-day time periods that govern the NDA review process. The first of these time periods, sometimes referred to as the "review clock," begins at the date of submission of the application and ends 180 days later (assuming that the application is not determined to be ineligible for filing within the first 60 days after submission) [21 CFR § 314.100(a)]. At the end of the 180 days FDA must send the applicant an "action letter." The types of action letters are described in the next section.

The other 180-day time period is sometimes referred to as the "filing clock" [21 CFR § 314.101(a)(2)]. This time period begins 60 days after FDA's receipt of the application unless the agency determines that the application is facially deficient and should not be filed. Such a determination is called a refuse-to-file (RTF) action. The agency must either approve the application or issue a notice of an opportunity for hearing by the expiration of the 180-day filing clock. The 60 days for the filing decision plus the 180 days gives FDA a total of 240 days to process an NDA. This time limit is frequently extended by the applicant filing supplements to the NDA, but even properly extended filing clock time limits can be challenging for FDA to meet.

***FDA Decisions on the Application*** The FDA's official decision on an NDA is termed an "action letter." There are three types of action letters: (1) an approval letter, (2) an approvable letter, or (3) a not approvable letter.* An approval letter indicates final positive action on the NDA, although it may be issued on the basis of draft labeling (21 CFR § 314.105).

An approvable letter is provided for an application that substantially meets the requirements for approval but has relatively minor deficiencies (23 CFR § 314.110). Such minor deficiencies typically relate to the final labeling. An applicant must respond to the approvable letter within 10 days, otherwise the application will be considered withdrawn. The applicant's response can take one of several forms [21 CFR § 314.110(a)]:

- File a resubmission (or amendment) or a notice of intent to file a resubmission to satisfy the remaining requirements for approval (the FDA has 45 days to review an amendment) [21 CFR § 314.110(a)(1)].
- Withdraw the application.
- Request a hearing on whether grounds exist for denying approval when conditions for marketing in the approvable letter are unacceptable.
- Agree to a review period extension to provide more time for the applicant to decide which of the above options to pursue.

An applicant also has 10 days to inform FDA of its intentions when it receives a not approvable letter. Options for the applicant [21 CFR § 314.120(a)(1)–(3)] include:

- Amend the NDA.
- Withdraw the application without prejudice to later refilling.
- Request an opportunity for a hearing on the basis of the agency's failure to approve the application (the agency must issue a notice of an opportunity for a hearing within 60 days of the date of the not approvable letter) [21 CFR § 314.120(a)(3)].

*FDA issued a proposed rule that would amend the regulations on marketing approval of new drugs to discontinue the use of approvable and not approvable letters when taking action on a marketing application and instead use "complete response" letters (69 *Fed Reg* 43357, July 20, 2004). The agency has yet to issue the final rule.

- Agree to a review period extension to provide more time for the applicant to decide which of the above options to pursue.

### 3.2 Application Content

The contents of an NDA are both detailed and extensive, with application lengths running into the tens of thousands of pages. A full NDA contains the following information.

***Safety and Effectiveness Data*** Safety and effectiveness data are the heart of the NDA. Safety data required for the submission includes the following:

1. Pharmacologic properties relating to potential adverse reactions [21 CFR § 314.50(d)(2)(i)]
2. Animal toxicology studies [21 CFR § 314.50(d)(2)(ii)]
3. Studies related to a drug's particular mode of action or conditions of use [21 CFR § 314.50(d)(2)(ii)]
4. Reproduction and teratology studies [21 CFR § 314.50(d)(2)(iii)]
5. Studies on drug absorption, distribution, metabolism, and excretion in animals [21 CFR § 314.50(d)(2)(iv)]
6. Human safety data, including:
   a. Pharmacokinetic and bioavailability data [21 CFR § 314.50(d)(3)]
   b. Clinical pharmacology data [21 CFR § 314.50(d)(5)(i)]
   c. Other safety information such as adverse reactions and drug–drug interactions [21 CFR § 314.50(d)(5)(vi)(a)]

As mentioned earlier in this chapter, the effectiveness data required for approval must be obtained from "adequate and well-controlled . . . clinical investigations" [FDCA § 505(d)]. In the past, FDA interpreted the plural "studies" to mean that two studies were required for approval, but current law permits one study when FDA considers one study to be sufficient. The characteristics of an adequate and well-controlled study are set out in an FDA regulation (21 CFR § 314.126). Well controlled refers to the comparison group relative to the group taking the experimental drug. The control group can either take a placebo (sugar pill) or another drug that is an established treatment for the underlying disease (active control). In other circumstances, dose comparisons, no treatment controls, or, rarely, historical controls may be accepted. The protocol, statistical analysis plan, and an analysis of the data must be submitted with the NDA [21 CFR § 314.50(d)(5)(ii)]. Any additional information and data relevant to safety or effectiveness, including data from uncontrolled studies, must also be submitted as part of a complete NDA.

***Patent Information*** All NDAs are required to include patent information (number and expiration date) for every patent that claims either the drug covered by the NDA or a method using that drug—that could form the basis of an infringement claim if a generic copy of the drug were marketed [FDCA § 505(b)(1)]. Patents for methods of manufacturing the drug are, however, not to be submitted [21 CFR § 314.53(b)]. Other types of patents that should not be submitted in an NDA include: (1) drug packaging patents, unless part of the "finished dosage form" approved in the NDA [9]; (2) patents on metabolites of the drug; (3) patents on intermediates; and (4) method of use patents for methods not approved in the NDA [21 CFR § 314.53(b)(1)].

Patents covering use, the drug product, formulation, and composition must claim the specific drug described in the NDA [21 CFR § 314.53(b)(1)]. However, patent holders can list patents for alternative polymorphs* [10] provided that they certify that the following information exists for the polymorph at the time of submission of the patent information [21 CFR § 314.53(b)(2)(v)].†

- A full description of the polymorphic form
- The executed batch record of the polymorphic form and documentation that the batch was manufactured under cGMP (current good manufacturing practices) requirements
- Demonstration of bioequivalence between the polymorph and the form of the drug product described in the NDA
- A list of all components used in making the polymorph and other manufacturing information
- Comparative in vitro dissolution testing on 12 units each of the batch of drug containing the polymorph and the form of the drug covered by the NDA

The decision as to whether to perform this testing for a polymorph depends upon the perceived advantages of listing the patent for the polymorph (in addition to all of the other patents listed).

*Polymorphs are different physical forms of a pharmaceutical solid. Polymorphic forms usually refer to crystalline and amorphous forms as well as solvate and hydrate forms.

†These data must exist in the patent holders file but need not be submitted to FDA.

### *Risk Evaluation and Mitigation Strategies*

*New REMS Requirements under FDAAA* Section 901 of FDAAA provides a statutory framework for integrating risk evaluation and mitigation strategies (REMS) into NDA reviews and postmarket pharmacovigilance (FDCA § 505-1). These provisions are an evolution from PDUFA III, which provided funding for development of FDA risk management guidance documents and review of voluntary risk minimization plans. Many of the risk minimization tools described in FDAAA are already in use in existing drug approvals [i.e., under Subpart H/Risk Minimization Action Plans (RiskMAPs)].

Unlike earlier versions of the legislation, the final FDAAA provisions do not impose a REMS on *all* drugs and biologics prior to approval. Under the new law, the secretary (in consultation with the office responsible for reviewing the drug and the office responsible for the drug's postapproval safety) may determine that an REMS is necessary to ensure that the benefits of the drug outweigh its risks [FDCA § 505-1(a)(1)]. (In making decisions relating to REMS, the FDA decision makers must be at a level at or above that of an official authorized to approve the NDA.) [FDCA § 505-1(a)(2)(B)(4)]. FDA may also require an REMS for new drug and biologic license applications, drugs and biologics that have already been approved, and for supplemental applications seeking approval of a new indication [FDCA § 505-1(a)(2)(A)]. The holder of a covered application has 120 days to submit its proposed REMS from the time it is notified by FDA that an REMS is required [FDCA § 505-1(a)(2)(B)].

An REMS must include a timetable for the submissions of assessments of the REMS (assessments after 18 months, 3 years, and 7 years, although FDA may eliminate assessments after the 3-year period if the serious risks of the drug are being adequately managed) [FDCA § 505-1(d)]. Additional REMS elements may also be required by FDA: (1) a medication guide and/or patient package insert for distribution to patients [FDCA § 505-1(e)(2)] and (2) a communication plan to health care providers (e.g., sending letters, disseminating information to health care providers directly or through professional organizations) [FDCA § 505-1(e)(3)]. FDA may also require "such elements as are necessary" to assure a drug's safe use if, because of its "inherent toxicity or potential harmfulness," it is associated with a serious adverse drug experience and could be approved only if (or withdrawn unless) the elements are made part of the REMS [FDCA § 505-1(f)(1)].

Furthermore, these REMS elements must [FDCA § 505-1(f)(2)]:

- Be commensurate with the drug's risk.
- Be publicly posted, with an explanation by FDA, within 30 days of the date they are imposed.
- Conform with elements of other drugs with similar serious risks.
- Be designed to be consistent with established systems for distributing and dispensing drugs.

The REMS elements are not to be unduly burdensome on patient access, especially with respect to patients with serious or life-threatening diseases or conditions, or patients who have difficulty accessing health care. The elements must also include one or more goals to mitigate a specific risk listed in the drug's labeling and may require that [FDCA § 505-1(f)(3)]:

- Prescribers have certain training or certification.
- Settings that dispense the drug be certified.
- The drug be dispensed only to patients in certain settings (e.g., hospitals).
- The drug be dispensed with documentation of safe-use conditions (e.g., laboratory tests).
- The patient be subject to monitoring.
- Each patient using the drug be enrolled in a registry.

The elements may also include a system in which the applicant would be required to take reasonable steps to monitor, evaluate, and improve implementation of required elements by providers and other parties in the health care system responsible for implementing the elements [FDCA § 505-1(f)(4)]. FDAAA provides a waiver of these requirements for a qualified countermeasure in the case of a public health emergency [FDCA § 505-1(f)(7)]. Under the new law, the Drug Safety and Risk Management Advisory Committee will seek input from patients and providers about the elements, as well as evaluate the elements of at least one drug, to ensure access and minimize the burdens [FDCA § 505-1(f)(5)]. The committee will also use the input and evaluations to issue guidance and modify elements [FDCA § 505-1(f)(5)(B)].

*FDA Action on REMS* The FDA, after consultation with the office responsible for reviewing the drug and the office responsible for postapproval safety of the drug, must promptly review all proposed REMS and each assessment of an approved REMS [FDCA § 505-1(h)]. No later than 60 days from the submission of the REMS (or 30 days if FDA determines there may be cause for withdrawal of the drug), FDA must initiate

discussions with the responsible person [FDCA § 505-1(h)(2)]. FDA will either describe any required REMS as part of the action letter on the application or issue an order within 90 days of beginning discussions. Any action letter or order will be made publicly available [FDCA § 505-1(h)(3)].

When a concern about a serious risk of a drug may be related to that drug's pharmacological class, FDA may defer assessments of the approved REMS until having convened one or more public meetings to consider responses to the class-based concern [FDCA § 505-1(h)(7)]. FDA must provide notice of the deferral to the holder within 5 days, publish notice in the *Federal Register*, and publish notice of any public meetings [FDCA § 505-1(h)(7)(B)(ii), (iii)]. The meetings may include discussion with the responsible person, meetings of advisory committees, or workshops of scientific experts and other stakeholders [FDCA § 505-1(h)(7)(C)]. After considering the discussions in the meetings, FDA may announce a planned regulatory action in the *Federal Register*, seek public comment, and then issue an order [FDCA § 505-1(h)(7)(D)].

*Dispute Resolution for REMS* FDAAA also sets up REMS dispute resolution procedures [FDCA § 505-1(h)(4), (5)]. If the REMS was submitted at initial approval and there is a dispute about it, then major dispute resolution procedures previously established under PDUFA and FDA guidance would apply. For all other disputed REMS, the responsible person may request that the Drug Safety Oversight Board (DSOB) review the dispute [FDCA § 505-1(h)(5)(B), § 505-1(j)]. The DSOB is composed of federal employees who are scientists and health care practitioners and appointed by the Secretary and will meet at least once each month. Although a determination by FDA that an REMS is required is not subject to review, the particular elements of an REMS may still be reviewed. Under this procedure, FDA must publish notice that a dispute is to be reviewed by the DSOB. Entering into dispute resolution does not preclude further discussions or the use of FDA's administrative appeals, although the responsible person must either withdraw the request for review or terminate the use of administrative appeals. Both parties make a written or oral presentation to the DSOB. The meeting is to be recorded and then made public within 90 days. The DSOB may add members with relevant experience from FDA or other public health agencies for a meeting. No more than 5 days after the meeting, the DSOB will provide a written recommendation to FDA on how to resolve the dispute. The recommendation must also be made public. FDA must issue an action letter or order that resolves the dispute before the action deadline for the action letter on the application or 7 days after receiving the DSOB recommendation.

A drug will be considered misbranded if the responsible person does not comply with an REMS requirement [FDCA § 502(y), (z)]. The law establishes civil monetary penalties of $250,000 per violation, not to exceed $1 million [FDCA § 801(b)(4)]. If a violation continues after FDA provides written notice to the responsible person, a civil monetary penalty of $250,000 for the first 30-day period may be imposed, doubling for each subsequent 30-day period, not to exceed $1 million in a 30-day period and $10 million for all violations adjudicated in a single proceeding. FDAAA requires that the FDA Commissioner report to Congress on how best to communicate to the public the risks and benefits of new drugs. The Commissioner may consider the possibility of using a symbol indicating that the drug was recently approved in the labeling and in any direct-to-consumer (DTC) advertisements.

***Pediatric Assessments*** FDAAA reauthorized the Pediatric Research Equity Act (PREA), so it now will expire in 2012. Under the PREA, an NDA or supplement must be accompanied by a pediatric assessment [FDCA § 505B(2)]. If the assessment indicates that the treatment for the disease has a similar course in all adult and pediatric populations, FDA may conclude that data supporting pediatric effectiveness can be extrapolated from adequate and well-controlled studies in adults or from one pediatric age group to another.

In addition to a reauthorization of the existing requirements, FDAAA significantly expands the PREA's procedures for deferring pediatric assessments. In order for an applicant to obtain a deferral, the applicant must submit a timeline for completion of pediatric studies. Following a deferral, the applicant will each year provide FDA with information detailing the progress made in conducting the studies, and if no progress has been made, evidence and documentation will be provided that such studies will be conducted with due diligence at the earliest possible time. The information will be available to the public through FDA's website. FDA retains the power to grant a full or partial waiver from the PREA assessment for good cause.

The FDAAA also formalizes dispute resolution procedures followed when FDA requests a labeling change to incorporate pediatric information [FDCA § 505B(g)(1)]. If the sponsor disputes the proposed change and the FDA Commissioner and sponsor are unable to reach an agreement on the appropriate changes to the labeling for the drug, then not later than 180 days after the date of the submission, the Commissioner will request that the sponsor make any labeling change that the Commissioner determines to be appropriate. If

the sponsor does not agree within 30 days to make the requested change, the Commissioner will refer the matter to the Pediatric Advisory Committee. Within 90 days of receiving the referral, the Pediatric Advisory Committee will review the pediatric study reports and make a recommendation to the Commissioner. Within 30 days of receiving the Pediatric Advisory Committee's report, the Commissioner will make a renewed request for a labeling change. If the sponsor does not agree to make a labeling change, the commissioner may deem the drug to be misbranded.

Not later than 210 days after the date of submission of a pediatric assessment, FDA must make available medical, statistical, and clinical pharmacology reviews on FDA's website [FDCA § 505B(h)(1)].

Sponsors of assessments that result in labeling changes are required to distribute such information to physicians and health care providers. In addition, sponsors must submit all adverse-event reports connected to the labeling change to the Office of Pediatric Therapeutics. FDA will also make public an annual report of all such labeling changes [FDCA § 505B(f)(6)(H)].

***Other Information*** In addition to the information described above, the NDA must contain the following additional information for the drug under review:

- A full list of articles used as components
- A full statement of the composition
- A full description of the methods used in, and the facilities and controls used for, the manufacture, processing, and packing of the drug
- Samples of the drug and articles used as components, as FDA may require
- Proposed labeling

An NDA must contain a debarment certification that states that the applicant did not and will not use in any capacity the services of a person who has been debarred from such activities in connection with that application [FDCA § 306(k)(1)]. It must also contain a certification of compliance with the requirement to submit information concerning clinical trials to the director of the National Institutes of Health (NIH) for publication in the NIH clinical trial registry [FDCA § 505(b)(6)].

## 4 SECTION 505(B)(2) APPLICATIONS

A 505(b)(2) application is an NDA described in § 505(b)(2) of the FDCA. FDA guidance states that a 505(b)(2) NDA can rely on data not developed by the applicant. For example, a 505(b)(2) application can rely on FDA's finding of safety or effectiveness for another applicant's approved drug product or it can rely on published literature. The FDA's interpretation of the statute in its guidance has been controversial and has been challenged at the agency level and in court by industry—thus far unsuccessfully.

The fundamental difference between a full NDA and a 505(b)(2) application is that unlike a full NDA, a 505(b)(2) relies for approval on one or more investigations that "were not conducted by or for the applicant and for which the applicant has not obtained a right of reference or use from the person by or for whom the investigations were conducted" [FDCA § 505(b)(2)].

A 505(b)(2) application, under the statute, must contain a patent certification as to any patents listed in the NDA for the drug for which the studies relied on were performed. That certification must state that the patents have expired, that the applicant is willing to have approval delayed until they do expire, or that the patents are invalid, unenforceable, or not infringed. The applicant must notify the patent holder and the applicant for the listed drug of any claim that a patent is invalid or not infringed [FDCA § 505(b)(2)]. As discussed below with respect to ANDAs, this notification can lead to a patent infringement lawsuit, which may in turn result in a delay of up to 30 months in approval of the 505(b)(2) application.

505(b)(2) applications may not be submitted for duplicates of approved products that are eligible for approval under 505(j) [see also 21 CFR § 314.101(d)(9)]. [505(b)(2) applications may in some cases be submitted for drugs that are eligible for consideration pursuant to a suitability petition under FDCA § 505(j)(2)(C). For further explanation of suitability petitions, refer to the next section of this chapter.]

## 5 ABBREVIATED NEW DRUG APPLICATIONS (ANDAS)

Abbreviated new drug applications (ANDAs) are NDAs for generic copies of FDA-approved innovator products that are not required to contain full safety and effectiveness data, hence the term "abbreviated" [FDCA § 505(j)]. FDCA § 505(j) governs ANDAs.

### 5.1 Eligibility for ANDA Consideration

An ANDA may be filed for a generic copy of any "listed drug," that is, any drug listed in FDA's *Approved Drug Products with Therapeutic Equivalence Evaluations*

(also referred to as the Orange Book).* Listed drugs are drugs for which a new drug application has been approved for safety and effectiveness [FDCA § 505(j)(7)]. Listed drugs include those approved under an ANDA but do not include drugs that were approved under NDAs that were approved solely on safety grounds prior to 1962 and whose effectiveness has not been affirmed in the Drug Efficacy Study Implementation (DESI) process [11].

The information for each listed drug includes [FDCA § 505(j)(7)(A)]*:

- The official name
- The proprietary name
- The date of approval if the drug is approved after 1981
- The number assigned to the listed drug's NDA
- Patent information

## 5.2 Contents of an ANDA and ANDA Process

### *Equivalence or Similarity to Listed Drug*

*Bioequivalence* The essential scientific determination that must be proven in the ANDA is that the generic drug product is "bioequivalent" to the listed product [FDCA § 505(j)(2)(A)(iv)], which means that the application must contain data that show that the generic drug that is the subject of the application will produce an effect in patients equivalent to that produced by the listed drug upon which the ANDA relies [see 21 CFR § 320.1(e)]. The only exception to showing bioequivalence is in certain cases in which the ANDA is submitted pursuant to a suitability petition, which are discussed in greater detail below. For such drugs, instead of having to show bioequivalence, the statute requires that: (1) the active ingredients of the ANDA drug are "of the same pharmacological and therapeutic class as those of the listed drug"; and (2) the ANDA drug is expected to have the same therapeutic effect as the listed drug [FDCA § 505(j)(2)(A)(iv)]. FDA regulations require a showing of bioequivalence for active ingredients that are the same in the ANDA drug as in the listed drug [21 CFR § 314.94(a)(7)(ii)(B) (2007)].

There are certain circumstances in which FDA may waive the requirement for data to demonstrate bioequivalence. These are generally when, as with a solution for injection or oral administration, bioequivalence is considered to be self-evident.

*Same Active Ingredient* For listed products that contain only one active ingredient, the ANDA must show that the active ingredient of the generic product is the same as that of the listed drug [FDCA § 505(j)(2)(A)(ii)(I)]. For products with more than one active ingredient, the ANDA must usually demonstrate that all of the active ingredients are the same as those of the listed drug. In limited circumstances in which a suitability petition, as discussed below, is filed for a combination product, the ANDA must show that either all of the active ingredients are the same as those of the listed drug or that FDA has approved an ANDA may be permitted to substitute one such ingredient [FDCA § 505(j)(2)(A)(ii)(I), (II)].

*Labeling* The ANDA labeling must show that the conditions of use for the ANDA drug have been approved for the listed drug [FDCA § 505(j)(2)(A)(i)]. There is also a requirement that the ANDA labeling be the same as that of the innovator [FDCA § 505(j)(2)(A)(v)]. The only differences permitted by the statute between the ANDA drug labeling and the listed drug labeling are changes required due to differences approved under a suitability petition and different information related to the identity of a manufacturer [FDCA § 505(j)(2)(A)(v)]. There may, however, be differences reflecting the fact that the ANDA may not, because of exclusivity or patent protection, be approved for all of the conditions of use of the innovator it copies.

***Dosage Form, Strength, Route of Administration—The Role of Suitability Petitions*** Suitability petitions request permission from the FDA to file an ANDA for a drug that differs in specified ways from the listed drug upon which the ANDA would be based [FDCA § 505(j)(2)(C)]. Deviations from the listed drug may only be (1) a different active ingredient in a combination product in which the other active ingredients match those of the listed drug; (2) different route of administration; (3) a different dosage form; or (4) a different strength.

There are restrictions on a different active ingredient that can be substituted for an ingredient in a combination product subject to a suitability petition. First, the substituted active ingredient must be of the "same pharmacologic or therapeutic class as those of the listed drug" [FDCA § 505(j)(2)(A)(iv)]. Second, the substituted active ingredient in a combination product must be either an active ingredient of a listed drug or an active ingredient of a drug that is not considered a "new

*FDA's *Approved Drug Products with Therapeutic Equivalence Evaluations* is called the Orange Book because of the color of the cover of the hard-copy published versions and the color of the welcome page for the Internet edition.

*The statute also requires that FDA publish, for each drug, a statement regarding the type of bioequivalence studies required for ANDAs that refer to the listed drug. FDA apparently considers its regulations on bioequivalence requirements fulfill this obligation and does not publish a separate statement for each approved NDA.

drug" by FDA [FDCA § 505(j)(2)(A)(ii)(III)]. FDA is to grant a suitability petition only if it concludes that clinical trials are not necessary to show the safety or effectiveness of the changed product [FDCA § 505(j)(2)(C)(i)]. FDA is to act on each petition within 90 days of its submission [FDCA § 505(j)(2)(C)].

***Patent and Exclusivity Considerations*** An ANDA must include a patent certification for each patent for the listed drug that the applicant seeks to copy [FDCA § 505(j)(2)(A)(iv)]. These certifications must be one of four types or, if no patents claim the drug, FDA requests that the ANDA applicant certify that to the best if its knowledge no patents are applicable [see 21 CFR § 314.94(a)(12)(ii)]. The four types of certifications are as follows:

*Patent Information Not Filed* This type of certification states that patent information has not been filed for the listed drug [FDCA § 505(j)(2)(A)(vii)(I)]. This requirement presumes that patents exist that claim the listed drug or use of the listed drug but have not been filed. FDA recommends, but the FDCA does not require, that applicants perform a search for unlisted patents.

*Patent Has Expired or Date of Expiration* The next two types of certifications are self-explanatory. The ANDA applicant certifies that the patent on the listed drug has expired or certifies the date on which the patent will expire [FDCA § 505(j)(2)(A)(vii)(II), (III)]. The patent expiration dates are published in the Orange Book, and an ANDA with this type of certification cannot be approved until the date of patent expiration [FDCA § 505(j)(5)(B)(i), (ii)].

*Patent is Invalid, Unenforceable, or Will Not Be Infringed* The fourth type of certification must state that the patent claim on the listed drug is invalid, unenforceable, or will not be infringed by the manufacture, use, or sale of the drug covered by the ANDA [FDCA § 505(j)(2)(A)(vii)(IV)]. This type of certification must also include a statement that the ANDA applicant will provide notice to the owner (or designated representative) of the patent and to the NDA owner (or designated representative) for the listed drug. This notice must be sent within 20 days of the postmark of the notice from the applicant to FDA advising that the ANDA has been filed [FDCA § 505(j)(2)(B)(ii)(I)]. Such notice to the owner of the pioneer NDA is required even if the ANDA applicant lists a generic version of the drug.

If the patent holder commences patent litigation within 45 days after the date of notice receipt by the patent owner and the NDA holder, then the approval date of the ANDA will be delayed until a decision of the trial court finding the patent invalid, unenforceable, or not infringed, or until 30 months after the date of the notice, whichever is earlier [FDCA § 505(j)(5)(B)(iii)]. The notice must include information detailed in 21 CFR § 314.95(c). Importantly, for a patent alleged not to be infringed, the notice must contain a theory of noninfringement [21 CFR § 314.95(c)(6)(i)].

***REMS Requirements for ANDAs*** A drug that is subject to an abbreviated new drug application (ANDA) is subject to the elements of a medication guide or patient package insert for the applicable listed drug and elements to assure safe use if required for the listed drug [FDCA § 505-1(i)]. FDA may waive the requirement that the drug subject to the ANDA use the same single, shared system as the listed drug if the burden of creating a shared system outweighs its benefits or if an aspect of the elements for the listed drug is claimed by a patent and the applicant for the ANDA certifies that it has sought a license for use of an aspect of the elements but was unable to obtain one.

## REFERENCES


1. Weinberger v. Hynson, Westcott & Dunning, Inc., 412 U.S. 609, 629–630 (1973).
2. Guidance for Industry, Submitting and Reviewing Complete Responses to Clinical Holds, October 2000, p. 5, Available: http://www.fda.gov/cder/guidance/3082fnl.pdf.
3. 71 *Fed Reg* 75147, 75151 (December 14, 2006).
4. CDER's Manual of Policies and Procedures (MaPP) 6030.4, INDs: Screening INDs, p. 2. Available: http://www.fda.gov/cder/mapp/6030-4.pdf.
5. 62 *Fed Reg* 66113, 66116 (December 17, 1997).
6. FDA Amendment Acts of 2007 (FDAAA) § 901.
7. 21 CFR § 312.42; MaPP 6030.1.
8. U.S. Food and Drug Administration. Draft Guidance for Industry, Applications Covered by Section 505(b)(2). Available: http://www.fda.gov/cder/guidance/2853dft.htm.
9. 68 Fed. Reg. 36676, 36680 (June 18, 2003).
10. Draft Guidance for Industry. ANDAs: Pharmaceutical Solid Polymorphism, Chemistry, Manufacturing, and Controls Information. Available: http://www.fda.gov/ohrms/dockets/98fr/2004d-0524-gdl0001.doc.
11. 57 *Fed Reg* 17956 (April 28, 1992).

# FROM CONCEPT TO THE CLINICS: DEVELOPMENT OF CANCER THERAPEUTICS

Balveen Kaur,[1] Gregory B. Lesinski,[2] and Abhik Ray Chaudhary[3]

[1]*Dardinger Laboratory for Neuro-oncology and Neurosciences, Department of Neurological Surgery, College of Medicine,*
[2]*Department of Molecular Virology, Immunology and Medical Genetics,*
[3]*Department of Pathology*
*The Ohio State University Medical Center, Columbus, Ohio*

## 1 INTRODUCTION

Development of a therapeutic agent is a long and laborious process commencing with the discovery of a "druggable target" underlying the pathology of a disease. This finding is then exploited to identify molecules/compounds that can activate/antagonize this target. Rigorous testing for the specificity and potency of the identified agents then leads to the identification of a lead molecule/compound. The pharmacological formulation of the compound is then optimized for efficient and effective drug delivery. Initial laboratory testing is subsequently followed by detailed animal studies confirming its in vivo absorption, distribution, metabolism, elimination, efficacy, and toxicology. Following these initial laboratory observations, the drug then undergoes thorough pharmacological testing to evaluate its pharmacokinetic and pharmacodynamic properties. During this time, it may return back to the laboratory for additional modifications to modulate its half-life, solubility, bioavailability and the like. The modified agent then goes through rigorous preclinical development in good manufacturing practice (GMP)-certified laboratories, followed by studies geared to demonstrate its bioequivalency and efficacy relative to the initial molecule. The candidate agent is then carefully tested for toxicity in appropriate animal models in multiple species to define a dose that can be safely administered to human patients without adverse effects. After the successful completion of all these preclinical studies, the developer of the experimental drug typically completes an IND (investigational new drug) application to be approved for testing in human patients. After obtaining IND approval, the drug is then tested for safety and efficacy in clinical trials. Phase I clinical trials are typically dose escalation studies meant to evaluate the safety, adverse events, and maximum tolerated dose in human patients. Agents that appear to be reasonably safe then progress to phase II and eventually phase III studies to evaluate their efficacy prior to government approval and marketing. In this chapter we will discuss three recently approved agents that have been through this process and are now being tested in combination with other therapies and one experimental agent that is actively engaged in the developmental process.

## 2 BEVACIZUMAB (AVASTIN)

Bevacizumab is a humanized antivascular endothelial growth factor (VEGF) monoclonal antibody that, in combination with chemotherapy, was the first anti-angiogenic agent approved by the Food and Drug Administration (FDA) as first-line treatment for metastatic colorectal cancer. In this section we will describe the discovery of VEGF, its function and role played

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

in tumor angiogenesis, followed by a discussion of the preclinical and clinical studies leading to the development of bevacizumab.

### 2.1 Tumor Angiogenesis

Angiogenesis/vasculogenesis is a highly regulated process controlling the generation of new blood vessels during development, wound healing, and ovulation in adults. This process commences with increased production of angiogenic factors and a concomitant reduction in angiostatic factors. The secreted angiogenic factors then bind to and activate specific receptors on endothelial cell receptors, initiating a signaling cascade leading to increased proliferation, migration, and invasion of endothelial and/or vascular progenitor cells. The process that transforms this nascent vasculature into mature vasculature is exquisitely regulated, turning "on" and "off" discrete sets of molecular events resulting in branching, remodeling, pruning and recruitment of perivascular supporting cells [1]. Dysregulation of this process manifests itself as new vessel development in several pathological states such as diabetic retinopathy, chronic inflammatory disorders, tumor growth, and metastasis. The development of a functional vascular network is considered a critical prerequisite for the progressive growth and metastasis of neoplasms. In fact, histological examination of autopsies from patients dying of unrelated events has revealed that most of us carry in situ tumors that will not develop into disease. These observations could be interpreted to imply that dormant microscopic tumors are kept in check by the body's ability to prevent the tumors from establishing a new blood supply essential to provide them with oxygen and nutrients [2–4].

The hypothesis that antiangiogenic strategies to attack the tumor's blood supply could be exploited as an effective anticancer therapy was first postulated by Folkman in 1971 [2]. Since then several angiogenic and antiangiogenic factors involved in tumoral angiogenesis have been described, and various strategies to counter the pathways they activate are currently being investigated as potential anticancer therapies.

### 2.2 Vascular Endothelial Growth Factor

Vascular endothelial growth factor is one member of a family of growth factors that includes VEGF-A, VEGF-B, VEGF-C, VEGF-D, VEGF-E, and placenta-like growth factor (PLGF) (Fig. 1). VEGF-A was the first growth factor found to have endothelial-cell-specific mitogenic properties. It was purified to homogeneity from media conditioned by bovine pituitary follicular cells as a growth factor with remarkable specificity for endothelial cells [5]. It had also been described as a vascular permeability factor (VPF), which facilitated

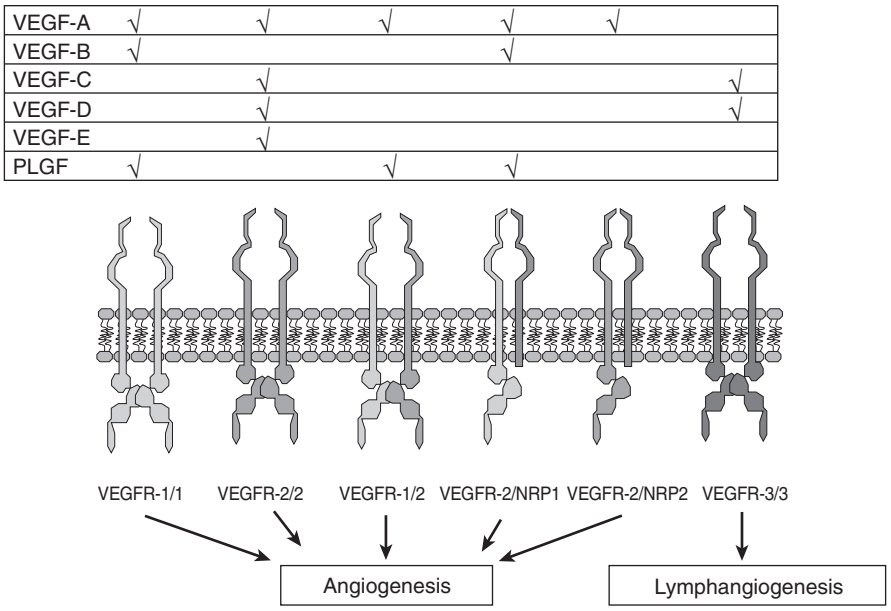


**Figure 1** VEGF family members (VEGF-A, VEGF-B, VEGF-C, VEGF-D, VEGF-E, and PLGF) bind differentially to the VEGF receptors VEGFR-1, VEGFR-2, VEGFR-3, NRP1, or NRP2, resulting in receptor dimerization and activation. The nonsignaling NRP1 and NRP2 function by hetrodimerizing with VEGFR-2 leading to its activation. Lymphangiogenesis is initiated by VEGF-C and VEGF-D binding to VEGFR-3, leading to its homodimerization and subsequent activation.

vascular leakage into the skin [6]. Subsequent isolation and cloning of VEGF and VPF revealed them to be identical [7, 8]. VEGF is a homodimeric heparin-binding secreted protein that functions as a potent stimulator of growth and migration of vascular endothelial cells derived from arteries, veins, and even lymphatics (reviewed in [9]) (Fig. 1). VEGF-A binds and activates VEGF receptor-1 (VEGFR-1/Flt-1) and VEGF receptor 2 (VEGFR-2/KDR) expressed primarily on endothelial cells. Both VEGFR-1 and VEGFR-2 are membrane-bound tyrosine kinase receptors. VEGF binding to VEGFR-2 initiates receptor dimerization and activates a signaling cascade resulting in increased endothelial cell proliferation, while its binding to VEGFR-1 results in endothelial cell migration and differentiation (reviewed in [10]). VEGF-A plays a very significant role in embryonic vessel development wherein targeted genetic disruption of even a single allele of VEGF results in abnormal blood vessel development and lethality observed in mice [11, 12]. Decades of research have uncovered the extremely significant role played by VEGF in the regulation of both normal and pathological angiogenesis.

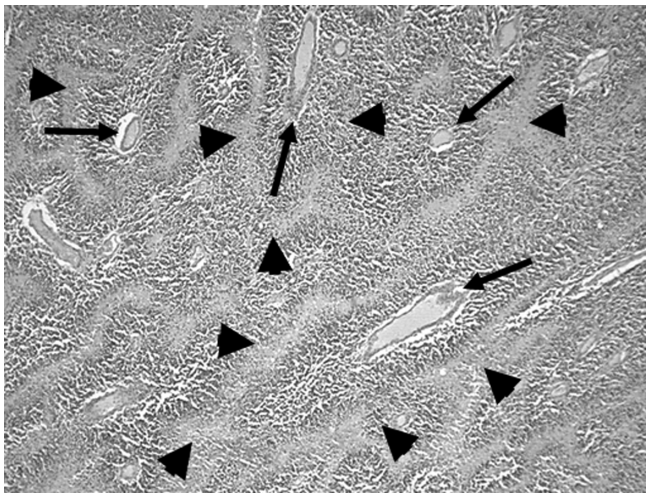

**Figure 2** Glioblastoma multiforme with geographic necrosis (pseudopalisading of tumor cells surrounding linear necrotic zones). Hematoxylene and Eosin images of glioblastoma specimen at 40× magnification. Areas distant from tumor vasculature (arrows) appear to be devoid of tumor cells and are referred to as hypoxic necrotic areas (arrow heads). These hypoxic necrotic areas appear to form a ring around the tumor vasculature and represent areas with poor perfusion.

### 2.3 VEGF Regulation

VEGF expression in human cancers has been shown to correlate with increased tumor vascularity and in many instances a worse prognosis. Overexpression of VEGF in human tumors promotes neovascularization and malignant growth. VEGF gene expression is rapidly induced by hypoxia and its increased expression in solid tumors is primarily a consequence of poor tissue oxygenation due to aberrant tumor vasculature. This is perhaps most apparent in pathological specimens derived from glioblastoma multiforme (GBM) grade IV malignant astrocytomas, in which areas of hypoxia/necrotic tissue are thought to be the result of poor tissue oxygenation [13] (Fig. 2). The resulting tissue hypoxia results in stabilization and activation of hypoxia inducible factor (HIF1). HIF1 is a hetrodimeric transcription factor that activates transcription of numerous hypoxia-responsive genes including VEGF [10]. Apart from increased gene expression, hypoxia-responsive 3′ untranslated region (3′UTR) in VEGF messenger ribonucleic acid (mRNA) can also function to increase its stability. Consistent with these observations, increased expression of VEGF is often observed adjacent to hypoxic necrotic areas in tissue specimens (Fig. 3). Increased VEGF in GBM often results in the appearance of hyperplastic or rapidly dividing endothelial and pericyte cell aggregates called *"glomeruloid bodies"*. These are often located in proximity to hypoxic necrotic centers within the tissue [14]. Figure 3 shows an example of a GBM specimen displaying hypoxic necrotic tissue cores and glomeruloid bodies. Hypoxia-independent activation of signaling pathways resulting from mutations in oncogene and/or tumor suppressor genes can also trigger increased levels of VEGF and subsequent angiogenesis [15]. Hence VEGF dysregulation appears to be a common factor in many different malignancies.

### 2.4 Preclinical Studies Targeting Tumor VEGF as a Therapeutic Target

Research over the last decade has established VEGF as a key endothelial-cell-specific mitogen that is essential for pathological and normal vascularization. Anti-VEGF murine monoclonal antibodies were initially generated to identify biologically relevant VEGFR-binding regions of human VEGF [16]. One of these VEGF-binding antibodies (A.4.6.1) abrogated the ability of VEGF to bind to and activate both VEGFR-1 and VEGFR-2 on endothelial cells. This murine monoclonal antibody was then evaluated for its ability to suppress angiogenesis and tumor growth in athymic mice bearing subcutaneous rhabdomyosarcoma and glioblastoma cells. The results indicated that treatment of tumor-bearing mice with A.4.6.1 inhibited tumor growth with a concurrent decrease in blood vessel density [17]. Ensuing studies using numerous human cancer cell lines confirmed the antitumorigenic potential of targeting VEGF using A.4.6.1 and other VEGF antibodies, soluble dominant negative VEGFR-2

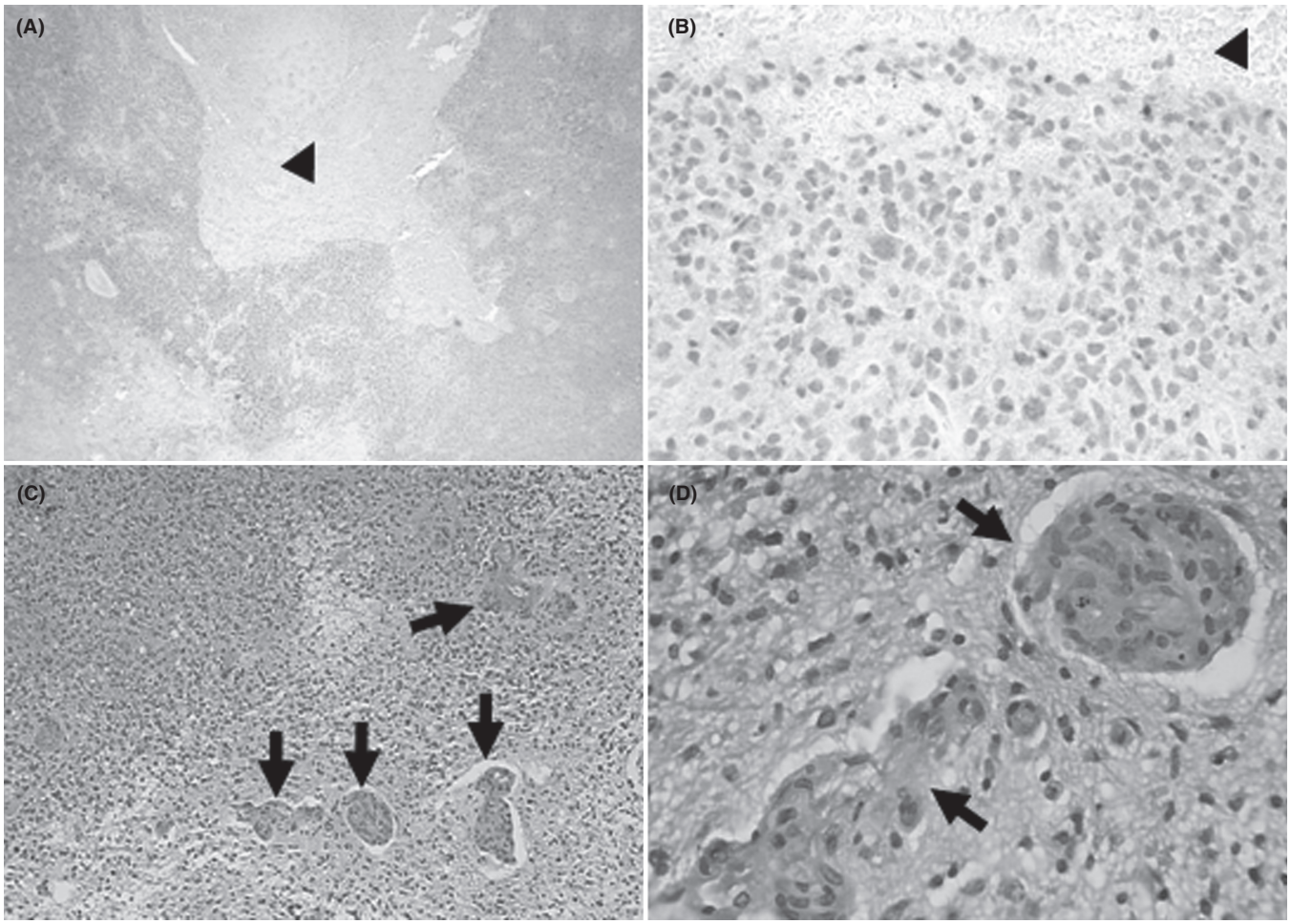


**Figure 3** Increased VEGF in hypoxic tumor areas: (A and B): Sections derived from human glioblastoma tissue showing tumor cells adjacent to a necrotic zone (arrows) showing positive reaction with VEGF immunostain (brown staining) at low (A) and high magnification (B). The positive cells display brown cytoplasmic staining. Note the increased expression of VEGF adjacent to hypoxic necrotic areas (arrow head) in the tumor. (C and D): Hematoxin and Eosin staining of glioblastoma multiforme tissue showing microvascular proliferation at low (C) and high magnification (D). Presence of tufted aggregates of proliferating endothelial and pericytes, called glomeuloid bodies, is indicated by arrows.

receptors, small-molecule inhibitors, and anti-VEGFR-2 antibodies in vivo (Table 1) [18].

### 2.5 Pharmacological Formulation of Bevacizumab

Interfering with VEGF signaling by delivery of the murine monoclonal antibody A.4.6.1 was found to have a therapeutic advantage for anticancer therapy in many different animal models (Table 1). However, a major limitation in the use of murine monoclonal antibodies in human patients is the anti-immunoglobulin response initiated against the antibody. Humanization of monoclonal antibodies is a powerful technique used to develop antibodies that closely resemble the human IgG sequence. This technique is an effective means to avoid an antiglobulin response against itself upon administration to human patients. Antibodies are composed of two identical heavy (H) polypeptide chains and two identical light (L) chains, which are covalently bonded to each other by interchain cysteine–cysteine disulfide (S–S) linkages (Fig. 4). Humanization of the murine A.4.6.1 was achieved by transferring six complementary determining regions (CDR) from the murine A.4.6.1 into a consensus human framework, consisting of the human heavy-chain subgroup III and the light-chain subgroup κI [19]. However, simple interchange of the CDRs from murine to human antibody framework reduced the VEGF-binding affinity of the humanized antibody by more than 1350-fold as compared to the murine A.4.6.1. Seven non-CDR framework residues were then mutagenized to achieve the recombinant humanized monoclonal antibody: bevacizumab

**TABLE 1 Preclinical Studies Demonstrating Antitumor Efficacy by Targeting Tumor VEGF**

| | Preclinical Tumor Model Tested | Mechanism of VEGF Inhibition | Noteworthy Observations |
|---|---|---|---|
| 1 | Human rhabdomyosarcoma (A673), glioblastoma multiforme (G55), and leiomyosarcom (SK-LMS-1) cells implanted as subcutaneous xenograft tumors in athymic nude mice | Antihuman VEGF mouse monoclonal antibody (A.4.6.1). | Inhibition of tumor growth and vascularity in vivo [17] |
| 2 | Colon carcinoma (HM7) cells grown as experimental hepatic metastases in athymic nude mice | Antihuman VEGF mouse monoclonal Antibody (A.4.6.1). | Reduction in the number and size of hepatic metstases [157] |
| 3 | Human prostate cancer (DU145) cells grown as subcutaneous xenografts in athymic nude mice | Antihuman VEGF mouse monoclonal antibody (A.4.6.1). | Tumor growth and formation of distant metastases in lungs was suppressed [158] |
| 4 | Human fibrosarcom (HT1080) cells | Antihuman VEGF mouse monoclonal antibody (MV303). | Inhibition of primary and metastatic tumor growth in nude mice [159] |
| 5 | Human glioblastoma (C6) cells grown as subcutaneous tumors in athymic nude mice | Retrovirus encoding dominant negative mutant of VEGFR/Flk-1 (Flk-1TM) | Suppression of tumor xenograft growth [160] |
| 6 | Epidermal carcinoma (A431), colon carcinoma (Ls174T and HT-29), prostate carcinoma (PC-3, DU-145, and CWR-22) cells grown as subcutaneous tumors in mice | PTK787, a small molecule inhibitor of VEGF receptor tyrosine kinase | Suppression of tumor growth and development of distant metastases [161] |
| 7 | Breast (MDA-MB-231) Lung (Calu-3), prostate (PC-3), and ovarian (SKOV-3) tumors grown as xenografts in nude mice | ZD4190, a small-molecule inhibitor of VEGF receptor tyrosine kinase. | Significant antitumor efficacy observed in prostate and lung cancer cell line derived tumors in mice [162] |
| 8 | Human colon (SW620), lung (Calu-6), prostate (PC3), Breast (MDA-MB-231), and ovary (SKOV-3) cells grown as subcutaneous tumors in mice | AZD2171, a small-molecule inhibitor of VEGF receptor tyrosine kinase. | Significant inhibition of tumor growth observed in all tumor models [163] |
| 9 | Human epidermoid (A431), renal cell carcinoma (SK-RC-29), pancreatic (BxPC-3), or glioblastoma (GBM-18) tumors grown as subcutaneous tumors in nude mice, and Lewis lung, B16 melanoma, and 4T1 breast tumors established in C57BL/6 or BALB/c mice | Anti-VEGFR/Flt-1 mouse monoclonal antibody (DC101). | Tumor suppression observed in several mouse and human tumor models [164] |

(Avastin) (RhuMAb). Bevacizumab consisted of a sequence that was 93% homologous to human immunoglobulin (the entire framework) and 7% homologous to murine immunoglobulin (the region in A.4.6.1 that binds specifically to human VEGF). Analysis of the binding affinity kinetics of bevacizumab and A.4.6.1 revealed that both antibodies had very similar binding affinity to human VEGF. Additionally, the two antibodies were indistinguishable in their ability to antagonize VEGF bioactivity both in endothelial cell cultures (ED50s against endothelial cell proliferation were $50 \pm 5$ and $48 \pm 8$ ng/mL, respectively) in vitro and in their ability to antagonize tumor growth in vivo in mice [19]. Numerous subsequent studies further confirmed the antitumor efficacy of the humanized anti-VEGF antibody alone and in combination with other agents in mouse models of tumorigenesis (Table 2). This engineering increased the half-life of bevacizumab while maintaining its high specificity for VEGF.

For clinical studies 100 mg of RhuMAb is provided in glass vials containing histidine, trehalose, polysorbate 20, and sterile water ready for injection. The vials contain no preservative and are intended for single use. The vial is diluted into 5% dextrose ($D_5W$) prior to infusion.

### 2.6 Safety Studies

Detailed preclinical safety evaluation of bevacizumab consisted of examining tissue reactivity, investigation of its hemolytic potential and blood compatibility studies, and multiple-dose toxicity studies in adult cynomologus monkeys. Absence of nontarget tissue binding was confirmed by immunohistochemistry of tissue derived from human donors, cynomologous monkey, and rabbit tissue. No tissue cross reactivity of nontarget tissue was observed with low (10 μg/mL) or high (400 μg/mL) doses of bevacizumab [20]. The compatibility of bevacizumab with human serum was further tested by incubating

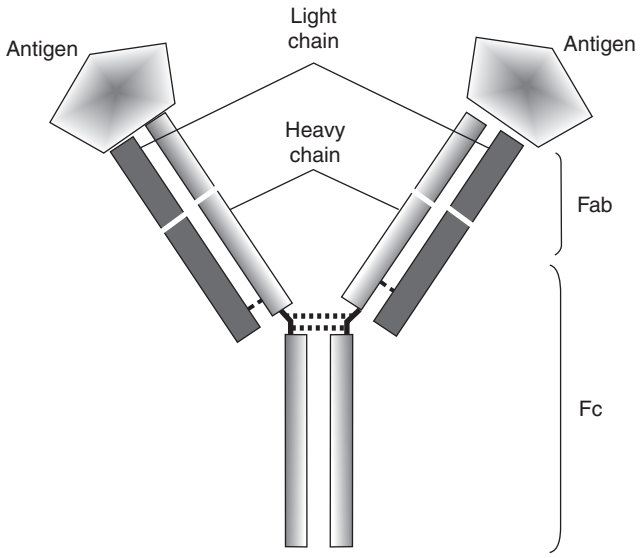


**Figure 4** Structure of an antibody: Each antibody contains two light chains (dark) and two heavy chains (light), held together by disulfide linkages (dotted lines). Each chain has a variable region and a constant region. The variable regions together constitute the antigen recognition domain within the antibody. Cleavage by papain at pH 7.0, results in the fragmentation of the antibody into two Fab fragments and one Fc fragment.

human serum with the antibody. No hemolysis, precipitation, or coagulation of human or monkey serum was observed indicating its serum/plasma compatibility [20]. Bevacizumab (0, 2, 10, and 50 mg/kg) was administered twice weekly intravenously to cynomologus monkeys for 4 or 13 weeks to determine its toxicity. Animals were monitored for weight loss, appetite, blood pressure measurements, electrocardiograms, physical activity, ophthalmic exams, and clinical pathology (hematology, clinical chemistry, and urinalysis). No bevacizumab-related clinical signs of toxicity were observed upon administration of the recombinant antibody. Bevacizumab treatment also led to inhibition of physiological angiogenesis into the growth plate and ovary and was reversible as shown by the resumption of angiogenesis when the treatment was stopped. This is consistent with its known function and was not considered an adverse event [20].

## 2.7 Clinical Studies with Bevacizumab

Based on the safety profile of bevacizumab, Genetech initiated phase I clinical trials with bevacizumab as a single agent and in combination with chemotherapy. In the first phase I study, 25 patients with metastatic cancer were given 0.1–10 mg/kg bevacizumab on days 0, 28, 35, and 42 by a 90-min intravenous infusion. Overall the infusions were well tolerated with no grade III or grade IV treatment-related adverse events [21]. Following these results several phase II studies with bevacizumab given as a single agent alone or in combination with chemotherapy were initiated [18]. A randomized phase II study comparing bevacizumab and fluorouracil (F)/leucovorin (L) with F/L treatment alone for patients with previously untreated metastatic colorectal cancer was initiated. One hundred and four patients were randomly assigned to be treated with either F/L alone (36 patients) or F/L + high-dose bevacizumab (10 mg/kg) (33 patients) or F/L + low-dose bevacizumab (5 mg/kg) [22]. Treatment with bevacizumab in combination with F/L showed an improvement in time to progression and overall survival for both the doses. Interestingly, patients treated with low-dose bevacizumab showed a better response than patients treated with the higher dose [22]. Fatal thrombosis was observed in one patient, and the lower dose was recommended for further studies in colorectal cancer patients. In a subsequent phase III trial, 813 patients were randomly segregated to receive irrinotecan, fluorouracil, and leucovorin (IFL) + bevacizumab (IFL + bevacizumab: 402 patients) or IFL + placebo (411 patients). The primary endpoint was to evaluate changes in overall survival, and the secondary endpoints were to determine if bevacizumab treatment induced progression-free survival, response rate, duration of response, safety, and quality of life. The addition of bevacizumab to IFL resulted in a statistically significant and clinically meaningful improvement in survival of patients with metastatic colorectal cancer. In 2004, administration of bevacizumab in combination with FL-based therapies was approved by the U.S. FDA for the treatment of patients with metastatic colorectal cancer. Administration of bevacizumab in combination with other chemotherapeutic drugs is being tested for efficacy as a front-line treatment for a variety of other tumor types. A recent randomized phase III study investigating bevacizumab in combination with interferon-α for patients with metastatic clear-cell renal-cell carcinoma showed improvements in progression-free survival and overall response rates [23].

## 2.8 Adverse Events Related to Bevacizumab Treatment

Hypertension was the most frequent bevacizumab-related adverse event that was noted in the studies testing combined therapy with bevacizumab + IFL in colorectal cancer patients. The underlying mechanism of hypertension is thought to be related to reduction in the levels of nitric oxide production as a consequence of VEGF blockade [24]. Other complications associated with bevacizumab treatment include gastrointestinal (GI) perforation, complications in wound healing,

**TABLE 2 Summary of Preclinical Studies for Antitumor Efficacy of the Humanized Anti-VEGF Antibody (RhuaVEGF)/ Bevacizumab**

| Agent | Tumor Type | Cell Line (Graft Location/Mouse Strain) | Dosage | Noteworthy Observation |
|---|---|---|---|---|
| *Single Agent* | | | | |
| RhuaVEGF | Androgen-independent prostate cancer | CWR223 (s.c/nude athymic BALB/c mice) | 5 mg/kg, given ip twice weekly for 4 weeks | >85% growth inhibition of established tumors with reduced MVD [165] |
| RhuaVEGF | Neuroblastoma | SK-N-AS<br>IMR-32<br>SH-SY5Y (s.c immunodeficient mice) | 5 mg/kg, given ip twice weekly once tumor volume was >0.3 mL | Significant reduction in NB tumor growth in vivo with 30–63% reduction of angiogenesis [166] |
| Bevacizumab | Non-small-cell lung cancer cells | SCH1299 (athymic nude BALB/c mice) | 100 μg/mouse ip twice weekly | Significant reduction in tumor growth [167] |
| *Combination with chemotherapy* | | | | |
| RhuaVEGF + Palcitaxel | Androgen-independent prostate cancer | CWR223 (s.c/nude athymic BALB/c mice) | RhuaVEGF: 5 mg/kg, ip twice weekly for 4 weeks, and Palcitaxel: 6.25 mg/kg s.c 5 times/week for 3 weeks | 98% growth inhibition of established tumors [165] |
| Bevacizumab + Ad-mda7 | Non-small-cell lung cancer cells | SCH1299 (athymic nude BALB/c mice) | Bevacizumab: 100 μg/mouse ip twice weekly for 4 weeks<br>Ad-mda7: $5 \times 10^9$ virus particles/dose intratumorally twice weekly for 4 weeks | Significant reduction in tumor growth, compared to Ad-mda7 treated alone or bevacizumab treated alone animals [167] |
| Bevacizumab + Ad5/3-Δ24 | Metastatic renal cancer | ACHN cells injected ip (SCID mice) | Bevacizumab: 50 μg/mouse once weekly for 5 weeks.<br>Ad5/3-Δ24: 108 virus particles once weekly for 3 weeks | No toxicity but no enhancement observed with co-treatment compared to treatment with the oncolytic adenovirus alone [168] |
| Bevacizumab + yeast-derived β-glucan | Ovarian carcinoma | SKOV-3 injected s.c (SCID mice) | Bevacizumab: 0.2 mg every third day, β-glucan: 1.2 mg every third day | Mice treated with combination had significantly smaller tumors than bevacizumab treated alone mice, and 86% of mice treated with combination were long-term survivors compared to 43% bevacizumab treated alone mice [169]. |

hemorrhage, thromboembolic events, proteinurea, and congestive heart failure. Careful monitoring and discontinuation of the drug is advised in such cases. More serious and even fatal pulmonary hemorrhage was noted in patients with non-small cell lung cancer (NSCLC) of squamous cell histology who were treated with bevacizumab [25]. Due to its impact on wound healing, bevacizumab therapy also cannot be initiated for at least 28 days before and after a surgery. Since angiogenesis during embryogenesis is essential for normal fetal development, bevacizumab is rated pregnancy category C. Due to its long half-life, patients are advised to avoid getting pregnant soon after discontinuing the drug.

## 3 PROTEASOME INHIBITORS

### 3.1 Proteasome as Therapeutic Target

The ubiquitin proteasome pathway regulates intracellular protein turnover and its function is critical to

cellular homestasis [26]. The 26s proteasome is a large multienzyme complex that functions to degrade or process intracellular proteins, including those involved in cell cycle progression and apoptosis [27, 28]. The proteasome is appreciated as playing a critical role in degrading intracellular proteins through both ubiquiti–dependent and ubiquitin-independent proteolysis [26, 29]. The proteasome consists of a 20S catalytic cylindrical complex capped at both ends by 19S regulatory subunits and is typically present within both the nucleus and cytoplasm of eukaryotic cells [30, 31]. Although other intracellular proteolytic processes are mediated by the immunoproteasome, the 26S proteasome mediates a substantial degree of protein turnover. The proteasome is central to the ubiquitin–proteasome pathway for ordered, temporal degradation of cyclins, cyclin dependent kinase (CDK) inhibitors (p21 and p27), and other proteins required for cell cycle progression in malignant cells as well as normal cells [32]. Because of its actions on cell cycle proteins, the proteasome represents a rational target for targeted therapy against neoplastic cells. Furthermore, the proteasome is known to play a role in regulation of apoptosis and/or cell survival through its effects on key mediators such as caspases, the Bcl2 family of proteins and nuclear factor κB (NF-κB) [32–34]. Based on these observations, a class of antitumor agents, termed *proteasome inhibitors* has emerged as a novel therapeutic option for the treatment of malignancies.

### 3.2 Discovery of Bortezomib as a Proteasome Inhibitor

To date, proteasome inhibitors have been separated into five classes based on their chemical structure: peptide aldehydes, peptide vinyl sulfones, peptide boronates, peptide epoxiketones, and β-lactones. The first agents to be identified as inhibitors of the proteasome were the natural inhibitor lactacystin and synthetic peptide aldehydes [27]. However, only two compounds (bortezomib/VELCADE/PS-341 and PS-519/MLN519) are currently in clinical development [27]. The latter compound, PS-519/MLN519, has been tested in healthy volunteers to determine safety, tolerability, and pharmacodynamic profiles [35]. Based on its protective ability in animal models of cerebral ischemia, it is thought that this compound may be an effective neuroprotective agent [36]. The former compound, bortezomib, has been primarily employed in the treatment of malignancy and will be the focus of the present discussion. Bortezomib is a small-molecule proteasome inhibitor (Fig. 5) developed by Millennium Pharmaceuticals, Inc., as a novel agent to treat malignancy. This drug is currently approved by the U.S. FDA for the treatment of patients

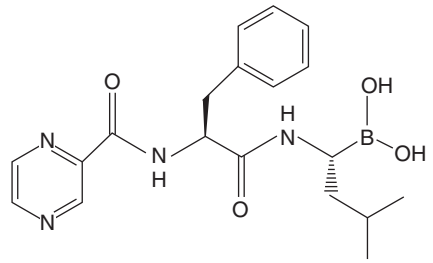


**Figure 5** Structure of bortezomib (reproduced from [171] with permission).

with multiple myeloma [37]. The rationale for developing bortezomib was that substitution of the aldehyde with boronic acid would create a compound that could form covalent and reversible complexes with the proteasome, resulting in increased potency and selectivity compared with their corresponding aldehydes [27]. As a result, a series of dipeptidyl boronic-acid proteasome inhibitors was developed and screened for activity against the standard National Cancer Institute (NCI) panel of 60 human tumor cell lines [38]. In these assays, bortezomib demonstrated effective in vitro cytotoxic activity against the tumor cell lines and was subsequently pursued in further studies for clinical development. Through its ability to inhibit the proteasome, bortezomib affects multiple signaling pathways.

### 3.3 Pharmacological Formulation of Bortezomib

Bortezomib is a modified dipeptidyl boronic acid formulated as an antineoplastic agent for intravenous injection. The chemical name for bortezomib, the monomeric boronic acid, is [(1R)-3-methyl-1-[(2S)-1-oxo-3-phenyl-2-[(pyrazinylcarbonyl) amino]propyl] amino]butyl] boronic acid. The product is provided as a mannitol boronic ester that, in reconstituted form, consists of the mannitol ester in equilibrium with its hydrolysis product, the monomeric boronic acid. The drug substance exists in its cyclic anhydride form as a trimeric boroxine. Typically, the compound is provided as a sterile lyophilized powder [Velcade (borezomib) for injection (from the package insert)].

### 3.4 Bortezomib Preclinical Studies and Mechanism of Action

Bortezomib-mediated reductions in the degradation of proteins involved in cell cycle progression and transcription factor activation leads to arrested growth of malignant cells via multiple mechanisms [33]. The first evidence that proteasome inhibition was a rational approach to cancer treatment arose from in vitro studies showing that proteasome inhibitors induced apoptosis

in leukemic cell lines, which were resistant to chemotherapy and radiation and in vivo studies using an animal model of non-Hodgkin's lymphoma [39–42]. Importantly, the apoptosis observed in these initial experiments appeared to be limited to transformed cells.

The antineoplastic effect of bortezomib likely involves several distinct mechanisms, including inhibition of cell growth and survival pathways, induction of apoptosis, and inhibited expression of genes that control cellular adhesion, migration, and angiogenesis. Thus, the mechanisms by which bortezomib elicits its antitumor activity may vary among tumor types, and the extent to which each affected pathway is critical to the inhibition of tumor growth could also differ. Bortezomib has a novel pattern of cytotoxicity against the NCI panel of 60 human tumor cell lines [33]. In addition, bortezomib has cytotoxic activity that is preferential to neoplastic cells in a variety of xenograft tumor models [32, 43, 44].

One particular pathway by which bortezomib is thought to exert its antitumor activity is via its effects on NF-κB [45]. NF-κB activity maintains cell viability through the transcription of factors that inhibit the apoptotic process induced by environmental stress and cytotoxic agents [46]. The proteasome promotes the activation of the NF-κB signaling pathway via degradation of its inhibitory protein IκB. Stabilization of IκB (e.g., by proteasome inhibition with bortezomib) and blockade of NF-κB activity renders cells more susceptible to apoptosis [38]. Reduced NF-κB activity also inhibits the expression of cell adhesion molecules such as intercellular adhesion molecule-1 (ICAM-1) and vascular cell adhesion protein-1 (VCAM-1), which are important for tumor metastasis and angiogenesis in vivo [27].

Despite the broad effects downstream of NF-κB, bortezomib and other proteasome inhibitors work through a number of mechanisms separate from its inhibition of NF-κB. For example, in the presence of bortezomib, cyclin and CDK inhibitor proteins required for cell cycle progression are stabilized and lead to cell cycle arrest [32]. Proteasome inhibition with bortezomib may also inhibit tumor progression via the inhibited release of pro-angiogenic factors such as VEGF and basic fibroblast growth factor (bFGF) [47]. Notably, bortezomib induces apoptosis in cells that overexpress bcl-2, a genetic trait that confers unregulated growth and resistance to conventional chemotherapeutics [34] and is believed to induce apoptosis by a p53-independent mechanism [28, 32, 38, 48]. Furthermore, bortezomib has been reported to activate both caspase-8 and caspase-9, suggesting involvement of both intrinsic and extrinsic apoptotic pathways. Notably, the mechanism of bortezomib-induced apoptosis in transformed cells appears to be highly variable, across tumor cells of different origin. For example, the differential induction of the pro-apoptotic Noxa protein has been shown to occur in melanoma cells upon exposure to bortezomib [49, 50]. Most recently, proteasome inhibition has been reported to induce endoplasmic reticulum stress and apoptosis through the unfolded protein response in human pancreatic cancer cell lines [51–53]. Together these data highlight the wide spectrum of pro-apoptotic events that are induced in response to proteasome inhibition.

### 3.5 Clinical Experience with Bortezomib

As an anticancer agent, bortezomib has shown particular promise in the setting of multiple myeloma. It has recently been approved by the FDA for patients with multiple myeloma who have received at least two prior therapies and have demonstrated disease progression on their last therapy [54]. Bortezomib is thought to be efficacious in multiple myeloma via its inhibition of NF-κB activation, its attenuation of interleukin-6 (IL-6)-mediated cell growth, a direct apoptotic effect, and possibly antiangiogenic and other effects [32]. To date, only one trial has tested bortezomib as a single agent in the treatment of metastatic melanoma and no complete or partial responses were observed [55]. However, bortezomib has shown clinical activity in other solid tumors, including renal cell carcinoma, non-small-cell lung carcinoma, androgen-independent prostate carcinoma, mantle cell lymphoma, and follicular non-Hodgkin's lymphoma [54, 56, 57].

An acceptable overall therapeutic index has been documented with bortezomib; however, there remains some risk of toxicity from this agent [37]. The most common dose-limiting side effects of bortezomib observed in phase I studies were fatigue, malaise, GI effects such as diarrhea, nausea, and loss of appetite, fluid retention, hypokalemia, hyponatremia, orthostasis, thrombocytopenia, and sensory neuropathy [58]. Subsequent clinical studies in patients with multiple myeloma have demonstrated that toxicity was observed in approximately 10% of patients receiving single agent bortezomib and that a majority of the dose-limiting toxicity associated with this drug are likely due to thrombocytopenia and neuropathy [58]. More detailed information on the pharmacokinetics and pharmacodynamics of bortezomib are beyond the scope of this review but is readily available from the product insert or the U.S. FDA [37, 58].

### 3.6 Future Studies: Combination Therapy with Bortezomib

The wide range of pro-apoptotic properties and acceptable toxicity profile of bortezomib in patients suggests

**TABLE 3 Clinical Trials Employing Bortezomib in Combination with Other Agents [170]**

| Other Agents Combined with Bortezomib | Tumor Type | Reference |
|---|---|---|
| Carboplatin | Ovarian or primary peritoneal | 59 |
| Thalidomide and dexamethasone | Multiple myeloma | 170 |
| Melphalan | Multiple myeloma | 62 |
| PEGylated liposomal doxorubicin | Multiple myeloma | 63 |
| Dexamethasone | Multiple myeloma | 65, 66 |
| Paclitaxel and carboplatin | Advanced malignancies | 67 |
| Melphalan and prednisone | Multiple myeloma | 68 |
| Docetaxel | Advanced solid tumors | 69 |
| Pegylated liposomal doxorubicin | Multiple myeloma | 70, 71 |
| Gemcitabine | Advanced solid tumors | 73 |
| Irinotecan | Advanced solid tumors | 73 |
| Gemcitabine and cisplatin | Advanced solid tumors | 74 |
| Thalidomide and dexamethasone | Multiple myeloma | 75 |

that this agent may be best utilized in combination with other anticancer therapeutics. Indeed bortezomib has been administered clinically in combination with a variety of other therapeutic agents (Table 3). A number of preclinical and clinical studies have demonstrated that bortezomib is an effective chemosensitizing agent [59–76]. Furthermore, preclinical work has demonstrated that combined treatment of malignant cells with bortezomib and receptor tyrosine kinase inhibitors, such as sorafenib, leads to enhanced apoptosis in vitro [77]. Proteasome inhibition has been recognized as an effective means for sensitizing tumor cells to tumor necrosis factor-related apoptosis-inducing ligand (TRAIL)-mediated apoptosis [78] and also appears to be a rational means of enhancing the antitumor activity of immune enhancing cytokines. Khan et al. demonstrated enhanced antitumor activity of bortezomib when administered together with interleukin-12 or interleukin-2 in murine models of neuroblastoma [79]. Other studies have demonstrated that combined administration of bortezomib and interferon-α can promote apoptosis and inhibit angiogenesis in human bladder cancer cell lines [78] and induce synergistic apoptosis in human melanoma cells [80]. Recent studies of B16 melanoma and multiple myeloma models also indicated that bortezomib may be an effective means to sensitize tumor cells to the lytic effects of dendritic cell (DC)-activated immune effector cells [81, 82]. The effects of proteasome inhibition on immune effector cells, therefore, remains an important area of investigation, as other reports have shown that proteasome inhibition can also lead to apoptosis in DC and inhibit their stimulatory capacity [83–85] and suppress $CD4^+$ T-cell function [86]. Finally, second- and third-generation proteasome inhibitors are currently under evaluation, and the available data suggests that these compounds will likely have enhanced specificity and anti-tumor properties [87, 88]. It will, therefore, be critical to evaluate the toxicity profile of these new-generation inhibitors as single agents and when given in combination with other anticancer and immune modulating therapies.

## 4 INTERFERON-A

### 4.1 History of Interferon-α (IFN-α)

The biological phenomenon known as "*interference*" was first described by Isaacs and Lindenmann in 1957 [89]. This phenomenon was defined as protection against viral infection in cells that had previously been exposed to another virus. This concept was examined because of its public health implications during a time in which there was an ongoing search for an "antiviral penicillin." In their research examining the molecular mechanisms of this phenomenon, Isaacs and Lindenmann discovered interferon (IFN) [89, 90]. Crude protein fractions, consisting of less than 1% of IFN and unsuccessfull attempts to purify and produce IFN in large quantities was a mjor limitation in the field [91]. It was not until the 1960s when IFN was recognized as a family of distinct structurally related proteins rather than a single molecule [92]. In this same decade, additional interest in the biological properties of IFN led to the discovery of its cytostatic, antitumor, and immunomodulatory properties [92, 93]. IFN research was later advanced by the work leading to its purification to homogeneity in solution. This facilitated production of the cytokine in sufficient amounts to study its physicochemical properties [94] and later for the cloning of the first IFN gene [95]. Competition between the pharmaceutical industry and academic research eventually led to the production of IFN for use in patients, and this substance was one of the first natural proteins to become available for the clinical treatment of malignancies [96]. Human IFNs are a heterogenous family of proteins that function endogenously to protect the host from viral infection. Type I IFN genes are clustered on human chromosome 9p21 [97] and mediate their biological effects through a common receptor, IFN-αβR, which is ubiquitously expressed throughout the body. The IFN gene family consists of 13 IFN-α members, a single member each of IFN-β, IFN-κ, IFN-ω, and IFN-ε [98–100]. Despite the large number of endogenously produced IFNs, IFN-α is the subtype that is used

most frequently in clinical medicine and, therefore, will remain the focus of the present chapter.

### 4.2 Pharmacologic Formulation of IFN-α

Interferon-α is available from two commercial sources. Either Roferon-A (Interferon-alfa-2a, recombinant; Roche, Inc.) or INTRON-A (Interferon-alpha-2b; Schering-Plough, Inc.). Both formulations are manufactured by recombinant deoxyribonucleic acid (DNA) technology that employs a genetically engineered *Escherichia coli* bacterium containing DNA that encodes for either the human inteterferon-α-2a or interferon-α-2b protein (165 amino acids; molecular weight =19,000 Da). Fermentation is carried out in a defined nutrient medium containing the antibiotic tetracycline hydrochloride; however, the presence of the antibiotic is undetectable in the final, purified product. Roferon-A is supplied both as an injectable solution in a vial or a prefilled syringe. Roferon-A formulations typically contain Interferon-alpha-2a ranging from 3 to 36 million international units (IU), and contain polysorbate 80, benzyl alcohol, and ammonium acetate as a preservative. Intron-A is supplied as a sterile lyophilized powder for reconstitution and injection, which is formulated with glycine, sodium phosphate dibasic, sodium phosphate monobasic, and human albumin. Intron-A is also supplied in solution for injection or as a multidose pen for injection. These formulations contain sodium chloride, sodium phosphate dibasic, sodium phosphate monobasic, edentate disodium, polysorbate 80, and *m*-cresol as a preservative.

### 4.3 Clinical Experience with IFN-α as a Treatment for Malignancy

Cytokine therapy with IFN-α is regularly used as a treatment for metastatic melanoma and renal cell carcinoma. IFN-α is also FDA-approved for melanoma patients as an adjuvant following the surgical resection of thick primary tumors and dissection of regional lymph nodes [101–103]. Historically this cytokine has also been used to treat chronic myelogenous leukemia and Kaposi's sarcoma [104, 105]. The recombinant form of this cytokine became available for use in human clinical trials during the 1980s [106]. In the setting of metastatic melanoma, the response rate to IFN-α typically ranges from 10 to 20%, where, in general, complete responses represent one-third of total responses [108, 109]. Likewise, a 15% objective response rate and a significantly prolonged overall survival of patients with metastatic renal cell carcinoma (RCC) was observed in response to single-agent IFN-α in randomized clinical trials [110, 111]. Subsequently, a series of clinical trials evaluated the use of IFN-α in the adjuvant setting for melanoma patients who underwent complete surgical resection of their tumor but were at high risk for recurrent disease. An initial phase III trial was conducted in which patients were randomized to receive either high-dose IFN-α (20 MU/m$^2$/d intravenously) for 1 month followed by 10 MU/m$^2$/thrice weekly subcutaneously or observation [112]. The results of this trial indicated an increase in both the overall and relapse-free survival of IFN-α-treated patients [112]. Subsequent trials of high-dose and low-dose IFN-α regimens (3 MU thrice weekly for 2 years) also demonstrated an increase in relapse-free survival of high-dose IFN-α-treated patients. However, these studies did not report a difference in overall survival in a similar population of surgically treated patients. Despite these conflicting results, an adjuvant trial comparing an antiganglioside vaccine to high-dose IFN-α was stopped as the result of an early analysis that showed a clear survival benefit to patients in the IFN-α arm [103]. Overall, IFN-α remains the only FDA-approved adjuvant therapy for patients with melanoma who have undergone surgical resection of high-risk lesions. There remains no alternative therapy that has shown efficacy in randomized trials in this setting, and IFN-α will likely continue as the standard of care in the adjuvant setting and will continue to be given in combination with other therapeutic agents. Accordingly, there remains a need to identify relevant biomarkers that might be useful for predicting clinical response or toxicity in response to this therapy.

### 4.4 Dosage and Toxicity Considerations

As with many other cytokine therapies, the use of IFN-α is tolerated quite well by many patients but results in a marked degree of toxicity in others. Particularly when given at high-doses, IFN-α is associated with a unique spectrum of toxicities that often limit patient compliance and have tempered enthusiasm for using this therapy as a first-line agent for metastatic disease [113]. The most common adverse reactions to high-dose IFN-α are flulike symptoms including fever, fatigue, headache, myalgias, chills, and arthralgias. In some cases, IFN-α treatment may result in depression; therefore, IFN-α is not recommended in patients with a pre-existing psychiatric condition. Gastrointestinal effects, mild proteinuria, intermittent or transient skin rashes, and myelosuppression are also commonly seen. Hepatic toxicity is manifested primarily as liver enzyme disturbances, which are generally mild to moderate and transient. Pulmonary complications are relatively unusual and may include dyspnea, cough, or more rarely pneumonitis or pneumonia. Edema, hypotension, arrhythmia, and tachycardia have been observed and

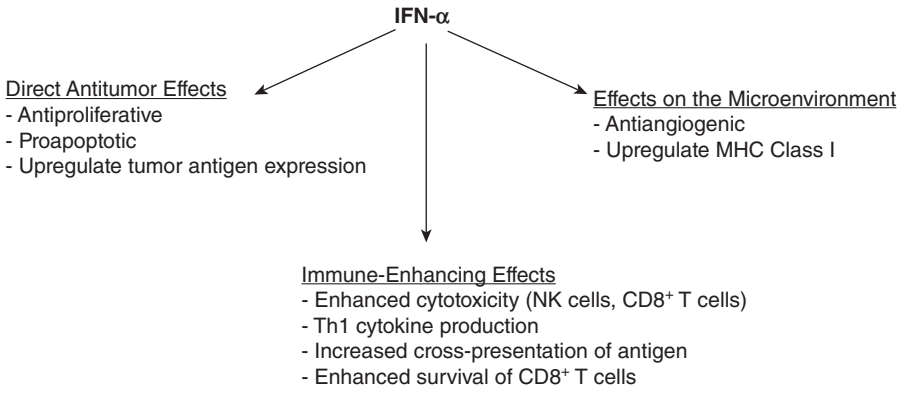


**Figure 6** Mechanism of interferon-α antitumor action. The molecular targets critical for the antitumor properties of IFN-α have not been precisely defined. However, IFN-α is known to act directly on the tumor cell, on the microenvironment, and via its ability to stimulate the host immune system.

may be related in some cases to the febrile reaction that occurs frequently during initial therapy. IFN-α could potentially lead to a depression of the hepatic clearance of certain drugs that use the hepatic microsomal cytochrome P-450 pathway [113].

In contrast, lower or moderate doses of this cytokine are often well tolerated and in rare cases can be administered in a sustained manner for several months or even years [101, 114]. The dosages employed in early phase I and phase II trials ranged from 3 to 100 MU per square meter ($m^2$) and would be considered very high dose regimens according to the current guidelines. The route of IFN-α administration does not appear to be a critical factor influencing clinical response. Indeed, investigators have reported complete clinical responses following both intravenous (i.v.) or intramuscular (i.m.) routes of administration in sequential clinical trials conducted at the same institution [108]. Despite this variability in dosage and in subject populations of the early trials, the response rates ranged between 5 and 30%. Although complete clinical responses have been observed most frequently upon the administration of higher doses (10–50 MU/$m^2$), complete responses have been shown to occur in response to very low doses of IFN-α (1–3 MU/$m^2$) [108, 113]. Recent data have also demonstrated that escalating clinical doses of IFN-α are no more effective at initiating signal transduction and gene expression in immune effector cells [115, 116]. Indeed, recent experience with tumor-targeted antibodies or other biologic agents demonstrate that higher doses do not necessarily exert the optimal biologic effect. These data are intriguing in that they suggest elevated doses of cytokines associated with greater toxicity may not necessarily be producing a greater clinical benefit. Therefore, it remains of value to explore strategies to individualize dosing of IFN-α and other immunotherapeutic cytokines on a patient-by-patient basis.

### 4.5 Mechanism of Action

The precise mechanism by which IFN-α mediates its antitumor activity remains undefined. Because IFN-α can bind to its receptor present upon both tumor cells and host cells, this cytokine can mediate pleiotropic effects in a biological system (Fig. 6). Indeed, in vitro studies have demonstrated that treatment of tumor cells with IFN-α leads to decreased cell proliferation and in some cases can promote apoptosis or programmed cell death [117–119]. Thus, the direct actions of IFN-α on the tumor cell were thought to constitute a major mechanism of action. Furthermore, other studies have confirmed a role for IFN-α in inhibiting the secretion of pro-angiogenic molecules from tumor cells (e.g., basic fibroblast growth factor) [120, 121] and upregulating the expression of major histocompatibility complex class I (MHC class I) proteins on tumor cells [122, 123].

Despite the apparent direct effects of IFN-α on the tumor cell, Belardelli and Gresser have conducted numerous eloquent studies investigating the mechanisms of antitumor effects of type I IFN in mice transplanted with Friend leukemia cells (FLC) [92, 124–129]. These tumor cells express Friend leukemia virus antigens on the cell surface and are metastatic to the liver and spleen following injection into syngeneic DBA/2 mice. Belardelli and Gresser demonstrated that IFN-α induced antitumor activity in mice transplanted with an IFN-resistant FLC clone. From these and subsequent studies [125, 130, 131], it was demonstrated that IFN-α acted to elicit antitumor activity through host-dependent mechanisms.

A series of more recent studies have further dissected the mechanism of host-dependent, IFN-α-mediated antitumor activity in murine models of melanoma. We found that the antitumor effects of IFN-α were dependent upon signal transducer and activator of transcription-1 (STAT1)-mediated signal transduction in the host immune cells, rather than the tumor cell [132, 133]. Subsequent studies have further demonstrated that exogenous IFN-α has profound antitumor activity when administered to mice lacking the negative regulatory suppressor of cytokine signaling-1 (SOCS1) protein (an inhibitor of STAT1 signal transduction) [134]. This finding is particularly intriguing as it suggests that modulating the negative regulators of IFN-α-induced signal transduction might be a means of sensitizing a cell to enhanced effects of IFN-α.

Clinical studies have shown that infiltration of $CD8^+$ T cells in cutaneous lesions from melanoma patients was associated with a favorable clinical response to IFN-α therapy [135]. Finally, studies indicate that melanoma patients who develop autoantibodies and clinical manifestations of autoimmunity during the course of adjuvant IFN-α therapy are less likely to experience disease recurrence. In a report by Gogas et al., autoantibodies were detected in 26% of melanoma patients following adjuvant IFN-α therapy ($n = 52$ patients total). Univariate and multivariate regression analyses indicated that autoimmunity was an independent prognostic marker for improved relapse-free survival and overall survival [136]. The development of autoimmunity may signify that exogenous IFN-α can break the tolerance of host T lymphocytes to host antigens as well as melanoma-associated antigens and further suggests that a principal mechanism of its antitumor action is immunologic in nature.

### 4.6 Future Studies with IFN-α

The extensive clinical experience with IFN-α and the fact that it can bind to receptors on both host immune effectors and tumor cells makes it an attractive candidate for combination therapy. Over the last several years, IFN-α has been administered in combination with multiple chemotherapeutic agents, cytokines, or tumor-targeting antibodies [23, 137–139]. In addition, more stable, PEGylated forms (PEG-INTRON) of IFN-α have become available for clinical use [140]. It is thought that this formulation might potentially allow for prolonged stability of the drug in vivo and fewer toxic side effects. As with most other cytokines, there remains a critical need to identify the molecular determinants that differentiate optimal dosing, patient toxicity, and clinical responsiveness to this drug. Advances in our mechanistic understanding in these areas could lead to important improvements in choosing which patients may be most likely to gain benefit from this potentially toxic therapy and to identify rational combinations for clinical use.

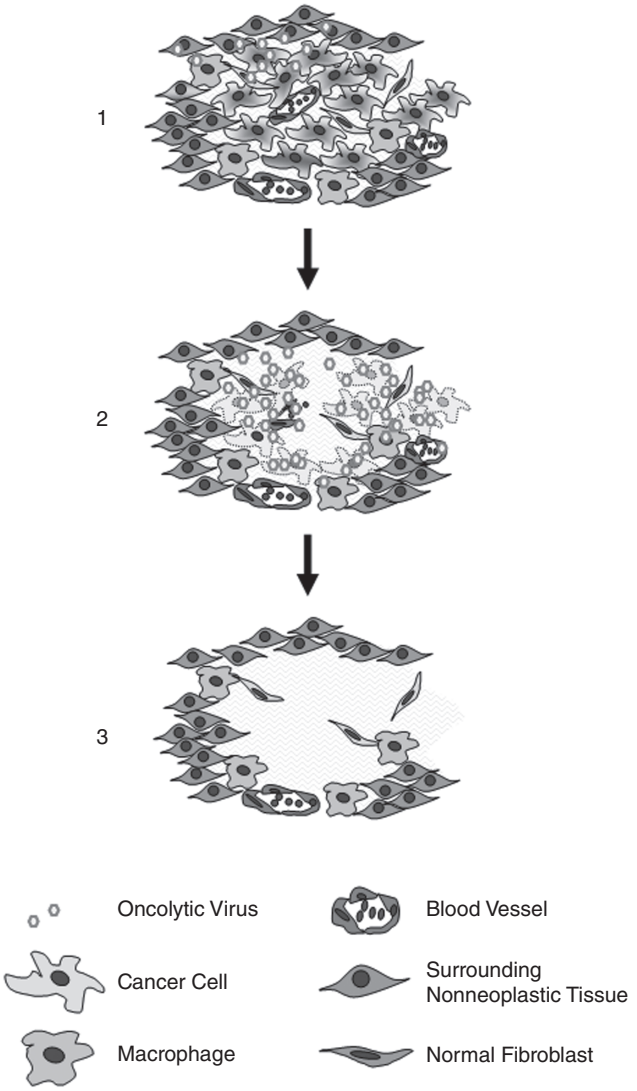


**Figure 7** Mechanism of anticancer effect of oncolytic viral therapy: Step 1: Initial infection of neoplastic tissue by oncolytic virus (OV) causes infection of both cancer cells and tumor associated normal fibroblasts and stromal cells. Step 2: The conditionally replication-competent virus is able to replicate only in the permissive cancer cells and proliferating tumor endothelium. Viral particles that infect normal cells are unable to replicate in the nonpermissive environment and are eventually cleared. Step 3: Tumor-specific replication of the oncolytic virus leads to lytic destruction of the infected cancer cells and the tumor supporting vasculature.

## 5 ONCOLYTIC VIRUS

Oncolytic virus (OV) is a biological cancer therapy based on the natural or engineered propensity of some viruses to replicate in and destroy cancer cells (Fig. 7). While the concept of using live viruses to infect and destroy tumors dates back nearly/over a century, advances in molecular biology and virology have accelerated the development of OV in the last two decades [141]. Therapeutic viruses used for the treatment of cancer can broadly be classified as replication-deficient and conditionally replication-competent viruses. Replication-deficient viruses are mostly used as gene therapy vehicles to deliver a therapeutic or cytotoxic gene to the cancer cells. Conditionally, replication-competent viruses are viruses whose replication potential is compromised in normal cells but can manipulate the aberrant molecular/genetic pathways in tumors to replicate efficiently in cancer cells. This leads to the lytic destruction of the infected malignant cell. Oncolytic viruses currently being tested for safety and efficacy include herpes simplex virus (HSV), and adenovirus, and RNA viruses such as Newcastle disease virus (NDV), measles virus, vaccinia virus (VV), and vesicular stomatitis virus (VSV). In this section we will describe the conception and testing of oncolytic viruses derived by genetically manipulating wild-type HSV-1-derived viruses.

### 5.1 Oncolytic Viruses Derived from Herpes Simplex Virus-1

HSV-1 is an enveloped, double-stranded DNA virus containing a large, well-characterized, fully sequenced

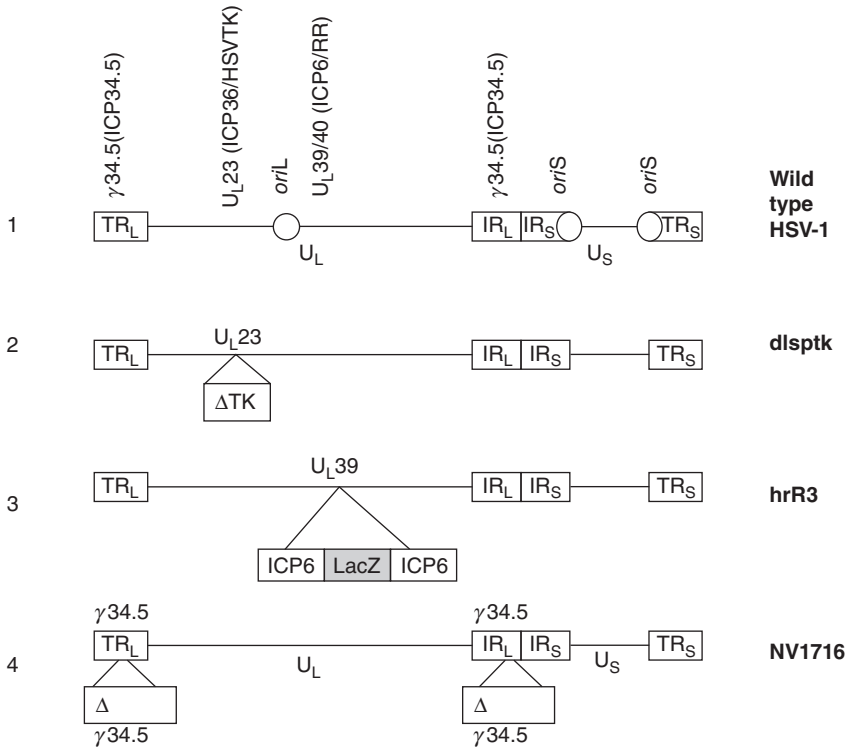


**Figure 8** Structure of the HSV-1 wild-type genome and some of the oncolytic viruses generated from it. (1) The HSV-1 genome consists of a linear double-stranded DNA of about 152 kb, encoding for more than 80 genes. Four inverted repeat sequences: Terminal repeat long ($TR_L$), terminal repeat small ($TR_S$), internal repeat long ($IR_L$), and internal repeat small ($IR_S$) divide the genome into unique long ($U_L$) and a unique small ($U_S$) segment. The three origins of replication within the HSV genome are indicated by circles. Locations of genes that are disrupted/deleted for the creation of tumor-specific replication competent viruses are indicated. (2) dlsptk is generated by a 360-bp deletion within the locus $U_L23$, encoding for viral HSVTK. (3) hrR3 is generated by an in-frame gene disrupting insertion of the *E. coli β-galactosidase* gene into the $U_L39$ locus encoding for large subunit of viral ribonucleotide reductase. (4) NV1716 is generated by deletion of both copies of the gene encoding for ICP34.5 within the terminal repeat sequences of the HSV genome.

genome of about 152 kb of DNA that encodes for more than 80 genes (Fig. 8). While the large size makes genetic manipulations cumbersome, it also provides ample opportunities to remove genes that are not essential for replication (estimated to be about 30 kb). These genes can be replaced to allow for the insertion of therapeutic transgenes within the viral backbone. Furthermore, the ability of HSV-1 to remain as an episome avoids the possibility of any insertional mutagenesis of the infected cell. These properties make HSV-1 a very desirable vector for therapeutic applications. In addition, HSV is sensitive to several antiherpetic drugs such as acyclovir (ACV) and gancyclovir (GCV), which are readily available to keep viral replication in check if needed. Together these factors make HSV an excellent candidate for oncolytic therapy.

### 5.2 HSVTK-Deleted Oncolytic Virus

HSV thymidine kinase (HSVTK) is a homidimeric protein encoded by $U_L23$ (ICP36) gene of the HSV genome. Unlike its mammalian counterpart, the viral enzyme has a very broad substrate specificity, permitting it to phosphorylate pyrimidine and purine nucleosides along with several nucleoside analog molecules. Viral HSVTK is required for nucleotide metabolism essential for DNA replication and in nondividing cells that do not express the mammalian counterpart of this enzyme. Based on these observations was born the hypothesis that an HSV virus deficient in HSVTK activity would be restricted for its replication in dividing cells, and this could be exploited for the generation of an oncolytic HSV. Subsequently, the first genetically altered HSV was created by a 360-bp (base pair) deletion within the viral gene infected cell protein 36 (ICP36) encoding for the enzyme thymidine kinase (HSVTK) [142] (Fig. 7). This mutant HSV (dlsptk) was tested for oncolytic efficacy in human glioblastoma cell lines and cultures in vitro. Furthermore, dlsptk treatment of human gliomas implanted into immune-deficient mice showed that treated tumors were smaller and mice survived longer than untreated mice with gliomas [142]. Further analysis of the pathogenesis of corneal infection of severe combined immunodeficiency (SCID) mice by dlsptk revealed persistent infection, which resulted in death of inoculated mice within $80.2 \pm 21.3$ days. The lack of sensitivity of dlsptk to commonly used antiherpetic drugs such as ACV and GCV (which exploit the virus-expressed HSVTK), combined with its neurovirulence precluded the use of this virus in human patients. However, the efficacy seen in preclinical animal models underscored the need for the development of safer oncolytic HSV vectors, which retain the antiherpetic drug-sensitizing HSVTK gene and are also less virulent.

### 5.3 Ribonucleotide Reductase-Deficient OV

Ribonucleotide reductase (RR) is required for the de novo synthesis of deoxyribonucleotides and is essential for viral DNA synthesis and replication. HSV-1 encodes for its own RR, composed of a large and small subunit encoded by $U_L39$ and $U_L40$, respectively. The proteins encoded by $U_L39$ and $U_L40$ together make up an $\alpha2\beta2$ tetramer constituting the functional RR. Like the HSVTK, viral RR is essential for viral replication in nondividing cells. However, the cellular enzyme compensates for this defect in cycling cells. Hence, an HSV lacking its own RR would be replication competent only in mitotic cancer cells but not in nonreplicating normal cells. To test this hypothesis a mutant virus, deficient for the large subunit of RR (ICP6/RR) was created by the insertion of an *E. coli LacZ* gene within the $U_L39$ gene locus [143] (Fig. 7). Cytopathic and drug sensitivity assays using this attenuated virus (hrR3) revealed efficient cancer cell killing in vitro and also efficient tumor killing in animal models of tumorigenesis [144, 145]. Figure 7 shows the destruction of brain tumors in rat with brain tumors treated with hrR3. Additionally, drug sensitivity assays revealed that the attenuated hrR3 was 10-fold more sensitive to antiherpetic drug GCV. HSVTK results in phosphorylation of GCV to a toxic nucleotide precursor, which results in death of infected cells and also uninfected bystander cells. This suggested that hrR3-infected glioma cells would be sensitive to killing by GCV. Consistent with this treatment of rats bearing intracranial tumors with GCV in combination with hrR3 significantly enhanced the antitumor effect of hrR3 [146]. More recent evidence also indicates that RR mutant HSV may be "molecularly targeted" to replicate more efficiently in cells with specific oncogenic mutations such as homozygous deletion of the *p16* gene [147].

### 5.4 ICP34.5-Deficient Oncolytic HSV

Interferon-induced, double-stranded RNA-activated protein kinase (PKR) is one of the principal cellular defenses against viral infection. Cellular PKR is maintained at low levels in uninfected cells but is induced by interferon. Activated PKR phosphorylates $\alpha$ subunit of eIF-2$\alpha$ leading to a total shutoff of host protein synthesis. HSV gene $R_L$-1 encodes for ICP34.5 protein, which binds to protein phosphatase 1$\alpha$ and leads to the dephosphorylation of $\alpha$ subunit of eIF-2$\alpha$. HSV-1 ICP34.5 was identified to be the neurovirulence gene and attenuation of both the copies of this gene led to

reduction of neurovirulence associated with HSV-1 in vitro and in vivo [148, 149]. More significantly this mutant virus retained the ability to replicate in actively dividing mouse embryo cells but not in confluent mouse embryo cultures [149]. Consistent with these results an RL-1 null HSV-1 mutant (1716), which fails to produce ICP34.5 was able to efficiently replicate in and cause subsequent cell death in a majority of glioma cell lines and primary tumor-derived cells [150] (Fig. 8). The in vivo antitumor efficacy of 1716 was tested in mice bearing experimental brain tumors, and treatment resulted in a statistically significant increase in survival [151, 152]. Pathological examination of mice injected intracranially with 1716, revealed a finite, self-limiting host response to 1716, highlighting the potential of using it for therapeutic purposes.

### 5.5 Pharmacological Formulation

The clinical-grade HSV 1716 used in human trials was prepared by amplifying low-passage laboratory stocks in BHK-21/C13 cells from a master seed cell bank under cGMP conditions (by Q-One Biotech Ltd., Glasgow, UK). The amplified clinical grade viral stocks were tested for contamination and evaluated for titers and stored at –70°C. The virus was stored at the appropriate dilution in 1 mL of human serum albumin, (0.5–1.0%) in isotonic phosphate-buffered saline (PBS). The titers of the virus were reconfirmed a week prior to the injection date. All studies were approved by the UK gene therapy advisory committee.

### 5.6 Clinical Testing in Human Patients

***HSV Mutant 1716*** Based on its safety profile in animals the mutated HSV 1716 was tested in UK for safety in human patients with malignant brain tumors [153]. This clinical study represented the first foray of inoculating mutated oncolytic HSV vectors into the brains of human patients with malignant brain tumors, and for concerns of safety, only a very moderate dosage of the virus was permitted to be tested. In this study, nine patients with malignant glioma were treated with doses of $10^3$, $10^4$, and $10^5$ plaque forming units (pfu) of HSV 1716 by direct intratumoral injection. All the patients were diagnosed with recurrent malignant glioma (eight glioblastoma and one anaplastic astrocytoma) that had relapsed following radical treatment, including surgery and radiation therapy. Patients were closely monitored and no adverse clinical events that could be attributed to the administration of 1716 were observed. No evidence of encephalitis or ongoing lytic replication was found in any treated patient. Buccal sawbs and serum samples from patients were examined weekly for presence of HSV, and no evidence of viral shedding or reactivation of endogenous latent virus was seen in any patient [153]. Five of these nine patients underwent a subsequent tumor progression, and biopsy samples from these patients showed no signs of 1716 or wild-type HSV by polymerase chain reaction (PCR) [153]. No evidence of efficacy was observed with inoculation of this dosage in patients, nevertheless the absence of any significant toxicity that could be attributed the administered OV was encouraging and subsequently two clinical trials to test therapeutic efficacy of HSV mutant 1716 were initiated. In a subsequent report by the investigators, it was noted that two of these nine patients were alive and stable at 4 years and 4 months and 3 years and 7 months after treatment [154].

Next, the therapeutic potential of HSV 1716 was also assessed in 5 patients with metastatic melanoma. In each patient paired distinct nodules were injected with either dilutent or $10^3$ pfu of HSV 1716. The very low dosage given in the first clinical study was based on the hypothesis that the oncolytic virus should be able to replicate and hence produce a higher dose within the tumor. In the second clinical study evidence of viral replication in human high-grade glioma was examined. Twelve patients with high-grade malignant glioma received $10^5$ pfu of HSV mutant 1716 by direct intratumoral stereotactic injection [155]. Tumor resection was carried out 4–9 days after treatment, and the tissue was examined for the presence of HSV DNA by PCR, Southern blotting immunohistochemistry, and the presence of infectious viral particles by plaque assay. Results from plaque assays revealed the presence of infectious viral particles in resected tumors from 2 out of 12 patients. Southern blot analysis confirmed that the recovered virus was HSV mutant 1716. HSV DNA was detected by PCR at the site of injection in 10 out of the 12 patients, and specimens from distal tumor sites of only 4 patients were found to be positive for HSV DNA [155]. Immunohistochemistry of tissue samples revealed positive staining for HSV proteins in tumor tissues from 2 of the 12 patients. The absence of detectable viral DNA protein or infectious particles in some specimens could be attributed to tissue sampling errors owing to nonuniform spread of the injected virus throughout the tumor tissue, or simply reflect the need to improve current methods used for detection. Nevertheless, the study demonstrated evidence of HSV presence in tissue without any obvious signs of toxicity [155].

Based on these studies, another clinical trial in the UK was initiated to test the therapeutic efficacy of the HSV mutant 1716 in human glioma patients. Despite the lack of toxicity seen in those trials, due to paramount safety concerns the therapeutic dose given to the patients was not increased as would have been done for clinical trials with a chemotherapeutic agent. A total of 12 patients who had been diagnosed with high-grade

recurrent (6 patients) or newly diagnosed (6 patients) glioma underwent gross tumor resection. After the initial tumor debulking, $10^5$ pfu of HSV mutant 1716 was injected into 8–10 sites within the tissue surrounding tumor cavity. The patients were carefully monitored with a comprehensive clinical work-up including immunological, hematological (full blood count with white cell count differential), biochemical (urea, electrolytes, liver function tests, and bone profile), and neurological examinations. The lack of any virus-related toxicity in this study considerably strengthened the safety profile of this approach even though the virus had been injected into tissue involving functioning brain. The expected median survival for patients with GBM is one year from the time of initial diagnosis and only 3–6 months following diagnosis of a recurrent tumor. One patient with a recurrent tumor and two patients diagnosed with primary GBM survived for more than 22, 18, and 15 months, respectively. The patient who survived for more than 22 months was diagnosed with a recurrent tumor but was disease free at 22 months. However, a causal relationship between response and viral treatment could not be established. This study demonstrated that HSV 1716 can be injected into the normal brain of individuals without any ensuing toxicity.

### 5.7 Future of HSV1716

HSV1716 is currently being pursued by Crusade Laboratories (Glasgow, UK). The company has been granted orphan drug status in Europe for "the use of HSV mutant 1716 in the treatment of glioma. The company has received regulatory approvals to initiate a phase III trial for glioma patients. Glioma patients with a first recurrence after surgery will be randomized into one of two treatment arms: HSV1716 or conventional chemotherapy. Satisfactory results in this trial could lead to a license and marketing authorization in glioma. The company is also pursuing phase I/II safety and biological activity studies in patients with ovarian carcinoma and unresectable hepatocellular carcinoma.

### 5.8 Second-Generation Multimutated Oncolytic Viruses: G207

In parallel with the HSV mutant 1716 studies going on in the UK, a doubly attenuated oncolytic HSV has also been investigated in clinics in the United States. Owing to the overriding concerns of safety arising from the use of singly mutated viruses, oncolytic viruses with multiple genetic defects were created for the treatment of cancer. G207 is one such virus that has deletions at both gamma 34.5 (RL1) loci and an in-frame and gene-disrupting insertion of *E. Coli* β-galactosidase gene within the ICP6 gene ($U_L39$), making it replication competent in dividing cells. This minimizes the chances of G207 reverting to wild type and maintains an attenuated neurovirulence, temperature sensitivity, as well as ganciclovir hypersensitivity [156]. The insertion of β-galactosidase permits easy detection of cells harboring the virus in infected tissue. Preclinical testing of this virus was performed in nude mice harboring subcutaneous or intracerebral gliomas. Intraneoplastic administration of the virus resulted in slower tumor growth and/or prolonged survival [156].

Preclinical toxicity studies of G207 also found it to be avirulent upon intracerebral inoculation of mice and HSV-sensitive nonhuman primates. Shedding and biodistribution of G207 was determined in New World owl monkeys (*Aotus nancymae*) after intracerebral inoculation of $3 \times 107$ pfu of clinical-grade virus. Even after 2 years of inoculation G207 DNA was found to be restricted to the brain, although no infectious viral particles were obtained. Histopathology of the necrospied tissue revealed normal brain even at the site of injection. Absence of virus DNA in tears, saliva, and vaginal secretions indicated lack of viral shedding and suggested that strict biohazard management of G207 patients may not be required. Together these results implicate the possibility of considering mutimutated oncolytic viruses such as G207 for clinical evaluation for the treatment of cancer.

Safety of intracerebral inoculation of G207 in humans was determined in a dose escalation study in patients suffering from recurrent malignant glioma. The protocol was approved by the recombinat DNA advisory committee of the National Institutes of Health and the Food and Drug Administration. The trial was initiated with the administration of $1 \times 10^6$ pfu of G207 into the brains of three patients posttumor resection. After 28 days of observation to confirm lack of acute toxicity, three patients were then recruited into the next cohort and were treated with the higher dosage. In the last cohort the patients were treated with $1 \times 10^9$ p.f.u of G207 inoculated into five sites after surgical resection of the brain tumor. No toxicity that could be directly attributed to the infectious virus was noted in this study. This study suggested that G207 can be safely inoculated into human brain tumors up to doses of $1 \times 10^9$ pfu without any adverse events. This agent is currently being investigated for efficacy in clinical trials. Future studies will reveal the efficacy of combining oncolytic viruses with radiation and chemotherapy in patients.

## REFERENCES

1. Jain RK (2003). Molecular regulation of vessel maturation. *Nat Med 9*, 685–693.

2. Folkman J (1971). Tumor angiogenesis: Therapeutic implications. *N Engl J Med 285*, 1182–1186.
3. Folkman J (1990). What is the evidence that tumors are angiogenesis dependent? *J Natl Cancer Inst 82*, 4–6.
4. Folkman J, Kalluri R (2004). Cancer without disease. *Nature 427*, 787.
5. Ferrara N, Henzel WJ (1989). Pituitary follicular cells secrete a novel heparin-binding growth factor specific for vascular endothelial cells. *Biochem Biophys Res Commun 161*, 851–858.
6. Senger DR, Galli SJ, Dvorak AM, Perruzzi CA, Harvey VS, Dvorak HF (1983). Tumor cells secrete a vascular permeability factor that promotes accumulation of ascites fluid. *Science 219*, 983–985.
7. Keck PJ, Hauser SD, Krivi G, Sanzo K, Warren T, Feder J, Connolly DT (1989). Vascular permeability factor, an endothelial cell mitogen related to PDGF. *Science 246*, 1309–1312.
8. Leung DW, Cachianes G, Kuang WJ, Goeddel DV, Ferrara N (1989). Vascular endothelial growth factor is a secreted angiogenic mitogen. *Science 246*, 1306–1309.
9. Ferrara N (2004). Vascular endothelial growth factor: Basic science and clinical progress. *Endocr Rev 25*, 581–611.
10. Kaur B, Khwaja FW, Severson EA, Matheny SL, Brat DJ, Van Meir EG (2005). Hypoxia and the hypoxia-inducible-factor pathway in glioma growth and angiogenesis. *Neurooncology 7*, 134–153.
11. Carmeliet P, Ferreira V, Breier G, Pollefeyt S, Kieckens L, Gertsenstein M, Fahrig M, Vandenhoeck A, Harpal K, Eberhardt C, Declercq C, Pawling J, Moons L, Collen D, Risau W, Nagy A (1996). Abnormal blood vessel development and lethality in embryos lacking a single VEGF allele. *Nature 380*, 435–439.
12. Ferrara N, Carver-Moore K, Chen H, Dowd M, Lu L, O'Shea KS, Powell-Braxton L, Hillan KJ, Moore MW (1996). Heterozygous embryonic lethality induced by targeted inactivation of the VEGF gene. *Nature 380*, 439–442.
13. Brat DJ, Van Meir EG (2004). Vaso-occlusive and prothrombotic mechanisms associated with tumor hypoxia, necrosis, and accelerated growth in glioblastoma. *Lab Invest 84*, 397–405.
14. Brat DJ, Castellano-Sanchez A, Kaur B, Van Meir EG (2002). Genetic and biologic progression in astrocytomas and their relation to angiogenic dysregulation. *Adv Anat Pathol 9*, 24–36.
15. Mizukami Y, Kohgo Y, Chung DC (2007). Hypoxia inducible factor-1 independent pathways in tumor angiogenesis. *Clin Cancer Res 13*, 5670–5674.
16. Kim KJ, Li B, Houck K, Winer J, Ferrara N (1992). The vascular endothelial growth factor proteins, Identification of biologically relevant regions by neutralizing monoclonal antibodies. *Growth Factors 7*, 53–64.
17. Kim KJ, Li B, Winer J, Armanini M, Gillett N, Phillips HS, Ferrara N (1993). Inhibition of vascular endothelial growth factor-induced angiogenesis suppresses tumour growth in vivo. *Nature 362*, 841–844.
18. Ferrara N, Hillan KJ, Novotny W (2005). Bevacizumab (Avastin), a humanized anti-VEGF monoclonal antibody for cancer therapy. *Biochem Biophys Res Commun 333*, 328–335.
19. Presta LG, Chen H, O'Connor SJ, Chisholm V, Meng YG, Krummen L, Winkler M, Ferrara N (1997). Humanization of an anti-vascular endothelial growth factor monoclonal antibody for the therapy of solid tumors and other disorders. *Cancer Res 57*, 4593–4599.
20. Ryan AM, Eppler DB, Hagler KE, Bruner RH, Thomford PJ, Hall RL, Shopp GM, O'Neill CA (1999). Preclinical safety evaluation of rhuMAbVEGF, an antiangiogenic humanized monoclonal antibody. *Toxicol Pathol 27*, 78–86.
21. Gordon MS, Margolin K, Talpaz M, Sledge GW Jr, Holmgren E, Benjamin R, Stalter S, Shak S, Adelman D (2001). Phase I safety and pharmacokinetic study of recombinant human anti-vascular endothelial growth factor in patients with advanced cancer. *J Clin Oncol 19*, 843–850.
22. Kabbinavar F, Hurwitz HI, Fehrenbacher L, Meropol NJ, Novotny WF, Lieberman G, Griffing, S. Bergsland E (2003). Phase II, randomized trial comparing bevacizumab plus fluorouracil (FU)/leucovorin (LV) with FU/LV alone in patients with metastatic colorectal cancer. *J Clin Oncol 21*, 60–65.
23. Escudier B, Pluzanska A, Koralewski P, Ravaud A, Bracarda S, Szczylik C, Chevreau C, Filipek M, Melichar B, Bajetta E, Gorbunova V, Bay JO, Bodrogi I, Jagiello-Gruszfeld A, Moore N (2007). Bevacizumab plus interferon alfa-2a for treatment of metastatic renal cell carcinoma: A randomised, double-blind phase III trial. *Lancet 370*, 2103–2111.
24. Shih T, Lindley C (2006). Bevacizumab: An angiogenesis inhibitor for the treatment of solid malignancies. *Clin Ther 28*, 1779–1802.
25. Bergsland E, Dickler MN (2004). Maximizing the potential of bevacizumab in cancer treatment. *Oncologist 9*(Suppl 1), 36–42.
26. Orlowski RZ, Kuhn DJ (2008). Proteasome inhibitors in cancer therapy: Lessons from the first decade. *Clin Cancer Res 14*, 1649–1657.
27. Adams J (2004). The proteasome: A suitable antineoplastic target. *Nat Rev Cancer 4*, 349–360.
28. Adams J, Kauffman M (2004). Development of the proteasome inhibitor Velcade (Bortezomib). *Cancer Invest 22*, 304–311.
29. Demartino GN, Gillette TG (2007). Proteasomes: Machines for all reasons. *Cell 129*, 659–662.
30. Brooks P, Fuertes G, Murray RZ, Bose S, Knecht E, Rechsteiner MC, Hendil KB, Tanaka K, Dyson J, Rivett J (2000). Subcellular localization of proteasomes and their regulatory complexes in mammalian cells. *Biochem J 346*(Pt 1), 155–161.

31. Peters JM, Cejka Z, Harris JR, Kleinschmidt JA, Baumeister W (1993). Structural features of the 26 S proteasome complex. *J Mol Biol 234*, 932–937.
32. Hideshima T, Richardson P, Chauhan D, Palombella VJ, Elliott PJ, Adams J, Anderson KC (2001). The proteasome inhibitor PS-341 inhibits growth, induces apoptosis, and overcomes drug resistance in human multiple myeloma cells. *Cancer Res 61*, 3071–3076.
33. Adams J, Palombella VJ, Elliott PJ (2000). Proteasome inhibition: A new strategy in cancer treatment. *Invest New Drugs 18*, 109–121.
34. Bold RJ, Virudachalam S, McConkey DJ (2001). Chemosensitization of pancreatic cancer by inhibition of the 26S proteasome. *J Surg Res 100*, 11–17.
35. Shah IM, Lees KR, Pien CP, Elliott PJ (2002). Early clinical experience with the novel proteasome inhibitor PS-519. *Br J Clin Pharmacol 54*, 269–276.
36. Zhang L, Zhang ZG, Zhang RL, Lu M, Adams J, Elliott PJ, Chopp M (2001). Postischemic (6-hour) treatment with recombinant human tissue plasminogen activator and proteasome inhibitor PS-519 reduces infarction in a rat model of embolic focal cerebral ischemia. *Stroke 32*, 2926–2931.
37. Kane RC, Farrell AT, Sridhara R, Pazdur R (2006). United States Food and Drug Administration approval summary: Bortezomib for the treatment of progressive multiple myeloma after one prior therapy. *Clin Cancer Res 12*, 2955–2960.
38. Adams J, Palombella VJ, Sausville EA, Johnson J, Destree A, Lazarus DD, Maas J, Pien CS, Prakash S, Elliott PJ (1999). Proteasome inhibitors: a novel class of potent and effective antitumor agents. *Cancer Res 59*, 2615–2622.
39. Delic J, Masdehors P, Omura S, Cosset JM, Dumont J, Binet JL, Magdelenat H (1998). The proteasome inhibitor lactacystin induces apoptosis and sensitizes chemo- and radioresistant human chronic lymphocytic leukaemia lymphocytes to TNF-alpha-initiated apoptosis. *Br J Cancer 77*, 1103–1107.
40. Imajoh-Ohmi S, Kawaguchi T, Sugiyama S, Tanaka K, Omura S, Kikuchi H (1995). Lactacystin, a specific inhibitor of the proteasome, induces apoptosis in human monoblast U937 cells. *Biochem Biophys Res Commun 217*, 1070–1077.
41. Orlowski RZ, Eswara JR, Lafond-Walker A, Grever MR, Orlowski M, Dang CV (1998). Tumor growth inhibition induced in a murine model of human Burkitt's lymphoma by a proteasome inhibitor. *Cancer Res 58*, 4342–4348.
42. Shinohara K, Tomioka M, Nakano H, Tone S, Ito H, Kawashima S (1996). Apoptosis induction resulting from proteasome inhibition. *Biochem J 317*(Pt 2), 385–388.
43. LeBlanc R, Catley LP, Hideshima T, Lentzsch S, Mitsiades CS, Mitsiades N, Neuberg D, Goloubeva O, Pien CS, Adams J, Gupta D, Richardson PG, Munshi NC, Anderson KC (2002). Proteasome inhibitor PS-341 inhibits human myeloma cell growth in vivo and prolongs survival in a murine model. *Cancer Res 62*, 4996–5000.
44. Teicher BA, Ara G, Herbst R, Palombella VJ, Adams J (1999). The proteasome inhibitor PS-341 in cancer therapy. *Clin Cancer Res 5*, 2638–2645.
45. Berenson JR, Ma HM, Vescio R (2001). The role of nuclear factor-kappaB in the biology and treatment of multiple myeloma. *Semin Oncol 28*, 626–633.
46. Hayden MS, Ghosh S (2008). Shared principles in NF-kappaB signaling. *Cell 132*, 344–362.
47. Nawrocki ST, Bruns CJ, Harbison MT, Bold RJ, Gotsch BS, Abbruzzese JL, Elliott P, Adams J, McConkey DJ (2002). Effects of the proteasome inhibitor PS-341 on apoptosis and angiogenesis in orthotopic human pancreatic tumor xenografts. *Mol Cancer Ther 1*, 1243–1253.
48. Chauhan D, Li G, Shringarpure R, Podar K, Ohtake Y, Hideshima T, Anderson KC (2003). Blockade of Hsp27 overcomes Bortezomib/proteasome inhibitor PS-341 resistance in lymphoma cells. *Cancer Res 63*, 6174–6177.
49. Fernandez Y, Verhaegen M, Miller TP, Rush JL, Steiner P, Opipari AW Jr, Lowe SW, Soengas MS (2005). Differential regulation of noxa in normal melanocytes and melanoma cells by proteasome inhibition: Therapeutic implications. *Cancer Res 65*, 6294–6304.
50. Qin JZ, Ziffra J, Stennett L, Bodner B, Bonish BK, Chaturvedi V, Bennett F, Pollock PM, Trent JM, Hendrix MJ, Rizzo P, Miele L, Nickoloff BJ (2005). Proteasome inhibitors trigger NOXA-mediated apoptosis in melanoma and myeloma cells. *Cancer Res 65*, 6282–6293.
51. Nawrocki ST, Carew JS, Dunner K Jr, Boise LH, Chiao PJ, Huang P, Abbruzzese JL, McConkey DJ (2005). Bortezomib inhibits PKR-like endoplasmic reticulum (ER) kinase and induces apoptosis via ER stress in human pancreatic cancer cells. *Cancer Res 65*, 11510–11519.
52. Nawrocki ST, Carew JS, Pino MS, Highshaw RA, Andtbacka RH, Dunner K Jr, Pal A, Bornmann WG, Chiao PJ, Huang P, Xiong H, Abbruzzese JL, McConkey DJ (2006). Aggresome disruption: A novel strategy to enhance bortezomib-induced apoptosis in pancreatic cancer cells. *Cancer Res 66*, 3773–3781.
53. Nawrocki ST, Carew JS, Pino MS, Highshaw RA, Dunner K Jr, Huang P, Abbruzzese JL, McConkey DJ (2005). Bortezomib sensitizes pancreatic cancer cells to endoplasmic reticulum stress-mediated apoptosis. *Cancer Res 65*, 11658–11666.
54. Papandreou CN, Daliani DD, Nix D, Yang H, Madden T, Wang X, Pien CS, Millikan RE, Tu SM, Pagliaro L, Kim J, Adams J, Elliott P, Esseltine D, Petrusich A, Dieringer P, Perez C, Logothetis CJ (2004). Phase I trial of the proteasome inhibitor bortezomib in patients with advanced solid tumors with observations in androgen-independent prostate cancer. *J Clin Oncol 22*, 2108–2121.
55. Markovic SN, Geyer SM, Dawkins F, Sharfman W, Albertini M, Maples W, Fracasso PM, Fitch T, Lorusso P, Adjei AA, Erlichman C. (2005). HYPERLINK "http://www.ncbi.nlm.nih.gov/pubmed/15887220" A phase II

study of bortezomib in the treatment of metastatic malignant melanoma. *Cancer Jun 15*, *103*(12), 2584–2589.

56. Aghajanian C, Soignet S, Dizon DS, Pien CS, Adams J, Elliott PJ, Sabbatini P, Miller V, Hensley ML, Pezzulli S, Canales C, Daud A, Spriggs DR (2002). A phase I trial of the novel proteasome inhibitor PS341 in advanced solid tumor malignancies. *Clin Cancer Res 8*, 2505–2511.
57. Orlowski RZ, Stinchcombe TE, Mitchell BS, Shea TC, Baldwin AS, Stahl S, Adams J, Esseltine DL, Elliott PJ, Pien CS, Guerciolini R, Anderson JK, Depcik-Smith ND, Bhagat R, Lehman MJ, Novick SC, O'Connor OA, Soignet SL (2002). Phase I trial of the proteasome inhibitor PS-341 in patients with refractory hematologic malignancies. *J Clin Oncol 20*, 4420–4427.
58. Bross PF, Kane R, Farrell AT, Abraham S, Benson K, Brower ME, Bradley S, Gobburu JV, Goheer A, Lee SL, Leighton J, Liang CY, Lostritto RT, McGuinn WD, Morse DE, Rahman A, Rosario LA, Verbois SL, Williams G, Wang YC, Pazdur R (2004). Approval summary for bortezomib for injection in the treatment of multiple myeloma. *Clin Cancer Res 10*, 3954–3964.
59. Aghajanian C, Dizon DS, Sabbatini P, Raizer JJ, Dupont J, Spriggs DR (2005). Phase I trial of bortezomib and carboplatin in recurrent ovarian or primary peritoneal cancer. *J Clin Oncol 23*, 5943–5949.
60. Amiri KI, Horton LW, LaFleur BJ, Sosman JA, Richmond A (2004). Augmenting chemosensitivity of malignant melanoma tumors via proteasome inhibition: Implication for bortezomib (VELCADE, PS-341) as a therapeutic agent for malignant melanoma. *Cancer Res 64*, 4912–4918.
61. Barlogie B, Anaissie E, van Rhee F, Haessler J, Hollmig K, Pineda-Roman M, Cottler-Fox M, Mohiuddin A, Alsayed Y, Tricot G, Bolejack V, Zangari M, Epstein J, Petty N, Steward D, Jenkins B, Gurley J, Sullivan E, Crowley J, Shaughnessy JD Jr (2007). Incorporating bortezomib into upfront treatment for multiple myeloma: Early results of total therapy 3. *Br J Haematol 138*, 176–185.
62. Berenson JR, Yang HH, Sadler K, Jarutirasarn SG, Vescio RA, Mapes R, Purner M, Lee SP, Wilson J, Morrison B, Adams J, Schenkein D, Swift R (2006). Phase I/II trial assessing bortezomib and melphalan combination therapy for the treatment of patients with relapsed or refractory multiple myeloma. *J Clin Oncol 24*, 937–944.
63. Biehn SE, Moore DT, Voorhees PM, Garcia RA, Lehman MJ, Dees EC, Orlowski RZ (2007). Extended follow-up of outcome measures in multiple myeloma patients treated on a phase I study with bortezomib and pegylated liposomal doxorubicin. *Ann Hematol 86*, 211–216.
64. Canfield SE, Zhu K, Williams SA, McConkey DJ (2006). Bortezomib inhibits docetaxel-induced apoptosis via a p21-dependent mechanism in human prostate cancer cells. *Mol Cancer Ther 5*, 2043–2050.
65. Harousseau JL, Attal M, Leleu X, Troncy J, Pegourie B, Stoppa AM, Hulin C, Benboubker L, Fuzibet JG, Renaud M, Moreau P, Avet-Loiseau H (2006). Bortezomib plus dexamethasone as induction treatment prior to autologous stem cell transplantation in patients with newly diagnosed multiple myeloma: Results of an IFM phase II study. *Haematologica 91*, 1498–1505.
66. Jagannath S, Durie BG, Wolf J, Camacho E, Irwin D, Lutzky J, McKinley M, Gabayan E, Mazumder A, Schenkein D, Crowley J (2005). Bortezomib therapy alone and in combination with dexamethasone for previously untreated symptomatic multiple myeloma. *Br J Haematol 129*, 776–783.
67. Ma C, Mandrekar SJ, Alberts SR, Croghan GA, Jatoi A, Reid JM, Hanson LJ, Bruzek L, Tan AD, Pitot HC, Erlichman C, Wright JJ, Adjei AA (2007). A phase I and pharmacologic study of sequences of the proteasome inhibitor, bortezomib (PS-341, Velcade), in combination with paclitaxel and carboplatin in patients with advanced malignancies. *Cancer Chemother Pharmacol 59*, 207–215.
68. Mateos MV, Hernandez JM, Hernandez MT, Gutierrez NC, Palomera L, Fuertes M, Diaz-Mediavilla J, Lahuerta JJ, de la Rubia J, Terol MJ, Sureda A, Bargay J, Ribas P, de Arriba F, Alegre A, Oriol A, Carrera D, Garcia-Larana J, Garcia-Sanz R, Blade J, Prosper F, Mateo G, Esseltine DL, van de Velde H, San Miguel JF (2006). Bortezomib plus melphalan and prednisone in elderly untreated patients with multiple myeloma: Results of a multicenter phase 1/2 study. *Blood 108*, 2165–2172.
69. Messersmith WA, Baker SD, Lassiter L, Sullivan RA, Dinh K, Almuete VI, Wright JJ, Donehower RC, Carducci MA, Armstrong DK (2006). Phase I trial of bortezomib in combination with docetaxel in patients with advanced solid tumors. *Clin Cancer Res 12*, 1270–1275.
70. Orlowski RZ, Nagler A, Sonneveld P, Blade J, Hajek R, Spencer A, San Miguel J, Robak T, Dmoszynska A, Horvath N, Spicka I, Sutherland HJ, Suvorov AN, Zhuang SH, Parekh T, Xiu L, Yuan Z, Rackoff W, Harousseau JL (2007). Randomized phase III study of pegylated liposomal doxorubicin plus bortezomib compared with bortezomib alone in relapsed or refractory multiple myeloma: Combination therapy improves time to progression. *J Clin Oncol 25*, 3892–3901.
71. Orlowski RZ, Voorhees PM, Garcia RA, Hall MD, Kudrik FJ, Allred T, Johri AR, Jones PE, Ivanova A, Van Deventer HW, Gabriel DA, Shea TC, Mitchell BS, Adams J, Esseltine DL, Trehu EG, Green M, Lehman MJ, Natoli S, Collins JM, Lindley CM, Dees EC (2005). Phase 1 trial of the proteasome inhibitor bortezomib and pegylated liposomal doxorubicin in patients with advanced hematologic malignancies. *Blood 105*, 3058–3065.
72. Ryan DP, Appleman LJ, Lynch T, Supko JG, Fidias P, Clark JW, Fishman M, Zhu AX, Enzinger PC, Kashala O, Cusack J Jr, Eder JP (2006). Phase I clinical trial of bortezomib in combination with gemcitabine in patients with advanced solid tumors. *Cancer 107*, 2482–2489.
73. Ryan DP, O'Neil BH, Supko JG, Rocha Lima CM, Dees EC, Appleman LJ, Clark J, Fidias P, Orlowski RZ, Kashala

O, Eder JP, Cusack JC Jr (2006). A Phase I study of bortezomib plus irinotecan in patients with advanced solid tumors. *Cancer 107*, 2688–2697.

74. Voortman J, Smit EF, Honeywell R, Kuenen BC, Peters GJ, van de Velde H, Giaccone G (2007). A parallel dose-escalation study of weekly and twice-weekly bortezomib in combination with gemcitabine and cisplatin in the first-line treatment of patients with advanced solid tumors. *Clin Cancer Res 13*, 3642–3651.
75. Wang M, Giralt S, Delasalle K, Handy B, Alexanian R (2007). Bortezomib in combination with thalidomide-dexamethasone for previously untreated multiple myeloma. *Hematology 12*, 235–239.
76. Weigert O, Pastore A, Rieken M, Lang N, Hiddemann W, Dreyling M (2007). Sequence-dependent synergy of the proteasome inhibitor bortezomib and cytarabine in mantle cell lymphoma. *Leukemia 21*, 524–528.
77. Yu C, Friday BB, Lai JP, Yang L, Sarkaria J, Kay NE, Carter CA, Roberts LR, Kaufmann SH, Adjei AA (2006). Cytotoxic synergy between the multikinase inhibitor sorafenib and the proteasome inhibitor bortezomib in vitro: Induction of apoptosis through Akt and c-Jun NH2-terminal kinase pathways. *Mol Cancer Ther 5*, 2378–2387.
78. Papageorgiou A, Kamat A, Benedict WF, Dinney C, McConkey DJ (2006). Combination therapy with IFN-alpha plus bortezomib induces apoptosis and inhibits angiogenesis in human bladder cancer cells. *Mol Cancer Ther 5*, 3032–3041.
79. Khan T, Stauffer JK, Williams R, Hixon JA, Salcedo R, Lincoln E, Back TC, Powell D, Lockett S, Arnold AC, Sayers TJ, Wigginton JM (2006). Proteasome inhibition to maximize the apoptotic potential of cytokine therapy for murine neuroblastoma tumors. *J Immunol 176*, 6302–6312.
80. Lesinski GB, Raig ET, Guenterberg K, Brown L, Go MR, Shah NN, Lewis A, Quimper M, Hade E, Young G, Chaudhury AR, Ladner KJ, Guttridge DC, Bouchard P, Carson WE 3rd. (2008). HYPERLINK "http://www.ncbi.nlm.nih.gov/pubmed/18922907" IFN-alpha and bortezomib overcome Bcl-2 and Mcl-1 overexpression in melanoma cells by stimulating the extrinsic pathway of apoptosis. *Cancer Res. Oct 15*, *68*(20), 8351–8360.
81. Schumacher LY, Vo DD, Garban HJ, Comin-Anduix B, Owens SK, Dissette VB, Glaspy JA, McBride WH, Bonavida B, Economou JS, Ribas A (2006). Immunosensitization of tumor cells to dendritic cell-activated immune responses with the proteasome inhibitor bortezomib (PS-341, Velcade). *J Immunol 176*, 4757–4765.
82. Spisek R, Charalambous A, Mazumder A, Vesole DH, Jagannath S, Dhodapkar MV (2007). Bortezomib enhances dendritic cell (DC)-mediated induction of immunity to human myeloma via exposure of cell surface heat shock protein 90 on dying tumor cells: Therapeutic implications. *Blood 109*, 4839–4845.
83. Nencioni A, Garuti A, Schwarzenberg K, Cirmena G, Dal Bello G, Rocco I, Barbieri E, Brossart P, Patrone F, Ballestrero A (2006). Proteasome inhibitor-induced apoptosis in human monocyte-derived dendritic cells. *Eur J Immunol 36*, 681–689.
84. Nencioni A, Schwarzenberg K, Brauer KM, Schmidt SM, Ballestrero A, Grunebach F, Brossart P (2006). Proteasome inhibitor bortezomib modulates TLR4-induced dendritic cell activation. *Blood 108*, 551–558.
85. Straube C, Wehner R, Wendisch M, Bornhauser M, Bachmann M, Rieber EP, Schmitz M (2007). Bortezomib significantly impairs the immunostimulatory capacity of human myeloid blood dendritic cells. *Leukemia 21*, 1464–1471.
86. Berges C, Haberstock H, Fuchs D, Miltz M, Sadeghi M, Opelz G, Daniel V, Naujokat C (2008). Proteasome inhibition suppresses essential immune functions of human CD4(+) T cells. *Immunology 124*(2), 234–246.
87. Chauhan D, Catley L, Li G, Podar K, Hideshima T, Velankar M, Mitsiades C, Mitsiades N, Yasui H, Letai A, Ovaa H, Berkers C, Nicholson B, Chao TH, Neuteboom ST, Richardson P, Palladino MA, Anderson KC (2005). A novel orally active proteasome inhibitor induces apoptosis in multiple myeloma cells with mechanisms distinct from Bortezomib. *Cancer Cell 8*, 407–419.
88. Kuhn DJ, Chen Q, Voorhees PM, Strader JS, Shenk KD, Sun CM, Demo SD, Bennett MK, van Leeuwen FW, Chanan-Khan AA, Orlowski RZ (2007). Potent activity of carfilzomib, a novel, irreversible inhibitor of the ubiquitin-proteasome pathway, against preclinical models of multiple myeloma. *Blood 110*, 3281–3290.
89. Isaacs A, Lindenmann J (1957). Virus interference I: The interferon. *Proc R Soc Lond Ser B 147*, 258–267.
90. Isaacs A, Lindenmann J, Valentine RC (1957). Virus interference, II: Some properties of interferon. *Proc R Soc Lond 147B*, 268–273.
91. Burke DC, Isaacs A (1958). Some factors affecting the production of interferon. *Br J Exp Pathol 39*, 452–458.
92. Gresser I, Bourali C (1969). Exogenous interferon and inducers of interferon in the treatment Balb-c mice inoculated with RC19 tumour cells. *Nature 223*, 844–845.
93. Paucker K, Cantell K, Henle W (1962). Quantitative studies on viral interference in suspended L cells, III: Effect of interfering viruses and interferon on the growth rate of cells. *Virology 17*, 324–334.
94. Rubinstein M, Rubinstein S, Familletti PC, Gross MS, Miller RS, Waldman AA, Pestka S (1978). Human leukocyte interferon purified to homogeneity. *Science 202*, 1289–1290.
95. Taniguchi T, Fujii-Kuriyama Y, Muramatsu M (1979). Construction and identification of a bacterial plasmid containing the human fibroblast interferon gene sequences. *Proc Jpn Acad 55B*, 464–469.
96. Hall, SS (1997). The cloning of interferon and other mistakes. In: *A commotion in the blood*. Henry Holt and Company, Inc: New York, NY. p. 178–208.

97. Owerbach D, Rutter WJ, Shows TB, Gray P, Goeddel DV, Lawn RM (1981). Leukocyte and fibroblast interferon genes are located on human chromosome 9. *Proc Natl Acad Sci USA 78*, 3123–3127.
98. Allen G, Diaz MO (1996). Nomenclature of the human interferon proteins. *J Interferon Cytokine Res 16*, 181–184.
99. Conklin DC, Grant FJ, Rixon MW, Kindsvogel W (2002). Interferon-e. U.S. Patent 6329175.
100. LaFleur DW, Nardelli B, Tsareva T, Mather D, Feng P, Semenuk M, Taylor K, Buergin M, Chinchilla D, Roshke V, Chen G, Ruben SM, Pitha PM, Coleman TA, Moore PA (2001). Interferon-kappa, a novel type I interferon expressed in human keratinocytes. *J Biol Chem 276*, 39765–39771.
101. Hancock BW, Wheatley K, Harris S, Ives N, Harrison G, Horsman JM, Middleton MR, Thatcher N, Lorigan PC, Marsden JR, Burrows L, Gore M (2004). Adjuvant interferon in high-risk melanoma: The AIM HIGH Study—United Kingdom Coordinating Committee on Cancer Research randomized study of adjuvant low-dose extended-duration interferon Alfa-2a in high-risk resected malignant melanoma. *J Clin Oncol 22*, 53–61.
102. Kirkwood JM, Ibrahim JG, Sondak VK, Richards J, Flaherty LE, Ernstoff MS, Smith TJ, Rao U, Steele M, Blum RH (2000). High- and low-dose interferon alfa-2b in high-risk melanoma: First analysis of intergroup trial E1690/S9111/C9190. *J Clin Oncol 18*, 2444–2458.
103. Kirkwood JM, Ibrahim JG, Sosman JA, Sondak VK, Agarwala SS, Ernstoff MS, Rao U (2001). High-dose interferon alfa-2b significantly prolongs relapse-free and overall survival compared with the GM2-KLH/QS-21 vaccine in patients with resected stage IIB-III melanoma: Results of intergroup trial E1694/S9512/C509801. *J Clin Oncol 19*, 2370–2380.
104. Burchert A, Neubauer A (2005). Interferon alpha and T-cell responses in chronic myeloid leukemia. *Leuk Lymphoma 46*, 167–175.
105. Krown SE (1998). Interferon-alpha: Evolving therapy for AIDS-associated Kaposi's sarcoma. *J Interferon Cytokine Res 18*, 209–214.
106. Dorr RT, Salmon SE, Robertone A, Bonnem E (1988). Phase I-II trial of interferon-alpha 2b by continuous subcutaneous infusion over 28 days. *J Interferon Res 8*, 717–725.
107. Agarwala S (2003). Improving survival in patients with high-risk and metastatic melanoma: Immunotherapy leads the way. *Am J Clin Dermatol 4*, 333–346.
108. Kirkwood JM, Ernstoff MS, Davis CA, Reiss M, Ferraresi R, Rudnick SA (1985). Comparison of intramuscular and intravenous recombinant alpha-2 interferon in melanoma and other cancers. *Ann Intern Med 103*, 32–36.
109. Legha SS (1986). Interferons in the treatment of malignant melanoma. A review of recent trials. *Cancer 57*, 1675–1677.
110. Pyrhonen S, Salminen E, Ruutu M, Lehtonen T, Nurmi M, Tammela T, Juusela H, Rintala E, Hietanen P, Kellokumpu-Lehtinen PL (1999). Prospective randomized trial of interferon alfa-2a plus vinblastine versus vinblastine alone in patients with advanced renal cell cancer. *J Clin Oncol 17*, 2859–2867.
111. Steineck G, Strander H, Carbin BE, Borgstrom E, Wallin L, Achtnich U, Arvidsson A, Soderlund V, Naslund I, Esposti PL, et al. (1990). Recombinant leukocyte interferon alpha-2a and medroxyprogesterone in advanced renal cell carcinoma. A randomized trial. *Acta Oncol 29*, 155–162.
112. Kirkwood JM, Strawderman MH, Ernstoff MS, Smith TJ, Borden EC, Blum RH (1996). Interferon alfa-2b adjuvant therapy of high-risk resected cutaneous melanoma: The Eastern Cooperative Oncology Group Trial EST 1684. *J Clin Oncol 14*, 7–17.
113. Kirkwood J (2002). Cancer immunotherapy: The interferon-alpha experience. *Semin Oncol 29*, 18–26.
114. Schuchter LM (2004). Adjuvant interferon therapy for melanoma: High-dose, low-dose, no dose, which dose? *J Clin Oncol 22*, 7–10.
115. Lesinski GB, Kondadasula SV, Crespin T, Shen L, Kendra K, Walker MJ, Carson WE (2004). Multiparametric flow cytometric analysis of inter-patient variation in STAT1 phosphorylation following interferon alfa immunotherapy. *J Natl Cancer Inst 96*, 1331–1342.
116. Zimmerer JM, Lehman AM, Ruppert AS, Noble CW, Olencki T, Walker MJ, Kendra K, Carson WE 3rd (2008). IFN-alpha-2b-induced signal transduction and gene regulation in patient peripheral blood mononuclear cells is not enhanced by a dose increase from 5 to 10 megaunits/$m^2$. *Clin Cancer Res 14*, 1438–1445.
117. Levy DE, Gilliland DG (2000). Divergent roles of STAT1 and STAT5 in malignancy as revealed by gene disruptions in mice. *Oncogene 19*, 2505–2510.
118. Thyrell L, Erickson S, Zhivotovsky B, Pokrovskaja K, Sangfelt O, Castro J, Einhorn S, Grander D (2002). Mechanisms of Interferon-alpha induced apoptosis in malignant cells. *Oncogene 21*, 1251–1262.
119. Thyrell L, Hjortsberg L, Arulampalam V, Panaretakis T, Uhles S, Dagnell M, Zhivotovsky B, Leibiger I, Grander D, Pokrovskaja K (2004). Interferon alpha-induced apoptosis in tumor cells is mediated through the phosphoinositide 3-kinase/mammalian target of rapamycin signaling pathway. *J Biol Chem 279*, 24152–24162.
120. Dinney CP, Bielenberg DR, Perrotte P, Reich R, Eve BY, Bucana CD, Fidler IJ (1998). Inhibition of basic fibroblast growth factor expression, angiogenesis, and growth of human bladder carcinoma in mice by systemic interferon-alpha administration. *Cancer Res 58*, 808–814.
121. Ezekowitz RA, Mulliken JB, Folkman J (1992). Interferon alfa-2a therapy for life-threatening hemangiomas of infancy. *N Engl J Med 326*, 1456–1463.
122. Martin-Henao GA, Quiroga R, Sureda A, Gonzalez JR, Moreno V, Garcia J (2000). L-Selectin expression is low on CD34+ cells from patients with chronic myeloid leukemia and interferon-a up-regulates this expression. *Haematologica 85*, 139–146.

123. von Stamm U, Brocker EB, von Depka Prondzinski M, Ruiter DJ, Rumke P, Broding C, Carrel S, Lejeune FJ (1993). Effects of systemic interferon-alpha (IFN-alpha) on the antigenic phenotype of melanoma metastases. EORTC melanoma group cooperative study No. 18852. *Melanoma Res 3*, 173–180.

124. Belardelli F, Gresser I, Maury C, Duvillard P, Prade M, Maunoury MT (1983). Antitumor effects of interferon in mice injected with interferon-sensitive and interferon-resistant Friend leukemia cells. III. Inhibition of growth and necrosis of tumors implanted subcutaneously. *Int J Cancer 31*, 649–653.

125. Belardelli F, Gresser I, Maury C, Maunoury MT (1982). Antitumor effects of interferon in mice injected with interferon-sensitive and interferon-resistant Friend leukemia cells. II. Role of host mechanisms. *Int J Cancer 30*, 821–825.

126. Belardelli F, Gresser I, Maury C, Maunoury MT (1982). Antitumor effects of interferon in mice injected with interferon-sensitive and interferon-resistant Friend leukemia cells. I. *Int J Cancer 30*, 813–820.

127. Gresser I (1961). Production of inteferon by suspensions of human leukocytes. *Proc Soc Exp Biol Med 108*, 799–803.

128. Gresser I, Belardelli F, Maury C, Tovey MG, Maunoury MT (1986). Anti-tumor effects of interferon in mice injected with interferon-sensitive and interferon-resistant friend leukemia cells. IV. Definition of optimal treatment regimens. *Int J Cancer 38*, 251–257.

129. Gresser I, Carnaud C, Maury C, Sala A, Eid P, Woodrow D, Maunoury MT, Belardelli F (1991). Host humoral and cellular immune mechanisms in the continued suppression of Friend erythroleukemia metastases after interferon alpha/beta treatment in mice. *J Exp Med 173*, 1193-1203.

130. Gresser I, Maury C, Belardelli F, Maunoury MT, Machover D (1988). Effectiveness of mouse interferon alpha/beta compared to single-agent chemotherapy in increasing survival time of mice after intravenous inoculation of Friend erythroleukemia cells. *J Natl Cancer Inst 80*, 126–131.

131. Gresser I, Maury C, Carnaud C, De Maeyer E, Maunoury MT, Belardelli F (1990). Anti-tumor effects of interferon in mice injected with interferon-sensitive and interferon-resistant Friend erythroleukemia cells. VIII. Role of the immune system in the inhibition of visceral metastases. *Int J Cancer 46*, 468–474.

132. Badgwell B, Lesinski GB, Magro C, Abood G, Skaf A, Carson WE (2003). The anti-tumor effects of interferon-alpha are maintained in mice challenged with a STAT1-deficient murine melanoma cell line. *J Surg Res 116*, 129–136.

133. Lesinski GB, Anghelina M, Zimmerer J, Bakalakos T, Badgwell B, Parihar R, Hu Y, Becknell B, Abood G, RayChaudhury A, Magro C, Durbin J, Carson WE (2003). The anti-tumor effects of interferon-alpha are abrogated in a STAT1-deficient mouse. *J Clin Invest 112*, 170–180.

134. Zimmerer JM, Lesinski GB, Kondadasula SV, Karpa VI, Lehman A, Raychaudhury A, Becknell B, Carson WE 3rd (2007). IFN-alpha-induced signal transduction, gene expression, and antitumor activity of immune effector cells are negatively regulated by suppressor of cytokine signaling proteins. *J Immunol 178*, 4832–4845.

135. Moschos SJ, Edington HD, Land SR, Rao UN, Jukic D, Shipe-Spotloe J, Kirkwood JM (2006). Neoadjuvant treatment of regional stage IIIB melanoma with high-dose interferon alfa-2b induces objective tumor regression in association with modulation of tumor infiltrating host cellular immune responses. *J Clin Oncol 24*, 3164–3171.

136. Gogas H, Ioannovich J, Dafni U, Stavropoulou-Giokas C, Frangia K, Tsoutsos D, Panagiotou P, Polyzos A, Papadopoulos O, Stratigos A, Markopoulos C, Bafaloukos D, Pectasides D, Fountzilas G, Kirkwood JM (2006). Prognostic significance of autoimmunity during treatment of melanoma with interferon. *N Engl J Med 354*, 709–718.

137. Atkins MB (2006). Cytokine-based therapy and biochemotherapy for advanced melanoma. *Clin Cancer Res 12*, 2353s–2358s.

138. Melichar B, Koralewski P, Ravaud A, Pluzanska A, Bracarda S, Szczylik C, Chevreau C, Filipek M, Delva R, Sevin E, Négrier S, McKendrick J, Santoro A, Pisa P, Escudier B (1008). HYPERLINK "http://www.ncbi.nlm.nih.gov/pubmed/18408224" First-line bevacizumab combined with reduced dose interferon-alpha2a is active in patients with metastatic renal cell carcinoma. *Ann Oncol 2008 Aug*, *19*(8), 1470–1476.

139. Varker KA, Biber JE, Kefauver C, Jensen R, Lehman A, Young D, Wu H, Lesinski GB, Kendra K, Chen HX, Walker MJ, Carson WE 3rd (2007). A randomized phase 2 trial of bevacizumab with or without daily low-dose interferon alfa-2b in metastatic malignant melanoma. *Ann Surg Oncol 14*, 2367–2376.

140. Bukowski RM, Tendler C, Cutler D, Rose E, Laughlin MM, Statkevich P (2002). Treating cancer with PEG Intron: Pharmacokinetic profile and dosing guidelines for an improved interferon-alpha-2b formulation. *Cancer 95*, 389–396.

141. Liu TC, Galanis E, Kirn D (2007). Clinical trial results with oncolytic virotherapy: A century of promise, a decade of progress. *Nat Clin Pract Oncol 4*, 101–117.

142. Martuza RL, Malick A, Markert JM, Ruffner KL, Coen DM (1991). Experimental therapy of human glioma by means of a genetically engineered virus mutant. *Science 252*, 854–856.

143. Goldstein DJ, Weller SK (1988). Herpes simplex virus type 1-induced ribonucleotide reductase activity is dispensable for virus growth and DNA synthesis: Isolation and characterization of an ICP6 lacZ insertion mutant. *J Virol 62*, 196–205.

144. Kasuya H, Nishiyama Y, Nomoto S, Hosono J, Takeda S, Nakao A (1999). Intraperitoneal delivery of hrR3 and ganciclovir prolongs survival in mice with disseminated pancreatic cancer. *J Surg Oncol 72*, 136–141.

145. Mineta T, Rabkin SD, Martuza RL (1994). Treatment of malignant gliomas using ganciclovir-hypersensitive, ribonucleotide reductase-deficient herpes simplex viral mutant. *Cancer Res 54*, 3963–3966.

146. Boviatsis EJ, Park JS, Sena-Esteves M, Kramm CM, Chase M, Efird JT, Wei MX, Breakefield XO, Chiocca EA (1994). Long-term survival of rats harboring brain neoplasms treated with ganciclovir and a herpes simplex virus vector that retains an intact thymidine kinase gene. *Cancer Res 54*, 5745–5751.

147. Aghi M, Visted T, Depinho RA, Chiocca EA (2008). Oncolytic herpes virus with defective ICP6 specifically replicates in quiescent cells with homozygous genetic mutations in p16. *Oncogene 27*(30), 4249–4254.

148. Chou J, Kern ER, Whitley RJ, Roizman B (1990). Mapping of herpes simplex virus-1 neurovirulence to gamma 134.5, a gene nonessential for growth in culture. *Science 250*, 1262–1266.

149. Bolovan CA, Sawtell NM, Thompson RL (1994). ICP34.5 mutants of herpes simplex virus type 1 strain 17syn+ are attenuated for neurovirulence in mice and for replication in confluent primary mouse embryo cell cultures. *J Virol 68*, 48–55.

150. McKie EA, MacLean AR, Lewis AD, Cruickshank G, Rampling R, Barnett SC, Kennedy PG, Brown SM (1996). Selective in vitro replication of herpes simplex virus type 1 (HSV-1) ICP34.5 null mutants in primary human CNS tumours—Evaluation of a potentially effective clinical therapy. *Br J Cancer 74*, 745–752.

151. Lasner TM, Kesari S, Brown SM, Lee VM, Fraser NW, Trojanowski JQ (1996). Therapy of a murine model of pediatric brain tumors using a herpes simplex virus type-1 ICP34.5 mutant and demonstration of viral replication within the CNS. *J Neuropathol Exp Neurol 55*, 1259–1269.

152. Kesari S, Randazzo BP, Valyi-Nagy T, Huang QS, Brown SM, MacLean AR, Lee VM, Trojanowski JQ, Fraser NW (1995). Therapy of experimental human brain tumors using a neuroattenuated herpes simplex virus mutant. *Lab Invest 73*, 636–648.

153. Rampling R, Cruickshank G, Papanastassiou V, Nicoll J, Hadley D, Brennan D, Petty R, MacLean A, Harland J, McKie E, Mabbs R, Brown M (2000). Toxicity evaluation of replication-competent herpes simplex virus (ICP 34.5 null mutant 1716) in patients with recurrent malignant glioma. *Gene Ther 7*, 859–866.

154. Harland J, Papanastassiou V, Brown SM (2002). HSV1716 persistence in primary human glioma cells in vitro. *Gene Ther 9*, 1194–1198.

155. Papanastassiou V, Rampling R, Fraser M, Petty R, Hadley D, Nicoll J, Harland J, Mabbs R, Brown M (2002). The potential for efficacy of the modified (ICP 34.5(-)) herpes simplex virus HSV1716 following intratumoural injection into human malignant glioma: A proof of principle study. *Gene Ther 9*, 398–406.

156. Mineta T, Rabkin SD, Yazaki T, Hunter WD, Martuza RL (1995). Attenuated multi-mutated herpes simplex virus-1 for the treatment of malignant gliomas. *Nat Med 1*, 938–943.

157. Warren RS, Yuan H, Matli MR, Gillett NA, Ferrara N (1995). Regulation by vascular endothelial growth factor of human colon cancer tumorigenesis in a mouse model of experimental liver metastasis. *J Clin Invest 95*, 1789–1797.

158. Melnyk O, Zimmerman M, Kim KJ, Shuman M (1999). Neutralizing anti-vascular endothelial growth factor antibody inhibits further growth of established prostate cancer and metastases in a pre-clinical model. *J Urol 161*, 960–963.

159. Asano M, Yukita A, Matsumoto T, Kondo S, Suzuki H (1995). Inhibition of tumor growth and metastasis by an immunoneutralizing monoclonal antibody to human vascular endothelial growth factor/vascular permeability factor 121. *Cancer Res 55*, 5296–5301.

160. Millauer B, Shawver LK, Plate KH, Risau W, Ullrich A (1994). Glioblastoma growth inhibited in vivo by a dominant-negative Flk-1 mutant. *Nature 367*, 576–579.

161. Wood JM, Bold G, Buchdunger E, Cozens R, Ferrari S, Frei J, Hofmann F, Mestan J, Mett H, O'Reilly T, Persohn E, Rosel J, Schnell C, Stover D, Theuer A, Towbin H, Wenger F, Woods-Cook K, Menrad A, Siemeister G, Schirner M, Thierauch KH, Schneider MR, Drevs J, Martiny-Baron G, Totzke F (2000). PTK787/ZK 222584, a novel and potent inhibitor of vascular endothelial growth factor receptor tyrosine kinases, impairs vascular endothelial growth factor-induced responses and tumor growth after oral administration. *Cancer Res 60*, 2178–2189.

162. Wedge SR, Ogilvie DJ, Dukes M, Kendrew J, Curwen JO, Hennequin LF, Thomas AP, Stokes ES, Curry B, Richmond GH, Wadsworth PF (2000). ZD4190: An orally active inhibitor of vascular endothelial growth factor signaling with broad-spectrum antitumor efficacy. *Cancer Res 60*, 970–975.

163. Wedge SR, Kendrew J, Hennequin LF, Valentine PJ, Barry ST, Brave SR, Smith NR, James NH, Dukes M, Curwen JO, Chester R, Jackson JA, Boffey SJ, Kilburn LL, Barnett S, Richmond GH, Wadsworth PF, Walker M, Bigley AL, Taylor ST, Cooper L, Beck S, Jurgensmeier JM, Ogilvie DJ (2005). AZD2171: A highly potent, orally bioavailable, vascular endothelial growth factor receptor-2 tyrosine kinase inhibitor for the treatment of cancer. *Cancer Res 65*, 4389–4400.

164. Prewett M, Huber J, Li Y, Santiago A, O'Connor W, King K, Overholser J, Hooper A, Pytowski B, Witte L, Bohlen P, Hicklin DJ (1999). Antivascular endothelial growth factor receptor (fetal liver kinase 1) monoclonal antibody inhibits tumor angiogenesis and growth of several mouse and human tumors. *Cancer Res 59*, 5209–5218.

165. Fox WD, Higgins B, Maiese KM, Drobnjak M, Cordon-Cardo C, Scher HI, Agus DB (2002). Antibody to vascular endothelial growth factor slows growth of an androgen-independent xenograft model of prostate cancer. *Clin Cancer Res 8*, 3226–3231.
166. Segerstrom L, Fuchs D, Backman U, Holmquist K, Christofferson R, Azarbayjani F (2006). The anti-VEGF antibody bevacizumab potently reduces the growth rate of high-risk neuroblastoma xenografts. *Pediatr Res 60*, 576–581.
167. Inoue S, Hartman A, Branch CD, Bucana CD, Bekele BN, Stephens LC, Chada S, Ramesh R (2007). mda-7 In combination with bevacizumab treatment produces a synergistic and complete inhibitory effect on lung tumor xenograft. *Mol Ther 15*, 287–294.
168. Guse K, Ranki T, Ala-Opas M, Bono P, Sarkioja M, Rajecki M, Kanerva A, Hakkarainen T, Hemminki A (2007). Treatment of metastatic renal cancer with capsid-modified oncolytic adenoviruses. *Mol Cancer Ther 6*, 2728–2736.
169. Salvador C, Li B, Hansen R, Cramer DE, Kong M, Yan J (2008). Yeast-derived beta-glucan augments the therapeutic efficacy mediated by anti-vascular endothelial growth factor monoclonal antibody in human carcinoma xenograft models. *Clin Cancer Res 14*, 1239–1247.
170. Pineda-Roman M, Zangari M, van Rhee F, Anaissie E, Szymonifka J, Hoering A, Petty N, Crowley J, Shaughnessy J, Epstein J, Barlogie B (2008). VTD combination therapy with bortezomib-thalidomide-dexamethasone is highly effective in advanced and refractory multiple myeloma. *Leukemia 22*(7), 1419–1427.
171. Ling YH, Liebes L, Zou Y, Perez-Soler R (2003). Reactive oxygen species generation and mitochondrial dysfunction in the apoptotic response to Bortezomib, a novel proteasome inhibitor, in human H460 non-small cell lung cancer cells. *J Biol Chem 278*, 33714–33723.

# DEVELOPMENT OF COUNTERMEASURES FOR BIOTERRORISM IN THE UNITED STATES

Gregory H. Levine and Jeffrey L. Handwerker
*Ropes and Gray LLP, Washington, D.C.*

## 1 INTRODUCTION

Shortly after the terrorist attacks of September 11, 2001, 22 people contracted anthrax infections from *Bacillus anthracis* spores deliberately disseminated through the U.S. postal system. Five of them died [1]. The anthrax killings, together with terrorist groups' demonstrated intention to inflict catastrophic casualties on civilian populations, focused the attention of the U.S. public, lawmakers, and government officials on the acute dangers of bioterrorism.

Bioterrorism presents unique threats to the security of the United States and other nations. Attacks involving contagious biological weapons, in particular, pose a risk of widespread disease outbreak. Such attacks can mimic naturally occurring disease, potentially delaying recognition and response [2].* In addition, once a terrorist has the capability of producing a biological weapon, there are few if any limits on the quantity of that weapon the terrorist can produce. In the right environment, a pathogenic microbe can reproduce on its own. This characteristic distinguishes biological weapons from other weapons that terrorists might use to cause mass casualties, such as chemical, nuclear, and radiological weapons, which are limited by the availability and cost of critical production materials and equipment [3].

Biological weapons also present unique challenges due to their ability to evade countermeasures, either through natural alteration or genetic manipulation. Recent advances in biotechnology and the life sciences, including the spread of expertise needed to create modified or novel organisms, have the potential to expand the spectrum of biological agents that might be used as weapons (HSPD 10). In 2000, for example, Australian scientists inadvertently discovered that introducing the interleukin-4 gene into an *ectromelia* virus (the cause of mousepox) suppressed primary antiviral immune responses in recently immunized genetically resistant mice [4]. Others have shown that it is possible to create strains of *B. anthracis* that are resistant to powerful antibiotics, including fluoroquinolones and macrolides [5]. These and other findings suggest that the bioterror threat will require increasingly sophisticated responses over time.

This chapter focuses on one element of biodefense—the development of medical countermeasures to be used in advance of, or in response to, a bioterrorist attack. Unlike other areas of medicine, for which private market forces are generally sufficient to spur private industry to develop and market useful products, there are few market incentives to develop and market vaccines, diagnostics, and therapeutics for biodefense purposes. To the contrary, the disincentives to such investment in the absence of government intervention far exceed the incentives. First, the market for biodefense products is unpredictable. Even assuming that a manufacturer could find a buyer for a biodefense countermeasure, the product may be stockpiled but never used, and the demand for the product ultimately will

*Homeland Security Presidential Directive 10 will be hereafter appear as HSPD 10.

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

depend on whether biological attacks occur or are expected to occur. The market demand is, therefore, not amenable to the return-on-investment analysis upon which pharmaceutical and biotechnology companies typically make business decisions [6]. Second, the market is not likely to offer attractive profit margins. Biotechnology and pharmaceutical companies anticipate operating margins far higher than those achieved by traditional defense contractors [6]. Third, industry has perceived biodefense products as posing disproportionate product liability risks, largely as a consequence of the unfamiliarity of the exposure and the manufacturer's lack of control over how government agencies deploy the countermeasure [7].

Since 9/11, Congress has enacted various federal laws to address the insufficiency of traditional market forces and to foster private industry's involvement in developing biodefense countermeasures, as well as countermeasures for naturally occurring pathogens that also can be used for bioterror ends. These laws focus on creating so-called push-and-pull incentives, with countermeasure-related research grants (primarily through the National Institutes of Health) providing the push with research grants and regulatory streamlining and countermeasure procurement programs providing the pull [8]. Most notably for pull mechanisms, the Project Bioshield Act of 2004 created a $5.6 billion fund for the purchase of biodefense countermeasures. Along with these incentives, various measures have been adopted to reduce the regulatory burdens on companies willing to invest in the biodefense space.

The majority of this chapter discusses the laws, regulations, and actions taken to date in the United States to spur the development of diagnostic, prophylactic, and therapeutic agents for biodefense. Initial efforts to encourage the development of such products have been slow to bear fruit and have been beset by well-publicized setbacks, such as the cancellation of a $877 million Project Bioshield contract with VaxGen Corporation to produce a new-generation anthrax vaccine. These setbacks have led to calls in some circles for further action to spur research and development in the biodefense field.

Although the measures described in this chapter generally reflect regulatory retrenchment intended to encourage research and development activities, the U.S. government also has concerns about research materials with potential bioterror uses falling into terrorists hands, and in particular the risk that more widespread participation in biodefense research will increase the opportunities for such diversion. Consequently, a number of laws and regulations intended to prevent such occurrences have been adopted or tightened since 9/11. Because any person entering into research or development in this field must be aware of these regulatory restrictions, the last section of this chapter briefly describes the most important of these restrictions, including select agent restrictions, biosafety laboratories, and export controls.

## 2 INCENTIVES TO CONDUCT RESEARCH AND DEVELOPMENT (R&D) OF BIOTERRORISM COUNTERMEASURES

Increased research funding, primarily through the National Institute of Allergy and Infectious Disease (NIAID) within the National Institutes of Health (NIH) is the primary push tool the U.S. government has employed in an attempt to encourage biodefense R&D. The government has also attempted to develop a pull incentive by streamlining and accelerating the regulatory pathway for biodefense products. This section describes the efforts of Congress and the executive branch of the government in these areas.

### 2.1 Research Incentives

Congress responded to the 9/11 and anthrax attacks by increasing funding earmarked for civilian biodefense research. From a negligible amount in fiscal year (FY) 2001, Congress appropriated $3.2 billion in FY 2002, and steadily increased the amount through FY 2005, when biodefense research spending peaked at $7.6 billion. Though funding has declined since FY 2005, it remains substantial and well ahead of pre-2001 levels [9]. The effect has been the dedication of billions of federal dollars allocated to a research enterprise that was essentially nonexistent before 2001.

The lead grant-making body for civilian biodefense is the NIAID. The NIAID's biodefense research priorities, which cover basic and translational research to foster the development of diagnostics, therapeutics, and vaccines, are based largely on categories established by the Centers for Disease Control and Prevention (CDC), as modified by NIAID [10]. The CDC's list of biologic agents is prioritized into three categories (categories A, B, and C, with category A representing the highest priority threats) based on their potential public health impact, potential for dissemination, and special needs for effective public health intervention. Examples of CDC categories A, B, and C agents are shown below listed in Table 1.

NIAID produces a regularly updated Funding Opportunities List that lists Requests for Applications (RFAs), which are NIAID-sponsored initiatives to stimulate targeted research by requesting grant applications in a well-defined scientific area [11]. NIAID also pub-

**TABLE 1 CDC Pathogen Categories**

| Category A | Category B | Category C |
|---|---|---|
| Anthrax (*B. anthracis*)<br>Botulism (*Clostridium botulinum* toxin)<br>Plague (*Y. pestis*)<br>Smallpox (variola major)<br>Tularemia (*Francisella tularensis*)<br>Viral hemorrhagic fevers (e.g., Ebola, Marburg) | Ricin toxin<br>Staphylococcal enterotoxin B<br>Typhus fever<br>Viral encephalitis<br>Food safety threats (e.g., *Salmonella*, *Escherichia coli*, *Shigella*)<br>Water safety threats (e.g., *Vibrio cholerae*) | Emerging infectious diseases such as Nipah virus and hantavirus |

lishes Requests for Proposals (RFPs), which are initiatives sponsored by NIAID that request proposals for a contract to meet a specific need, such as the development of an animal model [12].

All applications for funding must either have direct approval of the director of NIAID or be in response to an RFA or RFP [13]. Requirements for funding applications vary according to the RFP or RFA. As an example, the RFP for biodefense vaccines for categories A and B pathogens requires a demonstration of safety and proof-of-concept efficacy in an animal model and a well-defined and feasible product development plan for advancing the vaccine candidate to achieve certain specified objectives [14].

Funding applications are reviewed and accepted within about 6 months of receipt of the application. Peer reviewers are assigned applications within the first 6 weeks. The applications are given a priority score, with some receiving an expedited second level of review and others deferred for another few months. Once a second review is accepted, NIAID's main advisory council must approve funding. After approval, funding is paid out within 1 or 2 months. Even for the deferred applications, the approval process takes less than a year [15].

Outside of NIH, funding may be available through the Department of Defense (DoD) and other federal agencies. For example, in FY 2006, the DoD's Chemical and Biological Defense Program budget dedicated to basic and applied research was more than $332 million [16]. These agencies issue solicitations for projects on the federal business opportunities website (http://www.fedbizopps.gov).

## 2.2 Regulatory Pathway Incentives

Regulatory barriers present unique challenges in the biodefense field. Classic scientific techniques for drug development, which normally involve large clinical studies of a drug's effectiveness in patients with the target disease, will not usually be ethical or feasible for countermeasures to bioweapons. Nor does the typical 10-year pathway to develop a new drug in response to a pathogen fit well in a world characterized by rapidly mutating and possibly intentionally engineered threats [17].

Lawmakers and regulators have undertaken a number of initiatives to address some of these burdens, primarily focusing on regulation of new products by the Food and Drug Administration (FDA). FDA's dual mission of protecting the public health by ensuring the safety and effectiveness of medical products, while simultaneously promoting the public health by "promptly and efficiently" reviewing clinical research and acting upon marketing applications, routinely confronts the agency with difficult, high-stakes challenges [18]. If the agency permits the marketing of a product that later turns out to be unsafe, it is criticized as being too lax. If the agency delays or blocks the marketing of a product by demanding adherence to a high standard of scientific proof, the agency is often criticized as depriving patients of the ability to avail themselves of potentially beneficial therapies. This dilemma is even more stark in the area of biodefense, where it is often difficult or impossible to develop the type of clinical evidence the agency would normally require. This section describes measures FDA has taken to address this challenge, as well as other steps it has taken to encourage the development of medical countermeasures.

***Animal Rule and Development of Animal Models***

Generally, FDA requires manufacturers of new drugs and biologics to establish the safety and effectiveness of their products in one or more pivotal clinical trials. For biodefense countermeasures, however, such studies may be impossible because one cannot ethically expose healthy research subjects to potentially lethal bioterrorist agents. The challenge, therefore, is to develop a standard of proof that sufficiently protects the public health while still permitting approval of important biodefense countermeasures.

To its credit, FDA began considering this issue several years before 2001. In July 1997, the agency published a Request for Comments in which it asked what evidence of efficacy, other than that from human trials, would be appropriate to demonstrate the safety and efficacy of products that may provide protection against toxic chemical, biological, radiological, or nuclear substances [19].* After receiving responses, on October 5, 1999, FDA published a proposed rule under which it would

*Reference 19 will hereafter appear as # Proposed Rule.

grant marketing approval for a new drug or biological product on the basis of adequate and well-controlled trials in animals when it is "scientifically reasonable to expect that the effect of the drug or biological product in animals is reasonably likely to predict clinical benefit in humans" [19, at 53964]. This rule would apply only to drug or biological products for which "[d]efinitive human efficacy studies cannot be conducted because it would be unethical to deliberately expose healthy human volunteers to a lethal or permanently disabling toxic biological, chemical, radiological, or nuclear substance without a proven treatment; and field trials to study the product's efficacy after an accidental or hostile exposure are not feasible" [19, at 53968].

Further, under the proposed rule of October 1999, FDA would rely on evidence from studies in animals only where:

- There is a reasonably well understood pathophysiological mechanism of the toxicity of the substance and its prevention or substantial reduction by the product.
- The effect is independently substantiated in multiple animal species.
- The animal study endpoint is clearly related to the desired benefit in humans.
- The data or information on the kinetics and pharmacodynamics of the product or other relevant data or information allows selection of an effective dose in humans [19, at 53968].

The proposed rule also outlined requirements for post-marketing studies, approval with restrictions to assure safe use, and information necessary to distribute to patients and potential patients [19, at 53968]. In addition, the proposed rule allowed for expedited withdrawal of the approval if the product was later shown to be unsafe, to have no clinical benefit under its intended conditions of use, or if its safe use cannot be ensured by restricted distribution [19, at 53970].

At the time of the anthrax attacks in the fall of 2001, FDA had not issued the final animal rule. Consequently, in one of the first significant post-9/11 laws, commonly known as the Bioterrorism Act of 2002, Congress instructed the Secretary of Health and Human Services (HHS) to complete the process of rulemaking that was commenced with the October 1999 proposed rule by promulgating a final rule [20].*

FDA issued the final rule in May 2002 [21].† While the proposed rule would have required such substantiation of a product's effectiveness in more than one animal species, the final rule provides that the effect must be demonstrated in more than one animal species "unless the effect is demonstrated in a single animal species that represents a sufficiently well-characterized animal model for predicting the response in humans" [22]. FDA implemented this change because the multiple-species requirement may be inappropriate or unnecessary in certain situations [23]. One such circumstance is when only one species is capable of reacting with a response predictive for humans, such as when there is only one nonhuman host for the targeted microorganism [23]. Another may be when particular species are specifically well recognized as predictors of effectiveness in humans [23]. Therefore, single-species efficacy studies will generally be limited to situations where the study model is sufficiently well recognized so as to render studies in multiple species unnecessary. It is anticipated that single-species efficacy studies will most often be done for anti-infective drug products because the pathophysiological mechanisms of infectious diseases are usually well understood, and animal models for infectious diseases have been studied for years and are well characterized [23].

A product approved through the animal rule must be tested for safety in human clinical trials in accordance with existing requirements. The animal rule relaxes the normal evidentiary burden only with regard to effectiveness, for the simple reason that safety studies, unlike effectiveness studies, require exposure only to the experimental product, not to the toxic agent the countermeasure is intended to counteract.

FDA issued its first approval under the so-called Animal Efficacy Rule in February 2003. The approved drug, pyridostigmine bromide, increases survival after exposure to Soman nerve gas poisoning [24]. Despite its limited use to date, the Animal Efficacy Rule has potential applicability to a broad variety of agents, including anthrax, smallpox, botulism, plague, tularemia, and others. At the time of this writing, trade press reports indicated that one company was planning to seek FDA approval under the animal rule for its novel respiratory antibiotic shown to be protective against a lethal dose of inhalation anthrax when tested in primates [25].

To encourage the development of countermeasures under the animal rule, FDA has led or participated in efforts to foster discussion and consensus on appropriate animal models for effectiveness testing. Even before the animal rule was finalized, FDA and NIH hosted a workshop titled Anthrax Vaccine: Efficacy Testing and Surrogate Markers of Immunity [26]. That workshop focused on the substantial effect of the disease in multiple animal species, the degree of similarity to the human disease, and whether the immune response in the animal resembles the human immune response [26].

*Reference 20 will hereafter appear as Bioterrorism Act.
†Reference 21 will hereafter appear as Final Rule.

In October 2004, FDA's Center for Biologics and Evaluation, the NIAID, and the HHS Office of Research Development and Coordination convened a public workshop on Animal Models and Correlates of Protection for Plague Vaccines [27]. One purpose of this session was to discuss human disease and relevant animal models, including what constitutes good animal models for plague and whether there is sufficient understanding of the pathophysiology of the plague to qualify it under the animal rule. The FDA also announced that it would hold a workshop in November 2007 titled Anthrax Vaccines; Bridging Correlates of Protection in Animals to Immunogenicity in Humans to discuss possible strategies for bridging animal efficacy data to human immunogenicity data for investigational anthrax vaccines [28]. Despite these early efforts, much more scientific work needs to be done toward the development of acceptable animal models before the animal rule can be utilized to its full potential.

***Emergency Use Authorization*** Recognizing that approved countermeasures might not be available at the time of a terrorist attack, Congress included in the Project Bioshield Act of 2004 a new streamlined Emergency Use Authorization (EUA) process by which FDA can authorize the distribution of bioterrorism countermeasures on an emergency basis [29]. Under the EUA authority, FDA may authorize the use of an unapproved medical product or an unapproved use of an approved medical product during a "declared emergency involving a heightened risk of attack on the public or U.S. military forces" [30].

In the absence of such an emergency authorization procedure, a drug or biological product that has potential utility as a countermeasure, but that is not approved for such a use, could be distributed from a stockpile for countermeasure use only pursuant to an Investigational New Drug (IND) application. FDA's IND regulations, however, require ongoing institutional review board oversight over clinical research, detailed research protocols, signed informed consent of all study subjects, reporting of trial-related adverse events within specified time frames, and the like (21 CFR Part 312). Such requirements are generally incompatible with the distribution of a countermeasure in an emergency, as the military learned in trying to distribute anthrax vaccine to the armed forces under an IND at the start of the Iraq war [31]. By substituting for full marketing approval, the EUA permits the distribution of an unapproved countermeasure without complying with the IND regulations.

Before an EUA may be issued, the Secretary of HHS must declare an emergency. Such an emergency declaration must be based on: (1) a determination by the Secretary of Homeland Security that there is a domestic emergency, or the potential for a domestic emergency, involving a heightened risk of attack with a specified biological, chemical, radiological, or nuclear agent or agents; (2) a determination by the Secretary of Defense that there is a military emergency, or a significant potential for a military emergency, involving a heightened risk to U.S. military forces of attack with a specified biological, chemical, radiological, or nuclear agent or agents; or (3) a determination by the Secretary of HHS of a public health emergency that affects or may affect national security and that involves a specified biological, chemical, radiological, or nuclear agent or agents [32]. Once a declaration is in effect, the Secretary of HHS may issue an EUA if, after consulting with the directors the NIH and the CDC, the Secretary concludes that:

- The agent in question can in fact cause a serious or life-threatening disease or condition.
- It is reasonable to believe that the product for which emergency use is requested may be effective in diagnosing, treating, or preventing the disease or condition.
- The known and potential benefits of the product outweigh the known and potential risks of the product.
- There is no adequate, approved, and available alternative to the product [21 USC § 360bbb-3(c)].

An EUA terminates either when the circumstances necessitating the emergency cease to exist or one year from the date of the authorization, whichever is earlier [21 USC § 360bbb-3(b)(2)]. The Secretary may, at his or her discretion, renew the authorization. The Secretary also must promptly publish a notice of each authorization and termination or revocation of each authorization, along with an explanation of reasons for the action, in the *Federal Register* [21 USC § 360bbb-3(h)].

The first and, to date, only use of EUA authority was in 2005, involving Anthrax Vaccine Adsorbed (AVA) for the prevention of inhalation anthrax in military personnel.* Although AVA had been licensed since 1970, it was not originally contemplated as a biowarfare or bioterrorism countermeasure, and some members of the Armed Forces objected to the DoD mandatory Anthrax Vaccine Immunization Program (AVIP). In March 2003, six plaintiffs filed suit in the U.S. District Court for the

*Although FDA has worked with CDC to identify potential medical countermeasures in the Strategic National Stockpile as candidates for EUA for an unapproved use, there have been no authorizations apart from AVA.

District of Columbia seeking an injunction against AVIP and a declaration of AVA as an investigational drug when used for protection against inhalation anthrax [33]. The court declared that using AVA for inhalation anthrax constituted an unapproved use of an approved product and required a public notice and comment period before FDA could issue a determination that the AVA license included use for prevention of inhalation anthrax [34]. The DoD and FDA responded by issuing an EUA for the use of AVA to prevent inhalation anthrax for individuals between 18 and 65 years of age deemed by the DoD to be at heightened risk of exposure due to attack with anthrax (70 *Fed Reg* 5453). Not surprisingly, this issuance of the EUA for AVA proved controversial on Capitol Hill [35]. Before the issue could ultimately be resolved, however, FDA issued a final notice finding AVA safe for prevention of inhalation anthrax, thus eliminating the need for an EUA [36].

***Priority Review, Accelerated Approval, and Fast Track*** Beginning in the 1980s, in response to demands of acquired immunodeficiency syndrome (AIDS) and cancer patients, FDA began developing programs intended to accelerate access to drugs and medical devices intended to treat serious and life-threatening diseases. In the drug area, current programs that are potentially available to developers of biodefense products include:

- Priority Review, an FDA designation for drugs and biologicals that offer a major advance in current treatment or provide a treatment where no satisfactory therapy currently exists. Priority Review moves a product's application for marketing approval up in the FDA review queue and subjects it to a shorter (6-month) target review timeline under the Prescription Drug User Fee Act (PDUFA) [37].
- Accelerated Approval of drugs and biologicals intended to treat serious or life-threatening illnesses that provide meaningful therapeutic benefit to patients over existing treatments. FDA Accelerated Approval regulation—often referred to as the Subpart H regulation—provides for approval based on a surrogate endpoint of clinical benefit (21 CFR § 314.500).
- Fast Track treatment for new drugs intended to treat serious or life-threatening conditions and that demonstrate the potential to address unmet medical needs. FDA's Fast Track program, in addition to Priority Review and Accelerated Approval, include FDA review of portions of a new drug application before the entire application is complete, as well as close consultation and collaboration with the agency throughout the development process [38].

In the Bioterrorism Act of 2002, Congress expressly authorized the Secretary of HHS to designate certain bioterrorism countermeasures as "fast track" products [39]. The Bioterrorism Act also provided that drugs submitted for approval under the Animal Efficacy Rule may be designated as fast track products [39 at § 122(b)]. While this statutory provision did not confer any meaningful new authority or requirements upon FDA, it provided a clear signal of Congress's desire for FDA to give developers of biodefense products extra support, assistance, and high-priority treatment.

***Orphan Drug Designations*** The Orphan Drug Act, initially signed into law in 1983, offers various incentives to potential product sponsors, including tax incentives, grants for clinical research, and 7 years of marketing exclusivity if a product that FDA designates as an orphan drug or biological obtains FDA approval for the orphan use [40]. Orphan drugs are products intended to treat rare diseases, defined by statute as diseases affecting fewer than 200,000 Americans [FFDCA § 526(a)(2); 21 USC § 360bb(a)(2)]. As with medical countermeasures for bioterrorism generally, the fear regarding the orphan drug market is that there is not enough of a profit incentive to drive production without legislative or regulatory remedies. Orphan drug exclusivity, in particular, is a powerful incentive. Unlike other forms of nonpatent exclusivity under the Federal Food, Drug, and Cosmetic Act (FFDCA), orphan drug exclusivity blocks FDA approval not only of so-called abbreviated new drug applications, which do not require original safety and effectiveness data, but also full new drug applications (NDAs) containing original clinical data [FFDCA § 527(a); 21 USC 360cc(a)]. The only way to break orphan drug exclusivity is to prove that a subsequent version of the drug is clinically superior or that the holder of the orphan drug approval cannot supply enough product to meet market demand [21 CFR §§ 316.3(b)(13), 316.25(a)(3), 316.36].

One might expect many bioterrorism countermeasures not to qualify for orphan drug status, in that they would be used in a large number of people in the event of a bioterrorist attack. Certain other countermeasures, however, may be expected to have more limited use and therefore can qualify for orphan drug status. Indeed, FDA has approved several orphan drugs for countermeasure uses. The first such products were pentetate calcium trisodium injection (Ca-DTPA) and pentetate zinc trisodium injection (Zn-DTPA), approved for marketing in August 2004 [41]. Both drugs treat radioactive

contamination from materials such as plutonium, americium, or curium. FDA also granted approval to vaccinia immune globulin intravenous (VIGIV) as an orphan drug in February 2005 [42]. VIGIV, a specialty antibody made from human plasma, was approved for treating and preventing severe reactions, such as a severe infection of the skin, that may be brought on by administration of the smallpox vaccine [43]. The drug was developed by DynPort Vaccine Company LLC in partnership with the U.S. Department of Defense Joint Vaccine Acquisition Program for use as a biological defense vaccine [44].

Given the general tendency toward regulatory flexibility to encourage biodefense countermeasures, product developers should consider whether their products might qualify for orphan treatment. One possible objective measure of the potential market size for a countermeasure might be number of doses or courses of therapy the government has indicated it is seeking to procure for the Strategic National Stockpile, if such information is available. The operation of the Project Bioshield program, which governs the procurement of countermeasures for the Strategic National Stockpile, is discussed below.

## 3 ADVANCED DEVELOPMENT, PRODUCT PROCUREMENT, AND MARKETING

New attention to the need for medical countermeasures also has had a widespread effect on how the government has chosen to organize itself and do business with pharmaceutical and biotechnology companies developing countermeasures. As noted, the Project Bioshield Act is the primary pull incentive for development of biodefense countermeasures. The Weapons of Mass Destruction Medical Countermeasures (WMD MCM) Subcommittee, a subset of the Committee on Homeland and National Security of the National Science and Technology Council (NSTC) that predated the Project Bioshield Act [45], has been designated as the lead agency responsible for Project Bioshield implementation and policy planning [46].*

In 2006, the HHS established the Public Health and Emergency Medical Countermeasures Enterprise (PHEMCE) to implement the administration's strategy for development and procurement of medical countermeasures to bioterrorist threats. Although intended to replace the WMD MCM Subcommittee for implementation of Project Bioshield, the WMD MCM Subcommittee has in fact survived the establishment of the PHEMCE and continues to work under the NSTC. Later in 2006, President Bush signed into law the Pandemic and All-Hazards Preparedness Act (PAHPA), which, among other things, established the Biomedical Advanced Research and Development Authority (BARDA). BARDA will eventually assume oversight of the PHEMCE. Questions remain after this reshuffling about how the interagency community will coordinate its efforts and whether authorities have been clearly delegated and delineated.

Despite the lack of clarity regarding interagency oversight, Project Bioshield and other related legislation have focused on providing incentives to the private sector to aid the government in establishing a strategic national stockpile of medical countermeasures. The Project Bioshield Act created this new market. After some criticism regarding government oversight, unanticipated complications in the market and poor incentive programs, follow-on legislation has attempted to fill in those gaps. Major focuses of subsequent legislation were the funding gap created by the push-and-pull system, intellectual property ownership, and liability protection.

### 3.1 Project Bioshield: Procuring a National Stockpile of Medical Countermeasures

***The Project Bioshield Act of 2004*** The Project Bioshield Act of 2004 [47] was designed to address the lack of a commercial market for research, development, and deployment of medical countermeasures against biological, chemical, radiological, and nuclear weapons. It created incentives for private firms to invest in such countermeasures using three separate policy devices. First, it authorized an appropriation of $5.6 billion for procurement with the aim of creating a Strategic National Stockpile of medical countermeasures by the year 2013. The act also directed the NIH to use expedited and simplified peer review and contracting procedures for bioterrorism-related research and development. Finally, the Bioshield Act empowered FDA to use emergency use authority as discussed in Section 2.2.

The central feature of the Project Bioshield Act is the creation of a $5.6 billion reserve fund for the procurement of medical countermeasures against the use of biological, chemical, radiological, and nuclear weapons. This fund permits the federal government to contract for the purchase of medical countermeasures in advance of full product development and FDA approval. Such advance contracting is intended to provide incentives for private companies to initiate research and development of medical countermeasures.

Procurements for the Strategic National Stockpile require interagency collaboration and several determinations. First, the Secretary of the Department of

*Reference 46 will hereafter appear as BioShield CRS Report.

Homeland Security (DHS) must determine what chemical, biological, radiological, and nuclear agents present a "material threat." If a countermeasure for such a threat is not currently available, the secretaries of DHS and HHS can submit a joint proposal to the president to develop that countermeasure. The secretaries must also determine that the countermeasure is a "specific security countermeasure," that is, that it is specifically appropriate for inclusion in the Strategic National Stockpile. Payments made from the $5.6 billion reserve fund must await presidential approval. To date, HHS has obligated about $1.5 billion of the $5.6 billion national stockpile procurement fund (Table 2).

In April 2007, HHS published its plan for implementing the PHEMCE for chemical, biological, radiological, and nuclear threats [48]. The PHEMCE plan contained a list of threats and medical countermeasures that will become the principal procurement priorities under Project Bioshield [48]. The threats on which PHEMCE will focus are, among others, anthrax, botulism, hemorrhagic fever (filoviruses, e.g., Ebola and Marburg), tularemia radiological/nuclear agents, typhus, smallpox (variola virus), volatile nerve agents, and the plague [48, at 10]. The projected future top-priority medical countermeasure programs for PHEMCE include, for example, anthrax antitoxins, anthrax vaccine, acute radiation sickness medical countermeasures, broad-spectrum antibiotics and antivirals, diagnostics, smallpox antivirals, smallpox vaccine, and a volatile nerve agent single antidote [48, at 13]. Of those medical countermeasures, in the next 2 years, PHEMCE hopes to procure broad-spectrum antibiotics, anthrax vaccines, smallpox vaccines, and acute radiation sickness medical countermeasures [48]. Countermeasure procurements for threats posed by the various forms of influenza are not addressed in the PHEMCE implementation plan because HHS has issued a comprehensive HHS Pandemic Influenza Plan [48, at 2]. The HHS Pandemic Influenza Plan has its own provisions about procurement and stockpiling of antivirals and vaccines that target influenza specifically [49].

***Challenges and Criticism for Project Bioshield*** Project Bioshield was enacted with the implicit assumption that its financial incentives would entice large pharmaceutical companies to participate in the development of medical countermeasures. That assumption has not proven correct. As a result, HHS has turned to smaller biotechnology companies. Those companies may, however, lack the capital to implement and sustain a successful research and development program. The cancellation of the large contract with VaxGen for the anthrax rPA vaccine, for instance, has sparked concern within industry regarding the detrimental effects that the HHS contract decision-making process and bureaucratic delay can have on smaller companies. Some companies have also expressed doubt about HHS's ability to manage Bioshield contracts in an effective manner [50].

In its 2007 report on Project BioShield, the Congressional Research Service (CRS) observed that the process for developing countermeasures was moving too slowly [51]. According to CRS, DHS's process for determining material threats is lagging. Meanwhile, HHS has not contracted for medical countermeasures associated with existing material threats. The majority of the funding obligated for stockpile procurement, which began in 2004, has not yet resulted in products moving in substantial volumes into the stockpile.

On September 10, 2007, seven members of Congress wrote the Secretary of HHS, Michael O. Leavitt, to express their frustration with the delays in implementa-

**TABLE 2 Project Bioshield Contract Awards**

| Material Threat | Product | Company | Cost ($ millions) |
|---|---|---|---|
| Anthrax | rPA vaccine | VaxGen, Inc. | 877.5 (Cancelled) |
| Anthrax | AVA vaccine | Emergent BioSolutions | 242.7 |
| Anthrax | ABthrax | Humane Genome Sciences | 165.2 |
| Anthrax | Anthrax immune globulin | Cangene Corp. | 143.8 |
| Botulinum toxin | Botulinum antitoxin (Heptavalent) | Cangene Corp. | 362.6 |
| Radiological/nuclear | Potassium iodide (lquid) | Fleming & Company | 17.5 |
| Radiological/nuclear | Ca-DTPA and Zn-DTPA | Akorn, Inc. | 21.9 |
| Smallpox | Vaccinia Ankara (MVA) smallpox vaccine | Bavarian Nordic A/S | 500.0 |
| | | | Total Cost: |
| | | | 2,331.2 million |
| | | | –877.5 million |
| | | | **$1,453.7 million** |

*Sources*: Adapted from ref. [46], at 11, and Department of Health and Human Services, News Release: HHS Buys Next Generation Smallpox Vaccine, June 4, 2007. Available at: http://www.hhs.gov/news/press/2007pres/06/pr20070604a.html.

tion of Bioshield. They cited companies who asserted that "the Department's inability to effectively manage contracts have [sic] caused them to reconsider participating in Project Bioshield" [52]. Congress hoped that some of their organizational solutions to the above problems would also alleviate bureaucratic delay and increase transparency. Hearings on Capitol Hill have continued to highlight congressional attention to these issues and the belief of members of Congress that Project Bioshield implementation has not been as efficient as they would have hoped [53].

In 2005, the Senate considered a proposed BioShield II bill [54], aimed at addressing the major criticisms and gaps created by the Project Bioshield Act. The key observation of the Senate was that the private sector lacked adequate incentives to invest in medical countermeasures. Among other things, the bill would have established a new procurement fund under the auspices of DHS. The bill also attempted to improve procurement procedures by streamlining the process, requiring and encouraging regulatory improvements and naming DHS as the lead agency. In addition to new funding, other enticements for businesses included new research and development tax incentives and grants and an antitrust exemption for certain activities. The bill attempted to resolve major concerns about intellectual property incentives and product liability protection as well. Although the bill died on the Senate floor, it provided a roadmap for areas of improvement on the Project Bioshield Act. Since then, Congress has passed some new laws to remedy obvious problems and help accelerate procurement processes. Nevertheless, it is clear that criticisms regarding pace and transparency will remain.

### 3.2 Medical Countermeasures Deemed Commercial Items: Contracting Flexibility

The Homeland Security Act of 2002, which established the DHS, provided "special streamlined acquisition authority" for DHS designed to improve incentives for commercial companies to invest in bioterrorism research and to streamline the procurement process [55]. Specifically, section 883 of the Act empowered the Secretary of Homeland Security to deem a DHS procurement to be a "commercial item" procurement if the Secretary "determines in writing that the mission of the Department . . . would be seriously impaired without the use of such authorities" [55, at § 833(a)].

This "deemed commercial item procurement" authority substantially increased DHS's authority in contracting because commercial item procurements are substantially more streamlined than traditional government procurement processes. For example, commercial item procurements are: excepted from certain intellectual property provisions that apply to traditional government contracts [56], exempt from many socioeconomic provisions, such as affirmative action [56, at 52.212-4], permitted to use commercial financing procedures [56, at 12.210], simplified in terms of the contract requirements [56, at 12.403] and more flexible in terms of negotiation [56, at 12.302].

To date, DHS has not exercised its § 833 authority to deem an otherwise noncommercial item contract as a commercial item procurement [57]. DHS officials recently informed the Government Accountability Office (GAO) that the authority to deem a procurement a "commercial item" was "not worth the time and effort required to justify their use given the relatively modest increase in flexibility" [57]. On September 30, 2007, § 833 thus expired without action from Congress. Accordingly, homeland security contractors must comply fully with the requirements imposed on traditional government contractors.

### 3.3 Creation of BARDA and Filling the Valley of Death Funding Gap

One measure to address at least some of the Project Bioshield inadequacies was the Pandemic and All-Hazards Preparedness Act (PAHPA), passed in December 2006, which created the Biomedical Advanced Research and Development Authority (BARDA) [58]. A central purpose of the bill was to allow medical countermeasure makers to receive earlier payments for their medicines. The small biotechnology companies who were participating in Project Bioshield often lacked the capital to fund advanced research and development programs. Because of this capital deficit, the funding gap between the initial grant and the projected procurement, nicknamed the "valley of death," was a serious impediment for those smaller companies.

BARDA is the new Project Bioshield manager and will also oversee PHEMCE research, development, and procurement. It is the lead agency for all advanced development and procurement of medical countermeasures for biological, chemical, radiological, and nuclear threats, as well as emerging infectious diseases. BARDA specifically noted in its draft strategy plan that it was created in part to address the valley of death by supporting products with funding before they are ready for procurement [59]. The FY 2008 budget allocates $189 million toward that goal [60]. PAHPA endows BARDA with the authority to pay countermeasure manufacturers "an advance payment, partial payment for significant milestones, or payment to increase manufacturing capacity" on the authority of the Secretary of HHS as necessary to ensure success of their projects, thus helping to fill the funding gap [61]. Therefore, medical

countermeasures in theory will be more mature when they become subjects of Project Bioshield procurements, significantly lessening the risk involved in their procurement as well [62]. In addition, BARDA has authority to fund advanced development of medical countermeasures that are not eligible for the national stockpile under Project Bioshield [63]. BARDA also aims to incorporate and execute the HHS Pandemic Influenza Plan and the PHEMCE implementation plan, which prioritized pathogens and countermeasures for funding and procurement, into its strategy.

### 3.4 Addressing Intellectual Property Questions

Certain special rules regarding defense and homeland security procurements allow companies greater flexibility in negotiating contracts with the government than most government contractors would face. For example, the Homeland Security Act (HSA) of 2002 [64] and the PAHPA provide DHS and BARDA, respectively, with so-called other transaction authority. Other transactions are transactions that are not constrained by the same rigid procurement rules and regulations that contractors would normally face under the Federal Acquisition Regulations and the Bayh–Dole Act of 1980. The Federal Acquisition Regulations provide numerous contracting requirements, including competition, pricing, and cost accounting standards that generally must be satisfied. In the area of medical countermeasures, many of these requirements would be onerous and some of the competition provisions may even be unfeasible. The Bayh–Dole Act (35 USC § 200–212) generally governs the allocation of rights in patentable inventions first conceived or reduced to practice under federally funded agreements (such as procurement contracts, grants, or cooperative agreements) and grants the government a nonexclusive license to use the invention for governmental purposes.

Other transaction (OT) authority was first developed for the Defense Advanced Research Projects Agency (DARPA) in the 1980s. It was then extended to other parts of the Department of Defense. In general, there are two types of OTs: research OTs and prototype OTs. To be categorized as a research OT, DoD must establish the following three criteria:

- It is not "feasible or appropriate" to use a standard contract, grant, or agreement to carry out the project. A company's reluctance to contract under the intellectual property rules of the Bayh-–Dole Act will satisfy this condition.
- The research project does not duplicate other DoD research.
- To the extent practicable, the contractor must contribute at least half the funding for the project (10 USC § 2371).

DoD contracts for prototype OTs must comply with three conditions as well:

- The research does not duplicate other DoD research.
- To the maximum extent practicable DoD must use competitive procedures.
- The contractor must pay at least one-third of the cost of the project, subject to a few exceptions (10 USC § 23.71 note).

The HSA extended DoD's OT authority to DHS through a 5-year pilot program (6 USC § 391). The three requirements for research OTs are the same for DHS as for DoD. DHS must satisfy additional criteria before implementing a prototype OT. In addition to the three requirements on DoD prototype OT contracts, DHS requires that any contract valued at more than $5 million must include a clause permitting the GAO to examine the contractor's books and records. The wording of the HSA is unclear, but it may also require that the Secretary of DHS determine that it is not "feasible or appropriate" to use a standard procurement contract, grant, or agreement for the project. DHS's pilot program for OT authority expired at the beginning of 2008 and was subsequently extended.

The PAHPA, which in 2006 established BARDA, also granted BARDA OT authority but did not expand that authority to any other part of HHS. Section 401 of PAHPA also cited to 10 USC §2371, DoD's OT authority statute, to establish BARDA's OT authority. PAHPA specifies that OT authority for BARDA should be exercised for research or prototype OT contracts in accordance with the then existing DOD rules for prototype OTs. The senior procurement executive for HHS must sign off on any other transaction valued at more than $20 million dollars.

OT authority does not extend to any agency outside of DHS and DoD other than BARDA. Although the Project Bioshield Act provides some guidance on procurement generally, it does not specifically address OT authority, nor does it have protections for companies with intellectual property concerns. As mentioned above, the Project Bioshield Act does contain expedited procurement authority and creates some exceptions to the standard procurement rules. For instance, the act allows "other than full and open competition" and expedited peer review. For procurement for the national stockpile itself, the act requires certain contract terms

to be included in the procurement contract. Those specified terms address payment, contract duration, storage of the medical countermeasure, product approval, and export controls. However, important industry concerns about liability were also not addressed.

### 3.5 Liability Protection

One of the greatest concerns for potential Project Bioshield participants has been the risk of legal liability involving medical countermeasures. Were a biological terrorist attack to occur, it is likely that hundreds or even thousands of Americans would require medical countermeasures. In the unfortunate event that a medical countermeasure did not protect against the biological threat, or if it had an unanticipated side effect, the company that developed and manufactured the drug would theoretically be exposed to potentially ruinous liability. Such risks present major disincentives for companies, who have already displayed skepticism about the risk of failure and low profit margins that characterize the medical countermeasures industry.

A variety of laws limit liability exposure for drug companies interested in producing medical countermeasures. The most significant is the Pandemic Readiness and Emergency Preparedness (PREP) Act, which provides extensive liability protection, but thus far has been applied only to the avian flu vaccine [65]. Other relevant laws include the Support Anti-terrorism by Fostering Effective Technologies (SAFETY) Act [66] and some disease-specific legislation.

***P.L. 85-804: The Beginning of Government Indemnification*** The earliest law to consider liability protection for national defense contracting was passed nearly 50 years ago. In Public Law 85-804, Congress authorized the president to contract for facilitation of the national defense. Executive Order 10789 (as amended) provided that the government could partially indemnify contractors for risks defined in the contract as "unusually hazardous or nuclear" [FAR 52.250-1(c)]. Shortly after September 11, 2001, Executive Order (EO) 13232 extended that contract authority to HHS.

Indemnity under PL 85-804 and EO 10789 (as amended) covers claims for death, personal injury, or loss of or damage to property if the contract contains language specifying that a particular risk is "unusually hazardous or nuclear." Contractors are not indemnified for claims within deductible amounts of insurance. Additionally, protection under PL 85-804 extends only to negligent acts or omissions; indemnification is not available for willful misconduct or bad faith. To be eligible for indemnification, the contractor must notify the United States of any claims against it, and the United States reserves a right to "control or assist" the settlement or defense. Therefore, although this indemnification grant was revolutionary at the time it became law, it only applies in limited circumstances: claims for negligence that arise from contractually defined unusually hazardous or nuclear activities.

***The Smallpox Act*** After September 11, 2001, with attention focused on contracting with the private sector to facilitate terrorism prevention, relevant liability protection laws were passed in an attempt to address particular concerns related to terrorism. The HSA in 2002 included a provision commonly referred to as the Smallpox Act, which covered manufacturers and distributors of medical countermeasures [67]. Under the Smallpox Act, the Secretary of DHS must declare that an actual or potential bioterrorist incident or public health emergency exists and designate certain countermeasures to be covered. If those conditions were satisfied, the United States would then retain liability for any claim arising out of the administration of those designated countermeasures by a qualified person.

***The SAFETY Act*** Also included in the 2002 HSA was the SAFETY Act. The SAFETY Act empowers the Secretary of DHS to designate certain products as "qualified anti-terrorism technologies" (QATTs) [66, at § 863(1)]. The act limits liability when those technologies are deployed in defense against, response to, or recovery from an act of terrorism by creating an exclusive federal cause of action and indemnifying sellers of the technologies from liability to third parties arising out of an act of terrorism.

Like PL 85-804, the SAFETY Act indemnifies sellers of QATTs only for liability amounts above a specified level of insurance coverage that is set forth in an individual contract. If the SAFETY Act protections for QATTs are applicable, plaintiffs may not seek punitive damages, recovery is limited to economic damages, and noneconomic damages are available only in proportion to the percentage of responsibility of the defendant, reduced by any other compensation received by the plaintiff. SAFETY Act protections do not apply to willful misconduct or fraudulent actions by the contractor [66, at § 863(b)].

In addition to potential designation of a technology as a QATT, the SAFETY Act provides further protection for technologies that are certified as eligible for the government contractor defense. Certified QATTs have the benefit of a rebuttable presumption that the government contractor defense, which shields contractors from liability for third-party tort claims, applies to third-party tort claims resulting from an act of terrorism. Only a

showing of fraud or willful misconduct will allow a plaintiff's claim to go forward [66, at § 863(d)].

***The PREP Act*** The Public Readiness and Emergency Preparedness (PREP) Act, part of the DoD authorization bill for 2006, provides liability protection to manufacturers, distributors, and program planners for pandemic and epidemic products and medical countermeasures [68]. The Secretary of HHS can designate a product to be covered by the act. To do so, the Secretary must find that a "disease or other health condition or other threat to health constitutes a public health emergency" or that there is a "credible risk" of such an emergency in the future [66, at § 319F-3(b)]. By its terms, the PREP Act appears only to cover pandemic disease scenarios and does not cover terrorism, but it is conceivable that a bioterrorism attack would constitute a pandemic or epidemic public health emergency and therefore fall under the purview of the PREP Act.

Once the Secretary of HHS has declared a medical countermeasure to be a pandemic countermeasure to an existing threat to public health, PREP Act liability protections apply. Liability protection in the form of immunity from suit exists for losses that were "caused by, arising out of, relating to, or resulting from" use of a medical countermeasure [66, at § 319F-3(a)] by the distributor, manufacturer, or program planner [66, at § 319F-3(c)]. The sole exception to this immunity is for federal causes of action regarding death or serious physical injury proximately caused by willful misconduct [66, at § 319F-3(d)]. In those cases, damages would be reduced by any collateral source benefits [66, at § 319F-3(e)(8)] and noneconomic damages could only be recovered according to the percentage of responsibility of the defendant [66, at § 319F-3(e)(7)].

The PREP Act also creates a fund to cover compensation for injuries that result directly from administration or use of a covered countermeasure under a declaration by the Secretary of HHS. While some critics were hoping for clarity as to which injuries would be covered, the PREP Act does not provide much guidance. It refers to authority in existing statutes that require the Secretary of HHS to compile tables of known side effects of vaccines and to provide a complete list of covered injuries [66, at § 319F-4(b)(4) & (5)]. According to the statute, when the Secretary makes a declaration under the PREP Act, a table of eligible side effects or injuries should also be issued.

There has only been one declaration under the PREP Act of a pandemic or epidemic threat to public health. In early 2007, the Secretary of HHS issued a declaration in the *Federal Register* that the avian flu virus posed a potential future pandemic threat if it were to evolve into human influenza [69]. The declaration recommended "manufacturing, testing, development and distribution . . . of the pandemic countermeasure influenza A (H5N1) vaccine," which targets avian flu. Therefore, although the PREP Act provides significant liability protection, the only people it currently covers are manufacturers, distributors, and program planners for one medical countermeasure: the avian flu vaccine.

## 4 REGULATORY RESTRICTIONS AFFECTING RESEARCH

This section describes key regulatory restrictions applicable to those who conduct research involving biodefense products. This discussion is intended only as a general introduction to this topic. Those who intend to engage in biodefense-related research must ensure that they are familiar with the many complex laws and regulations that apply in this field and take appropriate steps to comply.

### 4.1 Select Agent Regulations

In the wake of the Oklahoma City bombing of the Alfred E. Murrah Building in April 1995, Congress passed the Antiterrorism and Effective Death Penalty Act of 1996 in October 1996 (AEDPA). As part of the implementation of this act, the first list of regulated biological agents was created [70]. This became known as the "select agent" list.

After the terrorist events of September and October 2001, Congress passed the Uniting and Strengthening America by Providing Appropriate Tools Required to Intercept and Obstruct Terrorism Act of 2001 (USA PATRIOT Act or PATRIOT Act). This was followed in June 2002 by passage of the Public Health Security and Bioterrorism Preparedness and Response Act of 2002 (Bioterrorism Act), which expanded the Select Agent Program. Both the PATRIOT Act and the Bioterrorism Act contained provisions to tighten the regulation of the possession, use, and transfer of "select agents," biological agents, and toxins with the potential to pose a severe threat to public health and safety, animal and plant health, or animal and plant products [71].

The PATRIOT Act placed restrictions on persons who possess select agents and established criminal penalties for illegal possession of such agents [72]. Section 202(a) of the Bioterrorism Act required that all persons possessing biological agents or toxins deemed a threat to public health provide notice to the Secretary of HHS [73], while section 213(b) required all persons possessing biological agents or toxins deemed a threat to animal or plant health and to animal or plant products to notify the Secretary of the U.S. Department of Agriculture (USDA) [73, at § 213(b)].

**TABLE 3 Examples of HHS and USDA Select Agents**[a]

| HHS Select Agents | USDA Select Agents | Overlap Agents |
|---|---|---|
| Cercopithecine herpesvirus (Herpes B virus) | African horse sickness virus | *Bacilius anthracis* |
| Conotoxins | African swine fever virus | Botulinum neurotoxins |
| Crimean-Congo hemorrhagic fever virus | Avian influenza virus (highly pathogenic) | *Brucella* (*abortus*, *melitensis*, and *suis*) |
| Ebola viruses | Bovine spongiform encephalopathy (BSE) agent | Burkholderia |
| Marburg virus | Camel pox virus | *Clostridium perfringens* epsilon toxin |
| Monkeypox virus | Classical swine fever virus | *Coccidiodes immitis* |
| Reconstructed 1918 influenza virus | Foot-and-mouth disease virus | *Coxiella burnetii* |
| Ricin | Japanese encephalitis virus | Eastern Equine Encephalitis virus |
| *Rickettsia rickettsii* | Peste des petits ruminants virus | *Francisella tularensis* |
| Variola major virus (Smallpox) | Rinderpest virus | Rift Valley fever virus |
| *Yersinia pestis* (plague) | Swine vesicular disease virus | Staphylococcal enterotoxins |

[a]For the complete and up-to-date lists, refer to the HHS and USDA regulations, currently found at 42 CFR § 73.7(a); 9 CFR § 121.7(a).

HHS and USDA published final rules in March 2005 implementing the select agent provisions of the PATRIOT Act and the Bioterrorism Act [74]. The regulations establish the agencies' respective lists of select agents (Table 3) and provide that "an individual or entity shall not possess, use, or transfer any . . . select agent or toxin without a certificate of registration issued by the . . . Secretary [or Administrator]" [42 CFR § 73.7(a); 9 CFR § 121.7(a)]. The regulations also establish a process to document successful transfer of agents; audit, quality control, and accountability mechanisms; and agent disposal requirements.

To receive the requisite registration certificate, the HHS Secretary or Administrator of the USDA's Animal and Plant Health Inspection Service (APHIS) must approve, based on a security risk assessment by the Attorney General, the individual or entity, responsible official, and any individual who owns or controls the entity [42 CFR § 73.7(c)(1); 9 CFR § 121.7(c)(1)]. The issuance of a certificate of registration may be contingent upon inspection or submission of additional information, such as a security plan, biosafety plan, or incident response plan [42 CFR § 73.7(f); 9 CFR § 121.7(f)]. An individual who transfers select agents to a person the transferor knows or has reason to believe is not registered according to the provisions in 42 CFR § 73 is subject to criminal penalties [18 USC § 175b(b)]. Similarly, any person who knowingly possesses a select agent without the requisite registration shall be fined, or imprisoned, or both [18 USC § 175b(c)].

Some select agents or toxins are excluded from regulation. One such exception applies to toxins under the control of a principal investigator, treating physician or veterinarian, or commercial manufacturer or distributor, so long as the aggregate amounts do not exceed those provided in the regulations [42 CFR § 73.4(d)(3)]. An example of this is botulinum neurotoxins, for which an exemption is provided if the aggregate amount under the control of a principal investigator does not, at any time, exceed 0.5 mg [42 CFR § 73.4(d)(3)]. The exception is intended to "balance the regulatory oversight of agents and toxins that have the potential to pose a severe threat to public health and safety while maintaining availability of these agents and toxins for research and educational activities" (70 *Fed Reg* 13297). Other exemptions apply to select agents used in products approved, licensed, or registered under federal laws including the FFDCA, section 351 of the Public Health Service Act (pertaining to biological products), and the Federal Insecticide, Fungicide, and Rodenticide Act [42 CFR § 73.5(c)].

The serious consequences of violating the select agent rules was highlighted by the case of Dr. Thomas Butler, a researcher at Texas Tech University. Dr. Butler brought plague samples back from Tanzania for research purposes, did not declare them at Customs, and drove from Lubbock, Texas, to a CDC laboratory in Colorado to have the samples tested, all without the necessary government permits [75]. Dr. Butler later sent samples back to Tanzania and flew with the samples from Lubbock to Washington D.C., again without the necessary paperwork. After Dr. Butler reported to his local health department that 30 plague vials were missing, the FBI arrested Dr. Butler [75]. Dr. Butler was eventually convicted on 47 of 69 criminal charges, resulting in a 2-year prison term [76]. Included in his conviction were 3 counts relating to the transportation of human plague bacteria, *Yersinia pestis*, an HHS select agent.

### 4.2 Biosafety Laboratories

The CDC and NIH prepared the Biosafety in Microbiological and Biomedical Laboratories (BMBL) as a guidance document for working with particular select agents. The BMBL states that biosafety procedures must be incorporated into the laboratory's standard operating procedures or biosafety manual;

personnel must be advised of special hazards and are required to read and follow instructions on practices and procedures; and personnel must receive training on the potential hazards associated with the work involved and the necessary precautions to prevent exposure [77].

The anthrax attacks in 2001 highlighted the shortage of high-containment lab capacity available to conduct research leading to the development of medical countermeasures [78]. To address this concern, the Bush Administration and Congress responded by providing increased funding for biodefense research and for additional Biosafety Level 3 (BSL-3) and Biosafety Level 4 (BSL-4) labs in the private sector, especially in university settings [78].

As a result, since September 11, 2001, there has been a "major proliferation" of high-containment BSL-3 and BSL-4 laboratories in the United States across all sectors—federal, academic, state, and private [78]. However, concerns have been raised about the oversight of these labs. No single federal agency is accountable for tracking the number of all BSL-3 and BSL-4 labs within the United States, and no agency knows how many such labs there are in the United States or the locations of the labs [78]. Information on expansion is available about high-containment labs that are registered with the CDC–USDA's Select Agent Program and/or are federally funded, but much less is known about the expansion of labs outside the Select Agent Program and the nonfederally funded labs, including location, activities, and ownership [78].

Before 1990, all BSL-4 labs were federal labs. Today, while expansion is taking place within the federal sector as well, there are also BSL-4 labs at universities, as part of state response, and in the private sector [78]. The expansion of BSL-3 labs is also taking place within all sectors. However, the only definitive data available are on labs registered with the CDC–USDA Select Agent Program. Within that program, two-thirds of registered BSL-3 labs are outside the federal sector [78].

This growth in government and privately owned high-containment laboratories has led to concerns about the potential for unintentional releases of dangerous agents. Recent incidents in the United States and England indicate the hazards associated with these labs. One NIAID-funded facility, at Texas A&M University—registered with CDC's Select Agent Program and approved for work on several select agents—was found to have allowed unauthorized access to high-security laboratories and to have misplaced vials of infectious agents [79]. A 21-page report issued by the CDC and released in late August 2007 followed months of scrutiny over the university's failure to report laboratory workers being exposed to dangerous biological agents [79]. As a result of its failures, the lab was banned from select agent research and required to undertake corrective actions [79].

Also in the summer of 2007, contamination of foot-and-mouth disease was discovered at several local farms near Pirbright, England, the site of several high-containment labs that work with live food-and-mouth disease virus, one of the most highly infectious livestock diseases [78]. An investigation of the site found evidence of long-term damage and leakage of the drainage system servicing the site. Although the definitive cause of the release has not been determined, it is suspected that contaminated wastewater from Pirbright's labs leaked into the surrounding soil from the deteriorated drainage pipes and that live virus was then carried offsite by vehicles splashed with contaminated mud [78]. The investigation revealed that monitoring and testing for the preventative maintenance of pipework for the drainage system was not a regular practice on site [78]. This incident shows that beyond initial design and construction, ongoing maintenance plays a critical role in ensuring that high-containment labs operate safely and securely over time, because even the smallest of defects can affect safety.

Perhaps in response to the Texas A&M violations and the incident in England, the House Committee on Energy and Commerce held a hearing in October 2007 on potential risks from BSL-3 and BSL-4 laboratories in the United States [80]. The public and congressional attention on this issue is likely to refocus U.S. regulators' attempts to ensure compliance with applicable biosafety standards and may make it more difficult in the future to obtain registration of a new biosafety laboratory.

### 4.3 Export Controls

U.S. law imposes strict controls on the export of sensitive biological agents, products designed to produce and defend against such agents, as well as related technology. U.S. export controls apply not only to transfers and international collaborations across borders, but also to transfers to "foreign nationals" within the United States.

Most open research work on university campuses in the United States falls under fundamental research exclusions and does not trigger licensing requirements for access by non-U.S. scientists in the United States. The research must be eligible for publication without restriction by the university or the project sponsors, and eligible for open access by all. These exclusions generally cover work on campus and, frequently, do not extend to international collaborations.

Any research and development work not eligible for the fundamental research protections falls under the International Traffic in Arms Regulations (ITAR)

or the Export Administration Regulations (EAR). Category XIV of the ITAR controls toxicological agents, including chemical agents, biological agents, and associated equipment (22 CFR § 121.1). The biological agents controlled are narrow and specific and only cover those agents developed or modified to cause harm.* Any medical countermeasures specifically designed or modified for the sole purpose of protecting against the biological agents controlled in ITAR Category XIV would require a State Department license for export from the United States or access to related technical data in the United States by anyone other than a U.S. citizen or permanent resident. Similarly, any equipment specifically designed or modified for military operations and compatibility with military equipment would fall under these stringent controls.

Most research and development work in the bioterrorism area falls under the jurisdiction of the Commerce Department's Export Administration Regulations. The most sensitive human and zoonotic pathogens and toxins, as well as certain sensitive animal pathogens and their genetic elements, require licenses for export from the United States to every destination. Similarly, only U.S. citizens and permanent residents may access in the United States without a license the technology for the development, production, and disposal of the most sensitive agents [81]. Equipment designed to detect, monitor, and defend against biological agents may trigger a license requirement not only for export to certain destinations but also for access to the technology for the development and production, and in some cases use, of the related technology [82].

## 5 CONCLUSION

This chapter reviews the significant efforts of Congress and the executive branch of the U.S. government to create the right incentives for the private sector to develop medical countermeasures. These measures, together with important changes in public health surveillance and delivery, are designed to enhance the ability of the United States to reduce the risk of, and improve its ability to recover from, a bioterrorist attack. While there have been impressive gains made in this area, most observers believe that the intensity and pace of change needs to accelerate to counter the growing threat of bioterrorism. These same programs and efforts also help in the fight against important naturally occurring infectious diseases and pandemics. Researchers or entities considering development programs in these therapeutic areas should stay abreast of future measures that may be of benefit to their development programs.

*Category XIV(b) controls "Biological agents and biologically derived substances specifically developed, configured, adapted or modified for the purpose of increasing their capability to produce casualties in humans or livestock, degrade equipment or damage crops."

## ACKNOWLEDGMENT

The authors would like to thank their colleagues John Barker, Jennifer Newberger, and Lindsay Rodman for their assistance in preparing this chapter.

## REFERENCES

1. Rhodes K, Chief Technologist Center for Technology and Engineering Applied Research and Methods (2007). High-containment biosafety laboratories, preliminary observations on the oversight of the proliferation of BSL-3 and BSL-4 laboratories in the United States. GAO-08-108T. Washington, DC: General Accounting Office, October 4, 2007.
2. Homeland Security Presidential Directive 10 (2004). Biodefense for the 21st century. April 28, 2004. Available: http://www.whitehouse.gov/homeland/20040430.html. Accessed September 6, 2007.
3. Falkenrath RA (2006). Statement before the United States Senate Committee on Health, Education, Labor and Pensions. March 16, 2006. Citing Danzig R (October 2005). Towards a long-term strategy for coping with the risk of bioterrorism. The Defense Science Office, Arlington, VA.
4. Jackson R, et al. (February 2001). Expression of mouse interleukin-4 by a recombinant ectromelia virus suppresses cytolytic lymphocyte responses and overcomes genetic resistance to mousepox. *J Virol 75*, 1205. Finkel E (2001). Engineered mouse virus spurs bioweapon Fears. *Science 291*, 585.
5. Price LB, et al. (2003). In vitro selection and characterization of *Bacillus anthracis* mutants with high-level resistance to ciprofloxacin. *Antimicrob Agents Chemother 47*, 2362. Choe CH, et al. (2000). In vitro development of resistance to ofloxacin and doxycycline in *Bacillus anthracis Sterne*. *Antimicrob Agents Chemother 44*, (2000). Brook I, et al. (2001). In vitro resistance of *Bacillus anthracis Sterne* to doxycycline, macrolides, and quinolones. *Int J Antimicrob Agents 18*, 559.
6. Smith B, Inglesby T, O'Toole T (2003). Biodefense R&D: Anticipating future threats, establishing a strategic environment. *Biosecurity Bioterrorism 1*, 193, 197.
7. Read M (2003). Testimony on behalf of the Biotechnology Industry Organization, before the House Energy and Commerce Subcommittee on Health and the House Select Committee on Homeland Security Subcommittee on

Emergency Preparedness and Response, regarding Project Bioshield Act of 2003, March 27, 2003.

8. Fauci A (August 2005). The global challenge of infectious diseases: The evolving role of the National Institutes of Health in basic and clinical research. *Nature Immunol 6*, 734, 747.
9. Franco C, Dietch S (2007). Billions for biodefense. *Biosecurity and Bioterrorism: Biodefense Strategy, Practice and Science 5*, 117, 118.
10. NIAID Strategic Plan for Biodefense Research—2007 update (September 2007). Available: http://www3.niaid.nih.gov/topics/BiodefenseRelated/Biodefense/PDF/biosp2007.pdf.
11. National Institute of Allergies and Infectious Diseases. NIAID Funding Opportunities List. Available: http://www.niaid.nih.gov/ncn/budget/opps.htm.
12. National Institute of Allergies and Infectious Diseases. NIAID Requests for Proposals. Available: http://www.niaid.nih.gov/contract/rfps.htm.
13. National Institute of Allergies and Infectious Diseases (1994). NIAID Award and Application Policies. *NIH Guide 23*(40), November 18, 1994. Available: http://grants.nih.gov/grants/guide/notice-files/not94-331.html.
14. Federal Business Opportunities, A—Biodefense Vaccine Enhancement. Available: http://www.fbo.gov/spg/HHS/NIH/NIAID/BAA%2DNIH%2DBARDA%2DNIAID%2DDMID%2DAI2007007/SynopsisP.html.
15. National Institute of Allergies and Infectious Diseases. Receipt to Funding Timeline for NIAID. Available: http://www.niaid.nih.gov/ncn/grants/charts/timeline_review.htm.
16. DoD fiscal year 2008/2009 budget estimates, Research, Development, Test, and Evaluation, Defense-Wide, Vol. 4, Chemical and Biological Defense Program (CBDP), at vii (February 2007).
17. Dimasi J, Hansen R, Grabowski H (2003). The price of innovation: New estimates of drug development costs. *J Health Econ 22*, 151.
18. 21 United States Code (USC) § 903(b). FDA mission statement.
19. New drug and biological drug products. 64 *Fed Reg* 53960 (October 5, 1999).
20. Public Health Security and Bioterrorism Preparedness and Response Act of 2002. Public Law No. 107-188, § 123, 116 Stat. 594, 613 (2002).
21. New drug and biological drug products. 67 *Fed Reg* 37988, May 31, 2002.
22. 21 CFR § 314.610(a)(2) (2002).
23. Final Rule, ref. 19, at 37991.
24. Center for Drug Evaluation and Research, Food and Drug Administration, U.S. Department of Health and Human Services. Improving public health through human drugs. Available: http://www.fda.gov/cder/reports/rtn/2002/rtn2002.htm#Counterterrorism.
25. News-Medical.net. Novel respiratory antibiotic cethromycin, shown to be effective against anthrax. Available: http://www.news-medical.net/?id=24655. Accessed May 6, 2007.
26. U.S. Food and Drug Administration. Transcript, anthrax vaccine: Efficacy testing and surrogate markers of immunity. Available: http://www.fda.gov/cber/minutes/anthrax0402.htm. Accessed September 21, 2007.
27. Transcript, Public Workshop on Animal Models and Correlates of Protection for Plague Vaccines. Available: http://www.fda.gov/cber/minutes/plague101304t.pdf. Accessed September 21, 2007.
28. Biomedical Market Newsletter. Anthrax Vaccines Public Workshop to be Held by FDA, September 21, 2007. Available: http://www.biomedical-market-news.com/cgilocal/SoftCart.exe/online-store/scstore/shophome.html?E+scstore.
29. Project BioShield Act of 2004. Public Law No. 108-276, § 4, 118 Stat. 835, 853–859 (2004).
30. Authorization of emergency use of anthrax vaccine adsorbed. 70 *Fed Reg* 5452 (February 2, 2005).
31. Pharma Marketletter. War, terror and terrorism overshadow normally sanguine FDLI meeting, April 21, 2003. Available: 2003 WLNR 5700724.
32. Federal Food, Drug and Cosmetic Act, § 564(b)(1), 21 USC § 360bbb-3(b)(1) (2004).
33. Authorization of emergency use of anthrax vaccine adsorbed. 70 *Fed Reg* 5454 (February 2, 2005).
34. Nightingale S, Prasher JM, Simonson S. U.S. Department of Health and Human Services (July 2007). *Emergency Use Authorization (EUA) to Enable Use of Needed Products in Civilian and Military Emergencies, United States*. Available: http://www.cdc.gov/ncidod/ts/print.asp.
35. Ruppe D (2004). *Pentagon Looks to Declare Anthrax Risk Emergency, Resume Mandatory Vaccinations*. Global Security Newswire, December 17, 2004. Available: http://www.nti.org/d%5Fnewswire/issues/2004/12/17/9cf4b007%2D984a%2D4d2d%2Db546%2Dc272035c9b1e.html.
36. Final Order. 70 *Fed Reg* 75180 (December 19, 2005).
37. U.S. Food and Drug Administration (FDA) Center for Drug Evaluation and Research. *Manual of Policies and Procedures*. MAPP 6020.3R, July 16, 2007.
38. U.S. Food and Drug Administration (FDA). Guidance for industry, fast Track drug development Programs—Designation, Development, and Application Review, January 2006.
39. Bioterrorism Act § 122(a).
40. Orphan Drug Act. Public Law No. 97-414, as amended.
41. Kugler M. About.com. *Orphan Drugs Approved in 2004*. Available: http://rarediseases.about.com/od/orphandrugs/a/010305.htm.
42. Kugler M. About.com. *Orphan Drugs Approved in 2005*. Available: http://rarediseases.about.com/od/orphandrugs/a/122005.htm.
43. MedScape Today. Available: http://www.medscape.com/viewarticle/499964_3. Accessed September 20, 2007.
44. Kugler M. About.com. *Orphan Drugs Approved in 2005*. Available: http://rarediseases.about.com/od/orphandrugs/a/122005.htm.

45. Executive Order No. 12881.
46. Congressional Research Service. Project BioShield: Appropriations, acquisitions, and policy implementation issues for Congress, RL33907, March 8, 2007, at 5.
47. Project BioShield Act of 2004. Public Law No. 108-276, § 4, 118 Stat. 835, 853–859 (2004).
48. Office of the Assistant Secretary for Preparedness and Response. HHS Public Health Emergency Medical Countermeasures Enterprise Implementation Plan for Chemical, Biological, Radiological and Nuclear Threats. Available: http://www.hhs.gov/aspr/barda/documents/phemce_implplan_041607final.pdf.
49. Department of Health and Human Services. HHS Pandemic Influenza Plan, November 2005. Available: http://www.hhs.gov/pandemicflu/plan/pdf/HHSPandemicInfluenzaPlan.pdf.
50. Letter from BG Thompson, JR Langevin, RD Holt, LL Bostwell, D Loebsack, BL Braley, and S King, U.S. Representatives, to MO Leavitt, Secretary of the Department of Health and Human Services, September 10, 2007.
51. BioShield CRS Report, ref. 47.
52. Letter from U.S. Representatives to Leavitt, ref. 50.
53. Six years after anthrax: Are we better prepared to respond to bioterrorism: Hearing before the S. Comm. *On Homeland Security and Governmental Affairs, 110th Congress* (October 2007).
54. Project BioShield II Act of 2005, S. 975, 109th Congress (1st Session 2005).
55. Homeland Security Act of 2002, Public Law No. 107-296 § 833, 116 Stat. 2135 (2002).
56. 48 CFR § 12.211.
57. Government Accountability Office, DHS Special Streamlined Acquisition Authority, GAO-06-783R, June 20, 2006 at 2.
58. Pandemic and All-Hazards Preparedness Act of 2006. Public Law No. 109-417, 120 Stat. 2831 (2006).
59. U.S. Department of Health and Human Services Public Health Emergency Medical Countermeasures Enterprise Biomedical Advanced Research and Development Authority. Draft BARDA Strategic Plan for Medical Countermeasure Research, Development and Procurement. Available: http://www.hhs.gov/aspr/barda/documents/draftbardaplan.pdf.
60. News Release. HHS buys next generation smallpox vaccine. April 26, 2007.
61. Pandemic and All-Hazards Preparedness Act of 2006. Public Law No. 109-417 § 406, 120 Stat. 2831 (2006).
62. Six years after anthrax: Are we better prepared to respond to bioterrorism. Hearing before the S. Comm. On Homeland Security and Governmental Affairs, 110th Congress (October 2007). Testimony of Dr. GW Parker, Principal Deputy Assistant Secretary, Office of the Assistant Secretary for Preparedness and Response, Department of Health and Human Services.
63. Government Accountability Office, Project Bioshield, GAO-08-88, October 23, 2007, at 9.
64. Homeland Security Act of 2002. Public Law No. 107-296, 116 Stat. 2135 (2002).
65. Public Readiness and Emergency Preparedness Act. Public Law No. 109-148, 119 Stat. 2680, 42 USC § 247d-6d (2005).
66. Support Anti-terrorism by Fostering Effective Technologies Act of 2002. Public Law No. 107-296 § 851 et seq., 116 Stat. 2135 (2002).
67. Smallpox Act. Public Law No. 107-296 § 304, 116 Stat. 2135 (2002).
68. Public Readiness and Emergency Preparedness Act. Public Law No. 109-148, 119 Stat. 2680 (2005).
69. Pandemic Countermeasures; Declaration Under the Public Readiness and Emergency Preparedness Act. 72 Fed Reg 4710 (February 1, 2007).
70. Antiterrorism and Effective Death Penalty Act of 1996 in October 1996. Public Law No. 104-132, § 511(d)(1) (1996).
71. Uniting and Strengthening America by Providing Appropriate Tools Required to Intercept and Obstruct Terrorism (USA PATRIOT) Act Of 2001. Public Law No. 107-56, § 175b, 115 Stat. 272, 386 (2001); Public Health Security and Bioterrorism Preparedness and Response Act of 2002. Public Law No. 107-188, Title II, 116 Stat. 594, 637–662 (2002).
72. PATRIOT Act, § 175b.
73. Bioterrorism Act, § 202(a).
74. Possession, use, and transfer of select agents and toxins. 70 *Fed Reg* 13294 (March 18, 2005); Agricultural Bioterrorism Protection Act of 2002. 70 *Fed Reg* 13242 (March 18, 2005).
75. Enzerink M, Malakoff D (2003). The trials of Thomas Butler. *Science*, December 19, 2003. Available: 2003 WLNR 886687.
76. United States v. Butler, 429 F.3d 140, 144 (2005).
77. Centers for Disease Control and Prevention, National Institutes of Health. (February 2007). *Biosafety in Microbiological and Biomedical Laboratories*, Section III, 5th ed. Available: http://www.cdc.gov/od/ohs/biosfty/bmbl5/bmbl5toc.htm.
78. Government Accountability Office, Statement of Keith Rhodes, GAO-08-108T, October 4, 2007.
79. Ramshaw E (2007). CDC reprimands Texas A&M over lab safety. D*allas Morning News*, September 4, 2007. Available: http://www.dallasnews.com/sharedcontent/dws/news/healthscience/stories/090507dntexcdc.9f66c8cc.html.
80. House Committee on Energy and Commerce (2007). News release. *Dingell, Stupak Announce Hearing on Biosafety Labs: Committee to Examine Risks Associated with Labs*, August 9, 2007. Available: http://energycommerce.house.gov/Press_110/110nr57.shtml.
81. 15 CFR, Supplement No. 1 to Part 774, §1C350, 351, 352, 1E001, 1E350 and 1E351.
82. 15 CFR, Supplement No. 1 to Part 774, 1E001 and 2E301.

# DEVELOPMENT AND APPLICATION OF CONFORMATION-SELECTIVE INTEGRIN ANTIBODIES

Nathan S. Astrof,[1,2] Dan Peer,[1,2] and Motomu Shimaoka[1,3]

[1]*Immune Disease Institute, Children's Hospital Boston, and Harvard Medical School, Boston, Massachusetts*
[2]*Laboratory of Nanomedicine, Dept. of Cell Research and Immunology, George S. Wise Faculty of Life Sciences, Tel Aviv University, Tel Aviv, Israel*
[3]*Department of Molecular Pathobiology and Cell Adhesion Biology, Mie University Graduate School of Medicine, Mie, Japan*

## 1 INTRODUCTION

### 1.1 Antibodies Constitute Natural Host Defense

Living organisms are continually challenged by pathogenic infections that, if allowed to persist, may result in severe disease and death. The role of the immune system is to identify, and bind to, the molecular determinants (called *epitopes*) of such pathogens, leading to their elimination from the host [1]. The challenge faced by the immune system is to identify a relatively small number of foreign substances, or antigens, against the much broader background of self and otherwise harmless or even beneficial agents (including ingested foods and pharmaceuticals). Failure to obtain sufficient discrimination between host and foreign tissue results in a wide spectrum of autoimmune pathologies, while cancer may arise, in part, due to the inability of the body to distinguish malignant from healthy cells.

The humoral adaptive immune response is mediated by soluble immunoglobulin molecules (antibodies) secreted by B cells that bind to foreign antigens and, through interactions with other components of the immune system, play a pivotal role in eliminating them from the organism [1]. The surfaces of pathogens, such as bacteria and viruses, have high molecular complexity with multiple potential (overlapping) epitopes. Each B cell secretes an antibody with a single specificity (monoclonal). A polyclonal response requires the participation of many B cells to provide effective coverage of the pathogen surface. The affinity of each specific antibody may differ, even against the same epitope. Through a process of clonal selection, the highest affinity antibodies are selected and become dominant in the serum.

### 1.2 Therapeutic Applications of Antibodies

The immune system can be modulated by the therapeutic application of antibodies with reactivity to host cell proteins. Monoclonal antibodies to human molecules can be obtained by immunization of laboratory species (such as mice) or by in vitro selection from synthetic libraries of antibody sequences displayed on the surfaces of microbes, such as filamentous bacteriophage [2, 3]. The origin of the antibody diversity is critical; antibodies from foreign organisms (and even purely human antibodies) can function as antigens when injected, prompting a rapid clearance and/or full immune response. Genetic engineering of the antibody sequence can be used to reduce the immunogenicity of antibodies. Native antibodies function both by blocking critical protein–protein interactions necessary for pathogen attachment and by recruiting effector cells to mediate destruction and clearance of the invading organism. Therapeutic antibodies may also be used to trigger a

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

native immune response against a disease-causing agent, either foreign or native. For some applications, such as diagnostic reagents or as targeting agents for drug delivery vehicles, the effector domains are unnecessary (or potentially deleterious) and are removed from the protein sequence, either pre- or posttranslation.

Antibodies that bind to foreign antigens are (generally) not highly cross-reactive with host proteins (although there are exceptions in autoimmune pathologies). However, application of therapeutic antibodies directed against host cell adhesion proteins faces the unavoidable dilemma that the same proteins that contribute to the disease pathology have normal physiological functions [4]. For example, nonselective blockade of immune adhesion molecules, such as integrins, can contribute to broad-based immune deficiency, necessitating a trade-off between the need for disease suppression and the potential for (severe) side effects.

### 1.3 Cell Adhesion Molecules as Therapeutic Targets

Cell adhesion is a critical component of all disease mechanisms; for example, foreign organisms must establish a foothold in the host to survive and propagate. Tumor cells use endogenous cell surface molecules to spread throughout the organism, and autoimmune disorders require cell adhesion molecules for the trafficking and accumulation of immune cells at inflammatory sites [5–7]. Microbial adhesins are inviting targets to the host humoral immune system given the functional requirement that they be displayed on the pathogen surface in an accessible conformation [8]. Autoimmune disorders as well as cancers present more difficult targets for the immune system because the host cell adhesion and migration machinery is intimately involved in disease progression but is intrinsically recognized as self, putting it beyond the reach of endogenous immunoremediation.

### 1.4 Integrins Undergo Substantial Activation-Dependent Conformational Changes

Of the major classes of adhesion receptors, including the immunoglobrin domain superfamily (IgSF), the cadherins, and other transmembrane adhesion receptors (e.g., claudins and occludins), integrins are unique in achieving rapid and reversible (second time scale) upregulation from a basal, inactive state to a maximally adhesive state, upon cell stimulation [5–7]. Other adhesion receptors require in-membrane diffusion (both active and passive) and/or regulation of cell surface expression levels to modulate the magnitude of adhesion. The dramatic increase in affinity is obtained by a series of conformational changes that increase the affinity of the integrin for ligand and enhance the accessibility of the ligand-binding site [9, 10] (Figs. 1 and 2).

A subset of anti-integrin monoclonal antibodies (mAbs) selectively recognizes the active (adhesive) conformation; some of these antibodies also simultaneously block integrin–ligand interactions [11]. In vivo, these antibodies are unlikely to bind to appropriately functional integrin since in these receptors adhesion is rapidly up- and then downregulated. By comparison, pathological adhesion is characterized by long-lived and sustained activation [12]. We have suggested, and subsequently demonstrated, that it is possible to selectively target persistently activated integrin in a physiological background [13]. In this chapter, we describe how such antibodies are derived, their mechanism of action, and their potential therapeutic application.

## 2 INTEGRIN STRUCTURE OVERVIEW

### 2.1 Integrins Are α/β Heterodimeric and Multidomain Cell Surface Molecules

Integrins are noncovalently associated heterodimers that consist of 1 (of 18) α and 1 (of 8) β subunits with 24 known combinations (excluding splice variants) [14]. Both α and β are single-pass type-I transmembrane glycoproteins with complex, multidomain extracellular architecture and a generally (excluding $\beta_4$) short cytoplasmic sequence (Fig. 1). High-resolution structural techniques [crystallography and nuclear magnetic resonance (NMR)], electron microscopy (EM), and biochemical/bioinformatics analysis (i.e., disulfide mapping and multiple sequence alignment) have been used to assemble a structural model of integrin domains and their role during the transition from the inactive to active state of the integrins [15, 16].

The β-propeller domain and the I domain in the α subunit and the I-like domain in the β subunit form a globular "integrin head" that contains the ligand-binding function (Fig. 1). The integin headpiece, which includes the integin head, the thigh domain in the α subunit, and plexins, semaphornins, and integrins (PSI) and hybrid domains in the β subunit, is connected via the integrin legpiece to the plasma membrane (Fig. 1).

### 2.2 Integrin Extracellular Domains in the α Subunit

Integrin α subunits are approximately 1000 amino acids in length, organized in 4 or 5 extracellular domains (Fig. 1).

***β-Propeller Domain*** The N-terminal domain of the α subunit is a β propeller; each blade of the propeller is a

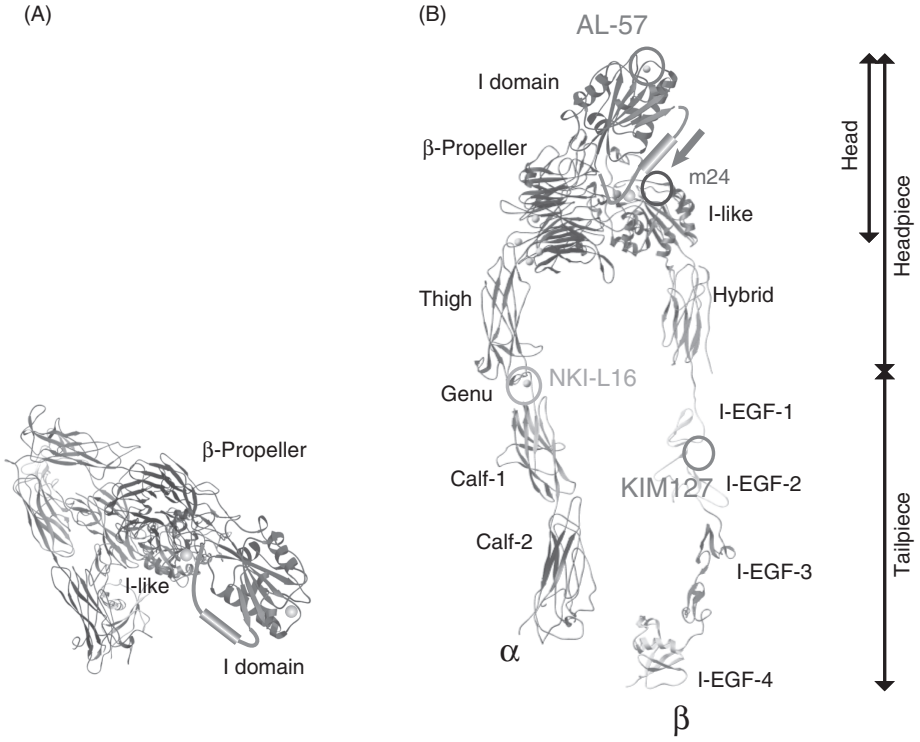


**Figure 1** Model for global conformational changes of the extracellular portion of $\alpha_L\beta_2$ integrin: (A) bent conformation (low affinity) and (B) extended conformation with open headpiece and open high-affinity I domain. The C-terminal I domain α helix is represented by a cylinder, and its downward shift upon activation is shown by an arrow. Integrin α and β subunits and integrin domains are labeled by black, location of epitopes of activation-dependent mAbs (i.e., AL-57, m24, KIM127, and NKI-L16) are shown by circles and labeled. (This figure is modified from [86].)

4-stranded antiparallel β sheet. In 9 of 18 α subunits, a domain of about 200 amino acids is inserted in a loop between blades 2 and 3 (the *I domain*). When present, the I domain forms all or part of the ligand-binding interface [17, 18].

***I Domain, A Major Ligand-Binding Domain*** The I-domain structure is a variant of the Rossmann dinucleotide binding fold; seven α helices wrap five parallel and one antiparallel β strands; the helices and strands generally alternate in the sequence topology (Figs. 1 and 2). Five amino acid side chains form a divalent cation-binding site (the metal ion dependent adhesion site, MIDAS) on the top face of the I domain. A ring of hydrophobic residues (the hydrophobic cup) surrounds the MIDAS. The I domain is joined to the β propeller on the opposite, bottom face where the N and C termini are in close proximity [10].

The I domain can assume two distinct conformations (termed *closed* and *open*) that differ in the coordination structure of the MIDAS and the relative position of secondary structural elements (Fig. 2) [10]. Compared to the closed conformation, the C-terminal helix (α7) in the open conformation is shifted downward by two turns. Also, a water molecule replaces an aspartic acid in the primary coordination sphere of a Mg(II) at the MIDAS. The loss of the direct interaction with the aspartic increases the electrophilicity of the metal. An acidic amino acid residue from a ligand or a neighboring molecule in the crystal lattice (in one structure of the αM I domain) displaces the bound water, compensating for the loss of the directly coordinating negatively charged aspartic acid. The interaction of a ligand acidic amino acid residue with the MIDAS in the open conformation is a canonical feature of integrin–ligand interactions [10].

The wild-type I domain from $\alpha_L\beta_2$ does not crystallize in the open conformation with a ligand-mimetic contact [19], consistent with the failure of the isolated I domain to constitutively bind the ligand ICAM-1 ($K_D$ I domain: ICAM-1 ~1 m*M*) [20]. Using the open structure of the $\alpha_M$ I domain as a template, an engineered disulfide bond was introduced into the αL domain sequence to stabilize the α7 helix in the open conformation [20]. Surface plasmon resonance (SPR, commonly referred to as BIACORE) was used to obtain the binding kinetics of the wild-type and locked-open I domains for immobilized ligand ICAM-1. The affinity of the $\alpha_L$ I domain for

**Figure 2** Activation-dependent conformational changes of the $a_M$ I domain. The regions with significant conformational changes are dark colored, while those with similar backbones are light colored. $Mg^{2+}$ ions (spheres) and side chains of metal-coordinating residues (ball-and-stick representations) are shown. A downward shift of the C-terminal helix (dashed arrow) and the conversion of the MIDAS to the high-affinity configuration (boxed). (This figure is modified from [87].)

the ligand increased $10^4$-fold over the wild-type sequence when the C-terminal helix is locked in the downward position by the engineered disulfide bond [20]. A crystal structure of the "high affinity" locked-open I domain reveals that movement of the C-terminal helix triggers conformational changes that were observed in the open conformation of the $\alpha_M$, including a reorganization of the coordination sphere surrounding the MIDAS and repositioning of surrounding loops to increase complementarity with the bound ligand. In the co-crystal with domains 1–2 of ICAM-1, the side-chain carboxylate E-34 of domain 1-ICAM-1 coordinates the MIDAS metal. Additional polar and electrostatic interactions surround the hydrophobic cup and make complimentary contacts on the surface of ICAM-1 [21].

***Thigh, and Calf Domains*** C terminal to the β propeller is the *thigh*, a β-sandwich domain that is topologically similar to an immunogloblin fold (Fig. 1) [15]. The *genu*, a short disulfide bonded loop that includes a calcium coordinating site is situated between the thigh and *calf-1* domain. Both calf-1 and *calf-2* domains are also β-sandwich domains. In most integrins that lack an I domain in the β propeller ($\alpha_3$, $\alpha_5$, $\alpha_6$, $\alpha_8$, $\alpha_V$, and $\alpha_{IIb}$ but not $\alpha_4$ or $\alpha_9$) there is a furin protease cleavage site in calf-2 that separates the α subunit into heavy and light chains.

### 2.3 Integrin Extracellular Domains in the β Subunit

The β subunit contains about 640 amino acids organized into 8 extracellular domains (Fig. 1).

***PSI and Hybrid Domains*** The N terminus of the β subunit is a 54-residue cysteine-rich sequence with significant homology to plexins and semaphorins and has been accordingly named the PSI domain [15]. The sequence folds into a 2-stranded antiparallel β sheet flanked by a pair of short α helices that are linked to the sheet by disulfide bonds; a short N-terminal segment may also form an α-helix. Significantly, the fourth disulfide bond of the PSI is to a stretch of amino acids that is C terminal to the adjacent hybrid domain; thus, the hybrid domain is actually inserted within the PSI sequence. The *hybrid domain* is also an immunogloblin domain.

***I-Like Domains and a Ligand Binding and/or Key Regulatory Domain*** Inserted into the B–C loop of the hybrid domain is the *β I-like (βA)* domain, which is structurally homologous to the I domain found in half of all α subunits. In addition to the MIDAS there are two additional metal coordination sites that share coordinated residues with the MIDAS and regulate the affinity of the βA domain for ligand, the adjacent to MIDAS (ADMIDAS) and the ligand-induced metal binding site (LIMBS) [22, 23].

The structure of αvβ3 bound to the cyclic pentapeptide {Arg-Gly-Asp-{D-Phe}-*N-methyl*-Val} P5 was obtained in the presence of the activating cation Mn(II), revealing many of the crucial features of integrin–ligand interactions [22, 23]. P5 binds in a crevice between the β propeller and the βA domain, burying 355Å (45% of the peptide surface area). The arginine and aspartic acid side chains of the peptide are oppositely disposed. The arginine is pointed toward the β propeller, inserting into a narrow groove formed from the D3–A3 and D4–A4 loops, held in place by a bidentate salt bridge to a pair of aspartic acids within the groove. The ligand Asp points toward the βA domain where one of the carboxylate oxygens completes the coordination sphere of the Mn(II)-bound MIDAS. While the ligand arginine is partially solvent exposed, the ligand aspartic acid is completely buried in the complex, with the second oxygen forming a network of polar interactions with backbone amides and the aliphatic side chain making hydrophobic contacts with protein side chains. The glycine sits at the

interface between α and β subunits, making hydrophobic contacts with residues in the α chain. Notably, the bond conformation of the RGD segment is similar to free conformation of the natural ligand echistatin, validating the physiological relevance of the structural analysis of the small-molecule ligand mimetic. The two additional residues in the peptide point away from the α/β interface.

The βA domain undergoes structural changes homologous to the α I domain upon binding ligand including a repositioning of the β6–α7 loop and downward piston-like motion of the C-terminal α7 helix. As a result of the downward motion, the hybrid and PSI domains are swung away from the α/β interface by ~62°. The motion of the hybrid domain is presumably the mechanism by which conformational changes in the ligand-binding site are propagated to C-terminal domains [23].

***Integrin–Epidermal Growth Factor (EGF) (I-EGF Domains)*** C terminal to the βA domain, four cysteine-rich segments fold into *EGF domains 1–4* (Fig. 1). The canonical EGF domain is 30–40 amino acids, with two β sheets, the major sheet at the N terminus and the minor sheet at the C terminus, and three disulfide bonds, two of which (at the N terminus) are interleaved in the sequence [24]. The integrin EGF repeats differ in having a fourth intradomain disulfide bond (all I-EGF-1 and I-EGF-3 of $\beta_4$ and $\beta_8$ also have a distinct arrangement of disulfide bonds from other I-EGF domains) [25].

***The β Tail Domain*** The *beta terminal domain (βTD)* is the most C-terminal extracellular domain of the β subunit, just prior to the transmembrane domain. The βTD is a novel fold, an N-terminal α helix facing a mixed parallel and antiparallel four-stranded β sheet.

### 2.4 Transmembrane Domains of α- and β Subunits

The *integrin transmembrane domains* are the critical point in transmitting extracellular ligand binding across the membrane into the cytoplasm. Solution NMR studies of the isolated $\beta_3$ transmembrane domains in detergent micelles suggest a structure that is largely helical [26]. A specific interface between $\alpha_{IIb}$ and $\beta_3$ in intact cell membranes was identified using disulfide cross-linking [27]. Sequence homology with other integrin transmembrane domains suggests that the interface is a conserved feature of all integrins in the resting state.

### 2.5 Cytoplasmic Domains of α and β Subunits

The *cytoplasmic tail domains* ("tails") are generally short; for example, the $\beta_3$ tail is 47 amino acids (aa) in length; the 2 associated α subunits ($\alpha_{IIb}$ and $\alpha_V$) are 20 aa and 32 aa, respectively. The cytoplasmic tail complex of $\alpha_{IIb}\beta_3$ has been studied by NMR spectroscopy and remains the only integrin tail structure to be determined at atomic resolution [28]. The $\alpha_{IIb}$ tail forms a short N-terminal α helix followed by a turn and a C-terminal loop that folds back onto the positive charged N-terminal helix. The structure of the $\alpha_{IIb}$ tail is similar in the absence and presence of $\beta_3$ tail. The $\beta_3$ tail also forms an N-terminal α helix; the remainder of the tail has no persistent structure, although the C-terminal N-P-L-Y sequence transiently populates a turn. The interface between the two tails is composed of both hydrophobic and electrostatic interactions contributed by the membrane-proximal helices. Most α subunits have a conserved G-F-F-K-R sequence in the juxta-membrane region (although in α10 this is replaced by a G-A-H-K-K sequence). $\alpha_{IIb}$-R-995 forms a conserved salt bridge with $\beta_3$-D-723 that plays a crucial role in restraining integrins in the inactive state [29]. Multiple serine, threonine, tyrosine (and possibly histidine) residues on both α and β tails are available for phosphorylation, which both regulates the interaction between the tails and the avidity of the tails for other cytoplasmic proteins [30].

### 2.6 Integrin Structure and Activation Mechanism

The integrin heterodimer is generally divided into two distinct regions. The head (or headpiece) is formed from the α-subunit propeller and I domain and the β-subunit βA and PSI domains. The legs (sometimes known as the tailpiece, not to be confused with the cytoplasmic tails) is formed from, the calf-1, calf-2, and thigh domains of the α subunit and the EGF and β-tail domains of the β subunit.

In the crystal structure of $\alpha_V\beta_3$, the integrin is folded into a V-shaped conformation with a 135° angle between the head and tail domains [22]. The bend occurs between the thigh and calf-1 domains in the α chain and between the I-EGF-1 and I-EGF-2 domains in the β subunit (specifically, at the genu). As a result, the ligand-binding head is pointed toward the cell surface where it is not accessible to ligands in the extracellular matrix or on the opposing cell surface. The bend also results in the formation of extensive buried and partially buried inter- and intrasubunit interfaces and restricts the flexibility of interdomain motions, such as the ability of the hybrid domain to attain the swung-out conformation [31].

The *switchblade model of integrin activation* proposes that integrins exist in two global conformations: bent (as observed in the crystal structure of $\alpha_V\beta_3$) and extended (Fig. 1) [15, 25]. The bent conformation is stabilized by a number of intra and intersubunit interfaces. Prior to activation, the two conformations are in exchange, with the equilibrium shifted heavily toward the bent

conformation. Dissociation of the cytoplasmic domains occurs upon binding of cytoplasmic proteins (such as talin) to the tails, triggering separation of the transmembrane domains and the interface between the legs. Separation of the α/β legs further destabilizes the interface between the legs and the headpiece domains, permitting the receptor to attain the fully extended conformation. In the extended conformation, the head domains are well disposed to bind ligands. Once the integrin is extended and the legs fully separate, the hybrid domain can swing out, maximally activating the integrin for ligand binding. The switchblade model accounts for the binding of many conformationally selective mAbs since the binding epitopes for many of these mAbs are buried (or partially buried) in the bent conformation but become exposed in the extended and fully activated (extended with hybrid domain out) conformations [25, 31].

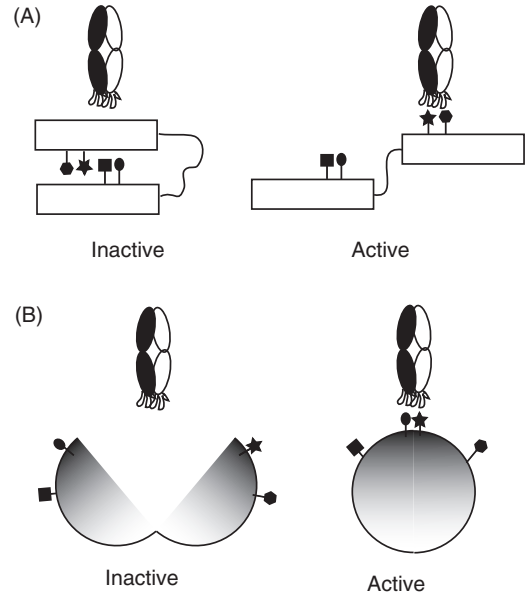


**Figure 3** Class I (A) and class II (B) conformational selective antibodies. (A) Antigenic epitopes are shielded from binding in one conformation but become (more) accessible after activation. (B) Antigenic epitopes that are formed as a result of conformational change(s).

## 3 CONFORMATION-SELECTIVE ANTI-INTEGRIN MABS

The transition from the bent to extended (and then fully activated) conformation integrins requires disruption of protein–protein interfaces between domains (i.e., separation of the α and β legs) and allosteric motions within domains (i.e., conversion of the closed-to-open conformation of the I domain). In this section, we review select members of two classes of mAbs defined by epitopes that are gained (or lost) via each transition. We specifically emphasize properties of each antibody pertinent to evaluating their potential as therapeutic agents.

### 3.1 Class I Conformation-Selective Integrin Antibodies

***Class I*** The first class of mAbs binds to pre-existing epitopes that are shielded in one conformation but become (more) accessible in a second conformation (Fig. 3). mAbs in the integrin legs that become exposed upon α/β dissociation and receptor extension are of this type. Most such antibodies are in the cysteine-rich region of the β subunit, although the binding of anti-$\alpha_L$ genu mAb NKI-L16 and the anti-α5 calf-1/calf-2 mAb SNAKA51 are enhanced in activating conditions. The swing-out motion of the hybrid domain also differentially exposes epitopes on the protein surface.

***Monocromal Antibodies (mAbs) Binding to the Integrin Legpiece*** KIM127, NKLI-16, and SNAKA51 are three activation-sensitive mAbs that bind to epitopes in the integrin legs. KIM127 reacts preferentially with free (unpaired) integrin $\beta_2$ subunits and can promote binding by $\beta_2$ integrins to surface-immobilized ligand. The binding of KIM127 is temperature sensitive with considerably enhanced binding at 37° versus 4° [32, 33]. The binding of KIM127 to $\alpha_L\beta_2$ with activating mutations (i.e., mutation or truncation of the conserved G-F-F-K-R sequence in the cytoplasmic tail) is significantly upregulated. Addition of the activating reagents phorbol myristyl acetate (PMA) or Mn(II) also leads to increased KIM127 exposure. This illustrates a critical property of many activation-dependent antibodies. Due to the equilibration between bent and extended conformations, an activation-sensitive epitope will be transiently exposed; consequently, most activation-dependent mAbs have some potential to activate integrins.

The modest activation potential of KIM127 is explained by the position of the binding epitope. The epitope for KIM127 is located in a long loop at the C terminus of the I-EGF2 domain in the β stalk and contains the residues G-504, L-506, and Y-508 [25, 34]. Structurally adjacent to this loop are residues that restrain activation by $\beta_2$ integrins by hydrophobic packing interactions with the α subunit. I-EGF2 and I-EGF-3 are wedged into a bend formed at the genu with the KIM127 epitope located between the α-subunit thigh domain and the calf-1 domain. The tight packing of the interface effectively blocks binding of KIM127 prior to separation of the integrin legs. The epitope for the KIM185, an antibody with greater activating poten-

tial, includes residues in the I-EGF4 domain that are partially accessible to solvent in the bent conformation. The greater exposure of the KIM185 interface permits rapid binding of the mAb during the transient exposure in the equilibration between bent and extended conformations. Once bound, the mAb serves to block reclosure of the interface, which favors conversion to the extended conformation.

Most activation-dependent antibodies recognize epitopes on the β subunit. However, the mAb NKI-L16 [35, 36] and SNAKA51 [37] bind (preferentially) to the extended conformation of $\alpha_L$ and $\alpha_5$, respectively. In the basal, inactive state NKI-L16 binding to the $\alpha_L$ subunit is absolutely dependent on Ca(II) in a dose-dependent manner at concentrations between 0.1 and 10 m*M* Ca(II). The binding of NKI-L16 is enhanced upon addition of drug XVA143, which induces the extended conformation of αL. Residues on the inner face of the thigh domain, buried in the closed conformation by the calf-1 domain in the α subunit and the I-EGF2 and 3 in the β subunit, are required for binding of NKIL-16. Two absolutely critical residues, A-602 and W-738, are proximate in the three-dimensional structure of the thigh domain, close to the genu. Residues forming the Ca(II) coordination site in the genu and calf-1 domains are also required for binding. However, there are no human–murine substitutions within the highly conserved genu, and it is not clear if the Ca(II) is explicitly part of the binding site or merely stabilizes the conformation recognized by the mAb.

The binding of SNAKA51 requires both the calf-1 and calf-2 domains. SNAKA51 stimulates adhesion of $\alpha_5\beta_1$ to fibronectin fragments in a dose-dependent manner. Conversely, SNAKA51 reports on integrin activation in the presence of increasing concentrations of fibronectin ligand. Notably, truncation of the $\beta_1$ subunit tail does not affect mAb binding; however, the ability to activate $\alpha_5\beta_1$ by SNAKA51 is lost when the integrin β1 tail is truncated. This suggests that SNAKA51 activates by binding to the transiently opened leg conformation and acts as a wedge to keep the α and β legs separate, promoting integrin extension.

***mAbs Binding to the Hybrid and I-Like Domains*** The binding of mAbs to epitopes that lie directly within, or adjacent to, the hybrid domain can modulate, or be modulated by, integrin activation. The epitope of SG/19 a function blocking anti-β1 mAb, for example, includes T-82 in the hybrid domain (Fig. 1), on the fully exposed side of the domain opposite the α/β interface [38]. The epitope of SG/19 is close, in a three-dimensional model of $\beta_1$, to the binding site of SG/7 and mAb13, both of which also function as blocking anti-$\beta_1$ mAbs. Binding of this group of mAbs is significantly attenuated in a β1 mutant with the hybrid domain locked in the active, out conformation. Negative stain EM images of Fab SG/19 bound to a recombinant integrin headpiece fragment conform that Fab binds to the boundary between the βA and hybrid domains on the opposite side of the α/β subunit interface, including segments of both the hybrid and βA domains. The binding interface is incompatible, due to steric interactions, with the swung-out conformation of the hybrid domain, confirming the conclusion that this class of antibodies works by inhibiting this pivotal motion.

A number of $\beta_2$-function blocking mAbs (i.e., GFR1, YFC51.1, YFC118.1, CLB LFA1/1, and CLB 54) map to epitopes that are buried or partially buried in the residues K-133, H-332, and Y-339, which are in a similar position as the SG/19 class of anti-β1 mAbs, suggesting that they might inhibit $\beta_2$ integrin function by a similar mechanism [39].

The epitopes for a reciprocal group of mAbs that includes 15/7 and HUTS-4 involve a triplet of spatially proximate residues (S-370 and E-371 on the C–D loop, and K-417 on the adjacent E–F loop) that are within the inner face of the hybrid domain, at a small interface with the α-subunit β propeller [40]. In the inactive conformation, K-417 is accessible (to a 20-Å sphere) but both S-370 and E-371 are buried. Upon activation [by divalent cations such as Mn (II) or activating mutations] the antibody epitopes become exposed. A similar activation reporter mAb, MEM148, is located on the inner face of the $\beta_2$-subunit hybrid domain [41]. In the inactive conformation, the most critical residue in the epitope, P-374, is located at a three-way interface between α and β subunits as well as the α/β headpiece/tailpiece interface (residues H370-Q373 also form part of the epitope). The presence of hybrid domain epitopes that both support and suppress integrin activation are a function of the role of the hybrid domain as a nexus of conformational change between the ligand binding and leg domains.

### 3.2 Class II Conformation-Selective Integrin Antibodies

***Class II*** The second class of mAbs is to antigens with epitopes formed as a result of conformational change(s) bring spatially distant amino acid side chains within the 100-Å footprint of the 6 mAb complementary determining region (CDR) loops (Fig. 3). mAbs to integrin I and βA domains (Figs. 1 and 2) are typically class II antibodies with epitopes that are formed (or lost) due to the intradomain allosteric transition. The extent of this conformational preference is directly related to the extent to which the binding epitope of the antibody binding epitope lies within more plastic segments of the I domain. Generally, since the closed conformation is the

default state of the I domain, most antibodies have weak to modest selectivity that disfavors binding to the open conformation (by about 2- to 10-fold). For example, the mAb HI111 binds the $\alpha_L$ I domain with a $K_D$ about 5-fold lower for the closed than the open conformation (as determined by SPR analysis of soluble αL I domain binding to randomly immobilized mAb) [42]. The epitope of HI111 includes K-268 at the top of the α5 helix, a residue located near the MIDAS, which undergoes substantial conformational change in the closed-to-open conformation.

***mAbs Binding to the $\alpha_M$ I Domain*** A number of conformationally selective mAbs that bind to the $\alpha_M$ I domain have been obtained using the immunization–hybridoma technique. Conformationally selective mAbs can bind to—and inhibit the activity of—the I domain either by direct competition with ligand or by inhibiting requisite conformational changes, stabilizing the closed conformation. Two mAbs that illustrate this property are mAb-107 and 44a.

*mAb-107* The mAb-107 is an antibody that was identified from a hybridoma screen for mAbs with divalent cation dependence for binding to the isolated $\alpha_M$ I domain [43]. In addition to blocking binding of $\alpha_M\beta_2$ (and the isolated I domain) to soluble and immobilized ligand, mAb-107 inhibits spreading and opsonization by neutrophils, both $\alpha_M\beta_2$-dependent processes. The $K_D$ of $\alpha_M$ I domain is about sevenfold higher for $\alpha_M$ in the open conformation than for the wild-type I domain. Of the residues found in the binding epitope (G-143, E-178, E-179, R-208, and F-246) only F-246 (equivalent to T-243 in $\alpha_L$) shifts position in the high-affinity conformation, and it is presumably through this residue that the conformational selectivity of the mAb is obtained.

*44a* The antibody 44a has even higher selectivity for the closed ($K_D$ 4.4 n*M*) versus open conformation (860 n*M*) of $\alpha_M$ [44]. The epitope for the mAb is distinct from the MIDAS binding site, mapping to residues M-153–T-159 (α1 helix), E-162–L-170 (β2–β3), and Q-309–E314 (α7). The C-terminal α7 helix undergoes an obligate conformational change between closed and open conformation, as described in the structural overview, and the ability to block ligand blocking by 44a likely involves allosteric inhibition of this transition.

*CBRM1/5* Although most mAbs bind with (at least modest) selectivity to the closed, default conformation, a subset preferentially recognizes the open I domain conformation. CBRM1/5 binds selectively to the open (active) conformation of the $\alpha_M$ I domain and blocks ligand binding [45]. The mAb was isolated from mice immunized with intact, purified $\alpha_M\beta_2$. CBRM1/5 was identified using a screen for mAbs that bound to PMA-activated, but not resting, neutrophils. The total number of epitopes on an N-formyl-methionine-leucine-phenylalanine–stimulated (fMLP-stimulated) neutrophil is about 10% of that obtained from a non-conformationally selective mAb (30% of neutrophils activated by PMA), suggesting that only a fraction of the receptors on the surface are in the active conformation after stimulation. Although CBRM1/5 binds to only a small fraction of αM on the cell surface, saturating concentrations of mAb completely block binding to ligands iC3b and fibrinogen.

The mAb binding epitope is formed from P-147, H-148, R-151, K-200, T-203, and L-206 [46]. The first three residues are in the loop and adjacent N-terminal part of α1 helix preceded by residues that form part of the MIDAS metal binding site. The helix and loop undergo an inward (toward the center of the I domain) motion during activation to reposition the MIDAS residue S-144 for ligand binding. The second group of three residues is in the loop connecting α3 and α4 and differs little in position in either closed or open conformation. However, all six residues are adjacent in the $\alpha_M$ I domain structure. The function blocking mAbs CBRM1/21 and CBRM1/22 bind to an overlapping epitope with CBRM1/5 on the MIDAS face. However, neither 1/21 nor 1/22, which were not explicitly selected to bind to activated neutrophils, are activation-specific antibodies, and neither incorporates R-151 in the binding epitope, suggesting that a single amino acid in the epitope is sufficient to account for the conformational selectivity of CBRM1/5 (although other amino acids, which do not differ between mouse and human $\alpha_M$, may contribute to this property as well).

*Class II $\alpha_M$ I Domain mAb Generated by Phage Display* Activation-specific scFv to the αM I domains were obtained using two fully human phage display libraries [47]. The first library was constructed from reverse transcription (RT) of ribonucleic acid (RNA) collected from peripheral blood lymphocytes of five donors and spleen biopsy material of an additional six donors. Variable heavy (VH) and variable light (VL) immunoglobulin domains of antibodies were amplified and randomly reconstructed by overlap extension polymerase chain reaction (PCR). The final library contained ~$2.0 \times 10^9$ clones. The second library (synthetic) was constructed using two human scFv as frameworks (one against estradiol, the second against hepatitis surface B antigen). The H-CDR3 of both scFv was replaced by the sequence C A K/R $X_{4\text{-}7}$ F E/D Y. The VH library contained ~$10^8$ clones. Additional diversity was added by grafting H-CDR-3 sequences from the natural library spleen and peripheral blood lymphocytes (PBLs) complementary deoxyribonucleic acid (cDNA) onto both frameworks to give four additional

libraries of ~$1 \times 10^8$ total clones. The VL library was similarly constructed by replaced L-CDR-3 of the estradiol scFv with the sequence Q V W $X_{4\text{-}7}$ V V and the hepatitis surface B antigen library with the sequence N S R $X_{4\text{-}6}$ V L. A total of ~$6.0 \times 10^7$ clones were obtained from the two libraries. Four additional VL sublibraries were created using PBL and spleen DNA, as was done for the VH, giving an additional ~$2.0 \times 10^7$ clones.

A double-subtraction panning strategy was performed to isolate scFv with the desired properties. The libraries were depleted on resting monocytes and enriched via binding to activated monocytes (which express high affinity $\alpha_M\beta_2$). Next, the library was further enriched for high-affinity binders via incubation with CHO cells transfected with the high-affinity G-F-F-K-R truncation mutant. This step was also subtractive as it depleted phage scFv clones that bound to other activation-specific markers on monocytes. One activation-specific scFv was obtained from the natural library (MAN-1 [monoclonal antibody natural-1]) and two from the synthetic library (MAS-1 [monoclonal synthetic-1] and MAS-2 [monoclonal synthetic-2]).

Activation-specific binding to $\alpha_M\beta_2$ for all three scFv was demonstrated using fluorescent activated flow cytometry (FACS) of CHO cells expressing wild-type (no binding) or G-F-F-K-R truncation mutation (binding) $\alpha_M\beta_2$. In addition, all three antibodies bound activated but not resting monocytes. Furthermore, MAN-1 was shown to inhibit $\alpha_M\beta_2$ binding to immobilized heparin, fibrinogen, and ICAM-1, although not iC3b (either soluble or immobilized) although a nonactivation-specific mAb could block adhesion to all four ligands. MAN-1 also blocked adhesion of PMA-activated but not unstimulated monocytes to immobilized fibrinogen in shear flow conditions. The ability to block some, but not all, $\alpha_M$–ligand interactions supports the mosaic model of ligand recognition, which accounts for the ability of $\alpha_M$ to bind a diverse array of structurally distinct ligands by the presence of distinct (but overlapping) binding sites on the MIDAS face. The ability to block preactivated but not unstimulated (wild-type) $\alpha_M$ binding suggests that the antibody does not efficiently compete with binding in the context of a receptor that rapidly undergoes transition to the active conformation after encountering ligand.

The MAN-1 epitope was located to the I domain by demonstrating the ability of the peptide derived from the fibrinogen binding surface on the I domain to compete with the I domain for mAb binding. However, no dual mAb competition or epitope mapping experiments were performed, and it is unclear if the peptide is structurally homologous to the authentic ligand-binding surface. Furthermore, divalent cation-dependent binding was not demonstrated and presumably is not a characteristic feature of the mAb antigen interaction. The paratope of all three scFv H-CDR-3 was mapped using alanine-scanning mutagenesis. In MAN-1, mutation of the first and last two residues resulted in a small decrease in binding, possibly due to loss of interactions with the flanking framework residues. A more significant decrease was observed with mutation of hydrophobic residues in the center of the loop (W in MAN-1 and MAS-2 and L in MAS-1). This suggests the importance of complementary hydrophobic surfaces in antigen recognition, although it is unclear if H-CDR-3 explicitly encompasses the activation-specific binding surface for αM.

***Some Class II I Domain mAbs Represent Ligand Mimetics*** A subset of this class is made up of *ligand mimetic antibodies*, which are defined by three requisite properties: (1) sharing a common binding site with the authentic ligand, (2) binding is divalent cation dependent, and (3) binding is activation dependent. We recently described AL-57, a ligand mimetic antibody, isolated by phage display to the active conformation of $\alpha_L$. Examples of ligand mimetic antibodies to integrins that do not contain the α I domain are the antihuman $\alpha_{IIb}\beta_3$ PAC-1 [48] and antihuman $\alpha_V\beta_3$ WOW-1 [49]. The H-CDR-3 of PAC-1 contains a R-Y-D sequence that is critical for binding, presumably through similar interactions as the canonical integrin-binding sequence R-G-D. WOW-1 was engineered by replacing the H-CDR-3 of PAC-1 with an integrin-binding loop from the adenovirus penton base protein.

***Class II I Domain mAbs to the $\alpha_L$ I Domain Generated by Phage Display*** mAbs selective for the active conformation of $\alpha_M$ have been obtained both by immunization–hybridoma and library (phage) screening, using either wild-type or high-affinity $\alpha_M\beta_2$ as antigen or bait. By comparison, there are no known mAbs selective for the high-affinity conformation of $\alpha_L$ identified by hybridoma screening. We collaborated with DYAX Inc. to obtain an active conformation-selective mAb to the high-affinity $\alpha_L$ I domain using a human phage display library of $1 \times 10^{10}$ unique Fab clones. (Fab-300). The rationale behind the construction of the library was to create a set of optimized antibody sequences that closely resemble the sequence and length diversity found in natural high-affinity mAbs *after* somatic hypermutation [50]. Specifically, most antibodies selected from libraries derived from unrearranged DNA (naïve libraries) or purely synthetic sequence are low affinity due to the restricted size of the library and limited diversity in both the length of the CDR and in key antigen paratope residues. These mAbs often require protracted affinity maturation protocols, as previously described, to obtain optimal antigen–antibody contact.

The library was constructed as a chimera of human immune cell DNA and synthetic sequence diversified in key positions of H-CDR-1 and H-CDR-2. The VH was constructed on the VH3-23 framework, which is commonly found in rearranged antibodies, and is well expressed and capable of pairing with all light chains. In a structural model of H-CDR-1, positions 31, 33, and 35 (Kabat standard numbering) were found to point toward the antigen-binding site. Each of these three positions was diversified to allow 19 or 20 amino acids (excluding cysteine), imparting greater diversity than that available in the germline sequence. In VH2, there is extensive variability in positions 50, 52, and 52a; amino acids 56 and 58 are often found to vary due to somatic hypermutation. In position 50 and 52, the diversity was limited to tyrosine, arginine, typtophan, valine, glycine, and serine. At position 52, only proline and serine were allowed (these are the highest frequency amino acids at this position). All amino acids except cysteine were allowed at positions 56 and 58.

The variability in sequence and length of H-CDR-3 is difficult to produce using synthetic sequence. Accordingly, the Fab-300 library design employs H-CDR-3 sequences captured from human autoimmune patient B cells expressing IgM. The H-CDR-3 length distribution of 4–26 residues is comparable to that found in naturally isolated antibodies. The sequence diversity was also comparable to authentic mAbs with tyrosine, aspartic acid, glycine, and serine being most common. Complete λ and κ light-chain repertoires were also obtained from autoimmune patient DNA.

A Fab specific for the high-affinity conformation of the $\alpha_L$ I domain was obtained from the Fab-300 library, using the soluble locked open $\alpha_L$ high-affinity mutant as bait [51, 52]. The I domain was randomly biotinylated on lysine residues and presented in solution for capture by streptavidin. The library was depleted using the soluble (inactive) wild-type $\alpha_L$ I domain as decoy bait (negative selection) to avoid selection of clones that bound to both conformations. One clone, AL-57, was identified after three rounds of selection and further characterized.

Using flow cytometry (FACS), we confirmed that AL-57 dose dependently bound to cells expressing $\alpha_L\beta_2$ with the I domain in the locked open conformation, but not wild-type $\alpha_L\beta_2$ [52]. IgG4-AL-57 (used to avoid competing binding with cell surface Fc receptors) bound to activated [10 m*M* Mg(II)/2 m*M* ethylene glycol tetraacetic acid (EGTA)] PBMC but not peripheral blood mononuclear cell (PBMC) in Ca(II) containing buffer; considerable upregulation of AL-57 binding was observed when PBMC was treated with the activating agent PMA. Transient upregulation of AL-57 binding was observed upon treatment of T-lymphocytes with the chemokine CXCL12. The total number of AL-57 epitopes was about 10% of the total number $\alpha_L\beta_2$ (determined using a nonconformational selective antibody) and was comparable to the number of KIM127 epitopes observed in the same conditions.

We sequenced the heavy and light chains of AL-57. The H-CDR-3 is 11 amino acids long and has some homology with the MAS/MAN series of anti-$\alpha_M$ activation-specific antibodies. We performed partial paratope and epitope mapping of AL-57, mutating each aspartic acid residues in each of the six CDR loops to alanine. Mutation of D101 in H-CDR-3 completely abolished AL-57 binding to activated $\alpha_L\beta_2$, while mutation of any other aspartic acid residue in the six CDRs had no effect on integrin binding. In addition, the binding of AL-57 to $\alpha_L$ was completely abrogated in the presence of ethylenediaminetetraacetic acid (EDTA). Taken together, this suggests that AL-57 is a member of a subclass of activation-dependent antibodies that functions as a ligand-mimetic, coordinating a MIDAS-bound Mg(II) with an aspartic acid in H-CDR-3. Significantly, the aspartic acid is part of a short motif, D–F-W conserved in MAN-1 and partially conserved in MAS-1 and MAS-2. However, mutation of the aspartic acid residue in the $\alpha_M$ mAbs had a minor, albeit detectable effect, on binding. Possibly, the $\alpha_M$ conformational selective mAbs have a weak dependence on binding to the MIDAS metal, although no divalent cation dependence for these antibodies was reported. Interestingly, there is a second conserved motif, I-G-S in L-CDR-1 shared between AL-57 and MAN-1.

We used SPR to determine the binding kinetics of AL-57 to soluble high-affinity, intermediate-affinity, and wild-type low-affinity I domain; for comparison, we also studied the kinetics of the conventional, activation-insensitive antibody MHM24. To avoid competing avidity effects, both antibodies were immobilized on the sensor surface over which we flowed high-affinity I domain at various concentrations. The binding of AL-57, but not MHM24, is divalent cation dependent, as we observed no binding of AL-57 to the I domain in the presence of EDTA. Notably, the off rate ($k_{off}$) of AL-57 is markedly slower than that of MHM24 ($k_d = 5.25 \times 10^{-5}$ versus $1.35 \times 10^{-3}\,s^{-1}$). This advantage in dissociation kinetics is offset by an association rate that is 2 orders of magnitude slower ($2.12 \times 10^3$ versus $2.16 \times 10^5\,M^{-1}\,s^{-1}$). Furthermore, while MHM24 is bound to each I domain with comparable affinity, we observed no binding of wild-type I domain to AL-57. Significantly, a mutant I domain that binds to ICAM-1 with micromolar affinity also bind at comparable strength to AL-57. This validates our hypothesis that AL-57 as a ligand mimetic that binds to the cell surface in proportion to the affinity state of cell surface LFA-1.

We compared the binding of soluble multivalent ICAM-1-Fcα to WT LFA-1 activated with Mg/EGTA and CBRLFA1/2 and LFA-1 containing the disulfide locked high-affinity I-domain by FACS in the presence of AL-57 or the blocking, conformation-insensitive antibody MHM24. Both antibodies were capable of antagonizing the binding of multivalent ICAM-1, although the $K_i$ of MHM24 was about twofold lower (WY, SP, NSA, TS, and MS, unpublished observations).

### *Class II mAbs to the β I-Like Domain*

*mAbs to the βA Domain* The βA domain and I domain share a common allosteric mechanism accordingly, or indirectly, via inhibition of critical allosteric motions. A particularly interesting class of mAbs is ligand mimetic to the $\beta_3$ integrins. Representative allosteric and ligand mimetic mAbs are presented below.

*12G10* The mAb 12G10 binds to the βA of $\beta_1$ integrins; binding is stimulated by Mn(II) or Mg(II), and inhibited by Ca(II) [53]. The affinity and binding of 12G10 is increased upon ligand (fibronectin or fibronectin-derived peptide sequence) binding to $\alpha_5\beta_1$, making 12G10 a prototypical ligand-induced binding site (LIBS) antibody. The MIDAS metal is required for high-affinity mAb binding since the binding to the MIDAS mutant D130A was comparable to the wild-type holoreceptor in the absence of cations; however, the MIDAS mutant was more sensitive to inhibition of 12G10 binding by Ca(II). The 12G10 epitope lies within residues 207–218 in the α2 helix βA, a region that is shared by other $\beta_1$ activation-sensitive antibodies. 12G10 is distinguished from other mAbs that bind to this region by the modulation of binding by divalent cations [54]. The divalent cation sensitivity is conferred by a unique contribution of the epitope formed from R-154 and R-155 in the α1 helix. While mutation of flanking residues (N-151, M-153, I-156) did not affect the binding of 12G10, mutation of either arginine residue reduces the binding by 65 and 30%, respectively (the binding of mAb to the double mutant is attenuated by >80%). There is, however, no effect of either single or combined arginine mutations on other activation-sensitive mAbs binding to the 207–218 segment. Since the arginine residues are structurally proximate to residues 207–218, they likely represent part of the 12G10 binding interface on the integrin, although an indirect effect on mAb binding cannot be excluded. Inhibition of ligand binding by 12G10 presumably occurs by blocking an obligate translational motion of the α1 helix that occurs during transition to the open βA. Significantly, the mAb24 epitope of $\beta_2$, which becomes exposed upon activation of Mn(II), or activating mutations, includes R-122 in the α1 helix, which may be sensitive to the same conformational change that occurs in the $\beta_1$ βA domain [55, 56].

*PAC-1* Ligand-mimetic antibodies to βA (specifically $\beta_3$) have been obtained, through hybridoma screening and phage display. The first such antibody identified, PAC-1 is an IgM (pentavalent) that binds selectively to $\alpha_{IIb}\beta_3$ on activated (i.e., adenosine diphosphate (ADP) stimulated), but not resting platelets [57]. The Fab of PAC-1 retains the same binding specificity, although the binding of the pentavalent form is tighter due to avidity effects [48]. Binding is inhibited by RDS peptide and in the absence of divalent cations. The H-CDR-3 of PAC-1 contains the sequence RSPSYY**RYD**GAGPYYAMDY. Binding of PAC-1 to $\alpha_{IIb}\beta_3$ is critically dependent on the central RYD; even the relatively conservative mutation D → E abolishes ligand binding. The molecular presentation of the RYD is also critical since inversion (RYD → DYR) also eliminates the functionality of the mAb. However, the central tyrosine can be mutated to a G (giving a canonical RGD sequence) without loss of affinity or specificity. This is significant since several other mAbs that recognized $\beta_3$ integrins with an RGD sequence show no such specificity between $\alpha_V$ and $\alpha_{IIb}$.

*OP-G2 and LJ-CP3* The mAbs OP-G2 and LJ-CP3 also bind to $\alpha_{IIb}\beta_3$ through an RYD in the H-CDR-3; however, unlike PAC-1, the two IgGs do not require preactivation for binding [58]. All three RYD containing mAbs use the same D gene (DSP 2.0 encoding YYRYD) and $J_H$ (JH4); the remainder of the CDR sequences are diverse with ~40% homology between each of the three VH sequences. The light-chain homology is higher with both OP-G2 and LJ-CP3 using the same VK segment ($V_{K5}$) and PAC-1 using the $V_{K1C}$ (~60% sequence conserved in the VK compared to OP-G2 and LJ-CP3). However, the D gene usage is not the sole determinant of integrin binding since mAbs HP20 and PCG1-1, neither of which binds to integrins, use the same D segment [59]. Like PAC-1, the central tyrosine can be replaced with a glycine with no change in specificity, affinity, or requirement for preactivation.

A prominent feature in the crystal structure of OP-G2 is the long, extended H-CDR-3 loop protruding away from the immunoglobin surface, ~12 Å from the common antibody–antigen recognition site. The RYD sequence is located at the tip of the protrusion; the absence of any defined cavity at the antigen-binding site suggests that the mAb interacts with $\alpha_{IIb}\beta_3$ by the extension of the loop into a binding pocket on the integrin. Superposition of the Cα of VH and fibronectin demonstrates that the conformation of the loop is similar to that of natural RGD containing ligands. The RYD sequence itself takes on two distinct conformations in

the crystal; in conformation 1 the critical aspartic acid is direct outward from the loop and may represent the ligand-binding conformation, while in the second conformation the Asp is bound to a solvent and may be the free solution conformation of the loop [60].

A high-resolution crystal structure has not been determined for LJ-CP3 or PAC-1, so it is not possible to conclusively describe the origin for the similarities and differences in activation based purely on structural constraints. However, the number and composition of the amino acid sequence flanking the central RYD are different in PAC-1 and OP-G2 and likely account for the difference in sensitivity to activation. Two H-CDR-3 chimeric derivatives of the RGD analogies of PAC-1 and OP-G2 were created around a central RGD sequence (PAC-RGD-OP and OP-RGD-PAC) and the binding specificity compared to the native sequences (OP-RGD-OP and PAC-RGD-PAC). Sequence N terminal to the RGD determined the activation selectivity (≥10:1) of the chimeric mAbs; PAC-RGD-OPG bound only to activated platelets while OPG-RGD-PAC bound equally well to active or inactive platelets [61].

Differences in the specificity for activation may be due to distinct contacts between the paratope and antigen. Alternatively, the specificity (or lack thereof) may arise due to constraints on the loop conformation that differ in the two mAbs, resulting in alternate presentation of the central RGD motif, as the N-terminal PAC-1 sequence is six residues as opposed to four for OP-G2 [61]. Difference in the length and composition could affect both the local conformation of the loop and the angle of its presentation with respect to the body of the immunoglobin [59]. The additional amino acids residues in PAC-1 may shift the RGD motif from the center of the CDR loop and move it closer to the antigen-binding face of the VH. Significantly, the binding of $\alpha_{IIb}\beta_3$ to peptides decorated on plastic beads was independent of activation state for peptides with long spacer arms but explicitly required activation for peptides with short spacer sequences [62]. Moving the central RGD off the center of the CDR loop would have a similar effect by lengthening the distance between the RGD and the binding site, imposing an activation requirement for mAb binding [61].

*Class II mAbs to the β I-Like Domain Generated by Phage Display* Ligand-mimetic mAbs to $\alpha_{IIb}\beta_3$ have also been obtained using phage display [63]. Differential subtractive selection was used to (separately) pan a pair of libraries ("natural" and "synthetic") as earlier described for $\alpha_M\beta_2$ [64]. The integrin was presented on two distinct backgrounds, ADP activated platelets and in a constitutively active form (G-F-F-K-R truncation mutant) on the surface of CHO cells, to eliminate binding to cell-type-specific proteins. Negative selections were performed using nonactivated platelets and wt–$\alpha_{IIb}\beta_3$ expressed in CHO cells. Elution of bound phage from cells was performed either using a nonselective low pH protocol (which dissociates phage scFv from target regardless of specificity) or saturating concentrations of blocking mAb c7E3. Using these protocols, two nonactivation-specific and six activation-specific clones were identified. Based on analysis of the CDR sequences of the recovered clones, it was determined that competitive elution did not improve the diversity of recovered scFv over what is obtained from low-pH elution both in the natural and synthetic libraries. Almost all activation-specific scFv recovered have an RXD sequence (including RGD). However, the binding of each of these antibodies was weak, possibly due to suboptimal selection due to restricted diversity (only in H-CDR-3).

A secondary selection was thus performed using a sublibrary with the randomization in the sequence in L-CDR-3. The panning strategy was modified to improve the diversity and affinity of the recovered clones. The depletion step was eliminated, smaller numbers of cells were used for panning (to promote competition between members of the library for a smaller number of binding sites), and the elution protocol included recovering phage during addition of increasing concentrations of the small-molecule inhibitor eptifabitide. Thirty clones were randomly selected using this protocol. Almost all were activation specific with affinity comparable to the clones obtained using the natural library. Analysis of the clones did not reveal preferential enrichment of particular sequences in L-CDR-3, although the average length of the L-CDR-3 (~10 aa) was shorter than the original framework L-CDR-3 (13 aa).

Three clones were selected for further characterization. The $IC_{50}$ of binding of fibrinogen to ADP-activated platelets of clones from the primary synthetic library (with an RSD sequence) was > 50 μg/mL, while a clone from the VL diversified library (also having an RSD sequence) inhibited with an $IC_{50}$ of 1.2 μg/mL clearly demonstrating the utility of diversity in the L-CDR-3. By comparison, the clinically used mAb c7E3 (abciximab) has an $IC_{50}$ of 0.75 μg/mL but reduced conformational selectivity. The aggregation of platelets by ADP, thrombin, or collagen was inhibited to similar extents by 15 μg/mL of each mAb and c7E3. In addition, the binding of each of the ligand mimetic scFv did not induce exposure of the LIBS epitope 145, which is associated with thrombocytopenia. Finally, mapping using both human mouse chimeras and alanine point mutations demonstrated that the mAbs bound at an interface that overlaps with the fibrinogen-binding site.

A similar approach was undertaken to the identification of a fully human, ligand-mimentic $\alpha_{IIb}\beta_3$ mAbs using a naïve Fab library derived from human bone

marrow cDNA (from heavy-chain DP-47) and an synthetic H-CDR-3 with a central RAD sequence and all 20 amino acids permitted at the 6 flanking residues (3 per side) [65]. The heavy chain was initially paired with the light-chain DPK-26 (κ). The Fabs were obtained in 7 rounds of panning on purified human αIIbβ3. Additional selections were performed using secondary libraries assembled with light-chain shuffling using both Vκ and Vλ sequences obtained from human bone marrow DNA, shuffled onto the heavy chain obtained in the primary selection. Ten clones with the highest affinity [as determined by phage enzyme-linked immunosorbent assay (ELISA)] were sequenced. Nine of 10 formed a disulfide constrained loop and shared the central sequence motif V(V/W)CRAD(K/R) RC (sequence number 10 was TSHRADRRE). Four of the 6 clones contained the original light chain while the remainder of the light chains were derived from the naïve chain shuffled library.

Each of the 10 Fabs showed specific binding to $\alpha_{IIb}\beta_3$ with no cross-reactivity to $\alpha_V\beta_3$, $\alpha_5\beta_1$, and $\alpha_V\beta_8$, both as purified proteins and integrins expressed on cell surfaces. Using a competitive ELISA, each of the 10 Fabs was able to inhibit the binding of purified, immobilized αIIbβ3 to fibrinogen. The $IC_{50}$ of the tightest binding Fab was determined to be $8.0 \times 10^{-8}\,M$ by comparison to abciximab ($9.0 \times 10^{-8}$.) The $K_D$ of the Fab $\alpha_{IIb}\beta_3$ interaction, determined by competitive ELISA, was $3.3 \times 10^{-9}\,M$ while that of abciximab was $1.1 \times 10^{-9}\,M$. The Fab also inhibited platelet aggregation (stimulated with 20 μ*M* of ADP) at a comparable level to abciximab ($EC_{50}$ 60 n*M* or 30 μg/mL versus 45 n*M* or 22.5 μg/mL).

## 4 CONFORMATIONALLY SELECTIVE MAB AS THERAPEUTICS: PROGRESS AND PROSPECTS

The use of conformationally selective mAbs as therapeutics is based on two premises: (1) nascent binding epitopes become exposed (or occluded) upon integrin and activation and (2) pathological versus physiological epitope exposure can be kinetically distinguished by the sustained signaling associated with disease states, which results in constitutive epitope exposure, unlike the transient upregulation of integrin activity observed in normal physiology [12]. Although the first postulate has been conclusively demonstrated, additional studies must be performed to conclusively validate the second.

Blocking protein–ligand interactions and drug targeting/delivery to specific cell types (i.e., activated but noninactive lymphocytes) are the two principal clinical applications of mAbs. One outstanding question is which class of epitope is most appropriate to use for each potential application. In this section, we review each application of activation-sensitive mAbs and discuss both the progress made and critical issues to be resolved for future therapeutic applications.

### 4.1 Selective Inhibition of Activated Integrins

Disrupting protein–ligand interactions is perhaps the most important and straightforward therapeutic application of mAbs. To date, all therapeutic mAbs block protein–ligand interactions without conformational selectivity. Improved pharmacokinetics and lower iatrogenic effects may be obtained when only the disease-active subset of integrin–ligand interactions is disrupted.

Although there have been no clinical studies on the use of blocking conformational selective mAbs, in vitro and in vivo studies using mouse models have offered compelling support for their potential [66]. A major limitation of existing antithrombosis agents that function as αIIbβ3 inhibitors is the tendency to prolong bleeding via nonselective binding to activated and inactive platelets. MA2, an anti-$\alpha_{IIb}\beta_3$ identified by phage display, was shown to block the adhesion of activated, but not inactive, $\alpha_{IIb}\beta_3$ expressed on CHO cells and platelets to immobilized fibrinogen under flow (in a parallel-plate flow chamber apparatus). The platelet function analyzer (PFA-100) assay (an in vitro human bleeding time assay) was used to compare the effect of c7E3 (abciximab) and MA2 on closure of human whole blood in cartridges sealed with epinephrine/collagen- or ADP/collagen-coated membranes. Despite the ability to block binding to immobilized and soluble fibrinogen, MA2 (but not c7E3) did not prolong the time for closure, even at relatively high concentrations of mAb (10 μg/mL).

MA2 does not cross-react with murine $\alpha_{IIb}\beta_3$, precluding the ability to use in vivo models of thrombus formation to validate the conformational selective strategy. However, mAb SCE5, obtained in a parallel phage display screen, does cross-react with human and murine $\alpha_{IIb}\beta_3$ and was used for in vivo studies [66]. Thrombus formation was induced by ferric chloride and the occlusion time measured using a nanoflow probe. Since (nonconformationally selective) mAb c7E3 also does not cross-react with murine $\alpha_{IIb}\beta_3$, it was not possible to directly compare the use of nonselective and selective mAbs. Instead, direct comparison of thrombus remediation and bleeding time prolongation was made between SCE5 and two clinically used integrin blocking agents, eptifabtide and tirofban. Treatment with mice administered ferric chloride with SCE5 decreased the thrombus occlusion time comparable to that obtained with the two approved drugs. However, unlike the small-molecule antagonists, there was no extension of the bleeding time, as measured both by two standard measures, the tail

transfection method and incision methods. This study was the first to demonstrate the superiority of conformationally selective therapy in remediation of a specific disease due to the reduction of pernicious side effects.

### 4.2 Selective Drug Delivery Targeting Activated Integrins

Targeted drug delivery describes the use of a specific ligand (either authentic ligand or antibody) conjugated to a drug molecule to mediate specific drug delivery to cells bearing the receptor with the appropriate specificity for the ligand [67–69]. The conjugation can be direct (covalent or noncovalent) or indirect (with drug encapsulated in a ligand-coated lipid vesicle). Ideally, the ligand should promote both accumulation and uptake (internalization) of drug. Since drug delivery is selective to cells requiring remediation, targeted drug delivery vehicles have two (related) advantages over passive delivery: (1) improved pharmacokinetics, since drug concentrations can be limited due to the restricted number of cells that uptake drug and (2) lower probably of systemic toxicity. Targeted drug delivery systems can be used to delivery a wide variety of drugs including catatonic (chemotherapeutic agents), radio nucleotides, and nucleic acids for genetically mediated therapy.

Genetically mediated therapy (*gene therapy*) is defined as a broad repertoire of techniques used to modulate the effect of a specific gene (typically one involved in disease) via the application of exogenous nucleic acids. Despite the large number of potential methods and applications of gene therapy, the technique can be divided into two classes: gene knockdown and gene replacement. In the case of the immune system, however, many receptors are present on both inactive and activated cells and selective targeting remains challenging. As described below, scFv can be used to enhance the selectivity of gene therapy for activated leukocytes without perturbing the majority of quiescent, unaffected cells.

*RNA interference (RNAi)* is the sequence-specific posttranslational gene-silencing mechanism mediated by double-stranded RNA (dsRNA) with sequence complementarily to a target gene (Fig. 4A) [70, 71]. RNAi is conserved in a wide range of organisms and functions as a part of the innate host defense against viruses and the mobilization of transposable genetic elements. Long dsRNA produced by replicating virus, endogenous micro-RNAs, or aberrant transgenic transcripts are processed by an RNAseIII-like enzyme (Dicer) into smaller, 21–25 mer dsRNA, referred to as small interfering RNA (siRNA). The siRNAs are subsequently unwound and assembled onto the RNA-induced silencing complex (RISC) that uses the siRNA as a guide to target mRNAs and direct their subsequent degradation.

RNAi can be achieved either by expressing siRNA precursors such as short hairpin RNA (shRNA) with viral vectors or by directly incorporating synthetic siRNAs into the cytoplasm of cells. The use of synthetic siRNAs as small-molecule drugs for gene silencing avoids clinical safety concerns associated with viral vectors. RNA interference can be activated by introducing synthetic short double-stranded RNA fragments, termed small interfere siRNAs into cells to silence genes bearing complimentary sequences. RNA is a powerful tool used to evaluate the role of specific genes in cellular and disease processes that also holds great promise for therapeutic remediation of metabolic, degenerative disease, viral infections, and cancer, as well as inflammatory diseases. The discovery that introduction of exogenous 21–25 mer siRNAs into mammalian cells leads to the degradation of the homologous target mRNAs without provoking interferon response cells has opened up a new avenue to utilize synthetic siRNA as a drug.

While cell-specific delivery can be obtained using direct or indirect conjugation of siRNA to antibody or ligand, only a small fraction of leukocytes are activated during inflammation, and it is desirable to selectively target this subpopulation for anti-inflammatory therapy. Peer and co-workers used an *antibody protamine fusion* approach [13, 72] to deliver siRNA selectively to activated leukocytes. A highly charged protamine moiety, genetically fused with an scFv appropriate of desired selectivity, binds to synthetic, negatively charged siRNA. The antibody guides the RNA to its target and promotes its uptake, via receptor-mediated endocytosis (Fig. 4A). As the conversion of $\alpha_L\beta_2$ integrin to the active conformation is a highly sensitive marker of activated leukocytes, the active conformation-selective mAb AL-57 was utilized to preferentially deliver siRNAs to aberrantly activated leukocytes. Peer et al. [13] used the scFv derivative of mAb AL-57, fused at the N terminus to the protamine fragment (AL-57-PF), to silence candidate genes only in activated leukocytes that express the active conformation of integrin $\alpha_L\beta_2$ , present in a larger population of inactive cells (Fig. 4B). In a proof-of-concept experiment, Peer et al. [13] were able to silence the *Ku-70* gene selectively in activated, but not unstimulated, leukocytes (Fig. 4B). Furthermore, Peer et al. [13] used siRNA against the *cyclin-D1* gene to selectively suppress proliferation in activated lymphocytes. In contrast, TS1/22-PF, which binds to $\alpha_L\beta_2$ integrin in a conformation-insensitive manner, delivered siRNA equally well to resting and activated leukocytes. These experiments were the first to demonstrate that a conformationally selective mAb fragment could

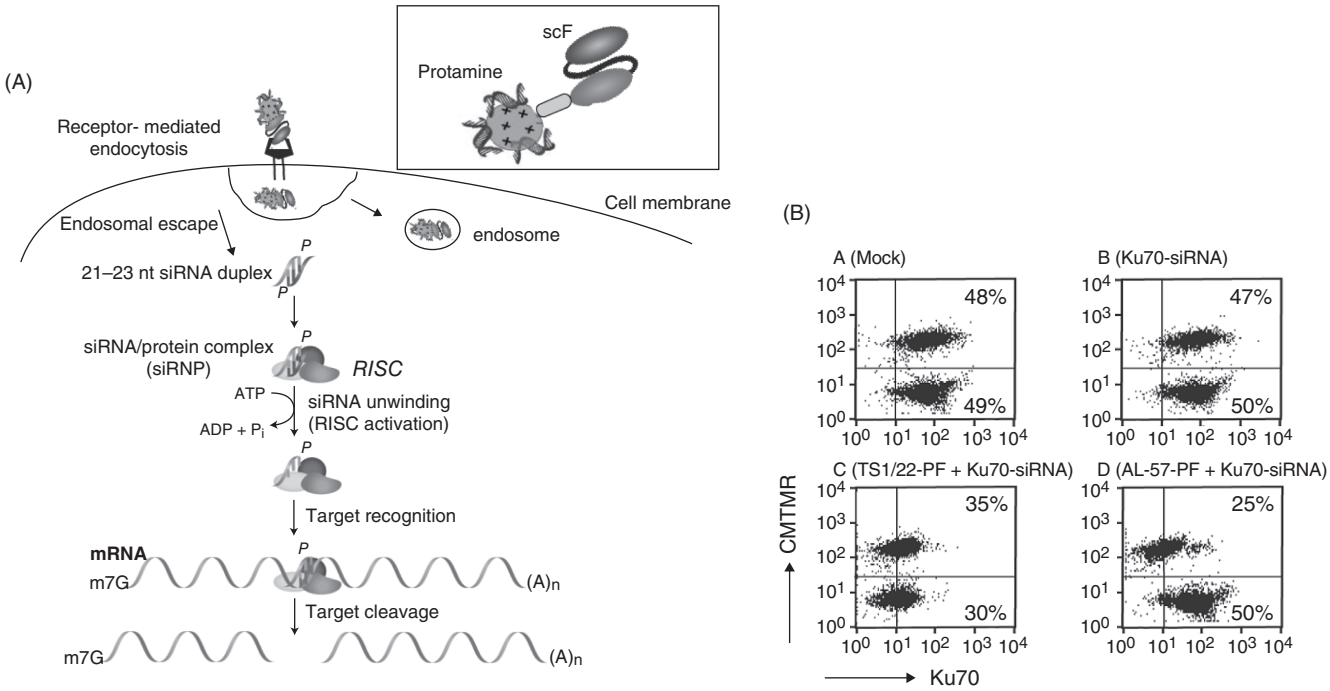


**Figure 4** Application of the active conformation-selective integrin antibody to siRNA delivery specific to activated leukocytes. (A) Receptor-mediated intracellular delivery of siRNA by scFv-protamine fusion protein (scFv-PF) and subsequent induction of gene silencing. ScFv-PF:siRNA complexes are directed via scFv to the target on the cell surface. Bound scFv-PF:siRNA complexes are internalized via receptor-mediated endocytosis, releasing siRNAs to the cytoplasm (potentially though an escape from the endosome). In the cytoplasm, siRNAs incorporated into RNA-induced silencing complex (RISC) leads to the cleavage of target messenger RNAs, inducing gene silencing. Inset: scFv-PF: protamine, a positively charged protein that binds to negatively charged siRNAs, is linked at the C termini of the scFv. (B) The active conformation-selective LFA-1 mAb AL-57 can be used to deliver siRNAs to induce gene silencing only in activated cells in mixed population. CMTMR (fluorescence dye)-labeled activated cells, expressing high-affinity LFA-1, were co-cultured with nonactivated (unlabeled) cells expressing low-affinity LFA-1. Three days after delivery of 1 nmol siRNA to Ku70, a reference protein, the co-cultures were analyzed. Whereas TS1/22-protamine fusion protein-delivered siRNAs silence Ku70 in both activated and nonactivated cells (panel C), AL-57-protamine fusion protein delivered siRNAs silence only the labeled activated cells (panel D). (This figure is modified from [13].)

be use to obtain modulation of gene expression and cell function in a therapeutically relevant context.

AL-57 does not cross-react with murine $\alpha_L\beta_2$. To demonstrate the ability of AL-57-PF to selectively silence gene expression in vivo, severe combined immunodeficiency (SCID) mice were engrafted with human erythroleukemia K562 cells stably transfected with wild- type (WT) and high-affinity (HA) $\alpha_L\beta_2$. After intravenous (I.V.) injection, the K562 cells formed small nodules in the mouse lung. Five days after injection, Cy3-conjugated siRNA complexed with AL-57-PF or TS1/22-PF was injected into the tail vein. At 5 h postinjection, the lung tissue was recovered and analyzed by FACS. TS1/22-PF delivered siRNA equally well to both WT and HA $\alpha_L\beta_2$-expressing K562 cells, while AL-57-PF delivered siRNA only to cells expressing the HA $\alpha_L\beta_2$-expressing K562 cells. Additional experiments will be required to demonstrate selective binding in a more authentic context (i.e., selective delivery to activated murine lymphocytes in an inflammation model). However, as there are no murine activation-specific $\beta_2$ mAbs, this must await development of the appropriate reagents.

Despite the promise of conformationally selective mAbs as therapeutics, a number of critical issues remain unresolved. Possibly the greatest bottleneck is the relative paucity of conformational selective antibodies for laboratory animal species, specifically rodents but also

pigs, dogs, and sheep. There are numerous, well-characterized animal models of human disease, including autoimmune diseases, cancer, susceptibility to bacterial infections, and transplantation. These would be ideal systems to investigate both basic integrin biology and the potential for conformational selective therapeutic remediation. Although great progress has been made in understanding the mechanism of integrin activation, the in vivo environment is much more complex with multiple competing (and overlapping) signaling pathways and ligands. Little is known about the mechanism by which integrin becomes activated in vivo, what is the spatiotemporal population of activated integrin in normal physiology, and how it differs in disease states. For ethical and practical reasons, these questions can clearly not be addressed using antihuman antibodies. Fortunately, the sequence for integrins from many laboratory animals have been published, and it should be possible to express cell surface and purified nonhuman integrins as the bait for antibody library selection. With these reagents in hand, it should be possible to address some of the most critical questions in the development of conformational selective mAb therapy.

## 5 APPENDIX

### 5.1 Overview of Monoclonal Antibodies

Antibodies are four-chain glycoproteins organized as a disulfide bonded homodimer of a disulfide-bonded heterodimer of one heavy (H) and one light (L) chain [stoichiometry $(HL)_2$] [73]. Both H and L chains are built from immunoglobulin fold (Ig) domains. There are five classes of H that differ in body tissue distribution, effector function, the position of disulfide bonds, interdomain flexibility, and number of Ig domains: IgA, IgD, IgE, IgG, and IgM. There are two classes of L, κ, and λ. Each antibody pairs one specific H chain with one L chain; there are no (naturally) occurring mixed H or L heterodimers. The IgG class has the highest concentration is serum (~14 mgs/mL), the highest circulating half-life (~21 days), and is the most clinically relevant.

There is (relatively) little sequence variation in the IgC domains within a given family. By comparison, there is extensive sequence variability in the IgV domains. Most of this variability (in both VL and VH) is concentrated in three loops [complementary determining regions (CDR) 1–3]. The specificity and affinity of antibodies for antigen is largely determined by the sequence and structure of the CDR loops. Interspersed between the CDR loops are the framework regions (FR 1–4) that maintain the structural orientation of the adjacent CDR regions. In the three-dimensional structure of the antibody, the six CDR loops are spatially proximate and form the binding interface with the antigen. The *paritope* is the set of amino acids on the antibody that contact the antigen (at the epitope). The hypervariable H-CDR-3 is not only the most variable in sequence but also in length, ranging from 6–14 aa, and generally contributes the most energetically significant antigen interactions to the interface.

The four N-terminal Ig domains preceding the hinge (VL-CL, VH-CH1) form the univalent *Fragment antigen binding (Fab)*. Fab fragments can be prepared by digestion of intact IgG with the cysteine protease papain which cleaves on the amino side of the disulfide bond; the second product is the *fragment crystallizable (Fc)* containing the CH2–CH3 domains, which interact with proteins in the activation of innate immunity.

The *Fv (fragment variable)*, the noncovalent dimer of VL and VH, is the minimal unit of antigen-binding specificity of (most) antibodies. Despite an extensive interface between the two domains, the heterodimer is relatively unstable and subject to dissociation on prolonged incubation. This technical restraint is overcome by the single-chain fusion of heavy- and light-chain variable domains (scFv) joined by a linker of 12–16 amino acid residues to span the 35–40 Å junction between the N and C termini of the two variable domains [74]. Linkers of various design have been used, although the most common are the trivial $(G_3S)_3$ and $(G_3S)_4$. The scFv can be assembled in two orientations VH–linker–VL and VL–linker–VH; generally both orientations are functional, although for some antibodies one orientation retains higher binding activity than the other. The small size of the scFv (ca. 20 kDa) is ideal for a conformation-specific binding reagent since steric interactions outside the binding epitope are minimized. Furthermore, the scFv is monomeric (circumventing avidity and cross-linking, which make quantitative analysis of binding problematic) and lacks the biological activity of the Fc region. However, for some applications, the avidity effects of multimerization may be desirable. Multimers of scFv can be obtained by altering the length of the linker between VH and VL or engineering multiple VH and VL sequences on a single DNA transcript [75, 76].

Unlike the Fab fragment, there is no ready preparative technique for obtaining scFv from intact antibodies, and it is necessary to either engineer the requisite gene de novo or screen a microbial library of scFv using bait antigen (vide infra). A novel cDNA encoding an scFv can be generated in four steps if the B cell (or B-cell derivative) that expresses the antibody is available:

1. RT-PCR of the antibody-expressing cell using degenerate primer sets that cover the appropriate

species heavy- and light-chain antibody repertoire.
2. Cloning and sequencing of the VL and VH domains.
3. Overlapping extension PCR to create a synthetic DNA that combines the VL and VH via an appropriate linker.
4. Cloning into an appropriate expression vector that includes a signal sequence and affinity tag for purification.

The relatively small size and limited number of disulfide bonds in the scFv permits the use of *Escherichia coli* periplasmic expression vectors to obtain high-level expression of correctly folded scFv. A $(His)_6$ tag is typically appended to the C terminus of the scFv to permit facile affinity purification using commercially available Ni(II) resins.

***Origin of Natural Antibody Diversity and Affinity*** The human immune system is capable of expressing at least $10^{11}$ antibodies with distinct specificity. Five mechanisms account for the remarkable size of the immune repertoire [1]. First, antibody heavy and light chains are combinatorially assembled from (distinct) multiple gene segments. The variable light chain is obtained by combining 1 of 40 Vκ or 1 of 30 Vλ segments (of 95–101 aa) with 1 of 5 Jκ or 1 of 4 Jλ segments (joining segments of up to 13 aa link the variable region to the constant domain). The heavy chain is built from 1 of 65 $V_H$ segments linked to 1 of 27 diversity (D) segments that are linked to 1 of 6 $J_H$. Second, in 5% of antibodies of a D–D fusion gives rise to a long H-CDR-3. A third source of diversity is the pairing of any one assembled light chain with any heavy chain. The enzymes that assemble the individual chains create a fourth source of diversity (in H-CDR-3) by adding or subtracting nucleotides during the assembly process. Finally, *somatic hypermutation* is the process whereby a large number of point mutations occur in the variable regions of both heavy and light chains of a single antibody, giving rise to many variants with differing affinity for antigen. B cells expressing the highest affinity antibodies are most likely to survive and become antibody-secreting cells. The combined process of somatic hypermutation and selection is known as *affinity maturation*.

Most laboratory antibody reagents and therapeutic agents have been obtained using *immunization–hybridoma technology* [77]. Mice produce a robust immune response following immunization with many protein antigens. A major difficulty in obtaining monoclonal antibodies of defined specificity following immunization is the inability to culture B cells. This limitation is overcome fusing spleen cells from an immunized animal with immortalized, non-antibody-producing myleoma cells to produce a *hybridoma*. The myeloma cell line is deficient in the enzyme hypoxanthinephosphoribosyltransferase (HPRT) and will not grow in media containing aminopterin. Spleen cells can use hypoxanthine and thymidine for RNA precursor synthesis. Fusion of the spleen cells with myeloma confers upon the hybridoma the ability to grow in aminopterin, allowing hybridoma cells to be selected from a mixture with unfused myeloma cells. The antibody that is secreted from monoclonal B cells is then screened for binding to the designed antigen. Care must be taken to ensure that the antigen is injected in the native state if conformation-sensitive antibodies are desired; isolated membrane fractions containing the protein are ideal for this purpose. Xenogenic antibodies that are intended for therapeutic application must be partially or fully humanized to avoid a human–antixeno (generally murine) response [78]. Humanization involves the grafting of all or part of the variable domains of an antibody onto a purely human framework or constant regions, followed by additional modifications that may be required to restore the affinity or specificity of the mAb.

### 5.2 Antibody Display Libraries

*Microbial surface display* of antibody libraries is an alternative to conventional immunization–hybridoma protocols for the isolation of new immunoaffinity reagents (Mondon et al., 2008). Antibodies with the desired selectivity are isolated by "panning" the library with desired antigen, retaining those that express favorable binding properties and rejecting non-binding and low-affinity clones. As the phenotype and genotype are "linked" by simultaneous presence on the particle, the DNA encoding the desired antibody can be recovered and used to obtain soluble protein for the desired application.

*Phage Display* is the most commonly used technique for the selection of affinity reagents from preexisting peptide and protein-fragment libraries [79, 80]. Filamentous bacteriophage (such as fd, f1, and Pf1) are single-stranded DNA viruses composed of ~2700 copies of a major coat protein (pVIII) encapsulating the single-stranded DNA (ssDNA) genome in a tube 1 μm by 6–7 nm. Approximately five copies of minor coat proteins pIII and pIX are located at opposite ends of the protein capsid.

Proteins (i.e., scFv and Fab) can be expressed in native form as fusions to the pIII protein where they are displayed on the surface of the phage. Libraries of scFv or Fab with sequence diversity in the CDR loops can be constructed by cloning the appropriate DNA into a

*phagemid vector*, which includes the packaging sequence for incorporation into the phage particle. An scFv or Fab library can be built using three sources of immune DNA [81]. The most unbiased source of DNA is obtained from the spleen, lymph nodes, or lymphocytes of laboratory animals or humans. When selection of antibodies against a particular antigen is desired, the animal can be pre-immunized to appropriately bias the immune repertoire. Finally, purely synthetic CDR sequences of defined length can be grafted onto standardized framework backbones. In a synthetic library, diversity is limited to selected positions that are hypervariable and/or participate at high frequency in the paratope of characterized antibody–antigen interactions. This approach maximizes the possibility of identifying an antibody that interacts with antigen in a relatively small library at the cost of not obtaining the highest potential affinity binders available in a more complex library. Transformation of the pool of vector DNA into the appropriate bacterial host (typically *E. coli*) is the final step in assembly of the library.

To obtain the desired affinity reagent, the library must be expanded and panned [82]. ScFv- (or Fab-) bearing phage are produced when the library cells are infected with a *helper phage* that carries the DNA for proteins required for phage synthesis and assembly. The desired antigen is immobilized, either randomly on plastic or in a given orientation (using affinity tags such as biotin - streptavidin ) and exposed to the phage library solution. After brief washing, only antigen-binding phage remain bound and are collected. The recovered pool of phage is used to reinfect bacterial cells. Each round of panning enriches the population of phage with specificity for a given antigen. The use of moderately biotinylated bait avoids protein denaturation that occurs upon binding plastic and is preferable when conformationally selective mAbs are desired.

An alternative source of bait is proteins in their native environment (i.e., in intact cells or cell membrane preparations) where the protein most likely maintains the native conformation and retains functionality. However, cell membranes contain many antigens, and nonspecific/off-target binding of phage is likely to occur, lowering the stringency of the screening protocol. Panning against the protein antigen expressed on two distinct cell line backgrounds (i.e., endogenous protein in T cells and CHO cell transfectants) increases the specificity of the selection protocol.

***Affinity Maturation*** Generally, it is desirable to obtain antibodies with the highest affinity to their cognate antigens. The affinity of library-selected antibodies is limited by the relatively small (by comparison to the immune system) diversity in the sequence and structure of the CDR. However, even mAbs selected by immunization have a wide range of affinities that may be insufficient for therapeutic application. The affinity can be further improved using in vitro affinity maturation and selection [83]. To affinity mature an mAb, it is necessary to diversify the sequence and structure of one or more CDR loops. The most straightforward approach is to use error-prone PCR to introduce one or more mutations into the cDNA encoding the antibody fragment (Fab or scFv). However, this process is relatively nonselective and inefficient (mutations are equally dispersed through the sequence instead of being targeted to key antigen-binding positions) and the bias of the error-prone polymerase does not necessarily match the optimal use of amino acids at key positions in the CDR sequence. A technically more challenging—but generally useful approach—is use overlapping extension PCR to reconstruct one or more CDR sequences using degenerate oligonucleotides with random sequences at positions found to vary in naturally occurring high-affinity antibodies [84]. If the paratope or crystal structure of the mAb–antigen complex is known, this approach can also be used to bias the randomized positions toward those that make favorable contacts with the antigen. Finally, the affinity can be improved by chain shuffling, pairing a library of light-chain domains with a heavy chain from the original selection [85].

***Antibody Humanization*** mAbs obtained from murine (or other species) hybridomas are immunogenic in humans. Even libraries assembled from purely human sequences are potential immunogens due to the novel sequences in the CDR loops that appear foreign to the immune system. For therapeutic applications, it is necessary to reduce or eliminate the xenogenic features of the mAb to avoid rapid clearance and a systemic inflammatory response [78]. A "foreign" mAb can be humanized by either (1) grafting the CDR loops from the murine mAb onto a human framework or (2) replacing surface residues (those most likely to provoke an immune response) with those found commonly in human mAbs . However, the binding affinity and specificity of the engineered mAb are not necessarily maintained. It is generally necessary to include one or more xenogenic framework mutations that contribute to the structure of the CDR loops into the humanized mAb. The affinity and specificity of the novel mAb may require reoptimization using phage display and affinity maturation.

## REFERENCES

1. Janeway CA, Travers P, Walport MJ, Shlomchik MJ (2005). *Immunobiology: The Immune System in Health and Disease*. New York: Garland Science Publishing.

2. Jain M, Kamal N, Batra SK (2007). Engineering antibodies for clinical applications. *Trends Biotechno 25*, 307–316.
3. ter Meulen J (2007). Monoclonal antibodies for prophylaxis and therapy of infectious diseases. *Expert Opin Emerg Drugs 12*, 525–540.
4. Bonnefoy-Berard N, Revillard JP (1996). Mechanisms of immunosuppression induced by antithymocyte globulins and OKT3. *J Heart Lung Transplant 15*, 435–442.
5. Butcher EC, Picker LJ (1996). Lymphocyte homing and homeostasis. *Science 272*, 60–66.
6. Ridley AJ, Schwartz MA, Burridge K, Firtel RA, Ginsberg MH, Borisy G, Parsons JT, Horwitz AR (2003). Cell migration: Integrating signals from front to back. *Science 302*, 1704–1709.
7. Springer TA (1994). Traffic signals for lymphocyte recirculation and leukocyte emigration: The multi-step paradigm. *Cell 76*, 301–314.
8. Hammerschmidt S (2006). Adherence molecules of pathogenic pneumococci. *Curr Opin Microbiol 9*, 12–20.
9. Carman CV, Springer TA (2003). Integrin avidity regulation: Are changes in affinity and conformation underemphasized? *Curr Opin Cell Biol 15*, 547–556.
10. Shimaoka M, Takagi J, Springer TA (2002). Conformational regulation of integrin structure and function. *Annu Rev Biophys Biomol Struct 31*, 485–516.
11. Humphries MJ (2004). Monoclonal antibodies as probes of integrin priming and activation. *Biochem Soc Trans 32*, 407–411.
12. Tanaka Y, Mine S, Figdor CG, Wake A, Hirano H, Tsukada J, Aso M, Fujii K, Saito K, van Kooyk Y, Eto S (1998). Constitutive chemokine production results in activation of leukocyte function-associated antigen-1 on adult T-cell leukemia cells. *Blood 91*, 3909–3919.
13. Peer D, Zhu P, Carman CV, Lieberman J, Shimaoka M (2007). Selective gene silencing in activated leukocytes by targeting siRNAs to the integrin lymphocyte function-associated antigen-1. *Proc Natl Acad Sci USA 104*, 4095–4100.
14. Hynes RO (1987). Integrins: A family of cell surface receptors. *Cell 48*, 549–554.
15. Luo BH, Carman CV, Springer TA (2007). Structural basis of integrin regulation and signaling. *Annu Rev Immunol 25*, 619–647.
16. Xiong J-P, Stehle T, Diefenbach B, Zhang R, Dunker R, Scott DL, Joachimiak A, Goodman SL, Arnaout MA (2001). Crystal structure of the extracellular segment of integrin αVβ3. *Science 294*, 339–345.
17. Leitinger B, Hogg N (2000). Effects of I domain deletion on the function of the β2 integrin lymphocyte function-associated antigen-1. *Mol Biol Cell 11*, 677–690.
18. Yalamanchili P, Lu C, Oxvig C, Springer TA (2000). Folding and function of I-domain deleted Mac-1 and LFA-1. *J Biol Chem 275*, 21877–21882.
19. Qu A, Leahy DJ (1995). Crystal structure of the I-domain from the CD11a/CD18 (LFA-1, $\alpha_L\beta2$) integrin. *Proc Natl Acad Sci USA 92*, 10277–10281.
20. Shimaoka M, Lu C, Palframan R, von Andrian UH, Takagi J, Springer TA (2001). Reversibly locking a protein fold in an active conformation with a disulfide bond: Integrin αL I domains with high affinity and antagonist activity in vivo. *Proc Natl Acad Sci USA 98*, 6009–6014.
21. Shimaoka M, Xiao T, Liu J-H, Yang Y, Dong Y, Jun C-D, McCormack A, Zhang R, Joachimiak A, Takagi J, Wang J-H, Springer, TA (2003). Structures of the αL I domain and its complex with ICAM-1 reveal a shape-shifting pathway for integrin regulation. *Cell 112*, 99–111.
22. Xiong JP, Stehle T, Zhang R, Joachimiak A, Frech M, Goodman SL, Arnaout MA (2002). Crystal structure of the extracellular segment of integrin αVβ3 in complex with an Arg-Gly-Asp ligand. *Science 296*, 151–155.
23. Xiao T, Takagi J, Wang J-H, Coller BS, Springer TA (2004). Structural basis for allostery in integrins and binding of ligand-mimetic therapeutics to the platelet receptor for fibrinogen. *Nature 432*, 59–67.
24. Wouters MA, Rigoutsos I, Chu CK, Feng LL, Sparrow DB, Dunwoodie SL (2005). Evolution of distinct EGF domains with specific functions. *Protein Sci 14*, 1091–1103.
25. Beglova N, Blacklow SC, Takagi J, Springer TA (2002). Cysteine-rich module structure reveals a fulcrum for integrin rearrangement upon activation. *Nat Struct Biol 9*, 282–287.
26. Li R, Babu CR, Valentine K, Lear JD, Wand AJ, Bennett JS, DeGrado WF (2002a). Characterization of the monomeric form of the transmembrane and cytoplasmic domains of the integrin β3 subunit by NMR spectroscopy. *Biochemistry 41*, 15618–15624.
27. Luo B-H, Springer TA, Takagi J (2004). A specific interface between integrin transmembrane helices and affinity for ligand. *PLoS Biol 2*, 776–786.
28. Vinogradova O, Velyvis A, Velyviene A, Hu B, Haas TA, Plow EF, Qin J (2002). A structural mechanism of integrin $\alpha_{IIb}\beta_3$"inside-out" activation as regulated by its cytoplasmic face. *Cell 110*, 587–597.
29. Hughes PE, Diaz-Gonzalez F, Leong L, Wu C, McDonald JA, Shattil SJ, Ginsberg MH (1996). Breaking the integrin hinge. *J Biol Chem 271*, 6571–6574.
30. Fagerholm SC, Hilden TJ, Gahmberg CG (2004). P marks the spot: Site-specific integrin phosphorylation regulates molecular interactions. *Trends Biochem Sci 29*, 504–512.
31. Nishida N, Xie C, Shimaoka M, Cheng Y, Walz T, Springer TA (2006). Three distinctive ectodomain conformations and their interconversion in integrins that contain I domains. *Immunity 25*, 583–594.
32. Andrew D, Shock A, Ball E, Ortlepp S, Bell J, Robinson M (1993). KIM185, a monoclonal antibody to CD18 which induces a change in the conformation of CD18 and promotes both LFA-1- and CR3-dependent adhesion. *Eur J Immunol 23*, 2217–2222.
33. Stephens P, Romer JT, Spitali M, Shock A, Ortlepp S, Figdor C, Robinson MK (1995). KIM127, an antibody that promotes adhesion, maps to a region of CD18 that includes cysteine-rich repeats. *Cell Adhes Commun 3*, 375–384.

34. Lu C, Ferzly M, Takagi J, Springer TA (2001). Epitope mapping of antibodies to the C-terminal region of the integrin β2 subunit reveals regions that become exposed upon receptor activation. *J Immunol 166*, 5629–5637.
35. Xie C, Shimaoka M, Xiao T, Schwab P, Klickstein LB, Springer TA (2004). The integrin α subunit leg extends at a $Ca^{2+}$-dependent epitope in the thigh/genu interface upon activation. *Proc Natl Acad Sci USA 101*, 15422–15427.
36. Keizer GD, Visser W, Vliem M, Figdor CG (1988). A monoclonal antibody (NKI-L16) directed against a unique epitope on the alpha-chain of human leukocyte function-associated antigen 1 induces homotypic cell-cell interactions. *J Immunol 140*, 1393–1400.
37. Clark K, Pankov R, Travis MA, Askari JA, Mould AP, Craig SE, Newham P, Yamada KM, Humphries MJ (2005). A specific alpha5beta1-integrin conformation promotes directional integrin translocation and fibronectin matrix formation. *J Cell Sci 118*, 291–300.
38. Luo B-H, Strokovich K, Walz T, Springer TA, Takagi J (2004). Allosteric β1 integrin antibodies that stabilize the low affinity state by preventing the swing-out of the hybrid domain. *J Biol Chem 279*, 27466–27471.
39. Huang C, Zang Q, Takagi J, Springer TA (2000). Structural and functional studies with antibodies to the integrin β2 subunit: A model for the I-like domain. *J Biol Chem 275*, 21514–21524.
40. Mould AP, Barton SJ, Askari JA, McEwan PA, Buckley PA, Craig SE, Humphries MJ (2003). Conformational changes in the integrin βA domain provide a mechanism for signal transduction via hybrid domain movement. *J Biol Chem 278*, 17028–17035.
41. Tang RH, Tng E, Law SK, Tan SM (2005). Epitope mapping of monoclonal antibody to integrin αLβ2 hybrid domain suggests different requirements of affinity states for intercellular adhesion molecules (ICAM)-1 and ICAM-3 binding. *J Biol Chem 280*, 29208–29216.
42. Ma Q, Shimaoka M, Lu C, Jing H, Carman CV, Springer TA (2002). Activation induced conformational changes in the I domain region of LFA-1. *J Biol Chem 277*, 10638–10641.
43. Li R, Haruta I, Rieu P, Sugimori T, Xiong JP, Arnaout MA (2002). Characterization of a conformationally sensitive murine monoclonal antibody directed to the metal ion-dependent adhesion site face of integrin CD11b. *J Immunol 168*, 1219–1225.
44. Zhang L, Plow EF (1999). Amino acid sequences within the alpha subunit of integrin alpha M beta 2 (Mac-1) critical for specific recognition of C3bi. *Biochemistry 38*, 8064–8071.
45. Diamond MS, Springer TA (1993). A subpopulation of Mac-1 (CD11b/CD18) molecules mediates neutrophil adhesion to ICAM-1 and fibrinogen. *J Cell Biol 120*, 545–556.
46. Oxvig C, Lu C, Springer TA (1999). Conformational changes in tertiary structure near the ligand binding site of an integrin I domain. *Proc Natl Acad Sci USA 96*, 2215–2220.
47. Eisenhardt SU, Schwarz M, Schallner N, Soosairajah J, Bassler N, Huang D, Bode C, Peter K (2007). Generation of activation-specific human anti-alphaMbeta2 single-chain antibodies as potential diagnostic tools and therapeutic agents. *Blood 109*, 3521–3528.
48. Abrams C, Deng YJ, Steiner B, O'Toole T, Shattil SJ (1994). Determinants of specificity of a baculovirus-expressed antibody Fab fragment that binds selectively to the activated form of integrin alpha IIb beta 3. *J Biol Chem 269*, 18781–18788.
49. Pampori N, Hato T, Stupack DG, Aidoudi S, Cheresh DA, Nemerow GR, Shattil SJ (1999). Mechanisms and consequences of affinity modulation of integrin αVβ3 detected with a novel patch-engineered monovalent ligand. *J Biol Chem 274*, 21609–21616.
50. Hoet RM, Cohen EH, Kent RB, Rookey K, Schoonbroodt S, Hogan S, Rem L, Frans N, Daukandt M, Pieters H, van Hegelsom R, Neer NC, Nastri HG, Rondon IJ, Leeds JA, Hufton SE, Huang L, Kashin, I, Devlin M, Kuang G, Steukers M, Viswanathan M, Nixon AE, Sexton DJ, Hoogenboom HR, Ladner RC (2005). Generation of high-affinity human antibodies by combining donor-derived and synthetic complementarity-determining-region diversity. *Nat Biotechnol 23*, 344–348.
51. Huang L, Shimaoka M, Rondon I, Roy I, Chang Q, Po M, Dransfield DT, Ladner RC, Edge AS, Salas A, Wood CR, Springer TA, Cohen EH (2006). Identification and characterization of a human monoclonal antagonistic antibody AL-57 specific for the high affinity form of lymphocyte function-associated antigen-1. *J Leuko Biol 80*, 905–914.
52. Shimaoka M, Kim M, Cohen E, Yang W, Astrof N, Peer D, Salas A, Ferrand A, Springer T (2006). AL-57, a novel ligand-mimetic antibody to integrin LFA-1, reveals chemokine-induced affinity upregulation in lymphocytes. *Proc Natl Acad Sci USA 103*, 13991–13996.
53. Mould AP, Garratt AN, Puzon-McLaughlin W, Takada Y, Humphries MJ (1998). Regulation of integrin function: Evidence that bivalent-cation-induced conformational changes lead to the unmasking of ligand-binding sites within integrin α5β1. *Biochem J 331*, 821–828.
54. Mould AP, Askari JA, Barton S, Kline AD, McEwan PA, Craig SE, Humphries MJ (2002). Integrin activation involves a conformational change in the α1 helix of the β subunit A-domain. *J Biol Chem 277*, 19800–19805.
55. Lu C, Shimaoka M, Zang Q, Takagi J, Springer TA (2001c). Locking in alternate conformations of the integrin αLβ2 I domain with disulfide bonds reveals functional relationships among integrin domains. *Proc Natl Acad Sci USA 98*, 2393–2398.
56. Kamata T, Tieu KK, Tarui T, Puzon-McLaughlin W, Hogg N, Takada Y (2002). The role of the CPNKEKEC sequence in the beta(2) subunit I domain in regulation of integrin alpha(L)beta(2) (LFA-1). *J Immunol 168*, 2296–2301.
57. Shattil SJ, Hoxie JA, Cunningham M, Brass LF (1985). Changes in the platelet membrane glycoprotein IIb-IIIa complex during platelet activation. *J Biol Chem 260*, 11107–11114.

58. Kunicki TJ, Annis DS, Felding-Habermann B (1997). Molecular determinants of arg-gly-asp ligand specificity for beta3 integrins. *J Biol Chem 272*, 4103–4107.
59. Tomiyama Y, Brojer E, Ruggeri ZM, Shattil SJ, Smiltneck J, Gorski J, Kumar A, Kieber-Emmons T, Kunicki TJ (1992). A molecular model of RGD ligands. Antibody D gene segments that direct specificity for the integrin αIIbβ 3. *J Biol Chem 267*, 18085–18092.
60. Kodandapani R, Veerapandian B, Kunicki TJ, Ely KR (1995). Crystal structure of the OPG2 Fab. An antireceptor antibody that mimics an RGD cell adhesion site. *J Biol Chem 270*, 2268–2273.
61. Kunicki TJ, Annis DS, Deng YJ, Loftus JC, Shattil SJ (1996). A molecular basis for affinity modulation of Fab ligand binding to integrin alphaIIb beta3. *J Biol Chem 271*, 20315–20321.
62. Beer JH, Springer KT, Coller BS (1992). Immobilized Arg-Gly-Asp (RGD) peptides of varying lengths as structural probes of the platelet glycoprotein IIb/IIIa receptor. *Blood 79*, 117–128.
63. Jacobin MJ, Laroche-Traineau J, Little M, Keller A, Peter K, Welschof M, Nurden A, Clofent-Sanchez G (2002). Human IgG monoclonal anti-alpha(IIb)beta(3)-binding fragments derived from immunized donors using phage display. *J Immunol. 168*, 2035–2045.
64. Schwarz M, Rottgen P, Takada Y, Le Gall F, Knackmuss S, Bassler N, Buttner C, Little M, Bode C, Peter K (2004). Single-chain antibodies for the conformation-specific blockade of activated platelet integrin alphaIIbbeta3 designed by subtractive selection from naive human phage libraries. *FASEB J 18*, 1704–1706.
65. Chung J, Rader C, Popkov M, Hur YM, Kim HK, Lee YJ, Barbas CF 3rd (2004). Integrin alphaIIbbeta3-specific synthetic human monoclonal antibodies and HCDR3 peptides that potently inhibit platelet aggregation. *FASEB J 18*, 361–363.
66. Schwarz M, Meade G, Stoll P, Ylanne J, Bassler N, Chen YC, Hagemeyer CE, Ahrens I, Moran N, Kenny D, Fitzgerald D, Bode C, Peter K (2006). Conformation-specific blockade of the integrin GPIIb/IIIa: A novel anti-platelet strategy that selectively targets activated platelets. *Circ Res 99*, 25–33.
67. Metselaar JM, Storm G (2005). Liposomes in the treatment of inflammatory disorders. *Expert Opin Drug Deliv 2*, 465–476.
68. Peer D, Karp JM, Hong SY, Farokhzad O, Margalit R, Langer R (2007). Nanocarriers: Emerging platforms for cancer therapy. *Nature Nanotech 2*(12), 751–760.
69. Torchilin VP (2005). Recent advances with liposomes as pharmaceutical carriers. *Nat Rev Drug Discov 4*, 145–160.
70. Behlke MA (2006). Progress towards in vivo use of siRNAs. *Mol Ther 13*, 644–670.
71. Dykxhoorn DM, Lieberman J (2006). RUNNING INTERFERENCE: Prospects and obstacles to using small interfering RNAs as small molecule drugs. *Annu Rev Biomed Eng 8*, 377–402.
72. Song E, Zhu P, Lee SK, Chowdhury D, Kussman S, Dykxhoorn DM, Feng Y, Palliser D, Weiner DB, Shankar P, Marasco WA, Lieberman J (2005). Antibody mediated in vivo delivery of small interfering RNAs via cell-surface receptors. *Nat Biotechnol 23*, 709–717.
73. Branden C, Tooze J (1991). *Introduction to Protein Structure*. New York and London: Garland Publishing.
74. Bird RE, Hardman KD, Jacobson JW, Johnson S, Kaufman BM, Lee SM, Lee T, Pope SH, Riordan GS, Whitlow M (1988). Single-chain antigen-binding proteins. *Science 242*, 423–426.
75. Dolezal O, Pearce LA, Lawrence LJ, McCoy AJ, Hudson PJ, Kortt AA (2000). ScFv multimers of the anti-neuraminidase antibody NC10: Shortening of the linker in single-chain Fv fragment assembled in V(L) to V(H) orientation drives the formation of dimers, trimers, tetramers and higher molecular mass multimers. *Protein Eng 13*, 565–574.
76. Hudson PJ, Kortt AA (1999). High avidity scFv multimers; diabodies and triabodies. *J Immunol Methods 231*, 177–189.
77. Köhler G, Milstein C (1975). Continuous cultures of fused cells secreting antibody of predefined specificity. *Nature 256*, 495–497.
78. Almagro JC, Fransson J (2008). Humanization of antibodies. *Front Biosci 13*, 1619–1633.
79. Kretzschmar T, von Ruden T (2002). Antibody discovery: Phage display. *Curr Opin Biotechnol 13*, 598–602.
80. Clackson T, Hoogenboom HR, Griffiths AD, Winter G (1991). Making antibody fragments using phage display libraries. *Nature 352*, 624–628.
81. Pini A, Giuliani A, Ricci C, Runci Y, Bracci L (2004). Strategies for the construction and use of peptide and antibody libraries displayed on phages. *Curr Protein Pept Sci 5*, 487–496.
82. Coomber DW (2002). Panning of antibody phage-display libraries. Standard protocols. *Methods Mol Biol 178*, 133–145.
83. Dufner P, Jermutus L, Minter RR (2006). Harnessing phage and ribosome display for antibody optimisation. *Trends Biotechnol 24*, 523–529.
84. Rajpal A, Beyaz N, Haber L, Cappuccilli G, Yee H, Bhatt RR, Takeuchi T, Lerner RA, Crea R (2005). A general method for greatly improving the affinity of antibodies by using combinatorial libraries. *Proc Natl Acad Sci USA 102*, 8466–8471.
85. Marks JD (2004). Antibody affinity maturation by chain shuffling. *Methods Mol Biol 248*, 327–343.
86. Salas A, Shimaoka M, Kogan AN, Harwood C, von Andrian UH, Springer TA (2004). Rolling adhesion through an extended conformation of integrin αLβ2 and relation to α I and β I-like domain interaction. *Immunity 20*, 393–406.
87. Shimaoka M, Springer TA (2003). Therapeutic antagonists and the conformational regulation of integrin structure and function. *Nat Rev Drug Disc 2*, 703–716.

# PAIN MANAGEMENT MARKET OVERVIEW

Anita Holdcroft
*Imperial College and Chelsea and Westminster Hospital, London, England*

## 1 INTRODUCTION

Worldwide an estimated one and a half billion people repeatedly endure moderate to severe pain. However, pain relief is not restricted to this population because pain is often of short duration. Both acute and chronic pain are associated with disability and distress that reduce quality of life. The International Association for the Study of Pain (IASP) has issued definitions (Table 1) and a classification of pain states based on temporal, site, and disease patterns [1]. The American Pain Foundation in 2002 through a Mason–Dixon Poll identified that more than 50 million Americans (18% of the total population, and set to increase to over 100 million in a decade) suffer chronic pain, thus making pain symptoms appear to be of epidemic proportions (http://www.painfoundation.org/page.asp?file=Library/PainSurveys.htm; accessed July 28, 2006). For example, 49% of 625 respondents had pain that varied from mild (31%), moderate (43%), moderate/severe (18%) to severe (7%), and of this group with pain, 40% of people experienced pain almost every day.

The definition of pain has recognized shortcomings because it is a sensation that is modified by biological and external influences. For example, the central nervous system actively alters the nerve transmission generated by a noxious stimulus, either by inhibition or excitation at different sites from the spinal cord to the cerebral cortex. In addition, social and cultural influences, beliefs, and biases can influence the sensation and may be linked to negative emotions such as anger, fear, and depression.

The European approach to this global reduction in quality of life in pain patients has been led by the European Federation of IASP Chapters (EFIC). This organization has called for recognition of pain not as merely a symptom but as an important health concern and a disease in its own right. It has sought partnership with groups, such as the International Alliance of Patients' Organisations (IAPO), and industry in order to represent patient interests and to build up a network of resources. These resources range from Web-based information to specialized journals and include education, curriculum development, evidence-based reports, and research tools. One of the best evidence-based reports is that from the Australian and New Zealand College of Anaesthetists and Faculty of Pain Medicine in 2005 on Acute Pain Management and it advocates acute pain management as a basic human right [2].

The response to surgery or trauma is an acute pain sensation that directly follows a noxious event and acts as a warning to seek help. When the pain persists, often the tissues have healed and different physiological systems are activated, leading to complex physical, behavioral, and psychological symptoms. These can result in dependence on medication, overuse and inappropriate use of health care, disability, poor sleep, isolation, and depression. The undertreatment of pain has been well recognized as a public health problem. Hence there have been professional and political pressures to make available and use opioid drugs to manage severe pain. Across the world the frequency of opioid use for pain relief has increased but, interestingly, at one U.S. medical center differences in the amount of prescribed

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

**TABLE 1 Definitions**

| |
|---|
| **Pain**: Pain is an unpleasant sensory and emotional experience, usually associated with actual or potential tissue damage or described in terms of such damage. The quality and quantity of pain sensations can only be determined from human reports. |
| **Acute pain**: Pain of recent onset and limited duration, usually having an identified temporal and causal relationship to injury or disease, e.g., postoperative pain, pain from sickle cell disease. |
| **Chronic pain**: Pain that persists beyond the point at which healing would be expected to be complete, or that occurring in disease processes in which healing does not take place. Chronic pain may be accompanied by psychological and social disturbance; there may be no evidence of tissue damage. |
| **Nociception**: This describes the physiological components of the neurological pathway used to detect damaging stimuli. It is a process that can be measured in animals and humans. |
| **Neuropathic pain**: Pain initiated or caused by a primary lesion or dysfunction in the nervous system. |
| **Prevalence**: Number of cases in a given population at a designated time. |
| **Incidence**: Number of new cases arising during a defined period of time. |
| **Number needed to treat (NNT)**: The number of people who have to be treated in order for one of them to achieve a specified level of benefit. |
| **Number needed to harm (NNH)**: The number of people who have to be treated in order for one of them to experience an adverse effect. |

opioids to various ethnic groups were measured. In this survey of cancer pain in over 13,000 patients, African Americans were 63% more likely than Caucasian Americans to receive no medications; daily pain was reported in 40% of African Americans but 25% of them received no analgesics [3]. Likewise local pharmacies in minority areas were less likely to have opioids available [4]. This raises issues about barriers to pain management from individual access to treatment to international regulation of analgesic supply and emphasizes the scale of inequalities in the undertreatment of pain.

The economic aspects of pain symptoms have been quantified for chronic pain. In the United States the costs are estimated to be $70 billion for time off work, medications, and health care services. Since quality of life suffers, this has effects on a patient's income, on work performance, on home making, on contribution to the household income, on welfare payments, as well as health care costs. Back pain is the most common site for pain in men, and in the United Kingdom alone this costs the economy £1632 million directly [5]. Of this, 31% is attributable to spending on hospital services, 14% to primary care, and 7% to medication. If indirect costs are included, such as workforce disruption, the total costs may be as high as £10668 million. Maximizing pain relief in this condition is, thus, an important benefit to the whole population when health services are nationally funded.

Procedures in health care facilities have been developed to recognize pain early in order to provide prompt and effective treatment. A clinical approach has been to advance the concept of pain as the fifth vital sign (the others are pulse, blood pressure, temperature, respiration) in order to ensure constant recognition and screening for pain symptoms. This approach has become standardized through a Veterans Health Administration Directive (number 2003-012) and was reviewed in the United States in 2008.

The market for analgesic preparations is thus very large yet diverse because of the multiple biopsychosocial mechanisms for initiating and maintaining the sensation of pain. In the past few years it has been subject to significant influences. One of these involved pharmacovigilance data from the cyclo-oxygenase-2 (COX-2) enzyme inhibitors (COXIBs) that reached a high profile in the media. This led to rofecoxib (Vioxx) being withdrawn in 2003. The COXIBs were developed to improve the gastrointestinal adverse event profile of nonsteroidal anti-inflammatory drugs (NSAIDs) that are prescribed worldwide to more than 80 million people. Then, after the introduction of the drug Celecoxib, clinical trial data for prevention of colonic polyps was reported [6]. The effect of the publication was to generate further substantial media hype in relation to a reported increase in cardiovascular risk. There were calls for drug withdrawal without consideration of both risks and benefits. It effectively left patients with a dramatically limited range of drugs for pain relief. The result was acutely detrimental to both the pharmaceutical industry and patients who required effective pain relief. The long-term effects of this event have been to reevaluate the use of COXIBs and to consider alternative therapies such as the nitro-NSAIDs and other inhibitors of the cycloxygenase enzyme system, for example, COX-3 inhibitors. The use of COX-2 inhibitors has decreased to 30% of the market in the United States and <10% in the UK community. The regulatory response from the European Commission was to initiate a wide ranging review and from the U.S. Food and Drug Administration (FDA) an alert. These have culminated in changes to the safety guidance issued when these type of drugs are prescribed.

Another influence on the UK market has been the effects of tighter regulation on opioid drugs following the Shipman inquiry. Doctors now fear that they can be held criminally responsible for accidental opioid overdose. In a recent poll of general practitioners one in three no longer carry opioid drugs to emergencies in the

community, and there are predications that pain relief in terminally ill patients will suffer. The proposal by England's Chief Medical Officer now out for consultation on doctor's fitness to practice has done nothing to alleviate doctors' apprehensions that have surfaced. Potential pressures on opioid use have also arisen from an increasing population prescribed long-term opioids for noncancer pain. This use of strong opioids has been facilitated first by the marketing of sustained-release oral formulations and transdermal preparations and second by the promotion of national and international (EFIC) guidance. This professionally led approach to therapeutic opioid use in chronic pain conditions is a key to their safe use [7, 8].

## 2 ANALGESIC USE

Population measures of analgesic use are available from different studies. The two population groups can be broadly divided into prescription medications and over-the-counter (OTC) purchases. One approach to quantify the data in Europe and North America has been to conduct surveys to measure painful conditions and associated analgesic use. Analgesics are the most commonly used nonprescription medications, and women report significantly more frequent analgesic use than men even when taking into account social and economic circumstances [9]. For example, 35% of women and 21% of men had used analgesics in the preceding 2 weeks in a Swedish population of 5405 (54% women). The prevalence in men was highest for pain in the back (21%), shoulders (15%), to help sleep (15%), and aching arms and legs (14%). The women had used them for pain in the shoulders (27%), in the back (26%), for sleep (23%), and arm and leg pains (19%). Pain associated with menstruation in the below-52-year-old age group was managed with analgesics in 12%; possibly over a month this figure would double. Although such self-reporting methods may have recall bias and other errors, in most surveys similar findings for the frequency of pain are reported.

The range of OTC includes aspirin, paracetamol (acetaminophen), NSAIDs, for example, ibuprofen, ketoprofen, naproxen, and combinations as well as topical preparations with local anaesthetics for ulcers and hemorrhoids. In recent studies of OTC analgesic use, a mail survey in Scotland determined the prevalence of OTC and prescription analgesic medication use. In the 2 weeks of the survey 37% of respondents had used OTC analgesics and these accounted for 59% of the market [10]. This high level of use was associated with evidence for inappropriate self-management with the potential for drug interactions with prescribed drugs. A similar concomitant use of OTC and prescription analgesics was highlighted in a Finnish survey and was related to daily or continuous pain and a high pain intensity [11]. The potential for increasing the risk of adverse effects has thus been identified and appears to be common to different study populations. In the United States, as part of a national survey, OTC use was prevalent in 74% of the population compared with prescription availability in 9% [12]. Females were more likely than males to use both OTC and prescription analgesics, and with increasing age, prescription use increased whereas OTC use decreased. In a one-month period about 30% of adults used multiple analgesics, and this again demonstrates the potential for adverse effects.

Various self-management strategies for chronic pain relief were investigated in Australia [13]. Analgesic use was defined as a passive strategy and was prevalent in 47%, while exercise was considered an active strategy and was reported by 26% of patients. The study found that use of more active strategies was associated with less disability and a reduction in the use of health care services, so these results were used to advocate community-based treatments for pain management. Overall, 68% of subjects used a combination of active and passive treatments and these patients had the best outcomes.

Analgesic costs in the Unites States have been estimated to be $23 billion for 2004 with a population of 281 million, and double-figure growth is expected in this market in the next decade, partly as a result of more effective drugs being marketed. In England the National Health Service (NHS) has established online reporting of drug costs and estimate their cost as £207 per person for 2004. Hospital pharmacies issue £2309 million of medicines, representing 22% of the total cost of medicines prescribed for 2004. In the community the prescription of analgesic drugs can be charted with some accuracy, and figures for the quantity and net ingredient costs are available online from England for 2005 (http://www.ic.nhs.uk/pubs/prescostanalysis2005). The data for analgesic drugs dispensed is listed alphabetically by chemical entity and the British National Formulary therapeutic class. Figures 1–6 are taken from this data set for opioid and nonopioid analgesics in order to demonstrate the patterns of recent prescription drug use. They amount to about £0.5 billion. Figure 1 shows the number of generic and proprietary formulations of the opioids, NSAIDs, and nonopioid analgesics. Only 4 opioids (pentazocine, codeine, diamorphine, and papaveretum) out of 17 have generic formulations compared with NSAIDs, where all have proprietary brands. Of the NSAID group, there is a large range that can be subdivided into a COX-2 group. Table 2 [14] shows the selectivity of different drugs for the cyclo-oxygenase

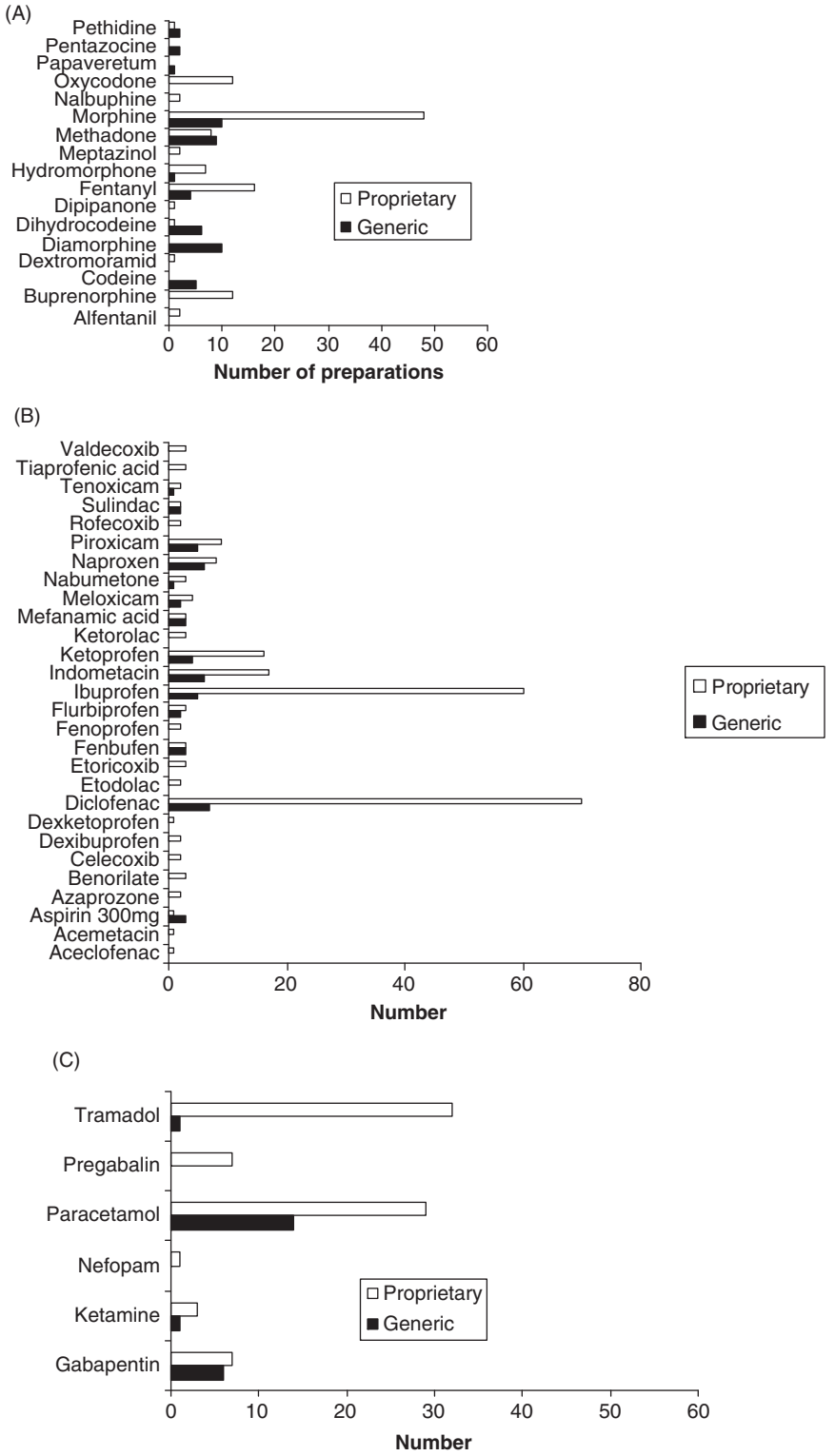


**Figure 1** Number of generic and proprietary preparations for (A) opioid drugs, (B) non-steroidal anti-inflammatory drugs (NSAIDs), and (C) other analgesics used in community dispensing in the NHS 2005.

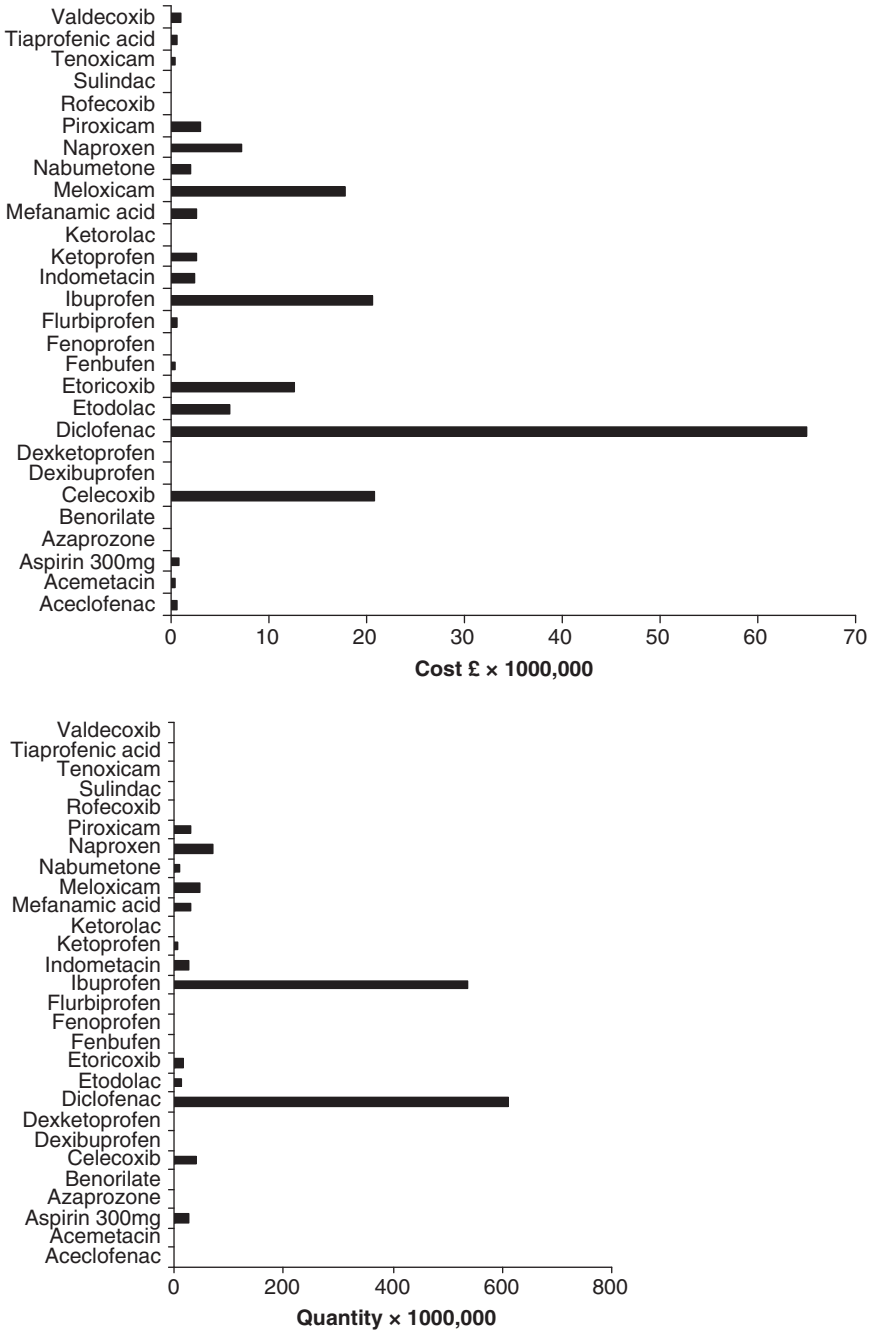


**Figure 2** Costs and quantities of NSAIDs used in community dispensing in the NHS 2005.

enzymes (COX-1 and COX-2). Of the NSAIDs, diclofenac is most commonly used, and this is reflected in the overall cost to the NHS (Fig. 2). In this figure the costs of proprietary formulations of diclofenac are five times higher than generic agents. The situation for tramadol is different (Fig. 3). Although there is one generic preparation, it has 34% of the market and in this formulation is prescribed in lower dose capsules compared to the proprietary brands. For gabapentin (Fig. 3), an antiepileptic drug widely used for neuropathic pain relief, 97% of all costs reflect generic use. It is interesting to speculate why there is this large variation in prescribing habits related to the choice of generic versus proprietary drugs for pain relief.

Figure 4 demonstrates the quantity and cost of opioid use. Fentanyl use in the community is predominantly via the transdermal route, and the cost for this type of formulation makes it the most expensive opioid

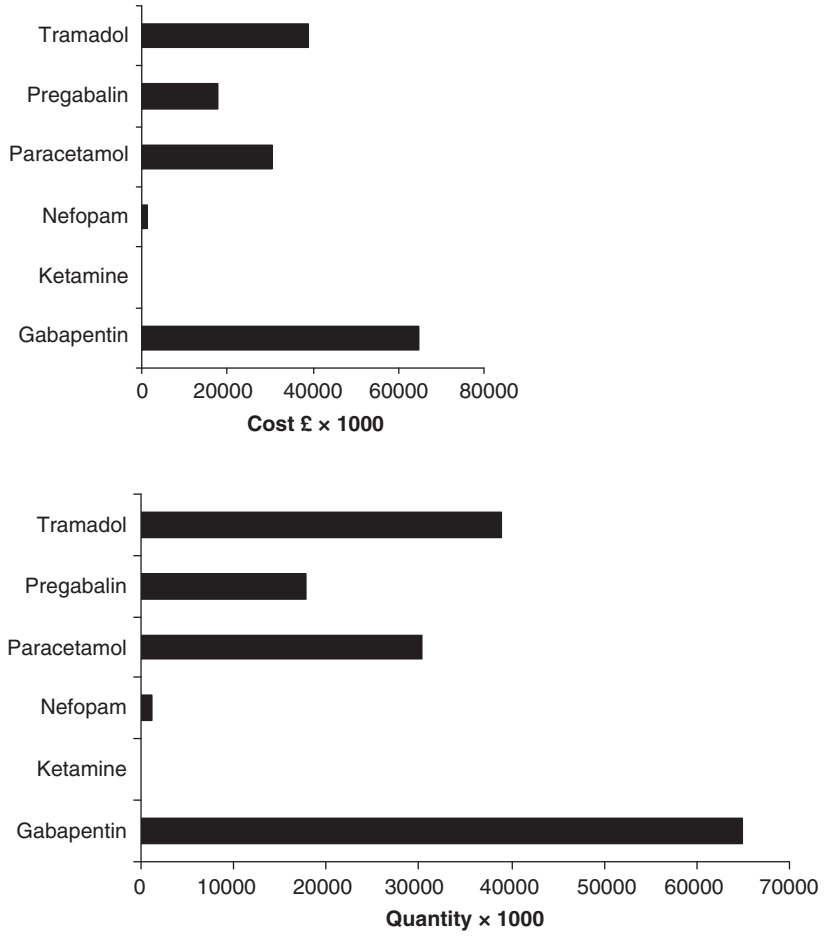


**Figure 3** Costs and quantities of other analgesics used in community dispensing in the NHS 2005.

analgesic, second only to the NSAID diclofenac. These benchmark figures provide basic data (excluding hospitals) across a national market. When efficacy is contrasted with the amount and cost of drug prescribed, codeine appears not to be cost effective. Nevertheless when used in combination with paracetamol it achieves a number needed to treat (NNT) of 2.2 (Table 3 [15]), but combined use (i.e., separate prescriptions for codeine and paracetamol) is not visible in the NHS figures unless the cocodamol preparation is prescribed (Fig. 5). The cocodamol formulation is the market leader within the small group of analgesic combinations, both for quantity of use and cost. Interestingly, it is the proprietary rather than the generic formulation that is the most widely prescribed item.

The range of formulations for NSAIDs and selected nonopioids is shown in Figure 6. Those for local application such as gels and mousse/foam are the most costly to prescribe, but the market value of these is small in comparison with overall drug costs. However, paracetamol (acetaminophen) suppositories are of higher cost than other suppositories, and so it has become economical to deliver this drug intravenously in hospitals. The most recently introduced preparation differs from its prodrug propacetamol, which can also be given intravenously after reconstitution. The solubilized preparation has less local effects and can be infused without preparation. In the community over 90% of the paracetamol used is generic.

## 3 ANALGESIC EFFICACY

One of the evidence-based approaches to analgesic use has been to compare the efficacy and side effect profile using a comparative rating based on the number needed to treat (NNT) and the number needed to harm (NNH). The ratings for NNT are calculated on the number of patients reporting at least 50% better pain relief than placebo in randomized, double-blind, single-dose studies in patients with at least moderate pain [15]. This measure of short-term efficacy has been criticized because it does

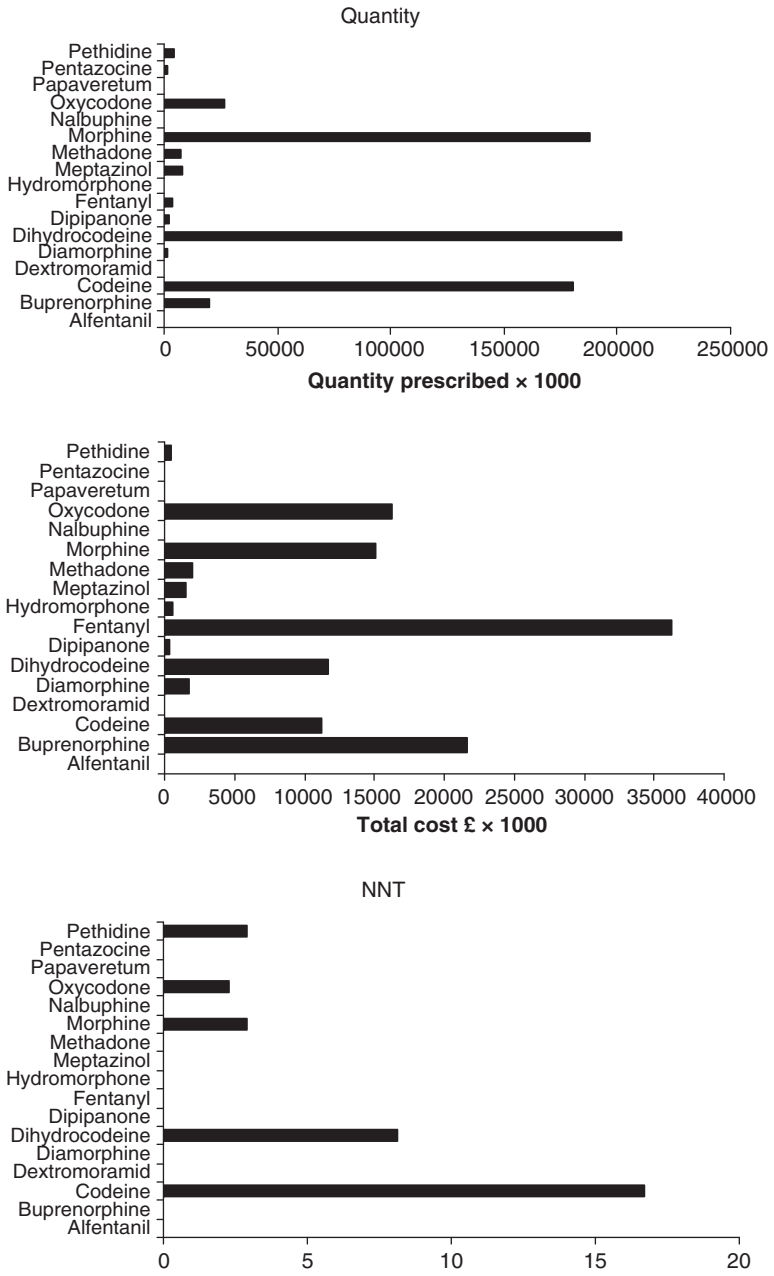


**Figure 4** Quantity of prescriptions dispensed in the community and the cost of opioid drugs in England in 2005 compared with acute pain relief efficacy as number needed to treat (NNT) [15].

not relate to pain management in clinical practice where drugs are given repeatedly, often in combination. As an alternative evidence-based approach, *Cochrane Reviews* may provide a broader analysis of pain management. For example, in chronic pain where NNTs have not been studied, the use of opioid drugs has been controversial because of their inherent dependency profile. The results of short-term use provide no evidence for opioid efficacy, but for an intermediate period of about 28 days, opioids did reduce neuropathic pain [16]. However, side effects of opioids in this review were common and the calculated NNH are shown in Table 4.

## 4 SELECTION OF ANALGESICS

### 4.1 Pain Measures

Selecting an appropriate analgesic regimen relies on pain measurement and diagnosis. The selection is then

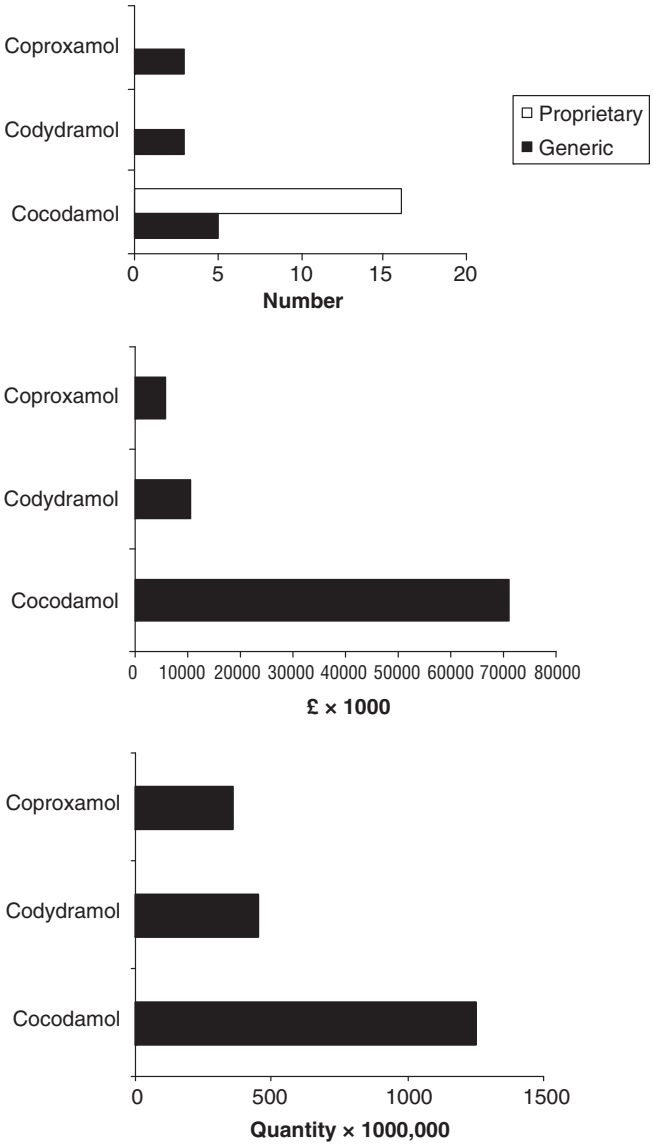


**Figure 5** Number, cost, and quantity of analgesic combinations used in community dispensing in the NHS 2005.

modified according to the patient's response. Nociceptive pain is classically described as originating either in somatic or visceral tissues (or both) mainly because there is a distinct clinical presentation. Somatic pain is described as being well localized, often sharp, whereas visceral pain may present as dull, colicky pain that is poorly localized and often referred to somatic segments and accompanied by autonomic activity such as nausea, sweating, and cardiovascular changes. Interestingly it is somatic pain that has most frequently been researched, and most of the research for NNTs has been related to somatic rather than visceral pain. However, they both present frequently and can be present in the same patient.

Neuropathic pain can be acute or chronic. It is described typically with words such as burning, shooting, and stabbing, indicating the paroxysmal and spontaneous onset of the pain with no clear precipitating features. In addition, there may be altered sensation in skin (e.g., allodynia, where a previously nonpainful stimulus becomes painful) and muscle (e.g., tenderness) or phantom pains when part of the body has been removed (e.g., limb or organ).

Nociceptive and neuropathic pains were thought to respond differently to different analgesics. However, the opioid drugs are effective in both pain types, and it has become acceptable for them to be used in the long term under tightly controlled conditions for pain that is noncancer related. Meanwhile, basic science research is unfolding the mechanisms for specific therapies in relation to receptor sites and nerve membrane activity (Table 5).

Apart from the pain type, the pain intensity is a valuable measure and it depends primarily on patient report. Thus, the influence of sleep, environment, mood, medications, and the like can alter the response to the following questions: How intense is your pain? Is it mild, moderate, or severe? Repeated measures using this standardized verbal questioning can monitor analgesic therapy. Pain can be measured at rest and during exercise, if appropriate, since recovery of function is often an acceptable outcome rather than an overall pain reduction. Verbal scales are simple and fast to use and can be applied to about three-quarters of all patients. Children and patients with special needs require modifications. Some pain measures characterize different elements of the pain sensation and are more applicable to chronic pain where behavioral and emotive responses are important to assess, for example, the McGill Pain Questionnaire.

In clinical trials pain intensity and pain relief are two standard measures taken over time. A change in pain intensity of at least 30% is considered clinically relevant, but most analgesics marketed have achieved 50% in selected patient groups. For patients an important outcome is the overall effect, including improved function, better sleep, and fewer adverse effects. These can be incorporated into a global assessment.

### 4.2 Step-up and Step-down Approach to Pain Management

The World Health Organisation (WHO) and the World Federation of Societies of Anaesthesia have transformed cancer pain relief through a structured therapeutic approach to pain intensity. A three-step approach is advocated that goes up and down the steps depending on pain intensity. For mild pain, the first step, nonopioid

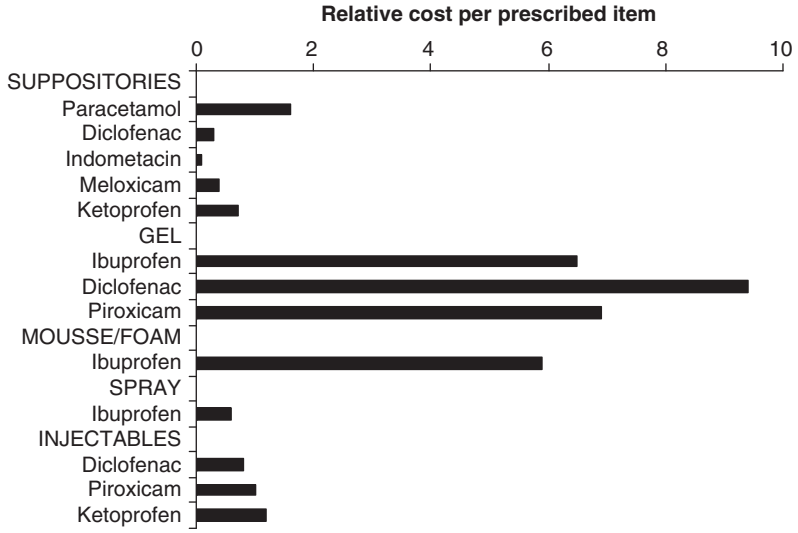


**Figure 6** Relative cost per prescribed item of nonoral formulations of NSAIDs and selected nonopioid drugs in NHS community dispensing 2005.

**TABLE 2 Examples of COX-1 and COX-2 Selectivity [14]**

| NSAID | COX-2:COX-1 Ratio |
|---|---|
| Celocoxib | 7.6:1 |
| Diclofenac | 3.0:1 |
| Etodolac | 2.4:1 |
| Etoricoxib | 106:1 |
| Ibuprofen | 0.2:1 |
| Indometacin | 0.4:1 |
| Meloxicam | 2.0:1 |
| Naproxen | 0.3:1 |
| Piroxicam | 0.08:1 |
| Rofecoxib | 35:1 |
| Valdecoxib | 30:1 |

analgesics are first-line medications. Step 2 involves weak opioids with or without nonopioids, and the third step includes strong opioids again with or without nonopioids. At each step, if needed, adjuvants can also be used and the intermediate step omitted. The outcome for pain management has been found satisfactory in up to 90% of patients. Those 10% of patients whose pain is not adequately managed need other strategies such as anticonvulsants, antidepressants, other adjuvants and alternative routes of drug administration, nerve blocks, or other invasive procedures.

Optimizing results has become important because the ultimate goal is to achieve pain relief quickly and with minimal adverse effects. Thus, the market has increased for anticonvulsant drugs as analgesics for

**TABLE 3 Ranked Order of Number Needed to Treat (NNT) of Different Groups of Analgesics Based on Acute Pain Studies Reported by Bandolier [15]**

| Drug | Number of Patients | NNT |
|---|---|---|
| *NSAIDs* | | |
| Valdecoxib | 473 | 1.6 |
| Ibuprofen | 76 | 1.6 |
| Ketorolac | 69 | 1.8 |
| Rofecoxib | 1900 | 1.9 |
| Diclofenac | 411 | 1.9 |
| Piroxicam | 30 | 1.9 |
| Naproxen | 257 | 2.3 |
| Aspirin | 5061 | 4.4 |
| *Combination preparations* | | |
| Paracetamol 1 g and codeine 60 mg | 197 | 2.2 |
| Paracetamol 500 mg and oxycodone 5 mg | 150 | 2.2 |
| Paracetamol 650 mg and dextropropoxyphene 65 mg | 963 | 4.4 |

**TABLE 4 Number Needed to Harm (NNH) for Opioids Used in Management of Neuropathic Pain [16]**

| Symptom | NNH |
|---|---|
| Nausea | 4.2 |
| Constipation | 4.2 |
| Drowsiness | 6.2 |
| Dizziness | 7.1 |
| Vomiting | 8.3 |

**TABLE 5 Analgesic Drugs and Their Sites of Action**

| Functional Site | | |
|---|---|---|
| *Receptors* | *Clinical Drug* | *Comments* |
| NMDA (*N*-methyl-d-aspartate) | Ketamine | Anesthetic—dose related. Dextromethorphan (NMDA antagonist) is in clinical trials for neuropathic pain. Memantine is in clinical trials as a glutamate receptor antagonist. |
| $GABA_A$ | Nitrous oxide | Gaseous (Entonox is 50:50 mixture with oxygen). |
| | Benzodiazepines | Sedation, muscle relaxation. |
| | Ethanol | Social use. |
| | Volatile anaesthetics | Used in low dose, sedative. |
| | Steroids | Also act as disease modifiers; ganaxalone as an anticonvulsant. |
| $GABA_B$ | Baclofen | Antispasmodic. |
| CGRP | None | Endogenous ligand is calcitonin gene-related peptide (CGRP). |
| VIP | None | Endogenous ligand is vasoactive intestinal peptide (VIP). |
| TrkA | None | Nerve growth factor (NGF) antagonists are potential analgesics. |
| Tachykinin | None | Endogenous ligand is substance P. |
| Kinin | None | Endogenous ligand is Bradykinin. BK1 antagonists are potential analgesics. |
| Acetylcholine (nicotinic) | None | Epibatidine, though toxic, is the lead compound. |
| 5HT | None | Endogenous ligand is serotonin (5 hydroxytryptamine). Site of action of antidepressants, e.g., amitriptyline, venlafaxine, duloxetine, milnacipram. |
| Purine (P2X) | None | Endogenous ligand is ATP. |
| Adrenergic $\alpha_2$ | Clonidine, dexmedetomidine | Xen2174, a chi-conopeptide, is a potential analgesic. |
| Opioid | Morphine, fentanyl, diamorphine (heroin), hydromorphone, codeine, sufentanil, oxycodone, oxymorphone, dextropropoxyphene, methadone, pethidine, etorphine, pentazocine,* buprenorphine,* nalbuphine* | *Partial kappa agonists. |
| Cannabinoid | Tetrahydrocannabinol (THC), Nabilone, CBME** | **Plant derived. |
| Vanilloid | Capsaicin | Receptor activated by heat. |
| *Nerve Membrane* | | |
| Sodium channels | Antiarrhythmics, e.g., lidocaine, mexilitene. Anticonvulsants: gabapentin, pregabalin, carbamazepine, phenytoin, valproate, oxcarbazepine, lamotrigine, topiramate, tiagabine | |
| Calcium channels | Ziconotide*** Nifedipine | ***Selective N-type blocker. |
| Denervation | Botulinum toxin | |
| *Peripheral to the Nerve* | | |
| Cyclooxygenase | COX-1 and COX-2 inhibitors | COX-3 inhibition may be the mechanism for paracetamol analgesia. |
| Nitric Oxide (NO) | NO-based drugs | NO- NSAIDs, NO-paracetamol are in preclinical tests. |
| *Cell Nucleus* | | |
| Gene therapies | None | Some inherited painful neuropathies may respond; activation of glia in neuropathic pain states. |
| *Synapse* | | |
| Synaptic vehicle exocytosis | Levetiracetam | Anticonvulsant limited clinical data in neuropathic pain. |

chronic pain conditions because of their sodium channel blocking effects on nerve membranes. In the United Kingdom about 20% of analgesic costs in the community are for gabapentin and pregabalin.

Some evidence now suggests that early application of invasive procedures in cancer pain may result in better pain relief and less side effects [17]. Thus, there are recommendations to find evidence for invasive procedures being an effective part of each step and not used as a last resort. It is in this context that a new intrathecal peptide analgesic, ziconotide, has entered the market. It is the first and only neuronal-type calcium channel blocker. At present there have been no comparative studies with other intrathecal analgesics. All invasive procedures carry a risk of infection, in this case meningitis, but the adverse effects specific to ziconotide include depression, cognitive impairment, and hallucinations. These relatively severe psychiatric risks preclude use of this drug for routine pain relief. It is, therefore, appropriate to consider the record of intrathecal drugs such as local anaesthetics and opioids (e.g., fentanyl, sufentanil) that are routinely used for relief of severe pain. Pharmacovigilance data for severe adverse effects to them is comparatively rare [18]. Furthermore, specific opioid antagonists for peripheral activity such as methylnaltrexone or alvimopan can reduce gastrointestinal dysfunction associated with acute opioid use.

The therapeutic opportunities for the WHO steps are shown in Table 6 for different painful conditions. Cancer pain affects about 1% of the U.S. population and requires a large range of analgesic agents. Opioids are the standard medication, but breakthrough pain is an important issue, so some drug formulations have been successfully developed to overcome this (e.g., Actiq). Postoperative pain is common with over 50 million procedures in the United States annually, yet many patients complain of inadequate relief or unwanted adverse effects. Opioids and NSAIDs are the standard analgesics used after surgery, and the use of more than one analgesic therapeutic group is essential in order to minimize pain. Postoperative pain varies with the type and site of surgery and is usually managed by intraoperative analgesia and local anaesthetic nerve blocks followed by regular parenteral and oral medications. The majority of pain management is on the WHO step-down basis from parenteral to oral medications. Neuropathic pain is associated with diseases that are increasing in frequency such as diabetes and human immunodeficiency virus (HIV) infection. This type of pain is not life-threatening and often continues throughout a lifetime, so this is an important market given the present lack of drug specificity and the high costs of drugs licensed for use in neuropathic pain. New agents being considered for long-term neuropathic pain management include those that prevent neural degeneration such as prosaptides [19].

The step-up and step-down approach to pain management identifies the opportunities for combined therapies, but only of analgesic drugs. Table 7 [20] lists the multiple strategies that can be used simultaneously to achieve adequate pain relief; they include pharmacological, physical, and psychobehavioral approaches. For analgesic drugs, the advantages of using more than one drug are that more than one neural site for pain transmission can be targeted, drug dosage can often be reduced and hence side effects, and if the formulation is in one preparation then costs of prescribing and distribution will be reduced—providing efficiency to patients and prescribers.

### 4.3 The At-Risk Patient

Over many years drug development has precluded at-risk groups of patients such as pregnant women and the young from clinical trials. This approach has led to many drugs being used off licence in these populations. In the next few years this situation should change with the introduction in the European Union of postmarketing pharmacovigilance. Another at-risk group includes the elderly because of the decline in normal renal and hepatic function. Pain is frequent in the elderly, and this group of patients is expected to increase in future because of longer life expectation.

Another approach to defining who is at risk is to consider new developments in personalized medicine. Women are overall more likely to suffer adverse drug effects than men, and this finding is not wholly related to physical differences. These effects can determine the pattern of prescription and OTC drug use and behavior toward analgesics. Hence, any method to improve compliance or drug selection would be welcomed by patients, prescribers, and pharmaceutical companies. Thus, the concept of designer analgesics is now advancing through stages such as pharmacogenetics where genetic factors such as the CYP450 (cytochrome P450) liver enzymes are identified, pharmacogenomics where the sites of action of drugs/diseases are characterized, and phenotyping where the metabolic profile of drugs can be determined, for example, the conversion of codeine to morphine.

Others at risk are those who are dependent on opioids. One approach has been the design of a new formulation of oxycodone, Remoxy, that is sticky, which prevents it being crushed or snorted. A market of over $1 million is forecast for this unique abuse-resistant product.

**TABLE 6 Choice of Standard Agents for Selected Painful Conditions**[a]

| | 2, 3 | 1, 2 | 1, 2 | 1, 2 | 1, 2, 3 | 1, 2, 3 | 1, 2, 3 | 1, 2, 3 |
|---|---|---|---|---|---|---|---|---|
| WHO Step | Opioids | Paracetamol | NSAIDs | COXIBs | Anticonvulsants | Antidepressants | Disease Specific and Modifying*, **, *** | Site Specific, e.g., Local Anaesthetics |
| Cancer pain | Y | Y | Y | Y | Y | Y | Y | Y |
| Postoperative pain | Y | Y | Y | | | | | Y |
| Neuropathic pain | (Y) | | (Y) | | Y | Y | | Y |
| Arthritis | (Y) | Y | Y | Y | | | Y | |
| Back pain | Y | | Y | | | Y | | Y |

[a]Y = in regular use; (Y) = used in selected patients; WHO steps 1 = mild, 2 = moderate, 3 = severe pain;
*colchicine for gout;
**aromatase inhibitors for endometriosis;
***aldose reductase inhibitors (ranirestat, fidarestat) and protein kinase C inhibitors (ruboxistaurin) for diabetic neuropathy.

**TABLE 7 Nonranked Potential List of Multidisciplinary Strategies Used in Combination for Pain Management [20]**

| Situational | Physical | Pharmacological |
|---|---|---|
| Cognitive | Thermal—cold, heat | Analgesics |
| Behavioral | Exercise | Opioids |
| Distraction | Traction | NSAIDs |
| Psychosocial | Injection—nerve blocks | Paracetamol |
| Relaxation | Surgery | Antidepressants |
| Music | TENS | Anticonvulsants |
| Hypnosis | Reflexology | Local anesthetics |
| Biofeedback | Massage | Tramadol |
| Self-help | Physiotherapy | Steroids |
| Networking | Osteopathic manipulation | Cannabinoids |
| Counseling | Acupuncture | Adrenergic antagonists |

**TABLE 8 List of Analgesic Drug Delivery Methods**

| Route | Special Qualities |
|---|---|
| Intravenous | No swallowing |
| Subcutaneous | Spill resistant |
| Intrathecal | Controlled release |
| Epidural | Pumps internal or external |
| Intramuscular | Iontophoresis |
| Oro/gastric | Transdermal |
| Oro/mucosal | |
| Nasal | |
| Pulmonary | |
| Topical | |
| Local | |
| Rectal | |

## 5 DRUG DEVELOPMENT

The mechanistic approach to development of analgesics has been well researched, but there are few compounds in clinical use compared with those available to laboratory workers for antinociceptive studies. However, in recent years innovative approaches to pain relief have included evidence-based professional guidelines [2], combining therapeutic strategies, improvements in drug formulations, and identification of patient and prescriber behaviors that limit effective drug use.

The opportunity to use old drugs in new formulations has made a welcome improvement in delivering drugs to patients, especially the opioids (Table 8). They have increased time between doses, removed messy delivery systems, and added secure and safer dosing regimes.

A structured approach to the analgesic market has been to identify the frequency and type of pain, how it is managed (e.g., mechanism-based and evidence-based approaches), and what combinations of therapies are used, what the unfulfilled needs are for OTC and prescribers, and how to improve therapeutic strategies to achieve maximum benefit for patients.

## REFERENCES


1. Merskey H, Bogduk N, Eds. (1994). *Classification of Chronic Pain: Descriptions of Chronic Pain Syndromes and Definitions of Pain Terms*, 2nd ed. Seattle, WA: IASP Press.
2. Australian and New Zealand College of Anaesthetists and Faculty of Pain Medicine (2005). *Acute Pain Management: Scientific Evidence Second Edition*. Victoria, Australia: Australian and New Zealand College of Anaesthetists. Available: www.nhmrc.gov.au/publications.
3. Bernabei R, Gambassi G, Lapane K, Landi F, Gatsonis C, Dunlop R, Lipsitz L, Steel K, Mor V (1998). Management of pain in elderly patients with cancer. SAGE Study Group. Systematic assessment of geriatric drug use via epidemiology. *JAMA 279*, 1877–1882.
4. Morrison RS, Wallenstein S, Natale DK, Senzel RS, Huang LL (2000). "We don't carry that"—Failure of pharmacies in predominantly nonwhite neighborhoods to stock opioid analgesics. *New Engl J Med 342*, 1023–1026.
5. Maniadakis N, Gray A (2000). The economic burden of back pain in the UK. *Pain 84*, 95–103.
6. Solomon SD, McMurray JJ, Pfeffer MA, Wittes J, Fowler R, Finn P, Anderson WF, Zauber A, Hawk E, Bertagnolli M, Adenoma Prevention with Celecoxib (APC) Study Investigators (2005). Cardiovascular risk associated with celecoxib in a clinical trial for colorectal adenoma prevention. *New Engl J Med 352*, 1071–1080.
7. Kalso E, Allan L, Dellemijn PLI, Faura CC, Ilias WK, Jensen TS, Perrot S, Plaghki LH, Zenz M (2003). Recommendations for using opioids in chronic non-cancer pain. *Eur J Pain 7*, 381–386.
8. The Pain Society (March 2004). Recommendations for the appropriate use of opioids for persistent non-cancer pain. A consensus statement prepared on behalf of the Pain Society, the Royal College of Anaesthetists, the Royal College of General Practitioners and the Royal College of Psychiatrists, London.
9. Isacson D, Bingefors K (2001). Epidemiology of analgesic use: A gender perspective. *Eur J Anaesthesiol 19*(Suppl 26), 5–15.
10. Porteous T, Bond C, Hannaford P, Sinclair H (2005). How and why are nonprescription analgesics used in Scotland? *Fam Pract 22*, 78–85.
11. Turunen JH, Mantyselka PT, Kumpusalo EA, Ahonen RS (2005). Frequent analgesic use at population level: Prevalence and patterns of use. *Pain 115*, 374–381.
12. Paulose-Ram R, Hirsch R, Dillon C, Losonczy K, Cooper M, Ostchega Y (2003). Prescription and non-prescription analgesic use among the US adult population: Results

from the third National Health and Nutrition Examination Survey (NHANES 111). *Pharmacoepidemiol Drug Saf 12*, 315–326.

13. Blyth FM, March LM, Nicholas MK, Cousins MJ (2005). Self-management of chronic pain: A population-based survey. *Pain 113*, 285–292.
14. Riendeau D, Percival MD, Brideau C, Charleson S, Dube D, Ethier D, Falgueyret JP, Friesen RW, Gordon R, Greig G, Guay J, Mancini J, Ouellet M, Wong E, Xu L, Boyce S, Visco D, Girard Y, Prasit P, Zamboni R, Rodger IW, Gresser M, Ford-Hutchinson AW, Young RN, Chan CC (2001). Etoricoxib (MK-0663): Preclinical profile and comparison with other agents that selectively inhibit cyclooxygenase-2. *J Pharmacol Exp Ther 296*, 558–566.
15. Bandolier. Available: http://www.jr2.ox.ac.uk/bandolier/booth/apinpag/Acutrev/Analgesics/lftab.html. Accessed July 28, 2006.
16. Eisenberg E, McNicol E, Carr D (2006). Opioids for neuropathic pain. *Cochrane Database Syst Rev Jul 19*(3), CD006146.
17. Wong GY, Schroeder DR, Carns PE, Wilson JL, Martin DP, Kinney MO, Mantilla CB, Warner DO (2004). Effect of neurolytic celiac plexus block on pain relief, quality of life, and survival in patients with unresectable pancreatic cancer: A randomized controlled trial. *JAMA 291*, 1092–1099.
18. Holdcroft (2007). UK drug analysis prints and anaesthetic drug reactions. *Pharmacoepidemiol Drug Saf 16*, 316–328.
19. Jolivalt CG, Ramos KM, Herbetsson K, Esch FS, Calcutt NA (2006). Therapeutic efficacy of prosaposin-derived peptide on different models of allodynia. *Pain 121*, 14–21.
20. Holdcroft A, Berkley KJ (2006). Sex and Gender Differences in Pain and Its Relief, 5th ed., *Wall and Melzack's Textbook of Pain*, McMahon S, Koltzenburg M, Eds. Edinburgh: Elsevier Churchill Livingstone, pp. 1181–1197.

# ANTIVIRAL MARKET OVERVIEW

Anita Arora, Natalia Mendoza, and Stephen K. Tyring
*Center for Clinical Studies, Houston, Texas*

## 1 INTRODUCTION

According to the latest figures published today in the UNAIDS/WHO 2006 AIDS Epidemic Update, an estimated 39.5 million people are living with the human immunodeficiency virus and 170 million people are thought to be infected with hepatitis C virus [1]. These figures just illustrate that the number of viral infections are rising and more and better antivirals are urgently required. Approximately 40 antiviral drugs have been officially approved for clinical use [1], and at least half of them are used in the treatment of human immunodeficiency virus (HIV) infections (acquired immuno deficiency syndrome; AIDS). The other antivirals that are readily available include those for the treatment of herpes simplex virus (HSV), varicella-zoster virus (VZV), cytomegalovirus (CMV), influenza virus, respiratory syncytial virus (RSV), and hepatitis B (HBV) and C virus (HCV) infections [2].

There are several other important virus infections for which no antiviral drugs have been developed; even for those infections for which antiviral drug therapy exists, there is still substantial room for advancement in terms of higher potency and/or increased selectivity or safety. The approaches toward the treatment of HIV, HBV, and HCV, herpesvirus, human papillomavirus (HPV), and poxvirus infections will be reviewed in this context.

## 2 ANTIVIRAL DRUGS FOR HUMAN IMMUNODEFICIENCY VIRUS INFECTIONS

Currently, there are over 20 antiretroviral drugs approved by the Food and Drug Administration (FDA) falling into 5 categories: nucleoside analogs (zidovudine, lamivudine, didanosine, zalcitabine, stavudine, abacavir, emtricitabine), nucleotide analogs (tenofovir), nonnucleoside analogs (nevirapine, delavirdine, efavirenz), protease inhibitors (saquinavir, indinavir, ritonavir, nelfinavir, amprenavir, fosamprenavir, atazanivir, lopinavir, tipranavir, darunavir), and fusion inhibitors (enfuvirtide). Five combination antiretrovirals exist to increase medication compliance: Combivir (zidovudine/lamivudine), Trizivir (zidovudine, lamivudine, abacavir), Truvada (tenofovir/emtricitabine), Epzicom (lamivudine/abacavir), and the latest approved (July 2006) Atripla (efavirenz, emtricitabine, tenofovir). Highlights of each class of drug will be described before development of new therapeutics agents is discussed further. The goals of using antiretroviral drugs are to reduce viral load for a prolonged period and delay disease progression. General guidelines for treatment initiation of antiretroviral therapy are reviewed in Table 1 and dosages are listed in Table 2.

### 2.1 Nucleoside Analogs

Nucleoside reverse transcriptase inhibitors (NRTIs) were the first family of antiretrovirals to be created against HIV. Zidovudine was the first FDA-approved (1987) drug for HIV infection. NRTIs are activated following intracellular phosphorylation. The triphosphorylated metabolites attach to viral reverse transcriptase and are integrated into the growing deoxyribonucleic and acid (DNA) chain where they bring about DNA termination. The end result is that they inhibit ribonucleic acid (RNA)-dependent DNA synthesis [3].

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

**TABLE 1 Overview of Guidelines for Initiation of Antiretroviral Therapy in Adults**

| |
|---|
| AIDS disease |
| Symptomatic infection |
| CD4 count < 200/mm$^3$ |
| Plasma HIV RNA levels of >55,000 copies/mL (RT-PCR) |
| Plasma HIV RNA levels of >30,000 copies/mL (RT-PCR) |
| Consider treatment in patients with CD4 count 200–350/mm$^3$ |

Resistance to one NRTI can often lead to resistance to all of the others.

### 2.2 Nucleotide Analogs

Tenofovir is a nucleotide analog that is used in combination therapy (Truvada, tenofovir/emtricitabine). Its mechanism of action is similar to nucleoside analogs, causing chain termination by competitive inhibition. Trials have demonstrated that the use of tenofovir in highly active antiretroviral therapy (HAART) is successful at suppressing RNA titers of resistant strains. Fortunately, cross resistance with nucleoside analogs has not yet been described [4–6].

### 2.3 Nonnucleoside Analogs

Nonnucleoside reverse transcriptase inhibitors (NNRTI) affect replication at the reverse transcriptase (RT) stage, like nucleoside analogs. However, they utilize a different mechanism of action by binding noncompetitively to the viral RNA-dependent DNA polymerase, altering its structure, and thus inhibiting its function. NNRTIs are active in their native state and do not require phosphorylation. As with other antiretrovirals, monotherapy is linked with the emergence of resistance and cross resistance exists among NNRTIs. Nonnucleoside analogs are approved for use in combination therapy against HIV-1 infections and are not active against HIV-2 infections [7, 8].

### 2.4 Protease Inhibitors

Protease inhibitors (PI) became widely available in 1996, and their mechanism of action blocks the protease enzyme involved in the final processing of viral proteins, subsequently resulting in decreased assembly of new progeny viruses from newly infected or chronically infected cells. This is a distinct advantage over RT inhibitors, which only inhibit replication in newly infected cells. PIs are currently used in combination with NRTI or NNRTI. This combination results in a 2-log decrease in viral load in the majority of HIV-infected individuals

**TABLE 2 Adult Dosages of FDA-Approved Antiretroviral Agents**

| Generic Name | Common Abbreviations | Adult Dosages[a] |
|---|---|---|
| *Nucleoside Reverse Transcriptase Inhibitors* | | |
| Zidovudine | AZT, ZDV | 300 mg b.i.d., 200 mg t.i.d. |
| Didanosine | ddI | 200 mg b.i.d., 400 mg q.d. |
| Zalcitabine | ddC | 0.75 mg t.i.d. |
| Stavudine | d4T | 20–40 mg b.i.d. |
| Lamivudine | 3TC | 150 mg b.i.d., 300 mg q.d. |
| Abacavir | ABC | 300 mg b.i.d. |
| Emtricitabine | FTC | 200 mg q.d. |
| Zidovudine + Lamivudine | AZT + 3TC | one b.i.d. |
| Zidovudine + Lamuvidine+ abacavir | AZT + 3TC + ABC | one b.i.d. |
| Lamivudine + Abacavir | 3TC + ABC | one q.d. |
| Tenofovir + Emtricitabine | TFV + FTC | one q.d. |
| Efavirenz + emtricitabine+ tenofovir[b] | EFV + FTC + TFV | one q.d. |
| *Nucleotide Reverse Transcriptase Inhibitor* | | |
| Tenofovir | TFV, TDF, PMPA | 300 mg q.d. |
| *Nonnucleoside Reverse Transcriptase Inhibitors* | | |
| Nevirapine | NVP | 200 mg b.i.d. |
| Delavirdine | DLV | 400 mg t.i.d. |
| Efavirenz | EFV | 600 mg q.d. |
| *Protease Inhibitors (PIs)* | | |
| Saquinavir | SQV | 400 or 1000 mg b.i.d. |
| Indinavir | IDV | 800 mg q8h |
| Ritonavir | RTV | 600 mg b.i.d. |
| Nelfinavir | NFV | 1250 or 750 mg t.i.d. |
| Amprenavir | APV | 1200 mg b.i.d. |
| Fosamprenavir | f-APV | 1400 mg b.i.d. |
| Atazanavir | ATV | 400 mg q.d. |
| Lopinavir + Ritonavir | LPV/r | Three b.i.d. |
| *Fusion Inhibitors (FIs)* | | |
| Enfuvirtide | T-20 | 90 mg s.c. b.i.d. |

[a]Frequently used dosages: b.i.d. = twice daily, t.i.d. = three times a day; q.d. = every day; q8h = every 8 hours; s.c. = subcutaneously.
[b]Nucleotide reverse transcriptase inhibitor.

while RT inhibitors used in monotherapy only produce a 0.5-log decrease [9].

### 2.5 Fusion Inhibitors

Fusion inhibitors, a new novel class of antiretroviral drugs, bind to proteins on the viral envelope and inhibit the conformational change required for fusion between the viral envelope and CD4 cells. Currently, the only drug in its class, enfuvirtide, has not demonstrated cross resistance with other antiviral medications. It is administered by abdominal subcutaneous injections twice daily.

### 2.6 Therapeutic Investigational Drugs

Although antiretroviral medications are fairly new in the armamentarium of medications against HIV infection, resistance to antiretrovirals can already be encountered in treatment-naïve patients [10]. HIV has generated more efforts toward antiviral drug development than any other virus. As more is identified in the cycle of viral replication, new targets for treatments are being studied. Each step in the entry of HIV into cells is now being targeted for therapeutic design (Table 3). Furthermore, new nucleoside analogs, NNRTIs, and PIs with increased activity against resistant strains are being studied for their use in combination therapy [11, 12].

**TABLE 3 Emerging Drug Targets for Antiretroviral Therapy**

| |
|---|
| Entry Inhibitors |
| • *Attachment* |
| Cyanovirin-N |
| Cyclotriazadisulfonamide analogs |
| PRO 2000 |
| TNX 355 |
| PRO 542 |
| BMS 806 |
| Peptide T |
| • *Co-receptor binding* |
| SCH-D (CCR5 antagonist) |
| Maraviroc (CCR5 antagonist) |
| Aplaviroc (CCR5 antagonist) |
| PRO 140 (CCR5 antagonist) |
| AMD 070 (CXCR4 antagonist) |
| • *Fusion* |
| Enfuvirtide (using needle-free injection device) |
| Tifuvirtide |
| Integrase Inhibitors |
| L870810 |
| JTK-303 |
| Maturation Inhibitors |
| PA 457 |

***New Targets: Entry*** The steps of viral entry present three unique targets for inhibition: (1) attachment of virions to the cell surface via nonspecific interactions and CD4 binding, (2) co-receptor binding, and (3) fusion. The envelope protein (Env) on the surface of virions regulates HIV's entry into target cells [13, 14]. Env is composed of an outer protein, glycoprotein (gp) 120, and a transmembrane protein, gp 41, which are brought together on the surface of the virion. Env interactions with cell surface molecules allow virions to attach to target cells, and then permit entry subsequent to specific, sequential interactions with its cell surface receptor CD4 and co-receptor CCR5 or CXCR4 [15, 16]. In addition, some entry inhibitors are being investigated as potential microbicides to prevent mucosal transmission of HIV [17, 18].

*Attachment Inhibitors* HIV attachment inhibitors include agents that interact with the virus or the target cell to inhibit virion absorption/attachment or Env–receptor interactions. Attachment inhibitors can be further categorized into those with nonspecific or specific modes of action. Nonspecific inhibitors of HIV adsorption/attachment to the cell surface include soluble polyanions, such as dextran sulfate and heparin, that aim at basic regions of gp120 to inhibit cell surface binding and co-receptor interactions [19, 20]. Specific inhibitors (targeting CD4) include agents that down-modulate CD4, ligands to CD4 (both of which have potential outcomes on immune function), and agents that bind to Env to prevent Env–CD4 binding or functional interactions. Several compounds are currently progressing through clinical trials [21].

*Co-receptor Binding* Inhibitors of co-receptor interactions characterize a second group of entry inhibitors that include proteins, small molecules, and antibodies. These co-receptor-binding inhibitors interact either with CCR5, CXCR4, or Env to preclude functional Env–co-receptor interactions. CCR5 using viruses (R5 tropic) are involved in HIV transmission. Subsequently in infection, approximately 20–50% of individuals acquire viruses that can use CXCR4 as well as CCR5 (dual tropic R5/X4). A small percentage of the population acquires viruses that use CXCR4 but not CCR5 (X4 tropic) [22, 23]. The chemokine macrophage inflammatory protein (MIP)-1α/CCL3, MIP-1β/CCL4, and RANTES/CCL5 bind to CCR5, and stromal cell-derived factor-1-α (SDF-1) binds CXCR4 to prevent viral infection, respectively [24–26]. Chemically adapted chemokine derivatives, which have been produced with more

effective HIV inhibitory activity, are currently being investigated in clinical trials.

*Fusion Inhibitors* Fusion inhibitors represent the first group of entry inhibitors for which there is an approved drug. Enfuvirtide, a 36-amino-acid peptide, was FDA approved in 2003 for use together with other antiretrovirals in treatment-experienced individuals with evidence of HIV-1 replication despite antiretroviral usage. Enfuvirtide targets a fusion intermediate conformation of Env [27–29] and has further been reported to interact with gp120 to inhibit CXCR4 binding [30]. Administration of enfuvirtide using a needle-free injection device is currently in development [31]. Other peptide-based and small-molecule inhibitors of fusion that aim at gp41 [32] are currently in clinical development. Small-molecule inhibitors hold the most potential for widespread use due to their decreased production costs and oral availability.

***Integrase Inhibitors*** Merging of the viral genome and host chromosomal DNA (cDNA) is the trademark of retroviral infection. This step of retroviral infection is performed by the integrase enzyme and is labeled integration. Like reverse transcriptase, integrase is a promising therapeutic target due to its distinctive enzymatic activity. However, unlike reverse transcriptase, this target continues to be clinically unused despite extensive drug development [33, 34]. Two distinctive targets exist for disrupting integration: (1) inhibiting the action of the integrase enzyme directly through conventional small-molecule inhibitors; or (2) designing agents that inhibit the formation of the preintegration complex (PIC), which occurs immediately after the uncoating of the viral capsid. Blocking the PIC may inhibit nuclear entry and/or significantly decrease the activity of integrase. Clinical trials are underway with small-molecule inhibitors through Merck Research Laboratories and Gilead Sciences [21]. Alongside reverse transcriptase, protease and entry inhibitors, integrase inhibitors denote the next class of antiretroviral agents most likely to be used for the treatment of HIV-1 infection.

***Inhibitors of Maturation*** Protease inhibitors are the essential component of HAART, and they prevent the necessary proteolytic processing of the gag and gag-pol polyproteins. The proteolytic processing allows the structural maturation of the virus particle and the activation of the viral enzymes along with viral budding. Resistance to protease inhibitors (PIs) occurs because of continued viral replication while the drug is present. PA 457 is an alternative drug that inhibits viral maturation at the substrate of the protease, unlike conventional PIs that act at the enzyme's active site. This alternative drug that targets viral maturation has retained its potency against isolates resistant to approved reverse transcriptase and PIs [35], and has been granted fast-track review status by the U.S. FDA while it is currently in phase II clinical trials.

## 3 ANTIVIRALS FOR HEPATITIS B AND C INFECTIONS

Individuals that are chronically infected with hepatitis B (HBV) and C (HCV) are at risk of developing cirrhosis and, subsequently, hepatocellular carcinoma. Antiviral therapy offers the only course of interrupting this progression. Until now, recombinant interferon α (IFN-α) and the nucleoside/nucleotide analogs lamivudine, adefovir dipivoxil, entecavir, and telbuvidine are the sole licensed drugs for the treatment of chronic hepatitis B. For hepatitis C, recombinant or PEGylated interferons (PEGIFN) in combination with ribavirin are used for chronic hepatitis C virus infections. Improvements in current therapies are needed for both hepatitis B and C due to viral resistance, effectiveness, significant cost, and adverse effects [36, 37].

### 3.1 Antivirals for Hepatitis B

Patients with chronic HBV infection who can benefit from therapeutic intervention are listed in Table 4 along with current practical recommendations. In surveillance of viral hepatitis serology in individuals with elevated alanine aminotransferase (ALT) or in screening for HBV in those that have risk factors, the finding of hepatitis B surface antigen (HBsAg) results in the diagnosis. This subsequently leads to the testing for hepatitis B e antigen (HBeAg), hepatitis B e antibody, and HBV DNA. Detection of hepatitis B surface antibody indicates immunity via vaccination or past infection. The goals of treatment of chronic hepatitis B infection are the initiation and maintenance of HBV suppression resulting in improved clinical outcome [38]. Loss of HBsAg along with or without seroconversion is the gold standard for resolving the HBV infection. Two potential mechanisms exist for the treatment of chronic HBV, an immune modulating approach and a completely antiviral approach.

***Interferons*** Interferons have immunomodulatory, antiproliferative, and antiviral effects; and furthermore they have been in use for years in the treatment of chronic HBV infection. The two IFN subgroup entities that have been most frequently used are IFN-α 2a and IFN-α 2b, and both are now present in the PEGylated form by conjugation to bis-monomethoxy polyethylene glycol (PEG-IFN). PEG IFN's efficacy

**TABLE 4 Practical Recommendations for Therapeutic Intervention of Chronic Hepatitis B**

| Clinical Findings | Therapeutic Intervention |
|---|---|
| HBsAg carrier | Observation[a] |
| Hbe Ag-positive or -negative (hepatitis B low HBV DNA ($\leq 10^4$ IU/mL) | PEG IFN[b] or adefovir[c] or entecavir[d] or lamivudine[e] or telbivudine[f] |
| Normal ALT | |
| HBe Ag positive | |
| Moderate to high HBV DNA ($\geq 10^5$ IU/mL) | Lamivudine or adefovir or entecavir or telbivudine or PEG IFN |
| Abnormal ALT (≥2× normal level) | |
| HBeAg negative | |
| Moderate to high HBV DNA ($\geq 10^4$ IU/mL) | Lamivudine or adefovir or entecavir or telbivudine |
| Abnormal ALT (≥2× normal level) | |
| HbeAg positive or negative | |
| HBeAg positive or negative | Lamivudine or adefovir or Entecavir or telbivudine |
| Cirrhosis, decompensated | |

[a]If immunosuppression planned, then recommended treatment is lamivudine.
[b]Five million units per day for a minimum of 16 months, given intramuscularly or subcutaneously [39].
[c]10 mg per day for 48 weeks [39].
[d]0.5 mg per day for 52 weeks (lamivudine-resistant patients are given a 1.0-mg daily dose for the same course) [40].
[e]Available in tablet or oral solution, and tablet form is 100 mg/day.
[f]600 mg of telbivudine daily for 52 weeks [41].

against HBV is currently under assessment, although it is already licensed for the treatment of chronic HCV infection [36].

***Lamivudine*** Lamivudine, a deoxycytidine analog, is an antiretroviral nucleoside reverse transcriptase inhibitor that results in premature chain termination subsequent to incorporating into newly synthesized HBV DNA. It is used in patients with positive and negative HBeAg chronic hepatitis and in place of IFN in those with decompensated cirrhosis and active liver inflammation. Drug-resistant viruses have been observed with extended treatment.

***Adefovir Dipivoxil*** Adefovir dipivoxil, an adenine nucleotide analog, is an antiretroviral nucleotide reverse transcriptase inhibitor that is used for the treatment of both HBeAg–positive and -negative individuals, including those individuals that are lamivudine resistant. The goal of therapy with adefovir is to slow the progression of HBV and improve liver inflammation and fibrosis [39].

***Entecavir*** Entecavir (ETV), a novel deoxyguanosine analog, differs from the other nucleoside/tide reverse transcriptase inhibitors in that it is >100-fold more potent against HBV in culture and displays no significant activity against HIV. Additionally, while lamivudine and adefovir are obligate DNA chain terminators, ETV halts HBV DNA elongation after incorporating a few additional bases. It is indicated for use in people infected with lamivudine-resistant HBV as well as patients suffering from persistently elevated levels of ALT/AST (asparte aminotransferase) [40].

***Telbivudine*** Telbivudine, the newest drug to enter the market against chronic HBV, is most unique in its mechanism of action in that it targets the DNA replication of the virus. It inhibits the HBV DNA polymerase's reverse transcriptase and results in subsequent inhibition of viral replication, improvement of liver inflammation, and suppression of the viral infection [41].

***Combination Therapy*** Due to restricted efficacy and drug resistance, combination therapy with two or more nucleos(t)ide analogs, or with IFN-α, is an alternative choice that might demonstrate higher sustained response rates.

***Therapeutic Investigational Drugs*** Several nucleos(t)ide analogs are currently in phase II/III clinical trials for the treatment of HBV. Among them are tenofovir, emtricitabine, and clevudine. The use of any of these analogs in combination therapy remains to be established [42, 43].

### 3.2 Antivirals against Hepatitis C

In the acute phase of infection with HCV, the virus escapes detection by the immune system in the vast number of the cases and subsequently leads to persistent infection in between 50 and 90% of cases [44]. Treatment is typically considered during the chronic phase of infection, which is defined as the presence of HCV RNA in the peripheral blood (detected by polymerase chain reaction, PCR) during a minimum 6-month period. The presence of elevated ALT also signifies chronic disease; however, 20–50% of patients have persistently normal aminotransferase levels and should not be excluded from therapy. The main goal of antiviral therapy is to prevent progression of liver fibrosis to cirrhosis, end-stage liver disease, and hepatocellular carcinoma (HCC).

***Interferons*** IFN-α has been the mainstay of antiviral therapy in HCV infection, and it inhibits viral replication indirectly through activation of natural killer (NK) cells and cytotoxic T cells and by way of altered cytokine synthesis [45]. IFN also directly inhibits viral RNA and

protein synthesis. In 2002, the FDA approved the use of PEG-IFNs as monotherapy for HCV patients that have not had prior treatment with IFN. Combining PEG to the IFN molecule has made significant advancement in antiviral therapy against chronic HCV. It has allowed an increase in absorption time and half-life and has decreased the clearance of the drug [46]. As a consequence of these events, a longer duration of action on the patient can occur while preventing viral replication between doses.

***Ribavirin*** Ribavirin, a nucleoside analog, is ineffective at reducing viral load as a monotherapy but has improved effectiveness when used in combination with IFN-α [47, 48]. The exact mechanism of action is unclear, but it is known that it enhances T-cell-mediated immunity against viral infection, suppresses viral RNA replication, directly inhibits HCV-encoded RNA-dependant RNA polymerase, and reduces the production of genomic viral RNA by way of increasing the mutation rate of RNA viruses [49].

***Combination Therapy of IFN and Ribavirin*** The HCV treatment has developed from the use of IFN alone to the combination of IFN plus ribavirin since IFN monotherapy has not yielded satisfactory results. PEG-IFN-α monotherapy is more effective than IFN-α monotherapy but less than the combination of IFN-α and ribavirin [50–54]. The current treatment of PEG-IFN in adjunct with ribavirin is now the best therapy for the treatment of chronic HCV [55]. This therapy is dependant upon the HCV genotype. Usually, genotype 1 requires 12 months of therapy with PEG-IFN once a week subcutaneously in combination with 1000 mg ribavirin. In patients with genotypes 2 and 3, 6 months of therapy with PEG-IFN and 800 mg ribavirin is adequate. Despite improvements in current therapy, less satisfactory results have been attained in particular patient populations (e.g., African-American ethnicity, older age, and possibly obesity) [56]. Furthermore, viral response to antiviral therapy fluctuates between genotypes, and ribavirin is associated with the frequent side effect of hemolytic anemia that subsequently limits dosing. In addition, the number of nonresponders to current therapy will continue to rise as more patients with chronic HCV are treated. The rising number of nonresponders emphasizes the necessity for the development of more efficacious antivirals against HCV.

***Therapeutic Investigational Drugs*** Emerging therapeutic agents are currently under clinical investigation (Table 5) include novel IFNs and IFN delivery systems, ribavirin analogs, antisense oligonucleotides, short interfering RNAs (siRNAs) and ribozymes, enzyme inhibitors, immunomodulatory agents, antifibrotic agents, therapeutic vaccine, and antibodies.

## 4 ANTIVIRALS FOR HUMAN HERPESVIRUS INFECTIONS

Human herpesviruses (HHV) include eight types of viruses that can produce a variety of illnesses. Among the HHV family are the herpes simplex virus type 1 and type 2 (HSV-1 and HSV-2), which the majority of the time result in herpes labialis (fever blisters/cold sores) and genital herpes, respectively. Both herpes labialis and genital herpes can be caused by either virus. Other disease entities caused by HSV-1 and HSV-2 include gingivostomatitis, ocular disease, herpes gladiatorum, eczema herpeticum, herpetic whitlow, neonatal herpes, lumbosacral herpes, herpetic keratoconjunctivitis, herpes encephalitits, cervicitis, and erythema multiforme [57]. HHV type 3 (varicella zoster virus, VZV) results in chickenpox in its primary form and herpes zoster or shingles in its recurrent form. The rest of the types of the HHV family and ensuing conditions are listed in Table 6. The three main antiviral agents that are effective against HSV-1, HSV-2, and VZV are acyclovir, valacyclovir, and famciclovir. The dosages of these antiviral agents against HSV-1, -2 and VZV, for both immunocompetent and immunocompromised patients, are listed in Table 7.

***Acyclovir*** Acyclovir (ACV), a guanosine analog, is the most familiar and widely used antiviral in the world [58]. ACV, once phosphorylated by herpes-specific thymidine kinase (TK), inhibits viral DNA polymerase by serving as an obligate chain terminator [59]. Indications for the use of ACV are found in Table 8. ACV can be given topically, orally, and intravenously. The oral form of administration is the most widely used for HSV infections. Intravenous ACV is utilized for severe illness and in the immunocompromised. Indications consist of disseminated HSV disease, complicated primary infection, neonatal infection, eczema herpeticum, herpes encephalitis, and HSV that is not treated successfully with oral therapy. Intravenous ACV is administered to individuals that are immunocompromised due its greater bioavailability and has been the mainstay of therapy in this patient population afflicted with primary or recurrent VZV infection [60].

***Valacyclovir*** Valacyclovir (VACV) is an oral prodrug of acyclovir and has an oral bioavailability three to five times greater than that of oral ACV, which is its main difference from ACV. When VACV is administered

**TABLE 5 Antiviral Agents for HCV in Current Clinical Investigation**

| Category | Name | Notes |
|---|---|---|
| Interferon | Omega Interferon | Genetically engineered IFN, oral prodrug formulation in development |
| | Multiferon | Natural IFN |
| | Albuferon | IFN fused to albumin molecules |
| | ANA 245 | Nucleoside analog that induces multiple cytokines, including IFN-α |
| | Resiquimod | Induced number of cytokines including IFN-α |
| Ribavirin analogs | Levovirin | L-isomer of ribavirin (linked with lesser degree of hemolytic anemia) |
| | Viramidine | Prodrug that is immediately converted into ribavirin in vivo |
| Antisense oligonucleotides | ISIS-14803 | Genetically inhibits production of disease-causing proteins |
| Ribozymes | Heptazyme | Cleaves RNA interrupting life cycle of virus[a] |
| Enzyme inhibitors | BILN 2061 | Serine protease inhibitor[a] |
| | VX-950 | Protease inhibitor |
| | NM-283 | RpRd polymerase inhibitor (oral nucleoside analog) |
| | JTK 003 | RpRd polymerase inhibitor, a nonnucleoside |
| | Merimebodib (VX-497) | IMPDH inhibitor[b] |
| | Mycophenolate mofetil | IMPDH inhibitor[b] |
| Immunomodulatory agents | Histamine dihydrochloride | Exhibits immunomodulatory effects |
| | Thymosin alfa-I | Stimulates immune system |
| | Amantadine | Current antiflu therapy |
| Antifibrotic agents | Interferon-γ | Result in reversal of liver fibrosis (by histological scoring system) |
| | Interleukin-10 | Reduces fibrosis |
| | Interleukin-12 | Minimal efficacy with significant toxicity |
| Therapeutic vaccine | Envelope protein 1 | Improvement in liver fibrosis |
| | Peptide IC41 | Induces IFN-γ secreting cytotoxic and helper T cells |
| Antibodies | Monoclonal antibodies | Reduces HCV RNA |
| | Immunoglobulin | Prevents HCV disease in HCV positive liver transplant patients |

[a]Trials stopped due to animal toxicity.
[b]Inosine monophosphate dehydrogenase.

**TABLE 6 Clinical Regimen for HHV Infections**

| Clinical Scenario | Acyclovir | Valacyclovir | Famciclovir |
|---|---|---|---|
| | *Immunocompetent Patients* | | |
| Herpes simplex—primary | 400 mg t.i.d. × 10 days | 1 g b.i.d. × 10 days | 250 mg t.i.d. × 10 days |
| Herpes simplex—recurrence | 400 mg t.i.d. × 5 days | 500 mg b.i.d. × 3 days | 1 g b.i.d. |
| Herpes simplex—suppression | 400 mg b.i.d. | 500 mg q.d.[a] | 250 mg b.i.d. |
| Herpes zoster—acute treatment | 800 mg 5×/day × 7–10 days | 1 g t.i.d. × 7–10 days | 500 mg t.i.d. × 7 days |
| Primary varicella—children | 20 mg/kg q.i.d. up to 800 mg/dose × 5–7 days | No exact recommendations | No exact recommendations |
| | *Immuncompromised Patients* | | |
| Herpes simplex—primary | 200–400 mg 5 ×/day × 10 days or 5 mg/kg i.v. q8h for 7–10 days | No studies reported | No studies reported |
| Herpes simplex—recurrent | Minimum of 400 mg t.i.d. × 7–10 days | 500 mg b.i.d. × 7 days | 500 mg b.i.d. × 7 days |
| Herpes simplex—suppression | Minimum 400 mg b.i.d. | 500 mg to 1 g q.d. | 500 mg b.i.d. |
| Herpes zoster—adults | 800 mg 5×/day × 7–10 days | 1 g t.i.d. × 7–10 days | 500 mg t.i.d. × 7–10 days |
| Primary varicella—children | 10 mg/kg i.v. q8h × 7–10 days | No studies reported | No studies reported |

[a]Recommended VACV 1000 mg daily suppressive dose (single dose) if 10 or more HSV recurrences per year.

**TABLE 7 HHV and Associated Diseases**

| HHV Type | Disease Entity |
|---|---|
| HHV-1 (HSV-1) | Herpes labialis, etc. |
| HHV-2 (HSV-2) | Genital herpes, etc. |
| HHV-3 (VZV) | Chicken pox, herpes zoster, etc. |
| HHV-4 (Epstein–Barr virus, EBV) | Mononucleosis, Burkitt's lymphoma |
| HHV-5 (Cytomegalovirus, CMV) | CMV retinitis |
| HHV-6 | Roseola infantum, etc. |
| HHV-7 | Pityriasis rosea, etc[a] |
| HHV-8 (Kaposi's sarcoma herpes virus, KSHV) | Kaposi's sarcoma (classic and epidemic) |

[a]The exact role of HHV-7 in pityriasis rosea has not been established.

**TABLE 8 ACV Indications in the Management of HHV Infections[a]**

| |
|---|
| *FDA-Approved Indications* |
| Herpes simplex infection [57, 59, 60, 137–143] |
| Primary episode |
| Recurrent episode |
| Suppressive therapy |
| Varicella–zoster infections |
| Chickenpox [57, 144–146] |
| Herpes zoster [57, 147] |
| Herpes simplex or varicella–zoster infections |
| Immunocompromised patients (e.g., HIV-infected patients) |
| *Other Uses* |
| 1° gingivostomatitis, recurrent herpes labialis, herpes gladiatorum, eczema herpeticum, herpetic whitlow, herpetic keratoconjunctivitis |
| Recurrent erythema multiforme (due to HSV [148–150]) |

[a]Pregnancy category B.

**TABLE 9 Valacyclovir Indications in the Management of HHV Infections[a]**

| |
|---|
| *FDA-Approved Indications* |
| Herpes simplex infection [57, 151–159] |
| Primary episode |
| Recurrent episode |
| Suppressive therapy |
| Varicella–zoster infections |
| Herpes zoster |
| Herpes simplex or varicella–zoster infections |
| Immunocompromised patients (e.g., HIV-infected patients) |
| *Other Uses* |
| 1° ginvivostomatitis, recurrent herpes labialis, herpes gladiatorum, eczema herpeticum, herpetic whitlow, herpetic keratoconjunctivitis |
| Recurrent erythema multiforme [159] |

[a]Pregnancy category B.

(orally), it is practically as potent as intravenous ACV [57]. Indications for its use are found in Table 9.

***Famciclovir*** Famciclovir (FCV) is the oral prodrug of penciclovir (PCV), an acyclic nucleoside. Similar to ACV, it is necessary for PCV to be phosphorylated for it to be active. Phosphorylated PCV has a much longer intracellular half-life (10–20 h in HSV-infected cells and 7 h in VZV-infected cells) than phosphorylated ACV (less than 1 h in HSV- and VZV-infected cells) [59]. Oral FCV has a 77% bioavailability in comparison to 15–30% biovailability of oral ACV and 55% for oral VACV [61, 62]. FCV is an effective oral agent that can be used as an alternative to intravenous antiviral therapy in immunocompromised individuals afflicted with herpes zoster. However, it is necessary to monitor patients very closely and administer intravenous ACV if any signs or symptoms of disseminated disease occur [63]. Indications for the use of FCV are listed in Table 10.

***Therapeutic Guidelines*** Overall, ACV, VACV, and FCV are equivalent in safety and efficacy. In deciding therapy against HSV infection, VACV and FCV provide the convenience of lower dosing frequency, and thus enhancing patient compliance; however, VACV and FCV are more costly than ACV. In the therapy of VZV, VACV and FCV are, to some extent, more effective in lowering the duration of zoster-associated pain than ACV due their superior bioavailability [64, 65].

Resistance to ACV can arise from uncontrolled replication of HSV, particularly in immunocompromised patients. ACV-resistant HSV is often cross resistant to VACV and FCV as well. Resistance to ACV regularly

**TABLE 10 Famciclovir Indications in the Management of HHV Infections[a]**

| |
|---|
| *FDA-Approved Indications* |
| Herpes simplex infections [57, 160–165] |
| Primary episode |
| Recurrent episodes |
| Suppressive therapy |
| Immunocompromised patients (e.g., HIV-infected patients) [165–167] |
| Varicella–zoster infections |
| Herpes zoster [64, 65, 168] |
| Immunocompromised patients (e.g., HIV-infected patients) [63] |
| *Other Uses* |
| 1° varicella |
| 1° gingivostomatitis, recurrent herpes labialis, herpes gladiatorum, eczema herpeticum, herpetic whitlow, herpetic keratoconjunctivitis |

[a]Pregnancy category B.

results from mutations in the viral thymidine kinase gene and less often with mutations in viral DNA polymerase [66, 67]. Patients that are not successfully treated with standard therapy with oral and intravenous acyclovir can be treated with foscarnet or cidofovir. Foscarnet and cidofovir are more toxic than ACV and thus are not administered as first-line therapy for HSV infections. Cidofovir can be compounded for topical use and neither drug requires phosphorylation by thymidine kinase to inhibit viral DNA polymerase.

***Therapeutic Investigational Drugs*** Emerging therapies against HHV infections include the newly discovered bicyclic furo(2,3-d) pyrimidine nucleoside analogs (BCNAs). BCNAs are extremely selective in their anti-VZV activity; more specifically, they only inhibit VZV replication and not the replication of any other viruses such as HSV [68, 69]. Resiquimod, an investigational immune response modifier and Toll-like receptor (TLR) 7 and 8 agonist, has been shown to reduce the frequency of HSV-2 mucosal reactivation [70, 71]. Novel anti-HSV agents that target the viral helicase-primase complex, such as thiazolylphenyl derivatives BILS 179BS and, particularly, BAY 57-1293 have been reported to have in vivo efficacy in animal models of HSV-1 and HSV-2 infections [72, 73]. The helicase-primase complex is composed of three viral proteins: HSV UL5, UL-8, and UL52 gene products. All three unwind the double-stranded viral DNA and produce the primer/templates for DNA synthesis by the viral DNA polymerase. It has been stated that BAY-1293, in its function as helicase-primase inhibitor, is superior to all other antiviral agents currently used for HSV infections [74]. This supports the continued pursuit of helicase-primase inhibitors against HSV. An extension of this approach involves compounds that interact with the formation of the HSV DNA polymerase complex (e.g., BP5) [75]. Targeting protein-to-protein interactions comprise a novel target for the generation of antiviral agents for continuing exploration as antiviral therapy.

## 5 ANTIVIRALS FOR HUMAN PAPILLOMAVIRUS (HPV)

HPV-associated diseases in humans have had an augmented significance over the years and are currently a foremost concern for public health. The majority of HPVs that cause significant diseases in humans are grouped into the genus α papillomavirus (all genital papillomaviruses, PVs), β papillomavirus (e.g., PVs that cause epidermodysplasia verruciformis), and γ PVs (most of the viruses responsible for cutaneous lesions) [76, 77]. The importance of HPV has been escalated through the definitive association with cervical cancer, which is the second most frequently occurring cancer in women worldwide. HPV infections are frequent, but only a small percentage of infected individuals will retain the virus and become chronic or persistent carriers, and thus leading the path for the virus to stimulate epithelial cell proliferation, whether malignant or not. Subsequent to the virus entering the cell and viral proteins being synthesized, there is an exact, yet intricate, control of the cell machinery by the virus. HPV's oncoproteins, E6 and E7, are able to interfere with cell regulatory proteins and the cascade controlled by them; further, the viral proteins E1 and E2 initiate viral DNA replication. Each step of the involved interaction between the virus and its host cell is a possible target for the development of therapeutics. In addition to surgery, which is the basic approach for the treatment of most frequently observed HPV lesions, adjuvant therapies have been proposed and tested to complete or replace surgery, when needed [78]. Therapeutic agents that interfere with viral replication of HPV will be further discussed.

***Imiquimod*** Imiquimod is a member of new class of immune response modifiers, the imidazoquinolines, and it enhances both the innate and acquired immune pathways (mainly through the T-helper cell type 1-mediated immune responses) resulting in antiviral, antitumor, and immunoregulatory activities. It was first approved in 1997 for the topical treatment of external genital and perianal warts. The mechanism of action of imiquimod involves cytokine induction in the skin, which subsequently activates the host's immune system to recognize the presence of viral infection (or tumor), eventually to eliminate the associated lesion. It is postulated that imiquimod specifically stimulates a cell-mediated immunity response against HPV, and the fairly long time to clinical efficacy by imiquimod supports this [79, 80].

Indication for licensure of imiquimod came from a number of recently published prospective studies, which were double-blind, randomized, vehicle-controlled trials with different dosing regimens (1 and 5%: daily or 3 times per week for 16 weeks or until lesions cleared). Total elimination of warts was seen in 37–52% of those treated with 5% cream, 14–21% with 1% cream, and 0–11% with vehicle-control by intent-to-treat analyses. Other recommended treatment methods for genital warts are nonspecific and mostly physical or cytotoxic ablative methods, and thus latent HPV can stay behind in the adjancent tissue. There is a relatively high rate of recurrence (20–70%) in these other methods in comparison to imiquimod (0–19%) [81–83]. Furthermore, initial studies show encouraging results for the use of imiquimod for the management of HPV-related intraepithelial dysplasias (bowenoid papulosis/vulvar

intraepithelial neoplasia) in both immunocompetent and immunocompromised patients [84, 85].

***Cidofovir*** Cidofovir is a nucleotide analog with broad spectrum antiviral activity against DNA viruses. Unlike other DNA viruses, HPV does not encode for its own DNA polymerase, the final target of cidofovir. Thus, it is not intuitive to use cidofovir for the treatment of HPV infection. Preliminary results of cidofovir in animal models indicated that cidofovir was able to provide some therapeutic results for the treatment of PV-induced lesions [86]. The first documented patient ever treated with cidofovir, in 1995, had hypopharyngeal/esophageal HPV lesions induced by types 16 and 18, which had failed laser treatment. The lesions regressed gradually and eventually disappeared with repeated cidofovir treatments [87]. Subsequent patients, with recurrent respiratory papillomatosis (RRP), have been treated with cidofovir with remarkable results [88–93]. Cidofovir has also been used for the local treatment of external HPV lesions in both immunocompetent and immunocompromised patients. Reported uses are limited series or case reports [94–102]. One placebo-controlled trial study revealed cidofovir superior to the placebo control in the treatment of condyloma acuminata in immunocompetent individuals [101]. Another study reported that cidofovir was able to partially or completely inhibit cervical dysplasia lesions only after three applications [103]. Other applications of cidofovir have included bowenoid papulosis, recalcitrant warts (including plantar warts), and other mucocutaneous HPV lesions that do not respond to conventional therapies [104]. Although further studies are needed to elucidate the exact mechanism of action of cidofovir in HPV-transformed cells, it has been demonstrated that cidofovir has an antiproliferative activity and it stimulates apoptosis in HPV-positive cell lines in vitro [105–109]. Topical cidofovir is not commercially available; however, it can be compounded at a local specialty pharmacy.

***Acyclovir and Ribavirin*** Both acyclovir and ribavirin have been used systemically (acyclovir) and injected locally (ribavirin) as therapeutic agents of severe RRP. Only limited reports exist and no control studies have been planned to date [89].

***Therapeutic Investigational Drugs*** In addition to cidofovir, there are several other acyclic nucleoside phosphonates, for instance, 9-(2-phosphonylmethoxyethyl) guanine (PMEG) and 9-(2-phosphonylmethoxyethyl)-$N^6$-cyclopropyl-2,6-diaminopurine (cPr PMEDAP), which, similar to cidofovir, are capable of specifically inducing apoptosis in HPV-infected cells. PMEG and cPr PMEDAP are being studied for the possibility of using them as therapeutic agents of HPV-associated papillomavirus lesions [105]. Molecules targeting HPV transcripts, including ribozymes, antisense oligonucleotides [110], and RNA interference-mediated genes are under investigation for their ability to inhibit the expression of HPV E6 or E7 oncogenes [111, 112]. Several therapeutic vaccines targeting the treatment of precancerous lesions and invasive carcinoma of the cervix, anus, and vulva, and juvenile respiratory papillomatosis are in different stages of development [113]. The rationale behind the vaccine is to target the oncoproteins since the continued expression of E6 and E7 is necessary to avoid apoptosis and cell cycle arrest and thus to sustain cells in a transformed state [78]. Results have been hopeful and at the same time discouraging [114]. A possible cause of the disappointing results could be that patients enrolled in the study were at an advanced stage of cervical disease, which is identified as being linked with genetic instability, viral immune escape, viral antigen tolerance, and downregulation of major histocompatability classes (MHC) class I alleles [114].

## 6 ANTIVIRALS FOR POXVIRUS INFECTIONS

The family of poxviruses includes the orthopoxviruses (e.g., variola, vaccinia, monkeypox, and camelpox), parapoxviruses (e.g., orf), and molluscivirus (e.g., molluscum contagiosum virus). Smallpox and molluscum contagiosum infect humans, and, although smallpox was eradicated in 1980, there are concerns now over bioterrorism [115, 116]. Furthermore, monkeypox, a rare zoonosis with a more recent outbreak in the midwestern United States, demonstrates the continuing need for development of effective therapies against poxviruses [117]. Antiviral drugs as a therapeutic regimen for poxviruses offer a combination of chemical stability and ease of delivery that is enticing from a public health viewpoint.

***Cidofovir*** Cidofovir, an acyclic nucleoside phosphonate, is currently licensed against cytomegalovirus (CMV) retinitis and is believed to work by inhibiting the cytomegalovirus DNA polymerase, a target common to both CMV and poxviruses. In addition, cidofovir has antiviral activity against poxviruses in vitro and against cowpox and vaccinia viruses in mice [118–120]. It is also thought that cidofovir inhibits the exonuclease activity of proofreading, and subsequently generating error-prone DNA synthesis during poxvirus replication [121]. This agent was used successfully in the treatment of molluscum contagiosum in AIDS patients, where it was administered topically or intravenously (at 2 or 5 mg/kg

every 2 weeks). This resulted in regression of lesions within 2–4 months of therapy [122, 123]. Cidofovir has also proved to be effective in the topical treatment of orf (ecthyma contagiosum or sheep pox) with complete resolution of the lesions subsequent to seven courses (2–3 months) [124].

***Therapeutic Investigational Drugs*** Several nucleoside analogs (e.g., S2242, 8-methyladenosine, and idoxuridine) and nucleotide analogs (e.g., HPMPDAP, (S)-HPMPO-DAPy, and adefovir) are effective in various animal models of poxvirus infections [125, 126]. Since smallpox has been eradicated and infections with monkeypox are very rare, it is difficult to study these selected agents. Thus, it is important to study these promising drugs in animal models and in related poxvirus infections (e.g., molluscum contagiosum or ecthyma) [125].

## 7 ANTIVIRALS AGAINST INFLUENZA VIRUS INFECTIONS

There are three influenza viruses: A, B, and C; only A is further subdivided and five of those subtypes are known to affect humans. The persistent epidemics of influenza A strains $H_3N_2$ and $H_1N_1$ and the possibility of a pandemic with avian influenza A strain $H_5N_1$ along with the fact that there is no existing vaccination for the latter has given much importance on treatment (and prophylaxis) of influenza virus infections. Four approved drugs from two drug classes are available as therapeutic and prophylactic agents in the United States. They include the adamantanes, amantadine and rimantadine, and the neuraminidase (NA) inhibitors, zanamivir and oseltamivir [127–129].

***Adamantanes*** Amantadine and rimantadine are matrix (M2) hydrogen ion blockers that have been used for many years. The target of the adamantanes, the M2 ion channel, is specific to influenza A viruses, and thus the adamantanes are active only against influenza A. The adamantanes have proven prophylactic efficacy in prior influenza pandemics and are more cost effective if the virus is susceptible. However, in the past there have been concerns of emergence of drug-resistant virus mutants, and thus the potential usefulness of adamantanes in future pandemics is unknown. Furthermore, amantadine, in particular, has adverse side effects on the central nervous system [127, 129].

***Neuraminidase Inhibitors*** NA inhibitors are active for the therapy and/or prophylaxis of both influenza A and B. Influenza virus has a distinct replication strategy where it uses one of its surface glycoproteins, hemagglutinin (H), to attach to the target cell receptor, which contains a terminal sialic acid (*N*-acetylneuraminic acid), and another surface glycoprotein, neuraminidase, to cleave off the terminal sialic acid of the host cell receptor, permitting the virus particles to leave the cells after viral replication has completed [130, 131]. NA inhibitors prevent the release of the newly formed virus particles, subsequently impeding the spread to further host cells. Advantages offered by NA inhibitors are considerable. In therapeutic terms, they lessen the illness in 1–2 days, reduce virus transmission to household or health care contacts, and lessen the frequency and severity of complications (e.g., sinusitis and bronchitis). When used prophylactically, they substantially lessen the number of new cases of influenza. Resistance to NA inhibitors can occur [132], but there is no direct evidence thus far [133]. Oseltamivir, which is conveniently administered as oral capsules, should be stockpiled to affront a possible influenza A epidemic in the future [127].

***Therapeutic Investigational Drugs*** Investigational antiviral agents include new NA inhibitors (peramivir and A-315675), long-acting topical NA inhibitors, conjugated sialidase, hemagglutinin inhibitors, small interfering RNA, polymerase inhibitors, and protease inhibitors. Topically applied long-acting NA inhibitors that exert activity for a longer duration in animal models are also under investigation. Furthermore, early in vitro studies assessing combination regimens demonstrated that ribavirin in addition to an M2 ion channel inhibitor had superior action against influenza A [134]. Subsequent combination therapy trials done, notably with rimatidine and subsequent zanamivir treatment, showed less adamantine resistance in the cohort receiving both agents [135]. However, the combination regimen of an NA inhibitor and adamantine would not be effective against viruses that are already adamantine resistant. Additional studies are needed to assess combination regimens, including those investigating dual NA inhibitors, and transcriptase inhibitors (e.g., ribavirin) combined with NA inhibitors. In addition to the already approved use of ribavirin in combination therapy in HCV infections, it is also indicated as small-droplet aerosol treatment of respiratory syncytial infections (RSV) in infants.

## 8 CONCLUSION

The past 15 years have seen significant progress in developing new antivirals. In 1990, there were just five drugs licensed as antiviral agents, whereas now there are approximately 40 drugs in the antiviral market

repertoire, most being for the treatment of HIV. Nonetheless, combination therapy has only augmented life expectancy for individuals with advanced AIDS from 1.53 to 2.91 years [136]. Evidently, novel therapies with different mechanisms of action and distinct resistance profiles continue to be greatly needed. In addition to HIV/AIDS, there is also an essential need for novel agents against other viral pathogens since treatment options tend to be more limited. In the investigational agents reviewed in each section, the most success is likely to be obtained with combination regimens, especially in the treatment of chronic viral infections, such as HIV, HBV and HCV [131]. Over the years, one of the main attributes sought in a new antiviral is the capacity to treat patients who have developed resistance to existing treatments, and numerous compounds are investigated in such advanced-stage patients. The main need for antiviral agents, past, present, and likely in the future continues to be in the world's poorest countries [136].

## REFERENCES


1. De Clercq E (2004). Antiviral drugs in current clinical use. *J Clin Virol 30*(2), 115–133.
2. WHO Website. http://www.who.int/en/. Accessed August 7, 2011.
3. Goldschmidt RH, Moy A (1996). Antiretroviral drug treatment for HIV/AIDS. *Am Fam Phys 54*(2), 574–580, 87–88.
4. Dando TM, Wagstaff AJ (2004). Emtricitabine/tenofovir disoproxil fumarate. *Drugs 64*(18), 2075–2082; discussion 83–84.
5. Squires K, Pozniak AL, Pierone G Jr, et al. (2003). Tenofovir disoproxil fumarate in nucleoside-resistant HIV-1 infection: a randomized trial. *Ann Intern Med 139*(5 Pt 1), 313–320.
6. Chang YC, Tyring SK (2004). Therapy of HIV infection. *Dermatol Ther 17*(6), 449–464.
7. Nunberg JH, Schleif WA, Boots EJ, et al. (1991). Viral resistance to human immunodeficiency virus type 1-specific pyridinone reverse transcriptase inhibitors. *J Virol 65*(9), 4887–4892.
8. Schmit JC, Weber B (1997). Recent advances in antiretroviral therapy and HIV infection monitoring. *Intervirology 40*(5–6), 304–321.
9. McDonald CK, Kuritzkes DR (1997). Human immunodeficiency virus type 1 protease inhibitors. *Arch Intern Med 157*(9), 951–959.
10. Weinstock HS, Zaidi I, Heneine W, et al. (2004). The epidemiology of antiretroviral drug resistance among drug-naive HIV-1-infected persons in 10 US cities. *J Infect Dis 189*(12), 2174–2180.
11. De Clercq E (2001). New developments in anti-HIV chemotherapy. *Curr Med Chem 8*(13), 1543–1572.
12. Rusconi S, Scozzafava A, Mastrolorenzo A, Supuran CT (2004). New advances in HIV entry inhibitors development. *Curr Drug Targets 4*(4), 339–355.
13. Doms RW, Moore JP (2000). HIV-1 membrane fusion: Targets of opportunity. *J Cell Biol 151*(2), F9–14.
14. Doms RW (2001). Chemokine receptors and HIV entry. *AIDS (Lond) 15*(Suppl 1), S34–35.
15. Berger EA, Murphy PM, Farber JM (1999). Chemokine receptors as HIV-1 coreceptors: roles in viral entry, tropism, and disease. *Annu Rev Immunol 17*, 657–700.
16. Reeves JD DR (2003). The role of chemokine receptors in HIV infection of host cells. In Bradshaw RA DE, Ed. *Handbook of Cell Signaling*. New York: Academic, pp. 191–196.
17. Moore JP, Shattock RJ (2003). Preventing HIV-1 sexual transmission–not sexy enough science, or no benefit to the bottom line? *J Antimicrob Chemother 52*(6), 890–892.
18. Shattock R, Solomon S (2004). Microbicides—Aids to safer sex. *Lancet 363*(9414), 1002–1003.
19. Callahan LN, Phelan M, Mallinson M, Norcross MA (1991). Dextran sulfate blocks antibody binding to the principal neutralizing domain of human immunodeficiency virus type 1 without interfering with gp120-CD4 interactions. *J Virol 65*(3), 1543–1550.
20. Moulard M, Lortat-Jacob H, Mondor I, et al. (2000). Selective interactions of polyanions with basic surfaces on human immunodeficiency virus type 1 gp120. *J Virol 74*(4), 1948–1960.
21. NIH AIDS Drug Information. http://aidsinfo.nih.gov/DrugsNew/Default.aspx?MenuItem=Drugs&Search=On. Accessed February 18, 2007.
22. Connor RI, Sheridan KE, Ceradini D, Choe S, Landau NR (1997). Change in coreceptor use correlates with disease progression in HIV-1—Infected individuals. *J Exper Med 185*(4), 621–628.
23. Moyle GJ, Wildfire A, Mandalia S, et al. (2005). Epidemiology and predictive factors for chemokine receptor use in HIV-1 infection. *J Infect Dis 191*(6), 866–872.
24. Alkhatib G, Combadiere C, Broder CC, et al. (1996). CC CKR5: a RANTES, MIP-1alpha, MIP-1beta receptor as a fusion cofactor for macrophage-tropic HIV-1. *Science 272*(5270), 1955–1958.
25. Bleul CC, Farzan M, Choe H, et al. (1996). The lymphocyte chemoattractant SDF-1 is a ligand for LESTR/fusin and blocks HIV-1 entry. *Nature 382*(6594), 829–833.
26. Oberlin E, Amara A, Bachelerie F, et al. (1996). The CXC chemokine SDF-1 is the ligand for LESTR/fusin and prevents infection by T-cell-line-adapted HIV-1. *Nature 382*(6594), 833–835.
27. Si Z, Madani N, Cox JM, et al. (2004). Small-molecule inhibitors of HIV-1 entry block receptor-induced conformational changes in the viral envelope glycoproteins. *Proc Nat Acad Sci USA 101*(14), 5036–5041.

28. Gallo SA, Puri A, Blumenthal R (2001). HIV-1 gp41 six-helix bundle formation occurs rapidly after the engagement of gp120 by CXCR4 in the HIV-1 Env-mediated fusion process. *Biochemistry 40*(41), 12231–12236.
29. He Y, Vassell R, Zaitseva M, et al. (2003). Peptides trap the human immunodeficiency virus type 1 envelope glycoprotein fusion intermediate at two sites. *J Virol 77*(3), 1666–1671.
30. Yuan W, Craig S, Si Z, Farzan M, Sodroski J (2004). CD4-induced T-20 binding to human immunodeficiency virus type 1 gp120 blocks interaction with the CXCR4 coreceptor. *J Virol 78*(10), 5448–5457.
31. Drug Monitor (2004). Needle-free administration of enfurvitide with Biojector 2000 (B2000) demonstrates pharmacokinetics bioequivalence to standard needle administration. In *HIV, DART: Frontiers in Drug Development of AntiRetroviral Therapies*. (Montego Bay).
32. Root MJ, Steger HK (2004). HIV-1 gp41 as a target for viral entry inhibition. *Curr Pharm Des 10*(15), 1805–1825.
33. De Clercq E (2004). HIV-chemotherapy and -prophylaxis: new drugs, leads and approaches. *Int J Biochem Cell Biol 36*(9), 1800–1822.
34. Condra JH, Miller MD, Hazuda DJ, Emini EA (2002). Potential new therapies for the treatment of HIV-1 infection. *Annu Rev Med 53*, 541–555.
35. Li F, Goila-Gaur R, Salzwedel K, et al. (2003). PA-457: A potent HIV inhibitor that disrupts core condensation by targeting a late step in Gag processing. *Proc Nat Acad Sci USA 100*(23), 13555–13560.
36. Bhopale GM, Nanda RK (2005). Emerging drugs for chronic hepatitis C. *Hepatol Res 32*(3), 146–153.
37. Mailliard ME, Gollan JL (2006). Emerging therapeutics for chronic hepatitis B. *Annu Rev Med 57*, 155–166.
38. Feld J, Lee JY, Locarnini S (2003). New targets and possible new therapeutic approaches in the chemotherapy of chronic hepatitis B. *Hepatology 38*(3), 545–553.
39. Tillmann HL (2007). Antiviral therapy and resistance with hepatitis B virus infection. *World J Gastroenterol 13*(1), 125–140.
40. Colonno RJ, Rose R, Baldick CJ, et al. (2006). Entecavir resistance is rare in nucleoside naive patients with hepatitis B. *Hepatology 44*(6), 1656–1665.
41. Lai CL, Lim SG, Brown NA, et al. (2004). A dose-finding study of once-daily oral telbivudine in HBeAg-positive patients with chronic hepatitis B virus infection. *Hepatology 40*(3), 719–726.
42. Marcellin P, Mommeja-Marin H, Sacks SL, et al. (2004). A phase II dose-escalating trial of clevudine in patients with chronic hepatitis B. *Hepatology 40*(1), 140–148.
43. van Bommel F, Wunsche T, Mauss S, et al. (2004). Comparison of adefovir and tenofovir in the treatment of lamivudine-resistant hepatitis B virus infection. *Hepatology 40*(6), 1421–1425.
44. Alter MJ, Kruszon-Moran D, Nainan OV, et al. (1999). The prevalence of hepatitis C virus infection in the United States, 1988 through 1994. *N Engl J Med 341*(8), 556–562.
45. Tilg H (1997). New insights into the mechanisms of interferon alfa: An immunoregulatory and anti-inflammatory cytokine. *Gastroenterology 112*(3), 1017–1021.
46. Harris JM, Martin NE, Modi M (2001). Pegylation: a novel process for modifying pharmacokinetics. *Clin Pharmacokinet 40*(7), 539–551.
47. Poynard T, Marcellin P, Lee SS, et al. (1998). Randomised trial of interferon alpha2b plus ribavirin for 48 weeks or for 24 weeks versus interferon alpha2b plus placebo for 48 weeks for treatment of chronic infection with hepatitis C virus. International Hepatitis Interventional Therapy Group (IHIT). *Lancet 352*(9138), 1426–1432.
48. McHutchison JG, Gordon SC, Schiff ER, et al. (1998). Interferon alfa-2b alone or in combination with ribavirin as initial treatment for chronic hepatitis C. Hepatitis Interventional Therapy Group. *N Engl J Med 339*(21), 1485–1492.
49. Graci JD, Cameron CE (2002). Quasispecies, error catastrophe, and the antiviral activity of ribavirin. *Virology 298*(2), 175–180.
50. Manns MP, McHutchison JG, Gordon SC, et al. (2001). Peginterferon alfa-2b plus ribavirin compared with interferon alfa-2b plus ribavirin for initial treatment of chronic hepatitis C: A randomised trial. *Lancet 358*(9286), 958–965.
51. Zeuzem S, Feinman SV, Rasenack J, et al. (2000). Peginterferon alfa-2a in patients with chronic hepatitis C. *N Engl J Med 343*(23), 1666–1672.
52. Heathcote EJ, Shiffman ML, Cooksley WG, et al. (2000). Peginterferon alfa-2a in patients with chronic hepatitis C and cirrhosis. *N Engl J Med 343*(23), 1673–1680.
53. Lindsay KL, Trepo C, Heintges T, et al. (2001). A randomized, double-blind trial comparing pegylated interferon alfa-2b to interferon alfa-2b as initial treatment for chronic hepatitis C. *Hepatology 34*(2), 395–403.
54. Fried MW, Shiffman ML, Reddy KR, et al. (2002). Peginterferon alfa-2a plus ribavirin for chronic hepatitis C virus infection. *N Engl J Med 347*(13), 975–982.
55. National Institutes of Health (2002). Consensus Development conference statement. Management of hepatitis C. *Hepatology 36*, 53–20.
56. Bruno R, Sacchi P, Maiocchi L, Zocchetti C, Filice G. (2004). Host factors affecting the outcome of treatment of hepatitis C. *Rev Gastroenterol Disord 4*(Suppl 1), S3–7.
57. Whitley R, Gnann, JW Jr (2002). Herpes simplex viruses. In Stephen K. Tyring, Ed. *Mucocutaneous Manifestations of Viral Diseases*. New York: Marcel Dekker, pp. 69–118.
58. Wagstaff AJ, Faulds D, Goa KL (1994). Aciclovir. A reappraisal of its antiviral activity, pharmacokinetic properties and therapeutic efficacy. *Drugs 47*(1), 153–205.
59. Herne K, Cirelli R, Lee P, Tyring SK (1996). Antiviral therapy of acute herpes zoster in older patients. *Drugs Aging 8*(2), 97–112.

60. Ioannidis JP, Collier AC, Cooper DA, et al. (1998). Clinical efficacy of high-dose acyclovir in patients with human immunodeficiency virus infection: A meta-analysis of randomized individual patient data. *J Infect Dis 178*(2), 349–359.
61. Tyring SK (1998). Advances in the treatment of herpesvirus infection: The role of famciclovir. *Clin Ther 20*(4), 661–670.
62. Pue MA, Pratt SK, Fairless AJ, et al. (1994). Linear pharmacokinetics of penciclovir following administration of single oral doses of famciclovir 125, 250, 500 and 750 mg to healthy volunteers. *J Antimicrob Chemother 33*(1), 119–127.
63. Tyring S, Belanger R, Bezwoda W, Ljungman P, Boon R, Saltzman RL (2001). A randomized, double-blind trial of famciclovir versus acyclovir for the treatment of localized dermatomal herpes zoster in immunocompromised patients. *Cancer Investig 19*(1), 13–22.
64. Tyring S, Barbarash RA, Nahlik JE, et al. (1995). Famciclovir for the treatment of acute herpes zoster: Effects on acute disease and postherpetic neuralgia. A randomized, double-blind, placebo-controlled trial. Collaborative Famciclovir Herpes Zoster Study Group. *Ann Intern Med 123*(2), 89–96.
65. Tyring SK, Beutner KR, Tucker BA, Anderson WC, Crooks RJ (2000). Antiviral therapy for herpes zoster: Randomized, controlled clinical trial of valacyclovir and famciclovir therapy in immunocompetent patients 50 years and older. *Arch Family Med 9*(9), 863–869.
66. Morfin F, Thouvenot D (2003). Herpes simplex virus resistance to antiviral drugs. *J Clin Virol 26*(1), 29–37.
67. Stranska R, Schuurman R, Nienhuis E, et al. (2005). Survey of acyclovir-resistant herpes simplex virus in the Netherlands: prevalence and characterization. *J Clin Virol 32*(1), 7–18.
68. De Clercq E (2003). Highly potent and selective inhibition of varicella-zoster virus replication by bicyclic furo[2,3-*d*]pyrimidine nucleoside analogues. *Med Res Rev 23*(3), 253–274.
69. Andrei G, Sienaert R, McGuigan C, De Clercq E, Balzarini J, Snoeck R (2005). Susceptibilities of several clinical varicella-zoster virus (VZV) isolates and drug-resistant VZV strains to bicyclic furano pyrimidine nucleosides. *Antimicrob Agents Chemother 49*(3), 1081–1086.
70. Spruance SL, Tyring SK, Smith MH, Meng TC (2001). Application of a topical immune response modifier, resiquimod gel, to modify the recurrence rate of recurrent genital herpes: A pilot study. *J Infect Dis 184*(2), 196–200.
71. Mark KE, Corey L, Meng TC, et al. (2007). Topical resiquimod 0.01% gel decreases herpes simplex virus type 2 genital shedding: A randomized, controlled trial. *J Infect Dis 195*(9), 1324–1331.
72. Crute JJ, Grygon CA, Hargrave KD, et al. (2002). Herpes simplex virus helicase-primase inhibitors are active in animal models of human disease. *Nat Med 8*(4), 386–391.
73. Kleymann G, Fischer R, Betz UA, et al. (2002). New helicase-primase inhibitors as drug candidates for the treatment of herpes simplex disease. *Nat Med 8*(4), 392–398.
74. Betz UA, Fischer R, Kleymann G, Hendrix M, Rubsamen-Waigmann H (2002). Potent in vivo antiviral activity of the herpes simplex virus primase-helicase inhibitor BAY 57-1293. *Antimicrob Agents Chemother 46*(6), 1766–1772.
75. Pilger BD, Cui C, Coen DM (2004). Identification of a small molecule that inhibits herpes simplex virus DNA Polymerase subunit interactions and viral replication. *Chem Biol 11*(5), 647–654.
76. Bernard HU (1994). Coevolution of papillomaviruses with human populations. *Trends Microbiol 2*(4), 140–143.
77. Bernard HU (2005). The clinical importance of the nomenclature, evolution and taxonomy of human papillomaviruses. *J Clin Virol 32*(Suppl 1), S1–6.
78. Snoeck R (2006). Papillomavirus and treatment. *Antivir Res 71*(2–3), 181–191.
79. Garland SM (2003). Imiquimod. *Curr Opin Infect Dis 16*(2), 85–89.
80. Tyring SK, Arany I, Stanley MA, et al. (1998). A randomized, controlled, molecular study of condylomata acuminata clearance during treatment with imiquimod. *J Infect Dis 178*(2), 551–555.
81. Beutner KR, Tyring SK, Trofatter KF Jr, et al. (1998). Imiquimod, a patient-applied immune-response modifier for treatment of external genital warts. *Antimicrob Agents Chemother 42*(4), 789–794.
82. Edwards L, Ferenczy A, Eron L, et al. (1998). Self-administered topical 5% imiquimod cream for external anogenital warts. HPV Study Group. Human PapillomaVirus. *Arch Dermatol 134*(1), 25–30.
83. Beutner KR, Spruance SL, Hougham AJ, Fox TL, Owens ML, Douglas JM Jr (1998). Treatment of genital warts with an immune-response modifier (imiquimod). *J Am Acad Dermatol 38*(2 Pt 1), 230–239.
84. van Seters M, Fons G, van Beurden M (2002). Imiquimod in the treatment of multifocal vulvar intraepithelial neoplasia 2/3. Results of a pilot study. *J Reproduct Med 47*(9), 701–705.
85. Petrow W, Gerdsen R, Uerlich M, Richter O, Bieber T (2001). Successful topical immunotherapy of bowenoid papulosis with imiquimod. *Br J Dermatol 145*(6), 1022–1023.
86. Christensen ND, Pickel MD, Budgeon LR, Kreider JW (2000). In vivo anti-papillomavirus activity of nucleoside analogues including cidofovir on CRPV-induced rabbit papillomas. *Antivir Res 48*(2), 131–142.
87. Van Cutsem E, Snoeck R, Van Ranst M, et al. (1995). Successful treatment of a squamous papilloma of the hypopharynx-esophagus by local injections of (*S*)-1-(3-

hydroxy-2-phosphonylmethoxypropyl)cytosine. *J Med Virol 45*(2), 230–235.

88. Snoeck R, Wellens W, Desloovere C, et al. (1998). Treatment of severe laryngeal papillomatosis with intralesional injections of cidofovir [(*S*)-1-(3-hydroxy-2-phosphonylmethoxypropyl)cytosine]. *J Med Virol 54*(3), 219–225.
89. Kimberlin DW (2004). Current status of antiviral therapy for juvenile-onset recurrent respiratory papillomatosis. *Antivir Res 63*(3), 141–151.
90. Lee AS, Rosen CA (2004). Efficacy of cidofovir injection for the treatment of recurrent respiratory papillomatosis. *J Voice 18*(4), 551–556.
91. Mandell DL, Arjmand EM, Kay DJ, Casselbrant ML, Rosen CA (2004). Intralesional cidofovir for pediatric recurrent respiratory papillomatosis. *Arch Otolaryngol Head Neck Surg 130*(11), 1319–1323.
92. Shehab N, Sweet BV, Hogikyan ND (2005). Cidofovir for the treatment of recurrent respiratory papillomatosis: A review of the literature. *Pharmacotherapy 25*(7), 977–989.
93. Silverman DA, Pitman MJ (2004). Current diagnostic and management trends for recurrent respiratory papillomatosis. *Curr Opin Otolaryngol Head Neck Surg 12*(6), 532–537.
94. Hengge UR, Tietze G (2000). Successful treatment of recalcitrant condyloma with topical cidofovir. *Sex Transm Infect 76*(2), 143.
95. Toro JR, Sanchez S, Turiansky G, Blauvelt A (2003). Topical cidofovir for the treatment of dermatologic conditions: Verruca, condyloma, intraepithelial neoplasia, herpes simplex and its potential use in smallpox. *Dermatol Clin 21*(2), 301–309.
96. Coremans G, Margaritis V, Snoeck R, Wyndaele J, De Clercq E, Geboes K (2003). Topical cidofovir (HPMPC) is an effective adjuvant to surgical treatment of anogenital condylomata acuminata. *Dis Colon Rectum 46*(8), 1103–1108; discussion 8–9.
97. Snoeck R, Van Ranst M, Andrei G, et al (1995). Treatment of anogenital papillomavirus infections with an acyclic nucleoside phosphonate analogue. *N Eng J Med 333*(14), 943–944.
98. Zabawski EJ, Jr., Sands B, Goetz D, Naylor M, Cockerell CJ (1997). Treatment of verruca vulgaris with topical cidofovir. *JAMA 278*(15), 1236.
99. Zabawski EJ, Jr., Cockerell CJ (1998). Topical and intralesional cidofovir: A review of pharmacology and therapeutic effects. *J Am Acad Dermatol 39*(5 Pt 1), 741–745.
100. Snoeck R, Van Laethem Y, De Clercq E, De Maubeuge J, Clumeck N (2001). Treatment of a bowenoid papulosis of the penis with local applications of cidofovir in a patient with acquired immunodeficiency syndrome. *Arch Intern Med 161*(19), 2382–2384.
101. Snoeck R, Bossens M, Parent D, et al. (2001). Phase II double-blind, placebo-controlled study of the safety and efficacy of cidofovir topical gel for the treatment of patients with human papillomavirus infection. *Clin Infect Dis 33*(5), 597–602.
102. Snoeck R, Andrei G, De Clercq E (2001). Cidofovir in the treatment of HPV-associated lesions. *Verhandelingen—Koninklijke Academie voor Geneeskunde van Belgie 63*(2), 93–120, discussion, 2.
103. Snoeck R, Noel JC, Muller C, De Clercq E, Bossens M (2000). Cidofovir, a new approach for the treatment of cervix intraepithelial neoplasia grade III (CIN III). *J Med Virol 60*(2), 205–209.
104. De Clercq E (2003). Clinical potential of the acyclic nucleoside phosphonates cidofovir, adefovir, and tenofovir in treatment of DNA virus and retrovirus infections. *Clin Microbiol Rev 16*(4), 569–596.
105. Andrei G, Snoeck R, Piette J, Delvenne P, De Clercq E (1998). Antiproliferative effects of acyclic nucleoside phosphonates on human papillomavirus (HPV)-harboring cell lines compared with HPV-negative cell lines. *Oncol Res 10*(10), 523–531.
106. Andrei G, Snoeck R, Piette J, Delvenne P, De Clercq E (1998). Inhibiting effects of cidofovir (HPMPC) on the growth of the human cervical carcinoma (SiHa) xenografts in athymic nude mice. *Oncol Res 10*(10), 533–539.
107. Andrei G, Snoeck R, Schols D, De Clercq E (2000). Induction of apoptosis by cidofovir in human papillomavirus (HPV)-positive cells. *Oncol Res 12*(9–10), 397–408.
108. Abdulkarim B, Sabri S, Deutsch E, et al. (2002). Antiviral agent cidofovir restores p53 function and enhances the radiosensitivity in HPV-associated cancers. *Oncogene 21*(15), 2334–2346.
109. Johnson JA, Gangemi JD (1999). Selective inhibition of human papillomavirus-induced cell proliferation by (*S*)-1-[3-hydroxy-2-(phosphonylmethoxy)propyl]cytosine. *Antimicrob Agents Chemother 43*(5), 1198–1205.
110. DiPaolo JA, Alvarez-Salas LM (2004). Advances in the development of therapeutic nucleic acids against cervical cancer. *Expert Opin Biol Ther 4*(8), 1251–1264.
111. Niu XY, Peng ZL, Duan WQ, Wang H, Wang P (2006). Inhibition of HPV 16 E6 oncogene expression by RNA interference in vitro and in vivo. *Int J Gynecol Cancer 16*(2), 743–751.
112. Tang S, Tao M, McCoy JP, Jr., Zheng ZM (2006). The E7 oncoprotein is translated from spliced E6*I transcripts in high-risk human papillomavirus type 16- or type 18-positive cervical cancer cell lines via translation reinitiation. *J Virol 80*(9), 4249–4263.
113. Scheurer ME, Tortolero-Luna G, Adler-Storthz K (2005). Human papillomavirus infection: Biology, epidemiology, and prevention. *Int J Gynecol Cancer 15*(5), 727–746.
114. Govan VA (2005). Strategies for human papillomavirus therapeutic vaccines and other therapies based on the e6 and e7 oncogenes. *Ann NY Acad Sci 1056*, 328–343.
115. Henderson DA, Inglesby TV, Bartlett JG, et al. (1999). Smallpox as a biological weapon: Medical and public

health management. Working Group on Civilian Biodefense. *JAMA 281*(22), 2127–2137.

116. Whitley RJ (2003). Smallpox: A potential agent of bioterrorism. *Antivir Res 57*(1–2), 7–12.
117. Lewis-Jones S (2004). Zoonotic poxvirus infections in humans. *Curr Opin Infect Dis 17*(2), 81–89.
118. Baker RO, Bray M, Huggins JW (2003). Potential antiviral therapeutics for smallpox, monkeypox and other orthopoxvirus infections. *Antivir Res 57*(1–2), 13–23.
119. Magee WC, Hostetler KY, Evans DH (2005). Mechanism of inhibition of vaccinia virus DNA polymerase by cidofovir diphosphate. *Antimicrob Agents Chemother 49*(8), 3153–3162.
120. Quenelle DC, Collins DJ, Kern ER (2003). Efficacy of multiple- or single-dose cidofovir against vaccinia and cowpox virus infections in mice. *Antimicrob Agents chemother 47*(10), 3275–3280.
121. Jahrling PB, Fritz EA, Hensley LE (2005). Countermeasures to the bioterrorist threat of smallpox. *Curr Mol Med 5*(8), 817–826.
122. Meadows KP, Tyring SK, Pavia AT, Rallis TM (1997). Resolution of recalcitrant molluscum contagiosum virus lesions in human immunodeficiency virus-infected patients treated with cidofovir. *Arch Dermatol 133*(8), 987–990.
123. Ibarra V, Blanco JR, Oteo JA, Rosel L (2000). Efficacy of cidofovir in the treatment of recalcitrant molluscum contagiosum in an AIDS patient. *Acta Dermato-Venereol 80*(4), 315–316.
124. Geerinck K, Lukito G, Snoeck R, et al. (2001). A case of human orf in an immunocompromised patient treated successfully with cidofovir cream. *J Med Virol 64*(4), 543–549.
125. De Clercq E, Neyts J (2004). Therapeutic potential of nucleoside/nucleotide analogues against poxvirus infections. *Rev Med Virol 14*(5), 289–300.
126. De Clercq E, Andrei G, Balzarini J, et al. (2005). Antiviral potential of a new generation of acyclic nucleoside phosphonates, the 6-[2-(phosphonomethoxy)alkoxy]-2,4-diaminopyrimidines. *Nucleosides Nucleotides Nucleic Acids 24*(5–7), 331–341.
127. De Clercq E (2005). Recent highlights in the development of new antiviral drugs. *Curr Opin Microbiol 8*(5), 552–560.
128. Tyring S (2005). General (non-antiretroviral antiviral drugs). In *Antiviral Agents, Vaccines, and Immunotherapies*. New York: Marcel Dekker, pp. 140–141.
129. Hayden FG, Pavia AT (2006). Antiviral management of seasonal and pandemic influenza. *J Infect Dis 194*(Suppl 2), S119–126.
130. De Clercq E (2004). Antivirals and antiviral strategies. *Nat Rev 2*(9), 704–720.
131. De Clercq E (2002). Strategies in the design of antiviral drugs. *Nat Rev Drug Discov 1*(1), 13–25.
132. Kiso M, Mitamura K, Sakai-Tagawa Y, et al. (2004). Resistant influenza A viruses in children treated with oseltamivir: Descriptive study. *Lancet 364*(9436), 759–765.
133. McKimm-Breschkin J, Trivedi T, Hampson A, et al. (2003). Neuraminidase sequence analysis and susceptibilities of influenza virus clinical isolates to zanamivir and oseltamivir. *Antimicrob Agents Chemother 47*(7), 2264–2272.
134. Hayden FG, Douglas RG Jr, Simons R (1980). Enhancement of activity against influenza viruses by combinations of antiviral agents. *Antimicrob Agents and Chemother 18*(4), 536–541.
135. Ison MG, Gnann JW Jr, Nagy-Agren S, et al. (2003). Safety and efficacy of nebulized zanamivir in hospitalized patients with serious influenza. *Antivir Ther 8*(3), 183–190.
136. Milroy D, Featherstone J (2002). Antiviral market overview. *Nat Rev Drug Discov 1*(1), 11–12.
137. Wald A, Zeh J, Barnum G, Davis LG, Corey L (1996). Suppression of subclinical shedding of herpes simplex virus type 2 with acyclovir. *Ann Intern Med 124*(1 Pt 1), 8–15.
138. Sheffield JS, Hollier LM, Hill JB, Stuart GS, Wendel GD (2003). Acyclovir prophylaxis to prevent herpes simplex virus recurrence at delivery: A systematic review. *Obstet Gynecol 102*(6), 1396–1403.
139. Goldberg LH, Kaufman R, Kurtz TO, et al. (1993). Long-term suppression of recurrent genital herpes with acyclovir. A 5-year benchmark. Acyclovir Study Group. *Arch Dermatol 129*(5), 582–587.
140. Spruance SL (1993). Prophylactic chemotherapy with acyclovir for recurrent herpes simplex labialis. *J Med Virol 41* (Suppl 1), 27–32.
141. Heng MC, Heng SY, Allen SG (1994). Co-infection and synergy of human immunodeficiency virus-1 and herpes simplex virus-1. *Lancet 343*(8892), 255–258.
142. Mole L, Ripich S, Margolis D, Holodniy M (1997). The impact of active herpes simplex virus infection on human immunodeficiency virus load. *J Infect Dis 176*(3), 766–770.
143. Reyes M, Shaik NS, Graber JM, et al. (2003). Acyclovir-resistant genital herpes among persons attending sexually transmitted disease and human immunodeficiency virus clinics. *Arch Intern Med 163*(1), 76–80.
144. Thandi A, Tyring, SK (1997). Newer aspects of herpesvirus infections. In McOrary M and OM Memar, Eds. *Current Opinions in Dermatology*. Philadelphia: Rapid Science, pp. 42–50.
145. Klassen TP, Hartling L, Wiebe N, Belseck EM (2005). Acyclovir for treating varicella in otherwise healthy children and adolescents. Cochrane Database of Systematic Reviews. Available: (online), (4), CD002980. www.cochrane.org/cochrane-reviews.
146. Smego RA Jr, Asperilla MO (1991). Use of acyclovir for varicella pneumonia during pregnancy. *Obstet Gynecol 78*(6), 1112–1116.

147. Huff JC, Drucker JL, Clemmer A, et al. (1993). Effect of oral acyclovir on pain resolution in herpes zoster: A reanalysis. *J Med Virol* (Suppl 1), 93–96.
148. Lemak MA, Duvic M, Bean SF (1986). Oral acyclovir for the prevention of herpes-associated erythema multiforme. *J Am Acad Dermatol 15*(1), 50–54.
149. Schofield JK, Tatnall FM, Leigh IM (1993). Recurrent erythema multiforme: Clinical features and treatment in a large series of patients. *Br J Dermatol 128*(5), 542–545.
150. Tatnall FM, Schofield JK, Leigh IM (1995). A double-blind, placebo-controlled trial of continuous acyclovir therapy in recurrent erythema multiforme. *Br J Dermatol 132*(2), 267–270.
151. Fife KH, Barbarash RA, Rudolph T, Degregorio B, Roth R (1997). Valaciclovir versus acyclovir in the treatment of first-episode genital herpes infection. Results of an international, multicenter, double-blind, randomized clinical trial. The Valaciclovir International Herpes Simplex Virus Study Group. *Sex Transm Dis 24*(8), 481–486.
152. Leone PA, Trottier S, Miller JM (2002). Valacyclovir for episodic treatment of genital herpes: A shorter 3-day treatment course compared with 5-day treatment. *Clin Infect Dis 34*(7), 958–962.
153. Corey L, Wald A, Patel R, et al. (2004). Once-daily valacyclovir to reduce the risk of transmission of genital herpes. *N Engl J Med 350*(1), 11–20.
154. Conant MA, Schacker TW, Murphy RL, Gold J, Crutchfield LT, Crooks RJ (2002). Valaciclovir versus aciclovir for herpes simplex virus infection in HIV-infected individuals: Two randomized trials. *Int J STD AIDS 13*(1), 12–21.
155. Reischig T, Jindra P, Mares J, et al. (2005). Valacyclovir for cytomegalovirus prophylaxis reduces the risk of acute renal allograft rejection. *Transplantation 79*(3), 317–324.
156. Spruance SL, Jones TM, Blatter MM, et al. (2003). High-dose, short-duration, early valacyclovir therapy for episodic treatment of cold sores: Results of two randomized, placebo-controlled, multicenter studies. *Antimicrob Agents Chemother 47*(3), 1072–1080.
157. Baker D, Eisen D (2003). Valacyclovir for prevention of recurrent herpes labialis: 2 double-blind, placebo-controlled studies. *Cutis; Cutaneous Medicine for the Practitioner 71*(3), 239–242.
158. Beeson WH, Rachel JD (2002). Valacyclovir prophylaxis for herpes simplex virus infection or infection recurrence following laser skin resurfacing. *Dermatol Surg 28*(4), 331–336.
159. Kerob D, Assier-Bonnet H, Esnault-Gelly P, Blanc F, Saiag P (1998). Recurrent erythema multiforme unresponsive to acyclovir prophylaxis and responsive to valacyclovir continuous therapy. *Arch Dermatol 134*(7), 876–877.
160. Diaz-Mitoma F, Sibbald RG, Shafran SD, Boon R, Saltzman RL (1998). Oral famciclovir for the suppression of recurrent genital herpes: A randomized controlled trial. Collaborative Famciclovir Genital Herpes Research Group. *JAMA 280*(10), 887–892.
161. Tyring SK, Diaz-Mitoma F, Shafran SD, Locke LA, Sacks SL, Young CL (2003). Oral famciclovir for the suppression of recurrent genital herpes: The combined data from two randomized controlled trials. *J Cutan Med Surg 7*(6), 449–454.
162. Spruance SL, Rowe NH, Raborn GW, Thibodeau EA, D'Ambrosio JA, Bernstein DI (1999). Peroral famciclovir in the treatment of experimental ultraviolet radiation-induced herpes simplex labialis: A double-blind, dose-ranging, placebo-controlled, multicenter trial. *J Infect Dis 179*(2), 303–310.
163. Alster TS, Nanni CA (1999). Famciclovir prophylaxis of herpes simplex virus reactivation after laser skin resurfacing. *Dermatol Surg 25*(3), 242–246.
164. Romanowski B, Aoki FY, Martel AY, Lavender EA, Parsons JE, Saltzman RL (2000). Efficacy and safety of famciclovir for treating mucocutaneous herpes simplex infection in HIV-infected individuals. Collaborative Famciclovir HIV Study Group. *AIDS (Lond) 14*(9), 1211–1217.
165. Schacker T, Hu HL, Koelle DM, et al. (1998). Famciclovir for the suppression of symptomatic and asymptomatic herpes simplex virus reactivation in HIV-infected persons. A double-blind, placebo-controlled trial. *Ann Intern Med 128*(1), 21–28.
166. Enzensberger R, Braun W, July C, Helm EB, Doerr HW (1991). Prevalence of antibodies to human herpesviruses and hepatitis B virus in patients at different stages of human immunodeficiency virus (HIV) infection. *Infection 19*(3), 140–145.
167. Schacker T, Ryncarz AJ, Goddard J, Diem K, Shaughnessy M, Corey L (1998). Frequent recovery of HIV-1 from genital herpes simplex virus lesions in HIV-1-infected men. *JAMA 280*(1), 61–66.
168. Tyring S, Engst R, Corriveau C, et al. (2001). Famciclovir for ophthalmic zoster: A randomised aciclovir controlled study. *Br J Ophthalmol 85*(5), 576–581.

# HIGH-THROUGHPUT SCREENING FOR SMALL-MOLECULE DRUG DISCOVERY

C. Parker and J.-H. Zhang

*Lead Finding Platform, Novartis Institute for Biomedical Research, Cambridge, Massachusetts*

## 1 INTRODUCTION

There is ample evidence that high-throughput screening (HTS) has become one of the main methods for generating leads for drug discovery [1]. The capacity and throughput for screening has been growing steadily for the last two decades. HTS has transformed biological testing from a process using test-tube and cuvette measurements to one using high-density, low-volume assay formats and screening systems. This transformation has been in response to rapid changes in biological target identification and validation, facilitated by advances in genomic and proteomic studies and compound synthetic methodologies. Particularly, the development and utilization of miniaturized, homogeneous assay technologies and various signal detection methodologies has made possible the screening of hundreds of thousands and even millions of compounds in a relatively short time period. Assay format standardization, instrumentation automation, assay miniaturization, and informatics tools for data analysis have greatly facilitate the screening process by increasing reliability and reducing personnel workload. Biomolecular screening has become a fast growing and critical field in early drug discovery. With the demands for ever-increasing productivity in drug discovery and development, one can only expect that the emphasis on discovering drug candidates faster and of higher quality will increase. These demands impose new challenges and also present new opportunities for the field of HTS and its related drug discovery activities.

There are many different drug discovery strategies where HTS can be applied, from the conceptually simple "screen everything we have" strategy to various strategies that call for the screening of subsets of compounds chosen to fulfill a defined hypothesis [2]. Although every strategy has its own advantages and disadvantages, they all require compounds or samples to be screened for biological activity in an efficient and reliable manner. The involvement of HTS activities in different phases of drug discovery and development are shown in Figure 1.

In a broader view, HTS should not be regarded as a specialty field, only performing assay development, method validation, and screening. HTS has become more of a collective term for a range of multidisciplinary activities, ranging from the creation of cell lines or recombinant enzymes for screening to computational and medicinal chemists designing and making screening compound libraries, and from engineers creating suitable robots and instrumentation allowing the automation of biological assays to statisticians and modelers who help analyze screening data to extract useful and predictive models of mode of interaction. Figure 2 illustrates the interconnection of scientific disciplines required for successful HTS and hit-to-lead optimization processes.

There are two major assay types in HTS, biochemical and cell-based assays, as will be illustrated in the following section. For any specific target, there are usually several assay formats available. In the past, assays involving the use of radioactive isotopes had been used

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

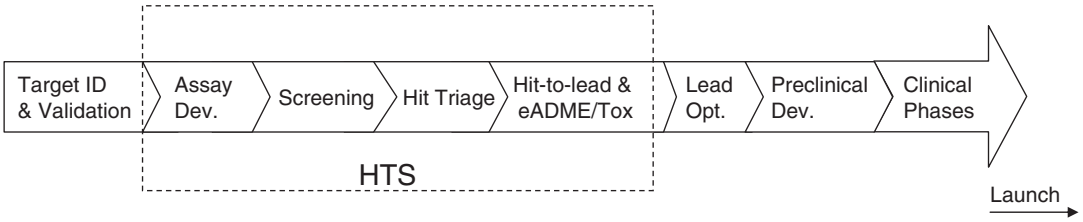


**Figure 1** Stages of drug discovery and development and the expanded involvement of HTS activities.

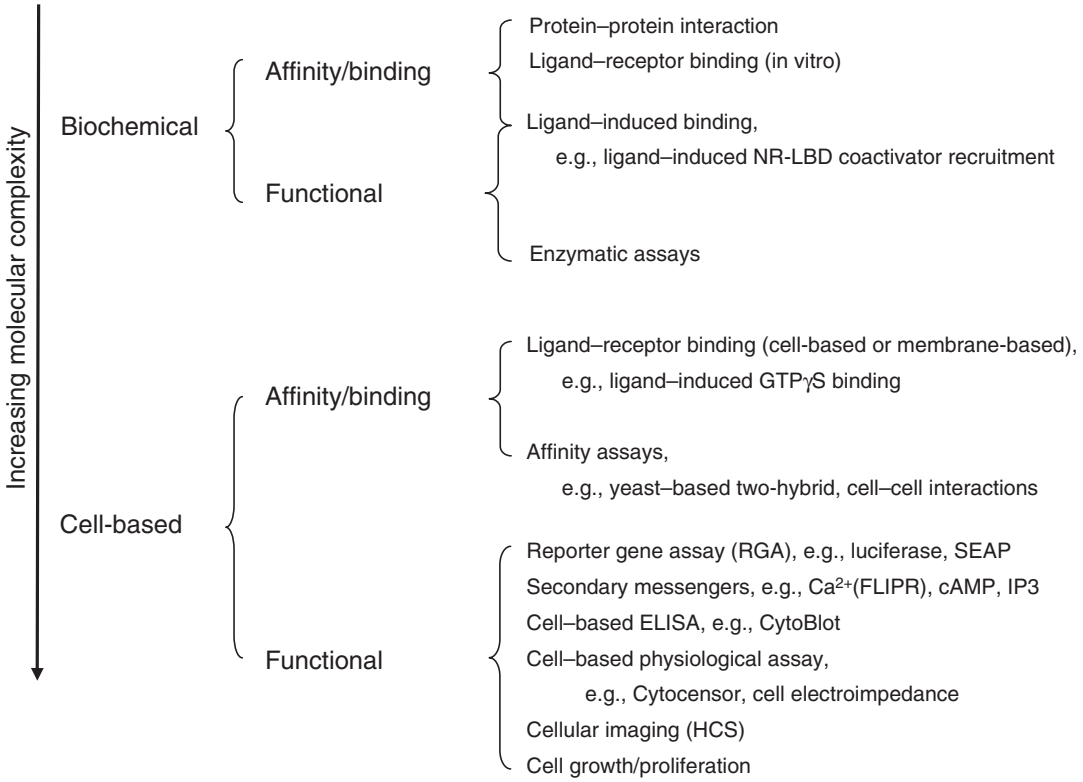


**Figure 2** Interconnection of scientific disciplines and activities involved in HTS.

routinely. Although radioactive assays are very sensitive and reliable, more and more nonradioactive assay formats are being employed in HTS in recent years to avoid the added safety issues and inconvenience of handling and disposing of radioactive materials. Assay formats not requiring separation steps, that is, homogeneous assays such as time-resolved fluorescence resonance energy transfer (TR-FRET) (or referred to as homogeneous time-resolved fluorescence [HTRF]), fluorescence polarization (FP), AlphaScreen, scintillation proximity assay (SPA), and chemiluminescence (e.g., HitHunter™ cAMP detection) have become popular formats in HTS. For cell-based assays, several common assay formats are reporter genes (luciferase, etc.), Fluorescence Imaging Plate Reader (FLIPR), Cytoblot™, and high content imaging (HCI). Some newer assay formats based on label-free detection [waveguide sensing, cell electronic impedance, liquid chromatography/mass spectrometry (LC/MS)] will emerge as potentially preferred assay formats in the coming years (see Section 6 of this chapter for more details).

The challenge facing HTS scientists is not only that an assay developed for a target must be biologically relevant and as specific as possible, but that HTS is also limited by practical constraints such as cost, scale, and throughput. An HTS must be properly designed and configured to maximize both screening efficiency and biological effectiveness. An improperly designed HTS assay may fail to detect the desired classes of hits even if it is trivial to execute. Moreover, drug screening can only be successful when the compound libraries interrogated are populated with druglike (or leadlike) molecules so that the identified active compounds can be evolved into lead compounds with suitable absorption, distribution, metabolism, and excretion (ADME)

properties for later drug development. In the last decade, the number of new chemical entities entering the medicinal marketplace has not increased as fast as would be expected given the increased capacity of HTS. Therefore, the HTS should be coupled more to pharmacological properties in order to make screening more relevant and predicative.

The aim of this chapter is to review some of the HTS processes in current drug discovery, including assay technologies and screening strategies, the steps taken in assay development and validation, the screening process, and commonly used quality control measures. Also highlighted in this chapter are the analysis of HTS data and some of the approaches used for data mining for finding lead series. Finally, the chapter reviews some of the emerging trends occurring in the science of biomolecular screening.

## 2 TYPES OF HTS ASSAYS

High-throughput assays can be used to probe different levels of molecular interactions in a continuum of complexity, ranging from simple molecular binding events to effects at the cellular or even whole organism level. These assays can be classified in several different types. Broadly speaking, there are two major assay types in HTS: in vitro biochemical and cell-based assays. The biochemical assays can be further divided into molecular binding (affinity) and in vitro functional assays (e.g., enzymatic reactions). There is sometimes, however, no clear distinction between the affinity binding and in vitro functional assays. For example, ligand-induced GTPγS binding to G proteins can both be regarded as a binding or a functional assay. A second example is the ligand-activated coactivator recruitment to the ligand-binding domain of a nuclear receptor. Cell-based assays can also further divided into cell- or membrane-based binding assays and functional cell-based assays (Fig. 3). This section is an overview of the assay technologies commonly used for HTS. Some newer trends will be discussed in Section 6 at the end of the chapter.

### 2.1 In Vitro Biochemical Assays

In modern HTS, in vitro biochemical assays have been used extensively against various targets, for example, enzymes, receptor–ligand, and protein–protein interactions. This reductionist approach has been facilitated by various biochemical and genetic studies for delineating complicated metabolic, signal transduction pathways, and biological systems involved in disease. It offers the advantage of clear drug–target interactions, leading to clean structure–activity relationship (SAR) during hit-to-lead optimization. The drawbacks of this approach include sometimes a lack of physiological context, compared to cell-based assays, which can lead to a lack of in vivo efficacy. A clear trend is to combine both approaches so as to gain the advantages of biochemical and cell-based screens while minimizing the risks.

Molecular binding interactions are fundamental to biological processes and together with enzymatic assays have formed the backbone of HTS assays for drug discovery. Targeting of specific enzymes and components in signal transduction, metabolic pathways, protein processing, and inflammatory cascades has proven to be a valid approach for addressing disease processes. Many classes of therapeutically relevant enzymes have been actively pursued using HTS for lead generation. The majority of these enzymatic assays have been adapted in miniaturized microplate format, automated, and used for compound screening or even mutant enzyme screening.

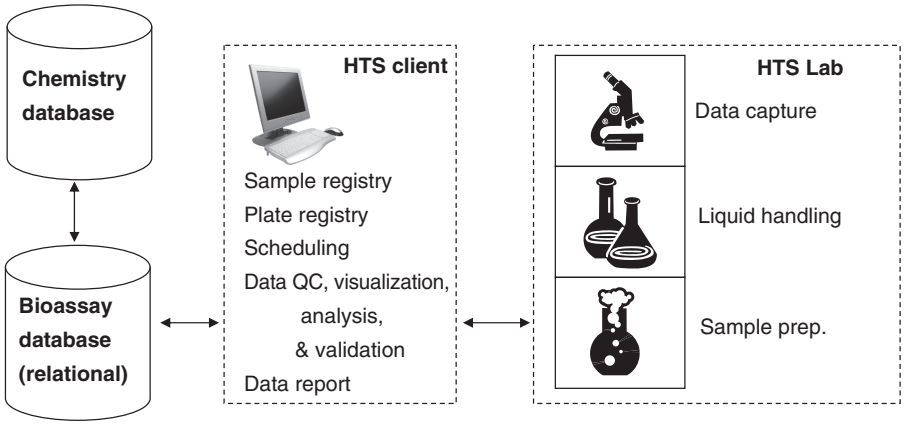


**Figure 3** Hierarchy of assays monitoring biological processes of increasing molecular complexity.

Several of the most commonly used high-throughput assay technologies based on direct molecular affinity binding events and enzymatic activities are highlighted below. Many of these assay and detection technologies can be used to measure both binding and enzymatic activities. Several reviews on HTS assays are available, for example, the books edited by Seethala and Fernandes [3] or Janzen [4] contain many excellent review articles.

***Fluorescence-Based Assays*** Currently, the most commonly used detection methods for high-throughput screening are fluorescence based, as these readouts are sensitive, usually homogeneous, and can be readily miniaturized, even down to the single molecule level [5]. Fluorescent signal can be detected by methods such as fluorescence intensity (FI), FP, fluorescence resonance energy transfer (FRET), TR-FRET, and single-molecule techniques such as fluorescence intensity lifetime and fluorescence correlation spectroscopy (FCS). These homogeneous assay formats are "mix and read" type of assays that allow high throughput and assay miniaturization.

*Fluorescence Intensity Assays* These assays measure fluorescence intensity change using fluorogenic substrates or internally quenched substrates. The enzymatic reaction generates a fluorescent product or releases quenching effects on the substrate. These assays are used mostly extensively for hydrolase assays, such as protease assays (e.g., [6]).

*FRET Assays* Fluorescence resonance energy transfer occurs when donor (D) and acceptor (A) molecules are brought in close proximity (e.g., 10–60 Å), with sufficient energy overlap and parallel in transition dipole orientation [7]. FRET can be measured as the quenching of donor fluorescence by an acceptor or as the sensitized fluorescence emission of an acceptor. Some D/A pairs are the 5-(2-aminoethylamino) naphthalene-1-sulfonic acid (EDANS)/4-(4-dimethylaminophenylazo) benzoic acid (DABCYL) with typical Föster radius ($R_0$) of 33 Å and the fluorescein/tetramethylrhodamine pair with an $R_0$ value of 55 Å. FRET assays have been developed for a variety of different target types, such as protease or kinase assays [8].

*TR-FRET Assays* TR-FRET assays have been commercialized under a number of different acronyms, including: HTRF (CisBio), LANCE (Perkin-Elmer), and LanthaScreen (Invitrogen): The assay principle is similar to FRET, that is, based on fluorescent resonance energy transfer from a fluorescent donor moiety (D) to an acceptor moiety (A) when A and D are brought into close proximity. The difference is that TR-FRET involves time-resolved fluorescence (TRF), based on the long fluorescence lifetime properties of certain lanthanide ions, for example, europium (Eu), terbium (Tb), and samarium (Sm). An added advantage of these lanthanide ion chelates is their relatively large Föster radius (~100 Å), allowing energy transfer over distances greater than for many other FRET pairs. The europium complex (D) paired with any of the allophycocyanine (APC), XL665, or Cy5 acceptor dyes or the terbium (D) paired with fluorescein (A) are frequently used to probe in vitro receptor–ligand and protein–protein interactions. The receptor–ligand or protein–protein interaction or other binding partners can be directly covalently linked to one of the A/D moieties or indirectly via A/D-conjugated antibodies or biotin/avidin binding. This homogeneous assay technology is widely used in binding and enzymatic assays [9–11] and can also be used for detection of cell metabolites for cell-based assays [12] .

*Fluorescence Polarization Assays* FP is a homogeneous assay format for the determination of molecular interactions in solution. When a fluorescent molecule is excited with polarized light, the degree to which the emitted light retains that polarization is proportional to the molecule's rotational relaxation time. Small fluorescent molecules rotate faster in solution and give a low polarization value, while large fluorescent molecules rotate slower and give a high polarization value. Both polarization and anisotrophy can be used to measure the molecular interaction in solution. In HTS, polarization (P) is defined as $(I_{\parallel} - I_{\perp})/(I_{\parallel} + I_{\perp})$, where $I_{\parallel}$ and $I_{\perp}$ represents the fluorescence intensity parallel and perpendicular to the incident light, respectively. The polarization value is usually expressed as milli-P (mP:1/1000 of 1 P unit). The FP signal is between 0 and 500 mP for solution samples and is in the range of 40–300 mP for many HTS assays experimentally. Even though in many cases the milli-P change is used to calculate the binding or product conversion, it should be noted that the FP signal is not directly proportional to the concentration of the large fluorescent moiety. Therefore, a standard curve is needed to convert the milli-P values to the concentration of the related specific fluorescent molecules. Fluorescence polarization is suitable as a detection method when the mass of the fluorescently labeled tracer changes during binding or enzymatic events [13]. Constraints on the lifetime of the fluorescence of most labels makes FP less suitable for probing interactions between two large macromolecules.

Some of the following fluorescent detection technologies have also been used in HTS, albeit less commonly than the above-mentioned technologies.

*Fluorescent Correlation Spectroscopy* In FCS, a laser beam focuses on a very small volume (femtoliter) in a solution and monitors spontaneously fluctuating fluorescence molecules in a statistical manner. The diffusion time and brightness of the fluorescence molecules reveal binding or reaction activity information. FCS is suitable for ultra-low-volume HTS (uHTS) assays [14].

*One- and Two-Dimensional Fluorescence Intensity Distribution Analysis (FIDA)* Similar to fluorescent correlation spectroscopy, one- and two-dimensional (1D and 2D) FIDA use Evotec's confocal fluorescence readouts. Besides being able to distinguish different fluorescent species (e.g., bound and unbound labeled molecules) in one sample, these technologies measure and simultaneously determine multiple, species-resolved parameters, such as particle number (concentration) or molecular brightness (i.e., fluorescence count rate per particle) or anisotropy. These parameters can be used to distinguish real hits from false positives, including artifacts induced by autofluorescent compounds, biological reagents, and technical errors [15]. Both in vitro receptor binding and enzymatic reactions can be detected in these formats.

*Fluorescent Microvolume Assay Technology (FMAT)* FMAT is a bead-based or cell-based fluorescent technology for nonseparation detection. In FMAT, a laser beam is focused on the bottom of the well of the assay plate and fluorescence intensity of beads (via binding to fluorescently labeled molecules) is detected over the unbound and background fluorescence in solution. FMAT can be configured as a bead-based or cell-based assay for detection of specific biomolecules in solution or receptors on the cell surface. Up to two different analytes can be detected with appropriate fluorophores and detection setup [16].

*Luminex Assays* This assay technology is a bead-based, nonseparation technology using the Luminex color-coded beads and detection systems (Luminex 100 IS or Luminex HT). The readers used for this assay format are based on the principle of flow cytometry. The system enables assays to be multiplexed, that is, to allow one to simultaneously measure different analytes. The Luminex HT system is compatible with 96- and 384-well microplates, allowing rapid and cost-effective bioassays [17]. The throughput of the reader is still currently a limiting factor for large-scale HTS.

*Heterogeneous Fluorescent Assays—TRF or Delfia Assay (Dissociation-Enhanced Lanthanide Fluoroimmuno Assay)* This assay format uses time-resolved fluorescence in combination with enzyme-linked immunosorbent assay (ELISA)-like analyte capture methods. After washing the captured fluorescent molecules on the surface or beads, the bound molecules can be detected by TRF. This assay format yields very high detection sensitivity with low background [18]. It is essentially an alternative version of ELISA or ELISA-like assays, without the use of conjugated enzymes for detection.

***Luminescence-Based Assays*** Luminescence detection has the advantage of very low background over fluorescent technologies. Hence, the sensitivity of the assay can be extremely high. The following section highlights several assay technologies based on luminescence, including chemiluminescence and electrochemiluminescence.

*AlphaScreen Assay* The AlphaScreen assay format is somewhat similar to TR-FRET assays and is another versatile, nonseparation screening technology. It is a bead-based, chemiluminescent assay technology. The donor beads contain a photosensitizer that absorbs light at 680 nm and converts molecular oxygen to its excited singlet state. The accepter beads contain thioxene derivatives that react with singlet oxygen, generating chemiluminescence at 370 nm, which then excites fluorophores in the same beads, emitting light at 520–620 nm. The effective distance from donor to acceptor beads is ~200 nm in aqueous solution, much larger than that in TR-FRET assays. Another feature of AlphaScreen is that the amplification of the signal is huge, offering a better sensitivity than TR-FRET assay format. However, the amplified signal may carry a higher variability compared to TR-FRET assay format [19]. Both TR-FRET and AlphaScreen assay formats can be used not only for many in vitro binding or enzymatic assays but also to probe cell lysates for detection of secondary messengers or specific proteins [20].

*Electrochemiluminescence (ECL) Assay* This is a bead-based electrochemiluminescence assay using the Origen technology platform. ECL utilizes ruthenium chelates conjugated to antibody as a tracer, which reacts with tripropylamine (TPA) and emits luminescence upon application of a low voltage. Magnetic beads are utilized to align the beads to the surface of an electrode. This assay has been mostly used in diagnostic bioassays but also developed for HTS use [21].

*Other Luminescence-Based Readouts* Chemiluminescence detection methods based on enzymes such as horseradish peroxidase have been used as improvements to standard ELISA-based assays. The number of different luminescence-based assays has been expanding. For example, the Kinase Glo assay format monitors

the levels of adenosine triphosphate (ATP) remaining in a reaction mix. While this is originally developed for assaying of kinase targets, it could be easily adapted to monitor any ATP-consuming (or generating) reaction [22]. Recently, additional luminescence assay formats have been developed using substrates that liberate luciferin upon action of the target enzyme and have been applied to proteases, such as caspases [23] and cytochrome P450s [24].

***Colorimetric and Chromogenic Assays*** Colorimetric assays are not used as frequently in high-throughput settings as fluorimetric assays for several reasons. First, colorimetric detection is usually less sensitive than the more widely used fluorimetric detection, due to the dependency of the readout on the path length of light passing through the liquid sample (Beer's law). This issue poses a limit to the miniaturization of assays into lower volumes. However, it should be noted that clever methods to increase the apparent path length of light have been used to allow absorbance-based assays to be miniaturized even into 1536-well plate formats with reaction volumes of only 10 μL [25]. The second constraint on colorimetric assays is that colored compounds that are present in most large compound libraries interfere with detection, leading to screening artifacts. This issue can be overcome by monitoring the enzyme reaction rate over time instead of using an endpoint readout. The change in reaction rate is not so greatly influenced by the presence of colored compounds, as has been shown in screening using the relatively weak absorbance of reduced nicotinamide adenine dinucleotide phosphate (NADPH) [26] or the weak absorption of nitrophenolate [27].

***Coupled Assays*** Coupled assays have been used to monitor a variety of different enzymatic assays and can be split into two types: chemically coupled and enzymatically coupled assays. In the former type, the product of the enzymatic reaction under study is detected by reaction with a reactive chemical to allow easy detection of the analyte of interest. Examples of such assay format include the so-called adrenaline assay for the detection of compounds that can react with sodium periodate, which then reacts with adrenaline to generate a colored product [28] and the detection of free thiol groups by Ellman's reagent [5,5′-dithiobis(2-nitrobenzoate) or DTNB].

In enzymatically coupled reactions, the products or substrates of the reaction of interest are acted upon by a second enzyme, creating a tangible readout. One advantage of this kind of assay format is that the development of such a standard detection method can be used to monitor the activity of a family of enzymes utilizing a common analyte. Examples of such coupled enzyme assays include the detection of S-adenosyl-L-homocysteine (SAH) [29] and the detection of phosphate or $H_2O_2$ via the generation of highly fluorescent resorufin from Amplex red [30]. Recently, this strategy has been extended with a coupled assay system that detects the generation of adenosine 5′-diphosphate (ADP) rather than just phosphate [31]. The disadvantage of these coupled assay systems are that rigorous counterscreens are required to eliminate hits not acting directly upon the primary enzymatic reaction of interest.

One of the areas of assay development that has seen increased application is the use of enzyme complementation assays. While this has primarily been used in cell-based assays, a number of interesting applications of enzyme complementation for in vitro assays have been reported as well [32]. We will also discuss some of these assays under Reporter Gene Assay below.

***MS-Based Detection*** Recently, mass spectrometry coupled with LC/MS-based detection has been shown to be a new and valuable label-free detection tool for HTS. By measuring mass-to charge ($m/z$) ratio of analytes, MS-based assays have the advantage of allowing generic, label-free detection of reaction substrates and products. A group at Biotrove has reported the use of LC/MS setups for HTS [33] , and Zhang [34] has recently reported the successful screening of several metabolic enzymes with LC/MS detection in a large industrial screening scale. Several strategies have been used to increase the speed of LC/MS analysis, which has been the bottleneck of this screening system. These include utilization of four parallel LC units in either multiplexed electrospray (MUX) or serial gradient configurations, and the use of orthogonally pooled matrices of a small number of compounds, for example, 10 compounds per mixture, where each compound is tested twice in two different mixtures. These strategies have been proven to work well, and the analysis of near a million compounds can be achieved in 6–8 weeks. So far, large-scale LC/MS-based screening has been limited to in vitro enzymatic reactions. Other desorption-based MS detection techniques for HTS applications are currently in early stages of development and hold great promise [35]. MS-based HTS will play an increasingly important role in the HTS paradigm in the coming years.

***Microcalorimetry and Thermostability Methods*** Microcalorimetry monitors molecular interactions such as ligand binding by monitoring the enthalpic response of interaction. Typically, isothermal titration calorimetry (ITC) and differential scanning calorimetry (DSC) have been used to measure the binding constants and ther-

mostability of the protein–ligand complex. However, these traditional microcalorimetric methods have a very limited use in screening due to low throughput and high protein consumption. Currently, only a few thermostability-based methods have been modified to allow screening of compounds with reasonable throughput. The patented ThermoFluor affinity screening method is based on the fluorescence change seen upon binding of a specific dye to native protein versus. a denatured form of the protein. This assay technology has been shown to have a higher throughput and relatively lower protein consumption than traditional microcalorimetry methods [36–38]. Other approaches measure the change of fluorescence during the denaturation of a protein at a set temperature and the effect of ligands in altering the rate of denaturation [39] or measure light scattering of these transitions [40]. Techniques to improve the throughput of these approaches include enthalpy arrays [41]. However, the current methods still suffer from the relatively high protein concentration (micromolar range) needed to monitor the ligand effect and their low throughput, compared to many other HTS assay formats.

***Radioisotope-Based Assays*** It has been anticipated for some time that biological assays based on radiolabeled substrates or ligand would be used less frequently in HTS [42]. However, the specificity with which substrates and ligands can be labeled with a radioisotope without perturbing their chemical properties and the high detection sensitivity of these assays continue to ensure that assays using radioactive tracers remain an important tool for HTS assays. A number of well-developed technologies based on scintillation proximity, such as SPA beads [43] or FlashPlates [44], allow complex biological assays to be conducted without the need for time-consuming separation steps (e.g., [45]). Technological advances such as the use of detection methods based on charge-coupled devices (CDD) that monitor the activity of a whole assay plate at once have reduced the time required for conventional scintillation counting. These CCD-based plate readers, such as the LeadSeeker and ViewLux, also have the advantage that they allow the miniaturization of assays to the 1536-well formats [46]. Other than the commonly used SPA and FlashPlates assay formats, traditional filter binding assays continue to be used for specific occasions.

***Chromatography-Based Assays*** *Affinity Selection Chromatography* The use of chromatography to separate the bound and unbound ligands constitutes an intuitively simple method for detection of ligand binding to a large biomolecule. Similar methods have been used previously for detection and characterization of small-molecule binding to serum albumin [47]. This general approach has been utilized and developed for HTS by several laboratories. One embodiment has been a method named SpeedScreen in which size exclusion chromatography (SEC) has been used to separate bound and unbound ligands from pools of small molecules [48]. The compounds bound to the protein of interest are then subjected to LC/MS analysis and identified by their mass. The SpeedScreen method has been automated, and increased throughput has been obtained by parallel processing of the SEC steps. A second method has taken a very similar idea; but instead of using SEC, it uses a size exclusion ultrafitration step such that the protein–ligand complex remains trapped in the filter plate rather than flowing through it [49]. Yet another slightly different approach has been reported [50] in which traditional SEC is used to separate bound and unbound ligands. This approach has not been formatted to allow parallel processing of multiple samples as is possible with the plate-based methods just described for Speedscreen. However, this screening method has been extended to allow screening of integral membrane proteins such as G-protein coupled receptors (GPCRs) [51].

***Immobilization-Based, Label-Free Detection Techniques*** The immobilization-based, label-free techniques have a variety of approaches and strategies: some in which binding is detected in a subsequent step by a labeled antibody or aptamer and some in which binding is detected by changes in the surface properties of the immobilized target. The first type of immobilized assays includes ELISA and ELISA-like methods. These methods require the availability of selective labeled binding partners such as antibodies or oligonucleotide aptamers for the protein of interest (for a review see [52] ).

One original embodiment of surface property-based detection methods was the Biacore instrumentation in which surface plasmon resonance (SPR) technology is used to detect molecular interactions with the immobilized target (reviewed by [52b]). In the last few years, label-free detection systems and technologies for in vitro and cell-based assays have emerged as a new detection paradigm in HTS and related applications [53]. These include the development of evanescent wave technologies that have been adapted to allow the detection of binding at the surface of microtiter plates, such as the Bind system from SRU, Corning's Epic system, and Fortèbio's Octet system. Still other methods have been developed, such as the use of acoustic sensing [the Resonant Acoustic Profiling (RAP) system from AccuBio] or wavelength-interrogated

optical sensing (WIOS, developed by CSEM Scientific). In these detection strategies, the change in surface molecular properties is dependent on the size of the ligand interacting at the immobilization surface [37]. Some of these technologies are already adapted into 384-well plate format for HTS applications.

Another approach that has been reported for monitoring the effect of small-molecule binding to proteins is frontal affinity chromatography (FAC) coupled with MS detection (FAC-MS), first reported by Hage [54] and recently reviewed by Slon-Usakiewicz and coworkers [55]. In this approach, the protein or target of interest is immobilized in the column matrix, and the ligands of interest are infused into the column with buffer and pumped through the column. If a compound has affinity for the immobilized protein, it will tend to be retained on the column, delaying the time it takes for the compound to be eluted. The higher the affinity the compound has for the immobilized protein, the longer it will take before it is eluted from the column. Compounds eluted from the column are detected by mass spectroscopy. In this approach it is possible to screen in mixtures, as long as the component compounds have unique mass-to-charge ratios for mass spectrometric detection, and compounds elute according to their relative affinity. Thus, the relative ranking of ligand affinities in the mixture can be assessed in one experiment. The drawback of this approach, as with all techniques requiring immobilization, is that sufficient protein has to be immobilized onto the column and immobilized in a way that will not inhibit or influence the binding of small molecules.

## 2.2 Cell-Based Assays

Cell-based assays have been important sources of drug candidates from the days of screening natural product extracts for antibacterial activity [56]. However, with the development of molecular biology and advanced methods for cell culture, many more ways to monitor specific biological processes in cultured cells have become available. The use of cell-based assays in drug discovery and in HTS has increased steadily in the past several years. Cell-based screening in which the target activity is directly assessed in its cellular environment can improve the quality and biological relevance of active compounds in comparison to isolated biochemical target-based screening. Various cell-based assays such as reporter gene assays, secondary messenger assays, cell-based ELISA, cell-based proximity assays, and pathway screening assays are all used extensively in screening. Other assays use reporters to monitor the state and location of cellular constituents, by using specific binding reagents, or use indirect technologies that monitor nonspecific aspects of cellular growth, shape, or electrical conductivity.

***Reporter Gene Assays*** Reporter gene assays are assays that monitor changes in the intracellular levels of a particular gene. This method uses specific enzymes or fluorescent proteins whose expression level are linked to a promoter or ribonucleic acid (RNA) sequence that is believed to be regulated by expression of a gene of interest. Initially, this experimental strategy was used in bacterial systems using reporters that could be used in selecting for antibiotic resistance or nonselective reporter systems such as β-galactosidase or chloramphenicol acetyl transferase. Multiple reporter gene and cell lines have been used in antibiotic drug discovery [57] as well as to help define the mechanism of action of compounds [58]. While the development of novel reporter systems for bacterial systems has not been a major area of research recently, technology development for reporter gene assays in mammalian cell systems has continued, with a wide range of technologies now being available. Some of the more commonly used reporter assay systems are highlighted below.

*β-Lactamase* One of the recent improvements in the available reporter gene systems for mammalian cell assays was the development of the bacterial β-lactamase, *bla*, in conjunction with the CCF2-AM dye (a FRET-based fluorescent substrate from Invitrogen), that is, the GeneBlazer system. This cephalosporin-based dye contains a set of acetate ester groups that are removed by intracellular esterases, resulting in intracellular accumulation of the dye. The dye molecule contains a FRET pair allowing ratiometric estimates of *bla* activity. The system has been shown to work with a wide variety of different pharmaceutical targets including GPCRs [59] or to allow detection of antiviral compounds [60].

*Nitroreductase* A novel reporter system based on the use of a bacterial nitroreductase (NTR) has been developed with a cell permeate cyanine fluorescent dye, CytoCy5S, as its substrate. The nitro groups in the substrate can be reduced by the enzyme to generate a fluorescent product that is then retained within the cell [61]. The purported advantages of this reporter are the use of a red-shifted substrate, leading to less interference due to compound and cell autofluorescence. Also, the substrate CytoCy5S is reportedly more photostable in solution and is more soluble in water. However, the biggest difference of this system from the *bla* system may be that the product of the nitroreductase can reportedly be fixed, which can simplify scheduling issues in screening. Assay plates can be read up to 48 h following fixation, thus making NTR compatible with high-content imaging readouts.

*Secreted Alkaline Phosphatase* Another reporter gene commonly used to monitor the activity of promoter constructs is secreted alkaline phosphatase (SEAP). Although such constructs do not allow monitoring of the intracellular location of the proteins, there are a number of very sensitive chemiluminescense-based detection assays for SEAP, allowing very low levels of gene expression to be monitored. The secretion of the reporter gene as well as the stability of the SEAP reporter also allows the accumulation of signal over time to be monitored. Recently, a SEAP reporter-based system has been developed for screening for effectors of the signal peptide-associated protein co-translational translocation process [62].

*Luciferase* The most commonly used reporter gene systems are those based on luciferases. These assays have been widely accepted for a number of reasons, including the sensitivity and large dynamic range of the assay and the availability of photon multiplier tube (PMT)-based readers capable of monitoring luciferase readouts in most biology laboratories. The assay format has also been widely used in the screening community with the development of sensitive CCD camera detection systems that allow rapid reading of assay plates.

Different luciferase reporters can be used for sequential detection of two different promoters, such as the Dual-Glo readouts. Such readouts allow the simultaneous normalization and correction of assay signals from the gene of interest and a control gene. Recently, there have been further developments in this area with the availability of the Chroma-Luc system in which two versions of the click beetle (*Pyrophorus plagiophalam*) luciferase have been cloned into expression vectors. These enzymes are essentially identical except for one amino acid, which causes one version to emit a predominantly green light while the second mutant emits a predominantly blue light. The result is that with the addition of one detection reagent it is possible to monitor both the specific target signal as well as the control signal. However, the luciferase method has the disadvantage that the readout, made after lysing the cells, is a single destructive event. However, some luciferase systems can use substrates that allow nondestructive monitoring of luciferase activity such as coelenterazine for the Renilla luciferase [63] or luciferin with firefly luciferase [64], allowing real-time, noninvasive measurement of luciferase activity. Luciferase can also be used to detect close proximity of protein partners using bioluminescence resonance energy transfer (BRET) [65].

*Enzyme Complementation as Reporters* As mentioned earlier, enzyme complementation has been used to develop cellular assays predominantly monitoring the intracellular location of different protein partners or secondary messengers. Thus, these assays can monitor translocation events causing both fragments of an enzyme to be present in the same cellular location, leading to formation of the active enzyme. This principle has been used with luciferase enzymes for development of sensitive translocation assays [66]. Recently, a novel split-GFP (green fluorescent protein) system that consists of a small (GFP β strand 11; GFP 11) and a complementary large fragment (GFP β strand 1–10; GFP 1–10) has been reported [67]. Neither fragment by itself displays fluorescence, but GFP fluorescence emerges upon self-association of the two complementary fragments. This split-GFP may also be useful for detecting protein–protein interaction in vivo, similar to the luciferase fragment complementation systems.

*Fluorescent or Bioluminescent Proteins* Green fluorescent protein has been widely used as fluorescent tracers and reporters for cellular imaging and for high-content screening. Some GFP derivatives (such as YFP and CFP) can be used in BRET assays [68]. For example, to monitor GPCR activation, GPCRs and β-arrestin are re-engineered to incorporate derivatives of Renilla luciferase (Rluc) and green fluorescent protein (GFP2), respectively. The interaction of the GFP2 arrestin fusion with an agonist-activated Rluc-GPCR fusion brings Rluc and GFP2 into close enough proximity to allow resonance energy transfer to occur [69].

***Cell-Based ELISA*** Chemiluminescence detection methods based on horseradish peroxidase have been used as improvements to standard ELISA-based assays. This detection method has been extended to allow the development of cell-based Western blot assays, also termed Cytoblot assays [70]. In fact, this assay format can be seen as the precursor to a number of newer assay technologies that seek to perform immunodetection of analytes in whole cells, including assay formats such as the SureFire and Licor technologies [71].

***Cellular Imaging Analyses*** Fluorescent-based cell imaging techniques include fluorescent microvolume assay technology (FMAT), fluorometric imaging plate reader (FLIPR), high-content imaging, and some luminescence readers. FMAT has been already mentioned in Section 2.1. FLIPR readers have been widely used for mobilized $Ca^{2+}$ flux measurement elicited by activation of GPCR. The development of nonwash dyes (e.g., Calcium-4) and FlexStation or $FLIPR^{tetra}$ systems have made it possible to miniaturize such assays to 384- or 1536-well formats [72, 73]. The use of voltage-sensitive indicator dyes on the FLIPR also allows screening assays for some ion channels [74]. However, while

showing good correlation with manual patch clamp data, ion channel activity from FLIPR membrane potential assays are sensitive to compound interference. Unfortunately, the chemicals interfering with a particular channel are highly unpredictable (see discussion on more ion channel HTS assays below).

High-throughput flash luminescence readers enable functional GPCR and calcium channel testing. For example, cell flash technology (Lumax Flash HT) uses an intensified CCD camera that is able to detect the signals from all wells of a plate, with a 16-fold parallel noncontact dispenser injecting cells or reagents into the assay wells.

In high-content screening (HCS), automated microscopes take images of cellular models where the proteins of interest are detected because they bear a fluorescent tag, such as the green fluorescent protein (GFP), or they may be identified by fluorescent antibodies. Many HCS instruments are (fluorescence) microscopes that are connected to some form of image analysis package, such as the Arrayscan (Cellomics), Incell 1000/3000 (GE/Amersham Biosciences), OPERA (Evotec Technologies), and ULTRA (Molecular Devices). Alternative instruments, such as the Acumen Explorer or IsoCyte's BlueShift readers, do not take microscope images of the cells, but rather use lasers to scan the bottom of the plate, recording the peaks of fluorescence intensity from which a subsequent image is generated.

The primary advantage of such cell-based assays is that the selection of active molecules is based on a cellular phenotype. Therefore, the screen does not require a priori knowledge of any specific biochemical targets that are being affected by the active compounds. The ability to rapidly perform an HTS campaign based on cellular phenotype may be a benefit for the rapid discovery of novel and unexpected compounds, but this screening paradigm does create the issue of further defining the mechainsm of action of such active compounds.

***High-Throughput Flow Cytometry*** High-throughput flow cytometry has been implemented for HTS applications. The reported HyperCyt system interfaces a flow cytometer and autosampler. As the sampling probe of the autosampler moves from one well to the next of a multiwell microplate, a peristaltic pump sequentially aspirates sample particle suspensions from each well. Between wells, the continuously running pump draws a bubble of air into the sample line. This results in the generation of a tandem series of bubble-separated samples for delivery to the flow cytometer. Likewise, the time-resolved data is collected in a continuous stream, with the accumulated data from all wells of a microplate representing a single data file. As with many screening systems, maximum throughput is achieved when the HyperCyt system is used for endpoint assays. The platform can be applied to general cell-based assays and particle-based multiplexed approaches [75]. The advantages this system offers are the ability to use many preexisting flow cytometry assays/antibodies and the rich multidimensional result sets that have been developed for flow cytometers. The challenges that remain are the need for cells in suspension for the assays, the limited throughput of the system.

***Cell-Based Label-Free Readouts*** Some recently developed cell-based, label-free monitoring systems include electro impedance-based sensors, for example, the CellKey system based on cellular dielectric spectroscopy (CDS) or the RT-CES (Real-Time Cell Electronic Sensing) System (ACEA Biosciences), which enable comprehensive pharmacological evaluation of endogenous cell surface receptors and even nuclear hormone receptors. These sensor systems use noninvasive impedance-based measurements to detect the presence, absence, or change of properties of cells or molecules affecting the electronic properties and ionic passage on the sensor surfaces, allowing label-free, real-time, kinetic measurement of several types of receptors including GPCRs and tyrosine kinase receptors. In a similar manner, there are reports of evanescent wave technologies being used to allow the detection of cellular changes at the surface of microtiter plates [76].

For all of these systems, cells are grown in the individual, sensor-containing wells of the microtiter plates. Changes to the biological status of the cells are measured automatically in real time by these systems. Although these systems have been adapted to 96- or even 384-well plate formats, they have only limited throughput at present. In addition, direct interpretation of the results in terms of the underlying molecular events is still not possible. These assays can be used in target validation and as orthogonal assays for primary HTS hit evaluation.

***Cell Growth*** The simplest cell-based assay is to monitor cell growth and proliferation. While the majority of screening reports have focused on growth inhibition for antibacterial and antifungal activities, such methods have successfully been used to screen for compounds with anticancer activity, as reviewed by Boyd [77]. However, growth inhibition for cytotoxicity assays are usually used as secondary screening assays to characterize compounds for possible off-target effects.

Some advances on these commonly used assays now include systems to monitor multitarget screens. For example antimicrobial screens have been implemented by using the protection afforded by the active compounds to rescue mammalian cell growth [78]. Other

similar systems have been developed using novel model organisms, such as the co-culture systems developed by companies such as Athelas and others [79, 80]. Such assay systems have the potential to not only identify compounds acting to inhibit microbial growth but also compounds that act by interfering with bacterial mechanisms essential for in vivo survival and pathogenesis of bacteria.

***Assays for Ion Channels*** Ion channels are an important family of drug targets. The gold standard for ion channel assays has always been conventional patch clamp experiments. However, due to its low throughput and the high technical skill requirement to conduct such experiments, this method is not suitable for drug screening. Functional HTS for channels and transporters was not available until a few years ago, when calcium or membrane potential dye was applied in FLIPR assays to measure ion channel activities causing intracellular calcium [81] or membrane potential [74] change, respectively. While significant for being the first major true screens for ion channel targets, the shortcomings of the FLIPR-type assay are also obvious: (1) It is considered indirect and therefore many other factors that cause internal calcium or membrane potential change will bring false positives or false negatives; (2) as FLIPR is a fluorescent dye-based assay, a large number of compounds may interfere with the readout and give false readings; and (3) lack of direct control of voltage make such assays less informative and, perhaps, to some channels, less relevant.

The recent breakthrough technologies in automated electrophysiology, such as IonWorks Quattro, PatchXpress (Molecular Devices), and QPatch (Sophion Inc.), have profoundly changed the ion channel research field. These new automated electrophysiology platforms are able to fill the gap between FLIPR type of assays and conventional patch clamp assay, not only by significantly increasing the throughput but also by using more relevant controlled testing conditions. Although the throughput for such technologies is not high enough for primary screening as conventional HTS, it is feasible to run screens of smaller or focused compound libraries for ion channel targets with direct electrophysiology criteria, which maximize potential gain yet minimize the liabilities.

## 3 ASSAY DEVELOPMENT AND VALIDATION FOR HTS

The previous section reviewed a variety of assay types that are commonly employed in HTS. Often there is a wide choice of assay technologies that can be used for design and implementation of any particular assay. An HTS must be properly designed and configured to maximize both screening efficiency and biological relevance. There are some generally accepted principles, or guidelines, for HTS assay development, including the biological relevance of the assay, assay specificity, assay sensitivity and signal dynamic range, linearity of the assay signal, technical ease (which will impact the throughput of the assay for large-scale screening), and related material supply and cost. These are all important factors that one needs to consider before and during assay development. As a reference, many of these considerations have been described in detail in a document made publicly available through the National Institutes of Health (NIH) Chemical Genomics initiative (http://www.ncgc.nih.gov/guidance/manual_toc.html).

### 3.1 Biological Relevance and Validation of Assay Pharmacology

The importance of the biological relevance of an assay used in screening cannot be overemphasized. The scientist developing an assay for HTS should investigate the details of the biological system before attempting to develop an assay for screening. Neglect or reluctance to put time upfront to study the biology and pharmacological aspects can sometimes lead to a catastrophic failure of an HTS campaign. When studying the biological system of the target, one should specifically consider some of the following:

- Target "drug-ability": What evidence is there to support the hypothesis that this target is amenable to modulation by small molecules, that is, "drugable"? Does the target belong to an enzyme or receptor class that has historically been a preferred drug target class or is this a completely novel target class based on new evidence?
- What is already known about the system or target itself? What are the properties of the target or the reaction involved (if the target is an enzyme)? What are the preferred binding partner(s) or substrate(s)? Which isoform or splice variant is more relevant? What are the specificity and selectivity issues? Are there any kinetic and stability issues? One needs also to consider the mode of inhibition (or activation) of interest, and biosafety concerns of running the assay as a HTS, and the like.
- What are the biologically relevant conditions and the environment of the biological system/target? These include buffer, pH, ionic strength, cofactor requirements, substrate concentrations, and other effectors.
- Will the proposed assay reflect the system being interrogated? What could be the biological and

experimental artifacts of the assay? What biological or pharmacologic experiments can be performed to validate the assay? Even factors such as the order of the reagent addition and reaction initiation can make a significant impact in regard to physiological relevance.

- Importantly, what are the alternatives for the primary screening assay? These alternatives need to be considered in selecting the primary assay for the HTS campaign.
- What would be the counterscreens and orthogonal assays to confirm the hits? In addition, secondary assays usually will be needed to further validate or classify the hits from the primary screen.

One typical example regarding biological relevance is the ATP concentration used in in vitro kinase assays. The physiological cellular concentration of ATP is in the low millimolar range. If the inhibitor of interest is non-ATP competitive, the ATP level can be set at a concentration close to the physiological concentration. On the other hand, if ATP competitive inhibitors or both ATP-competitive and noncompetitive inhibitors are sought, then a much lower concentration (at or below the $K_m$) of ATP should be used. Another example is from inhibitor screening for some blood coagulation factors. A desirable inhibitor should have the advantageous trait of fast inhibition kinetics, in addition to the potency and other preferred pharmacological profiles. Therefore, in this case, there is no need to preincubate the target enzyme with the test compounds before introducing the substrate.

### 3.2 Assay Format Selection and Development Considerations

Beyond the consideration of biological relevance, an assay for HTS has other requirements and considerations. Preferred HTS assays not only require robust signal, high throughput, and small assay volume, but also require adequate sensitivity, accuracy, and reproducibility in order to be able to discriminate among a very large number of compounds that span the entire range of activity. An improperly designed HTS assay may fail to detect the desired classes of hits even if it is trivial to execute.

***Assay Format Selection*** As illustrated in the previous section, there are two major assay types in HTS, biochemical and cell-based assays (Fig. 3). The selection of assay format should be driven by the need to develop a readout with high sensitivity, reliability, and biological predictiveness. For speed, ease of use, and reduction in data variability, assay procedures with as few steps as possible are preferred. It is desirable to use the lowest possible biological target concentration in an assay. Not only will this save reagent, but also the assay will be sensitive and will be able to discriminative among the most potent hits in an activity series. Of course, one has to strike a balance between assay sensitivity versus signal strength (or signal robustness), which will affect the reliability of the measurements and the overall data quality. For any specific target, there are usually several assay formats available to choose, from isolated in vitro biochemical assays to cell-based assays, and from heterogeneous formats to homogeneous formats. All being equal, homogeneous assays are preferred over heterogeneous assays, and nonradioactive assays are preferred over radioactive assays. Some homogeneous assays such as TR-FRET (HTRF), FP, AlphaScreen, SPA, and chemiluminescence (e.g., HitHunter cAMP detection) are popular formats in HTS for both in vitro assays and cell-based assays. For cell-based assays, several common assay formats are reporter genes (luciferase, etc.), FLIPR, Cytoblot, and high-content imaging. Some newer assay formats based on label-free detection (waveguide sensing, cell electronic impedance, LC/MS) emerge as newly preferred assay formats.

***Assay Optimization*** Some general considerations in assay development and optimization are given here. For biochemical binding assays, the dissociation constant ($K_d$) and maximum binding ($B_{\max}$) should be known or determined. Specific binding versus nonspecific binding will determine the specific assay signal strength, data reliability, and quality. Any nonspecific signal should be keep as low as possible, and the specific signal should be minimally several-fold larger than that of the nonspecific signal. Characterization of the binding kinetics (rate constants $k_{\text{on}}$, $k_{\text{off}}$) is also important for equilibrium or nonequilibrium binding behavior. Ligand binding as a function of receptor concentration should be in the linear or pseudolinear range.

For enzymatic assays, several excellent reviews [82, 83] detail the strategies and tactics used during assay development. The principles of enzymology are always a useful guide during assay development [84]. Briefly, an enzyme titration should be conducted to determine the minimum enzyme concentration that can be used, the $K_m$ values of substrates, and the turnover rate; all of these should also be assessed to ensure the linearity of the reaction kinetics. These parameters must be determined experimentally. For HTS assays, many researchers prefer a linear time window of 30–120 min to facilitate plate handling and throughput, while kinetic readings can also be implemented for some fast reactions. As will be discussed, the pharmacology of reference compounds should be used to validate the assay. The substrate con-

centration is preferably chosen to be at or below the $K_m$ value to ensure a sensitive and linear signal response. Sometimes, in order to ensure measurement accuracy and to generate a robust signal, substrate concentrations [S] at two or three times the $K_m$ is a common compromise. To exclude competitive inhibitors, the substrate can be set at $[S] > 10K_m$, or at as high a concentration as the substrate solubility permits, for example, using millimolar concentration of ATP in screening for non-ATP competitive inhibitors.

When an assay signal is very weak under conditions that still lie in the linear range of substrate titration, either substrate concentration or conversion of the substrate need to be increased to enhance the signal-to-background ratio. It is generally accepted that the conversion of product should be kept at <10% of the total substrate used when studying enzyme kinetics. When the substrate is poorly soluble, the researcher is left with only one choice, that is, performing the assay at >10% product conversion in order to enhance the robustness of the assay (as monitored using the $Z'$ factor, which will be discussed later). However, at >10% conversion, the observed inhibition will be influenced. Wu et al. [85] have described the relationship between the half-maximal inhibitory concentration ($IC_{50}$) value for a given inhibitor and the percentage of substrate conversion using a first-order kinetic model under conditions that obey Henri–Michaelis–Menten kinetics. The study concluded that the $IC_{50}$ and percent inhibition values are only slightly shifted to less potent values even with much higher substrate conversion (up to ~40% conversion to product). This analysis provides a sound rationale for using higher substrate-to-product conversion as a compromise in optimizing an enzymatic assay for the needs for robust signal and sensitivity to inhibitors.

Other components in the reaction buffer also play important roles in assay optimization. Cofactors or prosthetic groups of the enzyme, cationic or anionic ions, pH, detergents, carrier proteins, and so forth all need to be tested during assay development and optimization. In many cases, the optimization can be performed by varying one or two relevant factors at a time. Sometimes it may be more efficient to use the design of experiment (DOE) approach to vary more factors simultaneously [86]. Regardless of the use of the one-at-a-time optimization approach or the DOE optimization approach, preknowledge of the factors pertinent to assay signal is extremely useful. However, it should be noted that buffer optimization is only an intermediate step; once the final buffer conditions have been defined, the assay must be again checked to assure that appropriate reaction kinetics are occurring.

Assay development of cell-based assays offers a stark contrast to that of mechanism-based biochemical assays. There are many discussions in the literature and review articles on the pros and cons of cell-based assay versus biochemical assays. We do not intend to reiterate those here [87, 88]. The fundamental considerations in cell-based assay design are similar: that the signal is specific and linear with the ligand and cell number (or protein/receptor level within the cell). The cell/receptor and time response linearity are key in assay optimization. Sensitivity, robustness of the signal, and reliability of the reading are important assay quality indicators. False positives and false negatives due to other cellular events (toxicity, expression instability, etc.) are more important issues to be taken into account. For example, the use of dual luciferase reporters (e.g., firefly and Renilla luciferases) can discriminate real inhibition from cell toxicity and product instability. Even with such normalization, the correct interpretation of the signal can be confounded by nonspecific effects on the reporter gene promoter or on cell proliferation [89].

***Assay Pharmacological Validation*** Whenever possible, the assay system under development needs to be validated by some appropriate biological or pharmacological test. There are a number of ways to accomplish this. Typical validation experiments include the use of some known reference inhibitors, the nonlabeled substrate as a competitor to the labeled one, or some known active and inactive effectors, as appropriate. If some known inhibitors are used, the pharmacological profiles (potency and efficacy) should be the same or very similar to those previously reported. Usually, $IC_{50}$ values, or better yet, $K_i$ values within three-fold from well-documented values are usually acceptable in HTS assays when performed under the same or very similar conditions. The rank order of potency of some known activators/inhibitors should also be identical or at least very similar to the reported potency rank order [90]. For instance, in the development of a particular kinase assay, if known compounds such as staurosporine have been tested in other laboratories or reported previously for the enzyme, the HTS assay should yield similar or comparable results to assure the assay performance and its hit finding capability. This should be true in both biochemical and cell-based assay development. Assay quality assessment and statistical validation will be discussed in the next section.

### 3.3 Assay Data Quality and Statistical Validation

In most HTS programs, each compound is only evaluated in a single test during primary screening. A high degree of accuracy and sensitivity in the assay is therefore critical for identifying active compounds (or "hits"). Due to various unavoidable sources of error or

variability, the measurements from any assay will contain a certain degree of variability. Yet hits need to be identified in the presence of such variations in signal measurement. Reducing the measurement variation will provide higher fidelity to the screening data. Therefore, in the design and validation of HTS assays, assessment of the screening data variability, by metrics such as the standard deviation (SD) and coefficient of variation (CV) of the replicate data set, is critical in determining whether an assay can identify hits with high confidence.

***Z′ Factor in Screening Assay Quality Assessment*** The quality of an assay can be loosely expressed as signal-to-noise (S/N) ratio and/or signal-to-background (S/B) ratio. Both ratios reflect the assay signal strength and therefore are useful parameters in assay quality assessment. In HTS, compound activity is normally expressed as a percentage measured against some selected reference control signals, usually a high reference control and a low reference control. However, both S/N and S/B ratios fail to take into account information from data variability of the sample and reference controls, as well as the assay dynamic range at the same time. To overcome some of the pitfalls of S/N or S/B ratios, Zhang et al. [91] proposed a simple statistic, the *Z* factor (*Z′* factor) as defined below, for use in evaluation of assay performance and assay signal robustness in HTS.

The data quality of an assay can be estimated by the *Z′*-factor formula:

$$Z' = 1 - \frac{3\sigma_{c+} + 3\sigma_{c-}}{|\mu_{c+} - \mu_{c-}|}$$

where ($\sigma_{c+}$) and ($\sigma_{c-}$) are the standard deviations for data from the high reference controls and low reference controls, respectively, and $|\mu_{c+} - \mu_{c-}|$ is the absolute value of the difference of the two control signal means. The *Z′* factor defines the *usable* dynamic range (usually the linear range) of the assay. Thus, the *Z′* factor contains data variability information of both the high and low reference controls (the same high and low reference controls usually used for normalization of the measured raw data to a percentage activity) as well as the dynamic range information. Therefore, compared to CV, S/B, or S/N, the *Z′* factor is more appropriate for evaluating overall assay quality. It is a useful tool for the evaluation, comparison, and validation of any assays, in general, and has been widely used for quality control (QC) of HTS assay quality and suitability.

As originally proposed, *Z* factor refers to the specific *Z′* factor when the mean and standard deviation of one of the corresponding reference controls are replaced by those of the testing samples. Therefore, *Z* factor reflects the test sample in screening and *Z′* factor is for the assay. *Z* factor has been used as a screening quality assurance parameter for assay plates owing to its sensitivity toward signal strength and data variability of test samples on the plate as well as the hit rates on the plate. *Z* factor is also affected by the screening conditions such as the compound concentration.

Generally, a *Z′*-factor value of close to 1 is an ideal assay for HTS, $1 > Z' > 0.5$ is considered a good-quality assay; $0.5 > Z' > 0$ can be a moderate to a poor-quality assay; and assays with $Z' < 0$ cannot really be used in HTS. These provide, however, only a rough categorization of assay quality for HTS and should not be regarded as absolute rules. Also, this metric applies to single-point screening. When multiple (*n*) repeat measurements of a test sample are obtained in a screening, standard error (SE) should be used instead of standard deviation; therefore, assay quality can be improved accordingly by square root of *n*.

There are several limitations in using the *Z′*-factor statistics in HTS. First, *Z′* factor should be assessed using large data sets, and both high and low reference control signals should approximate a normal distribution. In some rare cases where either the data distributions deviate severely from normal distribution or ($\sigma_{c+}$) and ($\sigma_{c-}$) values are hugely different, the *Z′*-factor values can become less effective or even meaningless for indicating the assay data quality. Another limitation is that, due to its dependence on two reference control states of the assay, the *Z′* factor becomes a less defined measure of assay quality for stimulator or agonist screens if there are no known stimulator/agonist reference controls available. When a strong or natural agonist/stimulator is known, then the *Z′* factor can be rightfully defined.

***Other Assay Data Quality Indicators*** Besides the commonly used *Z′* factor, and S/B and S/N ratios, there are a few other assay performance statistical parameters used in the HTS literature. A "signal window" (SW) ratio had been reported to evaluate the performance of HTS assays [92]. The so-called assay variation ratio (AVR) was also used to evaluate assay quality [93]. AVR and *Z′* factor are interrelated and interchangeable with one another, except that the originally recommended values for the acceptable and unacceptable assay for screening were somewhat discrepant. Iversen et al. [94] conducted a simulation study and compared the performance of *Z′* factor with SW or AVR, and recommended *Z′* factor as a preferred assay performance measure. Many assay developers use *Z′* factor in conjunction with CV, S/B, or S/N to give a more comprehensive assessment of the assay variation and performance.

Several groups of researchers have also advocated the use of power analysis as an assay quality indicator

[95, 96]. While this approach is certainly statistically sound, it should be noted that power analysis is usually based on one reference state of the assay signal distribution, and, therefore, whether it will reflect the overall data quality of the assay over the entire dynamic range is an open question. Compared to power analysis, $Z'$ factor is a much simpler statistical parameter to use and can be used routinely during a screening campaign to track the assays quality.

***Detection Sensitivity of an Assay*** As with measures of HTS assay quality, there is more than one way to assess assay sensitivity. First, the assay sensitivity can be evaluated by the lower limit of detection (LLD) of the assay measurement. The LLD as assay sensitivity limits are evaluated in terms of the S/N ratio. The S/N ratio is a good indicator for strength of signal. An analyte concentration giving an S/N = 3 is conventionally regarded as the LLD of the assay measurement with regard to that analyte. Similarly, one can use an analyte concentration that gives $Z' = (-1)$ as LLD of the assay [91]. In HTS, an assay sensitivity can also be evaluated by the limit of hit potency detectable with the assay. For example, when seeking inhibitors of a particular enzyme, what is the potency limit of inhibitors that likely will be detected by the enzyme inhibition assay when screened at a certain compound concentration? In assessing the sensitivity of an assay with regard to detectable compound potency, Buxser and Vroegop [95] used Michaelis–Menten kinetics and Langmuir binding isotherms, combined with fundamental statistical principles, to deduce a measure of assay sensitivity for inhibitors. This approach takes into account the compound screening concentration, the data variability of the assay, the $K_m$ or $K_d$ of the reaction, and the substrate or ligand concentration used. Then, the threshold or cut-off values used for determining "hits" and their probabilities of being detected by the assay can be evaluated.

***Data Repeatability and Reproducibility*** Another aspect of HTS quality is data repeatability and reproducibility. Here, repeatability refers to the ability of the same assay to be "cloned" and run in multiple copies with identical or similar results in one experiment at one time. Reproducibility describes the same assay run at a different time or in different laboratories. Since each HTS campaign is a considerably large effort and most screens are only performed in a single test, the assay data repeatability and reproducibility thus become an important part of the assay validation process. Traditionally, and in our labs routinely, after the assay has been developed and optimized, several replicates of assay plates containing high and low reference controls or other selected known reference compounds are tested with the HTS instrumentation. The means (or medians) and standard deviations of the high and low reference data sets are calculated. Then, the corresponding CV, S/B, and $Z'$ factor are obtained from these test results. These data uniformity tests may be repeated (reproduced) at different times with independently prepared reagents. The same sets of statistical parameters are obtained and compared. The data from these rigorous uniformity tests give more reliable evaluation of assay data quality ($Z'$, CV, S/B, etc.) than from the tests performed at the bench-top with fewer data replicates.

In addition to assay quality, plate data uniformity, and reproducibility assessment, it is a good idea to run some further validation tests before the real large-scale screening starts. These include blinded tests for some known reference compound controls, or a small-scale pilot (initial) screening. In a blinded test, the known reference compounds (e.g., known inhibitors) are spiked randomly into some wells on the assay plates in several dosing levels or in a titration series. These blinded testing and initial screening results can be used to further assure the assay's hit-finding capability and reliability.

### 3.4 HTS Compatibility

A distinct difference between a general bioassay and an assay for use in HTS is that the latter must be suitable for a high-throughput process. Many assays which routinely produce acceptable or good results with reliable quality at the bench-top scale may not be suitable for HTS simply because the throughput is too low or too complex and cumbersome for automation; or the assay may simply be too costly for large-scale screening. Therefore, compared to a general bioassay, an HTS assay has more requirements and constraints. These requirements include throughput, ease of running and automation, cost-effectiveness, and the availability of a reliable reagent supply. These properties can be collectively referred to as the assay's HTS compatibility. One popular assay type is homogeneous (also called "mix and read") assay as it only involves liquid addition, some incubation period, and then reading of the signal directly. Many types of assays can be configured in a homogeneous assay format, while others may need to adopt heterogeneous formats, as described previously in Section 2.

## 4 SCREEN DESIGN, PROCESS, AND INSTRUMENTATION

If one addresses the assay biological relevance, assay quality, and HTS compatibility and feasibility, there

should not be any major hurdles at the screening implementation and execution stage. Execution of a successful HTS campaign needs careful planning and process optimization to ensure that the data generated from the screening is reliable and consistently of good quality. Conceptually, the HTS implementation has several phases as described below.

### 4.1 Screening Design and Logistics

No matter whether the assay is a biochemical or cell-based assay, after successful development and validation of the assay, the next task is to implement it for HTS. The HTS process can be quite different depending on assay type and facility. Specifically, the HTS implementation will be influenced by assay protocol, resources (personnel and instrument) availability, type of instruments and capacity, library size, and the means for data capture and data flow logistics. The HTS will also be influenced by other process issues such as scheduling (time and space), reagent supply, waste, project time line, and costs. A screening process flowchart with detailed assay protocol and instrumentation will always be helpful in designing and planning logistics for the process. For example, the assay type will usually affect the screening design and process profoundly. For homogeneous biochemical assays, the screen is usually simple to design and implement. For heterogeneous assays, more washing or separation steps may need to be accounted for. The instrumentation required for carrying out the specific assay must be carefully chosen and calibrated, and the whole process must be carefully validated. Furthermore, if the assay is to use a radioactive material or other potentially hazardous reagents, operational safety and waste management issues have to be worked out according to regulations.

***Plate Format and Plate Type*** Perhaps the major difference between the current HTS process and the low-throughout screening processes used before the arrival of HTS is the use of standardized microtiter plates or multiwell plates. The use of standardized microplates has not only enabled many different assays to be performed at high density and speed, but they have helped to standardize HTS instruments (liquid dispensers, detectors, etc.). The result of this standardization of instruments is that they can be used interchangeably or in a combined manner. A standard 96-well plate contains an $8 \times 12$ array of wells each typically capable of holding a volume of 200–300 μL. A 384-well microplate contains a $16 \times 24$ array of wells each typically being able to hold a maximal volume of ~120 μL. Low-volume 384-well plates have been developed that can hold volumes of ~25 μL per well. With the development of low-volume liquid handlers and readers, higher density plates such as 1536-well plates ($32 \times 48$ well array, assay volume 5–10 μL) are increasingly being used. Reagent savings in these low-volume microplate formats are substantial, besides the enhanced throughput due to reductions in plate manipulations.

Sporadically, other assays using plate formats such as 864-, 3456-, and even 9600-well plates have been reported [97]. These formats have not gained much popularity, presumably due to the lack of compatible liquid handlers and readers, as well as the associated difficulty of compound reformatting. The need for assay miniaturization is driven by needs for increasing throughput and reducing reagent usage and cost [98]. Other non-plate-based screening formats have been reported, but so far, they have not gained as wide acceptance in the screening community (for a review, see [99]). Beyond compound screening, some assay chip technologies are now used in genomic and gene expression screening. With the advances of screening technologies, such as noncontact picoliter dispensers, liquid spotters, and MS-based or imaging-based detectors, a convergence of compound and genetic screening formats and methodologies appears possible in the near future.

The majority of the microplates used in HTS are made of either polystyrene or polypropylene. Most 384- and 96-well plates are available with well shapes of circular, squared, or squared with rounded corners and in conical, flat, and round well bottoms. Clear, white, or opaque plates for well walls or bottoms are available for different detection purposes. Plates can be pretreated or coated with special materials for specific use, such as plates treated for cell culture, high/low protein binding, or FlashPlate for in-plate scintillation detection. Some of the new technologies in multiplexing, label-free detection, and acoustic energy based liquid transfer requires special plates with requirements such as protein-arrayed bottom, electrocircuited bottom, optical sensing bottom, and cyclic olefin copolymer (COC) plates for compatibility to the corresponding technologies.

***Plate Layout*** The positioning of test samples and reference control wells on the assay plate can be quite flexible and is dependent on the experimenter and affected by liquid handling process. From a statistical point of view, the controls and test samples should be placed in such a way that systematic errors can be minimized. Usually, statisticians suggest a "mixed" or randomized layout to minimize systematic errors. This would help to avoid systematic errors that can occur when controls are placed along a column, usually at one edge of the plate. Such a nonrandom layout may over-

estimate or underestimate the control mean if there is an edge effect of the assay on the plate.

In practice, however, there are several constraints in selection of the position and layout of the samples and controls. The mostly encountered constraint on the layout of the assay plate is that imposed by the layout of the compound storage plate, with the only flexibility available in the placing of the high and low reference controls in certain columns or rows. For example, in some common layouts, columns 23 and 24 of a 384-well plate may be the only 2 columns available for placing controls. Alternatively, columns 12 and 24 may be available for the controls. But there are still many possible ways of placing high/low controls and other reference/standards in these control columns. Several routinely used layouts are either one column for high control and another for low controls or half of a column for high and half for low controls in both columns. The selection of the control layout is often dictated by the convenience with which liquid handlers can add the high/low controls into the reference wells.

Another approach that has been taken by some laboratories is to eliminate control wells on each plate and instead to have a few "control plates" added at regular intervals between assay plates throughout a screening run. Such an approach has the advantage that systematic effects on the plates (such as center or edge effects) can be more easily detected. Yet other laboratories may choose to use plate median as the "inactive controls," given the majority of the compounds on the assay plate being inactive.

### 4.2 Instrumentation and Automation

***Instrumentation for Liquid Handling and Signal Detection*** The demand for higher throughput has driven the advancement of HTS instrumentation and also higher density microplates. Some of the earlier instruments for multichannel liquid handling in HTS labs were automated or semiautomated systems and workstations from Zymark, TomTeck, Beckman Coulter, Tecan, Robocon (CRS), CyBio, Matrix, TiterTek, for example. In recent years, many more advanced systems have become available, most notably from Perkin-Elmer, Velocity 11, Thermo Electron, and so forth, in addition to the aforementioned vendors. Meanwhile, several new picoliter to nanoliter liquid dispensing technologies have matured and are being increasingly used in HTS. Examples are noncontact liquid dispensing devices such as acoustic dispensing devices.

Many new signal detection methodologies and instrumentation have also become available in recent years, ranging from general multifunctional fluorescent and optical detectors to highly specialized plate readers (e.g., MesoScale) and instruments for high-content imaging (such as ArrayScan-II, In-Cell 3000, etc.). Label-free detection has shown much promise for new screening paradigms. Some examples are LC/MS-based screening, surface sensors, and cellular dielectric impedance changes. As long as the pressure for increasing drug discovery productivity continues, the use of advanced instrumentation and automation will continue to increase in HTS laboratories.

***Workstation versus Fully Integrated Systems*** Every HTS laboratory is designed based on its role within the drug discovery process and the organization it serves. The challenge for each screening laboratory is to balance the required throughput against expenditure, complexity, and the changing needs of their organization. As mentioned above, a wide range of microplate-based liquid handling and signal detection equipment are available in the marketplace. Some laboratories use full robotically controlled systems while others utilize standalone workstations. Each of these approaches for screening has its advantages and pitfalls.

Workstation mode has the clear advantages of simplicity, flexibility, low cost, and easy setup. In a modular workstation-based HTS lab, screens usually require microplates to be processed through several different instruments, each of which is programmed to perform a single task. Each instrument module functions independently from other modules and can be interconnected using simple robotic arms. Liquid handlers and readers are usually equipped with plate stackers or simple robotic arms. For instance, compound plates preparation, assay reagent dispensing, washing, and signal detection are carried out separately according to the assay protocol. However, with this process people have to be involved with executing each separate task in a screen. Human errors are therefore more likely to arise. Plates are usually processed in batch mode, sometimes leading to more pronounced batch-to-batch variation. The throughput of workstation screening is limited by the working hours of the people involved. With innovation on ultra-HTS and miniaturization, throughput issues are being addressed, and specialized workstations or semiautomated setups may be suitable options.

Modern fully integrated screening systems allow rapid processing of screening plates continuously throughout the day in an operator-independent or nearly independent manner. Even with fully integrated systems, operators are still required to "feed the machine" and to remove its waste as well as review and analyze the screening results it produces. Key elements of a typical fully integrated robotic screening system

include a working platform, plate-handling robotics (arm robots, cylindrical coordinate robots, or overhead Gantry robots) with intelligent plate gripers, linear servotrack, controller, the integrated barcode scanners, liquid-handling units (dispensers, aspirators, washing station, plate centrifuge, mixers/shakers), incubators, plate storage units for compound source plates, and/or assay plates. A software operating system with programming capabilities for scheduling of different tasks, error detection, warning, and fault rectification is also an indispensable component of the system. Some of the large HTS systems integrate online data analysis/visualization to allow quality control and database interfacing. The demand for increasing throughput requires that these systems perform reliably and can accommodate a variety of screening assays with less failures.

For fully integrated robotic systems, the major advantages includes alleviating people from tedious labor and repetitive work, allowing them to conduct "higher level" intellectual tasks. The robotic system can work for long hours with less human involvement. The data generated from a screen is generally in a strictly controlled manner and therefore usually more uniform in quality. The major disadvantage of fully integrated robotic systems are the long design and implementation times required, and the overall and upfront costs of fully automated screening systems relative to the workstation-based setups. In addition, there are additional costs associated with hiring specialist engineers for troubleshooting, maintenance, operating, and training others. These additional support labors (and costs) for fully automated robotic systems can offset the labor savings for screening personnel. Another disadvantage is that often more stringent biological reagent stability is required for implementing an assay on a fully automated system [100], potentiating the inflexibility of the systems. Sometimes it takes a longer time for troubleshooting and optimizing an assay on such a system than the real productive screening time on the robot.

Arguably, subjective matters such as different philosophies on managing HTS largely determine the instrumentation setup and thus affect the screening process design. In many cases, the sight of many robotic arms moving in a laboratory gives a strong visual impression of productivity and frequently creates inflated expectations in throughput [101]. The true challenge and measure of success of a robotic screening system, besides expense evaluation, is the time spent on productive use in generating quality data versus the time spent on repairing broken parts and retooling the system to accommodate the diverse application needs in HTS. Again, the best system may be flexible use of people and machines, depending on the diversity of the task and the screening needs.

### 4.3 Screening Process and Quality Control

***Pilot and Primary Screening*** The screening process involves several steps such as pilot run, primary screening, primary hit confirmation and validation ($IC_{50}$ determination and counterscreen), and associated data analysis. The first stage in the screening process is usually an initial, or pilot, screen. During a pilot screen, a subset of compound plates, for example, a small randomly selected compound plate set, a preselected diversity set, or a specific set of test compounds, such as the LOPAC set from Sigma-Aldrich, are used, and the assay plates are processed through the entire screening process, using the same liquid-handling and detection instrumentation and data-handling and analysis systems. The primary goal of such pilot screens is to validate the anticipated screening process. The pilot screen is useful not only in revealing some of the logistic issues the process may encounter, and so allow tuning of the process, but also in evaluating the test compound concentration, allowing an estimate of the hit rate for the full screening. Pilot screens also provide the first set of data to go through the data processing and analysis work flow. If everything looks normal or as expected, the full-fledged HTS campaign starts.

Regardless of whether the screen is performed with workstations or with a fully automated system, procedures should be kept consistent throughout the entire screening campaign. Additionally, as data is generated during the primary screening, it should be constantly monitored with suitable quality control measures. If problems arise during the screen, it will be necessary to pause the process for troubleshooting, before screening is resumed. Quality control and data analysis of screening results will be described and discussed in the following sections.

***Screening Quality Control (QC)*** In modern drug discovery, screening is an industrial-scale process used to survey hundreds of thousands to millions of compounds on selected biological targets. As described above, the process is dependent on a suitable bioassay with sophisticated instrumentation, automation, and various detection technologies. Many factors beyond the biology of the assay influence the quality and outcome of the screening. Thus, quality control is an integral part of the process, used to ensure good-quality data generated from screening.

Various statistical QC methods have been used in HTS. QC can be performed in an online or offline manner, depending on the setup of the HTS data acqui-

sition system. For workstation-based screening processes, QC is most likely carried out offline. For integrated HTS systems, on the other hand, if equipped with automated online QC software, a variety of readily interpretable graphs and tables of key QC parameters allow the operator to visualize HTS data quality and examine assay performance during the HTS process and to react to quality problems in a timely manner [102]. In most cases, QC in HTS is often focused on a limited set of control parameters such as S/B, S/N, $Z'$ factor, and $Z$ factor. Sometimes, more sophisticated QC analyses are used to reveal any systematic errors, patterns, and distribution deviations [103–107].

***Impact of Assay Technology Selection and Screening Efficiency*** It is fundamentally and pragmatically important to determine if a similar collection of active compounds are identified from the same chemical library using different assay technologies. Currently, there are usually many choices of assay formats available for various types of targets when developing an HTS assay. Results from several recent studies reveal that different assay and detection technologies can generate significantly different hit lists for the same biochemical interaction [19, 108–110]. For example, Wu et al. [19] have evaluated the same biochemical assay in a typical HTS setting with two different detection technologies. The results showed that the selection of assay technologies can make a significant difference in the screening results. However, there are other examples where the screening results, using different technologies, have been shown to be quite similar [111, 112]. The selection of assay detection technologies deserves careful justification and validation to maximize the hit-finding probability.

The issue of missing truly active compounds and collecting false positives from noisy screening data pose a serious problem for many HTS programs. Studies have begun to assess the intrinsic fidelity of primary screening results. Several approaches have been attempted to address this issue, from statistic models [113–115] to real case studies [116]. Usually, typical screens with reasonable activity cutoffs for selecting actives should have hit confirmation rates of 50% or above, and rates in the range of 60–80% are frequently seen. The hit confirmation, false-positive, and false-negative rates from a primary screen are mainly affected by (a) the HTS assay quality, (b) the hit limit selection, and (c) the chemical library used. Among these parameters, assay quality plays the primary role in determining the hit confirmation rate. However, since screening is a multistep process and errors can rise at any step in the process, many factors beyond the biology of the assay influence the quality and outcome of the screening process [105, 106].

### 4.4 Strategies of Compound Screening and Compound Handling

***Singleton versus Mixture Screening*** Compound samples, in most HTS campaigns, are only tested in singleton (each test well contains only one compound) at a single concentration. Synthetic compounds are usually created as single pure compounds by medicinal chemistry laboratories. However, some compounds, often originating from combinatorial synthesis, for example, split-and-pool synthesis, are only available as compound mixtures. In the latter case, active wells containing the compound mixture need to be deconvoluted in order to identify the actual active compound. In a similar manner, many natural product samples are only tested as crude extracts or semipurified fractions derived from different microorganisms or biological sources. Again, in this case samples identified as active by screening need to be further purified and analyzed to identify the active chemical structures. Natural products, as isolated and purified samples, can be treated the same way as pure synthetic compounds. The follow-up study of natural product hits is often more complicated and lengthy than for synthetic compounds. Nonetheless, natural products have proven to be valuable sources for active drug candidates for immunosuppressives, cholesterol-reducing agents, anticancer drugs, and antibiotics. In fact, most of the antibiotics currently being developed in the United States have been derived from natural products [117].

It is also worth mentioning one of the "old" strategies that has been applied to HTS: the strategy of compound matrix pooling [118, 119]. In this compound compression strategy, compound plates are pooled in orthogonal matrices (both horizontally and vertically pooled) across a small set of plates as mixtures, for example, mixture of 10 compounds. The result is that each compound is present twice in the pooled library, in two different mixtures. Therefore, compounds are tested essentially in duplicate, but the overall throughput still increases fivefold in the case of 10 compounds per mixture compared to testing single compounds. The primary hits can then be deduced by a self-deconvolution process, comparing the compounds present in the two different active samples. As long as the primary screening hit rate is not very high (e.g., <1%) and the assay signal is sufficiently robust, this strategy seems to work well. A recent comparison of this matrix pooling strategy with the singleton screening method yielded good concordance and comparability in hit finding [120]. This strategy was also used successfully in large-scale LC/MS-based functional screening to enhance throughput [34]. In other assay formats, particularly in affinity-based screening, larger compound mixtures of several

hundreds of compounds or more have been used [48, 50]. Some disadvantages commonly raised for mixture strategies are that compound mixing may show the so-called bystander effect (biological activity of the mixture may be affected by some compound interactions or interferences in the mixture) and that it may not work well for cell-based assays due to higher risk of compound interference and toxicity from the pooled mixtures. Therefore, this strategy is usually not recommended for cell-based assays. However, to our knowledge there is no published data to support or refute this assumption.

***Multiple Testing and Other Screening Strategies*** In addition to screening in singleton or mixtures, other compound screening strategies have been used in various situations. For example, when an assay quality is not sufficiently high, and the assay reagent and throughput are not limiting factors, the screening can be performed in duplicate to enhance hit-finding probability. Another common strategy is to use a two-tier hit selection in primary screening and hit confirmation/validation. Both strategies can be efficient in reducing false hits when applied appropriately. Some of the pros and cons of these approaches have been discussed previously [116].

When reagent supply and screening capacity is not a constraint, one can directly test each compound in multiple doses to eliminate the potential problems of data uncertainty with a single-dose test in routine screening campaigns. Inglese et al. [121] have recently reported a multidose compound screening approach, and called it "quantitative HTS" (qHTS). A similar strategy has been also used for some time by the National Cancer Institute (NCI) screening group [122]. A dose–response curve and $IC_{50}$ values can be directly deduced for each compound screened. This approach has the potential to reduce false-positive or false-negative hits. However, despite these advantages, it is still not clear if this approach is worthwhile in most cases, particularly when the hit rate is low, as is typical for a random compound library screen.

***Counterscreen*** All assay detection methods will generate some artifacts due to interference from compounds, resulting in false-positive hits in the primary hit lists. Different assay readouts have a tendency to generate different artifacts. Counterscreens have to be designed and used to eliminate most of those detection artifacts and nonspecific hits from the primary biological assay. For example, in a luciferase-based reporter gene assay, an inhibitor can either inhibit the specific activation of interest or reduce the signal by nonspecifically killing the cell, or even simply by inhibiting luciferase enzymatic activity. In this case, a counterscreen can be configured by employing a constitutively activated luciferase in the same parental cell line to weed out the toxic and nonspecific inhibitors (i.e., false positives) of the assay. An alternative strategy might be to use a different reporter gene with a different readout of activity, for example, β-lactamase activity or a dual luciferase system. It is noteworthy that a counterscreen in this sense is different from other orthogonal, secondary assays or a selectivity test for further characterization of the hits. A counterscreen may have slightly different meaning in different laboratories. It is also highly recommended to run orthogonal assays to further confirm the hits, given these assay technologies are well-validated themselves.

***Compound Diversity*** The selection of the compounds to be tested during a screening campaign is one of the most important determinants for the success of HTS, regardless of whether it is for a large diversity-based or an information-based (focused) compound library. Even when there is no prior knowledge about a target to guide compound selection, there is still a need to select the appropriate compounds for general screening. There has been much debate, but little real data, concerning the question of how best to design libraries for exploratory screening. One school of thought contends that libraries should be as diverse as possible, including as many molecular features as possible [123]. Others have suggested "informative libraries," where the compounds included in the test set are chosen because they allow the testing of concepts of how a compound may interact with a target [124]. Yet others have suggested that just blindly screening a random selection of compounds is just as appropriate a strategy as any other selection method in finding novel structures [125].

The discussion on the appropriate choice of compounds to test has been greatly influenced by the work of Lipinski and co-workers who retrospectively looked at molecular and chemical-physical properties of drug molecules and their effect on oral availability [126]. This work lead to the Lipinski "rule of five" as a general rule that is widely being used by chemists in the design of "druglike" small molecules for drug discovery. This ground-breaking work has lead to other researchers looking at additional molecular features, for example, number of rotatable bonds, as additional guides to compound selection and design. This concept of druglike molecules has been taken further by Proudfoot [127] with his retrospective analysis of lead to drug compounds, which showed that, often during the lead optimization process, compounds increase in molecular weight and lipophilicity, leading to the concept that screening libraries should not be druglike but rather

leadlike. This approach has been taken further in the context of fragment-based screening strategies and has lead to the "rule of three," which has been combined with the concepts of the ligand-binding efficiency in compound design [128].

***Sublibrary Screening (Focused and Small Diverse Libraries)*** Lead identification by high-throughput screening of large, randomly selected compound libraries is supplemented with strategies utilizing virtual and focused screening approaches. Such approaches attempt to reduce the number of compounds tested and thus to increase the efficiency of screening, by using information already known about the target or the compounds that interact with it. This is in contrast to the strategy of "screening everything," which is based on the assumption that testing large, random compound libraries is an efficient way of identifying novel leads. However, increasingly, there is already information about many of the targets being selected for drug discovery projects, making such rational strategies more attractive. Currently, there seems to be a trend toward the use of structure-based focused libraries [129–131]. Recent advances in high-throughput computational docking, nuclear magnetic resonance (NMR), and crystallography also have facilitated the application of focused library screening. For example, focused libraries have been used to help generate hits or lead compounds for kinases and other targets [132–135]. The focused libraries can be populated by either high-throughput cherrypicking or by combinatorial or focused synthesis [136, 137].

Generally, the scaffold diversity of a focused library is relatively low. Therefore, medium or small sized diverse sublibraries, designed using structural diversity and drug-likeness measures, is another approach for the rapid and efficient discovery of hits or leads [138]. As mentioned previously, the analysis of diversity is complex and subjective and has been a source of controversy [125, 139]. Whether a focused, screening strategy or screening a diverse compound library will do a better job in meeting both the scientific standards and business needs of discovering novel drugs remains to be seen, but it is an active area of current research in drug discovery.

***Fragment-Based Screening*** Fragment-based screening is a lead discovery approach in which lower molecular weight (120–250 Da) compounds than those usually present in the chemical libraries are tested in HTS campaigns. The number of compounds in fragment-based screening libraries are often significantly smaller than those used for HTS, and therefore the library design is even more important [140]. Even though these compound libraries are smaller, they are able to more efficiently sample available chemistry space with many of its possible combinations [141]. Further consideration to optimize the chemical tractability of the fragments present in generic fragment libraries have been included in library design strategies [142].

Compared with hits from HTS, the hits from fragment-based screening approaches usually have much lower affinity for their target (10 $\mu M$ to millimolars). As a consequence, compounds tested in fragment-based screens need to be present at higher testing concentrations and assayed using biophysical detection techniques such as NMR or crystallography rather than bioassays. However, hits from fragment-based screening methods typically possess high "ligand efficiency" (binding affinity per heavy atom) and so are highly suitable for optimization into lead candidates with good druglike properties [143]. However, it should be noted that these strategies do not lead to the identification of a unique solution to the identify and locations where the fragments that can bind, with fragmentlike compounds being able to interact in multiple ways at a particular proteinligand binding site. The consequence is that fragment-based screening could lead to the development of multiple leads [144].

***Compound Management and Integrity Characterization*** Chemical libraries routinely used for screening may contain large numbers of compounds, usually from hundreds of thousands up to one million or more. The increasing size of the compound collections used for screening has created a need for specialized hardware and software for automated compound management in order to cope with the increasing demands in HTS, as reviewed by Chan and Hueso-Rodriguez [145].

Compounds are usually stored as dry powders, in separate vials for each batch of compound synthesized, for long-term storage. Preparation of the compounds in a soluble form can be thought of as the initial step in the HTS. The preparation of compound solutions involves retrieving compounds from the compound store, distributing into microtubes, solubilization in dimethyl sulfoxide (DMSO), transferring and registration of the dissolved compounds into microplates (mother plate set), replica plating of multiple copies to intermediate plates (daughter plate sets), and then finally copying from these daughter plates into assay plates either with or without further dilution. These preparation steps are usually processed in a centralized compound management core facility. All of these different copies of compound plates and samples on the plates are registered with their attendant information being kept in electronic data files. Information

management systems are used to access and track the samples and link them to a variety of pertinent information that can help interpret HTS data.

Compound sample integrity and concentration accuracy should be monitored by appropriate QC procedures. Sample integrity and concentration accuracy in turn is linked to the quality and correct interpretation of the data generated from an HTS campaign. Various methodologies have been used for compound monitoring and characterization. Compound identity has been analyzed by LC/MS, NMR, and purity by LC/MS and ultraviolet (UV) spectrometry, evaporative light scattering, and nitrogen content [146]. Compound solubility and aggregation have been detected by light-scattering techniques, such as with a nephelometer. It has been proposed that "promiscuous inhibitors," possibly caused by colloidal precipitation, could be identified by enzymatic assays such as β-lactamase or chymotrypsin [147, 148]. Recently, a QC filter based on three 1D-proton NMR experiments has been proposed for rapid characterization of the identity, purity, solubility, and aggregation state of the active compound, for identifying false negatives and for detecting promiscuous inhibitors [149].

The commonly used compound solvent DMSO is hygroscopic and quickly absorbs water from the open atmosphere. Water presented in DMSO compound solutions can accelerate degradation and precipitation of some compounds. Understanding DMSO hydration in an HTS compound library can improve storage and preserve compound integrity [150]. A nondestructive, acoustic method compatible with HTS has been reported recently to measure water content in DMSO solutions. The performance of this acoustic method was compared with an optical technique and was reportedly in good agreement [151].

## 5 DATA HANDLING AND ANALYSIS

The relational database plays an essential role in an HTS laboratory. An integrated database enables compound registration, compound tracking, inventory control, assay/project registration, plate registration, version control, loading and storage of captured data from analytical instruments, data reduction/analysis/validation, and finally data mining for SAR or cross-target analysis. Figure 4 is a schematic overview of an HTS data management system.

This section focuses on the general data flow and aspects of HTS data analysis and data mining, which are the ultimate results of the HTS process.

### 5.1 HTS Data Analysis

***Data Flow*** A typical flowchart of HTS data analysis is shown in Figure 5. Data files captured directly from plate readers are first transferred to the bioassay data-

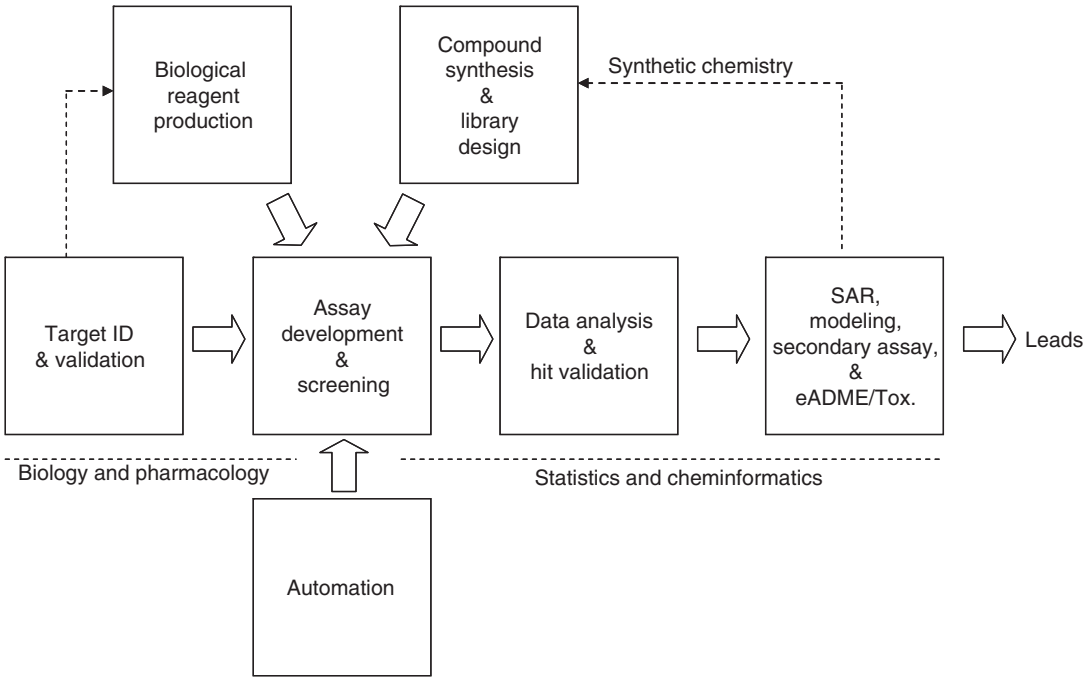


**Figure 4** Overview of an HTS data management system.

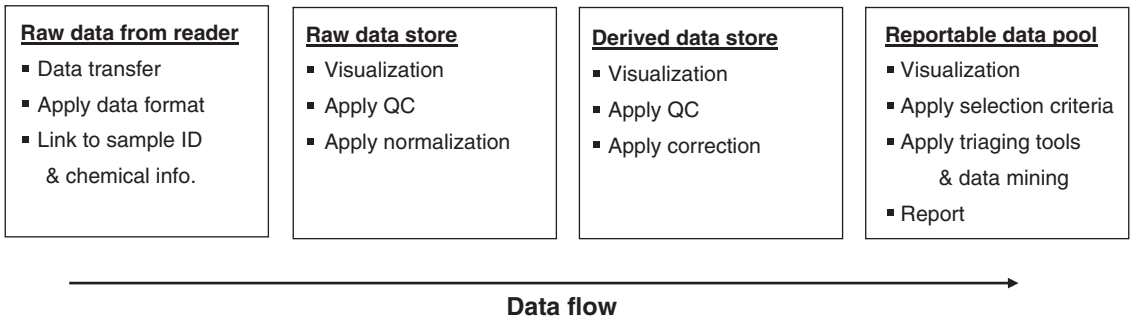


**Figure 5** Typical HTS data flow.

base and the data analysis system, where the data can be reformatted and linked uniquely with the compound sample identification number and other chemical information. These results reside in the system as raw data files. The raw data can be visualized, inspected, and checked with an appropriate set of quality control criteria. The plate data that passes these initial quality control filters is normalized by the designated reference values, usually by the mean positive and negative control signals. The normalized data can also be visualized, inspected, and checked again with a new set of more stringent quality control parameters. Systematic errors and other patterns detected by the analysis algorithm can be cautiously corrected if deemed appropriate. Finally, the resultant transformed data is ready for hit selection based on either statistical or activity criteria. The selected primary hits are subject to hit triage, data mining, or direct confirmation/validation testing.

***Data Normalization and Correction*** In most cases, the raw data is normalized platewise, using the mean signals from the positive and negative control wells present on each assay plate. In other cases, the positive and negative control wells may not reside on the same assay plate. It may be located on a QC plate that was run side by side with the assay plates during the same experimental run. Sometimes, rather than use the positive or negative control wells, the plate median signal is preferred to replace the corresponding control signal [107]. The argument for using plate median is that the control wells may not be representative of the sample wells, particularly when the control wells are not generated and treated exactly the same way as the rest of test samples. Depending on the specific assay configuration, one normalization method may produce slightly better results than the other. Generally, if there are no significant systematic errors or bias for the sample and control wells, these normalization methods should yield very similar results.

Data QC measures can be applied at several levels from wellwise to platewise to batchwise or even over an entire screening data set. If some systematic errors or patterns are detected, data analysis programs may apply a mathematic correction algorithm to compensate for the erroneous pattern in the data points (see [152] for a recent review) . It should be noted that any such correction should be cautiously justified before applying. As in any other scientific experiment, unless the systematic error and data pattern have a traceable, or at least an explainable cause, blind application of any correction can be very risky and is not recommended. Even if the systematic errors or patterns can be explained, as little data correction as possible should be applied, usually not exceeding 25–30% of the average signal in the screening data, as a rule of thumb. Any potentially problematic plates or data points flagged by QC parameters should be looked into carefully and retested if possible.

Visualization tools are indispensable parts of the data analysis system. One scatter plot display or a distribution histogram may offer more information about the quality of the screening or revealing potential issues than any data tables or statistical parameters. Some of the recommended visualization plots of HTS data were reviewed recently [152].

***Hit Limit Selection, Hit Rate, False Positives, and False Negatives in HTS*** Where is the best place to set the hit limit cutoff during primary screening of a chemical library? There is no straightforward answer. From a statistical point of view, it does not make a lot of sense if the hit limit setting is $<2 \times$ SD (standard deviations or modified equivalents) away from the activity mean of a random library. In many real screening campaigns, hit limits chosen at around $3 \times$ SD or greater or activity potency values (e.g., percent inhibition or stimulation) are used as the hit limit cutoffs. Since the hit limit is directly linked to primary hit rate, this means that in practice the chosen hit criteria is related to finding a

manageable number of compounds for hit validation and follow-up tests.

Since error is associated with any assay and screening process, the primary hits from a HTS will not be confirmed entirely. As mentioned earlier in this chapter, the hit confirmation rate, false-positive rate, and false-negative rate from a primary screen is influenced by not only the HTS assay quality but is also affected by hit limit selection and the primary hit profile from the screening library, as well as screening process and screening strategies. Several approaches have been used to assess false-positive and false-negative rates, from statistic models to real case studies [113, 114, 116, 153].

***Hit Confirmation and Validation/Dose–Response Curve Analysis*** The data analysis flow for hit confirmation testing is very similar to that of primary screening data. In a hit confirmation test, the hits are retested multiple times (e.g., 2–3 replicates) with the same assay used in the primary screen. The data for hit validation with dose–response analysis is normalized in a manner similar to primary screening results before curve fitting. The only difference in the treatment of the data is in applying any data correction. It is more difficult to detect and correct for any systematic errors and patterns on plates containing many dose–response tests. For this reason, separate set of control plates need to be tested along with the sample plates. The systematic errors or patterns found in the control plates are used for correction with sample plates. It can be very risky to apply such corrections and great care has to be exercised in doing so.

Dose–response curve analysis is usually carried out with common curve-fitting equations and programs. For most dose–response curves, a four-parameter, nonlinear equation is used for the curve fitting. Many kinds of competent curve-fitting software are readily available. Some of the challenging tasks for these analysis tools are outlier detection, curve-fitting model selection, and curve-fitting quality control. Visualization of the dose–response data with a fitted curve is an indispensable part of such analysis. Fully automatic curve-fitting algorithms that can satisfactorily fit curves to both inhibition and stimulation results are highly desirable.

### 5.2 Data Archive and Data Mining

Computational chemistry and cheminformatics have played important roles in successful HTS in the drug discovery setting. Several reviews have already summarized the use and impact of data mining on the discovery of compounds from high-throughput screens (e.g., [154–156]). There are a number of ways in which data mining can be performed to further improve the quality and analysis of HTS results.

***Chemotype Analysis and Hit Triage/Prioritization*** Cheminfomatics can make significant contributions to the screening process in at least several stages. First, cheminformatics has a significant impact on the compounds chosen for chemical library design. Library design can be as simple as removing unwanted compounds from the screening collection, such as removing reactive compounds or those containing moieties not believed to be of value to drug discovery (such as heavy-metal ion complexes; for a review see Rishton [157]). Another stage where cheminformatics impacts HTS outcome is when the primary hit list is evaluated to choose compounds for further validation and characterization. This hit triage process is needed especially in the cases where HTS campaigns identify more hits than those that can be practically progressed through validation and characterization.

Hit triaging can be considered as a series of steps:

1. The hits are filtered to remove those with reactive or "frequent hitter" characteristics. Cross-assay analysis and utilization of counterscreen data can reveal frequent hitters and nonspecific hits due to mechanisms other than the specific interaction on the target of interest.
2. The list usually is filtered again to remove compounds that are already known or known to be working by some inappropriate mechanism.
3. Apply appropriate "hitlike" or "leadlike" filters.
4. Cheminformatics then has the possibility to enrich the hit list by focusing on classes of compounds that are most likely to show reproducible activity using methods based either on modeling or clustering.
5. It can also ensure that a set of compounds with maximal diversity are selected for validation so that no one chemotype dominates the hit compound list.

***Modeling to Identify False Positives and Negatives*** High-throughput screening inherently carries the risk of generating false negatives as well as false positives. Significant effort has been invested in new technologies to help identify and deprioritize hits that are acting due to assay artifacts (false positives). However, other than independently retesting samples in a screen, there is little technology that can help to minimize false negatives. Data mining offers an alternative route for assessing and rescuing false negatives. A number of approaches have been proposed to use primary screening data to

generate models for predicting the activity of compound samples. Such models are then used to predict the activity of individual compounds so that the hits missed by screening can be rescued. A number of different model building methods have been applied in this way, including logistic regression [158], clustering [159], Bayesian modeling [160] and other sophisticated modeling methods. It has recently been shown that the greater amount of data offered by primary screening data can allow the derivation of models as predictive as models derived using validated $IC_{50}$ results [161]. Such methods also find application in genetic screens of small inhibitory RNAs (siRNAs) [162] and the evaluation of gene chip expression profiling [163]. All of these methods in a sense make the assumption that, by "averaging" the activity of compounds containing certain structural features (e.g., characteristics or substructures depending on the type of compound descriptors used), the potential activity of a compound can be predicted. However, a consequence of using such models is that "singleton" hits will be deprioritized for selection and hit follow-up, and therefore a balance in the application of such methods is needed. In addition, singleton false negatives cannot be rescued with such methods.

***Prediction of Mechanism and Off-Target Effects*** The modeling and mining of HTS data is not limited in use to identify active hits. Recently, such methods have been used to attempt to predict the possible mechanism of action of compounds, especially in cell-based assays where interaction with multiple targets may bring about the observed biological effect [164, 165]. As with approaches to predict activity against a single target, these attempts to predict the mechanism of action of a compound relies on clustering and Bayesian modeling methods. Essentially, these methods first create multiple models predicting the activity of compounds against a specific target or target class. Then, these models are used to predict the possible activity of any given compound in these multiple models; the models in which the compound is predicted to be active then suggest possible targets against which the molecule may be acting. Such methods also have been used to identify and explain the activity of compounds identified as frequent hitters in reporter gene assays [166] and so may be a way to help identify hits most suitable for lead optimization.

## 6 EMERGING TRENDS

Earlier sections have briefly reviewed and discussed the importance and different phases and aspects of the screening process from assay development and validation, design of the screen, quality control, data analysis, and the utilization of the HTS information to identify viable drug lead molecules. This section highlights several emerging trends that may play more important roles in the fields of HTS in early drug discovery, including fundamental approaches, technical advancements, and ongoing research interests in the HTS fields.

Currently, HTS capacity and throughput is often not perceived as a limiting factor in lead discovery. As a consequence, screening campaigns are no longer regarded as just an effort to screen as many compounds as possible (the "numbers" approach). The large number of "hits" generated from HTS campaigns in past years is slowly but surely starting to have a significant impact on the compounds coming to the clinic [167]. On the other hand, the "smart screening" approach of testing focused or designed sublibraries overall has also been of only limited success. How to construct the most effectively diverse or target-specific sublibraries is sometimes not even agreed upon by computational or medicinal chemists themselves. Thus, HTS is and will remain an important tool in early drug leads discovery for the foreseeable future, particularly for targets where little or no prior information is available or in cases where the goal of the project is to identify truly novel classes of inhibitors or ligands.

### 6.1 New HTS Approaches

HTS is facing new challenges and needs to adopt new approaches in order to improve efficiency and predictability to contribute to the early lead identification and drug discovery processes.

- ***Quality of HTS Assays Continually Improves*** This includes the continued development of new, high-quality assays that have better hit-finding capabilities due to improvements in biological relevance, reliability, sensitivity, and specificity. One of such approaches is the use of more direct readouts, such as MS-based assays. Another approach is the use of multiplexed readouts or well-validated orthogonal readouts, to enhance the efficiency of identifying "true" hits with desired activity profiles.
- ***Improvements in Screening Libraries*** The quality of the chemical library being tested has been shown to be at least as important as assay quality for a successful HTS campaign. The selection of a screening library with druglike and leadlike structures is a critical endeavor. The features of druglike and leadlike structures continues to be better defined. In addition, the diversity of druglike and leadlike molecular space continues to be explored

and categorized. The "preferred" or "privileged" structural scaffolds should be better characterized and, more importantly, better applied effectively in drug discovery and screening. Other areas of development will focus on the discovery and design of small molecules suitable for modulating protein–protein interactions and possibly a greater focus on natural productlike compounds.

- ***Combination of Potency Screening with Early ADME/Tox Testing*** The hits from potency-based screens should be evaluated, as early as possible, with a set of profiling (or selectivity) assays and toxicology testing. Such characterization will allow rapid triaging and filtering of HTS hits and, thus, maximize the chance of focusing on the chemotypes that may be more predictable and have lower attrition rates in the later stages of development.
- ***Combining and Exploring Novel Screening Strategies*** This may include evaluating strategies such as iterative screening in combination with strategies using lead modeling. Such iterative and combined screening strategies may reduce the apparent throughput and efficiency of screening (in terms of compounds screened per day) but may lead, overall, a more effective hit-to-lead discovery process. Similarly, HTS combined with other information-based screening strategies (various virtual screening and structure-based model building) will hopefully enhance the efficiency of hit expansion and evaluation. Therefore, HTS labs should also have medium throughput capability for screening higher value compounds in complex formats.
- ***Applying HTS in New Target Areas*** The needs of drug discovery will continue to drive the need for screening to be applied to new biological systems in order to support current drug discovery efforts. Some of the areas for which we see great potential for improvements for HTS is the development of additional assay formats for high-throughput ion channel readouts. High-throughput screening is a tool and an approach that has not just been limited to use in drug discovery but has also been applied in fields as diverse as genomic and proteomic discoveries, protein engineering, and the selection of chemical catalysts. Thus, it seems likely that the range of applications in which HTS is applied will continue to grow with increasing awareness of the power of this strategy.
- ***Increase in Academia Outreach and Collaboration*** HTS is no longer limited to industrial setting and is starting to play an increasingly important role in academic research [26, 168, 169]. With the new NIH initiatives to support a number of screening centers in the Molecular Libraries Screening Center Network (MLSCN) (http://nihroadmap.nih.gov/) for academic research use, it can be expected that the mutual impact of HTS and basic biology will increase, as well as the collaborative efforts in hit finding and target identification between academia and the pharmaceutical industry.

### 6.2 New Technologies

It is no doubt that new enabling tools will be important for future biomolecular screening. During the last decade, HTS has seen a paradigm shift enabled by advances in miniaturization, liquid handling, new assays, and improved detection technologies. Looking into the future, several trends emerge.

***Assay Miniaturization*** The pressures of cost constraints and throughput will continue to drive the need for assay miniaturization. Plate-based assay technologies will continue to be the dominant assay format. Non-plate-screening technologies have been developed and used at times in the past and will make increasingly more impact in future years. It is also possible that microfluidic assays and chip-based assay technologies will drive the miniaturization of HTS assays into new dimensions.

***Noncontact Liquid Handling*** With the continued drive for assay miniaturization, noncontact liquid-handling devices will play an increasingly important role. Currently, liquids are most conveniently handled by pipette tips for aspiration and dispensing into microtiter plates. Noncontact dispensing has been also used, such as Multidrop, FlexDrop, solenoid, and piezoelectic microdispensers. Recently, with the acoustic energy-driven liquid-dispensing devices (such as the ECHO550 or EDC system), liquid can be transferred in a small volume (nanoliter to microliter) accurately and directly from the wells on one plate into wells on another. So far, these devices have been best used for dispensing DMSO compound solutions, making it possible to run microliter volume assays without the need for tedious predilution steps. Increasing the speed and throughput of such transfer will further increase the number of assays that can be miniaturized. The transfer of aqueous solutions is also feasible with these devices. These nontip liquid transfer devices eliminate the need for tip washing and reduce the chance of liquid cross contamination.

***Label-Free Detection*** As mentioned above, the current signal detection methods in high-throughout

assays are predominantly fluorescent, colorimetric, or radioactive based. Such assay formats all require specifically labeled reagents and are subject to colored and fluorescent compound interference to various extents. In the last few years, new label-free detection systems based on surface sensors for in vitro and cell-based assays as a new detection paradigm in the HTS and related applications emerge [53]. A common requirement in these label-free detections is that the target or substrate have to be immobilized on a specialized surface for detection, with the exception of MS-based detection. One major advantage of MS-based functional screening is that it is not only label free, but it is a direct readout without any need of target or substrate immobilization. MS has been shown to be a powerful tool in HTS and has been used both in affinity-based screens and in functional-based HTS. Therefore, MS-based detection is a more generic detection method than many other label-free methods. Besides LC/MS-based HTS, other desorption-based MS detection technologies are also in development for use in HTS.

***High Content and Multiplexed Readouts*** Another emerging trend in HTS detection is multiplexed readouts. The traditional screening assay only monitors one or two signals of the testing system. With some specialized assay detection technologies, multiple readouts can be obtained. For example, a multimodule detector can simultaneously detect fluorescence intensity, fluorescent lifetime, and FP or FRET signals. Analysis of multiplexed readouts can provide information to discriminate a hit as either a true inhibitor or as an artifact caused by some interference with one of the detection modes. In a similar vein, high-content imaging analysis-based screening (HCS) also provides multiple readouts depicting cellular changes. With appropriate image analysis algorithms, a combination of the most relevant parameters, often deduced after unbiased analysis of all the possible readouts using tools such as principal component analysis, can be used to better assess the action of each hit and to eliminate false-positive hits. One of the challenges in multiplexed readouts or high-content imaging is the development of high-throughput methods to perform data analysis on more than a dozen different measurements with vastly different ranges of activity and activity distributions.

Another important set of assays with multiplexed readouts are assays to monitor messenger RNA (mRNA) transcript levels. The expression level of all the genes involved in a known signal transduction pathway or relevant selective genes can be monitored simultaneously as a way of monitoring the most relevant effects of a compound on a cell. The current technologies for multiple mRNA detection include quantitative reverse transcriptional polymerase chain reaction (PCR) (qRT-PCR), quantitative nuclease protection assay (qNPA), and Massarray assay technologies. The applications of such multiplexed in vitro and cell-based detection systems should provide more predicative information in hit and lead finding.

In summary, we reviewed the HTS processes and related methods and strategies in the above sections. HTS has become an indispensable tool for drug lead identification and discovery across the pharmaceutical industry. As a discovery tool, HTS still needs much improvement to be more efficient and predictive. In the foreseeable future, HTS will remain a fast growing field not only in pharmaceuticals but also in academic research with broader applications.

## ACKNOWLEDGMENTS

We would like to thank Dr. S. Siehler for help describing the reporter gene assay technologies, Dr. Q. Lu for help describing the ion channel assays, and Dr. M. E. Digan for help proofreading the manuscript.

## REFERENCES

1. Macarron R (2006). Critical review of the role of HTS in drug discovery. *Drug Discov Today 11*, 277–279.
2. Engels MFM, Venkatarangan P (2001). Smart screening: Approaches to efficient HTS. *Curr Opin Drug Discov Dev 4*, 275–283.
3. Seethala R, Fernandes P, Eds. (2001). *Handbook of Drug Screening*. New York: Marcel Dekker, pp. 525–562.
4. Janzen WP, Ed. (2002). *High Throughput Screening, Methods & Protocols, Methods in Molecular Biology*, Vol. 190. Totowa, NJ: Humana Press.
5. Dittrich PS, Manz A (2005). Single-molecule fluorescence detection in microfluidic channels—The Holy Grail in muTAS. *Anal Bioanal Chem 382*, 1771–1782.
6. Du MX, Sim J, Fang L, Yin Z, Koh S, Stratton J, Pons J, Wang JJ, Carte B (2004). Identification of novel small-molecule inhibitors for human transketolase by high-throughput screening with fluorescent intensity (FLINT) assay. *J Biomol Screen 9*, 427–433.
7. Clegg RM (1992). Fluorescence resonance energy transfer and nucleic acids. *Methods Enzymol 211*, 353–388.
8. Klumpp M, Boettcher A, Becker D, Meder G, Blank J, Leder L, Forstner M, Ottl J, Mayr LM (2006). Readout technologies for highly miniaturized kinase assays applicable to high-throughput screening in a 1536-well format. *J Biomol Screen 11*, 617–633.
9. Hemmila I, Webbs S (1997). Time-resolved fluorometry: An overview of the labels and core technologies for drug screening applications. *Drug Discov Today 2*, 373–381.

10. Park Y-W, Cummings RT, Wu L, Zhaeng S, Cameron PM, Woods A, Zaller DM, Marcy AL, Hermes JD (1999). Homogeneous proximity tyrosine kinase assays: Scintillation proximity assays versus homogeneous time-resolved fluorescence. *Anal Biochem 269*, 94–104.
11. Moshinsky DJ, Ruslim L, Blake RA, Tang F (2004). A widely applicable, high-throughput TR-FRET assay for the measurement of kinase autophosphorylation: VEGFR-2 as a prototype. *J Biomol Screen 8*, 447–452.
12. Gabriel D, Vernier M, Pfeifer MJ, Dasen B, Tenaillon L, Bouhelal R (2003). High throughput screening technologies for direct cyclic AMP measurement. *Assay Drug Dev Technol 1*(2), 291–303.
13. Nikiforov TT, Simeonov AM (2003). Application of fluorescence polarization to enzyme assays and single nucleotide polymorphism genotyping: Some recent developments. *Comb Chem High Throughput Screen 6*, 201–212.
14. Moore KJ, Turconi S, Ashman S, Ruediger M, Haupts U, Emerick VV, Pope AJ (1999). Single molecule detection technologies in miniaturized high throughput screening: Fluorescence correlation spectroscopy. *J Biomol Screen 4*, 335–354.
15. Wolcke J, Hunt N, Jungmann J, Ullmann D (2006). Early identification of false positives in high-throughput screening for activators of p53-DNA interaction. *J Biomol Screen 11*, 341–350.
16. Miraglia S, Swartzman EE, Mellentin-Michelotti J, Evangelista L, Smith C, Gunawan II, Lohman K, Goldberg EM, Manian B, Yuan PM (1999). Homogeneous cell-and bead-based assays for high throughput screening using fluorometric microvolume assay technology. *J Biomol Screen 4*, 193–204.
17. Seideman J, Peritt D (2002). A novel monoclonal antibody screening method using the Luminex-100 microsphere system. *J Immunol Methods 267*, 165–171.
18. Butcher H, Kennette W, Collins O, Demoor J, Koropatnick J (2003). A sensitive time-resolved fluorescent immunoassay for metallothionein protein. *J Immunol Methods 272*, 247–256.
19. Wu X, Sills MA, Zhang JH (2005). Further comparison of primary hit identification by different assay technologies and effects of assay measurement variability. *J Biomol Screen 10*, 581–589.
20. Lee HJ, Mun HC, Lewis NC, Crouch MF, Culverston EL, Mason RS, Conigrave AD (2007). Allosteric activation of the extracellular Ca 2+-sensing receptor by L-amino acids enhances ERK1/2 phosphorylation. *Biochem J 404*, 141–149.
21. Golla R, Seethala R (2004). A sensitive, robust high-throughput electrochemiluminescence assay for rat insulin. *J Biomol Screen 9*, 62–70.
22. Koresawa M, Okabe T (2004). High-throughput screening with quantitation of ATP consumption: A universal non-radioisotope, homogeneous assay for protein kinase. *Assay Drug Dev Technol 2*, 153–160.
23. Liu JJ, Wang W, Dicker DT, El-Deiry WS (2005). Bioluminescent imaging of TRAIL-induced apoptosis through detection of caspase activation following cleavage of DEVD-aminoluciferin. *Cancer Biol Ther 4*, 885–892.
24. Yueh MF, Kawahara M, Raucy J (2005). High volume bioassays to assess CYP3A4-mediated drug interactions: Induction and inhibition in a single cell line. *Drug Metab Dispos 33*, 38–48.
25. Zuck P, O'Donnell GT, Cassaday J, Chase P, Hodderr P, Strulovici B, Ferrer M (2005). Miniaturization of absorbance assays using the fluorescent properties of white microplates. *Anal Biochem 342*, 254–259.
26. Zolli-Juran M, Cechetto JD, Hartlen R, Daigle DM, Brown ED (2003). High throughput screening identifies novel inhibitorsa of *Escherichia coli* dihydrofolate reductase that are competitive with dihydrofolate. *Bioor Med Cehm Lett 13*, 2493–2496.
27. Iyer R, Barrese III AA, Parakh S, Parker CN, Tripp BC (2006). Inhibition profiling of human carbonic anhydrase II by high-throughput screening of structurally diverse, biologically active compounds. *J Biomol Screen 11*, 782–791.
28. Wahler D, Reymond J-L (2002). The adrenaline test for enzymes for enzymes. *Agnew Chem Int Ed Engl 41*, 1229–1232.
29. Wang C, Leffler S, Thompson DH, Hrycyna CA (2005). A general fuorescence-based coupled assay for *S*-adenosylmethionine-dependent methyltransferases. *Biochem Biophys Res Commun 331*, 351–356.
30. Zhou M, Diwu Z, Panchuk-Voloshina N, Haugland RP (1997). A stable nonfluorescent derivative of resorufin for the fluorometric determination of trace hydrogen peroxide: Applications in detecting the activity of phagocyte NADPH oxidase and other oxidases. *Anal Biochem 253*, 162–168.
31. Zhang B, Senator D, Wilson CJ, Ng S-C (2005). Development of a high-throughput robotic fluorescence-based assay for *Hs*Eg5 inhibitor screening. *Anal Biochem 345*(2), 326–335.
32. Naqvi T, Lim A, Rouhani R, Singh R, Eglen RM (2004). Beta galactosidase enzyme fragment complementation as a high-throughput screening protease technology. *J Biomol Screen 9*, 398–408.
33. Ozbal CC, LaMarr WA, Linton JR, Green DF, Katz A, Morrison TB, Brenan CJ (2004). High throughput screening via mass spectrometry: A case study using acetylcholinesterase. *Assay Drug Dev Technol 2*, 373–381.
34. Zhang JH (2006). Large scale LC/MS-based screening: Strategies, implementation and applications. Paper presented at the 12th Annual Conference of the Society of Biomolecular Science. Seattle, WA, September 17–21.
35. Shen Z, Go EP, Gamez A, Apon JV, Fokin V, Greig M, Ventura M, Crowell JE, Blixt O, Paulson JC, Stevens RC, Finn MG, Siuzdak G (2004). A mass spectrometry plate reader: Monitoring enzyme activity and inhibition with a

desorption/ionization on silicon (DIOS) platform. *Chembiochem 5*, 921–927.

36. Cummings MD, Farnum MA, Nelen MI (2006). Universal screening methods and applications of ThermoFluor. *J Biomol Screen 11*, 854–862.
37. Cunningham BT, Laing L (2006). Microplate-based, label-free detection of biomolecular interactions: Applications in proteomics. *Expert Rev Proteomics 3*, 271–281.
38. Pantoliano MW, Petrella EC, Kwasnoski JD, Lobanov VS, Myslik J, Graf E, Carver T, Asel E, Springer BA, Lane P, Salemme FR (2001). High-density miniaturized thermal shift assays as a general strategy for drug discovery. *J Biomol Screen 6*, 429–440.
39. Epps DE, Sarver RW, Roger JM, Herberg JT, Tomich PK (2001). The ligand affinity of proteins measured by isothermal denaturation kinetics. *Anal Biochem 292*, 40–50.
40. Vedadi M, et al. (2006). Chemical screening methods to identify ligands that promote protein stability, protein crystallization, and structure determination. *Proc Natl Acad Sci USA 103*, 15835–15840.
41. Torres FE, Kuhn P, De Bruyker D, Bell AG, Wolkin MV, Peeters E, Williamson JR, Anderson GB, Schmitz GP, Recht MI, Schweizer S, Scott LG, Ho JH, Elrod SA, Schultz PG, Lerner RA, Bruce RH (2004). Enthalpy arrays. *Proc Natl Acad Sci USA 101*, 9517–9522.
42. Zemanova L, Schenk A, Valler MJ, Nienhaus GU, Heilker R (2003). Confocal optics microscopy for biochemical and cellular high-throughput screening. *Drug Discov Today 8*, 1085–1093.
43. Lerner CG, Saiki AY (1996). Scintillation proximity assay for human DNA topoisomerase I using recombinant biotinyl-fusion protein produced in baculovirus-infected insect cells. *Anal Biochem 240*, 185–196.
44. Earnshaw DL, Pope AJ (2001). FlashPlate scintillation proximity assays for characterization and screening of DNA polymerase, primase, and helicase activities. *J Biomol Screen 6*, 39–46.
45. Crane K, Shih DT (2004). Development of a homogeneous binding assay for histamine receptors. *Anal Biochem 335*, 42–49.
46. Beveridge M, Park YW, Hermes J, Marenghi A, Brophy G, Santos A (2000). Detection of p56(lck) kinase activity using scintillation proximity assay in 384-well format and imaging proximity assay in 384- and 1536-well format. *J Biomol Screen 5*, 205–212.
47. Wong JT, Porter LS (1983). Ligand binding to protein measured by gel filtration on a dispersion column. *Anal Biochem 130*(2), 491–497.
48. Zehender H, Le Goff F, Lehmann N, Filipuzzi I, Mayr LM (2004). SpeedScreen: The "missing link" between genomics and lead discovery. *J Biomol Screen 9*(6), 498–505.
49. Comess KM, Schurdak ME, Voorbach MJ, Coen M, Trumbull JD, Yang H, Gao L, Tang H, Cheng X, Lerner CG, McCall JO, Burns DJ, Beutel BA (2006). An ultraefficient affinity-based high-throughout screening process: Application to bacterial cell wall biosynthesis enzyme MurF. *J Biomol Screen 11*(7), 743–754.
50. Annis DA, Nazef N, Chuang C-C, Scott MP, Nash HM (2004). A general technique to rank protein ligand binding affinities and determine allosteric versus direct binding site competition in compound mixtures. *J Am Chem Soc 126*, 15495–15503.
51. Whitehurst CE, Nazeef N, Annis DE, Hou Y, Murphy DM, Spacciapoli P, Yao Z, Ziebell MR, Cheng CC, Shipps GW, Felsch JS, Lau D, Nash HM (2006). Discovery and characterization of orthosteric and allosteric muscarinic $M_2$ acetylcholine receptor ligands by affinity selection-mass spectrocopy. *J Biomol Screen 11*, 194–207.
52. Nielsen UB, Geierstanger BH (2004). Multiplexed sandwich assays in microarray format. *J Immuno Meth 290*, 107–120.

52b Lofas S (2004). Optimizing the hit-to-lead process using SPR analysis. *Assay Drug Dev Technol 2*(4), 407–15.

53. Comley J (2005). Label-free detection: New biosensors facilitate broader range of drug discovery applications. *Drug Discov World Winter*(2004/5), 63–74.
54. Hage DS (1997). Chromatographic approaches to immunoassays. *J Clin Ligand Assay 20*, 293–301.
55. Slon-Usakiewicz JJ, Ng W, Dai JR, Pasternak A, Redden PR (2005). Frontal affinity chromatography with MS detection (FAC-MS) in drug discovery. *Drug Discov Today 10*, 409–416.
56. Newman DJ, Cragg GM, Snader KM (2000). The influence of natural products upon drug discovery. *Nat Prod Rep 17*, 215–234.
57. Fischer HP, Brunner NA, Wieland B, Paquette J, Macko L, Ziegelbauer K, Freiberg C (2004). Identification of antibiotic stress-inducible promoters: A systematic approach to novel pathway-specific reporter assays for antibacterial drug discovery. *Genome Res 14*, 90–98.
58. Goh EB, Yim G, Tsui W, McClure J, Surette MG, Davies J (2002). Transcriptional modulation of bacterial gene expression by subinhibitory concentrations of antibiotics. *Proc Natl Acad Sci USA 99*, 17025–17030.
59. Kunapuli P, Ransom R, Murphy KL, Pettibone D, Kerby J, Grimwood S, Zuck P, Hodder P, Lacson R, Hoffman I, Inglese J, Strulovici B (2003). Development of an intact cell reporter gene beta-lactamase assay for G protein-coupled receptors for high-throughput screening. *Anal Biochem 314*, 16–29.
60. Zuck P, Murray EM, Stec E, Grobler JA, Simon AJ, Strulovici B, Inglese J, Flores OA, Ferrer M (2004). A cell-based beta-lactamase reporter gene assay for the identification of inhibitors of hepatitis C virus replication. *Anal Biochem 334*, 344–355.
61. Thomas N (2001). CELL BASED ASSAYS—Seeing the light. *Drug Discov World Winter*(2001/2), 25–31.
62. Besemer J, Harant H, Wang S, Oberhauser B, Marquardt K, Foster CA, Schreiner EP, de Vries JE, Dascher-Nadel C, Lindley IJ (2005). Selective inhibition of

cotranslational translocation of vascular cell adhesion molecule 1. *Nature 436*(7048), 290–293.

63. Bhaumik S, Gambhir SS (2002). Optical imaging of *Renilla luciferase* reporter gene expression in living mice. *Proc Natl Acad Sci USA 99*, 377–382.
64. Paulmurugan R, Gambhir SS (2005a). Firefly luciferase enzyme fragment complementation for imaging in cells and living animals. *Anal Chem 77*, 1295–1302.
65. Prinz A, Diskar M, Herberg FW (2006). Application of bioluminescence resonance energy transfer (BRET) for biomolecular interaction studies. *Chembiochem 7*, 1007–1012.
66. Paulmurugan R, Gambhir SS (2005b). Novel fusion protein approach for efficient high-throughput screening of small molecule-mediating protein-protein interactions in cells and living animals. *Cancer Res 65*(16), 7413–7420.
67. Cabantous S, Terwilliger TC, Waldo GS (2005). Protein tagging and detection with engineered self-assembling fragments of green fluorescent protein. *Nat Biotechnol 23*, 102–107.
68. Boute N, Jockers R, Issad T (2002). The use of resonance energy transfer in high-throughput screening: BRET versus FRET. *Trends Pharmacol Sci 23*, 351–354.
69. Bertrand L, Parent S, Caron M, Legault M, Joly E, Angers S, Bouvier M, Brown M, Houle B, Menard L (2002). The BRET2/arrestin assay in stable recombinant cells: A platform to screen for compounds that interact with G protein-coupled receptors (GPCRS). *J Recept Signal Transduct Res 22*, 533–541.
70. Stockwell BR, Haggarty SJ, Schreiber SL (1999). High-throughput screening of small molecules in miniaturized mammalian cell-based assays involving post-translational modifications. *Chem Biol 6*, 71–83.
71. Olive DM (2004). Quantitative methods for the analysis of protein phosphorylation in drug development. *Expert Rev Proteomics 1*, 327–341.
72. Marshall IC, Boyfield I, McNulty S (2006). Ratiometric Ca2+ measurements using the FlexStation scanning fluorometer. *Methods Mol Biol 312*, 119–124.
73. Hodder P, Mull R, Cassaday J, Berry K, Strulovici B (2004). Miniaturization of intracellular calcium functional assays to 1536-well plate format using a fluorometric imaging plate reader. *J Biomol Screen 9*, 417–426.
74. Whiteaker KL, Gopalakrishnan SM, Groebe D, Shieh CC, Warrior U, Burns DJ, Coghlan MJ, Scott VE, Gopalakrishnan M (2001). Validation of FLIPR membrane potential dye for high throughput screening of potassium channel modulators. *J Biomol Screen 6*, 305–312.
75. Young SM, Bologa C, Prossnitz ER, Oprea TI, Sklar LA, Edwards BS (2005). High-throughput screening with HyperCyt flow cytometry to detect small molecule formylpeptide receptor ligands. *J Biomol Screen 10*, 374–382.
76. Dodgson K, Rawlings P, Dale I, Coldwell M, Hill S (2006). Early evaluation of the Corning Epic system for label-free functional GPCR assays. Poster presentation at the SBS Annual General Meeting, Seattle, WA.
77. Boyd MR (1997). The NCI in vitro anticancer drug discovery screen: Concept, implementation, and operation, 1985–1995. In Teicher BA, Ed. *Anticancer Drug Development Guide: Preclinical Screening, Clinical Trials, and Approval* Totowa, NJ: Humana Press, pp. 23–42.
78. Kleymann G, Werling HO (2004). A generally applicable, high-throughput screening-compatible assay to identify, evaluate, and optimize antimicrobial agents for drug therapy. *J Biomol Screen 9*, 578–587.
79. Benghezal M, Fauvarque MO, Tournebize R, Froquet R, Marchetti A, Bergeret E, Lardy B, Klein G, Sansonetti P, Charette SJ, Cosson P (2006). Specific host genes required for the killing of *Klebsiella* bacteria by phagocytes. *Cell Microbiol 8*, 139–148.
80. Moy TI, Ball AR, Anklesaria Z, Casadei G, Lewis K, Ausubel FM (2006). Identification of novel antimicrobials using a live-animal infection model. *Proc Natl Acad Sci USA 103*, 10414–10419.
81. Smart D, Gunthorpe MJ, Jerman JC, Nasir S, Gray J, Muir AI, Chambers JK, Randall AD, Davis JB (2000). The endogenous lipid anandamide is a full agonist at the human vanilloid receptor (hVR1). *Br J Pharmacol 129*, 227–230.
82. Chung TDY, Murphy JM (2001). Enzyme screens. In Seethala R, Fernandes P, Eds. *Handbook of Drug Screening*; New York: Marcel Dekke, pp. 283–312.
83. Wallace RW, Glodman ME (1997). Bioassay design and implementation. In Devlin JP, Ed. *High Throughput Screening: Discovery of Bioactive Substances*. New York: Marcel Dekker, pp. 279–306.
84. Copeland RA (2003). Mechanistic considerations in high-throughput screening. *Anal Biochem 320*, 1–12.
85. Wu G, Yuan Y, Hodge CN (2003). Determining appropriate substrate conversion for enzymatic assays in high-throughput screening. *J Biomol Screen 8*, 694–700.
86. Bruno S, Jenkins J, Hill WA, Parker CN (2005). Development and validation of a high throughput screen for inhibitors of hhai methyltransferase. Poster presented at the 11th Annual Conference for Biomolecular Screening. Geneva, September 11–15, 2005.
87. Manly SP (1997). In vitro biochemical screening. *J Biomol Screen 2*, 197–199.
88. Parandoosh Z (1997). Cell-based assays. *J Biomol Screen 2*, 201–202.
89. Turner BA, Parker CN, Grover GS, Lee PH (2004). Discovery, Evaluation and Characterization of Nuclear Receptor Antagonists. Poster presented at the American Society for Cell Biology, General Meeting. Washington DC, December 4–8, 2004.
90. Digan ME, Pou C, Niu H, Zhang J-H (2005). Evaluation of division-arrested cells for cell-based high throughput screening and profiling. *J Biomol Screen 10*, 615–623.
91. Zhang JH, Chung TD, Oldenburg KR (1999). A simple statistical parameter for use in evaluation and validation

of high throughput screening assays. *J Biomol Screen 4*, 67–73.

92. Sittampalam GS, Iversen PW, Boadt JA, Kahl SD, Bright S, Zock JM, Janzen WP, Lister MD (1997). Design of signal windows in high throughput screening assays for drug discovery. *J Biomol Screen 2*, 159–169.
93. Taylor PB, Stewart FP, Dunnington DJ, Quinn ST, Schulz CK, Vaidya KS, Kurali E, Lane TR, Xiong WC, Sherrill TP, Snider JS, Terpstra ND, Hertzberg RP (2000). Automated assay optimization with integrated statistics and smart robotics. *J Biomol Screen 5*, 213–226.
94. Iversen PW, Eastwood BJ, Sittampalam GS, Cox KL (2006). A comparison of assay performance measures in screening assays: Signal window, Z′ factor, and assay variability ratio. *J Biomol Screen 11*, 247–252.
95. Buxser S, Vroegop S (2005). Calculating the probability of detection for inhibitors in enzymatic or binding reactions in high-throughput screening. *Anal Biochem 340*, 1–13.
96. Sui Y, Wu Z (2007). Alternative statistical parameter for HTS assay quality assessment. *J Biomol Screen 12*, 229–234.
97. Oldenburg KR, Zhang J-H, Chen T, Maffia A, Blom KA, Coms AP, Chung TD (1998). Assay miniaturization for ultra-high throughput screening of combinatorial and discrete compound libraries: A 9600-well (0.2 microliter) assay system. *J Biomol Screen 3*, 55–62.
98. Oldenburg KR, Kariv I, Zhang J-H, Chung TDY, Lin S (2001). Assay miniaturization: Developing technologies and assay formats. In Seethala R, Fernandes P, Eds. *Handbook of Drug Screening*; New York: Marcel Dekker, pp. 525–562.
99. Marron BE, Jayawickreme CK (2003). Going to the well no more: Lawn format assays for ultra-high-throughput screening. *Curr Opin Chem Biol 7*, 395–401.
100. Imbert, P-E, Amoravain M, Parker CN, Gabriel D (2005). Evaluation of a new cAMP dynamic HTRF kit. Paper presented at the 11th Annual Conference for Biomolecular Screening. Geneva, September 11–15, 2005.
101. Babiac J (1997). Management and service Issues of a centralized robotic HTS core. In Devlin JP, Ed. *High Throughput Screening: The Discovery of Bioactive Substances*. New York: Marcel Dekker, pp. 461–470.
102. Gunter B, Brideau C, Pikounis B, Liaw A (2003). Statistical and graphical methods for quality control determination of high-throughput screening data. *J Biomol Screen 8*, 624–633.
103. Kevorkov D, Makarenkov V (2005). Statistical analysis of systematic errors in high-throughput screening. *J Biomol Screen 10*, 557–567.
104. Makarenkov V, Kevorkov D, Zentilli P, Gagarin A, Malo N, Nadon R (2006). HTS-Corrector: Software for the statistical analysis and correction of experimental high-throughput screening data. *Bioinformatics 22*, 1408–1409.
105. Gribbon P, Lyons R, Laflin P, Bradley J, Chambers C, Williams BS, Keighley W, Sewing A (2005). Evaluating real-life high-throughput screening data. *J Biomol Screen 10*, 99–107.
106. Padmanabha R, Cook L, Gill J (2005). HTS quality control and data analysis: A process to maximize information from a high-throughput screen. *Comb Chem High Throughput Screen 8*, 521–527.
107. Brideau C, Gunter B, Pikounis B, Liaw A (2003). Improved statistical methods for hit selection in high-throughput screening. *J Biomol Screen 8*, 634–47.
108. Sills MA,Weiss D, Pham Q, Schweitzer R, Wu X, Wu JJ (2002). Comparison of assay technologies for a tyrosine kinase assay generates different results in high throughput screening. *J Biomol Screen 7*, 191–214.
109. Wu X, Glickman JF, Bowen BR, Sills MA (2003). Comparison of assay technologies for a nuclear receptor assay screen reveals differences in the sets of identified functional antagonists. *J Biomol Screen 8*, 381–392.
110. Hentz N, Eckols K, Wegrzyn J, Dixon J, Weidner J (2003). Microfluidic chip-based vs. plate-based HTS: A comparison of data quality. Paper presented at the annual meeting of the Society for Biomolecular Screening. Portland, September 2003.
111. Zhu Z, Kim S, Chen T, Lin JH, Bell A, Bryson J, Dubaquie Y, Yan N, Yanchunas J, Xie D, Stoffel R, Sinz M, Dickinson K (2004). Correlation of high throughput pregnane X receptor (PXR) transactivation and binding assays. *J Biomol Screen 9*, 533–540.
112. Sun D, Whitty A, Papadatos J, Newman M, Donnelly J, Bowes S, Josiah S (2005). Adopting a practical statistical approach for evaluating assay agreement in drug discovery. *J Biomol Screen 10*, 508–516.
113. Zhang JH, Chung TDY, Oldenburg KR (2000). Confirmation of primary active substances from high throughput screening of chemical and biological populations: A statistical approach and practical considerations. *J Comb Chem 2*, 258–265.
114. Forgel P, Collette P, Dupront A, Garyantes T, Guedin D (2002). The confirmation rate of primary hits: A practical model. *J Biomol Screen 7*, 175–190.
115. Buxser S, Chapman DL (2006). Use of mixture distributions to deconvolute the behavior of "hits" and controls in high-throughput screening data. *Anal Biochem 361*, 197–209.
116. Zhang JH, Wu X, Sills MA (2005). Probing the primary screening efficiency by multiple replicate testing: A quantitative analysis of hit confirmation and false screening results of a biochemical assay. *J Biomol Screen 10*, 695–704.
117. Clardy J, Fischbach MA, Walsh CT (2006). New antibiotics from bacterial natural products. *Nat Biotechnol 24*, 1541–1550.
118. Snider M (1998). Screening of compound Libraries . . . Consomme or gumbo? *J Biomol Screen 3*, 169–170.
119. Schriemer DC, Hindsgaul O (1998). Deconvolution approaches in screening compound mixtures. *Comb Chem High Throughput Screen 1*, 155–170.

120. Ferrand S, Schmid A, Engeloch C, Glickman JF (2005). Statistical evaluation of a self-deconvoluting matrix strategy for high-throughput screening of the CXCR3 receptor. *Assay Drug Dev Technol 3*, 413–424.
121. Inglese J, Auld DS, Jadhav A, Johnson RL, Simeonov A, Yasgar A, Zheng W, Austin CP (2006). Quantitative high-throughput screening: A titration-based approach that efficiently identifies biological activities in large chemical libraries. *Proc Natl Acad Sci USA 103*(31), 11473–11478.
122. Bates SE, Fojo AT, Weinstein JN, Myers TG, Alvarez M, Pauli KD, Chabner BA (1995). Molecular targets in the National Cancer Institute drug screen. *J Cancer Res Clin Oncol 121*, 495–500.
123. Lajiness MS (1997). Dissimilarity-based compound selection techniques. *Perspect Drug Disc Dev 7/8*, 65–84.
124. Teig SL (1998). Informative libraries are more useful than diverse ones. *J Biomol Screen 3*, 85–88.
125. Spencer RW (1997). Diversity analysis in high throughout screening. *J Biomol Screen 2*, 69–70.
126. Lipinski CA (2000). Drug-like properties and the causes of poor solubility and poor permeability. *J Pharmacol Toxicol Methods 44*, 235–249.
127. Proudfoot JR (2002). Drugs, leads, and drug-likeness: an analysis of some recently launched drugs. *Bioorg Med Chem Lett 12*, 1647–1650.
128. Congreve M, Carr R, Murray C, Jhoti H (2003). A "rule of three" for fragment-based lead discovery? *Drug Discov Today 8*, 876–877.
129. Orry AJ, Abagyan RA, Cavasotto CN (2006). Structure-based development of target-specific compound libraries. *Drug Discov Today 11*, 261–266.
130. Alvesalo JK, Siiskonen A (2006). Similarity based virtual screening: A tool for targeted library design. *J Med Chem 49*(7), 2353–2356.
131. Kiselyov AS (2006). The concept of Focused Diversity: A new algorithm for building screening libraries. *IDrugs 9*, 193–198.
132. Sun D, Chuaqui C, Deng Z, Bowes S, Chin D, Singh J, Cullen P, Hankins G, Lee WC, Donnelly J, Friedman J, Josiah S (2006). A kinase-focused compound collection: Compilation and screening strategy. *Chem Biol Drug Des 67*, 385–394.
133. Keri G, Szekelyhidi Z, Banhegyi P, Varga Z, Hegymegi-Barakonyi B, Szantai-Kis C, Hafenbradl D, Klebl B, Muller G, Ullrich A, Eros D, Horvath Z, Greff Z, Marosfalvi J, Pato J, Szabadkai I, Szilagyi I, Szegedi Z, Varga I, Waczek F, Orfi L (2005). Drug discovery in the kinase inhibitory field using the Nested Chemical Library technology. *Assay Drug Dev Technol 3*(5), 543–551.
134. Miller JL (2006). Recent developments in focused library design: Targeting gene-families. *Curr Top Med Chem 6*, 19–29.
135. Goto Y, Arai-Otsuki S, Tachibana Y, Ichikawa D, Ozaki S, Takahashi H, Iwasawa Y, Okamoto O, Okuda S, Ohta H, Sagara T (2006). Identification of a novel spiropiperidine opioid receptor-like 1 antagonist class by a focused library approach featuring 3D-pharmacophore similarity. *J Med Chem 49*, 847–849.
136. Li J, Zhang J, Chen J, Luo X, Zhu W, Shen J, Liu H, Shen X, Jiang H (2006). Strategy for discovering chemical inhibitors of human cyclophilin a: Focused library design, virtual screening, chemical synthesis and bioassay. *J Comb Chem 8*, 326–337.
137. Chen G, Zheng S, Luo X, Shen J, Zhu W, Liu H, Gui C, Zhang J, Zheng M, Puah CM, Chen K, Jiang H (2005). Focused combinatorial library design based on structural diversity, druglikeness and binding affinity score. *J Comb Chem 7*, 398–406.
138. Krier M, Bret G, Rognan D (2006). Assessing the scaffold diversity of screening libraries. *J Chem Inf Model 46*(2), 512–524.
139. Wikel JH, Higgs RE (1997). Application of molecular diversity analysis in high throughout screening. *J Biomol Sreen 2*, 65–67.
140. Jacoby E, Davies J, Blommers MJ (2003). Design of small molecule libraries for NMR screening and other applications in drug discovery. *Curr Opin Med Chem 3*, 11–23.
141. Schuffenhauer A, Ruedisser S, Marzinzik AL, Jahnke W, Blommers M, Selzer P, Jacoby E (2005). Library design for fragment based screening. *Curr Top Med Chem 5*, 751–762.
142. Schuffenhauer A, Popov M, Schopfer U, Acklin P, Stanek J, Jacoby E (2004). Molecular diversity management strategies for building and enhancement of diverse and focused lead discovery compound screening collections. *Combi Chem High Throuput Screen 7*, 771–781.
143. Carr RA, Congreve M, Murray CW, Rees DC (2005). Fragment-based lead discovery: Leads by design. *Drug Discov Today 10*, 987–992.
144. Babaoglu K, Schoichet BK (2006). Deconstructing fragment-based inhibitor discovery. *Nat Chem Biol 2*, 720–723.
145. Chan JA, Hueso-Rodriguez JA (2002). Compound library management. *Methods Mol Biol 190*, 117–127.
146. Yurek DA, Branch DL, Kuo MS (2002). Development of a system to evaluate compound identity, purity, and concentration in a single experiment and its application in quality assessment of combinatorial libraries and screening hits. *J Comb Chem 4*, 138–148.
147. McGovern SL, Caselli E, Grigorieff N, Shoichet BK (2002). A common mechanism underlying promiscuous inhibitors from virtual and high-throughput screening. *J Med Chem 45*, 1712–1722.
148. McGovern SL, Helfand BT, Feng B, Shoichet BK (2003). A specific mechanism for nonspecific inhibition. *J Med Chem 46*, 4265–4272.
149. Dalvit C, Caronni D, Mongelli N, Veronesi M, Vulpetti A (2006). NMR-based quality control approach for the identification of false positives and false negatives in high throughput screening. *Curr Drug Discov Technol 3*, 115–124.

150. Kozikowski BA, Burt TM, Tirey DA, Williams LE, Kuzmak BR, Stanton DT, Morand KL, Nelson SL (2003). The effect of freeze/thaw cycles on the stability of compounds in DMSO. *J Biomol Screen 8*, 210–215.
151. Ellson R, Stearns R, Mutz M, Brown C, Browning B, Harris D, Qureshi S, Shieh J, Wold D (2005). In situ DMSO hydration measurements of HTS compound libraries. *Comb Chem High Throughput Screen 8*(6), 489–498.
152. Gubler H, Bookchapter (2007). Methods for statistical analysis, quality assurance and management of primary HTS data. In Hüser J, Ed. *High-Throughput Screening in Drug Discovery*, Vol. 28 of Methods and Principles in Medicinal Chemistry, Mannhold R, Kubinyi H, Folkers G, Series Ed. Weinheim: Wiley-VCH, pp. 511–205.
153. Woodward PW, Williams C, Sewing A, Benson N (2006). Improving the design and analysis of high-throughput screening technology comparison experiments using statistical modeling. *J Biomol Screen 11*, 5–12.
154. Davies JW, Glick M, Jenkins JL (2006). Streamlining lead discovery by aligning in silico and high-throughput screening, *Curr Opin Chem Biol 10*, 343–351.
155. Parker CN, Schreyer SK (2005). *Application of chemoinfomatics to high throughput screening: practical considerations*. In Bajorath J, Ed. *Chemoinformatics: Methods and Protocols*. Totowa, NJ: Humana.
156. Fara DC, Oprea TI, Prossnitz ER, Bologa CG, Edwards BS, Sklar LA (2006). Integration of virtual and physical screening. *Drug Discov Today Tech 3*, 377–385.
157. Rishton GM (1997). Reactive compounds and *in vitro* false positives in HTS. *Drug Discov Today 2*, 382–384.
158. Engels MF, Wouters L, Verbeeck R, Vanhoof G (2002). Outlier mining in high throughput screening experiments. *J Biomol Screen 7*, 341–351.
159. Yan SF, Asatryan H, Li J, Zhou Y (2005). Novel statistical approach for primary high-throughput screening hit selection. *J Chem Inf Model 45* 1784–1790.
160. Glick M, Klon AE, Acklin P, Davies JW (2004). Enrichment of extremely noisy high-throughput screening data using a naive Bayes classifier. *J Biomol Screen 9*, 32–36.
161. Glick M, Jenkins JL, Nettles JH, Hitchings H, Davies JW (2006). Enrichment of high-throughput screening data with increasing levels of noise using support vector machines, recursive partitioning, and Laplacian-modified naive Bayesian classifiers. *J Chem Inf Model 46*, 193–200.
162. Friedman N, Linial M, Nachman I, Pe'er D (2000). Using Bayesian networks to analyze expression data. *J Comput Biol 7*, 601–620.
163. Xu Y, Brereton RG (2005). Diagnostic pattern recognition on gene-expression profile data by using one-class classification. *J Chem Inf Model 45*, 1392–1401.
164. Nettles JH, Jenkins JL, Bender A, Deng Z, Davies JW, Glick M (2006). Bridging chemical and biological space: "Target fishing" using 2D and 3D molecular descriptors. *J Med Chem 49*, 6802–6810.
165. Yan SF, King FJ, He Y, Caldwell JS, Zhou Y (2006). Learning from the data: Mining of large high-throughput screening databases. *J Chem Inf Model 46*, 2381–2395.
166. Crisman TJ, Parker CN, Jenkins JL, Thoma M, Kang ZB, Davies JW, Glick M (2006). A chemogenomics approach to the prioritization of cell-based HTS data: A test case of reporter gene assays. *J Med Chem* (submitted).
167. Fox S, Farr-Jones S, Sopchak L, Boggs A, Nicely HW, Khoury R, Biros M (2006). High-throughput screening: Update on practices and success. *J Biomol Screen 11*, 864–869.
168. Young SM, Curry MS, Ransom JT, Ballesteros JA, Prossnitz ER, Sklar LA, Edwards BS (2004). High-throughput microfluidic mixing and multiparametric cell sorting for bioactive compound screening. *J Biomol Screen 9*, 103–111.
169. Tolliday N, Clemons PA, Ferraiolo P, Koehler AN, Lewis TA, Li X, Schreiber SL, Gerhard DS, Eliasof S (2006). Small molecules, big players: The National Cancer institute's Initiative for Chemical Genetics. *Cancer Res 66*, 8935–8942.

# MASS SPECTROMETRY

Robert D. English and Anthony M. Haag
*University of Texas Medical Branch, Mass Spectrometry Core Laboratory, Biomolecular Resource Facility, Galveston, Texas*

## 1 INTRODUCTION

The intent of this chapter is to provide a concise summary of informative topics related to the study of therapeutic agents by mass spectrometry (MS). Due to the vast number of published studies in this field alone, we have elected to restrict the cited literature to only the most recent as possible. Our objective is to highlight as many pertinent MS techniques as possible. However, it should be noted that this will not be an all-inclusive list of approaches due to the large volume of literature. Additionally, our intent is not to emphasize equations and detail about theory; a plethora of textbooks, training manuals, and journal articles concerning these details already exists. Theory will be discussed and referenced in this work only when its integration can make points more understandable.

This chapter is primarily divided into the areas of instrumentation, sample preparation, and applications in proteomics, metabolomics, and emerging fields in drug analysis. Each division will be discussed using examples from recent literature with the expectation that the reader will navigate to the appropriate topic for explanations and for direction to relevant citations. More time will be spent on instrumentation details in the first part in order to build a solid platform on which to discuss relevant applications in the latter portion of this review. Moreover, specific mass spectrometer designs used in the development and analysis of therapeutic agents will be given more attention based on their relevance and abundance in the literature.

## 2 INSTRUMENTATION

### 2.1 Sample Purification

In general, most MS experiments are easiest and results most accurate when a sample is as homogeneous as possible. Otherwise, the interpretation of mass spectra becomes less trivial, for example, when dealing with multiple-charge peaks in an electrospray ionization (ESI) experiment. Although software for deconvoluting ESI mass spectra accompanies most if not all commercial mass spectrometers nowadays, the software is usually unable to determine sample purity. The software attempts to identify the most probable protein(s) based on mass spectral comparison between experimental and theoretical data.

There are numerous methods employed in the fractionation, desalting, and specific selection (binding) of sample components prior to analysis and detection by MS. We will discuss several forms of sample treatment that make the analysis of pharmaceuticals and related drugs possible by MS.

***High-Performance Liquid Chromatography*** High-performance liquid chromatography (HPLC), also known as high-pressure liquid chromatography, is the most widely used (online) method of sample separation for analysis by MS. HPLC is so commonly used that, at the time of this writing in July 2011 when we executed a Scifinder Scholar [1] search (criteria: *dates 2003–2011, journal only, English language*), almost 95,000 hits were

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

registered. When the search was restricted to "HPLC and MS" using the same criteria, the list was decreased to 9901 hits. Further restriction of the search terms to "HPLC and MS and drugs" resulted in 6354 hits, "*containing all of the concepts*".

Choice of chromatographic separation depends on the relevant application. HPLC practice and theory is readily available [2, 3]; however, some common HPLC separation methods include reverse phase, normal phase, ion exchange, and size exclusion. HPLC typically uses pressures up to approximately 6000 psi in order to maintain a constant flow rate of approximately 1 mL/min; keeping a stable flow rate enables reproducible chromatography conditions. The stationary phase is packed into a column that must be physically capable of withstanding these high pressures. Reverse-phase (RP) chromatography is by far the most common means of separation used today both with and without MS. The separation mechanism relies on different solubilities in buffered aqueous and organic solvents. C18 (and C8) are the most common hydrocarbon-based alkyl groups bonded to the inner 5- or 10-μm porous silica surface. The linked carbon-based alkyl group is the bound, *nonpolar* surface on which analyte adsorption occurs. This nonpolar stationary phase allows for the most polar compound to elute first from the column; the opposite holds true in *normal-phase* chromatography where a stationary polar phase dictates that the least polar component is eluted first. More detailed information can be found about HPLC stationary phases in the literature [4].

Oftentimes a gradient elution profile for the mobile phase is used so that the ratio of organic solvent (such as acetonitrile in RP-HPLC) to water incrementally increases during the course of the chromatographic separation. As the various adsorbed compounds experience a changing organic content, the analytes of interest desorb and elute from the column based on their solubility with the mobile-phase composition. On the other hand, isocratic elution conditions can also be used, which leave the mobile-phase gradient unchanged.

Many commercial HPLC detectors are used in the detection of eluting compounds. Examples of these types of detectors include absorbance, fluorescence, light scattering, Fourier transform–infrared (FT-IR), electrochemical, refractive index, conductivity, and last, but certainly not least, mass spectrometry [5]. As a stand-alone technique, ultraviolet (UV) absorbance is the most widely used detection method of choice. A review article was recently published that, in part, surveys approximately 40 examples of HPLC assays for analyzing bulk pharmaceuticals and formulated drug products [6].

In order to take full advantage of HPLC, however, chromatography should be coupled to a stand-alone or post-UV detector mass spectrometer. Linking HPLC to MS combines the separation power of HPLC with the sensitivity of MS [7]. Both techniques can stand alone but are much more effective when coupled together as an LC/MS instrument, as evident by its increased attention and usage in research articles, commercial instrumentation, and work flows in pharmaceutical companies over the past two decades [8–10].

Without HPLC, identification of compounds in a mixture is sometimes possible but is certainly less trivial. Utilizing HPLC to allow "pure" components to be injected into the mass spectrometer takes full advantage of the sensitivity of MS because mass spectra of countless compounds are specific enough to allow identification with almost complete certainty, thereby making identification of numerous components in a mixture much more likely. Specific LC/MS examples will be addressed in more detail in later sections.

***Nanoflow HPLC*** Nanoflow refers to HPLC under much slower mobile-phase flow rates in the nanoliter/minute rather than the microliter/minute–milliliter/minute flow rate as in conventional HPLC. Instruments using nanoflow HPLC are becoming more widespread [11–16]. It must be mentioned, however, that there are several advantages as well as disadvantages to working in the nanoflow regime.

The disadvantages of nanoflow HPLC make it problematic under certain conditions. First, quantitation is very difficult due to spray instabilities into the mass spectrometer at this low of a flow rate; this problem can be circumvented, however, through appropriate use of internal standards as performed by Applied Biosystems [17]. Second, irregular ion production may occur due to desolvation of small droplets occurring too rapidly. Third, additional hardware such as splitters must be utilized in order to decrease the flow rate from conventional HPLC pumps used in this setup; this problem is normally addressed by the LC manufacturer unless a custom-built system is being used.

The advantages of nanoflow HPLC make it a very popular technique. First, less sample is consumed and injected during this process. Only 1 μL of sample volume is often required for injection into the mobile phase, although it is a concentration-dependent process. Second, ionization efficiencies approach 100% in the mass spectrometer due to the fact that the droplets formed during ionization are thought to contain only a single analyte molecule; in other words, there is less ionization suppression within smaller droplet sizes. Third, ions can be produced from high water-containing mobile phases. Fourth, the slower flow rate allows this

nanotechnique to be married to other separation techniques such as capillary electrophoresis (CE).

***Ultraperformance Liquid Chromatography*** The theory behind ultraperformance liquid chromatography (UPLC) is not as new as its popularity. UPLC uses sub-2-μm particles and (higher) pressures up to approximately 14,000 psi in order to gain improvements in resolution, analysis time, and detection sensitivity [18–20]. Additionally, UPLC can result in shorter run times via LC/MS/MS (MS/MS is a tandem MS technique that allows for precursor ion selection and deliberate fragmentation to give structural and sequence information) analysis of small molecules and peptides. Increased resolving power and absolute quantitation were demonstrated with the analysis of a serum digest on an Applied Biosystems 4000 QTrap hybrid triple-quadrupole linear ion trap mass spectrometer [21]. The advantages of UPLC were realized by Mazzeo and co-workers when they applied it to metabolite identification and quantitation [18]. UPLC/MS/MS demonstrated faster analysis speed and higher sensitivity compared to HPLC/MS/MS. A problem with UPLC, however, lies in the fact that nonlinear effects are present at these high pressures; these effects create solvent compressibility issues not typically encountered in traditional HPLC.

***Pulsed Ultrafiltration*** Pulsed ultrafiltration (PUF) is typically used in combination with MS and combinatorial libraries to screen chemical mixtures for compounds that bind to specific receptors [22–24]. In a typical PUF-MS setup, a solution chamber contains a porous membrane for molecular weight selectivity. A pulse containing an extract or a library of compounds is passed through the chamber where both ligand–receptor complexes and unbound components form. Following a wash cycle to eliminate the unbound compounds, the bound ligands are eluted into an HPLC column and then injected in their desalted forms into a mass spectrometer for analysis and identification.

Compared to other affinity-based techniques, PUF has advantages and disadvantages. Native conformations and binding interactions of ligands and receptors, respectively, are conserved because the reaction occurs in solution phase. Also, receptor compounds can be recovered and reused in a PUF experiment. Moreover, PUF can be used online with MS when volatile buffers are used, a scheme not common to all techniques. A disadvantage of PUF is that one needs to minimize nonspecific binding of molecules to the inner chamber surface and membrane.

A similar technique to PUF is ultrafiltration-MS, which uses centrifugal ultrafiltration as opposed to a membrane to separate bound and unbound complexes [25].

***Frontal-Affinity Chromatography*** Frontal-affinity chromatography (FAC) is based on the principle that compounds are in a dynamic equilibrium with an immobilized target (on a column) [26]. The unbound compounds are eluted earlier than the bound, with subsequent analysis by MS. FAC is used in many studies including protein–ligand characterization such as protein–protein, protein–drug, protein–small molecule, protein–DNA (deoxyribonucleic acid), and even protein–carbohydrate interactions. Additionally, FAC can be found in ligand discovery studies [27, 28]. In one study, the binding properties of inhibitors of the epidermal growth factor receptor were analyzed, and an approach was developed for quick screening of drug candidates from natural resources [28].

***Other Sample Purification Methods*** A recent review article describes many more affinity-based MS screening methods [22]. In it are discussed affinity size exclusion–MS [29], affinity capillary electrophoresis–MS [30], surface plasmon resonance–MS [31], affinity capture–MS [32], and noncovalent affinity–MS [33]. All of the techniques use MS as the final identification step of binding ligands, however, the approaches do slightly differ in the principle of separation of binding compounds from unbound components. Affinity size exclusion–MS uses fractionation in a spin column to achieve separation of target-ligand complexes. Affinity capillary electrophoresis–MS measures the mobility change of ligands upon reaction with target to identify bound ligands. Surface plasmon resonance, or SPR-MS, identifies the binders for an immobilized target by measuring the change in surface refractive index. Affinity capture–MS identifies bound ligands by mixing a bead-bound target with potential ligands. Noncovalent affinity–MS is the only affinity-based MS sample purification method that measures the mass of a ligand-target(s) complex by analyzing the target-ligand mixture directly with no prior separation or washing of bound and unbound compounds.

***Desalting*** Ionic compounds are often required to keep many biochemical samples stable in a buffered environment; however, they can hinder ionization in the mass spectrometer and/or cause the formation of adducts with the sample. This would not be a problem if it did not in turn decrease or inhibit the ionization for the protonated form of the analyte, which is most often the case. Many desalting approaches are used, some of which include ziptips, dialysis, spin columns, and desalting columns.

Ziptips (manufactured by Sigma) are a fast, simple, and effective method for cleaning up samples requiring routine matrix-assisted laser desorption/ionization (MALDI) and ESI mass spectrometric analysis. Ziptips are readily available with C4 or C18 nonpolar stationary phase bound to a resin in the end of a pipette tip. The desalting process takes only a few minutes and involves only a few solutions to concentrate a sample for subsequent mass analysis. Most often, acetonitrile is used to wet the resin. Next, 0.1% trifluoroacetic acid (TFA) is used to condition the tip. Then, the sample of interest is aspirated many times through the pipette tip in order to bind the sample to the resin. To wash the sample and remove salts, 0.1% TFA is used again. Finally, the sample is eluted from the resin with only a few microliters of a mixture of organic/water with a small amount of acid. The exact ratio of solvent composition varies depending on the peptides or proteins of interest because the process is based on sample-dependent, hydrophobic properties. A commonly used elution solvent in our laboratory is 50/50 or 70/30 acetonitrile/water in 0.1% TFA when using C18 ziptips.

Dialysis is a very cost-effective technique used both to separate differently sized molecules and to exchange a solute's solution (i.e., buffer exchange). Exchanging a nonvolatile buffer with a volatile buffer can be extremely beneficial for MS experiments because fewer problems are encountered with volatile buffers. In ESI, for example, nonvolatile buffers can clog the ionization tip in the mass spectrometer ionization source. Molecules are dialyzed by using size-selective, permeable tubing with high volumes of buffer washing. Some advantages include the low cost and reuse of the dialysis tubing, as well as that it is not a concentration-dependent technique. Disadvantages include long duration cycles and copious amounts of required washing.

Spin columns are a convenient means of desalting/purifying samples. They typically include some nonpolar, stationary phase (i.e., C18) for sample purification and subsequent concentration. For example, protein sample can simply be combined with a resin inside the column, and then the purified sample eluted following a washing step.

Desalting columns are another practical and easy approach for sample purification, and they function very similarly to spin columns. They are made out of a variety of materials with varying pore size so as to act as a molecular weight filter for the size of interest. These columns are commonly used for desalting large biomolecules such as proteins.

## 2.2 Ionization Sources

Information regarding many areas of mass spectrometry can be found in numerous analytical chemistry textbooks [5, 34], mass spectrometry textbooks [35–39], proteomics books [40], and Internet websites [41]. As of July 2011, a Google search of just the term "mass spectrometry" resulted in almost 11.6 million hits.

Before discussing any individual aspect of a mass spectrometer, it is useful to first briefly address the system as a whole. In simplest terms, mass spectrometers measure the mass-to-charge ($m/z$) ratio of a compound. The necessary components of any mass spectrometer include an ionization source, mass analyzer, and detector. Compounds must first be introduced into the mass spectrometer, whereupon they become: ionized and then accelerated by an electric and/or magnetic field; mass analyzed and separated according to their $m/z$ ratio; and detected at a device whereupon its signal is sent to and processed by a computer. If a gas-phase molecule cannot be ionized, then the compound is considered to be a neutral (zero charge), will not be affected by the field, and will not be detected.

An ionization source, the topic at hand, is vital to all mass spectrometers. Many ionization techniques are available, however, we will discuss four common types: electrospray ionization (ESI), atmospheric pressure chemical ionization (APCI), matrix-assisted laser desorption/ionization (MALDI), and atmospheric pressure photoionization (APPI).

***Electrospray Ionization*** The advent of ESI, one of many atmospheric pressure ionization sources, revolutionized the analysis of (large) biomolecules in the mid-1980s [42, 43]. Although techniques such as plasma desorption mass spectrometry (PDMS) [44], fast atom bombardment (FAB) [45], and secondary ion mass spectrometry (SIMS) [45–48] had been shown to desorb nonvolatile molecules, ESI became one of the most widely used ionization techniques after its inception because it produces gaseous ionized molecules from a liquid solution. In other words, ESI could easily be coupled to an HPLC system, which paved the way for new applications in biological research for mass spectrometry.

Briefly, desolvated ions form from charged liquid droplets. This process is realized by the creation of a spray of charged droplets in an electric field at a metal tip maintained at several thousand volts. Multiple-charge ions do commonly form during the ESI process. Although the mass analyzers on ESI instruments typically only pass molecules with $m/z$ values of 2000–3000, larger molecules (with molecular weights up to several hundred kilodaltons, but more routinely up to 70–100 kDa) can be analyzed because the ionization process can attach numerous protons, or charges, to molecules with sufficient sites, thereby pushing the $m/z$ of the compound(s) into the range of the mass analyzer.

ESI has many advantages. First, as previously mentioned, it encompasses a high practical mass range, primarily because of multiple-charge ions realizing the detection of macromolecules. Second, under the right conditions, it can be a highly sensitive technique (low picomole–femtomole). Third, ESI easily adapts to LC interfaces. Fourth, there is no matrix interference as with MALDI, meaning low-molecular-weight species (100–800 Da) are more easily identified. Fifth, ESI is a soft ionization technique that allows for the study of noncovalent interactions.

Disadvantages of ESI include low tolerance for mixtures (unless combined with LC), multiple charging (makes interpretation of spectra sometimes complex and in need of deconvolution software), and low salt tolerance.

***Atmospheric Pressure Chemical Ionization*** An atmospheric pressure chemical ionization (APCI) source is another ionization technique useful for a class of compounds sometimes not identified easily by ESI [49]. A corona discharge serves as the driving force behind forming and maintaining the ion-molecule reactions present in this plasma that ionize the neutral analyte molecules. Briefly, a corona discharge occurs when the field at the tip of an electrode ionizes the surrounding gas, where the vaporized solvent acts as the reagent gas. The analyte ions are subsequently transferred into the mass spectrometer in the same manner as with ESI; the interface between the ionization sources (APCI and ESI) and the mass analyzer is identical, the only difference being that APCI also requires a heated pneumatic nebulizer (along with a spray chamber and needle as in ESI) to be placed in front of the mass analyzer.

Some advantages of APCI (compared to ESI) include: low to slightly high polar compounds can be analyzed, higher LC flow rates (2 mL/min) can be used, and higher tolerance to buffers in the mobile phase can be tolerated.

APCI suffers from lower reproducibility, however, compared to other ionization techniques. Unlike desorption methods that can analyze nonvolatile species, APCI (and other atmospheric pressure ionization sources) requires that the analytes of interest be volatile. APCI does not function well at lower LC flow rates, and may contain more analyte–buffer adducts in mass spectra.

***Matrix-Assisted Laser Desorption/Ionization*** Matrix-assisted laser desorption/ionization (MALDI) [50–53] is yet another soft ionization technique but somewhat different from the aforementioned atmospheric methods. Briefly, a focused beam of light from a laser (UV, IR, or Nd:Yag) is directed onto a solid support containing a mixture of both analyte and matrix under vacuum. The sample/matrix is desorbed from the surface upon interaction with the pulsed light beam. Although the exact mechanisms of ion formation are still debatable today, the purpose of the matrix is threefold: to co-crystallize with the sample, to absorb at a similar wavelength or energy as the laser so as to only allow minimal or no fragmentation of the target molecule, and to donate or withdraw a proton from the sample in order to form the required sample ions needed for acceleration into a mass analyzer. Upon formation, ions (positive or negative charge) are accelerated into a mass analyzer [typically, a time-of-flight (TOF) field-free region] via an electric field. It is within the TOF region that ions separate according to their respective *m/z* ratio and are subsequently detected at the end of the drift tube at an ion detector. Reflectrons are often used in conjunction with the flight tube to bend the trajectory of the ions; this increases the length of the TOF region and corrects for different kinetic energies of ions with the same *m/z* ratio.

MALDI is advantageous for several reasons: practical mass range of several hundred kilodaltons, typical sensitivity in the low femtomole range, soft ionization with little fragmentation, and tolerance of millimolar concentrations of salt and other ionizing suppressing compounds. Disadvantages include matrix background (inhibits routine analysis of molecules less than 800 Da), photodegradation of sample by the laser desorption process, and cannot be connected to online HPLC.

***Atmospheric Pressure Photoionization*** Atmospheric pressure photoionization (APPI) is a newer ionization method compared to all methods described previously [54]. APPI is oftentimes used as an alternative to APCI. The primary difference from APCI is that rather than a corona discharge as the ionization vehicle, APPI utilizes a UV light source. APPI sometimes can ionize species not possible by APCI, just as APCI can ionize species sometimes not possible by ESI. All three techniques are often used in a complementary fashion because of both the advantages of each ionization source and the relative ease of interchanging the sources on the same instrument.

### 2.3 Analyzers

Mass analyzers are another one of the basic components to a mass spectrometer. After ionization occurs, the charged molecules must enter another region of the mass spectrometer in which separation of the ion mixture occurs. Nowadays, there are many mass analyzer designs and variations that accomplish this task. Some analyzers not only separate charged ions from each other but also store and filter the ions. In other words, one analyzer can function in several modes. This multifunctioning ability often makes it advantageous to

design mass spectrometers that utilize more than one type of analyzer in a single instrument. Mass analyzers measure the *m/z* ratio. This becomes especially important when compounds take on several charges (*z*) because the *m/z* will be much less than the actual mass (*m*) of the compound. Mass analyzers can directly affect instrument parameters such as resolution, mass accuracy, mass range, and overall sensitivity.

Many common types of analyzers include quadrupole, ion trap, magnetic sector, double focusing magnetic sector, time-of-flight, Fourier transform–ion cyclotron resonance, and numerous variations and combinations (i.e., hybrids) of these analyzers [55]. Several of these analyzers are discussed below.

***Quadrupole*** Quadrupole mass analyzers can be found in a plethora of instruments (such as triple-quadrupole mass spectrometers). In fact, triple-quad instruments have been the workhorse in pharmaceutical company laboratories and will continue to be widely utilized. A unique feature to the quadrupole analyzer is that several can be put in tandem within the same mass spectrometer in order to perform very specialized analyses involving both *m/z* and structural determination.

A typical quadrupole analyzer is also known as a mass filter because of its *m/z* selection capability. A direct current (DC) voltage and a superimposed radio-frequency (RF) potential is set across four parallel rods. The field across the poles dictates ion (*m/z*) transmission to the detector. Ions moving into the field inside the poles will oscillate based on their *m/z* values. Only ions of a specific *m/z* will pass through the filter given a particular RF potential, and only by scanning the RF field can a full mass spectrum be obtained. Quadrupoles (Q) can be used in tandem for MS/MS experiments to give not just *m/z* values but also much needed structural information. A specific *m/z* ion can be selected in Q1 and fragmented in Q2, and fragment (or product) ions can be scanned in Q3, for example. Q1, Q2, and Q3 are described in more detail in Section 2.4. Many hybrids using quadrupole analyzers are possible and will be addressed later.

***Ion Trap*** The behavior of ions inside an ion trap is very similar to that of a quadrupole analyzer. Ions can be stored, accumulated, and ejected from an ion trap to produce tandem MS data [56]. Tandem MS experiments can give the much sought-after molecular structural information. Because of its ability to conduct tandem MS experiments of complex mixtures of peptides (such as generated by trypsin digestion), the ion trap is perhaps the most common analyzer found in current mass spectrometers used for proteomics research. The linear ion trap (LIT) has been replacing many other ion storage analyzers because they have higher ion-trapping capacities to increase overall sensitivity, slower scanning functions to increase resolution, and $MS^n$ capabilities for structural elucidation.

***Quadrupole Ion Trap*** The first commercial quadrupole ion trap (QIT), a type of linear ion trap, was introduced in the early 1980s [57]. A QIT stores ions within a three-dimensional RF quadrupole field and has an *m/z* range of approximately 3000. The QIT consists of two endcaps and a ring electrode in replacement of one pair of conventional quadrupole rods. As in the other analyzers previously discussed, ions of a specific *m/z* (or *m/z* range) can be trapped by creating stable trajectories while simultaneously removing unwanted ions through the creation of unstable trajectories. The operation and fundamental concepts of the quadrupole ion trap are described elsewhere [34].

***Orbitrap*** The Orbitrap is patented technology by Thermo Electron Corporation (San Jose, CA) and is a newer analyzer that is capable of selection and storage of ions with a specific (or range) of *m/z* values through the separation of ions in an oscillating electric field. A schematic and details of the Orbitrap can be found elsewhere [58, 59]. A primary advantage to this analyzer over others is its mass accuracy and high-resolution capabilities that are beginning to rival those values conventionally only obtainable by Fourier transform–ion cyclotron resonance (FT-ICR), which requires the use of an expensive superconducting magnet.

***Time-of-Flight Analyzer*** The time-of-flight (TOF) analyzer became popular once again (on the heels of many successful years of the quadrupole analyzer in gas chromatography–mass spectrometry (GC-MS) instruments) after the introduction of MALDI, an ionization technique that produces pulsed packets of ions whose *m/z* analysis is very amenable to the TOF analyzer. A concise history of TOF analyzers can be found elsewhere [53]. Although Wiley and McLaren are the most popular for introducing a TOF mass spectrometer that addressed and improved the spatial and kinetic energy distributions on mass resolution [60], others too had either suggested and/or built TOF instruments [61–66].

In the most basic terms, the TOF analyzer is the simplest type of mass analyzer in mass spectrometry. The basic components necessary for a TOF-MS instrument include an ion source/extraction region, a drift region (tube) for separation of ions according to their velocity (proportional to their *m/z*), and a detector. Ion separation is based on their mass and velocity according to the kinetic energy equation. The lighter ions arrive at the detector sooner than the heavier ions because they have

higher velocities (the ions are given the same voltage, which directly translates to the same kinetic energy).

***Other Analzyers*** *FT-ICR* The Fourier transform–ion cyclotron resonance (FT-ICR) analyzer is becoming more commonplace in analytical laboratories ever since its introduction more than 30 years ago [67]. Commercial versions are now available from vendors such as Thermo Electron Corporation (San Jose, CA) and Bruker Daltronics (Billerica, MA). These instruments can be very costly. On the other hand, mass accuracies of 1–2 ppm and resolution greater than 100,000 are routine [68].

*Q-TOF* The TOF analyzer can often be found placed on the back end of a quadrupole (Q-TOF). The Q-TOF combines the mass selectivity of the quadruple with the resolution of the (nondisciminate mass selection) TOF analyzer. MS/MS data can be acquired with the Q-TOF instrument giving the much desired fragmentation (i.e., structural identification) of a compound.

*TOF/TOF* The TOF analyzer can often be found placed on the back end of another TOF (TOF/TOF) [69]. MS/MS data can be acquired via sequential steps. First, upon selection of a precursor ion that has been separated according to its *m/z* in the first TOF region (TOF1), subsequent fragmentation occurs in a collision-induced dissociation (CID) or postsource decay (PSD) region. In this region ions are decelerated, fragmented, and reaccelerated. These fragment ions are then analyzed in the second TOF region (TOF2) where they separate according to their *m/z*, pass through a reflectron, and impact a detector. The combination of MALDI with MS/MS capability has made the TOF/TOF a more powerful tool to the mass spectrometrist or proteomics researcher than ever before.

### 2.4 MS Scanning Acquisition Methods

There are a variety of modes in which to operate triple-quad mass spectrometers and hybrids such as the Q-Trap (quadrupole ion trap) as offered by Applied Biosystems (Foster City, CA) [70]. Oftentimes MALDI-based mass spectrometers are unable to confirm posttranslation modifications (PTMs) of compounds, and false-positive identifications are common as the number of PTMs increases (i.e., database size increases). Triple-quadrupole and hybrid versions, however, add a selectivity dimension because of the versatility of the analyzer(s). The array of scanning acquisition methods discussed herein are popular among the Q-Trap [70] and triple-quad mass spectrometers for targeted analysis [i.e., selected reaction monitoring (SRM) acquisition for PTM analysis] as such in the pharmaceutical and proteomics industries. Figure 1 lists many of these tandem MS modes along with pictorial descriptions concerning the ion behavior and selection within each quadrupole and/or ion trap [71].

***Q1, Q2, and Q3 Analyzers*** Q1 is the first quadrupole that ions will encounter. Typically, this quadrupole is used to scan all masses to yield a full mass spectrum. Q1 can also be used to select an ion of a specific *m/z* for further introduction into the second quadrupole, Q2, for CID. CID allows deliberate fragmentation of ions to study structures such as peptide sequence analysis and common moieties in pharmaceutical drugs and metabolites. Q3 is the third quadrupole and is used to scan the fragment ions resulting from the CID of a specific *m/z* ion in Q2, or it can be used to only analyze specified point masses from selected ion fragmentation pathways as input by the user.

A *Q1 scan* consists of a full scan using only the first quadrupole. Typically, the user sets a predetermined mass range, and a mass spectrum is produced of ion intensity verses *m/z* of all detected ions. A *Q1 multiple-ion scan* is a scan of only specified masses as input by the user. A *Q3* scan is analogous to the *Q1 scan* above, except that a full scan is performed using the third quadrupole Q3. A *Q3 multiple-ion scan* is similar to the *Q1 multiple-ion scan*, with the exception that the recorded mass spectrum is the result of a scan of the third quadrupole for specified masses.

***Precursor Ion*** Precursor ion scan mode is an MS/MS full scan in which Q1 sweeps a specified mass range, Q2 acts as a collision cell where precursor ions undergo fragmentation, and Q3 is fixed to only transmit one specific product ion to the detector. This can be used in many analyses when searching for an ion that generates a specific *m/z* fragment or product ion (e.g., functional group).

***Product Ion*** When searching for all the fragments or products of a specific precursor ion, a product ion scan (the most common tandem MS experiment) is implemented. This MS/MS scan involves fixing Q1 to only transmit a specific precursor ion. This chosen ion subsequently undergoes fragmentation in a collision cell (such as in Q2), and Q3 sweeps a mass range and analyzes the resulting product ions.

***Neutral Loss*** A neutral loss experiment is yet another tool to track an ion before and after the loss of a neutral group (i.e., phosphate group). This mode consists of both Q1 and Q3 sweeping a mass range with a fixed mass apart. Only an ion signal will be registered when

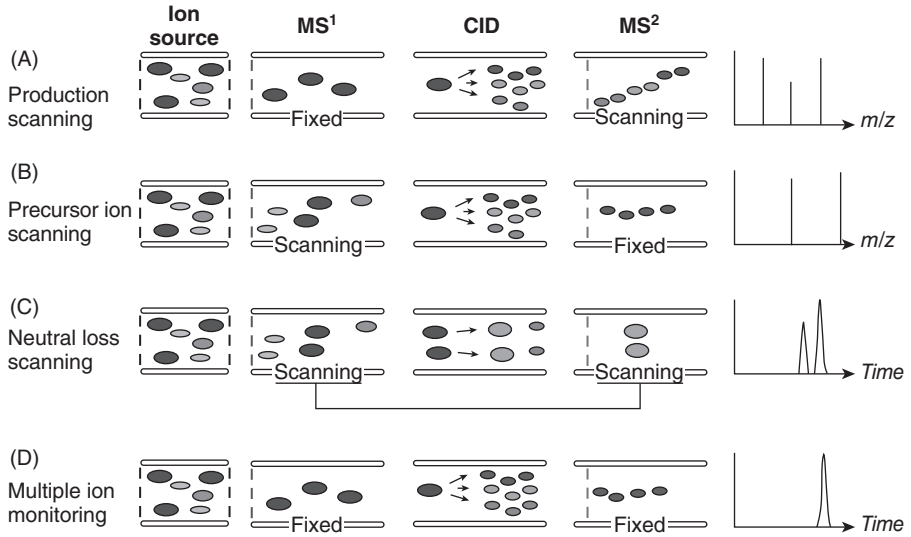


**Figure 1.** Schematic representation of various types of tandem MS experiments. (A) Product ion scanning generates fragment ion spectra for the identification of the amino acid sequence of specific peptides. (B) Precursor ion scanning is typically used to detect a subset of peptides in a sample that contains a specific functional group. (C) Neutral loss scanning also detects peptides that contain a specific functional group and is more often used in the detection of peptides with a specific PTM such as phorphorylation. (D) MRM consists of a series of short experiments in which one precursor ion and one characteristic fragment ion are selected by MS and $MS^2$, respectively. MRM is used for the detection of a specific analyte with known fragmentation properties in complex samples. (Reprinted with permission from AAAS and ref. 71.)

the chosen ion (by Q1) loses the specified neutral group input by the user. Oftentimes many possibilities exist for following a specific neutral loss because the charge of the ions dictates the mass of the neutral group; if there is question as to which charge state the ion has adopted (i.e., +1, +2, +3, +4, . . . ), then all the possibilities must be entered. For example, a phosphate group can be identified by the loss of mass 98, 49, 32.7, and 24.5 Da for charges of +1, +2, +3, and +4, respectively.

***Selected Reaction Monitoring/Multiple Reaction Monitoring*** Selected reaction monitoring (SRM), otherwise known as multiple reaction monitoring (MRM) in certain manufacturer's instruments, is as popular, if not more, than neutral loss scans because of its selectivity (precursor/fragment pair), sensitivity (lower chemical background), and quantitative capabilities. For drug discovery applications an SRM or MRM method is developed for a target compound. In proteomics this mode is often used to search for posttranslational modifications (PTMs). One caveat is that the user must predict or know beforehand the precursor and product ions. The method is optimized for a specific precursor ion (usually in positive ion mode, but can also be in negative ion mode) in Q1, and that specific ion is fragmented into smaller products inside Q2. One specific product ion of a particular *m/z* is chosen and selected inside Q3; only ions of the selected *m/z* are sent to the detector. The SRM/MRM technique can be very selective if designed to be informative as to the identity of the drug compound, and can dramatically decrease both analysis and method development time.

***Linear Ion Trap Modes*** There are numerous modes used in linear ion traps (LIT), which can be found in hybrid instruments also utilizing quadrupole analyzers, as is the case in the Q Trap. A few of the more widely used modes are discussed below.

*Enhanced MS (EMS)* This mode directs the ions from the ion source to Q3 for collection. Upon being scanned out of Q3, the ions are subsequently detected to constitute an enhanced mass spectrum.

*Enhanced Product Ion (EPI)* Precursor ions are selected in Q1 and transmitted into Q2 for CID with a reactant gas. The product ions are transmitted/collected in Q3, wherepuon they are scanned out of Q3 to constitute an enhanced product ion spectrum.

*Enhanced Resolution (ER)* Precursor ions are selected in Q1 and pass through Q2 and into Q3 without frag-

mentation. The slowest scan rate and smallest mass window are then selected to scan the precursor ions out of Q3 in order to produce a better resolved mass spectrum.

*MS/MS/MS (MS3)* After Q1 precursor ions are fragmented in Q2, the ions are transmitted and collected in Q3. Q3 ions then collide with a collision gas (such as nitrogen) and fragment, producing another generation of fragment ions, resulting in an MS/MS/MS experiment.

### 2.5 Common Mass Spectrometer Configurations

Mass spectrometry plays a vital role in numerous online, combined techniques where sample is first purified and separated on a column, with the separated components subsequently analyzed by MS. Mass spectrometers can also analyze material that has been purified offline, reducing the complexity of the mass spectrometer by removing the chromatography aspect. Both methods can be very successful, in any case, and configurations and examples are addressed below.

***Liquid Chromatography/Mass Spectrometry/Mass Spectrometry*** When a liquid chromatography system is placed inline and on the front end of a tandem mass spectrometer, the integrated technique becomes known as LC/MS/MS, one of the most widely used and most powerful MS techniques today. A SciFinder search of "LC/MS/MS and drugs" resulted in over 3600 hits in an October 2008 query. In our laboratory, we have used LC/MS/MS to identify up to 20 proteins in an equimolar protein mixture at approximately 100 fmol/protein using an Applied Biosystems 4000 Q-Trap (nanoflow) LC/MS/MS instrument [72]. LC/MS/MS is used not only in proteomics but also drug discovery [9, 10, 73, 74]. Following elution of analyte off of a column, analytes are then sprayed into the mass spectrometer and subjected to MS/MS analysis for identification.

For proteomics applications, typically the more abundant ions in each scan are subjected to MS/MS via CID (inside Q2) and transmitted into Q3; the product ions are then scanned out of Q3 into the detector. SRM or MRM experiments can be used to target specific post-translational modifications (PTMs) at specific sites of amino acids on peptides.

***Capillary Electrophoresis/Mass Spectrometry*** Although not as extensively used as LC/MS/MS, capillary electrophoresis/mass spectrometry (CE/MS) is another technique that has found a unique role in measuring chiral purity of pharmaceutical compounds. Briefly, components travel though a capillary tube containing an aqueous buffer by a DC potential imposed across the entire length of the capillary. Electrophoretic separations occur based on differences in solute mobilities. Small amounts/volumes (nanoliters) of material are typically injected into the electrically positive end of the capillary, and the charged analyte components migrate toward the negative electrode through the capillary. Typically, a UV detector is placed inline between the two electrodes near the negative electrode. The mass spectrometer is placed on the back end of the CE instrument such that analyte is eluted into the mass spectrometer.

Most studies to date make enantiomeric separation possible through the use of β-cyclodextrin or a derivatized form of β-cyclodextrin as the chiral selector.

***Gas Chromatography/Mass Spectrometry*** Gas chromatography/mass spectrometry, or GC/MS, is one of the older, simpler, and more robust MS techniques. GC/MS is widely used to analyze the molecular weight of organic and inorganic compounds that are thermally stable and volatile or can be made more volatile through reactions with the proper components. Liquid samples are injected into a heated sample port where the sample is vaporized and is carried by a gas such as helium onto the column. As the components elute from the column, the mass of the compounds is measured in the mass spectrometer. Ionization of the analyte is possible by bombarding the analyte with electrons, thereby forming ionized products that are then separated in the mass analyzer. Quantitation is also possible when using standards, where peak area varies linearly with concentration. In pharmaceutical analysis, GC/MS has been used to study dietary supplements and to quantify ephedrine alkaloids [75].

***Matrix-Assisted Laser Desorption Ionization–Time-of-Flight MS*** Matrix-assisted laser desorption ionization, better known as MALDI, is most commonly coupled to time-of-flight (TOF) analyzers. MALDI-TOF is the workhorse for most proteomic experiments in which large numbers of samples are analyzed with the intent of identifying proteins. MALDI-TOF allows users to analyze samples in a high-throughput manner, where analysis of a sample may take only a few seconds compared to an hour for a typical LC/MS run. Because only a little of the sample is consumed by this process and remains bound to the sample target plate onto which it is placed, the sample may be saved for reanalysis at a later date if needed. A plethora of studies are available in the literature regarding analysis of pharmaceutical compounds via MALDI-TOF MS. A SciFinder search for "MALDI-TOF and drugs" alone resulted in over 1400 hits in a search performed in October 2008.

***Liquid Chromatography-MALDI*** Offline coupling of liquid chromatography (LC) to MALDI-TOF is becoming increasingly popular. For example, a large protein mixture can be digested with a protease such as trypsin and then fractionated by LC, spotted at specific timed intervals on a MALDI plate, and the corresponding fractions/spots analyzed by MALDI-TOF MS/MS. As in LC/MS/MS, this fractionation process removes many competitive ionization processes among analyte ions that contributes to the preferential detection of the most abundant ions in conventional MALDI-TOF MS. LC-MALDI gives complementary data as LC/MS/MS with ESI, however, oftentimes, results in unique peptides not seen in ESI and vice versa. One advantage of LC-MALDI over LC/MS/MS with ESI is that the samples can be retained for reanalysis at a later date, whereas with conventional LC/MS/MS the sample is spent during the one-time analysis. LC-MALDI is a more recent technological advance, and a recent SciFinder search resulted in over 80 hits in an October 2008 query, much smaller compared to LC/MS/MS or MALDI of drugs.

***Surface-Enhanced Laser Desorption/Ionization–TOF MS*** Surface-enhanced laser desorption/ionization (SELDI) is a specialized form of MALDI-TOF MS in which chromatography is essentially performed on a chip (e.g., MALDI sample target plate) from which laser desorption occurs [76]. SELDI-TOF MS is primarily used as a screening tool to identify biomarkers associated with pathologies in biological samples or, in other words, used as a diagnostic tool in the early detection of diseases. In theory, each surface on the chip retains proteins according to their chemical properties. Briefly, samples originating from crude sources such as serum, plasma, urine, cell lysate, and so forth can be applied to the various surfaces and then analyzed by MALDI-TOF MS. Additionally, these types of biologically active surfaces enable molecular recognition studies such as receptor–ligand, protein–DNA, and antibody–antigen to name a few. A review article in 2003 outlined many successful studies with SELDI, some of which include SELDI as a diagnostic tool and its use in biological applications [77]. Although the technique itself has its advantages and limitations, SELDI generates patterns more so than the identification of specific proteins.

***Desorption/Ionization on Silicon Mass Spectrometry*** Desorption /Ionization on Silicon (DIOS) is a matrix-free technique where the analyte is adsorbed to a thick (micrometers), UV-absorbing porous surface on a silicon chip [78]. These surfaces can be engineered to have high UV absorptivity by using a specific etching procedure [79] and stabilizing hydrosilylation method. Some advantages of the DIOS technique include no matrix peaks in mass spectrum as with conventional MALDI, simple sample preparation, reusable sample surfaces, and low picomole detection limits [80]. Some drug analysis applications of DIOS include that of atmospheric pressure–DIOS, or AP-DIOS, and the detection of low picomole amounts of midazolam and propranolol [81] and monitoring drugs in biological fluids such as with staurosporine [82].

***Ion Mobility Spectrometry*** Ion mobility spectrometry (IMS) is a very popular technique used in national defense applications, such as in the detection of explosives and chemical agents, due to its compact size, high sensitivity, ease of use, and fast analysis speed (millisecond time scale) [83]. In pharmaceutical research, IMS is used in cleaning validation studies to confirm that a specific cleaning process has removed the active pharmaceutical drug from a specific piece of equipment.

An IMS instrument is very similar to a basic TOF design, except that drift gas is introduced in an opposite direction to that of the direction of the ion flight path. Ionization of the analyte is often initiated by a radioactive source such as $^{63}$Ni within a defined electric field and drift region. More recently, a high field asymmetric waveform ion mobility spectrometer (FAIMS) was coupled to nano-ESI for bioanalysis in drug discovery [84]. FAIMS measures differences in ion mobility at high electric fields, separating ions before introduction into the MS. This ability helps relieve the LC of needing high resolution, and in the above example LC was bypassed completely. Separation of the drug morphine from the metabolite glucuronide was accomplished prior to introduction into the MS, an important point because the metabolite can often convert back into the parent drug in the ion source of the MS [84]. Moreover, signal-to-noise (S/N) ratio saw a three-fold improvement.

***Accelerator MS*** Accelerator MS, or AMS, is best known for its use in $^{14}$C dating in archeology, for example. AMS uses very small sample amounts (sub-milligram), concentrations between $10^{-12}$ and $10^{-16}$, and analysis time of less than 1 h. Samples with ages of tens of thousand of years can be determined within these paramaters above. Other areas of importance utilizing AMS include atmospheric processes, ocean circulation, and even biomedical studies. Isotopes such as $^{10}$Be, $^{36}$Cl, $^{26}$Al, $^{41}$Ca, and $^{129}$I have also been used in dating rock (for Be, Cl, and Al), studying metabolic processes in living systems (Al), and tracing nuclear waste migration from nuclear storage and reprocessing plants as mentioned in recent review articles [85, 86]. In short, AMS most often uses an electrostatic tandem accelerator to count indi-

vidual atoms from rare isotopes from a sample following acceleration energies in the range of 0.2–40 MeV. The technical aspects and common applications of AMS have been published previously [87]. Sensitivity is dramatically increased due to improvement in S/N by decreasing the background levels substantially.

With regard to the current relevant application of biomedicine, isotopes such as $^{3}H$ and $^{14}C$ are common. The use of AMS is limited in this field, however, due to the cost and size of the instrumentation. There is a current push to reduce the size and complexity of AMS instrumentation, thereby making this technique more applicable to biomedical studies. One study showed that AMS provides a factor of $10^3$ improvement in sensitivity with milligram-sized $^{3}H$ samples compared to conventional decay counting [88]. Another study used $^{26}Al$ as an AMS isotope to study the metabolism of aluminum under physiological conditions [89]. Moreover, $^{41}Ca$ has been used as an isotope to study calcium metabolism in the skeleton [90].

## 3 SAMPLE PREPARATION

A plethora of preparation methodologies exist for treatment of pharmaceutical specimens prior to analysis in a mass spectrometer, depending on the sample complexity. The importance of sample preparation cannot be overemphasized because the cost and quality of the data are in fact dependent on the state of the sample prior to analysis. Discussed below are several preparations commonly used in chromatographic analysis and drug development [91]. However, the list is not intended to be an all-inclusive one containing every preparation in the literature. The techniques listed below are extraction processes. Preextraction and postextraction methods are not addressed in this section, however, they are equally important. Preextraction steps primarily deal with the addition of internal standards and sampling, and postextraction steps include sample treatment such as concentration, pH, and solubility.

### 3.1 Protein Precipitation

Protein precipitation (PPT) is a very established technique used, in many cases, to remove proteins from samples containing small molecules in preparation for analysis. PPT often uses organic solvents, heat, salt, or acid to reduce the protein concentration. A mixture of organic solvent(s) and treatment agent(s) is agitated to catalyze protein aggregation. The analyte (e.g., small molecule) remains immersed in the supernatant and can be separated from the mixture via filtration or centrifugation.

### 3.2 Solid-Phase Extraction

Solid-phase extraction (SPE) operates under the same principle as separation by chromatrography. As in chromatrography, the most common mode of SPE is reverse phase. The goal of SPE is the separation of analyte from biological matrix via retention and subsequent elution. In reverse-phase SPE, organic molecules bind to a solid support as an aqueous phase containing the organic molecules passes by. The most retentive molecules bind to the column, whereas the least retentive molecules are in higher abundance in the flowing mobile aqueous phase. Many reviews have been written detailing applications of SPE [92].

### 3.3 Liquid–Liquid Extraction

Liquid–liquid extraction (LLE) is applied to small-molecule recovery and identification. Efficient analyte extraction is obtained when a targeted small molecule diffuses between two immiscible organic layers upon mixing. Choice of organic solvents/additives, boiling points, containers, and automation are all important factors that should be carefully weighed before a final decision.

### 3.4 Other Sample Preparations

Turbulent-flow liquid chromatography (TF-LC) is a form of online chromatography that offers higher flow separations because it is not limited by diffusion as in traditional chromatography separations. The time scale of the experiment can be much faster, higher resolved peaks obtained, and less ionization suppression [93]. TF-LC often allows for little or no prior sample preparation methodologies other than filtration or centrifugation because complex samples can be injected onto the column. Other high-throughput, online sample cleanup strategies for metabolites include microdialysis [94] and liquid-phase membrane extraction [95].

## 4 ROLE OF PROTEOMICS IN PHARMACEUTICAL ANALYSIS

Pharmaceutical analysis encompasses a wide variety of targeted molecules and subsequent applications. One class of compounds is proteins. Mass spectrometry (e.g., MALDI and ESI) has revolutionized the study of proteins as seen by the shear number of articles in the literature in the past 15 years. Many review articles have been written in regard to numerous proteomes and biomarker discovery [96–98], drug discovery [99], PTMs [100], and quantitation [101, 102]. Approaches involving

both gel electrophoresis and gel-free techniques can be readily found in the literature [103, 104]. After a protein is identified, questions then arise as to relative abundance in control/diseased samples, PTMs at specific amino acids, and more generally how to process the amount of data that can be generated in a typical MS experiment. These are all important issues in pharmaceutical proteomic analysis.

### 4.1 Protein Identification

Protein identification is one of the first and most often requested items by researchers to personnel in our MS proteomics facility (and in most proteomics facilities), whether the sample is in the form of a stained gel band or in solution. Although gel electrophoresis of proteins gives information such as molecular weight range of the proteins and differentiates the presence of proteins between samples (e.g., diseased versus control), they do not offer the ability to identify each protein; techniques such as Edman sequencing and MS are required to either identify by protein sequence or with Web databases, respectively. N-terminal Edman sequencing can be very expensive, ranging in price from a few dollars to greater than $20.00 per amino acid for any given peptide or protein. MS is routinely used for protein identification, due to the low cost per sample (not per amino acid) and rapid turnaround time.

Bottom-up and top-down approaches (identification based on peptides or proteins, respectively) dominate proteomic research using MS, with more top-down approaches being introduced each year with improving commercial instrumentation [105, 106]. Identification by MS through bottom-up proteomics relies on the formation of peptides from enzymatic cleavage and their corresponding ion fragmentation patterns for amino acid sequence (MS/MS). The cost of identifying a protein by MS as opposed to Edman sequencing can be substantial. For example, a 50-amino-acid protein might cost up to $1000 for complete sequence information using Edman sequencing techniques, whereas it might cost only $50 for identification by MS. The assumption for MS analysis is made that the protein has previously been sequenced, or "known," and is found in a published protein database such as the National Center for Biotechnology Information (NCBI). Much more additional work would be required to positively identify an unknown protein by MS, whereas N-terminal protein sequencing is not dependent on databases and can positively identify protein sequence.

When surveying a large number of samples and wanting to quickly identify proteins in a cost-effective manner, MS (most often MALDI-TOF) is the preferred technique. More information is provided in Section 4.4 in regard to the various methods employed in MS protein identification via software and database matching.

In most cases, only after protein identification does PTM identification play a key role. In the section below, case studies are described that detail the identification of several common types of PTMs in identified proteins.

### 4.2 Posttranslational Modifications

The identification of posttranslational modifications (PTMs) is one of the more recent and biologically important advances in proteomics and mass spectrometry. No longer does an experiment have to conclude at the identification of a protein but rather can continue to the investigation of site-specific PTMs and, to some extent, the relative quantitation of these modifications. It is well known that PTMs play a vital role in protein recognition studies. One of the challenges is identifying PTMs in vivo when a heterogeneous mixture may be present and/or the PTM form is in low abundance relative to the unmodified form. Moreover, a protein may only have certain PTMs when it is in a certain biological or stressed state. Proteomics encompasses PTMs for many reasons, one being that PTMs can act as molecular switches to modify protein biological activity. The end result is that PTMs can be very difficult to identify and are dependent on the state of the protein or sample prior to extraction, isolation, and/or purification.

There are close to 200 mammamlian PTMs studied in the literature, however, we are focusing on 3 of the more popular ones, that is, phosphorylation, glycosylation, and nitration. Any PTM adds a mass shift, either positive or negative, to an amino acid. Mass accuracy is higher in the lower mass region of a mass spectrometer for peptides as opposed to whole proteins, thereby making the bottom-up proteomics approach the more common platform for protein and PTM identification.

***Phosphorylation*** Protein phosphorylation is one of the most abundant and most studied PTMs known in mammalian cells, as evident in a recent review article [107]. Phosphorylation occurs principally on serine, threonine, and tyrosine residues [108]. Routine identification of phosphorylation sites is usually accomplished via MS/MS, which is capable of identifying the amino acid modification site within a peptide or protein, as opposed to MS, which only gives the molecular weight. In the case of phosphorylation, the addition of a phosphate group results in a molecular weight increase of 80 Da ($HPO_3$). MS is useful in this case for identifying the mass shift, although MS/MS is useful in giving pseu-

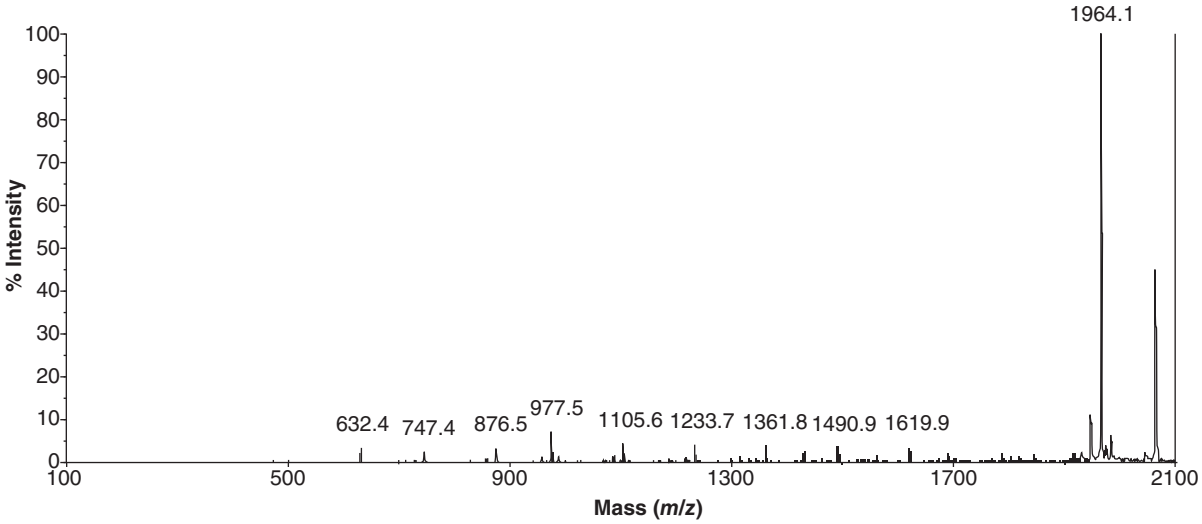


**Figure 2.** Predominant loss of phosphoric acid (98 Da) at *m/z* 1964.1 originates from a phosphorylated peptide at *m/z* 2062 under postsource decay (PSD) conditions in a MALDI-TOF/TOF mass spectrometer. Peptide sequence can be deduced from the fragment ions at multiple masses observed lower in the mass region.

dosequence information via fragment ions that ideally show the loss of the phosphate group from a specific amino acid site. Figure 2 demonstrates the typical predominant loss of phosphoric acid ($H_3PO_4$) (98 Da due to additional water loss of 18 Da) at *m/z* 1964.1 from the phosphorylated peptide at *m/z* 2062 during postsource decay (PSD) or collision-induced dissociation (CID) in a MALDI-TOF/TOF mass spectrometer. Peptide sequence can be deduced from the fragment ions observed at multiple masses lower in the mass region. Because of the likelihood of false positive using automated software programs to identify the phosphate group in an MS spectrum, it is still the norm to manually interpret the MS/MS spectrum for the loss of phosphoric acid as well as additional novel fragment ions such as phosphotyrosine at *m/z* 216.

A typical MS phosphoprotein standard is α/β-casein, which has several phosphorylation sites. In order to reduce complexity, it is very common to isolate/purify a phosphopeptide or phosphoprotein before MS analysis, and one method includes the use of immobilized metal affinity chromatography (IMAC) [109]. IMAC columns are the most common method of phosphopeptide enrichment to date and are incredibly simple to use. Selectivity of IMAC toward phosphate groups is governed by electrostatic interactions between the positively charged metal ions and the negatively charge phosphate groups. Most manufacturer protocols state that the entire process for phosphopeptide sample loading, binding, and eluting is on the order of approximately 10 min. Improvements to both offline and online IMAC methods have gradually been made over the years, ranging from choice of column material to choice of metal ion. The resultant phosphopeptide-enriched sample can then be analyzed by MALDI or ESI.

One alternative to IMAC is metal oxide affinity chromatography (MOAC) mircotips, which use support materials such as titanium dioxide [110] or zirconium dioxide [111] in combination with mass spectrometric analysis of the phosphopeptides. Another option for phosphopeptide enrichment is the use of strong cation exchange chromatography (SCX) [112].

Many other articles have been written detailing methods for identifying phosphorylated peptides through both unique sample preparation and MALDI-TOF or LC/MS analysis. Li and co-workers have investigated different ionic-liquid matrices for improved phospholipid analysis, where the extent of prompt fragmentation is decreased and alkali–metal–ion adducts are increased [113]. Asara and co-workers used ammonium salts to enhance the detection of phosphopeptides in MALDI MS [114]. Yang and co-workers have combined an alternative matrix, 2′, 4′, 6′-trihydroxyacetophenone (THAP) with diammonium citrate to aid in the detection of phosphopeptides generated from proteolytic digestion by MALDI MS [115]. The use of negative ion mode in MALDI, which detects negatively charged ions, has also been exploited to enhance detection of phosphopeptides [116]. Moyer and co-workers have investigated the fragmentation of phosphopeptides by atmospheric pressure MALDI and ESI/ion trap mass spectrometry and have proposed possible gas-phase mechanisms of dephosphorylation of phosphotryrosine [117]. A popular loss of 98 Da was observed

from phosphotyrosine, phosphoserine, and phosphothreonine. DeGnore investigated the fragmentation of phosphopeptides in an ion trap mass spectrometer as well, as seen in Figure 3, noting that the loss of phosphoric acid ($H_3PO_4$) from phosphoserine does not depend on the charge state [118].

***Glycosylation*** Another abundant PTM of proteins is that of glycosylation. Glycosylation is the process by which a glycan is covalently bound to a protein. According to Apweiler et al. [119], as many as 50% of all proteins are glycosylated. Fortunately, protein glycosylation occurs at specific and predictable amino acids. N-linked glycosylation occurs when a glycan is bound to the side chain of asparagines. However, this only occurs when the asparagine is of the consensus sequence Asn-X-Y where X is any amino acid except proline and Y is serine, threonine, or cysteine. O-linked glycosylation occurs when a glycan is covalently bound to serine or threonine. This can occur regardless of the amino acid sequence of the protein.

Unfortunately, the analysis of glycoproteins, particularly intact, is difficult due to the heterogeneity of the glycan moieties. Often there is no one individual method for performing mass analysis on glycoproteins, but rather a variety of different methods may be employed by the investigator. Both MALDI-TOF MS and ESI MS have been routinely used for analysis. ESI MS has the advantage in that it can be coupled to LC or CE.

Because glycopeptides often have lower ionization efficiency than their nonglycosylated counterpart, enrichment is often a requirement prior to analysis. Glycopeptides can be isolated with the use of lectin affinity methods [120]. LC columns can be fabricated in which the stationary phase contains a variety of lectins that allows not only for the capture of glycopeptides a nd glycoproteins but also allows for the separation of the heterogeneous glycopeptide/glycoprotein mixture [121, 122].

Due to the chemical complexity of glycans, higher order MS scans must often be performed in order to elucidate their chemical structure. MALDI-TOF/TOF and triple-quadrupole ESI mass spectrometers only allow for MS/MS analysis. Ion traps, however, allow for higher order MS analysis such as $MS^3$. Recently, hybrid mass spectrometers, such as the combination of triple-quad and linear ion traps, have become popular and allow for $MS^3$ analysis. This is especially popular in elucidating the structures of branched glycans [123].

***Nitration*** Lately there has been an interest in the biological role of tyrosine nitration to 3-nitrotyrosine. The nitro group on tyrosine is a PTM, which is the result of oxidative stress and therefore has the potential as a biomarker for oxidative stress of proteins [124]. 3-Nitrotyrosine has been associated with many human diseases including Alzheimer's disease [125], atherosclerosis [126], inflammatory diseases [127], and stroke [128]. Although there is still some debate as to the mechanisms by which nitrotyrosine is formed in vivo, it is believed to result from the presence of peroxidases in the presence of hydrogen peroxide and nitrite [129].

3-Nitrotyrosine-containing proteins that have been separated via two-dimensional gel electrophoresis can be visualized by performing a Western blot using commercially available 3-nitrotyrosine antibodies [130]. However, positive identification of 3-nitrotyrosine-containing proteins can only be confirmed by MS/MS analysis, which also allows for locating the specific tyrosine within a protein that has been nitrated.

Both ESI and MALDI-TOF MS have been used for the direct analysis of 3-nitrotyrosine-containing proteins [131]. Often 3-nitrotyrosine-containing proteins are first enriched in order to reduce the complexity of the sample and thereby improving the quality of the MS data. Enrichment may be performed using immunoaffinity methods. Typically proteins of interest are first digested with a protease and then the resulting 3-nitrotyrosine peptides are captured through the use of immunoaffinity beads [132]. 3-Nitrotyrosine peptides may also be selected for by first reducing 3-nitrotyrosine to 3-aminotyrosine, which is then coupled to a cleavable biotin affinity tag [133].

Nitration of tyrosine results in a molecular weight increase of 45 Da. The resulting molecular weight for 3-nitrotyrosine is not isobaric with any other amino acid and therefore allows one to easily identify and locate 3-nitrotyrosine within a peptide sequence via MS/MS. However, the choice of ionization will have a major effect on the mass spectral results. Analysis of 3-nitrotyrosine-containing peptides by LC/MS/MS, which incorporates the use of ESI, typically results in the nitro group on tyrosine remaining intact during ionization and CID [134]. The MS/MS of a 3-nitrotyrosine-containing peptide by ESI is illustrated in Figure 4. A mass difference of 208 Da is observed between fragment ions $y_9$ and $y_{10}$, which corresponds to 3-nitrotyrosine. Most protein database search software such as MASCOT and SEQUEST allow for the addition of modified amino acids.

Unlike ESI, MALDI ionization of 3-nitrotyrosine yields a variety of photodissociated products. This unique characteristic of 3-nitrotyrosine is believed to be the result of photodissociation caused by the wavelength of the laser light being very close to that of the absorption maximum wavelength for 3-nitrotyrosine as reported by Petersson [135]. MALDI MS typically yields ions of the type $[M+H]^+$ along with the photodis-

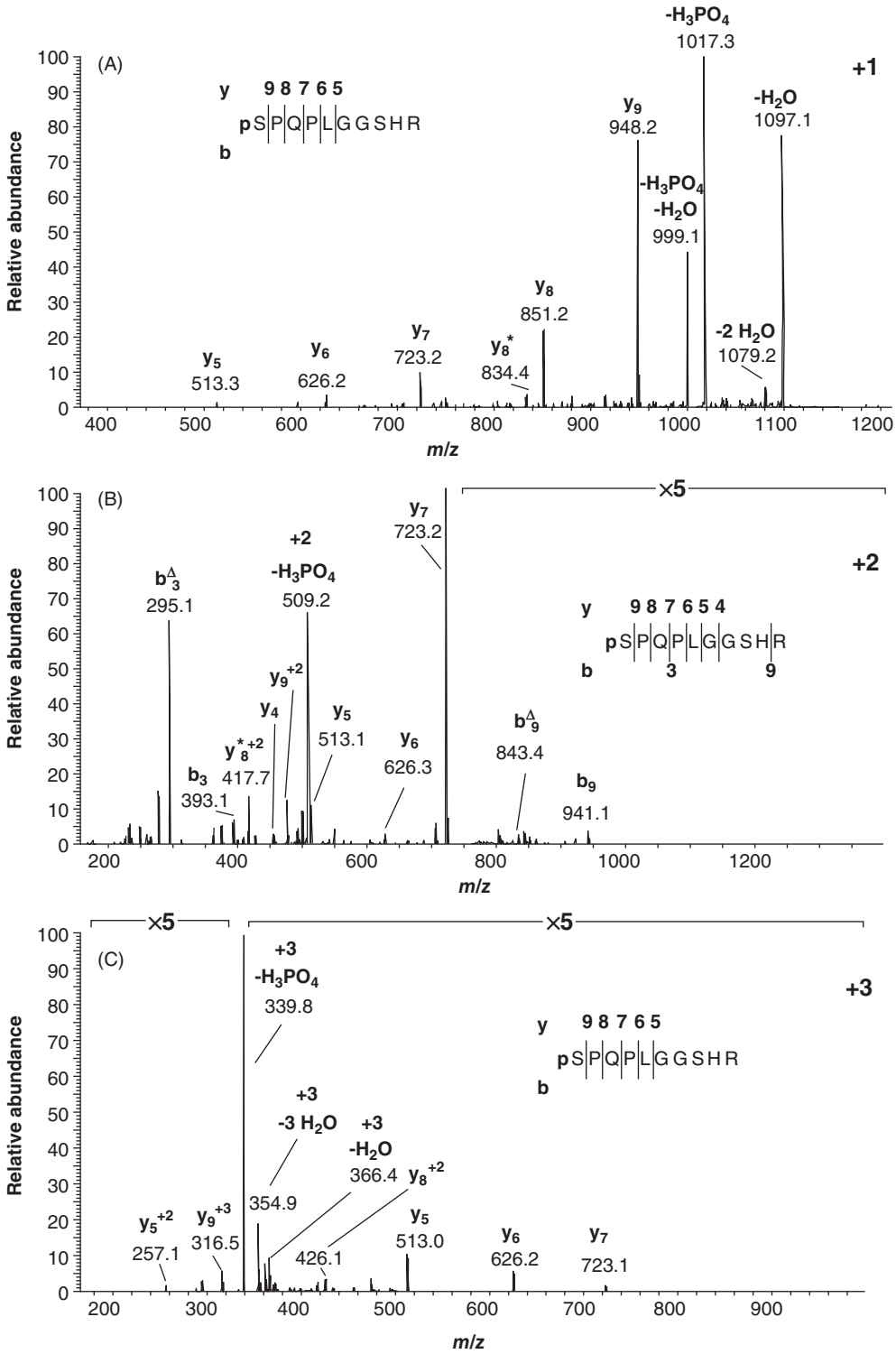


**Figure 3.** Loss of $H_3PO_4$ (98 Da) is not charge state dependent for phosphoserine-containing peptides. (A) $MS^2$ spectrum of the singly charged ion at *m/z* 1115.5. (B) $MS^2$ spectrum of the doubly charged ion at *m/z* 558.3. (C) $MS^2$ spectrum of the triply charged ion at *m/z* 372.5. Fragmentation of all charge states exhibits a loss of $H_3PO_4$. Fragments with an asterisk indicate loss of 17 Da (likely $NH_3$) from corresponding *b* or *y* fragment. (Reprinted with permission from Elsevier and ref. 118.)

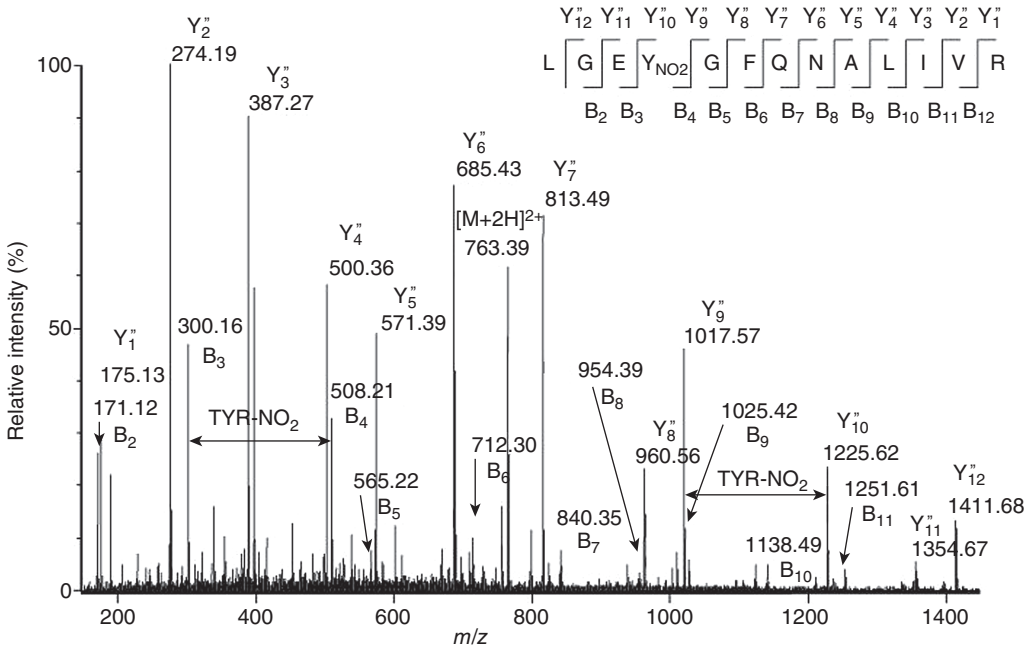


**Figure 4.** MS/MS of a 3-nitrotyrosine-containing peptide under ESI conditions. A mass difference of 208 Da is observed between fragment ions $y_9$ and $y_{10}$, which corresponds to 3-nitrotyrosine. (Reprinted with permission from John Wiley & Sons and ref. 135.)

sociated products $[M+H-16]^+$, $[M+H-30]^+$, and $[M+H-32]^+$, which corresponds to $NO_2 \rightarrow NO$, $NO_2 \rightarrow N$, and $NO_2 \rightarrow NH_2$ by the loss of oxygen as reported by Sarver [136]. This is illustrated in Figure 5, a MALDI-TOF mass spectrum of angiotensin II in which the tyrosine has been nitrated to 3-nitrotyrosine. The peak at *m/z* 1341.7 is that of the nitrated angiotensin and those at *m/z* 1309.8, 1311.8, and 1325.8 are those of the photodegraded products formed during the ionization process.

### 4.3 Quantitation

Mass spectrometry is not well known for supplying quantitation data because it most often depends on the ionization potential of compounds that can dramatically vary, regardless of the amount of each analyte in a mixture. Progress is being made, however, and the most common techniques utilizing mass spectrometry use isotope chemistry.

***Isotope-Coded Affinity Tags/Isobaric Tags for Relative and Absolute Quantitation*** Rapid identification of proteins has been one of the most important attributes of mass spectrometry for the advancement of proteomics and biological systems. However, in order to completely understand the functions of identified proteins in a system, it is also important to understand the expression of these proteins. Isotopic labeling of peptides and proteins followed by mass spectral analysis has become the most common method for quantifying peptides and proteins by mass spectrometry. Several methods such as isotope-coded affinity tags (ICAT) and isobaric tag for relative and absolute quantitation (iTRAQ) have been commercialized and are readily available as kits for the quantification of peptides and proteins. Recently, these methods have been adapted for the quantification of amino acids, thereby allowing one to perform amino acid analysis by mass spectrometry.

The use of ICAT reagents was first reported by Aebersold [137] in 2001. ICAT reagents are comprised of three parts: an iodoacetamide group for covalently binding the reagent to the thiol group of cysteine residues, an isotopically labeled linker for MS quantification, and a biotin tag to isolate the ICAT-reacted peptides. The linker of the ICAT reagents incorporates either hydrogen or deuterium as the stable isotope. Recently, ICAT has been reformulated such that nine carbons are exchanged in order to obtain a molecular weight difference of nine mass units. Figure 6 illustrates the chemical structure of this ICAT reagent. Peptides and proteins can be differentially labeled by labeling one set of proteins with the carbon-12 linker and another protein set with the carbon-13 linker. The relative ratios

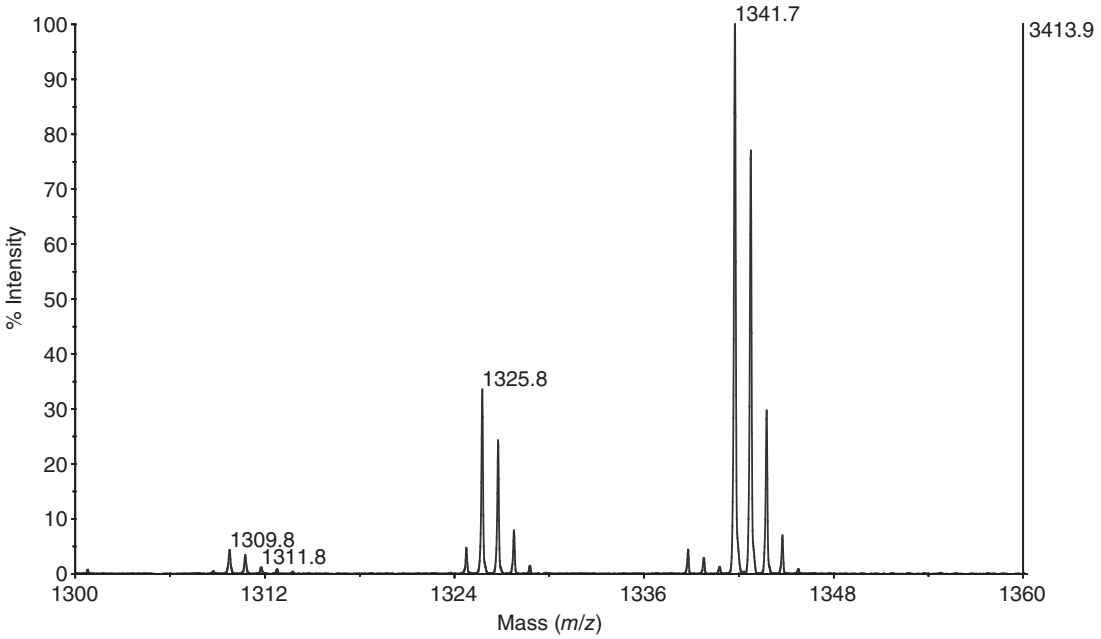


**Figure 5.** Nitrated angiotensin is observed at *m/z* 1341.7, and peaks at *m/z* 1309.8, 1311.8, and 1325.8 are those of the photodegraded products formed during the ionization process in a MALDI-TOF/TOF mass spectrometer in our MS laboratory at University of Texas Medical Branch.

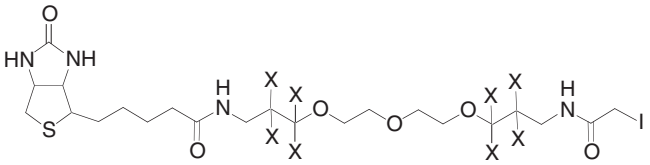

**Figure 6.** Chemical structure of an ICAT molecule.

of the $^{12}$C- to $^{13}$C-labeled peptides thereby correspond to the ratios of the peptides in the two different sets.

Recently, iTRAQ reagents have been replacing ICAT as the method of quantification [138, 139]. Unlike ICAT, which allows for only two sets of proteins to be differentially labeled, iTRAQ allows up to eight different sets of peptides or proteins to be labeled. This is accomplished by covalently attaching an isobaric tag to an amine reactive group of a peptide. The isobaric tag consists of a reporter group and a balance group. The reporter group ranges in mass from 113 to 121 mass units and the balance mass ranges from 192 to 184 mass units, thus giving a total mass of 305 mass units for each reagent. Because the total mass of the tag is identical for each reagent, identical peptides in two different sets that are differentially labeled will also have equal masses. However, during MS/MS fragmentation, quantification is measured by the difference in ion abundance of the reporter groups.

***Stable Isotope Labeling Amino Acids in Cell Culture*** Other methods for incorporating stable isotopes have been reported. Stable isotope labeling amino acids in cell culture (SILAC) allows for the incorporation of isotopically labeled amino acids into cell cultures [140, 141]. The growth media of one cell culture will contain normal amino acids while the growth media of another cell culture will contain an isotopically modified amino acid. Chemically identical peptides can be differentiated by the mass spectrometer because of their amino acid mass differences and therefore observed as peptide pairs. The ratio of peak intensities for the peptide pairs corresponds to the ratio of their protein abundance.

### 4.4 Bioinformatics

Bioinformatics is a very broad subject, therefore, we will briefly focus only on its application to MS data in the biopharmaceutical proteomics arena and, more specifically, in those topics that are more practical to the actual "users" of MS instrumentation.

Generally speaking, bioinformatics has progressed proteomics to the field it is today, that is, one that includes a large amount of data analysis, handling, interpretation, and storage. Advancements in the computer industry are largely responsible for the advent of bioinformatics in many areas of science. In the 1970s, the use

of mass spectrometers on a larger level was beginning to emerge. However, computers and user interfaces did not exist in the form we know and use today. Consequently, in the last 25 years mass spectrometry has grown not only as a result of MS instrumentation advancements but also as a result of computer hardware and software innovations.

***Database Platforms*** There are many options in the choice of database platform for protein identification, and many are dependent on mass spectrometer manufacturer. Some of these software packages include Mascot [142], PepProbe [143], OMSSA [144], RMET-Luck [145], SEQUEST [146], and Tandem [147]. Each platform usually requires various parameters to be completed by the user. Some parameters are dependent (e.g., mass tolerances) and others are independent (e.g., database choice such as NCBI, SwissProtein, MSDB; enzyme; variable/fixed modifications) on the instrumental conditions. Many novel platforms (i.e., algorithms) have been written for spectral peak analysis and for matching experimental data to theoretical data, and a more detailed description of the latest techniques can be found elsewhere [103].

***Standards and Data Management*** Because of the enormous amount of data generated by MS proteomic experiments, one typically only views a list of identified proteins and sometimes identified peptides in published research articles. The raw data is often omitted, and other critical data sets either buried in the supplemental data section online or also omitted from the print publication. The Association of Biomolecular Resource Facility (ABRF) (http://abrf.org) recently published guidelines it would prefer to see implemented in scientific journals, but each journal has its own set of guidelines for submission of publications. Moreover, from the standpoint of commercial MS instrumentation and software packages, many are unique in the data file extensions used, making it sometimes difficult to disseminate data in a universal fashion.

There has been an ongoing initiative throughout the last few years to improve the standardization of data distribution, which would simplify data sharing, an especially important factor in cases where large proteomic studies are being conducted by multiple institutions and where data is submitted and combined in online databases from various researchers with MS instrumentation from different manufacturers. The Proteomics Standards Initiative (PSI) (http://psidev.sf.net/), an extension of the Human Proteome Organization (HUPO), already has defined the use of software language, XML, as a standard for MS data [148]. XML offers the primary advantage of being shared across multiple software platforms. Other common formats include FASTA, mzXML, and pepXML. BioPAX (http://biopax.org) is an alternative to the PSI standard and is defined in both metabolic network and molecular interaction levels [97].

Other techniques have been introduced to reduce the amount of data gathered in a typical MS experiment in proteomic applications, and many of them involve improvements made to the very popular SEQUEST search engine. Beer introduced Pep-Miner for clustering MS/MS data in order to increase peptide identification with higher confidence [149]. There are a number of other statistical approaches also made to benefit SEQUEST for improving peptide/protein identification confidence while reducing false-positive rates [150, 151].

## 5 DRUG METABOLITES

Metabolite is a term applied to a compound that is typically pharmacologically less active than its parent compound, although this is not always the case. Over the years, mass spectrometry has grown tremendously in this field alone by allowing the analysis of trace amounts of compounds, such as metabolites, in complex mixtures. One of the earliest MS techniques of choice was GC/MS because of the MS requirement for molecules to be in the gas phase. As ionization techniques for MS evolved, however, the coupling of LC to MS has become the dominant analytical instrument for the study of pharmaceutical compounds such as metabolites in biological matrices. A recent review article details several LC/MS/MS applications for metabolite structural information such as ziprasidone for the treatment of schizophrenia and voriconazole as an antifungal agent to treat multiple pathogens [152].

Drug degradation products in pharmaceutical formulations are also widely studied by LC/MS, and numerous examples can be found in review articles [153]. Common degradation products originate from processes such as oxidation, rearrangement, hydrolysis, dehydration, and adduct formation, to name but a few, and are commonly analyzed in media such as human plasma. Oftentimes products can be formed that are combinations of the above-mentioned processes, further complicating analysis by MS. These cases then rely heavily on tandem MS ($MS^n$) techniques to elucidate structure. It is becoming more common to develop a single MS platform that simultaneously analyzes for more than one parent drug compound or its associated metabolites. Many case studies have been compiled in a recent review article [74].

Another review article was written on systematic LC/MS metabolite identification in the drug discovery

industry, focusing on tandem MS methods for the identification of compounds [154]. Many aspects are addressed, such as targeted product ion analysis, determination of modification sites, and identification of dosed compounds, with an emphasis on specific MS modes such as collecting precursor ion scan and neutral-loss data, product ion analysis of potential metabolites, structural determination by $MS^n$, and accurate mass measurement.

More recently an article was published detailing the use of the LTQ/Orbitrap, at a resolving power of 60,000 and high mass accuracy less than 2 ppm, for human liver microsomal metabolites of carvedilol [155]. Also, rat tissue sections from liver, spleen, and muscle were analyzed for erlotibiv and its metabolites with a MALDI quadrupole TOF mass spectrometer [156]. In addition, this study examined the feasibility of obtaining quantitative data from tissue sections analyzed by MALDI MS/MS.

More novel MS techniques have recently been applied to metabolite identification. Desorption electrospray ionization (DESI), a newer atmospheric pressure ionization technique introduced by Cooks and coworkers, can be used to quickly ascertain the molecular weight of metabolites from complex biological matrices such as urine [157]. Also, electrospray photoionization (ESPI) LC/MS was used for analyzing cyclodextrin and pharmaceuticals and their binding interactions [158]. All components of a mixed-polarity liquid sample were measured using this new ionization technique. Binding constants were also determined for cyclodextrin to various compounds.

## 6 EMERGING FIELDS FOR DRUG ANALYSIS BY MS

As is the case in all areas of science, there are specialty areas within fields that drive technology and engineering further into the future—the incorporation of mass spectrometry into biopharmaceutical science is no different. The advent of MS and its vital role in drug discovery is so common now that other applications are available for pharmaceutical compounds. Several budding biopharmaceutical fields utilizing mass spectrometry are discussed below, with the caveat that this section only serves as a brief introduction using several examples from the literature.

### 6.1 Environmental MS

Environmental MS often refers to the field of contaminants, whether they are present in soil, sludge, water, or atmosphere. Many reviews have been written on pharmaceuticals and the environment and several examples will be highlighted in this section [159, 160].

One field that garners a lot of public attention is the amount of antibiotics in wastewater; many cases have been studied where wastewater treatment did not completely remove antibiotics such as erythromycin, roxithromycin, to name but a few compounds. The technique of choice is often SPE-LC/MSMS, that is, solid-phase extraction (SPE) combined with LC/MS/MS. A mixture of antibiotics was analyzed in two studies that found concentrations in the low microgram/liter range in wastewater [161] and in the microgram/kilogram range in sewage sludge [162]. Another study found several antibiotics such as four macrolide antibiotics and six sulfonamides in the low nanogram/liter range [163]. LC/MS/MS has also been used to identify pharmaceuticals such as erythromycin, carbamazepine, fluoxetine, and diphenhydramine in reclaimed water used for irrigation [164]. Although there are a plethora of methods utilizing LC/MS/MS for pharamaceuticals, there are fewer using GC/MS and CE/MS; most GC/MS methods in the literature have comparable sensitivity limits to LC/MS/MS, however, CE/MS sensitivity is generally 1 to 2 orders of magnitude lower than most LC/MS/MS methods.

Other contaminants gaining notoriety in environmental MS analysis and discussed in detail elsewhere include sunscreens/organic UV filters, drinking water disinfection by-products, old and new flame retardants, benzotriazoles, naphthenic acids, pesticides, perchlorate, and last but not least, microorganisms [160]. Microorganisms have gained particular importance since the threat of bioterrorism became more widespread, and Fenselau has pioneered much of the MS-based identification of different organisms using both ESI and MALDI ionization techniques [165–167].

### 6.2 Forensic MS

A SciFinder search in July 2011 resulted in over 3600 hits when using the terms "MS," "forensics," and "drugs." Forensic analysis is a very broad field encompassing DNA analysis, trace evidence, and drugs [168]. Many reviews have covered these topics. The scope of this section, however, is to introduce some choice examples of drugs or pharmaceutical compounds relevant in forensic MS analysis. It is important to keep in mind that with most of these drug analyses in forensic MS, the medium in which the drugs are present can be very complex and may require one or more sample preparation strategies to extract and identify the compounds in the matrix (e.g., body fluids).

Forensic science has long used GC/MS for the analysis of volatile compounds, and even nonvolatile

compounds that are made more volatile through the use of chemical derivatization. The identification of ethanol in biological fluids such as blood is very common in postmortem procedures. The identification of narcotics such as morphine, heroin, cocaine, methadone, and their associated metabolites have been found in urine, blood, saliva, and even hair using LC/MS/MS techniques such as SIM/MRM in combination with sample preparation strategies such SPE and, more recently, solid-phase microextraction (SPME). Recently, anabolic steroids and their associated metabolites have been detected and quantitated in hair by GC/MS/MS [169].

### 6.3 Miniaturization

Miniaturization can increase throughput and drive down costs in many service facilities, however, the cost for developing miniature technologies can be high. The need to miniaturization mass spectrometer platforms is driven in many cases by the aspect of field portability. Miniature and field-portable MS instruments can be very useful in both military and civilian settings, either in a battlefield environment or in a hospital room. In the first case, chemical weapons testing can be done with miniature ion trap mass spectrometers made popular by Cooks at Purdue University who wrote a review article describing the various miniature MS analyzers available [170]. In the second case, Cotter and co-workers at Johns Hopkins University School of Medicine are designing miniature instruments that could be used either in a hospital room setting to detect potentially known biomarkers for disease [171] or in a mobile laboratory to identify bacterial organisms such as *Bacillus* spores [167] , or even for Mars surface analysis [172]. In all cases the application drives the need for miniaturization of existing technologies.

In the case of biopharmaceutical analysis, miniaturization many times takes the form of sample preparation methods that can increase throughput at lower costs per sample in a more timely fashion. This part can be extremely valuable in the case where drug candidates or potential biomarkers are being screened. Some key players in many miniature sample preparation platforms include microfluidics as well as chromatography. Other advances in MS instrumentation can also benefit pharmaceutical science by improving the mass accuracy, which can reduce false-positive rates for small-molecule and drug identification.

The most common form of miniaturization used across research laboratories is the use of 96-well sample plates fitted for automated microliter or even nanoliter volume extraction, increasing throughput while reducing cost. For example, less solvent waste is generated in this format, which will reduce laboratory waste disposal expenses. These well-plates can then be transferred, either directly or indirectly, into LC/MS or MALDI-TOF mass spectrometers for rapid analysis.

The advent of analyzing smaller amounts of material has also directly affected LC/MS instrument manufacturing, that is, in the form of nanospray ESI tips and nanoflow LC pumps (e.g., nano-LC/MS/MS). Identification can be made on an assortment of compounds through the injection of only a few microliters onto a capillary HPLC column, with solvent flow in the nanoliter/minute regime. The analysis of very small volumes of material is very convenient in cases where the user is searching for PTMs and may have to run several MS experiments to confirm the presence of a modified amino acid on a peptide, for example. Having enough sample for multiple runs also allows for experimental precision to be calculated, and the sample can be saved for analysis at a later time.

## 7 CONCLUSIONS AND FUTURE OUTLOOK

Mass spectrometry has experienced tremendous growth throughout numerous scientific disciplines in the last few decades. In the early days of MS, instruments were so large that they occupied entire rooms. Nowadays, there are many options in regard to the type and size of available instruments depending on the application. Advancements in MS technology have pushed the field into applied biological sciences and pharmaceutical research in such a way that was barely imaginable many years ago. Moreover, MS instrumentation is at the point where almost any researcher can quickly learn how to use a mass spectrometer because the software is so user friendly and the hardware is well engineered, much different than when MS instrumentation was primarily used by only experts in the field in the 1950s to early 1970s. The advent of MALDI and ESI ionization techniques, as well as TOF, quadrupole, and ion trap analyzers, have totally transfigured MS with the detection of higher mass molecules, which pushed fields such as proteomics to far horizons. In biopharmaceutical research, low concentrations of analytes can now be detected in very complex mixtures and environments. Clearly, the field of mass spectrometry will continue to play a starring role in this area as well as many other scientific disciplines in the foreseeable future.

## REFERENCES

1. (2011). *Scifinder Scholar*. Washington, DC: American Chemical Society.
2. Lindsay S (1992). *High Performance Liquid Chromatography*. Chichester: Wiley.

3. Meyer VR (1994). *Practical High Performance Liquid Chromatography*. Chichester: Wiley.
4. Krupczynska K, Buszewski B, Jandera P (July 1, 2004). Characterizing HPLC stationary phases. *Anal Chem 76*, 226A–234A.
5. Skoog DA, Leary JJ (1992). *Principles of Instrumental Analysis*, 4th ed. Fort Worth: Saunders College Publishing.
6. Gilpin RK, Gilpin CS (2007). Pharmaceuticals and related drugs. *Anal Chem 79*, 4275–4294.
7. Ardrey B (2003). *Liquid Chromatography-Mass Spectrometry: An Introduction*. West Sussex: Wiley.
8. Korfmacher WA, Cox KA, Ng KJ, Veals J, Hsieh Y, Wainhaus S, Broske L, Prelusky D, Nomeir A, White RE (2001). Cassettee-accelerated rapid rat screen: A systematic procedure for the dosing and liquid chromatography/atmospheric pressure ionization tandem mass spectrometric analysis of new chemical entities as part of new drug discovery. *Rapid Commun Mass Spectrom 15*, 335–340.
9. Xu X, Lan J, Korfmacher WA (October 1, 2005). Rapid LC/MS/MS method development for drug discovery. *Anal Chem 77*, 389A–394A.
10. Ermer J, Vogel M (2000). Applications of hyphenated LC-MS techniques in pharmaceutical analysis. *Biomed Chromatogr*, *14*, 373–383.
11. Song L, Choi K-S, Park Y-M, Kazim AL, Marlar K, Kong E-S, Park D-M, Kim YH, Koo K-H, Chae HZ (2005). Capillary-LC-uESI-MS/MS and nano-LC-nano ESI-MS/MS analysis using a single binary pump capillary LC system: Applications in proteomics. *J Liqu Chromatogr Related Technol 28*, 1271–1289.
12. Schmidt A, Karas M, Dulcks T (2003). Effect of different solution flow rates on analyte ion signals in nano-ESI MS, or: When does ESI turn into nano-ESI? *J Am Soc Mass Spectrom 14*, 492–500.
13. Shen Y, Zhao R, Berger SJ, Anderson GA, Rodriguez N, Smith RD (2002). High-efficiency nanoscale liquid chromatography coupled on-line with mass spectrometry using nanoelectrospray ionization for proteomics. *Anal Chem 74*, 4235–4249.
14. Wang Y, Zhang J, Liu C-L, Gu X, Zhang X-M (2005). Nano-flow multidimensional liquid chromatography with electrospray ionization time-of-flight mass spectrometry for proteome analysis of hepatocellular carcinoma. *Anal Chim Acta 530*, 227–235.
15. Wuhrer M, Koeleman CAM, Deelder AM, Hokke CH (2004). Normal-phase nanoscale liquid chromatography-mass spectrometry of underivatized oligosaccharieds at low-femtomole sensitivity. *Anal Chem 76*, 833–838.
16. Satomi Y, Shimonishi Y, Hase T, Takao T (2004). Site-specific carbohydrate profiling of human transferrin by nano-flow liquid chromatography/electrospray ionization mass spectrometry. *Rapid Commun Mass Spectrom 18*, 2983–2988.
17. Champion M (2008). Personal communication, College Station, TX.
18. Mazzeo JR, Neue UD, Kele M, Plumb RS (December 1, 2005). Advancing LC performance with smaller particles and higher pressure. *Anal Chem 77*, 460A–467A.
19. Swartz M (2004). Ultra performance liquid chromatography: Tomorrow's HPLC technology today. *LabPlus Int 18*, 6–9.
20. Nguyen DT-T, Guillarme D, Rudaz S, Veuthey J-L (2006). Fast analysis in liquid chromatography using small particle size and high pressure. *Separ Sci 29*, 1836–1848.
21. Kay RG, Gregory B, Grace PB, Pleasance S (2007). The application of ultra-performance liquid chromatography/tandem mass spectrometry to the detection and quantitation of apolipoproteins in human serum. *Rapid Commun Mass Spectrom 21*, 2585–2593.
22. Geoghegan KF, Kelly MA (2005). Biochemical applications of mass spectrometry in pharmaceutical drug discovery. *Mass Spectrom Rev*, *24*, 347–366.
23. Johnson BM, Nikolic D, Breeman RBV (2002). Applications of pulsed ultrafiltration-mass spectrometry. *Mass Spectrom Rev 21*, 76–86.
24. Zhao Y-Z, Breemen RBV, Nikolic D, Huang C-R, Woodbury CP, Schilling A, Venton DL (1997). Screening solution-phase combinatorial libraries using pulsed ultrafiltration/electrospray mass spectrometry. *J Med Chem 40*, 4006–4012.
25. Weiboldt R, Zweigenbaum J, Henion J (1997). Immunoaffinity ultrafiltration with ion spray HPLC/MS for screening small-molecule libraries. *Anal Chem 69*, 1683–1691.
26. Schriemer DC, Bundle DR, Li L, Hindsgaul O (1998). Micro-scale frontal affinity chromatography with mass spectrometric detection: A new method for the screening of compound libraries. *Angew Chem Int Ed 37*, 3383–3387.
27. Schriemer DC (December 1, 2004). Biosensor alternative: Frontal affinity chromatography. *Anal Chem 76*, 441A–448A.
28. Zhu L, Chen L, Luo H, Xu X (2003). Frontal affinity chromatography on-line with mass spectrometry: A tool for the binding study of different epidermal growth factor receptor inhibitors. *Anal Chem 75*, 6388–6393.
29. Kaur S, McGuire L, Tang D, Dollinger G, Huebner V (1997). Affinity selection and mass spectrometry-based strategies to identify lead compounds in combinatorial libraries. *J Prot Chem 16*, 505–511.
30. Chu YH, Dunayevskiy YM, Kirby DP, Vouros P, Karger BL (1996). Affinity capillary electrophoresis mass spectrometry for screening combinatorial libraries. *J Am Chem Soc*, *118*, 7827–7835.
31. Sonksen CP, Nordhoff E, Jansson O, Malmqvist M, Roepstorff P (1998). Combining MALDI mass spectrometry and biomolecular interaction analysis using a biomolecular interaction analysis instrument. *Anal Chem 70*, 2731–2736.
32. Kelly MA, LIang HB, Sytwu II, Vlattas I, Lyons NL, Bowen BR, Wennogle LP (1996). Characterization of

SH2-ligand interactions via library affinity selection with mass spectrometric detection. *Biochemistry 35*, 11747–11755.

33. Hofstadler SA, Sannes-Lowery KA, Crooke ST, Ecker DJ, Sasmor H, Manalili S, Griffey RH (1999). Multiplexed screening of neutral mass-tagged RNA targets against ligand libraries with electrospray ionization FTICR MS: A paradigm for high-throughput affinity screening. *Anal Chem 71*, 3436–3440.
34. Gross JH (2004). *Mass Spectrometry: A Textbook*. Berlin: Springer.
35. (1990). Doolittle R, ed. *Methods in Enzymology*. San Diego: Academic Press.
36. (1994). Russell D, ed. *Topics in Mass Spectrometry*. New York: Plenum.
37. Chapman JR (1993). *Practical Organic Mass Spectrometry: A Guide for Chemical and Biochemical Analysis*, 2nd ed. Chichester/New York: Wiley.
38. McLafferty FW, Turecek F (1993). *Interpretation of Mass Spectra*, 4th ed. Mill Valley, CA: University Science Books.
39. Biemann K (1962). *Mass Spectrometry: Organic Chemical Applications*. New York: McGraw-Hill.
40. Hamdan M, Righetti PG (2005). *Proteomics Today: Protein Assessment and Biomarkers Using Mass Spectrometry, 2D Electrophoresis, and Microarray Technology*. Hoboken, NJ: Wiley-Interscience.
41. ASMS. *What is Mass Spectrometry*, www.asms.org.
42. Fenn JB, Mann M, Meng CK, Wong SF, Whitehouse CM (1989). Electrospray ionization for mass spectrometry of large biomolecules. *Science 246*, 64–71.
43. Yamashita M, Fenn JB (1984). Electrospray ion source. Another variation on the free-jet theme. *J Phys Chem 88*, 4451–4459.
44. MacFarlane RD, Torgerson DF (1976). Californium-252-plasma desorption time-of-flight mass spectrometry. *Int J Mass Spectrom Ion Process 21*, 81–92.
45. Barber M, Bordoli RS, Sedgwick RD, Tyler AN (1981). Fast atom bombardment of solids (F.A.B.): A new ion source for mass spectrometry. *Nature 293*, 270–275.
46. Lehrle RS, Robb JC, Thomas DW (1962). A modified time-of-flight mass spectrometer for studying ion-molecule or neutral particle-molecule interactions. *J Sci Instrum 39*, 458–463.
47. Benninghoven A, Jaspers D, Sichtermann W (1976). Secondary-ion emission of amino acids. *Appl Phys 11*, 35–39.
48. Chait BT, Standing KG (1981). A time-of-flight mass spectrometer for measurement of secondary ion mass spectra. *Int J Mass Spectrom Ion Process 40*, 185–193.
49. Bruins AP (1991). Mass spectrometry with ion sources operating at atmospheric pressure. *Mass Spectrom Rev*, *10*, 53–77.
50. Tanaka K, Ido Y, Akita S, Yoshida T (1987). Protein and polymer analysis up to m/z 100,000 by laser desorption time-of-flight mass spectrometry. Proceedings of the Second Japan-China Symposium on Mass Spectrometry, Osaka, Japan, Matduda H, Liang, X-T, eds., Bando Press, Osaka pp. 185–188.
51. Tanaka K, Ido Y, Akita S, Yoshida Y, Yoshida T (1988). Protein and polymer analyses up to m/z 100,000 by laser ionization time-of-flight mass spectrometry. *Rapid Commun Mass Spectrom 2*, 151–153.
52. Karas M, Hillenkamp F (1987). Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. *Anal Chem 60*, 2299–2301.
53. Cotter RJ (1997). *Time-of-Flight Mass Spectrometry*. Washington, DC: American Chemical Society.
54. Robb DB, Covey TR, Bruins AP (2000). Atmospheric pressure photoionization: An ionization method for liquid chromatography-mass spectrometry. *Anal Chem 72*, 3653–3659.
55. Glish GL, Burinsky DJ (2008). Hybrid mass spectrometers for tandem mass spectrometry. *J Am Soc Mass Spectrom 19*, 161–172.
56. Cooks RG, Glish GL, McLuckey SA, Kaiser RE (1991). Ion trap mass spectrometry. *Chem Eng News 25*, 26–41.
57. Stafford GCJ, Kelley PE, Syka JEP, Reynolds WE, Todd JFJ (1941). Recent improvements in and analytical applications of advanced ion trap technology. *Int J Mass Spectrom Ion Process 60*, 85–98.
58. Hardmanand M, Makarov A (2003). *Anal Chem 75*, 1699.
59. Scigelova M, Makarov A (2006). Education & training: Orbitrap mass analyzer—Overview and applications in proteomics. *Proteomics 6*, 16–21.
60. Wiley WC, McLaren IH (1955). Time-of-flight mass spectrometer with improved resolution. *Rev Sci Instr 26*, 1150–1157.
61. Stephens WE (1946). A pulsed mass spectrometer with time dispersion. *Phys Rev 69*, 691.
62. Stephens WE (1952). U.S. Patent 2,612,607.
63. Cameron AE, Eggers DF (1948). Ion "velocitron." *Rev Sci Instrum 19*, 605–607.
64. Keller R (1949). Mass spectra obtained from measurement of the time of flight. *Helv Phys Acta 22*, 386–388.
65. Wolff MM, Stephens WE (1953). Pulsed mass spectrometer with time dispersion. *Rev Sci Instrum 24*, 616–617.
66. Katzenstein HS, Friedland SS (1955). New time-of-flight mass spectrometer. *Rev Sci Instrum 26*, 324–327.
67. Comisarow MB, Marshall AG (1974). Fourier transform ion cyclotron resonance spectroscopy. *Chem Phys Lett 25*, 282–283.
68. Zhou M, Veenstra TD (2008). Mass spectrometer: *m/z* 1983–2008. *Biotechniques 44*, 667–670.
69. Medzihradszky KF, Campbell JM, Baldwin MA, Falick AM, Juhasz P, Vestal ML, Burlingame AL (2000). The characteristics of peptide collision-induced dissociation using a high performance MALDI-TOF/TOF tandem mass spectrometer. *Anal Chem 72*, 552–558.

70. AppliedBiosystems (2003). 4000 Q Trap LC/MS/MS System Operator's Manual.
71. Domon B, Aebersold R (2006). Mass spectrometry and protein analysis. *Science 312*, 212–217.
72. Haag AM (2008). *Unpublished work*. Galveston, TX: University of Texas Medical Branch.
73. Dixon SP, Pitfield ID, Perrett D (2006). Comprehensive multi-dimensional liquid chromatographic separation in biomedical and pharmaceutical analysis: A review. *Biomed Chromatogr 20*, 508–529.
74. Srinivas NR (2006). Applicability of bioanalysis of multiple analytes in drug discovery and development: Review of select case studies including assay development considerations. *Biomed Chromatogr 20*, 383–414.
75. Marchei E, Pelligrini M, Pacifici R, Zuccaro P, Pichini SJ (2006). A rapid and simple procedure for the determination of ephedrine alkaloids in dietary supplements by gas chromatography-mass spectrometry. *J Pharm Biomed Anal 41*, 1633–1641.
76. Hutchens TW, Yip TT (1993). New desorption strategies for the mass spectrometric analysis of macromolecules. *Rapid Commun Mass Spectrom 7*, 576–580.
77. Issaq HJ, Conrads TP, Prieto DA, Tirumalai R, Veenstra TD (2003). SELDI-TOF MS for diagnostic proteomics. *Anal Chem 75*, 149A–155A.
78. Wei J, Buriak JM, Siuzdak G (1999). Desorption-ionization-MS on porous silicon. *Nature 399*, 243–246.
79. Shen Z, Thomas JJ, Averbuj C, Broo KM, Engelhard M, Crowell JE, Finn MG, Siuzdak G (2001). Porous silicon as a versatile platform for laser desorption/ionization-MS. *Anal Chem 73*, 612–619.
80. Laiko VV, Taranenko NI, Berkhout VD, Musselman BD, Doroshenko VM (2002). Atmospheric pressure laser desorption/ionization on porous silicon. *Rapid Commun Mass Spectrom* 1737–1742.
81. Huikko K, Ostman P, Sauber C, Mandel F, Grigoras K, Franssila S, Kotiaho T, Kostiainen R (2003). Feasibility of atmospheric pressure desorption/ionization on silicon mass spectrometry in analysis of drugs. *Rapid Commun Mass Spectrom 17*, 1339–1343.
82. Trauger SA, Go EP, Shen Z, Apon JV, Compton BJ, Bouvier ESP, Finn MG, Siuzdak G (2004). High sensitivity and analyte capture with desorption/ionization mass spectrometry on silylated porous silicon. *Anal Chem 76*, 4484–4489.
83. Eiceman GA, Stone JA (November 1, 2004). Ion mobility spectrometers in national defense. *Anal Chem 76*, 390A–397A.
84. Hatsis P, Brockman AH, Wu J-T (2007). Evaluation of high-field asymmetric waveform ion mobility spectrometry coupled to nanoelectrospray ionization for bioanalysis in drug discovery. *Rapid Commun Mass Spectrom 21*, 2295–2300.
85. Hellborg R, Skog G (2007). Accelerator mass spectrometry. *Mass Spectrom Rev 27*, 398–427.
86. Kutschere W (2005). Progress in isotope analysis at ultra-trace level by AMS. *Int J Mass Spectrom Ion Process 242*, 145–160.
87. Matteson S (2007). Issues and opportunities in accelerator mass spectrometry for stable isotopes. *Int J Mass Spectrom Ion Processes*, *27*, 470–484.
88. Ognibene TJ, Bench G, Brown TA, Vogel JS (2004). The LLNL accelerator mass spectrometry system for biochemical $^{14}$C-measurements. *Nucl Instrum Methods B 223/224*, 12–15.
89. Yumoto S, Nagal H, Kobayashi K, Tada W, Horikawa T, Matsuzaki H (2004). $^{26}$Al incorporation into the tisues of suckling rats through maternal milk. *Nucl Instrum Methods B 223/224*, 754–758.
90. Jiang S, He M, Dong K, Yue D, Wu S, Xu G, Qin J, You Q, Zheng Y, Guan Y, Liang Q, Zhang G, Zhao X, Wang Q, Liu S (2004). The measurement of $^{41}$Ca and its application for the cellular $Ca^{2+}$ concentration fluctuation caused by carcinogenic substances. *Nucl Instrum Methods B 223/224*, 750–753.
91. Chang MS, Ji Q, Zhang J, El-Shourbagy TA (2007). Historical review of sample preparation for chromatographic bioanalysis: Pros and cons. *Drug Dev Res 68*, 107–133.
92. Hennion MC (1999). Solid phase extraction: Method development, sorbents, and coupling with liquid chromatography. *J Chromatogr A 856*, 3–54.
93. Herman JL (2005). The use of turbulent flow chromatography and the isocratic focusing effect to achieve on-line clean-up and concentration of neat biological samples for low-level metabolite analysis. *Rapid Commun Mass Spectrom 19*, 696–700.
94. Bergstrom SK, Markides KE (2002). On-line coupling of microdialysis to packed capillary column liquid-chromatography-tandem mass spectrometry demonstrated by measurement of free concentrations of ropivacaine and metabolite from spiked plasma samples. *J Chromatogr B 775*, 79–87.
95. Kuuranne T, Kotiaho T, Pedersen-Bjergaard S, Rasmussen KE, Leinonen A, Westwood S, Kostiainen R (2003). Feasibility of a liquid-phase microextraction sample clean-up and LC-MS/MS screening method for selected anabolic steroid glucuronides in biological samples. *J Mass Spectrom 38*, 16–26.
96. Brotz-Oesterhelt H, Bandow JE, Labischinski H (2005). Bacterial proteomics and its role in antibacterial drug discovery. *Mass Spectrom Rev 24*, 549–565.
97. Smith JC, Lambert J-P, Elisma F, Figeys D (2007). Proteomics in 2005/2006: Developments, applications and challenges. *Anal Chem 79*, 4325–4344.
98. Graves PR, Haystead TAJ (2002). Molecular biologist's guide to proteomics. *Microbiol Mol Biol Rev 66*, 39–63.
99. Figeys D (August 1, 2002). Proteomics approaches in drug discovery. *Anal Chem 74*, 413A–419A.
100. Bonilla LE, Means GD, Lee KA, Patterson SD (2008). The evolution of tools for protein phosphorylation

site analysis: From discovery to clinical application. *Biotechniques 44*, 671–679.

101. Hopfgartner G, Bourgogne E (2003). Quantitative high-throughput analysis of drugs in biological matrices by mass spectrometry. *Mass Spectrom Rev 22*, 195–214.
102. MacCoss MJ, Matthews DE (August 1, 2005). Quantitative MS for proteomics: Teaching a new dog old tricks. *Anal Chem 77*, 295A–302A.
103. Lambert J-P, Ethier M, Smith JC, Figeys D (2005). Proteomics: From gel based to gel free. *Anal Chem 77*, 3771–3788.
104. Issaq HJ, Veenstra TD (2008). Two-dimensional polyacrylamide gel electrophoresis (2D-PAGE): advances and perspectives. *Biotechniques 44*, 697–700.
105. Kelleher NL (June 1, 2004). Top-down proteomics. *Anal Chem 76*, 196A–203A.
106. Zubarev RA, Kelleher NL, McLafferty FW (1998). Electron capture dissociation of multiply charged protein cations. A nonergodic process. *J Am Chem Soc 120*, 3265–3266.
107. Temporini C, Calleri E, Massolini G, Cacciaianza G (2008). Integrated analytical strategies for the study of phosphorylation and glycosylation in proteins. *Mass Spectrom Rev 27*, 207–236.
108. Hunter T (1991). Protein kinase classification. *Methods Enzymol 200*, 3–37.
109. Porath J, Carlsson J, Olsson I, Belfrage G (1975). Metal chelate affinity chromatography, a new approach to protein fractionation. *Nature 258*, 598–599.
110. Larsen MR, Thingholm TE, Jensen ON, Roepstorff P, Jorgensen TJD (2005). Highly selective enrichment of phosphorylated peptides from peptide mixtures using titanium dioxide microcolumns. *Mol Cell Proteomics 4*, 873–886.
111. Kweon HK, Hakansson K (2006). Selective zirconium dioxide-based enrichment of phosphorylated peptides for mass spectrometric analysis. *Anal Chem 78*, 1743–1749.
112. Ballif BA, Ville J, Beausoleil SA, Schwartz D, Gygi SP (2004). Phosphoproteomic analysis of the developing mouse brain. *Mol Cell Proteomics 3*, 1093–1101.
113. Li YL, Gross ML, Hsu F (2005). Ionic-liquid matrices for improved analysis of phospholipids by MALDI-TOF mass spectrometry. *J Am Soc Mass Spectrom 16*, 679–682.
114. Asara JM, Allison J (1998). Enhanced detection of phosphopeptides in matrix-assisted laser desorption/ionization mass spectrometry using ammonium salts. *J Am Soc Mass Spectrom 10*, 35–44.
115. Yang X, Wu H, Kobayashi T, Solaro RJ, Breemen RBC (2004). Enhanced ionization of phosphorylated peptides during MALDI TOF mass spectrometry. *Anal Chem 76*, 1532–1536.
116. Janek K, Wenschuh H, Bienert M, Krause E (2001). Phosphopeptide analysis by positive and negative ion matrix-assisted laser desorption/ionization mass spectrometry. *Rapid Commun Mass Spectrom 15*, 1593–1599.
117. Moyer SC, Cotter RJ, Woods AS (2001). Fragmentation of phosphopeptides by atmospheric pressure MALDI and ESI/ion trap mass spectrometry. *J Am Soc Mass Spectrom 13*, 274–283.
118. DeGnore JP, Qin J (1998). Fragmentation of phosphopeptides in an ion trap mass spectrometer. *J Am Soc Mass Spectrom 9*, 1175–1188.
119. Apweiler R, Hermjakob H, Sharon N (1999). On the frequency of protein glycosylation, as deduced from analysis of the SWISS-PROT database. *Biochim Biophys Acta 1473*, 4–8.
120. Drake RR, Schwegler EE, Malik G, Diaz J, Block T, Metha A, Semmens OJ (2006). Lectin capture strategies combined with mass spectrometry for the discovery of serum glycoprotein biomarkers. *Mol Cell Proteomics 5*, 1957–1967.
121. Yang Z, Hancock WS (2004). Approach to the comprehensive analysis of glycoproteins isolated from human serum using a multi-lectin affinity column. *J Chromatogr A 1053*, 79–88.
122. Yang Z, Haycock WS (2005). Monitoring glycosylation pattern changes of glycoproteins using multi-lectin affinity chromatography. *J Chromatogr A 1070*, 57–64.
123. Sandra K, Devreese B, Beeumen JV, Stals I, Claeyssens M (2003). The Q-Trap mass spectrometer, a novel tool in the study of protein glycosylation. *J Am Soc Mass Spectrom 15*, 413–423.
124. Ischiropoulos H, Beckman JS (2003). Oxidative stress and nitration in neurodegeneration: Cause, effect, or association? *J Clin Invest 111*, 163–169.
125. Castegna A, Thongboonkerd V, Klein JB, Lynn B, Markesbery WR, Butterfield DA (2003). Proteomic identification of nitrated proteins in Alzheimer's disease brain. *J Neurochem 85*, 1394–1401.
126. Beckman JS, Ye YZ, Anderson PG, Chen J, Accavitti MA, Tarpey MM, White CR (1994). Extensive nitration of protein tyrosines in human atherosclerosis detected by immunohistochemistry. *Biol Chem hoppe Seyler 375*, 81–88.
127. Aluk KS, Miyagi M, Yan L, West KA, Massillon D, Crabb JW, Stuehr DJ (2001). Proteomic method identifies proteins nitrated in vivo during inflammatory challenge. *Proc Natl Acad Sci 98*, 12056–12061.
128. Oury TD, Piantadosi CA, Crapo JD (1993). Cold-induced brain edema in mice: Involvement of extracellular superoxide dismutase and nitric oxide. *J Biol Chem 268*, 15394–15398.
129. Brennan ML, Wu W, Fu X, Shen Z, Song W, Frost H, Vadseth C, Narine L, Lenkiewicz E, Borchers MT, Lusis AJ, Lee JJ, Lee NA, Abu-Soud HM, Ischiropoulos H, Hazen SL (2002). A tale of two controversies: Defining both the role of peroxidases in nitrotyrosine formation in vivo using eosinophil peroxidase and myeloperoxidase-deficient mice, and the nature of

peroxidase-generated reactive nitrogen species. *J Biol Chem 277*, 17415–17427.

130. Zhan X, Desidero DM (2004). The human pituitary nitroproteome: detection of nitrotyrosyl-proteins with two-dimentional western blotting, and amino acid sequence determination with mass spectrometry. *Biochem Biophys Res Commun 325*, 1180–1186.
131. Kanski J, Schöneich C (2005). Protein nitration in biological aging: Proteomic and tandem mass spectrometric characterization of nitrated sites. *Methods Enzymol 396*, 160–171.
132. MacMillan-Crow LA, Thompson JA (1999). Immunoprecipitation of nitrotyrosine-containing proteins. *Methods Enzymol 301*, 135–145.
133. Nikov G, Bhat V, Wishnok JS, Tannenbaum SR (2003). Analysis of nitrated proteins by tyrosine-specific affinity probes and mass spectrometry. *Anal Biochem 320*, 214–222.
134. Lee SJ, Lee JR, Kim YH, Park YS, Park SI, Park HS, Kim KP (2007). Investigation of tyrosine nitration and nitrosylation of angiotensin II and bovine serum albumin with electrospray ionization mass spectrometry. *Rapid Commun Mass Spectrom 21*, 2797–2804.
135. Petersson A-S, Steen H, Kalume DE, Caidahl K, Roepstorff P (2001). Investigation of tyrosine nitration in proteins by mass spectrometry. *J Mass Spectrom 36*, 616–625.
136. Sarver A, Scheffler NK, Shetlar MD, Gibson BW (2001). Analysis of peptides containing nitrotyrosine by matrix-assisted laser desorption/ionization mass spectrometry. *J Am Soc Mass Spectrom 12*, 439–448.
137. Han DK, Eng J, Zhou H, Aebersold R (2001). Quantitative profiling of differentiated-induced microsomal proteins using isotope-coded affinity tags and mass spectrometry. *Nat Biotechnol 19*, 946–951.
138. Zieske LR (2006). A perspective on the use of iTRAQ reagent technology for protein complex and profiling studies. *J Exp Bot 57*, 1501–1508.
139. Ross PL, Huang YN, Marchese JN, Williamson B, Parker K, Hattan S, Khainovski N, Pillai S, Dey S, Daniels S, Purkayastha S, Juhasz P, Martin S, Bartlet-Jones M, He F, Jacobson A, Pappin DJ (2004). Multiplexed protein quantitation in the *Saccaromyces cerevisiae* using amin-reactive isobaric tagging reagents. *Mol Cell Proteomics 3*, 1145–1153.
140. Ong S, Blagoev B, Kratchmarova I, Kristensen D, Steen H, Pandey A, Mann M (2002). Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol Cell Proteomics 1*, 376–386.
141. Ong S, Kratchmarova I, Mann M (2002). Properties of $^{13}$C-sustituted arginine in stable isotope labeling by amino acids in cell culture (SILAC). *J Proteome Res 2*, 173–181.
142. Perkins DN, Pappin DJ, Creasy DM, Cottrell JS (1999). Probability-based protein identification by researching sequence databases using mass spectrometry data. *Electrophoresis 20*, 3551–3567.
143. Sadygov RD, Yates JR (2003). A hypergeometric probability model for protein identification and validation using tandem mass spectral data and protein sequence databases. *Anal Chem 75*, 3792–3798.
144. Geer LY, Markey SP, Kowalak JA, Wagner L, Xu M, Maynard DM, Yang X, Shi W, Bryant SH (2004). Open mass spectrometry search algorithm. *J Proteome Res 3*, 958–964.
145. Fridman T, Razumovskaya J, Verberkmoes N, Hurst G, Protopopescu V, Xu Y (2005). The probability distribution for a random match between an experimental-theoretical spectral pair in tandem mass spectrometry. *J Bioinf Comput Biol 3*, 455–476.
146. Link AJ, Eng J, Schieltz DM, Carmack E, Mize GJ, Morris DR, Garvik B, Yates JR (1999). Direct analysis of protein complexes using mass spectrometry. *Nature Biotechnology 17*, 676–682.
147. Craig R, Beavis RC (2004). TANDEM: Matching proteins with tandem mass spectra. *Bioinformatics 20*, 1466–1467.
148. Orchard S, Hermjakob H, Julian RK, Runte K, Sherman D, Wojcik J, Zhu W, Apweiler R (2004). Common interchange standards for proteomics data: Public availability of tools and schema. *Proteomics 4*, 490–491.
149. Beer I, Barnea E, Ziv T, Admon A (2004). Improving large-scale proteomics by clustering of mass spectrometry data. *Proteomics 4*, 950–960.
150. Anderson DC, Li W, Payan DG, Noble WS (2003). A new algorithm for the evaluation of shotgun peptide sequencing in proteomics: Support vector machine classification of peptide MS/MS spectra and SEQUEST scores. *J Proteome Res 2*, 137–146.
151. Lopez-Ferrer D, Martinex-Bartolome S, Villar M, Campillos M, Martin-Maroto F, Vazquez J (2004). Statistical model for large-scale peptide identification in databases from tandem mass spectra using SEQUEST. *J Anal Chem 76*, 6853–6860.
152. Prakash C, Shaffer C, Nedderman A (2007). Analytical strategies for identifying drug metabolites. *Mass Spectrom Rev 26*, 340–369.
153. Wu Y (2000). The use of liquid chromatography-mass spectrometry for the identification of drug degradation products in pharmaceutical formulations. *Biomed Chromatogr 14*, 384–396.
154. Clarke NJ, Rindgen D, Korfmacher WA, Cox KA (August 1, 2001). Systematic LC/MS metabolite identification in drug discovery. *Anal Chem 73*, 430A–439A.
155. Lim H-K, Chen J, Sensenhauser C, Cook K, Subrahmanyam V (2007). Metabolite identification by data-dependent accurate mass spectrometric analysis at resolving power of 60 000 in external calibration mode using an LTQ/orbitrap. *Rapid Commun Mass Spectrom 21*, 1821–1832.
156. Signor L, Varesio E, Staack RF, Starke V, Richter WF, Hopfgartner G (2007). Analysis of erlotinib and its

metabolites in rat tissue sections by MALDI quadrupole time-of-flight mass spectrometry. *J Mass Spectrom 42*, 900–909.

157. Kauppila TJ, Wiseman JM, Ketola RA, Kotiaho T, Cooks RG, Kostiainen R (2006). Desorption electrospray ionization mass spectrometry for the analysis of pharmaceuticals and metabolites. *Rapid Commun Mass Spectrom 20*, 387–392.
158. Short L, Syage JA (2008). Electrospray photoionization (ESPI) liquid chromatography/mass spectrometry for the simultaneous analysis of cyclodextrin and pharmaceuticals and their binding interactions. *Rapid Commun Mass Spectrom 22*, 541–548.
159. Petrovic M, Hernando MD, Diaz-Cruz MS, Barcelo DJ (2005). Liquid chromatography-tandem mass spectrometry for the analysis of pharmaceutical residues in environmental samples: A review. *J Chromatogr A 1067*, 1–14.
160. Richardson SD (2006). Environmental mass spectrometry: Emerging contaminants and current issues. *Anal Chem 78*, 4021–4045.
161. Yang S, Carlson KH (2004). Solid-phase extraction-high performance liquid chromatography-ion trap mass spectrometry for analysis of trace concentrations of macrolide antibiotics in natural and waste water matrices. *J Chromatogr A 1038*, 141–155.
162. Gobel A, Thomsen A, McArdell CS (2005). *J Chromatogr A 1085*, 179–189.
163. Gobel A, McArdell CS, Suter MJF, Giger W (2004). *Anal Chem 76*, 4756–4764.
164. Kinney CA, Furlong ET, Werner SL, Cahill JD (2006). Presence and distribution of wastewater-derived pharmaceuticals in soil irrigated with reclaimed water. *Environ Toxicol Chem 25*, 317–326.
165. Swatkoski S, Russell SC, Edwards N, Fenselau C (2006). Rapid chemical digestion of small acid-soluble proteins for analysis of Bacillus spores. *Anal Chem 78*, 181–188.
166. Warscheid B, Fenselau C (2003). Characterization of Bacillus spore species and their mixtures using post-source decay with a curved-field reflectron. *Anal Chem 75*, 5618–5627.
167. English RD, Warscheid B, Fenselau C, Cotter RJ (2003). Bacillus spore identification via proteolytic peptide mapping with a miniaturized MALDI TOF mass spectrometer. *Anal Chem 75*, 6886–6893.
168. Brettell TA, Butler JM, Almirall JR (2007). Forensic science. *Anal Chem 79*, 4365–4384.
169. Gambelunghe C, Sommavilla M, Ferranti C, Rossi R, Aroni K, Manes N, Bacci M (2007). Analysis of anabolic steroids in hair by GC/MS/MS. *Biomed Chromatogr 21*, 369–375.
170. Badman ER, Cooks G (2000). Miniature mass analyzers. *J Mass Spectrom 35*, 659–671.
171. Prieto MC, Kovtoun VV, Cotter RJ (2002). Miniaturized linear time-of-flight mass spectrometer with pulsed extraction. *J Mass Spectrom 37*, 1158–1162.
172. Evans-Nguyen T, Becker L, Doroshenko V, Cotter RJ (2008). Development of a low power, high mass range mass spectrometer for Mars surface analysis. *Int J Mass Spectrom Ion Process 278*, 170–177.

# MOLECULAR SCREENING FOR THERAPEUTIC AGENTS

R. Damoiseaux
*Molecular Screening Shared Resource, California NanoSystems Institute, University of California, Los Angeles, California*

## 1 INTRODUCTION

### 1.1 Historical Overview

Sequencing of the human genome in 2001 provided a conceptual infrastructure for understanding the cellular processes fundamental to life and health [1]. Following the advances in decoding the human genome at the sequence level, deciphering the function of individual proteins encoded by the genome and elucidating their respective biochemical pathways is now well on its way. As a result, cellular processes and their involvement in disease and discovery of therapeutic compounds for the various disease indications are becoming better understood. Consequently, the postgenomic era filled the drug target pipeline of the pharmaceutical industry with an unprecedented number of targets [2, 3].

In the pregenomic era, screening for therapeutic compounds of up to 10,000 compounds per diem was the scientific standard for drug discovery until the 1980s and 1990s. As more and more drug targets became available, drug discovery companies began pushing for more intelligent, productive, high-precision discovery [4]. A solution to these needs was molecular screening, the combination of molecular biology and drug screening. By means of molecular biology, it became possible to engineer assay systems that mimic a disease system and subject it to screening for therapeutic compounds [5].

Molecular screening soon became automated and enabled the use of very complex assay systems without the personnel burden normally associated with screening for bioactive compounds. [6, 7] In addition to saving labor, automation offered other compelling advantages. Experiments could be reproduced perfectly and performed on a much larger, more expansive scale. It became feasible to transfer experiments from low- to high-throughput platforms for screening of large compound sets and then transfer the experiments back again for hit validation and characterization purposes.

At the beginning of the transition to automation, most of the assays were biochemical. These assays were typically performed on single isolated proteins. Rapidly, the technology expanded to accommodate assay systems based on cells. Now, even whole organism platforms [8] are a promising new possibility, and current automation platforms are capable of handling complicated experimental protocols, opening the door to the use of molecular screening in general basic research.

These research applications of screening include functional genomics screens using small interfering ribonucleic acid (siRNAs) [9], short hairpin RNA (shRNAs) [10], and overexpression libraries [11], just to name a few. This chapter gives an introduction to the general concepts of molecular screening as they relate to discovery of therapeutic agents and highlight some research applications of molecular screening. It will give a general introduction to the concepts of molecular screening, provide an overview of the general work flow, outline different screening approaches, give an overview of the assay development and validation process, and discuss the basic readouts for assays. Then we will have a close look at compound libraries that can be

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

paired with the assays during screening. We will have a look at the results, how to explore the properties of the resulting hits and how to deal with false positives. Later on, target discovery will be examined, in particular of using small molecules and siRNAs, shRNAs, and complementary deoxyribonucleic acid (cDNAs), which rely on similar concepts and could be said to use molecular screening for research purposes. Last, but not least, we shall have a look at how the data acquired during screening can be stored, managed and combined, and mined for maximal efficiency.

### 1.2 Screening for New Chemical Entities (NCEs)

The main objective of screening for NCEs is to identify a therapeutic compound out of a collection of compounds by acquiring high-quality data of therapeutic biological relevance—in the shortest period of time and for the least amount of money possible.

The general historic work flow is to select a therapeutic protein target (not pathway or phenotype), then design an assay around it, execute the screen, validate the hits, and develop lead compounds. This work flow is often abandoned as a result of the completion of the genome project. We have more potential targets than ever available, targets where purified protein might not be available (as is the case with many membrane proteins) or all scenarios where players of the pathway are not known. Consequently, it is unavoidable to augment the arsenal of assays with, for example, pathway-specific screening strategies based on cell-based assays. The development away from the biochemical assays toward more inclusive, cell-based platforms culminated in the emergence of black-box approaches, which rely on, for example, a phenotype observed in cells or even whole organisms as readout. With the increasing complexity of the assay system, it has become equally more difficult to elucidate the mode of action or the target of a hit compound. However, not only is the screening for compounds now automated but also the target discovery, which makes these approaches more attractive than at the beginnings of molecular screening.

### 1.3 Screening as a Research Tool

Screening has made its evolution from a very specialized tool in the drug discovery of the pharmaceutical industry into a research tool that is now found not only in industry but as well at some universities [12]. This transformation has changed molecular screening on the scientific as well as the practical level.

On the scientific level of this evolution, the advent of siRNA, shRNA, and other techniques added to what is now called functional genomics: an interdisciplinary field in which biological questions are addressed by using molecular screening and reagents modulating gene function as a tool. Chemical genomics is a very similar approach; however, it relies on specific small molecules modulating protein function and molecular screening as a research tool. Prior to the development of functional and chemical genomics, it was a *conditio sine qua non* that the protein and its precise function had to be known before a screen for a drug aimed at this target was considered. This has changed to such a degree that screens without a known target are now considered since chemical and functional genomic principles have simplified postscreening target discovery [13, 14]. For example, isolating the target of a small-molecule drug after it has been found in a screen using a genome-wide siRNA collection in the presence of the compound in question is one of many applications for screening as a research tool [15].

On the practical level there has been a transformation of the equipment from expert systems into turn-key systems that are usable to any scientist. Current generations of automation platforms such as liquid handlers have become more and more user/scientist friendly through simplified interfaces and the incorporation of configuration wizards. For example, the early Mini Trak system from Packard was only programmable in scripting language. Since then, a more comfortable graphical user interface (GUI) has been introduced that makes use of these very powerful instruments much easier and enables scientists to be more productive in their research.

## 2 MOLECULAR SCREENING

### 2.1 Low-, Medium-, and High-Throughput Screening and Their Plate Formats

Screening can best be defined for our purposes as the repetitive execution of a given assay in order to find a "hit" compound with a desired biological activity. Consequently, one of the most fundamental properties of any given screen is its throughput, that is, how many compounds or other effectors (siRNA, shRNA, etc.) can be screened per day. There is a general scientific delineation, rather arbitrary in nature but widely accepted, that defines low-throughput screening as up to 1000 compounds per diem, medium-throughput is up to 10,000 compounds per diem, and high-throughput screening as any throughput above 10,000 compounds. The complexity of the assay procedure typically determines the number of wells possible to screen per day. The assay development begins with a full factorial design of the assay. The design will determine the work flow and if any manual labor is necessary. As a rule of thumb, assays that require any form of manual manipulation of the

individual wells will be low-throughput since they have to run in 96-well plates. Medium-throughput assays will run in 384 plates and plate manipulations will be automated, but the assay plates might be moved from workstation to workstation by hand or the assay might run on a fully automated system. Consistent timing of all steps for all plates is the biggest issue in this setting if individual workstations are employed. High-throughput (HT) screening will typically run in 384 or 1536 well plates without manual assistance beyond "feeding the system" or emptying the waste containers.

It is worth having a closer look at the three plate formats and their uses: the 96-well format still has its place for storing larger amounts of compounds samples and in the genomics arena, and the deep-well versions are being used for growing bacteria for the purpose of harvesting plasmid DNA, which can be used for HT transfection. Another application for 96-well plates is organism-based screening—for example, zebrafish-based screening. The main advantage is that its relatively large size can accommodate the space needs for whole-organism-based screens and is accessible for manipulations by hand. Low-volume versions of the 96-well plates exist and have their use in proof of concept experiments. Otherwise the 96-well plates are almost obsolete for screening purposes.

The main workhorse in screening has been the 384-well plate, which still makes the biggest part of all consumables used in screening. Its advantage is its small size—which lends itself to conserving assay reagents compared to 96-well plates—but not so small that it poses any major technical issues in terms of dispensing, plate washing, or reading. It will accommodate a good size number of cells (depending on the size of the cell and incubation time, about 1000–4000 at 50% confluency), and the assay volume can be trimmed from about 15 to 80 μL, depending on the plate subtype used (e.g., regular or low-volume plates). Most plate readers can read 384-well plates at good accuracy and sensitivity. Although evaporation can be an issue, it is generally easily dealt with by (a) minimizing the assay time, (b) using specialty plate lids, or (c) using a good incubator for the plates, which keeps the air at very high humidity or a combination of all the above. Low-volume versions of this plate type are frequently used to save on expensive reagents. Frequently, columns 1 and 24 are not used for experiments as they exhibit evaporation issues.

The 1536-well plates have rapidly become the workhorse of larger pharmaceutical companies [16] since their reduced assay volume (2–8 μL) offers great cost savings, but they suffer from a series of drawbacks. The dispensing and washing process for 1536-well plates requires special equipment. Although there are various plate dispensers for this format commercially available, the quality of those varies considerably from very good to barely acceptable. The lack of reasonably priced plate washers that perform within the very small error ranges permissible for molecular screening makes the general use of the 1536-well format challenging. Another problem with this format is evaporation, which can become rapidly problematic—3- to 4-day assays are usually the longest incubation time reasonable with this plate format. A partial solution is the use of custom lids for the 1536-well plate. Assays that rely on rare cellular events that require the evaluation of a larger number of cells in order to gain usable statistical data will also have issues with the 1536-well format. Furthermore, the signal-to-noise ratio is typically one order of magnitude less than in 384-well plates. Last but not least, it is difficult to preplate reagents such as siRNA's or cDNA containing plasmids in this plate type. These applications are better served in 384-well plates.

### 2.2 Molecular Screening Work Flow

The concept of molecular screening is quite different from performing traditional bench work in the life sciences. Screening is a numbers game and a balancing act. In terms of a numbers game, screening can be compared to playing the lottery, only with buying a lot of the tickets. The chances to win (find a compound with therapeutic properties) increase with the number of tickets bought (compounds screened). In terms of a balancing act, effective screening requires the researcher to balance speed and accuracy. A common challenge for scientists is choosing a process that is neither too slow nor too inaccurate.

Screening is the repetition of an assay (opening the lottery tickets), executed as fast as possible, and winning tickets (hits) are kept and losing tickets (inactive compounds) discarded. Since no assay can achieve perfect results, some winning tickets will be discarded and some losing tickets will be kept—all for the sake of speed or "throughput." However, as long as a sufficient number of the winning tickets are kept, the screen will be a success.

The final consideration in screening is a practical one: In any setting (academic or industrial), the researcher must balance financial constraints (time, resources, and reagents) with the scientific quality of a screening campaign. See Achytuan and Whitten [17] for an introduction to the terminology and concepts needed for the statistical basics needed in development and validation of an assays/screen. First of all, it is important to familiarize yourself with the general layout/work flow of molecular screening. (See Fig. 1.)

At the beginning of the drug discovery process, a strategic decision on a novel target has been made. The

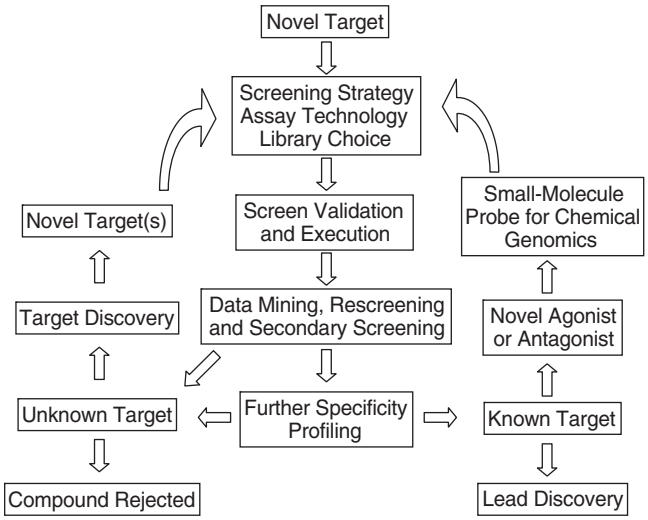


**Figure 1** General work flow of molecular screening.

target could be a protein, a pathway, or even a phenotype. This novel target has been validated to be of high therapeutic and biological relevance and ideally relies on a protein, which falls into one of the classes that are traditionally considered druggable [18]. Experience shows that the majority of current drugs work on G-protein coupled receptors (GPCRs), ion channels, kinases, or proteases. One can argue that the conservative view of drug development, which restricts drug development to those protein classes, is outdated since drug development should encompass novel target classes to respond to novel needs, especially considering the host of novel targets stemming from the human genome project [19]. After selecting the target, the next decision will concern the screening strategy. This could be single-target screening, pathway-specific screening, or black-box approach. Interestingly, all three approaches seem to be equally successful in terms of downstream success [20]. If the screening approach is not target based, the screening platform needs to be selected next. An obvious choice is a mammalian cell system. Other options include screening in bacteria, yeast, or even whole-organism-based screening such as zebrafish.

The screening approach and platform dictate the choice of appropriate assay technology (readouts, reagents, etc.), which is chosen based on its scientific appropriateness, efficiency, and financial considerations. It is absolutely necessary to consider the feasibility of secondary assays/necessary counterscreens and orthogonal readouts at this point, although they are only needed later on. For screens using known pathways or targets, assays for the validation of the hits need to be considered at this point. These assays are vital to the discovery process since they are needed to filter desirable therapeutic compounds from false positives and compounds, which do not have the potential to become a lead due to, for example, their toxicity or lack of specificity.

Together with the assay technology, the assay format has to be determined. Typically, a 384- or 1536-well plate will be used—the 96-well plate is still used when an assay needs to be manipulated by hand or the readout does not miniturize. At this stage small-scale experiments are carried out in order to optimize the assay, test the feasibility of the screen, and confirm the choice of controls and assay format. Next, the validation and screening libraries for this novel target have to be selected.

Validation libraries typically consist of annotated compounds (compounds for which the target and mode of action is known) and other compounds that are known to have activity against a multitude of targets or pathways.

Through the use of compounds that cast a wide net around the biology to be probed, either by directly interacting with the target or by modulating the target through known (or at least conceivable) interactions, the employed assay is subjected to a biological/pharmacological validation. The validation libraries are then employed in order to gain a first indication of the biological/pharmacological behavior of the screen. Validation libraries offer valuable insights into thorny, essential questions, such as whether a given assay has the hoped-for relevance. Potential problems with the small-molecule screening will also become apparent at this stage. At this point the assay is properly configured for the format and the throughput necessary, and plate statistics have validated the assay's accuracy and screening of larger library sets can be executed. A larger-scale screening campaign involves more diverse library sets that cover a broad range of chemical space—this might consist of targeted libraries, a standard set of diverse compounds or a selected part of the compound collection available (smart library).

The data from these large-scale screens is collected and examined for any plate artifacts/trends. The artifacts/trends are removed, the processed data are used to generate a hit list, and the individual hit compounds are then condensed into so-called cherry plates. Secondary assays (i.e., assays that have an orthogonal readout but rely on the same target) and specificity profiling (e.g., on multiple cell lines or homologous proteins) will subsequently indicate the usefulness of the individual hit compounds and aid in flagging false positives. After this point, compounds will either have a known or an unknown target through which they exert their effects. For compounds with unknown targets, target discovery is possible, and potentially novel targets that

are discovered along the way can be fed back into the target discovery cycle. Compounds with a known target and a proper specificity profile can be either advanced into lead discovery or used as novel agonists or antagonists in chemical genomics applications for basic research.

### 2.3 Screening Approaches

***Single-Target Screening*** For genuine single-target screening (on a purified protein or reconstituted pathway) it is necessary to isolate the target. Biochemical assays are the sole method of screening with single targets. Such assays typically contain only the protein in question and substrate or binding partner(s)—and a detection reagent if necessary for the readout. The merits of this approach are simplicity, control over the parameters of the assay itself, ease of use, and speed of the screening process. Since the assay system is simple, the resulting hit compounds can be easily filtered for real and false positives. Additionally, specificity profiling on homologous/related proteins gives an indication of the therapeutic value of a hit compound.

In contrast to target-based screening, assay systems based on cell extracts or cells contain not only the target but also other directly or indirectly related targets that could potentially interact with the primary target in a direct or indirect manner.

A central caveat of target-based screening is that a hit may or may not be relevant in cell-based assays. Compounds that exert an effect on purified proteins may not have the proper membrane permeability to enter the cell or may be overcome by the redundancy of the biological pathway. If a target-based approach is taken, it is essential to establish a cell-based version of the biochemical assay as a secondary screening tool.

***Pathway-Specific Screening*** Pathway-specific screens rely on an assay measuring the functional state of a target pathway. They are by necessity cell based. A typical example for a pathway-specific screen is the use of a reporter gene assay. In this kind of assay, the expression of a reporter gene (e.g., luciferase), which is expressed under the control of the promoter of the target gene in question, is monitored. Therefore, the functional state of the pathway is being gauged using the reporter gene construct. Considering this fact, we can observe that all of the possible targets influencing the functional state of the pathway in question are compromised within the assay system. This works as a multiplier in terms of throughput. Effectively, there it is not only one screen being conducted but many since all the compounds influencing the activity of a reporter will appear as hit. As is self-evident, hits from a screen using this kind of assay can modulate the activity of the reporter gene either directly or indirectly.

Most of the postscreening effort will be directed into looking carefully at questions such as which compounds are direct hits on the target and which compounds obtain their activity by a different means. This type of screen has the potential to deliver more hits than a single target-based screen, which will yield a solution focused only on the target protein(s) and will miss other solutions that might be equally relevant in a more complex pathway screen.

***Black-Box Approach*** The black-box approach is based on a cell-based assay using a phenotype as readout rather than measuring the functional state of a pathway. The advantage is that no a priori knowledge of a target or pathway is necessary. This readout can be very broad—for example, life or death of a cell—in which case it is truly a black-box approach in which no a priori knowledge is assumed or more reduced in scope. For example, it is possible to screen for modulators of nuclear export by tagging the Rev protein, which contains a strong heterologous nuclear export signal (NES), with green fluorescent protein (GFP) in order to make it traceable and track it during the screen using automated microscopy [21]. Such a screen should find compounds that are direct modulators of the nuclear export process but will most likely find as well compounds that have a different mode of action, for example, compounds that are modulators of target gene expression. Generally speaking, in a screen like this, hit compounds may or may not act on the pathway of the target. The obvious downside of the black-box approach is—much like the pathway-specific approach—the lack of knowledge of the mode of action and/or target of the hit compounds.

### 2.4 Cell-Based and Higher-Organism-Based Screening Platforms

***Cell-Based Screening Platforms*** Cell-based assays are frequently but not exclusively based on mammalian cells. For example, nonmammalian cell systems might be necessary for the discovery of antibiotics [22], and yeast can be a great system for heterologous expression of target proteins [23, 24] and thus offer an easily customizable screening platform.

Therefore, cells used in a screen can be prokaryotic or eukaryotic, and each cell-based system will have its own set of necessary precautions. Generally speaking, it is crucial to establish the identity and purity of the cell population before assay development. For bacteria, possible contaminations with bacteriophage come to mind such as M13.[25, 26]. These bacteriophages will not lyse

the cells; instead they will appear to grow at a reduced rate. This can be readily addressed using the top agar method. For yeast things get a little bit more complicated. Contaminations with bacteria—especially *Bacillus cereus*—are a concern. *B. cereus* is very common in the environment and is especially problematic since it is spore forming. It is a close relative of *Bacillus athracis* but has no toxin. The decontamination process is typically long and difficult. Yeast strains contaminated with *B. cereus* are best discarded. Other considerations when screening in yeast-based systems are whether to keep the drug pumps—which are quite active in yeast—intact or knock them out. Both of these approaches have their merit. However, it is necessary to adjust the drug concentration to a higher level for the screen should the drug pumps be kept. On the other hand, the yeast might be less robust if the drug pumps are absent.

For mammalian cells, *Mycoplasma* contamination is one of the most overlooked problems in screens. Tests for *Mycoplasma* are rapid and cheap and there is no excuse for not performing the test before the assay development. Other parameters that are critical are the passage number and the confluency at which the cells are being grown and what confluency they reach during the assay. Generally speaking, the passage number should be small and sufficient backups of the cells should be at hand. The response of the cells in a given assay will most likely diminish with increasing passage number. The same is generally true for confluency of mammalian cells and their response, for example, in reporter gene assays. It is also possible to freeze a large number of cells and use those cells directly in a screen [27].

***Higher-Organism-Based Assays Platforms*** High-throughput screening (HTS) and whole-organism approaches can be combined quite successfully if the organism model fits into the physical dimensions of 96- or 384-well plates. *Caenorhabditis elegans*, *Drosophila*, and zebrafish (embryos) come to mind as suitable organisms for this type of approach, since they are the right size to grow in microformat plates. Zebrafish are notable among vertebrate animals for their production of large numbers of transparent embryos, which makes them more likely candidates for use in the drug discovery process.

The main question as to whether compounds found by zebrafish screening have relevance in mammals remains to be answered. Zebrafish have only been used in drug discovery rather recently, and scientists have yet to develop new clinically viable drugs based on these experiments. However, there is evidence suggesting that some compounds behave very similarly in both humans and zebrafish. For example, Milan et al. reported that 22 out of 23 drugs known to prolong the QT interval (a part of the cardiac cycle) in human prolonged the QT interval in zebrafish as well [28]. Chan et al. reported that a well-characterized antiangiogenic small-molecule PTK787/ZK222584 blocked blood vessel formation in zebrafish [29]. Langheinrich reviews other compounds regulating cholesterol synthesis, coagulation, and vasodilation with analogous effects in both human and zebrafish [30]. It is therefore quite reasonable to conclude that compounds with an activity in conserved biological processes would behave very similarly in human and zebrafish.

Working with zebrafish has several advantages, including an admirable simplicity in terms of procedures performed: the embryos (3–6) are placed in the 96- or 384-well plates and then compounds are added at the usual concentration. Using a microscope or some form of automated imaging system, one can examine the embryos for interesting phenotypes. In addition to being a straightforward process, this sort of whole-organism discovery is also a unique platform for the identification of in vivo modifiers of the wide range of phenotypes present during embryogenesis. By definition, target-based approaches cannot offer this possibility of discovery: The only modifiers identified are those preselected by virtue of offering only one target. This is a compelling advantage for the whole-organism platform, which can enrich our knowledge beyond our ability to anticipate our findings. Whole-organism approaches free us from the limits of our expectations, and they also offer a valuable window into the development process of the embryo. From one embryo, a wealth of morphological data can be gathered by scrutinizing multiple organs, such as skin, hearts, eyes, and brain, at the same time for, for example, angiogenensis [8]. Internal organs can also be examined by issuing tissue-specific GFP transgenic lines. There is a broad context for this data, a physiological framework, offered by the zebrafish as a whole organism. This context excludes compounds with toxic effects at the beginning of screening, which is also quite useful.

## 2.5 Assay Development

Plate-reader-based assays fall into two different categories: gain-of-signal and loss-of-signal assays. Generally speaking, a gain-of-signal screen will deliver more consistent/robust data than a loss-of-signal screen. This holds true for most readout modes with one exception: in the case of fluorescence, gain-of-signal screens will have the added issue of finding fluorescent compounds as hits.

These two are to be differentiated from agonist or antagonist mode screens, which refer to the biological

function of the compounds screened. An agonist is defined as an agent that provokes a reaction, whereas an antagonist is defined as an agent that inhibits a reaction. However, in either case the reaction could be the generation or the loss of signal. Therefore, not all antagonist screens will be loss-of-signal screens and vice versa. Furthermore, screens can be executed in endpoint and kinetic mode. Most screens in HTS are in endpoint mode since throughput is one of the highest priorities, whereas medium- to low-throughput screens—which utilize complicated biological systems—can benefit greatly from kinetic reads. Kinetic reads or data might be collected after the addition of the last component of an enzymatic reaction within several minutes or, for example, every 12 h in a cell-based assay. The advantage of kinetic reads is that the slope rather than the absolute value can be used, and that might come in handy for the elimination of false positives.

Let's examine the main differences between executing an assay in a normal laboratory setting and molecular screening. It will soon become apparent that these differences must be considered carefully, as they will make or break a successful assay for a drug discovery campaign.

First of all, assays are normally performed in singlet, which means that one has only a single chance of catching a bioactive compound. It naturally follows that statistical validation of the prospective assay is a key step. In the ideal case, the validation will be performed on the very machines used in the screening campaign. This saves the later anguish of revalidating an assay on a different set of equipment. Second, all of the plate handling steps during the execution of the assay should be automated, if such automation is at all feasible. In the large majority of cases, each assay step requires the use of a different device, since they all have different capabilities, and for the sake of speed since "double duty" of devices tends to slow the screening process. Each device comes with its own sets of strengths and limitations, which need to be examined in great detail. Each additional step contributes, of course, to the overall likelihood of error. It is therefore quite valuable to streamline the process as much as possible and to limit the number of steps to the absolute minimum. Minimizing the number of steps is also prudent when one considers the logistics of handling large numbers of plates. A large number of plates typically require a fully automated platform and scheduling software in order to keep track of incubation times for each assay step of each plate.

Generally speaking, it is easier to add a liquid to an assay plate, rather than to withdraw the liquid or even to wash the plate. Assays that require only addition steps, steps in which the liquid is added to the assay plate and no separation steps, are considered the easiest to automate and are called "homogenous." We shall now look at the assay development work flow (Fig. 2).

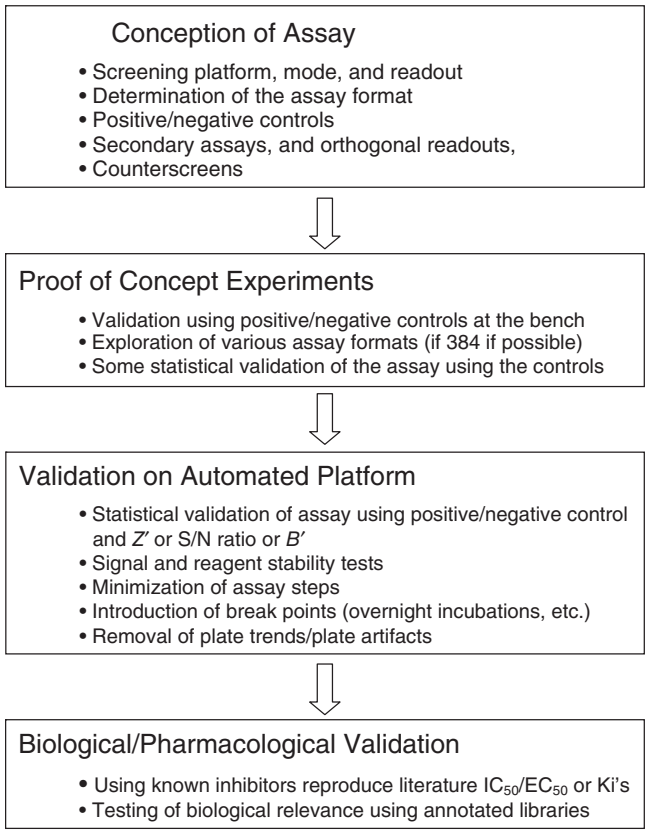


**Figure 2** Work flow for assay development.

***Assay Development Work Flow*** It is very difficult, if not impossible, to give a recipe for successful assay development for screening of therapeutic agents. This is not due to problems with the screening or assay development itself but rather with the disease relevance of the biological target and the model system used for screening. If the target does not have the necessary relevance in the disease, even the best compounds will not make it into drugs since in the latest steps the compounds will fail in the clinical trials—which is extremely expensive. This is outside the control of the screener who cannot—by necessity—be an expert in all disease indications. However, it is possible for the screener to follow a generic work flow for assay development and perform at least a pharmacological validation of the target, which will aid in making sure that the assay has at least the hoped for biological relevance in the model system used.

The steps during the assay development work flow (as summarized in Fig. 2) are: assay conception, proof of concept experiments, validation of the assay on the

automated platform, and pharmacological validation of the assay. Last but not least, we will give a list of the most common error sources. This list is not comprehensive due to the individuality of assays and should be used as a starting point during troubleshooting of an assay rather than a checklist.

*Assay Conception* After a target of therapeutic value has been set, a suitable assay needs to be conceived. The readout for the assay will depend on the selected screening strategy, screening platform, and screening mode. For example, if a gene has been set as a target and the screening strategy has been determined to be a pathway-specific approach, a suitable reporter gene construct could be cloned and cell lines bearing this construct are made. Part of the initial conception of the assay is to determine if the screening approach and the assay are adequate for this target. For example, the reporter gene could be GFP, which would be a fluorescent readout or luciferase, thus generating a luminescent readout. In selecting the readout for the assay, maximal linear range and sensitivity should be emphasized—however, cost can be a major concern as well. Then assay format will be determined for the initial tests and the campaign that is to follow. Other secondary assays should be considered and they should have an orthogonal readout (e.g., using nitroreductase as reporter gene). This is necessary since false positives, which we will discuss later, can be the direct result of the selected readout, which makes it important to have the same assay with a second, independent readout at hand for the rescreening of the hits. Last but not least, counterscreens have to be considered. Typically, these counterscreens eliminate, for example, cytotoxic compounds, which will influence the reporter gene activity by simply killing the cells. Other counterscreens would contain, for example, variations of the promoter in order to get a sense of the specificity profile of the potential hit compounds. During assay conception, reference substances such as known agonists or antagonists as well as siRNAs or shRNAs (or cDNAs) come in handy, as they allow for building a model system consisting of a positive and a negative control even if no reference compounds are available. When conceiving an assay, it is necessary to select from—and frequently experiment with—a multitude of detection platforms and readouts. The two most commonly used assay detection platforms are multimode plate readers and devices for high-content screening. These two platform types have their own sets of advantages and disadvantages, which need to be considered carefully during assay development.

*Plate-Reader-Based Assays* Plate-reader-based assays, generally speaking, rely on one of the five standard readouts, which are fluorescence [31], fluorescence polarization [32], time-resolved fluorescence [33], luminescence [34], and absorbance [35]. The main advantage of plate-reader-based assays is speed, since an entire plate can be read in the matter of minutes, and flexibility, since most plate readers offer many, if not all, of these readouts. Handling the data from these readouts is normally very easy as well. On the minus side, rare cellular events are typically not addressable using plate-reader-based assays since the signal-to-noise ratio is not high enough. Another downside is the generation of false positives, which depends on the reading mode used. We will now discuss the read modes in detail.

Fluorescence intensity is historically one of the earlier readouts for molecular screening. Fluorescent molecules used for screening purposes can be divided into synthetic dyes such as fluoresceine, fluorescent proteins such as GFP [36], and fluorescent inorganic molecules such as lanthanide chelates [33] and quantum dots. All of these classes have their pluses and minuses. Synthetic dyes have their main applications as latent fluorphores, which can be converted by enzymes such as proteases, hydrolases, horseradish peroxidase, and alkaline phosphatase to their fluorescent form. Other applications are the labeling of antibodies for immunostaining. Fluorescent proteins such as cyan fluorescent protein (CFP), GFP, and red fluorescent protein (RFP) are mostly used as reporter genes, with the exception of very bright fluorescent proteins such as, for example, allophycocyanine (APC), which has been used for labeling of secondary antibodies and other detection reagents. The phenomenon of fluorescence energy transfer (FRET) in which the emission spectrum of a donor and excitation wavelength of an acceptor molecule are in close proximity has as well its uses. The acceptor can be another fluorescent dye or just a quencher, which does not emit light but converts the light energy into kinetic energy. In the case of lanthanides the emission response to excitation is much longer lived than in standard dyes or fluorescent proteins. It is in the order of microseconds instead of nanoseconds. Therefore, they have uses in time-resolved fluorescence (TRF): After excitation of the sample, the plate reader waits a given period of time until the emission response of the standard fluorophores is exhausted and then starts measuring the emission of the lanthanides. This reduces the background greatly, leading to a much cleaner signal. Lanthanides can be combined as donors with other acceptor fluorophores to TR-FRET. The lanthanide chelate is chemically easy to modify and, therefore, the system is very useful for measuring the levels of small-molecule analytes such as cyclic guanosine monophosphate (cGMP) [37]. Strongly chelating agents such as ethylenediaminetetraacetic

acid (EDTA) should be especially avoided for lanthanide-based screens as they can possibly strip the lanthanide out of its organic chelator and destroy the detection system. Quantum dots are clusters of inorganic elements that typically are much brighter than traditional dyes or fluorescent proteins, and they exhibit as well an extended emission time allowing for time-resolved applications [38]. Last but not least, excitation with polarized light can be used in order to determine if a fluorophore has been bound by a larger molecule [39]. Small molecules (say, around 1000 Da) tend to rotate very quickly in solution. Therefore, if they are excited by polarized light, they will relax quickly, leading to a reduced anisothrophy, where the emitted light loses its *z* polarization. Larger molecules (e.g., 20,000 Da) do rotate slower; therefore, their emitted light retains the anisotrophy. It follows that the difference in anisothrophy between samples is only big if the difference between molecular weight in bound and unbound states is big. Since kinetic movement of molecules is involved in the measurement principle, it is self-evident that this assay will be strongly dependent on the viscosity of the sample and the temperature at which the assay is run. A difference of 5°C in ambient temperature can be expected to have a considerable effect on the assay results.

As a general precaution it should be mentioned that some fluorophores are strongly dependent on the pH (e.g., fluoresceine); therefore, it is key to check compatibility for the buffer pH and optimal fluorescence performance and, furthermore, some solvents like dimethyl sulfoxide (DMSO) are excellent quenchers of fluorescence themselves.

To complicate matters further, fluorescent compounds (see false positives) can be a problem [31]. This is the case for blue fluorescence, green fluorescence, but not as much for red fluorescence. In my personal experience, in the case of the first two wavelengths about 0.5–1% of all compounds might come up as false positives, whereas in the case of red fluorescence 0.1–0.2% can be expected. Workarounds include the use of kinetic reads instead of endpoint reads. Furthermore, other constituents of assay samples such as phenol red in the cell culture media, some batches of fetal bovine serum (FBS), or bacterial media can be fluorescent and interfere with this readout. In terms of plates, it should be noted that some plastic plates have a much higher background compared with glass bottom plates.

In my own experience, luminescence can now be considered the workhorse of molecular screening [34]. It has basically no background and the sensitivity and linear range are excellent, and in my experience it is more sensitive then fluorescence.

Luminescense has applications as readout for cell viability and reporter gene assays. In the case of cell viability a luminescence mix is added that contains a lysis buffer, firefly luciferase, and luminol but no adenosine 5′-triphosphate (ATP), which is necessary to fuel the reaction. The ATP is supplied from the live cells in the sample and the luminescence signal is proportional to the number of cells—down to about 50 cells per well in an average plate reader. In the case of a reporter gene assay, the luciferase reagent used contains ATP, luciferol, and a lysis reagent but no luciferase, which is supplied by the cells, thus measuring the expression level of the reporter gene.

Assays based on horseradish peroxidase (HRP)—for example, using a secondary antibody labeled with HRP—can be frequently converted to luminescence by using a luminescent substrate. The enhanced sensitivity comes in handy, especially in whole-cell enzyme-linked immunosorbent assay (ELISA) where the expression of a cell surface protein is to be monitored.

Colorimetric or optical-density-based assays are undoubtedly some of the cheapest readouts and therefore very attractive, especially in an academic setting. In an industrial setting, however, colorimetric assays are typically avoided since they can be used for HTS only if the signal difference between negative and positive controls is large enough to be significant in a high-throughput environment—which is rare. The main problem with this readout is that the linear range is very narrow, which is largely a result of the very small path length in a standard 384- or 1536-well plate. Colored compounds can on some occasions be a problem and can be dealt with by taking a baseline read before substrate addition or by using kinetic reads instead of endpoint measurements.

However, measuring optical density (OD) has its definite applications in multiplexing—especially in bacteria-based screens where an $OD_{595}$ reading can be used in order to gain additional information since no reagent has to be added in order to perform this readout.

*High-Content Screening* High-content screening (HCS) is a comparatively novel platform for drug discovery. It has already gained great importance in a short period of time [40, 41]. While plate readers read an entire well (or even plate) at once, in HCS an automated microscope takes one or more pictures for each wavelength of the cells in a given well on a plate. The pictures from one location in a well are then overlaid and automatically evaluated using image analysis software—multiple pictures per well add to the accuracy of the results. A very simple assay would include, for example, a 4′,6-diamidino-2-phenylindole (DAPI) stain for the nucleus—thus enabling the counting of cells—and the

protein of interest tagged with, for example, GPF or RFP in order to follow its fate and score cells as positive or negative. Therefore, high-content-based assays escape the gain of signal/loss of signal classification, which makes them uniquely suited for difficult applications such as functional genomic screens.

It is as well possible to use phase contrast and follow the morphology of the cells. This is technically more challenging since the signal-to-noise ratio is much lower than in fluorescent-based assays and requires more postprocessing, thus adding an additional layer of complexity. The fluorescent-based high-content assays will allow for phenotype-based screens such as translocation assays [21], but still some knowledge of the tagged target protein is required. The phase-contrast-based assay will allow for a true black-box screening. Other applications such as GPCR-based screening are possible as well with the transfluor technology [42, 43].

Since high-content screening is based on the evaluation of pictures of cells, it allows for reading more than one parameter at a time—thus the name high-content screening. HCS has the advantage that toxicity issues of hits will be apparent early on—provided that the data from the high-content screening are analyzed for it [44]. HCS is especially suited for difficult chemical screens with low signal-to-noise ratios. Other applications for HCS are functional genomic screens such as siRNA-, shRNA-, and cDNA-based screens. One of the reasons why HCS is extremely well suited for functional genomics-based screens is that it allows for the resolution of data on a single-cell level. The advantage lies in having the results from each cell as individual data point that can be averaged to a per-well result. This process reduces deviations to a level at which such screens, for which the readout is difficult to monitor or rely on, to a rare cellular event that can be comfortably run. This benefits, of course, the standard screening for NCEs. The other advantage is that it allows for unexpected phenotypes to be recorded (pictures can be analyzed repeatedly), and it does not rely on an assumed knowledge of the mode of action of a target. The disadvantages of HCS are the relatively low throughput—reading can take up to an hour per plate, depending on the system and how many pictures per well are acquired—and the amount of image data that needs to be managed—which necessitates investment into larger computer infrastructure. Furthermore, specialized software is necessary to evaluate the images and configuring the software to score hits correctly may need the help of a biostatistician.

*Proof-of-Concept Experiments* The next step in assay development is to execute small-scale experiments at the bench. The goal is to establish a basic assay using the positive and negative controls, work out basic assay parameters such as reagents, concentrations, and incubation times and the like. This is also the time to explore various assay formats. I typically go directly to 384-well plates as they are still—with a little bit of exercise—amenable to manipulation by hand, but they can be used already for high-throughput screens and it is not such a big step to 1536-well plates after this. It is a good idea to use the detection system (e.g., plate reader and reagent), which would be normally used during the screen, as this saves time, which otherwise is being spent reoptimizing an assay. Here the positive and negative controls are being used in order to explore if a given assay is at all viable. Typically, a $Z'$ value (the formula for $Z'$ is given below) of >0.5 should be obtainable.

*Validation on the Automated Platform* After a low-throughput (manual) assay has been developed and the proof-of-concept experiments have been run, the assay has to be adapted to automation. Since every assay is different, it is difficult to give a general blueprint for assay optimization. However, it is easy to explain what the general direction of optimization is: maximization of the signal-to-noise ratio and minimization of the variance of the signal and noise values. The two main statistical parameters to gauge progress are the standard deviation σ and $Z'$ [45] (see below), however, some people still like to use the signal-to-noise ratio.

$$Z' = 1 - \frac{3(\sigma_- + \sigma_+)}{\left|\overline{\mathrm{CTRL}}_+ - \overline{\mathrm{CTRL}}_-\right|}$$

*In this formula, σ is the standard deviation of either the positive or negative control, CTRL refers to the value for the positive or negative control.*

The basic necessity for an automated assay is a $\sigma^- < 20\%$ and $\sigma^+ < 20\%$ and $Z' > 0.5$. There are as well other statistical parameters that can be used such as $B'$ [46], but I personally prefer $Z'$ during assay development. Especially for loss of signal screens, the σ for the half maximal effective concentration ($EC_{50}$) of a reference compound should be established and needs to be below 20%. An excellent overview of the statistics and procedures for initial assay validation is given in the *Assay Guidance Manual* by Eli Lilly and the National Institutes of Health (NIH) Chemical Genomics Center, version 5.0, 2008. It can be found at http://www.ncgc.nih.gov/guidance/manual_toc.html (last accessed on 9/11/08).

The necessary assay optimization will usually include an optimization of each of the reagents used in a given assay and all steps performed during the execution of the assay. This seems at first somewhat daunting but

becomes very manageable if one parameter is varied at a time.

*Plate Uniformity* Plate uniformity needs to be addressed as one of the key issues for the development of an automated assay. Lack of plate uniformity will make any potentially successful screen impossible.

There are two major kinds of phenomena causing plate uniformity issues: edge effects and plate trends. Edge effect refers to a phenomenon in which the rim of the plate—especially its corners—shows a different signal than the middle of the plate. There are various causes for this and not all are well understood: evaporation on the edges of plates seems to be a major contributor to edge effects since it alters the liquid level at the rim of the plate but not the middle. Consequently, a more concentrated solution found now on the edges of plates can lead to, for example, faster enzymatic reaction. This in turn will increase the signal of a given readout. Or concentrated growth factors in cell culture media will lead to an elevated growth, leading to an increased signal at the edges of the plates. A less understood phenomenon occurs during the plating of cells for a cell-based assay: If a plate is filled with cells in cell culture media and transferred to the incubator immediately, the cells will collect in the corners of the wells at the edge of the plate but will be normally distributed in the center wells of the plate [47]. A workaround is to incubate the plates with the cells for 1–2 h at room temperature before transferring them to the incubator. It is thought that the edge plates warm up after transfer to an incubator at a higher rate than the middle, therefore leading to a different distribution of cells within the wells. This problem will be encountered as well if a cold reagent is being distributed to a plate. The edges warm up faster than the middle leading to a temperature gradient within the plate, which can translate into different enzymatic reaction rates, thus an edge effect. Edge effects are best observed using heat map plots of a given plate.

Plate trends fall into two categories: patterns on a plate or trends in which one side of a plate exhibits a higher signal than the other side. A plate pattern is best detected using a heat map of a plate, whereas plate trends are easily detected using scatter plots in which the data are sorted by row and then column or by column and then row. The reasons for plate patterns can be normally traced back to the impact of individual devices used during the assay run, whereas plate trends are more difficult to trace but frequently are the result of differences during plate incubation. The troubleshooting typically involves taking readings of the assay after each assay step and correlating the occurring signal anomaly with the assay step at which point changes to the protocol can be made. For obvious reasons it is key to run at least one half of plates per condition as the assessment of plate uniformity is one of the major necessities for making a good automated assay and at least one half of plate is necessary in order to detect edge effects.

*Signal Stability and Reproducibility* The assessment of signal stability and reproducibility is vital. This needs to be performed using intraday and interday testing. Sentinel plates, which include only the positive and negative controls and DMSO but no compounds at the beginning, end, and middle of the screen, have great use in this as well.

Automation has the big advantage that a given assay can be run 24/7. This puts reagent stability into the center stage of prescreening validation. All reagents (including cell lines) have to be (a) stable over the entire projected run time of the assay and (b) DMSO compatible. Furthermore, it is imperative to keep a record of all the lot numbers of all consumables—whether this is an in-house produced resource or a commercially available product like plates or detection reagents. The reagent validation needs to be performed for each reagent (and each lot used) separately in order to gain an understanding of the weak points of an assay. Particular care needs to be exerted on validation of reagent storage, which includes validation of freeze–thaw cycles and the validation for mixing different lots of a reagent.

If a given reagent is not stable over the projected run time, it is normally possible to break the run time into smaller parts and use each time a new aliquot of the reagent. This can counter most of the issues of performance degradation at the cost of more manpower. It is interesting to note that reagents tend to degrade in the tubing of liquid dispensers. Therefore, a good way to counter reagent degradation is to back-flush the reagent automatically after each dispense cycle into its chilled reservoir and autoprime it before addition to the next plates.

If the compatibility testing with DMSO—which is normally carried out for up to 4% DMSO final concentration—fails for a given reagent even at 1% DMSO in the final assay buffer composition, the amount of DMSO can be lowered to, for example, 0.1% by using a pin tool set with a smaller volume and higher concentrated compound stocks. However, this will raise the compound consumption since it necessitates having a second deck (i.e., compound collection) and will use more compound during cherry picking since higher concentrated stocks have to be used. Last but not least, it cannot be overemphasized how much the passage number of cells can influence the signal-to-noise ratio.

*Minimization of Steps and Introduction of Breakpoints* Every step in an assay will contribute to its variability. Therefore, it is imperative to simplify the assay as much as possible. For example, two reagents might be added together, thus saving one assay step. It is also necessary to know precisely where breakpoints of a given assay are, that is, points at which an assay can be stopped for any extended period of time without loss in performance.

A good example for simplification by rearranging the order of reagent addition and the introduction of a breakpoint is the use of reverse transfection strategy for screening of siRNAs. For normal transfection experiments, mammalian cells are first grown in a plate, and then the media is removed and replaced with a mixture of siRNA and transfection reagent. Finally, typically after 24–72 h, the plates are read. This work flow necessitates emptying a plate and immediately adding the siRNA/transfection reagent mix using a pipettor. This process is very cumbersome and agonizingly slow if larger sets of siRNAs are to be screened. In reverse transfection, the siRNA solutions are aliquoted into dry plates. This enables one to produce multiple copies of any given siRNA plate set. The aliquoted plates can be frozen and thawed directly before use—thus making it possible to introduce a breakpoint. After thawing of the plates containing siRNA some media is added to resuspend the siRNA, then the transfection reagent is added and finally the cells are dispensed on top: All additions can now be done with a noncontact bulk dispenser, which leads to a vastly improved throughput. The rest of the assay works as in forward transfection strategies. The simple change in the order of reagent addition and the introduction of a breakpoint leads to a 5- to 10-fold increase in throughput.

*Biological/Pharmacological Validation* After the automated assay is established, it is imperative to check if the current assay has the biological relevance hoped for when the original decision was made to pursue this target. The assay development included already known inhibitors—as far as they are available—or possibly siRNAs, cDNAs, or shRNAs. However, this does not protect from surprises, especially if the target is new. A straightforward way to functionally check the biological relevance of a given assay is to run small annotated libraries consisting of a variety of agonists, antagonists of relevant pathways, enzyme inhibitors, ion channels ligands, protease inhibitors, kinase inhibitors, and the like.

The results from this validation screen will indicate whether the assay fits into the biological context expected or if unusual or even abnormal hits are being found—thus making it possible to gain a sense of confidence in the screening results before the actual campaign starts. Furthermore, for novel targets the results from this explorative screen will aid in gaining a better understanding of its basic biology. After the pharmacological validation is complete, the screen can be run against libraries.

*Assay Troubleshooting* When an assay is not performing as envisioned, troubleshooting can be somewhat tricky since many different causes can potentially contribute to poor assay performance. Some causes are plain errors; others just add to the variability of the assay results, thus diminishing the assay performance. Some errors are common, some are specific to biochemical assays, and some are specific in cell-based assays. Common errors are liquid handling errors, reagent adhesion (e.g., peptide adhesion in the tubing of dispensers), insufficient mixing, problems with the incubation conditions, and human error.

Biochemical assays have in addition to the common errors as well solvent effects, reagent stability/solubility, pH, temperature, and the order of reagent additions as potential problems.

Cell-based assays can have all of the above as error sources and as well passage of the number of the cells, cell cycle, confluency, serum/media, cell culture plastics, and infection (e.g., with *Mycoplasma*) as potential problems.

For cell-based assays it should be considered that the biological relevance of an assay system can be a function of time. For example, inflammation can have a direct response that can be measured after 4–6 h or a long-term response that could be measured after 48 h. They are mediated through different pathways but react to the same stimulus. Another related example of this is screening with agonists in an immune background. When looking at short period responses, experiments capture the immediate pathway inductions. At later time points such as 24 h, responses can be convoluted by feedback loops of the stimulated pathways via interferon secretion of the cells in the well.

### 2.6 Library Choices

Molecular screening can be viewed as the interplay between a given assay and the libraries used in screening. Both need to be of high quality in order to deliver compounds with therapeutic value. It would be desirable to match a given assay with the right libraries in order to maximize the chances of successful screening. In practice, this can be very daunting as it is often totally unclear what the best library would be, especially if a black-box approach is used for screening and the molecular target is open. However, having good-quality librar-

ies and sampling as much chemical diversity as possible while using a reasonable count of compounds screened is a reasonable recipe for success [48].

There are various compound types that should be avoided when selecting libraries. A noncomprehensive list can be found in Figure 3. Reviews by Shoichet are a very helpful introduction to the field [49].

There are many different types of libraries to choose from and we will now look at the individual libraries in detail.

***Shortcomings of Compound Libraries from the 1990s and What New High-Quality Compound Libraries Are Doing Better*** Generally speaking, the combination of high-throughput screens and combinatorial chemistry so far has not filled the pipelines in the quantities hoped for—quantities that would be proportional to the compounds screened. This has lead to some disenchantment with molecular screening on the whole [50, 51].

It can be speculated that one of the reasons was—and might still be—poor library quality, and very modest coverage of the chemical space, which had its sources in the limitations of the organic chemistry employed while making these compounds and purity issues of the individual compounds. There are now more high-quality libraries available. Compound purity, which always had been an issue with the first compound libraries that came to market, is being addressed by efforts to use liquid chromatography/mass spectrometry (LC/MS) to purify all compounds going into the screening or at least performing LC/MS on the hits. On the issue of intrinsic quality of the compounds, such as diversity, there has been some progress with diversity-oriented synthesis, which is supposed to mimic natural scaffolds but still be synthesis friendly enough to be used in a combinatorial style. Furthermore, smart libraries are more and more employed. These smart libraries generally try to probe a wide array of scaffolds, searching a wide chemical space, with a minimal number of compounds. Targeted libraries go into the same direction, which now most vendors of compounds offer. These libraries are based on predictive computation and experience in medicinal chemistry, and are skewed toward a given target class, for example, kinases. In reality it cannot be ascertained that these compounds will indeed hit the target class intended, but, generally speaking, these compounds are of much higher quality than the classical, diverse sets that were available before. The newest generation of libraries takes as well toxicological problems into consideration that have traditionally been only considered in phase I and II clinical trials, and anticipates and eliminates these problems from the get-go, thus leading to higher success rates.

***Smart Libraries*** In smart library screening a large chemical space is covered by screening fewer compounds per space unit than normal. The results from screening this diversity sampling set are then used for selecting "smart libraries" around the hit, followed by rescreening of these libraries. Hits from the smart library can then be used in order to generate a small focused subset, that is, smart library, to further refine the location of chemical space screened and expand or "drill down" into the next set of hits. The testing/expansion cycle can be repeated until the best hits have been found.

It appears that large-scale screens are being reduced in their number in favor of smart screens [52]. Some companies have even abandoned large-scale screens altogether, largely because the benefit is not believed to justify the cost and labor involved for these companies. Smart screening strategies have some conceptual problems, however.

One of the main conceptual problems is that smart screening would enrich for false positives as well as real positives. Hits from a smart screen contain real and false positives since every readout type carries its own set of false positives as we have seen above. In order to avoid expanding into the chemical space of the false positives, not only rescreening in the already used assay is necessary before expansion into smart libraries but assays with orthogonal readouts as well are needed to filter the desirable compounds from the undesirables. This means that a screening campaign does not consist of the serial use of one assay for the primary screening and then a few assays for secondary screening but cycling between a single assay and a multitude of assays, which are necessary for filtering, thus slowing down the screening process. Another reason why it is imperative to "filter" for false positives is that chemotypes with undesirable properties tend to have broader structure–activity relationships (SARs)—quite a few of false positives depend on only one structural feature of the molecule—for example, reactive groups such as a Michael acceptor—while compounds specific for the target and thus desirable tend to have narrow SAR. Therefore, without filtering, smart screening could possibly enrich faster for compounds with undesirable properties and a broad SAR than enriching for compounds that have a desirable but narrow SAR.

Another challenge is to determine the number of compounds to screen initially to ensure that the chemical space is adequately represented in the original diversity set. This is very problematic since diversity sets are always selected in reference to a starting point (a single compound). It is unclear at this point how to objectively pick the starting points for the diversity-based sublibraries, which make up the initial diversity set.

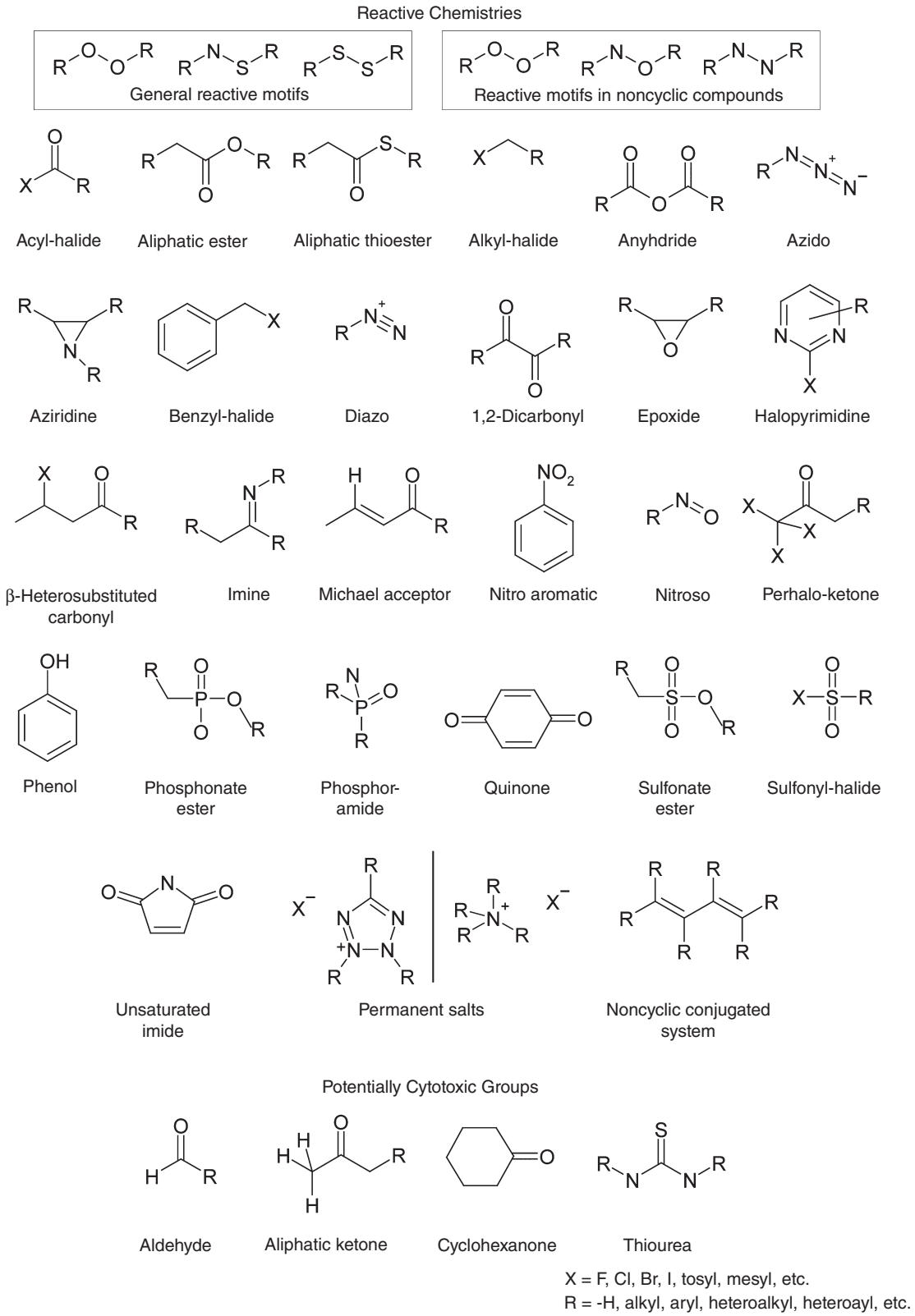


**Figure 3** List of chemical moieties with undesirable properties.

In an internal analysis at the Genomics Institute of the Novartis Research Foundation [53], it has been noted that "sampling diversity" is at least for some targets not a suitable approach and that the best compounds might be missed. In this study a retroactive analysis was performed on a larger data set. The initial actual project delivered the usual hit rate and one outstanding hit. For smart-screen purposes, the screened collection set was broken down into small diversity sets followed by iterations with expansions into the initial diversity set. On average, the strategy worked: With only a small subset of the compound deck "screened," it was possible to obtain double the amount of hits found; the screen delivered twice the number of hits found by traditional means. However, the one outstanding hit was missed frequently. Of course, it might be possible to skew the initial diversity set to achieve higher hit rates [54, 55].

Only time will tell whether full deck or smart screening will work best—and this might depend on the target and how amenable the target is to, for example, in silico screening and rational approaches. It can be expected for the foreseeable future that screening large libraries will remain necessary.

***Annotated Libraries*** Annotated libraries play a crucial role in pharmacological assay validation since their use enable the screener to gauge the pharmacological/biological relevance of a given assay system. Various libraries are available. Biomol sells, for example, ion chanel ligand, orphan receptor, enzyme inhibitor, and bioactive lipid libraries. The Library of Pharmacologically Active Compounds (LOPAC) library from Sigma Aldrich is well worth mentioning since it is used very widely [56]. Particular care should be taken that a validation library contains drugs that are specific against certain targets and have a very narrow specificity profile (e.g., certain kinase inhibitors such as the PKA inhibitor HA-89) as well as compounds with broad specificity against a whole target class (e.g., thiol protease inhibitors such as E64). Furthermore, it should be noted that the concentration of the compounds used is very crucial as specificity is a function of concentration: for example, kinase inhibitors can become very promiscuous if used at concentrations much higher than their half maximal inhibitory concentration ($IC_{50}$), which then in turn could lead to wrong conclusions during assay validation. Annotated libraries will, therefore, use each compound in a concentration range, which makes physiological sense.

***FDA Drug Collections*** Every molecular screener's dream is to find a compound directly from a screen that does not need any medicinal chemistry, no absorption, distribution, metabolism, and excretion (ADME) and toxicological (Tox)–Tox issues, no clinical trials, no Food and Drug Admistration (FDA) approval process, and can be used in humans.

At first this seems unrealistic, but, recently, more and more compounds that are already approved find novel uses [57]. There are even some companies that specialize in screening-approved drugs for novel uses. This is very lucrative since the novel use of an already existing drug can be captured in a "use patent," which is not quite as strong as a "composition of matter" patent but does grant the company sufficient time to make profit from their discovery. In terms of libraries, the NIH recently made a collection of compounds that have a history of clinical trial available for a nominal fee (www.nihclinicalcollection.com, last accessed 9/13/08). Other library manufacturers are Prestwick and Microsource who offer libraries containing already approved drugs. In the case of Prestwick, the drugs are off patent, which is especially useful when rapid commercialization is the goal. Another approach is to start with a known drug and develop the side activity into a new scaffold [58]. In this context it is interesting to note that sometimes completely unsuspected activities can be found in already approved drugs. For example, cardioglycosides have been shown to influence splicing at a concentration in the nanomolar range, which is their normal working concentration for cardiotonic uses [59]. Viagra would be another example. Its original use was for hypertension, however, the "side effects" became more lucrative than the originally intended use.

***Natural Products Libraries*** The main attractions of natural products lie in their diversity and typically good biocompatibility. They frequently violate the rule of 5 [60] but still have high oral bioavailability [61]. This might make sense since humans coevolved with these organisms and are therefore adapted to exposure to these compounds. These properties make them ideal candidates for drug leads [62], especially anticancer drugs [63]. However, currently, the most screened compound libraries are of synthetic origin, while about 40% of the approved drugs find their origin in natural products [64]. This disparity was originally due to the greater availability and manufacturability of compounds generated by combinatorial chemistry and the shortened cycle times between testing and synthesizing of analogs in modern molecular screening.

Natural products are more and more reincorporated into the drug discovery process as the bottlenecks for natural product discovery are solved, that is, the generation of libraries and derivatives [50, 62, 65]. Since a big part of the currently approved drugs have their origin in the natural product arena, it should be expected that natural products will add to the rapid population of discovery pipelines with candidate molecules. This is

especially interesting considering that the number of natural products screened was a lot smaller than the number of synthetic compounds screened [66].

Natural product libraries come in various flavors. The compounds can be a crude extract, partially purified—frequently pools of five compounds are being used—or fully purified single compounds. It is imperative to retest or deconvolute the hits from a screen on a natural product library to make sure that the active ingredient is properly identified or to determine the active ingredient. High-throughput LC-MS [67] and high-throughout nuclear magnetic resonance (NMR) [68] are the tools of choice for this process. The main disadvantages of natural products lie in the cost for the compounds and the possibility of discovering already known substances—a problem that can be countered by identifying the active compound very early on—directly after a plate location with activity has been found [22].

Natural compound libraries can be generally classified by the source organism: plant, bacteria, marine organism, fungus, and so on. It has been found that the same plant—just grown in a different climate—can have different compositions; therefore, care is advised. In light of the pharmacological usefulness of the biodiversity on this planet, we would do well in preserving it. As rapidly as species are vanishing, it seems a wise idea to purchase compound libraries for which the supplier either has the donor microorganism or enough material banked for resupply.

***Targeted Compound Sets*** Most compound suppliers offer compound sets that are targeted toward a certain class of proteins or enzymes in an attempt to match an assay or target with the most promising library. This works very well, for example, for proteases in which case peptidomimetics can be used [69, 70]. Another example would be ATP analog libraries for the inhibition of kinases. In this case even some rational design can be performed, with excellent results [71]. It becomes more problematic with targets that are members of a protein class that can bind to a huge array of diverse ligands such as GPCRs where no overall sequence homology is found [72] or ion channels. In both cases the target class is considered highly "druggable," but the ligands bound have considerable variation. For example, in the case of GPCRs the ligands bound can range from lipid derivatives [73] to opoids to muscarinic acetyl-choline [74].

Using a targeted library during screening does not guarantee that a targeted library will hit the class of proteins it is designed to hit. This is even apparent when mining targeted collections for their annotation of some manufacturers: It will be found that the classes have some overlap in their compounds, therefore the "targeting" is hypothetical and the used models do not allow to design a compound ab initio as only targeting a certain class of proteins.

Using these libraries, however, does guarantee that the compound set has been designed in consideration of medicinal chemistry principles, which makes for a high-class compound library. If cost is no objective, one is better off buying only (compounds) made from high-quality medicinal chemistry rather than diverse compound collections that might consist in part of compounds that have not been made with medicinal applications in mind at all. It is nowadays possible to find targeted libraries against most target classes. A noncomprehensive overview of suppliers and libraries is given in Table 1.

***Diverse Compound Sets*** Diverse compound sets as we know them originated from the collection efforts of various compound suppliers after the breakup of the Soviet Union. These collections had an enormous size of over 600,000, even millions of compounds. From these collections, subsets were selected that contained compounds that seemed more or less druglike and its members were more or less well characterized by NMR or LC-MS. These libraries were a boon to the discovery progress of many drug companies as they contained both novel chemistry and very interesting structures. Unfortunately, for quite a few of these compounds resupply was not guaranteed and resynthesis was often not possible due to either missing protocols or the scientists who originally synthesized these compounds could not be found. Other problems included impurities, compounds in different salt forms depending on the batch received, and so on. This ultimately led to much frustration when trying to find out why an active compound became inactive when reordered. Through the feedback of customers these compound collections have been improved and can now be considered fairly reliable sources of screening quality compounds, which can be used to cover a great deal of chemical space—on a budget.

***Library Integrity*** Since the product of molecular screening is basically just a plate number and well coordinated, the continued success depends completely on the quality of the compound in the well that was screened. Therefore, library integrity is crucial to the success of any molecular screening. Almost every person involved in drug discovery can tell anecdotes about the best hits from a screening campaign that turned out to be a very bioactive contaminant in the library screened that could never be identified. A reduction of these contaminants is therefore highly desirable [75–77].

The use of compounds follow a general work flow: First, compounds are selected from compound libraries

**TABLE 1 Overview of Suppliers and Compound Libraries**

| Name | Libraries/Core Offerings | Ordering | MedChem Support? | AdmeTox Services? |
|---|---|---|---|---|
| Ambinter (www.ambinter.com) | Large-scale druglike compound provider | Online library search | No | Yes |
| Acros Organics (part of Thermo Fisher) (www.acros.com) | General druglike compounds | Online compound reordering<br>Online library search with compound availability | Yes | No |
| Albany Molecular Research (http://www.amriglobal.com/) | Large-scale druglike compound provider<br>Natural product library | | Yes | No |
| Alfa Aesar (www.alfa.com) | General druglike compounds<br>Metal-binding compounds | Online library search with compound availability | Yes | No |
| AnalytiCon Discovery GmbH (www.ac-discovery.com) | Natural product library<br>Semisynthetic natural product library | | No | No |
| Asinex (www.asinex.com) | Large-scale druglike compound provider<br>Druglike fragment library | Online compound reordering<br>Online library search with compound availability | Yes | Yes |
| Apollo Scientific (www.apolloscientific.co.uk) | General druglike compounds | Online compound reordering<br>Online library search with compound availability | Yes | No |
| AsisChem (www.asischem.com) | Large-scale druglike compound provider<br>Building block library | | Yes | Yes |
| Aurora Fine Chemicals (www.aurorafinechemicals.com) | Large-scale druglike compound provider<br>Focused/targeted libraries:<br>Analgesic agents<br>Antibacterial agents<br>Anticancer compounds<br>GPCR modulators<br>Ion channel modulators<br>Kinase library<br>Nuclear docking library<br>Nuclear similarity library<br>Cellular pathways modulators<br>Signaling pathway modulators | Online library search | Yes | Yes |
| BioMol (www.biomol.com) | Focused/targeted libraries:<br>Bioactive lipids<br>Endocannabinoids<br>Neurotransmitters<br>Nuclear receptor ligands<br>Orphan ligands<br>Institute of Chemistry and Cell Biology (ICCB) known bioactive compounds<br>Kinase inhibitors<br>Ion channel ligands<br>Protease inhibitor library natural product library | Online compound reordering | Yes | No |
| Chemical Block (www.chemblock.com) | Large-scale druglike compound provider | Online library search | Yes | |
| Exclusive Chemistry (www.exchemistry.com) | General druglike compound library<br>Building block library<br>Combinatorial libraries | Online library search | Yes | No |

(*Continued*)

**TABLE 1** ***Continued***

| Name | Libraries/Core Offerings | Ordering | MedChem Support? | AdmeTox Services? |
|---|---|---|---|---|
| Cerep (www.cerep.fr) | Customized assay and in vitro target solutions | | No | Yes |
| Chembridge (www.chembridge.com) | Large-scale druglike compound provider<br>Building block library<br>Focused/targeted libraries:<br>Kinase inhibitors<br>GPCR modulators | Online compound reordering<br>Online library search with compound availability | Yes (library synthesis only) | No |
| ChemDiv (www.chemdiv.com) | Large-scale druglike compound provider<br>Building block library<br>Focused/targeted libraries:<br>Kinase inhibitors<br>GPCR modulators<br>Phosphatase inhibitors<br>Ion channel modulators<br>Protease inhibitors<br>Nuclear receptor modulators<br>Antibacterial compounds<br>Antiviral compounds<br>Hsp90 modulators<br>CB2 modulators<br>NaPi2b cotransporter modualtors<br>Recognition motif and peptidomimetics compounds<br>Peptide mimic compounds<br>Natural products<br>Antimitotic compounds | ? | Yes | Yes |
| Enamine (http://www.enamine.net) | General druglike compound library<br>Focused/targeted libraries<br>Fragment-based library<br>Carboxylic acids library | Online compound reordering<br>Online library search with compound availability | Yes | No |
| Evotec | Focused library synthesis<br>Custom library solutions | | Yes | Yes |
| InterBioScreen (www.ibscreen.com) | Large-scale druglike compound provider<br>Focused/targeted libraries:<br>Analgesics<br>Antiasthmatic agents<br>Cardiovasculants<br>Antibacterial agents<br>Antidiabetic agents<br>Antidiuretic agents<br>Anti-inflammatory agents<br>Antipyretic agents<br>Antirheumatic agents<br>Antiulcer agents<br>Cancerostatic agents<br>Central nervous system (CNS) regulators<br>Fungistatic agents<br>Immunostimulant agents<br>Parasitocides<br>Psychostimulants and psychomimetic agents<br>Tuberculostatic agents<br>Virulicides<br>Natural product library<br>Building block libraries | | Yes | No |

**TABLE 1** ***Continued***

| Name | Libraries/Core Offerings | Ordering | MedChem Support? | AdmeTox Services? |
|---|---|---|---|---|
| Interchim (www.interchim.com) | Large-scale druglike compound provider<br>Building block/intermediate compound libraries | Online compound reordering<br>Online library search | Yes | No |
| Key Organics/Bionet (http://www.keyorganics.ltd.uk) | General druglike compound library | | Yes | No |
| LeadQuest (formerly Tripos) | General druglike compound library<br>Scaffold/building block libraries<br>Focused/targeted libraries:<br>Ion channel blockers<br>GPCR modulators | | Yes (focused libraries only) | No |
| LaboTest (www.labotest.com) | Large-scale druglike compound provider<br>Building block/intermediate library | | No | No |
| Life Chemicals (www.lifechemicals.com) | Large-scale druglike compound provider<br>Building block library<br>Druglike fragment library<br>Focused/targeted libraries:<br>Kinase inhibitors<br>Ion channel modulators<br>GPCR modulators<br>Nuclear receptor modulators<br>Protease inhibitors<br>Anticancer agents<br>Adenosine receptor agonists<br>Human fibroblast growth factor receptor 1 tyrosine kinase (FGFR1K) modulators<br>3-Phosphoinositide-dependent kinase-1 (PDK-1) modulators<br>cAMP-dependent protein kinase (PKA) modulators<br>Phosphatidylinositol 3-kinase (PI3K) modulators<br>Human Janus kinase 2 (JAK 2) modulators<br>Aurora B Kinase receptor-based modulators<br>Vascular endothelial growth factor receptor (VEGFR) kinase modulator | Online library search | Yes | No |
| Maybridge (part of Thermo Fisher) (www.maybridge.com) | Large-scale druglike compound provider<br>Building block/intermediate library | Online compound reordering<br>Online library search | Yes | No |
| MicroSource Discovery Systems (www.msdiscovery.com) | Focused/targeted libraries:<br>U.S. clinical compound (USP drug) library<br>Known bioactive compound library<br>NINDS compound library<br>Natural product library<br>Cytotoxic compound library<br>Anticancer compound library | | No | No |
| NanoSyn (www.nanosyn.com) | Custom library synthesis<br>Lead development | | Yes | No/? |
| Oakwood Chemical (www.oakwoodchemical.com) | General druglike compounds with emphasis on fluorine and sulfur compounds | Online library search with compound availability | Yes | No |

(*Continued*)

**TABLE 1** ***Continued***

| Name | Libraries/Core Offerings | Ordering | MedChem Support? | AdmeTox Services? |
|---|---|---|---|---|
| Otava (www.otavachemicals.com) | Large-scale druglike compound provider<br>Focused/targeted libraries:<br>Death-associated protein kinase 2 (DAPK-2) inhibitors<br>p38 MAP kinase alpha inhibitors<br>Vascular endothelial growth factor receptor 2 (VEGFR2) inhibitors<br>Aurora B kinase inhibitors<br>Janus kinase 2 (Jak2) inhibitors<br>Phosphatidylinositol 3-kinase(PI3K) inhibitors<br>3-phosphoinositide-dependent kinase-1 (PDK-1) inhibitors<br>Estrogen receptor (α) inhibitors<br>Cyclin-dependent kinase 2 (CDK2) inhibitors<br>Glycogen synthase kinase-3 (GSK3) inhibitors<br>Acetylcholinesterase (AChE) inhibitors<br>Phospholipase A2 (PLA(2)) inhibitors<br>Fibroblast growth factor receptor 1 tyrosine kinase (FGFR1K) inhibitors<br>cAMP-dependent protein kinase (PKA) inhibitors<br>2,3-Oxidosqualene lanosterol cyclase inhibitors<br>Lipoxygenase inhibitors<br>Adenylate cyclase inhibitors<br>Aldose reductase inhibitors<br>ATPase inhibitors<br>Cathepsin B inhibitors<br>Cyclic AMP phosphodiesterase inhibitors<br>Cyclic GMP phosphodiesterase inhibitors<br>Cyclooxygenase inhibitors<br>Epidermal growth factor receptor kinase inhibitors<br>Lipid peroxidase inhibitors<br>Phosphodiesterase V inhibitors<br>Ribonucleoside diphosphate reductase inhibitors<br>Ribonucleoside triphosphate reductase inhibitors<br>Sodium channel blockers<br>Tyrosine kinase inhibitors | Online library search | Yes | No |
| Pharmeks (www.pharmeks.com) | Large-scale druglike compound provider | Online library search | Yes | No |
| Peakdale Molecular (www.peakdale.co.uk) | General druglike compound library<br>Focused/targeted libraries:<br>GPCR modulators<br>Kinase modulators<br>Building block/intermediate library | | Yes | Yes |

**TABLE 1** ***Continued***

| Name | Libraries/Core Offerings | Ordering | MedChem Support? | AdmeTox Services? |
|---|---|---|---|---|
| Prestwick Chemical Company (www.prestwickchemical.com) | Peptide library<br>Focused/targeted libraries:<br>FDA-approved drugs<br>Known bioactive compounds<br>Natural products<br>*C. elegans* bioactive compounds<br>Building block library | | Yes | Yes |
| Princeton BioMolecular Research (www.princetonbio.com) | Large-scale druglike compound provider<br>Focused/targeted libraries:<br>Kinases inhibitors<br>GPCR modulators<br>Enzyme modulators<br>Ion channel blockers<br>Custom library synthesis<br>Building block library | | Yes | No |
| Rare Chemicals (www.rarechemicals.de) | Large-scale druglike compound provider<br>Building block library | | Yes | No |
| Ryan Scientific (www.ryansci.com) | Multivendor compound sourcing company (USA) | Online library search | No | No |
| Sigma-Aldrich (www.sigmaaldrich.com) | LOPAC known bioactive library:<br>Apoptosis modulators<br>GPCR modulators<br>Cyclic nucleotides<br>Gene regulation and expression modulators<br>Ion channel modulators<br>Lipid signaling modulators<br>Multidrug resistance modulators<br>Neurotransmission modulators<br>Phosphorylation inhibitors<br>Building blocks libraries<br>Cardiotoxicity ligand library | Online compound reordering with compound availability | Partial | No |
| Specs (www.specs.net) | Large-scale druglike compound provider<br>Building blocks library<br>Focused/targeted libraries:<br>Natural products<br>Ion channel blockers<br>GPCR modulators<br>Kinase inhibitors<br>Protease inhibitors<br>Signal pathway modulators<br>Nuclear receptor modulators<br>Central nervous system modulators<br>Alzheimer's disease related modulators<br>Cytostatic agents<br>Antimicrobial agents<br>Antiviral agents | Online compound reordering with compound availability | Yes | No |
| Synchem (www.synchem.com) | General druglike compound library<br>Building block library | Online library search with compound availability | Yes | No |
| Synthonix (www.synthonix.com) | General druglike compound library | Online library search with compound availability | Yes | No |

(*Continued*)

**TABLE 1** ***Continued***

| Name | Libraries/Core Offerings | Ordering | MedChem Support? | AdmeTox Services? |
|---|---|---|---|---|
| Timtec (www.timtec.net) | Large-scale druglike compound provider<br>Combinatorial libraries<br>Building block library<br>Focused/targeted libraries:<br>Natural product derivatives<br>GPCR ligands<br>Kinase modulators<br>Protease inhibitors<br>Serpins inhibitors<br>Potassium channel modulators<br>Anti-infective agents<br>*O*-GlcNAc transferase inhibitors<br>Neutrophil activators<br>Anthrax Inhibitors<br>Signal pathway modulators<br>Known bioactive compounds | Online library search | Yes | No |
| Toslab (www.toslab.com) | General druglike compound library<br>Building block library<br>Focused/targeted libraries:<br>Natural product derivatives<br>Cycline-dependent kinase (CDK4) inhibitors<br>p38 MAP kinase inhibitors<br>HIV1 protease inhibitors<br>3D-QSAR Custom generated libraries | | Yes | No |
| Vitas-M Laboratory (www.vitasmlab.com) | Large-scale druglike compound provider<br>Building block library<br>Druglike fragment library<br>Focused/targeted libraries:<br>PPAR-γ modulators<br>PPAR-α modulators<br>PPAR-δ modulators<br>Dihydrofolate reductase modulators<br>HMG-CoA reductase modulators<br>Vitamin D receptor modulators<br>Phospholipase A2 modulators<br>Thyroid hormone receptor β modulators<br>Liver X receptor α modulators<br>Liver X receptor β modulators<br>Dihydroorotate dehydrogenase modulators<br>Leukotriene A4 hydrolase modulators<br>β-Lactamase modulators<br>Estrogen receptor β modulators<br>Farnesoid X receptor modulators<br>Glucocorticoid receptor modulators<br>Mineralocorticoid receptor<br>HIV-protease inhibitors<br>Natural product derivatives | | Yes | No |

using various filters and criterias. Screening compounds typically come in three forms: as loose powder in glass vials, as film in microtiter plates, and as DMSO solution in microtiter plates. These storage forms have different stabilities that need to be considered [78]. If powder was received, it needs to be weighed, dispensed into plates, and dissolved in DMSO.

From these master plates, which are typically at a compound concentration of 10 m*M*, one or more 1 m*M* daughter copies are made. The plates are normally stored under the exclusion of water at +4°C (if DMSO) or −40°C if powder. The active daughter plate is kept at room temperature, if possible under the exclusion of water, and retired after 3–6 months, depending on use. DMSO, which is the best solvent for compounds, is unfortunately extremely hygroscopic, which makes is hard to keep the library integrity high. Left unsealed, DMSO will take up about 30% water [79, 80] at which point the compound solubility is reduced significantly. Furthermore, compounds will decay much faster in the presence of water. The decay process is not well understood, and it is generally difficult to come up with one mechanism since compound collections tend to have up to millions of members that fall into many different scaffolds. It is an advantage to be able to store compounds under inert atmosphere, but water remains the main culprit in loss of library integrity [81].

### 2.7 Postscreening Work and Counterscreens

Recent advances in automation and technology have made the execution of a screen very rapid. Consequently, the size of screened compound collections has grown, putting more emphasis on the postscreening efforts, which consist of data analysis, calculation of hit rates, reconfirmation of hits, counterscreening, dose–response experiments, and specificity profiling of the obtained hits.

***Data Analysis*** Data analysis is a central part of the screening process. Errors are very hard to repair since they typically result in loss of information. Data analysis generally consists of data assessment, correction, and mining (see Fig. 4). The first step in data analysis should be the quality assessment for each plate run during the screen and the sentinel plates, if employed. Since each plate should contain a set of positive and negative controls, a $Z'$ value for each plate can be calculated and plates that might need rescreening can be flagged. The values for $Z'$ and the positive and negative controls can

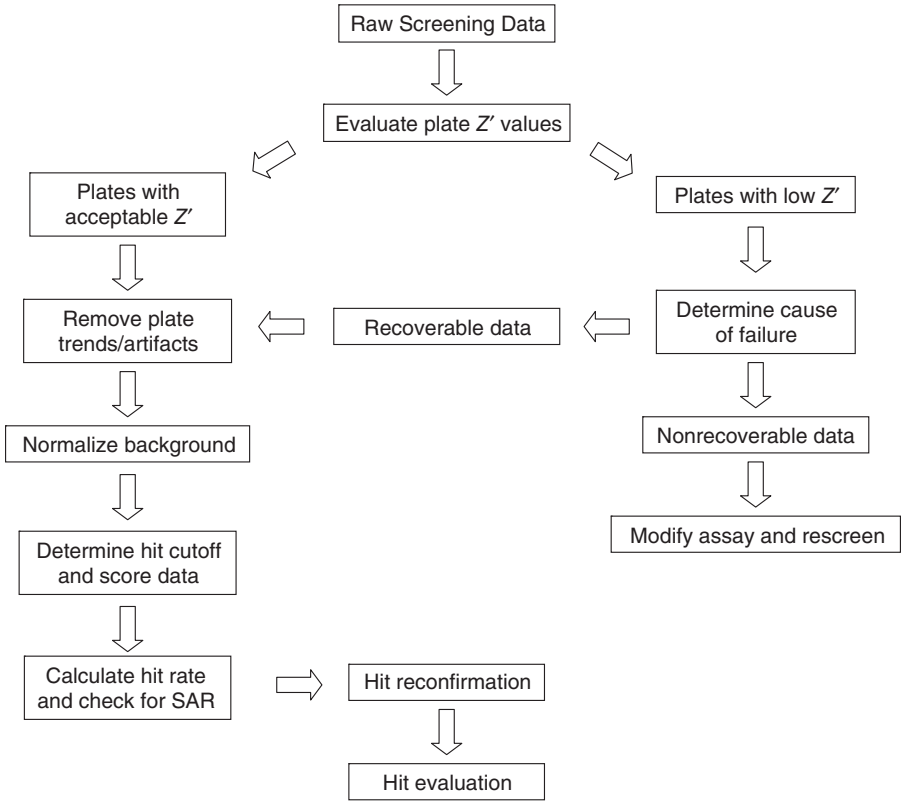


**Figure 4** General data mining work flow.

be compared over the entire screen and potential problems with, for example, drift of the signal due to signal instability during the screening will become apparent. The next step will be the evaluation of the plate background. The central idea is to check for plate trends and patterns. This can be very difficult work since the background can shift from day to day, reagent batch to reagent batch, or even plate to plate. Makarenov et al. have taken a very elegant approach in using *k*-means partitioning in their HTS corrector package, which allows for automatic assessment of the background and divides the plates into groups with similar background [82–84]. If strong plate trends or patterns are found, their causes need to be investigated before corrective action can be taken. After removal of the background, normalization using the positive and negative controls can be performed for each plate. This can be done as well on a per-screen basis but per plate is most commonly used.

At this point a cutoff needs to be defined in order to select a finite number of hits. Generally speaking, a cutoff that is the sum of three standard deviations plus the average of the control value can be used but, depending on the libraries (e.g., bioactive or focused libraries versus naïve libraries), this might be too much or not high enough. It is simple to experiment with thresholds using a database system or using a graphical representation of the screening data in order to select the cutoff. Useful programs include SpotFire, HTS corrector, GraphPad Prism, and Pipeline Pilot.

At this point it is good practice to examine the hit compounds for SAR series. Most commonly various pharmacophores will be found that have—depending on their moieties they are substituted with—various degrees of efficacy.

The total absence of these SAR clusters should be a strong warning sign that reevaluation of the results is necessary. The emergence of SAR contributes greatly to one's confidence in the screening results. Data from genomics approaches such as siRNA screens require a slightly modified approach toward data mining as the data type generated is different–please see Section 4.1 for details.

***Hit Rate and Reconfirmation*** After the hits are selected, the hit rate can be calculated. The hit rate is the percentage at which compounds are found to score positive during the primary screening. The hit rate is one of the first numbers to be examined directly after completion of a screen as it can give important information on how well the screen went.

A typical hit rate for a naïve library might be between 0.1 and 0.5% for target-based screens and 0.2–1% for cell-based screens. Depending on the assay, screening strategy and the libraries employed, these values fluctuate and sometimes even a hit rate somewhat higher than 1% can be found during cell-based screens.

A hit rate that is significantly higher than 1% indicates either problems with the libraries or the assay system used for screening. On the other hand, a very low hit rate might either indicate that the concentration of the compound is too low or that the target is very difficult to hit, or possibly the wrong type of library has been chosen. As we will see below, in the case of screens that have difficult targets (especially targeted assays) very clean libraries, that is, libraries that do not contain nuisance compounds, tend not to deliver many hits, resulting possibly in a hit rate that is much lower than average [85]. However, the few hits that are found tend to be specific and do normally warrant follow-up. The next step in screening is to consolidate the hit compounds into a custom library and reconfirm their activity in the assay used for screening. The reconfirmation rate is the percentage at which hit compounds that have been identified during primary screening repeat their activity during subsequent tests. Generally speaking, a high reconfirmation rate (>60–80%) indicates a well-functioning screen and a good execution on well-calibrated equipment.

***Evaluating Your Hits: What You Screened for Is What You Get—Precisely*** After reconfirmation of the hits in the original assay, the main question remaining is if hit compounds are doing *only* precisely what was desired or if they have been selected by the screen based on another property the screener is not aware of, and compounds have to be evaluated toward undesirable properties.

At this point it is important to keep in mind that you get what you screen for—if you are aware of what you are screening for or not. For example, a cell-based gain of signal screen for agonists using a reporter gene that is read after 72 h will find the following: agonist of the reporter gene and as well promitotic compounds that simply amplify the number of cells in a given well without changing the expression of the reporter gene. Or an antagonist screen for the same reporter will pick up any compounds which are cytotoxic or inhibitors of the reporter gene protein as well.

***Counterscreening*** It is the purpose of counterscreening to filter the true positive compounds from the false positives. The design of the necessary counterscreens should have been completed during the conception of the assay. Care in the design of the counterscreening strategy is especially necessary where a broader approach such as pathway-based screening or black-box screen has been chosen since not all reasons for false

positives are obvious. Counterscreens have to be designed such that the biological relevance of the hits can be ascertained and false positives can be minimized. Toward this end, counterscreens will generally assess the properties of the hit compounds in regards to, for example, inhibition of the reporter gene protein, activity promiscuity toward other proteins, cytotoxicity, and the like. We will now have a closer look at the reasons why a compound might appear as false positive during screening.

According to Lipinski [85], false positives can be divided into two categories: statistically or chemically caused false positives. Given a $Z' > 0.5$, the number of statistically caused false positives will be very small, since they are the statistical fluctuations or anomalies encountered during screening. Upon retesting, they will be weeded out since these false positives are random, they do not repeat, and have nothing to do with the compounds that were tested in the wells.

Chemically caused false positives, however, do reconfirm: They are caused by a problem related to the compound in the tested well. Compounds of this type are sometimes called nuisance compounds or "frequent hitters" [86] and chemically they fall into at least one of the following classes:

- Reactive compounds—such as reducing agents or covalent modifiers
- Promiscuous aggregators
- Detergents
- Optical interferers
- Cytotoxic or mitogenic agents
- Ion chelators
- Lysosomothrophic agents
- Inhibitors of reporter gene product proteins

It would be a good idea to have a medicinal chemist look through the entire compound deck before purchasing it in order to detect "junk" compounds. Most of these nuisance compounds have in common that they will come up in a multitude of assays. Bioactive compounds will too, but nuisance compounds will come up frequently in biologically unrelated screens, whereas bioactive compounds should exhibit an activity pattern that makes biological sense.

Ridding the panel of hits from nuisance compounds can be time intensive and challenging. General strategies include:

- Inclusion of, for example, detergent in the assay during retesting in order to diminish aggregator activity by scavenging the aggregating compound [87]
- Measuring "compound only" controls in order to detect optical interferers
- Quantifying viable cells per well to determine toxicity or mitogenicity

For dealing with the remaining ion chelators and reactive compounds, an experienced medicinal chemist's help is invaluable.

A further problem is typically the lack of purity of compounds. Even at a compound purity >90%—which is nowadays the industry standard—it is possible that a minor constituent in the compound solution is the active chemotype. This problem will be encountered after reordering compound from the supplier or making a larger, cleaner batch of compound and losing all bioactivity. In order to solve this problem the remaining compound from the library can fractionated by, for example, high-performance liquid chromatography (HPLC); and the active ingredient can be purified before being used in rescreening. Alternatively, all compounds on the deck can be by purified by preparative LC/MS [77]—however, the infrastructure and associated cost can be prohibitive.

A high number of false positives carry a heavy penalty in the labor involved for follow-up work. For every hit compound found, one compound needs to be replated, rescreened in multiple replicates, and if the compound passes the initial reconfirmation, it needs to be characterized.

This has the potential to offset the benefits of walkaway automation of the initial screening, as during the hit characterization the utilized assays tend to become more labor intensive. It is, therefore, crucial to automate the cherry picking and retesting of compounds and confirm the identity of the chemotype in order to keep the labor and possibilities for errors to a minimum.

***Dose–Response Experiments*** After hit confirmation, counterscreening, and confirmation of the chemical identity of the hit compounds, the potency of the hit compounds needs to be determined. The main tasks in potency determination include remaking of the compound stocks, performance of the dilution series, application of the dilution series, and performing of the assays. It should be noted that potency is not the same as efficacy—the first relating to the $IC_{50}$ or $EC_{50}$ and the later relating to the maximal quantity of the response caused by the compound.

Since the compound solvent DMSO is very hygroscopic the precise concentration of the compound in the screening plate set is not really known. Although nondestructive methods for determination of the water content are available [88], it is advisable to go back to compound powder if at all possible and remake the

stocks for potency testing, especially considering possible decay products from the compounds due to extended storage in DMSO. This labor-intensive process can be automated. Dry powder dispensers and automated weighting stations are available from various manufacturers. Alternatively, the purification of the compounds can be automated as well [77, 89]. The new compound stocks are then diluted, which is a simple task using automated liquid handlers. The dilution series can then be applied to the assay plates used for profiling using a pin tool. This strategy has the advantage that it becomes very easy to profile the hits on a larger range of assays without having to remake the dilution series, which saves compound and labor cost. However, it might be necessary to start with higher compound concentrations than the concentrations used during the initial screening, and in rare cases the compounds can stick to the pins. It is generally possible to clean the pin tools very efficiently using a combination of water, DMSO, ethanol, and methanol. An alternative strategy is to pipette the compounds into empty plates and add the assay constituents on top.

The $IC_{50}$ values or $EC_{50}$ values are then evaluated using curve-fitting software such as GraphPad Prism, SpotFire, Pipeline Pilot, CDD, and so on. It is imperative to check the quality of the curve fit, thus making the evaluation of the data a little labor intensive not only in terms of experimental labor but also in terms of data mining. After the dose–response experiments the initially observed SAR clusters of similar compounds should be confirmed and possibly broadened. It is good practice to perform SAR by inventory, that is, rescreen all compounds that are part a given SAR series available in the in-house libraries as well as compound suppliers, even if they appeared to have no effect in the screen.

### 2.8 Specificity Profiling of Compounds

Specificity profiling and mode of action studies are the next step in compound development after dose–response experiments. Compounds can be true positives with good potency and efficacy but still have undesirable secondary properties that render them useless for subsequent development into therapeutics. This goes beyond the concept of counterscreening, which addressed obvious problems such as compounds with frequent hitter profile such as toxic compounds. A major concern is specificity, especially for inhibitors of proteins where a closely related protein(s) has similar activities such as the cathepsins [90–93]. Or, for example, a kinase inhibitor, which has been found to inhibit BCR-ABL in a target-based assay, could still have significant inhibitory potential on the insulin receptor. This does not necessarily come as a surprise given the fact that most kinase inhibitors bind into the active site of the kinase. The active site of kinases is very well conserved since kinases operate using ATP as substrate, and thus the binding pocket is extremely conserved to the point that enables rational design within this pocket [94, 95]. Thus, even for a lead compound that generated from target-based screening the execution of specificity profiling is necessary. This can be fully automated [96]. Furthermore, it is becoming increasingly clear that most effective drugs have more than one target [96, 97], and there is as well more and more evidence from the clinics that certain indications, for example, oncology, require more than one target to be hit.

Another aspect of specificity is the genetic background in which a compound is tested. For example, in the area of oncology it would be desirable to study the compounds on cell lines that have different statuses in all relevant proto-oncogenes. It would be valuable to the drug development program to profile the compounds in a multitude of genetic backgrounds in order to get an understanding of how the compound will act functionally in these backgrounds. The fact that the human genome gains a big portion, if not most of its diversity, by alternative splicing—a process that is far from being wholly understood—makes a functional approach toward hit validation even more appealing [98].

## 3 TARGET DISCOVERY AND HIT VALIDATION FOR NCEs

The main challenge after hits are generated in a particular screen is to find their target(s) and the mode of action. In the specificity profiling examples in the previous chapter (see chapter 2.8), it is very simple since the hit compounds come from a target-based screen. To validate these compounds, one performs basic enzymology tests in order to determine the mode of action and test a panel of the most important kinases and close relatives of the enzyme targeted for off-target activity for biological hit validation. However, one generally does not have this ability with hits that were generated from a reporter gene assay screen or even a black-box screen: These hits were derived from cell-based assays and, therefore, involve a myriad of possible targets and many modes of action.

First of all, it might be a good idea to try to group the hits from a screen into compounds that have structural similarity. Typically, hit grouping will generate a small SAR series around one or more core structures from primary screens. In the ideal case SAR information with

potency and efficacy for each compound is available. Sometimes, single active compounds are found.

Second, these compounds can be combined into pairs. If their effect is cumulative, it is likely that they have different modes of action. The same strategy can be applied by pairing the novel compounds with reference compounds with known modes of action. In both cases knowledge of the $EC/IC_{50}$ values is of great help. Furthermore, secondary biochemical assays that look at phosphorylation states of key proteins (e.g., kinases) in the pathway of interest might give indications if the intended pathway is being affected. It may be necessary to use complicated means in order to assess the mode of action of a given hit such as proteomics, affinity chromatography, or protein microarrays to find the target [99–101]. Some methods will be discussed in greater detail in the following sections.

### 3.1 siRNA-, shRNA-, and cDNA-Based Screens

Loss of function (LOF) screens have recently gained much importance in drug and target discovery [15, 102], and it is to be expected that gain of function (GOF) screens will become equally important if they are not already. Besides the use of small-molecule libraries, LOF screens can be executed using siRNA or shRNA libraries [103]. siRNAs are short synthetic RNA duplexes that inhibit the transcription of their complementary mammalian genes. In the case of shRNA a plasmid containing an insert composed of a sequence that generates a stem-loop RNA driven by a polymerase II promoter can also produce siRNA via processing by Drosha in the nucleus and further processing in the cytoplasm (shRNA). At this time, the synthetic siRNA libraries are readily commercially available since their application in HTS has become routine [104]. A popular utilization strategy for siRNA in HTS is reverse transfection, where siRNA solutions are prespotted into 384-well plates—about 1–2 μL of a 1-μ*M* solution. These plates can then be immediately used or frozen for future use. The next step is the addition of a suitable transfection cocktail containing the transfection reagent and antibiotic and serum-free media in order to both resuspend the siRNA and form complexes. Typical incubation times range from 30 to 60 min. After the incubation time, cells are added in serum containing media. This scheme works extremely well with HEK-293T cells, and Hela cells. The main shortcoming of this system is that it relies on the transfectabillity of the cell line used. Therefore, it would be advisable to design assays based on cells that are compatible with downstream secondary screens using siRNA during the initial planning of the drug discovery campaign. Furthermore, transfection reagents tend to have a cytotoxic character; therefore, they should be removed after the transfection [105]. It should also be expected that the transfection reagent will interact to some degree with the NCE, necessitating additional controls in the secondary assays. Another shortcoming of this system is the transient nature of the gene interference. The amount of siRNA a cell has taken up will decrease over time due to the breakdown of siRNA and doubling of cells. Typically, the maximum time this kind of assay can be run for is about 72 h. If the primary drug screening requires longer incubation times, another technology is necessary. This technology could be, for example, shRNA. The main difference between siRNA and shRNA is that shRNAs can be delivered using a viral vector. This has two major consequences: first of all, shRNA enables—if, for example, a lentiviral vector is being used—the use of hard to transfect cell lines or even primary cell lines. Second, due to the integration of the vector payload into the genome, the resulting cells express the shRNA in a stable manner, that is, the knock-down of a given shRNA over time can be expected to be consistent. Overall, the greatest challenge with siRNA- or shRNA-based screens is the lack of control over how much knock-down of any given gene is being achieved using a single siRNA or shRNA. This is partly addressed in sh/siRNA libraries by having multiple constructs for any given gene. Furthermore, for siRNA, off-target effects have been observed and controls should be conducted before completing hit validation [106]. As a general rule of thumb, at least two constructs that target distinct regions of a given gene need to show activity in order to validate a hit.

Gain of function screens using cDNA in either an expression plasmid, which can be transfected into cells, or in a viral backbone can be used accordingly. Generally speaking, GOF screens tend to give more robust readouts; however, it is expected that gain and loss of function screens will be complementary to each other rather than delivering the same results.

### 3.2 Synthetic Genetic Array

Yeast is a very interesting model organism for compound profiling since many key proteins in human and yeast cells are conserved. Yeast is a unicellular eukaryotic organism that has been studied heavily by geneticists, cell biologist, and biochemists. It is easily altered by exogenous and genomic manipulation and a systematic gene deletion project was completed in 2001 [107]. The assembled library of strains, each with a single, nonessential gene deleted, is commercially available from Invitrogen and has allowed the development of

synthetic genetic array (SGA) screening (reviewed in [108]). Built on the concept of synthetic lethality, SGA screens search for cellular inviabilities that arise only when two or more elements/processes—each of which on its own is nonlethal—in a biological process are impaired at the same time, lethal knockout combinations imply that gene pairs have a related or linked function, for example, in a pathway. [109] Traditional SGA screening in yeast exploits the ability of yeast to propagate efficiently in both a diploid and haploid form. The system consists of a haploid query strain and the arrayed library of haploid yeast, which is crossed in order to generate an arrayed progeny, contain both mutations in the same host.

For compound profiling purposes, one replaces the query strain with the compound in question, and the arrayed library is either plated onto agar or grown in liquid media containing the compound. Observation of synthetic effects of the compound on the members of the library should enable mapping of interactions of the compound with pathways. Of course, an arrayed overexpression library would be attractive for screening in the presence of the compound as well. Such libraries of yeast overexpression plasmids are transformed easily and readily obtainable. A similar approach can be used with libraries of essential genes attenuated through conditional knockouts.

In theory, any colony-forming microorganism can be used with SGA if an array of genetic mutants are arrayed in an ordered system. Most SGA screens use either liquid in microtiter plates, solid support agar, or a combination of both formats. Manual SGA screening using a handheld pin tool is performed in 96-well/colony formats, while automated systems are able to perform in 384, 1536, or higher densities due to their precision manipulation. If pinning operations onto agar are performed, consideration should be given to create "buffer zones" in the outermost rows and columns because of agar wicking around the edge of the plate. Another consideration in SGA is the background of the query strain. The yeast knockout collection is provided in a S288C background, and it is recommended that query strains be generated in this background. Keeping an isogenic background will reduce unforeseen false positives and negatives due to strain variations. Such variation can also reduce strain mating efficiency, an essential component of successful SGA screens. Specialized hardware for SGA screening is offered for Genetix (Qbot, QPix) and Singer instruments. Scoring of "hit" synthetic lethal interactions has been described through automated methods but can also be done by eye [110]. Multiple repeats of the library screen are recommended to improve the accuracy of the results. A high-density screening format can cut consumable costs, screening time, and the physical space required to house the plates.

### 3.3 Affymetrix-Based and Other Chip-Based Assays

Transcription profiling for target discovery of compounds is very rapid and can be very rewarding [111, 112]. Transcriptional profiling relies on the isolation of RNA from a sample treated with the NCE, which is followed by cDNA synthesis during which a dye is incorporated into the cDNA. A second sample, which is treated only with the vector for the NCE is processed in identical manner. This time the cDNA is labeled with a different dye, however. Both samples are adjusted to the same cDNA concentration and hybridized onto an oligonucleotide array containing the sequences of the genome of the organism from which the samples had been derived. After washing to remove unbound cDNA, the arrays are scanned at two wavelenghts to detect cDNAs from the two differently labeled samples. This quantification of the two dyes will allow for ratiometric readouts for every gene represented in the array, thus giving a precise snapshot of the differential expression profile of the untreated versus the sample treated with the NCE. This differential profile should allow for at least an educated guess as to which pathway is being targeted by the NCE. Part of the appeal of this strategy is that various samples can be processed in parallel, so that, for example, multiple cell lines and backgrounds can be examined. The downside of this approach is that only RNA levels are being measured. RNA levels do not account for changes in protein levels or activation or inactivation of a protein by posttranslational modifications. Therefore, this method gives only a crude indication of the actual overall picture of the effects of an NCE. Furthermore, some targets are underrepresented in the transcriptome [113], and the data analysis can be challenging since genomewide data will be acquired and evaluated using this technology platform [114].

Another promising application of the microarray platform is drug profiling with yeast fitness tests. This yeast viability-based system is similar to SGA except that all of the heterozygous deletion strains are pooled into a single culture tube. After addition of a drug, all of the strains simultaneously complete for growth in the culture. Microarrays are used to assess the relative "fitness" of each knockout strain in the drug-containing media. Putative drug target hypotheses can be developed by checking for strains that are vastly over or underrepresented in the culture pool. The idea is that gene deletions that cause hypersensitivity toward a particular drug may be involved with the target of that drug [115, 116].

An interesting combination of microarray and siRNA/cDNA-based technology is the reversely transfected cell microarray (RTCM) technology that relies on spotting of these reagents on a slide and then performing standard reverse transfection. The benefit of this strategy is miniaturization and the possibility of parallel processing of many cell lines. This would enable functional profiling of drug candidates with less cost and effort than non-microarray-based methods [117].

## 4 DATA MANAGEMENT

Without the ability to store, mine, and explore data, molecular screening is unthinkable. Data and data management plays a central role in molecular screening, but little software is available to the public to cross the borders of various system platforms [118, 119]. Most of the tools used in the pharmaceutical industry are written in-house, which means that they are unavailable to the general public. A central role in the process of data mining is data visualization as it allows for the representation of a large set of data without overwhelming the scientist, who then can develop a "feel" for the quality of the screen, if there are any artifacts and the like, just by the appearance of the graphical representation. While data handling and representation is one end of the pipeline, tying these data to structures and managing the inventory of drugs is on the other. Most of the work of connecting data with chemical structures is done using specialized cheminformatics programs such as Accelrys HTS, IDBS, and Pipeline Pilot by cheminformatics specialists. An excellent overview of the products and possibilities is given by Bunin et al. [120]. We will first explore which data types we might encounter and then see what kind of databases we might use to make the most of the knowledge buried in the data.

### 4.1 Data Types

Data come in various forms and sizes in molecular screening. Most data will be generated by plate readers that then consist of one or more numerical values per plate coordinate. The next level of complexity is time-resolved data, such as those generated by fluorometric imaging plate reader (FLIPR) assays. In such an assay all wells of a plate are imaged at the same time with fluorescence values being generated for each time point for each well. One big step up from this situation is image-based data generated by HCS [121, 122]. HCS data can be per-well averages or per-cell measurements. This might include as many as 30 measurements, which are taken from each cell. Per-cell measurements might be biologically the most intriguing kind of data but are very difficult to manage. Furthermore, a database may also contain the original images from the HCS, which complicates the situation further. Depending on the readout, the numerical value will be very different from screen to screen, making it difficult to compare numerical value across different screens. One way around it would be to normalize to positive or negative controls for each screen. Another solution is to remove the background and then divide by the standard deviation—either by plate or by screen. However, dimensionless data might be harder to intuitively understand since the intrinsic biological context of the unit (e.g., percent survival, which ranges from 0 to 100) is missing. However, even if the numerical problem is solved, the challenge of making biological sense of screening data across different screens and screening platforms remains.

Data types can also be divided into data from singlets, one or more replicates, data from pooled compound libraries (i.e., libraries that contain more than one compound per well), and data derived from genomic approaches that contain frequently more than one well targeting a particular gene. Chemical compounds are normally screened in singlet for cost reasons, sometimes compounds are even pooled in order to save money and only active wells are deconvoluted later on. This seems to work reasonably well, even in an industrial setting [20]. Orthogonal pooling strategies, that is, strategies in which a multitude of compounds per well is used, and the deconvolution is already implemented by screening the same compound twice but each time together with different compounds, work as well [123]. In all of these cases, the ranking of the hits is based on activity of the compounds.

Data from genomics approaches, such as siRNA and shRNA, require a different way of evaluating the data. Konig et al. introduced redundant siRNA activity (RSA) analysis [124]. The classical data mining focuses on the activity of a compound. In contrast, RSA analysis takes advantage of the redundancy of the libraries, which often contain two to four sequences targeting one gene. RSA analysis uses all redundant sequences for calculation of the probability for a given gene to be a real positive. In addition, the activity data of the individual wells are weighted into the ranking of the hits. The resulting hit lists lead to higher reconfirmation rates than activity-based analysis alone.

### 4.2 Database Types

Given the number of data points typically collected during screening and the number of compounds these data need to be attached do, data storage in text files or Excel spreadsheets is not an option. The only viable solution is a relational database, which provides built-in

mechanisms for nonredundant data storage and logical associations between various types of data. The two most frequently used database engines are Oracle and MySQL. Oracle is a commercial, "industrial-strength" application, which is a good choice if the database needs advanced interactive features and the ability to be bulletproof even under the most severe circumstances—which is why all top financial institutions use Oracle. Oracle should be installed and maintained by a trained database administrator (DBA), and it is a rather expensive software package. MySQL is an open-source database engine and easier to set up, maintain, and program. It is less expensive than Oracle.

### 4.3 Accessing Data and Integrating Various Platforms for Maximal Efficiency

Before considering which database type the data should be stored in, it makes sense to think about whether to buy a software package for cheminformatics or hire a programmer as this might dictate which database type is necessary. The choice will depend on a variety of considerations such as whether customized applications are necessary or if an off-the-shelf product will perform the tasks needed within the budget available. Frequently, off-the-shelf software is also dependent on a particular database engine and, thereby, imposes specific restrictions on the underlying architecture. Therefore, it is important to evaluate the required feature set and make an informed decision as to which software–database combination to implement. An overview of the products and capabilities is given in Bunin's book and Ling's reviews [119, 120].

In order to answer the question of which kind of database interface to use, the question of who needs to access the data needs to be answered first. Is it a dedicated cheminformatics person or an end user? These two models necessitate different kinds of interfaces/databases. In the first case, an expert system could be used that requires a dedicated person to use it and make data available to the end user. On the other extreme is a system that is easy enough to use for a nonspecialist and enables the scientist to store data, mine, and make use of the infrastructure of the screening laboratory.

The question that stands behind this—is an expert system or a more user-friendly interface employed—is whether a screening laboratory becomes a "full-service" or a "self-service" entity. The advantages of full service, where screens are being dropped off and data are being returned, could be efficiency in the usage of resources. On the other hand, it is to be expected that only a scientist who is very familiar with a given biological system can ensure the best quality of hits from screening campaigns because he knows the biology best. The scientist will notice early on if the system is behaving strangely, for example, if the morphology of the cells is slightly off or similar things that might not be easily quantifiable and therefore difficult to mine for but are early indicators of either success or failure. The model of a visiting scientist works extremely well in my experience.

Frequently, data need to be shared between different locations or collaborators at different institutes. In this particular case, a Web-based interface is a good solution. In my experience, one of the best examples of easy-to-use, Web-based software for data storage and mining is made by Collaborative Drug Discovery (www.collaborativedrug.com, last accessed 9/13/08), which enables scientists who have a minimal knowledge of chemistry to archive, view, and mine their data. A novel trend is to integrate outside database content into internal databases. This is complicated by the fact that there is not one single search engine for this type of problem. While using Pubmed might become a central site for gathering this type of information [125], Chemspider (www.chemspider.com, last accessed 9/13/08) goes the other route. It is a metaengine for finding information on a given molecule in many databases. This search engine scours the Internet for information—like Google. However, it does not use a search phrase but a chemical structure.

As we have seen earlier in this chapter, data from genomic screens require different mining strategies than data from chemical screens. This is further complicated by the fact that compounds are organized by similarity while genomic libraries are organized by gene ontology. Currently, these two data types are normally kept in separate databases. However, accessing data from chemical and genome screens at the same time and integrating the data into one platform will make molecular screening and drug discovery much more efficient. Currently, the means are missing to integrate genomic and chemical screening data. Once cheminformatics and bioinformatics have merged [126], drug discovery will dramatically accelerate.

## 5 CONCLUSION

We have given an introduction to molecular screening, its underlying concepts, work flows, and novel applications as research tool and for target discovery. Molecular screening has changed from a very specialized tool that was primarily used in drug discovery to an indispensable tool for research. The integration of genomic and chemical screening data will lead to a better understanding of disease mechanisms and enable a more efficient drug discovery process.

## ACKNOWLEDGMENTS

I want to thank Etienne Damoiseaux and Dr. Peter Stoilov for proofreading, Sam Hasson for proofreading and his help with the SGA paragraph and figures, Loren Miraglia and Dr. Sasha Buzko for proofreading and many helpful discussions, Tony Orth, Jeff Janes, Achim Brinker, and Barry Bunin for helpful discussions.

## REFERENCES

1. Lander ES, et al. (2001). Initial sequencing and analysis of the human genome. *Nature 409*(6822), 860–921.
2. Lindsay MA (2003). Target discovery. *Nat Rev Drug Discov 2*(10), 831–838.
3. Workman P (2001). New drug targets for genomic cancer therapy: Successes, limitations, opportunities and future challenges. *Curr Cancer Drug Targets 1*(1), 33–47.
4. Pereira DA, Williams JA (2007). Origin and evolution of high throughput screening. *Br J Pharmacol 152*(1), 53–61.
5. Burch RM, Kyle DJ (1991). Mass receptor screening for new drugs. *Pharm Res 8*(2), 141–147.
6. Houston JG, Banks M (1997). The chemical-biological interface: Developments in automated and miniaturised screening technology. *Curr Opin Biotechnol 8*(6), 734–740.
7. Taylor PB, et al. (2000). Automated assay optimization with integrated statistics and smart robotics. *J Biomol Screen 5*(4), 213–226.
8. Wu X, et al. (2006). Mycophenolic acid is a potent inhibitor of angiogenesis. *Arterioscler Thromb Vasc Biol 26*(10), 2414–2416.
9. Aza-Blanc P, et al. (2003). Identification of modulators of TRAIL-induced apoptosis via RNAi-based phenotypic screening. *Mol Cell 12*(3), 627–637.
10. Silva JM, et al. (2005). Second-generation shRNA libraries covering the mouse and human genomes. *Nat Genet 37*(11), 1281–1288.
11. Chanda SK, et al. (2003). Genome-scale functional profiling of the mammalian AP-1 signaling pathway. *Proc Natl Acad Sci USA 100*(21), 12153–12158.
12. Verkman AS (2004). Drug discovery in academia. *Am J Physiol Cell Physiol 286*(3), C465–474.
13. Gaither LA (2007). Chemogenomics approaches to novel target discovery. *Expert Rev Proteomics 4*(3), 411–419.
14. Perrimon N, et al. (2007). Drug-target identification in *Drosophila* cells: Combining high-throughout RNAi and small-molecule screens. *Drug Discov Today 12*(1–2), 28–33.
15. Natt F (2007). siRNAs in drug discovery: Target validation and beyond. *Curr Opin Mol Ther 9*(3), 242–247.
16. Wunder F, et al. (2008). Functional cell-based assays in microliter volumes for ultra-high throughput screening. *Comb Chem High Throughput Screen 11*(7), 495–504.
17. Achyuthan KE, Whitten DG (2007). Design considerations for high-throughput screening and in vitro diagnostic assays. *Comb Chem High Throughput Screen 10*(6), 399–412.
18. Landry Y, Gies J-P (2008). Drugs and their molecular targets: An updated overview. *Fund Clin Pharmacol 22*(1), 1–18.
19. Sakharkar MK, Sakharkar KR (2007). Targetability of human disease genes. *Curr Drug Discov Technol 4*(1), 48–58.
20. Bender A, et al. (2008). Which aspects of HTS are empirically correlated with downstream success? *Curr Opin Drug Discov Dev 11*(3), 327–337.
21. Zanella F, et al. (2007). An HTS approach to screen for antagonists of the nuclear export machinery using high content cell-based assays. *Assay Drug Dev Technol 5*(3), 333–341.
22. Baltz RH (2006). Marcel Faber Roundtable: Is our antibiotic pipeline unproductive because of starvation, constipation or lack of inspiration? *J Ind Microbiol Biotechnol 33*(7), 507–513.
23. Ladds G, et al. (2003). Modified yeast cells to investigate the coupling of G protein-coupled receptors to specific G proteins. *Mol Microbiol 47*(3), 781–792.
24. Ladds G, Davey J (2004). Analysis of human GPCRs in fission yeast. *Curr Opin Drug Discov Devel 7*(5), 683–691.
25. Zinder ND, Boeke JD (1982). The filamentous phage (Ff) as vectors for recombinant DNA—A review. *Gene 19*(1), 1–10.
26. Rines DR, et al. (2006). High-content screening of functional genomic libraries. *Methods Enzymol 414*, 530–765.
27. Kunapuli P, et al. (2005). Application of division arrest technology to cell-based HTS: Comparison with frozen and fresh cells. *Assay Drug Dev Technol 3*(1), 17–26.
28. Milan DJ, et al. (2003). Drugs that induce repolarization abnormalities cause bradycardia in zebrafish. *Circulation 107*(10), 1355–1358.
29. Chan J, et al. (2002). Dissection of angiogenic signaling in zebrafish using a chemical genetic approach. *Cancer Cell 1*(3), 257–267.
30. Langheinrich U (2003). Zebrafish: a new model on the pharmaceutical catwalk. *Bioessays 25*(9), 904–912.
31. Gribbon P, Sewing A (2003). Fluorescence readouts in HTS: No gain without pain? *Drug Discov Today 8*(22), 1035–1043.
32. Nikiforov TT, Simeonov AM (2003). Application of fluorescence polarization to enzyme assays and single nucleotide polymorphism genotyping: Some recent developments. *Comb Chem High Throughput Screen 6*(3), 201–212.

33. Bazin H, et al. (2001). Homogeneous time resolved fluorescence resonance energy transfer using rare earth cryptates as a tool for probing molecular interactions in biology. *Spectrochim Acta A Mol Biomol Spectrosc 57*(11), 2197–2211.
34. Fan F, Wood KV (2007). Bioluminescent assays for high-throughput screening. *Assay Drug Dev Technol 5*(1), 127–136.
35. Lavery P, Brown MJ, Pope AJ (2001). Simple absorbance-based assays for ultra-high throughput screening. *J Biomol Screen 6*(1), 3–9.
36. Deo SK, Daunert S (2001). Luminescent proteins from *Aequorea victoria*: Applications in drug discovery and in high throughput analysis. *Fresenius J Anal Chem 369*(3–4), 258–266.
37. Bader B, et al. (2001). A cGMP-dependent protein kinase assay for high throughput screening based on time-resolved fluorescence resonance energy transfer. *J Biomol Screen 6*(4), 255–264.
38. Dahan M, et al. (2001). Time-gated biological imaging by use of colloidal quantum dots. *Opt Lett 26*(11), 825–827.
39. Smith DS, Eremin SA (2008). Fluorescence polarization immunoassays and related methods for simple, high-throughput screening of small molecules. *Anal Bioanal Chem 391*(5), 1499–1507.
40. Rausch O (2006). High content cellular screening. *Curr Opin Chem Biol 10*(4), 316–320.
41. Rausch O (2005). Use of high-content analysis for compound screening and target selection. *IDrugs 8*(7), 573–577.
42. Hudson CC, et al. (2006). High-content screening of known G protein-coupled receptors by arrestin translocation. *Methods Enzymol 414*, 63–78.
43. Garippa RJ, et al. (2006). High-throughput confocal microscopy for beta-arrestin-green fluorescent protein translocation G protein-coupled receptor assays using the Evotec Opera. *Methods Enzymol 414*, 99–120.
44. O'Brien PJ, et al. (2006). High concordance of drug-induced human hepatotoxicity with in vitro cytotoxicity measured in a novel cell-based model using high content screening. *Arch Toxicol 80*(9), 580–604.
45. Zhang, Chung, Oldenburg (1999). A simple statistical parameter for use in evaluation and validation of high throughput screening assays. *J Biomol Screen Offic J Soc Biomol Screen 4*(2), 67–73.
46. Malo N, et al. (2006). Statistical practice in high-throughput screening data analysis. *Nat Biotechnol 24*(2), 167–175.
47. Lundholt BK, Scudder KM, Pagliaro L (2003). A simple technique for reducing edge effect in cell-based assays. *J Biomol Screen 8*(5), 566–570.
48. Valler MJ, Green D (2000). Diversity screening versus focussed screening in drug discovery. *Drug Discov Today 5*(7), 286–293.
49. Shoichet BK (2006). Screening in a spirit haunted world. *Drug Discov Today 11*(13–14), 607–615.
50. Rishton GM (2008). Natural products as a robust source of new drugs and drug leads: Past successes and present day issues. *Am J Cardiol 101*(10A), 43D–49D.
51. Macarron R (2006). Critical review of the role of HTS in drug discovery. *Drug Discov Today 11*(7–8), 277–279.
52. Engels MF, Venkatarangan P (2001). Smart screening: Approaches to efficient HTS. *Curr Opin Drug Discov Devel 4*(3), 275–283.
53. Janes J (2008). *Retrospective Analysis of Several Full-Screening Lead Discovery Efforts*. La Jolla, CA: Damoiseaux.
54. Rester U (2008). From virtuality to reality—Virtual screening in lead discovery and lead optimization: A medicinal chemistry perspective. *Curr Opin Drug Discov Devel 11*(4), 559–568.
55. Schnur DM (2008). Recent trends in library design: "Rational design" revisited. *Curr Opin Drug Discov Devel 11*(3), 375–380.
56. Joesch C, et al. (2008). Use of FLIPR membrane potential dyes for validation of high-throughput screening with the FLIPR and μARCS technologies: Identification of ion channel modulators acting on the GABAA receptor. *J Biomol Screen 13*(3), 218–228.
57. Weisman JL, et al. (2006). Searching for new antimalarial therapeutics amongst known drugs. *Chem Biol Drug Des 67*(6), 409–416.
58. Bisson WH, et al. (2007). Discovery of antiandrogen activity of nonsteroidal scaffolds of marketed drugs. *Proc Natl Acad Sci USA 104*(29), 11927–11932.
59. Stoilov P, et al. (2008). A high-throughput screening strategy identifies cardiotonic steroids as alternative splicing modulators. *Proc Natl Acad Sci USA 105*(32), 11218–11223.
60. Lipinski CA (2003). Chris Lipinski discusses life and chemistry after the Rule of Five. *Drug Discov Today 8*(1), 12–16.
61. Ganesan A (2008). The impact of natural products upon modern drug discovery. *Curr Opin Chem Biol 12*(3), 306–317.
62. Harvey AL (2007). Natural products as a screening resource. *Curr Opin Chem Biol 11*(5), 480–484.
63. Tan G, Gyllenhaal C, Soejarto DD (2006). Biodiversity as a source of anticancer drugs. *Curr Drug Targets 7*(3), 265–277.
64. Muller-Kuhrt L (2003). Putting nature back into drug discovery. *Nat Biotechnol 21*(6), 602.
65. Lam KS (2007). New aspects of natural products in drug discovery. *Trends Microbiol 15*(6), 279–289.
66. Mueller-Kuhrt L (2007). Successful, but often unconventional: The continued and long-term contribution of natural products to healthcare. *Expert Opin Drug Discov 2*(3), 305–311.
67. Vu H, Pham NB, Quinn RJ (2008). Direct screening of natural product extracts using mass spectrometry. *J Biomol Screen 13*(4), 265–275.

68. Koehn FE (2008). High impact technologies for natural products screening. *Progress in Drug Research. Fortschritte Der Arzneimittelforschung. ProgrŠs Des Recherches Pharmaceutiques 65*, 175, 177–210.
69. Ripka AS, Rich DH (1998). Peptidomimetic design. *Curr Opin Chem Biol 2*(4), 441–452.
70. Randolph JT, DeGoey DA (2004). Peptidomimetic inhibitors of HIV protease. *Curr Top Med Chem 4*(10), 1079–1095.
71. Liu Y, Gray NS (2006). Rational design of inhibitors that bind to inactive kinase conformations. *Nat Chem Biol 2*(7), 358–364.
72. Surratt CK, Adams WR (2005). G protein-coupled receptor structural motifs: Relevance to the opioid receptors. *Curr Top Med Chem 5*(3), 315–324.
73. Parrill AL (2008). Lysophospholipid interactions with protein targets. *Biochim Biophys Acta 1781*(9), 540–546.
74. Pan HL, et al. (2008). Modulation of pain transmission by G-protein-coupled receptors. *Pharmacol Ther 117*(1), 141–161.
75. Guintu C, et al. (2006). Just-in-time purification: An effective solution for cherry-picking and purifying active compounds from large legacy libraries. *J Biomol Screen 11*(8), 933–939.
76. Yuan D, et al. (2005). A novel method for immediate post-purification purity determination of fractions collected during high-throughput purification. *J Chromatogr A 1092*(2), 228–234.
77. Isbell JJ, et al. (2005). Purifying the masses: Integrating prepurification quality control, high-throughput LC/MS purification, and compound plating to feed high-throughput screening. *J Comb Chem 7*(2), 210–217.
78. Bowes S, et al. (2006). Quality assessment and analysis of Biogen Idec compound library. *J Biomol Screen 11*(7), 828–835.
79. Rasmussen DH, MacKenzie AP (1968). Phase diagram for the system water-dimethylsulphoxide. *Nature 220*(5174), 1315–1317.
80. Catalan J, Diaz C, Garcia-Blanco F (2001). Characterization of binary solvent mixtures of DMSO with water and other cosolvents. *J Org Chem 66*(17), 5846–5852.
81. Oldenburg K, et al. (2005). High throughput sonication: Evaluation for compound solubilization. *Comb Chem High Throughput Screen 8*(6), 499–512.
82. Kevorkov D, Makarenkov V (2005). Statistical analysis of systematic errors in high-throughput screening. *J Biomol Screen 10*(6), 557–567.
83. Makarenkov V, et al. (2006). HTS-Corrector: Software for the statistical analysis and correction of experimental high-throughput screening data. *Bioinformatics (Oxford, England) 22*(11), 1408–1409.
84. Makarenkov V, et al. (2007). An efficient method for the detection and elimination of systematic error in high-throughput screening. *Bioinformatics 23*(13), 1648–1657.
85. Lipinksi C (2005). *Chemistry quality and the biology/chemistry interface*. Paper presented at the *6th Winter Conference on Medicinal and Bioorganic Chemistry*. Steamboat Springs, CO.
86. Crisman T-J, et al. (2007). Understanding false positives in reporter gene assays: In silico chemogenomics approaches to prioritize cell-based HTS data. *J Chem Inf Model 47*(4), 1319–1327.
87. Feng BY, Shoichet BK (2006). A detergent-based assay for the detection of promiscuous inhibitors. *Nat Protoc 1*(2), 550–553.
88. Ellson R, et al. (2005). In situ DMSO hydration measurements of HTS compound libraries. *Comb Chem High Throughput Screen 8*(6), 489–498.
89. Popa-Burke IG, et al. (2004). Streamlined system for purifying and quantifying a diverse library of compounds and the effect of compound concentration measurements on the accurate interpretation of biological assay results. *Anal Chem 76*(24), 7278–7287.
90. Alper PB, et al. (2006). Arylaminoethyl amides as noncovalent inhibitors of cathepsin S. Part 2: Optimization of P1 and N-aryl. *Bioorg Med Chem Lett 16*(6), 1486–1490.
91. Chatterjee AK, et al. (2007). Synthesis and SAR of succinamide peptidomimetic inhibitors of cathepsin S. *Bioorg Med Chem Lett 17*(10), 2899–2903.
92. Epple R, et al. (2007). Bicyclic carbamates as inhibitors of papain-like cathepsin proteases. *Bioorg Med Chem Lett 17*(5), 1254–1259.
93. Tully DC, et al. (2006). Synthesis and SAR of arylaminoethyl amides as noncovalent inhibitors of cathepsin S: P3 cyclic ethers. *Bioorg Med Chem Lett 16*(19), 5112–5117.
94. Shokat K-M (1995). Tyrosine kinases: Modular signaling enzymes with tunable specificities. *Chem Biol 2*(8), 509–514.
95. Shah K, et al. (1997). Engineering unnatural nucleotide specificity for Rous sarcoma virus tyrosine kinase to uniquely label its direct substrates. *Proc Natl Acad Sci USA 94*(8), 3565–3570.
96. Melnick JS, et al. (2006). An efficient rapid system for profiling the cellular activities of molecular libraries. *Proc Natl Acad Sci USA 103*(9), 3153–3158.
97. Frantz S (2005). Drug discovery: Playing dirty. *Nature 437*(7061), 942–943.
98. Hagiwara M (2005). Alternative splicing: A new drug target of the post-genome era. *Biochim Biophys Acta 1754*(1–2), 324–331.
99. Hart CP (2005). Finding the target after screening the phenotype. *Drug Discov Today 10*(7), 513–519.
100. Heal W-P, Wickramasinghe SR, Tate EW (2008). Activity based chemical proteomics: Profiling proteases as drug targets. *Curr Drug Discov Technol 5*(3), 200–212.
101. Tao SC, Chen CS, Zhu H (2007). Applications of protein microarray technology. *Comb Chem High Throughput Screen 10*(8), 706–718.
102. Bartz S, Jackson AL (2005). How will RNAi facilitate drug development? *Sci STKE 2005*(295), pe39.

103. Hannon GJ, Rossi JJ (2004). Unlocking the potential of the human genome with RNA interference. *Nature 431*(7006), 371–378.
104. Willingham AT, et al. (2004). RNAi and HTS: Exploring cancer by systematic loss-of-function. *Oncogene 23*(51), 8392–8400.
105. Axel DI, et al. (2000). Toxicity, uptake kinetics and efficacy of new transfection reagents: Increase of oligonucleotide uptake. *J Vasc Res 37*(4), 221–234; discussion 303–304.
106. Tschuch C, et al. (2008). Off-target effects of siRNA specific for GFP. *BMC Mol Biol 9*, 60.
107. Kelly DE, Lamb DC, Kelly SL (2001). Genome-wide generation of yeast gene deletion strains. *Comp Funct Genomics 2*(4), 236–242.
108. Boone C, Bussey H, Andrews BJ (2007). Exploring genetic interactions and networks with yeast. *Nat Rev Genet 8*(6), 437–449.
109. Ooi SL, et al. (2006). Global synthetic-lethality analysis and yeast functional profiling. *Trends Genet 22*(1), 56–63.
110. Tong AH, Boone C (2006). Synthetic genetic array analysis in *Saccharomyces cerevisiae*. *Methods Mol Biol 313*, 171–192.
111. Wang S, Cheng Q (2006). Microarray analysis in drug discovery and clinical applications. *Methods Mol Biol 316*, 49–65.
112. Clarke PA, et al. (2001). Gene expression microarray analysis in cancer biology, pharmacology, and drug development: Progress and potential. *Biochem Pharmacol 62*(10), 1311–1336.
113. Dechering KJ (2005). The transcriptome's drugable frequenters. *Drug Discov Today 10*(12), 857–864.
114. Chen JJ, et al. (2007). Selection of differentially expressed genes in microarray data analysis. *Pharmacogenomics J 7*(3), 212–220.
115. Parsons AB, et al. (2006). Exploring the mode-of-action of bioactive compounds by chemical-genetic profiling in yeast. *Cell 126*(3), 611–625.
116. Lum PY, et al. (2004). Discovering modes of action for therapeutic compounds using a genome-wide screen of yeast heterozygotes. *Cell 116*(1), 121–137.
117. Sturzl M, et al. (2008). High throughput screening of gene functions in mammalian cells using reversely transfected cell arrays: Review and protocol. *Comb Chem High Throughput Screen 11*(2), 159–172.
118. Yan SF, et al. (2006). Learning from the data: Mining of large high-throughput screening databases. *J Chem Inf Model 46*(6), 2381–2395.
119. Ling XB (2008). High throughput screening informatics. *Comb Chem High Throughput Screen 11*(3), 249–257.
120. Bunin BA, MG, Bajorath J, Siesel B (2007). *Chemoinformatics: Theory, Practice, and Products*. New York: Springer-Verlag.
121. Dunlay RT, Czekalski WJ, Collins MA (2007). Overview of informatics for high content screening. *Methods Mol Biol 356*, 269–280.
122. Low J, et al. (2008). Prioritizing hits from phenotypic high-content screens. *Curr Opin Drug Discov Devel 11*(3), 338–345.
123. Motlekar N, Diamond SL, Napper AD (2008). Evaluation of an orthogonal pooling strategy for rapid high-throughput screening of proteases. *Assay Drug Dev Technol 6*(3), 395–405.
124. Konig R, et al. (2007). A probability-based approach for the analysis of large-scale RNAi screens. *Nat Methods 4*(10), 847–849.
125. Rosania GR, et al. (2007). A cheminformatic toolkit for mining biomedical knowledge. *Pharm Res 24*(10), 1791–1802.
126. Sukumar N, Krein M, Breneman CM (2008). Bioinformatics and cheminformatics: Where do the twain meet? *Curr Opin Drug Discov Devel 11*(3), 311–319.

# ACCELERATOR MASS SPECTROMETRY AND POSITRON EMISSION TOMOGRAPHY IN HUMAN MICRODOSING

ZHENGDONG CHENG

*Artie McFerrin Department of Chemical Engineering, Professional Program in Biotechnology and Materials Science and Engineering Program, Texas A&M University, College Station, Texas*

## 1 INTRODUCTION

Human microdosing is a novel method that permits improved drug candidate selection by taking the investigational drugs into humans earlier to obtain human metabolism data. In 1991, there was a 40% failure rate associated with inappropriate drug metabolism and pharmacokinetics of candidate drug molecules in the traditional early clinical drug development (phase I) [1, 2]. Over the subsequent decade (1991–2000), attrition for poor pharmacokinetic properties decreased considerably and represented less than 10% of total attribution. More recently, lack of efficacy and/or inadequate safety (clinical or toxicological) have become the primary reasons for attrition [attrition was highest during phase II (62%) but still significant in phase III (45%) and at the time of registration (23%)] [3]. On average, it takes 10–12 years for a molecule from discovery to regulatory approval. The cost of drug development is now greater than $800 million per registered drug [4]. The costs of compounds abandoned during testing were linked to the costs of compounds that obtained marketing approval [5].

There are three main techniques that are used in microdosing studies: AMS (accelerator mass spectrometry), PET (positron emission tomography), and liquid chromatography–tandem mass spectrometry (LC-MS/MS). Both AMS and PET provide much greater sensitivity than LC-MS/MS, allowing the quantification of very low drug levels in plasma after microdosing. However, microdose/AMS or PET studies are labor and resource intensive compared to LC-MS/MS. PET can provide pharmacodynamic (PD) information, whereas AMS provides pharmacokinetic information (PK).

## 2 DEFINITIONS

*ADME* A drug's absorption, distribution, metabolism, and excretion characteristics. A drug's PK and PD define the ADME.

*Human Microdosing* A technique for studying the behavior of a drug candidate in humans through the administration of low dose at levels typically about 100 times lower than the therapeutic dosage predicted from animal in vitro models, and also not exceeding 100 μg, but high enough to evaluate PK properties of the candidate [6, 7]. This technique is also called trace-dose human ADME screening studies, or human phase 0 trials, in contrast to "phase 0 studies" or preclinical studies conducted in animals [8].

*Microdose* A subpharmacological active dose of drug: 1/100th of the projected pharmacologic dose in humans or a maximum of 100 μg [9].

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

*AMS* Accelerator mass spectrometry. AMS is an extension of ordinary MS by including a tandem accelerator to bring ions from tens or hundreds of kilo-electron volts to mgat electron volts of energy, ensuring the efficient removal of molecular isobars and providing the energy needs for other isobar separation [10]. It is a mass spectromic method for quantifying rare isotopes. It has had a great impact in geochronology and archeology and is now being applied in biomedicine. AMS measures radioisotopes such as $^{3}H$, $^{14}C$, $^{26}Al$, $^{36}Cl$, and $^{41}Ca$ with zepto- or attomole sensitivity and high precision and throughput, allowing safe human pharmacokinetic studies involving microgram doses, agents having low bioavailability, or toxicology studies where administered doses must be kept low ($<1\,\mu g\,kg^{-1}$) [11]. It is the enabling technology for human microdosing studies.

*PET* Positron emission tomography. It is a noninvasive nuclear imaging technology that can determine drug distribution and concentration in vivo in humans that utilizes pharmacological or biologically/biochemically active compounds labeled with short-lived positron-emitting radionuclides, such as $^{11}C$, $t_{1/2} = 20.4$ min, $^{18}F$, $t_{1/2} = 109.7$ min. Recent developments in the field of rapid synthesis of organic compounds labeled with positron-emitting radionuclides have allowed a substantial number of new drug candidates to be labeled and potentially used as probes in PET studies. PET microdosing in the drug development process became a method for selection or rejection of compounds based on performance in vivo in humans [12]. Comparing to AMS microdosing, PET offers the key advantage of providing direct quantitative assessment of drug distribution and kinetics in all organs of the human body.

## 3 HUMAN MICRODOSING

### 3.1 Concept

The methods to define the ADME characteristics of a molecule include [8] in silico modeling [13], X-ray crystallography studies of recombinant cytochrome P450s (one of the body's key enzyme systems in metabolizing drugs and foreign compounds) [14], in vitro metabolism studies including the use of animals and human liver microsomes, hepatocytes, or recombinant enzymes [15–17], and in vivo studies using a range of experimental animal models [18, 19]. These studies are used to predict how humans metabolize the drug.

The derivation from animal data to humans is done using a mathematical modeling process known as allometric scaling [20], which is only 60% predictive even though promising ADME data were obtained in animals. Therefore, there is a tremendous need to eliminate this misleading extrapolation of ADME data from animals to dosing in humans. On the other hand, a number of lead candidates might come out of a drug screening program with similar pharmacological activities and identical animal ADME parameters. One has to make the choice which of these leads would make the best drug.

From the middle of the 1990s to the earlier 2000s, research and development (R&D) expenditure has increased almost exponentially year to year, but the number of new molecular entities (NMEs) being registered for marketing is either static or declining [7], even though the number of new molecular entities starting the drug developing process has increased significantly since about 1990 due to the fast progress in biomedical and pharmaceutical research. Figure 1 plots the number of NMEs as a function of time (http://www.fda.gov/cder/rdmt/). There were only 15 NMEs in 2007, the lowest number of launches for 15 years registered by the U.S. Food and Drug Administration (FDA), compared with 36 in 2004 and 53 in 1996. This is an unsustainable situation in need of a fast effective solution. The situation is bad for patients who are looking for new treatments for life-threatening diseases such as cancer, and bad for the pharmaceutical industry who will generate insufficient profits to develop the drugs of tomorrow [21].

In *the traditional paradigm for drug development*, as illustrated in Figure 2A [8], upon the identification of an interesting compound, selected from animal and/or in vitro data, an investment is made into the development of a scaled-up synthetic process for a pilot plant synthesis of up to 10 kg of the drug substance. This substance is extensively characterized and used to assess

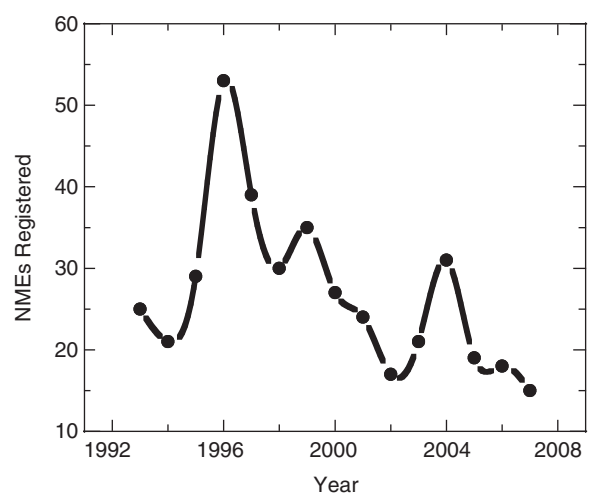


**Figure 1** NMEs registered by U.S. FDA.

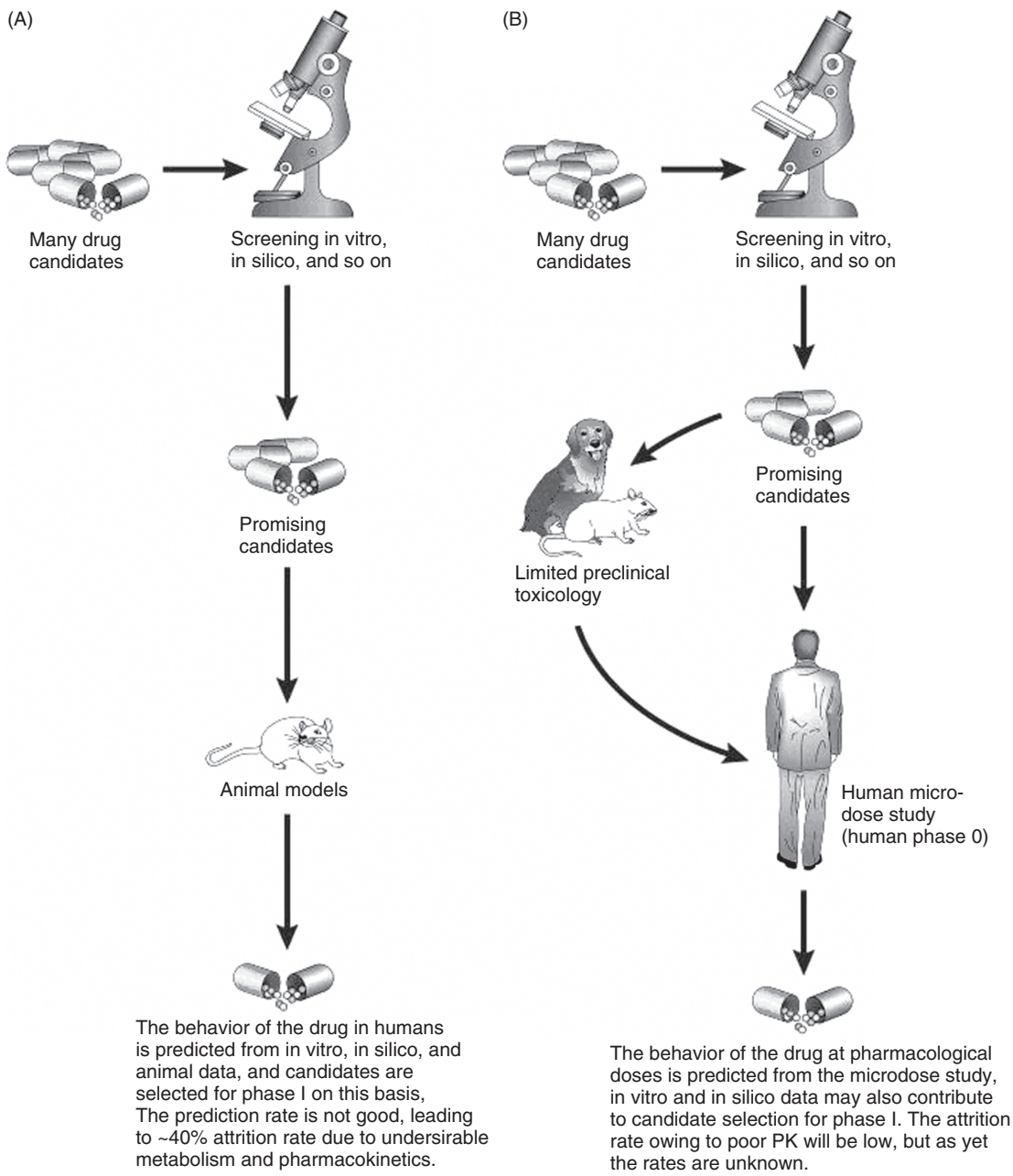


**Figure 2** Schematic of conventional and microdose approaches to drug selection [8]: (A) Path for a conventional phase I development program. The time taken to obtain human PK data for one new chemical entity can be up to 18 months. (B) Development path for a microdosing development program. Several new chemical entities can be taken through the program and human PK data obtained within 4 months.

safety issues through the conduct of 2- or 4-week toxicology studies in the rodent and nonrodent as well as gene toxicology and safety pharmacology studies. In addition, significant investment is made in the area of drug metabolism and in pharmaceutical development for the elaboration of a suitable delivery system for the compound. The time required for these activities to a first dose in humans is very compound and company dependent but typically would be 1.5–2 years.

The low success rate for drug development has been ascribed to the high number of new molecular entities with diverse mechanisms of action (MOA) to be tested for activity and efficacy combined with the lack of predictive preclinical models and inadequate

and complex clinical trial designs [22]. Overall 75% of the costs of drug development is associated with failures, mainly in the early stages of development. Often the toxicity is severe and poorly predictable. Furthermore, 40% of exits from phase I trials are caused by inappropriate pharmacokinetics of the test compound [1]. Several investigational new drugs fail in clinical testing because they do not behave pharmacologically as predicted in animal studies. Consequently, there is a need for additional methods to evaluate potential new drugs and to optimize existing development strategies.

In 2004, the U.S. FDA Critical Path Initiative [23], as well as the European Medicines Agency Road Map to 2010 [24], declared the urgent need for new toolkits to improve the critical development path that leads from scientific discovery to the patient [9]. In this scenario, *Phase 0* (zero) *trials* (or *human microdosing studies*) should allow an early evaluation in humans of pharmacokinetic and pharmacodynamic profiles of test compounds through administration of subpharmacological doses and for a short time period to a low number of humans (Fig. 2B). Typically, phase 0 studies have no therapeutic or diagnostic intent. Owing to the low doses administered and the low risk of toxicity, shorter preclinical packages to support these studies are required. Phase 0 trials have been proposed to help in making an early selection of promising candidates for further evaluation in phases I–III trials, providing a potentially useful instrument for drug discovery, particularly in the field of oncology. Phase 0 studies are expected to reduce costs of drug development, and to limit the preclinical in vitro and in vivo testing and the time period of drug development.

It is noteworthy that (a) a minimal toxicology package is required prior to a microdose study and hence only laboratory-scale quantities of drug substance are required, (b) the timelines to conduct a microdose study from commencement of the toxicology to obtaining human PK data are between 4 and 6 months in contrast to phase I study timelines of 12–18 months and (c) the costs of conducting a microdose study program are substantially less than a full phase I study program [25].

Human microdosing is a concept in line with the so-called translational science [26–29], the development of new concepts and tools that may provide more confidence in the selection of drug candidates early in clinical development. The process of discovery of new drugs and their full evaluation in the patients for whom they are designed involves at least five crucial extrapolations: (1) from physicochemical properties to biology (structure–activity relationships), (2) from in vitro to in vivo (within animal species and within humans), (3) from animals to healthy human volunteers (phase I—human safety and pharmacological proof of concept studies), (4) from single doses to multiple doses, and (5) from healthy volunteers to patients (including pathology and drug interaction considerations). Of these, the extrapolation from animals to humans is probably the most significant one. This step is now viewed as a critical component of translational science, which potentially embraces four of the five extrapolations listed above, excluding only the structure–activity relationships. Several nontraditional approaches are available to obtain an early assessment of pharmacokinetics and pharmacodynamics in first-in-human studies. Under the right circumstances, these methods may help early drug development decisions to be made more efficiently. Microdose studies are one of those approaches, but they will allow only assessment of pharmacokinetic properties [3].

### 3.2 Applications

The steps before human microdosing involve (Fig. 2B) identifying one or more drug candidates with high affinity to the target protein and demonstrating pharmacological activity in in vitro and animal models, conducting limited animal safety testing, and chemically synthesizing the appropriate isotopically labeled drug molecule [8].

In human phase 0 or microdosing studies, trace amount, subparmacological does of drugs are administrated to human subjects to obtain basic PK parameters such as clearance, volume of distribution, and $t_{1/2}$. An intravenous (IV) dose is often very useful, even for drugs that are intended purely for oral administration. (Because a drug achieves very low plasma concentrations after oral microdosing, the IV dose might show it is due to an unexpectedly high volume of distribution instead of lack of absorption.) Microdosing depends on the availability of ultrasensitive analytical methods, such as AMS and PET, that are able to measure drug and metabolite concentrations in the low pictogram to femotgram range. AMS is used for determining PK data by taking body samples over time, processing the samples in the laboratory and then analyzing their drug content. PET provides primarily PD data through real-time imaging and some limited PK data. In the latter technique PK data can be obtained for only some 2 h after drug administration (i.e., six decay half-lives), while PK data can be obtained using AMS for up to 100 days after drug administration [25]. The new path outlined in Figure 2B contrasts with the traditional approach (Fig. 2A), in that early in vivo human metabolism data make a key contribution to candidate selection. The information from microdose studies can be fed into in silico PK models to obtain a much better

estimate of the likely pharmacological dose in future efficacy studies [8].

Human microdosings, or phase 0 trials, are clinical studies conducted early in phase I, before the traditional dose escalation, safety, and tolerance studies. These first-in-human trials should involve a very limited number of normal volunteers or patients, exposed to a novel compound at a reduced dose compared to starting doses in phase I and for a short time period. Phase 0 clinical trials have no therapeutic or diagnostic purpose for the volunteer; in principle, they should allow researchers to quickly establish whether a novel compound has appropriate pharmacokinetic and pharmacodynamic profiles in humans. Phase 0 trials will not replace the traditional dose escalation, safety, and tolerance studies, and they will not indicate whether a candidate drug has a positive impact on the targeted disease. However, owing to the low doses administered, the limited number of humans treated and the reduced risk of toxicity, the phase 0 strategy would require fewer preclinical in vitro and in vivo studies than a typical phase I trial and a reduced amount of the experimental drug. Potentially, phase 0 clinical studies could help in eliminating candidate drugs before they reach phase I testing, thus reducing costs and time and improving the efficiency of drug development [22].

***Phase 0 Infrastructure*** The development and conduct of phase 0 trials requires a smoothly integrated infrastructure with a collaborative and highly motivated multidisciplinary team of professionals. Phase 0 trials provide an opportunity to establish feasibility and qualify assay methodology in limited numbers of human samples before embarking on studies that involve larger numbers of patients receiving higher, potentially toxic, doses. Phase 0 trials require the rigorous development of assays for target modulation and pharmacokinetic analysis before the initiation of the clinical trial, and real-time analysis of patient samples during the conduct of the study. This raises certain unique logistical, ethical, and statistical considerations, in addition to the intricacies of the required pharmacodynamic and pharmacokinetic studies that need to be carefully planned and managed before and during the conduct of phase 0 trials [30].

Microdosing can be used in a number of ways:

*1. Comparative Human Microdose Study* It aims to establish human PK for a number of molecules that are identified with good pharmacological activity but similar or different animal PK [25]. The obtained human PK data can be used to (a) assist in the drug candidate selection process, (b) determine the first dose for the subsequent phase I study on the selected candidate, (c) establish the likely pharmacological dose, and (d) calculate the likely cost of goods. This parallel way of conducting microdose studies is most appropriate when several drug candidates are available, perhaps with a common structural core where the radiolabel can be introduced into the core portion of the molecule. These parallel studies are best conducted on between two and five molecules using parallel human subject groups. Each molecule might be administered in a crossover design such as an intravenous dose followed after a suitable washout period with an oral dose. Thus, $V_d$ (volume of distribution) and CL (clearance) can be obtained as well as the other standard PK parameters.

*2. Iterative Microdosing Study Performed Compound by Compound* [25] Through each microdosing round, the PK and bioavailability properties of the molecule can be improved to, in the end, provide a molecule with the optimal desired PK properties. In one such study, three molecules were examined in sequence; with each microdosing round the systemic bioavailability was raised from a base value of 10% to a value of 80% in the third molecule. This iterative sequence was conducted in the space of 12 months, including the radiosynthesis, toxicology, clinical study, and bioanalysis.

*3. Microdosing with a Single Molecule* In some cases, the drug discovery process might only yield a single molecule. Microdosing can still be useful in such circumstances as it can quickly establish if it is worth taking the molecule forward prior to committing large-scale resources to a full phase I study. Sometimes a metabolic pathway is identified in human hepatocytes or liver microsomes, which is not seen in animals. Microdosing can be used to establish if the pathway occurs in vivo [25].

*4. Ensure Human Pharmacokinetics* For a *compound* in a precedent therapeutic area for which the target pharmacokinetic profile is known reasonably well, but prior data from related compounds indicate that there is too much *uncertainty in the prediction of human pharmacokinetics* [3]. This has been a rationale for a number of microdose studies at Pfizer.

*5. Applications beyond Drug Candidate Selection* Microdosing has been widely used to obtain phenotypic information on enzyme or transporter activity, for example, with the administration of small doses of erythromycin or midazolam to assess an individual's cytochrome P450 3A4 activity [31].

The current availability of advanced analytic equipment that permits detection of low concentrations of a drug in human matrices also enables microdosing to be

used in the application of previously unfeasible study designs in oncology and unanswerable questions to be investigated experimentally [32]. AMS microdosing provides a unique opportunity to revisit and reevaluate the drug interaction potential of imatinib at steady state. Microdosing should become an integrated approach to human studies aimed at *understanding* basic pharmacokinetic properties of (approved) oncology drugs that require continuous dosing. Another currently unexplored opportunity is the utility of microdosing to evaluate the contribution of germ-line variants in genes with a confirmed or suspected role in the pharmacokinetics of oncology drugs [32]. It is anticipated that in the near future microdosing will make a significant contribution in determining the relative role of genetic variation and environmental factors in the absorption and disposition of a wide variety of agents.

### 3.3 Pros and Cons of Human Microdosing

The database for microdosing studies is very small right now. It might take time to recognize its potential benefits and to be adopted. However, the adoption is accelerating, and the regulatory climate in Europe and the United States has changed, permitting microdose studies to be conducted with a minimal toxicology package. Small to medium size biotech companies are conducting microdose studies earlier than big pharma, perhaps reflecting a greater willingness to adopt innovative approaches by these scientist-led organizations [25].

The phase 0 trials were designed to improve the drug development pathway [22]. Phase 0 trials may shorten time periods to first-in-human studies through an earlier evaluation of human pharmacology of a test substance. The big advantage of phase 0 trials is the possibility to explore PK and PD profiles of candidate drugs in humans (such as bioavailability, metabolism, tissue distribution), thus potentially helping in the early selection of promising compounds for further development before traditional phase I trials. Indeed, one of the major limitations in anticancer drug development is the poor correlation between preclinical and clinical PK and PD. The early evaluation in humans of the candidate drug could potentially help in accelerating the selection of promising compounds, thus increasing the efficiency of the drug development process. For an example, the bioreductive agent EO9 was developed in phase I; however, its pharmacokinetics in plasma showed extensive and rapid degradation to inactive metabolites, making the drug unsuitable for intravenous application [33]. The time-consuming and costly phase I study could have been prevented if a phase 0 study had have been performed. In addition, inactive treatment of patients who took part in that study could have been more limited. Considering the reduced risk of human toxicity owing to the low dose of the test drug administered and the short time period of drug exposure in phase 0 trials, the preclinical safety packages, animal use, amount of test drug, development time, and costs will be reduced [22].

The potential disadvantages of phase 0 studies are:

***1. Radiolabeling Difficulties*** Radiotracer PET assay has the disadvantages of short tracer half-lives and limited specificity (assay may include metabolites). For both PET and AMS, drugs must be labeled at metabolically stable sites [25]. The use of lower doses than of traditional anticancer agents, thus potentially creating significant problems regarding the bioanalytical techniques, especially when there are few validated biomarkers for predicting anticancer activity. Pharmacological, biological, or imaging measurements at very low exposure levels are required, but unfortunately few reliable and validated assays are currently available even for most approved targeted cancer drugs. Moreover, along with novel therapeutic targets, more potent anticancer molecules have been discovered recently.

***2. Time Delay Concern*** Phase 0 trials may get drugs faster into the clinic, but there are still concerns *whether they can significantly shorten the drug development process*, as they will not replace the traditional dose escalation, safety, tolerance, and efficacy studies. The value of obtaining human data earlier, with fewer resources, must be weighed against a program delay to complete additional preclinical work and file a traditional investigational new drug (IND) application if a compound is found suitable for further development. The probability of success based on prior information from related compounds is an important factor to consider. If that probability is relatively high, a traditional phase I clinical program probably provides the most expedient path forward for that compound. However, it is also important to consider the benefits in terms of the portfolio of compounds because nontraditional approaches may cause a delay in one program but benefit the overall portfolio strategy by allowing more compounds and mechanisms to be evaluated for a given amount of resources. Finally, it is essential to be willing to make "go/no-go" decisions based on the relatively limited data from the various nontraditional approaches [3].

***3. Ethical Considerations*** Other limitations relate to the lack of therapeutic intent of phase 0 trials. The lack of therapeutic benefit is justified by the risk for the patient that is clearly less in a single-dose phase 0 trial

than in a standard dose-escalation phase I study. However, in phase I trials, although developed without therapeutic intent, there is on average a low response rate of about 5% [34], whereas in phase 0 trials *no therapeutic gain can be expected*. Therefore, besides the ethical issue, the motivation of patients to participate in a phase 0 trial may significantly affect the success rate of these trials. Similar to several phase I studies, in phase 0 trials, serial tumor biopsies may need to be obtained or other invasive procedures may be necessary. Invasive procedures can be justified by the fact that more effort should be undertaken to develop decisive pharmacodynamic assays. However, it is difficult to envisage that an end-stage patient will consider to undergo multiple hospital admissions, multiple PK samplings, and serial tumor biopsies without any potential benefit. If safe enough, the evaluation of the pharmacokinetics of a drug candidate can be conducted in healthy volunteers; however, PD measurements in tumor tissues and studies with cytotoxic agents have to be evaluated in patients.

***4. Subject Recruiting*** Participation in a phase 0 trial might delay or exclude the patient from other clinical trials that may have some, although little, therapeutic intent. Although the total duration of a patient's participation in a phase 0 trial is short and is expected to last no longer than 1 week, not only the preparation for a subsequent clinical study may be delayed but also a wash-out period may need to be taken into consideration. These issues may significantly impair patient recruitment, thus influencing the feasibility of phase 0 trials.

***5. Pharmacodynamic Considerations*** A fundamental question concerning microdosing is whether there is a correlation between metabolism at a microdose and a pharmacological dose [25]. The U.S. FDA 2006 guidance states that microdose studies should be utilized in oncology, primarily for imaging endpoints. However, as intended by the European Medicines Agency's (EMEA's) position paper, microdose studies could be used to evaluate human plasma pharmacokinetics and/or distribution properties or the receptor selectivity profile of new drug candidates [35]. In this case, an important question is whether there is a good correlation between the pharmacokinetics of a microdose and of a therapeutically relevant dose. In a study conducted by the Consortium for Resourcing and Evaluating AMS Microdosing, five drugs for which it was difficult to predict human pharmacokinetics were administered to subjects at a microdose and at a therapeutic dose level [36]. Of the five drugs evaluated, microdose measurements predicted the pharmacokinetics of three compounds well when administered at therapeutic levels and gave important metabolism data for one drug. An approximately 70% correspondence between the pharmacokinetics of a microdose and of a pharmacologically relevant dose was found. Microdosing appears to be appropriate for compounds (e.g., small organic molecules, peptide and protein therapeutics) with linear pharmacokinetics, small molecules with short half-lives or compounds that are metabolized and that have a rapid dissociation of binding to their target. In contrast, microdosing was not predictive for molecules with nonlinear pharmacokinetics and/or with high-affinity binding to their targets [7, 8, 36–38]. A microdose may not predict the behavior of clinical doses, although we have a body of evidence that for many drugs linearity or near-linearity is approached [25].

In summary, human microdosing studies are most valuable if (1) the obtainment of pharmacokinetic information is important, and (2) microdosing gives better prediction than traditional methods [39].

### 3.4 Short History of Microdorsing

***1. Radioactive Drug Research Committee*** Since 1975 the FDA has allowed basic clinical Critical Path Initiative research involving radiolabeled drugs at microdose, or subpharmacologic levels, to be conducted without an FDA investigational new drug (IND) application. Although the concept of "microdosing" has only been introduced in 2004, this program, codified in 21 Code of Federal Regulations (CFR) Part 361.1 established a process by which a medical institution can conduct such research under an FDA-approved Radioactive Drug Research Committee (RDRC). The program's intent was to encourage basic human research. Currently, there are 83 committees overseeing such research in the United States. RDRC research is considered basic science research. This research is intended to obtain basic information regarding the metabolism (including kinetics, distribution, dosimetry, and localization) of a radioactive drug. It is *not* intended for therapeutic, diagnostic, or preventive benefit to the research subject, and is *not* intended to determine the safety and effectiveness of the drug. Such types of investigations would require an IND. RDRC research must be conducted in a facility licensed for the use of radioactive materials in humans and conducted by qualified study investigators. The research protocols must be approved by the associated institutional review board (IRB). Research subjects must be properly selected, informed, and any adverse events must be reported to the FDA. Quality assurance testing of the radioactive drugs must be performed, the pharmacologic dose of the radioactive drug to be administered must be low enough to not cause any clinically detectable pharmacologic effect in

humans, and specified organ and whole-body radiation dose limits must not be exceeded.

FDA's RDRC program has been allowing microdosing study since 1975, with certain restrictions. Research is not allowed on new molecular entities (NMEs), there must be "no clinically detectable pharmacologic effect," and radiation doses to the whole body and specific organs may not exceed certain limits. Elements of new initiatives to improve human research on drug development, such as FDA's Critical Path Initiative and Microdosing, can be seen in FDA's RDRC program. This program relies on a local committee of qualified experts to provide the necessary oversight to conduct such basic human research safely.

***2. EMA Microdosing 2004 Paradigm*** A new paradigm for first-in-human studies in drug development where compounds could be investigated more efficiently for their pharmacokinetic profile emerged as the *microdose* European Medicines Agency (EMA) position paper in 2004 [39]. Toxicology testing limited to the rodent as extended single-dose studies (via the IV route and the intended clinical route *as well as* gene toxicology and safety pharmacology studies) allowed entry into humans with a laboratory batch of drug substance.

***3. U.S. FDA expIND 2006 Guidance*** The next evolution of exploratory studies in humans, at least by a health authority guidance, was the publication of the FDA exploratory investigational new drugs (expIND) guidance in January, 2006 [40], which was linked to the Critical Path Initiative [23]. The aim of the guidance, developed by the Center for Drug Evaluation and Research (CDER) (2006), was to "clarify what preclinical and clinical approaches, as well as chemistry, manufacturing and controls information, should be considered when planning exploratory studies in humans, including studies of closely related drugs or therapeutic biological products, under an investigational new drug application" [22]. This document was guidance on how the IND process could be used with more flexibility illustrated by providing three examples, implicating that the FDA would consider a unique preclinical program for exploratory INDs. The first example in the guidance addressed microdosing. The concept evolved significantly from the EMA guidance in that *only the route intended for the clinical study needed to be used* in an extended single-dose rodent toxicology program and *no* gene toxicology or safety pharmacology data were required.

The second example accurately reflected a proposal made to the FDA by the Pharmaceutical Research and Manufacturers of America (PhRMA), the U.S. pharmaceutical trade association, illustrating the study of pharmacologically relevant doses [22]. They are aimed at exploring pharmacologic effects of a test molecule. These phase 0 studies have no therapeutic or diagnostic purpose and are not meant to evaluate the maximally tolerated dose. The support of *a preclinical safety package* is more extensive than that of microdose studies, but still less than is required for traditional phase I IND studies. The paradigm allowed entry into humans based on a full 2-week good laboratory practice (GLP) compliant rodent toxicology study in both genders as well as a nonrodent study in a very limited number of animals dosed at one dose level (the no-adverse-effect level (NOAEL) of the rodent study extrapolated based on body surface area) for the intended duration of the human trial. Dosing in humans is limited with a start dose criterion (one-fifty of the rodent NOAEL study extrapolated on body surface area) as well as stop dose criteria based on the dose and exposure observed in the toxicology program—one-fourth of the rodent NOAEL dose extrapolated by body surface area, one-half the exposure [area under the blood level curve (AUC)] at that same dose or the exposure (AUC) observed in the nonrodent study. Multiple-dose studies in humans were limited to 7 days. The filing is a single administrative procedure allowing more than one compound and protocol. This paradigm has been successfully employed by the industry since the issuance of the guidance for >25 INDs up to early 2008 [9].

The third example of clinical studies of MOAs are related to efficacy. They are designed to evaluate the pharmacodynamic effect of a new drug that is expected to be correlated with its clinical activity. Typically, these phase 0 trials present biomolecular endpoints, such as the rate of an enzyme inhibition or the degree of saturation of a receptor. These clinical studies of MOAs related to efficacy have been proposed as particularly useful in the oncology field for the evaluation of molecularly targeted anticancer agents. Indeed, according to the guidance, based on careful preclinical studies and appropriate assay development, a single dose of a targeted agent can be delivered with minimal risk (since such agents have been developed for chronic administration) at a level that has the potential to have a pharmacodynamic effect in tumor tissues in patients. This is expected to speed up drug development timelines by early evaluation of target effects and the usefulness of the assay systems for such targets [22].

***4. Belgian Health Authority 2007 Guidance for Exploratory Studies in Humans*** It was posted at the Belgian health authority website in June 2007 [41]. This national guidance incorporated the US F.D.A. 2006 guidance in terms of the first two examples with some changes.

On March 13, 2006, six healthy young volunteers took part in a clinical trial in London and became violently ill minutes after having been injected with the drug TGN1412 developed to fight autoimmune disease and leukemia. This case demonstrates that even those animals claimed to be the "best" models give false reassurance regarding human safety. TGN1412 had been tested in both rats and monkeys. The drug company (TeGenero, based in Würzburg, Germany) had mistakenly assumed that tests in monkeys would predict human reactions due to the fact that the antibody reacted with both its human target protein and a similar protein in cynomolgus monkey cells. But despite the fact that the human doses were 500 times smaller than in the primate tests, a catastrophic chain reaction occurred in the human volunteers that was entirely unanticipated by the monkey tests [42, 43].

As a direct consequence of the TGN1412 episode, a new paradigm evolved for setting the first dose in humans assuring that the dose would be below the minimum anticipated biological effect level (MABEL) [9]. MABEL is to be assessed from all appropriate preclinical data extrapolated to humans. In addition to the rat NOAEL start dose criterion, the Belgian 2007 guidance includes *the MABEL principle*. The second change from the U.S. FDA 2006 guidance allowed dosing in humans up to 2 full weeks. Provisions were made to allow dose escalation in humans above the FDA stop dose criteria (translated to focus solely on exposure and not the rodent NOAEL dose). Escalation could proceed if the toxicology of the substance was of a non-life-threatening nature and was monitorable. Third, the guidance also offered a new paradigm, which had been evolving through European Federation of Pharmaceutial Industries, the European trade association. It allowed entry into humans based on GLP-compliant 2-week rodent and nonrodent studies where the top dose in each species was capped at an exposure level 10-fold higher than the exposure desired in the clinical program. This *capping of the upper dose* offered the advantage of being able to predict compound requirements accurately for the complete early development program, a significant savings to the industry. In addition, if target organ toxicity was observed in either or both species, an opportunity was offered to move directly into a classical phase I program. This paradigm has yet to be used by the industry [9].

The fourth change in the Belgian 2007 National Guidance from the U.S. FDA 2006 guidance was in the area of microdosing. It allows the microdose to be given in divided doses and also allows for 100-μg doses to be given up to five times as separate doses with an appropriate washout period between doses. This latter approach requires a 7-day GLP rodent study to support it. These microdosing changes are considered to be an important step forward for imaging studies.

This is a single administrative procedure for more than one compound and protocol and Belgium has received four exploratory clinical trial applications (expCTAs) up to early 2008 [9].

## 4 ACCELERATOR MASS SPECTROMETRY

### 4.1 Short History of AMS

***Beginnings of Accelerator Mass Spectrometry*** Although its use in 1939 helped the discovery of $^{3}$He in helim at natural abundances by Alvarez and Cornog [44, 45], AMS was normally considered to emerge from low-energy nuclear physics laboratories until the late 1970s in the unsuccessful attempts made to ascertain the $^{14}$C/$^{12}$C ratio by conventional mass spectrometry [46]. Its development was driven by a need to radiocarbon-date smaller samples than were possible with the well-known decay-counting method, which required about 1 g of carbon. The underlying principles were articulated by Muller in 1977, who proposed using a cyclotron as the accelerator and reported the successful detection of $^{3}$H [47]. Later in the same year the first AMS observations of $^{14}$C from natural materials were reported simultaneously by two groups, one working at McMaster [48] and the other at Rochester [49]. These groups used another type of accelerator, the tandem van de Graaff, which proved to have some powerful advantages over the cyclotron (mainly due to the use of negative ion sources beneficial to the rejection of interfering molecular isobars). Subsequently, AMS has been conducted almost exclusively with tandem electrostatic accelerators [50].

The technique was rapidly embraced by the nuclear physics and $^{14}$C-dating communities. This led to the development of the first dedicated AMS systems based on the Mark I Tandetron [51], and the first three were installed at Arizona, Oxford, and Toronto in 1981–82. At the same time, several nuclear physics laboratories added an AMS component to their research programs. Techniques for other isotopes, in particular $^{36}$Cl [52], $^{129}$I [53], $^{10}$Be [54], and $^{26}$Al [55], were soon developed. Over the subsequent years, many more accelerators have been given over in whole or in part to AMS and a series of second-generation purpose-built systems are now beginning to make their mark. In 1998, there are about 45 accelerators worldwide that are actively involved in AMS [50].

In parallel with the growth in the number of facilities, applications of the technique have burgeoned. Traditional $^{14}$C dating was no longer the dominant application. Carbon-14 remains the most important isotope,

but much of the research is directed toward an understanding of global climate change via studies of oceanic circulation, atmospheric processes and past climates [56]. Cosmic ray exposure dating, which exploits the buildup of $^{10}$Be, $^{26}$Al, or $^{36}$Cl in surface rocks, was making significant contributions to studies of landscape evolution. Beryllium-10 has also been used in studies of soil processes, the deposition of ocean sediments, and the sources of volcanic rocks. Both $^{36}$Cl and $^{129}$I have been used in hydrology and in tracing the migration of nuclear waste, while $^{26}$Al and $^{41}$Ca have found application in biomedicine. New applications continued to surface, often spawning new developments in technique.

***Development of Biomedical AMS*** Biomedical applications of $^{14}$C have become increasingly important. It is the most commonly employed radioisotope in toxicology and pharmacology, although similar studies are now being conducted with $^{3}$H. Both of these isotopes can be easily incorporated into most types of organic molecules that might be of interest in biological research. Additionally, $^{26}$Al, $^{41}$Ca, $^{129}$I, and $^{10}$Be are finding new applications in biomedicine [57–59].

In 1994, AMS was first used biomedically in studies to detect small amounts of $^{14}$C-labeled carcinogens when they are bound to deoxyribonucleic acid (DNA) [60]. The benefits of using AMS for the analysis of samples derived from radiotracer studies with humans soon became apparent [46]. The maximum radioactive dose that can be administered to humans depends upon the residence time in the body and whether it is accumulated in specific tissues [61]. Under certain circumstances, for example, when there is a long pharmacokinetic half-life, the amount of radioactivity that can be administered is below the capabilities of liquid scintillation counting (LSC), the most commonly used method for the measurement of radioactivity for biomedical applications. Since the advent of biomedical AMS, it is possible to conduct radiotracer studies in humans, with the administration of such low levels of $^{14}$C that they are, from a regulatory point of view, considered nonradioactive. Biomedical AMS has been used in the study of (1) metabolism of xenobiotics in animals and humans, (2) pathways of drug metabolism, (3) biomarkers, (4) metabolism of endogenous molecules including vitamins, (5) DNA and protein binding studies, and (6) clinical diagnosis [61].

***AMS Microdosing*** Miscrodosing is the most recent innovation utilizing AMS technology. As we have discussed, choosing a drug for clinical trials from numerous promising candidates is very much a hit-and-miss business. Data are gathered from in vitro, in vivo, and in silico models in order to predict the drug's behavior in humans, but such methods are probably only about 60% predictive. Presented with a choice of good candidates, it would be better to take them all in humans. This would, however, be prohibitively expensive, as each compound would require a significant package of toxicological safety testing. Alternatively, using microdosing, each candidate drug could be given to human volunteers at very low levels of a few tens or at most a hundred micrograms. At these levels, only a limited toxicology package is required and in vivo human data can be acquired for candidate selection. Only AMS has the required sensitivity to conduct such studies at the low microgram level [61].

Currently, there are more than 50 accelerators around the world (Table 1) carrying out AMS. CBAMS Ltd. (now Xceleron, York, UK) was founded at the end of 1996 and is the first private limited company in the world to offer commercial analytical services through AMS with full quality management procedures [46]. Xceleron has been a company at the fore front of using microdoing in drug R&D. The first comprehensive study in microdosing, the Consortium for Resourcing and Evaluating AMS Microdosing (CREAM) trial, was concoducted at Xceleron. On January 31, 2006, as a proof of microdosing's potential in drug development, Xceleron was awarded €2.1 million to lead the European Union Microdose AMS Partnership Programme (EUMAPP) to promote this technology as a viable experimental technique, where 7 additional drugs will be tested via this 30-month project, which gathered together 10 organizations from 5 different countries (United Kingdom, Sweden, Netherlands, France, and Poland). So far, one drug has been registered with the microdosing study at Xceleron: Tekturna, an antihypertensive drug by Novartis registed March 5, 2007. The microdosing of Speedel SPP635, a rennin inhibitor for the treatment of hypertension, was also carried out at Xceleron. In September 2007, Xceleron struck a deal with GlaxoSmithKline (GSK) that has resulted in the drug giant commissioning the pharmaceutical industry's first in-house AMS facility. In November 2007, Xceleron opened AMS operation in Germantown, Maryland. On May 1, 2008, Corcept Therapeutics Incorporated (Menlo Park, California) announced that its lead selective GR-II antagonist, CORT 108297, produced promising results in a human microdosing study. The microdosing study was conducted at Pharmaceutical Profiles Ltd. (Nottingham, UK) under Medicines and Healthcare product Regulatory Agency (MHRA) regulatory authorization and was coordinated by Xceleron Ltd.

On January 2, 2008, U.S.-based MPI Research and China's Shanghai Medicilon formed a joint venture in preclinical services, Medicilon-MPI Preclinical Research. They will establish a facility that is expected to be fully operational by 2009 providing both GLP and non-GLP

**TABLE 1 Facilities for AMS Worldwide (2002)[a]**

| Country | Accelerator | Location |
|---|---|---|
| *Europe* | | |
| Austria | 3 MV Pelletron | Vienna |
| Denmark | 6 MV EN tandem | Aarhus |
| UK | 2.5 MV Tandetron | Oxford |
| | 5 MV Pelletron | York |
| | 5 MV Pelletron[b] | East Kilbride |
| France | 2.5 MV Tandetron | Gif-sur-Yvette |
| | 3 MV Pelletron[b] | Gif-sur-Yvette |
| Germany | 6 MV EN tandem | Erlangen |
| | 3 MV Tandetron | Kiel |
| | 14 MV MP tandem | München |
| | 3 MV Tandetron | Rossendorf |
| Israel | 14 MV Pelletron | Rehovat |
| Italy | 3 MV tandem | Napels |
| Netherlands | 3 MV Tandetron | Groningen |
| | 6 MV EN tandem | Utrecht |
| Poland | 0.5 MV Pelletron | Poznan |
| Sweden | 3 MV Pelletron | Lund |
| | 5 MV Pelletron | Uppsala |
| Switzerland | 6 MV EN Pelletron | Zürich |
| | 0.5 MV Pelletron | Zürich |
| *North America* | | |
| Canada | 2.5 MV Tandetron | Toronto |
| United States | 2.5 MV Tandetron | Toronto |
| | ATLAS Linac | Argonne |
| | 2.5 MV Tandetron | Tucson |
| | 3 MV Pelletron | Tucson |
| | 9.5 MV FN tandem | Livermore |
| | RFQ | Livermore |
| | 0.5 MV Pelletron | Livermore |
| | K 1200 cyclotron | East Lansing |
| | 3 MV Pelletron | Washington D.C. |
| | 3 MV Pelletron | Denton, Texas |
| | 9 MV FN tandem | West Lafayette |
| | 2.5 MV Tandetron | Woodshole Ocean. Inst. |
| | 25 MV Pelletron | Oak Ridge |
| | 0.5 MV Pelletron | Athens, Georgia |
| | 0.5 MV Pelletron[b] | Irvine, California |
| *South America* | | |
| Argentina | 20 MV Pelletron | Buenos Aires |
| | 9 MV FN tandem | Buenos Aires |
| Brazil | 9 MV Pelletron | San Paulo |
| *Asia* | | |
| China | 14 MV MP tandem | Peking |
| | 6 MV EN tandem | |
| | 6 MV tandem | Shanghai |
| | Mini cyclotron | Shanghai |
| Japan | 3 MV Tandetron | Mutsu |
| | 5 MV Pelletron | Toki |
| | 8 MV Pelletron | Kyoto |
| | 10 MV tandem | Fukuoka |
| | 2.5 MV Tandetron | Nagoya |
| | 3 MV Tandetron | Nagoya |
| | 5 MV Pelletron | Tsukuba |
| | 12 MV Pelletron | Tsukuba |
| | 5 MV sPelletron | Tokyo |
| | 3 MV tandem | Jaeri |
| | 3 MV Pelletron | Fukushima |
| India | 3 MV Pelletron | Bhubaneswar |
| Korea | 3 MV Tandetron | Seoul |
| *Australia & New Zealand* | | |
| Australia | 8 MV FN tandem | Sydney |
| | 2.5 MV Tandetron | Sydney |
| | 14 MV Pelletron | Canberra |
| New Zealand | 6 MV EN tandem | Lower Hutt |

[a]Model EN, FN, and MP are different, relatively old, tandem accelerators produced by HVEC Burlington, USA, during the 1960s and 1970s. Their original design voltages were 6, 7.5, and 10 MV, respectively. Pelletrons are produced by National Electrostatic Corporation, Madison, WI. Tandetrons were originally produced by General Ionex Corporation. They are now produced by High Voltage Engineering Europe, Amersfoort, The Netherlands.
[b]Installation during 2002.
*Source*: From ref. 10.

preclinical services including IND enabling studies and assistance with IND submissions and new drug applications (NDAs) for the US and other regulated markets.

Besides Xceleron, AMS service is accessible for biochemical quantitation from several other private companies: Corsidus Inc., Vitalea Science, and Accium BioSciences. Noncommercial access to AMS is available at the National Institutes of Health (NIH) Research Resource at Lawrence Livermore National Laboratory. It is reported that 75% of the 20 leading pharmaceutical companies have now used microdosing in drug development [62].

### 4.2 Principles of AMS

***Elements Suitable for AMS Analysis*** For an isotope to be measurable by AMS, it should be long-lived with a half-life in the range of 10 years to 100 million years. Short-lived radioisotopes, with half-lives less than about 1 year, can usually be measured with high sensitivity by conventional techniques, in which decay products are

| | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **H** | | | | | | | | | | | | | | | | | He |
| Li | **Be** | | | | | | | | | | | B | **C** | N | O | F | Ne |
| Na | Mg | | | | | | | | | | | **Al** | **Si** | P | S | **Cl** | **Ar** |
| K | **Ca** | Sc | **Ti** | V | Cr | **Mn** | **Fe** | Co | **Ni** | Cu | Zn | Ga | Ge | As | **Se** | Br | **Kr** |
| Rb | **Sr** | Y | **Zr** | **Nb** | **Mo** | **Tc** | Ru | Rh | **Pd** | **Ag** | Cd | In | **Sn** | Sb | Te | **I** | Xe |
| **Cs** | **Ba** | | **Hf** | Ta | W | **Re** | Os | **Ir** | **Pt** | Au | **Hg** | T1 | **Pb** | **Bi** | **Po** | At | Rn |
| Fr | **Ra** | | | | | | | | | | | | | | | | |
| | | | **La** | Ce | Pr | Nd | Pm | **Sm** | **Eu** | **Gd** | **Tb** | **Dy** | **Ho** | Er | Tm | Yb | Lu |
| | | | **Ac** | **Th** | **Pa** | **U** | **Np** | **Pu** | **Am** | **Cm** | **Bk** | **Cf** | Es | Fm | Md | No | Lr |

**Figure 3** Periodic table that highlights (in bold with white background) those elements that have isotopes with half-lives in the range 10 years–100 million years, which would be suitable for AMS analysis.

counted efficiently. Primordial radioisotopes with half-lives greater than about $10^9$ years are relatively abundant naturally (since they have not completely decayed over the life of the solar system) and are used for dating by measuring the buildup of stable decay products. Radioisotopes with half-lives in the intermediate range of $10^3$–$10^8$ years are difficult to measure with decay counting since only a small fraction of the atoms decay over a reasonable counting period of a few months or less, even though many of the interesting processes that occur on Earth and in the solar system have time scales that fall in this interval. However, they are suitable for AMS. Such isotopes have short half-lives in geological terms so that they have very low natural abundance. For example, the naturally occurring isotopic abundances of the stable isotopes $^{12}C$ and $^{13}C$ are 98.9 and 1.1%, respectively, compared to $1.2 \times 10^{-10}$% for $^{14}C$. Elements that have long-lived isotopes suitable for AMS tracing are shown in Figure 3 [63]. Elements of interest in biomedicine include C, Be, Na, Al, Ca, Cs, Hg, Si, Cl, l, Pt, Ra, U, Pu, Mn, Fe, Co, Ni, Se, Mo, Sn, Tc, and Pb [46].

The relatively long half-lives of AMS isotopes also guarantee that there is negligible decay during biological in vivo studies; decay is an important issue, particularly where administration to humans is concerned [63]. Not surprisingly, the amount of radioactivity that can be administered to humans is limited owing to the radiation exposure. Regulatory agencies across the world proscribe limits of radiation exposure based on the risk of biological damage that a given radioactive dose is likely to cause in humans [61]. The biological effect is related to the amount of energy that a given organ or tissue is likely to absorb. The biological effect dose is known as the *dose equivalent* and is defined as the absorbed dose multiplied by a factor ($Q$) depending upon the type of radiation. For low-energy β emitters, such as $^{14}C$, $Q = 1$. Units of dose equivalents are sieverts (Sv), where 1 Sv equals 1 joule of kinetic energy absorbed per kilogram bodyweight. A joule is a measure of kinetic energy over time, and therefore the radioactive dose received depends upon the amount of radioactivity and its residence time in the body. There are also weighting factors applied depending upon the tissue or organ exposed to the radioactivity. For example, a larger weighting factor is used for the gonads than for the liver. Thus, a radiolabeled compound, which is eliminated rapidly from the body, without any retention in tissues or organs, can be administered using higher levels of radioactivity than a compound that is slowly eliminated and accumulates in the testes.

Experiments have to be conducted on a compound-by-compound basis to determine the maximum amount of radioactivity that can be administered to humans. These dosimetry studies are typically performed in a pigmented rat model. A pigmented rat is used to assess the extent of binding of the compound to melanin. Such binding can, for example, lead to a localized concentration of radioactivity in the eye. For a typical drug with a plasma half-life of a few hours and that does not bind to melanin, a radioactive dose of 2–3.5 MBq per volunteer might be administered.

***Count the Atoms Directly Using AMS*** AMS overcame fundamental limitations of both decay counting and conventional mass spectrometry [50]. Using $^{14}C$ as an example, there are $6 \times 10^{10}$ atoms of $^{14}C$ in a 1-g sample of modern carbon (the carbon after the era of the above-ground nuclear tests in several countries between 1955 and 1963, which dramatically increased the amount of carbon-14 in the atmosphere and subsequently in the biosphere), which decays 14 times per minute. In order to achieve 0.5% statistical precision with decay counting, it would be necessary to count for 48 h. In contrast,

typical $^{14}C$ counting rates in an AMS system are ~100 $s^{-1}$, where the same precision can be reached in 10 min, and a sample size of 1 mg is all that is required.

The half-life of $^{14}C$ is 5730 years, so it takes that long for half the $^{14}C$ atoms in a sample to decay. In any one year, therefore, only 0.012% of the $^{14}C$ atoms decay. For the rate of one disintegration in a second $4.3 \times 10^9$ atoms of $^{14}C$ must be present. It is therefore a far more sensitive method to directly detect $^{14}C$ atoms themselves, rather than detecting the radiation emitted from decaying $^{14}C$ atoms [8] since the detector is waiting the decay almost all that time.

AMS offers $10^3$–$10^9$ fold increases in sensitivity over the decay counting methods [63]. Decay counting indirectly determines the number of isotope nuclei present by measuring decay events, which is dependent on the measurement time and amount of sample analyzed. AMS, which directly quantifies each individual isotopic nucleus, is independent of these variables, therefore, can dramatically reduce analysis time and amount of sample required.

A conventional mass spectrometer (MS) can in principle count a much higher fraction of the $^{14}C$ atoms in a sample than decay counting, but in practice the $^{14}C$ ions are masked by relatively intense fluxes of ions of the $^{14}N$ isobar, by background from the stable isotopes $^{12}C$ and $^{13}C$, and also by molecular ions such as $^{13}CH$, $^{12}CH_2$, $^{12}CD$, and $^{7}Li_2$. Mass difference between isobars (atoms or molecules having the same mass number $M$) can be very small. The required mass resolution is $\Delta M/M = 84{,}000$ for $^{14}C$–$^{14}N$. In such cases isotopes can be resolved with sector MS only when very narrow slits are used, and this results in low efficiency for the transmission of ions through the instrument. *AMS obtains much better sensitivity by opening the slits wide for maximum transmission and finding other ways to separate atomic and molecular isobars.* It is possible in principle to build a mass spectrometer that could resolve these isobaric and molecular interferences, however, detected $^{12}C$ currents in such a device would be limited to 1 nA or less, and the corresponding flux of $^{14}C$ ions would be less than $10^{-2}\,s^{-1}$, that is, the rate would be comparable to decay counting. In contrast, detected $^{12}C$ currents in a typical AMS system are ~20 $\mu$A [50]. The lowest isotope ratios measurable by conventional mass spectrometers are ca. $10^{-10}$ (rare/abundant). This is about five orders of magnitude above measurements using AMS [46].

MS is an analytical method that makes use of the fact that each atomic and many molecular species have unique masses. MS typically involves four separate steps: (i) the formation of a charged atomic or molecular ion; (ii) acceleration through an electrostatic (ES) potential of a few kilovolts; (iii) separation of the ions based on their mass-to-charge ratio using some combination of the four basic analysis stages: (a) magnetic field, to distinguish mass × energy/charge$^2$, (b) electrostatic field to distinguish energy/charge, (c) velocity selector (also called a Wien or E × B filter, consists of perpendicular magnetic and electrostatic fields that provide zero net force on particles of the selected velocity, time-of-flight detectors also select ions based on their mass-to-energy ratio) to distinguish mass/energy, and (d) cyclotron resonance to distinguish mass/charge. Any two of these methods placed in series will select partciels that have unique mass-to-charge and energy-to-charge ratios; and (iv) measurement of the number of ions per second in a detector (or the electric current in a Faraday cup). AMS uses these four steps but includes an acceleration to megaelectron volt energies. *High energies* together with the analysis of *negative ions* allow the separation and measurement of isotopes at abundance ratios far lower than possible at low (kiloelectron volt) energies [64]. The high energy available is most suitable for isobar separation by ionization counter measurements [65].

The sensitivity limitations of conventional mass spectrometry are mainly the result of [46]:

1. Molecular interferences that cannot be distinguished from the wanted atom, except on the basis of differences in mass due to mass defects
2. Small-angle elastic scattering of ions of the intense mass into adjacent weak mass peak
3. Metastable decomposition and large-angle scattering from the walls or by residual gas molecules causing unwanted particles to arrive at the final detector
4. Isobaric ions of the same A, but different Z, which cannot be distinguished from the wanted ions (in combinations of electric and magnetic fields), except by the use of extremely high mass resolution
5. Particles whose kinematic properties (E/Z; $m/z$) mimic the deflection kinematics of the wanted particle at each point through the instrument unless high mass resolution is employed
6. Low probability charge-changing sequences that modify the kinematic properties of the unwanted ions at appropriate points through the instrument so they arrive in the detector volume

The very high sensitivity of AMS is obtained by an appropriate combination of energy, mass, and charge spectrometry, as the ions have at all times characteristic ratios of energy divided by charge ($E/q$) and mass divided by charge ($M/q$) that may change in a controlled manner during their passage through the spectrometer [65].

### 4.3 Procedure of Human Microdosing Study Using AMS

For a microdosing study, patients or volunteers are administered a drug that is isotopically enriched with $^{14}C$. At time intervals after drug administration, blood, urine, feces, and, very rarely biopsy samples are collected for analysis. Extreme care must be taken in preparing samples for AMS analysis to prevent any contamination and, therefore, to ensure that the isotope detected in the sample is actually associated with the labeled compound under investigation, or that the sources and amounts of any carbon introduced during processing are known. Consequently, many precautions are taken, and include sample handling in a designated laboratory and the use of disposable labwear to minimize the potential for contamination [66]. It is also important to ensure that the amount of isotope present in a sample does not exceed the dynamic range of the spectrometer because that excess could contaminate the instrument [63].

Before these biological samples can be analyzed by AMS, the carbon in the sample has to be converted to graphite. The advantages of analyzing solid samples derived from biological material include a decreased risk of cross contamination in the ion source and shorter memory effects when a sample that contains high $^{14}C$ concentrations is encountered. Furthermore, graphite can be prepared at any location and sent to an AMS facility for analysis; that process affords multiple researchers who conduct a wide variety of experiments access to an instrument at the same time.

Graphite for analysis of $^{14}C$ is obtained in a two-step process via oxidation of the crude sample [e.g., DNA, protein, or high-performance liquid chromatography (HPLC) fraction] by cupric oxide to a gaseous mixture of carbon dioxide, water, and nitrogen. The carbon dioxide is cryogenically extracted and reduced to yield filamentous graphite [67]. An improved procedure [68] involves transfer of combustion products from a sealed tube that contains the oxidized biological sample to a septa-sealed vial that contains zinc dust and catalytic iron powder. The vial is kept in liquid nitrogen to cryogenically trap the carbon dioxide and water (which act as a source of hydrogen gas), whereas noncondensing gases are removed by vacuum pump. After heating at 500°C, the resulting graphite/iron mixture is packed into a sample holder and is ready for AMS analysis. Generally, three to seven replicate measurements of a single graphite sample are taken to obtain at least a 3% precision. The measured isotope ratios are normalized against graphitized ANU (Australian National University) sucrose standards that are analyzed at the same time and have an accepted $^{14}C/^{12}C$ ratio of 1.5081 Modern [63].

AMS enables the measurements of drug concentrations in plasma and body fluids. Samples of blood and excreta can also be analyzed by chromatographic methods such as HPLC in order to separate the metabolites of the drug. The HPLC eluate is collected as a series of fractions that are then analyzed by AMS. Although AMS does not provide any direct information on chemical identity, the extent of metabolism can be assessed from the radio-chromatographic profile and, if authenticated standards are available, some metabolite comparison can be undertaken on the basis of chromatographic retention time [8].

Ideally, approximately 0.5–1 mg of total carbon is required for AMS analysis. Samples that yield this amount of carbon upon graphitization (e.g., tissue, protein) can be measured directly, whereas smaller samples (e.g., HPLC fractions, DNA) can be supplemented with a known amount of a carrier to provide sufficient carbon for optimal sample preparation. Carriers should contain depleted levels of $^{14}C$, such as the commonly used tributyrin, a petroleum-based nonvolatile liquid, 1 mL of which provides 0.6 mg carbon. In determining the accurate amount of $^{14}C$ in a sample by AMS, it is vital that the complete carbon inventory is known. This information can be obtained by conducting elemental analysis or alternatively, published reference values for a particular tissue type can be used. Results from AMS analysis are expressed in terms of "fraction modern" or "percent of modern carbon" (pMC), and can be converted to more relevant forms by subtracting the amount of $^{14}C$ contributed by the natural radioisotopic abundance of the sample plus any added carrier and taking into account the specific activity of the labeled compound [63].

### 4.4 Instrumentation of AMS

The principle of AMS analysis is to isolate the radioisotope of interest from all other interfering molecular and nuclear isobars to enable the direct quantitation of the radioisotope relative to a chosen stable isotope (isotope ratio). It achieves this by [50]:

1. Using negative ions ($C^-$, in the case of $^{14}C$, are generated in the ion source by cesium sputtering from a graphite target; target from various samples can be flexibly changed, which permits high throughput). The negative ions are separated based on mass by passing through a low-field magnet. Ions such as $^{13}CH^-$ and $^{12}CH_2$ are also present at this stage.
2. Dissociating molecular ions and stripping atomic ions to high positive charge states after a first-stage acceleration. The tandem electrostatic accel-

erator generates a high voltage (commonly 1–10 MV) to accelerate $C^-$ ions. At the positive terminal, they are stripped of several electrons, as a result of either passing through a thin molecular dissociator foil or a localized gas. Importantly, it is this process that eliminates molecular isobars because these are destroyed in the charge-changing collisions. The number of electrons removed increases with ion velocity; $^{14}C$ is most typically measured in the 3+ or 4+ charge state [63].

3. Using a second stage of acceleration followed by identification of individual ions. The positively charged ions are repelled by the positive high terminal voltage and are accelerated toward the exit and leave the accelerator (the term tandem accelerator is derived from this two-stage "pull–push" effect). They are further separated based on energy and charge through an electrostatic analyzer prior to entering a mass spectrometer that selects for the appropriate ions (e.g., $^{14}C^{4+}$, or $^{13}C^{4+}$). Individual ions are counted in a detector (e.g., gas-ionization chamber), which identifies isotopes by their characteristic pattern of energy loss, and therefore discriminates against any nuclear isobars. AMS measurements are an isotope ratio, for example, $^{14}C/^{13}C$ or $^{14}C/^{12}C$; therefore, a stable isotope of the same element ($^{13}C$ or $^{12}C$) is selected sequentially and is measured as an electrical current in an off-axis Faraday cup after acceleration and mass analysis since it is $10^{12}$ times more abundant than the rare isotope [63].

The van de Graaff tandem accelerators have some powerful advantages over the cyclotron. Tandem accelerators use negative ions in the first acceleration stage. Mass spectrometry of negative ions is a powerful method for rejecting isobars (such as $^{14}N^-$, $^{26}Mg^-$, $^{55}Mn^-$, $^{129}Xe^-$). The negative ion is not stable for some atoms and molecules (including N, Mg, $KH_3$, and noble gases), and this has made possible AMS of $^{14}C$, $^{26}A1$, $^{41}Ca$, $^{55}Fe$, and $^{129}I$ [10, 64]. Tandem accelerators have the following characteristics: (i) the ion source and detector are located at ground potential; (ii) they do not require pulsed beams; (iii) the electron-stripping step needed to eliminate molecules is an integral step in the operation of the accelerator; and (iv) transmission through the accelerator and subsequent analyzers can be made independent of small changes in the terminal voltage.

Figure 4 illustrates the essential features of an AMS system. AMS can in principle be described by the following steps [10]: negative ion production, selection of ion mass by a magnet, acceleration of the negative ions, charge exchange of the ions from negative to positive, acceleration of the now positive ions, selection of [mass × velocity/(charge)$^2$] and/or [velocity] and/or [charge/energy] in one or several analyzing units, and finally particle detection. The many steps from ion source to detector cause a strong reduction of the background. Three characteristics of AMS with tandem accelerators provide the possibility to measure low isotope ratios, for example, for $^{14}C/^{12}C$ down to $10^{-15}$.

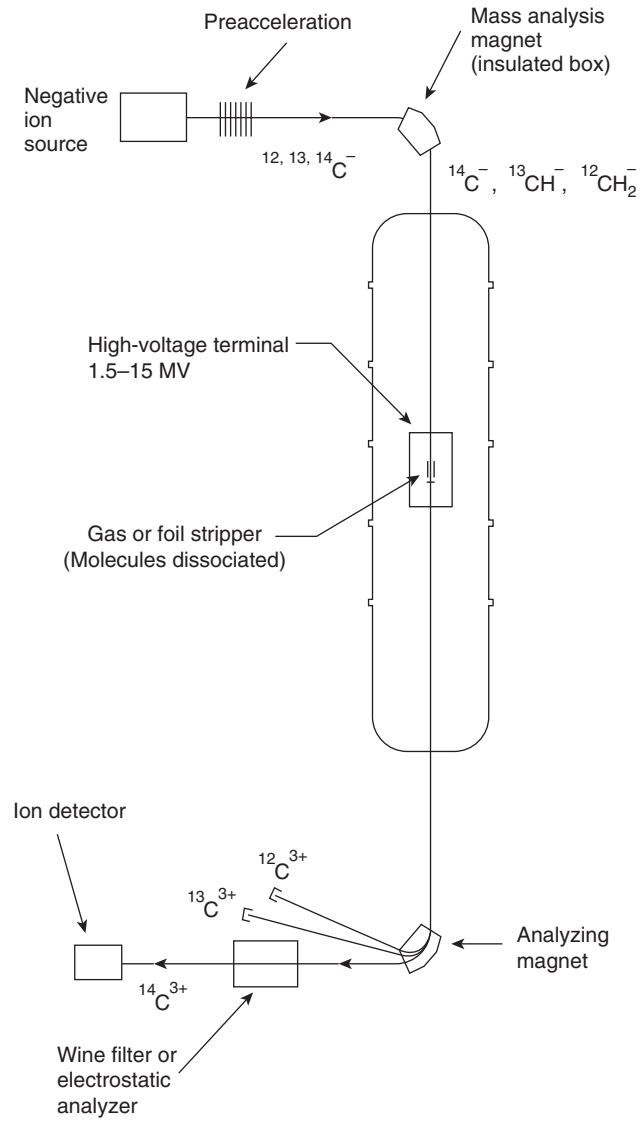


**Figure 4** Essential features of an AMS system. The first magnet is shown as transmitting mass-14 negative ions. It is also set periodically to transmit mass-12 and mass-13 ions, in which cases the accelerated ions are detected in the off-axis cups after the analyzing magnet [63].

Because biomedical studies are concerned with measuring much higher concentrations of $^{14}C$, it is possible to achieve the required degree of precision and sensitivity by measuring $^{14}C$ in the 1+ or 2+ charge states. This

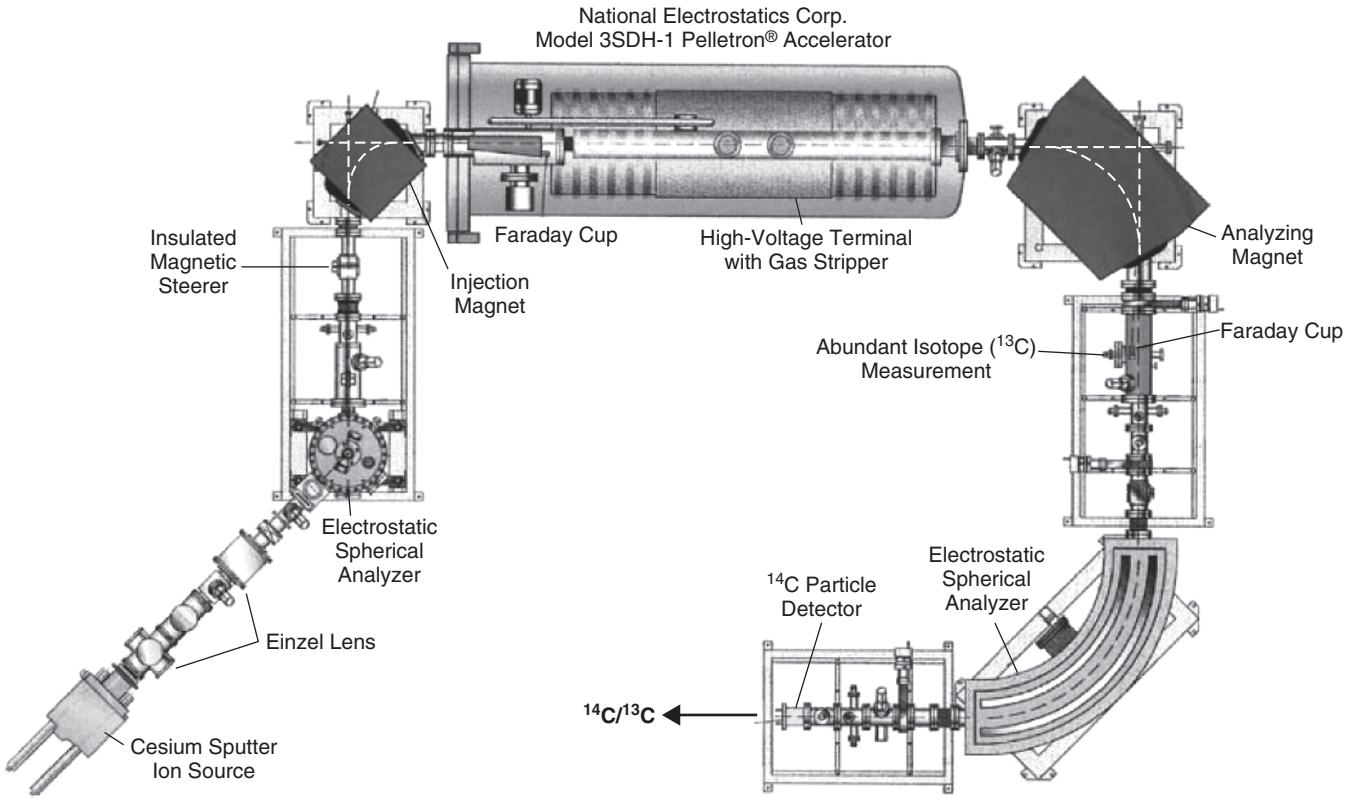


**Figure 5** Schematic layout of the National Electrostatic Corporation Model 3SDH-1 1 MV Pelletron tandem AMS instrument at the Center for Accelerator Mass Spectrometry and part of the NIH National Resource for Biomedical AMS located at Lawrence Livermore National Laboratory. This system has been in routine operation for the analysis of $^{14}C$ in biological samples since April 2001 [63].

lower charge state permits the use of a 1-MV accelerator [69–71]. Consequently, much smaller, cheaper, and lower maintenance instruments became available for the analysis of biological specimens.

Figure 5 shows the layout of the AMS system operated at Lawrence Livermore National Laboratory [11]. The carbon from the reduced sample is pounded into cylindrical target cathodes, which are mounted in a 64-sample wheel and placed inside a cesium sputter ion source. Atomic Cs vapor is produced from a heated cesium reservoir and sprayed onto a heated ionizer surface, producing $Cs^+$ ions that are accelerated toward the target held at −8 kV. The $Cs^+$ ions sputter carbon atoms and ions from the target that are ionized to $C^-$ ions as they pass through a condensed cesium layer on the cathode. Negative ions at $m/z$ 13 ($^{13}C^-$) and 14 ($^{14}C^-$) are pulsed through an injection magnet into a tandem electrostatic van der Graaff accelerator. Negative $C^-$ ions are accelerated toward the high-voltage terminal (+518 kV) in the center of the accelerator where they are converted to positive ions, $C^+$ being the most abundant, in an argon gas electron stripper. The positive ions are repelled toward the high-energy exit of the accelerator held at 0 V. $^{13}C^+$ and $^{14}C^+$ ions are then separated by momentum using a high-energy analyzing magnet (the analyzing magnet in Fig. 5). The beam currents of the abundant stable isotopes $^{12}C$ and $^{13}C$ are measured in Faraday cups. The $^{12}C$ cup is positioned after the injection magnet and before the accelerator, whereas $^{13}C$ cup is positioned between the analyzing magnet and the 90° electrostatic spherical analyzer (ESA). The $^{14}C^+$ ions are energy analyzed and focused in the ESA, detected, and counted by a silicon surface barrier particle detector using total energy deposition per ion to set the measurement gate. Molecular isobars at $m/z$ 14, such as $^{13}CH^-$ and $^{12}CH_2^-$, are the most abundant mass 14 ions and also pass through the injection magnet but dissociate in the electron stripper within the accelerator. $^{14}N^-$ ions decay on a femtosecond time scale and do not interfere with the analysis.

The normalized $^{14}C$ concentrations are easily converted into moles of $^{14}C$ per gram of carbon, becquerels

(Bq) per gram of carbon, or any other convenient unit. Since AMS was first applied to $^{14}C$ dating, the Modern unit is often used. One Modern is defined as the radiocarbon activity of a sample of 1950 wood growing in the Northern Hemisphere, 13.56 dpm per gram of carbon, which is equivalent to 6.11 fCi $mg^{-1}$ or 97.8 amol $^{14}C$ $mg^{-1}$ carbon. This is the atmospheric $^{14}C$ concentration due to cosmic radiation. Since the midnineteenth century, two anthropogenic sources have significantly changed the atmospheric $^{14}C$ levels: the burning of fossil fuels (decreasing the $^{14}C$ concentration) followed by atmospheric nuclear weapons tests, doubling the concentration of $^{14}C$ in the atmosphere by 1963. This excess $^{14}C$ continues to dissipate into the oceans with a half-life of 15 years, and the current atmosphere has a $^{14}C$ concentration of 1.04 Modern. This rapid increase and subsequent dissipation of atmospheric $^{14}C$, known as the "bomb pulse" (Fig. 6), can be used to measure age or molecular turnover rates on the time scale of years or decades in organisms that have lived after the1960s. This is of particular interest in neuroscience since turnover rates in the human brain and deposition rates of pathological structures extend to these time scales, or the lifetime of the organism. The use of $^{14}C$ to determine the age of biological samples assumes that an organism's biosynthesis is in isotopic equilibrium with its carbon source(s). Since the $^{14}C$ concentration intercepts the bomb curve at least twice, dates derived by this method are ambiguous and resolving this ambiguity requires additional information.

Figure 7 shows the 5-million-volt van der Graaff tandem accelerator mass spectrometer based at Xceleron Ltd., York, UK. Figure 8 is its floor layout.

## 5 POSITRON EMISSION TOMOGRAPHY

### 5.1 Short History of PET

***Discovery of the Positron and Its Medical Applications*** P.M. Dirac postulated the existence of positive electrons (positrons) in1932 based on the equations of quantum mechanics and Einstein's theory of relativity [72]. Anderson observed positrons in the cosmic ray in 1932 [73]. The first cyclotron was

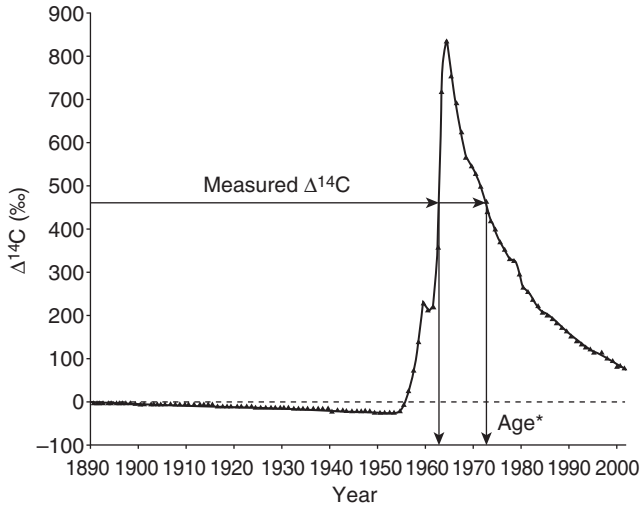


**Figure 6** Atmospheric nuclear weapons tests in the 1950s and 1960s added significant amounts of $^{14}C$ to the environment. This bomb curve can be used for dating recent biological material. Superimposed on the bomb spike is the addition of $^{14}C$-depleted $CO_2$ from burning fossil fuels [11].

**Figure 7** The 5-million-volt van der Graaff tandem accelerator mass spectrometer based at Xceleron Ltd, York, UK. (Photo source: Nature review 2003).

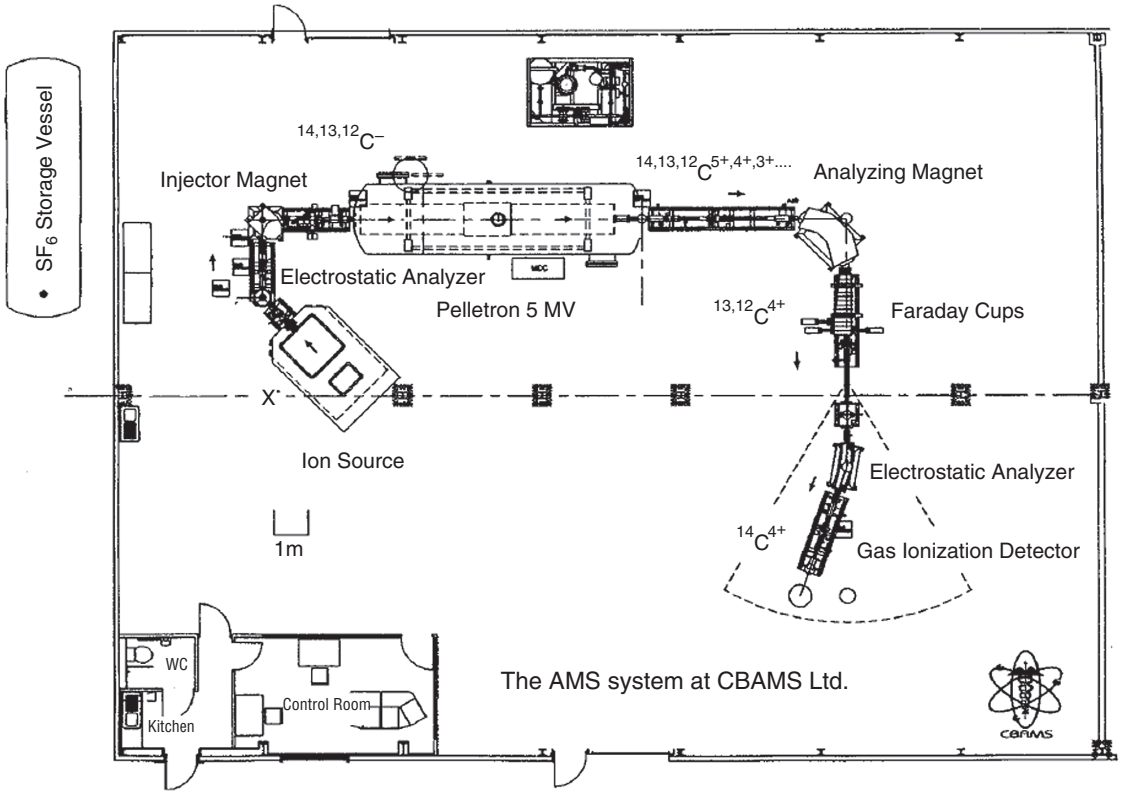


**Figure 8** AMS system at Xceleron (former CBAMS) Ltd, the first private limited company in the world to offer commercial analytical services through AMS [46].

developed by Ernest Lawrence and his team around this time in Berkeley, California: two D-shaped magnets capable of accelerating particles to produce progressively higher energy protons and deuterons that could then bombard elements to explore the nature of the atomic nucleus [74]. By 1938 Lawrence, Livingston, and others had designed a "medical cyclotron" [75]. The fusion of the theory and experimental observation led to the development of larger cyclotrons able to produce large quantities of artificial radioisotopes such as carbon-11, nitrogen-13, oxygen-15, and fuorine-18, isotopes that found many uses in medical and physiologic research. The introduction of F-18 deoxyglucose represents another major step toward practical clinical use of positron-emitting tracers. We have now achieved the transition from the postulation of the existence of positrons to their use in a wide variety of diseases [75].

***Evolution of Positron Emission Tomography*** In1951, the first medical applications for the positron were made and reported by Sweet at Massachusetts General Hospital (MGH), which utilized coincidence to localize brain tumor [76, 77]. In the same year, Wrenn et al. published studies using annihilation for localizing brain tumors [78]. In 1954, David Kuhl developed a photorecording system for radionuclide scanning [79] that has become the standard for all nuclear medicine diagnostic procedures throughout the world. In 1962, he developed radionuclide emission tomography with Roy Edwards [80], leading to the first "computer tomography," which ultimately led to the development of the X-ray computed tomography (CT) scanning as well as positron emission tomography (PET) scanning. Their work later led to the design and construction of several tomographic instruments at the University of Pennsylvania. Tomographic imaging techniques were further developed by Michel Ter-Pogossian, Michael E. Phelps, and others at the Washington University School of Medicine, where the first PET tomograph (PETT I, 1973) and first human PET tomograph (PETT III, 1974) was built. The first commercial PET scanner was designed at EG&G ORTEC in collaboration with Phelps and Hoffman as a commercial version of PETT III. The historical development of PET is marked by numerous significant technological accomplishments driven by an unprecedented

**TABLE 2 Characteristics of Scintillation Crystals Used or Developed Specifically for the Design of Current Generation PE Imaging Systems**

| Scintillator | BGO | LSO | GSO | LuAP | $LaBr_3$ | LYSO |
|---|---|---|---|---|---|---|
| Formula | $Bi_4Ge_3O_{12}$ | $Lu_2SiO_5$:Ce | $Gd_2SiO_5$:Ce | $LuA1O_3$:Ce | $LaBr_3$:Ce | $LuYSiO_5$:Ce |
| Density (g/cm$^3$) | 7.13 | 7.4 | 6.71 | 8.34 | 5.3 | 7.1 |
| Light yield (photons/keV) | 9 | 25 | 8 | 10 | 61 | 32 |
| Effective Z | 75 | 66 | 60 | 65 | 46.9 | 64 |
| Principal decay time (ns) | 300 | 42 | 60 | 18 | 35 | 48 |
| Peak wavelength (nm) | 480 | 420 | 440 | 365 | 358 | 420 |
| Index of refraction | 2.15 | 1.82 | 1.95 | 1.95 | 1.88 | 1.8 |
| Photofraction (%)[a] | 41.5 | 32.5 | 25 | 30.6 | 15 | 34.4 |
| Attenuation length (cm)[a] | 1.04 | 1.15 | 1.42 | 1.05 | 2.13 | 1.12 |
| Energy resolution (%)[a] | 12 | 9.1 | 7.9 | 11.4 | 3.3 | 7.1 |
| Hygroscopic | No | No | No | No | Yes | No |

[a]@511 keV.

*Source*: From ref. 81.

collaboration between multidisciplinary groups of investigators with backgrounds in medical sciences, physics, chemistry, mathematics, bioengineering, and computer science [77].

The scintillation detector material used in PET is the determining factor in the sensitivity. The scintillation process involves the conversion of photons (511-keV γ rays) into visible light via interaction with a scintillating material. Increased light yield, faster rise and decay times, greater stopping power and improved energy resolution, linearity of response with energy, in addition to low cost, availability, mechanical strength, moisture resistance, and machinability are the desired characteristics of scintillation crystals (Table 2) [81]. The replacement of NaI(Tl) crystals with bismuth–germanate-oxide (BGO) crystals, which is very dense, non-hygroscopic, and has a high effective atomic number, greatly improved the detection efficiency for the high-energy 511-keV γ ray. The first evaluation of BGO for use in PET was performed in the late 1970s by Cho and Farukhi [82], and Derenzo [83]. Lutetium oxyorthosilicate ($Lu_2SiO_5$, LSO) crystal has a slightly greater density, slightly lower effective atomic number, and has 5 times more light output than BGO. Also the light output of LSO has 7.5 times faster decay (a key requirement for high counting rates) than BGO. This LSO performance results in a combination speed and light output improvement of 37.5 over BGO [77]. Tomographs using LSO promise to dominate PET in the 2000 with patient throughput improvement of a factor of 5 or greater and with no increase in cost over BGO tomographs [77].

In 1978, Tatsuo Ido et al. at the Brookhaven National Laboratory first described the synthesis of $^{18}$F-tagged deoxyglucose ($^{18}$F-FDG), the most commonly used PET scanning isotope carrier [84]. The compound was first administered to two normal human volunteers by Abass Alavi in August 1976 at the University of Pennsylvania. Brain images obtained with an ordinary (non-PET) nuclear scanner demonstrated the concentration of FDG in that organ. Hamacher et al. [85] at Julich in Germany later developed a new synthesis method for FDG using a nucleophilic reaction (F-18 ion), which has become the synthesis of choice for FDG today.

In 1990s, the entrance of two major image companies (General Electric and Siemens) validate PET for clinic applications [77]. PET as the molecular imaging technology of molecular medicine may well exceed the incredible success and contributions made by CT (X-ray computed tomography) and MRI (magnetic resonance imaging) [77, 86, 87]. Figure 9 demonstrates the ability of PET observation of various biological processes [88]. The advantages of PET is that PET imaging (1) can show the extent of the disease; (2) can help in monitoring treatment and shows its effectiveness; (3) reveals disease earlier, can diagnose faster; (4) can detect whether a mass is benign of malignant; and (5) can detect abnormalities before there is an anatomical change. In contrast, CT and MRI detect changes in body structure, and simply confirm the presence of a mass. Figure 10 demonstrates that PET imaging is sensitive to the biological process of disease [88]. One survey in 2003 indicated the average cost per scan of PET in six U.S. veterans affairs hospitals and two academic medical centers was $1885 using manufactured FDG and $1898 using purchased FDG.

Since the beginnings of PET, the technique has been applied to many areas of medicine. It is especially useful in imaging neurological and psychiatric disorders, heart

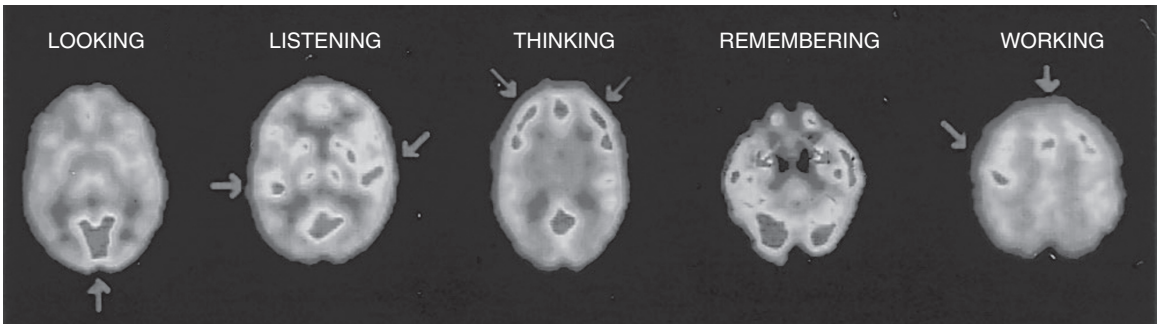


**Figure 9** PET studies of glucose metabolism to map the human brain's response in performing different tasks. Subjects looking at a visual scene activated visual cortex (arrow), listening to a mystery story with language and music activated left and right auditory cortices (arrows), counting backwards from 100 by 7's activated frontal cortex (arrows), recalling previously learned objects activated hippocampus bilaterally (arrows), and touching thumb to fingers of right hand activated left motor cortex and supplementary motor system (arrows). Images are cross sections with the front of brain at top. Highest metabolic rates are dark, with lower values in grays [88].

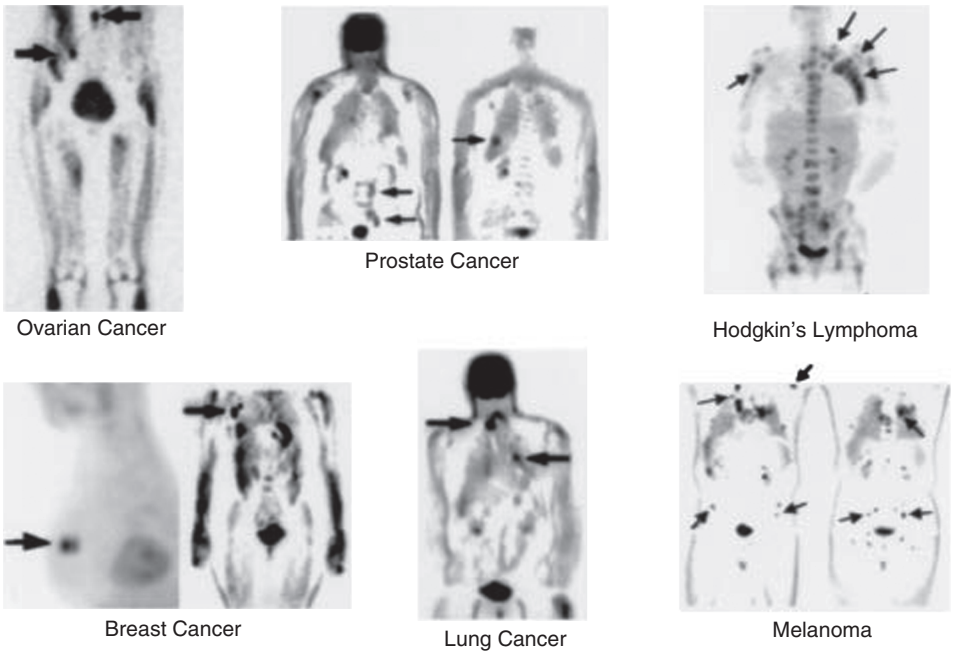


**Figure 10** PET images of glucose metabolism in various types of cancers. Study illustrates that increased glycolysis is a common property of cancer, independent of organ of origin. In breast example, a 6-mm lesion is just behind a 10-mm one. Mammogram was normal, and tumor had high expression of HER-2/neu oncogene. Arrows point to some tumors [88].

disease, and cancer [89]. Despite the obstacles in its development and practice, the reason people were drawn to PET in the first place remains: its ability to provide high-quality images of a biologic process in vivo with minimal dose to the patient. Hence PET will be the natural choice for human microdosing [12, 90–92].

### 5.2 Principles of PET

***Procedure*** The general procedure to perform a PET is (1) production of positron-emitting isotope in a cyclotron; (2) chemistry of labeling compound with positron emitter and preparing compound in a form suitable for

administration in humans; (3) transport of labeled compound from chemistry group to camera group; (4) administration (injection intravenously) of tracer compound and data acquisition with PET camera; (5) processing of data from PET camera to extract information related to the tracer's kinetics in the body: tissue concentration of the probe and labeled product over time. These data are combined with a measure of the time course of the plasma probe concentration, representing its delivery to tissue, and processed with a compartmental model containing equations describing the transport and reaction processes the probe undergoes [88]; and, finally, (6) interpretation of result. Common tissue concentrations of PET probes are in the range of pico- to femtomoles per gram. Over 500 molecular imaging probes have been developed and consist of various labeled enzyme and transporter substrates, ligands for receptor systems, hormones, antibodies, peptides, drugs (medical and illicit), and oligonucleotides [88].

***Positron–Electron Annihilation and γ-Ray Coincidence Detection*** Figure 11 diagrammatically represents the principles behind PET. It relies on the unique physics of radionuclides that decay via positron ($\beta^+$) emission. Within 1–2 mm, the positron collides with an electron present in the atoms making up surrounding tissues. On collision, the positron and electron annihilate each other, emitting two γ rays each of 511 keV energy, at 180° apart. These high-energy γ rays have an average path length of about 10 cm in tissue before they are scattered or absorbed and thus have a high probability of escaping the body. If both γ rays can be externally detected and localized in space, this forms a valid "coincidence" event and determines a line passing through the object that intersects the position at which the electron–positron annihilation took place. This, in turn, closely approximates (typically within tenths of a millimeter for low-energy positron emitters) the location at which the positron was emitted. A PET scan consists of the collection of large numbers of these coincidence events. Using mathematical reconstruction methods and appropriate correction factors for phenomena such as γ-ray attenuation and scatter, it is then possible to compute cross-sectional images through the object that quantitatively reflect the concentration of the positron-emitting radionuclide inside an object. By synthesizing compounds tagged with positron-emitting radionuclides and injecting these into the body, the regional concentration of the labeled compound (in $\mu Ci/cm^3$) can be imaged in vivo as a function of time [93]. The technique is therefore noninvasive and quantitatively measures the dynamic three-dimensional (3D) distribution of the drug in the body.

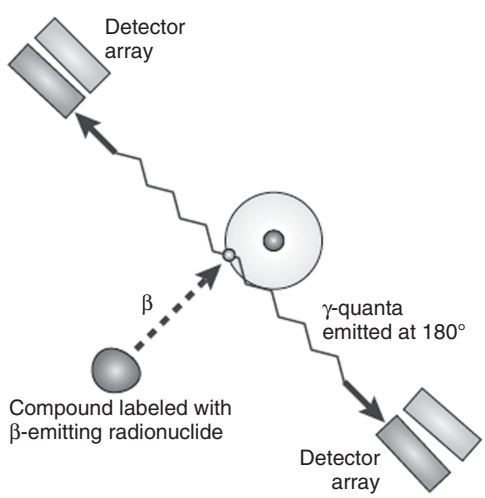


**Figure 11** Principle of positron tomography (PET) [8]. A labeled ligand emits a positron ($\beta^+$).Within a few tenths of a millimetre, the positron collides with an electron present in the atoms making up surrounding tissues. On collision, the positron and electron annihilate each other, emitting two γ rays at 180° apart. The γ rays are detected by a detector array. Opposite detectors must register the γ rays (coincidence detection) for the signal to be accepted as real.

***Positron-Emitting Radionuclides and Radiopharmaceuticals*** There are a large number of available positron-emitting radionuclides, most of which are produced by accelerators [94]. Some of the more frequently used positron-emitting radionuclides for biological applications are shown in Table 3.

Table 3 emphasizes one of the key strengths of PET, in that elements that are commonly found in molecules of biological interest, such as oxygen, nitrogen, and carbon, have positron-emitting isotopes. There is no positron emitter of hydrogen, so fluorine-18 is used as

**TABLE 3 List of Positron-Emitting Radionuclides Used in PET Applications**

| Radionuclide | Half-Life | Maximum Energy of Emitted Positrons |
|---|---|---|
| **C-11** | **20.3 min** | **0.96 MeV** |
| **N-13** | **9.97 min** | **1.19 MeV** |
| **O-15** | **122 sec** | **1.72 MeV** |
| **F-18** | **109.8 min** | **0.64 MeV** |
| Cu-62 | 9.74 min | 2.93 MeV |
| Cu-64 | 12.7 h | 0.65 MeV |
| Ga-68 | 68.1 min | 1.83 MeV |
| Br-76 | 16.1 h | 3.68 MeV |
| Rb-82 | 1.27 min | 3.3 MeV |
| I-124 | 4.17 days | 2.1 MeV |

*Source*: From ref. 93.

a hydrogen substitute. Positron emitters of Cu, Zn, K, Br, Rb, I, P, Fe, Ga, and others are also used [88]. Thus, in theory, any biological molecule can be labeled with a positron emitter by direct substitution, resulting in a labeled compound with virtually identical biochemical properties to the native compound. However, the half-lives of these radionuclides are very short, necessitating on-site production with a small biomedical cyclotron and also posing challenges in relation to extremely rapid synthesis of labeled compounds.

The four radionuclides in bold in Table 3 are produced in compact biomedical cyclotrons and are the most commonly used PET radionuclides. The positron energy determines the distance a positron will travel in tissue prior to annihilation and can limit spatial resolution of PET images in very high-resolution systems.

Fluorine-18 is a commonly used positron-emitting radionuclide, as its half-life of 110 min allows it to be produced and distributed by regional facilities (so that on-site cyclotrons are no longer a requirement), and this also gives sufficient time for more complex synthetic pathways to produce labeled compounds. Fluorine-18 can be used to substitute for H atoms, creating labeled analogs, whose properties are often similar to that of the native unlabeled compound, depending on the perturbation at the labeling site and its biological consequences. As PET becomes more widespread, other radionuclides are being investigated, particularly metals such as gallium-68 and copper-62, which can be produced in generator form and therefore do not require a nearby cyclotron [95].

Many hundreds of different compounds have been synthesized with positron-emitting labels, with carbon-11 and fluorine-18 by far the most widely used radionuclides [96]. A sample listing of the different biological processes that can be measured in vivo by PET using direct substitution and analog tracers is shown in Table 4. With appropriate labeling and understanding of the in vivo behavior of the labeled compound, just about any biological process of interest is amenable to in vivo investigation by PET [93].

*Fluorodeoxyglucose* $^{18}$FDG is used in neurology, cardiology, and oncology to study glucose metabolism. FDG is potentially useful in differentiating benign from malignant forms of stimulated osteoblastic activity because of the high metabolic activity of many types of aggressive tumors.

*Oxygen* $^{15}$O-labeled water and oxygen are being evaluated for the quantification of myocardial oxygen consumption and oxygen extraction fraction. $^{15}$O-labeled oxygen can also be used to measure tumor necrosis.

**TABLE 4 Small sample of PET Radiopharmaceuticals and the Biological Process That They Are Used to Measure**

| Desired Measurement | Examples of PET Radiopharmaceuticals |
|---|---|
| Blood volume | $^{11}$CO, C$^{15}$O |
| Blood flow | $^{15}$O-water, $^{15}$O-butanol, $^{13}$NH$_3$, $^{54}$Cu-PTSM |
| Hypoxia | $^{18}$F-misonidazole, $^{64}$Cu-ATSM |
| Substrate metabolism | $^{18}$F-fluorodeoxyglucose, $^{15}$O$_2$, $^{11}$C-acetate, $^{11}$C-palmitic acid |
| Protein synthesis | $^{11}$C-leucine, $^{11}$C-methionine, $^{11}$C-tyrosine |
| DNA synthesis (proliferation) | $^{18}$F-deoxyfluorothymidine, $^{11}$C-thymidine |
| Enzyme activity | $^{11}$C-deprenyl, $^{18}$F-deoxyuracil |
| Receptor ligands | $^{11}$C-raclopride, $^{18}$F-fluoroethylspiperone, $^{11}$C-carfentanil, $^{11}$C-WAY100635, etc. |
| Drug distribution/ kinetics | $^{11}$C-cocaine, $^{13}$N-cisplatin, $^{18}$F-fluorouracil, $^{18}$F-fluconazole |
| Receptor gene expression | $^{18}$F-fluoropenciclovir, $^{124}$I-FIAU |

*Source*: From ref. 93.

*Ammonia* $^{13}$N-labeled ammonia can be used to measure blood flow.

*Leucine* $^{11}$C-labeled methionine and leucine can be used to evaluate amino acid uptake and protein synthesis, providing an indicator of tumor viability.

*Fluorine Ion* Radiolabeled fluorine ion $^{18}$F$^-$ was once a standard agent for clinical bone scanning.

***Sensitivity of PET Assays*** A major strength of PET is that detection of radioactivity can be an extremely sensitive assay. Combined with high specific activities of the labeled compounds (typically in the range of 103–104 Ci/mmol for fluorine-18 and carbon-11), the total mass of compound that needs to be injected to create images is typically of the order of $10^{-6}$–$10^{-8}$ grams. Within the spectrum of macroscopic medical imaging, sensitivity ranges from the detection of millimolar to submillimolar concentrations of contrast media with CT and MRI, respectively, to picomolar concentrations with PET, a $10^8$–$10^9$ difference. The sensitivity of in vivo tracer studies is achieved par excellence with PET, which uses electronic collimation and thereby operates with a wide acceptance angle (particularly when operating in a fully three-dimensional mode) for detecting emitted annihilation photons. Consequently, the sensitivity of PET per disintegration, with comparable axial fields of view, is two orders of magnitude greater than

that of single-photon emission computed tomography (SPECT). Thus, PET studies in humans are, almost without exception, tracer kinetic experiments in which the underlying biological systems are not being perturbed by mass effects. This may not always be true in small-animal experiments in which mass effects need to be carefully considered under some circumstances because of the ~2500-fold difference in body weight between the human and the mouse [93].

### 5.3 Instrumentation of PET

Figure 12 illustrates historical developments highlighting the improvement in image quality and spatial resolution of clinical brain PET images resulting from the improvement of both the intrinsic performance of PET scanners and image correction and reconstruction algorithms. Both the academic community and the nuclear medicine industry maintain a continuous learning cycle and assessment of product quality to advance the technology and the operational capabilities of PET tomographs [97].

The presentation and comparison of the state-of-the-art dedicated PET systems actually available on the market in terms of physical performance and technical features can be found in refs. 98 and 99.

#### *Hardware*

*Typical PET Configuration* Modern PET scanners for clinical or human research typically consist of rings of position-sensitive scintillation detectors, with up to 25,000 individual scintillator elements of the order of 3–6 mm on a side, by 3 cm long, arranged into modular detectors (Figs. 13 and 14) [100, 101]. Detector resolution is largely determined by the width of the crystal elements. This leads to a final image resolution of ~4 mm in the best systems and an absolute detection efficiency of ~2%–3%, limited largely by solid-angle coverage constraints, which in turn are limited by cost considerations. Because of this relatively low detection efficiency and limits on radiation dose that may be given to normal human subjects, the image resolution achievable in human PET images is almost always limited by the number of detected events and the statistical noise associated with the counting statistics. To achieve satisfactory signal to noise, images are typically reconstructed at 4–8 mm resolution for brain images and perhaps 8–16 mm for whole-body images.

For very high resolution applications (detector sizes <4 mm), the physics of the positron emission and annihilation process can start to become noticeable [102]. Image resolution is then also affected by the positron range (the distance the positron travels before annihilation) and the noncollinearity of the two 511-keV $\gamma$ rays, which in practice are emitted with a small distribution of angles around 180°. The resolution in the final PET image is a convolution of these three factors (detector dimensions, positron range, noncollinearity), and for very small detector sizes, blurring due to positron range and noncollinearity can become significant and ultimately limiting. Positron range depends on the energy of the emitted positrons (see Table 3, higher energy positrons have longer ranges in tissue prior to annihilation) and varies from an average range of approximately

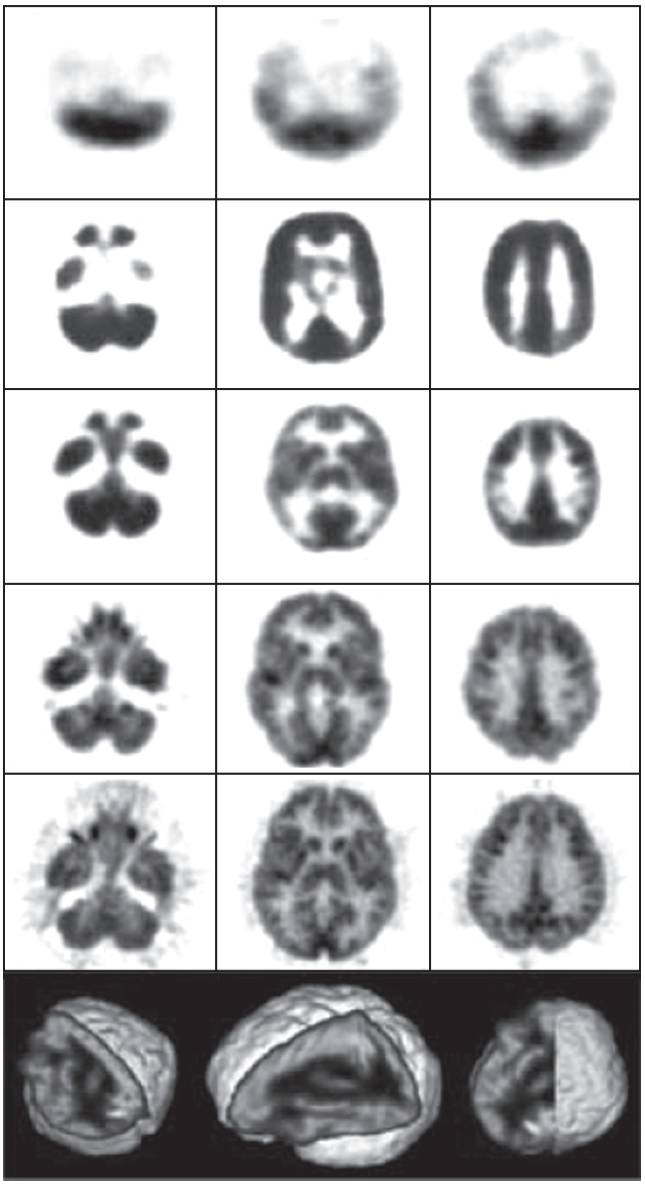

**Figure 12** Illustration of the significant improvement in clinical $^{18}$F-FDG brain PET image quality and spatial resolution resulting from the improvement in scanner performance for each generation during the last three decades [97].

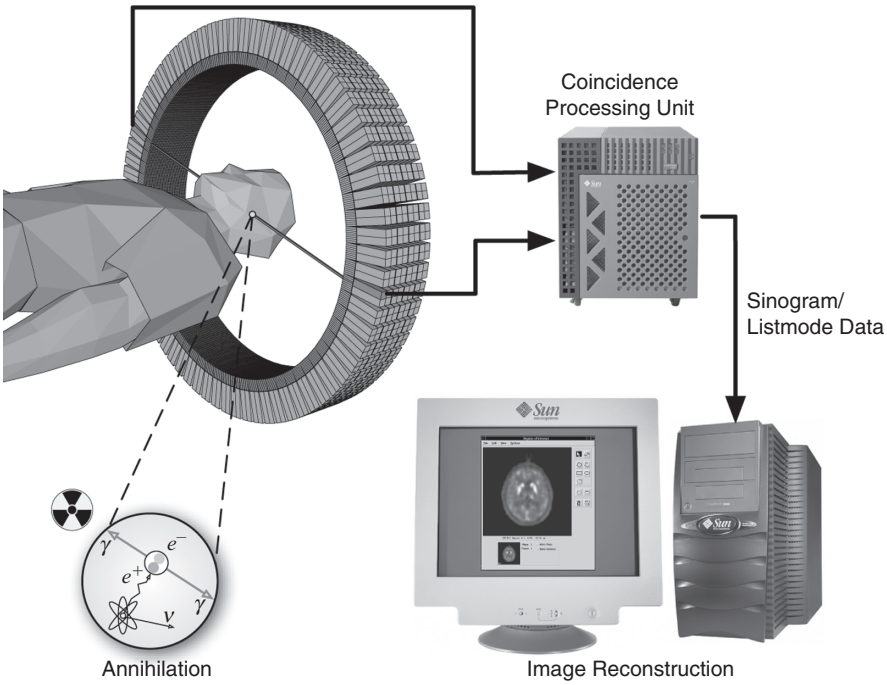


**Figure 13** Schematics of a PET acquisition process [107, 108].

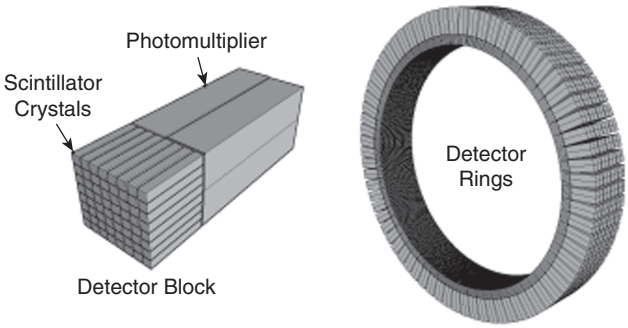


**Figure 14** Schematic view of a detector block and ring of a PET scanner [108].

0.25 mm for fluorine-18 up to 2–3 mm for high-energy emitters such as rubidium-82. The additional blurring due to noncollinearity depends linearly on the diameter of the detector ring in the PET scanner. These two factors place a limit of about 2.5 mm on the image resolution attainable in a whole-body human PET scanner with fluorine-18 labeled compounds. In small-animal PET scanners, because of the smaller ring diameter of the scanner (reducing blurring due to noncollinearity), it may be possible to reach submillimeter image resolution in vivo [93].

Phoswich (photodetector sandwich) detectors received considerable attention for the design of high-resolution scanners dedicated for brain, female breast, and small-animal imaging. It may be implemented with solid-state photodiode readouts, which also allows electronically collimated coincidence counting. Such a design has been implemented on the ECAT high-resolution research tomograph (HRRT) with LSO crystals and photomultiplier tube (PMT)-based readout [103]. Figure 15 illustrates the principle of the conventional detector block [104] and the phoswich approach [105], where two detectors are assembled in a sandwich-like design; the difference in decay time of the light is used to estimate depth in the crystal where the interaction occurred. The recently proposed axial arrangement of a matrix of long scintillator crystals readout on both extremities by hybrid photon detectors with matched segmentation is also shown [106].

In the conventional detector block concept, each block (e.g., $8 \times 8$ crystals) is readout by an arrangement of four standard PMTs optically coupled to the block by means of a light guide, which spreads the light from the individual crystals over all four PMTs. The centroid corresponds to the true interaction point of the annihilation photon, in case of conversion by photoelectric effect, or to the energy-averaged mean position of the interaction in case of occurrence of Compton scattering in the detector. The phoswich technique is able to roughly halve the uncertainty of the depth of interaction (DOI), and, hence, significantly reduces the parallax error, which is inherent to all radial geometries (such as the ring in Fig. 13). However, the phoswich approach is still only a compromise between the maximum crystal

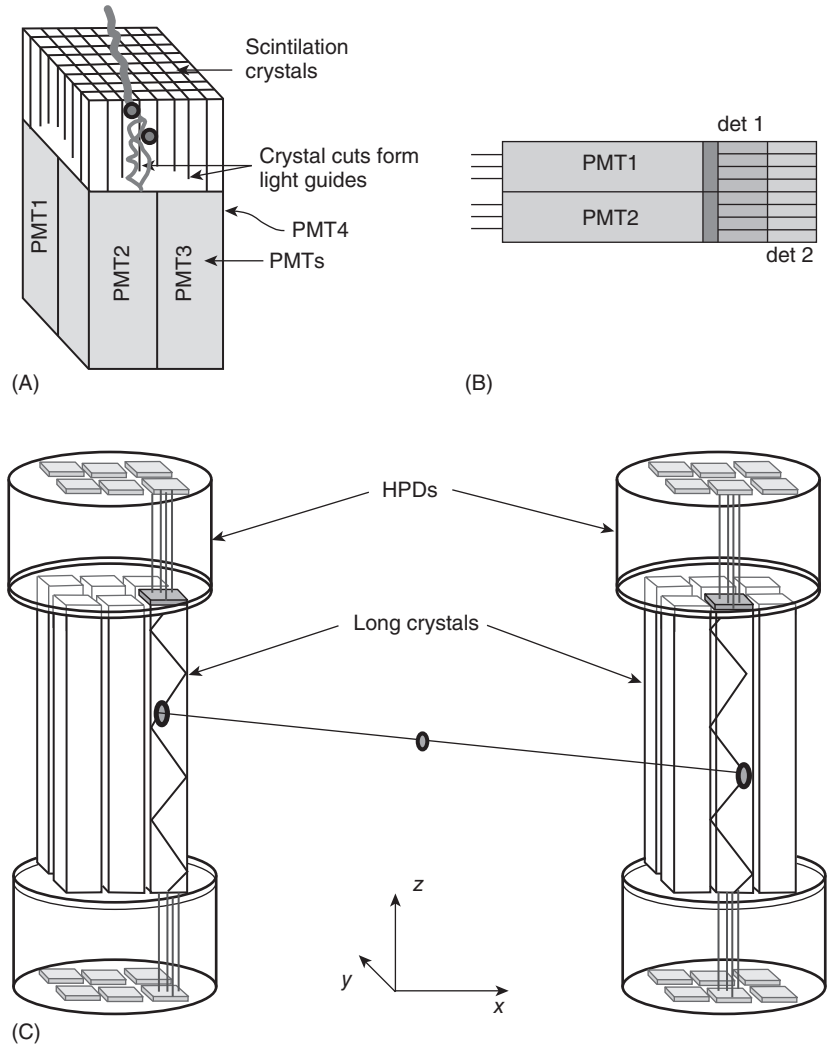


**Figure 15** Sketch of the range of different geometric arrangements of PET detector modules. (A) Conventional radial arrangement of block detectors consisting of a set of crystals having cuts of different depths acting as light guides and segmenting the block into 64 (8 × 8) detection elements in this example. The block is optically coupled to four photomultiplier tubes at the back, and the crystal in which photon absorption occurs is identified by comparing the outputs of the four photomultiplier tubes (Anger logic). (B) Detector block consisting of a phoswich (detector 1 and 2) with depth-of-interaction measurement capability. (C) Axial arrangement of a matrix of long scintillator crystals readout by hybrid photon detectors (HPDs) with matched segmentation. In this prototype configuration, two round HPD photodetectors are coupled to the matrix of scintillators [97].

length, which can be tolerated for parallax error reasons, and the minimum length, which is required to achieve high detection efficiency. In addition, the phoswhich approach demands a delicate pulse shape discrimination of the analog signal delivered by the PMTs in order to identify the hit crystal layer. The readout electronics becomes inevitably more complex and possibly limits the data acquisition rate. The axial arrangement of long finely segmented arrays of long high-Z scintillation crystals coupled to highly pixelated photodetectors leads to an essentially parallax-free geometry, which allows reconstructing the interaction point of the annihilation photon in true 3D reconstruction (Fig. 16) [97, 109].

Figure 17 shows a typical PET facility, and Figure 18 shows the maximum intensity projection (MIP) of a typical F-18 FDG whole-body PET acquisition.

*Transformable PET camera (HOTPET)* A high-resolution, flexible transformable PET camera has been developed [110] that can be configured into either a

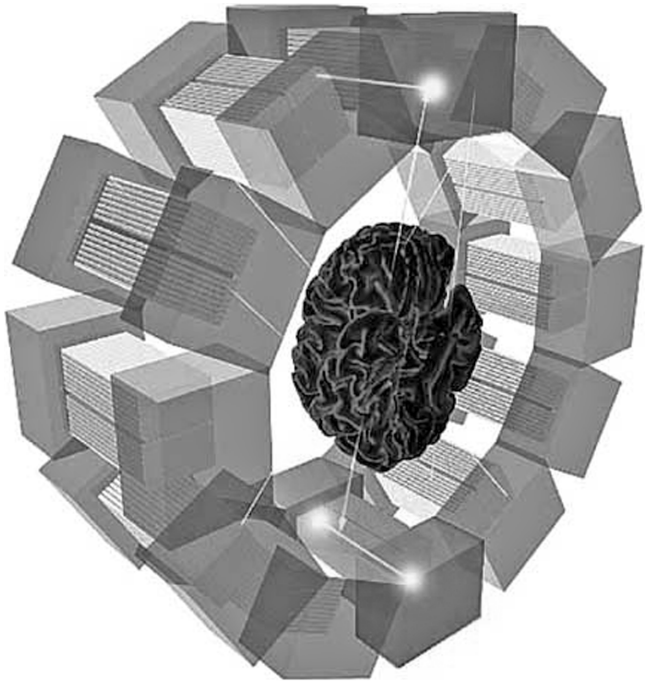

**Figure 16** Illustration of principles of data acquisition for the proposed axial 3D brain PET camera using the novel design based on 12 camera modules consisting of long scintillation crystals readout on both sides by HPDs shown in Figure 15C [106].

**Figure 18** Maximam intensity projection (MIP) of a typical F-18 FDG whole-body PET acquisition [108].

**Figure 17** Typical PET facility [108].

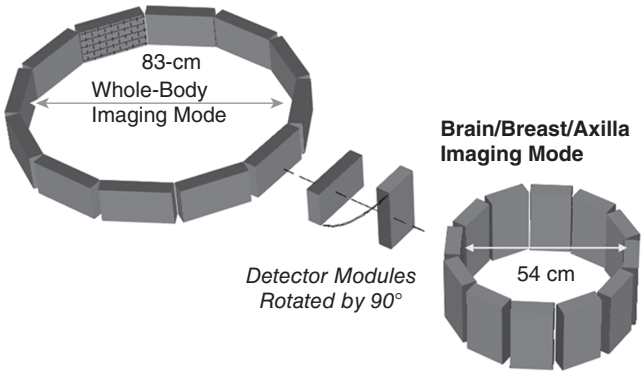


**Figure 19** Transformable detector ring diameter design with individually rotatable detector modules.

whole-body mode (83-cm detector ring diameter, 13-cm axial field of view (AFOV)), or a brain/breast/axilla (54-cm detector ring diameter, 21-cm AFOV) mode (Figs. 19 and 20). The transformation time is 1.5 h. The camera has 12 rectangular detector modules, of which each has a large detection area (13 cm × 21 cm) and three 168 bismuth BGO detector crystals of $2.7 \times 2.7 \times 18\,mm^3$ with built-in front-end electronics. To improve spatial sampling, the gantry on which the camera is mounted can be rotated 30° with a fine step size of 1°, and the high-speed electronics and detector modules also rotate along with the gantry. The system uses only 924 PMTs by applying a low-cost high-resolution PMT-quadrant-sharing detector design. To

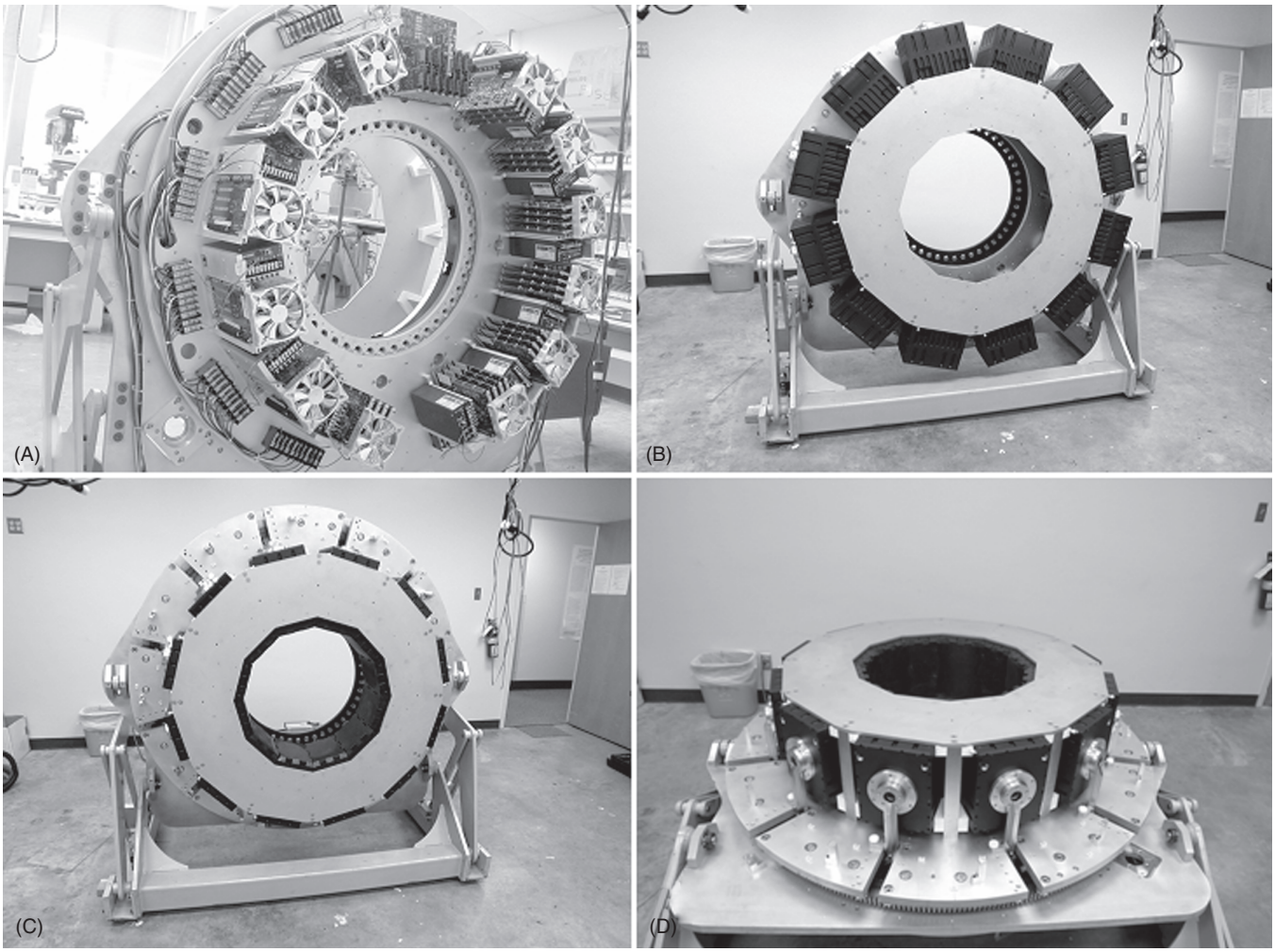


**Figure 20** HOTPET in different configurations. (A) Electronics that can rotate with the gantry, (B) whole-body mode, (C) brain imaging mode, and (D) breast image mode.

compensate for the slower scintillation timing of BGO detectors, the camera is equipped with the high-yield pile-up event recovery (HYPER) front-end electronics. A fast light-emitting diode (LED) cross-reference PMT tuning method is used to maintain the PMT gain balance. The system sensitivity is 4.2% for whole-body and 9.2% for brain/breast/axilla mode measured with Na-22 at the center of the field of view. For brain/breast/axilla modes, the transaxial image resolutions at 0 and 10 cm were found to be 2.7 and 4.0 mm, respectively, while for whole-body mode, the image resolutions were 3.3 and 3.9 mm. The resolution is higher than current commercial PET, especially for brain and breast imaging. The slice thickness resolution ranged from 2.6 to 3.3 mm for the middle seven detector rings in brain/breast/axilla modes. High-quality Hoffman brain phantom images were obtained in both brain and whole-body modes.

*Combined PET/CT and PET/MR* The introduction of dual-modality PET/CT systems in the clinical setting in the late 1990s revolutionized the practice of diagnostic imaging [111]. Combined PET/CT systems have been operational for almost a decade since their inception. The complementarity between the intrinsically aligned anatomic (CT) and functional or metabolic (PET) information provided in a "one-stop shop" and the opportunity to use virtually noise-free CT images for attenuation correction of the PET data has contributed to the success of this technology and its wide adoption by the medical imaging community [112, 113].

On the other hand, combining PET with MRI in a single gantry is technically more challenging owing to the strong magnetic fields. First-generation systems have been developed offering opportunities for integrated structural-functional image in vio [114, 115].

***Quantification and Reconstruction*** In principle, PET images can quantitatively reflect the concentration of the labeled compound in the body as a function of time [116]. To achieve absolute quantification, a number of important corrections need to be applied to the PET data. These include corrections for γ ray attenuation in the tissue, scatter in both the tissue and the detectors, accidental coincidence events (two γ rays from unrelated decays that strike the detector within the coincidence time window), detector dead time, and detector efficiency corrections. Furthermore, the raw PET data simply reflect the activity seen along the line joining each detector pair and therefore must be reconstructed into three-dimensional image volumes using either analytic filtered back-projection techniques or iterative reconstruction algorithms [117].

PET images are displayed on a computer monitor for inspection, allowing the identification of areas of the body with different degrees of uptake of the tracer. Comparisons are made between images obtained at different times after administration. Images are analyzed to obtain quantitative values on the drugs/tracers concentration–time profile in different organs or subregions of organs [12]. Therefore, regions of interest (ROI) are manually outlined in the images to represent targeted regions for analysis. Within these ROIs, the radioactivity concentration is determined for each frame in the sequence, resulting in a time–activity data set, displayed as a *time–activity curve*. The radioactivity is rapidly decaying according to the half-life of the radionuclide, so compensation should be applied for the correction. It is common in the PET field to recalculate radioactivity concentration into *standardized uptake value* (SUV), a dimensionless entity (actually milliliter/gram but with an assumed tissue density of 1 g/mL interpreted as dimensionless; the transmission scan can determine the actual density allowing a correct determination), which compensates for differences in administered radioactivity and body weight.

$$\text{SUV} = \frac{\text{local radioactivity concentration}}{\text{administered radioactivity/body weight}}$$

The time–activity data express the tracer's kinetics in a region and can in turn be used for the assessment of the drug concentration if the study is performed in such a way that tracer and drug are mixed and administered together. In this instance the relationship between cold compound and radioactivity is constant with time, if compensation is made for radioactive decay. Thus, also within each region, the relationship between radioactivity concentration and drug concentration is constant and equal to that between administered radioactivity and administered drug.

$$\text{Drug concentration} = \text{SUV} \times \left(\frac{\text{administered drug}}{\text{body weight}}\right)$$

For example, in a coinjection study, the subject of 75 kg was given 300 MBq-labeled compound mixed with 100 mg cold compound. The PET study measured an SUV in the brain of 1.2. This means the drug concentration the brain was $1.2 \times (100/75)$ mg/kg = 1.6 mg/kg = 1.6 μg/g.

***Application of PET in Drug Development*** General speaking, for drug development, PET can be used to: (1) provide in vivo biological characterization of disease (Imaging samples the entire cancer. It can quantitatively measure tumor heterogeneity in space and time. Serial assay is possible with imaging [118]). (2) Titrate drug to disease target in tissue for accurate dosing using a labeled form of the drug. (3) Provide pharmacodynamic and pharmacokinetic characterization of drugs and probes. (4) Determine whether the drug properly modified the biological process of disease or restored a normal process affected by disease [88].

### 5.4 Human Microdosing PET

PET microdosing studies are designed to determine the plasma and tissue pharmacokinetic profile of an investigational drug in vivo. Such an approach is also referred to as "pharmacokinetic imaging" [119, 120]. Direct assessment of the pharmacokinetics of a candidate drug in the target tissue, as provided by PET microdosing, has not been feasible using any other method. The concept of PET microdosing includes assessment of safety aspects, preclinical validations, and performance and analysis of human trials. In PET microdosing studies, trace amounts of a drug labeled with a short-lived positron emitter, such as carbon-11 or fluorine-18, are administered, usually by intravenous injection (oral administration is uncommon because of the need for high gastrointestinal dose rates). Dynamic PET images of various organ systems are subsequently recorded, and arterial or venous blood sampling is simultaneously performed to quantify the radiolabeled drug and its metabolites in plasma, using chromatographic methods such as high-performance liquid chromatography. In "receptor occupancy studies" (not microdosing studies, as they involve the administration of therapeutic doses of drugs that fail to fulfill the key requisite of microdosing studies), which aims to describe the interaction of a drug molecule with its pharmacodynamic target (e.g., receptor, transporter protein, or enzyme) by coadministering a validated PET tracer with different doses of the unlabeled drug under investigation. The level of

binding site occupancy induced by the unlabeled drug can then be monitored via PET imaging.

PET microdosing studies are gaining increasing importance in anticancer drug development. They are useful in understanding drug distribution in tumor tissue and can thus widen the narrow therapeutic index of cytotoxic therapy. Furthermore, the concept of radiolabeling anticancer drugs is appealing in view of the rising emphasis on personalized medicine. PET enables the measurement of pharmacokinetic parameters in individual patients in a noninvasive manner, prior to initiating full-dose treatment [90].

The radiation exposure of one PET scan is of approximately the same order as annual exposure from natural background irradiation (1–2 mSv/year). The radiation exposure of healthy volunteers and patients (5–10) in clinical studies with PET tracers is not uniformly regulated across different countries. In some countries, such as in the United States, there are clearly defined radiation dose limits, which might make whole-body dosimetry estimations mandatory in, for example, nonhuman primates, prior to initiating a phase 0 PET microdosing study. In other countries, however, no formal radiation dose limits are set for clinical research protocols, and risk-benefit evaluations are assessed on a case-by-case basis by local investigational review boards and/or competent authorities.

In contrast to AMS microdosing, PET microdosing is usually conducted subject by subject. There are two approaches to PET microdosing. First, the use of an isotope-labeled ligand, which is displaced by the drug substance under investigation. This is particularly useful for drugs designed to treat central nervous system disorders and establishes not only whether the drug reaches the brain, but whether it binds to the specific receptor it is designed to target. The rate of in vivo synthesis of dopamine, for example, can be investigated using $^{11}$C-dihydroxyphenylalanine ($^{11}$CDOPA) [121]. Therapeutic dose effects of a drug can be studied using labeled precursors in a metabolic pathway of interest. For example, the glucose analog $^{18}$FDG can be used to study effects on glucose metabolism. Similarly to glucose, $^{18}$FDG is metabolized to FDG-6-phosphate, which is then trapped in the cells. The rate of glucose uptake is, therefore, equivalent to the rate of glucose metabolism. Protein synthesis can be studied by following the uptake of $^{11}$C-tyrosine, whereas oxygen consumption and water uptake can be studied using $H_2{}^{18}O$. The selective effects of candidate drugs on metabolism in selective tissues—tumours, for example—can assist in candidate selection [93].

Alternatively, the drug itself can be labeled and its ability to reach target tissues can then be studied. Although feasible, this approach is used less frequently because of the difficulties of chemically synthesizing the radiolabeled drug [8].

What is appealing about the use of PET microdosing for assessing the pharmacokinetics of a compound is that the pattern of anatomical distribution of a radiolabeled drug might provide the first evidence of the interaction of the drug with its pharmacological target. The ability of PET to visualize drug–target site interactions is related to the fact that PET tracers are manufactured at high specific activities, implying that the amount of unlabeled carrier is low. The tissue concentration levels achieved may therefore be in the same range as the affinity of a drug (i.e., $K_i$, $K_d$) for its pharmacological target. This is of particular relevance as there is an increasing tendency to develop therapeutic drugs with high target site binding affinities (in the nanomolar or subnanomolar range). PET imaging with drugs labeled at high specific activity, therefore, has the potential to demonstrate whether a drug reaches its therapeutic target or not. It is believed that CNS (central nervous system) agents are particularly eligible for PET microdosing studies [90]. The human BBB (blood–brain barrier) represents a unique and selective filter for insufficiently lipophilic drugs [122], which preclinical models sometimes fail to predict. What PET cannot provide is chemical resolution, as the parent drug and its metabolites produce the same signal. Therefore, PET imaging should preferably be performed with compounds that are metabolically stable or whose metabolites are not taken up or retained in the tissue being investigated. The BBB impedes radiolabeled hydrophilic drug metabolites from penetrating into brain tissue and thus increases the signal-to-noise ratio of PET imaging.

***Added Value of PET Microdosing in Drug Development***

PET using trace amounts of radiolabeled drugs can answer questions that are crucial for clinical drug development such as: Does the drug reach its target organ? If so, does it attain functionally and therapeutically important concentrations? Is the residence time of the drug in its target tissue sufficient? Is there any accumulation in other organs that might affect the safety of the candidate drug? Moreover, by providing human pharmacokinetic data, even before extensive preclinical testing has been started, PET microdosing can provide guidance with respect to the selection of appropriate animal species for its detailed pharmacological and toxicological evaluation. The availability of such information at the interface of preclinical and clinical drug research helps to save time and money, as candidate drugs with unfavorable pharmacokinetic properties can be discarded before they enter into considerably more expensive phase I/II clinical trials. On the other hand,

preclinical PET could be a good method for assessing the dose linearity of a drug—an issue that is often discussed as a limitation of the microdosing concept [35, 90].

***Dose Linearity*** However, the PET method allows the recording of tissue kinetics and plasma kinetics at the same time and, therefore, allows a modeling of this relationship. The exchange of drug between tissue and plasma is usually little affected by dose and more conserved in-between species [12]. So, dose linearity means that the tissue-to-plasma relationship is unaffected by dose. In order to allow a correct extrapolation from a PET microdosing study, it is necessary to make an assessment of dose linearity. Going from a tracer-alone administration to a therapeutic dose mixed with the radiolabeled counterpart. Plasma kinetics and organ kinetics are recorded and additionally the fraction of nonmetabolized tracer in plasma is measured. With dose linearity, the time–activity data of each of the doses would be the same.

***Costs*** PET is a complicated and relatively costly technique that requires considerable infrastructure. In addition to a PET scanner, a cyclotron and radiochemistry facilities are necessary for the manufacture of the short-lived radiotracers [90]. PET studies can, if well planned in the drug development process, be very cost effective, although the price of PET may seem high. A very rough estimate of the costs of this PET microdosing package suggest an overall cost of about $300,000, subdivided into chemistry development $40,000, safety evaluation $50,000, preclinical validation $50,000, PET studies $110,000, and CRO (clinical research organization) costs $50,000 [12]. The number of PET centers, however, is continuously increasing; there are currently more than 500 facilities worldwide. The rising number of industry-owned and joint industry–academic PET centers emphasizes the increasing importance of this tool for drug development.

Table 5 presents the comparison of three analytical techniques used in human microdosing studies. An obvious advantage of PET microdosing compared with AMS is that it is able to provide invaluable information on drug distribution and kinetics in all organs of the human body, most importantly the brain [90]. AMS has, in some instances, been used to quantify drug concentration in peripheral body compartments by means of tissue biopsy. However, it is obvious that such samples cannot be collected from all organs of the human body. Moreover, repeated biopsy sampling for assessment of drug concentrations in tissue over time is not feasible for ethical reasons. On the other hand, because of the abundance of carbon, the majority of PET microdosing experiments have relied on [$^{11}$C]-labeled tracer molecules despite the rather impracticable half-life of this isotope [12]. Thus, one of the limitations of PET microdosing studies is that typical examination times for a [$^{11}$C]-labeled drug last only about 1.5–2 h. Drugs with a terminal plasma elimination half-life that is longer than the sampling duration are therefore considered to be inappropriate for plasma pharmacokinetic evaluation in a PET study [123]. For AMS, the fate of a [$^{14}$C]-labeled drug in plasma can be followed for much longer periods (several days to months). It thus becomes evident that if a drug of interest can be labeled with both a positron-emitting radionuclide, such as carbon-11, and the long-lived β-emitter carbon-14, then combined analysis by

**TABLE 5 Comparison of PET, AMS, and LC-MS/MS in Microdosing Studies**

| PET | AMS | LC-MS/MS |
|---|---|---|
| Highly sensitive ($10^{-32}$ to $10^{-18}$ g) | | Sensitive ($10^{-12}$ g) |
| Plasma and tissue pharmacokinetics | Plasma pharmacokinetics only (e.g., urine, cerebrospinal fluid, biopsy when accessible, feces) | |
| Relatively low radiation burden | Very low radiation burden | No radiation burden |
| Radiolabeled drug required | | No radiolabeled drug required |
| $\beta^+$-emitting, short-lived radioisotopes | $\beta^+$-emitting, long-lived radioisotopes | |
| No chemical separation of tissue signal | Chemical separation of drug-related material in plasma. Metabolites quantifiable without analytical standards | Chemical separation of drug-related material in plasma; structural elucidation of drug metabolites but quantification extremely difficult without analytical standards |
| Intravenous administration only | Intravenous and oral administration | |
| Limited sampling periods owing to short radioactive half-lives | Sampling period is essentially unlimited and depends only on the plasma clearance of drug | |
| Relatively high costs, limited availability | | Low costs, better availability |
| Analysis needs to be done on-site | Analysis can be outsourced | |

both PET and AMS could provide more precise tissue and plasma pharmacokinetic data than either technique alone [90].

## 6 EXAMPLES OF HUMAN MICRODOSING

### 6.1 First Microdosing Data

The data were acquired incidentally when xenobiotics with deoxyribonucleic acid (DNA) binding potential were administrated to laboratory animals and humans at environmental relevant doses (around 1990) [124]. AMS was used as the analytical technique. In general, the DNA-binding data were scalable between the doses [125], and it was also shown that the animal models were not necessarily good predictors for humans [126, 127].

### 6.2 Prediction of Pharmacokinetics of a Pharmacologic Dose from a Microdose

Will the pharmacokinetics be linear between a microdose and a therapeutic dose? It is largely because of this question that the microsdoing concept is a controversial subject. Even though that perfect linearity kinetics might not be necessary in order to make a selection for a series of drug candidates. Hopefully, with the accumulating knowledge surrounding the microdosing concept, we can use microdosing intelligently as one of the tools in the toolbox for drug candidate selection [128–131].

The first (and most up-to-date) comprehensive human study comparing microdose and therapeutic dose pharmacokinetics was completed in 2006 [35, 124]. This study is under the support of the Consortium for Resourcing and Evaluating AMS Microdosing (CREAM). The consortium comprised Xceleron, Pharma Bio-Research, Servier, Roche, Schering AG, and Eli Lilly. This study contributes to solving the paradox that many pharmaceutical companies show an interest in microdosing but are hesitant to test the concept (or publish their microdosing data since the studies are conducted at the critical stage in drug development when commercial confidentiality is paramount). The conclusion for this study is that human microdosing offers the potential to aid in early drug candidate selection. Views from the other side of the aisle stated that only for two of the five drugs was the pharmacokinetics following the microdose in agreement with that for the therapeutic dose, when both were administered by the same route. For another two drugs, even with the ultrasensitive AMS analytical method, plasma concentrations following an oral microdose were undetectable at all sampling times; and in one additional case, the terminal half-life for the microdose was more than fivefold longer than that for the therapeutic dose, which was attributed to saturable tissue binding at the higher dose. Therefore, the results highlight the limitations of microdose studies in predicting pharmacokinetics at therapeutic doses in a significant proportion of the drugs tested and thus the risk of making the wrong decision when based on microdose data [3]. Precisely due to this kind of debate, more knowledge around human microdosing study is welcome [128–131].

Each compound was selected to represent a situation in which *prediction* of human pharmacokinetics from nonclinical data *might* be considered *problematic*. There, a total of 5 compounds of which 4 are marketed drugs (warfarin, diazepam, miazolam, and erythromycin) and 1 drug (ZK253 by Schering) was rejected during development because of inadequate PK characteristics. AMS was used to quantify the low plasma concentrations anticipated after microdose administration. The volunteers include 30 subject, 24 men (18–55 years old) and 6 women (45–80 year old).

Midazolam was tested since it is a cytochrome P450 (CY)3A4 substrate that undergoes extensive first-pass metabolism that makes the pharmacokinetics difficult to predict using in vitro and animal methods. Another reason were that there were suggestions that the pharmacokinetics of high-clearance drugs, such as midazolam, are unlikely to be linear from a microdose to a therapeutic dose due to possible enzyme saturation effects [132]. The result in Figure 21 shows linearity. Results of midazolam showed predicative bioavialability due to first pass and predictive PK within a fator of 2.

Diazepam, a basic drug that is extensively distributed, also exhibited linear kinetics between a microdose and therapeutic dose (Fig. 21). It showed predictive PK.

ZK253 is interesting, which had an approximate twofold difference in half-life and clearance between microdose and therapeutic dose that resulted in the same volume of distribution for both doses, a better view of the pharmacokinetic linearity was obtained by looking at the pharmacokinetic curves (Fig. 21). Results of ZK253 showed low oral bioavailability and predictive PK within a factor of 2.

The case for warfarin is special. The plasma concentration for the microdose dropped more sharply before displaying a longer terminal phase comparing to its therapeutic dose (Fig. 21). However, estimates of clearance were similar (Table 6). Microdose data would have been useful in predicting clearance for a therapeutic dose. The situation probably is due to the fact the warfarin has a low-capacity, high-affinity binding site along with its small volume of distribution so that the biding site probably saturates at the higher dose, which has

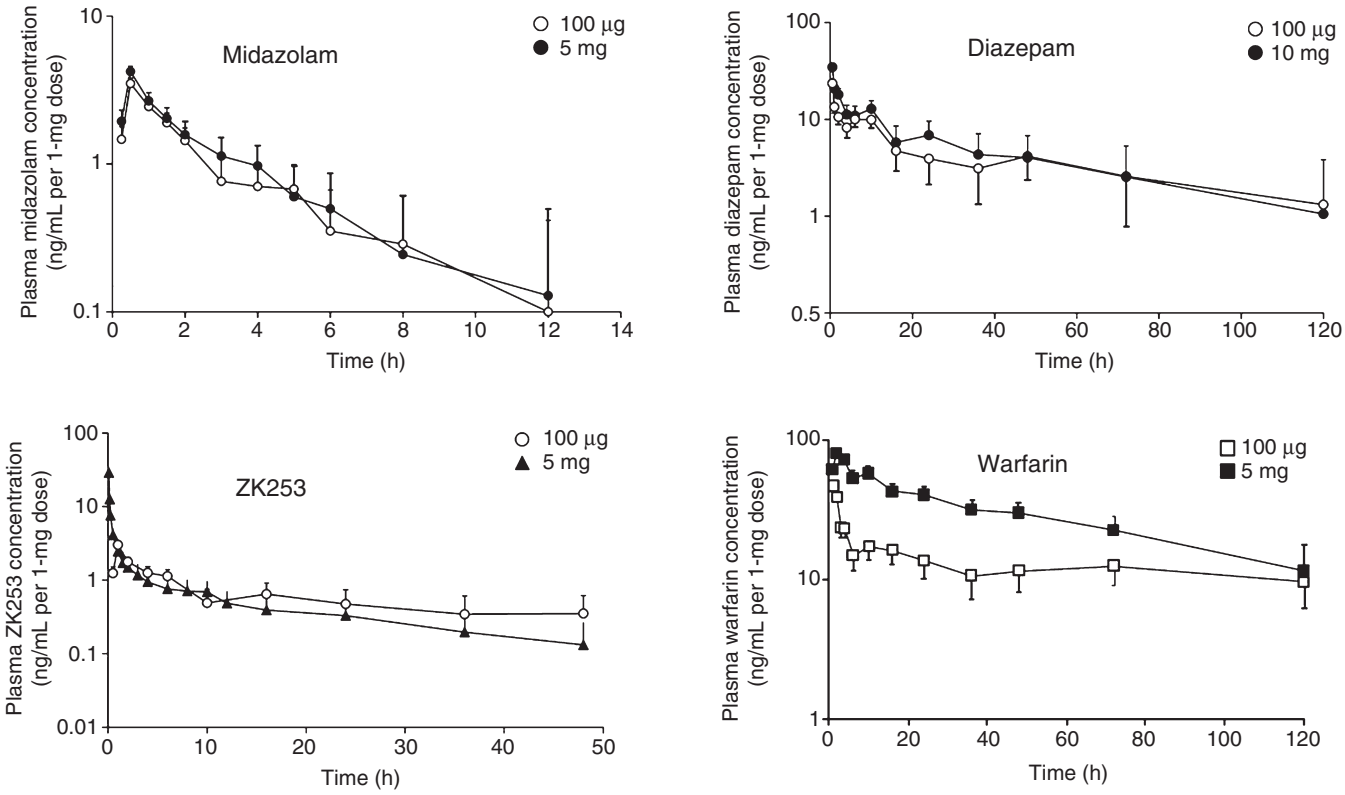


**Figure 21** Plasma concentrations after dosing. Data are normalized to a 1-mg dose. Three drugs (midazolam, diazepam, and ZK253) demonstrated dose linearity through these overlapping pharcokinetic curves, while warfarin did not due to its a low-capacity high-affinity binding site along with its small volume of distribution [124].

**TABLE 6 Summary of Pharmacokinetic Parameters Achieved for a Microdose and a Therapeutic Dose for Four Drugs**

| Drug | Dose | Route | $t_{1/2}$ (h) | CL (L/h) | V (L) | $AUC_{\infty}$[c] (ng · h/mL) | $C_{max}$ (ng) | F (%) |
|---|---|---|---|---|---|---|---|---|
| Warfarin | 100 µg | O | 274 | 0.17 | 67.3 | 571 | 5.04 | |
| | 5 mg | O | 48.6 | 0.26 | 17.9 | 20801 | 493 | |
| ZK253 | 100 µg | IV | 56.2 | 14.8 | 1207 | 7.15 | 1.52 | <1% |
| | 50 mg[a] | O | 31.1 | 29.5 | 1257 | | | <1% |
| Diazepam | 100 µg | IV | 45.1 | 1.38 | 90.1 | 65.5 | 4.70 | |
| | 10 mg | IV | 35.7 | 1.30 | 123 | 5577 | 322 | |
| Midazolam | 100 µg | IV | 4.87 | 20.4 | 75.1 | 4.68 | 2.97 | 22.1 |
| | 7.5 mg | O | 3.31 | 27.7[b] | 77[b] | 87.0 | 34.0 | 22.8 |

[a]From previous phasel I study (intravenous and oral routes of administration).
[b]From literature [25, 26].
[c]$AUC_{\infty}$ = area under the concentration–time curve to infinity; CI = clearance; $C_{max}$ = maximum drug concentration; $F$ = bioavailability; IV = intravenous routes of administration; O = oral; $t_{1/2}$ = half-life; V = volume of distribution.

*Source*: From ref. 124.

been observed for wafarin in rat [133]. Therefore, some caution would be justified for drugs similar to warfarin if using microdosing for candidate selection. Therefore, the results of warfarin showed predictive clearance (CL) but not predictive distribution from microdosing study. The lesson learned from this study is to be cautious of drugs that exhibit high-affinity binding to low-capacity biding site. Indications are that there are no specific reasons why drugs with high protein binding will not be predictive.

The study also included the P-glycoprotein substrate erythromycin (an *acid-labile drug*). Due to gastric lability of the labeled drug, nomeaningful comparison could be made between microdose and the therapeutic dose [36]. The study highlights the need to evaluate *drug stability* in gastric and intestinal fluids when one is con-

templating performing oral microdosing studies. Acid-labile drugs dosed orally have to be protected from stomach acid. Indications are that there are no specific reasons why P-glycoprotein (Pg-p) substrates will not be predictive.

### 6.3 PET Microdosing

A PET microdosing study [92] has been carried out with an investigational, carbon-11-labeled antiamyloid drug, 1,1-methylene-di-(2-naphthol) (ST1859) with three healthy volunteers and six patients with Alzheimer's disease. The study aimed to assess the distribution and local tissue pharmacokinetics of the study drug in its target organ, the human brain. Before PET studies were performed in humans, the toxicologic characteristics of ST1859 were investigated by an extended single-dose toxicity study according to guidelines of the FDA and European Medicines Agency, which are relevant for clinical trials with a single microdose.

After intravenous bolus injection of 341 ± 21 MBq [$^{11}$C]ST1859 (containing <11.4 nmol of unlabeled ST1859), peripheral metabolism was rapid, with less than 20% of total plasma radioactivity being in the form of unchanged parent drug at 10 min after administration. In both the control and patient groups, uptake of radioactivity into the brain was relatively fast (time to reach maximum concentration, 9–17 min) and pronounced [maximum concentration (standardized uptake value), 1.3–2.2] (Table 7, Fig. 22). In both healthy volunteers and patients, there was a rather uniform distribution of radioactivity in the brain (Fig. 23), including both amyloid-β–rich and amyloid-β–poor regions, with slow washout of radioactivity (half-life, 82–185 min).

In summary, these data provide important information on the blood–brain barrier penetration and metabolism of an investigational antiamyloid drug and suggest that the PET microdosing approach is a useful method to describe the target organ pharmacokinetics of radiolabeled drugs in humans.

### 6.4 List of Human Microdosing Studies in Order of Published Date

***1975*** FDA Radioactive Drug Research Committee allowed "microdosing" study with certain restriction.

***1990*** First microdosing data (AMS studies of environmental carcinogens $^{14}$C-MelQx by binding to DNA) [125].

***1996*** The antidepressant, BMS-181101, one of a select number of investigational CNS pharmaceuticals, was labeled with carbon-11 and administered to healthy volunteers at a microdose level to determine drug distribution in brain tissue [134]. Further clinical development of BMS-181101 was abandoned because PET imaging, using both high and low specific activity radiotracers, failed to reveal saturable target-site-specific binding of this investigational drug [135]. The PET microdosing approach selected in the case of BMS-181101 was judicious, as insufficient or fluctuating tissue concentrations of the investigational drug might considerably decrease occupancy of binding sites and affect therapy outcome [90].

***2001*** In an anticancer microdosing study [136], PET imaging was used to determine tumor tissue distribution of the [$^{11}$C]-labeled topoisomerase I/II inhibitor *N*-[(dimethylamino)ethyl]acridine-4-carboxamide, at one-thousandth of the phase I clinical trial starting dose. This study formed part of a series of studies examining novel anticancer agents designed around the use of PET in early clinical drug development [137]. In a more recent study using the same radiotracer, it was demonstrated

**TABLE 7 Pharmacokinetic Parameters of [$^{11}$C]ST1859 in Venous Plasma and Brain[a,b]**

| | Venous Plasma | | Isocortex | | Basal Frontal Lobe | | Pons | |
|---|---|---|---|---|---|---|---|---|
| Parameter | Control Subjects | AD Patients | Control Subjects | AD Patients | Control Subjects | AD Patients | Control Subjects | AD Patients |
| $C_{max}$ (SUV)[a] | 4.7 ± 1.6 | 8.6 ± 3.2 | 1.3 ± 0.01 | 1.6 ± 0.2 | 1.3 ± 0.2 | 1.8 ± 0.4 | 1.6 ± 0.1 | 2.2 ± 0.4 |
| $I_{max}$ (min) | 1.0 ± 0.0 | 1.0 ± 0.0 | 14.2 ± 3.1 | 11.8 ± 2.2 | 16.5 ± 1.7 | 12.0 ± 2.1 | 13.5 ± 1.7 | 10.7 ± 1.8 |
| $t_{1/2}$ (min) | 12.6 ± 8.6 | 10.3 ± 9.1 | 112.8 ± 6.5 | 82.4 ± 13.5 | 139.9 ± 49.0 | 103.8 ± 23.8 | 120.2 ± 23.8 | 105.2 ± 18.7 |
| $AUC_{O-1}$ (min)[d] | 23.8 ± 3.3 | 35.8 ± 9.6 | 96.9 ± 2.4 | 107.5 ± 17.7 | 99.2 ± 6.4 | 125.1 ± 26.0 | 103.8 ± 7.5 | 148.5 ± 32.7 |

[a]Data represent mean ± SD pharmacokinetic parameters of radioactivity in venous plasma (metabolite-corrected) and three selected brain regions after intravenous injection of [$^{11}$C]ST1859 into control subjects ($n$ = 3) and AD patients ($n$ = 6).
[b]ST1859. 1.1′-Methylene-di-(2-naphthol): AD, Alzheimer's disease; $C_{max}$, maximum concentration; SUV, standardized uptake value; $t_{max}$, time to reach maximum concentration; $t_{1/2}$, elimination half-life; $AUC_{O-1}$, area under time–activity curve from time 0 to last observation point.
[c]Radinactivity concentrations (in kilobecquerels per gram of tissue) were normalized to the injected radiotracer amount per body weight and expressed as dimensionless SUV.
[d]The unit for AUC (minutes) is different from conventional units because the concentration is expressed as dimensionless SUV.

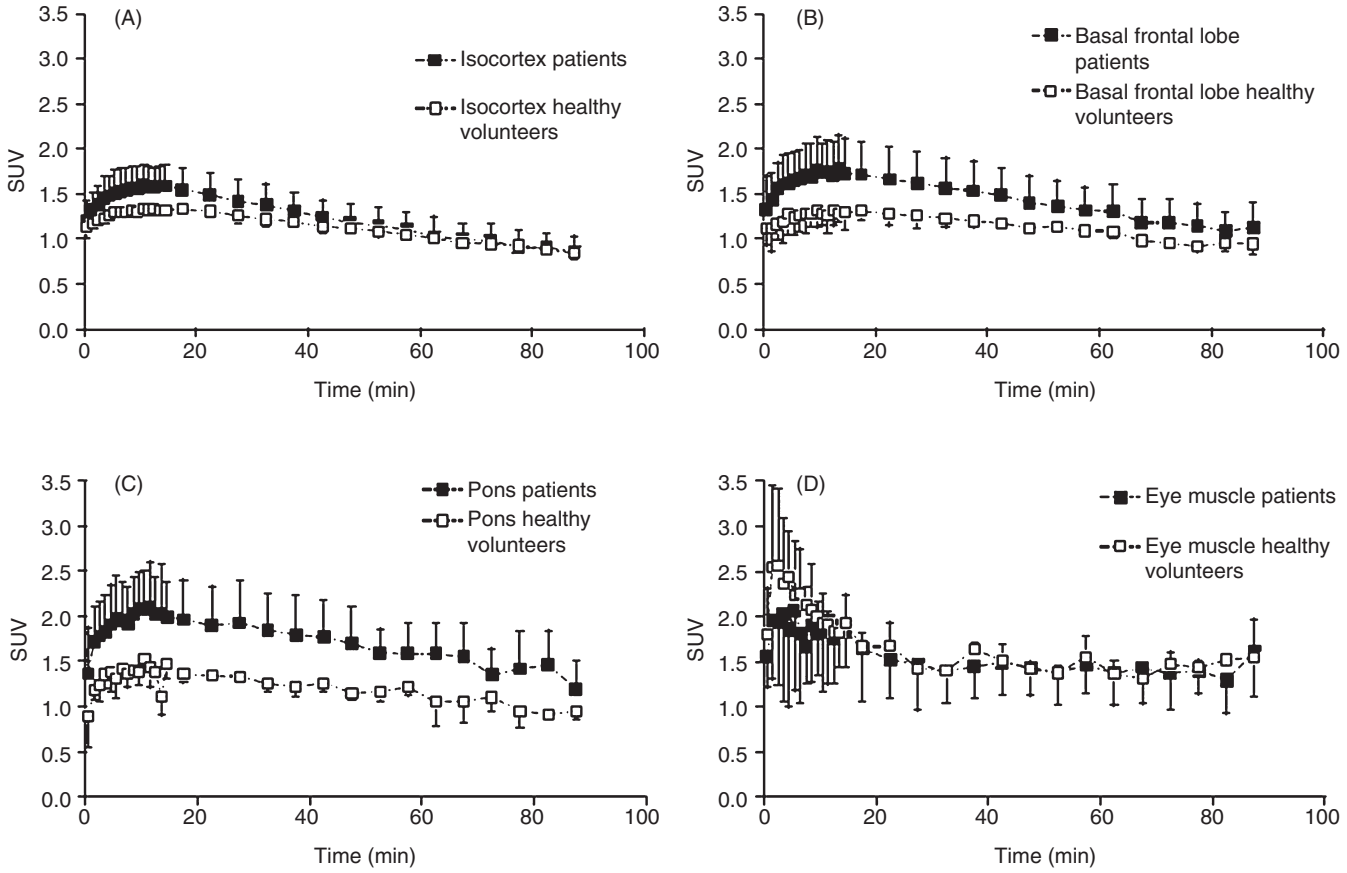


**Figure 22** Time–activity curves measured in three different brain regions of interest [isocortex (A), basal frontal lobe (B), and pons (C)] and one extracerebral region of interest [left lateral eye muscle (D)] after intravenous administration of [$^{11}$C]ST1859 to three healthy volunteers and six patients.

that tracer studies performed with plasma levels that were severalfold lower than the Michaelis–Menten constant ($K_m$) for metabolizing enzymes could still provide valuable information on drug clearance from plasma for the entire concentration range [35, 123].

***2003*** 1. An orally active $\alpha_{1A}$-adrenoceptor antagonist for the treatment of benign prostatic hyperplasia was administered to six healthy male volunteers with doses of 5, 50, and a pharmacological dose of 500 μg. $^{14}$C levels were measured in plasma by AMS, and the plasma pharmacokinetics was demonstrated to be linear between doses [8].

2. PET microdosing was first introduced [12, 90].

***2004*** 1. U.S. FDA Critical Path Initiative [23].

2. The microdose European Medicines Agency position paper [39].

3. (Microdosing in dogs) AMS was used to study the pharmacokinetics and disposition in dogs of a preclinical drug candidate (7-deaza-2′-C-methyladenosine) after oral and intravenous administration [138]. The primary objective of this study was to *examine* whether the drug displayed *linear kinetics* across subpharmacological (microdose) and pharmacological dose ranges in an animal model *before initiation of a human microdose study.* The AMS-derived disposition properties of the drug were comparable to data obtained via conventional techniques such as (LC-MS/MS) and liquid scintillation counting analyses (confirmed the use of AMS). The drug displayed multiphasic kinetics and exhibited low plasma clearance (5.8 mL/min/kg), a long terminal elimination half-life (17.5 h), and high oral bioavailability (103%). This study thus provided the first description of the full pharmacokinetic profile of a drug candidate assessed under subpharmacological and pharmacological doses. The pharmacokinetics in the dog appears to be linear across this 50-fold dose range.

***2005*** 1. A successful use of AMS in the generation of valuable pharmacokinetic information for nelfinavir in an absolute bioavailability study in healthy volunteers

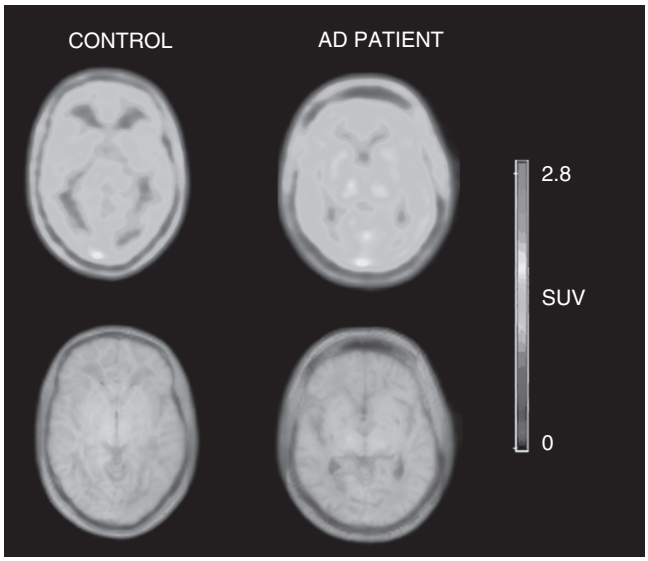


**Figure 23** Transaxial magnetic resonance–coregistered PET summation images recorded from 20 to 90 min after intravenous injection of [$^{11}$C]ST1859 into one control subject (HV 1) (*left*) and one AD patient (patient 9) (*right*). The *lower row* shows the same PET images with the transparency of the superimposed magnetic resonance scan set to 50%. The radioactivity concentration is normalized for injected radioactivity per body weight and expressed as the standardized uptake values (SUV).

that included the administration of an IV microdose of $^{14}$C-nelfinavir [139]. Because the moderate oral bioavailability of nelfinavir is due primarily to metabolic reasons, the results from this study indicated that traditional formulation improvement aimed at better gastrointestinal absorption would be unlikely to improve the oral bioavailability of nelfinavir in chronic treatment. This allowed the sponsor to make an appropriate cost–benefit decision and avoid allocating significant pharmaceutical science resources in developing a new oral formulation.

The author also concluded that human absolute bioavailability studies with administration of an IV microdose simultaneously with a conventional pharmacological oral dose of the same drug can yield important information about the gastrointestinal absorption versus first-pass metabolic turnover. Ultrasensitive and robust analytical techniques such as AMS can measure parent drug and metabolite concentrations with sufficient precision to characterize the full pharmacokinetic profile of a microdose. Administration of microdoses of $^{14}$C-labeled drugs, in conjunction with the sensitivity of the AMS approach, can reduce potential safety issues arising when dosing radioactivity to humans and prevent the need for formulation of high IV doses for absolute bioavailability studies.

2. (News Report [140]) Animal tests were made a legal requirement following the thalidomide tragedy 40 years ago, in the hope of preventing another such disaster. But have they lived up to their promise? Recently withdrawn arthritis drug Vioxx was safe and even beneficial to animal hearts but caused as many as 140,000 heart attacks and strokes in people—the biggest drug catastrophe in history. Shockingly, adverse reactions to prescription medicines (all tested for safety on animals) are now the fourth leading cause of death in the western world: killing over 10,000 people a year in the United Kingdom and costing the National Health Service (NHS) £466 million. New human-based safety tests before and during clinical trials (such as microdosing) could prevent many of these deaths.

3. First Human Data on Next Generation of Renin Inhibitors [141]. SPP630 and SPP635, compounds with low nanomolar potency being evaluated by Speedel (Basel, Switzerland, and Bridgewater, New Jersey) have been tested in microdosing studies with human volunteers and in preclinical studies with transgenic rats that express the human genes for renin and angiotensinogen. These studies show that these renin inhibitors demonstrated bioavailability in humans to 30% and in rats of 70–90%; have a half-life of more than 30 h in humans, making them suitable for once-daily dosing; show larger tissue distribution than previous renin inhibitors; and not only reduce blood pressure in transgenic rats over 24 h, but also slow the progression of renal damage after oral treatment. Both SPP630 and SPP635, together with other compounds from the SPP600 series, will be evaluated in toxicology studies before a decision is made about which compound(s) will continue in classic phase I studies with single-ascending and multiple-ascending doses.

***2006*** 1. U.S. FDA expIND guidance [40].

2. January 31, 2006, Xceleron news announcement regarding he European Union Microdose AMS Partnership Programme (EUMAPP), "This programme follows on from the successful CREAM trial and will add *7 more drugs* to the growing portfolio of compounds tested by Xceleron," said Professor Colin Garner, Xceleron's CEO. "Theorists can think of a number of situations whereby the predictivity of microdosing might be questioned, however in practice and taking into account many 'difficult' cases Xceleron has shown that microdosing should have an 80 per cent plus level of predictivity", said the spokesperson.

3. A feasibility study with conventional LC-MS/MS technology for oral microdosing assessment in rats (three drugs were studied) [132]. It was demonstrate that pharmacokinetics can be readily characterized by LC-MS/MS at microdose of 1 μg/kg for nonlabeled

drugs fluconazole and tolbutamide in rats, and, hence, LC-MS/MS should be adequate in human microdosing studies. The studies also exhibit linearity in exposure between the microdose and ≥1000-fold higher doses in rats for these drugs, which are known to show linear dose-exposure relationship in the clinic, further substantiating the potential utility of LC-MS/MS in defining pharmacokinetics from the microdose of drugs. These data should increase confidence on the use of LC-MS/MS in microdose pharmacokinetics studies of new chemical entities in humans. Application of this approach is also described for an investigational compound MLNX where the pharmacokinetics in rats were determined to be nonlinear, suggesting that MLNX pharmacokinetics at microdoses in humans also might not reflect those at the therapeutic doses.

4. The first (and most up-to-date) comprehensive human study comparing microdose and therapeutic dose pharmacokinetics was completed under the support of the Consortium for Resourcing and Evaluating AMS Microdosing (CREAM) [35, 124].

5. A PET microdosing study [92] has been carried out with an investigational antiamyloid drug (ST1859) with three healthy volunteers and six patients with Alzheimer's disease (see PET microdosing example section for details).

***2007*** 1. AMS microdosing study of the pharmacokinetics and metabolism of imatinib in patients on long-term therapy for chronic myelogenous leukemia [142]. This is possibly the first human study involving an anticancer drug that investigates the pharmacokinetic profile of an individual dose under steady-state conditions [32]. Six patients who were receiving imatinib (400 mg/day) as part of their ongoing treatment for chronic myeloid leukemia (CML) received a dose containing a trace quantity (13.6 kBq) of C-14-imatinib. Blood samples were collected from patients before and at various times up to 72 h after administration of the test dose and were processed to provide samples of plasma and peripheral blood lymphocytes (PBL). Samples were analyzed by AMS, with chromatographic separation of parent compound from metabolites. In addition, plasma samples were analyzed by LC/MS. Analysis of the AMS data indicated that imatinib was rapidly absorbed and could be detected in plasma up to 72 h after administration. Imatinib was also detectable in PBL at 24 h after administration of the C-14-labeled dose. Comparison of plasma concentrations determined by AMS with those derived by LC/MS analysis gave similar average estimates of area under plasma concentration time curve ($26 \pm 3$ versus $27 \pm 11\,\mu g/mL \cdot h$), *confirming* the predictability of microdosing study, but with some variation within each individual. The demonstration that AMS can be applied to investigate the pharmacokinetics of a single dose of imatinib in the context of chronic therapy provides a unique opportunity to revisit and reevaluate the drug interaction potential of imatinib at steady state [32]. Similar approaches can also be applied using a radiolabeled intravenous tracer dose of imatinib administered simultaneously with a nonlabeled oral dose [143] to determine absolute bioavailability in individual patients and discriminate between absorption and elimination as a potential cause for the occurrence of pharmacokinetic resistance to imatinib treatment.

2. (Not a microdosing study) PET studies were carried out in three baboons to measure pharmacokinetics in the brain and peripheral organs and to determine the effect of a therapeutic dose of imatinib [144]. Imatinib mesylate (Gleevec) is a well-known drug for treating chronic myeloid leukemia and gastrointestinal stromal tumors. its active ingredient, imatinib ([4-[(4-methyl-l-piperazinyl)methyl]-*N*-[4-methyl-3-[[4-(3-pyridyl)-2-pyrimidinyl]amino]phenyl]benzamide), blocks the activity of several tyrosine kinases. This work labeled imatinib with carbon-11 as a tool for determining the drug distribution and pharmacokinetics of imatinib. The results showed that [N-C-11-methyl]imatinib uptake in the brain is negligible (consistent with P-glycoprotein-mediated efflux); it peaks and clears rapidly from the heart, lungs and spleen. Peak uptake and clearance occur more slowly in the liver and kidneys, followed by accumulation in the gallbladder and urinary bladder. Pretreatment with imatinib did not change uptake in the heart, lungs, kidneys and spleen, and increased uptake in the liver and gallbladder. Therefore, [N-C-11-methyl]imatinib has *potential* for assessing the regional distribution and kinetics of imatinib in the human body to determine whether the drug targets tumors and to identify other organs to which the drug or its labeled metabolites distribute. Paired with tracers such as 2-deoxy-2-[F-18]fluoro-D-glucose ($^{18}$FDG) and 3′-deoxy-3′-[F-18]fluorothymidine ((FLT)-F-18), [N-C-11-methyl]imatinib may be a useful radiotracer for planning chemotherapy, for monitoring response to treatment, and for assessing the role of drug pharmacokinetics in drug resistance.

3. Belgian national guidance for exploratory studies in humans [41].

4. (News report) Microdosing with human tissues [145]. Robert Schmouder, executive director of translational medicine at Novartis, said at a Cambridge Healthtech Institute conference translational medicine that preclinical microdosing studies aimed at gauging the activity of investigational drugs in humans on a metabolic and pharmacokinetic level, commonly referred to as "phase 0" studies, won't predict whether a drug candidate will succeed or fail at higher doses and in broader populations. These studies rely on "a much

smaller dose than you are going to give actually, so you are really restricted in the amount of drug you can give. . . . We see a lot of examples where [microdose studies] don't give you the data that you want to have" before going into clinical development. As a company that has performed between 5 and 10 of the eINDs, Novartis is "still assessing what the utility of the eINDs are," according to Schmouder. However, a different type of service, called Phase Zero and provided by Asterand, relies on human tissues rather than human beings, and may be right up pharma's alley as companies increasingly try to employ predictive technologies to make go/no-go decisions in preclinical studies. Using human tissue samples rather than human beings, Asterand's Phase Zero Drug Discovery Service investigates a drug candidate's efficacy, pharmacology, safety profiling, and metabolism by using a variety of techniques, including gene expression profiling and protein expression profiling. While some in the industry may question the utility of FDA's phase 0 studies, Asterand's version may be more attractive to big pharma, particularly at a time when the industry is embracing R&D strategies that provide greater clarity about the likelihood of a drug candidate's success prior to entering full-fledged clinical development. According to Amanda Woodrooffe of Asterand's department of biochemical pharmacology, *18 out of the top 20 pharma companies, as well as key biotech companies, are using the company's Phase Zero services.* "I think the pharmaceutical industry is really buying into managing risk and helping build confidence that they are making the right decision early on." Asterand's Phase Zero services could be attractive to industry because they allow companies to use human tissue approaches in their early discovery phase to weed out the compounds that have lower chances for success. In this way "the companies can actually focus their budgets on investing in the compounds that are more likely to have success once they do get into the clinic."

According to the FDA, 9 out of 10 compounds fail in the clinic because they were advanced to human subjects based on how they metabolized in animal studies.

5. September 5, 2007, Closing remark by Dr. James H. Doroshow of Nation Cancer Institute workshop on "Phase 0 Trials in Oncologic Drug Development" [146].

Are phase 0 studies feasible? (a) Phase 0 recruitment is feasible but depends on patient altruism and prior physician–patient relationship. (b) Statistical methodology is more challenging in phase 0 trials, which involve novel PD endpoints and a small number of samples. (c) Require additional upfront investment in imaging or assay development, in addition to that needed for standard PK.

In what circumstances are phase 0 studies applicable for early drug development? Phase 0 trials provide the opportunity to develop PD assays measuring target function in tumors in humans. It (a) allows for the assessment of target inhibition as a potential primary endpoint in subsequent trials; (b) develops assay procedures, from tissue acquisition to the assay itself, that actually work in patients in the clinic.

Does conducting a phase 0 study save clinical development time? Critical question is time. It will take years to assess, but likely that imaging-driven PD will make a major impact soon.

6. Results from anonymous Xceleron studies [38]: A basic drug (sleep disorder) showed PK scalability between microdose and therapeutic dose in a 50-fold dose range. A basic drug with long $t_{1/2}$(ca. 100 h, sedative) showed PK scalability between microdose and therapeutic dose in a 1000-fold dose range. A drug with extensive phase II metabolism (analgesic) showed PK scalability between microdose and therapeutic dose in a 50,000-fold dose range. A rennin inhibitor showed PK scalability between microdose and therapeutic dose in a 500-fold dose range. A CNS drug with very short $t_{1/2}$(<1 h) showed PK scalability between microdose and therapeutic dose in a 1250-fold dose range.

***2008*** 1. April 5, 2008. Linear pharmacokinetics was demonstrated between a microdose and the therapeutic dose of the PgP substrate fexofenadine in humans, one of the drugs listed in the EUMAPP studies. It was presented by Prof. Colin Garner, CEO of Xceleron, as a late-breaking abstract at ASCPT Orlando, FL. Fexofenadine has been IV microdosed along with an oral therapeutic dose. The mean oral absolute bioavailability is 33% in contrast to the 11% estimation based on unchanged drug in urine, indicating a higher degree of biliary excretion than originally thought.

2. May 1, 2008, Corcept Therapeutics Incorporated (NASDAQ: CORT) announced that its lead selective GR-II antagonist, CORT 108297, produced promising results in a human microdosing study. "The compound was extremely well absorbed, demonstrated good bioavailability and had a half-life that appears compatible with once-a-day oral dosing in this 'first-in-man' study." Corcept had previously announced its discovery of three series of structurally different compounds that are potent blockers of the cortisol receptor but do not block the progesterone or any of the other steroid receptors in in vitro assays. "These microdosing results increase our confidence that we will be able to develop cortisol antagonists that are not abortifacients." Corcept is studying disease states associated with excess production of cortisol, the "stress" hormone.

In summary, there are 20 drugs of microdosing studies that have been carried out so far (16 in humans, 1 in dogs, 3 in rats), 6 additional drugs of microdosing studies are currently being carried out under the European

Union Microdose AMS Partnership Programme (EUMAPP).

## 7 CONCLUSIONS

Human microdosing study is one type of exploratory phase I investigation supported by both U.S. and European regulatory agencies. Very small subpharmacologic doses of a drug are used in humans to obtain an early assessment of pharmacokinetic properties in humans. This has the potential of reducing the number of preclinical animal studies, essential for conducting early clinical human studies, differentiating between good and bad candidate drugs early, and speeding up the early phases of drug development. Human microdosing is a newly established and still evolving concept, around which knowledge begins to accumulate.

Microdosing depends on the availability of ultrasensitive analytical methods, such as AMS and PET that are able to measure drug and metabolite concentrations in the low pictogram to femotgram range. AMS is used for determining PK data by taking body samples over time, processing the samples in the laboratory and then analyzing their drug content. PET provides primarily PD data through real-time imaging and organ-specific PK data.

Due to the high expenditure and a disappointingly low number of compounds achieving market authorization each year, the pharmaceutical industry has undergone a paradigm shift toward a more mechanistic and predictive scientific model for drug development. AMS and PET microdosing are two of the innovative techniques that were employed for rational selection of drug candidates in the strong drive to improve overall success rates of the drug development process.

## REFERENCES


1. DiMasi JA (2001). Risks in new drug development: Approval success rates for investigational drugs. *Clin Pharmacol Ther 69*(5), 297–307.
2. Prentis RA, Lis Y, Walker SR (1988). Pharmaceutical Innovation by the 7 UK-owned pharmaceutical companies (1964–1985). *Br J Clin Pharmacol 25*(3), 387–396.
3. Boyd RA, Lalonde RL (2007). Nontraditional approaches to first-in-human studies to increase efficiency of drug development, will microdose studies make a significant impact? *Clin Pharmacol Ther 81*(1), 24–26.
4. Dickson M, Gagnon JP (2004). Key factors in the rising cost of new drug discovery and development. *Nat Rev Drug Discov 3*(5), 417–429.
5. DiMasi JA, Hansen RW, Grabowski HG (2003). The price of innovation: New estimates of drug development costs. *J Health Econ 22*(2), 151–185.
6. McLean MA, Tam CYJ, Baratta MT, Holliman CL, Ings RM, Galluppi GR (2007). Accelerating drug development: Methodology to support first-in-man pharmacokinetic studies by the use of drug candidate microdosing. *Drug Devel Res 68*(1), 14–22.
7. Garner RC (2005). Less is mores: The human microdosing concept. *Drug Discov Today 10*(7), 449–451.
8. Lappin G, Garner RC (2003). Big physics, small doses: The use of AMS and PET in human microdosing of development drugs. *Nat Rev Drug Discov 2*(3), 233–240.
9. Robinson WT (2008). Innovative early development regulatory approaches: expIND, expCTA, microdosing. *Clin Pharmacol Ther 83*(2), 358–360.
10. Hellborg R, Faarinen M, Kiisk M, Magnusson CE, Persson P, Skog G, Stenstrom K (2003). Accelerator mass spectrometry—An overview. *Vacuum 70*(2–3), 365–372.
11. Palmblad M, Buchholz BA, Hillegonds DJ, Vogel JS (2005). Neuroscience and accelerator mass spectrometry. *J Mass Spectrom 40*(2), 154–159.
12. Bergstrom M, Grahnen A, Langstrom B (2003). Positron emission tomography microdosing: A new concept with application in tracer and early clinical drug development. *Eur J Clin Pharmacol 59*(5–6), 357–366.
13. Bajorath F (2002). Integration of virtual and high-throughput screening. *Nat Rev Drug Discov 1*(11), 882–894.
14. Blundell TL, Jhoti H, Abell C (2002). High-throughput crystallography for lead discovery in drug design. *Nat Rev Drug Discov 1*(1), 45–54.
15. Renaud JP, Peyronneau MA, Urban P, Truan G, Cullin C, Pompon D, Beaune P, Mansuy D (1993). Recombinant yeast in drug-metabolism. *Toxicol 82*(1–3), 39–52.
16. Bachmann KA, Ghosh R (2001). The use of in vitro methods to predict in vivo pharmacokinetics and drug interactions. *Curr Drug Metab 2*(3), 299–314.
17. Testino SA, Patonay G (2003). High-throughput inhibition screening of major human cytochrome P450 enzymes using an in vitro cocktail and liquid chromatography-tandem mass spectrometry. *J Pharm Biomed Anal 30*(5), 1459–1467.
18. Mahmood I (1998). Interspecies scaling: Predicting volumes, mean residence time and elimination half-life. Some suggestions. *J Pharm Pharmacol 50*(5), 493–499.
19. Jang GR, Harris RZ, Lau DT (2001). Pharmacokinetics and its role in small molecule drug discovery research. *Med Res Rev 21*(5), 382–396.
20. Mahmood I (1999). Allometric issues in drug development. *J Pharm Sci 88*(11), 1101–1106.
21. Bolten BM, DeGregorio T (2002). Trends in development cycles. *Nat Rev Drug Discov 1*(5), 335–336.
22. Marchetti S, Schellens JHM (2007). The impact of FDA and EMEA guidelines on drug development in relation to Phase 0 trials. *Br J Cancer 97*(5), 577–581.

23. U.S. Department of Health and Human Services, Food and Drug Administration (FDA) (March 2004). *Innovation/stagnation: Challenges and opportunities on the critical path to new medical products*. Washington, DC: FDA.
24. European Medicines Agency (EMA) (March 2005). *The European Medicines Agency road map to 2010: Preparing the ground for the future*. Washington, DC: EMA.
25. Garner RC, Lappin G (2006). The phase 0 microdosing concept. *Br J Clin Pharmacol 61*(4), 367–370.
26. Curry SH (2008). Translational science: Past, present, and future. *Biotechniques 44*(2), II–VIII.
27. FitzGerald GA (2005). Opinion—Anticipating change in drug development: The emerging era of translational medicine and therapeutics. *Nat Rev Drug Discov 4*(10), 815–818.
28. Vail DM (2007). Cancer clinical trials: Development and implementation. *Vet Clin North Am Small Anim Pract 37*(6), 1033–1057.
29. Rocchetti M, Simeoni M, Pesenti E, De Nicolao G, Poggespi I (2007). Predicting the active doses in humans from animal studies: A novel approach in oncology. *Eur J Cancer 43*(12), 1862–1868.
30. Kummar S, Kinders R, Rubinstein L, Parchment RE, Murgo AJ, Collins J, Pickeral O, Low J, Steinberg SM, Gutierrez M, Yang S, Helman L, Wiltrout R, Tomaszewski JE, Doroshow JH (2007). Compressing drug development timelines in oncology using phase '0' trials. *Nat Rev Cancer 7*(2), 131–139.
31. Eap CB, Buclin T, Cucchia G, Zullino D, Hustert E, Bleiber G, Golay KP, Aubert AC, Baumann P, Telenti A, Kerb R (2004). Oral administration of a low dose of midazolam (75 mu g) as an in vivo probe for CYP3A activity. *Eur J Clin Pharmacol 60*(4), 237–246.
32. Sparreboom A (2007). Unexplored pharmacokinetic opportunities with microdosing in oncology. *Clin Cancer Res 13*(14), 4033–4034.
33. Schellens JHM, Planting AST, Vanacker BAC, Loos WJ, Deboerdennert M, Vanderburg MEL, Koier I, Krediet RT, Stoter G, Verweij J (1994). Phase-I and pharmacological study of the novel indoloquinone bioreductive alkylating cytotoxic drug E09. *J Nat Cancer Inst 86*(12), 906–912.
34. Vonhoff DD, Turner J (1991). Response rates, duration of response, and dose-response effects in phase-I studies of antineoplastics. *Invest New Drugs 9*(1), 115–122.
35. European Medicines Agency (EMA) (June 2004). *Position paper on non-clinical safety studies to support clinical trials with a single microdose*. Washington, DC: EMA.
36. Lappin G, Kuhnz W, Jochemsen R, Kneer J, Chaudhary A, Oosterhuis B, Drijfhout WJ, Rowland M, Garner RC (2006). Use of microdosing to predict pharmacokinetics at the therapeutic dose: Experience with 5 drugs. *Clin Pharmacol Ther 80*(3), 203–215.
37. Wilding IR, Bell JA (2005). Improved early clinical development through human microdosing studies. *Drug Discov Today 10*(13), 890–894.
38. Singh SS (2006). Preclinical pharmacokinetics: An approach towards safer and efficacious drugs. *Curr Drug Metab 7*(2), 165–182.
39. Lappin G (2007). *A general introduction to microdosing*. Paper presented at the Translation Medicine Microdosing Workshop, Xceleron. Berlin, 1 July.
40. U.S. Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research (CDER) (January 2006). *Guidance for industry, investigators, and reviewers. Exploratory IND studies*. Washington, DC: CDER.
41. Federal Agency for Medicines and Health Products in Belgium (FAMHP) (June 2007). *Guidance to the conduct of exploratory trials in belgium*. FAMHP. Belgium.
42. http://www.uncaged.co.uk/news/2006/tgn1412.htm.
43. http://www.i-sis.org.uk/LDTC.php.
44. Alvarez LW, Cornog R (1939). Helium and hydrogen of mass 3. *Phys Rev 56*(6), 613–613.
45. Alvarez LW, Cornog R (1939). He-3 in helium. *Phys Rev 56*(4), 379–379.
46. Barker J, Garner RC (1999). Biomedical applications of accelerator mass spectrometry isotope measurements at the level of the atom. *Rapid Commun Mass Spectrom 13*(4), 285–293.
47. Muller RA (1977). Radioisotope dating with a cyclotron. *Sci 196*(4289), 489–494.
48. Nelson DE, Korteling RG, Stott WR (1977). C-14-Direct detection at natural concentrations. *Sci 198*(4316), 507–508.
49. Bennett CL, Beukens RP, Clover MR, Gove HE, Liebert RB, Litherland AE, Purser KH, Sondheim WE (1977). *Radiocarbon* dating using electrostatic accelerators—Negative-ions provide key. *Sci 198*(4316), 508–510.
50. Fifield LK (1999). Accelerator mass spectrometry and its applications. *Rep Prog Phys 62*(8), 1223–1274.
51. Purser KH, Liebert RB, Russo CJ (1980). Macs—An accelerator-based radioisotope measuring system. *Radiocarbon 22*(3), 794–806.
52. Elmore D, Fulton BR, Clover MR, Marsden JR, Gove HE, Naylor H, Purser KH, Kilius LR, Beukens RP, Litherland AE (1979). Analysis of Cl-36 in environmental water samples using an electrostatic accelerator. *Nature 277*(5691), 22–25.
53. Elmore D, Gove HE, Ferraro R, Kilius LR, Lee HW, Chang KH, Beukens RP, Litherland AE, Russo CJ, Purser KH, Murrell MT, Finkel RC (1980). Determination of I-129 using tandem accelerator mass-spectrometry. *Nature 286*(5769), 138–140.
54. Klein J, Middleton R, Tang H (1982). Modifications of an Fn tandem for quantitative Be-10 measurement. *Nucl Instrum Methods Phys Res 193*(3), 601–616.

55. Middleton R, Klein J, Raisbeck GM, Yiou F (1983). Accelerator mass-spectrometry with Al-26. *Nucl Instrum Methods Phys Res 218*(1–3), 430–438.
56. Vogel JS, Turteltaub KW, Finkel R, Nelson DE (1995). Accelerator mass-spectrometry—Isotope quantification at attomole sensitivity. *Anal Chem 67*(11), A353–A359.
57. Vogel JS, McAninch J, Freeman S (1997). Elements in biological AMS. *Nucl Instrum Methods Phys Res Sect B Beam Interact Mater Atoms 123*(1–4), 241–244.
58. Turteltaub KW, Vogel JS (2000). Bioanalytical applications of accelerator mass spectrometry for pharmaceutical research. *Curr Pharm Des 6*(10), 991–1007.
59. Chiarappa-Zucca ML, Finkel RC, Martinelli RE, McAninch JE, Nelson DO, Turteltaub KW (2004). Measurement of beryllium in biological samples by accelerator mass spectrometry: Applications for studying chronic beryllium disease. *Chem Res Toxicol 17*(12), 1614–1620.
60. Felton JS, Turteltaub KW (1994). Accelerator mass-spectrometry for measuring low-dose carcinogen binding to DNA. *Environ Health Perspect 102*(5), 450–452.
61. Lappin G, Garner RC (2004). Current perspectives of C-14-isotope measurement in biomedical accelerator mass spectrometry. *Anal Bioanal Chem 378*(2), 356–364.
62. http://en.wikipedia.org/wiki/Microdosing.
63. Brown K, Tompkins EM, White INH (2006). Applications of accelerator mass spectrometry for pharmacological and toxicological research. *Mass Spectrom Rev 25*(1), 127–145.
64. Elmore D, Phillips FM (1987). Accelerator mass-spectrometry for measurement of long-lived radioisotopes. *Sci 236*(4801), 543–550.
65. Litherland AE (1987). Fundamentals of accelerator mass-spectrometry. *Philos Trans R Soc Lond A Mathe Phys Eng Sci 323*(1569), 5–21.
66. Buchholz BA, Freeman S, Haack KW, Vogel JS (2000). Tips and traps in the C-14 bio-AMS preparation laboratory. *Nucl Instrum Methods Phys Res Sect B Beam Interact Mater Atoms 172*, 404–408.
67. Vogel JS (1992). Rapid production of graphite without contamination for biomedical Ams. *Radiocarbon 34*(3), 344–350.
68. Ognibene TJ, Bench G, Vogel JS, Peaslee GF, Murov S (2003). A high-throughput method for the conversion of CO2 obtained from biochemical samples to graphite in septa-sealed vials for quantification of C-14 via accelerator mass spectrometry. *Anal Chem 75*(9), 2192–2196.
69. Hughey BJ, Klinkowstein RE, Shefer RE, Skipper PL, Tannenbaum SR, Wishnok JS (1997). Design of a compact 1 MV AMS system for biomedical research. *Nucl Instrum Methods Phys Res Sect B Beam Interact Mater Atoms 123*(1–4), 153–158.
70. Hughey BJ, Shefer RE, Klinkowstein RE, Zhao XL, Kieser WE, Litherland AE (1997). Experimental investigation of the destruction of 1 MeV (CH2)-C-12 molecules in single and double stripping foils. *Nucl Instrum Methods Phys Res Sect B Beam Interact Mater Atoms 123*(1–4), 186–192.
71. Ognibene TJ, Bench G, Brown TA, Peaslee GF, Vogel JS (2002). A new accelerator mass spectrometry system for C-14-quantification of biochemical samples. *Int J Mass Spectrom 218*(3), 255–264.
72. Dirac PAM (1930). A theory of electrons and protons. *Proc R Soc Lond A Containing Pap Mathe Phys Character 126*(801), 360–365.
73. Anderson CD (1932). Energies of cosmic-ray particles. *Phys Rev 41*(4), 405–421.
74. Anderson H, Price P (2000). What does positron emission tomography offer oncology? *Eur J Cancer 36*(16), 2028–2035.
75. Wagner HN (1998). A brief history of positron emission tomography (PET). *Semin Nucl Med 28*(3), 213–220.
76. Sweet WH (1951). The uses of nuclear disintegration in the diagnosis and treatment of brain tumor. *N Engl J Med 245*(23), 875–878.
77. Nutt R (2002). The history of positron emission tomography. *Mol Imag Biol 4*, 11–26.
78. Wrenn FR, Good ML, Handler P (1951). The use of positron-emitting radioisotopes for the localization of brain tumors. *Sci 113*(2940), 525–527.
79. Kuhl DE, Chamberlain RH, Hale J, Gorson RO (1956). A high-contrast photographic recorder for scintillation counter scanning. *Radiology 66*(5), 730–739.
80. Kuhl DE, Edwards RQ (1962). Body section radioisotope scanning. *J Nucl Med 3*(3), 199–199.
81. Derenzo SE, Weber MJ, Bourret-Courchesne WE, Klintenberg MK (2003). The quest for the ideal inorganic scintillator. *Nucl Instrum Methods Phys Res Sect A Acceler Spectrom Detect Assoc Equip 505*(1–2), 111–117.
82. Cho ZH, Farukhi MR (1977). Bismuth germanate as a potential scintillation detector in positron cameras. *J Nucl Med 18*(8), 840–844.
83. Derenzo SE (1981). Monte-Carlo calculations of the detection efficiency of arrays of Nai(Tl), Bgo, Csf, Ge, and plastic detectors for 511-Kev photons. *IEEE Trans Nucl Sci 28*(1), 131–136.
84. Ido T, Wan CN, Casella V, Fowler JS, Wolf AP, Reivich M, Kuhl DE (1978). Labeled 2-deoxy-D-glucose analogs-F-18-labeled 2-deoxy-2-fluoro-D-glucose, 2-deoxy-2-fluoro-D-mannose and C-14–2-deoxy-2-fluoro-D-glucose. *J Label Comp Radiopharm 14*(2), 175–183.
85. Hamacher K, Coenen HH, Stocklin G (1986). Efficient stereospecific synthesis of no-carrier-added 2-[F-18]-fluoro-2-deoxy-D-glucose using aminopolyether supported nucleophilic-substitution. *J Nucl Med 27*(2), 235–238.
86. Jones T (1996). The role of positron emission tomography within the spectrum of medical imaging. *Eur J Nucl Med 23*(2), 207–211.

87. Jones T (1996). The imaging science of positron emission tomography. *Eur J Nucl Med 23*(7), 807–813.
88. Phelps ME (2000). Positron emission tomography provides molecular imaging of biological processes. *Proc Nat Acad Sci USA 97*(16), 9226–9233.
89. Tilyou SM (1991). The evolution of positron emission tomography. *J Nucl Med 32*(4), N15–N26.
90. Wagner CC, Muller M, Lappin G, Langer O (2008). Positron emission tomography for use in microdosing studies. *Curr Opin Drug Discov Dev 11*(1), 104–110.
91. Langer O, Kuntner C, Wanek T, Bauer M, Mandler M, Angelberger P, Muller M, Kvaternik H (2006). MicroPET imaging with the beta-amyloid probe [F-18]FDDNP in a transgenic mouse model of Alzheimer's disease. *Pharmacol 78*(3), 153–153.
92. Bauer M, Langer O, Dal-Bianco P, Karch R, Brunner M, Abrahim A, Lanzenberger R, Hofmann A, Joukhadar C, Carminati P, Ghirardi O, Piovesan P, Forloni G, Corrado ME, Lods N, Dudczak R, Auff E, Kletter K, Muller M (2006). A positron emission tomography microdosing study with a potential antiamyloid drug in healthy volunteers and patients with Alzheimer's disease. *Clin Pharmacol Ther 80*(3), 216–227.
93. Cherry SR (2001). Fundamentals of positron emission tomography and applications in preclinical drug development. *J Clin Pharmacol 41*(5), 482–491.
94. McCarthy TJ, Wetch MJ (1998). The state of positron emitting radionuclide production in 1997. *Semin Nucl Med 28*(3), 235–246.
95. Welch MJ, McCarthy TJ (2000). The potential role of generator-produced radiopharmaceuticals in clinical PET. *J Nucl Med 41*(2), 315–317.
96. Tewson TJ, Krohn KA (1998). PET radiopharmaceuticals: State-of-the-art and future prospects. *Semin Nucl Med 28*(3), 221–234.
97. Zaidi H, Montandon ML (2006). The new challenges of brain PET imaging technology. *Curr Med Imag Rev 2*(1), 3–13.
98. Tarantola G, Zito F, Gerundini P (2003). PET instrumentation and reconstruction algorithms in whole-body applications. *J Nucl Med 44*(5), 756–769.
99. Zanzonico P (2004). Positron emission tomography: A review of basic principles, scanner design and performance, and current systems. *Semin Nucl Med 34*(2), 87–111.
100. Lewellen TK, Kohlmyer SG, Miyaoka RS, Kaplan MS (1996). Investigation of the performance of the general electric ADVANCE positron emission tomograph in 3D mode. *IEEE Trans Nucl Sci 43*(4), 2199–2206.
101. Adam LE, Zaers J, Ostertag H, Trojan H, Bellemann ME, Brix G (1997). Performance evaluation of the whole-body PET scanner ECAT EXACT HR+ following the IEC standard. *IEEE Trans Nucl Sci 44*(3), 1172–1179.
102. Budinger TF (1998). PET instrumentation: What are the limits? *Semin Nucl Med 28*(3), 247–267.
103. Wienhard K, Schmand M, Casey ME, Baker K, Bao J, Eriksson L, Jones WF, Knoess C, Lenox M, Lercher M, Luk P, Michel C, Reed JH, Richerzhagen N, Treffert J, Vollmar S, Young JW, Heiss WD, Nutt R (2002). The ECAT HRRT: Performance and first clinical application of the new high resolution research tomograph. *IEEE Trans Nucl Sci 49*(1), 104–110.
104. Casey ME, Nutt R (1986). A multicrystal 2-dimensional Bgo detector system for positron emission tomography. *IEEE Trans Nucl Sci 33*(1), 460–463.
105. Dahlbom M, MacDonald LR, Schmand M, Eriksson L, Andreaco M, Williams C (1998). A YSO/LSO phoswich array detector for single and coincidence photon imaging. *IEEE Trans Nucl Sci 45*(3), 1128–1132.
106. Braem A, Llatas MC, Chesi E, Correia JG, Garibaldi F, Joram CX, Mathot S, Nappi E, da Silva MR, Schoenahl F, Seguinot J, Weilhammer P, Zaidi H (2004). Feasibility of a novel design of high resolution parallax-free Compton enhanced PET scanner dedicated to brain research. *Phys Med Biol 49*(12), 2547–2562.
107. Langner J (2003). *Development of a parallel computing optimized head movement correction method in positron-emission-tomography*. Dresden, Germany: University of Applied Sciences.
108. http://en.wikipedia.org/wiki/Positron_emission_tomography.
109. Shimizu K, Ohmura T, Watanabe M, Uchida H, Yamashita T (1988). Development of 3-D detector system for positron Ct. *IEEE Trans Nucl Sci 35*(1), 717–720.
110. Li HD, Wong WH, Baghaei H, Uribe J, Wang Y, Zhang YX, Kim SS, Ramirez R, Liu JG, Liu ST (2007). The engineering and initial results of a transformable low-cost high-resolution PET camera. *IEEE Trans Nucl Sci 54*(5), 1583–1588.
111. Beyer T, Townsend DW, Brun T, Kinahan PE, Charron M, Roddy R, Jerin J, Young J, Byars L, Nutt R (2000). A combined PET/CT scanner for clinical oncology. *J Nucl Med 41*(8), 1369–1379.
112. Zaidi H, Alavi A (2007). Current trends in PET and combined (PET/CTand PET/MR) systems design. *PET Clin 2*, 109–123.
113. Boss DS, Olmos RV, Sinaasappel M, Beijnen JH, Schellens JHM (2008). Application of PET/CT in the development of novel anticancer drugs. *Oncologist 13*(1), 25–38.
114. Cherry SR (2006). Multimodality in vivo imaging systems: Twice the power or double the trouble? *Annu Rev Biomed Eng 8*, 35–62.
115. Cherry SR, Louie AY, Jacobs RE (2008). The integration of positron emission tomography with magnetic resonance imaging. *Proc IEEE 96*(3), 416–438.
116. Anguelov Z (1999). Quantitative anatomy: Power beyond the images. *Theor Med Bioethics 20*(6), 501–516.
117. Ollinger JM, Fessler JA (1997). Positron-emission tomography. *IEEE Signal Process Mag 14*(1), 43–55.
118. Mankoff DA (2007). *Imaging to guide early drug trials (NIH)*. NIH Phase 0 Trials in Oncologic

Drug Development. Available: http://dctd.cancer.gov/MajorInitiatives/Sep0507Phase0Workshop/workshop.htm.

119. Wolf W (1995). Imaging can be much more than pretty pictures. *Pharm Res 12*(12), 1821–1822.
120. Weber WA, Czernin J, Phelps ME, Herschman HR (2008). Technology Insight: Novel imaging of molecular targets is an emerging area crucial to the development of targeted drugs. *Nat Clin Pract Oncol 5*(1), 44–54.
121. Hartvig P, Bergstrom M, Langstrom B (2001). Use of positron emission tomography in analysing receptor function in vivo. *Toxicol Lett 120*(1–3), 243–251.
122. Reichel A (2006). The role of blood-brain barrier studies in the pharmaceutical industry. *Curr Drug Metab 7*(2), 183–203.
123. Saleem A, Aboagye EO, Matthews JC, Price PM (2008). Plasma pharmacokinetic evaluation of cytotoxic agents radiolabelled with positron emitting radioisotopes. *Cancer Chemother Pharmacol 61*(5), 865–873.
124. Lappin G, Garner RC (2006). A review of human phase 0 (microdosing) clinical trials following the US Food and Drug Administration exploratory investigational new drug studies guidance. *Int J Pharm Med 20*(3), 159–165.
125. Turteltaub KW, Felton JS, Gledhill BL, Vogel JS, Southon JR, Caffee MW, Finkel RC, Nelson DE, Proctor ID, Davis JC (1990). Accelerator mass-spectrometry in biomedical dosimetry—Relationship between low-level exposure and covalent binding of heterocyclic amine carcinogens to DNA. *Proc Nat Acad Sci USA 87*(14), 5288–5292.
126. Mauthe RJ, Dingley KH, Leveson SH, Freeman S, Turesky RJ, Garner RC, Turteltaub KW (1999). Comparison of DNA-adduct and tissue-available dose levels of MeIQx in human and rodent colon following administration of a very low dose. *Int J Cancer 80*(4), 539–545.
127. Turteltaub KW, Mauthe RJ, Dingley KH, Vogel JS, Frantz CE, Garner RC, Shen N (1997). MeIQx-DNA adduct formation in rodent and human tissues at low doses. *Mutat Res Fund Mol Mechan Mutagen 376*(1–2), 243–252.
128. Bertino JS, Greenberg HE, Reed MD (2007). American College of Clinical Pharmacology position statement on the use of microdosing in the drug development process. *J Clin Pharmacol 47*(4), 418–422.
129. Rowland M (2007). Commentary on ACCP position statement on the use of microdosing in the drug development process. *J Clin Pharmacol 47*(12), 1595–1596.
130. Bertino JS, Greenberg HE, Reed MD (2007). Response to comments on ACCP position statement on microdosing. *J Clin Pharmacol 47*(12), 1597–1598.
131. Twombly R (2006). Slow start to phase 0 as researchers debate value. *J Nat Cancer Inst 98*(12), 804–806.
132. Balani SK, Nagaraja NV, Qian MG, Costa AO, Daniels JS, Yang H, Shimoga PR, Wu JT, Gan LS, Lee FW, Miwa GT (2006). Evaluation of microdosing to assess pharmacokinetic linearity in rats using liquid chromatography-tandem mass spectrometry. *Drug Metab Dispos 34*(3), 384–388.
133. Levy G, Mager DE, Cheung WK, Jusko W (2003). Comparative pharmacokinetics of coumarin anticoagulants L: Physiologic modeling of S-warfarin in rats and pharmacologic target-mediated warfarin disposition in man. *J Pharm Sci 92*(5), 985–994.
134. Christian BT, Livni E, Babich JW, Alpert NM, Dischino DD, Ruediger E, Salazar DE, Ford NF, Fischman AJ (1996). Evaluation of cerebral pharmacokinetics of the novel antidepressant drug, BMS-181101, by positron emission tomography. *J Pharmacol Exp Ther 279*(1), 325–331.
135. Salazar DE, Fischman AJ (1999). Central nervous system pharmacokinetics of psychiatric drugs. *J Clin Pharmacol 39*, 10S–12S.
136. Saleem A, Harte RJ, Matthews JC, Osman S, Brady F, Luthra SK, Brown GD, Bleehen N, Conners T, Jones T, Price PM, Aboagye EO (2001). Pharmacokinetic evaluation of *N*-[2(dimethylamino)ethyl]acridine-4-carboxamide in patients by positron emission tomography. *J Clin Oncol 19*(5), 1421–1429.
137. Tilsley DWO, Harte RJA, Jones T, Brady F, Luthra SK, Brown G, Price PM (1993). New techniques in the pharmacokinetic analysis of cancer drugs. 4. Positron Emission Tomography. *Cancer Surv 17*, 425–442.
138. Sandhu P, Vogel JS, Rose MJ, Ubick EA, Brunner JE, Wallace MA, Adelsberger JK, Baker MP, Henderson PT, Pearson PG, Baillie TA (2004). Evaluation of microdosing strategies for studies in preclinical drug development: Demonstration of linear pharmacokinetics in dogs of a nucleoside analog over a 50-fold dose range. *Drug Metab Dispos 32*(11), 1254–1259.
139. Sarapa N, Hsyu PH, Lappin G, Garner RC (2005). The application of accelerator mass spectrometry to absolute bioavailability studies in humans: Simultaneous administration of an intravenous microdose of C-14-nelfinavir mesylate solution and oral nelfinavir to healthy volunteers. *J Clin Pharmacol 45*(10), 1198–1205.
140. Europeans for Medical Progress (Agust 2005). *ANIMAL TESTING—MPs, GPs and scientists demand evaluation*. Available: http://www.curedisease.net/news/050525.shtml.
141. First human data on next generation of renin inhibitors, new drugs development in the united states (2005). *Medscape Cardiol 9*(1). Available: http://www.medscape.com/viewarticle/501272_10.
142. Boddy AV, Sludden J, Griffin MJ, Garner C, Kendrick J, Mistry P, Dutreix C, Newell DR, O'Brien SG (2007). Pharmacokinetic investigation of imatinib using accelerator mass spectrometry in patients with chronic myeloid leukemia. *Clin Cancer Res 13*(14), 4164–4169.
143. Gardner ER, Burger H, van Schaik RH, van Oosterom AT, de Bruijn EA, Guetens G, Prenen H, de Jong FA,

Baker SD, Bates SE, Figg WD, Verweij J, Sparreboom A, Nooter K (2006). Association of enzyme and transporter genotypes with the pharmacokinetics of imatinib. *Clin Pharmacol Ther 80*(2), 192–201.

144. Kil KE, Ding YS, Lin KS, Alexoff D, Kim SW, Shea C, Xu Y, Muench L, Fowler JS (2007). Synthesis and positron emission tomography studies of carbon-11-labeled imatinib (Gleevec). *Nucl Med Biol 34*(2), 153–163.
145. Ray T (March 14, 2007). FDA's "Phase 0" microdosing studies not useful preclinical tool, pharma exec says. *Pharmacogenom Reporter*.
146. http://dctd.cancer.gov/MajorInitiatives/Sep0507Phase0Workshop/workshop.htm.

# DEVELOPMENT OF ANTIMICROBIAL PEPTIDES AS THERAPEUTIC AGENTS

Robert S. Hodges, Ziqing Jiang, James Whitehurst, and Colin T. Mant

*Department of Biochemistry and Molecular Genetics, University of Colorado Denver, School of Medicine, Aurora, Colorado*

## 1 INTRODUCTION

### 1.1 Resistance to Classical Antibiotics

The great success of the pharmaceutical industry in creating new antibiotics over the past four decades has perhaps caused the general public and portions of the scientific community to become complacent about the potential for bacterial resistance. Indeed, antibiotic resistance due to the extensive clinical use of classical antibiotics has become a great concern in recent years to medical practitioners and scientists in the field of infectious disease. For instance, over the past 5 years, the rate of resistance to penicillin and cefotaxime has increased by more than 300 and 1000%, respectively (Fig. 1; http://www.cdc.gov/drugresistance/community/files/ads/resis_an.htm). Classical antibiotics work by inhibiting bacterial cell wall synthesis, protein synthesis, DNA (deoxyribonucleic acid) replication, or by modifying metabolism. Bacteria can resist antibiotics as a result of chromosomal changes (mutation or inductive expression of a latent chromosomal gene) or the exchange of genetic material via plasmids and transposons [1]. Antibiotics are rendered inactive by three major mechanisms: (1) antibiotic inactivation by destruction or modification, (2) prevention of access of the antibiotic to the target, and (3) alteration of the target site of the antibiotic [2]. Bacteria have a remarkable ability to overcome each new reagent synthesized as a potential classical antibiotic. Consequently, the current priority is the development of alternative drugs and/or the isolation of native molecules that would allow the consistent and proper control of pathogen-caused diseases [3]. Ideally, these molecules should be active over a wide range of pathogens, easily produced, and not prone to induce resistance.

### 1.2 Antimicrobial Peptides (AMPs) as Antibiotics

Clearly, the development of a new class of antibiotics with very different modes of action compared to classical antibiotics is urgently required. AMPs have emerged as promising candidates for such required novel therapeutic agents. Although the exact mode of action of AMPs has not been established, it is generally accepted that the cytoplasmic membrane is the main target of many AMPs, whereby peptide accumulation in the membrane causes increased permeability and a loss of barrier function, resulting in the leakage of cytoplasmic components and cell death. The development of resistance to membrane-active peptides whose sole target is the cytoplasmic membrane is thought to be considerably reduced compared with that of many current antibiotics, which have more specific molecular targets [4]. This prediction has been substantiated in several studies [5–7]. However, the major barrier to the use of AMPs as antibiotics is their toxicity or ability to lyse eukaryotic

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

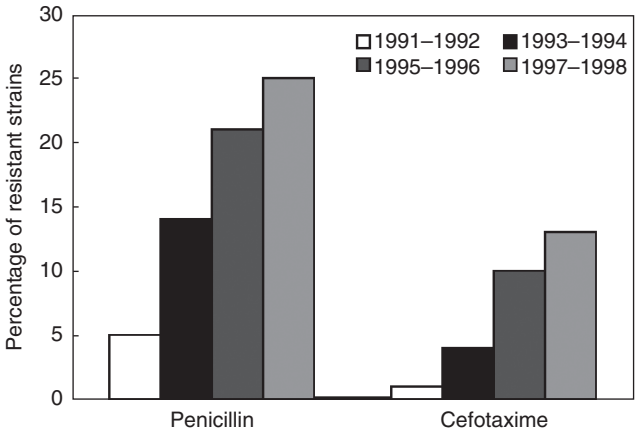


**Figure 1** Proportion of *S. pneumoniae* clinical isolates resistant to penicillin or cefotaxime from 1991–1998. (Adapted from the Center for Disease Control and Prevention website: http://www.cdc.gov/drugresistance/community/files/ads/resis_an.htm.)

cells, normally expressed as hemolytic activity (toxicity to human red blood cells). This is the main reason thus far preventing their applications as injectable therapeutics.

### 1.3 Distribution of Antimicrobial Peptides

Researchers use the term AMPs to describe gene-encoded, typically relatively short (12–100 amino acids) peptides, which have two unique features: (1) a net positive charge of +2 or greater, owing to an excess of basic amino acids (Lys, Arg) over acidic amino acids (Asp, Glu) and (2) an amphipathic nature [4, 8, 9], which segregates basic and hydrophobic residues. AMPs are found in organisms that are evolutionarily quite distant, ranging from single-cell microorganisms to insects and other invertebrates, plants, amphibians, birds, fish, and mammals, including humans [4]. Recently, such peptides were even reported to exist in human immunodeficiency virus (HIV; lentivirus lytic peptide 1 or LLP-1) [10, 11]. To date, more than 500 AMPs have been discovered [12], and at least 50% were identified in invertebrates, predominantly within insects [13]. AMPs are generally present in those parts of animals that are most likely to be in contact with environmental pathogens, that is, skin, ear, eye, tongue, trachea, lungs, gut, testes, and bone marrow [9]. AMPs have unusually broad spectra of activity, including an ability to kill or neutralize gram-negative and gram-positive bacteria, fungi (including yeasts), parasites (including planaria and nematodes), cancer cells, and even enveloped viruses like HIV and herpes simplex virus [3].

### 1.4 Diversity of Structure of Antimicrobial Peptides

Due to the diversity of these molecules, Koczulla and Bals [14] noted that it is difficult to categorize them in a generally accepted classification, although useful group classifications have been utilized for various reviews of the field [4, 13–24]. Overall, AMPs tend to be linear or cyclic peptides containing α-helical and/or β-sheet structure, together with unstructured sequence and frequently stabilized by disulfide bridges. For the purposes of the present review, we believe the most straightforward grouping of peptides is as follows: (a) α helical (helix, helix with hinge, helical with other secondary structure), (b) β sheet with one or more disulfide bonds, (c) cyclic β sheet, and (d) AMPs with no defined secondary structure. Representative examples of each group are discussed below, with specific structures shown in Figure 2.

***α-Helical AMPs*** As noted by Bulet et al. [16], the family of α-helical linear AMPs is diverse, being found in a variety of invertebrates and vertebrates, and representatives of this family may number several hundred.

*α-Helical AMPs with Uninterrupted Helical Regions* One of the most studied of α-helical AMPs is magainin, specifically magainin 1 and 2, which are linear 23-residue peptides from the African clawed toad *Xenopus laevis* (Table 1; Fig. 2). Two-dimensional nuclear magnetic resonance (2D-NMR) studies have shown that magainin 2 is unstructured in a fully aqueous environment but nearly fully helical in 25% aq. trifluoroethanol (TFE), an α-helix-inducing solvent [25]. Later studies [26] carried out two-dimensional NMR in the presence of detergent micelles, since the resulting peptide-containing micelles are regarded as more reliable for structure determinations than mixed organic solvents. Such studies showed magainin 2 to be fully helical in dodecylphosphocholine (DPC) detergent micelles and aqueous TFE solutions, confirming earlier studies. It is interesting to note that such studies were in contrast to suggestions from other researchers using solid-state NMR and Fourier transform infrared (FTIR) [27] studies that a population of magainin molecules exist in β conformation. Magainin 2, the most thoroughly studied member of the magainin family, has been shown to form pores in the membrane [28, 29]. However, recently, Ulrich and co-workers determined the location and alignment of the α-helical AMP, PGLa, a member of the magainin family, in membranes as monitored by solid-state $^{2}$H-NMR ([30] and references therein). Their very thorough structure studies show that PGLa adopts an α-helical structure upon binding to bilayers and lies flat on the

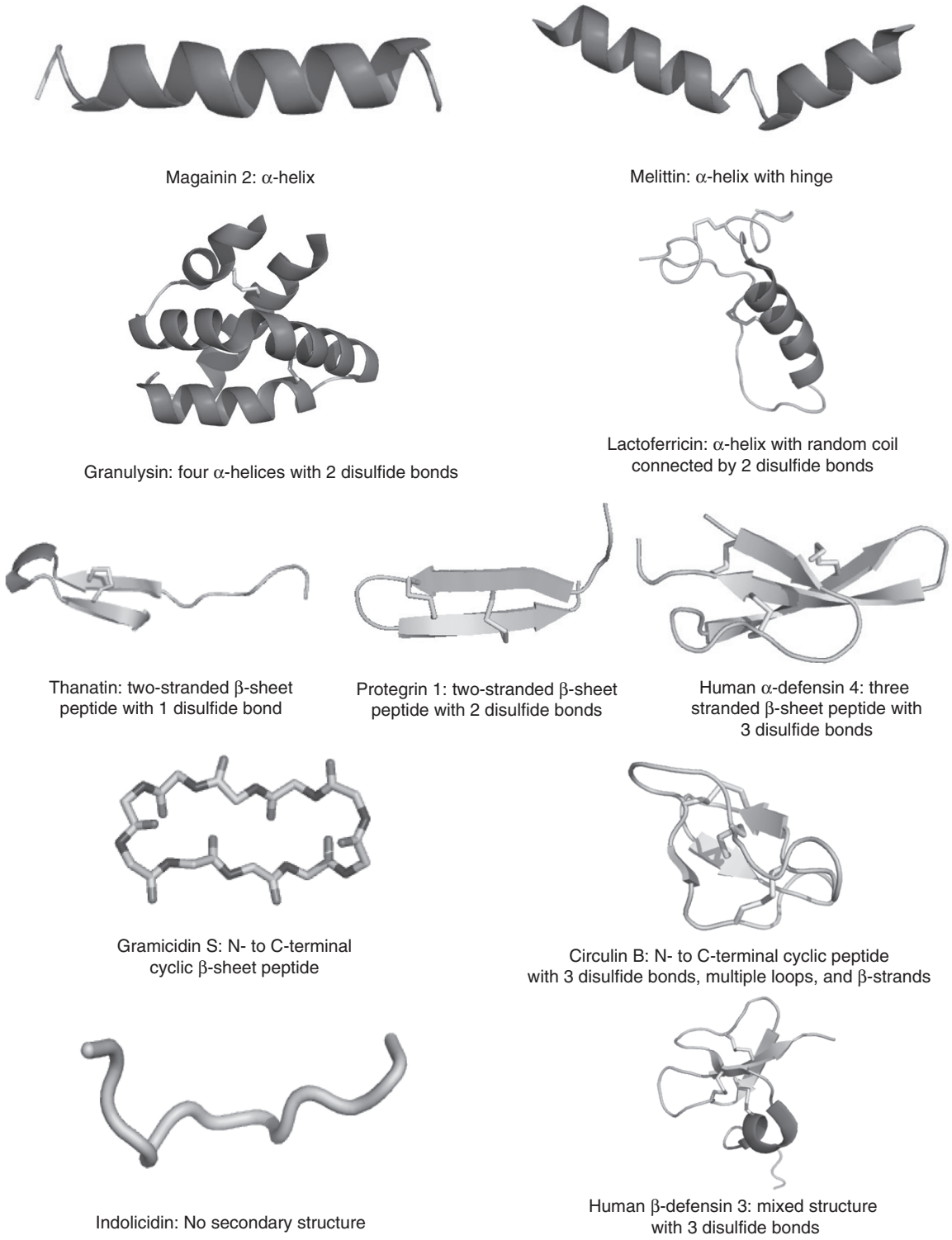


**Figure 2** Secondary structure representations of select antimicrobial peptides. Helices are colored dark gray, β sheets are light gray, disulfide bonds are off-white strings, and unstructured regions are light gray, with the exception of the gramicidin S structure. For gramicidin S, carbon is colored light gray, nitrogen is medium gray, and oxygen is dark gray. Magainin 2 (PDB code 2MAG) [26]; melittin (PDB code 2MLT) [324]; granulysin (PDB code 1L9L) [78]; lactoferricin (PDB code 1Z6V) [80]; thanatin (PDB code 8TFV) [89]; protegrin 1 (PDB code 1PG1) [90]; human α-defensin 4 (PDB code 1ZMM) [93]; gramicidin S (PDB code 1TK2); circulin B (PDB code 2ERI) [325]; indolicidin (PDB code 1G89) [110]; human β-defensin 3 (PDB code 1KJ6) [87]. Illustrations were prepared using PyMOL [326].

**TABLE 1A Representive Examples of Native Amphipathic Antimicrobial Peptides (22–27 Residues in Length)**

| Name | Source | Length | Sequence | C-Amide | Number of Pro Residues |
|---|---|---|---|---|---|
| Ascaphin-1 | Tailed frog *Ascaphus truei* | 23 | GFRDVLKGAAKAFVKTVAGHIAN | Y | 0 |
| Ascaphin-5 | Tailed frog *A. truei* | 24 | GIKDWIKGAAKKLIKTVASHIANQ | | 0 |
| BMAP-27 | Ox *Bos taurus* | 27 | GRFKRFRKKFKKLFKKLSPVIPLLHLG | Y | 2 |
| Blp-3 | Asian toad *Bombina orientalis* | 25 | GIGAAILSAGKSALKGLAKGLAEHF | | 0 |
| Chrysophsin-1 | Red sea bream *Pagrus major* | 25 | FFGWLIKGAIHAGKAIHGLIHRRRH | Y | 0 |
| Chrysophsin-2 | Red sea bream *P. major* | 25 | FFGWLIRGAIHAGKAIHGLIHRRRH | Y | 0 |
| Clavaspirin | Sea squirt, tunicate ascidian *Styela clava* | 23 | FLRFIGSVIHGIGHLVHHIGVAL | | 0 |
| Dermaseptin s4 | Sauvage's leaf frog *Phyllomedusa sauvagei* | 27 | ALWMTLLKKVLKAAAKALNAVLVGANA | Y | 0 |
| Dermaseptin s9 | Sauvage's leaf frog *P. sauvagei* | 24 | GLRSKIWLWVLLMIWQESNKFKKM | | 0 |
| Dermaseptin b1 | South American frog, boreal frog, two-colored leaf frog *Phyllomedusa bicolor* | 27 | DVLKKIGTVALHAGKAALGAVADTISQ | Y | 0 |
| eCATH-1 | Horse *Equus caballus* | 26 | KRFGRLAKSFLRMRILLPRRKILLAS | | 1 |
| eCATH-2 | Horse *E. caballus* | 27 | KRRHWFPLSFQEFLEQLRRFRDQLPFP | | 3 |
| Fowlicidin-1 | Chicken *Gallus gallus* | 26 | RVKRVWPLVIRTVIAGYNLYRAIKKK | | 1 |
| Fowlicidin-3 | Chicken *G. gallus* | 27 | KRFWPLVPVAINTVAAGINLYKAIRRK | | 2 |
| Histatin 5 | Human *Homo sapiens* | 24 | DSHAKRHHGYKRKFHEKHHSHRGY | | 0 |
| Lycotoxin I | Wolf spider *Lycosa carolinensis* | 25 | IWLTALKFLGKHAAKHLAKQQLSKL | Y | 0 |
| Magainin 1 or I | South African clawed frog *Xenopus laevis* skin | 23 | GIGKFLHSAGKFGKAFVGEIMKS | Y | 0 |
| Magainin 2 or II | South African clawed frog *X. laevis* skin | 23 | GIGKFLHSAKKFGKAFVGEIMNS | Y | 0 |
| Maximin 1 | Giant fire-bellied toad, Chinese red belly frog *Bombina maxima* | 27 | GIGTKILGGVKTALKGALKELASTYAN | Y | 0 |

| Number of Positive Charges | Number of Negative Charges | Net Charge (pH 7) | Net Charge Density (pH 7) | Number of Hydrophobic Residues | Hydrophobic Density | Amphipathicity | Segment Amphipathicity | Hemolytic Activity |
|---|---|---|---|---|---|---|---|---|
| 5 | 1 | 4 | 0.17 | 7 | 6.6 | 4.79 | 1–8 (7.29) | 50% hemolysis at >200 μ*M* |
| 5 | 1 | 4 | 0.17 | 7 | 6.8 | 4.05 | 1–8 (7.72), 9–24 (4.39) | 50% hemolysis at >200 μ*M* |
| 11 | 0 | 11 | 0.41 | 13 | 11.2 | 5.23 | 1–18 (8.30) | 3.5% hemolysis at 3 μ*M*, >80% hemolysis at 30 μ*M* |
| 3 | 1 | 2 | 0.08 | 7 | 7.0 | 4.64 | | No appreciable hemolysis (concentration unknown) |
| 6 | 0 | 6 | 0.24 | 9 | 9.3 | 3.62 | | 60% hemolysis at 1 μ*M* |
| 6 | 0 | 6 | 0.24 | 9 | 9.3 | 3.37 | | 80% hemolysis at 1 μ*M* |
| 1 | 0 | 1 | 0.04 | 12 | 11.7 | 6.08 | 1–10 (6.61) | Strong hemolysis as melittin |
| 5 | 0 | 5 | 0.19 | 11 | 9.0 | 3.59 | | 50% hemolysis at 1–1.5 μ*M*, 100% hemolysis at 3 μ*M* |
| 5 | 1 | 4 | 0.17 | 13 | 13.4 | 1.87 | 6–21 (2.66) | 50% hemolysis at 175 μ*M* |
| 4 | 2 | 2 | 0.07 | 8 | 6.1 | 2.79 | | No hemolysis at 367 μ*M* |
| 9 | 0 | 9 | 0.35 | 12 | 10.8 | 2.54 | 1–17 (4.65) | 4.6% hemolysis at 314 μg/mL |
| 6 | 3 | 3 | 0.11 | 13 | 11.6 | 3.76 | 1–6 (6.52), 7–24 (5.68) | 21% hemolysis at 314 μg/mL |
| 8 | 0 | 8 | 0.31 | 13 | 9.7 | 2.48 | 21–25 (4.17) | 50% hemolysis at 1.5 μ*M* |
| 6 | 0 | 6 | 0.22 | 13 | 9.7 | 2.67 | 22–26 (3.90) | 50% hemolysis at 9 μ*M* |
| 7 | 2 | 5 | 0.21 | 3 | 2.6 | 2.26 | | 50% hemolysis at 39 μ*M* |
| 6 | 0 | 6 | 0.24 | 9 | 9.3 | 1.17 | 1–9 (6.26), 10–25 (2.35) | 55% hemolysis at 200 μ*M* |
| 5 | 1 | 4 | 0.17 | 8 | 8.4 | 5.24 | | 3% hemolysis at 500 μg/mL, 23% hemolysis at 50 μ*M* |
| 5 | 1 | 4 | 0.17 | 8 | 8.4 | 5.15 | | 2% hemolysis at 200 μg/mL |
| 5 | 1 | 4 | 0.15 | 8 | 6.5 | 4.18 | | Little hemolytic activity at 50 μg/mL |

(*Continued*)

**TABLE 1A** ***Continued***

| Name | Source | Length | Sequence | C-Amide | Number of Pro Residues |
|---|---|---|---|---|---|
| Maximin 3 | Giant fire-bellied toad, Chinese red belly frog *B. maxima* | 27 | GIGGKILSGLKTALKGAAKELASTYLH | | 0 |
| Maximin 4 | Giant fire-bellied toad, Chinese red belly frog *B. maxima* | 27 | GIGGVLLSAGKAALKGLAKVLAEKYAN | Y | 0 |
| Maximin 5 | Giant fire-bellied toad, Chinese red belly frog *B. maxima* | 27 | SIGAKILGGVKTFFKGALKELASTYLQ | | 0 |
| Melittin B | Honeybee *Apis mellifera* venom | 26 | GIGAVLKVLTTGLPALISWIKRKRQQ | Y | 1 |
| MRP | Tago's brown frog *Rana tagoi* skin | 23 | AIGSILGALAKGLPTLISWIKNR | | 1 |
| Pandinin 2 | African scorpion *Pandinus imperator* | 24 | FWGALAKGALKLIPSLFSSFSKKD | | 1 |
| Piscidin 1 (moronecidin) | White bass *Morone saxatilis* × *Morone chrysops* | 22 | FFHHIFRGIVHVGKTIHRLVTG | | 0 |
| Piscidin 3 | White bass *M. saxatilis* × *M. chrysops* | 22 | FIHHIFRGIVHAGRSIGRFLTG | | 0 |
| Pleurocidin | Winter flounder *Pleuronectes americanus* skin mucous | 25 | GWGSFFKKAAHVGKHVGKAALTHYL | Y | 0 |
| PMAP-23 | Pig *Sus scrofa* | 23 | RIIDLLWRVRRPQKPKFVTVWVR | Y | 2 |
| Ponericin W1 | Fire ant *Pachycondylas goeldii* | 25 | WLGSALKIGAKLLPSVVGLFKKKKQ | | 1 |
| Pseudin-2 | Paradoxical frog *Pseudis paradoxa* | 24 | GLNALKKVFQGIHEAIKLINNHVQ | | 0 |

| Number of Positive Charges | Number of Negative Charges | Net Charge (pH 7) | Net Charge Density (pH 7) | Number of Hydrophobic Residues | Hydrophobic Density | Amphipathicity | Segment Amphipathicity | Hemolytic Activity |
|---|---|---|---|---|---|---|---|---|
| 4 | 1 | 3 | 0.11 | 8 | 6.8 | 3.49 | | Little hemolytic activity at 50 μg/mL |
| 5 | 1 | 4 | 0.15 | 9 | 7.1 | 3.48 | | Little hemolytic activity at 50 μg/mL |
| 4 | 1 | 3 | 0.11 | 10 | 8.7 | 3.91 | | Little hemolytic activity at 50 μg/mL |
| 6 | 0 | 6 | 0.23 | 11 | 9.2 | 4.63 | 1–13 (6.36) | 100% hemolytic at 1 μg/mL, MHC = 0.78 μ*M* |
| 3 | 0 | 3 | 0.13 | 10 | 10.1 | 6.17 | 1–13 (7.57) | 50% hemolysis at 8 μ*M* |
| 4 | 1 | 3 | 0.13 | 10 | 10.6 | 5.64 | 1–14 (5.71), 15–24 (6.29) | 51.4% hemolysis at 22.2 μ*M*, 90.5% at 44.5 μ*M* |
| 3 | 0 | 3 | 0.14 | 10 | 10.4 | 5.32 | | 50% hemolysis at approx. 30 μg/mL, 73–78% hemolysis at 100 μg/mL, 100% hemolysis at 20–50 μ*M* |
| 3 | 0 | 3 | 0.14 | 9 | 10.1 | 4.97 | | <30% hemolysis at >120 μg/mL, 15–50% hemolysis at 100 μg/mL |
| 5 | 0 | 5 | 0.20 | 8 | 7.5 | 3.08 | | Very low hemolytic activity at 100 μ*M* |
| 8 | 1 | 7 | 0.30 | 13 | 11.8 | 3.00 | 1–11 (4.07), 16–23 (6.55) | No hemolysis at 100 μ*M* |
| 6 | 0 | 6 | 0.24 | 11 | 10.0 | 4.66 | 1–13 (6.11) | Very hemolytic |
| 3 | 1 | 2 | 0.08 | 9 | 8.4 | 5.21 | 1–2 (7.47) | 50% hemolysis at 320–550 μ*M* |

**TABLE 1B Antimicrobial Activity for the Peptides (22–27 Residues in Length)**

| Name | Gram Positive | Gram Negative | Antifungal or Others | Reference |
|---|---|---|---|---|
| Ascaphin-1 | **MIC:** *S. aureus* NCTC 8325 > 50, *S. epidermidis* RP62A = 50, *E. faecalis* ATCC29212 > 100, *Streptococcus* Group B HNTCC 80130 = 13 μ*M* | **MIC:** *E. coli* ATCC 25922 = 3–6, *P. aeruginosa* ATCC 27853 = 25, *E. cloacae* NHTCC 53001 = 50, *K. pneumoniae* KK3 9904 = 50 μ*M* | **Fungi** (**MIC:** *C. albicans* ATCC 90028 > 100 μ*M*) | 175 |
| Ascaphin-5 | **MIC:** *S. aureus* NCTC 8325 = 50, *S. epidermidis* RP62A = 13, *E. faecalis* ATCC29212 > 100, *Streptococcus* Group B HNTCC 80130 = 6 μ*M* | **MIC:** *E. coli* ATCC 25922 = 6, *P. aeruginosa* ATCC 27853 = 13, *E. cloacae* NHTCC 53001 = 6, *K. pneumoniae* KK3 9904 = 13 μ*M* | **Fungi** (**MIC:** *C. albicans* ATCC 90028 > 100 μ*M*) | 175 |
| BMAP-27 | **MIC:** *S. aureus* ATCC 25923 = 2, *S. aureus* Cowan 1 = 2, *S. aureus* (methicillin R) = 4, *S. epidermidis* ATCC 12228 = 1, *B. megaterium* Bm11 = 2, *E. faecium* = 2–16, *S. agalactiae* = 2–4 μ*M* | **MIC:** *E. coli* ATCC 25922 = 1, *E. coli* ML35 = 1, *E. coli* D21 = 0.25, *S. typhimurium* ATCC 14028 = 1, *P. aeruginosa* ATCC 27853 = 1, *S. marcescens* ATCC 8100 = 2, *S. marcescens* = 1–8 μ*M*, **$IC_{50}$:** *C. trachomatis* = 80, *C. pneumoniae* > 80, *C. suis* = ≥ 80 μg/mL | **Fungi** (**MIC:** *C. neoformans* ATCC 52816 = 0.25, *C. neoformans* ATCC 52817 = 1, *C. neoformans* L1 = 0.5, *C. albicans* C2 = 32-16, *C. albicans* L21 = 8, *C. famata* (amphotericin B & itraconazole R) = 4–8, *C. famata* (amphotericin B & fluconazole R) = 2, *C. parapsilosis* = 16, > 32, *C. glabrata* = ≥ 32, *C. tropicalis* = 4–16, *C. guillermondii* = 4, *C. krusei* = 4, *C. humicola* = 8, *C. lusitaniae* = 16, *C. dubliniensis* = 32, *P. etchellsi* > 32, *P. carsonii* = 2, *R. rubra* = 0.5–4, *S. cerevisiae* = 4–8, *K. apis* = 16–32, *Aspergillus spp.* > 32, *Penicillium spp.* = ≥ 32 μ*M*) | 152, 184, 292, 328, 329 |
| Blp-3 | **LC:** *S. aureus* CowanI = 1.7, *B. megaterium* BmII = 0.3 μ*M* | **IC:** *N. meningitidis* = 20, *N. gonorrhoeae* = 20, *N. lactamica* = 20, *N. cinerea* = 20 μg/mL, **LC:** *E. coli* D21 = 1, *E. coli* D22 = 0.9, *Y. pseudotubercolosis* = 0.6, *P. aeruginosa* ATCC 15692 = 9.2 μ*M* | **LC:** *C. albicans* = 0.4 μ*M* | 37, 174 |
| Chrysophsin-1 | **MLC:** *B. subtilis* ATCC 6633 = 0.25, *L. garvieae* ATCC 49156 = 10 and *S. iniae* F-8502 = 2.5 μ*M* | **MLC:** *E. coli* wt-2 = 0.25, *V. anguillarum* ATCC 19264 = 2.5, *V. penaeicida* KHA = 10, *V. harveyi* ATCC 14126 = 2.5, *V. vulnificus* ATCC 33148 = 5, *A. salmonicida* NCMB 1102 = 10, and *P. putida* ATCC 12633 = 40 μ*M* | | 139 |

| | | | | |
|---|---|---|---|---|
| Chrysophsin-2 | **MLC:** *B. subtilis* ATCC 6633 = 0.25, *L. garvieae* ATCC 49156 = 5 and *S. iniae* F-8502 = 1.5 μ*M* | **MLC:** *E. coli* wt 2 = 0.25, *V. anguillarum* ATCC 19264 = 2.5, *V. penaeicida* KHA = 10, *V. harveyi* ATCC 14126 = 2.5, *V. vulnificus* ATCC 33148 = 5, *A. salmonicida* NCMB 1102 = 5, *P. putida* ATCC 12633 > 40 μ*M* | | 139 |
| Clavaspirin | **MEC:** *L. monocytogenes, S. aureus* = 1–3 (pH 5.5) μg/mL | **MEC:** *E. coli, P. aeruginosa* = 1–3 (pH 5.5) μg/mL | **Fungi** (**MEC:** *C. albicans* = 1–3 (pH 5.5) μg/mL) | 330 |
| Dermaseptin s4 | **MIC:** *S. aureus* B38302 = 25 μ*M*, *N. gonorrhoeae* = 10–100 μg/mL | **$IC_{50}$:** *E. coli* TG1 = 40, *E. coli* U16318 > 50 μ*M*, *Y. kristensenii* ATCC 33639, *P. aeruginosa* ATCC 27853 | **Fungi** (**MIC:** *C. neoformans* B3501 = 4.5 μ*M*), **Protozoa** (**$LC_{50}$:** *L. major* = 1.5–2 μ*M*, *L. mezicana*, *P. falciparum*), **Virus** (**$IC_{50}$:** P4-CCR5 cells by HIV-1LAI = 2 μ*M*) | 224, 331–335 |
| Dermaseptin s9 | **MIC:** *S. haemolyticus* = 12.5, *B. megaterium* = 0.4, *L. monocytogenes* = 0.4 μ*M* | **MIC:** *E. coli* 363 = 3.1, *S. typhimurium* = 3.1, *K. pneumoniae* = 12.5 μ*M* | | 33 |
| Dermaseptin b1 | **MIC:** *E. faecalis* IP103214 = 60, *S. aureus* IP76-25 > 75, *N. brasiliensis* IP16-80 = 60 μ*M* | **MIC:** *A. caviae* IP67-16T = 60, *P. aeroginusa* IP76110 > 75, *E. coli* IP76-24 = 28.3–50 μ*M* | **Fungi** (**MIC:** *C. neoformans* IP960-67 = 60, *C. neoformans* IP962-67 = 60, *C. albicans* IP884-65 = 60, *C. albicans* IP886-65 = 60, *M. canis* IP1194 = 20, *T. rubrum* IP1400-82 = 40, *T. rubrum* IP2043-92 = 40, *T. mentagrophytes* IP877-71 = 40, *A. simii* IP1063-74 = 40, *A. simii* IP903-65 = 40, *A. fumigatus* IP1025-70 = 60, *A. niger* IP218-53 = 40 μ*M*) | 336, 337 |
| eCATH-1 | **MIC:** *S. aureus* ATCC 25923 = 25, *S. aureus* (methicillin R) = 50, *S. epidermidis* ATCC 12228 = 6, *S. equinus* = 25, *B. megaterium* Bm11 = 12.5 μg/mL | **MIC:** *E. coli* ATCC 25922 = 3, *E. coli* ML35 = 3, *S. enterica* serovar *Typhimurium* ATCC 14028 = 6, *S. enterica* serovar *Enteritidis* = 6, *P. aeruginosa* ATCC 27853 = 12.5, *S. marcescens* ATCC 8100 = 12.5, *K. pneumoniae* ATCC 13883 = 3, *K. pneumoniae* SK1* = 6 μg/mL | **Fungi** (**MIC:** *C. neoformans* = 6–12.5, *C. neoformans* ATCC 90113 = 6–12.5, *C. neoformans* ATCC 52817 = 6–12.5, *R. rubra* = 3–50 μg/mL) | 338, 339 |
| eCATH-2 | **MIC:** *S. aureus* ATCC 25923 = 57, *S. aureus* (methicillin R) = 114, *S. epidermidis* ATCC 12228 = 28.5, *S. equinus* > 114, *B. megaterium* Bm11 = 14 μg/mL | **MIC:** *E. coli* ATCC 25922 = 14, *E. coli* ML35 = 14, *S. enterica* serovar *Typhimurium* ATCC 14028 = 114, *S. enterica* serovar *Enteritidis* = 114, *P. aeruginosa* ATCC 27853 ≥ 114, *S. marcescens* ATCC 8100 = 28.5, *K. pneumoniae* ATCC 13883 = 28.5, *K. pneumoniae* SK1* = 114 μg/mL | **Fungi** (**MIC:** *C. neoformans* = 21, > 114, *C. neoformans* ATCC 90113 = 28.5, > 114, *C. neoformans* ATCC 52817 = 28.5, > 114, *R. rubra* = 57, > 114 μg/mL) | 338, 339 |

(*Continued*)

**TABLE 1B** ***Continued***

| Name | Gram Positive | Gram Negative | Antifungal or Others | Reference |
|---|---|---|---|---|
| Fowlicidin-1 | **MIC:** *L. monocytogenes* ATCC 19115 = 0.8–2, *S. aureus* ATCC 25923 = 0.8–2, *S. aureus* (methicilin R) BAA-39 = 0.4–2, *S. aureus* (methicilin R) ATCC 43300 = 0.4–1 μ*M* | **MIC:** *E. coli* ATCC 25922 = 1.59–4, *E. coli* O157:H7 ATCC 700728 = 0.8, *S. typhimurium* ATCC 14028 = 0.8–2, *S. typhimurium* DT104 ATCC 700408 = 0.4–2, *S. enteritidis* ATCC 13076 = 2, *K. pneumoniae* ATCC 13883 = 0.4–1, *P. aeruginosa* ATCC 27853 = 3.18 μ*M* | | 51, 53, 58 |
| Fowlicidin-3 | **MIC:** *L. monocytogenes* ATCC 19115 = 2, *S. aureus* ATCC 25923 = 1, *S. aureus* (methicilin R) BAA-39 = 1, *S. aureus* (methicilin R) ATCC 43300 = 1 μ*M* | **MIC:** *E. coli* ATCC 5922 = 2, *S. typhimurium* ATCC 14028 = 2, *S. typhimurium* DT104 ATCC 700408 = 2, *S. enteritidis* ATCC 13076 = 2, *K. pneumoniae* ATCC 13883 = 1 μ*M* | | 51 |
| Histatin 5 | **MIC (Fragment 4-15):** *S. aureus** = 6.3–12.5 μg/mL | **MIC (Fragment 4-15):** *P. aeruginosa* ATCC 19142 = 3.1, *P. aeruginosa* ATCC 19143 = 3.1, *P. aeruginosa* ATCC 27853 = 6.3, *P. aeruginosa* ATCC 43393 = 6.3, *P. aeruginosa* ATCC 51674 = 1.6, *P. aeruginosa** = 1.6, > 100, *S. maltophilia** = 12.5–25, *A. xylosoxidans** = 6.3, *B. cepacia** = 50, *H. influenzae** = 3.1–12.5 μg/mL | **Fungi** (**MIC:** *C. glabrata* ATCC 90030 = 62.5, *C. krusei* 355 ATCC 6258 = 500, *C. neoformans* RIVM 39231 = 31.4 μg/mL, *S. cerevisiae* = 25–50 μ*M*; **Histatin 5 4-15 MIC**: *C. albicans* ATCC 10231 = 3.1, *C. albicans* ATCC 44505 = 3.1, *C. albicans* (HIV patient isolate) = 3.1, *C. glabrata* ATCC 98-2229 = 1.6, *C. glabrata** = 0.8–3.1, *C. krusei* ATCC 14243 = 1.6, *C. kefyr* ATCC 66028 = 6.3, *C. parapsilosis** = 3.1, *C. tropicalis** = 1.6–3.1 μg/mL; **$LD_{50}$:** *C. albicans* ATCC 44505 Blastospores = 2.0, Germ tubes = 9.7, Germination = 49.2 μ*M*; **$IC_{50}$:** *C. albicans* ATCC 10231 = 2.3, *C. pseudotropicalis* RIVM 4135 = 1.0, *C. albicans* ATCC 10231 = 2.4, *C. neoformans* RIVM 39231 = 0.7, *C. krusei* ATCC 6258 = 1.2, *C. parapsilosis* ATCC 90018 = 2.0, *C. glabrata* 359 = 29, *C. glabrata* B57149 (fluconazole R) = 1.5 μ*M*; **$ED_{50}$**: *C. albicans* = 6.68, *C. albicans* (azole R) = 6.4, *C. neoformans* = 3.71, *C. neoformans* (amphotericin B R) = 3.72, *C. glabrata* ATCC 90030 = 38.70, *C. glabrata* (azole R) = 85.35, *C. krusei* = 6.47, *S. cerevisiae* = 74.03 μ*M*; *C. tropicalis, C. guilliermondii*) | 38, 39, 340–350 |
| Lycotoxin I | **MEC:** *B. thuringiensis* subsp. *Israelensis* < 5.0 μ*M* | **MEC:** *E. coli* D31 = 10.0–20.0, *E. coli* DH5 = 80.0–150.0 μ*M* | **Fungi** (**MEC:** *C. glabrata* ATCC 2001 = 100.0–150.0, *C. albicans** = 100.0–200.0 μ*M*) | 351 |

| | | | | |
|---|---|---|---|---|
| Magainin 1 or I | **MIC:** *S. agalactiae* ATCC 12386 = 75, *S. faecalis* ATCC 212 > 500, *S. faecalis* ATCC 19433 > 100, *S. aureus* ATCC 29213 > 500, *S. aureus* ATCC 25923 > 500, *S. pyogenes* ATCC 19615 = 7–35 μg/mL | **MIC:** *K. pneumoniae* ATCC 13883 = 250, *P. aeruginosa* ATCC 27853 > 500, *E. coli* D31 = 1–7, *E. coli* ATCC 25922 = 7–35, *A. calcoaceticus* ATCC 19606 = 7–35, *S. sonnei* ATCC 25931 = 7–35, *E. cloacae* ATCC 23355 = 7–35, *S. flexneri* ATCC 12022 = 35–70, *C. freundii* ATCC 8090 > 100, *E. aerogenes* ATCC 13048 > 100, *S. epidermidis* ATCC 12228 > 100, *S. typhimurium* ATCC 14028 > 100, *P. vulgaris* ATCC 13315 > 100, *S. marcescens* ATCC 8100 > 100 μg/mL | **Fungi** (**MIC:** *S. cerevisiae* ATCC X2180 = 35–70 μg/mL); **Potazoa** (*A. castellanii* trophozoites 50% inhibition at 100 μg/mL for 24 h); **Virus** (**$EC_{50}$:** HSV-1 = 36.59, HSV-2 = 33.20 μ*M*) | 158, 171, 352–356 |
| Magainin 2 or II | **MIC:** *S. epidermidis* = 10, *S. aureus* = 50, *S. aureus* ATCC 25923 = 35–70, *S. aureus* ATCC 29213 = 25, *S. agalactiae* ATCC 12386 = 100, *S. faecalis* ATCC 212 > 100, *S. faecalis* ATCC 19433 = 35–70, *S. faecalis* ATCC 29212 = 400, *S. pyogenes* ATCC 19615 = 1–7, *S. aureus* (methicillin S) = 16–128, *S. aureus* (methicillin S) = 16, > 128, *S. pneumoniae* = 4–128, *S. sanguis* = 16–128, *Gemella* spp. = 8–64, *Lactococcus* spp. = 8–64, *E. faecalis* > 128, *R. equi* = 32–128 μg/mL | **MIC:** *A. caloaceticus* ATCC 19606 = 1–7, *C. freundii* ATCC 8090 = 7–35, *E. aerogenes* ATCC 13048 = 7–35, *E. cloacae* ATCC 23355 = 1–7, *E. coli* ATCC 25922 = 1–100, *E. coli* D31 = 5, *K. oxytoca* = 50, *K. pneumoniae* ATCC 13883 = 7–100, *P. aeruginosa* ATCC 27853 = 35–100, *P. aerugrnosa* ATCC 27953 = 6.25, *P. mirabilis* > 400, *P. putida* = 10, *P. vulgaris* ATCC 13315 > 100, *S. anatum* = 50, *S. epidermidis* ATCC 12228 = 7–35, *S. flexneri* ATCC 12022 = 7–35, *S. marcescens* ATCC 8100 > 100, *S. sonnei* ATCC 25931 = 1–7, *S. typhimurium* ATCC 14028 = 35–70, *E. coli* = 0.25–1, *K. pneumoniae* = 1–4, *E. cloacae* = 1–8, *P. mirabilis* = 8–32, *S. marcescens* = 2–32, *S. typhi* = 1–16, *Salmonella* spp. = 2–32, *Shigella* spp. = 2–32, *Y. enterocolitica* = 1–8, *P. aeruginosa* = 8–128, *Pseudomonas* spp. = 2–128, *S. maltophilia* = 16, > 128, *Acinetobacter* spp. = 1–8, *Brucella* spp. = 0.5–4, *P. aeruginosa* ATCC 27853 = 8, *P. aeruginosa** = 4–128 μg/mL | **Fungi** (**MIC:** *C. albicans* = 80, *S. cerevisiae* ATCC X2180 = 7–35 μg/mL, *R. solani* LR172 = 30.2 μ*M*); **Protozoa** (**IC:** *P. caudatum* = 10 μg/mL, *A. castellanii* trophozoites 50% inhibition at 100 μg/mL for 24 h, *C. parvum* = 33.2% reduction at 50 μ*M*); **Cancer cells** (**$IC_{50}$:** Bladder cancer cell = 198.1 μ*M* (WST-1 assay), 75.2 μ*M* (BrdU assay)); **Virus** (**$EC_{50}$:** HSV-1 = 22.16, HSV-2 = 19.80 μ*M*, **$ED_{50}$:** HSV-1 = 99.9, HSV-2 = 116.5 μg/mL); **Insect** (**$LD_{50}$:** *D. melanogaster* = 107.4–147.5 pmol/mg) | 26, 154, 158, 171, 263, 352, 354–364 |
| Maximin 1 | **MIC:** *S. aureus* ATCC 2592 = 19.5, *B. pyocyaneus* CMCCB 10104 = 19.5, *B. megatherium* = 19.5, *B. dysenteriae* = 2.7, *K. pneumoniae* = 9 μg/mL | **MIC:** *E. coli* ATCC 25922 = 19.5 μg/mL | **Fungi** (**MIC:** *C. albicans* ATCC 2002 = 3 μg/mL); **Virus** (**$EC_{50}$:** HIV-1 = 21.4 μg/mL); **Cancer cells** (**$IC_{50}$:** C8166 = 15.3, Molt-4 = 24.3, BIU-87 = 20.5, T24 = 35.4 μg/mL) | 168 |

(*Continued*)

**TABLE 1B** ***Continued***

| Name | Gram Positive | Gram Negative | Antifungal or Others | Reference |
|---|---|---|---|---|
| Maximin 3 | **MIC:** *S. aureus* ATCC 2592 = 3.1, *B. pyocyaneus* CMCCB 10104 = 1.5, *B. megatherium* = 0.9, *B. dysenteriae* = 0.9, *K. pneumoniae* = 3.1 μg/mL | **MIC:** *E. coli* ATCC 25922 = 0.9 μg/mL | **Fungi** (**MIC:** *C. albicans* ATCC 2002 = 15 μg/mL); **Virus** (**$EC_{50}$:** HIV-1 = 1.5 μg/mL); **Cancer cells** (**$IC_{50}$:** C8166 = 11.4, Molt-4 = 25.2, BIU-87 = 28, T24 = 18 μg/mL) | 168 |
| Maximin 4 | **MIC:** *S. aureus* ATCC 2592 = 2.7, *B. pyocyaneus* CMCCB 10104 = 2.7, *B. megatherium* = 1.5, *B. dysenteriae* = 2.7, *K. pneumoniae* = 15 μg/mL | **MIC:** *E. coli* ATCC 25922 = 2.7 μg/mL | **Fungi** (**MIC:** *C. albicans* ATCC 2002 = 15 μg/mL); **Virus** (**$EC_{50}$:** HIV-1 = 21.9 μg/mL); **Cancer cells** (**$IC_{50}$:** C8166 = 24.2, Molt-4 = 35.4, BIU-87 > 50, T24 > 50 μg/mL) | 168 |
| Maximin 5 | **MIC:** *S. aureus* ATCC 2592 = 3.6, *B. pyocyaneus* CMCCB 10104 = 7.2, *B. megatherium* = 12, *B. dysenteriae* = 12, *K. pneumoniae* = 3.6 μg/mL | | **Fungi** (**MIC:** *C. albicans* ATCC 2002 = 1.2, *A. flavus* IFFI4015 = 12, *P. uticale* IFFI2001 = 12 μg/mL); **Virus** (**$EC_{50}$:** HIV-1 = 39.8 μg/mL); **Cancer cells** (**$IC_{50}$:** C8166 = 34.4, Molt-4 > 50, BIU-87 > 50, T24 > 50 μg/mL) | 168 |
| Melittin B | **MIC:** *B. subtilis* KCTC 3068 = 1–2, *S. epidermidis* KCTC 1917 = 1, *S. aureus* KCTC 1621 = 0.5–1, *B. megaterium* Bm11 = 0.3, *M. luteus* ATCC 9341 = 0.4, *B. subtilis* ATCC 6051 = 0.4–3.1, *B. subtilis* IP 5832 = 12.5, *L. monocytogenes* NCTC 11994 = 6.3, *M. luteus* ATCC 9341 = 3.1, *E. faecalis* ATCC 19433 = 6.3, *S. aureus* ATCC 29213 = 6.3, *S. pneumoniae* ATCC 33400 = 6.3, *N. asteroides* ATCC 3308 = 6.3, *S. aureus* (methicillin R) = 1, *S. aureus* NCTC 8325 = 3, *S. citri* R8A2 = 0.39, *S. floricola* BNR1 = 3.12, *S. melliferum* BC3 = 1.56 μ*M*, *S. aureus* 8530 = 10–15, *B. subtilis* = 2–4 μg/mL | **MIC:** *E. coli* KCTC 1682 = 2, *S. typhimurium* KCTC 1926 = 2, *P. aeruginosa* KCTC 1637 = 2, *E. coli* D21 = 5, *A. calcoaceticus* Ac11 = 2–20, *P. aeruginosa* ATCC 27853 = 25, *S. typhimurium* LT2 = 13, *E. coli* ATCC 25922 = 4–25, *E. coli* DH5α = 3.9–50, *S. marcescens* ATCC 133880 > 50, *P. aeruginosa* ATCC 257853 > 50, *K. pneumoniae* ATCC 13833 = 50, *S. choleraesuis* ATCC 13311 > 50, *H. influenzae* ATCC 19418 > 50, *P. aeruginosa* (R) = 8, *A. laidlawii* A-EF22 = 0.78, *M. gallisepticum* S6 = 6.25, *M. mycoides* ssp. *mycoides* SCKH3J = 12.5 μ*M*, *E. coli* W 160-37 = 5–10, *P. putida* NCIM 2102 = 25–35 μg/mL | **Fungi** (**$IC_{50}$:** *N. crassa* = 0.8, *B. cinerea* = 3.1, *F. culmorum* = 0.8, *S. cerevisiae* = 5 μ*M*, **MIC:** *C. albicans* TIMM 1768 = 3.125, *T. beigelii* = 3.125–6.25, *S. cerevisiae* = 3.125, *P. digitatum* Conidia = 4.0, *P. digitatum* Mycelium = 8.0, *C. albicans* ATCC 90028 = 4 μ*M*) | 138, 176, 190, 289, 293, 298, 365–370 |

| | | | | |
|---|---|---|---|---|
| MRP | **MIC:** *S. aureus* NCTC 8325 = 3.5 μ*M* | **MIC:** *E. coli* ATCC 25922 = 20 μ*M* | **Fungi** (**MIC:** *C. albicans* ATCC 90028 = 19 μ*M*) | 189 |
| Pandinin 2 | **MIC:** *E. faecalis* ATCC 19433 = 2.4, *B. subtilis* ATCC 6051 = 4.8, *S. aureus* IAM 1098 = 2.4, *S. epidermidis* SAM 0020 = 4.8 μ*M* | **MIC:** *E. coli* ATCC 11775 = 19.1, *P. aeruginosa* ATCC 10145 = 38.2 μ*M* | **Fungi** (**MIC:** *C. albicans* ATCC 18804 = 19.1 μ*M*) | 76 |
| Piscidin 1 | **MIC:** *S. iniae* = 3.1, *L. garviae* (R) = 3.1, *S. aureus* = 3.1, *S. faecalis* = 3.1 μg/mL; *E. faecalis* ATCC 51299 (vancomycin R) = 5–10, *E. faecalis* ATCC 29212 = 2.5–5, *L. monocytogenes* ATCC 7616 = 2.5–5, *M. luteus* ATCC 49732 = 10–20, *S. aureus* ATCC 33591 (methicillin R) = 1.25–2.5, *S. epidermidis* ATCC 12228 = 5–10, *S. saprophiticus* ATCC 49907 = 5–10, *S. xylosus* ATCC 49148 > 20, *S. agalactiae* ATCC 12386 = 1.25–2.5, *S. bovis* ATCC 49147 = 1.25–2.5, *S. equisimilis* ATCC 12388 = 2.5–5, *S. mitis* ATCC 6249 = 1.25–2.5, *S. pneumoniae* ATCC 49619 = 1.25–2.5, *S. pyogenes* ATCC 19615 = 1.25–2.5, *S. iniae* KST740ak = 1.25–2.5, *S. iniae* KSTSi 6P = 1.25–2.5, *S. aureus* ATCC 25923 = 0–2, *B. cereus* ATCC 25923 = 0–2, *B. subtilis* = 2 μ*M* | **MIC:** *E. coli* = 3.1, *K. pneumoniae* = 3.1, *S. flexneri* = 3.1, *P. aeruginosa* = 12.5, *A. salmonicida* = 3.1, *A. hydrophila* = 0.8, *V. alginolyticus* = 3.1 μg/mL; *A. punctata* spp. *punctata* ST78-3-3 = 40–80, *A. sobria* CR79-1-1 = 10–20, *B. cepacia* ATCC 17765 > 20, *C. columnare* G4 = 40–80, *E. aerogenes* ATCC 35029 = 10–20, *E. cloacae* ATCC 35030 = 10–20, *E. coli* ATCC 25922 = 2–10, *E. coli* ATCC 35150 = 5–10, *E. coli* DH5α = 5–10, *K. oxytoca* ATCC 49131 = 5–10, *K. pneumoniae* ATCC 10031 = 2.5–5, *M. catarrhalis* ATCC 25340 = 2.5–5, *N. gonorrhea* ATCC 43069 > 20, *P. aeruginosa* ATCC 35032 = 5–10, *P. fluorescens* 56-12-10 = 10–20, *P. fluorescens* biotype I W81-11 = 10–20, *P. vulgaris* ATCC 13315 = 2–10, *S. arizonae* ATCC 13314 = 10–20, *S. choleraesuis* ATCC 14028 = 10–20, *S. flexneri* ATCC 12022 = 2.5–5, *S. marcescens* ATCC 8100 > 20, *S. sonnei* ATCC 9290 = 5–10, *S. typhimurium* ATCC 13311 = 10–20, *V. anguillarum* E-3-11 = 20–40, *V. anguillarum* SGL8542 = 40–80, *V. cholera* ATCC 10957 = 2.5–5, *Y. enterocolitica* ATCC 23715 = 2.5–5, *Y. ruckeri* SC90-2-4 = 20–40 μ*M* | **Fungi** (**MIC:** *N. crassa* = 1.56–3.12, *A. fumigatus* = 50–100, *F. oxysporum* = 0.78–1.56, *F. culmorum* = 0.39–0.78, *C. albicans* ATCC 66027 = 10–20, *C. glabrata* ATCC 66032 = 10–20, *C. lusitania* ATCC 66035 = 10–20, *C. tropicalis* ATCC 66029 = 10–20 μ*M*), **Virus** (**$IC_{50}$:** channel catfish virus (CCV) = 4, frog virus 3 (FV3) = 13 μ*M*) | 42, 43, 45, 290, 371, 372 |
| Piscidin 3 | **MIC:** *S. iniae* = 3.1, *L. garviae* (R) = 3.1–6.3, *S. aureus* = 3.1, *S. faecalis* = 12.5 μg/mL | **MIC:** *E. coli* = 3.1, *K. pneumoniae* = 6.3, *S. Flexneri* = 6.3, *P. aeruginosa* = 25.0, *A. salmonicida* = 12.5, *A. hydrophila* = 1.6, *V. alginolyticus* = 6.3 μg/mL | **Virus** (**$IC_{50}$:** channel catfish virus (CCV) = 11, frog virus 3 (FV3) = 16 μ*M*) | 45, 371 |

(*Continued*)

**TABLE 1B** ***Continued***

| Name | Gram Positive | Gram Negative | Antifungal or Others | Reference |
|---|---|---|---|---|
| Pleurocidin | **MIC:** *L. mucor* > 35.0, *B. subtilis* = 1.1–2.2, *S. aureus* = 17.7–35.0, *B. thuringiensis* ATCC 10792 = 1.3–2.5, *B. subtilis* ATCC 6633 > 80, *E. faecalis* LMG2708 = 20–40, *E. avicum* LMG2810 = 20–40, *L. corenyformis* subsp. *torquens* LMG2308 = 0.6–1.2, *L. ivanovii* Li4 = 0.6–1.2, *S. aureus* ATCC 6538 = 5, *S. epidermis* ATCC 14990 = 0.6–1.2 μ*M*, *S. aureus** = 62.5–74.7, *S. aureus* (methicillin R) = 8, *S. epidermidis** = 8–64 μg/mL | **MIC:** *E. coli* D31 = 2.2–3.3, *A. salmonicida* ATCC 49385 = 17.7–35.0, *C. aquatilis* ATCC 29551 = 2.2–4.4, *P. haemolytica* = 4.4–8.8, *S. marcescens* > 35.0, *P. aeruginosa* > 35.0, *S. typhimurium* I = 8.8–17.7, *S. typhimurium* II = 8.8–17.7, *A. salmonicida* subsp. *salmonicida* 1851/97 V 3819 = 5–10, *E. coli* ATCC 14763 = 2.5, *E. coli* DH5α = 0.6–1.2, *E. coli* pHK22 = 0.6–1.2, *P. flourescences* F283-V5 = 5, *Pseudomonas* sp. 92 = 5, *S. soneii* ATCC 11060 = 1.3–2.5, *V. anguillarum* 1190/973767 = 1.2–2.5, *V. anguillarum* HI-644 = 10–20, *V. anguillarum* HI-610 = 0.6–1.2, *V. salmonicida* HI-651 = 10–20, *Y. ruckeri* V3775 > 80 μ*M*, *P. aeruginosa** = 28–62.5, *K. pneumoniae** = 12.6–43.8, *V. anguillarum* (salmonid pathogen) = 16, *A. salmonicida* (salmonid pathogen) = 2, *S. enterica* (serovar Typhimurium)14028s = 32, *S. enterica* (serovar Typhimurium)MS7953s < 0.5, *P. aeruginosa* K799 = 8–64, *P. aeruginosa* Z61 = 4–16 μg/mL | **Fungi** (**MIC:** *C. albicans** = 15.4–24.8 μg/mL, *C. albicans* TIMM 1768 = 5, *S. cerevisiae* KCTC 7296 = 5, *T. beigelii* KCTC 7707 = 2.5 μ*M*) | 46, 147, 288, 290, 373, 374 |
| PMAP-23 | **MIC:** *B. megaterium** = 2, *S. aureus* ATCC 25923 = 4, *B. subtilis* KCTC 1918 = 1.56, *S. pyogenes* KCTC 3096 = 1.56, *S. aureus* KCTC 1621 = 3.125, *E. faecalis* KCTC 2011 = 4–8, *B. subtilis* KCTC 3068 = 2 μ*M* | **MIC:** *E. coli* ATCC 12795 = 0.4, *P. multocida* = 4.3, > 6.8, *E. coli* ML 35 = 2, *E. coli* ATCC 25922 = 4, *S. typhimurium* ATCC 14028 = 8, *P. aeruginosa* ATCC 27853 = 16, *E. coli* KCTC 1682 = 1.56, *S. typhimurium* KCTC 1926 = 1.56, *P. aeruginosa* KCTC 1637 = 1.56, μ*M* | **Fungi** (**MIC:** *C. albicans* TIMM 1768 = 2.5–20, *T. beigelii* KCTC 7707 = 2.5–20, *S. cerevisiae* KCTC 7296 = 5–20 μ*M*); **Cancer cells** (**$IC_{50}$:** Jurkat = 35, SNU 601 = 47 μ*M*); **Nematode** (*C. elegans* worm = 25 μ*M* decrease motility by 80%) | 75, 299, 375–378 |
| Ponericin W1 | **MIC:** *L. lactis* ssp. *cremoris* I 116 < 0.125, *S. aureus* CIP 677 = 2–4 μ*M* | **MIC:** *E. coli* $L^-$ = 4–8, *K. pneumoniae* CIP 8291 = 4–8, *P. putida* LMA = 4–8, *S. enterica* CIP 813 = 32–64 μ*M* | **Insect** (**$LD_{50}$:** *A. domesticus* = 244.9 μg/g) | 379 |
| Pseudin-2 | **MIC:** *S. aureus* NCTC 8325 = 80, *S. epidermidis* RP62A = 80, *Streptococcus* Group HNTCC 80130 = 80, *E. faecalis* ATCC 29212 > 80 μ*M* | **MIC:** *E. coli* ATCC 25922 = 2.5–20, *E. cloacae* HNTCC 53001 = 80, *K. pneumoniae* KK3 9904 = 40, *P. aeruginosa* ATCC 27853 = 80, *P. mirabilis* ATCC 25933 > 80 μ*M* | **Fungi** (**MIC:** *C. albicans* ATCC 90028 = 130 μmolar) | 181, 380 |

S denotes antibiotic sensitive strain; R denotes antibiotic resistant strain.

membrane surface at a peptide/lipid (P/L) ratio of 1:200 with the lysine side chains pointing up toward the water. They refer to the surface bound state as the S state. However, at higher peptide concentrations (P/L of 1:50), PGLa inserts obliquely into the membrane with a shallow tilt angle of about 125°, which they called the T state [30]. The T state is rationalized in terms of peptide dimers. These authors conclude that PGLa is adsorbed to the membrane surface, which is accompanied by a change in secondary structure from unordered to α helical and that this is a slow process on the NMR time scale. In contrast, when the local peptide concentration on the membrane surface reaches a threshold value, the PGLa switches from an S state to a tilted T state, presumably as a result of dimerization, and this process is faster on the NMR time scale. This study concludes that lysis of membranes by this peptide may occur via a different mechanism and not involve a transmembrane alignment and pore formation as observed for other peptides [30].

Frog skin is a particularly rich source of AMPs, including α-helical AMPs [31–37], of which the dermaseptin S family (8 closely related amphipathic α-helical peptides) represents an excellent example. These peptides are polycationic (Lys-rich), α helical, and amphipathic, with their polar/charged and nonpolar amino acids on opposing faces along the long axis of the helix cylinder and a conserved Trp residue at position 3. A recently identified new member of the dermaseptin S family, the 24-residue dermaseptin S9 (Table 1), is an α-helical peptide with a hydrophobic core and cationic termini (also, there is no Trp residue at position 3) [33]. Circular dichroism (CD) and 2D-NMR demonstrated this peptide to be aggregated in water but showed a monomeric nonamphipathic α-helical conformation (mainly residues 6–21) when stabilized by the addition of TFE [33]. Interestingly, even though the peptide has a continuous stretch of hydrophobic residues (residues 6–15; IWLWVLLMIW), it is still classed as amphipathic: a value of 2.73 for residues 6–15 and 2.66 for residues 6–22 (Table 1), due to the difference in relative nonpolarity of the two faces of the helix.

Histatins are a family of neutral and basic histidine-rich peptides present in human parotid and submandibular secretions [38, 39]. Examination of the solution structure of human salivary 24-residue histatin 5 (Table 1) by 2D-NMR indicated that this peptide is unstructured in aqueous solution at pH 3–8 but prefers a largely α-helical conformation in dimethyl sulfoxide solution [39]. Histatin 5 only contains three major hydrophobes (Y10, F14, and Y24) yet is still amphipathic (a value of 2.26) as a result of relative polarity difference between the two faces of the helix.

Piscidins are the first amphipathic, cationic, AMPs to be found in the mast cells of fish [40–45] (Table 1). Recent solid-state NMR measurements showed that the 22-residue piscidin 1 adopts an amphipathic α-helical conformation at water–membrane interfaces, similar to many other linear AMPs [42]. CD measurements also showed that it was unstructured in water but had a high α-helix content in micelles [40]. This was subsequently confirmed by $^{1}$H-NMR data in water and TFE. The 25-residue AMP, pleurocidin (Table 1), also found in fish [46–48], has also been shown by CD and $^{1}$H-NMR to have a random coil conformation in aqueous solution but assumes an amphipathic, α-helical structure (about 16–25% α helix) in TFE and micelles [48]. Further, from $^{15}$N relaxation studies, the helix is a rigid structure in the latter membrane-mimicking environment [48].

CD experiments performed in different solvents indicated the presence of a significant content of helical structures in organic solvent for the 29-residue AMP, ceratotoxin A, derived from the medfly *Ceratitis capitata* [49, 50]. 2D-NMR in methanol confirmed amphipathic, α-helical behavior for the 8- to 25-residue region of the peptide [50].

*α-Helical AMPs with Interrupted Helical Regions*

Cathelicidins represent a major family of host defense peptides that have been identified in fish, birds, and mammals [51–55], and three chicken cathelicidins (fowlicidins 1–3) have been recently identified [56–58] (Table 1). CD showed that the 26-residue fowlicidin 1 was largely unstructured in aqueous solution but exhibited significant α-helical content (up to 81%) in the presence of TFE or sodium dodecyl sulfate (SDS) [53]. Furthermore, 2D-NMR studies determined that fowlicidin 1 does indeed adopt an α-helical conformation but with a slight kink (or "hinge") induced by glycine (residue 16) close to the center, in addition to a short, flexible unstructured region near the N terminus [53]. It should also be noted that the α helix (residues 8–23) of the peptide is much more hydrophobic and less amphipathic than most of the mammalian cathelicidins [53]. Similar CD and 2D-NMR results were observed for the 27-residue fowlicidin 3, which adopts a predominantly α-helical structure (residues 9–25) with a slight kink (hinge) induced by glycine at position 17 [51].

Melittin (Table 1; Fig. 2) is a much studied 26-residue AMP and is the principal component of honeybee (*Apis mellifera*) venom. CD [56–58] and $^{1}$H-NMR [59] showed that the aqueous melittin monomer has no detectable secondary structure. However, the X-ray structure of tetrameric melittin crystallized from aqueous solution [60], as well as 2D-NMR melittin structures in the presence of methanol [61] and micelles [61–64], all indicate that the molecule consists of two

α-helical segments (residues 2–10 and 13–26) connected by a hinge (residues 11 and 12; threonine and glycine, respectively) to form a bent α-helical rod with the hydrophilic and hydrophobic sides facing opposite directions.

Cecropins represent a family of linear 34- to 39-residue AMPs found in the Cecropia moth [65, 66]. CD studies have indicated that cecropins have a largely random coil conformation in water alone but adopt a mainly α-helical (60–80%) conformation in the presence of membrane-mimicking organic solvents [67]. 2D-NMR studies of the 37-residue cecropin A indicated a similar structure to that of melittin (see above; Fig. 2), that is, two amphipathic α-helical regions (from residues 5 to 21 and from residues 24 to 37), with a hinge region (Gly-Pro) connecting the two helices [68]. A similar structure was assessed by 2D-NMR for 34-residue cecropin B1, a derivative of cecropin B, in a membrane-mimicking environment: two stretches of helices (residues 3–22 and residues 26–33) connected by a hinge region formed by Gly 23 and Pro 24 [69] (Table 5).

It is worth noting that, although cecropins and melittin have a similar conformational design (two helices separated by a hinge region), they have reversed polarity [65]: at neutral pH, the first 21 residues of cecropin A include eight positive and two negative charges, whereas the sequence of residues 22–37 only has a single positive charge; in contrast, the N-terminal domain of 20 amino acids in melittin only has two positive charges but residues 21–26 include four positive charges in a row. Subsequent cecropin/melittin hybrids were designed in an attempt to improve antimicrobial activity while minimizing hemolytic tendencies [70–73] (Table 5). For instance, 2D-NMR studies of a hybrid peptide formed by connecting the first 13 residues of each peptide [CA (1–13) M (1–13) = CAM (1–26)] indicated two α-helices, separated by a flexible hinge section, involving residues 4–12 and 16–26, with the latter helix being much more stable than the former [72]. In contrast, 2D-NMR studies with hybrid CA (1–8) M (1–12) indicated an α-helical structure of about three turns in the melittin domain and a flexible structure with one turn in the cecropin domain, connected by a flexible hinge region [70]. An interesting result arose from 2D-NMR studies of a cecropin/melittin hybrid made up of residues 1–8 of cecropin A and residues 1–18 of melittin [CA (1–8) M (1–18)], which indicated a β-pleated structure in water but a right-handed α-helix conformation in a membrane-mimicking environment [73]. Structural studies have also been carried out on a hybrid peptide derived from combining sequences of magainin (uninterrupted α-helix; Fig. 2) and cecropin A (two α-helices linked by a hinge region). Thus, NMR characterization of residues 1–8 of cecropin A and residues 1–12 of magainin [CA (1–8) MA (1–12)] (Table 5) indicated a hydrophobic $3_{10}$ helix of just two turns combined with one short helix at the N terminus, with a flexible hinge region (Gly-Ile-Gly) in between [70]. These researchers reported "a severely bent structure in the middle of the peptide."

PMAP-23 (pig myeloid antibacterial peptide-23; Table 1) is a 23-residue cathelicidin-derived AMP identified from porcine leukocytes [74, 75]. CD studies indicated that this peptide has no appreciable structure in aqueous buffer but exhibits appreciable α-helical structure in the presence of TFE or micelles [74]. Subsequent NMR elucidation of the peptide indicated an α-helical domain from residues 1–10 and a short α-helix from residues 17–23, these two domains being connected by a flexible region containing two proline residues (PXXP) [74].

Pandinins 1 (44 residues) and 2 (24 residues) (Table 1) are α-helical polycationic AMPs from venom of the scorpion *Pandinus imperator* [76, 77]. Both peptides showed an unordered structure in aqueous solution but considerable α-helicity in the presence of TFE or micelles [76]. One-dimensional and 2D-NMR studies in DPC micelles subsequently showed that, though both peptides are essentially helical, they differ in their overall structure: pandinin 1 consists of two distinct amphipathic α-helices separated by a coil region of higher flexibility incorporating a Pro residue; in contrast, pandinin 2 is composed of a single amphipathic α helix with a predominantly hydrophobic N-terminal sequence [76].

*α-Helical AMPs with Disulfide Bonds* Granulysin (Fig. 2) is a 74-residue basic protein produced by human cytolytic T lymphocytes and natural killer cells [78, 79]. The crystal structure of granulysin revealed a five-helix bundle with two disulfide bonds connecting helices 1 and 5 and 2 and 3 [78]. These authors also noted that the granulysin molecule is a small, hydrophobic core, helical bundle, only two helices thick, and is almost entirely amphipathic.

Lactoferricin is generated by gastric pepsin cleavage of lactoferrin [80–82]. 2D-NMR studies of 49-residue human lactoferricin, LfcinH, in a membrane mimetic solvent indicated an amphipathic helix was present between residues 14 and 29 with the remainder of the molecule being of random coil nature. Disulfide bonds are present between Cys 20 and Cys 37 and Cys 10 and Cys 46 [80]. Interestingly, the peptide is partially folded in aqueous solution, with a helical region from Pro 15 to Thr 18 and a coiled backbone to Gln 24, a turn, and then an antiparallel alignment for the remaining sequence of the peptide [80]. These authors also noted

the absence in the peptide solution structure of well-defined β structure, in addition to the helical region, reported for the intact protein crystal structure.

Defensins are small (3–5 kDa) cationic proteins with molecules stabilized by several (usually 3) disulfide bonds [83, 84] and are classified into 2 categories, α and β defensins, based on the arrangement of the disulfide bridges [85, 86]. NMR analysis of solution structures of the 3 human β defensins all exhibit similar mixed conformation tertiary structures, that is, a short helical segment preceding a 3-stranded antiparallel β sheet [87]. For the 45-residue human β-defensin 3 (HBD3) (Fig. 2), disordered structure of the first 6 residues is followed by the short helical segment of residues 10–14 packed against one face of the β-sheet, with the 3 β strands being formed by residues 17–19, 27–31, and 39–43 [87]. Disulfide bonds are present between Cys 23 and Cys 41, Cys 11 and Cys 40, and Cys 18 and Cys 33.

***β-Sheet AMPs with One or More Disulfide Bonds*** Thanatin is a 21-residue peptide that appears in the hemipteran insect *Podisus maculiventris* [88, 89]. 2D-NMR demonstrated a well-defined antiparallel β-sheet structure in water from residue 8 to the C terminus, including the single disulfide bridge between Cys 11 and Cys 18 (Fig. 2). The 2-stranded β-sheet, involving residues 8–12 and residues 17–21, is slightly twisted, with the 2 strands connected by a standard β–β hairpin turn [89].

Protegrins are a family of peptides isolated from porcine leukocytes [90–92] and composed of 16–18 residues, including 4 Cys residues forming 2 disulfide linkages [91]. 2D-NMR studies of the 18-residue protegrin 1 demonstrated a well-defined structure in solution composed mainly of a two-stranded antiparallel β-sheet (residues 4–8 and 13–17) with strands connected by a β-turn [90] (Fig. 2). Disulfide bridges are formed between Cys 6 and Cys 15 and Cys 8 and Cys 13.

Crystal structures of human α-defensins HNP4 (Fig. 2; 34-residue human α-defensin 4), HD5, and HD6 have shown that their tertiary structures are quite similar, that is, 3 β-strands arranged in antiparallel β sheet, a β-hairpin formed by strands β2 and β3, a 7- to 8-residue long loop connecting β1 and β2, and 2 termini (1–4 residues) located close to each other [93]. Three disulfide bridges also serve to stabilize the final conformation of the peptides.

***Cyclic β-Sheet AMPs*** Gramicidin S (GS) (Fig. 2) is a cationic cyclic and amphipathic decapeptide from *Bacillus brevis*. The structure of GS is known from X-ray crystallography [94] as well as $^{1}$H-NMR studies in various solvents [95, 96]: It has an antiparallel β-sheet conformation with a 2-fold symmetry, where 2 equivalent short strands are connected by β-turns consisting of D-Phe and Pro residues. Solid-state $^{19}$F-NMR was also used to study the membrane-bound structure of a $^{19}$F-labeled GS analog [97]. Such studies indicated that the analog is aligned symmetrically with respect to the membrane normal, that is, with its cyclic β-sheet backbone lying flat in the plane of the bilayer, consistent with its amphipathic character.

Circulin A and circulin B (Fig. 2) (CirA, CirB), cyclopsychotride (Cpt), and kalata are large end-to-end cyclic peptides of 29–31 amino acids from plants of the Rubiacease family [98–101]. As a group, these peptides, which contain 6 Cys residues, are reported to be the largest cyclic peptides found in nature [102]. In addition to their cyclic nature, these peptides all exhibit a cystine knot motif in their disulfide connectivity [100, 102–106], characterized by 1 disulfide bond threading through the other 2 to give a knotted motif. 2D-NMR of kalata and CirA determined that they are compact, fairly rigid structures exhibiting a global fold of distorted triple-stranded β-sheets [100, 103]. The rigidity is a result of the peptide backbone being extensively folded back on itself, as well as the cross-bracing of the disulfide bonds.

***AMPs with No Defined Secondary Structure*** Most AMP structures determined to date have included α-helical peptides and β-sheet peptides (including cyclic peptides), with few structures having been determined for other classes of cationic peptides such as extended peptides with no appreciable secondary structure. Tritrpticin is a member of the cathelicidin family and is a 13-residue peptide with 3 consecutive Trp residues (unique for AMPs), as well as 2 Pro and 4 Arg residues [107]. 2D-NMR of the peptide bound to a membrane mimetic indicates that it forms 2 adjacent turns around the 2 Pro residues, that is, a distinct fold for peptide–membrane interaction [108]. The first turn involves residues 4–7, followed by a $3_{10}$-helical turn involving residues 8–11. In addition, the hydrophobic residues are clustered together and separated from the basic Arg residues, thus resulting in an amphipathic structure.

Unlike tritrpticin, which only has a relatively weak antimicrobial activity [108], indolicidin (from bovine neutrophils [109]) has a broad activity spectrum [110, 111]. The 13-residue peptide contains the highest Trp fraction (5 out of 13 residues; 39%) ever observed for a protein sequence. The structure of indolicidin (Fig. 2) was determined using NMR methods and was shown to have a unique membrane-associated peptide structure [110], in contrast to unordered structure in aqueous solution [112]. Thus, the backbone structure, when bound to phospholipid vesicles, was extended and well-defined between residues 3 and 11, with 2 half turns at residues Lys 5 and Trp 8 (Fig. 2). The peptide had a

central hydrophobic core composed of Pro and Trp, bracketed by positively charged regions near the termini. With one exception, the Trp side chains folded flat against the peptide backbone, giving the molecule a wedge shape (Fig. 2).

### 1.5 Target Diversity of Antimicrobial Peptides

In addition to membrane permeabilization and cell lysis, several intracellular mechanisms have been suggested in recent years to explain AMP activity. Indirect evidence of nonmembrane targets consists of differential antimicrobial activity of select all-D versus all-L peptide enantiomers. AMPs kill by membrane pore formation function irrespective of chiral form. However, studies of D-apidaecin showed significant reduction in *Escherichia coli* inhibitory activity compared to the L-enantiomer and indicate an intracellular macromolecular target [113]. To date, the major identified intracellular mechanisms include inhibition of DNA synthesis [114–117], inactivation of protein synthesis [114, 116–119], and disruption of chaperone function and protein folding [120, 121]. Furthermore, many AMPs possess chemotactic properties and aid in stimulating an immune response. For a review of the immunological role of AMPs see Oppenheim et al. [122].

## 2 NATIVE AMPHIPATHIC α-HELICAL ANTIMICROBIAL PEPTIDES

Due to the enormous diversity of AMPs in structure, size, and amino acid composition, we decided to focus this review on amphipathic α-helical AMPs. We screened the literature for all native amphipathic α-helical AMPs that met the following criteria: (1) The peptides had to show broad spectrum antimicrobial activity, at least against gram-negative and gram-positive bacteria; (2) the peptides had to have a minimum calculated hydrophobicity and amphipathicity; and (3) native peptide homologs had to have significant differences, for example, in the number of positive charges, number of negative charges, number of proline residues, or number of hydrophobic residues. With this approach, we came up with a list of 136 native peptide sequences that we subsequently evaluated. The data were tabulated for the 136 peptides as follows: name; source; length; sequence; whether the peptide had a C-terminal amide; the number of Pro residues; the number of positively charged residues; the number of negatively charged residues; the net charge on the peptide at pH 7 assuming His residues are fully deprotonated; the net charge density at pH 7, which is calculated by dividing the net charge by the number of residues in the peptide (values range from 0 to 1); the number of hydrophobic residues, which includes only the major hydrophobic residues Pro, Val, Tyr, Met, Ile, Leu, Phe, and Trp; the hydrophobic density, which is the sum of the hydrophobicity values for these eight residues in the peptide divided by the number of residues in the peptide; amphipathicity of the whole native sequence and any segment of the sequence if its amphipathicity is significantly greater than the native sequence; and, finally, hemolytic activity. The second part of the table includes the antimicrobial activity data where the minimum inhibitory concentrations (MIC values) are reported in micrograms/milliliter or micromolar for each gram-positive and gram-negative organism reported. Also included in the table is antifungal activity or other activities (e.g., activity against cancer cells). Finally, the references from where we obtained the biological data are also included. The size of this table is too large to include in this review; thus, we have made the data available on our website (sc3id.ucdenver.edu/antimicrobialpeptidereview) and refer to this table as the online table in this text. We have included a representative table (Table 1) of the most abundant group of native amphipathic AMPs, that is, 22–27 residues in length and where the native peptide had measured hemolytic activity reported along with the antimicrobial activities. Thus, Table 1 includes the information for 31 peptides.

### 2.1 General Features of Amphipathic α-Helical Antimicrobial Peptides

***Length Distribution*** As shown in Figure 3, the length distribution of native amphipathic α-helical AMPs in

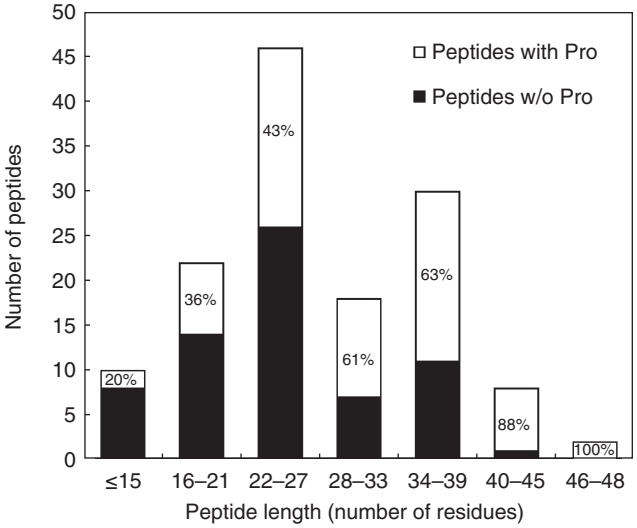


**Figure 3** Length distribution of native amphipathic α-helical antimicrobial peptides included in our database. The percent of peptides of a given length containing at least one Pro is shown in open bars.

our database range from 10 to 48 residues with 85% of the peptides ranging from 16 to 39 residues in length. Only ~7% of the peptides are ≤15 residues in length and only ~7% of the peptides are ≥ 40 residues in length. Interestingly, the largest number of peptides are in the 22- to 27-residue group (34%).

***Proline-Containing Native AMPs*** Of the AMPs in our database, ~50% contain Pro residues and, as shown in Figure 3, the number of Pro residues in the peptides is length dependent. As peptide length increases, the number of peptides containing Pro increases from only 20% of peptides ≤15 residues in length to 43% of peptides in the 22- to 27-residue group to 63% in the 34- to 39-residue group to 88% in the 40- to 45-residue group. Of the peptides containing Pro, most (95%) contain 1 or 2 Pro residues (57% contain a single Pro residue and 38% contain 2 Pro residues). This is interesting, since it is well documented that Pro can disrupt the α-helical structure by putting a kink in the polypeptide chain. In fact, in a short amphipathic model α-helical peptides (18 residues) where Pro was substituted in the center of the nonpolar face, approximately 50% of the α-helical content was lost even in a hydrophobic environment (50% trifluoroethanol, a helix inducing solvent and a mimic of the hydrophobicity of the bilayer; [123]). Pro has the lowest α-helical propensity of all 20 amino acid side chains [123]. This observation fits with the fact that the occurrence of Pro in short α-helical AMPs is less than in the longer peptides (Fig. 3). In fact, all peptides in our database containing multiple Pro residues contain greater than 23 residues.

In general, all the Pro-containing AMPs are unstructured in aqueous conditions but partial α-helical structure is induced in a hydrophobic membrane-mimetic solvent like trifluoroethanol. When NMR structures have been obtained, the Pro residues are in unstructured regions. For example, fowlicidin-1 (26 residues) has a single Pro residue at position 7 and fowlicidin 3 (27 residues) has 2 Pro residues at positions 5 and 8; regions 1–7 and 1–8, respectively, are unstructured in the presence of TFE, and the remaining portions of the molecules are α-helical in 50% TFE [51]. The role Pro residues can play in amphipathic α-helical AMPs is discussed later.

***Charge*** Of the 136 AMPs in our database, we tabulated the number of peptides containing negatively and positively charged residues. As shown in Table 2, 68% of the AMPs contained negatively charged residues and 32% of the AMPs had no negatively charged residues. The number of negatively charged residues per peptide is mainly between 1 and 3 (Table 2). On the other hand, all the AMPs in our database contain positively charged residues varying between 1 and 14, with the vast majority of peptides containing between 3 and 10 positively charged residues (Table 2). Since native AMPs vary dramatically in length, probably the most important parameters with regard to charge are negative, positive, and net charge density. Charge density is defined as the number of charges or net charge in the peptide divided by the number of residues in the peptide. Thus, charge density varies between 0 and 1. However, as shown in Figure 4, the negative charge density for most AMPs is below 0.1, which implies a maximum of 1 negatively charged residue for every 10 residues in the peptide. In comparison, the positive charge density for most AMPs is between 0.1 and 0.3, that is, 1–3 positively charged residues for every 10 residues in the peptide (Fig. 4). Perhaps the most important parameter may be the net charge density, which varies between mainly +0.05 and +0.25 (Fig. 4). The domination of positively charged residues in AMPs is understandable since these peptides have to be attracted to the negatively charged surface of the prokaryotic cell membranes. The reason why so many AMPs contain negatively charged residues remains unknown. There is a relationship between the number of negative charges, the number of positive charges, and net charge with peptide length. This is shown in the

**TABLE 2 Charge Distribution in AMPs**

| Number of negative charges[a] | Number of peptides | Number of positive charges[b] | Number of peptides | Net charge[c] | Number of peptides |
|---|---|---|---|---|---|
| **0** | 44 | | | **0** | 3 |
| **1** | 44 | **1** | 4 | **+1** | 9 |
| **2** | 26 | **2** | 8 | **+2** | 18 |
| **3** | 11 | **3** | 22 | **+3** | 24 |
| **4** | 4 | **4** | 19 | **+4** | 19 |
| **5** | 5 | **5** | 16 | **+5** | 12 |
| **6** | 1 | **6** | 12 | **+6** | 18 |
| **7** | 0 | **7** | 9 | **+7** | 12 |
| **8** | 0 | **8** | 11 | **+8** | 0 |
| **9** | 1 | **9** | 12 | **+9** | 14 |
| | | **10** | 13 | **+10** | 4 |
| | | **11** | 5 | **+11** | 2 |
| | | **12** | 3 | **+12** | 0 |
| | | **13** | 1 | **+13** | 0 |
| | | **14** | 1 | **+14** | 1 |
| | **136** | | **136** | | **136** |

[a]The number of negatively charged groups in AMPs range from 0 to 9. For example, 26 AMPs have 2 negatively charged groups.
[b]The number of positively charged groups in AMPs range from 1 to 14. For example, 22 AMPs have 3 positively charged groups.
[c]The net charge in AMPs range from 0 to +14. For example, 24 AMPs have a net charge of +3.

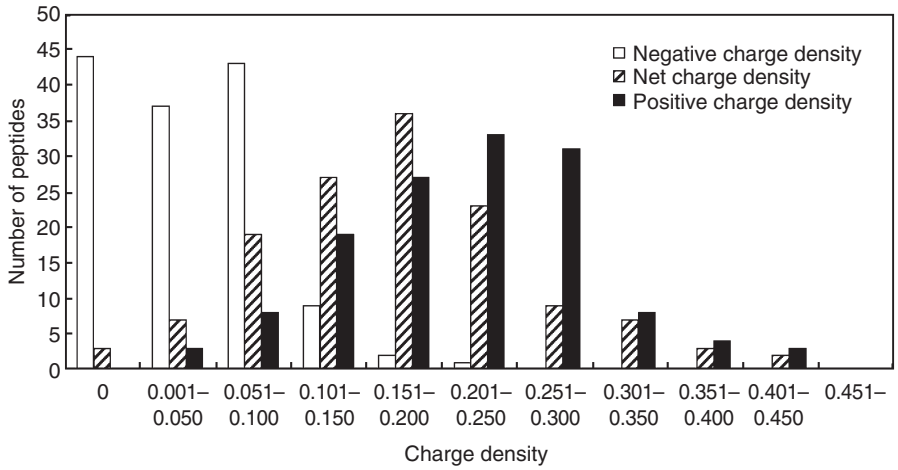


**Figure 4** Frequency distribution of peptide negative charge densities, positive charge densities, and net charge densities included in our database. Charge density is calculated as the number of charges or overall net charge divided by the number of residues.

three-dimensional plots of Figure 5. With regard to the number of positively charged residues at a peptide length of ≤15 or 16–21 residues, the greatest number of peptides in these 2 groups contain 3–4 positively charged residues. In contrast, in the 22- to 27-residue group of AMPs, the greatest number of peptides in this group have 3–6 positively charged residues; the 28- to 33-residue group is dominated by the greatest number of peptides having 7–8 positively charged residues; and the 34- to 39-residue group is dominated by the greatest number of peptides having 9–10 positively charged residues. Thus, as the peptides become longer, there is an increase in the number of positively charged residues. In contrast, except for the peptides of ≤15 residues, the greatest number of peptides in each peptide length range (from 16 to 46 residues) contain 1–2 negatively charged residues. Note that the greatest number of peptides containing 3–4 negatively charged residues is in the 34- to 39-residue length. The comparison of net charge with peptide length is shown in the bottom panel of Figure 5. The greatest number of peptides in the 22- to 27-residue group have a net charge of +3 to +4, whereas in the 34- to 39-residue group, the greatest number of peptides have a net charge of +7 to +8. Clearly, net charge is an important parameter for AMPs over the +2 to +9 range (Table 2), where a net charge of +7 to +8 dominates the peptide length of 34–39 residues and a net charge of +3 to +4 dominates the peptide length of 22–27 residues (Fig. 5).

***Hydrophobicity*** Amphipathic α-helical AMPs must have a certain minimum hydrophobicity to penetrate the hydrophobic membrane of prokaryotic cells. Thus, the relative hydrophobicity of all 20 common amino acid side chains becomes an important parameter in understanding the mechanism of action of AMPs. We have reviewed all hydrophobicity scales reported in the literature and selected the most accurate scale for the relative differences in side-chain hydrophobicity in peptides as that reported by Kovacs et al. [124]. The side-chain hydrophobicities relative to Gly are shown in Table 3. Kovacs et al. [124] showed that the intrinsic hydrophilicity/hydrophobicity of 17 amino acid side chains (Gly, Ala, Cys, Pro, Val, nVal, Leu, nLeu, Ile, Met, Tyr, Phe, Trp, Ser, Thr, Asn, and Gln) was independent of pH and buffer conditions; in contrast, the intrinsic hydrophilicity/hydrophobicity of potentially charged side chains (Orn, Lys, His, Arg, Asp and Glu) was dependent on pH and buffer conditions, including the type of salt or ion-pairing reagent (see [124] for a detailed discussion).

In the calculation of hydrophobic density for all the AMPs, we used the sum of the hydrophobicity values in a peptide for the nonpolar residues, Pro, Val, Tyr, Met, Ile, Leu, Phe, and Trp (note that Cys does not occur in the AMPs in our database). All other side chains have considerably lower hydrophobicity values compared to these nonpolar side chains (Table 3). The sum of these values was divided by the number of residues in the AMP to give the hydrophobic densities listed in Table 1. As shown in Figure 6, the hydrophobic density varies between 6.1 and 12.0 for greater than 90% of the AMPs in our database. To put this in perspective, an AMP would require one leucine residue for every two residues to give a hydrophobic density of 12. We next compared the effect of peptide length and hydrophobic density (Fig. 7). AMPs with high hydrophobic densities of 9.1–12.0 are in the 16- to 21- and

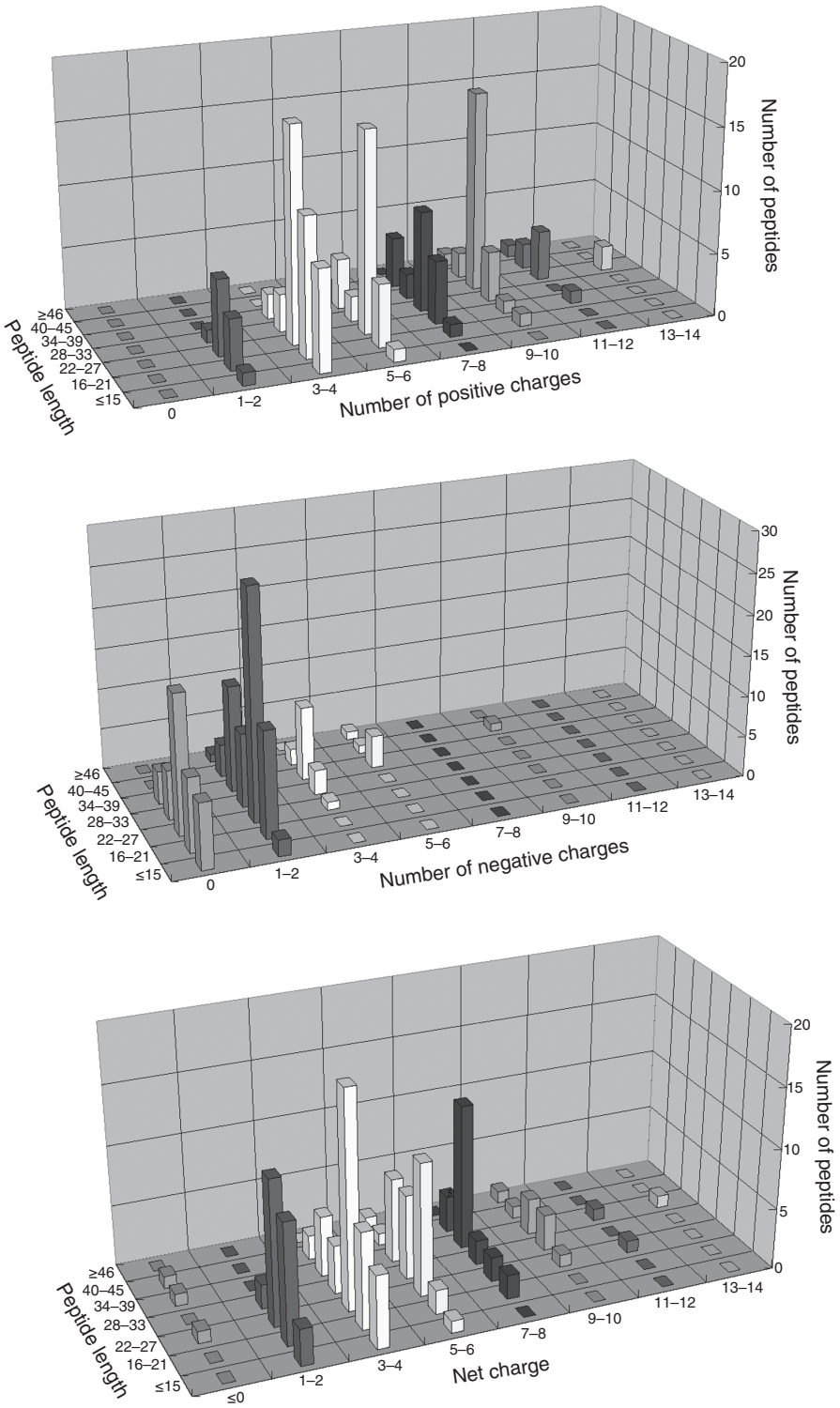


**Figure 5** Correlation of peptide length and number of peptides in our database with the number of positive charges (top panel), number of negative charges (middle), and net charge (lower panel).

**TABLE 3 Hydrophobicity and α-Helical Propensities of Amino Acid Side Chains**

| Amino Acid Side Chain[a] | Relative Hydrophobicity at pH 7[b] | Amino Acid Side Chain | α-Helical Propensity $\Delta\Delta G$ (kcal/mol)[d] |
|---|---|---|---|
| Trp | 33.0 | Ala | –0.96 |
| Phe | 30.1 | Arg | –0.90 |
| *n*-Leu[c] | 25.9 | Leu | –0.81 |
| Leu | 24.6 | Lys | –0.70 |
| Ile | 22.8 | Met | –0.67 |
| Met | 17.3 | Gln | –0.61 |
| *n*-Val[c] | 16.9 | Ile | –0.59 |
| Tyr | 16.0 | Trp | –0.49 |
| Val | 15.0 | Phe | –0.48 |
| Pro | 10.4 | Tyr | –0.43 |
| Cys | 9.1 | Cys | –0.43 |
| His | 4.7 | Val | –0.42 |
| Ala | 4.1 | Asn | –0.33 |
| Thr | 4.1 | Ser | –0.33 |
| Arg | 4.1 | His | –0.33 |
| Gln | 1.6 | Glu | –0.32 |
| Ser | 1.2 | Thr | –0.28 |
| Asn | 1.0 | Asp | –0.21 |
| Gly | 0 | Gly | 0 |
| Glu | –0.4 | Pro | 0.83 |
| Asp | –0.8 | | |
| Lys | –2.0 | | |
| Orn[c] | –2.0 | | |

[a]Side-chain hydrophobicity was determined by reversed-phase high-performance liquid chromatography of model synthetic peptides with L-amino acid substitutions at position X in the peptide sequence Ac-X-G-A-K-G-A-G-V-G-L-amide [124].
[b]Conditions, 10 m*M* phosphate buffer, pH 7, containing 50 m*M* NaCl.
[c]*n*-Leu, *n*-Val, and Orn denote norleucine, norvaline, and ornithine, respectively.
[d]Determined in 50 m*M* phosphate buffer, pH 7, containing 0.1 *M* KCl [123].

22- to 27-residue length groups. The 22 to 27-residue length has the greatest number of peptides with this high hydrophobic density. When the AMPs get longer (34–39 residues), the vast majority of the peptides of this length have lower hydrophobic densities (6.1–9.0). There are much fewer peptides of this length with a hydrophobic density greater than 9.0 or less than 6.1 (Fig. 7).

***Amphipathicity*** Amphipathic α-helical AMPs have a nonpolar face with a concentration of nonpolar amino acid residues and a polar face with a concentration of polar and charged amino acid residues. Amphipathicity of the AMPs included in this review was determined by the calculation of hydrophobic moment [125] using the software package Jemboss version 1.2.1 [126] modified to include the hydrophobicity scale shown in Table 3 [124].

As shown in Figure 8, the majority of native amphipathic α-helical AMPs have amphipathicity values ≥3.0 (~70%) and the number of AMPs with amphipathicity values between 3 and 6 is ~56%. The AMPs with the highest amphipathicity values (>6.0) were the 16- to 21-residue group. The greatest number of AMPs with amphipathicity values between 3 and 6 were the 22- to 27-residue group, and the greatest number of AMPs with amphipathicity values between 1.0 and <3 were the 34- to 39-residue group (Fig. 8). Thus, there is a general trend of decreasing peptide amphipathicity with increasing peptide length.

An example linking calculated amphipathicity with peptide sequence is shown in Figure 9. Thus, the sequence of K13, a 26-residue AMP, is shown in Figure 9 (left) presented as a helical wheel diagram. Even with a lysine residue in the center of the nonpolar face, the peptide has a high amphipathicity value of 4.92. To place this value in perspective, the sequence can be shuffled at all positions other than lysine 13 to give a minimum amphipathic value of 0.14 for a nonamphipathic sequence (Fig. 9, middle) up to a maximum value of 6.74 when residues are shuffled to optimize amphipathicity (Fig. 9, right).

***Peptide Self-Association*** Peptide self-association is an extremely important parameter in understanding the mechanism of action of amphipathic α-helical AMPs. These peptides have a very hydrophobic nonpolar face that promotes dimerization/oligomerization in aqueous solution. Since no method existed to measure peptide self-association of small amphipathic molecules with the sensitivity to measure, for example, the effect of a single substitution on the nonpolar face of these molecules, we developed a novel technique referred to as "temperature profiling in reversed-phase high-performance liquid chromatography (RP-HPLC)," based on the observation of the conformation-dependent response of peptides to temperature variations during RP-HPLC.

We have applied this technique to several types of peptides, including amphipathic cyclic β-sheet AMPs [127], monomeric α-helices and α-helices that dimerize [128], as well as α-helices that dimerize to form α-helical coiled coils [129]. Although peptides are eluted from a reversed-phase column mainly by an adsorption/desorption mechanism, even a peptide strongly bound to a hydrophobic stationary phase will partition between the matrix and the mobile phase when the acetonitrile content becomes high enough during gradient elution. The proposed mechanism of action for temperature profiling of α-helical peptides in RP-HPLC has been explained in detail by Mant et al. [128]. In summary, the

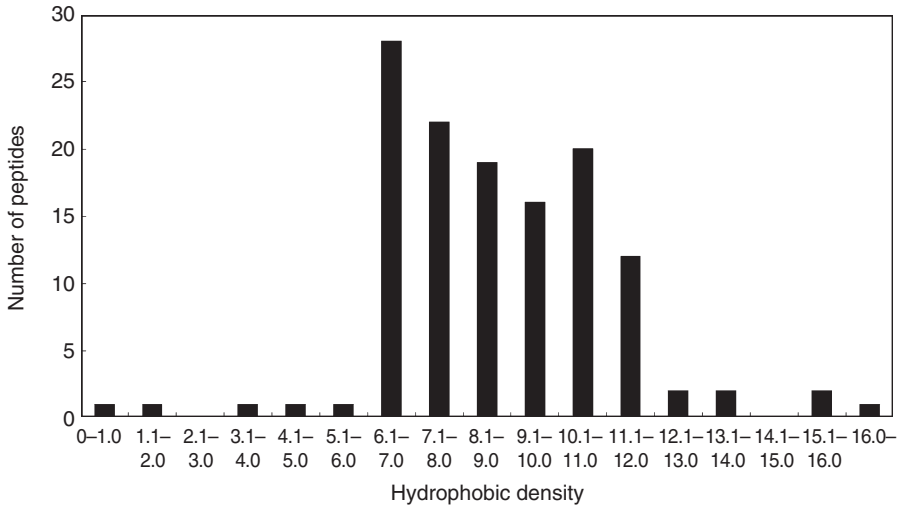


**Figure 6** Distribution of hydrophobic densities of peptides in our database. Hydrophobic density is calculated using the relative amino acid hydrophobicity scale [124] in Table 3. The hydrophobicity values for the nonpolar residues Pro, Val, Tyr, Met, Ile, Leu, Phe, and Trp is summed and the total is divided by the number of residues in the peptide to produce the density value.

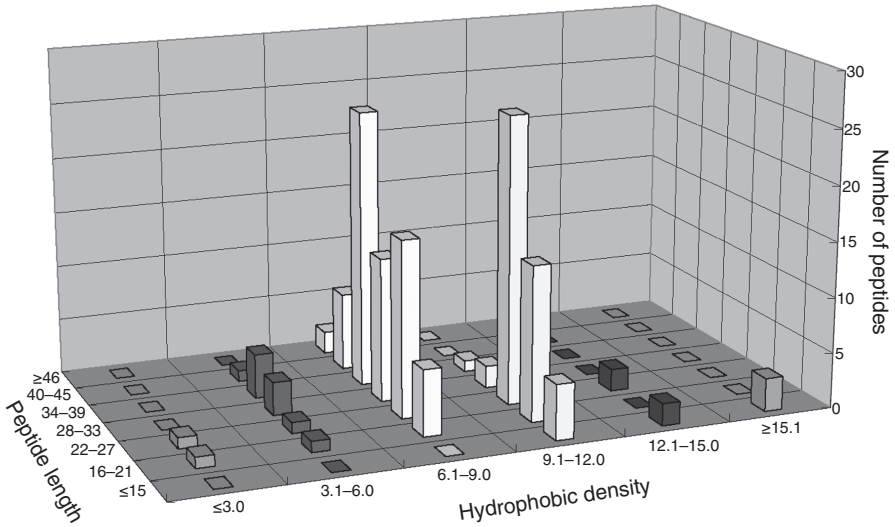


**Figure 7** Correlation of peptide length and number of peptides in our database with calculated hydrophobic density.

mechanism is based on four assumptions: (i) at low temperature, just as an amphipathic α-helical peptide is able to self-associate in aqueous solution (through its hydrophobic, nonpolar face), it will dimerize/oligomerize in solution during partitioning in RP-HPLC; (ii) at higher temperatures, the monomer–dimer equilibrium favors the monomer as the dimer is disrupted; (iii) at sufficiently high temperatures, only the monomer is present in solution; and (iv) the amphipathic α-helical peptide is always bound in its monomeric helical form by its preferred binding domain (the nonpolar face) to the hydrophobic stationary phase, that is, the peptide dimer can only be present in solution, and disruption of the dimer is required for rebinding to the reversed-phase matrix during partitioning.

Figure 10 shows the temperature profiles of the AMP, K13, shown in Figure 9, along with the peptide where position 13 is valine rather than lysine during RP-HPLC from 5 to 80°C. The maximal value of the change of peptide retention times with temperature

normalized to 5°C (Fig. 10A) and subsequently normalized to that of a control peptide with negligible secondary structure (Fig. 10B) is defined as the peptide association parameter ($P_A$) to quantify the association ability of peptide analogs in solution. Figure 10 shows that the substitution of Val by Lys dramatically reduces the ability of the peptide to self-associate. For details of this method and its value in monitoring self-association of amphipathic α-helical AMPs, see [130–133]. The $P_A$ values of peptides with nonpolar face substitutions were of the same order as their relative hydrophobicity, indicating that the hydrophobicity on the hydrophobic face of the amphipathic helix was essential for dimerization, since the dimers are formed by the binding together of the nonpolar faces of two amphipathic molecules. In general, the greater the hydrophobicity on the nonpolar face the greater the amphipathicity and the greater the association of peptides by their nonpolar faces. The importance of peptide self-association in the mechanism of action of AMPs is discussed below.

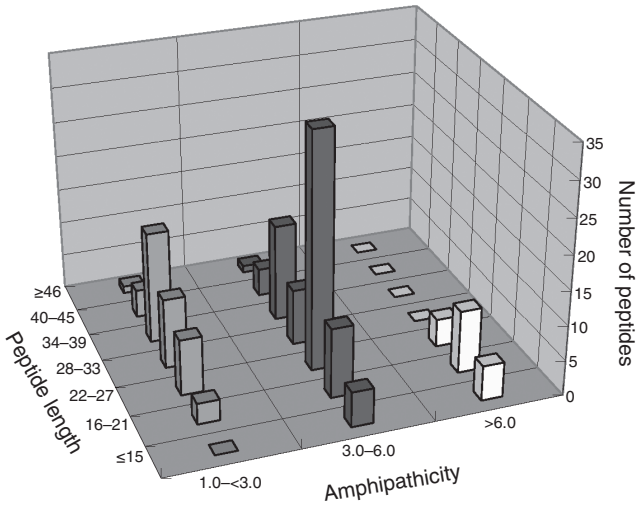


**Figure 8** Correlation of peptide length and number of peptides in our database with calculated peptide amphipathicity. Amphipathicity is derived by calculating the hydrophobic moment [125] using the Jemboss [126] software package and the hydrophobicity scale [124] shown in Table 3.

## 2.2 Biological Activity

Among our 136 representative α-helical AMPs, 53 are from amphibians, 31 are from insects, 20 are from mammals, 10 are from fish, 7 are from spiders, 6 are from scorpions, 3 are from birds, and 6 are from other organisms. All of them exhibit high levels of antibiotic activity.

***Antibacterial Activity*** Most studies on AMPs to date have been carried out on those with antibacterial activity [134], with no obvious conserved structural domains giving rise to antibacterial activity [4]. The antibacterial activities of AMPs are in the micrograms/milliliter or micromolar level. Among the 136 representative α-helical AMPs, 130 have anti-gram-negative bacterial

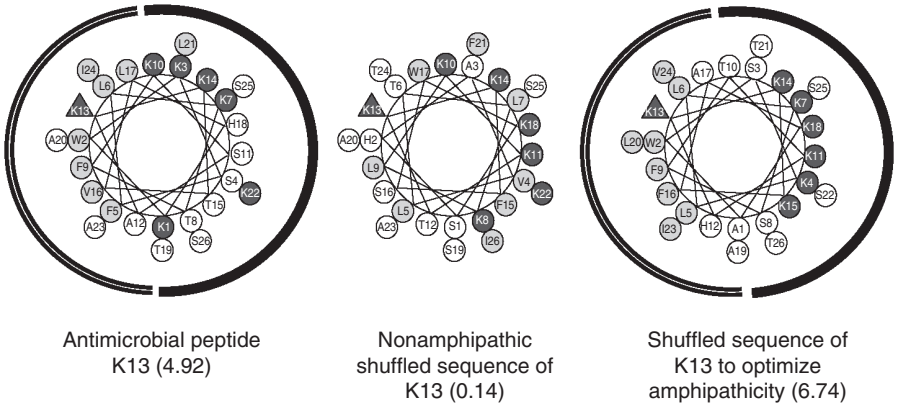


**Figure 9** Helical wheel [327] representation of the 26 residue K13 peptide. The nonpolar face is indicated by the open arc and the polar face is represented by the closed arc. Residues contributing to the polar face are circled white for Ser, Thr, and His residues. Positively charged residues are shaded dark gray and hydrophobic residues are shaded light gray. Amphipathicity values are in parentheses. (Left) Highly amphipathic native K13 peptide. (Middle) A nonamphipathic peptide constructed by shuffling the sequence at all positions except K13. (Right) K13 peptide optimized for high amphipathicity by shuffling residue positions.

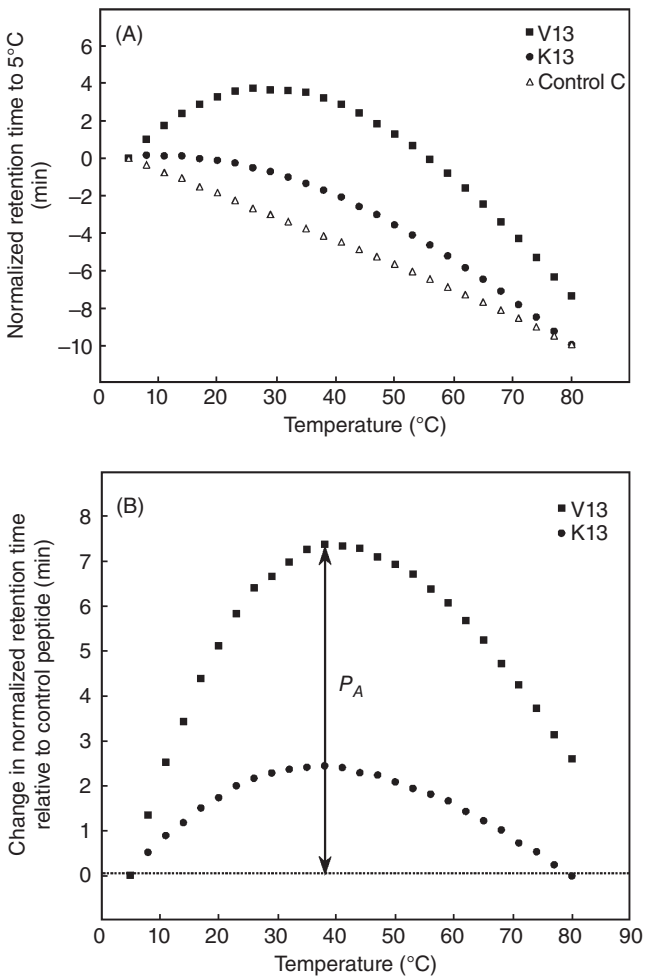


**Figure 10** Measurement of K13 and V13 peptide self-assembly using RP-HPLC temperature profiling. (A) The retention time for each peptide at a given temperature is normalized to the retention time at 5°C using the expression ($t_R^t$-$t_R^5$), where $t_R^t$ is the retention time at a given temperature and $t_R^5$ is the retention time at 5°C. (B) The retention properties of each peptide were further normalized against a control peptide with minimal secondary structure using the expression ($t_R^t$-$t_R^5$ for K13) - ($t_R^t$-$t_R^5$ for control peptide). The maximal change in retention relative to the control peptide is defined as the peptide association parameter ($P_A$).

activity reported, while 135 have anti-gram-positive bacterial activity reported (online table). *E. coli* and *Pseudomonas aeruginosa* are the commonly used gram-negative bacteria for activity assays. Among those anti-gram-negative-bacterial peptides, 119 have been proved to be efficient against *E. coli*, while 81 have activity toward *P. aeruginosa* (online table), some of which act even against clinical isolates of *Pseudomonas* [135, 136]. Citropin 1.1 [135], IsCT [137], and melittin B [138] have effects on antibiotic-resistant *P. aeruginosa* strains. Some special microorganisms were used to evaluate the activity of certain AMPs from rare sources. For example, *Vibrio anguillarum* and *Vibrio vulnificus*, gram-negative fish pathogens, were treated with chrysophsins, a peptide family from red sea bream—*Pagrus major* [139]. One hundred and ten α-helical AMPs have antibacterial activity against *Staphylococcus aureus*, the gram-positive bacterium frequently used for activity assays. Methicillin-resistant *S. aureus* (MRSA), a resistant variation discovered in 1961 [140], has evolved an ability to survive treatment with many antibiotics including gentamicin, tobramycin, netilmicin, and amikacin [141]. MRSA is especially troublesome in hospital-acquired (nosocomial) infections. Twenty-three α-helical AMPs were shown to kill MRSA effectively. Almost half of them (10 of 23) are mammalian cathelicidins (an AMP family, which share a well-conserved N-terminal segment, the cathelin domain, that is proteolytically cleaved to generate the mature, active peptide contained within the C terminus [4]) (online table). Vancomycin-resistant *Enterococcus* (VRE) is a group of bacterial species of the genus *Enterococcus* that is resistant to the glycopeptide antibiotic vancomycin. VRE was discovered in 1985 [142] and is particularly dangerous to immunocompromised individuals. VRE is also troublesome in hospital-acquired infections. Eight of 136 representative AMPs can inhibit VRE in vitro and half of them (4 of 8) are from amphibians (which have proved to be a rich source of AMPs [4]) (online table).

***Antifungal Activity*** Peptides with mainly antifungal activity (e.g., many isolated from plants) are generally rich in polar and neutral amino acids, which suggests a unique structure–activity relationship [143]. Interestingly, in a similar manner to antibacterial AMPs, no obvious conserved structural domains that give rise to antifungal activity are present [4]. Indeed, the mechanism of action of some antifungal peptides is still not clear [4]. Among 69 representative α-helical AMPs with antifungal activity, 57 of them have candidacidal activity (online table). *Candida albicans*, a diploid fungus, is a causal agent of opportunistic oral and genital infections in humans [144]. Systemic fungal infections (fungemias) have emerged as important causes of morbidity and mortality in immunocompromised patients, including AIDS (acquired immunodeficient syndrome) [145, 146]. In addition, hospital-related infections in patients have become a cause of major health concern. Seven peptides can kill the clinical isolates of *C. albicans* and 4 of them are cathelicidins (online table). For example, pleurocidin, a 25-residue α-helical AMP from the winter flounder *Pleuronectes americanus* can kill four *C. albicans* clinical isolates with MIC (minimum inhibition concentration) values ranging from 15.4 to 24.8 μg/mL [147].

Fifteen α-helical AMPs (7 from amphibians) kill some *Aspergillus* species in vitro (online table), which cause serious disease in humans and animals. The most common causes of invasive disease are *A. fumigatus* [148] and *A. flavus* [149]. Eleven and 9 α-helical AMPs can kill these species, respectively (online table). Life-threatening infections caused by the encapsulated fungal pathogen *Cryptococcus neoformans* have been increasing steadily because of the onset of AIDS and the expanded use of immunosuppressive drugs [150]. Twenty-four *C. neoformans*–killing AMPs have been reported (online table). PMAP-36 (pig myeloid antibacterial peptide-36) [151] and SMAP-29 (sheep myeloid antibacterial peptide-29) [152], cathelicidins from pig and sheep, respectively, have activity against clinical isolates of *C. neoformans.*

***Antiviral Activity*** AMP antiviral activity frequently appears to be related to the viral adsorption and entry process [153] or results from a direct effect on the viral envelope [154, 155]. Despite diverse AMP structures, many of these peptides exhibit similar antiviral modes of action [156, 157], indicating that, despite large structural differences, these peptides are able to interact with their targets perhaps due to the fact that AMPs are known to adopt amphipathic conformations intrinsic to antibacterial and antiviral activity [4]. Interestingly, although the viral target of these peptides appears to vary, the demonstrated antiviral effects are quite similar. Sixteen α-helical AMPs (10 from amphibians) displayed antiviral activity to different human viruses, including human immunodeficiency virus (HIV) and herpes simplex virus (HSV) (online table).

***Antiprotozoan Activity*** Magainin 2 was one of the first AMPs demonstrated to display antiprotozoan activity [158]. Although several AMPs possess antiprotozoan activity, structure–activity relationship studies suggest that antiprotozoan and antiviral activities are mediated by different mechanisms [159]. Indeed, it is likely that antiprotozoan activity is due to the presence of peptide motifs fundamentally different from those typical of activity against bacteria, virus, and fungi. Twelve native α-helical AMPs (7 from amphibians) showed antiprotozoa activity besides magainin 2 (online table). *Leishmania* is a genus of trypanosome protozoa and is the parasite responsible for the disease leishmaniasis (kala azar, black fever) [160]. Four of 136 α-helical AMPs exhibit inhibition activity against *Leishmania* (online table).

***Antitumor Activity*** Tumor cells and nontransformed cells may differ in membrane composition. For example, there is a 3- to 7-fold higher phosphatidylserine content in melanoma and carcinoma cell membranes compared to normal human keratinocytes [161]. Such differences in membrane composition can result in higher susceptibility of tumor cells to membrane-permeabilizing peptides [162]. Thirteen α-helical AMPs with antitumor activity have been reported, 10 of them from amphibians. The tumor cells tested are derived from a variety of different organs. For example, aurein 1.2, a 13-residue α-helical AMP from the Australian frog *Litoria aureus* was active against lung, colon, prostate, and breast cancer cells with $LC_{50}$ (minimum concentrations required to kill 50% of the cancer cells) values around 0.01 μ*M* [163].

***Other Antibiotic Activity*** Among the nine insecticidal AMPs, four of them are from spiders and four of them are from insects (online table). Besides insecticidal activity, AMPs also present additional properties [162], such as mitogenic activity [164], participation in signal transduction pathways [165], and adaptive immune responses [166].

***Hemolytic Activity*** Erythrocyte hemolysis is a commonly used in vitro assessment of AMP toxicity. The basic assay consists of exposing red blood cells (RBCs) to AMP-containing solutions for a given period of time, centrifugation to separate lysed cells, and ultraviolet (UV) measurement of hemoglobin released into the supernatant. Many variations of the assay exist, and inconsistencies in RBC source, peptide exposure time, and reporting of peptide concentrations impede comparison of AMP toxicity across structural classes and research groups. Each parameter is discussed below and recommendations are suggested for a common hemolysis assay.

*Red Blood Cell Source* A number of RBC sources including human [132], canine [167], sheep [76], and rabbit [168] have been used to assess hemolysis. Any source is acceptable for obtaining a hemolysis profile of a series of peptide analogs. However, nonhuman RBC results should not be compared to human RBC hemolysis data. Membrane lipid composition varies among mammalian species, and vastly different hemolysis results have been obtained across species [167, 169, 170]. Crowell demonstrated that the sphingomyelin content of canine, human, and sheep erythrocyte membranes (12.4, 25.0, and 53.1%, respectively) was inversely correlated with RBC vulnerability to a pore-forming peptide derived from *P. aeruginosa*, and the $EC_{50}$ (effective concentration of cytotoxin required for 50% loss of the intracellular cation) using human RBCs was approximately 52-fold that of canine RBCs and less than 25% that of sheep RBCs [170]. Similarly, analysis of the

hemolytic ability of SMAP-29 at 80 μ*M* showed only 3% lysis of sheep RBCs versus 67% lysis for human [169]. Given such drastically different results using erythrocytes from different species and the desire to develop AMPs for human clinical use, we recommend using human RBC for hemolysis assays.

*AMP Exposure Time* The length of time erythrocytes are exposed to AMPs during the hemolysis assay is the least standardized parameter of the method. Reported protocols have exposures ranging from 10 min [171] to 24 h [171, 172]. The most commonly cited times are 30 min [42, 173, 174] and 1 h [11, 175, 176]. Such short exposures provide valuable information about relative acute toxicity across a peptide series. However, higher exposure times are necessary to evaluate the longer term toxicity that could result if AMPs are not fully metabolized and cleared within 1 h in vivo. For example, time-course studies using the D-V13$A_L$ peptide showed substantially greater hemolysis when the exposure is increased from 1 to 8 h (Fig. 11, [131]), and similar examination of the L-V13$A_D$ peptide showed a dramatic shift in minimal hemolytic concentration from 250 μg/mL at 12 h [130] to 31.3 μg/mL at 18 h [131]. Therefore, we suggest hemolysis should be measured using a time-course approach extending to at least 8 h of exposure time (Fig. 11).

*Peptide Concentration* AMP concentration is routinely reported in either micromolar or micrograms/milliliter. The majority of published studies investigate and report the hemolytic ability of a series of peptide analogs containing the same number of residues and very similar molecular masses. Under such circumstances, the unit of concentration has negligible effects on data interpretation. However, a comparison of hemolytic ability of two AMPs with significantly different molecular masses should use a unit of molarity to avoid a mass bias. Consider the example of the hemolysis induced by bPaAP and bPcAP peptides (bullfrog pepsinogen A or C-derived peptide). The peptide solutions used in a hemolysis assay were reported using microgram/milliliter concentrations despite the fact that the molecular masses are approximately 1865 and 3692, respectively [177]. The mass difference produced nearly a twofold greater number of bPaAP molecules used at each assay point compared to the number of bPcAP molecules. Such a significant difference would have confounded interpretation if differential hemolysis had been observed. Therefore, we advocate the reporting of hemolysis data using a unit of molarity to describe AMP concentration.

***Cytotoxicity*** Although erythrocyte hemolysis is a useful method for initially evaluating AMP toxicity and screening peptide analogs, more extensive cytotoxicity testing using additional cell types should be performed to improve biological significance. The anticipated therapeutic site dictates the cell type most appropriate for use in cytotoxicity assays. For example, AMPs designed to treat bacterial infections in the lungs of cystic fibrosis patients should be tested using lung epithelial cells.

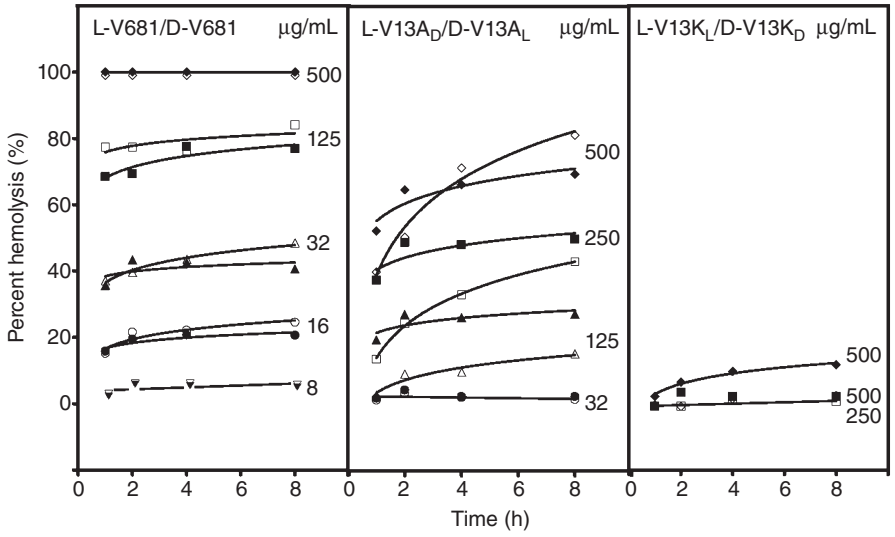


**Figure 11** Time-course study of the effect of AMP exposure time on hemolysis. Different peptide concentrations are shown using different symbols. Open symbols represent D-peptides, and closed symbols represent L-peptides.

AMPs with intended systemic circulation and parenteral administration should be tested additionally for endothelial and lymphocyte toxicity. Similar to hemolysis, no common assay conditions exist. The most popular assays for cell viability are based on metabolic activation of a dye or reduction of a tetrazolium salt by viable cells and subsequent fluorescent or UV spectrophotometric measurement. Common examples include alamarBlue (Invitrogen, USA) [51, 58], MTT (Invitrogen, USA) [168, 178, 179], and WST-1 (Roche Applied Science, USA) [180]. In an alternative method, lactate dehydrogenase (LDH) released into the media by lysed cells is used to reduce a tetrazolium salt reagent and to produce a colorimetric response proportional to the number of lysed cells. The LDH method, sold as CytoTox96 by Promega Corporation, USA, successfully measured the cytotoxicity of pseudin-2 in mouse fibroblast cells [181]. Targeted dyes such as trypan and propidium iodide (PI) [182] provide information about cell membrane condition without metabolic activation. Live cells with intact membranes exclude trypan or PI and remain unstained, unlike dead cells with compromised membranes. Trypan exclusion [172, 183] and PI staining [184] have been successfully used for AMP cytotoxicity assessment. The comparability of these assay methods has been questioned [185, 186]. Assays that require cell mixing or transfer show significantly higher cell death compared to assays that measure viability in the original seeding well [185]. Therefore, we recommend that all assays attempt to eliminate or limit cell transfer from the original growth plate. Regardless of the assay chosen, the time cells are exposed to AMPs must be considered. Reported AMP exposure time ranges from 15 min [187] to 4 days [188]. In order to calculate a relevant therapeutic index, the cytotoxicity exposure time should reflect the time scale used to assess antimicrobial activity. As many antimicrobial assays use 18–24 h [130, 189, 190], we suggest 18–24 h for cytotoxicity assessment.

***Resistance to AMPs*** Pathologically relevant bacteria constantly encounter host defense systems, including AMPs found within neutrophil granules. Not surprisingly, a wide range of microorganisms including *P. aeruginosa*, *S. aureus*, and *Neisseria gonorrhea* have developed resistance mechanisms against host AMPs and analogs. To date, characterized resistance systems include modifications to lipopolysaccharide (LPS), non-LPS membrane alterations, efflux pumps, and protease-mediated degradation.

*Alterations in LPS* LPS is the outermost component of gram-negative bacteria and the first structure encountered by an AMP. Composed of an outer O-specific side-chain oligosaccharide, a polysaccharide core, and an inner lipid A (endotoxin) particle, LPS possesses a negative charge conducive to electrostatic interactions with cationic AMPs. Reduction of the LPS negative charge offers one common mechanism of AMP resistance found in several species. *Salmonella typhimurium* and *P. aeruginosa* use the PhoP/PhoQ genetic regulatory system to induce changes in LPS charge and defend against AMPs. PhoP/PhoQ activates the PmrA–PmrB system and results in the addition of aminoarabinose to lipid A [191–194]. The subsequent charge reduction has been shown to confer resistance to polymyxin [191, 193, 194] and magainin [195]. A similar pattern of aminoaribinose addition to lipid A has been observed in *Proteus mirabilis* and induces resistance to polymyxin and protegrin [196]. Additionally, PhoP/PhoQ can activate pagP and subsequently acylate lipid A, thereby neutralizing the surface charge and providing resistance to the α-helical antimicrobial C18G peptide [197]. Other significant LPS modifications that create AMP resistance include the loss of the O-specific side chains in *Bodetella bronchioseptica* [198] and the inclusion of phosphorylcholine as a LPS component in *Haemophilus influenzae* [199].

*Other Membrane and Capsular Modifications* Several other membrane modifications have been reported to provide AMP resistance. Teichoic acids are negatively charged components of gram-positive cell walls that create favorable electrostatic potentials for cationic AMP binding. Peschel has shown that wild-type *S. aureus* can reduce the negative charge of teichoic acids by conjugation with D-alanine using a *dlt* operon-mediated process [200]. Confirming the role of the D-Ala modification in resistance, *dlt* deletion mutants showed marked magainin susceptibility in contrast to wild-type strains [200]. *S. aureus* also reduces the negative membrane charge by addition of the positively charged amino acid L-lysine to phosphotidylglycerol, and mutants incapable of lysine addition show increased susceptibility to defensins, magainin II, and HNP-1 [201, 202]. Another AMP defense is invoked by *Klebsiella pneumonia.* In response to polymyxin, *Klebsiella* increases the capsule polysaccharide thickness, limits the antimicrobial agent's access to the cell membrane, and decreases susceptibility [203].

*Efflux Pumps* Bacterial efflux pumps can move damaging antibiotics, dyes, and detergents across the cell membrane, and in cases of gram-negative bacteria these agents are also transported across the outer membrane. See Nikaido for a detailed mechanistic review [204]. *Yersinia enterocolitica* actively removes polymyxin using the RosA/RosB efflux pump [205]. Similarly, *N. gonorrhea*'s resistance to the human α-helical AMP, LL-37,

has been demonstrated to be dependent on the *mtr* efflux pump [206].

*Protease-Mediated Degradation* Extracellular proteases capable of cleaving and inactivating AMPs have been reported in several bacterial species. *E. coli* and *P. mirabilis* express the ZapA metalloprotease that has been shown to degrade LL-37 and diminish antimicrobial activity [207]. The metalloprotease autolysin has also been found to inactivate LL-37 and confers resistance in *S. aureus* [208]. In addition to the LPS modifications mentioned previously, induction of the PhoP/PhoQ regulatory system increases PgtE protease activity in the outer membrane of *S. typhimurium* and renders the bacteria resistant to LL-37 and C18G [209].

*Overview of Resistance to AMPs* Despite these means of generating AMP resistance, the practical clinical relevance remains unclear. Repetitive passaging of 27 clinical isolates of *S. aureus* in the presence of pexiganin, a magainin analog, at subinhibitory concentrations showed less than a twofold increase in MIC, whereas similar exposure to the small-molecule antibiotic mupirocin produced an eightfold reduction in susceptibility [6]. Similarly, Steinberg reported difficulty producing methicillin-resistant *S. aureus* (MRSA) or *P. aeruginosa* strains resistant to protegrin-1 using more than 10 passages at 50% MIC [5]. *These results suggest that although resistance is possible, it is by no means widespread or easily developed, and AMPs remain viable candidates for therapeutic development.*

## 3 MECHANISM OF ACTION

### 3.1 Cell Membrane

In order to understand the mechanism of action of AMPs, a thorough knowledge of the structure of the major target cell membrane is necessary. Thus, bacterial membranes are negatively charged with lipids bearing phospholipid headgroups: phosphatidylglycerol (PG), cardiolipin (CL), or phosphatidylserine (PS). In contrast, mammalian membranes are enriched in zwitterionic phospholipids (neutral in net charge): phosphatidylethanolamine (PE), phosphatidylcholine (PC), or sphingomyelin (SM) [210, 211]. Most of such anionic lipids of mammalian membranes are sequestered on the cytoplasmic side of the membrane while being exposed to the external medium with microbial membranes [17]. In addition, according to Matsuzaki, the presence of cholesterol, a major constituent of mammalian cellular membranes, can reduce the activity of AMPs by stabilizing the lipid bilayer or directly interacting and neutralizing them [212].

### 3.2 Models of Membrane Disruption

As previously described, the main target of AMPs is the cytoplasmic membrane. However, their interaction is nonspecific, an observation supported by the following evidence: (1) AMPs have broad-spectrum activity; (2) their speed of action is often within minutes in vitro; (3) both D- and L-peptide enantiomers of many AMPs have activity that indicates that there is no stereoselective interaction with a chiral enzyme, lipid, or protein receptor [213]. Generally, AMP action is almost universally dependent upon interaction with and permeabilization of the target cell membrane, with peptide net positive charge, hydrophobicity, and flexibility having been discovered as the driving physical forces [4, 214–216]. Regardless of the time required, or the specific antimicrobial mechanism, four specific and consecutive steps must occur in order to induce bacterial killing: attraction, attachment or self-promoted uptake, permeabilization, and membrane leakage.

***Attraction*** Attraction is considered to be the electrostatic interactions between peptides' cationic residues and negatively charged groups in the outer bacterial envelope. These interactions are effective over relatively long distances. Lipoteichoic acids, teichuronic acid, and amino acid carboxyl groups in the multilayered peptidoglycan [217] are the main source of negative charges of gram-positive bacterial membranes [4]. For gram-negative bacterial membranes, negative charges arise from phosphate groups within the LPS [4]. As noted previously, LPS can be subdivided into three specific regions: the O-antigen, the core, and lipid A. Lipid A molecules are negatively charged, and the outer membrane leaflet structure is held together in part by the binding of divalent cations. LPS, in particular lipid A, plays an important role in the pathophysiology of gram-negative bacterial sepsis [17].

***Attachment or Self-Promoted Uptake*** Once close to the microbial surface, AMPs must traverse capsular polysaccharides (LPS or outer membrane) before they can interact with the cytoplasmic membrane of gram-negative bacteria; on the other hand, AMPs have to traverse capsular polysaccharides, teichoic acids, and lipoteichoic acids in order to interact with the membrane of gram-positive bacteria [218]. On the surface of gram-negative bacteria, AMPs first remove or displace $Ca^{2+}$ or $Mg^{2+}$, which destabilizes the outer membrane [4]. AMPs then disrupt noncovalent bridging interactions between LPS molecules in a process driven by hydrophobic interactions, thus altering the ordered supramolecular arrangement of this leaflet, and leading to a disturbance of the outer membrane structure [4]. This

action possibly involves folding of the AMPs into a membrane-associated structure, which is different than the aqueous structure [213]. Many linear AMPs only adopt amphipathic structures when they are in contact with membranes, membrane-mimicking environments [25, 171], or when they are aggregated [219]. This transition of linear to structured peptide produces about 50% of the energy needed for membrane association [213, 220].

***Permeabilization*** For most AMPs, permeabilization of microbial membranes is an early and necessary step for killing microbes [221–223]. The molecular mechanism of membrane perturbation and disruption clearly depends on a number of parameters, for example, the nature of the peptides, temperature, environmental conditions, and peptide-to-lipid ratio [221]. In Huang's two-state model [213], he called the state observed at low peptide (P)/lipid (L) ratio the S state (functionally inactive, peptide embedded into the lipid head group), and the state observed at P/L ratio above the threshold P/L* the I state (functionally active, peptide inserted into the bilayer), for both α-helical and β-sheet peptides. He proposed that the action of AMPs is effected by the transition from the S state to the I state. Given the same peptide concentration, cells with a low P/L* threshold will be lysed while the cells with a high P/L* threshold may survive. Self-assembly of peptides was also observed for natural and synthetic AMPs [224–226].

To date, there are four major models (barrel-stave model, toroidal-pore model, carpet model, and aggregate Model; Table 4, Fig. 12) to explain the mechanism of permeabilization. All of them have two common features: (1) amphipathicity as a key characteristic and (2) parallel orientation on the surface of the membrane as a first step [4].

**TABLE 4 Summary of Permeabilization Models**

| Model Name | Orientation to the Membrane | Pore Size | Example |
|---|---|---|---|
| Aggregate model | No particular orientation | — | Polyphemusin |
| Toroidal-pore model (wormhole) | Perpendicular | Varies | Magainins, melittin, LL-37, protegrin-1, MSI-78 |
| Barrel-stave model (helical-bundle) | Perpendicular | Consistent | Alamethicin, gramicidin S |
| Carpet model | Parallel | Varies | Ovispirin, dermaseptin S, cecropin, melittin, caerin 1.1 |

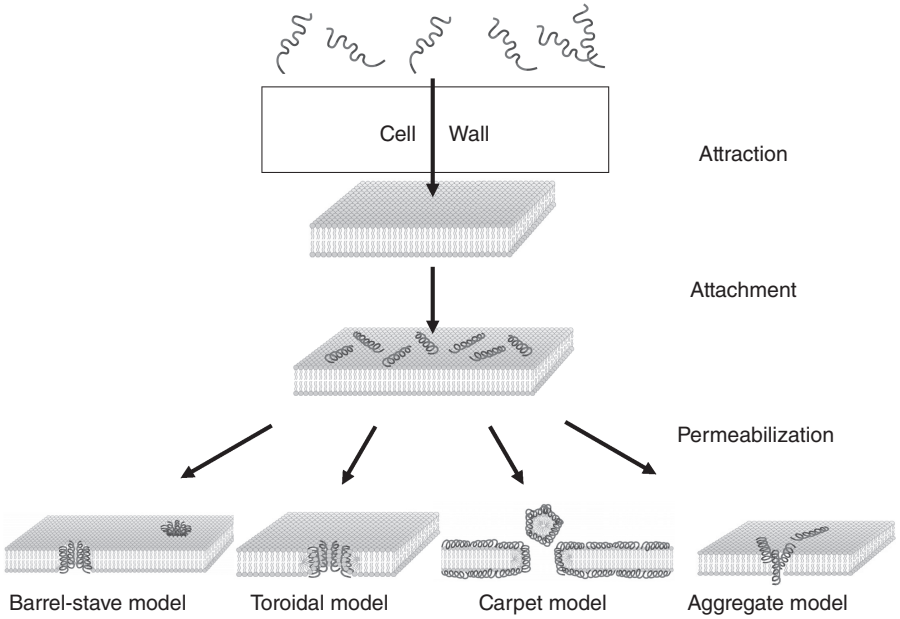


**Figure 12** Models of membrane disruption by antimicrobial peptides. (Adapted from Brogden [218].)

*Barrel-Stave or Helical-Bundle Model* This model was originally based on alamethicin, a 20-residue α-helical antimicrobial peptaibol from the fungus *Trichoderma viride* [227, 228]. In this model, peptides insert in a perpendicular orientation to the plane of the bilayer, forming the "staves" in a "barrel"-shaped cluster, with the hydrophilic regions of the peptides facing the lumen of the pore (point inward producing an aqueous pore) and the hydrophobic regions interacting with the lipid bilayer [4, 210, 218]. A disadvantage of this model is that it cannot explain the variety of pore sizes observed [4], although this model does appear to apply to cytotoxic peptides (toxins) such as pardaxin [217]. However, a recent hypothesis provides an explanation for this problem. When bound peptide reaches a threshold concentration, peptide monomers self-aggregate and insert deeper into the hydrophobic membrane core. The pore is a dynamic structure, and, hence, the continued recruitment or loss of peptide monomers results in a relatively polydisperse pore size [210]. α-Helical peptides require a minimal length of about 22 amino acids in order to traverse the lipid bilayers; alternatively, 8 amino acids are required if the peptide adopts a β-sheet structure [229]. Although proposed more than a decade ago, experimental evidence still supports this mechanism of membrane permeabilization for gramicidin S, a 10-residue cyclodecapeptide from gram-positive bacterium *B. brevis* [230].

*Toroidal-Pore or Wormhole Model* This model was first suggested by Matsuzaki and colleagues based on the well-known AMP magainin 2, a 23-residue α-helical peptide from the frog *X. laevis* [28, 29]. In this model, AMPs insert in a perpendicular orientation to the plane of the bilayer, with the hydrophilic regions of the peptides associating with the phospholipid head groups and the hydrophobic regions associating with the lipid core [4]. In this process, peptide and lipid together form well-defined pores and the membrane also curves inward to form a hole [210]. The lipids in these holes tilt from the lamellar normal and connect the two leaflets of the membrane, thus forming a continuous bend from top to bottom in the fashion of a toroidal hole. The pore is lined by both the peptides and the lipid head groups, which are likely to screen and mask cationic peptide charges [218]. Water-filled cavities, resembling "wormholes," are present [221]. It should be noted that the toroidal model differs from the barrel-stave model in that the peptides are always associated with the lipid head groups even when they are perpendicularly inserted into the lipid bilayer [218]. Other examples for this model are melittin, a 26-residue α-helical AMP from the honeybee *A. mellifera* [231], and LL-37, a 37-residue α-helical AMP from humans [232].

*Carpet Model* This model was proposed for the first time to describe the mode of action of dermaseptin I, a 34-residue α-helical AMP from the frog *Phyllomedusa sauvageii* [233]. In this model, AMPs aggregate parallel to the lipid bilayer, coating local areas in a "carpet"-like fashion. It is thought that this results in a detergent-like activity at a given threshold concentration, with subsequent formation of micelles and membrane pores [4]. Stepwise mechanisms involved in this model are: (1) preferential binding of the peptide monomers to the phospholipid head groups [234]; (2) alignment of peptide monomers on the surface of the membrane such that their hydrophobic faces are embedded relatively shallowly within the hydrophobic region of the membrane, and their positive charges are directed toward the hydrophilic region [233]; and (3) disintegration of the membrane by the peptide through disruption of the bilayer curvature [234]. The presence of negatively charged lipids helps in the formation of a dense peptide carpet, as they help to reduce the repulsive electrostatic forces between positively charged peptides [221]. There is more in vitro evidence for the carpet model than in vivo observations. Despite differences in secondary structure, the overall shape of each AMP is roughly cylindrical with hydrophobic side chains on one side of the cylindrical axis and hydrophilic side chains on the other. The minimum free energy configuration is for a single peptide adsorbing on the bilayer's hydrophilic–hydrophobic interface with the cylindrical axis more or less parallel to the plane of the bilayer [213]. Equilibrium between monomeric peptides and peptide aggregates that result in concentration-dependent effects also must be taken into account [221, 235]. This mechanism has also been proposed for ovispirin, a rationally designed α-helical AMP based on the sheep cathelicidin SMAP-29 [236].

*Aggregate Model* This model, which has some similarity to the toroidal-pore model, was first developed by Wu and colleagues based on the study of all four structural classes of AMPs [237]. In this model, AMPs reorient to span the membrane as an aggregate with micellelike complexes of peptides and lipids, but without adopting any particular orientation [4]. The advantage of this model is that it can explain both membrane permeabilization (generating channels with a variety of sizes and forms) and translocation across the bilayer. These are known to occur for several peptides [4], such as polyphemusin, an 18-residue β-hairpin AMP from the horseshoe crab *Limulus polyphemus* [238].

***Membrane Leakage*** Membrane leakage may involve: (1) membrane depolarization [217]; (2) leakage of essential metabolites [217]; (3) loss of compositional

specificity in the membrane due to an increased rate of phospholipid flip-flopping [29] and, for gram-negative bacteria, exchange with the inner leaflet of the outer membrane [239]; (4) translocation of the peptides to the cytoplasmic side of the membrane [29, 240], where they can interfere with important cellular mechanisms and/or trigger enzymes that cause peptidoglycan autolysis [241]; and (5) a passage of ions lowering the proton gradient and destroying the membrane potential, thus stopping adenosine triphosphate (ATP) production and all cellular metabolism and, hence, causing cell death [213]. The formation of pores affects cellular respiration, deprives sensitive organisms of their source of energy by disrupting the electrochemical gradient across free-energy transducing membranes, and results in increased water and ion flow across the membrane, concomitant with cell swelling and osmolysis [221]. Characteristics that may determine the rate of leakage are membrane charge, membrane curvature, and other factors [17].

***Methodology*** A variety of methods [218] have been employed to study the mechanism of action of AMPs including: (1) microscopy (confocal laser-scanning microscopy [242], scanning and transmission electron microscopy [243], and immunoelectron microscopy [244]); (2) studies with model membranes (X-ray crystallography [245], NMR spectroscopy [26, 246], FTIR spectroscopy [247, 248], Raman [249], and fluorescence [250]); (3) fluorescent dyes (measured by the release of internal fluorescent-labeled dextran, immunoglobulins, calcein, or other probes [251–255]); (4) ion channel formation (the ability of AMPs to attach to and penetrate the bilayer is measured by the conductivity of an electrical current generated through the subsequently formed pore [256–258]); (5) circular dichroism and orientated circular dichroism [130, 137, 259]; (6) solid-state NMR spectroscopy [260, 261]; (7) neutron diffraction [262–265]; (8) lamellar X-ray diffraction [137, 266]; (9) synchrotron-based X-ray scattering [267]; (10) isothermal titration calorimetry [268, 269] and (11) differential scanning calorimetry [270].

## 4 STRUCTURE-ACTIVITY RELATIONSHIPS IN AMPHIPATHIC α-HELICAL ANTIMICROBIAL PEPTIDES

### 4.1 Hybrid Amphipathic α-Helical Antimicrobial Peptides

Cecropin A, a 37-residue peptide from moths, was the first insect antibacterial peptide to be reported [66]. Other peptides with potent antimicrobial activity are the 26-residue melittin B from the honeybee *A. mellifera* and the 23-residue magainin-2 from the frog *X. laevis*. To develop a wider spectrum of activity and improve potency without undesirable cytotoxic effects, researchers began a program of designing hybrid peptides of cecropin A and either magainin-2 or melittin B [271–279]. Representative examples of hybrid peptides are shown in Table 5. Interestingly, melittin B was extremely hemolytic (100% hemolysis at 1 μg/mL), whereas the hybrid peptide consisting of the N-terminal basic region of cecropin A followed by the hydrophobic N terminal of melittin B referred to as cecropin A (1–8) melittin B (1–18) showed no hemolysis (Table 5). Shorter hybrids, for example, cecropin A (1–8) melittin B (1–12) still retained hemolytic activity on human red blood cells under physiological conditions [274] (Table 5). However, substitution of Thr 18 and Thr 19 with lysine residues caused a dramatic reduction in hemolytic activity while providing a twofold higher antifungal activity ([274], Table 5). Figure 13 shows a helical wheel presentation of the hybrid peptides as one continuous helix and as two α-helical segments because of the hinge region in the GIG segment between cecropin A 1–8 and melittin 1–18, melittin 1–12, or magainin-2 (1–12) (to be discussed below). The amphipathicity is substantially higher when considered as two segments for these hybrid peptides. For example, the amphipathicity of cecropin A (1–8) melittin (1–12) is 2.07 when calculated as one continuous α-helix, whereas the amphipathicity for each segment is 6.51 and 5.59 for residues 1–9 and residues 10–20, respectively. The values in brackets in Table 5 denote the calculated average of the amphipathicities for the two segments (N-terminal helical segment and the C-terminal helical segment). These results suggest that the true amphipathicity is much greater when considered as two segments joined by a flexible linker. We have selected a series of cecropin A 1–8 magainin 2 (1–12) hybrids that show significant structure activity relationships [280]. As shown in Figure 13, the helical wheel diagram places Ser 16 in cecropin A (1–8) magainin 2 (1–12) at the interface between the hydrophilic/polar face and the nonpolar face of the C-terminal amphipathic α helix. The effects on hemolytic activity of substitution of Ser 16 by a series of residues with increasing hydrophobicity (A < V < L < F < W) is shown in Table 5 (see Table 3 for the relative hydrophobicity differences of these residues). A large hydrophobe at position 16 can form $i$ to $i + 3$ and $i$ to $i + 4$ interactions with F13 and F20 to increase the hydrophobic surface area. As expected, the hemolytic activity increases with increasing hydrophobicity of the substituted residue (0% hemolysis for Ser and Ala at position 16 to 20.4% hemolysis at 100 μM for S16 W) (Table 5). Interestingly, the analog S16A, K12L compared to S16A increases hemolytic activity from

0% for S16A to 94.3% hemolysis at 100 μg/mL for S16A, K12L. S16A, K12L increases dramatically the hydrophobic surface area on the nonpolar face (Fig. 13), again demonstrating the effect of increasing hydrophobicity on hemolytic activity. All these peptides revealed a random coil structure in aqueous buffer but adopted α-helical structure in the presence of hydrophobic environments of TFE and SDS micelles. The three hydrophobic analogs at position 16 (Leu, Phe, and Trp) had potent antibacterial and antitumor activity but also concomitant with hemolytic activity [278]. The replacement of GIG by GPG in the hinge region of cecropin–melittin or cecropin–magainin hybrids had no effect on hemolytic activity but caused a drastic decrease in both antibacterial and antitumor activity [278] (Table 5). This suggests the importance of the conformation of the hinge region on these activities. On the other hand, replacing GIG in the hinge region with Leu in the S16L analog dramatically increased hemolysis from 11.8 to 88.6% hemolysis at 100 μ*M*. Replacement of GIG with Ser lowered hemolytic activity to 50% at 100 μ*M*, while replacement of GIG with Pro lowered the hemolytic activity to 3.6% at 100 μ*M* and retained both good antibacterial activity and antitumor activity. Thus, the GIG sequence and Pro must result in similar conformations of the hinge region.

### 4.2 De novo Designed Amphipathic α-Helical Antimicrobial Peptides

We have included in Table 6 representative examples of de novo designed amphipathic α-helical AMPs that emphasize key points about structure–activity relationships. The role of hydrophobicity/amphipathicity is clearly demonstrated by comparing the 12-residue peptides KAAAKWAAKAAK (peptide 1) and KLLLKWLLKLLK (peptide 2), where 7 alanine residues are substituted by 7 leucine residues [281]. In aqueous conditions, peptide 1 adopts a random-coil structure, whereas peptide 2 adopts substantial α-helical structure. When these two peptides were added to PE/PG vesicles, only the Leu-containing peptide adopted a predominantly α-helical structure. These results show the importance of hydrophobicity on the nonpolar face of a potential amphipathic α-helix to maintain α-helical structure. The amphipathicity also dramatically increases from 2.99 to 5.82 by increasing the hydrophobicity from Ala to Leu residues. The Ala peptide is essentially inactive against gram-negative and gram-positive organisms, whereas the Leu peptide dramatically enhances antimicrobial activity. The consequence of increased antimicrobial activity is a dramatic increase in hemolytic activity from 0 to 100% hemolysis at 100 μ*M* (Table 6). This demonstrates the importance of hydrophobicity and helical structure on antimicrobial and hemolytic activity, but peptide 2 shows no selectivity between prokaryotic and eukaryotic cells. Comparison of peptide 2 and peptide 3, where 4 D-Leu residues are substituted for 4 L-Leu residues, results in maintaining antimicrobial activity but significantly reducing hemolytic activity (<20% hemolysis at 100 μ*M*). The 2 peptides have the same inherent hydrophobicity; however, the mixing of D- and L-amino acid residues distorts the α-helical structure and amphipathicity. This suggests that the amphipathic α-helix is critical for disrupting eukaryotic membranes but not critical for activity in prokaryotic cells. The diastereomeric form of the peptide with a distorted helix in a hydrophobic environment may still be able to orient the hydrophobic face toward the hydrophobic part of the membrane and elicit antimicrobial activity [281]. Similarly, the 15-residue peptide (peptide 4) consisting of only Leu and Lys residues is extremely amphipathic (8.51) with excellent broad-spectrum activity against gram-positive and gram-negative organisms but no selectivity between prokaryotic and eukaryotic cells (100% hemolysis at 100 μ*M*) (Table 6) [282]. However, the incorporation of 5 D-amino acid residues (peptide 5) for the same L-amino acid residues results in excellent broad-spectrum activity and no hemolysis. In other words, distortion of α-helical structure has created outstanding selectivity. The same result is shown by comparing peptides 6 and 7 where peptide 6 has poor antimicrobial activity and is very hemolytic (80% hemolysis at 100 μ*M*) while peptide 7 with the incorporation of 4 D-Leu residues for 4 L-Leu residues results in removing hemolytic activity completely and dramatically enhancing antimicrobial activity [283]. Overall, these results suggest that AMPs can have high hydrophobic and positive charge density with excellent antimicrobial activity and cell selectivity as long as α-helical structure is distorted by incorporation of D-amino acid residues. The incorporation of D-amino acid residues also prevents peptide oligomerization and keeps the peptide monomeric. As discussed later, these results have implications for the mechanism of action of AMPs. Peptides 8–14 in Table 6 show a second approach to increasing cell selectivity by incorporation of a single Pro residue. Peptide 8 is the most amphipathic α-helix shown in Table 6 (9.70) with excellent antimicrobial activity but lacking cell selectivity (80% hemolysis at 50 μ*M*) [176]. Interestingly, substituting a Pro residue on the nonpolar face for a single Leu residue completely removes the hemolytic activity (peptide 9, Table 6). Similar results were observed when comparing peptide 10 with peptides 11 and 12 [284]. However, substituting a Pro residue for a Lys residue on the polar face of peptide 10 had no effect on reducing hemolytic activity (peptide 13). Hemolytic activity was reduced by using

**TABLE 5A Representative Examples of Hybrid Amphipathic Antimicrobial Peptides**

| Lead Peptide | Substitution | Length | Sequence | C-Amide | Number of Pro Residues | Number of Positive Charges |
|---|---|---|---|---|---|---|
| Cecropin A | | 37 | KWKLFKKIEKVGQNIRDGIIKAGPAV AVVGQATQIAK | Y | 1 | 9 |
| Melittin B | | 26 | GIGAVLKVLTTGLPALISWIKRKRQQ | Y | 1 | 6 |
| Cecropin A (1–8) + melittin B (1–18) | (CEME) | 26 | KWKLFKKIGIGAVLKVLTTGLPALIS | Y | 1 | 6 |
| Cecropin A (1–8) + melittin B (1–12) | | 20 | KWKLFKKIGIGAVLKVLTTG | Y | 0 | 6 |
| Cecropin A (1–8) + melittin B (1–12) | GIG→GPG | 20 | KWKLFKKIGPGAVLKVLTTG | Y | 1 | 6 |
| Cecropin A (1–8) + melittin B (1–12) | T18K, T19K | 20 | KWKLFKKIGIGAVLKVLKKG | Y | 0 | 8 |
| Cecropin A (1–8) + melittin B (1–18) derivative | (V681) | 26 | KWKSFLKTFKSAVKTVLHTALKAISS | Y | 0 | 7 |
| Cecropin A (1–8) + melittin B (1–18) derivative | (V681: V13K) | 26 | KWKSFLKTFKSAKKTVLHTALKAISS | Y | 0 | 8 |
| Magainin 2 | | 23 | GIGKFLHSAKKFGKAFVGEIMNS | Y | 0 | 5 |
| Cecropin A (1–8) + magainin 2 (1–12) | | 20 | KWKLFKKIGIGKFLHSAKKF | Y | 0 | 8 |
| Cecropin A (1–8) + magainin 2 (1–12) | GIG→GPG | 20 | KWKLFKKIGPGKFLHSAKKF | Y | 1 | 8 |
| Cecropin A (1–8) + magainin 2 (1–12) | S16A | 20 | KWKLFKKIGIGKFLHAAKKF | Y | 0 | 8 |
| Cecropin A (1–8) + magainin 2 (1–12) | S16V | 20 | KWKLFKKIGIGKFLHVAKKF | Y | 0 | 8 |
| Cecropin A (1–8) + magainin 2 (1–12) | S16L | 20 | KWKLFKKIGIGKFLHIAKKF | Y | 0 | 8 |
| Cecropin A (1–8) + magainin 2 (1–12) | S16F | 20 | KWKLFKKIGIGKFLHFAKKF | Y | 0 | 8 |
| Cecropin A (1–8) + magainin 2 (1–12) | S16W | 20 | KWKLFKKIGIGKFLHWAKKF | Y | 0 | 8 |
| Cecropin A (1–8) + magainin 2 (1–12) | S16A, K12L | 20 | KWKLFKKIGIGLFLHAAKKF | Y | 0 | 7 |
| Cecropin A (1–8) + magainin 2 (1–12) | S16L, GIG→L | 18 | KWKLFKKILKFLHLAKKF | Y | 0 | 8 |
| Cecropin A (1–8) + magainin 2 (1–12) | S16L, GIG→S | 18 | KWKLFKKISKFLHLAKKF | Y | 0 | 8 |
| Cecropin A (1–8) + magainin 2 (1–12) | S16L, GIG→P | 18 | KWKLFKKIPKFLHLAKKF | Y | 1 | 8 |

[a]The amphipathicity is calculated in two segments, residues 1–9 and residues 10–18 or 10–20 or 10–26. The amphipathicity value for the two segments was averaged and reported in brackets. The average value divided by the total number of residues in the peptides gives the amphipathicity per residue.
[b]Minimum peptide concentration that produces hemolysis.
[c]Maximum peptide concentration that results in no hemolysis.

| Number of Negative Charges | Net Charge (pH 7) | Charge Density (pH 7) | Number of Hydrophobic Residues | Hydrophobic Density | Amphipathicity | Amphipathicity per Residue | Hemolytic Activity |
|---|---|---|---|---|---|---|---|
| 2 | 7 | 0.19 | 13 | 7.4 | 1.84 | 0.05 | 13% hemolysis to sheep RBC at 580 μg/mL |
| 0 | 6 | 0.23 | 11 | 9.2 | 4.63 | 0.18 | 100% hemolytic at 1 μg/mL, MHC[b] = 0.78 μ*M* |
| 0 | 6 | 0.23 | 13 | 11.3 | 2.08 (6.28) | 0.08 (0.24) | 0% hemolysis to sheep RBC at >600 μ*M* |
| 0 | 6 | 0.30 | 9 | 10.6 | 2.07 (6.05) | 0.10 (0.30) | 14.1% hemolysis at 200 μg/mL, 15.5% hemolysis at 100 μ*M*, 51.0% hemolysis at 100 μg/mL |
| 0 | 6 | 0.30 | 9 | 10.0 | 2.08 (6.28) | 0.10 (0.31) | 0% hemolysis at 100 μ*M* |
| 0 | 8 | 0.40 | 9 | 10.6 | 2.42 (6.35) | 0.12 (0.32) | 0.5% hemolysis at 50 μg/mL, 1.0% hemolysis at 200 μg/mL |
| 0 | 7 | 0.27 | 9 | 8.5 | 5.55 | 0.21 | MHC[c] = 7.8 μg/mL |
| 0 | 8 | 0.31 | 8 | 7.9 | 4.92 | 0.19 | MHC[c] = 250 μg/mL |
| 1 | 4 | 0.17 | 8 | 8.4 | 5.15 | 0.22 | 2% hemolysis at 200 μg/mL |
| 0 | 8 | 0.40 | 8 | 10.9 | 3.28 (6.75) | 0.16 (0.34) | 0% hemolysis at 100 μg/mL |
| 0 | 8 | 0.40 | 8 | 10.3 | 3.22 (6.28) | 0.16 (0.31) | 0% hemolysis at 100 μ*M* |
| 0 | 8 | 0.40 | 8 | 10.9 | 3.39 (6.75) | 0.17 (0.34) | 0% hemolysis at 100 μ*M* |
| 0 | 8 | 0.40 | 9 | 11.7 | 3.83 (6.82) | 0.19 (0.34) | 0.9% hemolysis at 100 μ*M* |
| 0 | 8 | 0.40 | 9 | 12.0 | 4.16 (6.94) | 0.21 (0.35) | 11.8% hemolysis at 100 μg/mL |
| 0 | 8 | 0.40 | 9 | 12.4 | 4.48 (7.02) | 0.22 (0.35) | 15.3% hemolysis at 100 μ*M* |
| 0 | 8 | 0.40 | 9 | 12.6 | 4.61 (7.08) | 0.23 (0.35) | 20.4% hemolysis at 100 μ*M* |
| 0 | 7 | 0.35 | 9 | 12.1 | 3.88 (6.21) | 0.19 (0.31) | 94.3% hemolysis at 100 μg/mL |
| 0 | 8 | 0.44 | 9 | 13.6 | 3.81 (7.96) | 0.21 (0.44) | 88.6% hemolysis at 100 μ*M* |
| 0 | 8 | 0.44 | 8 | 12.2 | 4.16 (7.08) | 0.23 (0.39) | 50.4% hemolysis at 100 μ*M* |
| 0 | 8 | 0.44 | 9 | 12.8 | 3.98 (7.40) | 0.22 (0.41) | 3.6% hemolysis at 100 μ*M* |

**TABLE 5B Antimicrobial Activity for Hybrid Peptides**

| Lead Peptide | Substitution | Gram Positive | Gram Negative | Antifungal or Others | Reference |
|---|---|---|---|---|---|
| Cecropin A | | **LC:** *B. megaterium* Bm11 = 0.51–0.7, *B. subtilis* Bs11 = 6.4–61, *B. thuringiensis* Bt11 > 80, *S. faecalis* AD-4 = 15, *S. faecalis* DS16 = 74, *M. luteus* Ml11 = 1.4–4.6 μ*M*, **MIC:** *B. subtilis* KCTC 3068 = 1.56, *S. pyogenes* KCTC 3096 = 6.25, *S. aureus* KCTC 1621 < 50 μ*M* | **LC:** *E. coli* D21 = 0.19–0.4, *E. coli* D31 = 0.21, *E. cloacae* β12 = 0.4, *S. marcescens* Db10 = 4.2, *S. marcescens* Db1108 = 2.5, *S. marcescens* Db1121 = 4.4, *S. marcescens* Db11 > 127, *S. marcescens* Db1140 > 127, *S. marcescens* Strain 122 = 3.2, *P. aeruginosa* OT97 = 1.5–4.8, *X. nematophilus* Xn21 = 1.4, *K. pneumoniae* = 0.33 μ*M*, **MIC:** *E. coli* KCTC 1682 = 1.56, *S. typhimurium* KCTC 1926 = 1.56, *P. aeruginosa* KCTC 1637 = 3.13 μ*M*, **MIC50:** *R. monacensis* = 12–15 μ*M* | **Fungi** (**MIC:** *A. flavus* = 25.0, *A. fumigatus* = 6.25–12.5, *A. niger* = 12.5–25.0, *F. moniliforme* = 1.5–3.0, *F. oxysporum* = 1.5–3.0, *S. cerevisiae* W308 = 1.1–2.1, *G. candidum* = 2.1–4.3, *M. anisopliae* = 2.1–4.3, *B. bassianaa* > 17, *C. albicans* ATCC 10231 = 0.6, *C. albicans* ATCC 10261 = 1.3, *C. albicans* Sc 5314 = 2.5, *C. albicans* Dsm 6659 = 2.5 μ*M*, **LC:** *S. cerevisiae* W308 = 1, *S. cerevisiae* S289C = 1.3, *D. uninucleata* = 0.4, *G. candidum* = 1.4, *M. anisopliae* = 1.1, *B. bassianaa* > 10 μ*M*), **Potozoa** (*L. aethiopica*); **Virus** (**$ID_{50}$:** HIV-1 = 2–3 μ*M*, **$EC_{50}$:** JV = 3.4 μ*M*) | 68, 356, 381–394 |
| Melittin B | | **MIC:** *B. subtilis* KCTC 3068 = 1–2, *S. epidermidis* KCTC 1917 = 1, *S. aureus* KCTC 1621 = 0.5–1, *B. megaterium* Bm11 = 0.3, *M. luteus* ATCC 9341 = 0.4, *B. subtilis* ATCC 6051 = 0.4–3.1, *B. subtilis* IP 5832 = 12.5, *L. monocytogenes* NCTC 11994 = 6.3, *M. luteus* ATCC 9341 = 3.1, *E. faecalis* ATCC 19433 = 6.3, *S. aureus* ATCC 29213 = 6.3, *S. pneumoniae* ATCC 33400 = 6.3, *N. asteroides* ATCC 3308 = 6.3, *S. aureus* (methicillin R) = 1, *S. aureus* NCTC 8325 = 3 μ*M*, *S. aureus* 8530 = 10–15, *B. subtilis* = 2–4 μg/mL | **MIC:** *E. coli* KCTC 1682 = 2, *S. typhimurium* KCTC 1926 = 2, *P. aeruginosa* KCTC 1637 = 2, *E. coli* D21 = 5, *A. calcoaceticus* Ac11 = 2–20, *P. aeruginosa* ATCC 27853 = 25, *S. typhimurium* LT2 = 13, *E. coli* ATCC 25922 = 4–25, *E. coli* DH5α = 3.9–50, *S. marcescens* ATCC 133880 > 50, *P. aeruginosa* ATCC 257853 > 50, *K. pneumoniae* ATCC 13833 = 50, *S. choleraesuis* ATCC 13311 > 50, *H. influenzae* ATCC 19418 > 50, *P. aeruginosa* (R) = 8, *A. laidlawii* A-EF22 = 0.78, *M. gallisepticum* S6 = 6.25, *M. mycoides* ssp. *mycoides* SCKH3J = 12.5, *S. citri* R8A2 = 0.39, *S. floricola* BNR1 = 3.12, *S. melliferum* BC3 = 1.56 μ*M*, *E. coli* W 160-37 = 5–10, *P. putida* NCIM 2102 = 25–35 μg/mL | **Fungi** (**$IC_{50}$:** *N. crassa* = 0.8, *B. cinerea* = 3.1, *F. culmorum* = 0.8, *S. cerevisiae* = 5 μ*M*, **MIC:** *C. albicans* TIMM 1768 = 3.125, *T. beigelii* = 3.125–6.25, *S. cerevisiae* = 3.125, *P. digitatum* Conidia = 4.0, *P. digitatum* Mycelium = 8.0, *C. albicans* ATCC 90028 = 4 μ*M*) | 138, 176, 190, 289, 293, 298, 365–370 |

| | | | | | |
|---|---|---|---|---|---|
| Cecropin A (1–8) + melittin B (1–18) | (CEME) | **LC:** *B. subtilis* Bs11 = 0.4, *B. megaterium* Bm11 = 0.4, *S. aureus* Cowan I = 1, *S. pyogenes* Sp1 = 2 μ*M* | **LC:** *E. coli* D21 = 0.3, *P. aeruginosa* OT97 = 0.7 μ*M*, **MIC:** *E. cloacae* 218R1 (β-Lactam R) = 2.4, *E. cloacae* 218S = 2.4, *E. coli* DC2 (Polymyxin S) = 0.6, *E. coli* SC9251 = 1.2, *E. coli* SC9252 (Polymyxin R) = 1.2, *E. coli* UB1005 = 1–2.4, *P. aeruginosa* 1008OCR01 = 2, *P. aeruginosa* H309 = 2.4, *P. aeruginosa* K385 = 2, *P. aeruginosa* K799 = 4.8, *P. aeruginosa* PA-83-48 = 2, *P. aeruginosa* PAO1 = 2, *P. aeruginosa* PAO963 = 2, *P. aeruginosa* Z61 (S) = 1.2–2, *S. typhimurium* 14028s = 2–2.4, *S. typhimurium* MS7953s (Defensin S) = 0.6–1 μg/mL, *A. baumannii* Ac157* = 2 μ*M*, **MIC90:** *A. baumannii** = 4 μg/mL | | 272, 273, 275–277, 279 |
| Cecropin A (1–8) + melittin B (1–12) | | **LC:** *B. subtilis* Bs11 = 0.3, *B. megaterium* Bm11 = 0.2, *S. aureus* Cowan I = 2, *S. pyogenes* Sp1 = 0.1 μ*M*, **MIC:** *B. subtilis* ATCC 6633 = 0.78–1.56 μ*M*, *B. subtilis* ATCC 6633 = 6.25 μg/mL | **LC:** *E. coli* D21 = 0.4, *P. aeruginosa* OT97 = 0.9 μ*M*, **MIC:** *E. coli* ATCC 33694 = 6.25–12.5 m*M*, *E. coli* ATCC 33694 = 25.0 μg/mL | **Fungi** (**MIC:** *T. beigelii* = 5.0 μg/mL), **Cancer cells** (**$IC_{50}$:** Human SCLC H-148 = 9.0, Human SCLC H-126 = 5.0, Human SCLC H-69 = 8.0 μ*M*) | 272, 274, 278, 280 |
| Cecropin A (1–8) + melittin B (1–12) | GIG→GPG | **MIC:** *B. subtilis* ATCC 6633 = 3.125 μ*M* | **MIC:** *E. coli* ATCC 33694 = 25.0 μ*M* | **Cancer cells** (**$IC_{50}$:** Human SCLC H-148 > 60.0, Human SCLC H-126 > 60.0, Human SCLC H-69 > 60.0 μ*M*) | 278 |
| Cecropin A (1–8) + melittin B (1–12) | T18K, T19K | | | **Fungi** (**MIC:** *T. beigelii* = 2.5 μg/mL) | 274 |
| Cecropin A (1–8) + melittin B (1–18) derivative | (V681) | **MIC:** *S. aureus* ATCC25923 = 4, *S. epidermidis** = 4, *B. subtilis* C971 = 2.2–3.1, *C. xerosis* C875 = 2.5–3.1, *E. faecalis* C625 = 6.3–16.0, *S. aureus* 25923 = 16.0, *S. aureus* 622 = 3.1–6.3, *S. aureus* 623 = 6.3, *S. aureus* SAP0017 = 9.0, *S. epidermidis* C621 = 5.0–6.3 μg/mL | **MIC:** *E. coli* UB1005 = 0.5, *E. coli* KF130 = 1, *E. coli* KF111 = 1, *S. typhimurium* 14028s = 4, *P. aeruginosa* K799 = 2, *P. aeruginosa* H374 = 2, *P. aeruginosa* H744 = 2, *E. coli* DC2 = 4.5, *E. coli* C498 = 6.3, *E. coli* C500 = 3.1, *E. coli* UB1005 = 7.1, *P. aeruginosa* H187 = 6.3, *P. aeruginosa* H188 = 3.1–20.2, *S. typhimurium* C587 = 12.5–20.2, *S. typhimurium* C610 = 3.1–5.7, *S. typhimurium* C610 = 5.7, *S. typhimurium* C587 = 12.5 μg/mL | **Fungi** (**MIC:** *C. albicans* = 12.5 μg/mL) | 130, 131, 300, 395 |

(*Continued*)

**TABLE 5B** ***Continued***

| Lead Peptide | Substitution | Gram Positive | Gram Negative | Antifungal or Others | Reference |
|---|---|---|---|---|---|
| Cecropin A (1–8) + melittin B (1–18) derivative | (V681: V13K) | **MIC:** *B. subtilis* C971 = 1.6–3.1, *C. xerosis* C875 = 1.3–1.6, *E. faecalis* C625 = 50.0–64.0, *S. aureus* 25923 = 64.0, *S. aureus* 622 = 12.5, *S. aureus* 623 = 12.5, *S. aureus* SAP0017 = 64.0, *S. epidermidis* C621 = 5.0–6.3 μg/mL | **MIC:** *E. coli* DC2 = 1.6, *E. coli* C498 = 3.1, *E. coli* C500 = 3.1, *E. coli* UB1005 = 2.5, *P. aeruginosa* H187 = 6.3, *P. aeruginosa* H187 = 8.0, *P. aeruginosa* H188 = 3.1, 5.0, *S. typhimurium* C610 = 1.3, *S. typhimurium* C610 = 3.1, *S. typhimurium* C587 = 4.0, *S. typhimurium* C587 = 6.3 μg/mL | **Fungi** (**MIC:** *C. albicans* = 50.0 μg/mL) | 130, 131 |
| Magainin 2 | | **MIC:** *S. epidermidis* = 10, *S. aureus* = 50, *S. aureus* ATCC 25923 = 35–70, *S. aureus* ATCC 29213 = 25, *S. agalactiae* ATCC 12386 = 100, *S. faecalis* ATCC 212 > 100, *S. faecalis* ATCC 19433 = 35–70, *S. faecalis* ATCC 29212 = 400, *S. pyogenes* ATCC 19615 = 1–7, *S. aureus* (methicillin S) = 16–128, *S. aureus* (methicillin S) = 16, > 128, *S. pneumoniae* = 4–128, *S. sanguis* = 16–128, *Gemella* spp. = 8–64, *Lactococcus* spp. = 8–64, *E. faecalis* > 128, *R. equi* = 32–128 μg/mL | **MIC:** *A. caloaceticus* ATCC 19606 = 1–7, *C. freundii* ATCC 8090 = 7–35, *E. aerogenes* ATCC 13048 = 7–35, *E. cloacae* ATCC 23355 = 1–7, *E. coli* ATCC 25922 = 1–100, *E. coli* D31 = 5, *K. oxytoca* = 50, *K. pneumoniae* ATCC 13883 = 7–100, *P. aeruginosa* ATCC 27853 = 35–100, *P. aerugrnosa* ATCC 27953 = 6.25, *P. mirabilis* > 400, *P. putida* = 10, *P. vulgaris* ATCC 13315 > 100, *S. anatum* = 50, *S. epidermidis* ATCC 12228 = 7–35, *S. flexneri* ATCC 12022 = 7–35, *S. marcescens* ATCC 8100 > 100, *S. sonnei* ATCC 25931 = 1–7, *S. typhimurium* ATCC 14028 = 35–70, *E. coli* = 0.25–1, *K. pneumoniae* = 1–4, *E. cloacae* = 1–8, *P. mirabilis* = 8–32, *S. marcescens* = 2–32, *S. typhi* = 1–16, *Salmonella* spp. = 2–32, *Shigella* spp. = 2–32, *Y. enterocolitica* = 1–8, *P. aeruginosa* = 8–128, *Pseudomonas* spp. = 2–128, *S. maltophilia* = 16, > 128, *Acinetobacter* spp. = 1–8, *Brucella* spp. = 0.5–4, *P. aeruginosa* ATCC 27853 = 8, *P. aeruginosa** = 4–128 μg/mL | **Fungi** (**MIC:** *C. albicans* = 80, *S. cerevisiae* ATCC X2180 = 7–35 μg/mL, *R. solani* LR172 = 30.2 μ*M*); **Protozoa** (**IC:** *P. caudatum* = 10 μg/mL, *A. castellanii* trophozoites 50% inhibition at 100 μg/mL for 24 h, *C. parvum* = 33.2% reduction at 50 μ*M*); **Cancer cells** (**$IC_{50}$:** Bladder cancer cell = 198.1 μ*M* (WST-1 assay), 75.2 μ*M* (BrdU assay)); **Virus** (**$EC_{50}$:** HSV-1 = 22.16, HSV-2 = 19.80 μ*M*, **$ED_{50}$:** HSV-1 = 99.9, HSV-2 = 116.5 μg/mL); **Insect** (**$LD_{50}$:** *D. melanogaster* = 107.4–147.5 pmol/mg) | 26, 154, 158, 171, 263, 352, 354–364 |

| | | | | | |
|---|---|---|---|---|---|
| Cecropin A (1–8) + magainin 2 (1–12) | | **MIC:** *B. subtilis* KCTC 1918 = 0.78–3.12, *S. epidermidis* KCTC 3096 = 1.56–6.25, *S. pyogenes* KCTC 3096 = 0.78, *S. aureus* KCTC 1621 = 3.125, *B. subtilis* ATCC 6633 = 0.78–1.56 μ*M*, *B. subtilis* ATCC 6633 = 6.25 μg/mL | **MIC:** *P. aeruginosa* KCTC 1637 = 1.56, *S. typhimurium* KCTC 1926 = 0.19–0.78, *E. coli* KCTC 1682 = 3.125, *P. vulgaris* KCTC 2433 = 3.125, *E. coli* ATCC 33694 = 6.25 μ*M*, *E. coli* ATCC 33694 = 12.5 μg/mL | **Fungi** (**MIC:** *C. albicans* TIMM 1768 = 12.5, *T. beigelii* KCTC 7707 = 6.25 μ*M*); **Cancer cells** (Calu-6, SNU601, Jurkat, **$IC_{50}$:** Human chronic myelogenous leukemia K-562 ATCC CCL-243 = 65.0, Acute T-cell leukemia Jurkat ATCC TIB-152 = 30.0, Human lung carcinoma cancer cells K-549 ATCC CCL-185 = 60.0, Breast adenocarcinoma cells MDA-MB-361 ATCC HTB-27 = 44.0, Human SCLC H-148 = 13.0, Human SCLC H-126 = 9.0, Human SCLC H-69 = 16.0 μ*M*) | 280, 396–398 |
| Cecropin A (1–8) + magainin 2 (1–12) | GIG→GPG | **MIC:** *B. subtilis* ATCC 6633 = 3.125 μ*M* | **MIC:** *E. coli* ATCC 33694 = 12.5 μ*M* | **Cancer cells** (**$IC_{50}$:** Human SCLC H-148 > 60.0, Human SCLC H-126 > 60.0, Human SCLC H-69 > 60.0 μ*M*) | 278 |
| Cecropin A (1–8) + magainin 2 (1–12) | S16A | **MIC:** *B. subtilis* ATCC 6633 = 1.56 μ*M* | **MIC:** *E. coli* ATCC 33694 = 3.125 μ*M* | **Cancer cells** (**$IC_{50}$:** Human SCLC H-148 = 9.0, Human SCLC H-126 = 4.0, Human SCLC H-69 = 8.0 μ*M*) | 278 |
| Cecropin A (1–8) + magainin 2 (1–12) | S16V | **MIC:** *B. subtilis* ATCC 6633 = 0.78–1.56 μ*M* | **MIC:** *E. coli* ATCC 33694 = 3.125–6.25 μ*M* | **Cancer cells** (**$IC_{50}$:** Human SCLC H-148 = 5.5, Human SCLC H-126 = 3.0, Human SCLC H-69 = 4.0 μ*M*) | 278 |
| Cecropin A (1–8) + magainin 2 (1–12) | S16L | **MIC:** *B. subtilis* ATCC 6633 = 0.78–1.56 μ*M* | **MIC:** *E. coli* ATCC 33694 = 1.56–3.125 μ*M* | **Cancer cells** (**$IC_{50}$:** Human SCLC H-148 = 4.0, Human SCLC H-126 = 2.0, Human SCLC H-69 = 4.0 μ*M*) | 278, 280 |
| Cecropin A (1–8) + magainin 2 (1–12) | S16F | **MIC:** *B. subtilis* ATCC 6633 = 1.56 μ*M* | **MIC:** *E. coli* ATCC 33694 = 3.125 μ*M* | **Cancer cells** (**$IC_{50}$:** Human SCLC H-148 = 2.5, Human SCLC H-126 = 1.5, Human SCLC H-69 = 2.0 μ*M*) | 278 |
| Cecropin A (1–8) + magainin 2 (1–12) | S16W | **MIC:** *B. subtilis* ATCC 6633 = 0.78–1.56 μ*M* | **MIC:** *E. coli* ATCC 33694 = 1.56 μ*M* | **Cancer cells** (**$IC_{50}$:** Human SCLC H-148 = 4.0, Human SCLC H-126 = 1.5, Human SCLC H-69 = 2.0 μ*M*) | 278 |
| Cecropin A (1–8) + magainin 2 (1–12) | S16A, K12L | **MIC:** *B. subtilis* ATCC 6633 = 50 μg/mL | **MIC:** *E. coli* ATCC 33694 = 100 μg/mL | | 280 |

(*Continued*)

**TABLE 5B** ***Continued***

| Lead Peptide | Substitution | Gram Positive | Gram Negative | Antifungal or Others | Reference |
|---|---|---|---|---|---|
| Cecropin A (1–8) + magainin 2 (1–12) | S16L, GIG→L | **MIC:** *B. subtilis* KCTC 1918 = 1.56, *S. aureus* KCTC 1621 = 3.125, *B. megaterium* KCTC 1096 = 1.56 μ*M* | **MIC:** *E. coli* KCTC 1682 = 12.5, *P. aeruginosa* KCTC 1637 = 3.125, *P. vulgaris* = 6.25 μ*M* | **Fungi** (**MIC:** *C. albicans* KCTC 7121 = 25.0, *T. beigelii* KCTC 7251 = 12.5, *A. flavus* KCTC 7375 = 25.0, *F. oxysporum* KCTC 6084 = 25.0 μ*M*), **Cancer cells** (**$IC_{50}$:** Human chronic myelogenous leukemia K-562 ATCC CCL-243 = 3.0, Acute T-cell leukemia Jurkat ATCC TIB-152 = 9.0, Breast adenocarcinoma cells MDA-MB-361 ATCC HTB-27 = 3.0 μ*M*) | 399, 400 |
| Cecropin A (1–8) + magainin 2 (1–12) | S16L, GIG→S | **MIC:** *B. subtilis* KCTC 1918 = 1.56, *S. aureus* KCTC 1621 = 6.25, *B. megaterium* KCTC 1096 = 3.125 μ*M* | | **Cancer cells** (**$IC_{50}$:** Human chronic myelogenous leukemia K-562 ATCC CCL-243 = 3.0, Acute T-cell leukemia Jurkat ATCC TIB-152 = 7.0, Breast adenocarcinoma cells MDA-MB-361 ATCC HTB-27 = 4.0 μ*M*) | 399 |
| Cecropin A (1–8) + magainin 2 (1–12) | S16L, GIG→P | **MIC:** *B. subtilis* KCTC 1918 = 0.78–1.56, *S. aureus* KCTC 1621 = 1.56–3.125, *B. megaterium* KCTC 1096 = 0.78, *B. subtilis* KCTC 3068 = 0.78–2, *S. pyogenes* KCTC 3096 = 0.78–6.25, *S. aureus* KCTC 1916 = 1, *S. aureus** (methicillin R) = 2, *S. epidermidis* KCTC 1917 = 1, *E. faecalis** (Vancomycin R) = 4, *E. faecium** (Vancomycin R) = 1 μ*M*, *B. subtilis* ATCC 6633 = 0.78–1.56 μg/mL | **MIC:** *E. coli* KCTC 1682 = 4–6.25, *P. aeruginosa* KCTC 1637 = 1.56–3.125, *P. vulgaris* = 3.125, *S. typhimurium* KCTC 1926 = 0.78–3.13, *S. typhimurium* KCTC 1925 = 2, *S. marcescens* KCTC 1299 = 6.25, *S. maltophilia** = 4 μ*M*, *E. coli* ATCC 33694 = 3.125 μg/mL | **Fungi** (**MIC:** *C. albicans* KCTC 7121 = 12.5, *T. beigelii* KCTC 7251 = 12.5, *A. flavus* KCTC 7375 = 25.0, *F. oxysporum* KCTC 6084 = 25.0, *C. albicans* KCTC 7965 = 4, *C. albicans** (azole R) = 1–2 μ*M*); **Cancer cells** (**$IC_{50}$:** Human chronic myelogenous leukemia K-562 ATCC CCL-243 = 8.0–35.1, Acute T-cell leukemia Jurkat ATCC TIB-152 = 4.0–20.0, Human lung carcinoma cancer cells K-549 ATCC CCL-185 = 25.0, Breast adenocarcinoma cells MDA-MB-361 ATCC HTB-27 = 3.5–28.0, Human small-cell lung cancer cell lines = 2.0–4.0 μ*M*) | 278, 399–402 |

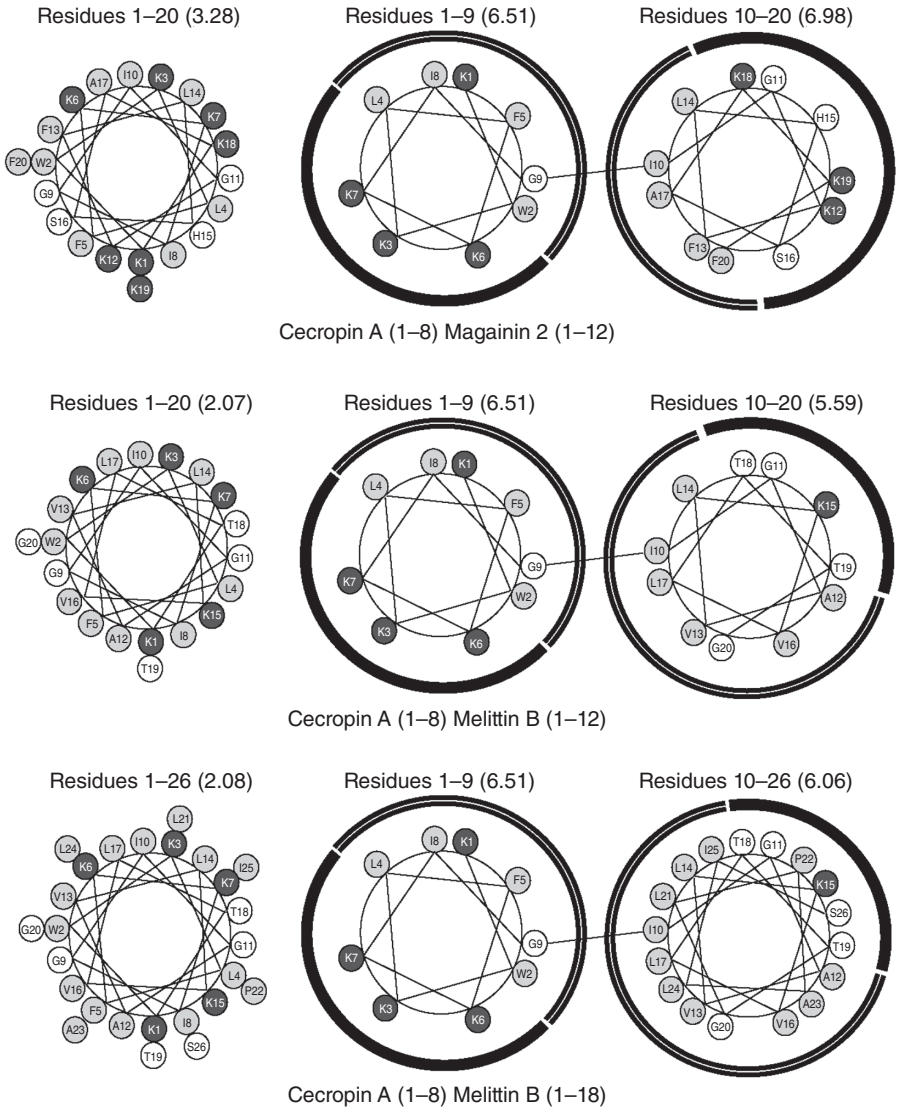


**Figure 13** Helical wheel representation of cecropin A hybrid α-helical antimicrobial peptides. Amphipathicity is indicated in brackets. The diagrams on the left show the hybrid peptides as a continuous helix and calculate the amphipathicity as a single helix. The figures on the right represent the hybrid peptides as two helical segments linked by a hinge region, and the amphipathicity values are shown for each segment. Gray denotes hydrophobic amino acids, dark gray denotes positively charged Lys residues, and white denotes hydrophilic residues, Gly, Thr, Ser and His.

D-Pro instead of L-Pro (peptide 14). Pro is known to disrupt α-helical structure in aqueous conditions and Pro has the lowest α-helical propensity value of all 20 amino acid residues (Table 3). Thus, distortion of α-helical structure by D-amino acid substitutions along the helix or a Pro substitution on the nonpolar face of an amphipathic α-helix are both effective at creating cell selectivity.

The last interesting result we observed on examining the activity profiles of de novo designed nonamphipathic cationic peptides was the dramatic enhancement of antimicrobial activity for both gram-negative and gram-positive organisms while maintaining cell selectivity when replacing 6 lysine residues by 6 arginine residues (peptides 15 and 16, Table 6) [285]. These effects of Arg are most likely due to its resonance-stabilized guanidinium group, which could lead to easier desolvation and more efficient ion-pairing with anions in the medium [124]. The difference in activity may be linked to differences in membrane lipid selectivity of Arg versus Lys side chains [286]. Interestingly, the occurrence of Arg and Lys in native antimicrobials in our database was 28

**TABLE 6 Representative Examples of de novo Designed Amphipathic Antimicrobial Peptide**

| Number | Name | Sequence[a] | Gram Positive | Gram Negative | Hemolytic Activity | Length | C-Amide |
|---|---|---|---|---|---|---|---|
| 1 | K4A7W | KAAAKWAAKAAK | **MIC:** *B. subtilis* ATCC 6051 = 200, *S. aureus* ATCC 25922 = 200 μ*M* | **MIC:** *E. coli* D21 = 200, *A. calcoaceticus* Ac11 = 200, *P. aeruginosa* ATCC 27853 = 200 μ*M* | No | 12 | Y |
| 2 | K4L7W | KLLLKWLLKLLK | **MIC:** *B. subtilis* ATCC 6051 = 1.25, *S. aureus* ATCC 25922 = 18.1 μ*M* | **MIC:** *E. coli* D21 = 12.5, *A. calcoaceticus* Ac11 = 3, *P. aeruginosa* ATCC 27853 > 200 μ*M* | 100% hemolysis at 100 μ*M* | 12 | Y |
| 3 | K4L3l4W | KLLLKWLLKLLK | **MIC:** *B. subtilis* ATCC 6051 = 1.25, *S. aureus* ATCC 25922 = 30 μ*M* | **MIC:** *E. coli* D21 = 15, *A. calcoaceticus* Ac11 = 15, *P. aeruginosa* ATCC 27853 = 7.5 μ*M* | <20% hemolysis at 100 μ*M* | 12 | Y |
| 4 | Amphipathic-1L | LKLLKKLLKKLLKLL | **MIC:** *M. luteus* = 2, *S. simulans* = 2, *S. aureus* II ATCC 6538P = 3 μ*M* | **MIC:** *E. coli* ATCC 25922 = 13, *E. coli* D21 = 13–50, *S. epidermidis* (methicillin R) LT1324 = 4, *P. aeruginosa* ATCC 27853 = 25, *A. baumannii* ATCC 19606 = 1–12 μ*M*, *P. aeruginosa* ATCC 27853 = 45.0, *A. baumannii* (Gentamicin R) = 11.2 μg/mL | 100% hemolysis at 100 μ*M*, 180 μg/mL | 15 | Y |

| Number of Pro Residue | Number of Positive Charges | Number of Negative Charges | Net Charge (pH 7) | Charge Density (pH 7) | Number of Hydrophobic Residues | Hydrophobic Density | Amphipathicity | Amphipathicity per Residue | Reference |
|---|---|---|---|---|---|---|---|---|---|
| 0 | 5 | 0 | 5 | 0.42 | 1 | 2.8 | 2.99 | 0.25 | 281 |
| 0 | 5 | 0 | 5 | 0.42 | 8 | 17.1 | 5.82 | 0.48 | 281 |
| 0 | 5 | 0 | 5 | 0.42 | 8 | 17.1 | — | — | 281 |
| 0 | 7 | 0 | 7 | 0.47 | 9 | 14.8 | 8.51 | 0.57 | 282, 403–405 |

(*Continued*)

**TABLE 6** ***Continued***

| Number | Name | Sequence[a] | Gram Positive | Gram Negative | Hemolytic Activity | Length | C-Amide |
|---|---|---|---|---|---|---|---|
| 5 | Amphipathic-1D | LKLLKKLLKKLLKLL | **MIC:** *S. aureus* II ATCC 6538P = 13, *S. simulans* = 1, *M. luteus* = 0.5, *E. faecium* I-11305b = 4, *E. faecium* I-11054 > 4, *S. haemolyticus* (coagulase-negative) I-10925 = 1, *S. epidermidis* (methicillin-resistant [R]) LT1324 = 1, *S. aureus* (methicillin S) 5185 = 4, *S. aureus* (methicillin S) I-11574 = 4, *S. aureus* (methicillin R) LT1338 > 4, *S. aureus* (methicillin R) LT1334 > 4 μ*M* | **MIC:** *E. coli* ATCC 25922 = 7, *E. coli* D21 = 4–6, *S. epidermidis* (methicillin R) LT1324 = 1, *P. aeruginosa* ATCC 27853 = 3, *A. baumannii* ATCC 19606 = 6–13, *C. freundii* I-11090 = 2, *K. pneumoniae* I-10910 = 4, *E. coli* I-11276b = 2, *E. coli* O-19592 = 1, *S. maltophila* O-16451 > 4, *S. maltophila* I-10717 = 4, *P. aeruginosa* 4991 > 4, *P. aeruginosa* I-10968 > 4 μ*M*, *P. aeruginosa* ATCC 27853 = 5.6, *A. baumannii* (Gentamicin R) = 5.6 μg/mL | 0% hemolysis at 100 μ*M* | 15 | Y |
| 6 | K5L7 | KLLLKLKLKLLK | | **MIC:** *E. coli* ATCC 25922 > 100, *S. sonnei* ATCC 25931 > 100, *S. choleraesuis* ATCC 14028 > 100 μ*M* | 80% hemolysis at 100 μ*M* | 12 | Y |
| 7 | 4D-K5L7 | KLLLKLKLKLLK | | **MIC:** *E. coli* ATCC 25922 = 5, *S. sonnei* ATCC 25931 = 4, *S. choleraesuis* ATCC 14028 = 8 μ*M* | 0% hemolysis at 100 μ*M* | 12 | Y |
| 8 | K6L6W | KWKKLLKKLLKLL | **MIC:** *B. subtilis* KCTC 3068 = 4, *S. epidermidis* KCTC 1917 = 4, *S. aureus* KCTC 1621 = 8 μ*M* | **MIC:** *E. coli* KCTC 1682 = 8, *S. typhimurium* KCTC 1926 = 8, *P. aeruginosa* KCTC 1637 = 8 μ*M* | 80% hemolysis at 50 μ*M* | 13 | Y |
| 9 | K6L5WP | KWKKLPKKLLKLL | **MIC:** *B. subtilis* KCTC 3068 = 2, *S. epidermidis* KCTC 1917 = 2, *S. aureus* KCTC 1621 = 2 μ*M* | **MIC:** *E. coli* KCTC 1682 = 4, *S. typhimurium* KCTC 1926 = 4, *P. aeruginosa* KCTC 1637 = 4 μ*M* | 0% hemolysis at 100 μ*M* | 13 | Y |

| Number of Pro Residue | Number of Positive Charges | Number of Negative Charges | Net Charge (pH 7) | Charge Density (pH 7) | Number of Hydrophobic Residues | Hydrophobic Density | Amphipathicity | Amphipathicity per Residue | Reference |
|---|---|---|---|---|---|---|---|---|---|
| 0 | 7 | 0 | 7 | 0.47 | 9 | 14.8 | — | — | 282, 403–405 |
| 0 | 6 | 0 | 6 | 0.50 | 7 | 14.4 | 3.62 | 0.30 | 283 |
| 0 | 6 | 0 | 6 | 0.50 | 7 | 14.4 | — | — | 283 |
| 0 | 7 | 0 | 7 | 0.54 | 7 | 13.9 | 9.70 | 0.75 | 176 |
| 1 | 7 | 0 | 7 | 0.54 | 7 | 12.8 | 8.92 | 0.69 | 176 |

(*Continued*)

**TABLE 6** ***Continued***

| Number | Name | Sequence[a] | Gram Positive | Gram Negative | Hemolytic Activity | Length | C-Amide |
|---|---|---|---|---|---|---|---|
| 10 | KLW | KWKKLLKKLLKLLKKLLK | **MIC:** *B. subtilis* KCTC 3068 = 8, *S. epidermidis* KCTC 1917 = 4, *S. aureus* KCTC 1621 = 4 μ*M* | **MIC:** *E. coli* KCTC 1682 = 4, *S. typhimurium* KCTC 1926 = 8, *P. aeruginosa* KCTC 1637 = 4 μ*M* | 50% hemolysis at 50 μ*M* | 18 | Y |
| 11 | KLW-L9P | KWKKLLKKPLKLLKKLLK | **MIC:** *B. subtilis* KCTC 3068 = 2, *S. epidermidis* KCTC 1917 = 1, *S. aureus* KCTC 1621 = 2 μ*M* | **MIC:** *E. coli* KCTC 1682 = 2, *S. typhimurium* KCTC 1926 = 2, *P. aeruginosa* KCTC 1637 = 1 μ*M* | <10% hemolysis at 50 μ*M* | 18 | Y |
| 12 | KLW-L9p | KWKKLLKKPLKLLKKLLK | **MIC:** *B. subtilis* KCTC 3068 = 1, *S. epidermidis* KCTC 1917 = 1, *S. aureus* KCTC 1621 = 2 μ*M* | **MIC:** *E. coli* KCTC 1682 = 2, *S. typhimurium* KCTC 1926 = 1, *P. aeruginosa* KCTC 1637 = 1 μ*M* | <5% hemolysis at 50 μ*M* | 18 | Y |
| 13 | KLW-K11P | KWKKLLKKLLPLLKKLLK | **MIC:** *B. subtilis* KCTC 3068 = 4, *S. epidermidis* KCTC 1917 = 2, *S. aureus* KCTC 1621 = 2 μ*M* | **MIC:** *E. coli* KCTC 1682 = 4, *S. typhimurium* KCTC 1926 = 4, *P. aeruginosa* KCTC 1637 = 4 μ*M* | 50% hemolysis at 50 μ*M* | 18 | Y |
| 14 | KLW-K11p | KWKKLLKKLLPLLKKLLK | **MIC:** *B. subtilis* KCTC 3068 = 1, *S. epidermidis* KCTC 1917 = 1, *S. aureus* KCTC 1621 = 2 μ*M* | **MIC:** *E. coli* KCTC 1682 = 2, *S. typhimurium* KCTC 1926 = 2, *P. aeruginosa* KCTC 1637 = 1 μ*M* | <20% hemolysis at 50 μ*M* | 18 | Y |
| 15 | F17 | KKKAAAFAAWAAFAKKK | **MIC:** *C. xerosis* C875 = 4, *S. epidermidis* C621 > 32, *S. aureus* ATCC 25923 > 32, *E. faecalis* ATCC 29212 > 32, *B. subtilis* ATCC 6633 = 16 μ*M* | **MIC:** *E. coli* DC2 = 16, *E. coli* UB1005 = 32, *E. coli* ATCC 25922 ≥ 32, *P. aeruginosa* ATCC 27853 > 32 μ*M* | 0% hemolysis at 200 μ*M* | 17 | Y |
| 16 | F17-R | RRRAAFAAWAAFAARRR | **MIC:** *C. xerosis* C875 = 1, *S. epidermidis* C621 = 4, *S. aureus* ATCC 25923 > 32, *E. faecalis* ATCC 29212 > 32, *B. subtilis* ATCC 6633 = 2 μ*M* | **MIC:** *E. coli* DC2 = 1, *E. coli* UB1005 = 2, *E. coli* ATCC 25922 = 8, *P. aeruginosa* ATCC 27853 = 16 μ*M* | 0% hemolysis at 200 μ*M* | 17 | Y |

[a]Underlined amino acid residues denote D-amino acids. S denotes antibiotic sensitive strain; R denotes antibiotic resistant strain.

| Number of Pro Residue | Number of Positive Charges | Number of Negative Charges | Net Charge (pH 7) | Charge Density (pH 7) | Number of Hydrophobic Residues | Hydrophobic Density | Amphipathicity | Amphipathicity per Residue | Reference |
|---|---|---|---|---|---|---|---|---|---|
| 0 | 10 | 0 | 10 | 0.56 | 9 | 12.8 | 8.98 | 0.50 | 284 |
| 1 | 10 | 0 | 10 | 0.56 | 9 | 12.0 | 8.24 | 0.46 | 284 |
| 1 | 10 | 0 | 10 | 0.56 | 9 | 12.0 | — | — | 284 |
| 1 | 9 | 0 | 9 | 0.50 | 10 | 13.3 | 8.29 | 0.46 | 284 |
| 1 | 9 | 0 | 9 | 0.50 | 10 | 13.3 | — | — | 284 |
| 0 | 7 | 0 | 7 | 0.41 | 3 | 5.5 | 3.27 | 0.19 | 285 |
| 0 | 7 | 0 | 7 | 0.41 | 3 | 5.5 | 3.47 | 0.20 | 285 |

and 72%, respectively. These results suggest that the incorporation of Arg residues in amphipathic α-helical peptides instead of Lys residues may have an advantageous effect. Deslouches et al. [287] designed a 24-residue AMP composed of Arg (13 residues), Val (8 residues), and Trp (3 residues) with the sequence RRWVRRVRRWVRRVVRVVRRWVRR, arranged to form an idealized helical and amphipathic structure with a high net positive charge and high hydrophobicity. This peptide had high antimicrobial activity and minimal toxicity to mammalian cells (no hemolysis at 50 μ*M*). In animal studies, this peptide had a 50% lethal dose between 12 and 16 mg/kg. When *Pseudomonas*-infected mice [intrperitoneal (i.p.) with strain PAO1 at $1 \times 10^7$ to $1.5 \times 10^7$ colony-forming units (CFU)] were treated intravenously (i.v.) with 3 mg of peptide per kilogram body weight 30–45 min after administration of *P. aeruginosa*, no fatalities were observed; however, all untreated animals died by 40 h.

### 4.3 Enhancing Antimicrobial Activity and Removing Toxicity to Eukaryotic Cells

The major barrier to the use of AMPs as antibiotics is their toxicity or ability to lyse eukaryotic cells. To be useful as a broad-spectrum antibiotic, it would be necessary to dissociate antieukaryotic activity from antimicrobial activity, that is, increasing the antimicrobial activity and reducing toxicity to normal cells. An excellent measure of specificity is a parameter called therapeutic index, which is defined as the ratio of MHC (maximum peptide concentration that results in no hemolysis) and MIC (minimal inhibitory concentration or lowest peptide concentration that inhibited growth). A high value for the therapeutic index (MHC/MIC) means cell selectivity or specificity for bacteria over erythrocytes. For example, an MHC value of 500 μg/mL and an MIC value of 10 μg/mL would give a therapeutic index of 50. The therapeutic index (MHC/MIC) can be increased in one of the following three ways: increasing antimicrobial activity, decreasing hemolytic activity while maintaining antimicrobial activity, or a combination of both, that is, increasing antimicrobial activity and decreasing hemolytic activity. The vast majority of native AMPs are very hemolytic. However, hemolytic activity in amphipathic α-helical AMPs is related in a complex manner to hydrophobicity, amphipathicity, and net positive charge density (Table 1). With regard to amphipathicity, we have calculated overall amphipathicity based on the complete native sequence; however, it is clear that certain segments within the native sequence may be much more amphipathic. For example, BMAP-27 (bovine myeloid antimicrobial peptide-27) (1–27) has an overall amphipathicity of 5.23; however, its segment 1–18 has an amphipathicity of 8.30. Similarly, eCATH-1 (horse myeloid cathelicidin 1), a 26-residue AMP has an overall amphipathicity of 2.54, but the segment 1–17 has an amphipathicity of 4.65. Lycotoxin 1, a 25-residue AMP, has an overall amphipathicity of 1.17; however, segments 1–9 and 10–25 have amphipathicities of 6.26 and 2.35, respectively. Thus, structural determination is critical to determine if native AMPs contain one continuous α-helix in hydrophobic environments or have α-helical segments within the native sequence that are more amphipathic. Only with this information will it be possible to correlate amphipathicity with biological activity.

AMPs with low amphipathicity can have high hemolytic activity (e.g., histatin 5, amphipathicity of 2.26 and 50% hemolysis at 39 μ*M*) or low hemolytic activity (e.g., dermaseptin b1, amphipathicity of 2.79 and no hemolysis at 367 μ*M*). The relationship of net positive charge density and hemolytic activity is also complex. For example, BMAP-27 has a very high net positive charge density of 0.41 (+11 per 27 residues) while clavaspirin has a very low net positive charge density of 0.04 (+1 per 23 residues), yet both peptides are very hemolytic. On the other hand, PMAP-23 has a net positive charge density of 0.30 (+7 per 23 residues) with no hemolysis at 100 μ*M* (Table 1). Dermaseptin S9 has an interesting sequence. Its overall amphipathicity is low (1.87); however, the continuous hydrophobic segment 6–15 (IWLWVLLMIW) has an amphipathicity of 2.73. This peptide has the highest hydrophobic density (13.4) of all the AMPs listed in Table 1 and this may explain its high hemolytic activity (50% hemolysis at 175 μ*M*). Peptides with high hydrophobic density, high amphipathicity, and high charge density are very hemolytic. For example, BMAP-27 has a hydrophobic density of 11.2, an amphipathicity of 8.30 for 1–18, and a high net positive charge density of 0.41 and is very hemolytic (>80% hemolysis at 30 μ*M*) (Table 1). Clavaspirin has a very high hydrophobic density (11.7) and high amphipathicity of 6.61 for segment 1–10, which may explain its very high hemolytic activity, yet it has the lowest net charge density (0.04) (Table 1).

***C-Terminal Amidation Increases Antimicrobial Activity*** Removal of the C-terminal negative charge by amidation has been shown to improve antimicrobial activity for many native peptides [112, 288, 289]. For example, pleurocidin, a 25-residue peptide with broad-spectrum activity against a wide range of gram-positive and gram-negative bacteria (Table 1) when C terminally amidated showed a dramatic enhancement of antimicrobial activity [288]. On the other hand, though the amidated form of ascaphin-8 (19 residues) dramatically enhanced antimicrobial activity, hemolytic activity was

also enhanced (the concentration of peptide producing 50% hemolysis changed from >200 to 55 μ*M* for the C-terminal amidated form) [175]. This may be explained by the hydrophobicity of the C terminal (-VLF) of ascaphin-8 where the C-terminal negative charge would decrease the overall hydrophobicity, thus making the peptide less hemolytic.

***All-D Form of AMPs Increases Antimicrobial Activity*** The all-D form of AMPs has been shown to enhance antimicrobial activity in vitro and in vivo due to stability against proteolytic degradation [290, 291].

***Decreasing Hemolytic Activity by Shortening the Peptide Sequence of a Native AMP While Maintaining Antimicrobial Activity*** Examples of this approach are shown with bovine myeloid cathelicidins, BMAP-27 and BMAP-28 (bovine myeloid antimicrobial peptides-28). BMAP-27 is shown in Table 1. By removing the hydrophobic C-terminal tail and maintaining the complete N-terminal helix (residues 1–18), antimicrobial activity was maintained but the lytic effects on human red blood cells was lost in the case of BMAP-27 (1–18) [184]. These researchers also substituted the hydrophobic residues in the C-terminal tail of BMAP-28 from IIVPIIAI-amide with more hydrophilic ones (QATPATRQ-amide) and showed that this peptide did not have any significant lytic activity toward human red blood cells (3.3 versus 94.1% lysis at 100 μ*M* peptide). Their results showed a dramatic reduction of the lytic activity against mammalian cells when the C-terminal tail was removed or when substitutions in the C-terminal tail were carried out to decrease hydrophobicity [184]. Interestingly, these short peptides (1–18) were less effective in vivo than the full length peptides [292]. A second example of this approach is the shortening of the 26-residue melittin sequence. Melittin exhibits potent antimicrobial activity and potent hemolytic activity (Table 1) (100% hemolysis at 1 μg/mL). The 15-residue synthetic peptide corresponding to the C terminus (GLPALISWIKRKRQQG-amide) had about 5–7 times less antibacterial activity but 300 times less hemolytic activity (100% hemolysis at 300 μg/mL). Interestingly, changing the sequence to GL<u>KK</u>LISWIKR<u>AA</u>QQG-amide restored comparable antimicrobial activity to native melittin with only a marginal increase in hemolytic activity (100% hemolysis at 100 μg/mL) [293]. This degree of hemolysis is still unacceptable for a potential therapeutic but this result clearly shows the value of this approach.

***Substitution of Peptoid Residues on the Nonpolar Face of α-Helical Amphipathic AMPs to Reduce Hemolysis*** Peptoid residues are imino acids with their side chains shifted from the α carbon to the nitrogen position. These residues are thought to have a general helix-disrupting tendency due to the lack of an amide proton [294, 295]. Song et al. [296] showed that the incorporation of two alanine peptoid residues into the hydrophobic face of a model α-helical AMP significantly reduced hemolytic activity but not antimicrobial activity. In a recent study [138], peptoid residues were introduced into the nonpolar face of melittin at positions Leu 6, Leu 13, and Ile 20. Melittin has a minimal peptide concentration that produces hemolysis of 0.78 μ*M*. The peptoid residue-containing analogs of melittin caused no hemolysis at 100 μ*M* and were able to maintain antimicrobial activity within two-fold of native melittin. Thus, the therapeutic index was increased from 0.6 for melittin to 71 for the best peptoid analog (119-fold improvement). Certain peptoid analogs did not exhibit any cytotoxicity to mammalian cell lines. These peptoid analogs also had no α-helical structure in aqueous buffer, SDS micelles, or small unilamellar vesicles (SUVs) and showed no self-association ([138] and references therein) compared to melittin, which shows a concentration-dependent formation of an α-helical conformation. This reduction in hemolytic activity was due to a loss of structure and not a reduction in hydrophobicity since substitution of three Phe peptoid residues will have similar hydrophobicity compared to the original residues in melittin (Leu 6, Leu 13, and Ile 20). It is interesting that reducing hydrophobicity by substituting these three large hydrophobes in melittin with Lys peptoid residues and, thus, increasing the net charge from +6 to +9, maintained antimicrobial activity, while this peptide still had no hemolysis at 100 μM. Thus, the key seems to be disruption of α-helical structure since substitution by either Phe peptoid residues or Lys peptoid residues maintains antimicrobial activity.

***Changing Net Charge on the Polar Face*** Recently, Jiang et al. [297] showed that both the number of positively charged residues on the polar face and net charge are important for both antimicrobial activity and hemolytic activity. The most dramatic observation was the sharp transition of hemolytic activity on increasing one positive charge on the polar face from +8 to +9 in a 26-residue amphipathic α-helical AMP. The change from +8 to +9 resulted in a greater than 32-fold increase in hemolytic activity (250 μg/mL to <7.8 μg/mL, respectively). On the other hand, increasing the number of positively charged residues from 7 to 8 maintained the hemolytic activity at 250 μg/mL (maximal peptide concentration that produces no hemolysis of human red blood cells after 18 h) while showing an enhancement of antimicrobial activity [297]. Decreasing the number of postively charged residues on this 26-residue AMP from

7 to 4 made the peptide inactive against 6 different strains of *P. aeruginosa,* showing the importance of a minimum positive charge density for antimicrobial activity [297]. Similarly, adding 3 negatively charged residues to the polar face while maintaining the 7 positively charged residues (net charge +4) dramatically decreased antimicrobial activity while having no effect on hemolytic activity. This result emphasizes the importance of a minimum net positive charge density on antimicrobial activity [297]. The dramatic increase in hemolytic activity on increasing the net charge from +8 to +9 can be explained by the increase in net charge favoring the formation of transmembrane channels/pores in eukaryotic cells. The increase in net positive charge beyond +8 for the 26-residue peptide could stabilize the pore by increasing hydrophilicity of the pore [297].

***Decreasing Hemolytic Activity by Reducing Hydrophobicity While Maintaining Antimicrobial Activity*** The substitution of two Leu residues on the nonpolar face of melittin (Leu 6 and Leu 13) to Ala residues dramatically decreased hemolytic activity without impairing gram-positive and gram-negative activity. This analog is still induced into the α-helical structure in the hydrophobic environment of lipid vesicles but did not assemble into aggregates like the native melittin [298]. Decreasing hydrophobicity on the nonpolar face disrupts self-association and eliminates hemolytic activity. On the other hand, a peptide with excellent antimicrobial activity and extremely low hemolytic activity (no hemolysis at 500 μg/mL after 8 h; [132]) shows a systematic increase in hemolytic activity with a systematic increase in hydrophobicity of the nonpolar face (Fig. 17) [132]. Increasing hydrophobicity also increases antimicrobial activity until an optimum hydrophobicity is reached (Fig. 17, [132]) (see discussion later in Section 4.4, a general specificity determinant for amphipathic α-helical AMPs). Similarly, Lee et al. [299] designed analogs of PMAP-23 (Table 1) to increase fungicidal activity against pathogenic fungal cells. Three hydrophilic residues in the native peptide were substituted with Trp residues (R10W, Q13W, and K14W). These substitutions increased overall hydrophobicity (see hydrophobicity scale of [124]) and caused a dramatic increase in the fungicidal activity and antitumor activity. However, these authors did not report the hemolytic activity for this analog, which normally increases with increasing hydrophobicity, as shown in Fig. 17. Interestingly, two of the Trp substitutions lie between two Pro residues (PQKP in native PMAP-23 to PWWP in the analog). Since Pro residues disrupt the α-helical structure, the effect on amphipathicity cannot be calculated.

***Incorporation of Hinge Regions and Pro Residues Can Decrease Hemolytic Activity While Maintaining Antimicrobial Activity*** Hinge sequences like GIG or Pro have been shown to reduce or eliminate hemolytic activity dramatically (see Section 4.1 on hybrid amphipathic α-helical AMPs and de novo designed amphipathic α-helical AMPs). Hancock and co-workers [300] studied a series of α-helical AMPs without proline or with one or two prolines. Though the analogs with and without Pro were not identical in the remainder of the sequence, they were highly homologous. The peptide that lacked Pro exhibited the highest antimicrobial activity, with MICs of 0.5–4 μg/mL against all tested gram-negative and gram-positive bacteria, and was very hemolytic at 8 μg/mL (concentration required for lysis after 4 h was taken as the minimal hemolytic concentration). The peptide with a single Pro at position 12 had a hemolytic activity of >128 μg/mL with some loss in antimicrobial activity [300]. The native fowlicidin-1 and fowlicidin-3 have had their NMR structures determined in 50% trifluoroethanol. The peptides adopt a predominately α-helical structure with a slight kink induced by a Gly residue at position 16 of fowlicidin-1 and position 17 of fowlicidin-3 [51, 53]. However, these peptides still have very high hemolytic activity (Table 1). Obviously, Gly does not have the same effect on hemolytic activity compared to a hinge region like GIG or Pro.

***Reduction or Complete Removal of Hemolytic Activity by the Incorporation of D-Amino Acid Residues While Maintaining Antimicrobial Activity*** Melittin B, a 26-residue amphipathic α-helical AMP, is extremely hemolytic (100% hemolysis at 1 μg/mL) (Table 1). Substitution of 4 D-amino acids into the melittin sequence (D – $V^5$, $V^8$, $I^{13}$, and $K^{21}$) for their L-residues dramatically reduced α-helical structure in 40% TFE from 73% α-helical content for melittin to 15% for the above D-analog [190]. The D-analog lost its cytotoxic effects on mammalian cells and retained its antibacterial activity. NMR structural analysis on this D-analog revealed an amphipathic α-helix only in the C-terminal region of the diastereomer in TFE/water and methanol solutions and in DPC/DMPG (dimyristoyl phosphatidylglycerol) micelles [301]. ATR–FTIR experiments showed that melittin and the D-analog are oriented parallel to the membrane surface. They concluded that the N-terminal helix in melittin is crucial for binding and insertion into zwitterionic membranes and for hemolytic activity [301]. Thus, with a disrupted N-terminal helix hemolytic activity is lost. Chen et al. [302] made 19 L- and D- single amino acid substitutions in the center of a model amphipathic α-helix to determine the variation in helix disruption by each substitution in aqueous and hydrophobic conditions. Each single L-amino acid

substitution has a profound effect on α-helical structure under aqueous conditions. In contrast, in the presence of the α-helix inducing solvent TFE, the maximum α-helical content of the L-analogs was essentially the same, excluding Pro. All of the D-substituted peptides exhibited negligible secondary structure under aqueous conditions underlining the helix-disrupting properties of single D-amino acid substitutions. However, in the presence of 50% TFE these D-analogs (except for Pro) all exhibited considerable induction of α-helical structure similar to their L-counterparts. Thus, the disruption of α-helical structure by D-amino acid substitutions in a hydrophobic environment depends on the number of substitutions and the original helical propensity of the α-helix. The results of Chen et al. [302] showed the feasibility of controlled disruption of α-helical peptides by D-amino substitution in aqueous medium, while still allowing full folding in a more hydrophobic environment. We have also discussed the value of D-amino acid substitutions to remove hemolytic activity in the de novo designed amphipathic α-helical AMPs section above. In addition, we utilized single D-amino acid substitutions in amphipathic cyclic β-sheet and α-helical AMPs in the sections below.

***Substitution of a Single Positively Charged Residue in the Center of the Nonpolar Face of Amphipathic AMPs to Reduce Hemolysis*** This is discussed in Section 4.4.

### 4.4 A General Specificity Determinant of Cell Selectivity

Our approach was to show that a single substitution in the sequence of amphipathic cyclic β-sheet and α-helical AMPs with strong toxicity to human cells (as measured by hemolytic activity to erythrocytes) could generate specificity for prokaryotic cells.

***Amphipathic Cyclic β-Sheet AMPs*** Gramicidin S (GS) is a 10-residue, head-to-tail amphipathic cyclic peptide with the sequence Val-Orn-Leu-D-Phe-Pro-Val-Orn-Leu-D-Phe-Pro, consisting of 2 type II' β turns defined by Xxx-D-Phe-Pro-Xxx sequence with an amide bond between the N and C termini (Fig. 14). GS has excellent antimicrobial activity against gram-negative and gram-positive bacteria and yeast, as well as strong hemolytic activity (100% hemolysis at 12.5 μg/mL) [303]. In order to dissociate antimicrobial and hemolytic

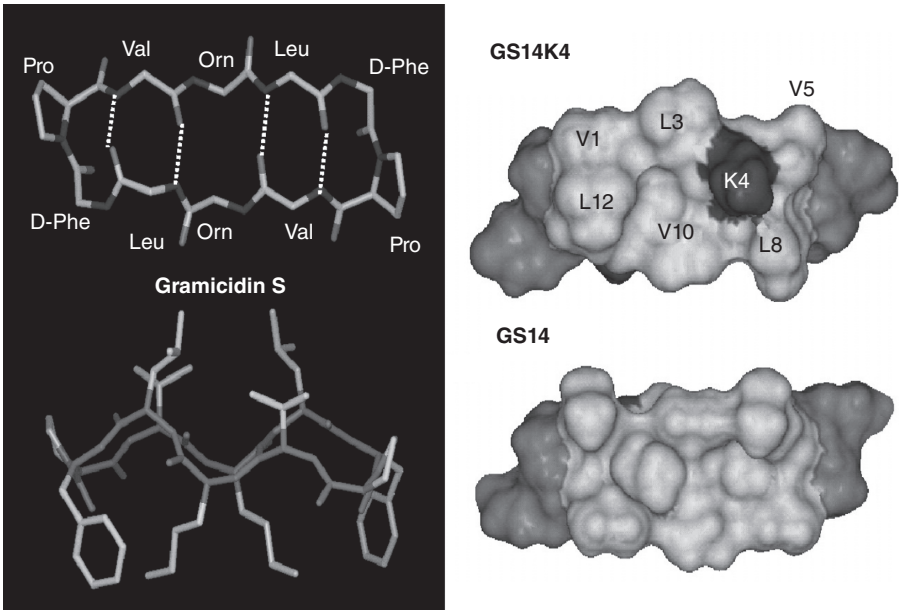


**Figure 14** Structure of Gramicidin S and related analogs. (Left) Stick diagrams of cyclic β-sheet antimicrobial peptide gramicidin S. Dotted lines represent intramolecular hydrogen bonds. The bottom left shows the 4 hydrophobes (2 Val and 2 Leu residues on the top and 2 Orn residues on the bottom). (Right) Connolly surface representation of ring analogs GS14 (bottom) and GS14K4 (top) that has D-Lys substituted for L-Lys at position 4 (dark gray). Green represents the hydrophobic face (light gray).

activities, we developed a ring size analog of GS called GS14. This 14-residue amphipathic cyclic β-sheet peptide with the sequence Val-Lys-Leu-D-Tyr-Pro-Val-Lys-Leu-Lys-Val-D-Tyr-Pro-Leu-Lys had 6 large hydrophobes (3 Val and 3 Leu residues) on the nonpolar face and 4 lysine residues on the polar face. This peptide showed very weak antibiotic activity (>200 μg/mL) and extremely strong hemolytic activity (100% hemolysis at 1.5 μg/mL) for a very low undesired therapeutic index (MHC/MIC = 1.5 μg/mL/>200 μg/mL = <0.01). This analog represented an important milestone in our rational drug design approach in that we were able to dissociate antimicrobial and hemolytic activity. However, a true test of our ability to design an analog with potential clinical applications would be to invert the properties of GS14 and produce a compound with excellent antimicrobial activity but weak hemolytic activity (high therapeutic index). Nevertheless, this important study represented the first demonstration of the dissociation of antimicrobial and hemolytic activity in GS analogs, demonstrating that modulation of the biological activity of cyclic GS analogs could be achieved by variations in ring size [304]. Interestingly, GS14 bound to LPS with the highest affinity of the ring size analogs we prepared and with similar affinity to polymixin B [304]. The high affinity to LPS and the low antimicrobial activity suggested that the rigid β-sheet structure, high amphipathicity, and high positive charge of GS14 locked the peptide on the negatively charged surface of bacteria, thus preventing penetration of the peptide into the hydrophobic bilayer needed for antimicrobial activity. We reasoned that since amphipathicity is intimately linked to β-sheet structure in GS14, any change in β-sheet structure, for example, through disruption by substitution of residues with their enantiomers, would directly affect the amphipathicity of the molecule. We made a series of single enantiomeric substitutions where the peptides retained the same sequence, intrinsic hydrophobicity, and positive charge, differing only in structure. The extent of disruption of the β-sheet structure in the GS14 diastereomers was confirmed/demonstrated by both NMR and CD analyses of GS14 and GS14 analogs [303, 305]. This study was highly significant in demonstrating that a pre-formed highly amphipathic molecule in aqueous medium is not desirable in these cyclic peptides, since this results in decreased specificity and a low therapeutic index of GS14 as well as increased interactions with outer membrane components [303]. We identified a peptide called GS14K4 (D-Lys substitution for L-Lys at position 4 in the cyclic peptide) with a high therapeutic index (65 against *P. aeruginosa*, which is a 6500-fold improvement over GS14, which had a therapeutic index <0.01). There was a 133-fold decrease in hemolytic activity for GS14K4 compared to GS14. We were able to maintain or improve antimicrobial activity of our analog GS14K4 compared to GS while obtaining large increases in the therapeutic index (16- to 35-fold for gram-negative organisms and 8- to 33-fold for gram-positive organisms and 10-fold for the yeast *C. albicans* [303]. Structural studies by our group on GS14K4 in water and in the membrane mimetic solvent, trifluoroethanol, clearly showed the induction of β-sheet character in a membrane environment and the positioning of the charged side chain of D-Lys on the hydrophobic face of the cyclic framework (Fig. 14), which explains the decrease in amphipathicity and the decrease in effective hydrophobicity [305]. We then showed that it was possible to improve the therapeutic index of GS14K4 further, that is, improving the antimicrobial activity and/or decreasing the hemolytic activity or both by modulating the hydrophobicity on the nonpolar face [306]. These studies showed that the placement of a single positive charge in the center of the nonpolar face of a amphipathic cyclic β-sheet peptide dramatically altered specificity and decreased dimerization/oligomerization [127, 307]. The next question to be addressed was the generality of this specificity determinant (positively charged residue in the center of the nonpolar face). Therefore, we decided to take the same approach with amphipathic α-helical AMPs.

***Amphipathic α-Helical AMPs*** Our objectives were threefold: First, to test the hypothesis that disruption of α-helical structure in aqueous conditions by D-amino acid substitutions or substitutions of hydrophilic/positively charged L-amino acids on the nonpolar face can dramatically alter specificity in a similar manner to our results with cyclic β-sheet AMPs [303, 308]; second, to observe if these substitutions would enhance antimicrobial activity, decrease or eliminate toxicity, and improve antimicrobial specificity while maintaining broad-spectrum activity for gram-negative and gram-positive bacteria; and third, to demonstrate the importance of peptide self-association or lack of self-association in the activity of amphipathic α-helical AMPs. To test the above objectives, we selected an ideal candidate, a helical peptide with excellent antimicrobial activity and strong hemolytic activity from previous published research. The peptide referred to as V681 (Table 5A) resulted from semirandom DNA mutagenesis of an analog of a cecropin A (1–8) melittin B (1–18) hybrid [300, 309]. V681, a 26-residue peptide, was highly hemolytic (lysed human red blood cells at 8 μg/mL) and exhibited a very high activity against all gram-negative and gram-positive bacteria that were tested (MIC values ranging from 0.5 to 4 μg/mL) [300]. We previously demonstrated that the helix-destabilizing properties of D-amino acids offered a systematic approach to the controlled alteration of the hydrophobicity, amphipathicity, and helicity of amphipathic α-helical model pep-

tides. By single substitutions of different D-amino acids into the center of the nonpolar face of an amphipathic α-helical model peptide, we demonstrated that different D-amino acids disrupted α-helical structure to different degrees, while the destabilized structure could still be induced to fold into an α-helix in hydrophobic medium [302]. The advantage of the approach of single D- or L-amino acid substitutions at a specific site in the peptide is that it enables a greater understanding of the relationship of structure and mechanism of action of amphipathic α-helical AMPs. Chen et al. (2005) [130] chose D-/L-amino acid substitutions sites in the center of the nonpolar face (position 13) and polar face (position 11) of the helix of V681 (Figs. 15 and 16). The amino acid residues substituted at these positions represented a wide range of hydrophobicity from Leu > Val > Ala > Gly > Ser > Lys [124]. Leucine was used to replace the valine residue on the nonpolar face of V681 to increase peptide hydrophobicity and amphipathicity; alanine was selected to reduce peptide hydrophobicity/amphipathicity while maintaining high helicity; a hydrophilic amino acid, serine, was selected to decrease the hydrophobicity/amphipathicity of V681 in the nonpolar face; positively charged lysine was used to decrease further peptide hydrophobicity and amphipathicity. In our previous studies, we successfully utilized D-amino acid substitutions to remove the hemolytic activity of the cyclic β-sheet analogs [303, 305, 308]. In the V681 study, we also used D-enantiomers of the five L-amino acid residues Leu, Val, Ala, Ser, and Lys to change not only peptide hydrophobicity and amphipathicity but, more importantly, to disrupt peptide helical structure [130]. The results clearly showed that, for the nonpolar face substituted peptides, hemolytic activity was correlated with side-chain hydrophobicity of the substituting amino acid residues ($L_L > V_L > A_L > S_L > K_L$), that is, the more hydrophobic the substituting amino acid, the more hemolytic the peptide, consistent with our previous study on β-sheet AMPs [306]. For example, the maximum peptide concentration to show no detectable hemolytic activity (MHC value) for the $L_L$ substitution at position 13 (Fig. 15) was 7.8 μg/mL; in contrast, the MHC value was decreased, parallel with the reduction

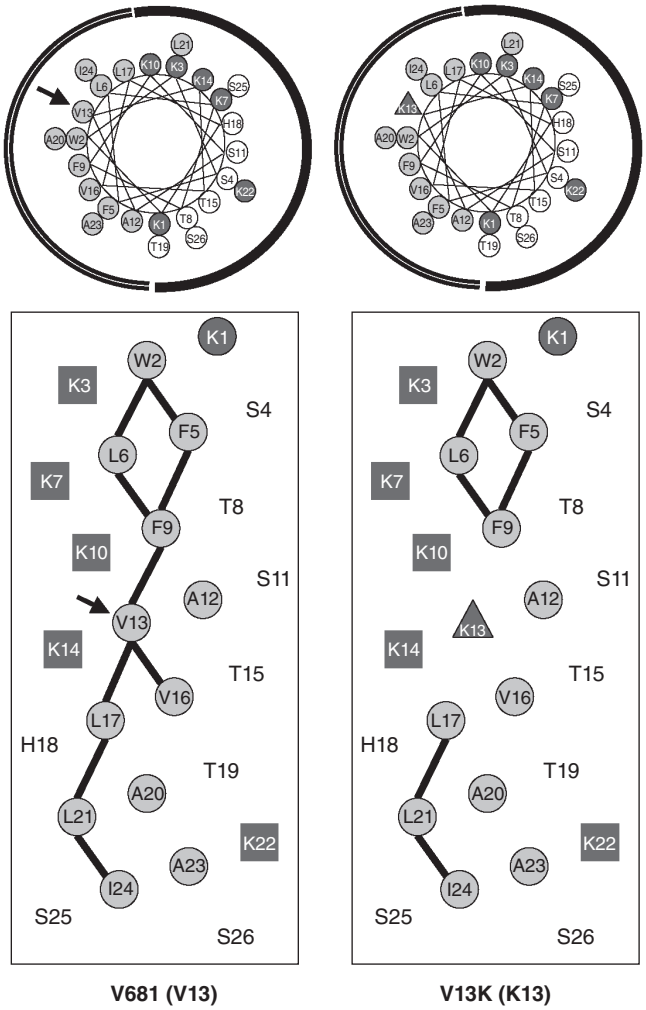


**Figure 15** Helical wheels/helical nets representations of α-helical antimicrobial peptides V681 and V13K. The nonpolar face is indicated by circled hydrophobes (light gray) down the center of the helical net. The black bars between large hydrophobes indicate i to i + 3/i to i + 4 hydrophobic interactions. The positively charged Lys residues are boxed. The black arrow denotes position 13 the site for substitution.

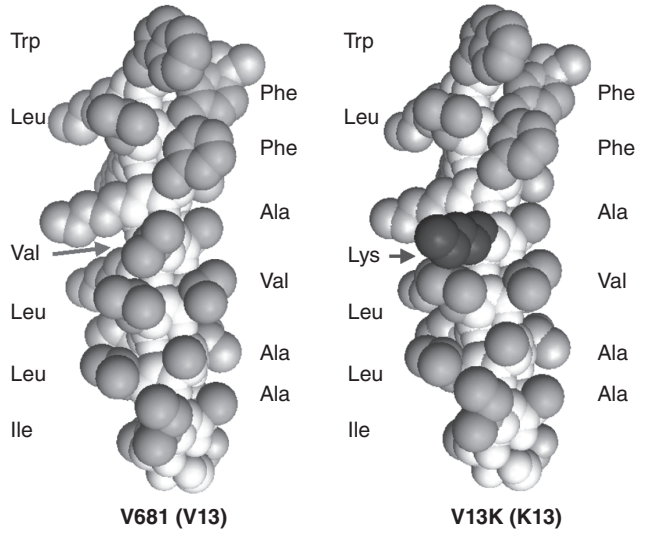


**Figure 16** Space filling model of α-helical antimicrobial peptides V681 and V13K. White represents the peptide backbone; gray shows hydrophobic residues on the nonpolar face; light gray indicates amino acids included in the polar face; and the V13K substitution on the nonpolar face is colored dark gray.

of hydrophobicity, to an undetectable level for the peptide V13K (Fig. 15), a >32-fold decrease in hemolytic activity for V13K compared to V681 at 15.6 μg/mL. Peptide hydrophobicity and amphipathicity on the nonpolar face were also correlated with peptide self-associating ability; thus, peptides with less self-association in aqueous conditions also exhibited less hemolytic activity against eukaryotic cells [130]. In addition, peptide helicity seemed to have an additional effect on hemolytic activity. In general, on both the polar and nonpolar faces, D-amino acid substituted peptides were less hemolytic than their L-diastereomers. For example, $A_L$ at position 13 on the nonpolar face had an MHC value of 31.2 μg/mL compared to $A_D$ with a value of 250 μg/mL, an 8-fold decrease in hemolytic activity. This phenomenon generally correlated with peptide self-association ability since D-diastereomeric analogs exhibited weaker self-associating ability than L-analogs [130]. That is, D-substitutions disrupt helicity, which, in turn, disrupts hydrophobicity on the nonpolar face, which disrupts dimerization and decreases hemolytic activity. Thus, peptide structure is important in cytotoxicity toward mammalian cells, although these disrupted helices in aqueous conditions can still be induced into helical structure in the hydrophobic bilayer and still maintain antibacterial activity.

The antimicrobial activity of the two best peptides with nonpolar face amino acid substitutions against a range of gram-negative and gram-positive microorganisms is shown in Table 7. The two analogs with $K_L$ or $A_D$ substitutions on the nonpolar face showed a significant improvement (~3-fold) in antimicrobial activity against gram-negative bacteria when compared to peptide V681, which already had potent antimicrobial activity. For gram-negative bacteria, disruption of peptide helicity seemed to outweigh other factors in the improvement of antimicrobial activity, that is, in most cases, the peptides with D-amino acid substitutions showed better antimicrobial activity than L-diastereomers [130]. Compared to peptide V681, most of the peptide analogs with increased antimicrobial activity against gram-positive microorganisms were D-amino acid substituted peptides (data not shown; see [130]). The major change in activity profile of these two analogs was the reduction in hemolytic activity (Table 7), which has a dramatic effect on the therapeutic index. Peptide V13K showed a 90-fold and 17-fold improvement in the therapeutic index against gram-negative and gram-positive bacteria, respectively (Table 7). Peptide V13$A_D$ showed a 42-fold and 23-fold improvement in therapeutic index against gram-negative and gram-positive bacteria, respectively. From the helicity data of these peptides, it was clear that all peptides with improved therapeutic indices are those showing less stable helical structure in aqueous conditions (either D-amino acid substituted peptides or the hydrophilic/positively charged amino acid substituted peptides on the nonpolar face). It was noteworthy that the hemolytic activity of these two peptides was not detectable or was extremely weak. As noted above, we developed a novel technique to measure the self-association of small amphipathic molecules referred to as "temperature profiling in RP-HPLC." This technique was developed because we felt that dimerization/oligomerization of amphipathic AMPs in aqueous solution was an important factor in understanding overall antimicrobial activity. That is, the dimer/oligomer would have to dissociate to monomer to enter the hydrophobic environment of the bilayer, and it is the monomer that generates the antimicrobial activity. If the self-association ability of a peptide in aqueous media is too strong (forming dimers and burying the nonpolar face), it could decrease the ability of the peptide to dissociate and penetrate into the biomembrane and to kill target cells. In addition, if we can prevent dimerization in aqueous solution but still maintain the amphipathicity/hydrophobicity of the peptide in the bilayer, we would enhance activity. High peptide hydrophobicity, amphipathicity, and helicity usually

**TABLE 7 Effect of Amino Acid Substitutions on Biological Activity of V681**

| | Gram negative | | | | | | Gram Positive | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Peptide[a] | MIC[b] (μg/mL) | Fold[c] | Ther. Index[e] | Fold[c] | MHC[d] (μg/mL) | Fold[c] | MIC[b] (μg/mL) | Fold[c] | Ther. Index[e] | Fold[c] |
| **$V_L$(V681)** | 8.8 | 1.0 | 1.8 | 1.0 | 15.6 | 1.0 | 6.3 | 1.0 | 2.5 | 1.0 |
| **$K_L$(V13K)** | 3.1 | 2.8 | 163.0 | 90.6 | >250.0 | 32.1 | 11.8 | 0.5 | 42.3 | 16.9 |
| **$A_D$(V13$A_D$)** | 3.3 | 2.7 | 75.7 | 42.1 | 250.0 | 16.0 | 4.3 | 1.5 | 57.8 | 23.1 |

[a]Substitutions at position 13 in the center of the nonpolar face.
[b]MIC values reported are the geometric mean of 6 gram-negative and 6 gram-positive microorganisms tested [130].
[c]Fold improvement in MIC, therapeutic index, and MHC compared to V681.
[d]Hemolytic activity (MHC) was determined on human red blood cells and is the maximum peptide concentration that results in no hemolysis.
[e]Therapeutic (Ther.) index = MHC (μg/mL)/geometric mean of MIC (μg/mL). Larger values indicate greater antimicrobial specificity.

result in strong hemolytic activity. The controlled disruption of the α-helical structure (disruption under aqueous conditions and inducible in hydrophobic conditions) seems to be also related to strong antimicrobial activity in a variety of gram-negative and gram-positive bacteria.

It is generally accepted that increasing hydrophobicity of the nonpolar face of amphipathic α-helical peptides increases antimicrobial activity [310–312]. The striking result from our antibacterial assays was that hydrophobicity has two effects on peptide antimicrobial activity: at a relatively lower level of hydrophobicity, increasing peptide hydrophobicity caused an improvement in antimicrobial activity until an optimal hydrophobicity was reached; in contrast, peptide antimicrobial activity was weakened dramatically with further increases of hydrophobicity beyond the optimum, even resulting in dramatic loss of antimicrobial activity (Fig. 17) [132]. The reason for loss of antimicrobial activity at high hydrophobicity is that high hydrophobicity causes stronger peptide self-association/dimerization in solution, which in turn results in the monomer–dimer equilibrium favoring the dimer conformation. Peptide dimers are in their folded α-helical conformation and would be inhibited from passing through the capsule and cell wall to reach the target membranes compared to an unstructured monomer. In contrast, there is no polysaccharide-based cell wall in eukaryotic cells and increased hydrophobicity continues to increase hemolysis of human red blood cells where hydrophobicity of the bilayer causes rapid dissociation of dimers to monomers and entry into the bilayer to form channels/pores. In addition, the higher content of zwitterionic phospholipids and large amount of cholesterol in eukaryotic cell membranes compared to bacterial membranes may also supply a more hydrophobic environment, which would promote dimer-to-monomer dissociation and enhance hemolytic activity. The peptides used in Figure 17 are described in Chen et al. [132] where systematic increases in hydrophobicity were made by changing A12, A20, and A23 on the nonpolar face of V13K (Fig. 15) to Leu residues as single, double, and triple substitutions. Clearly, the placement of a lysine residue in the center of the nonpolar face destabilizes transmembrane helix formation in human red blood cells. However, the effect of this specificity determinant can be overridden by increasing the hydrophobicity of the nonpolar face of V13K (Fig. 15) by changing A12, A20, and A23 to Leu residues. The hemolytic activity then increases by 62.5-fold (no hemolysis at 250 μg/mL for V13K [132]. Increasing hydrophobicity allows transmembrane penetration of the helix by stabilizing pore/channel formation by increasing hydrophobic interactions of the hydrophobic surface of the pore/channel with the

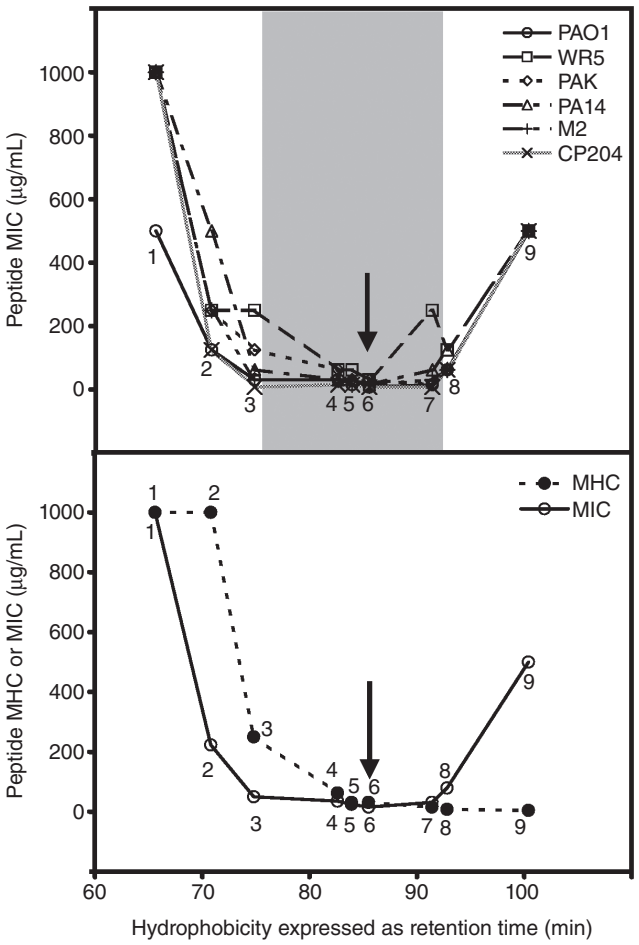


**Figure 17** Relationship between peptide hydrophobicity, hemolysis (MHC), and antibacterial activity (MIC). (Top) The MIC values for six different strains of *P. aeruginosa*. The shaded panel shows the optimal hydrophobicity zone for antimicrobial activity. (Bottom) The MIC value plotted is based on the geometric mean of the MICs for the six *P. aeruginosa* clinical isolates versus peptide hydrophobicity. The arrow denotes the optimal antimicrobial activity. MHC is the maximal peptide concentration that caused no hemolysis of human erythrocytes after 18 h. MIC is the lowest peptide concentration that inhibits growth after 24 h [132].

hydrophobic lipid. Similarly, increasing hydrophilicity/positive charge on the polar face can also stabilize transmembrane pore formation [297]. Thus, it seems reasonable to propose that both surfaces of a transmembrane pore/channel could contribute to formation and stability of such a pore/channel. Stability of such a pore/channel is dependent on the hydrophobicity of the nonpolar face and hydrophilicity of the polar face.

Factors believed to be important for antimicrobial activity have been identified and include peptide hydrophobicity, the presence of positively charged residues on the polar face, amphipathicity that segregates basic and

hydrophobic residues to opposite sides of the molecule and secondary structure. Our laboratory has added to this list, which includes: (1) the importance of lack of structure in aqueous conditions but inducible structure in the presence of hydrophobic environment of the membrane; (2) the presence of a positively charged residue in the center of the nonpolar face of amphipathic cyclic β-sheet and amphipathic α-helical AMPs as a determinant for locating peptides to the interface region of prokaryotic membranes and decreasing transmembrane penetration into eukaryotic membranes; (3) the importance of peptide self-association in aqueous environment to biological activity; and (4) we have recently shown that if the surface hydrophobicity is too large the peptides actually become less active, that is, increasing hydrophobicity decreases antimicrobial activity at some point, which is due to increased dimerization/self-association [132]. It is interesting that nature did not design AMPs for selectivity (i.e., to be nontoxic to human cells). Our de novo designed specificity determinant (positively charged Lys residue in the center of the nonpolar face) achieves three goals at once: (a) prevents dimerization in aqueous solution; (b) disrupts the secondary structure in aqueous solution; and (c) provides specificity for prokaryotic cells. We hypothesize that the disruption of secondary structure and keeping the peptide monomeric allows the peptide to more easily pass through the capsule and cell wall to reach the membrane and allows the secondary structure to be induced in the hydrophobic environment of the bilayer, which is critical for antimicrobial activity. Our de novo designed determinant of specificity has not been observed in any AMPs to date, and that is why we believe it is such a key discovery to the development of AMPs as therapeutics.

***Proposed Mechanism for Cell Selectivity of Amphipathic Cyclic β-Sheet and α-Helical AMPs*** We have proposed a "membrane discrimination mechanism" of action for amphipathic cyclic β-sheet and amphipathic α-helical AMPs whose sole target is the biomembrane to explain peptide specificity between prokaryotic and eukaryotic membranes. Peptide specificity depends upon the compositional difference in the lipids between the two types of membranes where eukaryotic cell membranes, in contrast to prokaryotic membranes, are generally characterized by zwitterionic phospholipids, a relatively large amount of cholesterol and sphingomyelin, and the absence of the high, inside negative transmembrane potential that is present in prokaryotic membranes. The outer surfaces of bacterial membranes are devoid of cholesterol and contain significant amounts of negatively charged (anionic) phospholipids in the outer leaflet of the lipid bilayer [268, 313, 314]. Our membrane discrimination mechanism is based on a "barrel-stave" mechanism [228] in eukaryotic cells and a "carpet" mechanism [233, 315] in prokaryotic cells. This model proposes that hemolysis of human red blood cells by amphipathic α-helical peptides requires insertion of the peptides into the membrane, perpendicular to the membrane surface, and interaction of the nonpolar face of the amphipathic α-helix with the hydrophobic lipid core of the bilayer while the hydrophilic charged surface of the polar face points inward, producing an aqueous transmembrane channel/pore (barrel-stave mechanism). In contrast, amphipathic α-helical peptides with cell selectivity produce antimicrobial activity in prokaryotic cells by lying at the membrane interface parallel with the membrane surface; interaction of the nonpolar face of the amphipathic α-helix with the hydrophobic component of the lipid and interaction of polar/positively charged face with the negatively charged head groups of the phospholipids causes membrane disruption in a detergent-like mechanism as described in the carpet model. In support of this proposal, we have selected two examples from the study of Chen et al. [130]. Peptide V13K (in its all-D form) (Figs. 15 and 16), which has Lys in the center of the nonpolar face shows no hemolysis at 500 μg/mL after 8 h [131] (Fig. 11) yet has excellent broad-spectrum antimicrobial activity [130, 131]. If hemolysis of human red blood cells requires transmembrane insertion of the peptide in the membrane, it seems reasonable to assume that disruption of the hydrophobic surface with a positive charge in the center of the nonpolar face would prevent penetration into the hydrophobic core of the membrane. Thus, no hemolysis for V13K. On the other hand, if the mechanism for antimicrobial activity in prokaryotic cells does not require transmembrane insertion of the peptide into the hydrophobic bilayer, then the peptide can locate to a much lower hydrophobic environment in the interface region and still effectively disrupt the membrane. In contrast to V13K, this same sequence with a Leu substitution at position 13 instead of Lys results in poorer antimicrobial activity (MIC values of 12.7 μg/mL of V13L versus 3.1 μg/mL for V13K against gram-negative bacteria) and stronger hemolytic activity (7.8 μg/mL for V13L versus no detectable hemolytic activity for V13K), which can also be explained by our combined model. Peptide V13L has a continuous nonpolar face and greater hydrophobicity of the nonpolar face than V13K. Thus, it can insert transmembrane into the red blood cell membrane, causing significant hemolysis. Because of the properties of V13L, it actually interacts less with the water–lipid interface region of the bacterial membrane (it is not restricted to this preferred location), hence the antimicrobial activity is fourfold weaker than V13K against gram-negative bacteria. This

supports the view that the carpet mechanism is essential for strong antimicrobial activity and, if there is a preference by the peptide for deep penetration into the hydrophobic core of the bilayer, the antimicrobial activity will actually decrease.

### 4.5 Enhancing Stability

In 1990, Merrifield and co-workers [279] were the first to show that the all-D enantiomers of naturally occurring AMPs, cecropin A, magainin 2 amide, and melittin along with cecropin–melittin hybrid peptides were structurally identical to their L-counterparts as shown by circular dichroism measurements, but the ellipticities were of opposite sign. In addition, in all the isomeric pairs of peptides, the antimicrobial activity of the D-enantiomer was quantitatively equivalent to that of the L-enantiomer. This suggested that the antimicrobial activity does not involve a stereoselective interaction with a chiral enzyme or lipid or protein receptor. The major advantage of the D-enantiomers is their resistance to enzyme degradation. In cases where the L-enantiomers have been shown to be significantly less active than the D-enantiomer, the differences implicated the stability of the D-enantiomer to proteolytic enzymes produced by the organisms. The above results have been well documented for many other AMPs ([131, 279, 291, 316–322]). From Fig. 18, the CD spectra are shown for two pairs of L- and D-enantiomers. The substitution of a lysine residue in the center of the nonpolar face of the peptide V681 completely abolished α-helical structure in aqueous conditions (50 m*M* potassium phosphate buffer containing 100 m*M* KCl) but allowed the α-helical structure to be induced in the presence of a hydrophobic environment, 50% trifluoroethanol in aqueous buffer. As discussed earlier, the incorporation of D-amino acids into an L-peptide can be effective at inducing selectively between eukaryotic and prokaryotic cells, but the all D-enantiomers are much preferred because of their stability to proteolysis.

Casteels and Tempost [113] showed that the all-D enantiomer of apidaecin is completely devoid of antibacterial activities, which suggested that the effects of apidaecin involve a stereoselective recognition of a chiral cellular target. This peptide was functionally unique among insect antibacterials. Apidaecin, an 18-residue peptide, had a unique sequence of 33% proline residues (GNNRPVYIPQPRPPHPRL).

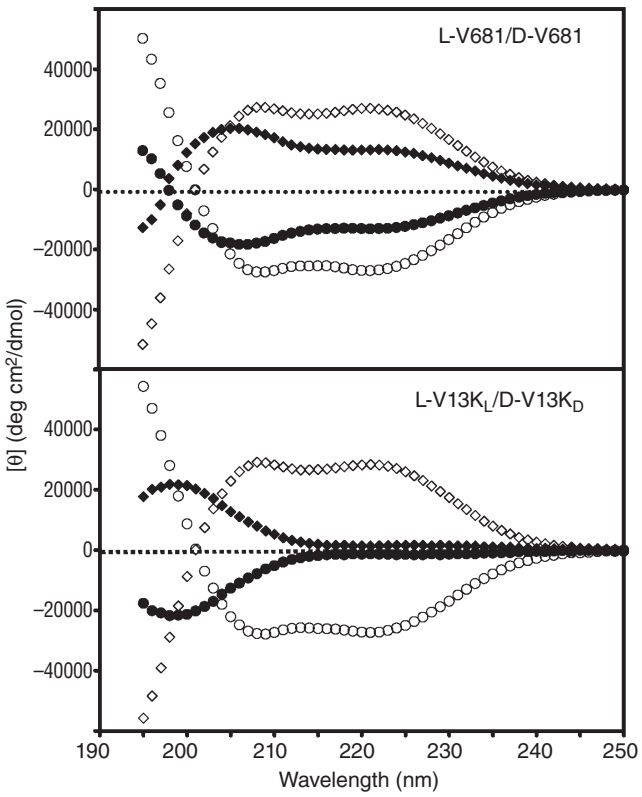


**Figure 18** Circular dichroism (CD) spectra of V681 and D-V681 (Top) and V13K$_L$ and D-V13K$_D$ peptides (Bottom). Circles denote L-peptides and diamonds denote D-peptides. Solid symbols show spectra obtained for peptides in 50 m*M* aq. potassium phosphate buffer containing 100 m*M* KCl without triflouroenthanol (TFE), and open symbols show spectra collected in the presence of aqueous buffer containing 50% TFE. (See [131] for more detail.)

## 5 THERAPEUTIC DEVELOPMENT (PRECLINICAL AND CLINICAL STUDIES)

During 2005, 8987 cases of invasive MRSA with 1598 associated deaths were reported in the United States [323]. New classes of antibiotics are needed to protect against the growing MRSA threat and other challenging infections such as *P. aeruginosa* in cystic fibrosis patients and candidiasis in persons with AIDS. AMPs represent a relatively new therapeutic class with potential impact in these areas. Not surprisingly, academic and industrial interest in therapeutic development of AMPs continues to grow. The development record of AMPs has been mixed to date (Table 8). Late-stage trials of Intrabiotics' peptide Iseganan for ventilator-associated pneumonia were halted for safety concerns, and Xoma's rBPI$_{21}$ failed to reach statistically significant efficacy in a phase III trial. However, select AMPs have successfully completed clinical development, gained regulatory approval, and entered the clinic. Polymyxin B is currently used in multiple topical and ophthalmic therapies including Polysporin First Aid Antibiotic Ointment and Neosporin Ophthalmic Solution. In the antiviral area, Trimeris' peptide Enfuvirtide defined the viral fusion

**TABLE 8 AMPs in Clinical Development**

| Peptide/Product | Company | Indication | Development Stage |
|---|---|---|---|
| | | *Marketed* | |
| Enfuvirtide/Fuzeon | Trimeris | HIV | Approved |
| Polymyxin B | Pfizer | Bacterial eye infection, *P. aeruginosa* infection | Approved |
| | | *Clinical Development* | |
| MBI-226 | Migenix/Cadence Pharmaceuticals | Wound/burn/device-related infection | Phase III completed |
| $rBPI_{21}$/Neuprex | Xoma | Severe meningococcal sepsis | Phase III inconclusive |
| Iseganan | Intrabiotics/Pacgen | Ventilator-associated pneumonia | Failed phase III |
| P113 | Demegen/Pacgen | Lung infections in CF, Oral candidiasis in immunocompromised | Phase II complete |
| hLF-11 | AM-Pharma Holding | 1. Prevention of infections during hematopoietic stem cell transplant | 1. Phase IIa |
| | | 2. MRSA/MRSE | 2. Phase I completed |
| | | 3. Candida | 3. Phase I completed |
| | | *Preclinical* | |
| PTX005 | PepTx | *P. aeruginosa* infection, LPS neutralization | Preclinical |
| Mersacidin | Novacta Biosystems | MRSA | Unknown |

inhibitor therapeutic class to block HIV entry into CD4 T cells. Given the variety of structural peptide motifs available for rational drug design and the low natural bacterial resistance, preclinical and clinical investigations will continue to explore possible AMP therapeutic applications.

## ACKNOWLEDGMENT

This work was supported by an NIH grant from National Institute of Allergy and Infectious Diseases (NIAID) (RO1 AI067296) and the John Stewart Chair in Peptide Chemistry to Robert S. Hodges. The content is solely the responsibility of the authors and does not necessarily represent the official view of NIAID and NIH.

## REFERENCES

1. Neu HC (1992). The crisis in antibiotic resistance. *Science 257*, 1064–1073.
2. Bryan LE (1988). General mechanisms of resistance to antibiotics. *J Antimicrob Chemother 22*(Suppl A), 1–15.
3. Marshall SH, Arenas G (2003). Antimicrobial peptides: A natural alternative to chemical antibiotics and a potential for applied biotechnology. *Electron J Biotechnol 6*, 271–284.
4. Jenssen H, Hamill P, Hancock RE (2006). Peptide antimicrobial agents. *Clin Microbiol Rev 19*, 491–511.
5. Steinberg DA, Hurst MA, Fujii CA, Kung AH, Ho JF, Cheng FC, Loury DJ, Fiddes JC (1997). Protegrin-1: A broad-spectrum, rapidly microbicidal peptide with in vivo activity. *Antimicrob Agents Chemother 41*, 1738–1742.
6. Ge Y, MacDonald DL, Holroyd KJ, Thornsberry C, Wexler H, Zasloff M (1999). In vitro antibacterial properties of pexiganan, an analog of magainin. *Antimicrob Agents Chemother 43*, 782–788.
7. Zhang L, Parente J, Harris SM, Woods DE, Hancock RE, Falla TJ (2005). Antimicrobial peptide therapeutics for cystic fibrosis. *Antimicrob Agents Chemother 49*, 2921–2927.
8. Hancock RE, Diamond G (2000). The role of cationic antimicrobial peptides in innate host defences. *Trends Microbiol 8*, 402–410.
9. Hancock RE, Scott MG (2000). The role of antimicrobial peptides in animal defenses. *Proc Natl Acad Sci USA 97*, 8856–8861.
10. Tencza SB, Creighton DJ, Yuan T, Vogel HJ, Montelaro RC, Mietzner TA (1999). Lentivirus-derived antimicrobial peptides: Increased potency by sequence engineering and dimerization. *J Antimicrob Chemother 44*, 33–41.
11. Tencza SB, Douglass JP, Creighton DJ Jr, Montelaro RC, Mietzner TA (1997). Novel antimicrobial peptides derived from human immunodeficiency virus type 1 and other lentivirus transmembrane proteins. *Antimicrob Agents Chemother 41*, 2394–2398.

12. Wang Z, Wang G (2004). APD: The Antimicrobial Peptide Database. *Nucleic Acids Res 32*, D590–D592.
13. Bulet P, Hetru C, Dimarcq JL, Hoffmann D (1999). Antimicrobial peptides in insects; structure and function. *Dev Comp Immunol 23*, 329–344.
14. Koczulla AR, Bals R (2003). Antimicrobial peptides: Current status and therapeutic potential. *Drugs 63*, 389–406.
15. Andreu D, Rivas L (1998). Animal antimicrobial peptides: An overview. *Biopolymers 47*, 415–433.
16. Bulet P, Stocklin R, Menin L (2004). Anti-microbial peptides: From invertebrates to vertebrates. *Immunol Rev 198*, 169–184.
17. Epand RM, Vogel HJ (1999). Diversity of antimicrobial peptides and their mechanisms of action. *Biochim Biophys Acta 1462*, 11–28.
18. Gallo RL, Huttner KM (1998). Antimicrobial peptides: An emerging concept in cutaneous biology. *J Invest Dermatol 111*, 739–743.
19. Nissen-Meyer J, Nes IF (1997). Ribosomally synthesized antimicrobial peptides: Their function, structure, biogenesis, and mechanism of action. *Arch Microbiol 167*, 67–77.
20. Risso A (2000). Leukocyte antimicrobial peptides: Multifunctional effector molecules of innate immunity. *J Leukoc Biol 68*, 785–792.
21. Sitaram N, Nagaraj R (2002). Host-defense antimicrobial peptides: Importance of structure for activity. *Curr Pharm Des 8*, 727–742.
22. Toke O (2005). Antimicrobial peptides: New candidates in the fight against bacterial infections. *Biopolymers 80*, 717–735.
23. Vizioli J, Salzet M (2002). Antimicrobial peptides from animals: Focus on invertebrates. *Trends Pharmacol Sci 23*, 494–496.
24. Weinberg A, Krisanaprakornkit S, Dale BA (1998). Epithelial antimicrobial peptides: Review and significance for oral applications. *Crit Rev Oral Biol Med 9*, 399–414.
25. Marion D, Zasloff M, Bax A (1988). A two-dimensional NMR study of the antimicrobial peptide magainin 2. *FEBS Lett 227*, 21–26.
26. Gesell J, Zasloff M, Opella SJ (1997). Two-dimensional 1H NMR experiments show that the 23-residue magainin antibiotic peptide is an alpha-helix in dodecylphosphocholine micelles, sodium dodecylsulfate micelles, and trifluoroethanol/water solution. *J Biomol NMR 9*, 127–135.
27. Hirsh DJ, Hammer J, Maloy WL, Blazyk J, Schaefer J (1996). Secondary structure and location of a magainin analogue in synthetic phospholipid bilayers. *Biochemistry 35*, 12733–12741.
28. Matsuzaki K, Murase O, Fujii N, Miyajima K (1995). Translocation of a channel-forming antimicrobial peptide, magainin 2, across lipid bilayers by forming a pore. *Biochemistry 34*, 6521–6526.
29. Matsuzaki K, Murase O, Fujii N, Miyajima K (1996). An antimicrobial peptide, magainin 2, induced rapid flip-flop of phospholipids coupled with pore formation and peptide translocation. *Biochemistry 35*, 11361–11368.
30. Tremouilhac P, Strandberg E, Wadhwani P, Ulrich AS (2006). Conditions affecting the re-alignment of the antimicrobial peptide PGLa in membranes as monitored by solid state 2H-NMR. *Biochim Biophys Acta 1758*, 1330–1342.
31. Apponyi MA, Pukala TL, Brinkworth CS, Maselli VM, Bowie JH, Tyler MJ, Booker GW, Wallace JC, Carver JA, Separovic F, Doyle J, Llewellyn LE (2004). Host-defence peptides of Australian anurans: Structure, mechanism of action and evolutionary significance. *Peptides 25*, 1035–1054.
32. Conlon JM, Kolodziejek J, Nowotny N (2004). Antimicrobial peptides from ranid frogs: Taxonomic and phylogenetic markers and a potential source of new therapeutic agents. *Biochim Biophys Acta 1696*, 1–14.
33. Lequin O, Ladram A, Chabbert L, Bruston F, Convert O, Vanhoye D, Chassaing G, Nicolas P, Amiche M (2006). Dermaseptin S9, an alpha-helical antimicrobial peptide with a hydrophobic core and cationic termini. *Biochemistry 45*, 468–480.
34. Nascimento AC, Fontes W, Sebben A, Castro MS (2003). Antimicrobial peptides from anurans skin secretions. *Protein Pept Lett 10*, 227–238.
35. Nicolas P, Vanhoye D, Amiche M (2003). Molecular strategies in biological evolution of antimicrobial peptides. *Peptides 24*, 1669–1680.
36. Rinaldi AC (2002). Antimicrobial peptides from amphibian skin: An expanding scenario. *Curr Opin Chem Biol 6*, 799–804.
37. Simmaco M, Mignogna G, Barra D (1998). Antimicrobial peptides from amphibian skin: What do they tell us? *Biopolymers 47*, 435–450.
38. Raj PA, Soni SD, Levine MJ (1994). Membrane-induced helical conformation of an active candidacidal fragment of salivary histatins. *J Biol Chem 269*, 9610–9619.
39. Raj PA, Marcus E, Sukumaran DK (1998). Structure of human salivary histatin 5 in aqueous and nonaqueous solutions. *Biopolymers 45*, 51–67.
40. Campagna S, Saint N, Molle G, Aumelas A (2007). Structure and mechanism of action of the antimicrobial peptide piscidin. *Biochemistry 46*, 1771–1778.
41. Chekmenev EY, Jones SM, Nikolayeva YN, Vollmar BS, Wagner TJ, Gor'kov PL, Brey WW, Manion MN, Daugherty KC, Cotten M (2006). High-field NMR studies of molecular recognition and structure-function relationships in antimicrobial piscidins at the water-lipid bilayer interface. *J Am Chem Soc 128*, 5308–5309.
42. Chekmenev EY, Vollmar BS, Forseth KT, Manion MN, Jones SM, Wagner TJ, Endicott RM, Kyriss BP, Homem LM, Pate M, He J, Raines J, Gor'kov PL, Brey WW, Mitchell DJ, Auman AJ, Ellard-Ivey MJ, Blazyk J, Cotten M (2006). Investigating molecular recognition

and biological function at interfaces using piscidins, antimicrobial peptides from fish. *Biochim Biophys Acta 1758*, 1359–1372.

43. Lauth X, Shike H, Burns JC, Westerman ME, Ostland VE, Carlberg JM, Van Olst JC, Nizet V, Taylor SW, Shimizu C, Bulet P (2002). Discovery and characterization of two isoforms of moronecidin, a novel antimicrobial peptide from hybrid striped bass. *J Biol Chem 277*, 5030–5039.
44. Lee SA, Kim YK, Lim SS, Zhu WL, Ko H, Shin SY, Hahm KS, Kim Y (2007). Solution structure and cell selectivity of piscidin 1 and its analogues. *Biochemistry 46*, 3653–3663.
45. Silphaduang U, Noga EJ (2001). Peptide antibiotics in mast cells of fish. *Nature 414*, 268–269.
46. Cole AM, Weis P, Diamond G (1997). Isolation and characterization of pleurocidin, an antimicrobial peptide in the skin secretions of winter flounder. *J Biol Chem 272*, 12008–12013.
47. Mason AJ, Chotimah IN, Bertani P, Bechinger B (2006). A spectroscopic study of the membrane interaction of the antimicrobial peptide Pleurocidin. *Mol Membr Biol 23*, 185–194.
48. Syvitski RT, Burton I, Mattatall NR, Douglas SE, Jakeman DL (2005). Structural characterization of the antimicrobial peptide pleurocidin from winter flounder. *Biochemistry 44*, 7282–7293.
49. Marchini D, Bernini LF, Marri L, Giordano PC, Dallai R (1991). The female reproductive accessory glands of the medfly *Ceratitis capitata*: Antibacterial activity of the secretion fluid. *Insect Biochem 21*, 597–605.
50. Ragona L, Molinari H, Zetta L, Longhi R, Marchini D, Dallai R, Bernini LF, Lozzi L, Scarselli M, Niccolai N (1996). CD and NMR structural characterization of ceratotoxins, natural peptides with antimicrobial activity. *Biopolymers 39*, 653–664.
51. Bommineni YR, Dai H, Gong YX, Soulages JL, Fernando SC, Desilva U, Prakash O, Zhang G (2007). Fowlicidin-3 is an alpha-helical cationic host defense peptide with potent antibacterial and lipopolysaccharide-neutralizing activities. *FEBS J 274*, 418–428.
52. Lehrer RI, Ganz T (2002). Cathelicidins: A family of endogenous antimicrobial peptides. *Curr Opin Hematol 9*, 18–22.
53. Xiao Y, Dai H, Bommineni YR, Soulages JL, Gong YX, Prakash O, Zhang G (2006). Structure-activity relationships of fowlicidin-1, a cathelicidin antimicrobial peptide in chicken. *FEBS J 273*, 2581–2593.
54. Zaiou M, Gallo RL (2002). Cathelicidins, essential gene-encoded mammalian antibiotics. *J Mol Med 80*, 549–561.
55. Zanetti M (2004). Cathelicidins, multifunctional peptides of the innate immunity. *J Leukoc Biol 75*, 39–48.
56. Lynn DJ, Higgs R, Gaines S, Tierney J, James T, Lloyd AT, Fares MA, Mulcahy G, O'Farrelly C (2004). Bioinformatic discovery and initial characterisation of nine novel antimicrobial peptide genes in the chicken. *Immunogenetics 56*, 170–177.
57. van Dijk A, Veldhuizen EJ A, van Asten AJ AM, Haagsman HP (2005). CMAP27, a novel chicken cathelicidin-like antimicrobial protein. *Vet Immunol Immunopathol 106*, 321–327.
58. Xiao Y, Cai Y, Bommineni YR, Fernando SC, Prakash O, Gilliland SE, Zhang G (2006). Identification and functional characterization of three chicken cathelicidins with potent antimicrobial activity. *J Biol Chem 281*, 2858–2867.
59. Lauterwein J, Brown LR, Wuthrich K (1980). High-resolution 1H-NMR studies of monomeric melittin in aqueous solution. *Biochim Biophys Acta 622*, 219–230.
60. Terwilliger TC, Weissman L, Eisenberg D (1982). The structure of melittin in the form I crystals and its implication for melittin's lytic and surface activities. *Biophys J 37*, 353–361.
61. Brown LR, Wuthrich K (1981). Melittin bound to dodecylphosphocholine micelles. H-NMR assignments and global conformational features. *Biochim Biophys Acta 647*, 95–111.
62. Brown LR, Braun W, Kumar A, Wuthrich K (1982). High resolution nuclear magnetic resonance studies of the conformation and orientation of melittin bound to a lipid-water interface. *Biophys J 37*, 319–328.
63. Inagaki F, Shimada I, Kawaguchi K, Hirano M, Terasawa I, Ikura T, Go N (1989). Structure of melittin bound to perdeuterated dodecylphosphocholine micelles as studied by two-dimensional NMR and distance geometry calculations. *Biochemistry 28*, 5985–5991.
64. Ikura T, Go N, Inagaki F (1991). Refined structure of melittin bound to perdeuterated dodeclylphoscholine micelles as studied by 2D-NMR and distance geometry calculation. *Proteins: Structure, Function and Bioinformatics 9*, 81–89.
65. Boman HG, Hultmark D (1987). Cell-free immunity in insects. *Annu Rev Microbiol 41*, 103–126.
66. Steiner H, Hultmark D, Engstrom A, Bennich H, Boman HG (1981). Sequence and specificity of two antibacterial proteins involved in insect immunity. *Nature 292*, 246–248.
67. Steiner H (1982). Secondary structure of the cecropins: Antibacterial peptides from the moth Hyalophora cecropia. *FEBS Lett 137*, 283–287.
68. Holak TA, Engstrom A, Kraulis PJ, Lindeberg G, Bennich H, Jones TA, Gronenborn AM, Clore GM (1988). The solution conformation of the antibacterial peptide cecropin A: A nuclear magnetic resonance and dynamical simulated annealing study. *Biochemistry 27*, 7620–7629.
69. Srisailam S, Arunkumar AI, Wang W, Yu C, Chen HM (2000). Conformational study of a custom antibacterial peptide cecropin B1: Implications of the lytic activity. *Biochim Biophys Acta 1479*, 275–285.
70. Oh D, Shin SY, Kang JH, Hahm KS, Kim KL, Kim Y (1999). NMR structural characterization of cecropin

A(1–8)-magainin 2(1–12) and cecropin A (1–8)-melittin (1–12) hybrid peptides. *J Pept Res 53*, 578–589.

71. Oh D, Shin SY, Lee S, Kang JH, Kim SD, Ryu PD, Hahm KS, Kim Y (2000). Role of the hinge region and the tryptophan residue in the synthetic antimicrobial peptides, cecropin A(1–8)-magainin 2(1–12) and its analogues, on their antibiotic activities and structures. *Biochemistry 39*, 11855–11864.
72. Sipos D, Chandrasekhar K, Arvidsson K, Engstrom A, Ehrenberg A (1991). Two-dimensional proton-NMR studies on a hybrid peptide between cecropin A and melittin. Resonance assignments and secondary structure. *Eur J Biochem 199*, 285–291.
73. Tauro S, Coutinho E, Srivastava S (2004). Conformation of the cecropin-melittin hybrid, cecropin A(1–8)-melittin (1–18) antibacterial Peptide. *Protein Pept Lett 11*, 115–124.
74. Park K, Oh D, Shin SY, Hahm KS, Kim Y (2002). Structural studies of porcine myeloid antibacterial peptide PMAP-23 and its analogues in DPC micelles by NMR spectroscopy. *Biochem Biophys Res Commun 290*, 204–212.
75. Zanetti M, Storici P, Tossi A, Scocchi M, Gennaro R (1994). Molecular cloning and chemical synthesis of a novel antibacterial peptide derived from pig myeloid cells. *J Biol Chem 269*, 7855–7858.
76. Corzo G, Escoubas P, Villegas E, Barnham KJ, He W, Norton RS, Nakajima T (2001). Characterization of unique amphipathic antimicrobial peptides from venom of the scorpion *Pandinus imperator*. *Biochem J 359*, 35–45.
77. Nomura K, Corzo G, Nakajima T, Iwashita T (2004). Orientation and pore-forming mechanism of a scorpion pore-forming peptide bound to magnetically oriented lipid bilayers. *Biophys J 87*, 2497–2507.
78. Anderson DH, Sawaya MR, Cascio D, Ernst W, Modlin R, Krensky A, Eisenberg D (2003). Granulysin crystal structure and a structure-derived lytic mechanism. *J Mol Biol 325*, 355–365.
79. Krensky AM (2000). Granulysin: A novel antimicrobial peptide of cytolytic T lymphocytes and natural killer cells. *Biochem Pharmacol 59*, 317–320.
80. Hunter HN, Demcoe AR, Jenssen H, Gutteberg TJ, Vogel HJ (2005). Human lactoferricin is partially folded in aqueous solution and is better stabilized in a membrane mimetic solvent. *Antimicrob Agents Chemother 49*, 3387–3395.
81. Vorland LH, Ulvatne H, Andersen J, Haukland H, Rekdal O, Svendsen JS, Gutteberg TJ (1998). Lactoferricin of bovine origin is more active than lactoferricins of human, murine and caprine origin. *Scand J Infect Dis 30*, 513–517.
82. Vorland LH, Ulvatne H, Andersen J, Haukland HH, Rekdal O, Svendsen JS, Gutteberg TJ (1999). Antibacterial effects of lactoferricin B. *Scand J Infect Dis 31*, 179–184.
83. Ganz T (2002). Immunology. Versatile defensins. *Science 298*, 977–979.
84. Lehrer RI, Ganz T (2002). Defensins of vertebrate animals. *Curr Opin Immunol 14*, 96–102.
85. Selsted ME, Harwig SS (1989). Determination of the disulfide array in the human defensin HNP-2. A covalently cyclized peptide. *J Biol Chem 264*, 4003–4007.
86. Tang YQ, Selsted ME (1993). Characterization of the disulfide motif in BNBD-12, an antimicrobial beta-defensin peptide from bovine neutrophils. *J Biol Chem 268*, 6649–6653.
87. Schibli DJ, Hunter HN, Aseyev V, Starner TD, Wiencek JM, McCray PB Jr, Tack BF, Vogel HJ (2002). The solution structures of the human beta-defensins lead to a better understanding of the potent bactericidal activity of HBD3 against *Staphylococcus aureus*. *J Biol Chem 277*, 8279–8289.
88. Fehlbaum P, Bulet P, Chernysh S, Briand JP, Roussel JP, Letellier L, Hetru C, Hoffmann JA (1996). Structure-activity analysis of thanatin, a 21-residue inducible insect defense peptide with sequence homology to frog skin antimicrobial peptides. *Proc Natl Acad Sci USA 93*, 1221–1225.
89. Mandard N, Sodano P, Labbe H, Bonmatin JM, Bulet P, Hetru C, Ptak M, Vovelle F (1998). Solution structure of thanatin, a potent bactericidal and fungicidal insect peptide, determined from proton two-dimensional nuclear magnetic resonance data. *Eur J Biochem 256*, 404–410.
90. Fahrner RL, Dieckmann T, Harwig SS, Lehrer RI, Eisenberg D, Feigon J (1996). Solution structure of protegrin-1, a broad-spectrum antimicrobial peptide from porcine leukocytes. *Chem Biol 3*, 543–550.
91. Harwig SS, Swiderek KM, Lee TD, Lehrer RI (1995). Determination of disulphide bridges in PG-2, an antimicrobial peptide from porcine leukocytes. *J Pept Sci 1*, 207–215.
92. Kokryakov VN, Harwig SS, Panyutich EA, Shevchenko AA, Aleshina GM, Shamova OV, Korneva HA, Lehrer RI (1993). Protegrins: Leukocyte antimicrobial peptides that combine features of corticostatic defensins and tachyplesins. *FEBS Lett 327*, 231–236.
93. Szyk A, Wu Z, Tucker K, Yang D, Lu W, Lubkowski J (2006). Crystal structures of human alpha-defensins HNP4, HD5, and HD6. *Protein Sci 15*, 2749–2760.
94. Hull SE, Karlsson R, Main P, Woolfson MM, Dodson EJ (1978). The crystal structure of a hydrated gramicidin S-urea complex. *Nature 275*, 206–207.
95. Gibbs AC, Kondejewski LH, Gronwald W, Nip AM, Hodges RS, Sykes BD, Wishart DS (1998). Unusual beta-sheet periodicity in small cyclic peptides. *Nat Struct Biol 5*, 284–288.
96. Xu Y, Sugar IP, Krishna NR (1995). A variable target intensity-restrained global optimization (VARTIGO) procedure for determining three-dimensional structures of polypeptides from NOESY data: Application to gramicidin-S. *J Biomol NMR 5*, 37–48.
97. Salgado J, Grage SL, Kondejewski LH, Hodges RS, McElhaney RN, Ulrich AS (2001). Membrane-bound

structure and alignment of the antimicrobial beta-sheet peptide gramicidin S derived from angular and distance constraints by solid state 19F-NMR. *J Biomol NMR 21*, 191–208.

98. Gran L (1973). On the effect of a polypeptide isolated from "Kalata-Kalata" (Oldenlandia affinis DC) on the oestrogen dominated uterus. *Acta Pharmacol Toxicol (Copenh) 33*, 400–408.
99. Gustafson KR, Sowder RC, Henderson LE, Parsons IC, Kashman Y, Cardellina JH, McMahon JB, Buckheit RW, Pannell LK, Boyd MR (1994). Circulins A and B. Novel human immunodeficiency virus (HIV)-inhibitory macrocyclic peptides from the tropical tree *Chassalia parvifolia*. *J Am Chem Soc 116*, 9337–9338.
100. Saether O, Craik DJ, Campbell ID, Sletten K, Juul J, Norman DG (1995). Elucidation of the primary and three-dimensional structure of the uterotonic polypeptide kalata B1. *Biochemistry 34*, 4147–4158.
101. Witherup KM, Bogusky MJ, Anderson PS, Ramjit H, Ransom RW, Wood T, Sardana M (1994). Cyclopsychotride A, a biologically active, 31-residue cyclic peptide isolated from Psychotria longipes. *J Nat Prod 57*, 1619–1625.
102. Tam JP, Lu YA, Yang JL, Chiu KW (1999). An unusual structural motif of antimicrobial peptides containing end-to-end macrocycle and cystine-knot disulfides. *Proc Natl Acad Sci USA 96*, 8913–8918.
103. Daly NL, Koltay A, Gustafson KR, Boyd MR, Casas-Finet JR, Craik DJ (1999). Solution structure by NMR of circulin A: A macrocyclic knotted peptide having anti-HIV activity. *J Mol Biol 285*, 333–345.
104. Derua R, Gustafson KR, Pannell LK (1996). Analysis of the disulfide linkage pattern in circulin A and B, HIV-inhibitory macrocyclic peptides. *Biochem Biophys Res Commun 228*, 632–638.
105. Tam JP, Lu, Y-A (1997). Synthesis of large cyclic cystine-knot peptide by orthogonal coupling strategy using unprotected peptide precursor. *Tetrahedron Lett 38*, 5599–5602.
106. Tam JP, Lu YA (1998). A biomimetic strategy in the synthesis and fragmentation of cyclic protein. *Protein Sci 7*, 1583–1592.
107. Lawyer C, Pai S, Watabe M, Borgia P, Mashimo T, Eagleton L, Watabe K (1996). Antimicrobial activity of a 13 amino acid tryptophan-rich peptide derived from a putative porcine precursor protein of a novel family of antibacterial peptides. *FEBS Lett 390*, 95–98.
108. Schibli DJ, Hwang PM, Vogel HJ (1999). Structure of the antimicrobial peptide tritrpticin bound to micelles: A distinct membrane-bound peptide fold. *Biochemistry 38*, 16749–16755.
109. Selsted ME, Novotny MJ, Morris WL, Tang YQ, Smith W, Cullor JS (1992). Indolicidin, a novel bactericidal tridecapeptide amide from neutrophils. *J Biol Chem 267*, 4292–4295.
110. Rozek A, Friedrich CL, Hancock RE (2000). Structure of the bovine antimicrobial peptide indolicidin bound to dodecylphosphocholine and sodium dodecyl sulfate micelles. *Biochemistry 39*, 15765–15774.
111. Subbalakshmi C, Bikshapathy E, Sitaram N, Nagaraj R (2000). Antibacterial and hemolytic activities of single tryptophan analogs of indolicidin. *Biochem Biophys Res Commun 274*, 714–716.
112. Falla TJ, Karunaratne DN, Hancock RE (1996). Mode of action of the antimicrobial peptide indolicidin. *J Biol Chem 271*, 19298–19303.
113. Casteels P, Tempst P (1994). Apidaecin-type peptide antibiotics function through a non-poreforming mechanism involving stereospecificity. *Biochem Biophys Res Commun 199*, 339–345.
114. Boman HG, Agerberth B, Boman A (1993). Mechanisms of action on *Escherichia coli* of cecropin P1 and PR-39, two antibacterial peptides from pig intestine. *Infect Immun 61*, 2978–2984.
115. del Castillo FJ, del Castillo I, Moreno F (2001). Construction and characterization of mutations at codon 751 of the *Escherichia coli* gyrB gene that confer resistance to the antimicrobial peptide microcin B17 and alter the activity of DNA gyrase. *J Bacteriol 183*, 2137–2140.
116. Patrzykat A, Friedrich CL, Zhang L, Mendoza V, Hancock RE (2002). Sublethal concentrations of pleurocidin-derived antimicrobial peptides inhibit macromolecular synthesis in *Escherichia coli*. *Antimicrob Agents Chemother 46*, 605–614.
117. Subbalakshmi C, Sitaram N (1998). Mechanism of antimicrobial action of indolicidin. *FEMS Microbiol Lett 160*, 91–96.
118. Friedrich CL, Rozek A, Patrzykat A, Hancock RE (2001). Structure and mechanism of action of an indolicidin peptide derivative with improved activity against gram-positive bacteria. *J Biol Chem 276*, 24015–24022.
119. Park CB, Kim HS, Kim SC (1998). Mechanism of action of the antimicrobial peptide buforin II: Buforin II kills microorganisms by penetrating the cell membrane and inhibiting cellular functions. *Biochem Biophys Res Commun 244*, 253–257.
120. Kragol G, Lovas S, Varadi G, Condie BA, Hoffmann R, Otvos L Jr (2001). The antibacterial peptide pyrrhocoricin inhibits the ATPase actions of DnaK and prevents chaperone-assisted protein folding. *Biochemistry 40*, 3016–3026.
121. Otvos L Jr, O I, Rogers ME, Consolvo PJ, Condie BA, Lovas S, Bulet P, Blaszczyk-Thurin M (2000). Interaction between heat shock proteins and antimicrobial peptides. *Biochemistry 39*, 14150–14159.
122. Oppenheim J, Biragyn A, Kwak LW, Yang D (2003). Roles of antimicrobial peptides such as defensins in innate and adaptive immunity. *Ann Rheum Dis 62*(Suppl 2), ii17–ii21.
123. Zhou NE, Kay CM, Hodges RS (1994). α-helical propensities of amino acids in the hydrophobic face of an amphipathic α-helix. *Protein and Peptide Letters 1*, 114–119.

124. Kovacs JM, Mant CT, Hodges RS (2006). Determination of intrinsic hydrophilicity/hydrophobicity of amino acid side chains in peptides in the absence of nearest-neighbor or conformational effects. *Biopolymers 84*, 283–297.

125. Eisenberg D, Weiss RM, Terwilliger TC (1982). The helical hydrophobic moment: A measure of the amphiphilicity of a helix. *Nature 299*, 371–374.

126. Carver T, Bleasby A (2003). The design of Jemboss: A graphical user interface to EMBOSS. *Bioinformatics 19*, 1837–1843.

127. Lee DL, Hodges RS (2003). Structure-activity relationships of de novo designed cyclic antimicrobial peptides based on gramicidin S. *Biopolymers 71*, 28–48.

128. Mant CT, Chen Y, Hodges RS (2003). Temperature profiling of polypeptides in reversed-phase liquid chromatography. I. Monitoring of dimerization and unfolding of amphipathic alpha-helical peptides. *J Chromatogr A 1009*, 29–43.

129. Mant CT, Tripet B, Hodges RS (2003). Temperature profiling of polypeptides in reversed-phase liquid chromatography. II. Monitoring of folding and stability of two-stranded alpha-helical coiled-coils. *J Chromatogr A 1009*, 45–59.

130. Chen Y, Mant CT, Farmer SW, Hancock RE, Vasil ML, Hodges RS (2005). Rational design of alpha-helical antimicrobial peptides with enhanced activities and specificity/therapeutic index. *J Biol Chem 280*, 12316–12329.

131. Chen Y, Vasil AI, Rehaume L, Mant CT, Burns JL, Vasil ML, Hancock RE, Hodges RS (2006). Comparison of biophysical and biologic properties of alpha-helical enantiomeric antimicrobial peptides. *Chem Biol Drug Des 67*, 162–173.

132. Chen Y, Guarnieri MT, Vasil AI, Vasil ML, Mant CT, Hodges RS (2007). Role of peptide hydrophobicity in the mechanism of action of alpha-helical antimicrobial peptides. *Antimicrob Agents Chemother 51*, 1398–1406.

133. Jiang Z, Vasil AI, Hale JD, Hancock RE, Vasil ML, Hodges RS (2008). Effects of net charge and the number of positively charged residues on the biological activity of amphipathic α-helical cationic antimicrobial peptides. *Biopolymers (Peptide Science) 90*, 369–383.

134. Faber C, Stallmann HP, Lyaruu DM, Joosten U, von Eiff C, van Nieuw Amerongen A, Wuisman PI (2005). Comparable efficacies of the antimicrobial peptide human lactoferrin 1–11 and gentamicin in a chronic methicillin-resistant *Staphylococcus aureus* osteomyelitis model. *Antimicrob Agents Chemother 49*, 2438–2444.

135. Giacometti A, Cirioni O, Kamysz W, Silvestri C, Licci A, Riva A, Lukasiak J, Scalise G (2005). In vitro activity of amphibian peptides alone and in combination with antimicrobial agents against multidrug-resistant pathogens isolated from surgical wound infection. *Peptides 26*, 2111–2116.

136. Ali MF, Soto A, Knoop FC, Conlon JM (2001). Antimicrobial peptides isolated from skin secretions of the diploid frog, *Xenopus tropicalis* (Pipidae). *Biochim Biophys Acta 1550*, 81–89.

137. Lee MT, Chen FY, Huang HW (2004). Energetics of pore formation induced by membrane active peptides. *Biochemistry 43*, 3590–3599.

138. Zhu WL, Song YM, Park Y, Park KH, Yang ST, Kim JI, Park IS, Hahm KS, Shin SY (2007). Substitution of the leucine zipper sequence in melittin with peptoid residues affects self-association, cell selectivity, and mode of action. *Biochim Biophys Acta 1768*, 1506–1517.

139. Iijima N, Tanimoto N, Emoto Y, Morita Y, Uematsu K, Murakami T, Nakai T (2003). Purification and characterization of three isoforms of chrysophsin, a novel antimicrobial peptide in the gills of the red sea bream, *Chrysophrys major*. *Eur J Biochem 270*, 675–686.

140. Barber M (1961). Methicillin-resistant staphylococci. *J Clin Pathol 14*, 385–393.

141. Maple PA, Hamilton-Miller JM, Brumfitt W (1989). World-wide antibiotic resistance in methicillin-resistant *Staphylococcus aureus*. *Lancet 1*, 537–540.

142. Dixson S, Brumfitt W, Hamilton-Miller JM (1985). In vitro activity of six antibiotics against multiresistant staphylococci and other gram-positive cocci. *Eur J Clin Microbiol 4*, 19–23.

143. Lustig F, Hoebeke J, Ostergren-Lunden G, Velge-Roussel F, Bondjers G, Olsson U, Ruetschi U, Fager G (1996). Alternative splicing determines the binding of platelet-derived growth factor (PDGF-AA) to glycosaminoglycans. *Biochemistry 35*, 12077–12085.

144. McCullough MJ, Ross BC, Reade PC (1996). *Candida albicans*: A review of its history, taxonomy, epidemiology, virulence attributes, and methods of strain differentiation. *Int J Oral Maxillofac Surg 25*, 136–144.

145. Meunier-Carpentier F, Kiehn TE, Armstrong D (1981). Fungemia in the immunocompromised host. Changing patterns, antigenemia, high mortality. *Am J Med 71*, 363–370.

146. Rex JH, Rinaldi MG, Pfaller MA (1995). Resistance of *Candida* species to fluconazole. *Antimicrob Agents Chemother 39*, 1–8.

147. Cole AM, Darouiche RO, Legarda D, Connell N, Diamond G (2000). Characterization of a fish antimicrobial peptide: Gene expression, subcellular localization, and spectrum of activity. *Antimicrob Agents Chemother 44*, 2039–2045.

148. Latge JP (1999). *Aspergillus fumigatus* and aspergillosis. *Clin Microbiol Rev 12*, 310–350.

149. Hedayati MT, Pasqualotto AC, Warn PA, Bowyer P, Denning DW (2007). *Aspergillus flavus*: Human pathogen, allergen and mycotoxin producer. *Microbiology 153*, 1677–1692.

150. Buchanan KL, Murphy JW (1998). What makes *Cryptococcus neoformans* a pathogen? *Emerg Infect Dis 4*, 71–83.

151. Scocchi M, Zelezetsky I, Benincasa M, Gennaro R, Mazzoli A, Tossi A (2005). Structural aspects and

biological properties of the cathelicidin PMAP-36. *FEBS J 272*, 4398–4406.

152. Benincasa M, Scocchi M, Pacor S, Tossi A, Nobili D, Basaglia G, Busetti M, Gennaro R (2006). Fungicidal activity of five cathelicidin peptides against clinically isolated yeasts. *J Antimicrob Chemother 58*, 950–959.
153. Belaid A, Aouni M, Khelifa R, Trabelsi A, Jemmali M, Hani K (2002). In vitro antiviral activity of dermaseptins against herpes simplex virus type 1. *J Med Virol 66*, 229–234.
154. Aboudy Y, Mendelson E, Shalit I, Bessalle R, Fridkin M (1994). Activity of two synthetic amphiphilic peptides and magainin-2 against herpes simplex virus types 1 and 2. *Int J Pept Protein Res 43*, 573–582.
155. Robinson WE Jr, McDougall B, Tran D, Selsted ME (1998). Anti-HIV-1 activity of indolicidin, an antimicrobial peptide from neutrophils. *J Leukoc Biol 63*, 94–100.
156. Jenssen H, Andersen JH, Mantzilas D, Gutteberg TJ (2004). A wide range of medium-sized, highly cationic, alpha-helical peptides show antiviral activity against herpes simplex virus. *Antiviral Res 64*, 119–126.
157. Jenssen H, Andersen JH, Uhlin-Hansen L, Gutteberg TJ, Rekdal O (2004). Anti-HSV activity of lactoferricin analogues is only partly related to their affinity for heparan sulfate. *Antiviral Res 61*, 101–109.
158. Zasloff M (1987). Magainins, a class of antimicrobial peptides from *Xenopus* skin: Isolation, characterization of two active forms, and partial cDNA sequence of a precursor. *Proc Natl Acad Sci USA 84*, 5449–5453.
159. Roch P, Beschin A, Bernard E (2004). Antiprotozoan and antiviral activities of non-cytotoxic truncated and variant analogues of mussel defensin. *Evid Based Complement Alternat Med 1*, 167–174.
160. Herwaldt BL (1999). Leishmaniasis. *Lancet 354*, 1191–1199.
161. Utsugi T, Schroit AJ, Connor J, Bucana CD, Fidler IJ (1991). Elevated expression of phosphatidylserine in the outer membrane leaflet of human tumor cells and recognition by activated human blood monocytes. *Cancer Res 51*, 3062–3066.
162. Kamysz W, Okroj M, Lukasiak J (2003). Novel properties of antimicrobial peptides. *Acta Biochim Pol 50*, 461–469.
163. Rozek T, Wegener KL, Bowie JH, Olver IN, Carver JA, Wallace JC, Tyler MJ (2000). The antibiotic and anticancer active aurein peptides from the Australian Bell Frogs *Litoria aurea* and *Litoria raniformis* the solution structure of aurein 1.2. *Eur J Biochem 267*, 5330–5341.
164. Murphy CJ, Foster BA, Mannis MJ, Selsted ME, Reid TW (1993). Defensins are mitogenic for epithelial cells and fibroblasts. *J Cell Physiol 155*, 408–413.
165. Charp PA, Rice WG, Raynor RL, Reimund E, Kinkade JM Jr, Ganz T, Selsted ME, Lehrer RI, Kuo JF (1988). Inhibition of protein kinase C by defensins, antibiotic peptides from human neutrophils. *Biochem Pharmacol 37*, 951–956.
166. Territo MC, Ganz T, Selsted ME, Lehrer R (1989). Monocyte-chemotactic activity of defensins from human neutrophils. *J Clin Invest 84*, 2017–2020.
167. Sang Y, Teresa Ortega M, Rune K, Xiau W, Zhang G, Soulages JL, Lushington GH, Fang J, Williams TD, Blecha F, Melgarejo T (2007). Canine cathelicidin (K9CATH): Gene cloning, expression, and biochemical activity of a novel pro-myeloid antimicrobial peptide. *Dev Comp Immunol 31*, 1278–1296.
168. Lai R, Zheng YT, Shen JH, Liu GJ, Liu H, Lee WH, Tang SZ, Zhang Y (2002). Antimicrobial peptides from skin secretions of Chinese red belly toad *Bombina maxima*. *Peptides 23*, 427–435.
169. Skerlavaj B, Benincasa M, Risso A, Zanetti M, Gennaro R (1999). SMAP-29: A potent antibacterial and antifungal peptide from sheep leukocytes. *FEBS Lett 463*, 58–62.
170. Crowell KM, Lutz F (1989). *Pseudomonas aeruginosa* cytotoxin: The influence of sphingomyelin on binding and cation permeability increase in mammalian erythrocytes. *Toxicon 27*, 531–540.
171. Chen HC, Brown JH, Morell JL, Huang CM (1988). Synthetic magainin analogues with improved antimicrobial activity. *FEBS Lett 236*, 462–466.
172. Mor A, Nicolas P (1994). The NH2-terminal alpha-helical domain 1–18 of dermaseptin is responsible for antimicrobial activity. *J Biol Chem 269*, 1934–1939.
173. Blazyk J, Wiegand R, Klein J, Hammer J, Epand RM, Epand RF, Maloy WL, Kari UP (2001). A novel linear amphipathic beta-sheet cationic antimicrobial peptide with enhanced selectivity for bacterial lipids. *J Biol Chem 276*, 27899–27906.
174. Gibson BW, Tang DZ, Mandrell R, Kelly M, Spindel ER (1991). Bombinin-like peptides with antimicrobial activity from skin secretions of the Asian toad, *Bombina orientalis*. *J Biol Chem 266*, 23103–23111.
175. Conlon JM, Sonnevend A, Davidson C, Smith DD, Nielsen PF (2004). The ascaphins: A family of antimicrobial peptides from the skin secretions of the most primitive extant frog, *Ascaphus truei*. *Biochem Biophys Res Commun 320*, 170–175.
176. Shin SY, Hahm KS (2004). A short alpha-helical antimicrobial peptide with antibacterial selectivity. *Biotechnol Lett 26*, 735–739.
177. Minn I, Kim HS, Kim SC (1998). Antimicrobial peptides derived from pepsinogens in the stomach of the bullfrog, *Rana catesbeiana*. *Biochim Biophys Acta 1407*, 31–39.
178. Jones S, Howl J (2004). Charge delocalisation and the design of novel mastoparan analogues: Enhanced cytotoxicity and secretory efficacy of [Lys5, Lys8, Aib10]MP. *Regul Pept 121*, 121–128.
179. Pardo J, Perez-Galan P, Gamen S, Marzo I, Monleon I, Kaspar AA, Susin SA, Kroemer G, Krensky AM, Naval J, Anel A (2001). A role of the mitochondrial apoptosis-inducing factor in granulysin-induced apoptosis. *J Immunol 167*, 1222–1229.

180. Imura Y, Nishida M, Matsuzaki K (2007). Action mechanism of PEGylated magainin 2 analogue peptide. *Biochim Biophys Acta 1768*, 2578–2585.
181. Pal T, Sonnevend A, Galadari S, Conlon JM (2005). Design of potent, non-toxic antimicrobial agents based upon the structure of the frog skin peptide, pseudin-2. *Regul Pept 129*, 85–91.
182. Krishan A (1975). Rapid flow cytofluorometric analysis of mammalian cell cycle by propidium iodide staining. *J Cell Biol 66*, 188–193.
183. Javadpour MM, Juban MM, Lo WC, Bishop SM, Alberty JB, Cowell SM, Becker CL, McLaughlin ML (1996). De novo antimicrobial peptides with low mammalian cell toxicity. *J Med Chem 39*, 3107–3113.
184. Skerlavaj B, Gennaro R, Bagella L, Merluzzi L, Risso A, Zanetti M (1996). Biological characterization of two novel cathelicidin-derived peptides and identification of structural requirements for their antimicrobial and cell lytic activities. *J Biol Chem 271*, 28375–28381.
185. Kubo H, Loegering DA, Tohda Y, Bankers-Fulbright J, Weiler CR, Nakajima H, Thomas LL, Adolphson CR, Gleich GJ (1999). Discordant and anomalous results among cytotoxicity assays: The confounding properties of eosinophil granule major basic protein on cell viability assays. *J Immunol Methods 227*, 1–15.
186. Glavin GB, Szabo S, Johnson BR, Xing PL, Morales RE, Plebani M, Nagy L (1996). Isolated rat gastric mucosal cells: Optimal conditions for cell harvesting, measures of viability and direct cytoprotection. *J Pharmacol Exp Ther 276*, 1174–1179.
187. King MA, Wu QX, Donovan GR, Baldo BA (1998). Flow cytometric analysis of cell killing by the jumper ant venom peptide pilosulin 1. *Cytometry 32*, 268–273.
188. Ohsaki Y, Gazdar AF, Chen HC, Johnson BE (1992). Antitumor activity of magainin analogues against human lung cancer cell lines. *Cancer Res 52*, 3534–3538.
189. Conlon JM, Sonnevend A, Patel M, Camasamudram V, Nowotny N, Zilahi E, Iwamuro S, Nielsen PF, Pal T (2003). A melittin-related peptide from the skin of the Japanese frog, *Rana tagoi*, with antimicrobial and cytolytic properties. *Biochem Biophys Res Commun 306*, 496–500.
190. Oren Z, Shai Y (1997). Selective lysis of bacteria but not mammalian cells by diastereomers of melittin: Structure-function study. *Biochemistry 36*, 1826–1835.
191. Ernst RK, Yi EC, Guo L, Lim KB, Burns JL, Hackett M, Miller SI (1999). Specific lipopolysaccharide found in cystic fibrosis airway *Pseudomonas aeruginosa*. *Science 286*, 1561–1565.
192. Groisman EA, Parra-Lopez C, Salcedo M, Lipps CJ, Heffron F (1992). Resistance to host antimicrobial peptides is necessary for *Salmonella* virulence. *Proc Natl Acad Sci USA 89*, 11939–11943.
193. Gunn JS, Lim KB, Krueger J, Kim K, Guo L, Hackett M, Miller SI (1998). PmrA-PmrB-regulated genes necessary for 4-aminoarabinose lipid A modification and polymyxin resistance. *Mol Microbiol 27*, 1171–1182.
194. Macfarlane EL, Kwasnicka A, Hancock RE (2000). Role of *Pseudomonas aeruginosa* PhoP-phoQ in resistance to antimicrobial cationic peptides and aminoglycosides. *Microbiology 146*(Pt 10), 2543–2554.
195. Fields PI, Groisman EA, Heffron F (1989). A *Salmonella* locus that controls resistance to microbicidal proteins from phagocytic cells. *Science 243*, 1059–1062.
196. McCoy AJ, Liu H, Falla TJ, Gunn JS (2001). Identification of *Proteus mirabilis* mutants with increased sensitivity to antimicrobial peptides. *Antimicrob Agents Chemother 45*, 2030–2037.
197. Guo L, Lim KB, Poduje CM, Daniel M, Gunn JS, Hackett M, Miller SI (1998). Lipid A acylation and bacterial resistance against vertebrate antimicrobial peptides. *Cell 95*, 189–198.
198. Banemann A, Deppisch H, Gross R (1998). The lipopolysaccharide of *Bordetella bronchiseptica* acts as a protective shield against antimicrobial peptides. *Infect Immun 66*, 5607–5612.
199. Lysenko ES, Gould J, Bals R, Wilson JM, Weiser JN (2000). Bacterial phosphorylcholine decreases susceptibility to the antimicrobial peptide LL-37/hCAP18 expressed in the upper respiratory tract. *Infect Immun 68*, 1664–1671.
200. Peschel A, Otto M, Jack RW, Kalbacher H, Jung G, Gotz F (1999). Inactivation of the dlt operon in *Staphylococcus aureus* confers sensitivity to defensins, protegrins, and other antimicrobial peptides. *J Biol Chem 274*, 8405–8410.
201. Kristian SA, Durr M, Van Strijp JA, Neumeister B, Peschel A (2003). MprF-mediated lysinylation of phospholipids in *Staphylococcus aureus* leads to protection against oxygen-independent neutrophil killing. *Infect Immun 71*, 546–549.
202. Peschel A, Jack RW, Otto M, Collins LV, Staubitz P, Nicholson G, Kalbacher H, Nieuwenhuizen WF, Jung G, Tarkowski A, van Kessel KP, van Strijp JA (2001). *Staphylococcus aureus* resistance to human defensins and evasion of neutrophil killing via the novel virulence factor MprF is based on modification of membrane lipids with l-lysine. *J Exp Med 193*, 1067–1076.
203. Campos MA, Vargas MA, Regueiro V, Llompart CM, Alberti S, Bengoechea JA (2004). Capsule polysaccharide mediates bacterial resistance to antimicrobial peptides. *Infect Immun 72*, 7107–7114.
204. Nikaido H (1996). Multidrug efflux pumps of gram-negative bacteria. *J Bacteriol 178*, 5853–5859.
205. Bengoechea JA, Skurnik M (2000). Temperature-regulated efflux pump/potassium antiporter system mediates resistance to cationic antimicrobial peptides in *Yersinia*. *Mol Microbiol 37*, 67–80.
206. Shafer WM, Qu X, Waring AJ, Lehrer RI (1998). Modulation of *Neisseria gonorrhoeae* susceptibility to vertebrate antibacterial peptides due to a member of the resistance/nodulation/division efflux pump family. *Proc Natl Acad Sci USA 95*, 1829–1833.

207. Belas R, Manos J, Suvanasuthi R (2004). *Proteus mirabilis* ZapA metalloprotease degrades a broad spectrum of substrates, including antimicrobial peptides. *Infect Immun 72*, 5159–5167.
208. Sieprawska-Lupa M, Mydel P, Krawczyk K, Wojcik K, Puklo M, Lupa B, Suder P, Silberring J, Reed M, Pohl J, Shafer W, McAleese F, Foster T, Travis J, Potempa J (2004). Degradation of human antimicrobial peptide LL-37 by *Staphylococcus aureus*-derived proteinases. *Antimicrob Agents Chemother 48*, 4673–4679.
209. Guina T, Yi EC, Wang H, Hackett M, Miller SI (2000). A PhoP-regulated outer membrane protease of *Salmonella enterica* serovar typhimurium promotes resistance to alpha-helical antimicrobial peptides. *J Bacteriol 182*, 4077–4086.
210. Giuliani A, Pirri, G, Nicoletto, S. (2007). Antimicrobial peptides: An overview of a promising class of therapeutics. *Central Eur J Biol 2*, 1–33.
211. Yeaman MR, Yount NY (2003). Mechanisms of antimicrobial peptide action and resistance. *Pharmacol Rev 55*, 27–55.
212. Matsuzaki K (1999). Why and how are peptide-lipid interactions utilized for self-defense? Magainins and tachyplesins as archetypes. *Biochim Biophys Acta 1462*, 1–10.
213. Huang HW (2000). Action of antimicrobial peptides: Two-state model. *Biochemistry 39*, 8347–8352.
214. Dathe M, Wieprecht T (1999). Structural features of helical antimicrobial peptides: Their potential to modulate activity on model membranes and biological cells. *Biochim Biophys Acta 1462*, 71–87.
215. Hancock RE, Patrzykat A (2002). Clinical development of cationic antimicrobial peptides: From natural to novel antibiotics. *Curr Drug Targets Infect Disord 2*, 79–83.
216. Hancock RE, Rozek A (2002). Role of membranes in the activities of antimicrobial cationic peptides. *FEMS Microbiol Lett 206*, 143–149.
217. Tossi A, Sandri L, Giangaspero A (2000). Amphipathic, alpha-helical antimicrobial peptides. *Biopolymers 55*, 4–30.
218. Brogden KA (2005). Antimicrobial peptides: Pore formers or metabolic inhibitors in bacteria? *Nat Rev Microbiol 3*, 238–250.
219. Vogt TC, Bechinger B (1999). The interactions of histidine-containing amphipathic helical peptide antibiotics with lipid bilayers. The effects of charges and pH. *J Biol Chem 274*, 29115–29121.
220. Wieprecht T, Beyermann M, Seelig J (1999). Binding of antibacterial magainin peptides to electrically neutral membranes: Thermodynamics and structure. *Biochemistry 38*, 10377–10387.
221. Bechinger B, Lohner K (2006). Detergent-like actions of linear amphipathic cationic antimicrobial peptides. *Biochim Biophys Acta 1758*, 1529–1539.
222. Lehrer RI, Barton A, Daher KA, Harwig SS, Ganz T, Selsted ME (1989). Interaction of human defensins with *Escherichia coli*. Mechanism of bactericidal activity. *J Clin Invest 84*, 553–561.
223. Westerhoff HV, Juretic D, Hendler RW, Zasloff M (1989). Magainins and the disruption of membrane-linked free-energy transduction. *Proc Natl Acad Sci USA 86*, 6597–6601.
224. Kustanovich I, Shalev DE, Mikhlin M, Gaidukov L, Mor A (2002). Structural requirements for potent versus selective cytotoxicity for antimicrobial dermaseptin S4 derivatives. *J Biol Chem 277*, 16941–16951.
225. Oren Z, Shai Y (2000). Cyclization of a cytolytic amphipathic alpha-helical peptide and its diastereomer: Effect on structure, interaction with model membranes, and biological function. *Biochemistry 39*, 6103–6114.
226. Stella L, Mazzuca C, Venanzi M, Palleschi A, Didone M, Formaggio F, Toniolo C, Pispisa B (2004). Aggregation and water-membrane partition as major determinants of the activity of the antibiotic peptide trichogin GA IV. *Biophys J 86*, 936–945.
227. Baumann G, Mueller P (1974). A molecular model of membrane excitability. *J Supramol Struct 2*, 538–557.
228. Ehrenstein G, Lecar H (1977). Electrically gated ionic channels in lipid bilayers. *Q Rev Biophys 10*, 1–34.
229. Shai Y (2002). Mode of action of membrane active antimicrobial peptides. *Biopolymers 66*, 236–248.
230. Zhang L, Rozek A, Hancock RE (2001). Interaction of cationic antimicrobial peptides with model membranes. *J Biol Chem 276*, 35714–35722.
231. Yang L, Harroun TA, Weiss TM, Ding L, Huang HW (2001). Barrel-stave model or toroidal model? A case study on melittin pores. *Biophys J 81*, 1475–1485.
232. Henzler Wildman KA, Lee DK, Ramamoorthy A (2003). Mechanism of lipid bilayer disruption by the human antimicrobial peptide, LL-37. *Biochemistry 42*, 6545–6558.
233. Pouny Y, Rapaport D, Mor A, Nicolas P, Shai Y (1992). Interaction of antimicrobial dermaseptin and its fluorescently labeled analogues with phospholipid membranes. *Biochemistry 31*, 12416–12423.
234. Reddy KV, Yedery RD, Aranha C (2004). Antimicrobial peptides: Premises and promises. *Int J Antimicrob Agents 24*, 536–547.
235. Terwilliger TC, Eisenberg D (1982). The structure of melittin. I. Structure determination and partial refinement. *J Biol Chem 257*, 6010–6015.
236. Yamaguchi S, Huster D, Waring A, Lehrer RI, Kearney W, Tack BF, Hong M (2001). Orientation and dynamics of an antimicrobial peptide in the lipid bilayer by solid-state NMR spectroscopy. *Biophys J 81*, 2203–2214.
237. Wu M, Maier E, Benz R, Hancock RE (1999). Mechanism of interaction of different classes of cationic antimicrobial peptides with planar bilayers and with the cytoplasmic membrane of *Escherichia coli*. *Biochemistry 38*, 7235–7242.
238. Powers JP, Martin MM, Goosney DL, Hancock RE (2006). The antimicrobial peptide polyphemusin localizes

to the cytoplasm of *Escherichia coli* following treatment. *Antimicrob Agents Chemother 50*, 1522–1524.

239. Oh JT, Van Dyk TK, Cajal Y, Dhurjati PS, Sasser M, Jain MK (1998). Osmotic stress in viable *Escherichia coli* as the basis for the antibiotic response by polymyxin B. *Biochem Biophys Res Commun 246*, 619–623.
240. Fattal E, Nir S, Parente RA, Szoka FC Jr (1994). Pore-forming peptides induce rapid phospholipid flip-flop in membranes. *Biochemistry 33*, 6721–6731.
241. Sahl HG, Bierbaum G (1998). Lantibiotics: Biosynthesis and biological activities of uniquely modified peptides from gram-positive bacteria. *Annu Rev Microbiol 52*, 41–79.
242. Park CB, Yi KS, Matsuzaki K, Kim MS, Kim SC (2000). Structure-activity analysis of buforin II, a histone H2A-derived antimicrobial peptide: The proline hinge is responsible for the cell-penetrating ability of buforin II. *Proc Natl Acad Sci USA 97*, 8245–8250.
243. Henk WG, Todd WJ, Enright FM, Mitchell PS (1995). The morphological effects of two antimicrobial peptides, hecate-1 and melittin, on *Escherichia coli*. *Scanning Microsc 9*, 501–507.
244. Kalfa VC, Jia HP, Kunkle RA, McCray PB Jr, Tack BF, Brogden KA (2001). Congeners of SMAP29 kill ovine pathogens and induce ultrastructural damage in bacterial cells. *Antimicrob Agents Chemother 45*, 3256–3261.
245. Wimley WC, Selsted ME, White SH (1994). Interactions between human defensins and lipid bilayers: Evidence for formation of multimeric pores. *Protein Sci 3*, 1362–1373.
246. Prenner EJ, Kiricsi M, Jelokhani-Niaraki M, Lewis RN, Hodges RS, McElhaney RN (2005). Structure-activity relationships of diastereomeric lysine ring size analogs of the antimicrobial peptide gramicidin S: Mechanism of action and discrimination between bacterial and animal cell membranes. *J Biol Chem 280*, 2002–2011.
247. Seto GW, Marwaha S, Kobewka DM, Lewis RN, Separovic F, McElhaney RN (2007). Interactions of the Australian tree frog antimicrobial peptides aurein 1.2, citropin 1.1 and maculatin 1.1 with lipid model membranes: Differential scanning calorimetric and Fourier transform infrared spectroscopic studies. *Biochim Biophys Acta 1768*, 2787–2800.
248. Prenner EJ, Lewis RN, Jelokhani-Niaraki M, Hodges RS, McElhaney RN (2001). Cholesterol attenuates the interaction of the antimicrobial peptide gramicidin S with phospholipid bilayer membranes. *Biochim Biophys Acta 1510*, 83–92.
249. Jackson M, Mantsch HH, Spencer JH (1992). Conformation of magainin-2 and related peptides in aqueous solution and membrane environments probed by Fourier transform infrared spectroscopy. *Biochemistry 31*, 7289–7293.
250. Tang J, Signarvic RS, Degrado WF, Gai F (2007). Role of helix nucleation in the kinetics of binding of mastoparan x to phospholipid bilayers. *Biochemistry 46*, 13856–13863.
251. Hristova K, Selsted ME, White SH (1997). Critical role of lipid composition in membrane permeabilization by rabbit neutrophil defensins. *J Biol Chem 272*, 24224–24233.
252. Kang JH, Shin SY, Jang SY, Lee MK, Hahm KS (1998). Release of aqueous contents from phospholipid vesicles induced by cecropin A (1–8)-magainin 2 (1–12) hybrid and its analogues. *J Pept Res 52*, 45–50.
253. Ladokhin AS, Selsted ME, White SH (1997). Sizing membrane pores in lipid vesicles by leakage of co-encapsulated markers: Pore formation by melittin. *Biophys J 72*, 1762–1766.
254. Matsuzaki K, Yoneyama S, Miyajima K (1997). Pore formation and translocation of melittin. *Biophys J 73*, 831–838.
255. Zhao H, Mattila JP, Holopainen JM, Kinnunen PK (2001). Comparison of the membrane association of two antimicrobial peptides, magainin 2 and indolicidin. *Biophys J 81*, 2979–2991.
256. Christensen B, Fink J, Merrifield RB, Mauzerall D (1988). Channel-forming properties of cecropins and related model compounds incorporated into planar lipid membranes. *Proc Natl Acad Sci USA 85*, 5072–5076.
257. Kagan BL, Selsted ME, Ganz T, Lehrer RI (1990). Antimicrobial defensin peptides form voltage-dependent ion-permeable channels in planar lipid bilayer membranes. *Proc Natl Acad Sci USA 87*, 210–214.
258. Lockey TD, Ourth DD (1996). Formation of pores in *Escherichia coli* cell membranes by a cecropin isolated from hemolymph of *Heliothis virescens* larvae. *Eur J Biochem 236*, 263–271.
259. Wu Y, Huang HW, Olah GA (1990). Method of oriented circular dichroism. *Biophys J 57*, 797–806.
260. Lam YH, Wassall SR, Morton CJ, Smith R, Separovic F (2001). Solid-state NMR structure determination of melittin in a lipid environment. *Biophys J 81*, 2752–2761.
261. Ramamoorthy A, Thennarasu S, Lee DK, Tan A, Maloy L (2006). Solid-state NMR investigation of the membrane-disrupting mechanism of antimicrobial peptides MSI-78 and MSI-594 derived from magainin 2 and melittin. *Biophys J 91*, 206–216.
262. He K, Ludtke SJ, Huang HW, Worcester DL (1995). Antimicrobial peptide pores in membranes detected by neutron in-plane scattering. *Biochemistry 34*, 15614–15618.
263. Ludtke SJ, He K, Heller WT, Harroun TA, Yang L, Huang HW (1996). Membrane pores induced by magainin. *Biochemistry 35*, 13723–13728.
264. Yang L, Harroun TA, Heller WT, Weiss TM, Huang HW (1998). Neutron off-plane scattering of aligned membranes. I. Method of measurement. *Biophys J 75*, 641–645.
265. Yang L, Weiss TM, Harroun TA, Heller WT, Huang HW (1999). Supramolecular structures of peptide assemblies

in membranes by neutron off-plane scattering: Method of analysis. *Biophys J 77*, 2648–2656.

266. Chen FY, Lee MT, Huang HW (2003). Evidence for membrane thinning effect as the mechanism for peptide-induced pore formation. *Biophys J 84*, 3751–3758.
267. Spaar A, Munster C, Salditt T (2004). Conformation of peptides in lipid membranes studied by x-ray grazing incidence scattering. *Biophys J 87*, 396–407.
268. Abraham T, Lewis RN, Hodges RS, McElhaney RN (2005). Isothermal titration calorimetry studies of the binding of a rationally designed analogue of the antimicrobial peptide gramicidin s to phospholipid bilayer membranes. *Biochemistry 44*, 2103–2112.
269. Abraham T, Marwaha S, Kobewka DM, Lewis RN AH, Prenner EJ S, Hodges R, McElhaney RN (2007). The relationship between the binding to and permeabilization of phospholipid bilayer membranes by GS14dK4, a designed analog of the antimicrobial peptide gramicidin S. *Biochim Biophys Acta (BBA) Biomembr 1768*, 2089–2098.
270. Prenner EJ, Lewis RN, Kondejewski LH, Hodges RS, McElhaney RN (1999). Differential scanning calorimetric study of the effect of the antimicrobial peptide gramicidin S on the thermotropic phase behavior of phosphatidylcholine, phosphatidylethanolamine and phosphatidylglycerol lipid bilayer membranes. *Biochim Biophys Acta 1417*, 211–223.
271. Boman HG, Wade D, Boman IA, Wahlin B, Merrifield RB (1989). Antibacterial and antimalarial properties of peptides that are cecropin-melittin hybrids. *FEBS Lett 259*, 103–106.
272. Andreu D, Ubach J, Boman A, Wahlin B, Wade D, Merrifield RB, Boman HG (1992). Shortened cecropin A-melittin hybrids. Significant size reduction retains potent antibiotic activity. *FEBS Lett 296*, 190–194.
273. Friedrich C, Scott MG, Karunaratne N, Yan H, Hancock RE (1999). Salt-resistant alpha-helical cationic antimicrobial peptides. *Antimicrob Agents Chemother 43*, 1542–1548.
274. Lee DG, Park JH, Shin SY, Lee SG, Lee MK, Kim KL, Hahm KS (1997). Design of novel analogue peptides with potent fungicidal but low hemolytic activity based on the cecropin A-melittin hybrid structure. *Biochem Mol Biol Int 43*, 489–498.
275. Piers KL, Brown MH, Hancock RE (1994). Improvement of outer membrane-permeabilizing and lipopolysaccharide-binding activities of an antimicrobial cationic peptide by C-terminal modification. *Antimicrob Agents Chemother 38*, 2311–2316.
276. Rodriguez-Hernandez MJ, Saugar J, Docobo-Perez F, de la Torre BG, Pachon-Ibanez ME, Garcia-Curiel A, Fernandez-Cuenca F, Andreu D, Rivas L, Pachon J (2006). Studies on the antimicrobial activity of cecropin A-melittin hybrid peptides in colistin-resistant clinical isolates of *Acinetobacter baumannii*. *J Antimicrob Chemother 58*, 95–100.
277. Saugar JM, Alarcon T, Lopez-Hernandez S, Lopez-Brea M, Andreu D, Rivas L (2002). Activities of polymyxin B and cecropin A-,melittin peptide CA(1–8)M(1–18) against a multiresistant strain of *Acinetobacter baumannii*. *Antimicrob Agents Chemother 46*, 875–888.
278. Shin SY, Kang JH, Hahm KS (1999). Structure-antibacterial, antitumor and hemolytic activity relationships of cecropin A-magainin 2 and cecropin A-melittin hybrid peptides. *J Pept Res 53*, 82–90.
279. Wade D, Boman A, Wahlin B, Drain CM, Andreu D, Boman HG, Merrifield RB (1990). All-D amino acid-containing channel-forming antibiotic peptides. *Proc Natl Acad Sci USA 87*, 4761–4765.
280. Shin SY, Kang JH, Lee MK, Kim SY, Kim Y, Hahm KS (1998). Cecropin A–magainin 2 hybrid peptides having potent antimicrobial activity with low hemolytic effect. *Biochem Mol Biol Int 44*, 1119–1126.
281. Avrahami D, Oren Z, Shai Y (2001). Effect of multiple aliphatic amino acids substitutions on the structure, function, and mode of action of diastereomeric membrane active peptides. *Biochemistry 40*, 12591–12603.
282. Papo N, Oren Z, Pag U, Sahl HG, Shai Y (2002). The consequence of sequence alteration of an amphipathic alpha-helical antimicrobial peptide and its diastereomers. *J Biol Chem 277*, 33913–33921.
283. Papo N, Shai Y (2005). A molecular mechanism for lipopolysaccharide protection of Gram-negative bacteria from antimicrobial peptides. *J Biol Chem 280*, 10378–10387.
284. Song YM, Yang ST, Lim SS, Kim Y, Hahm KS, Kim JI, Shin SY (2004). Effects of L- or D-Pro incorporation into hydrophobic or hydrophilic helix face of amphipathic alpha-helical model peptide on structure and cell selectivity. *Biochem Biophys Res Commun 314*, 615–621.
285. Stark M, Liu LP, Deber CM (2002). Cationic hydrophobic peptides with antimicrobial activity. *Antimicrob Agents Chemother 46*, 3585–3590.
286. Dennison SR, Wallace J, Harris F, Phoenix DA (2005). Amphiphilic alpha-helical antimicrobial peptides and their structure/function relationships. *Protein Pept Lett 12*, 31–39.
287. Deslouches B, Islam K, Craigo JK, Paranjape SM, Montelaro RC, Mietzner TA (2005). Activity of the de novo engineered antimicrobial peptide WLBU2 against *Pseudomonas aeruginosa* in human serum and whole blood: Implications for systemic applications. *Antimicrob Agents Chemother 49*, 3208–3216.
288. Jia X, Patrzykat A, Devlin RH, Ackerman PA, Iwama GK, Hancock RE (2000). Antimicrobial peptides protect coho salmon from *Vibrio anguillarum* infections. *Appl Environ Microbiol 66*, 1928–1932.
289. Oren Z, Shai Y (1996). A class of highly potent antibacterial peptides derived from pardaxin, a pore-forming peptide isolated from Moses sole fish *Pardachirus marmoratus*. *Eur J Biochem 237*, 303–310.
290. Jung HJ, Park Y, Sung WS, Suh BK, Lee J, Hahm KS, Lee DG (2007). Fungicidal effect of pleurocidin by

membrane-active mechanism and design of enantiomeric analogue for proteolytic resistance. *Biochim Biophys Acta 1768*, 1400–1405.

291. Vunnam S, Juvvadi P, Rotondi KS, Merrifield RB (1998). Synthesis and study of normal, enantio, retro, and retroenantio isomers of cecropin A-melittin hybrids, their end group effects and selective enzyme inactivation. *J Pept Res 51*, 38–44.
292. Benincasa M, Skerlavaj B, Gennaro R, Pellegrini A, Zanetti M (2003). In vitro and in vivo antimicrobial activity of two alpha-helical cathelicidin peptides and of their synthetic analogs. *Peptides 24*, 1723–1731.
293. Subbalakshmi C, Nagaraj R, Sitaram N (1999). Biological activities of C-terminal 15-residue synthetic fragment of melittin: Design of an analog with improved antibacterial activity. *FEBS Lett 448*, 62–66.
294. Tang YC, Deber CM (2002). Hydrophobicity and helicity of membrane-interactive peptides containing peptoid residues. *Biopolymers 65*, 254–262.
295. Tang YC, Deber CM (2004). Aqueous solubility and membrane interactions of hydrophobic peptides with peptoid tags. *Biopolymers 76*, 110–118.
296. Song YM, Park Y, Lim SS, Yang ST, Woo ER, Park IS, Lee JS, Kim JI, Hahm KS, Kim Y, Shin SY (2005). Cell selectivity and mechanism of action of antimicrobial model peptides containing peptoid residues. *Biochemistry 44*, 12094–12106.
297. Jiang Z, Vasil AI, Hale JD, Hancock RE, Vasil ML, Hodges RS (2008). Effects of net charge and the number of positively charged residues on the biological activity of amphipathic α-helical cationic antimicrobial peptides. *Biopolymers (Peptide Science) 90*, 369–383.
298. Asthana N, Yadav SP, Ghosh JK (2004). Dissection of antibacterial and toxic activity of melittin: A leucine zipper motif plays a crucial role in determining its hemolytic activity but not antibacterial activity. *J Biol Chem 279*, 55042–55050.
299. Lee DG, Kim PI, Park Y, Woo ER, Choi JS, Choi CH, Hahm KS (2002). Design of novel peptide analogs with potent fungicidal activity, based on PMAP-23 antimicrobial peptide isolated from porcine myeloid. *Biochem Biophys Res Commun 293*, 231–238.
300. Zhang L, Benz R, Hancock RE (1999). Influence of proline residues on the antibacterial and synergistic activities of alpha-helical peptides. *Biochemistry 38*, 8102–8111.
301. Sharon M, Oren Z, Shai Y, Anglister J (1999). 2D-NMR and ATR-FTIR study of the structure of a cell-selective diastereomer of melittin and its orientation in phospholipids. *Biochemistry 38*, 15305–15316.
302. Chen Y, Mant CT, Hodges RS (2002). Determination of stereochemistry stability coefficients of amino acid sidechains in an amphipathic alpha-helix. *J Pept Res 59*, 18–33.
303. Kondejewski LH, Jelokhani-Niaraki M, Farmer SW, Lix B, Kay CM, Sykes BD, Hancock RE, Hodges RS (1999). Dissociation of antimicrobial and hemolytic activities in cyclic peptide diastereomers by systematic alterations in amphipathicity. *J Biol Chem 274*, 13181–13192.
304. Kondejewski LH, Farmer SW, Wishart DS, Kay CM, Hancock RE, Hodges RS (1996). Modulation of structure and antibacterial and hemolytic activity by ring size in cyclic gramicidin S analogs. *J Biol Chem 271*, 25261–25268.
305. McInnes C, Kondejewski LH, Hodges RS, Sykes BD (2000). Development of the structural basis for antimicrobial and hemolytic activities of peptides based on gramicidin S and design of novel analogs using NMR spectroscopy. *J Biol Chem 275*, 14287–14294.
306. Kondejewski LH, Lee DL, Jelokhani-Niaraki M, Farmer SW, Hancock RE, Hodges RS (2002). Optimization of microbial specificity in cyclic peptides by modulation of hydrophobicity within a defined structural framework. *J Biol Chem 277*, 67–74.
307. Lee DL, Mant CT, Hodges RS (2003). A novel method to measure self-association of small amphipathic molecules: Temperature profiling in reversed-phase chromatography. *J Biol Chem 278*, 22918–22927.
308. Lee DL, Powers JP, Pflegerl K, Vasil ML, Hancock RE, Hodges RS (2004). Effects of single D-amino acid substitutions on disruption of beta-sheet structure and hydrophobicity in cyclic 14-residue antimicrobial peptide analogs related to gramicidin S. *J Pept Res 63*, 69–84.
309. Zhang L, Falla T, Wu M, Fidai S, Burian J, Kay W, Hancock RE (1998). Determinants of recombinant production of antimicrobial cationic peptides and creation of peptide variants in bacteria. *Biochem Biophys Res Commun 247*, 674–680.
310. Avrahami D, Shai Y (2002). Conjugation of a magainin analogue with lipophilic acids controls hydrophobicity, solution assembly, and cell selectivity. *Biochemistry 41*, 2254–2263.
311. Dathe M, Wieprecht T, Nikolenko H, Handel L, Maloy WL, MacDonald DL, Beyermann M, Bienert M (1997). Hydrophobicity, hydrophobic moment and angle subtended by charged residues modulate antibacterial and haemolytic activity of amphipathic helical peptides. *FEBS Lett 403*, 208–212.
312. Wieprecht T, Dathe M, Beyermann M, Krause E, Maloy WL, MacDonald DL, Bienert M (1997). Peptide hydrophobicity controls the activity and selectivity of magainin 2 amide in interaction with membranes. *Biochemistry 36*, 6124–6132.
313. Lugtenberg B, Van Alphen L (1983). Molecular architecture and functioning of the outer membrane of *Escherichia coli* and other gram-negative bacteria. *Biochim Biophys Acta 737*, 51–115.
314. Zilberstein D, Schuldiner S, Padan E (1979). Proton electrochemical gradient in *Escherichia coli* cells and its relation to active transport of lactose. *Biochemistry 18*, 669–673.
315. Shai Y (1999). Mechanism of the binding, insertion and destabilization of phospholipid bilayer membranes by alpha-helical antimicrobial and cell non-selective

membrane-lytic peptides. *Biochim Biophys Acta 1462*, 55–70.

316. Bland JM, De Lucca AJ, Jacks TJ, Vigo CB (2001). All-D-cecropin B: Synthesis, conformation, lipopolysaccharide binding, and antibacterial activity. *Mol Cell Biochem 218*, 105–111.
317. De Lucca AJ, Bland JM, Vigo CB, Jacks TJ, Peter J, Walsh TJ (2000). D-cecropin B: Proteolytic resistance, lethality for pathogenic fungi and binding properties. *Med Mycol 38*, 301–308.
318. Cribbs DH, Pike CJ, Weinstein SL, Velazquez P, Cotman CW (1997). All-D-enantiomers of beta-amyloid exhibit similar biological properties to all-L-beta-amyloids. *J Biol Chem 272*, 7431–7436.
319. Hamamoto K, Kida Y, Zhang Y, Shimizu T, Kuwano K (2002). Antimicrobial activity and stability to proteolysis of small linear cationic peptides with D-amino acid substitutions. *Microbiol Immunol 46*, 741–749.
320. Elmquist A, Langel U (2003). In vitro uptake and stability study of pVEC and its all-D analog. *Biol Chem 384*, 387–393.
321. Hong SY, Oh JE, Lee KH (1999). Effect of D-amino acid substitution on the stability, the secondary structure, and the activity of membrane-active peptide. *Biochem Pharmacol 58*, 1775–1780.
322. Wakabayashi H, Matsumoto H, Hashimoto K, Teraguchi S, Takase M, Hayasawa H (1999). N-Acylated and D enantiomer derivatives of a nonamer core peptide of lactoferricin B showing improved antimicrobial activity. *Antimicrob Agents Chemother 43*, 1267–1269.
323. Klevens RM, Morrison MA, Nadle J, Petit S, Gershman K, Ray S, Harrison LH, Lynfield R, Dumyati G, Townes JM, Craig AS, Zell ER, Fosheim GE, McDougal LK, Carey RB, Fridkin SK (2007). Invasive methicillin-resistant *Staphylococcus aureus* infections in the United States. *JAMA 298*, 1763–1771.
324. Bazzo R, Tappin MJ, Pastore A, Harvey TS, Carver JA, Campbell ID (1988). The structure of melittin. A 1H-NMR study in methanol. *Eur J Biochem 173*, 139–146.
325. Koltay A, Daly, NL, Gustafson, KR, Craik, DJ (2005). Structure of circullin B and implications for antimicrobial activity of the cyclotides. *Int J Pept Protein Res 11*, 99–106.
326. DeLano WL (2002). *The PyMOL Molecular Graphics System*. DeLano Scientific, San Carlos, CA, USA.
327. Schiffer M, Edmundson AB (1967). Use of helical wheels to represent the structures of proteins and to identify segments with helical potential. *Biophys J 7*, 121–135.
328. Donati M, Di Leo K, Benincasa M, Cavrini F, Accardo S, Moroni A, Gennaro R, Cevenini R (2005). Activity of cathelicidin peptides against *Chlamydia* spp. *Antimicrob Agents Chemother 49*, 1201–1202.
329. Donati M, Di Francesco A, Gennaro R, Benincasa M, Magnino S, Pignanelli S, Shurdhi A, Moroni A, Mazzoni C, Merialdi G, Baldelli R, Cevenini R (2007). Sensitivity of *Chlamydia suis* to cathelicidin peptides. *Vet Microbiol 123*, 269–273.
330. Lee IH, Zhao C, Nguyen T, Menzel L, Waring AJ, Sherman MA, Lehrer RI (2001). Clavaspirin, an antibacterial and haemolytic peptide from *Styela clava*. *J Pept Res 58*, 445–456.
331. Mor A, Nicolas P (1994). Isolation and structure of novel defensive peptides from frog skin. *Eur J Biochem 219*, 145–154.
332. Feder R, Dagan A, Mor A (2000). Structure-activity relationship study of antimicrobial dermaseptin S4 showing the consequences of peptide oligomerization on selective cytotoxicity. *J Biol Chem 275*, 4230–4238.
333. Krugliak M, Feder R, Zolotarev VY, Gaidukov L, Dagan A, Ginsburg H, Mor A (2000). Antimalarial activities of dermaseptin S4 derivatives. *Antimicrob Agents Chemother 44*, 2442–2451.
334. Lorin C, Saidi H, Belaid A, Zairi A, Baleux F, Hocini H, Belec L, Hani K, Tangy F (2005). The antimicrobial peptide dermaseptin S4 inhibits HIV-1 infectivity in vitro. *Virology 334*, 264–275.
335. Zairi A, Tangy F, Ducos-Galand M, Alonso JM, Hani K (2007). Susceptibility of *Neisseria gonorrhoeae* to antimicrobial peptides from amphibian skin, dermaseptin, and derivatives. *Diagn Microbiol Infect Dis 57*, 319–324.
336. Mor A, Amiche M, Nicolas P (1994). Structure, synthesis, and activity of dermaseptin b, a novel vertebrate defensive peptide from frog skin: Relationship with adenoregulin. *Biochemistry 33*, 6642–6650.
337. Strahilevitz J, Mor A, Nicolas P, Shai Y (1994). Spectrum of antimicrobial activity and assembly of dermaseptin-b and its precursor form in phospholipid membranes. *Biochemistry 33*, 10951–10960.
338. Scocchi M, Bontempo D, Boscolo S, Tomasinsig L, Giulotto E, Zanetti M (1999). Novel cathelicidins in horse leukocytes(1). *FEBS Lett 457*, 459–464.
339. Skerlavaj B, Scocchi M, Gennaro R, Risso A, Zanetti M (2001). Structural and functional analysis of horse cathelicidin peptides. *Antimicrob Agents Chemother 45*, 715–722.
340. Oppenheim FG, Xu T, McMillian FM, Levitz SM, Diamond RD, Offner GD, Troxler RF (1988). Histatins, a novel family of histidine-rich proteins in human parotid secretion. Isolation, characterization, primary structure, and fungistatic effects on *Candida albicans*. *J Biol Chem 263*, 7472–7477.
341. Xu T, Telser E, Troxler RF, Oppenheim FG (1990). Primary structure and anticandidal activity of the major histatin from parotid secretion of the subhuman primate, *Macaca fascicularis*. *J Dent Res 69*, 1717–1723.
342. Xu T, Levitz SM, Diamond RD, Oppenheim FG (1991). Anticandidal activity of major human salivary histatins. *Infect Immun 59*, 2549–2554.
343. Tsai H, Bobek LA (1997). Human salivary histatin-5 exerts potent fungicidal activity against *Cryptococcus neoformans*. *Biochim Biophys Acta 1336*, 367–369.

344. Helmerhorst EJ, Reijnders IM, van't Hof W, Simoons-Smit I, Veerman EC, Amerongen AV (1999). Amphotericin B and fluconazole-resistant *Candida* spp, *Aspergillus fumigatus*, and other newly emerging pathogenic fungi are susceptible to basic antifungal peptides. *Antimicrob Agents Chemother 43*, 702–704.

345. Situ H, Bobek LA (2000). In vitro assessment of antifungal therapeutic potential of salivary histatin-5, two variants of histatin-5, and salivary mucin (MUC7) domain 1. *Antimicrob Agents Chemother 44*, 1485–1493.

346. van't Hof W, Reijnders IM, Helmerhorst EJ, Walgreen-Weterings E, Simoons-Smit IM, Veerman EC, Amerongen AV (2000). Synergistic effects of low doses of histatin 5 and its analogues on amphotericin B anti-mycotic activity. *Ant Leeuwenhoek 78*, 163–169.

347. Rothstein DM, Spacciapoli P, Tran LT, Xu T, Roberts FD, Dalla Serra M, Buxton DK, Oppenheim FG, Friden P (2001). Anticandida activity is retained in P-113, a 12-amino-acid fragment of histatin 5. *Antimicrob Agents Chemother 45*, 1367–1373.

348. Sajjan US, Tran LT, Sole N, Rovaldi C, Akiyama A, Friden PM, Forstner JF, Rothstein DM (2001). P-113D, an antimicrobial peptide active against *Pseudomonas aeruginosa*, retains activity in the presence of sputum from cystic fibrosis patients. *Antimicrob Agents Chemother 45*, 3437–3444.

349. Brewer D, Lajoie G (2002). Structure-based design of potent histatin analogues. *Biochemistry 41*, 5526–5536.

350. Nikawa H, Jin C, Fukushima H, Makihira S, Hamada T (2001). Antifungal activity of histatin-5 against non-albicans *Candida species*. *Oral Microbiol Immunol 16*, 250–252.

351. Yan L, Adams ME (1998). Lycotoxins, antimicrobial peptides from venom of the wolf spider *Lycosa carolinensis*. *J Biol Chem 273*, 2059–2066.

352. Zasloff M, Martin B, Chen HC (1988). Antimicrobial activity of synthetic magainin peptides and several analogues. *Proc Natl Acad Sci USA 85*, 910–913.

353. Cruz-Chamorro L, Puertollano MA, Puertollano E, de Cienfuegos GA, de Pablo MA (2006). In vitro biological activities of magainin alone or in combination with nisin. *Peptides 27*, 1201–1209.

354. Moore KS, Bevins CL, Brasseur MM, Tomassini N, Turner K, Eck H, Zasloff M (1991). Antimicrobial peptides in the stomach of Xenopus laevis. *J Biol Chem 266*, 19851–19857.

355. Feldman ST, Speaker M, Cleveland P (1991). Effect of magainins on *Acanthamoeba castellanii*. *Rev Infect Dis 13*(Suppl 5), S439.

356. Albiol Matanic VC, Castilla V (2004). Antiviral activity of antimicrobial cationic peptides against Junin virus and herpes simplex virus. *Int J Antimicrob Agents 23*, 382–389.

357. Bessalle R, Kapitkovsky A, Gorea A, Shalit I, Fridkin M (1990). All-D-magainin: Chirality, antimicrobial activity and proteolytic resistance. *FEBS Lett 274*, 151–155.

358. Lehmann J, Retz M, Sidhu SS, Suttmann H, Sell M, Paulsen F, Harder J, Unteregger G, Stockle M (2006). Antitumor activity of the antimicrobial peptide magainin II against bladder cancer cell lines. *Eur Urol 50*, 141–147.

359. Bechinger B, Zasloff M, Opella SJ (1993). Structure and orientation of the antibiotic peptide magainin in membranes by solid-state nuclear magnetic resonance spectroscopy. *Protein Sci 2*, 2077–2084.

360. Giacometti A, Cirioni O, Greganti G, Quarta M, Scalise G (1998). In vitro activities of membrane-active peptides against gram-positive and gram-negative aerobic bacteria. *Antimicrob Agents Chemother 42*, 3320–3324.

361. Giacometti A, Cirioni O, Barchiesi F, Fortuna M, Scalise G (1999). In-vitro activity of cationic peptides alone and in combination with clinically used antimicrobial agents against *Pseudomonas aeruginosa*. *J Antimicrob Chemother 44*, 641–645.

362. Giacometti A, Cirioni O, Barchiesi F, Ancarani F, Scalise G (2000). In vitro anti-cryptosporidial activity of cationic peptides alone and in combination with inhibitors of ion transport systems. *J Antimicrob Chemother 45*, 651–654.

363. Oard S, Rush MC, Oard JH (2004). Characterization of antimicrobial peptides against a US strain of the rice pathogen *Rhizoctonia solani*. *J Appl Microbiol 97*, 169–180.

364. Kuhn-Nentwig L, Muller J, Schaller J, Walz A, Dathe M, Nentwig W (2002). Cupiennin 1, a new family of highly basic antimicrobial peptides in the venom of the spider Cupiennius salei (Ctenidae). *J Biol Chem 277*, 11208–11216.

365. Bucki R, Pastore JJ, Randhawa P, Vegners R, Weiner DJ, Janmey PA (2004). Antibacterial activities of rhodamine B-conjugated gelsolin-derived peptides compared to those of the antimicrobial peptides cathelicidin LL37, magainin II, and melittin. *Antimicrob Agents Chemother 48*, 1526–1533.

366. Oren Z, Hong J, Shai Y (1997). A repertoire of novel antibacterial diastereomeric peptides with selective cytolytic activity. *J Biol Chem 272*, 14643–14649.

367. Moerman L, Bosteels S, Noppe W, Willems J, Clynen E, Schoofs L, Thevissen K, Tytgat J, Van Eldere J, Van Der Walt J, Verdonck F (2002). Antibacterial and antifungal properties of alpha-helical, cationic peptides in the venom of scorpions from southern Africa. *Eur J Biochem 269*, 4799–4810.

368. Lee DG, Kim PI, Park Y, Jang SH, Park SC, Woo ER, Hahm KS (2002). HP (2–20) derived from the amino terminal region of helicobacterpylori ribosomal protein L1 exerts its antifungal effects by damaging the plasma membranes of *Candida albicans*. *J Pept Sci 8*, 453–460.

369. Munoz A, Lopez-Garcia B, Marcos JF (2006). Studies on the mode of action of the antifungal hexapeptide PAF26. *Antimicrob Agents Chemother 50*, 3847–3855.

370. Beven L, Castano S, Dufourcq J, Wieslander A, Wroblewski H (2003). The antibiotic activity of cationic linear amphipathic peptides: Lessons from the action of

leucine/lysine copolymers on bacteria of the class Mollicutes. *Eur J Biochem 270*, 2207–2217.

371. Chinchar VG, Bryan L, Silphadaung U, Noga E, Wade D, Rollins-Smith L (2004). Inactivation of viruses infecting ectothermic animals by amphibian and piscine antimicrobial peptides. *Virology 323*, 268–275.
372. Moon DO, Park SY, Choi YH, Kim ND, Lee C, Kim GY (2008). Melittin induces Bcl-2 and caspase-3-dependent apoptosis through downregulation of Akt phosphorylation in human leukemic U937 cells. *Toxicon 51*, 112–120.
373. Douglas SE, Patrzykat A, Pytyck J, Gallant JW (2003). Identification, structure and differential expression of novel pleurocidins clustered on the genome of the winter flounder, *Pseudopleuronectes americanus* (Walbaum). *Eur J Biochem 270*, 3720–3730.
374. Birkemo GA, Luders T, Andersen O, Nes IF, Nissen-Meyer J (2003). Hipposin, a histone-derived antimicrobial peptide in Atlantic halibut (*Hippoglossus hippoglossus* L.). *Biochim Biophys Acta 1646*, 207–215.
375. Brogden KA, Nordholm G, Ackermann M (2007). Antimicrobial activity of cathelicidins BMAP28, SMAP28, SMAP29, and PMAP23 against *Pasteurella multocida* is more broad-spectrum than host species specific. *Vet Microbiol 119*, 76–81.
376. Kang JH, Shin SY, Jang SY, Kim KL, Hahm KS (1999). Effects of tryptophan residues of porcine myeloid antibacterial peptide PMAP-23 on antibiotic activity. *Biochem Biophys Res Commun 264*, 281–286.
377. Lee DG, Kim DH, Park Y, Kim HK, Kim HN, Shin YK, Choi CH, Hahm KS (2001). Fungicidal effect of antimicrobial peptide, PMAP-23, isolated from porcine myeloid against *Candida albicans*. *Biochem Biophys Res Commun 282*, 570–574.
378. Park Y, Jang SH, Lee DG, Hahm KS (2004). Antinematodal effect of antimicrobial peptide, PMAP-23, isolated from porcine myeloid against *Caenorhabditis elegans*. *J Pept Sci 10*, 304–311.
379. Orivel J, Redeker V, Le Caer JP, Krier F, Revol-Junelles AM, Longeon A, Chaffotte A, Dejean A, Rossier J (2001). Ponericins, new antibacterial and insecticidal peptides from the venom of the ant *Pachycondyla goeldii*. *J Biol Chem 276*, 17823–17829.
380. Olson L, 3rd, Soto AM, Knoop FC, Conlon JM (2001). Pseudin-2: An antimicrobial peptide with low hemolytic activity from the skin of the paradoxical frog. *Biochem Biophys Res Commun 288*, 1001–1005.
381. Hultmark D, Engstrom A, Bennich H, Kapur R, Boman HG (1982). Insect immunity: Isolation and structure of cecropin D and four minor antibacterial components from *Cecropia pupae*. *Eur J Biochem 127*, 207–217.
382. Andreu D, Merrifield RB, Steiner H, Boman HG (1983). Solid-phase synthesis of cecropin A and related peptides. *Proc Natl Acad Sci USA 80*, 6475–6479.
383. Andreu D, Merrifield RB, Steiner H, Boman HG (1985). N-terminal analogues of cecropin A: Synthesis, antibacterial activity, and conformational properties. *Biochemistry 24*, 1683–1688.
384. Li ZQ, Merrifield RB, Boman IA, Boman HG (1988). Effects on electrophoretic mobility and antibacterial spectrum of removal of two residues from synthetic sarcotoxin IA and addition of the same residues to cecropin B. *FEBS Lett 231*, 299–302.
385. Steiner H, Andreu D, Merrifield RB (1988). Binding and action of cecropin and cecropin analogues: Antibacterial peptides from insects. *Biochim Biophys Acta 939*, 260–266.
386. Boman HC, Boman IA, Andreu D, Li ZQ, Merrifield RB, Schlenstedt G, Zimmermann R (1989). Chemical synthesis and enzymic processing of precursor forms of cecropins A and B. *J Biol Chem 264*, 5852–5860.
387. Kimbrell DA (1991). Insect antibacterial proteins: Not just for insects and against bacteria. *Bioessays 13*, 657–663.
388. Chen HM, Wang W, Smith D, Chan SC (1997). Effects of the anti-bacterial peptide cecropin B and its analogs, cecropins B-1 and B-2, on liposomes, bacteria, and cancer cells. *Biochim Biophys Acta 1336*, 171–179.
389. DeLucca AJ, Bland JM, Jacks TJ, Grimm C, Cleveland TE, Walsh TJ (1997). Fungicidal activity of cecropin A. *Antimicrob Agents Chemother 41*, 481–483.
390. Wachinger M, Kleinschmidt A, Winder D, von Pechmann N, Ludvigsen A, Neumann M, Holle R, Salmons B, Erfle V, Brack-Werner R (1998). Antimicrobial peptides melittin and cecropin inhibit replication of human immunodeficiency virus 1 by suppressing viral gene expression. *J Gen Virol 79*(Pt 4), 731–740.
391. Ekengren S, Hultmark D (1999). *Drosophila cecropin* as an antifungal agent. *Insect Biochem Mol Biol 29*, 965–972.
392. Andra J, Berninghausen O, Leippe M (2001). Cecropins, antibacterial peptides from insects and mammals, are potently fungicidal against *Candida albicans*. *Med Microbiol Immunol 189*, 169–173.
393. Shin SY, Lee KW, Kim Y, Kim JI, Hahm KS, Kang SW (2002). Structure-antibacterial activity relationship of cecropin A derivatives. *Protein Pept Lett 9*, 487–493.
394. Baldridge GD, Kurtti TJ, Munderloh UG (2005). Susceptibility of *Rickettsia monacensis* and *Rickettsia peacockii* to Cecropin A, Ceratotoxin A, and lysozyme. *Curr Microbiol 51*, 233–238.
395. Friedrich CL, Moyles D, Beveridge TJ, Hancock RE (2000). Antibacterial action of structurally diverse cationic peptides on gram-positive bacteria. *Antimicrob Agents Chemother 44*, 2086–2092.
396. Park Y, Lee DG, Jang SH, Woo ER, Jeong HG, Choi CH, Hahm KS (2003). A Leu-Lys-rich antimicrobial peptide: Activity and mechanism. *Biochim Biophys Acta 1645*, 172–182.
397. Shin SY, Kang JH, Jang SY, Kim Y, Kim KL, Hahm KS (2000). Effects of the hinge region of cecropin A(1–8)-magainin 2(1–12), a synthetic antimicrobial peptide, on liposomes, bacterial and tumor cells. *Biochim Biophys Acta 1463*, 209–218.

398. Park Y, Park SN, Park SC, Shin SO, Kim JY, Kang SJ, Kim MH, Jeong CY, Hahm KS (2006). Synergism of Leu-Lys rich antimicrobial peptides and chloramphenicol against bacterial cells. *Biochim Biophys Acta 1764*, 24–32.
399. Shin SY, Lee SH, Yang ST, Park EJ, Lee DG, Lee MK, Eom SH, Song WK, Kim Y, Hahm KS, Kim JI (2001). Antibacterial, antitumor and hemolytic activities of alpha-helical antibiotic peptide, P18 and its analogs. *J Pept Res 58*, 504–514.
400. Lee DG, Hahm KS, Shin SY (2004). Structure and fungicidal activity of a synthetic antimicrobial peptide, P18, and its truncated peptides. *Biotechnol Lett 26*, 337–341.
401. Lee SH, Lee DG, Yang ST, Kim Y, Kim JI, Hahm KS, Shin SY (2002). Antibiotic activity of reversed peptides of alpha-helical antimicrobial peptide, P18. *Protein Pept Lett 9*, 395–402.
402. Shin SY, Yang ST, Park EJ, Eom SH, Song WK, Kim Y, Hahm KS, Kim JI (2002). Salt resistance and synergistic effect with vancomycin of alpha-helical antimicrobial peptide P18. *Biochem Biophys Res Commun 290*, 558–562.
403. Braunstein A, Papo N, Shai Y (2004). In vitro activity and potency of an intravenously injected antimicrobial peptide and its DL amino acid analog in mice infected with bacteria. *Antimicrob Agents Chemother 48*, 3127–3129.
404. Pag U, Oedenkoven M, Papo N, Oren Z, Shai Y, Sahl HG (2004). In vitro activity and mode of action of diastereomeric antimicrobial peptides against bacterial clinical isolates. *J Antimicrob Chemother 53*, 230–239.
405. Rosenfeld Y, Papo N, Shai Y (2006). Endotoxin (lipopolysaccharide) neutralization by innate immunity host-defense peptides. Peptide properties and plausible modes of action. *J Biol Chem 281*, 1636–1643.

# ANTIBODIES FOR HUMAN DISEASES

Guo-Liang Yu[1] and Jian Ni[2]
[1]Epitomics, Inc., Burlingame, California, [2]Human Antibodomics Limited, Shanghai, China

## 1 PRINCIPLE OF MONOCLONAL ANTIBODY THERAPY

The immune system is our defense system that prevents us from illness and also works to help our recovery from many illnesses. The immune system plays a central role in protecting the body against various diseases such as infectious diseases, inflammation, and cancer. It is also the major force in the body in combating diseases that have already developed. The body's ability to develop an immune reaction to tumors may help determine which patients are cured of cancer using conventional therapies, including surgery, radiation, and drugs. The concept of immunotherapy is based on the body's natural defense system, which protects us against a variety of diseases [1].

Many years of research have resulted in the successful examples of immunotherapies for many diseases including cancers and inflammatory diseases. These new treatments referred to as biological therapies—such as monoclonal antibodies and vaccine therapies. The recent technological advances in monoclonal antibody (mAb) development caused significant investment and interest in the pharmaceutical and biotech industry. It is estimated that nearly half of all clinical trials are conducted on monoclonal antibodies [2]. Antibodies are complex protein molecules that are a component of the immune system. These molecules belong to the immunoglobulin (Ig) class of proteins. They are a major component of the body's adaptive immune system, which is involved in mounting a specific attack against "foreign" molecules, cells, and organisms by recognizing and binding to antigens. The cells that produce antibodies are known as B cells.

There are several classes of antibodies in humans including the IgG class (IgG1, IgG2, IgG3, and IgG4), IgM, IgA class, IgD, and IgE. IgG is most abundant class of antibodies with serum concentration of 10%. All therapeutic antibodies are IgG molecules, typically IgG1 or IgG2. The structure of antibodies is Y-shaped, consisting of two light chains and two heavy chains, which are linked by disulfide bonds (Fig. 1). The light and heavy chains contain a variable region, also known as the Fab (fragment antigen binding) region, and a constant region, which is also known as the Fc (fragment crystallizable) region. The Fab region is the main part of the molecule involved in antigen recognition. This region is encoded by genes that can be rearranged to achieve the extensive diversity of antigen recognition. The constant region, on the other hand, is involved in inducing the effector functions of antibodies and mediating an immune response [3].

High specificity, high affinity to a disease target and relatively low or no effect on other substance except the target antigen make mAb ideal molecules for therapeutic agents. Monoclonal antibodies are suited to target different target molecules and may offer different action mechanisms. In addition, it is possible to modify and improve an antibody drug candidate by genetic engineering methods and establish large-scale production using recombinant protein technology.

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

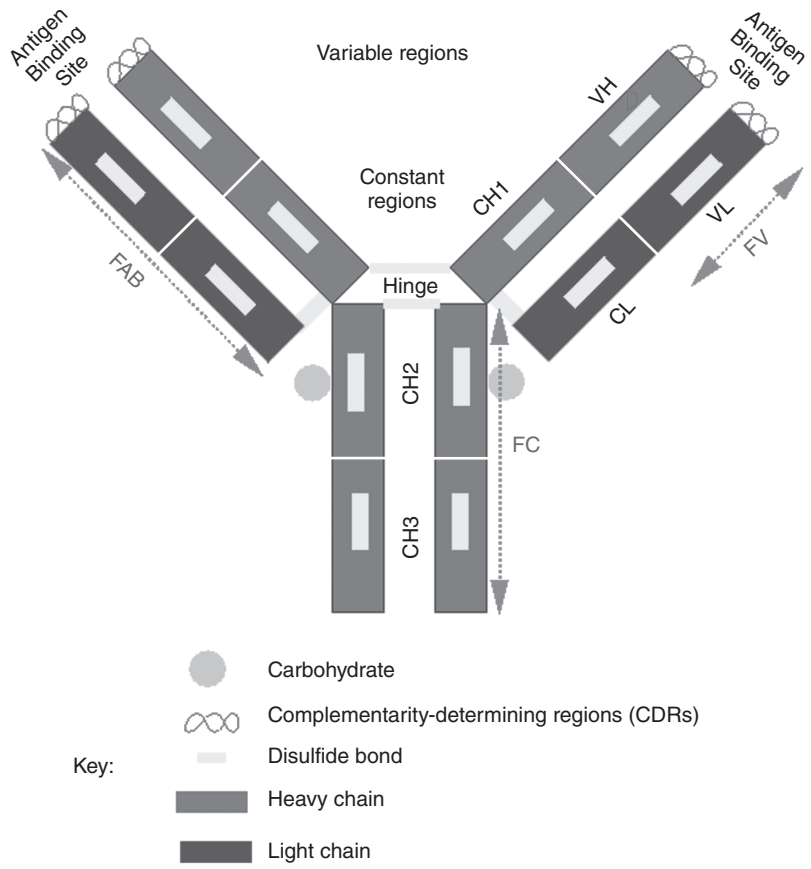


**Figure 1** Structure of an antibody, immunoglobulin (IgG).

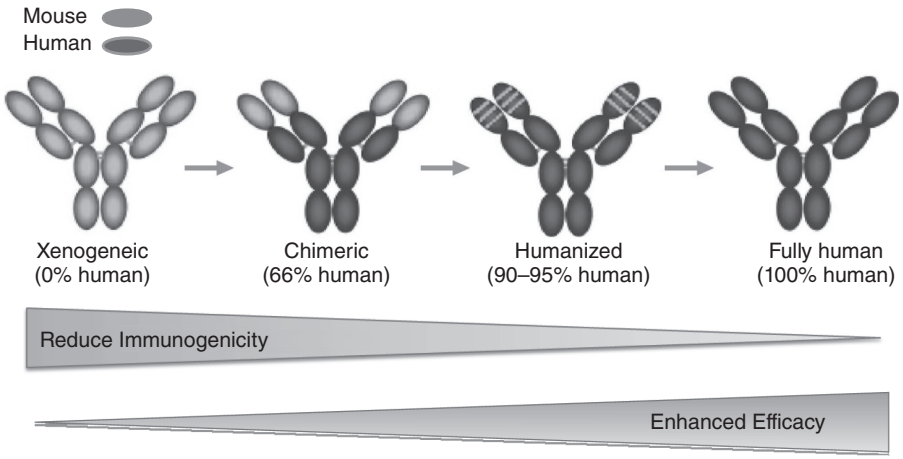


**Figure 2** Evolution of therapeutic antibody products.

## 2 HISTORY OF ANTIBODY THERAPEUTICS

The history of antibody therapeutics is an evolution of antibody engineering to produce antibodies with less and less immunogenicity reaction in humans. Figure 2 illustrates the four types of monoclonal antibodies from pure murine to partial human and from partial human to full human. Köhler and Milstein developed the hybridoma method of murine antibody production in 1975 [4], which won the Nobel Prize and allowed the production of the first mAb to market. In 1986, Johnson & Johnson developed Orthoclone OKT3 (muromonab) for the treatment of renal allograft rejection resistant to steroid and antilymphocyte globulin [5]. However,

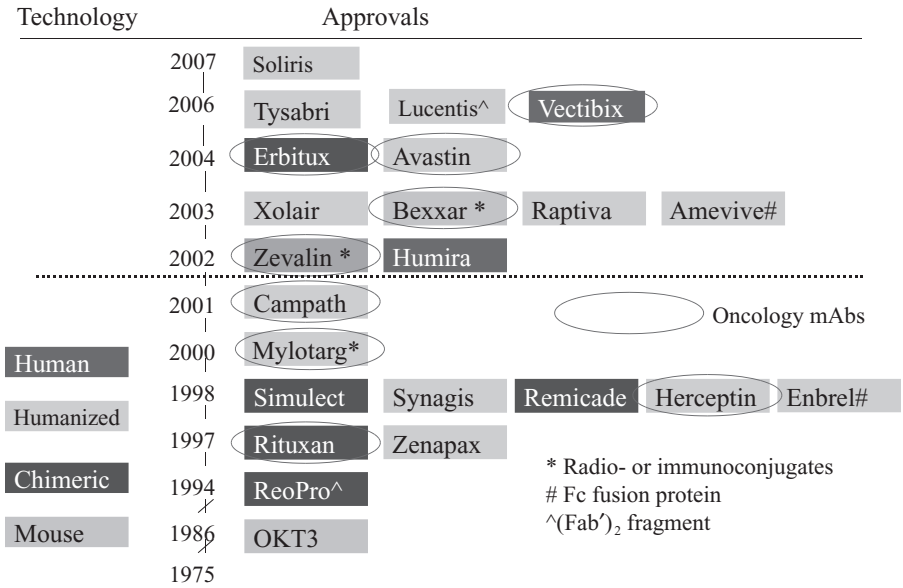


**Figure 3** U.S. approved therapeutic mAbs.

murine antibodies are from mice and the application of murine antibodies was extremely limited in humans because it caused an immunogenicity reaction, known as the HAMA (human antimouse antibodies) response. When administered to humans, murine antibodies cause serious side effects, including anaphylaxis and hypersensitivity reactions. This immunogenicity reaction is triggered by the human immune system recognizing the molecules as foreign because murine antibodies are different from natural human antibodies. In addition, murine monoclonal antibodies consequently have a short half-life relative to human antibodies.

In 1984, the first chimeric antibodies, where 30–35% of the molecule is derived from mouse antibody sequences and around 65–70% is derived from human antibody sequences, were reported [6, 7]. Chimeric monoclonal antibodies are generated by using recombinant deoxyribonucleic acid (DNA) technology, where the murine Fab region is fused to the human Fc region of an antibody. Chimeric antibodies turned out to be more successful therapeutic agents than murine monoclonal antibodies in overcoming the immunogenicity problems. The two most successful monoclonal antibody therapies in 2004, Biogen Idec/Genentech/Roche's Rituxan (rituximab) and Johnson & Johnson/Schering-Plough's Remicade (infliximab), are both chimeric. Their success has been attributed to the lack of immunogenicity reactions, as well as successful target selection. However, as a substantial portion of chimeric monoclonal antibodies are murine, there is still a considerable risk of the human immune system recognizing the molecules as foreign and eliciting an immunogenicity reaction, known as HACA (human antichimeric antibody) responses [3].

To mitigate this, the next technological development after chimeric antibodies to minimize immunogenicity reactions was humanization technology [3, 8]. Using protein engineering technologies, it can generate monoclonal antibodies where 90–95% of the sequences are of human origin. Human antibody generates about 10% diversity in the antibodies protein sequences during the antibody maturation process by gene splicing and mutations. Therefore, the remaining 5–10% nonhuman portions in the humanized antibodies become much tolerable in humans. The humanization technique initially uses genetic engineering to transfer the murine antigen-binding sequences, known as the complementarity-determining region (CDR), to a human antibody scaffold. The CDR region is part of the Fab region, which carries the antigen-binding specificity.

The CDR sequences transferred are usually optimally selected to those that confer the binding specificity and biological function, keeping murine sequences to a minimum in the resultant monoclonal antibodies. A number of techniques are now used for humanization or deimmunization. As a result of the fewer murine sequences, these molecules carry a lower risk of triggering an immunogenicity reaction than chimeric and murine antibodies. Humanized antibodies, like chimeric monoclonal antibodies, carry the ability to effectively induce immune responses because of the Fc region is humans. Most of antibody drugs on the market and in clinical development are humanized antibodies.

The initial methods for humanization are time consuming and costly. Advanced technologies such as Epitomics' "HuRAb technology" bring significant improvement for the humanization process (unpublished). HuRAb technology utilizes the sequence information in the antibody evolution to correlate rabbit monoclonal antibody sequences and their biological functions. Alignment of sequences of antibodies that derive from the same antibody maturation process and possess the same biological function allows the identification of residues that are critical to the functions of the antibodies and systematically convert the noncritical residues to human residues, resulting in a higher degree of homology to human antibody. Such a process can be accomplished within 2 weeks by one researcher compared with the almost 1-year effort using traditional methods.

To further minimize the effort to humanize murine antibodies, technologies were developed to generate human antibodies directly. Transgenic mice have had their innate antibody genes replaced by human antibody genes, while leaving intact the other components of the mouse immune system [3, 9]. As a result the transgenic mice produce human antibodies in response to immunization with an antigen. Once the mouse has been immunized, mAbs can either be generated through traditional hybridoma generation or by using technologies that involve the harvesting, plating, and screening of B cells, followed by isolation of mAb genes and cloning into production cell lines. One example of such success is panitumumab (anti-epidermal growth factor [EGF]) of Amgen (Abgenix). Another approach to generate human antibody is to screen a phage library generated using human antibody genes. Phage display involves the engineering of bacteriophages to display human monoclonal antibodies on their surface. Genes from human antibody gene libraries are inserted into a library of phage. Each phage carries the genes for a different antibody and displays a different antibody on its surface. Antibodies are selected for binding to a target antigen. The corresponding genes that encode the antibody protein can be recovered. Adalimumab [tumor necrosis factor-α (anti-TNF-α)] is the first human antibody developed by Abbot and AstraZeneca (Cambridge Antibody Technology) using phage display technology.

The traditional therapy areas for mAb therapy are oncology and autoimmune and inflammatory disorders. It is also anticipated that mAb therapy will play a role in infectious diseases, cardiovascular diseases, as well as degenerative diseases. The high level of innovation, as demonstrated by advancements in antibody engineering with the introduction of chimeric, human mAbs, lead the way for this new class of therapeutics to become a major class of modern medicine. Other innovations in antibody technology include advancements in developing small-sized antibody molecules, enhanced affector function, and multifunctional antibodies such as bispecific antibodies lead to a huge boost in developing drugs that utilize these technologies [9]. Table 1 list all monoclonal antibodies launched on the market by time and by class.

In summary, the driver of antibody industry growth to be considered is the evolution of antibody technology. Although murine monoclonal antibodies were the first class of mAbs to enter clinical trials, these molecules have failed to become a major class of therapeutic antibodies due to the immunogenicity problems they cause. As a result, chimeric and humanized antibodies have dominated the industry's marketed products. There is a level of interest at the research and development (R&D) level across all phases to develop human mAbs. While technology of antibodies has been a key factor in the success of the market to date, the therapeutic suitability of antibodies for the treatment of immune-related diseases, cancer, and infectious disease is also expected to continue to drive new innovations. Historically, the monoclonal antibody development has been led by new antibody engineering technologies.

## 3 ANTIBODY TECHNOLOGY

### 3.1 Murine Hybridoma

The first type of monoclonal antibodies that could be generated is murine monoclonal antibodies. Köhler and Milstein developed the Nobel-Prize-winning hybridoma technique in 1975, which can be used to produce cell lines capable of producing identical antibodies indefinitely [3, 4]. The technology generates cells called hybridomas that can multiply indefinitely having the characteristics of B cells, which produce antibodies, and myeloma cells, which can divide indefinitely. A hybridoma is created by fusing a B cell with a myeloma cell using either a chemical [polyethylene glycol (PEG)] or physical (electrofusion) method. This development marked a major breakthrough in the development of monoclonal antibodies as therapeutics, diagnostics, and research reagents.

The ability to produce monoclonal antibodies led to the investigation of these molecules as therapeutic agents, using them to either mimic natural human antibodies or to interfere with biochemical pathways through receptor interaction. Unfortunately, it was found that murine mAbs caused an immunogenicity reaction, known as the HAMA (Human Anti-Mouse Antibodies) response which causes serious side effects including anaphylaxis and hypersensitivity reactions,

**TABLE 1 U.S. FDA Approved Therapeutic mAbs**

| Ab and Target | Product | Company | Preparation | Approved | Indication |
|---|---|---|---|---|---|
| Muromomab-CD3 (anti-CD3) | Orthoclone OKT3 | Ortho Biotech J&J | Murine hybridoma | 1986.6 | Kidney-transplant rejection |
| Abciximab (anti-GPIIb/IIIa) | ReoPro | Centocor | NSO (Chimeric) | 1994.12 | Coronary angioplasty |
| Rituximab (anti-CD20) | Rituxan Mabthera | Roche/IDEC/Genentech | CHO (Chimeric) | 1997.11 | Non-Hodgkin's lymphoma |
| Daclizumab (anti-CD25) | Zenapax | Hoffman–La Roche | CHO (Humanized) | 1997.12 | Kidney-transplant rejection |
| Basiliximab (anti-CD25) | Simulect | Novartis | NS0 (Chimeric) | 1998.5 | Kidney-transplant rejection |
| Palivizumab (anti-F protein of RSV) | Synagis | Medlmmune | NSO (Humanized) | 1998.6 | RSV |
| Infliximab (anti-TNFα) | Remicade | Centocor | NSO (Chimeric) | 1998.8 | Crohn's disease, ra |
| Trastuzumab (anti-HER-2) | Herceptin | Roche/Genentech | CHO (Humanized) | 1998.9 | Breast Cancer |
| Etanercept (anti-TNFα) | Enbrel | Immunex/Wyeth | CHO (Fc fusion) | 1998.11 | RA |
| Gemtuzumab ozogamicin (CD33) | Mylotarg | Wyeth/AHP | NSO (Humanized) | 2000.5 | CD33+ acute myeloid leukemia |
| Alemtuzumab (anti-CD52) | Campath | Millennium/ILEX | CHO (Humanized) | 2001.5 | B-cell chronic lymphocytic leukemia |
| Ibritumomab tiuxetan (anti-CD20) | Zevalin | Biogen IDEC | CHO (Murine) | 2002.2 | Non-Hodgkin's lymphoma |
| Adalimumab (anti-TNFα) | Humira | Abbot Laboratories/CAT | CHO (Phage display, human) | 2002.12 | RA |
| Alefacept | Amevive | Biogen IDEC | CHO (Fc Fusion) | 2003.1 | Plaque psoriasis |
| $^{1}$131 Tositumomab (anti-CD20) | Bexxar | Corixa Corp. and GlaxoSmithKline | NS0 (Murine) | 2003.6 | Non-Hodgkin's lymphoma |
| Omalizumab (anti-IgE) | Xolair | Genentech/Tanox/Novartis | CHO (Humanized) | 2003.6 | Persistent asthma |
| Efalizumab (anti-CD11a) | Raptiva | Xoma/Genentech | CHO (Humanized) | 2003.1 | Plaque psoriasis |
| Cetuximab (anti-EGFR) | Erbitux | Imclone/BMS | NS0 (Chimeric) | 2004.2 | Metastatic colorectal cancer |
| Bevacizumab (anti-EGFR) | Avastin | Genentech | CHO (Humanized) | 2004.2 | Metastatic colorectal cancer |
| Natalizumab (α-4-integrin) | Tysabri | Biogen Idec/Elan Pharmaceuticals | NS0 (Humanized) | 2004.11 | Multiple sclerosis |
| Ranibizumab (anti-VEGF) | Lucentis | Genentech | E. coli (Humanized) | 2006.6 | Neovascular (wet) age-related macular degeneration |
| Panitumumab (anti-EGFR) | Vectibix | Amgen | CHO (Xenomice, human) | 2006.9 | Metastatic colorectal cancer |
| Eculizumab | Soliris | Alexion | NS0 (humanized) | 2007.3 | Paroxysmal nocturnal hemoglobinuria (PHN) |

when administered to humans. This immunogenicity reaction is triggered by the human immune system recognizing the molecules as foreign because the murine antibodies are different from natural human antibodies, at structure and sequence levels. In addition, murine monoclonal antibodies consequently have a short half-life because the human immune system recognizes murine antibodies as foreign. As a result of these issues, none of the many murine monoclonal antibodies was able to be used in clinic and there is only one marketed

murine monoclonal antibody, muromonab (OrthoClone OKT3). Consequently, there is no interest in using this technology to develop therapeutic agents, especially with the advancement in technology to produce more human-like antibodies. The use of murine antibodies is now largely focused on use as diagnostic and research agents, where immunogenicity issues are not an issue.

### 3.2 Humanized Antibodies

The next technological development after chimeric antibodies to minimize immunogenicity reactions was humanization technology. This can be used to generate monoclonal antibodies where 90–95% of the sequences are of human origin and 5–10% are of murine origin. The technique uses genetic engineering to transfer the murine antigen-binding sequences, known as the complementarity-determining region (CDR), to a human antibody scaffold [3, 8]. The CDR is part of the Fab, which carries the antigen-binding specificity. The CDR sequences transferred are usually optimally selected to those that confer the binding specificity and strength, keeping murine sequences to a minimum in the resultant monoclonal antibodies. As a result of the fewer murine sequences, these molecules carry a lower risk of triggering an immunogenicity reaction than chimeric and murine antibodies.

There are a number of antibody technology platforms that can be licensed to develop humanized antibodies. The key players include Protein Design Labs' SMART humanization technology, Aeres Biomedical's humanization technology, and Xoma's human engineering technology. However, developing humanized antibodies can be time consuming and engineering humanized monoclonal antibodies can be expensive. Most importantly, humanization can reduce binding affinity because certain murine sequences involved in binding strength may not be transferred to the final antibody molecule. As a result of this, sequence selection needs to be carefully done to ensure that binding affinity and biological activities are not reduced substantially.

Recently, a new humanization technology was developed by Epitomics termed HuRAB technology, which is designed to overcome the limitation of traditional humanization technology. The HuRAb technology utilizes information of genetic evolution of the rabbit immune system: B-cell maturation, to optimize and at the same time convert a rabbit monoclonal antibody to human antibody molecule. The cost and time to engineer an antibody is significantly reduced. Furthermore the resulting antibodies remain and sometimes have improved binding capability than their parental antibody molecules.

Developing humanized antibodies can be time consuming because murine antibodies specific to the antigen need to be developed before humanization can be carried out. Secondly, as a result of this, engineering humanized monoclonal antibodies can be expensive. Thirdly, and most importantly, humanization can reduce binding affinity because certain murine sequences involved in binding strength may not be transferred to the final antibody molecule. As a result of this, sequence selection needs to be carefully done to ensure that binding affinity and biological activities are not reduced substantially.

### 3.3 Rabbit Hybridoma

Rabbits are one of the best sources of high-affinity polyclonal antibodies. Traditionally, it has been difficult to raise monoclonal antibodies in the rabbit. The most efficient and practical way to immortalize antibody-secreting B lymphocytes is by fusing lymphocytes obtained from spleen or lymph node with myeloma cells. Myeloma cells are derived from a B-cell tumor that arises in mice and humans, but not in rabbits. These fusions result in hybrid cells, called hybridomas, which often maintain the differentiated properties of antibody-secreting plasma cells, while gaining the ability to multiply indefinitely in culture (Fig. 4). Since myeloma-like tumors were not available in the rabbit, it has been impossible to use this approach to raise rabbit monoclonal antibodies. Several laboratories developed mouse–rabbit heterohybridomas, but this technology has had limited success. The earliest mouse–rabbit heterohybridomas were unstable and some only secreted light chains [10–13]. Recombinant antibody approaches reported that rabbit monoclonal antibody genes were isolated from single-cell polymerase chain reaction (PCR) or phage display [14, 15]. However, its general application is limited by the difficulty of single-cell selection and PCR, or by technically demanding protocols of phage display.

In 1995, Katherine Knight and colleagues succeeded in producing a myeloma-like tumor (plasmacytoma) in transgenic rabbits expressing the *v-abl* and *c-myc* oncogenes under the control of the immunoglobulin heavy- and light-chain enhancers, respectively [16] . These researchers isolated a plasmacytoma cell line, termed 240E, and showed that fusions of 240E cells with rabbit lymphocytes produced stable hybridoma.

The availability of rabbit monoclonal antibodies is highly desirable [17]. First, antisera from rabbits are generally of higher affinity than the equivalents in mice. Preliminary result from Epitomics indicated that the binding affinity of rabbit monoclonal antibodies can be 100 times better (unpublished data). Second, rabbit monoclonal antibodies are expected to recognize many

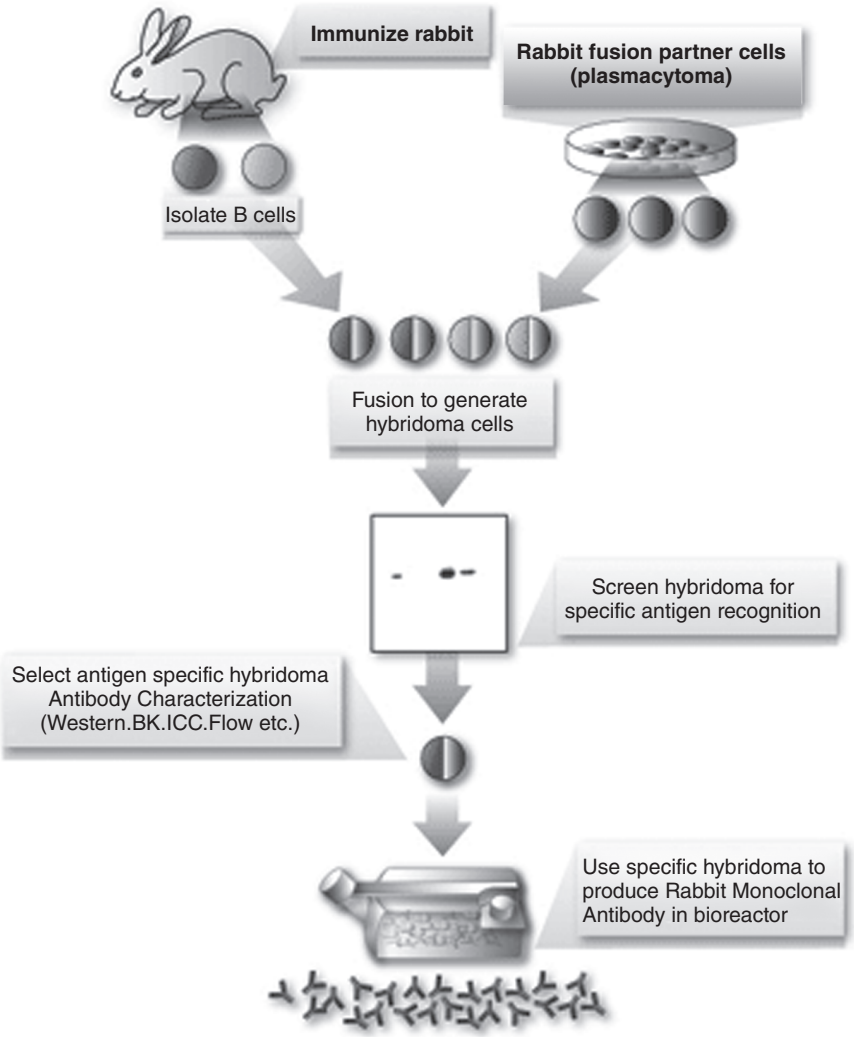


**Figure 4** Rabitt hybridoma.

antigens that are not immunogenic in mice, including mouse proteins [18–22]. For example, Bystryn et al. [19] compared rabbit and mouse antibodies directed against human melanoma cells and showed that they recognize different epitopes [19]. Third, Because the size of the rabbit spleen, more fusion experiments can be performed, making hybridoma generation at industrial scale a feasible task. Epitomics, Inc. has exclusively licensed 240E cell line from Loyola University and has developed a 240E derivative cell line that is highly stable and productive.

### 3.4 Human Hybridoma

Human antibodies are derived either from vectors carrying human antibody genes or from human cells. Human antibodies, therefore, theoretically have minimal immunogenicity. The difference between humanized and human mAbs is very small, since humanized antibodies already go virtually undetected by the human immune system because the antibodies produced by human differ by 10% themselves. Human monoclonal antibodies are still perceived to have favorable safety profiles and be eliminated less rapidly from the human body, potentially reducing the frequency and amount of dosing required. However, this is based on the assumption that the human antibody repertoire is sufficient to produce an antibody that is efficacious enough to be a good drug. Nevertheless, a further advantage is that human mAbs do not require humanization, a process that could be challenging and time consuming, and can result in antibodies with lowered affinities for their respective targets. Currently, two types of technologies are used to create human antibodies: transgenic mice and phage display of libraries of human antibodies [9].

After 20 years of effort, Karpas et al. in the University of Cambridge reported a human myeloma cell line (Karpas707H) that can give rise to stable human hybridomas capable of producing human monoclonal antibodies [23]. The cell line is useful for the generation of stable human hybridomas. It can easily be fused with ouabain-sensitive Epstein–Barr virus-transformed cells as well as with fresh tonsil and blood lymphocytes, giving rise to stable hybrids that continuously secrete very large quantities of human immunoglobulins. The derived hybrids do not lose immunoglobulin secretion over many months of continuous growth. The availability of this cell line should enable the in vitro immortalization of human antibody-producing B cells that are formed in vivo. The monoclonal antibodies produced may have advantages in immunotherapy. The cell line has the potential to immortalise most of the antibodies the human body produces to fight infections, cancer, acquired immunodeficiency syndrome (AIDS), and other pathological conditions [23]. Further molecular characterization study suggest that the hybridoma-derived antibodies are representative of antibodies from populations of human lymphocytes and at different stages in the maturation of the response; the use of Karpas707H myeloma for human hybridoma fusions may therefore provide a valuable tool for analysis of the human antibody responses [24]. Recently, Kametani et al. tried to prepare designed antigen-specific antibodies of completely human origin using immunodeficient mouse and human hybridoma [5]. Nonobese diabetic/severe combined immunodeficient/interleukin 2 (IL-2) receptor α null mouse (NOG) mouse was used to reconstitute the human immune system with umbilical cord blood hematopoietic stem cells (CB-NOG mouse) and to prepare human-derived Her-2-epitope-specific antibodies. Hybridoma lines were prepared by fusing the human myeloma cell line Karpas707H. Hybridoma lines were successfully prepared with spleen B cells obtained from the immunized CB-NOG mouse. One of these cell lines produced human IgM against the epitope peptide that can recognize the surface Her-2 molecule [25].

### 3.5 Transgenic Mice

XenoMouse or HuMouse are transgenic mice that have had their innate antibody genes replaced by human antibody genes, while leaving intact the other components of the mouse immune system [26–28]. As a result the transgenic mice produce human antibodies in response to immunization with an antigen. Once the mouse has been immunized, mAbs can either be generated through traditional hybridoma generation or by using genetic approaches to isolate B cells that produce the functional antibodies, followed by isolation of mAb genes and cloning into production cell lines. Within the field of transgenic mice, the two leading technology platforms are Medarex's HuMAb-Mouse and Abgenix's XenoMouse.

Medarex's UltiMAb platform uses transgenic mice, where mouse genes are suppressed and replaced with the expression of human antibody genes, to generate human antibodies. The technology platform includes three key antibody technologies; Medarex's HuMAb-Mouse technology; Kirin Brewery's TC Mouse technology; and the KM-Mouse technology, a crossbred mouse that has a combination of the characteristics of the HuMAb-Mouse and Kirin's TC Mouse technologies.

In the transgenic mice of the HuMAb-Mouse technology, mouse genes encoding antibodies have been inactivated and replaced with human antibody genes. These human genes encode unrearranged human antibody genes for both the heavy and light chains. Kirin's TC Mouse technology, on the other hand, has been developed so that the transgenic mice contain the complete set of variable and constant genes, which are found at the human immunoglobulin loci. In addition, the TC Mouse technology is capable of generating all antibody isotypes, as the mouse's chromosomes have been inactivated and replaced with human chromosomes that encode these. By combining the key features of these mice to create the KM-Mouse technology, Medarex and Kirin have generated a platform capable of making a broad range of antibodies because it has 100% of the human chromosomes for antibody genes. The ability to generate varied isotypes, in particular, allows a greater flexibility in designing therapeutic mAbs, as a particular isotype determines which effector functions of the immune response are induced.

Abgenix's XenoMouse and XenoMax are human antibody technologies using transgenic mice to create therapeutic monoclonal antibodies. Abgenix's XenoMouse is a transgenic mouse in which the mouse's antibody gene expression is suppressed and replaced with human antibody genes. However, the remainder of the mouse's immune system is left intact. The mice engineered through this process have approximately 80% of the human heavy-chain antibody genes as well as a significant amount of the light-chain genes. Abgenix has developed multiple strains of XenoMouse animals, each of which produces a different class of antibody to perform different therapeutic functions.

### 3.6 Phage Display

Phage display involves the engineering of bacteriophages to display human monoclonal antibodies on their surface [3, 28]. This is an important tool in studies aimed at identifying molecules, such as antibodies, that

bind to a specific target. Genes from extensive human antibody gene libraries are inserted into a population of phage. Each phage carries the genes for a different antibody and thus displays a different antibody on its surface. Antibodies are selected for binding to a target antigen, and the corresponding genes can be recovered and used in the development and potential manufacture of an antibody therapeutic product. The benefits of phage display are the huge size of the libraries. The larger the library, the more likely it is that an antibody can be isolated that binds effectively to the selected target. Certain libraries contain the genes for over 100 billion distinct antibodies. The recombination of the genes that code for the variable parts of an antibody allows further expansion of these libraries.

A number of companies are engaged in the development of human mAbs through phage display [29, 30] including Cambridge Antibody Technology (CAT) now part of AstraZenica, Dyax, and MorphoSys. CAT has created large libraries of billions of antibody genes collected from the blood of healthy individuals. CAT has used these gene libraries as the basis for the production of phage antibodies, created by phage display. By mixing the genes that code for the variable parts of an antibody (V genes) to give further combinations and inserting them into phage, CAT has created phage antibody libraries that now contain over 100 billion distinct phage antibodies. Six fully human therapeutic antibodies developed by CAT are at various stages of clinical trials, and Humira (adalimumab) was the first phage display antibody to reach the market, launched by Abbott in January 2003. Dyax's phage display libraries contain bacteriophages that display the Fab fragments of antibodies. Dyax developed its phage display libraries by taking gene fragments from human donors that encode the light-chain variable region and part of the heavy-chain variable region.

Advantages of phage display are the speed of identifying the initial antibody, which avoids the need for immunization in animals which usually takes several months [28]. However, the phage antibodies isolated in the first round of display are typically of very low affinity, and optimization (affinity maturation) of the antibody is necessary. Affinity maturation is the process by which B cells produce antibodies with increased affinity for antigen during the course of an immune response. With repeated exposures to the same antigen, a host will produce antibodies of successively greater affinities. A secondary response can elicit antibodies with several logfold greater affinities than in a primary response. Such affinity maturation process can be applied in test tubes by selecting a random mutated antibody population. Typically, in vitro affinity maturation takes a few months too.

Ribosome display involves the use of ribosomes to display functional antibody proteins in a test tube [29, 30]. Ribosome technology eliminates the need for phage particles and bacteria to generate antibodies and, therefore, represents a novel variation of phage display technology, with the potential to build even larger libraries. While the phage display technology is currently the more commercially important of the two technologies, ribosome display may be used as a supplement to phage display. Yeast display is a technique initially used in the field of protein engineering [31, 32]. A protein of interest is displayed as a fusion to the Aga2p protein on the surface of yeast. The Aga2p protein is naturally used by yeast to mediate cell–cell contacts during yeast cell mating. As such, display of a protein via Aga2p projects the protein away from the cell surface, minimizing potential interactions with other molecules on the yeast cell wall. The use of flow cytometry in conjunction with a yeast display library is a highly effective method to isolate high-affinity protein ligands (antigens) against nearly any receptor (antibodies) through directed evolution.

Advantages of yeast display over other in vitro evolution methods include eukaryotic expression and processing, quality control mechanisms of the eukaryotic secretory pathway, minimal avidity effects, and quantitative library screening through fluorescent-activated cell sorting (FACS). Disadvantages include smaller mutagenic library sizes compared to alternative methods and differential *glycosylation* in yeast compared to mammalian cells. It should be noted that these disadvantages have not limited the success of yeast display for a number of applications, including engineering the highest monovalent ligand-binding affinity reported to date for an engineered protein [33].

Mammalian cells such as Chinese hamster ovary cells are the dominant system used for the production of the majority of therapeutic antibodies in the pharmaceutical industry. Therefore, mammalian cells should be used in antibody selection to avoid problems caused by selection in microorganisms. A random peptide library fused to CCR5 and expressed on the mammalian cell surface has been reported [35]. The potential of this display system has been demonstrated for small peptides of 9 or 13 amino accids (aa). Ho et al. reported that human cells can be engineered for cell surface display of functional single-chain variable-fragment (scFv) antibodies and is suitable for affinity maturation. They used the anti-CD22 scFv fused to the transmembrane domain of human platelet-derived growth factor receptor (PDGFR) displayed on human embryonic kidney (HEK) 293T cells [34]. The mammalian cell scaffold allows hundredfold enrichment of a rare higher affinity mutant from a large excess of only slightly lower affinity initial antibody.

## 4 ANTIBODY ENGINEERING

### 4.1 Immunoconjugated Cancer Drugs

Monoclonal antibodies can be chemically conjugated to other molecules to create a different type of therapeutic agent. One way to achieve this is to use monoclonal antibodies as carriers of a small molecule or biologic drug to deliver drugs to a specific site, resulting in a localized and more direct action of the effector drug. This would limit the undesired effects on healthy cells that result from systemic treatment regimens, which is particularly problem with many cancer treatments available. At least three companies with key conjugation technology platforms: Seattle Genetics' antibody-drug conjugate (ADC) and antibody-directed enzyme prodrug therapy (ADEPT) technologies, ImmunoGen's tumor-activated prodrug technology, as well as Peregrine Pharmaceuticals' tumor necrosis therapy [36].

ImmunoGen's tumor-activated prodrug (TAP) technology is designed to conjugate a small-molecule prodrug to a given monoclonal antibody. In this situation the mAb acts as a carrier system, delivering the prodrug to specific cells by the recognition of a particular marker on cells. ImmunoGen's TAP technology creates a stable link between the mAb and the prodrug. The prodrug remains in its inactive state until it enters the specific cancer cell, when the link between the antibody and the prodrug is broken, releasing the prodrug. ImmunoGen is focused on using this technology to develop treatments for cancer, using mAbs that bind to specific markers associated with certain types of cancer.

Seattle Genetics has developed two conjugated antibody platforms: ADC and ADEPT technologies. The key components of the ADC technology are the stable, enzyme-cleavable linkers, such as hydrazones, and the highly potent, synthetic cytotoxic drugs. By using a rapidly internalizing mAb, the drug payload can be delivered inside target cells. The linkers are stable in the bloodstream, but the environment inside the target cell cleaves the linker, thereby releasing the drug payload. Until released, the drug is inactive, thereby sparing non-target cells many of the toxic effects of traditional chemotherapy. This is expected to translate into a more targeted, effective therapy and a better side effect profile for ADC products. The linkers used in ADC include an acid-labile hydrazone and protease-sensitive dipeptides such as valine-citrulline peptides. These have been shown in preclinical models to be much more stable in blood than conventional means of attaching drugs to mAbs. In vitro and in vivo testing has shown good stability of peptide-linked mAb conjugates with monomethylauristatin E, and good in vitro specificity and lower in vivo toxicity than corresponding hydrazone conjugates. These led to uniformly substituted conjugates that efficiently released active drug in the antigen-positive tumor cells. In vivo studies demonstrated that the peptide-linked conjugates induced regressions of established tumor xenografts with therapeutic indices (ratio of maximum tolerated dose to minimum curative dose) as high as 60-fold, indicating high tolerability. ADEPT on the other hand uses antibodies with tumor specificity that do not internalize but instead remain at the cell surface. These mAbs can be genetically engineered as fusion proteins with enzymes that then cleave nontoxic prodrugs and release chemotherapeutically active drugs at the tumor site. The ADEPT technology is a two-step process that combines two relatively nontoxic agents to achieve potent antitumor activity specifically within tumor tissue. In the first step, a tumor-targeting mAb is used to target catalytic enzymes to the surface of cancer cells. In the second step, inactive forms of anticancer drugs (prodrugs) are administered and converted to their fully active cytotoxic form by the enzyme that has been localized to the tumor microenvironment. This approach results in enzymatic conversion of prodrug to drug specifically in tumor tissue, thus reducing exposure of normal tissue to the drug while maximizing concentrations in tumor tissue [36].

There are currently four marketed conjugated antibody products, where the antibody acts as a carrier agent for the drug: Mylotarg (gemtuzumab ozogamicin), Biogen Idec/Schering AG's Zevalin (ibritumomab tiuxetan), Cotara and Bexxar (tositumomab-iodine 131), and a number of other such products in development. However, these marketed products have much lower sales compared with other antibody drugs. It is thought due to their limited efficacy and the complexity in administering such drugs, which require a radiologist and oncologist to oversee the administration, making them expensive and time-consuming treatment regimens [2]. Nonetheless, using monoclonal antibodies as carrier agents to deliver drugs to a specific site, resulting in a more direct action of the drug, has considerable therapeutic potential. This strategy can circumvent problems in treating complex diseases, such as cancer. Therapeutically effective conjugated antibodies should be eventually commercially viable and would have high potential for life-threatening conditions where there is high unmet need, such as cancer, despite their complex and expensive administration.

### 4.2 Fc Engineering

Scientists at Kyowa Hakko Kogyo Co. Ltd. (KHK) discovered that when an antibody had a reduced amount of fucose in its sugar chains, it exhibited much higher antigen-dependent cellular cytotoxicity (ADCC) activ-

ity as compared to a highly fucosylated conventional antibody. The elimination of fucose from sugar chains on an antibody has demonstrated enhanced therapeutic effects of monoclonal antibodies for cancer [37]. The ability to create fucose-free monoclonal antibodies allows enhancement of ADCC, the critical factor in anti-tumor activity. It was demonstrated that the mechanism behind the enhanced ADCC of a low/no-fucose antibody was its increased affinity to FcγRIIIa (CD16), the major Fc receptor for ADCC in humans. To take advantage of the above finding, scientists at Biowa Inc. knocked out a gene called *FUT*8 that is responsible for the fucose addition to sugar chains in order to create a new production method for fucose-free antibodies in Chinese hamster ovary (CHO) cells. The *FUT*8 knock-out CHO cells (Potelligent cells) produces fucose-free, thereby ADCC-enhanced antibodies (Potelligent maBs), with unchanged basic properties (stability, growth rate, productivity, and scalability) compared to parental CHO cells.

### 4.3 Bi-specific Antibodies

Two monoclonal antibodies can also be combined to create a bi-specific monoclonal antibody. These molecules are still at an early stage of development but have potential to harness the human immune system more effectively than single monoclonal antibodies because its mechanism of action is more alike to the natural immune system, which relies on polyclonal antibodies to generate an effective response. As a result, bi-specific molecules may be effective to treat more complex diseases [38, 39]. In a report Jimenez et al. constructed and produced a bispecific antibody, a diabody, using the variable domain genes of two neutralizing antibodies, *IMC-1121* to VEGFR2 and *hF4-3C5* to VEGFR3. The diabody binds to both VEGFR2 and VEGFR3 in a dose-dependent manner, and blocks interaction between VEGF/VEGFR2, VEGF-C/VEGFR2, and VEGF-C/VEGFR3. In cell-based assays, the diabody neutralized both VEGF and VEGF-C-stimulated activation of VEGFR2, VEGFR3, and p44/p42 mitogen-activated protein kinase in endothelial cells. Furthermore, the diabody was able to inhibit both vascular endothelial growth factor (VEGF) and VEGF-C-induced migration of endothelial cells. Taken together, these results suggest that a dual blockade of both VEGFR2 and VEGFR3 simultaneously may represent a more potent approach to effective cancer therapy [40].

### 4.4 Antibody Fragments

Antibody fragment is another key emerging area for mAbs therapeutic development. Fragments are small and, therefore, have a number of potential advantages over whole antibodies. The small size allows more extensive penetration of tissues such as solid tumors or eyes. Other advantages of antibody fragments are their lower cost of development and production often utilizing microbial expression systems. In addition it is possible to adjust the drug's half-life within the body.

The antigen-binding region of an antibody can be separated from the constant (Fc) portion to create a much smaller molecule (antigen-binding fragment or Fab) that contains the entire antigen-binding site and yet still binds specifically to its target. Removal of the Fc portion of an antibody may also decrease nonspecific binding between the Fc portion and Fc receptor bearing cells, which may increase specificity. Eliminating the Fc region can also lower immunogenicity, while maintaining specific antigenic properties. However, while this kind of microbial antibody fragment manufacturing system has some advantages, not all antibodies may be produced in this way. The fragments lack an Fc-binding region, which is responsible for certain functions such as ADCC. This makes fragments inappropriate for many antibody drug mechanisms, such as that used by Rituxan and Herceptin.

While fragments would generally be cleared more rapidly from the body than full antibodies, the half-life can be extended by modifications such as PEGylation. PEGylation is the process of attaching one or more chains of PEG to a protein molecule, which slow down the rapid clearance of the drug by the body. Current preclinical research is using multi-PEG technology, which allows numerous PEG molecules to be attached to Fab fragments, with the advantage of reduction in manufacturing costs through less wastage of PEG and a greater variability in half-life dependant on the size of the PEG fragments attached. The system for manufacturing Fab fragments in microbial systems allows fragments to be produced on a large scale at a lower cost.

Single-chain variable fragments are another category of antibody fragments that consist only of the variable heavy (VH) and variable light (VL) regions of a full antibody, connected by a synthetic peptide linker. Fab fragments differ from scFvs by containing not only the variable domains but also the constant regions of the light chain (Cl) and part of the heavy chain (Ch1), linked by a disulfide bridge. scFvs have similar binding properties but a smaller molecular size. Enzon Pharmaceuticals has developed the single-chain antibody (SCA) technology to develop drug candidates both independently and with partners. The most advanced SCA-based compound is pexelizumab, being developed by Enzon's licensee Alexion together with Procter & Gamble. Fabs are approximately one-third of the size of whole antibodies, while single-chain antibodies are about one-sixth of the size [41].

Domain Antibodies technology platform can be used to generate small fragments of antibodies, which correspond to the variable regions of either the (VH) or VL chains. The antibody fragments generated from this platform have a molecular weight of approximately 13 kDa, compared to a full antibody, which has a molecular weight of around 150 kDa. Domantis, acquired by GlaxoSmithKline, has created libraries that contain around 10 billion variants of fully human heavy and light chains and uses these libraries to select high-affinity antibody fragments for a given antigen [41]. Domain Antibodies technology has several key features compared with conventional antibodies [42]: The antibody fragments are well expressed in bacterial, yeast, and mammalian cells, which provides greater flexibility in manufacturing, including lower cost methods such as yeast and bacterial cells; the libraries are highly stable, and the antibody fragments can therefore withstand harsh treatment during testing and manufacturing; as the antibody fragments have low molecular weights compared to conventional antibodies, they can be manufactured at higher molar quantities per gram. Therefore, lower doses should be required compared with conventional antibodies, with lower costs per dose; domain antibodies can be delivered by injection and noninjectable routes, such as oral administration because the molecules are highly soluble and resistant to proteolysis during transit through the digestive system. In addition, the Domain Antibodies technology has the flexibility that two fragments can be combined to form a dimer, which can increase the specificity of the molecule binding to a specific antigen.

The commercial potential of antibody fragments is substantial, particularly in niche therapeutic areas that have not traditionally been targeted by antibody therapy. Genentech/Novartis's Lucentis (ranibizumab) for macular degeneration was put on the market recently. Two Fabs are in phase III development: CDP 870 (certolizumab pegol), UCB's antibody fragment for the treatment of rheumatoid arthritis; and Alexion/Procter & Gamble's pexelizumab for cardiopulmonary bypass graft surgery.

### 4.5 Fc-Based Protein Traps

Cytokines are among the best validated of the therapeutic targets. Cytokines can be blocked by the use of soluble receptors. Ebrel/etanercept is a soluable receptor that blocks TNF-α and is used for treating rheumatoid arthritis, ankylosing spondylitis, and psoriatic arthritis. TNF-α promotes inflammation and its associated fever and signs (pain, tenderness, and swelling) in several inflammatory conditions, including rheumatoid arthritis and ankylosing spondylitis. Etanercept reduces the signs and symptoms of rheumatoid arthritis, the arthritis of psoriasis, and ankylosing spondylitis. It prevents the progressive destruction of the joints in patients with rheumatoid arthritis and the arthritis of psoriasis. A method called "cytokine traps" consisting of fusions between the constant region of IgG and the extracellular domains of two distinct cytokine receptors was developed by Economides et al. [43]. Traps potently block cytokines in vitro and in vivo and represent a substantial advance in creating novel therapeutic candidates for cytokine-driven diseases.

## 5 ANTIBODIES IN CLINICS

Monoclonal antibodies as therapies are better suited to certain therapy areas, such as immunodisorders and oncology, where they mimic the natural function of antibodies in the human immune system. The therapeutic application of monoclonal antibodies are mostly used the specificity of these molecules to block receptor–ligand interactions on cell surfaces or binding to receptors and modulating downstream biochemical pathways. However, the applicability of the antibody technologies is largely limited by the small number of disease targets where an extensive knowledge of extracellular receptor–ligand interaction is known to be involved in the pathogenicity of the disease. Despite the efforts in human genomics and protoemics, where numerous disease-related proteins are identified, most of the antibody targets are still limited to cytokines and growth factors and their respective receptors. For central nervous system (CNS) diseases, the concern is that antibodies cannot pass the blood–brain barrier. Nonetheless, respiratory and CNS are emerging as new therapy areas in the monoclonal antibody therapeutic areas. Genentech/Novartis's Xolair (omalizumab) was launched in 2003 by Genentech for the treatment of asthma. In addition, Biogen Idec and Elan launched Tysabri(natalizumab) in 2004 for multiple sclerosis. There are many diseases, beyond oncology and immunodisorders, where therapeutic antibodies can used to treat diseases. The high cost of therapeutic antibodies, however, restricts the commercial viability in those areas where small-molecule drugs are available to sufficiently treat the disease. Therefore, developing monoclonal antibodies targeting those diseases with high unmet need, where there is and will be a lack of competition from small molecules, have greater commercial viability.

Oncology and immunodisorders therapy areas are the therapeutic focus of the monoclonal antibody drug development industry because these molecules are well suited to treat such diseases. The oncology therapy area

is expected to grow significantly in the next 5–10 years. Much of the growth is forecast to be driven by the four best-selling oncology monoclonal antibodies in 2006, Rituxan, Herceptin, Avastin, and Erbitux. There are also a number of antibodies in development for oncology indications. The immunodisorders therapy is forecast to grow strongly over the next 5–10 years. The key growth drivers are expected to be three key rheumatoid arthritis and Crohn's disease antibody treatments: Remicade, Humira, and CDP 870. There is considerable room for further expansion of the immunodisoders product range beyond these markets and greater value can be extracted from monoclonal therapies. The third major focus is infectious disease, mainly attributed by Synagis and its follow-up product, Numax, pending FDA approval. Antibodies have the potential to treat many other infectious diseases, especially other viral infections, and thus there is considerable effort in R&D. Although a number of companies have begun developing antibodies for other viral infections, many of these are at early stages.

The monoclonal antibody is also encompassing a number of other therapy areas, including respiratory and ophthalmology, although at early stages. The therapeutic potential of antibodies outside of the oncology, immunodisorders, and infectious disease therapy areas, is limited due to their extracellular mechanism of action. Nonetheless, antibodies have the potential to treat some diseases that required intracellular activity, as by binding to external cell antigens, antibodies can interfere with downstream intracellular signaling pathways and thus convey a therapeutic benefit. However, extensive knowledge of a disease mechanism involving signaling pathways will be required to achieve this, which unfortunately still remains elusive for many diseases.

Since the mid-1990s, antibodies have emerged as an important new class of drug for therapeutic use across diverse clinical settings, including oncology, chronic inflammatory diseases, transplantation, infectious diseases, and cardiovascular medicine. The Food and Drug Administration (FDA) approved 23 antibody therapeutics, including 16 unmodified IgG molecules, 2 radioimmunoconjugates, 1 antibody–drug conjugate, 2 Fabs, and 2 Fc-fusion proteins, which comprise four different types: 3 murine, 5 chimeric, 11 humanized, and 2 human. At least 150 additional antibodies are in various stages of clinical development. One of the strengths of antibody therapeutics is that they belong to a well-established drug class that has a high success rate from the first use in humans to regulatory approval, the success rate amounting to 29% for chimeric antibodies and 25% for humanized antibodies. This compares favorably with the 11% success rate for small-molecule drugs. The global sale of monoclonal antibodies in 2006 were $21 billion. Remicade had annual sales $3.6 billion in 2006. Rituxan with 2006 sales of $4.7 billion was the best selling monoclonal antibody and biological product and the sixth among the top selling medicinal brands. Seven monoclonal antibody brands in 2006 had sales exceeding $1 billion. Humanized or fully human monoclonal antibodies with low immunogenicity, enhanced antigen binding, and reduced cellular toxicity provide better clinical efficacy. The higher technical and clinical success rate, overcoming of technical hurdles in large-scale manufacturing, low cost of market entry and investigational new drug (IND) filing, and use of fully human and humanized monoclonal antibodies have attracted funds and resources toward R&D. The antibody bandwagon has been joined by 200 companies with hundreds of new projects and targets and has attracted billions of dollars in R&D investment, acquisitions, and licensing deals leading to the current Monoclonal Antibody Gold Rush.

Tables 1 list all U.S. FDA-approved antibody drugs to date. Included in the table is the names of targets and antibodies, the names of products and companies, the source of the products, the approval date, and indications of the drugs. We now provide the detail description of each antibody drug.

**Orthoclone OKT3 (muromonab-CD3)** is a murine monoclonal antibody to the cell differentiation cluster 3 (CD3) antigen of human T cells that functions as an immunosuppressant [44]. The antibody is a biochemically purified IgG2a immunoglobulin with a heavy chain of approximately 50,000 Da and a light chain of approximately 25,000 Da. The proper name, muromonab-CD3, is derived from the descriptive term murine monoclonal antibody. The CD3 designation identifies the specificity of the antibody as the cell differentiation (CD) cluster 3 defined by the First International Workshop on Human Leukocyte Differentiation Antigens.

Orthoclone OKT3 reverses graft rejection, probably by blocking the function of T cells, which play a major role in acute allograft rejection. Orthoclone OKT3 reacts with and blocks the function of a 20,000-Da (CD3) in the membrane of human T cells that has been associated in vitro with the antigen recognition structure of T cells and is essential for signal transduction. In in vitro cytolytic assays, Orthoclone OKT3 blocks both the generation and function of effector cells. Binding of Orthoclone OKT3 to T lymphocytes results in early activation of T cells, which leads to cytokine release, followed by blocking T-cell functions. After termination of Orthoclone OKT3 therapy, T-cell function usually returns to normal within one week. In vivo, Orthoclone OKT3 reacts with most peripheral blood T cells and T cells in body tissues but has not been found to react with other hematopoietic elements or other tissues of the

body. A rapid and concomitant decrease in the number of circulating.

CD3 positive cells, including those that are CD2, CD4, or CD8 positive, has been observed in patients studied within minutes after the administration of Orthoclone OKT3. T-cell activation results in the release of numerous cytokines/lymphokines, which are felt to be responsible for many of the acute clinical manifestations seen following Orthoclone OKT3 administration. Orthoclone OKT3 is indicated for the treatment of acute allograft rejection in renal transplant patients. Orthoclone OKT3 is indicated for the treatment of steroid-resistant acute allograft rejection in cardiac and hepatic transplant patients.Orthoclone OKT3 should not be given to patients who are hypersensitive to products of murine origin, have uncontrolled hypertension, and have a history of seizures, or are predisposed to seizures.

**Abciximab, ReoPro**, is the Fab fragment of the chimeric human-murine monoclonal antibody 7E3 [45]. The chimeric 7E3 antibody is produced by continuous perfusion in mammalian cell culture. The 47,615-Da Fab fragment is purified from cell culture supernatant by a series of steps involving specific viral inactivation and removal procedures, digestion with papain, and column chromatography. Abciximab binds to the intact platelet GPIIb/IIIa receptor, which is a member of the integrin family of adhesion receptors and the major platelet surface receptor involved in platelet aggregation. Abciximab inhibits platelet aggregation by preventing the binding of fibrinogen, von Willebrand factor, and other adhesive molecules to GPIIb/IIIa receptor sites on activated platelets. The mechanism of action is thought to involve steric hindrance and/or conformational effects to block access of large molecules to the receptor rather than direct interaction with the RGD (arginine–glycine–aspartic acid) binding site of GPIIb/IIIa. Abciximab binds with similar affinity to the vitronectin receptor, also known as the avb3 integrin. The vitronectin receptor mediates the procoagulant properties of platelets and the proliferative properties of vascular endothelial and smooth muscle cells. Abciximab also binds to the activated Mac-1 receptor on monocytes and neutrophils. Abciximab is indicated as an adjunct to percutaneous coronary intervention for the prevention of cardiac ischemic complications in patients undergoing percutaneous coronary intervention and in patients with unstable angina not responding to conventional medical therapy when percutaneous coronary intervention is planned within 24 h. Safety and efficacy of abciximab use in patients not undergoing percutaneous coronary intervention have not been established. Abciximab is intended for use with aspirin and heparin and has been studied only in that setting. Because abciximab may increase the risk of bleeding, Abciximab is contraindicated in the following clinical situations: active internal bleeding; recent (within 6 weeks) gastrointestinal (GI) or genitourinary (GU) bleeding of clinical significance; history of cerebrovascular accident (CVA) within 2 years, or CVA with a significant residual neurological deficit; bleeding diathesis; administration of oral anticoagulants within 7 days unless prothrombin time is 1.2 times control; thrombocytopenia (<100,000 cells/μL); recent (within 6 weeks) major surgery or trauma; intracranial neoplasm, arteriovenous malformation, or aneurysm; severe uncontrolled hypertension; presumed or documented history of vasculitis; and use of intravenous dextran before percutaneous coronary intervention (PCI), or intent to use it during an intervention. Abciximab is also contraindicated in patients with known hypersensitivity to any component of this product or to murine proteins.

**Rituxan (rituximab)** is a genetically engineered chimeric murine/human monoclonal antibody directed against the CD20 antigen found on the surface of normal and malignant B lymphocytes [46]. The antibody is an IgG1κ immunoglobulin containing murine light- and heavy-chain variable region sequences and human constant region sequences. Rituximab is composed of two heavy chains of 451 amino acids and two light chains of 213 amino acids and has an approximate molecular weight of 145 kDa. Rituximab has a binding affinity for the CD20 antigen of approximately 8.0 nM. The chimeric anti-CD20 antibody is produced by mammalian cell (CHO) suspension culture. The anti-CD20 antibody is purified by affinity and ion exchange chromatography. The purification process includes specific viral inactivation and removal procedures. Rituximab binds specifically to the antigen CD20 (human B-lymphocyte-restricted differentiation antigen, Bp35), a hydrophobic transmembrane protein with a molecular weight of approximately 35 kDa located on pre-B and mature B lymphocytes. The antigen is also expressed on >90% of B-cell non-Hodgkin's lymphomas (NHL) but is not found on hematopoietic stem cells, pro-B cells, normal plasma cells, or other normal tissues. CD20 regulates an early step(s) in the activation process for cell cycle initiation and differentiation and possibly functions as a calcium ion channel. B cells are believed to play a role in the pathogenesis of rheumatoid arthritis (RA) and associated chronic synovitis. In this setting, B cells may be acting at multiple sites in the autoimmune/inflammatory process, including through production of rheumatoid factor (RF) and other autoantibodies, antigen presentation, T-cell activation, and/or proinflammatory cytokine production. Rituxan (rituximab) is indicated for the treatment of patients with relapsed or refractory, low-grade or follicular, CD20-positive,

B-cell, non-Hodgkin's lymphoma; the first-line treatment of follicular, CD20-positive, B-cell non-Hodgkin's lymphoma in combination with CVP (cyclophosphamide + vincristine + prednisone) chemotherapy; the treatment of low-grade, CD20-positive, B-cell non-Hodgkin's lymphoma in patients with stable disease or who achieve a partial or complete response following first-line treatment with CVP chemotherapy; the first-line treatment of diffuse large B-cell, CD20-positive, non-Hodgkin's lymphoma in combination with CHOP (cyclophosphamide + hydroxydauno [doxorubicin] + Oncovin [vincristine] + prednisone) or other anthracycline-based chemotherapy regimens; in combination with methotrexate to reduce signs and symptoms in adult patients with moderately to severely active rheumatoid arthritis who have had an inadequate response to one or more TNF antagonist therapies. There is no contraindication for Rituxan.

**Zenapax (daclizumab)** is an immunosuppressive, humanized IgG1 monoclonal antibody produced by recombinant DNA technology that binds specifically to the α subunit (p55 α, CD25, or Tac subunit) of the human high-affinity IL-2 receptor that is expressed on the surface of activated lymphocytes [47]. Daclizumab is a composite of human (90%) and murine (10%) antibody sequences. The human sequences were derived from the constant domains of human IgG1 and the variable framework regions of the Eu myeloma antibody. The murine sequences were derived from the complementarity-determining regions of a murine anti-Tac antibody. The molecular weight predicted from the DNA sequence is 144 kDa.

Daclizumab functions as an IL-2 receptor antagonist that binds with high affinity to the Tac subunit of the high-affinity IL-2 receptor complex and inhibits IL-2 binding. Daclizumab binding is highly specific for Tac, which is expressed on activated but not resting lymphocytes. Administration of Zenapax inhibits IL-2-mediated activation of lymphocytes, a critical pathway in the cellular immune response involved in allograft rejection. Zenapax is indicated for the prophylaxis of acute organ rejection in patients receiving renal transplants. It is used as part of an immunosuppressive regimen that includes cyclosporine and corticosteroids. The efficacy of Zenapax for the prophylaxis of acute rejection in recipients of other solid organ allografts has not been demonstrated. Zenapax is contraindicated in patients with known hypersensitivity to daclizumab or to any components of this product.

**Simulect (basiliximab)** is a chimeric (murine–human) monoclonal antibody (IgG1-κ), produced by recombinant DNA technology, that functions as an immunosuppressive agent, specifically binding to and blocking the interleukin-2 receptor α chain (IL-2Rα, also known as CD25 antigen) on the surface of activated T lymphocytes [48]. Based on the amino acid sequence, the calculated molecular weight of the protein is 144 kDa. It is a glycoprotein obtained from fermentation of an established mouse myeloma cell line genetically engineered to express plasmids containing the human heavy- and light-chain constant region genes and mouse heavy- and light-chain variable region genes encoding the RFT5 antibody that binds selectively to the IL-2Rα.

Basiliximab functions as an IL-2 receptor antagonist by binding with high-affinity ($K_a = 1 \times 1010$ M-1) to the α chain of the high-affinity IL-2 receptor complex and inhibiting IL-2 binding. Basiliximab is specifically targeted against IL-2Rα, which is selectively expressed on the surface of activated T lymphocytes. This specific high-affinity binding of Simulect (basiliximab) to IL-2Rα competitively inhibits IL-2-mediated activation of lymphocytes, a critical pathway in the cellular immune response involved in allograft rejection. Simulect (basiliximab) is indicated for the prophylaxis of acute organ rejection in patients receiving renal transplantation when used as part of an immunosuppressive regimen that includes cyclosporine, United States Pharmacopeia (USP) (Modified), and corticosteroids. The efficacy of Simulect for the prophylaxis of acute rejection in recipients of other solid-organ allografts has not been demonstrated. Simulect (basiliximab) is contraindicated in patients with known hypersensitivity to basiliximab or any other component of the formulation.

**Synagis (palivizumab)** is a humanized monoclonal antibody (IgG1 produced by recombinant DNA technology), directed to an epitope in the A antigenic site of the F protein of respiratory syncytial virus (RSV) [49]. Synagis is a composite of human (95%) and murine (5%) antibody sequences. The human heavy-chain sequence was derived from the constant domains of human IgG1 and the variable framework regions of the VH genes *Cor* and *Cess*. The human light-chain sequence was derived from the constant domain of *C*κ and the variable framework regions of the VL gene *K104* with *J*κ-*4*. The murine sequences were derived from a murine monoclonal antibody, Mab 1129, in a process that involved the grafting of the murine complementarity determining regions into the human antibody frameworks. Synagis is composed of two heavy chains and two light chains and has a molecular weight of approximately 148,000 Da.

Synagis exhibits neutralizing and fusion-inhibitory activity against RSV. These activities inhibit RSV replication in laboratory experiments. Although resistant RSV strains may be isolated in laboratory studies, a panel of 57 clinical RSV isolates were all neutralized by

Synagis. The in vivo neutralizing activity of the active ingredient in Synagis was assessed in a randomized, placebo-controlled study of 35 pediatric patients tracheally intubated because of RSV disease. In these patients, Synagis significantly reduced the quantity of RSV in the lower respiratory tract compared to control patients. Synagis is indicated for the prevention of serious lower respiratory tract disease caused by RSV in pediatric patients at high risk of RSV disease. Safety and efficacy were established in infants with bronchopulmonary dysplasia (BPD), infants with a history of premature birth (≤35 weeks gestational age), and children with hemodynamically significant congenital heart disease (CHD). Synagis should not be used in pediatric patients with a history of a severe prior reaction to Synagis or other components of this product.

**Remicade** is a chimeric IgG1-κ monoclonal antibody with an approximate molecular weight of 149,100 Da [50]. It is composed of human constant and murine variable regions. Infliximab binds specifically to human TNF-α with an association constant of 1010 $M^{-1}$. Infliximab is produced by a recombinant cell line cultured by continuous perfusion and is purified by a series of steps that includes measures to inactivate and remove viruses.

Infliximab neutralizes the biological activity of TNF-α by binding with high affinity to the soluble and transmembrane forms of TNF-α and inhibits binding of TNF-α with its receptors. Infliximab does not neutralize TNF-β (lymphotoxin α), a related cytokine that utilizes the same receptors as TNF-α. Biological activities attributed to TNF-α include induction of proinflammatory cytokines such as interleukins (IL) 1 and 6, enhancement of leukocyte migration by increasing endothelial layer permeability and expression of adhesion molecules by endothelial cells and leukocytes, activation of neutrophil and eosinophil functional activity, induction of acute-phase reactants and other liver proteins, as well as tissue-degrading enzymes produced by synoviocytes and/or chondrocytes. Cells expressing transmembrane TNF-α bound by infliximab can be lysed in vitro or in vivo. Infliximab inhibits the functional activity of TNF-α in a wide variety of in vitro bioassays utilizing human fibroblasts, endothelial cells, neutrophils, B and T lymphocytes, and epithelial cells. The relationship of these biological response markers to the mechanism(s) by which Remicade exerts its clinical effects is unknown. Anti-TNF-α antibodies reduce disease activity in the model of collagen-induced arthritis. Infliximab prevents disease in transgenic mice that develop polyarthritis as a result of constitutive expression of human TNF-α, and when administered after disease onset, allows eroded joints to heal.

Remicade, in combination with methotrexate, is indicated for reducing signs and symptoms, inhibiting the progression of structural damage, and improving physical function in patients with moderately to severely active rheumatoid arthritis.

Remicade is indicated for reducing signs and symptoms and inducing and maintaining clinical remission in adult and pediatric patients with moderately to severely active Crohn's disease who have had an inadequate response to conventional therapy.

Remicade is indicated for reducing the number of draining enterocutaneous and rectovaginal fistulas and maintaining fistula closure in adult patients with fistulizing Crohn's disease.

Remicade is indicated for reducing signs and symptoms in patients with active ankylosing spondylitis.

Remicade is indicated for reducing signs and symptoms of active arthritis, inhibiting the progression of structural damage, and improving physical function in patients with psoriatic arthritis.

Remicade is indicated for the treatment of adult patients with chronic severe (i.e., extensive and /or disabling) plaque psoriasis who are candidates for systemic therapy and when other systemic therapies are medically less appropriate. Remicade should only be administered to patients who will be closely monitored and have regular follow-up visits with a physician.

Remicade is indicated for reducing signs and symptoms, inducing and maintaining clinical remission and mucosal healing, and eliminating corticosteroid use in patients with moderately to severely active ulcerative colitis who have had an inadequate response to conventional therapy.

Remicade at doses >5 mg/kg should not be administered to patients with moderate to severe heart failure. In a randomized study evaluating Remicade in patients with moderate to severe heart failure [New York Heart Association (NYHA) Functional Class III/IV], Remicade treatment at 10 mg/kg was associated with an increased incidence of death and hospitalization due to worsening heart failure.

Remicade should not be readministered to patients who have experienced a severe hypersensitivity reaction to Remicade.

Additionally, Remicade should not be administered to patients with known hypersensitivity to inactive components of the product or to any murine proteins.

**Herceptin (trastuzumab)** is a recombinant DNA-derived humanized monoclonal antibody that selectively binds with high affinity in a cell-based assay ($K_d = 5\,nM$) to the extracellular domain of the human epidermal growth factor receptor 2 protein, HER2 [51]. The antibody is an IgG1κ that contains human frame-

work regions with the complementarity-determining regions of a murine antibody (4D5) that binds to HER2. The humanized antibody against HER2 is produced by a mammalian cell (CHO) suspension culture. The HER2 (or c-erbB2) proto-oncogene encodes a transmembrane receptor protein of 185 kDa, which is structurally related to the epidermal growth factor receptor. HER2 protein overexpression is observed in 25–30% of primary breast cancers. HER2 protein overexpression can be determined using immunohistochemistry (IHC). The presence of HER2 overexpression may also be inferred when HER2 gene amplification is identified using fluorescence in situ hybridization (FISH) on fixed tumor blocks. Trastuzumab has been shown, in both in vitro assays and in animals, to inhibit the proliferation of human tumor cells that overexpress HER2. Trastuzumab is a mediator of ADCC. In vitro, Herceptin-mediated ADCC has been shown to be preferentially exerted on HER2 overexpressing cancer cells compared with cancer cells that do not overexpress HER2. Herceptin (trastuzumab), as part of a treatment regimen containing doxorubicin, cyclophosphamide, and paclitaxel, is indicated for the adjuvant treatment of patients with HER2-overexpressing, node-positive breast cancer. Herceptin as a single agent is indicated for the treatment of patients with metastatic breast cancer whose tumors overexpress the HER2 protein and who have received one or more chemotherapy regimens for their metastatic disease. Herceptin in combination with paclitaxel is indicated for treatment of patients with metastatic breast cancer whose tumors overexpress the HER2 protein and who have not received chemotherapy for their metastatic disease. There is no contraindication for Herceptin.

**Enbrel (etanercept)** is a dimeric fusion protein consisting of the extracellular ligand-binding portion of the human 75-kDa (p75) tumor necrosis factor receptor (TNFR) linked to the Fc portion of human IgG1 [52]. The Fc component of etanercept contains the CH2 domain, the CH3 domain and hinge region, but not the CH1 domain of IgG1. Etanercept is produced by recombinant DNA technology in a CHO mammalian cell expression system. It consists of 934 amino acids and has an apparent molecular weight of approximately 150 kDa. Etanercept binds specifically to TNF and blocks its interaction with cell surface TNF receptors. TNF is a naturally occurring cytokine that is involved in normal inflammatory and immune responses. It plays an important role in the inflammatory processes of RA, polyarticular-course juvenile rheumatoid arthritis (JRA), and ankylosing spondylitis (AS) and the resulting joint pathology. In addition, TNF plays a role in the inflammatory process of plaque psoriasis. Elevated levels of TNF are found in involved tissues and fluids of patients with RA, psoriatic arthritis, AS, and plaque psoriasis. Two distinct receptors for TNF (TNFRs), a 55-kDa protein (p55) and a 75-kDa protein (p75), exist naturally as monomeric molecules on cell surfaces and in soluble forms. Biological activity of TNF is dependent upon binding to either cell surface TNFR. Etanercept is a dimeric soluble form of the p75 TNF receptor that can bind to two TNF molecules. It inhibits the activity of TNF in vitro and has been shown to affect several animal models of inflammation, including murine collagen-induced arthritis. Etanercept inhibits binding of both TNFa and TNFb [lymphotoxin alpha [LTa]] to cell surface TNFRs, rendering TNF biologically inactive. Cells expressing transmembrane TNF that bind Enbrel are not lysed in vitro in the presence or absence of complement. Etanercept can also modulate biological responses that are induced or regulated by TNF, including expression of adhesion molecules responsible for leukocyte migration (i.e., E-selectin and to a lesser extent intercellular adhesion molecule-1 [ICAM-1]), serum levels of cytokines (e.g., IL-6), and serum levels of matrix metalloproteinase-3 (MMP-3 or stromelysin). Enbrel is indicated for reducing signs and symptoms, inducing major clinical response, inhibiting the progression of structural damage, and improving physical function in patients with moderately to severely active rheumatoid arthritis. Enbrel can be initiated in combination with methotrexate (MTX) or used alone. Enbrel is indicated for reducing signs and symptoms of moderately to severely active polyarticular-course IRA in patients who have had an inadequate response to one or more disease-modifying antirheumatic drugs (DMARDs). Enbrel is indicated for reducing signs and symptoms, inhibiting the progression of structural damage of active arthritis, and improving physical function in patients with psoriatic arthritis. Enbrel can be used in combination with methotrexate in patients who do not respond adequately to methotrexate alone. Enbrel is indicated for reducing signs and symptoms in patients with active ankylosing spondylitis. Enbrel is indicated for the treatment of adult patients (18 years or older) with chronic moderate to severe plaque psoriasis who are candidates for systemic therapy or phototherapy. Enbrel should not be administered to patients with sepsis or with known hypersensitivity to ENBREL or any of its components.

**Mylotarg (gemtuzumab ozogamicin)** is a chemotherapy agent composed of a recombinant humanized IgG4κ, antibody conjugated with a cytotoxic antitumor antibiotic, calicheamicin, isolated from fermentation of a bacterium, *Micromonospora echinospora* sp. *Calichensis* [53]. The antibody portion of Mylotarg binds

specifically to the CD33 antigen, a sialic-acid-dependent adhesion protein found on the surface of leukemic myeloblasts and immature normal cells of myelomonocytic lineage, but not on normal hematopoetic stem cells. The anti-CD33 hP67.6 antibody is produced by mammalian cell suspension culture using a myeloma NS0 cell line and is purified under conditions that remove or inactivate viruses. Mylotarg contains amino acid sequences of which approximately 98.3% are of human origin. The constant region and framework regions contain human sequences while the CDRs are derived from a murine antibody (p67.6) that binds CD33. This antibody is linked to *N*-acetyl-γ-calicheamicin via a bifunctional linker. Gemtuzumab ozogamicin has approximately 50% of the antibody loaded with 4–6 mol calicheamicin per mole of antibody. The remaining 50% of the antibody is not linked to the calicheamicin derivative. Gemtuzumab ozogamicin has a molecular weight of 151–153,000 Da. Gemtuzumab ozogamicin binds to the CD33 antigen. This antigen is expressed on the surface of leukemic blasts in more than 80% of patients with acute myeloid leukemia (AML). CD33 is also expressed on normal and leukemic myeloid colony-forming cells, including leukemic clonogenic precursors, but it is not expressed on pluripotent hematopoietic stem cells or on nonhematopoietic cells. Binding of the anti-CD33 antibody portion of Mylotarg with the CD33 antigen results in the formation of a complex that is internalized. Upon internalization, the calicheamicin derivative is released inside the lysosomes of the myeloid cell. The released calicheamicin derivative binds to DNA in the minor groove resulting in DNA double-strand breaks and cell death. Gemtuzumab ozogamicin is cytotoxic to the CD33-positive HL-60 human leukemia cell line. Gemtuzumab ozogamicin produces significant inhibition of colony formation in cultures of adult leukemic bone marrow cells. The cytotoxic effect on normal myeloid precursors leads to substantial myelosuppression, but this is reversible because pluripotent hematopoietic stem cells are spared. In preclinical animal studies, gemtuzumab ozogamicin demonstrates antitumor effects in the HL-60 human promyelocytic leukemia xenograft tumor in athymic mice. Mylotarg is indicated for the treatment of patients with CD33-positive AML in first relapse who are 60 years of age or older and who are not considered candidates for cytotoxic chemotherapy. The safety and efficacy of Mylotarg in patients with poor performance status and organ dysfunction has not been established. There are no controlled trials demonstrating a clinical benefit, such as improvement in disease-related symptoms or increased survival, compared to any other treatment. Mylotarg is contraindicated in patients with a known hypersensitivity to gemtuzumab ozogamicin or any of its components: anti-CD33 antibody (hP67.6), calicheamicin derivatives, or inactive ingredients.

**Campath (alemtuzumab)** is a recombinant DNA-derived humanized monoclonal antibody (Campath-1H) that is directed against the 21- to 28-kD cell surface glycoprotein, CD52 [54]. The Campath-1H antibody is an IgG1κ with human variable framework and constant regions, and CDRs from a murine (rat) monoclonal antibody (Campath-1G). The Campath-1H antibody has an approximate molecular weight of 150 kDa. Campath is produced in mammalian cell (CHO) suspension culture. Alemtuzumab binds to CD52, a non-modulating antigen that is present on the surface of essentially all B and T lymphocytes, a majority of monocytes, macrophages, and natural killer (NK) cells, and a subpopulation of granulocytes. Analysis of samples collected from multiple volunteers has not identified CD52 expression on erythrocytes or hematopoetic stem cells. The proposed mechanism of action is anti-body-dependent lysis of leukemic cells following cell surface binding. Campath-1H Fab binding was observed in lymphoid tissues and the mononuclear phagocyte system. A proportion of bone marrow cells, including some CD34+ cells, express variable levels of CD52. Significant binding was also observed in the skin and male reproductive tract (epididymis, sperm, seminal vesicle). Mature spermatozoa stain for CD52, but neither spermatogenic cells nor immature spermatozoa show evidence of staining. Campath is indicated for the treatment of B-cell chronic lymphocytic leukemia (B-CLL) in patients who have been treated with alkylating agents and who have failed fludarabine therapy. Determination of the effectiveness of Campath is based on overall response rates. Comparative, randomized trials demonstrating increased survival or clinical benefits such as improvement in disease-related symptoms have not yet been conducted.

Campath is contraindicated in patients who have active systemic infections, underlying immunodeficiency [e.g., seropositive for (human immunodeficiency virus (HIV)], or known type I hypersensitivity or anaphylactic reactions to Campath or to any one of its components.

**Zevalin (ibritumomab tiuxetan)** is the immunoconjugate resulting from a stable thiourea covalent bond between the monoclonal antibody ibritumomab and the linker-chelator tiuxetan [*N*-[2-bis(carboxymethyl) amino]-3-(p-isothiocyanatophenyl)-propyl]-[*N*-[2-bis(carboxymethyl)amino]-2-(methyl)-ethyl]glycine [54]. This linker-chelator provides a high affinity, conformationally restricted chelation site for indium-111 or yttrium-90. The approximate molecular weight of ibri-

tumomab tiuxetan is 148 kDa. The antibody moiety of Zevalin is ibritumomab, a murine IgG1κ monoclonal antibody directed against the CD20 antigen, which is found on the surface of normal and malignant B lymphocytes. Ibritumomab is produced in CHO cells and is composed of two murine γ 1 heavy chains of 445 amino acids each and two κ light chains of 213 amino acids each.

Ibritumomab tiuxetan binds specifically to the CD20 antigen (human B-lymphocyte-restricted differentiation antigen, Bp35). The apparent affinity ($K_d$) of ibritumomab tiuxetan for the CD20 antigen ranges between approximately 14 on 18 n*M*. The CD20 antigen is expressed on pre-B and mature-B lymphocytes and on >90% of B-cell NHL. The CD20 antigen is not shed from the cell surface and does not internalize upon antibody binding.

*Mechanism of Action*: The complementarity-determining regions of ibritumomab bind to the CD20 antigen on B lymphocytes. Ibritumomab, like rituximab, induces apoptosis in CD20+ B-cell lines in vitro. The chelate tiuxetan, which tightly binds In-111 or Y-90, is covalently linked to the amino groups of exposed lysines and arginines contained within the antibody. The β emission from Y-90 induces cellular damage by the formation of free radicals in the target and neighboring cells. Zevalin, as part of the Zevalin therapeutic regimen, is indicated for the treatment of patients with relapsed or refractory low-grade, follicular, or transformed B-cell non-Hodgkin's lymphoma, including patients with rituximab refractory follicular non-Hodgkin's lymphoma. Determination of the effectiveness of the Zevalin therapeutic regimen in a relapsed or refractory patient population is based on overall response rates. The effects of the Zevalin therapeutic regimen on survival are not known. The Zevalin therapeutic regimen is contraindicated in known type I hypersensitivity or anaphylactic reaction proteins or to any component of this product, including yttrium chloride and indium chloride.

**Humira (adalimumab)** is a recombinant human IgG1 monoclonal antibody specific for human tumor necrosis factor [55]. Humira was created using phage display technology resulting in an antibody with human-derived heavy- and light-chain variable regions and human IgG1κ constant regions. Adalimumab is produced by recombinant DNA technology in a mammalian cell expression system and is purified by a process that includes specific viral inactivation and removal steps. It consists of 1330 amino acids and has a molecular weight of approximately 148 kDa. Adalimumab binds specifically to TNF-α and blocks its interaction with the p55 and p75 cell surface TNF receptors. Adalimumab also lyses surface TNF expressing cells in vitro in the presence of complement. Adalimumab does not bind or inactivate lymphotoxin (TNF-β). TNF is a naturally occurring cytokine that is involved in normal inflammatory and immune responses. Elevated levels of TNF are found in the synovial fluid of rheumatoid arthritis, psoriatic arthritis, and ankylosing spondylitis patients and play an important role in both the pathologic inflammation and the joint destruction that are hallmarks of these diseases. Adalimumab also modulates biological responses that are induced or regulated by TNF, including changes in the levels of adhesion molecules responsible for leukocyte migration (ELAM-1 [endothelial-leukocyte adhesion molecule 1], VCAM-1 [vascular-cell adhesion molecule 1], and ICAM-1 [intercellular-adhesion molecule 1] with an $IC_{50}$ of $1–2 \times 10^{-10} M$). Humira is a TNF blocker indicated for treatment of RA: Reducing signs and symptoms, inducing major clinical response, inhibiting the progression of structural damage, and improving physical function in adult patients with moderately to severely active disease. For the treatment of psoriatic arthritis: Reducing signs and symptoms of active arthritis, inhibiting the progression of structural damage, and improving physical function. For the treatment of ankylosing spondylitis: reducing signs and symptoms in patients with active disease. For treatment of Crohn's disease: reducing signs and symptoms and inducing and maintaining clinical remission in adult patients with moderately to severely active Crohn's disease who have had an inadequate response to conventional therapy. Reducing signs and symptoms and inducing clinical remission in these patients if they have also lost response to or are intolerant of infliximab. There are no contraindications for Humira.

**Xolair (omalizumab)** is a recombinant DNA-derived humanized IgG1κ monoclonal antibody that selectively binds to human immunoglobulin E (IgE). The antibody has a molecular weight of approximately 149,000 Da. Xolair is produced by a CHO cell suspension culture [56]. Xolair inhibits the binding of IgE to the high-affinity IgE receptor (FcεRI) on the surface of mast cells and basophils. Reduction in surface-bound IgE on FcεRI-bearing cells limits the degree of release of mediators of the allergic response. Treatment with Xolair also reduces the number of FcεRI receptors on basophils in atopic patients. A variety of studies have found that omalizumab decreases inflammatory mediators. The result of blocking the IgE high-affinity Fc receptor is manifolds. Omalizumab treatment downregulates dendritic cell FcεRI expression. The combined effect of specific immunotherapy (SIT) and omalizumab is associated with prevention of nasal eosinophilic cationic protein increase and decreased tryptase levels in nasal

secretions. Anti-IgE therapy reduces leukotriene release of peripheral leukocytes stimulated with allergen in children with allergic rhinitis undergoing allergen immunotherapy independent of the type of SIT allergen used. Xolair (omalizumab) is indicated for adults and adolescents (12 years of age and above) with moderate to severe persistent asthma who have a positive skin test or in vitro reactivity to a perennial aeroallergen and whose symptoms are inadequately controlled with inhaled corticosteroids. Xolair has been shown to decrease the incidence of asthma exacerbations in these patients. Safety and efficacy have not been established in other allergic conditions. Xolair should not be administered to patients who have experienced a severe hypersensitivity reaction to Xolair.

**Bexxar (tositumomab and iodine $^{131}$I tositumomab)** is an antineoplastic radioimmunotherapeutic mAb-based regimen composed of the monoclonal antibody tositumomab and the radiolabeled monoclonal antibody iodine $^{131}$I tositumomab [57, 58]. Tositumomab is a murine IgG2aλ monoclonal antibody directed against the CD20 antigen, which is found on the surface of normal and malignant B lymphocytes. Tositumomab is produced in an antibiotic-free culture of mammalian cells and is composed of two murine γ2a heavy chains of 451 amino acids each and two λ light chains of 220 amino acids each. The approximate molecular weight of tositumomab is 150,000 Da. Iodine $^{131}$I tositumomab is a radio-iodinated derivative of tositumomab that has been covalently linked to $^{131}$iodine. Tositumomab binds specifically to the CD20 (human B-lymphocyte-restricted differentiation antigen, Bp 35 or B1) antigen. This antigen is a transmembrane phosphoprotein expressed on pre-B lymphocytes and at higher density on mature-B lymphocytes. The antigen is also expressed on >90% of B-cell NHL. The recognition epitope for tositumomab is found within the extracellular domain of the CD20 antigen. CD20 does not shed from the cell surface and does not internalize following antibody binding. Possible mechanisms of action of the Bexxar therapeutic regimen include induction of apoptosis, complement-dependent cytotoxicity (CDC) and ADCC mediated by the antibody. Additionally, cell death is associated with ionizing radiation from the radioisotope. The Bexxar therapeutic regimen (tositumomab and iodine $^{131}$I tositumomab) is indicated for the treatment of patients with CD20 antigen-expressing relapsed or refractory, low-grade, follicular, or transformed NHL, including patients with rituximab-refractory NHL. Determination of the effectiveness of the Bexxar therapeutic regimen is based on overall response rates in patients whose disease is refractory to chemotherapy alone or to chemotherapy and rituximab. The effects of the Bexxar therapeutic regimen on survival are not known. The Bexxar therapeutic regimen is not indicated for the initial treatment of patients with CD20-positive NHL. The Bexxar therapeutic regimen is intended as a single course of treatment. The safety of multiple courses of the Bexxar therapeutic regimen, or combination of this regimen with other forms of irradiation or chemotherapy, has not been evaluated. The Bexxar therapeutic regimen is contraindicated in patients with known hypersensitivity to murine proteins or any other component of the Bexxar therapeutic regimen.

**Raptiva (efalizumab)** is an immunosuppressive recombinant humanized IgG1κ isotype monoclonal antibody that binds to human CD11a [59]. Efalizumab has a molecular weight of approximately 150,000 Da and is produced in a CHO mammalian cell expression system. Raptiva binds to CD11a, a subunit of leukocyte function antigen-1 (LFA-1), which is expressed on all leukocytes, and decreases cell surface expression of CD11a. Raptiva inhibits the binding of LFA-1 to intercellular adhesion molecule-1 (ICAM-1), thereby inhibiting the adhesion of leukocytes to other cell types. Interaction between LFA-1 and ICAM-1 contributes to the initiation and maintenance of multiple processes, including activation of T lymphocytes, adhesion of T lymphocytes to endothelial cells, and migration of T lymphocytes to sites of inflammation including psoriatic skin. Lymphocyte activation and trafficking to skin play a role in the pathophysiology of chronic plaque psoriasis. In psoriatic skin, ICAM-1 cell surface expression is upregulated on endothelium and keratinocytes. CD11a is also expressed on the surface of B lymphocytes, monocytes, neutrophils, NK cells, and other leukocytes. Therefore, the potential exists for Raptiva to affect the activation, adhesion, migration, and numbers of cells other than T lymphocytes. Raptiva (efalizumab) is indicated for the treatment of adult patients (18 years or older) with chronic moderate to severe plaque psoriasis who are candidates for systemic therapy or phototherapy. Raptiva should not be administered to patients with known hypersensitivity to Raptiva or any of its components.

**Amevive (alefacept)** is an immunosuppressive dimeric fusion protein that consists of the extracellular CD2-binding portion of the human LFA-3 linked to the Fc (hinge, CH2, and CH3 domains) portion of human IgG1 [60]. Alefacept is produced by recombinant DNA technology in a CHO mammalian cell expression system. The molecular weight of alefacept is 91,400 Da. Amevive interferes with lymphocyte activation by specifically binding to the lymphocyte antigen, CD2, and inhibiting LFA-3/CD2 interaction. Activation of T lym-

phocytes involving the interaction between LFA-3 on antigen-presenting cells and CD2 on T lymphocytes plays a role in the pathophysiology of chronic plaque psoriasis. The majority of T lymphocytes in psoriatic lesions are of the memory effector phenotype characterized by the presence of the CD45RO marker1, express activation markers (e.g., CD25, CD69) and release inflammatory cytokines, such as interferon γ.

Amevive also causes a reduction in subsets of CD2+ T lymphocytes (primarily CD45RO+), presumably by bridging between CD2 on target lymphocytes and immunoglobulin Fc receptors on cytotoxic cells, such as NK cells. Treatment with Amevive results in a reduction in circulating total CD4+ and CD8+ T-lymphocyte counts. CD2 is also expressed at low levels on the surface of NK cells and certain bone marrow B lymphocytes. Therefore, the potential exists for Amevive to affect the activation and numbers of cells other than T lymphocytes. In clinical studies of Amevive, minor changes in the numbers of circulating cells other than T lymphocytes have been observed. Amevive is indicated for the treatment of adult patients with moderate to severe chronic plaque psoriasis who are candidates for systemic therapy or phototherapy. Amevive should not be administered to patients infected with HIV. Amevive reduces CD4+ T-lymphocyte counts, which might accelerate disease progression or increase complications of disease in these patients. Amevive should not be administered to patients with known hypersensitivity to Amevive or any of its components.

**Erbitux (cetuximab)** is a recombinant, human/mouse chimeric monoclonal antibody that binds specifically to the extracellular domain of the human epidermal growth factor receptor (EGFR) [61]. Cetuximab is composed of the Fv regions of a murine anti-EGFR antibody with human IgG1 heavy- and κ light-chain constant regions and has an approximate molecular weight of 152 kDa. Cetuximab is produced in mammalian (murine myeloma) cell culture. The epidermal growth factor receptor (EGFR, HER1, c-ErbB-1) is a transmembrane glycoprotein that is a member of a subfamily of type I receptor tyrosine kinases including EGFR, HER2, HER3, and HER4. The EGFR is constitutively expressed in many normal epithelial tissues, including the skin and hair follicle. Expression of EGFR is also detected in many human cancers, including those of the head and neck, colon, and rectum. Cetuximab binds specifically to the EGFR on both normal and tumor cells, and competitively inhibits the binding of epidermal growth factor (EGF) and other ligands, such as transforming growth factor-α. In vitro assays and in vivo animal studies have shown that binding of cetuximab to the EGFR blocks phosphorylation and activation of receptor-associated kinases, resulting in inhibition of cell growth, induction of apoptosis, and decreased matrix metalloproteinase and vascular endothelial growth factor production. In vitro, cetuximab can mediate ADCC against certain human tumor types. While the mechanism of cetuximab's antitumor effect(s) in vivo is unknown, all of these processes may contribute to the overall therapeutic effect of cetuximab. In vitro assays and in vivo animal studies have shown that cetuximab inhibits the growth and survival of tumor cells that express the EGFR. No antitumor effects of cetuximab were observed in human tumor xenografts lacking EGFR expression. The addition of cetuximab to radiation therapy, irinotecan, or irinotecan plus 5-fluorouracil in human tumor xenograft models in mice resulted in an increase in antitumor effects compared to radiation therapy or chemotherapy alone. Erbitux (cetuximab), in combination with radiation therapy, is indicated for the treatment of locally or regionally advanced squamous cell carcinoma of the head and neck. Erbitux as a single agent is indicated for the treatment of patients with recurrent or metastatic squamous cell carcinoma of the head and neck for whom prior platinum-based therapy has failed.

Erbitux, used in combination with irinotecan, is indicated for the treatment of EGFR-expressing, metastatic colorectal carcinoma in patients who are refractory to irinotecan-based chemotherapy. Erbitux administered as a single agent is indicated for the treatment of EGFR-expressing, metastatic colorectal carcinoma in patients who are intolerant of irinotecan-based chemotherapy. The effectiveness of Erbitux for the treatment of EGFR-expressing, metastatic colorectal carcinoma is based on objective response rates. Currently, no data are available that demonstrate an improvement in disease-related symptoms or increased survival with Erbitux for the treatment of EGFR-expressing, metastatic colorectal carcinoma. There are no contraindications for Erbitux.

**Avastin (bevacizumab)** is a recombinant humanized monoclonal IgG1 antibody that binds to and inhibits the biologic activity of human VEGF in in vitro and in vivo assay systems [62]. Bevacizumab contains human framework regions and the complementarity-determining regions of a murine antibody that binds to VEGF. Bevacizumab is produced in a CHO mammalian cell expression system and has a molecular weight of approximately 149,000 Da. Bevacizumab binds VEGF and prevents the interaction of VEGF to its receptors (Flt-1 and KDR) on the surface of endothelial cells. The interaction of VEGF with its receptors leads to endothelial cell proliferation and new blood vessel formation in in vitro models of angiogenesis. Administration of

bevacizumab to xenotransplant models of colon cancer in nude (athymic) mice caused reduction of microvascular growth and inhibition of metastatic disease progression. Avastin, in combination with intravenous 5-fluorouracil-based chemotherapy, is indicated for first- or second-line treatment of patients with metastatic carcinoma of the colon or rectum. Avastin, in combination with carboplatin and paclitaxel, is indicated for first-line treatment of patients with unresectable, locally advanced, recurrent or metastatic nonsquamous, non-small-cell lung cancer. There are no contraindications for Avastin.

**Tysabri (natalizumab)** is a recombinant humanized IgG4κ monoclonal antibody produced in murine myeloma cells [63]. Natalizumab contains human framework regions and the complementarity-determining regions of a murine antibody that binds to α4-integrin. The molecular weight of natalizumab is 149,000 Da. Tysabri binds to the α4 subunit of α4β1 and α4β7 integrins expressed on the surface of all leukocytes except neutrophils, and inhibits the α4-mediated adhesion of leukocytes to their counterreceptor(s). The receptors for the α4 family of integrins include vascular cell adhesion molecule-1 (VCAM-1), which is expressed on activated vascular endothelium, and mucosal address in cell adhesion molecule-1 (MadCAM-1) present on vascular endothelial cells of the gastrointestinal tract. Disruption of these molecular interactions prevents transmigration of leukocytes across the endothelium into inflamed parenchymal tissue. In vitro, anti-α4-integrin antibodies also block α4-mediated cell binding to ligands such as osteopontin and an alternatively spliced domain of fibronectin, connecting segment-1 (CS-1). In vivo, Tysabri may further act to inhibit the interaction of α4-expressing leukocytes with their ligand(s) in the extracellular matrix and on parenchymal cells, thereby inhibiting further recruitment and inflammatory activity of activated immune cells. The specific mechanism(s) by which Tysabri exerts its effects in multiple sclerosis have not been fully defined. In multiple sclerosis, lesions are believed to occur when activated inflammatory cells, including T lymphocytes, cross the blood–brain barrier (BBB). Leukocyte migration across the BBB involves interaction between adhesion molecules on inflammatory cells and their counterreceptors present on endothelial cells of the vessel wall. The clinical effect of natalizumab in multiple sclerosis may be secondary to blockade of the molecular interaction of α4β1 integrin expressed by inflammatory cells with VCAM-1 on vascular endothelial cells, and with CS-1 and/or osteopontin expressed by parenchymal cells in the brain. Data from an experimental autoimmune encephalitis animal model of multiple sclerosis demonstrate reduction of leukocyte migration into brain parenchyma and reduction of plaque formation detected by magnetic resonance imaging (MRI) following repeated administration of natalizumab. The clinical significance of these animal data is unknown. Tysabri is indicated as monotherapy for the treatment of patients with relapsing forms of multiple sclerosis to delay the accumulation of physical disability and reduce the frequency of clinical exacerbations. The safety and efficacy of Tysabri beyond 2 years are unknown. Because Tysabri increases the risk of progressive multifocal leukoencephalopathy (PML), an opportunistic viral infection of the brain that usually leads to death or severe disability, Tysabri is generally recommended for patients who have had an inadequate response to, or are unable to tolerate, alternate multiple sclerosis therapies. Safety and efficacy in patients with chronic progressive multiple sclerosis have not been studied. Tysabri should not be administered to patients with known hypersensitivity to Tysabri or any of its components. Tysabri is contraindicated in patients who have or have had PML.

**Lucentis (ranibizumab injection)** is a recombinant humanized IgG1κ isotype monoclonal antibody fragment designed for intraocular use [64, 65]. Ranibizumab binds to and inhibits the biologic activity of all VEGF-A isoforms and the biologically active degradation product, VEGF110. Ranibizumab has a molecular weight of approximately 48,000 Da and is produced by an *Escherrchia coli* expression system. The parent molecule of ranibizumab is the murine full-length monoclonal antibody A4.6.1. A first-generation humanized Fab of A4.6.1 was affinity matured using phage display technology to generate the Y0317 variant, also known as rhuFab V2 or ranibizumab, which increased its affinity for VEGF-A 140-fold. Ranibizumab binds to the receptor binding site of active forms of VEGF-A, including the biologically active, cleaved form of this molecule, VEGF110. VEGF-A has been shown to cause neovascularization and leakage in models of ocular angiogenesis and is thought to contribute to the progression of the neovascular form of age-related macular degeneration (AMD). The binding of ranibizumab to VEGF-A prevents the interaction of VEGF-A with its receptors (VEGFR1 and VEGFR2) on the surface of endothelial cells, reducing endothelial cell proliferation, vascular leakage, and new blood vessel formation. Lucentis is indicated for the treatment of patients with neovascular (wet) AHD. Lucentis is contraindicated in patients with ocular or periocular infections.

**Vectibix (panitumumab)** is a recombinant, human IgG2κ monoclonal antibody that binds specifically to the human EGFR [66, 67]. Panitumumab has an approximate molecular weight of 147,000 Da. Panitumumab is produced in genetically engineered mammalian CHO cells. The EGFR is a member of a subfamily of type I receptor tyrosine kinases, including EGFR (HER1, c-ErbB-1), HER2/neu, HER3, and HER4. EGFR is a transmembrane glycoprotein that is constitutively expressed in many normal epithelial tissues, including the skin and hair follicle. Overexpression of EGFR is also detected in many human cancers, including those of the colon and rectum. Interaction of EGFR with its normal ligands (e.g., EGF, transforming growth factor-α) leads to phosphorylation and activation of a series of intracellular tyrosine kinases, which in turn regulate transcription of molecules involved with cellular growth and survival, motility, proliferation, and transformation. Panitumumab binds specifically to EGFR on both normal and tumor cells and competitively inhibits the binding of ligands for EGFR. Nonclinical studies show that binding of panitumumab to the EGFR prevents ligand-induced receptor autophosphorylation and activation of receptor-associated kinases, resulting in inhibition of cell growth, induction of apoptosis, decreased proinflammatory cytokine and vascular growth factor production, and internalization of the EGFR. In vitro assays and in vivo animal studies demonstrate that panitumumab inhibits the growth and survival of selected human tumor cell lines expressing EGFR. Vectibix is indicated for the treatment of EGFR-expressing, metastatic colorectal carcinoma with disease progression on or following fluoropyrimidine-, oxaliplatin-, and irinotecan-containing chemotherapy regimens. The effectiveness of Vectibix for the treatment of EGFR-expressing, metastatic colorectal carcinoma is based on progression-free survival. Currently, no data are available that demonstrate an improvement in disease-related symptoms or increased survival with Vectibix. Panitumumab has been approved in the United States and is currently awaiting registration in Australia, Canada, the European Union, and Switzerland for the treatment of patients with EGFR-expressing metastatic colorectal cancer who have disease progression on or following fluoropyrimidine-, irinotecan- and oxaliplatin-containing chemotherapy regimens. It is under clinical investigation as both monotherapy and in combination with other anticancer agents for the treatment of patients with various types of cancer.

**Soliris (eculizumab)** is a formulation of eculizumab, which is a recombinant humanized monoclonal IgG2/4κ antibody produced by murine myeloma cell culture and purified by standard bioprocess technology [68]. Eculizumab contains human constant regions from human IgG2 sequences and human IgG4 sequences and murine complementarity-determining regions grafted onto the human framework light- and heavy-chain variable regions. Eculizumab is composed of two 448-amino-acid heavy chains and two 214-amino-acid light chains and has a molecular weight of approximately 148,000 Da. Eculizumab, the active ingredient in Soliris, is a monoclonal antibody that specifically binds to the complement protein C5 with high affinity, thereby inhibiting its cleavage to C5a and C5b and preventing the generation of the terminal complement complex C5b-9. Soliris inhibits terminal complement-mediated intravascular hemolysis in paroxysma nocturnal hemoglobinuria (PNH) patients. A genetic mutation in PNH patients leads to the generation of populations of abnormal red blood cells (RBCs) (known as PNH cells) that are deficient in terminal complement inhibitors, rendering PNH RBCs sensitive to persistent terminal complement-mediated destruction. The destruction and loss of these PNH cells (intravascular hemolysis) results in low RBC counts (anemia), and also fatigue, difficulty in functioning, pain, dark urine, shortness of breath, and blood clots. Soliris is indicated for the treatment of patients with PNH to reduce hemolysis. Do not initiate Soliris therapy in patients with unresolved serious *Neisseria meningitidis* infection and patients who are not currently vaccinated against *N. meningitidis*.

## 6 MONOCLONAL ANTIBODIES UNDER CLINICAL DEVELOPMENT

We estimated at least 300 monoclonal antibodies under development for the treatment of various human diseases. The majority of antibodies in development unsurprisingly are in preclinical development. There are many more undisclosed antibodies in preclinical development. Table 2 listed mAbs that are currently in clinical trials phase 1–3. phase 1 and phase 1/2 consists of 61 clinical trials. The number of phase 2 programs is 95 slightly larger that that of phase 1. Table 3 lists the fewest programs are in phase III trials and in registration with many are already approved drugs on the market. The active antibody development programs have a wider therapeutic focus, spanning many therapy areas. However, the main focus still remains on the two therapy areas with the highest sales, oncology and immunodisorders.

**TABLE 2 mAb under Clinical Development**

| Development Stage | Targets | Brand/Code | Generic | Company | Partners | Therapy area | Indication | Antibody class |
|---|---|---|---|---|---|---|---|---|
| Phase 1 | Blys | LymphoStat-B | Anti-BLyS | CAT | HGS | Immuno-Disease | Lupus Erythematosis | Fully human |
| Phase 1 | C6 | Eculizumab | 0 | Alexion | Lonza | Immuno-Disease | Paroxysmal nocturnal hemoglobinuria | Unknown |
| Phase 1 | CanAg | Cantuzumab mertansine, huC242-DM1 | 0 | ImmunoGen | 0 | Oncology | Cancers | Conjugated |
| Phase 1 | CCR2 | MLN1202 | Anti-CCR2 | Millennium | 0 | Immuno disease | Rheumatoid arthritis | Humanized |
| Phase 1 | CD20 | Pretarget lymphoma | Anti-CD20 | NeoRx Corp | 0 | Oncology | Lymphoma B-cell | Conjugated |
| Phase 1 | CD3/EpCAM | Removab | Anti-EpCAM × anti-CD3 | TRION Pharma GmbH | 0 | Oncology | Ovarian carcinoma | Other |
| Phase 1 | CD4 | TNX-355 | Anti-CD4 | Tanox | Biogen | Infectious diseases | HIV | Humanized |
| Phase 1 | CD4 | TRX1 | Anti-CD4 | TolerRx | Genentech | Immuno disease | IDI | Humanized |
| Phase 1 | CD44v6 | Bivatuzumab mertansine | 0 | ImmunoGen | Boehringer Ingelheim | Oncology | Tumors | Conjugated |
| Phase 1 | CD56 | huN901-DM1 | 0 | ImmunoGen | British Biotech | Oncology | Small-cell lung cancer | Conjugated |
| Phase 1 | CTLA-4 | MDX-010 | 0 | Medarex | 0 | Infectious diseases | HIV | Fully Human |
| Phase 1 | CTLA-4 | MDX-010 + melanoma peptides | 0 | Medarex | Biosite | Oncology | Melanoma | Fully Human |
| Phase 1 | CTLA4-Ig | Anti-CTLA4-Ig | 0 | RepliGen | 0 | Immuno disease | Thrombocytopenic purpura | Unknown |
| Phase 1 | EGFR | ABX-EGF | Anti-EGFR | Amgen | Amgen | Oncology | Various tumors | Fully Human |
| Phase 1 | Ep-CAM | ING-1 | heMAb | XOMA | 0 | Oncology | Adenocarcinoma | Humanized |
| Phase 1 | Her2 | rexomab | Anti-Her2/neu | TRION Pharma GmbH | 0 | Oncology | non-small-cell lung cancer | Other |
| Phase 1 | HER-2 | Omnitarg (rhuMAb-2C4) | Pertuzumab | Genentech | Roche | Oncology | Multiple solid tumors | Humanized |
| Phase 1 | HM1.24 | AHM | | Chugai | 0 | Oncology | myeloma | Humanized |
| Phase 1 | I-131 | Cotara | 0 | Peregrine Pharma | 0 | Oncology | Colorectal cancer | Conjugated |
| Phase 1 | IgE | rhuMab-E26 | 0 | Genentech | Tanox | Respiratory | Asthma | Humanized |
| Phase 1 | IL-12/IL-23 | CNTO 1275 | 0 | Medarex | Centocor (J&J) | Immuno disease | Anti-Inflammatory (Anti-TNFa) | Fully Human |
| Phase 1 | Integrin a5b1 | Anti-a5b1integrin (M200) | 0 | Designed HorsepowerLabs | 0 | Oncology | Solid tumors | Chimeric |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Phase 1 | Integrin alpha-v / beta-3 | Vitaxin | Anti-A-V/B-3 integrin | ELI LILY | MedImmune | Immuno disease | Rheumatoid arthritis | Humanized |
| Phase 1 | Integrin αvβ3 | Vitaxin | Anti-αvβ3 | MedImmune | 0 | Immuno disease | RA | Humanized |
| Phase 1 | Integrin αvβ3 | Vitaxin | Anti-αvβ3 | MedImmune | 0 | Oncology | Malanoma | Humanized |
| Phase 1 | Integrin αvβ3 | Vitaxin | Anti-αvβ3 | MedImmune | 0 | Oncology | prostate cancer | Humanized |
| Phase 1 | K3 | Anti-tissue factor | 0 | Genentech | Roche | Cardiovascular | Acute coronary syndrome | Fragment |
| Phase 1 | KDR receptor | IMC-1C11 | Anti-KDR | ImClone | 0 | Oncology | Metastatic colorectal carcinoma | Humanized |
| Phase 1 | Lewis Y | IGN301 | Anti-Lewis Y | Igeneon | In-licensed from Novartis | Oncology | Epithelial cancers | Antibody vaccine |
| Phase 1 | MUC-1 | Therex, R1550 | huHMFG1 | Antisoma | Roche | Oncology | Breast cancer | Humanized |
| Phase 1 | MUC18 | ABX-MA1 | Anti-MUC18 | Amgen | 0 | Oncology | Metastatic melanoma | Fully human |
| Phase 1 | PSMA | J591 | Anti-Prostate-Specific Membrane Antigen | Biovation (Merck KGaA) | Agen BioMedical | Oncology | Prostate cancer | Chimeric |
| Phase 1 | PSMA | MLN2704 | 0 | ImmunoGen | Millenium | Oncology | Cancer | Conjugated |
| Phase 1 | PSMA | MDX-070 | 0 | Medarex | 0 | Oncology | Prostate Cancer | Fully human |
| Phase 1 | T-cells | Nuvion | Visilizumab | NuvionNuvion Labs | 0 | Immuno disease | Ulcerative Colitis | Humanized |
| Phase 1 | T cell | MT103 | Bi-specific T cell engagers | MedImmune | Micromet AG | Oncology | NHL | Other |
| Phase 1 | T cell | MT104 | Bi-specific T cell engagers | MedImmune | Micromet AG | Oncology | CLL | Other |
| Phase 1 | T cell | MT103 | Bi-specific antibody derivative | Micromet AG | MedImmune | Oncology | B cell tumors | Other |
| Phase 1 | TAG-72 | Pretarget carcinoma | Anti-TAG-72 fusion protein | NeoRx Corp | 0 | Oncology | Carcinoma | Conjugated |
| Phase 1 | TNF-α | CNTO 148 | Golimumab | Medarex | Centocor (J&J) | Immuno disease | Anti-Inflammatory (Anti-TNFa) | Fully human |
| Phase 1 | TRAIL-R1 | TRAIL-R1 | Anti-TRAIL Receptor-1 | CAT | HGS, Takeda | Oncology | Tumor (misc) | Fully human |
| Phase 1 | TRAIL-R1 | TRAIL-R1 | 0 | HGS | CAT/Takeda | Oncology | solid tumor | Fully human |
| Phase 1 | TRAIL-R2 | TRAIL-R2 mAb | 0 | CAT | HGS | Oncology | Several cancers | Fully human |
| Phase 1 | TRAIL-R2 | TRAIL Receptor-2 | 0 | HGS | CAT/Kirin Brewery | Oncology | Advanced tumors | Fully human |

(*Continued*)

**TABLE 2** ***Continued***

| Development Stage | Targets | Brand/Code | Generic | Company | Partners | Therapy area | Indication | Antibody class |
|---|---|---|---|---|---|---|---|---|
| Phase 1 | Vascular Adhesion Protein-1 | HUVAP | Vapaliximab | BioTie | Boehringer Ingelheim | Immuno disease | Inflammation | Chimeric |
| Phase 1 | Fibrin D-Dimer | 3B6 | 0 | Biovation (Merck KGaA) | Agen BioMedical | Cardiovascular | Thrombosis | Chimeric |
| Phase 1 | PTHrP | CAL | 0 | Chugai | 0 | Oncology | 0 | Humanized |
| Phase 1 | CTA1 | KSB310 | 0 | KS Biomedix | Intracel Corporation | Oncology | Ovarian cancer | Fully human |
| Phase 1 | Undisclosed | NVS Antibody 1 | Novartis Pharma AG | Medarex | Novartis | Immuno disease | Autoimmune Disease | Fully human |
| Phase 1 | Undisclosed | NVS Antibody 2 | 0 | Medarex | Novartis | Immuno disease | Autoimmune Disease | Fully human |
| Phase 1 | β-2 Subunit of integrins | MLN2201 | Anti-beta-2 subunit of integrins | Millennium | XOMA Ltd | Immuno disease | vascular inflammation | Humanized |
| Phase 1 | β-2 Subunit of integrins | MLN2201 | 0 | XOMA | Millennium | Immuno disease | Vascular inflammation | Humanized |
| Phase 1 | Protective antigene | Abthrax | Against anthrax infection | HGS | CAT | Infectious diseases | Anthrax infection | Fully human |
| Phase 1 | Calicheamicin | CMC-544 | Ab-calicheamicin conjugate | Celltech | Wyeth | Oncology | NHL | Fragment |
| Phase 1 | GFR | CDP 791 | Anti-GFR Ab fragment | Celltech | 0 | Oncology | Cancer | Conjugated |
| Phase 1 | PAM4 | hPAM4 | 0 | Immunomedics | 0 | Oncology | Pancreatic cancer | Humanized |
| Phase 1 | IL-12 | Anti-IL12 | 0 | Protein Design Labs | 0 | Immuno disease | Autoimmune disease | Humanized |
| Phase 1 | Integrin Alpha4 | CDP-323 | Alpha4 integrin antagonist | Celltech | 0 | Immuno disease | Inflammatory disease | Fragment |
| Phase 1 | MHC class II molecules | 1D09C3 | Anti-MHC class II | GPC Biotech AG | 0 | Oncology | Cancer (lymphoid tumors) | Fully human |
| Phase 1 | TGF β | GC 1008 | Anti-TGF beta | Genzyme | CAT | Cardiovascular | pulmonary fibrosis | Unknown |
| Phase 1/2 | CEA | CEA-Cide | labetuzumab | Immunomedics | 0 | Oncology | Colorectal/breast cancer | Humanized |
| Phase 1/2 | CTLA-4 | MDX-018 | HuMax-Inflam | Medarex | Genmab | Immuno disease | Autoimmune disease | Fully human |
| Phase 1/2 | PSMA | MLN2704 | Anti-PSMA | Millennium | 0 | Oncology | Metastatic androgen-independent prostate cancer | Conjugated |
| Phase 1/2 | CD20 | HuMax-CD20 | Ant-CD20 | Genmab | 0 | Oncology | Non Hodgkin's Lymphoma | Fully human |
| Phase 1/2 | Humanized CD20 | hCD20 | 0 | Immunomedics | 0 | Oncology | NHL | Humanized |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Phase 2 | Anti-Lewis Y (Le(Y) | SGN-15 + Taxotere | BR96-doxorubicin | Seattle Genetics | 0 | Oncology | Lung | Conjugated |
| Phase 2 | A-V/B-3 | Vitaxin | Anti-A-V/B-3 integrin | ELI LILY | MedImmune | Oncology | Colorectal cancer / Several tumors | Humanized |
| Phase 2 | C 5 | Eculizumab | Anti-C5 | Alexion | Lonza | Immuno disease | Rheumatoid Arthritis / nephritis / Dermatomyositis | Unknown |
| Phase 2 | CA125 | OvaRex | Anti-CA125 | AltaRex | United Therapeutics | Oncology | Ovarian cancer | Murine |
| Phase 2 | Capsular polysaccharides | Altastaph | 0 | NABI Pharma | 0 | Infectious diseases | Staph infections | Other |
| Phase 2 | CD11a | Raptiva | Efalizumab anti-CD11a | XOMA | Genentech/ Roche | Immuno disease | Moderate-to-severe rheumatoid arthritis / Psoriatic arthritis | Humanized |
| Phase 2 | CD14 | IC14mAb | Anti-CD14 | ICOS | 0 | Infectious diseases | Sepsis | Unknown |
| Phase 2 | CD14 | IC14mAb | Anti-CD14 | ICOS | 0 | infectious diseases | Pneumonia | Unknown |
| Phase 2 | CD20 | Rituxan | Rituximab | Biogen-IDEC | Roche/ Genentech | Immuno disease | Idiopathic Thrombocytopenic Purpura | Chimeric |
| Phase 2 | CD20 | Rituxan | Rituximab | Biogen-IDEC | Roche/ Genentech | Immuno disease | Rheumatoid arthritis | Chimeric |
| Phase 2 | CD20 | Rituxan | Rituximab | Genentech | Roche/IDEC | Immuno disease | Idiopathic Thrombocytopenic Purpura | Chimeric |
| Phase 2 | CD20 | Rituxan | Rituximab | Genentech | Roche/IDEC | Immuno disease | Rheumatoid arthritis | Chimeric |
| Phase 2 | CD20 | Lymphocide-Y-90 | Epratuzumab-Y-90 | Immunomedics | 0 | Oncology | NHL B-cell lymphomas | Conjugated |
| Phase 2 | CD23 | IDEC-152 | Anti-CD23 | Biogen-IDEC | Seikagaku corporation | Respiratory | Asthma and allergic disorders | Other |
| Phase 2 | CD3 | Nuvion | Visilizumab | Protein Design Labs | 0 | Immuno disease | Graft vs Host disease | Humanized |
| Phase 2 | CD3 | TRX4 | Anti-CD3 | TolerRx | 0 | Diabetes and endocrinology | Type I diabetes | Humanized |
| Phase 2 | CD30 | MDX-060 | 0 | Medarex | 0 | Oncology | CD30+ lymphomas | Fully human |
| Phase 2 | CD30 | SGN-30 | Anti-CD30 | Seattle Genetics | 0 | Oncology | Hodgkin's disease | Unknown |

(*Continued*)

**TABLE 2** ***Continued***

| Development Stage | Targets | Brand/Code | Generic | Company | Partners | Therapy area | Indication | Antibody class |
|---|---|---|---|---|---|---|---|---|
| Phase 2 | CD4 | IDEC-151 | Clenoliximab Anti-CD4 | Biogen-IDEC | 0 | Immuno disease | RA | Other |
| Phase 2 | CD4 | HuMax CD4 | Anti-CD4 | Genmab | 0 | Immuno disease | Psoriasis | Fully human |
| Phase 2 | CD4 | HuMax CD4 T-Cell Lymphoma | Anti-CD4 | Genmab | 0 | Oncology | T-cell lymphoma | Fully human |
| Phase 2 | CD4 | HuMaxTM-CD4 | Anti-CD4 | Medarex | GenMab | Immuno disease | Psoriasis | Fully human |
| Phase 2 | CD44v6 | BIWA4 | Bivatuzumab mertansine | Boehringer Ingelheim | ImmunoGen | Oncology | Head and neck cancer | Humanized |
| Phase 2 | CD5 | HuMaxTM-CD5 | 0 | Medarex | GenMab | Immuno disease | Immuno-Disease | Fully human |
| Phase 2 | CD53 | Campath | Alemtuzumab anti-CD53 | ILEX Oncology | Berlex, Schering AG | Oncology | NHL | Humanized |
| Phase 2 | CD54 | Campath | Alemtuzumab anti-CD54 | ILEX Oncology | Berlex, Schering AG | CNS | MS | Humanized |
| Phase 2 | CD81 | IDEC-115/ rituximab | Anti-CD81 plus rituximab | Biogen-IDEC (ID) | Mitsubishi Pharma (Japan) | Oncology | relapsed NHL | Other |
| Phase 2 | CEA | CEA-Cide-Y-90 | Labetuzumab-Y-90 | Immunomedics | 0 | Oncology | Colorectal/pancreatic | Conjugated |
| Phase 2 | CEA | CEA-Cide-I-131 | Labetuzumab-I-131 | Immunomedics | 0 | Oncology | Metastatic colorectal cancer | Conjugated |
| Phase 2 | CEA | KSB309 | 0 | KS Biomedix | Intracel Corporation | Oncology | Head and neck cancers | Fully human |
| Phase 2 | CEA | KSB303 | Anti-CEA Mab | KS Biomedix | YM Biosciences | Oncology | Late-stage colon and pancreatic cancer | Unknown |
| Phase 2 | Clostridium difficile | DiffGAM | Anti-Clostridium difficile immunoglobulins | Immucell | 0 | Infectious diseases | gastric infections | Other |
| Phase 2 | CTLA-4 | MDX-010 + gp100 | 0 | Medarex | 0 | Oncology | Melanoma | Fully human |
| Phase 2 | CTLA-4 | MDX-010 +/- chemotherapy | 0 | Medarex | 0 | Oncology | Melanoma | Fully human |
| Phase 2 | CTLA-4 | MDX-010 +/- chemotherapy | 0 | Medarex | 0 | Oncology | Prostate Cancer | Fully human |
| Phase 2 | EGFR | ABX-EGF | Anti-EGFr | Amgen | Amgen | Oncology | Various cancers | Fully human |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Phase 2 | EGFR | HuMax-EGFr | Anti-EGFr | Genmab | 0 | Oncology | Head and neck cancer | Fully human |
| Phase 2 | Eotaxin-1 | CAT-213 | Anti-Eotaxin-1 | CAT | 0 | Respiratory | Allergies | Fully human |
| Phase 2 | Ep-CAM | MT201 | Anti-EpCAM | Micromet AG | Novuspharma SpA | Oncology | Prostate cancer | Fully human |
| Phase 2 | FAP.α | BIBH1 | Sibrotuzumab | Boehringer Ingelheim | 0 | Oncology | Metastatic colorectal cancer | Humanized |
| Phase 2 | GD3 | KW-2871 | Anti-GD3 | Kyowa Hakko | 0 | Oncology | Melanoma | Chimeric |
| Phase 2 | gp120 | PRO 542 | Antibody-like molecule to gp120 | Progenics | Genzyme Transgenics | Infectious diseases | HIV | other |
| Phase 2 | HER | Omnitarg (rhuMAb-2C4) | Pertuzumab | Genentech | Roche | Oncology | Prostate cancer | Humanized |
| Phase 2 | HER | Omnitarg (rhuMAb-2C4) | Pertuzumab | Genentech | Roche | Oncology | Ovarian cancer | Humanized |
| Phase 2 | HER | Omnitarg (rhuMAb-2C4) | Pertuzumab | Genentech | Roche | Oncology | Lung cancer | Humanized |
| Phase 2 | HER | Omnitarg (rhuMAb-2C4) | Pertuzumab | Genentech | Roche | Oncology | Breast cancer | Humanized |
| Phase 2 | HLA-DR | Oncolym | IgG2a Mab/Lym-1 | Peregrine Pharma | 0 | Oncology | B-cell lymphomas | Conjugated |
| Phase 2 | IFN-γ | HuZaf | Anti-IFN-gamma | Protein Design Labs | 0 | Immuno disease | Crohn's disease | Humanized |
| Phase 2 | IFN-γ | HuZaf | Anti-IFN-gamma | Protein Design Labs | 0 | Immuno disease | Psoriasis | Humanized |
| Phase 2 | IgE | TNX-901 | Anti-IgE | Tanox | 0 | Respiratory | Allergy | Unknown |
| Phase 2 | IgM | H11 scFv | Anti-IgM | Viventia Biotech Inc | 0 | Oncology | Breast/prostate cancers, melanoma | Fragment |
| Phase 2 | IL-2R(α) | Zenapax | Daclizumab Anti-CD25 | Protein Design Labs | 0 | Immuno disease | Graft vs Host disease | Humanized |
| Phase 2 | IL-12 | ABT-874 (former J695) | 0 | CAT | Abbott | Immuno disease | Crohn's Disease | Fully human |
| Phase 2 | IL15 | HuMax IL15 | Anti-IL15 | Genmab | Amgen | Immuno disease | Inflammatory, autoimmune diseases | Fully human |
| Phase 2 | IL15 | HuMaxTM-IL15 | Anti-IL15 | Medarex | GenMab | Oncology | T-cell lymphoma | Fully human |
| Phase 2 | IL-2R | Zenapax | Daclizumab Anti-CD25 | Protein Design Labs | 0 | CNS | Multiple sclerosis | Humanized |
| Phase 2 | IL-2R | Zenapax | Daclizumab Anti-CD25 | Protein Design Labs | 0 | Diabetes and endocrinology | Type 1 diabetes | Humanized |
| Phase 2 | IL-2R | Zenapax | Daclizumab Anti-CD26 | Protein Design Labs | 0 | Immuno disease | Ulcerative colitis | Humanized |
| Phase 2 | IL-2R | Zenapax | Daclizumab Anti-CD25 | Protein Design Labs | 0 | Immuno disease | Uveitis | Humanized |

(*Continued*)

**TABLE 2** ***Continued***

| Development Stage | Targets | Brand/Code | Generic | Company | Partners | Therapy area | Indication | Antibody class |
|---|---|---|---|---|---|---|---|---|
| Phase 2 | IL-2R | Zenapax | Daclizumab Anti-CD25 | Protein Design Labs | 0 | Respiratory | Asthma | Humanized |
| Phase 2 | IL-4 | Anti-IL4 | Pascolizumab | Protein Design Labs | 0 | Respiratory | Steriod-naïve asthma | Humanized |
| Phase 2 | IL-5 (human) | 240563 | Mepolizumab | GSK | 0 | Respiratory | 0 | Humanized |
| Phase 2 | Integrin Alpha4beta7 heterodimer (gut-specific) | MLN-02 | 0 | Genentech | Roche/ Millennium | Immuno disease | Crohn's disease | Humanized |
| Phase 2 | Integrin Alpha4beta7 heterodimer (gut-specific) | MLN-02 | 0 | Genentech | Roche/ Millennium | Gastrointestinal | Ulcerative colitis | Humanized |
| Phase 2 | Integrin α4β7 | MLN02 | Anti-alpha4beta7 | Millennium | Genentech/ Roche | Immuno disease | Crohn's disease/ ulcerative colitis | Humanized |
| Phase 2 | Integrin αvβ3 | Vitaxin | Anti-αvβ3 | MedImmune | 0 | Oncology | Colorectal cancer | Humanized |
| Phase 2 | Integrin αvβ3 | Vitaxin | Anti-αvβ3 | MedImmune | 0 | Oncology | Refractory solid tumors | Humanized |
| Phase 2 | Lewis Y | IGN311 | Anti-Lewis Y | Igeneon | In-licensed from PDL | Oncology | Epithelial cancers | Humanized |
| Phase 2 | Mimics CEA/ HMFG | CeaVac/TriAb | Mimics the HMFG Ag | Titan Pharmaceuticals | 0 | Oncology | NSCLC | Unknown |
| Phase 2 | Mimics the GD2 | 4B5 | Anti-GD2 | Viventia Biotech Inc | 0 | Oncology | Melanoma, neuroblastoma, SCCLung | Fully human |
| Phase 2 | MUC1 | BrevaRex | Anti-MUC1 | AltaRex | 0 | Oncology | Myeloma | Murine |
| Phase 2 | MUC-1 | R1550 | Pemtumomab | Antisoma | Roche | Oncology | Gastric cancer | Conjugated |
| Phase 2 | PDGF (beta receptor for Platelet Derived Growth Factor) | CDP 860 | PEG-anti-PDGF-b-Ab fragment | Nektar | Celltech | Oncology | Solid tumors | Fragment |
| Phase 2 | PDGF-b | CDP 860 | Anti-PDGF-b Ab fragment | Celltech | 0 | Oncology | Cancer | Conjugated |
| Phase 2 | TGFß1 | CAT-192 | Anti-TGFß1 | CAT | Genzyme | CNS | Diffuse systemic sclerosis | Fully human |
| Phase 2 | TGFß1 | CAT-192 | Anti-TGFß1 | Genzyme | CAT | CNS | Diffuse systemic sclerosis | Fully human |
| Phase 2 | TNFa | CDP 870 | Anti-TNFa Ab | Celltech | Pfizer | Immuno disease | Crohn's disease | Fragment |
| Phase 2 | TNF-α | D2E7 | Adalimumab | Eisai | Abbott | Immuno disease | Rheumatoid arthritis | Fully human |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Phase 2 | VEGF | Lucentis (rhufab V2) | Ranibizumab | Genentech | Roche/Novartis | Immuno disease | age-related macular degeneration | Fragment |
| Phase 2 | PSMA | MLN591RL | 0 | Millennium | 0 | Oncology | prostate cancer | Conjugated |
| Phase 2 | Interleukin-12 p40 protein | ABT-874 (former J695) | 0 | Abbott | CAT | Immuno disease | Autoimmune disease | Fully human |
| Phase 2 | CD44v6 | BIBW2 | Bivatuzumab mertansine | Boehringer Ingelheim | ImmunoGen | Oncology | Skin cancer | Humanized |
| Phase 2 | Undisclosed | HuMax Inflam | 0 | Genmab | Medarex | Immuno disease | Autoimmune disease | Fully human |
| Phase 2 | α-fetoprotein | AFP-Cide-Y-90 | 0 | Immunomedics | 0 | Oncology | Liver/germ-cell | Conjugated |
| Phase 2 | Hepatitis C virus | Civacir | 0 | NABI Pharma | 0 | Infectious diseases | Hepatitis C | Other |
| Phase 2 | An adenocarcinoma antigen | COU-1 | Recognizes adenocarcinomas of colon, lung, breast, and ovary origin | NatImmune | 0 | Oncology | Adenocarcinomas (colorectal) | Fully human |
| Phase 2 | S. mutans SigA | CaroRx | Against the. bacterium Streptococcus mutans | Planet Biotechnology | 0 | Infectious diseases | Tooth decay | Other |
| Phase 2 | Ovarian Cancer | SGN-15 + Gemzar | BR96-Gemzar | Seattle Genetics | 0 | Oncology | Ovarian | Conjugated |
| Phase 2 | Idiotype | TriaAb/TriGem | 0 | Titan Pharmaceuticals | 0 | Oncology | SCLC | Unknown |
| Phase 2 | Hepatits C virus | HepeX-C (XTL002) | 0 | XTL Biopharma | 0 | Infectious diseases | Hepatitis C | Fully human |
| Phase 2 | Hepatits B virus | HepeX-B (XTL001 | 0 | XTL Biopharma | 0 | Infectious diseases | Hepatitis B | Fully human |
| Phase 2 | Complement (C5) | 5G1.1 transgenic | Anti-C5 | GTC Biotherapeutics | Alexion | Immuno disease | Rheumatoid Arthritis, Nephritis | Fully human |
| Phase 2 | HLA-DR beta | Remitogen | Apolizumab | Protein Design Labs | 0 | Oncology | Solid tumors | Humanized |
| Phase 2 | Radioactive molecule | Cotara | 0 | Peregrine Pharma | 0 | Oncology | Brain cancer | Conjugated |
| Phase 2 | A renal cancer cell antigen | WX-G250RIT | 0 | Wilex | 0 | Oncology | Cholangiocarcinoma/ gall bladder carcinoma | Conjugated |
| Phase 3 | C5 complement | Pexelizumab | Anti-C5 complement\ | Alexion | Procter & Gamble | Cardiovascular | Cardiopulmonary bypass / heart attack | Unknown |
| Phase 3 | CA IX/MN | Rencarex WX-G250 | 0 | Wilex | Centocor/J&J | Oncology | Renal cell carcinoma | Chimeric |
| Phase 3 | CD20 | Rituxan | Rituximab | Biogen-IDEC (ID) | Roche/ Genentech | Oncology | Aggressive non-Hodgkin's lymphoma | Chimeric |

(*Continued*)

**TABLE 2** ***Continued***

| Development Stage | Targets | Brand/Code | Generic | Company | Partners | Therapy area | Indication | Antibody class |
|---|---|---|---|---|---|---|---|---|
| Phase 3 | CD20 | Rituxan | Rituximab | Genentech | Roche/IDEC | Oncology | Aggressive non-Hodgkin's lymphoma | Chimeric |
| Phase 3 | CD-22 | Lymphocide | Epratuzumab | Amgen | Immunomedics | Oncology | B-cell non-Hodgkin's lymphoma | Humanized |
| Phase 3 | CD-22 | Lymphocide (hCD22) | Epratuzumab | Immunomedics | Amgen | Oncology | NHL | Humanized |
| Phase 3 | CTLA4-Ig | Anti-CTLA4-Ig | MRI | RepliGen | BMS | Immuno disease | Rhematiod arthritis | Unknown |
| Phase 3 | EpCAM | IGN-101 | Anti-Ep-CAM | Igeneon | 0 | Oncology | Epithelial cancers | Unknown |
| Phase 3 | HER –2/neu | Osidem (IDM-1) | Anti-HER-2/neu | Medarex | IDM | Oncology | HER-2+ Ovarian Cancer | Conjugated |
| Phase 3 | human IgE | Xolair | Omalizumab | Tanox | 0 | Respiratory | Allergy | Humanized |
| Phase 3 | IL-2R | Zenapax | Daclizumab Anti-CD25 | Protein Design Labs | 0 | Immuno disease | Heart transplantation | Humanized |
| Phase 3 | IL-6 receptor | MRA | Anti-IL-6 receptor antibody | Chugai | 0 | Immuno disease | 0 | Humanized |
| Phase 3 | IL-6 receptor | MRA | Anti-IL-6 receptor antibody | Chugai | 0 | Oncology | 0 | Humanized |
| Phase 3 | Integrin $\alpha 4$ $\beta 1$ intergrin | ANTEGREN | Natalizumab | Biogen-IDEC | 0 | CNS | MS | Humanized |
| Phase 3 | mimics the GD3 epitope | BEC2 | Mitumomab | ImClone | Merck KGaA | Oncology | small cell lung carcinoma | Murine |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Phase 3 | MUC1 | R1549 | Pemtumomab | Antisoma | Roche | Oncology | Ovarian cancer | Conjugated |
| Phase 3 | TGF-beta2 | Trabio | Lerdelimumab/ CAT-152/ anti-TGFß2 | CAT | Integra LifeSciences | Other | Cataract (Eye Disorders) Glaucoma | Fully human |
| Phase 3 | TNF-a | CDP 870 | Anti-TNFa Ab | Celltech | Pfizer | Immuno disease | Rheumatoid arthritis | Fragment |
| Phase 3 | TNF-a | CDP 870 | PEG-anti-TNF-a Ab fragment | Nektar | Pfizer Celltech | Immuno disease | Rheumatoid arthritis | Fragment |
| Phase 3 | TNF-a | CDP870 | 0 | Pfizer | Celltech | Immuno disease | Rheumatoid arthritis | Fragment |
| Phase 3 | VEGF | Avastin | Bevacizumab-rhuMAb-VEGF | Genentech | Roche | Oncology | Metastatic colorectal cancer | Humanized |
| Phase 3 | VEGF | Avastin | Bevacizumab-rhuMAb-VEGF | Genentech | Roche | Oncology | Renal cell carcinoma | Humanized |
| Phase 3 | VGEF | Lucentis | Ranibizumab | Novartis | Genentech/ Roche | Immuno disease | Macular degeneration | Humanized |
| Phase 3 | VGEF | Lucentis | Ranibizumab | Novartis | Genentech/ Roche | Immuno disease | Macular degeneration | Humanized |
| Phase 3 | VGEF | Lucentis | Ranibizumab | Novartis | Genentech/ Roche | Immuno disease | Macular degeneration | Humanized |
| Phase 3 | Idiotype | CeaVac | Mimics the carcinoembryonic antigen | Titan Pharmaceuticals | 0 | Oncology | metastatic colorectal cancer | Unknown |
| Phase 3 | Integrin α4 β1 intergrin | Antegren transgenic | anti α4 intergrin | GTC Biotherapeutics | Elan | Immuno disease | MS, Crohn's disease | Fully human |

**TABLE 3 Current Phase III Chlinc Monoclonal Antibody Trials in the United States**

| Condition | Trial |
|---|---|
| Lymphome | Chemotherapy and Peripheral Stem Cell Transplant with or without Monoclonal Antibody therapy in Treating Patients with Non-Hodgkin's Lymphoma |
| Kidney cancer | Monoclonal Antibody Therapy (RencarexA) in Treating Patients Who Have Undergone Surgery for Non-Metastatic Kidney Cancer |
| Brain and central nervous system tumors | Isotretinoin with or without Monochlonal Antibody, Interleukin-2, and Sargramostim Following Stem Cell Transplantation in Treating Patients with neuroblastoma |
| Brain and central nervous system tumors | Combination Chemotherapy with or without Filgrastim before Surgery, High-Dose Chemotherapy, and Radiation therapy Followed by Isotretinoin with or without Monoclonal Antibody in Treating Patients with Neuroblastoma |
| Lymphoma | Combination Chemotherapy with Monoclonal Antibody Therapy in Treating Patients with Newly diagnosed Non-Hodgkin's Lymphoma |
| Lymphoma | Combination Chemotherapy Plus filgrastim with or without Rituximab in Treating Older Patients with Non-Hodgkin's Lymphoma |
| Leukemia | Combination Chemotherapy with or without Monoclonal Antibody Therapy Followed by Stem Cell Transplant in Treating Patients with Acute Myeloid Leukemia |
| Ulcerative colitis | Efficacy and Safety of Adalimumab in Subjects with Moderately to Severely Acute Ulcerative Colitis |
| Lymphoma | Cyclophosphamide, Doxorubicin, Etoposide, and Prednisone OR Cyclophosphamide, Doxorubicin, Vincristine, Prednisone, and Rituximab Followed by Vaccine Therapy Plus Sargramostim in Treating Patients with Stage II, Stage III, or Stage IV Follicular Lymphoma |
| Breast cancer | Combination Chemotherapy with or without Capecitabine and/orTrastuzumab Before Surgery in Treating Women With Stage I, Stage II, or Stage III Breast Cancer |
| Ulcerative colitis | Efficacy and Safely of Adalimumab in Subjects with Moerately to Severely Acute Ulcerative Colitis |
| Crohn's disease | Open-Lable Study of Adalimumab for the induction and Maintonance of Clinical Remission in Subjects with Moderate to Severe Conhn's Disease |
| Crohn's disease | Efficacy and Safely of Adalimumab in Pediatric Subjects with Moderate to Severe Conhn's Disease |
| Breast cancer | Trastuzumab with or without Paclitaxel in treating Women with Metastatic Breast cancer That Overexpresses HER2 |
| Leukemia | Cytarabine and Daunorubicin with or without Gemtuzumab Ozogamicin Followed by High-Dose Cytarabine Followed by Either Gemtuzumab Ozogamicin or No Additional Therapy in Treating Patients with previously Untreated De Novo Acute Myeloid Leukemia |
| Colorectal cancer | Comparison of Combination Chemotherapy Regimens with or without Cetuximab in Treating Patients Who Have Undergone Surgery for Stage III Colon Cancer |
| Colorectal cancer, metastatic cancer | Combination Chemotherapy with or without Cetuximab before and after Surgery in Treating Patients with Reaectable Liver Metastases Caused by Colorectal Cancer |
| Colorectal cancer | Cetuximab and/or Bevacizumab Conbined with Combination Chemotherapy in Treating Patients with Metastatic Colorectal Cancer |
| Renal cell carcinoma (kidney) | cG250 Monocional Antibody vs Placebo for Renal Cell Carcinoma |
| Rheumatoid arthritis | Safety and Efficacy of Intravenous ACZ885 and Oral Methotrexate Therapy in Patients with Early Rheumatoid Arthritis |
| Muchle–Wells syndrome | Efficacy, Safety, and Tolerability of ACZ885 in Patients with Muckle-Wells Syndrome |
| Lymphoma | Rituximab in Treating Patients with Low Tumor Burden Indolent Non-Hodgkin's Lymphoma |
| Lymphoma | Combination Chemotherapy with or without Rituximab in Treating Older Patients with Non-Hodgkin's Lymphoma |
| Lymphoma | Rituximab in Treating Patients with Newly Diagnosed Stage II, Stage III, or Stage IV Follicular Non-Hodgkin's Lymphoma |
| Breast cancer | Trastuzumab for 6 Months or 1 Year in Treating Women with Nonmetastatic Breast Cancer That Can Be Removed by Surgery |
| Lymphoma | Rituximab in Treating Patients with Flooicular Non-Hodgkin's Lymphoma |
| Leukemla | Combination Chemotherapy with or without Gemtuzumab in Treating Yong Patients with Newly Diagnosed Acute Myeloid Leukemia |
| Lymphoma | Rituximab and Combination Chemotherapy in Treating Patients with Non-Hodgkin's Lymphoma |

**TABLE 3** ***Continued***

| Condition | Trial |
|---|---|
| Colorectal cancer | Combination Chemotherapy with or without Cetuximab as First-Line Therapy in Treating Patients with Metastatic Colorectal Cancer |
| Leukemia; myelodysplastic syndromes | Combination Chemotherapy with or without Gemtuzumab Ozogamicin or Tipifamib in Treating Patients with Acute Myeloid Leukemia or High-Risk Myelodysplastic Syndromes |
| Gastrointestinal stromal tumor | Imatinib Mesylate with or without Bevacizumab in Treating Patients with Metastatic or Unresectable Gastrointestinal Stromal Tumor |
| Breast cancer | Two Different Combination Chemotherapy Regimens Given with Trastuzumab in Treating Women with Palpable Breast Cance That Can Be Removed by Surgery |
| Leukemia | Combination Chemotherapy with or without Gemtuzumab Followed by Tretinoin, Mercaptopurine, and Methotexate or Observation in Treating Patients with Acute Promyelocytic Leukemia |
| Lymphoma | Rituximab and Combination Chemotherapy in Treating Patients with Diffuse Large B-Cell Non-Hodgkin's Lymphoma |
| Gastric cancer | Combination Chemotherapy with or without Bevacizumab in Treating Patients with Previously Untreated Stomach Cancer or Gastroesophageal Junction Cancer That Can Be Removed by Surgery |
| Prostate cancer | Docetaxel and Prednisone with or without Bevacizumab in Treating Patients with Prostate Cancer That Did Not Respond to Hormone Therapy |
| Colorectal cancer | Irinotecan and Cetuximab with or without Bevacizumab in Treating Patients with Metastatic Colorectal Cancer That Progressed During First-Line Therapy |
| Lymphoma | Rituximab and Combination Chemotherapy with or without Radiation Therapy in Treating Patients with B-Cell Non-Hodgkin's Lymphoma |
| Leukemia | Rituximab, Fludarabine, and Cyclophosphamide or Observation Alone in Treating Patients with Stage 0, Stage I, or Stage II Chronic Lymphocytic Leukemia |
| Leukemia | Early or Delayed Fludarabine and Rituximab in Treating Patients with Previously Untreated Chronic Lymphocytic Leukemia |
| Falloplan tube cancer, ovarian cancer; peritoneal cavity cancer | Carboplatin and Paclitaxel with or without Bevacizumab in Treating Patients with Stage III or Stage IV Ovarian Epithelial or Primary Peritoneal Cancer |
| Breast cancer | First-Line Chemotherapy and Trastuzumab with or without Bevacizumab in Treating Patients with Metastatic Breast Cancer That Overexpresses HER2/New |
| Colorectal cancer | Irinotecan with or without Panitumumab or Cyclosporine in Treating Patients with Advanced or Metastatic Colorectal Cancer That Did Not Respond to Fluorouracil |
| Esophageal cancer | Cisplatin, Capecitabine , and Radiation Therapy with or without Cetuximab in Treating Patients with Esophageal Cancer |
| Fallopian tube cancer, ovarian cancer; peritoneal cavity cancer | Carboplatin and Paclitaxel with or without Bevacizumab in Treating Patients with Newly Diagnosed Ovarian Epithelial Cancer, Fallopian Tube Cancer, or Primary Peritoneal Cavity Cancer |
| Lung cancer | Radiation Therapy and Cisplatin with or without Cetuximab in Treating Patients with Stage III or Stage IV Head and Neck Cancer |
| Colorectal cancer | Combination Chemotherapy with or without Bevacizumab in Treating Patients who Have Had Surgery for Stage II or Stage III Rectal Cancer |
| Breast cancer | Doxorubicin, Cyclophosphamide, and Paclitaxel with or without Bevacizumab in Treating Patients with Lymph Node-Positive or High-Risk, Lymph Node-Negative Breast Cancer |
| Lung cancer | Chemotherapy with or without Bevacizumab in Treating Patients with Stage IB, Stage II, or Stage IIA Non-Small Lung Cancer That Was Removed by Surgery |
| Colorectal cancer | Combination Chemotherapy and Bevacizumab with or without Eriotinib in Treating Patients with Metastatic Colorectal Cancer That Cannot Be Removed by Surgery |
| Type 1 diabetes mellitus | The Protégé Study—Clinical Trial of MGA031 in Children and Adults with Recent-Onset Type 1 Diabetes Mellitus |
| Lymphoma | Comparison of Two Induction Combination Chemotherapy Regimens Followed by High-Dose Consolidation Chemotherapy and Autologous Stem Cell |
| Colorectal cancer | Combination Chemotherapy with or without Cetuximab in Treating Patients with Stage III Colon Cancer That Was Completely Removed by Surgery |

(*Continued*)

**TABLE 3** ***Continued***

| Condition | Trial |
|---|---|
| Breast cancer | Chemotherapy with or without Bevacizumab in Treating Women with Stage I, Stage II or Stage IIA Breast Cancer That Can Be Removed by Surgery |
| Lymphoma | Comparison of Two Salvage Chemotherapy Regimens before Autologous Stem Cell Transplantation with or without Maintonance Rituximab in Treating Patients with Relapsed or Refractory Aggressive Non-Hodgkin's Lymphoma |
| Colorectal cancer | Oxaliplatin, Leucovorin, and Fluorouracil with or without Bevacizumab in Treating Patients Who Have Undergone Surgery for Stage II Colon Cancer |
| Mysosis fungoides; sezary | Study of Human Monoclonal Antibody to Treat Mycosis Fungoides and Sezary Syndrome |
| Lymphoma | Chemoradiotherapy and Peripheral Stem Cell Transplantation Compared with Combination Chemotherapy in Treating Patients with Non-Hodgkin's Lymphoma |
| Breast cancer | Adjuvant Doxorubicin, Cyclophosphamide, and Paclitaxel in Treating Patients with Breast Cancer |
| Melanoma; metastases | MDX-010 Antibody, MDX-1379 Melanoma Vaccine, or MDX-010/MDX-1379 Combination Treatment for Patients with Melanoma |
| Ulcerative colitis | Open-Label Adalimumab for Ulcerative Colitis Patients |
| Crohn's disease | Study of Adalimumab for the Maintenance of Clinical Remission in Japanese Subjects with Crohn's Disease |
| Crohn's disease | Study of Adalimumab for the Induction of Clinical Remission in Japanese Subjects with Crohn's Disease |
| Crohn's disease | Effects of Adalimumab on Mucosal Healing in Subjects with Crohn's Disease Involving the Colon |
| Lupus erythematosus, systemic; Lupus nephritis | TNF Blockade with Remicade in Active Lupus Nephritis WHO Class V (TRIAL) |
| Non-Hodgkin's lymphoma | Zevalin and BEAM High-Dose Chemotherapy Compared with BEM Alone as Conditioning Regimen in Patients with Chemosensitive Repapse of Non-Hodgkin's Lymphoma |
| Polyneuropathy associated with anti-MAG IgM gamopathy | RIMAG Study: Trial of Rituximab versus Placebo in Polyneuropathy Associated with Anti-MAG IgM Monoclonal Gammopathy |
| Membranoproliferative | Pilot Study of Rituximab for Membranoproliferative Glomerulonephritis |
| Renal transplantation | Efficacy and Safety of Enteric-Coated Mycophenolate Sodium and Cyclosporine in Combination with and without Steroids, in Adult Renal Transplant Recipients |
| Arthritis | Intraarticular Injection of Infliximab |
| Systemic lupus erythematosus | Study of Belimumab, a Fully Human Monoclonal Anti-BLyS Antibody, in Subjects with Systemic Lupus Erythematosus (SLE) |
| Systemic lupus erythematosus | Study of Belimumab, a Fully Human Monoclonal Anti-BLyS Antibody, in Subjects with Systemic Lupus Erythematosus (SLE) |
| Follicular Lymphoma | HuMax-CD20 in Follicular Lymphoma (FL) Patients Refractory to Rituximab |
| B-cell chronic lymphocytic | HuMax-CD20 in B-Cell Chronic Lymphocytic Leukemia (B-CLL) Patients Failing Fludarabine and Alemtuzumab |
| Head and Neck cancer; squamous cell cancer | Zalutumumab in Patients with Non-Curable Head and Neck cancer |
| Acute myeloid leukemia; myelodysplastic syndrome | To Determine the Role of Adding Campath-1H or ATG Given in-Vivo in Addition to Fludarabine and Low Dose Busulfex on Outcome in Patients Treated with Reduced Intensity Conditioning |
| Liver transplantation | Tacrolimus and Daclizumab versus Tacrolimus and Steroides in Liver Recipients Receiving Aub-Optimal Grafts |
| Non-Hodgkin's lymphoma | Study of Rituximab versus Iodine I 131 Tositumomab Therapy for Patients with Non-Hodgkin's Lymphoma |
| Large-cell, immunoblastic; lymphoma, Non-Hodgkin | Comparison of Rituxan versus Bexxar When Combined with Carmustine, Etoposied, Cytarabine and Melphalan (BEAM) with Autologous Hematopoietic Stem Cell Transplantation (ASCT) |
| Leukemia, myelocytic, acute | Treatment of Patients with Newly Diagnosed Acute Myeloid Leukemia or Myelodyspiasia |
| Hypereosinophilic syndrome | Compassionate Use of Mepolizumab in Subjects with Hypereosinophilic Syndrome (HES) |
| Metastatic breast neoplasm | Vinorelbine+Trastuzumab vs Docetaxel+Trastuzumab as 1. Line Treatment for HER2+ Metastatic Breast Cancer |

**TABLE 3** ***Continued***

| Condition | Trial |
|---|---|
| Macular degeneration; choroidal neovascularization | Efficacy and Safety of Verteporfin Photodynamic Therapy Administered with Ranibizumab in Patients with Subfoveal Choroidal Neovascularization |
| Acute myeloid leukemia | Efficacy/Safety of Verteporfin Photodynamic Therapy and Ranibizumab Compared with Ranibizumab in Patients with Subfoveal Choroidal Neovascularization |
| Acute myeloid leukemia | NOPHO-AML 2004 Study for Children with Acute Myeloid Leukemia |
| Diabetic macular edema; visual acuity; dlabetic retinopathy | Zevalin-Beam for Aggressive Lymphoma |
| Diabetic macular edema; visual acuity; dlabetic retinopathy | Treatment of Refractory Diabetic Macular Edema with Infliximab |

## REFERENCES


1. Waldmann TA (2003). Immunotherapy: Past, present and future. *Nat Med 9*, 269–277.
2. Weiner LM (2007). Building better magic bullets—Improving unconjugated monoclonal antibody therapy for cancer. *Nat Rev Cancer 7*, 701–706.
3. Brekke OH, Sandlie I (2003).Therapeutic antibodies for human diseases at the dawn of the twenty-first century. *Nat Rev Drug Discov 2*, 52–62.
4. Köhler G, Milstein C (1975). Continuous cultures of fused cells secreting antibody of predefined specificity. *Nature 256*, 495–497.
5. Oh CS, Sollinger HW, Stratta RJ, Kalayoglu M, Belzer FO (1988). Delayed response to orthoclone OKT3 treatment for renal allograft rejection resistant to steroid and antilymphocyte globulin. *Transplantation 45*, 65–67.
6. Boulianne GL, Hozumi N, Shulman MJ (1984). Production of functional chimaeric mouse/human antibody. *Nature 312*, 643–646.
7. Morrison SL, Johnson MJ, Herzenberg LA, Oi VT (1984). Chimeric human antibody molecules: Mouse antigen-binding domains with human constant region domains. *Proc Natl Acad Sci USA 81*, 6851–6855.
8. Jones PT, Dear PH, Foote J, Neuberger MS, Winter G (1986). Replacing the complementarity-determining regions in a human antibody with those from a mouse. *Nature 321*, 522–525.
9. Weiner LM (2006). Fully human therapeutic monoclonal antibodies. *J Immunother 29*, 1–9.
10. Yarmush ML, Gates FT 3rd, Dreher KL, Kindt TJ (1981). Serologic and structural characterization of immunoglobulin chains secreted by rabbit-mouse hybridomas. *J Immunol 126*, 2240–2244.
11. Yarmush ML, Mandy WJ, Kindt TJ (1980). Evidence for linked expression of latent allotypes of the heavy chain constant and variable regions. *J Immunol 124*, 2864–2869.
12. Dreher KL, Sogn JA, Gates FT 3rd, Kuo MC, Kindt TJ (1983). Allotype-defined mRNA for rabbit immunoglobulin H and L chains isolated from rabbit-mouse hybridomas. *J Immunol 130*, 442–448.
13. Kuo MC, Sogn JA, Max EE, Kindt TJ (1985). Rabbit-mouse hybridomas secreting intact rabbit immunoglobulin. *Mol Immunol 22*, 351–359.
14. Babcook JS, Leslie KB, Olsen OA, Salmon RA, Schrader JW (1996). Isolation of functional antibody genes from single lymphocytes of defined antigen-specificity. *Proc Natl Acad Sci USA 93*, 7843–7848.
15. Rader C, Ritter G, Nathan S, Elia M, Gout I, Jungbluth AA, Cohen LS, Welt S, Old LJ, Barbas CF 3rd (2000). The rabbit antibody repertoire as a novel source for the generation of therapeutic human antibodies. *J Biol Chem 275*, 13668–13676.
16. Spieker-Polet H, Sethupathi P, Yam PC, Knight KL (1995). Rabbit monoclonal antibodies: Generating a fusion partner to produce rabbit-rabbit hybridomas. *Proc Natl Acad Sci USA 92*, 9348–9352.
17. Rossi S, Laurino L, Furlanetto A, Chinellato S, Orvieto E, Canal F, Facchetti F, Dei Tos AP (2005). Rabbit monoclonal antibodies: A comparative study between a novel category of immunoreagents and the corresponding mouse monoclonal antibodies. *Am J Clin Pathol 124*, 295–302.
18. Krause RM (1970) The search for antibodies with molecular uniformity. *Adv Immunol 12*, 1–56.
19. Bystryn JC, Jacobsen JS, Liu P, Heaney-Kieras J (1982) Comparison of cell-surface human melanoma-associated antigens identified by rabbit and murine antibodies. *Hybridoma 1*, 465–472.
20. Norrby E, Mufson MA, Alexander H, Houghten, RA (1987). Site-directed serology with synthetic peptides representing the large glycoprotein G of respiratory. *Proc Natl Acad Sci USA 84*, 6572–6576.
21. Weller A, Meek J, Adamson ED (1987). Preparation and properties of monoclonal and polyclonal antibodies to mouse epidermal growth factor (EGF) receptors:

evidence for cryptic EGF receptors in embryonal carcinoma cells. *Development 100*, 351–363.

22. Raybould TJ, Takahashi M (1988). Production of stable rabbit-mouse hybridomas that secrete rabbit mAb of defined specificity. *Science 240*. 1788–1790.
23. Karpas A, Dremucheva A, Czepulkowski BH (2001). A human myeloma cell line suitable for the generation of human monoclonal antibodies. *Proc Natl Acad Sci USA 98*, 1799–1804.
24. Vaisbourd M, Ignatovich O, Dremucheva A, Karpas A, Winter G (2001). Molecular characterization of human monoclonal antibodies derived from fusions of tonsil lymphocytes with a human myeloma cell line. *Hybrid Hybridomics 20*, 287–292.
25. Kametani Y, Shiina M, Katano I, Ito R, Ando K, Toyama K, Tsukamoto H, Matsumura T, Saito Y, Ishikawa D, Taki T, Ito M, Imai K, Tokuda Y, Kato S, Tamaoki N, Habu S (2006). Development of human-human hybridoma from anti-Her-2 peptide-producing B cells in immunized NOG mouse. *Exp Hematol 34*, 1240–1248.
26. Jakobovits A, Green LL, Hardy MC, Maynard-Currie CE, Tsuda H, Louie DM, Mendez MJ, Abderrahim H, Noguchi M, Smith DH, et al. (1995). Production of antigen-specific human antibodies from mice engineered with human heavy and light chain YACs. *Ann NY Acad Sci 764*, 525–535.
27. Green LL, Hardy MC, Maynard-Currie CE, Tsuda H, Louie DM, Mendez MJ, Abderrahim H, Noguchi M, Smith DH, Zeng Y, et al. (1994). Antigen-specific human monoclonal antibodies from mice engineered with human Ig heavy and light chain YACs. *Nat Genet 7*, 13–21.
28. Carter PJ (2006). Potent antibody therapeutics by design. *Nat Rev Immunol 6*, 343–357.
29. Dufner P, Jermutus L, Minter RR (2006). Harnessing phage and ribosome display for antibody optimisation. *Trends Biotechnol 24*, 523–529.
30. Hoogenboom HR (2005). Selecting and screening recombinant antibody libraries. *Nat Biotechnol 23*, 1105–1116.
31. Chao G, Lau WL, Hackel BJ, Sazinsky SL, Lippow SM, Wittrup KD (2006). Isolating and engineering human antibodies using yeast surface display. *Nat Protoc 1*, 55–768.
32. Colby DW, Kellogg BA, Graff CP, Yeung YA, Swers JS, Wittrup KD (2004). Engineering antibody affinity by yeast surface display. *Methods Enzymol 388*, 348–358.
33. Boder ET, Wittrup KD (2000). Yeast surface display for directed evolution of protein expression, affinity, and stability. *Methods Enzymol 328*, 430–444.
34. Ho M, Nagata S, Pastan I (2006). Isolation of anti-CD22 Fv with high affinity by Fv display on human cells. *Proc Natl Acad Sci USA 103*, 9637–9642.
35. Roland Wolkowicz, Gina C. Jager, and Garry P. Nolan (2005). A random peptide library fused to CCR5 for selection of mimetopes expressed on the mammalian cell surface via retroviral vectors. *J Biol Chem 280*: 15195–15201.
36. Schrama D, Reisfeld RA, Becker JC (2006). Antibody targeted drugs as cancer therapeutics. *Nat Rev Drug Discov 5*, 147–159.
37. Vaccaro C, Zhou J, Ober RJ, Ward ES (2005). Engineering the Fc region of immunoglobulin G to modulate in vivo antibody levels. *Nat Biotechnol 23*, 1283–1288.
38. Lum LG, Davol PA, Lee RJ (2006). The new face of bispecific antibodies: Targeting cancer and much more. *Exp Hematol 34*, 1–6.
39. Shen J, Vil MD, Jimenez X, Zhang H, Iacolina M, Mangalampalli V, Balderes P, Ludwig DL, Zhu Z (2007). Single variable domain antibody as a versatile building block for the construction of IgG-like bispecific antibodies. *J Immunol Methods 318*, 65–74.
40. Jimenez X, Lu D, Brennan L, Persaud K, Liu M, Miao H, Witte L, Zhu Z (2005). A recombinant, fully human, bispecific antibody neutralizes the biological activities mediated by both vascular endothelial growth factor receptors 2 and 3. *Mol Cancer Ther 4*, 427–434.
41. Holliger P, Hudson PJ (2005). Engineered antibody fragments and the rise of single domains. *Nat Biotechnol 23*, 1126–1136.
42. Holt LJ, Herring C, Jespers LS, Woolven BP, Tomlinson IM (2003). Domain antibodies: Proteins for therapy. *Trends Biotechnol 21*, 484–490.
43. Economides AN, Carpenter LR, Rudge JS, Wong V, Koehler-Stec EM, Hartnett C, Pyles EA, Xu X, Daly TJ, Young MR, Fandl JP, Lee F, Carver S, McNay J, Bailey K, Ramakanth S, Hutabarat R, Huang TT, Radziejewski C, Yancopoulos GD, Stahl N (2003). Cytokine traps: Multicomponent, high-affinity blockers of cytokine action. *Nat Med 9*, 47–52.
44. Smith SL (1996). Ten years of Orthoclone OKT3 (muromonab-CD3). *J Transpl Coord 6*, 109–119.
45. Ibbotson T, McGavin JK, Goa KL (2003). Abciximab: An updated review of its therapeutic use in patients with ischaemic heart disease undergoing percutaneous coronary revascularisation. *Drugs 63*, 121–1163.
46. Fietz T, Thiel E (2007). Antibody therapy in non-Hodgkin's lymphoma: The role of rituximab, 90Y-ibritumomab tiuxetan, and alemtuzumab. *Recent Results Cancer Res 176*, 153–163.
47. Waldmann TA (2007). Daclizumab (anti-Tac, Zenapax) in the treatment of leukemia/lymphoma. *Oncogene 26*, 3699–3703.
48. Ramirez CB, Marino IR (2007). The role of basiliximab induction therapy in organ transplantation. *Expert Opin Biol Ther 7*, 137–148.
49. Cardenas S, Auais A, Piedimonte G (2005). Palivizumab in the prophylaxis of respiratory syncytial virus infection. *Expert Rev Anti Infect Ther 3*, 719–726.
50. Scott DL, Kingsley GH (2006). Tumor necrosis factor inhibitors for rheumatoid arthritis. *N Engl J Med 355*, 704–712.
51. Hudis CA (2007). Trastuzumab—Mechanism of action and use in clinical practice. *N Engl J Med 357*, 39–51.

52. Scott DL, Kingsley GH (2006). Tumor necrosis factor inhibitors for rheumatoid arthritis. *N Engl J Med 355*, 704–712.
53. Pagano L, Fianchi L, Caira M, Rutella S, Leone G (2007). The role of Gemtuzumab Ozogamicin in the treatment of acute myeloid leukemia patients. *Oncogene 26*, 3679–3690.
54. Fietz T, Thiel E (2007). Antibody therapy in non-Hodgkin's lymphoma: The role of rituximab, 90Y-ibritumomab tiuxetan, and alemtuzumab. *Recent Results Cancer Res 176*, 153–163.
55. Scott DL, Kingsley GH (2006). Tumor necrosis factor inhibitors for rheumatoid arthritis. *N Engl J Med 355*, 704–712.
56. Chiang DT, Clark J, Casale TB (2005). Omalizumab in asthma: Approval and postapproval experience. *Clin Rev Allergy Immunol 29*, 3–16.
57. Smith S, Sweetenham JW (2007). Iodine 131 tositumomab in the treatment of non-Hodgkin's lymphoma. *Future Oncol 3*, 255–262.
58. Davies AJ (2007). Radioimmunotherapy for B-cell lymphoma: Y90 ibritumomab tiuxetan and I(131) tositumomab. *Oncogene 26*, 3614–3628.
59. Shear NH, Langley RG, Ho V (2006). Efalizumab, a reversible T-cell modulator for psoriasis. *J Cutan Med Surg* (Suppl 1), 4–9.
60. Langley RG, Cherman AM, Gupta AK (2005). Alefacept: An expert review concerning the treatment of psoriasis. *Expert Opin Pharmacother 6*, 2327–2333.
61. Galizia G, Lieto E, De Vita F, Orditura M, Castellano P, Troiani T, Imperatore V, Ciardiello F (2007). Cetuximab, a chimeric human mouse anti-epidermal growth factor receptor monoclonal antibody, in the treatment of human colorectal cancer. *Oncogene 26*, 3654–3660.
62. Grossi F, Aita M (2007). Bevacizumab and non-small-cell lung cancer: Starving the enemy to survive. *Expert Opin Biol Ther 7*, 1107–1119.
63. Sweet BV (2007). Natalizumab update. *Am J Health Syst Pharm 64*, 705–716.
64. Narayanan R, Kuppermann BD, Jones C, Kirkpatrick P (2006). Ranibizumab. *Nat Rev Drug Discov 5*, 815–816.
65. Blick SK, Keating GM, Wagstaff AJ (2007). Ranibizumab. *Drugs 67*, 1199–1206.
66. Saltz L, Easley C, Kirkpatrick P (2006). Panitumumab. *Nat Rev Drug Discov 5*, 987–988.
67. Cohenuram M, Saif MW (2007). Panitumumab the first fully human monoclonal antibody: From the bench to the clinic. *Anticancer Drugs 18*, 7–15.
68. Parker CJ, Kar S, Kirkpatrick P (2007). Eculizumab. *Nat Rev Drug Discov 6*, 515–516.

# COMBINATORIAL COMPOUNDS AND DRUG DISCOVERY

James C. Bigelow
*Department of Biomedical and Pharmaceutical Sciences, College of Pharmacy, Idaho State University, Pocatello, Idaho*

## 1 INTRODUCTION

### 1.1 Overview

***What Is a Drug?*** The U.S. Food and Drug Administration (FDA) defines a drug as: "(A) articles recognized in the official United States Pharmacopeia, official Homeopathic Pharmacopeia of the United States, or official National Formulary, or any supplement to any of them; and (B) articles intended for use in the diagnosis, cure, mitigation, treatment, or prevention of disease in man or other animals; and (C) articles (other than food) intended to affect the structure or any function of the body of man or other animals; and (D) articles intended for use as a component of any articles specified in clause (A), (B), or (C)" [http://www.fda.gov/opacom/laws/fdcact/fdcact1.htm, Chapter II—Definitions, (g)(1)].

While this rather broad definition serves well for regulatory purposes for this chapter, the term *drug* must be narrowed to that more appropriate to the disciplines of pharmacology, pharmacy, and medicine. In this context, we will define a "drug" as a specific chemical species able to cause specific changes in the functioning of living things. More specifically, we will further narrow the definition to relatively small (<1500 Da) organic and occasionally inorganic compounds. This narrowed definition is not meant to indicate that the many macromolecular agents such as polyclonal or monoclonal antibodies or enzymes are of less value therapeutically, but merely recognizes the great difference between macromolecules and smaller discrete organic or inorganic molecules.

***Pharmacokinetics, ADME, and Drug Properties*** Drugs, essentially by definition, are chemicals that bring about specific, often very specific and very potent, biological responses in their target organisms. However, it is insufficient for a chemical to be either potent or specific. Drugs must have other properties. When a drug is administered, it will follow a reasonably predictable course of ***a****bsorption* into the body, ***d****istribution* throughout the body, possible enzymatic ***m****etabolism* of the drug, and ultimate ***e****xcretion* (ADME). These processes, together with drug ***t****oxicity*, constitute the common pharmaceutical acronym, *PK/ADMET*, with PK standing for *pharmacokinetics*, which is to say the time course of a drug into, through, and eventually out of the body; it is what the body does to the drug (versus *pharmacodynamics*, which is what the drug does to the body).

***Polarity and Drug Properties*** Pharmaceutical companies require certain properties of the drugs they discover, develop, and manufacture. Most importantly, drugs must be safe and effective. This is a trade-off. For some life-threatening diseases such as cancer or an acute serious infection, it is more important that a drug be effective than safe. On the other hand, drugs used very widely such as over-the-counter pain relievers or drugs taken regularly over years or decades for chronic conditions such as hypertension or osteoporosis are held to a much higher standard of safety.

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

Beyond safety and efficacy, drug manufacturers desire that drugs be chemically stable, reasonably inexpensive, and easy to manufacture (this is often a fairly minor consideration since production costs often represent only a small fraction of the consumer costs of a drug), have favorable pharmacokinetic and metabolic profiles, and, importantly, be soluble to some extent in water or some aqueous formulation.

This aspect of drugs is dictated by the fact that they must function in living organisms, which are overwhelmingly composed of water. While there are many means of drug administration, the two most important ones are intravenous injection and most importantly, oral administration. For intravenous injection, a drug must either be soluble in a water-based formulation or must be able to be formulated as a stable suspension or colloid. For drugs taken orally, they ideally should have moderate solubility and be of moderate polarity. Highly polar molecules are generally very water soluble but pass very poorly through biological membranes via passive diffusion while very nonpolar molecules pass readily through membranes but are essentially insoluble in the fluids of the stomach and intestines where drug absorption occurs.

## 1.2 Drug Discovery and the *Pipeline*

Modern drug discovery is an extremely expensive and very long process with some sources stating that it takes an *average of* $800 million and 11–15 years to move a drug through ultimate FDA (in the United States) approval and final marketing. (There is some controversy about the numbers involved, but it is certainly both long and expensive.) The drug discovery and development process is illustrated in Figure 1 and is often referred to in the pharmaceutical industry as the *pipeline*.

The pipeline usually starts with identifying an appropriate disease state on which to focus corporate resources (itself a very complex process) and find a corresponding *target* for that disease state. The target might, for instance, be an important enzyme present in a pathogenic bacteria or a receptor associated with a neurological disorder. An *assay* is developed, usually an in vitro assay, that will allow many synthetic or natural product chemicals to be tested or *screened* until a fairly small number of *hits* are found to have favorable properties in terms of potency and selectivity. At this point, however, it is unlikely that any one of those hits will have appropriate PK/ADMET properties in and of themselves. Thus, they will have to be chemically modified to improve PK/ADMET and other properties (ease and cost of synthesis, chemical stability, etc.). This process converts hits into *leads*, or compounds that are likely to be good drugs. The leads are evaluated against a large battery of tests, both in vivo and in vitro, such as toxicity in animals, likelihood of crossing the gut wall (for oral drugs), the *blood–brain barrier* (BBB) (a specialized brain capillary structure that prevents most chemical species in the blood such as drugs and many endogenous species to reach the brain while still allowing essential species such as glucose to reach the brain). This is generally an iterative process and is nearly always very expensive and time consuming. Eventually a single lead structure will be identified as a drug *candidate* that will be moved forward into human clinical trials. Thus, the whole process of drug discovery and development involves a constant winnowing, wherein many, sometimes millions, compounds are tested and only one or a very few emerge as approved drugs.

***Sources of Drugs*** Drugs are used everywhere in our society and throughout the world. They have been used in one form or another for all recorded history (as herbs or herbal preparations). Their uses range from legitimate medical use (where the pharmaceutical industry prefers the term "medicine," believing it to have a less pejorative connotation) to illegal "recreational" use, as well as veterinary, animal husbandry, and sports enhancement (both legal and illegal) uses. Currently, small-molecule drugs can be broadly classified as (1) natural product drugs, (2) synthetics drugs, and (3) semisynthetic drugs. Natural product drugs were the first drugs used. This was well before the concept that dis-

**Figure 1** *Pipeline* of drug discovery starts with identifying a disease-related *target* for which a quantitative assay can be developed such as a relevant enzyme or receptor. A bioassay is developed and a great many chemicals are assayed until one or more *hit* chemicals are found, which is to say a chemical structure that is likely to be both potent and selective as a drug. These compounds can be synthetically modified to improve PK/ADMET properties to generate *lead* structures, only a very few of which will be chosen as drug *candidates* to be advanced into human clinical trials.

crete chemicals even existed, and they continue to be used in impure preparations by indigenous healers worldwide. Natural product drugs are produced by various organisms such as microbes and fungi but also plants and animals with a more recent focus on marine organisms. Synthetic drugs are produced wholly by chemical synthesis, and semisynthetic drugs represent a fusion of the two, as either a synthetic chemical modification of an existing natural product or a de novo chemical synthesis of a current natural product or of a related structure.

***Natural Product Drug Discovery*** Natural products and their derivatives are the single most important source of anticancer agents. Natural product drug discovery programs reflect the philosophy that chemical compounds derived from nature "far exceed the imagination and synthetic capabilities of medicinal chemists," a view born out by the great structural complexity and wide range of mechanisms of action found among these agents [1]. To date, more than 40% of all drugs are either natural products or semisynthetic derivatives of natural products. Indeed, 61% of small-molecule *new chemical entities* (NCEs, an active chemical species that has never been previously approved in any form) introduced as drugs from 1981–2002 can be traced to natural products. For certain categories this is higher such as for anticancer agents where it is 74% and antibacterial agents where it is 79% [2].

Natural products exhibit truly remarkable diversity in chemical structure. The overwhelming majority of natural products with pharmacological activity are *secondary metabolites* (SeMes) [3, 4]. Secondary metabolites are cellular chemical products with no recognized essential metabolic function in growth or reproduction, but they play important roles in the ecological interactions of an organism and are probably ubiquitous in nature [5, 6]. Rather, secondary metabolites are expressed during times of microbial stress and are thought to provide the microbe protection from predators, assisting the microbe by protecting a nutritive source from foreign microbial invasion as well as playing many other important biological roles.

Secondary metabolites can be grouped into large chemical classes. Cavalier-Smith [7] distinguishes 13 major classes of SeMes derived from primary metabolic pathways (acetyl-Co-A, amino acids, purines, etc.), which are distinct and specifically distributed among 8 major phylogenetic clades. A similar classification is found in antibiotic classification with 10 antibiotic families and approximately 500 discrete groupings [8]. In a broader natural product classification of 17 classes over 500 subgroupings have been defined [9]. Classification of antibiotics and natural products reflect a compromise in classifying a great array of chemical species according to their primary metabolic source, evolutionary development, and chemical features.

Secondary metabolites are synthesized by specialized, well-coordinated multigene systems that are often clearly derived from some past primary metabolic systems in response to prolonged evolutionary pressure [10]. Because of these features, production of any given SeMe exhibits very narrow taxonomic distribution with a given chemical substance rarely being produced by more than a few related genera and frequently by only a single species or strain in a genus. This narrow taxonomic range has driven drug discovery into ecosystems that exhibit great and/or novel biodiversity. Thus, one sees efforts in tropical and temperate rainforest, caves, coral reefs, other exotic or "extreme" ecosystems. For instance, the very important anticancer agent paciltaxel was isolated by Wall and Wani in 1967 from the bark of the Pacific yew tree (*Taxus brevifolia*), which grows in North American temperate rainforests [11].

Discovery of natural problem drugs is very difficult. The isolation procedures often are tedious, labor intensive, poorly amenable to automation, and highly customized for each natural product. They ultimately depend on bioassays. Such assays require biological activity be identified in an iterative fashion with multiple separation fractions and the assays are often slow and show poor reproducibility. Also, natural products must be isolated from biological material. This often does not represent a problem for antibiotics or other agents that can be grown in very large amounts in fermentation vats but may be very problematic in other cases where the source material may be present in very limited quantities. Indeed, there was concern that isolation of paciltaxel in sufficient quantities to treat large numbers of patients would drive the Pacific yew tree to extinction (paciltaxel is no longer obtained from the yew tree). Also, it can take very large quantities of starting material to obtain small quantities of active material. Again, paciltaxel is illustrative; it took almost 1200 kg of bark to yield 10 g of pure material [12]. Accessing biodiverse sites in development nations can also be problematic. Developing, and indeed most, nations are justifiably very protective of their biodiverse heritage and fear that large American and European pharmaceutical companies may exploit the genetic and biodiversity without compensation. The UN Convention on Biodiversity restricts access to areas such as tropical rainforests without appropriate negations and agreement [13].

Even after biological material is obtained and a pure compound is isolated, elucidating the full chemical structure is not trivial even with high-resolution mass spectroscopy and advance multidimensional nuclear

magnetic resonance (NMR). Even with a fully elucidated structure in hand, there is another serious problem: *dereplication*. Dereplication refers to identifying and eliminating from any further drug discovery activities active agents that have been previously identified. This is a key step in natural product drug discovery. As early as 1985, Berdy reported more than 95% of chemicals that were isolated from surface-soil microbes had been previously recorded [14]. Despite the great success of natural products as drugs, given all the above limitations, natural products began to lose their allure for pharmaceutical companies.

***Traditional Synthetic Drug Discovery*** Historically, the birth of drug discovery dates to 1846 when the explosive and antianginal agent nitroglycerin was first synthesized by the Italian chemist A. Sobrero by adding glycerol to a mixture of concentrated nitric and sulfuric acids. He observed that a small amount placed on the tongue gave a severe headache. In 1847 C. Hering developed the sublingual dosage form. Subsequently, in 1857 an Edinburgh physician, T. L. Brunton, administered amyl nitrite, a known vasodepressor, by inhalation and noted transient anginal pain relief. In what may be the first instance of true medicinal chemistry, W. Murrell deduced that the action of nitroglycerin mimicked that of amyl nitrite and helped establish the use of sublingual nitroglycerin for relief of the acute anginal attack. Ironically, Alfred Nobel developed dynamite in the 1860s by combining liquid nitroglycerin with an inert porous material. Later he would use nitroglycerin for anginal pain. Nobel died in 1896 of a cerebral hemorrhage and left the bulk of his fortune in trust to establish the Nobel Prizes. In 1998, the Nobel Prize in Medicine or Physiology was awarded to Furchgott, Ignarro, and Murad for recognizing the importance of nitric oxide (NO), which derives from nitroglycerin, as a signaling molecule. After its early beginning, synthetic chemical drug discovery became a fully developed arm of drug development. It drew heavily from the coal gas (and related coal tar) industry and the dye industry, particularly in Germany. It had a great (although very checkered) success in development of the first antibiotics in the 1930s, the *sulfa* or *sulfonamide drugs,* which well predated the use (but not discovery) of penicillin.

Synthetic drug synthesis is, of course, a subset of organic chemical synthesis. That should be well known to any person who has sat through an organic chemistry lecture or laboratory classes. As a means of making and discovering drugs, traditional chemical synthesis has many of the same limitations of natural product chemistry. As with natural product chemistry, even with modern instrumentation, it remains time and labor intensive, is generally poorly amenable to automation, and is also highly customized. While some parallel synthesis is common, many compounds must be made essentially "one at a time." As such, the number of chemical species that can be made, purified, and verified to have the desired structure remains limited. As with natural product isolation, synthetic chemists are highly experienced and often highly specialized, having focused much of their careers on synthesizing a limited number of molecular scaffoldings. These difficulties of both traditional synthetic chemistry and natural product chemistry provided a strong incentive to the development of combinatorial chemistry in the 1990s.

***Combinatorial Drug Discovery*** The basic concept of combinatorial chemistry is straightforward: A large (often very large) number of structurally different, but chemically related species are synthesized essentially simultaneously. To accomplish this, very similar or identical reaction conditions and reaction apparatus must be used. To perhaps better understand this, one can look back to the 1960s when Merrifield developed solid-phase peptide synthesis, an important antecedent of combinatorial chemistry [15].

Briefly, In Merrifield's *solid-phase peptide synthesis* (SPPS), fairly large (20–50 μm), chemically inert resin beads have the C-terminal amino acid residue of a nascent desired peptide chain covalently bound via a linker to the bead (beads with appropriated prelinked residues are readily commercially available). The α-amino group is protected with a labile group [commonly *tert*-butyloxycarbonyl (*t*-Boc) or 9-flourenylmethloxycarbonyl (Fmoc). The amine can be made available for further chemical coupling reactions with other residues by deprotection. The beads are loosely packed into small glass columns. A series of organic solvents, with or without subsequent residues or reactants, are pumped through the column in a sequential fashion in a series of steps as illustrated in Figure 2.

1. *Deprotection* The amine-protecting group is removed via a chemical reagent exposing the reactive amine.
2. *Wash* The deprotecting reagent (and removed protecting groups) is washed out of the column.
3. *Coupling* An amine(s)-protected residue is cycled through the column and a coupling reaction occurs between the amine of the unprotected immobilized reside and the carboxyl moiety of the amine-protected free residue. Steps 1–3 are repeated until the entire desired peptide is completed.

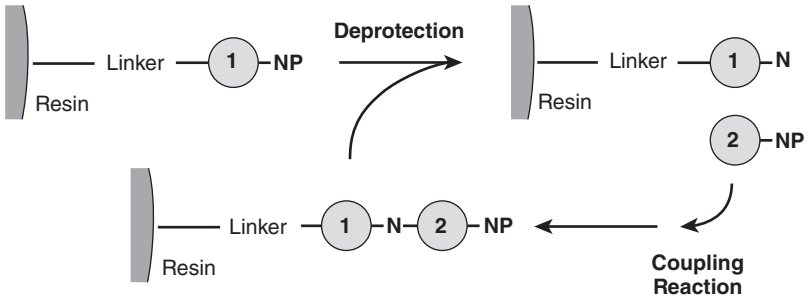


**Figure 2** In Merrifield's solid-phase peptide Synthesis (SPPS), an amino acid residue (represented as a circle) is bound covalently to a solid support via a linker. The α-amino group (N) is protected (represented as NP). The protecting group is removed chemically and a second protected residue is allowed to react and couple. The protecting group of this residue can then be removed and a third protected residue can be added. The process is repeated until the desired peptide is made, after which it can be cleaved from the resin, purified, and characterized.

4. *Cleavage, Deprotection, and Purification* At the end of synthesis, the peptide is cleaved from the bead, all amines are deprotected, and the peptide is purified, often by semipreparative high-performance liquid chromatography (HPLC).

Merrifield's SPPS (and other procedures like it) can only make one peptide at a time. However, in combinatorial chemical synthesis, a great many different (in this case peptides) chemical species can be made essentially at once. For instance, instead of just starting with beads that had just one residue, the chemist could start with a selection of beads that had all 20 natural occurring amino acids immobilized on them. In turn, these beads could be exposed to, and reacted with, all 20 amino acids at each cycle of the growing peptide. At the end one would have, at least in principle, all possible peptides of the given length.

Combinatorial chemistry (the term was first coined in the early 1990s, although research in combinatorial chemistry had been ongoing at Glaxo Wellcome Inc. under the direction of H. M. Geysen since 1984) did at first primarily focus on peptide and oligonucleotide *chemical or compound libraries.* A chemical or compound library is a collection, often very large, of stored chemicals usually used for drug screening purposes (and also other industrial purposes such the search for better catalysts). The library can consist of both synthetic and pure natural products or plant, microbial, or animal extracts and should always have a database with critical information such as quantity, structure, and purity associated with it.

The methods used for combinatorial synthesis are diverse and no longer limited to peptides or oligonucleotides. However, the concept remains the same; simultaneous synthesis of many new chemical species. Combinatorial chemistry gave a promise to the pharmaceutical industry of providing many NCEs to reduce the time and costs required for drug discovery.

## 2 CHEMICAL DIVERSITY

Chemical diversity represents the idea that, for new drug discovery, one should examine a great many small organic molecules that are different in as many chemical properties as possible. It is based on the observation that structurally similar compounds exhibit similar physicochemical and biological properties. Thus, if one wishes to discover new drugs with novel mechanism of action, one must examine as many structurally diverse molecules as possible.

### 2.1 Permutations, Combinations, and Possibilities

The concept, while simple in principle, is daunting if we actually examine how many molecules are indeed possible. To do this we need to look at the permutations, combinations, and possibilities available in small organic molecules.

Each different arrangement of all or part of a set is a permutation. The number of *permutations* of $n$ things taken $r$ at a time is

$$nPr = \frac{n!}{(n-r)!} \tag{1}$$

For instance, if one wants to determine the number of ways to arrange three amino acids, such as aspartic acid (Asp), phenylalanine (Phe), and alanine (Ala) in a

**TABLE 1 Combinations, Permutations, and Possibilities for Peptides of Various Lengths Made from the 20 Naturally Occurring Amino Acids**

| | Peptide Length | | |
|---|---|---|---|
| | 5 | 10 | 15 |
| Combinations | $1.6 \times 10^4$ | $1.9 \times 10^5$ | $1.6 \times 10^4$ |
| Permutations | $1.9 \times 10^6$ | $6.7 \times 10^{11}$ | $2.0 \times 10^{16}$ |
| Possibilities | $3.2 \times 10^6$ | $1.0 \times 10^{13}$ | $3.2 \times 10^{19}$ |

dipeptide, where order *does* matter and no amino acid can be repeated, there would be six permutations.

Each arrangement of all or part of a set without regard to order is a combination permutation. The number of *combinations* of $n$ things taken $r$ at a time is

$$nPr = \frac{n!}{r!(n-r)!} \tag{2}$$

For the same example, the number of ways to arrange the three amino acids in a dipeptide where order *does not* matter and no amino acid can be repeated would give three combinations. Of course, in biology order definitely does matter. One can note, for instance, that aspartame, the artificial sweetener, is the methyl ester of the dipeptide of the amino acids aspartic acid and phenylalanine, while the corresponding ester of the dipeptide of phenylalanine and aspartic acid is not a sweetener. Finally, all possible outcomes where a unit such as an amino acid can be repeated, which is to say the number of possibilities of $n$ things taken $r$ at a time is, $n^r$.

Consideration of this leads to insight on chemical diversity. For short peptides, much less proteins, made with the 20 naturally occurring amino acids, the numbers can rise very quickly, as seen in Table 1 wherein are shown the combinations, permutations, and possibilities for short peptides of various lengths. The task of examining chemical diversity is daunting indeed.

### 2.2 Chemical Space and Molecular Descriptors

Chemical space is a theoretical concept wherein all possible chemical species occupy an $n$-dimensional topography. Each "dimension" is represented by a *molecular descriptor*, which is quantitative values that transform the chemical information of a molecule into a number. These can be obtained from experimentally derived values such as $pK_a$, polarity (a log P value is the logarithm of the partition coefficient of a molecule between $n$-octanol and water and the M log P value refers to the experimentally measured value versus the calculated or C log P value), and molecular weight and the like. These descriptors represent properties of complete molecules. Other descriptors can be derived from various algorithms that perform two- or three-dimensional structural analysis of molecules. The former would include simple counts of molecular features such as the number and types of ring systems, the number of rotatable bonds, and so forth. An example of the latter would be spatial relationships between atoms, rings, and other planer structures. This requires generation of three-dimensional conformational models and generally must take into account molecular flexibility. Such modeling, particularly with the large data sets, can be expensive and is very computationally intensive. Any given type of descriptor (experimental, or two- or three-dimensional) is unlikely to be useful alone in drug discovery, and thus they need to be used in combination. In the context of drug discovery, the desire is to gain insight into how chemically diverse a given library is. A library is mapped into chemical space using its molecular descriptors, and it is generally highly desirable to have a very diverse library, which is to say one that occupies a large volume of chemical space. In three dimensions, a library can be visualized, and at least a qualitative appreciation of the chemical diversity of that library can be obtained. Figure 3 shows a three-dimensional expression of chemical space wherein is illustrated a diverse (Fig. 3B) and less diverse (Fig. 3A) library. Beyond three dimensions, measures of the chemical diversity of a library cannot be visualized, and various computational algorithms must be employed to give a measure of that diversity.

### 2.3 Lipinski's Rule of Five

Lipinski's rule of five is a specific set of molecular descriptors used to evaluate if a given chemical compound has properties that would make it likely to be an orally active drug in terms of PK and ADME properties but says nothing about the pharmacological activity of a chemical. It was based the observation of Christopher Lipinski that many drugs were relatively small lipophilic molecules with certain particular chemical properties [16]. In drug development, when a pharmacologically hit or lead structure needs to be optimized for PK and ADME properties, the rule is very useful and states that, in general, an orally active drug should have:

1. No more than 5 hydrogen bond donors
2. No more than 10 hydrogen bond acceptors
3. A molecular weight under 500 g/mol
4. A log P value of less than 5

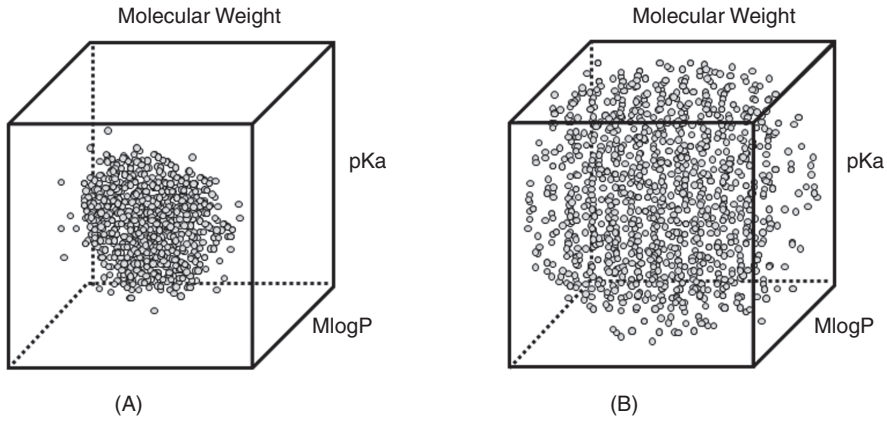


**Figure 3** Chemical space is a theoretical concept. It represents the "space" that can contain all possible chemically stable small organic molecules. Molecules are mapped into chemical space by various molecular descriptors. The degree to which molecules in a chemical library spread through chemical space gives an indication of the diversity of that library. There can be many descriptors. In the example above only three descriptors are used: molecular weight, p$K_a$, and measured lop P (M log P). The library in (B) has greater chemical diversity than the library in (A) and would be a better library for a drug search because of this.

## 3. PRINCIPLES OF COMBINATORIAL CHEMISTRY

As noted previously, combinatorial chemistry is a means by which large numbers of structurally different compounds are made at a given time. The general goal is to generate a very large chemical library for drug discovery with the hope (often not realized) to cover a large region of chemical space. To accomplish this, many simultaneous chemical synthetic reactions must be done. This requires identical reaction conditions, including temperature, reagents, and solvents for many synthetic reactions that are being done in parallel. The principles of combinatorial chemistry are best illustrated with the following example.

### 3.1 Split-Pool Combinatorial Synthesis

There are many combinatorial methods including those that are proprietary. To illustrate the power of combinatorial chemistry, we will focus on a single method, *split-pool combinatorial synthesis*. Split-pool synthesis is a solid-phase method and is one of the best methods rapidly synthesizing large number of compounds. In split-pool synthesis, a large library can be made quickly from a relatively few initial chemical building blocks, as illustrated in Figure 4. In the figure, three different building blocks (X, Y, and Z) are immobilized to a resin bead in a manner analogous to Merrifield's solid-phase peptide synthesis. Following immobilization, the beads are pooled then split into three (or as many building

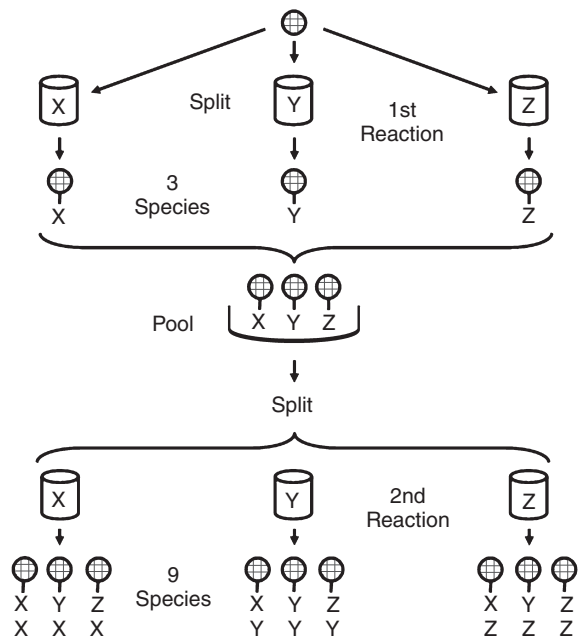


**Figure 4** Split-pool combinatorial chemistry. In this combinatorial method, initial combin-atorial building blocks, X, Y, and Z in this case, are first immobilized on a solid support (resin beads) by the first chemical reaction in separate reaction vessels. The beads are then pooled, thoroughly mixed, and then split into new reaction vessels. To these reaction vessels are added the building blocks, which are allowed to couple to the previous building blocks, thus forming a growing chain. This process can be repeated through many cycles. The final number of theoretical individual chains can be very large. For a 10-mer polypeptide it would be approximately $1.0 \times 10^{13}$ unique peptides.

blocks as are being utilized) and reacted again with the building blocks. By this point there are now 6 unique products. After the next round of pooling, splitting, and reacting, there are 27 unique products. The cycles can be repeated as desired. In principle, all possible products of a given length, as determined by the number of cycles, should be synthesized. For a 10-mer with 15 building blocks, this would be $15^{10}$ ($n$ things taken $r$ at a time or $n^r$) or approximately $6 \times 10^{11}$ products; a truly enormous number.

The method has a number of limitations, the most important of which is determining which structure is on which bead; this will be addressed later in this chapter in the discussion of encoding. Depending on the synthesis, it can be irreproducible and have limited yield. Further, since no given coupling reaction is ever 100% efficient, every bead will have a heterogeneous mix of products representing the desired product as well as products where one or more coupling steps failed. For a 10-step synthesis with a 95% reaction efficiency, the final yield will be only approximately 60%. Thus, any observed bioactivity in a screen may not be in fact the result of the desired product but of one of the products where one or more coupling steps failed. Finally, because the building blocks must be synthetically compatible and thus probably structurally similar, the number of products produced may be very large, but the actual chemical diversity of the library may actually be quite limited.

### 3.2 High-Throughput Screening

The development of combinatorial chemistry was accompanied by the development *high-throughput screening* (HTS) techniques in the pharmaceutical industry. Clearly, if it is possible to make tens of thousands of compounds but there is no means to test them quickly and relatively inexpensively for bioactivity, combinatorial chemistry becomes an exercise in futility. High-throughput screening allows for tens to hundreds of thousands of small-molecule chemicals to be tested in days or weeks against a bioassay. High-throughput screening is characterized by high sample throughput and reduction of assay size so that assays can be run in microtiter plates. Microtiter plates are flat plates, usually of some form of plastic, and disposable. They have multiple wells that function essentially as very small test tubes. For HTS, the plates typically have 96, 384, or even 1536 sample wells arranged in a 2:3 rectangular array (6- and 24-well plates are also available and have many experimental uses). The wells have sample volumes that range up to several hundred microliters. In HTS, robotics and near to full automation is necessary for sample handling and sample assay. Near continuous operation is also highly desirable. Because of the amount of data HTS produces, most data processing and archiving are also highly automated. It is beyond the scope of this chapter to discuss HTS in detail, but it is important to note that HTS is a billion-dollar industry and represents a sea change in the means through which drugs are now discovered.

## 4 SOLID-PHASE SYNTHESIS

As noted previously, solid-phase combinatorial synthesis can be traced back to Merrifield's research in the 1960s where peptides, immobilized on resin beads, could be generated by sequential step-by-step reactions. In contrast to conventional organic liquid–liquid synthesis, reactant and solvents are readily removed by filtration, decanting, centrifugation, and the like. While initially applied to peptides, solid-phase synthesis (SPS) was then used for synthesis of other directionally aligned (i.e., 3′ to 5′ for oligonucleotides) biopolymers such as deoxyribonucleic acid (DNA), ribonucleic acid (RNA), and related polymers such as phosporothioate oligonucleotides that are used for antisense chemotherapy. It was later used in combinatorial chemistry and has been largely responsible for the great growth in the field for both development of new drugs but also for synthesis of nondrug species such as catalysts.

The basic scheme in SPS is illustrated in Merrifield's method. Building blocks of a desired polymer such as amino acids with at least two functional groups are used. A starting building block is covalently linked to a solvent-insoluble polymeric support, which will withstand the synthetic reaction conditions. The binding is done via a linker that optimizes the reaction, for instance, by providing greater access to solvents and reactants to the growing polymer chain. The linker must also be such that it can be selectively cleaved from the resin at the end of the synthesis.

Of the two functional groups, one must usually be protected by a protecting group. At the start of the reaction, the protecting group of the first building block is removed, excess reactant is washed away, and the second building block is introduced and allowed to couple the first building block. The coupling reagents and reactants are washed away and the process is repeated until the desired polymer is completed. The polymer is then cleaved from the resin, purified, and characterized. Of note is that at each step of the polymer formation, errors will be introduced. As noted before, even under optimized reaction conditions, any given building block may fail to successfully couple to a previous one.

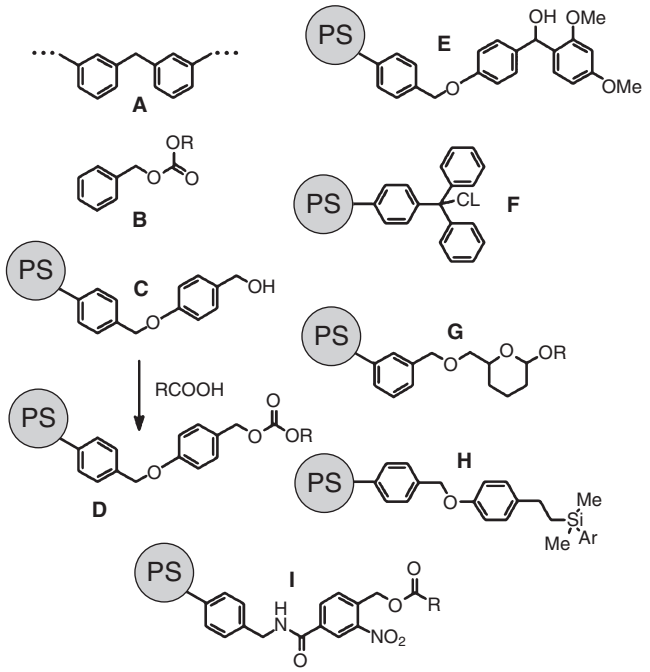


**Figure 5** Chemical structures associated with linkers for solid-phase combinatorial chemistry: (A) polystyrene, (B) amino acid reacted with chloromethylbenzene, (C) and (D) wang linker (C) before and (D) after attachment of the first building block, (E) rink linker, (F) trityl linker, (G) tetrahydropyran linkers, (H) silicon linker, and (I) photocleavable linker.

### 4.1 Resins

***Polystyrene*** Polystyrene (PS) (polymerized styrene; Fig. 5A) is the most common resin employed in SPS. The SPS supports are small spheres that are reasonably rigid and resistant to compression (unlike, e.g., agarose beads), insoluble to reaction solvents, and chemically stable to the reaction and porous to allow for much greater synthetic surface area. They are also functionalized (typically with chloromethyl or aminomethyl groups) so that linkers and starting building block molecules can be covalently attached. Commonly, styrene is lightly cross-linked (1–2%) with divinyl benzene (PS-DVB) or oxymethylene to produce poly(styrene-oxyethylene) graft copolymers. Polystyrene resins swell considerably on exposure to many organic solvents that commonly used SPS. This opens up the structure of the bead and allows easier diffusion of reactants, which both speeds the reaction and increases yield. The major exceptions to this are polar solvents such as alcohols, which do not cause appreciable swelling. This lack of swelling does not allow reaction sites to be at a sufficient distance from a hydrophobic resin core of PS-DVB beads.

***Polyacrylamide*** Polyacrylamide supports have been utilized with the polar solvents commonly used in peptide synthesis. These resins have amide functional groups that are similar to the peptide bonds present in proteins.

***Poly(Styrene-Oxyethylene) Copolymers*** Poly(styrene-oxyethylene) graft copolymers are also used commonly in SPS, also with a view to allow synthesis with more protic solvents. They are marketed under the name TentaGel and consist of polyethylene glycol attached via an ether linkage to cross-linked polystyrene. This class of resin swells in a wide variety of polar and non-polar solvents and is available commercially with a wide variety of linker groups. The polyethylene glycol gives the resin a polar character, but they retain the favorable features of polystyrene. The polymers have long flexible ethylene oxide chains branching off from the polystyrene backbone, which allows the synthetic reaction site to be distant from the hydrophobic bead core.

### 4.2 Linking Groups

The final element in making a good SPS support resin is the linking group. Liking groups connect the combinatorial building block to the solid support. The linking group must not cleave prematurely during the steps of the synthesis and must protect the connecting functional group(s) of the building block. Only at the end of the synthesis must the linker be cleaved, preferably easily and at high yield. The linking group should also allow ready attachment of the building block. Many linkers (at least 200) have been devised over the years. Given this, no attempt will be made at completeness and only a few very important or interesting linkers will be discussed here. Guillier et al. have produced an excellent and comprehensive review and should be consulted by interested readers [17].

***Strong and Mild Acid-Cleaved Linkers*** Carboxylic acid linkers were first used by Merrifield. Merrifield's resin beads were PS with a nitrated chloromethylpolystyrene. The free carboxylic acid of an *N*-protected amino reacted at the chlorine, giving a stable ester bond linkage of the first amino acid residue to the resin. Figure 5B shows the result of a reaction of an amino acid residue to chloromethylbenzene. At the end of the synthesis, HF was commonly used for cleavage.

Use of HF is very hazardous, although it does give good yields and purity. Partially to avoid the use of HF, a number of other methodologies were developed that required less acidic (relatively) conditions. Ease of cleavage of the linker under conditions of acid cleavage is related to the stability of the carbocation that is

formed upon cleavage. It was reasoned that addition of groups able to donate electrons would decrease the acid strength necessary for cleavage. This resulted in the development of the Wang and SASRIN (superacid sensitive resin) and Rink linkers [18–20, respectively].

Wang linkers are generally used with PS, TentaGel, and polyacrylamide resins. With its activated benzyl alcohol design, cleavage can be done in a less acidic environment using trifluoracetic acid (TFA; 50% in $CH_2Cl_2$ for 30 min). The Wang linker is the standard support for much peptide synthesis. Figures 5C and 5D show the Wang linker before (Fig. 5C) and after (Fig. 5D) attachment of the first building block or residue. The SASRIN linker is structurally very similar to the Wang linker but is made more acid-labile by the addition of one or more methoxy groups to enhanced stabilization of the carbonium ion formed during cleavage and can be cleaved with 1%–5% TFA. Finally, introduction of alkoxy groups gives the Rink linker (Fig. 5E), which can be cleaved under very gentle conditions (10% acetic acid or 0.2% TFA).

***Trityl and Tetrahydropyran Linkers*** Trityl linkers (Fig. 5F) were initially developed in the 1970s [21, 22]. The trityl group is an acid-labile protecting group for many heteroatoms. It has been used to anchor one alcohol group of a diol and has been used to synthesize a library of β-mercaptoketones. In peptide chemistry, the problem of the formation of racemates can be avoided during addition of the first residue using these linkers. Additionally, cleavage with dry TFA, HBr, or HCl is not difficult. A hydroxyl linker based on the tetrahydropyran protecting group was developed by Thompson and Ellmann [23] (Fig. 5G). Alcohols readily couple to dihydropyran, and the resulting tetrahydropyran protecting group is stable to strong bases and nucleophiles. It is easily cleaved by employing TFA/water (95:5, 15 min) and produces reasonable yields (66 and 95%) for primary and secondary alcohols but poor yields (10%) for tertiary alcohols.

***Silicon Linkers*** Organic compounds incorporating silicon (organosilyl compounds) groups have become widely used since their introduction in the early 1970s. In combinatorial chemistry where there is a desire to make libraries of low-molecular-weight species that are not peptides, these linkers are useful. Although mainly used with heteroatoms, they can also be employed with carbon and are compatible with many synthetic methods. One such linker is shown in Figure 5H [24]. Because they do not require attachment via a carboxyl group as in peptide chemistry, they are sometime referred to by the unfortunate and inaccurate term as "traceless linkers" because the final compound may give no definitive "trace" to the linkage site on the solid phase. The availability of agents specific for silicon such as fluoride species also adds to their utility. Acidic cleavage of the C–Si bond is generally quantitative and selective.

***Photocleavable Linkers*** Using light as a means of bond cleavage (photolysis) is very attractive. In principle, it should be gentle, can be used at neutral pH, and offers a means of deprotecting functional groups independently of one another (orthogonal deprotection). Photolysis has been widely used with biopolymers such as peptides, carbohydrates, and oligonucleotides. Use in nonoligomeric combinatorial chemistry is limited to some extent because many small organic compounds of medicinal interest absorb light or are photolabile at the wavelength(s) needed for cleavage. Ideally, the wavelengths used should only be absorbed by the linker and not cause generalized photodegradation. One of photocleavable inkers is shown in Figure 5I. It was found to be suitable for the preparation of protected peptides and could be cleaved by irradiation at 350 nm in an orthogonal fashion without degradation of the aromatic amino acids [25].

## 5 ENCODING

The ability of combinatorial methods to synthesize many thousands of discrete compounds very quickly has, as in all things, both pluses and minuses. While a great many compounds can be made for subsequent HTS, the chemical structure of any hit must, of course, be able to be determined. In combinatorial chemistry, this is not a trivial problem. For instance, in split-pool synthesis, each resin bead will, in principle, have a single unique compound attached to it but the structure of that compound will not be known. To determine that structure, various methods have been developed that allow the information relating to the actual synthesized species to be encoded on each bead.

### 5.1 Positional Encoding

Encoding can perhaps best be illustrated with an example. Positional encoding was first used by Geysen et al. in combinatorial synthesis [26]. The solid-phase support was an array of pins fixed in a simple matrix on a polypropylene rack. Unlike split-pool synthesis, the pins were not randomized, and thus the actual physical position of each pin served as the code for each compound. Thus, if pin 1 went through a series of steps, components A, B, C, and so on were added in sequential order, whereas pin 2 had A, C, and B added in that order, then after cleavage, the product from pin 1 should

be A–B–C while that from pin 2 should be A–C–B (provided in each case the synthesis proceeded correctly). The cleaved products from each pin would then be screened and any hit would be resynthesized as conformation. This method is one of the oldest and still the most widely used in automated robotic synthesizers. However, it has a number of limitations, the greatest of which is that only a relatively few number of compounds can be made owing to space limitations. While first employed with peptide libraries, this type of encoding is the method currently used for DNA microarrays where it has found its greatest utility.

### 5.2 DNA-Based Tagging

DNA-based tagging or encoding employs a strategy of dual synthesis on a support. The support resin has linkers on to which both an oligonucleotide and the combinatorial compound can be built. Orthogonal protecting groups allow a DNA tag to be made while the combinatorial compounds are also made. A polymerase chain reaction (PCR) primer is used as the linker between the support matrix and the DNA tag. With each new synthesis step, a new nucleotide is added to the DNA tag. At the end of the synthesis, a second PCR primer is added to the tag. After bioassays identify hits, the tag can be enzymatically cleaved and the tag can be amplified and sequenced. Because sequencing DNA is a very mature technology, the DNA sequence can be used to encode the combinatorial structure. Thus, if at the first split-pool step, components X were paired with adenosine [as adenosine 5′-triphosphate (ATP) for instance], Y with guanine, and Z with cytosine, every bead with an X would have a corresponding A, Y would have a corresponding G, and Z would have corresponding C. As the split-pool process continued, the growing oligonucleotide sequence would reflect the growing combinatorial sequence. If more than four possible combinatorial components are desired, multiple nucleotides can be incorporated with each reaction step. This method and all chemical tagging methods have the disadvantage that all bioactivity may result from the tag or some synergistic effect between the tag and the combinatorial compound.

### 5.3 Peptide Tagging

Peptides can also be utilized as tags in a way very analogous to DNA-based encoding with the tag sequences that can be decoded via Edman sequencing, and such an encoding strategy was developed by Kerr et al. at Chiron [27]. A lysine residue attached via a Rink amide linker to the solid support served as a bifunctional linker. Parallel synthesis of the combinatorial compounds and the coding peptides was possible by using orthogonal protecting groups. A base-labile Fmoc group was used on the α amine for the compound peptide while the mildly acid-labile protecting group, α, α-dimethyl-3,5-dimethoxybenzyloxycarbonyl (Ddz), was used on the ε amine. Four amino acids in a sequence of three residues ($4^3$ or 64 possible codes) were used to code for each combinatorial building block. After cleavage and identification of hits, sequencing allowed the hit structure to be decoded. The authors measured $IC_{50}$ (half maximal inhibitory concentration) bioactivity of both the product and tag to establish that the tag did not have activity. They tested the tag and product in combination and guarded against synergistic effects by noting the activity of product was similar to that of the product/tag combination.

### 5.4 Haloaromatic Binary Encoding

Binary encoding is an alternative to the oligomeric encoding of DNA-based and peptide tagging. Discrete chemical tags are used as binary "bits" to label each step of the combinatorial synthesis. This can perhaps best be understood by reference to Figure 6. In this example, pairs of tags are used to generate a series of binary identifiers. The presence of the first tag ($T_1$) and absence of the second tag ($T_2$) is encoded as 1, 0 and the presence of both $T_1$ and $T_2$ is encodes as 1, 1. New discrete chemical tags are incorporated at each step in the combinatorial cycle. At the end of the combinatorial synthesis, the tags are cleaved from the solid support and analyzed, for their presence or absence, to generate binary codes. This method has the advantage in that only a small proportion of a resin bead need be functionalized with a tag. However, the tags must be very chemically stable, not interconvertible, and readily and unambiguously detectable by appropriate means.

One method of binary tagging uses haloaromatic molecules (also called "hard" tags for their chemical stability). These tags were reported by Ohlmeyer et al. [28]. The tags were generated by the reaction of halophenols with bromo-alkanols. A great advantage of these tags was their ability to be separated via gas chromatography (GC) and detected with great sensitivity (0.1 pmol) with electron capture detection (ECD). This allowed the tags to be present at approximately 1% of the product, which greatly reduces the possibility of the tag interfering with the bioassay. One approach to enhance separation was to attach the tag via a photo-labile bond. After irradiation (366 nm), the released tags were chemically trapped via reaction with bis(trimethylsilyl)acetamide, which resulted in chemical derivatization of the tags (via reaction with the trimethylsilyl group, a common strategy used in GC to increase

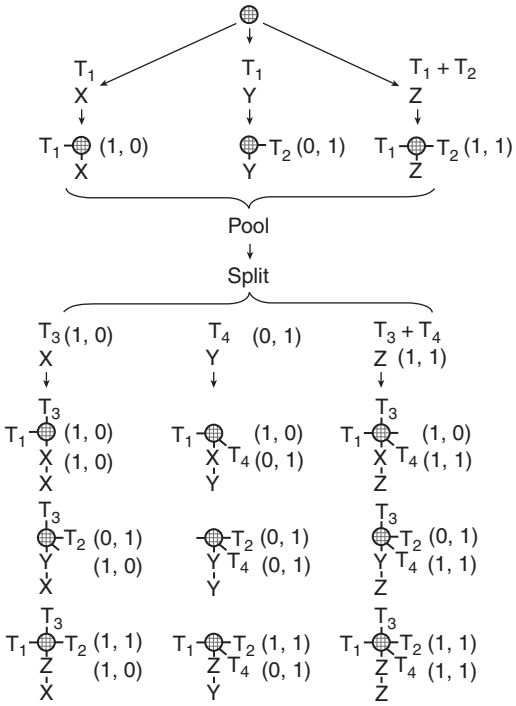


**Figure 6** Binary encoding with split-pool synthesis. In this example, split-pool synthesis proceeds through two cycles generating nine products from three (X, Y, and Z) building blocks. To determine which product resides on which bead, information is encoded on each bead in the form of a series of chemically distinct binary tags ($T_1/T_2$ and $T_3/T_4$ as shown) with $T_1/T_2$ being used for the first cycle and $T_3/T_4$ used for the second cycle. (Subsequent cycles would need further pairs of distinct tags, e.g., $T_5/T_6$, $T_7/T_8$, etc.) At the end of the synthesis, the tags are cleaved from the solid support and appropriate methods (GC with ECD, e.g., with haloaromatic tags) are used to determine if a given tag is present on a given bead. The presence or absence of the tags is a binary code that can be used to decode the corresponding combinatorial products on the bead.

volatility of the analytes) and allowed direct analysis of the derivatized tags by GC-ECD.

### 5.5 Radiofrequency Encoding

Radiofrequency (RF) encoding or electronic encoding avoids problems associated with various chemical encoding schemes such as synthetic limitations or bioassay interferences arising from the presence of chemical tags [29]. This type of encoding uses very small electronic devices called radio frequency (RF) memory tags. These tags are in the millimeter range and are encased in sturdy plastic or glass sheaths. The tags are passive; they have no internal power source. Rather they contain an internal rectifying circuit that can absorbed RF energy and converted it to direct current (DC) electrical power. By appropriately irradiating the tag with external RF radiation and placing an antenna interfaced to a computer to record and analyze the passive signal, the tag can be "interrogated" as to its nature. Each tag can be so manufactured as to give a signal signature unique from all other tags. A solid-phase combinatorial structure such as a bead or bag of beads carrying an RT tag can be explicitly and automatically tracked through each reaction step. This allows the entire combinatorial sequence to be reconstructed and the products to be deduced. This technology is by no means exotic, difficult, or overwhelmingly expensive. Similar tags are used to track pets, livestock, pallets of merchandise, and the like and may greatly simplify combinatorial chemistry.

## 6. FUTURE OF COMBINATORIAL CHEMISTRY AND DRUG DISCOVERY

Combinatorial chemistry is a technology for creating large, sometimes extraordinarily large, numbers of small potentially druglike molecules for use in screening. Its roots lie, at least in part, in the increasing difficulty of traditional synthetic and natural product chemistry to generate new drugs in the 1980s and 1990s. However, the ability to generate $10^6$ (or even $10^7$) new chemicals does not mean that those molecules will be well distributed across chemical space, a fact generally not well appreciated at the time that combinatorial chemistry was rapidly developing. Virtually every large pharmaceutical company has invested in combinatorial chemistry, often very substantially, and the fruits of the money and effort have been impressive in terms of both chemistry and technology. Combinatorial chemistry has not been without its success, but it has not opened a flood gate to a wealth of new drugs as some of its early proponents once dreamed. It is not the "magic bullet" of drug therapy any more than liposomes (two drugs on the market), antisense oligonucleotides (one drug on the market), gene therapy (zero drugs on the market), or nano-this or nano-that (no drug even close to market) are. For those familiar with the early days of natural product drug discovery, there was also euphoria at that time that a great wellspring of new drugs had been discovered. Indeed natural products have been a much more generous source of drugs than combinatorial chemistry has been so far, but that spring is not infinite and the discovery of new natural products will require great effort. Combinatorial chemistry, along with natural product and traditional synthetic and medicinal chemistry, is a most welcome and important tool, and one

long to remain in our arsenal in the fight against disease. However, it cannot stand alone. It is highly doubtful we will ever find the next penicillin in a neglected Petri dish; those days are long gone. In new drug discovery, we must depend on careful painstaking science and hard thought, work, and decisions; serendipity and sophisticated technology have their place, but they are no substitutes.

## 7 KEY POINTS

1. The two main sources of drugs are natural products and synthetic drugs.
2. The PK/ADME properties of a drug must be considered along with the drug's efficacy and toxicity.
3. Combinatorial chemistry is a means of generating a great many new synthetic compounds, usually by massive parallel synthesis.
4. For drug discovery, the best chemical libraries are those that cover as much of chemical space as possible.
5. Combinatorial chemistry and HTS could not survive without one another.
6. Some means of encoding information on the synthetic steps of a combinatorial synthesis is critical to determining the resultant product structure.

## 8 REVIEW QUESTIONS

1. List advantages and disadvantages of natural product drugs.
2. Do the same for synthetic drugs.
3. What is meant by chemical diversity and chemical space?
4. Calculate the permutations, combinations, and possibilities for 12-mer polypeptides.
5. List the rule of five. What is it used for?
6. Explain Merrifield's solid-phase peptide synthesis; use diagrams as needed.
7. Explain split-pool synthesis; use diagrams as needed.
8. Which is, and why, the best resin to be used with polar solvents?
9. What is the purpose of a linker?
10. What is encoding and why is it important?
11. Explain oligomeric tagging.
12. Explain binary tagging.

## REFERENCES


1. Grever M, Schepartz S, Chabner B (1992). The National Cancer Institute: Cancer drug discovery and development program. *Semin Oncol 19*, 622–638.
2. Newman D, Cragg G, Snader K (2003). Natural products as sources of new drugs over the period 1981–2002. *J Nat Prod 66*, 1022–1037.
3. Demain A (1983). New applications of microbial products. *Science 219*, 709–714.
4. Balandrin M, Klocke J, Wurtele E, Bollinger W (1985). Natural plant chemical: Sources of industrial and medicinal materials. *Science 228*, 1154–1160.
5. Vining I (1990). Functions of secondary metabolites. *Annu Rev Microbiol 44*, 395–427.
6. Maplestone R, Stone M, Williams D (1992). The evolutionary role of secondary metabolites—A review. *Gene 115*, 151–157.
7. Cavalier-Smith T (1992). Origins of secondary metabolism. *Ciba Found Ser 17*, 64–87.
8. Aszalos A, Bostian M, McNitt K (1980–1987). *Handbook of Antibiotic Compounds*, Vols. 1–14, Boca Raton, FL: CRC Press.
9. Buckingham J (1994). *Dictionary of Natural Products* Vols 1–7, London: Chapman and Hall.
10. Vining I (1992). Secondary metabolism inventive evolution and biochemical diversity—A review. *Gene 115*, 135–140.
11. Wani M, Taylor H, Wall M, Coggon P, McPhail A (1971). Plant antitumor agents. VI. The isolation and structure of taxol, a novel antileukemic and antitumor agent from *Taxus brevifolia*. *J Am Chem Soc 93*, 2325–2327.
12. Goodman J, Walsh V (2001). *The Story of Taxol: Nature and Politics in the Pursuit of an Anti-Cancer Drug*. Cambridge: Cambridge University Press, p. 81.
13. The Convention on Biological Diversity. http://www.cbd.int/.
14. Bérdy J (1985). Screening, classification and identification of microbial products. In *Discovery and Isolation of Microbial Products*, M.S. Verrall ed. Chichester, England: Ellis Horwood Limited, pp. 9–28.
15. Merrifield R (1963). Solid phase peptide synthesis. I. The synthesis of a tetrapeptide. *J Am Chem Soc 85*, 2149.
16. Lipinski C, Lombardo F, Dominy B, Feeney P (2001). Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv Drug Dev Rev 46*, 3–26.
17. Guillier F, Orain D, Bradley M (2000). Linkers and cleavage strategies in solid-phase organic synthesis and combinatorial chemistry. *Chem Rev 100*, 2091–2158.
18. Wang S (1973). *p*-Alkoxybenzyl alcohol resin and *p*-alkoxybenzyloxycarbonyl-hydrazide resin for solid phase synthesis of protected peptide fragments. *J Am Chem Soc 95*, 1328–1333.

19. Mergler M, Tanner R, Gosteli J, Grogg P (1988). Peptide synthesis by a combination of solid-phase and solution methods: A new very acid-labile anchor group for the solid phase synthesis of fully protected fragments. *Tetrahedr Lett 29*, 4005–4008.
20. Rink H (1987). Solid-phase synthesis of protected peptide fragments using a trialkoxy-diphenyl-methylester resin. *Tetrahedr Lett 28*, 3787–3790.
21. Fyles T, Leznoff C (1976). The use of polymer supports in organic synthesis. V. The preparation of monoacetates of symmetrical diols. *Can J Chem 54*, 935–942.
22. Fréchet J, Haque K (1975). Use of polymers as protecting groups in organic synthesis. II. Protection of primary alcohol functional groups. *Tetrahedr Lett 16*, 3055–3056.
23. Thompson L, Ellman J (1994). Straightforward and general method for coupling alcohols to solid supports. *Tetrahedr Lett 35*, 9333–9336.
24. Chenera B, Finkelstein J, Veber D (1995). Protodetachable arylsilane polymer linkages for use in solid phase organic synthesis. *J Am Chem Soc 117*, 11999.
25. Rich D, Gurwara S (1975). Preparation of a new *o*-nitrobenzyl resin for solid-phase synthesis of *tert*-butyloxycarbonyl-protected peptide acids. *J Am Chem Soc 97*, 1575–1579.
26. Geysen H, Meloen R, Barleling S (1984). Use of peptide synthesis to probe viral antigens for epitopes to a resolution of a single amino acid. *Proc Nat Acad Sci USA 81*, 3998–4002.
27. Kerr J, Banville S, Zuckermann R (1993). Encoded combinatorial peptide libraries containing non-natural amino acids. *J Am Chem Soc 115*, 2529–2531.
28. Ohlmeyer H, Swanson R, Dillard L, Reader J, Asouline G, Kobayashi R, Wigler M, Still W (1993). Complex synthetic chemical libraries indexed with molecular tags. *Proc Nat Acad Sci USA 90*, 10922–10926.
29. Czarnik A. (1997). Encoding strategies in combinatorial chemistry. *Proc Nat Acad of Sci USA 94*, 12738–102739.

# LOCOREGIONAL CHEMORADIOTHERAPY OF CANCER USING HYDROGEL

David J. Yang, Ali Azhdarinia, Hiroaki Kurihara, Jerry Bryant, and E. Edmund Kim
*Division of Diagnostic Imaging, The University of Texas M.D. Anderson Cancer Center, Houston, Texas*

## 1 INTRODUCTION

### 1.1 Values of Molecular Imaging in Cancer Treatment

Drug discovery and development are accelerating due to rapid synthesis of potential drugs and development of high-throughput in vitro tests. The development of radiolabeled biochemical compounds and imaging devices to detect the radioactivity by external imaging has expanded the use of nuclear medicine studies in drug development. Agents developed for positron emission tomography (PET) and single photon emission computed tomography (SPECT) are molecular imaging agents because they can characterize the disease. Since molecular imaging plays a major role in drug development because of its ability to quantify drug properties in vivo, it is beginning to be used in all phases of drug discovery and development. PET and SPECT agents show high specific activities because they are made through a nuclear transformation and use carrier-free forms of isotopes. PET and SPECT agents do not produce detectable pharmacologic effects but provide important information concerning the characterization of varieties of tumors (vascular angiogenesis, hypoxia, apoptosis, cellular signaling, and transcriptional activity) [1–8]. Computed tomography (CT), magnetic resonance imaging (MRI), and ultrasound are prognostic tools because they do not provide cellular target information. Thus, assessment of the effectiveness of cancer therapy is not optimal. However, combining the anatomical and morphological location from CT, PET/CT and SPECT/CT are able to accurately evaluate posttherapy anatomic alteration. In addition, PET and SPECT agents may assist in the determination of optimal therapeutic dosing, differential diagnosis between inflammation/infection and recurrence, sensitive or resistant to treatment response, grading of tumors, and the prediction of treatment response by selecting patients who may respond to therapy.

### 1.2 Radiotheranostic Isotopes for Image-Guided Locoregional Chemoradiotherapy

Radiation therapy (XRT) is a form of cancer treatment that uses ionizing radiation to destroy cancer cells. Roughly 60% of all cancer patients receive XRT for multiple cancers including lung, breast, prostate, and head and neck. XRT can be delivered via external beams (photon, proton, neutron, and boron–neutron capture therapy) or internally using brachytherapy seeds. Conventional external beam radiation plays a limited role in the management of patients with certain nonresectable tumors because the maximum tumor dose is limited by the normal tissue tolerance [9]. Internal radiation therapy (IRT) with iodine-125 ($^{125}$I) (150 keV β emission, $t_{1/2} = 59.4$ days) brachytherapy seeds are the preferred method for localized prostate cancer treatment. However, several drawbacks exist with such a procedure. First, brachytherapy seeds can only be used once and the procedure cannot be repeated

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

within a patient. Second, the long half-life of $^{125}$I may result in radiation-induced tissue damage to surrounding tissues [10]. Also, due to the inability of the low-energy γ rays of $^{125}$I to escape the body as well as the geometry of the seeds, radiation dosimetry estimation is difficult to perform. Additionally, brachytherapy seed preparation involves labeling with $^{125}$I using copper or silver wires in a method that has low yield [11] and more potential radiation exposure, thus a simple and efficient IRT method is needed. Such factors have lead investigators to explore a multitude of therapeutic isotopes that may be suitable for IRT.

Yttrium-90 ($^{90}$Y) (2.28 MeV $\beta^-$ emission, $t_{1/2} = 64$ h) labeled ibritumomab tiuxetan (Zevalin) is the first radioimmunotherapy to receive Food and Drug Administration (FDA) approval and is used for the treatment of relapsed or refractory low-grade, follicular, or transformed B-cell non-Hodgkin's lymphoma (NHL). Although extensive clinical and preclinical studies have been undertaken with different $^{90}$Y molecules, one major drawback exists: It is a pure β em itter and $^{111}$In was used for imaging, biodistribution, and assessing dosimetry [12–14]. However, the estimated dosimetry obtained from $^{111}$In for $^{90}$Y is not accurate enough. Thus, much attention has been directed toward practical application of radiotheranostic radionuclides that have imaging and therapeutic capabilities for more accurate dosimetry calculations [15–21].

Among the rhenium radionuclides, the medium-energy ß emitter $^{186}$Re, with a physical half-life of 3.8 days, makes it a suitable candidate for radiotherapy. However, this reactor-produced $^{186}$Re contains a large amount of carrier, which is the cause of its low reported specific activity (5.92–8.88 GBq/mg, 160–240 mCi/mg Re) [22, 23]. The $^{188}$W/$^{188}$Re generator-produced radionuclide, $^{188}$Re, on the other hand, is carrier free and might offer higher specific activity as in the case of $^{99}$Mo/$^{99m}$Tc (37,000–74,000 MBq/mg; 1000–2000 mCi/mg $^{99m}$Tc), which makes it very convenient for clinical use [24]. $^{188}$Re has good characteristics for imaging and for potential therapeutic use because of its ß energy (2.1 MeV) with tissue penetration of 3.5 mm, its short physical half-life (16.9 h) and its 155-keV γ-ray emission (15%) for dosimetric and imaging purposes. The short physical half-life of $^{188}$Re allows for higher doses compared with long-lived radionuclides. Furthermore, the short half-life reduces the problems of radioactive waste handing and storage. $^{188}$Re is of interest for a variety of therapeutic applications including radioimmunotherapy [25, 26] and use of $^{188}$Re aminex microspheres and colloids for the potential treatment of rheumatoid arthritis of the synovial joints [27, 28] and hepatoma [29, 30]. Other applications that may be pursued include treatment of bone pain with $^{188}$Re-labeled phosphates [31–34]. Recently, an initial feasibility study with $^{188}$Re-hydroxyethylidene diphosphonate ($^{188}$Re-HEDP) using different $^{188}$Re sources demonstrated that application of $^{188}$Re-HEDP is safe and results in pain relief palliation [34]. In addition, dose escalation studied with $^{188}$Re-HEDP in prostate cancer patients showed that a 3.3-GBq (~90-mCi) dose is the maximum tolerated [35].

### 1.3 Selection of Chemistry Chelators for Radiotheranostic Applications

Several chelators have been reported such as $N_4$ (e.g., Cyclam-14, tetraazacyclododecane-1,4,7,10-tetraacetic acid, DOTA), $N_3S$ (e.g., mertiatide, MAG-3), $N_2S_2$ (e.g., ethylene dicysteine, ECD), $NS_3$, $S_4$ (e.g., sulfur colloid), diethylenetriamine pentaacetic acid (DTPA), $O_2S_2$ (e.g., dimercaptol succinic acid, DMSA), and hydrazinenicotinamide (HYNIC) [5, 36–39]. $^{188}$Re is in the same family as $^{99m}$Tc. Thus, $^{99m}$Tc chelates were used for the testing of stability for $^{188}$Re. Among these chelators, DTPA has faster wash out and forms less stable complexes with $^{99m}$Tc. $^{99m}$Tc-HYNIC has been shown to be useful in imaging, but labeling HYNIC with $^{99m}$Tc requires two chemicals, thiphenylphosphine and tricine, which are inconvenient for the kit preparation. The nitrogen and sulfur combination has been shown to be a stable chelator for $^{99m}$Tc-bis-aminoethanethiol tetradentate ligands, also called diaminodithiol compounds, and are known to form very stable Tc(V)O complexes on the basis of efficient binding of the oxotechnetium group to two thiosulfur and two amine nitrogen atoms.

### 1.4 Locoregional Chemoradiotherapy Using Hydrogel

Local administration, coupled with slow release, has demonstrated the ability to deliver effective doses of chemotherapy while significantly reducing commonly associated systemic toxicities [40]. While slow release has favorable characteristics in chemotherapy, it is not ideal when using radionuclides due to leakage of radioactive material and irradiation of nontarget tissues. Current hydrogel technologies are well established for slow-release chemotherapy [41, 42] and are under investigation for radionuclide retention [43–45]. The hydrogel is comprised of two components, a biodegradable polymer such as sodium alginate (SA) and a cross-linker such as calcium chloride or tin chloride, delivered separately from a dual-lumen-type syringe. SA is a hydrophilic gelling polysaccharide extracted from giant brown seaweed that has previously been described as having satisfactory hemocompatibility. Alginates have not been observed to accumulate in any of the major organs, and some evidence for in vivo degradation has even been described [46]. Alginate beads or microspheres can be instantly formed by cross-linking with divalent cations

such as calcium. The reaction is simple, and no toxic material is produced as described below: In addition, SA has a carboxylic group in the repeating unit, which makes it possible to chelate with radiometallic isotopes. In the present study, the use of a hydrogel to deliver combined therapeutic modalities, radiation and chemotherapy, administered locoregionally in tumor-bearing animal models was evaluated.

## 2 MATERIALS AND METHODS

### 2.1 In Situ Formation of Hydrogel

Cisplatin was added to a sodium alginate water solution (5% w/w). For in vitro release assays, the cisplatin-containing solution was added dropwise to a calcium chloride water solution (8% w/w). The hydrogel was formed immediately. The aqueous solution was decanted and the hydrogel was air dried. Preparation of hydrogel for in vivo studies required two injections through a single puncture site in the tumor using a single-lumen syringe. First, sodium alginate was injected intratumorally, followed by injection of calcium chloride. Mild pressure was applied to the injection site to aid in mixing and expedite hydrogel formation. This method was employed with and without the therapeutic components, cisplatin and $^{188}$Re, as dictated by the study. Preparation of $^{188}$Re consisted of elution of radioactivity from a $^{188}$W/$^{188}$Re generator (Oak Ridge National Laboratories, 250 mCi) using 5 mL 0.9% saline. The eluant was passed through a wetted alumina SepPak (Waters) to remove any tungsten-188 breakthrough and reduced in the presence of tin(II) chloride (2 mg) and heat (75°C, 20 min). This yielded reduced $^{188}$Re particles. The $^{188}$Re-containing solution was added dropwise to a calcium chloride water solution (5% w/w) and the hydrogel was formed. For in vivo studies, hydrogel loaded with cisplatin (CDDP) and $^{188}$Re was prepared in situ through a dual-lumen syringe.

### 2.2 Slow-Release Assays

Cisplatin-loaded hydrogel (SA-CDDP) was incubated in phosphate-buffered saline (PBS) in tubes at 37°C. Samples were collected in aliquots at various time intervals for 15 h, and the amount of cisplatin released from the hydrogel was determined by ultraviolet analysis ($\lambda_{max}$ = 230 nm). Cisplatin powder was used as a control.

### 2.3 Cell Viability Assays

Breast tumor cells were obtained from RBA CRL-1747 rat breast cancer cell line (American Type Culture Collection, ATCC< Rockville, MD). This cell line was derived from a tumor induced in Fischer-344 rats by giving an oral dose of 7,12-dimethylbenz[a]anthracene. The cells were cultured in Eagle's medium (MEM) with Earle's balanced salt solution (BSS) (90%) and fetal bovine serum (10%): 50,000 cells were plated into each well of 12-well plates and incubated at 37°C for 48 h to form monolayers. In order to prepare various radioactivity levels for the study, a fixed amount of hydrogel was determined and reacted with varying amounts of $^{188}$Re solution stock in test tubes to allow formation of $^{188}$Re-hydrogel. $^{188}$Re-hydrogel was then added to the cells with varying radioactivity ranging from 25 to 100 μCi/well to determine the relationship between cell viability and increased radiation dose. After 2 h incubation, the cells were harvested and prepared for a protein assay using a modified Lowry's protein assay kit (Sigma). Protein concentration was then determined for each sample using a dual-wavelength spectrophotometer [47, 48]. Hydrogel without $^{188}$Re was used as the control group. Each data point represents an average of three measurements.

### 2.4 In Vitro Radiation Dose Estimates from $^{188}$Re-Hydrogel

To estimate dosimetry emitted from $^{188}$Re-hydrogel, $^{188}$Re-hydrogel (5 mCi) was formulated as previously described in the text. In situ hydrogel with or without streaking forms were performed. Both $^{188}$Re-hydrogel forms were placed in 3-mL vials against X-ray films (MD55 Radiochromic film, Kodak). The films were exposed for 16 h. The color scales were used to estimate radiation dose (Gy) emitted from $^{188}$Re-hydrogel.

### 2.5 Treatment of Tumor-Bearing Animal Models with Hydrogel

The animals were housed in The University of Texas M. D. Anderson Cancer Center facility. All protocols involving animals are approved by the M. D. Anderson Animal Use and Care Committee and adhered to the principles of laboratory animal care. Fischer-344 Rats (150 ± 25 g) (Harlan Sprague–Dawley, Indianapolis, IN) or athymic nude mice were inoculated subcutaneously with rat breast adenocarcinoma cells (13762, $10^6$ cells/rodent) or human prostate tumor cells (PC-3) into the lumber region in legs using 25-gauge needles. The studies were performed 10–12 days after inoculation when tumors were palpable and could be measured. Treatment groups for mammary tumor-bearing rats were designated as control ($n$ = 6), CDDP ($n$ = 5, 5.4 mg/mL, 3 mg/kg body weight), CDDP-hydrogel ($n$ = 5, CDDP equivalent 3 mg/kg body weight), $Na^{188}ReO_4$ ($n$ = 6, 0.5 and 1 mCi/rat), $^{188}$Re-hydrogel ($n$ = 12, 0.5 and 1 mCi/rat), and $^{188}$Re-hydrogel loaded

with CDDP ($n = 3$, 5.4 mg/mL, 3 mg/kg body weight, 0.5–1 mCi/rat). In human prostate tumor-bearing nude mice treated with hydrogel studies, nude mice were inoculated with PC-3 tumor cells and allowed to reach an average tumor volume of 100 $mm^3$ on day 0. Treatment groups for nude mice bearing human prostate tumors were designated as control ($n = 10$), $Na^{188}ReO_4$ ($n = 10$, 1 mCi/mouse), $^{188}$Re-hydrogel ($n = 10$, 1 mCi/mouse), and $^{188}$Re-hydrogel loaded with CDDP ($n = 10$, 5.4 mg/mL, 3 mg/kg body weight, 1 mCi/mouse). Each animal received two injections: sodium alginate (with or without CDDP) followed by calcium chloride (with or without $Na^{188}ReO_4$). All injections were intratumoral and administered through the same puncture site. Slight pressure was applied to the injection site to aid in mixing of the injected chemicals and hydrogel formation. Tumor volume and body weight were recorded daily for 60 days. Tumor volumes were measured as [length ($l$) × width ($w$) × thickness ($h$)]/2. Loss of body weight of 15% was considered a chemical-induced toxic effect. Blood chemical assays [blood urea nitrogen (BUN) and serum creatinine, serum glutamate oxaloacetate transaminase (SGOT), serum glutamic pyruvic transaminase (SGPT)] were performed to detect hydrogel-induced renal or hepatotoxicity. A Student $t$ test was used to compare whether there is a significant difference between groups. To demonstrate the efficacy of hydrogel therapy, tumor specimens from rodents were excised at 48 h postadministration for histopathologic examination.

### 2.6 Assessment of Tumor Growth by Planar Scintigraphy

Scintigraphic images were obtained using an M-camera from Siemens Medical Systems (Hoffman Estates, IL). The camera was equipped with a medium-energy parallel-hole collimator. The field of view is 53.3 cm × 38.7 cm. The intrinsic spatial resolution is 3.2 mm and the pixel size is 19.18 mm (32 × 32, zoom = 1) to 0.187 mm (1024 × 1024, zoom = 3.2).

Animals with injections containing $^{188}$Re were imaged at 24 and 48 h postinjection. To determine tumor growth rate posthydrogel therapy, $^{99m}$Tc-ethylenedicysteine glucosamine ($^{99m}$Tc-EC-DG), a glycolysis marker, was selected for tumor localization. We previously have reported that $^{99m}$Tc-EC-DG is a marker for tumor growth assessment [49]. One rat was selected from each treatment group and was imaged with $^{99m}$Tc-EC-DG (300 μCi/rat, intravenously, 1 h postinjection) on day 9 posttreatment to visualize tumor mass. Tumors were excised and measurements expressed as tumor size (mm) and tumor volume ($mm^3$).

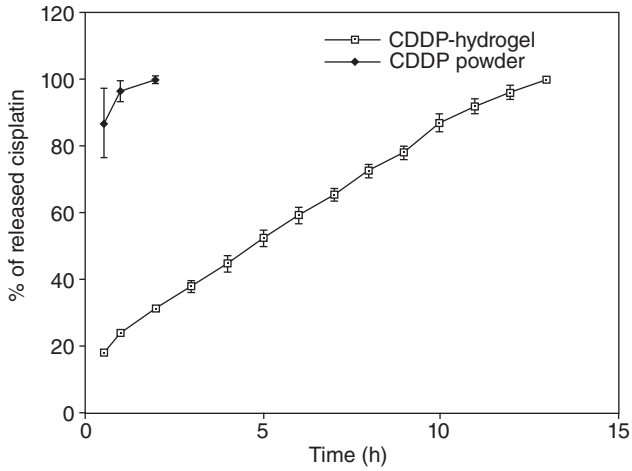


**Figure 1** Alginate hydrogel, loaded with 5.4 mg of CDDP, was placed in a test tube containing PBS and incubated at 37°C for up to 24 h. A 100-μL aliquot was removed at a designated interval. CDDP was quantified by ultraviolet absorbance.

## 3 RESULTS

### 3.1 Slow-Release Assays

Ultraviolet (UV) analysis of 100-μL aliquots showed complete release from the hydrogel within 15 h. The integrity of the hydrogel matrix was maintained through the endpoint of the study. No physical deterioration or degradation was observed. In the absence of hydrogel use, rapid release of CDDP-powder was observed and complete release was achieved within 2.5 h. (Fig. 1).

### 3.2 Cell Viability Assays

After incubation with different doses of $^{188}$Re-hydrogel, protein concentration was determined and showed that cell viability decreased as a function of increasing dose of the radionuclide. The data showed a significant decrease in absorbance in the 50-μCi group as compared to controls (Fig. 2). The 100-μCi group had >80% decrease in absorbance and may serve as a marker for the dose needed to kill tumor cells in vivo.

### 3.3 In Vitro Radiation Dose Estimates from $^{188}$Re Hydrogel

$^{188}$Re-hydrogel was placed in a 3-mL vial against X-ray films for 16 h as shown in Figure 3. Treatment of $^{188}$Re-hydrogel with or without streaking forms was performed. The color scales showed estimate radiation dose (Gy) emitted from $^{188}$Re-hydrogel (Fig. 4). Both forms had homogenous emission. The dose density in streaking form was less than that with hydrogel without streaking form.

### 3.4 Treatment of Tumor-Bearing Animal Models with Hydrogel

Animals were successfully injected with sodium alginate and $CaCl_2$ (in the presence of CDDP or $^{188}$Re) using the same injection site for each component. A resulting subcutaneous gel-like ball was formed near the injection site and designated that cross-linking had occurred and was also used as a marker to demonstrate that the hydrogel was intact. In mammary tumor-bearing rats treated with hydrogel studies, repeated measurements of tumor volume showed a clear delay in tumor growth in the $^{188}$Re-hydrogel group as compared to $^{188}$Re and control groups (Fig. 5), prolonging survival more than 2 times that of untreated rats. Low-dose $^{188}$Re-hydrogel (0.5 mCi) loaded with CDDP showed more anticancer effects than $^{188}$Re-hydrogel alone (Fig. 5). Rats given higher doses (1 mCi/rat) of $^{188}$Re/$^{188}$Re-hydrogel showed increased delay in tumor growth (Fig. 6) and those treated with CDDP-hydrogel demonstrated complete remission of tumors (Fig. 7A) while the untreated group experienced rapid increase in tumor size within 4 days, leading to euthanization due to excessive tumor burden (Fig. 7B). No marked difference in tumor volume was observed between $^{188}$Re-hydrogel (0.5 mCi) loaded with CDDP and CDDP-hydrogel (Fig. 7A). This effect might be due to the dose of $^{188}$Re used. Assessment of renal toxicity in CDDP-hydrogel-treated rats was shown to be minimal as demonstrated by minor variations in BUN and serum creatinine levels following treatment (Fig. 8). In nude mice bearing human prostate tumors administered with $^{188}$Re, $^{188}$Re-hydrogel, and $^{188}$Re-CDDP-hydrogel, the $^{188}$Re-CDDP-hydrogel showed more pronounced anticancer effect than free $^{188}$Re or gel alone (Fig. 9). Pathological stating of prostate tumor tissue post$^{188}$Re-CDDP-hydrogel showed cell apoptosis at 48 h (Fig. 10).

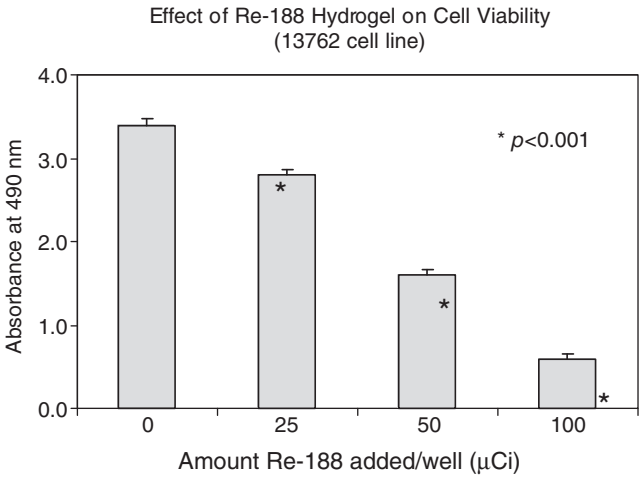


**Figure 2** Increasing radioactivity was added to each well and the cytotoxicity was assessed using a protein assay kit to determine UV absorbance. $*p < 0.05$, $t$ test compared to control group.

### 3.5 Assessment of Tumor Growth by Planar Scintigraphy

One of the groups treated with $^{188}$Re-hydrogel was monitored with planar scintigraphic imaging (Fig. 11). Images taken at 24 and 48 h postinjection of $^{188}$Re showed very poor retention of the radionuclide at the

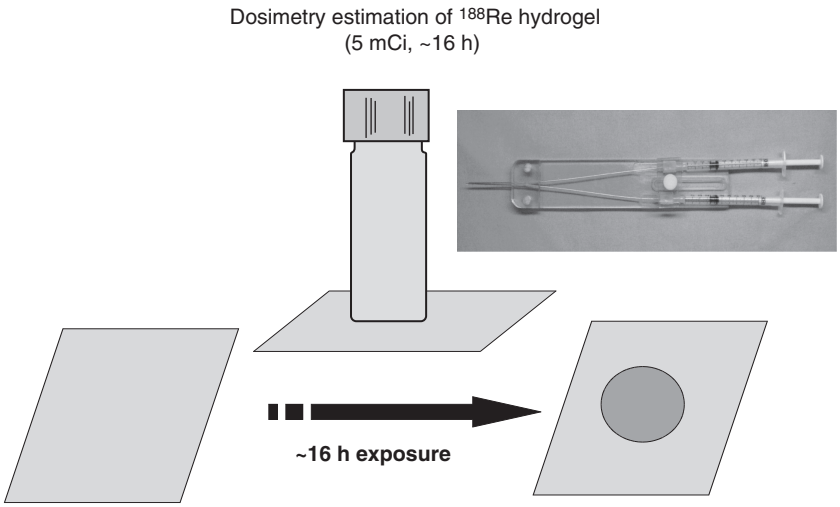


**Figure 3** Hydrogel with or without streaking forms were performed. Hydrogel was placed in a 3-mL vial against X-ray films (MD55 Radiochromic film, Kodak) for 16 h.

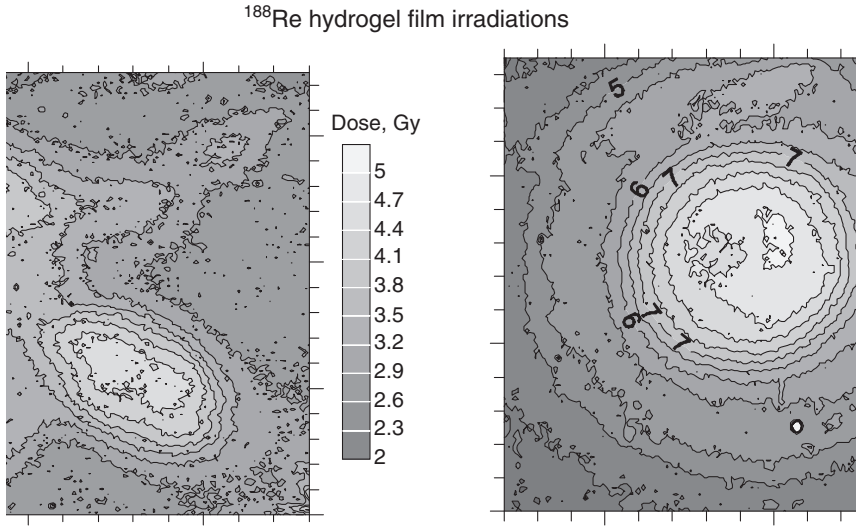


**Figure 4** $^{188}$Re-hydrogel with or without streaking forms were performed. The color scales showed estimate radiation dose (Gy) emitted from $^{188}$Re-hydrogel. Both forms had homogenous emission. The dose density in streaking form was less than that hydrogel without streaking form.

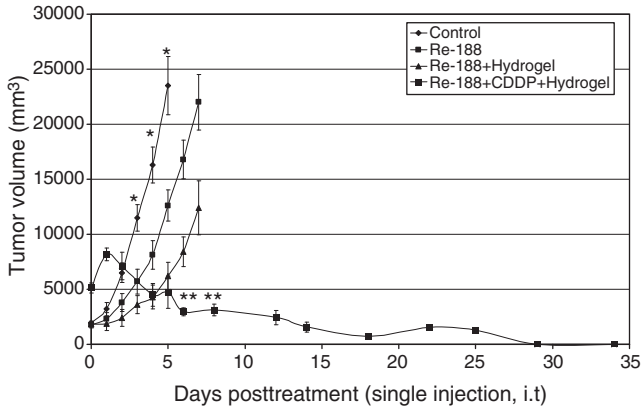


**Figure 5** Mammary tumor-bearing rats were randomized into three groups and injected with $^{188}$Re and $^{188}$Re-hydrogel (single injection, intratumorally, 0.5 mCi/rat). Tumor volume was measured and compared to the untreated group. $p < 0.05$, $t$ test: *compared to treated groups; **compared to $^{188}$Re and $^{188}$Re-hydrogel.

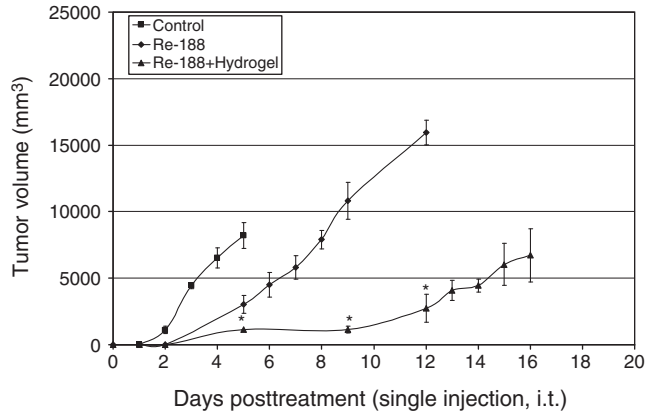


**Figure 6** Mammary tumor-bearing rats were treated with $^{188}$Re and $^{188}$Re-hydrogel (single injection, intratumorally, 1 mCi/rat) and tumor volume was monitored for up to 16 days posttreatment. Rats given higher doses (1 mCi/rat) of $^{188}$Re-hydrogel showed increased delay in tumor growth. *$p < 0.05$, $t$ test compared to $^{188}$Re.

injection site, with diffusion throughout the whole body and localization in the stomach of one rat. Imaging of the $^{188}$Re-hydrogel group demonstrated that the hydrogel can effectively trap the radionuclide at 48 h postinjection. No thyroid or stomach uptake was observed, signifying the absence of free $^{188}$Re from systemic circulation. Prior to sacrificing of the animals, imaging with $^{99m}$Tc-EC-DG allowed for clear visualization of tumors and qualitatively demonstrated significant differences in tumor size of $^{188}$Re and $^{188}$Re-hydrogel samples, respectively, based on count density in the tumor region (Fig. 12). Excision of the tumors and measurement of tumor volume showed correlation between count density and tumor volume.

## 4 DISCUSSION

Clinical use of ionizing radiation commonly involves dealing with external sources such as X-ray, proton, and neutron beams. Internal use of radiation is performed using brachytherapy, which takes advantage of the close

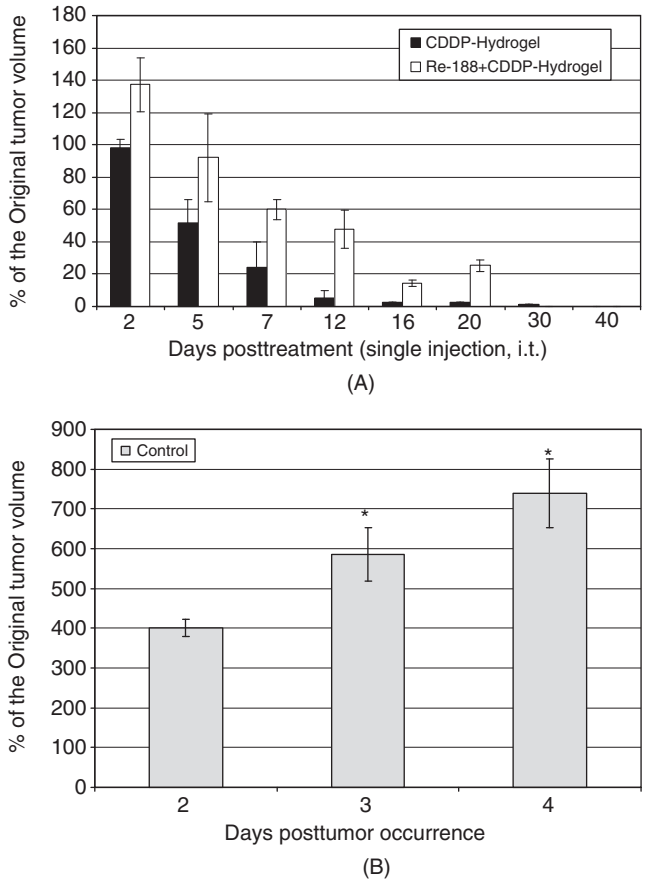


**Figure 7** (A) Mammary tumor-bearing rats were treated with CDDP-hydrogel (single injection, intratumorally, 3 mg/kg) and $^{188}$Re-CDDP-hydrogel (single injection, intratumorally, 0.5 mCi) and tumor volume was monitored until no longer palpable. (B) Tumor volume of untreated mammary tumor-bearing rats was monitored until tumor burden became too large and animals were euthanized. $*p < 0.05$, *t* test compared to day 2.

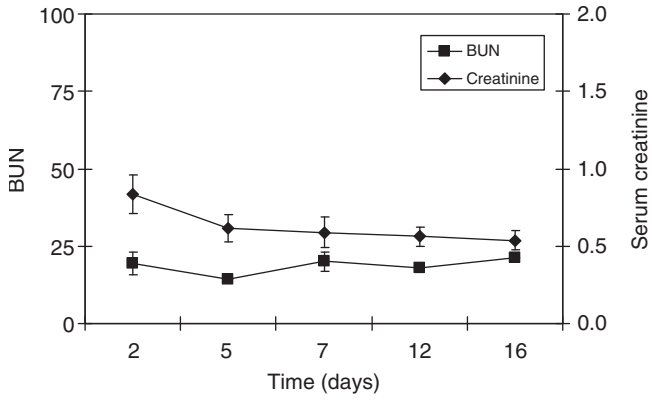


**Figure 8** Effect of CDDP-hydrogel on renal function of treated rats was assessed by monitoring BUN and serum creatinine levels for 16 days posttreatment.

proximity of the radioactive seeds to the tumor and delivers a predetermined radioactive dose for treatment.

Brachytherapy is an attractive treatment option for prostate cancer since the seeds are deposited intratumorally and spare surrounding healthy tissues from damage. $^{125}$I (half-life = 59.4 days) seeds are the preferred source but palladium-103 ($^{103}$Pd, half-life = 16.9 days) seeds are in use as well. Researchers are now seeking to expand the use of brachytherapy to multiple tumor types such as head and neck, breast and lung, among others. Some drawbacks, however, exist when using brachytherapy for treatment. First, brachytherapy seeds are nonbiodegradable and will remain in the area where they are injected permanently. Therefore, a patient that already received a seed implant is very unlikely to receive another one and must seek alternative treatment should tumor recurrence occur. Second, although these seeds are radioactive, they lack the ability to incorporate chemotherapy into their existing application.

While it has been shown that combined chemoradiation therapy has synergistic effects in various tumor types, such applications are not clinically available with brachytherapy but may be feasible as suggested by the hydrogel findings presented in this chapter. Figure 9 shows a comparison of some key differences between hydrogel and brachytherapy. The hydrogel is composed of a biodegradable, hemocompatible polymer that is virtually nontoxic. This polymer, once cross-linked with calcium chloride, forms with a 'gel-like' consistency and can be prepared in multiple configurations (thin tracks, globular, etc.). It must be noted that administration of the gel into the tumor will have an adverse effect on tumor volume, initially due to an increase in interstitial pressure and general enlargement of the area, followed by reduction in volume resulting from the therapeutic modality delivered (Fig. 7A). Though our in vivo imaging studies have not reached the necessary degradation time for alginate hydrogel, the data indicates that $^{188}$Re-gel is stable and localized at least up to 4 days (data not shown). Others have reported that 40% of the gel weight dissolved after 15 weeks, suggesting that degradation can occur in time frames ranging from 3 weeks to more than 4 months by simply varying the number of covalent and ionic cross-links [50]. The longevity with which the hydrogel remains intact in vivo coincides well with the selection of $^{188}$Re since, generally, >99% of a radionuclide is decayed after 10 half-lives, which is 7.1 days in the case of $^{188}$Re. This ensures that by the time breakdown of the hydrogel occurs no radioactivity remains and toxicity to surrounding tissue will be negligible. Additionally, the loading dose of the hydrogel can be varied based on the needs of the patient, shape

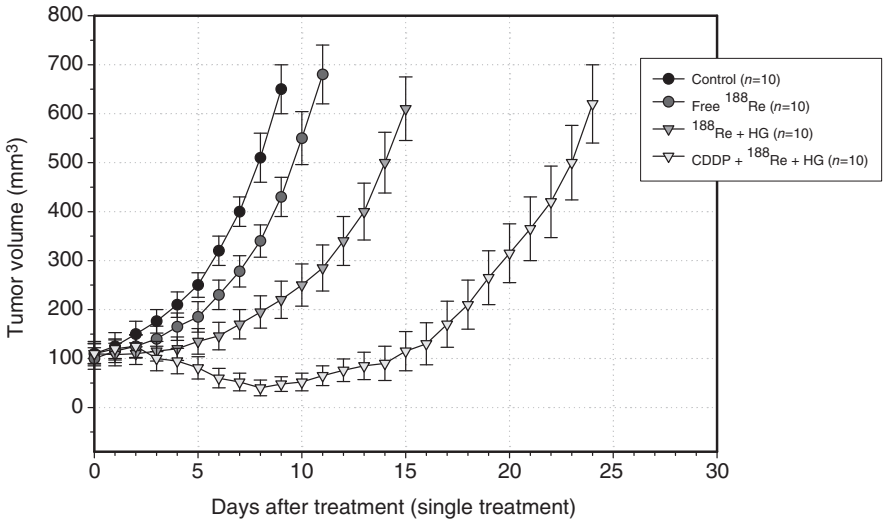


**Figure 9** Nude mice were inoculated with PC-3 tumor cells and allowed to reach an average tumor volume of 100 $mm^3$ on day 0. The mice were injected [intratumoral (i.t.)] with $^{188}$Re, $^{188}$Re-hydrogel, and $^{188}$Re-cisplatin (CDDP)-hydrogel (1 mCi, 250 μg). The tumor volume was monitored until tumor burdens became excessive. $^{188}$Re-CDDP-hydrogel showed more pronounced anticancer effect than free $^{188}$Re or gel alone.

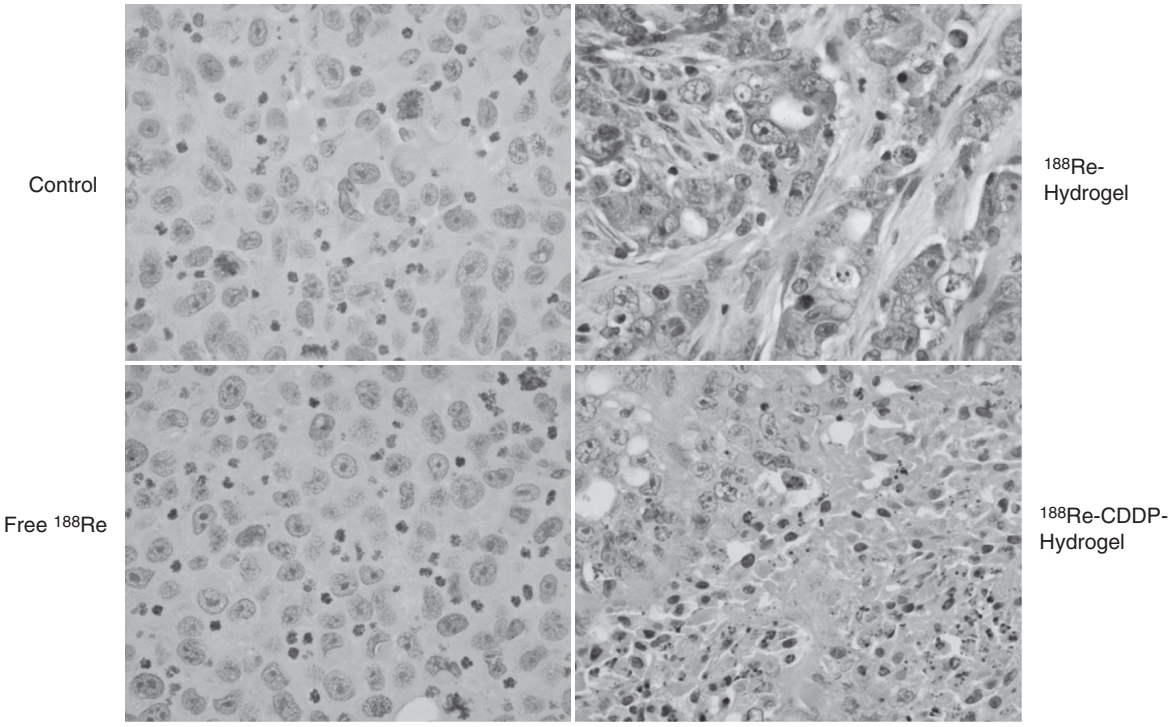


**Figure 10** H and E staining of prostate tumor tissue post-188Re-hydrogel and 188Re-CDDP-hydrogel showed cell apoptosis at 48 h (40×).

and location of the tumor, susceptibility of surrounding healthy tissue to radiation, and planned number of treatments. Increasing amounts of hydrogel and accompanying radioactivity can be delivered to certain sites as required based on delivery from a dual-lumen syringe, which can be administered under computed tomography (CT) or ultrasound (US) guidance using an automated step motor to ensure accurate and reproducible dose distribution. Ongoing studies have shown promising data using radiochromic film (not shown) for

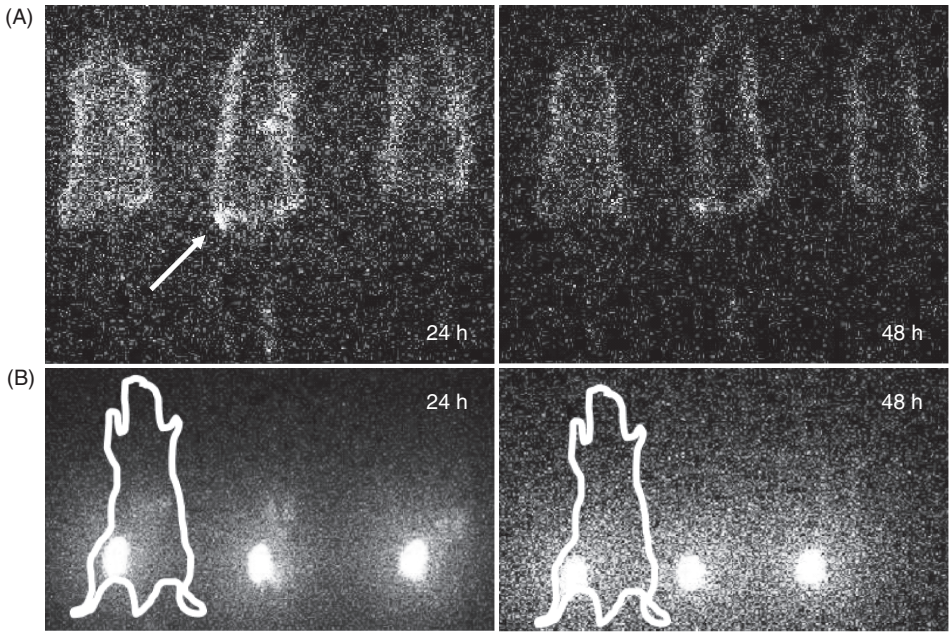


**Figure 11** Rats were injected with $Na^{188}ReO4$ (A) and $^{188}$Re-hydrogel (B) (1 mCi/rat, i.t.) in the hind limb and imaged at 24 and 48 h postinjection. Medium-energy collimator was used and the images demonstrate that $^{188}$Re-hydrogel provides locoregional radionuclide therapy and prevents whole-body distribution. Injection site shown by arrow.

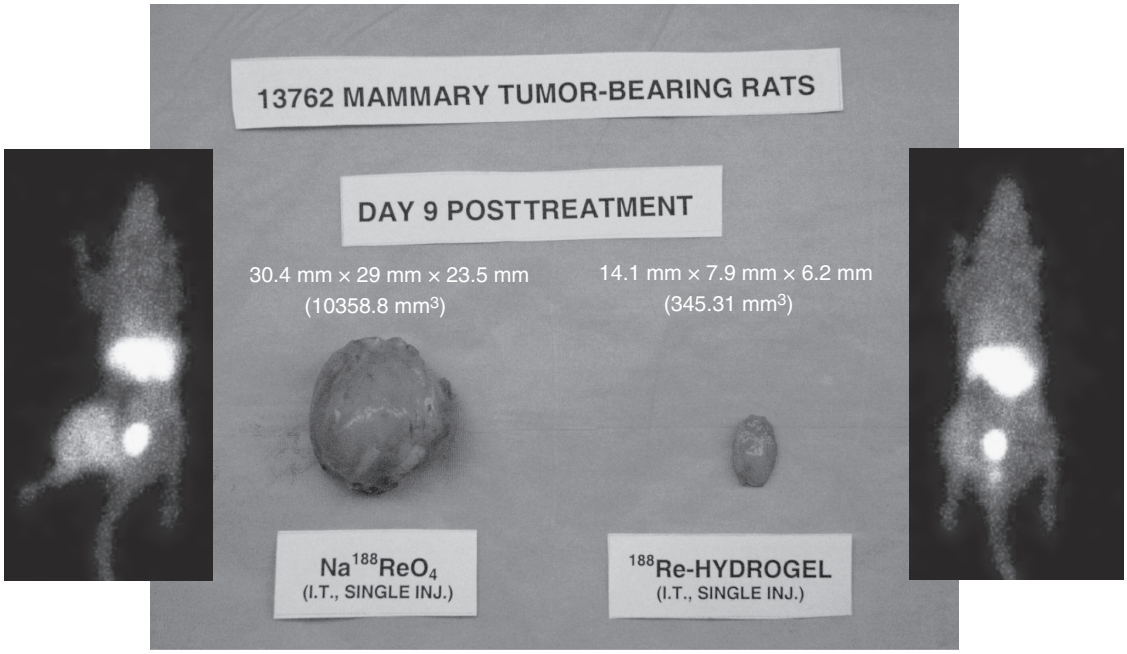


**Figure 12** Mammary tumor-bearing rats treated with $Na^{188}ReO_4$ (left) and $^{188}$Re-hydrogel (right) were imaged with 99mTc-EC-DG (a glycolysis marker, at 1 h) on day 9 posttreatment to visualize tumor mass. Tumors were excised and measurements are shown as tumor size (mm) and tumor volume ($mm^3$).

determining dosimetry and absorbed dose spatially, and may serve as a powerful tool in treatment planning.

In an attempt to incorporate well-known slow-release methodologies for drug delivery with radiotherapy, we developed a hydrogel that plays two roles in cancer treatment: (1) slow release of chemotherapeutic agents and (2) trapping of therapeutic radionuclides. While many hydrogels exist with similar polymeric composition and drug delivery capability, the uniqueness of this hydrogel lies in the chemical properties associated with the cross-linking agent calcium chloride. Calcium chloride acts bifunctionally in this system: forming a complex with $^{188}$Re and cross-linking with sodium alginate. This leads to gel formation and slow release of CDDP coupled with trapping of $^{188}$Re at the site of injection, which can be monitored by imaging. $^{188}$Re was chosen for this particular study because of its therapeutic potential, however, different contrast materials such at gadolinium-containing compounds for MRI, or iodine-containing compounds for CT could also be incorporated into the hydrogel framework and serve as a high-resolution guide to demonstrate where the therapy is actually being delivered, as well as how long the hydrogel is able to maintain its integrity in vivo.

In summary, we have demonstrated that in situ hydrogel can be used for local-regional delivery of chemotherapy and radiotherapy while reducing associated toxicities found with traditional delivery methods. The findings demonstrate the potential uses of image-guided therapy in cancer treatment.

## ACKNOWLEDGMENTS

This work was supported in part by Cell > Point L.L.C (MDA LS01-212) and the John S. Dunn Foundation. The animal research and NMR facility were supported by M. D. Anderson Cancer Center (CORE) Grant NIH CA-16672.

## REFERENCES

1. Zareneyrizi F, Yang DJ, Oh CS, et al. (1999). Synthesis of 99mTc-ethylenedicysteine-colchicine for evaluation of antiangiogenic effect. *Anticancer Drugs 10*, 685–692.
2. Yang DJ, Kim KD, Schechter NR, et al. (2002). Assessment of antiangiogenic effect using 99mTc-EC-endostatin. *Cancer Biother Radiopharm 17*, 233–245.
3. Yang DJ, Bryant J, Chang JY, et al. (2004). Assessment of COX-2 expression with 99mTc-labeled celebrex. *Anti-Cancer Drugs 15*, 255–263.
4. Ito M, Yang DJ, Mawlawi O, et al. (2006). PET and planar imaging of tumor hypoxia with labeled metronidazole. *Academic Radiol 13*, 598–609.
5. Ohtsuki K, Akashi K, Aoka Y, et al. (1999). Technetium-99m HYNIC-annexin V: A potential radiopharmaceutical for the in-vivo detection of apoptosis. *Eur J Nucl Med 26*, 1251–1258.
6. Vriens PW, Blankenberg FG, Stoot JH, et al. (1998). The use of technetium 99mTc annexin V for in vivo imaging of apoptosis during cardiac allograft rejection. *J Thorac Cardiovasc Surg 116*, 844–853.
7. Kubota K, Ishiwata K, Kubota R, et al. (1991). Tracer feasibility for monitoring tumor radiotherapy: A quadruple tracer study with fluorine-18-fluorodeoxyglucose or fluorine-18-fluorodeoxyuridine, L-[methy1-14C]methionine,[6–3*H*] thymidine, and gallium-67. *J Nucl Med 32*, 2118–2123.
8. Schechter NR, Yang DJ, Azhdarinia A, et al. (2003). Assessment of epidermal growth factor receptor with 99mTc-ethylenedicysteine-C225 monoclonal antibody. *Anticancer Drugs 14*, 49–56.
9. Marples B, Greco O, Joiner MC, Scott SD (2002). Molecular approaches to chemoradiotherapy. *Eur J Cancer 38*, 231–239.
10. Nath R (1993). New directions in radionuclide sources for brachytherapy. *Semin Radiat Oncol 3*, 278–289.
11. Mathew C, Majali MA, Balakrishnan SA (2002). A novel approach for the adsorption of iodine-125 on silver wire as matrix for brachytherapy source for the treatment of eye and prostate cancer. *Appl Radiat Isot 57*, 359–67.
12. DeNardo SJ, Williams LE, Leigh BR, Wahl RL (2002). Choosing an optimal radioimmunotherapy dose for clinical response. *Cancer 94*, 1275–1286.
13. Fisher D, Rajon D, Breitz H, Goris M, Bolch W, Knox S (2004). Dosimetry model for radioactivity localized to intestinal mucosa. *Cancer Biother Radiopharm 19*, 293–307.
14. Wiseman GA, Leigh BR, Erwin WD, Sparks RB, Podoloff DA, Schilder RJ, Bartlett NL, Spies SM, Grillo-Lopez AJ, Witzig TE, White CA (2003). Radiation dosimetry results from a Phase II trial of ibritumomab tiuxetan (Zevalin) radioimmunotherapy for patients with non-Hodgkin's lymphoma and mild thrombocytopenia. *Cancer Biother Radiopharm 18*, 165–178.
15. Novak-Hoferand I, Schubiger PA (2002). Copper-67 as a therapeutic nuclide for radioimmunotherapy. *Eur J Nucl Med Mol Imaging 29*, 821–830.
16. Delaloye AB, Delaloye B, Buchegger F, Vogel CA, Gillet M, Mach JP, Smith A, Schubiger PA (1997). Comparison of copper-67- and iodine-125-labeled anti-CEA monoclonal antibody biodistribution in patients with colorectal tumors. *J Nucl Med 38*, 847–853.
17. Srivastava SC, Atkins HL, Krishnamurthy GT, Zanzi I, Silberstein EB, Meinken G, Mausner LF, Swailem F, D'Alessandro T, Cabahug CJ, Lau Y, Park T, Madajewicz S (1998). Treatment of metastatic bone pain with tin-117m stannic diethylenetriaminepentaacetic acid: a phase I/II clinical study. *Clin Cancer Res 4*, 61–68.
18. Bugaj JE, Erion JL, Johnson MA, Schmidt MA, Srinivasan A (2001). Radiotherapeutic efficacy of (153)Sm-CMDTPA-

Tyr(3)-octreotate in tumor-bearing rats. *Nucl Med Biol 28*, 327–334.

19. Kwekkeboom DJ, Bakker WH, Kooij PP, Konijnenberg MW, Srinivasan A, Erion JL, Schmidt MA, Bugaj JL, deJong M, Krenning EP (2001). [177Lu-DOTAOTyr3] octreotate: Comparison with [111In-DTPAo]octreotide in patients. *Eur J Nucl Med 28*, 1319–1325.
20. Iznaga-Escobar N, Ramirez IL, Izquierdo JC, Suarez L, Morales D, Perez-Rodriguez R (2003). 188Re-labeled anti-epidermal growth factor receptor humanized monoclonal antibody h-R3: Labeling conditions, in vitro and in vivo stability. *Methods Find Exp Clin Pharmacol 25*, 703–711.
21. Luo TY, Hsieh BT, Wang SJ, Lin WY, Lee TW, Shen LH, Su MJ (2004). Preparation and biodistribution of rhenium-188 ECD/Lipiodol in rats following hepatic arterial injection. *Nucl Med Biol 31*, 671–677.
22. Bisunadan MM, Blower PJ, Clarke SEM, et al. (1991). Synthesis and characterization of [186Re]rhenium(V) dimercaptosuccinic acid: A possible tumour radiotherapy agent. *Appl Radiat Isot 42*, 167–171.
23. Blower PJ, Lam ASK, O'Doherty MJ, Kettle AG, Coakley AJ, Knapp Jr FF (1998). Pentavalent rhenium-188 dimercaptosuccinic acid for targeted radiotherapy: Synthesis and preliminary animal and human studies. *Eur J Nucl Med 25*, 613–621.
24. Kotzerke J, Glatting G, Seitz U, et al. (2000). Radioimmunotherapy for the intensification of conditioning before stem cell transplantation: Differences in dosimetry and biokinetics of 188Re- and 99mTc-labeled anti-NCA-95 MAbs. *J Nucl Med 41*, 531–537.
25. Bernhardt P, Benjegard SA, Koelby L, et al. (2000). Dosimetric comparison of radionuclides for therapy of somatostain receptor-expressing tumors. *Nucl Med Commun 21*, 566.
26. Buchman I, Bunjes D, Seitz U, et al. (2000). Radioimmunotherapy for myeloablation prior to stem cell transplantation with Re-188 CD 66 a,b,c,d,e antibody in high risk leukemia patients. *Eur J Nucl Med 27*, S81.
27. Wang SJ, Lin WY, Chen MN, et al. (1998). Rhenium-188 microsphere: A new radiation synovectomy agent. *Nucl Med Commun 19*, 427–433.
28. Wang SJ, Lin WY, Hsieh BT, et al. (1995). Rhenium-188 sulfur colloid as a radiation synovectomy agent. *Eur J Nucl Med 22*, 505–507.
29. Wang SJ, Lin WY, Chen MN, et al. (1996). Radiolabelling of Lipiodol with generator-produced 188Re for hepatic tumor therapy. *Appl Radiat Isot 47*, 267–271.
30. Wang SJ, Lin WY, Chen MN, et al. (1998). Intratumoral injection of rhenium-188 microspheres into an animal model of hepatoma. *J Nucl Med 39*, 1572–1577.
31. deKlerk JMH, vanDijk A, vanhetSchip AD, et al. (1994). Dose escalation study of rhenium-186-hydroxyethylidene diphosphonate in patients with metastatic prostate cancer. *Eur J Nucl Med 21*, 1114–1120.
32. Hsieh BT, Hsieh JF, Tsai SC, et al. (1999). Comparision of various rhenium-188 labeled diphosphonates for the treatment of bone metastases. *Nucl Med Biol 26*, 973–976.
33. Lin WY, Hsieh JF, Lin CP, et al. (1999). Effect of reaction conditions of rhenium-188 hydroxyethylidene diphosphonate complexes. *Nucl Med Biol 26*, 455–459.
34. Maxon HR, Schroder LE, Washburn LC, et al. (1998). Rhenium-188(Sn)HEDP for treatment of osseous metastases. *J Nucl Med 39*, 659–663.
35. Palmedo H, Guhlke S, Bender H, et al. (2000). Dose escalation study with rhenium-188 hydroxyethylidene diphosphonate in prostate cancer patients with osseous metastases. *Eur J Nucl Med 27*, 123–130.
36. Canet EP, Casali C, Desenfant A, et al. (2000). Kinetic characterization of CMD-A2-Gd-DOTA as an intravascular contrast agent for myocardial perfusion measurement with MRI. *Magn Reson Med 43*, 403–409.
37. Van Nerom CG, Bormans GM, DeRoo MJ, et al. (1993). First experience in healthy volunteers with technetium-99m L,L-ethylenedicysteine, a new renal imaging agent. *Eur J Nucl Med 20*, 738–746.
38. Laissy JP, Faraggi M, Lebtahi R, et al. (1994). Functional evaluation of normal and ischemic kidney by means of gadolinium-DOTA enhanced TurboFLASH MR imaging: A preliminary comparison with 99mTc-MAG3 dynamic scintigraphy. *Magn Reson Imaging 12*, 413–419.
39. Wu HC, Chang CH, Lai MM, et al. (2003). Using Tc-99m DMSA renal cortex scan to detect renal damage in women with type 2 diabetes. *J Diabetes Complications 17*, 297–300.
40. Colombo A, Drzewiecki J, Banning A, Grube E, Hauptmann K, Silber S, Dudek D, Fort S, Schiele F, Zmudka K, Guagliumi G, Russell ME (2003). Randomized study to assess the effectiveness of slow- and moderate-release polymer-based paclitaxel-eluting stents for coronary artery lesions. *Circulation 108*, 788–794.
41. Konishi M, Tabata Y, Kariya M, Suzuki A, Mandai M, Nanbu K, Takakura K, Fujii S (2003). In vivo anti-tumor effect through the controlled release of cisplatin from biodegradable gelatin hydrogel. *J Control Release 92*, 301–313.
42. Blanchette J, Kavimandan N, Peppas NA (2004). Principles of transmucosal delivery of therapeutic agents. *Biomed Pharmacother 58*, 142–151.
43. Ozeki M, Ishii T, Hirano Y, Tabata Y (2001). Controlled release of hepatocyte growth factor from gelatin hydrogels based on hydrogel degradation. *J Drug Target 9*, 461–471.
44. Quand X, Weinberger J (2002). Deposition of (90)YPO(4) and (144)CePO(4) radioisotopes on polymer surfaces for radiation delivery devices. *J Biomed Mater Res 63*, 98–105.
45. Quand X, Weinberger J (2001). Encapsulation of isotope on novel beta-emitting poly(ethylene terephthalate) surfaces. *J Biomed Mater Res 57*, 619–623.
46. Tonnesenand HH, Karlsen J (2002). Alginate in drug delivery systems. *Drug Dev Ind Pharm 28*, 621–630.

47. Matsuoka K, Watanabe M, Toriyama F (1997). In vitro analysis for cellular toxicity of polychlorinated biphenys (PCBs) on HeLa cellular proliferation (IV)—The effect of Kanpo preparations on cellular toxicity. *Fukuoka Igaku Zasshi 88*, 207–210.
48. Kim EJ, Sampathkumar SG, Jones MB, Rhee JK, Baskaran G, Goon S, Yarema KJ (2004). Characterization of the metabolic flux and apoptotic effects of *O*-hydroxyl- and *N*-acyl-modified *N*-acetylmannosamine analogs in Jurkat cells. *J Biol Chem 279*, 18342–18352.
49. Yang DJ, Kim CG, Schechter NR, Azhdarinia A, Yu DF, Oh CS, Bryant JL, Won JJ, Kim EE, Podoloff DA (2003). Imaging with 99mTc ECDG targeted at the multifunctional glucose transport system: Feasibility study with rodents. *Radiology 226*, 465–473.
50. Bouhadir KH, Alsberg E, Mooney DJ (2001). Hydrogels for combination delivery of antineoplastic agents. *Biomaterials 22*, 2625–2633.

# RNAi-MEDIATED THERAPEUTICS

Núria Morral and Scott R. Witting

*Department of Medical and Molecular Genetics and Center for Diabetes Research, Indiana University School of Medicine, Indianapolis, Indiana*

## 1 INTRODUCTION

Ribonucleic acid interference (RNAi) is the phenomenon by which small noncoding RNAs suppress the expression of a gene. In the early 1990s, experiments to block protein expression using antisense technology [i.e., RNA complementary to the target messenger RNA (mRNA)] indicated that introduction of the sense strand could also promote silencing [45, 55], suggesting that the silencing effect was the result of an independent mechanism. Later on, Fire and collaborators observed that in *Caenorhabditis elegans*, administration of double-stranded RNA (dsRNA) was at least two orders of magnitude more potent than the antisense or sense strands alone [46]. Furthermore, the interference effect spread from the tissue in which the dsRNA was injected to other tissues, and it was passed from the parental animal to the next generation through the germ line [46].

This seminal work triggered interest in determining whether RNA interference would be observed in other organisms. Present in most eukaryotes, including plants, *Neurospora*, *Drosophila*, and mammals [42, 57, 122, 140, 167], it is now clear this is a universal mechanism used to modulate gene expression. RNAi is triggered by dsRNA of variable lengths, which are processed to short RNA duplexes, approximately 21–28 nucleotides (nt) long. The first RNA silencing mechanisms were discovered as part of an antiviral response believed to protect organisms from viral infection. It was later discovered that cells have the capacity to synthesize small RNAs known as microRNA (miRNA), which regulate hundreds of genes, coordinating complex developmental as well as physiological responses. Based on the fact that miRNA molecules modulate gene expression in all mammalian tissues, RNAi is being exploited as a system for the treatment of human diseases as well as to study gene function.

## 2 MECHANISMS OF GENE SILENCING

### 2.1 MicroRNA

MicroRNAs (miRNAs) are small RNA molecules that regulate expression of genes involved in cellular metabolic pathways, development, cell differentiation, proliferation, and death [14, 22, 78, 138, 185]. It has been estimated that in eukaryotes, approximately 0.5–1% of predicted genes encode a microRNA, which in humans, regulates 30% of genes at an average of 200 genes per miRNA [86, 160]. MicroRNAs are tissue specific and differentially expressed during development [5, 80]. Although the gene targets of the majority of miRNA are still unknown, some have been determined. One such example is miR-122, which is abundantly expressed in liver, where it regulates a large number of cholesterol synthesis genes [78]. Another example is brain-specific miR-134, which negatively modulates the size of dendritic spines [148]. MicroRNAs have become targets for the treatment of cancer, as many are dysregulated in

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

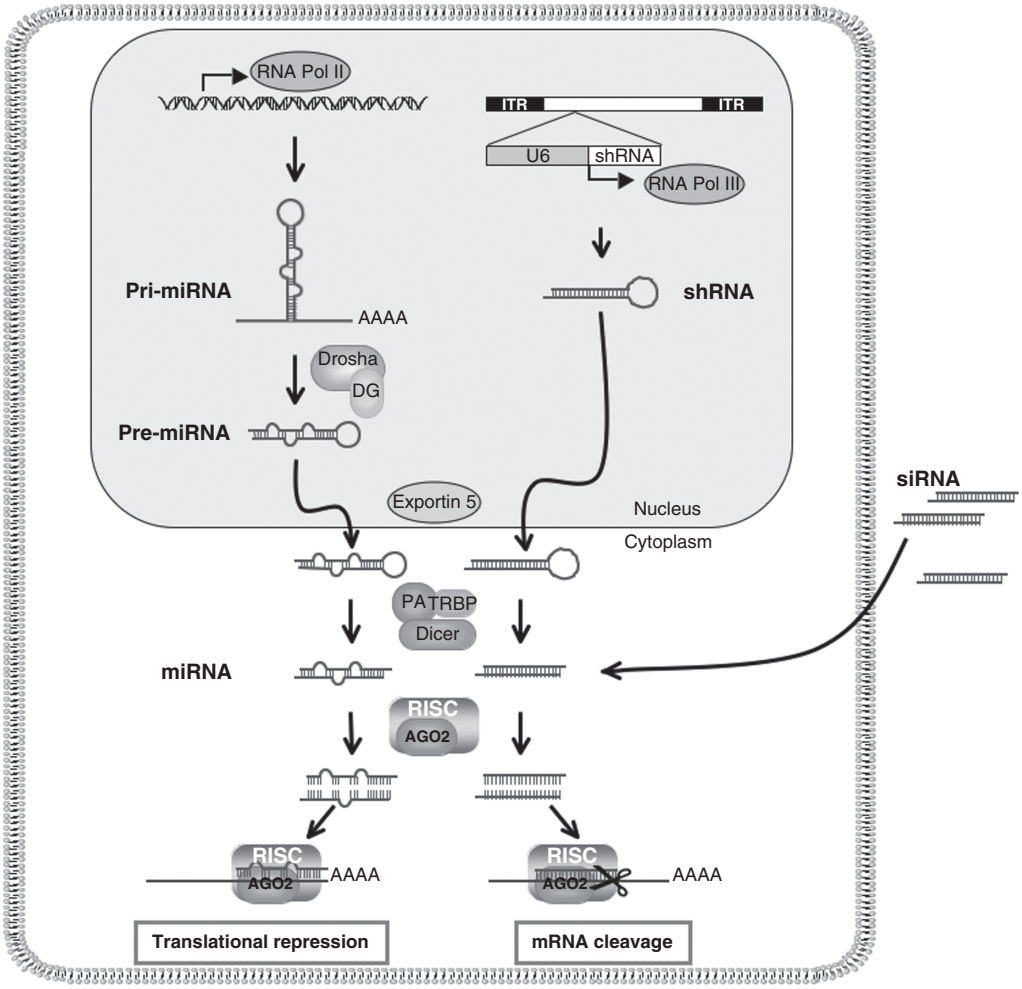


**Figure 1** RNAi machinery. In the miRNA pathway (top left), pri-miRNA generated by RNA polymerase II are processed to pre-miRNA by Drosha-DGCR8, and transported to the cytoplasm by exportin-5. Dicer-TRBP-PACT cleaves the loop sequence, generating ~22 nt sequences with 2-nt 3′ overhangs. RISC-AGO2 unwinds the dsRNA molecule and selects one strand for degradation while the other is used to direct silencing of the target mRNA. In animal cells, the majority of the mRNA is translationally repressed rather than degraded, although deadenylation and subsequent 5′ decapping, exposing the mRNA to exonuclease degradation, have also been observed. Expression of shRNA from a viral vector (top right) results in synthesis of molecules that enter the RNAi pathway at the exportin-5 level. shRNA are typically designed to completely match the mRNA target sequence, thereby inducing mRNA cleavage by AGO2. Chemically synthesized siRNA do not require Dicer activity and are directly processed by the RISC complex, triggering mRNA degradation. DG, DiGeorge syndrome critical region 8; PA, protein activator of PKR (PACT); TRBP, TAR RNA-binding protein; RISC, RNA-induced silencing complex; AGO2, endonuclease Argonaute 2.

oncogenic processes. For example, let-7 is a tumor-supressor miRNA, regulating the HMGA2 oncogene [85]. Its expression in breast cancer cell lines reduced cell proliferation [188].

MicroRNAs are transcribed by RNA polymerase II [18, 84] (and polymerase III to a lower extent [10]) as long primary transcripts (pri-miRNA). Pri-miRNA are cleaved by the microprocessor complex [34, 51], composed of the ribonuclease III Drosha and the dsRNA-binding protein DiGeorge syndrome critical region 8 (DGCR8), to generate precursor structures known as pre-miRNA (Fig. 1). Pre-miRNA are stem-loop molecules approximately 70-nt long, with a 5′ phosphate group and a 2-nt 3′ overhang [83]. Pre-miRNAs are

bound by karyopherin exportin-5 and transported out of the nucleus [96, 187]. In the cytoplasm, pre-miRNA are cleaved by the RNase III enzyme Dicer, which removes the loop sequence, leaving ~22-nt dsRNA molecules with 2-nt overhangs in the 3′ end of each of the two strands. This structure is recognized by the RNA-induced silencing complex (RISC), which selects one of the two strands for degradation (referred to as the passenger strand) and loads the guide strand, which is then used to bind the mRNA with a complementary sequence [50, 99, 134]. The core subunit of RISC is Argonaute 2, an endonuclease that is responsible for binding the miRNA and cleaving the target mRNA. The level of complementarity between the guide strand and the mRNA sequence determines whether protein synthesis will be inhibited or the mRNA will be degraded. Binding to the target mRNA with only partial homology blocks protein translation without inducing cleavage of the mRNA, while perfect or nearly perfect complementarity induces mRNA degradation. In animals, the guide strand usually binds to the 3′ untranslated region (UTR) of mRNAs imperfectly, resulting in protein translation inhibition and exonucleolytic mRNA degradation in cytoplasmic compartments known as processing bodies (P bodies) [94, 130].

Multiple studies have provided evidence that the 5′ region of miRNA is critical for gene silencing. In particular, nucleotide positions 2–7 or 2–8 from the 5′ end ("seed" region) (Fig. 2) are the most conserved among metazoan miRNAs [87], playing a key role in determining target specificity. Alteration of a nucleotide in these positions results in a significant reduction of silencing efficacy.

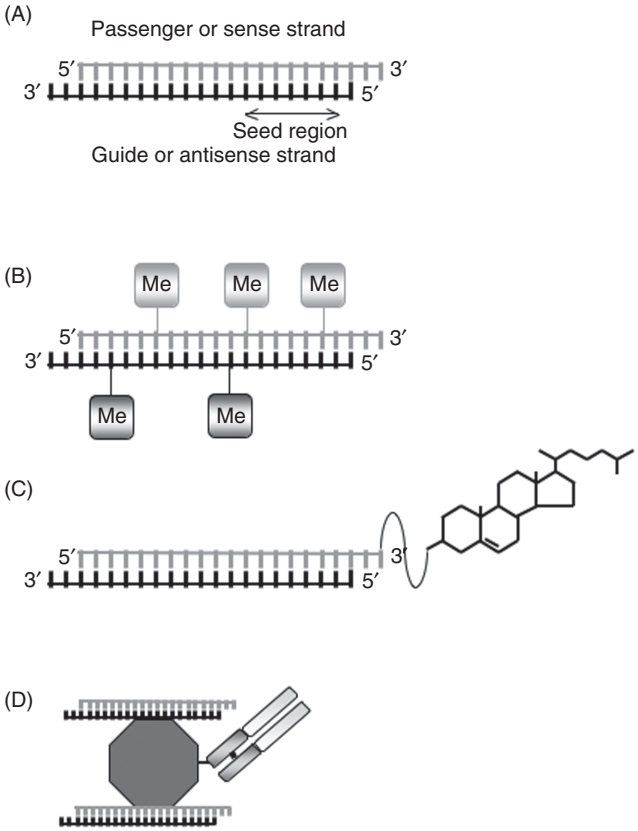


**Figure 2** Examples of siRNA modifications/delivery systems. (A) Unmodified siRNA; (B) modified nucleotides in the 2′ position of the pentose sugar stabilizes the dsRNA in vivo, making it less susceptible to be degraded by nucleases. In the example, 2′-*O*-methyl modifications have been incorporated. The optimal number of nucleotides that are modified needs to be tested by the investigator. (C) Cholesterol-conjugated siRNA provides a way to enhance uptake by tissues. (D) Protamine-complexed siRNA. An antibody fragment that recognizes a cell-specific receptor can be incorporated for targeted delivery.

## 2.2 siRNA and shRNA

The discovery of RNAi as a cellular pathway to modulate expression of endogenous genes was exploited to generate small interfering RNA (siRNA) to silence a gene of interest. siRNAs are designed to perfectly match the target sequence. The sense and antisense strands are chemically synthesized and after annealing, a double-stranded RNA with 2-nt 3′ overhangs is generated. Thus, the structure is similar to that of miRNA generated by Dicer (Fig. 1). Upon entering the cell, siRNA molecules are processed by RISC in the cytoplasm, which cleaves the target mRNA between nucleotides 10 and 11 of the guide strand, relative to its 5′ end [43].

siRNA can block gene expression for a few days, and readministration is necessary for a sustained silencing effect. Alternatively, short hairpin RNA (shRNA) can be used to provide a continuous source of silencing molecules. To generate shRNA, expression cassettes are engineered using RNA polymerase II or RNA polymerase III promoters. The sequences are usually designed for perfect complementarity to the mRNA. The shRNA is transcribed in the nucleus as a linear molecule that folds to generate a stem of approximately 21 nt, connected with a loop sequence of variable length. Thus, shRNAs have a similar structure to pre-miRNA and are transported to the cytoplasm similarly to pre-miRNA, that is, through the exportin-5 pathway. Once in the cytoplasm, shRNAs are processed by Dicer and subsequently by RISC, inducing endonucleolytic degradation of the target mRNA, as described for siRNA. shRNA expression cassettes can be delivered to cells

using plasmids or viral vectors such as lentivirus, adenovirus, or adeno-associated virus (AAV).

## 3 OFF-TARGET AND UNSPECIFIC EFFECTS

### 3.1 Off-Target Effects

An unwanted consequence of RNAi is the silencing of genes other than the one intended to target (known as off-target effects). High concentrations of siRNA lead to dose-dependent nonspecific gene downregulation patterns [129, 151] and toxicity [44]. These effects are sequence dependent in addition to concentration dependent. As little as 11 consecutive nucleotides matching an mRNA can result in cleavage [67]. In a recent study no correlation was found between the extent of target silencing and the number of off-target genes affected [67]. The UGGC motif, in particular, was found to be toxic and, therefore, should be avoided in any siRNA or shRNA construct [44]. The seed region of the RISC-entering strand should not overlap with the 3′ UTR of untargeted mRNA or silencing may take place [9]. In a similar manner, RNAi can block translation with as little as three to four mismatches (G-U "wobble") in the coding regions of mRNA [144]. Indeed, this mechanism is used to silence genes via the endogenous miRNA processing pathway in animals [73]. Matches in the seed region are more critical when using endogenously expressed shRNA as siRNA can be chemically modified with a 2′-*O*-methyl substitution at position 2 of the ribose in the guide strand to avoid off-target effects [68].

### 3.2 Unspecific Effects

In mammalian cells, a defense mechanism against viral infections is the recognition of foreign single-stranded (SS) or double-stranded RNA. Most cell types respond to the presence of foreign RNA by activating at least three RNA-binding proteins: dsRNA-dependent protein kinase (PKR), 2′,5′-oligoadenylate synthetases (OASs), and Toll-like receptors (TLRs) [157, 158]. The presence of dsRNA in the cell induces PKR autophosphorylation and subsequent phosphorylation of its substrates, resulting in general protein translation inhibition [162]. In addition, PKR activation initiates signal transduction events resulting in the production of interferons (IFNs), resulting in upregulation of interferon-stimulated genes (ISG), mediators of proapoptotic events. Interferons represent part of the innate immune system and act as a first line of defense against infections. Production of these cytokines is induced by pathogens such as bacteria, viruses, and fungi, and they function in a paracrine manner to protect cells from infection [7]. Two main families of interferons have been described: type I (IFN-α/β) and type II (IFN-γ). Type I interferons are induced by most cell types, in addition to plasmacytoid dendritic cells, leukocytes, and fibroblasts [82], while type II interferons are secreted by T lymphocytes and natural killer cells [162].

A second RNA sensor is 2′,5′-oligoadenylate synthetase (OAS), which is activated by dsRNA, inducing the conversion of adenosine 5′-triphosphate (ATP) into short 2′,5′-oligoadenylates that bind to and activate RNase L. This leads to cleavage of cellular and viral RNA, resulting in inhibition of protein translation [89, 162]. The third RNA sensor involves Toll-like receptors, signaling proteins expressed in the cell membrane or in intracellular membranes [152]. Each of the 11 members of this family responds to a specific ligand. Detection of foreign RNAs in the endosomes/plasma membrane are mediated by Toll-like receptor-3 (TLR3) [4, 71] and Toll-like receptor-7 (TLR7), and the downstream effect is the activation of IFNα/β production [38]. Generation of the signaling cascade leading to IFN expression can be mediated by other cytosolic ssRNA/dsRNA sensors, including retinoic acid inducible gene 1 (RIG1) and melanoma differentiation associated gene-5 (MDA-5), which detect foreign RNAs in the cytosol and initiate a signal to express IFNs [106].

Initial studies using small interfering RNA intermediates of 21–23 nt were found to bypass these responses [42] and were used by multiple investigators to study the effects of inhibiting gene expression in cell culture [6, 16, 42, 164] as well as in vivo [62, 76, 88, 100, 103, 181]. As a result of the increasing number of research applications using RNA interference, it has recently become clear that siRNA molecules have the capacity to induce unspecific effects. Indeed, multiple investigators have observed stimulation of interferon-induced genes using both siRNA [70, 71, 156] and shRNA [15, 47, 71, 127]. This discrepancy is likely because the early reports centered on PKR and predated the characterization of other important RNA-sensing proteins (i.e., RIG-1, MDA-5, TLRs) [48]. Although some controversy exists on whether or not siRNA/shRNA activate the complete interferon pathway [147], it is apparent that some downstream effectors/genes are shared [129]. Sequence-dependent mechanisms of interferon response induction exist. For example, activation of TLR7 in plasmacytoid dendritic cells involves a GUCCUUCAA motif [63]. A similar study found the UGUGU motif to be a potent inducer of the interferon response, although the role of TLR7 was not explored [70]. Therefore, both motifs should be avoided when using shRNA in vivo. With siRNA, it was found that incorporating 2′-*O*-methyl (2′O-Me) uridine or guanosine nucleosides into the antisense strand prevented an interferon response [69].

A major limitation in the use of shRNA-based vectors in vivo will be the potential for interferon response induction. When using helper-dependent adenovirus, we have found that expression of shRNA in liver requires careful titration of the dose to avoid the interferon response [177]. In general, in vivo studies using shRNA should include a pilot study to determine optimum vector doses in the tissue of interest that lead to the highest silencing without interferon induction. Analysis of the target tissue can be performed by examining the tissue histologically and looking for signs of inflammation and apoptosis. Also, the levels of known interferon-stimulated genes, such as OAS [157] and interferon-stimulated gene 20 (ISG20) [115], can be evaluated by Northern blot or quantitative real-time PCR.

Another issue specific to the use of shRNA constructs involves the protein exportin-5, which mediates the transport of unprocessed miRNA and shRNA out of the nucleus [187]. This pathway can become saturated when high levels of shRNA are expressed leading to cellular toxicity [186]. A study performed in mice found that saturation of exportin-5 was likely the cause of severe morbidity and mortality when shRNA was expressed in the liver. Longer hairpins [over 19 base pairs (bp)] tended to be more toxic [53]. To avoid this problem, the optimal dose of vector needs to be determined to reach efficient target silencing in the absence of unspecific events. Choosing the construct with the highest efficacy in vitro will facilitate reducing the dose. Alternatively, shRNA expression could be controlled with an inducible promoter, as discussed below. Monitoring the target tissue for levels of at least two abundantly expressed endogenous miRNAs (such as miR122 and let7a in liver), will help determine whether the amount of shRNA expressed results in exportin-5 saturation [177].

## 4 CHEMICALLY SYNTHESIZED SMALL INTERFERING RNA (siRNA)

### 4.1 siRNA Design

Conventionally designed siRNA are 21 nt with 2-nt 3′ overhangs. The design of the siRNA starts by selecting a suitable 19- to 21-nt sequence from the mRNA. The target transcript sequence can be obtained from the Entrez database provided by the National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov/Entrez). The open reading frame (ORF) is used to select sequences that will be targeted by siRNA. If the ORF is short, the 5′ and 3′ untranslated regions can also be included. Algorithms are currently available to choose a sequence that can potentially induce efficient gene silencing, and several companies (Ambion, Dharmacon, GenScript Corporation, Invitrogen) offer them online for custom design of optimal sequences. For the beginner, these are very useful tools, as the user can introduce the target mRNA sequence, and the system returns selected sequences based on their algorithms.

Multiple factors affect the potency of a silencing construct, including siRNA length, duplex stability, sequence asymmetry of the overhangs, as well as nucleotide composition in specific positions of the siRNA, with end stability being the most important feature [8, 128]. RISC preferentially loads the strand that displays the less stable 5′ end in the duplex [66]. Once the sequence to be targeted is selected, it is important to look for possible homologies to mRNAs other than the target. This is typically done using the Basic Local Alignment Search Tool (BLAST) (www.ncbi.nlm.nih.gov/blast/Blast.cgi), which compares a query sequence with a gene database. Sequences with the least number of matching nucleotides in genes are potentially less susceptible to give rise to off-target effects. Table 1 shows the considerations to be taken into account to generate effective siRNAs, based on current knowledge of critical elements that influence siRNA performance. For detailed information on each of the aspects affecting siRNA potency, an excellent protocol has been recently published by Birmingham and colleagues [8].

**TABLE 1 Considerations to Generate Potent siRNA**

| |
|---|
| Select all possible 19-nt to 21-nt sequences that are not located in areas with known polymorphisms. |
| Remove sequences that contain stretches with >5 single nucleotides, as these may lead to increased level of off-targeting [8]. |
| Remove sequences with a GC content > 64% (which affects helicase activity to unwind the dsRNA), or lower than 30% [8, 128]. |
| Remove sequences that may form internal hairpins (which may decrease silencing efficacy) and palindromes [126]. |
| Select sequences that have a thermodynamically favorable composition in the antisense strand. The strand with the less stable 5′ ends (usually A+U enriched) will be loaded into RISC [66, 72, 149]. |
| To enforce specificity, after a BLAST analysis, select sequences with at least three mismatches in positions 2–19 to undesired mRNAs. Avoid sequences with an antisense strand seed region that perfectly matches the 3′ UTR of mRNA. A Web-based algorithm to identify mRNAs with seed matches to the RNAi sequence is available from Dharmacon (http://www.dharmacon.com/seedlocator/default.aspx) [92, 128]. |
| Remove sequences with known toxic motifs that can trigger unspecific effects, such as 5′-GUCCUUCAA-3′ [63]. |

After several siRNA sequences have been identified, the investigator needs to test a minimum of four to determine their potency, as well as test for the presence of unspecific effects such as activation of the interferon response. This can be done by transfecting the siRNA in a cell line known to express the target gene. In some instances, though, this is not possible, as some cell lines are very difficult to transfect. A couple of approaches can be used in this case. One is to generate a construct consisting of the target sequence fused to the mRNA of a reporter protein, such as luciferase. This fusion construct can be co-transfected into HEK293 cells (which are easily transfected using commercially available reagents), together with the target-specific siRNA and a plasmid containing a control gene for normalization. The degree of luciferase activity allows determining the level of knock-down. Alternatively, it is also possible to generate an expression cassette with the comphementary DNA (cDNA) of the target gene, and quantify its mRNA or protein upon co-transfection with the siRNA [177]. Finally, the psiCHECK system (Promega, Madison, WI) can be used to not only test efficacy but also RISC selection of the sense and antisense strand. High levels of sense strand loading may decrease efficacy and, more importantly, lead to potential unspecific effects.

### 4.2 siRNA Stability

One of the major hurdles to effective RNAi is the limited stability of the siRNA molecules. In vitro, chemically synthesized, unmodified dsRNA have a half-life of 3–5 min [110]. These are suitable for most applications because cells quickly internalize the molecules. However, upon systemic injection into animals, nucleic acids are rapidly degraded by endo- and exonucleases. This limitation makes siRNA very ineffective for therapeutic purposes. Thus, the real challenge when it comes to a clinically relevant siRNA drug, is the delivery and stability of the siRNA molecule, rather than its efficacy at knocking down its target. Extensive research has been conducted to increase the stability of siRNAs so that they circulate for a longer time in blood and, therefore, allow for a more potent effect.

As described above, once the siRNA has entered a cell, unwinding by helicase activity takes place, and the sense or passenger strand is degraded while the antisense or guide strand is incorporated into RISC, mediating the silencing effect. The siRNA is designed such that the antisense strand is favored over the sense strand for loading onto RISC. It has been shown that single-stranded RNAs are less efficacious at silencing than dsRNA. Thus, the sense strand seems to provide stability to the molecule, making it less susceptible to degradation, and providing a template for degradation by RISC, which allows selection of the antisense strand to effect gene silencing. Because the process of mRNA recognition and cleavage is highly specific compared to strand selection for uploading into the RISC complex, it is more feasible to introduce changes in the sense strand than in the antisense strand. The position along the 21-nt sequence that is modified is also important, and perturbing certain nucleotides results in loss of functionality.

The starting point for chemical modifications has been incorporating backbone modifications (Fig. 2). For example, a phosphorothioate linkage provides exonuclease protection while retaining potency [25, 56]. A second possible modification is changing the functional group in position 2′ of the sugar ring. These modifications are introduced to protect the siRNA from endonucleases and are more likely to interfere with gene silencing efficacy. In one study, inclusion of 2′-fluoro substitutions protected the molecule against endonuclease activity and increased its stability in serum but did not result in a prolonged silencing effect [81]. In a related study, addition of 2′-fluoro substitutions in all pyrimidines (sense and antisense strands) in addition to other substitutions resulted in 900-fold increase in stability and higher potency in a mouse model of hepatitis B virus replication [110]. Addition of a 2′-*O*-methyl group in a limited number of nucleotides was also well tolerated and increased the performance of the siRNA [30, 61]. Other modifications that are possible in the 2′ position include locked nucleic acid (LNA), 2′-*O*-(2,4-dinitrophenyl) (2-DNP), 2′-deoxy, and 2′-*O*-(2-methoxylethyl) (2′-*O*-MOE), which are more extensively described in recent reviews [1, 192]. Progress in this area is slow because after testing in cells, it is necessary to try the modified siRNA in animal models in order to generate a therapeutic drug.

### 4.3 siRNA Delivery

Delivery of the siRNA drug remains the most challenging aspect of siRNA therapeutics. When systemically administered, siRNAs accumulate in blood, liver, lung, kidney, spleen, and intestine [12, 173]. The mechanism used by cells to take siRNA is not yet clear, but it is possible that it occurs through non-receptor-mediated endocytosis. Reaching the appropriate tissue can be accomplished by conjugating the siRNA molecule to specific ligands. The route of delivery is another important factor.

***Natural Ligands, Proteins, and Aptamers*** As mentioned above, naked siRNAs are highly unstable in blood and are degraded quickly by nucleases. siRNA can be conjugated with molecules that protect them

from degradation and at the same time allow targeted delivery. In general, terminal modifications affecting the 3′ end of the sense or antisense strand are well tolerated and do not interfere with the ability of the antisense strand to effect target mRNA degradation. Modifications in the 5′ end of the antisense strand are not well tolerated and usually reduce the silencing activity, while modifications in the 5′ of the sense strand do not affect siRNA activity. The most widely used siRNA conjugates used include natural ligands, peptides, and aptamers (RNA molecules that can bind specific receptors) [1].

It has been shown in mice that conjugation of siRNA to cholesterol to target ApoB is a feasible strategy to reduce its mRNA by 55% in the liver and 70% in the jejunum [161], resulting in a drop of plasma cholesterol of approximately 70%. Conjugation to cholesterol provided enhanced pharmacokinetics, presumably through binding to serum proteins [161]. Tissue biodistribution included liver, heart, kidney, adipose, and lung [161]. Wolfrum and colleagues [179] have shown that conjugation of siRNA to cholesterol as well as other lipophilic compounds, such as bile acids or fatty acids, results in uptake by cells expressing specific lipoprotein receptors. siRNA conjugated with high-density lipoproteins (HDL) were taken up preferentially by liver, kidney, adrenal gland, ovary, and intestine, in an SR-BI receptor-dependent manner, while low-density lipoproteins (LDL)-conjugated siRNA were mainly taken up by the liver, through the LDL receptor [179]. Interestingly, the uptake was independent from lipoprotein endocytosis, suggesting a model where the particles bind to the receptors, and uptake is dependent on a distinct transmembrane transporter [179].

The tolerance of the sense strand for chemical modifications has been exploited to conjugate siRNA to membrane permeant peptides, such as penetratin or transportan [112]. These conjugates reduce transgene expression in multiple cell lines. RNA aptamers have been used to direct cellular uptake in vitro and in animal models. For example, prostate cancer cells express high levels of prostate-specific membrane antigen (PSMA), a cell surface receptor. McNamara and colleagues showed that aptamer-conjugated siRNA to silence polo-like kinase 1 (PLK1) significantly reduced tumor size in a PMSA expression-dependent manner in a xenograft model of prostate cancer [105].

***Cationic Delivery Systems*** Cationic lipids, polymers, and dendrimers (branched polymers) have been used as vectors for siRNA delivery to tissues (also referred to as lipoplexes, polyplexes, and dendriplexes, respectively). These positively charged molecules facilitate complex formation with nucleic acids and interaction with the negatively charged cell membrane. Examples of cationic lipids that are commercially available are Lipofectamine (Invitrogen), Oligofectamine (Invitrogen), or TransIT TKO (Mirus Bio Corporation). Polymers used include polyethyleneimine (PEI), Arg-Gly-Asp (RGD) peptide, and protamine. siRNA uptake occurs through endosome formation. To facilitate endosomal escape, PEI or pH-sensitive peptides such as poly-histidine-lysine, have been included in the complex to facilitate endosomal disruption and release of the siRNA complex into the cytoplasm [2]. Complexes can be designed to include an antibody or other ligand for targeting to a specific tissue (Fig. 2). For example, a heavy-chain antibody fragment (Fab) fused to protamine has been complexed to siRNA to selectively target cells expressing human immunodeficiency virus (HIV-1) envelope protein gp160 [159]. In a mouse model of melanoma, a cocktail of Fab-conjugated siRNA against c-myc, MDM2, and vascular endothelial growth factor resulted in reduced tumor size only in cells expressing the glycoprotein gp160 [159].

The optimal delivery system should be one that results in highest RNAi efficacy at the lowest dose and does not induce off-target or unspecific effects. In this regard, a few studies have shown by microarray analysis that the transfection reagent per se can induce alterations in gene expression [119, 120]. For example, cationic lipids can generate such responses, in addition to enhancing immune responses triggered by siRNA [97]. Thus, when planning on using a delivery system, caution needs to be taken to avoid these potential unspecific events.

## 5 EXPRESSION OF SHORT HAIRPIN RNA (shRNA) CONSTRUCTS

### 5.1 Hairpin Design

Two possible approaches can be used to express shRNA. The first one consists of generating a structure that has a stem with complete homology to the target sequence [16, 121]. The second involves a longer structure, with the features of a microRNA [104, 155, 190] (Fig. 3). There is some discrepancy as to which system is more efficacious at knocking down expression [90, 155, 163]. Hence, choosing one over the other is a matter of personal preference.

A successful shRNA design should contain many of the features discussed in the siRNA section. These include avoidance of known toxic motifs and BLAST analysis to confirm target specificity. However, recent studies have found siRNA algorithms are only modestly successful at predicting functional shRNA constructs [90, 169]. Based on a set of 190 shRNAs to 19 targets, Li and colleagues [90] found several nucleotide

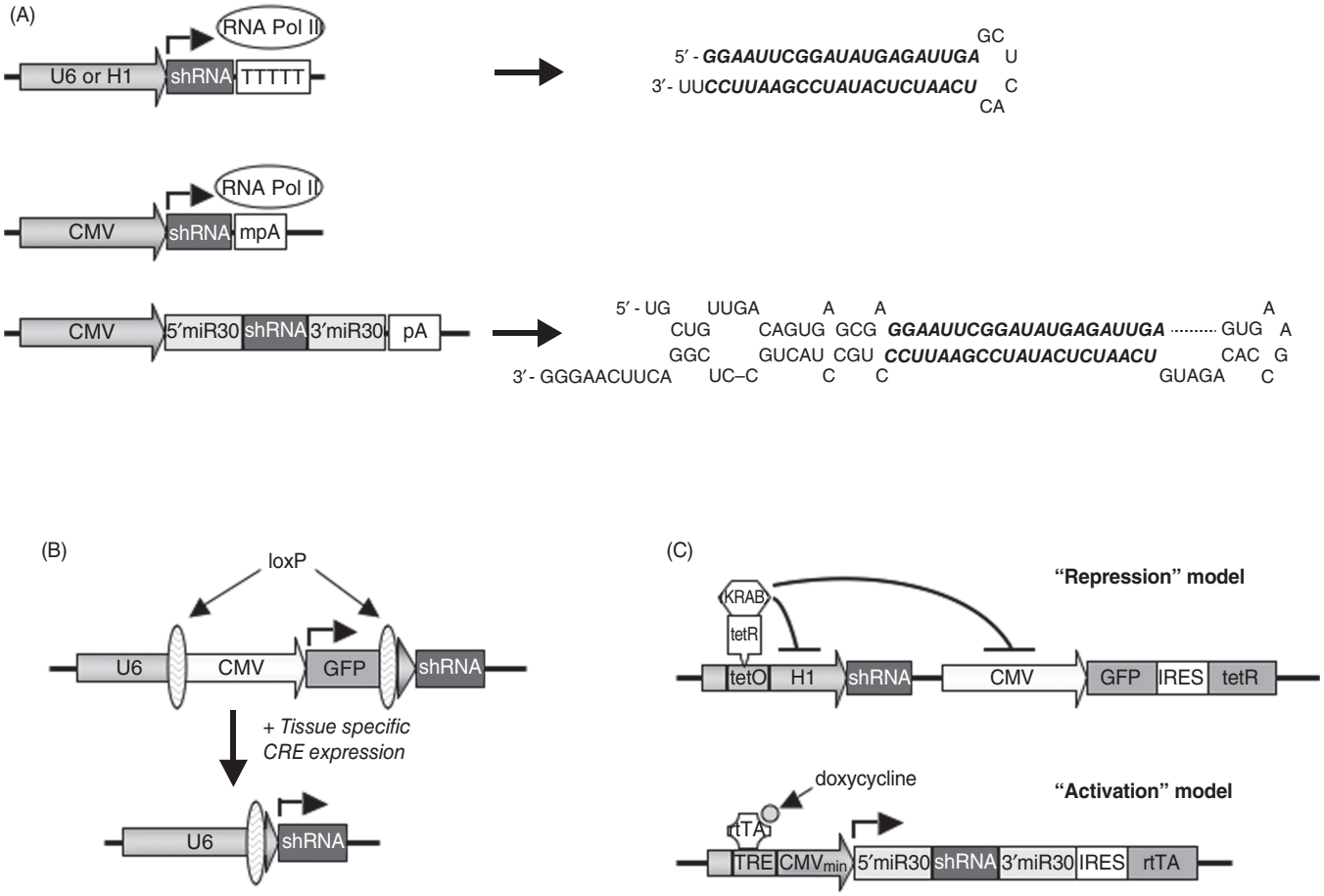


**Figure 3** Expression of shRNA using constitutive or inducible promoters. Example designs of shRNA constructs. (A) Three basic one-promoter systems using RNA polymerase II and III promoters. In the third construct, expression as part of the miR30 microRNA is accomplished by replacing the miR30 stem sequence with the shRNA. The shRNA sequence shown is for illustration purposes only; it has not been designed to knock down any particular gene. (B) Cre-mediated regulation of shRNA with pSico constructs in tissues expressing Cre recombinase. The loxP sites flank a stuffer sequence (in this case, GFP). Cre-mediated excision creates a functional U6 promoter allowing shRNA expression. Alternatively, the loxP sites may flank U6-shRNA instead of CMV-GFP for "off" regulation by Cre excision (pSicoR, not shown). (C) "Repression" model: regulation of shRNA expression by tet-inducible systems. A fusion protein of KRAB and tetR binding to tetO causes inhibition of shRNA and GFP expression. The presence of doxycycline releases KRAB-tetR (by binding to tetR) and permits transgene expression. "Activation" model: the presence of doxycycline induces rtTA binding and allows for transgene expression. See text for more detail. mpA, minimal polyadenylation signal; pA, polyadenylation signal; GFP, green fluorescent protein; IRES, internal ribosome entry site; tetR, tetracycline repressor; tetO, tet operator; TRE, tetracycline response element; rtTA, tetracycline transactivator; KRAB, Kruppel-associated box domain.

preferences in potent, functional shRNAs that may aid in hairpin design (Table 2). Additionally, there are other design parameters specific for shRNA constructs. The most widely used stem lengths are 19–21 bp to avoid potential toxic effects. In fact, it has been proposed that a single nucleotide mismatch created in the middle of the sense strand may further curb toxicity by limiting dsRNA length to ≤ 11 bp [29]. Loop lengths are 4–10 nt, although there is evidence that a 9-nt loop may yield a more potent shRNA in the context of a 19-bp stem [90]. If using a RNA polymerase III promoter (see below), a string of 4≤ thymidines should be avoided to prevent premature transcription termination.

Once shRNA hairpins are designed, it is necessary to test efficacy in vitro before proceeding to in vivo work. At least four or five candidates should be tested target-

**TABLE 2 Design Considerations for Effective shRNAs**

| |
|---|
| **Basic optimal parameters** |
| 19-nt stem [16, 53, 90, 139] |
| 9-nt loop [90] |
| 1-nt mismatch in position 10 of "sense" strand [29] |
| **Strong nucleotide preferences in stems of potent shRNAs** |
| Starting with G or GG [90] |
| GC-rich at positions 14–16 [90] |
| AU-rich at positions 3–7,17–19 [90] |
| **Known sequence motifs to avoid** |
| 4 ≤ thymidines (when using RNA polymerase III promoters) |
| 5′-UGGC-3′ (uncharacterized toxicity) [44] |
| 5′-GUCCUUCAA-3′ (shown to stimulate the interferon response) [63] |
| 5′-UGUGU-3′ (shown to stimulate the interferon response) [70] |

ing different sites of the mRNA transcript. The methods to accomplish this are identical to those described to test siRNA potency.

### 5.2 Promoter Selection

The in vitro expression of shRNA from both RNA polymerase II [35, 190] and III [16, 121, 164] promoters has been reported. The RNA polymerase III U6 and H1 are the most commonly used promoters as they possess a number of advantages. First, these ubiquitous promoters allow for a high level of expression in most cell types. Second, U6 and H1 contain well-defined transcriptional start sites and stop signals (four to six consecutive thymidines) that yield a 2-nt, 3′ overhang. Lastly, both have been used successfully to express shRNA in vivo [21, 36, 65, 117, 177, 184]. One study directly compared U6 and H1 silencing efficiency in mouse brain and found that U6 was superior [98]. The disadvantage of using RNA polymerase III promoters is the lack of tissue-specific expression. However, this problem may be circumvented by appropriate vector selection or use of a *Cre*-mediated activation or deactivation strategy [154, 171, 175]. For example, the pSico vector system is designed such that loxP sites flank a green fluorescent protein (GFP) expression cassette, located 5′ to the shRNA sequence (Fig. 3) [175]. When used in transgenic mice expressing *Cre* recombinase driven by a tissue-specific promoter, shRNA expression is exclusive to this tissue [175].

Transcripts generated by RNA polymerase II are structurally more complex than those generated by RNA polymerase III, as they include 5′ end caps and 3′ polyadenylated tails, which could potentially interfere with processing of the shRNA. However, discovering that cellular microRNA are transcribed from RNA polymerase II promoters, opened the possibility to use this type of promoter for expression of shRNA. The cytomegalovirus (CMV) promoter has been used to express shRNA directly, using a minimal polyA sequence [181], and in the context of the miR30 transcript [163, 189, 191]. Replacing CMV with a tissue-specific promoter is a tangible design. Experimental design may require temporal silencing of a gene to study functionality. To this end, an inducible construct to express shRNA can be used. Most inducible shRNA vectors used in vivo are responsive to ecdysone [135] or are based on the *Escherichia coli* tetracycline resistance operon. In this case, expression of shRNA is controlled by treatment with a tetracycline derivative (usually doxycycline) in several possible ways, depending on vector design. In the "repression" model, the placement of a tetracycline operator (tetO) in the shRNA promoter allows for the binding of a tetracycline repressor protein (tetR) and, consequently, transcription is inhibited in the absence of doxycycline [24, 31, 101, 102, 174]. However, this design has been shown to exhibit leaky expression in a number of studies [60, 93, 102]. Inserting a second tetO sequence [60] or using a codon-optimized tetR gene [150] appears to abrogate the leaky expression. A further adjustment to the "repression" model involves fusing the Kruppel-associated box (KRAB) domain to tetR [165, 178] (Fig. 3). The KRAB domain induces the local formation of heterochromatin, effectively blocking gene transcription with the added benefit of tighter expression compared to tetR-only vectors. Since this blockage affects both RNA polymerase II and III promoters, an internal marker, such as GFP, can be included in the vector for monitoring shRNA expression [178]. More importantly, any protein (i.e., a mutated variant) could be expressed concomitant with the shRNA to examine more complex conditions [178].

An alternative tetracycline "activation" model can be used where instead of tetR, a reverse tetracycline transactivator (rtTA) is expressed. Binding of rtTA to a tet-responsive promoter element (TRE) induces shRNA expression in the presence of doxycycline [37, 132, 163]. However, activation model constructs may suffer from greater leakiness when compared to repression model constructs [132].

### 5.3 Vector Selection

Delivery of the shRNA-expressing construct to the cell type or tissue of interest requires careful consideration. While in vitro work can utilize chemical transfection reagents and plasmid DNA, researchers have turned to viral vectors as a convenient delivery method for use in vivo. The most popular and well-characterized viral vectors are discussed here (Table 3).

**TABLE 3 Vector Systems for shRNA Expression at a Glance**

| Vector | Major Advantages | Major Disadvantages | In vivo Application Reference |
|---|---|---|---|
| E1-deleted adenovirus | Rapid production | High immunogenicity in vivo | 114, 143, 168, 181 |
| Helper-dependent adenovirus (HDAd) | Low toxicity *in* vivo<br>High cloning capacity | Difficult production methods | 177 |
| Adeno-associated virus (AAV) | Low toxicity in vivo<br>Wide range of serotypes for tissue specificity and repeated treatments | Lag in shRNA expression<br>Toxicity has been observed when expressing shRNA in liver | 21, 53, 125, 176 |
| Lentivirus | Integration of virus genome into the cell genome<br>Cell division is not necessary for transduction | Integration-induced mutagenesis | 11, 27, 33, 49, 77, 142 |

Adeno-associated viruses (AAV) are small, single-stranded DNA viruses highly used in the field of gene therapy [19]. While the best characterized serotype, AAV2, transduces a wide range of cell types, several other serotypes were shown to transduce certain specific tissues with greater efficiency. For detailed information on AAV biology and tissue specificity, we recommend reading recent reviews [52, 180]. As is the case with most viral vectors, true tissue-specific expression requires the use of a construct containing a tissue-specific promoter. Site-specific injection of the optimal AAV serotype is also an option and works well in the context of a self-contained organ, such as the eye or brain, limiting vector distribution. Disadvantages of this vector in the context of shRNA expression include a significant delay in transgene expression on the order of weeks—depending on serotype [180]. Also, AAV is known to form high-molecular-weight DNA concatemers [20]. This property could make it difficult to determine optimal vector doses and result in toxicity as extremely high amounts of shRNA are expressed. Nevertheless, AAV shRNA-expressing vectors have been used successfully in a number of in vivo studies [21, 53, 125, 176].

Adenovirus vectors have the capability of transducing a wide variety of cell types and are easily produced [123]. Earlier first-generation versions of adenovirus vectors were engineered with deletions in E1 and E3 genes. However, low levels of remnant viral gene expression lead to cytotoxicity and a cellular immune response, which precluded the use of these vectors for long-term studies in vivo [32, 108, 116]. A vast improvement in toxicity was seen when all viral genes were deleted, leaving only the inverted terminal repeats (ITRs) and packaging sequence. These vectors became known as helper-dependent (HDAd), high-capacity, or "gutless" adenovirus and were shown in numerous in vivo studies to yield long-term transgene expression [109, 111, 118, 137, 145]. Although highly efficient transduction of the liver is achieved with intravenous injection, localized applications of HDAd in the brain [170, 183, 194], muscle [26, 41], and lung [172] have also seen reasonable success. A few studies have shown it is possible to obtain a significant level of gene silencing using first-generation vectors [114, 143, 168, 181]. HDAd remains a better choice for expressing shRNA constructs in vivo [65, 177], and we have successfully used this system in liver [177]. We do not recommend the use of first-generation adenovirus vectors to express shRNA in vivo. In addition to the toxicity and lack of persistence in tissues associated with this vector system, first-generation vectors express VA1 noncoding RNA known to interfere with the RNAi processing pathway [95].

Lentiviruses, a member of the retrovirus family, have become useful tools for gene delivery. Transduction of both dividing and nondividing cells are possible [74, 113]—particularly when pseudotyped with the vesicular stomatitis virus G (VSV-G) envelope [3]. Other pseudotypes with improved tissue specificity are also available but are not as well characterized (reviewed in [28]). Viral-mediated cellular toxicity is largely not an issue as lentivirus vectors do not express viral proteins. Lentiviruses have been used in conjunction with shRNA for in vivo [23, 27, 49, 98] and ex vivo [11, 117] experiments. In contrast to AAV and adenovirus vectors, lentivirus vectors efficiently integrate into the host cell genome and do not require cell division for integration. This characteristic not only eliminates vector dilution effects by cell division but allows for an alternative to traditional knockouts. Several examples exist of success using the lentivirus/shRNA as a system to generate transgenic mice with silenced genes [33, 77, 142, 171, 182]. However, potential difficulties with epigenetic vector silencing and genetic mosaicism may impact the usefulness of this strategy [124].

## 6 APPLICATIONS

### 6.1 Treatment of Human Diseases

The field of RNAi therapeutics is advancing at a very rapid pace, and there are currently several clinical trials

at either phase I or phase II/III (for detailed information visit www.clinicaltrial.gov). Other treatments are at a preclinical stage, and proof of concept has been provided from studies in animal models.

***Infectious Diseases*** In therapies aimed at treating viral diseases, the siRNA is directed against a viral RNA transcript, replication intermediate, or a cellular component necessary for the virus to replicate. Clinical trials have started for the treatment of respiratory syncytial virus (RSV). Alnylam Pharmaceuticals has recently finished a phase II study in approximately 88 volunteers with experimental RSV infection (i.e., patients were infected with a clinical strain of the the virus). ALN-RSV01, an siRNA to target a component of the nucleocapsid of RSV, was administered intranasally once a day for 5 days. Antiviral activity in the absence of toxicity was observed (press release). A phase II study in patients with natural infection with RSV is planned for the near future. Another clinical trial to be started in the near future involves a combinatorial therapy using lentivirus vectors to treat HIV infection [91, 141]. In this ex vivo approach, hematopoietic progenitor cells are transduced with a lentivirus expressing an shRNA directed against tat/rev (essential proteins for virus replication), a ribozyme against the cellular co-receptor CCR5, and a nucleolar localizing transactivation responsive (TAR) RNA decoy [91].

***Central Nervous System (CNS) Diseases*** Using viral vectors, several studies in mice have shown considerable success at silencing the genes involved in gain-of-function conditions such as Huntington's disease and familial amyotrophic lateral sclerosis (ALS) [58, 107, 133, 136]. Huntingon's disease is a neurodegenerative disorder caused by the presence of polyglutamine repeat expansions in the huntingtin protein, resulting in a gain of function. Harper and colleagues have used AAV1 to knock down a mutant form of huntingtin in mice, which led to improved behavioral and neuropathological abnormalities [58]. Similarly, lentiviral vector-mediated shRNA expression in animal models of ALS, a disease caused by mutations in the superoxide dismutase gene, has also proved the concept that RNAi may represent a valuable treatment for this gain-of-function disease [133, 136]. Although no clinical trials have started yet, Alnylam Pharmaceuticals has conducted preclinical studies in mice showing that cholesterol-conjugated siRNA to target huntingtin by intrastriatum infusion is a valid approach to attenuate the neuronal and behavioral abnormalities characteristic of the disease [39].

Delivery of siRNA to the CNS, however, remains a challenge. Several groups have shown it is possible to silence a gene through direct injection of siRNA in the brain. There have been efforts to administer the siRNA by stereotaxic surgery, but usually this route results in transfection being restricted to the area close to the injection site. Nevertheless, this route is unacceptably invasive for the treatment of human conditions. Intrathecal siRNA infusion has been used successfully to reduce levels of the pain-related cation channel $P2X_3$, reducing the response to pain [40]. Intravenous administration represents a more viable delivery route, but poses the challenge of crossing the blood–brain barrier. A possibility to overcome this limitation is to take advantage of the fact that neurotropic viruses can cross the blood–brain barrier. Peptides from their envelopes can be used for targeted siRNA delivery. A recent study has shown that this is a feasible approach by using siRNA complexed to a cell-penetrating peptide and a rabies virus glycoprotein (RVG) peptide. The RVG peptide confers neuron specificity by binding to acetylcholine receptors, while the cell-penetrating peptide provides a way to bind the siRNA through charge interaction. The siRNA was found widely distributed in all areas of the brain and was not detectable in liver or spleen. Approximately 50% gene silencing was observed [79].

***Age-Related Macular Degeneration*** Age-related macular degeneration (AMD) is a disease for which gene therapy approaches had already been pursued, and some of the molecular pathways involved were relatively well known. AMD is characterized by damage to the central area of the retina (macula), resulting in loss of vision. Abnormal growth of vessels causes the less common form of the condition, called wet AMD. Opko Health, Inc., developed Cand5 (Bevasiranib), a siRNA to target vascular endothelial growth factor (VEGF), a molecule involved in neovascularization of the retina. The reagent is administered by direct injection into the vitreous. Cand5 displayed good safety profiles and signs of biological activity in phase II trials in patients with AMD and in patients with diabetic macular edema (Opko Health, Inc., press release). A second siRNA to target the VEGF pathway is Sirna-027 [153], which was developed by Sirna Therapeutics (currently owned by Merck & Co., Inc.). This siRNA targets VEGF receptor 1 (VEGFR1) and has been tested in phase I trials. Eight weeks after one single injection of the drug, 100% of patients showed visual acuity stabilization and a few patients had improved visual acuity (press release). A phase II trial demonstrated good efficacy (clinicaltrial.gov ID: NCT00363714). A third product, RTP-801i (Silence Therapeutics, Inc.), is currently being tested in a phase I trial (press release). Its target is the hypoxia-induced gene DNA damage-inducible transcript 4 (*DDIT4*) [13].

***Cardiovascular Diseases*** siRNA have the potential to become drugs for the treatment of hypercholesterolimia. In one study, administration of a cholesterol-conjugated siRNA to target apolipoprotein B (ApoB) was infused intravenously in mice. The siRNA conjugate was widely distributed and was found in liver, kidney, heart, adipose tissue, and lung [161]. In the study, it was shown that ApoB mRNA was decreased in liver and jejunum, and that ApoB protein in plasma was lower. As a consequence of the RNAi treatment, mice had a 50% decrease in chylomicron levels, a 40% reduction in the LDL concentration and a 25% decrease in HDL levels, together with a reduction in total plasma cholesterol. No effect in very low density lipoproteins (VLDL) was observed [161]. In a follow-up study, liposome-conjugated siRNA was delivered intravenously to nonhuman primates. A dose-dependent reduction in ApoB mRNA levels was observed, concurrent to a decrease in serum cholesterol and LDLs [193]. One dose was sufficient to silence ApoB for 11 days.

***Cancer*** siRNA drugs can be developed to interfere with tumor biology in multiple ways. One of the most pursued approaches has been to obstruct the angiogenesis process by targeting VEGF or its receptor, to reduce formation of blood vessels, necessary for tumor growth. In one study, an siRNA to VEGF receptor 2 (VEGF R2) complexed to an RGD peptide was delivered systemically. The RGD motif binds to integrins in endothelial cells of the vasculature, and the siRNA inhibits expression of the VEGF receptor. This approach resulted in selective tumor uptake and reduction of neovasculature and tumor size [146]. In a second study, the siRNA was directed to VEGF, and it was tested in an animal model of prostate cancer. The siRNA was conjugated to atelocollagen, which was expected to protect it against nucleases, prolonging the circulation time and increasing uptake. The siRNA was given intratumorally every 10 days at the initial stages of tumor development. At day 50, tumors were significantly smaller and had much lower levels of vessel density than those treated with a control siRNA [166]. In a more recent study, Kim and colleagues showed that anti-VEGF siRNA conjugated to cholesteryl oligo-D-arginine reduced tumor and neovascularization in a model of colon adenocarcinoma [75]. The use of siRNA as a tool to prevent the onset of metastasis has been addressed in a study to knock down the Ewing sarcoma breakpoint region 1-Friend leukemia integration 1 transcription factor (EWS-FLI1) fusion gene product in a mouse model of metastatic Ewing's sarcoma [64].

Another approach has been to target genes that are overexpressed in tumor cells. One example includes generating an aptamer-siRNA chimera to target prostate cancer cells. The aptamer portion bound the prostate-specific membrane antigen (PSMA), while the siRNA was designed to target the genes polo-like kinase 1 (PLK1) or the proto-oncogene B-cell leukemia/lymphoma 2 (BCL2), which are overexpressed in tumors [105]. In a xenograft model of prostate cancer, the chimeras prevented tumor growth in PSMA-positive cells but not in PSMA-negative cells [105]. Glioblastoma multiforme is one of the most aggressive brain tumors, with a poor prognosis. The growth factor pleiotrophin is overexpressed in these tumors, which can be used as a siRNA drug target. In a mouse model of brain cancer, administration of a polyethyleneimine-complexed siRNA against pleiotrophin reduced the tumor size by 50% [54]. Finally, treatment with chitosan-coated siRNA to target a low-molecular-weight GTP/GDP-binding GTPase, Ras homologous A (RhoA), reduced the tumor size by 90% in a model of breast cancer [131].

### 6.2 Gene Function Studies and Drug Target Identification

RNAi has become the preferred tool for identification of new targets and to study gene function. Traditional technologies to validate the role of genes in disease involve the development of animal models with null alleles for the target gene. In some instances, absence of a gene is incompatible with normal development, leading to death in utero. In addition, for particular studies it may be necessary to knock down gene function in the adult animal (instead of at the stem cell level), thereby providing a model similar to the human condition. One example of such an application would be to study the effect of reducing a gene product that is known to be upregulated in animal models of diet-induced obesity.

The limitations of using RNAi for gene function studies and target validation are similar to those described for the use of this technology as a therapeutic treatment, that is, presence of unspecific and off-target effects that may mask the real impact of knock-down on the gene of interest. Induction of unspecific effects such as the interferon response, is siRNA concentration dependent [129], and cell type dependent [139]. Therefore, data generated in one tissue may not apply to another, and pilot studies should be conducted in the tissue of interest. Unintended targeting can be mediated by any of the dsRNA strands that are generated from shRNA processing. Thus, carrying out experiments with more than one siRNA/shRNA sequence is desirable, to confirm that the impact of silencing the gene of interest in the cell/animal is due to on-target rather than off-target effects.

Numerous gene function studies have been conducted using RNAi. For example, to investigate the role of the gene defect in acute myeloid leukemia (AML), Heidenreich and colleagues used an RNAi-mediated approach [59]. Translocation between chromosomes 8 and 21 results in the fusion gene *AML1/MTG8*. By designing an siRNA to specifically target the fusion protein (leaving the wild-type protein unaffected), the authors were able to show that leukemic cells transfected with siRNA to the fusion protein were susceptible to the action of tumor growth factor β1, inducing cell differentiation, and providing evidence for a role for *AML1/MTG8* in preventing differentiation and promoting cell division [59].

RNAi is also being used in target validation studies. Brusselmans and colleagues explored the possibility of using acetyl-CoA carboxylase α (ACC-α) as the target for prostate cancer treatment, taking advantage of the fact that de novo lipogenesis is upregulated in many cancer cell types, while noncancerous cells have low levels of the enzymes involved in this process [17]. Treatment of LNCaP cancer cells with a siRNA against ACC-α resulted in reduced phospholipid and triglyceride synthesis levels, followed by subsequent induction of caspase-mediated apoptosis. Although the siRNA molecule could also induce gene silencing of the target genes in normal fibroblasts, apoptosis was not observed in these cells [17], suggesting that ACC-α is a good target for treatment of prostate cancer. Thus, RNAi can be used to conduct studies testing the role of genes in disease and simultaneously provide the proof of concept for target validation.

## ACKNOWLEDGMENTS

Our research is funded by grants from the NIDDK (DK069432-01 and DK078595), American Diabetes Association (1-08-RA-135), INGEN (Indiana Genomics Initiative of Indiana University supported in part by Lilly Endowment Inc.), and by the American Heart Association (fellowship to SRW).

## REFERENCES

1. Akhtar S, Benter IF (2007). Nonviral delivery of synthetic siRNAs in vivo. *J Clin Invest 117*, 3623–3632.
2. Akinc A, Thomas M, Klibanov AM, Langer R (2005). Exploring polyethylenimine-mediated DNA transfection and the proton sponge hypothesis. *J Gene Med 7*, 657–663.
3. Akkina RK, Walton RM, Chen ML, Li QX, Planelles V, Chen IS (1996). High-efficiency gene transfer into CD34+ cells with a human immunodeficiency virus type 1-based retroviral vector pseudotyped with vesicular stomatitis virus envelope glycoprotein G. *J Virol 70*, 2581–2585.
4. Alexopoulou L, Holt AC, Medzhitov R, Flavell RA (2001). Recognition of double-stranded RNA and activation of NF-kappaB by Toll-like receptor 3. *Nature 413*, 732–738.
5. Ambros V (2000). Control of developmental timing in *Caenorhabditis elegans*. *Curr Opin Genet Dev 10*, 428–433.
6. Arts GJ, Langemeijer E, Tissingh R, Ma L, Pavliska H, Dokic K, Dooijes R, Mesic E, Clasen R, Michiels F, van der Schueren J, Lambrecht M, Herman S, Brys R, Thys K, Hoffmann M, Tomme P, van Es H (2003). Adenoviral vectors expressing siRNAs for discovery and validation of gene function. *Genome Res 13*, 2325–2332.
7. Barber GN (2005). The dsRNA-dependent protein kinase, PKR and cell death. *Cell Death Differ 12*, 563–570.
8. Birmingham A, Anderson E, Sullivan K, Reynolds A, Boese Q, Leake D, Karpilow J, Khvorova A (2007). A protocol for designing siRNAs with high functionality and specificity. *Nat Protoc 2*, 2068–2078.
9. Birmingham A, Anderson EM, Reynolds A, Ilsley-Tyree D, Leake D, Fedorov Y, Baskerville S, Maksimova E, Robinson K, Karpilow J, Marshall WS, Khvorova A (2006). 3′ UTR seed matches, but not overall identity, are associated with RNAi off-targets. *Nat Methods 3*, 199–204.
10. Borchert GM, Lanier W, Davidson BL (2006). RNA polymerase III transcribes human microRNAs. *Nat Struct Mol Biol 13*, 1097–1101.
11. Bot I, Guo J, Van Eck M, Van Santbrink PJ, Groot PH, Hildebrand RB, Seppen J, Van Berkel TJ, Biessen EA (2005). Lentiviral shRNA silencing of murine bone marrow cell CCR2 leads to persistent knockdown of CCR2 function in vivo. *Blood 106*, 1147–1153.
12. Braasch DA, Paroo Z, Constantinescu A, Ren G, Oz OK, Mason RP, Corey DR (2004). Biodistribution of phosphodiester and phosphorothioate siRNA. *Bioorg Med Chem Lett 14*, 1139–1143.
13. Brafman A, Mett I, Shafir M, Gottlieb H, Damari G, Gozlan-Kelner S, Vishnevskia-Dai V, Skaliter R, Einat P, Faerman A, Feinstein E, Shoshani T (2004). Inhibition of oxygen-induced retinopathy in RTP801-deficient mice. *Invest Ophthalmol Vis Sci 45*, 3796–3805.
14. Brennecke J, Hipfner DR, Stark A, Russell RB, Cohen SM (2003). Bantam encodes a developmentally regulated microRNA that controls cell proliferation and regulates the proapoptotic gene hid in *Drosophila*. *Cell 113*, 25–36.
15. Bridge AJ, Pebernard S, Ducraux A, Nicoulaz AL, Iggo R (2003). Induction of an interferon response by RNAi vectors in mammalian cells. *Nat Genet 34*, 263–264.
16. Brummelkamp TR, Bernards R, Agami R (2002). A system for stable expression of short interfering RNAs in mammalian cells. *Science 296*, 550–553.

17. Brusselmans K, De Schrijver E, Verhoeven G, Swinnen JV (2005). RNA interference-mediated silencing of the acetyl-CoA-carboxylase-alpha gene induces growth inhibition and apoptosis of prostate cancer cells. *Cancer Res 65*, 6719–6725.
18. Cai X, Hagedorn CH, Cullen BR (2004). Human microRNAs are processed from capped, polyadenylated transcripts that can also function as mRNAs. *RNA 10*, 1957–1966.
19. Carter BJ (2005). Adeno-associated virus vectors in clinical trials. *Hum Gene Ther 16*, 541–550.
20. Carter BJ (2003). Gene delivery technology: Adeno-associated virus. *In* AaS Rolland SM, Ed. *Pharmaceutical Gene Delivery Systems*. New York: Marcel Dekker.
21. Chen CC, Ko TM, Ma HI, Wu HL, Xiao X, Li J, Chang CM, Wu PY, Chen CH, Han JM, Yu CP, Jeng KS, Hu CP, Tao MH (2007). Long-term inhibition of hepatitis B virus in transgenic mice by double-stranded adeno-associated virus 8-delivered short hairpin RNA. *Gene Ther 14*, 11–19.
22. Chen CZ, Li L, Lodish HF, Bartel DP (2004). MicroRNAs modulate hematopoietic lineage differentiation. *Science 303*, 83–86.
23. Chen Y, Lin MC, Yao H, Wang H, Zhang AQ, Yu J, Hui CK, Lau GK, He ML, Sung J, Kung HF (2007). Lentivirus-mediated RNA interference targeting enhancer of zeste homolog 2 inhibits hepatocellular carcinoma growth through down-regulation of stathmin. *Hepatology 46*, 200–208.
24. Chen Y, Stamatoyannopoulos G, Song CZ (2003). Down-regulation of CXCR4 by inducible small interfering RNA inhibits breast cancer cell invasion in vitro. *Cancer Res 63*, 4801–4804.
25. Choung S, Kim YJ, Kim S, Park HO, Choi YC (2006). Chemical modification of siRNAs to improve serum stability without loss of efficacy. *Biochem Biophys Res Commun 342*, 919–927.
26. Clemens PR, Kochanek S, Sunada Y, Chan S, Chen HH, Campbell KP, Caskey CT (1996). In vivo muscle gene transfer of full-length dystrophin with an adenoviral vector that lacks all viral genes. *Gene Ther 3*, 965–972.
27. Couturier C, Sarkis C, Seron K, Belouzard S, Chen P, Lenain A, Corset L, Dam J, Vauthier V, Dubart A, Mallet J, Froguel P, Rouille Y, Jockers R (2007). Silencing of OB-RGRP in mouse hypothalamic arcuate nucleus increases leptin receptor signaling and prevents diet-induced obesity. *Proc Natl Acad Sci USA 104*, 19476–19481.
28. Cronin J, Zhang XY, Reiser J (2005). Altering the tropism of lentiviral vectors through pseudotyping. *Curr Gene Ther 5*, 387–398.
29. Cullen BR (2006). Enhancing and confirming the specificity of RNAi experiments. *Nat Methods 3*, 677–681.
30. Czauderna F, Fechtner M, Dames S, Aygun H, Klippel A, Pronk GJ, Giese K, Kaufmann J (2003). Structural variations and stabilising modifications of synthetic siRNAs in mammalian cells. *Nucleic Acids Res 31*, 2705–2716.
31. Czauderna F, Santel A, Hinz M, Fechtner M, Durieux B, Fisch G, Leenders F, Arnold W, Giese K, Klippel A, Kaufmann J (2003). Inducible shRNA expression for application in a prostate cancer mouse model. *Nucleic Acids Res 31*, e127.
32. Dai Y, Schwarz EM, Gu D, Zhang WW, Sarvetnick N, Verma IM (1995). Cellular and humoral immune responses to adenoviral vectors containing factor IX gene: Tolerization of factor IX and vector antigens allows for long-term expression. *Proc Natl Acad Sci USA 92*, 1401–1405.
33. Dann CT, Alvarado AL, Hammer RE, Garbers DL (2006). Heritable and stable gene knockdown in rats. *Proc Natl Acad Sci USA 103*, 11246–11251.
34. Denli AM, Tops BB, Plasterk RH, Ketting RF, Hannon GJ (2004). Processing of primary microRNAs by the Microprocessor complex. *Nature 432*, 231–235.
35. Denti MA, Rosa A, Sthandier O, De Angelis FG, Bozzoni I (2004). A new vector, based on the PolII promoter of the U1 snRNA gene, for the expression of siRNAs in mammalian cells. *Mol Ther 10*, 191–199.
36. Dentin R, Benhamed F, Hainault I, Fauveau V, Foufelle F, Dyck JR, Girard J, Postic C (2006). Liver-specific inhibition of ChREBP improves hepatic steatosis and insulin resistance in ob/ob mice. *Diabetes 55*, 2159–2170.
37. Dickins RA, Hemann MT, Zilfou JT, Simpson DR, Ibarra I, Hannon GJ, Lowe SW (2005). Probing tumor phenotypes using stable and regulated synthetic microRNA precursors. *Nat Genet 37*, 1289–1295.
38. Diebold SS, Kaisho T, Hemmi H, Akira S, Reis e Sousa C (2004). Innate antiviral responses by means of TLR7-mediated recognition of single-stranded RNA. *Science 303*, 1529–1531.
39. DiFiglia M, Sena-Esteves M, Chase K, Sapp E, Pfister E, Sass M, Yoder J, Reeves P, Pandey RK, Rajeev KG, Manoharan M, Sah DW, Zamore PD, Aronin N (2007). Therapeutic silencing of mutant huntingtin with siRNA attenuates striatal and cortical neuropathology and behavioral deficits. *Proc Natl Acad Sci USA 104*, 17204–17209.
40. Dorn G, Patel S, Wotherspoon G, Hemmings-Mieszczak M, Barclay J, Natt FJ, Martin P, Bevan S, Fox A, Ganju P, Wishart W, Hall J (2004). siRNA relieves chronic neuropathic pain. *Nucleic Acids Res 32*, e49.
41. Dudley RW, Lu Y, Gilbert R, Matecki S, Nalbantoglu J, Petrof BJ, Karpati G (2004). Sustained improvement of muscle function one year after full-length dystrophin gene transfer into mdx mice by a gutted helper-dependent adenoviral vector. *Hum Gene Ther 15*, 145–156.
42. Elbashir SM, Harborth J, Lendeckel W, Yalcin A, Weber K, Tuschl T (2001). Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. *Nature 411*, 494–498.
43. Elbashir SM, Lendeckel W, Tuschl T (2001). RNA interference is mediated by 21- and 22-nucleotide RNAs. *Genes Dev 15*, 188–200.

44. Fedorov Y, Anderson EM, Birmingham A, Reynolds A, Karpilow J, Robinson K, Leake D, Marshall WS, Khvorova A (2006). Off-target effects by siRNA can induce toxic phenotype. *RNA 12*, 1188–1196.

45. Fire A, Albertson D, Harrison SW, Moerman DG (1991). Production of antisense RNA leads to effective and specific inhibition of gene expression in *C. elegans* muscle. *Development 113*, 503–514.

46. Fire A, Xu S, Montgomery MK, Kostas SA, Driver SE, Mello CC (1998). Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature 391*, 806–811.

47. Fish RJ, Kruithof EK (2004). Short-term cytotoxic effects and long-term instability of RNAi delivered using lentiviral vectors. *BMC Mol Biol 5*, 9.

48. Gantier MP, Williams BR (2007). The response of mammalian cells to double-stranded RNA. *Cytokine Growth Factor Rev 18*, 363–371.

49. Georgievska B, Jakobsson J, Persson E, Ericson C, Kirik D, Lundberg C (2004). Regulated delivery of glial cell line-derived neurotrophic factor into rat striatum, using a tetracycline-dependent lentiviral vector. *Hum Gene Ther 15*, 934–944.

50. Gregory RI, Chendrimada TP, Cooch N, Shiekhattar R (2005). Human RISC couples microRNA biogenesis and posttranscriptional gene silencing. *Cell 123*, 631–640.

51. Gregory RI, Yan KP, Amuthan G, Chendrimada T, Doratotaj B, Cooch N, Shiekhattar R (2004). The Microprocessor complex mediates the genesis of microRNAs. *Nature 432*, 235–240.

52. Grimm D, Kay MA (2003). From virus evolution to vector revolution: Use of naturally occurring serotypes of adeno-associated virus (AAV) as novel vectors for human gene therapy. *Curr Gene Ther 3*, 281–304.

53. Grimm D, Streetz KL, Jopling CL, Storm TA, Pandey K, Davis CR, Marion P, Salazar F, Kay MA (2006). Fatality in mice due to oversaturation of cellular microRNA/short hairpin RNA pathways. *Nature 441*, 537–541.

54. Grzelinski M, Urban-Klein B, Martens T, Lamszus K, Bakowsky U, Hobel S, Czubayko F, Aigner A (2006). RNA interference-mediated gene silencing of pleiotrophin through polyethylenimine-complexed small interfering RNAs in vivo exerts antitumoral effects in glioblastoma xenografts. *Hum Gene Ther 17*, 751–766.

55. Guo S, Kemphues KJ (1995). *par-1*, a gene required for establishing polarity in *C. elegans* embryos, encodes a putative Ser/Thr kinase that is asymmetrically distributed. *Cell 81*, 611–620.

56. Hall AH, Wan J, Spesock A, Sergueeva Z, Shaw BR, Alexander KA (2006). High potency silencing by single-stranded boranophosphate siRNA. *Nucleic Acids Res 34*, 2773–2781.

57. Hannon GJ (2002). RNA interference. *Nature 418*, 244–251.

58. Harper SQ, Staber PD, He X, Eliason SL, Martins IH, Mao Q, Yang L, Kotin RM, Paulson HL, Davidson BL (2005). RNA interference improves motor and neuropathological abnormalities in a Huntington's disease mouse model. *Proc Natl Acad Sci USA 102*, 5820–5825.

59. Heidenreich O, Krauter J, Riehle H, Hadwiger P, John M, Heil G, Vornlocher HP, Nordheim A (2003). AML1/MTG8 oncogene suppression by small interfering RNAs supports myeloid differentiation of t(8;21)-positive leukemic cells. *Blood 101*, 3157–3163.

60. Henriksen JR, Lokke C, Hammero M, Geerts D, Versteeg R, Flaegstad T, Einvik C (2007). Comparison of RNAi efficiency mediated by tetracycline-responsive H1 and U6 promoter variants in mammalian cell lines. *Nucleic Acids Res 35*, e67.

61. Holen T, Amarzguioui M, Babaie E, Prydz H (2003). Similar behaviour of single-strand and double-strand siRNAs suggests they act through a common RNAi pathway. *Nucleic Acids Res 31*, 2401–2407.

62. Hommel JD, Sears RM, Georgescu D, Simmons DL, DiLeone RJ (2003). Local gene knockdown in the brain using viral-mediated RNA interference. *Nat Med 9*, 1539–1544.

63. Hornung V, Guenthner-Biller M, Bourquin C, Ablasser A, Schlee M, Uematsu S, Noronha A, Manoharan M, Akira S, de Fougerolles A, Endres S, Hartmann G (2005). Sequence-specific potent induction of IFN-alpha by short interfering RNA in plasmacytoid dendritic cells through TLR7. *Nat Med 11*, 263–270.

64. Hu-Lieskovan S, Heidel JD, Bartlett DW, Davis ME, Triche TJ (2005). Sequence-specific knockdown of EWS-FLI1 by targeted, nonviral delivery of small interfering RNA inhibits tumor growth in a murine model of metastatic Ewing's sarcoma. *Cancer Res 65*, 8984–8992.

65. Huang B, Schiefer J, Sass C, Landwehrmeyer GB, Kosinski CM, Kochanek S (2007). High-capacity adenoviral vector-mediated reduction of huntingtin aggregate load in vitro and in vivo. *Hum Gene Ther 18*, 303–311.

66. Hutvagner G (2005). Small RNA asymmetry in RNAi: Function in RISC assembly and gene regulation. *FEBS Lett 579*, 5850–5857.

67. Jackson AL, Bartz SR, Schelter J, Kobayashi SV, Burchard J, Mao M, Li B, Cavet G, Linsley PS (2003). Expression profiling reveals off-target gene regulation by RNAi. *Nat Biotechnol 21*, 635–637.

68. Jackson AL, Burchard J, Leake D, Reynolds A, Schelter J, Guo J, Johnson JM, Lim L, Karpilow J, Nichols K, Marshall W, Khvorova A, Linsley PS (2006). Position-specific chemical modification of siRNAs reduces "off-target" transcript silencing. *RNA 12*, 1197–1205.

69. Judge AD, Bola G, Lee AC, MacLachlan I (2006). Design of noninflammatory synthetic siRNA mediating potent gene silencing in vivo. *Mol Ther 13*, 494–505.

70. Judge AD, Sood V, Shaw JR, Fang D, McClintock K, MacLachlan I (2005). Sequence-dependent stimulation of the mammalian innate immune response by synthetic siRNA. *Nat Biotechnol 23*, 457–462.

71. Kariko K, Bhuyan P, Capodici J, Weissman D (2004). Small interfering RNAs mediate sequence-independent gene suppression and induce immune activation by signaling through toll-like receptor 3. *J Immunol 172*, 6545–6549.
72. Khvorova A, Reynolds A, Jayasena SD (2003). Functional siRNAs and miRNAs exhibit strand bias. *Cell 115*, 209–216.
73. Kim VN (2005). MicroRNA biogenesis: Coordinated cropping and dicing. *Nat Rev Mol Cell Biol 6*, 376–385.
74. Kim VN, Mitrophanous K, Kingsman SM, Kingsman AJ (1998). Minimal requirement for a lentivirus vector based on human immunodeficiency virus type 1. *J Virol 72*, 811–816.
75. Kim WJ, Christensen LV, Jo S, Yockman JW, Jeong JH, Kim YH, Kim SW (2006). Cholesteryl oligoarginine delivering vascular endothelial growth factor siRNA effectively inhibits tumor growth in colon adenocarcinoma. *Mol Ther 14*, 343–350.
76. Kishida T, Asada H, Gojo S, Ohashi S, Shin-Ya M, Yasutomi K, Terauchi R, Takahashi KA, Kubo T, Imanishi J, Mazda O (2004). Sequence-specific gene silencing in murine muscle induced by electroporation-mediated transfer of short interfering RNA. *J Gene Med 6*, 105–110.
77. Kissler S, Stern P, Takahashi K, Hunter K, Peterson LB, Wicker LS (2006). In vivo RNA interference demonstrates a role for Nramp1 in modifying susceptibility to type 1 diabetes. *Nat Genet 38*, 479–483.
78. Krutzfeldt J, Rajewsky N, Braich R, Rajeev KG, Tuschl T, Manoharan M, Stoffel M (2005). Silencing of microRNAs in vivo with "antagomirs". *Nature 438*, 685–689.
79. Kumar P, Wu H, McBride JL, Jung KE, Kim MH, Davidson BL, Lee SK, Shankar P, Manjunath N (2007). Transvascular delivery of small interfering RNA to the central nervous system. *Nature 448*, 39–43.
80. Lagos-Quintana M, Rauhut R, Yalcin A, Meyer J, Lendeckel W, Tuschl T (2002). Identification of tissue-specific microRNAs from mouse. *Curr Biol 12*, 735–739.
81. Layzer JM, McCaffrey AP, Tanner AK, Huang Z, Kay MA, Sullenger BA (2004). In vivo activity of nuclease-resistant siRNAs. *RNA 10*, 766–771.
82. Le Bon A, Schiavoni G, D'Agostino G, Gresser I, Belardelli F, Tough DF (2001). Type i interferons potently enhance humoral immunity and can promote isotype switching by stimulating dendritic cells in vivo. *Immunity 14*, 461–470.
83. Lee Y, Ahn C, Han J, Choi H, Kim J, Yim J, Lee J, Provost P, Radmark O, Kim S, Kim VN (2003). The nuclear RNase III Drosha initiates microRNA processing. *Nature 425*, 415–419.
84. Lee Y, Kim M, Han J, Yeom KH, Lee S, Baek SH, Kim VN (2004). MicroRNA genes are transcribed by RNA polymerase II. *Embo J 23*, 4051–4060.
85. Lee YS, Dutta A (2007). The tumor suppressor microRNA let-7 represses the HMGA2 oncogene. *Genes Dev 21*, 1025–1030.
86. Lewis BP, Burge CB, Bartel DP (2005). Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell 120*, 15–20.
87. Lewis BP, Shih IH, Jones-Rhoades MW, Bartel DP, Burge CB (2003). Prediction of mammalian microRNA targets. *Cell 115*, 787–798.
88. Lewis DL, Hagstrom JE, Loomis AG, Wolff JA, Herweijer H (2002). Efficient delivery of siRNA for inhibition of gene expression in postnatal mice. *Nat Genet 32*, 107–108.
89. Li G, Xiang Y, Sabapathy K, Silverman RH (2004). An apoptotic signaling pathway in the interferon antiviral response mediated by RNase L and c-Jun NH2-terminal kinase. *J Biol Chem 279*, 1123–1131.
90. Li L, Lin X, Khvorova A, Fesik SW, Shen Y (2007). Defining the optimal parameters for hairpin-based knockdown constructs. *RNA 13*, 1765–1774.
91. Li MJ, Kim J, Li S, Zaia J, Yee JK, Anderson J, Akkina R, Rossi JJ (2005). Long-term inhibition of HIV-1 infection in primary hematopoietic cells by lentiviral vector delivery of a triple combination of anti-HIV shRNA, anti-CCR5 ribozyme, and a nucleolar-localizing TAR decoy. *Mol Ther 12*, 900–909.
92. Lin X, Ruan X, Anderson MG, McDowell JA, Kroeger PE, Fesik SW, Shen Y (2005). siRNA-mediated off-target gene silencing triggered by a 7 nt complementation. *Nucleic Acids Res 33*, 4527–4535.
93. Lin X, Yang J, Chen J, Gunasekera A, Fesik SW, Shen Y (2004). Development of a tightly regulated U6 promoter for shRNA expression. *FEBS Lett 577*, 376–380.
94. Liu J, Valencia-Sanchez MA, Hannon GJ, Parker R (2005). MicroRNA-dependent localization of targeted mRNAs to mammalian P-bodies. *Nat Cell Biol 7*, 719–723.
95. Lu S, Cullen BR (2004). Adenovirus VA1 noncoding RNA can inhibit small interfering RNA and MicroRNA biogenesis. *J Virol 78*, 12868–12876.
96. Lund E, Guttinger S, Calado A, Dahlberg JE, Kutay U (2004). Nuclear export of microRNA precursors. *Science 303*, 95–98.
97. Ma Z, Li J, He F, Wilson A, Pitt B, Li S (2005). Cationic lipids enhance siRNA-mediated interferon response in mice. *Biochem Biophys Res Commun 330*, 755–759.
98. Makinen PI, Koponen JK, Karkkainen AM, Malm TM, Pulkkinen KH, Koistinaho J, Turunen MP, Yla-Herttuala S (2006). Stable RNA interference: Comparison of U6 and H1 promoters in endothelial cells and in mouse brain. *J Gene Med 8*, 433–441.
99. Matranga C, Tomari Y, Shin C, Bartel DP, Zamore PD (2005). Passenger-strand cleavage facilitates assembly of siRNA into Ago2-containing RNAi enzyme complexes. *Cell 123*, 607–620.

100. Matsuda T, Cepko CL (2004). Electroporation and RNA interference in the rodent retina in vivo and in vitro. *Proc Natl Acad Sci USA 101*, 16–22.
101. Matsukura S, Jones PA, Takai D (2003). Establishment of conditional vectors for hairpin siRNA knockdowns. *Nucleic Acids Res 31*, e77.
102. Matthess Y, Kappel S, Spankuch B, Zimmer B, Kaufmann M, Strebhardt K (2005). Conditional inhibition of cancer cell proliferation by tetracycline-responsive, H1 promoter-driven silencing of PLK1. *Oncogene 24*, 2973–2780.
103. McCaffrey AP, Meuse L, Pham TT, Conklin DS, Hannon GJ, Kay MA (2002). RNA interference in adult mice. *Nature 418*, 38–39.
104. McManus MT, Petersen CP, Haines BB, Chen J, Sharp PA (2002). Gene silencing using micro-RNA designed hairpins. *RNA 8*, 842–850.
105. McNamara JO 2nd, Andrechek ER, Wang Y, Viles KD, Rempel RE, Gilboa E, Sullenger BA, Giangrande PH (2006). Cell type-specific delivery of siRNAs with aptamer-siRNA chimeras. *Nat Biotechnol 24*, 1005–1015.
106. Melchjorsen J, Jensen SB, Malmgaard L, Rasmussen SB, Weber F, Bowie AG, Matikainen S, Paludan SR (2005). Activation of innate defense against a paramyxovirus is mediated by RIG-I and TLR7 and TLR8 in a cell-type-specific manner. *J Virol 79*, 12944–12951.
107. Miller TM, Kaspar BK, Kops GJ, Yamanaka K, Christian LJ, Gage FH, Cleveland DW (2005). Virus-delivered small RNA silencing sustains strength in amyotrophic lateral sclerosis. *Ann Neurol 57*, 773–776.
108. Morral N, O'Neal W, Zhou H, Langston C, Beaudet A (1997). Immune responses to reporter proteins and high viral dose limit duration of expression with adenoviral vectors: Comparison of E2a wild type and E2a deleted vectors. *Hum Gene Ther 8*, 1275–1286.
109. Morral N, Parks R, Zhou H, Langston C, Schiedner G, Quinones J, Graham FL, Kochanek S, Beaudet AL (1998). High doses of a helper-dependent adenoviral vector yield supraphysiological levels of $a_1$-antitrypsin with negligible toxicity. *Hum Gene Ther 9*, 2709–2716.
110. Morrissey DV, Blanchard K, Shaw L, Jensen K, Lockridge JA, Dickinson B, McSwiggen JA, Vargeese C, Bowman K, Shaffer CS, Polisky BA, Zinnen S (2005). Activity of stabilized short interfering RNA in a mouse model of hepatitis B virus replication. *Hepatology 41*, 1349–1356.
111. Morsy MA, Gu MC, Motzel S, Zhao J, Su Q, Allen H, Franlin L, Parks RJ, Graham FL, Kochanek S, Bett AJ, Caskey CT (1998). An adenoviral vector deleted for all viral coding sequences results in enhanced safety and extended expression of a leptin transgene. *Proc Natl Acad Sci USA 95*, 7866–7871.
112. Muratovska A, Eccles MR (2004). Conjugate for efficient delivery of short interfering RNA (siRNA) into mammalian cells. *FEBS Lett 558*, 63–68.
113. Naldini L, Blomer U, Gallay P, Ory D, Mulligan R, Gage FH, Verma IM, Trono D (1996). In vivo gene delivery and stable transduction of nondividing cells by a lentiviral vector. *Science 272*, 263–267.
114. Narvaiza I, Aparicio O, Vera M, Razquin N, Bortolanza S, Prieto J, Fortes P (2006). Effect of adenovirus-mediated RNA interference on endogenous microRNAs in a mouse model of multidrug resistance protein 2 gene silencing. *J Virol 80*, 12236–12247.
115. Nguyen LH, Espert L, Mechti N, Wilson DM 3rd (2001). The human interferon- and estrogen-regulated ISG20/HEM45 gene product degrades single-stranded RNA and DNA in vitro. *Biochemistry 40*, 7174–7179.
116. O'Neal WK, Zhou H, Morral N, Aguilar-Cordova E, Pestaner J, Langston C, Mull B, Wang Y, Beaudet AL, Lee B (1998). Toxicological comparison of E2a-deleted and first-generation adenoviral vectors expressing alpha1-antitrypsin after systemic delivery. *Hum Gene Ther 9*, 1587–1598.
117. Ohmori T, Kashiwakura Y, Ishiwata A, Madoiwa S, Mimuro J, Sakata Y (2007). Silencing of a targeted protein in in vivo platelets using a lentiviral vector delivering short hairpin RNA sequence. *Arterioscler Thromb Vasc Biol 27*, 2266–2272.
118. Oka K, Pastore L, Kim IH, Merched A, Nomura S, Lee HJ, Merched-Sauvage M, Arden-Riley C, Lee B, Finegold M, Beaudet A, Chan L (2001). Long-term stable correction of low-density lipoprotein receptor-deficient mice with a helper-dependent adenoviral vector expressing the very low-density lipoprotein receptor. *Circulation 103*, 1274–1281.
119. Omidi Y, Hollins AJ, Benboubetra M, Drayton R, Benter IF, Akhtar S (2003). Toxicogenomics of non-viral vectors for gene therapy: A microarray study of lipofectin- and oligofectamine-induced gene expression changes in human epithelial cells. *J Drug Target 11*, 311–323.
120. Omidi Y, Hollins AJ, Drayton RM, Akhtar S (2005). Polypropylenimine dendrimer-induced gene expression changes: The effect of complexation with DNA, dendrimer generation and cell type. *J Drug Target 13*, 431–443.
121. Paddison PJ, Caudy AA, Bernstein E, Hannon GJ, Conklin DS (2002). Short hairpin RNAs (shRNAs) induce sequence-specific silencing in mammalian cells. *Genes Dev 16*, 948–958.
122. Pal-Bhadra M, Bhadra U, Birchler JA (1997). Cosuppression in *Drosophila*: Gene silencing of alcohol dehydrogenase by white-Adh transgenes is Polycomb dependent. *Cell 90*, 479–490.
123. Palmer DJ, Ng P (2005). Helper-dependent adenoviral vectors for gene therapy. *Hum Gene Ther 16*, 1–16.
124. Park F (2007). Lentiviral vectors: Are they the future of animal transgenesis? *Physiol Genomics 31*, 159–173.
125. Paskowitz DM, Greenberg KP, Yasumura D, Grimm D, Yang H, Duncan JL, Kay MA, Lavail MM, Flannery JG, Vollrath D (2007). Rapid and stable knockdown of an endogenous gene in retinal pigment epithelium. *Hum Gene Ther 18*, 871–880.

126. Patzel V, Rutz S, Dietrich I, Koberle C, Scheffold A, Kaufmann SH (2005). Design of siRNAs producing unstructured guide-RNAs results in improved RNA interference efficiency. *Nat Biotechnol 23*, 1440–1444.

127. Pebernard S, Iggo RD (2004). Determinants of interferon-stimulated gene induction by RNAi vectors. *Differentiation 72*, 103–111.

128. Pei Y, Tuschl T (2006). On the art of identifying effective and specific siRNAs. *Nat Methods 3*, 670–676.

129. Persengiev SP, Zhu X, Green MR (2004). Nonspecific, concentration-dependent stimulation and repression of mammalian gene expression by small interfering RNAs (siRNAs). *RNA 10*, 12–18.

130. Pillai RS, Bhattacharyya SN, Filipowicz W (2007). Repression of protein synthesis by miRNAs: How many mechanisms? *Trends Cell Biol 17*, 118–126.

131. Pille JY, Li H, Blot E, Bertrand JR, Pritchard LL, Opolon P, Maksimenko A, Lu H, Vannier JP, Soria J, Malvy C, Soria C (2006). Intravenous delivery of anti-RhoA small interfering RNA loaded in nanoparticles of chitosan in mice: Safety and efficacy in xenografted aggressive breast cancer. *Hum Gene Ther 17*, 1019–1026.

132. Pluta K, Diehl W, Zhang XY, Kutner R, Bialkowska A, Reiser J (2007). Lentiviral vectors encoding tetracycline-dependent repressors and transactivators for reversible knockdown of gene expression: A comparative study. *BMC Biotechnol 7*, 41.

133. Ralph GS, Radcliffe PA, Day DM, Carthy JM, Leroux MA, Lee DC, Wong LF, Bilsland LG, Greensmith L, Kingsman SM, Mitrophanous KA, Mazarakis ND, Azzouz M (2005). Silencing mutant SOD1 using RNAi protects against neurodegeneration and extends survival in an ALS model. *Nat Med 11*, 429–433.

134. Rand TA, Petersen S, Du F, Wang X (2005). Argonaute2 cleaves the anti-guide strand of siRNA during RISC activation. *Cell 123*, 621–629.

135. Rangasamy D, Tremethick DJ, Greaves IK (2008). Gene knockdown by ecdysone-based inducible RNAi in stable mammalian cell lines. *Nat Protoc 3*, 79–88.

136. Raoul C, Abbas-Terki T, Bensadoun JC, Guillot S, Haase G, Szulc J, Henderson CE, Aebischer P (2005). Lentiviral-mediated silencing of SOD1 through RNA interference retards disease onset and progression in a mouse model of ALS. *Nat Med 11*, 423–428.

137. Reddy PS, Sakhuja K, Ganesh S, Yang L, Kayda D, Brann T, Pattison S, Golightly D, Idamakanti N, Pinkstaff A, Kaloss M, Barjot C, Chamberlain JS, Kaleko M, Connelly S (2002). Sustained human factor VIII expression in hemophilia A mice following systemic delivery of a gutless adenoviral vector. *Mol Ther 5*, 63–73.

138. Reinhart BJ, Slack FJ, Basson M, Pasquinelli AE, Bettinger JC, Rougvie AE, Horvitz HR, Ruvkun G (2000). The 21-nucleotide let-7 RNA regulates developmental timing in *Caenorhabditis elegans*. *Nature 403*, 901–906.

139. Reynolds A, Anderson EM, Vermeulen A, Fedorov Y, Robinson K, Leake D, Karpilow J, Marshall WS, Khvorova A (2006). Induction of the interferon response by siRNA is cell type- and duplex length-dependent. *RNA 12*, 988–993.

140. Romano N, Macino G (1992). Quelling: Transient inactivation of gene expression in *Neurospora crassa* by transformation with homologous sequences. *Mol Microbiol 6*, 3343–3353.

141. Rossi JJ, June CH, Kohn DB (2007). Genetic therapies against HIV. *Nat Biotechnol 25*, 1444–1454.

142. Rubinson DA, Dillon CP, Kwiatkowski AV, Sievers C, Yang L, Kopinja J, Rooney DL, Zhang M, Ihrig MM, McManus MT, Gertler FB, Scott ML, Van Parijs L (2003). A lentivirus-based system to functionally silence genes in primary mammalian cells, stem cells and transgenic mice by RNA interference. *Nat Genet 33*, 401–406.

143. Sakamoto N, Tanabe Y, Yokota T, Satoh K, Sekine-Osajima Y, Nakagawa M, Itsui Y, Tasaka M, Sakurai Y, Cheng-Hsin C, Yano M, Ohkoshi S, Aoyagi Y, Maekawa S, Enomoto N, Kohara M, Watanabe M (2008). Inhibition of hepatitis C virus infection and expression in vitro and in vivo by recombinant adenovirus expressing short hairpin RNA. *J Gastroenterol Hepatol 23*, 1437–1447.

144. Saxena S, Jonsson ZO, Dutta A (2003). Small RNAs with imperfect match to endogenous mRNA repress translation. Implications for off-target activity of small inhibitory RNA in mammalian cells. *J Biol Chem 278*, 44312–44319.

145. Schiedner G, Morral N, Parks R, Wu Y, Koopmans SC, Langston C, Graham FL, Beaudet AL, Kochanek S (1998). Genomic DNA transfer with a high-capacity adenovirus vector results in improved *in vivo* gene expression and decreased toxicity. *Nat Genet 18*, 180–183.

146. Schiffelers RM, Ansari A, Xu J, Zhou Q, Tang Q, Storm G, Molema G, Lu PY, Scaria PV, Woodle MC (2004). Cancer siRNA therapy by tumor selective delivery with ligand-targeted sterically stabilized nanoparticle. *Nucleic Acids Res 32*, e149.

147. Schlee M, Hornung V, Hartmann G (2006). siRNA and isRNA: Two edges of one sword. *Mol Ther 14*, 463–470.

148. Schratt GM, Tuebing F, Nigh EA, Kane CG, Sabatini ME, Kiebler M, Greenberg ME (2006). A brain-specific microRNA regulates dendritic spine development. *Nature 439*, 283–289.

149. Schwarz DS, Hutvagner G, Du T, Xu Z, Aronin N, Zamore PD (2003). Asymmetry in the assembly of the RNAi enzyme complex. *Cell 115*, 199–208.

150. Seibler J, Kleinridders A, Kuter-Luks B, Niehaves S, Bruning JC, Schwenk F (2007). Reversible gene knockdown in mice using a tight, inducible shRNA expression system. *Nucleic Acids Res 35*, e54.

151. Semizarov D, Frost L, Sarthy A, Kroeger P, Halbert DN, Fesik SW (2003). Specificity of short interfering RNA

determined through gene expression signatures. *Proc Natl Acad Sci USA 100*, 6347–6352.

152. Sen GC, Sarkar SN (2005). Transcriptional signaling by double-stranded RNA: Role of TLR3. *Cytokine Growth Factor Rev 16*, 1–14.
153. Shen J, Samul R, Silva RL, Akiyama H, Liu H, Saishin Y, Hackett SF, Zinnen S, Kossen K, Fosnaugh K, Vargeese C, Gomez A, Bouhana K, Aitchison R, Pavco P, Campochiaro PA (2006). Suppression of ocular neovascularization with siRNA targeting VEGF receptor 1. *Gene Ther 13*, 225–234.
154. Shukla V, Coumoul X, Deng CX (2007). RNAi-based conditional gene knockdown in mice using a U6 promoter driven vector. *Int J Biol Sci 3*, 91–99.
155. Silva JM, Li MZ, Chang K, Ge W, Golding MC, Rickles RJ, Siolas D, Hu G, Paddison PJ, Schlabach MR, Sheth N, Bradshaw J, Burchard J, Kulkarni A, Cavet G, Sachidanandam R, McCombie WR, Cleary MA, Elledge SJ, Hannon GJ (2005). Second-generation shRNA libraries covering the mouse and human genomes. *Nat Genet 37*, 1281–1288.
156. Sledz CA, Holko M, de Veer MJ, Silverman RH, Williams BR (2003). Activation of the interferon system by short-interfering RNAs. *Nat Cell Biol 5*, 834–839.
157. Sledz CA, Williams BR (2004). RNA interference and double-stranded-RNA-activated pathways. *Biochem Soc Trans 32*, 952–956.
158. Sledz CA, Williams BR (2005). RNA interference in biology and disease. *Blood 106*, 787–794.
159. Song E, Zhu P, Lee SK, Chowdhury D, Kussman S, Dykxhoorn DM, Feng Y, Palliser D, Weiner DB, Shankar P, Marasco WA, Lieberman J (2005). Antibody mediated in vivo delivery of small interfering RNAs via cell-surface receptors. *Nat Biotechnol 23*, 709–717.
160. Sood P, Krek A, Zavolan M, Macino G, Rajewsky N (2006). Cell-type-specific signatures of microRNAs on target mRNA expression. *Proc Natl Acad Sci USA 103*, 2746–2751.
161. Soutschek J, Akinc A, Bramlage B, Charisse K, Constien R, Donoghue M, Elbashir S, Geick A, Hadwiger P, Harborth J, John M, Kesavan V, Lavine G, Pandey RK, Racie T, Rajeev KG, Rohl I, Toudjarska I, Wang G, Wuschko S, Bumcrot D, Koteliansky V, Limmer S, Manoharan M, Vornlocher HP (2004). Therapeutic silencing of an endogenous gene by systemic administration of modified siRNAs. *Nature 432*, 173–178.
162. Stark GR, Kerr IM, Williams BR, Silverman RH, Schreiber RD (1998). How cells respond to interferons. *Annu Rev Biochem 67*, 227–264.
163. Stegmeier F, Hu G, Rickles RJ, Hannon GJ, Elledge SJ (2005). A lentiviral microRNA-based system for single-copy polymerase II-regulated RNA interference in mammalian cells. *Proc Natl Acad Sci USA 102*, 13212–13217.
164. Sui G, Soohoo C, Affar el B, Gay F, Shi Y, Forrester WC (2002). A DNA vector-based RNAi technology to suppress gene expression in mammalian cells. *Proc Natl Acad Sci USA 99*, 5515–5520.
165. Szulc J, Wiznerowicz M, Sauvain MO, Trono D, Aebischer P (2006). A versatile tool for conditional gene expression and knockdown. *Nat Methods 3*, 109–116.
166. Takei Y, Kadomatsu K, Yuzawa Y, Matsuo S, Muramatsu T (2004). A small interfering RNA targeting vascular endothelial growth factor as cancer therapeutics. *Cancer Res 64*, 3365–3370.
167. Tang G, Reinhart BJ, Bartel DP, Zamore PD (2003). A biochemical framework for RNA silencing in plants. *Genes Dev 17*, 49–63.
168. Taniguchi CM, Ueki K, Kahn R (2005). Complementary roles of IRS-1 and IRS-2 in the hepatic regulation of metabolism. *J Clin Invest 115*, 718–727.
169. Taxman DJ, Livingstone LR, Zhang J, Conti BJ, Iocca HA, Williams KL, Lich JD, Ting JP, Reed W (2006). Criteria for effective design, construction, and gene knockdown by shRNA vectors. *BMC Biotechnol 6*, 7.
170. Thomas CE, Schiedner G, Kochanek S, Castro MG, Lowenstein PR (2000). Peripheral infection with adenovirus causes unexpected long-term brain inflammation in animals injected intracranially with first-generation, but not with high-capacity, adenovirus vectors: Toward realistic long-term neurological gene therapy for chronic diseases. *Proc Natl Acad Sci USA 97*, 7482–7487.
171. Tiscornia G, Tergaonkar V, Galimi F, Verma IM (2004). CRE recombinase-inducible RNA interference mediated by lentiviral vectors. *Proc Natl Acad Sci USA 101*, 7347–7351.
172. Toietta G, Koehler DR, Finegold MJ, Lee B, Hu J, Beaudet AL (2003). Reduced inflammation and improved airway expression using helper-dependent adenoviral vectors with a K18 promoter. *Mol Ther 7*, 649–658.
173. van de Water FM, Boerman OC, Wouterse AC, Peters JG, Russel FG, Masereeuw R (2006). Intravenously administered short interfering RNA accumulates in the kidney and selectively suppresses gene function in renal proximal tubules. *Drug Metab Dispos 34*, 1393–1397.
174. van de Wetering M, Oving I, Muncan V, Pon Fong MT, Brantjes H, van Leenen D, Holstege FC, Brummelkamp TR, Agami R, Clevers H (2003). Specific inhibition of gene expression using a stably integrated, inducible small-interfering-RNA vector. *EMBO Rep 4*, 609–615.
175. Ventura A, Meissner A, Dillon CP, McManus M, Sharp PA, Van Parijs L, Jaenisch R, Jacks T (2004). Cre-lox-regulated conditional RNA interference from transgenes. *Proc Natl Acad Sci USA 101*, 10380–10385.
176. Wang X, Skelley L, Cade R, Sun Z (2006). AAV delivery of mineralocorticoid receptor shRNA prevents progression of cold-induced hypertension and attenuates renal damage. *Gene Ther 13*, 1097–1103.
177. Witting SR, Brown M, Saxena R, Nabinger S, Morral N (2008). Helper-dependent adenovirus-mediated short hairpin RNA expression in the liver activates the interferon response. *J Biol Chem 283*, 2120–2128.

178. Wiznerowicz M, Szulc J, Trono D (2006). Tuning silence: Conditional systems for RNA interference. *Nat Methods 3*, 682–688.

179. Wolfrum C, Shi S, Jayaprakash KN, Jayaraman M, Wang G, Pandey RK, Rajeev KG, Nakayama T, Charrise K, Ndungo EM, Zimmermann T, Koteliansky V, Manoharan M, Stoffel M (2007). Mechanisms and optimization of in vivo delivery of lipophilic siRNAs. *Nat Biotechnol 25*, 1149–1157.

180. Wu Z, Asokan A, Samulski RJ (2006). Adeno-associated virus serotypes: Vector toolkit for human gene therapy. *Mol Ther 14*, 316–327.

181. Xia H, Mao Q, Paulson HL, Davidson BL (2002). siRNA-mediated gene silencing in vitro and in vivo. *Nat Biotechnol 20*, 1006–1010.

182. Xie SY, Ren ZR, Zhang JZ, Guo XB, Wang QX, Wang S, Lin D, Gong XL, Li W, Huang SZ, Zeng F, Zeng YT (2007). Restoration of the balanced alpha/beta-globin gene expression in beta654-thalassemia mice using combined RNAi and antisense RNA approach. *Hum Mol Genet 16*, 2616–2625.

183. Xiong W, Goverdhana S, Sciascia SA, Candolfi M, Zirger JM, Barcia C, Curtin JF, King GD, Jaita G, Liu C, Kroeger K, Agadjanian H, Medina-Kauwe L, Palmer D, Ng P, Lowenstein PR, Castro MG (2006). Regulatable gutless adenovirus vectors sustain inducible transgene expression in the brain in the presence of an immune response against adenoviruses. *J Virol 80*, 27–37.

184. Xu H, Wilcox D, Nguyen P, Voorbach M, Suhar T, Morgan SJ, An WF, Ge L, Green J, Wu Z, Gimeno RE, Reilly R, Jacobson PB, Collins CA, Landschulz K, Surowy T (2006). Hepatic knockdown of mitochondrial GPAT1 in ob/ob mice improves metabolic profile. *Biochem Biophys Res Commun 349*, 439–448.

185. Xu P, Vernooy SY, Guo M, Hay BA (2003). The *Drosophila* microRNA Mir-14 suppresses cell death and is required for normal fat metabolism. *Curr Biol 13*, 790–795.

186. Yi R, Doehle BP, Qin Y, Macara IG, Cullen BR (2005). Overexpression of exportin 5 enhances RNA interference mediated by short hairpin RNAs and microRNAs. *RNA 11*, 220–226.

187. Yi R, Qin Y, Macara IG, Cullen BR (2003). Exportin-5 mediates the nuclear export of pre-microRNAs and short hairpin RNAs. *Genes Dev 17*, 3011–3016.

188. Yu F, Yao H, Zhu P, Zhang X, Pan Q, Gong C, Huang Y, Hu X, Su F, Lieberman J, Song E (2007). let-7 regulates self renewal and tumorigenicity of breast cancer cells. *Cell 131*, 1109–1123.

189. Zeng Y, Cai X, Cullen BR (2005). Use of RNA polymerase II to transcribe artificial microRNAs. *Methods Enzymol 392*, 371–380.

190. Zeng Y, Wagner EJ, Cullen BR (2002). Both natural and designed micro RNAs can inhibit the expression of cognate mRNAs when expressed in human cells. *Mol Cell 9*, 1327–1333.

191. Zeng Y, Yi R, Cullen BR (2005). Recognition and cleavage of primary microRNA precursors by the nuclear processing enzyme Drosha. *EMBO J 24*, 138–148.

192. Zhang HY, Du Q, Wahlestedt C, Liang Z (2006). RNA Interference with chemically modified siRNA. *Curr Top Med Chem 6*, 893–900.

193. Zimmermann TS, Lee AC, Akinc A, Bramlage B, Bumcrot D, Fedoruk MN, Harborth J, Heyes JA, Jeffs LB, John M, Judge AD, Lam K, McClintock K, Nechev LV, Palmer LR, Racie T, Rohl I, Seiffert S, Shanmugam S, Sood V, Soutschek J, Toudjarska I, Wheat AJ, Yaworski E, Zedalis W, Koteliansky V, Manoharan M, Vornlocher HP, MacLachlan I (2006). RNAi-mediated gene silencing in non-human primates. *Nature 441*, 111–114.

194. Zou L, Yuan X, Zhou H, Lu H, Yang K (2001). Helper-dependent adenoviral vector-mediated gene transfer in aged rat brain. *Hum Gene Ther 12*, 181–191.

# CRYSTALLOGRAPHY

Carl H. Schwalbe
*School of Life & Health Sciences, Aston University, Birmingham, United Kingdom*

## 1 INTRODUCTION

In the hierarchy of methods for structure determination, crystallography is the monarch. A technically correct crystal structure determination on high-quality material approaches as close to infallibility as any human effort can do. This degree of certainty comes at a cost: It requires a properly crystalline sample and uses sophisticated expensive apparatus. This chapter first aims to give a general overview of what crystallography can contribute to drug discovery. A discussion of practical matters follows, covering required features of the crystalline sample, the radiation source, and the detector. Next, the essentials of the underlying theory are introduced so that the rationale for the steps in a crystal structure determination can be understood. Although crystallography is underpinned by mathematics, the treatment here will be as intuitive as possible with limited recourse to mathematics. The narrative then returns to specific examples of the use (and nonuse) of crystallography in drug discovery.

### 1.1 Discovering an Unknown Structure

For many years crystallography has provided the definitive proof of structure for a novel small molecule. Only one sample crystal is usually needed, and this specimen crystal remains available for further studies. Such economy of material is of particular value if, for instance, a very small quantity of a natural product has been isolated. If just the connectivity of the molecule is required, spectroscopy [especially nuclear magnetic resonance (NMR)] is generally sufficient, but crystallography removes any doubt and additionally provides geometric data like bond distances and angles. Crystallographic location of hydrogen atoms is less certain since X rays respond weakly to hydrogen atoms. Nevertheless, hydrogen atom positions can frequently be inferred from the molecular geometry and subsequently confirmed in a map. For instance, a carboxyl group in the free acid form has a short C=O bond and a longer C–OH bond, while a carboxylate ion has two similar C–O distances. If precise details of hydrogen atom geometry are necessary to understand the activity of a particular drug, neutron diffraction provides this information.

Crystallography is of key importance for macromolecules. Revolutionary technical advances have improved the productivity of macromolecular crystallography from a few structures per year in the 1960s to many thousands now. This is fortunate since the Human Genome Project has provided a wealth of new macromolecules whose structure and function need to be investigated. Once again, NMR provides a complementary technique. However, it has the great disadvantage of requiring inference, which could be wrong, although it does have the advantage of sampling all populated conformations in solution. Crystal growing necessarily freezes out most of the motion and selects one or a few conformations that pack together well.

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

### 1.2 Structure-Based Drug Design

In the ideal case the detailed three-dimensional structure is known for a target macromolecule whose function needs to be inhibited or enhanced by therapeutic intervention. Likewise the structures are known for some lead compounds, chosen by molecular modeling or high-throughput screening. Additional crystal structures are determined for the drug–target complexes, and the important binding contacts are identified, along with any regions of poor steric fit. Using such information in an iterative process, successive generations of drugs are made to have stronger binding and greater selectivity.

If the structure or even the identity of the target macromolecule is not known, crystallography can still be useful. From the structures of a series of active drugs and closely related inactive molecules, a *pharmacophore model* can be developed, comprising chemical groups (e.g., hydrogen bond donors, hydrogen bond acceptors, or aromatic rings) and the spatial relationships between them. Such a model is especially useful to screen out hopeless candidate molecules before synthetic effort is wasted on them.

### 1.3 Relating Crystal Structure to Properties

There are obvious correlations between molecular geometry and chemical reactivity; for instance, a longer bond between a defined pair of elements is a weaker bond, likely to be more reactive. Crystal structures also help to address formulation issues for solid dosage forms. Two conflicting objectives have to be met. A long shelf life correlates with a stable solid; but good bioavailability requires adequate aqueous solubility, which is more easily achieved with an unstable solid. The same drug may exist in multiple crystal forms (*polymorphs*) having identical chemical composition but different structure, leading to differences in stability, melting point, solubility, and handling properties. Aqueous solubility in a series of salts of a drug does not always correlate with hydrophilicity of the counterion, showing the importance of structure. The pharmaceutical industry has invested heavily in robotic technology to screen new drugs for polymorphs and find the most suitable counterion in a salt. If neither the neutral drug nor a simple salt has the desired properties, the investigation can be extended to co-crystals with other drugs.

### 1.4 Crystallography without a Crystal: Data Mining

Pharmaceutical use of crystallography has been greatly enhanced by the compilation of two databases: the Cambridge Structural Database (CSD) with hundreds of thousands of small molecules (www.ccdc.cam.ac.uk) and the Protein Data Bank (PDB) with tens of thousands of macromolecules (www.rcsb.org/pdb/). One crystal structure of a flexible drug molecule may show a conformation that is dictated by crystal packing and unrepresentative of the preferred conformation in solution or at a receptor. However, if that same conformation occurs in crystal structures of related molecules in the CSD, it is likely to be stable. If the structure of a target macromolecule is available in the PDB, it provides a sound basis for drug design. Frequently, the structure is only available for a related protein or the same one from a different organism. If so, *homology modeling* can be attempted. The backbone is retained, the amino acid side chains are adjusted to match the sequence of the new target, and strains are relieved by optimization with molecular mechanics calculations.

### 1.5 Crystal Structure Prediction

If crystal structures could be reliably predicted, the effort of determining them experimentally could be avoided. It is fairly straightforward to predict the crystal structure of a simple small molecule. Unfortunately, the correct structure is usually accompanied by several dozen incorrectly predicted structures that seem just as good according to the common-sense criteria of low energy and high packing density. A major research effort is underway to develop prediction methods that discriminate better between correct and incorrect structures. Progress has been assessed in a series of blind tests. With the authors' consent publication of a few novel crystal structures has been held back for a few months. Diagrams of the molecular connectivity have been sent to participating research groups, who have been asked to apply their prediction software during this period and to return their best three predicted crystal structures for each molecule. The results of the third blind test (2004) showed that good progress has been made in predicting crystal structures built from a single copy of a small rigid molecule, but flexible molecules still cause insuperable difficulty [1].

## 2 PRACTICAL MATTERS

### 2.1 The Sample Crystal

The definition of a crystal is one source of misunderstanding between crystallographers and organic chemists. Organic chemists tend to judge the macroscopic external appearance, but to be suitable for crystallographic use the sample crystal has to possess an internal structure that is a regular repeating array of atoms, mol-

**TABLE 1 Summary of Techniques Used for Crystal Growing**

| Method | Requirements | Advantages | Disadvantages | Uses |
|---|---|---|---|---|
| Slow evaporation | Moderately soluble sample | Easy, gentle | Long exposure to air; drying out leaves a crust | Small molecules |
| Slow cooling | Large variation of solubility with temperature | Easy | Cooling must be controlled; initial heating risks decomposition | Small molecules |
| Vapor diffusion | Sample dissolves in one solvent; vapor of poorer solvent can diffuse | System can be protected by sealing and cooling | Solvent pair must have correct solvent power and volatility | Macromolecules, small molecules |
| Liquid diffusion | Sample in good solvent; less dense poor solvent layered on top | System can be protected by sealing and cooling | Solvent pair must have correct solvent power and density | Small molecules, macromolecules |
| Sublimation | Volatile sample evaporates, then crystallizes on cooler surface | No unpredictable interactions with solvents | Risk of decomposition | Small molecules |

ecules, or ions. It must have smooth edges and no reentrant angles, that is, "bays" flanked by "headlands," which are indicative of more than one crystal stuck together. A polarizing microscope is a cheap and extremely useful part of any crystallography laboratory. All but the most highly symmetrical crystals rotate the plane of polarized light. If such a crystal is viewed between crossed polarizers on a slowly rotating microscope stage, both the crystal and its surroundings will appear black until the appropriate angle is reached where the crystal rotates the light so that it can pass the second polarizer. At that point the crystal will flash into brightness and soon after subside back into darkness. If the changes are sharp and uniform, the crystal is expected to be single and of high quality. Uniform but more diffuse changes suggest lower quality, while illumination of only part of the crystal shows that it is not single.

***Crystal Growing Techniques*** Crystal structure determination relies on the ability of a crystal, irradiated with a parallel beam of X rays or neutrons, to produce diffracted beams that can be measured. Specimen crystals must be of high quality and about 0.4 mm across to match the typical X-ray beam size, or >1 mm across if used for a neutron diffraction experiment. The most appropriate method for growing them depends greatly on the properties of the particular material, but some general principles apply. The initial sample should be reasonably pure. While recrystallization is a standard technique for purifying chemicals since the growing crystals exclude impurities, such impurities are likely to impede growth on certain faces leading to small size and distorted crystal morphology. Glassware should be well but not obsessively cleaned. Dust particles often provide nucleation sites for crystal growth: Too many and the crystals will be too small; none at all and crystals may not form. Patience is a great virtue: Provided the material is stable, slow growth will yield the largest and most perfect crystals.

The most used methods for growing "X-ray quality" crystals are enumerated in Table 1.

Public-spirited individuals have placed helpful guidance notes on the Web. Two that can be recommended are "Growing Crystals That Will Make Your Crystallographer Happy" by Paul D. Boyle [2] and paired contributions both entitled "Crystal Growing Tips" by Maarten Dinger and Jerzy Klosin [3]. In turn these notes cite useful books. Techniques for improving protein crystals used in diffraction experiments have been comprehensively surveyed [4].

### *Crystal Pathology*

*Twinning* A common problem with crystal growing is that crystals may grow together or stick together. If a large high-quality specimen crystal has a small randomly oriented "flea" stuck to it, it may be wiser to tolerate the situation than to risk cutting the crystals apart. The small flea will give weak diffracted X-ray beams, and most of them will not coincide with diffracted beams from the main crystal. Major difficulties arise if two or possibly more adhering crystals are nearly equal in size and their orientations are related by rotation or reflection. Such crystals are *twinned*. Twinning usually arises when crystal axes are exactly or approximately at right angles to each other. If the axes are exactly perpendicular, 180° rotation about one axis will leave the other two coincident but pointing in opposite directions. In this case, *merohedral twinning*, measuring

the diffracted beams is not difficult since the beams from the two crystals coincide, but interpreting them is much more difficult. If the axes are nearly but not quite perpendicular, the diffracted beams will no longer coincide but may overlap. This case, *nonmerohedral twinning*, causes problems in both measurement and interpretation. Modern hardware and software can cope with twinned specimen crystals, but it is far better to avoid them if possible.

*Thermal Motion and Disorder* Even though the *average* positions of molecules in a perfect crystal are fixed in an orderly array, the molecules and their component parts are vibrating. The atoms are properly represented not as points but as the ellipsoids they trace out, which are displayed in the familiar "ORTEP plots" [5] in crystal structure papers. Besides degrading the precision with which the molecular geometry can be determined, such vibrations can introduce a small systematic error into bond distances. Because it is hard to stretch a bond but easy to bend one, thermal vibration tends to make a bonded atom trace out an arc around its bonded partner atom. Since the centroid of the arc lies within the radius of the arc, the bond distance appears slightly too small. To reduce such effects and also to promote crystal stability, data collection is increasingly being carried out at low temperature. Low temperature is particularly important for macromolecular crystals, which tend to be very sensitive to radiation damage. While simply freezing such a crystal is likely to cause even more damage, initial treatment with a cryoprotectant such as 25–35% glycerol followed by flash-freezing stabilizes the crystal and provides improved data.

A side group on a molecule, such as a pendant ethyl or isopropyl group, may not be tightly surrounded by neighbor atoms within the crystal on all sides. Such a group in one molecule may flip back and forth between alternative conformations (*dynamic disorder*), or molecules in different domains within the crystal may adopt one or the other alternative conformation (*static disorder*). The disordered atoms have to be treated as occupying more than one site, with each one having fractional occupancy.

*Crystal Size versus Beam Size* A crystal structure determination relies on the measurement of X rays or neutrons diffracted by the sample crystal. Because a very small crystal can only provide weak diffracted beams with a poor signal/noise ratio, there is an understandable desire to work with large specimen crystals. However, the maximum allowable size of crystal is limited by several sources of serious error. First of all, the crystal must fit within the uniformly intense portion of the beam irradiating it. If part of the crystal projects beyond this region, then as the crystal is turned during the measuring process, the mass of material interacting with the beam is likely to vary, introducing errors.

*Absorption* Irradiated crystals not only diffract the incoming beam, they tend to absorb some of it too. "Heavy" atoms, actually those with a high atomic number, are effective at absorbing X rays as well as diffracting them. From data in the International Tables for X-ray Crystallography, also stored in popular software such as the SHELX suite [6, 7], it is possible to calculate an absorption factor $\mu$. As a rule of thumb, if no dimension of the crystal exceeds $1/4\mu$, absorption will not be much of a problem. For a crystal composed only of C, H, N, and O atoms, this will allow the 0.4 mm size mentioned earlier for X rays. If heavy atoms are present, the optimum size may be much smaller. Larger crystals will require an absorption correction, supported by measurements if at all possible.

*Extinction* This phenomenon requires mathematical analysis, but an intuitive explanation is that when a large perfect crystal is efficiently producing a diffracted beam, the incident beam is weakened before it reaches deeper layers of the crystal. Thus, these deeper layers produce less than their expected contribution to the diffracted beam. The discrepancy is most pronounced when diffraction is most efficient. Good correction routines are available in many software packages.

***Powder Samples and Their Usefulness*** As will be shown subsequently in the brief section devoted to theory, the crystal must be set at exactly the appropriate angle to the incoming beam in order to generate a particular diffracted beam, and this operation must be repeated many times to obtain a full set of measurements. This requirement adds to the complexity of the apparatus but makes it clear which diffracted beam is currently being measured. If the sample is a powder instead of a single crystal, it will contain a huge number of tiny crystallites. For any possible diffracted beam some of these crystallites will be at the appropriate angle to the incoming beam, and thus a large number of diffracted beams can be measured all at once. The measurement process is both cheaper and quicker. However, the attribution of measured intensity to particular diffracted beams is lost, and interpretation requires pattern matching, much like the fingerprint region in an infrared spectrum. If reliable standard powder patterns exist for a particular drug, this technique can be valuable for quality control. The Powder Diffraction File, maintained by the International Centre for Diffraction Data (www.icdd.com), provides a database of such measure-

ments. In particular, unlike chemical analysis, powder diffraction can make clear which polymorph is present in a sample.

Recently, methods have been developed that make it possible to carry out a full three-dimensional structure determination from a powder sample. These are not routine procedures and may depend on good fortune in initial trial-and-error attempts to find a starting model, but they are of great value when all attempts to produce larger crystals end in failure.

### 2.2 Radiation Source

***Sealed Tube*** The humble X-ray tube is a technological marvel! As shown in Figure 1, a focused beam of electrons is accelerated by a high voltage, typically around 50 kV, and smashes into a metal anode, usually made of copper or molybdenum. In the sort of fine-focus tube generally used for single crystals of drugs, the same power as a domestic electric heater (1.5 or 2 kW) falls upon a spot $0.4 \times 8$ mm. A rapid circulation of cooling water is essential to carry away the heat produced; and if this circulation fails, the tube will be a write-off. The water circuit is protected by an internal filter in the tube and possibly by an external filter as well. Cleaning these filters is a necessary chore most laboratory staff try to avoid, but restricted water flow may lead to overheating and distortion of the anode that will impair the beam quality. Delivering a focused electron beam requires a vacuum in the tube. Molecules from the air slowly leak in but are driven to the walls by collisions with the electron beam. A stored X-ray tube should be operated every few months so that this self-cleaning mechanism can work.

Many of the electrons impacting the anode are stopped in their tracks. An accelerated (in this case decelerated) charge emits radiation, and for the energies described here the radiation comes in the form of X rays. This "white" radiation has a continuous range of wavelengths and energies and is useless for most purposes. Some of the bombarding electrons knock out inner shell electrons from the metal atoms in the anode. Outer shell electrons fall in to fill the vacancy, emitting X rays of defined energy and wavelength that are characteristic of the target metal. This wavelength is approximately 1.54 Å for copper and 0.71 Å for molybdenum (1 Å = $10^{-10}$ m). The critical excitation voltage that is just sufficient to enable a bombarding electron to knock out an inner electron is 9 kV for copper and 20 kV for molybdenum. Greater efficiency is achieved at higher voltage; hence 40–55 kV is typically used.

The long narrow focal spot is viewed almost edge-on at a take-off angle of around 6° by a collimator typically 0.4 to just over 1 mm in diameter, directing a parallel beam of X rays at the sample. Unwanted wavelengths of radiation are removed by a monochromator containing a crystal, usually graphite, which reflects them in directions where they miss the collimator and get absorbed by the metal housing. The desired radiation is Kα, meaning that an electron has been ejected from the innermost K shell of the target and replaced from the adjacent L shell. Kβ is produced when the replacement electron comes from one shell farther out, the M shell. Along with all the white radiation, it needs to be removed by the monochromator.

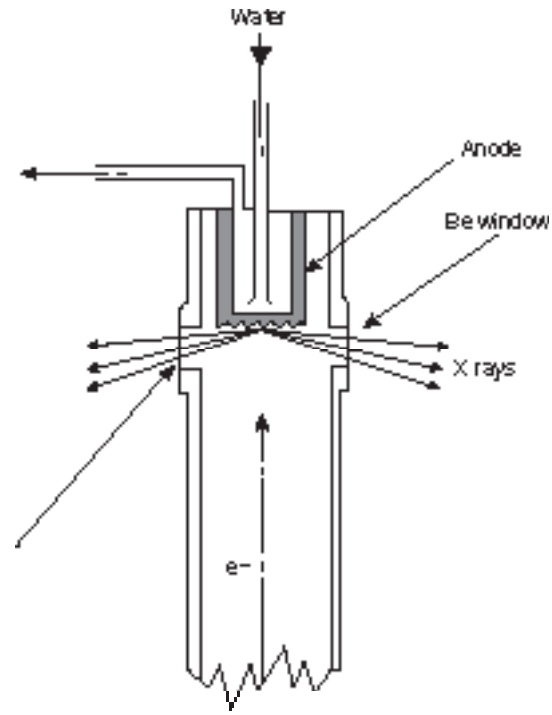


**Figure 1** Schematic diagram of an X-ray tube used for crystallography [8]. (Reprinted with permission from IUCr.)

***Rotating Anode*** Higher power can be achieved by replacing the static anode with a rotating disk so that the heat production is spread out along the circumference. The enhanced beam intensity is particularly useful in protein crystallography, where a large quantity of data must be collected before the often unstable crystal degrades. However, the provision of cooling and vacuum seals for a rotating object is much more difficult. An active pumping system is required to maintain the vacuum, and acquisition and maintenance costs are higher.

***Focusing Optics*** The operation of an X-ray tube is inherently wasteful, with much of the energy input going into the production of heat or X rays of the wrong

wavelength. Furthermore, even those X rays with the correct wavelength may be emitted in the wrong direction. Devices to redirect these wayward X rays toward the sample can provide a useful enhancement of beam intensity.

***Synchrotron*** Once again the principle that an accelerated charge emits radiation is applicable. The electrons constrained in a circular path in a synchrotron undergo never-ending centripetal acceleration and therefore emit radiation. In the form of X rays this radiation posed a safety hazard for the early accelerator physicists. More recently, publicly funded synchrotrons have been purposely built in most major countries to serve X-ray scientists. Insertion devices (wigglers or undulators) bend the beam to generate X rays of a desired wavelength at a predetermined location. Access to such beamlines is now available to many scientists by competitive application, and some pharmaceutical companies have purchased block bookings of beam time for their researchers.

The main advantage of a synchrotron source is that the limitation on intensity posed by the heat tolerance of a target is removed. The much more intense beam enables data to be collected much more rapidly. Another great advantage is that the X-ray wavelength is not fixed by the choice of target metal but instead can be tuned. This permits the optimal use of anomalous scattering, which will be described later.

***Neutron Diffraction*** All the radiation sources discussed so far produce X rays. The first step in the process of X-ray diffraction, to be explained in more detail later, is where the X rays are absorbed and then reemitted. In ordinary matter this process is carried out most efficiently by the electrons, and the final product of the analysis is a map of electron density in the molecule. For nearly all elements the electron density around an atom is virtually spherical and the location of that atom is easily inferred. However, the only electron possessed by a covalently bonded hydrogen atom is shared with another atom. The apparent position of such a hydrogen atom deduced from the electron density is too close, by roughly 0.1 Å, to the other atom, and the location is poorly defined. Unfortunately, in many pharmaceutical molecules this location is exactly the information most needed since hydrogen bonding strongly influences their structure and mechanism and it may be important to know whether a hydrogen atom has been transferred from an acid to a base. Neutrons, on the other hand, interact with the atomic nuclei and reveal their location accurately, even for hydrogen atoms.

The only practical neutron sources are at large centrally operated facilities, either nuclear reactors or spallation sources. Fission reactors operate by means of a neutron reaction (SNS) in which a fissionable nucleus absorbs a neutron, becomes unstable, and splits into smaller fragments including a number of neutrons that can attack other fissionable nuclei. Some of these surplus neutrons can be led away from the reactor core. Although we usually think of neutrons as particles, they also have wave properties that can be exploited for diffraction. After some contact with materials at room temperature their energy is reduced to the point that their wavelength is suitable for crystallography. To increase the intensity of the neutron beam, fission reactors have been designed for maximum power density and hence neutron production within a small core. Such design evolution runs up against the need to maintain cooling and control of the core, along with the nervousness of residents living near the reactor.

Many recent high-intensity sources, for instance, the American facility at Oak Ridge, TN, commissioned in 2006, use the alternative spallation technique. A pulsed beam of protons from an accelerator is fired at a heavy metal target with neutron-rich nuclei. As the high-energy protons strike the nuclei, neutrons are released. Where a solid metal target is used, the risk of damage to the target has limited the proton flux that could be directed at it. The Oak Ridge SNS uses liquid mercury as the target for the first time. With constant recirculation of the mercury, a much higher intensity can be tolerated.

Because of the weakness of the available beams, the application of neutron diffraction has up to now been limited to large crystals typically greater than 1 mm on a side compared to one-tenth that length for X-ray diffraction. It is hoped that the new facilities will provide satisfactory neutron data from more normal-sized crystals.

### 2.3 Detectors

The development of detectors for X rays has followed a pendulum process, alternating between point and area detectors. When the Bragg family, father and son, made their first accurate measurements, they used an ionization spectrometer positioned to intercept one diffracted beam at a time. Subsequently, this temperamental piece of apparatus was replaced by photographic films that could record many diffracted beams at once in the form of spots. However, the intensity of each spot had to be laboriously evaluated by visual comparison with calibration spots. Robust and reliable scintillation counters were introduced in the early models of diffractometers, providing automatic readout of each diffracted beam in succession. Such one-at-a-time measurement inevitably takes days or weeks for collection of a full set of data

and often exceeds the lifetime of an irradiated protein crystal. The need for greater speed has prompted a return to area detectors in modern diffractometers.

Charge-coupled device (CCD) detectors generally have a thin phosphor screen that absorbs the energy of incident X rays and reemits it as visible light. A tapered optical fiber conveys the light to a doped semiconductor chip that provides a readout of intensity versus position by a process related to television camera technology. Multiple CCDs can be placed side by side in a matrix to enhance the area covered. Multiwire proportional counters provide a similar readout by using an alternative process of ion-pair generation in a gas-filled tube. Image plates are analogous to films. Diffracted X rays striking a phosphor excite it into a metastable state. The exposed plate is scanned with a laser operating at one wavelength, which stimulates release of the trapped energy at a different wavelength with an intensity that is proportional to the intensity of the original X-ray beam that struck that place on the plate. After scanning, the plate can be reused.

## 3 BRIEF HIGHLIGHTS OF NECESSARY THEORY

In the short sections that follow it is impossible to convey the full depth of knowledge needed by a practicing crystallographer. For this purpose the excellent courses run by national or regional crystallographic societies are recommended. Teaching materials from the British Crystallographic Association have been published in book form [9]. Around two dozen teaching pamphlets written by experts on specific topics for a carefully defined audience, usually scientists in other fields and students, are available online from the International Union of Crystallography (http://www.iucr.org/ and then select Teaching Pamphlets). The goal in this chapter is to enable understanding of the steps in a crystal structure determination and the ability to read critically a paper describing a crystal structure.

### 3.1 Introduction to Lattice and Space Group Symmetry

A crystal is composed of a regular repeating array of atoms, molecules, or ions. If these units are conceptually simplified to points, the repetition generates a three-dimensional *lattice*. Adjacent lattice points in three dimensions define a block called the *unit cell*, with lengths $a$, $b$, $c$ along the $x$, $y$, $z$ axes; and the crystal can be generated by placing a myriad of unit cells side by side with edges in register. There is no requirement that $a$, $b$, $c$ should be the same. Neither do the $x$, $y$, and $z$ axes generated by vectors from one lattice point to its neighbors have to be perpendicular. This completely general case is called the triclinic *system*. Other systems are more restricted: monoclinic has $y$ perpendicular to $x$ and $z$, orthorhombic has all three axes mutually perpendicular, tetragonal additionally has $a = b$, hexagonal also has $a = b$, but at a 120° angle, rhombohedral has all three axes equal in length but nonperpendicular, and cubic has all three axes equal and perpendicular. With lattice points just at the corners of the unit cell, the arrangement is called primitive, but it may give more efficient packing to have *centering*, that is, extra lattice points in the middle of faces or at the midpoint of the entire unit cell. There are 14 permissible combinations of crystal systems and centering types, termed the Bravais lattices.

Another way in which molecular crystals may attain improved packing is to place additional molecules into the unit cell in different orientations. These molecules may be related to the original molecule by a strict application of symmetry, or they may be independent (the so-called $Z' > 1$ situation). Racemic mixtures frequently do not separate upon crystallization. Instead, both enantiomers occupy one unit cell, related by an inversion center. Two molecules related by rotation or by mirror reflection may also occupy the unit cell. Simple rotations or reflections are rare because they often cause one part of the rotated or reflected molecule to bump into part of the original molecule. However, screw axes (rotation accompanied by translation through ½ or another fraction of a unit cell length) and glide planes (reflection accompanied by translation through ½ of a unit cell length) are common because the interleaving they produce is conducive to good packing. The combination of such symmetry elements determines the *space group*. The space group is essential information in any crystal structure publication. The most common space group is $P2_1/c$, containing all types of symmetry element just mentioned. The symbol $P$ signifies a primitive lattice, the $2_1$ refers to a screw axis, and $c$ to a glide plane. The combined operation of $2_1$ and $c$ implicitly generates an inversion center.

### 3.2 Physical Process: Diffraction

Being electromagnetic radiation, an X ray impinging upon an atom can accelerate an electron, which will then reemit the X radiation. The X rays reemitted by different electrons in the atom interfere with each other. This also applies to those from different atoms in the molecule, different molecules in the unit cell, and the contents of different unit cells. The end result is the production of an extremely complicated three-dimensional diffraction pattern that strains human comprehension. The important point to remember is that

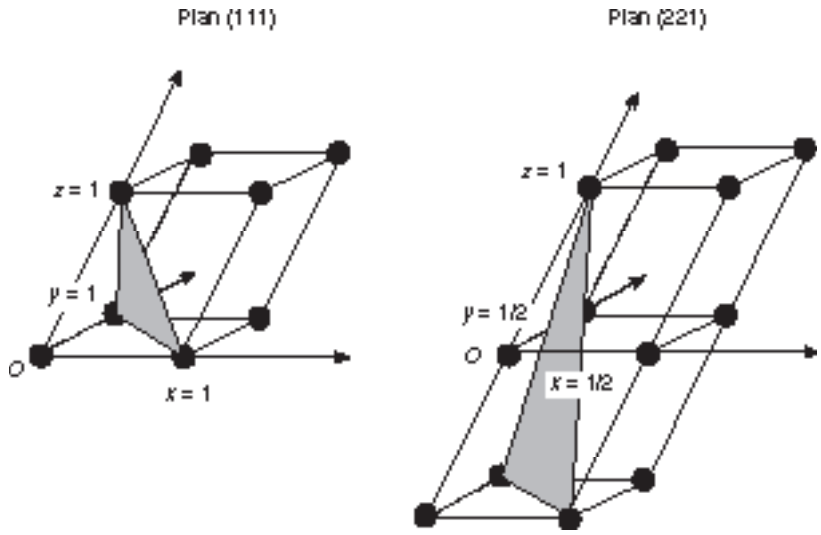


**Figure 2** Illustration of the reference planes for calculation of Miller indices for the 111 and 221 stacks of planes [10].

diffraction of X rays is effectively caused by electrons, and the ultimate result of computer analysis of a measured diffraction pattern is a map of the electron density.

### 3.3 Understandable Analogy: Reflection from Crystal Planes. Miller Indices, Bragg's Law, and the Meaning and Importance of Resolution

The Braggs devised a much more comprehensible picture. Imagine stacks of planes cutting through the unit cell. These stacks are capable of reflecting X rays, and together they create the diffraction pattern. Since we may need to consider thousands of stacks differing in spacing and orientation, a clear system of nomenclature is essential. This requirement is satisfied by the Miller indices, which are defined next.

Arrange a stack of planes so that one plane goes through the origin of the unit cell. The next plane in the stack is the reference plane. This plane is shown in both diagrams in Figure 2. Note at what fraction of a unit cell length $a$, $b$, or $c$ the reference plane cuts the $x$, $y$, or $z$ axes. (If it is not a simple fraction, the stack is of no importance.) Now take the reciprocal of each fraction to obtain the Miller indices. It may happen that the reference plane is parallel to one of the cell axes. In that case the intercept is said to be $\infty$, and the Miller index is $1/\infty = 0$. When algebra is to be done on general values of the Miller indices, the symbols $hkl$ are customarily used.

Reflection of X rays from a stack of planes shows some similarities and some differences compared with what we see reflected in a hand mirror. As with the hand mirror, angle of incidence = angle of reflection.

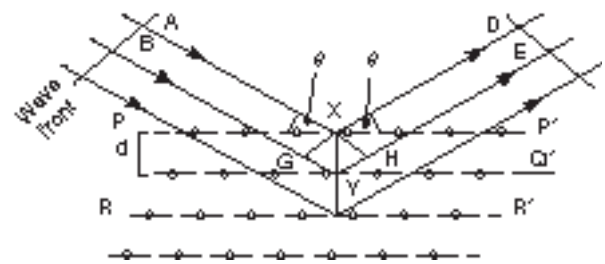


**Figure 3** Diagram of incident and reflected beam paths for reflection from successive planes in a stack [8]. (Reprinted with permission from IUCr.)

Crystallographers measure these angles between the ray and the plane surface and assign the symbol θ. However, while the hand mirror provides a reflected view whatever its orientation, this is no longer true for a stack of mirrors. Reflected waves from successive planes in the stack must interfere constructively, and this can only happen if the extra distance traversed by waves reflected from the lower mirror, GYH in Figure 3, is equal to an integral number of wavelengths λ. This is the basis of Bragg's law:

$$n\lambda = 2d\sin\theta$$

where $n$ is an integer and $d$ the perpendicular spacing between planes in the stack.

Usually, we work with first-order diffraction so that $n = 1$, and the wavelength λ is most commonly 1.5418 Å for a copper target or 0.7107 Å for a molybdenum target (1 Å = $10^{-10}$ m). Thus, the left-hand side of the equation is constant during an experiment. For a typical crystal

of a small-molecule drug, several thousand stacks of planes have to be set to the appropriate angle θ so that the intensity of reflection can be measured.

This principle leads to the important concept of resolution. To maintain a constant left-hand side, as *d* becomes smaller, sin θ must increase. However, sin θ can never exceed 1, so this determines the minimum spacing between planes that can be observed. This physical limit may not actually be reached due to poor crystal quality or limited availability of measuring time. The smallest spacing between planes actually measured in an experiment is called the *resolution*, and it determines the level of detail that can be observed. Consider the phenyl ring in the amino acid phenylalanine in a protein. The distance across this ring is 2.8 Å, while a carbon–carbon bond length is 1.4 Å. Quite a few protein crystal structures have been measured to 2.8 Å resolution, which would only show the phenyl ring as a vaguely hexagonal blob. Improving the resolution to 1.4 Å should begin to make it possible to see atoms. Advances in cryoprotection of protein crystals and the wider availability of synchrotron beam time have provided increasing numbers of protein crystal structures to resolution even better than 1.4 Å, and "small-molecule" structures of druglike molecules should beat 0.8 Å. The key point is that "high-resolution" is associated with a low number for the *d* spacing, and "low resolution" with a high number.

### 3.4 Relationships between Measured Intensities, Structure Factors, and Electron Density

The measured intensities have to be corrected for background scattering of X rays not due to the diffraction process and by factors relating to the geometry of data collection. The square root of the corrected intensity of reflection from each stack of planes with Miller indices *hkl* is called the *structure factor F(hkl)*. A structure factor will be large if a high proportion of the electron density lies in the planes but small if the electron density is smeared out between planes. Thus, a relationship exists between the full set of structure factors and the electron density at any point within the unit cell. Atoms appear as peaks in a map of electron density. The standard mathematical technique of Fourier transformation, implemented in crystallographic computer programs like the SHELX suite [6, 7], is capable of generating an electron density map from the measured structure factors. However, it requires a key piece of information for each structure factor that is missing from the measurements: the *phase*. The phase of a wave determines whether, at a reference location and a reference time, there is a crest, a trough, or a point in between.

### 3.5 Phase Problem and How It Can Be Solved

There are three main ways to supply the missing phase information, that is, to solve the phase problem. The first method is to set all phases to zero and supply squared structure factors to the Fourier transformation. This produces not a map showing atoms as electron density peaks but rather a map showing vectors between atoms (a Patterson map). For a simple diatomic or triatomic molecule it is not difficult to deduce the locations of the atoms from the pattern of vectors, and some early crystal structures were solved this way. Even for a molecule the size of an aspirin, though, the Patterson map becomes too complicated a molecule to interpret easily. Fortunately, the strength of a vector between two atoms is proportional to the product of their atomic numbers. If just a few atoms are much "heavier" (actually, have a much higher atomic number) than the others, the vectors between them will stand out from the others. This fact makes the *heavy-atom method* feasible. Many early crystal structures of basic drugs were solved as hydrobromide salts, the bromine acting as a heavy atom. In the case of the rubidium salt of benzylpenicillin, the vectors involving rubidium atoms did not stand out enough. However, both the potassium and rubidium salts had the same unit cell dimensions and mode of packing, that is, they were *isomorphous*. In the 1940s Dorothy Hodgkin collected data on both salts and compared the vector maps [11]. Only the vectors involving the cations changed in intensity between the two data sets, and thus *single isomorphous replacement* enabled the heavy atoms to be located. Due to the large number of atoms in a much bigger protein molecule, even single isomorphous replacement may be inadequate, and multiple isomorphous derivatives may have to be made. Fortunately, protein crystals typically contain 40–80% water [12], and it is often possible to diffuse positively and negatively charged ions into the crystal, where they displace water and bind to different sites on the protein molecule without disturbing the packing.

Having located the heavy atom(s), one still needs to find the rest of the structure, which is the drug or protein of actual interest! Starting from the electron density of the heavy atoms alone and running the Fourier transformation backward from electron density to structure factors yields crude values for the magnitude and phase of each structure factor. The roughly estimated magnitudes can be replaced with the measured values while the approximate phases must be kept, for lack of anything better. Running the Fourier transformation forward now produces an electron density map that, thanks to the use of correct structure factor magnitudes, reveals additional atoms. The additional atoms make it possible to calculate a better approximation to the

phases. By such a bootstrapping process the entire structure can be teased out.

This "Fourier refinement" technique for structure improvement is useful for completing the structure starting from approximate phase information, however determined. It also is the basis for the second method of solving the phase problem, *molecular replacement*, which has greatly facilitated the determination of protein structures. For instance, if the structure of an enzyme found in the mouse has been laboriously solved by multiple isomorphous replacement, it is likely that the human enzyme structure will not differ very much. A set of phases calculated from the mouse structure, rotated and translated when necessary, applied to structure factors measured for a crystal of the human enzyme may provide an instant starting point for the location of atoms.

The third, very popular, procedure for phase determination involves the so-called *direct methods*. If the Miller indices *hkl* for three stacks of planes designated 1, 2, 3 have the relation

$$h1 + h2 = h3,\ k1 + k2 = k3,\ l1 + l2 = l3$$

then it is likely though not certain that the phase of reflection from the third stack is the sum of the phases for stacks 1 and 2 (modulo 360°). Thus, knowing a few phases, one can calculate more. Of course, at the start, no phases at all are known. In practice, some phases have to be systematically or randomly guessed, and the right guess will lead to consistent predictions while wrong guesses lead to chaos. By direct methods implementations such as SHELXS [6], it is routine to solve the structures of drugs containing 100 atoms or more, and small proteins are yielding to this approach as well.

### 3.6 Completion of a Structural Model

In the small-molecule context an experienced crystallographer is unlikely to make significant mistakes. Increased automation is encouraging nonspecialists also to determine crystal structures, and most of the time they come up with satisfactory results. However, even in small-molecule crystallography the interpretation of peaks on an electron density map is not always straightforward. While the peak heights ought to be proportional to the number of electrons, someone working hastily may still confuse the location of nitrogen atoms and CH groups in a heterocycle. Chemical common sense and clues such as hydrogen bond acceptance or donation should always be respected.

Another issue concerns hydrogen atoms. Before the 1970s structures were often reported without such atoms. As equipment improved and computing power increased, hydrogen atoms were explicitly located. Sometimes the geometry was strange, but one could be sure a peak had appeared at the site. The modern practice is to place hydrogen atoms in calculated positions based on standard bond distances and angles. The software is very convenient; but given erroneous instructions, it will obediently add nonexistent hydrogen atoms.

The generally lower resolution obtainable with macromolecular crystals increases the dependence on models for the location of atoms. Standard bond lengths and angles in amino acids have been compiled from relevant small-molecule structures for use in refining protein structures [13]. Sometimes the path of the main chain itself may initially be traced erroneously. It is valuable to carry out checks by deleting portions of the molecule and verifying that they come back in an electron density map.

### 3.7 Refinement

Inevitably the atomic positions found in an initial map can be improved so as to give a better fit to the data. Thermal vibration is represented by displacement parameters, assuming that the atom traces out either a sphere (isotropic, one parameter) or an ellipsoid (anisotropic, six parameters). Positions and displacement parameters are refined by a nonlinear weighted least-squares technique. To avoid dominance by a few very large structure factors or degradation of the model by numerous weak ones with a poor signal-to-noise ratio, a weighting scheme is applied to the structure factors. The refinement process does not converge in a single cycle but requires iteration until the shifts in parameters become small, typically <0.01 times their estimated uncertainty.

### 3.8 Anomalous Scattering

When the wavelength of X rays striking an atom is systematically varied, near a certain wavelength called the *absorption edge*, the absorbing power of the atom changes drastically. The scattering of X rays is also affected. Near the absorption edge an electron affected by an X-ray behaves like a damped oscillator, and the phase of the reemitted radiation is changed. The effect on structure factors is that *F(hkl)* is no longer equal to *F(-h,-k,-l)*, and this renders the experiment sensitive to inversion. If a single enantiomer of a chiral drug molecule is present in the crystal, either one form of the structure or its inverse (i.e., left or right hand) will give a slightly better fit to the data and is therefore correct. For greatest certainty all structure factors, both *F(hkl)* and *F (-h,-k,-l)*, ought to be measured and compared; but the Flack parameter [14] has been devised to give

an indication of correctness during refinement of an ordinary data set.

Anomalous scattering has an additional use in macromolecular crystallography. On a synchrotron it is possible to change the wavelength of the X rays to encompass values near to and far from an absorption edge of a heavy atom. The altered scattering power of the heavy atom makes it act like two different elements in the two experiments, thus simulating isomorphous replacement with zero disturbance.

### 3.9 Expressing the Quality of a Crystal Structure Determination

The reader of a crystal structure paper should take note of the quality measures that usually are published. An excellent guide is the *R factor* or *discrepancy index*. For each structure factor, the observed value $F_o$ and the calculated value $F_c$ are compared, and the absolute values of the differences are added up for all the measured data. This sum is divided by the sum of all $F_o$. A small discrepancy between $F_o$ and $F_c$ values confirms both good-quality data ($F_o$) and an appropriate structural model ($F_c$). The perceived importance of the $R$ factor has made it the subject of some cosmetology. It is customary to calculate it only for structure factors that have a value greater than twice their estimated standard deviation, thus eliminating measurements with a poor signal-to-noise ratio. Suppression of any other data gives grounds for suspicion. Expected $R$ factors for small-molecule structures are <0.05 for excellent data and <0.10 for respectably mediocre data. For macromolecular structures much higher values are acceptable. After least-squares refinement the weighted $R$ factor $wR(F^2)$ is also reported. Based on squared structure factors and covering all data, it is more sensitive to discrepancies than the conventional $R$ factor and frequently attains a value more than twice as great.

Of direct relevance to the description of molecular geometry is the standard uncertainty of a typical C–C bond length, which in the small-molecule context should be <0.005 Å for excellent work and should not exceed 0.01 Å except perhaps for bonds affected by problems such as disorder. For macromolecular structures the uncertainty is much greater, usually exceeding 0.1–0.3 Å. Other quality measures that are published in crystal structure papers include the goodness of fit, which should be near 1 if the weights used in refinement are suitable, and the residual electron density, which should be as low as possible. Residuals of ±0.3 electron / $Å^3$ merely indicate harmless noise (larger magnitudes are tolerable near heavy atoms), but significant positive peaks suggest that an atom may have been missed and significant negative regions suggest wrongly positioned atoms. The International Union of Crystallography (www.iucr.org) offers useful software online to help authors check the validity of their results. The multifunctional program PLATON is particularly helpful.

## 4 DETAILED APPLICATIONS AND CASE STUDIES

### 4.1 Determination of Unknown Structures

The work on benzylpenicillin by Dorothy Hodgkin [11], which established the previously controversial β-lactam structure, has already been mentioned. More recent accomplishments include determining the structure of a precursor of the anticancer drug paclitaxel [15] and then paclitaxel itself [16], complicated drugs made from natural products, for which NMR studies had left some issues unresolved. In the macromolecular context studies on human immunodeficiency virus (HIV) protease not only elucidated this previously unknown structure but also provided the springboard for the design of an important class of new drugs.

### 4.2 Salt Selection and Its Importance

Flurbiprofen is a nonsteroidal anti-inflammatory agent with a carboxylic acid group (Fig. 4). Its aqueous solubility is only 0.03 mg mL$^{-1}$ for the free acid. Amine salts of flurbiprofen were evaluated as possible means to attain higher solubility.

In Table 2 it is evident that aqueous solubility for this series with a constant anion does not depend simply on the hydrophilicity of the cation. There must be differences in stability of the crystals formed: crystals of the *n*-butylammonium salt, which has the most hydrophilic cation, must be exceptionally stable to account for its

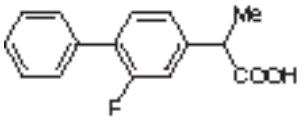


**Figure 4** Structural formula of flurbiprofen.

**TABLE 2 Aqueous Solubility of (aryl)alkylammonium Salts of Flurbiprofen [17]**

| Ammonium Cation | Solubility (mg/mL) |
|---|---|
| *n*-Butyl | 0.19 |
| *n*-Pentyl | 7.51 |
| *n*-Hexyl | 2.8 |
| Cyclohexyl | 0.37 |
| Benzyl | 2.85 |

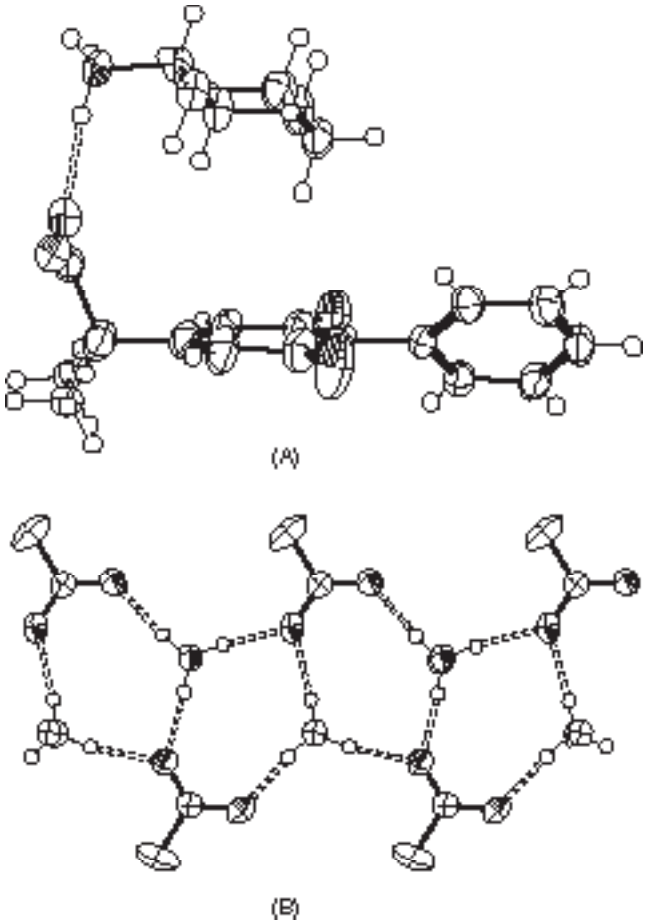


**Figure 5** (A) Structures of the cyclohexylammonium cation and flurbiprofen anion showing their interaction in the crystalline salt [17]. (Reprinted with permission from Taylor & Francis, http://www.informaword.com.) (B) Ladder motifs produced by hydrogen bonding between ammonium and carboxylate moieties. For clarity only the carbon atoms directly attached to these groups are shown.

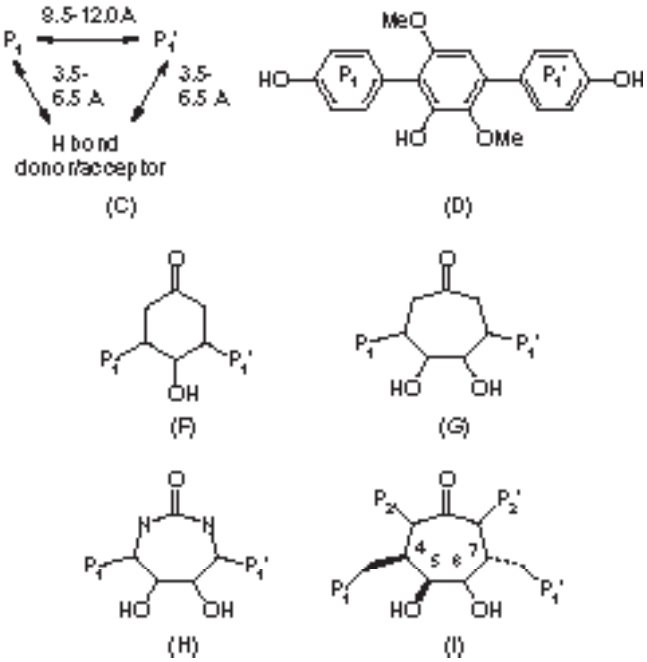


**Figure 6** Structures illustrating steps in the development of potent inhibitors of HIV protease [18]. (Partial figure with permission from AAAS.)

low solubility. The crystalline cyclohexylammonium salt must be more stable than the *n*-hexylammonium variant. Crystals of the cyclohexylammonium salt were of sufficient quality for a structure determination [17] and demonstrated a tidy segregation into nonpolar hydrocarbon domains (Fig. 5A) and hydrophilic, hydrogen-bonded regions that make optimal use of the three hydrogen atoms on the ammonium group (Fig. 5B). Such favorable intermolecular interactions are expected to lower the energy of the crystal and would be lost upon dissolution.

### 4.3 Structure-Based Drug Design

The discovery of new life-saving drugs is arguably the most inspiring application of the entire science of crystallography. HIV protease inhibitors are a notable success story. The study by Lam et al. [18] shows with particular clarity the iterative application of crystallography, modeling, and synthesis, which is conceptually summarized in Figure 6. These authors noted in previously published structures of protease–inhibitor complexes that a structural water molecule, unique to retroviral proteases, accepts two hydrogen bonds from backbone amide groups of the protease and donates two hydrogen bonds to carbonyl oxygen atoms of the inhibitor. They sought a novel inhibitor incorporating a feature that binds to HIV protease like this water molecule, reasoning that with no loss of binding energy it would achieve an entropy gain by freeing the structural water molecule. Furthermore, toxicity should be limited since mammalian proteases lacking such a removable water molecule would provide a poor fit.

The authors began by building a computer model of a fairly simple diol with twofold symmetry docked into the active site, from which they derived a pharmacophore model (Fig. 6C). In this simple pharmacophore hydrophobic groups $P_1$ and $P_1'$ were placed 8.5–12.0 Å apart, each one 3.5–6.5 Å from a common hydrogen bond donor/acceptor. A search of their small-molecule database, a subset of the CSD, revealed a molecule (Fig. 6D) that not only met the criteria but also could position an oxygen atom at the water site. Because the rigid spine of its terphenyl skeleton provided little

opportunity for "tweaking" substituent positions, a 3,4,5-trisubstituted cyclohexanone skeleton (Fig. 6F) was proposed as an alternative offering more chance for development. The carbonyl oxygen atom was expected to accept hydrogen bonds from the protease in the same way as the original water molecule. The 4-hydroxycyclohexanone was conceptually enlarged to a cycloheptanone (Fig. 6G) incorporating two adjacent hydroxyl groups like the original diol. This ring system was altered to a cyclic urea (Fig. 6H) in order to feed extra electron density into the carbonyl oxygen atom and make it a stronger hydrogen bond acceptor. Attention now shifted to the optimal positioning of substituents on the otherwise unoccupied carbon and nitrogen atoms of the ring (Fig. 6I). It was predicted that *4R*, *5S*, *6S*, *7R* stereochemistry would give best complementarity to the protease. After synthesis of such a molecule, small-molecule crystallography confirmed its preferred conformation, and biochemical studies verified tight binding. The research then entered a cyclic phase of synthesis, biochemical evaluation, and crystal structure determination of the drug–protease complex, each cycle providing ideas for the further optimization of the substituents. A total of 10 structures were thus determined, $K_i$ values as low as $0.27 \pm 0.05$ nM were attained, and no inhibition of representative cellular aspartyl proteases was observed at drug concentrations 3000 times the $K_i$ for HIV protease.

On the other hand, imatinib (Gleevec), a "silver bullet" specifically directed against chronic myeloid leukemia, was not and probably could not have been discovered by a structure-based approach. Imatinib blocks the adenosine 5′-triphosphate (ATP) binding site in a kinase encoded by the chimeric Philadelphia chromosome that does not occur in normal cells. Because every cell needs ATP, it is difficult to imagine a committee allocating funds for the design of an ATP blocker. In any case, the structure of imatinib shown in Figure 7 is so different from that of ATP that no analogy would have been spotted.

After its discovery the crystal structure of imatinib bound to its target has been determined [19]. This structure still supplies useful information since it suggests how certain mutations of the target enzyme confer resistance, and how such resistance might be circumvented by structural modifications.

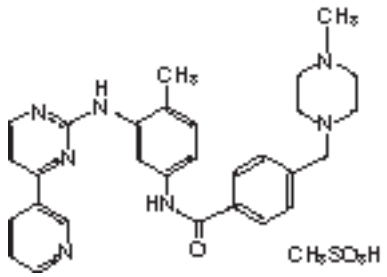


**Figure 7** Structural formula of imatinib.

The diffusion of heavy ions into macromolecular crystals to provide isomorphous derivatives as a means of structure solution has already been mentioned. Because of the high water content of protein crystals, solvent channels through the crystal may actually be wide enough to allow candidate drugs to diffuse [20]. Such soaking experiments do not always work. The extra bulk of the bound drug may cause the crystal to crack. Finding a solvent that dissolves the drug but does not damage the crystal may be difficult. It may be necessary to habituate the crystal to gradually increasing concentrations of the solvent. However, the results can be very rewarding. Pre-formed crystals can be soaked in a solution of the drug, left for hours or days to reach equilibrium, and used for X-ray data collection. With the protein structure remaining largely undisturbed, a trial structure can be rapidly obtained by the molecular replacement method. Thus, the entire process from ligand binding to structure determination can achieve relatively high throughput.

It has even been possible to determine binding constants by this technique [20]. Several soaking experiments in different concentrations of drug are allowed to reach equilibrium. By determining the structure of crystals from each batch, it is possible to refine a fractional occupancy parameter for the drug in its binding site. This value, along with knowledge of the ligand and total protein concentrations, enables a dissociation constant to be calculated.

### 4.4 Use of Databases

By comparing structures of related drug and nondrug molecules, it may be possible to identify structural features required for activity and establish a pharmacophore, even in the absence of the macromolecular target structure. Because of the immense therapeutic importance of the β-lactam antibiotics, such studies were carried out in advance of and in parallel with structure determination of the bacterial transpeptidase that they inhibit. The important moiety in penicillins, cephalosporins, and other β lactams is shown in Figure 8.

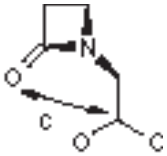


**Figure 8** Core structure of a β-lactam antibiotic illustrating the geometrical relationship between important functional groups.

The amide carbonyl carbon atom is attacked by a serine residue from the enzyme, and the nearby carboxylate ion provides an additional anchor point. Two geometric descriptors have been proposed: (1) the Woodward height-of-pyramid *h* measuring the deviation of the amide N atom from the plane of its substituents, which modulates its influence on the carbonyl group, and (2) the Cohen distance *c*, which represents the distance between the carbonyl oxygen and carboxylate carbon atoms. The Desiraju group subsequently showed [21] that consideration of both parameters in tandem improves their reliability, two separate "islands" of activity appearing at low *h*–low *c* and high *h*–high *c*.

Another application [17] involved a series of antimycobacterial carboxamidrazones whose general formula is shown in Figure 9. Because the spread of resistant strains of *Mycobacterium tuberculosis* necessitates long courses of treatment with multiple drugs, discovery of new classes of antimycobacterials would be very advantageous. Although the biological target is unknown, the best drugs in the amidrazone series showed similar minimum inhibitory concentrations [MIC] against the test organism *Mycobacterium fortuitum* to the clinically important drug isoniazid.

Crystal structure determinations on 15 compounds in this series provided structural data on 20 independent molecules. For most molecules an obvious long axis runs through the heterocycle, the central C-N8-N9-C unit, and the aryl ring (Fig. 10). It was postulated that the heterocyclic nitrogen atom accepts a hydrogen bond that anchors the drug in a receptor pocket running parallel to this long axis. With a bulky but short *t*-butyl substituent on Ar, both 2- and 4-pyridyl derivatives are active (MIC 8–16 μg/mL). When Ar is *p*-benzyloxy-*m*-methoxyphenyl with its greater length, the 2-pyridyl derivative is highly active (MIC 4–8 μg/mL) while the 4-pyridyl is inactive. These observations suggest a maximum permissible length and guide synthesis away from overelaboration of the long axis.

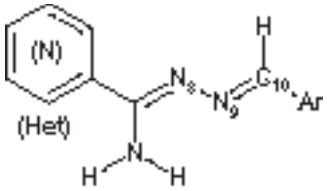


**Figure 9** Generic structure of the antimycobacterial carboxamidrazones.

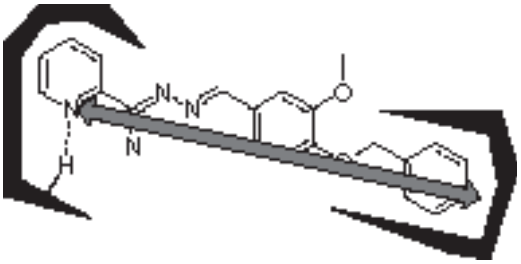


**Figure 10** Binding model deduced from crystallographic information and activity data for the antimycobacterial carboxamidrazones [17]. (Reprinted with permission from Taylor & Francis, http://www.informaworld.com.)

Of fundamental concern both for structure-based drug design and, eventually, for crystal structure prediction is the correct assignment of conformation in flexible drug molecules. The structure of a candidate drug molecule from molecular modeling is often calculated for the isolated molecule in a vacuum. If the structure is obtained by small-molecule crystallography, it represents the molecule surrounded by others of like kind. These differing environments could induce different conformations from the one adopted by the drug when bound to its target. The discrepancies were evaluated for a series of 33 small-molecule ligands from a wide variety of drug classes where model, small-molecule, and protein-bound structures were available [22]. The results were discouraging: Conformations differed widely, and root-mean-square (rms) discrepancies increased along with the number of flexible bonds.

More recently, a pair of review articles [23, 24] examined the situation specifically for dihydrofolate reductase (DHFR) inhibitors, for which many structures are available. A particularly well-studied molecule is the antibacterial drug trimethoprim (Fig. 11). Its pyrimidine and phenyl rings are linked by two bonds that can rotate, C4-C5-C7-C1′ and C5-C7-C1′-C2′. Steric hindrance precludes values of 0° or 180° for both together.

In Figure 12 the small-molecule data (squares) fall into two regions identified as I and II. Evidently the macromolecular target selects from this manifold of conformations since the great majority of such structures, whether binary (DHFR with drug) or ternary (DHFR + drug + cofactor) and whether determined by X-ray crystallography or NMR in solution, fall within region I, and the remaining three structures are close to II. These results for trimethoprim and its derivatives

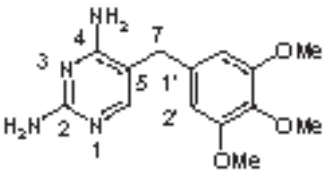


**Figure 11** Structural formula and numbering scheme for trimethoprim.

support a more encouraging conclusion: Although any one small-molecule structure may be unrepresentative of the protein-bound form, the consensus of many structures can be revealing.

Figure 13 shows a very important interaction between inhibitors and DHFRs: a hydrogen-bonded salt link from the protonated N1 and the adjacent 2-amino group of the drug to the side-chain carboxylate group of an active-site aspartate or glutamate residue in DHFR.

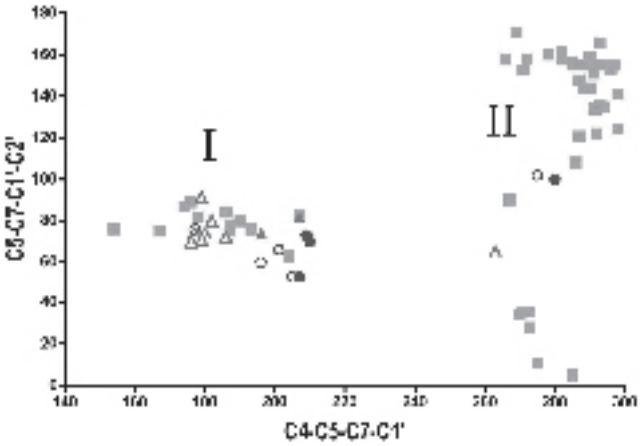


**Figure 12** Plot of bridging torsion angles for trimethoprim and its derivatives from small-molecule structures (squares) and bound to DHFR (triangles—ternary complexes with NADPH, circles—binary complexes; open—X-ray data; filled—NMR data) [24]. (Reprinted with permission from Taylor & Francis, http://www.informaworld.com.)

Crystal structures of small-molecule carboxylate salts enable the geometrical relationships between the protonated drug and the anion to be precisely defined. Comparison of N1 . . . O1 and N7 . . . O2 distances in Figure 14 (data from [23]) shows that N1 . . . O1 is usually the closer contact, presumably representing stronger attraction. Two points that stand apart from the main cluster of data are worthy of some attention. One point represents unusually long (>2.8 Å) values of both distances, N1 . . . O1 and N7 . . . O2. This structure [25], assigned code WOYPIJ in the Cambridge Structural Database, is a trifluoroacetate salt. Diversion of negative charge to the fluorine atoms may lessen the ability of the carboxylate group to attract the cationic drug, leading to longer contact distances. The second outlier, coded ULAXUA [26], has a short N1 . . . O1 distance

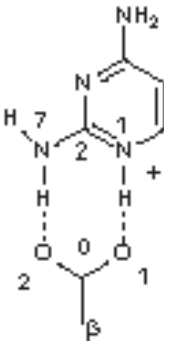


**Figure 13** Structural formula and numbering scheme for the hydrogen-bonded salt link between a protonated diaminopyrimidine antifolate drug and a carboxylate anion.

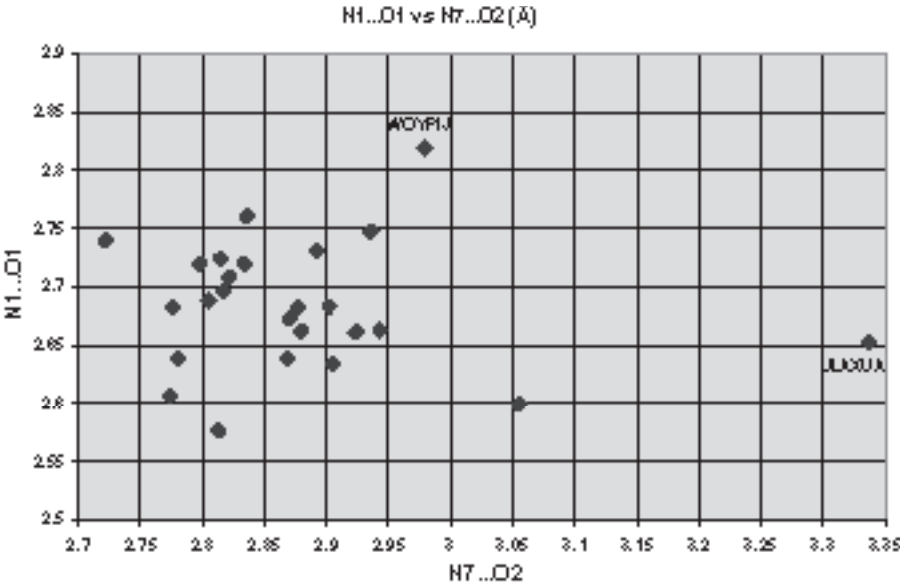


**Figure 14** Scatter plot of N1 . . . O1 and N7 . . . O2 distances in carboxylate salts of protonated diaminopyrimidine antifolate drugs.

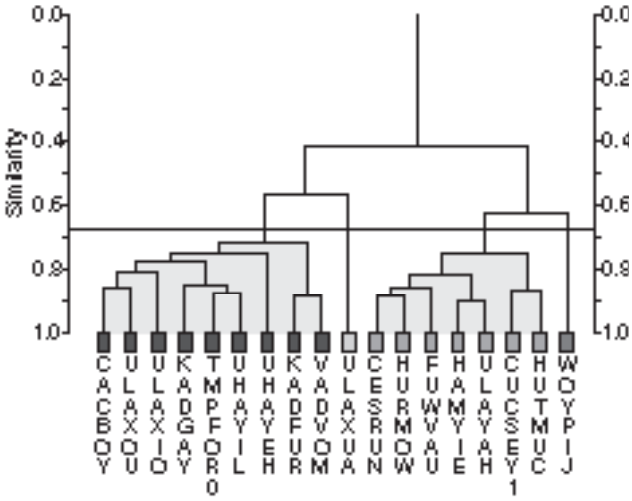


**Figure 15** Dendrogram comparing the geometry of carboxylate salts of protonated diaminopyrimidine antifolate drugs [23]. (Reprinted with permission from Taylor & Francis, http://www.informaworld.com.)

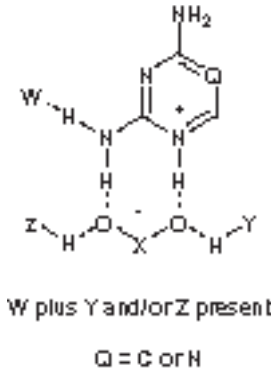


**Figure 16** Schematic drawing of the interaction of pyrimidine- and 1,3,5-triazine-based antifolate drugs with counterions and solvent molecules at the 1 and 2 positions [23]. (Reprinted with permission from Taylor & Francis, http://www.informaworld.com.)

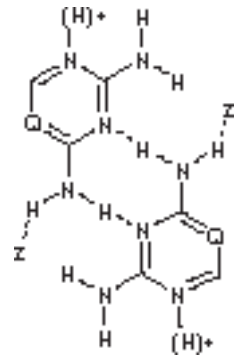

**Figure 17** Schematic drawing of base pairing of antifolate drugs at the 3 and 4 positions [23]. (Reprinted with permission from Taylor & Francis, http://www.informaworld.com.)

but a very long N7 . . . O2 (>3.3 Å), suggesting a large relative rotation of the ions.

A mathematical analysis of the geometrical relationship between the diaminopyrimidine (drug) and carboxylate moieties has been published [23]. Starting from a situation where these moieties are coplanar with equal N1 . . . O1 and N7 . . . O2 distances, the axial rotations needed to match the observed arrangements have been calculated. By defining the limits on distortions found in precisely determined small-molecule structures, it should give the drug designer an idea of permissible distortions in the drug–DHFR complex.

The technique of cluster analysis, implemented in the program dSNAP [27], provides an objective classification of similarities and differences. A variety of plot styles can be viewed interactively, but a simple dendrogram usually suffices to show groups of clearly similar structures and to isolate the outliers. Its application to an earlier version of the diaminopyrimidine carboxylate data set discussed previously [23] is shown in Figure 15. Two broad groups are evident; and, as in Figure 14, ULAXUA and WOYPIJ appear as outliers.

The DHFR review [23] went on to catalog the supramolecular hydrogen-bonded interactions of protonated pyrimidine and 1,3,5-triazine-based inhibitors in crystals with a variety of anions, including sulfonate and halide as well as carboxylate ions. The preferred motifs for interaction with oxyanions at the 1 and 2 positions (Fig. 16) are the same in both small-molecule and protein-bound situations.

In the small-molecule structures the unprotonated N3 and the 4-amino group were frequently, though by no means universally, involved in base pairing (Fig. 17), an interaction that does not occur when the inhibitor is bound to DHFR. The motif at the 3,4 positions of an inhibitor bound to DHFR [24] is the same as the second most frequent motif in the small-molecule context (Fig. 18) [23]. Once again, the existence of a large volume of structural data is of key importance. It enables the propensity for intermolecular interactions, as well as features of molecular geometry, to be evaluated.

Although the PDB contains an order of magnitude fewer macromolecular structures than there are small-molecule structures in the CSD, comparison of related macromolecular structures can still be very worthwhile. The recent review of DHFR structures [24] reported the existence of 129 structures obtained from 15 different species as of February 2007. Sequence comparison shows high (75–95%) homology among vertebrate

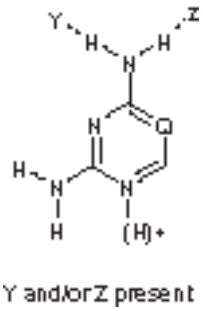


**Figure 18** Schematic drawing of the second most frequent interaction of antifolate drugs with counterions and solvent molecules at the 3 and 4 positions [23]. (Reprinted with permission from Taylor & Francis, http://www.informaworld.com.)

species but only about 30% homology between vertebrate and bacterial species. Although all DHFR enzymes interact with dihydrofolate, their main substrate, in very similar ways, there are sufficient structural differences at and near the active site to allow selective inhibition to be achieved. Ligand-induced conformational flexibility in loop regions near the active site was an unexpected property of DHFR deduced from structural comparison. It must be considered during the drug design process.

### 4.5 Future Prospects

On the experimental side, the chances of obtaining a usable data set from a sample are improving. Many tedious and time-consuming aspects of crystal growing have recently been automated. Robots can simultaneously conduct crystal-growing experiments on small molecules in a wide variety of solvents with a wide variety of counterions and co-crystal formers. The optimum conditions for growing macromolecular crystals can be systematically investigated. Computer vision systems and computer-controlled spectrometers monitor the success of such efforts. Once the development costs have been amortized, such robotic equipment should become available to less well-funded laboratories. Recently developed microfocus X-ray sources direct narrow intense beams onto crystals, alleviating the minimum size requirements. Even when all attempts to grow adequately sized single crystals of a drug fail, it may still be possible to determine its three-dimensional structure from powder diffraction data. At almost every crystallographic meeting innovative applications of powder diffraction and improved methods for determining structures from powder diffraction data are reported.

The innovation now being applied to crystal structure prediction can also be expected to bear fruit. More groups now take into account the charge distribution on the heteroatoms in drug molecules and its influence on hydrogen bonding, not just treating them as uniform spheres. Prior knowledge from the Cambridge Structural Database is being incorporated into some methods. Kinetic factors are being considered as a reason why certain dense, low-energy, predicted crystal forms do not materialize. Before structure prediction techniques become reliable enough for wide use ab initio, they may link up with powder diffraction by providing trial structures as starting points. Even an occasional correct prediction would be helpful, and the diffraction data would be available to verify its validity.

Indeed it is an exciting time to be applying crystallography to the development of therapeutic agents!

## REFERENCES


1. Day GM, et al. (2005). A third blind test of crystal structure prediction. *Acta Cryst B61*, 511–527.
2. Boyle PD. Growing crystals that will make your crystallographer happy. Available: http://www.xray.ncsu.edu/GrowXtal.html. Accessed October 23, 2007.
3. Dinger M, Klosin J. Crystal growing tips. Available: http://xray.chem.ufl.edu/growing%20tips.htm. Accessed October 23, 2007.
4. Newman J (2006). A review of techniques for maximizing diffraction from a protein crystal *in stilla*, *Acta Cryst D62*, 27–31.
5. Johnson CK (1976). *ORTEPII*. Report ORNL-5138. Oak Ridge, TN: Oak Ridge National Laboratory.
6. Sheldrick GM (1997). *SHELXS. Program for the Solution of Crystal Structures*. University of Göttingen, Germany.
7. Sheldrick GM (1997). *SHELXL. Program for Crystal Structure Refinement*. University of Göttingen, Germany.
8. Laing M. An introduction to the scope, potential and applications of X-ray analysis. Available: http://www.iucr.org/iucr-top/comm/cteach/pamphlets.html. Accessed October 23, 2007.
9. Clegg W, Ed. (2001). *Crystal Structure Analysis: Principles and Practice*. Oxford: Oxford University Press.
10. Chan CDN. Definition of the Miller indices of a plane. Available: http://en.wikipedia.org/wiki/Miller_indices. Accessed October 23, 2007.
11. Hodgkin DC (1972). The X-ray analysis of complicated molecules. In *Nobel Lectures, Chemistry, 1963–1970*. Amsterdam: Elsevier.
12. Matthews BW (1968). The solvent content of protein crystals. *J Mol Biol 33*, 491–497.
13. Engh RA, Huber R (1991). Accurate bond and angle parameters for X-ray protein structure refinement. *Acta Cryst A47*, 392–400.
14. Flack HD (1983). On enantiomorph-polarity estimation. *Acta Cryst A39*, 876–881.

15. Gueritte-Voegelein F, Guenard D, Mangatal L, Potier P, Guilhem J, Cesario M, Pascard C (1990). Structure of a synthetic taxol precursor: *N-tert*-butoxycarbonyl-10-deacetyl-*N*-debenzoyltaxol. *Acta Cryst C46*, 781–784.
16. Mastropaolo D, Camerman A, Luo Y, Brayer GD, Camerman N (1995). Crystal and molecular structure of paclitaxel (Taxol). *Proc Natl Acad Sci USA 92*, 6920–6924.
17. Schwalbe CH (2005). From small molecule structures to medicines. *Crystallogr Rev 11*, 49–60.
18. Lam PYS, Jadhav PK, Eyermann CJ, Hodge CN, Ru Y, Bacheler LT, Meek JL, Otto MJ, Rayner MM, Wong YN, Chang C-H, Weber PC, Jackson DA, Sharpe TR, Erickson-Viitanen S (1994). Rational design of potent, bioavailable, nonpeptide cyclic ureas as HIV protease inhibitors. *Science 263*, 380–384.
19. Cowan-Jacob SW, Fendrich G, Floersheimer A, Furet P, Liebetanz J, Rummel G, Rheinberger P, Centeleghe M, Fabbro D, Manley PW (2007). Structural biology contributions to the discovery of drugs to treat chronic myelogenous leukaemia. *Acta Cryst D63*, 80–93.
20. McNae IW, Kan D, Kontopidis G, Patterson A, Taylor P, Worrall L, Walkinshaw MD (2005). Studying protein-ligand interactions using protein crystallography. *Crystallogr Rev 11*, 61–71.
21. Nangia A, Biradha K, Desiraju GR (1996). Correlation of biological activity in β-lactam antibiotics with Woodward and Cohen structural parameters—A Cambridge database study. *J Chem Soc Perkin Trans 2*, 943–952.
22. Nicklaus MC, Wang S, Driscoll JS, Milne GW (1995). Conformational changes of small molecules binding to proteins. *Bioorg Med Chem 3*, 411–428.
23. Schwalbe CH, Cody V (2006). Structural characteristics of small-molecule antifolate compounds. *Crystallogr Rev 12*, 267–300.
24. Cody V, Schwalbe CH (2006). Structural characteristics of antifolate dihydrofolate reductase interactions. *Crystallogr Rev 12*, 301–333.
25. Francis S, Muthiah PT, Bocelli G, Righi L (2002). Hydrogen-bonding patterns in trimethoprim trifluoroacetate. *Acta Cryst E58*, o717–o719.
26. Sethuraman V, Stanley N, Muthiah PT, Sheldrick WS, Winter M, Luger P, Weber M (2003). Isomorphism and crystal engineering: Organic ionic ladders formed by supramolecular motifs in pyrimethamine salts. *Cryst Growth Des 3*, 823–828.
27. Barr G, Dong W, Gilmore CJ, Parkin A, Wilson CC (2005). *dSNAP*: A computer program to cluster and classify Cambridge Structural Database searches. *J Appl Cryst 38*, 833–841.

# PEGYLATION

Sanju Dhawan, Deepak N. Kapoor, Rishi Kapil, Manoj Kumar, and Ankur Bansal
*UGC Centre of Advanced Studies, University Institute of Pharmaceutical Sciences, Panjab University, Chandigarh, India*

## 1 PEGYLATION

### 1.1 History

The pioneering first steps in PEGylation were taken in late 1970 in the laboratory of Frank Davis of Rutgers University [1]. In late 1970s, PEGylation was first reported by Davies and Abuchowski in two key research papers on albumin and catalase modification [2, 3]. This was an important milestone because at that time it was not conceivable to modify an enzyme while maintaining its activity. Since then, the research on the PEGylation process expanded and developed tremendously [4]. The covalent attachment of polyethylene glycol (PEG) to enzymes, proteins, peptides, and nonpeptides, a process known as PEGylation, has received increased attention [5] nowadays.

Abuchowski et al., [2] altered the immunological properties of bovine serum albumin (BSA), which was achieved by the covalent attachment of PEG. With this achievement, the exploration of PEG chemistry started [6].

### 1.2 Definition

Different researchers have defined PEGylation in their own words. PEGylation involves the masking of the surface of proteins and peptides by covalent coupling with soluble PEG [1]. It is a method for optimizing pharmacokinetic and pharmacodynamic properties of therapeutic small drug molecules such as peptides and proteins [7].

Thus, it can be concluded that PEGylation is the modification of the surface properties of proteins and peptide or nonpeptide molecules via covalent attachment using a chemical modifier, PEG, either alone or in combination with other moieties. PEGylation has been exploited to increase the half life ($t_{1/2}$), solubility, stability and decrease of the immunogenicity, and toxicity of enzymes, proteins, peptides, and nonpeptide molecules.

### 1.3 PEGnology

PEGnology, as the name indicates, is the study of the modifications of proteins, peptides, or nonpeptides by associating them with PEG [8]. Frank F. Davis, of Rutgers University, New Jersey, is known as "father of PEGnology." While browsing in medical journals, he came across an article that explained how physicians infused a solution containing a mild surfactant, a block copolymer of PEG and polypropylene glycol into the blood of patients undergoing major vessel surgery in order to prevent the formation of lipid embolisms [8]. This was when he focused on PEG [8] and started his work with methoxy PEG.

## 2 PEGYLATION OF THERAPEUTIC AGENTS

### 2.1 Proteins and Peptides

Many differences exist among proteins/peptides and conformational low-molecular-weight (MW) compounds,

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

which may affect their predicted stability profiles. The distinction between peptides and proteins is somewhat arbitrary. Typically, peptides contain fewer than 20 amino acids and proteins contain 50 or more amino acids, between these two categories are polypeptides, which contain about 20–50 amino acids. Polypeptides consist of regularly repeating amino acid units with distinctive side chains that interact with each other to contribute to the three-dimensional structure of the protein [9]. In terms of MW, peptides have a MW less than 5000 Da and proteins have a MW more than 5000 Da.

***Primary Structure*** The sequence of covalently bonded amino acids in the polypeptide chain is known as the primary structure. The primary structure of proteins is determined genetically by the sequence of nucleotides in deoxyribonucleic acid (DNA). Once a polypeptide chain is formed, there is only one residue with a free amino group and only one residue with a free carboxylic group [10].

***Secondary Structure*** The turns and loops of the polypeptide chain constitute the secondary structure. The α helices and the β sheets are the common conformations of the secondary structure. In the α helix, a single protein chain twists like a coiled spring. Each turn contains 3.6 amino acid residues, with the side chain pointing out from the helix. All NH and CO groups are hydrogen bonded except the first NH and the last CO group, which occur at the end of helices. Therefore, the ends of helices are polar and are likely to be at the surface of protein molecules [10].

In the β-sheet conformation, the protein backbone is nearly fully extended. The β sheets can be either parallel or antiparallel. In the parallel β sheet, two or more peptide chains align their backbone in the same direction. The antiparallel β sheet has adjacent chains having parallel but opposite orientations. A protein can contain about 30% of the residues in α helices and 30% in β sheets and the rest in other conformations such as the random coils or β bends. A fully extended polypeptide chain of about 60 residues may be about 200 long, but the folded globular protein may actually be just 30 in diameter [10].

***Tertiary Structure*** The tertiary structure of protein is the overall packing in space of the various elements of the secondary structure. Protein folds in a highly specific manner, but it is believed that the specific folding is determined by the amino acid sequence or the primary structure of the protein. The overall shape is spherical; polar groups are on the surface, and hydrophobic groups are buried in the interior; the interior is closely packed. Guided by these rules, protein folds into a unique structure [11].

The forces that may stabilize the folded structure of proteins may be covalent or noncovalent. Disulfide bridges provide a covalent linkage that can hold together two chains or two parts of the same chain. The disulfide bonds are very hydrophobic compared to the thiol groups. The noncovalent forces that can stabilize proteins include hydrogen bonding, salt bridges, and hydrophobic interactions between the amino acids. Typically, a protein would have maximum conformational stability at its isoelectric point, but this may shift because of the presence of counterions [12].

***Quaternary Structure*** The quaternary structure of a protein refers to the specific association of separate protein chains (subunits or monomers) to form a well defined structure. The subunits are held together by noncovalent forces such as hydrophobic interactions, hydrogen bonds, or van der Waals interactions. Protein–protein associations usually involve the more hydrophobic portions on the surface of a molecule [13]. Normally, the hydrophobic residues are buried in the interior, but they can also be closer to the exterior. This is especially true when the protein has a quaternary structure because the subunits form part of a larger aggregate with geometry that influences the location of the hydrophobic residues, such as tryptophan in α crystallin.

The use of proteins and peptides as human therapeutics has expanded in recent years due to: (1) the discovery of novel peptides and proteins, (2) a better understanding of the mechanism of action in vivo, (3) improvements in expression or synthesis of proteins and peptides that closely resemble human proteins and peptides, (4) improvements in formulation or molecule-altering technologies that have the ability to deliver polypeptides in vivo with improved pharmacokinetic and pharmacodynamic properties [14].

Currently, there are more than 100 recombinant proteins (RP) in phase I clinical trials or beyond, and several dozen have received Food and Drug Administration (FDA) approval. The available delivery methods for proteins can critically be evaluated by considering ease of manufacturing, impact on conformational stability and biological activity, bioavailability, and toxicity. Therefore, the formulation design, which ensures an efficient and a safe delivery of proteins and peptides in a conformationally stable and biologically active form, is the key to commercial success of current and future biotechnology products [15].

Many biopharmaceuticals have problems that are typical of polypeptide therapeutics, including short circulating half-life, immunogenicity, proteolytic degradation, and low solubility [14]. Several strategies have emerged as ways to improve the pharmacokinetic and pharmacodynamic properties of biopharmaceuticals.

These include (1) manipulation of amino acid sequence to decrease immunogenicity and proteolytic cleavage, (2) fusion or conjugation to immunoglobulins and serum proteins, such as albumin, (3) incorporation into drug delivery vehicles for protection and slow release, and (4) conjugating natural or synthetic polymers [3, 16–18]. Those in the biomedical, biotechnological, and pharmaceutical communities have become quite familiar with the improved pharmacological and biological properties that are associated with the covalent attachment of PEG to therapeutically useful polypeptides. For instance, PEG conjugation shield antigenic epitopes of the polypeptide, thus reducing clearance by the reticuloendothelial system (RES), recognition by the immune system, and reduced degradation by proteolytic enzymes. PEG conjugation also increases the apparent size of the polypeptide, thus reducing renal filtration and altering biodistribution [14]. The main modifications exerted by PEGylation include steric hindrance, changes in conformation, changes in electrostatic binding properties resulting in an increase in polypeptide molecular size, and the shielding of its surface from the approach of proteolytic enzymes, antibodies, or antigen-processing cells [7]. Perhaps the most significant modification to improve circulatory $t_{1/2}$ of proteins is PEGylation [19].

In the early 1990s, Altus Biologics introduced cross-linked enzyme crystals, which were dimers of common enzymes hardwired together by covalent chemical bonds [19]. It is also a feasible technique to alter the membrane disruption behavior of a molecular carrier by manipulating its hydrophobic stabilization [20].

PEGylation significantly improves several classes of protein drugs, such as enzymes, cytokines, and antibodies. Table 1 compiles the most important examples of protein conjugates, exploiting the advantages of PEGylation and PEG derivatives that are useful therapeutically. In general, the improvements are an increased retention time in the body, a reduction of immunogenicity, and increased stability toward metabolic enzymes [4].

Unfortunately, the PEGylation of proteins is often accompanied by loss of biological activity. However, this is compensated for by the prolonged body residence time as a result of the increased stability and higher hydrodynamic volume. A typical example is the PEGylated α interferon, which retains only 7% of the antiviral activity of the native protein but still shows a greatly improved performance in vivo, compared with the unmodified enzyme because of improved pharmacokinetics [21]. A further instructive case comes from the PEGylation of the tumor necrosis factor (TNF) and its lysine-deficient mutant proteins, as reported by several laboratories [22–24]. It was demonstrated that body residence time, PEG mass and shape, and the stability of PEG–protein bonds are all important factors for the antitumor activity, probably because they play a role, in this specific case, in releasing the native protein [25] or in maintaining the activity of the protein inside the tumor mass [4].

Many approaches for protein delivery have been examined, including protein entrapment in insoluble matrices [1] and liposomes and immobilization onto polymer resins for use with extracorporeal shunts through which blood could flow [26]. By far the most successful approach that has been used to mask the protein surface by covalent coupling of soluble PEG is PEGylation [1]. This approach has been applied successfully to develop various marketed therapeutic protein formulations such as Amgen's Neulasta (Pegfilgrastim), a PEGylated version of G-CSF (granulocyte colony-stimulating factor) used to boost white blood cells, and Roche's Pegasys, PEGylated α interferon for treating hepatitis. In both cases, PEGylation has improved the efficacy of the protein and reduced the dosing. Most importantly, PEGylation demonstrated

**TABLE 1 Approved PEG Conjugates [4]**

| PEG Conjugates | Type of PEGylation | Disease |
|---|---|---|
| PEG–asparaginase (Oncaspar) | Random, linear PEG | Acute lymphoblastic leukemia |
| PEG–adenosine deaminase (Adagen) | Random, linear PEG | Severe combined immunodeficiency disease (SCID) |
| PEG–interferon-α2a (Pegasys) | Random, branched PEG 40 kDa | Hepatitis C |
| PEG–interferon-α2b (PEG–Intron) | Random, linear PEG 12 kDa | Hepatitis C and clinical trials for cancer, multiple sclerosis, HIV/AIDS |
| PEG–G–CSF (Pegfilgrastim, Neulasta) | Selective, linear PEG 20 kDa | Treating of neutropenia during chemotherapy |
| PEG–growth hormone receptor antagonist (Pegvisomant, Somavert) | Random, linear PEG 5 kDa (genetic modified protein) | Acromegaly |

that chemical modification can make complex protein drugs more effective [19].

### 2.2 Drugs

Common problems encountered with the use of small molecular weight drugs especially the antitumor agents, are their low solubility, rapid excretion, and untargeted biodistribution [4]. PEGylation can address all these factors. Increased solubility, modification of pharmacokinetics, and targeting for important drugs such as taxol, camptothecin, *cis*-platinum, and doxorubicin had been described only after the introduction of PEGylation as a new chemical/biochemical modification technique [27]. The conjugation chemistry of nonprotein drugs faces fewer problems because of the reduced number of functional groups present on the molecule, absence of conformational constraints, and easier purification and characterization steps for the drug–polymer conjugates [4]. For small-molecule drugs many researchers have exploited PEGylation for conferring a passive targeting to solid tumors, by enhancing permeability and retention effects [28], or simply by improving the drug's pharmacokinetic profile by slowing the drug clearance from the body [4].

A general problem encountered in PEGylation of small–molecular weight drugs is the low loading of the polymer, which possesses only one or two activated hydroxyl terminal groups that can be activated [29]. To overcome this limitation, some research groups have proposed the construction of dendrimeric structures at the level of the polymer termini, obtaining PEG dendrons, called forked PEGs. The implementation of the number of active groups is reached by using branching molecules, such as bicarboxylic amino acid, which allows an increase in the drug–polymer loading ratio [29].

Doxil is a representative commercialized PEGylated liposomal formulation of doxorubicin [30]. In this formulation, a PEGylated liposome, that is, STEALTH liposome, enhances therapeutic effect and decreases serious side effects such as cardiotoxicity of doxorubicin by improving its biodistribution [31].

### 2.3 Gene Delivery through Nonviral Polymeric Vectors

Gene therapy is a technology employing a gene as a therapeutic agent to treat a disease. The potential goals of gene therapy include replacement of an inactive gene, reactivating inactive gene, turning off gene causing disease, as observed in oncogenes, turning on further naturally protective genes, or adding a gene characteristic to cells, such as increased susceptibility of cancer cells to chemotherapeutic drugs [32]. Thus, it is important to develop a technology and methodology of a drug delivery system to enhance the level of protein expression accompanied with gene transfection [33]. The objective of gene therapy is to allow a gene to express the coded protein in the target cells and consequently to treat disease by the protein secreted from transfected cells.

For instance, gelatins have been grafted with PEG (e.g., PEG-cationized gelatin) to function as a nonviral gene carrier [34]. Following intravenous injection, the PEG-introduced gelatin was retained in the blood circulation for a longer period than the original gelatin [34]. It is likely that gelatin is covered with the PEG molecule and protected from the enzymatic attack and exclusion by RES [33]. The presence of PEG molecules for the PEG-cationized gelatin–plasmid DNA complex protects the plasmid DNA from the enzymatic attack because of the "stealth effect" of PEG [35]. The level of gene expression was significantly increased when the PEG-cationized gelatin–plasmid DNA complex was intramuscularly administered and compared with the injection of plasmid DNA solution. Therefore, PEG-cationized gelatin is a promising nonviral gene carrier to enhance gene expression in vivo [33].

A variety of nonviral polymeric gene vectors have received much attention in the past [36–38] for the delivery of genetic materials to the targeted cells in an effective and safe manner. Especially, the polyplexes formed by the electrostatic interaction between plasmid DNA (pDNA) and polycations have been designed to condense pDNA by shielding its negative charges, protect pDNA from rapid nucleolytic degradation, and facilitate its cellular uptake in order to achieve effective gene delivery [39, 40]. When the complex of a plasmid DNA with a nonviral gene carrier is given to cells or injected into the body in the solution form, the plasmid DNA is rapidly and readily degraded and inactivated by enzymes or cells. As one practically possible way to minimize the enzymatic degradation of plasmid DNA, the surface of the plasmid DNA can be modified with PEG or PEG-like polymers [33].

A typical PEGylated polyplex is a core–shell type of micelle (polyplex micelle) formed through the electrostatic interaction between pDNA and PEG-*block*-polycation copolymers. One of the advantages of polyplex systems is the possibility of various structural modifications to improve the stability and transfection efficiency of the polyplexes. Among such modifications, PEGylation of polycations is a promising way to realize systemic gene delivery due to the improved stability of polyplexes in biological media [33]. Kakizawa and Kataoka [41] have reported

delivery systems of genes and related compounds with PEG-based polymer micelles. The research data revealed that the formation of PEG-based polymer micelles protected the plasmid DNA from DNase degradation [33].

Polyplex micelles have also been demonstrated to show high colloidal stability under biological media and substantial transfection activity against various cells even after preincubation with serum proteins [42, 43]. Moreover, polyplex micelles showed prolonged blood circulation and in vivo gene transfer to liver [40, 44]. Polyplex micelles from PEG-*block*-poly(*N*-substituted asparagine) copolymers having the *N*-(2-aminoethyl)-2-aminoethyl group in the side chain, for example, [45] showed efficient and nontoxic transfection to several primary cells including endothelial and smooth muscle cells and, thus, is expected to be a potent nonviral vector for in vivo gene delivery [45].

## 3 POLYETHYLENE GLYCOL

PEGs comprise a series of low-to-medium MW, wholly synthetic water-soluble polymers [46]. Chemically, they can be regarded as polyether diols, made by the stepwise addition of ethylene oxide to water or ethylene glycol starters to form long, linear chains of oxyethylene ($-CH_2CH_2O-$) units having primary hydroxyl groups at each end. Depending upon chain length, they range in physical form at room temperature from transparent to white viscous liquids (MW 200–700 Da), through waxy semisolids (MW 1000–2000 Da), to hard, waxlike solids (MW 300–20,000 Da and above). All PEGs are completely soluble in water, very low in toxicity, quite bland, and nonirritating. In addition, they possess good solvent action, good stability, and good lubricity. This unusual combination of properties has enabled them to find a very wide range of commercial applications in fields such as ointments, suppositories, cosmetic creams, lotions, and dressings; paper coating lubricants; softeners and modifiers; metal working lubricant detergent modifiers; and wood impregnants [46]. PEG is nontoxic, nonimmunogenic, nonantigenic, highly soluble in water, and FDA approved [4].

### 3.1 PEG Moiety Properties

***Chemical Nature of PEG*** All the PEG molecular chains, regardless of length, are in fact diols, with primary hydroxyl groups at each end. Thus, they can enter into all the chemical reactions characteristic of aliphatic hydroxyl groups, for example, esterification, urethane formation, cyanoethylation, and the like. Commercially, the most useful chemical derivatives of the PEGs are the fatty acid mono and diesters, which are valuable as emulsifiers, detergents, softeners, and lubricants [46]. In its most common form, PEG is a linear or branched polyether terminated with hydroxyl groups and having the general following structure [14]:

$$HO-(CH_2CH_2O)_n-CH_2CH_2-OH$$

***Synthesis of PEG [14]*** PEG is synthesized by anionic ring-opening polymerization of ethylene oxide initiated by nucleophilic attack of a hydroxide ion on the epoxide ring. Most useful for polypeptide modification is monomethoxy PEG (mPEG) having the following general structure:

$$CH_3O-(CH_2CH_2O)_n-CH_2CH_2-OH$$

***Synthesis of mPEG [14]*** mPEG is synthesized by anionic ring-opening polymerization initiated with methoxide ions. Commercially available mPEG contains a considerable amount of diol PEG due to the presence of trace amounts of water during polymerization. This diol PEG has relatively high MW due to polymerization at both ends of the polymer. The amount of diol PEG can exceed 15% of the composition of mPEG.

***Physical Nature of PEG*** A number of grades of PEGs are commercially available. They are listed in Tables 2 and Table 3. When blends of different grades of PEG are being considered, it should be realized that the

**TABLE 2 A Properties of Liquid PEGs**

| Grade Number | 200 | 300 | 400 | 500 |
|---|---|---|---|---|
| Avg. molecular weight range | 190–210 | 285–315 | 380–420 | 570–630 |
| Specific gravity, 20/20°C | 1.127 | 1.127 | 1.128 | 1.128 |
| Freezing range, °C | Supercools | –15 to –8 | 4–8 | 20–25 |
| Viscosity at 25°C (77°F), cSt | 40 | 69 | 90 | 131 |
| Viscosity at 98.9°C (210°F), cSt | 4.3 | 5.8 | 7.3 | 10.5 |
| Water solubility | Complete | Complete | Complete | Complete |

**TABLE 3 Properties of Solid PEGs**

| Grade Number | 1000 | 1500 | 4000 | 6000 | 20,000 |
|---|---|---|---|---|---|
| Avg. molecular weight range | 950–1050 | 1300–1600 | 3000–3900 | 6000–8500 | 18,500 |
| Density, 25°C, g/cm$^3$ | 1.17 | 1.210 | 1.212 | 1.212 | 1.215 |
| Freezing range, °C | 37–41 | 43–47 | 53–60 | 57–63 | 56–64 |
| Viscosity at 98.9°C (210°F), cSt | 17–19 | 25–32 | 75–110 | 580–900 | 6,900 |
| Water solubility (g/100 mL) | ~74 | ~70 | ~62 | ~53 | ~50 |

average molecular weight of the blends must be calculated on the basis of the number of moles of each grade, not the actual proportion by weight. For example, an equimolar blend of grades PEG 1000 and PEG 4000 would have an average MW of 2500 Da, and would correspond to a 20:80, PEG 1000:PEG 4000 mixture on a weight basis. The 50:50 weight basis mixtures of these two would correspond to a 4:1 mixture on a molar basis and would have an average MW of 1600 Da [46].

*Solubility in Water* All grades of PEGs are soluble in water. Aqueous solutions of higher MW grades may form gels [47].

*Solubility in Organic Solvents* PEGs are soluble in most of the organic solvents with higher polarity, such as alcohols, ether alcohols, glycols, esters, ketones, aromatic hydrocarbons, nitroparaffins, and the like. They are neither soluble in the aliphatic or cycloaliphatic hydrocarbons nor in other solvents of low polarity [46]. In addition, the PEGs are also soluble in a number of liquids not usually regarded as solvents, such as the lower aldehydes, amines, organic acids and anhydrides, and polymerizable monomers. They are not soluble in vegetable or mineral oils or similar liquid compounds. The individual PEGs differ considerably in their solubilities, with the lower-molecular-weight grades being generally much more soluble than the higher grades. Often, raising the temperature high enough to melt the solid PEGs greatly improves their solubility [46].

*Hygroscopicity* The ability of the lower PEGs to take up and retain moisture from the atmosphere, coupled with their good permanence and good plasticizing action for other substances, makes them industrially useful as humectants for a variety of hydrophilic products [46].

***Toxicological Properties*** PEG shows little toxicity and is eliminated from the body intact either by the kidneys (for PEGs < 30 kDa) or as digestive wastes (for PEGs > 20 kDa). PEG is rapidly cleared in vivo without structural change and clearance is dependent on molecular weight [14]. PEG lacks immunogenicity, and antibodies to PEG are generated in rabbits only if PEG is combined with highly immunogenic proteins. No reports were found on the generation of antibodies to PEG under routine clinical administration of PEGylated proteins. PEG is only weakly immunogenic even at high molecular weights [14]. However, under extreme experimental conditions, antibodies to PEG have been generated in animals as a result of the injection of PEG–protein conjugates [48]. Antibodies to PEG have also been generated when attached to a highly immunogenic molecule under an immunization protocol with Freund's adjuvant [14].

Under the Food Additives Amendment to the Federal Food, Drugs and Cosmetics Act, the *Food Chemicals Codex* grades of the PEGs are cleared for various direct and indirect food additive uses, provided that the amounts used do not exceed those required to produce the intended effect. The FDA has approved PEG for use as a vehicle or base in foods, cosmetics, and pharmaceuticals, including injectable, topical, rectal, and nasal formulations [48]. Under EPA (Environmental Protection Agency) regulations, PEGs are exempted from any requirement of residue tolerance when used as inert ingredients in pesticide formulations that are applied to growing crops or harvested agricultural products [46].

### 3.2 Polyethylene Glycol versus Polyethylene Oxide

PEGs are viscous liquids or waxy solids with a MW less than ~25,000 Da [49]. The PEGs should be distinguished from high and very high MW polymers of ethylene oxide, which are also water soluble [50]. These high MW polymers of ethylene oxide are commonly known as poly ethylene oxide (PEO) [51]. PEO resins are made commercially by the catalytic polymerization of ethylene oxide in the presence of metallic catalyst systems. They are available with average molecular weights from 200 to $7 \times 10^6$ [49] and form strong, thermoplastic films [50]. They are represented by the formula $(OCH_2CH_2)_n$

in which *n* represents the average number of oxyethylene groups (~2000 to more than 100,000).

### 3.3 PEG Reagents for Protein Modification

Protein modification reagents that incorporate repeating ethylene oxide/ethylene glycol subunits onto the surface of a protein are often referred to as PEGylation reagents. Some important modification reagents are discussed next.

***Alkylating PEG*** This is a convenient way for conjugation when the amino positive charge is critical for the retention of biological activity. Alkylating PEG derivatives maintain the charge of the native protein in the final conjugate [4]. These products do not modify the charge of amino residues [52]. One of the example, that is, PEG-aldehyde, gives a permanent linkage after Schiff base formation followed by cyanoborohydride reduction (Fig. 1A). However, the reaction rate for the Schiff base formation is relatively low (approx. 24 h), thus leading to inactivation of labile molecules. The reaction pH is critical for selectivity; α-amino terminal modification can be achieved at around pH 5 [53]. PEG-tresyl chloride activation is an alternative method to maintain positive charge [54] (Fig. 1B). However, the chemistry of conjugation and the conjugation products are not unique and well defined [55]. The use of epoxy PEG has been used, but, again, as with PEG-aldehyde, the reactivity is low and, furthermore, the specificity is not certain since hydroxyl groups may also react (Fig. 1C). The popular alkylating PEG derivatives have been enlisted in Table 4.

***Acylating PEGs*** These are the PEG derivatives that, after amino coupling, lead to a loss of positive charge in the final conjugate with respect to starting protein [4].

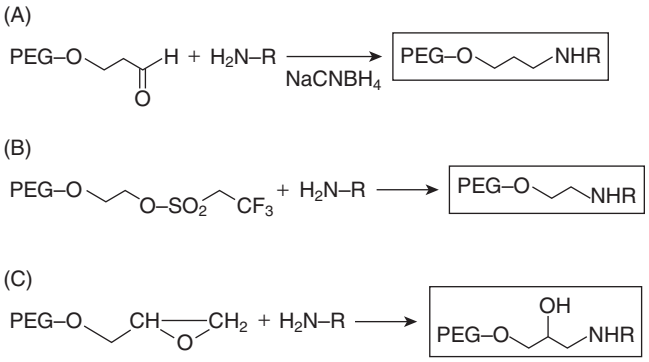


**Figure 1** Alkylating PEGs: (A) PEG aldehyde: Best reaction takes place with α- and ε-amino groups in protein at neutral or mild alkaline pHs, while good selectivity for the α-amino terminal aminoacid at pH 5–6 was found. (B) Tresylated-PEG: Best reaction takes place with amines around pH 8 and low temperature. The positive amino charge is maintained. The reaction rate is dependent upon the properties of the peptide, but doubts remain on the structure of the reaction products. (C) PEG epoxide: Reaction at high pH 8.5–9.5, is preferred for amino conjugation, although binding to hydroxyl, imidazole and thiol groups also may take place [52].

**TABLE 4 Alkylating PEGs [4]**

| Structure | Alkylating PEGs | Properties |
|---|---|---|
| 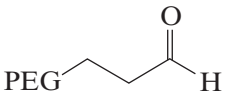 | PEG–aldehyde | A two-step reaction; the first product (a Shiff base) is reduced by $NaCNBH_3$. When the coupling reaction is carried out at low pH, i.e., 4.5–5, it labels only the α-amino group. |
| $PEG–O–SO_2–CH_2CF_3$ | PEG–tresyl or tosyl | Not much used because the chemistry leads to a mixture of products. |
| 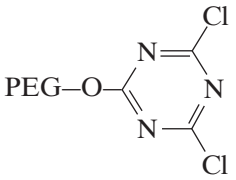 | PEG–dichlorotriazine or cholortriazine | Now they are abandoned for therapeutic application because of their toxicity. |
| 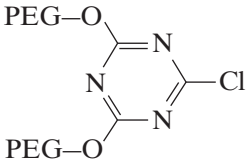 | | |
| | PEG–epoxide | Slowly reactive, rarely used. |

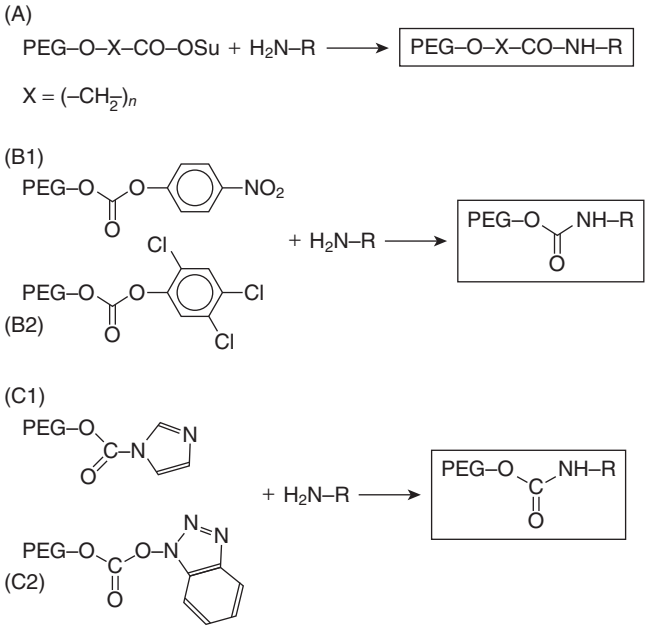


**Figure 2** Acylating PEGs: (A) Activated alkyl acids: These compounds are highly reactive toward amino groups, but the reaction rate may change significantly from the X structure. (B) PEG-*p*-nitrophenylcarbonate (B1) and PEG–trichlorophenylcarbonate (B2): The reaction rate is much slower than the OSu-activated carboxylate–PEG. Carbamate derivatives are obtained. (C) PEG–oxycarbonylimidazole (C1) and PEG–benzotriazole carbonate (C2): Both reagents yield carbamate linkages with amino groups, but with different reaction rates [52].

Most of these are hydroxysuccinimidyl esters (–OSu) of carboxylated PEGs (Fig. 2A) [52]. The distance between the active ester (–COOSu) and the last PEG ether can vary in different available products by up to four methylene units, and this has profound influence on the reaction toward amino groups as well as toward water. As an example, the $t_{1/2}$ hydrolysis rate of PEG–O–$CH_2$–$CH_2$–$CH_2$–COOSu is 23 h, while that of PEG–O–$CH_2$–COOSu is 0.75 h [52].

In another acylating mechanism, the terminal PEG hydroxy group is activated by chloroformates or carbonylimidazole. PEG-p-nitrophenylcarbonate or PEG-trichlorophenylcarbonates exhibit lower reaction rates than the –OSu-activated PEGs, and this property allows exploitation of different reactivity among the amino groups of the macromolecule by stopping the reaction at the desired time of modification. In this way some selectivity in the conjugation based on amino group availability can be achieved. Furthermore, reactions in aqueous-organic solvents are also feasible. There is an interesting difference between these two carbonate-activated PEGs: The first one has a pale yellow color, whose intensity increases with the conjugation reaction for the formation of phenol as a leaving group, while the second is a colorless product [52].

A urethane linkage with amino groups is also obtained with other electrophilically activated PEGs. Among these are PEG-oxycarbonylimidazole [56] and PEG-benzotriazole carbonate (Fig. 2C). The first one is characterized by a lower reaction rate as compared to choloroformates, while the second one is very reactive, although less than the succinimidyl-activated carboxyl PEGs [52]. Some acylating PEG derivatives are listed in Table 5.

To avoid the mixtures of PEGylated products, the evolution of PEGylation chemistry has also allowed site-specific amino modification, thus helping the purification and characterization procedures. It is also feasible to protect the active site of enzymes or the recognition area of proteins by carrying out the PEGylation in the presence of an inhibitor, a substrate, or a specific ligand, with the aim to cover the reactive groups close to sensitive areas. These ligands might be free in the reaction solution or linked to an insoluble resin. In the latter case, not only the binding site but also its surroundings are protected from conjugation [4].

***PEG with Amino Acid Arm*** The presence of an amino acid or peptide arm between PEG and the attached macromolecule gives several advantages due to the variability of properties that may be introduced using a suitable amino acid or peptide. For the convenience of synthesis, the amino acid or peptide can be linked through its amino group by urethane linkage to PEG, while the carboxyl group, properly activated, is bound to the amine of the macromolecule [52] (Fig. 3). Among these arms, norleucine is of interest for the advantage in analysis of PEGylated proteins. A second modification is the so-called branched PEG or PEG2. This PEG derivative is characterized by two linear PEG chains linked together through two functions of a trifunctional spacer, while the third function is used to bind the protein. Lysine is the trifunctional amino acid spacer and the two PEG chains are linked to its α and ε amino groups while the carboxylic group is activated as –OSu for protein binding (Fig. 4). This PEG has the advantage of a lower inactivation of the enzymes during conjugation. Furthermore, due to its "umbrella-like" structure it is more elective in protecting proteins from proteolysis, in the approach of antibodies, and in reducing immunogenicity [52] (Fig. 4).

***Branched-Chain PEG*** In addition to the linear structure of the PEG molecule shown in Figure 5, branched structures have been proved useful for protein and peptide modification. The first branched PEG structure, 2,4-bis(methoxyPEG)-6-chloro-*s*-triazine (mPEG-2

**TABLE 5 Acylating PEGs [4]**

| Structure[a] | PEG–Carboxilates | Properties |
|---|---|---|
| PEG–O–(–$CH_2$–)$_n$–C(=O)–OSu | Several PEG derivatives with one or more $CH_2$ groups between the PEG and the carboxylic group | The carboxylic group is activated as *N*-hydroxy succinimidyl ester, imidazole or benzotriazole. The kinetic rate of conjugation depends on the numbers, and eventual ramification of $CH_2$ groups linked to the carboxyl group. |
| PEG–O–C(=O)–$CH_2CH_2$–C(=O)–OSu | PEG–succinimidyl succinate | The ester bond between succininc acid and PEG is easily hydrolyzed |
| PEG–X–C(=O)–OSu | PEG–amino acid-succinimidyl ester | Norleucine (Nle) or β-Alanine (βAla) as amino acid moiety allows an easy quantification of the number of linked PEG chains by amino acid analysis. |
| PEG–(–X–)$_n$–C(=O)–OSu | PEG–peptide-succinimidyl ester | The Met-Nle or Met-βAla allows the removal of PEG by CNBr treatment for an easy localization of PEGylation site. Lysosomal cleavable sequences, as H-Gly-Phe-Lue-Gly-OH, allow the release of the bound drug inside the cell. |

| Structure[a] | PEG–Carbonates | Properties |
|---|---|---|
| PEG–O–C(=O)–O–C$_6H_4$–$NO_2$ | PEG–*p*-nitrophenyl carbonate | Slowly reactive, yield a urethane linkage with amine. |
| PEG–O–C(=O)–O–N (benzotriazolyl) | PEG–benzotriazolyl carbonate | Slowly reactive, yield a urethane linkage with amine. |
| PEG–O–C(=O)–O–C$_6H_2$(Cl)$_3$ | PEG–2,3,5-trichlorophenyl carbonate | Slowly reactive, yield a urethane linkage with amine. |
| PEG–O–C(=O)–OSu | PEG–succinimidyl carbonate | Slowly reactive, yield a urethane linkage with amine. |

[a]Su, succinimide; X, amino acid linker.

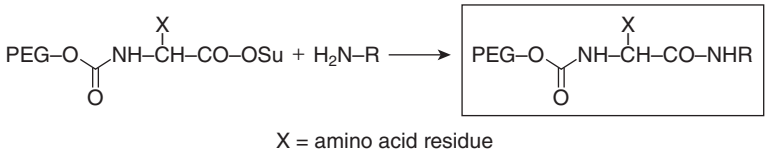


**Figure 3** PEG amino acids: A variety of properties may be conveyed to the PEG conjugate using different amino acids or peptides. Examples may be: Nle for analytical purposes, $^{14}C$ or tritium-labeled Gly for pharmacokinetic studies, Trp for fluorescence labeling, Lys for branching, Met-Nle or Met-bAla for PEG removal by BrCN treatment [52].

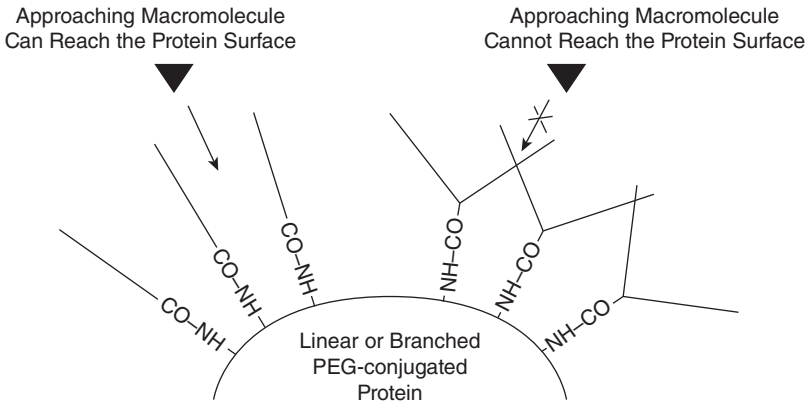


**Figure 4** The "umbrella-like" structure of branched PEG explains the higher capacity in rejecting approaching molecules or cells as compared to linear PEG of the same size [52].

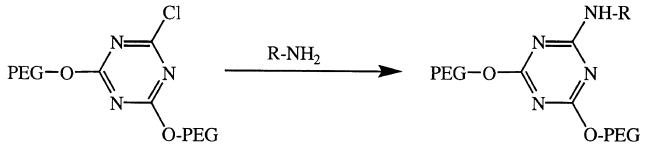


**Figure 5** Branched PEG (PEG2) based on PEG-triazine [14].

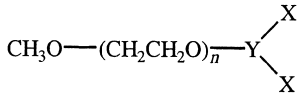


**Figure 7** Linear forked PEG (Y is a group having a carbon branching moiety and X is a molecule, linker, and/or a functional group) [14].

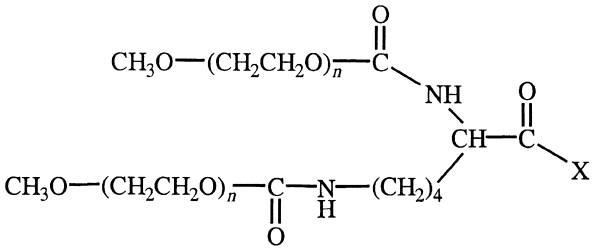


**Figure 6** Branched PEG (PEG2) (X is a molecule, linker, and/or a functional group) [14].

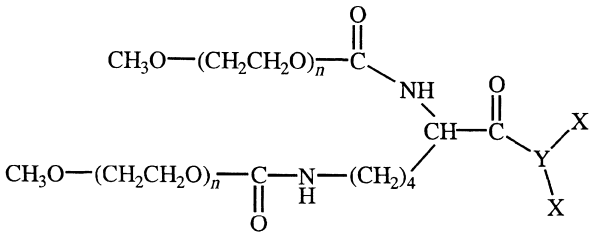


**Figure 8** Branched forked PEG (Y is a group having a carbon branching moiety and X is a molecule, linker, and/or a functional group) [14].

chlorotriazine), was based on a triazine core and synthesized by Inada et al., as shown in (Fig. 5) [14].

Yamasaki et al., first synthesized a more useful branched PEG structure, based on a lysine core. A highly purified branched PEG or PEG2 (Fig. 6) was constructed by Veronese et al., using two linear PEG–BTC (PEG–benzotriazole carbonate) chains linked to the α- and ε-amino groups of lysine. This construct allowed a large molecular weight (upward of 60 kDa) and highly pure PEG to be synthesized with a single reactive end group [14].

PEG2 turns out to be a more effective protein PEGylation reagent because of its unique characteristics when compared to linear PEGs. For example, PEG2 attached to proteins "acts" to be much larger than a corresponding linear mPEG of the same MW. This structure of PEG also has the advantage of adding two PEG chains per attachment site on the protein, therefore reducing the chance of protein inactivation [14].

Another branched PEG is the forked PEG. Instead of having a single functional group at the end of two PEG chains, as with PEG2, forked PEG has two reactive groups at one end of a single PEG chain (Fig. 7) or branched PEG (Fig. 8).

*[Methyl PEG$_{12}$]$_3$–PEG$_4$–NHS Ester* The amine-reactive, methyl-capped PEO-containing modification reagent is characterized by its triply branched, $(PEO)_{12}$-containing chains built off a $(PEO)_4$ stem that terminates in an *N*-hydroxyl succinimide (NHS) ester [57].

A branched (methyl-PEO$_{12}$)$_3$–PEO$_4$–NHS can efficiently cover the surface of a protein. Each branched

reagent contributes three, long polyoxyethylene glycol chains per protein modification site. These chains impart all the benefits known to be associated with proteins modified by PEG-containing reagents, including improved stabilization, with potentially fewer sites requiring modification.

(Methyl-$PEO_{12}$)$_3$–$PEO_4$–NHS ester is unique because the branched spacers are discrete with respect to the number of PEG groups indicated. The reagent is synthesized as a homogeneous compound of defined molecular weight, characterized by discrete chain lengths, providing a greater ability to optimize and characterize surface protein modifications. Each reagent contains a single NHS ester group that reacts with primary amines (e.g., ε-amino groups of lysine and the amino terminus of proteins) at pH 7.0–9.0 to form covalent amide bonds. Typically, PEGylation reagents are a heterogeneous mixture composed of a distribution of chain lengths in proximity to a specific MW or number of PEG subunits [57].

*[Methyl-$PEG_{12}$]$_3$–$PEG_4$–maleimide* This sulfhydryl-reactive, methyl-capped PEO-containing modification reagent is similar to that of amine-reactive reagent except the $PEO_4$ stem that terminates in a sulfhydryl-reactive maleimide group [57].

Since SH groups, occurring naturally or as reduced disulfides, appear at a substantially lower frequency than ε-amino groups, for example, in protein sequences, a branched (methyl-$PEO_{12}$)$_3$–$PEO_4$–maleimide increases the effectiveness of the PEG modification. At each –SH locus modified with this reagent, three, long polyoxyethylene glycol chains are covalently inserted. These branched chains can impart the benefits associated with proteins PEGylated with single-chain maleimide reagents, including improved stabilization, with fewer sites requiring modification.

(Methyl-$PEO_{12}$)$_3$–$PEO_4$–maleimide contains a single maleimide group that reacts with available sulfhydryl (–SH) groups on proteins and peptides at pH 6.5–7.5 forming a covalent thioether bond. The maleimide group is highly specific to free –SH group being 1000× more reactive at pH 7, as compared to primary amines. The properties of [methyl-$PEG_{12}$]$_3$–$PEG_4$–NHS ester and [methyl-$PEG_{12}$]$_3$–$PEG_4$–maleimide are listed in Table 6.

### *Linear PEG*

*Methyl-$PEG_n$–NHS Ester (n = 4, 8, 12)* These amine-reactive, methyl-capped PEO-containing modification reagents are provided as a series of defined chain-length reagents containing 4, 8, and 12 ethylene glycol subunits and an NHS group [57].

*Methyl-$PEG_{12}$–Maleimide* Methyl-$PEO_{12}$–maleimide is a methyl-capped PEO-containing sulfhydryl (SH)-reactive protein modification reagent. This PEGylation reagent has a defined chain length of 12 ethylene glycol/ethylene oxide subunits in the chain, terminating on one

**TABLE 6 Properties of Branched PEGylation Reagents**

| | Amine-Reactive (Methyl-$PEO_{12}$)$_3$-$PEO_4$-NHS Ester[a] | Sulfhydryl-Reactive (Methyl-$PEO_{12}$)$_3$-$PEO_4$-Maleimide |
|---|---|---|
| Alternative names | (m-$PEO_{12}$)$_3$-$PEO_4$-NHS<br>NHS-d$PEG_4$-(m-d$PEG_{12}$)$_3$ester<br>(methyl-$PEG_{12}$)$_3$-$PEG_4$-NHS | (m-$PEO_{12}$)$_3$-$PEO_4$-Mal<br>Mal-d$PEG_4$-(m-d$PEG_{12}$)$_3$, |
| Molecular formula | $C_{108}H_{206}O_{51}N_6$ | $C_{106}H_{202}O_{50}N_6$ |
| Molecular weight | 2420.80 | 2360.75 |
| Spacer arm length | 77.5 Å | 79.6 Å |
| Physical form | Waxy solid | Waxy solid |
| Solvent compatability | • Dry DMF<br>• DMSO<br>• Methylene chloride<br>• Water | • Dry DMF<br>• DMSO<br>• Methylene chloride<br>• Water |
| Storage conditions | –20°C, desiccated | 4°C, desiccated |
| Reactive group | NHS ester, reacts with primary amines at pH 7.0–9.0 | Maleimide; reacts with sulfhydryl (-SH) groups with high specificity between pH 6.5 and 7.5 to form a stable thioether bond |
| Spacer composition | Branched portion: 3 × methyl-capped $(PEO)_{12}$; 12 ethylene oxide subunits/branch<br>Stem portion: $(PEO)_4$; 4 ethylene oxide units | Branched portion: 3 × methyl-capped $(PEO)_{12}$; 12 ethylene oxide subunits/branch.<br>Stem portion: $PEO_4$; 4 ethylene oxide units |

[a]DHF: dimethylformamide; DMSO: dimethyl/sulfoxide.

end with a methyl group and on the other end with a maleimide-reactive group [57].

The maleimide group reacts with available –SH groups from naturally occurring or mutagenically derived cysteine residues. Sulfhydryl groups can be incorporated artificially by way of Traut's Reagent (2-imionthiolane), which converts amino groups to sulfhydryl groups. The properties of linear PEGylation reagents are listed in Table 7.

***Sources of PEG*** Union Carbide introduced commercial grades of PEGs to the United States market in 1939. Some of the manufacturers are the Dow Chemical Company, Jefferson Chemical Company, BASF Wyandotte Corporation, Huels, Imperial Chemical Industries, Kuhlmann, Montedison, and Sanyo. Others are NOF, Nektar Inc., Schering-Plough Corp., and Hoffman-La Roche Inc.

Fluka Chemie AG (Germany) is also a source of PEGs of different molecular weights [52]. Polymer Laboratories Ltd. (England) supplies ultrapure PEG or methoxy PEG of low polydispersity to use as standards and small quantities of PEG or functionalized PEG for R&D purposes. Shearwater Polymers Inc. (Huntsville, Al) is a company specialing in PEG that supplies a great variety of products activated toward any amino acid residues as well as heterofunctional PEGs and PEG conjugates [52].

## 4 ACTIVE SITE PROTECTION IN PEGYLATION

Rejection of a substrate from the protein active site attenuates the efficacy of PEGylated molecule. Thus, it becomes important to prevent PEGylation at the active site of the peptide or protein molecule. This can be done by various methods reported by Veronese et al., [52], namely, by using an active site protecting agent during conjugation, which could be an insoluble-resin-linked, pH- or ionic strength-sensitive inhibitor with reversible affinity toward enzymes; a pH-sensitive polypeptide inhibitor or a branched PEG protecting the active site by structural steric hindrance. Modifying the stoichiometric deficiency of PEG by selective isolation of high biologically active PEG isomer or chemical reversible site-directed peptide protection may also solve the purpose.

## 5 IDENTIFICATION OF PEGYLATION SITE

The identification of the PEGylation site in proteins is an essential requirement in order to characterize PEGylated protein and retain their biological activity to the maximum [52].

### 5.1 PEGylation Site Identification in Peptide Conjugates by Edman Degradation

Site identification by Edman degradation is possible only after cleavage of large proteins to smaller peptides [58]. Smaller peptides can be more easily identified than the larger conjugates. However, this method is not very effective for proteins due to prevention of specific cleavage by hindrance of bound polymer chains or difficulty in isolation and purification of smaller chains. However, Monkarsh et al., [59] characterized no-PEGylated interferon-2α by Edman degradation.

**TABLE 7 Properties of Linear PEGylation Reagents**

| | Methyl-$PEO_4$-NHS Ester | Methyl-$PEO_{12}$-Maleimide |
|---|---|---|
| Alternative names | m-$PEO_4$-NHS<br>NHS-m-dPEG<br>Methyl-$PEG_4$-NHS<br>NHS-m-$dPEG_4$ | m-$PEO_{12}$-maleimide, methyl-$dPEG_{12}$-Mal and m-$dPEG_{12}$-Mal |
| Molecular formula | $C_{14}H_{23}O_8N$ | $C_{32}H_{58}N_2O_{15}$ |
| Molecular weight | 333.33 | 710.81 |
| Spacer arm length | 16.4 Å | 51.9 Å |
| Physical form | Liquid | Off-white solid |
| Solvent compatability | • DMSO<br>• Methylene chloride<br>• Water | DMSO |
| Storage conditions | –20°C, desiccated | –20°C, desiccated |
| Reactive group | NHS ester, reacts with primary amines at pH 7.0–9.0 | Maleimide groups react with -SH groups at pH 6.5–7.5 to form stable thioether bonds |
| Spacer composition | PEO/4 ethylene oxide subunits | PEOS/12 ethylene oxide subunits |

### 5.2 PEGylation Site Identification in Protein Conjugates

A method reported by Sartore et al., [58, 60] using PEGylated lysozyme and PEGylated insulin with methionine in the peptide arm overcomes the difficulty of site identification in larger protein conjugation. In this method, methionine (Met-X) is present in the peptide arm with PEG for protein modification where X is an unnatural amino acid, norleucine (Nle), or β-alanine (β-Ala). Its presence is useful to remove bulky PEG polymer chains from the protein on treatment with cyanogens bromide (BrCN). These de-PEGylated proteins may undergo any chemical or enzymatic cleavage at this point for sequence evaluation. They are also separated easily into positional isomers due to lower hindrance by PEG chains. The cleaved and short peptides are easily identified on the based on linked unnatural amino acids, Nle, and β-Ala as reporter groups which can be detected by amino acid analysis after acid hydrolysis, electrospray mass spectrometry, MALDI-TOF (matrix-assisted laser desorption ionization – time of flight) mass spectrometry or Edman degradation.

### 5.3 One-Step Method for PEGylation Site Identification in Protein Conjugates

A one-step process for identification of PEGylation sites in lysozyme by using a heterobifunctional PEG derivative, biotin–PEG–NHS, was reported by Lee et al., [61]. Lysozyme (MW of 14.3 kDa) and biotin–PEG–$CO_2$–NHS (MW of 3400 Da) were reacted in the presence of 0.1 *M* Tris HCl, pH 8.0, to obtain final lysozyme-to-PEG molar ratio of 5.9:1. PEGylated lysozyme conjugate so formed had a biotin molecule at each PEG chain end and was digested by trypsin to obtain biotin–PEG-conjugated peptide fragments. These fragments were purified and separated by avidin affinity chromatography and reverse-phase high-performance liquid chromatography (RP-HPLC). The eluted fractions of the peptide were analyzed for mass and identification of PEGylation sites by MALDI-TOF mass spectrometry. Extent of PEGylation at each site was determined by RP-HPLC. This method may be applied to other therapeutic proteins also, but without closely located PEGylation site as they might remain unidentified. Furthermore, only biotinylated PEG derivatives can be used for identification of PEGylation sites. The method is time saving and devoid of any tedious procedures.

### 5.4 PEGylation Site Identification Method by Peptide Mapping

Peptide mapping is a method based on digestion of large conjugates with enzymes to smaller peptides and their analysis using RP-HPLC, based on their relative hydrophobicities. The chromatogram patterns of digested PEGylated and mature proteins are then compared to identify the PEGylated site. Capillary electrophoresis can also be used as a complementary separation technique to determine the site of PEGylation and its relative reactivity. Sodium dodecyl sulfate–capillary gel electrophoresis (SDS–CGE) method using a hydrophilic replaceable polymer network matrix can separate linear 12-kDa PEG-conjugated interferons (IFNs) with high resolution [62].

### 5.5 PEGylation Site Identification Method by Tag Labeling

Tag labeling method for protein identification involves a hydrolyzable succinic ester linker that cleaves the protein molecule and PEG chain ends. The labeled succinic acid group in protein can be determined by MALDI-TOF mass spectrometry. Kopecky et al., [63] carried out a study to identify suitable sites on peptide 238/S1 for chemical reaction or modification without affecting its bioactivity. Peptide 238/S1 contains the sequence NH (2)-PYWKWQYKYD-COOH and is capable of blocking any inhibitory antibodies produced against blood clotting factor VIII (FVIII). The highly reactive peptide was identified by mutational analysis. Coupling PEG with N terminus of peptide led to loss of its neutralization property, thus identifying its importance as a site for bioactivity or chemical reaction.

## 6 SITE-SPECIFIC PEGYLATION

During protein modification, a reactive functional group is introduced to link the polymer to the ε-amino group of lysine residues. The process generally aims to prepare an activated PEG with a reactive functional group, which can be conjugated, to the lysine groups on the protein. The number and location of lysine groups vary greatly in native protein, which leads to difficulties in conjugation chemistry of proteins. The stoichiometry of the conjugates and the attachment sites of the polymer occur randomly, which might lead to unexpected reactions in proteins and compromise with protein stability and function. Various approaches to conjugate the polymer with protein at a specific site have been reported, which prevent multiple PEGylation and

maintain the biological activity of the proteins. Each method has its own advantages and disadvantages, and each of them attempt to overcome the deficiencies of the other, which has resulted in an array of different methods for site-specific PEGylation of peptides and proteins.

Gaertner et al., [64, 65] presented a method in which a reactive carbonyl group was formed, instead of a terminal amine group. In case a protein has N-terminal serine or threonine, very mild periodate oxidation is required, which generates a glyoxylyl group. But, another method which is less restricted by the nature of N-terminal residue and requires harsh metal-catalyzed transamination, generates a keto group. At the N terminal of the polypeptide, the aminoxy-functionalized PEG reacts specifically with the generated aldehyde group to form a stable oxime bond. This property has been used for site-specific labeling of interleukin-8 (IL-8). In another study, Gaertner et al., [66] applied this approach to form site-specific attachment of several functionalized polymers varying in size and shape by an oxime linkage to three proteins: IL-8, G-CSF, and IL-1ra. In all cases the reactive carbonyl group reacts with the aminoxy group of functionalized PEG. Various conjugates of the proteins, with different size and shape of linear or branched PEG polymer, were characterized for bioactivity and pharmacokinetic behavior in rats. This approach protected the bioactivity of the proteins as only a single chain of PEG is placed at a specific site on the protein.

However, studies carried out by Balan et al., [67] on site-specific PEGylation using two disulfide bonds and a three-carbon bridge of proteins revealed an interesting fact that when a three-carbon bridge was used to attach PEG, biological activity of the PEGylated protein was found to be independent of the length of the PEG. It was, therefore, suggested that peptides, proteins, enzymes, and antibody fragments can be site specifically PEGylated across a native disulfide bond using three-carbon bridges without destroying their tertiary structure or affecting their biological activity. The stoichiometric efficiency of this approach also enabled recycling of an unreacted protein.

Youn et al., [68] developed and optimized a one-pot, two-step site-specific PEGylation method. This method employs water-soluble organic phosphine where a polypeptide disulfide bridge is first reduced followed by subsequent reaction of two thiols *in situ*. A sCT derivative FMOC-sCT (9-fluorenylmethoxycarbonyl salmon calcitonin) was used for this study, which was FMOC protected at $Cys^{1}$- and $Lys^{11}$-amines among three PEGylation sites including $Lys^{18}$-amine. Site-specific and optimized reactions resulted in the high production yield (86% approx.) of mono-PEGylated salmon calcitonin ($Lys^{18}$-PEG-sCT), compared with that from conventional non-specific PEGylation (18% approx.). The prepared $Lys^{18}$-PEG-sCT conjugate showed improved biological stability and required biological activity. Examples of a few PEG reagents used for site-specific PEGylation are [21] PEG-vinyl sulfone (via free cysteine), PEG–iodoacetamide (via free cysteine), PEG–maleimide (via free cysteine), PEG–orthopyridyl disulfide (via free cysteine), PEG–Hydrazide (via oligosaccharide), and PEG–isocyanate (via alcohol or amino group).

### 6.1 Thiol Selective PEGylation

Thiol being one of the most useful functional groups in bioconjugate chemistry, thiol-selective PEGylation of an introduced cysteine residue by using site-directed mutagenesis, and either maleimide-based or haloacetic-based PEG derivatives, is considered as the most popular method among different site-specific PEGylation methods [14, 69]. The specificity for a suitable site on the protein molecule and the rate of reaction are comparatively better than other protein modifications. However, its use is restricted due to certain limitations [70].

These limitations include the following:

1. As compared to the native protein, the modified or unmodified form of a genetically engineered mutant may have a different structure and function.
2. The yield of properly folded form would reduce if a cysteine mutant forms disulfide isomers with cysteine residues.
3. Although, *N*-maleimide derivative is sulfhydryl group specific, it may react at a very slow rate with amino and imidazoyl groups at pH 7–8.

Zalipsky et al. [71] were the first to report a detailed method to prepare a hetero functional PEG with a thiol group on one terminus (PEG–SH), yielding 80–95% transformation of the functional groups. The synthesis involved glycine amination of [ethoxy-(thiocarbony1)-thiol]-PEG to yield PEG-SH (Fig. 9) [72]. Continuous oxygen removal and anaerobic condition maintenance ensured maximum yield of the thiol group. However, the reaction formed disulfide, which was sometimes the only product formed. Roberts et al., [14] and Veronese [52] have also suggested different thiol-reactive PEGs and reactions to form mPEG–thiol conjugates, which are shown in Figure 10.

### 6.2 PEGylation by Transglutaminase

To overcome the limitations of the earlier site-specific protein PEGylation methods, Sato [69] developed a

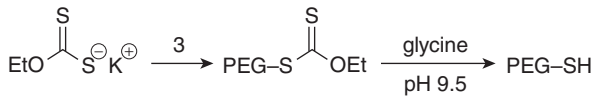


**Figure 9** Thiol group introduction on PEG [72].

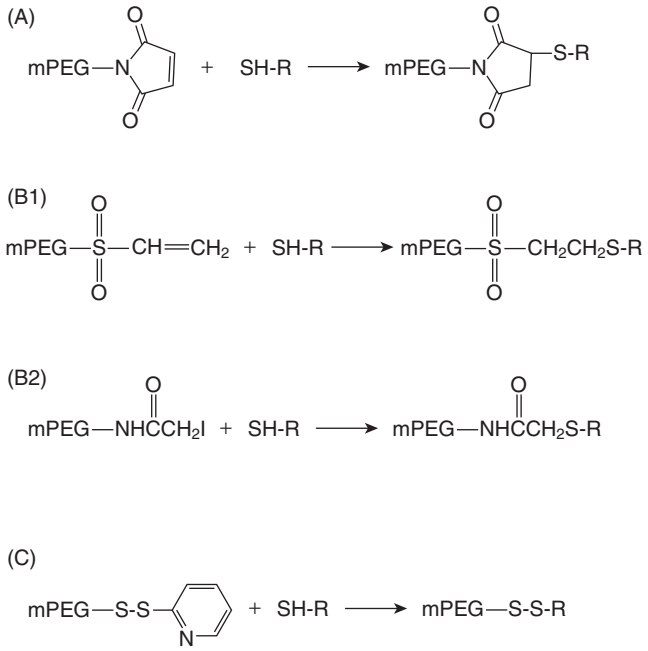


**Figure 10** Thiol reactive PEGs: (A) PEG orthopyridyl disulfide, (B1) PEG maleimide, (B2) PEG vinyl sulfone, and (C) PEG iodoacetamide [14].

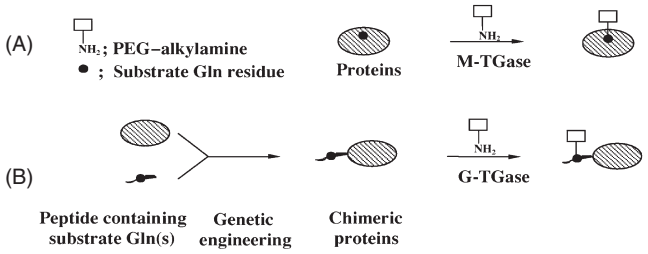


**Figure 11** Schematic representation of site-specific PEGylation of proteins by using (A) M-TGase and (B) G-TGase [69].

novel method using transglutaminase (TGase). According to this method, alkylamine derivatives of PEG are site specifically incorporated into chimeric and intact proteins with higher structural and biological efficiencies. It is comparatively more useful than thiol-selective and the amino-terminus-selective PEG incorporations as the incorporation site of the TGase-catalyzed modification is limited to the substrate Gln residues, the reaction conditions are mild, and dual and site-specific incorporation is possible by utilizing a special substrate sequence (TG2 or FibN) as the appending tag. Figure 11 shows the schematic representation of the site-specific PEGylation of proteins by the use of M-TGase and G-TGase. The authors have elaborately reviewed site-specific PEGylation of proteins in two sections—(i) by using microbial transglutaminase (M-TGase) and (ii) by using guinea pig liver transglutaminase.

## 7 CHEMISTRY OF PEGYLATION

The PEG polymer is required to be activated at one or both of its terminals before it is attached to a molecule [53]. PEG is functionalized by attachment of reactive moieties at both ends. In case the reactive moieties are similar at both terminals, the PEG derivative is termed as "homobifunctional" and if different from each other then it is known as "heterobifunctional" [48]. Reactive amino acids, which can be used as suitable functional groups, are lysine, cysteine, histidine, arginine, aspartic acid, glutamic acid, serine, threonine, and tyrosine [14]. The N-terminal amino group and the C-terminal carboxylic acid can also be used. Lysine and N-terminal amino groups are the most common reactive groups. The presence of lysine leads to replacement of many amines and many lysine modifications, due to which number of PEG molecules conjugate with a single protein molecule, which are also known as "PEGmers" [53]. The number of attachments depends upon the number of ε- and α-amino groups in the proteins. As there is increase in the degree of modification, number of positional isomers ($P$) formed is also increased. This can be calculated by the following equation:

$$P = \frac{N!}{(N-k)! \times k!}$$

where $N$ is the number of possible sites and $k$ is the number of sites modified. Higher degree of modification increases the circulation half-life and decreases the possibility of antigenicity [73]. This type of PEGylation is non-site-specific in nature, whereas site-specific PEGylation could provide better, specific, and reproducible advantages.

A site-directed method of joining proteins to PEG was presented by Bentley et al., [74], which allowed for the preparation of essentially homogeneous PEG–protein derivatives with a single PEG chain conjugated to the amine terminus of the protein. This selectivity is achieved by conducting the reductive alkylation of proteins with PEG–aldehydes at lower pH. Various PEGylation reaction chemistries involve the use of electrophilic activated PEG and ε-amino group of lysine or N-terminal amino group of protein. Some of the examples are, *N*-hydroxysuccinimide-activated esters (amide

bond), PEG–epoxide (alkyl bond), PEG–carbonyl imidazole (urethane bond), PEG–nitrophenyl carbonates (urethane bond), PEG–tresylate (alkyl bond), and PEG–aldehyde (N-terminus, Schiff's base) [21].

### 7.1 First Generation

To maximize the pharmacological benefits of PEGylation, a stable bond is formed between the PEG polymer and polypeptide drug of choice [48]. Chemical activation of the PEG polymer to react with a polypeptide drugs and to maximize the pharmacological efficacy is an advantage [75]. As described previously, ε- and α-amino groups of lysine is most common in proteins and, thus, is utilized mostly for first-generation protein PEGylation. The first-generation PEGylation techniques generally use linear PEGs with lower molecular masses (<12 kDa) forming multiple PEG isomers [76]. PEG used for PEGylation was methoxy-PEG (mPEG), with high PEG diol impurity leading to protein cross-linking and formation of inactive aggregates [77]. These were the key reasons for PEG–drug conjugate degradation, antigenicity, and ultimately stability and clinical problems. First-generation PEG polymers contained groups that were reactive with groups such as anhydrides, acid chlorides, chloroformates, and carbonates [78]. First-generation PEG derivatives that are involved in protein PEGylation are of two types:

1. Alkylating reagents, for example, PEG dichlorotriazine, PEG tresylate, $mPEG_2$-chlorotriazine
2. Acylating reagents, for example, PEG succinimyl carbonate, PEG *p*-nitrophenyl carbonate, PEG trichlorophenyl carbonate, PEG carbonyl imidazole, and PEG succinimidyl carbonate

Examples of first-generation PEGylated proteins approved by the FDA are PEGademase (Adagen), PEGaspargase (Oncaspar), and PEGintron.

### 7.2 Second Generation

The chemistry of second-generation PEGylation is focused on overcoming the limitations or drawbacks of first-generation PEGylation chemistry, namely, diol contamination, restriction to low MW mPEG, unstable linkages, side reactions, and lack of selectivity [48]. Gary Walsh [79] has described various therapeutically applied second-generation biopharmaceuticals. Roberts et al., [14] have discussed second-generation PEGylation by classifying them into different chemistries on the basis of protein modifications:

1. PEG chemistry for amine conjugation
2. PEG chemistry for cysteine modification
3. PEG chemistry for oxidized carbohydrates or N terminus
4. PEG chemistry for reversible PEGylation
5. Heterobifunctional PEG chemistry

***PEG Chemistry for Amine Conjugation*** Primary amine conjugation with aldehydes proceeds through Schiff's base, which reduces to give stable secondary amine linkage (Fig. 12) [14]. A noted example of second generation is mPEG-propionaldehyde, substituted to PEG-acetaldehyde, which is prone to dimerization by aldol condensation [48, 80]. Under acidic conditions (approx. pH 5), the aldehyde of mPEG–propionaldehyde is highly selective for N terminal α amine, which has lower $pK_a$ compared to other nucleophiles. Longer storage stability and higher purity could be obtained by using acetal derivative of propionaldehyde or aldehyde instead of their free forms [81, 82].

Acylating reagents used frequently for protein modification are active esters of PEG carboxylic acid. Ion exchange chromatography is used to purify PEG, achieving high (approx. 97%) purity levels. Other carboxylic acid derivatives developed subsequently over the years, with longer half-lives and lesser reactivity are propionic acid (PEG–O–CHCH–COOH) and butanoic acid (PEG–O–CHCHCH–COOH) derivatives of PEG, their succinimidyl esters (propionic acid-PEG, SPA–PEG, and butanoic acid-PEG, SBA–PEG) and α-methyl branched (PEG–NHS) derivatives [83].

***PEG Chemistry for Cysteine Modification*** Cysteine residue PEGylation is a common strategy for site-specific modification of proteins [52]. The number of free cysteines present on the protein surface is lesser than lysine residues or may be absent too; and in case they are absent, free cysteines can be added on the protein surface by genetic engineering techniques [84]. Such addition may, improve biological activity but may

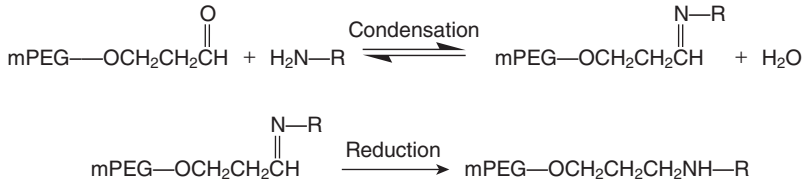


**Figure 12** Reductive amination using PEG propionaldehyde [14].

also form incorrect disulfide or cause protein dimerization. Some PEG derivatives suitable for PEGylation of cysteine residues are PEG–maleimide (PEG–MAL), PEG–vinylsulfone (PEG–VS), PEG–iodoacetamide (PEG–IA), and PEG–orthopyridyl disulfide (PEG–OPSS) [14]. Generally, these PEGs are thiol reactive and form stable thioether linkage with the protein. However, their stability is pH dependent, different at different (acidic or basic) conditions. Roberts et al., have developed a highly active, long circulating, and stable conjugate of IFN-b using a two-step method with PEG–OPSS [85]. In order to stabilize disulfide linkage in plasma protein between PEG and protein, a low-MW di-OPSS PEG (MW 2000) was coupled to interferon, and the remaining terminal OPSS group was then coupled to PEG thiol [86].

***PEG Chemistry for Oxidized Carbohydrate or N Terminus*** Site-directed PEGylation of proteins can also be carried out by oxidation (enzymatic, e.g., glucose oxidase; or chemical, e.g., sodium periodate) of carbohydrates and N-terminal serine or threonine [87]. Oxidation of carbohydrate residues forms multiple reactive aldehydes which react with either PEG-hydrazide to form hydrazone linkage or PEG-amine to form a reversible Schiff's base. Oxidation of N-terminal serine or threonine in the presence of sodium periodate forms a glyoxylyl derivative that can be further coupled with aminooxy and hydrazide PEG derivatives [88].

***PEG Chemistry for Reversible PEGylation*** PEGylation of peptide and protein drugs can improve their short in vivo circulatory half-life. However, the conjugation between the molecules may not be permanent, which would adversely affect the stability of the PEGylated protein. Development of several PEGylated protein conjugates with stable linkages has various advantages such as easier purification, improvement of circulatory half-life, longer shelf life, and reduction of protein activity possibly due to the presence of the PEG chain at the active or binding site of the protein or steric hinderance at the active or binding site [21]. Development of several protein conjugates with stable linkages can gain such properties.

However, PEGylation often leads to reduction or complete loss of biological and pharmacological activity of the protein. The PEG chain may also restrict the penetration of the bulky molecule only to selective body tissues because of the steric hindrance. These deficiencies of PEGylated peptides and proteins can be overcome by designing a $PEG_{40}$–protein conjugate with a hydrolyzable or enzymatically (serum proteases, esterases) cleavable functional group attached to the PEG chain [21]. Such a type of $PEG_{40}$–protein conjugate is capable of producing native protein at slow and constant rate under physiological conditions with desirable pharmacokinetic parameters. This property, called "reversible PEGylation," was investigated by Peleg-Shulman et al., [89] to improve the short circulatory half-life of the protein by coupling $PEG_{40}$–FMS (2 sulf-9-fluorenylmethoxy-carbonyl) to interferon-α2 (IFN-α2) as $PEG_{40}$–IFN-α2 is inactive in nature. The conjugate was activated by constant regeneration of the native protein by controlled hydrolysis of the FMS. Serum pH and nucleophilicity determine the rate of hydrolysis. Although reversible PEGylation overcomes the problems of conventional PEGylation, the concept itself presents various drawbacks. As for in vivo advantages, this method may not provide any improvement for in situ applications because the cleaving enzymes, such as proteases or esterases, may not be available at a rate-determining step, unpredictability, impracticable utilization, and nonspecific PEGylation leading to complex mixture formation and immunogenic reactions [90, 91].

Shechter et al., presented a new approach of reversible PEGylation. The authors studied the antidiabetic peptide exendin-4, insulin, and human growth hormone (hGH) as model drugs [92]. "Exendin-4 is an insulinotropic glucagon like peptide-1 agonist that associates with β-pancreatic cells, elevates endogenous cAMP levels, enhance secretion of Insulin, and lowers circulating glucose levels" [92, 93]. The basic hypothesis was to entrap a pharmacologically silent conjugate in the circulatory system releasing the parent peptide/protein through a nonenzymatic hydrolyzing pathway, which was exclusively dependent on pH, temperature, and protein content, and preserved in mammalian tissue fluids under strict homeostasis [89]. Several $PEG_{40}$–FMS–peptide/protein conjugates were prepared, and their efficacy in facilitating prolonged effects in mice, following a single subcutaneous administration, was studied. All conjugates were highly soluble in aqueous neutral media as inclusion of highly hydrophilic $PEG_{40}$-chain within the conjugate compensated the lowered solubility gained by the covalently linked FMOC (9-fluorenylmethoxycarbonyl) moieties (Fig. 13). The prodrug character of prepared conjugates allowed the administration of higher doses with no risk of toxicity, downregulation, or desensitization. Slow release of the covalently linked polypeptide drug maintained prolonged action over many hours. Thus, reversible PEGylation reduced the tendency of conventional PEGylation to inactivate short-lived therapeutic peptide and protein and extended their half-life, bioavailability, and efficacy. In another study, Shechter et al. [94] developed a reversibly PEGylated $PYY_{3\text{-}36}$ (peptide $YY_{3\text{-}36}$) derivative by coupling it to a 40-kDa PEG through a spontaneously cleavable linker. The resulting conjugate ($PEG_{40}$–FMS–$PYY_{3\text{-}36}$), a 1:1 conjugate of $PEG_{40}$–FMS

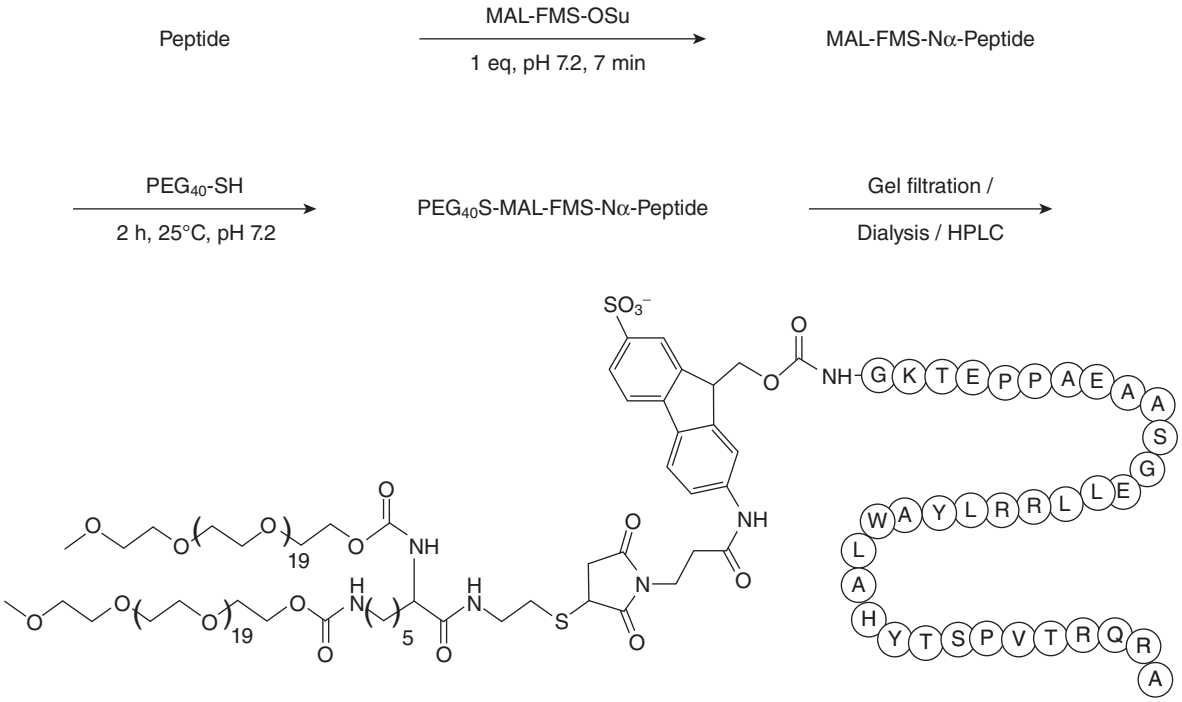


**Figure 13** Preparation of monomodified PEG40–FMS–peptide [89, 94].

covalently attached to $PYY_{3\text{-}36}$, gradually released unmodified $PYY_{1\text{-}36}$ in vivo, exhibiting an eightfold increase in its functional half-life, to approximately 24 h.

Tsunoda et al. developed another novel method for polymer-conjugated cytokines using TNF-α (tumour necrosis factor-α) as a model, with higher efficiencies, and investigated the usefulness of the resultant conjugate in vivo [94]. TNF-α and PEG were used for polymer conjugation, by using a pH-reversible, amino-protective reagent, dimethylmaleic manhydride (DMMAn). Significant improvements in antitumor activity were observed in vivo as compared to the changes in specific activity. Native TNF caused dose-dependen body weight loss in mice, whereas no obvious side effects were observed in any mice treated with PEG–TNF-α. PEG–TNF-α(+) had 20–30% higher specific activity than PEG–TNF-α(−), which was not treated with DMMAn.

***Heterobifunctional PEG Chemistry*** Hydrophilic, flexible, and biocompatible PEG derivatives with different terminal reactive groups used as spacers to link two macromolecular moieties, targeting of drugs, liposomes, or viruses to specific tissues are known as heterobifunctional PEGs [14]. Bentley et al. [95] prepared heterobifunctional PEGs by first initiating PEG polymerization with benzyl alcohol to produce benzyl PEG (Bz–PEG–OH). The hydroxyl group of the intermediate polymer is converted to a first reactive functional group. The benzyl group can then be removed by a hydrogenolysis functional group, thus making available a new terminal hydroxyl group for the addition of a second functional group. A promising strategy for generating heterobifunctional PEGs is the polymerization approach. Initiating polymerization of ethylene oxide with an anion that ultimately becomes the end group of PEG is the most direct route for heterobifunctional synthesis. Terminating the reaction with water generates a hydroxyl group at the second terminal end that can undergo activation by chemical methods. For example, this method has been used by Yokoyama et al., [96] to prepare a PEG with a hydroxyl group at one terminus and an amino group at the other. This strategy has its limitations. Only those anions that are desirable as end groups and suitable for initiating polymerization are useful for synthesis of heterobifunctional PEG by this route. This method is also limited by the fact that rigorous exclusion of water is necessary to prevent the formation of diol. This problem becomes more severe as PEG molecular weight increases. If these drawbacks are resolved, generating heterobifunctional PEG reagents by anionic polymerization may be the most efficient and cost-effective method. These PEG derivatives are also useful in immunassays, biosensors, probes, and bioconjugation techniques. Several heterobifunctional PEG derivatives can be prepared by methods such as anionic polymerization of ethylene oxide onto an initiator containing a protected functional group and terminated by alkylation, for example, α-amino-ω-hydroxy–PEG

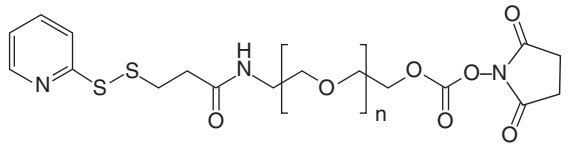

**Figure 14** PEG-based heterobifunctional *N*-succinimidyl-3-(2 pyridyldithio) propionate (SPDP) analog [72].

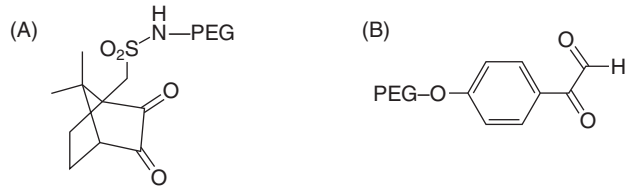


**Figure 15** PEG reagents for arginine modification [72].

$[NH_2(O–CH_2–CH_2)_nOH]$ for bioconjugation applications [71]. Zalipsky et al. reported a method to prepare pure heterobifunctional PEGs with two different functional groups at their end terminals, containing an amino or carboxyl groups at one of the termini [82].

Allen et al., [72, 98] used a macromolecular analog of SPDP [*N*-succinimidyl-3-(2 pyridyldithio) propionate], a PEG-based heterobifunctional reagent (Fig. 14) to prepare long circulating immunoliposomes. This reagent is used to link ligands to biomaterials and for protein–protein cross-linking.

Oligomeric heterobifunctional derivatives containing maleimide, iodoacetate, dithiopyridyl, and succinimidyl ester groups were prepared and used for preparation of protein conjugates interlinked through flexible and well water solvated oligo ethylene oxide spacers.

### 7.3 Carboxyl Group Modification

Carboxyl groups can be introduced onto PEG by various methods. Carboxymethylation of PEG with a haloacetic acid or its esters is a simple method for quantitative carboxylation of the polymer's end groups. Displacement of the bromide in ethyl bromoacetate with PEG–alkoxide, followed by saponification of the ester can also be carried out easily [72, 99].

Strong oxidizing agents, for example, permanganate [100], can cause polyether chain degradation and should be causiously used with PEG. Carboxylic acid groups at PEG terminals can also be substituted by modification with succinic (or glutaric) anhydride. Various organic solvents and specific pH conditions are utilized at different steps of synthesis. Formation of the final product depends upon the presence and type of catalyst used in the reaction, that is, the presence of only dicyclohexylcarbodiimide (DCC) as a catalyst does not lead to the formation of any ester or amide but forms *N*-acylurea derivative. However, DCC when used along with *N*-hydroxybenzotriazole (HoSu) forms amides and esters of various primary amines and alcohols [101]. Succinimidyl esters of carboxymethylated and succinylated PEGs are the most widely used reagents for modification of proteins, lipids, and other substrates [102, 103].

Kynclova et al., [104] reported a new method for covalent attachment of PEG to carboxylic groups of proteins (lipase from *Candida rugosa*) using *O*,*O*-bis-(2-aminopropyl)–PEG and carbodiimide/*N*-hydroxysuccinimide. Mild acidic pH conditions were favorable for higher protein activity, identification of preferred site, and most dominant isomer.

### 7.4 Arginine Modification

PEG reagent for selective modification of arginine residues [72] (Fig. 15A) was prepared by coupling camphorquinone 10-sulfonylchloride to amino-PEG [105]. The synthesized PEG reagent forms PEG–peptide conjugates in borate at pH 9, at 37°C. Such linkage between the components of the conjugates was cleavable by a treatment with *o*-phenylenediamine and was also stable to a variety of acidic and nucleophilic reagents. Thus, arginine modification was found to be suitable for enzymatic and chemical alterations of peptides, taking advantage of the properties of PEG for convenient purification of the final modified peptide and intermediate products. Another PEG reagent applied for selective modification of arginine residues of polypeptides is phenylglyoxal [106]. Attachment of its mPEG analog (Fig. 15B) to various proteins was carried out at room temperature and pH range of 5.5–9.3. It was characterized by a decrease in arginine content as determined by amino acid analysis of the conjugates.

### 7.5 OH Group Conjugation

The terminal primary hydroxy group of PEG is the most common site for drug–PEG conjugation reactions [107]. All the conjugates prepared from PEG–OH and free carboxylic acids under relatively mild conditions are summarized in Table 8. One of the most common approaches, that is, the conjugation of dicyclohexylcarbodiimide in the presence of methylaminopyridine as the carboxylic acid transformation into PEG esters occurs at room temperature, utilizing less time (2–4 h) with better results [108, 109]. In aqueous conditions, such esters may undergo hydrolysis, thus dissociating the conjugate components. The rates of hydrolysis depend on the nature of the carboxylic acid, molecular

**TABLE 8 Application of PEG–OH and Carboxylic Acid Conjugation**

| Sl No. | Drug | MW of PEG | Reagent/Condition[a] | Comments |
|---|---|---|---|---|
| 1. | Ibuprofen | 200, 900, 2000 | CDI in $CHCl_3$ at 60°C | Longer-lasting anti-inflammatory activity of PEG derivatives was observed. |
| 2. | Ursodeoxycholic acid | 1000, 2000 | CDI in $CHCl_3$ at 60°C | An efficient method for PEG esterification applicable to bile acids was demonstrated. |
| 3. | 2-Bezoxazolon-3-yl acetic acid | 200–1000 | DCC in THF | Conjugate of PEG-400 was used as a substrate in penicillin amidase-mediated synthesis of new biologically active cephem. |
| 4. | Benzoylacrylic acids | mPEGs 550–1900 | DCC/pyridine in $CH_2Cl_2$ | Antimicrobial activities were studied in vitro. |
| 5. | Penicillin V | mPEG-2000, 3000, 4000; poly(PEG-Lys) 50,000 | DCC/DMAP | Efficient esterification demonstrated. The conjugate is readily hydrolyzed releasing an active antibiotic. |
| 6. | Aspirin | 400, 1000, 3000 | DCC/DMAP | Aspirin esterification used as a model reaction for efficient synthesis of PEG–ester conjugates. |

[a]CDI: *N*,*N*´-carbonyl diimidazole; DCC: *N*,*N*´-dicyclohexylcarbodiimide; DMAP: 4-dimethylaminopyridine; THF: tetrahydrofuran.

weight of PEG, pH, and temperature. Predicted half-life times of this ester hydrolysis can range between several hours at pH 7.4 in phosphate buffer. Accordingly, it could be applied as an approach for slow release of an active drug under physiological conditions. However, use of appropriate spacers, such as succinate, can allow the use of PEG esters of a larger selection of drugs. Atropine and quinidine were successfully linked via ester linkages to PEG-succinates of various molecular weights [108].

Direct attachment of drugs to hydroxyl groups of the polymer is also possible by forming ether, carbonate, and urethane bonds via the appropriate alkylation and acylation reactions. For example, Khue and Galin [110] prepared carbonate-linked conjugates of hydrocortisone and dexamethasone by reacting their chloroformate derivatives with HO–PEG, where the derivatives were selectively formed on the primary hydroxyl groups. The yields of the coupling reactions were low to moderate. Anti-inflammatory steroids from the conjugates due to hydrolysis of the carbonate linkage was pH, temperature, and steric hindrance dependent [111]. Hydrolytic rates of PEG-steroid carbonates indicated the reaction to be base catalyzed. In a few cases, attachment of biologically relevant molecules to PEG was achieved directly by utilizing the reactivity of terminal primary OH groups of the commercially available polymer. In most cases, however, suitable functionalization of the polymer was observed to be an essential first step during formation of final conjugate. In the biological literature this process is often referred to as "activation" of the polymer [107].

Two principal approaches used to prepare functional derivatives of PEGs are [72] (1) direct transformation of hydroxyls to the new target functionality and (2) reaction of the polymer with a bifunctional molecule so that one function forms an attachment to the polymer and the other one remains available for further chemical transformations. Both approaches are often used, and each has its merits, which should be considered when planning a specific conjugation strategy. Table 9 shows various PEG conjugates and their applications.

Reaction of primary hydroxyl groups of PEG with organic sulfonyl chlorides was used extensively as a method for placement of good leaving-group residues at the polymer terminals and as an alternative to halogenation of PEG–OH [72]. Sulfonate esters display higher reactivity than halides and are often used as convenient starting materials for preparation of functionalized PEGs namely amino, thiol [71], and phenylglyoxal [112] to name a few. Tosylate-PEG is the most commonly used sulfonate ester. Several variations on the synthesis of this derivative claimed to provide complete transformation of the functional groups [56, 60].

## 8 LYOPHILIZATION OF PEGYLATED PROTEINS

Freeze-drying is the separation of liquid water from a solution frozen to ice by vacuum sublimation, leaving the solutes in an anhydrous or almost anhydrous state [113]. Lyophilization maintains bioactivity and stability of protein formulations over their shelf life. Stability of protein formulations can be improved by the addition of a lyoprotectant or cryoprotectant such as a carbohydrate, for example, sucrose and trehalose, during the lyophilization process [114]. These cryoprotectants form

**TABLE 9 Various PEG Conjugates and Their Applications**

| Sl. No. | PEG Conjugates of | Useful Properties and Applications |
|---|---|---|
| 1. | Drugs | Improved solubility, controlled permeability through biological barriers, longevity in bloodstream, controlled release |
| 2. | Affinity ligands | Used in aqueous two-phase partitioning systems for purification and analysis of biological macromolecules, cells |
| 3. | Cofactors | Bioreactors with continuous regeneration and recycling of macromolecular cofactors |
| 4. | Peptides | Improved solubility, conformational analysis |
| 5. | Proteins | Resistance to proteolysis, reduced immunogenicity and antigenicity, longevity in bloodstream. Uses: therapeutics, organic-soluble reagents, bioreactors |
| 6. | Saccharides | New biomaterials, drug carriers |
| 7. | Oligonucleotides and analogs | Improved solubility, resistance to nucleases, cell membrane permeability |
| 8. | Lipids | Used for preparation of PEG-grafted liposomes |
| 9. | Liposomes and particulates | Longevity in bloodstream, RES evasion |
| 10. | Biomaterials | Reduced thrombogenicity, reduced protein and cell adherence |

an amorphous or partially amorphous stabilizing matrix in the protein formulation, which in turn restrict mobility and reactivity, reduce aggregation [115], and enhance molecular interactions by hydrogen bonding between carbohydrates and proteins [116].

Mosharraf et al. [117] carried out a detailed study on lyophilization of PEGylated growth hormone antagonist (PEGvisomant) as a model protein. They determined the effect of formulation and process parameters on the solid-state properties such as degree of solid-state disorder (amorphicity), protein–PEG–excipient interactions, moisture-induced recrystallization of amorphous samples, cooling rates for the solid-state properties, and dissolution characteristics of the resultant lyophilized cakes. Interactions between sucrose and PEGylated protein molecules were demonstrated by cooling rates, decrease in $T_m$, and melting enthalpy of PEG, which were determined by differential scanning calorimetry (DSC) and isothermal microcalorimetry (IMC). Moisture-induced recrystallization of the lyophilized cakes was not prevented by PEGylation. Ratio of protein to lyoprotectant or stabilizer also affected the stability of the lyophilized formulation. In a study carried out by Tattini et al. [118] different ratios (1:0.25 and 1:0.50) of bovine serum albumin to polyethylene glycol (BSA:PEG) were used. BSA–PEG in a ratio of 1:0.25 was found to have better structural stability and optimized conditions for lyophilization and protein storage as compared to that of 1:0.50. Another protein stabilizer or lyoprotectant reported to prevent protein aggregation caused during lyophilization is inulin. It has been used to stabilize PEGylated nanoparticles and liposomes. Inulins and dextrans can prevent protein aggregation of plasmid DNA and PEGylated or non-PEGylated polyethylenimine (PEI) complexes [119]. However, inulins are found to be more effective over dextrans for PEGylated proteins or protein formulations due to their compatibility with PEG. Freeze-drying and freeze-thawing of 1,2-dioleoyl-3-trimethylammonium-propane (DOTAP)/dioleoyl-phosphatidyl-ethanolamine (DOPE) liposomes and egg phosphatidyl choline (EPC)/cholesterol (CHOL) liposomes can be stabilized by using inulin [120].

Another mechanism explaining the degradation of proteins during freeze-drying or lyophilisation was reported by Heller et al. [5]. In this study, the effect of PEGylation on the conformational stability and partitioning behavior of therapeutic proteins in two-phase systems (PEG–dextran) during lyophilisation process were determined. PEGylated and non PEGylated forms of hemoglobin and brain-derived neurotrophic factor (BDNF) were partitioned separately in PEG–dextran phase separating systems. Partition coefficients of both PEGylated and non-PEGylated proteins were obtained from the ratio of their concentrations in PEG and dextran phases, respectively. Protein concentration in each phase for BDNF and PEGylated BDNF was measured by ultraviolet (UV) analysis at 280 nm, whereas haemoglobin and PEGylated haemoglobin were analyzed by BCA total protein assay. For primary drying the samples were freeze dried at –20°C for 1000 min at 60 mT vacuum after annealing at –7°C for different time lengths followed by cooling below –30°C for 120 min and for secondary drying, 0, 10, and 25°C for 120 min at 60 mT vacuum. Both the proteins behaved differently in PEG–dextran two phase systems. PEGylated haemoglobin and BDNF were found to have greater affinity toward the PEG-rich phase than the dextran-rich phase. In case of PEGylated BDNF, the affinity for the PEG-rich phase is enhanced, whereas haemoglobin exhibited more affinity toward the dextran-rich phase. These results confirm the alteration in partitioning behavior of

proteins after PEGylation, thus its surface characteristics and interaction with other surfaces, interfaces, and cosolutes [120]. Therefore, depending upon the properties of given therapeutic protein and its compatibility with added excipients, PEGylation can have favorable or unfavorable effect on lyophilization and ultimately on its stability.

## 9 CHARACTERIZATION OF PEGYLATED PRODUCT

### 9.1 Physicochemical Characterization of PEGylated Products

Solid-state properties such as internal cake structure and porosity, residual moisture, weight ratios of protein:PEG:sucrose (w/w/w), PEG melting endotherm, sucrose crystallization exotherm, sucrose melting endotherm, sucrose degree of disorder, PEG degree of disorder (amorphicity) of lyophilized PEGylated protein formulations, determined by Mosharraf et al. [117], have already been introduced briefly in Section 8. Different methods to determine these solid-state properties of "lyophilized" PEGylated proteins are presented next.

***Internal Cake Structure and Porosity*** The study carried out by Mosharraf et al. [117] was the first to report the effect of cooling rate on the pore size of a lyophilized PEGylated protein. Each lyophilized cake was sliced into three pieces. The middle slice was then scanned using a scanning electronic microscope under an acceleration voltage of 20 kV with 75× and 300× magnifications. Scanning electron microscopy photomicrographs revealed that the cakes, which were lyophilized at a cooling rate of 0.11°C/min had a smaller pore size and a more homogeneous pore distribution within the cake structure than in the corresponding formulations cooled 10 times faster at 1.1°C/min. As the size and number of ice nuclei are related to the degree of supercooling, higher degree of supercooling leads to smaller crystals. It was also concluded that increase in sucrose concentration (added to the protein as lyoprotectant) decreased the pore size. Some scientists [5, 121], however, reported that faster cooling rates lead to higher rate of nucleation and hence smaller ice crystals. When colyophilized rhGH analog/sucrose formulations were compared at different amounts of sucrose (5, 25, and 50 mg/mL) at the same cooling rate (0.11 or 1.1°C/min), a decrease in pore size was observed with increasing amounts of sucrose.

***Determination of Residual Moisture*** As stated in Mosharraf and co-workers: "The percentage of residual moisture in each formulation was determined using a 684 Karl Fischer coulometer. A solution of methanol and formamide (70:30) was used as solvent, to extract the moisture from the solid cake. For each formulation, three cartridges were analyzed, and the average percent residual moisture and standard deviation (S.D.) were calculated" [117, p. 220].

Residual moisture contents in lyophilized sucrose formulations were approximately 1.7–3.5%, whereas in all other lyophilized formulations, the water content was <1%, indicating efficient drying. It was observed that cooling rate did not affect the moisture content, but the additio n of PEG in the formulation reduced the moisture content of the lyophilized cake.

It was also observed that PEG and protein influence the amount of residual moisture in formulations containing sucrose. On comparing water content in the PEGylated protein formulations with that in the non-PEGylated protein formulations, it was found that PEGylation caused a decrease in water content. PEGylation may alter the ability of protein to bind to water, thus increasing the tendency of water to leave the lyophile during secondary drying.

***X-Ray Powder Diffraction (XRPD)*** Diffraction patterns of the materials were obtained using a diffractometer, with CuKα radiation at 45 kV and 40 mA [117]. The samples were scanned in steps from 5° to 45° (2θ), the crystalline ordered structure of sucrose. The XRPD pattern for crystalline sucrose showed sharp peaks indicative of long-range order. The patterns for the lyophilized formulation of PEGylated rhGH analog in 50 mg/mL sucrose and quenched sucrose both have broad, diffuse spectra typical of a highly disordered (amorphous) structure.

***Differential Scanning Calorimetry*** As stated in Mosharraf and co-workers: "DSC experiments were performed on a DSC 220. Special attention was paid to minimizing the exposure of the samples to moisture prior to DSC analysis. Rubber-stoppered cartridges containing samples were opened in a nitrogen flushed chamber. Immediately after opening, the freeze-dried cake was gently disintegrated and a sample of approximately 2.5–3.5 mg was accurately weighed in an aluminium pan and sealed hermetically. An empty pan was used as reference. The sample and the reference were exposed to a linear heating ramp in the temperature range 20–225°C at a heating rate of 5°C/min in an atmosphere of (100%) nitrogen. For each formulation, duplicate samples were analyzed ($n = 2$). Exothermic signals were given positive values" [117, p. 221].

From the DSC curve obtained, it was found that at slow cooling rates the degree of solid-state disorder of

sucrose was dependent on sucrose concentration in the pre-lyo solution. Lyophilization of PEGylated protein without excipient at a slow cooling rate resulted in partially crystalline cakes with higher degree of crystallinity, which could be due to the presence of PEG molecules. Thus, it can be concluded that PEG increases the crystalline character of the cakes [122]. The presence of crystalline character in the samples was determined by interpretation of DSC scans, which revealed characteristic crystal melting endotherm [5]. Both IMC and DSC results confirmed the partial crystalline characteristic of the lyophilized pegylated protein. Furthur, it was also observed that PEG–protein conjugation altered the physicochemical properties of PEG as well as protein.

***Isothermal Micro Calorimetry*** As stated by Sebhatu and co-workers: "The IMC experiments were performed in a thermal activity monitor (TAM) using the ampoule method" [123, p. 137]. Three to six freeze-dried cakes were removed from the cartridges, disintegrated, and pooled. A sample of about 100 mg of the disintegrated cake (powder) was placed into a vial, weighed carefully, and analyzed in the microcalorimeter. Samples of crystalline sucrose and quenched sucrose were also tested. Storage of disordered material at high relative humidity leads to water vapor absorption and recrystallization due to which low relative humidity (RH) (57.5%) was maintained during the experiment by placing a glass tube containing saturated NaBr solution into the vials, which were then sealed. An empty glass tube in a sealed vial was used as reference. The sample and reference vials were equilibrated in the TAM for 15 min before starting the experiment. All experiments were performed at 25°C. The heat flow signal ($dQ/dt$ in microwatts) was monitored as a function of time. The heat flow curve obtained for a pure amorphous reference usually consists of a moisture absorption phase and a recrystallization phase. Because of overlapping moisture absorption and recrystallization peaks, the heat of interaction with water ($Q$, measured in joules) was calculated by integrating the total area under the heat flow curve [124]. The values were normalized for eventual weight differences. These values were used to calculate the degree of solid-state disorder.

Lyophilized sucrose and lyophilized PEGylated protein formulation were not analyzed by this method because of sample limitations. The degree of disorder in the remaining formulations was calculated using the $Q$ value of quenched sucrose as the 100% amorphous reference. The 100% amorphicity of this sample was verified by XRPD. The degree of disorder was determined by dividing the $Q$ value obtained for the sample by that of quenched sucrose and multiplying by 100. Except for crystalline sucrose, all the samples interacted with water vapor resulted in exothermic peaks. In all cases except that of crystalline sucrose, the samples interacted with water vapor, resulting in exothermic peaks demonstrating the amorphous character of the lyophilized and quenched samples. The $Q$ values were considered as heat of recrystallization and were calculated by integration of heat flow curves [124] and normalized for the weight of the samples. The degree of disorder in the formulations containing sucrose was calculated using the $Q$ value for quenched sucrose (41.5 ± 9.2 mJ/mg) as the 100% amorphous reference. The crystalline sucrose sample did not react with moisture, confirming a totally crystalline structure for this sample.

***Ion Exchange Chromatography (IEC)*** High-performance ion exchange chromatography (HP-IEC) can be performed on an analytical scale Mono S column, using a Protein Purification System [125]. The column is previously equilibrated with 0.04 *M* sodium acetate, pH 5.0 (buffer A) for 10 column volumes, at a flow rate of 0.2 mL/min at room temperature. Samples are loaded and washed with one column volume of buffer A. Subsequently, the column is eluted with a linear gradient to 100% of the same buffer containing 0.15 *M* sodium chloride (buffer C, pH 5.0) over 15 column volumes, monitoring the absorbance at 280 nm. The HP-IEC results of two monomer conjugates, mono-PEGylated rhIL-1ra with SC-$PEG_{5k}$ and mono-PEGylated rhIL-1ra with SC-$PEG_{10k}$ revealed that rhIL-1ra was composed of 153 amino acids with an MW of 17.5 kDa. It was also observed that the attachment of polymer was prone to occur at imidazole groups of histidine, which was independent on peptides or proteins [126]. The same ionic properties of purified monomeric conjugates on analytical Mono S column and the same hydrophobic behaviors on analytical C4 column suggested that rhIL-1ra might be site-specifically PEGylated.

***Zeta Potential Measurement*** Surface charge of PEG–PIBCA (poly-iso-butyl cyanoacrylate) nanoparticles prepared at different pHs (0.75–1.5) was studied by zeta potential measurement in phosphate-buffered saline (PBS) containing KCl and also repeated with water to evaluate the influence of ionic strength on the zeta potential [127]. Surface hydrophobicity of PEG–PIBCA nanoparticles was investigated by HIC (hydrophobic interaction chromatography). HIC is a technique of column chromatography capable of separating material on the basis of its interaction with a hydrophobic gel matrix [128]. The sample suspension was injected into a column filled with propyl agarose and connected with a pump. Elution was performed using PBS containing NaCl. Washing was carried out with a buffer containing Triton X-100. Optical density of the eluted samples was

determined spectrophotometrically at 400 nm. HIC was performed on PIBCA and PEG-coated PIBCA nanoparticles, prepared at pH 0.75–1.5.

PEG–PIBCA nanoparticles prepared at pH 1–1.5 presented, in PBS, a mean zeta potential of 218 mV, whereas when the pH of the polymerization medium was 0.75, the surface charge was much more negative (230 mV). When suspended in water, nanoparticles prepared at any pH exhibited the same zeta potential (230 mV). Hydrophilic surface properties (determined by HIC) showed that only particles produced at pH 1–1.5 presented a more hydrophilic surface. Nanoparticles prepared at pH 0.75, as well as noncoated PIBCA nanoparticles, interacted with the propyl-agarose and were eluted through the column only after washing with Triton X-100.

Interestingly, data relative to surface charge can be compared to PEG association to nanoparticles as a function of the pH. At pH 0.75 a very negative surface charge corresponded to a very low PEG association, whereas at pH 1 surface charge was less negative and PEG content higher. This suggested at least a partial PEG disposition on the nanoparticle surface, even if the PEG total content was reduced in nanoparticles prepared at pH 1.25 and 1.5 when compared to nanoparticles prepared at pH 1. No differences were observed by varying the pH of the polymerization medium, suggesting a strong influence of the higher ionic strength of the medium (PBS/KCl).

### 9.2 Biochemical and Biological Characterization of PEGylated Proteins

***Protein Determination [21]*** Commonly used methods to determine the concentration of protein in a sample are, for example, bicinchoninic assay (BCA) [129], enzyme linked immunosorbant assay (ELISA) [130, 131], ultraviolet spectroscopy, and Bradford assay. Amino acid composition analysis is also used and considered an accurate method for protein analysis.

According to the Bradford method [132], protein concentrations can be determined spectrophotometrically using Coomassie Brilliant Blue G-250. Crystalline BSA was used to prepare standard curve. Absorbance was read at 595 nm.

***SDS–PAGE (Sodium Dodecyl Sulfate–Polyacrylamide Gel Electrophoresis)*** SDS is an anionic detergent that denatures proteins by "wrapping around" the polypeptide backbone, and SDS binds to proteins fairly specifically in a mass ratio of 1.4:1. In doing so, SDS confers a negative charge to the polypeptide in proportion to its length, that is, the denatured polypeptides become "rods" of a negative charge cloud with equal charge or charge densities per unit length. It is usually necessary to reduce disulfide bridges in proteins before they adopt the random-coil configuration necessary for separation by size: This is done with 2-mercaptoethanol or dithiothreitol. At a given pH and under nondenaturing conditions, the electrophoretic separation of proteins is determined by both size and charge of the molecules [133].

All proteins have a large negative charge, which means they will all migrate toward the positive pole when placed in an electric field [134]. In denaturing SDS–PAGE separations, therefore, migration is determined not by intrinsic electrical charge of the polypeptide but by molecular weight gels of given acrylamide concentration that separates proteins effectively within a characteristic range. Very large polypeptides cannot penetrate far into a gel, and thus their corresponding bands may be too compressed for resolution. Polypeptides below a particular size are not restricted at all by the gel, and regardless of mass they all move at the same pace along with the tracking dye. Gel concentration should be selected so that the proteins of interest are resolved.

SDS–PAGE analysis of PEGylated proteins can be performed according to the method described by Laemmli [135]. The gels can be stained specifically for protein or PEG. Staining for PEG is done according to a modified procedure of Kurfurst [136]. SDS–PAGE gels are rinsed with distilled water and placed in a 5% barium chloride solution for 10 min. After rinsing the gel with distilled water to remove barium chloride, the gel is placed in 0.1 *M* Titrisol for another 10 min. After washing away the yellow Titrisol, the gel is stored in distilled water in a heat-sealed bag in the dark.

The polyacrylamide concentration was 5% in the stacking gel and 12% in the separation gel, using a Miniprotein III cell. Each sample was mixed with the fivefold reducing buffer, boiled for 5 min, and loaded into each well. Silver staining was used to visualize the proteins [125].

***Determinaton of Molecular Weight of PEGylated Proteins*** Several methods have been used to determine the molecular weight of PEGylated proteins, namely mass spectroscopy (electrospray ionization mass spectroscopy or MALDI-TOF mass spectroscopy) [137], SDS–PAGE analysis, and CE (capillary electrophoresis) analysis [138], each method having its own advantages and disadvantages.

***Identification of PEGylation Site and Positional Isomers*** The process to identify a suitable PEGylation site and its positional isomers involves various steps (as discussed in Section 5, and thus different analytical

methods may be required at each step. Some common methods used to identify PEGylation sites on protein molecules and their positional isomers are ion exchange chromatography, size exclusion chromatography (SEC) [139], high-performance liquid chromatography (HPLC), reversed-phase HPLC, peptide mapping, Edman degradation, amino acid analysis, MALDI-TOF mass spectroscopy [137], and Raman spectroscopy and nuclear magnetic resonance (NMR) [138].

***Estimation of Grafting Density/Degree of PEGylation/Extent of PEGylation/Modification Extent*** The determination of polymer content in PEGylated proteins, that is, modification extent, is essential in order to control the half-life, maintain manufacturing uniformity, and satisfy regulatory concerns [140]. Various methods have been reported to determine modification extent. Some of the useful methods are discussed next.

*TNBS (Trinitrobenzene Sulfonic Acid) Method [141]* The TNBC method is the most commonly used method to obtain modification extent of PEG compounds due to its speed and simplicity. However, this colorimetric method detects only modified lysine reactivity and not actual numbers of PEG in the protein [142]. According to this method, serial dilutions of the native RP (recombinant protein) and RP–PEG are made in 100 m*M* sodium phosphate (pH 8.3), followed by the addition of the TNBS reagent. The reactions are then heated to 40°C for 2 h. The absorbance is determined and absorbance versus protein concentration graph is plotted from which a slope of the line is calculated. The number of PEG molecules attached to each molecule of RP = 1 − (slope of native RP/slope of RP–PEG) × 29, where 29 represents the number of primary amines present in RP (28 lysines + 1 N terminal).

*Microfluidics Assay* Yu et al. investigated a simple CE-based method to separate PEGylated conjugates and quantitate the separated species [142]. Separation is based on relative molecular size in the presence of SDS. The method is consistent and fast (90 s/sample analysis), and it consumes a limited sample (up to 4 μL). Even more advantageous, this method does not require post-electrophoretic handling such as the staining and destaining of PAGE. Each protein 200 LabChip contains an interconnected set of microchannels in which proteins (mixed with a fluorescent dye and SDS) are sieved by size as they are pulled through a polymer matrix by electrical current. At the end of the channel, separated proteins are mixed with a destaining solution and quantitated with a sensitive fluorescence detector that is connected to a computer for run control and automated data analysis. Each species of PEGylated protein has its size and relative concentration reported. It can provide data output in either an electropherogram or a simulated slab gel picture format, and it is well suited for electronic record storage.

An Agilent Bioanalyzer equipped with Chip Priming Station, Protein 200 plus LabChip kit, and the Agilent software was used for analysis. Typically, concentrated protein RP–PEG samples were diluted to 0.5–1 mg/mL in PBS saline and free RP was diluted to 0.1 mg/mL to acquire peaks distinguishable from the baseline. Standard sample preparation involved mixing 2 μL of 100 μg/mL RP or 1000 μg/mL RP–PEG with 4 μL reduced sample buffer. Samples were then incubated at 95°C for 5 min, and then 84 μL of distilled water was added followed by loading of 6 μL of the mixture onto the Protein 200 Chip (along with other reagents) to the appropriate wells. Two internal markers are mixed with each test sample and with a molecular weight reference "ladder." The quantitation is calculated by comparing corrected peak area with either an upper internal marker or a standard reference.

Other methods that have been used for determination of modification extent are NMR, amino acid analysis, capillary zone electrophoresis (CZE), matrix-assisted laser desorption/ionization–mass spectrometry (MALDI–MS), Fourier-transform infrared spectroscopy (FTIR), Raman spectroscopy, online size-exclusion high-performance liquid chromatography (SE-HPLC), and light scattering differential refractometry methods [137, 139, 143–145]. Although these methods are much more accurate than TNBS and fluorimetric methods, they are less popular as these methods require tedious operations, sophisticated equipments, and skilled operators. Secondly, these methods consume PEGylated proteins, which are generally costly. Thus, Gong et al. [140] have reported another alternative method, which is comparatively simpler and cost effective. The free PEG in the given sample was determined according to the procedure of Sims and Snape [146] with some modifications.

Briefly, samples were diluted with distilled water to final concentrations of PEG within the range of 0–7.5 μg/mL. To an 800-μL sample, 200 μL barium chloride solution and 100 μL iodine solution were added subsequently. The reagent blank was prepared as above with distilled water instead of sample. Solutions were agitated to ensure adequate mixing. Color was allowed to develop for 15 min at room temperature, and then absorbance was read at 535 nm. Inactivated PEG, which has the same molecular weight as that of free PEG, was used to prepare a standard curve. The concentration of free PEG was estimated from the standard curve. Modification extent was calculated according to

$$\mathrm{ME} = \left(\frac{m_a - m_c}{m_b \times N_b}\right) \times 100 \quad (1)$$

where ME is the modification extent, $m_a$ the molar amount of PEG used for modification, $m_b$ the molar amount of protein, $m_c$ the molar amount of free PEG, and $N_b$ is the amount of target group in a one-protein molecule.

***In Vitro Release Studies of PEGylated Products*** Dissolution studies of lyophilized PEGylated proteins can be carried out by a method described by Mosharraf et al. [117]. According to this method, the rear compartment of a dual-chamber cartridges was filled with 1 mL water for injection (WFI) and sealed. The cakes were then reconstituted using a threaded mixing device. The cartridges were turned over every 30 s manually until a clear solution was obtained. The time taken for dissolution of the cake was measured. For each formulation, dissolution time was recorded for three cakes and the mean dissolution time and standard deviation (SD) were calculated.

Lyophilized sucrose without additives dissolved very quickly. Lyophilized free PEG also dissolved rapidly. The dissolution time for cakes containing PEGylated rhGH analog was reduced to a third of that for cakes containing non-pegylated protein and free PEG (rhGH/PEG). A decrease in dissolution time in these cases is likely to be caused by PEGylation. It is known that PEGylation changes the surface properties of protein and increases protein solubility. The dissolution time of co-lyophilized PEGylated protein (PEGylated rhGH analog/sucrose (30:30:25) was shorter than co-lyophilized non-pegylated protein (rhGH) in the same amount of sucrose [rhGH/sucrose (30:0:25)], with the former being 1.5 times more disordered than the latter. Hence, the results showed that PEGylation reduces the dissolution time of the cake by increasing the degree of solid-state disorder. The higher the heat of interaction with water (i.e., degree of amorphicity), the shorter the dissolution time, until the dissolution time seemed to reach a plateau, when the solid was completely disordered. Parameters such as surface area and solubility are two of the most important parameters for dissolution rate (and thus dissolution time) of drugs. It has been suggested that the extent of solubility enhancement is closely related to the amount of amorphous material in the product [147]. An increase in apparent solubility would lead to an increase in initial dissolution rate (and thus a decrease in initial dissolution time).

A study carried out by Park et al. [148] compared the in vitro release property of mono-PEGylated growth hormone releasing peptide-6 (mono-PEG-GHRP-6) microspheres with native GHRP-6. The study was performed in 0.1 *N* phosphate buffer, pH 7.4, containing 0.02% Tween 80 and sodium azide at 37 or 55°C. The release profile at 55°C was carried out to observe drug release under accelerated conditions. At a specific time intervals samples were withdrawn, centrifuged, and the supernatant was analyzed by HPLC. It was found that the rate of drug release from microspheres decreased with increasing size of the PEG attached to GHRP-6.

***Pharmacokinetics and Biodistribution*** PEGylation alters physicochemical properties of proteins and modifies their pharmacokinetic parameters which in turn affect pharmacodynamic response. These modifications are utilized to deliver therapeutic proteins more effectively with better absorption, sustained plasma concentration, longer half life, reduced immunogenicity, decreased adverse reactions, and improved patient compliance. Alteration in size of PEG chain also alters the biodistribution behavior of PEGylated protein. Thus, increased PEG molecular weight decreases renal clearance, restricts biodistribution, and improves bioactivity. However, it is limited to a particular mass after which no increase in bioactivity is observed. Influence of pharmacokinetics and pharmacodynamics effect of protein PEGylation is given in Table 10 [149].

A potential effective antidiabetic agent, glucagon-like peptide-1-(7–36) (GLP-1) is a glucose-dependent insulinotropic hormone. However, a major limitation of GLP-1 is its very short half-life in plasma, which is due to its degradation by dipeptidyl peptidase IV (DPP-IV) enzyme. Thus, in an attempt to improve its therapeutic efficacy and reduce the dose and dosage frequency, a novel PEGylated GLP-1 conjugate was prepared by Lee et al. [150]. In vitro characterization and in vivo pharmacokinetic parameters of mono-PEGylated GLP-1 were also determined. The concentration of GLP-1 in plasma samples was analyzed by ELISA. The pharmacokinetic parameters were determined by non-compartmental and two-compartmental analysis after subutaneous (sc) and intravenous (iv) administration to Sprague–Dawley rats. Half-life of GLP-1 after iv administration was found to be 2.6 min, whereas mono-PEGylated GLP-1 showed 10-fold increase in half life, 1.5 fold reduction in renal clearance, and 16-fold increase in mean residence time. Similarly, after sc administration mono-PEGylated GLP-1 showed 2-fold increase in $C_{max}$, 7.5-fold increase in AUC (area under the curve), and 2.5-fold increase in half life as compared to GLP-1.

Owens et al. [151] reviewed the process of opsonization, biodistribution, and pharmacokinetics of polymeric nanoparticles and reported PEGylation to be the generally preferred method to prevent the nanoparticle detection by mononuclear phagocytic system (MPS) of

**TABLE 10 Pharmacokinetic and Pharmacodynamic Effect of Protein PEGylation**

| Sl. No. | Molecule | Pharmacokinetic Effect | Pharmacodynamic Effect |
|---|---|---|---|
| 1. | Interferon-α2a | Sustained absorption (subcutaneous first-order reaction)<br>Increased half-life (from 3–8 h to 65 h)<br>Decreased volume of distribution (from 31–73 to 8–12 L)<br>Decreased systemic clearance (from 6.6–29.2 to 0.06–0.10 L/h) | In vivo antiviral activity increased 12–135 times<br>Antitumor activity increased 18-fold<br>Improved sustained response to chronic hepatitis C |
| 2. | Interleukin-6 | Increased half-life (from 2.1 to 206 min) | Thrombopoietic potency increased 500 times<br>Decreased IgG1 production |
| 3. | Tumour necrosis factor (TNF) | Increased half-life (from 3 to 45–136 min) | Antitumor potency increased from 4 to 100 times<br>Decreased adverse events |
| 4. | Growth hormone releasing factor | | Potency increased 12- 55-fold, enhanced duration of action |
| 5. | Megakaryocyte growth and development factor | Increased half-life (10-fold) | In vitro activity increased 20-fold |
| 6. | Brain-derived neurotrophic factor | Increased half-life 5-fold (from 10 to 50 min)<br>Decreased clearance 2.6-fold (from 5.5 to 1.9 mL/min/kg)<br>Increased volume of distribution 1.7-fold (from 80 to 135 mL/kg) | |
| 7. | Asparaginase | Increased half-life 18-fold (from 20 to 357 h)<br>Unchanged volume of distribution<br>Increased AUC 26-fold (from 0.4 to 10.2 IU/mL/day<br>Decreased clearance 17-fold (from 2196 to 128 mL/m$^2$/day | |
| 8. | Suproxide dismutase | Increased half-life > 150-fold (from 3.5 to 540–990 min) | |
| 9. | Lactoferrin | Increased half-life 5- to 20-fold (from 3 to 15–60 min) | |
| 10. | Streptokinase | Increased half-life 1.7- to 5-fold (from 4 to 7–20 min) | Decreased antigenicity |
| 11. | Interleukin-2 | Increased half-life up to 6-fold (from 44 to 57–256 min)<br>Decreased clearance up to 10-fold (from 1.15 to 0.11–0.97 mL/min)<br>Unchanged volume of distribution | |

the body. The authors have discussed various methods to analyze in vitro and in vivo biodistribution and clearance of PEGylated conjugates. The in vitro methods are flow cytometry and cell-associated fluorescence measurements whereas in vivo biodistribution studies in animal models employ $^{14}$C radiolabeling or encapsulated indium histological studies.

***Stability of PEGylated Product*** Changes in the surface structure by chemical modification can impart useful and required properties to proteins. Highly flexible, well-defined chemistry and amphipathic properties of PEG make it a suitable polymer for carrying out chemical modifications in proteins [152]. Many methods, including immobilization, site-directed mutagenesis, chemical modification, and cross-linking have been described for protein modification [153].

Christakopoulos et al. [154] modified the surface of the low-molecular-weight endoglucanase of *Fusarium oxysporum* by covalent protein PEGylation with PEG

derivatives specific for lysine or cysteine attachment. The authors studied the effect of these modifications on pH stability.

The pH stability was determined by incubating the enzyme in the pH region 7.0–12.0, at 30°C for 24 h and measuring the residual activity at pH 7. The following buffering systems were used: 0.1 *M* phosphate (pH 6–8), 0.1 *M* Tris–HCl (pH 9.0), 0.1 *M* glycine–NaOH (pH 10), 0.1 *M* KCl–NaOH (pH 11–12).

*Effect of pH on PEGylated Protein Stability* At pH 12, 25°C temperature, and 24 h incubation time, the native enzyme (endoglucananase) was completely inactivated while 66 and 73% of the initial activity remained in two different PEG-modified endoglucanase. The modified endoglucanase showed enhanced stability due to the buffering action created against denaturation of proteins by the long PEG chains present with the copolymer modifiers such as SS–PEG (succinimidyl succinate–PEG) and Mal–PEG (maleimide–PEG) [155].

The effect of reaction pH on PEGylation of rh-interleukin-10 with histidine was investigated by Wylie et al. [126, 156]. In this study, mono-PEGylated interferon was prepared by reaction at pH 6.5 and 8.6. The sample created at pH 6.5 was stored at pH 6.8 and ambient temperature for 30 days. By SE-HPLC, only 4% was dePEGylated to unmodified interferon. At pH 8.6, however, 22% of mono-PEGylated interferon was dePEGylated only after 24 h. Thus, whereas the PEGylated isomer is more labile under basic conditions, it is stable at physiologic pH. Similarly, rh-interleukin-10 was PEGylated with succinimidyl carbonyl-PEG (SC–PEG) at pH 6.5 and 8.6. The products were assayed for stability in the presence of neutral hydroxylamine. For the pH 6.5 reaction, 50–60% of the product was cleaved by aminolysis with neutral hydroxylamine. But for the reaction at pH 8.6, only 10% was labile to this reagent. The combined results of interferon and rh-interleukin-10 suggested that the effect of reaction pH is largely unaffected by protein structure. Thus, control of reaction pH provides an effective means for rational design of PEGylated biological drugs.

*Effect of PEGylation Site on Protein Stability* Small structural changes in polypeptides can influence their physicochemical properties, biological activity, and potency. Most of the peptides or proteins are sensitive and prone to degradation by variations in chemical, thermal, and enzymatic changes leading to shorter biological half-life, poor efficacy, loss of potency, and changes in their physicochemical properties. PEGylation is widely applied to improve the resistance of polypeptides toward degradation, thus increasing their half life, improving their efficacy, reduced immunogenicity, increased solubility, and low toxicity. On PEGylation, even receptor binding affinity and distribution of polypeptides from blood to the target tissues and restricted due to increase in the molecular size by PEGylation. Therefore, proteins or peptides with selective and specific sites for PEGylation are indispensible for stable, potent, and effective therapeutic agents.

Lee et al. [157] separated, purified, characterized, and evaluated the effects of PEGylation site on the biological stability of three positional isomers of mono-PEGylated salmon calcitonins (mono-PEG-sCTs) in rat kidney homogenates. Stability analysis of sCT and mono-PEG-sCTs was carried out by radiolabeling (chloramine T method) and column switching reverse-phase HPLC. The degradation half life of these positional isomers increased in the order of the N terminus (125.5) < $Lys^{11}$ (157.3) < $Lys^{18}$ residue-modified mono-PEGylated sCT (281.5 min) over the native sCT (4.8 min). Hence, $Lys^{18}$ residue-modified mono-PEGylated sCT was found to be most stable which suggests that a selective and specific modified site on the protein molecule leads to higher stability as compared to those with many available sites for protein PEGylation.

*Effect of PEG Molecular Weight on Protein Stability* Lee et al. [61] evaluated the effect of PEGylation on in vivo and in vitro stability and biological activity of epidermal growth factor (EGF). For in vivo studies native EGF, PEG5k-EGF, and PEG2k-EGF were injected intravenously in the tail vein of mice. After a specified time both mouse kidneys were isolated, washed, and homogenized in PBS. The homogenate was centrifuged and the supernatant was analyzed for residual EGF as a function of time using a human ELISA kit. It was observed that PEGylation increased the circulation half-life about fourfold with PEG5k-EGF having the longest circulation time followed by PEG2k-EGF and native EGF. This order could be dependent on the chain length of PEG as a higher molecular weight of PEG decreased the glomular filteration rate and resisted proteolytic digesion. In this case, PEG5k-EGF and PEG2k-EGF were prepared by using PEG-propionaldehyde with molecular weights of 5000 and 2000 Da, respectively. The amount of residual protein in vitro was determined by incubating EGF and PEGylated EGF with kidney homogenates of mice for a specific time and consequently analyzing the supernatant samples. The residual amount of PEG5k-EGF and PEG2k-EGF were 88 and 50%, respectively, whereas native EGF degraded rapidly. Also, PEG5k-EGF maintained its original structure even after 10 h incubation with tissue homogenates. These results suggested higher stability of PEGylated EGF both in vivo and in vitro.

*Effects of PEG Conjugation on Insulin Properties* According to a study carried out by Hinds et al. [156], electrophilically activated derivatives of low-molecular-weight mPEG were chemically coupled to insulin, through its amino groups, at phenylalanine-B1 or lysine-B29 positions, with an amide bond being formed between the polymer and protein. The site specific attachment of mPEG to insulin significantly enhanced its resistance to aggregation. In addition, the PEGylated insulin was almost free from immunogenicity, allergenicity, and antigenicity. The modified conjugates remained in the systemic circulation for longer duration than unmodified insulin on sc administration.

*Effect of Thermal Modification on PEGylated Protein Stability* Two methods have been reported by Abuchowsky et al. to compare thermal stability of proteins and their PEGylated forms [3]. For this study, catalase was covalently attached to mPEG-1900 with 43% modification of amino groups to form PEG-1900–catalase. Similarly, catalase with PEG-5000 attached to 40% of the available amino groups formed PEG-5000–catalase. According to the first method, catalase and PEG-1900–catalase were maintained at a fixed temperature for 5 min followed by rapid cooling to room temperature and assay. The results revealed denaturation of catalase and its PEG-conjugated forms at around 55°C with PEG having no role in prevention or delay of denaturation at high temperature.

In the second method, catalase and PEG-5000–catalase were exposed to different temperature conditions, and samples from each condition were assayed by the perborate method. Maximum activity was shown by catalase at around 40°C, higher than that of PEG-5000–catalase. In this case, PEG-5000–catalase seemed to destabilize catalase structure, whereas low temperature near 0°C increased the activity of catalase and PEG-5000–catalase.

*Application of the Bioanalyzer to Stability Studies of PEGylation* In a different stability study [142], aliquots of RP–PEG5000 (recombinant protein–PEG5000), stored at −70°C for either 1 or 13 months, were compared. PEGylated product stability was preserved during the extended storage period. Both of the predominant peaks after 1 and 13 months contained 9 and 10 PEGs. Further MS data with these same samples confirmed the stability of frozen PEGylated RP. In this manner, the utility of the bioanalyzer for assessing PEG stability on the RP–PEG5000 molecule was demonstrated.

## 10 BIOPEGYLATION

### 10.1 Biosynthesis of PHA–PEG Natural Synthetic Hybrids (Polyhydroxyalkanoates)

Polyhydroxyalkanoates (PHAs) are polyesters family synthesized microbially and sequestered intracellularly under environmental stress and excess carbon conditions [158]. They are white refractile inclusion bodies that act as carbon reserves and ion sinks. A well known PHA derivative is polyhydroxybutyrate (PHB) which is a homopolymer with 3-hydroxybutyrate monomer (HBA). These are ketone bodies, which are also synthesized in mammalian cells during prolonged or periodic starvation and diabetes mellitus. Trace amount of PHB is also found in mammalian cell membranes involved in transport functions of the cells. It has been studied as a biomaterial for medical devices and failed in spite of its biocompatibility. This was due to its inherent physicochemical properties such as high crystallinity, brittle nature, and in vivo degradation tendency. Modifications with copolymers to overcome these drawbacks have been attempted successfully. Copolymerisation of PHB with hydroxyvalerate (HV) resulted in a biocompatible and biodegradable polymer, P (HB-*co*-HV), which when synthesized in different microorganisms produced an array of PHAs with different monomeric components. These provided many choices to select from brittle and crystalline to flexible and elastomeric. However, formation of different PHAs depends on the species of microorganisms as some produced PHAs with short chain length, whereas others produced medium or longer chain lengths. As an alternative, Gross et al. used PEG to increase 4-HBA units and to form novel PHAs with diverse monomeric units, considering biocompatibility and utility of PEGs in biomedical applications. PEG 200 (1–4%) was loaded in Erlenmeyer flasks with *Ralstonia eutropha* cultures to produce P (3HB-*co*-4HB). The results provided control over PHA monomeric composition with increase in the main carbon source, 4-HBA in sclPHA (short chain length polyhydroxyalkanoates). It was also observed that the lower the PEG molecular weight the greater was the PHA molecular weight reduction. Thus, PEGs also offered some control over PHA molecular weight. PEG molecules also promote PHA chain scission. The $H^1$ and $C^1$ NMR analysis of PEG-modulated PHAs confirmed PHA chain termination by PEG. This formed a new class of natural synthetic hybrid block copolymers with PHA chains "end capped" with PEG. In a review on PHA and PEG conjugation, Foster proposed the term "BioPEGylation" for such copolymerization to distinguish their biological incorporation and to relate their functional and structural similarity with

"PEGylation," the term that was coined by Harris et al. [48].

### 10.2 Applications of PEGylated PHA: BioPEGylation

BioPEGylation is up and coming field in terms of PEG–PHA hybrids with high potential to control the composition and molecular mass. It is also a simple method to control physicochemical properties of PHAs. Even crystallization properties of PHAs are also altered by PEG–PHA bioconjugation, which leads to formation of disordered microporous honeycomb films. These films have been reported to have biotechnological and biomedical applications such as biomolecular immobilization. BioPEGylation also reduce surface hydrophobicity of PHB determined by significant decrease in the contact angle. Thus, solubility of hydrophobic pharmaceutical agents can also be improved by the BioPEGylation technique.

## 11 PEGYLATION IN NOVEL DRUG DELIVERY SYSTEMS

### 11.1 PEGylation of Microspheres

Gbadamosi et al. [159] generated a heterogeneous population of surface PEGylated polystyrene microspheres [159]. The procedure followed was a two step process by which the polystyrene particles were PEGylated with mPEG-5000. In this process, the polystyrene particles were first treated with BSA and then followed by the conjugation of the nucleophiles of the adsorbed BSA monolayer with the molar excess quantities of cyanuric chloride-activated mPEG-5000. On characterizing the PEGylated particles by hydrophobic interaction chromatography, particle electrophoretic mobility was measured in a defined field. It was observed that the formulation was composed of two different entities of similar size distribution but with different surface hydrophilicity. The minor population (20%) consisted of the entities with predominant surface mPEGs in a mushroom regime. This population of the particles produced was prone to phagocytosis and hence did not exhibit prolonged circulation times in the blood. On the other hand, the major population (80%) consisting of microsphere entities having a predominant surface of mPEG molecules as high-density mushroom brush intermediate or brush configuration were most resistant to phagocytosis and poorly activated the human complement system [159].

Out of the various methods reported, therapeutic embolization is a well-estabilished method for the treatment of many vesicular lesions, including hemorrhages, arteriovenous malformations of various origins, and hypervascular tumors. This method is based on the principle of keeping the embolic region away from the blood circulation. This method has a limitation due to the development of a vascular network around the pathologic area, which has a tendency to nourish the isolated region. This revascularization is due to the presence of giant cells (macrophages). The most effective results for the obliteration of blood vessels have been obtained by using calibrated, soft, compressible, and relatively hydrophilic microspheres of controlled shape and dimensions. The nonbiodegradeable polymer used for the preparation of microspheres was derived from polyacrylamide called "Trisacryl." Madani et al. [160] PEGylated microspheres designed for embolization, defined as GF2000-Trisacryl MS (GF-MS) and DEAE-Trisacryl MS (DEAE-MS), using (3-amino propyl)triethoxy silane as the coupling agent. Indomethacin was loaded into both PEGylated and non-PEGylated DEAE-MS, displaying ion exchange ability. A complement activation test was performed in this case similarly as mentioned above. It was observed that the PEGylated GF-MS and DEAE-MS activated the complement system less than the naked microspheres. It was also observed that loading PEGylated DEAE-MS with indomethacin suppressed the complement activation. This inhibiting role showed that PEGylation as well as loading the microspheres with anti-inflammatory drug had a compact effect on the interaction of microspheres with blood proteins [161].

### 11.2 PEGylation of Liposomes

PEGylation of liposomes has provided several advantages in the development and design of novel drug delivery systems (NDDS). Liposomes without PEG were cleared rapidly from the blood and accumulated preferentially in the liver and the spleen. In a study reported by Lee et al. [162] it was observed that the modification of liposomes using PEG resulted in steric hindrance to the phagocyte system and hence prolonged blood circulation time. A reduction in radiolabeling efficiency (RE) was also observed when the glutathione method was used for the radiolabeling of liposomes. Different liposomal formulations were prepared with different extent of PEGylation, and their effect on in vivo biodistribution of liposomes in Wistar rats, and RE with technetium-(99m) ($^{99m}$Tc) was observed. Liposomes were prepared and PEGylated with egg phosphatidylcholine (egg PC, 1.85 mol%), cholesterol (1.0 mol%), and distearoylphosphatidylethanolamine-*N*-[polyethylene glycol] (DSPE–PEG). Preformed liposomes were radio-

labeled with $^{99m}$Tc-hexamethylpropylene-amine oxime [(99m)Tc-HMPAO] complexes. Sephadex G-15 column was used to check the labeling efficiency and stability. Biodistribution studies of $^{99m}$Tc-liposomes after iv injection was observed in Wistar rats [162].

Vyas et al. [163] observed that encapsulation of IFN-α in multivesicular liposomes (MVLs) and PEGylating it further affects the half-life of the IFN-α. They conjugated the IFN-α with mPEG and then purified them with size exclusion chromatography. A double emulsification technique followed by evaporation of organic solvents from chloroform ether spherules suspended in water was the method used for the preparation of PEGylated IFN-α encapsulated multivesicular liposomes. In vitro antiviral activity of PEGylated IFN-α-2a was found to be 87.9% of the unmodified IFN-α. This study, however, confirmed that multivesicular liposomes can be exploited as a long-acting or sustained-release delivery system using a combination of PEGylation and encapsulation technique for controlled delivery of IFN-α [163].

Very recently, work is going on for the development of photodynamic therapy (PDT) with a photosensitizer and Laser for the treatment of various types of cancers such as lung cancer, gastric cancer, bladder cancer, and cervical cancer. PDT is a novel approach for the destruction of cancerous cells or the malignant tissue. This approach involves the systemic or topical administration of nontoxic dyes known as photosensitisers. Then the lesion is exposed to the light in the visible range (usually red). The photosensitizers on the absorption of light produce free oxygen radicals, which are cytotoxic and thus kill malignant tissue. Sadzuka et al. [164] prepared photofrin liposomes (PF-Lip) with dimyristoylphosphatidylcholine, dimyristoylphosphatidylglycerol, and cholesterol, and they were further PEGylated (PF-PEG-Lip) to enhance the efficacy of PDT. It was observed that the trapping ratio of PF and particle size was unaffected by the PEGylation of liposomes. Phototoxicity was enhanced by liposomalization, especially PEG modification. However, it was also observed that the PF-PEG-Lip inhibited the uptake of photofrin into tumor cells. The amount of singlet oxygen from the photofrin solution (PF-sol) and each liposome was PF-PEG-Lip ≒ PF-Lip > PF-sol. PF-PEG-Lip showed a lower release level of photofrin than that of PF-Lip. Formation of a fixed aqueous layer on the liposome membrane by PEGylation made it stable in the state of PF-PEG-Lip [164].

Another study has been carried out on the PEGylated liposomal vesicles for the delivery of oligonucleotides intracellularly. Remaut et al. [165] observed that the PEGylated liposomal vesicles provided a better way to avoid the unspecific uptake by the reticulo endothelial system (RES). PEGylation of the liposomes, by incorporating PEG at the surface, avoided the unspecific uptake by the RES. PEGylation was shown to have a clear benefit on the systemic level, but its benefits on the delivery efficacy of liposomes at the intracellular level is yet to be proved. The authors compared the intracellular distribution of non-PEGylated and PEGylated liposomes, with special attention to the integrity of the oligonucleotides they were carrying. After uptake in the cells, the non-PEGylated liposomes efficiently escaped from the endosomes, thereby releasing phosphodiester oligonucleotides (PO-ONs) in the cytoplasm of the cells. The PO-ONs were, however, rapidly degraded in the intracellular environment. In contrast to non-PEGylated liposomes, PEGylated liposomes failed in protecting the PO-ONs they were carrying, leading to rapid degradation of the PO-ONs in the endosomal compartment. Furthermore, the PEG chains inhibited the endosomal escape of the degraded ONs. These intracellular findings explained the reasons for the failure of PEGylated liposomes in establishing an antisense effect.

### 11.3 PEGylation of Charged Polymer Photosensitizer Conjugate

As discussed earlier, PDT is a novel approach for the destruction of cancerous cells or malignant tissue. In another study Hamblin et al. [166] studied the conjugation of photosensitizers to macromolecular delivery vehicles that are designed to improve their performance by increasing specificity and uptake in tumors or other pathological lesions, favorably altering pharmacokinetics or biodistribution or decreasing phototoxicity to normal tissue. The authors investigated the effect of PEGylation on the uptake and phototoxicity of poly-L-lysine chlorine conjugates with either cationic or anionic charges in two types of cell lines, that is, human ovarian cancer cells and mouse macrophages. It was observed that the PEGylated cationic conjugate was less aggregated, consumed less oxygen, and had reduced cellular uptake. But it was also observed that the phototoxicity corrected for cellular uptake increased three- to fivefold. On the contrary, the anionic succinylated conjugate on PEGylation became more aggregated, consumed similar amounts of oxygen, and had higher cellular uptake. The anionic conjugate showed the highest relative phototoxicity toward both the cell lines and it decreased most toward the macrophages after PEGylation. The authors concluded that the PEGylation reduced the amount of oxygen consumed per chlorine molecule when photosensitized cells were illuminated. These in vitro studies suggested that PEGylation altered the phototoxicity of polystyrene (PS) conjugates

depending on the effect produced on the aggregation state [166].

### 11.4 PEGylation of Micelles

Various studies have been carried out to develop a micellar system for various drugs to enhance their circulation time and thereby improve their pharmacokinetic and therapeutic properties. Dhanikula and Panchagnula [167] prepared the micelles of paclitaxel by two methods. The first method involved physical incorporation of paclitaxel into poloxamer micelles wherein poloxamer 407 and paclitaxel were dissolved in chloroform. The solvent was removed by rotary evaporation at 60°C under vacuum. The resulting matrix film was hydrated to yield micelles containing paclitaxel. The other method involved an initial generation of oligomer with acid chloride groups to esterify both paclitaxel and poloxamer, leading to the formation of a conjugate. The oligomer (300–500 μL in dry acetonitrile) was transferred into a 2-mL microcentrifuge tube containing poloxamer 407 and paclitaxel dissolved in dry acetonitrile. After closing the tube to prevent access of moisture, it was left for 10–12 h at room temperature with occasional agitation for conjugation of poloxamer as well as drug with acid functional groups of oligomer. About 500 μL of water was added in order to quench the reaction and precipitate unreacted oligomer. After centrifugation to settle the precipitate, supernatant (acetonitrile–water phase) was separated and dried at 45°C for 12–15 h by centrifugal evaporation. The product thus obtained on drying was reconstituted in phosphate buffer (pH 7.4), which yielded paclitaxel-loaded micelles. It was also observed that the physical stability and payloads of colloidal carrier systems are generally low while dealing with molecules of high molecular weight such as paclitaxel. It was also observed that the PEGylation of the micelles loaded with the drug improved the aqueous solubility of the drug [167].

### 11.5 PEG-Coated Emulsions

Various studies have revealed that PEGylation of the oil-in-water emulsions prolongs the circulation time of drug entrapped in emulsions and shows enhanced accumulation in the carcinoma cells. In a study conducted by Rossi et al. [168] oil-in-water emulsions were developed (100–120 nm in diameter), and the surface properties of the emulsions were correlated with blood residence time and accumulation into neoplastic tissues by passive targeting. The authors investigated the effect of various phospholipid and sphingolipid emulsifiers, hydrogenated soyabean phosphatidylcholine (HSPC) and egg sphingomyelin (ESM), in combination with polysorbate 80 (PS-80) and 1,2-distearoyl-sn-glycero-3-phosphatidylethanolamine (DSPE)-PEG lipids of various PEG chain lengths and structures in prolonging circulation time and enhancing accumulation into B16 melanoma or C26 colon adenocarcinoma. The relationship between amphiphile molecular packing at the air–water interface on emulsion stability upon dilution in albumin and circulation longevity in vivo was also explored. PEGylation of the droplet surface with 10–15 mol% of DSPE–PEG 2000 or 5000 enhanced the circulation time of the emulsions. However, accumulation was only observed in the C26 tumor model. The tighter molecular packing observed with ESM/PS-80 monolayers at the air–water interface compared to HSPC/PS-80 correlated with improved emulsion stability in vitro, however, enhanced circulation time in vivo was not observed. The authors concluded that the understanding of the relationships between composition and performance will result in improved emulsion-based drug delivery vehicles for cancer therapy [168].

### 11.6 PEGylated Nanoparticles

Cationic polylactic acid (PLA) nanoparticles (NPs) have emerged as an important tool to replace viral vectors for gene delivery because of their biocompatibility and biodegradability. The only limitation faced by them is the fact that they are not capable of prolonging gene transfer and high transfection efficiency. Chen et al. [169] used copolymer methoxypolyethyleneglycol–PLA (MePEG–PLA), PLA, and chitosan (CS) to prepare MePEG–PLA–CS NPs and PLA–CS NPs by a diafiltration method and prepared NPs/DNA complexes through the complex coacervation of NPs with the pDNA. The authors evaluated the transfection efficiency of MePEG–PLA–CS versus PLA–CS NPs. The MePEG–PLA–CS NPs had a zeta potential of 15.7 mV at pH 7.4 and size under 100 nm, while the zeta potential of PLA–CS NPs was only 4.5 mV at pH 7.4. Electrophoretic analysis suggested that NPs could protect the DNA from nuclease degradation. The cell viability assay showed that NPs exhibited a low cytotoxicity to normal human liver cells. The pDNA being carried by MePEG–PLA–CS NPs, PLA–CS NPs, and lipofectamine could enter and express in COS7 cells. However, the transfection efficiency of MePEG–PLA–CS/DNA complexes was better than PLA–CS/DNA and lipofectamine/DNA complexes as observed by inversion fluorescence microscope and flow cytometry. Also, the transfection activity of PEGylated complexes was improved. The NPs were also tested for their ability to transport across the gastrointestinal mucosa in vivo in mice. In vivo experiments showed that MePEG–PLA–CS/DNA complexes had higher gene expression

in stomach and intestine of BALB/C mice (an albino strain of laboratory mouse from which a number of common substrains are derived) compared to PLA–CS/DNA and lipofectamine/DNA complexes. These results suggested that MePEG–PLA–CS NPs have favorable properties for nonviral gene delivery [169].

Many of the ocular diseases are due to the gene defect in the retina, and ocular gene delivery has appeared as a promising tool for the treatment of these disorders. It is well known that the blood–retina barrier and sclera prevent hydrophilic and high-MW drugs to reach the retina after systemic or topical application. Sanders et al. [170] overviewed the behavior of nonviral nucleic acid NPs (lipoplexes) in vitreous and on PEGylation strategies that improve the behavior in vitreous, without compromising with their transfection capacity. It was observed that the intravitreal injection of nonviral nucleic acid NPs was safe and a promising approach in ocular gene transfer.

Yoncheva et al. [171] performed the experiments to test the bioadhesive potential and intestinal transport of PEGylated poly(anhydride) NPs. The authors PEGylated the NPs based on poly(methyl vinyl ether-*co*-maleic anhydride) with two hydroxyl-functionalized PEGs (PEG and methoxy PEG, mPEG) and two amino-PEGs (DAE–PEG and DAP–PEG). The resulted NPs were observed to have reduced negative charge in comparison to the non-PEGylated particles when in vivo tests were performed. Higher bioadhesive potential was observed in the case of PEG-NP and DAE–PEG-NP, which was attributed to the flexibility and specific properties of the surface "brush" layer of these particles [171].

## 11.7 PEG-Based Micelles as Diagnostics

***Diacyl PEG Micelles*** A number of studies have revealed that the stability of micelles depends on the strength of van der Waal's interactions between hydrophobic blocks and the molecular size of the hydrophilic block. Use of lipid moieties as hydrophobic blocks capping PEG chains has the advantage of particle stability in comparison to the conventional amphiphilic polymer micelles. This is due to the existence of two fatty acid acyls that act as a major contributor to an increase in the hydrophobic interactions between the polymeric chains in the micelles core.

Diacyllipid-PEG conjugates were introduced as polymeric surface modifiers for liposomes. The diacyllipid-PEG molecule represents a characteristic amphiphilic polymer with a bulky hydrophilic (PEG) portion and short but extremely hydrophobic diacyllipid part, and diacyllipid-PEG conjugates were found to form micelles in an aqueous environment [172].

***Polymeric Micelles*** Polymeric micelles are a special type of micelles formed from copolymers consisting of hydrophilic as well as hydrophobic monomer units. The micelles are composed of the core of the hydrophobic chains stabilized by the corona of hydrophilic polymeric chains. The decrease of free energy of the system due to removal of hydrophobic fragments from the aqueous surroundings with the formation of micelle core stabilized with hydrophilic blocks exposed into water is the major driving force behind self-association of amphiphilic copolymers. The lower is the critical micellar concentration (CMC) value of a given amphiphilic polymer, the more stable are micelles even at low concentration of an amphiphile in the medium [172].

PEG is frequently used to build hydrophilic block(s) of micelle-forming copolymers since this polymer is water soluble and serves as an efficient stearic protector for various microparticulates (such as micelles, liposomes, nanoparticles, and nanocapsules) in biological media. Though a number of other hydrophilic polymers may be used to make corona blocks, PEG remains the hydrophilic block of choice [172].

Bae and co-workers [173] studied the in vivo behavior and biodistribution of therapeutic polymeric micelles formed by the copolymer of PEG and poly(aspartic acid) (PEG-*b*-PAA) with covalently bound adriamycin. The approach involved the covalent attachment of adriamycin to PAA of PEG-*b*-PAA (PEG–PE) copolymer resulting in the formation of an amphiphilic AB-type copolymer. This copolymer was able to form polymeric micelles with the hydrophobic block making up particle's core, while the PEG block formed a surrounding water-soluble shell. Micelles formed from this conjugate were stable in vivo and their disintegration took hours. AB-type amphiphilic block copolymers of PEG (MW of 12 kDa) and poly(aspartyl adriamycin) with 20 aspartyl adriamycin units demonstrated blood half-life in mice >5 h, indicating thus a substantial increase in micelle stability [173].

## 11.8 Use of PEGylation in Scaffold for Tissue Engineering

Hydrogels formed by photochemical-induced cross-linking contain large water contents and have mechanical properties similar to those of soft tissues. In addition, hydrogels have high permeability for oxygen, nutrients, and other water-soluble metabolites. Another advantage of photogenerated hydrogels as opposed to natural physical gels such as alginate is that their material properties can be more easily adjusted to fit the desired application.

These have emerged as a significant biomaterial in applications such as scaffolds for tissue engineering,

drug delivery carriers, thrombosis prevention devices, postoperative adhesion promoters, and coating for biosensors. The photopolymerization process allows the hydrogel to be generated in vitro or in vivo from a low viscosity solution of monomer, oligomer, or low molecular mass functional polymer (macromer) by a free radical pathway in a minimally invasive manner.

***Tyrosine PEGylated Polyether Carbonate (PEC)*** Polymeric scaffolds for tissue reconstruction, development of artificial organs, degradable membranes for the prevention of surgical adhesions, and intra-arterial coatings for the prevention of thrombosis or restenosis are some of the examples for new applications of medical implants. These advanced applications require degradable biomaterials whose physicomechanical, chemical, and biological properties can be closely matched to a wide range of narrowly defined, application-specific requirements.

Biomedical engineers and medical device designers have employed almost exclusively degradable polyesters derived of α-hydroxy acids such as poly(glycolic acid), poly(lactic acid), and polydioxanone. Among these polymers, tyrosine-derived polycarbonates have been studied most extensively and have been found to be tissue compatible, strong, tough, hydrophobic materials that degrade slowly under physiological conditions. These materials appear to be most suitable for use in orthopedic implants [174].

Other medical applications such as drug delivery, nonthrombogenic coatings, vascular grafts, scaffolds for wound healing, and artificial skin require materials that are more hydrophilic, softer, and that degrade faster than the available tyrosine-derived polycarbonates. To address these needs, tyrosine-derived diphenolic monomers were copolymerized with blocks of PEG, resulting in a new class of poly(ether carbonate)s.

Yu et al. [125] prepared tyrosine–PEG–derived poly(ether carbonate)s by condensation copolymerization with phosgene. The resulting polymers were random copolymers with average molecular weights from 40,000 to 200,000 Da. Chemical structure and purity were confirmed by NMR and FTIR spectral analysis. They established the general structure–property correlations. The glass transition temperature decreased with increasing PEG content and increasing pendent chain length. When higher molecular weight PEG blocks was used, the glass transition temperature increased relative to identical polymers having shorter PEG blocks. The tensile modulus increased with decreasing PEG content, decreasing pendent chain length, and when longer PEG blocks were used. Water uptake and the rate of backbone degradation increased with increasing PEG content [174].

***PEG–Fibrinogen Hydrogel*** Various types of biological materials and synthetic polymers are used to fabricate tissue engineering scaffolds. Biological materials provide biofunctional signals and interact well with cells, whereas, synthetic polymers provide precise control over their structural properties. A biosynthetic hybrid scaffold comprised of a fibrinogen backbone and cross-linked with difunctional PEG side chains was prepared by Almany and Seliktar [175]. Denatured fibrinogen fragments were PEGylated with PEG-diacrylates, mixed with photoinitiator, and exposed to UV light to form a hydrogel material in the presence of a cell suspension. This method of hydrogel preparation was unique because its mechanical properties were highly malleable while the biological functionality was maintained by the backbone of the polymeric network. The PEG–fibrinogen material represents one of a new generation of tissue regeneration scaffolds, which are suitable for forming three-dimensional structures with distinct biological and structural properties. The design of the material will also permit precise control over degradability, biofunctionality, and structural properties, depending on the clinical need [175]. The elastic modulus of the PEG–fibrinogen hydrogel is dependent on the molecular weight of the PEG constituent and proportional to the percent polymeric composition. Smooth muscle cells demonstrate the ability to proteolytically penetrate through the hydrogel material and form interconnecting networks of cells.

***PEG Urethane Dimethacrylate Hydrogels*** Several types of photopolymerizable hydrogels, including those derived from PEG diacrylate and dimethacrylate derivatives and PEG copolymerized with fumarate groups, have been investigated for use as biomaterials. Well characterized PEG dimethacrylates (PEGDM) and similar reactive PEG derivatives are some of the interesting model systems because, although PEG alone is bioinert, it can be easily modified to become bioactive. In addition, hydrogels from PEGDM and their copolymers and derivatives have been shown to be promising as scaffold materials.

The physical properties of hydrogels are still difficult to predict by theories due to nonidealities of gel formation. Well-defined model materials are necessary for the preparation of hydrogels with highly reproducible and easily adjustable properties [176].

***Single-Chain Glyco sFv Proteins*** Single-chain sFv proteins have become the antigen-binding format of choice because of its single polypeptide design. They normally lack asparagine-linked glycosylation. Nowadays many therapeutic agents employ the advantageous conjugation chemistry or the targeting proper-

ties provided by the glycosylation. Wang et al. [177] found the newly formed sFv proteins on evaluation in *Pichia pastoris* for glycosylation efficiency, expression level, oligosaccharide chain length and composition, and affinity. It was also observed that the engineered attachment of N-glycans conveniently near the polypeptide C terminus was found to produce the optimal results. The glycosylated sFv mannose chains could be effectively conjugated to PEG, and the resulting conjugate displayed a 10-fold increased circulating life in mice. The potential to control polymer:sFv or drug:sFv molar ratios by site-specific conjugation may substantially improve the therapeutic efficacy of these minimal antigen-binding molecules [177].

## 11.9 PEGylation in Gene Delivery

***PEGylated Adenovirus*** Gene therapy is proposed as an alternative treatment for cancers of gastrointestinal tract, especially colorectal cancer, which has become the leading cause of cancer mortality in western countries. "Adenovirus is a suitable gene delivery vector for cancer because it is capable of producing high levels of gene expression rapidly and transiently and has been used in several clinical gene therapy protocols for this purpose" [178, p. 1444]. Formulation of adenovirus significantly improved the viral transduction of differentiated intestinal epithelial cells in vitro and in vivo. Reduction of intestinal motility and distension of isolated intestinal segments with large fluid loads have also improved adenovirus transduction efficiency in vivo.

Oral administration of adenoviral vectors for the gene delivery to the intestine is not popular despite of the above mentioned improvements. This is because of the following problems associated with it:

1. Adenoviral vectors are unstable at acidic pH and significant amount of dose loss had been observed, when it comes into contact with stomach contents.
2. Enzymes present in the gastrointestinal tract (GIT) may inactivate the virus because viral capsid proteins are prone to digestion in vitro.
3. Bile salts induce aggregation of adenoviral particles making them ineffective.

PEGylation of the adenoviral vectors comes with the solution to the above mentioned problem. Recombinant adenovirus was modified with mono-mPEG activated by cyanuric chloride, succinimidyl succinate, and tresyl chloride. Transduction efficiency was assessed on Caco-2 cells. In vitro stability of viruses in simulated gastric fluid, pancreatic fluid, and bile was assessed by serial dilution on 293 cells. Oral administration of $1 \times 10^{12}$ particles of unmodified or PEGylated virus to fasted Sprague–Dawley rats determined transduction efficiency in vivo. Results have shown that the titers of unmodified virus declined to undetectable levels after 40 min in simulated gastric fluid while the infectious titer of the modified vectors did not change for 3 h. Similar results were seen with simulated pancreatic fluid. PEGylation also enhanced adenoviral transduction efficiency in Caco-2 cells by a factor of 20. PEGylation-enhanced adenovirus transduction efficiency 10- to 40-fold in vivo in intestinal segments that do not express significant amounts of adenovirus receptors (jejunum, colon) with transgene expression located in the crypt regions. Hence, the authors concluded that the PEGylated adenoviruses are suitable gene delivery vehicles for oral administration [178].

***PEGylated Vesicular Stomatitis Virus G*** Lentivirus belongs to the family of retroviruses that includes human immunodeficiency virus types 1 and 2 (HIV-1 and HIV-2), simian immunodeficiency virus, feline immunodeficiency virus, and bovine immunodeficiency and equine infectious anemia viruses. They have generated great interest as vectors for gene therapy. In terms of gene delivery, they have all the conventional attributes of other retrovirus vectors, including stable integration of the transgene into the target cell genome. One disadvantage of vesicular stomatitis virus G (VSV-G) pseudotyped lentivirus vectors for clinical application is inactivation of the vector by human serum complement. To prevent this, Croyle et al. [179] conjugated monomethoxypoly(ethylene) glycol to a VSV-G-HIV vector expressing *Escherichia coli* β-galactosidase. The modification did not affect transduction efficiency in vitro and protected the vector from inactivation in complement-active human and mouse sera. PEGylation extended the circulation half-life of active vector by a factor of 5 and reduced the rate of vector inactivation in the serum by a factor of 1000. Pharmacokinetic profiles for the total number of virus particles present in the circulation were unaffected by PEGylation. Modification of the vector with PEG significantly enhanced transduction efficiency in the bone marrow and in the spleen 14 days after systemic administration of the virus. These results, in concert with the pharmacokinetic profiles, indicated that PEGylation does protect the virus from inactivation in the serum and, as a result, improves the transduction efficiency of VSV-G pseudotyped lentivirus vectors in susceptible organs in vivo [179].

***PEGylated Antibodies*** Recently, work has been done by various scientists on the PEGylation of antibodies and antibody fragments. PEGylation of antibodies has been used to reduce immunogenicity and increase the circulation half-lives. It has also been observed that the PEGylation of antibodies improved the targeting to tumors. The site-specific attachment of PEG to antibody fragments avoids the common and substantial losses in antigen binding and may lead to further improvements by maintaining binding affinity and homogeneity of product suitable for large-scale manufacturing [180].

## 12 RESEARCH UNDER PROGRESS

### 12.1 PEG-Acetylcholinesterase

Acetylcholinesterase (AChE) is a neuroenzyme that aids in the termination of impulse transmission in cholinergic synapses by rapid hydrolysis of the neurotransmitter acetylcholine. AChE is found in cholinergic neurons (dendrites, perikarya, and axons) with high concentration at the postsynaptic end-plate of the neuromuscular junction [181]. PEGylation has been used as a tool to extend the circulatory residence time of AChE by Kronman [182]. The authors attached PEG chains to primary amines in AChE using succinimidyl propionate-activated methoxy-PEG. Briefly, purified rHuAChE (5 μM) was incubated with PEG-20000 at a ratio of 30:1 (mol/mol) $[PEG]_0/[AChE\ primary\ amines]_0$ in 50 mM borate buffer, pH 8.5, for 2 h at room temperature. The chemically modified products were dialyzed extensively against PBS and analyzed on 6% SDS/PAGE gels. This resulted in a protective shield around the enzyme, thus impeding physical interactions between the PEGylated AChE and membrane-bound protein receptors. The antigenicity studies were carried out by incubating PEGylated-HuAChE in the presence of different amounts of mouse polyclonal anti-ChE antiserum and then precipitating with protein G–Sepharose beads. Residual ChEs activity in the soluble nonbound fraction was determined. It was observed that more than 90% of the PEGylated AChE activity remained in the supernatant following immune precipitation with the same volume of antiserum, attesting to the poor interactions of PEGylated AChE with anti-AChE antibodies. Thus, it can be concluded that the efficient appendage of PEG moieties to AChE indeed provided the enzyme with a protective "shield," which prevents its immunorecognition by antibodies. Moreover, this PEG shield conferred circulatory longevity to enzyme molecules, which was confirmed by administering the PEGylated enzyme to mice and monitoring its pharmacokinetic profile by determining residual AChE activity in blood samples removed at various time points. PEGylation of AChE affected elimination processes not only by increasing the overall molecular size of the modified protein but also by masking the modified enzyme in a manner that prevented molecular interactions with other proteins. In addition to these advantages, PEGylated enzyme retained the native property of AChE that is, protection against organophosphate poisoning. To confirm this, mice challenged with more than twice the lethal dose of Soman (organophosphate) were treated with AChE and PEGylated AChE, and it was observed that animals in both the groups survived. Moreover, the circulatory long-lived PEGylated AChE can confer superior protection to mice against organophosphate poisoning. These findings established PEGylation as an effective and safe means of improving the circulatory residence time and reducing the immunogenic potential of AChE [182].

### 12.2 PEG–Interferon

Interferon (IFN), a class of cytokines displaying a wide range of antiviral, antitumor, and immunomodulatory activities, has been successfully used as a therapeutic tool for a variety of diseases such as skin malignancies, hematologic proliferative diseases, viral infections, multiple sclerosis, and hepatitis B and C [183]. Interferons have been PEGylated in order to improve the pharmacokinetic properties and minimize the loss of IFN bioactivity. PEGylation was carried out by mixing IFN and PEG at 1:3 molar ratio in 50 m*M* sodium borate buffer (pH 8.0). The reaction was stopped after 2 h by adjusting the pH of the mixture to 4.8 with phosphoric acid. The pharmacokinetics of IFN and PEG–IFN were evaluated after a single intravenous or subcutaneous injection in rats. Blood samples were withdrawn at specified time points, and the serum samples were analyzed for interferon activity by cytopathic effect (CPE) assay. It was observed that the plasma half life of PEG-IFN increased as compared to native IFN due to a decreased rate of renal clearance of PEG–IFN. The in vitro antiviral activity of PEG–IFN was determined by CPE assay using Madin–Darby bovine kidney cells (MDBK) cells challenged with vesicular stomatitis virus. The PEGylated IFN showed reduced anti-viral activity vis-à-vis native IFN, but the prolonged in vivo half-life of PEGylated IFN has overcompensated for the decreased bioactivity, eventually leading to a markedly enhanced in vivo therapeutic efficacy [183].

The elimination $t½$ of IFN is short; hence the number of injections of the molecule required per day is more. Asselah et al. [184] observed that PEGylation of IFN increases the half-life, which in turn leads to reduction in the number of daily injectons, thus greatly enhancing the patient compliance. Souvignet et al. [185] also

reported that the conventional IFN plus ribavirin therapy for treatment of hepatitis C can be rendered more efficacious by employing a combination of PEGylated IFN plus ribavirin. The results were so promising that PEG–IFN and ribavirin combination therapy is currently the recommended standard for treatment of patients with chronic hepatitis C. The recommended treatment duration is 48 weeks for hepatitis C genotype I and 24 weeks for genotypes II or III [185]. Gupta et al. [186] recommended that the dose of PEGylated interferon needs to be adjusted in patients with renal insufficiency. The multiple-dose pharmacokinetic study showed that exposure to PEG–IFN is increased in patients with renal insufficiency due to reduced clearance. Hence, the doses of the drug should be reduced by 50% in patients with severe renal insufficiency and by 25% in those with moderate insufficiency [186].

The PEGylated interferons have also been reported to nullify the risk of developing thyroid dysfunction associated with the treatment with IFN plus ribavirin. In a study carried out by Tran et al. [187] monthly thyrotropin levels of 24 patients undergoing PEG–IFN and ribavirin therapy were observed and it was concluded that there were no symptoms of hypo- or hyperthyroidism. It was concluded that the PEG–IFN and ribavirin therapy is safe in hepatitis C.

### 12.3 PEG-Asparaginase

L-Asparaginase is an important chemotherapeutic agent used for the treatment of a variety of lymphoproliferative disorders and lymphomas, acute lymphoblastic leukemia. Some of the drawbacks associated with asparaginase therapy are high immunogenicity ranging from mild allergic reactions to anaphylactic shocks and very short half-life. Narta et al. [188] tried a number of chemical modifications of asparaginase to reduce such adverse effects, but he concluded that PEGylation was the best option. Antileukemic activities of the PEGylated preparation were tested in the L5178Y tumor-bearing BDF mouse model. Conjugation of enzyme to PEG succeeded in abolishing the drug's immunogenicity. The biochemical properties of PEG-asparaginase markedly differ from the native enzyme, namely apparent molecular weight is higher, and reactivity with specific antibodies is very low but increases if the drug is subjected to freeze–thawing cycles. In addition, the plasma half-life of 20 h of the native enzyme can be successfully increased to 357 h using PEGylation approach. Due to this property, the frequency of injection of the enzyme can be reduced from once in every 2 days to once in every 2 weeks [188]. The improvement in half lives of the enzyme upon PEGylation as observed in various laboratory animals is depicted in Table 11.

**TABLE 11 Effect of PEGylation of l-Asparaginase on in Vivo Half-lives in Experimental Animal Models**

| | Half-Life (h) | |
|---|---|---|
| Animal Model | Native Enzyme | PEGylated Enzyme |
| Rats | 2.9 | 5.6 |
| Mice | 3.35 | 32.9 |
| Rabbits | 20 | 144 |

### 12.4 PEG–Calcitonin

Calcitonin is a hypocalcemic hormone that lowers calcium and phosphate concentrations in patients with hypercalcemia. This effect results from decreased bone resorption and is greater in patients in whom bone turnover rates are high. Calcitonin is effective in disorders of increased skeletal remodeling, such as Paget's disease, and in patients with osteoporosis [181]. Calcitonin is PEGylated in order to improve its enzymatic stability, in vivo hypocalcemic efficacy, and pharmacokinetics. The PEGylation is carried out by adding PEG and calcitonin to 50 m*M* phosphate buffer solution (pH 7.5) [189]. The mixture is gently shaken at room temperature for 30 min, and the reaction is quenched by adjusting the pH to around 4.0 with phosphoric acid. Youn et al. [189] examined the PEGylated enzyme for its stability in various intestinal enzymes, namely, the isolated pancreatic peptidases trypsin and chymotrypsin, the isolated brush-border peptidases leucine aminopeptidase and dipeptidyl peptidase IV, and rat intestinal homogenate. The authors observed significantly increased resistance to all intestinal enzymes applied in case of PEG–calcitonin. The degradation half-lives of PEG–calcitonin were 4.5, 2.9, and 3.1 times greater than those of free cacitonin in trypsin, chymotrypsin, and rat intestine homogenate solutions, respectively. Thus, the problem of poor systemic bioavailability of enzyme on being orally administered due to high susceptibility of enzyme to intestinal enzymatic attack was solved using a PEGylated approach. The authors further observed that liver and kidney accumulations of PEG–calcitonin at 15 and 30 min after subcutaneous injection were significantly lower than those of calcitonin by about fourfold and twofold, respectively. This, in turn, lowers the possibility of the enzyme being exposed to metabolic enzymes of liver and kidney and thus prolongs the half-life of the enzyme. Furthermore, site-specific PEGylation can preserve peptide bioactivity and maximize in vivo therapeutic efficacy. It was observed that at calcitonin equivalent doses of 100 and 200 IU/kg, total calcium decreases ($D$%) induced by PEG–calcitonin were 5.8 and 3.0 times greater than those of unmodified calcitonin, respectively. In particular, at the higher dose of

200 IU/kg, calcium levels after the intestinal administration of PEG–calcitonin decreased continuously and reached a maximum of 52% at 360 min. In contrast, unmodified calcitonin showed a maximum concentration of only 76% in plasma at 120 min [189]. Having viewed the numerous advantages of PEGylated calcitonin, Youn et al. [190] developed a method of producing a higher yield of the PEGylated product. The procedure involved conjugation of 9-fluorenylmethoxycarbonyl–calcitonin (FMOC-sCT) with PEG, and then deprotection of FMOC groups from the PEGylated FMOC-sCT. This unique site-specific PEGylation method for calcitonin using an FMOC protection/deprotection strategy produced only target isomer of PEG–calcitonin without any amino acid change in the protein, with a markedly increased production yield (86%). The findings suggested that PEGylation using FMOC-protected peptides have great potential for the mass production of site specifically PEGylated peptides with promising pharmaceutical properties [190].

## 12.5 PEG–Catalase

Catalase is a well-conserved enzyme that detoxifies hydrogen peroxide into water and oxygen [191]. The enzyme is among the most efficient known, with rates approaching 200,000 catalytic events/second/subunit (near the diffusion-controlled limit). However, the enzyme needs to be delivered to sites where the level of hydrogen peroxide increases near metastasizing tumor cells in order to achieve catalase-based inhibition/prevention of tumor metastasis. PEGylation of the enzyme was preformed with a view to prolong its retention in the blood circulation after intravenous injection [191]. This increase in blood circulation time was attributed to the altered physicochemical properties of the enzyme upon PEGylation (Table 12). The relatively higher molecular weight of PEGylated catalase made it inaccessible for the renal metabolizing enzymes, thus increasing its plasma half life, which in turn compensated for the lower enzymatic activity.

To demonstrate the antitumor activity of PEG–catalase, the authors developed clones of colon26 cells and mouse melanoma B16-BL6 cells by transfecting the firefly luciferase gene to obtain colon26/Luc and B16-BL6/Luc cells, respectively. Administration of PEG–catalase just before tumor inoculation reduced the number of B16-BL6/Luc cells in the lung 24 h after tumor inoculation. Therefore, it was concluded that PEG–catalase inhibited the adhesion of tumor cells to the endothelium in the early steps of metastasis. As a result, the ability of PEGylated catalase to inhibit pulmonary metastasis was greatly improved vis-à-vis the parent enzyme. Furthermore, the daily injection of PEG–catalase after tumor inoculation significantly increased the survival rate [191].

**TABLE 12 Physicochemical Characteristics of Catalase and PEG–Catalase**

| Physicochemical Property | Catalase | PEG–Catalase |
|---|---|---|
| Number of amino groups (modified)[a] | 112 | 26 (86) |
| Apparent molecular weight[b] | 235,000 | 665,500 |
| Enzymatic activity (%)[c] | 100 | 96.2 |
| AUC (0–24 h)[d] | 16.2 | 817.5 |

[a]The number of amino groups was determined using trinitrobenzene sulfonic acid with glycine as a standard.
[b]The apparent molecular weight was estimated by HPLC using a G4000SWXL column with the calibration curve obtained using marker proteins.
[c]The enzymatic activity was measured by monitoring the ability to degrade hydrogen peroxide.
[d]The area under the plasma concentration–time curve (AUC) up to 24 h was calculated by fitting a monoexponential equation to the experimental data of the plasma concentration of radioactivity after injection of $^{111}$In–catalase or $^{111}$In–PEG–catalase.

## 12.6 PEG–Chitosan

Chitosan, a polysaccharide derived by partial deacetylation of chitin, is widely used in pharmaceutical and medical areas, due to its favorable biological properties such as biodegradability, biocompatibility, low toxicity, hemostatic, bacteriostatic, fungistatic, anticarcinogen, and anticholesteremic properties [192, 193]. Despite these merits, the use of chitosan in formulation development is somewhat compromised due to its low water solubility [194]. Chan et al. [195] exploited PEGylation to find out the solution to this problem. To achieve this, chitosan was dissolved in acetic acid solution with gradual addition of sodium hydroxide. PEG was then added to the solution, and the reaction was conducted for 24 h at room temperature. The synthesized product was dialyzed against de-ionized water using dialysis tubing. As chitosan is widely used as a vector for gene delivery, the effect of PEGylation on the DNA binding capacity of chitosan was a major concern for the authors. So, the authors induced PEG–chitosan complexes to self-assemble by incubating DNA with the polymer. First, DNA and various amounts of chitosan or its derivative were diluted separately in 0.2 *M* sodium acetate/acetic acid buffer (pH 5.6), and an appropriate amount of polymer solution was then added to the solution of plasmid, and the resulting solution was mixed by vortex and left for 20 min at room temperature. The ability of both PEG–chitosan and chitosan to form complexes with DNA was confirmed by gel retardation assay. The

authors reported that PEGylation did not cause any alteration in the DNA binding capacity of chitosan. In addition, the presence of PEG did not add any significant cytotoxicity to chitosan, which was confirmed by estimating the relative viability of the complex in HEK 293 cells using Alamar blue assay. The fluorescence was recorded on a microplate reader with excitation wavelength at 530 nm and emission wavelength at 590 nm. Besides improving the aqueous solubility of chitosan, PEG coating around the chitosan nanocarrier systems improved their stability in the biological fluids. Consequently, the transport of bioactive macromolecules across the intestinal and nasal epithelia was facilitated [196]. The authors prepared the nanoemulsions by adding an organic phase, composed of 125 μL of miglyol, 40 mg of lecithin dissolved in 10 mL of acetone, to an aqueous phase containing Poloxamer 188 (0.25% w/v). Then, the solvents were evaporated under vacuum. Finally, 4 mL of the nanoemulsion was incubated for 1 h with 1 mL of chitosan–PEG aqueous solution (0.5% w/v) with different PEGylation degrees of chitosan (0.5 and 1%), leading to the formation of chitosan–PEG 0.5% and chitosan–PEG 1% nanocapsules, respectively. Fifty microliters of an aqueous solution of salmon calcitonin were added to the organic phase containing the oil and phospholipids. As it is well known that nanocarriers are susceptible to aggregation in media with high ionic strength, extreme pH or high enzyme/protein content, the authors investigated the effect of PEG at improving the stability of the nanoemulsions and chitosan nanocapsules in the gastrointestinal fluids. It was observed that the presence of PEG, chitosan, or chitosan–PEG attached to the oily core was found to improve the stability of the nanoemulsion, which otherwise aggregated massively upon dilution in the incubation medium. Additionally, the ability of chitosan–PEG nanocapsules in promoting the absorption of salmon calcitonin and, hence improving its pharmacological response, was studied using rats. It was observed that the $AAC_{0-24\,h}$ (area above the hypocalcemic effect curve) of the serum calcium levels–time curve after oral administration of both chitosan–PEG nanocapsules, 0.5 and 1%, was significantly higher than the nanoemulsion and salmon calcitonin aqueous solution, thus confirming the pharmacological superiorty of PEGylated chitosan carriers over untreated nanocarriers [196].

## 12.7 PEG–Cholecystokinin

Cholecystokinin (CCK), a peptide hormone secreted by the endocrine cells within the duodenal and jejunal mucosa in the gastrointestinal tract, is known to induce meal termination, delay gastric emptying, and induce satiety behavior [181]. It is widely employed for the treatment of anxiety, diarrhea, inflammatory bowel disease, and sometimes to reduce pain. However, the extremely rapid inactivation of the peptide in circulation severely hampers the unraveling of its working mechanisms. Therefore, PEGylation of CCK remains a viable and successful option to overcome this problem. The main modifications exerted by PEGylation include steric hindrance, changes in conformation, changes in electrostatic binding properties resulting in an increase in polypeptide molecular size, the reduction of its renal ultrafiltration, and the shielding of its surface from the approach of proteolytic enzymes, antibodies, or antigen-processing cells. Leon-Tamariz et al. [7] achieved PEGylation by dissolving CCK in phosphate or borate buffer and adding PEG in 1:1 or 2:1 ratio. The reaction was allowed to proceed either at 4 or 20°C under constant $N_2$ flow, and the solvents were eliminated after the completion of reaction by lyophilization [7]. The PEGylated product showed remarkable stability toward enzymatic degradation, which was evaluated by incubated CCK and PEGylated CCK in 1 mL of fresh rat plasma at 37°C. Aliquots of 20 mL were sampled every 20 s for CCK-9 and every 30 min for the conjugates. CCK-9 incubated in fresh rat plasma at 37°C showed a half-life of 2–3 min. On the other hand, HPLC chromatograms obtained from PEG–CCK conjugates, incubated under the same conditions, did not show any significant change during a 24-h period. Apart from being more stable, PEGylated CCK showed prolongation of anorectic activity too. Using rats as the animal model, the authors showed that the cumulative food intake reduction (a measure of anorectic activity), which is 1–2 h in case of nonmodified CCK, could be extended up to 6–23 h, depending upon the grade of PEG employed for PEGylation of CCK [7].

## 12.8 PEG–Antigen-Binding Fragments

Antigen-binding fragments (Fab) of antibodies can be used as therapeutics due to their monovalent antigen binding, lack of innate effector functions, and ability to be produced in a range of large-scale expression hosts such as *E. coli*. However, due to rapid renal filtration and lack of receptor-mediated recycling, Fab fragments have a very short circulating half-life in mammals. Humphreys et al. [197] applied PEGylation approach to overcome this problem. The authors di-PEGylated Fab with 20 kDa linear PEG using tris(2-carboxyethyl) phosphine (TCEP) as reductant. Reduction was carried out at room temperature for 30 min, and the product was desalted on a PD-10 desalting column (a prepacked, disposable column containing Sephadex medium for group separation of high- from low-MW substances by desalting and buffer exchange) followed by mixing with

a fivefold molar excess of PEG–maleimide over Fab. The antigen-binding capacity of Fab was not altered on PEGylation. This was confirmed by subcutaneously injecting $^{125}$I-labeled Fab–PEG and unmodified Fab in rats and analyzing the blood samples using a COBRA Autogamma counter. PEGylation conferred a serum half-life, approaching that of immunoglobulins to facilitate infrequent therapeutic dosing regimes. Apart from this, PEGylation conferred increased solubility and decreased immunogenicity, enzymatic degradation, and drug toxicity to the protein [197].

## 12.9 PEG–Glucagon

Glucagon, a 29-amino-acid single-chain peptide, is used to treat severe hypoglycemia, particularly in diabetic patients. Plasma concentrations of glucagon are elevated in poorly controlled diabetic patients. In view of its capacity to enhance gluconeogenesis and glycogenolysis, glucagon exacerbates the hyperglycemia of diabetes [181]. However, glucagon forms fibrils in aqueous solution and this property limits its therapeutic use. In a study carried out by Stignases et al. [198] glucagon was PEGylated by dissolving it in sodium hydroxide and *N*,*N*-dimethylformamide followed by addition of PEG predissolved in borate buffer [199]. The PEGylated product showed significant resistance toward fibrillation compared to glucagon. The fibrillation of glucagon and glucagon–PEG was examined using the Thioflavin T assay, a probe that interacts rapidly with amyloid fibrils inducing a strong characteristic fluorescence [199]. Apart from this, PEGylation provided conformational stability to the protein. Native glucagon has a very flexible secondary structure, which varies with the peptide concentration and solvent to give three different conformational states. In dilute aqueous solution (less than 1 mg/mL) glucagon is in a random coil with reported α-helix content of 10–15%. Structural changes can easily occur because of the low physical stability in solution of glucagon, leading to intermolecular β-sheet structure. Freeze-drying of glucagon is also known to result in an amorphous structure, which is less ordered and less crystalline-like, resulting in formation of intermolecular β sheets. Using Fourier Transform Infrared (FTIR) and circular dichroism (CD) spectroscopy as diagnostic tools, the authors remarked that PEGylation protects the internal structure of glucagon from being converted into β sheets. The authors concluded that PEGylation can be considered as a viable option to improve the stability of glucagon [199].

## 12.10 PEG–Hemoglobin

Cell-free hemoglobin (Hb) derivatives are currently being developed as promising tools for a wide range of applications, including their use as artificial oxygen carrier instead of allogenic transfusion and as oxygen transporting tools for oxygen therapeutics [200]. PEGylation of Hb is performed with a major objective of avoiding extravasation (*to force the flow of blood or lymph from a vessel out into surrounding tissue*) and to prolong the circulatory half-life of Hb. Asanuma et al. PEGylated Hb by adding activated ester of PEG-bis(succinimidyl succinate) (2.5 m*M*) very slowly to pyridoxalated Hb with stirring. To examine whether PEG–Hb increased coronary blood flow in ischemic canine heart, the authors infused PEG–Hb (10 mg/kg/min, Hb 10% solution, $n = 7$) into the coronary vein of dogs along with untreated Hb to serve as a control. It was observed that myocardial oxygen consumption in the group undergoing treatment with PEGylated Hb increased significantly vis-à-vis a control group, thus indicating that the coronary blood flow had increased [200].

In another study carried out by Li et al. [201], a new protocol, namely thiolation-mediated maleimide-chemistry-based PEGylation (extension arm facilitated PEGylation) was developed. In this process, 0.5 m*M* Hb in PBS, pH 7.4, was incubated with 5 m*M* 2-iminothilane and 10 m*M* maleimidophenyl-PEG at 4°C overnight and the reaction mixture was dialyzed extensively against PBS, pH 7.4. Purification of PEGylated Hb was carried out on a Superose 12 Prep Grade column. PEGylated Hb prepared by this method had higher oxygen affinity, a desired property for an oxygen carrier [201]. Apart from Hb, red blood cells (RBCs) have also been PEGylated with a view to increase the oxygen-carrying capacity [202]. This approach is also under investigation as a tool to generate universal red blood cells for transfusion. The authors incubated RBCs (at 5% hematocrit in the incubation buffer) with 5 m*M* dithiobis (sulfosuccinimidyl propionate) at room temperature for 1 h. After washing off the excess reagent, the cells were reincubated with 20 m*M* TCEP for 30 min in the same buffer at room temperature. The cells were washed to remove excess TCEP, incubated with 15 m*M* Mal-Phe-PEG-5000 for 2 h at room temperature, and washed again to remove excess reagents. PEGylation employing the standard blood typing test using antibodies against the respective antigens tested RBCs for blockage of the blood group antigens. Agglutination was evaluated at the macroscopic level using an agglutination viewer. It was observed that PEG chains could mask the D antigen sufficient to prevent agglutination with the respective antibodies at the microscopic level. The oxygen affinity of the PEGylated cells was measured to determine the influence of PEGylation on this property of RBCs. The authors reported significant increase in oxygen-carrying capacity of RBC on being PEGylated [202].

### 12.11 PEG–Histaminase

Diamine oxidases, originally referred as histaminases, are homodimeric copper amine oxidases which catalyze the oxidative deamination of several primary biogenic amines. For this reason, histaminases present good perspectives for therapeutic applications in myocardial ischemia treatment [203]. The major objective of PEGylation of histaminases is to reduce the high degree of immunogenicity associated with histaminase treatment and prolongation of plasma half life. Federico et al. [203] prepared PEGylated histaminase by adding linear PEG activated as $pNO_2$ phenyl formate at room temperature in small portions and under rapid stirring to diamine oxidase (0.5 mg/mL) dissolved in 0.1 *M* phosphate buffer, pH 8.5. The conjugation reaction was carried out for 4 h, while the pH was maintained with 0.05 *M* NaOH. In order to evaluate the immunological properties of either native or modified histaminase, the ability to elicit antihistaminase IgG or IgM immune response after mouse boosting with native histaminase, PEG histaminase, was measured. The authors reported that PEG derivatives elicited a slower immune response and the maximal IgM concentration reached after the fifth immunization was about 50–80% lower, when compared with the native protein. The pharmacokinetic behavior of native and PEGylated histaminase was evaluated by determining the enzymatic activity in serum after single intravenous administration to mice. It was observed that plasma half-life of PEGylated product increased and its clearance decreased vis-à-vis the native enzyme [203].

### 12.12 PEG–Horseradish Peroxidase

Horseradish peroxidase (HRP) is a versatile enzyme that finds applications in chemical, environmental, pharmaceutical, and biotechnological industries [204]. However, the nonrecovery of the enzyme in certain reactions is a retarding step in its development, which has been sorted out by PEGylation technology. A two-step process was adopted for the conjugation of PEG to HRP. First, PEG was activated with *p*-nitrophenyl chloroformate and then the activated PEG was mixed with HRP in molar ratios ranging from 1:10 to 1:500. The PEG–HRP conjugates were apparently more tolerant toward organic inactivation than the native HRP. Jeng et al. [204] reported that while the PEG–HRP conjugates exhibited 70% residual activity at acetonitrile concentrations up to at least 70%, the residual activity of native HRP droped significantly at acetonitrile concentrations above 40%. The authors further examined the capability of PEGylated HRP to catalyze the polymerization of 2-napthol. Fifteen micrograms of PEG–HRP were added to 2-napthol and dissolved in aqueous buffer (at a concentration of 0.15 *M*) containing the designated amount of acetonitrile. The authors observed that the yield of product increased significantly using PEGylated enzyme instead of native enzyme. Apart from this, the PEGylation approach allowed successful recovery of the enzyme at the end of reaction, so that it could be used in other reactions [204].

### 12.13 PEG–Glucagon-like Peptide

Glucagon-like peptide-1 (GLP-1) is a proglucagon-derived hormone secreted from enteroendocrine L cells of the intestine in response to oral ingestion which stimulates insulin secretion and inhibits glucagon secretion in a glucose-dependent manner [205]. The biological half-life of GLP-1 in plasma is only a few minutes (~2 min) due to severe proteolytic attack on being released into circulation. PEGylation can increase the therapeutic potential of PEG–glucagon-like peptide by enhancing the proteolytic resistance, diminishing glomerular filtration, thereby prolonging the biological life-time. Youn et al. [205] performed PEGylation of GLP by adding aldehyde derivative of PEG to GLP (2 mg/mL) in a 50-m*M* acetate buffer solution (pH 5.5) containing 20 m*M* sodium cyanoborohydride (molar ratio of GLP-1 to PEG = 1:2). The reaction was allowed to continue at 4°C for 2 h and was quenched by adjusting the pH to ~2.2 with 1% deionized water. The authors assessed insulinotropic activities of GLP-1 and PEGylated GLP-1 using isolated rat pancreatic islets. Twenty islets were incubated for 2 h in 2 mL of Krebs ringer bicarbonate (KRB) buffer at 37°C. The concentration of insulin released was determined using an enzyme immunoassay (EIA) kit. The results of these experiments revealed that insulinotropic activity of GLP was well preserved on being PEGylated. It was also observed that PEGylation protected the peptide from proteolytic attack. The shielding of metabolic sites due to increased streric hindrance is the underlying mechanism by which PEG protects GLP-1 from proteolytic attack. Thus it can be concluded that PEGylated GLP can be used as a viable option for treating type II diabetes due to the well-preserved insulinotropic activity, the increased proteolytic stability, and thereby the improved glucose-stabilizing capability [205].

### 12.14 PEG Islets

Insulin-dependent diabetes mellitus (IDDM) or type I diabetes mellitus is a severe form associated with ketosis in the untreated state. It occurs most commonly in the juveniles but occasionally in adults [206, 207]. Itis a catabolic disorder in which circulating insulin is

virtually absent, plasma glucagon is elevated, and the pancreatic β cells fail to respond to all insulinogenic stimuli. Exogenous insulin is required to reverse the catabolic state, prevent ketosis, and reduce hyperglucagonemia and elevated blood glucose levels [206]. One of the widely used methods, if exogenous insulin is not successful, to treat type I diabetes mellitus is pancreatic islet transplantation [208]. This treatment is based on the hypothesis that the transplanted islets would start secreting insulin and also trigger other islets to secrete it. But the greatest risk faced during any type of transplant is autoimmune rejection. In case of islets transplantation, PEGylation technology has emerged out as a viable option to combat immune response. In a process described by Lee et al. [209], the islets were first washed with Hank's balanced salt solution (HBSS) and then incubated in the HBSS (pH 8.0) containing 0.25% w/v active PEG–succinimidyl propionic acid for 1 h. Finally, the PEGylated islets were washed twice and cultured in the culture medium for one more day. In order to examine whether PEGylation reduces immunorejection of islets, the authors devised a unique procedure. A capsulotomy was performed on the caudal outer surface of the left kidney of diabetic rats and unmodified islets or PEGylated islets were injected (1200 islets/recipient). The tail vein blood glucose level of these rats, which were kept under free feeding condition, was measured daily with a portable glucometer for 100 days. Transplantation was considered successful if the blood glucose level returned to normal (<120 mg/dL) and remained normal for the first 2 days after islet transplantation. Islet rejection was diagnosed when the blood glucose level became hyperglycemic (blood glucose >200 mg/dL) for 2 consecutive days. The authors also examined whether there was any dose reducing effect of PEGylation of islets on the dose of immunosuppressant agent, cyclosporine A. It was observed that unmodified islets in all the recipients were completely rejected within 2 weeks, although cyclosporine A was continuously administered for up to 2 weeks. On the other hand, 3 out of 7 recipients of PEGylated islets survived for 100 days although cyclosporine A treatment was stopped after 2 weeks [209].

### 12.15 PEG–Lysozyme

Lysozyme (1,4-β-*N*-acetylmuramidase) is an enzyme that plays an important role in the prevention of bacterial infections. It acts by attacking a specific component of certain bacterial cell walls, peptidoglycan. Peptidoglycan is composed of the repeating amino sugars, *N*-acetylglucosamine (NAG) and *N*-acetylmuramic acid (NAM), crosslinked by peptide bridges. Lysozyme acts by hydrolyzing the bond between NAG and NAM, increasing the bacteria's permeability and causing the bacteria to burst [210]. The biggest challenge in formulation development of proteins and peptides is their tendency to aggregate. Surface-induced clustering can lead to the generation of a pool of aggregated and inactive protein. Consequently, these aggregated molecules lead to a more severe immune response. However, covalent grafting of proteins to PEG can reduce their aggregation tendencies. Lysozyme has been successfully PEGylated with a view to stabilize its formulation. Daly et al. [211] examined the effect of PEGylation on tendency of lysozyme to self-aggregate. The authors passed a mixture of di-PEGylated lysozyme along with thioflavin T through the TIRF (total internal reflection flourescence) cell which is used to study the self aggregation behavior of peptides. The authors observed that di-PEGylation of lysozyme decreased the number of protein–protein contacts in an adsorbed layer. Thus, it can be stated that PEG grafting might improve the functional stability of therapeutic proteins during manufacture, storage, anddelivery by reducing surface-induced aggregation [212].

### 12.16 PEG–Phenylalanine Ammonia-lyase

Phenylalanine ammonia-lyase (PAL) is an enzyme used for first-line treatment of phenylketonuria, a metabolic disorder that results from impaired activity of hepatic phenylalanine hydroxylase. This enzyme is responsible for the metabolism of the majority of phenylalanine intake [213]. Patients with phenylketonuria display impaired neurophysiological functioning and reduced cognitive development. Gamez et al. PEGylated PAL by coupling linear 20-kDa methoxy-PEG-SPA to PAL. The authors proposed that the uniform coverage of the protein by PEG will help to mitigate an immune response. The identified immunogenic regions of PAL can be covered through a process of site-specific mutagenesis to introduce additional lysine residues followed by PEGylation [213]. Besides reduced immunogenicity, PEGylated PAL possesses better specific activity and prolonged half-life [214].

### 12.17 PEG–Lysine

Lysine is an essential amino acid important for proper growth, and it plays an essential role in the production of carnitine, a nutrient responsible for converting fatty acids into energy and helping to lower cholesterol [181]. Lysine appears to help the body absorb and conserve calcium, and it plays an important role in the formation of collagen, a substance important for bones and connective tissues including skin, tendon, and cartilage.

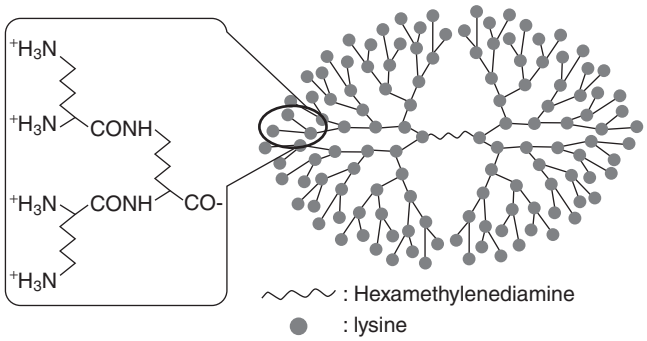


**Figure 16** Structure of dendritic poly (L-lysine) [31].

Pharmacologically, lysine is used to treat herpes, shingles, osteoporosis, and mild inflammation [181]. PEGylation has been widely employed to modify proteinaceous pharmaceuticals to reduce toxicity and immunogenicity and increase water solubility, stability, and plasma half-life. Nowadays, a variety of small and narrow sized drug carriers are utilized to deliver drugs to small sized tumors. One such carrier gaining widespread attention is lysine dendrimer (Fig. 16). Since lysine dendrimers have many reactive primary amino groups, they can be modified by several functional groups. One such modification is PEGylation, aimed at achieving sustained blood levels of drugs after intravenous administration. Okuda et al. [31] PEGylated the lysine dendrimers by dissolving them in borate buffer (pH 9.8) and reacting with PEG at 4°C. After an overnight incubation, the reaction mixture was ultrafiltrated and the purified PEGylated lysine dendrimer was then lyophilized. The authors employed mice as experimental animals to study the plasma residence time and biodistribution of PEGylated lysine dendrimers. It was reported that while the intact lysine was rapidly eliminated from the circulation within a few minutes, the retention of PEGylated lysine derivatives was enhanced by modification with PEG chains up to 1 h. The biodistribution characteristics of PEGylated lysine derivatives and intact lysine were evaluated until 60 min postintravenous injection into mice by measuring the radioactivity in the organs and urine. It was observed that there was no accumulation of PEGylated lysine in liver or kidney, thus confirming their presence in blood circulation. To further investigate whether the PEGylated lysine shows tumor tissue accumulativeness, PEG–lysine, which showed a better blood etention, was radiolabeled and was subjected to tumor-bearing mice. Although the intact lysine hardly showed tumor accumulativeness, PEG–lysine gradually and effectively accumulated in tumor tissue [31].

## 12.18 PEG–Fibroblast Growth Factor

Human acid fibroblast growth factor (rhaFGF) is a mitogenic cytokine that can stimulate mesoderm- and neuroectoderm-originated cell proliferation. However, the widespread use of haFGF is restricted by short half life and high immunogenic activity. Huang et al. [215] PEGylated rhaFGF by incubating it with 30-fold molar excess of 5-kDa PEG–maleimide in PBS at 25°C for 4 h to allow efficient reaction. The reaction mixtures were directly loaded on a Sephadex G-25 column to eliminate the free salts. The desalted PEG–rhaFGF conjugate was loaded on a CM-Sepharose column previously equilibrated with 20 m*M* phosphate buffer (pH 7.0), and eluted with buffer B (0.6 mol/L NaCl in 20 mmol/L PB, pH = 7.2). To assess the biostability of the PEGylated product, the authors incubated PEGylated rhaFGF with mouse serum under different temperatures. It was reported that compared to the native rhaFGF, PEGylated rhaFGF was resistant to incubation-decreased cell proliferating action. To assess the immunogenicity of PEGylated rhaFGF, BALB/c mice were given, by intraperitoneal (ip) injection of rhaFGF at a dose of 1.9 μmol/mouse in 50 mmol/L PBS. The PEGylated product elicited only a mild immune response compared to native peptide which was due to decreased uptake of PEGylated protein by antigen-presenting cells in vivo. Another important finding of the present study was to physiologically confirm that PEGylated rhaFGF provided a more effective protection than native rhaFGF, against infrared (IR)-induced retinal damage in the animal model. Using a rat model, PEGylated rhaFGF showed a better protection than native rhaFGF against I/R-induced retina damage, examined not only by morphological observations but also by biochemical measurement of lipid peroxidation [215].

## 12.19 PEGylated–Granulocyte-Macrophage–Colony-Stimulatinng Factor (PEG–GM–CSF)

GM–CSF belongs to a family of colony-stimulating factors that regulate proliferation and differentiation of hematopoietic cells. Recently, GM–CSF has shown promising results in treatment of Crohn's disease, a severe chronic inflammatory gastrointestinal disorder. Lysine-directed PEGylation-modified GM–CSF to address its limited half-life and rapid renal clearance. Sainathan et al. [216] performed the amine-directed PEGylation of mGM–CSF using mPEG–succinimidyl propionate MW 5000 (mPEG–SPA) [216]. GM–CSF was derivatized using a 5 : 1 molar excess of mPEG–SPA in PBS at pH 8.2 for 30 min at 25°C. To study the half-life of recombinant PEGylated GM–CSF, mice were injected intraperitoneally with a single dose (5 μg) of unmodified

recombinant GM–CSF or PEGylated GM–CSF. Blood was collected at different time points and serum GM–CSF concentration was estimated using an ELISA kit. Unmodified GM–CSF was rapidly cleared ($t_{1/2}$: 2–6 h). Animals treated with PEGylated GM–CSF had similar peak levels and a prolonged $t_{1/2}$ (>48 h) [216].

### 12.20 PEGylated–Granulocyte–Colony-Stimulating Factor (PEG–G–CSF)

G–CSF is a growth factor useful in the reatment of cancer-induced nutropenia. It is commercially synthesized in *E. coli* as recombinant methionyl human granulocyte-colony stimulating factor (r-metHuG–CSF), which is also known as Filgrastim marketed as NEUPOGEN by Amgen Inc. [217]. It is a 175-residue protein folded into a 4 helical bundle containing one cysteine residue at C17 and two disulfide bonds that stabilize the molecule at low pH. Filgrastime has a short median serum half-life of 3.5–3.8 h that lead to the development of pegfilgrastim or PEGylated r-metHuG–CSF (NEULASTA) [218]. Pegfilgrastim is an improved version of filgrastim with longer half-life, increased physical stability, higher solubility, and reduced dosing frequency. It has better patient acceptability as only one sc injection is required per chemotherapy schedule compared to the daily dosing of filgrastim. As Pegfilgrastim is a large molecule formed by covalent conjugation of 20 kDa PEG to α-amino group of N-terminal methionine group of filgrastim, it has poor renal clearance and, thus, increased mean plasma residence time [53].

### 12.21 PEG–SN38 (EZN-2208)

As a carrier, PEG has poor drug-loading property due to the presence of only two reactive groups in each polymer chain, which could be improved by imparting a dendrimer-like formation to the PEG end chain [219]. Enzon Pharmaceuticals is working on a similar strategy, where SN-38, a CPT (camptothecin) metabolite is conjugated to a four-armed 40-kDa PEG forming EZN-2208. Phase I clinical trials of iv EZN-2208 are being carried out on patients with advanced solid tumors.

### 12.22 PEGylated–Cyanovirin-N (PEG-CV-N)

CV-N is a potent virucidal bacterial protein isolated from the cyanobacterium *Nostoc ellipsosporum*. It is an entry inhibitor of HIV and permanently inactivates various clinical isolates and laboratory strains of HIV-1, HIV-2, SIV (simian immunodeficiency virus), FIV (feline immunodeficiency virus), herpes virus-6, hepatitis-C virus, influenza virus, and Ebola virus [220]. As CV-N is a bacterial protein, it presents various unwanted effects such as immunological reactions, rapid renal clearance, and inflammation of mucous membrane [221]. Thus, to overcome these unwanted effects, PEGylation was employed as the method of choice. Site-directed mutagenesis of CV-N and various mutants were tested. A mutant in which amino acid residue replacement did not interfere with its high carbohydrate binding capacity was selected [CV-N(Q62C)] for PEGylation [222]. Molecular cloning, protein expression, and purification of CV-N were performed. Purified CV-N(Q62C) was PEGylated with 20 or 30 kDa mPEG–maleimide at pH 7.2. SDS–PAGE and SEC (size exclusion chromatography) methods were used to carry out qualitative analysis and separation of PEGylated and free CV-N(Q62C). The yield, purity, and molecular weight of eluted PEG conjugates were determined by reversed-phase HPLC, gel permeation HPLC, and MALDI-TOF mass spectrometry, respectively [223]. The result of an in vitro anti-HIV assay showed that the mutant selected for PEGylation (Q62C) displayed reasonable activity and maintained efficacy equivalent to that of AZT (azidothymidine). An ELISA-type endpoint titer assay determined that immunogenicity of PEGylated–CV-N was lower then non-PEGylated form of the protein [224].

### 12.23 PEG–Methioninase (PEG–METase)

Methioninase (methionine-α-deamino-γ-mercaptomethane-lyase; METase) is a homotetramer pyridoxal-L-phosphate-dependent enzyme composed of 398 amino acids and isolated from *Psudomonas putida*. It is a low toxic enzyme with anticancer properties [225]. However, METase is immunogenic in nature and has a short half-life of 80–120 min on iv injection into the body. PEGylation of METase with succinimidylpropionate or succinimidylglutarate activated 5000 Da methoxy-PEG reduced its immunogenicity [226]. PEGylation also increased its half-life (12–30 h) and duration of methionine depletion in serum (24–72 h) depending upon PEG–methioninase molar ratios.

### 12.24 PEG–Arginine Deiminase (PEG–ADI)

Arginine deiminase (ADI) is an enzyme that degrades arginine into citrulline and ammonia. Some cancer cells cannot synthesize arginine and its depletion causes cell apoptosis [225]. Thus, arginine-depleting enzymes such as ADI are useful agents for killing arginine-auxotrophic cancer cells [227]. However, continuous low enzyme levels are required for extended periods to ensure cancer cell death, whereas ADI has high immunogenic-

ity and short serum half-life (5 h). PEGylation of recombinant ADI with PEG chain of 12,000 or 20,000 Da activated as succinimidyl carbonate decreased the immunogenicity and increased the plasma mean residence time of the enzyme up to 30-fold in mice [225]. A single iv dose of PEG–ADI (either 12,000 or 20,000 Da) reduced plasma levels of arginine below detectable limits within 1 h of administration and maintained the same for 10 days.

## REFERENCES


1. Veronese FM, Harris JM (2002). Introduction and overview of peptide and protein pegylation. *Adv Drug Deliv Rev 54*, 453–456.
2. Abuchowski A, van Es T, Palczuk NC, Davis FF (1977). Alteration of immunological properties of bovine serum albumin by covalent attachment of polyethylene glycol. *J Biol Chem 252*, 3578–3581.
3. Abuchowski A, McCoy JR, Palczuk NC, van Es T, Davis FF (1977). Effect of covalent attachment of polyethylene glycol on immunogenicity and circulating life of bovine liver catalase. *J Biol Chem 252*, 3582–3586.
4. Veronese FM, Pasut G (2005). PEGylation, successful approach to drug delivery. *Drug Discov Today 10*, 1451–1458.
5. Heller MC, Carpenter JF, Randolph TW (1999). Conformational stability of lyophilized PEGylated proteins in a phase-separating system. *J Pharm Sci 88*, 58–64.
6. Greenwald RB (2001). PEG drugs: An overview. *J Control Release 74*, 159–171.
7. Leon-Tamariz F, Verbaeys I, Van Boven M, De Cuyper M, Buyse J, Clynen E, Cokelaere M (2007). PEGylation of cholecystokinin prolongs its anorectic effect in rats. *Peptides 28*, 1003–1011.
8. Davis FF (2002). The origin of pegnology. *Adv Drug Deliv Rev 54*, 457–458.
9. Bummer PM, Koppenol S (2000). In McNally EJ, Ed. *Protein Formulation and Delivery, Drugs and the Pharmaceutical Sciences*. New York: Marcel Dekker, pp. 15–18.
10. Vyas SP, Dixit VK (2004). In *Pharmaceutical Biotechnology*. New Delhi: CBS Publishers & Distributors, pp. 565–614.
11. Sadana A (1987). Enzyme folding and unfolding. *J Biotechnol 6*, 107–133.
12. Sali A, Shakhnovich E, Karplus M (1994). Kinetics of protein folding. A lattice model study of the requirements for folding to the native state. *J Mol Biol 235*, 1614–1636.
13. Young L, Jernigan RL, Covell DG (1994). A role for surface hydrophobicity in protein-protein recognition. *Protein Sci 3*, 717–729.
14. Roberts MJ, Bentley MD, Harris JM (2002). Chemistry for peptide and protein PEGylation. *Adv Drug Deliv Rev 54*, 459–476.
15. Cleland JL, Daughtery A, Mrsny R (2001). Emerging protein delivery methods. *Curru Opinion Biotechnal 12*, 212–219.
16. Mateo C, Lombardero J, Moreno E, Morales A, Bombino G, Coloma J, Wims L, Morrison SL, Perez R (2000). Removal of amphipathic epitopes from genetically engineered antibodies: Production of modified immunoglobulins with reduced immunogenicity. *Hybridoma 19*, 463–471.
17. Lyczak JB, Morrison SL (1994). Biological and pharmacokinetic properties of a novel immunoglobulin-CD4 fusion protein. *Arch Virol 139*, 189–196.
18. Cohen S, Yoshioka T, Lucarelli M, Hwang LH, Langer R (1991). Controlled delivery systems for proteins based on poly(lactic/glycolic acid) microspheres. *Pharm Res 8*, 713–720.
19. Morhet J (2008). Formulation: Enhance Therapeutic Antibodies. Pharmaceutical Formulation & Quality. pp. 34–38. Available at *www.pharmaquality.com*
20. Chen R, Yue Z, Eccleston ME, Williams S, Slater NK (2005). Modulation of cell membrane disruption by pH-responsive pseudo-peptides through grafting with hydrophilic side chains. *J Control Release 108*, 63–72.
21. Bailon P, Palleroni A, Schaffer CA, Spence CL, Fung WJ, Porter JE, Ehrlich GK, Pan W, Xu ZX, Modi MW, Farid A, Berthold W, Graves M (2001). Rational design of a potent, long-lasting form of interferon: A 40 kDa branched polyethylene glycol-conjugated interferon alpha-2a for the treatment of hepatitis C. *Bioconjug Chem 12*, 195–202.
22. Shibata H, Yoshioka Y, Ikemizu S, Kobayashi K, Yamamoto Y, Mukai Y, Okamoto T, Taniai M, Kawamura M, Abe Y, Nakagawa S, Hayakawa T, Nagata S, Yamagata Y, Mayumi T, Kamada H, Tsutsumi Y (2004). Functionalization of tumor necrosis factor-alpha using phage display technique and PEGylation improves its antitumor therapeutic window. *Clin Cancer Res 10*, 8293–8300.
23. Yoshioka Y, Tsutsumi Y, Ikemizu S, Yamamoto Y, Shibata H, Nishibata T, Mukai Y, Okamoto T, Taniai M, Kawamura M, Abe Y, Nakagawa S, Nagata S, Yamagata Y, Mayumi T (2004). Optimal site-specific PEGylation of mutant TNF-alpha improves its antitumor potency. *Biochem Biophys Res Commun 315*, 808–814.
24. Li YP, Pei YY, Ding J, Shen ZM, Zhang XY, Gu ZH, Zhou JJ (2001). PEGylated recombinant human tumor necrosis factor alpha: Preparation and anti-tumor potency. *Acta Pharmacol Sin 22*, 549–555.
25. Greenwald RB, Yang K, Zhao H, Conover CD, Lee S, Filpula D (2003). Controlled release of proteins from their poly(ethylene glycol) conjugates: Drug delivery systems employing 1,6-elimination. *Bioconjug Chem 14*, 395–403.

26. Davis FF, Kazo GM, Nucci ML, Abuchowski A (1990). In Lee VHL, Ed. *Peptide and Protein Drug Delivery*. New York: Marcel Dekker.
27. Pasut G (2004). Protein, peptide and non-peptide drug PEGylation fot therapeutic application: A review. *Exp Opin Ther Patents 14*, 859–894.
28. Maeda H, Wu J, Sawa T, Matsumura Y, Hori K (2000). Tumor vascular permeability and the EPR effect in macromolecular therapeutics: A review. *J Control Release 65*, 271–284.
29. Choe YH, Conover CD, Wu D, Royzen M, Gervacio Y, Borowski V, Mehlig M, Greenwald RB (2002). Anticancer drug delivery systems: Multi-loaded N4-acyl poly(ethylene glycol) prodrugs of ara-C. II. Efficacy in ascites and solid tumors. *J Control Release 79*, 55–70.
30. James ND, Coker RJ, Tomlinson D, Harris JR, Gompels M, Pinching AJ, Stewart JS (1994). Liposomal doxorubicin (Doxil): An effective new treatment for Kaposi's sarcoma in AIDS. *Clin Oncol (R Coll Radiol) 6*, 294–296.
31. Okuda T, Kawakami S, Akimoto N, Niidome T, Yamashita F, Hashida M (2006). PEGylated lysine dendrimers for tumor-selective targeting after intravenous injection in tumor-bearing mice. *J Control Release 116*, 330–336.
32. Evens RP (2007). In Swarbrick J, Ed. *Encyclopedia of Pharmaceutical Technology*. New York: Informa Healthcare, pp. 258–280.
33. Kushibiki T, Tabata Y (2005). Preparation of poly(ethylene glycol)-introduced cationized gelatin as a non-viral gene carrier. *J Biomater Sci Polym Ed 16*, 1447–1461.
34. Kushibiki T, Matsuoka H, Tabata Y (2004). Synthesis and physical characterization of poly(ethylene glycol)-gelatin conjugates. *Biomacromolecules 5*, 202–208.
35. Vittaz M, Bazile D, Spenlehauer G, Verrecchia T, Veillard M, Puisieux F, Labarre D (1996). Effect of PEO surface density on long-circulating PLA-PEO nanoparticles which are very low complement activators. *Biomaterials 17*, 1575–1581.
36. Wu GY, Wu CH (1987). Receptor-mediated in vitro gene transformation by a soluble DNA carrier system. *J Biol Chem 262*, 4429–4432.
37. Boussif O, Lezoualc'h F, Zanta MA, Mergny MD, Scherman D, Demeneix B, Behr JP (1995). A versatile vector for gene and oligonucleotide transfer into cells in culture and in vivo: *Polyethylenimine. Proc Natl Acad Sci USA 92*, 7297–7301.
38. Pack DW, Hoffman AS, Pun S, Stayton PS (2005). Design and development of polymers for gene delivery. *Nat Rev Drug Discov 4*, 581–593.
39. Kabanov AV, Astafieva IV, Maksimova IV, Lukanidin EM, Georgiev GP, Kabanov VA (1993). Efficient transformation of mammalian cells using DNA interpolyelectrolyte complexes with carbon chain polycations. *Bioconjug Chem 4*, 448–454.
40. Harada-Shiba M, Yamauchi K, Harada A, Takamisawa I, Shimokado K, Kataoka K (2002). Polyion complex micelles as vectors in gene therapy—Pharmacokinetics and in vivo gene transfer. *Gene Ther 9*, 407–414.
41. Kakizawa Y, Kataoka K (2002). Block copolymer micelles for delivery of gene and related compounds. *Adv Drug Deliv Rev 54*, 203–222.
42. Itaka K, Harada A, Nakamura K, Kawaguchi H, Kataoka K (2002). Evaluation by fluorescence resonance energy transfer of the stability of nonviral gene delivery vectors under physiological conditions. *Biomacromolecules 3*, 841–845.
43. Itaka K, Yamauchi K, Harada A, Nakamura K, Kawaguchi H, Kataoka K (2003). Polyion complex micelles from plasmid DNA and poly(ethylene glycol)-poly(L-lysine) block copolymer as serum-tolerable polyplex system: Physicochemical properties of micelles relevant to gene transfection efficiency. *Biomaterials 24*, 4495–4506.
44. Miyata K, Kakizawa Y, Nishiyama N, Yamasaki Y, Watanabe T, Kohara M, Kataoka K (2005). Freeze-dried formulations for in vivo gene delivery of PEGylated polyplex micelles with disulfide crosslinked cores to the liver. *J Control Release 109*, 15–23.
45. Han M, Bae Y, Nishiyama N, Miyata K, Oba M, Kataoka K (2007). Transfection study using multicellular tumor spheroids for screening non-viral polymeric gene vectors with low cytotoxicity and high transfection efficiencies. *J Control Release 121*, 38–48.
46. Powell GM (1980). In *Handbook of Water Soluble Gums and Resins*. New York: Mc-Graw-Hill, pp. 18.11–18.31.
47. Price JC (1994). In Wade A, Weller PJ, Eds. *Handbook of Pharmaceutical Excipients*. American Pharmaceutical Association, Washington, and The Pharmaceutical Press, London. pp. 355–361.
48. Harris JM, Chess RB (2003). Effect of pegylation on pharmaceuticals. *Nat Rev Drug Discov 2*, 214–221.
49. Dhawan S, Varma M, Sinha VR (2005). High molecular weight poly(ethylene oxide)-based drug delivery systems. *Pharm Technol 29*(5): 72–80.
50. Braun DB (1980). In Davidson RL, Ed. *Handbook of Water Soluble Gums and Resins*. New York: Mc-Graw-Hill, pp. 19.11–19.33.
51. Dhawan S, Dhawan K, Varma M, Sinha VR (2005). Applications of poly(ethylene oxide) in drug delivery systems. *Pharm Technol 29*, 82–92.
52. Veronese FM (2001). Peptide and protein PEGylation: A review of problems and solutions. *Biomaterials 22*, 405–417.
53. Kinstler OB, Brems DN, Lauren SL, Paige AG, Hamburger JB, Treuheit MJ (1996). Characterization and stability of N-terminally PEGylated rhG-CSF. *Pharm Res 13*, 996–1002.
54. Delgado C, Patel JN, Francis GE, Fisher D (1990). Coupling of poly(ethylene glycol) to albumin under very mild conditions by activation with tresyl chloride: Characterization of the conjugate by partitioning in aqueous two-phase systems. *Biotechnol Appl Biochem 12*, 119–128.

55. Sperinde JJ, Martens BD, Griffith LG (1999). Tresyl-mediated synthesis: Kinetics of competing coupling and hydrolysis reactions as a function of pH, temperature, and steric factors. *Bioconjug Chem 10*, 213–220.
56. Beauchamp CO, Gonias SL, Menapace DP, Pizzo SV (1983). A new procedure for the synthesis of polyethylene glycol-protein adducts; effects on function, receptor recognition, and clearance of superoxide dismutase, lactoferrin, and alpha 2-macroglobulin. *Anal Biochem 131*, 25–33.
57. Kozlowski A, Harris JM (2001). Improvements in protein PEGylation: Pegylated interferons for treatment of hepatitis C. *J Control Release 72*, 217–224.
58. Veronese FM, Sacca B, Schiavon O, Caliceti P, Orsatti L, Orschin P (1999). PEG peptide protein pegylation: A procedure to identify the pegylation site. *Proc Int Symp Control Bioact Mater 26*, 106–107.
59. Monkarsh SP, Ma Y, Aglione A, Bailon P, Ciolek D, DeBarbieri B, Graves MC, Hollfelder K, Michel H, Palleroni A, Porter JE, Russoman E, Roy S, Pan YC (1997). Positional isomers of monopegylated interferon alpha-2a: Isolation, characterization and biological activity. *Anal Biochem 247*, 434–440.
60. Sartore L, Caliceti P, Schiavon O, Monfardini C, Veronese FM (1991). Accurate evaluation method of the polymer content in monomethoxy(polyethylene glycol) modified proteins based on amino acid analysis. *Appl Biochem Biotechnol 31*, 213–222.
61. Lee H, Jang IH, Ryu SH, Park TG (2003). N-terminal site-specific mono-PEGylation of epidermal growth factor. *Pharm Res 20*, 818–825.
62. Na DH, Park EJ, Youn YS, Moon BW, Jo YW, Lee SH, Kim WB, Sohn Y, Lee KC (2004). Sodium dodecyl sulfate-capillary gel electrophoresis of polyethylene glycolated interferon alpha. *Electrophoresis 25*, 476–479.
63. Kopecky EM, Greinstetter S, Pabinger I, Buchacher A, Romisch J, Jungbauer A (2006). Effect of oriented or random PEGylation on bioactivity of a factor VIII inhibitor blocking peptide. *Biotechnol Bioeng 93*, 647–655.
64. Gaertner HF, Offord RE, Cotton R, Timms D, Camble R, Rose K (1994). Chemo-enzymic backbone engineering of proteins. Site-specific incorporation of synthetic peptides that mimic the 64-74 disulfide loop of granulocyte colony-stimulating factor. *J Biol Chem 269*, 7224–7230.
65. Gaertner HF, Rose K, Cotton R, Timms D, Camble R, Offord RE (1992). Construction of protein analogues by site-specific condensation of unprotected fragments. *Bioconjug Chem 3*, 262–268.
66. Gaertner HF, Offord RE (1996). Site-specific attachment of functionalized poly(ethylene glycol) to the amino terminus of proteins. *Bioconjug Chem 7*, 38–44.
67. Balan S, Choi, J-W, Godwin A, Teo I, Laborde CM, Heidelberger S, Zloh M, Shaunak S, Brocchini S (2007). Site-specific PEGylation of protein bisulfide bonds using a three-carbon bridge. *Bioconjug Chem 18*, 61–76.
68. Youn YS, Jung JY, Oh SH, Yoo SD, Lee KC (2006). Improved intestinal delivery of salmon calcitonin by Lys18-amine specific PEGylation: Stability, permeability, pharmacokinetic behavior and in vivo hypocalcemic efficacy. *J Control Release 114*, 334–342.
69. Sato H (2002). Enzymatic procedure for site specific PEGylation of proteins. *Adv Drug Deliv Rev 54*, 487–504.
70. Kogan TP (1992). The synthesis of substituted methoxy-poly-(ethylene glycol) derivatives suitable for selective protein modification. *Synth Commun 22*, 2417–2424.
71. Zalipsky S, Albericio F, Slomczynska U, Barany G (1987). A convenient general method for synthesis of *N* alpha- or *N* omega-dithiasuccinoyl (Dts) amino acids and dipeptides: Application of polyethylene glycol as a carrier for functional purification. *Int J Pept Protein Res 30*, 740–783.
72. Zalipsky S (1995). Functionalized poly (ethylene glycol) for preparation of biologically relevant conjugates. *Bioconjug Chem 6*, 150–165.
73. Clark R, Olson K, Fuh G, Marian M, Mortensen D, Teshima G, Chang S, Chu H, Mukku V, Canova-Davis E, Somers T, Cronin M, Winkler M, Wells JA (1996). Long-acting growth hormones produced by conjugation with polyethylene glycol. *J Biol Chem 271*, 21969–21977.
74. Bentley MD, Roberts MJ, Harris JM (1998). Reductive amination using poly(ethylene glycol) acetaldehyde hydrate generated in situ: Applications to chitosan and lysozyme. *J Pharm Sci 87*, 1446–1449.
75. Katre NV (1993). The conjugation of proteins with polyethylene glycol and other polymers: Altering properties of proteins to enhance their therapeutic potential. *Adv Drug Deliv Rev 10*, 91–114.
76. Zalipsky S, Lee CM (1992). Use of functionalized poly (ethylene glycol)s for modification of polypeptides. In: J.M. Harris, S. Zalipsky (Eds.), *Polyethylene Glycol Chemistry, Biotechnical and Biomedical Applications*, New York: Plenum, pp. 347–370.
77. Dust JM, Fang ZH, Harris JM (1990). Proton NMR characteristics of polyethylene glycol and derivatives. *Macromolecules 23*, 3742–3746.
78. Zalipsky S, Seltzer R, Nho K (1991). In Dunn RL, Ottenbrite RM, Eds. *Polymeric Drugs and Drug Delivery Systems*. Washington, DC: American Chemical Society, pp. 91–100.
79. Walsh G (2004). Second-generation biopharmaceuticals. *Eur J Pharm Biopharm 58*, 185–196.
80. Bentley MD, Harris JM (1999). Polyethylene glycol aldehyde hydrates and related polymers and applications in modifying. U.S. Patent 5,990,237.
81. Harris JM, Herati RM (1993). Preparation and use of polyethylene glycol propionaldehyde. U.S. Patent. 5,252,714.
82. Zalipsky S, Barany G (1986). Preparation of polyethylene glycol derivatives with two different functional groups at the termini. *Polym Preprints 27*, 1–2.

83. Harris JM, Kozlowski A (1997). Poly (ethylene glycol) and related polymers monosubstituted with propionic or butanoic acids and functional derivatives thereof for biotechnical applications, US Patent 5,672,662.
84. Kuan CT, Wang QC, Pastan I (1994). Pseudomonas exotoxin A mutants. Replacement of surface exposed residues in domain II with cysteine residues that can be modified with polyethylene glycol in a site-specific manner. *J Biol Chem 269*, 7610–7616.
85. El Tayar N, Roberts MJ, Harris JM, Sawlivich W (1999). Polyol-IFN-β-conjugates, WO99/55377.
86. Karpusas M, Nolte M, Benton CB, Meier MA, Lipscomb WN, Goelz S (1997). The crystal sructure of human interferon beta at 2.2-A resolution. *Proc Natl Acad Sci USA 94*, 11813–11818.
87. Zalipsky S, Menon-Rudolph S (1997). In Harris JM, Zalipsky S, Eds. *Poly(ethylene glycol) Chemistry and Biological Applications*. Washington, DC: American Chemical Society, pp. 318–340.
88. Goodson RJ, Katre NV (1990). Site-directed pegylation of recombinant interleukin-2 at its glycosylation site. *Biotechnology (NY) 8*, 343–346.
89. Peleg-Shulman T, Tsubery H, Mironchik M, Fridkin M, Schreiber G, Shechter Y (2004). Reversible PEGylation: A novel technology to release native interferon alpha2 over a prolonged time period. *J Med Chem 47*, 4897–4904.
90. Garman AJ, Kalindjian SB (1987). The preparation and properties of novel reversible polymer-protein conjugates. 2-omega-Methoxypolyethylene (5000) glycoxymethylene-3-methylmaleyl conjugates of plasminogen activators. *FEBS Lett 223*, 361–365.
91. Zalipsky S, Qazen M, Walker JA 2nd, Mullah N, Quinn YP, Huang SK (1999). New detachable poly(ethylene glycol) conjugates: Cysteine-cleavable lipopolymers regenerating natural phospholipid, diacyl phosphatidylethanolamine. *Bioconjug Chem 10*, 703–707.
92. Shechter Y, Mironchik M, Saul A, Gershonov E, Precido-Patt L, Sasson K, Tsubery H, Mester B, Kapitkovsky A, Rubinraut S, Vachutinski Y, Fridkin G, Fridkin M (2007). New technologies to prolong life-time of peptide and protein drugs in vivo. *Int J Peptide Res Ther 13*, 105–117.
93. Shechter Y, Tsubery H, Fridkin M (2003). [2-Sulfo-9-fluorenylmethoxycarbonyl]3-exendin-4-a long-acting glucose-lowering prodrug. *Biochem Biophys Res Commun 305*, 386–391.
94. Shechter Y, Tsubery H, Mironchik M, Rubinstein M, Fridkin M (2005). Reversible PEGylation of peptide $YY_{3-36}$ prolongs its inhibition of food intake in mice. *FEBS Lett 579*, 2439–2444.
95. Tsunoda S, Ishikawa T, Yamamoto Y, Kamada H, Koizumi K, Matsui J, Tsutsumi Y, Hirano T, Mayumi T (1999). Enhanced antitumor potency of polyethylene glycolylated tumor necrosis factor-alpha: A novel polymer-conjugation technique with a reversible amino-protective reagent. *J Pharmacol Exp Ther 290*, 368–372.
96. Bentley MD, Harris JM, Kozlowski A (2001). Heterobifunctional polyethylene glycol derivatives and methods for their preparation. In *WO 01/26692 A1*.
97. Yokoyama M, Okano T, Sakurai Y, Kikuchi A, Ohsako N, Nagasaki Y, Kataoka K (1992). Synthesis of poly(ethylene oxide) with heterobifunctional reactive groups at its terminals by an anionic initiator. *Bioconjug Chem 3*, 275–276.
98. Allen TM, Brandeis E, Hansen CB, Kao GY, Zalipsky S (1995). A new strategy for attachment of antibodies to sterically stabilized liposomes resulting in efficient targeting to cancer cell. *Biochim Biophys Acta 1237*, 99–108
99. Buckmann A, Morr M, Johansson G (1981). Functionalization of poly(ethylene glycol) and monomethoxy-poly(ethylene glycol). *Makromol Chem 182*, 1379–1384.
100. Johansson G (1986). Effects of poly(ethylene glycol) bound alcohols and amines on the partition of albumins and thylakoid membranes in an aqueous two phase system. *J Chromatogr 368*, 309–317.
101. Hemmasi B, Bayer E (1977). Use of 1-hydroxybenzotriazole in peptide synthesis. *Tetrahedron Lett 19*, 1599–1602.
102. Zalipsky S, Lee CM (1992). In Harris JM, Zalipsky S, Eds. *Poly(ethylene glycol) Chemistry, Biotechnical and Biomedical Applications*. New York: Plenum, pp. 347–370.
103. Blume G, Cevc G, Grommelin MDJA, Bakker-Woudenberg IAJM, Kluft C, Storm G (1993). Specific targeting with poly(ethylene glycol) modified liposomes: Coupling of homing devices to the ends of the polymeric chains combines effective target binding with long circulation times. *Biochim Biophys Acta 1149*, 180–184.
104. Kynclova E, Elsner E, Kopf A, Hawa G, Schalkhammer T, Pittner F (1996). Novel method for coupling of poly(ethyleneglycol) to carboxylic acid moieties of proteins. *J Mol Recognit 9*, 644–651.
105. Glass JD, Miller R, Wesolowski G (1981). A stable, water soluble polymer derivative which specifically and reversibly attaches to the side chain of arginine residues. In *Peptides: Structure and Function, Proceedings of the Seventh American Peptide Symposium*. Rockford, IL: Pierce Chemical Company.
106. Atassi MZ (1977). In Atassi MZ, Ed. *Immunochemistry of Proteins*. New York: Plenum, pp. 1–161.
107. Zalipsky S (1995). Chemistry of poly ethylene glycol conjugates with biologically active molecules. *Adv Drug Deliv Rev 16*, 157–182.
108. Zalipsky S, Gilon C, Zilkha A (1983). Attachment of drugs to polyethylene glycols. *Eur Polym J 19*, 1177.
109. Zalipsky S, Gilon C, Zilkha A (1984). Esterification of polyethylene glycols. *J Macromol Sci Chem A21*, 839–845.
110. Khue NV, Galin JC (1985). Antiinflammatory polymer bound steroids for topical applications I. Synthesis and characterization. *J Appl Polym Sci 30*, 2761–2778.

111. Khue NV, Coupin LJG, Poindron P (1986). Antiinflammatory polymer bound steroids for topical applications II. Controlled release of the steroids. *J Polym Sci Polym Chem 24*, 359–373.

112. Sano A, Maeda H, Kai Y, Ono K (1989). Polyethylene glycol derivatives, modified peptides and production thereof. European Patent Application 0340741.

113. Frokjaer S, Otzen DE (2005). Protein drug stability: A formulation challenge. *Nat Rev Drug Discov 4*, 298–306.

114. Izutsu K, Yoshioka S, Takeda Y (1991). The effects of additives on the stability of freeze dried beta-galactosidase stored at elevated temperature. *Int J Pharm 71*, 137–146.

115. Liu R, Langer R, Klibanov AM (1991). Moisture induced aggregation of lyophilised proteins in the solid state. *Biotechnol Bioeng 37*, 177–184.

116. Carpenter JF, Crowe JH (1989). An infrared spectroscopic study of the interactions of carbohydrates with dried proteins. *Biochemistry 28*, 3916–3922.

117. Mosharraf M, Malmberg M, Fransson J (2007). Formulation, lyophilization and solid-state properties of a pegylated protein. *Int J Pharm 336*, 215–232.

118. Tattini V Jr, Parra DF, Polakiewicz B, Pitombo RN (2005). Effect of lyophilization on the structure and phase changes of PEGylated-bovine serum albumin. *Int J Pharm 304*, 124–134.

119. Hinrichs WL, Sanders NN, De Smedt SC, Demeester J, Frijlink HW (2005). Inulin is a promising cryo- and lyoprotectant for PEGylated lipolexes. *J Control Release 103*, 465–479.

120. Hinrichs WL, Mancenido FA, Sanders NN, Braeckmans K, De Smedt SC, Demeester J, Frijlink HW (2006). The choice of a suitable oligosaccharide to prevent aggregation of PEGylated nanoparticles during freeze thawing and freeze drying. *Int J Pharm 311*, 237–244.

121. Strambini GB, Gabellieri E (1996). Proteins in frozen solutions: Evidence of ice-induced partial unfolding. *Biophys J 70*, 971–976.

122. Carpenter JF, Prestrelski SJ, Arakawa T (1993). Separation of freezing- and drying-induced denaturation of lyophilized proteins using stress-specific stabilization. I. Enzyme activity and calorimetric studies. *Arch Biochem Biophys 303*, 456–464.

123. Sebhatu T, Angberg M, Ahlneck C (1994). Assessment of the degree of disorder in crystalline solids by isothermal microcalorimetry. *Int J Pharm 104*, 135–144.

124. Buckton G, Darcy P (1999). Assessment of disorder in crystalline powders—A review of analytical techniques and their application. *Int J Pharm 179*, 141–158.

125. Yu P, Zhang G, Bi J, Lu X, Wang Y, Su Z (2007). Facile purification of mono-PEGylated interleukin-1 receptor antagonist and its characterization with multi-angle laser light scattering. *Process Biochem 42*, 971–977.

126. Wylie DC, Voloch M, Lee S, Liu YH, Cannon-Carlson S, Cutler C, Pramanik B (2001). Carboxyalkylated histidine is a pH-dependent product of pegylation with SC-PEG. *Pharm Res 18*, 1354–1360.

127. Peracchia MT, Vauthier C, Puisieux F, Couvreur P (1997). Development of sterically stabilized poly(isobutyl 2-cyanoacrylate) nanoparticles by chemical coupling of poly(ethylene glycol). *J Biomed Mater Res 34*, 317–326.

128. Carstensen H, Muller BW, Muller RH (1991). Adsorption of ethoxylated surfactants on nanoparticles. I. Characterization by hydrophobic interaction chromatography. *Int J Pharmacol 67*, 29–37.

129. Stoscheck CM (1990). Quantitation of protein. *Methods Enz 182*, 50–69.

130. Gardiner C, Mackie IJ, Machin SJ (1998). Simultaneous assay of free and total protein S by ELISA using monoclonal and polyclonal antibodies. *Clin Lab Haematol 20*, 41–45.

131. Rinaldi S, Kaaks R, Zeleniuch-Jacquotte A, Arslan AA, Shore RE, Koenig KL, Dossus L, Riboli E, Stattin P, Lukanova A, Toniolo P (2005). Insulin-like growth factor-I, IGF binding protein-3, and breast cancer in young women: A comparison of risk estimates using different peptide assays. *Cancer Epidemiol Biomarkers Prev 14*, 48–52.

132. Bradford MM (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal Biochem 7*, 248–254.

133. Hjelmeland LM, Chrambach A (1981). Electrophoresis and electrofusing in detergent containing media: A discussion of basic concepts. *Electrophoresis 2*, 1–11.

134. Jones AJ (1993). Analysis of polypeptides and proteins. *Adv Drug Deliv Rev 10*, 29–90.

135. Laemmli UK (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage $T_4$. *Nature 227*, 680–685.

136. Kurfürst MM (1992). Detection and molecular weight determination of polyethylene glycol-modified hirudin by staining after sodium dodecyl sulfate-polyacrylamide gel electrophoresis. *Anal Biochem 200*, 244–248.

137. Monkarsh SP, Ma Y, Aglione A, Bailon P, Ciolek D, DeBarbieri B, Graves MC, Hollfelder K, Michel H, Palleroni A, Porter JE, Russoman E, Roy S, Pan YC (1997). Positional isomers of monopegylated interferon alpha-2a: Isolation, characterization, and biological activity. *Anal Biochem 247*, 434–440.

138. Bullock J, Chowdhury S, Severdia A, Sweeney J, Johnston D, Pachla L (1997). Comparison of results of various methods used to determine the extent of modification of methoxy polyethylene glycol 5000-modified bovine cupri-zinc superoxide dismutase. *Anal Biochem 254*, 254–262.

139. Kendrick BS, Kerwin BA, Chang BS, Philo JS (2001). Online size-exclusion high performance liquid chromatography light scattering and differential refractometry methods to determine degree of polymer conjugation to proteins and protein-protein or protein-ligand association states. *Anal Biochem 299*, 136–146.

140. Gong XW, Wei DZ, He ML, Xiong YC (2007). Discarded free PEG-based assay for obtaining the modification extent of pegylated proteins. *Talanta 71*, 381–384.
141. Habeeb AF (1996). Determination of free amino groups in proteins by trinitrobenzenesulfonic acid. *Anal Biochem 14*, 328–336.
142. Yu T, Traina JA, Erno Pungor J, McCaman M (2006). Precise and comparative pegylation analysis by microfluidics and mass spectrometry. *Anal Biochem 359*, 54–62.
143. Bullock J, Chowdhury S, Johnston D (1996). Characterization of poly(ethylene glycol) modified superoxide dismutase. *Anal Chem 68*, 3258–3263.
144. Na DH, Lee KC (2004). Capillary electrophoretic characterization of PEGylated human parathyroid hormone with matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Anal Biochem 331*, 322–328.
145. Youn YS, Na DH, Yoo SD, Song SC, Lee KC (2004). Chromatographic separation and mass spectrometric identification of positional isomers of polyethylene glycol-modified growth hormone-releasing factor (1–29). *J Chromatogr A 1061*, 45–49.
146. Sims GEC, Snape TJ (1980). A method for the estimation of polyethylene glycol in plasma protein fractions, *Anal Biochem 107*, 60–63.
147. Mosharraf M, Nystrom C (2003). Apparent solubility of drugs in partially crystalline systems. *Drug Dev Ind Pharm 29*, 603–622.
148. Park EJ, Na DH, Lee KC (2007). In vitro release study of mono-PEGylated growth hormone-releasing peptide-6 from PLGA microspheres. *Int J Pharm 343*, 281–283.
149. Harris JM, Martin NE, Modi M (2001). Pegylation: A novel process for modifying pharmacokinetics. *Clin Pharmacokinet 40*, 539–551.
150. Lee SH, Lee S, Youn YS, Na DH, Chae SY, Byun Y, Lee KC (2005). Synthesis, characterization, and pharmacokinetic studies of PEGylated glucagon-like peptide-1. *Bioconjug Chem 16*, 377–382.
151. Owens DE 3rd, Peppas NA (2006). Opsonization, biodistribution, and pharmacokinetics of polymeric nanoparticles. *Int J Pharm 307*, 93–102.
152. Pasta P, Riva S, Carrea G (1988). Circular dichroism and fluorescence of polyethylene glycol-subtilisin in organic solvents. *FEBS Lett 236*, 329–332.
153. Shami EY, Rothstein A, Ramjeesingh M (1989). Stabilization of enzymes by their specific antibodies, *Trends Biotechnol 7*, 186–190.
154. Christakopoulos P, Kourentzi E, Hatzinikolaou DG, Claeyssens M, Kekos D, Macris BJ (1998). Enhancement of pH-stability of a low molecular mass endoglucanase from *Fusarium oxysporum* by protein PEGylation. *Carbohydr Res 314*, 95–99.
155. Park JW, Kajiuchi T (1995). Development of effective modified cellulase for cellulose hydrolysis process, *Biotechnol Bioeng 45*, 366–372.
156. Hinds KD, Kim SW (2002). Effects of PEG conjugation on insulin properties. *Adv Drug Deliv Rev 54*, 505–530.
157. Lee KC, Moon SC, Park MO, Lee JT, Na DH, Yoo SD, Lee HS, DeLuca PP (1999). Isolation, characterization, and stability of positional isomers of mono-PEGylated salmon calcitonins. *Pharm Res 16*, 813–818.
158. Foster LJ (2007). Biosynthesis, properties and potential of natural-synthetic hybrids of polyhydroxyalkanoates and polyethylene glycols. *Appl Microbiol Biotechnol 75*, 1241–1247.
159. Gbadamosi JK, Hunter AC, Moghimi SM (2002). PEGylation of microspheres generates a heterogeneous population of particles with differential surface characteristics and biological performance. *FEBS Lett 532*, 338–344.
160. Madani F, Bessodes M, Lakrouf A, Vauthier C, Scherman D, Chaumeil J-C (2007). PEGylation of microspheres for therapeutic embolization: Preparation, characterization and biological performance evaluation. *Biomaterials 28*, 1198–1208.
161. Madani F, Bessodes M, Lakrouf A, Vauthier C, Scherman D, Chaumeil JC (2007). PEGylation of microspheres for therapeutic embolization: Preparation, characterization and biological performance evaluation. *Biomaterials 28*, 1198–1208.
162. Lee CM, Choi Y, Huh EJ, Lee KY, Song HC, Sun MJ, Jeong HJ, Cho CS, Bom HS (2005). Polyethylene glycol (PEG) modified 99mTc-HMPAO-liposome for improving blood circulation and biodistribution: The effect of the extent of PEGylation. *Cancer Biother Radiopharm 20*, 620–628.
163. Vyas SP, Rawat M, Rawat A, Mahor S, Gupta PN (2006). Pegylated protein encapsulated multivesicular liposomes: A novel approach for sustained release of interferon alpha. *Drug Dev Ind Pharm 32*, 699–707.
164. Sadzuka Y, Tokutomi K, Iwasaki F, Sugiyama I, Hirano T, Konno H, Oku N, Sonobe T (2006). The phototoxicity of photofrin was enhanced by PEGylated liposome in vitro. *Cancer Lett 241*, 42–48.
165. Remaut K, Lucas B, Braeckmans K, Demeester J, De Smedt SC (2007). Pegylation of liposomes favours the endosomal degradation of the delivered phosphodiester oligonucleotides. *J Control Release 117*, 256–266.
166. Hamblin MR, Miller JL, Rizvi I, Loew HG, Hasan T (2003). Pegylation of charged polymer-photosensitiser conjugates: Effects on photodynamic efficacy. *Br J Cancer 89*, 937–943.
167. Dhanikula AB, Panchagnula R (2005). Preparation and characterization of water-soluble prodrug, liposomes and micelles of Paclitaxel. *Curr Drug Deliv 2*, 75–91.
168. Rossi J, Giasson S, Khalid MN, Delmas P, Allen C, Leroux JC (2007). Long-circulating poly(ethylene glycol)-coated emulsions to target solid tumors. *Eur J Pharm Biopharm 67*, 329–338.
169. Chen J, Tian B, Yin X, Zhang Y, Hu D, Hu Z, Liu M, Pan Y, Zhao J, Li H, Hou C, Wang J (2007). Preparation, characterization and transfection efficiency of cationic PEGylated PLA nanoparticles as gene delivery systems. *J Biotechnol 130*, 107–113.

170. Sanders NN, Peeters L, Lentacker I, Demeester J, De Smedt SC (2007). Wanted and unwanted properties of surface PEGylated nucleic acid nanoparticles in ocular gene transfer. *J Control Release 122*, 226–235.

171. Yoncheva K, Guembe L, Campanero MA, Irache JM (2007). Evaluation of bioadhesive potential and intestinal transport of pegylated poly(anhydride) nanoparticles. *Int J Pharm 334*, 156–165.

172. Torchilin VP (2002). PEG-based micelles as carriers of contrast agents for different imaging modalities. *Adv Drug Deliv Rev 54*, 235–252.

173. Bae Y, Nishiyama N, Kataoka K (2007). In vivo antitumor activity of the folate-conjugated pH-sensitive polymeric micelle selectively releasing adriamycin in the intracellular acidic compartments. *Bioconjug Chem 18*, 1131–1139.

174. Yu C, Kohn J (1999). Tyrosine-PEG-derived poly(ether carbonate)s as new biomaterials. Part I: Synthesis and evaluation. *Biomaterials 20*, 253–264.

175. Almany L, Seliktar D (2005). Biosynthetic hydrogel scaffolds made from fibrinogen and polyethylene glycol for 3D cell cultures. *Biomaterials 26*, 2467–2477.

176. Lin-Gibson S, Bencherif S, Cooper JA, Wetzel SJ, Antonucci JM, Vogel BM, Horkay F, Washburn NR (2004). Synthesis and characterization of PEG dimethacrylates and their hydrogels. *Biomacromolecules 5*, 1280–1287.

177. Wang M, Lee LS, Nepomich A, Yang JD, Conover C, Whitlow M, Filpula D (1998). Single-chain Fv with manifold *N*-glycans as bifunctional scaffolds for immunomolecules. *Protein Eng 11*, 1277–1283.

178. Cheng X, Ming X, Croyle MA (2003). PEGylated adenoviruses for gene delivery to the intestinal epithelium by the oral route. *Pharm Res 20*, 1444–1451.

179. Croyle MA, Callahan SM, Auricchio A, Schumer G, Linse KD, Wilson JM, Brunner LJ, Kobinger GP (2004). PEGylation of a vesicular stomatitis virus G pseudotyped lentivirus vector prevents inactivation in serum. *J Virol 78*, 912–921.

180. Chapman AP (2002). PEGylated antibodies and antibody fragments for improved therapy: A review. *Adv Drug Deliv Rev 54*, 531–545.

181. Goodman and Gilman's Manual of Pharmacology and Therapeutics (2008). *Acetylcholinesterase Inhibitors, Chapter 8*, New York: Mc-Graw Hill, pp. 126–134. Ebook link *http://www.accessmedicine.com/*

182. Kronman C, Cohen O, Raveh L, Mazor O, Ordentlich A, Shafferman A (2007). Polyethylene-glycol conjugated recombinant human acetylcholinesterase serves as an efficacious bioscavenger against soman intoxication. *Toxicology 233*, 40–46.

183. Jo YW, Youn YS, Lee SH, Kim BM, Kang SH, Yoo M, Choi EC, Lee KC (2006). Long-acting interferon-alpha 2a modified with a trimer-structured polyethylene glycol: Preparation, in vitro bioactivity, in vivo stability and pharmacokinetics. *Int J Pharm 309*, 87–93.

184. Asselah T, Ripault, M-P, Castelnau C, Giuily N, Boyer N, Marcellin P (2005). The current status of antiviral therapy of chronic hepatitis B. *J Clin Virol 34*, S115–S124.

185. Souvignet C, Lejeune O, Trepo C (2007). Interferon-based treatment of chronic hepatitis C. *Biochimie 89*, 894–898.

186. Gupta SK, Swan SK, Marbury T, Smith W, Schwartz S, Kolz K, Cutler DL (2007). Multiple-dose pharmacokinetics of peginterferon alfa-2b in patients with renal insufficiency. *Br J Clin Pharmacol 6*, 6.

187. Tran HA, Attia JR, Jones TL, Batey RG (2007). Pegylated interferon-alpha2beta in combination with ribavirin does not aggravate thyroid dysfunction in comparison to regular interferon-alpha2beta in a hepatitis C population: Meta-analysis. *J Gastroenterol Hepatol 22*, 472–476.

188. Narta UK, Kanwar SS, Azmi W (2007). Pharmacological and clinical evaluation of L-asparaginase in the treatment of leukemia. *Crit Rev Oncol Hematol 61*, 208–221.

189. Youn YS, Jung JY, Oh SH, Yoo SD, Lee KC (2006). Improved intestinal delivery of salmon calcitonin by Lys18-amine specific PEGylation: Stability, permeability, pharmacokinetic behavior and in vivo hypocalcemic efficacy. *J Control Release 114*, 334–342.

190. Youn YS, Na DH, Lee KC (2007). High-yield production of biologically active mono-PEGylated salmon calcitonin by site-specific PEGylation. *J Control Release 117*, 371–379.

191. Nishikawa M, Hyoudou K, Kobayashi Y, Umeyama Y, Takakura Y, Hashida M (2005). Inhibition of metastatic tumor growth by targeted delivery of antioxidant enzymes. *J Control Release 109*, 101–107.

192. Dhawan S, Singla AK (2003). Nifedipine loaded chitosan microspheres prepared by emulsification phase-separation. *Biotech Histochem 78*, 243–254.

193. Singla AK, Sharma ML, Dhawan S (2001). Nifedipine loaded chitosan microspheres: Characterization of internal structure. *Biotech Histochem 76*, 165–171.

194. Dhawan S, Singla AK, Sinha VR (2004). Evaluation of mucoadhesive properties of chitosan microspheres prepared by different methods. *AAPS PharmSciTech 5*, e67.

195. Chan P, Kurisawa M, Chung JE, Yang, Y-Y (2007). Synthesis and characterization of chitosan-*g*-poly(ethylene glycol)-folate as a non-viral carrier for tumor-targeted gene delivery. *Biomaterials 28*, 540–549.

196. Prego C, Torres D, Fernandez-Megia E, Novoa-Carballal R, Quinoa E, Alonso MJ (2006). Chitosan-PEG nanocapsules as new carriers for oral peptide delivery. Effect of chitosan pegylation degree. *J Control Release 111*, 299–308.

197. Humphreys DP, Heywood SP, Henry A, Ait-Lhadj L, Antoniw P, Palframan R, Greenslade KJ, Carrington B, Reeks DG, Bowering LC, West S, Brand HA (2007). Alternative antibody Fab′ fragment PEGylation strategies: Combination of strong reducing agents, disruption

of the interchain disulphide bond and disulphide engineering. *Protein Eng Des Sel 20*, 227–234.

198. Stigsnaes P, Frokjaer S, Bjerregaard S, Weert Mvd, Kingshott P, Moeller EH (2007). Characterisation and physical stability of PEGylated glucagon. *Int J Pharm 330*, 89–98.
199. Stigsnaes P, Frokjaer S, Bjerregaard S, van de Weert M, Kingshott P, Moeller EH (2007). Characterisation and physical stability of PEGylated glucagon. *Int J Pharm 330*, 89–98.
200. Asanuma H, Nakai K, Sanada S, Minamino T, Takashima S, Ogita H, Fujita M, Hirata A, Wakeno M, Takahama H, Kim J, Asakura M, Sakuma I, Kitabatake A, Hori M, Komamura K, Kitakaze M (2007). *S*-Nitrosylated and pegylated hemoglobin, a newly developed artificial oxygen carrier, exerts cardioprotection against ischemic hearts. *J Mol Cell Cardiol 42*, 924–930.
201. Li D, Manjula BN, Ho NT, Simplaceanu V, Ho C, Acharya AS (2007). Molecular aspects of the high oxygen affinity of non-hypertensive hexa pegylated hemoglobin, [(SP-PEG5K)(6)-Hb]. *Artif Cells Blood Substit Immobil Biotechnol 35*, 19–29.
202. Nacharaju P, Friedman JM, Prabhakaran M, Acharya SA, Manjula BN (2007). Combining the influence of two low O2 affinity-inducing chemical modifications of the central cavity of hemoglobin. *Biochemistry 46*, 4554–4564.
203. Federico R, Cona A, Caliceti P, Veronese FM (2006). Histaminase PEGylation: Preparation and characterization of a new bioconjugate for therapeutic application. *J Control Release 115*, 168–174.
204. Jeng F-Y, Lin S-C (2006). Characterization and application of PEGylated horseradish peroxidase for the synthesis of poly(2-naphthol). *Process Biochem 41*, 1566–1573.
205. Youn YS, Chae SY, Lee S, Jeon JE, Shin HG, Lee KC (2007). Evaluation of therapeutic potentials of site-specific PEGylated glucagon-like peptide-1 isomers as a type 2 anti-diabetic treatment: Insulinotropic activity, glucose-stabilizing capability, and proteolytic stability. *Biochem Pharmacol 73*, 84–93.
206. Adhikari PM, Mohammed N, Pereira P (1997). Changing profile of diabetic ketosis. *Indian Med Assoc 95*, 540–542.
207. Denisenko TV (1990). Glycosylated lipoproteins as atherogenic factor in diabetes (review of the literature). *Vopr Med Khim 36*, 5–10.
208. Lee DY, Park SJ, Lee S, Nam JH, Byun Y (2007). Highly poly(ethylene) glycolylated islets improve long-term islet allograft survival without immunosuppressive medication. *Tissue Eng 13*, 2133–2141.
209. Lee DY, Park SJ, Nam JH, Byun Y (2006). A combination therapy of PEGylation and immunosuppressive agent for successful islet transplantation. *J Control Release 110*, 290–295.
210. Wisconsin, UO (2007). *Biological Magnetic Resonance Data Bank*. Available: http://www.bmrb.wisc.edu/.
211. Daly SM, Przybyciena TM, Tilton RD (2007). Aggregation of lysozyme and of poly(ethylene glycol)-modified lysozyme after adsorption to silica. *Colloids and Surfaces B: Biointerfaces 57*, 81–88.
212. Daly SM, Przybycien TM, Tilton RD (2007). Aggregation of lysozyme and of poly(ethylene glycol)-modified lysozyme after adsorption to silica. *Colloids Surf B Biointerfaces 57*, 81–88.
213. Gamez A, Wang L, Sarkissian CN, Wendt D, Fitzpatrick P, Lemontt JF, Scriver CR, Stevens RC (2007). Structure-based epitope and PEGylation sites mapping of phenylalanine ammonia-lyase for enzyme substitution treatment of phenylketonuria. *Mol Genet Metab 91*, 325–334.
214. Sarkissian CN, Gamez A (2005). Phenylalanine ammonia lyase, enzyme substitution therapy for phenylketonuria, where are we now? *Mol Genet Metab 86*, S22–26.
215. Huang Z, Zheng Q, Wu X, Su Z, Xu H, Tan Y, Feng W, Li X, Cai L (2007). Enhanced protection of modified human acidic fibroblast growth factor with polyethylene glycol against ischemia/reperfusion-induced retinal damage in rats. *Toxicol Lett 170*, 146–156.
216. Sainathan SK, Tu L, Bishnupuri KS, Han M, Li A, Newberry RD, McDonald KG, Crimmins DL, Houchen C, Anant S, Dieckgraefe BK (2005). PEGylated murine granulocyte-macrophage colony-stimulating factor: Production, purification, and characterization. *Protein Expr Purif 44*, 94–103.
217. Piedmonte DM, Treuheit MJ (2008). Formulation of Neulasta (pegfilgrastim). *Adv Drug Deliv Rev 60*, 50–58.
218. Amgen I (2006). *NEULASTA (pegfilgrastim) Prescribing Information*.
219. Li C, Wallace S (2008). Polymer-drug conjugates: Recent development in clinical oncology. *Adv Drug Deliv Rev 60*, 886–898.
220. Boyd MR, Gustafson KR, McMahon JB, Shoemaker RH, O'Keefe BR, Mori T, Gulakowski RJ, Wu L, Rivera M, Laurencot CM, Currens MJ, Cardellina JH 2nd, Buckheit RW Jr, Nara PL, Pannell LK, Sowder RC 2nd, Henderson LE (1997). Discovery of cyanovirin-N, a novel human immunodeficiency virus-inactivating protein that binds viral surface envelope glycopraotein gp120: Potential applications to microbicide development. *Antimicrob Agents Chemother 41*, 1521–1530.
221. Mori T, Gustafson KR, Pannell LK, Shoemaker RH, Wu L, McMahon JB, Boyd MR (1998). Recombinant production of cyanovirin-N, a potent human immunodeficiency virus-inactivating protein derived from a cultured cyanobacterium. *Protein Expr Purif 12*, 151–158.
222. Zappe H, Snell ME, Bossard MJ (2008). PEGylation of cyanovirin-N, an entry inhibitor of HIV. *Adv Drug Deliv Rev 60*, 79–87.
223. Gustafson KR, Sowder RC 2nd, Henderson LE, Cardellina JH 2nd, McMahon JB, Rajamani U, Pannell LK, Boyd MR (1997). Isolation, primary sequence determination, and disulfide bond structure of cyanovirin-N,

an anti-HIV (human immunodeficiency virus) protein from the cyanobacterium *Nostoc ellipsosporum*. *Biochem Biophys Res Commun 238*, 223–228.

224. Reeves JD, Piefer AJ (2005). Emerging drug targets for antiretroviral therapy. *Drugs 65*, 1747–1766.
225. Pasut G, Sergi M, Veronese FM (2008). Anti-cancer PEG-enzymes: 30 years old, but still a current approach. *Adv Drug Deliv Rev 60*, 69–78.
226. Tan Y, Sun X, Xu M, An Z, Tan X, Han Q, Miljkovic DA, Yang M, Hoffman RM (1998). Polyethylene glycol conjugation of recombinant methioninase for cancer therapy. *Protein Expr Purif 12*, 45–52.
227. Savoca KV, Abuchowski A, van Es T, Davis FF, Palczuk NC (1979). Preparation of a non-immunogenic arginase by the covalent attachment of polyethylene glycol. *Biochim Biophys Acta 578*, 47–53.

# GOLD STANDARD ANIMAL MODELS

Carol Castaneda and Michael S. LaPointe
*Indiana University Northwest, Gary, Indiana*

## 1 INTRODUCTION: IMPORTANCE OF AND DANGERS IN USING ANIMAL MODELS FOR DRUG DEVELOPMENT

Using data from animal models for designing and approving human trials is necessary but must be done cautiously. Animal models should be characterized carefully and compared to similar human systems including in vitro studies using human cells. For example, CD4 is expressed on human macrophages but not on murine macrophages. Using multiple animal models is important and no data from any of these systems should be ignored. When animal models are used to examine the efficacy of treatment, it is very important that treatment in animal models start after clinical symptoms become evident. Moreover, drugs should also be tested using in vitro human systems when available. Well-characterized animal models can also be used to examine the ability of certain prophylactic treatments to prevent the onset of disease or pathologic symptoms. Finally, human trials should be based on sound scientific evidence using animal models and other appropriate model systems. Even when human trials go awry, we must investigate the results openly and carefully in order to determine why there were differences between the human trials and the animal models. This will lead to the development of better models and thus faster and more effective treatments in the future.

Testing of biologics has dramatically expanded the horizon of possible treatments; a successful example being infliximab, a tumor necrosis factor-alpha (TNF-α) inhibitor used to treat rheumatoid arthritis. Caution must be taken when using antibody reagents. These reagents may be fully or partial agonistic or antagonistic, leading to very different treatment results. Often these antibodies are thought to block or activate the function of cytokines or specific cellular receptors. It is important, therefore, to rigorously test the effects of these reagents in several animal models due to the complex intricacy and redundancy of all living systems and therefore the many possible dangers.

Monoclonal antibody reagents have often been used to block a specific signaling receptor. However, it has been shown in animal models that a monoclonal antibody (mAb) can cross-link the receptor it binds and thus result in signaling rather than blocking the signal [1, 2]. This cross-linking is especially powerful if the fragment crystallizable (Fc) region of the treatment antibody binds to Fc receptors on the patient's own cells, resulting in extensive cross-linking. Complement activation and other inflammatory responses are, therefore, very real possibilities as a result of monoclonal antibody treatment. Although there are dangers in using these treatments, there is also the exciting potential in transplantation, cancer, and autoimmune disease. Some evidence suggests that short-term treatment with the antibodies that block T-cell activation may result in long-term autoantigen-specific tolerance and remission from clinical disease.

## 2 IMMUNOMODULATION

For activation, T lymphocytes require antigen-specific signals through the binding of their T-cell receptor (TCR) complex and co-stimulatory signals including

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

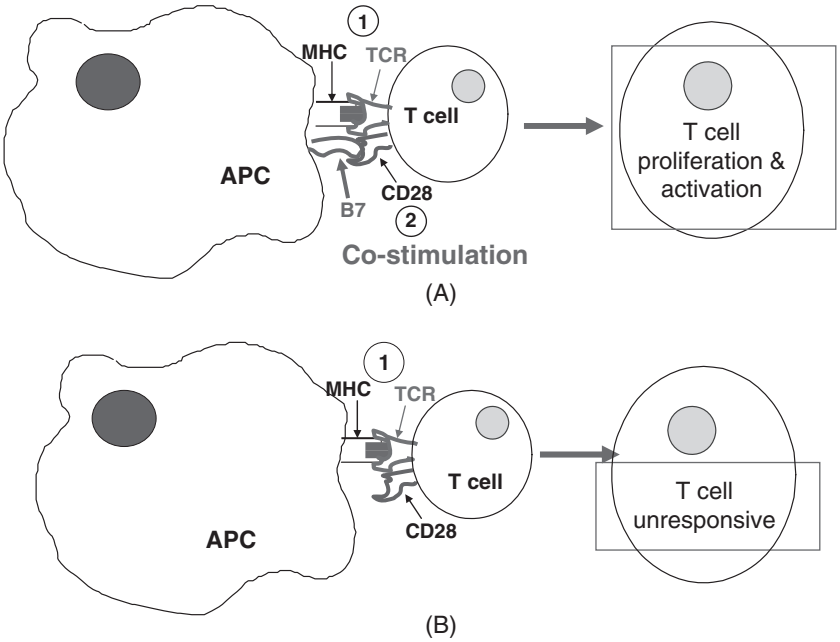


**Figure 1** For activation, T lymphocytes require antigen-specific signals through the binding of their T cell receptor (TCR) complex and costimulatory signals including signals through their CD28. CD28 on the T cell binds CD80 and CD86 on an antigen-presenting cell (APC).

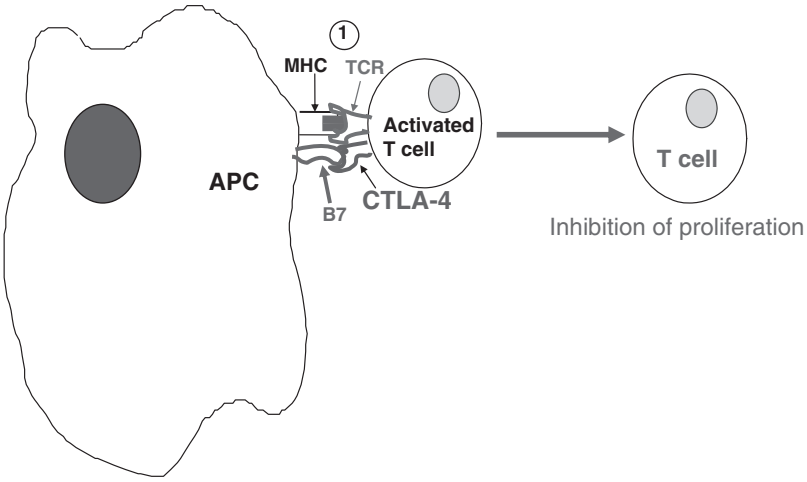


**Figure 2** CD80 and CD86 on the APC also bind CTLA-4 on activated T cells. Binding of CTLA-4 transduces a negative signal, causing the activated T cell to stop proliferating.

signals through their CD28 and CD154 receptors. CD28 on the T cell binds CD80 (B7-1) and CD86 (B7-2) on an antigen-presenting cell (APC). In fact, binding of the antigen-specific TCR alone actively induces anergy (T-cell unresponsiveness), leaving these T cells unable to respond even to an APC expressing co-stimulatory receptors (Fig. 1). CD80 and CD86 can also bind to a negative regulator of T-cell proliferation, CTLA-4 (CD152), found in higher concentration on activated lymphocytes (Fig. 2). CD154 on T lymphocytes binds CD40 on dendritic cells (APCs) and on B lymphocytes (Fig. 3). This interaction is necessary for the activation of B cells and maturation of dendritic cells. Antibodies designed to bind any of these receptors can modulate the immune response and can be antagonists, partial agonists, or agonists.

Once T lymphocytes are activated, these effector cells must travel to and enter the area of infection/inflammation. Receptors such as very late antigen (VLA)-4 on activated T cells are required for entrance into the inflamed tissue and are therefore targets for blocking mAb therapy. Activated T lymphocytes can be

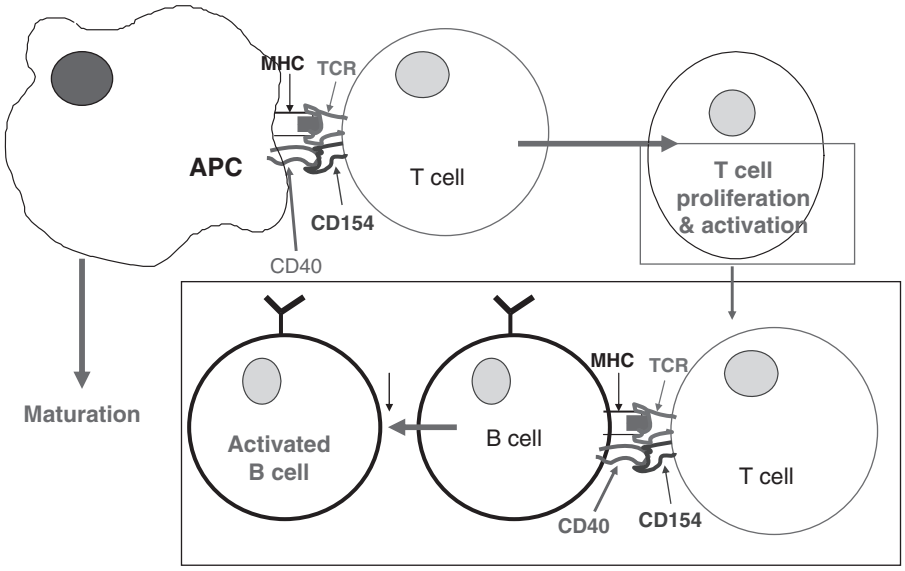


**Figure 3** CD40/CD154 interaction signals activation/maturation for dendritic (APCs), T helper cells and B lymphocytes.

functionally diverse, including $CD4^+$ helper T cells (type 1, type 2, or type 17), $CD8^+$ cytotoxic lymphocytes, or $CD4^+CD25^+$ regulatory T cells, which can be powerfully immunosuppressive.

Using two mouse models of the human demyelinating disease multiple sclerosis (MS), Remitting experimental autoimmune encephalomyelitis (R-EAE) and Theiler's murine encephalomyelitis virus-induced demyelinating disease (TMEV-IDD), along with the NOD mouse model of type I diabetes, we will demonstrate the importance of understanding the structure and function of the drug being tested and the pathology in the animal models being used.

## 3 RELAPSING REMITTING EXPERIMENTAL AUTOIMMUNE ENCEPHALOMYELITIS

Relapsing remitting experimental autoimmune encephalomyelitis (R-EAE) in SJL mice is an artificial but useful animal model of the human demyelinating disease multiple sclerosis. Myelin is the lipid insulation found on many nerve axons. The presence of myelin covering these axons is necessary for efficient propagation of the neuronal signal. Central nervous system (CNS) demyelinating disease is induced in genetically identical SJL mice by subcutaneous immunization with 100 μL complete Freund adjuvant emulsion containing *Mycobacterium tuberculosis* H37Ra plus a myelin antigen such as proteolipid protein (PLP) or myelin basic protein (MBP) or specific PLP or MBP peptides. Disease in these mice follows a relapsing/remitting course.

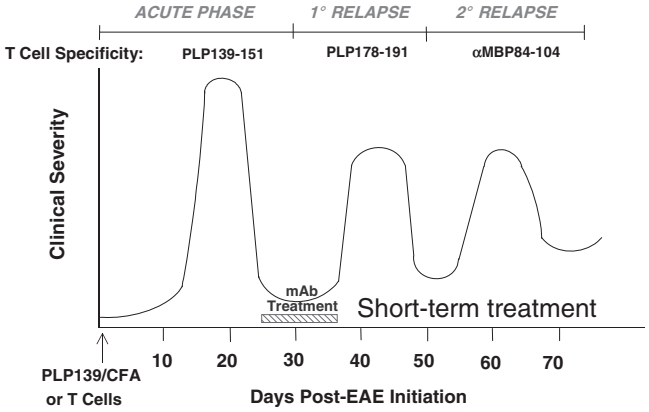


**Figure 4** Blockade of costimulation during remission in PLP 139–151-induced R-EAE effects epitope spreading.

The immune response along with the resulting clinical disease has been carefully characterized in this murine R-EAE model of MS. We are able to induce disease with a specific myelin epitope and verify that acute disease is due to T-helper type 1 lymphocytes targeting this epitope, for example, PLP 139–151. During the acute phase of disease and into remission, T cells specific to other epitopes on myelin are becoming activated (epitope spreading), probably due to interferon-gamma (IFN-γ)-induced tissue damage at the inflammatory site (Fig. 4). These newly activated autoreactive T cells are responsible for clinical relapse [3]. This animal model demonstrates the difficulty when searching for an antigen-specific treatment for human autoimmune disease since the target of the T-cell

response changes with time. Human autoimmune diseases such as MS and type I diabetes confirm this phenomenon of epitope spreading in human disease.

## 4 BLOCKADE OF T-CELL CO-STIMULATION USING ANTI-B7 MONOCLONAL ANTIBODIES IN R-EAE

Using the mouse model of MS, relapsing R-EAE in SJL mice, anti-B7-1 monoclonal antibodies were used to treat mice during remission after the acute phase of disease. The rational for this treatment was that co-stimulation blockade would anergize autoreactive T lymphocytes being activated in vivo during initial myelin damage and would therefore block relapses. Instead of blocking further relapses as was expected, this treatment resulted in exacerbation of disease relapses [1]. Further examination of the effects of whole antibody treatment demonstrated that CNS-infiltrating $CD4^+$ T-helper type 1 (TH1) lymphocytes bound anti-CD80 mAb; that increased activity and survival of these autoreactive lymphocytes correlated with exacerbation of disease; and in vitro studies demonstrated that cross-linking CD80 on these T cells with the addition of TH1 cytokines increased T-cell production of T-bet, IFN-γ, and Bcl-x [4]. These findings are interesting since T-bet is a transcription factor essential for the development of TH1 lymphocytes and the transcription of the IFN-γ gene while Bcl family members are important in down-regulation of an immune response via apoptosis of effector T cells. This data may be relevant to human MS since during disease the expression of B7-1 in the CNS is high both in this mouse model and in human MS.

Subsequent treatment under the same experimental conditions with anti B7-1 F(ab) fragments, which contain only one binding site and therefore cannot cross-link, resulted in profound and extended inhibition of disease relapses [2]. Epitope spreading results from the activation of T lymphocytes. Therefore, blocking co-stimulation during the activation of these newly recruited autoreactive T cells should stop epitope spreading. As demonstrated in R-EAE, treatment with anti-B7-1 F(ab) not only blocked clinical disease but also halted epitope spreading. The T-cell response to the spread epitopes such as PLP 178–191 was blocked (Fig. 4). This work demonstrates the importance of understanding the treatment reagent and the animal model being used.

## 5 BLOCKADE OF T-CELL CO-STIMULATION WITH CTLA4-IG (ABATACEPT AND BELATSCEPT)

CTLA4-Ig is a fusion protein that includes the B7-binding domain of CTLA-4 (abatacept) and therefore binds both B7-1 and B7-2. Treatment with CTLA4-Ig blocks both B7-1 and B7-2 interactions, in contrast to the anti-B7-1 F(ab) treatment described above. This biologic has been shown to block allograft rejection and autoimmune disease in various animal models, even resulting in long-term tolerance [5]. In SJL R-EAE, treatment with CTLA4-Ig downregulates epitope spreading and mediates remission from disease. No noticeable adverse effects were seen [6]. In human trials, abatacept treatment reduced psoriatic lesions in patients with psoriasis vulgaris and demonstrated significant improvements in the signs and symptoms of rheumatoid arthritis (RA) including long-term inhibition of structural damage. No significant detrimental side effects were seen unless RA patients received both abatacept and anti-TNF agents. Abatacept has been approved by the Food and Drug Administration (FDA) for treating rheumatoid arthritis.

Development of belatacept is an excellent example of using animal models carefully and to our advantage. Although CTLA4-Ig (abatacept) induces prolonged graft survival in rodent models of transplantation, this biologic did not prove effective in nonhuman primate models of renal transplantation. During these studies, it was recognized that abatacept binds preferentially to B7-1. In response to this knowledge, the second-generation CTLA4-Ig, belatacept, was produced. Belatacept binds more strongly to both B7s; in nonhuman primate studies belatecept demonstrated better efficacy and is showing promise in phase II human trials [5].

## 6 THEILER'S MURINE ENCEPHALOMYELITIS VIRUS–INDUCED DEMYELINATING DISEASE

Theiler's murine encephalomyelitis virus—Induced demyelinating disease (TMEV–IDD) is a virally induced demyelinating disease with clinical disease progression and CNS histopathology much like the progressive form of human MS. The virus remains chronic in the CNS of SJL mice, however, myelin damage appears to be due to a T-helper type 1 inflammatory response targeting myelin not virus-infected cells. In fact, within 30 days of infection, the target of the immune response appears to include myelin components, and specific autoreactive T lymphocytes are found in the CNS along with activated macrophages, which are stripping and engulfing myelin from the neuronal axons. Epitope spreading, as characterized in the simpler more artificial MS model R-EAE, progresses in TMEV–IDD from viral epitopes to the immunodominant epitope PLP 139–151 and finally to other epitopes on PLP, MBP, and additional myelin proteins [7].

As a mouse model of MS, TMEV–IDD is more complex than R-EAE. As one might expect, attempts to block T-cell activation and the resulting macrophage-induced demyelination can lead to increased levels of virus and do not result in significant and long-lasting inhibition of disease progression.

## 7 BLOCKADE OF CD154/CD40 SIGNALING IN R-EAE AND TMEV-IDD

A second co-stimulatory signal important to the maturation of dendritic cells, B lymphocytes, and T lymphocytes is the CD154/CD40 interaction (Fig. 3). This co-stimulatory pathway is a good example of the complex functionality of many signaling molecules and the important variation of expression in different animal models compared to the expression on human cells. CD154 is found on activated T lymphocytes in both mice and men. However, this receptor is also found in humans on activated dendritic cells, mast cells, monocytes/macrophages, neutrophils, and platelets. CD40 is expressed on mouse and human dendritic cells, monocyte/macrophage, and B lymphocytes. CD154 blockade has been used successfully to inhibit disease progression in many autoimmune animal models including R-EAE, TMEV-IDD [8] and NOD [9].

Anti-CD154 treatment in both R-EAE and TMEV-IDD resulted in reduced severity of clinical signs and CNS demyelination. Treatment of R-EAE in SJL mice with a total of four injections of blocking anti-CD154 given every other day beginning at the peak of the acute phase or during remission (Fig. 4) reduced the severity of clinical disease, the number and severity of disease relapses, and the infiltration of IFN-γ-producing T cells into the CNS. No significant adverse reactions were noted [10].

CD154 blockade using a similar protocol in TMEV-IDD resulted in a temporary reduction of disease progression along with the reduction of CNS infiltration by T cells and macrophage [11]. The temporary quality of this inhibition may be due to the chronic nature of the viral infection of the CNS. Since a previous or concurrent viral infection is hypothesized as a trigger for human MS, results of treatment in this virally induced MS model are pertinent to human trials. No adverse reactions were noted in these experiments.

Although CD154 blockade in many animal models of autoimmunity and transplantation has been successful and exposed no adverse reaction, clinical trials in human autoimmune disease were halted due to a low but significant incidence of thrombotic events—inappropriate blood clotting leading to tissue damage as in coronary thrombosis. This is no doubt connected in some way to the expression of CD154 on human platelets.

## 8 BLOCKADE OF CD154/CD40 SIGNALING IN NONOBESE DIABETIC MICE (NOD)

Anti-CD154 treatment of NOD mice demonstrates the variation in clinical outcomes dependent on the time of treatment relative to the pathology. Female NOD mice spontaneously develop inflammation of the pancreas at around 3–5 weeks of age. This insulitis is followed by diabetes due to the inflammatory destruction of the β cells of the pancreas, which produce insulin. Diabetes, defined by the increase in blood glucose, occurs between 3 and 7 months of age in the NOD mice. This model of insulin-dependent diabetes mellitus (IDDM) closely resembles human IDDM in histopathology and clinical presentation. Infiltration of the pancreas by T lymphocytes and activated macrophage set the stage for destruction of the pancreatic β cells exclusively. No other pancreatic cells are destroyed [12]. In the mouse model and in human IDDM, the epitope spreading phenomenon has been proposed and much data supports this proposal.

Treatment with anti-CD154 mAb at 3 weeks of age, prior to onset of diabetes, completely prevented insulitis and diabetes. This time-dependent efficacy can also be seen when these mice are treated with other biologics such as anti-CD3 antibody [12]. Treatment of human patients prior to clinical disease onset will present some problems. However, since IDDM tends to run in families, early treatment may be possible. Interestingly, even though T-cell involvement is unquestioned in both mouse model and human IDDM, depletion of B cells by treatment with anti-CD20 (retuximab) is in human diabetes trials and has been tested and studied in a transgenic NOD mouse model in which the human CD20 molecule is expressed on mouse B lymphocytes [13].

## 9 ANTI-CD3 (TCR) IN EAE AND NOD

Responses to drugs can be very different depending on the animal model used, nature of the biologic, dose given, and when treatment begins. The clinical outcome of treatment with antibodies against the T-cell receptor in murine and human autoimmunity differ greatly depending on the Fc portion of the antibodies used. Although Fc receptor-binding (mitogenic) anti-CD3 mAB were used to successfully treat many transplantation and autoimmunity animal models, the initial cytokine burst induced due to massive T-cell activation is a potential problem. Short-term treatment with Fc receptor nonbinding (nonmitogenic) anti-CD3 (teplizumab) in both EAE and NOD resulted in disease reversal and long-term tolerance without the initial cytokine burst [14–16].

A short, 5-day treatment protocol with nonmitogenic anti-CD3 given after the peak of the acute phase in R-EAE (Fig. 4) resulted in reversal of disease, inhibition of disease relapse, and reduction in type 1 cytokines like IFN-γ, Interleukin-12 (IL-12), and TNF-α [16]. This treatment protocol in the NOD mouse had little effect when started after clinical disease [14]. No adverse reactions were noted.

Teplizumab trials in new onset diabetic patients have been encouraging; however, improvements were seen mainly in patients with less β-cell damage at treatment onset—mirroring the outcome in the NOD animal model. Disease progression stopped in these patients for over a year after treatment and insulin requirements were reduced. Although side effects were much reduced compared to the mitogenic anti-CD3, some minor flulike side effects were experienced due to a limited cytokine release at the time of injection [14].

## 10 A SUPERAGONISTIC ANTI-CD28 MONOCLONAL ANTIBODY (VISILIZUMAB) IN LEWIS RAT EAE AND ADJUVANT ARTHRITIS ANIMAL MODEL

CD28 is expressed on all T lymphocytes (Fig. 1), including helper, cytotoxic, and regulatory T lymphocytes. In both a Lewis rat model of multiple sclerosis, experimental autoimmune encephalomyelitis [17], and in a rat model of rheumatoid arthritis, adjuvant arthritis [18], a superagonistic anti-CD28 monoclonal antibody proved to be an effective treatment. Characterization of the immune response in treated rats suggested that protection from disease was due to the activation of regulatory T lymphocytes. Activation of such regulatory T cells has been an exciting area of research, promising new ways to reduce inappropriate immune responses including autoimmunity.

It is important to note that the treatment in these animal models occurred prior to or immediately upon onset of clinical disease and treatment with this antibody apparently did not require antigen-specific TCR signals for T-cell activation. Both of these details demonstrate the need for caution. Unfortunately, a human trial of an anti-CD28 superagonist TGN1412 resulted in the hospitalization of six human volunteers [19, 20]. It became apparent that in humans this antibody treatment resulted in indiscriminant activation of helper T cells with a resulting "cytokine storm."

Understanding the biologic reagent used may be the important insight in this case. Antibodies to be used in human trials are usually bioengineered to include the human constant region in order to extend their half-life. This Fc domain is the functional end of the antibody, binding to Fc receptors on many cells in the body and inducing various responses, including secretion of inflammatory cytokines. The human Fc portion of this biologic may not bind to the Fc receptors on the cells in the animal models tested. Once introduced into humans, the unforeseen binding of the Fc portion of the antibody to the Fc receptors on the human T-helper lymphocytes and APCs probably induced the massive secretion of these potent cytokines.

## 11 T-CELL MIGRATION BLOCKADE WITH ANTI-VLA-4 (NATALIZUMAB)

Migration of encephalitogenic $CD4^+$ T cells into the CNS during the induction of R-EAE in SJL mice is dependent on the binding of VLA-4 (α4β1), expressed by the autoactivated T cell, to the endothelial cell ligand VLA-4 calcium-dependent adhesion molecule (VCAM)-1. Blockade of this migration with an anti-VLA-4 antibody, or a small-molecule VLA-4 antagonist (monovalent), prior to clinical disease, delays and inhibits disease onset. However, treatment after onset of clinical disease does not significantly inhibit disease. In fact, treatment with either VLA-4 antagonist during the first remission (Fig. 4) results in enhanced disease and in all treatment protocols, once treatment is halted, significant disease exacerbation occurs [21, 22].

Interestingly, even using protocols that inhibited disease, treatment with VLA-4 blockers enhanced proliferation and IFN-γ production by myelin-specific splenic T cells. These T cells are not only specific for the priming epitope (PLP 139–151) but also to newly spread epitopes (PLP 178–191). Enhanced proliferation and epitope spreading may explain the posttreatment exacerbation seen in all VLA-4 blocking treatment protocols. This posttreatment exacerbation of disease corresponds to significant increase in the histopathology in the CNS, including enhanced lymphoid cell infiltration, amplified expression of VCAM-1, and increased proinflammatory cytokine messenger ribonucleic acid (mRNA).

Unfortunately, extension of the treatment to 44 days in R-EAE did not improve the outcome. Disease exacerbation began even before treatment ended, resulting in accumulation of highly activated T cells circulating in the periphery and invading the CNS. These studies suggest multiple functions for VLA-4 binding and advise caution when bringing these biologics into clinical trials.

Based on the work done in R-EAE, expectations for VLA-4 blockers might be low. In fact, the results of clinical trials in relapsing MS were encouraging, resulting in 50% reduction in relapses in these patients [23, 24]. This led to early approval of natalizumab

(TYSABRI) by the FDA in 2004, prior to the completion of these phase III trials. As in our R-EAE model, MS patients on TYSABRI did exhibit increased indicators of peripheral T-cell activation including significantly increased numbers of cells secreting proinflammatory cytokines TNF-α and IFN-γ [25]; however, to date, thousands of patients have been receiving TYSABRI for well over 2 years, presenting with few serious side effects, with one puzzling exception [26, 27]. Two patients with MS and one patient with Crohn's disease (CD) were diagnosed with progressive multifocal leukoencephalopathy (PML).

PML is a demyelinating disease caused by the JC virus. This virus is a common human virus persisting in renal and CNS tissue of approximately 80% of the world population. This virus can be reactivated in severely immunocompromised patients (PML) and may be associated with gastrointestinal cancers [28].

Although initially pulled off the market, due to the great promise of natalizumab, in 2005 the FDA reapproved this biologic for use in relapsing MS. In addition, thousands of Crohn's patients are receiving natalizumab with promising results. In the future it will be important to monitor both the increased peripheral activation of proinflammatory cells and the possible reactivation of latent viruses in patients receiving reagents designed to block migration of activated cells into an inflammatory region such as the CNS (MS) or the gastrointestinal tract (CD).

## 12 MODELS OF CARDIOVASCULAR DISEASE AND METABOLISM

Numerous animal models exist for studying blood pressure regulation, stroke, and metabolic disease. Some models have developed spontaneously in mice, rats, or other species. Some have been created through selective inbreeding and others through modern genetic manipulations to create transgenics, consomics, and knockouts. Still other models use surgical, chemical, or dietary alterations to create a disease state. The gold standard animal models listed in Table 1 were chosen based upon the extent of their use in biomedical research and their proven utility. With the exception of the streptozotocin-induced type diabetes, the models chosen were created through inbreeding or via spontaneous mutations. It is the opinion of the authors, that such models best represent the heterogeneity associated with human disease states as compared to transgenic and knockout models.

References listed in Table 1 are, for the most part, review articles, describing phenotypic or genotypic alterations in the animal model or intervention studies used to treat the disease state. Table 2 lists some of the major suppliers of animals for research purposes. Neither table is exhaustive. Suppliers listed in Table 2 are those with which the authors have experience. In addition to these and other commercial vendors, animal colonies, especially transgenics and knockouts, may be obtained from individual investigators or research facilities. Below is a brief synopsis of selected models. Some of the vendors' websites contain additional information about the strains and other information that the reader will find useful for selecting or developing an animal model that fits specific genotypic or phenotypic characteristics.

The spontaneously hypertensive rat (SHR) is by far the most widely used animal model for studying the development and treatment of hypertension. It was developed by selective litter mate pairing for hypertension from the wister kyoto rat (WKY) in the 1960s at the Okmato, Kyoto School of Medicine in Japan. Colonies have been in the United States since the late 1960s and can be obtained as an inbred or outbred rat from several commercial vendors. The WKY is the traditional normotensive control rat used when studying the SHR. Hypertension in these rats is multifactorial because of the breeding and selecting for high blood pressure—similar to most human cases of essential hypertension. Although the WKY is the traditional control rat used with the SHR, it is not ideal. During the development of the strain, hypertension was selected for and inbred, whereas the normotensive WKY parent colony was kept as an outbred colony. Thus, there is more genetic diversity within the WKY than SHR. Moreover, chance selection of nonhypertensive genes due to the bottleneck of inbreeding has lead to selection of genotypic and phenotypic features unrelated to hypertension. Another distraction from the SHR is that the kidney shows limited signs of hypertensive-related nephropathy. Thus, the SHR can be considered a model of renal protection against hypertensive nephropathy as compared to surgical or pharmacological models of comparable blood pressures.

Substrains of the SHR have other interesting and useful characteristics. A few substrains exhibit a high incidence of cerebral hemorrhaging and are thus termed stroke-prone SHR. The incidence of stroke is mainly seen in males, where the incidence rate can be as high as 82%.

The SHR is considered a salt-insensitive form of hypertension since blood pressure increases under a normal salt diet and is affected little by increases or decreases of salt in the diet. It also exhibits many signs of diabetes, such as insulin resistance, and hyperinsulemia. This appears to be a recurring theme in hypertensive models, suggesting that the development of

**TABLE 1 Commonly used animal models for human disease**

| Therapeutic Area | Animal Models | References |
|---|---|---|
| Cancer | Human xenografts in nude mice | 29 |
| | Multifocal, established prostate cancer using prostrate-specific membrane antigen (PSMA)$^+$ prostate cancer (PC)-3 tumor cells | 30 |
| | Various mouse models of established allograft tumurs | 31 |
| Immunology | | |
| Allergy | Systemic anaphylactic mouse model | 35 |
| | Scratching behavior mouse model | 35 |
| | Murine rhinovirus-induced asthma exacerbation | 36 |
| Autoimmunity | | |
| Lupus | 2,6,10,14-tetramethylpentadecane (TMPD)-induced lupus mouse model | 37 |
| | Pristine-induced mouse model of systemic lupus erythematosus (SLE) | 38 |
| Arthritis | Inbred mouse strain BXD2 of spontaneous arthritis | 39 |
| | Adjuvant arthritis (rats) | 40 |
| MS | EAE (various mouse and rat strains) | |
| | R-EAE (SJL mice)[a] | 1–4, 6, 16 |
| | EAE (Lewis rate) | 17 |
| | TMEV-IDD (MS)[a] | 7, 11 |
| Type I diabetes | NOD (type I diabetes)[a] | 9, 13 |
| | Transgenic NOD with human CD20[a] | 13 |
| Inflammatory bowel disease (IBD) | Murine Il-2 & Il-10 knockouts (IBD)[a] | 42 |
| | Transgenic human leukocyte antigen (HLA)-B27 rat IBD model | 42 |
| Transplantation | Rat graft models | 43 |
| Deficiency | Chronic granulomatous disease | 44 |
| Neurology | Various behavior models | 75 |
| Addiction | Cocaine-, nicotine-, alcohol-, heroin-seeking behavior models | 55 |
| Cognition | Models of learning and memory | 56–58 |
| | Social recognition models | 59 |
| Alzheimer's | Rodent models | 60 |
| Schizophrenia | Various animal models | 65 |
| Depression | Some of the most commonly applied rat models of depression | 66 |
| | Rodent models of depression | 67 |
| Parkinson's | Neurotoxin models of Parkinson's disease | 68 |
| Pain | Behavior testing (paw pressure and formalin test) | 62 |
| | Persistent pain models | 63, 64 |
| Cardiology | | |
| Hypertensive | Spontaneously hypertensive rat (SHR) | 126–136 |
| | Stroke-prone SHR | 94–102 |
| | Borderline hypertensive rat (BHR) | 82–93 |
| | Dahl salt-sensitive rat | 76–81, 103 |
| Hyperlipidemia | Scd1-deficient mice | 32 |
| Anti-coagulants | Two thrombosis models in rats containing a silk thread in the aorta and vena cava, respectively | 33 |
| Ischemia | Rat model of transient cerebral artery occlusion | 34 |
| Antimicrobial | | |
| Antibacterial | Mouse transnasal murine model of *Staphylocuccus aureus* pneumonia | 45 |
| | Mouse model of pneumonia by *Klebsiella pneumoniae* | 46 |
| | Mouse bacteremia, skin abscess and peritonitis models | 47 |
| Antivirals | Mouse models of rhinovirus infection | 48 |
| Vaccines | Novel animal models of staphylococcal biofilm-associated infection | 49 |
| Wound healing | Surgical wound infection model | 50 |
| Metabolic and endocrine | | |
| Obesity | Göttingen minipig | 51 |
| | Zucker rat (spontaneous) | 69–74 |
| Obesity and Type II diabetes | OB/OB mouse (spontaneous) | 104–110 |
| | DB/DB mouse (spontaneous) | 111–119 |

**TABLE 1** ***Continued***

| Therapeutic Area | Animal Models | References |
|---|---|---|
| Type I diabetes | NOD mouse (spontaneous)[a] | 9, 13 |
| | Biobreeding (BB) rat (spontaneous) | 41 |
| | Streptozotocin-induced type I diabetes | 120–125 |
| Muscular | | |
| Muscular dystrophy | C57 BL/10ScSN-Dmd $^{mdx/j}$(X-linked muscular dystrophy—Duchenne's) | 54 |
| Mitochondrial disorders | Rpe65-mutant rd12 mice | 52 |
| | Murine Opa3 gene models human Costeff syndrome | 53 |
| | Mouse model of mitochondrial hepatopathies | 61 |

[a]Indicates animal models discussed in the text.

**TABLE 2 Major Providers of Animals for Research**

| Provider | Website |
|---|---|
| Charles River Laboratories | http://www.criver.com |
| Covance Inc. | http://www.covance.com |
| Harlan Sprague–Dawley Inc. (Harlan Teklad) | http://www.harlan.com |
| Hilltop Lab Animals Inc. | http://www.hilltoplabs.com |
| Jackson Laboratories | http://www.jax.org |
| Taconic Farms Inc. | http://www.taconic.com |
| United States National Institutes of Health (NIH)—National Institute on Aging | http://www.nia.nih.gov |

hypertension may have an underlying metabolic component. Although the SHR is considered to be salt insensitive, crossing of the SHR with normotensive rats, such as the WKY, results in an F1 generation, the borderline hypertensive rat (BHR) that is salt sensitive and hyperreactive to stimuli. The BHR will develop increased blood pressure in response to various stresses, such has a high-salt diet, noise, crowding, or mild air jets blown at it. Kidneys of the F1 generation show more signs of renal injury due to increased blood pressure than the parent SHR strain.

The Dahl salt-sensitive (DSS) and salt-resistant rat (DSR) strains were originally developed Lewis Dahl and then maintained as an inbred colony by John Rapp in the early 1970s by selectively inbreeding litter mates after selection of salt sensitivity or salt resistance to increased blood pressure. Unlike the SHR, both the DSS and DSR were maintained as inbred colonies as the strains were being developed. The DSS rat shows normal or slightly elevated blood pressure when maintained on a normal rat chow diet. When exposed to a high-salt diet (typically 8%), the blood pressure rapidly and progressively increases. Many of the rats will exhibit cerebral hemorrhaging if kept on a high-salt diet for prolonged periods. They are also subjective to moderate-to-severe renal nephropathy in response to the high-salt diet.

Harlan Laboratories was the sole commercial provider of Dahl rats. In the mid-1990s there was a breeding mishap in one of the expansion colonies resulting in a DSS hybrid [103]. The vendor has since re-derived the strain from unaffected colonies. Charles River also maintains a colony of DSS and DSR rats that originated from the Harlan colony prior to the mishap and has been cleared of genetic impurities.

The OB/OB and DB/DB mouse strains are models used for obesity and type II diabetes. They are spontaneous mutations of the leptin gene and the leptin receptor gene, respectively. Both mutations end up creating a nonfunctional protein. Although leptin has been shown to be central in maintaining a lean body mass and normal metabolism in mice, its role is more limited in humans. Nonetheless, both mouse models exhibit the classic phenotypic signs associated with obesity and type II diabetes in humans. Thus, while the etiology of obesity and type II diabetes is more heterogenic than in these mouse models, the phenotypic similarities makes them useful targets for the study of the onset of secondary pathologic conditions associated with diabetes or for pharmacologic intervention.

## 13 KNOCKOUT AND TRANSGENIC ANIMALS AS ANIMAL MODELS OF HUMAN DISEASE

With the development of DNA (deoxyribonucleic acid) technology, large numbers of designer mice and rats can now be created. These animals can serve as interesting and informative animal models of human disease and range from addition of a human gene (transgenic) to deletion of a functional gene (knockouts). Use of these animals in testing pharmaceuticals must be done with caution.

As scientists learn more about the molecular etiology of certain diseases, the opportunity to create animals with similar molecular abnormalities opens a new avenue of research. Alison Abbott, in *Nature*, summarizes some of the exciting new animal models in the study of several psychiatric disorders such as depression and schizophrenia [75]. Many such disorders are diagnosed and described by very nonobjective criteria. This includes overlapping symptoms and little characterized pathology in terms of anatomy or molecular etiology. This vague characterization has led to the use of animal models that seem to some to be inadequate as models for human disease.

The advent of increased knowledge of the molecular deficiencies and the identification of candidate susceptibility genes in various human diseases have lead to the creation and development of more useful mouse models for study. The *Disc* gene is a strong candidate gene for schizophrenia. Transgenic mice with a disrupted *Disc-1* gene show anatomical similarity to human schizophrenia, with enlarged brain ventricles. The Sapap3 knockout mice groom themselves until they bleed, modeling compulsive-behavior syndromes. Because these mice demonstrate physical and molecular similarities to human psychiatric disorders, scientists anticipate their use in more successful pharmaceutical trials in the future.

Immunology is another focus for creation of transgenic and knockout animal models. Many cytokine knockout mice and major histocompatability class (MHC) I and II transgenic mice have been created, and many of these genetically manipulated mice present with autoimmune disease (13, 42). Similarly, various transgenic and knockout mice have been developed by numerous investigators examining hypertension, obesity, diabetes, and other disease processes. Many more models have been created strictly for basic research, examining the function of a specific protein or other factor. In fact, so many genetically manipulated animal models exist that we would be unable to list them all in this chapter. In addition to the major providers of traditional animals, Taconic Farms Inc. produces thousands of transgenic and knockout mice used as animal models of disease (Table 2).

These genetically manipulated animal models exemplify the hope that basic research will lead to the discovery of new mechanisms of disease. Characterizing diverse molecular pathological pathways may result in new families of drugs targeting very different aspects of disease. The goal of this research is to increase the number of different drugs tested in animals that go on to be successful and safe in human trials. In most, if not all cases, transgenic and knockout models do not represent the human condition. Moreover, the phenotypic expression of a transgenic or knockout animal is often dependent upon the animal strain used. Thus, the phenotype is dependent upon the genetic background on which they were created. They can be of extreme value and function for studying the function of a particular factor but have limited utility as a "gold standard" animal model for pharmaceutical or therapeutic purposes.

## 14 ARE ANIMAL MODELS USEFUL?

Many groups of citizens, physicians, and scientists are demanding change in preclinical drug testing. Often these groups focus on using techniques other than animal models, such as human cell/tissue methods and computer modeling. As is often the case when change is required, throwing out old methods for the new is popular but not practical or wise. The use of computer models is limited by the existing knowledge used to create those models. The more information gained from biological and chemical discovery, the more useful the computer models. Their value in drug development is unquestionable, once the structure of the target compound is known, or analysis of several known agonists or antagonists reveals key structural or chemical features that can be used to create newer and potentially better drugs. Once created, however, they must be tested in vitro and in vivo. What often looks good in a computer model has unanticipated results or side effects when tested in a real-world environment.

As discussed in this chapter, much can be gained by making smart use of several good animal models. However, these animal experiments must be augmented with human-based in vitro experiments, good toxicity databases, and computer modeling technology. Differences in animal and human cells do result in differences in the efficacy and toxicology of certain drugs. The old industrial adage does hold true for the development of drugs and other therapeutics—"cheaper, faster, better." It is accomplished by the use of computer models, in vitro testing using human cells or cell lines, and gold standard animal models, before human trials begin.

## REFERENCES

1. Vanderlugt CL, Karandidar NJ, Lenschow DJ, Dal Canto MC, Bluestone JA, Miller SD (1997). Treatment with intact anti-B7-1 mAb during disease remission enhances epitope spreading and exacerbates relapses in R-EAE. *J Neuroimmunol 79*, 113–118.
2. Miller SD, Vanderlugt CL, Lenschow DJ, Pope JG, Karandidar NJ, Dal Canto MC, Bluestone JA (1995).

Blockade of CD28/B7 interaction prevents epitope spreading and clinical relapses of murine EAE. *Immunity 3*, 739–745.

3. Vanderlugt CL, Neville KL, Nidcevich KM, Eagar TN, Bluestone JA, Miller SD (2000). Pathologic role and temporal appearance of newly emerging autoepitopes in relapsing experimental autoimmune encephalomyelitis. *J Immunol 164*, 670–678.
4. Podojil JR, Kohm AP, Miller SD (2006). $CD4^+$ T cell expressed CD80 regulates central nervous system effector function and survival during experimental autoimmune encephalomyelitis. *J Immunol 177*, 2948–2958.
5. Vincenti F (2008). Costimulation blockade in autoimmunity and transplantation. *J Allergy Clin Immunol 121*, 299–306.
6. Karandikar NJ, Eagar TN, Vanderlugt CL, Bluestone JA, Miller SD (2000). CTLA-4 downregulated epitope spreading and mediates remission in relapsing experimental autoimmune encephalomyelitis. *J Neuroimmunol 109*, 173–180.
7. Miller SD, VanderLugt-Castaneda CL (2004). Theiler's murine encephalomyelitis virus-induced demyelinating disease (TMEV_IDD) and autoimmunity. In Shoenfeld Y, Rose N, Eds. *Infections and Autoimmunity*. pp. 253–263. Elsevier, Amsterdam - Boston - Heidelberg - London - New York - Oxford - Paris - San Diego - San Francisco - Singapore - Sydney.
8. Howard LM, Kohm AP, Castaneda CL, Miller SD (2005). Therapeutic blockade of TCR signal transduction and co-stimulation in autoimmune disease. *Curr Drug Targets Inflamm Allergy 4*, 85–94.
9. Balasa B, Krahl T, Patstone G, Lee J, Tisch R, McDevitt HO, Sarvetnick N (1997). CD40 ligand-CD40 interactions are necessary for the initiation of insulitis and diabetes in nonobese diabetic mice. *J Immunonal 159*, 4620–4627.
10. Howard LM, Miga AJ, Vanderlugt CL, Dal Canto MC, Laman JD, Noelle RJ, Miller SD (1999). Mechanisms of immunotherapeutic intervention by anti-CD40L (CD154) antibody in an animal model of multiple sclerosis. *J Clin Invest 103*, 281–290.
11. Howard LM, Neville KL, Haynes LM, Dal Canto MC, Miller SD (2003). CD154 blockade results in transient reduction in Theiler's murine encephalomyelitis virus-induced demyelinating disease. *J Virol 77*, 2247–2250.
12. Shoda LKM, Young DL, Ramanujan S, Whiting CC, Atkinson MA, Bluestone JA, Eisenbarth GS, Mathis D, Rossini AA, Campbell SE, Kahn R, Kreuwel HTC (2005). A comprehensive review of interventions in the NOD mouse and implications for translation. *Immunity 23*, 115–126.
13. Bour-Jordan H, Bluestone JA (2007). B cell depletion: A novel therapy for autoimmune diabetes? *J Clin Invest 117*, 3642–3645.
14. Chatenoud L, Bluestone JA (2007). CD3-specific antibodies: A portal to the treatment of autoimmunity. *Nature Rev Immunol 7*, 622–632.
15. Chatenoud L (2003). CD3-specific antibody-induced active tolerance: From bench to bedside. *Nature Rev Immunol 3*, 123–131.
16. Kohm AP, Williams JS, Bickford AL, McMahon JS, Chatenoud L, Bach JF, Bluestone JA, Miller SD (2005). Treatment with nonmitogenic anti-cd3 monoclonal antibody induces CD4+ T cell unresponsiveness and functional reversal of established experimental autoimmune encephalomyelitis. *J Immunol 174*, 4525–4534.
17. Tischner D, Weishaupt A, van den Brandt J, Muller N, Beyersdorf N, Wang Ip C, Toyka KV, Hung T, Gold R, Kerkau T, Reichardt HM (2006). Polyclonal expansion of regulatory T cells interferes with effector cell migration in a model of multiple sclerosis. *Brain 129*, 2635–2647.
18. Rodriguez-Palmero M, Franch A, Castell M, Pelegri C, Perez-Cano FJ, Kleinschnitz C, Stoll G, Hunig T, Castellote C (2006). Effective treatment of adjuvant arthritis with a stimulatory CD28-specific monoclonal antibody. *J Rheumatol 33*, 110–118.
19. Hopkin M (2006). Can super-antibody drugs be tamed? *Nature 440*, 338–339.
20. Beyersdorf N, Gaupp S, Balbach K, Schmidt J, Toyka KV, Lin CH, Hanke T, Hunig T, Kerkau T, Gold R (2005). Selective targeting of regulatory T cells with CD28 superagonists allows effective therapy of experimental autoimmune encephalomyelitis. *J Exp Med 202*, 445–455.
21. Theien BE, Vanderlugt CL, Eagar TN, Nickerson-Nutter C, Nazareno R, Kuchroo VK, Miller SM (2001). Discordant effects of anti-VLA treatment before and after onset of relapsing experimental autoimmune enceophalomyelitis. *J Clin Invest 107*, 995–1006.
22. Theien BE, Vanderlugt CL, Nickerson-Nutter C, Cornebise M, Scott DM, Perper SJ, Whalley ET, Miller SD (2003). Differential effects of treatment with a small-molecule VLA-4 antagonist before and after onset of relapsing EAE. *Blood 102*, 4464–4471.
23. Ghosh S, Goldin E, Gordon FH, et al. (2003). Natalizumab for active Crohn's disease. *N Engl J Med 348*, 24–32.
24. Miller DN, Khan OA, Sheremata WA, et al. (2003). International Natalizumab Multiple Sclerosis Trial Group. A controlled trial of natalizumab for relapsing multiple sclerosis. *N Engl J Med 348*, 15–23.
25. Khademi M, Stol D, Olsson T, Wallström E (2008). Induction of systemic TNFalpha in natalizumab-treated multiple sclerosis. *Eur J Neurol 15*, 309–312.
26. Clerico M, Rivoiro C, Contessa G, Viglietti D, Durelli L (2008). The therapy of multiple sclerosis with immunemodulating or immunosuppressive drug: A critical evaluation based upon evidence based parameters and published systematic review. *Clin Neurol Neurosurg (9)*, 878–885.
27. Van Assche G, Van Ranst M, Sciot R, Dubois B, Vermeire S, Norman M, Verbeeck J, Geboes K, Robberecht W, Rutgeerts P (2005). Progressive multifocal leukoencephalopathy after natalizumab therapy for Crohn's disease. *N Engl J Med 353*, 362–368.

28. Yao K, Gagnon S, Akhyani N, Williams E, Fotheringham J, Frohman E, Stuve O, Monson N, Racke MK, Jacobson S (2008). Reactivation of human herpesvirus-6 in natalizumab treated multiple sclerosis patients. *PLoS ONE 3*(4), 1–8. e2028.
29. Wang YJ, et al. (2002). Ketoconazole potentiates the antitumor effects of nocodazole: In vivo therapy for human tumor xemografts in nude mice. *Mol Carcinog 34*, 199–210.
30. Stephan MT, et al. (2007). T cell-encoded CD80 and 4-1BBL induce auto- and transcostimulation, resulting in potent tumor rejection. *Nature Med 13*, 1440–1449.
31. Henke E, Perk J, Vider J, de Candia P, Chin Y, Solit DB, Ponomarev V, Cartegni L, Manova K, Rosen N, Benezra R (2008). Peptide-conjugated antisense oligonucleotides for targeted inhibition of a transcriptional regulator in vivo. *Nature Biotechnol 26*, 91–100.
32. Flowers MT, Ntambi JM (2008). Role of stearoyl-coenzyme A desaturase in regulating lipid metabolism. *Curr Opin Lipidol 19*, 248–256.
33. Vogel GM, van Amsterdam RG, Zandberg P, van Houwelingen P, Kop WJ, van Mensvoort FW, Meuleman DG (1997). Two new closely related rat models with relevance to arterial thrombosis—Efficacies of different antithrombotic drugs. *Thromb Haemost 77*, 183–189.
34. Lukic-Panin V, et al. (2007). Prevention of neuronal damage by calcium channel blockers with antioxidative effects after transient focal ischemia in rats. *Brain Res 1176*, 143–150.
35. Jiang S, et al. (2007). Inhibitory effects of Moutan cortex on immediate allergic reactions. *Biol Pharm Bull 30*, 1701–1710.
36. Bartlett NW, Walton RP, Edwards MR, Aniscenko J, Caramori G, Zhu J, Glanville N, Choy KJ, Jourdan P, Burnet J, Tuthill TJ, Pedrick MS, Hurle MJ, Plumpton C, Sharp NA, Bussell JN, Swallow DM, Schwarze J, Guy B, Almond JW, Jeffery PK, Lloyd CM, Papi A, Killington RA, Rowlands DJ, Blair ED, Clarke NJ, Johnston SL (2008). Mouse models of rhinovirus-induced disease and exacerbation of allergic airway inflammation. *Nature Med 14*, 199–204.
37. Nacionales DC, et al. (2007). Deficiency of the type I interferon receptor protects mice from experimental lupus. *Arthritis Rheum 56*, 3770–3783.
38. Thibault DL, Chu AD, Graham KL, Balboni I, Lee LY, Kohlmoos C, Landrigan A, Higgins JP, Tibshirani R, Utz PJ (2008). IRF9 and STAT1 are required for IgG autoantibody production and B cell expression of TLR7 in mice. *J Clin Invest 118*, 1417–1426.
39. Hsu HC, Yang P, Wang J, Wu Q, Myers R, Chen J, Yi J, Guentert T, Tousson A, Stanus AL, Le TV, Lorenz RG, Xu H, Kolls JK, Carter RH, Chaplin DD, Williams RW, Mountz JD (2008). Interleukin 17-producing T helper cells and interleukin 17 orchestrate autoreactive germinal center development in autoimmune BXD2 mice. *Nat Immunol 9*, 124–126.
40. Rodríguez-Palmero M, Franch A, Castell M, Pelegrí C, Pérez-Cano FJ, Kleinschnitz C, Stoll G, Hünig T, Castellote C (2006). Effective treatment of adjuvant arthritis with a stimulatory CD28-specific monoclonal antibody. *J Rheumatol 33*, 110–118.
41. Whalen BJ, Mordes JP, Rossini AA (2001). The BB rat as a model of human insulin-dependent diabetes mellitus. *Curr Protoc Immunol* Chapter 15, Unit 15.3.
42. Shih DQ, Targan SR (2008). Immunopathogenesis of inflammatory bowel diseases. *World J Gastroenterol 14*, 390–400.
43. Smith JA, Tang Q, Bluestone JA (1998). Partial TCR signals delivered by FcR-nonbinding anti-CD3 monoclonal antibodies differentially regulate individual Th subsets. *J Immunol 160*, 4841–4849.
44. Romani L, Fallarino F, De Luca A, Montagnoli C, D'Angelo C, Zelante T, Vacca C, Bistoni F, Fioretti MC, Grohmann U, Segal BH, Puccetti P (2008). Defective tryptophan catabolism underlies inflammation in mouse chronic granulomatous disease. *Nature 451*, 211–215.
45. Bubeck Wardenburg J, Patel RJ, Schneewind O (2007). Surface proteins and exotoxins are required for the pathogenesis of *Staphylococcus aureus* pneumonia. *Infect Immunonol 75*, 1040–1044.
46. Aujla SJ, Chan YR, Zheng M, Fei M, Askew DJ, Pociask DA, Reinhart TA, McAllister F, Edeal J, Gaus K, Husain S, Kreindler JL, Dubin PJ, Pilewski JM, Myerburg MM, Mason CA, Iwakura Y, Kolls JK (2008). IL-22 mediates mucosal host defense against Gram-negative bacterial pneumonia. *Nat Med 14*, 247–249.
47. Wang R, Braughton KR, Kretschmer D, Bach TH, Queck SY, Li M, Kennedy AD, Dorward DW, Klebanoff SJ, Peschel A, DeLeo FR, Otto M (2007). Identification of novel cytolytic peptides as key virulence determinants for community-associated MRSA. *Nat Med 13*, 1418–1420.
48. Bartlett NW, Walton RP, Edwards MR, Aniscenko J, Caramori G, Zhu J, Glanville N, Choy KJ, Jourdan P, Burnet J, Tuthill TJ, Pedrick MS, Hurle MJ, Plumpton C, Sharp NA, Bussell JN, Swallow DM, Schwarze J, Guy B, Almond JW, Jeffery PK, Lloyd CM, Papi A, Killington RA, Rowlands DJ, Blair ED, Clarke NJ, Johnston SL (2008). Mouse models of rhinovirus-induced disease and exacerbation of allergic airway inflammation. *Nat Med 14*, 199–204.
49. Otto M (2008). Staphylococcal biofilms. *Curr Top Microbiol Immunol 322*, 207–228.
50. Rittenhouse S, Singley C, Hoover J, Page R, Payne D (2006). Use of the surgical wound infection model to determine the efficacious dosing regimen of retapamulin, a novel topical antibiotic. *Antimicrob Agents Chemother 50*, 3886–3888.
51. Raun K, von Voss P, Knudsen LB (2007). Liraglutide, a once-daily human glucagon-like peptide-1 analog, minimizes food intake in severely obese minipigs. *Obesity (Silver Spring) 15*, 1710–1716.

52. Roman AJ, Boye SL, Aleman TS, Pang JJ, McDowell JH, Boye SE, Cideciyan AV, Jacobson SG, Hauswirth WW (2007). Electroretinographic analyses of Rpe65-mutant rd12 mice: Developing an in vivo bioassay for human gene therapy trials of Leber congenital amaurosis. *Mol Vis 13*, 1701–1710.
53. Davies VJ, Powell KA, White KE, Yip W, Hogan V, Hollins AJ, Davies JR, Piechota M, Brownstein DG, Moat SJ, Nichols PP, Wride MA, Boulton ME, Votruba M (2007). A missense mutation in the murine Opa3 gene models human Costeff syndrome. *Brain 131*, 368–380.
54. Darabi R, Gehlbach K, Bachoo RM, Kamath S, Osawa M, Kamm KE, Kyba M, Perlingeiro RC (2008). Functional skeletal muscle regeneration from differentiating embryonic stem cells. *Nat Med 14*, 134–143.
55. Micheli F, Bonanomi G, Blaney FE, Braggio S, Capelli AM, Checchia A, Curcuruto O, Damiani F, Fabio RD, Donati D, Gentile G, Gribble A, Hamprecht D, Tedesco G, Terreni S, Tarsi L, Lightfoot A, Stemp G, Macdonald G, Smith A, Pecoraro M, Petrone M, Perini O, Piner J, Rossi T, Worby A, Pilla M, Valerio E, Griffante C, Mugnaini M, Wood M, Scott C, Andreoli M, Lacroix L, Schwarz A, Gozzi A, Bifone A, Ashby CR Jr, Hagan JJ, Heidbreder C (2007). 1,2,4-Triazol-3-yl-thiopropyl-tetrahydrobenzazepines: A series of potent and selective dopamine D(3) receptor antagonists. *J Med Chem 50*, 5076–5089.
56. Bitner RS, Bunnelle WH, Anderson DJ, Briggs CA, Buccafusco J, Curzon P, Decker MW, Frost JM, Gronlien JH, Gubbins E, Li J, Malysz J, Markosyan S, Marsh K, Meyer MD, Nikkel AL, Radek RJ, Robb HM, Timmermann D, Sullivan JP, Gopalakrishnan M (2007). Broad-spectrum efficacy across cognitive domains by alpha7 nicotinic acetylcholine receptor agonism correlates with activation of ERK1/2 and CREB phosphorylation pathways. *J Neurosci 27*, 10578–10587.
57. Grilly DM, Simon BB, Levin ED (2000). Nicotine enhances stimulus detection performance of middle- and old-aged rats: A longitudinal study. *Pharmacol Biochem Behav 65*, 665–670.
58. Levbin ED, Simon BB (1998). Nicotinic acetylcholine involvement in cognitive function in animals. *Psychopharmacol (Berl) 138*, 217–230.
59. Van Kampen M, Selbach K, Schneider R, Schiegel E, Boess F, Schreiber R (2004). AR-R 17779 improves social recognition in rats by activation of nicotinic alpha7 receptors. *Psychopharmacol (Berl) 172*, 375–383.
60. Dodart JC, May P (2005). Overview on rodent models of Alzheimer's disease. *Curr Protoc Neurosci* Chapter 9, Unit 9.22.
61. Diaz F, Garcia S, Hernandez D, Regev A, Rebelo A, Oca-Cossio J, Moraes CT (2008). Pathophysiology and fate of hepatocytes in a mouse model of mitochondrial hepatopathies. *Gut 57*, 232–242.
62. Ji J, Bunnelle WH, Anderson DJ, Faltynek C, Dyhring T, Ahring PK, Rueter LE, Curzon P, Buckley MJ, Marsh KC, Kempf-Grote A, Meyer MD (2007). A-366833: A novel nicotinonitrile-substituted 3,6-diazabicyclo[3.2.0]-heptane alpha4beta2 nicotinic acetylcholine receptor selective agonist: Synthesis, analgesic efficacy and tolerability profile in animal models. *Biochem Pharmacol 74*, 1253–1262.
63. Blackburn-Munro G, Erichsen HK (2005). Antiepileptics and the treatment of neuropathic pain: Evidence from animal models. *Curr Pharm Des 11*, 2961–2976.
64. Rode F, Thomsen M, Brøløs T, Jensen DG, Blackburn-Munro G, Bjerrum OJ (2007). The importance of genetic background on pain behaviours and pharmacological sensitivity in the rat spared serve injury model of peripheral neuropathic pain. *Eur J Pharmacol 564*, 103–111.
65. Powell SB, Geyer MA (2007). Overview of animal models of schizophrenia. *Curr Protoc Neurosci* Chapter 9, Unit 9.24.
66. Mu J, Yang ZS, Xie P (2008). Proteomic analysis of a rat model of depression. *Expert Rev Proteomics 5*, 315–320.
67. Anisman H, Merali Z (2001). Rodent models of depression: Learned helplessness induced in mice. *Curr Protoc Neurosci* Chapter 8, Unit 8.10C.
68. Stack EC, Ferro JL, Kim J, Del Signore SJ, Goodrich S, Matson S, Hunt BB, Cormier K, Smith K, Matson WR, Ryu H, Ferrante RJ (2008). Therapeutic attenuation of mitochondrial dysfunction and oxidative stress in neurotoxin models of Parkinson's disease. *Biochim Biophys Acta 1782*, 151–162.
69. Argilés JM (1989). The obese Zucker rat: A choice for fat metabolism 1968–1988: Twenty years of research on the insights of the Zucker mutation. *Prog Lipid Res 28*, 53–66.
70. Kasiske BL, O'Donnell MP, Keane WF (1992). The Zucker rat model of obesity, insulin resistance, hyperlipidemia, and renal injury. *Hypertension 19*(1, Suppl), I110–115.
71. Van Zwieten PA, Kam KL, Pijl AJ, Hendriks MG, Beenen OH, Pfaffendorf M (1996). Hypertensive diabetic rats in pharmacological studies. *Pharmacol Res 33*(2), 95–105.
72. Plotkin BJ, Paulson D (1996). Zucker rat (fa/fa), a model for the study of immune function in type-II diabetes mellitus: Effect of exercise and caloric restriction on the phagocytic activityof macrophages. *Lab Anim Sci 46*(6), 682–684.
73. Stevenson FT, Kaysen GA (1999). Hyperlipidemia and renal disease: The use of animal models in understanding pathophysiology and approaches to treatment. *Wien Klin Wochenschr 111*(8), 307–314.
74. Mayers RM, Leighton B, Kilgour E (2005). PDH kinase inhibitors: A novel therapy for Type II diabetes? *Biochem Soc Trans 33*(Pt 2), 367–370.
75. Abbott A (2007). Model behaviour. *Nature 450*, 6–7.
76. Knorr A, Garthoff B, Kazda S, Stasch JP (1991). Effects of different antihypertensive drug classes on survival in animal models. *J Cardiovasc Pharmacol 17*(Suppl 2), S94–100.
77. Rapp JP (1982). Dahl salt-susceptible and salt-resistant rats. A review. *Hypertension 4*(6), 753–763.

78. Zeng C, Sanada H, Watanabe H, Eisner GM, Felder RA, Jose PA (2004). Functional genomics of the dopaminergic system in hypertension. *Physiol Genomics 19*(3), 233–246.
79. Kunes J, Hojná S, Kadlecová M, Dobesová Z, Rauchová H, Vokurková M, Loukotová J, Pechánová O, Zicha J (2004). Altered balance of vasoactive systems in experimental hypertension: The role of relative NO deficiency. *Physiol Res 53*(Suppl 1), S23–34.
80. Lee SJ, Cicila GT (2002). Functional genomics in rat models of hypertension: Using differential expression and congenic strains to identify and evaluate candidate genes. *Crit Rev Eukaryot Gene Exp 12*(4), 297–316.
81. Kato N (1999). Genetic analysis in Dahl salt-sensitive rats. *Clin Exp Pharmacol Physiol 26*(7), 539–540.
82. Sanders BJ, Lawler JE (1992). The borderline hypertensive rat (BHR) as a model for environmentally-induced hypertension: A review and update. *Neurosci Biobehav Rev 16*(2), 207–217.
83. Lawler JE, Barker GF, Hubbard JW, Schaub RG (1980). Pathophysiological changes associated with stress-induced hypertension in the borderline hypertensive rat. *Clin Sci (Lond) 59*(Suppl 6), 307s–310s.
84. Lawler JE, Cox RH (1985). The borderline hypertensive rat (BHR): A new model for the study of environmental factors in the development of hypertension. *Pavlov J Biol Sci 20*(3), 101–115.
85. Sanders BJ, Gray MJ (1997). Early environmental influences can attenuate the blood pressure response to acute stress in borderline hypertensive rats. *Physiol Behav 61*(5), 749–754.
86. Coste SC, Qi Y, Brooks VL, McCarron DA, Hatton DC (1995). Captopril and stress-induced hypertension in the borderline hypertensive rat. *J Hypertens 13*(12, Pt 1), 1391–1398.
87. Xiao J, Pang PK (1995). Hypertension is not related to suppressed lymphocyte proliferation but to elevated NO synthesis in vascular smooth muscle cells of borderline hypertensive rat. *Blood Press 4*(4), 249–256.
88. Winecoff AP, Tackett RL, Fuchs LC (1995). Insulin attenuates alpha-adrenergic aortic contraction in normotensive but not borderline hypertensive rats. *Clin Exp Hypertens J17*(5), 787–801.
89. DiBona GF, Jones SY (1992). Effect of dietary potassium chloride in borderline hypertensive rats. *J Am Soc Nephrol 3*(2), 188–195.
90. Melby CJ, Sanders BJ, Johnson AK (1991). The effects of exercise training on salt-induced hypertension in the borderline hypertensive rat. *Clin Exp Hypertens A 13*(5), 865–872.
91. Orlov SN, Adragna NC, Adarichev VA, Hamet P (1999). Genetic and biochemical determinants of abnormal monovalent ion transport in primary hypertension. *Am J Physiol 276*(3, Pt 1), C511–536.
92. Cowley AW Jr (1997). Genetic and nongenetic determinants of salt sensitivity and blood pressure. *Am J Clin Nutr 65*(2, Suppl), 587S–593S.
93. Sanders PW (1996). Salt-sensitive hypertension: lessons from animal models. *Am J Kidney Dis 28*(5), 775–782.
94. Nabika T, Cui Z, Masuda J (2004). The stroke-prone spontaneously hypertensive rat: How good is it as a model for cerebrovascular diseases? *Cell Mol Neurobiol 24*(5), 639–646.
95. Rubattu S, Stanzione R, Gigante B, Volpe M (2004). Role of genetic factors in the etiopathogenesis of cerebrovascular accidents: From an animal model to the human disease. *Cell Mol Neurobiol 24*(5), 581–588.
96. Baumbach GL, Chillon JM (2000). Effects of angiotensin-converting enzyme inhibitors on cerebral vascular structure in chronic hypertension. *J Hypertens Suppl 18*(1), S7–11.
97. Richer C, Vacher E, Fornes P, Giudicelli JF (1997). Antihypertensive drugs in the stroke-prone spontaneously hypertensive rat. *Clin Exp Hypertens 19*(5–6), 925–936.
98. Yamori Y, Nara Y, Mizushima S, Murakami S, Ikeda K, Sawamura M, Nabika T, Horie R (1992). Gene-environment interaction in hypertension, stroke and atherosclerosis in experimental models and supportive findings from a world-wide cross-sectional epidemiological survey: A WHO-cardiac study. *Clin Exp Pharmacol Physiol Suppl 20*, 43–52.
99. Razzetti R, Acerbi D (1995). Pharmacokinetic and pharmacologic properties of delapril, a lipophilic nonsulfhydryl angiotensin-converting enzyme inhibitor. *Am J Cardiol 75*(18), 7F–12F.
100. Nishida Y, Takahashi Y, Sugahara-Kobayashi M, Ishikawa K, Asai S (2008). Decreased expression of angiotensin II type 1 and type 2 receptors in the brain after long-term administration of antihypertensive drugs in stroke-prone spontaneously hypertensive rat. *J Pharmacol Sci 106*(4), 663–666.
101. Pu Q, Brassard P, Javeshghani DM, Iglarz M, Webb RL, Amiri F, Schiffrin EL (2008). Effects of combined AT1 receptor antagonist/NEP inhibitor on vascular remodeling and cardiac fibrosis in SHRSP. *J Hypertens 26*(2), 322–333.
102. Nagaoka A, Iwatsuka H, Suzuoki Z, Okamoto K (1976). Genetic predisposition to stroke in spontaneously hypertensive rats. *Am J Physiol 230*(5), 1354–1359.
103. Lewis JL, Russell RJ, Warnock DG (1994). Analysis of the genetic contamination of salt-sensitive Dahl/Rapp rats. *Hypertension 24*(3), 255–259.
104. Herberg L, Coleman DL (1977). Laboratory animals exhibiting obesity and diabetes syndromes. *Metabolism 26*(1), 59–99.
105. Jeanrenaud B (1979). Animal obesities. *Ann Nutr Aliment 33*(1), 27–38.
106. Koteish A, Mae Diehl A (2002). Animal models of steatohepatitis. *Best Pract Res Clin Gastroenterol Oct 16*(5), 679–690.

107. Ktorza A, Bernard C, Parent V, Penicaud L, Froguel P, Lathrop M, Gauguier D (1997). Are animal models of diabetes relevant to the study of the genetics of non-insulin-dependent diabetes in humans? *Diabetes Metab 23*(Suppl 2), 38–46.
108. Lindström P (2007). The physiology of obese-hyperglycemic mice [ob/ob mice]. *Sci World 7*, 666–685.
109. Samad F, Loskutoff DJ (1997). The fat mouse: A powerful genetic model to study elevated plasminogen activator inhibitor 1 in obesity/NIDDM. *Thromb Haemost 78*(1), 652–655.
110. Tschöp M, Heiman ML (2001). Rodent obesity models: An overview. *Exp Clin Endocrinol Diabetes 109*(6), 307–319.
111. Allen TJ, Cooper ME, Lan HY (2004). Use of genetic mouse models in the study of diabetic nephropathy. *Curr Diab Rep 4*(6), 435–440.
112. Aoki K, Nakajima A, Mukasa K, Osawa E, Mori Y, Sekihara H (2003). Prevention of diabetes, hepatic injury, and colon cancer with dehydroepiandrosterone. *J Steroid Biochem Mol Biol 85*(2–5), 469–472.
113. Bahary N, Leibel RL, Joseph L, Friedman JM (1990). Molecular mapping of the mouse db mutation. *Proc Natl Acad Sci USA 87*(21), 8642–8646.
114. Carley AN, Severson DL (2005). Fatty acid metabolism is enhanced in type 2 diabetic hearts. *Biochim Biophys Acta 1734*(2), 112–126.
115. Levine DZ (2006). Hyperfiltration, nitric oxide, and diabetic nephropathy. *Curr Hypertens Rep 8*(2), 153–157.
116. Roesler WJ, Pugazhenthi S, Khandelwal RL (1990). Hepatic glycogen metabolism in the db/db mouse. *Mol Cell Biochem 92*(2), 99–106.
117. Sharma K, McCue P, Dunn SR (2003). Diabetic kidney disease in the db/db mouse. *Am J Physiol Renal Physiol 284*(6), F1138–1144.
118. Sun G (2007). Application of DNA microarrays in the study of human obesity and type 2 diabetes. *OMICS 11*(1), 25–40.
119. Ye M, Wysocki J, Naaz, P, Salabat MR, LaPointe MS, Batlle D (2004). Increased ACE2 and decreased ACE protein in renal tubules from diabetic (db/db) mice: A renoprotective combination? *Hypertension 43*(5), 1120–1125.
120. An D, Rodrigues B (2006). Role of changes in cardiac metabolism in development of diabetic cardiomyopathy. *Am J Physiol Heart Circ Physiol l291*(4), H1489–1506.
121. Rees DA, Alcolado JC (2005). Animal models of diabetes mellitus. *Diabet Med 22*(4), 359–370.
122. Tesch GH, Allen TJ (2007). Rodent models of streptozotocin-induced diabetic nephropathy. *Nephrology (Carlton) 12*(3), 261–266.
123. Tesch GH, Nikolic-Paterson DJ (2006). Recent insights into experimental mouse models of diabetic nephropathy. *Nephron Exp Nephrol 104*(2), e57–62.
124. Van Zwieten PA, Kam KL, Pijl AJ, Hendriks MG, Beenen OH, Pfaffendorf M (1996). Hypertensive diabetic rats in pharmacological studies. *Pharmacol Res 33*(2), 95–105.
125. Wu KK, Huan Y (2007). Diabetic atherosclerosis mouse models. *Atherosclerosis 191*(2), 241–249.
126. Amenta F, Di Tullio MA, Tomassoni D (2003). Arterial hypertension and brain damage—Evidence from animal models (review). *Clin Exp Hypertens 25*(6), 359–380.
127. Batlle D, Redon J, Gutterman C, LaPointe M, Saleh A, Sharma A, Rombola G, Ye M, Alsheikha W, Gomez L, et al. (1994). Acid-base status and intracellular pH regulation in lymphocytes from rats with genetic hypertension. *J Am Soc Nephrol 5*(5, Suppl 1), S12–22.
128. Bing OH, Conrad CH, Boluyt MO, Robinson KG, Brooks WW (2002). Studies of prevention, treatment and mechanisms of heart failure in the aging spontaneously hypertensive rat. *Heart Fail Rev 7*(1), 71–88.
129. Kunes J, Hojná S, Kadlecová M, Dobesová Z, Rauchová H, Vokurková M, Loukotová J, Pechánová O, Zicha J (2004). Altered balance of vasoactive systems in experimental hypertension: The role of relative NO deficiency. *Physiol Res 53*(Suppl 1), S23–34.
130. LaPointe MS, Sodhi C, Sahai A, Batlle D (2002). $Na^+/H^+$ exchange activity and NHE-3 expression in renal tubules from the spontaneously hypertensive rat. *Kidney Int 62*(1), 157–165.
131. LaPointe MS, Ye M, Bacallao R, Batlle D (1997). Expression of NHE-1 protein in vascular smooth muscle and lymphocytes from SHR and WKY rats. *Hypertension 30*(4), 880–885.
132. Lerman LO, Chade AR, Sica V, Napoli C (2005). Animal models of hypertension: An overview. *J Lab Clin Med 146*(3), 160–173.
133. Phillips MI (2001). Gene therapy for hypertension: sense and antisense strategies. *Expert Opin Biol Ther 1*(4), 655–662.
134. Reckelhoff JF, Fortepiani LA (2004). Novel mechanisms responsible for postmenopausal hypertension. *Hypertension 3*(5), 918–923.
135. Shimamoto K, Ura N (2006). Mechanisms of insulin resistance in hypertensive rats. *Clin Exp Hypertens 28*(6), 543–552.
136. Skov K, Fenger-Grøn J, Mulvany MJ (1996). Effects of an angiotensin-converting enzyme inhibitor, a calcium antagonist, and an endothelin receptor antagonist on renal afferent arteriolar structure. *Hypertension 28*(3), 464–471.

# DISCOVERY OF VACCINE ADJUVANTS

Fiona A. Sharp[1] and Ed C. Lavelle[2]

[1]*Fahmy Lab, Department of Biomedical Engineering, Yale University, New Haven, Connecticut*; [2]*Adjuvant Research Group, School of Biochemistry and Immunology, Trinity College, Dublin, Ireland*

## 1 INTRODUCTION

Vaccination has had a profound effect in alleviating human morbidity and mortality since the approach was widely introduced in the nineteenth century. The objective of vaccination is to administer a vaccine in such a way as to induce a protective immune response against the corresponding pathogen. Live attenuated and inactivated viral and bacterial vaccines were successful against a number of major pathogens including smallpox, polio, measles, and pertussis. However, together with the major advances in molecular genetics and immunology, safety concerns associated with live and whole-cell killed vaccines have driven a quest for "subunit" vaccines that are based on isolated components of microbes. Despite their advantages in terms of safety, these vaccines are generally ineffective when administered alone, necessitating the use of substances (adjuvants) that can enhance immunity. Alum is the most widely used vaccine adjuvant and is very effective at promoting antibody responses and Th2 type T cell responses. However, it is now recognized with a number of major diseases, including tuberculosis and malaria, that strong cell-mediated immunity is important for protection so adjuvants that can effectively drive such responses and not only antibody responses are required [1]. The global vaccine market is projected to grow very strongly over the next decade so there will be an increasing demand for a range of potent and effective vaccine adjuvants.

## 2 HISTORY OF ADJUVANT DISCOVERY

Adjuvants were originally defined as "substances used in combination with a specific antigen that produce more immunity than the antigen alone" [2]. The modern field of adjuvant research emanated from studies carried out by Gaston Ramon at the Pasteur Institute in the early 1920s that demonstrated that the injection of an eclectic variety of substances including tapioca, inorganic salts, oil, and pyogenic bacteria with diphtheria toxoid enhanced antibody responses to the toxin [3, 4]. Soon afterward Glenny and co-workers reported that injection of diphtheria toxoid as an alum precipitate induced significantly higher antibody responses than injection of the toxoid alone [5–7]. This seminal observation on the adjuvant properties of aluminum compounds ushered in the use of these mineral salts, and these still form the basis of the most widely used adjuvants today. Indeed these remain the only vaccine adjuvants currently licensed by the U.S. Food and Drug Administration (FDA).

The next significant advance in adjuvant research was the description by Jules Freund of a water-in-mineral-oil emulsion containing killed mycobacteria [8]. This is commonly known as Freunds complete adjuvant (FCA) and is a highly potent adjuvant although associated with unacceptable side effects. As a result, a formulation without the killed bacteria was developed and has been used clinically [9, 10]. However, newer emulsions have since been developed and Freunds

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

incomplete adjuvant (FIA) is not used in prophylactic vaccination. FCA was the first adjuvant that deliberately included killed pathogens or pathogen-derived factors and thus may be considered the first "immunostimulatory" adjuvant.

The ability of pathogen-derived and other substances to promote inflammatory responses can be traced back to much earlier studies. Lydia Rabinowitsch in 1897 demonstrated that injection of *Mycobacterium butyricum* led to local granulomatous reactions at the injection site and that the extent of this reaction was exacerbated by the addition of paraffin oil or butter [11]. This was supported by subsequent work showing that killed mycobacteria exerted nonspecific immunostimulatory effects [12]. However, the description of the adjuvant properties of bacterial lipopolysaccharide (LPS) [13, 14] and mycobacterial peptidoglycan [15] was a major advance as it demonstrated the immunostimulatory properties of isolated pathogen-derived factors.

The field of vaccine adjuvants is going through a "golden age" at present due in part to the developments in immunology. After almost 100 years of largely empirical developments in adjuvant research and development largely due to a poor understanding of how adjuvants work, the field is now developing on a new more rational plane.

## 3 REQUIREMENTS OF ADJUVANTS

Adjuvants must satisfy a number of criteria to be considered for clinical use. Safety is a central factor, and there is frequently a fine balance between potency and toxicity. Local reactions that have been described with adjuvants have included pain, inflammation, necrosis, lymphadenopathy, granulomas, ulcers, and sterile abscesses [16, 17]. These have mainly been associated with Freunds complete and incomplete adjuvants while aluminum hydroxide and aluminum phosphate generally do not lead to serious local pain or damage. The effects of the adjuvants are influenced by the associated antigens; thus, it is now recognized that the safety of adjuvant–antigen combination must be tested to establish safety on a case-by-case basis [18].

In terms of the immune responses induced, adjuvants can achieve a number of objectives that include enhancement of antibody and T-cell responses, increased duration of response, induction of memory responses, reduction in vaccine dose, facilitate vaccination by alternative routes, and exert control over the type of T-cell or antibody response induced [19]. The ability of adjuvants to reduce the dose of vaccine required to induce protective immunity "dose sparing" has been highlighted recently in relation to the production of sufficient vaccines to enable global coverage with avian influenza vaccines [20]. In many cases vaccines are not optimally effective in neonates and the elderly. For example, there is evidence that influenza vaccines are less effective in aged populations [21]; here again adjuvants may be used to allow successful vaccination of these age groups.

## 4 IMMUNOLOGICAL REQUIREMENTS OF VACCINES

In the past the efficacy of adjuvants, including alum, emulsions, liposomes, and microparticles, was attributed to a "depot effect" enabling sustained presentation of vaccines to the immune system. It is now clear that activation of innate immune responses is essential to drive adaptive immune responses including helper T cells and antibody responses. Dendritic cells (DCs) occupy a central position in studies of the mechanism of adjuvant action as these cells are key in priming naïve T cells [22]. The activation and polarization of T-helper-cell responses by dendritic cells requires three signals: Signal 1 is provided by the interaction between specific peptides presented in the context of major histocompatability complex (MHC) class II to the T-cell receptor, signal 2 comprises the interaction of co-stimulatory molecules on activated DCs with their cognate ligands on T cells, and signal 3 represents polarizing cytokines released by DCs [23]. T-cell polarization is strongly influenced by cytokines produced by DCs but also by other cells including natural killer cells. Interleukin-12 (IL-2) production promotes the differentiation of T helper cells type 1 (Th1 cells), which produce interferon-γ and IL-2 and promote cell-mediated immunity, activation of $CD8^+$ cytotoxic T cells, and complement fixing antibody [24]. Th2 cells are promoted by IL-4 and characterized by production of cytokines including IL-4, IL-5, and IL-13 and provide help for B-cell production of IgG1, IgE, and IgA antibodies (IgE and IgG4 in humans) [24]. A recently described population of Th17 cells produce the proinflammatory cytokine IL-17 and play roles in antimicrobial immunity, particularly in promoting neutrophil infiltration [25]. The differentiation of these cells is promoted by the cytokines IL-23, IL-1, transforming growth factor (TGF)-β, and IL-6. Regulatory T cells including antigen-specific Tr1 and Th3 cells are promoted by IL-10 and TGF-β and play roles in regulating inflammatory responses and maintaining the balance between tolerance and immunity [26].

## 5 ACTIVATION OF INNATE AND ADAPTIVE IMMUNITY BY ADJUVANTS: STRANGER AND DANGER MODELS

Despite the enormous body of work on vaccine-specific immunity and adjuvants, the vital role of innate immunity in the mechanism of adjuvant action and in promoting protective immune responses was not appreciated until relatively recently. In particular, it was the insight of Charles Janeway who observed that adjuvants, which he termed immunologists' "dirty little secret," activated innate immune responses and dictated the ensuing adaptive immune response, which provided an important conceptual framework for the mechanism of adjuvant action [27]. Janeway proposed that conserved elements of infectious organisms, or *pathogen-associated molecular patterns* (PAMPs), were recognized by the immune system by specific receptors, *pattern recognition receptors* (PRR). Many of the adjuvants now in preclinical and clinical trials are based on such PAMPs. The interaction of PAMPs with PRR on dendritic cells leads to activation of the cells and licences them to provide signals 1, 2, and 3 to drive T-cell activation and polarization. The discovery of PRR, including Toll-like receptors, Nod-like receptors, and Rig I-like receptors [28] has shone tremendous light on the mechanism of action of a number of adjuvants.

A competing or possibly complementary theory for how the immune system is activated was formulated in Polly Matzinger's "danger model," which proposed that endogenous mediators released by stressed or damaged cells activated innate and subsequently adaptive immunity [29]. This latter theory may explain how adjuvants and delivery systems that do not derive from PAMPs or bind to specific receptors on immune cells can serve as effective adjuvants. The potential role of danger as proposed by Matzinger on the efficacy of vaccines may be traced back to the 1920s when Ramon made the observation that higher antidiphtheria toxoid antibody titers were induced in horses that developed an abscess at the inoculation site and that generation of an abscess by the injection of unrelated substances with the diphtheria toxoid increased antibody responses against the toxoid [30]. Further support for a role of "danger" in promoting immunity came from a study by Dienes and Schoenheit showing that injection of protein antigens into guinea pigs infected with *Mycobacterium tuberculosis* induced a strong delayed-type hypersensitivity reaction after sensitization with the same antigen [31]. In 1937, Jules Freund reported elevated antibody titers in tuberculous guinea pigs injected with sheep erythrocytes [32]. This prompted studies into the effect of emulsifying an aqueous solution of the antigen ovalbumin in paraffin oil containing killed tubercle bacilli with the aid of a surfactant [32, 33]. The work culminated in the development of Freund's complete adjuvant, a water-in-oil emulsion of mineral oil, the surfactant mannide monooleate, and heat-killed *M. tuberculosis* or derived bacterial components [8]. This work established the efficacy of emulsions in vaccine delivery and demonstrated the added potency of incorporating pathogen-derived factors to the system. This series of studies demonstrated the requirement for pathogen-derived danger signals and/or local inflammatory responses to promote strong adaptive immune responses to antigens.

## 6 TYPES OF ADJUVANTS

There have been many attempts to classify adjuvants into groups, but this is not an easy exercise since many "delivery systems" have immunostimulatory effects and some systems can fulfill a number of functions. However, for the purposes of this chapter, we classified adjuvants into *immunostimulatory adjuvants*, which can in most cases be shown to interact directly with specific receptors on antigen-presenting cells, and *particulate delivery systems*, which are defined principally by their particulate nature. However, an increasing body of evidence indicates that many of these systems also exert specific effects on innate immunity.

## 7 IMMUNOSTIMULATORY ADJUVANTS

### 7.1 Pathogen Derived

***Toll-Like Receptor (TLR) and Nod-Like Receptor (NLR) Agonists*** Toll-like receptors recognize microbial associated molecular patterns and trigger innate immune responses. To date 13 of these receptors have been identified in mammalian cells [28]. Among the first microbial products investigated for immunostimulant activity were components of the tubercle bacillus and other related mycobacteria [34]. A series of biochemical studies ultimately led to the identification and isolation of trehalose dimycolate (TDM), also known as cord factor, which has adjuvant properties [35]. Ribi adjuvants, which are highly potent but only used experimentally, contain this high-molecular-weight glycolipid, which is extracted from saprophytic, noncording *Mycobacterium phlei* cells [36]. The immunostimulatory activity of released cell wall lipids in Bacillus Calmette-Guérin (BCG) is principally associated with trehalose mycolates [37]. Cell wall skeleton (CWS), a mycobacterial cell wall extract from which the loosely bound lipids,

proteins, and carbohydrates have been removed is a potent immunostimulator that works independently of free TDM [38]. The immunostimulatory activities of CWS and whole mycobaterial cells can be at least partly attributed to the presence of a polymerized form of muramyl dipeptide (MDP). MDP is recognized by the Nod-like receptor Nod2, leading to the induction of intracellular signaling cascades and inflammatory responses [39]. MDP has been shown to promote antigen-specific antibody production and cell-mediated immunity in addition to increased nonspecific immunity to bacteria, viruses, fungi, and parasites [40]. MDP was successfully used as an adjuvant in a veterinary vaccine against leishmaniasis [41].

***Lipopolysaccharide-Derived Adjuvants*** Lipopolysaccharide (LPS) is a component of endotoxin in the cell wall of gram-negative bacteria, such as *Escherichia coli*. The discovery that LPS could enhance antibody responses to coadministered antigens was discovered over 60 years ago [14], and the ability of killed bacteria and LPS to mediate nonspecific immunostimulatory effects has a similarly long history [42], although the full significance of these findings has only been appreciated relatively recently. LPS is recognized as a very effective adjuvant for protein antigens and can promote both humoral and cell-mediated immunity. However, endotoxins are unacceptable for human use because of their toxicity. The lipid A component of LPS is responsible for most of its biological activities [43]. Lipid A and certain synthetic derivatives with diminished toxicity resulting from the substitution of fatty acids and removal of phosphate groups increase humoral and cell-mediated immune responses [44, 45]. Monophosphoryl lipid (MPL) A is derivative of LPS that retains immunostimulatory activity but with considerably reduced toxicity and has been widely used clinically [46]. Due to its effective adjuvant properties and safety record in clinical trials, MPL is likely to be approved as an adjuvant in the near future for clinical use and is currently used together with alum in the Fendrix, hepatitis B vaccine from GlaxoSmithKline (GSK), which is approved in Europe [47]. Second-generation, low-toxicity derivatives of LPS are also under study such as RC529 [48], a derivative of which has been included in a candidate HIV (human immunodeficiency virus) vaccine clinical trial [49].

***Other TLR Agonist-Based Adjuvants*** A wide range of other pathogen-derived immunostimulatory adjuvants are being investigated as adjuvants. These include Pam3Cys (TLR1/2 agonist), poly I:C (TLR 3), imidazoquinolines including imiquimod and resiquimod (R848) (TLR7/8 agonists), and unmethylated CpG oligodeoxynucleotides (TLR9). Imidazoquinolines and unmethylated CpG oligonucleotides induce Th1 responses following signaling initiated via TLR4, TLR7/8, and TLR9, respectively. In addition to the well-established use of resiquimod and imiquimod, new synthetic TLR7/8 agonists including 3M-019 have been produced and are being tested [50]. Currently, a CpG adjuvanted vaccine against malaria is also being tested in humans. [51] This field is going through a period of tremendous growth with multiple synthetic ligands for pathogen recognition receptors currently being tested. A full discussion of this area is outside the scope of this chapter; the area has been excellently reviewed recently [52].

***QS21*** QS21 is an immunostimulatory adjuvant but of plant origin, and, in contrast to a number of the foregoing examples, no specific receptor for QS21 has yet been identified. Saponins are a heterogeneous group of sterol glycosides and tripernoid glycosides distributed throughout the bark, stem, roots, and flowers of many plant species. Humans have used saponins for many applications for hundreds of years. For example, saponins were used in Africa to kill snails infected with schistosomiasis [53]. The adjuvant property of saponins was first demonstrated in the early 1920s by Gaston Ramon [4]. In the1950s Espinet included saponins in a veterinary vaccine against foot-and-mouth disease [54]. The adjuvant-active fraction of saponins from the bark of a tree found in South America, *Quilaja saponaria* Molina, was identified by Daalsgard and named Quil A [55]. Further analysis of Quil A showed that it was a heterogeneous mixture of saponins, making it unsuitable for human use. In 1991 Kensil and co-workers isolated 22 fractions from Quil A using reverse-phase high-performance liquid chromatography (RP-HPLC). QS21 [56] has proved to be the most promising as a vaccine adjuvant due to its strong immunostimulatory activity and reduced toxicity. Studies involving QS21 suggest that it acts as an immunostimulatory adjuvant rather than an antigen depot [57, 58]. Murine studies using QS21 as an adjuvant have shown that it has the ability to induce the Th1-associated antibody isotype, IgG2a, and potent cytotoxic T lymphocyte (CTL) activity, giving it an advantage over a number of other adjuvants that lack the ability to induce strong cell-mediated immunity. [59]. QS21 also has the ability to induce mucosal immunity following intranasal immunization [60]. The mechanism of action of Quil A is not currently well understood. With its strong adjuvant activity, low toxicity, and ability to induce cell-mediated responses, QS21 has significant potential for use in human vaccines. A phase I/II trial involving QS21 and melanoma antigens is currently underway in the United States [61]. QS21 has also been tested in a therapeutic vaccine for

advanced pancreatic cancer and colorectal cancer. Phase I has just been completed and results are pending [62]. Thousands of subjects have received QS21, and only minor side effects have been registered such as pain around the injection site. This work is focused principally on therapeutic vaccines for cancer. However, the possible application of QS21 in infectious diseases is also under active investigation. In fact, QS-21 is currently being evaluated in more than 20 vaccine indications, several in late-stage clinical trials [63].

### 7.2 Particulate Vaccine Adjuvants and Delivery Systems

***Mineral Salts*** Aluminum phosphate, aluminum hydroxide (generally referred to as alum), and calcium phosphate are used clinically to enhance immunity to vaccines. Alum has been widely used in human vaccination for more than half a century during which time billions of doses have been administered, particularly in diphtheria, tetanus, and pertussis vaccine (DPT), *Haemophilus influenzae* B and, IPV vaccines [64]. The early work by Glenny and co-workers focused on alum precipitates [5]. However, efficacy with such preparations was variable and pre-formed aluminum-hydroxide hydrated gels were developed that are effective in protein adsorption, can be pre-formed, and act reproducibly [65]. Aluminum phosphate is also used and a pre-formed aluminum phosphate gel preparation is an effective adsorbant and adjuvant [16], and calcium phosphate has also been established as comparable to alum in terms of efficacy [66]. Alum is generally given by intramuscular injection, and it has been reported that granulomas occur if the adjuvant is administered via subcutaneous or intradermal routes [67].

Despite decades of widespread use, the mechanism by which alum activates the immune system is not well understood. Alum is a very effective inducer of antibody responses but is less effective in the induction of cell-mediated immunity. The strong induction of Th2-type responses indicates that the adjuvant may exert specific effects on the immune system in addition to acting as an antigen "depot". Injection of alum promotes the secretion of IL-4 by splenic Gr1-positive cells [68], and its induction of T-cell proliferation relies on IL-1 [69]. Indeed, it was recently found that alum can promote the secretion of IL-1 and IL-18 by antigen-presenting cells ([70] and our unpublished data).

Injection of alum in humans causes adverse reactions, including local pain and fever [66, 71]; however, whether this enhances vaccine efficacy is unknown. The "danger theory" proposes that tissue and cell damage releases intracellular factors with powerful immunostimulatory properties [72]. We have preliminary data demonstrating that tissue necrosis occurs around the alum injection site and that high concentrations of alum are toxic to murine cells (manuscript in preparation). Thus, there may be a role for endogenous danger signals in the mechanism of action of alum and related adjuvants.

Alum is a highly effective and safe adjuvant for a number of vaccines, but it is not universally effective. The adjuvant is not optimally effective for use with peptide vaccines [64], and poor efficacy was documented using alum with influenza and typhoid fever vaccines [73, 74]. Furthermore, due to its Th2 bias in promoting adaptive immunity, alum is not suitable for vaccines against diseases such as tuberculosis, HIV, and malaria where potent cell-mediated immunity is critical. However, there are indications that when alum is combined with other adjuvants, it is more effective in promoting Th1-type T-helper-cell responses [75, 76].

***Emulsions*** The use of emulsions in vaccines to boost immune responses can be traced back to work by LeMoignic Pinoy in 1916 who used the first oil emulsion, water, and vaseline, with inactivated *Salmonella typhimurium*. [77]. Less then 10 years later Ramon demonstrated that starch oil increased antibody responses to diphtheria toxoid [4]. In the 1930s Freund developed paraffin-oil-based emulsions with or without heat-killed *Mycobacteria*, producing Freund's complete or incomplete adjuvant, respectively [32]. Freund's adjuvant is an example of water-in-oil (W/O) emulsion, which is made up of antigen-containing water droplets in a continuous oil phase. Oil-in-water (O/W) emulsions are composed of antigen in a water phase with oil droplets in the solution, and double emulsions can be produced with water droplets trapped within larger oil droplets dispersed in a water phase (W/O/W) [78]. Herbert showed that O/W emulsions have the ability to retain an antigen at the site of injection, allowing prolonged exposure to the immune system [79]. Oil-in-water emulsions have shown the best safety profile, while mineral oil emulsions have proven too toxic for human use. Syntax adjuvant formulation, or SAF, is an oil-in-water adjuvant developed in the 1980s as a possible replacement for Freund's adjuvant. SAF contains the biodegradable oil, squalene, but was too toxic for clinical use due to the presence of muramyl dipeptide [80]. The toxic nature of SAF led to the development of a safer emulsion for adjuvant use known as MF59, a microfluidized O/W emulsion comprising squalene oil with Tween 80 and Span 85 as surfactants but with no additional immunostimulators [81]. MF59 has been shown to be a potent adjuvant with an acceptable safety profile. A large number of subjects have been immunized with MF59, and the adjuvant was well tolerated and was safe in newborns and the elderly [82]. FLUAD is an MF59 adjuvanted influenza vaccine

currently licensed for use in the elderly by 12 European health authorities and in some non–European Union EU countries, with 25 million administered doses to data. [83] Novartis is also using MF59 in a candidate H5N1 vaccine [84] and claims that it can significant reduce the required "dose sparing" and also induce cross-serotype specific immunity. Furthermore, Sanofi-Pasteur claims that its proprietary emulsion-based adjuvant for their candidate H5N1 vaccine facilitated the use of 10-fold less antigen then conventional vaccines [85]. There is strong evidence from a number of companies that emulsion based vaccines are effective for pandemic flu vaccines [86].

***ISCOMs*** In 1971 immune-stimulating complexes, or ISCOMs, were first visualized by electron microscopy in a preparation of virus particles treated with saponin [87]. However, it was not until 1984 that Morein et al. first described the formation of these saponin-containing complexes with viral membrane proteins [88]. Negative stain electron transmission shows these well-defined particles as spherical, cagelike structures of approximately 40 nm in diameter [89]. Their formation occurs spontaneously upon mixing cholesterol, *Quilaja saponaria* saponins, and phospholipids. It is the interaction of the cholesterol and saponins that forms the matrix of these particles, and phospholipids are added to produce a less tightly compressed particle, allowing for the insertion of amphiphatic molecules [90, 91]. ISCOMs are manufactured by and registered to Isotec AB, which has also developed ISCOMATRIX, which consists of cholesterol, phospholipids, and saponins but lacks the viral membrane proteins allowing for the insertion of antigens into the particles [92]. ISCOM-based vaccines have been shown to induce both antibody and Th1-polarized cell-mediated responses and are particularly potent in the induction of cytotoxic T cells [89]. Watson et al. demonstrated that the trafficking of ISCOMs from the site of injection to the spleen and the local draining of lymph nodes was faster than protein alone [93]. However, the exact mode of action of ISCOMs remains largely unknown. Several theories have been suggested to explain their adjuvant activities: (a) delivery of the antigen directly to antigen presenting cells, (b) delivery of the antigen into the cytosol of the cell, allowing access to the MHC I pathway, which is required for $CD8^+$ T-cell development, and (c) induction of cytokine production by antigen-presenting cells; for example, ISCOMs have been shown to induce IL-6 and IL-1 production in macrophages [53, 94]. It has also been found that IL-12 and interferon-γ are key to the adjuvant activity of ISCOMs. [95]. However, to date no vaccine containing ISCOMs or ISCOMATRIX has been licensed for human use. Despite this, over 1000 people have been injected intramuscularly with ISCOM-based vaccines, and it was reported that they were safe and generally well tolerated, and side effects were found to be self-limiting [89]. Several ISCOM-based vaccine formulations are currently in development, including viral proteins from influenza, purified protein recombinants such as HPV16 E6E7 fusion protein, and NY-ESO-1, an antigen expressed in a wide variety of cancers [96, 97]. Results in trials so far involving these antigens have been positive, with 69–92% of subjects developing $CD4^+$ T-cell responses and virtually all subjects producing antibody responses irrespective of preexposure [96, 97].

Although ISCOMs have been widely used for veterinary patients, their reactogenicity is perceived as a problem in relation to clinical application. However, recent work demonstrated that GPI-0100, semisynthetic derivatives of deacylated Quilaja saponins (DS saponins) modified at the glucuronic acid residue retained immunostimulatory properties, indicating that synthetic derivatives with strong immunostimulatory properties may be developed [98]. Two different fractions induced Th1 and Th2 responses, respectively, suggesting that formulations may be tailored to the desired type of response.

***Biodegradable Microparticles*** Biodegradable, particularly poly(lactide-*co*-glycolide) (PLG) microparticles have been used for decades in sutures but also for drug delivery and controlled-release vaccine delivery systems [99]. The ability to enhance the immunogenicity of antigens by encapsulation into PLG microparticles was first demonstrated in the early 1990s [100, 101]. Microparticles can be formulated from a range of different polymers; however, PLG microparticles are the prime candidates for development as vaccine adjuvants [102]. PLG microparticles are biodegradable in vivo, undergoing nonenzymatic hydrolysis of the ester bonds to produce lactic and glycolic acids. The half-life of PLG microparticles is dependent on the molecular weight of the particle and the lactic-to-glycolic ratio and can range from days to weeks and possibly years. The optimal size range for vaccination is likely to depend on the application, but it has been proposed that microparticles of 1–3 μm are optimal in some settings [99]. During the microencapsulation process, antigens can be exposed to various potentially damaging conditions such as low pH, high shear, and organic solvents [103, 104]. This and the susceptibility of encapsulated antigen to acid-mediated hydrolysis [105, 106] encouraged scientists to develop alternative means of associating antigens with microparticles. O'Hagan et al. described the preparation of charged microparticles using sodium dodecyl sulfate (SDS) to produce

anionic PLG microparticles onto which it was easier to adsorb proteins [107]. Factors such as protein charge and hydrophobicity affect antigen adsorption and the optimal adsorption conditions are protein specific [108]. Cationic particles were developed by the same group that are optimized for the delivery of deoxyribonucleic acid (DNA) vaccines [109].

The exact mode of action of PLG microparticles is unclear; however, uptake by macrophages and dendritic cells seems to play an important role in their adjuvanticity [101, 110]. Injection of microparticles leads to uptake by inflammatory monocytes and migration and the differentiation of significant numbers of these cells into dendritic cells in draining lymph nodes [111]. Microparticles facilitate multimeric presentation of antigens on their surface, which is required for optimal B-cell activation. [112]. In addition to promoting strong antibody responses, studies in rodents indicate that PLG microparticles also effectively promote cell-mediated immunity [113, 114]. Cationic PLG microparticles have been tested in nonhuman primates as a delivery system for a hepatitis C vaccine and induced seroconversion in all animals vaccinated [115]. In addition cationic microparticles are currently being evaluated in a clinical trial with a candidate HIV vaccine [116].

***Liposomes*** Liposomes were developed in 1965 by Alec Bangham, who extracted lipids from egg yolk and found that when they were submerged into an aqueous solution they spontaneously rearranged as double-layered bubbles approximately the same size of a cell [117]. These "bubbles" were defined as liposomes, which are particulate structures composed of either natural or synthetic phospholipids. The biomedical applications of liposomes include drug delivery, vaccines, red blood cell substitutes for hemorrhagic injuries. Allison and Gregoriadis were the first to show that encapsulation of an antigen into a liposome led to increased antibody responses [118]. *Conventional liposomes* consist of a mixture of phospholipids and have been shown to be rapidly cleared by phagocytic cells of the reticuloendothelial system and macrophages [119, 120]. This discovery prompted the development of "stealthlike" liposomes that contain a small proportion of sterically stabilizing phospholipids and sometimes cholesterol, to further boost membrane stability and rigidity [121]. The increased stability of these stealthlike liposomes allows them to persist for as long as 2 days in the bloodstream after injection [121]. Physicochemical properties such as charge and fluidity of liposomes have been reported to have an effect on their immunogenicity. It may be possible to use liposomes to target specific cells of the immune system. For example, anionic liposomes with a bilayer composed of phosphatidylcholine, cholesterol, and phosphatidylglycerol or phosphatidylserine have been reported to interact specifically with human and murine dendritic cells [122]. *Immunoliposomes* allow for more specific targeting to cells as they contain a cell-specific monoclonal antibody incorporated into their membrane [122]. It has also been reported that monocytes express receptors for phosphatidylserine and that charged lipids bind selectively to dendritic cells. Findings to date also show that liposomes have the ability to promote access of antigens into both MHC class I and class II antigen presentation pathways, allowing the activation of both $CD4^+$ and $CD8^+$ T cells [123]. In 1992 the first demonstration of the effects of liposomal vaccines in humans was reported. Alving et al. immunized a group of 30 healthy males with a liposome-encapsulated recombinant protein ($R32NS1_{81}$) containing epitopes from *Plasmodium falciparium* [124]. The liposomes used also contained the immunostimulator monophosphoryl lipid A and were adsorbed to aluminum hydroxide. Overall, the vaccine was found to increase R32-specific antibody production compared with previous studies, and all reactions in volunteers were moderate [124].

Four liposome formulations are currently on the market for drug delivery in the treatment of Kaposi's sarcoma, metastatic breast cancer, and fungal infections in immunocompromised individuals [125]. The idea of lipid-based carrier systems for the delivery of antigens in vaccines has lead to the development of several alternatives to pure liposomes. Almeida et al. were the first to describe the intercalation of viral envelope proteins into conventional liposomes in 1975 [126]. Immunopotentiating reconstructed influenza virosomes, or IRIV, have proved to be a promising system for the delivery of antigens. These virosomes contain a fusion-active Hemagglutinin A (HA) protein within the membrane, which mediates pH-induced membrane fusion upon acidification of the endosome [127]. There are currently two marketed virosome-based vaccines in 29 countries worldwide. Inflexal V is a trivalent influenza vaccine, and Epaxal contains inactivated hepatitis A particles [128, 129]. Other lipid-based carrier systems include archeasomes, which contain archaebacterial lipid membranes, and proteasomes formed from reconstituted bacterial outer membrane proteins and cochleates [130]. A vaccine against follicular lymphoma using proteoliposomal technology has recently been used clinically. The proteoliposomes are constructed using proteins directly extracted from autologous tumor cells and lipids producing a patient-specific vaccine that can be formulated within one day. The vaccine was found to be safe and showed promise for this type of technology in the future [131].

## 8 FUTURE OF ADJUVANT RESEARCH AND DEVELOPMENT

Adjuvant research has a long history and many currently used adjuvants or products from which they were derived were identified prior to a mechanistic understanding of their modes of action. With developments in immunology and biochemistry and particularly an appreciation of the central role of innate immunity in adjuvant action, adjuvant research is now progressing on a more rational plain and is going through a golden age. Despite these developments, alum remains the only FDA-approved adjuvant, although this is likely to change and a number of additional adjuvants have been approved in Europe (Table 1). While many new adjuvants are likely to be identified through approaches such as TLR and NLR agonist screens [132] recent evidence indicates that identification even of relatively potent immunostimulatory molecules only constitutes the first step (Fig. 1). There are a number of promising particulate delivery systems in place, and one key approach is identifying synergistic interactions between these systems and immunostimulatory adjuvants. A number of recent clinical studies have demonstrated that the combination of particulate adjuvants such as alum and emulsions with TLR agonists or QS21 (AS02) induces significantly enhanced responses than these adjuvants alone (reviewed in [1]). An example of this is the use of alum and MPL (AS04) in Fendrix to prevent HepB in high-risk patients [47] and human papillomavirus 16/18 virus-like particles (Cervarix) to prevent cervical cancer [133].

While this review has concentrated on parenteral routes of vaccination, there are also great strides forward being made in transcutaneous and mucosal vaccine delivery. In particular, bacterial toxins or their derivatives are potent adjuvants for transcutaneous vaccine delivery [134, 135]. These toxins are also very effective adjuvants for mucosal vaccination, although clinical application has been delayed due to concerns over potential neurotoxic effects [136]. A number of other approaches are also being tested to enhance the efficacy of mucosal delivery of nonliving vaccines, including combinations of particulates and immunostimulatory molecules, protection of antigens from proteolysis, and specific targeting of vaccines to immune inductive sites [19]. In conclusion, this is an exciting time for adjuvant research, and we can expect the increased activity in the field to generate improved adjuvants for clinical use.

**TABLE 1 Examples of Adjuvants That Have Been or Are Currently Being Tested in Humans[a]**

| | Receptor | Applications | Status |
|---|---|---|---|
| | | *Immunostimulatory Adjuvants* | |
| RC529 | TLR4 | HIV | Phase III complete [49] |
| QuilA | ? | Advanced pancreatic and colorectal cancer | Phase I [62] |
| | | Melanoma antigens | Phase I/II [61] |
| MPL | TLR4 | Fendrix, Hep B mixed with alum | Licensed in EU [47] |
| | | Cervarix, HPV | Phase III [123] |
| CpG | TLR9 | Malaria | Phase I [51] |
| | | *Particulate Adjuvants/Delivery Systems* | |
| Mineral salts: Alum | ? | Diphtheria | Licensed [64] |
| | | *Haemophilus influenzae* B | Licensed [64] |
| | | IPV | Licensed [64] |
| Emulsions: MF59 | ? | FLUAD | Licensed [83] |
| | | H5N1 | Phase III [84] |
| ISCOMs | ? | HPV 16 E6E7 | Phase Ib [96] |
| | | NY-ESO-1 | Phase II [97] |
| Microparticles | ? | HIV Clade B | Phase I [116] |
| IRIV-based vaccines | ? | Inflexal V trivalent influenza vaccine | Licensed [128] |
| Proteoliposomes | ? | Epaxal hep A particles | Licensed [129] |
| | | Follicular lymphoma | Phase I complete [130] |

[a]The question marks in the second column refer to the fact that for many commonly used vaccine adjuvants no specific receptor has been identified on cells of the innate immune system.

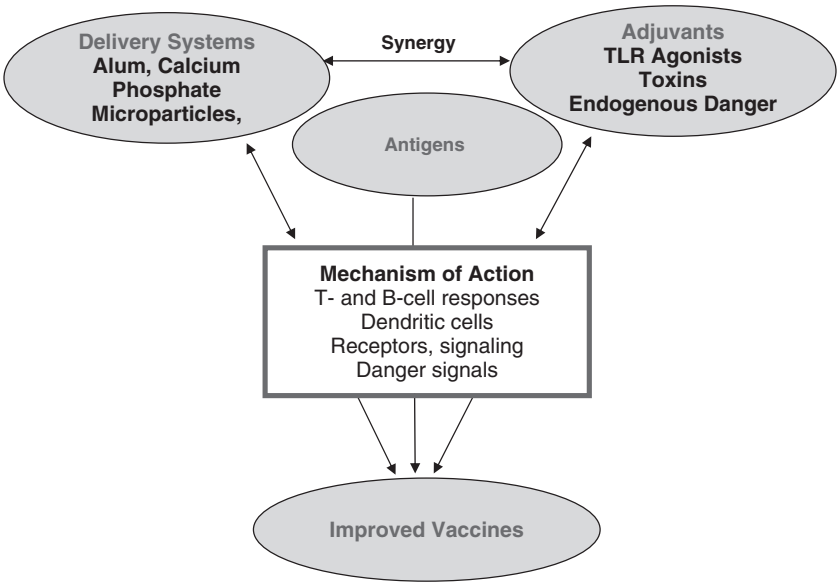


**Figure 1 Strategy for the optimisation of adjuvants for vaccines.** By optimising the choice of immunostimulatory adjuvant and particulate delivery system, optimal, directed activation of innate immunity can be achieved. The type of system can be tailored to the specific requirements of a particular vaccine.

## REFERENCES


1. Guy B (2007). The perfect mix: Recent progress in adjuvant research. *Nat Rev Microbiol 5*, 505–517.
2. Ramon G (1924). Sur la toxine et sur l'anatoxine diphtheriques. *Ann Inst Pasteur 38*, 1–10.
3. Ramon G (1925). Sur l'augmentation anormale de l'antitoxine chez les chevaux producteurs de serum antidiphterique. *Bull Soc Centr Med Vet 101*, 227–234.
4. Ramon G (1926). Procedes pour accroître la production des antitoxins. *Ann Inst Pasteur 40*, 1–10.
5. Glenny A, Pope C, Waddington H, Falacce U (1926). The antigenic value of toxoid precipitated by potassium alum. *J Pathol Bacteriol 29*, 31–40.
6. Glenny A, Pope C, Waddington H, Wallace U (1926). Immunological notes XVII to XXIV. *J Pathol 29*, 31–40.
7. Glenny A, Buttle G, Stevens M (1931). Rate of disappearance of diphtheria toxoid injected into rabbits and guinea-pigs: Toxoid precipitated with alum. *J Pathol 34*, 267–275.
8. Freund J (1951). The effect of paraffin oil and mycobacteria on antibody formation and sensitization: A review. *Am J Clin Pathol 21*, 645–656.
9. Stuart-Harris CH (1969). Adjuvant influenza vaccines. *Bull World Health Organ 41*, 617–621.
10. Bourdette DN, Edmonds E, Smith C, Bowen JD, Guttmann CR, Nagy ZP, Simon J, Whitham R, Lovera J, Yadav V, Mass M, Spencer L, Culbertson N, Bartholomew RM, Theofan G, Milano J, Offner H, Vandenbark AA (2005). A highly immunogenic trivalent T cell receptor peptide vaccine for multiple sclerosis. *Mult Scler 11*, 552–561.
11. Rabinowitsch L (1897). Zur Frage des Vorkommens von Tuberkelbacillen in der Marktbutter. *Zeitschrift fuer Hygiene 26*, 90–111.
12. Friedmann F (1903). Der Schildkroetentuberkelbacilllus, seine Zuechtung, Biologie und Pathogenitaet. *Centralblatt für Bakteriologie 34*, 793–804.
13. Condie RM, Zak SJ, Good RA (1955). Effect of meningococcal endotoxin on the immune response. *Proc Soc Exp Biol Med 90*, 355–360.
14. Johnson AG, Gaines S, Landy M (1956). Studies on the O antigen of *Salmonella typhosa*. V. Enhancement of antibody response to protein antigens by the purified lipopolysaccharide. *J Exp Med 103*, 225–246.
15. Adam A, Ciorbaru R, Ellouz F, Petit JF, Lederer E (1974). Adjuvant activity of monomeric bacterial cell wall peptidoglycans. *Biochem Biophys Res Commun 56*, 561–567.
16. Gupta RK, Relyveld EH, Lindblad EB, Bizzini B, Ben-Efraim S, Gupta CK (1993). Adjuvants—A balance between toxicity and adjuvanticity. *Vaccine 11*, 293–306.
17. Aguilar JC, Rodriguez EG (2007). Vaccine adjuvants revisited. *Vaccine 25*, 3752–3762.
18. Sesardic D (2006). Regulatory considerations on new adjuvants and delivery systems. *Vaccine 24*, S2-86–87.
19. Lavelle EC (2005). Generation of improved mucosal vaccines by induction of innate immunity. *Cell Mol Life Sci 62*, 2750–2770.

20. Poland GA (2006). Vaccines against avian influenza—A race against time. *N Engl J Med 354*, 1411–1413.
21. Targonski PV, Jacobson RM, Poland GA (2007). Immunosenescence: Role and measurement in influenza vaccine response among the elderly. *Vaccine 25*, 3066–3069.
22. Steinman RM (2007). Dendritic cells: Versatile controllers of the immune system. *Nat Med 13*, 1155–1159.
23. Kapsenberg ML (2003). Dendritic-cell control of pathogen-driven T-cell polarization. *Nat Rev Immunol 3*, 984–993.
24. Kalinski P, Moser M (2005). Consensual immunity: Success-driven development of T-helper-1 and T-helper-2 responses. *Nat Rev Immunol 5*, 251–260.
25. Steinman L (2007). A brief history of T(H)17, the first major revision in the T(H)1/T(H)2 hypothesis of T cell-mediated tissue damage. *Nat Med 13*, 139–145.
26. Belkaid Y (2007). Regulatory T cells and infection: A dangerous necessity. *Nat Rev Immunol 7*, 875–888.
27. Janeway CA Jr (1989). Approaching the asymptote? Evolution and revolution in immunology. *Cold Spring Harb Symp Quant Biol 54*, 1–13.
28. Creagh EM, O'Neill LA (2006). TLRs, NLRs and RLRs: A trinity of pathogen sensors that co-operate in innate immunity. *Trends Immunol 27*, 352–357.
29. Matzinger P (1994). Tolerance, danger, and the extended family. *Annu Rev Immunol 12*, 991–1045.
30. Ramon G (1959). [Certain works presented at the Academie Nationale de Medecine (Paris) from 1925 to 1950.] *Rev Immunol Ther Antimicrob 23*, 359–401.
31. Dienes L, Schoenheit EW (1927). Local hypersensitiveness. *J Immunol 14*, 9–74.
32. Freund J, Casals J, Hosmer EP (1937). Sensitization and antibody formation after injection of tubercle bacilli and paraffin oil. *Proc Soc Exp Biol 37*, 509–513.
33. Freund J, McDermott K (1942). Sensitization to horse serum by means of adjuvants. *Proc Soc Exp Biol Med 49*, 548–553.
34. Bekierkunst A, Levij IS, Yarkoni E, Vilkas E, Adam A, Lederer E (1969). Granuloma formation induced in mice by chemically defined mycobacterial fractions. *J Bacteriol 100*, 95–102.
35. Noll H, Bloch H, Asselineau J, Lederer E (1956). The chemical structure of the cord factor of *Mycobacterium tuberculosis*. *Biochim Biophys Acta 20*, 299–309.
36. Rudbach JA, Cantrell JL, Ulrich JT (1995). Methods of immunization to enhance the immune response to specific antigens in vivo in preparation for fusions yielding monoclonal antibodies. *Methods Mol Biol 45*, 1–8.
37. Geisel RE, Sakamoto K, Russell DG, Rhoades ER (2005). In vivo activity of released cell wall lipids of *Mycobacterium bovis* bacillus Calmette-Guerin is due principally to trehalose mycolates. *J Immunol 174*, 5007–5015.
38. Azuma I, Yamamura Y, Ribi E (1974). Preparation of an adjuvant-active, tuberculin-free peptidoglycolipid from human tubercle bacilli. *Jpn J Microbiol 18*, 327–332.
39. Kanneganti TD, Lamkanfi M, Nunez G (2007). Intracellular NOD-like receptors in host defense and disease. *Immunity 27*, 549–559.
40. Johnson AG (1994). Molecular adjuvants and immunomodulators: New approaches to immunization. *Clin Microbiol Rev 7*, 277–289.
41. Lemesre JL, Holzmuller P, Cavaleyra M, Goncalves RB, Hottin G, Papierok G (2005). Protection against experimental visceral leishmaniasis infection in dogs immunized with purified excreted secreted antigens of *Leishmania infantum* promastigotes. *Vaccine 23*, 2825–2840.
42. Whitby JL, Michael JG, Woods MW, Landy M (1961). Symposium on bacterial endotoxins. II. Possible mechanisms whereby endotoxins evoke increased nonspecific resistance to infection. *Bacteriol Rev 25*, 437–446.
43. Takada H, Kotani S (1989). Structural requirements of lipid A for endotoxicity and other biological activities. *Crit Rev Microbiol 16*, 477–523.
44. Chiller JM, Skidmore BJ, Morrison DC, Weigle WO (1973). Relationship of the structure of bacterial lipopolysaccharides to its function in mitogenesis and adjuvanticity. *Proc Natl Acad Sci USA 70*, 2129–2133.
45. Kotani S, Takada H, Takahashi I, Ogawa T, Tsujimoto M, Shimauchi H, Ikeda T, Okamura H, Tamura T, Harada K et al. (1986). Immunobiological activities of synthetic lipid A analogs with low endotoxicity. *Infect Immun 54*, 673–682.
46. Mata-Haro V, Cekic C, Martin M, Chilton PM, Casella CR, Mitchell TC (2007). The vaccine adjuvant monophosphoryl lipid A as a TRIF-biased agonist of TLR4. *Science 316*, 1628–1632.
47. Kundi M (2007). New hepatitis B vaccine formulated with an improved adjuvant system. *Expert Rev Vaccines 6*, 133–140.
48. Baldridge JR, McGowan P, Evans JT, Cluff C, Mossman S, Johnson D, Persing D (2004). Taking a toll on human disease: Toll-like receptor 4 agonists as vaccine adjuvants and monotherapeutic agents. *Expert Opin Biol Ther 4*, 1129–1138.
49. Safety of and Immune Response to a New HIV Vaccine: HIV CTL MEP Trial started 2004, last updated 2009. http://www.clinicaltrials.gov/ct/show/NCT00076037.
50. Johnston D, Zaidi B, Bystryn JC (2007). TLR7 imidazoquinoline ligand 3M-019 is a potent adjuvant for pure protein prototype vaccines. *Cancer Immunol Immunother 56*, 1133–1141.
51. Safety of and Immune Response to a Malaria Vaccine (MSP1 42-C1) with or without CPG 7909 Adjuvant Trial started 2006, last updated 2008. http://www.clinicaltrials.gov/ct/show/NCT00320658.
52. Kwissa M, Kasturi SP, Pulendran B (2007). The science of adjuvants. *Expert Rev Vaccines 6*, 673–684.
53. Barr IG, Mitchell GF (1996). ISCOMs (immunostimulating complexes): The first decade. *Immunol Cell Biol 74*, 8–25.

54. Espinet RG (1951). Nouveau vaccin antiaphteux a complexe glucoviral. Gac. Vet. (5. Aires). *Gac Vet 13*, 268–270.
55. Dalsgaard K (1974). Saponin adjuvants. 3. Isolation of a substance from *Quillaja saponaria* Molina with adjuvant activity in food-and-mouth disease vaccines. *Arch Gesamte Virusforsch 44*, 243–254.
56. Kensil CR, Patel U, Lennick M, Marciani D (1991). Separation and characterization of saponins with adjuvant activity from *Quillaja saponaria Molina* cortex. *J Immunol 146*, 431–437.
57. Kensil CR (1996). Saponins as vaccine adjuvants. *Crit Rev Ther Drug Carrier Syst 13*, 1–55.
58. Kensil CR, Soltysik S, Wheeler DA, Wu JY (1996). Structure/function studies on QS-21, a unique immunological adjuvant from *Quillaja saponaria*. *Adv Exp Med Biol 404*, 165–172.
59. Newman MJ, Wu JY, Gardner BH, Anderson CA, Kensil CR, Recchia J, Coughlin RT, Powell MF (1997). Induction of cross-reactive cytotoxic T-lymphocyte responses specific for HIV-1 gp120 using saponin adjuvant (QS-21) supplemented subunit vaccine formulations. *Vaccine 15*, 1001–1007.
60. Sasaki S, Sumino K, Hamajima K, Fukushima J, Ishii N, Kawamoto S, Mohri H, Kensil CR, Okuda K (1998). Induction of systemic and mucosal immune responses to human immunodeficiency virus type 1 by a DNA vaccine formulated with QS-21 saponin adjuvant via intramuscular and intranasal routes. *J Virol 72*, 4931–4939.
61. Ragupathi G, Gardner JR, Livingston PO, Gin DY (2011). Natural and synthetic saponin adjuvant QS21 for vaccines against cancer. *Exp Rev Vaccines 10*(4), 463–470.
62. Vaccine therapy plus QS21 in treating patients with advanced pancreatic and colorectal cancer. http://clinicaltrials.gov/ct2/show/NCT00006387
63. Garçon N, Van Mechelen M (2011). Recent clinical experience with vaccines using MPL- and QS-21-containing adjuvant systems. *Expert Rev Vaccines 10*, 471–486.
64. Lindblad EB (2004). Aluminium adjuvants—In retrospect and prospect. *Vaccine 22*, 3658–3668.
65. Lindblad EB (2004). Aluminium compounds for use in vaccines. *Immunol Cell Biol 82*, 497–505.
66. Aggerbeck H, Wantzin J, Heron I (1996). Booster vaccination against diphtheria and tetanus in man. Comparison of three different vaccine formulations–III. *Vaccine 14*, 1265–1272.
67. Butler NR, Voyce MA, Burland WL, Hilton ML (1969). Advantages of aluminium hydroxide adsorbed combined diphtheria, tetanus, and pertussis vaccines for the immunization of infants. *Br Med J 1*, 663–666.
68. Jordan MB, Mills DM, Kappler J, Marrack P, Cambier JC (2004). Promotion of B cell immune responses via an alum-induced myeloid cell population. *Science 304*, 1808–1810.
69. Grun JL, Maurer PH (1989). Different T helper cell subsets elicited in mice utilizing two different adjuvant vehicles: The role of endogenous interleukin 1 in proliferative responses. *Cell Immunol 121*, 134–145.
70. Li H, Nookala S, Re F (2007). Aluminum hydroxide adjuvants activate caspase-1 and induce IL-1beta and IL-18 release. *J Immunol 178*, 5271–5276.
71. Gupta RK (1998). Aluminum compounds as vaccine adjuvants. *Adv Drug Deliv Rev 32*, 155–172.
72. Matzinger P (2002). The danger model: A renewed sense of self. *Science 296*, 301–305.
73. Davenport FM, Hennessy AV, Askin FB (1968). Lack of adjuvant effect of A1PO4 on purified influenza virus hemagglutinins in man. *J Immunol 100*, 1139–1140.
74. Cvjetanovic B, Uemura K (1965). The present status of field and laboratory studies of typhoid and paratyphoid vaccines with special reference to studies sponsored by world health organization. *Bull World Health Organ 32*, 29–36.
75. Vernacchio L, Bernstein H, Pelton S, Allen C, MacDonald K, Dunn J, Duncan DD, Tsao G, LaPosta V, Eldridge J, Laussucq S, Ambrosino DM, Molrine DC (2002). Effect of monophosphoryl lipid A (MPL) on T-helper cells when administered as an adjuvant with pneumococcal-CRM197 conjugate vaccine in healthy toddlers. *Vaccine 20*, 3658–3667.
76. Pollock KG, Conacher M, Wei XQ, Alexander J, Brewer JM (2003). Interleukin-18 plays a role in both the alum-induced T helper 2 response and the T helper 1 response induced by alum-adsorbed interleukin-12. *Immunology 108*, 137–143.
77. Le Moignic P (1916). Les vaccines en emulsion dans les corps gras ou "lipovaccins". *Comptes Rendus de la Societe de Biologie 79*, 201–203.
78. Jansen T, Hofmans MP, Theelen MJ, Manders F, Schijns VE (2006). Structure- and oil type-based efficacy of emulsion adjuvants. *Vaccine 24*, 5400–5405.
79. Herbert W (1968). The mode of action of mineral-oil emulsion adjuvants on antibody production in mice. *Immunology 14*, 301–318.
80. Allison A, Byars N (1986). An adjuvant formulation that selectively elicits the formation of antibodies of protective isotypes and of cell-mediated immunity. *J Immunol Methods 95*, 157–168.
81. Ott G, Barchfeld GL, Van Nest G (1995). Enhancement of humoral response against human influenza vaccine with the simple submicron oil/water emulsion adjuvant MF59. *Vaccine 13*, 1557–1562.
82. O'Hagan DT (2001). Recent developments in vaccine delivery systems. *Curr Drug Targets Infect Disord 1*, 273–286.
83. O'Hagan DT (2007). MF59 is a safe and potent vaccine adjuvant that enhances protection against influenza virus infection. *Expert Rev Vaccines 6*, 699–710.
84. Stephenson I, Bugarini R, Nicholson KG, Podda A, Wood JM, Zambon MC, Katz JM (2005). Cross-reactivity to

highly pathogenic avian influenza H5N1 viruses after vaccination with nonadjuvanted and MF59-adjuvanted influenza A/Duck/Singapore/97 (H5N3) vaccine: A potential priming strategy. *J Infect Dis 191*, 1210–1215.

85. Baldwin S, Fox CB, Pallansch MA, Coler RN, Reed SG, Friede M (2011). Increased potency of an inactivated trivalent polio vaccine with oil-in-water emulsions. *Vaccine 29*(4), 644–649.
86. Vogel FR, Caillet C, Kusters IC, Haensler J (2009). Emulsion-based adjuvants for influenza vaccines. *Expert Rev Vaccines 8*, 483–492.
87. Horzinek M, Mussgay M (1971). Studies on the substructure of togaviruses. Effect of urea, deoxycholate and saponin on the Sindbus virion. *Arch Gesamte Virusforsch 33*, 296–305.
88. Morein B, Sundquist B, Hoglund S, Dalsgaard K, Osterhaus A (1984). Iscom, a novel structure for antigenic presentation of membrane proteins from enveloped viruses. *Nature 308*, 457–460.
89. Sanders MT, Brown LE, Deliyannis G, Pearse MJ (2005). ISCOM-based vaccines: The second decade. *Immunol Cell Biol 83*, 119–128.
90. Kersten GF, Teerlink T, Derks HJ, Verkleij AJ, van Wezel TL, Crommelin DJ, Beuvery EC (1988). Incorporation of the major outer membrane protein of *Neisseria gonorrhoeae* in saponin-lipid complexes (iscoms): Chemical analysis, some structural features, and comparison of their immunogenicity with three other antigen delivery systems. *Infect Immun 56*, 432–438.
91. Lovgren K, Morein B (1988). The requirement of lipids for the formation of immunostimulating complexes (iscoms). *Biotechnol Appl Biochem 10*, 161–172.
92. Cox JC, Sjolander A, Barr IG (1998). ISCOMs and other saponin based adjuvants. *Adv Drug Deliv Rev 32*, 247–271.
93. Watson DL, Watson NA, Fossum C, Lovgren K, Morein B (1992). Interactions between immune-stimulating complexes (ISCOMs) and peritoneal mononuclear leucocytes. *Microbiol Immunol 36*, 199–203.
94. Villacres-Eriksson M, Bergström-Mollaoglu M, Kåberg H, Lövgren K, Morein B (1993). The induction of cell-associated and secreted IL-1 by iscoms, matrix or micelles in murine splenic cells. *Clin Exp Immunol*, *93*(1), 120–125.
95. Emery DL, Rothel JS, Wood PR (1990). Influence of antigens and adjuvants on the production of gamma-interferon and antibody by ovine lymphocytes. *Immunol Cell Biol 68*(Pt 2), 127–136.
96. Frazer IH, Quinn M, Nicklin JL, Tan J, Perrin LC, Ng P, O'Connor VM, White O, Wendt N, Martin J, Crowley JM, Edwards SJ, McKenzie AW, Mitchell SV, Maher DW, Pearse MJ, Basser RL (2004). Phase 1 study of HPV16-specific immunotherapy with E6E7 fusion protein and ISCOMATRIX adjuvant in women with cervical intraepithelial neoplasia. *Vaccine 23*, 172–181.
97. Davis ID, Chen W, Jackson H, Parente P, Shackleton M, Hopkins W, Chen Q, Dimopoulos N, Luke T, Murphy R, Scott AM, Maraskovsky E, McArthur G, MacGregor D, Sturrock S, Tai TY, Green S, Cuthbertson A, Maher D, Miloradovic L, Mitchell SV, Ritter G, Jungbluth AA, Chen YT, Gnjatic S, Hoffman EW, Old LJ, Cebon JS (2004). Recombinant NY-ESO-1 protein with ISCOMATRIX adjuvant induces broad integrated antibody and CD4(+) and CD8(+) T cell responses in humans. *Proc Natl Acad Sci USA 101*, 10697–10702.
98. Marciani DJ, Reynolds RC, Pathak AK, Finley-Woodman K, May RD (2003). Fractionation, structural studies, and immunological characterization of the semi-synthetic Quillaja saponins derivative GPI-0100. *Vaccine 21*, 3961–3971.
99. O'Hagan DT, Singh M (2003). Microparticles as vaccine adjuvants and delivery systems. *Expert Rev Vaccines 2*, 269–283.
100. Challacombe SJ, Rahman D, Jeffery H, Davis SS, O'Hagan DT (1992). Enhanced secretory IgA and systemic IgG antibody responses after oral immunization with biodegradable microparticles containing antigen. *Immunology 76*, 164–168.
101. Eldridge JH, Staas JK, Meulbroek JA, Tice TR, Gilley RM (1991). Biodegradable and biocompatible poly(DL-lactide-*co*-glycolide) microspheres as an adjuvant for staphylococcal enterotoxin B toxoid which enhances the level of toxin-neutralizing antibodies. *Infect Immun 59*, 2978–2986.
102. O'Hagan DT (1998). Microparticles and polymers for the mucosal delivery of vaccines. *Adv Drug Deliv Rev 34*, 305–320.
103. Shenderova A, Burke TG, Schwendeman SP (1999). The acidic microclimate in poly(lactide-*co*-glycolide) microspheres stabilizes camptothecins. *Pharm Res 16*, 241–248.
104. Singh M, Ugozzoli M, Kazzaz J, Chesko J, Soenawan E, Mannucci D, Titta F, Contorni M, Volpini G, Del Guidice G, O'Hagan DT (2006). A preliminary evaluation of alternative adjuvants to alum using a range of established and new generation vaccine antigens. *Vaccine 24*, 1680–1686.
105. Lavelle EC, Yeh MK, Coombes AG, Davis SS (1999). The stability and immunogenicity of a protein antigen encapsulated in biodegradable microparticles based on blends of lactide polymers and polyethylene glycol. *Vaccine 17*, 512–529.
106. Takahata H, Lavelle EC, Coombes AG, Davis SS (1998). The distribution of protein associated with poly(DL-lactide *co*-glycolide) microparticles and its degradation in simulated body fluids. *J Control Release 50*, 237–246.
107. Kazzaz J, Neidleman J, Singh M, Ott G, O'Hagan DT (2000). Novel anionic microparticles are a potent adjuvant for the induction of cytotoxic T lymphocytes against recombinant p55 gag from HIV-1. *J Control Release 67*, 347–356.

108. Singh M, Chesko J, Kazzaz J, Ugozzoli M, Kan E, Srivastava I, O'Hagan DT (2004). Adsorption of a novel recombinant glycoprotein from HIV (Env gp120dV2 SF162) to anionic PLG microparticles retains the structural integrity of the protein, whereas encapsulation in PLG microparticles does not. *Pharm Res 21*, 2148–2152.
109. Singh M, Briones M, Ott G, O'Hagan D (2000). Cationic microparticles: A potent delivery system for DNA vaccines. *Proc Natl Acad Sci USA 97*, 811–816.
110. O'Hagan DT, Jeffery H, Davis SS (1993). Long-term antibody responses in mice following subcutaneous immunization with ovalbumin entrapped in biodegradable microparticles. *Vaccine 11*, 965–969.
111. Randolph GJ, Inaba K, Robbiani DF, Steinman RM, Muller WA (1999). Differentiation of phagocytic monocytes into lymph node dendritic cells in vivo. *Immunity 11*, 753–761.
112. Bachmann MF, Zinkernagel RM (1997). Neutralizing antiviral B cell responses. *Annu Rev Immunol 15*, 235–270.
113. Maloy KJ, Donachie AM, O'Hagan DT, Mowat AM (1994). Induction of mucosal and systemic immune responses by immunization with ovalbumin entrapped in poly(lactide-*co*-glycolide) microparticles. *Immunology 81*, 661–667.
114. Nixon DF, Hioe C, Chen PD, Bian Z, Kuebler P, Li ML, Qiu H, Li XM, Singh M, Richardson J, McGee P, Zamb T, Koff W, Wang CY, O'Hagan D (1996). Synthetic peptides entrapped in microparticles can elicit cytotoxic T cell activity. *Vaccine 14*, 1523–1530.
115. O'Hagan DT, Singh M, Dong C, Ugozzoli M, Berger K, Glazer E, Selby M, Wininger M, Ng P, Crawford K, Paliard X, Coates S, Houghton M (2004). Cationic microparticles are a potent delivery system for a HCV DNA vaccine. *Vaccine 23*, 672–680.
116. Safety of and Immune Response to a Combination HIV Vaccine Regimen in HIV Uninfected Adults Trial dates—2003–2010. http://www.clinicaltrials.gov/show/NCT00073216.
117. Bangham A, Standish M, Miller N (1965). Cation permeability of phospholipid model membranes: Effect of narcotics. *Nature 208*, 1295–1297.
118. Allison AG, Gregoriadis G (1974). Liposomes as immunological adjuvants. *Nature 252*, 252.
119. Gabizon A, Papahadjopoulos D (1992). The role of surface charge and hydrophilic groups on liposome clearance in vivo. *Biochim Biophys Acta 1103*, 94–100.
120. Lee KD, Hong K, Papahadjopoulos D (1992). Recognition of liposomes by cells: In vitro binding and endocytosis mediated by specific lipid headgroups and surface charge density. *Biochim Biophys Acta 1103*, 185–197.
121. Maruyama K, Kennel SJ, Huang L (1990). Lipid composition is important for highly efficient target binding and retention of immunoliposomes. *Proc Natl Acad Sci USA 87*, 5744–5748.
122. Foged C, Arigita C, Sundblad A, Jiskoot W, Storm G, Frokjaer S (2004). Interaction of dendritic cells with antigen-containing liposomes: Effect of bilayer composition. *Vaccine 22*, 1903–1913.
123. Sharma SK, Dube A, Nadeem A, Khan S, Saleem I, Garg R, Mohammad O (2006). Non PC liposome entrapped promastigote antigens elicit parasite specific CD8+ and CD4+ T-cell immune response and protect hamsters against visceral leishmaniasis. *Vaccine 24*, 1800–1810.
124. Fries LF, Gordon DM, Richards RL, Egan JE, Hollingdale MR, Gross M, Silverman C, Alving CR (1992). Liposomal malaria vaccine in humans: A safe and potent adjuvant strategy. *Proc Natl Acad Sci U S A 89*, 358–362.
125. Felnerova D, Viret JF, Gluck R, Moser C (2004). Liposomes and virosomes as delivery systems for antigens, nucleic acids and drugs. *Curr Opin Biotechnol 15*, 518–529.
126. Almeida JD, Edwards DC, Brand CM, Heath TD (1975). Formation of virosomes from influenza subunits and liposomes. *Lancet 2*, 899–901.
127. Bungener L, Huckriede A, Wilschut J, Daemen T (2002). Delivery of protein antigens to the immune system by fusion-active virosomes: A comparison with liposomes and ISCOMs. *Biosci Rep 22*, 323–338.
128. Mischler R, Metcalfe IC (2002). Inflexal V a trivalent virosome subunit influenza vaccine: Production. *Vaccine 20*(Suppl 5), B17–23.
129. Zurbriggen R, Novak-Hofer I, Seelig A, Gluck R (2000). IRIV-adjuvanted hepatitis A vaccine: In vivo absorption and biophysical characterization. *Prog Lipid Res 39*, 3–18.
130. Perez O, Lastre M, Cabrera O, del Campo J, Bracho G, Cuello M, Balboa J, Acevedo R, Zayas C, Gil D, Mora N, Gonzalez D, Perez R, Gonzalez E, Barbera R, Fajardo EM, Sierra G, Solis RL, Campa C (2007). New vaccines require potent adjuvants like AFPL1 and AFCo1. *Scand J Immunol 66*, 271–277.
131. Neelapu SS, Gause BL, Harvey L, Lee ST, Frye AR, Horton J, Robb RJ, Popescu MC, Kwak LW (2007). A novel proteoliposomal vaccine induces antitumor immunity against follicular lymphoma. *Blood 109*, 5160–5163.
132. Zhou XY, Shibusawa N, Naik K, Porras D, Temple K, Ou H, Kaihara K, Roe MW, Brady MJ, Wondisford FE (2004). Insulin regulation of hepatic gluconeogenesis through phosphorylation of CREB-binding protein. *Nat Med 10*, 633–637. Epub 2004 May 2016.
133. Paavonen J, Jenkins D, Bosch FX, Naud P, Salmeron J, Wheeler CM, Chow SN, Apter DL, Kitchener HC, Castellsague X, de Carvalho NS, Skinner SR, Harper DM, Hedrick JA, Jaisamrarn U, Limson GA, Dionne M, Quint W, Spiessens B, Peeters P, Struyf F, Wieting SL, Lehtinen MO, Dubin G (2007). Efficacy of a prophylactic adjuvanted bivalent L1 virus-like-particle vaccine against infection with human papillomavirus types 16 and 18 in young women: An interim analysis of a phase III double-blind, randomised controlled trial. *Lancet 369*, 2161–2170.

134. McKenzie R, Bourgeois AL, Frech SA, Flyer DC, Bloom A, Kazempour K, Glenn GM (2007). Transcutaneous immunization with the heat-labile toxin (LT) of enterotoxigenic *Escherichia coli* (ETEC): Protective efficacy in a double-blind, placebo-controlled challenge study. *Vaccine 25*, 3684–3691.

135. Glenn GM, Villar CP, Flyer DC, Bourgeois AL, McKenzie R, Lavker RM, Frech SA (2007). Safety and immunogenicity of an enterotoxigenic *Escherichia coli* vaccine patch containing heat-labile toxin: Use of skin pretreatment to disrupt the stratum corneum. *Infect Immun 75*, 2163–2170.

136. van Ginkel FW, Jackson RJ, Yoshino N, Hagiwara Y, Metzger DJ, Connell TD, Vu HL, Martin M, Fujihashi K, McGhee JR (2005). Enterotoxin-based mucosal adjuvants alter antigen trafficking and induce inflammatory responses in the nasal tract. *Infect Immun 73*, 6892–6902.

# ORAL DELIVERY OF MACROMOLECULES FOR THE DEVELOPMENT OF THERAPEUTIC AGENTS

Hye J. Lee and Chanmi Choi
*Hopkins Bio Research Center, Co., Seoul, South Korea*

## 1 INTRODUCTION

Oral delivery of macromolecules as therapeutic agents is still an unthinkable project for pharmaceutical scientists. However, considering the fact that developing peptides and proteins (major macromolecules) as medicines was also a mission impossible not too long ago, delivering macromolecules orally could not be just dreaming. Currently, over 160 protein-based therapeutics are available on the U.S. market and about 400 compounds are being developed [1]. Expenditures in the protein therapeutics market are predicted to continue to expand by 15–20% annually [2]. This is astonishing progress and could be attributed to the tremendous advances in recombinant DNA (deoxyribonucleic acid) techniques and biotechnologies. The dosage forms for peptide and protein drugs rely on their parenteral injection, which requires frequent administration, and are urging the development of more convenient dosage forms for non-invasive administration. An oral dosage form is the most preferable route because of the accompanying advantages of patient comfort, ease of administration, decreased medical costs, and increased patient compliance. Significant efforts in academic and commercial laboratories have been expended, but major breakthroughs in oral peptide and protein formulation achieving predictable and reproducible drug profiles in therapeutic doses without wasting up to 99% of the drug have not been achieved yet. The major barriers to developing oral formulations for peptides and proteins include poor intrinsic permeability, luminal and cellular enzymatic degradation, rapid clearance, and chemical and conformational stability [3]. Conventional approaches to address these barriers have not readily translated into effective peptide and protein formulations [4].

This review will provide an overview of the existing approaches, the recent advances that have been made in oral delivery as related to proteins, and some future directions to consider followed by reviewing factors affecting oral protein absorption. It is likely that effective oral formulations for peptides and proteins will remain highly compound specific, but the principal concept into preserve protein integrity before its reaches systemic circulation.

## 2 FACTORS AFFECTING ORAL PROTEIN ABSORPTION

Mechanisms of the absorption of peptides and proteins by epithelial cells are well described by Goldberg and Gomez-Orellana [5]. The process of receptor-mediated or non-receptor-mediated endocytosis (also called pinocytosis, when the material taken up is entirely soluble) is the most likely route of uptake and possible translocation of an orally administered therapeutic protein. Recent findings on many transporters add a transporter-mediated uptake as a possible absorption mechanism for proteins in addition to the paracellular absorption

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

mechanism [6, 7]. Oral delivery of particulates in which the protein drugs were embedded to be protected from the gut enzymes may require more understanding of uptake and/or absorption. Many groups believe that uptake of these particulates is a function carried out by membraneous/microfold (M) cells in Peyer's patches of the mammalian gut. However, other groups claim that uptake across the villous epithelium occurs via damaged regions or by a paracellular route [8]. Thomas et al. [9] found that the absorption of microparticles across the Peyer's patches of rats, mice, and rabbits can occur rapidly as microparticles can be detected in the lymph soon after intestinal delivery.

Numerous research studies during the last decade describe factors affecting low oral bioavailability of protein drugs [10–14]. The most advanced findings are summarized below. Effective and successful oral protein delivery requires further development in understanding these complex mechanisms in order to maximize the therapeutic potential of this class of compounds.

### 2.1 Absorption Site

In addition to metabolic differences, the anatomical, physiological, and biochemical differences between the various part of the gastrointestinal (GI) tract can cause significant variation in protein drug oral absorption [15, 16]. Among the physiological factors, regional differences in composition and thickness of mucus, pH, surface area, enzyme activity, bile, pancreatic juice, mucus, and fluid volume and content can modify dissolution rates, solubility, transit times, metabolism, and membrane transport of the drug molecules. The lipid/protein composition of the enterocyte membrane along the GI tract can alter binding and transport of drugs. The location and number of Peyer's patches, which are most abundant in the ileum, can also be important in the absorption of large molecules and particulate matters. Consequently, the existence of a (better) absorption window could not be a new discovery for peptide and protein drugs administered orally. For peptides, cyclosporine [17] and desmopressin [18] exhibited maximum absorption in the duodenum in humans and rabbits, respectively, while octreotide, a somatostatin analog, in the lower GI tract in humans [19]. Although little information exists on site-to-site variation in the protein absorption, the absorptive cells of neonatal rats have been shown to capture nerve growth factor [20], epidermal growth factor [21], milk proteins [22], ferritin and its analogs [23], and Follicle Stimulating Hormone-Fragment crystallizable region (FSH-Fc) fusion proteins by receptor-mediated endocytosis [24]. Some of the polymeric materials induce the opening of a tight junction between cells, which enables paracellular absorption to take place [5, 25–27]. Other enzyme systems, cytochrome P450 3A4 (CYP3A4) and catalytic activity decrease longitudinally along the small intestine [28], while P-glycoprotein (P-gp) messenger ribonucleic acid (mRNA) levels increase longitudinally along the intestine, with lowest levels at the stomach and highest levels in the colon [29].

### 2.2 Metabolism

A comprehensive review on the intestinal enzyme systems has been written by Woodley [14]. The greatest threat to therapeutic proteins lies in the lumen of the small intestine, which contains gram quantities of proteases secreted from the pancreas, as well as cellular proteases from the mucosal cells, which are constantly sloughed off from the villie. The second major enzymatic barrier is the brush-border membrane of the epithelial cells, which contains at least 15 peptidases that together have a broad specificity and can degrade both proteins and peptides. Woodley [30] suggested that an orally administered protein might encounter more than 40 different enzymes during passage through the small intestine. This is a qualitative assessment difficult to translate into quantitative information. Furthermore, lysosomal peptidases, estimated at over 60 in number, will also present a barrier to any peptides or proteins endocytosed by the epithelial cells. Considering the 30-amino-acid B chain of insulin and just 2 of the lysosomal peptidases, capthepsins D and B, exposure of the peptide to just these enzymes might lead to it being split at 20 sites out of a total of 29 [31]. Site-variable enzymatic degradation of luteinizing hormone-releasing hormone (LHRH) [32] and insulin-like growth factor (IGF-I) [33] were profiled.

Interestingly, Garrido et al. [15] have found that the hydrolytic activity for all enzymes (including the subcellular fractions of the aminopeptidase activities) and absorption rates from both the free amino acid and peptide solutions were reduced in bypassed jejunal segments. When large peptides and protein drugs were administered to the ileum, there was significant increase in absorption as compared to those drugs administered orally by gavage, supporting the possibility for heterogeneous distribution of the corresponding enzymes in the intestine [34, 35]. The colon has received some attention as a possible site for protein delivery; however, evidence shows that the lumen of the colon contains substantial amounts of peptidase activity, largely because of enzyme production by microorganisms. Therefore, intestinal metabolism acts as a concerted barrier to drug absorption. From knowledge of the enzymatic barrier, the strategies for oral protein delivery of enzyme inhibition and the synthesis of enzyme-

resistant protein analogs are logical developments. The latter approach seems to be a more realistic approach for protein drug oral delivery.

In addition, CYP3A4, the major phase I drug metabolizing enzyme in humans, and the *MFR-1* gene product P-gp are to be considered in designing the delivery system, which are present at high concentration in villus tip enterocytes of the small intestine and share a significant overlap in substrate specificity [36–38]. Cyclosporin, saquinavir, and a new cystein protease inhibitor of K02 (Morpholine-Urea-Phe-Hphe-Vinyl Sulfone; Axys Pharmaceutical) were metabolized by the CYP3A4 and P-gp system [36]. Fluvoxamine has prominent affinity for the CYP12 isozyme, lesser affinity for the CYP3A4 and CYP2C isozymes, and minimal affinity for CYP2D6 [39]. This profile suggests the need for careful dosage adjustment when used together with some drugs that have a narrow therapeutic range in order to minimize inhibiting their metabolism. Once they survive the intestinal enzymatic barriers, protein drugs have to face another presystemic metabolism, blood and hepatic extraction, although the transit time is shorter and the enzyme capacity could be smaller than those of the gut [40].

### 2.3 Permeability

The generally poor permeability of peptide and protein drugs across biological membranes can be attributed to their hydrophilic structure and large molecular size [3, 11, 41]. The mucosal cell permeability determines the rate of entry into the mucosa across the brush-border membrane. The permeation process can be carrier mediated and a paracellular or transcellular route. Although no quantitative cross correlations are established between human and animal permeability, tissue culture (e.g., the Caco-2 cell), permeabilities, and physicochemical properties such as partition coefficient, advancement in research in these areas has been exploitied by many investigators, indicating the availability of an informative database in the near future. Poor membrane permeability of hydrophilic proteins might be overcome by structurally modifying the compounds, thus increasing their membrane partition characteristics and/or their affinity to carrier proteins [42]. A variety of permeability enhancers for peptides [3], including salicylates, mixed bile salt–fatty acid micelles [43], chelators [44], fatty acids, bile salts [44], acylcamitines, medium-chain glycerides, surfactants [45, 46], and melittin [47] have been shown in cellular and animal models to increase the absorption of the peptides.

Cylcodextrins (CD) are cyclic oligomers of glucose that form inclusion complexes with drugs and cause alveolar epithelial membrane to increase the permeability based on their capacity to alter fluidity [44]. The surface charge of a polymer loaded on active compounds should be appropriately selected for the ability of permeability [48]. These enhancers are not warranted to work for protein drugs, however, it can be helpful to find a starting point for protein drug delivery.

### 2.4 Transporters

The distribution and characteristics of the intestinal transporters exhibit regional differences along the intestine, implying diverse physiological functions and in some cases pathological responses. There have been several studies on the mechanism and substrate structure–affinity relationship for this transport system [36, 49–52]. The cloning, expression, and characterization of individual peptide transporters will give additional insight into the rational design of drug molecules as well as the delivery system-especially for prodrug approaches [37, 53]. Functional expression of the multidrug resistance gene product, P-gp, as a primary active transporter in the intestinal brush-border membrane leads to net secretion of some drugs such as anticancer agents in the blood-to-luminal direction, serving as a secretory detoxification and as a part of the absorption barrier in the intestine [54, 55].

### 2.5 Low Solubility

For certain proteins with low solubility such as human growth hormone (18 mg/ml), poor absorption was believed to be a major cause for the low solubility. In an attempt to improve the solubility-limited bioavailability associated with these types of compounds, formulators have turned to the use of lipid excipients in which the compounds can be solubilized prior to oral administration. For selection of lipid-based delivery, the Biopharmaceutical Classification System (BCS) can serve as a useful preliminary guide [56]. Further elucidation of the physiological aspects of lipid excipients, for instance, the inhibition of gut wall efflux mediated by multidrug-resistant (MDR) P-gp and presystemic metabolism by CYP3A4 bound to the gut wall, can be expected to define further candidate selection based on the BCS [36, 37, 56]. Nanocapsules and nanoparticulates could help protein molecules to bypass the limited solubility, thus penetration and proteolytic barriers, by taking up the epithelial mucosa via the paracellular pathway [13].

### 2.6 Molecular Size

In particular peptides and proteins displaying a molecular size greater than 30 kDa, the intestinal membrane becomes an important rate-limiting factor for drug

absorption [57, 58]. Size limitation for protein drug oral absorption could be modulated by a formulation type and an absorption site. The particle sizes employed for current protein oral delivery can be classified into two: 10–1000 nm of nanoparticle and 1–1000 μm of microparticulates [10]. The uptake of particles within the intestine increases with decreasing particle size and increasing hydrophobicity [8]. Furthermore, this size dependency in absorption could be site dependent. One study, which employed a series of polyethylene glycols (PEG), suggests that a compound with a molecular weight up to 4000 Da might be significantly absorbed [42], assuming appropriate partition behavior and stability. However, several reports on insulin [human growth hormone (hGH); erythropoietin (EPO), and interferon (IFN)] and many other proteins with a molecular weight ranging from 6 to 30 kDa showed absorption when administered orally, indicating site-dependent absorption [59]. These findings may need further clarification.

### 2.7 Chemical Stability

Physical properties of proteins often present significant formulation problems not encountered with conventional organic drug molecules [3], for example, self-aggregation. The tendency of insulin to form hexamers is well known, and the absorption of hexamers is different than that of the monomer. Human calcitonin is also known for self-organization into fibrillar structures with reduced biological activity [60].

### 2.8 Drug Interaction

Drug interactions can occur at any step from absorption to elimination of a drug and can induce adverse as well as beneficial effects [61]. The main potential for drug interaction is with drugs or compounds that are metabolized by or affect CYP3A4, including imidazoles, grapefruit juice, erythromycin, mibefradil, and others [62]. Glutathione may play an important role in the inactivation of orally given peptides and proteins by thiol-disulfide exchange reactions [63].

### 2.9 Gender Difference

Gender-specific differences can be explained by differences in body composition between men and women and/or physiological changes during the menstrual cycle as well as differences in plasma protein binding secondary to hormonal characteristics [64]. In addition, gender differences in drug metabolism may be involved in the higher incidence of adverse reactions in women compared with men [65]. However, it has not been fully elucidated in the literature whether there is any quantitative and or qualitative difference in proteolytic enzymes and physiology of the absorptive cells in the intestine, which may produce a significant impact on protein drug oral absorption. Gender–related differences could be shown in phase I (cytochrome P450) as well as phase II (especially glucuronidation) reactions [66]. Frequent and sometimes clinically relevant gender differences could be identified for drug elimination processes and were predominantly linked to the gender–specific expression of metabolic enzyme systems, for example, CYP3A4 and CYPlA2. With regard to pharmacodynamics (PD), gender differences have been observed in baseline characteristics as well as in drug response, which might both, at least in part, be the consequence of modulation by sex hormones. Since most protein drugs are from endogenous sources, it would be important to understand baseline pharmacodynamics in oral drug delivery. Some of the most striking examples identified were in pain therapy and perception, glucose management, and arrhythmia susceptibility. CYP3A4, responsible for the metabolism of over 50% of therapeutic drugs, exhibits higher activity in women than in men [67], whereas the activity of many other systems involved in drug metabolism may be higher in men than in women [65]. Women and men also show different PD responses to a variety of drugs. While the clinical significance of these gender differences remains to be determined, it is anticipated to be most important in the administration of drugs that have a narrow therapeutic range.

### 2.10 Others

Particle uptake seems to be greater in older animals [10]. The exact mechanism of this was not fully explained. The presence of food seems to be another enhancing factor for particle uptake possible by increasing intestinal transit time. Hepatic insufficiency reduce the plasma clearance of a number of drugs eliminated by biotransformation and/or biliary excretion, but it can also affect plasma protein binding, which in turn could influence the processes of distribution and elimination [68]. In addition, reduced liver blood flow in patients with chronic disease will decrease the systemic clearance of flow-limited (high extraction) drugs, and portal-systemic shunting may substantially reduce their presystemic elimination (first-pass effect) following oral administration. When selecting a drug and its dosage regimen for a patient with liver disease, additional considerations such as altered pharmacodynamics and impaired renal excretion (hepatorenal syndrome) of drugs and metabolites should also be taken into account.

Consequently, dosage reduction is necessary for many drugs administered to patients with chronic

liver disease such as liver cirrhosis. Protein drug pharmacokinetics (PK) are not appreciably affected by age, gender, or race when differences in renal function and body mass and composition are taken into account [10, 12].

## 3 APPROACHES FOR ORAL DELIVERY

The strategy employing enzyme inhibitor(s) and absorption enhancers to protect from the various enzymes and to help to overcome its own hurdles, that is, hydrophilicity, molecular size, charge, and the like, seems to be the most popular approach for protein oral delivery [40]. Second, realistic approach could be modifying protein structures to more stable analogs by chemical synthesis. Targeting the specific absorption site and dosage form modification, that is, lipid vesicles, colloidal carrier systems with and without mucoadhesive polymers could be another approachs to improve protein absorption. A novel and perhaps less complex oral protein drug technology could be based on a combination of several approaches that are described below.

### 3.1 Enzyme Inhibiton

Due to the nature of enzyme distribution and quantities, the use of only enzyme inhibitor(s) for protein oral delivery will be extremely inefficient. However, some successful results have been reported for insulin administered with sodium glycocholate, camostat mesilate, and bacitracin to rats by Yamamato et al. [69] and with FK-448 and other protease inhibitors by other investigators [57, 70]. The CYP3A4 inhibitor showed marked increase in the oral bioavailability of cyclosporin [71]. To investigate the possible use of casein as an adjuvant for oral delivery of peptide drugs, the inhibitory activity of casein on proteases in rat small intestine was examined [72]. Utilization of both inhibitor cocktails and a specific absorption site with less enzyme distribution showed improvement in the absorption of large peptides, cholecystokinin and enkephalin analogs [34, 73], and protein drugs [35].

A new and interesting possibility for protein drug oral delivery makes use of a polymer and enzyme inhibitor conjugate system [74, 75] in which the embedded therapeutic agent is protected from enzymatic degradation [76]. The covalent attachment of aprotinin to chitosan exhibit mucoadhesive properties, which allows a drug release at the same intestinal segment as the protease inhibition takes place. Contrary, coadministered unbound inhibitors are highly diluted during GI passage, and, therefore, high amounts of the inhibitor are necessary in order to achieve an improved drug uptake [77]. Chitosan and its derivatives showed multiple effects of enhancer, enzyme inhibitor, and mucoadhesion to improve the absorption of insulin, calcitonin, and buserelin following oral administration [58]. Several ovomucoid species effectively stabilize salmon calcitonin (sCT) against serine proteases [78].

In similar work, the enzyme inhibitor, pepstatin analogs [79], were covalently joined to mucoadhesive polymers so that the inhibitor molecules were freely accessible to the enzyme as a substrate. This system provided advantages toward increasing contact time with the GI mucosa and maintained a controlled and sustained drug release from the adhered polymer and reducing unwanted toxic effects of the inhibitor because of its attachment to the nonabsorbable polymer backbone. However, from a practical point of view, the utility of this approach may be limited by high manufacturing cost [59].

### 3.2 Carrier Holecules or Enhancers

Avoiding temporary to long-lasting membrane damage, carrier molecules or enhancers improving the epithelial cell permeability of protein drugs were synthesized by Leone-Bay et al. [59, 80], which were low-molecular-weight, peptide-like compounds that interact covalently with the protein drug. This approach was successful for hGH oral absorption with a mean peak concentration of 55 ng/mL in primates [80]. Another group's research employing the delivery agent resulted in the bioavailability increase of 50–104% for hGH, interferon-α-2b, and insulin [81]. Permeability increases in the presence of carrier molecules for hGH, IFN, and insulin were 0, 350, and 800%, respectively, indicating a proportional increase of bioavailability to permeability [81]. The ability to increase oral absorption of granulocyte colony–stimulating factor (G-CSF) and EPO by covalent coupling to vitamin $B_{12}$ has been investigated [82, 83]. This absorption occurs via receptor-mediated endocytosis, and an essential requirement is the presence of the intrinsic factor of a mucoprotein.

The mucoprotein complex reaches the ileum, where resorption is mediated by specific receptors. Possible disadvantages of the vitamin-$B_{12}$-mediated delivery are the limited capacity of this active transport mechanism and the interference with vitamin $B_{12}$ absorption in chronic therapy [84].

The conjugated carrier of oligomeric transferrin (Tf) to insulin causes a 6- to 10-fold increase of the stability of insulin against chymotrypsin digestion [85]. Oral route of mini-proinsulin-transgenic *Ganoderma lucidum* can be effective in reducing blood glucose level less than or equal to 64% in 80% of diabetic rats [86].

### 3.3 Prodrugs and Analogs

Altering physicochemical properties of protein drugs seems to be the easiest approach [14, 87], warranting unaltered or intensified efficacy, although it requires the synthesis of a new chemical entity [59]. Changes can be made in lipophilicity, charge, molecular size, solubility, configuration, isoelectric point, chemical stability, enzyme liability, and affinity to carriers to enhance the absorption and systemic circulation. Two specific approaches made for peptide oral delivery by Samanen et al. [87]—the chemical modification of peptide amide bond to enhance intestinal permeability and the design of compounds bearing nonpeptide templates—were more amenable to the discovery of compounds with oral activity, from peptide pharmacophore models, for example, SB 208651 designed from the peptides SK&F 106760 and SK&F 107260. Computational modification of hGH attempts to improve its thermostability and oral absorption [88]. The permeability of 10 amino-acid ester prodrugs of levovirin (LVV) is increased [6.9-fold increase of the area under the curve (AUC) in Caco-2 cells [89].

### 3.4 Optimal Absorption Site

Many investigators [13, 90] suggested that identification of the optimal absorption site for a given peptide or protein is the first step toward the design of a delivery system that maximizes absorption. Regional variations in the penetration barriers to peptides may result in regional differences in their absorption. For instance, M cells located on the dome epitheliam of gut-associated lymphoid tissue are known to be capable of sampling macromolecular antigens from the ileum through an endocytotic pathway [91]. Controlling the absorption site to deliver proteins and large peptide drugs has been attempted with some success in improving oral bioavailability [34, 35, 73]. Researchers searched for absorption sites with less enzyme distribution and avoiding the hepatic first-pass effect and found, thus far, the oral cavity (buccal) and colon as alternative sites for protein oral delivery.

### 3.5 Buccal Delivery

The oral mucosa has greater permeability and perfusion than the skin while the oral cavity provides an environment almost free from the acidity and protease activity encountered elsewhere in the GI tract [92]. In addition, blood vessels of the oral mucosa drain directly into the jugular vein, avoiding first-pass extraction by the liver. Buccal delivery permits easy removal in the event of adverse reactions. Recent study revealed the presence of aminopetidase activity on the buccal mucosa and its inhibition by enzyme inhibitors [93]. Complete review on the regional variation in oral mucosal drug permeability written by Kurosaki and Kimura [94] is available. In this report, the structure and composition of the mucosa at different sites in the GI tract, factors affecting mucosal permeability, and formulation factors including penetration enhancement relevant to the design of systems targeting oral mucosal delivery were reviewed. Many oral mucosal delivery systems for numerous peptide/protein drugs have been described since the first attempt at a bioadhesive system by Nagai in the early 1980s [95]. Interactions of randomly methylated β–cyclodextrin (RAMEB) with buccal mucosa could be diffused through cell layers by paracellular pathway. Two and 5% RAMEB do not induce tissue damages even after 5 days [96].

Pituitary adenylate cyclase-activating polypeptide (PACAP) with the thiolated polymer chitosan combined with GSH and the nonionic polyoxyethylene surfactant Brij 35 led to an absolute bioavailability of 1% by the buccal route of porcine [97] α-interferon. With 5% lysalbinic acid, a product of the alkaline hydrolysis of egg albumin and a mild detergent, resulted in a five-fold increase in permeability through the mucous in 10 min, whereas, lysalbinic acid at a concentration of 0.3% increased buccal permeability of fluorescein isothiocyanate-labeled insulin (FITC-insulin) for above five times just in 10 min [98]. A relative bioavailability of $43.8 \pm 10.9\%$ was achieved after buccal application of sCT-loaded thin-film composites (TFCs) in rabbits [99].

Transforming growth factor β3 (TGF-53) has shown that it can temporarily arrest oral mucosal basal cell proliferation [100]. Delivery of calcitonin, LHRH, and glucagon-like peptide-I as an adhesive tablet showed the bioavailability of 37% [101], 100% [102], and 41% [103], respectively. Exciting challenges remain to influence the bioavailability of drugs across the oral mucosa. Nonetheless, there are still many issues to resolve before this effective and convenient route of drug delivery can be thoroughly and safely utilized [92].

### 3.6 Colon Delivery

A common belief is that colonic delivery for orally administered protein drugs is possible because of the postulated low proteolysis activity in the large intestine, an assumption that requires further verification [104]. Yet, other opportunities for colonic delivery of drugs also exist. Some recent examples include bypassing small intestine metabolism, achieving constant absorption rates for some molecules, and delivering cationized antioxidant enzymes to the colonic epithelium. Colonic delivery of insulin after coating with azo-cross-linked copolymer of styrene and hydroxyethyl methacrylate(HEMA), aiming for activation by micro-

flora of large intestine and intestinal delivery of enterocoated cyclosporin, are examples of site-specific delivery cases [13, 105, 106]. Colon-specific drug delivery systems (CDDSs) have been investigated for model drugs, bovine serum albumin (BSA) with resistant starch-coated films [107, 108], and bee venom peptide with calcium-coated alginate-gel-beads-entrapped liposome [109].

Baluom et al. [110] also achieved about 15% of bioavailability after colonic administration of sCT using submicron emulsions containing Carbopol 940. Another problem with protein delivery to the colon is various transit times; ranging from a few hours to 10 h [111]. Thiamine-coated nanoparticles showed a strong capacity to be captured in Peyer's patch of rat colon by fluorescence microscopy images [112].

### 3.7 Targeting Transporter

Recent advances in gene cloning [113], molecular biology techniques, and computational methods [114] make it possible to study the characteristics and distribution of transporters [115] at the molecular level based on molecular characterizations of membrane transporters and accumulated biochemical data on their specificities and kinetics [113], structural modification [116], and targeting of a specific transporter is a promising strategy for the design of drugs that improve bioavailability and tissue distribution [37, 117].

The peptide transporter is limited to peptides of less than 4 amino acids [82]. Its physiological and pathophysiological conditions are reviewed by Nielsen and Brodin [118]. The bile acid transporter is limited to peptides of less than 400 Da moleadar weight (HW), while the vitamin $B_{12}$ transporter is limited to pharmaceuticals that require less than 40 mol per day.

### 3.8 Colloidal Carrier Systems

Various colloidal systems have been studied for absorption enhancement of proteins, such as submicron emulsions [108, 119], lipid suspensions, liposome, polymeric nano/microparticles, and freeze-dried suspension [120].

The bioavailability of a stable self-emulsifying formulation for insulin is up to 15.2% at an oral dose of 2.5 IU/kg in comparison with the hypoglycemic effect of native insulin (0.5 IU/kg) delivered by subcutaneous injection in diabetic beagle dog [121].

Controversy still exists on physicochemical factors governing gastrointestinal uptake, including size, size distribution, consistency, hydrophobicity, and surface properties [84]. Prolonged contact of nanoparticles (NP) with absorptive cells may be achieved using bioadhesive materials. Bioadhesion could be followed by NP uptake in a second step. Therefore, biomaterials with both adhesive and protective properties would be desirable for oral protein delivery in order to assure the drug stability and thus improved bioavailability [122].

### 3.9 Nanoparticles

Nanoparticles, defined as solid particles with a size in the range of 10–1000 nm, allow encapsulation of the protein drugs inside a matrix, protecting them against enzymatic and hydrolytic degradation. Various biomaterials of polymers, lipids, lectins, and the like could be employed to make NP using techniques of emulsion polymerization, interfacial polymerization, emulsification evaporation, solvent displacement [123, 124], salting out, emulsification diffusion, and desolvation [125, 126]. The solvent displacement and salting out have received increasing importance because they provide less stress to protein drugs. The physicochemical properties of NP, for example, portion of the mucoadhesive polymers [127] and their behavior on exposure to physiological media, for example, formulation pH [128], are greatly dominated by their chemical structures and surface characteristics [84, 129, 130]. The mucus layer and Peyer's patches [131] were considered to be uptake sites for NPs. Nauoparticles and cyclodextrins are biocompatibility as delivery systems [132]. The PEGylation of chitosan reduced the cytotoxicity of the nanocapsules [123]. Studies on the novel formulation of alginate-blended chitosan NPs have been made to improve the controlled release [133, 134] and try to optimize the alginate : chitosan mass ratio [135].

Surface compositions of NPs that greatly dominate the intensity and stability of the response are newly developed by using glucomannan-coated chitosan [136], chitosan-coated lipid [137, 138], NP film consisting of carboxylation chitosan-grafted nanoparticles (CCGNs), bilaminated films [139], and Polydextran aldehyde with particle core [140]. Some peptides such as helodermin confirm the high permeability of the NP system across in vitro follicle-associated epithelium model [141].

Erythropoietin loaded short fiber length carbon nanotubes deliver more EPO and absorption enhancer to the absorptive cells of the rat small intestine [142]. The structural integrity of insulin after being entrapped into chitosan/alginate nanoparticles produced by ionotropic polyelectrolyte pre-gelation was preserved as a result of assessment by Fourier transform infrared (FTIR) and circular dichroism (CD) [143].

### 3.10 Bio(muco)adhesion

Mucoadhesive oral drug delivery systems have received considerable attention over the last decade. Utilizing bioadhesive polymers to prolong and/or to intensify the

contact between controlled-release dosage forms and the stomach or gut mucosa has also shed new light on the potential of mucoadhesive polymers. In a series of cationic, anionic, and neutral polymers, anionic poly(acrylic acid) (PAA) was found by Park and Robinson [144]. Poly(methacrylic acid) grafted–poly(ethylene glycol) (MAA-*g*-EG) hydrogels [145] show the highest mucoadhesive characteristic. PAA could be identified as a potent inhibitor of proteolytic enzymes as well. There is also increasing evidence that the interaction between various types of bio(muco) adhesive polymers and epithelial cells has direct influence on the permeability of mucosal epithelia. By choice of polymer concentration and, more importantly, solvent type, organogels exhibited a wide range of viscoelastic and flow properties [108].

It was demonstrated that the adhesion or binding of proteins to the intestinal epithelial cell surface led to enhanced uptake by nonspecific receptor-mediated endocytosis [146]. One area to improve this technology is to increase the localization or retention in any desired regions [13]. The inability of mucoadhesive systems to remain for longer periods at the site of attachment is partly due to the high turnover and sloughing rates of the mucus to which the system attached (40–270 min) in rats [147] and partly due to the possible displacement of the delivery system from binding to the mucus layer by soluble mucus in the luminal contents [148]. Rather than being just adhesives, mucoadhesive polymers may, therefore, be considered as a novel class of multifunctional macromolecules with a number of desirable properties for their use as biologically active drug delivery adjuvant [40]. Novel bioadhesives allow for the capacity of the enhanced adhesion along with sitespecific ligands such as lectins [149–152], RGD peptides [153], and immunoregulatory properties with lectins [150, 154–156].

Mucoadhesive tablets are examined using Carbopol 974P, Labrasol, for EPO [157]. Microfabrication technology combined with appropriate surface chemistry may offer the highly localized or unidirectional release of drugs, permeation enhancers, and promoters [152, 158].

### 3.11 Liposomes

Liposomes have been abandoned due to insufficient stability under in vivo condition [8]. In an application to vaccines, homogeneous unilamellar liposomes of ~100 nm were reported to be taken up into M cells [159]. However, studies on the vaccine liposome led to the overall conclusion that liposomes are ineffective as adjuvants for the nonparticulate, naturally weak immunogens used in the investigation. A mucoadhesive liposomal system prepared by coating a negatively (phosphatidyl choline) or positively charged (salicylic acid) liposome suspension with mucoadhesive polymer solutions, such as chitosan and carbopol, showed some success in significant absorption of protein drugs, for example, insulin and calicitonin [160]. Chitosan-coated liposomes [161, 162], polysaccharide-coated liposomes [163], and PEG-coated liposomes [164] were characterized by the mucoadhesiveness of liposomes and resultant pharmacological effects of the liposomal protein and/or peptide drug.

Recently, the liposome technology has been revisited [165] and evaluated. Various cytokines in liposomes [166] and multivesicular liposomes characterized by their unique structure of multiple nonconcentric aqueous chambers surrounded by a network lipid [167] have been tested and obtained satisfied pharmacodynamic effects of model protein drugs.

Oral delivery of antigens in liposomes modulates oral tolerance to the antigens [168].

A combination assay of PAMPA and unidirectional (apical to basal) Caco-2 permeability model can synergistically provide invaluable permeability/absorption assessment of liposomal formulations of serratiopeptidase [169].

Because of a liposome's instability, alternative nonionic surfactants vesicles, "niosomes," have been investigated. Using this technique, it has been synthesized and examined for delivery of insulin [170] and adjuvant carrier system [171].

### 3.12 Microencapsulation

Poly(lactide1gycolic acid) (PLGA) polyester microspheres have been used for long-acting injectable sustained-release formulations of protein drugs [172] due to its biodegradable and biocompatible property. PLGA microparticles separated from the magnetic material are providing an opportunity for long-term efficacy and high bioavailability of orally delivered protein drugs [173, 174].

This PLGA microsphere has also been investigated as a potential antigen delivery system for oral vaccines of staphylococcal enterotoxin B (SEB) [175]. Viruses were also successfully encapsulated in PLGA microspheres for oral delivery [176]. Zein, one of the typical prolamines and a major storage protein of corn [177, 178], is insoluble in aqueous media yet is degraded over time by protease enzymes to constituent peptides and amino acids showing mucoadhesive properties. Alkermes, Inc. (Cambridge, MA) was used successfully to encapsulate calcitonin, EPO, sesmopressin, vasopressin, and insulin into the zein microsphere system for oral administration [159].

Because of different compositions [179, 180], additives of microparticles [181] influence the degree of releasing drugs. PEGylated insulin undergoes about 10% degradation with elastase, while native insulin undergoes about 70% degradation after 1 h incubation [182].

Calcium phosphate particles containing insulin keep from degradation while passing through the acidic environment of the GI track until being released in the less acidic environment of the intestines [183]. Serratiopeptidase (STP)-loaded nanocore was further encapsulated into alginate gel for enzyme protection [184].

### 3.13 Emulsions

Using this technology the method of transport in the gastrointestinal mucosa was shown to be transcellular by enhancing membrane permeability with oil in the formulation [185]. The ability of water-in-oil (W/O) emulsion containing ovalbumin (OVA), a model antigen, efficiently enhances the induction of Th2-dominated imbalance [186].

Calcitonin and hGH in microemulsion for an oral formulation had been developed and tested in humans for commercialization by Cortecs Int. Ltd. (Middlesex, UK) and Affinity Biotech, Inc. (Aston, PA) [159, 187].

### 3.14 Inclusion/Association Complexes

Dextran-derived biomaterials have been considered to be compatible matrices for protein and bioactive drugs because of their hydrophilic properties and ability to control drug dissolution and permeability. Cyclodextrins (CD) are potential candidates for use as delivery systems because of no effect of antigenicity, tolerogenicity, or immunogenicity [188].

Protection of the peptide by complex formation with α-cyclodextrin was also studied with respect to their ability to protect the peptide against proteolytic degradation by the exopeptidase, γ-glutamyl transpeptidase [189]. Cyclodextrin–insulin complex encapsulated PMAA–based NPs are examples of satisfied pharmacodynamic effects of model protein drugs [190].

### 3.15 Gastrointestinal Patch Systems

The use of a multilayered patch system is proposed to induce greater levels of absorption and stability at the intestinal epithelium. Patches comprise layers of thin, flexible membranes: an impermeable backing, a drug reservoir, a rate-controlling membrane, and an adhesive [191]. The pharmacological availability of model drugs in a multilayered patch system, for example, insulin [192], human G-CSF [193], interferon-α (IFN-α) [194], and EPO [195] had been tested.

### 3.16 Hydrogels

A hydrogel is a network of hydrophilic polymers that can swell in water and hold a large amount of water while maintaining its structure [196]. Hydrogel is sensitive and responsive to environmental variables. Thus, a pH-sensitive hydrogel is the practical application for delayed-release systems. The total amount of the drugs released should be determined to compare the rate between erosion rate and release rate [197]. Poly(2-hydroxyethyl methacrylate) (PHEMA) hydrogels can modify drug release profiles, reducing the early burst effect [198] and can make layers together with poly(methyacrylic acid) (PMAA), resulting in unidirectional release [199]. Encapsulation of cross-linked subtilisin microcrystals in hydrogel beads is used for controlled-release applications [200]. Dual cross-linkage effectively promoted the stability of beads under gastrointestinal tract conditions [26, 201].

Complexation hydrogels assessed transport-enhancing effects on the insulin-loaded polymer [202–204]. The oral administration of insulin using pH-responsive hydrogels has a similar effect on subcutaneous delivery [205, 206].

A series of poly[methacrylic acid-grafted-poly(ethylene glycol)] [P(MAA-*g*-EG)] with different polymer compositions are newly designed complexation hydrogels [207–209].

In addition, a hydrogel group of superporous hydrogels (SPH) and SPH composite (SPHC) have also been developed for prolonging the residence time of delivery systems in the GI tract with a new mechanism of action [210]. rhEGF loaded on Eudispert hv (poly(ethylacrylate/methacrylic acid) 50:50) hydrogel with sodium lauryl sulfate (SLS) prolongs the release from the formulation at the site of action [211].

### 3.17 Lipidization

The lipidization strategy involves the addition of lipid-like molecules through the modification of the hydrophilic moieties on the drug structure. [212]. The AUC of oral fatty acid–polypeptide conjugation of sCT was at least 19 times higher than that of sCT [213].

### 3.18 Oral Vaccination

Currently, the oral vaccine formulations approved by the Food and Drug Administration (FDA) are enteric coated adenovirus tablets, typhoid fever bacteria in gelatin capsules, polio virus in liquid, and the lyophilized rotavirus [214]. The successful use of live attenuated

viral and bacterial vaccines depends not only on the proper choice and delivery of the microorganisms but also on maintaining the sufficient potency required for an immune response [215]. *Lactobacillus* is utilized as a potent oral vaccine carrier. Vaccination purposes have been obtained with tetanus toxin fragment C (TTFC) as the model antigen [216]. De Jonge et al., [217] use an amorphous inulin matrix as a stabilizer to generate influenza virosomes as a stable freeze-dry (lyophilization) powder formulation. The living rate of engineered yeasts secreting active compounds directly in the digestive tract is more than 80% [218] and the effects of using cryoprotectants on them were assessed by an artificial digestive system [219]. The effective delivery of antigens via the oral route is an extremely desirable goal. Mucosal delivery of antigens stimulates mucosal and systemic immunity without affecting maternal antibodies and reduces the need for sterile needles or trained personnel [220].

Two of the most promising potential approaches for oral vaccination are the use of lectins [221] to target vaccines to the epithelium and the association of vaccines with biodegradable microparticles [217, 222–229]. Attending to the need for new vaccine adjuvants, a highly stable cochleate structure from Proteoliposomes is suggested for antimeningococcal vaccination [230]. Lipid vesicles containing bile salts (Bilosome) can successfully induce systemic and mucosal immune responses, characterized by IgG and IgA antibodies, respectively. [231–233]

## 4 FUTURE WORK

### 4.1 Vitamin $B_{12}$ Conjugates

Vitamin $B_{12}$ conjugates to deliver protein drugs (e.g., G-CSF, EPO, and insulin) by the oral route have been pursued by a few investigators [82, 83, 234]. This system relies upon the natural uptake mechanism for vitamin $B_{12}$ to co-transport peptides and proteins linked to the vitamin $B_{12}$, from the intestine to the circulatory system. In an exciting extension of the technology, it has been found that it is also possible to transport nanoparticles, linked to the vitamin $B_{12}$, into the circulatory system. Such nanoparticles can potentially be loaded with peptides or proteins of choice, and so protect these molecules from degradation in the intestine, while simultaneously transporting them into the circulatory system.

### 4.2 Potent Inhibitors/Enhancers

Ideal protease inhibitors are effective at preventing or suppressing enzymatic degradation in the intestine. Bowman-Birk inhibitor (BBI) is a protein with a molecular weight of 8000 Da with a well-characterized ability to inhibit trypsin and chymotrypsin [235, 236]. Developing reversal agents for P-gp, which can substantially increase oral absorption [237], could be an alternate approach to aid the inhibitor strategy. Promising strategy would be the use of polymeric inhibitors because they would offer the advantage of remaining more concentrated in the GI tract at the site of drug absorption and of not being absorbed per se, thus avoiding systemic toxic side effects [238]. Enhanced mucoadhesive properties of thiolated chitosan increase the inhibition of P-gp substrates [238, 239], pharmacological efficacy of insulin [240], and pharmacological efficacy of sCT [241].

### 4.3 Assay Development

An increased understanding of the many factors that can alter the analysis and pharmacokinetics of protein drugs and the development of simple yet sensitive and specific analysis methods are essential for protein drug oral delivery. Enzyme-linked immunosorbent assays (ELISA) have become the most common method of detection, and commercial kits are available for a wide variety of protein drugs. Monoclonal antibody products are sensitive, have minimal cross-reactivity, and are relatively inexpensive when compared with high-performance liquid chromatography (HPLC). However, the primary limitation of these assays is their inability to measure biologically active proteins. Conversely, bioassays do measure a biological event, that is, proliferation or cytotoxicity but are generally not used because of their high cost, long assay completion time, lack of specificity, poor sensitivity, and influence of environmental conditions on the outcome.

## 5 CONCLUDING REMARKS

The GI enzymatic barrier for protein oral delivery is formidable. Overcoming this barrier is one of the technical challenges for the pharmaceutical scientist. Other obstacles associated with protein oral delivery remain challenging as well. However, significant progress has been made with each of the obstacles and with certain proteins. The development of the composite formulation, which improves bioavailability yet meets regulatory requirements for reproducibility and intra and intersubject variability, and reasonable manufacturing costs, will be difficult, but the future of this infant field is bright.

## REFERENCES


1. Oh EJ, et al. (2007). Drug delivery systems for protein therapeutics. *Polym Sci Technol 18*(5), 444–449.
2. Dudzinski DM, Kesselheim AS (2008). Scientific and legal viability of follow-on protein drugs. *N Engl J Med 358*(8), 843–849.
3. Fix JA (1996). Oral controlled release technology for peptides: Status and future prospects. *Pharm Res 13*(12), 1760–1764.
4. Cleland JL, Daugherty A, Mrsny R (2001). Emerging protein delivery methods. *Curr Opin Biotechnol 12*(2), 212–219.
5. Goldberg M, Gomez-Orellana I (2003). Challenges for the oral delivery of macromolecules. *Nat Rev Drug Discov 2*(4), 289–295.
6. Yang CY, Dantzig AH, Pidgeon C (1999). Intestinal peptide transport systems and oral drug availability. *Pharm Res 16*(9), 1331–1343.
7. Lee VH (2000). Membrane transporters. *Eur J Pharm Sci 11*(Suppl 2), S41–S50.
8. Kreuter J (1991). Peroral administration of nanoparticles. *Adv Drug Deliv Rev 7*, 71–86.
9. Thomas NW, et al. (1996). Particle uptake and translocation across epithelial membranes. *J Anat 189*(Pt 3), 487–490.
10. Fasano A (1998). Novel approaches for oral delivery of macromolecules. *J Pharm Sci 87*(11), 1351–1356.
11. Fasano A (1998). Modulation of intestinal permeability: An innovative method of oral drug delivery for the treatment of inherited and acquired human diseases. *Mol Genet Metab 64*(1), 12–18.
12. Fasano A (1998). Innovative strategies for the oral delivery of drugs and peptides. *Trends Biotechnol 16*(4), 152–157.
13. Kompella UB, Lee VH (2001). Delivery systems for penetration enhancement of peptide and protein drugs: Design considerations. *Adv Drug Deliv Rev 46*(1–3), 211–245.
14. Woodley JF (1994). Enzymatic barriers for GI peptide and protein delivery. *Crit Rev Ther Drug Carrier Syst 11*(2–3), 61–95.
15. Garrido AB Jr, Freeman HJ, Kim YS (1981). Amino acid and peptide absorption in bypassed jejunum following jejunoileal bypass in rats. *Dig Dis Sci 26*(2), 107–112.
16. Kararli TT (1995). Comparison of the gastrointestinal anatomy, physiology, and biochemistry of humans and commonly used laboratory animals. *Biopharm Drug Dispos 16*(5), 351–380.
17. Grevel J, et al. (1986). Pharmacokinetics of oral cyclosporin A (Sandimmun) in healthy subjects. *Eur J Clin Pharmacol 31*(2), 211–216.
18. Lundin S, Vilhardt H (1986). Absorption of 1-deamino-8-D-arginine vasopressin from different regions of the gastrointestinal tract in rabbits. *Acta Endocrinol (Copenh) 112*(3), 457–460.
19. Kohler E, et al. (1987). Absorption of an aqueous solution of a new synthetic somatostatin analogue administered to man by gavage. *Eur J Clin Pharmacol 33*(2), 167–171.
20. Owen RL, Bhalla DK (1983). Cytochemical analysis of alkaline phosphatase and esterase activities and of lectin-binding and anionic sites in rat and mouse Peyer's patch M cells. *Am J Anat 168*(2), 199–212.
21. Gonnella PA, et al. (1987). Transepithelial transport of epidermal growth factor by absorptive cells of suckling rat ileum. *J Clin Invest 80*(1), 22–32.
22. Cornell R, Padykula HA (1969). A cytological study of intestinal absorption in the suckling rat. *Am J Anat 125*(3), 291–315.
23. Gonnella PA, Neutra MR (1984). Membrane-bound and fluid-phase macromolecules enter separate prelysosomal compartments in absorptive cells of suckling rat ileum. *J Cell Biol 99*(3), 909–917.
24. Low SC, et al. (2005). Oral and pulmonary delivery of FSH-Fc fusion proteins via neonatal Fc receptor-mediated transcytosis. *Hum Reprod 20*(7), 1805–1813.
25. Kitchens KM, et al. (2007). Endocytosis and interaction of poly (amidoamine) dendrimers with Caco-2 cells. *Pharm Res 24*(11), 2138–2145.
26. Foss AC, Peppas NA (2004). Investigation of the cytotoxicity and insulin transport of acrylic-based copolymer protein delivery systems in contact with Caco-2 cultures. *Eur J Pharm Biopharm 57*(3), 447–455.
27. Torres-Lugo M, et al. (2002). Physicochemical behavior and cytotoxic effects of p(methacrylic acid-*g*-ethylene glycol) nanospheres for oral delivery of proteins. *J Control Release 80*(1–3), 197–205.
28. Paine MF, et al. (1997). Characterization of interintestinal and intraintestinal variations in human CYP3A-dependent metabolism. *J Pharmacol Exp Ther 283*(3), 1552–1562.
29. Fojo AT, et al. (1987). Expression of a multidrug-resistance gene in human tumors and tissues. *Proc Natl Acad Sci U S A 84*(1), 265–269.
30. Woodley JF (1992). Peptidase enzymes of G.I. tract; barriers to peptide delivery, but potential for controlled release. *Proc Intl Symp Control Rel Bioact Mater 19*, 2–5.
31. Vaes G (1973). Digestive capacity of lysosomes. In Hers HG, Van Hoof F. Eds. *Lysosomes and Storage Diseases*. London: Academic, pp. 43–59.
32. Wen JY, et al. (2002). Enzymatic degradation of luteinizing hormone releasing hormone (LHRH) by mucosal homogenates from the intestine of the common brushtail possum (*Trichosurus vulpecula*). *Life Sci 71*(25), 3019–3030.
33. Anderle P, et al. (2002). In vitro assessment of intestinal IGF-I stability. *J Pharm Sci 91*(1), 290–300.

34. Su SF, Amidon GL, Lee HJ (2002). Intestinal metabolism and absorption of cholecystokinin analogs in rats. *Biochem Biophys Res Commun 292*(3), 632–638.
35. Lee HJ, Lee MG (2002). Controlling absorption site, metabolism, and membrane permeability in the intestine to develop strategies for protein drug oral delivery. Abstract presented at the 16th Annual Meeting of the AAPS. Toronto, Canada, November 2002.
36. Wacher VJ, et al. (1998). Role of P-glycoprotein and cytochrome P450 3A in limiting oral absorption of peptides and peptidomimetics. *J Pharm Sci 87*(11), 1322–1330.
37. Wacher VJ, Salphati L, Benet LZ (2001). Active secretion and enterocytic drug metabolism barriers to drug absorption. *Adv Drug Deliv Rev 46*(1–3), 89–102.
38. Schellens JH, et al. (2000). Modulation of oral bioavailability of anticancer drugs: From mouse to man. *Eur J Pharm Sci 12*(2), 103–110.
39. DeVane CL, Gill HS (1997). Clinical pharmacokinetics of fluvoxamine: Applications to dosage regimen design. *J Clin Psychiatry 58*(Suppl 5), 7–14.
40. Lehr CM (1994). Bioadhesion technologies for the delivery of peptide and protein drugs to the gastrointestinal tract. *Crit Rev Ther Drug Carrier Syst 11*(2–3), 119–160.
41. Watts TL, Fasano A (2000). Modulation of intestinal permeability: A novel and innovative approach for the oral delivery of drugs, macromolecules and antigens. *Biotechnol Genet Eng Rev 17*, 433–453.
42. Lipka E, Crison J, Amidon GL (1996). Transmembrane transport of peptide type compounds: Prospects for oral delivery. *J Control Release 39*(2–3), 121–129.
43. Lane ME, O'driscoll CM, Corrigan OI (2005). Quantitative estimation of the effects of bile salt surfactant systems on insulin stability and permeability in the rat intestine using a mass balance model. *J Pharm Pharmacol 57*(2), 169–175.
44. Shah RB, Palamakula A, Khan MA (2004). Cytotoxicity evaluation of enzyme inhibitors and absorption enhancers in Caco-2 cells for oral delivery of salmon calcitonin. *J Pharm Sci 93*(4), 1070–1082.
45. Sang Yoo H, Gwan Park T (2004). Biodegradable nanoparticles containing protein–fatty acid complexes for oral delivery of salmon calcitonin. *J Pharm Sci 93*(2), 488–495.
46. Ke WT, et al. (2005). Physical characterizations of microemulsion systems using tocopheryl polyethylene glycol 1000 succinate (TPGS) as a surfactant for the oral delivery of protein drugs. *J Control Release 102*(2), 489–507.
47. Maher S, et al. (2007). Melittin as a permeability enhancer II: In vitro investigations in human mucus secreting intestinal monolayers and rat colonic mucosae. *Pharm Res 24*(7), 1346–1356.
48. Kitchens KM, et al. (2006). Transport of poly(amidoamine) dendrimers across Caco-2 cell monolayers: Influence of size, charge and fluorescent labeling. *Pharm Res 23*(12), 2818–2826.
49. Arimori K, Nakano M (1998). Drug exsorption from blood into the gastrointestinal tract. *Pharm Res 15*(3), 371–376.
50. Fricker G, Drewe J (1996). Current concepts in intestinal peptide absorption. *J Pept Sci 2*(4), 195–211.
51. Fromm MF (2000). P-glycoprotein: A defense mechanism limiting oral bioavailability and CNS accumulation of drugs. *Int J Clin Pharmacol Ther 38*(2), 69–74.
52. Suzuki H, Sugiyama Y (2000). Role of metabolic enzymes and efflux transporters in the absorption of drugs from the small intestine. *Eur J Pharm Sci 12*(1), 3–12.
53. Oh DM, Han HK, Amidon GL (1999). Drug transport and targeting. Intestinal transport. *Pharm Biotechnol 12*, 59–88.
54. Liang XJ, Aszalos A (2006). Multidrug transporters as drug targets. *Curr Drug Targets 7*(8), 911–921.
55. Varma MV, Perumal OP, Panchagnula R (2006). Functional role of P-glycoprotein in limiting peroral drug absorption: Optimizing drug delivery. *Curr Opin Chem Biol 10*(4), 367–373.
56. Wasan KM (2001). Formulation and physiological and biopharmaceutical issues in the development of oral lipid-based drug delivery systems. *Drug Dev Ind Pharm 27*(4), 267–276.
57. Lee VHL, Dodda-Kashi S, Rubas GMGaW (1991). Protease inhibitors and penetration enhancers as approaches to modify peptide absorption. *Journal of Control Release 13*, 213–223.
58. Bernkop-Schnurch A (2000). Chitosan and its derivatives: Potential excipients for peroral peptide delivery systems. *Int J Pharm 194*(1), 1–13.
59. Leone-Bay A, Paton DR, Weidner JJ (2000). The development of delivery agents that facilitate the oral absorption of macromolecular drugs. *Med Res Rev 20*(2), 169–186.
60. Merkle HP (1994). New aspects of pharmaceutical dosage forms for controlled drug delivery of peptide and proteins. *Eur J Pharm Sci 2*, 19–21.
61. Gregg CR (1999). Drug interactions and anti-infective therapies. *Am J Med 106*(2), 227–237.
62. Miller DB, Spence JD (1998). Clinical pharmacokinetics of fibric acid derivatives (fibrates). *Clin Pharmacokinet 34*(2), 155–162.
63. Schmitz T, Huck CW, Bernkop-Schnurch A (2006). Characterisation of the thiol-disulphide chemistry of desmopressin by LC, mu-LC, LC-ESI-MS and Maldi-Tof. *Amino Acids 30*(1), 35–42.
64. Beierle I, Meibohm B, Derendorf H (1999). Gender differences in pharmacokinetics and pharmacodynamics. *Int J Clin Pharmacol Ther 37*(11), 529–547.
65. Harris RZ, Benet LZ, Schwartz JB (1995). Gender effects in pharmacokinetics and pharmacodynamics. *Drugs 50*(2), 222–239.
66. Thurmann PA, Hompesch BC (1998). Influence of gender on the pharmacokinetics and pharmacodynamics of drugs. *Int J Clin Pharmacol Ther 36*(11), 586–590.

67. Tanaka E (1999). Gender-related differences in pharmacokinetics and their clinical significance. *J Clin Pharm Ther 24*(5), 339–346.
68. Verbeeck RK, Horsmans Y (1998). Effect of hepatic insufficiency on pharmacokinetics and drug dosing. *Pharm World Sci 20*(5), 183–192.
69. Yamamoto A, et al. (1994). Effects of various protease inhibitors on the intestinal absorption and degradation of insulin in rats. *Pharm Res 11*(10), 1496–1500.
70. Fujii S, et al. (1985). Promoting effect of the new chymotrypsin inhibitor FK-448 on the intestinal absorption of insulin in rats and dogs. *J Pharm Pharmacol 37*(8), 545–549.
71. Foradori A, et al. (1998). Modification of the pharmacokinetics of cyclosporine A and metabolites by the concomitant use of Neoral and diltiazem or ketoconazol in stable adult kidney transplants. *Transplant Proc 30*(5), 1685–1687.
72. Ohtani S, et al. (2003). Evaluation of inhibitory activity of casein on proteases in rat intestine. *Pharm Res 20*(4), 611–617.
73. Lee HJ, Amidon GL (2002). The effect of enzyme inhibitor and absorption site following [D-ala2, D-leu5]enkephalin oral administration in rats. *Biopharm Drug Dispos 23*(4), 131–141.
74. Bernkop-Schnurch A, Bratengeyer I, Valenta C (1997). Development and in vitro evaluation of a drug delivery system protecting from trypsinic degradation. *Int J Pharm 157*, 17–25.
75. Bernkop-Schnurch A, Schwarz GH, Kratzel M (1997). Modified mucoadhesive polymers for the peroral administration of mainly elastase degradable therapeutic (po1y)peptidesase *N. J Control Rel 47*, 113–121.
76. Smoum R, Rubinstein A, Srebnik M (2006). Chitosan-pentaglycine-phenylboronic acid conjugate: A potential colon-specific platform for calcitonin. *Bioconjug Chem 17*(4), 1000–1007.
77. Werle M, et al. (2007). Design and evaluation of a chitosan-aprotinin conjugate for the peroral delivery of therapeutic peptides and proteins susceptible to enzymatic degradation. *J Drug Target 15*(5), 327–333.
78. Shah RB, Khan MA (2004). Protection of salmon calcitonin breakdown with serine proteases by various ovomucoid species for oral drug delivery. *J Pharm Sci 93*(2), 392–406.
79. Kratzel M, Hiessbock R, Bernkop-Schnurch A (1998). Auxiliary agents for the peroral administration of peptide and protein drugs: Synthesis and evaluation of novel pepstatin analogues. *J Med Chem 41*(13), 2339–2344.
80. Leone-Bay A, et al. (1996). 4-[4-[(2-Hydroxybenzoyl) amino]phenyl]butyric acid as a novel oral delivery agent for recombinant human growth hormone. *J Med Chem 39*(13), 2571–2578.
81. Stoll BR, et al. (2000). A mechanistic analysis of carrier-mediated oral delivery of protein therapeutics. *J Control Release 64*(1–3), 217–228.
82. Russell-Jones GJ (1998). Use of vitamin B12 conjugates to deliver protein drugs by the oral route. *Crit Rev Ther Drug Carrier Syst 15*(6), 557–586.
83. Russell-Jones GJ, Arthur L, Walker H (1999). Vitamin B12-mediated transport of nanoparticles across Caco-2 cells. *Int J Pharm 179*(2), 247–255.
84. Jung T, et al. (2000). Biodegradable nanoparticles for oral delivery of peptides: Is there a role for polymers to affect mucosal uptake? *Eur J Pharm Biopharm 50*(1), 147–160.
85. Lim CJ, Shen WC (2005). Comparison of monomeric and oligomeric transferrin as potential carrier in oral delivery of protein drugs. *J Control Release 106*(3), 273–286.
86. Ni T, et al. (2007). Oral route of mini-proinsulin-expressing *Ganoderma lucidum* decreases blood glucose level in streptozocin-induced diabetic rats. *Int J Mol Med 20*(1), 45–51.
87. Samanen J, et al. (1996). Chemical approaches to improve the oral bioavailability of peptidergic molecules. *J Pharm Pharmacol 48*(2), 119–135.
88. Filikov AV, et al. (2002). Computational stabilization of human growth hormone. *Protein Sci 11*(6), 1452–1461.
89. Li F, et al. (2006). Transport of levovirin prodrugs in the human intestinal Caco-2 cell line. *J Pharm Sci 95*(6), 1318–1325.
90. Zhang N, et al. (2006). Transport characteristics of wheat germ agglutinin-modified insulin-liposomes and solid lipid nanoparticles in a perfused rat intestinal model. *J Nanosci Nanotechnol 6*(9–10), 2959–2966.
91. Keljo DJ, Hamilton JR (1983). Quantitative determination of macromolecular transport rate across intestinal Peyer's patches. *Am J Physiol 244*(6), G637–644.
92. Senel S, et al. (2001). Delivery of bioactive peptides and proteins across oral (buccal) mucosa. *Curr Pharm Biotechnol 2*(2), 175–186.
93. Walker GF, Langoth N, Bernkop-Schnurch A (2002). Peptidase activity on the surface of the porcine buccal mucosa. *Int J Pharm 233*(1–2), 141–147.
94. Kurosaki Y, Kimura T (2000). Regional variation in oral mucosal drug permeability. *Crit Rev Ther Drug Carrier Syst 17*(5), 467–508.
95. Nagai T (1988). [Drug delivery systems by controlled release]. *Yakugaku Zasshi 108*(7), 613–624.
96. Boulmedarat L, et al. (2005). Evaluation of buccal methyl-beta-cyclodextrin toxicity on human oral epithelial cell culture model. *J Pharm Sci 94*(6), 1300–1309.
97. Langoth N, et al. (2006). Thiolated chitosans: Design and in vivo evaluation of a mucoadhesive buccal peptide drug delivery system. *Pharm Res 23*(3), 573–579.
98. Starokadomskyy PL, Dubey IY (2006). New absorption promoter for the buccal delivery: Preparation and characterization of lysalbinic acid. *Int J Pharm 308*(1–2), 149–154.
99. Cui Z, Mumper RJ (2002). Buccal transmucosal delivery of calcitonin in rabbits using thin-film composites. *Pharm Res 19*(12), 1901–1906.

100. Spijkervet FK, Sonis ST (1998). New frontiers in the management of chemotherapy-induced mucositis. *Curr Opin Oncol 10*(Suppl 1), S23–27.
101. Alur HH, et al. (1999). Evaluation of a novel, natural oligosaccharide gum as a sustained-release and mucoadhesive component of calcitonin buccal tablets. *J Pharm Sci 88*(12), 1313–1319.
102. Nakane S, et al. (1996). Oramucosal delivery of LHRH: Pharmacokinetic studies of controlled and enhanced transmucosal permeation. *Pharm Dev Technol 1*(3), 251–259.
103. Gutniak MK, et al. (1996). Potential therapeutic levels of glucagon-like peptide I achieved in humans by a buccal tablet. *Diabetes Care 19*(8), 843–848.
104. Rubinstein A (1995). Approaches and opportunities in colon-specific drug delivery. *Crit Rev Ther Drug Carrier Syst 12*(2–3), 101–149.
105. Sinha VR, Kumria R (2003). Microbially triggered drug delivery to the colon. *Eur J Pharm Sci 18*(1), 3–18.
106. Saffran M, et al. (1986). A new approach to the oral administration of insulin and other peptide drugs. *Science 233*(4768), 1081–1084.
107. Chen L, et al. (2007). Resistant starch as a carrier for oral colon-targeting drug matrix system. *J Mater Sci Mater Med 18*(11), 2199–2203.
108. Jones DS, et al. (2007). An examination of the rheological and mucoadhesive properties of poly(acrylic acid) organogels designed as platforms for local drug delivery to the oral cavity. *J Pharm Sci 96*(10), 2632–2646.
109. Xing L, et al. (2003). Oral colon-specific drug delivery for bee venom peptide: Development of a coated calcium alginate gel beads-entrapped liposome. *J Control Release 93*(3), 293–300.
110. Baluom M, Friedman DI, Rubinstein A (1997). Absorption enhancement of calcitonin in the rat intestine by carbopolcontaining submicron emulsions. *Int J Pharm 154*, 235–243.
111. Macfarlane GT, et al. (1989). Influence of retention time on degradation of pancreatic enzymes by human colonic bacteria grown in a 3-stage continuous culture system. *J Appl Bacteriol 67*(5), 520–527.
112. Salman HH, et al. (2007). Bioadhesive capacity and immunoadjuvant properties of thiamine-coated nanoparticles. *Vaccine 25*(48), 8123–8132.
113. Metzner L, Brandsch M (2006). Influence of a proton gradient on the transport kinetics of the H+/amino acid cotransporter PAT1 in Caco-2 cells. *Eur J Pharm Biopharm 63*(3), 360–364.
114. Biegel A, et al. (2005). Three-dimensional quantitative structure-activity relationship analyses of beta-lactam antibiotics and tripeptides as substrates of the mammalian H+/peptide cotransporter PEPT1. *J Med Chem 48*(13), 4410–4419.
115. Calcagno AM, et al. (2006). Comparison of drug transporter levels in normal colon, colon cancer, and Caco-2 cells: Impact on drug disposition and discovery. *Mol Pharm 3*(1), 87–93.
116. Metzner L, et al. (2008). Mutational analysis of histidine residues in the human proton-coupled amino acid transporter PAT1. *Biochim Biophys Acta 1778*(4), 1042–1050.
117. Tsuji A, Tamai I (1996). Carrier-mediated intestinal transport of drugs. *Pharm Res 13*(7), 963–977.
118. Nielsen CU, Brodin B (2003). Di/tri-peptide transporters as drug delivery targets: Regulation of transport under physiological and patho-physiological conditions. *Curr Drug Targets 4*(5), 373–388.
119. Takeuchi H, et al. (2005). Effectiveness of submicron-sized, chitosan-coated liposomes in oral administration of peptide drugs. *Int J Pharm 303*(1–2), 160–170.
120. Donini C, et al. (2002). Preparation of poly(methacrylic acid-*g*-poly(ethylene glycol)) nanospheres from methacrylic monomers for pharmaceutical applications. *Int J Pharm 245*(1–2), 83–91.
121. Ma EL, et al. (2006). In vitro and in vivo evaluation of a novel oral insulin formulation. *Acta Pharmacol Sin 27*(10), 1382–1388.
122. Sakuma S, Hayashi M, Akashi M (2001). Design of nanoparticles composed of graft copolymers for oral peptide delivery. *Adv Drug Deliv Rev 47*(1), 21–37.
123. Prego C, et al. (2006). Chitosan-PEG nanocapsules as new carriers for oral peptide delivery. Effect of chitosan pegylation degree. *J Control Release 111*(3), 299–308.
124. Prego C, et al. (2006). Efficacy and mechanism of action of chitosan nanocapsules for oral peptide delivery. *Pharm Res 23*(3), 549–556.
125. Quintanar-Guerrero D, et al. (1998). Preparation techniques and mechanisms of formation of biodegradable nanoparticles from preformed polymers. *Drug Dev Ind Pharm 24*(12), 1113–1128.
126. Allemann E, Gurney R, Doelker E (1993). Drug loaded nanopaticles-preparation methods and drug targeting issues. *Eur J Pharm Biopharm 39*, 173–191.
127. Cui F, Qian F, Yin C (2006). Preparation and characterization of mucoadhesive polymer-coated nanoparticles. *Int J Pharm 316*(1–2), 154–161.
128. Ma Z, Yeoh HH, Lim LY (2002). Formulation pH modulates the interaction of insulin with chitosan nanoparticles. *J Pharm Sci 91*(6), 1396–1404.
129. Chen F, et al. (2008). In vitro and in vivo study of *N*-trimethyl chitosan nanoparticles for oral protein delivery. *Int J Pharm 349*(1–2), 226–233.
130. Sakuma S, et al. (2002). Optimized chemical structure of nanoparticles as carriers for oral delivery of salmon calcitonin. *Int J Pharm 239*(1–2), 185–195.
131. Jani P, et al. (1989). The uptake and translocation of latex nanospheres and microspheres after oral administration to rats. *J Pharm Pharmacol 41*(12), 809–812.
132. Soares AF, Carvalho Rde A, Veiga F (2007). Oral administration of peptides and proteins: Nanoparticles and

cyclodextrins as biocompatible delivery systems. *Nanomed 2*(2), 183–202.

133. Lin YH, et al. (2005). Physically crosslinked alginate/N,O-carboxymethyl chitosan hydrogels with calcium for oral delivery of protein drugs. *Biomaterials 26*(14), 2105–2113.
134. Li T, et al. (2007). Quaternized chitosan/alginate nanoparticles for protein delivery. *J Biomed Mater Res A 83*(2), 383–390.
135. Sarmento B, et al. (2007). Insulin-loaded nanoparticles are prepared by alginate ionotropic pre-gelation followed by chitosan polyelectrolyte complexation. *J Nanosci Nanotechnol 7*(8), 2833–2841.
136. Cuna M, et al. (2006). Development of phosphorylated glucomannan-coated chitosan nanoparticles as nanocarriers for protein delivery. *J Nanosci Nanotechnol 6*(9–10), 2887–2895.
137. Garcia-Fuentes M, Torres D, Alonso MJ (2005). New surface-modified lipid nanoparticles as delivery vehicles for salmon calcitonin. *Int J Pharm 296*(1–2), 122–132.
138. Garcia-Fuentes M, et al. (2005). A comparative study of the potential of solid triglyceride nanostructures coated with chitosan or poly(ethylene glycol) as carriers for oral calcitonin delivery. *Eur J Pharm Sci 25*(1), 133–143.
139. Cui F, et al. (2007). Nanoparticles incorporated in bilaminated films: A smart drug delivery system for oral formulations. *Biomacromolecules 8*(9), 2845–2850.
140. Prokop A, et al. (2002). Water-based nanoparticulate polymeric system for protein delivery: Permeability control and vaccine application. *Biotechnol Bioeng 78*(4), 459–466.
141. des Rieux A, et al. (2007). Helodermin-loaded nanoparticles: Characterization and transport across an in vitro model of the follicle-associated epithelium. *J Control Release 118*(3), 294–302.
142. Ito Y, et al. (2007). Effect of fiber length of carbon nanotubes on the absorption of erythropoietin from rat small intestine. *Int J Pharm 337*(1–2), 357–360.
143. Sarmento B, et al. (2007). Probing insulin's secondary structure after entrapment into alginate/chitosan nanoparticles. *Eur J Pharm Biopharm 65*(1), 10–17.
144. Park K, Robinson JR (1984). Bioadhesive polymers as platforms for oral-controlled drug delivery: Method to study bioadhesion. *Int J Pharm 19*, 107–127.
145. Peppas NA (2004). Devices based on intelligent biopolymers for oral protein delivery. *Int J Pharm 277*(1–2), 11–17.
146. Carreno-Gomez B, Woodley JF, Florence AT (1999). Studies on the uptake of tomato lectin nanoparticles in everted gut sacs. *Int J Pharm 183*(1), 7–11.
147. Lehr CM, et al. (1990). Intestinal transit of bioadhesive microspheres in an in situ loop in the rat. A comparative study with polymers and blends based on poly(acrylic acid). *J Control Rel 13*, 51–62.
148. Allen A, et al. (1984). Studies on gastrointestinal mucus. *Scand J Gastroenterol Suppl 93*, 101–113.
149. Smart JD, et al. (2002). A quantitative evaluation of radiolabelled lectin retention on oral mucosa in vitro and in vivo. *Eur J Pharm Biopharm 53*(3), 289–292.
150. Yin Y, et al. (2007). Lectin-conjugated PLGA nanoparticles loaded with thymopentin: Ex vivo bioadhesion and in vivo biodistribution. *J Control Release 123*(1), 27–38.
151. Kim BY, et al. (2005). Bioadhesive interaction and hypoglycemic effect of insulin-loaded lectin-microparticle conjugates in oral insulin delivery system. *J Control Release 102*(3), 525–538.
152. Tao SL, Lubeley MW, Desai TA (2003). Bioadhesive poly(methyl methacrylate) microdevices for controlled drug delivery. *J Control Release 88*(2), 215–228.
153. Garinot M, et al. (2007). PEGylated PLGA-based nanoparticles targeting M cells for oral vaccination. *J Control Release 120*(3), 195–204.
154. Lavelle EC, et al. (2004). Immunological implications of the use of plant lectins for drug and vaccine targeting to the gastrointestinal tract. *J Drug Target 12*(2), 89–95.
155. Yin Y, et al. (2006). Preparation and evaluation of lectin-conjugated PLGA nanoparticles for oral delivery of thymopentin. *J Control Release 116*(3), 337–345.
156. Gupta PN, et al. (2007). M-cell targeted biodegradable PLGA nanoparticles for oral immunization against hepatitis B. *J Drug Target 15*(10), 701–713.
157. Venkatesan N, et al. (2006). Pharmacokinetic and pharmacodynamic studies following oral administration of erythropoietin mucoadhesive tablets to beagle dogs. *Int J Pharm 310*(1–2), 46–52.
158. Ahmed A, Bonner C, Desai TA (2002). Bioadhesive microdevices with multiple reservoirs: A new platform for oral drug delivery. *J Control Release 81*(3), 291–306.
159. DiBiase MD, Morrel EM (1997). Oral delivery of microencapsulated proteins. *Pharm Biotechnol 10*, 255–288.
160. Takeuchi H, Yamamoto H, Kawashima Y (2001). Mucoadhesive nanoparticulate systems for peptide drug delivery. *Adv Drug Deliv Rev 47*(1), 39–54.
161. Thongborisute J, et al. (2006). Properties of liposomes coated with hydrophobically modified chitosan in oral liposomal drug delivery. *Pharmazie 61*(2), 106–111.
162. Thongborisute J, et al. (2006). The effect of particle structure of chitosan-coated liposomes and type of chitosan on oral delivery of calcitonin. *J Drug Target 14*(3), 147–154.
163. Carafa M, et al. (2006). Novel *O*-palmitoylscleroglucan-coated liposomes as drug carriers: Development, characterization and interaction with leuprolide. *Int J Pharm 325*(1–2), 155–162.
164. Li H, et al. (2003). Polyethylene glycol-coated liposomes for oral delivery of recombinant human epidermal growth factor. *Int J Pharm 258*(1–2), 11–19.
165. Lasic DD, Papahadjopoulos D (1995). Liposomes revisited. *Science 267*(5202), 1275–1276.

166. Anderson PM, et al. (1994). Cytokines in liposomes: Preliminary studies with IL-1, IL-2, IL-6, GM-CSF and interferon-gamma. *Cytokine 6*(1), 92–101.
167. Ye Q, et al. (2000). DepoFoam technology: A vehicle for controlled delivery of protein and peptide drugs. *J Control Release 64*(1–3), 155–166.
168. Masuda K, et al. (2002). Oral delivery of antigens in liposomes with some lipid compositions modulates oral tolerance to the antigens. *Microbiol Immunol 46*(1), 55–58.
169. Sandhya KV, Devi GS, Mathew ST (2008). Liposomal formulations of serratiopeptidase: In vitro studies using PAMPA and Caco-2 models. *Mol Pharm 5*(1), 92–97.
170. Pardakhty A, Varshosaz J, Rouholamini A (2007). In vitro study of polyoxyethylene alkyl ether niosomes for delivery of insulin. *Int J Pharm 328*(2), 130–141.
171. Jain S, Sharma RK, Vyas SP (2006). Chitosan nanoparticles encapsulated vesicular systems for oral immunization: Preparation, in-vitro and in-vivo characterization. *J Pharm Pharmacol 58*(3), 303–310.
172. Lee HJ, et al. (1997). In vivo characterization of sustained-release formulations of human growth hormone. *J Pharmacol Exp Ther 281*(3), 1431–1439.
173. Teply BA, et al. (2008). The use of charge-coupled polymeric microparticles and micromagnets for modulating the bioavailability of orally delivered macromolecules. *Biomaterials 29*(9), 1216–1223.
174. Cheng J, et al. (2006). Magnetically responsive polymeric microparticles for oral delivery of protein drugs. *Pharm Res 23*(3), 557–564.
175. Eldridge JH, et al. (1990). Controlled vaccine release in the gut-associated lymphoid tissues. I. Orally administered biodegradable microsphere target the Peyers patches. *J Control Release 11*, 205–214.
176. Ray R, et al. (1993). Microencapsulated human parainfluenza virus induces a protective immune response. *J Infect Dis 167*(3), 752–755.
177. Larkins BA, et al. (1984). Zein gene organization in maize and related grasses. *J Mol Biol 20*(3–4), 330–340.
178. Hurtado-Lopez P, Murdan S (2006). Zein microspheres as drug/antigen carriers: A study of their degradation and erosion, in the presence and absence of enzymes. *J Microencapsul 23*(3), 303–314.
179. Shi XW, et al. (2005). Ionically crosslinked alginate/carboxymethyl chitin beads for oral delivery of protein drugs. *Macromol Biosci 5*(9), 881–889.
180. Reddy T, Tammishetti S (2002). Gastric resistant microbeads of metal ion cross-linked carboxymethyl guar gum for oral drug delivery. *J Microencapsul 19*(3), 311–318.
181. Rawat M, Saraf S, Saraf S (2007). Influence of selected formulation variables on the preparation of enzyme-entrapped Eudragit S100 microspheres. *AAPS PharmSciTech 8*(4), E116.
182. Calceti P, et al. (2004). Development and in vivo evaluation of an oral insulin-PEG delivery system. *Eur J Pharm Sci 22*(4), 315–323.
183. Morcol T, et al. (2004). Calcium phosphate-PEG-insulin-casein (CAPIC) particles as oral delivery systems for insulin. *Int J Pharm 277*(1–2), 91–97.
184. Rawat M, et al. (2008). Development and in vitro evaluation of alginate gel-encapsulated, chitosan-coated ceramic nanocores for oral delivery of enzyme. *Drug Dev Ind Pharm 34*(2), 181–188.
185. Torres-Lugo M, Peppas NA (2000). Transmucosal delivery systems for calcitonin: A review. *Biomaterials 21*(12), 1191–1196.
186. Masuda K, et al. (2003). Oral-antigen delivery via a water-in-oil emulsion system modulates the balance of the Th1/Th2 type response in oral tolerance. *Pharm Res 20*(1), 130–134.
187. New RRC, et al. (1994). Administration of calcitonin to humans: Comparison of intra-nasal and oral routes. Paper presented at the Third European Symposium on Controlled Release Drug Delivery. Noordwijkaan Zee, The Netherlands, April 6–8, 1994.
188. Kamphorst AO, et al. (2004). Association complexes between ovalbumin and cyclodextrins have no effect on the immunological properties of ovalbumin. *Eur J Pharm Biopharm 57*(2), 199–205.
189. Garcia-Fuentes M, Trapani A, Alonso MJ (2006). Protection of the peptide glutathione by complex formation with alpha-cyclodextrin: NMR spectroscopic analysis and stability study. *Eur J Pharm Biopharm 64*(2), 146–153.
190. Sajeesh S, Sharma CP (2006). Cyclodextrin-insulin complex encapsulated polymethacrylic acid based nanoparticles for oral insulin delivery. *Int J Pharm 325* (1–2), 147–154.
191. Tao SL, Desai TA (2005). Gastrointestinal patch systems for oral drug delivery. *Drug Discov Today 10*(13), 909–915.
192. Whitehead K, Shen Z, Mitragotri S (2004). Oral delivery of macromolecules using intestinal patches: Applications for insulin delivery. *J Control Release 98*(1), 37–45.
193. Eiamtrakarn S, et al. (2002). Gastrointestinal mucoadhesive patch system (GI-MAPS) for oral administration of G-CSF, a model protein. *Biomaterials 23*(1), 145–152.
194. Ito Y, et al. (2005). Absorption of interferon alpha from patches in rats. *J Drug Target 13*(6), 383–390.
195. Venkatesan N, et al. (2006). Gastro-intestinal patch system for the delivery of erythropoietin. *J Control Release 111*(1–2), 19–26.
196. Qiu Y, Park K (2001). Environment-sensitive hydrogels for drug delivery. *Adv Drug Deliv Rev 53*(3), 321–339.
197. Nam K, Watanabe J, Ishihara K (2004). Modeling of swelling and drug release behavior of spontaneously forming hydrogels composed of phospholipid polymers. *Int J Pharm 275*(1–2), 259–269.
198. Wu L, Brazel CS (2008). Surface crosslinking for delayed release of proxyphylline from PHEMA hydrogels. *Int J Pharm 349*(1–2), 1–10.

199. He H, Guan J, Lee JL (2006). An oral delivery device based on self-folding hydrogels. *J Control Release 110*(2), 339–346.
200. Simi CK, Emilia Abraham T (2007). Encapsulation of crosslinked subtilisin microcrystals in hydrogel beads for controlled release applications. *Eur J Pharm Sci 32*(1), 17–23.
201. Xu Y, et al. (2007). Preparation of dual crosslinked alginate-chitosan blend gel beads and in vitro controlled release in oral site-specific drug delivery system. *Int J Pharm 336*(2), 329–337.
202. Perakslis E, Tuesca A, Lowman A (2007). Complexation hydrogels for oral protein delivery: An in vitro assessment of the insulin transport-enhancing effects following dissolution in simulated digestive fluids. *J Biomater Sci Polym Ed 18*(12), 1475–1490.
203. Kavimandan NJ, Losi E, Peppas NA (2006). Novel delivery system based on complexation hydrogels as delivery vehicles for insulin-transferrin conjugates. *Biomaterials 27*(20), 3846–3854.
204. Sajeesh S, Sharma CP (2006). Novel pH responsive polymethacrylic acid-chitosan-polyethylene glycol nanoparticles for oral peptide delivery. *J Biomed Mater Res B Appl Biomater 76*(2), 298–305.
205. Madihally SV, Pantelogianis A, Toner M (2006). Antiproteolytic action of orally delivered insulin using pH-responsive hydrogels in a rat burn model. *J Surg Res 135*(1), 187–194.
206. Besheer A, et al. (2006). Loading and mobility of spin-labeled insulin in physiologically responsive complexation hydrogels intended for oral administration. *J Control Release 111*(1–2), 73–80.
207. Kim B, Peppas NA (2003). In vitro release behavior and stability of insulin in complexation hydrogels as oral drug delivery carriers. *Int J Pharm 266*(1–2), 29–37.
208. Nam KW, Watanabe J, Ishihara K (2002). pH-modulated release of insulin entrapped in a spontaneously formed hydrogel system composed of two water-soluble phospholipid polymers. *J Biomater Sci Polym Ed 13*(11), 1259–1269.
209. Ichikawa H, Peppas NA (2003). Novel complexation hydrogels for oral peptide delivery: In vitro evaluation of their cytocompatibility and insulin-transport enhancing effects using Caco-2 cell monolayers. *J Biomed Mater Res A 67*(2), 609–617.
210. Dorkoosh FA, et al. (2004). Feasibility study on the retention of superporous hydrogel composite polymer in the intestinal tract of man using scintigraphy. *J Control Release 99*(2), 199–206.
211. Park JS, et al. (2003). Buccal mucosal ulcer healing effect of rhEGF/Eudispert hv hydrogel. *Arch Pharm Res 26*(8), 659–665.
212. Garcia-Garcia E, et al. (2005). Colloidal carriers and blood-brain barrier (BBB) translocation: A way to deliver drugs to the brain? *Int J Pharm 298*(2), 274–292.
213. Wang J, et al. (2003). Reversible lipidization for the oral delivery of salmon calcitonin. *J Control Release 88*(3), 369–380.
214. Burke CJ, Hsu TA, Volkin DB (1999). Formulation, stability, and delivery of live attenuated vaccines for human use. *Crit Rev Ther Drug Carrier Syst 16*(1), 1–83.
215. Babiuk S, et al. (2000). Cutaneous vaccination: The skin as an immunologically active tissue and the challenge of antigen delivery. *J Control Release 66*(2–3), 199–214.
216. Seegers JF (2002). Lactobacilli as live vaccine delivery vectors: Progress and prospects. *Trends Biotechnol 20*(12), 508–515.
217. de Jonge J, et al. (2007). Inulin sugar glasses preserve the structural integrity and biological activity of influenza virosomes during freeze-drying and storage. *Eur J Pharm Sci 32*(1), 33–44.
218. Blanquet S, et al. (2004). Living recombinant *Saccharomyces cerevisiae* secreting proteins or peptides as a new drug delivery system in the gut. *J Biotechnol 110*(1), 37–49.
219. Blanquet S, et al. (2005). Effects of cryoprotectants on the viability and activity of freeze dried recombinant yeasts as novel oral drug delivery systems assessed by an artificial digestive system. *Eur J Pharm Biopharm 61* (1–2), 32–39.
220. Foster N, Hirst BH (2005). Exploiting receptor biology for oral vaccination with biodegradable particulates. *Adv Drug Deliv Rev 57*(3), 431–450.
221. Lambkin I, et al. (2003). Toward targeted oral vaccine delivery systems: Selection of lectin mimetics from combinatorial libraries. *Pharm Res 20*(8), 1258–1266.
222. Ahire VJ, et al. (2007). Chitosan microparticles as oral delivery system for tetanus toxoid. *Drug Dev Ind Pharm 33*(10), 1112–1124.
223. Bharali DJ, Mousa SA, Thanavala Y (2007). Micro- and nanoparticle-based vaccines for hepatitis B. *Adv Exp Med Biol 601*, 415–421.
224. Yeh M, Chiang C (2004). Inactive *Vibrio cholerae* whole-cell vaccine-loaded biodegradable microparticles: In vitro release and oral vaccination. *J Microencapsul 21*(1), 91–106.
225. van der Lubben IM, et al. (2003). Chitosan microparticles for mucosal vaccination against diphtheria: Oral and nasal efficacy studies in mice. *Vaccine 21*(13–14), 1400–1408.
226. Ren JM, et al. (2002). PELA microspheres loaded *H. pylori* lysates and their mucosal immune response. *World J Gastroenterol 8*(6), 1098–1102.
227. Yeh MK, et al. (2002). Oral immunogenicity of the inactivated *Vibrio cholerae* whole-cell vaccine encapsulated in biodegradable microparticles. *J Control Release 82* (2–3), 237–247.
228. Li XY, et al. (2007). Chitosan-alginate microcapsules for oral delivery of egg yolk immunoglobulin (IgY). *J Agric Food Chem 55*(8), 2911–2917.

229. Kim B, et al. (2002). Mucosal immune responses following oral immunization with rotavirus antigens encapsulated in alginate microspheres. *J Control Release 85*(1–3), 191–202.
230. Perez O, et al. (2004). Novel adjuvant based on a proteoliposome-derived cochleate structure containing native lipopolysaccharide as a pathogen-associated molecular pattern. *Immunol Cell Biol 82*(6), 603–610.
231. Mann JF, et al. (2006). Oral delivery of tetanus toxoid using vesicles containing bile salts (bilosomes) induces significant systemic and mucosal immunity. *Methods 38*(2), 90–95.
232. Mann JF, et al. (2004). Optimisation of a lipid based oral delivery system containing A/Panama influenza haemagglutinin. *Vaccine 22*(19), 2425–2429.
233. Singh P, et al. (2004). Cholera toxin B subunit conjugated bile salt stabilized vesicles (bilosomes) for oral immunization. *Int J Pharm 278*(2), 379–390.
234. Chalasani KB, et al. (2007). A novel vitamin B12-nanosphere conjugate carrier system for peroral delivery of insulin. *J Control Release 117*(3), 421–429.
235. Bell DS, Ovalle F (2001). Use of soy protein supplement and resultant need for increased dose of levothyroxine. *Endocr Pract 7*(3), 193–194.
236. Kennedy AR (1998). The Bowman-Birk inhibitor from soybeans as an anticarcinogenic agent. *Am J Clin Nutr 68*(6, Suppl), 1406S–1412S.
237. Spahn-Langguth H, Langguth P (2001). Grapefruit juice enhances intestinal absorption of the P-glycoprotein substrate talinolol. *Eur J Pharm Sci 12*(4), 361–367.
238. Foger F, et al. (2006). In vivo comparison of various polymeric and low molecular mass inhibitors of intestinal P-glycoprotein. *Biomaterials 27*(34), 5855–5860.
239. Foger F, Schmitz T, Bernkop-Schnurch A (2006). In vivo evaluation of an oral delivery system for P-gp substrates based on thiolated chitosan. *Biomaterials 27*(23), 4250–4255.
240. Krauland AH, Guggi D, Bernkop-Schnurch A (2004). Oral insulin delivery: The potential of thiolated chitosan-insulin tablets on non-diabetic rats. *J Control Release 95*(3), 547–555.
241. Guggi D, Krauland AH, Bernkop-Schnurch A (2003). Systemic peptide delivery via the stomach: In vivo evaluation of an oral dosage form for salmon calcitonin. *J Control Release 92*(1–2), 125–135.

# AEROSOL DRUG DELIVERY TO THE LUNGS

Andrew R. Martin and Warren H. Finlay
*Department of Mechanical Engineering, University of Alberta Edmonton, Alberta, Canada*

## 1 INTRODUCTION

The delivery of therapeutic agents locally to the lungs through the inspiratory route is, in the majority of cases, the preferred course of treatment of respiratory diseases. Indeed, inhaled aerosols have long been employed in the management of asthma, cystic fibrosis, and chronic obstructive pulmonary disease (COPD). Gas-borne suspensions of fine liquid droplets or solid particles are introduced into the patient's inspiratory airflow to carry drugs into the lungs. Over recent years, it has become apparent that systemic drug delivery through the lungs to the bloodstream is an attractive alternative to more traditional methods of delivery, such as oral or intravenous administration. Targeting of inhaled aerosols to the lungs' alveolar regions allows for transport of deposited drug across the alveolar epithelium into the capillaries. Accordingly, aerosol drug delivery has emerged as a viable option for systemic administration of a large variety of drugs and will likely play an important role in the development of new therapies over the next decade.

The development and evaluation of aerosol drug delivery systems requires a level of comfort with fields such as aerosol mechanics, fluid mechanics, and transport phenomena that may be unfamiliar to newcomers whose training backgrounds lie outside of the physical sciences. The intent of the first section of this chapter is to provide a very brief introduction to aerosol mechanics as related to the transport and deposition of inhaled drug particles throughout the respiratory tract. This familiarization will serve as requisite background for the discussion of aerosol particle sizing instruments in use for research and development, regulatory, and quality assurance purposes, as well as for descriptions of the different types of aerosol delivery systems currently available and in development.

## 2 RESPIRATORY TRACT DEPOSITION OF INHALED PARTICLES

The human respiratory tract is made up of a complicated series of branching, or bifurcating, airways. Phillips et al. [1] have described in detail much of the influential work done in measuring and modeling respiratory tract geometry. Despite its limitations, the most widely adopted model remains the highly idealized, symmetric model given by Weibel [2], and referred to as the Weibel A model. Under the Weibel A model, airways are classified by a generation number (beginning with zero at the trachea), which indicates the number of bifurcations separating the airway from the trachea. Airway diameters and lengths become progressively smaller as the generation number increases. Generations 0–16 make up the tracheobronchial region, through which air is conducted upon inhalation and exhalation, while generations 17–23 comprise the alveolar region, wherein gas exchange with the capillaries occurs. In addition, the extrathoracic region, consisting of the oral and

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

nasal cavities, the pharynx, and the larynx, must be included in a complete description of the respiratory tract. Further discussion of respiratory tract structure and function, in the context of aerosol drug delivery, may be found in ICRP Publication 66 [3], in Altiere and Thompson [4], in Finlay [5], and in Stocks and Hislop [6].

The site within the respiratory tract at which an inhaled particle deposits is a key determinant of its subsequent pharmacokinetics. For topical treatment of certain lung diseases, it may be preferable to target the tracheobronchial airways, while for systemic delivery to the bloodstream, deposition deep into the alveolar region is required. Deposition in the extrathoracic region is undesirable for pulmonary delivery, so that these upper airways are viewed as a barrier that filters out particles intended to reach the lungs.

The size and mass of an inhaled particle affect its entrainment in inspiratory and expiratory airflow, thereby encouraging deposition in particular regions of the respiratory tract. For this reason, it is the convention in aerosol science to express particle size and mass in terms of an *aerodynamic* diameter, given as

$$d_{ae} = \sqrt{\frac{\rho_{particle}}{\rho_{water}}} d \tag{1}$$

where $d_{ae}$ is the particle aerodynamic diameter, $d$ is the particle geometric diameter, $\rho_{particle}$ is the density of the particle, and $\rho_{water}$ is the density of water.

For irregular particles, the aerodynamic diameter standardizes shape to that of a sphere and density to that of water. The assumption of spherical shape is in general reasonable, given that liquid droplets *will* be spherical due to surface tensions, while solid particles are usually compact, so that their motion through the air is not too much different from that of a sphere of equivalent volume.

Figure 1 displays average total respiratory tract deposition of inhaled aerosol particles as a function of aerodynamic diameter for three different breathing patterns, as measured by Stahlhofen et al. [7] in human subjects during tidal mouth breathing. A broad minimum of total deposition exists around aerodynamic diameters of about 300 or 400 nm. As the curve rises to the left and right of the minimum value, deposition tends to shift toward more proximal (i.e., closer to the mouth) regions of the respiratory tract. For example, particles larger than approximately 5 μm deposit primarily in the mouth, throat, and upper branches of the bronchial tree. The shape of the total deposition curve can be understood through consideration of the different physical mechanisms through which respiratory tract deposition occurs, as will now be introduced.

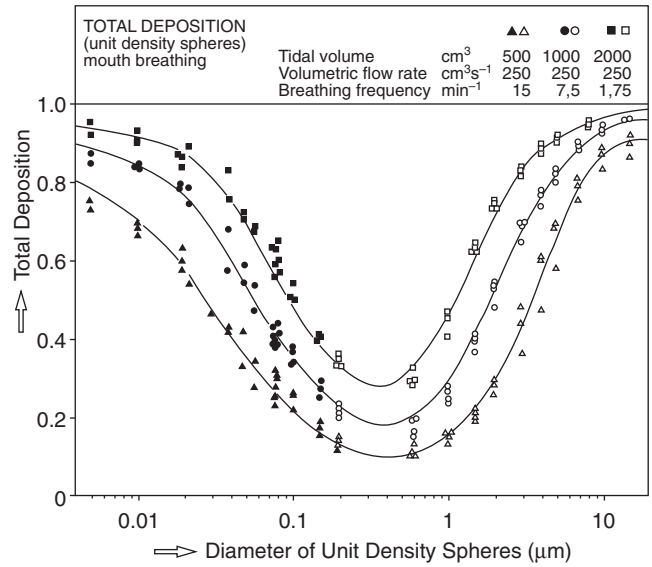


**Figure 1** Total respiratory tract deposition of inhaled particles as a function of the aerodynamic particle diameter for human subjects during tidal mouth breathing. Three different breathing patterns are shown. (From Stahlhofen et al. [7].)

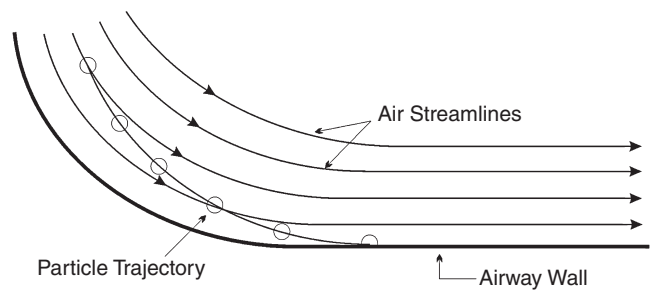


**Figure 2** Illustration of particle deposition by inertial impaction.

## 2.1 Factors Determining Deposition

***Particle Inertia*** Aerosol particles are carried into the lungs by inhaled airflow. Where air streamlines (i.e., lines everywhere tangent to the air velocity) are curved at bends in the respiratory airways, aerodynamic drag forces act on the particles to steer them along with the flow, around the bends. However, a particle's inertia can cause it to deviate from a streamline to the extent that it deposits on an airway wall, as illustrated in Figure 2. The situation can be thought of as analogous to an automobile that veers off the road as a result of traveling too quickly while attempting a sharp turn. The ratio between particle inertia and the drag force is described by the particle Stokes (Stk) number:

$$\mathrm{Stk} = \frac{U\rho_{\mathrm{particle}} d^2 C_C}{18\mu D} \tag{2}$$

where $U$ is the average air velocity in the airway, $\mu$ is the air viscosity, $D$ is the airway diameter, and $C_C$ is the Cunningham slip correction factor, which accounts for the fact that very small aerosol particles experience a reduced drag force owing to their ability to "slip" through spaces between gas molecules. For particles a few microns and larger in diameter, a typical size range for inhaled drug particles, $C_C$ is very close to unity and can be dropped from Equation (2) with little consequence. Further discussion of the slip correction factor for small aerosol particles can be found in Willeke and Baron [8] or in Hinds [9].

From the value of the Stokes number, it is possible to qualitatively describe the motion of aerosol particles through the airways. For Stk << 1, the drag force largely suppresses any particle inertia, and particle trajectories follow air streamlines very closely. However, as Stk grows larger, the relative influence of inertia on particle motion grows stronger, and particle trajectories increasingly stray from curved streamlines. Accordingly, numerous experimental and numerical studies indicate that the probability of inertial deposition of an inhaled particle on airway walls increases monotonically with the Stokes number [3, 10–17]. A summary and comparison of various formulas available in the literature to relate the probability of inertial deposition in the respiratory tract to the particle Stokes number can be found in Finlay [5]. In the interest of brevity, it is enough at present to note from Equation (2) that the Stokes number, and therefore the likelihood of inertial deposition, increases with the average air velocity, the particle density, and the square of the particle diameter. The latter, squared dependence indicates the strong effect of particle size on inertial deposition. The fact that the probability of inertial impaction also increases with air velocity leads to the conclusion that impaction occurs primarily in the mouth, throat, and upper generations of the tracheobronchial region, where flow rates are highest. Particles with aerodynamic diameters of approximately 10 μm and larger are predominantly filtered by inertial impaction in the extrathoracic region, and thus do not normally reach the lungs for typical inhalation flow rates. The bifurcating structure of the respiratory tract ensures that air velocities deep in the lungs are orders of magnitude smaller than those at more proximal levels, thereby very much reducing the probability that inertial impaction will occur into the alveolar region. However, for a given air velocity, the Stokes number increases with decreasing airway diameter. Such a relationship highlights the dependence of inertial deposition on an individual patient's respiratory tract geometry, which is responsible in part (along with variations in breathing maneuver) for large intersubject variability in extrathoracic and tracheobronchial deposition patterns [7].

***Gravitational Sedimentation*** Aerosol deposition in the respiratory tract can also occur when the influence of gravity causes particles to settle onto airway walls. The likelihood that a particle will deposit due to gravitational sedimentation will increase with its settling velocity, which is the terminal velocity of the particle in still air. The settling velocity may be expressed as

$$v_{\mathrm{settling}} = \frac{C_C \rho_{\mathrm{particle}} d^2 g}{18\mu} \tag{3}$$

where $g$ is the acceleration due to gravity. Finlay [5] has demonstrated that by approximating the respiratory airways as randomly orientated cylindrical tubes, it can be determined that the probability of particle deposition in an airway by sedimentation increases monotonically with the nondimensional parameter $t'$, given as

$$t' = \frac{v_{\mathrm{settling}} L}{UD} \tag{4}$$

where $L$ is the airway length. This result is comparable to that of previous researchers, where a parameter similar to $t'$, but including the angle of airway orientation, was employed [18, 19].

Equation (4) can be thought of as a ratio comparing the average residence time of a particle in an airway ($L/U$) with the approximate time required for the particle to settle onto an airway wall ($D/v_{\mathrm{settling}}$). For $t' \ll 1$, the particle passes through the airway in much less time than would be required for it to deposit by gravitational sedimentation; however, as $t'$ increases, so does the probability that the particle will come to rest on a wall prior to leaving the airway. As was the case for inertial deposition, the probability of deposition due to sedimentation increases with the particle density, and the square of the particle diameter (both of which increase the settling velocity). Conversely, the likelihood of sedimentation onto airway walls decreases with the average air velocity, an opposite trend to that discussed above for inertial deposition. As a result, deposition by sedimentation primarily occurs deep in the lungs, where air velocities are sufficiently low to give particles time to settle onto airway walls. It is important to note here that while Equations (3) and (4) indicate that the probability of deposition by sedimentation is greater for larger particles, particles exceeding a certain size (aerodynamic diameter of about 5 μm) will likely be filtered out in the

extrathoracic region and the upper part of the tracheobronchial region by impaction prior to reaching the bronchioles and the alveolar region where sedimentation is favored. This regional filtering will be discussed in more detail later; however, it can now be stated that the combined actions of gravitational sedimentation and particle inertia are responsible for the increase in total respiratory tract deposition with increasing aerodynamic particle diameter that occurs to the right of the minimum value as displayed in Figure 1.

***Brownian Diffusion*** The motion of very small aerosol particles (with diameters much less than 1 μm) is influenced by random collisions with air molecules. The trajectories of such particles change direction sharply at the occurrence of a collision and as a result are nondeterministic. This type of random-walk particle path is known as Brownian motion. Over time, Brownian motion can cause very small aerosol particles to diffuse onto airway walls, as is illustrated in Figure 3.

While the exact paths followed by particles subject to Brownian motion cannot be predicted, it is possible to estimate the net distance over a given time that a particle will be displaced due to molecular collisions. As such, the average deposition probability of particles in a cylindrical tube can be determined [20–23] and has been shown to increase monotonically with the parameter

$$\Delta = \frac{kTC_C}{3\pi\mu d_{\text{particle}}}\frac{L}{UD} \tag{5}$$

where $k = 1.38 \times 10^{-23}$ J/K is Boltzmann's constant, and $T$ is the air temperature expressed in kelvins. As was the case for gravitational sedimentation, the probability of deposition due to diffusion increases with decreasing air velocity; therefore, diffusional deposition might be expected to occur deep in the lungs, in the alveolar region. This is a reasonable assumption for particles no smaller than 100 nm; however, notice that the probability of particle deposition by diffusion increases with decreasing particle size, whereas the probabilities of deposition by impaction and sedimentation both increased with the square of particle diameter. Referring back to the total deposition curve displayed in Figure 1, deposition that occurs to the left of the minimum value results from Brownian diffusion. In the extreme case for particles less than 10 nm in diameter, total respiratory tract deposition can reach values over 90%, and as these particles are rapidly transported to airway walls by random collisions with air molecules, significant fractions of the particles deposit in the extrathoracic and tracheobronchial regions.

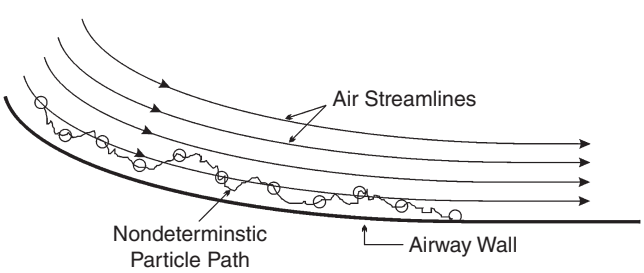


**Figure 3** Illustration of particle deposition by Brownian diffusion.

In aerosol drug delivery, diffusional deposition normally plays a minimal role. Typical pharmaceutical aerosols consist of particles a few microns in diameter, for which diffusive motion is largely negligible. Even when these aerosols do contain significant portions of particles with submicron diameters, for the particle size distributions produced by conventional inhalers the majority of the drug dose (by mass) is contained in larger, micron-scale particles.

***Electric Charge*** Aerosol particles produced by conventional delivery devices can contain significant amounts of electric charge [24–28]. As many devices are made from plastics that are typically susceptible to carrying surface charge, electrostatic effects can lead to aerosol deposition in the devices, reducing their output efficiencies [29]. Care must also be taken in performing bench tests of delivery devices so that any electrostatic effects occurring in plastic airway replicas are accounted for [30].

In the respiratory tract itself, analyses by Yu [31] and by Finlay [5] predict that the charge on aerosol particles must reach quite high levels before electrostatics plays a significant role in determining deposition. These analyses appear to be in good agreement with experimental measurements [32], though the possibility that high humidity in the respiratory tract mitigates electrostatic effects even further remains to be fully evaluated [5]. Many aerosols produced for drug delivery contain charges well below the levels predicted by Yu [31] and Finlay [5], so that electrostatic effects are not expected to influence deposition in the lungs for these aerosols. However, certain pharmaceutical aerosols [25] do contain high enough charge that deposition might be affected; therefore, electrostatic effects should be considered on a case-by-case basis.

## 2.2 Regional Deposition

***Mouth–Throat Deposition*** Whether the intention is to target deposition of an inhaled therapeutic aerosol to the tracheobronchial airways for treatment of lung

disease or to the alveolar region for systemic delivery to the bloodstream, deposition in the mouth and throat is rarely desirable. Deposition in the mouth–throat region (also referred to as the upper airways, or the extrathoracic or oropharyngeal region) not only reduces the dose delivered to the lungs but can also lead to undesirable side effects, both those that occur locally in the mouth and throat and systemic effects that can occur subsequent to swallowing. While the filtering efficiency of the oral cavity (the mouth) is lower than that of the nasal cavity, the combined geometry of the mouth, the oropharynx, and the larynx still presents a formidable barrier to lung delivery. Airway constrictions in the oropharynx and the larynx cause local acceleration of inhaled airflow and lead to inertial deposition of aerosol particles. In addition, deposition in the mouth is often augmented by drug delivery devices, particularly when aerosol exits the device mouthpiece at high speed. Mouth–throat deposition for commercial inhalers can often reach values higher than 75% of the dose emitted from the device [33–37].

Empirical correlations for predicting average mouth–throat deposition for given particle sizes and breathing conditions are available from multicenter in vivo studies of tidally inhaled aerosols [7, 38]. As is seen in Figure 4, intersubject variability around average deposition curves is very high [7], in large part owing to differences between individuals in mouth–throat geometry [39, 40]. Grgic et al. [40] have presented a correlation that allows for prediction of mouth–throat deposition in a given individual provided that the volume and path length of the mouth–throat airway are known. These correlations [7, 38, 40] are based on experimental data obtained with aerosols supplied to the mouth through straight, relatively large-diameter (~15–25 mm) tubes. In contrast, some delivery devices have smaller-diameter mouthpieces, which can increase deposition in the oral cavity [41–44]. DeHaan and Finlay [45] have presented a correlation that may be used to predict mouth–throat deposition for such devices where the mouthpiece diameter affects deposition.

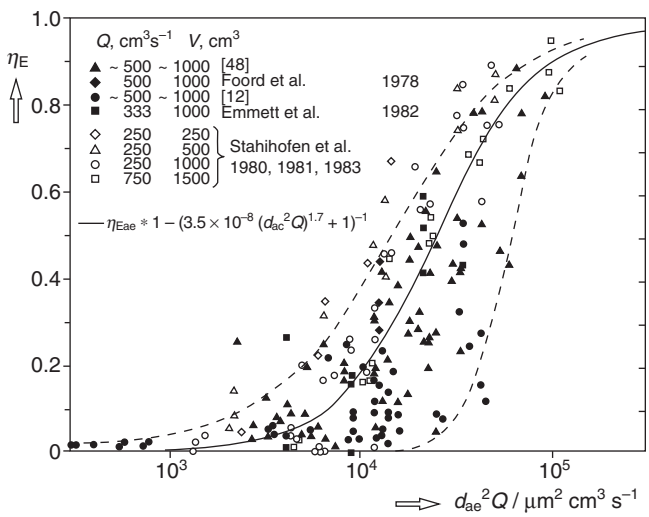


**Figure 4** Fraction of inhaled particles depositing in the mouth–throat (extrathoracic) region as a function of $d_{ae}Q$, where $d_{ae}$ is the aerodynamic particle diameter and $Q$ is the inhalation flow rate, for human subjects during mouth breathing. The solid line represent an empirical fit to the average of all data points, while the dashed lines represent the approximate range of the data. (From Stahlhofen et al. [7].)

***Lung Deposition*** Deposition in the lungs can be broadly divided into tracheobronchial and alveolar fractions. The tracheobronchial airways are covered with a surface layer of mucus, which sits atop of cilia attached to the airway walls. The cilia beat so as to set the mucus layer in continuous motion toward the throat. Thus, inhaled particles that deposit in the tracheobronchial region are cleared relatively quickly (within approximately 24 h) to the throat, where they are swallowed or expectorated. For this reason, in determining regional deposition patterns in human subjects, researchers typically associate the fast clearing deposition fraction of an inhaled aerosol (i.e., the fraction of total deposition that is cleared within 24 h) with the tracheobronchial deposition fraction. It is important to note, however, that the two are not precisely synonymous, as a small portion of tracheobronchial deposition is not cleared from the region within 24 h, particularly for particles a few microns or less in diameter [3].

The mucociliary clearance mechanism does not extend to the alveolar region. Slow-dissolving aerosol particles that deposit in the alveolar region will persist for much longer than 24 h, and may be subject to a variety of fates, but they are generally thought to be engulfed by alveolar macrophages and phagocytized and absorbed, or translocated to the mucociliary escalator, or the lymphatics [3, 46, 47]. Again, while experimentally determined slowly cleared fractions of inhaled aerosols may be associated with alveolar deposition, a small percentage of particles depositing in the tracheo bronchial airways will remain in the lungs for longer than 24 h.

Figure 5 displays average fast and slow clearing deposition efficiencies of radio-labeled aerosol particles measured in human subjects during oral, tidal breathing as summarized by Stahlhofen et al. [7]. These data were collected with a straight tube inserted into the mouth, so that the mouthpiece effects that can occur for some delivery devices, as discussed in the previous section, are not included. A broad maximum in deposition efficiency occurs for both

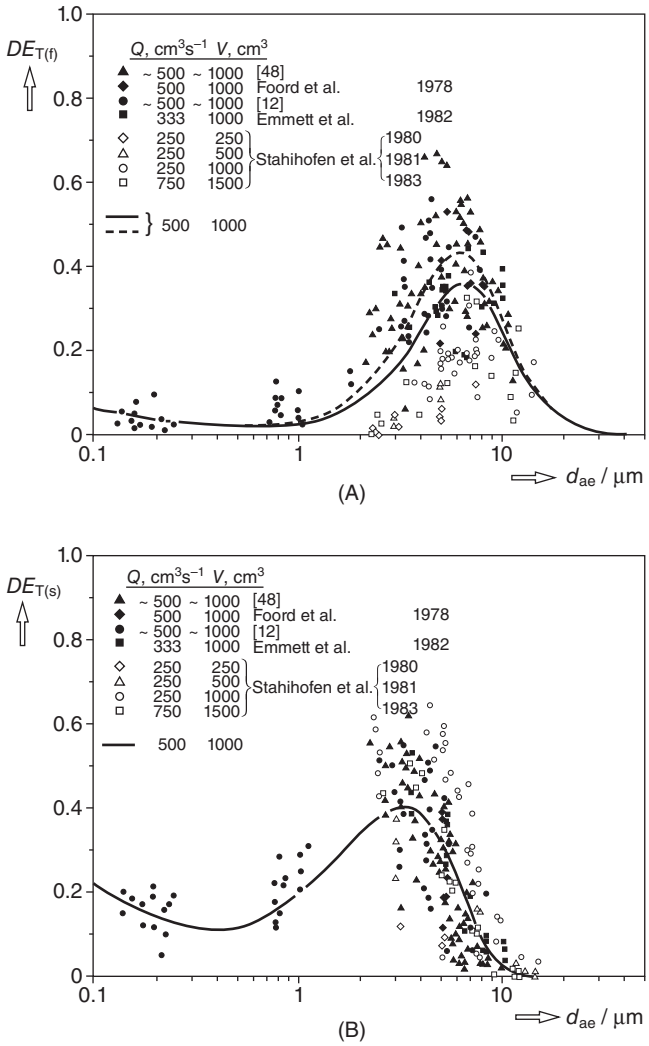


**Figure 5** Deposition efficiencies measured in human subjects during tidal mouth breathing for (A) particles cleared from the lungs within 24 h from inhalation, and (B) particles remaining in the lungs after 24 h from inhalation. The solid curves represent empirical fits to the average of the data, while the dashed curve in (A) is an average of the data from Lippman [48] and Chan and Lippmann [12]. (From Stahlhofen et al. [7].)

regions of the lung, with maximum tracheobronchial deposition at approximately 6 µm particle diameter and maximum alveolar deposition at approximately 3 µm particle diameter. As alluded to earlier, tracheobronchial deposition decreases with particle size above 6 µm or so as more and more particles are filtered out by the mouth and throat; likewise, alveolar deposition decreases above 3 µm due to regional filtering by the bronchial airways. While Figure 5 displays general regional deposition patterns averaged over a variety of subjects and tidal breathing conditions, these patterns may shift when specific breathing maneuvers are employed, as can occur with some aerosol delivery techniques. For example, researchers have attempted to improve targeting of pharmaceutical aerosols to specific lung regions through insertion of a short burst (bolus) of aerosol into the inhaled airflow at a prescribed point in the breathing cycle [49–51]. In addition, the data presented in Figure 5 were collected with stable aerosols that, unlike some pharmaceutical aerosols [52–54], do not undergo hygroscopic size changes during transit through the humid respiratory tract. Accordingly, it is best to view the curves presented in Figure 5 only as rough guidelines for estimating regional deposition patterns for inhaled pharmaceutical aerosols.

## 2.3 Deposition Modeling

One of the main benefits provided through an understanding of the mechanics of inhaled aerosols is the ability to predict the behavior and deposition pattern of a given aerosol in the respiratory tract. Deposition models vary greatly in their complexity and in the refinement of their predictions. While many bench tests are performed simply to estimate the fraction of aerosol filtered by the mouth and throat compared to that reaching the lungs, sophisticated computational simulations at times go so far as to determine localized deposition patterns within a single airway bifurcation. Clearly, the choice of type of model to employ depends on the requirements of a particular research or development program. Nevertheless, even in their simplest forms, deposition models can prove extremely useful both in guiding the early, preclinical stages of a project and afterward in evaluating contending explanations of in vivo results.

***Empirical and Semiempirical Models*** Perhaps the simplest models for predicting respiratory tract deposition are purely empirical models. Such models provide a set of algebraic equations fit to a given collection of in vivo experimental data and allow for estimation of regional deposition patterns based on parameters that were varied in vivo, such as particle size and breathing pattern. Empirical models are inherently limited by their inapplicability outside of the parameter space of the experimental data upon which they were based. As the dynamical equations of particle and fluid motion are not explicitly considered in formulating predictive algebraic equations, those latter equations do not lend themselves well to inclusion of dynamical effects such as hygroscopic particle size changes. That said, provided that they are not used outside of their range of applicability, empirical models can facilitate quick, accurate estimation of respiratory tract deposition. Examples of empirical models presented in the archival literature

include those of Davies [55], Rudolf et al. [38, 56], Yu et al. [57], and the ICRP [3].

A level up in complexity from purely empirical models are the so-called dynamical or semiempirical models. Such models predict deposition in a given lung generation through consideration of the underlying physical mechanisms of sedimentation, diffusion, and inertial impaction but may still rely on empirical correlations to estimate the probability of deposition due to a given mechanism (most often, deposition probabilities for sedimentation and diffusion are estimated analytically, whereas the deposition probability for inertial impaction is calculated from empirical correlations). In addition, empirical correlations are employed for predicting mouth–throat deposition. Methods of combining deposition probabilities for individual mechanisms into a combined deposition probability for a given lung generation have been investigated by several authors [3, 58–61].

Dynamical models allow for examination of aerosol deposition in conditions for which experimental data may be sparse. Several authors have incorporated hygroscopic particle size changes into dynamical deposition models [15, 62, 63]. Koblinger and Hofmann [19] have employed stochastic techniques to model asymmetric airway branching. Finlay et al. [64] have combined a dynamical deposition model with a mathematical model of the mucous thickness throughout the tracheobronchial region in order to predict the mucosal drug concentration in each generation. This model has been useful in evaluating the potential efficacy of inhaled antimicrobial agents [65–67], for which local drug concentrations upon delivery may be compared to the minimum inhibitory concentrations of bacteria causing respiratory infection. In addition, regional deposition patterns in infants and children have been modeled by various researchers [3, 64, 68–71] with both dynamical and empirical models, where the idealized lung geometry has been scaled according to age. Detailed review and classification of numerical models for predicting regional respiratory tract deposition for pharmaceutical aerosols may be found in Swift [72] and in Finlay [5, 73].

***Computational Fluid Dynamics*** Use of computational fluid dynamics (CFD) routines has become widespread in recent years over a large variety of industries. Partial differential equations governing fluid motion are discretized over a physical domain of interest and solved iteratively for a given set of boundary conditions to yield the fluid velocity at each point in the domain. Particles may then be numerically introduced and tracked through the domain through consideration of the dynamical particle equation of motion. Given the complexity of the algorithms contained in commercially available CFD software packages, errors in modeling can easily occur if the choice of simulation parameters is not made carefully. Furthermore, user-specified boundary conditions can dramatically affect the computational output. For these reasons, results garnered from CFD simulations should always be validated against experimental data to as best an extent as is possible.

In the context of inhaled aerosols, despite its potential pitfalls, CFD has proven useful in examining aerosol transport and predicting localized deposition to a very fine resolution in geometries such as the mouth–throat region [74–77] and airway bifurcations [78–82]. CFD has also been used in evaluating the design of specific pharmaceutical aerosol delivery devices [83–85]. In product development programs, this latter application of CFD can markedly reduce the time and cost required for evaluation of several prototypical designs.

***Bench Tests*** In many instances, aerosol drug delivery systems involve physical phenomena that are not understood or not characterized to the extent that they can be described by mathematical equations in dynamical or CFD models. Examples include the deagglomeration of powders; aerodynamic breakup and two-way coupled evaporation of droplets in ballistic sprays; triboelectric charging and resultant electrostatic effects; and homogeneous or nucleated condensation of supersaturated vapors. In these circumstances, researchers must turn to bench tests to evaluate the performance of the devices. Such tests also play a role in the regulatory process and are used in batch testing for quality control of aerosol products.

One of the most intuitive, and most useful, tests that can be performed *in vitro* on the bench is to collect the aerosol mass delivered to a filter placed distal to a replica of the extrathoracic airways. In such a manner, the fraction of the dose delivered to the lungs may be estimated. Figure 6, from Dolovich and Rhem [86], displays seven different "throats" that have been used for this purpose. Of these, the USP-1, USP-2, and twin impinger inlet have been the most popular [87]. Unfortunately, these inlets do not accurately represent the complex geometry of the mouth and throat, and are known to underpredict mouth-throat deposition [88, 89]. Replicas based on casts or images of individual mouth-throat geometries are more realistic, but their lack of generality, especially considering the large intersubject variability in mouth-throat deposition, is problematic. For these reasons, it is desirable to have an idealized mouth-throat model that matches average *in vivo* deposition data. For adult subjects, one such idealized geometry is the Alberta geometry, as displayed in Figure 7, which has been shown to match average *in vivo*

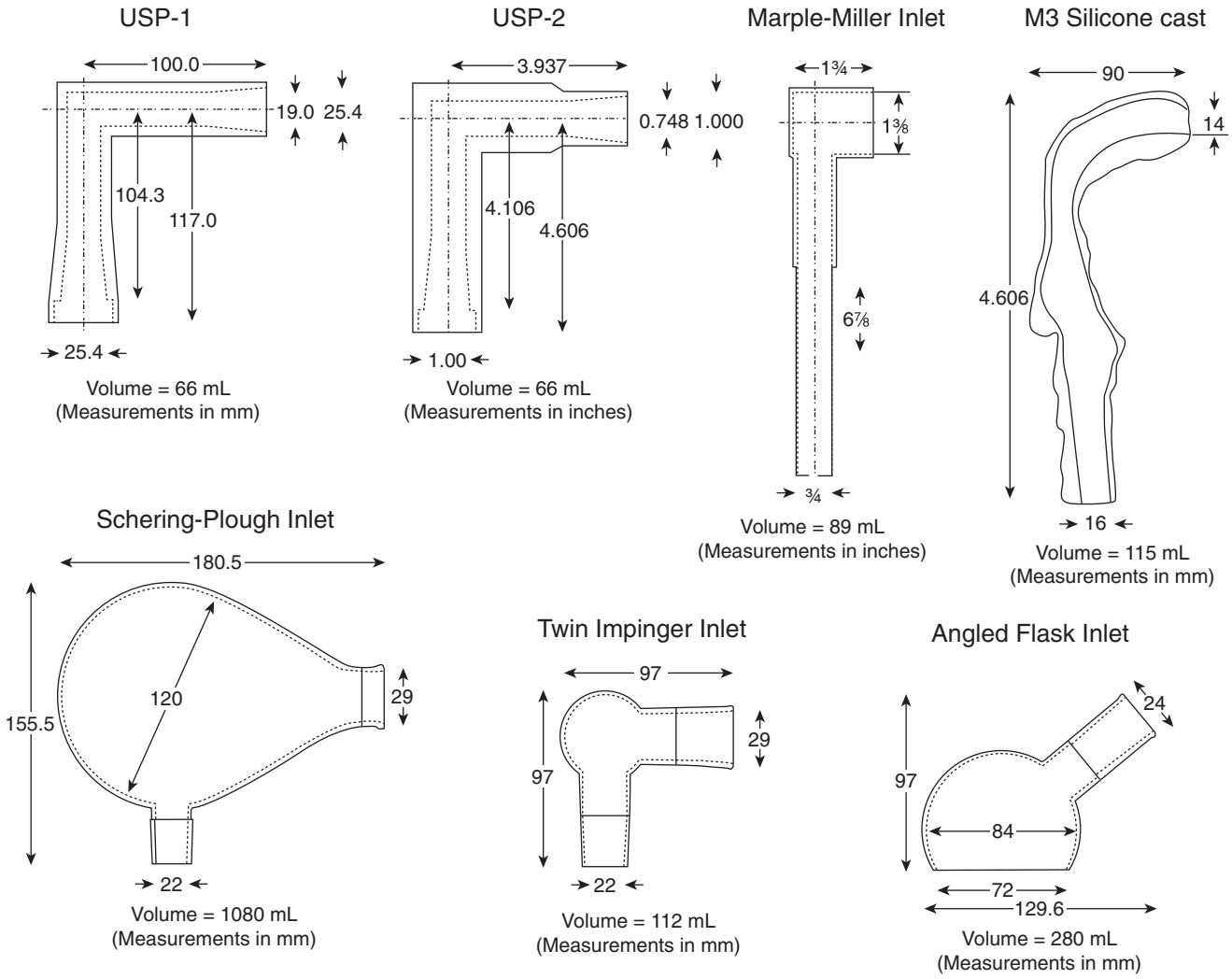


**Figure 6** Sagittal views of seven different throat models. (From Dolovich and Rhem [86].)

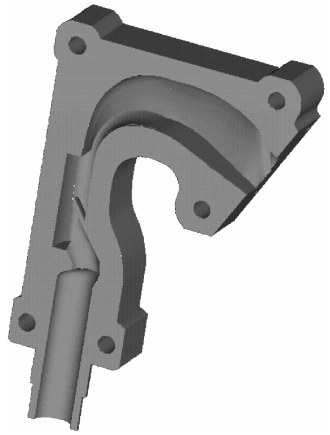

**Figure 7** Sagittal view of the Alberta mouth–throat geometry. (From DeHaan and Finlay [44].)

data for inhalation flow rates typical to many aerosol delivery devices [40, 44, 90].

A second approach to bench testing of aerosol delivery devices is to determine the distribution of particle sizes exiting the device mouthpiece and then to predict regional deposition patterns using either empirical or dynamical models such as those described above or using correlations developed for specific classes of device. In the latter approach, deposition is usually correlated with the fraction of the aerosol mass contained in particles below some cut-off diameter. This fraction is termed the respirable fraction or the fine-particle fraction. Clark and Borgstrom [87] review some of the correlations that have been developed for commonly used delivery devices. Given that these predictions are dependent on accurate determination of the particle size distribution delivered from the device, it is prudent to review several aerosol sizing instruments that have

found application in the pharmaceutical industry. This is the focus of the next section.

## 3 PARTICLE SIZING

The size of an inhaled particle is key in determining its site of deposition within the respiratory tract. Accordingly, much effort has gone into developing experimental instruments and procedures suitable for accurate determination of aerosol particle sizes. Unfortunately, the variety of properties exhibited by different pharmaceutical aerosols prohibits the application of any one, standard approach to particle sizing across all types of delivery devices. While the aerosols produced by some devices are stable in size, others are susceptible to hygroscopic growth or shrinkage. Some aerosol particles exit their delivery devices at ballistic speeds, while others are well entrained in inspiratory airflow. Liquid aerosol droplets are spherical due to surface tensions, while solid particles are in general not spherical. All these differences can affect particle size measurement; therefore, in choosing a sizing technique for a particular application, it is important to have a general understanding of the advantages and limitations of the different types of sizing instrument that are available.

### 3.1 Particle Size Distributions

The vast majority of aerosol drug delivery systems produce aerosols that are polydisperse, meaning that particle diameters vary over a fairly wide size range. Such distributions of size are best described by statistical means, typically under the assumption that the particles are spherical. As mentioned above, the assumption of sphericity is met for liquid droplets but not in general for solid particles. However, as it is the aerodynamic properties of pharmaceutical aerosol particles that govern their inertial and gravitational deposition in the respiratory tract, any compact particle (i.e., any particle upon which the aerodynamic drag force is close to that on a sphere) can be reasonably approximated as spherical. Consequently, the aerosols produced by many conventional delivery systems may be approximated as containing spherical particles.

In order to express particle size data statistically, a suitable distribution function must be found. For most single-source aerosols, the lognormal distribution function provides a reasonable fit to experimental data points [9]. Size distributions of pharmaceutical aerosols are conventionally expressed in terms of mass, as it is the deposited mass that constitutes the delivered drug dose. The normalized, lognormal mass distribution may be expressed as

$$m_{\text{normalized}}(x) = \frac{1}{x\sqrt{2\pi}\ln\sigma_g}\exp\left[\frac{-(\ln x - \ln \text{MMD})^2}{2(\ln\sigma_g)^2}\right] \tag{6}$$

where $x$ is the particle diameter, MMD is the mass median diameter (half of the total aerosol mass is contained in particles with diameter below the MMD), and $\sigma_g$ is the geometric standard deviation. Here, $m_{\text{normalized}}(x)\,dx$ gives the fraction of the total mass contained in particles with diameters between $x$ and $dx$. Of the total mass 68% is contained in particles with diameters between MMD/$\sigma_g$ and MMD*$\sigma_g$. A typical normalized lognormal mass distribution is displayed in Figure 8.

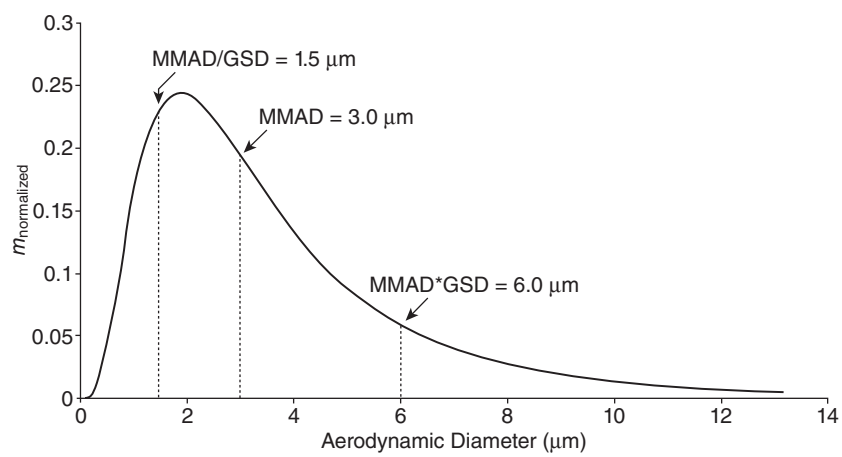


**Figure 8** Normalized lognormal mass distribution, with MMAD = 3 μm and GSD = 2. Of the total aerosol mass 68% is contained in particles with diameters between MMAD/GSD and MMAD*GSD.

In many cases, particle size distributions are described in terms of aerodynamic particle diameters, as related to geometric diameters by Equation (1). Where this occurs, the MMD appearing in Equation (6) is replaced by the mass median aerodynamic diameter (MMAD).

The lognormal distribution provides a convenient summary of particle size data: The size distribution is completely described by its MMD and $\sigma_g$. However, the technique through which the distribution function is fit to experimental data can influence how well the actual particle sizes are approximated [91]. Furthermore, some pharmaceutical aerosols, such as those containing aggregate clusters of particles mixed with deaggregated particles, are poorly described by lognormal distributions. In practice, when lognormal distributions are used to describe particle size distributions, measures should be taken to confirm that the distributions fit experimental data reasonably closely.

### 3.2 Inertial Impactors

Given that respiratory tract deposition of an inhaled pharmaceutical aerosol is determined in large part by its aerodynamic properties, it makes sense that the particle size distribution of the aerosol be determined from those same aerodynamic properties. This is indeed the case for a very popular type of size classification, known as inertial impaction. At its simplest, an inertial impactor can consist of a single nozzle and plate, as in Figure 9. The single nozzle/plate combination, or "stage," serves to separate particles above and below a cut-off aerodynamic diameter, the particle diameter for which the probability of collection on the plate is 50%. Particles with aerodynamic diameters larger than the cut-off carry sufficient inertia through the nozzle that their trajectories deviate from the flow streamlines to the point that the particles impact and collect on the plate. In contrast, particles smaller than the cut-off size are steered around the plate by the carrier flow and may be collected on a downstream filter.

In practice, inertial impactors commonly consist of a number of stages aligned in series and are referred to as multistage or cascade impactors. The cut-off diameters of the stages become progressively smaller moving through the device (from the inlet to the outlet), so that the particles collected on any one plate have a size range bounded at the upper end by the cut-off diameter of the previous stage and at the lower end by the cut-off diameter of the present stage. Such a design allows for a polydisperse aerosol entering the impactor to be sorted into a number of relatively narrow size bins. The cut-off diameters of the stages are typically supplied for one or more flow rates by the manufacturer, and for commonly used impactors calibrations for other flow rates are sometimes available in the archival literature. In cases where collected particles might possibly bounce off the impactor plates or be re-entrained from the plates back into the carrier flow (both of which can occur for powder aerosols), the plates should be coated with grease or otherwise wetted [92] in order to mitigate such effects.

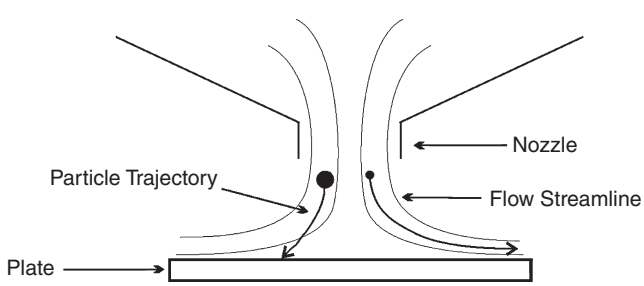


**Figure 9** Schematic diagram of a single impactor stage.

In addition to direct sizing of aerosols by aerodynamic particle diameters, an advantage of inertial impactors is that the drug mass deposited on each plate can be determined directly by chemical assay. This is especially important for multicomponent aerosols, such as powders containing inactive carrier particles, or suspensions of drug particles in liquid droplets, where not all particles contain equal proportions of drug [93]. Furthermore, direct measurement of the aerosol mass distribution is advantageous since for most delivery systems the mass contained within a particular size range is closely correlated to the delivered dose.

Two problems that sometimes arise in interpreting size distributions measured by intertial impaction stem from the fact that the measurements are not made in situ, but rather an airstream is diverted through the impactor. Therefore, the size distribution measured by the impactor is not necessarily the same as the size distribution, for example, exiting the mouthpiece of a delivery device. First, a significant fraction of the aerosol can be lost in the inlet port used to draw air into the impactor. For this reason, in testing delivery systems, some of the "throats" depicted earlier in Figure 6 are often used as inlets to inertial impactors, with the interpretation that the losses in the inlet are then representative of losses that might be expected in a patient's mouth and throat. Unfortunately, as discussed in Section 2.3, many of these inlets poorly model average mouth–throat deposition in humans, so that this interpretation is generally not accurate. This issue has been discussed in detail by Mitchell and Nagel [94].

A second problem can arise when the size distributions of unstable aerosols, such as evaporating droplets, are determined by inertial impaction. As the size distributions of these aerosols vary with time, they can be

expected to change in transit to and through the impactor, especially if the impactor is not in thermal equilibrium with the carrier flow. This issue is discussed further in Section 3.4.

Detailed review of the use of inertial impactors in sizing of pharmaceutical aerosols, including descriptions and cut-off sizes of some of the more popular commercially available cascade impactors, can be found in Marple et al. [95], and in Mitchell and Nagel [94].

### 3.3 Laser Methods

Though inertial impaction is the most widely adopted technique for measuring the size distributions of pharmaceutical aerosols, it is not the only technique available. In fact, a large and diverse range of instruments is available for determining the sizes of aerosol particles. Many instruments have been designed to measure aerosols produced during various natural and anthropogenic processes for which particle sizes are outside the range typical for pharmaceutical applications (broadly, 1–10 μm particle diameters); therefore, these will not be discussed at present. Instead, three light-scattering techniques, each of which has found application in the evaluation of aerosol drug delivery systems, will now be described.

***Time-of-Flight Instruments*** Time-of-flight instruments measure the transit time between two laser beams, separated by a known distance, of aerosol particles exiting a region of accelerating flow. The elapsed time between intersection of the particle with the first and second laser beam is correlated with the aerodynamic particle diameter. As with all single-particle counters, the maximum number density of particles in a volume of the carrier flow is limited by coincidence of two or more particles in the measurement space at the same time. For time-of-flight instruments, aerosol number densities produced by common delivery devices can exceed this maximum [96], so that dilution of the aerosol is required. Add-on dilution instruments are available for precisely this purpose; however, particle losses in the diluter must be accounted for or confirmed as negligible.

A second difficulty that arises with time-of-flight instruments is the presence of artificial background, or "phantom," particles at every particle size [97]. Because mass distributions must be calculated from measured number distributions, a very small number of large phantom particles (which contain a much larger fraction of the total aerosol mass) can skew mass distributions toward larger particle sizes. This can be accounted for in postprocessing calculations, which may be included in software packages accompanying commercial instruments. As for any case in which the mass distribution is a calculated quantity, attention should be paid to the steps taken in converting the raw, measured data to the final, reported distribution.

Finally, unlike chemical assay of inertial impactor plates, but like all light-scattering techniques, time-of-flight instruments do not in general discern between active drugs and inactive excipients. For example, where drug particles are suspended in liquid droplets, no distinction can be made between droplets of the same diameter that contain one drug particle, multiple particles, or no particles [98], making accurate determination of the fraction of drug mass contained in a given droplet size range difficult.

Due to the limitations outlined above, time-of-flight instruments are less attractive for use in determining particle sizes produced by aerosol drug delivery systems than are the other instruments described in the present chapter. However, in applications where all that is required is a quick, comparative measure of the particle size distribution from one device to another, such as in batch testing, time-of-flight instruments may be an attractive option [99].

***Phase Doppler Particle Analyzers*** Phase Doppler particle analyzers (PDPAs) measure the spatial frequency of Doppler-shifted light scattered by an aerosol particle that passes through a measurement volume defined by the intersection of two laser beams. A typical PDPA system setup is displayed schematically in Figure 10. The phase difference between frequency signals measured by two photodetectors spaced slightly apart is linearly related to the spherical particle diameter [100]. A third photodetector may be employed so as to overcome a trade-off between system resolution and range and to check for particle sphericity [101]. As particles must be spherical for accurate sizing to occur, PDPAs are most often employed to size liquid droplets. Unlike time-of-flight instruments, the measurement volume of a PDPA is sufficiently small that coincidence does not tend to interfere with accurate sizing of aerosols for the particle number densities typically produced by inhalation drug delivery systems.

A major advantage of PDPAs is that size measurements can be made in situ, without re-directing the flow of aerosol from a device. In situ measurement is especially useful in sizing volatile aerosols, where particle (droplet) sizes change with time [102, 103]. However, because the PDPA measures particle sizes at a precise location in the flow, determined by the intersection of the two laser beams, it is important either to perform measurements at several locations across the flow or to choose one location for which measurements are representative of the overall aerosol size distribution. In

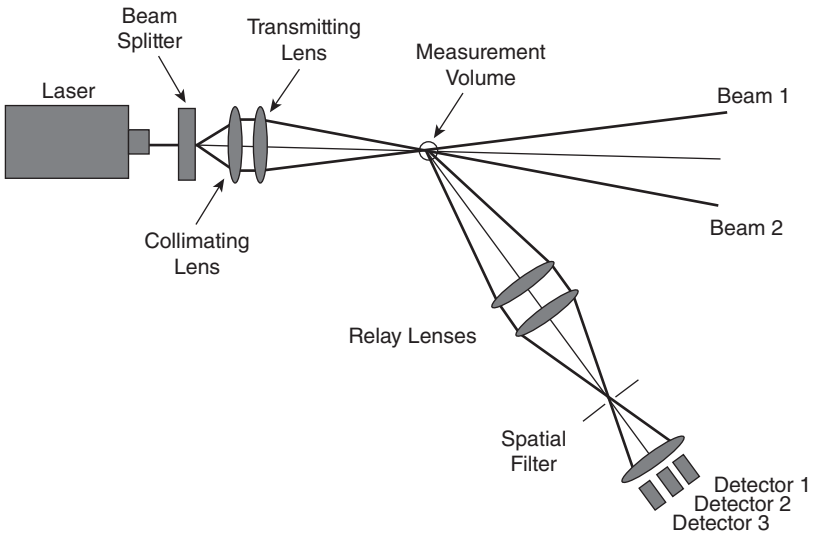


**Figure 10** Diagram of a phase Doppler particle analyzer (PDPA) system. (From Rader and O'Hearn [100].)

addition, as was discussed above for time-of-flight instruments, light-scattering techniques do not in general differentiate between drug (or drug-containing) particles and inactive, excipient particles; therefore, PDPA data must be carefully interpreted in cases where the proportion of drug in aerosol particles is not constant, such as when a drug is suspended in liquid droplets [93].

***Laser Diffraction Instruments*** Unlike the light-scattering techniques described in the previous two sections, where particles are sized one by one as they pass through a small measurement volume, laser diffraction instruments determine an ensemble size distribution of particles passing through a larger measurement region, produced by an expanded laser beam. Particles in the measurement region diffract light at different angles according to their size, with smaller diameter particles diffracting at higher angles. The cumulative effect of diffraction from a large number of particles is the formation of a Fraunhofer diffraction pattern, which when focused through a Fourier lens is visible as a set of concentric circles [104]. A laser diffraction instrument typically measures diffraction patterns using an array of photodetectors and, through a series of fairly complex computations [100], converts these to size distributions.

The algorithm employed by a particular diffraction instrument to calculate size distributions from measured diffraction patterns can strongly affect its accuracy. Some earlier instruments, based solely on Frauhofer diffraction theory, do not account for refraction of light through small liquid droplets or for scattering of light off more than one particle (which may occur when the aerosol number density is high). Corcoran et al. [104] have demonstrated that this type of instrument is inappropriate for measuring the size distributions of droplets produced by commercial nebulizers, whereas newer diffraction instruments, incorporating more complex algorithms, display much greater accuracy.

Laser diffraction instruments quickly and reproducibly determine particle size distributions and may be employed in situ to size volatile droplets as they exit delivery devices. Consequently, laser diffraction has been used widely to measure the sizes of aqueous droplets produced for aerosol drug delivery [105, 106]. As discussed above for other light-scattering techniques, laser diffraction is better suited for sizing of aerosol droplets produced from solution drug formulations, where the proportion of drug in the droplets is constant, than those produced from suspension formulations.

Laser diffraction has also been used extensively to size solid particles, such as those found in pharmaceutical powder formulations. Strictly speaking, such measurements are not rigorously accurate, as the basic principle of laser diffraction is valid for incident laser light impinging on a spherical particle [107], while solid particles are not generally perfect spheres. Application of a spherical model to nonspherical particles will result in unpredictable measurement errors if the shape of the nonspherical particles is not well defined [107a]. For example, a small deviation in particle shape from that of a sphere to a slightly elongated ellipsoid has been shown to significantly affect apparent particle diameters measured by laser diffraction [108]. While these apparent diameters can be interpreted, given prior knowledge of the particle shape, to provide a reasonable indication

of the physical particle dimensions, in many practical applications particles are neither spherical nor of known analytically specified shape but are rather of some unknown, irregular shape. In these cases, laser diffraction instruments are best suited for quality control purposes, whereas inertial sizing techniques should be used to determine equivalent aerodynamic spherical diameters from which to predict deposition patterns in the respiratory tract.

### 3.4 Evaporation and Condensation

In many inhalation drug delivery systems, aerosol is produced through atomization of a solution, or suspension, of drug in an aqueous medium. In these cases, aerosol droplets may decrease in size as they evaporate into the surrounding air, particularly when the output flow from the device mixes with drier ambient air, as occurs during inhalation. If the mass fraction of droplets in a volume of air is small, droplet evaporation can progress until only the drug and any nonvolatile excipients present in the formulation remain, and the rate at which droplet sizes change with time can be predicted from simplified hygroscopic theory [5, 9]. However, when the mass fraction of droplets is relatively large, evaporation increases the water vapor concentration in the air, which in turn slows further evaporation. This type of "two-way coupled" effect stabilizes droplet sizes when the vapor concentration and temperature of the surrounding air reach equilibrium values. A similar effect can occur for growth of droplets (or solid particles) by condensation in humid air, where the vapor concentration in air decreases as droplets grow. Collectively, evaporation and condensation are referred to as hygroscopic size changes. Finlay [53] has presented a method for classifying the type of hygroscopic size change undergone by aerosol droplets, and has demonstrated that for some nebulizers used to produce aqueous aerosols for drug delivery, the mass fraction of droplets produced is sufficiently high that hygroscopic size changes are negligible. That is to say, droplet evaporation increases the vapor concentration in the air so quickly for these aerosols that an equilibrium is established before the droplets significantly decrease in size. In contrast, aqueous aerosols produced by other nebulizers have smaller mass fractions, so that significant hygroscopic size changes do occur before an equilibrium is reached, if one is reached at all [53].

When droplet sizes do change with time, a dilemma exists in developing experimental sizing procedures. Where should the aerosol size distribution be measured, and what does the distribution represent? One approach is to measure droplet sizes as close to the mouthpiece of the delivery system as is possible [103, 105]. The measured droplet size distribution is then easy to interpret as consisting of the initial sizes of droplets as they exit the device, though sizing can only be performed using instruments capable of in situ measurement. A second approach is to sample the aerosol at some downstream position, typically using a cascade impactor (e.g., [109]). Here, care must be taken to ensure that hygroscopic size changes do not occur between the delivery system and the impactor. Furthermore, if the impactor is not cooled to the temperature of the aerosol, heating of the aerosol inside the impactor can lead to droplet evaporation within the instrument itself [110]. If hygroscopic size changes do occur during transit to and through the sizing instrument, the measured size distribution will be different from that delivered to a patient at the mouthpiece of the delivery system.

Droplet size changes also occur for a second type of pharmaceutical aerosol, produced by atomization of a pressurized, volatile propellant. For these devices, known as pressurized metered-dose inhalers (pMDIs, or simply MDIs), droplet evaporation occurs at a much faster rate than for aqueous aerosols. As a result, size measurements for pMDIs are often made under the assumption that the propellant has completely evaporated, and only solid particles (consisting of drug and nonvolatile excipients) remain. It is important to realize that these residual particle sizes cannot be used to predict deposition in the mouth and throat. When a pMDI is actuated into the mouth, the aerosol consists of rapidly evaporating droplets traveling at ballistic speeds. It is these droplet sizes and speeds that primarily determine deposition in the mouth and throat region. Because droplet diameters and speeds rapidly decrease, due, respectively, to evaporation and entrainment in the inhalation airflow, they are difficult to measure [102], and add a great deal of complexity to predictive, mathematical deposition models. For this reason, mouth–throat deposition for pMDIs is most often estimated from bench experiments using physical mouth–throat replicas, while the sizes of residual particles may be used with mathematical models or correlations to predict regional deposition patterns within the lungs.

## 4 AEROSOL DELIVERY DEVICES

In recent years, many new aerosol drug delivery systems have been developed and introduced to the market. In part, this surge in development has been spurred by the desire to deliver new forms of drug to the lungs. Drugs that have been traditionally delivered through the inhalation route, such as bronchodilators and corticosteroids, have large therapeutic windows, so that high intersubject and intrasubject variability in dosing are

tolerated. However, for many drugs not traditionally delivered through inhalation, including inhaled insulin for treatment of diabetes and cytotoxic agents for treatment of lung cancers, dosing must be controlled much more tightly. Consequently, novel aerosol drug delivery platforms have been invented, with the goal of overcoming some of the known shortcomings of the long-established delivery devices that are currently in wide use. It is not anticipated that this next generation of devices will render current devices obsolete any time soon. More likely, as the market for inhaled drugs grows, the range of available delivery systems will grow alongside it.

The following sections will describe several different types of aerosol drug delivery devices (or inhalers) that are presently in use and in development. This is not an exhaustive list, as new technologies are continuously emerging; however, the vast majority of inhalers will fall into one of the categories outlined below. In addition, we have attempted to avoid referring to commercial companies and products but instead to discuss families of similar devices. References are provided wherein discussion of specific commercial devices may be found.

### 4.1 Pressurized Metered-Dose Inhalers

The pressurized metered-dose inhaler (pMDI) was introduced in the mid 1950s (see [111] for a history of the pMDI and other devices) and remains the most widely used aerosol delivery device for treatment of asthma. The popularity of the pMDI stems largely from its portability and low manufacturing cost. On the other hand, the dose delivered to a patient is highly dependent on his or her inhalation technique, and deposition in the extrathoracic airways is often very high. For these reasons, add-on devices such as spacers are often used with pMDIs to improve delivery.

Briefly, a pMDI formulation consists of a suspension, or solution, of drug in a volatile liquid propellant, or mixture of propellants, held under pressure inside a valved canister. A co-solvent may be added to the formulation to enhance drug solubility. In addition, surfactants are typically added in low concentrations in order to stabilize suspensions and lubricate valves. The choice of formulation can have a significant effect on both the shelf life and performance of a pMDI. With the ratification of the 1989 Montreal Protocol, and the requirement to replace ozone-depleting chlorofluorocarbon (CFC) propellants in pMDIs, many conventional formulation techniques required reinventing with alternative propellants. Over time, hydrofluoroalkane (HFA) propellants have emerged as a viable replacement for CFCs. Review of current practices with respect to formulation of HFA pMDIs may be found in Vervaet and Byron [112] and in Smyth [113].

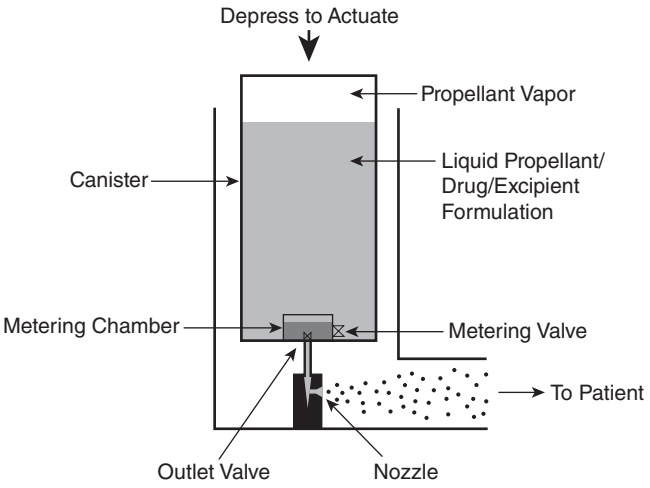


**Figure 11** Schematic diagram of a typical propellant metered-dose inhaler (pMDI).

A typical pMDI design is displayed schematically in Figure 11. The canister is placed inside a plastic actuator, which contains a spray orifice, or nozzle, and mouthpiece. Prior to actuation of the device, an outlet valve at the base of the canister is closed, while a metering valve inside the canister remains open, allowing the metering chamber to fill with a prescribed volume of the formulation (usually on the order of 50–100 μL). When the pMDI is actuated by depressing the canister, the outlet valve opens and the metering valve closes, causing the metered dose to exit rapidly through the nozzle. The flow of propellant through the nozzle generates a spray of volatile liquid droplets traveling at high velocities. The mechanics underlying this initial spray formation are complex but have been described in as much detail as possible by Dunbar et al. [114] and by Finlay [5]. Postnozzle evaporation quickly reduces droplet sizes, while aerodynamic drag rapidly decreases velocities. Ideally, these processes occur over a very short period of time, leaving behind a cloud of residual drug particles entrained in the patient's inspiratory airflow; however, in practice, the ballistic initial speeds of pMDI droplets lead to high deposition in the mouth and throat, thereby limiting lung deposition (e.g., see [115–118]).

A second issue that can reduce the efficiency of delivery from pMDIs is improper coordination between device actuation and patient inhalation. Difficulties with coordination are often reported anecdotally when pMDIs are used by children or seniors. Breath-actuated pMDIs, designed to release a metered dose at a specific point in the patient's inhalation maneuver, have been developed to overcome coordination problems [119, 120]. Another solution is to use add-on spacer or

holding-chamber devices, which fit between the pMDI mouthpiece and the patient. After the pMDI is actuated into a spacer, the patient may inhale the aerosol in one or more breaths. Spacers also serve the purpose of providing room (and therefore time) for pMDI droplets to evaporate and decelerate prior to reaching the patient, thereby reducing deposition in the mouth and throat [121, 122]. This reduction in mouth–throat deposition is often associated with deposition in the spacer itself; therefore, lung deposition does not necessarily increase when spacers are used with pMDIs. However, because deposition in plastic spacers often occurs due to electrostatic attraction of particles to spacer walls, coating spacers with ionic detergents (to remove static charge) has been predicted in vitro to increase lung deposition compared to that predicted for uncoated spacers [29]. In addition, nonelectrostatic metal and polymer spacers have recently become available [123, 124].

As with any aerosol drug delivery device, the efficacy of a pMDI can be evaluated in terms of a final clinical outcome. For example, the efficacy of an inhaled bronchodilator may be determined from the change in a patient's forced expiratory volume ($FEV_1$) from the baseline level [125]. Such clinical outcomes are, of course, paramount to the patient, but they provide little insight for researchers and developers into their underlying causes, as regional deposition patterns can only be inferred (i.e., more drug might be presumed to have deposited in the lungs when the clinical effect is greater). Regional deposition patterns can be measured in vivo by labeling the inhaled aerosol with radionucleotides and then imaging the respiratory tract after delivery [126, 127], or by measurement of pharmacokinetic parameters such as the absorption of drug into the bloodstream [125]. However, it is often desirable, especially in the early stages of development, to estimate regional deposition patterns without expending the time and costs associated with in vivo studies. In these circumstances, the performance of a device in vivo can be predicted using the various particle sizing techniques and mathematical deposition models describes earlier in this chapter.

As mentioned above in the context of particle sizing (Section 3.4), the sizes and speeds of pMDI droplets close to the spray nozzle are very difficult to measure. Unfortunately, it is these same sizes and speeds that to a large extent determine the division of deposition between the extrathoracic airways and the lungs. As such, one approach to estimating lung deposition from pMDIs is to actuate the device into a physical mouth–throat model and then to size the aerosol that passes through it ([87], and references therein). For the aerosol passing through the physical model, regional deposition may be predicted from the measured particle size distribution using a mathematical deposition model. The fraction of aerosol depositing in the physical model is then added to the mouth–throat deposition predicted mathematically, the assumption being that deposition in a patient's mouth and throat due to the initially ballistic speeds of pMDI sprays (which is not accounted for in the mathematical model) is accurately represented by the deposition in the physical model. Clark and Borgström [87] describe the success various researchers have had in predicting in vivo data using such an approach.

### 4.2 Dry Powder Inhalers

Another class of inhaler in wide use is the dry powder inhaler (DPI). These devices contain drug and excipients in the form of a fine powder, which is delivered to a patient during inhalation. DPIs are as portable as pMDIs, and in many cases have been developed as direct alternatives to pMDIs for treatment of asthma and COPD. The design of DPIs varies a great deal from device to device. Descriptions of a number of commercially available DPIs may be found in Dunbar et al. [128] and in Borgström et al. [129].

In general, the function of a DPI is to disperse a metered dose of powder into a patient's inspiratory airflow. This involves both fluidization of the powder from its bed and deaggregation of particle agglomerates. Normally, respirable size particles, with diameters on the order of a few microns, are manufactured by milling or spray drying. These particles readily adhere to solid surfaces, including device walls and neighboring particles, due to short-range intermolecular attractions (van der Waals forces). For respirable particle sizes, the aerodynamic forces associated with typical inhalation flow rates cannot overcome particle adhesion [5], so that powder dispersion is difficult to achieve. A solution to this problem is to add larger carrier particles, often lactose, to the powder formulation. The smaller drug particles adhere to the surface of the carriers to form agglomerate particles. Alternatively, small drug particles may be tumbled together to form a larger, agglomerate particle without the addition of excipient carriers. Agglomerate particles are typically formed with diameters ranging from 50 to 150 μm [129], so that the aerodynamic lift forces acting on them during inhalation are sufficient for fluidization [5]. A challenge for the formulator is to design powders for which the adhesion forces that hold agglomerate particles together during fluidization are sufficiently small that they can be overcome once the powder is entrained in the air, so as to return drug particles to respirable sizes. Several approaches can be taken to enhance deaggregation of airborne agglomerates, including generation of turbulence, and

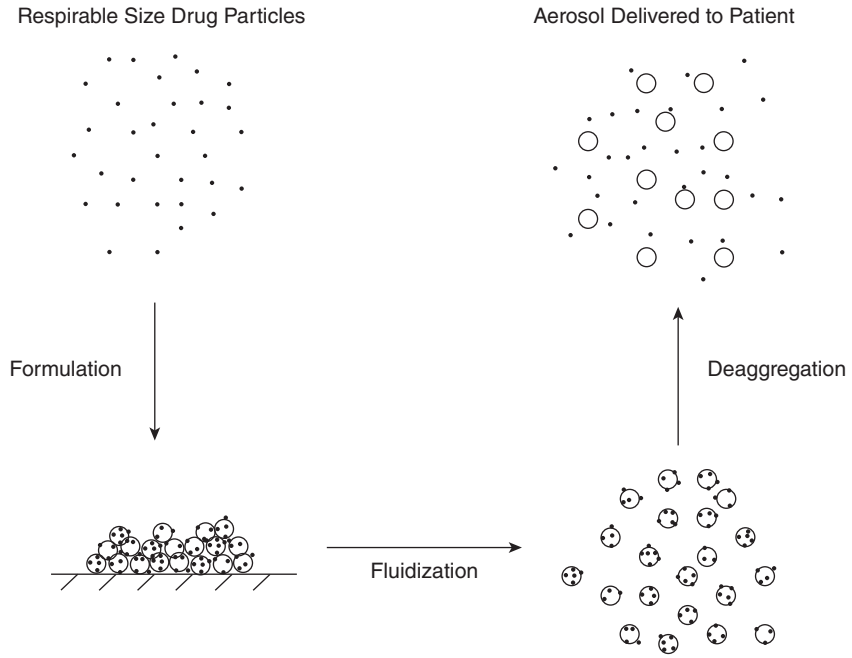


**Figure 12** Illustration of powder formulation, fluidization, and deaggregation for a dry powder inhaler (DPI).

impaction onto grid or mesh surfaces. Figure 12 provides a schematic diagram of the formulation, fluidization, and deaggregation of DPI powders.

As reviewed by Chan [130], a number of approaches have been taken to improve powder dispersion in DPIs. One of these is to fabricate powders consisting of large, porous particles, so as to enhance the aerodynamic lift and drag forces acting on particles while maintaining small enough particle mass to avoid inertial impaction in the upper airways [131]. Another approach is to include a spinning impeller in the DPI to assist deaggregation [132], while still another relies on compressed air. Protection of DPI formulations containing hydrophilic particles from ambient humidity will also aid powder dispersion, as capillary condensation can increase particle–particle and particle–wall adhesions for such formulations [5, 133].

Cascade impactors are commonly used to size DPI aerosols. Chemical assay of impactor plates allows for the mass distribution of drug to be distinguished from that of excipient, carrier particles. In many cases, a similar experimental approach is used for DPIs as for pMDIs, where the aerosol is delivered to the cascade impactor through a physical mouth–throat model. However, for DPIs, the emitted particle size distribution can be strongly dependent on inhalation flow rate, as this flow rate influences powder fluidization and deaggregation; therefore, ideally, the flow rate through the impactor should match the flow rate at which the dose is expected to be delivered to patients [134]. As the dose may be emitted at different flow rates due to intersubject and intrasubejct variability in inhalation maneuver, it is useful to evaluate the variability of a DPI over a range of flow rates. The pressure drop, or resistance to flow, across a DPI is also an important parameter. High-resistance inhalers limit the range of flow rates that can be achieved by patients, thereby reducing variability in dosing; however, some patients may have difficulty providing a sufficient inspiratory effort to overcome the resistance in the device.

In estimating mouth–throat and lung deposition from DPI size distribution data, it is necessary to account for jetlike effects that can occur for small-diameter mouthpieces [41–44]. Such effects enhance deposition in the oral cavity, leading to higher overall mouth–throat deposition than would be predicted using correlations developed for mouthpieces of larger diameter, for example, those by Stahlhofen et al. [7] and Rudolf et al. [38]. DeHaan and Finlay [45] presented a correlation that was used to predict mouth-throat deposition for a variety of DPIs by accounting for mouthpiece effects. Reasonably good agreement with in vivo data available in the archival literature was reported [45].

### 4.3 Aqueous Aerosols

Many compounds can be dissolved, or suspended colloidally, in water. One of the oldest and simplest methods

used to generate medicinal aerosols for inhalation is to atomize these solutions and suspensions into fine droplets. Indeed, numerous inhalation drug delivery systems currently in use create aerosols from aqueous formulations. While the jet nebulizer has found wide use for decades, many of new devices have only recently been introduced, many of which have been developed with the goal of overcoming limitations traditionally associated with nebulizers.

***Jet Nebulizers*** A jet nebulizer produces aerosol droplets by passing a flow of pressurized air over a liquid surface. Unlike pMDIs and DPIs, which deliver metered doses in a single inhalation, nebulizers are designed to operate during tidal breathing, over several minutes. In addition, jet nebulizers rely on an external source of pressurized air to drive aerosol production. As a result, jet nebulizers are normally less portable than pMDIs and DPIs. Even so, for many applications, and in many geographic locations, nebulizers remain standard for drug delivery to the lungs.

Though the construction of the device can vary, a basic jet nebulizer design is shown schematically in Figure 13. Air supplied by a compressor, or from a wall source, is accelerated to high speed through a nozzle. At the nozzle exit the pressure is lower than that in the liquid reservoir, thus liquid is drawn up through a feed tube toward the high-speed air. This interface between liquid and air constitutes the region of primary droplet production. Different explanations have been proposed to describe the instability of the interface that coincides with the formation of primary droplets; however, it is likely the case that this instability results from the discontinuity in viscosity between air and liquid across the interface [5]. The large droplets produced at the interface are directed toward a primary baffle, where they splash and produce smaller, respirable size droplets. These droplets then flow around secondary baffles, which filter out any remaining large droplets, and exit the nebulizer through a mouthpiece.

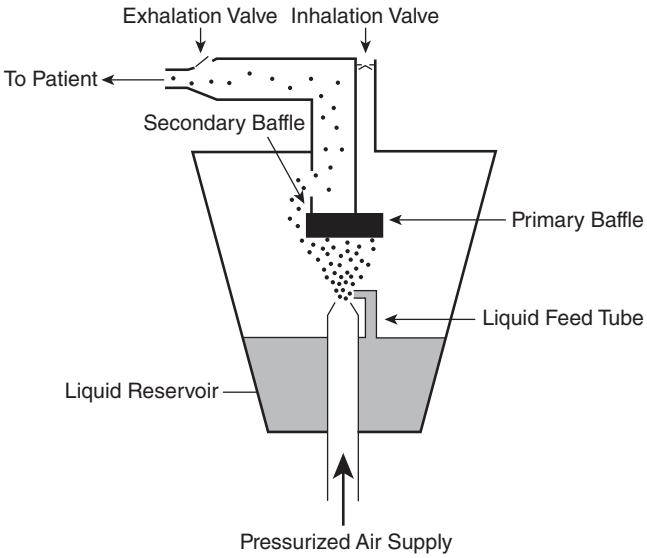


**Figure 13** Schematic diagram of a typical vented jet nebulizer.

An important parameter in assessing nebulizer performance is the nebulization efficiency, that is, the fraction of the nominal dose put in the nebulizer that is delivered to the patient. The nebulization efficiency is highly dependent on the design of the device, particularly on its ability to recycle droplets collected on the secondary baffles and nebulizer walls back into the liquid reservoir, to minimize residual liquid left in the reservoir after nebulization, and to limit aerosol losses during exhalation. Typically, the supply of pressurized air to the nebulizer is smaller than the patient's inhalation flow rate, so that additional ambient air is inhaled. If this additional air is drawn through the nebulizer, as in Figure 13, the nebulizer is referred to as a vented, breath-enhanced, or active venturi, nebulizer. One-way inhalation and exhalation valves are employed to limit the amount of aerosol that escapes the nebulizer during exhalation. If, instead, ambient air is mixed with the nebulizer output at the mouthpiece, and no valves are employed, the device is referred to as an unvented nebulizer. In general, with other factors being equal, unvented nebulizers are less efficient than vented nebulizers, as more aerosol is wasted during exhalation. For young children, whose inhalation flow rates are smaller than the rate at which pressurized air is supplied to the nebulizer, nebulization efficiency is reduced even further, as a fraction of the aerosol is also lost during inhalation.

Another parameter used to assess nebulizer performance is the output rate, defined as the amount of drug that is delivered over a given period of time. Low output rates can reduce patient compliance, as treatment times become longer. As a nebulizer is operated, the concentration of dissolved drug in the solvent will gradually increase as a result of solvent evaporation [135]. This effect can lead to increases in drug output rate as nebulization progresses [136]. However, jet nebulizers also become cooler during operation as a result of solvent evaporation [135], and the increase in solvent viscosity with decreasing temperature can be expected to reduce the drug output rate [5]. Such an effect may limit increases in output rate owing to increased drug concentration in some cases, so that the output rate remains relatively constant (e.g., [137]).

Jet nebulizers produce aqueous droplets that can undergo hygroscopic size changes; therefore, care must

be taken in order to accurately determine droplet sizes. Different approaches to sizing volatile aerosols were discussed earlier in Section 3.4. Briefly, the aerosol should be sized in situ as close to the nebulizer mouthpiece as possible, or else hygroscopic size changes should be negligible during transit from the mouthpiece to a sizing device located at some position downstream. Clark [105] has described the use of laser diffraction instruments for sizing nebulizer aerosols, while Smaldone and Lesouef [138] have described sizing of nebulizer aerosols using low flow rate cascade impactors. In addition, Finlay et al. [103] combined size measurements for a range of jet nebulizers made using a phase Doppler particle analyzer with a mathematical deposition model to accurately estimate in vivo lung deposition data. In predicting respiratory tract deposition from measured droplet size distributions, it must be ensured that any effect of breathing pattern on aerosol production, as occurs for some vented nebulizers [103], is accurately captured in the experimental setup.

Quite recently, a group of "smart nebulizers" were developed [139], with the aim of improving nebulization efficiency and targeting of inhaled aerosols. Many of these devices only produce aerosol during patient inhalation and, thereby, minimize losses during exhalation [140, 141]. Such devices are classified as breath-actuated nebulizers. Another device used in combination with conventional jet nebulizers aims to optimize aerosol deposition in prescribed lung regions by combining breath-actuated nebulization with control of patient inhalation maneuvers [140]. This latter type of delivery system has promising potential for tailoring aerosol drug delivery to individual patients.

***New Aqueous Aerosol Generators*** The traditional jet nebulizer is inherently inefficient given that a large fraction of its dose is wasted during exhalation. In addition, treatment times for jet nebulizers are much longer than those for single-breath inhalers such as pMDIs and DPIs. Not surprisingly, much work has been done in recent years toward improving existing nebulizer designs and in developing portable, single-breath aqueous aerosol generators to compete with pMDIs and DPIs. Many newer devices rely on mechanisms other than jet nebulization in order to produce aqueous aerosols. These include, but are not limited to, vibrating orifice-based nebulizers [142, 143], electrohydrodynamic aerosol generators [144], and devices that force liquid through machined nozzle systems or arrays to produce fine droplets [145, 146]. Some of these devices are capable of producing very high fine particle fractions in slow-moving aerosol clouds, for which extrathoracic deposition is much lower than that for most pMDIs and DPIs [147]. In addition, the relative ease of aqueous formulation is advantageous in developing delivery systems for new drug compounds. Reviews of new technologies for delivery of aqueous aerosols to the lungs are available in Dennis and Nerbrink [148] and in Lange and Finlay [149]. It is anticipated that these technologies will play a significant role in the future of aerosol drug delivery.

### 4.4 Condensation Aerosols

Delivery systems that form aerosols from condensation of vapors constitute another class of inhalers that has emerged fairly recently [150–152]. Liquid solutions of drug [152] or solid films of drug [150] can be rapidly heated to form vapors, which then condense upon cooling and mixing with ambient airflow to produce aerosols.

Though condensation aerosol generators have been used for several decades to produce test aerosols in laboratory settings [9, 153], adapting the technology to drug delivery presents two clear challenges. First, the heat required for vaporization can cause decomposition of the drug, reducing bioavailability and potentially producing harmful combustion products. Second, condensation often produces aerosol particles with submicron diameters [152], for which deposition in the lungs by gravitational sedimentation is less probable than for micron-sized particles. As a result, a significant fraction of the dose inhaled by patients would be expected to be exhaled, though this limitation would likely be balanced by reduced extrathoracic deposition compared to traditional single-breath inhalers.

Rabinowitz et al. [150] demonstrated that for a variety of drugs medically suited for systemic delivery through the lungs, thermal decomposition can largely be avoided by keeping the time period over which the drug is vaporized very short. Furthermore, aerosols consisting purely of drug were produced with MMADs between approximately 1 and 3 μm [150] suitable for deposition in the peripheral regions of the lungs. However, MMADs decreased with increasing flow rate of mixing air through the device [150], owing to dilution of the drug vapor prior to condensation, and the tested flow rates were lower than typical patient inhalation flow rates through single-breath inhalers. Consequently, further study is required to determine the size of particles delivered to patients.

Condensation-based delivery systems likely will not be suitable for all drugs, as some will undergo degradation due to heating. Given the small particle sizes that are produced, such devices appear to be particularly attractive for delivery to the peripheral regions of the lungs, with the intention that the dose be absorbed into the bloodstream for systemic action. Although

condensation-based systems are promising, further research will be required in order to optimize particle size distributions for maximum lung deposition, while minimizing losses upon exhalation.

## 5. CONCLUDING REMARKS

Inhaled pharmaceutical aerosols are already widely used for treatment of respiratory disease. As next-generation devices, designed for delivery of drugs through the lungs into the bloodstream, reach the market, aerosol drug delivery stands to transform the manner in which many patients take their medication. In order to maximize the therapeutic benefit to the patient, while minimizing the risks of side effects and variability in dosing, understanding the link between physical properties of pharmaceutical aerosols and outcomes measured in the clinic is crucial. In large part, the quality of an aerosol delivery device is determined by its ability to limit deposition of drug in the mouth and throat, while maintaining an adequate delivered dose to targeted regions of the lungs. Consequently, the ability to predict regional deposition patterns of inhaled pharmaceutical aerosols, gained through an understanding of the mechanics governing particle transport and deposition in the respiratory tract, has proven indispensable in the development and evaluation of aerosol drug delivery systems. In the future, progress toward targeting pharmaceutical aerosols to more localized regions within the lungs, and toward tailoring delivery systems to individual users, should expand the range of drugs delivered through the inhalation route.

## REFERENCES


1. Phillips CG, Kaye SR, Schroter RC (1994). A diameter-based reconstruction of the branching pattern of the human bronchial tree. *Respir Physiol 98*, 193–217.
2. Weibel ER (1963). *Morphometry of the Human Lung*. New York: Academic.
3. International Commission on Radiological Protection (ICRP) (1994). Publication 66. *Ann ICRP* 24(1–3). Tarrytown, NY: Pergamon/Elsevier.
4. Altiere RJ, Thompson DC (1996). Physiology and pharmacology of the airways. In Hickey AJ, Ed. *Inhalation Aerosols*. New York: Marcel Dekker.
5. Finlay WH (2001). *The Mechanics of Inhaled Pharmaceutical Aerosols*. London: Academic.
6. Stocks J, Hislop AA (2002). Structure and function of the respiratory system: Developmental aspects and their relevance to aerosol therapy. In Bisgaard H, O'Callaghan C, Smaldone GC, Eds. *Drug Delivery to the Lung*. New York: Marcel Dekker.
7. Stahlhofen W, Rudolf G, James AC (1989). Intercomparison of experimental regional aerosol deposition data. *J Aerosol Med 2*, 285–308.
8. Willeke K, Baron PA (1993). *Aerosol Measurement*. New York: Van Nostrand Reinhold.
9. Hinds WC (1999). *Aerosol Technology*, 2nd ed. New York: Wiley.
10. Landahl HD (1950). On the removal of air-borne droplets by the human respiratory tract: I. The lung. *Bull Math Biophys 12*, 43–56.
11. Taulbee DB, Yu CP (1975). A theory of aerosol deposition in the human respiratory tract. *J Appl Physiol 38*, 77–85.
12. Chan TL, Lippmann M (1980). Experimental measurement and empirical modeling of the regional deposition of inhaled particles in humans. *Am Ind Hyg Assoc J 41*, 399–409.
13. Yeh H-C, Schum GM (1980). Models of human lung airways and their application to inhaled particle deposition. *Bull Math Biol 42*, 461–480.
14. Cai FS, Yu CP (1988). Inertial and interceptional deposition of spherical particles and fibers in a bifurcating airway. *J Aerosol Sci 19*, 679–688.
15. Ferron GA, Kreyling WG, Haider B (1988). Inhalation of salt aerosol particles—II. Growth and deposition in the human respiratory tract. *J Aerosol Sci 19*, 611–631.
16. Balashazy I, Hofmann W, Martonen TB (1991). Inspiratory particle deposition in airway bifurcation models. *J Aerosol Sci 22*, 15–30.
17. Kim CS, Fisher DM, Lutz DJ, Gerrity TR (1994). Particle deposition in bifurcating airway models with varying airway geometry. *J Aerosol Sci 25*, 567–581.
18. Heyder J, Gebhart J (1977). Gravitational deposition of particles from laminar aerosol flow through inclined circular tubes. *J Aerosol Sci 8*, 289–295.
19. Koblinger L, Hofmann W (1990). Mont Carlo modeling of aerosol deposition in human lungs. Part I. Simulation of particle transport in a stochastic lung structure. *J Aerosol Sci 21*, 661–674.
20. Buchwald, E. (1921). Die Austrittswahrscheinlichkeit kolloidaler Teilchen und verwandte Begriffe. *Annalen der Physik*, *66*, 1–14.
21. Gormley PG, Kennedy K (1949). Diffusion from a stream flowing through a cylindrical tube. *Proc R Irish Soc 52A*, 163.
22. Fuchs NA (1964). *The Mechanics of Aerosols*. New York: Dover.
23. Ingham DB (1975). Diffusion of aerosols from a stream flowing through a cylindrical tube. *J Aerosol Sci 6*, 125–132.
24. Chow HY, Mercer TT (1971). Charges on droplets produced by atomization of solutions. *Am Ind Hyg Assoc J April*, 247–255.
25. Byron PR, Peart J, Staniforth JN (1997). Aerosol electrostatics I: Properties of fine powders before and after

aerosolization by dry powder inhalers. *Pharm Res 14*, 698–705.

26. Peart J, Magyar C, Byron PR (1998). Aerosol electrostatics—Metered dose inhalers (MDIs): Reformulation and device design issues. In Dalby RN, Byron PR, Farr SJ, Eds. *Respiratory Drug Delivery VI*. Buffalo Grove, Ill: Interpharm, pp. 227–233.
27. Glover W, Chan HK (2004). Electrostatic charge characterization of pharmaceutical aerosols using electrical low-pressure impaction (ELPI). *J Aerosol Sci 35*(6), 755–764.
28. Kwok PCL, Glover W, Chan HK (2005). Electrostatic charge characteristics of aerosols produced from metered dose inhalers. *J Pharm Sci 94*(12), 2789–2799.
29. Pierart F, Wildhaber JH, Vrancken I, Devadason SG, Le Souef PN (1999). Washing plastic spacers in household detergent reduces electrostatic charge and greatly improves delivery. *Eur Respir J 13*(3), 673–678.
30. Grgic B, Martin AR, Finlay WH (2006). The effect of unsteady flow rate increase on in vitro mouth-throat deposition of inhaled boluses. *J Aerosol Sci 40*, 361–370.
31. Yu CP (1985). Theories of electrostatic lung deposition of inhaled aerosols. *Ann Occup Hyg 29*, 219–227.
32. Melandri C, Tarroni G, Prodi V, Zaiacomo T De, Formignani M, Lombardi CC (1983). Deposition of charged particles in the human airways. *J Aerosol Sci 14*, 657–669.
33. Pitcairn G, Lunghetti G, Ventura P, Newman S (1994). A comparison of the lung deposition of salbutamol inhaled from a new dry powder inhaler, at two inhaled flow rates. *Int J Pharm 102*, 11–18.
34. Newman SP, Hollinworth A, Clark AR (1994). Effect of different modes of inhalation on drug delivery from a dry powder inhaler. *Int J Pharm 102*, 127–132.
35. Thorsson L, Kenyon C, Newman SP, Borgstrom L (1998). Lung deposition of budesonide in asthmatics: A comparison of different formulations. *Int J Pharm 168*(1), 119–127.
36. Newman SP, Pitcairn GR, Adkin DA, Vidgrin MT, Silvasti M (2001). Comparison of beclomethasone dipropionate delivery by easyhaler dry powder inhaler and pMDI plus large volume spacer. *J Aerosol Med 14*(2), 217–225.
37. Warren S, Taylor G, Smith J, Buck H, Parry-Billings M (2002). Gamma scintigraphic evaluation of a novel budesonide dry powder inhaler using a validated radiolabeling technique. *J Aerosol Med 15*(1), 15–25.
38. Rudolf G, Kobrich R, Stahlhofen W (1990). Modeling and algebraic formulation of regional land aerosol deposition in man. *J Aerosol Sci 21*(Suppl 1), S403–S406.
39. McRobbie DW, Pritchard S, Quest RA (2003). Studies of the human oropharyngeal airspaces using magnetic resonance imaging. 1. Validation of a three-dimensional MRI method for producing ex vivo virtual and physical casts of the oropharyngeal airways during inspiration. *J Aerosol Med 16*(4), 401–415.
40. Grgic B, Heenan AF, Burnell PKP, Finlay WH (2004). In vitro intersubject and intrasubject deposition measurements in realistic mouth-throat geometries. *J Aerosol Sci 35*, 1025–1040.
41. Clark AR, Egan M (1994). Modeling the deposition of inhaled powdered drug aerosols. *J Aerosol Sci 25*(1), 175–186.
42. Clark AR, Newman SP, Dasovich N (1998). Mouth and oropharyngeal deposition of pharmaceutical aerosols. *J Aerosol Med 11S*(1), S116–S120.
43. Lin T-C, Breysse PN, Laube B, Swift DL (2001). Mouthpiece diameter affects deposition efficiency in cast models of the human oral airways. *J Aerosol Med 14*, 335–341.
44. DeHaan WH, Finlay WH (2001). In vitro monodisperse aerosol deposition in a mouth and throat with six different inhalation devices. *J Aerosol Med 14*, 361–367.
45. DeHaan WH, Finlay WH (2004). Predicting extrathoracic deposition from dry powder inhalers. *J Aerosol Sci 35*, 309–331.
46. Hof VI, Patrick G (1994). Particle retention and clearance. *J Aerosol Med 7*(1), 39–47.
47. Hickey AJ, Swift D (2001). Measurement of pharmaceutical and diagnostic aerosols. In Baron PA, Willeke K, Eds. *Aerosol Measurement*. New York: Wiley.
48. Lippmann M (1977). Regional deposition of particles in the human respiratory tract. In Lee DHK, et al, Eds. *Hanbook of Physiology—Reaction to Environmental Agents*. Bethesda, MD: American Physiological Society, pp. 213–232.
49. Riley DJ, Weitz BW, Edelman NH (1976). The responses of asthmatic subjects to isoproterenol inhaled at differing lung volumes. *Am Rev Respir Dis 114*, 509–515.
50. Bennet WD (1991). Targeting respiratory drug delivery with aerosol boluses. *J Aerosol Med 4*(2), 69–78.
51. Groth ML, Bergofsky EH, Foster WM (1992). Aerosolized atropine sulfate: Influence of inhalation pattern on effective blockade of vagal tone. *Am Rev Respir Dis 145*, 215–219.
52. Hickey AJ, Martonen TB (1993). Behavior of hygroscopic pharmaceutical aerosols and the influence of hydrophobic additives. *Pharm Res 10*(1), 1–7.
53. Finlay WH (1998). Estimating the type of hygroscopic behavior exhibited by aqueous droplets. *J Aerosol Med 11*, 221–229.
54. Peng CG, Chow AHL, Chan CK (2000). Study of the hygroscopic properties of selected pharmaceutical aerosols using single particle levitation. *Pharm Res 17*(9), 1104–1109.
55. Davies CN (1982). Deposition of particles in the human lung as a function of particle size and breathing pattern: An empirical model. *Ann Occup Hyg 26*, 119–135.
56. Rudolf G, Gebhart J, Heyder J, Schiller ChF, Stahlhofen W (1986). An empirical formula describing aerosol deposition in man for any particle size. *J Aerosol Sci 17*, 350–355.

57. Yu CP, Zhang L, Becquemin MH, Roy M, Bouchikhi A (1992). Algebraic modeling of total and regional deposition of inhaled particles in the human lung of various ages. *J Aerosol Sci 23*, 73–79.
58. Heyder J, Gebhart J, Scheuch G (1985). *Interaction* of diffusional and gravitational transport in aerosols. *Aerosol Sci Tech 4*, 315–326.
59. Balashazy I, Martonen TB, Hofmann W (1990). Inertial impaction and gravitational deposition of aerosols in curved tubes and airway bifurcations. *Aerosol Sci Tech 13*, 303–321.
60. Balashazy I, Martonen TB, Hofmann W (1990). Simultaneous sedimentation and impaction of aerosols in two-dimensional channel bends. *Aerosol Sci Tech 13*, 20–34.
61. Asgharian B, Anjilvel S (1994). A Monte Carlo calculation of the deposition efficiency of inhaled particles in lower airways. *J. Aerosol Sci 4*, 711–721.
62. Persons DD, Hess GD, Muller WJ, Scherer PW (1987). Airway deposition of hygroscopic heterodispersed aerosols: Results of a computer calculation. *J Appl Physiol 63*, 1195–1204.
63. Finlay WH, Stapleton KW (1995). The effect on regional lung deposition of coupled heat and mass transfer between hygroscopic droplets and their surrounding phase. *J Aerosol Sci 26*, 655–670.
64. Finlay WH, Lange CF, King M, Speert DP (2000). Lung delivery of aerosolized dextran. *Am J Respir Crit Care Med 161*, 91–97.
65. Lange, CF, Hancock, RE (2001). In vitro aerosol delivery and regional airway surface liquid concentration of a liposomal cationic peptide. *Journal of Pharmaceutical Sciences. 90*(10), 1647–1657.
66. Desai, EW and Finlaly, WH. (2003). A novel approach to the delivery by inhalation of liposomal polymyxin B sulfate. *Pharm Res 20*, 442–447.
67. Sweeney et al. (2005).
68. Hofmann W (1982). Mathematical model for the postnatal growth of the human lung. *Respir Physiol 49*, 115–367.
69. Phalen RF, Oldham MJ, Beaucage CB, Crocker TT, Martensen JD (1985). Postnatal enlargement of human tracheo-bronchial airways and implications for particle deposition. *Anat Rec 212*, 368–380.
70. Xu GB, Yu CP (1986). Effects of age on deposition of inhaled aerosols in the human lung. *Aerosol Sci Tech 5*, 349–357.
71. Hofmann W, Martonen TB, Graham RC (1989). Predicted deposition of nonhygroscopic aerosols in the human lung as a function of subject age. *J Aerosol Med 2*, 49–68.
72. Swift DL (1996). Use of mathematical aerosol deposition models in predicting the distribution of inhaled therapeutic aerosols. In Hickey AJ, Ed. *Inhalation Aerosols*. New York: Marcel Dekker.
73. Finlay WH (2003). Lung deposition simulation. In Hickey AJ, Ed. *Inhalation Aerosol Technology*, 2nd ed. New York: Marcel Dekker.
74. Matida EA, DeHaan WH, Finlay WH, Lange CF (2003). Simulation of particle deposition in an idealized mouth with different small diameter inlets. *Aerosol Sci Tech 37*, 924–932.
75. Kleinstreuer C, Zhang Z (2003). Laminar-to-turbulent fluid-particle flows in a human airway model. *Int J Multiphase Flow 29*(2), 271–289.
76. Matida EA, Finlay WH, Lange CF, Grgic B (2004). Improved numerical simulation of aerosol deposition in an idealized mouth-throat. *J Aerosol Sci 35*, 1–19.
77. Zhang Z, Kleinstreuer C, Kim CS, Cheng YS (2004). Vaporizing microdroplet inhalation, transport, and deposition in a human upper airway model. *Aerosol Sci Tech 38*, 36–49.
78. Martonen TB, Yang Y, Xue ZQ (1994). Effects of carinal ridge shapes on lung airstreams. *Aerosol Sci Tech 21*, 119–136.
79. Balashazy I, Heistracher T, Hofmann W (1996). Air flow and particle deposition patterns in bronchial airway bifurcations: The effect of different CFD models and bifurcation geometries. *J Aerosol Med 9*(3), 287–301.
80. Ferron GA, Edwards DA (1996). Numerical simulation of air and particle transport in the conducting airways. *J Aerosol Med 9*(3), 303–316.
81. Zhang Z, Kleinstreuer C (2002). Transient airflow structures and particle transport in a sequentially branching lung airway model. *Phys Fluids 14*(2), 862–880.
82. van Ertbruggen C, Hirsch C, Paiva M (2005). Anatomically based three-dimensional model of airways to simulate flow and particle transport using computational fluid dynamics. *J Appl Phys 98*(3), 970–980.
83. Coates MS, Fletcher DF, Chan HK, Raper JA (2004). Effect of design on the performance of a dry powder inhaler using computational fluid dynamics. Part 1: Grid structure and mouthpiece length. *J Pharm Sci 93*, 2863–2876.
84. Matida E, Rimkus M, Grgic B, Lange CF, Finlay WH (2004). A new add-on spacer design concept for dry powder inhalers. *J Aerosol Sci 35*, 823–833.
85. Shakked T, Katoshevski D, Broday DM, Amirav I (2005). Numerical simulation of air flow and medical-aerosol distribution in an innovative nebulizer hood. *J Aerosol Med 18*, 207–217.
86. Dolovich M, Rhem R (1998). Impact of oropharyngeal deposition on inhaled dose. *J Aerosol Med 11*(1), S112–115.
87. Clark AR, Borgstrom L (2002). In vitro testing of pharmaceutical aerosols and predicting lung deposition from in vitro measurements. In Bisgaard H, O'Callaghan C, Smaldone GC, Eds. *Drug Delivery to the Lung*. New York: Marcel Dekker.

88. Srichana T, Martin G, Marriott C (2000). A human oral-throat cast integrated with a twin-stage impinger for evaluation of dry powder inhalers. *J Pharm Pharmacol 52*, 771–778.
89. Zhang Y, Finlay WH, Matida EA (2004). Particle deposition measurements and numerical simulation in a highly idealized mouth-throat. *J Aerosol Sci 35*, 789–803.
90. Grgic B, Finlay WH, Heenan AF (2004). Regional aerosol deposition and flow measurements in an idealized mouth and throat. *J Aerosol Sci 35*, 21–32.
91. Dunbar CA, Hickey AJ (2000). Evaluation of probability density functions to approximate particle size distributions of representative pharmaceutical aerosols. *J Aerosol Sci 31*(7), 813–831.
92. Dunbar C, Kataya A, Tiangbe T (2005). Reducing bounce effects in the Andersen cascade impactor. *Int J Pharm 301*(1–2), 25–32.
93. Finlay WH, Stapleton KW, Zuberbuhler P (1997). Predicting regional lung dosages of a nebulized suspension: Pulmicort® (budesonide). *Particulate Sci Tech 15*(3), 185–193.
94. Mitchell JP, Nagel MW (2003). Cascade impactors for the size characterization of aerosols from medical inhalers: Their uses and limitations. *J Aerosol Med 16*(4), 341–377.
95. Marple VA, Olson BA, Miller NC (1998). The role of inertial particle collectors in evaluating pharmaceutical aerosol delivery systems. *J Aerosol Med 11*, S139–S153.
96. Mitchell JP, Nagel MW, Cheng YS (1999). Use of the aerosizer aerodynamic particle size analyzer to characterize aerosols from pressurized metered-dose inhalers (pMDIs) for medication delivery. *J Aerosol Sci 30*(4), 467–477.
97. Baron PA, Heitbrink WA (1993). Factors affecting aerosol measurement quality. In Willeke K, Baron PA, Eds. *Aerosol Measurement*. New York: Van Nostrand Reinhold.
98. Mitchell JP, Nagel MW, Archer AD (1999). Size analysis of a pressurized metered dose inhaler-delivered suspension formulation by the API aerosizer time-of-flight aerodynamic particle size analyzer. *J Aerosol Med 12*(4), 255–264.
99. Mitchell JP, Nagel MW (1996). An assessment of the API Aerosizer(R) for the realtime measurement of medical aerosols from pressurized metered-dose inhaler (pMDI) systems. *Aerosol Sci Tech 25*(4), 411–423.
100. Rader DJ, O'Hern TJ (2001). Optical direct-reading techniques: In situ sensing. In Baron PA, Willeke K, Eds. *Aerosol Measurement*, 2nd ed. New York: Wiley.
101. Dantec (1991). *PDA User's Manual*. Dantec Measurement Technology, Skovlunde, Denmark.
102. Dunbar CA, Watkins AP, Miller JF (1997). An experimental investigation of the spray issued from a pMDI using laser diagnostic techniques. *J Aerosol Med 10*(4), 351–368.
103. Finlay WH, Stapleton KW, Zuberbuhler P (1998). Variations in predicted regional lung deposition of salbutamol sulphate between 19 nebulizer types. *J Aerosol Med 11*(2), 65–80.
104. Corcoran, TE, Chigier, N (2000). Characterization of the laryngeal jet using phase doppler interferometry. *Journal of Aerosol Medicine*, *13*(2), 125–137.
105. Clark AR (1995). The use of laser diffraction for the evaluation of the aerosol clouds generated by medical nebulizers. *Int J Pharm 115*, 69–78.
106. Vecellio None L, Grimbert D, Becquemin MH, Boissinot E, Le Pape A, Lemarié E, Diot P (2001). Validation of laser diffraction method as a substitute for cascade impaction in the European project for a nebulizer standard. *J Aerosol Med 14*(1), 107–114.
107. Matsuyama T, Yamamoto H (2004). Particle shape and laser diffraction: A discussion of the particle shape problem. *J Disp Sci Tech 25*(4), 409–416.
107a. Xu R, Di Guida OA (2003). Comparison of sizing small particles using different technologies. *Powder Tech 132*, 145–153.
108. Mühlenweg H, Hirleman ED (1998). Laser diffraction spectroscopy: Influence of particle shape and a shape adaptation technique. *Part Part Syst Charact 15*, 163–169.
109. Sangwan S, Condos R, Smaldone GC (2003). Lung deposition and respirable mass during wet nebulization. *J Aerosol Med 16*(4), 379–386.
110. Finlay WH, Stapleton KW (1999). Undersizing of droplets from a vented nebulizer caused by aerosol heating during transit through a cascade impactor. *J Aerosol Sci 30*(1), 105–109.
111. O'Callaghan C, Nerbrink O, Vidgren MT (2002). The history of inhaled drug therapy. In Bisgaard H, O'Callaghan C, Smaldone GC, Eds. *Drug Delivery to the Lung*. New York: Marcel Dekker.
112. Vervaet C, Byron PR (1999). Drug-surfactant-propellant interactions in HFA-formulations. *Int J Pharm 186*, 13–30.
113. Smyth HDC (2003). The influence of formulation variables on the performance of alternative propellant-driven metered dose inhalers. *Adv Drug Deliv Rev 55*(7), 807–828.
114. Dunbar CA, Watkins AP, Miller JF (1997). Theoretical investigation of the spray from a pressurized metered-dose inhaler. *Atom Sprays 7*, 417–436.
115. Biddiscombe MF, Melchor R, Mak VHF, Marriot RJ, Taylor AJ, Short MD, Spiro SG (1993). The lung deposition of salbutamol, directly labeled with TC-99M, delivered by pressurized metered-dose and dry powder inhalers. *Int J Pharm 91*(2–3), 111–121.
116. Thiel CG (1998). Can in vitro particle size measurements be used to predict pulmonary deposition of aerosols from inhalers? *J Aerosol Med 11*(1), S43–S51.
117. Newman S, Pitcairn G, Steed K, Harrison A, Nagel J (1999). Deposition of fenoterol from pressurized

metered dose inhalers containing hydrofluoroalkanes. *J Allergy Clin Immunol 104*(6), S253–S257.

118. Newman SP, Steed KP, Hooper G, Jones JI, Upchurch FC (1999). Improved targeting of beclomethasone diproprionate (250 mu g metered dose inhaler) to the lungs of asthmatics with the Spacehaler (TM). *Respir Med 93*(6), 424–431.
119. Chapman KR, Love L, Brubaker H (1993). A comparison of breath-actuated and conventional metered-dose inhaler inhalation techniques in elderly subjects. *Chest 140*(5), 1332–1337.
120. Leach CL, Davidson PJ, Hasselquist BE, Boudreau RJ (2005). Influence of particle size and patient dosing technique on lung deposition of HFA-beclomethasone from a metered dose inhaler. *J Aerosol Med 18*(4), 379–385.
121. Hirst PH, Pitcairn GR, Richards JC, Rohatagi S, Gillen MS, Newman SP (2001). Deposition and pharmacokinetics of an HFA formulation of triamcinolone acetonide delivered by pressurized metered dose inhaler. *J Aerosol Med 14*(2), 155–165.
122. Rahmatalla MF, Zuberbuhler PC, Lange CF, Finlay WH (2002). In vitro effect of a holding chamber on the mouth-throat deposition of QVAR® (hydrofluoroalkanebeclomethasone dipropionate). *J Aerosol Med 15*(4), 379–385.
123. Nagle MW, Schmidt JN, Doyle CC, Varallo VM, Mitchell JP (2003). *In vitro* performance of a new non-electrostatic transparent valved holding chamber (VHC). In *Drug Delivery to the Lungs 14*, 71–74.
124. Asmus MJ, Arya V, Hochhaus G (2003). An *in vitro* comparison of two non-electrostatic valved holding chambers with a fluticasone metered-dose inhaler. *Chest 124*(4), S106.
125. Girodet P-O, Molimard M (2005). Pharmacological approach to evaluate aerosol pulmonary deposition. *J Aerosol Med 18*(2), 183–192.
126. Everard ML, Dolovich M. (2002). In vivo measurement of lung dose. *J Aerosol Med 15*(2), 179–188.
127. Newman SP, (1993) Scintigraphic assessment of therapeutic aerosols. *Crit Rev Ther Drug Carrier Sys 10*, 65–109.
128. Dunbar CA, Hickey AJ, Holzner P (1998). Dispersion and characterization of pharmaceutical dry powder aerosols. *KONA*, *Powder and Particle 16*, 7–45.
129. Borgström L, Bisgaard H, O'Callaghan C, Pedersen S (2002). Dry powder inhalers. In Bisgaard H, O'Callaghan C, Smaldone GC, Eds. *Drug Delivery to the Lungs*. New York: Marcel Dekker.
130. Chan HK (2006). Dry powder aerosol delivery systems: Current and future research directions. *J Aerosol Med 19*(1), 21–27.
131. Edwards DA, Hanes J, Caponetti G, Hrkach J, BenJebria A, Eskew ML, Mintzes J, Deaver D, Lotan N, Langer R (1997). Large porous particles for pulmonary drug delivery. *Science 276*(5320), 1868–1871.
132. Warren S, Taylor G, Godfrey C, Cote G, Hill M (1999). Gamma scintigraphic evaluation of dry powder beclomethasone dipropionate (BDP) from an investigational Spiros((R)) inhaler: Effect of inspiratory flow profile on pulmonary drug deposition. *Am J Respir Crit Care Med 159*(3), 118–125.
133. Podczeck F, Newton JM, James MB (1997). Influence of relative humidity of storage air on the adhesion and autoadhesion of micronized particles to particulate and compacted powder surfaces. *J Coll Interface Sci 187*, 484–491.
134. Finlay WH, Gehmlich MG (2000). Inertial sizing of aerosol inhaled from two dry powder inhalers with realistic breath patterns versus constant flow rates. *Int J Pharm 210*(1–2), 83–95.
135. Mercer TT, Tillery MI, Chow HY (1968). Operating characteristics of some compressed-air nebulizers. *Am Ind Hyg J 29*, 66–78.
136. Diederik H, Le Brun PPH, Frijlink HW, Vitanyi PMB, Weda M, Barends DM (2003). Drug output of unvented jet nebulizers as a function of time. *Int J Pharm 257*(1–2), 33–39.
137. Smaldone GC, Dickinson G, Marcial E, Young E, Seymour J (1992). Deposition of aerosolized pentamidine and failure of pneumocystis prophylaxis. *Chest 101*, 32–87.
138. Smaldone GC, LeSouef PN (2002). Nebulization: The device and clinical considerations. In Bisgaard H, O'Callaghan C, Smaldone GC, Eds. *Drug Delivery to the Lung*. New York: Marcel Dekker.
139. Smaldone GC (2002). Smart Nebulizers. *Respir Care 47*, 1434–1441.
140. Brand P, Beckmann H, Mass-Enriquez M et al (2003). Peripheral deposition of alpha1-protease inhibitor using commercial inhalation devices. *Eur Respir J 22*, 263–267.
141. Rau JL, Ari A, Restripo RD (2004). Performance comparison of nebulizer designs. *Am J Respir Crit Care Med 169*, 822–828.
142. Stangl R, Waldner R, Seemann S, Knoch M (2003). In-vitro characterisation of the eFlow™ electronic inhaler. *J Aerosol Med 16*(2), 197.
143. Lipworth BJ, Sims EJ, Taylor K, Cockburn W, Fishman R (2005). Dose-response to salbutamol via a novel palm sized nebuliser (Aerodose (R) Inhaler), conventional nebuliser (Pari LC Plus) and metered dose inhaler (Ventolin (TM) Evohaler (TM)) in moderate to severe asthmatics. *Br J Clin Pharmacol 59*(1), 5–13.
144. Zimlich W, Ding J, Busick D, Pham S, Palmer D, Placke M (2002). Development of multiple clinical and commercial applications using Mystic™ inhalation delivery technologies. In Dalby RN, Byron PR, Peart J, Farr SJ, Eds. *Respiratory Drug Delivery VIII*. Surrey: David Horwood International Publishing, pp. 363–366.
145. Noymer P, Schuster J, Holst P, Rosell J, Ament B, Wilbanks T, Srinivasan S, Morishige R, Farr S (2004).

AERx®-Essence™: Efficiency and breath control without electronics. In Dalby RN, Byron PR, Peart J, Suman JD, Farr SJ, Eds. *Respiratory Drug Delivery IX*. River Grove, ILL: Davis Healthcare International Publishing, pp. 255–262.

146. Dalby R, Spallek M, Voshaar T (2004). A review of the development of Respimat soft mist inhaler. *Int J Pharm 283*(1–2), 1–9.
147. Pitcairn G, Reader S, Pavia D, Newman S (2005). Deposition of corticosteroid aerosol in the human lung by Respimat (R) Soft Mist (TM) inhaler compared to deposition by metered dose inhaler or by Turbuhaler (R) dry powder inhaler. *J Aerosol Med 18*(3), 264–272.
148. Dennis JH, Nerbrink O (2002). New nebulizer technology. In Bisgaard H, O'Callaghan C, Smaldone GC, Eds. *Drug Delivery to the Lung*. New York: Marcel Dekker.
149. Lange CF, Finlay WH (2006). Liquid atomizing: Nebulizing and other methods of producing aerosols. *J Aerosol Med 19*(1), 28–35.
150. Rabinowitz JD, Wensley M, Lloyd P, Myers D, Shen W, Lu A, Hodges C, Hale R, Mufson D, Zaffaroni A (2004). *J Pharmacol Exp Ther 309*, 769–775.
151. Shen X, Hindle M, Byron PR (2004). Effect of energy on propylene glycol aerosols using the capillary aerosol generator. *Int J Pharm 275*, 249–258.
152. Li X, Blondino FE, Hindle M, Soine WH, Byron PR (2005). Stability and characterization of perphenazine aerosols generated using the capillary aerosol generator. *Int J Pharm 303*, 113–124.
153. John W (1993). The characteristics of environmental and laboratory generated aerosols. In Willeke K, Baron PA, Eds. *Aerosol Measurement*. New York: Van Nostrand Reinhold.

## BIBLIOGRAPHY

Blanchard J, Coleman J, Hayling CD, Ghaderi R, Haeberlin B, Hart J, Jensen S, Malcolmson R, Mittelman S, Snodgrass-Pilla C, Sundahl M, Thompson G, Wolff R (2004). Foreign particles testing in orally inhaled and nasal drug products. *Pharm Res 21*(21), 2137–2147.

Matida EA, Finlay WH, Breuer M, Lange CF (2006). Improving prediction of aerosol deposition in an idealized mouth using large-eddy simulation". *J Aerosol Med 19*(3), 290–300.

Mitchell JP (2003). Regarding the development and practice of cascade impaction testing, including a mass balance failure investigation. *J Aerosol Med 16*(4), 433.

Nagao LM, Lyapustina S, Munos MK, Capizzi MD (2005). Aspects of particle science and regulation in pharmaceutical inhalation drug products. *Cryst Growth Des 5*(6), 2261–2267.

# RESISTANCE TO ANTIPLATELET THERAPY

Paul A. Gurbel and Udaya S. Tantry
*Sinai Center for Thrombosis Research, Sinai Hospital of Baltimore, Baltimore, Maryland*

## 1 INTRODUCTION

Atherothrombotic complications are the primary causes of cardiovascular mortality and morbidity. The common underlying pathological basis for the development of catastrophic ischemic events during cardiovascular diseases is the occlusive platelet-rich thrombus generation at the site of plaque rupture. Occlusive thrombus development is also responsible for microembolization, which results in microvascular dysfunction following stroke and myocardial infarction. Platelets are not only central to these thrombotic processes ("platelet hypothesis") but also play important roles in atherosclerosis, coagulation, and inflammation (Fig. 1) [1–3]. Therefore, pharmacological agents that suppress platelet function have proven to be beneficial during the acute settings of acute coronary syndrome (ACS) and percutaneous coronary interventions (PCI) where superior platelet inhibition is desired and also during long-term treatment of coronary artery disease, ischemic stroke, and peripheral arterial disease [4]. Since platelets are activated by multiple redundant pathways, simultaneous inhibition of two important secondary agonist pathways [thromboxane A2 by inhibiting cyclooxygenase-1 by aspirin and adenosine 5′-diphosphate (ADP) pathway by inhibiting $P2Y_{12}$ receptor by clopidogrel] is desired. Clinical efficacy of the dual antiplatelet therapy has been demonstrated in large-scale clinical trials. In these clinical trials, a "one size fits all" approach has been adopted without evaluating the platelet response in the individual patient. However, subsequent ex vivo laboratory evaluation of individual antiplatelet response indicated a wide response variability where a substantial number of patients were not receiving the desired benefits with the prescribed dosing regimens. This phenomenon is described as "antiplatelet resistance or nonresponsiveness" [5]. Moreover, continued occurrence of recurrent ischemic events among selected patients with high on-treatment platelet reactivity or among patients exhibiting antiplatelet resistance again highlighted the importance of limitations of uniform dual antiplatelet treatment strategy. Currently, 25 million people in United States use aspirin as an antiplatelet agent to prevent cardiovascular events, whereas, between 1998 and 2004, use of clopidogrel has increased over 7 times whereas the use of clopidogrel with aspirin has increased 16 times in the United States [6]. The "nonresponsiveness or resistance" to antiplatelet therapy is becoming a major concern. Therefore, numerous scientific meetings and hundreds of clinical publications have been focused on this phenomenon. However, the mechanism of antiplatelet resistance, and treatment strategies to address these concerns are still unclear.

## 2 ROLE OF PLATELETS IN THROMBOTIC EVENTS

Platelets are the smallest blood cells released from megakaryocytes in bone marrow and circulate in blood as disk-shaped anucleate fragments without attaching to the normal vessel wall. They play an important role during normal hemostasis by forming clots at the site of vascular injury to prevent excessive blood loss. In the

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

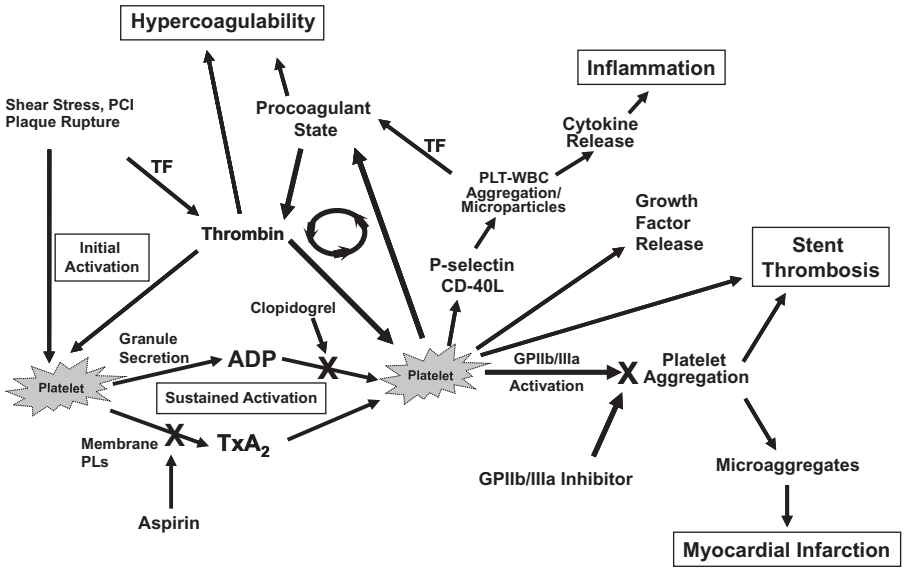


**Figure 1** The platelet hypothesis: The Nidus of evil.

setting of pre-existing dysfunctional endothelium and atheroinflammatory conditions, platelet activation is a central process precipitating the development of occlusive thrombus [7]. Platelets are activated by various pathophysiological processes such as altered shear forces and released soluble agonists at the site of plaque rupture. Endothelial denudation following plaque rupture leads to the exposure of collagen and von Willebrand's factor (vWF) as well as the synthesis of small quantities of thrombin generated by exposed tissue factor. Binding of collagen, vWF, and thrombin (primary platelet activating factors) to specific receptors on the platelet leads to the release of major secondary agonists such as thromboxane (Tx) $A_2$ from membrane phospholipids through cyclooxygenase COX-1 activity and ADP from dense granule secretion. These secondary agonists by paracrine and autocrine mechanism are responsible for the amplification of platelet activation and sustained platelet aggregation. TxA2 binds to thromboxane receptors, whereas ADP binds to specific G-protein coupled purinergic receptors, $P2Y_{12}$ and $P2Y_1$, which result in sustained platelet aggregation. Though both $P2Y_1$ and $P2Y_{12}$ pathways are needed for stable thrombus generation, the $P2Y_{12}$ pathway is a major pathway through which the activation of platelets by various primary and secondary agents converges in the activation of glycoprotein (GP) IIb/IIIa receptors ("final common pathway") and subsequent stable platelet–platelet aggregation (homotypic aggregation) [7–10]. Activation of platelet by these agonists also leads to surface expression of the activation-dependent molecules, P-selectin and CD-40L. These activation-dependent molecules facilitate the binding of platelets to leukocytes and endothelium (heterotypic aggregation) to further enhance inflammation and thrombin generation. Both homotypic and heterotypic aggregation lead to the release of microparticles that are important in the thrombosis, inflammation, and atherosclerosis processes [7, 8].

## 3 RATIONALE FOR ANTIPLATELET THERAPY

Platelet activation and high platelet reactivity to agonists in acute coronary syndromes and following percutaneous intervention have been associated with stent thrombosis, restenosis, inflammation, myocardial infarction, and other ischemic events (Fig. 1). Platelet activation is also associated with diabetes, hypertension, and hyperlipidemia [9]. Increased expressions of platelet surface markers and heightened platelet reactivity as measured by light transmittance aggregometry (LTA) have been demonstrated in patients with ACS [9]. Recent studies indicate that elevated posttreatment platelet reactivity after PCI is a risk factor for the development of long-term ischemic complications, stent thrombosis, and periprocedural myocardial necrosis [11–15]. Through crosstalk with the coagulation system, platelets play a fundamental role in generating a hypercoagulable state, which has been demonstrated as an important risk factor for ischemic events following stenting [16]. Thus, on one hand, platelets play an important first line of defense in the development of clot to arrest bleeding during normal physiological hemostasis, while on the other hand during pathological processes

the uncontrolled activation and high reactivity may result in various severe ischemic complications that position the platelet as a "nidus of evil" [7]. In these settings, the rationale for antiplatelet therapy is to prevent the development of thrombus formation, to arrest the procoagulant activity and finally to promote the dissolution of thrombus. The determination of optimal inhibition is dependent on the degree of ischemic risk and is counterbalanced by the risk of bleeding.

The meta-analysis of randomized trials of antiplatelet therapy have demonstrated a reduction in myocardial infarction (MI), stroke, and death with antiplatelet therapy (aspirin or another oral antiplatelet agent) in patients at "high risk" of cardiovascular diseases. This meta-analysis further demonstrated that low-dose aspirin (75–150 mg/day) is an effective antiplatelet agent for long-term treatment, whereas a higher initial dose of 150–300 mg aspirin may be required in acute settings of MI, ischemic stroke, and unstable angina where complete inhibition of thromboxane-mediated platelet aggregation is desired [17]. Based on these data, indefinite oral aspirin therapy for the secondary prevention of cardiovascular events is recommended [18, 19]. Addition of clopidogrel to aspirin is associated with significantly lower serious vascular events in the antithrombotic trialists' collaboration meta-analysis. Based on the marginal superiority over aspirin and less frequent gastrointestinal bleeding observed in the CAPRIE (clopidogrel versus aspirin in patients at risk of ischemic events) trial, United States Food and Drug Administration (FDA) approved the use of 75 mg/day of clopidogrel for patients with a history of recent heart attack, stroke, or established peripheral arterial diseases [20]. Subsequent clinical trials have demonstrated significant clinical efficacy of adding clopidogrel to aspirin in preventing cardiovascular events especially in high-risk patients [21–23]. A recent meta-analysis of randomized clinical trials that compared addition of clopidogrel to aspirin to aspirin plus placebo for the treatment of coronary artery disease demonstrated a small reduction in all-cause mortality in patients with ST-elevation myocardial infarction (6.3 vs. 6.7%, $p = 0.023$) and a modest reduction in myocardial infarction (2.7 vs. 3.3%, $p < 0.001$) and stroke (1.6 vs. 1.3%, $p = 0.002$) in patients with cardiovascular disease [24].

## 4 RESISTANCE TO ANTIPLATELET DRUGS

Despite significant benefits and widespread use of dual antiplatelet treatment, the continued occurrence of adverse ischemic events remains a serious clinical problem [1]. The major limitation of these clinical trials with dual antiplatelet therapy was the implementation of one size fits all principle while treating the wide range of cardiovascular patients. However, recent pharmacodynamic studies, based on the ex vivo measurement of antiplatelet response aggregation, have revealed various limitations of aspirin and clopidogrel treatment. There was a wide response variability with a certain percentage of patients exhibiting very limited response or complete nonresponsiveness. Moreover, small translation research studies correlated this nonresponsiveness to the occurrence of recurrent ischemic events. Other terms used to describe this phenomenon include "resistance" and "hyporesponsiveness." The phenomenon of resistance may be due to a (a) *pharmacokinetic* mechanism where a reduced absorption or conversion to the active metabolite leads to inadequate availability of drug/active metabolite to sufficiently block the respective receptor pathways, (b) *pharmacodynamic* mechanism where decreased response of platelets to the drug is observed and is due to variability in platelet response secondary to increased turnover of platelets, increased responsiveness of platelets to certain agonists, genetic variability of receptors or intracellular signaling mechanism, or interactions with other drugs at the active site of the receptors, and (c) *treatment failure* where patients may suffer recurrent vascular events despite inhibition of the platelet activation pathways by antiplatelet therapy [5].

Multiple signaling pathways mediate platelet activation and the occurrence of thrombotic events. Thus, a treatment strategy directed against a single pathway cannot be expected to prevent the occurrence of all events, and treatment failure following single antiplatelet strategy alone is not sufficient evidence of drug resistance. "Resistance" or nonresponsiveness to an antiplatelet is provided by persistent activity of the specific target of the antiplatelet agent. Since the active metabolite of clopidogrel irreversibly inhibits the $P2Y_{12}$ receptor, evidence for nonresponsiveness to clopidogrel is provided by high posttreatment $P2Y_{12}$ reactivity. For aspirin, the identification of resistance would use a laboratory technique that detects residual activity of COX-1 [5, 11]. The so-called resistance to antiplatelet drugs will be more meaningful only when there is a well-established laboratory method to measure the biochemical phenomenon of resistance that has a strong relation to the occurrence of adverse clinical outcomes. Thus, a standardized laboratory method and well-established criteria to define resistance are still unclear [25].

## 5 CLOPIDOGREL RESISTANCE

Activation of platelet $P2Y_{12}$ receptor by ADP is predominantly responsible for GPIIb/IIIa activation and sustained platelet aggregation [7, 8] (Fig. 2). The latter

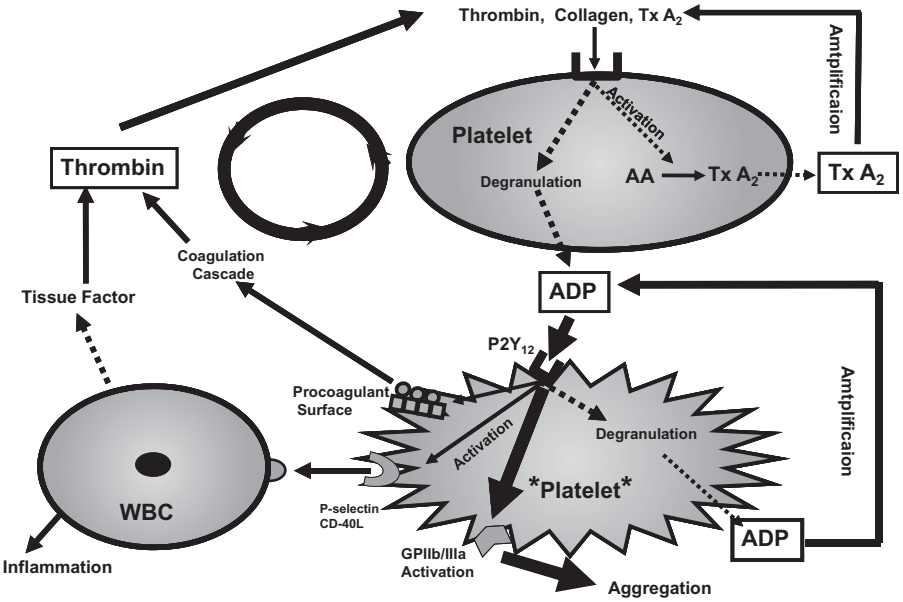


**Figure 2** $P2Y_{12}$- Pivotal Platelet Receptor.

concept is supported by: (1) demonstration of adverse ischemic events associated with high on-treatment ADP-induced platelet reactivity [11], (2) superior clinical outcomes with respect to ischemia observed in patients treated with the more potent $P2Y_{12}$ receptor blockers [26] (3) recent observation of clustering of adverse events in the initial 90 days after stopping clopidogrel among both medically treated and PCI-treated patients with ACS [27]. A detailed meta-analysis of clinical trials comparing clopidogrel with aspirin revealed that clopidogrel is comparatively superior to aspirin in reducing the odds of serious cardiovascular events, including stroke, MI, or vascular death in high-risk patients. Moreover, treatment with clopidogrel and aspirin is associated with superior clinical efficacy compared to aspirin alone [7]. Now dual antiplatelet therapy is a widely adopted treatment strategy and has created a huge market, which was in excess of U.S. $6 billion in 2006. Ticlopidine, the first thienopyridine antiplatelet agent, was found to be superior to aspirin in reducing thrombotic risks during different pathological conditions including peripheral arterial disease. However, neutropenia and other deleterious side effects associated with ticlopidine led to the development of a more potent second-generation thienopyridine, that is, clopidogrel, which was associated with fewer adverse effects [7]. However, subsequent laboratory investigations of clopidogrel responsiveness, based on the ex vivo measurement of ADP-induced platelet aggregation, revealed various limitations: (i) a delayed onset of pharmacodynamic response, (ii) a modest degree of mean platelet inhibition (~ 30–50%) following an a maintenance dose (MD) of 75 mg or a loading dose (LD) of 300 or 600 mg, (iii) wide response variability and unpredictable response with substantial percentage of patients exhibiting nonresponsiveness or resistance, (iv) the potential influence of drug–drug interactions especially with lipophilic statins (e.g., simvastatin, atorvastatin, and lovastatin), and (v) the recent demonstration of an association between clinical adverse events including stent thrombosis and clopidogrel nonresponsiveness.

## 6 PHARMACOKINETIC AND PHARMACODYNAMIC PROPERTIES

Clopidogrel is a prodrug and is rapidly absorbed from the intestine following oral administration. Approximately 85% of the drug is hydrolyzed to an inactive carboxylic acid metabolite (SR26334) by hepatic human carboxylesterase (HCE-1) [28]. The remaining 15% of clopidogrel is first converted to 2-oxo-clopidogrel and subsequently in a multistep pathway to an active thiol metabolite, R-130964 by hepatic cytochrome P450 isoenzymes CYP3A4 and 3A5 and to a smaller extent by CYP2B6, 1A2, 2C9 and 2C19 [29, 30]. The short-lived active metabolite permanently binds to the $P2Y_{12}$ receptor via a disulfide bridge between the reactive thiol group and two cysteine residues (Cys17 and Cys270) present in the extracellular domains of the $P2Y_{12}$ receptor [31]. The half-life of the platelet bound clopidogrel metabolite was found to be 11 days as demonstrated by radioligand studies.

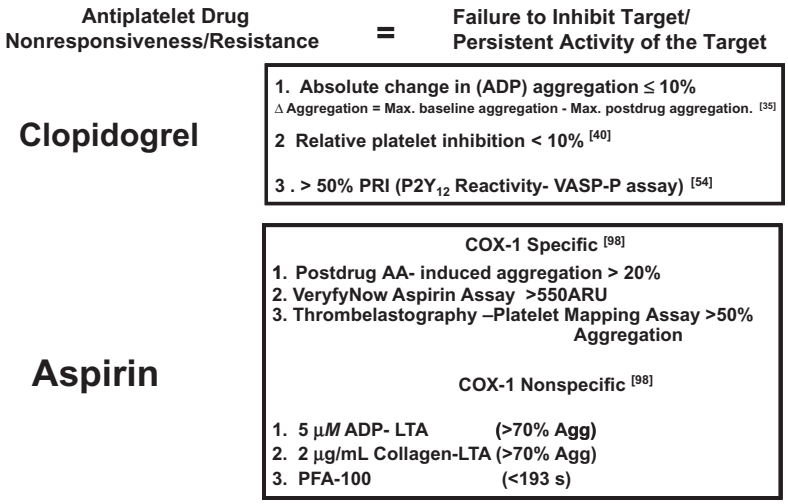


**Figure 3** Definition of Antiplatelet Resistance.

The inactive metabolite is a predominant form present in the systemic circulation following oral administration and its elimination half-life is 8 h after single and repeated administration. The binding of ADP to the $P2Y_{12}$ receptor is permanently inhibited and also the ADP-induced platelet aggregation in a time and dose dependent fashion [4]. There was a linear relationship between clopidogrel dose, absorption, active metabolite generation, and platelet inhibition. Therefore, repeated administration of 75 mg daily dose of clopidogrel leads to a steady-state inhibition of 40–50% in 4–7 days, whereas a loading dose of 300 and 600 mg clopidogrel leads to 40–50% steady-state inhibition at 24 h and 50–60% inhibition at 4–6 h, respectively. Clopidogrel is excreted in urine (50%) and feces (46%) [4, 7]. A wide interindividual variability was observed in clopidogrel absorption and also its hepatic conversion to active metabolite generation [7] (Fig. 3).

Clopidogrel mainly inhibits ADP-induced platelet aggregation. Clopidogrel also inhibits collagen- and thrombin-induced aggregation; however, the inhibitory effect on collagen- and thrombin-induced aggregation can be overcome by increased concentrations of these agonists. These findings suggest that clopidogrel indirectly inhibits the effect of these agonists via the attenuation of ADP-mediated amplification of the platelet response. Clopidogrel does not have any effect on ADP binding to the $P2Y_1$ receptor or the platelet response to $P2Y_1$ activation, which induces shape change and intracellular $Ca^{2+}$ mobilization [32]. Clopidogrel has also been reported to attenuate platelet-leukocyte aggregate formation and the levels of C-reactive protein (CRP), P-selectin, and CD-40L, as well as the rate of thrombin formation [7].

## 7 LABORATORY EVALUATION OF CLOPIDOGREL RESPONSIVENESS

A standardized laboratory method that simulates the in vivo platelet response to antiplatelet therapy is still lacking. Since clopidogrel specifically inhibits one of two ADP receptors, ex vivo measurement of ADP-induced maximum platelet aggregation by LTA has been the most commonly used laboratory method to evaluate clopidogrel responsiveness and is considered the gold standard [7]. (Recently, it was suggested that since antiplatelet drugs (especially clopidogrel), induce platelet disaggregation, the response to clopidogrel would be better demonstrated by measuring late platelet aggregation at 6 min after stimulation with ADP rather than maximum aggregation [33]. However, recent study based on the evaluation of both maximum and final aggregation from 100 consecutive patients undergoing stenting and treated with clopidogrel, indicated that both measurements were equivalent in determining the prevalence of clopidogrel nonresponsiveness [34]. Flow cytometric measurements of the expression of activated GPIIb/IIIa receptor and P-selectin expression after ADP stimulation can also identify clopidogrel nonresponsiveness and correlated with measurements of maximum aggregation stimulated by ADP [13, 35]. In addition, measurements of ADP-induced platelet-fibrin clot strength by whole-blood thrombelastography and the VerifyNow $P2Y_{12}$ receptor assay using ADP as the agonist can also be used to measure clopidogrel responsiveness as point-of-care assays [16, 36]. The platelet function analyzer (PFA)-100 method using cartridges based on collagen-ADP and whole-blood

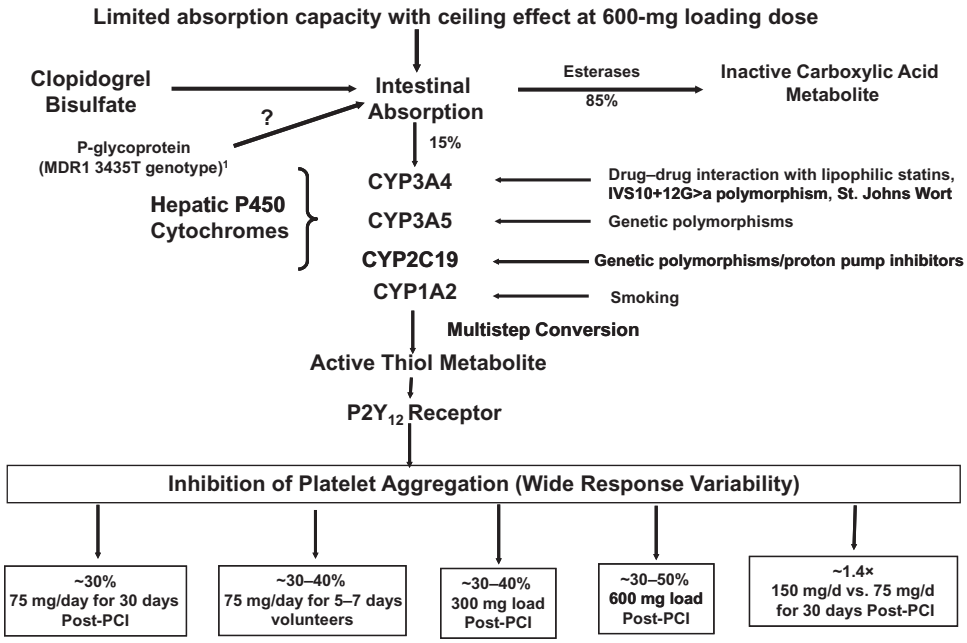


**Figure 4** Mechanism of Clopidogrel Response Variability.

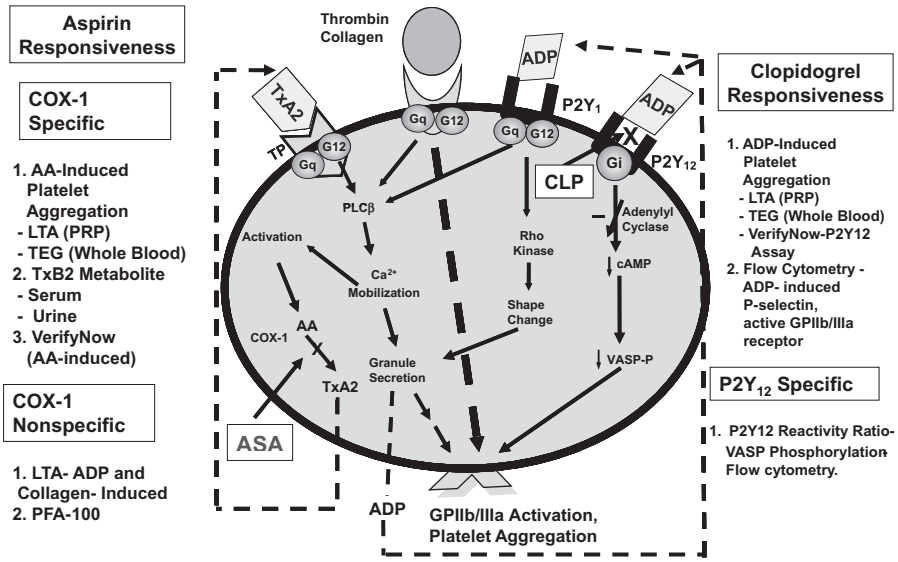


**Figure 5** Laboratory Measurements of Antiplatelet Responsiveness.

aggregometry are associated with inconsistent estimates of platelet reactivity to ADP. The phosphorylation state of vasodilator-stimulated phosphoprotein is a specific intracellular marker of residual $P2Y_{12}$ receptor reactivity in patients treated with clopidogrel and can be measured by flow cytometry. This technique is perhaps the most specific indicator of residual $P2Y_{12}$ activity in patients treated with a $P2Y_{12}$ inhibitor. However, the methodology is labor intensive and requires permeation of the platelet membrane and use of monoclonal antibodies specific for phospohrylated vasodilator-stimulated phosphoprotein [14, 34, 37] (Figs. 4 and 5).

## 8 CLOPIDOGREL RESPONSIVENESS AND TIME OF TREATMENT

Similar to other drugs, response variability and nonresponsiveness to clopidogrel has been demonstrated in patients following coronary stenting. In an early investigation from our center, 96 patients undergoing elective

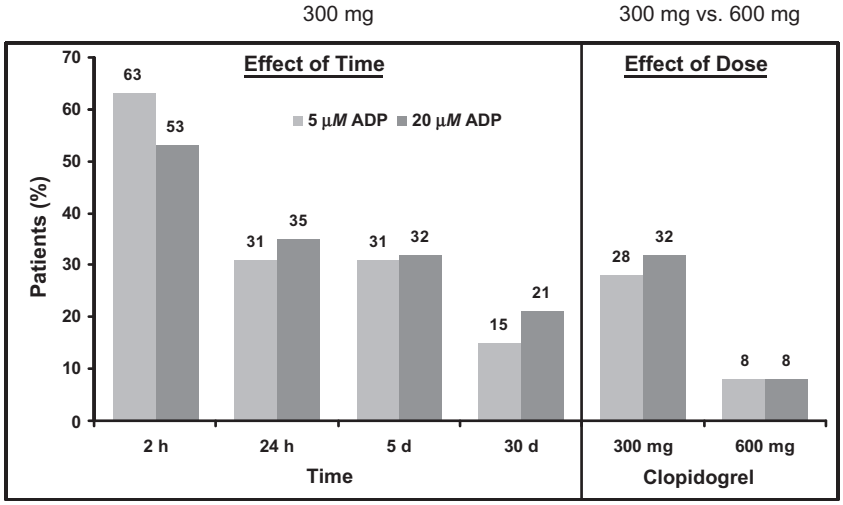


**Figure 6** Clopidogrel Resistance: The Importance of Time and Dose. (Adapted from [35] and [38].)

**TABLE 1 Clopidogrel Resistance**

| Investigators | n | Patients | Clopidogrel Dose (mg Load) | Resistance |
|---|---|---|---|---|
| Jaremo et al. [39] | 18 | PCI | 300 | 28% |
| Gurbel et al. [35] | 92 | PCI | 300 | 31–35% |
| Muller et al. [40] | 105 | PCI | 600 | 5–11% |
| Mobley et al. [41] | 50 | PCI | 300 | 30% |
| Lepantalo et al. [42] | 50 | PCI | 300 | 40% |
| Angiolillo et al. [43] | 48 | PCI | 300 | 44% |
| Matetzky et al. [44] | 60 | PCI | 300 | 25% |
| Dzieweierz et al. [45] | 31 | PCI | 300 | 23% |
| Gurbel et al. [38] | 192 | PCI | 300/600 | 8–32% |
| Lev et al. [46] | 150 | PCI | 300 | 24% |
| Total | 616 | | | 5–44% |

stenting were treated with a 300-mg clopidogrel loading dose in the catheterization laboratory followed by a 75-mg maintenance dose. ADP-induced platelet aggregation and activation-dependent platelet surface marker expression (P-selectin and activated GPIIb/IIIa) were assessed at baseline and serially for 30 days following stenting. Response variability to clopidogrel was demonstrated as measured by all markers, and a certain percentage of patients were found to have no demonstrable antiplatelet effect [35]. In the latter patients, the difference between pre- and posttreatment ADP-induced platelet aggregation was ≤10%. We defined these patients as clopidogrel "resistant" or "nonresponsive" to clopidogrel therapy. A subgroup of resistant patients exhibited platelet aggregation that was greater after stent implantation than at baseline despite clopidogrel therapy. These patients were defined as having heightened platelet reactivity to ADP.

In the latter study, 53–63% of patients were resistant to clopidogrel treatment at 2 h poststenting; ~ 30% were resistant at day 1 and day 5 poststenting; and 13–21% were resistant at day 30 poststenting [35] (Fig. 6). Therefore, clopidogrel "resistance" in this study appeared to be time dependent. Based on the results, it was hypothesized that the occurrence of clopidogrel resistance might be related to the inadequacy of a 300-mg loading dose to provide sufficient active metabolite generation to arrest platelet reactivity in selected patients, and that these resistant patients may be at particular risk for thrombotic complications, including periprocedural infarction and stent thrombosis [35].

Since this initial description, multiple investigators have confirmed the phenomenon of clopidogrel resistance [38–46]. The prevalence of clopidogrel nonresponsiveness has been reported at 5–44%. This wide variation in prevalence is primarily due to dosing and is less related to various definitions, laboratory methods, and the time at which blood samples were evaluated for responsiveness [38–46] (Table 1). As demonstrated in Table 1, the data are markedly concordant. The higher

resistance estimates are present following the 300-mg loading dose and the lower estimates occur after the 600-mg loading dose [38, 40]. Diabetic patients who were on long-term dual antiplatelet therapy had a higher number of clopidogrel nonresponders compared to nondiabetic patients in a recent study [47].

## 9 CLOPIDOGREL RESPONSIVENESS AND EFFECT OF DOSE

Subsequent investigations have unequivocally demonstrated that clopidogrel nonresponsiveness is dependent on dose. In the largest pharmacodynamic study comparing 300 and 600-mg clopidogrel loading doses, treatment with a 600-mg loading dose during elective PCI reduced clopidogrel nonresponsiveness to 8% compared to 28–32% after a 300-mg loading dose (Fig. 6). Moreover, the study demonstrated a narrower response profile following treatment with 600 mg compared to 300 mg clopidogrel [38]. Muller et al. also observed a time- and dose-dependent effect of clopidogrel in patients undergoing stenting [40]. A similar increased responsiveness was also observed in the Intracoronary Stenting and Antithrombotic Regimen: Choice between 3 High Oral Doses for Immediate Clopidogrel Effect (ISAR-CHOICE) study, where a ceiling effect of platelet inhibition was observed with a 600-mg clopidogrel loading dose, whereas a nonsignificant increase in platelet inhibition was observed with a 900-mg loading dose [48]. In addition, recent studies have demonstrated the comparatively better inhibition of ADP-induced platelet aggregation with both single 600-mg loading dose and repeated 150 mg clopidogrel daily [49–51].

## 10 RELATION OF CLOPIDOGREL NONRESPONSIVENESS TO ADVERSE CLINICAL EVENTS

Limited data are available to link clopidogrel nonresponsiveness to the occurrence of thrombotic events, including stent thrombosis and periprocedural adverse clinical events [13–16, 44, 46, 52–59] (Table 2). Matetzky et al. studied clopidogrel responsiveness in patients undergoing stenting for acute ST-elevation myocardial infarction and found that patients who exhibited the highest quartile of ADP-induced aggregation had a 40% probability for a recurrent cardiovascular event within 6 months [44]. In the PREPARE POST-STENTING (Platelet Reactivity in Patients and Recurrent Events POST-STENTING) study, patients suffering a recurrent ischemic event within 6 months of elective stenting had high poststent platelet reactivity to ADP compared to patients without ischemic events despite dual antiplatelet therapy [16]. In the CLEAR PLATELETS (Clopidogrel Loading with Eptifibatide to Arrest PLATELET Reactivity) and CLEAR PLATELETS Ib studies, a 600-mg clopidogrel loading dose used to treat patients undergoing elective stenting was associated with superior early platelet inhibition compared to a 300-mg loading dose, and this inhibition was accompanied by a decrease in release of myocardial necrosis and inflammation markers [13, 52]. Cuisset et al. demonstrated that patients with high posttreatment platelet reactivity had an increased risk of cardiovascular events. More importantly, these patients were resistant to both clopidogrel and aspirin treatment [53]. Similarly Lev et al. demonstrated that occurrence of

**TABLE 2 Studies Linking Clopidogrel Nonresponsiveness to Adverse Clinical Events**

| Study | Results | Clinical Relevance |
|---|---|---|
| Barragan et al. [54] | ↑ $P2Y_{12}$ reactivity ratio (VASP-P levels) | Stent Thrombosis |
| Ajzenberg et al. [55] | ↑ Shear- induced platelet aggregation | Stent Thrombosis |
| Gurbel et al. [14] | ↑ $P2Y_{12}$ reactivity ratio<br>↑ ADP- induced aggregation<br>↑ Stimulated GPIIb/IIIa expression | Stent Thrombosis |
| Matzesky et al. [44] | ↑ ADP-Induced platelet aggregation | Recurrent cardiac events (fourth quartile) |
| Gurbel et al. [13, 52] | ↑ Periprocedural platelet aggregation | Myonecrosis and inflammation marker release |
| Bliden et al. [15] | ↑ Platelet aggregation (pre-PCI) on chronic clopidogrel | 1 yr post-PCI events |
| Lev et al. [46] | Clopidogrel/ aspirin-resistant patients | Post-PCI myonecrosis |
| Cuisset et al. [53] | ↑ Platelet aggregation | 30-day post-PCI events 600 mg-less events |
| Geisler et al. [58] | Clopidogrel low responders | 3 months MACE and death |
| Hochholzer et al. [59] | ↑ Platelet aggregation (upper quartile) | 30 day MACE |

MACE = major adverse cardiovascular events.

creatinine kinase–myocardial band after stenting was more frequent in patients exhibiting aspirin and clopidogrel resistance [46]. Finally, significantly higher recurrent ischemic events within 6 months of the procedure were observed in patients who were on chronic clopidogrel therapy undergoing elective coronary stenting and had higher preprocedure ADP-induced platelet aggregation [15]. All these findings strongly suggest that a high platelet reactivity despite currently recommended antiplatelet therapy is a risk factor for ischemia in patients undergoing PCI.

Based on the analysis of flow cytometric measurements of intracellular vasodilator-stimulated phosphoprotein (VASP) phosphorylation levels, a specific intracellular marker of clopidogrel-induced $P2Y_{12}$ receptor inhibition, nonresponsiveness to clopidogrel treatment has been suggested as a risk factor for the occurrence of stent thrombosis [14, 54]. In the recent CREST (Clopidogrel Effect on Platelet Reactivity in Patients with Stent Thrombosis) study, elevated levels of ADP-induced platelet aggregation, ADP-stimulated expression of active GPIIb/IIIa expression and the $P2Y_{12}$ reactivity ratio measured by VASP phosphorylation were observed in patients with stent thrombosis compared to patients without stent thrombosis, indicating inadequate inhibition of $P2Y_{12}$ receptor [14]. Other investigators have reported that high ex vivo shear-induced platelet aggregation or prevalence of clopidogrel resistance despite dual antiplatelet therapy may be a risk factor for stent thrombosis [55–57].

## 11 MECHANISM OF CLOPIDOGREL RESISTANCE

The mechanisms responsible for clopidogrel response variability and resistance are incompletely defined. Differences in intestinal absorption, hepatic conversion to the active metabolite through cytochrome 3A4 (CYP3A4) activity, and platelet receptor polymorphisms have been suggested [5]. Only up to 30–50% inhibition of ex vivo ADP-induced platelet aggregation was demonstrated following a repeated daily dose of 75 mg in normal volunteers or loading doses of 300 or 600 mg in patients undergoing PCI [13, 35, 38]. This level of inhibition indicated an incomplete $P2Y_{12}$ receptor blockade and suboptimal inhibition of ADP-induced platelet aggregation. The repeated demonstrations that a high loading dose of 600 mg clopidogrel is associated with increased inhibition of ex vivo ADP-induced platelet aggregation in patients undergoing PCI and a decreased prevalence of nonresponders supports insufficient active metabolite generation as a major factor in clopidogrel resistance [13, 38, 40, 48].

Recent studies involving the measurement of hepatic cytochrome (CYP) P450 activity suggest that individual variations in the activity of this enzyme play a major role. In a landmark investigation, Lau et al. demonstrated that pharmacologic stimulation of CYP3A4 activity by rifampin enhances the inhibitory effect of clopidogrel, whereas agents that compete with clopidogrel for CYP3A4 activity (e.g., erythromycin) attenuate the antiplatelet effect of clopidogrel [60]. In addition, cigarette smoking, possibly by the effect of nicotine as an inducer of CYP1A2 activity and St. Johns Wort, an inducer of CYP3A4, have also been associated with enhanced clopidogrel responsiveness [61, 62]. However, coadministration of lipophilic statins (that compete with clopidogrel for CYP3A4) such as atorvastatin or simvastatin have been shown to attenuate the platelet inhibitory effect of clopidogrel in pharmacodynamic studies [63].

Controversy abounds in the literature with respect to clopidogrel–lipid soluble statin interactions. The retrospective and prospective studies reporting no relation of statin use to clopidogrel effect may be explained by (1) the low doses of atorvastatin used in the studies and (2) under power of the sample size of the non-CYPP3A4-metabolized statins when compared to the CYPP3A4-metabolized statins to demonstrate statistical significance. There is evidence that this drug–drug interaction may indeed be associated with increased adverse cardiac events. Although not statistically significant because of sample size, Weinberger et al. reported 26% increased odds ratio for mortality in patients on clopidogrel for an acute coronary syndrome prescribed atorvastatin versus other statin therapies [64]. Although the lipophilic statin–clopidogrel interaction has not been definitively demonstrated in retrospective analyses of large-scale clinical trials, a trend toward a better clinical outcome in patients coadministered clopidogrel and non-CYP3A4-metabolized statins compared to lipophilic statins has been reported [65] (Fig. 3).

Additional information supporting the pivotal role of CYPP450 in generating the active metabolite of clopidogrel was demonstrated in a small study involving human volunteers where the effects of the CYP3A4 inhibitor, ketoconazole, on the pharmacokinetics and the ex vivo platelet inhibitory effects of prasugrel and clopidogrel were investigated. Prasugrel is also a thienopyridine that requires conversion to an active metabolite by a hepatic cytochrome. In this study, ketoconazole coadministration did not have any effect on prasugrel-active metabolite generation or prasugrel-induced platelet inhibition, whereas clopidogrel-induced platelet inhibition was reduced following a loading dose as well as after a maintenance dose [66]. The latter effect on clopidogrel-induced platelet inhibition was accompanied by lesser active metabolite generation. Similarly,

in another study, prasugrel treatment was associated with superior active metabolite generation and platelet inhibition together with a lower incidence of nonresponsiveness compared to clopidogrel treatment [67].

In recent studies, the influence of CYP3A5 and CYP2C19 isoenzymes on clopidogrel metabolic activation and responsiveness has been demonstrated [68, 69]. Suh et al. demonstrated higher clopidogrel responsiveness among subjects with the CYP3A5 expressor genotype than subjects with the nonexpressor genotype. Moreover, worse outcomes were seen in patients undergoing stent implantation with the nonexpressor genotype following treatment with clopidogrel than in patients with the CYP3A5 expressor genotype [68]. Similarly, Hulto et al. [69] and Brandt et al. [69a] independently demonstrated the influence of the CYP2C19 genotype on clopidogrel responsiveness in healthy volunteers. To further support the drug–drug interaction at the CYP2C19 level, in this issue Gilard et al. [70] demonstrated that coadministration of proton pump inhibitors (omeprazole) with clopidogrel in patients undergoing stenting was associated with significantly less platelet inhibition. They measured platelet inhibition by a $P2Y_{12}$ receptor-specific vasodilator-stimulated phosphoprotein phosphorylation (VASP-P) assay at 7 days posttreatment rather than widely used ADP-induced platelet aggregation measured at 24 h posttreatment. Since, proton pump inhibitors (PPIs) are also metabolized by CYP2C19, it was logical to study the influence of PPIs on clopidogrel-induced effect on platelets. In this study, authors reported 4.31 times more chances of being a clopidogrel bad responder when patients were treated with omeprazole in addition to clopidogrel than clopidogrel alone. This study gains significant clinical interest because possible combination therapy of PPIs coadministered with clopidogrel to attenuate gastrointestinal bleeding [70].

Suboptimal platelet response to clopidogrel may be due to an increased number of platelet $P2Y_{12}$ receptors, or polymorphism of platelet receptors. Genetic polymorphisms of platelet GPIIb/IIIa, GPIa/IIa, or $P2Y_{12}$ receptors have been reported to affect platelet function and may influence clopidogrel response variability [71–73]. Recently, it was reported that an increased percentage of patients with peripheral arterial disease have the $P2Y_{12}$ receptor H2 haplotype [72]. However, in another study, the relation of this haplotype to clopidogrel responsiveness could not be demonstrated [73]. Since the relation of genetic polymorphisms to clopidogrel responsiveness is inconclusive, further studies are required to establish a correlation between receptor polymorphisms and clopidogrel nonresponsiveness.

It has been shown that patients with diabetes exhibit platelet activation and increased reactivity to agonists. The heightened platelet reactivity may be related to the increased prevalence of nonresponders and occurrence of ischemic events reported in patients with diabetes [74, 75]. It has also been reported that patients with a high body mass index (BMI) exhibited a suboptimal platelet response with the standard 300-mg loading dose [76].

All of the above data strongly support the importance of insufficient metabolite generation secondary to limitations in the intestinal absorption, drug–drug interaction at CYP3A4, or genetic polymorphisms of CYP isoenzymes as the primary explanations for resistance rather than genetic polymorphisms of platelet receptors or intracellular signaling mechanisms. The latter mechanisms may be relevant in those patients who remain resistant and with high platelet reactivity to ADP even after high dosing strategies.

## 12 MANAGEMENT OF CLOPIDOGREL RESISTANCE

### 12.1 Higher Doses

In recent clinical studies of patients undergoing stenting, a 600-mg clopidogrel loading dose was associated with a higher level of platelet inhibition, lower mean posttreatment reactivity to ADP, and a lower incidence of nonresponsiveness when compared to a 300-mg dose. Moreover, a 600-mg clopidogrel loading dose was associated with a narrower response profile [38, 40]. Kastrati et al. found that patients achieved additional platelet inhibition when a 75-mg/day clopidogrel maintenance dose was followed by an additional 600-mg loading dose [77]. In the CLEAR PLATELETS study a 600-mg loading dose was associated with a superior pharmacodynamic antiplatelet profile compared to a 300-mg clopidogrel loading dose [13, 52]. In the recent ISAR-CHOICE study, there was a ceiling effect in unchanged clopidogrel and clopidogrel metabolite levels and platelet inhibition with the 600-mg loading dose, and no significant additional effect was seen with the 900-mg loading dose. Based on the pharmacokinetic profile of the free drug and metabolites the investigators concluded that intestinal absorption was the major factor explaining response variability [48]. Recent studies have demonstrated the comparatively better inhibition of ADP-induced platelet aggregation with both single 600-mg loading dose and repeated 150-mg daily clopidogrel [49–51].

Thus, higher loading doses may be considered for selected patients exhibiting high platelet reactivity to ADP. However, the superiority of a high-dose regimen in reducing ischemic events and the associated risk profile compared to a standard dose has yet to be established in large-scale clinical trials. Despite these limitations, the current American College of Cardiology

(ACC)/American Heart Association (AHA) guidelines for percutaneous coronary intervention (PCI) provide a class IIa recommendation that "a regimen of greater than 300 mg is reasonable to achieve higher levels of antiplatelet activity more rapidly [18, p. e277] ." Finally, the ACC/AHA guidelines provide a class IIb recommendation that "in patients in whom subacute thrombosis may be catastrophic or lethal . . . platelet aggregation studies may be considered and the dose of clopidogrel increased to 150 mg per day if less than 50% inhibition of platelet aggregation is demonstrated" [18, p. e277]. The latter guidelines however, do not specify the methodology that should be used to assess inhibition. Moreover, there are very limited clinical data to support the cutpoint of 50% inhibition [18, 38]. In the CLEAR PLATELETS study we observed periprocedural myocardial infarction only in those patients with 5 μ*M* ADP-induced aggregation >50% [13]. In the CREST study the cutpoint for stent thrombosis was 20 μ*M* ADP-induced aggregation >40% [14].

### 12.2 New $P2Y_{12}$ Receptor Antagonists

New $P2Y_{12}$ receptor antagonists are currently undergoing investigation. AZD 6140 (Astra-Zeneca) and cangrelor (Medicines Company) are reversible, direct, and potent inhibitors of the $P2Y_{12}$ receptor [78–80]. AZD 6140 is an oral inhibitor whereas cangrelor is administered parentarally. Both of these agents exhibit more consistent and greater platelet inhibition compared to clopidogrel. The short onset and offset of action make these agents appealing adjunctive antiplatelet agents during PCI when maximum and rapid platelet inhibition of ADP-induced aggregation is desired [78–80]. Prasugrel is an irreversible inhibitor of $P2Y_{12}$ and, similar to clopidogrel, is a prodrug that requires metabolic activation [81–83]. In the JUMBO-TIMI-26 trial, prasugrel treatment was associated with a similar primary endpoint of significant bleeding compared to standard clopidogrel regimen (1.7 vs. 1.2), and the bleeding events were the same for all doses of prasugrel [81]. The pharmacodynamic profile of prasugrel is superior to clopidogrel and is associated with lesser incidences of nonresponsiveness, however, it is associated with increased bleeding events [82, 83]. All three of the above agents will undergo study in phase III clinical trials. Prasugrel is currently being compared to clopidogrel in an ACS trial in patients undergoing PCI.

## CONCLUSIONS

Clopidogrel use has increased over the last few years following its effectiveness together with aspirin in significantly reducing adverse events in large-scale clinical trials. At the same time, based on the laboratory evaluation of platelet response, wide response variability and nonresponsiveness in selected patients are also present. In the recent small studies, heightened platelet reactivity or clopidogrel nonresponsiveness in patients who were on clopidogrel treatment is associated with adverse thrombotic events including stent thrombosis. The primary reason is attributed to the suboptimal generation of active metabolite secondary to limitation in intestinal absorption, and drug-drug interaction at CYP3A4 and genetic polymorphism of hepatic cytochrome P450 isoenzymes. Use of higher loading or maintenance doses of clopidogrel or new and more potent $P2Y_{12}$ receptor blockers are potential alternative strategies. In addition, the treatment of dual antiplatelet therapies may be confined to the pharmacologic management of patients at high risk for arterial thrombotic events but not as a primary prevention strategy or as an alternative to anticoagulants.

## 13 ASPIRIN RESISTANCE

Aspirin (acetyl salicylic acid) was synthesized by Felix Hoffman in 1897 at the Bayer Pharmacological research Laboratories as a palatable alternative to widely used bitter taste painkiller salicylic acid. Sir John Vane in 1971, discovered the aspirin's antithrombotic potential [84]. Since then, it is consider as a wonder drug and is the most economical and effective antiplatelet drug prescribed for the treatment of cardiovascular and cerebrovascular diseases. The extensive analysis of primary and secondary clinical trials have indicated that aspirin treatment is associated with a 20–40% reduction in cardiovascular adverse events [17]. Based on the results of primary and secondary prevention trials, the U.S. Preventive Services Task Force (USPSTF) recommends that low-dose aspirin (75 mg/day) should be administered in patients whose 5-year risk of a first cardiovascular (CV) event is ≥3% and whose 10-year risk is ≥6%. The American Heart Association (AHA) recommends that 75–160 mg/day be administered to patients whose 10-year risk of a first CV event is ≥10%. Overall, the benefits of low-dose aspirin therapy for prevention of MI outweigh the risks of gastrointestinal and intracerebral bleeding in the latter risk groups [85]. A higher dose of aspirin may be also required during the conditions of high platelet turnover such a coronary bypass surgery. Since aspirin is ~150-fold more potent in inhibiting platelet COX-1 (antithrombotic effect) than COX-2 present in endothelial and other cells (anti-inflammatory effect), it is more effective as an antithrombotic agent at a 75–325 mg per day dose [4]. Despite these benefits, it has been reported that only slightly more than half of eligible patients receive the proper aspirin treatment.

Implementation of appropriate and widespread use of aspirin should be the focus of future primary and secondary prevention trials. Finally, whenever rapid and complete inhibition of thromboxane-$A_2$- induced platelet aggregation is desired, such as in the setting of acute MI, unstable angina, or PCI, a 150- to 325-mg aspirin loading dose is favored [18].

### 13.1 Pharmacokinetics and Pharmacodynamics

Aspirin is readily absorbed in the upper gastrointestinal tract with enteric-coated aspirin requires a longer time to absorb. Peak plasma levels of aspirin were observed within 40 min and platelet inhibition and also prolongation can be observed within an hour. The bioavailability of aspirin is 50%. Aspirin is mainly metabolized by intestinal and liver HCE-2 to acetic acetyl moiety and salicylate moiety [28]. Aspirin has a half-life of 15 min whereas salicylate has a half-life of 3–6 h. Platelet inhibition was observed even before the appearance of aspirin in systemic circulation with low-dose controlled-release aspirin indicated that acetylation of the platelet COX-1 takes place in the portal or prehepatic circulation. Therefore, thromboxane-related platelet inhibition or thromboxane metabolite measurement in serum is considered as a surrogate measure of aspirin bioavailability rather than measurement of salicylate levels in the venous blood. The antiplatelet effect of aspirin is primarily due to irreversible acetylation of a serine residue (Ser530) in COX-1 in platelets, which prevents the binding of arachidonic acid to the catalytic site. Subsequent generation of $TxA_2$ and $TxA_2$-induced platelet aggregation are inhibited for the life of platelets. Since life span of platelets in circulation is 10 days, it takes 8–10 days to achieve pretreatment level of platelet aggregation.

It was also suggested that normal hemostasis may be achieved with as little of 20% of the platelet having normal COX-1 activity and thromboxane synthesis. In this line, it was demonstrated that more than 95% inhibition of thromboxane synthesis is required to observe any clinical efficacy of aspirin treatment [4].

### 13.2 Laboratory Evaluation of Aspirin Resistance (Figure 5)

A reliable and specific laboratory method to identify aspirin resistance has not yet been uniformly accepted by investigators. Aspirin resistance has been regarded as the failure of aspirin to inhibit COX-1, the failure of aspirin to inhibit platelet function after stimulation by various agonists ("biochemical aspirin resistance"), or the failure to protect from cardiovascular events ("clinical aspirin resistance" or more specifically "treatment failure"). Based on different ex vivo methods and criteria to define aspirin resistance, various studies have reported a wide variability in the occurrence of aspirin resistance (<1–55%) [86–97] (Table 3). Laboratory methods including point-of-care methods using agonists such as ADP-, epinephrine-, or collagen-induced platelet aggregation do not solely indicate COX-1 activity and thus are fundamentally flawed in specifically measuring the platelet response to aspirin (Fig. 5). More specific methods that indicate residual COX-1 activity include arachidonic-acid-induced platelet aggregation and, more precisely, the measurement of the stable metabolite of $TxA_2$ in serum during aspirin therapy. Measurement of urinary thromboxane metabolite levels may not be indicative of platelet COX-1 inhibition exclusively.

### 13.3 Aspirin Resistance Is Assay Specific

Platelet aspirin resistance was found to be uncommon in complaint patients treated with high-dose aspirin (325 mg per day) when measured by cyclooxygenase-1 (COX-1) specific assays such as using arachidonic acid (AA)-induced light transmittance aggregometry (LTA) and thrombelastography [96]. In another study of patients with a history of myocardial infarction, 9% of patients were found to be noncompliant with aspirin therapy and only one patient was resistant to 325 mg aspirin therapy as measured by AA-induced LTA [97]. In a recent prospective, randomized, double-blind, double crossover investigation studying aspirin dosing using multiple COX-1 specific and nonspecific assays in patients with stable coronary artery disease ($n = 120$), it was found that aspirin resistance was rare (1–6%) using COX-1-specific methods (AA-induced LTA, VerifyNow Aspirin assay with AA cartridges, and AA-induced aggregation measured by thromelastography [TEG]) at all doses of aspirin compared to COX-1 nonspecific methods (1–27%). Moreover, a dose-dependent platelet response to aspirin treatment was observed using methods where collagen was used as an agonist such as collagen-induced LTA and PFA-100. The latter occurred in the presence of near complete inhibition of COX-1 enzyme activity as measured by AA-induced platelet aggregation, indicating that aspirin may have non-COX-1-mediated dose-dependent effects in platelets [98]. More importantly, these dose-dependent effects were more pronounced in patients with diabetes where treatment with higher doses was found to be effective in inhibiting platelet function [99]. Similar results reporting a low prevalence of aspirin resistance using COX-1 specific assays have been reported [100]. At the same time, further studies have found a correlation between

**TABLE 3 Studies Demonstrating Aspirin Resistance**

| Investigators | *n* | | ASA Dose (mg/day) | Time | Method | Criteria for Aspirin Resistance (AR) | % AR |
|---|---|---|---|---|---|---|---|
| Hurlen et al. [86] | 143 | AMI | 75–160 | 2–24 h | Platelet aggregation ratio (PAR) | PAR < 0.82<br>PAR < 0.82 after additional ASA | 9.8<br>1.4 |
| Buchanan et al. [87] | 40 | CABG | 325 | | Bleeding time platelet $TxA_2$, 12-HETE and platelet adhesion | No prolongation of bleeding time above baseline | 43 |
| Buchanan et al. [88] | 287 | CABG | 325 | 24 h<br>2 years follow-up | Bleeding time | No prolongation of bleeding time above baseline | 54.7 |
| Grotemeyer et al. [89] | 180 | Stroke | 1500 | 1 year follow-up | Platelet reactivity (PR) | Normal PR index (>1.25) at 2 or 12 h | 33.3 |
| Gum et al. [90, 91] | 325 | Stable CAD | 325 | ≥ 7 days<br>2 years follow-up | LTA- AA and ADP<br>PFA- 100 collagen/ADP or collagen/EPI | >70% ADP-induced aggregation + AA (0.5 mg/mL) induced >20% after ASA<br>Normal (<193) collagen/EPI closure time after ASA | 5.5<br>9.5 |
| Eikelboom et al. [92] (HOPE Study) | 488 | | 75–325 | 5 years follow-up | Urinary 11-dehydro $TxA_2$ | Elevated urinary 11-dehydro $TxB_2$—upper quartile | 25 |
| Wang et al. [93] | 422 | stable CAD | 81–325 | ≥7 days | RPFA | ARU > 550 | 23 |
| Chen et al. [94] | 151 | Non-urgent PCI | 81–325 | ≥7 days | RPFA<br>Myonecrosis (CK-MB +TnI) | ARU > 550 | 19.2 |
| Gonzalez-Conejero et al. [95] | 24 | HS | 100 and 500 | 2 weeks | PA—1 mM AA or 10 ug/mL collagen, 11-dehydro $TxA_2$.<br>PFA-100, genotyping. | <300 s closer time | 33.3 (100 mg)<br>0 (500 mg) |
| Tantry et al. [96] | 223 | PCI | 325 | Long term | LTA -1 mM AA<br>TEG- 1 mM AA | >20% aggregation<br>>50% | <1.0<br><1.0 |

ASA = aspirin; ARU = aspirin resistance units; RPF = rapid platelet function analyzer; AA = arachidonic acid; LTA = light transmittance aggregometry; TEG = thrombelastography; PCI = percutaneous coronary intervention; CAD = coronary artery disease; PA = platelet aggregation; CK-MB = creatinine kinase-MB; Txb2 = thromboxane B2; EPi = epinephrine; CABG = coronary artery bypass grafting; HS = healthy subjects; HETE = hydroxyeicosatetraenoic acid.

adverse clinical events and aspirin resistance measured by PFA-100 [100–106] (Table 4). Thus, taken together, all of these studies indicate that noncompliance, various assessment methods, and underdosing may all be important factors responsible for the reported variability in aspirin resistance estimates in clinical studies.

It has also been demonstrated in recent studies that aspirin resistance may be associated with clopidogrel resistance. Moreover, aspirin-resistant patients have been shown to exhibit high platelet reactivity induced by various agonists such as collagen and ADP in addition to arachidonic acid [46, 107]. Therefore, aspirin resistance as measured in these studies may mark a

**TABLE 4 Clinical Relevance of Aspirin Resistance**

| Investigators | *n* | Method | Clinical Outcome |
|---|---|---|---|
| Grotmeyer et al. [89] | 180 post-stroke | Platelet aggregates | 10× increase in vascular events |
| Mueller et al. [116] | 100 PCD | Platelet aggregation (whole blood) | 87% increase in incidence of reocclusion |
| Eikelboom et al. [92] | 976 HOPE trial | Urinary TxB2 | Increase in MI/stroke/death with increase in TxB2 |
| Gum et al. [91] | 325 stable CAD | LTA | 3.12× increase in MI/stroke/death |
| Tantry et al [96] | 223 elective | LTA and TEG- 2 mm AA | ASA-R 0.4%, 1 pt ST |
| Chen et al. [94] | 151 elective | RPFA- propyl gallate | 2.9× increase in myonecrosis |
| Lev et al. [46] | 150 PCI | RPFA (propyl gallate) LTA-1.6 mM AA | Elevated CK-MB |
| Marcucci et al. [101] | 146 MI/PCI | PFA-100 (coll/EPI) | ↑ 1 -year MACE |
| Gianetti et al. [102] | 175 (SA/AMI) | PFA-100 (Coll-EPI and ADP) | 6 months outcome-HR = 8.5- |
| Gulmez et al. [103] | 114 (PCI) | PFA -100 (Coll-EPI) | ASA-R vs CK-MB-No relation |
| Foussas et al. [104] | 612 (SA/UA) | PFA-100 (COLL-EPI) | 1 year CV death and rehosp-HR = 2.7 |
| Malek et al. [105] | 91 (PCI) | PFA-100 (COLL/EPI and ADP) | Unfavorable in-hospital outcome |
| Borna et al. [106] | 135 (ACS) | PFA-100 (COLL/EPI and ADP) | ASA-R associated with STEMI |

CAD = coronary artery disease; CVA = cerebrovascular incident; HOPE = heart outcomes prevention evaluation; LTA = light transmittance aggregometry; MI = myocardial infarction; PCI = percutaneous coronary intervention; PVD = peripheral vascular disease; RPFA = rapid platelet function analyzer; Tx = thromboxane; SA = stable angina; AMI = acute myocardial infarction; COLL/EPI = collagen/epinephrine; ADP = adenosine diphosphate; ACS = acute coronary syndrome; ASA-R = aspirin resistance; HR = hazard ratio; STEMI = ST-segment elevation myocardial infarcton; UA = unstable angina; CV = cardiovascular; MACE = major cardiovascular adverse events; ST = stent thrombosis..

platelet reactivity phenotype indicative of high risk for ischemic events.

In addition to the mechanisms described above to explain mechanism of antiplatelet resistance, role of other cardiovascular risk factors that are known to augment platelet reactivity such as smoking, diabetes, hyperlipidemia and hyperglycemia, nonadhenrence to the therapy and concomitant use of other nonsteroidal anti-inflammatory drugs especially observed with aspirin may affect the antiplatelet responsiveness, prevalence of antiplatelet resistance, and finally adverse clinical outcome [4, 108, 109].

### 13.4 Variability of Aspirin Efficacy in Reducing Ischemic Events

The benefit of aspirin treatment may be dependent on the type of clinical setting. Although a repeated dose of 30 mg per day is reportedly sufficient to fully inhibit COX-1 in platelets, a 75- to 150-mg daily dose for long-term prevention and a 150- to 300-mg per day dose to provide rapid and complete inhibition in high-risk cardiovascular patients is recommended [4, 17, 110, 111]. However, even higher doses have been recommended for patients with cerebrovascular disease to prevent ischemic complications [4, 17]. The benefit of aspirin therapy is more pronounced (up to 25% reduction in mortality) in high-risk patients, especially patients with previous myocardial infarction [110]. A similar benefit was reported in the Antithrombotic Trialists' Collaboration (ATT) analysis, whereas the Physician's Health Study demonstrated a 44% reduction in the incidence of first MI [17, 111]. In patients with acute ischemic stroke or in patients undergoing coronary bypass graft surgery, the aspirin effect is less pronounced, whereas a marked risk reduction has been reported in patients with advanced coronary artery disease [112]. The role of aspirin in the prevention of atherothrombotic complications in women and patients with diabetes remains unclear [113–115].

These findings suggest that different activators (thrombin, collagen, ADP, and thromboxane) may influence platelet aggregation depending on the type of disease encountered. Thus, aspirin may be less effective in some diseases where activation of the platelet is not dependent on thromboxane synthesis. This observation may be relevant to the occurrence of ischemic events despite aspirin treatment that is otherwise described as clinical aspirin resistance.

### 13.5 Clinical Relevance of Aspirin Resistance (Table 4)

In an earlier study, Grotemeyer et al. found that patients exhibiting aspirin resistance as measured by elevated

platelet aggregates ratio had a 10-fold increase in the risk of recurrent vascular events as compared to aspirin-sensitive patients (44 vs. 4.4%, $p < 0.001$) [9]. Using whole-blood aggregometry to evaluate aspirin resistance, Mueller et al. found an 87% increase in the incidence of reocclusion in "aspirin-resistant" patients who underwent balloon angioplasty and were treated with 100 mg aspirin daily for 18 months [116]. Using a more COX-1 specific method measuring urinary thromboxane (Tx), metabolite levels among patients enrolled in Heart Outcomes Prevention Evaluation (HOPE) trial for aspirin responsiveness, Eikelboom et al. found that the risk of MI, stroke, or cardiovascular death was greater in the highest quartile of urinary $TxB_2$ levels. These data suggest that suboptimal inhibition of COX-1 or nonresponsiveness to aspirin therapy is a major cardiovascular risk factor [92]. Gum et al. found a 5% incidence of aspirin resistance based on the criteria of ≥20% AA-induced aggregation and ≥70% ADP-induced aggregation in patients with stable cardiovascular disease treated with 325 mg aspirin daily for up to 2.5 years. Aspirin resistance was associated with a significant increase in the composite endpoint of death, MI, or stroke [90, 91]. Chen et al., using the Accumetrics point-of-care device with cationic propyl gallate as the agonist, found that despite treatment with clopidogrel and heparin, aspirin resistance as defined by this assay was associated with ~2.9-fold increase in the occurrence of myonecrosis following PCI [94]. Recent studies correlated adverse clinical outcomes in patients treated with aspirin in selected patients who were exhibiting aspirin resistance as indicated by PFA-100 methods [101–104].

### 13.6 Therapeutic Intervention of Aspirin Resistance

Initial studies have suggested that inhibition of COX-1 by aspirin can reach saturable levels (>95% inhibition) with a 100-mg daily dose [117]. In a double-blind, double-crossover study of outpatients with stable coronary artery disease, we investigated the antiplatelet effects of 81, 162, and 325 mg daily aspirin [98]. The antiplatelet effects of aspirin were dose independent when assessed by methods that directly indicate COX-1 activity. A higher dose (>75–325 mg/day) is not clearly associated with an increased clinical benefit but rather may be accompanied by an elevated risk of bleeding [4, 17, 110, 111]. However, some investigators have suggested that higher dosages may be more clinically effective [118]. Recent studies have suggested that 325 mg aspirin per day may be associated with lesser aspirin nonresponsiveness as measured by COX-1 specific methods [96, 97]. Thus, a higher dose and strict compliance to therapy may effectively overcome the occurrence of aspirin resistance in selected patients.

## 14 CONCLUSIONS

Reports of high incidences of aspirin resistance may be due to laboratory measurements based on nonspecific methods that do not isolate the response of platelet COX-1 to aspirin or an inadequate dose in selected patients. Noncompliance may also be an important factor related to aspirin resistance. Clopidogrel nonresponsiveness is a consistent phenomenon observed in research studies conducted at multiple sites utilizing various $P2Y_{12}$ receptor-specific assays. Data from small studies support that patients with high ex vivo platelet reactivity to ADP during and after percutaneous intervention may be at greatest risk for subsequent ischemic events.

However, at this time there are no uniformly established methods to quantify ex vivo platelet reactivity after clopidogrel and aspirin treatment or the extent of platelet inhibition by clopidogrel and aspirin. Therefore, specific treatment recommendations for patients exhibiting high platelet reactivity during clopidogrel or aspirin therapy or who have poor platelet inhibition by clopidogrel or aspirin are not established. A higher aspirin dose (≥ 325 mg/day) and strict compliance to therapy may overcome the occurrence of aspirin resistance in selected patients. A clopidogrel maintenance dose of 150 mg may be considered in patients exhibiting clopidogrel nonresponsiveness. In the near future new $P2Y_{12}$ receptor blockers will likely overcome the limitations of clopidogrel. The bleeding issue associated with prolonged dual antiplatelet therapy remains a major concern that has been highlighted by the recent results from the CHARISMA trial [119]. At this time, it is uncertain whether a threshold of platelet reactivity exists, beyond which ischemic event occurrence rises. Similarly, the threshold platelet reactivity for bleeding is also unknown.

In the future, routine measurement of platelet function in patients with cardiovascular disease may become the standard of care. The clinical utility of routine measurements of platelet function in patients treated with antiplatelet agents, akin to low-density lipoprotein (LDL) measurements in patients treated with statins, will be determined by definitive large-scale investigations addressing the link between insufficient platelet inhibition and cardiovascular thrombotic risk. Personalized antithrombotic treatment strategies may be determined by ex vivo measurements that identify critical pathways influencing thrombotic risk in the individual patient. Currently, treatment of patients who

exhibit heightened reactivity of these pathways with specific inhibitors is being investigated in small pilot trials.

## REFERENCES


1. Gurbel PA, Tantry US (2006). The relationship of platelet reactivity to the occurrence of post-stenting ischemic events: Emergence of a new cardiovascular risk factor. *Rev Cardiovasc Med 7*(Suppl 4), S20–28.
2. von Hundelshausen P, Weber C (2007). Platelets as immune cells: Bridging inflammation and cardiovascular disease. *Circ Res 100*, 27–40.
3. Lindemann S, Krämer B, Seizer P, Gawaz M (2007). Platelets, inflammation and atherosclerosis. *J Thromb Haemost 5*(Suppl 1), 203–211.
4. Patrono C, Coller B, FitzGerald GA, Hirsh J, Roth G (2004). Platelet-active drugs: The relationships among dose, effectiveness, and side effects: The Seventh ACCP Conference on Antithrombotic and Thrombolytic Therapy. *Chest 126*(3 Suppl), 234S–264S.
5. Tantry US, Bliden KP, Gurbel PA (2005). Resistance to antiplatelet drugs: Current status and future research. Review. *Expert Opin Pharmacother 6*, 2027–2045.
6. Brinker AD, Swartz L (2006). Growth in clopidogrel-aspirin combination therapy. *Ann Pharmacother 40*, 1212–1213.
7. Tantry US, Etherington E, Bliden KP, Gurbel PA (2006). Antiplatelet therapies; current strategies and future trends. *Future Cardiol 2*, 343–366.
8. Jackson SP (2007). The growing complexity of platelet aggregation. *Blood 109*, 5087–5095.
9. Gurbel PA, Bliden KP, Hayes KM, Tantry U (2004). Platelet activation in myocardial ischemic syndromes. *Expert Rev Cardiovasc Ther 2*, 535–545.
10. Gurbel PA, Kereiakes DJ, Dalesandro MR, Bahr RD, O'Connor CM, Serebruany VL (2000). Role of soluble and platelet-bound P-selectin in discriminating cardiac from noncardiac chest pain at presentation in the emergency department. *Am Heart J 139*, 320–328.
11. Gurbel PA, Becker RC, Mann KG, Steinhubl SR, Michelson AD (2007). Platelet function monitoring in patients with coronary artery disease. *J Am Coll Cardiol 50*, 1822–1834.
12. Matetzky S, Shenkman B, Guetta V, Shechter M, Bienart R, Goldenberg I, Novikov I, Pres H, Savion N, Varon D, Hod H (2004). Clopidogrel resistance is associated with increased risk of recurrent atherothrombotic events in patients with acute myocardial infarction. *Circulation 109*, 3171–3175.
13. Gurbel PA, Bliden KP, Zaman KA, Yoho JA, Hayes KM, Tantry US (2005). Clopidogrel loading with eptifibatide to arrest the reactivity of platelets: Results of the clopidogrel loading with eptifibatide to arrest the reactivity of platelets (CLEAR PLATELETS) study. *Circulation 111*, 1153–1159.
14. Gurbel PA, Bliden KP, Samara W, Yoho JA, Hayes KM, Tantry US (2005). The clopidogrel Resistance and Stent Thrombosis (CREST) study. *J Am Coll Cardiol 45*, 34A.
15. Bliden KP, Tantry U, Zaman K, Bassi A, Hayes K, Yoho J, Gurbel PA (2005). High platelet reactivity is a risk factor for post-discharge ischemic complications following elective coronary stenting. *J Am Coll Cardiol 45*, 33A.
16. Gurbel PA, Bliden KP, Guyer K, Cho PW, Zaman KA, Kreutz RP, Bassi AK, Tantry US (2005). Platelet reactivity in patients and recurrent events post-stenting: Results of the PREPARE POST-STENTING Study. *J Am Coll Cardiol 46*, 1820–1826.
17. Antithrombotic Trialists' Collaboration (2002). Collaborative meta-analysis of randomised trials of antiplatelet therapy for prevention of death, myocardial infarction, and stroke in high risk patients. *BMJ 324*, 71–86.
18. Smith SC Jr., Feldman TE, Hirshfeld JW Jr, et al. (2006). American College of Cardiology/American Heart Association Task Force on Practice Guidelines; ACC/AHA/SCAI Writing Committee to Update 2001 Guidelines for Percutaneous Coronary Intervention. ACC/AHA/SCAI 2005 guideline update for percutaneous coronary intervention: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines (ACC/AHA/SCAI Writing Committee to Update 2001 Guidelines for Percutaneous Coronary Intervention). *Circulation 113*, e166–286.
19. American College of Cardiology; American Heart Association Task Force on Practice Guidelines (2008). 2007 Focused update of the ACC/AHA/SCAI 2005 guideline update for percutaneous coronary intervention. A report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. *Catheter Cardiovasc Interv 71*, E1–40.
20. CAPRIE Steering Committee. (1996). A randomised, blinded, trial of clopidogrel versus aspirin in patients at risk of ischaemic events (CAPRIE). CAPRIE Steering Committee. *Lancet 348*, 1329–1339.
21. Yusuf S, Zhao F, Mehta SR, Chrolavicius S, Tognoni G, Fox KK; Clopidogrel in Unstable Angina to Prevent Recurrent Events Trial Investigators (2001). Effects of clopidogrel in addition to aspirin in patients with acute coronary syndromes without ST-segment elevation. *N Engl J Med 345*, 494–502.
22. Mehta SR, Yusuf S, Peters RJ, Bertrand ME, Lewis BS, Natarajan MK, Malmberg K, Rupprecht H, Zhao F, Chrolavicius S, Copland I, Fox KA (Aug. 18, 2001). Clopidogrel in Unstable angina to prevent Recurrent Events Trial (CURE) Investigators. Effects of pretreatment with clopidogrel and aspirin followed by long-term therapy in patients undergoing percutaneous coronary intervention: The PCI-CURE study. *Lancet 358*(9281), 527–533.

23. Steinhubl SR, Berger PB, Mann JT 3rd, Fry ET, DeLago A, Wilmer C, Topol EJ; CREDO Investigators (2002). Clopidogrel for the reduction of events during observation. Early and sustained dual oral antiplatelet therapy following percutaneous coronary intervention: A randomized controlled trial. *JAMA 288*, 2411–2420.
24. Helton TJ, Bavry AA, Kumbhani DJ, Duggal S, Roukoz H, Bhatt DL (2007). Incremental effect of clopidogrel on important outcomes in patients with cardiovascular disease: A meta-analysis of randomized trials. *Am J Cardiovasc Drugs 7*, 289–297.
25. Michelson AD, Cattaneo M, Eikelboom JW, Gurbel P, Kottke-Marchant K, Kunicki TJ, Pulcinelli FM, Cerletti C, Rao AK; Platelet Physiology Subcommittee of the Scientific and Standardization Committee of the International Society on Thrombosis and Haemostasis; Working Group on Aspirin Resistance (2005). Aspirin resistance: Position paper of the Working Group on Aspirin Resistance. *J Thromb Haemost 3*, 1309–1311.
26. Wiviott SD, Braunwald E, McCabe CH, Montalescot G, Ruzyllo W, Gottlieb S, Neumann FJ, Ardissino D, De Servi S, Murphy SA, Riesmeyer J, Weerakkody G, Gibson CM, Antman EM; TRITON-TIMI 38 Investigators (2007). Prasugrel versus clopidogrel in patients with acute coronary syndromes. *N Engl J Med 357*, 2001–2005.
27. Ho PM, Peterson ED, Wang L, Magid DJ, Fihn SD, Larsen GC, Jesse RA, Rumsfeld JS (2008). Incidence of death and acute myocardial infarction associated with stopping clopidogrel after acute coronary syndrome. *JAMA 299*, 532–539.
28. Tang M, Mukundan M, Yang J, Charpentier N, LeCluyse EL, Black C, Yang D, Shi D, Yan B (2006). Antiplatelet agents aspirin and clopidogrel are hydrolyzed by distinct carboxylesterases, and clopidogrel is transesterificated in the presence of ethyl alcohol. *J Pharmacol Exp Ther 319*, 1467–1476.
29. Savi P, Combalbert J, Gaich C, et al. (1994). The antiaggregating activity of clopidogrel is due to a metabolic activation by the hepatic cytochrome P450-1A. *Thromb Haemost 72*, 313–317.
30. Farid NA, Payne CD, Small DS, et al. (2007). Cytochrome P450 3A inhibition by ketoconazole affects prasugrel and clopidogrel pharmacokinetics and pharmacodynamics differently. *Clin Pharmacol Ther 81*, 735–741.
31. Ding Z, Kim S, Dorsam RT, Jin J, Kunapuli SP (2003). Inactivation of the human P2Y12 receptor by thiol reagents requires interaction with both extracellular cysteine residues, Cys17 and Cys270. *Blood 101*, 3908–3914.
32. Foster CJ, Prosser DM, Agans JM, et al. (2001). Molecular identification and characterization of the platelet ADP receptor targeted by thienopyridine antithrombotic drugs. *J Clin Invest 107*, 1591–1598.
33. Labarthe B, Theroux P, Angioi M, Ghitescu M (2005). Matching the evaluation of the clinical efficacy of clopidogrel to platelet function tests relevant to the biological properties of the drug. *J Am Coll Cardiol 46*, 638–645.
34. Gurbel PA, Bliden KP, Etherington A, Tantry US (2007). Assessment of clopidogrel responsiveness: Measurements of maximum platelet aggregation, final platelet aggregation and their correlation with vasodilator-stimulated phosphoprotein in resistant patients. *Thromb Res 21*, 107–115.
35. Gurbel PA, Bliden KP, Hiatt BL, O'Connor CM (2003). Clopidogrel for coronary stenting: Response variability, drug resistance, and the effect of pretreatment platelet reactivity. *Circulation 107*, 2908–2913.
36. von Beckerath N, Pogatsa-Murray G, Wieczorek A, Sibbing D, Schomig A, Kastrati A (2006). Correlation of a new point-of-care test with conventional optical aggregometry for the assessment of clopidogrel responsiveness. *Thromb Haemost 95*, 910–911.
37. Aleil B, Ravanat C, Cazenave JP, et al. (2005). Flow cytometric analysis of intraplatelet VASP phosphorylation for the detection of clopidogrel resistance in patients with ischemic cardiovascular diseases. *J Thromb Haemost 3*, 85–92.
38. Gurbel PA, Bliden KP, Hayes KM, et al. (2005). The relation of dosing to clopidogrel responsiveness and the incidence of high post-treatment platelet aggregation in patients undergoing coronary stenting. *J Am Coll Cardiol 45*, 1392–1396.
39. Jaremo P, Lindahl TL, Fransson SG, Richter A (2002). Individual variations of platelet inhibition after loading doses of clopidogrel. *J Intern Med 252*, 233–238.
40. Muller I, Besta F, Schulz C, Massberg S, Schonig A, Gawaz M (2003). Prevalence of clopidogrel non-responders among patients with stable angina pectoris scheduled for elective coronary stent placement. *Thromb Haemost 89*, 783–787.
41. Mobley JE, Bresee SJ, Wortham DC, Craft RM, Snider CC, Carroll RC (2004). Frequency of nonresponse antiplatelet activity of clopidogrel during pretreatment for cardiac catheterization. *Am J Cardiol 93*, 456–458.
42. Lepantalo A, Virtanen KS, Heikkila J, Wartiovaara U, Lassila R (2004). Limited early antiplatelet effect of 300 mg clopidogrel in patients with aspirin therapy undergoing percutaneous coronary interventions. *Eur Heart J 25*, 476–483.
43. Angiolillo DJ, Fernandez-Ortiz A, Bernardo E, Ramirez C, Barrera-Ramirez C, Sabate M, Hernandez R, Moreno R, Escaned J, Alfonso F, Banuelos C, Costa MA, Bass TA, Macaya C (2005). Identification of low responders to a 300-mg clopidogrel loading dose in patients undergoing coronary stenting. *Thromb Res 115*, 101–108.
44. Matetzky S, Shenkman B, Guetta V, Shechter M, Bienart R, Goldenberg I, Novikov I, Pres H, Savion N, Varon D, Hod H (2004). Clopidogrel resistance is associated with increased risk of recurrent atherothrombotic events in patients with acute myocardial infarction. *Circulation 109*, 3171–3175.

45. Dziewierz A, Dudek D, Heba G, Rakowski T, Mielecki W, Dubiel JS (2005). Inter-individual variability in response to clopidogrel in patients with coronary artery disease. *Kardiol Pol 62*, 108–117.
46. Lev EI, Patel RT, Maresh KJ, Guthikonda S, Granada J, DeLao T, Bray PF, Kleiman NS (2006). Aspirin and clopidogrel drug response in patients undergoing percutaneous coronary intervention: The role of dual drug resistance. *J Am Coll Cardiol 47*, 27–33.
47. Angiolillo DJ, Fernandez-Ortiz A, Bernardo E, Ramirez C, et al. (2005). Platelet function profiles in patients with type 2 diabetes and coronary artery disease on combined aspirin and clopidogrel treatment. *Diabetes 54*, 2430–2435.
48. von Beckerath N, Taubert D, Pogatsa-Murray G, Schomig E, Kastrati A, Schomig A (2005). Absorption, metabolization, and antiplatelet effects of 300-, 600-, and 900-mg loading doses of clopidogrel: Results of the ISAR-CHOICE (Intracoronary Stenting and Antithrombotic Regimen: Choose Between 3 High Oral Doses for Immediate Clopidogrel Effect) Trial. *Circulation 112*, 2946–2950.
49. Lotrionte M, Biondi-Zoccai GG, Agostoni P, Abbate A, Angiolillo DJ, Valgimigli M, Moretti C, Meliga E, Cuisset T, Alessi MC, Montalescot G, Collet JP, Di Sciascio G, Waksman R, Testa L, Sangiorgi G, Laudito A, Trevi GP, Sheiban I (2007). Meta-analysis appraising high clopidogrel loading in patients undergoing percutaneous coronary intervention. *Am J Cardiol 100*, 1199–1206.
50. Zhang L, Kanna B (2008). A double blind, randomized study on platelet aggregation in patients treated with a daily dose of 150 or 75 mg of clopidogrel for 30 days. *Eur Heart J 29*, 423–424.
51. von Beckerath N, Kastrati A, Wieczorek A, Pogatsa-Murray G, Sibbing D, Graf I, Schömig A (2007). A double-blind, randomized study on platelet aggregation in patients treated with a daily dose of 150 or 75 mg of clopidogrel for 30 days. *Eur Heart J 28*, 1814–1819.
52. Gurbel PA, Bliden KP, Tantry US (2006). Effect of clopidogrel with and without eptifibatide on tumor necrosis factor-alpha and C-reactive protein release after elective stenting: Results from the CLEAR PLATELETS 1b study. *J Am Coll Cardiol 48*, 2186–2191.
53. Cuisset T, Frere C, Quilici J, Barbou F, et al. (2006). High post-treatment platelet reactivity identified low-responders to dual antiplatelet therapy at increased risk of recurrent cardiovascular events after stenting for acute coronary syndrome. *J Thromb Haemost 4*, 542–549.
54. Barragan P, Bouvier JL, Roquebert PO, Macaluso G, Commeau P, Comet B, Lafont A, Camoin L, Walter U, Eigenthaler M (2003). Resistance to thienopyridines: Clinical detection of coronary stent thrombosis by monitoring of vasodilator-stimulated phosphoprotein phosphorylation. *Catheter Cardiovasc Interv 59*, 295–302.
55. Ajzenberg N, Aubry P, Huisse MG, Cachier A, et al. (2005). Enhanced shear-induced platelet aggregation in patients who experience subacute stent thrombosis: A case-control study. *J Am Coll Cardiol 45*, 1753–1756.
56. Buonamici P, Marcucci R, Miglironi A, et al. (2007). Impact of platelet reactivity after clopidogrel administration on drug-eluting stent thrombosis. *J Am Coll Cardiol 49*, 2312–2317.
57. Price JM, Wong GB, Valenica R, et al. (2006) Measurement of clopidogrel inhibition with a point-of-care assay identifies patients at risk for stent thrombosis after percutaneous coronary intervention. *Am J Cardiol 98*, 204M (abstract).
58. Geisler T, Langer H, Wydymus M, et al. (2006). Low response to clopidogrel is associated with cardiovascular outcome after coronary stent implantation. *Eur Heart J 27*, 2420–2425.
59. Hochholzer W, Trenk D, Bestehorn HP, et al. (2006). Impact of the degree of peri-interventional platelet inhibition after loading with clopidogrel on early clinical outcome of elective coronary stent placement. *J Am Coll Cardiol 48*, 1742–1750.
60. Lau WC, Gurbel PA, Watkins PB, Neer CJ, Hopp AS, Carville DG, Guyer KE, Tait AR, Bates ER (2004). Contribution of hepatic cytochrome P450 3A4 metabolic activity to the phenomenon of clopidogrel resistance. *Circulation 109*, 166–171.
61. Gurbel PA, Lawal L, Singla A, et al. (2007). Synergistic effect of smoking on platelet responsiveness to clopidogrel. Ab-87580. Paper presented at the European Society of Cardiology Meeting. Vienna.
62. Lau WC, Welch TD, Shields T, Rubenfire M, Tantry US, Gurbel PA (2011). The effect of St John's Wort on the pharmacodynamic response of clopidogrel in hyporesponsive volunteers and patients: Increased platelet inhibition by enhancement of CYP3A4 metabolic activity. *J Cardiovasc Pharmacol 57*, 86–93.
63. Lau WC, Waskell LA, Watkins PB, Neer CJ, Horowitz K, Hopp AS, Tait AR, Carville DG, Guyer KE, Bates ER (2003). Atorvastatin reduces the ability of clopidogrel to inhibit platelet aggregation: A new drug-drug interaction. *Circulation 107*, 32–37.
64. Wienbergen H, Gitt AK, Schiele R, Juenger C, Heer T, Meisenzahl C, Limbourg P, Bossaller C, Senges J; MITRA PLUS Study Group (2003). Comparison of clinical benefits of clopidogrel therapy in patients with acute coronary syndromes taking atorvastatin versus other statin therapies. *Am J Cardiol 92*, 285–288.
65. Neubauer H, Mugge A (2006). Thienopyridines and statins: Assessing a potential drug-drug interaction. *Curr Pharm Des 12*, 1271–1280.
66. Farid NA, Payne CD, Small DS, Winters KJ, Ernest CS 2nd, Brandt JT, Darstein C, Jakubowski JA, Salazar DE (2007). Cytochrome P450 3A inhibition by ketoconazole affects prasugrel and clopidogrel pharmacokinetics and pharmacodynamics differently. *Clin Pharmacol Ther 81*, 735–741.
67. Jakubowski JA, Payne CD, Brandt JT, Weerakkody GJ, Farid NA, Small DS, Naganuma H, Li GY, Winters KJ

(2006). The platelet inhibitory effects and pharmacokinetics of prasugrel after administration of loading and maintenance doses in healthy subjects. *J Cardiovasc Pharmacol 47*, 377–384.

68. Suh JW, Koo BK, Zhang SY, Park KW, Cho JY, et al. (2006). Increased risk of atherothrombotic events associated with cytochrome P450 3A5 polymorphism in patients taking clopidogrel. *CMAJ 174*, 1715–1722.
69. Hulot JS, Bura A, Villard E, Azizi M, Remones V, Goyenvalle C, Aiach M, Lechat P, Gaussem P (2006). Cytochrome P450 2C19 loss-of-function polymorphism is a major determinant of clopidogrel responsiveness in healthy subjects. *Blood. 108*, 2244–2247.

69a. Brandt JT, Close SL, Iturria SJ, et al. (2007). Common polymorphisms of CYP2C19 and CYP2C9 affect the pharmacokinetic and pharmacodynamic response to clopidogrel but not prasugrel. *J Thromb Haemost 5*, 2429–2436.

70. Gilard M, Arnaud B, Cornily JC, Le Gal G, Lacut K, Le Calvez G, Mansourati J, Mottier D, Abgrall JF, Boschat J (2008). Influence of omeprazole on the antiplatelet action of clopidogrel associated with aspirin: The randomized, double-blind OCLA (Omeprazole CLopidogrel Aspirin) study. *J Am Coll Cardiol 51*, 256–260.
71. Beer JH, Pederiva S, Pontiggia L (2000). Genetics of platelet receptor single-nucleotide polymorphisms: Clinical implications in thrombosis. *Ann Med 32*, 10–14.
72. Fontana P, Gaussem P, Aiach M, Fiessinger JN, Emmerich J, Reny JL (2003). P2Y12 H2 haplotype is associated with peripheral arterial disease: A case-control study. *Circulation 16*(108), 2971–2973.
73. Von Beckerath N, von Beckerath O, Koch W, Eichinger M, Schomig A, Kastrati A (2005). P2Y12 gene H2 haplotype is not associated with increased adenosine diphosphate-induced platelet aggregation after initiation of clopidogrel therapy with a high loading dose. *Blood Coagul Fibrinolysis 16*, 199–204.
74. Ferroni P, Basili S, Falco A, Davi G (2004). Platelet activation in type 2 diabetes mellitus. *J Thromb Haemost 2*, 1282–1291.
75. Angiolillo DJ, Fernandez-Ortiz A, Bernardo E, Ramirez C, Sabate M, Jimenez-Quevedo P, Hernandez R, Moreno R, Escaned J, Alfonso F, et al. (2005). Platelet function profiles in patients with type 2 diabetes and coronary artery disease on combined aspirin and clopidogrel treatment. *Diabetes 54*, 2430–2435.
76. Angiolillo DJ, Fernandez-Ortiz A, Bernardo E, Barrera Ramirez C, Alfonso F, Macaya C (2004). Platelet aggregation according to body mass index in patients undergoing coronary stenting: Should clopidogrel loading-dose be weight adjusted? *J Invas Cardiol 16*, 169–174.
77. Kastrati A, von Beckerath N, Joost A, Pogatsa-Murray G, Gorchakova O, Schomig A (2004). Loading with 600 mg clopidogrel in patients with coronary artery disease with and without chronic clopidogrel therapy. *Circulation 110*, 1916–1919.
78. Greenbaum AB, Ohman EM, Gibson CM, Borzak S, Stebbins AL, Lu M, Le May MR, Stankowski JE, Emanuelsson H, Weaver WD (2007). Preliminary experience with intravenous P2Y12 platelet receptor inhibition as an adjunct to reduced-dose alteplase during acute myocardial infarction: Results of the safety, tolerability and effect on patency in acute myocardial infarction (STEP-AMI) angiographic trial. *Am Heart J 154*, 702–709.
79. Husted S, Emanuelsson H, Heptinstall S, Sandset PM, Wickens M, Peters G (2006). Pharmacodynamics, pharmacokinetics, and safety of the oral reversible P2Y12 antagonist AZD6140 with aspirin in patients with atherosclerosis: A double-blind comparison to clopidogrel with aspirin. *Eur Heart J 27*, 1038–1047.
80. Cannon CP, Husted S, Harrington RA, Scirica BM, Emanuelsson H, Peters G, Storey RF; DISPERSE-2 Investigators (2007). Safety, tolerability, and initial efficacy of AZD6140, the first reversible oral adenosine diphosphate receptor antagonist, compared with clopidogrel, in patientswith non-ST-segment elevation acute coronary syndrome: Primary results of the DISPERSE-2 trial. *J Am Coll Cardiol 50*, 1844–1851.
81. Wiviott SD, Antman EM, Winters KJ, Weerakkody G, Murphy SA, Behounek BD, Carney RJ, Lazzam C, McKay RG, McCabe CH, Braunwald E; JUMBO-TIMI 26 Investigators (2005). Randomized comparison of prasugrel (CS-747, LY640315), a novel thienopyridine P2Y12 antagonist, with clopidogrel in percutaneous coronary intervention: Results of the Joint Utilization of Medications to Block Platelets Optimally (JUMBO)-TIMI 26 trial. *Circulation 28*(111), 3366–3373.
82. Jakubowski JA, Payne CD, Brandt JT, Weerakkody GJ, Farid NA, Small DS, Naganuma H, Li GY, Winters KJ (2006). The platelet inhibitory effects and Pharmacokinetics of prasugrel after administration of loading and maintenance doses in healthy subjects. *J Cardiovasc Pharmacol 47*, 377–384.
83. Wiviott SD, Braunwald E, McCabe CH, Montalescot G, Ruzyllo W, Gottlieb S, Neumann FJ, Ardissino D, De Servi S, Murphy SA, Riesmeyer J, Weerakkody G, Gibson CM, Antman EM; TRITON-TIMI 38 Investigators (2007). Prasugrel versus clopidogrel in patients with acute coronary syndromes. *N Engl J Med 357*, 2001–2015.
84. Vane JR, Botting RM (2003).The mechanism of action of aspirin. *Thromb Res 110*, 255–258.
85. Hennekens CH (2002). Update on aspirin in the treatment and prevention of cardiovascular disease. *Am J Manag Care 8*, S691–700.
86. Hurlen M, Seljeflot I, Arnesen H (1998). The effect of different antithrombotic regimens on platelet aggregation after myocardial infarction. *Scand Cardiovasc J 32*, 233–237.
87. Buchanan MR, Brister SJ (1995). Individual variation in the effects of ASA on platelet function: Implications for the use of ASA clinically. *Can J Cardiol 11*, 221–227.

88. Buchanan MR, Schwartz L, Bourassa M, Brister SJ, Peniston CM; BRAT Investigators (2000). Results of the BRAT study—A pilot study investigating the possible significance of ASA nonresponsiveness on the benefits and risks of ASA on thrombosis in patients undergoing coronary artery bypass surgery. *Can J Cardiol 16*, 1385–1390.

89. Grotemeyer KH, Scharafinski HW, Husstedt IW (1993). Two-year follow-up of aspirin responder and aspirin non responder. A pilot-study including 180 post-stroke patients. *Thromb Res 71*, 397–403.

90. Gum PA, Kottke-Marchant K, Poggio ED, et al. (2001). Profile and prevalence of aspirin resistance in patients with cardiovascular disease. *Am J Cardiol 88*, 230–235.

91. Gum PA, Kottke-Marchant K, Welsh PA, White J, Topol EJ (2003). A prospective, blinded determination of the natural history of aspirin resistance among stable patients with cardiovascular disease. *J Am Coll Cardiol 41*, 961–965.

92. Eikelboom JW, Hirsh J, Weitz JI, Johnston M, Yi Q, Yusuf S (2002). Aspirin-resistant thromboxane biosynthesis and the risk of myocardial infarction, stroke, or cardiovascular death in patients at high risk for cardiovascular events. *Circulation 105*, 1650–1655.

93. Wang JC, Aucoin-Barry D, Manuelian D, Monbouquette R, Reisman M, Gray W, Block PC, Block EH, Ladenheim M, Simon DI (2003). Incidence of aspirin nonresponsiveness using the Ultegra Rapid Platelet Function Assay-ASA. *Am J Cardiol 92*, 1492–1494.

94. Chen W-H, Lee P-Y, Ng W, et al. (2004) Aspirin resistance is associated with a high incidence of myonecrosis after non-urgent percutaneous coronary intervention despite clopidogrel pretreatment. *J Am Coll Cardiol 43*, 1122–1126.

95. Gonzalez-Conejero R, Rivera J, Corral J, Acuna C, Guerrero JA, Vicent V (2005). Biological assessment of aspirin efficacy on healthy individuals: Heterogeneous response or aspirin failure? *Stroke 36*, 276–280.

96. Tantry US, Bliden KP, Gurbel PA (2005). Overestimation of platelet aspirin resistance detection by thrombelastograph platelet mapping and validation by conventional aggregometry using arachidonic acid stimulation. *J Am Coll Cardiol 46*, 1705–1709.

97. Schwartz KA, Schwartz DE, Ghosheh K, Reeves MJ, Barber K, DeFranco A (2005). Compliance as a critical consideration in patients who appear to be resistant to aspirin after healing of myocardial infarction. *Am J Cardiol 95*, 973–975.

98. Gurbel PA, Bliden KP, DiChiara J, et al. (2007). Evaluation of dose-related effects of aspirin on platelet function: Results from the Aspirin-Induced Platelet Effect (ASPECT) study. *Circulation 115*, 3156–3164.

99. DiChiara J, Bliden KP, Tantry US, Hamed MS, Antonino MJ, Suarez TA, Bailon O, Singla A, Gurbel PA (2007). The effect of aspirin dosing on platelet function in diabetic and nondiabetic patients: An analysis from the aspirin-induced platelet effect (ASPECT) study. *Diabetes 56*, 3014–3019.

100. Lordkipanidze M, Pharand C, Schampaert E, Turgeon J, Palisaitis DA, Diodati JG (2007). A comparison of six major platelet function tests to determine the prevalence of aspirin resistance in patients with stable coronary artery disease. *Eur Heart J 28*, 1702–1708.

101. Marcucci R, Paniccia R, Antonucci E, Gori AM, Fedi S, Giglioli C, Valente S, Prisco D, Abbate R, Gensini GF (2006). Usefulness of aspirin resistance after percutaneous coronary intervention for acute myocardial infarction in predicting one-year major adverse coronary events. *Am J Cardiol 98*, 1156–1159.

102. Gianetti J, Parri MS, Sbrana S, Paoli F, Maffei S, Paradossi U, Berti S, Clerico A, Biagini A (2006). Platelet activation predicts recurrent ischemic events after percutaneous coronary angioplasty: A 6 months prospective study. *Thromb Res 118*, 487–493.

103. Gulmez O, Yildirir A, Kaynar G, Konas D, Aydinalp A, Ertan C, Ozin B, Muderrisoglu H (2008). Effects of persistent platelet reactivity despite aspirin therapy on cardiac troponin I and creatine kinase-MB levels after elective percutaneous coronary interventions. *J Thromb Thrombolysis 25*, 239–246.

104. Foussas SG, Zairis MN, Patsourakos NG, Makrygiannis SS, Adamopoulou EN, Handanis SM, Prekates AA, Fakiolas CN, Pissimissis EG, Olympios CD, Argyrakis SK (2007). The impact of oral antiplatelet responsiveness on the long-term prognosis after coronary stenting. *Am Heart J 154*, 676–681.

105. Malek LA, Spiewak M, Filipiak KJ, et al. (2007). Persistent platelet activation is related to very early cardiovascular events in patients with acute coronary syndromes. *Kardiol Pol 65*, 40–45.

106. Borna C, Lazarowski E, van Heusden C, et al. (2005). Resistance to aspirin is increased by ST-elevation myocardial infarction and correlates with adenosine diphosphate levels. *Thromb J 3*, 10.

107. Templin C, Schaefer A, Stumme B, et al. (2006). Combined aspirin and clopidogrel resistance associated with recurrent coronary stent thrombosis. *Clin Res Cardiol 95*, 122–126.

108. Rauch U, Osende JI, Chesebro JH, Fuster V, Vorchheimer DA, Harris K, Harris P, Sandler DA, Fallon JT, Jayaraman S, Badimon JJ (2000). Statins and cardiovascular diseases: The multiple effects of lipid-lowering therapy by statins. *Atherosclerosis 153*, 181–189.

109. Kronish IM, Rieckmann N, Shimbo D, Davidson KW (2007). Letter by Kronish et al regarding article, "Residual arachidonic acid-induced platelet activation via an adenosine diphosphate-dependent but cyclooxygenase-1- and cyclooxygenase-2-independent pathway: A 700-patient study of aspirin resistance." *Circulation 115*, e45; author reply e46.

110. ISIS-2 (Second International Study of Infarct Survival) Collaborative Group (1988). Randomised trial of intravenous streptokinase, oral aspirin, both, or neither among

17,187 cases of suspected acute myocardial infarction: ISIS-2. *Lancet 2*, 349–360.

111. Steering Committee of the Physicians' Health Study Research Group (1989). Final report on the aspirin component of the ongoing Physicians' Health Study. *N Engl J Med 321*, 129–135.
112. Zimmermann N, Wenk A, Kim U, Kienzle P, Weber AA, Gams E, Schror K, Hohlfeld T (2003). Functional and biochemical evaluation of platelet aspirin resistance after coronary artery bypass surgery. *Circulation 108*, 542–547.
113. Ridker PM, Cook NR, Lee IM, Gordon D, et al. (2005). A randomized trial of low-dose aspirin in the primary prevention of cardiovascular disease in women. *N Engl J Med 352*, 1293–1304.
114. Levin RI (2005). The puzzle of aspirin and sex. *N Engl J Med 352*, 1366–1368.
115. Colwell JA; American Diabetes Association (2004). Aspirin therapy in diabetes. *Diabetes Care 27*(Suppl 1), S72–73.
116. Mueller MR, Salat A, Stangl P, et al. (1997). Variable platelet response to low-dose ASA and the risk of limb deterioration in patients submitted to peripheral arterial angioplasty. *Thromb Haemost 78*, 1003–1007.
117. DUTCH Group (1991). A comparison of two doses of aspirin (30 mg vs. 283 mg a day) in patients after a transient ischemic attack or minor ischemic stroke. The Dutch TIA Trial Study Group. *N Engl J Med 325*, 1261–1266.
118. Tohgi H, Konno S, Tamura K, Kimura B, Kawano K (1992). Effects of low-to-high doses of aspirin on platelet aggregability and metabolites of thromboxane A2 and prostacyclin. *Stroke 23*, 1400–1403.
119. Bhatt DL, Fox KA, Hacke W, Berger PB, Black HR, Boden WE, et al.; CHARISMA Investigators (2006). Clopidogrel and aspirin versus aspirin alone for the prevention of atherothrombotic events. *N Engl J Med 354*, 1706–1717.

# ANTIBIOTIC DEVELOPMENT

Robert L. Bettiker[1] and Jason C. Gallagher[2]
[1]*Temple University School of Medicine, Philadelphia, Pennsylvania*
[2]*Temple University School of Pharmacy, Philadelphia, Pennsylvania*

## 1 BIRTH OF THE ANTIBIOTIC ERA

It is easily forgotten that antibiotics as we know them have only been available for about 70 years. Death from infection was a feared eventuality before quite recently, whereas now society almost seems to have the opposite problem—expecting easy cures from antibiotic therapy for nearly all infectious diseases. The early discoveries that were keystone events in antimicrobial development were due to combinations of luck, skill, ingenuity, and Nobel-Prize-winning research.

The antibiotic era was primed to begin after Alexander Fleming's famous observation in 1929 of a zone of inhibition around a *Penicillium notatum* mold on a Petri dish of staphylococci in his laboratory [1]. Fleming was wise enough to recognize that a natural substance was being secreted from the mold, and for the sake of convenience he named this substance *penicillin* to avoid calling it "mold broth filtrate," which is what his experiments began working with [1]. To Fleming, penicillin's greatest potential was to assist in the selection culturing of organisms that were penicillin resistant to prevent overgrowth of organisms such as streptococci and staphylococci on culture medium from samples taken in areas of mixed human flora. His paper only briefly mentioned its possible utility in treating infections, and the editor's first revision of the paper actually removed mention of this [1, 2].

The utility of penicillin as a therapeutic agent was first described by Chain et al. in 1940 [3]. By this point, sulfonamides had already been discovered and were being used therapeutically in the treatment of infections [4]. A group under the leadership of Gerhard Domagk experimented with a antimicrobial dye named prontosil that was eventually determined to be a prodrug of sulfanilamide [4]. Experiments in mice revealed prontosil's in vivo activity against streptococcal infections, and Domagk himself administered the drug successfully to his child when she developed a dangerous streptococcal infection. After the publication of Domagk's results, the discovery and development of many active sulfonamides followed with numerous compounds becoming commercially available [4]. Domagk won the Nobel Prize in Medicine in 1939 for his discovery of prontosil's antimicrobial effects, though he was prevented from receiving it until 1947 due to Hitler's policies [4, 5].

Chain et al. described penicillin's antimicrobial activity in a paper written in 1940 [3]. By this point, the active penicillin molecule had still not been isolated, but several researchers were working with active penicillin preparations in a similar manner to Fleming. Chain and co-workers were the first to investigate penicillin as a chemotherapeutic agent. They tested this by injecting groups of healthy mice with one of three bacteria: *Staphylococcus aureus*, *Streptococcus pyogenes*, or *Clostridium septicum* (formerly *C. septique*) [3]. Half of the mice were then given penicillin for different durations. Penicillin dramatically saved the lives of the treated mice, while all controls expired. This led to testing in human patients and the eventual availability of the refined drug for Allied soldiers in World War II. For their work in the discovery and development of

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

penicillin, Fleming and Chain and co-workers jointly received the Nobel Prize in Medicine in 1945 [2, 6].

The initial fortuitous discovery of penicillin and the development of the sulfonamides led to a massive expansion in the area of antimicrobial drug discovery. Another pioneer in this area was Salman Waksman at Rutgers University [7]. Waksman's laboratory sampled soil bacteria looking for naturally occurring "antibiotics," a phrase that he coined. After several failed compounds, the aminoglycoside streptomycin was discovered. Streptomycin was notable in that it was active against organisms that were resistant to both sulfonamides and penicillin. Most significantly, streptomycin has activity against *Mycobacterium tuberculum*. For developing the first agent active in the treatment of tuberculosis, Waksman was awarded the Nobel Prize in Medicine in 1952 [7].

These early advances paved the way for the development of further new antimicrobials, and many large pharmaceutical companies and university laboratories alike participated in their discovery and development [6]. As antimicrobials became utilized more frequently and bacteria were exposed to them, antibiotic resistance inevitably occurred. New antibiotics were needed to address the growing tide of resistant bacteria, and antibiotics that were active against the many organisms not treated by penicillin, sulfonamides, and streptomycin needed to be developed. Classes of antibiotics that have been developed include drugs active against cell-wall biosynthesis, bacterial protein synthesis, topoisomerases and bacterial deoxyribonucleic acid (DNA), folic acid synthesis, ribonucleic acid (RNA) polymerase, and several others. To examine how antibiotic development has progressed as bacteria have evolved and new needs have been met, we will examine the development of the penicillins.

### 1.1 Penicillin Development

Penicillin enjoyed tremendous use in the treatment of many infections after its potential was discovered, particularly in the treatment of wound infections caused by streptococci and staphylococci. Nevertheless, it was not long after it became widely used that clinicians began to notice that some infections that should have been cured with penicillin were no longer responding [6]. Initially, this was thought to be a problem of dosage, but eventually the production of penicillinases by different types of bacteria were discovered, most notably staphylococci [6, 8].

β-Lactamases are enzymes that destroy the β-lactam ring of penicillin molecules, rendering them inactive [6, 9]. The term penicillinase refers to a β-lactamase that is active against penicillins. These enzymes were uncommon initially, but, as staphylococci that could produce them gained a competitive advantage in a penicillin-filled world, they have proliferated. The vast majority of staphylococci are now penicillin resistant. To counteract the activity of penicillinases, researchers began chemical substitutions of the functional groups attached to the core of penicillin, 6-aminopenicillanic acid (6-APA) [6]. After rising to the challenge of producing 6-APA from penicillin, it was available as a nucleus for the substitution of multiple side chains to create many semisynthetic penicillins [6, 8]. Among the first of these semisynthetic penicillins was methicillin [6, 8]. Unlike penicillin, methicillin contains a large side chain that sterically inhibits penicillinase from reaching the β-lactam ring. Methicillin was the first of what would be known as the "antistaphylococcal penicillins," since resistance to penicillin was so common that it eventually became almost useless in staphylococcal infections [6, 8]. Methicillin was soon followed by other agents, including nafcillin and cloxacillin [6, 8]. Though methicillin was removed from the market due to concerns with nephrotoxicity from acute interstitial nephritis, its classmates remain in use in clinical practice.

Although the antistaphylococcal penicillins improved on the activity of penicillin against staphylococci, they did not improve on the poor gram-negative activity of penicillin. To address this gap, several classes of semisynthetic penicillins with a broader spectrum of activity were developed [6, 8]. The first of these were the aminopenicillins, beginning with ampicillin in 1961 [6, 8]. Aminopenicillins have a polar side chain that allows them to enter the porin channels in the outer membrane of some gram-negative bacteria, allowing for their activity. The introduction of ampicillin gave clinicians a drug that was more active against gram-negative organisms than penicillin [8]. This enhanced activity increased the ability to treat infections of the urinary and respiratory tract. However, ampicillin and other aminopenicillins were found to be inactive against some gram-negative bacilli, including *Pseudomonas aeruginosa* and the anaerobe *Bacteroides fragilis* [6, 8]. Also, they are as susceptible to penicillinases as penicillin and they are not active against penicillinase-producing staphylococci. As aminopencillins began to be used more frequently, resistance to them in previously susceptible organisms increased, and this was found to be due to penicillinase production in other organisms.

In 1967, carbenecillin was introduced to the antibiotic market. At the time, carbenecillin was unique in that it was the first penicillin with clinically useful activity against *P. aeruginosa*, a difficult-to-treat gram-negative bacillus [6, 8]. Carbenecillin and the other antipseudomonal penicillins also display activity against some Enterobacteraciae that were resistant to amino-

penicillins, such as the *Enterobacter* species. Though carbenecillin is active against *P. aeruginosa*, other antipseudomonal pencillins, such as ticarcillin and piperacillin, were found to have superior activity [8, 10]. However, the antipseudomonal penicillins are susceptible to many β-lactamases, including the penicillinases produced by staphylococci and many gram-negative bacilli. Hence, the search for even broader-spectrum penicillins continued.

Though improvements in the spectrum of activity of penicillin were made against staphylococci with the antistaphylococcal pencillins, and against gram-negative bacilli with the aminopenicillins and antipseudomonal pencillins, no single penicillin had been developed that encompassed all of these beneficial properties. The need for an expanded-spectrum penicillin that was stable to penicillinases was apparent. Scientists at Beecham Research Laboratories, who had taken part in the earlier development of the aminopenicillins and antipseudomonal penicillins, screened bacteria for the production of natural inhibitors of β-lactamase [9, 11]. They discovered that *Streptococcus clavuligerus* produced such a substance, which they named clavulanic acid [11]. Clavulanic acid structurally resembles a penicillin but has little antimicrobial activity. When given to a β-lactamase-producing organism, clavulanate irreversibly binds the β-lactamase [11]. If clavulanate and a β-lactam are given together to a β-lactamase-producing organism, the β-lactam will not be inactivated by the β-lactamase and is available to exert its pharmacologic effect on the microorganism, provided that it has intrinsic activity against it (such as with penicillinase-producing staphylococci and penicillins). The first penicillin–β-lactamase combination released was amoxicillin–clavulanate in 1981 [11]. Amoxicillin is an aminopenicillin closely related to ampicilllin, except for a hydroxyl substitution that allows for enhanced oral absorption [6, 8]. The co-formulation, named Augmentin, was significant in that it restored the intrinsic activity of amoxicillin against penicillinase-producing staphylococci as well as other penicillinase-producing organisms that had evolved resistance to aminopenicillins and some that intrinsically produce penicillinases, such as *B. fragilis*. The later combination of ticarcillin and clavulanate in one formulation rendered a broad-spectrum penicillin combination that was also active against *P. aeruginosa* [9].

Penicillins have played an important role in antimicrobial chemotherapy for decades, and they are an excellent example of how antimicrobial development progresses to follow the needs of patients. The relentless advance of bacterial resistance drives the continued need for new antimicrobials. Shortly after the introduction of methicillin in 1959, methicillin-resistant staphylococci were isolated [12, 13]. These methicillin-resistant organisms had evolved a change in the target enzyme of β-lactams, rendering all of them useless in the treatment of these pathogens [14]. Now methicillin-resistant *S. aureus* (MRSA) predominates in many U.S. hospitals and is increasing dramatically in the community setting [14]. Over 40 years after MRSA was first described, some anti-MRSA β-lactams are approaching the final stages of clinical development, of which the cephalosporin ceftaroline has been approved [15, 16]. It is inevitable that resistance to ceftaroline will emerge, and our pursuit of new antimicrobials will continue.

## 2 WHY DO WE NEED NEW ANTIBIOTICS?

Since the time of the first discovery of penicillin-resistant *S. aureus*, humans have been in a race with bacteria—developing effective antibiotics before the bacteria can become resistant to them. Antibiotic discovery and development is different from most other fields in two major ways. First, antibiotics must have differential effects on two very different organisms—the human host and the invading bacteria [17, 18]. The drug must harm the latter while passing harmlessly through the former. Enzymatic or metabolic targets must be selected such that the bacterial processes are poisoned, leaving a similar human process minimally impacted. Second, bacterial evolution far outpaces human evolution, and resistance development will change an antibiotic's effectiveness over time. The magic bullet penicillin that cured *S. aureus* endocarditis in the 1950s is rarely effective today [19]. This is not always the case in other fields: Although rarely used, the old "tried and true" propranolol is still available if we need it—not necessarily true with old antibiotics.

Approximately 2 million hospital-acquired infections occur annually in the United States [20, 21], causing some 90,000 deaths. The majority of these infections are caused by bacteria having at least some resistance [22]. The incidence of resistance in infections in the intensive care unit (ICU) caused by four of the most serious bacteria is on the rise (see Fig. 1). Clinicians in the United States, including the authors, are confronting *Acinetobacter* infections resistant to all standard antibiotics at an alarmingly increasing rate [23]. Many clinicians are forced to turn to the old drug colistin (polymyxin E)—once thought to be too toxic—to treat these infections [24, 25].

Resistant bacteria would not be the problem they are if we had antibiotics effective against them coming down the development pipeline. Such is not the case. Six new antibiotic classes were developed in the 1950s and 1960s, but only three new ones have since been approved

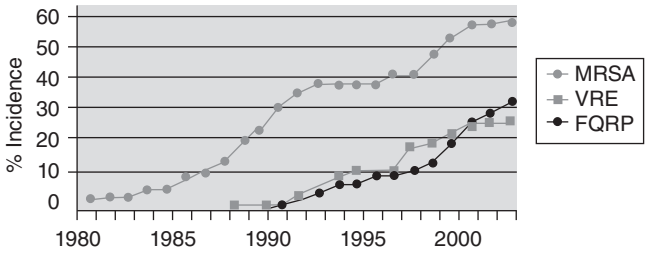


**Figure 1** Resistant strains spread rapidly [21]. This chart shows the increase in rates of resistance for three bacteria that are of concern to public health officials: methicillin-resistant *S. aureus* (MRSA), vancomycin-resistant enterococci (VRE), and fluoroquinolone-resistant *P. aeruginosa* (FQRP). These data were collected from hospital intensive care units that participate in the National Nosocomial Infections Surveillance System, a component of the Centers for Disease Control and Prevention (CDC). (Taken from: Infectious Disease Society of America, Bad Bugs, No Drugs: As antibiotic discovery stagnates, a public health crisis brews, July 2004, www.idsociety.org; accessed 17 May 2008.)

by the Food and Drug Administration (FDA): linezolid in 2000, daptomycin in 2003, and the topical retapamulin in 2007 [26]. All of these are agents against gram-positive bacteria only; no new novel drugs against gram-negatives have come to market in nearly half a century [27]. Ten other antibiotics have been approved since 1998, but these are modifications of existing drugs. For instance, tigecycline, a drug approved in 2005 for both gram-positive and gram-negative infections, is a derivative of the old drug minocycline [28]. Very real concerns exist that established resistance mechanisms will require little modification to inactivate these so-called "me-too" drugs.

Molecules can be combined into a single drug for increased efficacy or durability. Examples are numerous: ß-lactam/ß-lactamase inhibitors, trimethoprim-sulfamethoxazole, and quinupristin-dalfopristin [29]. Combinations of antibiotics marketed separately may be useful in treating resistant bacteria. For example, rifampin and colistin appear to help somewhat in multidrug-resistant *Acinetobacter* spp. infections [25], but this type of combination therapy has limited use clinically due to difficulties in evaluating in vitro susceptibilities to multiple simultaneous antibiotics.

How did we get to the twenty-first century with such a lack of effective antibiotics? Multiple problems exist along the entire drug discovery and development pathway. At the beginning of the pipeline, the source of new molecules to test has dropped to a trickle. Antibacterial screening programs are five times less likely to have a "hit" than one for mammalian cell targets [30]. As the oldest of living organisms, bacteria have had some 3 billion years of evolving in harsh environments and may be more resistant to molecular attack [19].

Once discovered, development costs for a new potential drug are enormous, some $800 million to $1.7 billion [21], and none but the largest pharmaceutical companies can invest in a compound that might never be approved. Smaller companies with promising molecules must entice a larger one to buy or license their drug. The exact pathway of development to approval is unclear as the FDA does not always have clear-cut guidelines on antibiotic approval [31]. The drug itself may be flawed for many reasons, for example, (1) toxicity to humans, (2) lack of bioavailablity pharmacokinetically, (3) inability to reach the molecular target due to difficulties in penetrating the bacterial cell wall or membrane, or to efflux pumps that remove the drug, or (4) lack of durability as the bacteria rapidly become resistant to the compound.

Incentives at the end of the pipeline are likewise feeble: Antibiotics are used for short durations, and some 250–500 persons need to be treated with an antibiotic to recoup the same return on investment as one person on a medication for a chronic disease [22, 31]. As resistance develops to a new drug—and it will—an antibiotic's effectiveness and earning potential will wane over time [32, 33]. Is it any wonder that, in 1990, half of all pharmaceutical companies in the United States and Japan cut back or cut out antibiotic research and development [21, 23]?

Bacterial resistance is such a public health problem that both the Infectious Diseases Society of America (IDSA) and the European Society of Clinical Microbiology and Infectious Diseases (ESCMID) have systematically expressed concern [21, 34, 35]. IDSA's "Bad bugs, no drugs" policy statement of 2004 listed legislative, FDA, and National Institute of Allergy and Infectious Diseases (NIAID) recommendations to address the problem (see Table 1). Response to the recommendations has been tepid.

## 3 ANTIMICROBIAL DISCOVERY

Two major aspects of locating a new chemical entity (NCE) with potential antibiotic properties are (1) a source of compounds to screen and (2) the screening process itself [36]. Pharmacokinetics and dynamics of the potential drug are two more areas of concern in antibiotic development. Although there are innumerable other steps in antibiotic development, the remainder of this chapter will focus on a few examples of these four areas.

Classically, sources of molecules include soil screening, fungal fermentation [36], and chemical compound

**TABLE 1 DSA's "Bad Bugs, No Drugs" Recommendations [21]**

| Congress | FDA | NAIAD |
|---|---|---|
| Establish a commission to prioritize antimicrobial discovery | Publish updated guidelines for antibiotic clinical trials | Aggressively encourage translational (bench to bedside) research |
| Restore patent time lost during FDA's review of priority antibiotics | Encourage imaginative clinical trial design that lead to insight of drug efficacy against resistant bacteria | Remove roadblocks to antibiotic R&D that exist in NIAID's structure, including unnecessary restrictions on companies' intellectual property rights |
| "Wild-Card" patent extension | Provide a clear definition of acceptable surrogate markers as endpoints for clinical trials | Develop collaborations with industry and the infectious diseases research community |
| Extend market exclusivity similar to pediatric and orphan drugs | Explore animal and in vitro infection models and surrogate markers to reduce the number of efficacy studies required for additional indications while maintaining safe and effective dose regimens | Sufficiently fund and rapidly launch NIAID's Drug Discovery and Mechanisms of Antimicrobial Resistance Study Section |
| Tax incentives for R&D of priority antibiotics | Grant priority antibiotics accelerated review status | Engage outside experts in research planning and ensure transparency in decision making |
| Measured liability protections | Explore with NIAID opportunities to streamline antibiotic drug development | Encourage research on topics related to conduct of clinical trials |
| Additional statutory flexibility at FDA regarding antibiotic approval | | Sponsor research into new rapid diagnostic tests for bacterial infections that, when available, could reduce the cost of clinical trials |
| Antitrust exemptions for certain company communications | | Encourage research on antibiotic use and resistance development |
| | | Fund placebo-controlled trials to evaluate the necessity of antibiotic therapy for selected diseases |
| | | Explore with FDA opportunities to streamline antibiotic drug development |

*Source*: Adapted from Infectious Disease Society of America, "Bad Bugs, No Drugs: As antibiotic discovery stagnates, a public health crisis brews," July 2004, www.idsociety.org, accessed 17 May 2008.

libraries. A possible new source is the use of genomics to look for potential antimicrobials among bacterial DNA [17]. Soil and fermentation sampling have the advantage of providing a vast number of molecules to screen simultaneously but often rediscover compounds already in use [36]. A large pharmaceutical company may have over 100,000 compounds available for an automated high-throughput screen (HTS). Genomic applications to antibiotic design may have the difficulty of interpreting what a particular piece of DNA does and has yet to pay off with a marketed antibiotic. If a "hit" occurs with a compound from any of the above sources, the molecule often must then be modified by medicinal chemists to decrease human toxicity, enhance bacterial killing, improve entry into the bacteria cell, prevent efflux out of the bacterial cell, or to broaden the spectrum of activity against bacteria.

Screens fall along two major classes: looking for growth inhibition from whole-cell bacteria and looking for inhibition of single molecular targets. A hybrid between source and screen is structural-based design [37], which uses the molecular structure of the bacterial target and computers to test various compounds for activity against it in silico [38]. Whole-cell screens have the advantage of only selecting products that have the ability to penetrate the bacterial cell wall and reach their metabolic target. Genomics is used for screens as well as bits of DNA are examined for potential targets, but suffer the same problem as when used as a source of molecules: What does a particular piece of DNA do and is it really applicable to the job at hand? The tried-and-true method—using soil samples as a molecular source in a whole-cell bacteria screen—worked well in the 1950s and 1960s when six new classes of antibiotics

were developed, but now no longer appears sufficient [19]. Novel ways of locating compounds and the use of more sensitive screens appear to be the next phase of antimicrobial drug discovery [22].

### 3.1 Sources of Compounds: Digging in Dirt

Bacteria themselves are the source of many of our current antibiotics [29]. One gram of soil can harbor some 1000 different species of bacteria [22], and many of them secrete substances toxic to their neighbors to get a survival edge. For example, chloramphenicol, clavulanic acid, erythromycin, gentamicin, rifampin, teicoplanin, tetracycline, and vancomycin all were isolated from soil bacteria [39]. All the easy ore, however, may have already been taken from this mine. Current soil screening laboratories often rediscover antibiotics already in use [17, 36], frustrating scientists looking for new compounds.

Many scientists are thinking of innovative ways to search the soil for potential agents. Bacteria are sought out in different locations, such as in lichen, on seaweed, in Alaskan soil, or in deep-sea mud. Advanced culture techniques are used to coax fastidious organisms to thrive in the test tube, since 99% of soil bacteria have not been cultured in vitro. Bacterial DNA can also be directly amplified from a soil sample, creating a genomics profile that is a potential blueprint for future antibiotics [17, 22].

### 3.2 Genomics: Stealing the Blueprints

The genomes of several bacteria have been mapped out. The science of genomics can use this or other information to sift through bacterial DNA looking for genes that may create molecules used against other bacteria [17]. Usually, scientists screen the DNA code by homology—looking for bits of DNA similar to the code of known antibiotics. Although no antibiotic in use has been discovered by this method, the science is still young and the hope is great.

## 4 THE SCREEN DOOR

The screen used to weed out useless compounds from potential great antibiotics may be the most important intellectual piece of a drug development program [18]. A screen too sensitive will swamp the development team with too many hits, yet a screen too specific will ignore compounds that, with some minor molecular tweaking, could have great potential.

### 4.1 Bacterial Whole-Cell Screens

The whole-cell screens—looking for compounds that inhibit growth of living bacteria—have the distinct advantage of finding only compounds that will reach their target (through the cell wall without being inactivated or pumped out) [19]. Whole-cell screens using resistant bacteria are likely to ignore known compounds already in use, allowing for focus on compounds useful against these problematic bacteria [22]. Nevertheless, such screens were not amenable to automated HTS, and molecular screens rose in popularity [17]. These in vitro screens often use a single metabolic reaction to test compounds against, often at an enormous rate. No antibiotics in use today arose from HTS [19]. HTS might not be amenable to antibiotic screening at all, as it is five times less likely to generate an antibiotic hit than when used as a screen for drugs for mammalian cells [30].

Merck has used a novel addition to the whole-cell screen to detect low levels of inhibitory compounds in soil or fermentation samples. Antisense RNA can lower levels of target enzymes in bacteria, making them more susceptible to minute levels of antibiotics that might have been previously missed in soil samples. Such efforts have yielded the potential drugs platensimycin and platencin, compounds that inhibit enzymes in the bacterial fatty acid synthase pathway. These molecules can kill MRSA in vitro and in the mouse model [22, 40–42]. These potential drugs were discovered by modifying the whole-cell screen in two fashions: (1) antisense DNA was used to weaken an enzyme system to make it more susceptible to minute levels of compounds and (2) the use of the fatty acid synthesis pathway as a screening target [42].

### 4.2 Targeting the Defenses: Inactivating Modes of Bacterial Resistance

Bacteria can become resistant to antibiotics via several pathways. They can degrade the drug, remove it from the inside of the cell via efflux pumps, modify the drug's target, reduce uptake of the drug, or overproduce the enzymes inhibited by the drug [29, 43–45]. One should not think of resistance as wholly a response to human use and overuse of antibiotics. Bacteria have been using them against each other for eons and already have countermeasures in place [39].

### 4.3 The ß-Lactamase Inhibitors

Targeting these countermeasures has worked spectacularly well for over 20 years in one specific area: the ß-lactamase inhibitors. As discussed above, bacterial

ß-lactamases cleave and inactivate penicillin and its derivatives. Inhibitors of these enzymes bind irreversibly and inactivate them. ß-Lactamase inhibitors are structurally closely related to the intended substrate of the enzyme—the ß-lactam antibiotic—possibly making it difficult for mutations to occur that will differentially ignore the inhibitor and still bind the substrate [46]. Perhaps this is why these drugs are still widely used and effective today.

Over 300 different ß-lactamases with different spectra of activity leading to more narrow or broader resistance have been described. They fall roughly into four categories: Class A, B, C, and D. Of the four classes, only class A is inhibited by clavulanic acid, sulbactam, and tazobactam (the commercially available ß-lactamase inhibitors) with any degree of consistency. New research into the field has yielded some prospects of inhibiting other classes of ß-lactamase inhibitors, but these currently are far from clinical trials [45].

### 4.4 The Efflux Pumps

Bacterial efflux systems pump harmful antibiotics and biocides out of the cytoplasm and periplasmic space and prevent them from reaching their targets [47, 48]. They were first recognized as mechanisms of antimicrobial resistance in 1980 [49]. Efflux pumps can protect bacteria by increasing their number in the membranes [50] or by mutations allowing them to pump out a broad array of antibiotics. The latter may confer multidrug resistance (MDR) status on a bacterium. Bacteria that produce antibiotics often use efflux pumps to render themselves immune to their own toxins [48]. Efflux pumps have contributed to resistance of many different antibiotics and biocides, including chloramphenacol [51, 52], metronidazole [53], quinolones [54, 55], macrolides [56], carbapenems [57, 58], streptogramins [59], tetracyclines [59, 60], linezolid [61], and chlorhexidine [62].

Poisoning bacterial efflux pumps appears to work in restoring antibiotic concentrations within the cell [63, 64] and successfully resensitizing bacteria in vitro to the antibiotic of concern [53, 58, 65]. Efflux pump inhibitors (EPIs) may also find a place in treatment of infections caused by relatively sensitive bacteria as a suppressor of resistance development [47, 66]. Few agents are approaching clinical trials, although the old antihypertensive drugs reserpine and verapamil [67], the H+/K+ adenosine triphosphatase (ATPase) pump inhibitors omeprazole and lansoprazole [68], and the two phenothiazines chlorpromazine and thioridazine appear to inhibit some efflux pumps [69, 70]. Unfortunately, drug levels required to inhibit efflux pumps are much higher than those normally used clinically [48].

Several screens have been employed to find EPIs that could prove useful in resensitizing resistant bacteria. Structural alterations of antibiotics to turn them into EPIs has met with some success in vitro but not clinically [71, 72]. A second system grows efflux-using bacteria with subinhibitory concentrations of antibiotics as a screen for EPIs that restore antibiotic sensitivity. Many compounds have been isolated via this method, but much work to further characterize them lies ahead.

One EPI coupled with an antibiotic has made it to human trials. The compound MP-601,205 is aerosolized and used to treat pulmonary exacerbations in cystic fibrosis patients along with ciprofloxacin. The company, Mpex Pharmaceuticals, has not revealed the structure or mode of action of its drug, which is in phase II trials [48].

EPIs may prove to be difficult to develop, as efflux pumps are widely diverse among the different bacteria. Antibiotic efflux may be mediated by more than one pump in a single organism [73–75]. Also, mutations do occur to render EPIs inactive [76], and an EPI may inhibit passage of only some antibiotics through a multidrug efflux pump and leave others unaffected [47]. Nevertheless, we await further information regarding this avenue of research that may extend the life span of both existing and future antibiotics.

### 4.5 Targeting Their Weapons: Inactivating Bacterial Virulence Factors

If a bacterium invades your body and does no damage, would you care? Virulence factors are an excellent target for use as antimicrobial screens, and there are an enormous number of them to shoot at. For instance, drugs that block the effects of edema factor from *Bacillus anthracis* could have a large impact on patient outcome [37, 77, 78]. Many bacteria can sense the presence of each other, and once they reach a critical mass—a quorum—various changes in virulence can occur. Interference with quorum sensing is another target [79]. Biofilms often occur when bacteria colonize artificial material, such as joint replacements or intravascular or urinary catheters, and render bacteria relatively resistant to current antimicrobials. Up to half of nosocomial infections in the United States involve indwelling devices, and molecules that inhibit biofilm formation could have a huge patient impact [80]. There are many other examples of virulence factors that could potentially be used as screens [23].

### 4.6 Structural-Based Design: The Source and Screen Hybrid

Structural-based design of drugs first became a possibility with the advent of X-ray crystallography, but it is

only recently that it has begun to pay off with medications. Several antiviral drugs, including some human immunodeficiency virus (HIV) protease inhibitors [81, 82] and the influenza neuraminidase inhibitor zanamavir [83] are examples. To date, no antibiotics have come from this method, but there are many prospects [37].

Structural-based design used X-ray crystallography structure of a known target—in the above cases, either HIV-1's protease or influenza's neuraminidase—to search for compounds with a potential fit in an active site. Each possible fit, or hit, was then used in silico or in vitro HTS of similar compounds to improve binding and inhibition. Combinatorial chemistry could be used to modify molecules based on their predicted structural interactions with the binding site [38].

Although there are many prospects, no antibacterial or antifungals have come to the clinics via this method. One difficulty lies in the difference between a compound's minimum inhibitory concentration (MIC) for the target in question versus the MIC for a whole bacterium. A molecule can poison a particular enzyme or metabolic process quite well in silico or in vitro, but if the bacteria block its entry, kick it out, or destroy it before it can reach the target, then it is quite useless [17]. Chemistry must alter the compound's structure to allow it to pass through the cell wall and membrane safely without being pumped out or destroyed [38].

Structural biology is thus another tool in drug design and development. It can help to narrow the range of compounds to try and even help to locate new targets on a large molecule, but every hit must still be vetted by the in vitro, in vivo, and, finally, clinical trials processes.

Nevertheless, advances in this field may make it a far more useful tool. Better protein crystallization, X-ray crystallography, and software make it easier to prepare compounds for examination, to view them with higher resolution, and to manipulate them more easily in silico [38]. Genomics can allow an entire bacterium's DNA to be examined at once to search for potential targets over a vast array of proteins [18, 23]. Prediction of protein structure from DNA sequences continues to improve, and homology from other known protein structures allow for prediction of function and of the level of criticalness for a bacterium [37].

The preceding examples are just a few of the potential sources and screens of antibiotics that can help clinicians in the fight against bacteria.

## 5 PHARMACOKINETICS

The pharmacokinetic considerations in antimicrobial therapy do not differ greatly from those of other drugs. Drug concentrations must always be considered at the site of the infection, and the clinician must remember that plasma concentrations are only a surrogate marker for concentrations at other sites in the body.

### 5.1 Absorption

The bioavailability of oral and intramuscular formulations of antibiotics must be considered when they are given through these routes of administration. High oral bioavailability with a potent antibiotic is particularly desirable since these drugs can be given easily on an outpatient basis. Nevertheless, this is not the case with many antibiotics currently on the market. For β-lactams in particular, very few agents available orally can achieve the high concentrations of their intravenous counterparts.

Transitioning patients with serious infections from intravenous to oral therapy is a common practice in the hospital. Transitions to oral therapy take several forms [84]. A transition to an equipotent oral drug, such as from an intravenous cephalosporin to an oral fluoroquinolone, can be considered a therapeutically equivalent switch [84]. These differ from step-down therapy, where the oral drug being used is less potent than the intravenous drug [84]. This can occur if the oral drug has a lower bioavailability or other inferior pharmacokinetic parameters to the intravenous one. An example of this would be transitioning a patient from intravenous ampicillin to oral ampicillin, where the lower absorption and decreased tolerance of the oral drug makes the therapy a step-down from the intravenous form. This may be acceptable if the infection has largely resolved, but clinicians need to be aware that a change is taking place [84]. Some antibiotics with very low bioavailability nonetheless have oral utility. Some infections within the alimentary canal are better treated with oral rather than intravenous antibiotics. The oral absorption of vancomycin is very low, but this allows for very high vancomycin concentrations in the gastrointestinal tract and the efficacious treatment of *Clostridium difficile*–associated disease. Oral ampicillin is preferable to intravenous ampicillin for gastroenteritis caused by *Salmonella* species.

### 5.2 Distribution

Distribution is probably the least understood pharmacokinetic variable for many clinicians when choosing antibiotic therapy. The ability of the antibiotic to achieve an adequate concentration at the site of infection must always be considered. The blood–brain barrier (BBB) is an effective obstruction to the penetration of many antibiotics into the central nervous system (CNS) [85]. When the BBB is inflamed, such as in meningitis, CNS

penetration is easier to achieve [85]. Some antibiotics are useful in treating CNS infections even though they can only penetrate the BBB when it is inflamed.

Abscesses are localized collections of microorganisms and leukocytes that are self-contained. In general, antibiotics do not penetrate abscesses and surgical procedures are often required. Organs that many antibiotics have difficulty penetrating include the eyes, prostate, and bones.

The distribution of antibiotics may be variable between patients. It may be limited by patient factors, such as poor circulation to peripheral tissues in diabetic patients. Certain patient populations have inherently larger volumes of distribution, such as burn patients and patients with cystic fibrosis. This is relatively insignificant for antibiotics with very large volumes of distribution, but significantly affects the utilization of some drugs, such as aminoglycosides. The pharmacokinetics of many antibiotics in obesity are not well characterized.

### 5.3 Metabolism

As a class, relatively few antibiotics undergo considerable hepatic metabolism. Of the agents that do, few require significant dosing adjustment in liver disease. Pharmacokinetic drug interactions with antibiotics are mainly due to effects on the cytochrome P450 system [86]. Induction or inhibition of these enzymes is common among agents in a few classes of antibiotics, including azole antifungals, rifamycins, macrolides, and ketolides [86].

Extra-hepatic metabolism occurs with some antibiotics, though not commonly to an appreciable extent. Imipenem, a carbapenem antibiotic, is degraded in the kidney by renal dihydropeptidase into a nephrotoxic metabolite [87]. To counter this reaction, imipenem is co-formulated with cilastatin, an agent that inhibits this enzyme and allows for greater in vivo imipenem concentrations without the development of the nephrotoxic metabolite [87]. Anidulafungin is an echinocandin antifungal that is not hepatically metabolized or renally eliminated, but slowly degrades to inactive products in plasma [88].

### 5.4 Excr\etion

Renal excretion is the predominant method of elimination for many, if not most, antibiotics. Variations in the rate of elimination between patients can lead to significant accumulation of active drugs, and doses often need to be adjusted to account for these differences. Careful phase I pharmacokinetic studies need to be performed to establish correct doses for further clinical study, particularly for drugs with dose-related toxicity. Imipenem, the carbapenem mentioned above, was originally given in higher doses than are currently used. After seizures were noted in some patients, further investigations led to more accurate dosing in patients with renal dysfunction, as it was determined that elevated concentrations of the drug were responsible for this adverse effect [89]. Seizures associated with imipenem are uncommon when it is dosed appropriately [89]. One therapeutic advantage of renally eliminated antibiotics is that they are often useful for the treatment of urinary tract infections.

Antibiotics may also be excreted through the biliary route. Many drugs eliminated this way have the advantage of not requiring dose adjustment in renal dysfunction, which is a much more prevalent problem than hepatic dysfunction. However, this can eliminate their therapeutic potential for the treatment of infections of the lower urinary tract, including the very common infection of cystitis. Moxifloxacin is a fluoroquinolone that, unlike other fluoroquinolones, is eliminated through the biliary tract [90]. It has not been studied formally for the treatment of urinary tract infections, but it is unlikely to reach concentrations high enough to be effective in the treatment of cystitis. This makes moxifloxacin easy to dose in patients with renal dysfunction but lowers its utility in treating cystitis, therefore missing an important target market. Some drugs have dual elimination routes, eliminated via both renal and hepatic excretion. Ceftriaxone is a third-generation cephalosporin that is eliminated this way [91]. One of the reasons for the popularity of ceftriaxone is that its dual elimination does not require adjustment in renal failure, allowing for ease in dosing. This is a potential advantage from the viewpoint of the prescriber, but not necessarily the pharmacy department since drugs that need to be renally adjusted are given in lower doses in patients with renal dysfunction, lowering direct drug costs. In addition to not needing dose adjustment in renal disease, ceftriaxone reaches concentrations in the kidney and bladder that are adequate for treating urinary tract infections [91].

## 6 PHARMACODYNAMICS

In addition to considering the pharmacokinetics of an antimicrobial agent, the pharmacodynamics need to be considered. Pharmacodynamics is the field of study that evaluates the effects of drugs at given concentrations on patients, diseases, and in the case of infectious diseases, on microorganisms [92, 93]. That is, where pharmacokinetics is the study of "what the body does to the drug," pharmacodynamics studies "what the drug does to the body."

The study of antimicrobial pharmacodynamics has expanded considerably in recent years. Earlier studies of infected patients established doses of antimicrobials for infections based on observed effects—when a drug was given at a certain dose, infections either were or were not cured. Advances in microbiology and pharmacodynamics have coincided, and the ability to monitor concentrations of drugs and relate them to their effects is what has given birth to the field of modern antimicrobial pharmacodynamics.

### 6.1 Bactericidal versus Bacteriostatic Effects

One primary pharmacodynamic effect of antimicrobials is whether a given antimicrobial exerts a bactericidal or bacteriostatic effect at commonly achieved concentrations in the body. Bactericidal (or fungicidal) drugs actively kill bacteria at achievable concentrations [93]. The concentration at which the number of bacteria in a sample is reduced by 99.9% is referred to as the *minimum bactericidal concentration* (MBC) [93]. Bactericidal drugs readily achieve these concentrations in vivo at the site of infection. This contrasts with bacteriostatic or fungistatic drugs, which inhibit bacterial or fungal growth [94]. These drugs rely on the function of the immune system to kill the bacteria, and for that reason are not preferred in disease states where the immune response in inadequate, such as in meningitis, febrile neutropenia, or endocarditis [95]. For most indications, however, bacteriostatic activity is adequate [95]. The lowest concentration at which bacterial growth is inhibited is called the *minimum inhibitory concentration* (MIC) [95]. All organism–drug combinations have both MICs and MBCs associated with them. For a drug to be active against a given organism in a given infection, it has to achieve concentrations above the MIC of the organism at the site of the infection. If a drug is bactericidal against an organism in an infection, it had to achieve concentrations above the MBC, which is always above the MIC [93, 95]. If the MBC and MIC are close, the drug is likely to be bactericidal. When MICs are relatively low and MBCs are relatively high, a drug is bacteriostatic. If the MIC for an organism is not achievable for a drug, then that drug will be inactive against that organism.

MICs are much more easily measured by microbiology labs than MBCs and are therefore the values used to guide pharmacotherapeutic decisions in clinical practice [95]. Hence, clinicians often rely on historical data to determine whether a drug is likely to be bactericidal against an organism, not patient-specific data. However, the ease in testing MICs makes susceptibility testing for most organisms relatively easy. Standards for the definition of "susceptible" are defined by regulatory agencies based on achievable drug concentrations. Microbiology labs compare the MICs of given clinical isolates to determine the likelihood of successful therapy by various antibiotics and report these to clinicians.

### 6.2 Concentration-Dependent and Time-Dependent Drugs

One of the most important pharmacodynamic considerations is the development and use of an antibiotic is to determine the type of activity the agent possesses [94]. *Concentration-dependent* drugs have antimicrobial therapy that is dependent on high concentrations in vivo relative to the MIC of the infecting organism [94]. Concentration-dependent drugs are maximized when maximum concentration to MIC ($C_{max}$:MIC) ratios are 10:1 or more [96]. This is achieved by giving high doses of the antibiotic, generally relatively infrequently. The dosing schedule is based on the rate of elimination of the drug ($K_e$) and the ability of the host (patient) to tolerate high concentrations of the antibiotic. Ideal $C_{max}$:MIC ratios may involve supratherapeutic concentrations that lead to undesirable effects in patients, though adverse effects are not necessarily concentration dependent.

Concentration-dependent drugs contrast with *time-dependent* drugs, which have activity that is best defined by the time that the drug concentration exceeds the MIC of the organism [94]. Time-dependent drugs are also called concentration-independent drugs since the concentration generally has minimal effect on efficacy as long as it remains above the MIC. The efficacy of time-dependent drugs is increased by maximizing the time that the concentration at the site of the infection is above the MIC (time > MIC) [92, 93]. High concentrations are not more effective than lower concentrations just above the MIC, at least not proportionally, though they may expose patients to higher incidences of unwanted adverse effects. The ideal method of administering a time-dependent drug is by continuous infusion, a regiment that is pharmacodynamically optimized but often impractical. To compensate for practicality, most time-dependent drugs are given frequently to imitate the effects of a continuous infusion. For many time-dependent antibiotics, the time > MIC needed for maximal efficacy is less than 100% [93, 97]. This varies for both each class of antimicrobials and the organisms that cause infection (see Fig. 2).

### 6.3 Postantibiotic Effects

The antimicrobial effect of some antibiotics extends beyond the time that the drug is present at the site of infection. The time that organism growth is inhibited

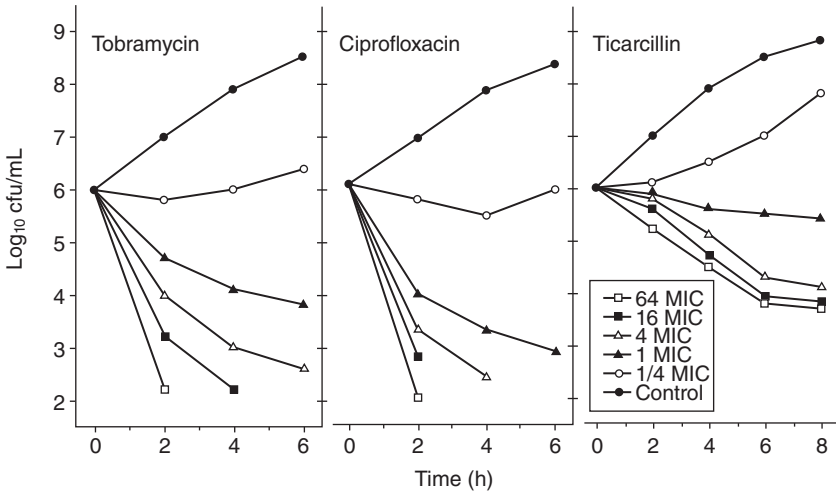


**Figure 2** Concentration versus time dependency [93]. Time–kill curves for *P. aeruginosa* ATCC (American Type Culture Collection) 27853 with exposure to tobramycin, ciprofloxacin, and ticarcillin at concentrations from one-fourth to 64 times the MIC.

despite drug concentrations below the MIC of the organism is referred to as the *postantibiotic effect* (PAE) [93]. During the PAE, the organisms are still recovering from the effects of the antibiotic exposure. PAEs may be concentration dependent, thus maximizing the dose of a drug may increase the PAE and allow for infrequent dosing, as with aminoglycosides [93, 96].

All antibacterials have a PAE against gram-positive bacteria, but PAEs against gram-negative organisms are less common [93]. PAEs measured in vitro are often less than in vivo PAEs due to the additional interactions of neutrophils and antimicrobial pharmacokinetics that do not exist in vitro [93]. Bacteria exhibiting the PAE after exposure to aminoglycosides and fluoroquinolones are more susceptible to destruction by leukocytes, thus prolonging the PAE in vivo [96, 98].

### 6.4 Examples of the Application of Antimicrobial Pharmacodynamics

***Aminoglycosides*** Aminoglycosides have been available since the discovery of streptomycin. The development and discovery of other aminoglycosides soon followed, and their broad-spectrum gram-negative activity and antipseudomonal nature led to much utilization of aminoglycosides for hospital-acquired infections. However, aminoglycosides have significant toxicities, most commonly dose-limiting nephrotoxicity [99]. Though aminoglycosides were known to be toxic above certain threshold trough concentrations and effective only if peak concentrations were reached in infected tissues, the ideal way of administering aminoglycosides was not known for many years.

These issues were improved upon by pharmacodynamic studies into both the toxicity and efficacy of aminoglycosides that followed their introduction years later. The uptake of aminoglycoside molecules into the nephrons in the kidney was found to be saturable [94]. Thus, once a threshold value of aminoglycoside exposure was reached within the nephron, greater concentrations had no effect. Also, aminoglycosides were found to be concentration-dependent drugs that were most efficacious with plasma concentrations 8–10 times greater than the MIC of the infecting organism (assuming adequate tissue penetration) [94, 99]. A dose-dependent postantibiotic effect was also noted with aminoglycosides [94, 99]. This led to a change in how aminoglycosides are dosed. Instead of multiple daily doses that lead to only moderately high peaks and somewhat low troughs, extended-interval dosing took advantage of the pharmacodynamics of the drugs. With extended-interval dosing, high doses of aminoglycoside are administred once a day or less [94, 99]. This maximizes the peak:MIC ratio and gives a trough concentration close to zero, allowing the kidney an aminoglycoside-free time period that seems to decrease nephrotoxicity [94, 99]. This dosing method has not become standardized and has not been studied in all patient populations, but it is common. Its efficacy and safety have been validated in multiple clinical trials and meta-analyses [94].

***Daptomycin*** Daptomycin is a cyclic lipopeptide antibiotic that is active against gram-positive cocci, including methicillin-resistant *Staphylococcus aureus* and vancomycin-resistant enterococci, two types of

organisms for which limited therapeutic options exist. It was originially developed in the late 1980s and early 1990s by Eli Lilly and Company and was studied at a dose of 4 mg/kg twice daily [100]. However, during clinical trials, high rates of elevations of creatine phosphokinase were found, indicating that skeletal muscle toxicity was occurring that could possibly lead to the dangerous condition of rhabdomyolysis [100]. Further development of the drug was halted until Cubist Pharmacueticals purchased the rights to study and market daptomycin years later. As daptomycin was found to be concentration dependent, Cubist studied the drug with a once-daily administration dosing regimen and did not find the toxicity described in early studies [100]. It was concluded that the likelihood of skeletal muscle toxicity was related to the frequency of administration, and the drug was approved for a once-daily regimen.

***Extended-Infusion β-Lactams*** β-Lactams are the largest class of antimicrobial drugs, encompassing the penicillins, cephalosporins, carbapenems, and monobactams. Antimicrobial resistance to β-lactams has become common among many bacteria, so methods to optimize the pharmacodynamics of these drugs have been sought out to maximize their utility [101]. Studies have been done with several β-lactams administered as either continuous infusions or extended infusions [101]. Continuous infusions are self-explanatory; during extended infusions the drug is administered over several hours instead of the customary 15 or 30 min. This prolongs the time that the concentration is above the MIC of an infecting organism, potentially increasing the efficacy of the drug in difficult-to-treat infections [101, 102]. An increase in efficacy with piperacillin/tazobactam in *Pseudomonas* infections has been reported, and other β-lactams are being studied with this administration technique to maximize the type of organisms and infections that they are effective in treating [101, 102].

## 7 PHARMACOGENETICS

Pharmacogenetics studies the interplay between human genetic differences and the effectiveness of drug therapy [103]. In the development of anti-infectives, pharmacogenetics is most useful in studying the variance in the elimination of drugs since the agents need to be effective against target sites in bacteria, not humans. Pharmacogenetics is a growing area of study and few antibiotics have undergone extensive pharmacogenetic profiling [103]. An exception to this is isoniazid, a hepatically metabolized drug with high interpatient variability [104]. Some individuals slowly eliminate isoniazid from the body and are termed "slow acetylators," while "fast acetylators" are the opposite. Certain patient populations are known to generally fall into one of these categories, but specific genetic testing has been studied to determine how an individual will metabolize isoniazid before it is administered [105]. This is not yet widely used but serves as an example of future testing that can be performed to determine which patients are most likely to suffer adverse effects from particular toxic drugs before they are given. In the world of HIV therapy, a test is now used to determine which patients are most likely to have a hypersensitivity reaction to abacavir, a commonly used antiretroviral drug [106]. Pharmacogenetic tests are likely to be most useful for drugs with high degrees of toxicity that are either dangerous or that have adverse effects that are difficult to monitor pharmacodynamically.

## 8 DEVELOPMENT AS A CHANGING NEED

As patient populations change, new infectious agents emerge, and antimicrobial resistance advances, the target areas for new antibiotics evolve. In this section, we highlight antifungal development as an example of where this has taken place, highlighting the need to continue drug development as patient populations change.

### 8.1 Antifungal Development

Systemic fungal infections continue to become increasingly prevalent [107, 108]. Once dismissed as a contaminant in blood cultures, *Candida* species have become the fourth most common cause of bloodstream infections [108]. This is due to many factors [107]. The patient population within hospitals has become sicker due to increased life spans, improved outpatient management, and managed care decisions [107]. Chemotherapy regimens have become more effective in treating cancer, but cancer patients are immunosuppressed for long periods of time [107]. Organ transplantation has become more common and effective, also leading to increased immunosuppression [107]. All of these factors have led to a patient population that is increasingly susceptible to infection by fungi who infrequently affect healthier patients.

Amphotericin B deoxycholate was the first systemic agent that was widely used for the treatment of systemic fungal infections. It is effective against many fungi but has notable toxicities that are intolerable for many patients [107]. Most notable among these are nephrotoxicity and infusion-related reactions. Years after its introduction, pharmaceutical manufacturers sought to attenuate these toxicities by producing lipid formula-

tions of amphotericin B that were less nephrotoxic, and some have decreased infusion-related reactions [107]. Another class of antifungals, the azoles, were also developed to combat the rising incidence of fungal infections. Early agents were hampered by poor bioavailability or limited effectiveness, but as manufacturers saw the market for antifungals increasing, they responded by developing superior agents [107]. The first of these was fluconazole, a drug with very high bioavailability, minimal adverse effects, and excellent activity against many types of yeast [107].

Able to be taken orally and much less toxic than amphotericin products, fluconazole was a true advance in antifungal pharmacotherapy. However, it was not active against molds, a class of fungi that mainly infects only highly immunosuppressed patients [107]. As this population grew, itraconazole was developed to overcome the lack of mold activity of fluconazole. Itraconazole is active against many common molds, but its limitations were mainly pharmacokinetic [107]. Itraconazole has poor bioavailability in the original capsule formulation, and an intravenous formulation was not introduced for years after its introduction [107].

To address the shortcomings of itraconazole, voriconazole was introduced in the early 2000s. Voriconazole has excellent bioavailability and was available intravenously from its introduction [107]. Its manufacturer studied its utility in aspergillosis (the most common invasive mold infections) and showed superiority to the gold standard regimen of amphotericin B deoxycholate [109]. Though voriconazole has changed the way many mold infections are treated, there are still fungi against which it is not active, and posaconazole is the agent most recently introduced to the antifungal market to address some of these [110].

During the 1990s, the echinocandin class of antifungals was developed to address the issues of the increasing incidence of fluconazole-resistant *Candida* and to offer an alternative to amphotericin B formulations in the treatment of aspergillosis [107]. Marked by low toxicity, few drug interactions, and a broad spectrum of activity against yeasts and some molds, echinocandins were welcome alternatives to an antifungal market that had not seen a truly novel mechanism of action in several decades [88, 107]. Caspofungin was the first echinocandin to be marketed [88, 107]. After several years of market exclusivity, two other echinocandins were introduced: micafungin and anidulafungin [88]. All three of these agents are similar in terms of efficacy and safety, so a very competitive pricing market has developed.

These advances in antifungal development were led by the pharmaceutical industry's response to a growing need and resultant market. Development in antifungals had stalled for years before the echinocandins were introduced. The expansion of the immunocompromised population promises to result in a continued need for antifungal development for many years.

## REFERENCES


1. Fleming A (1929). On the antibacterial action of cultures of a pencillium, with special reference to their use in the isolation of *B. influezae*. *Br J Exp Path 10*, 226–236.
2. Lax E (2004). *The Mold in Dr Florey's Coat: The Story of the Penicillin Miracle*, 1st ed. New York: H. Holt.
3. Chain E, Florey HW, Gardner AD, Heatley NG, Jennings MA, Orr-Ewing J, et al. (1940). Penicillin as a chemotherapeutic agent. *Lancet 236*(6104), 226–228.
4. Lesch JE (2006). *The First Miracle Drugs: How the Sulfa Drugs Transformed Medicine*. New York: Oxford University Press.
5. *Nobel Lectures: Physiology or Medicine 1922–1941*. Amsterdam: Elsevier Publishing Company, 1965.
6. Rolinson GN (1998). Forty years of beta-lactam research. *J Antimicrob Chemother 41*(6), 589–603.
7. *Nobel Lectures: Physiology or Medicine 1942–1962*. Amsterdam: Elsevier Publishing Company, 1964.
8. Rolinson GN, Sutherland R (1973). Semisynthetic penicillins. *Adv Pharmacol Chemother 11*, 151–220.
9. Rolinson GN (1991). Evolution of beta-lactamase inhibitors. *Rev Infect Dis 13*(Suppl 9), S727–732.
10. Rolinson GN (1986). Beta-lactam antibiotics. *J Antimicrob Chemother 17*(1), 5–36.
11. Geddes AM, Klugman KP, Rolinson GN (2007). Introduction: Historical perspective and development of amoxicillin/clavulanate. *Int J Antimicrob Agents 30*(Suppl 2), S109–112.
12. Barrett FF, McGehee RF Jr, Finland M (1968). Methicillin-resistant *Staphylococcus aureus* at Boston City Hospital. Bacteriologic and epidemiologic observations. *N Engl J Med 279*(9), 441–448.
13. Boyce JM, Cookson B, Christiansen K, Hori S, Vuopio-Varkila J, Kocagoz S, et al. (2005). Methicillin-resistant *Staphylococcus aureus*. [see comment]. *Lancet Infect Dis 5*(10), 653–663.
14. Sabol KE, Echevarria KL, Lewis JS 2nd (2006). Community-associated methicillin-resistant *Staphylococcus aureus*: New bug, old drugs. *Ann Pharmacother 40*(6), 1125–1133.
15. Bazan JA, Martin SI, Kaye KM (2011). Newer beta-lactam antibiotics: doripenem, ceftobiprole, ceftaroline, and cefepime. *Med Clin N Am 95*(4), 743–760.
16. Biek D, Critchley IA, Riccobene TA, Thye DA (2010). Ceftaroline fosamil: A novel broad-spectrum cephalosporin with expanded anti-gram-positive activity. *J Antimicrob Chemother* (65 Suppl 4), 9–16.

17. Monaghan RL, Barrett JF (2006). Antibacterial drug discovery—Then, now and the genomics future. *Biochem Pharmacol 71*(7), 901–909.
18. Zheng C, Han L, Yap CW, Xie B, Chen Y (2006). Progress and problems in the exploration of therapeutic targets. *Drug Discov Today 11*(9–10), 412–420. Erratum *11*(15–16), 717.
19. Projan SJ, Bradford PA (2007). Late stage antibacterial drugs in the clinical pipeline. *Curr Opin Microbiol 10*(5), 441–446.
20. Centers for Disease Control and Prevention (CDC) (2000). Monitoring hospital-acquired infections to promote patient safety—United States, 1990–1999. *MMWR Morb Mortal Wkly Rep 49*(8), 149–153.
21. IDSA (2004). *Bad bugs, no drugs: As antibiotic discovery stagnates, a public health crisis brews*. Available: http://idsociety.org/Content.aspx?id=5558. Accessed May 7, 2008.
22. von Bubnoff A (2006). Seeking new antibiotics in nature's backyard. *Cell 127*(5), 867–869.
23. Payne D, Tomasz A (2004). The challenge of antibiotic resistant bacterial pathogens: The medical need, the market and prospects for new antimicrobial agents. *Curr Opin Microbiol 7*(5), 435–438.
24. Zavascki AP, Goldani LZ, Li J, Nation RL (2007). Polymyxin B for the treatment of multidrug-resistant pathogens: A critical review. *J Antimicrob Chemother 60*(6), 1206–1215.
25. Falagas ME, Bliziotis IA (2007). Pandrug-resistant gram-negative bacteria: The dawn of the post-antibiotic era? *Int J Antimicrob Agents 29*(6), 630–636.
26. James JS (2004). Empty antibiotic pipeline critically endangers public: IDSA report. *AIDS Treat News Aug 27*(404), 7.
27. Nordmann P, Naas T, Fortineau N, Poirel L (2007). Superbugs in the coming new decade; multidrug resistance and prospects for treatment of *Staphylococcus aureus*, *Enterococcus* spp. and *Pseudomonas aeruginosa* in 2010. *Curr Opin Microbiol 10*(5), 436–440.
28. Bush K, Macielag M, Weidner-Wells M (2004). Taking inventory: Antibacterial agents currently at or beyond phase 1. *Curr Opin Microbiol 7*(5), 466–476.
29. Cottarel G, Wierzbowski J (2007). Combination drugs, an emerging option for antibacterial therapy. *Trends Biotechnol 25*(12), 547–555.
30. Payne DJ, Gwynn MN, Holmes DJ, Pompliano DL (2007). Drugs for bad bugs: Confronting the challenges of antibacterial discovery. *Nat Rev Drug Discov 6*(1), 29–40.
31. Pearson H (2006). Antibiotic faces uncertain future. *Nature 441*(7091), 260–261.
32. Clarke T (2003). Drug companies snub antibiotics as pipeline threatens to run dry. *Nature 425*(6955), 225.
33. Norrby SR, Nord CE, Finch R, European Society of Clinical Microbiology and Infectious Diseases (2005). Lack of development of new antimicrobial drugs: A potential serious threat to public health. *Lancet Infect Dis 5*(2), 115–119.
34. MacKenzie FM, Struelens MJ, Towner KJ, Gould IM, ARPAC Steering G, ARPAC Consensus Conference P (2005). Report of the consensus conference on antibiotic resistance; prevention and control (ARPAC). *Clin Microbiol Infect 11*(11), 938–954.
35. Talbot GH, Bradley J, Edwards JE Jr, Gilbert D, Scheld M, Bartlett JG, et al. (2006). Bad bugs need drugs: An update on the development pipeline from the antimicrobial availability task force of the infectious diseases society of America [see comment]. *Clin Infect Dis 42*(5), 657–668. Erratum *42*(7), 1065.
36. Baltz RH (2006). Marcel faber roundtable: Is our antibiotic pipeline unproductive because of starvation, constipation or lack of inspiration? *J Ind Microbiol Biotechnol 33*(7), 507–513.
37. Barker JJ (2006). Antibacterial drug discovery and structure-based design. *Drug Discov Today 11*(9–10), 391–404.
38. Scapin G (2006). Structural biology and drug discovery. *Curr Pharm Des 12*(17), 2087–2097.
39. Tomasz A (2006). Microbiology. Weapons of microbial drug resistance abound in soil flora. *Science 311*(5759), 342–343.
40. Wang J, Soisson SM, Young K, Shoop W, Kodali S, Galgoci A, et al. (2006). Platensimycin is a selective FabF inhibitor with potent antibiotic properties. *Nature 441*(7091), 358–361.
41. Wang J, Kodali S, Lee SH, Galgoci A, Painter R, Dorso K, et al. (2007). Discovery of platencin, a dual FabF and FabH inhibitor with in vivo antibiotic properties. *Proc Natl Acad Sci USA 104*(18), 7612–7616.
42. Wright HT, Reynolds KA (2007). Antibacterial targets in fatty acid biosynthesis. *Curr Opin Microbiol 10*(5), 447–453.
43. Havlichek D, Stein G (1999). Sulfonamides. In Yu VL, Merigan TC Jr, Barriere SL, Eds. *Antimicrobial Therapy and Vaccines*. Baltimore: Williams & Wilkins, pp. 972–977.
44. Nicolau DP, Quintiliani R Jr (1999). Aminoglycosides. In Yu VL, Merigan TC Jr, Barriere SL, Eds. *Antimicrobial Therapy and Vaccines*. Baltimore: Williams and Wilkins, pp. 621–638.
45. Buynak JD (2006). Understanding the longevity of the beta-lactam antibiotics and of antibiotic/beta-lactamase inhibitor combinations. *Biochem Pharmacol 71*(7), 930–940.
46. Varghese JN, Smith PW, Sollis SL, Blick TJ, Sahasrabudhe A, McKimm-Breschkin JL, et al. (1998). Drug design against a shifting target: A structural basis for resistance to inhibitors in a variant of influenza virus neuraminidase. *Structure (Camb) 6*(6), 735–746.
47. Lomovskaya O, Bostian KA (2006). Practical applications and feasibility of efflux pump inhibitors in the clinic—A vision for applied use. *Biochem Pharmacol 71*(7), 910–918.

48. Lynch AS (2006). Efflux systems in bacterial pathogens: An opportunity for therapeutic intervention? An industry view. *Biochem Pharmacol 71*(7), 949–956.
49. McMurry L, Petrucci RE Jr, Levy SB (1980). Active efflux of tetracycline encoded by four genetically different tetracycline resistance determinants in *Escherichia coli*. *Proc Natl Acad Sci USA 77*(7), 3974–3977.
50. Maniati M, Ikonomidis A, Mantzana P, Daponte A, Maniatis AN, Pournaras S (2007). A highly carbapenem-resistant *Pseudomonas aeruginosa* isolate with a novel blaVIM-4/blaP1b integron overexpresses two efflux pumps and lacks OprD. *J Antimicrob Chemother 60*(1), 132–135.
51. Hansen LH, Jensen LB, Sorensen HI, Sorensen SJ (2007). Substrate specificity of the OqxAB multidrug resistance pump in *Escherichia coli* and selected enteric bacteria. *J Antimicrob Chemother 60*(1), 145–147.
52. Ping Y, Ogawa W, Kuroda T, Tsuchiya T (2007). Gene cloning and characterization of KdeA, a multidrug efflux pump from *Klebsiella pneumoniae*. *Biol Pharm Bull 30*(10), 1962–1964.
53. Pumbwe L, Chang A, Smith RL, Wexler HM (2007). BmeRABC5 is a multidrug efflux system that can confer metronidazole resistance in *Bacteroides fragilis*. *Microb Drug Resist 13*(2), 96–101.
54. Perichon B, Courvalin P, Galimand M (2007). Transferable resistance to aminoglycosides by methylation of G1405 in 16S rRNA and to hydrophilic fluoroquinolones by QepA-mediated efflux in *Escherichia coli*. *Antimicrob Agents Chemother 51*(7), 2464–2469.
55. Shin JH, Jung HJ, Kim HR, Jeong J, Jeong SH, Kim S, et al. (2007). Prevalence, characteristics, and molecular epidemiology of macrolide and fluoroquinolone resistance in clinical isolates of *Streptococcus pneumoniae* at five tertiary-care hospitals in korea. *Antimicrob Agents Chemother 51*(7), 2625–2627.
56. Cai Y, Kong F, Gilbert GL (2007). Three new macrolide efflux (mef) gene variants in *Streptococcus agalactiae*. *J Clin Microbiol 45*(8), 2754–2755.
57. Fish D, Yu VL, Merigan TC Jr, Barriere SL, Eds. (1999). *Antimicrobial Therapy and Vaccines*. Baltimore: Williams and Wilkins, pp. 690–703.
58. Walsh F, Amyes SG (2007). Carbapenem resistance in clinical isolates of *Pseudomonas aeruginosa*. *J Chemother 19*(4), 376–381.
59. Perdue B, Standiford H (1999). Tetracyclines. In Yu VL, Merigan TC Jr, Barriere SL, Eds. *Antimicrobial Therapy and Vaccines*. Baltimore: Williams and Wilkins, pp. 981–995.
60. Pantosti A, Sanchini A, Monaco M (2007). Mechanisms of antibiotic resistance in *Staphylococcus aureus*. *Future Microbiol 2*, 323–334.
61. Escribano I, Rodriguez JC, Llorca B, Garcia-Pachon E, Ruiz M, Royo G (2007). Importance of the efflux pump systems in the resistance of *Mycobacterium tuberculosis* to *fluoroquinolones and linezolid*. *Chemotherapy 53*(6), 397–401.
62. DeMarco CE, Cushing LA, Frempong-Manso E, Seo SM, Jaravaza TA, Kaatz GW (2007). Efflux-related resistance to norfloxacin, dyes, and biocides in bloodstream isolates of *Staphylococcus aureus*. *Antimicrob Agents Chemother 51*(9), 3235–3239.
63. Schelz Z, Martins M, Martins A, Viveiros M, Molnar J, Amaral L (2007). Elimination of plasmids by SILA compounds that inhibit efflux pumps of bacteria and cancer cells. *In Vivo 21*(4), 635–639.
64. Schumacher A, Trittler R, Bohnert JA, Kummerer K, Pages JM, Kern WV (2007). Intracellular accumulation of linezolid in *Escherichia coli*, *Citrobacter freundii* and *Enterobacter aerogenes*: Role of enhanced efflux pump activity and inactivation. *J Antimicrob Chemother 59*(6), 1261–1264.
65. Michalet S, Cartier G, David B, Mariotte AM, Dijoux-franca MG, Kaatz GW, et al. (2007). *N*-Caffeoylphenalkylamide derivatives as bacterial efflux pump inhibitors. *Bioorg Med Chem Lett 17*(6), 1755–1758.
66. Lomovskaya O, Warren MS, Lee A, Galazzo J, Fronko R, Lee M, et al. (2001). Identification and characterization of inhibitors of multidrug resistance efflux pumps in *Pseudomonas aeruginosa*: Novel agents for combination therapy. *Antimicrob Agents Chemother 45*(1), 105–116.
67. Yu JL, Grinius L, Hooper DC (2002). NorA functions as a multidrug efflux protein in both cytoplasmic membrane vesicles and reconstituted proteoliposomes. *J Bacteriol 184*(5), 1370–1377.
68. Aeschlimann JR, Dresser LD, Kaatz GW, Rybak MJ (1999. Effects of NorA inhibitors on in vitro antibacterial activities and postantibiotic effects of levofloxacin, ciprofloxacin, and norfloxacin in genetically related strains of *Staphylococcus aureus*. *Antimicrob Agents Chemother 43*(2), 335–340.
69. Louie A, Brown DL, Liu W, Kulawy RW, Deziel MR, Drusano GL (2007). In vitro infection model characterizing the effect of efflux pump inhibition on prevention of resistance to levofloxacin and ciprofloxacin in *Streptococcus pneumoniae*. *Antimicrob Agents Chemother 51*(11), 3988–4000.
70. Kristiansen MM, Leandro C, Ordway D, Martins M, Viveiros M, Pacheco T, et al. (2006). Thioridazine reduces resistance of methicillin-resistant *Staphylococcus aureus* by inhibiting a reserpine-sensitive efflux pump. *In Vivo 20*(3), 361–366.
71. Nelson ML, Park BH, Andrews JS, Georgian VA, Thomas RC, Levy SB (1993). Inhibition of the tetracycline efflux antiport protein by 13-thio-substituted 5-hydroxy-6-deoxytetracyclines. *J Med Chem 36*(3), 370–377.
72. Nelson ML, Levy SB (1999). Reversal of tetracycline resistance mediated by different bacterial tetracycline resistance determinants by an inhibitor of the tet(B) antiport protein. *Antimicrob Agents Chemother 43*(7), 1719–1724.
73. Lee A, Mao W, Warren MS, Mistry A, Hoshino K, Okumura R, et al. (2000) Interplay between efflux pumps

may provide either additive or multiplicative effects on drug resistance. *J Bacteriol 182*(11), 3142–3150.

74. Pumbwe L, Piddock LJ (2000). Two efflux systems expressed simultaneously in multidrug-resistant *Pseudomonas aeruginosa*. *Antimicrob Agents Chemother 44*(10), 2861–2864.
75. Hocquet D, Llanes C, Patry I, El Garch F, Plesiat P (2004). Two efflux systems expressed simultaneously in clinical pseudomonas aeruginosa. *Pathol Biol (Paris) 52*(8), 455–461 (in French).
76. Ahmed M, Borsch CM, Neyfakh AA, Schuldiner S (1993). Mutants of the *Bacillus subtilis* multidrug transporter bmr with altered sensitivity to the antihypertensive alkaloid reserpine. *J Biol Chem 268*(15), 11086–11089.
77. Karginov VA, Nestorovich EM, Moayeri M, Leppla SH, Bezrukov SM (2005). Blocking anthrax lethal toxin at the protective antigen channel by using structure-inspired drug design. *Proc Natl Acad Sci USA 102*(42), 15075–15080.
78. Sherer K, Li Y, Cui X, Eichacker PQ (2007). Lethal and edema toxins in the pathogenesis of *Bacillus anthracis* septic shock: Implications for therapy. *Am J Respir Crit Care Med 175*(3), 211–221.
79. Wenzel RP (2004). The antibiotic pipeline—Challenges, costs, and values. *N Engl J Med 351*(6), 523–526.
80. Rice LB (2006). Unmet medical needs in antibacterial therapy. *Biochem Pharmacol 71*(7), 991–995.
81. Lapatto R, Blundell T, Hemmings A, Overington J, Wilderspin A, Wood S, et al. (1989). X-ray analysis of HIV-1 proteinase at 2.7 A resolution confirms structural homology among retroviral enzymes. *Nature 342*(6247), 299–302.
82. Miller M, Schneider J, Sathyanarayana BK, Toth MV, Marshall GR, Clawson L, et al. (1989). Structure of complex of synthetic HIV-1 protease with a substrate-based inhibitor at 2.3 A resolution. *Science 246*(4934), 1149–1152.
83. Varghese JN (1999). Development of neuraminidase inhibitors as anti-influenza virus drugs. *Drug Dev Res 46*(3–4), 176–196.
84. Cunha BA (2006). Oral antibiotic therapy of serious systemic infections. *Med Clin North Am 90*(6), 1197–1222.
85. Lutsar I, McCracken GH Jr, Friedland IR (1998). Antibiotic pharmacodynamics in cerebrospinal fluid. *Clin Infect Dis 27*(5), 1117–1127; Quiz 1128–9.
86. Pai MP, Momary KM, Rodvold KA (2006). Antibiotic drug interactions. *Med Clin North Am 90*(6), 1223–1255.
87. Rodloff AC, Goldstein EJC, Torres A (2006). Two decades of imipenem therapy. *J Antimicrob Chemother 58*(5), 916–929.
88. Cappelletty D, Eiselstein-McKitrick K (2007). The echinocandins. *Pharmacotherapy 27*(3), 369–388.
89. Koppel BS, Hauser WA, Politis C, van Duin D, Daras M (2001). Seizures in the critically ill: The role of imipenem. *Epilepsia 42*(12), 1590–1593.
90. Keating GM, Scott LJ (2004). Moxifloxacin: A review of its use in the management of bacterial infections. *Drugs 64*(20), 2347–2377.
91. Lamb HM, Ormrod D, Scott LJ, Figgitt DP (2002). Ceftriaxone: An update of its use in the management of community-acquired and nosocomial infections. *Drugs 62*(7), 1041–1089.
92. Craig W (1993). Pharmacodynamics of antimicrobial agents as a basis for determining dosage regimens. *Eur J Clin Microbiol Infect Dis 12*(Suppl 1), S6–8.
93. Craig WA (1998). Pharmacokinetic/pharmacodynamic parameters: Rationale for antibacterial dosing of mice and men. *Clin Infect Dis 26*(1), 1–10; Quiz 11–2.
94. McKinnon PS, Davis SL (2004). Pharmacokinetic and pharmacodynamic issues in the treatment of bacterial infectious diseases. *Eur J Clin Microbiol Infect Dis 23*(4), 271–288.
95. The importance of bactericidal drugs: Future directions in infectious disease. *Clinical Infectious Diseases 39*(9), 1314–1320.
96. Craig WA (1993). Post-antibiotic effects in experimental infection models: Relationship to in-vitro phenomena and to treatment of infections in man. *J Antimicrob Chemother 31*(Suppl D), 149–158.
97. Craig WA (2003). Basic pharmacodynamics of antibacterials with clinical applications to the use of beta-lactams, glycopeptides, and linezolid. *Infect Dis Clin North Am 17*(3), 479–501.
98. Craig WA, Redington J, Ebert SC (1991). Pharmacodynamics of amikacin in vitro and in mouse thigh and lung infections. *J Antimicrob Chemother 27*(Suppl C), 29–40.
99. Nicolau DP, Freeman CD, Belliveau PP, Nightingale CH, Ross JW, Quintiliani R (1995). Experience with a once-daily aminoglycoside program administered to 2,184 adult patients. *Antimicrob Agents Chemother 39*(3), 650–655.
100. Tedesco KL, Rybak MJ (2004). Daptomycin. *Pharmacotherapy 24*(1), 41–57.
101. Lodise TP, Lomaestro BM, Drusano GL, Society of Infectious Diseases P (2006). Application of antimicrobial pharmacodynamic concepts into clinical practice: Focus on beta-lactam antibiotics: Insights from the society of infectious diseases pharmacists. *Pharmacotherapy 26*(9), 1320–1332.
102. Lodise TP Jr, Lomaestro B, Drusano GL (2007). Piperacillin-tazobactam for *Pseudomonas aeruginosa* infection: Clinical implications of an extended-infusion dosing strategy. *Clin Infect Dis 44*(3), 357–363.
103. Eichelbaum M, Ingelman-Sundberg M, Evans WE (2006). Pharmacogenomics and individualized drug therapy. *Annu Rev Med 57*, 119–137.
104. Roy PD, Majumder M, Roy B (2008). Pharmacogenomics of anti-TB drugs-related hepatotoxicity. *Pharmacogenomics 9*(3), 311–321.
105. Vuilleumier N, Rossier MF, Chiappe A, Degoumois F, Dayer P, Mermillod B, et al. (2006). CYP2E1 genotype

and isoniazid-induced hepatotoxicity in patients treated for latent tuberculosis. *Eur J Clin Pharmacol 62*(6), 423–429.

106. Mallal S, Phillips E, Carosi G, Molina JM, Workman C, Tomazic J, et al. (2008). HLA-B*5701 screening for hypersensitivity to abacavir. *N Engl J Med 358*(6), 568–579.
107. Gallagher JC, MacDougall C, Ashley ES, Perfect JR (2004). Recent advances in antifungal pharmacotherapy for invasive fungal infections. *Expert Rev Anti Infect Ther 2*(2), 253–268.
108. Pfaller MA, Jones RN, Messer SA, Edmond MB, Wenzel RP (1998). National surveillance of nosocomial blood stream infection due to species of *Candida* other than *Candida albicans*: Frequency of occurrence and antifungal susceptibility in the SCOPE program. SCOPE participant group. Surveillance and control of pathogens of epidemiologic. *Diagn Microbiol Infect Dis 30*(2), 121–129.
109. Herbrecht R, Denning DW, Patterson TF, Bennett JE, Greene RE, Oestmann JW, et al. (2002). Voriconazole versus amphotericin B for primary therapy of invasive aspergillosis. *N Engl J Med 347*(6), 408–415.
110. Nagappan V, Deresinski S (2007). Reviews of anti-infective agents: Posaconazole: A broad-spectrum triazole antifungal agent. *Clin Infect Dis 45*(12), 1610–1617.

# ANTIMICROBIAL RESISTANCE IN BIOFILMS: A STICKY SITUATION

Shu Yeong Queck and Michael Otto
*National Institute of Allergy and Infectious Diseases, The National Institutes of Health, Bethesda, Maryland*

## 1 BIOFILMS—A CLINICAL AND INDUSTRIAL PERSPECTIVE

In many environments, microbial cells are found associated with surfaces and form conglomerate structures termed *biofilms* [1]. During human infections, the formation of biofilms has been implicated, for example, in periodontitis, otitis media, cystic fibrosis pneumonia, endocarditis, and osteomyelitis. In addition, biofilms are often found on indwelling medical devices such as prostheses and catheters, leading to a high incidence of infections associated with the use of such biomaterials [2]. It is estimated that at least 60% of all microbial infections involve biofilms [3]. For pathogens, the beneficial associations with biofilm formation include protection against stresses by antimicrobial agents and host defenses. As a consequence of its key role in pathogenesis, biofilm formation is a significant determinant of infectivity and persistence [3–5].

Outside the medical domain, biofilm formation causes serious problems for industries, particularly in food-processing and water distribution systems [6, 7]. For example, when biofilms are established on the interior of metal pipes in the production lines, accelerated corrosion of the pipes and contamination of pipe contents may occur. Additionally, increased tolerance of microorganisms in biofilms to disinfecting agents such as chlorine, peroxides, and quaternary ammonium compounds has been frequently reported. As a result, billions of dollars are spent annually to rectify biofilm-related repercussions in the industries [8].

The increased recognition of the impact that biofilm formation has in the clinical and industrial settings has led to an explosion in the field of biofilm research over the past two decades. These studies led to the generation of several hypotheses to account for the exceptional ability of biofilm populations to resist antimicrobial killings. In this review, the various mechanisms of antimicrobial resistance will be examined with a specific emphasis on the predominant pathogens involved in biofilm-associated infections, that is, staphylococci, *Pseudomonas aeruginosa*, and the fungal pathogen *Candida albicans*.

## 2 ANTIMICROBIAL RESISTANCE IN BIOFILMS DIFFERS FROM THAT IN PLANKTONIC CULTURE

It has become clear in several studies that mechanisms of resistance to antibiotics can differ significantly between biofilm and planktonic cells. A classic example is the resistance to β-lactam antibiotics such as ampicillin. In biofilms of *Klebsiella pneumoniae*, resistance to β-lactams was not correlated with the production of β-lactamase [9]. The penetration of ampicillin was impaired in the biofilms of wild-type cells, whereas biofilms formed by a β-lactamase mutant were fully penetrated by the antibiotic. Despite that, biofilm cells of the β-lactamase mutant were resistant to ampicillin, suggesting that mechanisms other than degradation by β-lactamase were responsible for the observed β-lactam resistance. The findings obtained in this and similar studies have prompted researchers to take a more detailed look at the specific mechanisms of antimicrobial resistance in biofilms (Fig. 1).

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

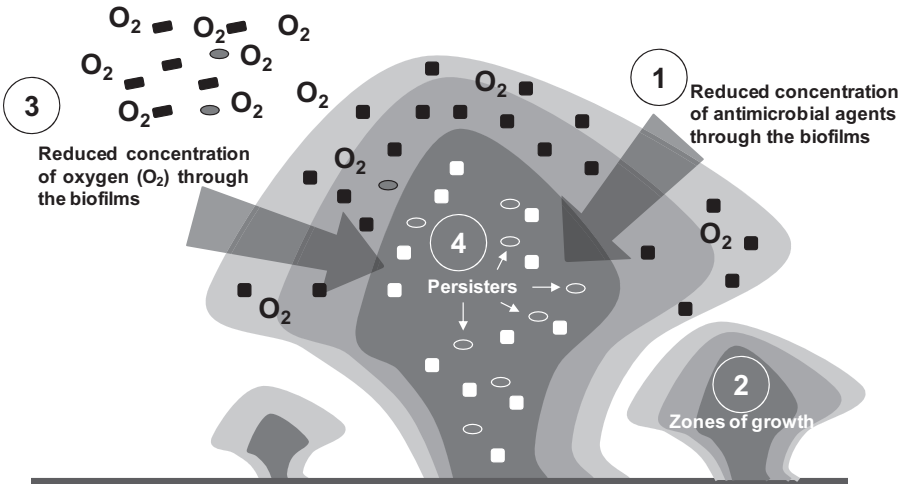


**Figure 1** Mechanisms of antimicrobial resistance in biofilms. (1) Penetration barrier. The penetration of some antimicrobials may be limited (reducing arrow) by their interactions with biofilm constituents (e.g. binding of antimicrobials to extracellular materials in the biofilm matrix and/or electrostatic repulsion between antimicrobials and the biofilm matrix). (2) Growth difference. Regions of biofilm exhibit different growth (light to dark; fast to slow growth), varying the efficacy of antimicrobials that predominantly target fast growing cells. (3) Biofilm physiology. The physiology of biofilm cells (■) is different from that of the planktonic (▬) counterparts (e.g. increased anaerobic respiration in biofilm cells), although some mechanisms of resistance may be similar to both modes of lifestyle. (4) Persister cells. The presence of persisters (⬭) that are resistant to a number of adverse environmental conditions.

## 3 RESTRICTIVE BARRIER TO ANTIMICROBIALS

The microbial biofilm matrix is a complex that consists primarily of water (up to 97%), microbial cells (2–5%), and a mixture of polymeric substances, which include polysaccharides, proteins, lipids, and nucleic acids [10, 11]. The matrix may act as a barrier in limiting the diffusion of molecules, including antimicrobial agents, into the biofilm, rendering biofilm cells resistant to antimicrobials. The presence of a diffusion barrier in biofilms has become evident in several early studies that showed limited penetration of antimicrobials into biofilms. For instance, the commonly used disinfectant chlorine was impaired in its penetration in a mixed *P. aeruginosa* and *K. pneumoniae* biofilm. Only 20% or less chlorine reached the deeper layers of the biofilm, compared to the bulk liquid [12]. Similarly, following treatment with monochloramine, a gradient of respiratory activity in the biofilm was observed. Cells in the deeper layer of the biofilm exhibited greater respiratory activity relative to lesser exposure to the biocide, compared to those at the biofilm–bulk liquid interface [13]. Finally, transport limitation of the fluoroquinolone antibiotic ciprofloxacin in *P. aeruginosa* biofilms was noted using attenuated total reflection/Fourier transform infrared (ATR/FT-IR) spectrometry [14]. However, the mechanism accounting for the limited penetration of the biocides into biofilms was not completely understood in these studies.

For some larger antimicrobials, the biofilm matrix may constitute a simple mechanical barrier. However, the penetration gradient of most antimicrobials in a biofilm is likely due to more specific interactions between antimicrobial agents with some of the biofilm constituents. For instance, sequestration by extracellular biofilm polymers may prevent the antimicrobial agents from penetrating to their targets. This has been reported for glucans in the periplasm of gram-negative bacteria [15]. A *P. aeruginosa* mutant disrupted in the synthesis of cyclic β-(1,3)-glucans formed biofilms that were more susceptible to the various antimicrobial agents tested. Furthermore, cationic aminoglycosides such as tobramycin have been shown to interact with the exopolysaccharide of *P. aeruginosa* [16], suggesting that the biofilm matrix may prevent the accumulation of antimicrobial agents by a sequestration mechanism that is likely dependent on electrostatic attraction.

In addition, electrostatic repulsion may account for limited penetration of antimicrobial agents in biofilms.

For example, resistance to cationic antimicrobial peptides (AMPs) in staphylococcal biofilms has been attributed to the unusual positive net charge of the staphylococcal exopolysaccharide. Polysaccharide intracellular adhesin (PIA) is a significant component of the staphylococcal biofilm matrix, consisting of β-1,6-linked glucosaminylglycan, which is synthesized and exported by the gene products of the intracellular adhesion (*ica*) operon [17, 18]. PIA production is important for biofilm formation, virulence, and protection of staphylococci against mechanisms of the human innate immune system, such as phagocytosis and killing by polymorphonuclear leucocytes (PMNLs), and human AMPs [19, 20]. Unlike most bacterial exopolymeric substances, PIA has an overall positive net charge due to the deacetylation of some of the *N*-acetylglucosamine monomers in the PIA polymer [20]. Due to the considerable addition of positive charges to the otherwise negatively charged staphylococcal cell surface by PIA, cationic antimicrobial agents (e.g., human AMPs LL-37 and β-defensin 3) are repelled, preventing their accumulation and access to cells within the biofilms. Together, these studies suggest that limited drug susceptibility in biofilms is at least in part due to interactions between biofilm constituents and antimicrobials, which prevent the penetration of drugs to their target site.

However, there are a number of studies to suggest that biofilm resistance is not solely due to a penetration barrier. For example, resistance to chlorine is likely not due to a penetration difference, as the diffusion coefficient of chlorine in biofilms is indistinguishable from that in pure water [21]. Similarly, biofilm resistance to antibiotics is not always due to limited penetration. For example, concentrations of the "last-resort" antibiotic vancomycin were detected in *Staphylococcus epidermidis* biofilms that exceeded the minimal inhibitory concentration (MIC) and minimal bactericidal concentration (MBC) for that organism [22]. Despite that, biofilm cells were not completely eradicated. Moreover, resistance to a variety of antifungal agents in biofilms of *Candida* species could not be explained by a penetration barrier [23]. Although the rate of penetration varied among the antifungal agents and the various biofilm species, all biofilms were fully penetrated by the antifungal agents tested, often at several folds above their MIC. Furthermore, drug resistance was evident in all the fungal biofilms examined. Therefore, the penetration of many antimicrobial agents is not limited in biofilms, and, thus, additional biofilm-specific mechanisms must be involved in the resistance of cells to those antimicrobials.

Taken together, the penetration of antimicrobial agents in biofilms is dependent on a multitude of factors, notably the structures of the biofilm matrix molecules and antimicrobial agents. Furthermore, it has become evident that the exceptional resistance of biofilm cells to antimicrobial agents is not only due to the restrictive barrier function of the biofilm matrix but that other mechanisms are involved.

## 4 A GRADIENT OF GROWTH

Cells in different zones of a biofilm grow at different rates. There is usually an inverse relationship between biofilm depth and growth activity. For example, biofilms of *K. pneumoniae* stained with acridine orange (nucleic acid stain) revealed spatial patterns of fluorescence [24], revealing that cells located within the deeper regions of the colony exhibited lower growth and vice versa. Not surprisingly, the availability of nutrient plays a major role for the development of such a gradient in growth. Sternberg et al. investigated the growth of cells in biofilms using a reporter fusion system. By tagging the growth-rate-regulated *rrnB*P1 promoter to a green fluorescent protein (GFP) reporter, the growth of *Pseudomonas putida* biofilms, ranging from individual cells to microcolony structures, was monitored. During the initial stages of biofilm development (e.g., attachment and surface colonization), similar intensity of fluorescence in cells was noted throughout the structures. As the biofilms matured, cells in the deeper regions exhibited reduced fluorescence (i.e., lower metabolic activities) relative to those closer to the biofilm–bulk fluid interface, displaying a gradient of growth differences. Part of this growth gradient observed was attributed to a depletion of nutrients as the biofilms expand. This hypothesis was supported by the fact that the gradient intensity became less pronounced upon introducing fresh nutrient [25]. These differences in growth may in part be due to the oxygen gradient that has been observed in biofilms. The levels of oxygen were significantly lower in the deeper regions of the biofilms compared to those closer to the bulk liquid interface [26].

As many antimicrobial agents are only active toward metabolically active cells, growth differences have a dramatic impact on antimicrobial resistance. In fact, several studies have correlated antimicrobial resistance with the metabolic status of cells in a biofilm. Such a correlation has been found both for those antimicrobial agents that result in the killing of the organism (e.g., by inhibition of cell wall synthesis such as by β-lactam antibiotics) and those that inhibit the organism's ability to propagate further (e.g., by inhibition of protein synthesis such as by tetracyclines). For instance, in *P. aeruginosa*, the regions in a biofilm in which antibiotics showed

the most pronounced activity corresponded to the sites of oxygen repletion and active metabolism, that is, the outer regions of the biofilm [27–29]. Interestingly, as the metabolic activity of biofilms of the fungus *C. albicans* increases rather than decreases with age, mature biofilms of *C. albicans* are more resistant to antimicrobials than younger biofilms—in contrast to what has been observed in bacteria [30]. However, these observations appeared to be rather specific for the *Candida* species among eukaryotic biofilms, as no such correlation could be established for biofilms of *Sacchromyces cerevisiae.* Thus, while the specific characteristics of growth-dependent biofilm resistance may differ from species to species, it appears that, in general, antimicrobial resistance in biofilms is due to a large extent to the reduced growth activity of the main biofilm body.

## 5 PHYSIOLOGICALLY DISTINCT POPULATION

We have seen that growth conditions and biofilm structure significantly affect biofilm resistance. In addition, recent global analysis of gene and protein expression has allowed us to determine specific factors that contribute to the resistance of biofilm communities to antimicrobials. Frequently, these factors determine the specific physiology of a biofilm and represent regulators that control the adaptation to the physiological requirements of biofilms. In addition, they may include resistance factors expressed during planktonic growth, but whose production or significance is greater in biofilms.

As a major result of transcriptome and proteome analyses, we now know that active, aerobic metabolism is reduced in bacterial biofilms, which is likely a consequence of the low levels of oxygen that have been reported in biofilms. Accordingly, in several studies, it has been found that genes involved in anaerobic respiration were upregulated in biofilms, while those associated with aerobic respiration were among those repressed [31–35]. As outlined above, the dormant state of cells in biofilms may mediate resistance to antimicrobial agents, particularly those that target growth-specific factors. In contrast to those earlier studies, however, biofilm resistance as a consequence of a shift in metabolism could now be pinpointed to specific genes and proteins. For instance, the enzyme formate dehydrogenase, a typical component of fermentative metabolism, is upregulated in biofilms of *Staphylococcus aureus*. The by-products of formate metabolism that are generated at an increased rate by formate dehydrogenase activity may generate an acidic milieu in biofilms, thereby reducing the efficacy of some antimicrobial agents [31, 34, 35]. In *P. aeruginosa*, the *nir* and *nap* gene clusters, which are required for respiratory nitrate reduction were upregulated in biofilms [33]. Anaerobic respiration is an important physiological adaptation in the pathogenesis of *P. aeruginosa*, in particular, for persistence in the hypoxic mucus environment of the airway lumens in cystic fibrosis [3, 36, 37]. Furthermore, the efficacy of aminoglycosides, such as tobramycin, is greatly reduced under anaerobic conditions. As we have noted already, in contrast to bacterial biofilms, cells in *C. albicans* biofilms are metabolically active [38–40]. Of particular interest in *C. albicans* biofilms is the increased expression of genes/gene products associated with sulfur metabolism, which is closely linked to the metabolism of glutathione, a molecule involved in the protection from oxidative stress in yeasts and filamentous fungi [41]. It has been suggested that glutathione may cause resistance to antifungal agents in *C. albicans* biofilms [39] as the alteration in glutathione levels has previously been correlated with the resistance of the eukaryotic pathogen *Plasmodium falciparum* to antimalarial drugs [42]. Hence, these studies suggest that the specific characteristics of biofilm physiology may reduce susceptibility to some antimicrobial agents.

Multidrug efflux pumps (MDPs) are involved in the extrusion of a wide spectrum of antibiotics such as aminoglycosides and cationic antimicrobial peptides [43]. Whereas in planktonic population their contribution to antimicrobial resistance has been well established, it is not known whether they contribute to antimicrobial resistance in biofilms. Although the genome of *P. aeruginosa* contains at least 10 MDPs [44], only 6 of them have been identified as important for antimicrobial resistance. Moreover, none of these 6 MDPs were differentially regulated in biofilms, suggesting that MDPs do not contribute to biofilm resistance in *P. aeruginosa* or resistance is due to yet uncharacterized MDPs [32, 45]. However, there appears to be a temporal role for MDPs in the resistance to antifungal agents in biofilms of *C. albicans* [46, 47]. For example, while there is no evidence for a role of MDPs in resistance to antifungal agents during the later stages of *C. albicans* biofilm development, MDP-mediated resistance to fluconazole is confined to planktonic and young (6 h) biofilm communities. In conclusion, however, there is currently no clear evidence to indicate a general involvement of MDPs in the antimicrobial resistance of biofilms, and, thus, these studies underline the notion that resistance mechanisms differ significantly between the planktonic and biofilm modes of growth.

Recent advances in our understanding of biofilm-specific physiology and gene expression suggest that the differential expression of regulatory factors plays a major role in biofilm resistance. Notably, several global regulators, such as those involved in the control of stress

response and expression of virulence determinants, are differentially regulated in biofilms. For instance, the environmental and infection stress response sensor, ClpC, which has previously been shown to be crucial for biofilm formation [48], was highly expressed in biofilms of *S. aureus*. ClpC and other Clp proteins belong to a family of closely related proteins, the Hsp100/Clp ATPases, which carry out diverse cellular functions. For example, they function as chaperones and are involved in the proteolysis of incorrectly folded protein substrates and the regulation of transcription [49]. In *S. aureus*, the Clp proteins have been shown to mediate stress responses (e.g., thermotolerance and salt tolerance) and replication in epithelial cells [48]. In addition to ClpC, the levels of other chaperones and oxidative stress proteins were higher in staphylococcal biofilms [34, 35]. Interestingly, major regulators of virulence in staphylocci, such as *agr* and *sarA* were downregulated in biofilms, causing lower levels of toxins and protease production in biofilms [34, 35, 50]. The loss in these regulator functions in staphylococci has previously been correlated with reduced activation of the host immune defense and resistance to antimicrobial agents [51–53]. In *P. aeruginosa*, the *Escherichia coli* stationary phase sigma factor, RpoS, was repressed in biofilms [32]. Although RpoS is responsible for the induction of starvation stress resistance and virulence response in *E. coli*, its role appears to control virulence in *P. aeruginosa* [54, 55]. RpoS predominantly induces genes coding for proteins involved in chemotaxis and two-component systems, whereas several heat shock proteins and secreted factors are among those repressed. Although the expression of RpoS has been correlated with the resistance of *P. aeruginosa* biofilms to tobramycin, the role of this sigma factor in biofilm resistance to antimicrobial agents is still poorly understood [32, 56]. Finally, recent studies have indicated a key role of cylic diguanosine monophosphate (c-di-GMP) in the control of switching bacterial physiology between the biofilm and planktonic modes of growth. Although the specific targets of c-di-GMP control remain yet to be determined, it is known that c-di-GMP is important for a number of cellular processes, including cell differentiation, multicellular behavior, and expression of virulence traits in a number of bacterial species [57]. The aminoglycoside response regulator (*arr*, formerly PA2818) of *P. aeruginosa* induces resistance to aminoglycosides and biofilm formation through the modulation of the intracellular levels of c-di-GMP [58]. Thus, it appears that increased levels of c-di-GMP lead to increased antimicrobial resistance by switching physiology to the typical physiological status of a biofilm. Taken together, these studies indicate that a variety of regulatory changes make biofilm cells better equipped than their planktonic counterparts to encounter adverse conditions.

Not surprisingly, a number of resistance mechanisms shown to protect planktonic cells from antimicrobial agents were also adopted in biofilms. However, in some cases there appears to be a more pronounced role of some mechanisms of resistance in the biofilm mode of growth, which may contribute to the overall higher antimicrobial resistance of biofilms. For instance, genes that code for specific proteins required for the synthesis of poly-γ-glutamate (PGA) complex in *S. epidermidis* were upregulated in biofilms [50]. PGA mediates high salt tolerance and protection from major components of the human immune system, including killing by antimicrobial peptides and PMNLs. Moreover, PGA is crucial for the persistence of *S. epidermidis* in device-related infections [59]. In *C. albicans*, resistance to antifungal agents in planktonic cells has been associated with alterations in the sterol composition of the cell membrane [60]. In biofilms, the low ergosterol levels and reduced transcript level of genes involved in ergosterol biosynthesis were found to correlate with increased resistance to antifungal agents such as amphotericin B and fluconazole [47, 61]. In addition, a number of proteins are expressed at higher levels in biofilms of clinical isolates that were resistant to azole antifungal agents [62]. Hence, these studies suggest that some resistance mechanisms that have previously been described in cells grown in planktonic culture play a crucial and possibly more important role in biofilm communities.

## 6 CONTRIBUTIONS OF THE PERSISTERS

A relatively new and controversial hypothesis claims that a subpopulation of dormant cells exists in biofilms that determines antimicrobial resistance. These cells have been called "persisters." Similar to spores and other resistant structures, persisters grow slower than the wild-type cells but are better at withstanding adverse environmental conditions such as those encountered during antibiotic treatment. Therefore, persisters may contribute to the continual propagation of the biofilm progeny following exposure to antimicrobial agents, resulting in recalcitrant infections [63]. Notably, persister cells are genetically identical to cells in the bulk population and are not exclusive to biofilms [64, 65]. Although they are typically present in higher numbers in stationary phase growth, cultures in earlier growth phases also contain persisters, emphasizing their specific and unique physiological status.

While there are many reports of persister cells in gram-positive and gram-negative bacteria, and recently *C. albicans* [66], little is known about the mechanism of

drug resistance conveyed by persisters and the regulation of persister cell formation. Recently, several multiple drug tolerance (MDT) genes were identified to be important for the generation of persister cells in *E. coli* [67, 68]. These genes include a presumable toxin/antitoxin module encoded by *hipBA* and genes involved in sn-glycerol-3-phosphate metabolism, in particular, the enzyme sn-glycerol-3-phosphate dehydrogenase, which is encoded by *glpD*. Importantly, the deletion of either of these gene loci resulted in a lower frequency of persister cells in stationary phase, while induced expression of their gene products enhanced the tolerance of the cell population to a range of antimicrobial agents [67, 68]. However, the modes of action of MDT genes are still elusive. According to one model, the toxin HipA offers a "protective" effect by binding to the cellular target sites of antimicrobial agents [63]. In this model, the binding of the toxin alters functionality of the target (e.g., ribosome), rendering it insensitive to the actions of the specific antimicrobial agent, thereby generating tolerance in cells. Furthermore, it has been suggested that the MDT genes involved in persister cell formation may be different among species. For example, although multiple drug-tolerant persister cells made up about 1% of the population in stationary phase *P. aeruginosa* cultures, no homologs of *hipBA* were found in this organism. Hence, more detailed investigation of the mechanism underlying the generation and MDT of persister cells in other organisms is warranted.

In *P. aeruginosa*, staphylococci and other bacterial species, another type of persister cells termed small colony variant (SCV) is frequently reported [69, 70]. Contrary to the persisters described by Lewis [63], SCVs of *P. aeruginosa* are heterogenous in growth and are typically isolated from biofilm cultures. Moreover, SCVs are often implicated in bacterial pathogenesis as they exhibit MDT levels several folds higher than the corresponding wild-type strains and may cause more persistent infections. Indeed, SCVs and biofilm formation have been made accountable for the persistence of *P. aeruginosa* in the lungs of cystic fibrosis (CF) patients despite long-term antibiotic therapy [3, 71–73]. Several recent studies have led to a better understanding of SCV development, the molecular mechanism underlying the generation of the RSCV phenotype, and its correlation with antibiotic resistance, and biofilm formation in *P. aeruginosa* [74]. The phenotype variant regulator, *pvrR*, was found to exhibit a multitude of functions, including the regulation of the switch in rough SCV (RSCV) to wild-type phenotype, attachment, and biofilm formation. More recently, Webb et al. proposed a model to explain the generation of SCVs in biofilms of *P. aeruginosa* by the involvement of the filamentous phage Pf4 [75]. The analysis of the Pf4 phage sequence revealed the presence of a toxin/antitoxin module and several genes that may contribute to the SCV phenotype. Similarly, small colony formation in *E. coli* has been attributed to filamentous phage infection [76]. Furthermore, transcriptional profiling of gene expression has revealed that a number of phage-related genes in *P. aeruginosa* were upregulated in biofilm communities [32, 33]. These findings suggest a general role for phage involvement in the generation of SCVs and bacterial biofilm development.

SCVs in staphylococci are different from those of *P. aeruginosa* and the persister cells described by Lewis [63]. Staphylococcal SCVs are slow growing and their emergence is often linked to mutations in metabolic pathways in response to challenges by antimicrobial agents, generating auxotrophs [77, 78]. The most commonly isolated and best characterized staphylococcal SCVs are those deficient in electron transport (ET) and thymidine biosynthesis. Compared to their isogenic parental strains, these SCVs are more resistant to cell wall specific agents such as β-lactam antibiotics and display reduced sensitivities to cationic agents such as aminoglycosides [79]. Deficiency in the ET pathway in staphylococcal SCVs can be caused by mutation in one of several single-gene loci [80–82]. For example, mutation in *menD* disrupts the synthesis of menadione, which in combination with isoprenoid lipid is required for the synthesis of menaquinone. Mutations in *hemB* and *ctaA* affect the synthesis of heme, which disrupts cytochrome biosynthesis. Comparatively, the genetic basis of thymidine-auxotrophic (TA) SCVs is still poorly characterized. TA SCVs are prevalent and highly persistent in the airway secretions of CF patients on long-term therapy with trimethoprim sulf-amethoxazole [83, 84].

Clinically, SCVs of staphylococci are involved in a multitude of persistent infections such as cystic fibrosis, osteomyelitis, and device-related infections [70]. SCV formation is crucial for staphylococcal pathogenesis, as SCVs exhibit enhanced persistence in host tissues [82, 85–87]. This is likely due to two factors that result in immune evasion: enhanced internalization into host cells and diminished expression of aggressive virulence factors such as hemolysins, leukocidins, and extracellular proteases [88, 89]. Additionally, in most of the staphylococcal TA SCVs, the expression of a number of major virulence regulators including *sarA*, the *agr* quorum sensing system, and the alternate sigma factor *sigB*, is decreased [90], which leads to a shift in virulence factor production. This loss of regulator function has been previously linked to reduced activation of host defenses and resistance of staphylococci to antimicrobial agents [51–53]. Interestingly, Vaudaux et al. showed that the observed enhanced internalization of staphylococcal SCVs by host cells may be due to higher expression of

surface adhesins, including fibronectin and fibrinogen-binding proteins, which play an important role during the internalization process [87]. Taken together, immune evasion by the generation of SCVs appears to represent one of the reasons for recalcitrant biofilm-associated staphylococcal infections.

## 7 NOVEL APPROACHES TO THE "STICKY PROBLEM"

Two approaches, namely prevention and eradication, have been proposed for the control of biofilm-related infections. First, making the surfaces less favorable for microbial colonization may prevent biofilm formation on medical devices. This may be achieved by the alteration of the physicochemical properties of the surface or by coating with substances that inhibit bacterial growth and/or biofilm formation. Second, biofilms may be disrupted after their formation, thereby converting bacterial communities to single cells, which can be more efficiently attacked by conventional antibiotics. This section will highlight some of the recent advances made with regard to preventing and removing microbial biofilms on surfaces.

### 7.1 Quorum Sensing Inhibitors

The term *quorum sensing* (QS) was initially coined to describe the regulation of bacterial gene expression in a cell-density-dependent fashion [91]. Recently, the QS phenomenon has also been described in fungi [92]. Notably, in many organisms QS has been shown to influence biofilm development. This has become clear in several studies that investigated phenotypic differences in the biofilms formed by QS mutants compared with their isogenic wild-type strains [93].

It has been frequently reported that a group of compounds termed furanones, which are produced by the marine algae *Delisea pulchra*, act by competing with acylated homoserine lactone (AHL)-type QS signals for binding to the cognate response regulator, thereby inhibiting QS controlled phenotypes [94–97]. Although the primary target of the furanones is the AHL regulatory system (AI-1) in gram-negative bacteria, they also appear to interfere with an alternative signaling system, AI-2, that is prevalent in gram-positive and gram-negative bacteria [98, 99]. In *P. aeruginosa*, furanone treatment resulted in the attenuation of QS-controlled virulence factor expression, causing a reduction in biofilm formation and the severity of infection [100, 101]. In addition, the pretreatment of biomaterials with furanones provided protection from microbial colonization and biofilm formation [102, 103]. Importantly, at least in gram negatives, the mechanisms of furanones are neither bacteriostatic nor bactericidal [104], which is believed to minimize the development of microbial resistance to these compounds.

The inhibition of QS control has also been suggested as a means to inhibit biofilm formation in eukaryotes. For example, the fungal pathogen *C. albicans* uses farnesol as a QS signal to regulate filamentation and biofilm formation [105, 106]. Transcriptional profiling of gene expression in *C. albicans* biofilms after treatment with farnesol revealed that many factors that may contribute to biofilm persistence, such as heat shock proteins, genes involved in cell wall maintenance, and genes associated with drug resistance were downregulated [107]. In addition, farnesol was shown to be active against other microbes [108, 109] and influenced, for example, the induction of apoptosis in *Aspergillus nidulans*, enhanced antimicrobial susceptibility, and inhibited biofilm formation in *S. aureus*. Thus, farnesol may be applicable in the control of biofilms formed by fungi and mixed bacterial–fungal populations.

While the idea of using QS antagonists offers a promising outlook in the combat of microbial biofilms, caution should be exercised when using such agents for therapeutic purposes. For example, furanones are structurally similar to the AHL molecules [104], which in addition to their regulatory role, possess immunomodulatory activity that can alter the outcome of an immune response [110–114]. Thus, the administration of high levels of furanones may potentially trigger undesirable immune consequences in the human host, leading to more severe development of certain diseases. Also, the effect of QS regulation in virulence and biofilm formation varies among microbes. For instance, the *agr* QS system in staphylococci negatively regulates biofilm formation [115, 116]. In addition to enhanced biofilm formation, interfering with QS signaling in staphylococci might drive undesirable consequences including persistence in device-related infections [117] and reduced susceptibility to antimicrobials [52, 53]. Hence, interfering with QS control as a therapeutic means to treat biofilm-associated infection should take into account the dosage of compounds administered and the specific pathogen and type of infections involved.

### 7.2 Enzymatic Treatment

Polysaccharides are among the most intensively studied components of the microbial biofilm matrix [10, 11]. Notably, they often represent an indispensable component of this matrix and thus are common targets in strategies aimed to disrupt microbial biofilms. An important example is the exopolysaccharide alginate of *P. aeruginosa*. *P. aeruginosa* strains that are isolated from

the lungs of CF patients frequently have a mucoid phenotype. The formation of mucoid has been attributed to the production of alginate, which appears to have a key role in pathogenesis, as it provides a protective barrier against the host's immune defense and antibiotic treatment [118]. The enzyme alginate lyase (AlgL), encoded as part of a 12-gene operon for alginate biosynthesis in *P. aeruginosa*, is likely involved in the degradation of periplasmic alginate polymers and their secretion to the outer membrane [119]. Mutation in *algL* led to rapid cell burst due to the accumulation of the alginate polymers in the periplasm. Therefore, alginate lyase has been suggested as a potential therapeutic for mucoid *P. aeruginosa* infection. In such an effort, Alkawash et al. showed recently that following treatment with exogenous alginate lyase, alginate was rapidly degraded in biofilms of mucoid *P. aeruginosa* [120]. The coadministration of alginate lyase with antibiotics also greatly enhanced the efficacy of the antimicrobial agents in the clearance of mucoid *P. aeruginosa* biofilms, suggesting that the incorporation of alginate lyase into an antibiotic regime may provide a better prognosis in the treatment of CF.

Staphylococci are among the most common group of microbes implicated in device-related infections [116, 121]. Although recent reports have suggested that in some cases, staphylococcal biofilms may be formed without the involvement of the exopolysaccharide PIA [122], the importance of PIA in staphylococcal biofilms and infections has been well established [123–125], making it an ideal target for antibiofilm therapy. One such approach is the use of the enzyme dispersin B, an *N*-acetyglucosaminidase produced by the gram-negative bacterium *Actinobacillus actinomycetemcomitans* [126]. Dispersin B specifically degrades PIA into its *N*-acetylglucosamine monomers [17, 18]. As a consequence, the addition of dispersin B resulted in the rapid disruption of intracellular adhesion within *S. epidermidis* biofilms, while pretreatment of surfaces with the enzyme prevented biofilm formation [126]. Interestingly, homologs of genes responsible for the biosyntesis of PIA have been found in other bacteria, such as *E. coli* [127] and *Yersinia* species [128], indicating that dispersin B may exhibit activity against other important biofilm forming organisms. As another example, therapeutic use of the cell wall lytic enzyme lysostaphin has been proposed for the control of *S. aureus* biofilms and infections. This enzyme specifically targets the pentaglycine cross bridges in the cell wall of *S. aureus* [129]. Although *S. epidermidis* has a certain degree of natural resistance to lysostaphin, which is due to the exchange of glycine residues with serine in the pentapeptide, biofilms of both *S. aureus* and *S. epidermidis* were disrupted by lysostaphin, and pretreatment of catheters with the enzyme prevented staphylococcal biofilm formation [130, 131]. Furthermore, the incorporation of lysostaphin in the treatment regime resulted in the eradication of diseases such as staphylococcal keratitis and endocarditis in several animal models [132–135]. However, although enzyme-based approaches may appear promising and elegant in vitro and in animal infection models, proteolytic instability and antigenicity may in general limit their applicability for the treatment of human infection. This may be the case specifically for lysostaphin, whose immunogenicity seems to limit the potential use of this enzyme for the treatment of staphylococcal infections [136].

### 7.3 Bacteriophage

The coexistence of bacteriophage and bacteria in biofilms has been established in a number of studies [137]. In addition, bacteriophages may cause the lysis of cells in biofilms of gram-positive and gram-negative bacteria [75, 138, 139], paving another alternative in biofilm control. Two recent reports have demonstrated the feasibility of using bacteriophage in combating biofilm infections. The pretreatment of catheters with an *S. epidermidis* bacteriophage resulted in a significant reduction in the formation of biofilms [140]. These results were encouraging and highlighted the prospect of bacteriophage coatings on biomaterial as a means to reduce device-related infections, although these initial findings need to be confirmed by in vivo studies. In the second study, a bacteriophage was used as a delivery tool for administering chloramphenicol to *S. aureus* [141]. Essentially, a high concentration of a chloramphenicol prodrug was conjugated to a bacteriophage, which remains inactive when bound to the phage. Upon interacting with the target *S. aureus* cells, a high dosage of chloramphenicol was delivered via enzymatic cleavage of the prodrug, resulting in severe retardation of *S. aureus* growth. As we have seen in previous chapters, the failure to eradicate microbial biofilms may in part be attributed to an insufficient concentration of antimicrobial agents that reaches the target [142, 143]. In that respect, such a use of biofilm-penetrating bacteriophage may be helpful in increasing the effective concentration of antibiotic and thus eradicating microbial biofilms.

## 8 CONCLUDING REMARKS

Antimicrobial resistance is one of the major reasons for which life in a biofilm is advantageous for microbes over the free-floating mode of growth. Notably, it allows pathogens to persist during human infection for prolonged periods and cause chronic infections that are often extremely difficult to eradicate.

We have outlined here how the formation of biofilms causes an increased resistance to antimicrobials, including microbicidal molecules of innate host defense and agents used during antibiotic therapy. We have seen that this increased resistance is due to a number of different mechanisms, of which presumably the most crucial and widely spread are the reduction of active metabolism and the production of a protective extracellular barrier. In addition, some microbes have specific mechanisms of resistance, which include for example the enzymatic degradation of antimicrobial substances.

The more recent advances in genomic and proteomic analyses of microbial gene and protein expression, and their application to the investigation of biofilm physiology, have already yielded a vast amount of data, and it is to be expected that similar studies will be performed to include more microorganisms and modes of biofilm growth. While the synoptic interpretation of these results still suffers somewhat from the lack of standardization with regard to biofilm models, we will certainly gain a much better understanding from these studies of how details of biofilm physiology affect antimicrobial resistance.

Current efforts to find treatments for biofilm-associated infections frequently aim to develop compounds for the inhibition of biofilm formation, which would be applied in combination with conventional antibiotics to kill biofilm-forming organisms. Directly interfering with the mechanisms of biofilm antibiotic resistance would represent a variation of this strategy. However, as we have seen, the mechanisms of biofilm resistance to antimicrobials are diverse, and, thus, biofilm resistance is probably best attacked by eliminating biofilm formation as a whole. Additionally, structures that are indispensable for the formation of a biofilm, such as the polysaccharide intercellular adhesin, are often also crucial for antimicrobial resistance. Therefore, understanding the molecular basis of biofilm resistance may assist us in providing potential lead substances in the endeavor to find new treatments for biofilm-associated infection.

## ACKNOWLEDGMENT

This work was supported by the Intramural Research Program of the National Institute of Allergy and Infectious Diseases (NIAID), NIH.

## REFERENCES

1. Hall-Stoodley L, Costerton JW, Stoodley P (2004). Bacterial biofilms: From the natural environment to infectious diseases. *Nat Rev Microbiol 2*(2), 95–108.
2. Thomas JG, Ramage G, Lopez-Ribot JL (2004). Biofilms and implant infections. In Ghannoum M, O'Toole GA, Eds. *Microbial Biofilms*. Washington, DC: ASM Press, pp. 269–293.
3. Costerton JW, Stewart PS, Greenberg EP (1999). Bacterial biofilms: A common cause of persistent infections. *Science 284*(5418), 1318–1322.
4. Hoyle BD, Costerton JW (1991). Bacterial resistance to antibiotics: The role of biofilms. *Prog Drug Res 37*, 91–105.
5. Fux CA, et al. (2005). Survival strategies of infectious biofilms. *Trends Microbiol 13*(1), 34–40.
6. Bower CK, Daeschel MA (1999). Resistance responses of microorganisms in food environments. *Int J Food Microbiol 50*(1–2), 33–44.
7. Berry D, Xi C, Raskin L (2006). Microbial ecology of drinking water distribution systems. *Curr Opin Biotechnol 17*(3), 297–302.
8. Costerton JW, Stewart PS (2001). Battling biofilms. *Sci Am 285*(1), 74–81.
9. Anderl JN, Franklin MJ, Stewart PS (2000). Role of antibiotic penetration limitation in *Klebsiella pneumoniae* biofilm resistance to ampicillin and ciprofloxacin. *Antimicrob Agents Chemother 44*(7), 1818–1824.
10. Sutherland IW (2001). The biofilm matrix–an immobilized but dynamic microbial environment. *Trends Microbiol 9*(5), 222–227.
11. Branda SS, et al. (2005). Biofilms: The matrix revisited. *Trends Microbiol 13*(1), 20–26.
12. De Beer D, Srinivasan R, Stewart PS (1994). Direct measurement of chlorine penetration into biofilms during disinfection. *Appl Environ Microbiol 60*(12), 4339–4344.
13. Huang CT, et al. (1995). Nonuniform spatial patterns of respiratory activity within biofilms during disinfection. *Appl Environ Microbiol 61*(6), 2252–2256.
14. Suci PA, et al. (1994). Investigation of ciprofloxacin penetration into *Pseudomonas aeruginosa* biofilms. *Antimicrob Agents Chemother 38*(9), 2125–2133.
15. Mah TF, et al. (2003). A genetic basis for *Pseudomonas aeruginosa* biofilm antibiotic resistance. *Nature 426*(6964), 306–310.
16. Nichols WW, et al. (1988). Inhibition of tobramycin diffusion by binding to alginate. *Antimicrob Agents Chemother 32*(4), 518–523.
17. Mack D, et al. (1996). The intercellular adhesin involved in biofilm accumulation of *Staphylococcus epidermidis* is a linear beta-1,6-linked glucosaminoglycan: Purification and structural analysis. *J Bacteriol 178*(1), 175–183.
18. Heilmann C, et al. (1996). Molecular basis of intercellular adhesion in the biofilm-forming *Staphylococcus epidermidis*. *Mol Microbiol 20*(5), 1083–1091.
19. Vuong C, et al. (2004). Polysaccharide intercellular adhesin (PIA) protects *Staphylococcus epidermidis* against major components of the human innate immune system. *Cell Microbiol 6*(3), 269–275.

20. Vuong C, et al. (2004). A crucial role for exopolysaccharide modification in bacterial biofilm formation, immune evasion, and virulence. *J Biol Chem 279*(52), 54881–54886.
21. Stewart PS, et al. (2001). Biofilm penetration and disinfection efficacy of alkaline hypochlorite and chlorosulfamates. *J Appl Microbiol 91*(3), 525–532.
22. Darouiche RO, et al. (1994). Vancomycin penetration into biofilm covering infected prostheses and effect on bacteria. *J Infect Dis 170*(3), 720–723.
23. Al-Fattani MA, Douglas LJ (2004). Penetration of *Candida biofilms* by antifungal agents. *Antimicrob Agents Chemother 48*(9), 3291–3297.
24. Wentland EJ, et al. (1996). Spatial variations in growth rate within *Klebsiella pneumoniae* colonies and biofilm. *Biotechnol Prog 12*(3), 316–321.
25. Sternberg C, et al. (1999). Distribution of bacterial growth activity in flow-chamber biofilms. *Appl Environ Microbiol 65*(9), 4108–4117.
26. De Beer D, et al. (1994). Effects of biofilm structures on oxygen distribution and mass transport. *Biotechnol Bioeng 43*, 1131–1138.
27. Xu KD, et al. (1998). Spatial physiological heterogeneity in *Pseudomonas aeruginosa* biofilm is determined by oxygen availability. *Appl Environ Microbiol 64*(10), 4035–4039.
28. Walters MC 3rd, et al. (2003). Contributions of antibiotic penetration, oxygen limitation, and low metabolic activity to tolerance of *Pseudomonas aeruginosa* biofilms to ciprofloxacin and tobramycin. *Antimicrob Agents Chemother 47*(1), 317–323.
29. Borriello G, et al. (2004). Oxygen limitation contributes to antibiotic tolerance of *Pseudomonas aeruginosa* in biofilms. *Antimicrob Agents Chemother 48*(7), 2659–2664.
30. Chandra J, et al. (2001). Biofilm formation by the fungal pathogen *Candida albicans*: Development, architecture, and drug resistance. *J Bacteriol 183*(18), 5385–5394.
31. Beenken KE, et al. (2004). Global gene expression in *Staphylococcus aureus* biofilms. *J Bacteriol 186*(14), 4665–4684.
32. Whiteley M, et al. (2001). Gene expression in *Pseudomonas aeruginosa* biofilms. *Nature 413*(6858), 860–864.
33. Hentzer M, Givskov M, Eberl L (2004). Quorum sensing in biofilms: Gossip in slime city. In Ghannoum M, O'Toole GA, Eds. *Microbial Biofilms*. Washington, DC: ASM Press, pp. 118–140.
34. Resch A, et al. (2005). Differential gene expression profiling of *Staphylococcus aureus* cultivated under biofilm and planktonic conditions. *Appl Environ Microbiol 71*(5), 2663–2676.
35. Resch A, et al. (2006). Comparative proteome analysis of *Staphylococcus aureus* biofilm and planktonic cells and correlation with transcriptome profiling. *Proteomics 6*(6), 1867–1877.
36. Worlitzsch D, et al. (2002). Effects of reduced mucus oxygen concentration in airway *Pseudomonas* infections of cystic fibrosis patients. *J Clin Invest 109*(3), 317–325.
37. Yoon SS, et al. (2002). *Pseudomonas aeruginosa* anaerobic respiration in biofilms: Relationships to cystic fibrosis pathogenesis. *Dev Cell 3*(4), 593–603.
38. Garcia-Sanchez S, et al. (2004). *Candida albicans* biofilms: A developmental state associated with specific and stable gene expression patterns. *Eukaryot Cell 3*(2), 536–545.
39. Murillo LA, et al. (2005). Genome-wide transcription profiling of the early phase of biofilm formation by *Candida albicans*. *Eukaryot Cell 4*(9), 1562–1573.
40. Thomas DP, et al. (2006). A proteomic-based approach for the identification of *Candida albicans* protein components present in a subunit vaccine that protects against disseminated candidiasis. *Proteomics 6*(22), 6033–6041.
41. Pocsi I, Prade RA, Penninckx MJ (2004). Glutathione, altruistic metabolite in fungi. *Adv Microb Physiol 49*, 1–76.
42. Muller S (2004). Redox and antioxidant systems of the malaria parasite *Plasmodium falciparum*. *Mol Microbiol 53*(5), 1291–1305.
43. Li XZ, Nikaido H (2004). Efflux-mediated drug resistance in bacteria. *Drugs 64*(2), 159–204.
44. Stover CK, et al. (2000). Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic pathogen. *Nature 406*(6799), 959–964.
45. De Kievit TR, et al. (2001). Multidrug efflux pumps: Expression patterns and contribution to antibiotic resistance in *Pseudomonas aeruginosa* biofilms. *Antimicrob Agents Chemother 45*(6), 1761–1770.
46. Ramage G, et al. (2002). Investigation of multidrug efflux pumps in relation to fluconazole resistance in *Candida albicans* biofilms. *J Antimicrob Chemother 49*(6), 973–980.
47. Mukherjee PK, et al. (2003). Mechanism of fluconazole resistance in *Candida albicans* biofilms: Phase-specific role of efflux pumps and membrane sterols. *Infect Immun 71*(8), 4333–4340.
48. Frees D, et al. (2004). Clp ATPases are required for stress tolerance, intracellular replication and biofilm formation in *Staphylococcus aureus*. *Mol Microbiol 54*(5), 1445–1462.
49. Schirmer EC, et al. (1996). HSP100/Clp proteins: A common mechanism explains diverse functions. *Trends Biochem Sci 21*(8), 289–296.
50. Yao Y, Sturdevant DE, Otto M (2005). Genomewide analysis of gene expression in *Staphylococcus epidermidis* biofilms: Insights into the pathophysiology of *S. epidermidis* biofilms and the role of phenol-soluble modulins in formation of biofilms. *J Infect Dis 191*(2), 289–298.
51. Vuong C, et al. (2004). Regulated expression of pathogen-associated molecular pattern molecules in *Staphylococcus epidermidis*: Quorum-sensing determines

pro-inflammatory capacity and production of phenol-soluble modulins. *Cell Microbiol 6*(8), 753–759.

52. Sakoulas G, et al. (2003). *Staphylococcus aureus* accessory gene regulator (agr) group II: Is there a relationship to the development of intermediate-level glycopeptide resistance? *J Infect Dis 187*(6), 929–938.
53. Sakoulas G, et al. (2005). Reduced susceptibility of *Staphylococcus aureus* to vancomycin and platelet microbicidal protein correlates with defective autolysis and loss of accessory gene regulator (agr) function. *Antimicrob Agents Chemother 49*(7), 2687–2692.
54. Venturi V (2003). Control of rpoS transcription in *Escherichia coli* and *Pseudomonas*: Why so different? *Mol Microbiol 49*(1), 1–9.
55. Schuster M, et al. (2004). The *Pseudomonas aeruginosa* RpoS regulon and its relationship to quorum sensing. *Mol Microbiol 51*(4), 973–985.
56. Cochran WL, et al. (2000). Role of RpoS and AlgT in *Pseudomonas aeruginosa* biofilm resistance to hydrogen peroxide and monochloramine. *J Appl Microbiol 88*(3), 546–553.
57. Romling U, Amikam D (2006). Cyclic di-GMP as a second messenger. *Curr Opin Microbiol 9*(2), 218–228.
58. Hoffman LR, et al. (2005). Aminoglycoside antibiotics induce bacterial biofilm formation. *Nature 436*(7054), 1171–1175.
59. Kocianova S, et al. (2005). Key role of poly-gamma-DL-glutamic acid in immune evasion and virulence of Staphylococcus epidermidis. *J Clin Invest 115*(3), 688–694.
60. Ghannoum MA, Rice LB (1999). Antifungal agents: Mode of action, mechanisms of resistance, and correlation of these mechanisms with bacterial resistance. *Clin Microbiol Rev 12*(4), 501–517.
61. Khot PD, et al. (2006). A small subpopulation of blastospores in *Candida albicans* biofilms exhibit resistance to amphotericin B associated with differential regulation of ergosterol and $\beta$-1,6-glucan pathway genes. *Antimicrob Agents Chemother 50*(11), 3708–3716.
62. Rogers PD, Barker KS (2003). Genome-wide expression profile analysis reveals coordinately regulated genes associated with stepwise acquisition of azole resistance in *Candida albicans* clinical isolates. *Antimicrob Agents Chemother 47*(4), 1220–1227.
63. Lewis K (2005). Persister cells and the riddle of biofilm survival. *Biochem (Mosc) 70*(2), 267–274.
64. Spoering AL, Lewis K (2001). Biofilms and planktonic cells of *Pseudomonas aeruginosa* have similar resistance to killing by antimicrobials. *J Bacteriol 183*(23), 6746–6751.
65. Keren I, et al. (2004). Persister cells and tolerance to antimicrobials. *FEMS Microbiol Lett 230*(1), 13–18.
66. Lafleur MD, Kumamoto CA, Lewis K (2006). *Candida albicans* biofilms produce antifungal-tolerant persister cells. *Antimicrob Agents Chemother 50*(11), 3839–3846.
67. Keren I, et al. (2004). Specialized persister cells and the mechanism of multidrug tolerance in *Escherichia coli*. *J Bacteriol 186*(24), 8172–8180.
68. Spoering AL, Vulic M, Lewis K (2006). GlpD and PlsB participate in persister cell formation in *Escherichia coli*. *J Bacteriol 188*(14), 5136–5144.
69. Haussler S (2004). Biofilm formation by the small colony variant phenotype of *Pseudomonas aeruginosa*. *Environ Microbiol 6*(6), 546–551.
70. Proctor RA, et al. (2006). Small colony variants: A pathogenic form of bacteria that facilitates persistent and recurrent infections. *Nat Rev Microbiol 4*(4), 295–305.
71. Haussler S, et al. (1999). Small-colony variants of *Pseudomonas aeruginosa* in cystic fibrosis. *Clin Infect Dis 29*(3), 621–625.
72. Haussler S, et al. (2003). Highly adherent small-colony variants of *Pseudomonas aeruginosa* in cystic fibrosis lung infection. *J Med Microbiol 52*(Pt 4), 295–301.
73. Singh PK, et al. (2000). Quorum-sensing signals indicate that cystic fibrosis lungs are infected with bacterial biofilms. *Nature 407*(6805), 762–764.
74. Drenkard E, Ausubel FM (2002). *Pseudomonas* biofilm formation and antibiotic resistance are linked to phenotypic variation. *Nature 416*(6882), 740–743.
75. Webb JS, Lau M, Kjelleberg S (2004). Bacteriophage and phenotypic variation in *Pseudomonas aeruginosa* biofilm development. *J Bacteriol 186*(23), 8066–8073.
76. Kuo MY, et al. (2000). High-frequency interconversion of turbid and clear plaque strains of bacteriophage f1 and associated host cell death. *Can J Microbiol 46*(9), 841–847.
77. Kahl BC, et al. (2003). Population dynamics of persistent *Staphylococcus aureus* isolated from the airways of cystic fibrosis patients during a 6-year prospective study. *J Clin Microbiol 41*(9), 4424–4427.
78. Schaaff F, et al. (2003). Mutations are involved in emergence of aminoglycoside-induced small colony variants of *Staphylococcus aureus*. *Int J Med Microbiol 293*(6), 427–435.
79. Proctor RA (1998). Bacterial energetics and antimicrobial resistance. *Drug Resist Updat 1*, 227–235.
80. von Eiff C, et al. (1997). A site-directed *Staphylococcus aureus* hemB mutant is a small-colony variant which persists intracellularly. *J Bacteriol 179*, 4706–4712.
81. Clements MO, et al. (1999). CtaA of *Staphylococcus aureus* is required for starvation survival, recovery, and cytochrome biosynthesis. *J Bacteriol 181*(2), 501–507.
82. Bates DM, et al. (2003). *Staphylococcus aureus* menD and hemB mutants are as infective as the parent strains, but the menadione biosynthetic mutant persists within the kidney. *J Infect Dis 187*(10), 1654–1661.
83. Gilligan PH, et al. (1987). Prevalence of thymidine-dependent *Staphylococcus aureus* in patients with cystic fibrosis. *J Clin Microbiol 25*(7), 1258–1261.
84. Kahl B, et al. (1998). Persistent infection with small colony variant strains of *Staphylococcus aureus* in

patients with cystic fibrosis. *J Infect Dis 177*(4), 1023–1029.

85. Pelletier LL Jr, Richardson M, Feist M (1979). Virulent gentamicin-induced small colony variants of *Staphylococcus aureus*. *J Lab Clin Med 94*(2), 324–334.
86. Balwit JM, et al. (1994). Gentamicin-resistant menadione and hemin auxotrophic *Staphylococcus aureus* persist within cultured endothelial cells. *J Infect Dis 170*(4), 1033–1037.
87. Vaudaux P, et al. (2002). Increased expression of clumping factor and fibronectin-binding proteins by hemB mutants of *Staphylococcus aureus* expressing small colony variant phenotypes. *Infect Immun 70*(10), 5428–5437.
88. Kohler C, et al. (2003). Physiological characterization of a heme-deficient mutant of *Staphylococcus aureus* by a proteomic approach. *J Bacteriol 185*(23), 6928–6937.
89. von Eiff C, et al. (2006). Phenotype microarray profiling of *Staphylococcus aureus* menD and hemB mutants with the small-colony-variant phenotype. *J Bacteriol 188*(2), 687–693.
90. Kahl BC, et al. (2005). Thymidine-dependent *Staphylococcus aureus* small-colony variants are associated with extensive alterations in regulator and virulence gene expression profiles. *Infect Immun 73*(7), 4119–4126.
91. Fuqua WC, Winans SC, Greenberg EP (1994). Quorum sensing in bacteria: The LuxR-LuxI family of cell density-responsive transcriptional regulators. *J Bacteriol 176*(2), 269–275.
92. Hogan DA (2006). Talking to themselves: Autoregulation and quorum sensing in fungi. *Eukaryot Cell 5*(4), 613–619.
93. Kjelleberg S, Molin S (2002). Is there a role for quorum sensing signals in bacterial biofilms? *Curr Opin Microbiol 5*(3), 254–258.
94. Manefield M, e^oo^t al. (1999). Evidence that halogenated furanones from *Delisea pulchra* inhibit acylated homoserine lactone (AHL)-mediated gene expression by displacing the AHL signal from its receptor protein. *Microbiology 145*(Pt 2), 283–291.
95. Manefield M, et al. (2002). Halogenated furanones inhibit quorum sensing through accelerated LuxR turnover. *Microbiology 148*(Pt 4), 1119–1127.
96. Koch B, et al. (2005). The LuxR receptor: The sites of interaction with quorum-sensing signals and inhibitors. *Microbiology 151*(Pt 11), 3589–3602.
97. de Nys R, et al. (2006). Furanones. *Prog Mol Subcell Biol 42*, 55–86.
98. Hentzer M, et al. (2003). Attenuation of *Pseudomonas aeruginosa* virulence by quorum sensing inhibitors. *Embo J 22*(15), 3803–3815.
99. Ren D, et al. (2004). Differential gene expression shows natural brominated furanones interfere with the autoinducer-2 bacterial signaling system of *Escherichia coli*. *Biotechnol Bioeng 88*(5), 630–642.
100. Hentzer M, et al. (2002). Inhibition of quorum sensing in *Pseudomonas aeruginosa* biofilm bacteria by a halogenated furanone compound. *Microbiology 148*(Pt 1), 87–102.
101. Wu H, et al. (2004). Synthetic furanones inhibit quorum-sensing and enhance bacterial clearance in *Pseudomonas aeruginosa* lung infection in mice. *J Antimicrob Chemother 53*(6), 1054–1061.
102. Hume EB, et al. (2004). The control of *Staphylococcus epidermidis* biofilm formation and in vivo infection rates by covalently bound furanones. *Biomaterials 25*(20), 5023–5030.
103. Baveja JK, et al. (2004). Furanones as potential antibacterial coatings on biomaterials. *Biomaterials 25*(20), 5003–5012.
104. Givskov M, et al. (1996). Eukaryotic interference with homoserine lactone-mediated prokaryotic signalling. *J Bacteriol 178*(22), 6618–6622.
105. Hornby JM, et al. (2001). Quorum sensing in the dimorphic fungus *Candida albicans* is mediated by farnesol. *Appl Environ Microbiol 67*(7), 2982–2992.
106. Ramage G, et al. (2002). Inhibition of *Candida albicans* biofilm formation by farnesol, a quorum-sensing molecule. *Appl Environ Microbiol 68*(11), 5459–5463.
107. Cao YY, et al. (2005). cDNA microarray analysis of differential gene expression in *Candida albicans* biofilm exposed to farnesol. *Antimicrob Agents Chemother 49*(2), 584–589.
108. Semighini CP, et al. (2006). Farnesol-induced apoptosis in *Aspergillus nidulans* reveals a possible mechanism for antagonistic interactions between fungi. *Mol Microbiol 59*(3), 753–764.
109. Jabra-Rizk MA, et al. (2006). Effect of farnesol on *Staphylococcus aureus* biofilm formation and antimicrobial susceptibility. *Antimicrob Agents Chemother 50*(4), 1463–1469.
110. Telford G, et al. (1998). The *Pseudomonas aeruginosa* quorum-sensing signal molecule *N*-(3-oxododecanoyl)-L-homoserine lactone has immunomodulatory activity. *Infect Immun 66*(1), 36–42.
111. Ritchie AJ, et al. (2003). Modification of in vivo and in vitro T- and B-cell-mediated immune responses by the *Pseudomonas aeruginosa* quorum-sensing molecule *N*-(3-oxododecanoyl)-L-homoserine lactone. *Infect Immun 71*(8), 4421–4431.
112. Smith RS, et al. (2002). The *Pseudomonas aeruginosa* quorum-sensing molecule *N*-(3-oxododecanoyl)homoserine lactone contributes to virulence and induces inflammation in vivo. *J Bacteriol 184*(4), 1132–1139.
113. Smith RS, et al. (2002). The *Pseudomonas* autoinducer *N*-(3-oxododecanoyl) homoserine lactone induces cyclooxygenase-2 and prostaglandin E2 production in human lung fibroblasts: Implications for inflammation. *J Immunol 169*(5), 2636–2642.
114. Hooi DS, et al. (2004). Differential immune modulatory activity of *Pseudomonas aeruginosa* quorum-sensing signal molecules. *Infect Immun 72*(11), 6463–6470.

115. Vuong C, et al. (2003). Quorum-sensing control of biofilm factors in *Staphylococcus epidermidis*. *J Infect Dis 188*(5), 706–718.
116. Kong KF, Vuong C, Otto M (2006). *Staphylococcus* quorum sensing in biofilm formation and infection. *Int J Med Microbiol 296*(2–3), 133–139.
117. Vuong C, et al. (2004). Increased colonization of indwelling medical devices by quorum-sensing mutants of *Staphylococcus epidermidis* in vivo. *J Infect Dis 190*(8), 1498–1505.
118. Ramsey DM, Wozniak DJ (2005). Understanding the control of *Pseudomonas aeruginosa* alginate synthesis and the prospects for management of chronic infections in cystic fibrosis. *Mol Microbiol 56*(2), 309–322.
119. Jain S, Ohman DE (2005). Role of an alginate lyase for alginate transport in mucoid *Pseudomonas aeruginosa*. *Infect Immun 73*(10), 6429–6436.
120. Alkawash MA, Soothill JS, Schiller NL (2006). Alginate lyase enhances antibiotic killing of mucoid *Pseudomonas aeruginosa* in biofilms. *Apmis 114*(2), 131–138.
121. Vuong C, Otto M (2002). *Staphylococcus epidermidis* infections. *Microbes Infect 4*(4), 481–489.
122. Qin Z, et al. (2007). Formation and properties of in vitro biofilms of ica-negative *Staphylococcus epidermidis* clinical isolates. *J Med Microbiol 56*(Pt 1), 83–93.
123. Rupp ME, et al. (1999). Characterization of the importance of polysaccharide intercellular adhesin/hemagglutinin of *Staphylococcus epidermidis* in the pathogenesis of biomaterial-based infection in a mouse foreign body infection model. *Infect Immun 67*(5), 2627–2632.
124. Rupp ME, et al. (1999). Characterization of *Staphylococcus epidermidis* polysaccharide intercellular adhesin/hemagglutinin in the pathogenesis of intravascular catheter-associated infection in a rat model. *Infect Immun 67*(5), 2656–2659.
125. Rupp ME, et al. (2001). Characterization of the importance of *Staphylococcus epidermidis* autolysin and polysaccharide intercellular adhesin in the pathogenesis of intravascular catheter-associated infection in a rat model. *J Infect Dis 183*(7), 1038–1042.
126. Kaplan JB, et al. (2004). Enzymatic detachment of *Staphylococcus epidermidis* biofilms. *Antimicrob Agents Chemother 48*(7), 2633–2636.
127. Wang X, Preston JF 3rd, Romeo T (2004). The pgaABCD locus of *Escherichia coli* promotes the synthesis of a polysaccharide adhesin required for biofilm formation. *J Bacteriol 186*(9), 2724–2734.
128. Darby C, et al. (2002). *Caenorhabditis elegans*: Plague bacteria biofilm blocks food intake. *Nature 417*(6886), 243–244.
129. Browder HP, et al. (1965). Lysostaphin: Enzymatic mode of action. *Biochem Biophys Res Commun 19*, 383–389.
130. Wu JA, et al. (2003). Lysostaphin disrupts *Staphylococcus aureus* and *Staphylococcus epidermidis* biofilms on artificial surfaces. *Antimicrob Agents Chemother 47*(11), 3407–3414.
131. Shah A, Mond J, Walsh S (2004). Lysostaphin-coated catheters eradicate *Staphylococccus aureus* challenge and block surface colonization. *Antimicrob Agents Chemother 48*(7), 2704–2707.
132. Climo MW, et al. (1998). Lysostaphin treatment of experimental methicillin-resistant *Staphylococcus aureus* aortic valve endocarditis. *Antimicrob Agents Chemother 42*(6), 1355–1360.
133. Patron RL, et al. (1999). Lysostaphin treatment of experimental aortic valve endocarditis caused by a *Staphylococcus aureus* isolate with reduced susceptibility to vancomycin. *Antimicrob Agents Chemother 43*(7), 1754–1755.
134. Dajcs JJ, et al. (2000). Lysostaphin treatment of methicillin-resistant *Staphylococcus aureus* keratitis in the rabbit(1). *Am J Ophthalmol 130*(4), 544.
135. Dajcs JJ, et al. (2001). Lysostaphin is effective in treating methicillin-resistant *Staphylococcus aureus* endophthalmitis in the rabbit. *Curr Eye Res 22*(6), 451–457.
136. Harrison EF, Fuquay ME, Zygmunt WA (1975). Antigenic response to topically applied proteins. *Infect Immun 11*(2), 309–312.
137. Sutherland IW, et al. (2004). The interaction of phage and biofilms. *FEMS Microbiol Lett 232*(1), 1–6.
138. Webb JS, et al. (2003). Cell death in *Pseudomonas aeruginosa* biofilm development. *J Bacteriol 185*(15), 4585–4592.
139. Resch A, et al. (2005). Phage release from biofilm and planktonic *Staphylococcus aureus* cells. *FEMS Microbiol Lett 252*(1), 89–96.
140. Curtin JJ, Donlan RM (2006). Using bacteriophages to reduce formation of catheter-associated biofilms by *Staphylococcus epidermidis*. *Antimicrob Agents Chemother 50*(4), 1268–1275.
141. Yacoby I, et al. (2006). Targeting antibacterial agents by using drug-carrying filamentous bacteriophages. *Antimicrob Agents Chemother 50*(6), 2087–2097.
142. Gilbert P, et al. (2001). Assessment of resistance towards biocides following the attachment of micro-organisms to, and growth on, surfaces. *J Appl Microbiol 91*(2), 248–254.
143. Grobe KJ, Zahller J, Stewart PS (2002). Role of dose concentration in biocide efficacy against *Pseudomonas aeruginosa* biofilms. *J Ind Microbiol Biotechnol 29*(1), 10–15.

# DROTRECOGIN ALFA ACTIVATED (RECOMBINANT HUMAN ACTIVATED PROTEIN C, XIGRIS)

Byungse Suh and Rafik Samuel
*Section of Infectious Diseases, Temple University School of Medicine, Philadelphia, Pennsylvania*

Our understanding of the immunopathophysiology of sepsis has made extraordinary progress during the past several decades, although therapeutic advances during the similar period have been extremely slow. During the past decade several clinical trials demonstrated promising results showing reduced morbidity and mortality in sepsis, although some had mixed outcomes. These trials have supported four major approaches: (1) maintenance of good tissue oxygenation by aggressive management of fluid and blood transfusion, (2) intensive insulin therapy to maintain the euglycemia, (3) low-dose corticosteroids, and (4) treatment with human recombinant activated protein C (drotrecogin alfa activated, Xigris, Eli Lilly and Company, Indianapolis, IN). These options will be briefly summarized below, but this chapter will concentrate on the epidemiology, pathophysiology, and management of sepsis with emphasis on the use of drotrecogin alfa activated (DAA) in the management of sepsis.

## 1 INTRODUCTION

In 1991, the American College of Chest Physicians (ACCP) and the Society of Critical Care Medicine (SCCM) held a "Consensus Conference" to define the terms *sepsis* and *organ failure* in a precise manner. At the conference, the term *systemic inflammatory response syndrome* (SIRS) was developed representing a clinical response arising from a wide variety of nonspecific insults [1]. The syndrome complex includes, but is not limited to, more than one of the following: (a) a body temperature of >38°C or <36°C; (b) a heart rate of >90 beats/min; (c) tachypnea, as manifested by a respiratory rate of >20 breaths/min or hyperventilation, as indicated by a $PaCO_2$ of 32 Torr (<4.3 kPa); and (d) an alteration of the white blood cell (WBC) count of >12,000 cells/$mm^3$, or <4000 cells/$mm^3$, or the presence of >10% immature neutrophils ("bands"). *Sepsis* is defined as having SIRS with proven or highly suspected infection, while sepsis associated with organ dysfunction, hypoperfusion, or hypotension is defined as *severe sepsis*. Furthermore, *septic shock* is referred to sepsis with hypotension [a systolic blood pressure (BP) of <90 mmHg or reduction of >40 mmHg from baseline], despite adequate fluid resuscitation, along with the evidence of tissue hypoperfusion abnormalities that may include, but are not limited to, lactic acidosis, oliguria, or an acute alteration in mental status [1]. The introduction and use of these definitions were not without criticism for making the situation more confusing, and the term SIRS is nonspecific and too sensitive [2]. A decade later, in 2001, the 1992 original document [1] was revisited by the expert members of the SCCM, the ACCP, the European Society of Intensive Care Medicine (ESICM), the American Thoracic Society (ATS), and the Surgical Infection Society (SIS), and found no evidence to support a change to the definitions [3]. Therefore, in this chapter the 1992 consensus definitions will be used.

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

## 2 EPIDEMIOLOGY OF SEPSIS

Review of the epidemiology of sepsis in the United States during the past several decades strongly indicates that the incidence of sepsis and the absolute number of sepsis-related deaths are increasing [4, 5]. There were 164,000 and 660,000 cases of sepsis in 1979 and 2000, respectively, which translate into approximately 8.7% annual increase in the incidence of sepsis [5]. A best estimate for the national annual incidence of severe sepsis has been reported to be 751,000 with a mortality rate of 28.6% (215,000) and the average costs per case was $22,100 with annual total costs of $16.7 billion [6]. Mortality for the more severe septic shock is 58% [7]. It is estimated that incidence of sepsis will be more than 1 million cases per year in the U. S. by 2020 [6, 8].

## 3 PATHOPHYSIOLOGY OF SEPSIS

Since sepsis is a clinical syndrome representing a set of signs and symptoms occurring together or in sequence due to infection rather than a distinct disease entity, it is likely that there may be multiple pathophysiologic mechanisms. In fact, there have been multiple mechanisms proposed and some of them have been tested for the syndrome; some with success and some with failure. Therefore, the precise pathophysiologic mechanism underlying the syndrome is still poorly understood and most likely entails more than one mechanism. The interaction between invading microbial agent and their products such as endotoxin (lipopolysaccharide, LPS) and the host responses can be classified into two large categories: (1) local defense mechanism: walling off and killing invading microbes, and (2) systemic response mechanism: keeping infection and inflammation localized [9].

The local defense mechanism utilizes the involvement of tissue–resident macrophages, mast cells, and dendritic cells. These cells, by secreting mediators that mobilize the inflammatory response, can mount the defense mechanism at the local level, helping to wall off the tissue infection and minimize bloodstream invasion. In addition, neutrophils and Kupffer cells of the liver phagocytose and help eliminate the invading organisms, and this response is facilitated by host's both immune systems, that is, innate and acquired immunity [10, 11].

The current concept on the systemic response of host to infection, injury, and other stresses emphasizes the suppression of an inflammatory process within the bloodstream. The systemic responses are believed to be primarily regulated by the brain and the liver [9]. The central nervous system (CNS), upon receiving signals of microbial invasion through nociceptive and vagal nerves, rapidly activates the hypothalamus–pituitary–adrenal (HPA) axis. Thus, the autonomic nervous system and the hypothalamic thermoregulatory center are primed properly to minimize the inflammatory process and maximally curtail the microbial growth. At the same time, blood-borne mediators [interleukin-1β (IL-1β), interleukin 6 (IL-6), tumor necrosis factor (TNF), interferons, prostaglandin $E_2$] are transported to reach the hypothalamus to augment the above processes [9, 12].

The liver and the spleen play an essential role in systemic responses to infection. The liver filters the blood that may harbor microbes, leading to collection and killing the blood-borne microorganisms. At the same time, it senses low level of circulating mediators and communicates this information to the CNS. The spleen is the principal organ to filter and eliminate the opsonized microorganisms.

As the CNS receives information on sepsis, injury, and other stresses, it triggers the mechanisms of systemic responses in five major categories: anti-infective, anti-inflammatory, procoagulant, metabolic, and thermoregulatory processes [9]. Each category of mechanisms in theory contributes to limit the infection or injury.

Leukocytosis, through demargination by epinephrine and cortisol, and through mobilization of marrow reserve neutrophils by granulocyte colony stimulating factor (G-CSF) and other cytokines, is an example of the initial anti-infective mechanism. Furthermore, anti-inflammatory responses keep the demarginated neutrophils from adhering to uninfected tissues, and increase in blood levels of anti-inflammatory mediators such as IL-1Ra, soluble TNF receptors, epinephrine, cortisol, α-melanocyte-stimulating hormone (α-MSH), adrenocorticotropic hormone (ACTH), IL-4, IL-6, IL-10, IL-13, transforming growth hormone (TGF)-ß, C-reactive protein (CRP), protease inhibitors, and antioxidants [9]. Recent evidence suggests that the pro-inflammatory response as determined by the level of IL-1ß and/or interferon-γ is localized to the infection or inflammatory site, whereas anti-inflammatory response—as measured by the levels of Anti-IL-1 and Anti-TNF—is systemic [13, 14].

Circulating cytokines trigger the brain's normal adaptive thermoregulatory response to systemic inflammation with fever [15–17]. This is manifested by shivering and redirecting blood flow from the skin and extremities to internal organs by means of vasoconstriction [9]. This increase in body temperature may favor the survival of the host while inhibiting bacterial growth. Bactericidal activity of neutrophils and macrophages may be increased during the febrile period [18].

## 4 MANAGEMENT OF SEVERE SEPSIS

Septic shock can be managed aggressively and effectively by identifying the possible etiologic organism(s) and the most likely source of infection. Antimicrobial therapy remains the mainstay of therapy to improve the outcome of severe sepsis. Therefore, efforts to correctly identify the source and possible etiologic agent(s) should be sought early in the process. Any necessary surgical drainage should be performed as soon as possible, and adequate cultures must be obtained. Then the attention should be concentrated on stabilizing the cardiovascular, respiratory, and other organ systems in an intensive care setting, often with aggressive fluid resuscitation.

Once these items, such as antimicrobials, fluid resuscitation to maintain the blood pressure, vasopressor use, and adequate oxygenation have been put in place, adjunctive therapeutic measures to reduce the morbidity and mortality based on pathophysiology of sepsis may be contemplated. Therapeutic modalities that have been tried to improve the outcomes of patients with septic shock will be described briefly; some with positive outcome and some with disappointing outcomes.

## 5 INTENSIVE INSULIN THERAPY

It is well recognized that among various metabolic derangements in the early phase of shock, the hyperglycemia, as a result of glycogenolysis and gluconeogenesis mediated by glucocorticoids, glucagon, and catecholamines as well as the decrease in insulin, is commonly observed. Also, hy*po*glycemia may ensue in the late phase of shock due to glycogen depletion or failure of glucose synthesis in the liver [19]. In a recent clinical trial in critically ill surgical patients, it was demonstrated that intensive insulin therapy to maintain blood glucose at or below 110 mg per deciliter improves the outcome with reduced morbidity and mortality [20]. However, in a more recent study, well designed with larger patient population that investigated the role of intensive insulin therapy in patients with severe sepsis, failed to demonstrate improved outcomes. Rather, the intensive insulin therapy group had more serious adverse effects related to hypoglycemia with no improvement in mortality [21]. Therefore intensive insulin therapy in sepsis will need to be reassessed.

## 6 USE OF STEROIDS

It is well understood that the host responds to stress/injury including that due to sepsis with an activation of the HPA axis that results in increased hydrocortisol levels. A possible poor prognostic indicator for patients in septic shock may be an inadequate rise in hydrocortisol after a corticotropin stimulation test [22]. Based on this hypothesis, a double-blind, placebo-controlled, randomized, and multicenter clinical study was performed to see whether low doses of corticosteroids (50 mg intravenously/6 h for 7 days and 50 μg of 9-α-fludrocortisone) in patients with septic shock would improve the 28-day survival [23]. When a total of 229 nonresponders were analyzed (placebo, 115; corticosteroid, 114), a lower mortality (53%) was seen in the corticosteroid group compared to the mortality (63%) in the placebo group. On the other hand, a similar analysis of 70 responders (placebo, 34; corticosteroid, 36) demonstrated no such effect; the 28-day mortality for the placebo and corticosteroid group were 53 and 61%, respectively. A meta-analysis of the effect of steroids on survival in septic shock during the period of 1963–2002 indicated that survival depends on the dose of steroid; high dose (e. g., methylprednisolone, 30 mg/kg of body weight, in four intravenous doses over 24 h) is detrimental, whereas low dose (e.g., 200–300 mg of hydrocortisone in four intravenous doses over a 24-h period for 5 days) increases the survival rate and shock reversal [24]. To validate the beneficial effect of low-dose steroids in patients in septic shock, a larger scale study on the patient population consisting of 254 corticotropin responders (hydrocortisone, 125; placebo, 108) and 233 nonresponders (hydrocortisone, 118; placebo, 136) has been undertaken [25]. The results, however, failed to demonstrate a beneficial effect of low-dose steroids in either patient population. Perhaps the use of corticosteroid may only be justified in those patients in septic shock which is resistant to high-dose vasopressors [23].

## 7 ENDOTOXIN-NEUTRALIZING PREPARATIONS

It is well recognized that endotoxin (LPS), a glycolipid outer membrane component of gram-negative bacteria, produces the typical features of gram-negative sepsis syndrome [26, 27]. Neutralization of endotoxin, therefore, should minimize the septic process in theory. HA-1A, a human monoclonal IgM antibody that binds to the lipid A domain of endotoxin, has been shown to reduce mortality in patients with septic shock from gram-negative bacteria from 57 to 33% when compared with placebo [28]. Nevertheless, a well controlled canine model of gram-negative septic shock failed to support the above study in that the HA-1A-treated group had a higher mortality (HA-1A group, 85%; control group, 43%) [29]. Another controlled clinical trial with E5

murine monoclonal IgM antibody to endotoxin revealed that only the patients who were not in shock seemed to have benefited from antibody therapy with improved survival and reversal of organ failure. A similar benefit was not seen in patients who were in shock [30]. Furthermore, in 2000 a study of over 900 patients with confirmed gram-negative sepsis investigated the efficacy of an E5 murine monoclonal antibody preparation. The study was prematurely stopped after an interim analysis showed no significant efficacy [31]. Current endotoxin monoclonal antibody preparations, either human or murine, have all failed to show benefit.

## 8 ANTI-INFLAMMATORY/ IMMUNOMODULATING AGENTS

In addition to low-dose corticosteroid use as described above, high-dose corticosteroids as anti-inflammatory agents have been studied in septic shock. No study showed benefit and some showed harm [32–35]. A randomized, double-blind, and placebo controlled study to investigate the effects of ibuprofen (10 mg per kilogram body weight given intravenously for 8 doses), a potent cyclooxygenase inhibitor and anti-inflammatory agent, in severe sepsis patients failed to improve the overall patient population [36]. However, when the subpopulation of patients with hypothermia was analyzed, there was a significant improvement in survival (ibuprofen group, 54%; control group, 90%). Several studies have shown that patients with severe sepsis who are hypothermic (<35°C) have almost twice as high mortality as patients with normal or elevated temperature [37, 38].

Since the early 1990s, immunomodulatory agents such as interleukin 1 receptor antagonist [39] and tumor necrosis factor receptor fusion protein [40] treatment have been studied in patients with septic shock and failed to improve survival. Also, in a severe sepsis animal model, mortality in complete knockout IL-6 (IL-6KO) animals was no different from wild-type animals [41]. These study results strongly indicate that our understanding of severe sepsis is incomplete. Clinical trials of immunomodulatory therapy in severe sepsis have recently been reviewed with the majority showing disappointing outcomes [42]; large clinical studies often contradict the promising preclinical data [43].

## 9 ANTICOAGULANTS

It has long been recognized that inflammation and coagulation are closely linked, and the coagulation system is invariably activated in the presence of inflammation. If uncontrolled, this activation of the coagulation cascade may lead to serious consequences such as disseminated intravascular coagulation or organ failure due to acute vascular events and acute vascular events [44, 45]. Therefore, the understanding and appropriate manipulation of those factor(s) involved in the activation of the coagulation cascade may benefit patients with severe sepsis. To this end, two major coagulation inhibitory factors, anithrombin III and recombinant tissue factor pathway inhibitor (TFPI; tifacogin) were studied in patients with severe sepsis [46, 47]. Again, the results were disappointing [46, 47], contrary to a preliminary study in volunteers, receiving endotoxin followed by TFPI, which showed dose-dependent inhibition of coagulation without influencing the fibrinolytic and cytokine response [48].

## 10 ACTIVATED PROTEIN C

At last in 2001, a recombinant human activated protein C (drotrecogin alfa activated, Xigris) was shown to improve the outcomes of severe sepsis in the single randomized controlled Recombinant Human Activated Protein C Worldwide Evaluation in Severe Sepsis (PROWESS) trial [49]. This has led to the approval of DAA in patients with severe sepsis from the U. S. Food and Drug Administration, although by an evenly divided vote. Controversy has followed this drug since then [50–52].

Protein C is activated when thrombin binds the endothelial surface protein thrombomodulin. Activated protein C promotes fibrinolysis and inhibits the generation of thrombin by inactivating factor Va and factor VIIIa [53, 54], and it also has substantial anti-inflammatory properties [55, 56] (Fig. 1). DAA is a recombinant form of human activated protein C and is a serine protease with the same amino acid sequence as human-plasma-derived activated protein C. DAA is a glycoprotein with a molecular weight of approximately 55 kDa, and it has the same site of glycosylation as the human-plasma-derived activated protein C.

The PROWESS study was the pivotal, double-blind, placebo controlled multinational phase III trial of DAA for patients with severe sepsis [49]. In this study, a total of 1690 patients were randomized; 840 in the placebo group and 850 in the DAA group. The baseline patient characteristics including the demography, underlying prior risk factors, APACHE II (Acute Physiology and Chronic Health Evaluation II) scores, and severity scores of the patients in the placebo and treatment groups were well matched.

The trial was originally designed to enroll 2280 patients with two planned interim analyses after 760 and 1520 patients had been enrolled. If the statistical analyses revealed that DAA group was found to be significantly more efficacious than placebo, the guidelines

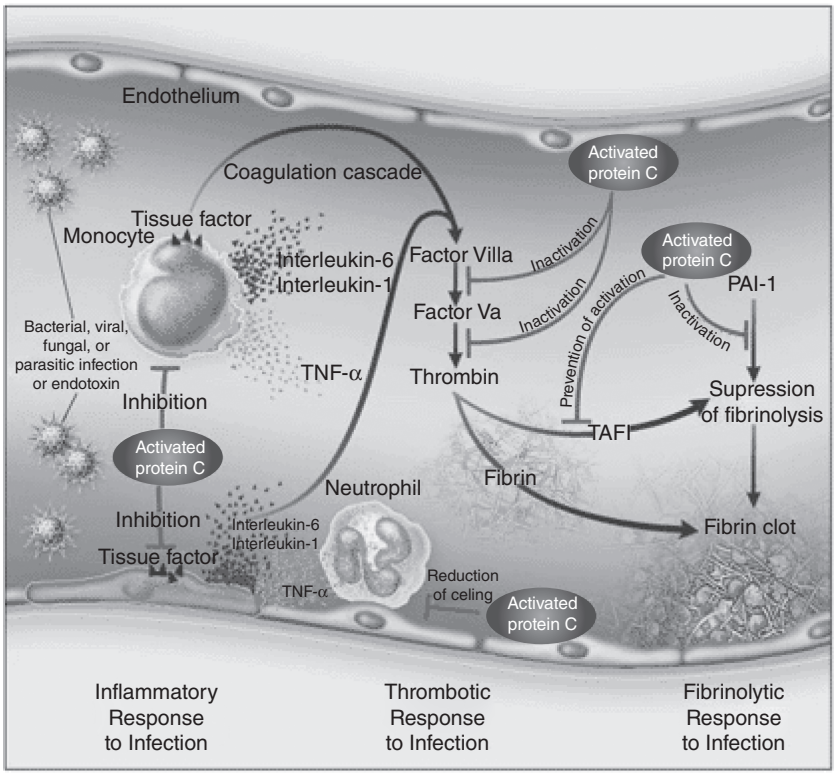


**Figure 1** Proposed actions of activated protein C in modulating the systemic inflammatory, procoagulant, and fibrinolytic host responses to infection. (Reprinted from [49] with permission.)

**TABLE 1 Summary of PROWESS Trial (Rates of Death from any Cause at 28 Days)**

| Variable | Placebo Group | Drotrecogin Alfa Activated Group | *p* Value | Absolute Reduction in Risk |
|---|---|---|---|---|
| *Treated patients* | | | | |
| Nonstratified analysis | 259/840 (30.8) | 210/850 (24.7) | 0.005 | 6.1 |
| Stratified analysis | | | 0.005 | 6.2 |
| Protein C deficiency | | | | |
| Yes | 215/670 (32.1) | 182/709 (25.7) | 0.009 | 6.4 |
| No | 28/105 (26.7) | 14/90 (15.6) | 0.06 | 11.1 |
| Unknown | 16/65 (24.6) | 14/51 (27.5) | 0.73 | –2.8 |
| *Randomized patients* | | | | |
| Nonstratified analysis | 268/857 (31.3) | 216/871 (24.8) | 0.003 | 6.5 |

were to suspend enrollment. At the time of the second interim analysis of 1520 patients, the difference in mortality rate between the two groups was significant to stop the trial.

In the trial, the treatment group received DAA at a dose of 24 μg per kilogram of body weight per hour, while the placebo group received 0.9% saline with or without 0.1% human serum albumin by continuous intravenous administration at a constant rate for 96 h. The results revealed that the mortality rate of 30.8% in the placebo group and 24.7% in the DAA group (Table 1), for an absolute reduction in the risk of death of 6.1%.

This outcome was accompanied by significantly increased incidence of serious bleeding in the DAA group (3.5%) compared to the placebo group (2.0%). In two patients in the DAA group, fatal intracranial hemorrhage occurred during the infusion period. Although one similar bleed occurred in one subject in the placebo group, it happened 6 days after the end of

infusion. These complications occurred despite very careful exclusion criteria were set forth in advance as described below.

Since DAA has anticoagulant properties, the patients with conditions that increase the risk of bleeding were excluded from the study. These exclusion criteria included were: age ≤18 years; platelet count of ≤30,000; surgery requiring general or spinal anesthesia within 12 h; a history of severe head trauma, intracranial surgery, or stroke within 3 months, intracerebral arteriovenous malformation, cerebral aneurysm, or presence of central nervous system mass lesions; a bleeding diatheses; a history of gastrointestinal bleeding within 6 weeks; a known hypercoagulable state; human immunodeficiency virus infection with cluster of differentiation 4 (CD4) counts of ≤50/mm$^3$; a history of solid organ or bone marrow transplantation; chronic renal failure requiring dialysis; a history of portosystemic hypertension with ascites; acute pancreatitis with no documented infection; and use of anticoagulant including heparin, aspirin, warfarin, and other anticoagulants.

In this study, the patients who received the DAA clearly had benefited with lower mortality due to all causes at 28 days compared to that of the placebo group. This benefit was observed regardless of whether or not the patients had a deficiency of protein C at the baseline, indicating that DAA possesses the additional pharmacologic effect(s) extending beyond the simple replacement of deficient activated protein C [49, 57]. Exact mechanisms underlying the role of DAA in reducing the mortality rate in patients with severe sepsis is unclear. The measurement of two parameters, d-dimer and inferteron-6 levels, during the first 7 days after the initiation of infusion begun, were monitored. The levels of both were significantly reduced in the treatment group, indicating that thrombus formation and the inflammatory process were reduced. Neutralizing antibodies against activated protein C were not detected in any patient, suggesting that the use of DAA may not be associated with antibody-mediated immune complex reactions.

One economic assessment on the entire above patient population that had participated in the trial concluded that the use of DAA in patient with severe sepsis with APACHE II score of 25 or more would be cost effective [58]. Another economic analysis on a similar patient population was not so favorable, although the economic attractiveness improved in patients with high APACHE II (≥25), it was not conclusive and further research needed [53]. It does appear that the use of DAA in patients with severe sepsis and high risk of death (APACHE II score ≥25) is relatively cost effective, but probably not in the lower risk patients (APACHE II score <25).

In the PROWESS study, patients over 75 years old (386 patients), were assessed separately for both short- and long-term survival outcomes [59]. The results revealed that the recipients of DAA had absolute risk reductions in 28-day and in-hospital mortality of 15.5 and 15.6%, respectively, compared with placebo recipients. Long-term follow-up data were available for 375 subjects, and survival rates as analyzed by the 24-month Kaplan–Meier survival curve for DAA recipients showed significantly higher 2-year survival [median 31 days for placebo group vs. 119 days for the DAA group ($p = 0.02$)]. This brings up the question whether or not we should pay special attention to this patient population in the era where the percentage of the population aged ≥65 years is continuously growing [60]. Pharmacokinetics and pharmacodynamics of many pharmacologic agents including DAA, although already being done to some extent with most new pharmacologic agents, should be investigated with special attention to older patients.

In 2005, the study results of adult patients with severe sepsis and a low risk of death, defined as an APACHE II score <25 or single organ failure, were published [61]. This was a randomized, double-blind, placebo-controlled, multicenter, and multinational (516 centers in 34 countries) trial. In this study, a total of 2640 patients were randomized and data from 2613 patients (1297 in placebo group and 1316 in DAA group) were collected. This study was prematurely terminated because of a low likelihood of meeting the statistically significant reduction in the 28-day mortality rate with the use of DAA. The results demonstrated that DAA offered no benefit in mortality rate in low-risk patients, and there was a greater incidence of bleeding (2.4% in the treatment group vs. 1.2% in the placebo group). Therefore, DAA should not be used in patients with severe sepsis with a low risk of death (APACHE II score <25).

In a recent study, only a pediatric population with ages ranging from 34 weeks' gestational age to 17 years with sepsis-induced cardiovascular and respiratory failure was investigated. They were randomized to receive either placebo or DAA (24 μg/kg/h for 96 h) [62]. A total of 477 patients (237 in placebo group and 240 in DAA group) were studied, and there was no significant difference in 28-day mortality rate (placebo 17.5% and DAA 17.2%) between the two groups. Although there was no significant difference in serious bleeding incidences, there were numerically more instances of central nervous system bleeding in DAA group (4.6%) than placebo group (2.1%), especially in children who were younger than 60 days.

An extended safety assessment of the use of DAA in adult patients with severe sepsis has been completed on all available safety data in 6777 patients—2786 patients

who enrolled in phase II and III trial studies, and 3991 patients who received the drug in commercial use [63]. This analysis validated the findings of the major PROWESS study outcome [49].

Based on the above discussion, review of the literature, and review of the package insert [64], the use of DAA in clinical medicine can be summarized as follows:

1. *Indications and Usage* Adult patients with severe sepsis with a high risk of death with an APACHE score >25. DAA is not indicated in patients with severe sepsis with a low risk of death (i.e., APACHE score <25). DAA is not indicated in pediatric patients.
2. *Contraindications* DAA is contraindicated in patients in whom bleeding could be associated with a high risk of mortality and/or morbidity. Conditions belonging to the category may include: active internal bleeding; recent (within 3 months) hemorrhagic stroke; recent (within 2 months) intracranial surgery, or severe head trauma; trauma with an increased risk of life-threatening bleeding; presence of an epidural catheter; intracranial neoplasm or mass lesion or evidence of cerebral herniation; and patients with hypersensitivity to DAA or any component of this product.
3. *Dosage and Administration* DDA should be administered intravenously at an infusion rate of 24 μg/kg infusion rate for a total duration of infusion of 96 h. Dose adjustment based on clinical or laboratory parameters is not recommended. If the infusion is interrupted, DDA should be restarted at the 24 μg/kg infusion rate. Dose escalation or bolus doses of DDA are not recommended. After preparation, the solution should be used at controlled room temperature of 20–25°C (68–77°F) within 12 h.
4. *Warnings* Bleeding is the most common serious adverse effect associated with DAA treatment. Individuals with one or more of the following conditions should be carefully considered prior to the treatment with DAA: patients requiring therapy with heparin for active thrombotic or embolic event; platelet count $<30{,}000 \times 10^6$/L; prothrombin time, international normalized ratio (INR) > 3.0; recent (within 6 weeks) gastrointestinal bleeding; recent (within 3 days) thrombolytic therapy; recent (within 7 days) oral anticoagulants or glycoprotein IIb/IIIa inhibitors; recent (within 7 days) administration of aspirin >650 mg per day; recent (within 3 months) ischemic stroke; intracranial arteriovenous malformation or aneurysm; known bleeding diathesis; chronic hepatic disease; and any other condition in which bleeding constitutes a significant hazard. Should clinically important bleeding occur, stop DAA infusion immediately. Once adequate hemostasis has been achieved, continued use of DAA may be reconsidered. If an invasive procedure that may cause major bleeding planned, DAA infusion should be discontinued 2 h prior to the invasive surgical procedure, and continued use of DAA may be reconsidered 12 h after the procedure.
5. *Precautions* Invasive procedures, including arterial and central venous punctures, should be minimized.
6. *Laboratory Tests* DAA may variably prolong the activated partial thromboplastin time (APTT); therefore, the APTT cannot be reliably used to assess the status of the coagulation during the DAA infusion. However, DAA has minimal effect on the thromboplastin time (PT), and the PT can be monitored to follow the coagulopathy.
7. *Immunogenicity* The incidence of antibody development in patients receiving DAA has not been adequately determined, as the assay sensitivity is inadequate to reliably detect all potential antibody responses. One patient in phase II trial developed antibodies to DAA without clinical sequelae.
8. *Carcinogenesis, Mutagenesis, Impairment of Fertility* Long-term studies in animals to evaluate potential carcinogenicity of DAA has not been performed. DAA is not mutagenic in vitro micronucleus study in mice or in an in vitro chromosomal aberration study in human peripheral blood lymphocytes with or without rat liver metabolic activation. The potential DAA to impair fertility has not been evaluated in male or female animals.
9. *Pregnancy Category C* It is not known whether DAA can cause fetal harm when administered to a pregnant woman or can affect reproduction capacity. DAA should be given to pregnant women only if clearly needed.
10. *Nursing Mothers* It is not known whether DAA is excreted in human milk or absorbed systemically after ingestion. A decision should be made whether to discontinue nursing or discontinue the drug, taking into account the importance of the drug to the mother.

11. *Pediatric Use* A placebo-controlled trial in pediatric patients did not establish the safety and effectiveness of DAA in the pediatric population.
12. *Geriatric Use* No overall differences in safety or effectiveness were observed between these patients and younger patients.
13. *Adverse Reactions* As mentioned above, bleeding is the most common adverse reaction with DAA. In PROWESS study, serious bleeding events were observed during the 28-day study period in 3.5% of DAA-treated patients and 2.0% of placebo-treated patients, respectively. Serious bleeding events included any intracranial hemorrhage, any life-threatening or fatal bleed, and any bleeding event requiring the administration of ≥3 units of packed red blood cells per day for 2 consecutive days, or any bleeding event assessed as a serious adverse event. As described above in the PROWESS study, two cases of intracranial hemorrhage occurred during the infusion period for DAA-treated patients and no cases were reported in the placebo patients. The incidence of intracranial hemorrhage during the 28-day study period was 0.2% for DAA-treated patients and 0.1% for placebo-treated patients. In that study, 25% of the DAA-treated patients and 18% of the placebo-treated patients experienced at least one bleeding event during the 28-day study period. In both treatment groups, the majority of bleeding events were ecchymoses or gastrointestinal tract bleeding. Additional studies with larger patient population were consistent with the PROWESS study findings. Low-dose heparin did not increase the risk of serious bleeding, although the risk of nonserious bleeding was increased compared with placebo over the treatment period of 0–6 days.

## 11 CONCLUSION

Sepsis carries with it a significantly high mortality rate despite tremendous improvements in the understanding of its pathogenesis. Many products designed to interfere with and prevent the sepsis cascade have unfortunately not been found to be helpful, and in some cases have been found to be detrimental. One agent, DAA, has been shown to be beneficial; however, the benefits of this agent is limited to only to a small number of individuals as outlined above. Further research and development of products for the treatment of sepsis must continue if we can try to improve the outcome of this disease state.

## REFERENCES


1. Members of the American College of Chest Physicians/Society of Critical Care Medicine Consensus Conference Committee (1992). American College of Chest Physicians/Society of Critical Care medicine Consensus Conference: Definitions for sepsis and organ failure and guidelines for the use of innovative therapies in sepsis. *Crit Care Med 20*, 864–874.
2. Vincent J (1997). Dear SIRS, I'm sorry to say that I don't like you. . . . *Crit Care Med 25*, 372–374.
3. Levy MM, Fink MP, Marshall JC, Abraham E, Angus D, Cook D, Cohen J, Opal SM, Vincent J-L, Ramsay G (2003). 2001 SCCM/ESICM/ACCP/ATS/SIS International Sepsis Definitions Conference. *Crit Care Med 31*, 1250–1256.
4. Angus DC, Wax RS (2001). Epidemiology of sepsis: An update. *Crit Care Med 29*(Suppl), S109–S116.
5. Martin GS, Mannino DM, Eaton S, Moss M (2003). The epidemiology of sepsis in the United States from 1979 through 2000. *N Engl J Med 348*, 1546–1554.
6. Angus DC, Linde-Zairble WT, Lidicker J, Clermont G, Carcillo J, Pinsky MR (2001). Epidemiology of severe sepsis in the United States: Analysis of incidence, outcome, and associated costs of care. *Crit Care Med 29*, 1303–1310.
7. Annane D, Aegerter P, Jars-Guinestre MC, Guider B (2003). Current epidemiology of septic shock for the CUB-Réa Network. *Am J Respir Crit Care Med 168*, 165–172.
8. O'Brien JM Jr, Ali NA, Aberegg SK, Abraham E (2007). Sepsis. *Am J Med 120*, 1012–1022.
9. Munford RS (2005). Sepsis, severe sepsis, and septic shock. In Mandell GL, Bennett JE, Dolin R, Eds. *Mandell, Douglas, and Bennett's Principles and Practice of Infectious Diseases*, 6th ed. Philadelphia, PA: Elsevier Churchill Livingstone, pp. 908–926.
10. Medzhitov R, Janeway CA Jr (2002). Decoding the patterns of self and nonself by the innate immune system. *Science 296*, 298–300.
11. Takeda K, Kaisho T, Akira S (2003). Toll-like receptors. *Annl Rev Immunol 21*, 335–376.
12. Netea MG, Kullberg BJ, Van der Meer JWM (2000). Circulating cytokines as mediators of fever. *Clin Infectt Dis 31*(Suppl), S178–S184.
13. Dugernier TL, Laterre P-F, Wittebole X, Roeseler J, Latinne D, Reynaert MS, Pugin J (2003). Compartmentalization of the inflammatory response during acute pancreatitis: Correlation with local and systemic complications. *Am J Respir Crit Care Med 168*, 148–157.
14. Rivera-Chavez FA, Wheeler H, Linberg G, Munford RS, O'Keefe GE (2003). Regional and systemic cytokine responses to acute inflammation of the vermiform appendix. *Ann Surg 237*, 408–416.
15. Dinarello CA, Cannon JG, Wolff SM (1988). New concepts on the pathogenesis of fever. *Rev Infect Dis 10*, 168–189.

16. Saper CB, Breder CD (1994). The neurologic basis of fever. *N Engl J Med 330*, 1880–1886.
17. Mackowiak PA, Bartlett JG, Borden EC, Godblum SE, Hasday JD, Munford RS, Nasraway SA, Stolley PD, Woodward TE (1997). Concepts of fever: Recent advances and lingering dogma. *Clin Infect Dis 25*, 119–138.
18. Rosenspire AJ, Kindzelskii AL, Petty HR (2002). Cutting edge: Fever-associated temperatures enhance neutrophil responses to lipopolysaccharide: A potential mechanism involving cell metabolism. *J Immunol 169*, 5396–5400.
19. Parrillo JE (2008). Approach to the patient with shock. In Goldman L, Ausiello D Eds. *Cecil Medicine*, 23rd ed. Philadelphia, PA: Saunders, pp. 742–750.
20. Van den Berghe G, Wouters P, Weekers F, Verwaest C, Bruyninckx F, Schetz M, Vlasselaers D, Ferdinande P, Lauwers P, Bouillon R (2001). Intensive insulin therapy in critically ill patients. *N Engl J Med 345*, 1359–1367.
21. Brunkhorst FM, Engel C, Bloos F, Meier-Hellman A, Ragaller M, Weiler N, Moerer O, Gruendling M, Oppert M, Grond S, Ofthoff D, Jaschinski U, John S, Rossaint R, Welte T, Schaefer M, Kern P, Kuhnt E, Kiehntopf M, Hartog C, Natanson C, Loeffler M, Reinhart K, for the German Competence Network Sepsis (SepNet) (2008). Intensive insulin therapy and pentastarch resuscitation in severe sepsis. *N Engl J Med 358*, 125–139.
22. Annane D, Sébile V, Troché G, Gajdos P, Bellissant E (2000). A 3-level prognostic classification in septic shock based on cortisol levels and cortisol response to corticotropin. *JAMA 283*, 1038–1045.
23. Annane D, Sébile V, Charpentier C, Bollaert P-E, François B, Korach J-M, Capellier G, Cohen Y, Azoulay E, Troché G, Chaumet-Riffaut P, Bellissant E (2002). Effect of treatment with low doses of hydrocortisone and fludrocortisones on mortality in patients with septic shock. *JAMA 288*, 862–871.
24. Minneci PC, Deans KJ, Banks SM, Eichacker PQ, Natanson C (2004). Meta-analysis: The effect of steroids on survival and shock during sepsis depends on the dose. *Ann Intern Med 141*, 47–56.
25. Sprung CL, Annane D, Keh D, Moreno R, Singer M, Freivogel K, Weiss YG, Benbenishty J, Kalenka A, Forst H, Laterre P-F, Reinhart K, Cuthbertson BH, Payen D, Briegel J, for the CORTICUS Study Group (2008). Hydrocortisone therapy for patients with septic shock. *N Engl J Med 358*, 111–124.
26. Lynn WA, Cohen J (1995). Adjunctive therapy for septic shock: A review of experimental approaches. *Clin Infect Dis 20*, 143–158.
27. Cohen J (2002). The immunopathogenesis of sepsis. *Nature 420*, 885–891.
28. Ziegler EJ, Fisher CJ Jr, Sprung EI, Straube RC, Sadoff JC, Foulke GE, Wortel CH, Fink MP, Dellinger P, Teng NNH, Allen IE, Berger HJ, Knatterud GL, LoBuglio AF, Smith CR, The HA-1A Sepsis Study Group (1991). Treatment of gram-negative bacteremia and septic shock with HA-1A human monoclonal antibody against endotoxin. *N Engl J Med 324*, 429–436.
29. Quezado ZMN, Natanson C, Aling DW, Banks SM, Koev CA, Elin RJ, Mossein JM, Bacher JD, Danner RL, Haffman WD (1993). A controlled trial of HA-1A in a canine model of gram-negative septic shock. *JAMA 269*, 2221–2227.
30. Greenman RL, Schein RM, Martin MA, Wenzel RP, Macintyre NR, Emmanuel G, Chmel H, Kohler RB, McCarthy M, Plouffe J, Russel JA, the XOMA Sepsis Study Group (1991). A controlled clinical trial of E5 murine monoclonal IgM antibody to endotoxin in the treatment of gram-negative sepsis. *JAMA 266*, 1097–1102.
31. Angus DC, Birmingham MC, Balk RA, Scannon PJ, Colins D, Kruse JA, Graham DR, Dedhia HV, Homann S, MacIntyre N, for the E5 Study Investigators (2000). E5 murine monoclonal anti-endotoxin antibody in gram-negative sepsis: A randomized controlled trial. *JAMA 283*, 1723–1730.
32. Sprung CL, Caralis PV, Marcial EH, Pierce M, Gelbard MA, Long WM, Duncan RC, Tendler MD, Karpf M (1984). The effects of high-dose corticosteroids in patients with septic shock: A prospective, controlled study. *N Engl J Med 311*, 1137–1143.
33. Bone RC, Fisher CJ Jr, Clemmer TP, Slotman GJ, Metz CA, Balk RA, The Methylprednisolone Severe Sepsis Study Group (1987). A controlled clinical trial of high-dose methylprednisolone in the treatment of severe sepsis and septic shock. *N Engl J Med 317*, 653–658.
34. The Veterans Administration Systemic Sepsis Cooperative Study Group (1987). Effect of high-dose glucocorticoid therapy on mortality in patients with clinical signs of systemic sepsis. *N Engl J Med 317*, 659–665.
35. Annane D, Bellissant E, Bollaert PE, Briegel J, Keh D, Kupfer Y (2004). Corticosteroids for severe sepsis and septic shock: A systematic review and meta-analysis. *Br Med J 329*, 480. BMJ, doi:10.1136/bmj.38181.4822222.55 (published 2 August 2004).
36. Pittet D, Thiévent B, Wenzel RP, Auckenthaler R, Suter PM (1996). Bedside prediction of mortality from bacteremic sepsis. A dynamic analysis of ICU patients. *Am J Respir Crit Care Med 153*, 684–693.
37. Arons MM, Wheeler AP, Bernard GR, Christman BW, Russell JA, Schein R, Summer WR, Steinberg KP, Fulkerson W, Wright P, Dupont WD, Swindell BB (1999). Effects of ibuprofen on the physiology and survival of hypothermic sepsis. Ibuprofen in Sepsis Study Group. *Crit Care Med 27*, 669–670.
38. Bernard GR, Wheeler AP, Russell JA, Schein R, Summer WR, Steinberg KP, Fulkerson WJ, Wright PE, Christman BW, Dupont WD, Higgins SB, Swindell BB, for the Ibuprofen in Sepsis Group (1997). The effects of ibuprofen on the physiology and survival of patients with sepsis. *N Engl J Med 336*, 912–918.
39. Fisher CJ, Dhainaut J-FA, Opal SM, Pribble JP, Balk RA, Slotman GJ, Iberti TJ, Rackow EC, Shapiro MJ, Greenman RL, Reines D, Shelly MP, Thompson BW, LaBrecque JF, Catalano MA, Knaus WA, Sadoff JC, for the Phase III rhl-Ira Sepsis Syndrome Study Group (1994). Recombinant

human interleukin 1 receptor antagonist in the treatment of patients with sepsis syndrome: Results from an randomized, double-blind, placebo-controlled trial. *JAMA 271*, 1836–1843.

40. Abraham E, Gauser MP, Butler T, Garbino J, Gelmont D, Laterre PF, Kudsk K, Bruining HA, Otto C, Tobin E, Zwingelstein C, Lesslauer W, Leighton A, for the Ro 45-2081 Study Group (1997). p55 tumor necrosis factor receptor fusion protein in the treatment of patients with severe sepsis and septic shock: A randomized controlled multicenter trial. *JAMA 277*, 1531–1538.
41. Remick DG, Bolgos G, Copeland S, Siddiqui J (2005). Role of interleukin-6 in mortality from and physiologic response to sepsis. *Infect Immun 73*, 2751–2757.
42. Vincent J-L, Sun Q, Dubois M-J (2002). Clinical trials of immunomodulatory therapies in severe sepsis and septic shock. *Clin Infect Dis 34*, 1084–1093.
43. Eichacker PQ, Parent C, Kalil A, Esposito C, Cui X, Banks SM, Gersptenberger EP, Fitz Y, Danner RL, Natanson C (2002). Risk and the efficacy of antiinflammatory agents: Retrospective and confirmatory studies of sepsis. *Am J Respir Crit Med 166*, 1197–1205.
44. Jurado RI (2001). Nonbleeding clotting: The role of the coagulation system in inflammation. *Infect Dis Clin Pract 10*, 415–421.
45. Dolinger RP (2003). Inflammation and coagulation: Implications for the septic patient. *Clin Infect Dis 36*, 1259–1265.
46. Warren BI, Eid A, Singer P, Pillay SS, Carl P, Novak I, Chapula P, Atherstone A, Pénzes I, Kübler A, Knaub S, Keineche H-O, Heinrichs H, Schindel F, Juers M, Bone RC, Opal SM, for the KyberSept Trial Study Group (2001). High-dose antithrombin III in severe sepsis: A randomized controlled trial. *JAMA 286*, 1869–1878.
47. Abraham E, Reinhart K, Opal S, Demeyer I, Doig C, Rodriguez AL, Beale R, Svoboda P, Laterre PF, Simon S, Light B, Spapen H, Stone J, Seibert A, Peckelsen C, De Deyne C, Postier R, Pettilä V, Sprung CI, Artigas A, Percell SR, Zwingelstein C, Tobias J, Poole L, Stolzenbach JC, Creasey AA, for the OPTIMIST Trial Study Group (2003). Efficacy and safety of tifacogin (recombinant tissue factor pathway inhibitor) in severe sepsis: A randomized controlled trial. *JAMA 290*, 238–247.
48. de Jonge E, Dekkers PEP, Creasey AA, Hack CE, Paulson SK, Karim A, Kesecioglu J, Levi M, van Denter SJH, van der Poll T(2000). Tissue factor pathway inhibitor dose-dependently inhibits coagulation activation without influencing the fibrinolytic and cytokine response during human endotoxemia. *Blood 95*, 1124–1129.
49. Bernard GR, Vincent J-L, Laterre P-F, LaRosa SP, Dhainaut J-F, Lopez-Rodriguez A, Steingrub JS, Garber GE, Helterbrand JD, Ely EW, Fisher CJ, for the Recombinant Human Activated Protein C Worldwide Evaluation in Severe Sepsis (PROWESS) Study Group (2001). Efficacy and safety of recombinant human activated protein C for severe sepsis. *N Engl J Med 344*, 699–709.
50. Walker FJ, Sexton PW, Esmon CT (1979). The inhibition of blood coagulation by activated protein C through the selective inactivation of activated factor V. *Biochem Biophys Acta 571*, 333–342.
51. Mackenzie AF (2005). Activated protein C: Do more survive? *Intens Care Med 31*, 1624–1626.
52. Smith B (2005). Drotrecogin alfa: Too early to give it early. *Emerg Med Austral 17*, 179–180.
53. Costa V, Brophy JM (2007). Drotrecogin alfa (activated) in severe sepsis: A systematic review and new cost-effective analysis. *BMC Anesthesiol 7*, 5–15.
54. Fulcher CA, Gardner JE, Griffin JH, Zimmerman TS (1984). Proteolytic inactivation of human factor VIII procoagulant protein by activated human protein C and its analog with factor V. *Blood 63*, 486–489.
55. Matthay MS (2001). Severe sepsis—A new treatment with both anticoagulant and anti-inflammatory properties. *N Engl J Med 344*, 759–762.
56. Hotchkiss RS, Karl IE (2003). The pathophysiology and treatment of sepsis. *N Engl J Med 348*, 138–150.
57. Matthay MA (2001). Severe sepsis—A new treatment with both anticoagulant and anti-inflammatory properties. *N Engl J Med 344*, 759–762.
58. Manns BJ, Lee H, Doig CJ, Johnson D, Donaldson C (2002). An economic evaluation of activated protein C treatment for severe sepsis. *N Engl J Med 347*, 993–1000.
59. Ely EW, Angus DC, Williams MD, Bates B, Qualy R, Bernard GR (2003). Drotrecogin alfa (activated) treatment of older patients with severe sepsis. *Clin Infect Dis 37*, 187–195.
60. High KP (2003). Why should the infectious diseases community focus on aging and care of older adult? *Clin Infect Dis 37*, 196–200.
61. Abraham E, Laterre P-F, Garg R, Levy H, Talwar D, Trzaskoma BL, François B, Guy JS, Brückmann M, Rea-Neto Á, Rossaint R, Perrotin D, Sablotzki A, Arkins N, Utterback BG, Macias WL, for the Administration of Drotrecogin Alfa (Activated) in Early Stage Severe Sepsis (ADDRESS) Study Group (2005). Drotrecogin alfa (activated) for adults with severe sepsis and a low risk of death. *N Engl J Med 353*, 1332–1341.
62. Nadel S, Goldstein B, Willaims MD, Dalton H, Peters M, Macias WL, Abd-Allah SA, Levy H, Angle R, Wand D, Sundin DP, Girior B, for the Researching Severe Sepsis and Organ Dysfunction in Children: A Global Perspective (RESOLVE) Study Group (2007). Drotrecogin alfa (activated) in children with severe sepsis: A multicenter phase III randomized controlled trial. *Lancet 369*, 836–843.
63. Bernard GR, Macias WL, Joyce DE, Williams MD, Bailey J, Vincent J-L (2003). Safety assessment of drotrecogin alfa (activated) in the treatment of adult patients with severe sepsis. *Crit Care 7*, 155–163.
64. Xigris (package insert) (2007). Indianapolis, IN: Eli Lilly and Company.

# ANTI-INFLAMMATORY THERAPEUTICS

Marion M. Chan[1] and Dunne Fong[2]

[1]Department of Microbiology and Immunology, Temple University School of Medicine, Philadelphia, Pennsylvania
[2]Department of Cell Biology and Neuroscience, Rutgers, The State University of New Jersey, Piscataway, New Jersey

## 1 INTRODUCTION

### 1.1 Definitions of Inflammation

Inflammation is the physiological response of the body to injury. In this sense, it serves as a defense mechanism. It is a process typical of vascular tissues, whereby fluid and leukocytes accumulate at the site of injury. The process normally leads to recovery from injury and to healing. However, if targeted destruction and assisted repair are not properly phased, inflammation can lead to persistent tissue damage [1].

*To inflame* means to "set fire." The word comes from *enflamme* in Middle English, *enflamm* in French, and *enflammare* in Latin [2]. In AD 30, Aulus Cornelius Celsus, in his work *De re medica*, recognized what is now known as the four cardinal signs of inflammation: "*rubor et tumor cum calore et dolor*" ("redness and swelling with heat and pain"). Later, a fifth sign, "*functio laesa*" ("disturbed function") was added by Rudolf Virchow, in his *Cellular Pathology* published in 1858. If an organ is inflamed, it does not function as it should. However, beneficial attributes of inflammation were also recognized by the Scottish surgeon-scientist John Hunter. In his 1793 work *Blood, Inflammation, and Gunshot Wounds*, he noted that inflammation is not a disease but rather a "salutary reaction" [3].

For inflammatory diseases, there are clinical needs for anti-inflammatories and analgesics. Simple definitions for these drugs are available. Analgesics are drugs that relieve pain. Anti-inflammatories are drugs used to reduce inflammation—the redness, heat, swelling, and increased blood flow found in infections and in many chronic noninfective diseases such as rheumatoid arthritis and gout.

### 1.2 Importance of Inflammation and Inflammatory Diseases

Inflammation is indispensable to the survival of organisms. Patients with genetic deficiencies in components of the inflammatory process, such as leukocyte adhesion deficiency, have an increased risk of life-threatening infections.

Inflammation serves to rid the body of the source of injury and repair the injured site. However, in situations where such resolution is not possible, for example, during persistent infections, inflammation lingers on and becomes chronic. Whereas acute inflammation only lasts for days, chronic inflammation can last for months and years. Chronic inflammation may be viewed as "frustrated repair." Complete removal of the irritant is essential for repair to progress to completion. Therefore, chronic inflammation is an immune response initiated by an irritant, the persistent antigen [2].

Chronic inflammation contributes to diseases. The connection is more clearly demonstrated in certain diseases. Chronic inflammation has been implicated in arthritis, atherosclerosis, and cancer. In the case for cancer, cellular products of chronic inflammation induce deoxyribonucleic acid (DNA) modifications

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

leading to somatic mutations that result in carcinogenesis. Therefore, there is demand for anti-inflammatories to treat chronic inflammation and its associated diseases.

### 1.3 History of Anti-inflammatory Therapeutics

The use of salicylate-containing plant extracts has a long history in human civilization. About 2400 years ago, Hippocrates recommended the use of the bark of the willow (*Salix alba*) for the treatment of pain in childbirth. Other plants with similar properties such as meadowsweet (*Spiraea ulmaria*) and wintergreen (*Gautheria procumbens*) have also been used as medicines since antiquity.

Historical development of salicylates is a classic story in the field of pharmacognosy—natural product-derived remedies giving rise to modern pharmacological agents. The development initiated the transformation of the German dye industry into the drug industry in the nineteenth century. A mass-scale chemical manufacture of salicylic acid was made by Frederick von Heyden and Herman Kolbe in 1874. The product was sold as treatment for rheumatism. This commercial endeavor was followed by the development of aspirin.

Because high dose of salicylic acid and sodium salicylate taken by patients led to side effects including nausea, vomiting, and other gastrointestinal symptoms, scientists at Friedrich Bayer and Company were told to synthesize a compound that would be competitively superior to the von Heyden salicylate acid. The young chemist Felix Hoffmann completed the synthesis of acetylsalicylic acid. Hoffman's father had rheumatoid arthritis and found taking several grams of salicylic acid daily unbearable. This synthesis was performed at the suggestion of Hoffmann's supervisor Arthur Eichengrun. The synthesis, previously performed by the French chemist Charles Frederich von Gerhardt in 1853, was the combination of salicylic acid with acetic anhydride to make acetylsalicylic acid.

The product acetylsalicylic acid was introduced as a drug by Bayer in 1899. The director of research at Bayer, Henrich Dreser, did not want the new drug to sound like the old salicylic acid, and named the drug aspirin. The trade name Aspirin came from: "a" for acetyl and "spirin" for natural salicylic acid from plants of the Spiraea family. A less believable version is that the aspirin name was from the early Neapolitan Bishop, St. Aspirinus, patron saint of headaches.

Today, aspirin rates second to alcohol as the most commonly consumed drugs in the world. It has been estimated that 50 billion tablets are taken worldwide annually. Despite intense competition, aspirin is still a well-known, cheap, and effective drug [4, 5]. However, other anti-inflammatory drugs are currently in use, although there is still a genuine need to develop novel agents.

### 1.4 Current Anti-Inflammatory Therapeutics and Their Problems

In 2005, Michael Whitehouse classified the development of anti-inflammatory agents into four types: (1) random discoveries, (2) programmed nonsteroid anti-inflammatory drugs (NSAIDs), (3) synthetic glucocorticosteroids, and (4) biologicals targeting pro-inflammatory cytokines [6]. Examples for random discoveries include the findings of antimicrobials that also relieve the patients' arthritis, such as chloroquine and hydroxychloroquine for malaria treatment, and tetracycline and minocycline for antibacterial treatment. The programmed NSAID development in many drug companies has led to the variety of anti-inflammatory agents currently in use, although some have been withdrawn. Glucocorticosteroid synthesis started with cortisone, followed by hundreds of cortisone analogs. Topical use on the skin and direct inhalation into the lung for asthma patients are examples for anti-inflammatory steroids. Biologicals are the latest products on the market. They are designed to neutralize cytokines or cytokine receptors, especially targeting tumor necrosis factor-α (TNFα) and interleukin-1β (IL-1β). Because biologicals are proteins requiring injections, monitoring, and expensive to manufacture, they cannot be considered as medicine for the masses.

The most common anti-inflammatory therapeutic agents are known as nonsteroid anti-inflammatory drugs. NSAIDs are anti-inflammatory, analgesic, and antipyretic. They act by inhibiting cyclooxygenases (COXs), also known as prostaglandin G/H synthases. The two enzymes are cyclooxygenases 1 and 2 (COX-1 and COX-2). Most NSAIDs act as reversible, competitive COX inhibitors and inhibit both enzymes with little selectivity. NSAIDs are useful in the treatment of rheumatoid arthritis and osteoarthritis. They provide symptomatic relief from pain and inflammation but do not affect disease progression. However, 20% of patients show adverse gastrointestinal side effects.

Among NSAIDs, aspirin is still the most widely consumed anti-inflammatory analgesic, and antipyretic agent and is the standard for comparison and evaluation of other agents. In use since 1893, acetaminophen (trade name Tylenol) is an effective alternative to aspirin as analgesic and antipyretic agent, but it has weak anti-inflammatory effect. Introduced in 1963, indomethacin is used for the treatment of rheumatoid arthritis. It is estimated to be 20 times more potent than aspirin. Indomethacin, in addition to COX inhibition, also inhib-

its leukocyte motility. The compound diclofenic is the most commonly used NSAID in Europe. It has more potent action against COX-2 than indomethacin. Diclofenic is used for long-term symptomatic treatment of rheumatoid arthritis and osteoarthritis. Compared to diclofenic, the most commonly used NSAID in the United States is ibuprofen (trade names Advil, Motrin, Nuprin). It is available without a prescription and used to treat arthritis and other symptoms [7].

The discovery of two COX enzymes and that COX-2 tends to be inducible led to a fashionable but misleading hypothesis that anti-inflammatory effects of NSAIDs might be primarily due to their effects on COX-2, but their principal side effects such as gastrotoxicity might be due to their effects on COX-1, with the corollary to find COX-2 selective agents [6]. These agents are known as COX-1 sparing/COX-2 selective NSAIDs, or "coxibs," and they have become blockbuster drugs.

The COX-2 inhibitor celecoxib (trade name Celebrex), launched in 1998 for indications of inflammatory pain and arthritis, is among the top 10 drugs in global sales, at $4 billion in 2002. On the other hand, another COX-2 inhibitor, rofecoxib (trade name Vioxx), launched in 1999, with $2.16 billion sale in 2000, was withdrawn in 2004 because of increased cardiovascular risk [8]. By inhibiting the COX-2-driven prostacyclin synthesis in the wall of blood vessels but not the COX-1-driven thromboxane-A2 in blood platelets/thrombocytes, the selective COX-2 inhibitor such as rofecoxib causes very serious cardiovascular complications, promoting platelet aggregation and precipitating strokes and many premature deaths. An estimate of a large number of needless deaths (>80,000 by the end of 2004) has been made [6]. The appropriate use of coxibs and NSAIDs has prompted a scientific statement from the American Heart Association to the clinicians [9]. For patients with a prior history of or at high risk for cardiovascular disease, the use of COX-2 inhibitors for pain relief should be limited to patients for whom there are no appropriate alternatives, and then only in the lowest dose and for the shortest duration. The coxibs gained Food and Drug Administration approval based on a superior side effect profile in gastrointestinal endoscopy studies compared with traditional NSAIDs. The problem becomes substituting one side effect for another. Thus, the demand for anti-inflammatory agents remains high, and an urgent need to develop novel therapeutics is warranted.

## 2 IMMUNOBIOLOGY OF INFLAMMATION

Before discussing potential targets for anti-inflammatory therapeutics, we briefly review the inflammatory process, noting the triggers and mediators, as well as the cells and tissues involved in inflammation. (See Fig. 1 for a schematic flowchart of inflammation.)

Inflammation is a complex integrated immune response to cell/tissue injury [10]. Triggered by many types of stimuli (physical, chemical, immunological, or microbial), the process normally initiates the elimination of noxious agent and disposal of damaged tissue in order to promote wound healing. A highly coordinated acute phase is directed toward the site of injury and aims to quickly remove the inflammatory stimulus. During acute inflammation, effector mechanisms of the innate and acquired immune response are activated, and multiple mediators are engaged. Mediators of inflammation vary from small molecules to proteins, and different ones are produced at different phases of inflammation. From plasma, inflammation activates the kinin system to release molecules such as bradykinin, the complement system for molecules such as C3 fragments, and the fibrinolytic system for ones such as fibrinopeptides. Other mediators are produced in cells and tissues. In response to stimulus, effector immune cells, polymorphonuclear neutrophils (PMNs, first line of defense) followed by macrophages, are propagated in increased numbers and recruited to the injured site. The activated macrophage is one main source of inflammation mediators [11]. At the inflamed site macrophages produce mediators that generate the local and systemic symptoms of inflammation [11, 12]. Local acute inflammatory response is demonstrable by the cardinal signs of inflammation, and the phenomenon is characterized by transient vasodilation and increased capillary permeability. The combined efforts of exudate and cells are expected to destroy the inflammatory trigger and restore the injured site of the body.

The local response is accompanied by a systemic reaction known as the acute-phase response. The systemic response is marked by induction of fever in the hypothalamus, increased synthesis of corticosteroids in the adrenal cortex, increased production of leukocytes (leukocytosis) in the bone marrow, and synthesis of acute-phase proteins such as the C-reactive protein in the liver. The systemic response is mediated by the pro-inflammatory cytokines TNF-α, IL-1β, and IL-6 produced by activated macrophages at the site of inflammation.

Whereas PMNs and macrophages are players at early inflammation, at later stages lymphocytes (T and B cells) may participate. Leukocyte trafficking is essential for the inflammatory process. They are transported via the blood vessel, and endothelial cells in the vicinity of the injured site are activated to allow passage of immune cells from blood vessels to the inflamed site [13]. Processes involving chemotaxis, cell adhesion,

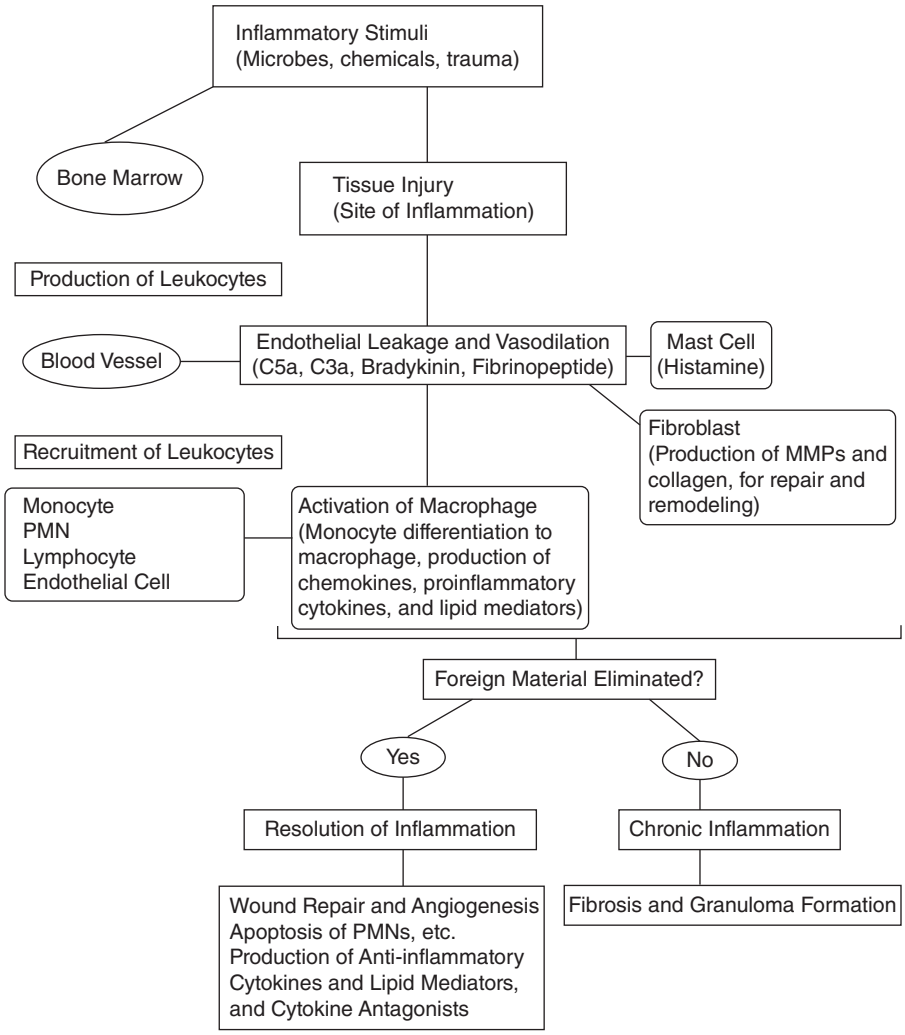


**Figure 1** Flowchart of inflammation. Major molecular and cellular events of the inflammatory process are presented in the scheme, with special emphasis on temporal changes at the site of inflammation.

and migration are active in this scenario. During inflammation, additional cells may be involved. In early inflammation, mast cells may produce vasoactive amine histamine. In late inflammation, fibroblasts may produce collagen for tissue remodeling. Ultimately, a successful immune response leads to elimination of stimulus/pathogen and resolution of inflammation [14].

If the initial inflammatory response fails to remove the stimulus/irritant and restore the damage, a chronic proliferative phase sets in. Tissue degeneration and fibrosis may occur. Chronic inflammation is evidenced by disease pathology and may require therapeutic intervention.

## 3 MOLECULAR TARGETS OF ANTI-INFLAMMATORY HERAPEUTICS

A multitude of molecular targets of anti-inflammatory therapeutics are known. The inflammatory stimulus signals the immune cell to activate its kinases (such as mitogen-activated protein kinase p38) and transcription-factors [nuclear factor kappa B (NFκB)], leading to the production of cytokines (IL-1β) and various effective molecules (nitric oxide and prostaglandins) to act at the site of inflammation [15]. For these numerous targets of the inflammatory response, our challenge is to place a value on one target relative to another and to develop strategies to modulate them [16]. Historically, too much

**TABLE 1 Examples of Molecules for the Activation and Resolution of Inflammation**

| Type | Activation | Resolution |
|---|---|---|
| Lipid mediators | Prostaglandins | 15d-$PGJ_2$[a] |
| | Leukotrienes | Lipoxins |
| | PAF | Resolvins |
| | | Protectins |
| Cytokines | TNF-α | TGFβ |
| | IL-1β | IL-10 |
| Transcription factors | p65/p50 NFκB | p50/p50 NFκB |
| Proteinases | TACE (ADAM-17) | Caspases for apoptosis |
| | ICE (Caspase-1) | |

[a]15d-$PGJ_2$ is a cyclopentenone prostanglandin.

emphasis has been put on monotherapy, treating inflammation with a single agent. With the current knowledge of multiple targets, a judicious combination may be a better approach: particularly when each component is targeted specifically at carefully chosen pro-inflammatory/catabolic proteins, their receptors or their induction mechanisms, perhaps programmed to be used sequentially [6]. Traditional therapeutics have been developed by blocking what initiates and drives inflammation. A novel approach is to design therapeutic agents based on the molecules necessary for the resolution of inflammation [17]. (See Table 1 for examples of molecules for activation and resolution of inflammation.)

### 3.1 Kinases

Reversible phosphorylation via kinase-phosphatase activities, essential for signal transduction, plays a critical role in the inflammatory process. Two kinases in particular have come under scrutiny as potential anti-inflammatory therapeutics.

***Mitogen-Activated Protein Kinase p38 (MAP Kinase p38)*** Originally identified as an enzyme that is phosphorylated and activated by lipopolysaccharide (LPS) in stimulated monocytes, the p38 serine/threonine protein kinase is involved in the synthesis of pro-inflammatory cytokines TNF-α and IL-1β. Four isoforms are known, p38α, p38β, p38γ, and p38δ, with p38α representing a point of convergence of multiple signaling processes during inflammation. This isoform is a potential key target for modulation of cytokine production, and over 100 different inhibitors have been generated [18]. All potent inhibitors are adenosine triphosphate (ATP) competitive and have a significant presence in the adenosine pocket, as seen from the enzyme's crystal structure. The orally active VX702 (Vertex), together with SCIO 469 and SCIO 323 (Johnson & Johnson), are undergoing early clinical trials. AMG 548 (Amgen) and BIRB 796 (Boehringer Ingelheim) will not be developed due to toxicity (elevation of liver enzymes). For a functional inhibitor to be anti-inflammatory, its selectivity over other kinases and the safety issue over long-term use need to be assured.

***Phosphotidylinositol 3-Kinase (PI3K)*** Among the three PI3K classes, the class IB phosphotidylinositol 3-kinase p110γ (PI3Kγ) is another kinase of importance in inflammation. Two isoforms are PI3Kα and PI3Kβ. PI3Kγ, activated by the seven-transmembrane-spanning G-protein-coupled receptors (GPCRs), is the downstream signaling component of a wide range of chemokine receptors. However, it only controls a fraction of the various cellular processes in which it is involved. Therefore, this can yield a "soft" mode of inhibition. Even the most potent inhibitor would only dampen down any given biochemical event. Many inhibitors have been generated, and TG100-115 (3-[2,4-diamino-6-(3-hydroxyphenyl)pteridin-7-yl]phenol), found selective for PI3K in a panel of over 190 kinases, is currently under preclinical development. Although PI3Kγ inhibition has been listed as "an aspirin of the 21st century," the challenge, similar to p38 kinase, is to strike the right balance between isoform selectivity and potency and to combine maximum efficacy with minimum side effects [19].

### 3.2 Transcription Factors

Transcription factors regulate gene expression. The ones implicated in inflammation are nuclear factor kappa B (NFκB), activating protein-1 (AP-1), nuclear factor of activated T cells (NFAT), and signal transduction and activators of transcription (STATs), with NFκB being the most notable [10].

***Nuclear Factor Kappa B*** NFκB is pivotal in the inflammatory response by regulating the expression of pro-inflammatory cytokines, chemokines, adhesion molecules, and inducible enzymes. It also controls the expression of antiapoptotic genes that might protect inflammatory leukocytes from programmed cell death, an important determinant for the resolution of inflammation. The transcription factor works either as hetero- or homodimer, after different combinations of p50, p52, c-Rel, RelA (p65), and RelB subunits. It binds to the DNA sequence known as the κB site and modulates target gene expression. However, under normal conditions, the dimer resides in the cytoplasm in an inactive state by association with the IκB inhibitor proteins.

Upon stimulation (e.g., by the pro-inflammatory cytokine TNF-α), the inhibitor protein is rapidly phosphorylated, ubiquitinated, and then degraded in the proteasome, leaving the activated transcription factor to enter the nucleus. Both phosphorylation and degradation of IκB inhibitor are key steps in the regulation of NFκB, and the IκB kinase (IKK) complex is involved. IKK contains two kinases, IKKα and IKKβ, plus an associated scaffold-like regulatory protein named NFκB essential modulator (NEMO, alias IKKγ). Furthermore, there are multiple pathways for the activation of NFκB, with the classical (canonical) pathway leading to p50/RelA, and the alternative (noncanonical) pathway to p52/RelB [17, 20].

Inhibitors of NFκB have been shown to act at various levels of the complex pathway. Over 750 inhibitors are known, including natural products, chemicals, metals, metabolites, synthetic compounds, peptides, and proteins (from various sources: cellular, viral, bacterial, and fungal). Some are antioxidants, engineered dominant-negative or constitutively active polypeptides, small ribonucleic acid (RNA)/DNA, and even NSAIDs. Aspirin inhibits activation of NFκB in cell culture, although at a higher concentration than the ones used for COXs. Druggable components of the NFκB pathway are actively being pursued, and some are in phase II clinical trials. The aim is to find molecules that block specific pathways leading to NFκB activation without interfering with other signaling cascades, and to discover the ones that can inhibit distinct NFκB complexes induced by select stimuli in specific cell types [21, 22].

### 3.3 Lipid Mediators

Lipid mediators are the most immediate targets of inflammation. The study of eicosanoids (from the Greek *eicosa*, twenty, for 20 carbon fatty acid derivatives) leads to potent eicosanoid lipid mediators prostaglandins and leukotrienes, both derived from the phospholipase-released arachidonic acid. The cyclooxygenases produce prostaglandins and the lipoxygenases produce leukotrienes, both involved in inflammation as well as homeostatic functions of the body. Lipid mediators cause the cardinal signs of inflammation. As an example, prostaglandins elicit increased sensitivity to touch, by sensitizing the free end of pain neurons in peripheral inflammation. Therefore, the lipid mediator pathway of inflammation continues to be a target for the development of anti-inflammatories and analgesics [23, 24].

***Cyclooxygenase (COX)*** Inhibitors of COX-1 and COX-2 are anti-inflammatory agents currently in use. COX-targeting NSAIDs are effective, and aspirin is the most common nonprescription drug. Although COX-2 targeting coxibs have suffered recalls, they nonetheless remain as important therapeutic agents [25, 26]. With careful monitoring of adverse cardiovascular side effects, the COX-2 inhibitors celecoxib (trade name Celebrex, from Pfizer) and parecoxib (trade name Dynastat, from Pfizer) are valuable options for the management of postoperative pain. Parecoxib, the water-soluble prodrug of valdexozib, is the first injectable COX-2-specific inhibitor approved in Europe for short-term treatment of moderate-to-severe postoperative pain [27]. However, because of market removal of rofecoxib (trade name Vioxx, from Merck) and valdezozib (trade name Bextra, from Pfizer) in 2004 and 2005, respectively, future development of anti-inflammatories based on the lipid mediator pathways most likely will be concentrating on other targets.

***Lipoxygenase (LOX)*** Dual-acting anti-inflammatory drugs currently under development are ones that inhibit both cyclooxygenases (COX-1 and COX-2) and 5-lipoxygenase (5-LOX) [28]. The enzyme 5-lipoxygenase (5-LOX) generates leukotrienes, and some leukotrienes are pro-inflammatory. As an example, leukotriene $B_4$ ($LTB_4$) is a potent chemotactic agent that induces chemotaxis of eosinophils, neutrophils, and monocytes to the inflamed site. The classic NSAIDs act by inhibiting the COXs. One of their main drawbacks is the curtailed production of gastroprotective prostaglandins, together with the concurrent increased production of gastro-damaging leukotrienes, because blockage of COX pathways shifts the eicosanoids to the LOX pathway. Dual inhibitors may prevent this from occurring [29]. Licofelone, a balanced dual inhibitor, reduces inflammation in a rabbit model of arthritis [30]. Besides inflammation, dual inhibitors show additional benefits. They are neuroprotective and anticarcinogenic [31].

***Phospholipase (PL)*** Whereas prostaglandins and leukotrienes are targeted by dual inhibitors against COX and LOX, another lipid mediator of inflammation, platelet-activating factor (PAF), a product from lysophospholipid, is still being generated. In order to inhibit PAF, it is necessary to develop inhibitors of phospholipase. Phospholipase $A_2$ ($PLA_2$) catalyzes hydrolysis of phospholipid to generate fatty acid and lysophospholipid. The fatty acid arachidonic acid becomes substrate for COX and LOX, and the lysophospholipid is converted into PAF. Therefore, inhibitors of $PLA_2$ offer the potential to stop the production of a more complete set of inflammatory lipid mediators through the blockade at the onset of the cascade of reactions that follow arachidonic acid release.

A total of 18 human $PLA_2$ are known. Divided into three major classes, the secretory phospholipase $A_2$

group IIA ($sPLA_2$ IIA) becomes the target for anti-inflammatory therapeutics, and many inhibitors have been generated. Whereas animal studies have shown efficacy in models of inflammatory disorders, clinical trials have been less successful. As an example, the potent, selective, and bioavailable $sPLA_2$ IIA inhibitor LY315920 (from Eli Lilly) offered no overall survival benefit to patients with sepsis, although this enzyme is known to be elevated during sepsis. Sepsis is an important medical condition, with estimates of 750,000 episodes annually in the United States, resulting in associated mortality of 30% after 28 days. Other groups of $sPLA_2$, such as group V and X, as well as the cytosolic enzyme ($cPLA_2$) may be compensatory [32]. Further study on this inflammatory drug target is warranted.

***Lipoxins, Resolvins, and Protectins*** Lipoxins, resolvins, and protectins are lipid mediators active at the resolution stage of acute inflammation. Stable analogs of these molecules may be potential anti-inflammatories. Current investigations suggest that the biosynthesis of lipid mediator during acute inflammation is biphasic, with a role for eicosanoids both at the initiation and termination of the inflammatory response, and there is eicosanoid class switching between the two stages [14, 17, 33].

Lipoxins (LXs) and their 15-epimers, aspirin-triggered lipoxins (ATL), are major promoters of resolution of inflammation. Lipoxin is an acronym for lipoxygenase interaction products. First discovered by incubation of human leukocytes with arachidonic acid, they stimulate monocyte recruitment and the uptake of apoptotic PMNs by macrophages and reduce vascular permeability to stop PMN infiltration. Therefore, lipoxins are proposed to act as an endogenous "braking signal" in inflammation; and aspirin jump starts resolution by triggering the biosynthesis of ATL [34, 35].

Besides lipoxins, resolvins and protectins are stereoselective, bioactive, local lipid mediators for the resolution of inflammation. Resolvin is an acronym for resolution-phase interaction products. They are derived from the polyunsaturated fatty acids (PUFAs) eicosapentaenoic acid (EPA) and docosahexaenoic acid (DHA). Protectins and neuroprotectins (protectins generated in neural tissues) are oxygenated products derived from DHA and have anti-inflammatory and neuroprotective actions in animal models. Thus, these molecules possess anti-inflammatory and immunoregulatory activities [14].

Stable analogs of lipoxins, resolvins, and protectins have been chemically synthesized and tested in animal models of inflammation. As an example, the resolvin E1 analog, 19-(*p*-fluorophenoxy)-Rev E1, resists rapid metabolic inactivation and retains bioactivity in reducing PMN infiltration and pro-inflammatory cytokine/chemokine production in vivo [36]. The synthetic stable versions of these proresolving lipid mediators may have clinical applications if proven safe and active.

### 3.4 Receptors

Several groups of receptors are potential targets for anti-inflammatory therapeutics. Cell surface receptors serve to transmit extracellular signals to cellular interior. Different receptors recognize specific signals and act accordingly. For example, both lipoxin $A_4$ and ATL 15-epi-lipoxin $A_4$ act on the G-protein-coupled $A_4$ lipoxin receptor (ALXR, alias FPRL1, or formyl peptide receptor-like 1) [17]. The discovery of hot pepper ingredient capsaicin action on the transient receptor potential vanilloid receptor 1 (TRPV1), a cation channel, opens new horizons in the field of analgesics and pharmacology of norciceptors [37].

***G-Protein-Coupled Receptors*** GPCRs are the most prevalent class of cell surface proteins in pharmaceutical research today. In fact, GPCR-targeting drugs account for one-tenth of worldwide pharmaceutical sales. Of the over 800 GPCRs in the human genome, about 30 have been identified as potential therapeutic targets for the treatment of inflammatory diseases. Examples are receptors for kinins such as bradykinin, complements such as C3a and C5a, and chemokines such IL-8. Nonpeptide ligands that bind peptide-activated GPCRs in inflammation have been produced. Because of their important roles as cellular sentries in inflammatory diseases, GPCRs deserve a great deal more attention as potential therapeutic targets [38].

***Glucocorticoid Receptor (GR)*** Glucocorticoids, cortisol in humans and corticosterone in rodents, are the first examples of endogenous anti-inflammatory molecules, and they have potent anti-inflammatory and immunosuppressive actions. These molecules interact with the receptor, which is a ligand-activated transcription factor. GR acts by masking the DNA-binding domain of other transcription factors such as NFκB, NFAT, AP-1, and STAT. Glucocorticoids are widely used in the clinical control of severe pathologies with an inflammatory etiology, but there are complications due to side effects [39].

Glucocorticoids (GCs) have been historically linked to annexin 1. Originally described as a GC-induced inhibitor of $PLA_2$ and prostaglandin generation in activated macrophages, annexin 1 shares the same receptor as the GPCR $LXA_4$ receptor. Therefore, GC, annexin, and lipoxin are all key players for the resolution of inflammation. Novel GCs with improved specificity

such as the nitro-steroids nitro-prednisolone, and annexino-mimetic, are currently being developed. Thus, the GR pathway provides specific targets for innovative anti-inflammatory drug discovery [40].

***Peroxisome Proliferator-Activated Receptors*** Peroxisome proliferator-activated receptors (PPARs) have strong links with inflammation. Classified into three isotypes, PPARα, PPAR β/δ, and PPARγ, they are ligand-activated transcription factors that act as heterodimer with another transcription factor named retinoid X receptor (RXR). Suppression of inflammation by PPARs has beneficial effects, as shown for the progression of inflammatory diseases in animal models. Currently, several PPAR ligands are in clinical use, notably fibrate for lipid disorder, and rosiglitazone (trade name Avandia, from GlaxoSmithKline) for diabetes. However, the use of PPAR ligands in inflammatory diseases awaits further development [41].

***Toll-Like Receptors*** Toll-like receptors (TLRs) are important in host defense against microbial pathogens. They function as primary sensors of microbial products and activate signaling pathways that induce the expression of immune and pro-inflammatory genes. This pathway must be tightly regulated because the immune system needs to constantly strike a balance between activation and inhibition to avoid detrimental and inappropriate inflammatory responses. Negative regulators of the TLR pathway may lead to novel therapeutic approaches to inflammatory diseases. As an example, the peptide RDP-58 that blocks TLR, is in phase II clinical trial for Crohn's disease [16, 42, 43].

### 3.5 Cytokines

Cytokines continue to be fruitful targets of anti-inflammatory therapeutics. Immune cells produce a diverse range of cytokines serving different functions. Initially, polymorphonuclear neutrophils and macrophages produce the cytokines TNF-α and IL-1β [12]. These pro-inflammatory cytokines are current targets for inflammatory diseases. Later stage cytokines produced by macrophages and T lymphocytes, such as IL-10 and transforming growth factor-β (TGF-β), have important functions in the suppression of inflammation [44–46].

***Tumor Necrosis Factor-α and Interleukin-1β*** TNF-α and IL-1β are situated at the apex of the cascade of inflammation. Therefore, their inhibition is effective for many inflammatory diseases including rheumatoid arthritis [47]. Although therapeutic proteins are expensive to manufacture, all existing agents are proteins. Three anti-TNF-α therapies are available: infliximab (trade name Remicade, from Johnson & Johnson) is a chimeric monoclonal antibody for TNF, etanercept (trade name Enbrel, from Amgen and Wyeth) is a soluble TNF-receptor construct, and adalimumab (trade name Humira, from Abbott) is a human anti-TNF monoclonal [48]. These anti-TNF agents have been touted as the biggest advance in the management of inflammatory arthritis in over 50 years [49]. An anti-IL-1 therapy is available: anakinra (trade name Kineret, from Amgen) is recombinant human IL-1 receptor, antagonist (IL-1RA). IL-1RA acts by binding to the IL-1 receptor, thus blocking the binding of IL-1β to the receptor so as to inhibit signal transduction. While anti-TNF therapy may be more effective, treatments can lead to expression of latent infections, and anti-IL-1 therapy may be preferable under such conditions.

Novel therapies are being developed targeting other cytokines, such as IL-6 and IL-17. Improvement in therapy with regards to cost, efficacy, and tolerability are being made. In addition, there is an urgent need for treatments that are remission inducing rather than just disease suppressing. As an example, in arthritis, disease-modifying antirheumatic drugs (DMARDs) are more attractive than ones merely for the alleviation of symptoms.

### 3.6 Proteinases

Proteinases are appealing targets for development of anti-inflammatories. Cytokines such as TNF-α and IL-1β are synthesized as precursor proteins, and functional cytokines come only after polypeptide processing by proteinases. Furthermore, in inflammatory diseases such as arthritis, proteinases are actively degrading the cartilage collagen proteins. Therefore, modulators of proteinase activity are being investigated.

***Tumor Necrosis Factor-α Converting Enzyme*** Tumor necrosis factor-α converting enzyme (TACE) is a zinc metalloproteinase. The enzyme processes the membrane-bound pro-TNF protein into the released form for circulation. It is postulated that small-molecule inhibitors against TACE will be effective and inexpensive to manufacture. From a medicinal chemistry standpoint, remarkable progress has been made, resulting in the design and synthesis of a large number of potent and differentially selective compounds. These molecules can decrease the level of TNF-α in cell and animal models of inflammatory disease. However, very few end up in clinical trials, with a couple discontinued after phase II, and one (TMI-005 from Wyeth) still in phase I. Concerns over TACE inhibitors include a possible buildup of membrane-bound TNF precursors and their functional

significance, as well as selectivity of the inhibitor toward other metalloproteinases [50].

***Interleukin-1β Converting Enzyme/Caspase-1*** Interleukin-1 converting enzyme (ICE) is a cysteine endoproteinase. Its active form is a heterotetramer, with two p20 and two p10 subunits. The enzyme processes the pro-IL-1 protein into mature functional cytokine that is released into the extracellular environment. Secreted IL-1 binds to IL-1 receptor of the target cell to initiate inflammatory response. Like TACE, small-molecule inhibitors have been sought, and their fate is also similar to TACE. For example, the orally bioactive inhibitor pralnacasan (from Vertex) showed efficacy in phase III trials for rheumatoid arthritis, osteoarthritis, and psoriasis but was discontinued in 2005 due to toxicity. One concern is that the inhibitors may react with other endogenous nucleophiles in the body, leading to toxic side effects associated with long-term use in chronic inflammation [50]. The current nomenclature has renamed ICE as caspase-1.

***Caspases and Inflammasomes*** Caspases are proteinases involved in programmed cell death or apoptosis. Apoptosis is active during the resolution phase of inflammation for the death of polymorphonuclear neutrophils. Inflammatory caspase-1 is activated in a multiprotein complex known as the inflammasome, a structure with associated regulatory proteins and is essential for the processing and secretion of IL-1β. Since current therapies are solely based on neutralization of cytokines that are already in circulation, targeting the inflammasome may prevent their generation and provide a novel approach to anti-inflammatory therapy [51].

***Matrix Metalloproteinases*** TACE is a zinc metalloproteinase, and there are many other metalloproteinases. Originally discovered for the degradation of extracellular matrix such as collagen, they become important enzymes in inflammation and cancer. A nomenclature overhaul changes their names from collagenase, gelatinase, and stromelysin to matrix metalloproteinases (MMPs). Besides the MMPs, there is another group named ADAMs (a disintegrin and metalloproteinases), multidomain metalloproteinases with additional disintegrin domains that are anchored in cell membranes. TACE is a member of ADAMs, alias ADAM-17. A third group is named ADAMTSs (ADAMs with a thrombospondin motif), with aggrecanase as an example. The aim is to develop highly selective and preferentially orally active MMP inhibitors. Unexpectedly, synthetic MMP inhibitors have been disappointing in clinical trials for cancer treatment. However, they may be useful for inflammatory diseases, acknowledging the fact that prolonged clinical application would be necessary for chronic inflammation [52]. For diseases such as arthritis, inhibitor selectivity may be less of a problem because it would be beneficial to inhibit both TACE/ADAM-17 together with other MMPs that degrade the extracellular matrix.

### 3.7 Free Radicals

Free radicals are active in various steps of inflammation. Major ones are reactive oxygen species (ROS) and reactive nitrogen species (RNS). Free radicals target redox-responsive transcription factors such as NFκB [53]. Products of ROS such as hydrogen peroxide have bactericidal properties, and the NADPH oxidase complex from neutrophils is responsible for their production. An important RNS is nitric oxide (NO).

***Nitric Oxide*** NO is produced by the enzyme nitric oxide synthase (NOS). Three known types are endothelial (eNOS), neuronal (nNOS), and inducible (iNOS). NO from iNOS is produced in high-volume output and serves antimicrobial and other functions, including modulation of COX enzymes [54]. The NO free radical was named Molecule of the Year in 1992 by *Science* magazine because of its multiple functions from microbicidal to vessel relaxation (NO was originally named endothelium-derived relaxing factor, or EDRF). However, the use of NO as target of anti-inflammatory can be tricky. The actions of NO in relation to inflammation are multifaceted and as a consequence it remains unclear if NO is pro- or anti-inflammatory; ultimately it is likely to be both, very much dependent on the local concentrations of NO and the prevailing immunological status of the host [55]. Therefore, NO has been described as the "double edge sword mediator," a phenomenon also referred to as the NO paradox. Design of anti-inflammatory therapeutics based on NO must have taken into account this dual nature of NO [56].

### 3.8 Other Targets

Because inflammation is a complex process, it offers ample targets for development of anti-inflammatories. In addition to the ones already mentioned, many others exist. The treatment of multiple sclerosis offers an example. Multiple sclerosis involves inflammation of white matter of the central nervous system, and there is a corresponding animal model, experimental autoimmune encephalomyelitis (EAE). Available current immunotherapeutic potentials of action include: (1) adhesion molecule blockade for specific integrins, (2) cytokines for TNF-α, TGF-β, interferon (IFN), (3) antigen-specific bystander suppression, (4) B-cell

depletion/reduction, (5) chemokine receptor blockade, (6) co-stimulation inhibition for cytotoxic T-lymphocyte-associated protein-4 (CTLA4), (7) depletion/inactivation of pathogenic T cells, (8) Fc-mediated tissue injury/antibody production, (9) general immunosuppression, (10) autologous stem cell rescue, and (11) increase pregnancy-associated hormone signaling [57]. Many of these targets are common to other inflammatory diseases; several targets are further discussed below.

***Chemokines*** Leukocyte trafficking is a continuous process during inflammation. Thus, adhesion molecules, chemotactic molecules, or chemokines and chemokine receptors are valuable targets for anti-inflammatories [58].

***Heat Shock Proteins*** Heat shock proteins (HSPs) are cellular stress proteins and their synthesis is upregulated during stress, including inflammation. Acting as molecular chaperones, they have paradoxical effects toward cell injury. HSP induction before a pro-inflammatory stimulus is beneficial but HSP induction after a pro-inflammatory stimulus is cytotoxic. They are abundant intracellular proteins and are also very immunogenic. In addition, specific TLRs may serve as HSP receptors. Modulating HSP levels and their receptors may be potential drug targets in inflammation [59, 60].

***More Proteins, Peptides, and Other Molecules*** At the early stage of inflammation, there is activation of the complement system. The products C3a and C5a are known as anaphylatoxins and are potent pro-inflammatory mediators [61]. Complement components are targets of anti-inflammatories. The pro-inflammatory cytokines include interleukin-1β. IL-1β binds to its receptor, and this is followed by receptor clustering into a multiprotein aggregate known as a cell adhesion complex. Blocking this process with peptides is another novel anti-inflammatory strategy [62]. The suppressor of cytokine signaling (SOCS) proteins regulate cytokines. Their use for suppressing cytokine signaling may be therapy for the treatment of inflammatory diseases [63]. In specific diseases such as arthritis, loss of cartilage and bone can occur. Under such conditions bone morphogenetic proteins (BMPs) may be applicable for repair [64]. In neuroinflammation, neuropeptide Y (NYY) and related peptides may have a role in clinical application [65]. Moving from proteins to carbohydrates, hyaluronan (HA), the long-chain polysaccharide composed of repeats of disaccharide units of glucuronic acid and *N*-acetylglucosamine as a component of extracellular matrix, seems to have a role in modulating inflammation. Interestingly, the effect is dependent on molecular size, with larger polysaccharide chains having anti-inflammatory activity and smaller ones having pro-inflammatory properties. This knowledge may assist investigators to harness the beneficial effects of HA in anti-inflammatory therapeutics [66].

## 4 EXPERIMENTAL METHODS FOR IDENTIFYING ANTI-INFLAMMATORIES

From in vitro cultured cell assays to in vivo animal models, various experimental approaches are available for the development of anti-inflammatory therapeutics. The benefits and limitations of these tools must be recognized, so the investigator can make an informed decision in choosing the most appropriate experimental system.

### 4.1 In Vitro Assays of Inflammation

In vitro assays are mostly based on cell cultures. They are attractive because experiments are fast and economical. The assays are relatively easy to perform, and many are also amendable to high-throughput formats (for mass screening). From the standpoint of anti-inflammatory therapeutics, activity of proteins and enzymes can be tested using cell lysates or recombinant proteins. Monocytes and macrophages are central in the inflammatory response, and rodent and human cell lines are readily available. In addition, normal, unstimulated monocytes/macrophages can be obtained from blood samples of human volunteers and from the mouse peritoneum. For example, we have used the human cell line MonoMac 6 to demonstrate the inhibition of NFκB activation by the phytochemical curcumin using the electrophoretic mobility shift assay (EMSA) [67]. However, EMSA is a cumbersome assay. A commercially available alternative is to test the activated transcription factor by an enzyme-linked immunosorbent assay (ELISA)-based nuclear binding assay, utilizing immobilized κB-specific oligonucleotide and nuclear lysates. In addition, other assays such as IκB kinase activity are available for the NFκB signaling pathway. To check the inhibition of transcription-factor-dependent gene expression, NFκB response element sequence can be linked to a reporter gene such as the firefly luciferase, yielding an easy and highly sensitive luminescence assay [68].

Assays should be selected depending on the targets of inflammation. For example, to test the secretion of lipid mediators and cytokines by immune cells, ELISA method for specific prostaglandins or cytokines can be used. Fluorescence-based immunoassays measuring multiple cytokines from single samples (multiplexing) are also available. Different cell types are appropriate

for specific purposes, for example, endothelial cells for adhesion molecule expression and macrophages for phagocytosis. We have isolated murine peritoneal exudate cells (macrophages) and showed that iNOS activity (stimulating the cells with lipopolysaccharide and interferon γ and testing NO production with Griess reagent) can be inhibited by phytochemicals [69]. Arthritis is an important inflammatory disease. Proteinases such as MMPs can be assayed with various peptide substrates that allow sensitive fluorometric detection. In vitro model of cartilage degradation is possible, by incubating isolated human and calf cartilage with cytokines such as IL-1β, and checking for the released degradation products hydroxyproline and glycosaminoglycan in culture medium. Potential inhibitors of cartilage breakdown can be added to monitor their effects [68].

### 4.2 In Vivo Models of Inflammation

In vivo models of inflammation involve laboratory animals. The process to assay potential anti-inflammatories normally progresses from in vitro to in vivo because it is often difficult to predict inhibitory potency and selectivity obtained from in vitro experiments for clinical usefulness. Additional concerns are delivery and bioavailability: how to send the therapeutic agent to the site of inflammation and how to prevent its modification and degradation by the body's metabolism of xenobiotics. For in vivo works, animal welfare is an increasingly significant issue. Regulations on this subject vary by country. In the United States, approval for animal protocols must be obtained from the Institutional Animal Care and Use Committee (IACUC) before initiating any experimentation [70].

Many animal models are available for testing anti-inflammatories in vivo. Rats and mice are more economical, but guinea pigs, rabbits, and dogs have established models for diseases such as arthritis [68, 70]. Rats and mice are rodents, with rats being larger in size for sampling and analysis, and mice demonstrating the advantage of genetics (from inbred strains to transgenic and knockout animals). For example, we have used a murine sepsis model by injecting lipopolysaccharide into animals to show in vivo effect of the phytochemical curcumin [71]. For studies on potential therapeutics for arthritis, a useful model is rodent collagen-induced arthritis (CIA). We first inject chicken collagen (type II) and adjuvant to the DBA strain of mice. CIA is then followed for several weeks by measuring the increase in paw thickness and the changes in inflammatory state by histology [70].

With the abundance of animal models for anti-inflammatory research, some more costly than others, judicious choice for the ones to be used should be based on specific objectives. An example for a simple model is carrageenan-induced rat paw edema (CPE), a test that can be completed in less than a day. CPE is involved in the launching of indomethacin, although delayed response of fibrosis has been neglected [6]. Another quick assay is acute inflammation of rodent ears after stimulants such as tetradecanoylphorbol acetate (TPA). A third model is the air-pouch: in 1953, Hans Selye originally described that injection of air into the dorsal surface of rats provides a cavity in which the inflammatory response can be studied. The model allows performance of biochemical analysis of both the exudate and inflammatory cells, together with histological analysis of the air-pouch lining. Modifications such as the murine carrageenan air pouch and the chronic granulomatous air pouch have been reported [68].

Specific animal models are available for different inflammatory diseases, such as asthma, skin and neuro-inflammation, and inflammatory bowel disease (IBD) [72]. For IBD, there are models using hamsters and mice by feeding dextran sulfate sodium (DSS solution) to animals to induce colitis, with treatment for 5 days for acute and 21 days for chronic colitis. The hypothesis is that DSS converts to hydrogen sulfide, which is toxic to gut epithelium.

A general strategy for testing anti-inflammatory therapeutics proceeds from cells to animal studies. However, for potential therapeutics such as phytochemicals already in use, a jump from in vitro assays to humans is possible. Hops (used in beer brewing) are a useful example. Screening of phytochemicals in vitro against COXs in cell lines generated potential therapeutics (using ibuprofen as positive control); testing in human volunteers leads to the conclusion that a hops extract may be useful as anti-inflammatory dietary supplement [73].

We end this section with a word of caution: Potential anti-inflammatories may not work in a specific animal model. For example, chloroquine and minocycline (currently in use for arthritis) are inactive in rodent arthritis models [6]. Shortfall of each animal model must be recognized. If the budget permits, multiple experimental models should be used so as to complement each other.

## 5 PIPELINE NOVEL ANTI-INFLAMMATORIES

For anti-inflammatory therapeutics, what comes next? Many compounds have been tested but then withdrawn at different stages of development. A bright spot for the immediate future will be the nitric-oxide-releasing non-steroid anti-inflammatory drugs (NO-NSAIDs). This is

a new class of anti-inflammatory and analgesic drugs, generated by adding nitroxybutyl or nitrosothiol moiety to the parent NSAID via a short-chain ester linkage. Two compounds, NO-aspirin and NO-naproxen, are undergoing phase III clinical trials [74]. Effectiveness has already been assessed in vitro, in vivo, and in human volunteers. NO released from NO-naproxen inhibits both COX-1 and COX-2 in cell cultures in vitro [75]. In animal models, NO-diclofenac works in carrageenan-induced paw edema, and NO-naproxan in rat adjuvant-induced arthritis, although potential activation of cartilage metabolism by NO is a concern that needs to be addressed [56, 76]. In human trials, NO-aspirin spares the gastric mucosa while inhibiting platelet aggregation and allowing the production of aspirin-triggered lipoxin (ATL) [77, 78]. Thus, the combination of balanced inhibition of COX-1 and COX-2 with the release of NO confers to NO-NSAIDs a reduction in gastrointestinal and cardiorenal toxicity. These NO-NSAIDs have been renamed; now they are known as cyclooxygenase inhibitor NO donors (CINODs) [74].

A footnote can be added to the CINOD discussion. Adding NO to other anti-inflammatory therapeutics has been achieved, such as the NO-releasing steroids NO-prednosolone [40]. Hydrogen sulfide ($H_2S$) can be used instead of NO because it also modulates inflammation at physiological concentration. The $H_2S$ donors increase resistance of the gastric mucosa to injury and accelerate repair. Hence $H_2S$-releasing NSAIDs, such $H_2S$-releasing derivatives of indomethacin and diclofenac, are being investigated [79].

## 6 NATURAL ANTI-INFLAMMATORY COMPOUNDS IN FOODS AS DIETARY THERAPEUTICS

Natural products have played an essential role in drug discovery. Recent examples are lovastatin for cholesterol lowering, artemisinin for malaria, and paclitaxel for cancer [80]. Starting from the story of willow bark leading to aspirin, phytochemicals commonly have anti-inflammatory activity. A total of 38 plant species has been identified in a review on patented anti-inflammatory plant drug development, with the botanicals showing diverse chemical structures, polysaccharides, terpenes, curcuminoids, alkaloids, and many others [81]. A proprietary avocado soybean product has been reported to be effective in the treatment of painful osteoarthritis and chronic lower back pain [82]. A future anti-inflammatory drug for rheumatoid arthritis and other diseases may come from the Chinese herb, *Tripterygium wilfordii*, locally known as "thunder god vine." Its terpene triptolide and other metabolites act on inflammation targets such as COX-2, TNF-α, iNOS, and NFκB [83].

Whereas phytochemicals continue to yield products with anti-inflammatory therapeutics, the variety of fruits, vegetables, beverages, meat, seafood, and spices consumed by the individual also contains compounds with such properties. These food-derived molecules have GRAS (generally regarded as safe) status. One group, collectively known as dietary polyphenols, has anti-inflammatory properties targeting transcription factors such as NFκB, enzymes such as iNOS and COX, as well as acting as antioxidants [84–87]. Examples are resveratrol, epigallocatechin 3-gallate, and curcumin. Another group, also associated with inflammation, is the omega-3 fatty acids. These compounds are discussed here as dietary therapeutics. (See Fig. 2 for their chemical structure.)

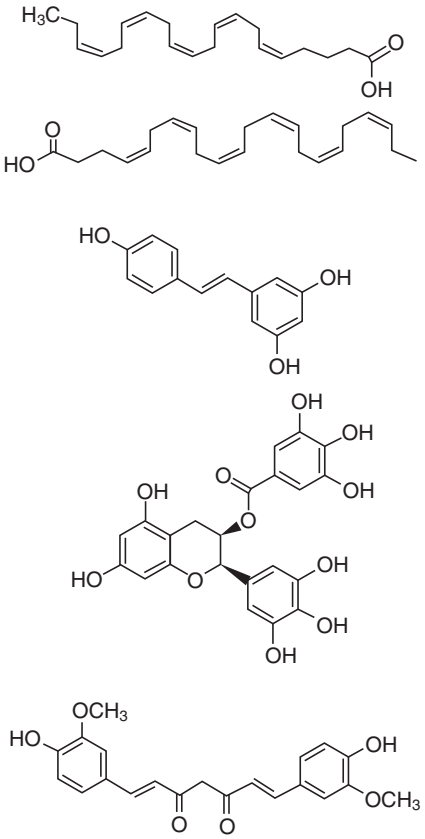


**Figure 2** Selected dietary compounds with anti-inflammatory properties. From top to bottom, the molecules shown are the omega-3 polyunsaturated fatty acids EPA and DHA, followed by the polyphenols resveratrol, EGCG, and curcumin.

### 6.1 Omega-3-Polyunsaturated Fatty Acids

Omega-3-polyunsaturated fatty acids (ω-3 PUFAs), including eicosapentaenoic acid (EPA) (20:5ω-3) and docosahexaenoic acid (DHA) (22:6ω-3), come from oily marine fish in the diet (or fish oil supplements). The typical fish oil may contain 18% EPA and 12% DHA. At doses of milligrams to grams per day, PUFAs show beneficial actions in many diseases, including inflammatory diseases [14, 88, 89]. Possible mechanisms on inflammation relief include competing with arachidonic acid to decrease the production of inflammatory eicosanoids and acting as substrates for the synthesis of the anti-inflammatory lipid mediators such as resolvins and protectins. Although in vitro and animal studies have demonstrated anti-inflammatory activity, the evidence from human trials is somewhat equivocal [90]. The cause may be genetic variation, differences in individual responsiveness to nutrient modulation of inflammation. Another reason may be appropriateness of the selected assay. Condition-specific clinical endpoints may be more sensitive markers of modulation by EPA and DHA than cytokines [91]. For clinical studies, a meta-analysis has concluded that PUFAs have analgesic effects for inflammatory joint pain, and another review of diet and rheumatoid arthritis also concluded that dietary supplementation with PUFAs can provide modest symptomatic benefit in patients with rheumatoid arthritis [92, 93]. Thus, a diet of marine fish may be beneficial to individuals with inflammatory diseases.

### 6.2 Resveratrol

Resveratrol comes from grapes, peanuts, pine nuts, and red wine in the diet. The molecule is known as a phytoalexin, the compound produced by plants against pathogen attack. We have shown that resveratrol is active against human dermal fungal pathogens and may be useful for treating athlete's foot [94]. This polyphenol has multiple health benefits, especially cardioprotection. It acts on the enzyme sirtuin deacetylase and can extend life spans of yeasts, nematodes, fruit flies, and fish [95, 96]. Resveratrol also shows anti-inflammatory activity. It downregulates the inflammatory response by targeting molecules such as COX and NFκB, and this activity may be the basis of its cardioprotective effect [97]. A phenomenon known as the "French paradox" refers to the French who enjoy a relatively low risk of cardiovascular disease despite a diet high in saturated fat; resveratrol and alcohol in red wine may be part of the reason [98]. Thus, a diet including red wine may be beneficial to individuals with cardiovascular and inflammatory diseases.

### 6.3 Epigallocatechin 3-Gallate

Epigallocatechin 3-gallate (EGCG) comes from green tea in the diet. Made from leaves of the plant *Camellia sinensis*, tea is the most popular beverage consumed in the world, second only to water. Approximately 20% of the world's tea is consumed as green tea, and 78% as black tea. EGCG is the major green tea polyphenol, accounting for 30–40% dry weight of the water-extractable material of green tea. It is converted to dimers and polymers during the production of black tea. EGCG has multiple health benefits, including cancer chemoprevention and anti-inflammatory activity. It acts on multiple cellular targets, from enzymes to signaling pathways [99–102]. We have shown its inhibition of iNOS at both gene expression and enzyme activity levels [103], and EGCG acts synergistically with the chemotherapeutic compound cisplatin to kill drug-resistant cancer cells [104]. Others have shown EGCG inhibition of MMPs [105], and green tea polyphenol prevents the onset of murine collagen-induced arthritis [106]. EGCG, like resveratrol, is antimicrobial, and may be useful in preventing dental caries (green tea polyphenol in toothpaste). Thus, a diet including green tea may be beneficial to individuals for cancer prevention and inflammatory diseases.

### 6.4 Curcumin

Curcumin comes from the spice turmeric in the diet. Turmeric is produced from rhizome of the plant *Curcuma longa*. Curcumin is also responsible for the yellow color of curry spice mix. Besides its use as spice, turmeric has been applied in traditional Asian Indian medicine from treating insect bites to wound healing. The polyphenol curcumin is the major compound of turmeric. Multiple and diverse bioactivities have been attributed to curcumin, including anti-inflammatory activity. It acts on targets such as COX, LOX, iNOS, NFκB, AP-1, and many others [107–111]. We have shown its inhibition of NFκB in vitro and its inhibition of iNOS in liver and spleen in a murine sepsis model in vivo [67, 71]. Thus, a diet including curry dishes may be beneficial to individuals for cancer prevention and inflammatory diseases.

As discussed above, omega-3 unsaturated fatty acids, resveratrol, epigallocatechin 3-gallate, and curcumin are dietary ingredients with experimental data demonstrating anti-inflammatory activities. This is an evidence-based approach toward dietary compounds, although a major concern is whether their disease-modifying level is achievable in the body [112]. The effects of diet on inflammation have been analyzed. Specific dietary

patterns, such as the Mediterranean diet, because of anti-inflammatory properties, are known as protective against cardiovascular disease [113]. It has been suggested that western dietary patterns warm up inflammation, while other dietary patterns may cool it down [114]. For specific diseases such as arthritis, a large number of inexpensive natural products that can modulate inflammatory responses, but lack side effects, constitute "gold mines" for disease treatment [115]. Hence we propose the idea of natural anti-inflammatory compounds in foods as dietary therapeutics.

## 7 FUTURE PROSPECTS

The development of anti-inflammatory therapeutics is an active area of pharmacological research. As we look forward to future discoveries, it also pays to look to the past for potentially useful but neglected therapeutic agents.

Older compounds may still be worthwhile for consideration as anti-inflammatories. One example is thalidomide, the drug that had caused a birth defect disaster. Another is cannabinoid, the molecule that has been associated with substance abuse. Alternatively, current drugs may have novel applications as anti-inflammatories. One example is the statins [116]. Some anti-inflammatories may be underutilized, such as the less toxic but slow-acting antimalarials. This may be due to patent expiration that caused the compounds to lose drug company sponsors [6].

Targets for anti-inflammatory therapeutics have been and will be discovered. We give two examples. Suberoylanilide hydroxamic acid (SAHA), the histone deacetylase inhibitor, prevents collagen-induced arthritis in rodents, thus suggesting an innovative therapeutic strategy for rheumatoid arthritis [117]. Polymorphonuclear neutrophils and macrophages produce catecholamines upon stimulation by lipopolysaccharide. As catecholamines enhance inflammatiory injury, discovery of this previously unrecognized source allows development of novel anti-inflammatory strategy (namely $\alpha_2$-adrenoreceptor blockade) [118]. Bringing initial observations to final therapeutic drugs takes judgment, luck, time, and effort, and we look forward to industrial and academic pharmacologists to achieve this goal.

## ACKNOWLEDGMENT

We thank Cara Nasello and Marisa Apicella for helpful comments. The writing of this chapter was supported, in part, by a grant from the National Institutes of Health (5R21AR051761-02) to Marion Chan, and a grant from the Department of Defense (DOD-W911QY-07-C-0007) to Marion Chan and Dunne Fong.

## REFERENCES

1. Nathan C (2002). Points of control in inflammation. *Nature 420*, 846–852.
2. Trowbridge HO, Emling RC (1997). *Inflammation: A Review of the Process*, 5th ed. Chicago: Quintessence.
3. Ryan GB, Majno G (1977). *Inflammation: A Scopes Publication*. Kalamazoo, MI: Upjohn Company.
4. Rainsford KD (2004). *Aspirin and Related Drugs*. London: Taylor & Francis.
5. Brune K, Hinz B (2004). The discovery and development of anti-inflammatory drugs. *Arthritis Rheum 50*, 2391–2399.
6. Whitehouse MW (2005). Drugs to treat inflammation: A historical introduction. *Curr Med Chem 12*, 2931–2942.
7. Brunton LL, Lazo JS, Parker KL, Eds. (2006). *Goodman & Gilman's The Pharmacological Basis of Therapeutics*, 11th ed. New York: McGraw-Hill.
8. Rang HP (2006). *Drug Discovery and Development*. New York: Elsevier.
9. Antman EM, Bennett JS, Daugherty A, Furberg C, Roberts H, Taubert KA (2007). Use of nonsteroidal anti-inflammatory drugs: An update for clinicians: A scientific statement from the American Heart Association. *Circulation 115*, 1634–1642.
10. Gallin JI, Snyderman R, Feron DT, Jaynes BF, Nathan C, Eds. (1999). *Inflammation: Basic Principles and Clinical Correlates*, 3rd ed. Philadelphia, PA: Lippincott.
11. Fujiwara N, Kobayashi K (2005). Macrophages in inflammation. *Curr Drug Targets Inflamm Allergy 4*, 281–286.
12. Kasama T, Miwa Y, Isozaki T, Odai T, Adachi M, Kunkel SL (2005). Neutrophil-derived cytokines: Potential therapeutic targets in inflammation. *Curr Drug Targets Inflamm Allergy 4*, 273–279.
13. Szekanecz Z, Koch AE (2005). Endothelial cells in inflammation and angiogenesis. *Curr Drug Targets Inflamm Allergy 4*, 319–323.
14. Serhan CN (2007). Resolution phase of inflammation: Novel endogenous anti-inflammatory and proresolving lipid mediators and pathways. *Annu Rev Immunol 25*, 101–137.
15. Hume DA, Fairlie DP (2005). Therapeutic targets in inflammatory disease. *Curr Med Chem 12*, 2925–2929.
16. O'Neill LA (2006). Targeting signal transduction as a strategy to treat inflammatory diseases. *Nat Rev Drug Discov 5*, 549–563.
17. Gilroy DW, Lawrence T, Perretti M, Rossi AG (2004). Inflammatory resolution: New opportunities for drug discovery. *Nat Rev Drug Discov 3*, 401–416.
18. Lee MR, Dominguez C (2005). MAP kinase p38 inhibitors: Clinical results and an intimate look at their interac-

tions with p38alpha protein. *Curr Med Chem 12*, 2979–2994.

19. Rückle T, Schwarz MK, Rommel C (2006). PI3Kgamma inhibition: Towards an "aspirin of the 21st century"? *Nat Rev Drug Discov 5*, 903–918.
20. Ghosh S, Eds. (2006). *Handbook of Transcription Factor NF-kappaB*. Boca Raton, FL: CRC Press.
21. Gilmore TD, Herscovitch M (2006). Inhibitors of NF-kappaB signaling: 785 and counting. *Oncogene 25*, 6887–6899.
22. Calzado MA, Bacher S, Schmitz ML (2007). NF-kappaB inhibitors for the treatment of inflammatory diseases and cancer. *Curr Med Chem 14*, 367–376.
23. Funk CD (2001). Prostanglandins and leukotrienes: Advances in eicosanoid biology. *Science 294*, 1871–1875.
24. Khanapure SP, Garvey DS, Janero DR, Letts LG (2007). Eicosanoids in inflammation: Biosynthesis, pharmacology, and therapeutic frontiers. *Curr Top Med Chem 7*, 311–340.
25. Mitchell JA, Warner TD (2006). COX isoforms in the cardiovascular system: Understanding the activities of non-steroidal anti-inflammatory drugs. *Nat Rev Drug Discov 5*, 75–86.
26. FitzGerald GA (2007). COX-2 in play at the AHA and the FDA. *Trends Pharmacol Sci 28*, 303–307.
27. Gajraj NM (2007). COX-2 inhibitors celecoxib and parecoxib: Valuable options for postoperative pain management. *Curr Top Med Chem 7*, 235–249.
28. González-Périz A, Clària J (2007). New approaches to the modulation of the cyclooxygenase-2 and 5-lipoxygenase pathways. *Curr Top Med Chem 7*, 297–309.
29. Leone S, Ottani A, Bertolini A (2007). Dual acting anti-inflammatory drugs. *Curr Top Med Chem 7*, 265–275.
30. Vidal C, Gomez-Hernandez A, Sanchez-Galan E, Gonzalez A, Ortega L, Gomez-Gerique JA, Tunon J, Egido J (2007). Licofelone, a balanced inhibitor of cyclooxygenase and 5-lipoxygenase, reduces inflammation in a rabbit model of atherosclerosis. *J Pharmacol Exp Ther 320*, 108–116.
31. Goossens L, Pommery N, Henichart JP (2007). COX-2/5-LOX dual acting anti-inflammatory drugs in cancer chemotherapy. *Curr Top Med Chem 7*, 283–296.
32. Reid RC (2005). Inhibitors of secretory phospholipase $A_2$ group IIA. *Curr Med Chem 12*, 3011–3026.
33. Fonteh AN, Wykle RL, Eds. (2004). *Arachidonic Remodeling and Inflammation*. Basel: Birkhauser.
34. Fiorucci S (2006). Editorial. *Inflamm Allergy Drug Targets 5*, 79–80.
35. Romano M (2006). Lipid mediators: Lipoxin and aspirin-triggered 15-epi-lipoxins. *Inflamm Allergy Drug Targets 5*, 81–90.
36. Arita M, Oh SF, Chonan T, Hong S, Elangovan S, Sun YP, Uddin J, Petasis NA, Serhan CN (2006). Metabolic inactivation of resolvin E1 and stabilization of its anti-inflammatory actions. *J Biol Chem 281*, 22847–22854.
37. Malmberg AB, Bley KR, Eds. (2005). *Turning up Heat in Pain: TRPV1 Receptors in Pain and Inflammation*. Basel: Birkhauser.
38. Blakeney JS, Fairlie DP (2005). Nonpeptide ligands that target peptide-activated GPCRs in inflammation. *Curr Med Chem 12*, 3027–3042.
39. Cato AC, Schacke H, Sterry W, Asadullah K (2004). The glucocorticoid receptor as target for classic and novel anti-inflammatory therapy. *Curr Drug Targets Inflamm Allergy 3*, 347–353.
40. Perretti M, D'Acquisto F (2006). Novel aspects of annexin 1 and glucocorticoid biology: Intersection with nitric oxide and the lipoxin receptor. *Inflamm Allergy Drug Targets 5*, 107–114.
41. Kostadinova R, Wahli W, Michalik L (2005). PPARs in diseases: Control mechanisms of inflammation. *Curr Med Chem 12*, 2995–3009.
42. Liew FY, Xu D, Brint EK, O'Neill LA (2005). Negative regulation of toll-like receptor-mediated immune responses. *Nat Rev Immunol 5*, 446–458.
43. O'Neill LA, Brint E, Eds. (2005). *Toll-Like Receptors in Inflammation*. Basel: Birkhauser.
44. Higgs GA, Henderson B, Eds. (2000). *Novel Cytokine Inhibitors*. Basel: Birkhauser.
45. Breit SN, Wahl SM, Eds. (2001). *TGF-β and Related Cytokines in Inflammation*. Basel: Birkhauser.
46. Taams LS, Akbar AN, Wauben MHM, Eds. (2005). *Regulatory T Cells in Inflammation*. Basel: Birkhauser.
47. Day RO, Furst DE, van Riel PLCM, Bresnihan B, Eds. (2005). *Antirheumatic Therapy: Actions and Outcomes*. Basel: Birkhauser.
48. Weinberg JM, Buchholz R, Eds. (2006). *TNF-alpha Inhibitors*, *Milestones in Drug Therapy Series*. Basel: Birkhauser.
49. Brown MA (2005). Antibody treatments of inflammatory arthritis. *Curr Med Chem 12*, 2943–2946.
50. Le GT, Abbenante G (2005). Inhibitors of TACE and caspase-1 as anti-inflammatory drugs. *Curr Med Chem 12*, 2963–2977.
51. Stehlik C, Dorfleutner A (2007). COPs and POPs: Modulators of inflammasome activity. *J Immunol 179*, 7993–7998.
52. Hu J, Van den Steen PE, Sang QX, Opdenakker G (2007). Matrix metalloproteinase inhibitors as therapy for inflammatory and vascular diseases. *Nat Rev Drug Discov 6*, 480–498.
53. Winyard PG, Blake DR, Evans CH, Eds. (2000). *Free Radicals and Inflammation*. Basel: Burkhauser.
54. Salvemini D, Billiar TR, Vodovotz Y, Eds. (2001). *Nitric Oxide and Inflammation*. Basel: Birkhauser.
55. Hobbs AJ, Moncada S (1999). Inducible nitric oxide synthase and inflammation. In Willoughby DA, Tomlinson A, Eds. *Inducible Enzymes in the Inflammatory Response*. Basel: Birkhauser, pp. 31–54.
56. Cirino G, Distrutti E, Wallace JL (2006). Nitric oxide and inflammation. *Inflamm Allergy Drug Targets 5*, 115–119.

57. Kanwar JR (2005). Anti-inflammatory immunotherapy for multiple sclerosis/experimental autoimmune encephalomyelitis (EAE) disease. *Curr Med Chem 12*, 2947–2962.
58. Moser B, Letts GL, Neote K, Eds. (2006). *Chemokine Biology—Basic Research and Clinical Application*, Vol. I: Immunobiology of Chemokines. Basel: Birkhauser.
59. Van Eden W, Ed. (2003). *Heat Shock Proteins and Inflammation*. Basel: Birkhauser.
60. Chen Y, Voegeli TS, Liu PP, Noble EG, Currie RW (2007). Heat shock paradox and a new role of heat shock proteins and their receptors as anti-inflammation targets. *Inflamm Allergy Drug Targets 6*, 91–100.
61. Haas PJ, van Strijp J (2007). Anaphylatoxins: Their role in bacterial infection and inflammation. *Immunol Res 37*, 161–175.
62. McCulloch CA, Downey GP, El-Gabalawy H (2006). Signalling platforms that modulate the inflammatory response: New targets for drug development. *Nat Rev Drug Discov 5*, 864–876.
63. Yoshimura A, Naka T, Kubo M (2007). SOCS proteins, cytokine signalling and immune regulation. *Nat Rev Immunol 7*, 454–465.
64. Vukicevic S, Sampath KT, Eds. (2002). *Bone Morphogenetic Proteins*. Basel: Birkhauser.
65. Zukowska Z, Feuerstein GZ, Eds. (2005). *The NYT Family of Peptides in Immune Disorder, Inflammation, Angiogenesis and Cancer*. Basel: Birkhauser.
66. Cantor JO, Nadkarni PP (2007). Hyaluronan: The Jekyll and Hyde molecule. *Inflamm Allergy Drug Targets 5*, 257–260.
67. Chan MM (1995). Inhibition of tumor necrosis factor by curcumin, a phytochemical. *Biochem Pharmacol 49*, 1551–1556.
68. Winyard PG, Willoughby DA, Eds. (2003). *Inflammation Protocols*. Totowa, NJ: Humana.
69. Chan MM, Ho CT, Huang HI (1995). Effects of three dietary phytochemicals from tea, rosemary and turmeric on inflammation-induced nitrite production. *Cancer Lett 96*, 23–29.
70. Stevenson CS, Marshall LA, Morgan DW, Eds. (2006). *In Vivo Models of Inflammation*, Vol. I, 2nd ed. Basel: Birkhauser.
71. Chan MM, Huang HI, Fenton MR, Fong D (1998). In vivo inhibition of nitric oxide synthase gene expression by curcumin, a cancer preventive natural product with anti-inflammatory properties. *Biochem Pharmacol 55*, 1955–1962.
72. Stevenson CS, Marshall LA, Morgan DW, Eds. (2006). *In Vivo Models of Inflammation*, Vol. II, 2nd ed. Basel: Birkhauser.
73. Lemay M (2006). Anti-inflammatory phytochemicals: In vitro and ex vivo evaluation. In Meskin MS, Bidlack WR, Randolph RK, Eds. *Phytochemicals: Nutrient-Gene Interactions*. Boca Raton, FL: Taylor & Francis, pp. 41–60.
74. Fiorucci S, Antonelli E (2006). NO-NSAIDs: From inflammatory mediators to clinical readouts. *Inflamm Allergy Drug Targets 5*, 121–131.
75. Young DV, Cochran ED, Dhawan V, Earl RA, Ellis JL, Garvey DS, Janero DR, Khanapure SP, Letts LG, Melim TL, Murty MG, Shumway MJ, Wey SJ, Zemtseva IS, Selig WM (2005). A comparison of the cyclooxygenase inhibitor-NO donors (CINOD), NMI-1182 and AZD3582, using in vitro biochemical and pharmacological methods. *Biochem Pharmacol 70*, 1343–1351.
76. Coruzzi G, Menozzi A, Dobrilla G (2004). Novel non-steroidal anti-inflammatory drugs: What we have learned from animal studies. *Curr Drug Targets Inflamm Allergy 3*, 43–61.
77. Fiorucci S, Santucci L, Gresele P, Faccino RM, del Soldato P, Morelli A (2003). Gastrointestinal safety of NO-aspirin (NCX-4016) in healthy human volunteers: A proof of concept endoscopic study. *Gastroenterology 124*, 600–607.
78. Fiorucci S, Santucci L, Wallace JL, Sardina M, Romano M, del Soldato P, Morelli A (2003). Interaction of a selective cyclooxygenase-2 inhibitor with aspirin and NO-releasing aspirin in the human gastric mucosa. *Proc Natl Acad Sci USA 100*, 10937–10941.
79. Wallace JL (2007). Hydrogen sulfide-releasing anti-inflammatory drugs. *Trends Pharmacol Sci 28*, 501–505.
80. Koehn FE, Carter GT (2005). The evolving role of natural products in drug discovery. *Nat Rev Drug Discov 4*, 206–220.
81. Darshan S, Doreswamy R (2004). Patented antiinflammatory plant drug development from traditional medicine. *Phytother Res 18*, 343–357.
82. Chrubasik JE, Roufogalis BD, Chrubasik S (2007). Evidence of effectiveness of herbal antiinflammatory drugs in the treatment of painful osteoarthritis and chronic low back pain. *Phytother Res 21*, 675–683.
83. Brinker AM, Ma J, Lipsky PE, Raskin I (2007). Medicinal chemistry and pharmacology of genus *Tripterygium* (Celastraceae). *Phytochemistry 68*, 732–766.
84. Shay NF, Banz WJ (2005). Regulation of gene transcription by botanicals: Novel regulatory mechanisms. *Annu Rev Nutr 25*, 297–315.
85. Surh YJ, Kundu JK, Na HK, Lee JS (2005). Redox-sensitive transcription factors as prime targets for chemoprevention with anti-inflammatory and antioxidative phytochemicals. *J Nutr 135*, 2993S–3001S.
86. Yoon JH, Baek SJ (2005). Molecular targets of dietary polyphenols with anti-inflammatory properties. *Yonsei Med J 46*, 585–596.
87. Rahman I, Biswas SK, Kirkham PA (2006). Regulation of inflammation and redox signaling by dietary polyphenols. *Biochem Pharmacol 72*, 1439–1452.
88. Nettleton JA (1995). *Omega-3 Fatty Acids and Health*. New York: Chapman & Hall.
89. Calder PC (2006). n-3 polyunsaturated fatty acids, inflammation, and inflammatory diseases. *Am J Clin Nutr 83*, 1505S–1519S.

90. Fritsche K (2006). Fatty acids as modulators of the immune response. *Annu Rev Nutr 26*, 45–73.
91. Sijben JW, Calder PC (2007). Differential immunomodulation with long-chain n-3 PUFA in health and chronic disease. *Proc Nutr Soc 66*, 237–259.
92. Goldberg RJ, Katz J (2007). A meta-analysis of the analgesic effects of omega-3 polyunsaturated fatty acid supplementation for inflammatory joint pain. *Pain 129*, 210–223.
93. Stamp LK, James MJ, Cleland LG (2005). Diet and rheumatoid arthritis: A review of the literature. *Semin Arthritis Rheum 35*, 77–94.
94. Chan MM (2002). Antimicrobial effect of resveratrol on dermatophytes and bacterial pathogens of the skin. *Biochem Pharmacol 63*, 99–104.
95. Baur JA, Sinclair DA (2006). Therapeutic potential of resveratrol: The in vivo evidence. *Nat Rev Drug Discov 5*, 493–506.
96. Aggarwal BB, Shishodia S, Eds. (2006). *Resveratrol in Health and Disease*. Boca Raton, FL: Taylor & Francis.
97. Das S, Das DK (2007). Anti-inflammatory responses of resveratrol. *Inflamm Allergy Drug Targets 6*, 168–173.
98. Watkins TR (1997). *Wine: Nutritional and Therapeutic Benefits*. ACS Symposium Series, Vol. 661. Washington, DC: American Chemical Society.
99. Hara Y (2001). *Green Tea: Health Effects and Applications*. New York: Marcel Dekker.
100. Zaveri NT (2006). Green tea and its polyphenolic catechins: Medicinal uses in cancer and noncancer applications. *Life Sci 78*, 2073–2080.
101. Khan N, Afaq F, Saleem M, Ahmad N, Mukhtar H (2006). Targeting multiple signaling pathways by green tea polyphenol (−)-epigallocatechin-3-gallate. *Cancer Res 66*, 2500–2505.
102. Yang CS, Sang S, Lambert JD, Hou Z, Ju J, Lu G (2006). Possible mechanisms of the cancer-preventive activities of green tea. *Mol Nutr Food Res 50*, 170–175.
103. Chan MM, Fong D, Ho CT, Huang HI (1997). Inhibition of nitric oxide synthase gene expression and enzyme activity by epigalloccatechin gallate, a natural product from green tea. *Biochem Pharmacol 54*, 1281–1286.
104. Chan MM, Soprano KJ, Weinstein K, Fong D (2006). Epigallocatechin-3-gallate delivers hydrogen peroxide to induce death of ovarian cancer cells and enhances their cisplatin susceptibility. *J Cell Physiol 207*, 389–396.
105. Dell'Aica I, Caniato R, Biggin S, Garbisa S (2007). Matrix proteases, green tea, and St. John's wort: Biomedical research catches up with folk medicine. *Clin Chim Acta 381*, 69–77.
106. Haqqi TM, Anthony DD, Gupta S, Ahmad N, Lee MS, Kumar GK, Mukhtar H (1999). Prevention of collagen-induced arthritis in mice by a polyphenolic fraction from green tea. *Proc Natl Acad Sci USA 96*, 4524–4529.
107. Sharma RA, Gescher AJ, Steward WP (2006). Curcumin: The story so far. *Eur J Cancer 41*, 1955–1968.
108. Duvoix A, Blasius R, Delhalle S, Schnekenburger M, Morceau F, Henry E, Dicato M, Diederich M (2005). Chemopreventive and therapeutic effects of curcumin. *Cancer Lett 223*, 181–190.
109. Maheshwari RK, Singh AK, Gaddipati J, Srimal RC (2006). Multiple biological activities of curcumin: A short review. *Life Sci 78*, 2081–2087.
110. Salvioli S, Sikora E, Cooper EL, Franceschi C (2007). Curcumin in cell death processes: A challenge for CAM of age-related pathologies. *Evid Based Complement Alternat Med 4*, 181–190.
111. Aggarwal BB, Surh YJ, Shishodia S, Eds. (2007). *The Molecular Targets and Therapeutic Uses of Curcumin in Health and Diseases*. Advances in Experimental Medicine and Biology, Vol. 595. New York: Springer.
112. Higdon J (2007). *An Evidence-Based Approach to Dietary Phytochemicals*. New York: Thieme Medical Publishers.
113. Kontogianni MD, Zampelas A, Tsigos C (2006). Nutrition and inflammatory load. *Ann NY Acad Sci 1083*, 214–238.
114. Giugliano D, Ceriello A, Esposito K (2006). The effects of diet on inflammation: Emphasis on the metabolic syndrome. *J Am Coll Cardiol 48*, 677–685.
115. Khanna D, Sethi G, Ahn KS, Pandey MK, Kunnumakkara AB, Sung B, Aggarwal A, Aggarwal BB (2007). Natural products as a gold mine for arthritis treatment. *Curr Opin Pharmacol 7*, 344–351.
116. Ghittoni R, Lazzerini PE, Pasini FL, Baldari CT (2007). T lymphocytes as targets of statins: Molecular mechanisms and therapeutic perspectives. *Inflamm Allergy Drug Targets 6*, 3–16.
117. Lin HS, Hu CY, Chan HY, Liew YY, Huang HP, Lepescheux L, Bastianelli E, Baron R, Rawadi G, Clément-Lacroix P (2007). Anti-rheumatic activities of histone deacetylase (HDAC) inhibitors in vivo in collagen-induced arthritis in rodents. *Br J Pharmacol 150*, 862–872.
118. Flierl MA, Rittirsch D, Nadeau BA, Chen AJ, Sarma JV, Zetoune FS, McGuire SR, List RP, Day DE, Hoesel LM, Gao H, Van Rooijen N, Huber-Lang MS, Neubig RR, Ward PA (2007). Phagocyte-derived catecholamines enhance acute inflammatory injury. *Nature 449*, 721–725.

# ANTI-INFLAMMATORY/ANALGESICS: AN IN-DEPTH LOOK AT P38 MAP KINASES

Kelly L. Brown
*Department of Rheumatology and Inflammation Research, Göteborg University, Göteborg, Sweden*

## 1 INTRODUCTION

### 1.1 Inflammatory Response and Inflammatory Pain

Inflammation in mammals is a complex biological response to the endogenous or exogenous products of injury or infection (danger signals), for example, pathogens, damaged cells, and inflammatory mediators. Cells in the immune system are equipped with receptors for these stimuli; notable examples are the interleukin-1 (IL-1), tumor necrosis factor-α (TNF-α), and Toll-like receptors. The cells in the innate branch of the immune response such as neutrophils, epithelial cells, and monocytes/macrophages are the first responders to such danger signals. Receptor ligation on these cells induces an acute cellular response that is characterized by the expression of a repertoire of early-response genes including genes that encode for pro-inflammatory cytokines (TNF-α, IL-6), chemokines (IL-8, Gro-α/CXCL1) and antibacterial agents (antimicrobial peptides, reactive oxygen/nitrogen species). The mediators that are produced from early-response gene transcription will prompt additional cellular and physiological events that collectively facilitate eradication of the irritant that initiated the response [1]. In doing so, however, the inflammatory mediators inescabably cause collateral damage to host tissues and can stimulate inflammatory and neuropathic pain [2]. Inflammatory pain is associated with swelling and damage in the inflected tissue via the natural influx of immune cells and subsequent release of proteases and toxic agents. Neuropathic pain results from the activation and sensitization of primary sensory nociceptors in the peripheral nervous system (PNS) and/or via dorsal horn and brain neurons in the central nervous system (CNS). Mediators of both peripheral and central sensitization to pain include prostaglandins, nitric oxide (NO), and cytokines (TNF-α and IL-1β). To minimize the negative effects that can be incurred during an inflammatory response, the process is designed to be inherently self-limiting. Receptor ligation inevitably induces an active anti-inflammatory response through the production of anti-inflammatory mediators [e.g., IL-10, transforming growth factor-β (TGFβ)] and the expression and/or activation of cellular proteins that calm the cellular response to stimuli. This culminates in physiological processes such as wound healing and a reversion to basal levels of cellular activity in which pro-inflammatory responses are constrained by negative regulation. As both inflammatory and neuropathic pain are linked to the inflammatory episode, sensations of pain should cease with the resolution of the inflammatory response.

The coordinated induction and resolution of the inflammatory response is a complex system that depends on the balanced expression and management of numerous positive and negative regulatory molecules. The family of mitogen-activated protein (MAP) kinases are key signaling molecules broadly expressed in cells and tissues of the immune system. MAP kinases profoundly influence many aspects of cellular physiology via regulation of gene transcription, protein synthesis, cell-cycle

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

regulation, cell death, and differentiation [1]. It is well established that MAP kinases along with the transcription factor nuclear factor κ B (NFκB) promote and coordinate the expression of inflammatory mediators. MAP kinases have also been implicated in the regulation of inflammatory and neuropathic pain. Signals mediated by MAP kinases and NFκB are under strict control to ensure that any activation is both timely and appropriate in magnitude and duration. A collapse of the regulatory measures either in the presence or absence of an obvious initiating stimulus can lead to some of the most serious threats to human health, including long-term pain, chronic inflammatory diseases, and endotoxic shock [3]. Millions of people worldwide suffer from chronic inflammatory diseases and pain. More than 2 million people in the United States alone suffer from the inflammatory disease rheumatoid arthritis (RA). This disease is characterized by pain, swelling, and stiffness in the joints that is driven by the overproduction of inflammatory mediators including those that are influenced by MAP kinases (MAPK).

Steroidal and nonsteroidal anti-inflammatory drugs (NSAIDs) are the most efficacious treatments currently available for inflammation and inflammatory pain. NSAIDs inhibit cycloxygenase (COX) to prevent the production of prostaglandins that mediate inflammatory and anaphylactic reactions. Steroidal drugs such as corticosteroids and glucocorticoids inhibit the transcription of inflammatory genes. The mechanism of their action is not fully understood, but evidence suggests that glucocorticoid receptor-mediated signaling can suppress MAPK activity, thereby inhibiting gene transcription [4–6]. While these treatments may be highly effective, they are also expensive, and, of greater concern, they can cause dire side effects. The advent of COX-2-selective NSAIDs have reduced but not eliminated the significant risk of mortality associated with earlier NSAID compounds. Immune surveillance and cell-mediated immunity can be significantly compromised in patients receiving glucocorticoid thereapy as critical cytokines that drive B- and T-cell proliferation, lymphocyte function, and $F_c$ receptor expression are not transcribed. Glucocorticoids have also been known to cause adrenal insufficiency, osteoporosis, inhibited growth, decreased skin resiliency, and hyperglycemia. Thus there is a considerable need to develop new anti-inflammatory and antianalgesic therapies that are efficacious, selective, and economically viable.

### 1.2 Protein Kinases

Protein kinases are enzymes that catalyze the transfer of phosphate from adenosine triphosphate (ATP) to a hydroxyl side chain on serine, threonine, or tyrosine residues within a protein. The phosphorylation reaction is reversed by enzymes in the corresponding family of protein phosphatases. Together kinases and phosphatases regulate the phosphorylation state of cellular proteins that can control various aspects of protein structure and function, for example, cellular localization, interaction partners, stability, and enzymatic activity. It has been estimated that phosphoproteins comprise one-third of all cellular proteins. Protein phosphorylation is a common aspect of cellular signal transduction that is integral to the maintenance of normal cell physiology and responsiveness to environmental (extracellular) cues. The transmission of an extracellular signal via a protein phosphorylation network often evokes gene transcription and a cellular response to the original input.

The introduction of deoxyribonucleic acid (DNA) cloning and sequencing in the 1970s and the discovery that protein kinases have a high degree of sequence similarity in the catalytic domain enabled the identification of nearly 100 protein kinases by the mid-1980s [7]. At that time, it was speculated that the human genome was comprised of 50,000–100,000 genes, and Hunter et al. predicted that as many as 1000 of these genes, or 1–2% of the entire genome, coded for protein kinases. The first published sequence of the entire human genome came a mere 15 year later [8], which facilitated a genetic analysis of gene families involved in cellular phosphorylation events and the construction of the human "kinome," a compilation of the full complement of protein kinase genes [9]. Consistent with the original prediction, 1.7% of the human genome (518 genes) codes for protein kinases, making this one of the largest families of genes [9].

### 1.3 MAP Kinases

The family of MAP kinases (EC 2.7.11.24) are proline-directed serine/threonine-specific kinases. MAPKs are part of a phosphorylation signaling cascade that is activated in response to extracellular stimuli, primarily cytokines and environmental stress. Central to classic MAPK signaling cascades is a three-tiered hierarchical module of serine/threonine and dual-specificity protein kinases (Fig. 1). MAP kinase kinase kinases (MKKK or MAP3K) form the first tier in the MAPK cascade. Stimuli-induced MAP3K phosphorylate serine and threonine residues on downstream dual-specificity MAP kinase kinases (MKK, MAP2K or MEK). In turn, the phosphorylated MAP2Ks activate MAPK through phosphorylation on threonine and tyrosine residues. The family of MAP kinases and the MAPK phosphorylation cascades have been well conserved in evolution. The mammalian MAPK family can be subdivided into

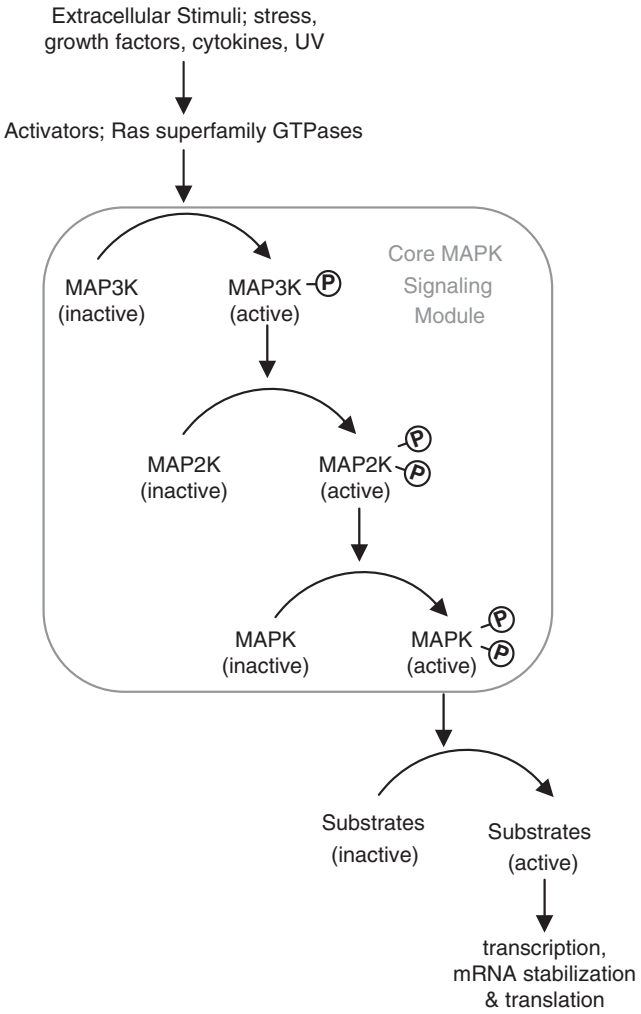


**Figure 1** Generic MAPK signal transduction cascade.

distinct groups, four of which have been extensively characterized: extracellular signal-regulated kinase 1, 2 and 5 (ERK1, ERK2, ERK5), c-Jun N-terminal kinase (JNK), and p38. The MAP kinases ERK3, ERK4, ERK7, and ERK8 are structurally related, atypical MAP kinases with distinct regulation and function compared to the other ERK family members. The ERK and JNK MAP kinases are described briefly below as well as in Table 1 and Figure 2. The reader is encouraged to refer to several recent reviews on MAP kinases for a more comprehensive discussion [10–12].

***ERK1/2*** In 1991, extracellular signal-regulated kinase-1 and -2 (ERK1, ERK2) were the first MAPK family members discovered; thus ERK1/2 are considered the prototypic MAP kinases [13]. ERK1 and ERK2 are 44 and 42 kDa, respectively, with 90% sequence identity. ERK1/2 are preferentially activated by growth factors and phorbol esters (tumor promoters), regulate cell proliferation and differentiation, and are therapeutic targets in cancer [14].

***ERK5*** While this kinase shares ~66% sequence identity with ERK1/2, it is significantly larger (98 kDa) [15]. Fittingly, it was also ascribed the name Big MAP kinase-1 (BMK1). Three isoforms of the protein exist, two of which may act in a dominant-negative fashion. With the exception of the liver, ERK5 is ubiquitously expressed and is responsive to growth factors, oxidative stress, and mitogenic stimuli [16, 17].

***ERK3/4*** ERK1/2 and ERK3/4 differ in the C-terminus (which is longer in ERK3/4) and in the highly conserved activation loop that is found between kinase subdomains VII and VIII. ERK3/4 are the only MAPK family members to contain a Ser-Glu-Gly motif instead of the characteristic Thr-Glu-Tyr motif, which is dually phosphorylated on the threonine and tyrosine residues. Consequently, ERK3/4 are not substrates for the same kinases that phosphorylate and regulate ERK1/2 [17, 18].

***ERK7/8*** ERK7 and ERK8 share approximately 82% amino acid identity in the kinase domain and 40% homology with ERK1/2. Like ERK3/4, ERK7/8 differ from ERK1/2 in the C-terminus and do not appear to be substrates of typical ERK1/2-activating kinases [17].

***JNK/SAPK*** Stress-activated protein kinases (SAPKs), also known as c-Jun N-terminal kinases (JNK), are responsive to stress stimuli such as cytokines, ultraviolet (UV) irradiation, heat shock, and osmotic shock. JNK1 and JNK2 are ubiquitously expressed while JNK3 expression is limited to the brain, heart, and testis. Differential splicing within the catalytic domain and C-terminus of JNK1-3 can create at least 10 isoforms of JNK. Biological roles for JNK have been linked to cellular apoptosis, cellular differentiation, and embryogenesis [19, 20].

The genuine in vivo function(s) of many MAP kinases in cellular and biological processes and disease pathogenesis remain elusive [10]. Difficulties arise in part from functional redundancy within the MAPK family and a high degree of interconnectivity with other signal transduction cascades. Mice deficient in a single MAPK are often either phenotypically indistinguishable from wild-type mice (e.g., ERK2, JNK3, and p38β,δ,γ isoforms) or not viable (e.g., ERK2, ERK5, and p38α isoform). While this outcome established a role for certain MAP kinases in development [21], it prevented an analysis of the role(s) of these kinases in mature animals [22].

**TABLE 1 Quick Reference Guide for the MAPK Family**

| MAPK → | ERK1/2 | SAPK/JNK | P38 | ERK5 | ERK3/4 | ERK7/8 |
|---|---|---|---|---|---|---|
| ***Alternate names*** | ERK1; p44, MAPK3<br>ERK2; p42, MAPK1 | JNK1; SAPK1, SAPKγ, MAPK8<br>JNK2; p54aSAPK, SAPKα, MAPK9<br>JNK3; p54bSAPK, SAPKβ, MAPK10 | P38α; MAPK14, CSBP, RK, SAPK2a<br>P38β; MAPK11, SAPK2b<br>P38γ; MAPK12, SAPK3, ERK6<br>P38δ; MAPK13, SAPK4 | MAPK7, BMK1 | ERK3; MAPK6<br>ERK; MAPK4 | ERK8; MAPK15 |
| ***Activation stimuli*** | *Mitogens and proliferative signals*<br>Growth factors and phorbol esters, e.g., EGF, FGF, PDGF, NGF, BDNF, NT3/4, LPS, CSF | *Environmental stress*<br>Cytokines; TNF-α, IL-1, TGFβ<br>UV irradiation<br>Heat shock, osmotic shock<br>FasL, LPS | | Growth factors (EGF), stress stimuli, serum, ROS | ERK3; constitutively active | ERK8; serum<br>ERK7; constitutively active |
| ***Location*** | Ubiquitous | JNK1/2; ubiquitous<br>JNK3: brain, heart, testis | P38α/β; ubiquitous<br>P38γ; skeletal<br>P38δ; testis, pancreas, small intestine, and $CD4^+$ T cells | Ubiquitous | Highly expressed in brain, lung, and kidney | nd[a] |
| ***Upstream kinase (MEKK)*** | A-Raf, Raf-1, B-Raf, c-Mos | TAK1, ASK1, MLK2/3, MEKK1-4, DLK, Tlp2/Cot, TAO-1/2 | | MEKK2, MEKK3 | nd | ERK8; Src |
| ***Activating kinase (MEK)*** | MKK1/MEK1, MKK2/MEK2 | MKK4/MEK4, MKK7/MEK7 | MKK3/MEK3, MKK6/MEK6, MKK4/MEK4 | MKK5/MEK5 | MKK5/MEK5 | Unknown |
| ***Substrates*** | Tau, STMN1, cPLA2, MSK1/2, RSK2, CREB, Elk-1, SAP1a, c-myc, SRF, c-fos, NFIL6, MNK1/2 | ATF-2, Elk1, p53, c-JUN, JunD, C/EBP, MSK2<br>NFAT-2, -4 (negative regulation) | ATF2, Elk1, p53, SAP1A, GADD153, Max, MEF2C, PRAK, MSK1/2, Cdc25b CREB, hsp27, MNK1/2, MAPKAPK2 | Nu77, CGK-1 | PRAK | ERK7; c-fos c-myc<br>ERK8; Hic5/ARA55 |
| ***Regulated functions*** | Proliferation and differentiation | Proliferation, differentiation, transcription regulation, apoptosis, inflammation, and development | | Cardiac & vascular development | Embryonic and postnatal development | nd |
| ***KO phenotype*** | ERK1; possible defect in T-cell development<br>ERK2: not viable | JNK1; altered Th2 differentiation<br>JNK2; altered Th1 differentiation<br>JNK3; no phenotype | P38α; not viable<br>P38β, γ and δ; no phenotype in single KO | Not viable | Not viable | nd |

[a]nd; not determined.

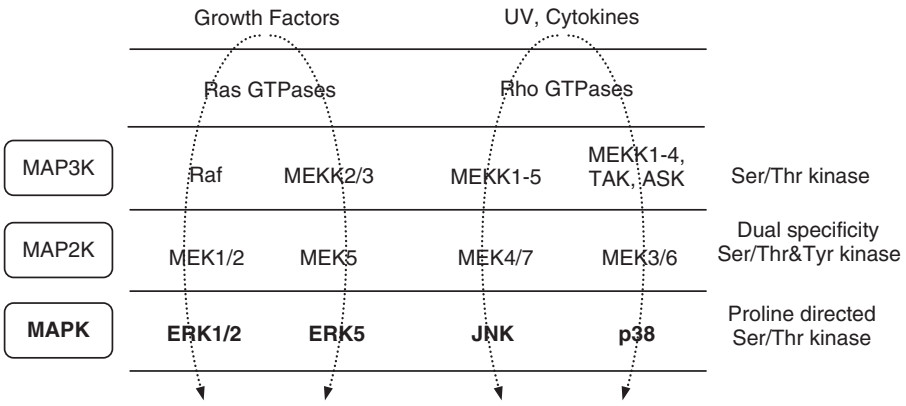


**Figure 2** Family of MAP kinases.

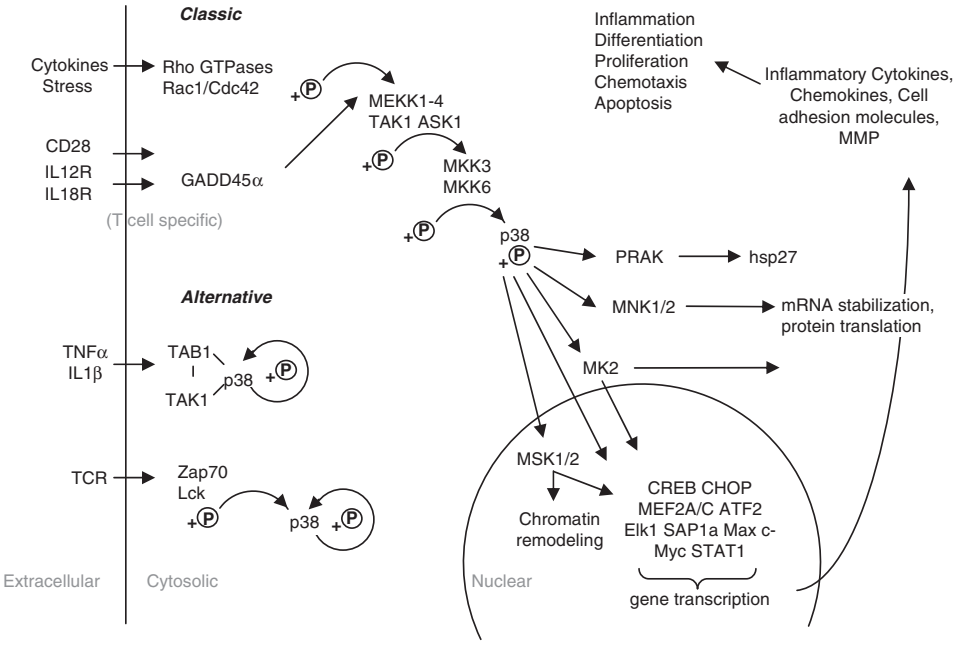


**Figure 3** Activation of p38 MAPK signal transduction pathways.

## 2 P38 MAP KINASE FAMILY

### 2.1 Discovery

The p38 MAPK family is highly conserved throughout evolution with mammalian homologs existing in yeast (Hog1 and Spc/Sty1), worm (pmk1-3), fruit fly (D-p38a/Mpk2, D-p38b), and frog (Mapk14a) [23–25]. Like JNK, p38 MAP kinases respond to stress stimuli and are implicated in cell survival, growth, differentiation, and apoptosis [14, 20]. Four isoforms of p38 (p38α, p38β, p38γ, and p38δ) have been discovered in mammals, the first of which, p38α, was described independently by several investigators in 1994 in the context of cellular responses to stress. Lee and colleagues at SmithKline & French discovered p38α as the molecular target responsible for the inhibition of TNF-α production in lipopolysaccharide (LPS)-stimulated monocytes treated with a cytokine suppressant anti-inflammatory drug (CSAID). Aptly, p38 was named a CSAID-binding protein (CSBP) [26]. Han et al. described a 38-kDa tyrosine phosphorylated polypeptide in human monocytes subjected to endotoxin and hyperosmolarity shock [27, 28]. Finally, two laboratories independently discovered that p38 [named p40 and reactivating kinase (RK)] regulated MAPK-activated protein kinase-2 (MAPKAPK2) and, in turn, heat shock protein 27 (hsp27) in response to stress [29–32]. The isoforms p38β, p38γ, and p38δ were

identified a few years after the discovery of p38α through expressed sequence tag (EST) database analysis and homology screening [33–38].

Two splice variants of p38α have also been described. The potential splice variant Mxi2 was identified by association with the transcription factor max and the C-terminus of c-myc in a yeast-2-hybrid screen. While the ATP-binding pocket and activation loop are conserved, Mxi2 lacks 65 amino acids at the C-terminus, is less sensitive to inhibition by CSAID, and does not have activators and substrates in common with p38α [39, 40]. Exon-Skp (Exip) binds to the IL-1R/Toll-like receptor (TLR)-associated signaling molecules Tollip (toll interacting protein) and IRAK (interleukin-1 receptor-associated kinase), and overexpression of Exip marginally impaired IL-1β-induced activation of NFκB [41, 42]. The functional relevance of these splice variants has not been elucidated.

### 2.2 Distribution

The p38α, β, and δ isoforms are widely expressed, although not to equivalent levels, whereas the expression of p38γ is restricted to skeletal muscle. p38α is highly expressed in human lymphoid tissues (spleen, lymph node) and in peripheral leukocytes, most prominently in monocytes, macrophages, $CD4^+$ T cells, neutrophils, and endothelial cells, the latter being the only cell type in which the expression of p38β is comparable to that of p38α [43]. p38δ is found in the salivary and adrenal glands, testis, pancreas, and small intestine as well as in cells of the immune system, namely $CD4^+$ T cells, macrophages, and neutrophils where it is expressed at levels comparable to p38α. In resting cells p38 is predominantly localized in the cytoplasm. Following cellular activation by stress or inflammatory cytokines, p38 may transit to the nucleus, although the evidence for this is often contradictory and may vary with cell type and state of cellular maturation or activation.

### 2.3 p38 MAPK Signal Transduction

***Classic Signaling Cascade*** The p38 MAPK signaling cascade is strongly activated in vivo in a variety of immune cells in response to a battery of environmental stresses and inflammatory agents (and less so by serum and growth factors that are known to activate ERK MAPK). Human neutrophils alone activate p38 in response to formyl-met-leu-phe (fMLP), granulocyte-macrophage–colony-stimulating factor (GM-CSF), phorbol 12-myristate 13-acetate (PMA), TNF-α, LPS, TGF-β, or ionomycin [44–47]. Stress-associated stimuli commonly trigger the intracellular Rho family of guanosine triphosphate (GTP)-binding proteins, namely Rac1, Cdc42, and Rho [48–51]. The mechanism by which the Rho proteins activate MAP3K has not been entirely elucidated, although it may involve a direct binding interaction that leads to (i) membrane recruitment of MAP3K by upstream activating proteins, (ii) homoligomerization with multiprotein complexes of regulatory components, and (iii) phosphorylation of MAP3K [46, 48–52].

MAP3K are promiscuous, diverse kinases with highly conserved catalytic domains that can interact with a number of regulatory proteins and catalyze the activation of multiple pathways (Fig. 2 and Fig. 3). MAP3K in the p38 cascade include TAK1, ASK1, Tpl-2/Cot, MLK2/3, and MEKK1-4. The abundance and usage of various MAP3K may vary with cell type and reflect particular extracellular stimuli. MAP3K family members activated in the p38 pathway show preference for phosphorylation of the dual specificity MAP2-kinases MKK3 (MEK3), MKK6 (MEK6), and to a lesser extent MKK4 (MEK4), which is usually upstream of JNK but can activate p38 (α and δ isoforms) in response to UV [38]. MKK6 is phosphorylated (Ser207 and Thr 211) in cells exposed to all p38-pathway stimuli, whereas MKK3 appears to be activated (phosphorylation at Ser189 and Thr193) specifically in cells under chemical and physical stress. This may reflect the aforementioned stimuli-specific activation of MAP3K. Consistent with their promiscuous nature, MAP3-kinases have activity toward multiple substrates including those outside the MAPK pathway. For example, MEKK1-3, Tpl-2/Cot, and TAK-1 phosphorylate IKK, an event that releases the transcription factor NFκB from negative regulation [53–55]. MAP2-kinases are also substrates for multiple MAP3Ks. These top two tiers of the cascade are equipped to receive and coordinate responses to multiple divergent inputs and activate multiple, possibly parallel, signaling cascades of both an MAPK-dependent and MAPK-independent nature. While such a system imparts a great deal of biological latitude, elucidation of this part of the pathway is a laborious task, and accurate illustration requires contextual information regarding specific cell types and stimuli.

Finally, MAP3K-activated MAP2K phosphorylate and activate p38 MAPK. MKK6 can activate all four isoforms of p38, whereas MKK3 preferentially catalyze phosphorylation on p38α/β [46, 56, 57].

***Alternative Pathways of Activation*** The activation of p38 in T cells downstream of CD28 or IL12R/IL18R can occur via the conventional MAPK phosphorylation cascade. Alternatively, $CD4^+$ T cells stimulated through the T-cell receptor (TCR) activate p38 independent of the core MAP3K-MAP2K cascade (Fig. 3). In this activation pathway, p38α autophosphorylates (on Thr180

and Tyr182), following initial tyrosine phosphorylation of Tyr323 by Src- or Syk-family kinases Lck and Zap70 [58, 59]. The phosphorylation on Tyr323 is sensitive to Src-tyrosine kinase inhibitors or the presence of DNA-damage-inducible protein 45α (Gadd45α). Gadd45α is an inhibitor of activation-induced T-cell proliferation. Mice deficient in Gadd45α develop T-cell hyperproliferation and autoimmunity and show chronic phosphorylation of p38 [60–62].

It has also been speculated that autophosphorylation of p38 can be mediated by an interaction with scaffold proteins. TAB1 is a cytoplasmic scaffold protein with no catalytic activity that plays a critical role in TLR, TNF-α, and IL-1 signaling pathways. In the conventional signaling pathway, an interaction of TAB1 with TAK1 results in TAK1 autophosphorylation and subsequent activation (phosphorylation) of MKK3/6 and, in turn, p38 (Fig. 3). Through a yeast-two-hybrid screen, TAB1 was identified as a binding partner for p38α. It is speculated that after engagement with TAB1, p38α, like TAK1, would autophosphorylate [63]. A similar phenomenon was observed in mouse embryonic fibroblasts and appeared to involve a disulfide complex with an unknown ~85-kDa protein [64]. It follows that scaffold proteins other than TAB1 may participate in the activation of p38 MAPK. Indeed, the scaffold protein β-arrestin-2 controls the activity of JNK3 [65]. The scaffold proteins JIP1-3, KSR, and MP1 also bind to MAP kinases and should be investigated for the ability to regulate kinase activity and dictate cellular localization [66–68].

***Down Regulation of p38 Activation*** The main players in the dephosphorylation of p38 are (i) dual specificity serine, threonine, and tyrosine phosphatases (DUSP), which dephosphorylate p38α/β, (ii) serine/threonine protein phosphatase type 2C (PP2C) and (iii) protein tyrosine phosphatases, which are responsible for the dephosphorylation of p38γ/δ [69]. Although a mechanism was not elucidated, it was demonstrated that GSK3 can also negatively regulate p38 [70].

DUSPs are present in eukaryotes, including yeast, worms, flies, plants, and mammals. Approximately 30 protein-coding DUSP genes are present in the human genome [71], 11 of which have a MAPK binding domain (MKB) and can dephosphorate MAP kinases. Therefore, these phosphatases are also referred to as MAP kinase phosphatases (MKP) [71–73]. DUSPs have strong amino acid sequence identity (37–75% identity) in their catalytic domains. Three amino acids (Cys, Arg, and Glu) in the highly conserved consensus sequence are essential for enzymatic activity [74]. Mammalian DUSP/MKP have distinct biochemical properties with respect to substrate specificity, subcellular localization, and regulation and are subdivided into three subgroups based on phylogenetics and substrate specificity. MAPK-specific DUSP, which are active toward p38, also have activity for JNK and include DUSP1, 2, 4, 5, 8, 10, and 16 and a low-molecular-mass dual-specificity phosphatase, LMW-DSP2, which lacks the conserved cdc25 homology domain [75–77]. In general, these DUSPs are widely distributed, although DUSP2 is predominantly expressed in hematopoietic tissues, DUSP8 in the brain, heart, and lung, and DUSP10 in the liver and skeletal muscle. The phosphatases have variable subcellular location with some members residing predominantly in the nucleus (p38/JNK-specific DUSP1, 2, 4, 5), others in the cytosolic domain (p38/JNK-specific DUSP16), and others in both the nucleus and cytoplasm (p38/JNK-specific DUSP8, 10). DUSPs are not ubiquitously expressed but are regulated in a cell-type specific manner and/or in response to cellular activation. DUSP1/MKP1 is the archetypal member of the family and an immediate early gene in the innate immune response. DUSP1 can be induced by, amoung others, inflammatory stimuli such as LPS or M-CSF and infection [78]. Mitogenic stimuli causes a rapid induction of high quantities of MKP with messenger ribonucleic acid (mRNA) levels increasing 10- to 100-fold within 15–60 min of exposure to the mitogen. The expression may be controlled posttranscriptionally as mRNA has a short half-life, and ERK-phosphorylation has been shown to increase DUSP1 protein stability [79, 80]. The time course for MKP expression correlates well with the rapid yet transient activation of p38; activity peaks at 15 min and returns to baseline in macrophage cell lines within the first hour following exposure to TLR ligands. Antirheumatic and anti-inflammatory drugs (e.g., glucocorticoids) and anti-inflammatory cytokines (e.g., IL-10) also increase and prolong DUSP1/MKP1 to downregulate p38 signaling and the p38-driven expression of inflammatory mediators [79, 81, 82].

### 2.4 Structure–Function Relationship

In the years following the discovery of p38, extensive in vitro biochemical assays coupled with X-ray crystallography, molecular modeling, and structure–function analyses of p38 and inhibitory CSAID compounds have fueled advances in our understanding about the selective nature of regulatory proteins and CSAID inhibitors for, and differential substrates of, the p38 MAPK family.

Despite a highly conserved activation structure and catalytic domain, the isoforms of p38 differ in amino acid sequence identity, with p38α and p38β sharing 74% amino acid sequence homology and p38γ and p38δ sharing, respectively, 63 and 61% amino acid sequence identity with p38α. Two independent groups

solved the X-ray crystal structure of the inactive, nonphosphorylated form of p38α in 1997–98 [83, 84]. Like other kinases, p38α is comprised of two distinct domains separated by a deep hydrophobic pocket for substrate binding. The N-terminal domain of the molecule is responsible for binding to ATP, whereas the catalytic residues in the C-terminus facilitate the transfer of phosphate from ATP to the substrate. Spanning amino acids 160–185 in the C-terminal kinase subdomain VIII is an activation loop containing a conserved Thr-X-Tyr motif, the phosphoacceptor sequence [46, 85]. This motif is conserved in all MAP kinases where only the X amino acid varies; X is a Glu in ERK, Pro in JNK, and Gly in p38. These amino acids in the context of the full sequence of the activation loop dictate specificity of the various upstream activating MAP2K. Despite a common activation motif in p38α and p38β and 80% homology between MKK3 and MKK6, MKK3 does not activate p38β [56, 57], suggesting that other determinants are involved. MAPK docking sites and other molecular determinants within the N-terminal region of MAP2K also contribute to MAP2K selectivity [86]. A mutant of MKK3 lacking the p38 docking site preferentially phosphorylated tyrosine over threonine residues in the T-G-Y motif on p38α. These data offer evidence that docking sites also influence enzymatic activity [86–89].

MAPK docking sites (also known as D-domains or D-boxes) are present in the upstream activation kinases (MAP2Ks) as well as in MAPK substrates and other MAPK regulator proteins such as phosphatases and scaffold proteins [89]. Competitive binding of activators, substrates, and suppressors to MAPK docking sites may be a mechanism of MAPK regulation, perhaps one that is evolutionarily conserved as docking sites are present in the yeast MAP2K homolog Ste7p. A typical D-domain is comprised of a cluster of positively charged amino acids that are surrounded by hydrophobic residues. D-domains interact with common docking (CD) motifs that consist of acidic and hydrophobic residues and are located outside the catalytic domain in the C-terminus of MAPK [88, 89]. The traditional p38 consensus docking sequence is Arg/Lys-$X_4$-Leu/Ile-X-Leu/Ile, although additional sequences have been identified in several MAPK substrates. For example, a docking site of Leu-X-$(Lys/Arg)_4$ (where X is a hydrophobic residue) was identified in p38 substrates PRAK (p38-regulated/activated protein kinase) and MAPKAPK2. The number of contiguous basic residues following the hydrophobic residues correlates with MAPK specificity. Insertion of the p38 docking site of MAPKAPK2 into the ERK1/2-restricted substrate RSK2 was sufficient to convert RSK2 to a stress-activated kinase in vivo [90].

Molecular modeling predicts that the relative orientation of the MAPK N- and C-terminal domains is altered upon phosphorylation to promote ATP binding, substrate binding, and phosphate transfer. Irrespective of the activation state, CSAIDs bind near the ATP-binding pocket of p38, effectively inhibiting enzymatic activity by exclusion of ATP. SB203580 is the most extensively characterized CSAID and shows specificity for p38α/β MAPK when used at low doses [27, 91]. Located in the hinge of the ATP-binding pocket is a tri-amino-acid sequence Thr106-His107-Leu108 with Thr106 as the key residue controlling selectivity and sensitivity of p38α and p38β to inhibition by CSAID. The CSAID-insensitive p38δ/γ isoforms contain a Met-Pro-Phe sequence in this same position. The introduction of Thr106 or Thr106-His107-Leu108 into p38δ/γ (and JNK) renders these isoforms sensitive to SB203580, and, conversely, substitution of methionine for threonine at position 106 results in a 10-fold reduction in p38α sensitivity to SB203580 [92–94].

Structure–function analyses have identified the molecular determinants that confer enzymatic activity and binding affinity to p38. In vivo, however, the regulation and function of p38 is heavily influenced by multiple factors that are unique to different cell types and their state of activation. Variables include the relative expression and cellular localization of p38 isoforms, phosphatases, kinases, and scaffold proteins. In vivo substrate specificity of activated p38 pathway kinases is in large part determined by cellular accessibility to substrates. The true physiological activity of p38 reflects both the intrinsic biochemical properties of p38 and a given cellular milieu. Scaffold proteins have been ascribed a critical, although perhaps underinvestigated, role in spatial-temporal organization of p38 signaling cascade. Given that p38 MAP kinases have nuclear substrates yet are devoid of a nuclear localization signal, cytoplasmic scaffold proteins and shuttle proteins likely participate in the distribution and regulation of MAP kinases. The D-domain in several p38 substrates (MSK1/2, MAPKAPK2/3, and PRAK) overlaps with their nuclear localization sequence. Thus the engagement of p38 can alter cellular localization of the complex. It has been suggested that both DUSP8 and DUSP16 translocate to the nucleus, then bind, dephosphorylate, and transit p38 to the cytoplasm. These events may also depend on cellular redox states as a reduced form of DUSP1/MKP1 in alveolar macrophages enhanced p38 activity and TNF-α expression [95]. It should be apparent that the p38 pathway is complex and capable of customizing itself to the input stimuli. Therefore, it is advisable to interpret in vivo and in vitro assays addressing mechanisms of MAPK regulation within the specific cellular and environmental conditions under which the experiments were performed.

**TABLE 2 Diversity of p38α/β Substrates**

| Transcription Factors | Protein Kinases | Other |
|---|---|---|
| MEF2A/2C, SAP1/2, ATF2/6, NFAT, CDX3, CHOP/GADD153, PGC-1, p53, USF-1, STAT-1, Elk-1 | MAPKAPK2/K3, MNK1/MNK2, MSK1/MSK2, PRAK | Cdc42, cPLA2, Tau, TAB1-3, NHE-1, keratin8, stathmin, $p47^{phox}$ |

**TABLE 3 Substrates of p38α/β-Activated Kinases**

| p38 MAPK-Activated Kinase | Transcription Factors and DNA Modifying Agents | Binding Proteins, Enzymes, Regulators of Protein Translation |
|---|---|---|
| MAPKAPK2/3 | CREB, CHOP1, ATF1, SRF, ER81, E47 | hsp25, hsp27, TTP, hnRNP, HuR, 14-3-3, 5-lipoxygenase, glycogen synthase, tuberin, eEF2 kinase, PABP1, p16-Arc, TIAR |
| MNK1/2 | ERβ | eIF-4E |
| MSK1/2 | CREB, ATF1, NFκBp65, ER81, STAT3, Histone H3, HMG-14 | Bad, Akt, 4E-BP1 |
| PRAK | | hsp27 |

## 3 EFFECTS MEDIATED BY P38 MAP KINASES

### 3.1 Cellular Effects

***Targets of p38 MAPK*** The substrate specificity of p38 is determined at a molecular level by serine and threonine target residues in the context of proline (e.g., Pro-X-Ser/Thr-Pro) and p38-docking motifs. MAPK-activated protein kinase 2 (MAPKAPK2, MK2), a serine/threonine protein kinase, was the first p38 substrate to be identified [31, 32, 96]. Other substrates include, but are not restricted to, PRAK, MNK, MSK, RSK, cPLA2, ATF2, ATF6, MEF2C, MEF2A, CHOP, $p47^{phox}$, stathmin, Elk1, and SAP1A (Table 2). Most substrates were first identified with the assistance of CSAID compounds and confirmed through the use of p38α heterozygous knockout cells or in p38α–/– embryonic stem cells. The use of p38 inhibitors has favored the identification of p38α/β substrates (over p38δ/γ). However, it does not distinguish between substrates that are unique to the p38α or p38β isoform [97]. Not surprising, p38α and p38β have similar substrate profiles [34], and while some substrates are shared with p38γ and p38δ (such as MBP, ATF2, Elk1) there are some distinctions. For example, p38γ and p38δ do not activate MAPKAPK2/3; however, p38γ can bind to PDZ domains, thereby allowing it to interact with scaffold proteins in plasma membrane-associated cytoskeletal structures such α1-syntrophin, SAP90/PSD95, and SAP97/hDlg [98].

Transcription factors and protein kinases are the major classes of p38 substrates with p38δ and p38γ showing particular preference for transcription factor substrates [99]. The family of MAPK-activated kinases consist of five subfamilies based on sequence homology: RSKs, MSKs, MNKs, MAPKAPK2/3 (MK2/3), and MAPKAPK5 (MK5). p38-induced activation of MAPKAPK results in the downstream engagement of both cytosolic and nuclear substrates, which are unique to MAPKAP-kinases, many of which are involved in regulating gene transcription and protein translation (Table 3). Some MAPKAP-kinases have nuclear localization signals and are exclusive residents of the nucleus, thereby enabling activation of transcription factors such as the basic-leucine zipper (bZIP) transcription factor CREB (cyclic adenosine monophocphate [cAMP] response element [CRE] binding protein). As the name implies, CREB binds to certain sequences called cyclic adenosine monophocphate (cAMP) response elements (CRE) in DNA and transactivates the expression of genes such as TNF-α, IL-1, and IL-6 [100, 101]. The p38-induced activation of protein kinases bestows a secondary set of substrates to p38 that participate in the regulation of inflammatory mediators [102, 103].

***p38 MAPK Control over the Expression of Inflammatory Mediators*** The expression of inflammatory mediators can be complex and requires the contribution of multiple factors that influence gene transcription and translation. The most prominent transcription factors and coactivators that are required for the transcription of inflammatory genes include NFκB, AP1, p38, and C/EBP (CCAAT-enhancer binding protein) [1]. The expression of inflammatory mediators, in particular, pro-inflammatory cytokines, is also heavily dependent on factors that promote mRNA stabilization and protein translation. AU-rich elements (AREs) are common elements of the 3′ untranslated region (3′UTR) of inflammatory protein mRNA [1]. When complexed with AU-binding proteins, the transcripts are destabilized and degraded. The shortest strongly destabilizing

sequence thus far identified is the 34-nucleotide TNF-α ARE that contains six AUUUA motifs [104]. The prototypic zinc finger protein ZFP36 (tristetraprolin/TTP) inhibits TNF-α secretion from murine macrophages by binding the ARE and destabilizing TNF-α mRNA [105–107]. Inhibition of any one regulatory factor often results in a partial, not complete, abrogation of inflammatory protein expression.

Inflammatory genes regulated downstream of p38α/β activation were discovered by the use of the CSAID as well as inactive and constitutively active mutants of upstream kinases and downstream substrates of p38. There is significant diversity in the genes that are activated by p38 MAPK, which likely reflects the diversity of substrates that can be activated by p38 and p38-activated kinases. There is evidence that p38 controls the expression of cytokines (TNF-α, IL-1, IL-6, RANKL), chemokines (IL-8, IP-10, MCP-1, MIP-1α/β), extracellular matrix (ECM)-modifying enzymes (MMP1, 3), inflammatory enzymes (COX-2, NO synthase), cell adhesion molecules (ICAM-1), and growth factors (GM-CSF) [99].

Inhibition of p38 can have differential effects on mRNA and protein levels, a phenomenon commonly described for the expression of TNF-α [104, 108]. While not all ARE-containing mRNAs are regulated by p38, evidence exists for p38-dependent stabilization of TNF-α, IL-1β, GROα, IL-6, GM-CSF, COX-2, c-Fos, and IL-8 mRNA [1, 99, 108]. p38-induced activation of MAPKAPK2 and MNK1/2 results in, respectively, the activation of tristetraprolin (TTP) and eukaryotic initiation factor 4E (eIF-4E). TTP is an mRNA destabilizing element and eIF-4E a component of the eIF-4F protein complex that unwinds mRNA secondary structure and promotes the initiation of protein translation. Consistent with these findings, MAPKAPK2-deficient mice have impaired protein synthesis of TNF-α but normal levels of TNF-α mRNA. Accordingly, TTP null mice spontaneously express TNF-α as do mice deficient in the TNF-α 3′UTR AU-element. TNF-α production in TTP and TNF-α ARE knockout, but not wild type, mice were insensitive to p38 MAP kinase inhibitors despite normal activation of the upstream components of the p38 MAPK pathway [109, 110]. While the details concerning the connectivity between AU-binding proteins such as AUF1, HuR, TTP and the p38 pathway are not well established, stimuli-induced activation of p38 and subsequently MAPKAPK2 may result in, for example, the phosphorylation and release of TTP from TNF-α mRNA and subsequent expression of TNF-α protein [106, 107, 111, 112]. Therefore the regulation of inflammatory protein expression by p38 includes both transcriptional and posttranscriptional mechanisms. In response to stimuli such as IL-1, TNF-α, TLR, and stress (UV), p38 MAPK can

1. *Directly activate transcription factors*, for example, Elk-1 and SAP-1 [1, 46, 99] that control the expression of genes for inflammatory agents.
2. *Promote gene transcription via indirect activation of transcription factors and induction of transactivating factors and modifications to DNA that promote transcription*. p38 via MSK1/2 can activate CREB, NFκB subunit p65, phosphorylate histone H3 (to recruit NFκB) and TATA box binding protein TFIIDa, and activate high mobility group protein 14 (HMG-14) to remodel chromatin and promote transcription [113–115].
3. *Promote mRNA stability and protein translation* to facilitate rapid control over the expression of inflammatory agents. This can be achieved indirectly through MAPKAPK2-mediated phosphorylation of AU-binding proteins and stabilization of ARE-containing mRNA and via MNK-mediated activation of eIF-4E [109, 116–119].

The activation of p38 through different pathways may manifest in the activation of distinct substrates and diverse cellular functions. Indeed, TAB1-dependent activation of p38 stimulated a set of substrates that were distinct from those downstream of p38 when activated by MAP2K [120] and has been linked specifically to the maturation of monocyte-derived dendritic cells, induction and maintenance of peripheral T-cell anergy and intracellular infection, parasite-induced IL-12 production, and myocardial ischemia [62, 120–123]. In addition, TAB1 may become a substrate of p38 following activation in complexes of TAB1-TAK1-TAB2/3. Phosphorylated TAB1 inhibits TAK1 that would mediate a negative feedback loop on the activation of p38, TAK1, and IKK/NFκB in response to inflammatory stimuli. Furthermore, sequestration of p38 to one pathway may inhibit participation in other signaling pathways, responsiveness to new divergent inputs, and altered availability to substrates and regulators [124].

### 3.2 Biological Effects

A single genetic knockout of murine p38β, p38γ, and p38δ as well as a double knockout of p38γ/δ had no observable impact on birth frequency, size, or health of these mice. On the other hand, p38α is without question an essential kinase in embryogenesis [22, 125–128]. In the absence of viable p38α-deficient mice and no discernable phenotype in mice deficient in the other p38 isoforms, p38-inhibitor compounds have been used extensively to evaluate the role(s) for p38 in vivo. New generations of p38 inhibitors have improved specificity, yet CSAID toxicity and off-target effects within an already complex, functionally redundant and cellularly diverse p38 signaling pathway, can shed doubt on the

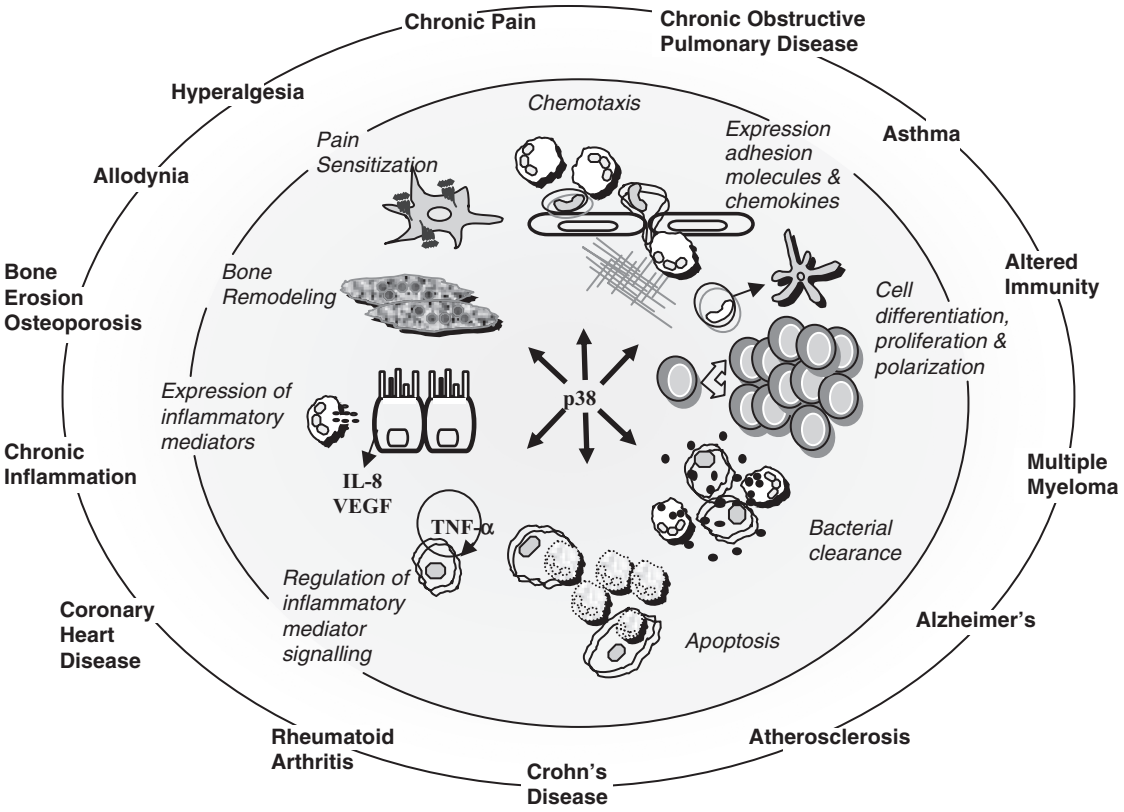


**Figure 4** Inflammatory roles of p38 MAPK. Dysregulation of any number of the cellular and physiological functions depicted in the core of the illustration can drive the development of adverse inflammatory disorders and diseases, examples of which are given on the periphery of the diagram.

authenticity of any physiological role(s) of p38 that are revealed using the chemical inhibitors. While the inhibitors are convenient to use for in vitro and in vivo experiments, biological roles should be confirmed using another system, for example, heterozygous knockout mice (p38α+/−), tissue-specific homozygous p38α depletion (e.g., in cardiac muscle), transgenic mice expressing a dominant negative form of p38α (dnp38a, T180A/Y182F) and/or in mice with genetic deletion of p38α substrates and activators. An ever increasing amount of in vitro and in vivo evidence using different model systems substantiates a role for p38α in inflammation, inflammation-mediated disease, and inflammatory pain including, but not limited to, rheumatoid arthritis (RA), inflammatory bowel disease (IBD), acute respiratory distress, ischemic heart disease, and sensitivity to pain (Fig. 4 and Table 4) [14, 129–140].

***Inflammatory Disease*** The chronic inflammatory disease RA is caused by an inflammatory mass of activated leukocytes in the synovium, 70% of which are $CD4^+$ T cells (50%) and macrophages (20%), with the remaining 30% of cells comprised of dendritic cells and B lymphocytes. The disease is driven by the overproduction of (p38-regulated) genes, including IL-1, TNF-α, RANKL, IL-6, GM-CSF, COX-2, collagenases, and MMPs that enhance inflammatory pain and promote bone and cartilage erosion. Badger et al. demonstrated that p38α/β and p38α-specific CSAID blocked LPS-induced TNF-α production in mice and reduced joint edema in the collagen-induced arthritic model. The p38α/β CSAID SB203580 was also effective at blocking IL-1-induced nitric oxide in bovine articular cartilage explants and chondrocytes and reduced serum TNF-α and protected against LPS/D-gal-induced endotoxic shock [141]. Subsequent experiments demonstrated that inhibitors with greater selectivity for the p38α isoform (SB242235, R130823 and SD282, indole-5-carboxamide) blocked paw swelling in the collagen-induced arthritic model and in the adjuvant arthritic rat model [133, 142, 143]. Since p38β-deficient mice show no defect in cytokine production [22, 144, 145], it has generally been accepted that p38α is the dominant isoform to mediate inflammation, and effects of CSAID are due to inhibition of the p38α rather than p38β isoform. O'Keefe et al. recently used a combination of specific genetic mutations and pharmacological techniques to conclusively demonstrate that only p38α is involved in acute

**TABLE 4 Sampling of in vitro and in vivo Evidence for p38α/β in the Inflammatory Response and Inflammatory Disease**

| Role | Assay System/ Model | | Observation | Reference |
|---|---|---|---|---|
| Inflammation; chemokine production | p38α/β inhibitor SB202190, SB203580, dominant negative MKK6 & constitutively active MKK6 | in vitro | Inhibitors and dominant negative MKK6 blocked TNF-α-induced MCP1 in HUVEC (and human mesangial cells)<br>Constitutively active MKK6 enhanced TNF-α-induced MCP1 in HUVEC | 259, 260 |
| | p38α/β inhibitor SB203580 | in vitro | Decreased IL-1 by bronchial epithelial cells exposed to IL-1β, TNF-α, or PAF | 261 |
| | p38α/β inhibitor SB202190 | ex vivo | Blocked LPS-induced TNF-α, MIP1α, IL-6, and EBV membrane protein-induced IP-10 by mRNA destabilization | 168, 262 |
| | p38α/β inhibitor SB202190 | in vivo | Decreased IL-2, 4, 5, 1β, 18, TNF-α, and IFNγ | 132 |
| Inflammation; chemotaxis | p38α/β inhibitor SB203580 | in vitro | Blocked LPS- and TNF-α-induced ICAM1 expression on pulmonary microvascular endothelial cells | 156 |
| | p38α inhibitor L-790070 | in vitro | Blocked serum-induced monocyte differentiation and RANTES-, MIP1α-, MCP1-, and fMLP-induced monocyte chemotaxis | 263 |
| | MAPKAPK2–/– | in vitro | Defective migration of MEF, smooth muscle cells and neutrophils (to fMLP)<br>Reduced expression of p38 | 264, 265 |
| | MAPKAPK2 RNA interference, MAPKAPK2–/–, hsp27–/– | ex vivo | Decreased migration of gene-deficient macrophages in response to HGF, PDGF, C5a, and CXCL12 | 266 |
| | $p38\alpha^{+/-}$ | in vivo | Decreased ear swelling, inhibition of cell infiltration by monocytes, neutrophils, and eosinophils in contact hypersensitivity model | 132 |
| | p38α inhibitor SB2439063 | in vivo | Inhibited IL-6, MMP9 activity, and neutrophil infiltration in BAL after exposure to LPS | 267 |
| Inflammation; cytokine production | p38α and p38α/β inhibitors | ex vivo | All inhibitors that have made it to clinical trials have shown reduced LPS-induced TNF-α in whole blood and human blood cells | 196, 198, 207 |
| | MAPKAPK2–/– | in vivo | Defective LPS-induced TNF-α, IL-6 and IFNγ production | 103, 265 |
| Inflammation | p38α/β inhibitor SB203580 and p38α inhibitor SD-282 | in vivo | Reduced peripheral tissue inflammation | 187 |
| | $p38\alpha^{+/-}$ | in vivo | Reduced ear swelling and infiltration of inflammatory cells (neutrophils, monocytes, eosinophils) in DNFB-induced contact hypersensitivity model | 132 |
| | $p38\beta^{-/-}$ and TNF-$ARE^{-/-}$ double knockout | in vivo | $p38\beta^{-/-}$ TNF-$ARE^{-/-}$ had equivalent predisposition to inflammatory diseases as TNF-$ARE^{-/-}$-deficient mice. p38β plays no critical role in predisposition to inflammatory diseases | 144 |
| | p38α/β inhibitor SC102 | in vivo | Protected in animal models of rheumatic disease | 268 |
| Inflammation; LPS challenge | p38α/β inhibitor SB203580 | in vivo | Protected against LPS/D-gal-induced endotoxic shock | 141 |
| | $p38\beta^{-/-}$ | in vivo | No phenotype in LPS-challenge model. p38β plays no critical role in inflammatory response to LPS | 144 |
| | $MAPKAPK2^{-/-}$ | in vivo | Protected against LPS-induced endotoxic shock | 103, 265 |

**TABLE 4** ***Continued***

| Role | Assay System/ Model | | Observation | Reference |
|---|---|---|---|---|
| Inflammation; Arthritis | Inflammed tissue from patients with RA | ex vivo | p38α and p38γ found are most prevalent isoforms | 269 |
| | p38α/β inhibitors | in vivo | Reduced paw inflammation in the adjuvant arthritic rat model and joint edema in the collagen-induced arthritis model | 141, 142 |
| | p38α inhibitor R130823 | in vivo | Efficacy in mouse and rat models of arthritis | 143 |
| | p38α inhibitor | in vivo | New synthesis of 2-amino-benzimidazole inhibitor with efficacy in rat collagen-induced arthritis model | 270 |
| | p38α inhibitor TAK-715 | in vivo | Efficacy in adjuvant-induced rat arthritis | 196 |
| | p38α inhibitor VGX1027 | in vivo | Reduced paw swelling in collagen- and polysaccharide-induced arthritis model | 194 |
| | p38α inhibitor SD-282 | in vivo | Reduction in LPS-induced inflammatory cytokine production, osteoclast formation, and alveolar bone loss | 133 |
| | $p38\alpha^{ki/ki}$ and $p38\beta^{ki/ki}$ T106M knock in mice are enzymatically active and insensitive to CSAID | in vivo | CSAID MRK48 reduced LPS-induced TNF-α and collagen antibody-induced arthritis in $p38\beta^{ki/ki}$ but not $p38\alpha^{ki/ki}$ mice p38α specifically involved in inflammation | 146 |
| | $MKK3^{-/-}$ | in vivo | Markedly decreased the severity of arthritis (production of inflammatory agents and cellular infiltration) in passive K/BxN model of arthritis or in response to LPS challenge | 233 |
| Inflammation; Asthma | p38α/β inhibitor SB203580 | in vivo | Attentuated BAL TNF-α production in rat model of allergen-induced asthma | 159 |
| | Aerosolized p38α antisense oligo (p38α-ASO) | in vivo | Inhibited ovalbumin-induced inflammatory cells, IL-4, IL-5, IL-13 in BAL, pulmonary eosinophilia, mucus hypersecretion, and airway hyperresponsiveness | 160 |
| Inflammation; COPD | Lung tissue samples from smokers with COPD | ex vivo | Increased phosphorylated p38 in alveolar macrophages | 135 |
| | p38α inhibitor SB2439063 | in vivo | Inhibited inflammatory mediators and bleomycin-induced right-ventricular hypertrophy in models of COPD and lung fibrosis | 267 |
| | p38α inhibitor SD-282 | in vivo | Inhibited infiltration of macrophages and neutrophils and inflammatory mediators (COX-2, IL6) in tobacco-smoke-induced pulmonary inflammation model | 161 |
| Inflammation; Acute lung | p38α and p38α/β inhibitors SD282 and SB239063 | ex vivo | Decreased LPS-induced TNF-α and GM-CSF in macrophages isolated from lung tissue of smokers, ex-smokers, emphysema patients | 162 |
| | p38α/β inhibitor SB203580 | in vivo | Reduced IL-1β and neutrophilia in rat aerosolized LPS challenge | 163 |
| | p38α/β inhibitor M39 | in vivo | Systemic inhibition decreased TNF-α and neutrophil accumulation in airspaces following intratracheal administration of LPS | 136 |
| Inflammation; atherosclerosis | p38α/β inhibitor SB203580 | in vitro | Attentuated thrombin-induced vascular smooth muscle cell migration and expression of VEGF | 271 |

(*Continued*)

**TABLE 4** ***Continued***

| Role | Assay System/ Model | | Observation | Reference |
|---|---|---|---|---|
| Inflammation; myocardial apoptosis and injury | Adenoviral expression of p38 in cardiomyocytes | in vitro | Activation of p38α caused apoptosis, whereas activation of p38β mediated protective hypertrophic response | 272 |
| | p38α inhibitor SD282 | in vivo | Improved cardiac function (stroke volume, cardiac output), reduced macrophage infiltration of inflammatory cells and abolished cardiomyocyte apoptosis in a rat model of myocardial injury | 137 |
| Inflammation; apoptosis | MAPKAPK2$^{-/-}$ | in vivo | Decreased IL-1β and focal ischemic brain injury | 164 |
| Inflammation; pain | p38α inhibitor R130823 | in vivo | Blocked hyperalgesia in rat adjuvant-induced arthritis | 143 |
| | p38α/β inhibitor SB203580 | in vivo | Reduced spinal nerve ligation induced mechanical allodynia (neuropathic pain) | 185, 186 |
| | p38α/β inhibitor SB203580 and p38α inhibitor SD-282 (Scios) | in vivo | Reduced thermal hyperalgesia and spinal activation | 187, 188 |
| Immunity; lymphocyte function | p38α inhibitor SCIO469 | in vitro | Suppressed IL-6 and VEGF and multiple myeloma proliferation | 138 |
| | Murine thymocytes | in vivo | Expresssion of p38 is required for early stages of T-cell development and induces apoptosis in CD8$^{+}$ but not CD4$^{+}$ T cells | 60, 273 |
| | p38α/β inhibitor SB203580 and dominant negative p38α (dnp38aT180A/Y 182F expressed under the lck promoter in transgenic B10Br) | in vivo | Chemical inhibitor blocked IFNγ production by Th1 cells but not IL4 by Th2 cells and dominant negative p38 impaired Th1 responses | 59, 273 |
| Immunity; bacterial/viral clearance | p38α/β inhibitor SB203580 | in vitro | Inhibited activation of neutrophil NADPH oxidase | 274 |
| | MAPKAPK2$^{-/-}$ | in vivo | Normal levels of IL-lβ, IL-6, NO, and ROS, decreased TNF-α and IFNγ production, enhanced bacterial load, and increased susceptibility to *Listeria* monocytogenes | 275 |
| | Murine models of infection | in vivo | Increased cytokine production and decreased bacterial clearance in mouse models of pneumococcal pneumonia and tuberculosis | 276 |
| | Various | in vitro in vivo | Antiviral and antiproliferative effects on hematopoiesis and tumorigenesis mediated by type I interferons | 277 |
| Inflammation and immunity | MKK3/MKK6 single and double KO | | MKK3/MKK6 double KO are not viable while single KO are viable; therefore likely have redundant roles | 278–280 |
| | MSK1 and MSK2 single and double KO | | No phenotype in single or double knockout | 103, 113 |
| | MK5$^{-/-}$ | | No phenotype and MK2/MK5 double knockout has the same phenotype as MK2 single knockout | 281 |

and chronic inflammation [146]. Based on the observation that Thr106Met substitution does not alter p38α/β kinase activity but does confer resistance to inhibition by CSAID [84, 93, 94], O'Keefe et al. constructed both p38α Thr106Met and p38β Thr106Met knock-in mice. p38α Thr106Met mice, but not p38β Thr106Met mice, were resistant to pharmacological inhibition (by CSAID MRK48) of LPS-induced TNF-α and collagen-induced arthritis. This concurs with a normal inflammatory response observed in p38β knockout mice, and results by Beardmore et al. of similar TNF-α production in wild type and p38β knockout mice in an LPS/D-gal challenge model and sensitivity of p38β knockout mice (in terms of TNF-α production) to SB203580 [144].

The inflammatory phenotypes observed in vivo using chemical inhibition have been substantiated using MAPKAP2, TNF-α, and DUSP1/MKP1 knockout mice. TNF-α ARE-deficient mice consitutitively produce high levels of TNF-α and develop pathologies similar to chronic inflammatory arthritis and Crohn's inflammatory bowel disease [110, 111, 147]. In human TNF-α transgenic mice, p38 activity contributed to the development of arthritis [148]. Mice deficient in MAPKAP2 are protected from endotoxic shock and are resistant to collagen-induced arthritis [103, 149]. As seen in vitro using pharmacological inhibition of DUSP1/MKP1 or overexpression of an MKP1 catalytic mutant, p38 (and JNK) activation was prolonged in vivo in MKP1–/– mice. TLR agonists also induced greater pro-inflammatory cytokines (TNF-α, IL-6, IL-10) and chemokines (MIP1α, MIP1β, MIP2) in MKP–/– animals [95]. The MKP1 knockout mice develop more severe collagen-induced arthritis, impaired kidney function, multiorgan dysfunction, and enhanced mortality [95, 150–153].

In chronic obstructive pulmonary disease (COPD) and asthma (airway inflammation), inflammatory cytokines and chemokines such as TNF-α, IFNγ, IL-4, IL-5, IL-8, RANTES, and eotaxin are important components of the disease [154, 155]. Pulmonary hypertension is believed to lead to chronic hypoxia [155], and hypoxic stimulation of pulmonary artery cells is also mediated by SAPKs [156]. An inhibitor of p38 blocked LPS- and TNF-α-induced CD54/ICAM1 expression on pulmonary microvascular endothelial cells [157], and bronchial epithelial cells exposed to a hyperosmolar medium produced IL-8 by a MAPK-dependent mechanism [158]. In a rat model of allergen-induced asthma, the p38α/β CSAID SB203580 blocked the production of TNF-α [159]. Along similar lines, aerosolized p38α antisense oligo was effective at inhibiting inflammatory mediators and airway hyperresponsiveness [160]. Various CSAID have proven effective at blocking cellular infiltration in chronic and acute lung models (Table 4) [136, 160–163], and an increased expression of phosphorylated p38 was detected in alveolar macrophages isolated from the lung tissue of smokers with COPD [135].

Heart, stroke, and atherosclerotic vascular disease are the leading cause of death in the United States each year. p38 plays a significant role in the control of apoptotic cell death and ischemia in cardiomyoctes, which appears to be isoform specific [164, 165]. The p38α isoform-specific inhibitor SD282 enhanced heart function and abolished apoptosis in a rat model of myocardial injury [137]. The role of p38 as a therapeutic target in heart disease was described extensively in a recent review by Clark et al. [166].

Multiple myeloma is a cancer affecting plasma B cells that is driven by cytokines and growth factors such as IL-6 and vascular endothelial growth factor (VEGF). Cross-linking CD40 on B cells stimulates p38 and MAPKAPK2 and results in NFκB activation, cell proliferation, and the induction of the cell adhesion molecule CD54/ICAM1. All inducible events were inhibited by CSAID SB203580 [167]. Another inhibitor of p38 (VX-745) blocked IL-6 and VEGF production by bone marrow stromal cells and subsequently inhibited paracrine-induced multiple myeloma cell growth [168]. An inhibitor of p38 in clinical trials for multiple myeloma (SCIO-469) increased susceptibility of the myelomas to killing by proteasome inhibitors [138, 169].

The cytokine RANKL and chemokine MIP1α are expressed in a p38-dependent manner by myelomas [170]. Also in a p38-dependent manner, RANKL drives the formation of osteolytic bone lesions that are associated with this type of cancer. An analogous process is at work in rheumatoid arthritis. Here, RANKL is produced by T cells and RANKL-induced osteoclast differentiation mediated by p38 causes bone erosion associated with RA [170–173].

These data indicate that p38α contributes to the inflammatory disease in vivo, and inhibition of p38 was sufficient to reduce the severity of disease.

***Pain*** Inflammatory mediators, via a p38-dependent mechanism, stimulate the production of inflammatory mediators and neuroactive substances. Neuroactive peptides and inflammatory cytokines are overexpressed in the spinal cord in peripheral inflammatory and neuropathic experimental models of pain. The synthesis of inflammatory mediators and neuroactive substances that act on nociceptive neurons is dependent on p38 MAPK. IL-1β-stimulated p38 in turn induced iNOS and the production of NO. The production of NO caused sensitization of spinal cord nociceptors and thermal hyperalgesia (heightened perceptions of pain) and allodynia (perceived pain when there should be none)

[174–176]. Inhibitors of IL-1β, p38, or iNOS block NO production and sensitization [177, 178]. Primary sensory neurons and p38 are also reactive to inflammation-induced neuronal growth factor (NGF). Here, p38 drives the expression of TRPV1/VR-1, a transient receptor potential/heat gated ion channel causing enhanced inflammatory heat hyperanalgesia and heightened perception of pain [179]. The negative controls on p38 in this system are not well known, and, unless the positive feedback loop is broken, chronically inflamed nociceptors continue to transmit a barrage of signals to the nervous system, causing chronic sensitization and heightened, continuous feelings of pain.

Flavonoids are present in plants and are believed to possess natural antioxidant and anti-inflammatory properties. It is known that flavinoids can inhibit protein kinases and reduce ERK and JNK phosphorylation in carcinoma cells. The flavonoid myricitrin has antinociceptive effects in experimental models of acute and chronic pain. This was shown previously to involve decreased PKC and NFκB activation, NO production, altered $Ca^{2+}$ and $K^{+}$, and activation of G protein pathways [180]. Meotti et al. further demonstrate that myricitrin inhibits p38, but not ERK1/2 or JNK1/2, phosphorylation in response to cytokine stimulation and prevented TNF-α- and IL-1β-induced biting behavior in mice [181].

Opioid peptides [e.g., Met-enkephalin (ENK) and B-endorphin (END)] bind to cognate receptors on peripheral sensory nerves and locally inhibit pain. Polymorphonuclear (PMN) cells are the major source of opioid-mediated pain control during the early phase of inflammation. Depletion of PMN abolishes antinociception induced by inflammatory chemokines (CXCL2/3). The effect can be reversed by a local injection of allogenic PMN [182, 183]. p38 MAPK mediate the chemkine-induced release of opioid peptide and subsequent pain control. Chemokine-induced release of opioid peptides from human PMN intracellular vesicles in vivo can be blocked by antipeptide antibodies as well as by inhibitors of p38 but not ERK1/2 [184, 185].

It appears then that p38 can have dual effects in the regulation of pain; contributing to pain through the stimulation of neuroactive peptides and inflammatory mediators and reducing pain by inducing neutrophil release of opioid peptides. In vivo, inhibition of p38 blocked hyperalgesia and mechanical allodynia in various rodent models of pain [143, 184–188], and, clinically, inhibition of p38 showed enhanced efficacy in pain control compared to ibuprofen, a hallmark discovery for the use of p38 as an analgesic [189]. Improved mechanistic information for the induction and maintenance of neural plasticity is required to further the development of effective drug regimes to control acute and chronic pain [2, 190].

## 4 P38 MAPK AS A THERAPEUTIC TARGET

Nonsteroidal anti-inflammatory drugs (NSAIDs) such as aspirin and ibuprofen have analgesic, antipyretic, and anti-inflammatory effects and are currently the most efficacious treatment to reduce inflammation, fever, and pain. Immunomodulatory agents, often referred to as biologic response modifiers (BRM) such as the TNF-α blocker drugs Enbrel (etanercept; soluble TNF-α receptor), Remicade (infliximab), and Humira (adalimumab; anti-TNF-α antibodies) as well as Kineret (anakinra), an IL-1 antagonist, have also seen success in the treatment of inflammatory diseases. The investigation into p38 as a potential therapeutic target began shortly following the fortuitous discovery of p38 coincident with the discovery of a crucial p38 substrate (MAPKAPK2) and a potent inhbitor that blocked inflammatory cytokine production. Several recent reviews have discussed the potential of p38 as a therapeutic agent for the treatment of a number of diseases [129, 130, 133, 134, 139, 140, 154, 176, 189, 191–198].

It is anticipated that compounds that inhibit p38 will confer a significant clinical benefit over current anti-inflammatory therapies at the very least for diseases in which anti-TNF-α therapy has already proven effective (e.g., RA, psoriasis, Crohn's disease). Unlike current therapies that generate a clinical anti-inflammatory effect by inhibition of a single inflammatory mediator (TNF-α, IL-1β, or COX-2), the suppression of p38 would simultaneously influence the production of multiple inflammatory agents. This feature should have superior efficicacy, which has already been observed clinically when anti-TNF-α and anti-IL-1β therapies were used in combination [199]. In addition to blocking the cellular expression of inflammatory mediators, anti-p38 compounds have the added benefit of blocking cellular responses to those mediators (including TNF-α, IL-1β, and RANKL). Inhibition of a multifunctional protein such as p38 may confer a significant advantage over current biologic therapies such as the ability to achieve positive results clinically with inhibitors that possess only moderate inhibitory potency towards p38. From a therapeutic perspective, it is also important to recognize that p38 participates in processes within both the immune and nervous system, and p38-regulated inflammatory agents mediate cross-talk between the two systems. Hence anti-p38 inhibitors are primed to have dual anti-inflammatory and anti-analgesic effects and could be administered for use in multiple clinical applications for the treatment of acute and chronic inflammatory diseases and pain. The oral bioavailability of p38 inhibitors significantly reduces the economic load associated with the therapy and imparts a convenient drug administration regime.

### 4.1 Previous Successes and Failures

The prototypic CSAID SB203580 has served as a template for the generation of other CSAID pryidinyl-imidazole inhibitors and new classes of small molecule inhibitors of p38. Much of this work has been the result of detailed molecular modeling and crystal/co-crystal structures of p38 and pryidinyl-imidazole group inhibitors [83, 84]. Pyridinyl-pyrazole scaffold inhibitors emerged as the second generation of p38 inhibitors. They were designed to have increased activity and selectivity with reduced toxic side effects and were generated by substitutions to the imidazole ring of pryidinyl-imidazole compounds. VX-745 (Vertex Pharmaceuticals) is an inhibitor in this class that achieved enhanced selectivity for p38 over JNK and ERK via the altered orientation of a small glycine amino acid that is located at position 110 in p38 but not in ERK or JNK. Both pyridinyl-imidazole and pyridinyl-pyrazole inhibitors lie in the ATP-binding pocket of p38 and rely on competitive inhibition of ATP to block p38 enzymatic activity [200, 201]. In contrast, the third class of p38 inhibitors to materialize, diaryl ureas, engage sites away from the ATP-binding pocket. Binding of these inhibitors to p38 disrupts the activation loop and blocks phosphorylation of p38, thereby locking it in an inactive conformation [202]. Several other classes of p38 inhibitors have been developed, including substituted benzamides, diaryl ketones (e.g., RO201195), and indole amides (e.g., SCIO469 and 323). The chemical structure and synthesis of p38 inhibitors has been recently reviewed by several groups [195, 198, 203–209].

By 2005, more than 20 potent p38 inhibitors, including BIRB-796, SCIO469, and VX702, had been designed with enhanced selectivity over the original pryidinyl-imidazole inhibitors. Preclinical studies demonstrated that each of these inhibitors reduced, in vitro and in vivo, inflammatory cytokine production and were effective in vivo in various models of animal disease (examples given in Table 4). With the significant promise bestowed on these compounds by the in vitro and in vivo studies, many inhibitors entered clinical trials to treat a range of inflammatory diseases. Preliminary studies demonstrated that AMG 548 (produced by Amgen) blocked LPS-induced TNF-α and IL-1β release and TNF-α-induced IL-8 release in whole blood, had 85% inhibition of TNF-α and IL-1β ex vivo, and was effective against collagen- and adjuvant-induced arthritis in rats. AMG 548 was suspended from future investigation as a human therapeutic due to random elevations in liver enzymes. Similarly, clinical trials with BIRB796 (Boehringer Ingelheim) were discontinued due to a suspected negative impact on liver enzyme production and, disappointingly, no observed efficacy against Crohn's disease [210] despite effectiveness in human endotoxin challenge. Unfortunately, many other

**TABLE 5 Clinical Trials with p38 MAPK Inhibitors**

| Target Disease(s) | Drug Name | Producer | Status/Results | Reference |
|---|---|---|---|---|
| RA, osteoporosis | AMG548 | Amgen | Discontinued<br>85% inhibition of ex vivo of LPS-induced TNF-α and IL1β<br>Altered liver enzymes | 198 |
| Rheumatoid arthritis | VX745 | Vertex | Discontinued<br>Increased ACR20 in approx 40% of patients (2.5 × placebo control)<br>Crossed blood brain barrier | 213 |
| | VX702 | Veretx | Phase II complete Feb 2006<br>Increased ACR20 in ~40% of patients (1.3 × placebo control)<br>Rash, infection, GI intolerance | 198 |
| | SCIO469 | Scios, J&J | Phase II complete<br>Increased ACR20 in ~50% of patients (2.3 × placebo control) | 212 |
| | RO3201195 | Hoffmann-LaRoche | Phase I in progress | 196 |
| | RO4402257 | Hoffmann-LaRoche | Phase II in progress Apr 2006 | www.clinicaltrails.gov |
| | PH797804 | Pfizer | Phase II in progress Dec 2006 | www.clinicaltrails.gov |
| | AVE-9940 | Sanofi-Aventis | Phase I discontinued | 139 |
| | VGX-1027 | VGX Pharma. | Phase I 2007 | VGX press release 28 Apr 2006 |
| | 856553 and 681323 | GlaxoSmithKline | Phase II | www.gsk.com |

(*Continued*)

**TABLE 5** ***Continued***

| Target Disease(s) | Drug Name | Producer | Status/Results | Reference |
|---|---|---|---|---|
| Antirheumatic | TAK-715 | Takeda | Phase I | 195 |
| RA, Crohns, psoriasis | RWJ67657 | Johnson & Johnson | Discontinued | 282 |
| Crohns | BIRB 796 | Boeringer Ingelheim | Discontinued<br>No observed efficacy against 284 patients with Crohn's disease despite effectiveness in human endotoxin challenge | 211 |
| COPD | 856553 and 681323 | GlaxoSmithKline | Phase II complete | www.clinicaltrails.gov |
| Atherosclerosis | BIRB 796 | Boeringer Ingelheim | Phase I<br>Inhibited LPS-induced coagulation in 24 healthy individuals | 283 |
| | 856553 and 681323 | GlaxoSmithKline | Phase II | www.gsk.com |
| | BMS582949 | Bristol-Myers Squibb | Phase II in progress Jan 2008 | www.clinicaltrails.gov |
| Heart disease | VX702 | Veretx | Phase II<br>Decreased serum levels of inflammatory biomarker CRP in acute coronary syndrome | Vertex press release 18 Oct 2004 |
| | 681323 | Glaxo SmithKline | Phase II in progress Feb 2006<br>Coronary heart disease with percutaneous intervention | www.clinicaltrails.gov |
| | Tacrolimus | Astellas Pharma Inc. | Phase III in progress<br>Heart transplantation and heart disease | www.clinicaltrials.gov |
| | SCIO323 | Scios, Johnson & Johnson | Phase I<br>stroke | www.siosinc.com |
| Pain | SCIO469 | Scios, Johnson & Johnson | Phase II complete<br>Compared to ibuprofen, SCIO469 extended dental pain tolerance for several hours in 263 individuals following extraction of impacted molars<br>First clinical demonstration of anti-nociceptive effects in acute pain | 205 |
| | 681323 | Glaxo SmithKline | Phase II in progress Aug 2006<br>Neuropathic pain following nerve trauma | www.clinicaltrails.gov |
| Cancer | SCIO469 | Scios, Johnson & Johnson | Phase II expected completion Apr 2006<br>Multiple myeloma | www.clinicaltrails.gov |
| | SCIO469 | Scios, Johnson & Johnson | Phase I and II in progress Apr 2005<br>Myelodysplastic syndromes | www.clinicaltrails.gov |
| Endotoxin challenge | RW767657 | Johnson & Johnson | Decreased TNF-α, IL-6, IL-8 and flu-like symptoms in endotoxaemia model in healthy human volunteers | 281 |
| Depression | 856553 | GlaxoSmithKline | Phase II in progress Dec 2007 | www.clinicaltrails.gov |
| Unspecified | ARRY162 and 797[a] | Array Biopharma Inc. | Phase I | Press release 30 June 2007 |

[a]Inhibitor of MEK; all other compounds are inhibitors of p38α or p38α/β.

p38 inhibitors have met the same fate and have failed to advance through human clinical trials (Table 5) [139, 140, 198, 211–213].

Deleterious unforeseen effects of p38 inhibitors that come to light in clinical trials are likely the result of insufficient biological information rather than failures associated with the structure–function relationship between p38 and the inhibitor. Such academic drawbacks may include a lack of sufficient knowledge regarding detailed MAPK pathways and cross-talk in the MAPK family under basal and disease conditions and/or incomplete understanding of the role(s) of MAPK in biological processes. Inhibitors were most commonly withdrawn from the clinic due to what appeared to be off-target, toxic effects. Pyridinyl imidazoles were discovered to interact with hepatic cytochrome p450 liver enzymes involved in drug metabolism [214], which is the most likely explanation for the observed liver toxicity with many of the original p38 inhibitors. This finding is also important as it demonstrates that the compound (pyridinyl imidazole) itself, not the target (p38), was the cause of toxicity, thereby preserving the reputation of p38 as a safe and viable drug target. The focus now is to build on the progress and failures of previous p38 inhibitors as therapeutic agents, honing selectivity to decrease off-target effects and toxicity. The most effective p38 inhibitors may share certain general properties (summarized in Table 6) and may be those that reduce but do not abrogate the production of all inflammatory mediators. In this manner, they can selectively modulate aspects of inflammation while maintaining normal physiological functions and a relatively normal balance between positive and negative regulatory pathways [215].

**TABLE 6 Properties of Safe and Efficacious Therapeutic Agents**

***Potent***
- Interact with inactive and active modules of the target protein
- Interact with a region on the target protein that has a selective advantage over in vivo competitors (e.g., targeting regions of p38 separate from those utilized by ATP)
- The therapeutic target is multifunctional, thereby inhibition (or activation) produces the desired physiological outcome through multiple avenues (e.g., p38 can inhibit cytokine production and cytokine-mediated signaling)

***Selective***
- Avoid targeting conserved domains or ubiquitously expressed molecules
- All biological roles of the target protein and any relevant splice variants have been deciphered
- Target molecules have a well-defined role in the disease of interest

***Safe***
- Aberrant function of the target protein does not provoke undesired effects in other systems (e.g., anti-inflammatory agents can compromise immunity)
- All biological roles of the target protein and any relevant splice variants have been deciphered
- Targets do not impair essential cellular processes (e.g., gene transcription)

### 4.2 Moving Forward

***New p38 Inhibitory Compounds*** Early counter-screens of CSAID cross-reactivity for other kinases were rather modest and suggested that these compounds had specific activity toward p38α/β. It has become apparent, however, that this is not the case. With a high degree of structural conservation in the adenosine binding pocket and functional redundancy within the approximately 520 kinases in the human genome [9], it is not wholly unsurprising that, while showing preference for some kinases over others, the CSAIDs are not monogamous inhibitors. The well-characterized and widely used SB2300580 CSAID has measurable enzymatic activity and binding reactivity with nonenzymatically active cellular proteins and cellular kinases, including enzymes from distantly related families, for example, GAK, CKII, JNK, Raf, Lck, GSK3β, PRKα, and RICK, with reactivity toward RICK equivalent to that for p38α/β [198, 216–218]. Some of these kinases are critical in the inflammatory and immune response and may have positively contributed to the anti-inflammatory action observed in vitro and in vivo with this inhibitor. On the other hand, it is also possible that inhibition of these kinases may cause immune suppression or negative effects on other organ systems in a clinical setting [219, 220]. Without a doubt it is an imposing task to achieve substrate specificity with kinase inhibitors that target the ATP-binding pocket. Nevertheless, some kinase inhibitors have been used successfully to treat cancer. Licensed drugs include ABL kinase (Gleevec) from Novartis and EGFR kinase (Iressa) from AstraZeneca [221], and synthetic blockers of MAP kinases are showing promise in the treatment of myelomas [138, 222, 223]. Current p38 inhibitors that bind the ATP-binding pocket can do so on both the inactive, nonphophorylated and active, phosphorylated forms of p38. In contrast, ATP can only engage the active form of p38, which gives the inhibitor a competitive advantage over ATP and eliminates the need for extremely high levels of potency in vivo in order to contend with the presence of millimolar levels of cellular ATP [224, 225]. Nevertheless, it may be advisable to avoid the highly conserved ATP-binding site and target an allosteric site in p38 MAP kinases to reduce

possible off-target effects on proteins with identical or similar functional domains. This should minimize laborious screening of the inhbitor for cross-reactivity with other kinases [140]. BIRB796 has been the prototype for inhibitors that do not engage the ATP-binding pocket and that possess slowed dissociation rates ($t_{1/2}$ of 24 h) compared to the ATP-binding pocket directed inhibitors ($t_{1/2}$ of 1 min). Moss et al. have recently reported new methods using synchronized crystallographic structure information, chemical properties, and molecular modeling to design new classes of inhibitors with BIRB-like features. The resulting compounds diminish TNF-α release from THP1 cells, whole human blood, and LPS-treated mice [208]. Natarajan et al. define a new subclass of pyridinyl-pyrazole p38α inhibitors, the 7-amino-naphthyridones. Selected molecules from this class demonstrated potent inhibition of p38 and blocked TNF-α release in human whole blood, in a LPS challenge model when given orally 27.5 h prior to LPS and in rats where inhibition of TNF-α was similar or better than that achieved with dexamethasone. In addition, the inhibitors were effective in models of disease, showing efficacy in a rat antigen-induced arthritis model [209]. A host of other computer- and structural biology-aided methods have been used to design inhibitors of p38 such as 3D- and 4D-QSAR modeling [226, 227]. There are a number of excellent publications providing detailed discussions of structure and activity relationships (SAR) and molecular modeling of p38 inhibitors [198, 203, 228–231]. This sort of combined approach using detailed structure–function analyses, X-ray crystallography, and site-directed mutagenesis studies will continue to provide a strong molecular basis to understand the selectivity of p38 inhibitors and to design new inhibitors. Lee and Dominguez suggest that every kinase-specific difference in structure should be exploited in order to discover inhibitors with true selectivity and specificity [198].

***Alternative Targets*** These same techniques can be applied to the design of inhibitors that target p38-interacting proteins in order to suppress p38 [232]. In this manner the inherent difficulties in explicitly targeting molecules with highly conserved regions of activity and functional redundancy can be avoided, and a new pool of proteins on the periphery of the central MAP3K-MAP2K-MAPK pathway can be considered for their potential as safe and efficacious targets to downregulate p38. Potential molecules of interest include MKK3, MKK6, MAPKAPK2, and PRAK. MKK3 deficiency markedly decreased the severity of synovial inflammation in spontaneous arthritis in the KRN transgenic mouse (K/BxN model of arthritis) [233]. MAPKAPK2 has been deemed the most important substrate to execute the inflammatory properties of p38. The knockout mice have an anti-inflammatory phenotype and, most conveniently, have decreased levels of p38α [103]. Efforts are underway to find natural MK2 inhibitors [234]. Small-molecule inhibitors that induce a conformational change in p38 also provide a structural basis for the rational design of inhibitors that induce an inactivating conformation, for example, one that excludes ATP binding or alters substrate interactions. Mutations in the hydrophobic core can cause a conformational change that allows autophosphorylation of p38 [235]. A similar conformational change is induced by TAB1 and may explain the observed autophosphorylation of p38. Small-molecule inhibitors that disrupt p38-TAB1 interactions (or p38 autophosphorylation) may yield specific benefits in myocardial ischemia and TLR-induced inflammatory responses [121]. Studies using MKP/DUSP knockout mice define a critical role of this family of phosphatases in controlling p38 activity. Very small changes in MKP expression is sufficient to drastically downregulate p38 activity and MKP1–/– mice are impaired in their ability to control inflammation. While MAPK phosphatases should not be excluded as potential targets to modulate p38 activity, consideration needs to be given to the impact of these phosphatases in particular on JNK MAPK and other non-MAPK substrates. As a general rule, any serious investigation of new targets to regulate p38 function should be restricted to molecules that have activity exclusively within the p38 pathway and a well-defined role in inflammatory disease.

***Natural Products*** Insight into novel targets to regulate p38 may also be gained from bacterial systems and natural products. Virulence factors can effectively subvert immune detection by manipulation of cellular signaling pathways. For example, YopJ from *Yersinia* catalyzes the acetylation of serine/threonine residues in MAP2K activation loops, thereby preventing phosphorylation and activation by MAP3K and activation of downstream substrates. The bacterial proteins themselves may be useful templates for therapeutic design as well as the process of acetylation of phosphorylated residues [236, 237]. Resveratrol, tangeretin, and ligustilide are natural products of wine, citrus peels, and ginseng that are natural inhibitors of SAPK/MAPK [238–240]. In addition, inflammatory cytokines TNF-α, IL-1β, IL-6, Th1 cytokines IFNγ, and IL-12, and phosphorylation of JNK and p38 were reduced in LPS-stimulated human peripheral blood mononuclear cells treated with crude and pure extracts from the plant *Boswellia serrata* [241]. These are examples of natural products that could be templates for small-molecule pan- or isoform-specific inhibitors of p38 [238–243]. Paul et al. describe in situ chemical modeling approaches

for the design of such products of natural origin that inhibit TNF-α signaling and MAPK signaling [244]. These natural inhibitors were then analyzed in silico for cell surface properties and computational prediction of in vivo absorption, distribution, metabolism, excretion, and toxicity in order to determine which compounds would be suitable as drug candidates [206]. Other natural and endogenous products such as host defense peptides are known to act on MAPK to suppress LPS-induced release of pro-inflammatory mediators while still protecting rodents from infection [245].

### 4.3 Biological Considerations

Effective inhibition at the molecular level must translate to effective inhibition in vitro and in vivo of cellular and physiological activities. Construction of specific inhibitors based on solid structural designs may still be prone to failure in the clinical setting. This may reflect gaps in our understanding of p38 biology compared with the highly advanced understanding of p38 on a molecular level. These disparate levels of understanding constitute a somewhat unharmonious situation when both are required for the rational design of a therapeutic agent.

As undesirable pharmacological activities of p38 are a formidable concern in the clinical applicability of p38 inhibitors, it would be ideal to understand the impact of p38 on all aspects of inflammation and immunity in advance. Particular attention should be devoted to understanding the biological processes that p38 controls, the mechanisms by which p38 regulates posttranscriptional control of protein expression and a clear delineation of the signaling pathways and role(s) of each of the isoforms in human disease. Complications associated with isoform-specific effects have become a greater concern as accumulating evidence suggests that the isoforms have different physiological roles. At present, the impact of p38γ on inflammatory disease has been somewhat ignored, even though this isoform is highly expressed in tissues and cells of the immune system that are responsible for the production of inflammatory mediators that drive disease. p38α and p38γ are the dominant isoforms found in inflamed RA joints [246]. A number of new in situ molecular modeling approaches are being utilized to develop highly specific inhibitors for the various isoforms of p38 [247, 248]. Based on the discovery that mutations in the hydrophobic core of p38 (Tyr69, Phe327, and Trp337) can create a constitutively active enzyme [235], Askari et al. mutated aspartic acid residues in the activation loop of p38 (Asp176 in p38α, β, δ and Asp179 in p38γ) to produce autoactivated p38 isoforms. Using the constructs, they demonstrated that AP-1-mediated transcription was stimulated by p38α but inhibited by p38γ [248]. These constructs will be extremely valuable tools for the dissection of p38 isoform-specific signal transduction and physiological roles attributed to each isoform.

Considerable work is necessary to discover the biological consequences of suppressing the p38 cellular pathway. Afflicted systems include those whose function depends on normal p38 function and/or the presence of mediators regulated by p38. Both glucocorticoids and anti-TNF-α therapy compromise human defense. Data suggests that p38 plays an important role in the activation of lymphocytes, thus it is reasonable to predict that inhibition of p38 in vivo would compromise innate and cell-mediated immunity [249]. Indeed the p38 MAPK pathway is important in the clonal expansion of lymphocytes [58]. CD28-, IL-2-, or IL-7-dependent T-cell proliferation and IL-2 production can be inhibited by SB203580 [250, 251]. Mice deficient in MKP5/DUSP10 had enhanced levels of pro-inflammatory cytokines during innate immune responses and exhibited greater T-cell activation [252–254]. p38 can also influence the expression of T-cell polarizing cytokines. CSAID SB203580 inhibited IFNγ production by Th1 cells and overexpression of dominant negative p38 impaired Th1 responses. Likewise, mice deficient in MKK3 or MKP1 were defective in IL-12 and IFNγ production [150, 255]. MKP1–/– mice succumb to challenge by both gram-positive and gram-negative microorganisms [150, 152]. p38α is also required for the expression of chemokines and chemotactic agents such as MCP-1, MIP-1α, MIP-1β, and IP10 as well as the cellular response (of T cells, monocytes, macrophages, and neutrophils) to MCP-1, MIP-1α, MIP-1β, RANTES, and fMLP. The expression of chemokines (MCP-1, MIP-1α, MIP-1β, IP10) was blocked in p38α+/– mice, resulting in inhibition of monocyte, neutrophil, and eosinophil infiltration. While this reduced ear swelling in a model of contact hypersensitivity [132], the same suppressive effects would also hamper immune responses to infection. While there was a small decrease in antibody response to antigen challenge, Badger et al. reported no other immunosuppressive effects of the p38 inhibitor in vitro or in vivo despite the potent anti-inflammatory activity [141]. Furthermore chimeric mice with p38α-deficient ES cell-derived lymphocytes have no developmental defect and responded to antigen and antibody [123]. These data suggest that perhaps specific inhibition of the p38α isoforms may not have negative effects on lymphocyte activity. Nevertheless, a thorough investigation into immune defense should be conducted and should involve multiple cell types as inhibition of p38 signaling will have differential effects in different immune cells.

Other potential systems that are prone to negative effects of p38 inhibition include reproduction and

development, coagulation and endothelial function, skeletal muscle function, hematopoiesis, central and peripheral nervous system function, and liver function [256]. p38 inhibitors have caused hepatotoxicity, CNS toxicity, cardiotoxicity, skin rashes, gastrointestinal (GI) tract symptoms, and increased the risk of infection. It will be important to explore approaches to investigate and predict adverse side effects and to determine whether these side effects are related to the target or to the compound. Investigation to achieve this would encompass detailed biochemical characterizations and systems biology approaches [257, 258]. The p38$\alpha^{ki/ki}$ mouse constructed by O'Keefe et al. may be useful for distinguishing between on- and off-target effects of p38α inhibitors as off-target effects of inhibitors would cause toxicity in both wild-type and knock-in mice [146]. Equally important, particular organ systems should be flagged and carefully monitored for toxicity when new drugs enter the clinic. This is particularly important when treating a chronic condition that requires lifetime therapy.

## 5 CONCLUDING REMARKS

Inflammation is a complex process controlled by numerous positive and negative regulators. The p38 MAP kinase pathway is integral to the production of mediators that drive inflammation and pain and the cellular responses to these mediators. If not properly controlled, it can cause dire effects in the inflammatory, immune, and nervous systems that manifest in the development of chronic pain and inflammatory disease. Early inhibitory compounds of p38 demonstrated significant clinical efficacy and served as proof of principle that p38 is an excellent anti-inflammatory therapeutic target. However, these inhibitors were also beset with undesirable off-target toxic effects. Newer versions of these molecular inhibitors must be designed in a manner that is more mindful of the complexity of p38 and the inflammatory response to reach their potential as an effective anti-inflammatory and analgesic therapy.

Abbreviatizons

| | |
|---|---|
| ABL | abelson tyrosine kinase |
| ACR20 | American College of Rheumatology 20% Improvement Criteria |
| ASK1 | apoptosis signal-regulating kinase 1 |
| ARE | AU-rich element |
| ATF | activating transcription factor |
| BAL | bronchoalveolar lavage |
| BDNF | brain-derived neurotrophic factor |
| Cdc42 | cell division cycle 42 |
| CFU-E | colony-forming unit-erythroid |
| CGK1 | cGMP-dependent protein kinase |
| CHOP | C/EBP homologous protein, aka GADD153 |
| CKII | casein kinase II |
| CML | chronic myelogenous leukemia |
| CRP | C-reactive protein |
| CSF | colony-stimulating factor |
| CXCL1 | chemokine (C-X-C motif) ligand 1 |
| D-gal | D-galactosamine |
| DLK | dual leucine zipper-bearing kinase |
| DNFB | 2,4-dinitro-1-fluorobenzene |
| E47 | E2A-encoded transcription factor |
| 4E-BP1 | 4E binding protein 1 |
| eIF-4 | eukaryotic initiation factor 4 |
| eEF2 kinase | eukaryotic elongation factor 2-kinase |
| EBV | Epstein–Barr virus |
| EGF(R) | epidermal growth factor (receptor) |
| Elk1 | ETS-domain transcription factor 1 |
| ER | estrogen receptor |
| EST | expressed sequence tag |
| FasL | Fas ligand |
| FGF | fibroblast growth factor |
| fMLP | formyl-met-leu-phe |
| GADD153 | growth arrest and DNA damage inducible protein 153. See CHOP. |
| GADD45α | growth arrest and DNA damage inducible 45 alpha |
| GAK | cyclin G associated kinase |
| GM-CSF | granulocyte-macrophage colony-stimulating factor |
| Groα | growth-related oncogene protein-alpha |
| GSK | glycogen synthase kinase |
| HGF | hepatocyte growth factor |
| hnRNP | heterogeneous nuclear ribonucleoprotein |
| hsp | heat shock protein |
| HuR | human antigen R |
| ICAM1 | intercellular adhesion molecule 1 also known as CD54 |
| IFNγ | interferon gamma |
| Ig | immunoglobulin |
| IKK | IκB kinase |
| IL | interleukin |
| IP10 | IFNγ inducible protein 10 also known as CXCL10 |
| IRAK | interleukin-1 receptor-associated kinase |
| JIP | JNK interacting protein |
| KSR | kinase suppressor of Ras |
| Lck | leukocyte-specific protein tyrosine kinase |
| LPS | lipopolysaccharide |
| LTB | lymphotoxin beta |
| MCP | monocyte chemotactic protein |

| | |
|---|---|
| MEF2A/2C | myocyte-specific enhancer factor 2A and C |
| MIP | macrophage inflammatory protein |
| MLK3 | mixed lineage kinase-3 |
| MMP | metallomatrix protein |
| MP1 | MEK partner 1 |
| MRSA | methicillin-resistant *S. aureus* |
| MST1 | mammalian STE20-like protein kinase-1 |
| NFAT | nuclear factor of activated T cell |
| NFκB | nuclear factor kappa B |
| NGF | neuronal growth factor |
| NIK | NFκB-inducing kinase |
| iNOS | inducible nitric oxide synthase |
| NO | nitric oxide |
| p16-Arc | p16 subunit of the seven-member Arp2/3 complex |
| PABP1 | poly(A)-binding protein 1 |
| PDGF | platelet-derived growth factor |
| cPLA2 | cytosolic phospholipase A2 |
| PRAK | p38-regulated/activated protein kinase |
| PRK | protein kinase C-related protein kinase |
| QSAR | quantitative structure–activity relationship |
| RANKL | receptor activator of NFκB ligand |
| RICK | RIP-like interacting CLARP kinase also known as RIP2 or CARDIAK |
| RIP | receptor interacting protein |
| RSK | p90 ribosomal S6 kinase |
| SAP1A | signaling lymphocytic activation molecule-associated protein 1A |
| SAP90/97 | synapse-associated protein 90 and 97 |
| SRF | serum response factor |
| STMN1 | stathmin |
| STAT | signal transducer and activator of transcription |
| TAB1 | TAK1 binding protein 1 |
| TAK1 | transforming growth factor β-actiating kinase |
| TAO | thousand-and-one amino acids |
| TCR | T-cell receptor |
| TFIID | transcription factor II D |
| TGFβ | transforming growth factor-β |
| TIAR | T-cell-restricted intracellular antigen-related protein |
| Tollip | toll interacting protein |
| Tpl | tumor progression locus 2 also known as Cot |
| TRAF | TNF receptor-associated factor |
| TTP | tristetraprolin |
| TRPV1/VR-1 | transient receptor potential vanilloid 1 |
| VRE | vancomycin-resistant *Enterococci* |
| ZFP | zinc finger protein |

## ACKNOWLEDGMENTS

I gratefully acknowledge the support of the Wenner-Gren Research Foundation, and I extend sincere thanks to C. Elsner for assistance with the manuscript preparation.

## REFERENCES

1. Kracht M, Saklatvala J (2002). Transcriptional and post-transcriptional control of gene expression in inflammation. *Cytokine 20*, 91–106.
2. Wieseler-Frank J, Maier SF, Watkins LR (2005). Immune-to-brain communication dynamically modulates pain: Physiological and pathological consequences. *Brain Behav Immun 19*, 104–111.
3. Zhong J, Kyriakis JM (2007). Dissection of a signaling pathway by which pathogen-associated molecular patterns recruit the JNK and p38 MAPK and trigger cytokine release. *J Biol Chem 282*, 24246–24254.
4. Lasa M, Abraham SM, Boucheron C, Saklatvala J, Clark AR (2002). Dexamethasone causes sustained expression of MKP1 and phosphatase-mediated inhibition of MAPK p38. *Mol Cell Biol 22*, 7802–7811.
5. Caelles C, González-Sancho HM, Muñoz A (1997). Nuclear hormone receptor antagonism with AP-1 by inhibition of the JNK pathway. *Genes Dev 11*, 3351–3364.
6. Newton R (2000). Molecular mechanisms of glucocorticoid action: What is important? *Thorax 55*, 603–613.
7. Hunter T (1987). One thousand and one protein kinases. *Cell 50*, 823–839.
8. Venter J, et al. (2001). The sequence of the human genome. *Science 291*, 1304–1351.
9. Manning G, Whyte DB, Martinez R, Hunter T, Sudarsanam S (2002). The protein kinase complement of the human genome. *Science 298*, 1912–1934.
10. Raman M, Chen W, Cobb MH (2007). Differential regulation and properties of MAPK. *Oncogene 26*, 3100–3112.
11. Nishimoto S, Nishida E (2006). MAPK signalling: ERK5 versus ERK1/2. *EMBO Repts 7*, 782–786.
12. Coulombe P, Meloche S (2007). Atypical mitogen-activated protein kinases: Structure, regulation and functions. *Biochim Biophys Acta 1773*, 1376–1387.
13. Boulton TG, Cobb MH (1991). Identification of multiple extracellular signal-regulated kinases (ERKs) with anti-peptide antibodies. *Cell Regul 2*, 357–371.
14. Roux PP, Blenis J (2004). ERK and p38 MAPK-activated protein kinases: A family of protein kinases with diverse biological functions. *Microbiol Mol Biol Rev 68*, 320–344.
15. Lee JD, Ulevitch RJ, Han J (1995). Primary structure of BMK1: A new mammalian map kinase. *Biochem Biophys Res Commun 213*, 715–724.

16. Wang X, Tournier C (2006). Regulation of cellular functions by the ERK5 signalling pathway. *Cell Signal 18*, 753–760.
17. Bogoyevitch MA, Court NW (2004). Counting on mitogen-activated protein kinases—ERKs 3, 4, 5, 6, 7 and 8. *Cell Signal 16*, 1345–1354.
18. Kant S, Schumacher S, Singh MK, Kispert A, Kotlyarov A, Gaestel M (2006). Characterization of the atypical MAPK ERK4 and its activation of the MAPK-activated protein kinase MK5. *J Biol Chem 281*, 35511–35519.
19. Weston CR, Davis RJ (2007). The JNK signal transduction pathway. *Curr Opin Cell Biol 19*, 142–149.
20. Imajo M, Tsuchiya Y, Nishida E (2006). Regulatory mechanisms and functions of MAP kinase signaling pathways. *IUBMB Life 58*, 312.
21. Krens SR, Spaink HP, Snaar-Jagalska BE (2006). Functions of the MAPK family in vertebrate-development. *FEBS Lett 580*, 4984–4990.
22. Gerits N, Kostenko S, Moens U (2007). *In vivo* functions of mitogen-activated protein kinases: Conclusions from knock-in and knock-out mice. *Transgen Res 16*, 281–314.
23. Brewster JL, de Valoir T, Dwyer ND, Winter E, Gustin MC (1993). An osmosensing signal transduction pathway in yeast. *Science 259*, 1760–1763.
24. Shiozaki K, Russell P (1995). Cell-cycle control linked to extracellular environment by MAP kinase pathway in fission yeast. *Nature 378*, 739–743.
25. Berman K, McKay J, Avery L, Cobb M (2001). Isolation and characterization of pmk-(1-3): Three p38 homologs in *Caenorhabditis elegans*. *Mol Cell Biol Res Commun 4*, 337–344.
26. Lee JC, Laydon JT, McDonnell PC, Gallagher TF, Kumar S, Green D, McNulty D, Blumenthal MJ, Heys JR, Landvatter SW, et al. (1994). A protein kinase involved in the regulation of inflammatory cytokine biosynthesis. *Nature 372*, 739–746.
27. Han J, Lee JD, Tobias PS, Ulevitch RJ (1993). Endotoxin induces rapid protein tyrosine phosphorylation in 70Z/3 cells expressing CD14. *J Biol Chem 268*, 25009–25014.
28. Han J, Lee JD, Bibbs L, Ulevitch RJ (1994). A MAP kinase targeted by endotoxin and hyperosmolarity in mammalian cells. *Science 265*, 808–811.
29. Freshney NW, Rawlinson L, Guesdon F, Jones E, Cowley S, Hsuan J, Saklatvala J (1994). Interleukin-1 activates a novel protein kinase cascade that results in the phosphorylation of Hsp27. *Cell 78*, 1039–1044.
30. Cuenda A, Rouse J, Doza YN, Meier R, Cohen P, Gallagher TF, Young PR, Lee JC (1995). SB 203580 is a specific inhibitor of a MAP kinase homologue which is stimulated by cellular stresses and interleukin-1. *FEBS Lett 364*, 229–233.
31. Stokoe D, Engel K, Campbell DG, Cohen P, Gaestel M (1992). Identification of MAPKAP kinase 2 as a major enzyme responsible for the phosphorylation of the small mammalian heat shock proteins. *FEBS Lett 313*, 307–313.
32. Rouse J, Cohen P, Trigon S, Morange M, Alonso-Llamazares A, Zamanillo D, Hunt T, Nebreda AR (1994). A novel kinase cascade triggered by stress and heat shock that stimulates MAPKAP kinase-2 and phosphorylation of the small heat shock proteins. *Cell 78*, 1027–1037.
33. Jiang Y, Chen C, Li Z, Guo W, Gegner JA, Lin S, Han J (1996). Characterization of the structure and function of a new mitogen-activated protein kinase (p38beta). *J Biol Chem 271*, 17920–17926.
34. Kumar S, McDonnell PC, Gum RJ, Hand AT, Lee JC, Young PR (1997). Novel homologues of CSBP/p38 MAP kinase: Activation, substrate specificity and sensitivity to inhibition by pyridinyl imidazoles. *Biochem Biophys Res Commun 235*, 533–538.
35. Stein B, Yang MX, Young DB, Janknecht R, Hunter T, Murray BW, Barbosa MS (1997). p38–2, a novel mitogen-activated protein kinase with distinct properties. *J Biol Chem 272*, 19509–19517.
36. Lechner C, Zahalka MA, Glot JF, Møller NP, Ullrich A (1996). ERK6, a mitogen-activated protein kinase involved in C2C12 myoblast differentiation. *Proc Nat Acad Sci USA 93*, 4355–4359.
37. Li Z, Jiang Y, Ulevitch RJ, Han J (1996). The primary structure of p38gamma: A new member of p38 group of MAP kinases. *Biochem Biophys Res Commun 228*, 334–340.
38. Jiang Y, Gram H, Zhao M, New L, Gu J, Feng L, Di Padova F, Ulevitch RJ, Han J (1997). Characterization of the structure and function of the fourth member of p38 group mitogen-activated protein kinases, p38delta. *J Biol Chem 272*, 30122–30128.
39. Zervos AS, Faccio L, Gatto JP, Kyriakis JM, Brent R (1995). Mxi2, a mitogen-activated protein kinase that recognizes and phosphorylates Max protein. *Proc Nat Acad Sci USA 92*, 10531–10534.
40. Faccio L, Chen A, Fusco C, Martinotti S, Bonventre JV, Zervos AS (2000). Mxi2, a splice variant of p38 stress-activated kinase, is a distal nephron protein regulated with kidney ischemia. *Am J Physiol Cell Physiol 278*, C781–C790.
41. Sudo T, Yagasaki Y, Hama H, Watanabe N, Osada H (2002). Exip, a new alternative splicing variant of p38 alpha, can induce an earlier onset of apoptosis in HeLa cells. *Biochem Biophys Res Commun 291*, 838–843.
42. Yagasaki Y, Sudo T, Osada H (2004). Exip, a splicing variant of p38alpha, participates in interleukin-1 receptor proximal complex and downregulates NF-kappaB pathway. *FEBS Lett 575*, 136–140.
43. Hale KK, Trollinger D, Rihanek M, Manthey CL (1999). Differential expression and activation of p38 mitogen-activated protein kinase alpha, beta, gamma, and delta in inflammatory cell lineages. *J Immunol 162*, 4246–4252.

44. Kyriakis JM, Avruch J (2001). Mammalian mitogen-activated protein kinase signal transduction pathways activated by stress and inflammation. *Physiol Rev 81*, 807–869.
45. Zhong J, Kyriakis JM (2007). Dissection of a signaling pathway by which pathogen-associated molecular patterns recruit the JNK and p38 MAPK and trigger cytokine release. *J Biol Chem 282*, 24246–24254.
46. Ono K, Han J (2000). The p38 signal transduction pathway: Activation and function. *Cell Signal 12*, 1–13.
47. Raingeaud J, Gupta S, Rogers JS, Dickens M, Han J, Ulevitch RJ, Davis RJ (1995). Pro-inflammatory cytokines and environmental stress cause p38 mitogen-activated protein kinase activation by dual phosphorylation on tyrosine and threonine. *J Biol Chem 270*, 7420–7426.
48. Marinissen MJ, Chiariello M, Gutkind JS (2001). Regulation of gene expression by the small GTPase Rho through the ERK6 (p38 gamma) MAP kinase pathway. *Genes Dev 15*, 535–553.
49. Sakabe K, Teramoto H, Zohar M, Behbahani B, Miyazaki H, Chikumi H, Gutkind JS (2002). Potent transforming activity of the small GTP-binding protein Rit in NIH 3T3 cells: Evidence for a role of a p38gamma-dependent signaling pathway. *FEBS Lett 511*, 15–20.
50. Zhang S, Han J, Sells MA, Chernoff J, Knaus UG, Ulevitch RJ, Bokoch GM (1995). Rho family GTPases regulate p38 mitogen-activated protein kinase through the downstream mediator Pak1. *J Biol Chem 270*, 23934–23936.
51. Bagrodia S, Dérijard B, Davis RJ, Cerione RA (1995). Cdc42 and PAK-mediated signaling leads to Jun kinase and p38 mitogen-activated protein kinase activation. *J Biol Chem 270*, 27995–27998.
52. Zhong J, Kyriakis JM (2007). Dissection of a signaling pathway by which pathogen-associated molecular patterns recruit the JNK and p38 MAPK and trigger cytokine release. *J Biol Chem 282*, 24246–24254.
53. Zhao Q, Lee FS (1999). Mitogen-activated protein kinase/ERK kinase kinases 2 and 3 activate nuclear factor-kappaB through IkappaB kinase-alpha and IkappaB kinase-beta. *J Biol Chem 274*, 8355–8358.
54. Lin X, Cunningham ET Jr, Mu Y, Geleziunas R, Greene WC (1999). The proto-oncogene Cot kinase participates in CD3/CD28 induction of NF-kappaB acting through the NF-kappaB-inducing kinase and IkappaB kinases. *Immunity 10*, 271–280.
55. Holtmann H, Enninga J, Lakble S, Thiefes A, Dorrie A, Broemer M, Winzen R, Wilhelm A, Ninomiya-Tsuji J, Matsumoto K, Resch K, Kracht M (2001). The MAPK kinase kinase TAK1 plays a central role in coupling the interleukin-1 receptor to both transcriptional and RNA-targeted mechanisms of gene regulation. *J Biol Chem 276*, 3508–3516.
56. Cuenda A, Alonso G, Morrice N, Jones M, Meier R, Cohen P, Nebreda AR (1996). Purification and cDNA coning of SAPKK3, the major activator of RK/p38 in stress- and cytokine-stimulated monocytes and epithelial cells. *EMBO J 15*, 4156–4164.
57. Cuenda A, Cohen P, Buée-Scherrer V, Goedert M (1997). Activation of stress-activated protein kinase-3 (SAPK3) by cytokines and cellular stresses is mediated via SAPKK3 (MKK6); comparison of the specificities of SAPK3 and SAPK2 (RK/p38). *EMBO J 16*, 295–305.
58. Dong C, Davis RJ, Flavell RA (2002). MAP kinases in the immune response. *Annu Rev Immunol 20*, 55–72.
59. Rincón M, Pedraza-Alava G (2003). JNK and p38 MAP kinases in CD4+ and CD8+ T cells. *Immunol Rev 192*, 131–142.
60. Salvador JM, Mittelstadt PR, Belova GI, Fornace AJ Jr, Ashwell JD (2005). The autoimmune suppressor Gadd45alpha inhibits the T cell alternative p38 activation pathway. *Nat Immunol 6*, 396–402.
61. Salvador JM, Mittelstadt PR, Buszczynski T, Copeland TD, Yamaguchi H, Appella E, Fornace AJ Jr, Ashwell JD (2005). Alternative p38 activation pathway mediated by T cell receptor-proximal tyrosine kinases. *Nat Immunol 6*, 390–395 .
62. Rudd CE (2005). MAPK p38: Alternative and nonstressful in T cells. *Nat Immunol 6*, 368–370.
63. Ge B, Gram H, Di Padova F, Huang B, New L, Ulevitch, RJ, Luo Y, Han J (2002). MAPKK-independent activation of p38alpha mediated by TAB1-dependent autophosphorylation of p38alpha. *Science 295*, 1291–1294.
64. Kang YJ, Seit-Nebi A, Davis RJ, Han J (2006). Multiple activation mechanisms of p38alpha mitogen-activatd protein kinases. *J Biol Chem 281*, 26225–26234.
65. Chen W, Kirkbride KC, How T, Nelson CD, Mo J, Frederick JP, Wang XF, Lefkowitz RJ, Blobe GC (2000). Beta-arrestin 2 mediates endocytosis of type III TGF-beta receptor and down-regulation of its signaling. *Science 290*, 1574–1577.
66. Kelkar N, Gupta S, Dickens M, Davis RJ (2000). Interaction of a mitogen-activated protein kinase signaling module with the neuronal protein JIP3. *Mol Cell Biol 20*, 1030–1104.
67. Kortum RL, Lewis RE (2004). The molecular scaffold KSR1 regulates the proliferative and oncogenic potential of cells. *Mol Cell Biol 24*, 4407–4416.
68. Schaeffer HJ, Catling AD, Eblen ST, Collier LS, Krauss A, Weber MJ (1998). MP1: A MEK binding partner that enhances enzymatic activation of the MAP kinase cascade. *Science 281*, 1668–1671.
69. Keyse SM (2000). Protein phosphatases and the regulation of mitogen-activated protein kinase signalling. *Curr. Opin Cell Biol 12*, 186–192.
70. Abell AN, Granger DA, Johnson GL (2007). MEKK4 stimulation of p38 and JNK activity is negatively regulated by GSK3beta. *J Biol Chem 282*, 30476–30484.
71. Alonso A, Sasin J, Bottini N, Friedberg I, Osterman A, Godzik A, Hunter T, Dixon J, Mustelin T (2004) Protein tyrosine phosphatases in the human genome. *Cell 117*, 699–711.

72. Lang R, Hammer M, Mages J (2006). DUSP meet immunology: Dual specificity MAPK phosphatases in control of the inflammatory response. *J Immunol 177*, 7497–7504.
73. Masuda K, Shima H, Watanabe M, Kikuchi K (2001). MKP-7, a novel mitogen-activated protein kinase phosphatase, functions as a shuttle protein. *J Biol Chem 276*, 39002–39011.
74. Stewart AE, Dowd S, Keyse SM, McDonald NQ (1999). Crystal structure of the MAPK phosphatase Pyst1 catalytic domain and implications for regulated activation. *Nat Struct Biol 6*, 174–181.
75. Sun H, Charles CH, Lau LF, Tonks NK (1993). MKP1 (3CH134), an immediate early gene product, is a dual specificity phosphatase that dephosphorylates MAP kinase *in vivo*. *Cell 75*, 487–493.
76. Lang R, Hammer M, Mages J (2006). DUSP meets immunology: Dual specificity MAPK phosphatases in control of the inflammatory response. *J Immunol 177*, 7497–7504.
77. Wang X, Liu Y (2007). Regulation of innate immune response by MAP kinase phosphatase-1. *Cell Signal 19*, 1372–1382.
78. Zhao Q, Sheperd EG, Manson ME, Nelin LD, Sorokin A, Liu Y (2005). The role of mitogen-activated protein kinase phosphatase-1 in the response of alveolar macrophages to lipopolysaccharide: Attenuation of proinflammatory cytokine biosynthesis via feedback control of p38. *J Biol Chem 280*, 8101–8108.
79. Hammer M, Mages J, Dietrich H, Schmitz F, Striebel F, Murray PJ, Wagner H, Lang R (2005). Control of dual-specificity phosphatase-1 expression in activated macrophages by IL-10. *Eur J Immunol 35*, 2991–3001.
80. Liu C, Shi Y, Du Y, Ning X, Liu N, Huang D, Liang J, Xue Y, Fan D (2005). Dual-specificity phosphatase DUSP1 protects overactivation of hypoxia-inducible factor 1 through inactivating ERK MAPK. *Exp Cell Res 309*, 410–418.
81. Imasato A, Desbois-Mouthon C, Han J, Kai H, Cato AC, Akira S, Li JD (2002). Inhibition of p38 MAPK by glucocorticoids via induction of MAPK phosphatase-1 enhances nontypeable *Haemophilus influenzae*-induced expression of Toll-like receptor 2. *J Biol Chem 277*, 47444–47450.
82. Kassel O, Sancono A, Krätzschmar J, Kreft B, Stassen M, Cato AC (2001). Glucocorticoids inhibit MAP kinase via increased expression and decreased degradation of MKP1. *EMBO J 20*, 7108–7116.
83. Wang Z, Canagarajah BJ, Boehm JC, Kassisà S, Cobb MH, Young PR, Abdel-Meguid S, Adams JL, Goldsmith EJ (1998). Structural basis of inhibitor selectivity in MAP kinases. *Structure 6*, 1117–1128.
84. Wilson KP, McCaffrey PJ, Hsiao K, Pazhanisamy S, Galullo V, Bemis GW, Fitzgibbon MJ, Caron PR, Murcko MA, Su MS (1997). The structural basis for the specificity of pyridinylimidazole inhibitors of p38 MAP kinase. *Chem Biol 4*, 423–431.
85. Jiang Y, Li Z, Schwartz EM, Lin A, Guan K, Ulevitch RJ, Han J (1997). Structure-function studies of p38 mitogen-activated protein kinase. Loop 12 influences substrate specificity and autophosphorylation, but not upstream kinase selection. *J Biol Chem 272*(17), 11096–11102.
86. Enslen H, Brancho DM, Davis RJ (2000). Molecular determinants that mediate selective activation of p38 MAP kinase isoforms. *EMBO J 19*, 1301–1311.
87. Enslen H, Davis RJ (2001). Regulation of MAP kinases by docking domains. *Biol the Cell 93*, 5–14.
88. Sharrocks AD, Yang SH, Galanis A (2000). Docking domains and substrate-specificity determination for MAP kinases. *Trends Biochem Sci 25*, 448–453.
89. Tanoue T, Adachi M, Moriguchi T, Nishida E (2000). A conserved docking motif in MAP kinases common to substrates, activators and regulators. *Nat Cell Biol 1*, 110–116.
90. Smith JA, Poteet-Smith CE, Lannigan DA, Freed TA, Zoltoski AJ, Sturgill TW (2000). Creation of a stress-activated p90 ribosomal S6 kinase. The carboxyl-terminal tail of the MAPK-activated protein kinases dictates the signal transduction pathway in which they function. *J Biol Chem 275*, 31588–31593.
91. Davies SP, Reddy H, Caivano M, Cohen P (2000). Specificity and mechanism of action of some commonly used protein kinase inhibitors. *Biochem J 351*(Pt 1), 95–105.
92. Kuma Y, Sabio G, Bain J, Shpiro N, Márquez R, Cuenda A (2005). BIRB-796 inhibits all p38 MAPK isoforms *in vitro* and *in vivo*. *J Biol Chem 280*, 19472–19479.
93. Eyers PA, Craxton M, Morrice N, Cohen P, Goedert M (1998). Conversion of SB 203580-insensitive MAP kinase family members to drug-sensitive forms by a single amino-acid substitution. *Chem Biol 5*, 321–328.
94. Gum RJ, McLaughlin MM, Kumar S, Wang Z, Bower MJ, Lee JC, Adams JL, Livi GP, Goldsmith EJ, Young PR (1998). Acquisition of sensitivity of stress-activated protein kinases to the p38 inhibitor, SB 203580, by alteration of one or more amino acids within the ATP binding pocket. *J Biol Chem 273*, 15605–15610.
95. Tephly LA, Carter AB (2007). Differential expression and oxidation of MKP1 modulates TNF-alpha gene expression. *Am J Respir Cell Mol Biol 37*, 366–374.
96. Allen M, Svensson L, Roach M, Hambor J, McNeish J, Gabel CA (2000). Deficiency of the stress kinase p38alpha results in embryonic lethality: Characterization of the kinase dependence of stress responses of enzyme-deficient embryonic stem cells. *J Exp Med 191*, 859–870.
97. Cohen P (1997). The search for physiological substrates of MAP and SAP kinases in mammalian cells. *Trends Cell Biol 7*, 353–361.
98. Hasegawa M, Cuenda A, Spillantini MG, Thomas GM, Buée-Scherrer V, Cohen P, Goedert M (1999). Stress-activated protein kinase-3 interacts with the PDZ domain of alpha1-syntrophin. A mechanism for specific substrate recognition. *J Biol Chem 274*, 12626–12631.

99. Zarubin T, Han J (2005). Activation and signaling of the p38 MAP kinase pathway. *Cell Res 15*, 11–18.
100. Deak M, Clifton AD, Lucocq LM, Alessi DR (1998). Mitogen- and stress-activated protein kinase-1 (MSK1) is directly activated by MAPK and SAPK2/p38, and may mediate activation of CREB. *EMBO J 17*, 4426–4441.
101. Pierrat B, Correia JS, Mary JL, Tomás-Zuber M, Lesslauer W (1998). RSK-B, a novel ribosomal S6 kinase family member, is a CREB kinase under dominant control of p38alpha mitogen-activated protein kinase (p38alphaMAPK). *J Biol Chem 273*, 29661–29671.
102. Prichett W, Hand A, Sheilds J, Dunnington D (1995). Mechanism of action of bicyclic imidazoles defines a translational regulatory pathway for tumor necrosis factor alpha. *J Inflamm 45*, 97–105.
103. Kotlyarov A, Neininger A, Schubert C, Eckert R, Birchmeier C, Volk HD, Gaestel M (1999). MAPKAP kinase 2 is essential for LPS-induced TNF-alpha biosynthesis. *Nat Cell Biol 1*, 94–97.
104. Brook M, Sully G, Clark AR, Saklatvala J (2000). Regulation of tumour necrosis factor alpha mRNA stability by the mitogen-activated protein kinase p38 signalling cascade. *FEBS Lett 483*, 57–61.
105. Carballo E, Lai WS, Blackshear PJ (1998). Feedback inhibition of macrophage tumor necrosis factor-alpha production by tristetraprolin. *Science 281*, 1001–1005.
106. Chrestensen CA, Schroeder MJ, Shabanowitz J, Hunt DF, Pelo JW, Worthington MT, Sturgill TW (2004). MAPKAP kinase 2 phosphorylates tristetraprolin on *in vivo* sites including Ser178, a site required for 14-3-3 binding. *J Biol Chem 279*, 10176–10184.
107. Rutault K, Hazzalin CA, Mahadevan LC (2001). Combinations of ERK and p38 MAPK inhibitors ablate tumor necrosis factor-alpha (TNF-alpha ) mRNA induction. Evidence for selective destabilization of TNF-alpha transcripts. *J Biol Chem 276*, 6666–66674.
108. Frevel MA, Bakheet T, Silva AM, Hissong JG, Khabar KS, Williams BR (2003). p38 Mitogen-activated protein kinase-dependent and -independent signaling of mRNA stability of AU-rich element-containing transcripts. *Mol Cell Biol 23*, 425–436.
109. Kontoyiannis D, Pasparakis M, Pizarro TT, Cominelli F, Kollias G (1999). Impaired on/off regulation of TNF biosynthesis in mice lacking TNF AU-rich elements: Implications for joint and gut-associated immunopathologies. *Immunity 10*, 387–398.
110. Kontoyiannis D, Kotlyarov A, Carballo E, Alexopoulou L, Blackshear PJ, Gaestel M, Davis R, Flavell R, Kollias G (2001). Interleukin-10 targets p38 MAPK to modulate ARE-dependent TNF mRNA translation and limit intestinal pathology. *EMBO J 20*, 3760–3770.
111. Mahtani KR, Brook M, Dean JL, Sully G, Saklatvala J, Clark AR (2001). Mitogen-activated protein kinase p38 controls the expression and posttranslational modification of tristetraprolin, a regulator of tumor necrosis factor alpha mRNA stability. *Mol Cell Biol 21*, 6461–6469
112. Stoecklin G, Stubbs T, Kedersha N, Wax S, Rigby WF, Blackwell TK, Anderson P (2004). MK2-induced tristetraprolin:14-3-3 complexes prevent stress granule association and ARE-mRNA decay. *EMBO J 23*, 1313–1324.
113. Wiggin GR, Soloaga A, Foster JM, Murray-Tait V, Cohen P, Arthur JS (2002). MSK1 and MSK2 are required for the mitogen- and stress-induced phosphorylation of CREB and ATF1 in fibroblasts. *Mol Cell Biol 22*, 2871–2881.
114. Vermeulen L, De Wilde G, Van Damme P, Vanden Berghe W, Haegeman G (2003). Transcriptional activation of the NF-kappaB p65 subunit by mitogen- and stress-activated protein kinase-1 (MSK1). *EMBO J 22*, 1313–1324.
115. Saccani S, Pantano S, Natoli G (2002). p38-Dependent marking of inflammatory genes for increased NF-kappa B recruitment. *Nat Immunol 3*, 69–75.
116. Soloaga A, Thomson S, Wiggin GR, Rampersaund N, Dyson MH, Hazzalin CA, Mahadevan LC, Arthur JS (2003). MSK2 and MSK1 mediate the mitogen- and stress-induced phosphorylation of histone H3 and HMG-14. *EMBO J 22*, 2788–2797.
117. Clark AR, Dean JL, Saklatvala J (2003). Post-transcriptional regulation of gene expression by mitogen-activated protein kinase p38. *FEBS Lett 546*, 37–44.
118. Winzen R, Kracht M, Ritter B, Wilheim A, Chen CY, Shyu AB, Müller M, Gaestel M, Resch K, Holtmann H (1999). The p38 MAP kinase pathway signals for cytokine-induced mRNA stabilization via MAP kinase-activated protein kinase 2 and an AU-rich region-targeted mechanism. *EMBO J 18*, 4969–4980.
119. Lasa M, Mahtani KR, Finch A, Brewer G, Saklatvala J, Clark AR (2000). Regulation of cyclooxygenase 2 mRNA stability by the mitogen-activated protein kinase p38 signaling cascade. *Mol Cell Biol 20*, 4265–4274.
120. Lu G, Kang YJ, Han J, Herschman HR, Stefani E, Wang Y (2006). TAB-1 modulates intracellular localization of p38 MAP kinase and downstream signaling. *J Biol Chem 281*(9), 6087–6095.
121. Matsuyama W, Faure M, Yoshimura T (2003). Activation of discoidin domain receptor 1 facilitates the maturation of human monocyte-derived dendritic cells through the TNF receptor associated factor 6/TGF-beta-activated protein kinase 1 binding protein 1 beta/p38 alpha mitogen-activated protein kinase signaling cascade. *J Immunol 171*, 3520–3532.
122. Ohkusu-Tsukada K, Tominaga N, Udono H, Yui K (2004). Regulation of the maintenance of peripheral T-cell anergy by TAB1-mediated p38 alpha activation. *Mol Cell Biol 24*, 6957–6966.
123. Kim L, Del Rio L, Butcher BA, Mogensen TH, Paludan SR, Flavell RA, Denkers EY (2005). p38 MAPK autophosphorylation drives macrophage IL-12 production during intracellular infection. *J Immunol 174*, 4178–4184.
124. Morrison DK, Davis RJ (2003). Regulationof MAP kinase signalling modules by scaffold proteins in mammals. *Annu Rev Cell Dev Biol 19*, 91–118.

125. Allen M, Svensson L, Roach M, Hambor J, McNeish J, Gabel CA (2000). Deficiency of the stress kinase p38alpha results in embryonic lethality: Characterization of the kinase dependence of stress responses of enzyme-deficient embryonic stem cells. *J Exp Med 191*, 859–870.
126. Tamura K, Sudo T, Senftleben U, Dadak AM, Johnson R, Karin M (2000). Requirement for p38alpha in erythropoietin expression: A role for stress kinases in erythropoiesis. *Cell 102*, 221–231.
127. Adams RH, Porras A, Alonso G, Jones M, Vintersten K, Panelli S, Valladares A, Perez L, Klein R, Nebreda AR (2000). Essential role of p38alpha MAP kinase in placental but not embryonic cardiovascular development. *Mol Cell 6*, 109–116.
128. Mudgett JS, Ding J, Guh-Siesel L, Chartrain NA, Yang L, Gopal S, Shen MM (2000). Essential role for p38alpha mitogen-activated protein kinase in placental angiogenesis. *Proc Nat Acad Sci USA 97*, 10454–10459.
129. Johnson GV, Bailey CD (2003). The p38 MAP kinase signaling pathway in Alzheimer's disease. *Exp Neurol 183*, 263–268.
130. Hollenbach E, Neumann M, Vieth M, Roessner A, Malfertheiner P, Naumann M (2004). Inhibition of p38 MAP kinase- and RICK/NF-kappaB-signaling suppresses inflammatory bowel disease. *FASEB J 13*, 1550–1552.
131. Boyle DL, Jones TL, Hammaker D, Svensson CI, Rosengren S, Albani S, Sorkin L, Firestein GS (2006). Regulation of peripheral inflammation by spinal p38 MAP kinase in rats. *PLoS Med 3*, e338.
132. Takanami-Ohnishi Y, Amano S, Kimura S, Asada S, Utani A, Maruyama M, Osada H, Tsunoda H, Irukayama-Tomobe Y, Goto K, Karin M, Sudo T, Kasuya Y (2002). Essential role of p38 mitogen-activated protein kinase in contact hypersensitivity. *J Biol Chem 277*, 37896–37903.
133. Kirkwood KL, Li F, Rogers JE, Otremba J, Coatney DD, Kreider JM, D'Silva NJ, Chakravarty S, Digar S, Higgins LS, Protter AA, Medicheria S (2007). A p38alpha selective mitogen-activated protein kinase inhibitor prevents periodontal bone loss. *J Pharmacol Exp Ther 320*, 56–63.
134. Duan W, Wong WS (2006). Targeting MAPK for asthma. *Curr Drug Targets 7*, 691–698.
135. Renda T, Baraldo S, Pelaia G, Bazzan E, Turato G, Papi A, Maestrelli P, Maselli R, Vatrella A, Fabbri L, Zuin R, Marsico SA, Saetta M (2007). Increased activation of p38 MAPK in chronic obstructive pulmonary disease. *Eur Respir J 31*, 62–69.
136. Nick JA, Young SK, Brown KK, Avdi NJ, Arndt PG, Suratt BT, Janes MS, Henson PM, Worthen GS (2000). Role of p38 mitogen-activated protein kinase in a murine model of pulmonary inflammation. *J Immunol 164*, 2151–2159.
137. Li Z, Ma JY, Kerr I, Chakravarty S, Dugar S, Schreiner G, Protter AA (2006). Selective inhibition of p38alpha MAPK improves cardiac function and reduces myocardial apoptosis in rat model of myocardial injury. *Am J Physiol Heart Circulatory Physiol 291*, H1972–H1977.
138. Nguyen AN, Stebbins EG, Henson M, O'Young G, Choi SJ, Quon D, Damm D, Reddy M, Ma JY, Haghnazari E, Kapoun AM, Medicheria S, Protter A, Schreiner GF, Kurihara N, Anderson J, Roodman GD, Navas TA, Higgins LS (2006). Normalizing the bone marrow microenvironment with p38 inhibitor reduces multiple myeloma cell proliferation and adhesion and suppresses osteoclast formation. *Exp Cell Res 312*, 1909–1923.
139. Schindler JF, Monahan JB, Smith WG (2007). p38 pathway kinases as anti-inflammatory drug targets. *J Dent Res 86*, 800–811.
140. Kumar S, Boehm J, Lee JC (2003). p38 MAP kinases: Key signalling molecules as therapeutic targets for inflammatory diseases. *Nat Rev Drug Discov 2*, 717–726.
141. Badger AM, Bradbeer JN, Votta B, Lee JC, Adams JL, Griswold DE (1996). Pharmacological profile of SB 203580, a selective inhibitor of cytokine suppressive binding protein/p38 kinase, in animal models of arthritis, bone resorption, endotoxin shock and immune function. *J Pharmacol Exp Ther 279*, 1453–1456.
142. Badger AM, Griswold DE, Kapadia R, Blake S, Swift BA, Hoffman SJ, Stroup GB, Webb E, Rieman DJ, Gowen M, Boehm JC, Adams JL, Lee JC (2000). Disease-modifying activity of SB 242235, a selective inhibitor of p38 mitogen-activated protein kinase, in rat adjuvant-induced arthritis. *Arthr Rheum 43*, 175–183.
143. Wada Y, Nakajima-Yamada T, Yamada K, Tsuchida J, Yasumoto T, Shimozato T, Aoki K, Kimura T, Ushiyama S (2005). R-130823, a novel inhibitor of p38 MAPK, ameliorates hyperalgesia and swelling in arthritis models. *Eur J Pharmacol 506*, 285–295.
144. Beardmore VA, Hinton HJ, Eftychi C, Apostolaki M, Armaka M, Darragh J, McIlrath J, Carr JM, Armit LJ, Clacher C, Malone L, Kollias G, Arthur JS (2005). Generation and characterization of p38beta (MAPK11) gene-targeted mice. *Mol Cell Biol 25*, 10454–10464.
145. Aouadi M, Binetruy B, Caron L, Le Marchand-Brustel Y, Bost F (2006). Role of MAPK in development and differentiation: Lessons from knockout mice. *Biochimie 88*(9), 1091–1098.
146. O'Keefe SJ, Mudgett JS, Cupo S, Parsons JN, Chartrain NA, Fitzgerald C, Chen SL, Lowitz K, Rasa C, Visco D, Luell S, Carballo-Jan E, Owens K, Zaller DM (2007). Chemical genetics define the roles of p38{alpha} and p38 in acute and chronic inflammation. *J Biol Chem 282*(48), 34663–34667.
147. Taylor GA, Carballo E, Lee DM, Lai WS, Thompson MJ, Patel DD, Schenkman DI, Gilkeson GS, Broxmeyer HE, Haynes BF, Blackshear PJ (1996). A pathogenetic role for TNF alpha in the syndrome of cachexia, arthritis, and autoimmunity resulting from tristetraprolin (TTP) deficiency. *Immunity 4*, 445–454.
148. Meyer LH, Pap T (2005). MAPK signalling in rheumatoid joint destruction: Can we unravel the puzzle? *Arthr Res Ther 7*, 177–178.

149. Hegen M, Gaestel M, Nickerson-Nutter CL, Lin LL, Telliez JB (2006). MAPKAP kinase 2-deficient mice are resistant to collagen-induced arthritis. *J Immunol 177*, 1913–1917.

150. Zhao Q, Wang X, Nelin LD, Yao Y, Matta R, Manson ME, Baliga RS, Meng X, Smith CV, Bauer JA, Chang CH, Liu Y (2006). MAP kinase phosphatase 1 controls innate immune responses and suppresses endotoxic shock. *J Exp Med 203*, 131–140.

151. Hammer M, Mages J, Dietrich H, Servatius A, Howells N, Cato AC, Lang R (2006). Dual specificity phosphatase 1 (DUSP1) regulates a subset of LPS-induced genes and protects mice from lethal endotoxin shock. *J Exp Med 203*, 15–20.

152. Salojin KV, Owusu IB, Millerchip KA, Potter M, Platt KA, Oravecz T (2006). Essential role of MAPK phosphatase-1 in the negative control of innate immune responses. *J Immunol 176*, 1899–1907.

153. Chi H, Barry SP, Roth RJ, Wu JJ, Jones EA, Bennett AM, Flavell RA (2006). Dynamic regulation of pro- and anti-inflammatory cytokines by MAPK phosphatase 1 (MKP1) in innate immune responses. *Proc Nat Acad Sci USA 103*, 2274–2279.

154. Barnes PJ (1998). Chronic obstructive pulmonary disease: New opportunities for drug development. *Trends Pharmacol Sci 19*(10), 415–423.

155. MacNee W (2007). Pathogenesis of chronic obstructive pulmonary disease. *Clin Chest Med 28*(3), 479–513.

156. Scott PH, Paul A, Belham CM, Peacock AJ, Wadsworth RM, Gould GW, Welsh D, Plevin R (1998). Hypoxic stimulation of the stress-activated protein kinases in pulmonary artery fibroblasts. *Am J Respir Crit Care Med 158*(3), 958–956.

157. Tamura DY, Moore EE, Johnson JL, Zallen G, Aiboshi J, Silliman CC (1998). p38 mitogen-activated protein kinase inhibition attenuates intercellular adhesion molecule-1 up-regulation on human pulmonary microvascular endothelial cells. *Surgery 124*, 403–408.

158. Hashimoto S, Matsumoto K, Gon Y, Maruoka S, Kujime K, Hayashi S, Takeshia I, Horie T (2000). p38 MAP kinase regulates TNF alpha-, IL-1 alpha- and PAF-induced RANTES and GM-CSF production by human bronchial epithelial cells. *Clin Exp Allergy 30*(1), 48–55.

159. Escott KJ, Belvisi MG, Birrell MA, Webber SE, Foster ML, Sargent CA (2000). Effect of the p38 kinase inhibitor, SB 203580, on allergic airway inflammation in the rat. *Br J Pharmacol 131*, 173–176.

160. Duan W, Chan JH, McKay K, Crosby JR, Choo HH, Leung BP, Karras JG, Wong WS (2005). Inhaled p38alpha mitogen-activated protein kinase antisense oligonucleotide attenuates asthma in mice. *Am J Respir Crit Care Med 171*, 571–578.

161. Medicheria S, Fitzgerald M, Spicer D, Woodman P, Ma JY, Kapoun AM, Chakravarty S, Dugar S, Protter AA, Higgins LS (2007). p38{alpha} Selective MAP kinase inhibitor, SD-282, reduces inflammation in a sub-chronic model of tobacco smoke-induced airway inflammation. *J Pharmacol Exp Ther 324*, 921–999.

162. Smith SJ, Fenwick PS, Nicholson AG, Kirschenbaum F, Finney-Hayward TK, Higgins LS, Giembycz MA, Barnes PJ, Donnelly LE (2006). Inhibitory effect of p38 mitogen-activated protein kinase inhibitors on cytokine release from human macrophages. *Br J Pharmacol 149*, 393–404.

163. Haddad EB, Birrell M, McCluskie K, Ling A, Webber SE, Foster ML, Belvisi MG (2001). Role of p38 MAP kinase in LPS-induced airway inflammation in the rat. *Br J Pharmacol 132*, 1715–1724.

164. Wang X, Xu L, Wang H, Young PR, Gaestel M, Feuerstein GZ (2002). Mitogen-activated protein kinase-activated protein (MAPKAP) kinase 2 deficiency protects brain from ischemic injury in mice. *J Biol Chem 277*, 43968–43972.

165. Wang Y, Huang S, Sah VP, Ross J Jr., Brown JH, Han J, Chien KR (1998). Cardiac muscle cell hypertrophy and apoptosis induced by distinct members of the p38 mitogen-activated protein kinase family. *J Biol Chem 273*, 2161–2168.

166. Clark JE, Sarafraz N, Marber MS (2007). Potential of p38-MAPK inhibitors in the treatment of ischaemic heart disease. *Pharmacol Ther 116*, 192–206.

167. Craxton A, Shu G, Graves JD, Saklatvala J, Crebs EG, Clark EA (1998). p38 MAPK is required for CD40-induced gene expression and proliferation in B lymphocytes. *J Immunol 161*(7), 3225–3236.

168. Hideshima T, Akiyama M, Hayashi T, Richardson P, Schlossman R, Chauhan D, Anderson KC (2003). Targeting p38 MAPK inhibits multiple myeloma cell growth in the bone marrow milieu. *Blood 101*, 703–705.

169. Hideshima T, Podar K, Chauhan D, Ishitsuka K, Mitsiades C, Tai YT, Hamasaki M, Raje N, Hideshima H, Schreiner G, Nguyen AN, Navas T, Munshi NC, Richardson PG, Higgins LS, Anderson KC (2004). P38 MAPK inhibition enhances PS-341 (bortezomib)-induced cytotoxicity against multiple myeloma cells. *Oncogene 23*, 8766–8776.

170. Wang SW, Pawlowski J, Wathern ST, Kinney SD, Lichensten HS, Manthey CL (1999). Cytokine mRNA decay is accelerated by an inhibitor of p38-mitogen-activated protein kinase. *Inflamm Res 48*, 533–538.

171. Wei S, Kitaura H, Zhou P, Ross FP, Teitelbaum SL (2005). IL-1 mediates TNF-induced osteoclastogenesis. *J Clin Investig 115*, 282.

172. Matsumoto M, Sudo T, Saito T, Osada H, Tsujimoto M (2000). Involvement of p38 mitogen-activated protein kinase signalling pathway in osteoclastogenesis mediated by receptor activator of NF-kappa B ligand (RANKL). *J Biol Chem 275*, 31155–31161.

173. Li X, Udagawa N, Itoh K, Suda K, Murase Y, Nishihara T, Suda T, Takahashi N (2002). p38 MAPK-mediated signals are required for inducing osteoclast differentiation but not for osteoclast function. *Endocrinology 143*, 3105–3113.

174. Ji RR (2004). Peripheral and central mechanisms of inflammatory pain, with emphasis on MAP kinases. Inhibitors of p38 decrease neuropathic pain and inflammation by reducing synthesis of cytokines. *Curr Drug Targets Inflamm Allergy 3*(3), 299–303.
175. Ji RR, Strichartz G (2004). Cell signaling and the genesis of neuropathic pain. *Science STKE 2004*(252), reE1.
176. Kaminska B (2005). MAPK signalling pathways as molecular targets for anti-inflammatory therapy—from molecular mechanisms to therapeutic benefits. *Biochim Biophys Acta 1754*, 253–262.
177. Sung CS, Wen ZH, Change WK, Chan KH, Ho ST, Tsai SK, Change YC, Wong CS (2005). Inhibition of p38 mitogen-activated protein kinase attenuates interleukin-1beta-induced thermal hyperalgesia and inducible nitric oxide synthase expression in the spinal cord. *J Neurochem 94*, 742–752.
178. Sung CS, Wong CS (2007). Cellular mechanisms of neuroinflammatory pain: The role of interleukin-1beta. *Acta Anaesthesiol Taiwanica 45*, 103–109.
179. Ji RR, Samad TA, Jin SX, Schmoll R, Woolf CJ (2002). p38 MAPK activation by NGF in primary sensory neurons after inflammation increases TRPV1 levels and maintains heat hyperalgesia. *Neuron 36*, 57–68.
180. Meotti FC, Luiz AP, Pizzolatti MG, Kassuya CA, Calixto JB, Santos AR (2006). Analysis of the antinociceptive effect of the flavonoid myricitrin: Evidence for a role of the L-arginine-nitric oxide and protein kinase C pathways. *J Pharmacol Exp Ther 316*, 789–796.
181. Meotti FC, Posser T, Missau FC, Pizzolatti MG, Leal RB, Santos AR (2007). Involvement of p38MAPK on the antinociceptive action of myricitrin in mice. *Biochem Pharma 74*, 924–931.
182. Brack A, Rittner HL, Machelska H, Leder K, Mousa SA, Schäfer M, Stein C (2004). Control of inflammatory pain by chemokine-mediated recruitment of opioid-containing polymorphonuclear cells. *Pain 112*(3), 229–238.
183. Rittner HL, Labuz D, Schaefer M, Mousa SA, Schulz S, Schäfer M, Stein C, Brack A (2006). Pain control by CXCR2 ligands through $Ca^{2+}$-regulated release of opioid peptides from polymorphonuclear cells. *FASEB J 20*(14), 2627–2629.
184. Rittner HL, Labuz D, Richter JF, Brack A, Schäfer M, Stein C, Mousa C (2007). CXCR1/2 ligands induce p38 MAPK-dependent translocation and release of opioid peptides from primary granules *in vitro* and *in vivo*. *Brain Behav Immun 21*(8), 1021–1023.
185. Jin SX, Zhuang ZY, Woolf CJ, Ji RR (2003). p38 mitogen-activated protein kinase is activated after a spinal nerve ligation in spinal cord microglia and dorsal root ganglion neurons and contributes to the generation of neuropathic pain. *J Neurosci 23*, 4017–4022.
186. Schäfers M, Svensson CI, Sommer C, Sorkin LS (2003). Tumor necrosis factor-alpha induces mechanical allodynia after spinal nerve ligation by activation of p38 MAPK in primary sensory neurons. *J Neurosci 23*, 2517–2521.
187. Svensson CI, Marsala M, Westerlund A, Calcutt NA, Campana WM, Freshwater JD, Catalano R, Feng Y, Protter AA, Scott B, Yaksh TL (2003). Activation of p38 mitogen-activated protein kinase in spinal microglia is a critical link in inflammation-induced spinal pain processing. *J Neurochem 86*, 1534–1544.
188. Svensson CI, Hua XY, Protter AA, Powell HC, Yaksh TL (2003). Spinal p38 MAP kinase is necessary for NMDA-induced spinal PGE(2) release and thermal hyperalgesia. *Neuroreport 14*, 1153–1157.
189. Tong S (2004). SCIO-469, a novel P38A MAPK inhibitor, provides efficacy in acute post-surgical dental pain. *Clin Pharmacol Ther 75*, 3.
190. Watkins LR, Milligan ED, Maier SF (2003). Glial proinflammatory cytokines mediate exaggerated pain states: Implications for clinical pain. *Adv Exp Med Biol 521*, 1–21.
191. Adcock IM, Chung KF, Caramori G, Ito K (2006). Kinase inhibitors and airway inflammation. *Eur J Pharmacol 533*(1–3), 118–132.
192. Sebolt-Leopold JS, Herrera R (2004). Targeting the mitogen-activated protein kinase cascade to treat cancer. *Nat Rev Cancer 4*, 937–947.
193. Mikalsen T, Gerits N, Moens U (2006). Inhibitors of signal transduction protein kinases as targets for cancer therapy. *Biotechnol Annu Rev 12*, 154–223.
194. Stanjanovic I, Cuzzocrea S, Mangano K, Mazzon E, Miljkovic D, Wang M, Donia M, Al Abed Y, Kim J, Nicoletti F, Stosic-Grujicic S, Claesson M (2007). *In vitro*, *ex vivo* and *in vivo* immunopharmacological activities of the isoxazoline compound VGX-1027: Modulation of cytokine synthesis and prevention of both organ-specific and systemic autoimmune diseases in murine models. *Clin Immunol 123*, 311–323.
195. Miwatashi S, Arikawa Y, Kotani E, Miyamoto M, Naruo K, Kimura H, Tanaka T, Asahi S, Ohkawa S (2005). Novel inhibitor of p38 MAP kinase as an anti-TNF-alpha drug: Discovery of *N*-[4-[2-ethyl-4-(3-methylphenyl)-1,3-thiazol-5-yl]-2-pyridyl]benzamide (TAK-715) as a potent and orally active anti-rheumatoid arthritis agent. *J Med Chem 48*, 5966–5979.
196. Goldstein DM, Gabriel T (2005). Pathway to the clinic: Inhibition of P38 MAP kinase. A review of ten chemotypes selected for development. *Curr Top Med Chem 5*, 1017–1029.
197. Mercer BA, D'Armiento JM (2006). Emerging role of MAP kinase pathways as therapeutic targets in COPD. *J Chronic Obstruct Pulm Dis 1*, 137–150.
198. Lee MR, Dominguez C (2005). MAP kinase p38 inhibitors: Clinical results and an intimate look at their interactions with p38alpha protein. *Curr Med Chem 12*, 2979–2994.
199. Feige U, Hu, YL, Gasser J, Campagnuolo G, Munyakazi L, Bolon B (2000). Anti-interleukin-1 and anti-tumor necrosis factor-alpha synergistically inhibit adjuvant arthritis in Lewis rats. *Cell Mol Life Sci 57*, 1457–1470.

200. Tong L, Pav S, White DM, Rogers S, Crane KM, Cywin CL, Brown ML, Pargellis CA (1997). A highly specific inhibitor of human p38 MAP kinase binds in the ATP pocket. *Nat Struct Biol 4*, 311–316.

201. Fitzgerald CE, Patel SB, Becker JW, Cameron PM, Zaller D, Pikounis VB, O'Keefe SJ, Scapin G (2003). Structural basis for p38alpha MAP kinase quinazolinone and pyridol-pyrimidine inhibitor specificity. *Nat Struct Biol 10*(9), 764–769.

202. Pargellis C, Tong L, Churchill L, Cirillo PF, Gilmore T, Graham AG, Grob PM, Hickey ER, Moss N, Pav S, Regan J (2002). Inhibition of p38 MAP kinase by utilizing a novel allosteric binding site. *Nat Struct Biol 9*, 268–272.

203. Wrobleski ST, Doweyko AM (2005). Structural comparison of p38 inhibitor-protein complexes: A review of recent p38 inhibitors having unique binding interactions. *Curr Top Med Chem 5*, 1005–1016.

204. de Dios A, Shih C, López de Uralde B, Sánchez C, del Prado M, Martin Cabrejas LM, Pleite S, Blanco-Urgoiti J, Lorite MJ, Nevill CR Jr, Bonjouklian R, York J, Vieth M, Wang Y, Magnus N, Campbell RM, Anderson BD, McCann DJ, Biera DD, Lee PA, Schultz RM, Li LC, Johnson LM, Wolos JA (2005). Design of potent and selective 2-aminobenzimidazole-based p38alpha MAP kinase inhibitors with excellent in vivo efficacy. *J Med Chem 48*, 2270–2273.

205. Graneto MJ, Kurumbail RG, Vaszquez ML, Shieh HS, Pawlitz JL, Williams JM, Stallings WC, Geng L, Naraian AS, Koszyk FJ, Stealey MA, Xu XD, Weier RM, Hanson GJ, Mourey RJ, Compton RP, Mnich SJ, Anderson GD, Monahan JB, Devraj R (2007). Synthesis, crystal structure, and activity of pyrazole-based inhibitors of p38 kinase. *J Med Chem 50*, 5712–5719.

206. Kadam RU, Garg D, Paul AT, Bhutani KK, Roy N (2007). Evaluation of pro-inflammatory cytokine pathway inhibitors for p38 MAPK inhibitory potential. *J Med Chem 50*, 6337–6342.

207. Lin TH, Metzger A, Diller DJ, Desai M, Henderson I, Ahmed G, Kimble EF, Quadros E, Webb ML (2006). Discovery and characterization of triaminotriazine aniline amides as highly selective p38 kinase inhibitors. *J Pharmacol Exp Ther 318*, 495–502.

208. Moss N, Breitfelder S, Betageri R, Cirillo PF, Fadra T, Hickey ER, Kirrane T, Kroe RR, Madwed J, Nelson RM, Pargellis CM, Aian KC, Regan J, Swinamer A, Torcellini C (2007). New modifications to the area of pyrazole-naphthyl urea based p38 MAP kinase inhibitors that bind to the adenine/ATP site. *Bioorg Med Chem Lett 17*, 4242–4247.

209. Natarajan SR, Liu L, Levorse M, Thompson JE, O'Neill EA, O'Keefe SJ, Vora KA, Cvetovich R, Chung JY, Carballo-Jane E, Visco DM (2006). p38 MAP kinase inhibitors. Part 5: Discovery of an orally bio-available and highly efficacious compound based on the 7-amino-naphthyridone scaffold. *Bioorg Med Chem Lett 16*, 5468–5471.

210. Schreiber S, Feagan B, D'Haens G, Colombel JF, Geboes K, Yurcov M, Isakov V, Golovenko O, Bernstein CN, Ludwig D, Winter T, Meier U, Yong C, Steffgen J (2006). BIRB 796 Study Group Oral p38 mitogen-activated protein kinase inhibition with BIRB 796 for active Crohn's disease: A randomized, double-blind, placebo-controlled trial. *Clin Gastroenterol Hepatol 4*, 325–334.

211. Goldstein DM, Gabriel T (2005). Pathway to the clinic: Inhibition of P38 MAP kinase. A review of ten chemotypes selected for development. *Curr Top Med Chem 5*, 1017–1029.

212. Nikas SN, Drosos AA (2004). SCIO-469 Scios Inc. *Curr Opin Investig Drugs 5*, 1205–1212.

213. Haddad JJ (2001). VX-745. Vertex pharmaceuticals. *Curr Opin Investig Drugs 2*, 1070–1076.

214. Adams JL, Boehm JC, Kassis S, Gorycki PD, Webb EF, Hall R, Sorenson M, Lee JC, Ayrton A, Griswold DE, Gallagher TF (1998). Pyrimidinylimidazole inhibitors of CSBP/p38 kinase demonstrating decreased inhibition of hepatic cytochrome P450 enzymes. *Bioorg Med Chem Lett 8*, 3111–3116.

215. Brown KL, Cosseau C, Gardy JL, Hancock RE (2007). Complexities of targeting innate immunity to treat infection. *Trends Immunol 28*, 260–266.

216. Fabian MA, Biggs WH 3rd, Treiber DK, Atteridge CE, Azimioara MD, Benedetti MG, Carter TA, Ciceri P, Edeen PT, Floyd M, Ford JM, Galvin M, Gerlach JL, Grotzfeld RM, Herrgard S, Insko DE, Insko MA, Lai AG, Lélias JM, Mehta SA, Milanov ZV, Velasco AM, Wodicka LM, Patel HK, Zarrinkar PP, Lockhart DJ (2005). A small molecule-kinase interaction map for clinical kinase inhibitors. *Nat Biotechnol 23*, 329–336.

217. Liverton NJ, Butcher JW, Claiborne CF, Claremon DA, Libby BE, Nguyen KT, Pitzenberger SM, Selnick HG, et al. (1999). Design and synthesis of potent, selective, and orally bioavailable tetrasubstituted imidazole inhibitors of p38 mitogen-activated protein kinase. *J Med Chem 42*, 2180–2190.

218. Godl K, Wissing J, Kurtenbach A, Habenberger P, Blencke S, Butbrod H, Salassidis K, Stein-Gerlach M, Missio A, Cotton, M, Daub H (2003). An efficient proteomics method to identify the cellular targets of protein kinase inhibitors. *Proc Nat Acad Sci USA 100*, 15434–15439.

219. Adams JL, Boehm JC, Gallagher TF, Kassis S, Webb EF, Hall R, Sorenson M, Garigipati R, Griswold DE, Lee JC (2001). Pyrimidinylimidazole inhibitors of p38: Cyclic N-1 imidazole substituents enhance p38 kinase inhibition and oral activity. *Bioorg Med Chem Lett 11*, 2867.

220. Davies SP, Reddy H, Caivano M, Cohen P (2000). Specificity and mechanisms of action of some commonly used protein kinase inhibitors. *Biochem J 351*, 95–105.

221. Milella M, Estrov Z, Kornblau SM, Carter BZ, Konopleva M, Tari A, Schober WD, Harris D, Leysath CE, Lopez-Berestein G, Huang Z, Andreeff M (2002). Synergistic induction of apoptosis by simultaneous disruption of the Bcl-2 and MEK/MAPK pathways in acute myelogenous leukemia. *Blood 99*, 3461–3464.

222. Sebolt-Leopold JS, Herrera R (2004). Targeting the mitogen-activated protein kinase cascade to treat cancer. *Nat Rev Cancer 4*, 937–947.
223. Mikalsen T, Gerits N, Moens U (2006). Inhibitors of signal transduction protein kinases as targets for cancer therapy. *Biotechnol Annu Rev 12*, 154–223.
224. LoGrasso PV, Frantz B, Rolando AM, O'Keefe SJ, Hermes JD, O'Neill EA (1997). Kinetic mechanisms for p38 MAP kinase. *Biochemistry 36*, 10422–10427.
225. Frantz B, Klatt T, Pang M, Parsons J, Rolando A, Williams H, Tocci MJ, O'Keefe SJ, O'Neill EA (1998). The activation state of p38 MAPK determines the efficiency of ATP competition for pyridinylimidazole inhibitor binding. *Biochemistry 37*, 12846–12853.
226. Romerio NC, Albuquerque MG, de Alencastro RB, Ravi M, Hopfinger AJ (2005). Construction of 4D-QSAR models for use in the design of novel p38-MAPK inhibitors. *J Computer-Aided Mol Des 19*, 385.
227. Sperandio da Silva GM, Sant'Anna CM, Barreiro EJ (2004). A novel 3D-QSAR comparative molecular field analysis (CoMFA) model of imidazole and quinazolinone functionalized p38 MAP kinase inhibitors. *Bioorg Med Chem Lett 12*, 3159–3166.
228. Lee JC, Kumar S, Griswold DE, Underwood DC, Votta BJ, Adams JL (2000). Inhibition of p38 MAP kinase as a therapeutic strategy. *Immunopharmacology 47*, 185–201.
229. Boehm JC, Bower MJ, Gallagher TF, Kassis S, Johnson SR, Adams JL (2001). Phenoxypyrimidine inhibitors of p38alpha kinase: Synthesis and statistical evaluation of the p38 inhibitory potencies of a series of 1-(piperidin-4-yl)-4-(4-fluorophenyl)-5-(2-phenoxypyrimidin-4-yl) imidazoles. *Bioorg Med Chem Lett 11*, 1123–1126.
230. Jackson PF, Bullington JL (2002). Pyridinylimidazole based p38 MAP kinase inhibitors. *Curr Top Med Chem 2*, 1011–1020.
231. Peifer C, Wagner G, Laufer S (2006). New approaches to the treatment of inflammatory disorders small molecule inhibitors of p38 MAP kinase. *Curr Top Med Chem 6*, 113–149.
232. Saklatvala J (2004). The p38 MAP kinase pathway as a therapeutic target in inflammatory disease. *Curr Opin Pharmacol 4*, 372–377.
233. Inoue T, Boyle DL, Corr M, Hammaker D, Davis RJ, Flavell RA, Firestein GS (2006). Mitogen-activated protein kinase kinase 3 is a pivotal pathway regulating p38 activation in inflammatory arthritis. *Proc Nat Acad Sci USA 103*, 5484–5489.
234. Almholt DL, Loechel F, Nielsen SJ, Krog-Jensen C, Terry, R, Bjørn SP, Pedersen HC, Praestegaard M, Møller S, Heide M, Pagliaro L, Mason AJ, Butcher S, Dahl SW (2004). Nuclear export inhibitors and kinase inhibitors identified using a MAPK-activated protein kinase 2 redistribution screen. *Assay Drug Dev Technol 2*, 7–20.
235. Diskin R, Askari N, Capone R, Engelberg D, Livnah O (2004). Active mutants of the human p38alpha mitogen-activated protein kinase. *J Mol Biol 365*, 66–76.
236. Orth K, Palmer LE, Bao ZQ, Stewart S, Rudoph AE, Bliska JB, Dixon JE (1999). Inhibition of the mitogen-activated protein kinase kinase superfamily by a *Yersinia* effector. *Science 285*, 1920–1923.
237. Mukherjee S, Keitany G, Li Y, Wang Y, Ball HL, Goldsmith EJ, Orth K (2006). *Yersinia* YopJ acetylates and inhibits kinase activation by blocking phosphorylation. *Science 312*, 1211–1214.
238. Kim AL, Zhu Y, Zhu H, Han, L, Kopelovich L, Bickers DR, Athar M (2006). Resveratrol inhibits proliferation of human epidermoid carcinoma A431 cells by modulating MEK1 and AP-1 signaling pathways. *Exp Dermatol 15*, 538–546.
239. Chen KH, Weng MS, Lin JK (2007). Tangeretin suppresses IL-1beta-induced cyclooxygenase (COX)-2 expression through inhibition of p38 MAPK, JNK, and AKT activation in human lung carcinoma cells. *Biochem Pharmacol 73*, 215–227.
240. Lu Q, Qiu TQ, Yang H (2006). Ligustilide inhibits vacular smooth muscle cells proliferation. *Eur J Pharmacol 542*, 136–140.
241. Gayathri B, Manjula N, Vinaykumar KS, Lakshmi BS, Balakrishnan A (2007). Pure compound from *Boswellia serrata* extract exhibits anti-inflammatory property in human PBMCs and mouse macrophages through inhibition of TNFalpha, IL-1beta, NO and MAP kinases. *Int Immunopharmacol 4*, 473–482.
242. Malemud CJ (2007). Inhibitors of stress-activatd protein/mitogen-activated protein kinase pathways. *Curr Opin Pharmacol 7*, 339–343.
243. Lin TH, Metzger A, Diller DJ, Desai M, Henderson I, Ahmed G, Kimble EF, Quadros E, Webb ML (2006). Discovery and characterization of triaminotriazine aniline amides as highly selective p38 kinase inhibitors. *J Pharmacol Exp Ther 318*, 495–502.
244. Paul AT, Gohil VM, Bhutani KK (2006). Modulating TNF-alpha signalling with natural products. *Drug Discov Today 11*, 725–732.
245. Scott MG, Dullaghan E, Mookherjee N, Glavas N, Waldbrook M, Thompson A, Wang A, Lee K, Doria S, Hamill P, Yu JJ, Li Y, Donini O, Guarna MM, Finlay BB, North JR, Hancock RE (2007). An anti-infective peptide that selectively modulates the innate immune response. *Nat Biotechnol 25*, 465–472.
246. Korb A, Tohidast-Akrad M, Cetin E, Axmann R, Smolen J, Schett G (2006). Differential tissue expression and activation of p38 MAPK alpha, beta, gamma, and delta isoforms in rheumatoid arthritis. *Arthr Rheum 54*, 2745–2756.
247. Revesz L, Blum E, Di Padova FE, Buhl T, Feifel R, Gram H, Hiestand P, Manning U, Rucklin G (2004). Novel p38 inhibitors with potent oral efficacy in several models of rheumatoid arthritis. *Bioorg Med Chem Lett 14*, 3595–3599.
248. Askari N, Diskin R, Avitzour M, Capone R, Livnah O, Engelberg D (2007). Hyperactive variants of p38alpha induce, whereas hyperactive mutants of p38gamma sup-

press, activating protein 1-mediated transcription. *J Biol Chem 282*, 91–99.

249. Zanin-Ahorov A, Tal-Lapidot G, Cahalon L, Cohen-Sfady M, Pevsner-Fischer M, Lider O, Cohen IR (2007). T cells respond to lipopolysaccharide innately via TLR signaling. *J Immunol 179*, 41–44.
250. Ward SG, Parry RV, Matthews J, O'Neill L (1997). A p38 MAP kinase inhibitor SB203580 inhibits CD28-dependent T cell proliferation and IL-2 production. *Biochem Soc Trans 25*(2), 304S.
251. Sen J, Kapeller R, Fragoso R, Sen R, Zon LI, Burakoff SJ (1996). Intrathymic signals in thymocytes are mediated by p38 mitogen-activated protein kinase. *J Immunol 156*, 4535–4538.
252. Tanoue T, Yamamoto T, Maeda R, Nishida E (2001). A novel MAPK phosphatase MKP-7 acts preferentially on JNK/SAPK and p38 alpha and beta MAPKs. *J Biol Chem 276*, 26629–26639.
253. Masuda K, Shima H, Watanabe M, Kikuchi K (2001). MKP-7, a novel mitogen-activated protein kinase phosphatase, functions as a shuttle protein. *J Biol Chem 276*, 39002–39011.
254. Zhang Y, Blattman JN, Kennedy NJ, Duong J, Nguyen T, Wang Y, Davis RJ, Greenberg PD, Flavell RA, Dong C (2004). Regulation of innate and adaptive immune responses by MAP kinase phosphatase 5. *Nature 430*, 793–797.
255. Lu HT, Yang DD, Wysk M, Gatti E, Mellman I, Davis RJ, Flavell RA (1999). Defective IL-12 production in mitogen-activated protein (MAP) kinase kinase 3 (Mkk3)-deficient mice. *EMBO J 18*(7), 1845–1857.
256. Dambach DM (2005). Potential adverse effects associated with inhibition of p38alpha/beta MAP kinases. *Curr Top Med Chem 5*, 929–939.
257. Kolch W, Calder M, Gilbert D (2005). When kinases meet mathematics: The systems biology of MAPK signalling. *FEBS Lett 579*, 1891–1895.
258. Elion EA (2006). Methods for analyzing MAPK cascades. *Methods 40*, 207–208.
259. Goebeler M, Killan K, Gillitzer R, Kunz M, Yoshimura T, Bröcker EB, Rapp UR, Ludwig S (1999). The MKK6/p38 stress kinase cascade is critical for tumor necrosis fator-alpha-induced expression of monocyte-chemoattractant protein-1 in endothelial cells. *Blood 93*, 857–865.
260. Rouin BH, Wilmer WA, Danne M, Dickerson JA, Dixon CL, Lu L (1999). The mitogen-activated protein kinase p38 is necesssary for interleukin 1beta-induced monocyte chemoattractant protein 1 expression by human mesangial cells. *Cytokine 11*, 118–126.
261. Matsumoto K, Hashimoto S, Gon Y, Nakayama T, Horie T (1998). Pro-inflammatory cytokine-induced and chemical mediator-induced IL-8 expression in human bronchial epithelial cells through p38 mitogen-activated protein kinase-dependent pathway. *J Allergy Clin Immunol 101*, 825–831.
262. Vockerodt M, Pinkert D, Smola-Hess S, Michels A, Ransohoff RM, Tesch H, Kube D (2005). The Epstein-Barr virus oncoprotein latent membrane protein 1 induces expression of the chemokine IP-10: Importance of mRNA half-life regulation. *Int J Cancer 114*, 598–605.
263. Ayala JM, Goyal S, Liverton NJ, Claremon DA, O'Keefe SJ, Hanlon WA (2000). Serum-induced monocyte differentiation and monocyte chemotaxis are regulated by the p38 MAP kinase signal transduction pathway. *J Leukocyte Biol 67*, 869–875.
264. Hannigan MO, Zhan L, Ai Y, Kotlyarov A, Gaestel M, Huang CK (2001). Abnormal migration phenotype of mitogen-activated protein kinase-activated protein kinase 2–/– neutrophils in Zigmond chambers containing formyl-methionyl-leucyl-phenylalanine gradients. *J Immunol 167*, 3953–3961.
265. Kotlyarov A, Yannoni Y, Fritz S, Laass K, Telliez JB, Pitman D, Lin LL, Gaestel M (2002). Distinct cellular functions of MK2. *Mol Cell Biol 22*, 4827–2835.
266. Rousseau S, Dolado I, Beardmore V, Shpiro N, Marquez R, Nebreda AR, Arthur JS, Case LM, Tessier-Lavigne M, Gaestel M, Cuenda A, Cohen P (2006). CXCL12 and C5a trigger cell migration via a PAK1/2-p38alpha MAPK-MAPKAP-K2-HSP27 pathway. *Cell Signal 18*, 1897–1905.
267. Underwood DC, Osborn RR, Bochnowicz S, Webb EF, Rieman DJ, Lee JC, Romanic AM, Adams JL, Hay DW, Griswold DE (2000). SB 239063, a p38 MAPK inhibitor, reduces neutrophilia, inflammatory cytokines, MMP-9, and fibrosis in lung. *Am J Physiol Lung Cell Mol Physiol 279*, L895–902.
268. Graneto MJ, Kurumbail RG, Vaszquez ML, Shieh HS, Pawlitz JL, Williams JM, Stallings WC, Geng L, Naraian AS, Koszyk FJ, Stealey MA, Xu XD, Weier RM, Hanson GJ, Mourey RJ, Compton RP, Mnich SJ, Anderson GD, Monahan JB, Devraj R (2007). Synthesis, crystal structure, and activity of pyrazole-based inhibitors of p38 kinase. *J Med Chem 50*, 5712–5719.
269. Korb A, Tohidast-Akrad M, Cetin E, Axmann R, Smolen J, Schett G (2006). Differential tissue expression and activation of p38 MAPK alpha, beta, gamma, and delta isoforms in rheumatoid arthritis. *Arthr Rheum 54*, 2745–2756.
270. de Dios A, Shih C, López de Uralde B, Sánchez C, del Prado M, Martin Cabrejas LM, Pleite S, Blanco-Urgoiti J, Lorite MJ, Nevill CR Jr, Bonjouklian R, York J, Vieth M, Wang Y, Magnus N, Campbell RM, Anderson BD, McCann DJ, Biera DD, Lee PA, Schultz RM, Li LC, Johnson LM, Wolos JA (2005). Design of potent and selective 2-aminobenzimidazole-based p38alpha MAP kinase inhibitors with excellent in vivo efficacy. *J Med Chem 48*, 2270–2273.
271. Wang Z, Castresana MR, Newman WH (2004). Reactive oxygen species-sensitive p38 MAPK controls thrombin-induced migration of vascular smooth muscle cells. *J Mol Cell Cardiol 36*, 49–56.

272. Wang X, Xu L, Wang H, Young PR, Gaestel M, Feuerstein GZ (2002). Mitogen-activated protein kinase-activated protein (MAPKAP) kinase 2 deficiency protects brain from ischemic injury in mice. *J Biol Chem 277*, 43968–43972.

273. Rincón M, Enslen H, Raingeaud J, Recht M, Zapton T, Su MS, Penix LA, Davis RJ, Flavell RA (1998). Interferon-gamma expression by Th1 effector T cells mediated by the p38 MAP kinase signaling pathway. *EMBO J 17*, 2817–2829.

274. Lal AS, Clifton AD, Rouse J, Segal AW, Cohen P (1999). Activation of the neutrophil NADPH oxidase is inhibited by SB 203580, a specific inhibitor of SAPK2/p38. *Biochem Biophys Res Commun 259*, 465–470.

275. Lehner MD, Schwoebel F, Kotlyarov A, Leist M, Gaestel M, Hartung T (2002). Mitogen-activated protein kinase-activated protein kinase 2-deficient mice show increased susceptibility to *Listeria monocytogenes* infection. *J Immunol 168*, 4667–4673.

276. van den Blink B, Juffermans NP, ten Hove T, Schultz MJ, van Deventer SJ, van der Poll T, Pepelenbosch MP (2001). p38 mitogen-activated protein kinase inhibition increases cytokine release by macrophages in vitro and during infection in vivo. *J Immunol 166*, 582–587.

277. Platanias LC (2003). The p38 mitogen-activated protein kinase pathway and its role in interferon signaling. *Pharmacol Ther 98*, 129–142.

278. Lu HT, Yang DD, Wysk M, Gatti E, Mellman I, Davis RJ, Flavell RA (1999). Defective IL-12 production in mitogen-activated protein (MAP) kinase kinase 3 (Mkk3)-deficient mice. *EMBO J 18*(7), 1845–1857.

279. Tanaka N, Kamanaka M, Enslen H, Dong C, Wysk M, Davis RJ, Flavell RA (2002). Differential involvement of p38 mitogen-activated protein kinase kinases MKK3 and MKK6 in T-cell apoptosis. *EMBO Repts 3*, 785–791.

280. Wysk M, Yang DD, Lu HT, Flavell RA, Davis RJ (1999). Requirement of mitogen-activated protein kinase kinase 3 (MKK3) for tumor necrosis factor-induced cytokine expression. *Proc Nat Acad Sci USA 96*, 3763–3768.

281. Shi Y, Kotlyarov A, Laabeta K, Gruber AD, Butt E, Marcus K, Meyer HE, Friedrich A, Volk HD, Gaestel M (2003). Elimination of protein kinase MK5/PRAK activity by targeted homologous recombination. *Mol Cell Biol 23*, 7732–7741.

282. Faas MM, Moes H, Fijen JW, Kobold AC, Tulleken JE, Zijlstra JG (2002). Monocyte intracellular cytokine production during human endotoxaemia with or without a second in vitro LPS challenge: Effect of RWJ-67657, a p38 MAP-kinase inhibitor, on LPS-hyporesponsiveness. *Clin Exp Immunol 127*, 337–343.

283. Branger J, van den Blink B, Weijer S, Gupta A, van Deventer SJ, Hack CE, Peppelenbosch MP, van der Poll T (2003). Inhibition of coagulation, fibrinolysis, and endothelial cell activation by a p38 mitogen-activated protein kinase inhibitor during human endotoxemia. *Blood 101*, 4446–4448.

# CARDIOVASCULAR DISEASE: IN-DEPTH LOOK

Nicholas B. Norgard[1] and Mazen S. Abu-Fadel[2]

[1]*Department of Pharmacy, University of Oklahoma Health Sciences Center, Oklahoma City, Oklahoma*
[2]*Section of Cardiovascular Medicine, University of Oklahoma Health Sciences Center, Oklahoma City, Oklahoma*

## 1 INTRODUCTION

Cardiovascular disease has topped the list of killer diseases every year since 1900 (except during the influenza epidemic of 1918) [1]. According to the American Heart Association, one American dies from cardiovascular disease every 36 seconds and nearly 80 million Americans currently have some form of cardiovascular disease. These diseases are costing Americans more than $430 billion a year. However, remarkable progress has been made in the field of cardiology. The development of new drug treatments has helped cut death rates from cardiovascular disease by almost half since 1950. Drug discovery is vital in maintaining this momentum. There are currently over 250 drugs in development for treating heart disease and stroke.

There are many interesting stories in the development of cardiovascular drugs. In 1785, William Withering made the chance observation of the improvement in a patient with severe heart failure after ingesting an herbal remedy [2]. He went on to identify the active ingredient in this remedy to be extracts from the foxglove plant. Over the ensuing 9 years he carefully tried out different preparations from various parts of the plant and documented 156 cases where he described the effects and the best way of using it. He believed that digitalis had a diuretic effect in patients with a weak and irregular pulse. Withering had no real insight into the pharmacologic and physiologic effects of digitalis, and it was actually the courting of his wife that enticed him to advance his expertise in botany that allowed him to make such an observation that inspired the development of digoxin.

The discovery of warfarin is linked to an observation made in the early 1920s of cattle in the northern United States and Canada that would die of uncontrollable bleeding from very minor injuries [3]. It was determined that the cattle were ingesting moldy silage made from sweet clover that functioned as a potent anticoagulant by reducing the amount of functioning prothrombin. The anticoagulant substance in moldy sweet clover was determined to be the coumarin derivative, 4-hydroxycoumarin. Over the next few years, numerous similar chemicals were developed that had the same anticoagulant properties, which eventually resulted in the development of warfarin in 1948. Warfarin was first registered for use as a rodenticide in the United States in 1952 and in 1954 was approved for medical use in humans.

Today cardiovascular drug development depends less on chance observations and more on logical scientific approaches. Most cardiovascular drugs used today were developed using the process known as "forward pharmacology" [4]. This approach begins by taking a pathophysiological or physiological process and carefully examining the role of individual mediators in that process in a disease. This leads to the characterization of entities responsible for a biological activity and to the identification of specific drug targets such as a receptor or enzyme. This process can be inefficient. The concept of the rennin–angiotensin–aldosterone system (RAAS) was first hypothesized in the 1930s. It took almost

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

four decades before the therapeutic significance of angiotensin-converting enzymes (ACEs) became apparent and an inhibitor of the enzyme was developed. Nearly 50 years after the initial work had been done on the RAAS, captopril (Capoten) reached the market [5]. The development of beta-blockers has a similar story. Approximately 60 years after the characterization of the parasympathetic and sympathetic nervous systems, the beta-blockers became available for human use. In this day and age 50–60 years is not an acceptable amount of time to bring a product to market.

The approach to cardiovascular target identification has changed since the advent of the Human Genome Project. Use of the human genome has allowed for the formation of private and proprietary deoxyribonucleic acid (DNA) databases from which novel cardiovascular drug targets can be identified [6]. Our current pharmacological therapies are directed toward <500 human proteins, but thousands of potential targets are estimated to exist [7]. Exploring the human genome has resulted in the identification of numerous novel drug targets, some of which were identified before anything was known about their biological significance [6]. This is in absolute contrast to the forward pharmacology method used to develop ACE inhibitors and beta-blockers.

New drugs can be designed to take advantage of the deficiencies of the previously developed drugs in order to provide more efficacious and safer pharmacotherapeutic options. For instance, there may be drugs already available to clinicians that are known to affect a previously established drug target. The low-molecular-weight heparins (LMWHs), for example, were developed to take advantage of the limitations associated with unfractionated heparin (UFH). UFH use is accompanied with unreliable pharmacokinetics imposing the need for diligent monitoring and an anticoagulant effect that is modified by platelets, fibrin, vascular surfaces, and plasma proteins. LMWHs are derived through depolymerization of UFH to yield an average fragment size that is approximately one-third the molecular weight. Like UFH, LMWHs continued to produce their anticoagulant effect by activating antithrombin; however, the reduction in fragment size resulted in superior pharmacokinetics, a more predictable response, and more convenient dosing. There are also numerous examples of extended-release formulations and two drug combinations that have been developed from previously launched drugs that allow for more convenient dosing. Additionally, active metabolites derived from prodrugs have been formulated to increase efficacy and eliminate drug interactions. Although improving our current pharmacotherapeutic options is important, an emphasis should be to discover new therapeutic drug targets. The focus of this chapter is to explore current and novel pharmacological therapies in development for the treatment of cardiovascular diseases.

## 2 CORONARY ARTERY DISEASE

### 2.1 Antianginal Agents

Coronary artery disease (CAD) is the leading cause of death of adults in the United States. Death rates are two to three times higher in men than women; however, when compared to death rates prior to 1975, Cardiovascular (CV) death has fallen in most developed countries by 24–50%. Approximately one-half of this reduction is due to improvement in the risk factors of CAD, including reduction in cholesterol levels, blood pressure, smoking, and physical inactivity. The other half is due to improved secondary prevention measures after coronary events, better initial and long-term treatment of acute coronary syndrome, congestive heart failure, and better revascularization procedures [8].

In addition, in-hospital and long-term mortality rates following myocardial infarction have fallen substantially with improvement in medical therapy. This has been noted in the Atherosclerosis Risk In Communities (ARIC) study [9] as well as the National Registry of Myocardial Infarctions (NRMI) registry [10].

Even with improvements in medical therapy and revascularization, including coronary artery bypass surgery and percutaneous coronary intervention, there continues to be patients with angina refractory to all these treatments. Such patients are a challenge to treat and are the rational for novel therapeutic strategies. These strategies include but are not limited to inhibition of fatty acid oxidation, ischemic preconditioning, endothelin receptor blockers, and therapeutic angiogenesis.

The conventional treatment of CAD includes a wide variety of agents, including beta-blockers, nitrates, calcium channel blockers, antiplatelet agents, lipid-lowering drugs, and others. Discussing all these agents is beyond the scope of this chapter, but we will briefly touch on some.

***Beta-Blockers*** Beta-blockers are first-line therapy for angina. Their mechanism of action involves competitive inhibition of catecholamines from binding to the beta-adrenergic receptors. These receptors increase in density after chronic beta-blocker use and thus sudden withdrawal of the medications results in transient supersensitivity to catecholamines, which may precipitate angina, myocardial infraction (MI), or death.

The cardiovascular effects of beta-blockers are mediated by their ability to decrease myocardial oxygen demand. This is achieved by decreasing heart rate, con-

tractility, and left ventricular wall stress by dropping the systemic blood pressure. Cardioselective beta-blockers that mediate their effect by preferentially blocking the beta-1 receptor can decrease unwanted side effects but their selectivity is lost at high doses [11–13].

***Nitrates*** Nitroglycerin was the first medication used in 1879 by Williams Murrell for the treatment of angina pectoris. The mechanism of action works by relaxing smooth muscle cells, causing dilatation of veins of arteries including coronary arteries. This action is mediated via the formation of nitric oxide and subsequent activation of guanylate cyclase and the generation of cyclic guanosine monophosphate (cGMP). Their primary action is venodilatation, causing a decrease in left ventricular end diastolic pressure, resulting in decreased wall stress and decrease myocardial oxygen demand. The clinical effect of direct coronary vasodilation is uncertain, especially in patients with flow-limiting lesions of a coronary artery, which by itself causes maximal vasodilation during ischemia.

A major problem with nitrates is the development of tolerance after chronic usage. The mechanism of tolerance is due to at least the following:

- Impaired bioconversion and formation of nitric oxide
- Reduced bioactivity of nitric oxide
- Activation of the vasoconstrictor renin–angiotensin–aldosterone and sympathetic nervous system in response to nitrate-induced vasodilation

Nitrate tolerance can be presented by intermittent therapy with an adequate nitrate-free period [14–16].

***Statins*** The discovery of the key cholesterol synthesis enzyme 3-hydroxy-3-methylglutaryl coenzyme A reductase and subsequently its inhibitors (statins or HMG-CoA inhibitors) ranks as one of the most important advances of the twentieth century in the prevention of morbidity and death from atherosclerosis. Pleiotropic effects of statins may exist, but the effectiveness of statins in preventing major cardiovascular events in direct proportion to low-density lipoprotein (LDL) cholesterol reduction is well established.

Statins reduce LDL cholesterol from 15 to 60%; have variable effects on total high-density lipoprotein (HDL) cholesterol, raising it 0–15% and reducing triglycerides by 5–30%. Many disingenuous trials compare different formulations and starting doses, attempting to establish that a particular product is better at achieving National Cholesteral Eduction Program (NCEP) goals. Clinical practice should be based on those tested in clinical trials with clinical outcome data.

Rhabdomyolysis is seen in less than 1 : 1000 and when seen is usually in those with coexisting organ dysfunction or those on combination therapy with other hyperlipidemia drugs. Minor muscle pain may be more frequent. The supplement coenzyme Q10 has been used, but it remains of unproven benefit or safety. Transaminitis occurs in 0.1–2% of patients, typically in proportion to potency. Transaminase levels should be measured prior to therapy and then at 12 weeks. The long-term risk of transaminitis is unclear, but if levels remain >3 times upper limit of normal, then statin therapy should be withheld. Once levels return to normal, a different and/or lower dose statin may be tried.

Special care is required when using combination therapy of fibrates and statins. Definite need should be documented and then therapy started at low doses with avoidance of maximum doses of either agent. Future or ongoing trials may determine how low to go with LDL and whether other markers such as C-reactive protein should be measured to determine if maximal benefit is being obtained.

***Ranolazine*** Ranolazine recently became the first novel antianginal medication made available in over 20 years. The most important mechanism of action of ranolazine is decreasing diastolic tension by preventing calcium overload in ischemic myocytes. Its efficacy has been extensively tested especially in patients with chronic stable angina. It is effective for angina relief and significantly increases symptom-limited exercise duration and time to onset of angina. However, ranolazine failed to show any benefit when used in patients with acute coronary syndromes.

A dose-dependant QTc prolongation is an important adverse effect of this medication. However, torsades de pointes has not been reported with ranolazine and, in fact, ranolazine has been shown to have antiarrhythmic traits [17–19].

Multiple newer therapies for CAD and angina pectoris are being developed. Table 1 summarizes some of these medications and their mechanism of action [20].

Another novel therapeutic option for patients with end-stage angina and even peripheral arterial disease is angiogenesis. The goal of this therapy is the induction of new coronary arterial vessels that can effectively provide blood supply to an ischemic area of the myocardium. A number of angiogenic growth factors stimulate blood vessel growth. These include fibroblast growth factors (FGFs), vascular endothelial growth factors (VEGFs), and platelet-derived growth factors (PDEGF) [21].

The experience with these agents is still actively growing, and solutions to many problems are not completely resolved. These problems include patient

**TABLE 1 Novel Therapeutic Agents for the Treatment of Angina**

| Medication | Mechanism of Action |
|---|---|
| Trimetazidine and perihexiline | Inhibition of fatty acid oxidation that increases cardiac metabolic efficency |
| Nicorandil | Potassium channel activator that causes arterial and venous dilatation and mimics a natural process of ischemic preconditioning |
| Ivabradine | Inhibition of a specific sinus node current causing a decreased heart rate |
| Fasudil | Inhibits Rho kinase involvement in vascular smooth muscle contractile response |
| Testosterone | Improves endothelial dysfunction |

selection, mechanisim of delivery of the growth factors, and protein vs. gene therapy. In addition, there are a number of complications associated with this treatment, including but not limited to:

Angiogenesis in nontarget tissues
Increased vascular permeability
Uncontrolled neoplasm growth

### 2.2 Antiplatelet Agents

One of the main goals in CAD treatment is to prevent adverse outcomes from disease. Platelets play a central role in the development of atherosclerosis and clinical expression of thrombotic events. The use of antiplatelet agents has resulted in significant improvements in cardiovascular morbidity and mortality. Antiplatelet therapies, such as aspirin, clopidogrel, and the G-protein (GP) IIb/IIIa receptor inhibitors, are central to the treatment of patients with acute coronary syndrome (ACS). Despite their proven efficacy, many patients receiving these therapies continue to experience major coronary adverse events, or, on the other side of the spectrum, patients receiving these therapies experience major bleeding episodes that have severe consequences on morbidity and mortality. Thus, there is a need to investigate new antiplatelet agents that could potentially offer patients improved safety and efficacy. There are many novel antiplatelet targets in a wide range of developmental stages being investigated for therapeutic benefit in coronary artery disease.

Aspirin exerts its antiplatelet effects by inhibiting the cyclooxygenase-1 (COX-1) enzyme, which limits the production of the potent platelet agonist and vasoconstrictor, thromboxane ($TXA_2$). However, beneficial prostanoids are also produced via the COX-1 enzyme. Therefore, efforts to more specifically block $TXA_2$ may prove fruitful. Several $TXA_2$ receptor antagonists have entered clinical development in recent years. These include ifetroban, sulotroban, and GR32191. Despite the demonstration of antithrombotic effects of these agents in animal studies, phase II/III clinical trials have yielded disappointing results, which has forced development discontinuation of many of these agents. Terutroban is a $TXA_2$ receptor antagonist that exerts antiplatelet effects similar in magnitude to clopidogrel and superior to aspirin under low and high sheer conditions. This drug is now in phase III trials in stroke patients.

Dually acting agents that inhibit $TXA_2$ synthase and antagonize the $TXA_2$ receptor have shown promise in humans. These agents have been shown to reduce platelet aggregation and $TXA_2$ generation. Ridogrel failed to improve coronary flow in the setting of acute myocardial infarction compared to aspirin [22]. Terbogrel showed no benefit in the treatment of primary pulmonary hypertension [23]. Picotamide significantly reduced overall mortality compared to aspirin in diabetic patients with peripheral artery disease with a lower incidence of bleeding [24].

The P2Y12 receptor is predominantly located on platelets and promotes platelet activation through signal activation by adenosine diphosphate (ADP). Ticlopidine and clopidogrel represent the first- and second-generation P2Y12 receptor antagonists, respectively. Due to their limitations, however, other P2Y12 receptor antagonists, both oral and intravenous, are being evaluated for the treatment of patients with atherothrombotic disease. These include prasugrel, a thienopyridine; cangrelor, an adenosine triphosphate (ATP) analog; and ticagrelor, a cyclopentyl-triazolo-pyrimidine (CPTP), all of which have entered phase III development.

Prasugrel differs from the other irreversible thienopyridine P2Y12 receptor antagonists by having an ester group close to the reactive thiol group and a fluorine atom replacing the chlorine atom. Like clopidogrel, prasugrel is a prodrug that requires hepatic metabolism to form its active metabolite, which irreversibly inhibits the P2Y12 receptor [25]. However, the active metabolite of prasugrel is generated more efficiently and at higher concentrations compared with clopidogrel [26]. As a result prasugrel has shown to improve inhibition of ADP-induced platelet aggregation and fewer nonresponders, as measured by inhibition of platelet aggregation [27].

Reversible P2Y12 receptor antagonists, such as cangrelor and AZD6140, have chemical structures that differ substantially from those of the thienopyridines. Cangrelor is an ATP analog characterized by the pres-

ence of substituents in the 2-position of the adenine ring that increase the P2Y12 receptor affinity properties and β,γ-methylene substitutions in the triphosphate part of the molecule that increases the metabolic stability [28]. Cangrelor is administered intravenously and does not require metabolism for therapeutic activity, thus inhibiting platelet aggregation with rapid onset and offset of action, offering potential benefits over clopidogrel [28].

AZD6140 is an oral antagonist of the P2Y12 receptor that does not require metabolic activation [28]. The drug belongs to the same family as cangrelor of stable ATP analogs with high P2Y12 receptor affinity. AZD6140 is less potent than cangrelor but has a rapid onset of action (2–4 h) and a half-life of approximately 12 h [28]. As such, it is the first rapidly acting, reversible oral P2Y12 receptor antagonist in development for the treatment of ACS.

Thrombin is believed to be the most powerful platelet agonist known [29]. It plays a key role in coagulation and hemostasis partially through platelet activation via stimulation of protease-activated receptors (PAR) on the platelet surface. PAR1 and PAR4 are expressed on the surface of human platelets, with the PAR1 receptor proposed as the principal thrombin receptor [30]. Blocking PAR4 has little effect on inhibiting platelet aggregation [31]. Inhibiting the actions of thrombin on platelets by targeting PAR1 receptors is a potential antiplatelet therapy. At least two oral PAR1 receptor antagonists are currently in development: vorapaxar and atopaxar. Pepducins are cell-penetrating peptides that bypass the PAR receptors and modulate thrombin/PAR-induced intracellular signaling [32]. An anti-PAR4 pepducin has been shown to reduce platelet aggregation. This is a novel mechanism for inhibiting receptor function outside of the standard receptor antagonists and antibodies.

Phosphoinositide 3-kinase (PI3K) p110β is important in modulating intracellular calcium flux and GPIIb\IIIa bond stability [33]. PI3K p110β is predominantly required for sustained platelet aggregation under high shear conditions. TGX-221 is an inhibitor of PI3K p110β that has been shown, in animal studies, to inhibit platelet aggregation in vitro and in vivo under high sheer conditions [34]. Administration of TGX-221 caused no changes in bleeding times even at high doses [33].

The binding of von Willebrand factor (VWF) to subendothelial GPIb receptors is required for platelet blinding to injured vasculature [35]. Blockade of this interaction is a potential mechanism to inhibit platelet adhesion and subsequent platelet activation upstream of aggregation [35]. There are several approaches under investigation that target this interaction. A monoclonal antibody directed at GPIb successfully inhibited platelet aggregation and reduced thrombosis [36]. However, it also decreased platelet formation, causing an immediate and irreversible thrombocytopenia. There are two GPIb antagonists derived from snake venoms, crotalin and mamushigin. Crotalin was shown to decrease thrombus formation but was also shown to increase bleeding [37]. Mamushigin reduced aggregation at high sheer stress but stimulated aggregation at low sheer stress [38]. Rather than blocking the receptor another approach is to interfere with the interaction at the level of VWF. The VWF antibody, AJvW-2, was shown to inhibit platelet aggregation and thrombus formation in hamsters and guinea pigs without increasing bleeding times [39]. Aurin tricarboxylic acid binds to VWF, preventing the interaction between GPIb and VWF [40]. In a rabbit model, aurin tricarboxylic acid was more effective at inhibiting thrombus formation than a thromboxane synthase/thromboxane receptor blocker, a GPIIb/IIIa antagonist, and an antibody against tissue factor [40]. Finally, a recombinant VWF peptide, VCL, is able to block the GPIb–VWF interaction [41, 42]. Recombinant vWF fragment Leu504-Ser728 (VCL) does not directly antagonize GPIb or VWF but rather saturates the GPIb receptor to block the interaction. The peptide was shown to reduce aggregation and thrombus formation in primate and guinea pig models.

Leptin is increased in obesity and has been shown to enhance platelet aggregation through the effect of leptin receptors on the platelet surface [43]. Obese individuals may benefit from a leptin receptor antagonist.

The Arp 2/3 complex regulates the shape change platelets undergo upon activation [44]. The shape change is characterized by the formation of filopodia, lamelipodia, stress fibers, and contractile rings. This process is necessary for stable platelet aggregation to occur. Pretreatment with recombinant Arp was shown to arrest the platelets in the early rounded stage of activation while keeping other critical platelet functions intact [44].

Adenosine triphosphate diphosphohydrase (ATPase) or CD39 (Cluster of Differentiation 39) is an enzyme that does not act directly on the platelet but is a dominant contributor to the hydrolysis of the platelet agonists ADP and ATP [45, 46]. The enzyme prevents prothrombotic platelet activation and mitigates excessive platelet recruitment to a growing thrombus. Due to its enzymatic activity, a recombinant soluble CD39 was able to inhibit platelet aggregation in mice. Also using a recombinant ATPase in mice altered to have 100-fold increased ATPase activity, platelet aggregation was reduced and existing platelet aggregates were dissolved.

The CD40/CD40 ligand system is a transmembrane protein that plays an important role in inflammation [47]. The CD40 ligand is shed as a result of platelet

activation and increases production of chemokines, E-selection, vascular cell adhesion protein (VCAM), and intercellular adhesion molecule (ICAM) triggering leukocyte recruitment. Decreasing CD40 ligand shedding or blocking ligand–receptor interaction represents a novel therapeutic approach.

RANTES (Regulated upon Activation, Normal T Cell Expressed and Presumably Secreted) is a chemokine that interacts with P-selectin and mediates monocyte/macrophage infiltration into atherosclerotic lesions [48, 49]. The peptide antagonist Met-RANTES reduced macrophage infiltration in areas of atherosclerosis and decreased neointima formation in mice.

The development of a platelet plug begins when platelets adhere to the site of vascular injury via interactions with subendothelial collagen. Platelets interact with collagen either directly, through glycoprotein VI (GPVI), or indirectly, through VWF. C1q-TNF-related protein-1 (CTRP-1) is an endogenous inhibitor of collagen-induced platelet aggregation [50]. CTRP-1 blocks collagen-dependent platelet function predominantly through the disruption of the interaction between VWF and collagen. CTRP-1 has been shown to prevent platelet thrombosis in vivo in nonhuman primates. CTRP-1 prevents collagen-induced platelet aggregation by specific blockade of VWF binding to collagen.

### 2.3 Conclusion

The biology of coronary artery disease is becoming very advanced. Although many of the targets are rudimentary in development, the number of potential treatment options for the disease is substantial. A topic not focused on in this chapter but that warrants mention is the area of CAD prevention. Therapies targeted at obesity, diabetes, hypertension, and metabolic syndrome are important in preventing the onset of CAD.

## 3 HEART FAILURE

Heart failure is a complex clinical syndrome that results from any structural or functional cardiac disorder that impairs the ability of the heart to fill with or pump a sufficient amount of blood through the body [51]. The prevalence of the disease is reaching epidemic proportions with approximately 5.2 million Americans living with heart failure (HF) [1]. At the age of 40, the lifetime risk of developing HF is 1 in 5. In the last decade there has been a dramatic increase in HF death rate. It is estimated that 80% of men and 70% of women under age 65 who have HF will die within 8 years of diagnosis. HF also imposes a substantial economic burden with an estimated direct and indirect cost of $33.2 billion in the United States for 2007 [1].

Over the past 2 decades, investigators have learned more about the pathophysiologic changes that occur in systolic and diastolic ventricular dysfunction. Expansion of our knowledge of heart failure has increased the gamut of possible targets in HF therapy. However, in comparison with coronary artery disease, the development of HF pharmacotherapeutic options is in its infancy. Although advances have been made in our understanding of the pathogenesis and management of this complex disorder, we are still lacking substantial information regarding fundamental abnormalities, particularly at the cellular level.

The treatment of HF focuses on treating the symptoms and signs of HF and preventing the progression of disease [51]. Drugs that interfere with the excessive activation of RAAS and the adrenergic nervous system improve HF with a depressed ejection fraction (EF) by stabilizing and/or reversing cardiac remodeling. In this regard ACE inhibitors (ACEIs), angiotensin receptor blockers (ARB), and beta-blockers have emerged as cornerstones of modern HF therapy for patients with a reduced EF [51]. Numerous large-scale clinical trials with a wide range of causes and severity of HF have established their importance, and the evidence for the first-line use of these agents is compelling. The advent of neurohormonal active therapies has been a major advance in chronic HF treatment and prevention. However, in the full scope of HF there remain true unmet needs. Currently, there is a paucity of controlled clinical trials concerning the optimal treatment of HF with preserved EF and acute decompensated HF.

### 3.1 Cardiac Remodeling

Heart failure should be viewed as a continuum of developmental stages [51] (Table 2).

A major focus of patients in stages A and B is preventing the progression to stage C or clinical HF. Left ventricular remodeling is a pivotal element is the development of heart failure. Progressive remodeling results in HF and its extent is associated with clinical outcome [52, 53]. A range of cellular and biochemical processes encompass the remodeling process, and each of these processes is a potential therapeutic target.

Cardiac hypertrophy develops in response to elevated wall stress that eventually progresses to failure. At the cellular level multiple signaling pathways interact to drive hypertrophy. Two well-known drivers of cardiac hypertrophy are angiotension II and norepinephrine, which accounts for the utility of RAAS antagonists and beta-blockers in the attenuation of ventricular hypertrophy. Many receptor-activated calcium-dependent and mitogen-activated protein (MAP) kinase-dependent signaling pathways converge toward

**TABLE 2 Stages of Heart Failure**

| HF Stages | Description of Stage | Example of Patient in Stage |
|---|---|---|
| Stage A | High risk for developing HF, but without structural heart disease or symptoms of HF | Patients with diabetes or hypertension |
| Stage B | Structural heart disease but without symptoms of HF | Patients with a previous myocardial infarction (MI) and asymptomatic left ventricular (LV) dysfunction |
| Stage C | Structural heart disease with symptoms of HF | Patients with a previous MI with shortness of breath and fatigue |
| Stage D | Patients with refractory HF | Patients who are awaiting cardiac transplantation |

a final common hypertrophic molecular program regulated by histone deacetylases (HDACs) [54]. Inhibition of HDACs has been shown to attenuate cardiac hypertrophy in animals.

Likewise, Rho-kinase has been identified as one of the effectors of the small guanosine triphosphate (GTP)-binding protein Rho [55]. Accumulating evidence has demonstrated that the Rho/Rho-kinase pathway plays an important role in various cellular functions [55]. Including the pathogenesis of cardiovascular remodeling and hypertrophic development. Fasudil, a clinically available Rho-knase inhibitor, has been shown to inhibit left ventricular hypertrophy after myocardial infarction in mice [56]. There are many other small-molecule inhibitors available that affect a number of hypertrophic pathways. For instance, cyclosporine and rapamycin have been shown to attenuate ventricular hypertrophy [57–60]. However, their lack of cardioselectivity and associated toxicities limits the utility of these agents [61].

A recombinant protein known as periostin has been shown to induce cardiomyocyte proliferation following myocardial infarction, resulting in improved cardiac function. In a rat model of myocardial infarction, this protein, normally expressed during cardiac development, was bound to a biodegradeable extracellular matrix and injected epicardially at the time of injury [62]. Periostin-injected rats had significantly higher proportion of cycling cardiomyocytes and cardiomyocyte nuclei, which translated into improved ventricular remodeling, better myocardial function, and reduced infarct size [62].

### 3.2 Inflammation Targets

Heart failure has been associated with an increase in a range of inflammatory cytokines such as tumor necrosis factor-α (TNF-α) and various interleukins [63]. The potential utility of TNF-α antagonists is based on a correlation between HF severity and plasma TNF-α and the finding that overexpression of cardiac-specific TNF-α results in severe dilated cardiomyopathy [64]. However, the TNF-α antagonist, enterecept, offered negative results in HF patients [65]. Several mechanisms have been proposed for this negative outcome. It may be that accumulation of inflammatory markers is a consequence of HF, but disrupting the inflammatory cascade does not influence the HF process. Reduced renal clearance associated with HF may prevent elimination of the cytokines. Alternatively, these targets may be intimately involved in HF pathophysiology but is much more complex than currently understood. Despite previous failures, there remains a considerable amount of interest in immune modulation as a therapeutic strategy in HF. For instance, intravenous immunoglobin (IVIG), which is used in a number of inflammatory diseases, has been shown to significantly increase ejection fraction in patients with heart failure [66, 67].

### 3.3 Contractile Targets

Failing myocardium is linked with a reduced myocardial contractility triggered by defective handling of intracellular calcium [68]. Key components of intracellular calcium dysregulation include protein kinase A-mediated hyperphosphorylation of the ryanodine receptor, which causes diastolic calcium leak and reduced expression of the sarcoplasmic reticulum $Ca^{2+}$ ATPase (SERCA), which impairs calcium reuptake by the sarcoplasmic reticulum [69]. The ryanodine receptor stabilizer, JTV519, inhibited diastolic calcium leakage and improved contractility [70]. However, intracellular calcium overload can lead to electrical instability putting patients at risk for arrhythmias. Rather than increasing intracellular calcium supplies, another approach is to use the intracellular calcium supplies more efficiently. Levosimendan is classified as a "calcium sensitizer" that stabilizes the molecular conformation of troponin-C when it binds calcium. This improves contractility and has been shown to improve hemodynamics in acute decompensated heart failure [71]. Cardiac myosin activators have also been developed to increase contractility without changing intracellular calcium levels [72].

Another potential inotropic target is the Na/K-ATPase enzyme. Inhibition of Na/K-ATPase has been shown to augment myocardial contractility by

stimulating calcium entry via the sarcolemmal Na/Ca exchanger [73]. Istaroxime is a Na/K-ATPase inhibitor chemically unrelated to digoxin [74]. Preliminary data suggests that istaroxime permits intracellular calcium accumulation while avoiding proarrhythmic consequences. The mechanism behind this is unclear.

### 3.4 Metabolic Modulation

Increasing metabolic efficiency has become another potential target in HF therapy. Metabolism of fatty acids is the main source of ATP production in the heart [75]. Fatty acids are not as efficient as glucose as an energy source requiring more oxygen to produce an equivalent amount of ATP [76]. The heart seems to pump more efficiently if the fuel is glucose. There is evidence that enhancement of glucose oxidation in heart failure increases cardiac contractility and that reliance on fatty acid oxidation in heart failure may contribute to systolic dysfunction. Cardiac metabolic agents switch metabolism from fatty acid oxidation to glucose oxidation for more efficient ATP generation. Perhexiline, a carnitine palmitoyl transferase-1 (CPT-1) inhibitor, and trimetazidine, an inhibitor of mitochondrial fatty acid B-oxidation, long-chain 3-ketoacyl coenzyme A thiolase, are two of a number of drugs that may act in this way. Both drugs have shown favorable effects in HF thus far [77]. In patients with HF, perhexiline improved peak exercise oxygen consumption, left ventricular ejection fraction, HF symptoms, resting and peak stress myocardial function, and skeletal muscle energetics [78]. Treatment with trimetazidine, for 18 months, was associated with improvements in New York Heart Association (NYHA) functional class, left ventricular ejection fraction, and other indices of cardiac remodeling [79].

### 3.5 Neurohormonal Targets

Activation of endogenous neurohormonal systems plays a major role in the progression of HF. Neurohormonal factors such as angiotensin II, norepinephrine, aldosterone, endothelin, and vasopressin adversely affect the structure and function of the heart. They work in concert to increase hemodynamic stress by causing sodium retention and peripheral vasoconstriction and exert direct toxic effects on the myocardium. In HF, there are many successful treatment options that antagonize neurohormonal pathways and our growing knowledge of these pathways has produced a number of potential therapeutic targets.

The RAAS plays an important role in the pathophysiology of HF, and inhibition of the RAAS has become a target for several cardiovascular drugs [80]. ACEIs are the best studied drugs in this class [81, 82]. However, concern regarding incomplete inhibition of the RAAS with ACEIs has led to the use of newer agents such as ACEIs and aldosterone antagonists that when added to ACEIs provide more complete inhibition of the RAAS.

Renin blockade may provide superior neurohormonal inhibition by blocking the system upstream rather than downstream like ACE inhibitors and ARB [83]. Upstream blockade avoids reflex activation of renin, angiotensin I, and angiotensin II; however, this may also diminish the potential beneficial effects seen from angiotensin II activation of the angiotensin subtype $AT_2$ receptor [84]. The direct renin inhibitor, aliskiren, is approved in the United States for the treatment of hypertension and will be extensively studied in the HF population [83]. There are also other renin inhibitors currently under development.

Vasopressin is critical in the regulation of fluid balance, vascular tone, and may play a role in myocardial remodeling [85]. The vasopressin system comprises $V1_A$ receptors, which primarily mediate vasoconstriction, and $V_2$ receptors, which promote free water reabsorption [86]. Vasopressin antagonists, tolvaptan and conivaptan, are undergoing clinical development for the treatment of HF. Tolvaptan is a selective $V_2$ antagonist that significantly promoted diuresis but thus far has not been shown to reduce major HF-endpoints such as cardiovascular mortality or HF-related hospitalizations [87, 88]. In addition to tolvaptan, other selective $V_2$ antagonists are under development [85, 89]. Conivaptan provides simultaneous blockade of $V1_A$ and $V_2$ receptors and has been shown to improve select parameters of cardiac performance [85, 90, 91]. Nonselective vasopressin blockade is theoretically advantageous; however, it is unclear whether it will confer greater clinical benefit.

Urodilatin is an endogenous peptide released by the kidney in response to atrial stretch that regulates renal sodium and water excretion [92]. Its syntheic equivalent, ularitide, demonstrated enhanced diuresis and hemodynamic improvements in acute decompensated HF patients [93].

Adrenomedullin is an endogenous vasorelaxing and natriuretic peptide that plays an important role in cardiorenal regulation [94]. Based on its vasodilatory and antifibrotic actions, it is thought that adrenomodullin may potentially have a therapeutic effect on heart failure and renal failure. The administration of adrenomedullin has been shown to result in hemodynamic and renal benefits when added to traditional therapies [95, 96].

The peptide, urotensin II, is regarded as the most potent vasoconstrictor described and may be contributory to increased vascular tone in HF patients [97].

Increased gene expression of the ligand and the receptor has been observed in failing myocardium of HF patients [98]. Urotensin II receptor blockers have been developed and may be a potential therapeutic option in HF [99].

### 3.6 Conclusion

The potential therapies listed in this chapter are not all-inclusive, illustrating that great progress is being made in the identification of novel pharmacological targets in HF. However, the pace that therapies move toward clinical use is slow. It has been estimated that the cost of conducting a large-scale HF trial ranges between $102 and $207 million, and recent development failures have discouraged the industry from investing large amounts of money to develop an HF drug in a broad population [100].

**TABLE 3 Vaughan Williams Classification**

| Class | Drugs | Mechanism of Action |
|---|---|---|
| Ia | Quinidine<br>Procainamide<br>Disopyramide | Block $Na^+$ channels<br>Intermediate onset/offset kinetics |
| Ib | Lidocaine<br>Phenytoin<br>Mexilitine | Block $Na^+$ channels<br>Rapid onset/offset kinetics |
| Ic | Flecainide<br>Propafenone | Block $Na^+$ channels<br>Slow onset/offset kinetics |
| II | Beta-Blockers | Antagonize β-Adrenergic receptors |
| III | Amiodarone<br>Sotalol<br>Ibutilide<br>Dofetilide<br>Bretylium | Block-delayed rectifier $K^+$ channels |
| IV | Verapamil<br>Bretylium | Block the L-type $Ca^{2+}$ channels |

## 4 ARRHYTHMIAS

### 4.1 Current Concepts in Pharmacotherapy

The incidence of ventricular and supraventricular arrhythmias has increased dramatically in recent decades. This is due, in part, to the aging of our population and the increasing prevalence of HF and CAD. Accounting for one-third of cardiac rhythm disturbance hospitalizations, atrial fibrillation is estimated to affect 2.3 million people in the United States and 4.5 million in Europe [101]. Over the past two decades hospital admissions due to atrial fibrillation has increased by 66% [101]. Although the failure of discrimination from other arrhythmias has complicated calculations, the estimated prevalence of supraventricular arrhythmias is 6 to 8 per 1000 people [102]. Sudden cardiac death (SCD), mostly from ventricular arrhythmias, has been estimated to be responsible for up to 18.5% of all deaths, accounting for approximately 300,000–350,000 deaths annually [103, 104].

Antiarrhythmic drug therapy has traditionally been the mainstay of treatment for both ventricular and supraventricular arrhythmias. In the past decade, the development of nonpharmacologic therapies, such as targeted ablation of arrhythmogenic tissues and implantable cardiac defibrillators, as well as the limited efficacy and the significant adverse effects of antiarrhythmic agents, has dramatically altered the treatment of cardiac arrhythmias [105]. In addition, recent studies have suggested a diminished need for maintaining normal sinus rhythm in patients with atrial fibrillation. This has led some to question the future of antiarrhythmic drugs. These agents, however, will remain an important complement to nonpharmacologic methods of arrhythmia treatment.

Most antiarrhythmics are ion channel blockers and are classified by their predominate site of action [106] (Table 3). Class I agents are those that block $Na^+$ channels. This class is further divided into three subgroups according to their onset and offset kinetics. β-Adrenergic receptor antagonists or beta-blockers make up class II. Class III agents are $K^+$ channel blockers that block one specific potassium current, termed $I_{Kr}$ (the rapid component of the delayed $K^+$ current). Class IV is made up of agents that block L-type $Ca^{2+}$ channels. Although antiarrhythmic drugs in the same class share important electrophysiologic properties, most drugs block multiple ion channels, which limits the value of the simple Vaughan Williams classification scheme. In addition, in most cases, the choice of antiarrhythmic drug therapy is not mechanism based but rather remains empiric and is made on an individual basis by trial and error and consideration of side effects.

Trials comparing rate and rhythm control strategies in patients with persistent and paroxysmal atrial fibrillation (AF) [107–111] found no reduction in death, disabling stroke, hospitalizations, new arrhythmias, or thromboembolic complications when providing therapy to maintain normal sinus rhythm. These studies show that controlling the ventricular rate with atrioventricular (AV) node blocking agents is just as effective at reducing events as treating patients with antiarrhythmics to maintain normal sinus rhythm on an intention-to-treat basis. Many patients, however, eventually need antiarrhythmic drug therapy to suppress symptoms, improve exercise capacity and hemodynamic function, and prevent tachycardia-induced cardiomyopathy due

to AF. Additionally, further analysis shows that a majority of events in the trials comparing rate and rhythm control strategies occurred while patients were in AF, suggesting that the rhythm "strategy" is not at fault but rather that antiarrhythmic medications are not effective enough at maintaining normal sinus rhythm for a benefit to be seen. Current antiarrhythmic agents will maintain sinus rhythm in 23–79% of patients after 12–21 months of follow-up [112].

Ventricular fibrillation (VF) and ventricular tachycardia (VT) are most commonly the consequence of coronary heart disease [113]. They differ fundamentally from supraventricular tachycardias in that they are a common cause of SCD. No class of antiarrhythmic agent has clearly been demonstrated to reduce total and SCD mortality in patients at risk for SCD [114–117], and as a group the antiarrhythmics have limited efficacy for preventing VT [118, 119]. Their present role is as an adjunct to implantable cardioverter-defibrillator (ICD) implantation to reduce the number of shocks or in patients who do not wish to have an ICD. Sixty to 70% of patients with ICDs implanted for spontaneous VT will have therapy delivered for recurrent VT or VF during follow-up [120]. In addition, 20–25% of patients will have "inappropriate shocks," most commonly for rapid atrial fibrillation. For these reasons, additional therapy is often required for ICD recipients for preventing recurrent arrhythmias and ICD firings following device implantation. When used, empiric amiodarone is usually the drug of choice even though most studies show no clear advantage of amiodarone over placebo. [118, 119]. In addition, chronic administration of amiodarone is associated with complex drug interactions and a host of adverse side effects involving the lung, liver, thyroid, and skin. A number of clinical trials have evaluated the efficacy of class Ia, Ib, Ic, and III drugs in the long-term prevention of ventricular arrhythmias. A consistent message from these trials is that antiarrhythmic drugs may be proarrhythmic, especially in patients with coronary disease [121].

Research focused on improving the understanding of the mechanisms of arrhythmias will allow identification of molecular and cellular targets for new antiarrhythmic treatment options. Electrophysiological studies and molecular cloning have detailed the functional properties of $Na^+$, $Ca^{2+}$, and $K^+$ currents in cardiomyocytes and have demonstrated that distinct molecular entities give rise to the various cardiac ion channels and shape the myocardial action potential. Reconstruction of the critical components of cardiac ion channels will permit the development of more selective and effective antiarrhythmics. These recent advances in our understanding of the molecular and cellular basis of arrhythmias suggest that novel drugs can be developed that modulate not only traditional ion channels ($Na^+$, $Ca^{2+}$, and $K^+$ channels) with less proarrhythmic potential but also new molecular targets such as those involved in disease-associated electrical remodeling [122].

### 4.2 Antiarrhythmics under Development

A number of antiarrhythmic agents are currently under development (Table 4). The list includes agents that are modifications of existing compounds and agents with novel mechanisms of action. Three agents that are advanced in their development are vernakalant, azimilide, and dronedarone.

Vernakalant is an atrial-selective $Na^+$ and $K^+$ channel blocker under development for the acute conversion and chronic treatment of atrial fibrillation [123]. This antiarrhythmic blocks $I_{Kur}$, $I_{to}$, $I_{KAch}$, and I-Na currents allowing it to selectively prolong atrial refractoriness with minimal electrophysiologic or proarrhythmic effect on ventricular tissue.

Reverse-use dependence is a characteristic shared by most class III antiarrhythmics [124]. This is where a drug exerts greater effects at slower heart rates than at faster ones. This characteristic is opposite to what an ideal antiarrhythmic agent would do. An ideal antiarrhythmic would exert its effects at fast rates to interrupt or prevent tachycardias rather than at slow rates where prolongation of refractoriness is more prone to precipitate torsades de pointes. Azimilde is a novel class III antiarrhythmic being developed for the treatment of supraventricular and ventricular arrhythmias that does not exhibit reverse-use dependence [125]. This antiarrhythmic reversibly blocks both the rapid acting component ($I_{kr}$) and slowly acting component ($I_{ks}$) of the delayed rectifier potassium channel ($I_k$) [126]. The most significant adverse effect of azimilide is torsades de pointes, although the fact that it does not exhibit reverse-use dependence makes the occurrence less common with this agent than with other class III medications.

Dronedarone is an antiarrhythmic similar to amiodarone. Like amiodarone, dronedarone is considered a class III antiarrhythmic with its predominant mechanism of action consisting of blockade of $I_{kr}$, $I_{kl}$, $I_{KACh}$, and $I_{SUS}$ channels and with additional effects on $Na^+$ and $Ca^{2+}$ channels [127]. Dronedarone was developed to have a similar efficacy as amiodarone but with an improved safety and tolerability profile [128–130].

### 4.3 Conclusion

With advancing technology in surgical and device therapies, the future management of arrhythmias will most

**TABLE 4 Investigational Antiarrhythmic Drug in Development[a]**

| Modification of Existing Compound | Novel Mechanism Of Action |
|---|---|
| Compound (site of action) | Compound (site of action) |
| *Amiodarone analogs* | *Serotonin type 4 antagonists* |
| Dronedarone (IKr IKs β1 ICa Ito INa) | Piboserod |
| Celivarone (IKr IKs β1 ICa Ito INa) | RS100302 |
| ATI-2042 (IKr IKs β1 ICa Ito INa) | SB203186 |
| ATI-2001 (IKr IKs β1 ICa Ito INa) | *Atrial selective repolarization delaying agents* |
| GYKI-16638 (IKr IKI INa) | AZD 7009 (IKr INa IKur) |
| KB 130015 (IKAch INa ICa IKATP) | AVE 0118 (IKur Ito) |
| *Conventional class III agents* | AVE 1231 (IKur Ito) |
| Azimilide (IKr IKs) | Vernakalant (IKur Ito INa IAch) |
| Tedisamil (IKr Ito IKATP IKur INa) | Almokalant (IKur Ito INa IAch) |
| Bertosamil (IKr Ito IKATP IKur INa) | Terikalant (IKur Ito INa IAch) |
| SB-237376 (IKr) | Nifekalant (IKur Ito INa IAch) |
| NIP-142 (IKur IKAch) | S-9947 (IKur) |
| L-768673 (IKs) | S-20951 (IKur) |
| HMR-1556 (IKs) | *Miscellaneous compounds* |
| HMR-1402(IKs. IATP) | ZP-123 (GAP 486) |
| *Miscellaneous compounds* | AAP 10 (connexin modulator) |
| Ersentilide (IKr b) | GsMtx4 (stretch receptor) |
| Trecetilide (IKr b) | |
| CP060S (INa ICa) | |
| KB-R7943 (INa ICa) | |
| Cariporide (INa IH) | |
| JTV-519 (INa IKr ICa) | |

[a]Abbreviations: β, β-adrenergic antagonist; ICa, inward calcium current; IKach, Ach-sensitive inward potassium current; IKATP, ATP-sensitive inward potassium current; IKI, inward potassium rectifier; IKr, rapid component of the delayed rectifier potassium inward current; IKs, slow component of the delayed rectifier potassium inward current; IKur, ultra-rapid component of the delayed rectifier potassium inward current; INa, inward sodium current; Ito, transient outward potassium current.

*Sources*: Data from refs. 131–133.

likely evolve with even less reliance on antiarrhythmic drugs. However, the development of safer and more efficacious antiarrhythmic drugs remains desirable with several potential agents in the pipeline. In addition, the evolving knowledge about the molecular basis of arrhythmias will continue to provide novel drug targets.

## 5 CONCLUSION

Heart disease is responsible for more deaths than any other disease. Recent advances in pharmacotherapy have helped to lessen the burden of heart disease but new technology is needed to maintain our progress. There are numerous potential cardiovascular therapeutic targets in various stages of development for the treatment of CAD, heart failure, and arrhythmias. With advances in genetics and molecular biology the list will most definitely grow.

## REFERENCES


1. Rosamond W, Flegal K, Friday G, et al. (2007). Heart disease and stroke statistics—2007 update: A report from the American Heart Association Statistics Committee and Stroke Statistics Subcommittee. *Circulation 115*, e169–171.
2. Norman JN (1985). William Withering and the purple foxglove: A bicentennial tribute. *J Clin Pharmacol 25*, 479–483.
3. Mueller RL, Scheidt S (1994). History of drugs for thrombotic disease: Discovery, development, and directions for the future. *Circulation 89*, 432–449.
4. Douglas SA, Ohlstein EH, Johns DG (2004). Techniques: Cardiovascular pharmacology and drug discovery in the 21st century. *Trends in Pharmacological Sciences 25*(4), 225–233.
5. Cushman DW, Cheung HS, Sabo EF, et al. (1981). Angiotensin converting enzyme inhibitors: Evolution of a new class of antihypertensive drugs. In Horovitz ZP, Ed.

*Angiotensin Converting Enzyme Inhibitors*. Baltimore, MD: Urben and Schwarzenberg, pp. 3–25.

6. Lee RT (2001). Functional Genomics and Cardiovascular Drug Discovery. *Circulation 104*, 1441–1446.
7. Drews J (2000). Drug discovery: A historical perspective. *Science 287*, 1960–1964.
8. Ford ES, Ajani UA, Croft JB, et al. (2007). Explaining the decrease in U.S. deaths from coronary disease, 1980–2000. *N Engl J Med 356*, 2388.
9. Rosamond WD, Chambless LE, Folsom AR, et al. (1998). Trends in the incidence of myocardial infarction and in mortality due to coronary heart disease, 1987 to 1994. *N Engl J Med 339*, 861.
10. Rogers WJ, Canto JG, Lambrew CT, et al. (2000). Temporal trends in the treatment of over 1.5 million patients with myocardial infarction in the US from 1990 through 1999: The National Registry of Myocardial Infarction 1, 2 and 3. *J Am Coll Cardiol 36*, 2056.
11. Lefkowitz RJ, Caron MG, Stiles GL (1984). Mechanism of membrane-receptor regulation. Biochemical, physiological, and clinical insights derived from studies of the adrenergic receptors. *N Engl J Med 310*, 1570.
12. Frishman WH (1987). Beta-adrenergic blocker withdrawal. *Am J Cardiol 59*, 26F.
13. Psaty BM, Koepsell TD, Wagner EH, et al. (1990). The relative risk of incident coronary heart disease associated with recently stopping the use of ß-blockers. *JAMA 263*, 1653.
14. Parker JD, Parker JO (1998). Nitrate therapy for stable angina pectoris. *N Engl J Med 338*, 520.
15. Abrams J (1985). Hemodynamic effects of nitroglycerin and long-acting nitrates. *Am Heart J 110*, 216.
16. Chen Z, Zhang J, Stamler JS (2002). Identification of the enzymatic mechanism of nitroglycerin bioactivation. *Proc Natl Acad Sci USA 99*, 8306.
17. Chaitman BR (2006). Ranolazine for the treatment of chronic angina and potential use in other cardiovascular conditions. *Circulation 113*, 2462.
18. Chaitman BR, Skettino SL, Parker JO, et al. (2004). Antiischemic effects and long-term survival during ranolazine monotherapy in patients with chronic severe angina. *J Am Coll Cardiol 43*, 1375.
19. Morrow DA, Scirica BM, Karwatowska-Prokopczuk E, et al. (2007). Effects of ranolazine on recurrent cardiovascular events in patients with non-ST-elevation acute coronary syndromes: The MERLIN-TIMI 36 randomized trial. *JAMA 297*, 1775.
20. Roger L, Michael S (2007). New therapies for angina pectoris. In Rose BD, Ed. Waltham, MA: UpToDate.
21. Michael S (2007). Therapeutic angiogenesis for management of refractory angina. In Rose BD, Ed. Waltham, MA: UpToDate.
22. Randomized trial of ridogrel, a combined thromboxane A2 synthase inhibitor and thromboxane A2/prostaglandin endoperoxide receptor antagonist, versus aspirin as adjunct to thrombolysis in patients with acute myocardial infarction. The Ridogrel Versus Aspirin Patency Trial (RAPT). *Circulation 89*(2), 588–595, (1994).
23. Langleben D, Christman BW, Barst RJ, et al. (2002). Effects of the thromboxane synthetase inhibitor and receptor antagonist terbogrel in patients with primary pulmonary hypertension. *Am Heart J 143*(5), E4.
24. Neri Serneri GG, Coccheri S, Marubini E, Violi F (2004). Picotamide, a combined inhibitor of thromboxane A2 synthase and receptor, reduces 2-year mortality in diabetics with peripheral arterial disease: The DAVID study. *Eur Heart J 25*(20), 1845–1852.
25. Sugidachi A, Asai F, Ogawa T, Inoue T, Koike H (2000). The in vivo pharmacological profile of CS-747, a novel antiplatelet agent with platelet ADP receptor antagonist properties. *Br J Pharmacol 129*(7), 1439–1446.
26. Sugidachi A, Ogawa T, Kurihara A, et al. (2007). The greater in vivo antiplatelet effects of prasugrel as compared to clopidogrel reflect more efficient generation of its active metabolite with similar antiplatelet activity to that of clopidogrel's active metabolite. *J Thromb Haemost 5*(7), 1545–1551.
27. Wiviott S, Antman E, Winters K, et al. (2005). Randomized comparison of prasugrel (CS-747, LY640315), a novel thienopyridine P2Y12 antagonist, with clopidogrel in percutaneous coronary intervention: Results of the Joint Utilization of Medications to Block Platelets Optimally (JUMBO)-TIMI 26 trial. *Circulation 111*(25), 3366–3373.
28. van Giezen J, Humphries R (2005). Preclinical and clinical studies with selective reversible direct P2Y12 antagonists. *Semin Thromb Hemost 31*(2), 195–204.
29. Davey MG, Luscher EF (1967). Actions of thrombin and other coagulant and proteolytic enzymes on blood platelets. *Nature 216*(5118), 857–885.
30. Kahn ML, Nakanishi-Matsui M, Shapiro MJ, Ishihara H, Coughlin SR (1999). Protease-activated receptors 1 and 4 mediate activation of human platelets by thrombin. *J Clin Invest 103*(6), 879–887.
31. Ma L, Hollenberg MD, Wallace JL (2001). Thrombin-induced platelet endostatin release is blocked by a proteinase activated receptor-4 (PAR4) antagonist. *Br J Pharmacol 134*(4), 701–704.
32. Covic L, Misra M, Badar J, Singh C, Kuliopulos A (2002). Pepducin-based intervention of thrombin-receptor signaling and systemic platelet activation. *Nat Med 8*(10), 1161–1165.
33. Stein RC, Waterfield MD (2000). PI3-kinase inhibition: A target for drug development? *Mol Med Today 6*(9), 347–357.
34. Jackson SP, Schoenwaelder SM, Goncalves I, et al. (2005). PI 3-kinase p110beta: A new target for antithrombotic therapy. *Nat Med 11*(5), 507–514.
35. Bhatt DL, Topol EJ (2003). Scientific and therapeutic advances in antiplatelet therapy. *Nat Rev Drug Discov 2*(1), 15–28.
36. Wu D, Meiring M, Kotze HF, Deckmyn H, Cauwenberghs N (2002). Inhibition of platelet glycoprotein Ib, glycopro-

tein IIb/IIIa, or both by monoclonal antibodies prevents arterial thrombosis in baboons. *Arterioscler Thromb Vasc Biol 22*(2), 323–328.

37. Chang MC, Lin HK, Peng HC, Huang TF (1998). Antithrombotic effect of crotalin, a platelet membrane glycoprotein Ib antagonist from venom of *Crotalus atrox*. *Blood 91*(5), 1582–1589.
38. Sakurai Y, Fujimura Y, Kokubo T, et al. (1998). The cDNA cloning and molecular characterization of a snake venom platelet glycoprotein Ib-binding protein, mamushigin, from *Agkistrodon halys blomhoffii* venom. *Thromb Haemost 79*(6), 1199–1207.
39. Yamamoto H, Vreys I, Stassen JM, Yoshimoto R, Vermylen J, Hoylaerts MF (1998). Antagonism of vWF inhibits both injury induced arterial and venous thrombosis in the hamster. *Thromb Haemost 79*(1), 202–210.
40. Golino P, Ragni M, Cirillo P, et al. (1995). Aurintricarboxylic acid reduces platelet deposition in stenosed and endothelially injured rabbit carotid arteries more effectively than other antiplatelet interventions. *Thromb Haemost 74*(3), 974–979.
41. Azzam K, Garfinkel LI, Bal dit Sollier C, Cisse Thiam M, Drouet L (1995). Antithrombotic effect of a recombinant von Willebrand factor, VCL, on nitrogen laser-induced thrombus formation in guinea pig mesenteric arteries. *Thromb Haemost 73*(2), 318–323.
42. McGhie AI, McNatt J, Ezov N, et al. (1994). Abolition of cyclic flow variations in stenosed, endothelium-injured coronary arteries in nonhuman primates with a peptide fragment (VCL) derived from human plasma von Willebrand factor-glycoprotein Ib binding domain. *Circulation 90*(6), 2976–2981.
43. Konstantinides S, Schafer K, Koschnick S, Loskutoff DJ (2001). Leptin-dependent platelet aggregation and arterial thrombosis suggests a mechanism for atherothrombotic disease in obesity. *J Clin Invest 108*(10), 1533–1540.
44. Li Z, Kim ES, Bearer EL (2002). Arp2/3 complex is required for actin polymerization during platelet shape change. *Blood 99*(12), 4466–4474.
45. Enjyoji K, Sevigny J, Lin Y, et al. (1999). Targeted disruption of cd39/ATP diphosphohydrolase results in disordered hemostasis and thromboregulation. *Nat Med 5*(9), 1010–1017.
46. Pinsky DJ, Broekman MJ, Peschon JJ, et al. (2002). Elucidation of the thromboregulatory role of CD39/ectoapyrase in the ischemic brain. *J Clin Invest 109*(8), 1031–1040.
47. Henn V, Slupsky JR, Grafe M, et al. (1998). CD40 ligand on activated platelets triggers an inflammatory reaction of endothelial cells. *Nature 391*(6667), 591–594.
48. Schober A, Manka D, von Hundelshausen P, et al. (2002). Deposition of platelet RANTES triggering monocyte recruitment requires P-selectin and is involved in neointima formation after arterial injury. *Circulation 106*(12), 1523–1529.
49. von Hundelshausen P, Weber KS, Huo Y, et al. (2001). RANTES deposition by platelets triggers monocyte arrest on inflamed and atherosclerotic endothelium. *Circulation 103*(13), 1772–1777.
50. Specht S (2007). ZymoGenetics reports new findings on anti-thrombotic activities of CTRP1; novel protein prevents platelet thrombosis without causing bleeding 7/17/2003. Available: http://www.zymogenetics.com/ir/newsItem.php?id=432558. Accessed December 18, 2007.
51. Hunt SA, et al. (2005). ACC/AHA 2005 guideline update for the diagnosis and management of chronic heart failure in the adult. *Circulation 112*, e154–e235.
52. St John Sutton M, et al. (1994). Quantitative two-dimensional echocardiographic measurements are major predictors of adverse cardiovascular events after acute myocardial infarction. The protective effects of captopril. *Circulation 89*, 68–75.
53. St John Sutton M, et al. (2003). Left ventricular remodeling and ventricular arrhythmias after myocardial infarction. *Circulation 107*, 2577–2582.
54. Frey N, Olson EN (2003). Cardiac hypertrophy: The good, the bad, and the ugly. *Annu Rev Physiol 65*, 45–79.
55. Brown JH, Del Re DP, Sussman MA (2006). The Rac and Rho hall of fame: A decade of hypertrophic signalling hits. *Circ Res 98*, 730–742.
56. Hattori T, et al. (2004). Long-term inhibition of Rho-kinase suppresses left ventricular remodeling after myocardial infarction in mice. *Circulation 109*, 2234–2239.
57. Leinwand LA (2001). Calcineurin inhibition and cardiac hypertrophy: A matter of balance. *Proc Natl Acad Sci USA 98*, 2947–2949.
58. Taigen T, De Windt LJ, Lim HW, Molkentin JD (2000). Targeted inhibition of calcineurin prevents agonist-induced cardiomyocyte hypertrophy. *Proc Natl Acad Sci USA 97*, 1196–1201.
59. Vega RB, et al. (2003). Dual roles of modulatory calcineurin-interacting protein 1 in cardiac hypertrophy. *Proc Natl Acad Sci USA 100*, 669–674.
60. Gao XM, et al. (2006). Inhibition of mTOR reduces chronic pressure-overload cardiac hypertrophy and fibrosis. *J Hypertens 24*, 1663–1670.
61. Kee HJ, et al. (2006). Inhibition of histone deacetylation blocks cardiac hypertrophy induced by angiotensin II infusion and aortic banding. *Circulation 113*, 51–59.
62. Katsuragi N, Morishita R, Nakamura N, Ochiai T, Taniyama Y, Hasegawa Y, Kawashima K, Kaneda Y, Ogihara T, Sugimura K (2004). Periostin as a novel factor responsible for ventricular dilation. *Circulation 110*, 1806–1813.
63. Aukrust P, Gullestad L, Ueland T, Damas JK, Yndestad A (2005). Inflammatory and anti-inflammatory cytokines in chronic heart failure: Potential therapeutic implications. *Ann Med 37*, 74–85.
64. Malave HA, Taylor AA, Nattama J, Deswal A, Mann DL (2003). Circulating levels of tumour necrosis factor

correlate with indexes of depressed heart rate variability: A study in patients with mild-to-moderate heart failure. *Chest 123*, 716–724.

65. Mann DL, et al. (2004). Targeted anticytokine therapy in patients with chronic heart failure: Results of the Randomized Etanercept Worldwide Evaluation (RENEWAL). *Circulation 109*, 1594–1602.
66. Gullestad L, et al. (2001). Immunomodulating therapy with intravenous immunoglobulin in patients with chronic heart failure. *Circulation 103*, 220–225.
67. McNamara DM, et al. (2001). Controlled trial of intravenous immune globulin in recent-onset dilated cardiomyopathy. *Circulation 103*, 2254–2259.
68. Colucci WS, Braunwald E (2005). In Zipes DP, Libby P, Bonow RO, Braunwald E, Eds. *Braunwald's Heart Disease: A Textbook of Cardiovascular Medicine*. Philadelphia: Elsevier Saunders, pp. 509–538.
69. Wehrens XH, Lehnart SE, Marks AR (2005). Intracellular calcium release and cardiac disease. *Annu Rev Physiol 67*, 69–98.
70. Wehrens XH, Lehnart SE, Marks AR (2005). Ryanodine receptor-targeted anti-arrhythmic therapy. *Ann NY Acad Sci 1047*, 366–375.
71. Earl GL, Fitzpatrick JT (2005). Levosimendan: A novel inotropic agent for treatment of acute, decompensated heart failure. *Ann Pharmacother 39*, 1888–1896.
72. Shen Y, Vatner S, Morgans D, Malik F (2006). Activating cardiac myosin, a novel inotropic mechanism to improve cardiac function in conscious dogs with congestive heart failure. *J Card Fail 12* (Suppl 1), S587.
73. Rocchetti M, Besana A, Mostacciuolo G, Ferrari P, Micheletti R, Zaza A (2003). Diverse toxicity associated with cardiac Na /K pump inhibition: Evaluation of electrophysiological mechanisms. *J Pharmacol Exp Ther 305*, 765–771.
74. Adamson PB, Vanoli E, Mattera GG, et al. (2003). Hemodynamic effects of a new inotropic compound, PST-2744, in dogs with chronic ischemic heart failure. *J Cardiovasc Pharmacol 42*, 169–173.
75. Stanley WC, Lopaschuk GD, Hall JL, McCormack JG (1997). Regulation of myocardial carbohydrate metabolism under normal and ischaemic conditions. Potential for pharmacological interventions. *Cardiovasc Res 33*, 243–257.
76. Liedtke AJ (1981). Alterations of carbohydrate and lipid metabolism in the acutely ischemic heart. *Prog Cardiovasc Dis 23*, 321–336.
77. Lee L, Horowitz J, Frenneaux M (2004). Metabolic manipulation in ischaemic heart disease, a novel approach to treatment. *Eur Heart J 25*, 634–641.
78. Lee L, Campbell R, Scheuermann-Freestone M, et al. (2005). Metabolic modulation with perhexiline in chronic heart failure: A randomized, controlled trial of short-term use of a novel treatment. *Circulation 112*, 3280–3288.
79. Fragasso G, et al. (2006). Effects of metabolic modulation by trimetazidine on left ventricular function and phosphocreatine/adenosine triphosphate ratio in patients with heart failure. *Eur Heart J 27*, 942–948.
80. Francis GS, et al. (1990). Comparison of neuroendocrine activation in patients with left ventricular dysfunction with and without congestive heart failure. A substudy of the studies of left ventricular dysfunction. *Circulation 82*, 1724–1729.
81. CONSENSUS Trial Study Group (1987). Effects of enalapril on mortality in severe congestive heart failure. Results of the Cooperative North Scandinavian Enalapril Survival Study (CONSENSUS). The CONSENSUS Trial Study Group. *N Engl J Med 316*, 1429–1435.
82. Cohn JN, et al. (1991). A comparison of enalapril with hydralazine-isosorbide dinitrate in the treatment of chronic congestive heart failure. *N Engl J Med 325*, 303–310.
83. Stanton A (2003). Therapeutic potential of renin inhibitors in the management of cardiovascular disorders. *Am J Cardiovasc Drugs 3*, 389–394.
84. Matsubara H (1998). Pathophysiological role of angiotensin II type 2 receptor in cardiovascular and renal diseases. *Circ Res 83*, 1182–1191.
85. Lee CR, Watkins ML, Patterson JH, et al. (2003). Vasopressin: A new target for the treatment of heart failure. *Am Heart J 146*, 9–18.
86. Sanghi P, Uretsky BF, Schwarz ER (2005). Vasopressin antagonism: A future treatment option in heart failure. *Eur Heart J 26*, 538–543.
87. Gheorghiade M, Gattis WA, O'Connor CM, et al. (2004). Effects of tolvaptan, a vasopressin antagonist, in patients hospitalized with worsening heart failure: A randomized controlled trial. *JAMA 291*, 1963–1971.
88. Gheorghiade M, Orlandi C, Burnett JC, et al. (2005). Rationale and design of the multicenter, randomized, double-blind, placebo-controlled study to evaluate the efficacy of vasopressin antagonism in heart failure: Outcome study with tolvaptan (EVEREST). *J Card Fail 11*, 260-269.
89. Tang WH, Francis GS (2003). Novel pharmacological treatments for heart failure. *Expert Opin Investig Drugs 12*, 1791–1801.
90. Wada K, Fujimori A, Matsukawa U, et al. (2005). Intravenous administration of conivaptan hydrochloride improves cardiac hemodynamics in rats with myocardial infarction-induced congestive heart failure. *Eur J Pharmacol 507*, 145–151.
91. Wada K, Tahara A, Arai Y, et al. (2002). Effect of the vasopressin receptor antagonist conivaptan in rats with heart failure following myocardial infarction. *Eur J Pharmacol 450*, 169–177.
92. Forssmann W, Meyer M, Forssmann K (2001). The renal urodilatin system: Clinical implications. *Cardiovasc Res 51*, 450–462.

93. Mitrovic V, et al. (2005). Effects of the renal natriuretic peptide urodilatin (ularitide) in patients with decompensated chronic heart failure: A double-blind, placebo-controlled, ascending-dose trial. *Am Heart J 150*, 1239.
94. Rademaker MT, et al. (2003). Adrenomedullin and heart failure. *Regul Pept 112*, 51–60.
95. Nakamura MT, et al. (1997). Potent and long-lasting vasodilatory effects of adrenomedullin in humans. Comparisons between normal subjects and patients with chronic heart failure. *Circulation 95*, 1214–1221.
96. Ishimitsu T, Ono H, Minami J, Matsuoka H (2006). Pathophysiologic and therapeutic implications of adrenomedullin in cardiovascular disorders. *Pharmacol Ther 111*, 909–927.
97. Gilbert RE, Douglas SA, Krum H (2004). Urotensin-II as a novel therapeutic target in the clinical management of cardiorenal disease. *Curr Opin Investig Drugs 5*, 276–282.
98. Douglas SA, Tayara L, Ohlstein EH, Halawa N, Giaid A (2002). Congestive heart failure and expression of myocardial urotensin II. *Lancet 359*, 1990–1997.
99. Jin J, et al. (2005). Aminoalkoxybenzyl pyrrolidines as novel human urotensin-II receptor antagonists. *Bioorg Med Chem Lett 15*, 3229–3232.
100. Eisenstein EL, et al. (2005). Reducing the costs of phase III cardiovascular clinical trials. *Am Heart J 149*, 482–488.
101. Fuster V, et al. (2006). ACC/AHA/ESC 2006 Guidelines for the Management of Patients with Atrial Fibrillation: A report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines and the European Society of Cardiology Committee for Practice Guidelines. *Circulation 114*(7), e257–354.
102. Blomström-Lundqvist C, et al. (2003). ACC/AHA/ESC guidelines for the management of patients with supraventricular arrhythmias—Executive summary: A report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines and the European Society of Cardiology Committee for Practice Guidelines. *Circulation 108*(15), 1871–1909.
103. Myerburg RJ, Kessler KM, Castellanos A (1993). Sudden cardiac deathepidemiology, transient risk, and intervention assessment. *Ann Intern Med 119*, 1187–1197.
104. de Vreede-Swagemakers JJ, Gorgels AP, Dubois-Arbouw WI, et al. (1997). Out-of-hospital cardiac arrest in the 1990's: A population-based study in the Maastricht area on incidence, characteristics and survival. *J Am Coll Cardiol 30*, 1500–1505.
105. Zipes DP, Camm AJ, et al. (2006). ACC/AHA/ESC 2006 guidelines for management of patients with ventricular arrhythmias and the prevention of sudden cardiac death: A report of the American College of Cardiology/American Heart Association Task Force and the European Society of Cardiology Committee for Practice Guidelines. *J Am Coll Cardiol 48*, 247–346.
106. Vaughan Williams EM (1970). The experimental basis for the choice of an antiarrhythmic drug. *Adv Cardiol 4*, 275–289.
107. Van Gelder IC, Hagens VE, Bosker HA, et al. (2002). A comparison of rate control and rhythm control in patients with recurrent persistent atrial fibrillation. *N Engl J Med 347*, 1834–1840.
108. Carlsson J, Miketic S, Windeler J, et al. (2003). Randomized trial of rate-control versus rhythm-control in persistent atrial fibrillation: The Strategies of Treatment of Atrial Fibrillation (STAF) study. *J Am Coll Cardiol 41*, 1690–1696.
109. Opolski G, Torbicki A, Kosior DA, et al. (2004). Rate control vs. rhythm control in patients with nonvalvular persistent atrial fibrillation: The results of the Polish How to Treat Chronic Atrial Fibrillation (HOT CAFE) Study. *Chest 126*, 476–486.
110. Hohnloser SH, Kuck KH, Lilienthal J (2000). Rhythm or rate control in atrial fibrillation—Pharmacological Intervention in Atrial Fibrillation (PIAF): A randomised trial. *Lancet 356*, 1789–1794.
111. Cooper HA, Bloomfield DA, Bush DE, et al. (2004). Relation between achieved heart rate and outcomes in patients with atrial fibrillation (from the Atrial Fibrillation Follow-up Investigation of Rhythm Management [AFFIRM] Study). *Am J Cardiol 93*, 1247–1253.
112. Crijns HJ, Gosselink AT (1994). Prophylactic drug therapy after cardioversion of AF or atrial. *Flutter Cardio 11*, 31–34.
113. Demirovic J, Myerburg RJ (1994). Epidemiology of sudden cardiac death: An overview. *Prog Cardiovasc Dis 37*, 39–48.
114. The AVID Investigators. (1997). A comparison of antiarrhythmic drug therapy with implantable defibrillators in patients resuscitated from near-fatal ventricular arrhythmias. *N Engl J Med 337*, 1576–1584.
115. Kuck KH, Cappato R, Siebels J, et al. (2000). Randomized comparison of antiarrhythmic drug therapy with implantable defibrillators in patients resuscitated from cardiac arrest. The cardiac arrest study Hamburg (CASH). *Circulation 102*, 748–754.
116. Connolly SJ, Gent M, Roberts RS, et al. (2000). Canadian implantable defibrillator study (CIDS): A randomized trial of the implantable cardioverter defibrillator against amiodarone. *Circulation 101*, 1297–1302.
117. Bardy GH, Lee KL, Mark DB, et al. (2005). Amiodarone or an implantable cardioverter-defibrillator for congestive heart failure. *N Engl J Med 253*, 225–237.
118. The CASCADE Investigators. (1993). Randomised antiarrhythmic therapy in survivors of cardiac arrest (the CASCADE study). *Am J Cardiol 72*, 280.
119. Mason JW (1993). A comparison of electrophysiologic testing with Holter monitoring to predict antiarrhythmic drug efficacy for ventricular arrhythmias. *N Engl J Med 329*, 445–451.

120. Klein RC, Raitt MH, Wilkoff BL, et al. (2003). Analysis of implantable cardioverter defibrillator therapy in the Antiarrhythmics Versus Implantable Defibrillators (AVID) Trial. *J Cardiovasc Electrophysiol 149*, 940–948.
121. Echt DS, Liebson PR, Mitchell LB (1991). Mortality and morbidity in patients receiving encainide, flecainide, or placebo. The cardiac arrhythmia suppression trial. *N Engl J Med 324*, 781–788.
122. Darbar D, Roden DM (2006). Future of antiarrhythmic drugs. *Curr Opin Cardiol 21*, 361–367.
123. Fedida D (2007). Vernakalant (RSD1235): A novel, atrial-selective antifibrillatory agent. *Expert Opin Investig Drugs 16*(4), 519–532.
124. Libby P (2008). *Braunwald's Heart Disease: A Textbook of Cardiovascular Medicine*, 8th ed. Philadelphia: Saunders Elsevier.
125. Salata JJ, Brooks R (1997). Pharmacology of azimilide dihydrochloride (NE-10064). *Cardiovasc Drug Rev 15*, 137–156.
126. Karam R, et al. (1998). Azimilide dihydrochloride, a novel antiarrhythmic agent. *AJC 81*, 40D–46D.
127. Gautier P, Marion A, Bertrand JP, Tourneur Y, Nisato D (1997). Electrophysiological characteristics of dronedarone (SR 33589), a new amiodarone-like agent, in cardiac ventricular myocytes. *Eur Heart J 18*(Suppl), 269–278.
128. Sun W, Sarma JSM, Singh BN (1999). Electrophysiological effects of dronedarone (SR 33589), a noniodinated benzofuran derivative in the rabbit heart: Comparison with amiodarone. *Circulation 100*, 2276–2283.
129. Sun W, Sarma JSM, Singh BN (2002). Chronic and acute effects of dronedarone on the action potential of rabbit atrial muscle preparations: Comparison with amiodarone. *J Cardiovasc Pharmacol 39*, 677–684.
130. Singh BN, Connolly SJ, Crijns HJ, et al. (2007). Dronedarone for maintenance of sinus rhythm in atrial fibrillation or flutter. *N Engl J Med 357*, 987–999.
131. Goldstein RN, Stambler BS (2005). New antiarrhythmic drugs for prevention of atrial fibrillation. *Prog Cardiovasc Dis 48*(3), 193–208.
132. Pecini R, Elming H, Pedersen OD, et al. (2005). New antiarrhythmic agents for atrial fibrillation and atrial flutter. *Expert Opin Emerg Drugs 10*(2), 311–322.
133. Musco S, Conway EL, Kowey PR (2008). Drug therapy for atrial fibrillation. *Med Clin N Am 92*, 121–141.

# EZETIMIBE AND CHOLESTEROL ABSORPTION

Stefan Oswald and Werner Siegmund
*Department of Clinical Pharmacology, University of Greifswald, Greifswald, Germany*

## 1 INTRODUCTION

Coronary heart disease (CHD) remains the leading cause of morbidity and mortality in the western world. More than the half of atherosclerosis-associated deaths were due to CHD (in the United States: ~450,000 in 2004), which is responsible for 83% of deaths in individuals over 65 years of age [1]. There is convincing evidence from many large epidemiological studies that elevated serum low-density lipoprotein cholesterol (LDL-C) concentrations are associated with an increased risk for the development of atherosclerosis and CHD leading to such potential life-threatening events as myocardial infraction and stroke [2–5].

The pathophysiological mechanisms behind these observations are complex and currently explained with LDL-C oxidation, incorporation of these lipids into macrophages, their conversion to foam cells, and subsequent development of atherosclerotic lesions in the vessel wall, which results in manifold dysfunctions of the endothelium [6].

Recent treatment guidelines such as the Third National Cholesterol Education Program (NCEP) emphasize the need for aggressive LDL-C-lowering therapy in individuals at substantial risk or patients with cardiovascular diseases [7].

For pharmacological intervention several classes of lipid-lowering drugs have been approved in the last two decades, such as hydroxy-methylglutary-coenzyme A (HMG-CoA) reductase inhibitors (statins), fibrates, nicotinic acids, and bile-acid-binding resins. In particular for statins, the most potential and most frequently prescribed of the cholesterol lowering drugs, the benefits of lipid-lowering therapy have been clearly demonstrated in several clinical trials for primary and secondary prevention [8–10]. However, many hypercholesterolemic patients fail to reach the recommended LDL-C target concentrations on statin therapy [11, 12]. Potential reasons include poor patient compliance, insufficient drug response, drug intolerability, as well as inadequate stain dose caused by concerns about safety risks at higher satin doses. With regard to the latter, the relatively high doses of statins required to enhance their cholesterol-lowering effects are connected with the incidence of serious adverse side effects (e.g., myopathy, rhabdomyolysis) [13]. However, even at maximum statin dose titration up to 80 mg, approximately one-third of patients did not achieve the recommended LDL-C levels [14]. Compared to the high LDL-C-lowering potency of 30–40% by low-dose satins (10–20 mg/day), doubling the daily dose only results in an additional cholesterol-lowering benefit of about 5%. Although combining lipid-lowering drugs that act via complementary pathways may provide more pronounced reduction of LDL-C, comedication with available agents may increase the risk for drug interactions or side effects [15]. Specifically, the combination therapy with fibrates was shown to increase the risk of myotoxicity, which caused the worldwide withdrawal of cerivastatin in August 2001 due to numerous deaths [16]. Thus, there was the urgent need for other compounds and principles that would offer supportive strategies.

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

Accordingly, the demand for additional agents has directed research toward the inhibition of intestinal cholesterol absorption. The rationale for this therapeutic strategy is based on the known positive correlation between the serum LDL-C levels and the efficiency of intestinal cholesterol absorption in humans [17]. Consequently, a 90–100% reduction of intestinal cholesterol absorption achieved a lowering of total serum cholesterol and LDL-C by 35–62% in hypercholesterolemic individuals [18]. With the approval of ezetimibe in the fall of 2002, the first selective inhibitor of intestinal cholesterol absorption became available and has markedly advanced the treatment of hypercholesterolemia.

This chapter will focus on the discovery and development of ezetimibe, its pharmacological and pharmacokinetic properties, as well as its therapeutic implications. Furthermore, the molecular mechanisms of cholesterol absorption, whose understanding was completely revolutionized by ezetimibe-associated research, and novel therapeutic approaches affecting this promising pharmacological mechanism will be highlighted.

## 2 INTESTINAL CHOLESTEROL ABSORPTION

The serum concentration of cholesterol is maintained by a complex and highly regulated balance of endogenous de novo synthesis in the liver, removal and catabolism of lipoproteins from the systemic circulation, biliary secretion of cholesterol and bile acids, as well as intestinal (re-)absorption of bile acids and of dietary and biliary cholesterol [19]. From the current point of view, intestinal cholesterol absorption is a multistep process involving several enzymes and proteins (Fig. 1). The total cholesterol content of an adult human body (70 kg) is about 140 g of which less than 1% is subjected to daily turnover [20]. The biliary secretion of cholesterol into the alimentary canal ranges from 0.5 to 2 g/day of which nearly 50% is reabsorbed. Additionally, the intestinal mucosal cell turnover contributes ~300 mg/day of intraluminal cholesterol. Thus, from the physiological point of view humans do not have a dietary requirement for cholesterol because the body is fully capable to synthesize all the cholesterol that it needs. Nevertheless, a typical western diet delivers 300–500 mg/day of additional cholesterol into the intestinal tract [21]. In humans, cholesterol absorption primarily occurs in the duodenum and proximal jejunum of the small intestine and is characterized by a pronounced interindividual variation (20–80%). Dietary cholesterol intake is usually associated with fat consumption, and digestion and absorption begins in the stomach when lingual and gastric lipases are mixed with chyme and hydrolyze dietary triglycerides. This process continues within the small intestinal lumen, where bile and pancreatic juices are added to the crude emulsion. The juices contain several hydrolytic enzymes, including cholesterol esterase, that hydrolyzes the small portion of dietary cholesterol esters (5–15%) [22]. Under the influence of bile salts, phospholipids, fatty acids, and monoglycerides, mixed micelles are formed in which unesterified cholesterol is incorporated. This step is critical for the enteral absorption of cholesterol due to its nearly water insolubility ($2.3 \times 10^{-6}$ g/L at pH 1–10). Before interaction and subsequent transport of cholesterol molecules across the mucosal cell membrane, an unstirred water layer and an aqueous surface mucous coat, both diffusion barriers that are located at the intestinal lumen–membrane interface, must be passed. This emphasizes the need of bile salts as biologic amphipathic detergents. Plant sterols and stanols as ß-sitosterol and ß-sitostanol are thought to interrupt the formation of these mixed micelles, leading to a significant reduction of cholesterol absorption, which is the rationale behind marketing of plant sterol or stanol-enriched margarines [23].

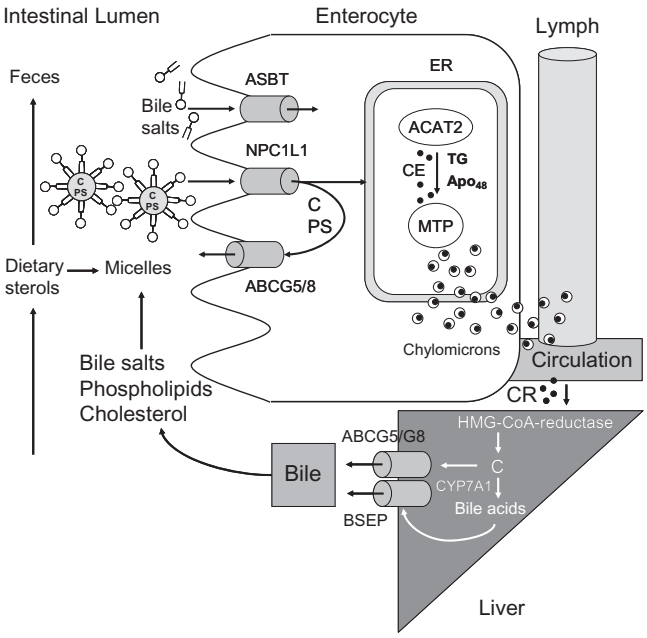


**Figure 1** Intestinal and hepatic determinants involved in intestinal sterol absorption. Abbreviations: ABC, ATP-binding cassette; ACAT, acyl CoA:cholesterol acyl transferase; $Apo_{48}$, apolipoprotein B48; ASBT, apical sodium-dependent bile salt transporter; BSEP, bile salt exporting pump; C, cholesterol; CE, cholesterol esters; CR, chylomicron remnants; CYP, cytochrome P450 enzyme; ER, endoplasmic reticulum; HMG-CoA, hydroxy-methylglutary-coenzyme A; MTP, microsomal triglyceride transfer protein; NPC1L1, Niemann–Pick C1 like 1; PS, plant sterols; TG, triglycerides.

Thereafter, the micellar fraction undergoes disaggregation at the brush border membrane of the enterocytes. Referring to the following absorption step, it was assumed for decades that lipophilic cholesterol (log $P = 9.83$) undergoes only passive diffusion uptake from the intestinal lumen. There was a significant gap between the knowledge of well-defined cholesterol biosynthesis and clearance pathways and the molecular mechanisms controlling cholesterol absorption in the gut wall, which was supposed to play only a negligible role in overall cholesterol disposition. Contrary to this, there was ample evidence that cholesterol absorption might be facilitated by a certain uptake carrier because (1) cholesterol absorption is characterized by a second-order kinetics, (2) marked differences between the absorption of structure-related plant sterols and cholesterol were detected, (3) the transport process is sensitive to proteases, and (4) small molecules such as neomycin, an aminoglycoside antibiotic, or SCH48461 are capable to interrupt intestinal sterol absorption, whereas the identity of this putative cholesterol transporter remained elusive [20, 23]. Several potential candidates such as the adenosine triphosphate (ATP)-binding cassette (ABC) A1, the scavenger receptor type B1 (SR-B1), the sterol-efflux pumps ABCG5/G8, the annexin-2/caveolin-1 complex, or the aminopeptidase N (CD36) were postulated to be involved in cholesterol uptake. However, none of these variables could be shown to facilitate cholesterol uptake in vivo [24, 25].

In March 2004, Altmann et al. from the Schering-Plough laboratories overcame the paradigm of passive cholesterol absorption by showing that with ezetimibe cholesterol lowering is obviously caused by the inhibition of the Niemann–Pick C 1 like 1 (NPC1L1) protein [26]. NPC1L1 was shown to be mainly expressed in the proximal jejunum, that is, the intestinal site of cholesterol absorption. In this report, it was assessed that NPC1L1 knockout mice showed significantly lowered cholesterol absorption compared to wild-type animals, which was similar to the effects of ezetimibe-treated wild-type mice and resistant to further reduction by ezetimibe. These findings led to the suggestion that NPC1L1 is critical for the intestinal sterol uptake and most likely the molecular target of ezetimibe.

However, cholesterol that enters the enterocytes and that is not effluxed via the half transporters ABCG5/G8 moves to the endoplasmic reticulum, where it is esterified by the acyl CoA:cholesterol acyl transferase (ACAT) 2. It is now well established that ABCG5/G8, which resides on the apical membrane of the enterocytes and prevents sterol accumulation, occurs a higher efflux capacity for plant sterols compared to cholesterol, which is responsible for their poor absorption (5–15%) and their markedly lower serum concentrations in humans [21, 22]. Consistently, mutations in one of these transporter genes leads to the rare hereditary disease of sitosterolemia, which is characterized by hyperabsorption of plant sterols and cholesterol, which accelerates the development of atherosclerosis and premature CHD [27].

In the following event, esterified cholesterol, triglycerides, and apolipoprotein $B_{48}$ (apoB) are assembled to nascent chylomicrons via the microsomal triglyceride transfer protein (MTP) [19, 22]. The importance of MTP activity is clarified by reports of patients with abetalipoproteinemia, a disease caused by *MTP* gene mutations, who exhibit serve steatorrhoea, fatty liver, neurologic symptoms, and only low serum cholesterol levels. Finally, the chylomicrons are delivered into lymph and reach the systemic circulation via the thoracic duct. After the degrading influence of endothelial lipoprotein lipases, the resulting cholesterol-enriched chylomicron remnants are taken up by the liver to be processed to various lipoproteins, which enable cholesterol transport to the periphery to fulfill its numerous physiological functions (e.g., synthesis of bile acids, vitamin D, and steroid hormones) [20, 22]. Figure 1 summarizes the intestinal and hepatic variables involved in cholesterol absorption.

In reference to the intestine, hepatobiliary secretion of sterols is mediated by ABCG5/G8, which initiates an enterohepatic recirculation. Bile salts are predominately excreted into gut lumen via the bile salt exporting pump (BSEP). Cholesterol and bile that escape intestinal reabsorption are excreted with feces as neutral sterols or sterol acids, which represent the major sterol elimination pathway from the body.

Despite that under certain circumstances (e.g., disease, drug therapy) one of the mentioned steps involved in intestinal cholesterol trafficking (e.g., lipase activity, formation of mixed micelles, sterol efflux, esterification) can become rate limiting, the NPC1L1-mediated absorption appears to be the most important and specific step of the sterol absorption machinery without influence on the absorption of other lipids as triglycerides, fatty acids, and lipophilic vitamins.

Beyond the described cascade of enzymes and proteins that mediate in concert the intestinal cholesterol absorption, the whole-body cholesterol homeostasis is controlled by a highly regulated network of nuclear receptors to counteract excessive cholesterol uptake. Thus, in many studies even massive intake of dietary cholesterol caused no concomitant elevation of serum cholesterol, which led some authors to the conclusion that there is no association between dietary cholesterol intake and atherosclerosis [28]. For instance, dietary sterols that are taken up from the intestine undergo cytochrome P450 (CYP) enzyme-mediated

biotransformation to oxysterols, which are strong inducers of the nuclear transcription factor lipid X receptor (LXR), leading to increased expression of intestinal and hepatic ABCG5/G8 and ABCA1 but decreasing levels of HMG-CoA synthetase and reductase as well as NPC1L1 [29, 30]. Thus, cholesterol activates its own intestinal and hepatic kick-out transport by ABCG5/G8 but slows down its intestinal NPC1L1-mediated uptake and its endogenous synthesis. Furthermore, oxysterols activate CYP7A1 (via binding to LXR), which metabolizes cholesterol to bile acids. These degradation products are subjected to active biliary secretion by BSEP. Interestingly, the expression of BSEP is enhanced by bile acids due to activation of the farnesoid X receptor (FXR), which downregulates on the other side CYP7A1 and the ileal bile acid uptake transporter (apical sodium-dependent bile salt transporter, ASBT). Altogether, this elevates cholesterol elimination but also prevents systemic accumulation of cytotoxic bile acids. In contrast to this, inhibition of HMG-CoA reductase by statins was shown to increase the intestinal NPC1L1 expression and cholesterol absorption [31, 32].

Finally, there are several functional feedback mechanisms that control intestinal absorption, endogenous biosynthesis, enzymatic degradation, and active excretion of cholesterol. This regulatory network implies that modulation of one of these steps (e.g., cholesterol synthesis or absorption) will cause compensatory effects that demand the use of combination therapy.

## 3 PHARMACOLOGICAL PRINCIPLE OF CHOLESTEROL ABSORPTION INHIBITION BEFORE EZETIMIBE

Based on the relevance of intestinal cholesterol absorption to the overall body cholesterol homeostasis as described before, drug-mediated blockade of this process appears to be a promising pharmacological approach. Thus, it is expected to reduce the incorporation of cholesterol into chylomicrons, which in turn diminishes the delivery of cholesterol to the liver [23, 33]. The consequences of reduced hepatic cholesterol supply include decreased LDL production, increased LDL clearance, thereby reducing serum cholesterol and, subsequently, may reduced atherogenesis.

Before ezetimibe was introduced to the market, this mechanism was already realized by bile-acid-binding resins. These drugs, including cholstyramine, colestipol, and colesevelam, are not absorbed from the intestine and reduce LDL-C by unspecific binding of biliary secreted bile acids, which interrupts their reabsorption from the ileum and leads to subsequent fecal elimination [23, 33]. This results in depletion of the bile acid pool, which is regenerated by catabolism of serum cholesterol, increased LDL receptor expression, and lowering of serum LDL-C of about 15–20%. Despite their proven reduction of LDL-C and CHD events, however, these compounds are associated with several unpleasant side effects such as constipation, steatorrhoea, bloating, flatulence, and hypovitaminosis, which limit patient compliance. Furthermore, this is encouraged by the relatively high doses that must be consumed for drug efficacy (5–20 g/day). Due to their unspecific binding capacity for anionic and lipophilic compounds, the absorption of many coadministered drugs, including furosemide, L-thyroxine, digoxin, and warfarin, is markedly influenced [34]. Therefore, bile-salt binding resins remain only as compounds of the second choice in cases of insufficient LDL-C lowering by statins or drug intolerance.

Another approach to lower intestinal cholesterol absorption is realized by some functional food products such as low-fat spreads or yogurt drinks containing esters of plant sterols and stanols such as sitosterol, campesterol, or stiostanol (e.g., Benecol, Becel) [21, 35]. Plant sterols, of which sitosterol and campesterol are the most abundant compounds, are naturally present in all vegetable food products and structurally related to cholesterol (campesterol = 24-metyl cholesterol, sitosterol = 24-ethyl cholesterol). They are not synthesized in the human body and exclusively of dietary origin. Their intestinal uptake is facilitated via the same NPC1L1-mediated mechanism as is known for cholesterol. The daily ingestion of plant sterol is comparable to cholesterol (300–500 mg), but the subsequent serum concentrations are usually less than 1% of that of serum cholesterol [21]. This poor intestinal absorption (5–15%) is caused by their high-affinity reverse transport into gut lumen by the intestinal and hepatic sterol efflux pumps ABCG5/G8. The absorption of plant stanols (5α-saturated plant sterols) is considerably lower than that of plant sterols, and these compounds are even more potent in cholesterol absorption inhibition. It is assumed that plant sterols and stanols may inhibit the micellar incorporation of cholesterol within the gut lumen. Hence, the serum LDL-C is reduced at about 10%. Several reports imply that a daily intake of less than 3 g stanols/sterols is capable to reduce significantly the serum cholesterol level [35]. Short-term studies estimated a reduction of CHD by plant sterols/stanols of about 20–25%, whereas no serious side effects were described. Consequently, the U.S. Food and Drug Administration (FDA) have concluded that foods containing plant sterols/stanols may reduce the risk of CHD. On the other side, increased serum levels of plant sterols as caused by enriched dietary products was reported to possibly increase the incidence of CHD [35].

Referring to this, more data are required to prove the long-term safety of these nutritional products. A novel development of plant stanol derivatives represents FM-VP4. This water-soluble agent is a mixture of the ascorbyl phosphate esters of campestanol and sitostanol, which are not absorbable from the gut and very potent inhibitors of the intestinal cholesterol absorption. FM-VP4 is now under clinical investigation in phase II trials [33].

The discovery of ezetimibe introduced a novel and very selective principle of cholesterol absorption inhibition, which opens a new door to the treatment of hypercholesterolemia. There are numerous other cholesterol absorption inhibitory drugs in development or even under clinical investigation, which will be highlighted in Section 7.

## 4 DISCOVERY AND DEVELOPMENT OF EZETIMIBE

The discovery program that finally led to ezetimibe was started in the early 1990s at the Schering-Plough laboratories and initially aimed at the identification and development of an ACAT inhibitor [36, 37]. The rationale for this was due to the fact that ACAT is apparently involved in cholesterol trafficking in rodents [38]. In cholesterol-fed hamsters, the investigated predecessors of ezetimibe were capable to reduce the dramatically increased levels of hepatic cholesterol ester, whereas the reduction of serum cholesterol was only less pronounced (e.g., 0–10%) [39, 40]. Furthermore, these studies assessed a considerable divergence between the in vivo activity and the in vitro inhibition of ACAT. Despite the lack of knowledge of the distinct molecular target, a broad chemical program was initiated that generated several ß-lactam scaffold structures [41]. However, none of the gained compounds was a potent ACAT inhibitor but paradoxically showed in vivo activities, including a modest to strong lowering effect on serum cholesterol in the cholesterol-fed hamster model (10–43%) [36]. These experiments led to the conclusion that the tested compounds inhibit cholesterol absorption via a novel mechanism of action, which is apart of ACAT. Already at this point, various data according to structure–activity relationship (SAR) were available but the target protein was still unknown. A comprehensive discussion of the related SAR is beyond the scope of this contribution but given elsewhere by excellent reviews [37, 42]. In consequence that there was no suitable in vitro assay available, and the ongoing research program was strictly dependent on in vivo experiments in appropriate animal models. It is worth noting that such an in vivo optimization process is a very challenging medicinal chemistry approach since the intrinsic activity of a compound is indistinguishable from differences in bioavailability and metabolism. Nevertheless, in 1994 this effort culminated in the discovery of the 2-azetidinone SCH48461, which was reported to be the most effective compound in vivo and reduced serum cholesterol in hamsters, rats, and rhesus monkeys at considerable low effective dose ($ED_{50}$) values of 1, 2, and 0.2 mg/kg per day, respectively [36, 43]. In the same study it was assessed that SCH48461 inhibited the accumulation of cholesterol and cholesterol esters in the serum of cholesterol-fed hamsters, whereas the specific ACAT inhibitor CI-976 only minimized serum cholesterol esters but had no effect on serum cholesterol levels. Thus, it was again suggested that cholesterol lowering by ezetimibe-related compounds such as SCH48461 represents a novel and so far undiscovered mechanism that is located upstream from ACAT within the cholesterol uptake machinery.

It was found that SCH48461 is subjected to extensive metabolism including oxidation and glucuronidation, making it difficult to identify exactly the active species. To address this issue, Van Heek et al. performed animal experiments using bile-duct-cannulated rats. In this study, the cholesterol-lowering activity of SCH48461 itself was compared with its metabolites containing bile, which was collected via a bile duct cannula from animals after previous SCH48461 administration (Fig. 2) [44]. Interestingly, SCH48461 was capable of inhibiting intestinal $^{14}$C-cholesterol absorption by 70%, whereas the "metabolite bile" inhibited the absorption by greater than 95%. These data implied that metabolite formation plays an important role in drug activity. The major metabolite and concurrently the most active moiety in bile was identified as the C4-phenolic glucuronide of the demethylated parent compound (SCH53695). Consistently, the in vivo activity of the accordant biliary metabolite fraction was identical to that of authentic SCH53695. The predominating metabolite remained mainly in the intestinal lumen when delivered intraduodenally, that is, the glucuronide is not absorbable from the upper small intestine. This led to the premature assumption by the authors that the glucuronide is the active moiety, which was refuted later. Altogether, this in vivo metabolism-based approach led to the conclusion that drug glucuronidation facilitates extensive drug recycling to the side of action via an enterohepatic circulation. In addition, it was clarified that dealkylation of the parent structure SCH48461 increases the cholesterol-lowering activity that directed research again to a substantial chemistry synthesis program, which finally yielded a potent second-generation cholesterol absorption inhibitor, SCH58235, later known as ezetimibe (Fig. 3) [45]. Thus, ezetimibe represents a premetabolized and optimized structure of SCH48461 that showed a 50 and

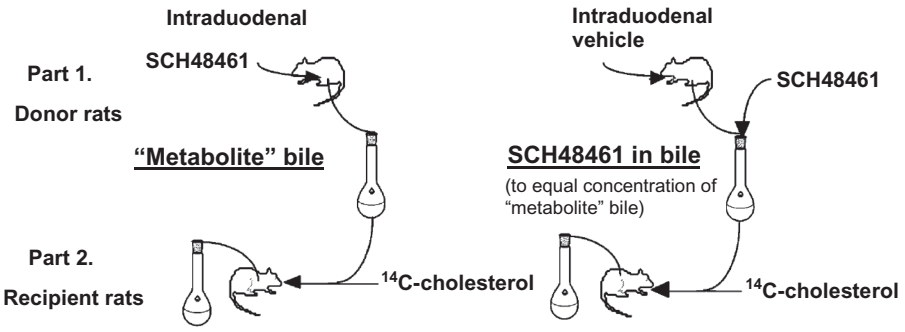


**Figure 2** Schematic overview of the performed in vivo metabolism-based approach for the discovery of more potent cholesterol absorption inhibitors [54]. Within the first part of the study, blank bile or "metabolite" bile, which was generated after oral administration of SCH48461, was collected via a bile duct cannula. In the second part, "metabolite" bile or blank bile mixed with original SCH48461 was given intraduodenally to bile-duct-cannulated and intact rats. The inhibitory effect on intestinal cholesterol absorption was observed by determination of $^{14}$C-cholesterol levels in liver and plasma. (Reproduced from [36], with kind permission of the publisher.)

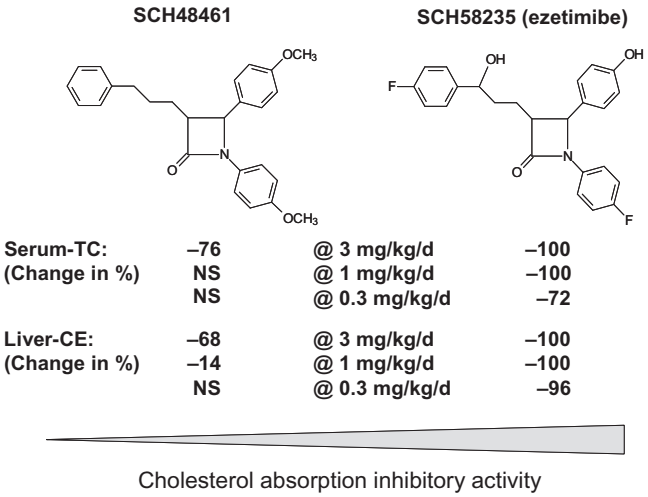


**Figure 3** Dose–response effects of ezetimibe and its predecessor SCH48461 on total serum cholesterol (TC) and liver cholesterol esters (CE) in the 7-day cholesterol-fed hamster model (NS, not significant) [45].

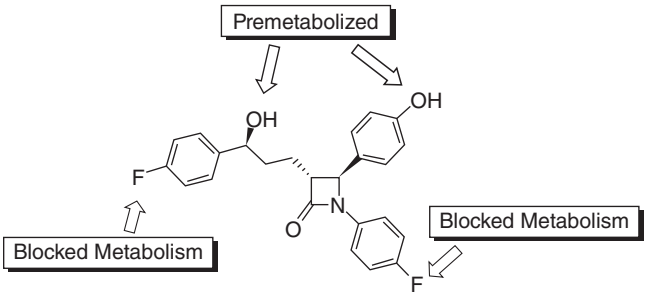


| $ED_{50}$ | SCH48461 | Ezetimibe |
|---|---|---|
| Hamster | 2.2 mg/kg | 0.04 mg/kg |
| Rat | 2.0 mg/kg | 0.03 mg/kg |
| Monkey | 0.2 mg/kg | 0.0005 mg/kg |
| Dog | 0.1 mg/kg | 0.007 mg/kg |

**Figure 4** Comparison of structure and in vivo effects of SCH48461 and its optimized derivative ezetimibe. (Reproduced from [36], with kind permission of the publisher.)

more than 400 times increased cholesterol-lowering activity in hamsters and rhesus monkey, respectively [44, 45]. Subsequently, ezetimibe proved its markedly cholesterol and LDL-lowering effect in various animal species such as mice, rats, hamsters, dogs, and rhesus monkeys with substantially increased activity compared to its predecessor (Fig. 4). Later on, numerous clinical trials proved the efficacy and safety of ezetimibe in LDL-C-lowering therapy in hypercholesterolemic patients.

In conclusion, the ezetimibe-associated research program enabled deeper insights into the physiological mechanisms of intestinal sterol absorption. The discovery of ezetimibe was an elegant exception of today's typical rational drug design, which is characterized by novel target-related technologies such as genomics, proteomics, high-throughput screening, and molecular structure modeling [46]. Thus, it is expected to promote a more directed approach toward drug discovery but requires detailed knowledge about the respective molecular target. On the contrary, it is often reported that rational drug design sometimes leads down a blind alley, that is, compounds with insufficient in vivo effects or with unacceptable pharmacokinetic properties are yielded. Referring to ezetimibe, neither the molecular receptor nor the explicit mechanisms of intestinal cho-

lesterol absorption were known. In accordance with this, none of the "modern" tools of drug discovery played a significant role in the development of this novel lipid-lowering agent. Nonetheless, the tremendous effort of creative animal experiments and SAR-based in vivo and chemical optimization enabled the establishment of this new and promising drug principle. Thus, one could learn from the ezetimibe story that the development of a successful pharmaceutical drug does not necessarily require accurate understanding of the molecular target but needs creative solutions and "old-fashioned" experimental know-how.

## 5 EZETIMIBE—A PORTRAIT

### 5.1 Physicochemical Properties

From the chemical point of view, ezetimibe (IUPAC name:1-(4-fluorophenyl)-3(*R*)-[3-(4-fluorophenyl)-3(*S*)-hydroxypropyl]-4(*S*)-(4-hydroxyphenyl)-2-azetidinone; empirical formula: $C_{24}H_{21}F_2NO_3$; molecular weight: 409.43 g/mol) belongs to the class of 2-acetidinones, which are characterized by a ß-lactam scaffold structure (Fig. 4). Thus, in comparison to several penicillins, ezetimibe appears to be unstable at alkaline pH values as caused by cleavage of the ß-lactam ring [47]. It is a white, crystalline powder that is freely soluble in methanol, ethanol, acetonitrile, and acetone but practically insoluble in water (0.007 g/L, pH 2–9). The drug is stable at ambient temperature and melts at about 163–166°C [48]. Because of its distinctive lipophilicity (calculated log P = 3.49) a bioconcentration factor (ratio of the substance concentration within a cell compared to the concentration of the surrounding water) was estimated to be above 250 at pH 2–8. Hence, it is suggested that ezetimibe may be easily absorbed from the intestinal lumen into the enterocytes. On the other side, its lack of water solubility counteracts this theoretical prediction of drug uptake whereby ezetimibe belongs to the class II compounds of the biopharmaceutical classification system (low solubility, high permeability) [49]. Due to its free phenolic OH group, ezetimibe is only a very weak acid ($pK_a = 9.72$), which means that under physiological conditions (pH 7.4) the drug is predominately present in its molecular form (99.5%), whereas in the gut (pH 8–9), it undergoes increasing ionization along the intestine (2–16%). The latter may limit passive drug absorption of ezetimibe.

### 5.2 Pharmacodynamics

As introduced above, ezetimibe represents the first compound of a new class of lipid-lowering drugs that selectively inhibits the intestinal absorption of biliary and dietary cholesterol without influencing the absorption of other dietary lipids such as triglycerides, fat-soluble vitamins, or bile acids [50]. Animal experiments conducted in rodents clearly verified that the mechanism of ezetimibe is not explained by inhibition of pancreatic lipases, bile acid sequesting, and without influence on the intestinal sterol-efflux pumps ABCG5/G8 and the acute hepatic or intestinal cholesterol synthesis [51–53]. Using radiolabeled molecules, it has been demonstrated that ezetimibe acts at the luminal membrane of the enterocytes along the small intestine [54]. The exact mode of its pharmacodynamic action was recently elucidated as inhibition of the putative sterol-transporting protein NPC1L1. NPC1L1 features nearly 50% amino acid homology to NPC, a protein that functions in intracellular cholesterol trafficking and that is defective in the hereditary cholesterol storage disease Niemann–Pick type C. The origin of the hypothesis that NPC1L1 may be involved in the intestinal cholesterol absorption came from a genomic-bioinformatics approach. In brief, human and rodent complementary deoxyribonucleic acid (cDNA) libraries were screened for anticipated target characteristics of a potential sterol transporter as transmembrane domains, extracellular signal sequences, and cholesterol sensing motifs that identified the credible candidate NPC1L1 [26]. As expected, the small intestine, as the known site of cholesterol absorption, showed the highest expression levels of NPC1L1. Thereafter, it was convincingly demonstrated by an animal study using NPC1L1-deficient mice that this protein may be the target of ezetimibe and the major determinant of intestinal cholesterol absorption. In the absence of NPC1L1, the mice exhibit a substantial reduction in absorbed cholesterol (70%), which was similar to the effects observed in ezetimibe-treated wild-type animals (68–73%) and intractable to further reduction by additional ezetimibe administration (Fig. 5) [26]. Together, these observations led to the suggestion that NPC1L1 is essential for the intestinal sterol uptake and may be the molecular target of ezetimibe. However, this study failed to demonstrate the direct binding of ezetimibe to NPC1L1 in overexpressing non-enterocyte cells.

Preclinical animal studies showed that ezetimibe effectively inhibited intestinal cholesterol absorption and reduced both plasma cholesterol and LDL-C in a dose-dependent manner in rhesus monkeys ($ED_{50} = 0.0005$ mg/kg day), dogs ($ED_{50} = 0.007$ mg/kg day), rats ($ED_{50} = 0.03$ mg/kg day), hamsters ($ED_{50} = 0.04$ mg/kg day), and mice ($ED_{50} = 0.7$ mg/kg day) [55]. In chow-fed dogs it was assessed that comedication of ezetimibe with HMG-CoA reductase inhibitors resulted in a synergistic and markedly increased serum cholesterol reduction compared to monotherapy

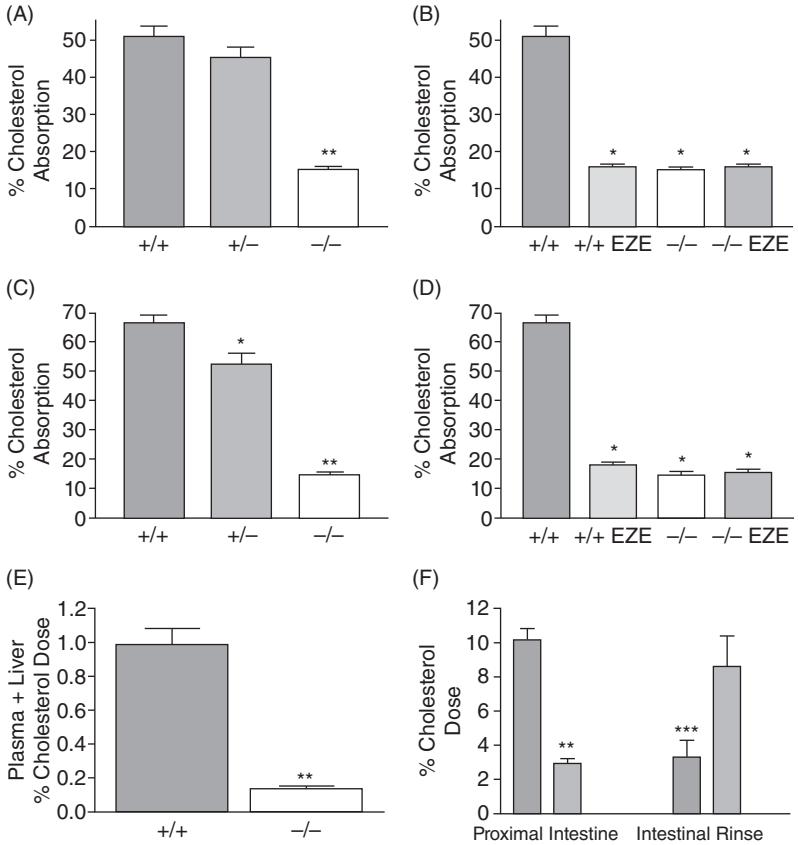


**Figure 5** Cholesterol absorption in homozygous (–/–) and heterozygous (–/+) NPC1L1 knockout mice as well as wild-type animals (+/+) with or without ezetimibe treatment (EZE, 10 mg/kg). Mice were fed a standard chow (A and B) or a diet containing 0.1% sodium cholate (C and D) and orally gavaged with $^{14}$C-cholesterol. Acute cholesterol uptake (2 h) was studied in (–/–) and (+/+) mice (E and F). Asterisks indicate significance levels [*$p < 0.001$ for comparison with (+/+) and (+/–) mice; **$p < 0.001$ and ***$p < 0.05$ for comparison with (+/+) only]. (Reproduced from [26], with kindly permission of the publisher.)

with either agent. Dose-finding studies including several hundred patients revealed a daily dose of 10 mg to be sufficient for maximal cholesterol-lowering effect [56]. In humans, this ezetimibe dose reduced cholesterol absorption by 54%, whereas its lowers LDL-C by about 15–20% [57–59]. On the other side, the increase of high-density lipoprotein cholesterol (HDL-C) was only low. Compared to statins, the LDL-C lowering effect is equivalent or even less pronounced [50]. Consequently, the relevance of ezetimibe monotherapy is limited (e.g., statin-intolerant patients). Furthermore, this is encouraged by the fact that ezetimibe treatment results in compensatory increase of endogenous cholesterol synthesis, which can be antagonized by the combination with statins [12, 60]. The combination therapy with statins leads to a dual and complementary cholesterol-lowering action since both major sources of body cholesterol are modulated (i.e., intestinal absorption and endogenous synthesis). Interestingly, statins were shown to increase the intestinal cholesterol uptake [32]. Therefore, both ezetimibe and statins break down each others compensatory mechanisms, which affect their therapeutic effects, that is, enhanced cholesterol synthesis by ezetimibe and the increase in cholesterol absorption caused by statins. Thus, the main indication for ezetimibe is a combination therapy with statins. Consequently, this combination provides an additional LDL-C reduction of 18–20%, which is distinctly above the effect that is generated by doubling the statin dose (~5%) [61–63]. For instance, when a combination of 10 mg ezetimibe with 80 mg simvastatin was used, the average LDL-C reduction was 60%, which is greater than can be attained

**TABLE 1 Effects of Ezetimibe and Statins on Serum Levels of Total Cholesterol, LDL-Cholesterol (LDL-C), Triglycerides (TG), and HDL-Cholesterol (HDL-C) in Patients with Primary Hypercholesterolemia[a]**

| Treatment | *N* | Total Cholesterol | LDL-C | TG | HDL-C |
|---|---|---|---|---|---|
| Ezetimibe 10 mg (EZE) | 65 | –14 | –20 | –5 | +4 |
| Atorvastatin 10 mg | 60 | –26 | –37 | –21 | +6 |
| Atorvastatin 20 mg | 60 | –30 | –42 | –23 | +4 |
| Atorvastatin 40 mg | 66 | –32 | –45 | –24 | +4 |
| Atorvastatin 80 mg | 62 | –40 | –54 | –31 | +3 |
| Atorvastatin 10 mg + EZE | 65 | –38 | –53 | –31 | +9 |
| Atorvastatin 20 mg + EZE | 62 | –39 | –54 | –30 | +9 |
| Atorvastatin 40 mg + EZE | 65 | –42 | –56 | –34 | +5 |
| Atorvastatin 80 mg + EZE | 62 | –46 | –61 | –40 | +7 |
| Ezetimibe 10 mg (EZE) | 61 | –13 | –19 | –11 | +5 |
| Simvastatin 10 mg | 70 | –18 | –27 | –14 | +8 |
| Simvastatin 20 mg | 61 | –26 | –36 | –18 | +6 |
| Simvastatin 40 mg | 65 | –27 | –38 | –24 | +6 |
| Simvastatin 80 mg | 67 | –32 | –45 | –23 | +8 |
| Simvastatin 10 mg + EZE | 67 | –32 | –46 | –26 | +9 |
| Simvastatin 20 mg + EZE | 69 | –33 | –46 | –25 | +9 |
| Simvastatin 40 mg + EZE | 73 | –40 | –56 | –32 | +11 |
| Simvastatin 80 mg + EZE | 65 | –41 | –58 | –31 | +8 |

[a]Values are given in mean % change from baseline [73].

with any statin alone (Table 1) [64]. Given this, the pharmacological effects of statins are markedly bolstered by ezetimibe, which requires only lower doses of statins and may lead to a smaller incidence of side effects. In experimental animal models ezetimibe profoundly minimized diet-induced atherosclerosis [65]. A similar outcome benefit is expected in humans as impressively shown for statins [8–10]. Until now, a confusing multiplicity of clinical trials employing ezetimibe and statins were published. However, outcome studies that investigate ezetimibe with or without statins are not available or still in progress, and no results are anticipated for several years [66].

Beyond the well-documented efficacy of ezetimibe in LDL-C lowering, also the enteral absorption of the phytosterols campesterol and sitosterol, molecules that are highly structurally related to cholesterol, is effectively reduced by ezetimibe [67]. In patients with sitosterolemia, a rare, autosomal recessively inherited sterol storage disease characterized by excessive absorption of these plant sterols, ezetimibe decreased the serum concentrations of sitosterol and campesterol by 21 and 24%, respectively [68]. Thus, ezetimibe offers a valuable therapeutic option for this potential life-threatening disease. Since these plant sterols are not endogenously synthesized as is the case for cholesterol, their reduction by ezetimibe demonstrates a direct drug effect that can be used as a surrogate for ezetimibe efficacy.

An open question associated with the pharmacodynamics of ezetimibe is whether the parent compound or its glucuronide represents the active moiety and whether the pharmacologic binding site of the postulated molecular target (NPC1L1) is located within the enterocytes or at the apical (luminal) membrane.

With regard to the latter question, Davies et al. detected that the NPC1L1 protein is expressed intracellulary in transfected HepG2 cells [69]. Contradictory to this, in vitro studies by the manufacturer described an apical target localization to emphasize the nonsystemic mode of ezetimibe action [70, 71]. Finally, Yu et al. clarified these diverging findings [72]. They found that NPC1L1 undergoes subcellular translocation in depending on the cellular content of cholesterol, that is, dietary cholesterol excess leads to intracellular existence of NPC1L1, which is the case in a high-fat western diet. From the current point of view only unchanged ezetimibe appears to be the active principle, which is discussed in detail in Section 5.3.

Ezetimibe was shown to be well tolerated in monotherapy and in combination with statins in thousands of patients [73]. Adverse drug effects were reported in ≥2% of patients and thus at a greater incidence compared to placebo including fatigue, abdominal pain, diarrhea, infections, and muscle pain. Additionally, very recent postmarketing studies assessed

that for ezetimibe when coadministered with statins, the incidence of significantly elevated liver transaminases was more markedly increased than in patients treated with HMG-CoA inhibitors alone (1.3 vs. 0.4%). Also very rare cases of myopathy and rhabdomyolysis, even during ezetimibe monotherapy, were described [74, 75].

After all, ezetimibe inhibits intestinal cholesterol and plant sterol absorption, which leads to clinically relevant reduction of serum LDL-C and plant sterols and offers a novel therapeutic approach for patients with hyercholesterolemia or sitosterolemia. The mainly well-tolerated combination of ezetimibe with statins inhibits the compensatory increase in cholesterol synthesis and leads to additional reduction in serum cholesterol levels.

### 5.3 Pharmacokinetics

After oral administration, ezetimibe is rapidly absorbed into the enterocytes after liberation from its immediate release dosage form (Zetia, Ezetrol) along the upper intestine, that is, the site of highest NPC1L1 expression and cholesterol absorption. Even a few minutes after intraduodenal delivery, the drug was detected in portal blood flow of rats [54]. Despite its high lipophilicity and its nearly water insolubility, it is expected that ezetimibe is rather poorly absorbed from the gastrointestinal tract. Accordingly, the absolute oral bioavailability has not been determined due to the absence of a parenteral application form. Furthermore, for the systemic ezetimibe exposure, no dose proportionality was assessed, which implies an involvement of active carriers that may limit intestinal drug absorption [76]. With regard to this, ezetimibe was identified as an in vivo and in vitro substrate of the multidrug transporting efflux pump P-glycoprotein (P-gp) [77]. However, once in the enterocytes, ezetimibe undergoes nearly complete intestinal glucuronidation to its major phenolic glucuronide (SCH60663), which accounts for approximately 90% of the systemic drug exposure (Fig. 6) [54, 78]. Additionally, traces of a benzylic glucuronide (SCH488128) were assessed. Later on, the uridine diphosphate (UDP)-glucuronosyltransferases (UGTs) 1A1, 1A3, and 2B15 were demonstrated to be the relevant isoenzymes for human metabolism [79]. The parent compound and its conjugated metabolite were shown to be subjected to hepatic uptake from the portal blood and subsequent biliary secretion [54]. The multidrug resistance-associated protein (MRP) 2, an ATP-consuming efflux transporter expressed in liver, kidney, and intestine, was identified as the major determinant for intestinal and heptobiliary secretion of the glucuronide [80]. Consequently, Mrp2 deficiency in rats led to diminished glucuronide elimination via feces but dramatically increased serum levels of the respective phase II metabolite. Altogether this creates an enterohepatic recirculation that enables repeated delivery of ezetimibe predominately in its glucuronized form back to the

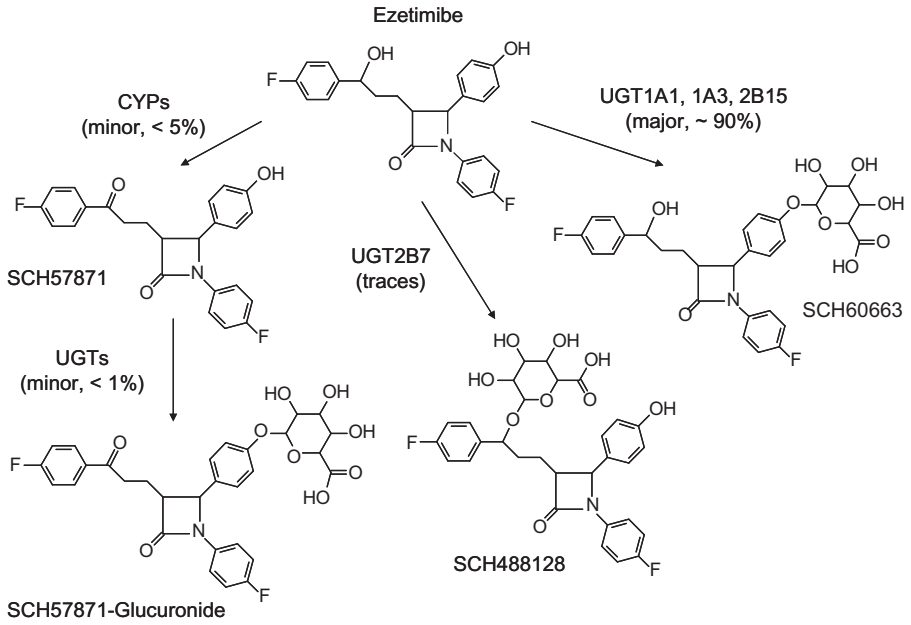


**Figure 6** Biotransformation pattern of ezetimibe in humans (CYP, cytochrome P450 enzyme; UGT, UDP- glucuronosyltransferase).

site of action an in turn results in a long-lasting sterol-lowering effect and a high terminal half-life of about 20–24 h [76, 78]. Thus, it is sufficient to dose the agent once daily. Using a mathematical approach based on a population of a pharmacokinetic model, Ezzet et al. showed that the recycled amount of ezetimibe accounts for nearly 15–20% of the administered dose [81]. Nevertheless, the glucuronide fraction, which is secreted into the intestinal lumen, is not directly capable to inhibit the NPC1L1 protein because it is not absorbable from the upper parts of the gut due to its hydrophilicity (log $P = 1.18$) and ionic character ($pK_a = 2.78 \rightarrow$ pH 8–9, >99.9%) as well as the intracellular localization of the NPC1L1 protein [69]. This theoretical assumption was supported by an animal study [54]. In the lower intestine (i.e., terminal ileum and colon), the metabolite is degraded by cleavage enzymes from the bacterial flora to the parent compound. Thus, unchanged ezetimibe is now available for reabsorption from the proximal colon but not for immediate cholesterol absorption inhibition because this site of drug release is not the site of intestinal cholesterol absorption. Consequently, the expression level of NPC1L1 is only marginal in this intestinal segment [26, 69]. Therefore, the systemic drug concentration, which continuously provides ezetimibe to the enterocytes via the venous blood and not exclusively the intestinal drug amounts, dictate the drug efficacy of ezetimibe [56]. The latter seems only relevant during the initial period drug absorption, whereas unexceptional free ezetimibe from the digested tablet is on hand and invariably responsible for NPC1L1 modulation from the luminal site of the gut.

The nonsecreted portion of the drug appears via the portal vein after passing through the liver in the central compartment, which produces a first serum maximum within one hour. In the blood, ezetimibe and the glucuronide are highly bound to plasma proteins (>90%). The serum concentration–time profiles of ezetimibe ($C_{max} = 3$–$10$ ng/mL) and its glucuronide ($C_{max} = 45$–$75$ ng/mL) are characterized by multiple peaks that reliably appear at the time of food ingestion (Fig. 7). This is most likely caused by food-triggered emptying of the gall bladder and boluslike translocation of glucuronide containing chyme to the hydrolytic environment of the colon by the meal-induced gastroileocecal reflex [77, 82]. Thus, cumulative amounts of repetitively released ezetimibe are reabsorbed from the colon, which is enforced by the low expression levels of P-gp, MRP2, and the UGTs in this region of the bowel. Thereby, the second and third serum peaks of unchanged ezetimibe are even higher than the initial one (Fig. 7).

In general, the pharmacokinetics of ezetimibe is characterized by a high intersubject variability [35–60% for the area under the curve (AUC) values] the reasons of which are so far unknown. In a study of eight healthy subjects using radiolabeled $^{14}$C-ezetimibe the fecal and renal elimination 240 h after drug administration was approximately 78 and 11% of the administered 20-mg dose, respectively [78]. In urine, predominately the glucuronide (95%) is excreted, whereas in stool it is the unchanged ezetimibe (85%) as caused by bacterial hydrolysis [77, 83]. Hence, a substantial part of the dose (~10%) undergoes unknown metabolism and elimination pathways or remains in the body.

In contrast to statins, ezetimibe is only negligibly metabolized by CYP enzymes, which accounts for turnover of below 5% of the applied dose leading to a ketone structure (Fig. 6). In accordance with this, drug

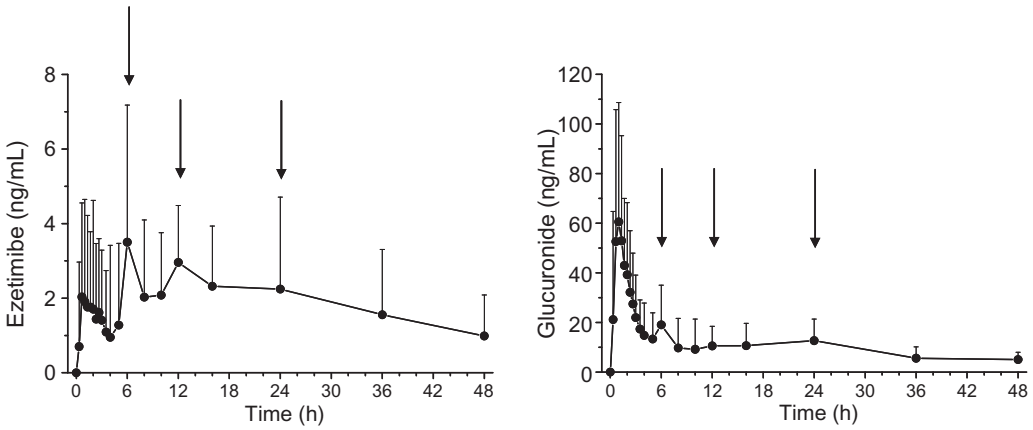


**Figure 7** Serum concentration–time profiles of ezetimibe (left) and its glucuronide (right) after a single oral dose of 20 mg in eight healthy subjects. Arrows indicate time of food ingestion [83].

interaction studies with several well-established CYP substrates including caffeine (CYP1A2), tolbutamide (CYP2C8), warfarin (CYP2C9), dextromethorphan (CYP2D6), midazolam (CYP3A4), and cimetidin (CYP inhibitor) revealed no influence on the pharmacokinetics of each comedicated drug [76, 84]. This lack of oxidative metabolism by the CYP pathway suggested safety when combined with other agents such as statins that undergo mostly intensive CYP-mediated biotransformation. On the other side, unclear pharmacokinetic drug interactions with cyclosporine, gemfibrozil, and fenofibrate were described that led in part to several-fold increased serum concentrations of ezetimibe [76]. Very recently, inhibition of the efflux transporters P-gp and MRP2 by the mentioned interaction partners was elucidated as the molecular mechanism for these drug interactions. In accordance with this, transporter modulation caused markedly increased ezetimibe bioavailability, whereas induction of P-gp and MRP2 led to dramatically minimized serum levels of ezetimibe and in parallel to significantly reduced sterol-lowering effects [77, 83]. Hence, one could hypothesize that upregulation of these active elimination carriers by other known inducers such as St. John's wort, carbamazepine, or dexamethasone may deplete ezetimibe serum levels and its inhibitory effects on the intestinal sterol uptake. On the contrary, inhibition of P-gp and MRP2 by well-established substrates and/or modulating drugs such as statins, human immunodeficiency virus (HIV) protease inhibitors, verapamil, or ciclosporine most likely augment the systemic drug concentrations but may shorten the mean residence time of the drug. Accordingly, the combination of ezetimibe with simvastatin, lovastatin, and rosuvastatin caused elevated serum concentrations of ezetimibe and the respective statin by about 10–20% [85–87].

However, the clinical importance of both scenarios needs further clinical investigation. Nevertheless it cannot be excluded that enhanced serum concentrations due to pharmacokinetic drug–drug interactions may increase the risk for the described myotoxicity. Furthermore, nothing is known about the in vivo meaning of modulated ezetimibe glucuronidation which was shown in vitro by diclofenac and flunitrazepam [79].

Finally, ezetimibe possesses unique pharmacokinetic characteristics including extensive glucuronidation but practically no CYP-mediated biotransformation, which minimize its potential for drug–drug interactions with CYP substrates. However, the significant involvement of the efflux pumps P-gp and MRP2 in ezetimibe disposition is responsible for other pharmacokinetic interactions the impact of which needs further clarification.

## 6 NPC1L1—THE REAL TARGET OF EZETIMIBE?

The identification of the molecular basis of the ezetimibe effects has been an area of intensive research since its discovery in the 1990s. Modulation of pancreatic enzymes and inhibition of the hitherto described candidate proteins SR-BI, ABCA1, and ABCG5/G8 were excluded as potential mode of ezetimibe action [51, 53, 88].

The already mentioned report by Altmann et al. demonstrated very cogently that NPC1L1 might be the molecular target of ezetimibe and in turn that this protein may be the key player for intestinal cholesterol absorption [26]. In contrast to this study, during the ongoing research work, it was successfully shown that ezetimibe directly binds to HEK cells expressing NPC1L1 [70]. This is supported by a very recent paper from the Schering-Plough laboratories that detected a good correlation between the assessed in vitro binding affinity and the known in vivo potency of ezetimibe in different animal species. In addition, even SAR considerations could be addressed in this study [89].

In humans, functional relevant polymorphisms in the *NPC1L1* gene were described that were shown to affect the cholesterol-lowering potency of ezetimibe and thus underline that sterol lowering of ezetimibe seems to be exclusively realized by modulation of this putative sterol uptake carrier [90]. Also functional in vitro assays, which studied the cellular cholesterol and sitosterol uptake into Caco-2 cells in dependence on NPC1L1 expression and ezetimibe-mediated inhibition, clearly highlighted NPC1L1 as the molecular target of ezetimibe [91]. A further strong argument for this target represents an impressive study by Davis et al. that could show that deficiency of NPC1L1 in apoE null mice prevents atherosclerosis [92]. This was shown in the same manner for the treatment with ezetimibe [65].

Finally, the coordinate transcriptional regulation of NPC1L1 within the whole-body cholesterol homeostasis indicates an outstanding role of the NPC1L1 protein for cholesterol absorption. Referring to this, cholesterol and other ligands of the "lipid sensors" LXR and peroxisome proliferators-activated receptor (PPAR) δ were shown to suppress NPC1L1 expression, which limits excessive cholesterol uptake from the intestine [67, 93, 94]. Interestingly, it was also reported that ezetimibe itself is capable to downregulate NPC1L1 [95]. On the other side, inhibition of endogenous cholesterol synthesis by HMG-CoA inhibitors leads to an increased intracellular need of cholesterol that consequently results in a compensatory enhanced expression of NPC1L1 and subsequent higher intestinal cholesterol absorption [31, 32]. Given the high degree of homology

of NPC1L1 and other sterol-sensing domain-containing proteins (e.g., HMG-CoA reductase, ABCG5/G8), the emerging role of these proteins in cellular sterol trafficking is suggested. However, the distinct molecular regulation by nuclear receptors controlling NPC1L1 expression remains elusive.

Despite these investigations, some other possible target proteins and molecular mechanisms have been described for ezetimibe. In parallel to the discovery of NPC1L1 as the potential ezetimibe target, it was speculated that annexin-2 and caveolin-1, both cytoskeletal proteins, might be involved in concert in cholesterol absorption. In vitro, ezetimibe was capable to prevent the formation of the heterocomplex of both proteins, which was assumed to mediate the intracellular cholesterol transport [96]. Nevertheless, caveolin-1 knockout mice have a normal percentage of cholesterol absorption that disqualifies this protein complex to be of major relevance for the ezetimibe effects [97].

In another manufacturer-independent investigation study using membrane-impermeable photoreactive ezetimibe derivatives, it was demonstrated that ezetimibe binds to a 145-kDa integral membrane protein of the brush border membrane of small intestinal enterocytes identified as aminopeptidase N (CD13) [98]. However, this enzyme, which is normally involved in the endocytosis of viruses, showed no direct cholesterol absorption. Thus, it remains uncertain whether aminopeptidase N is an in vivo target of ezetimibe or is even involved in the cholesterol uptake cascade.

In addition, Seedorf et al. revealed that ezetimibe inhibits the in vitro uptake of oxidized LDL-C into human macrophages in a concentration-dependent manner, which could represent an interesting systemic mode of ezetimibe action that may prevent the formation of atherogenic foam cells [99].

In conclusion, there is ample evidence by numerous in vitro investigations, animal studies, and pharmacogenomic data from humans that NPC1L1 is indeed the molecular target of ezetimibe and the rate-limiting carrier for the intestinal absorption of cholesterol and plant sterols.

## 7 DRUGS OF THE FUTURE

Since it was demonstrated that inhibition of intestinal cholesterol absorption appears to be a promising pharmacological drug target to lower serum cholesterol, in parallel to the discovery of ezetimibe several other auspicious compounds affecting different steps within the intestinal cholesterol uptake cascade were developed. In accordance with the predominate use of ezetimibe in combination therapy, these compounds may be of interest as further options to achieve a higher degree of LDL-C lowering. This Section summarizes the recent proceedings in this area.

### 7.1 ACAT Inhibitors

As described in Section 2, the acyl CoA:cholesterol acyl transferase (ACAT) is a membrane-bound enzyme involved in intracellular esterification of cholesterol. ACAT is located in the rough endoplasmic reticulum of various tissues such as liver, intestine, macrophages, and blood vessels. Two isoforms, ACAT1 and ACAT2, were identified [100]. The latter is exclusively expressed in the liver and the intestine and might be of great significance for pharmacological intervention in order to block cholesterol uptake from the gut. Beside re-esterification of dietary cholesterol in the enterocytes, hepatic ACAT2 is involved in the secretion of very low density lipoprotein (VLDL) particles into the systemic circulation. Mice lacking ACAT2 showed nearly halved serum cholesterol levels compared to their wild-type counterparts [101]. Otherwise, ACAT1 in the vessel wall is responsible for macrophage-mediated storage of cholesterol esters. Thus, inhibition of ACAT was expected (1) to lower intestinal cholesterol absorption by limiting the assembly of chylomicrons, (2) to decrease the appearance of VLDL cholesterol in the blood, and (3) to diminish cholesterol accumulation into macrophages, which is critical for the development of plaque lesions. Numerous ACAT-inhibiting candidates, including DuP-128, F-1349, CL-277082, pactimibe (CS-505), and avasimibe (CI-1011), were evaluated. ACAT inhibitors were shown to decrease plasma cholesterol in various animal models [23, 102]. However, due to the low bioavailability of most ACAT inhibitors, these drugs seem only to interrupt the intestinal cholesterol absorption and in turn demonstrated only marginal effects on serum cholesterol in humans [33]. Furthermore, the unselective modulation of both isoenzymes by current ACAT inhibitors led to significant toxicological effects as emesis, changes in red blood cell morphology, and hepatic and adrenotoxicity. Substantial clinical data are only available for avasimibe and pactimibe. Both drugs are orally available but nonspecific inhibitors of ACAT. Avasimibe-treated minipigs exhibited a markedly increased fractional clearance rate of postprandial triglyceride-rich lipoproteins (VLDL), whereas an LDL-C-lowering effect was only observed at markedly higher doses. In humans, avasimibe reduced triglycerides by 16–23% but neither affected total cholesterol or LDL-C concentrations nor favorably altered coronary atherosclerosis [103, 104]. Similar results were

demonstrated for pactimibe by a very recent study including over 400 patients with coronary disease [105]. Taken together, current ACAT inhibitors failed to demonstrate substantial effects on LDL-C levels and atherosclerosis in humans. Despite the promising molecular target of ACAT and the fact that selective ACAT2 inhibitors may feature improved pharmacologic characteristics, one might raise the question if this class of drugs could even generate clinically significant effects.

### 7.2 MTP Inhibitors

The microsomal triglyceride transfer protein (MTP) is a heterodimer protein that catalyzes the transfer of triglycerides, cholesterol esters, and apolipoproteins across membranes and is therefore critical for the assembly of chylomicrons in the enterocytes and of VLDL particles in the liver. The exact molecular mechanism of lipid binding and transport by MTP remains so far uncertain. Abetalipoproteinemia is a rare inherited disease caused by genetic defects of this protein and characterized by hypocholesterolemia, malabsorption of fat-soluble vitamins, and the absence of apoB-containing lipoprotein particles. The inhibition of MTP is expected to reduce intestinal cholesterol absorption by modulating chylomicron secretion into lymph and to decrease hepatic VLDL secretion leading to decreased plasma levels of VLDL-C, LDL-C, and triglycerides. Several experimental MTP inhibitors such as BMS-201038 (lomitapide), BMS-212122, BMS-200150, JTT-130, CP-346086, and implitapide (BAY-139953) have been developed and investigated in vitro and in animal studies [23, 33, 106]. Nearly all of these compounds showed distinctive reduction of serum cholesterol and triglycerides in small animal models. However, only very little is known about the clinical effects of MTP inhibitors in humans. Implitapide, which was shown to prevent atherosclerosis in apoE knockout mice, was studied in 188 patients with primary hypercholesterolemia in a placebo-controlled, randomized, double-blind trial. It was shown that implitapide is capable to reduce LDL-C in a dose-dependent manner (8–55%) [23]. On the other side, this drug frequently caused adverse side effects as increased liver transaminases and gastrointestinal symptoms, especially at high doses, and consequently this compound is no longer in development. For CP-346086, an even more potent and orally active inhibitor of MTP, a substantial dose-dependent lowering of total serum cholesterol (47%), LDL-C (72%), and triglycerides (75%) with no influence on HDL was revealed in healthy human volunteers [107]. Treatment of patients with homozygous familial hypercholesterolemia with BMS-201038 (AEGR-733) led to a marked reduction of LDL-C and apoB-containing lipoproteins by up to 51 and 56% [106]. Under consideration of these observations, MTP-modulating drugs seem to be most suitable in patients with hepatic hypersecretion of apoB, including the metabolic syndrome or type II diabetes mellitus. Despite the impressive effects on serum lipids, the clinical impact of these drugs for the treatment of hypercholesterolemia remains still unclear. However, inhibition of MTP by different drugs was associated, as expected, with malabsorption of fat-soluble vitamins and hepatic steatosis, which is in line with reports of patients with ABL and MTP-deficient animals. The mentioned safety aspects have to be addressed in furture studies in order to exclude long-term toxicity (e.g., avitaminosis and hepatic and intestinal fat accumulation).

### 7.3 Bile Acid Reabsorption Inhibitors (BARI)

Due to the fact that bile acids represent a substantial part of the whole-body cholesterol pool, interruption of their intestinal reabsorption appears to be a promising pharmacological approach that is so far targeted by bile-acid-binding resins as noted before. Bile acids are actively taken up from the intestinal lumen of the ileum by the apical sodium-dependent bile salt transporter (ASBT [IBAT—ileal bile acid transporter]). Thus, specific inhibition of ASBT represents an alternative option to minimize bile acid reuptake. Interruption of this enterohepatic recirculation activates endogenous bile salt resynthesis in the liver by upregulation of cholesterol 7α-hydroxylase (CYP7A1), which mediates the rate-limiting turnover of systemic cholesterol to bile salts. Consequently, LDL receptors at the hepatocyte cell surface are upregulated to increase lipoprotein catabolism, and, subsequently, the serum cholesterol concentrations decrease. In addition, hepatic VLDL secretion is reduced, whereas the HMG-CoA reductase-catalyzed cholesterol synthesis is increased. As described for the other novel classes of cholesterol absorption inhibitors, also for BARI such as SC-435, S-8921, 264-W94, TA-7552, and HMR-1453, an explicit cholesterol-lowering effect was demonstrated in small animals [33, 108]. As expected, animal treatment with BARI was associated with increased fecal bile acid excretion and concomitantly enhanced hepatic activity of CYP7A but a decrease of serum LDL and total cholesterol levels. Moreover, combination of BARI with HMG-CoA reductase inhibitors elevated the LDL-C-lowering effects due to complementary modes of action. However, until now no clinical data are available for the use of BARI in humans. It was hypothesized that the nearly doubled fecal loss of bile acids, which was assessed in animal studies, may cause gastrointestinal side effects. Most likely this could be antagonized with low dose of bile-acid-binding resins.

### 7.4 Other NPC1L1 Inhibitors

During the development of ezetimibe, several structurally related compounds were discovered [37, 109]. Using in vitro binding experiments in NPC1L1-expressing cells and animal studies, some of these candidates showed an even higher potency than the parent compound [37, 89, 110]. It is worth noting that the detected in vitro binding affinity of these compounds represents no direct surrogate for their expected in vivo efficacy because the established NPC1L1 overexpressing cell lines differ markedly from the situation in human enterocytes (e.g., apical localization NPC1L1). Furthermore, using such in vitro assays, the pharmacokinetic fate of each compound is not considered but of immense impact on *in vivo* efficacy as discussed for ezetimibe. Nevertheless, it is expected that Schering-Plough and other companies will follow the proven strategy of NPC1L1 inhibition and design more potent cholesterol absorption inhibitors. Ultimately, the future work in this area will dictate how ezetimibe therapy is accepted into clinical practice and if ezetimibe in mono- or combination therapy will yield clinically significant outcome benefits.

## 8 CONCLUSIONS

Ezetimibe is the first compound of lipid-lowering drugs that selectively inhibits the intestinal cholesterol absorption. Modulation of the sterol uptake carrier NPC1L1 appears to be the molecular mode of ezetimibe action. The associated research project has revolutionized our current understanding of intestinal cholesterol trafficking. The drug was marketed without any knowledge of its distinct molecular target. Accordingly, the discovery of ezetimibe was an explicit exception to today's drug design. Nevertheless, despite the fact that no in vitro assay was available, by using very sophisticated animal studies and broad chemical synthesis programs the optimal structure of ezetimibe was acquired. In clinical trials, ezetimibe was shown to reduce serum LDL-C by about 20%. In combination with statins an additional cholesterol-lowering effect of 15–20% was generated. Ezetimibe possesses unique pharmacokinetic properties, including comprehensive glucuronidation, efflux transport by P-gp and MRP2, and enterohepatic recirculation but no CYP-mediated metabolism. Thus, ezetimibe exhibits a long terminal half-life of nearly 24 h, which enables a once-daily dosing regimen. Despite several pharmacokinetic drug–drug interactions reported, ezetimibe appears to be a safe and well-tolerated drug. Although ezetimibe markedly lowers intestinal cholesterol absorption and serum LDL-C in hypercholesterolemic patients, especially in combination with statins, significant clinical benefits in the form of a reduced morbidity and mortality must be proven. The promising strategy of cholesterol absorption inhibition is focused on numerous new classes of compounds that are under intensive investigation but still far from being introduced to the market.

## REFERENCES


1. Rosamond W, Flegal K, Friday G, Furie K, Go A, Greenlund K, et al. (2007). Heart disease and stroke statistics—2007 update: A report from the American Heart Association Statistics Committee and Stroke Statistics Subcommittee. *Circulation 115*, e69–171.
2. Anderson KM, Castelli WP, Levy D (1987). Cholesterol and mortality. 30 years of follow-up from the Framingham study. *JAMA 257*, 2176–2180.
3. Levine GN, Keaney JF Jr, Vita JA (1995). Cholesterol reduction in cardiovascular disease. Clinical benefits and possible mechanisms. *N Engl J Med 332*, 512–521.
4. Law MR, Wald NJ (1994). An ecological study of serum cholesterol and ischaemic heart disease between 1950 and 1990. *Eur J Clin Nutr 48*, 305–325.
5. Stamler J, Wentworth D, Neaton JD (1986). Is relationship between serum cholesterol and risk of premature death from coronary heart disease continuous and graded? Findings in 356,222 primary screenees of the Multiple Risk Factor Intervention Trial (MRFIT). *JAMA 256*, 2823–2828.
6. Choy PC, Siow YL, Mymin D, O K (2004). Lipids and atherosclerosis. *Biochem Cell Biol 82*, 212–224.
7. Grundy SM, Cleeman JI, Merz CN, Brewer HB Jr, Clark LT, Hunninghake DB, et al. (2004). Implications of recent clinical trials for the National Cholesterol Education Program Adult Treatment Panel III guidelines. *Circulation 110*, 227–239.
8. Pedersen T, Kjekshus J, Berg K, Haghfelt T, et al. Randomised trial of cholesterol lowering in 4444 patients with coronary heart disease: The Scandinavian Simvastatin Survival Study (4S). *Lancet 344*, 1383–1389 (1994).
9. Prevention of cardiovascular events and death with pravastatin in patients with coronary heart disease and a broad range of initial cholesterol levels. The Long-Term Intervention with Pravastatin in Ischaemic Disease (LIPID) Study Group. *N Engl J Med 339*, 1349–1357 (1998).
10. Shepherd J, Cobbe SM, Ford I, Isles CG, Lorimer AR, MacFarlane PW, et al. (1995). Prevention of coronary heart disease with pravastatin in men with hypercholesterolemia. West of Scotland Coronary Prevention Study Group. *N Engl J Med 333*, 1301–1307.
11. Foley KA, Simpson RJ Jr, Crouse JR III, Weiss TW, Markson LE, Alexander CM (2003). Effectiveness of statin titration on low-density lipoprotein cholesterol

goal attainment in patients at high risk of atherogenic events. *Am J Cardiol 92*, 79–81.

12. Ballantyne CM (2005). Rationale for targeting multiple lipid pathways for optimal cardiovascular risk reduction. *Am J Cardiol 96*, 14K–19K.
13. Thompson PD, Clarkson P, Karas RH (2003). Statin-associated myopathy. *JAMA 289*, 1681–1690.
14. Andrews TC, Ballantyne CM, Hsia JA, Kramer JH (2001). Achieving and maintaining National Cholesterol Education Program low-density lipoprotein cholesterol goals with five statins. *Am J Med 111*, 185–191.
15. Neuvonen PJ, Niemi M, Backman JT (2006). Drug interactions with lipid-lowering drugs: Mechanisms and clinical relevance. *Clin Pharmacol Ther 80*, 565–581.
16. Staffa JA, Chang J, Green L (2002). Cerivastatin and reports of fatal rhabdomyolysis. *N Engl J Med 346*, 539–540.
17. Kesaniemi YA, Miettinen TA (1987). Cholesterol absorption efficiency regulates plasma cholesterol level in the Finnish population. *Eur J Clin Invest 17*, 391–395.
18. Gylling H, Miettinen TA (1995). The effect of cholesterol absorption inhibition on low density lipoprotein cholesterol level. *Atherosclerosis 117*, 305–308.
19. Hui DY, Howles PN (2005). Molecular mechanisms of cholesterol absorption and transport in the intestine. *Semin Cell Dev Biol 16*, 183–192.
20. Turley SD, Dietschy JM (2003). Sterol absorption by the small intestine. *Curr Opin Lipidol 14*, 233–240.
21. von Bergmann K, Sudhop T, Lutjohann D (2005). Cholesterol and plant sterol absorption: Recent insights. *Am J Cardiol 96*, 10D–14D.
22. Lammert F, Wang DQ (2005). New insights into the genetic regulation of intestinal cholesterol absorption. *Gastroenterology 129*, 718–734.
23. Sudhop T, von Bergmann K (2002). Cholesterol absorption inhibitors for the treatment of hypercholesterolaemia. *Drugs 62*, 2333–2347.
24. Sudhop T, Lutjohann D, von Bergmann K (2005). Sterol transporters: Targets of natural sterols and new lipid lowering drugs. *Pharmacol Ther 105*, 333–341.
25. Plosch T, Kosters A, Groen AK, Kuipers F (2005). The ABC of hepatic and intestinal cholesterol transport. *Handb Exp Pharmacol 170*, 465–482.
26. Altmann SW, Davis HR Jr, Zhu LJ, Yao X, Hoos LM, Tetzloff G, et al. (2004). Niemann-Pick C1 Like 1 protein is critical for intestinal cholesterol absorption. *Science 303*, 1201–1204.
27. Berge KE, Tian H, Graf GA, Yu L, Grishin NV, Schultz J, et al. (2000). Accumulation of dietary cholesterol in sitosterolemia caused by mutations in adjacent ABC transporters. *Science 290*, 1771–1775.
28. Kratz M (2005). Dietary cholesterol, atherosclerosis and coronary heart disease. *Handb Exp Pharmacol 170*, 195–213.
29. Ory DS (2004). Nuclear receptor signaling in the control of cholesterol homeostasis: Have the orphans found a home? *Circ Res 95*, 660–670.
30. Chawla A, Repa JJ, Evans RM, Mangelsdorf DJ (2001). Nuclear receptors and lipid physiology: Opening the X-files. *Science 294*, 1866–1870.
31. Alrefai WA, Annaba F, Sarwar Z, Dwivedi A, Saksena S, Singla A, et al. (2007). Modulation of human Niemann-Pick C1-like 1 gene expression by sterol: Role of sterol regulatory element binding protein 2. *Am J Physiol Gastrointest Liver Physiol 292*, G369–G376.
32. Miettinen TA, Gylling H (2003). Synthesis and absorption markers of cholesterol in serum and lipoproteins during a large dose of statin treatment. *Eur J Clin Invest 33*, 976–982.
33. Burnett JR, Huff MW (2006). Cholesterol absorption inhibitors as a therapeutic option for hypercholesterolaemia. *Expert Opin Investig Drugs 15*, 1337–1351.
34. Bays H, Stein EA (2003). Pharmacotherapy for dyslipidaemia—Current therapies and future agents. *Expert Opin Pharmacother 4*, 1901–1938.
35. Gylling H, Miettinen TA (2005). The effect of plant stanol- and sterol-enriched foods on lipid metabolism, serum lipids and coronary heart disease. *Ann Clin Biochem 42*, 254–263.
36. Clader JW (2004). The discovery of ezetimibe: A view from outside the receptor. *J Med Chem 47*, 1–9.
37. Clader JW (2005). Ezetimibe and other azetidinone cholesterol absorption inhibitors. *Curr Top Med Chem 5*, 243–256.
38. Schnitzer-Polokoff R, Compton D, Boykow G, Davis H, Burrier R (1991). Effects of acyl-CoA: Cholesterol *O*-acyltransferase inhibition on cholesterol absorption and plasma lipoprotein composition in hamsters. *Comp Biochem Physiol A 99*, 665–670.
39. Clader JW, Berger JG, Burrier RE, Davis HR, Domalski M, Dugar S, et al. (1995). Substituted (1,2-diarylethyl) amide acyl-CoA:cholesterol acyltransferase inhibitors: Effect of polar groups on in vitro and in vivo activity. *J Med Chem 38*, 1600–1607.
40. Vaccaro W, Amore C, Berger J, Burrier R, Clader J, Davis H, et al. (1996). Inhibitors of acyl CoA:cholesterol acyltransferase. *J Med Chem 39*, 1704–1719.
41. Burnett DA, Caplen MA, Davis HR Jr, Burrier RE, Clader JW (1994). 2-Azetidinones as inhibitors of cholesterol absorption. *J Med Chem 37*, 1733–1736.
42. Burnett DA (2004). Beta-lactam cholesterol absorption inhibitors. *Curr Med Chem 11*, 1873–1887.
43. Salisbury BG, Davis HR, Burrier RE, Burnett DA, Bowkow G, Caplen MA, et al. (1995). Hypocholesterolemic activity of a novel inhibitor of cholesterol absorption, SCH 48461. *Atherosclerosis 115*, 45–63.
44. van Heek M, France CF, Compton DS, McLeod RL, Yumibe NP, Alton KB, et al. (1997). In vivo metabolism-based discovery of a potent cholesterol absorption inhibitor, SCH58235, in the rat and rhesus monkey through

the identification of the active metabolites of SCH48461. *J Pharmacol Exp Ther 283*, 157–163.

45. Rosenblum SB, Huynh T, Afonso A, Davis HR Jr, Yumibe N, Clader JW, et al. (1998). Discovery of 1-(4-fluorophenyl)-(3*R*)-[3-(4-fluorophenyl)-(3*S*)-hydroxypropyl]-(4*S*)-(4-hydroxyphenyl)-2-azetidinone (SCH 58235): A designed, potent, orally active inhibitor of cholesterol absorption. *J Med Chem 41*, 973–980.
46. Balakin KV, Kozintsev AV, Kiselyov AS, Savchuk NP (2006). Rational design approaches to chemical libraries for hit identification. *Curr Drug Discov Technol 3*, 49–65.
47. Singh S, Singh B, Bahuguna R, Wadhwa L, Saxena R (2006). Stress degradation studies on ezetimibe and development of a validated stability-indicating HPLC assay. *J Pharm Biomed Anal 41*, 1037–1040.
48. Chemical Abstract Service. SciFinder Scholar. 2006. Available at: https://scifinder.cas.org.
49. Wu CY, Benet LZ (2005). Predicting drug disposition via application of BCS: Transport/absorption/ elimination interplay and development of a biopharmaceutics drug disposition classification system. *Pharm Res 22*, 11–23.
50. Darkes MJ, Poole RM, Goa KL (2003). Ezetimibe. *Am J Cardiovasc Drugs 3*, 67–76.
51. Repa JJ, Dietschy JM, Turley SD (2002). Inhibition of cholesterol absorption by SCH 58053 in the mouse is not mediated via changes in the expression of mRNA for ABCA1, ABCG5, or ABCG8 in the enterocyte. *J Lipid Res 43*, 1864–1874.
52. van Heek M, Farley C, Compton DS, Hoos LM, Smith-Torhan A, Davis HR (2003). Ezetimibe potently inhibits cholesterol absorption but does not affect acute hepatic or intestinal cholesterol synthesis in rats. *Br J Pharmacol 138*, 1459–1464.
53. van Heek M, Farley C, Compton DS, Hoos L, Davis HR (2001). Ezetimibe selectively inhibits intestinal cholesterol absorption in rodents in the presence and absence of exocrine pancreatic function. *Br J Pharmacol 134*, 409–417.
54. van Heek M, Farley C, Compton DS, Hoos L, Alton KB, Sybertz EJ, et al. (2000). Comparison of the activity and disposition of the novel cholesterol absorption inhibitor, SCH58235, and its glucuronide, SCH60663. *Br J Pharmacol 129*, 1748–1754.
55. van Heek M, Davis H (2002). Pharmacology of ezetimibe. *Eur Heart J Suppl 4*, J5–J8.
56. Ezzet F, Wexler D, Statkevich P, Kosoglou T, Patrick J, Lipka L, et al. (2001). The plasma concentration and LDL-C relationship in patients receiving ezetimibe. *J Clin Pharmacol 41*, 943–949.
57. Sudhop T, Lutjohann D, Kodal A, Igel M, Tribble DL, Shah S, et al. (2002). Inhibition of intestinal cholesterol absorption by ezetimibe in humans. *Circulation 106*, 1943–1948.
58. Dujovne CA, Ettinger MP, Mcneer JF, Lipka LJ, LeBeaut AP, Suresh R, et al. (2002). Efficacy and safety of a potent new selective cholesterol absorption inhibitor, Ezetimibe, in patients with primary hypercholesterolemia. *Am J Cardiol 90*, 1092–1097.
59. Knopp RH, Gitter H, Truitt T, Bays H, Manion CV, Lipka LJ, et al. (2003). Effects of ezetimibe, a new cholesterol absorption inhibitor, on plasma lipids in patients with primary hypercholesterolemia. *Eur Heart J 24*, 729–741.
60. Ness GC, Holland RC, Lopez D (2006). Selective compensatory induction of hepatic HMG-CoA reductase in response to inhibition of cholesterol absorption. *Exp Biol Med (Maywood) 231*, 559–565.
61. Davidson MH, Ballantyne CM, Kerzner B, Melani L, Sager PT, Lipka L, et al. (2004). Efficacy and safety of ezetimibe coadministered with statins: Randomised, placebo-controlled, blinded experience in 2382 patients with primary hypercholesterolemia. *Int J Clin Pract 58*, 746–755.
62. Ballantyne CM, Blazing MA, King TR, Brady WE, Palmisano J (2004). Efficacy and safety of ezetimibe co-administered with simvastatin compared with atorvastatin in adults with hypercholesterolemia. *Am J Cardiol 93*, 1487–1494.
63. Ballantyne CM, Houri J, Notarbartolo A, Melani L, Lipka LJ, Suresh R, et al. (2003). Effect of ezetimibe coadministered with atorvastatin in 628 patients with primary hypercholesterolemia: A prospective, randomized, double-blind trial. *Circulation 107*, 2409–2415.
64. Feldman T, Koren M, Insull W Jr, McKenney J, Schrott H, Lewin A, et al. (2004). Treatment of high-risk patients with ezetimibe plus simvastatin co-administration versus simvastatin alone to attain National Cholesterol Education Program Adult Treatment Panel III low-density lipoprotein cholesterol goals. *Am J Cardiol 93*, 1481–1486.
65. Davis HR Jr, Compton DS, Hoos L, Tetzloff G (2001). Ezetimibe, a potent cholesterol absorption inhibitor, inhibits the development of atherosclerosis in ApoE knockout mice. *Arterioscler Thromb Vasc Biol 21*, 2032–2038.
66. Baigent C, Landry M (2003). Study of Heart and Renal Protection (SHARP). *Kidney Int Suppl 84*, S207–S210.
67. Davis HR, Zhu LJ, Hoos LM, Tetzloff G, Maguire M, Liu J, et al. (2004). Niemann-pick C1 Like 1 (NPC1L1) is the intestinal phytosterol and cholesterol transporter and a key modulator of whole body cholesterol homeostasis. *J Biol Chem 279*, 33586–33592.
68. Salen G, von Bergmann K, Lutjohann D, Kwiterovich P, Kane J, Patel SB, et al. (2004). Ezetimibe effectively reduces plasma plant sterols in patients with sitosterolemia. *Circulation 109*, 966–971.
69. Davies JP, Scott C, Oishi K, Liapis A, Ioannou YA (2005). Inactivation of NPC1L1 causes multiple lipid transport defects and protects against diet-induced hypercholesterolemia. *J Biol Chem 280*, 12710–12720.
70. Garcia-Calvo M, Lisnock J, Bull HG, Hawes BE, Burnett DA, Braun MP, et al. (2005). The target of ezetimibe is Niemann-Pick C1-Like 1 (NPC1L1). *Proc Natl Acad Sci USA 102*, 8132–8137.

71. Iyer SP, Yao X, Crona JH, Hoos LM, Tetzloff G, Davis HR Jr, et al. (2005). Characterization of the putative native and recombinant rat sterol transporter Niemann-Pick C1 Like 1 (NPC1L1) protein. *Biochim Biophys Acta 1722*, 282–292.
72. Yu L, Bharadwaj S, Brown JM, Ma Y, Du W, Davis MA, et al. (2006). Cholesterol-regulated translocation of NPC1L1 to the cell surface facilitates free cholesterol uptake. *J Biol Chem 281*, 6616–6624.
73. Merck/Schering-Plough Pharmaceuticals (2006). Zetia (ezetimibe) tablets. Prescribing information.
74. Fux R, Morike K, Gundel UF, Hartmann R, Gleiter CH (2004). Ezetimibe and statin-associated myopathy. *Ann Intern Med 140*, 671–672.
75. Simard C, Poirier P (2006). Ezetimibe-associated myopathy in monotherapy and in combination with a 3-hydroxy-3-methylglutaryl coenzyme A reductase inhibitor. *Can J Cardiol 22*, 141–144.
76. Kosoglou T, Statkevich P, Johnson-Levonas AO, Paolini JF, Bergman AJ, Alton KB (2005). Ezetimibe: A review of its metabolism, pharmacokinetics and drug interactions. *Clin Pharmacokinet 44*, 467–494.
77. Oswald S, Haenisch S, Fricke C, Sudhop T, Remmler C, Giessmann T, et al. (2006). Intestinal expression of P-glycoprotein (ABCB1), multidrug resistance associated protein 2 (ABCC2), and uridine diphosphate-glucuronosyltransferase 1A1 predicts the disposition and modulates the effects of the cholesterol absorption inhibitor ezetimibe in humans. *Clin Pharmacol Ther 79*, 206–217.
78. Patrick JE, Kosoglou T, Stauber KL, Alton KB, Maxwell SE, Zhu Y, et al. (2002). Disposition of the selective cholesterol absorption inhibitor ezetimibe in healthy male subjects. *Drug Metab Dispos 30*, 430–437.
79. Ghosal A, Hapangama N, Yuan Y, Achanfuo-Yeboah J, Iannucci R, Chowdhury S, et al. (2004). Identification of human UDP-glucuronosyltransferase enzyme(s) responsible for the glucuronidation of ezetimibe (Zetia). *Drug Metab Dispos 32*, 314–320.
80. Oswald S, Westrup S, Grube M, Kroemer HK, Weitschies W, Siegmund W (2006). Disposition and sterol-lowering effect of ezetimibe in multidrug resistance-associated protein 2-deficient rats. *J Pharmacol Exp Ther 318*, 1293–1299.
81. Ezzet F, Krishna G, Wexler DB, Statkevich P, Kosoglou T, Batra VK (2001). A population pharmacokinetic model that describes multiple peaks due to enterohepatic recirculation of ezetimibe. *Clin Ther 23*, 871–885.
82. Schiller C, Frohlich CP, Giessmann T, Siegmund W, Monnikes H, Hosten N, et al. (2005). Intestinal fluid volumes and transit of dosage forms as assessed by magnetic resonance imaging. *Aliment Pharmacol Ther 22*, 971–979.
83. Oswald S, Giessmann T, Luetjohann D, Wegner D, Rosskopf D, Weitschies W, et al. (2006). Disposition and sterol-lowering effect of ezetimibe are influenced by single-dose coadministration of rifampin, an inhibitor of multidrug transport proteins. *Clin Pharmacol Ther 80*, 477–485.
84. Zhu Y, Statkevich P, Kosoglou T, Zambas D, Patrick J, Cayen MN, et al. (2000). Effect of SCH 58235 on the activity of drug metabolizing enzymes in vivo. *Clin Pharmacol & Ther 67*, 152.
85. Kosoglou T, Meyer I, Veltri EP, Statkevich P, Yang B, Zhu Y, et al. (2002). Pharmacodynamic interaction between the new selective cholesterol absorption inhibitor ezetimibe and simvastatin. *Br J Clin Pharmacol 54*, 309–319.
86. Kosoglou T, Statkevich P, Yang B, Suresh R, Zhu Y, Boutros T, et al. (2004). Pharmacodynamic interaction between ezetimibe and rosuvastatin. *Curr Med Res Opin 20*, 1185–1195.
87. Reyderman L, Kosoglou T, Boutros T, Seiberling M, Statkevich P (2004). Pharmacokinetic interaction between ezetimibe and lovastatin in healthy volunteers. *Curr Med Res Opin 20*, 1493–1500.
88. Altmann SW, Davis HR Jr, Yao X, Laverty M, Compton DS, Zhu LJ, et al. (2002). The identification of intestinal scavenger receptor class B, type I (SR-BI) by expression cloning and its role in cholesterol absorption. *Biochim Biophys Acta 1580*, 77–93.
89. Hawes BE, O'neill KA, Yao X, Crona JH, Davis HR Jr, Graziano MP, et al. (2007). In vivo responsiveness to ezetimibe correlates with Niemann-pick C1 Like-1 (NPC1L1) binding affinity: Comparison of multiple species NPC1L1 orthologs. *Mol Pharmacol 71*, 19–29.
90. Huff MW, Pollex RL, Hegele RA (2006). NPC1L1: Evolution from pharmacological target to physiological sterol transporter. *Arterioscler Thromb Vasc Biol 26*, 2433–2438.
91. Yamanashi Y, Takada T, Suzuki H (2007). Niemann-Pick C1-Like 1 overexpression facilitates ezetimibe-sensitive cholesterol and beta-sitosterol uptake in CaCo-2 cells. *J Pharmacol Exp Ther 320*, 559–564.
92. Davis HR Jr, Hoos LM, Tetzloff G, Maguire M, Zhu LJ, Graziano MP, et al. (2007). Deficiency of Niemann-Pick C1 Like 1 prevents atherosclerosis in apoE–/– mice. *Arterioscler Thromb Vasc Biol 27*, 841–849.
93. Duval C, Touche V, Tailleux A, Fruchart JC, Fievet C, Clavey V, et al. (2006). Niemann-Pick C1 Like 1 gene expression is down-regulated by LXR activators in the intestine. *Biochem Biophys Res Commun 340*, 1259–1263.
94. van der Veen JN, Kruit JK, Havinga R, Baller JF, Chimini G, Lestavel S, et al. (2005). Reduced cholesterol absorption upon PPARdelta activation coincides with decreased intestinal expression of NPC1L1. *J Lipid Res 46*, 526–534.
95. During A, Dawson HD, Harrison EH (2005). Carotenoid transport is decreased and expression of the lipid transporters SR-BI, NPC1L1, and ABCA1 is downregulated in Caco-2 cells treated with ezetimibe. *J Nutr 135*, 2305–2312.
96. Smart EJ, De Rose RA, Farber SA (2004). Annexin 2-caveolin 1 complex is a target of ezetimibe and regu-

lates intestinal cholesterol transport. *Proc Natl Acad Sci USA 101*, 3450–3455.

97. Valasek MA, Weng J, Shaul PW, Anderson RG, Repa JJ (2005). Caveolin-1 is not required for murine intestinal cholesterol transport. *J Biol Chem 280*, 28103–28109.
98. Kramer W, Girbig F, Corsiero D, Pfenninger A, Frick W, Jahne G, et al. (2005). Aminopeptidase N (CD13) is a molecular target of the cholesterol absorption inhibitor ezetimibe in the enterocyte brush border membrane. *J Biol Chem 280*, 1306–1320.
99. Seedorf U, Engel T, Lueken A, Bode G, Lorkowski S, Assmann G (2004). Cholesterol absorption inhibitor Ezetimibe blocks uptake of oxidized LDL in human macrophages. *Biochem Biophys Res Commun 320*, 1337–1341.
100. Leon C, Hill JS, Wasan KM (2005). Potential role of acyl-coenzyme A:cholesterol transferase (ACAT) Inhibitors as hypolipidemic and antiatherosclerosis drugs. *Pharm Res 22*, 1578–1588.
101. Buhman KK, Accad M, Novak S, Choi RS, Wong JS, Hamilton RL, et al. (2000). Resistance to diet-induced hypercholesterolemia and gallstone formation in ACAT2-deficient mice. *Nat Med 6*, 1341–1347.
102. Burnett JR, Wilcox LJ, Huff MW (1999). Acyl coenzyme A: Cholesterol acyltransferase inhibition and hepatic apolipoprotein B secretion. *Clin Chim Acta 286*, 231–242.
103. Insull W Jr, Koren M, Davignon J, Sprecher D, Schrott H, Keilson LM, et al. (2001). Efficacy and short-term safety of a new ACAT inhibitor, avasimibe, on lipids, lipoproteins, and apolipoproteins, in patients with combined hyperlipidemia. *Atherosclerosis 157*, 137–144.
104. Tardif JC, Gregoire J, L'Allier PL, Anderson TJ, Bertrand O, Reeves F, et al. (2004). Effects of the acyl coenzyme A:cholesterol acyltransferase inhibitor avasimibe on human atherosclerotic lesions. *Circulation 110*, 3372–3377.
105. Nissen SE, Tuzcu EM, Brewer HB, Sipahi I, Nicholls SJ, Ganz P, et al. (2006). Effect of ACAT inhibition on the progression of coronary atherosclerosis. *N Engl J Med 354*, 1253–1263.
106. Burnett JR, Watts GF (2007). MTP inhibition as a treatment for dyslipidaemias: Time to deliver or empty promises? *Expert Opin Ther Targets 11*, 181–189.
107. Chandler CE, Wilder DE, Pettini JL, Savoy YE, Petras SF, Chang G, et al. (2003). CP-346086: An MTP inhibitor that lowers plasma cholesterol and triglycerides in experimental animals and in humans. *J Lipid Res 44*, 1887–1901.
108. Kramer W, Glombik H (2006). Bile acid reabsorption inhibitors (BARI): Novel hypolipidemic drugs. *Curr Med Chem 13*, 997–1016.
109. Clader JW, Burnett DA, Caplen MA, Domalski MS, Dugar S, Vaccaro W, et al. (1996). 2-Azetidinone cholesterol absorption inhibitors: Structure-activity relationships on the heterocyclic nucleus. *J Med Chem 39*, 3684–3693.
110. Kvaerno L, Werder M, Hauser H, Carreira EM (2005). Synthesis and in vitro evaluation of inhibitors of intestinal cholesterol absorption. *J Med Chem 48*, 6035–6053.

# CARDIOLOGY—IN-DEPTH LOOK: KINASES AS THERAPEUTIC TARGETS IN THE HEART

Karin Mauer and J. Emilio Exaire
*Department of Medicine, Cardiovascular Section, University of Oklahoma Health Sciences Center, Oklahoma City, Oklahoma*

## 1 KINASES AS PHARMACOLOGICAL TARGETS IN THE HEART—A NEW ERA IN MOLECULAR CARDIOVASCULAR THERAPEUTICS

Cardiovascular disease is the leading cause of morbidity and mortality in modern society, resulting in enormous health care costs. The spectrum of diseases encompasses a group of clinically heterogeneous ailments of the heart (e.g., ischemic heart disease or coronary artery disease, congestive heart failure, and cardiomyopathy) and of blood vessels (e.g., hypertension and atherosclerosis) with multifactorial origin, strongly influenced by race, gender, and environmental aspects. In recent years there has been extensive progress toward understanding the pathophysiology of cardiovascular disease, which has shown multiple cell signaling cascades. Anomalous activation of diverse neurohormonal systems (e.g., catecholamines, angiotensin II, endothelin-1, natriuretic peptides, cytokines, and growth factors), which are crucial for homeostasis under physiological conditions, play a role in the establishment and progression of cardiovascular disease. Furthermore, there is evidence that specific signaling pathways are involved in acute and chronic cardiovascular disease, and the molecules comprising these pathways are targets for future development of therapies. In fact, the efficacy of drugs such as angiotensin II and adrenergic receptor blockers provides confirmation that targeting these signaling pathways is valuable for the treatment of cardiovascular disease. Most of the signaling pathways consist of a cascade of protein kinases that transduce signals from a receptor on the cell membrane to the nucleus of the cell. This series of protein phosphorylation leads to specific cellular responses (e.g., gene expression, growth, and/or proliferation), and hypothetically it could be possible to modulate these events in order to achieve a desirable cellular effect. Since most of these signaling pathways and protein phosphorylation are mediated by these enzymes, changes in protein kinase activity by pharmacological agents could provide a therapeutic benefit in cardiovascular disease [1]. Even though there is an extensive selection of effective treatments that target cardiovascular disease (e.g., statins, β-blockers, antagonists of the renin–angiotensin system, calcium channel blockers, thrombolytic agents, and antiplatelet drugs), mortality associated to it remains high. Therefore, there is a need to develop new therapeutic agents that will positively impact quality of life and prolong survival.

Several protein kinases have been implicated in human cardiovascular disease by the successful use of G-protein-coupled receptor (GPCR) antagonists (e.g., protein kinase A and CaMKII mediate β-adrenergic receptor action) and in some cases by evidence of altered expression in diseased human heart tissue. In fact, the successful use of GPCR blockers have shown to reduce mortality from cardiovascular disease and suggests that a larger number of downstream kinases mediating the actions of these receptors are also

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

**TABLE 1 Protein Kinase Involvement in Heart Disease**[a]

| | | Experimental Models | |
|---|---|---|---|
| Kinase | Human Disease | Overexpression or Increased Activation | Knockout, Inhibition or Decreased Activation |
| AKT | Decreased activity (after LVAD) | Preserved function after I/R, but with increased MI size | Reduced recovery after ischemia (DN) |
| CaMKII | Increased expression (HF) | Hypertrophy, dysfunction, sudden death | Reduced remodeling and improved function after MI and isoproterenol |
| ERK | No change (HF secondary to ischemic disease) | Hypertrophy and increased contraction | NA |
| JAK | Decreased JAK2 activity (DCM)<br>Increased JAK1 activity (DCM) | Promotes cell survival<br>Prevents remodeling after MI | Reduced ischemic preconditioning |
| JNK | Increased activity (HF secondary to ischemic disease)<br>Decreased activity (after LVAD) | Atrial enlargement | Increased hypertrophy with pressure overload and aging (DN) |
| p 38 MAPK | Increased activity (HF secondary to ischemic disease)<br>Increased activity (after LVAD) | Fibrosis, dysfunction, and sudden death without hypertrophy | 1. Increased hypertrophy with pressure overload, Ang II, and isoproterenol<br>2. Hypertrophy without fibrosis with pressure overload<br>3. Reduced I/R-induced apoptosis and preserved cardiac function |
| PKA | Increased activity (HF) | Hypertrophy, dysfunction, sudden death | NA |
| PKC | Increased activity and expression (HF) | Improved function after ischemia (PKCβ) | Protection from ischemia (PKCε translocation) |
| PI3Kγ | | | 1. Impaired ADP-induced, but not collagen-induced platelet aggregation (protection against ADP-induced thromboembolic vascular occlusion)<br>2. Myocardial damage after chronic pressure overload<br>3. Reduced phosphorylation of NOS3, blunted NO production, and protection from PAF-induced negative inotropism<br>4. Improved contractile recovery after I/R<br>5. Attenuated hypertrophy and interstitial fibrosis in response to isoproterenol<br>6. Protection from isoproterenol-induced HF<br>7. Protection against Ang-II-mediated hypertensive response |
| AMPK | | Hypertrophy associated with glycogen accumulation and ventricular pre-excitation | 1. Insulin resistance, no overt cardiac phenotype (total α1 or 2-subunit KO)<br>2. Decreased glucose uptake, increased ATP depletion, and increased left ventricular dysfunction during ischemia (mutated α2 subunit)<br>3. Normal cardiac function during exercise, without ATP depletion (mutated α2 subunit)<br>4. No cardiac hypertrophy; during I/R decreased glucose uptake, glycolysis, and failure to increase fatty acid oxidation; impaired cardiac function and increased cardiac injury following reperfusion (mutated α2 subunit)<br>5. Cardiac hypertrophy associated with glycogen accumulation and ventricular pre-excitation |

**TABLE 1** ***Continued***

| Kinase | Human Disease | Experimental Models: Overexpression or Increased Activation | Experimental Models: Knockout, Inhibition or Decreased Activation |
|---|---|---|---|
| ROCK | | | 1. Normal basal and Ang-II-stimulated blood pressure, heart rate, and cardiac dimensions, but reduced perivascular fibrosis<br>2. Decreased perivascular and interstitial fibrosis after pressure overload<br>3. Reduced I/R injury, infarct size, myocardial fibrosis, and cardiac hypertrophy<br>4. Lower cardiac myocyte apoptosis and improved postischemic cardiac function |

[a]AKT, kinase originally encoded by akt murine thymoma viral oncogene; LVAD, left-ventricular assist device; I/R, ischemia/reperfusion; MI, myocardial infarction; DN, dominant negative mutant; CaMKII, multifunctional calcium/calmodulin-dependent protein kinase II; HF, heart failure; ERK, extracellular-signaling-related kinases; NA, not available; JAK, Janus kinases; DCM, dilated cardiomyopathy; JNK, jun N-terminal kinase; MAPK, mitogen-activated protein kinase; ang II, Angiotensin II; PKA, protein kinase A; PKC, protein kinase C; NOS, nitric-oxide synthase; NO, nitric oxide; PAF, platelet-activating factor; AMPK, AMP-activated protein kinase; KO, knockout.

*Sources*: Adapted from refs. 1, 2, 4, 5, 13, 14, and 21.

**TABLE 2 Drug Development for Protein Kinase Modulators in Cardiovascular Disease[a]**

| Indication | | Kinase Targeted | Compound | Stage of Testing |
|---|---|---|---|---|
| ACS | | p38 MAPK | VX-702 (Vertex Pharmaceuticals Inc.) | Phase II |
| | | p38 MAPK | SB239063 | |
| | Acute MI | PI3Kγ | TG100-115 (TargeGen Inc.) | Phase I/II |
| | MI, adjunct to angioplasty | PKCδ | KAI-9803 (KAI Pharmaceuticals Inc.) | Phase II |
| | Angina | ROCK | Fasudil[b] | Phase II |
| HF | | MAPK and c-raf | Semapimod | Preclinical[c] |
| | | Grk2 | NA | Preclinical |
| | | HDAC kinase | NA | Preclinical |
| | | JAK2 | NA | Preclinical |
| | | PKC β | NA | Preclinical |
| Arrhythmia | | CaMKII | NA | Preclinical |
| I/R injury | | JNK | JNK9395 | Preclinical |
| Stroke, diabetes | | p38 MAPK | SCIO-323 (Scios Inc.) | Phase I |
| Acute cardioprotection | | Cdk9 | NA | Preclinical |
| Cardiomyocyte survival | | HGK | NA | Preclinical |
| Cardiovascular disease | | Chk2 | NA | Preclinical |
| | | PKC | NA | Preclinical |

[a]ACS, acute coronary syndromes; MAPK, mitogen-activated protein kinase; MI, myocardioal infarct; PI3K, phosphatidylinositide 3 kinase; PKC, protein kinase C; ROCK, Rho-associated kinase; HF, heart failure; c-raf, a serine/threonine kinase activated by RAS binding; Grk2, G-protein-coupled-receptor kinase 2; NA not available; HDAC kinase, histone deacetylase kinase; JAK2, janus kinase 2; CaMKII, calcium/calmodulin-dependent protein kinase II; I/R, ischemia/reperfusion; JNK, jun N-terminal kinase; Cdk9, cyclin-dependent kinase 9; HGK, hepatocyte progenitor kinase-like/germinal center kinaselike kinase; Chk2, checkpoint kinase.

[b]Approved for treatment of cerebral vasospasm in Japan.

[c]Phase II for inflammatory disease.

*Sources*: Adapted from refs. 2, 4, and 5.

potential therapeutic targets [2]. Most, if not all, intracellular signaling pathways depend on the phosphorylation status, and thus activity of involved proteins and all cellular functions (e.g., gene expression, cell adhesion, cell cycle progression, and differentiation) are modulated by the complex and specific interplay of kinases and phosphatases [3]. There are over 500 different protein kinases, and overall they facilitate the phosphorylation of diverse protein substrates within a given cell, which serves as a common form of signal

transduction. The kinase transfers a phosphate group from adenosine triphosphate (ATP) onto hydroxyl residues of serine or threonine (serine/threonine protein kinases), or of tyrosine (tyrosine protein kinases) on specific proteins, thereby modulating their function, which has an impact on different cellular processes. Protein kinase activity is regulated by phosphorylation itself or second messenger levels [e.g., cyclic adenosine monophosphate (cAMP) and calcium] controlled by GPCRs in response to extracellular signals [1, 2]. Protein kinases interrelate through a process known as cross talk, which reflects an interacting network rather than a sequential activation linked to one GPCR. Since GPCRs are positioned at the beginning of signaling pathways, activation of one receptor can modify multiple second messengers and mediators downstream of the cascade. Thus, antagonism at receptor level might not select between the beneficial and harmful effects of aberrant signaling in disease, and blocking the signaling cascade further downstream has more potential for greater selectivity [2]. Selective targeting of signaling pathways using protein kinase modulators could have a potential advantage over receptor blockers [4], and choosing the right biological target is one of the most critical decisions made in pharmaceutical research and development [5]. Recently, a lot of effort has focused on compounds that modulate different protein kinases in multiple settings. In fact, the cancer field has led the way in drug development directed both at normal and mutated protein kinases [4], and there is growing evidence showing that regulation of these enzymes has therapeutic interventional value in cardiovascular disease.

### 1.1 Key Events in the Development of Protein Kinase Inhibitors [6]

In the early 1980s the first protein kinase inhibitors were developed. Certain antagonists of the calcium-binding protein calmodulin (naphthalene sulfonamides) at higher concentrations were found to inhibit several protein kinases. In addition, when the naphthalene ring was replaced by isoquinoline, it was observed that the compound was no longer a calmodulin antagonist but retained the ability to inhibit protein kinases. Since these compounds were ATP competitive and cell permeable, a potential use in vivo was foreseen. Even though many of the compounds developed subsequently are of relatively low potency and selectivity, one of these compounds, fasudil hydrochloride (also known as HA1077 or AT877) progressed to human clinical trials in the early 1990s. This drug was approved in Japan in 1995 for the treatment of cerebral vasospasm after surgery for subarachnoid hemorrhage and associated cerebral ischemic symptoms.

The discovery that the antifungal agent staurosporine was an inhibitor of protein kinase C (PKC) at nanomolar concentrations led to the development of compounds such as Roche's Ro31-8220 and Goedke's Go6850 (also called GF109203X) used in cell-based assays. However, these and several other drugs were later shown to lack specificity. Nevertheless, several have progressed to human clinical trials, but it is unclear whether their efficacy results from the inhibition of PKC and/or other protein kinases. It has been shown that some of these also block cell-cycle progression by inhibiting the cell-cycle checkpoint-control kinase CHK1. In fact, the compound PKC412 by Novartis has been tested in human clinical trials for the treatment of advanced cancers. Furthermore, it also seems to inhibit retinal and choroidal neovascularization in murine models and is being evaluated in clinical trials for the treatment of ischemic retinopathy.

The compound LY333531 inhibits the PKC-β isoform more potently than other PKC isoforms and has been said to normalize PKC activity in the retina and kidneys of diabetic rats, and it has been tested in phase III clinical trials for the treatment of diabetic microvascular disease.

Interestingly, some of the initial drugs to be approved that inhibit a specific protein phosphatase or protein kinases were cyclosporin and rapamycin. However, the development of these compounds was achieved without knowing their exact mechanism of action and focusing only on their clinical efficacy. Most of the effort to develop agents that target specific protein kinases has been focused in the field of oncology. Actually, some of the most promising drugs being developed as anticancer drugs are tyrosine kinase inhibitors. Key studies carried out 10–15 years ago, which identified main compound classes of tyrosine kinase inhibitors, such as quinazolines and tyrphostins, have greatly influenced progress in this area. However, it was not until May 2001 that imatinib was approved for clinical use as a protein kinase inhibitor [6]. By the year 2002, at least 3 protein kinase inhibitors (fasudil, rapamycin, and Gleevec) and 1 phosphatase inhibitor (cyclosporin) had been approved for clinical use in specific pathologies, over 20 other protein kinase inhibitors were being tested in human clinical trials and many more were at the preclinical stage [6]. By the year 2007 at least 8 drugs that target kinases have received regulatory approval for specific oncologic conditions, and over 100 additional compounds that block protein kinase activity are being evaluated [1]. Consequently, protein kinases are the second largest group of pharmaceutical targets after GPCRs, comprising between 20 and 30% of the drug discovery programs. Since protein kinases encompass one of the largest enzyme families encoded by the

human genome, there are many potential targets for drug development. Much of the effort so far has been directed toward oncologic conditions in which protein kinase inhibitors are being well tolerated compared to conventional chemotherapeutic treatments. The safety profile seen so far in human clinical trials is encouraging since several of these protein kinase inhibitors are not very specific. In order to determine whether a drug is exerting its effect through blocking of a particular protein kinase, it needs to be evaluated when a drug-resistant mutant is overexpressed or replaces the wild-type enzyme, which has been shown in only a few cases (e.g., mutations in Abelson tyrosine kinase, or ABL, can make it resistant to Gleevec, which causes a relapse in patients who have an advanced stage of chronic myelogenous leukemia) [6] .

There is a need to use different ways to screen for and/or design inhibitors of particular protein kinases based on a detailed understanding of their catalytic and regulatory properties to fully understand the essential role in the pathogenesis of disease as well as for research purposes. In fact, specific protein kinase inhibitors that cannot be used as drugs because of toxicity, pharmacology, or solubility (e.g., PD98059 and SB203580) have helped broaden our understanding in specific therapeutic areas [6]. The majority of the kinase inhibitors currently in development are focused on diverse oncologic conditions, and kinase involvement in cancer seems to be different than in cardiovascular disease. In cancer cells, the aberrant behavior of kinases is associated with genetic or somatic mutations, while in cardiovascular disease this is because of enhanced stimulation by activated neurohormonal systems. The successful use of kinase inhibitors in cancer supports its application in the treatment of cardiovascular disease [1]. Currently, there are two kinase antagonists under development targeting the receptors either by antibodies or small molecule. Antibodies target receptor tyrosine kinases and behave similar to receptor blockers while small-molecule kinase inhibitors target signal transduction [1].

### 1.2 Mitogen-Activated Protein Kinases (MAPKs)

Small G-proteins are low-molecular-weight monomeric G-proteins (e.g., Ras, Rho, Ran, Rab, Arf, and Kir/Rem/Rad) that act as a molecular "switch" controlled by a guanosine diphosphate/triphosphate (GDP/GTP)-binding cycle, and regulate several intracellular signaling events, such as cell proliferation, apoptosis, migration, gene expression, cytoskeletal reorganization, intracellular vesicle trafficking, and microtubule organization during the cell cycle. These important cellular responses are mediated partly by their downstream effectors known as MAPKs [i.e., extracellular signal regulated kinase (ERK), c-Jun $NH_2$-terminal kinase (JNK), and p38 MAPK] [7]. There are several members in the MAPK family that can be classified into four groups: (1) the classical MAPKs (ERK1 and ERK2), (2) JNKs (JNK1, JNK2, and JNK3), (3) p38 MAPKs (p38α, p38β, p38γ, and p38δ), and (4) atypical MAPKs (ERK3, ERK5, and ERK8). Traditionally, JNKs and p38 MAPKs are considered as stress-activated MAPKs and have been implicated in cardiovascular disease. In animal models of cardiovascular disease, such as hypertension and atherosclerosis, MAPKs appear to be activated in vascular tissues by mechanical and oxidative stress, vasoconstrictors, and growth factors [7]. JNK1 and JNK2 are ubiquitously expressed and JNK3 appears to be neural specific. Some of JNK's substrates include transcription factors (e.g., AP-1, p53 and c-Myc), as well as nontranscription factors that are related to apoptotic cell death factors (e.g., Bcl-2 family members). In cultured vascular smooth muscle cells (VSMCs), JNKs are activated by environmental stress (e.g., oxidative stress and pro-inflammatory cytokines) and growth factors [7]. ERK1/2, JNK, and p38MAPK signaling cascades play an important role in multiple processes, including cell differentiation, proliferation, and death, as well as homeostasis and acute response to hormones. Despite being ubiquitous, their activation mechanism, substrate preference, and functions are often cell specific. ERK1/2 activation depends on growth stimuli (e.g., ET-1, phenylephrine, and Ang II in cardiac myocytes) and mediates growth and cytoprotective functions. In contrast, JNK and p38MAPK are activated by cellular stress (e.g., ischemia, pro-inflammatory cytokines, and hypertrophic stimuli). There is controversy whether activation plays a causal or consequential role in cardiac disease since conflicting results may be associated with the existence of multiple isoforms and nonselectivity of current inhibitors [1]. Selective oral inhibition of p38α isoform with SD-282 in animal models of myocardial injury improved cardiac function, stopped cardiac myocyte apoptosis, and decreased macrophage infiltration. In cultured rat cardiomyocytes exposed to hypoxia/reoxygenation, inhibition of p38α with SB203580 notably avoided induced cell death. Furthermore, estrogen has proven to decrease p38α activation and concomitantly favor p38β activity, which is barely affected by hypoxia/reoxygenation. This implies that estrogen differentially modulates the two major p38 isoforms and that cell death is mainly regulated by p38α. There is evidence that there are differences in myocardial responsiveness to ischemia/reperfusion (I/R) between the murine and the porcine models, since p38 activation in the mouse participates in acute cellular injury and death, while in the pig it seems not to have a contributing effect on these parameters. Thus, species specificity is important

to have in mind when using p38 as an aim for intervention [1].

MAPK p38 is a point of convergence for numerous signaling processes that have impact on different events that are essential in inflammation [8]. Over 100 p38 inhibitors have been reported, and most of them compete directly with ATP for binding to a kinase domain, which is structurally conserved in all catalytically active kinases to facilitate the transfer of the γ phosphate from ATP to the substrate. A few allosteric inhibitors do not fill the adenosine pocket, but their binding induces a conformational change that prevents the kinase from binding ATP. These blockers have varying degrees of selectivity, and the impact this might have on the safety profile of each compound after chronic treatment in humans needs to be better established [8].

### 1.3 G-Protein-Coupled Receptor Kinases (GRKs)

There are seven members in the GRK family that contribute to the regulation and signal transduction via GPCRs. GRK levels change in pathological settings such as heart failure, hypertrophy, and hypertension. GRK2 (or β-ARK1) modulates β-adrenergic receptors (β-ARs) in the heart, which are overexpressed in human heart failure. In fact, the attenuated contractile response to β-adrenergic stimulation in cardiac dysfunction has been linked to increased β-ARK1 expression and activity. Indeed, there is data to suggest that cardiac β-ARK1 expression and activity increases with disease severity. In addition, in patients with a left ventricular assist device restitution of β-AR signaling in the heart has been associated with decreased myocardial β-ARK1 activity. Pressure and volume overload appear to upregulate β1-ARs associated with subcellular redistribution of GRKs in the failing heart. In animal studies, β blockers have been described to reverse GRK-mediated desensitization and downregulation of cardiac β-ARs. Furthermore, β blockers seem to prevent increased β-ARK1 activity in the failing human heart [1]. Interestingly, exercise training in a model of spontaneous hypertension improves the inotropic and lusitropic response to β-AR stimulation, which seems to correlate with lower β-ARK1 and enhanced phosphorylation of calcium cycling proteins [1].

### 1.4 Phosphatidylinositol 3-Kinase (PI3K)

PI3Ks have central roles as lipid and protein kinases in several intracellular signaling events. In response to diverse biological stimuli, PI3Ks primarily phosphorylate phospholipids containing inositol, which have an important role as second messengers that work as a docking platform for lipid-binding domains of proteins triggering downstream kinase cascades [5]. Class IA members (i.e., PI3Kα, β, and δ) act in response to hormones, growth factors, cytokines, integrins, and other extracellular stimuli to regulate diverse processes, including cell metabolism, proliferation, growth, survival, and motility. The only member of class IB (i.e., PI3Kγ) is activated by GPCRs and Ras and has been shown to regulate MAPK cascades. There is the possibility that antagonism of such a central signaling molecule is likely to be accompanied by unwanted side effects [5]. While PI3Kα and PI3Kβ are expressed ubiquitously, PI3Kγ and PI3Kδ expression is mostly limited to the hematopoietic system, even though PI3Kγ can also be detected in the heart, endothelium, and brain. Thus, a PI3Kγ inhibitor will affect the signaling network of any given GPCR only when they are coexpressed and activated in the same cell type and tissue. Apparently, PI3Kγ shows preference to certain GPCRs and Gβγ-PI3Kγ interaction is only one event among many that is triggered by the dissociation of the trimeric G proteins subsequent to GPCR activation. Thus, even when PI3Kγ inhibition is complete, the signal coming from an activated GPCR is expected to be just reduced, rather than abolished. In that way, PI3Kγ antagonism constitutes a more subtle mode of intervention than the complete block of all downstream signals that is achieved by inhibitors that act directly on the GPCR [5]. The catalytic subunits of class IA isoforms (p110α, p110β, and p110δ) are activated by receptor tyrosine kinases, while the one in class IB members (p110γ) is activated by the βγ subunit of heterotrimeric G proteins. It has been described that PI3K has several biological roles associated with cancer, inflammation and autoimmunity, as well as cardiovascular functions. For example, the PI3K-Akt-GSK3β pathway increases cardiac protein synthesis and thus promotes hypertrophy [1].

### 1.5 Rho-dependent Serine–threonine Kinases

Rho is part of a family of small GTP-binding proteins that contribute to intracellular signal transduction, which regulates gene expression and cytoskeletal reorganization and which is triggered by activation of growth factor receptors and GPCRs. In resting cells, Rho is found primarily in the cytosol in its inactive GDP-bound form, and when certain agonists stimulate the cell, the Rho-GTP compound is translocated to the cell membrane, where it interacts with its specific targets. It has been shown that in the cardiovascular system, Rho modulates contraction of vascular smooth muscle by regulating calcium sensitivity. Rho kinase (ROCK), with its two highly homologous isoforms ROCK1 and ROCK2, is one of the downstream effectors of Rho [4, 9]. Even

though both isoforms are ubiquitously expressed, ROCK1 is mostly expressed in the kidney, lung, liver, spleen, and testis, whereas ROCK2 is preferentially present in the heart and the brain. ROCK substrates include different proteins such as myosin light-chain phosphatase, adducin, intermediate filaments (such as vimentin and desmin), the Na/H exchanger, the ezrin, radixin, and moesin family and LIM domain-containing kinase (LIM-kinase) [9]. ROCKs seem to be involved in the effects of several vasoactive substances, including angiotensin II, thrombin, 5-hydroxytryptamine (5-HT), platelet-derived growth factor (PDGF) among others, all of which have been implicated in cardiovascular disease [9] . Furthermore, it has been suggested that some effects of statins are mediated by their inhibitory effects on Rho and thus the resultant block of ROCKs. Although, it has been shown that statins promote and activate endothelial nitric oxide synthase (eNOS) expression through inhibition of Rho, the extent to which therapeutic levels of statins inhibit Rho activity and, thereby, ROCK activity remains to be established [9].

Rho GTPases are associated with cell adhesion, motility, and contraction and seem to regulate cell proliferation and hypertrophy of the cardiomyocyte, gene expression, ion channel activity, endothelial permeability, and production of reactive oxygen species. The effects of certain members of the Rho family (i.e., RhoA) are mediated by Rho-dependent serine–threonine kinases (ROCK1 and ROCK2). This ROCK activity results in increased myosin phosphorylation and contraction, which could be associated with hypertension or coronary and vascular spasm [1]. Moreover, ROCK interaction with other signaling pathways results in PI3K inhibition, eNOS modulation, and activation of plasminogen activator inhibitor-1 (PAI1), all of which may contribute to endothelial dysfunction and atherogenesis [1]. Interestingly, some of the cardioprotective effects of statins that are independent of cholesterol could be related to ROCK inhibition [1]. Recent evidence also suggests that the phosphodiesterase inhibitor sildenafil inhibits functions mediated by ROCK in the pulmonary artery, having a beneficial effect on pulmonary hypertension. There is data to imply a beneficial effect of ROCK inhibition with Y-27632 and fasudil in I/R injury and myocardial infarction. In fact, ROCK antagonism in different experimental models of myocardial infarction, chronic hypertension, and congestive heart failure appears to reduce ischemia/reperfusion injury, infarct size, myocardial fibrosis, and cardiac hypertrophy. Additionally, it appears that ROCK inhibition leads to activation of the PI3-kinase/Akt/eNOS pathway, which results in cardiovascular protection. ROCK inhibition also prevents cardiac myocyte apoptosis and improves postischemic cardiac function [1].

In recent studies, ROCK1(+/–) haplo-insufficient mice had comparable basal and Ang II-stimulated blood pressure, heart rate, and cardiac dimensions in contrast to control mice, but showed considerably less perivascular fibrosis. Furthermore, after aortic banding ROCK1(–/–) homozygous mice revealed less perivascular and interstitial fibrosis associated with decreased expression of extracellular matrix proteins and fibrogenic cytokines [e.g., transforming growth factor-β2 (TGF-β2) and connective tissue growth factor]. These experimental models suggest that ROCK1 in response to pathological stimuli may contribute to the development of cardiac fibrosis but not hypertrophy [1].

### 1.6 Protein Kinase C (PKC) and Protein Kinase A (PKA)

The PKC family encompasses closely related serine–threonine kinases implicated in cell survival, proliferation, and apoptosis and plays a role in vasoconstriction, ischemic preconditioning, and cardiac hypertrophy. There is a tissue-specific expression pattern for the PKC isoforms, which have different subcellular distribution, targets, and cell function, as well as variable calcium and lipid sensitivity. However, the importance of each isoform in the cardiovascular system remains to be elucidated. It is known that PKC signaling pathways are set in motion by GPCRs with established cardiovascular effects, such as receptors for angiotensin II, Endothelin 1 (ET-1), phenylephrine, and neuron-hormonal ligands. Activation of some PKC isoforms could be beneficial, since PKC-ε may have a cytoprotective function in cardiac myocytes and PKC-ζ limits fibroblast hypoxia-induced proliferation by stopping ERK1/2 activation [1].

PKA is a ubiquitous serine–threonine protein kinase activated by cAMP that is of major importance for cellular function [10].

### 1.7 Calcium/Calmodulin-Dependent Protein Kinase II (CaMKII)

CaMKII is one of several specialized proteins that modulates or responds to calcium signaling in myocytes, and there is evidence that associates CaMKII to both physiological and pathological conditions in the heart. It has been shown that CaMKII phosphorylates multiple proteins that share consensus motifs that contain serine or threonine. Thus, CaMKII articulates changes in calcium cycling at multiple levels, including (1) regulation of L-type calcium channels (LTCC) and thus calcium entry into the cell, (2) modulation of ryanodine receptors (RyR) and thus calcium release from the sarcoplasmic

reticulum (SR) stores, and (3) changes of the sarcoplasmic/endoplasmic reticulum calcium-ATPase (SERCA) and thus SR calcium uptake. There are four genes encoding CaMKII subunits (α, β, γ, and δ) with highly identical catalytic and regulatory domains, but so far only γ and δ have been detected in the heart, being CaMKIIδ the predominant isoform in the mammalian heart [11].

Certain parts of the regulatory domain bind to the catalytic domain in the absence of calcium, occluding thereby the binding of nucleotides and protein substrates. When calcium/calmodulin binds to the regulatory domain, apparently there is a change in protein conformation and this activation occurs to a different extent depending on the frequency, amplitude, and duration of the calcium transients. The kinase capacity to sustain its activity through autophosphorylation, even in the absence of increased calcium/calmodulin, renders notable flexibility for extending CaMKII-dependent regulation over time. Thus, CaMKII integrates and regulates calcium cycling in the cell and translates these changes to transcriptional tasks that are crucial for the contractile apparatus in cardiac myocytes [11].

### 1.8 Mammalian Target of Rapamycin (mTOR)

The serine–threonine kinase mTOR is activated through stimulation by ATP, hypoxia, growth factors, and other elements that modulate cell growth and division. A number of signaling pathways that are upstream of mTOR, such as ERK1/2 and PI3K, are involved in cardiac hypertrophy. In fact, mTOR inhibition by rapamycin appears to attenuate and overturn cardiac hypertrophy in experimental pressure overload models [1].

### 1.9 Glycogen Synthase Kinase-3β (GSK3β)

GSK3 phosphorylates serine–threonine residues, which can change the biological activity of the substrate or can prime some substrates (e.g., β-catenin and c-MYC) for proteasome destruction. GSK3 has several physiological substrates, such as transcription factors and regulators of metabolism. As with other kinases, GSK3 activity is controlled by it own phosphorylation at a tyrosine residue, and in mammalian cells, this seems to be mediated by an autophosphorylating event that is important for enzyme stabilization. Even when it has been suggested that the tyrosine phosphorylation of GSK3 changes in response to extracellular signals, it is now said that GSK3 is inhibited by a variety of agonists because of phosphorylation of a serine residue [12]. GSK3β positively regulates apoptosis and negatively regulates hypertrophy in the heart. In endothelial cells GSK3β signaling is associated with angiogenesis. In contrast to other protein kinases, phosphorylated GSK3β is inactive and vice versa. Active GSK3β stimulates factors involved in protein synthesis (e.g., elF2B) and expression of genes related to hypertrophy (e.g., c-MYC, GATA4, β-catenin). There is evidence that GSK3β inhibition reduces myocardial cell death after I/R [1].

GSK3β inhibitors (such as lithium and SB216763) can mimic angiogenesis and cardioprotection mediated by ischemic preconditioning in rats. In addition, GSK3β antagonism could have therapeutic applications requiring angiogenesis since inactivation of the GSK3β gene in human endothelial progenitor cells appears to increase angiogenic cytokine secretion, such as vascular endothelial growth factor (VEGF) and interleukin-8, and is associated with enhanced survival, proliferation, and differentiation of these cells. Furthermore, genetically modified cell transplantation into ischemic hind limbs of mice has been shown to improve blood flow, limb salvage, and tissue capillary density. On the other hand, ischemia followed by administration of volatile anesthetics seems to activate salvage kinases and lessen reperfusion injury. Isoflurane protects against reperfusion damage by blocking GSK3β. In addition, it has been described that local kallikrein gene delivery in I/R involves GSK3β, which promotes cardiac myocyte survival. Short-term antagonism of GSK3β in myocardial ischemia may be useful, and further investigation in this field is required [1]. There has been some concern that chronic inhibition of GSK3β might promote the appearance of cardiac hypertrophy since this kinase has been described as having an antihypertrophic role. However, there is data suggesting that chronic GSK3β inhibition with lithium does not appear to favor cardiac hypertrophy. A number of selective and potent inhibitors of GSK3β have been identified, and further evidence is needed to define their potential in this specific therapeutic setting [1].

### 1.10 5′-AMP-Activated Protein Kinase (AMPK)

Since the early 1970s, AMPK has been recognized as an AMP-dependent kinase that inactivates 3-Hydroxy-3-methyl-glutaryl-Coenzime A (HMGCoA) reductase and acetyl-CoA carboxylase (ACC), which works as a cellular fuel sensor, that is activated when energy availability is reduced and AMP levels are relatively high, and decreases anabolic processes such as lipogenesis while enhancing glucose uptake [4]. Recently, AMPK has been recognized as an important regulator of energy homeostasis in the cell and coordinator of multiple catabolic and anabolic pathways in the heart. Several pathophysiological conditions such as ischemia, hypertrophy, diabetes, and certain forms of cardiomyopathies are associated with alterations in fatty acid and glucose metabolism, partly secondary to alterations in regulatory proteins. It is thought that in the heart, chronic

variations in energy substrate utilization involve shifts in acute control mechanisms, such as phosphorylation/dephosphorylation of key proteins, being AMPK centrally involved. In fact, it has been shown that acute increases in workload appear to activate AMPK due to increases in AMP levels, which seems to favor energy-generating pathways to maintain or restore intracellular ATP levels. Under acute conditions, such as mild ischemia or short intervals of severe ischemia, cardiac AMPK activation appears to stimulate ATP generation via increased glycolysis and accelerated fatty acid oxidation, thus being necessary for myocyte function and survival. AMPK seems to mediate the uptake and utilization of fatty acids by the heart via a number of processes, such as recruitment of cardiac lipoprotein lipase to the coronary lumen, translocation of the fatty acid transport protein CD36 (cluster of differentiation 36) from intracellular compartments to the plasma membrane, phosphorylation and subsequent inhibition of ACC, thus favoring the import of fatty acyl CoA into the mitochondria, which results in enhanced β-oxidation of fatty acids and subsequent ATP generation. Therefore, AMPK is a key mediator of fatty acid utilization in the heart by promoting fatty acid availability, fatty acid transport, and fatty acid oxidation [13, 14]. Several upstream AMPK kinases (AMPKKs) that potentially regulate its activity have been identified, such as serine/threonine kinase 11 (LKB1) and the calcium/calmodulin-dependent protein kinase kinases (CaMKKs α and β) among others. Defining their role in the heart will contribute to an increased understanding of cardiac AMPK regulation [13].

While AMPK activation may be crucial for adjustment of cardiac energy metabolism during acute and/or minor metabolic stresses, it is not known whether activation of this kinase turns out to be maladaptive in certain chronic disease states and/or during excessive energetic stress. In fact, changes in cardiac AMPK activity have been associated with cardiac hypertrophy, ischemia, glycogen storage cardiomyopathy, and Wolff–Parkinson–White syndrome. Since it is known that heart failure, myocardial ischemia and hypertrophy, diabetic cardiomyopathy, and lipotoxic heart disease are associated with changes in cardiac energy metabolism, and it has been established that AMPK is an important modulator of cardiac energy metabolism, directed AMPK regulation may be relevant in these pathological settings. Different stimuli, such as hyperosmotic stress, exercise, hypoxia, ischemia, leptin, adiponectin, alfa-adrenergic receptors, metformin, and long-chain fatty acids, activate AMPK [13, 14].

Basically, AMPK stimulation seems to be advantageous to the heart through the activation of energy-generating pathways and inhibition of energy-consuming pathways. In order to maintain constant ATP levels for proper cardiac function, activated AMPK increases energy-production by increasing fatty acid uptake and oxidation, stimulating glucose uptake, and favoring glycolysis, while decreasing protein synthesis and possibly lipid synthesis. Inhibition of energy-consuming pathways in a healthy heart that is metabolically stressed probably plays a minor role in maintaining mechanical function, whereas it may be more relevant in chronically or severely stressed hearts. AMPK regulates cardiac glucose utilization by stimulating glucose uptake, glycolysis, and possibly glycogen metabolism. Activation of cardiac AMPK appears to promote glucose transporter (GLUT) translocation to the myocyte membrane in an insulin-independent manner, and it may also favor GLUT4 to be retained in the sarcolemma, thus increasing overall GLUT4 levels in the plasma membrane [13, 14]. Inhibition of energy-consuming pathways that are not essential in an energetically compromised heart, such as protein synthesis, allows directing ATP toward crucial cellular processes such as contractile function and ion homeostasis [13].

### 1.11 Other Protein Kinases

The family of lysine deficient protein kinases (WNK) serine–threonine kinases has been shown to play an essential role in the regulation of electrolyte homeostasis [15]. It appears that WNK1 and WNK4 control diverse transcellular ion transport pathways and inhibit the thiazide-sensitive sodium/chloride cotransporter and the renal outer medullar potassium channel in the kidney. Furthermore, WNK4 seems to be localized to tight junctions and might also regulate paracellular ion flux. In the kidney, WNKs have been described to act as a molecular switch that allows alternating between sodium/chloride reabsorption and potassium secretion. In extrarenal epithelia, WNK4 seems to regulate other ion transporters and exchangers, suggesting pleiotropic roles for this kinase [15, 16].

## 2 PROTEIN KINASE-MEDIATED SIGNALING IN CARDIAC PATHOLOGY: A NOVEL THERAPEUTIC TARGET?

In response to diverse stresses, there are changes in cellular architecture, and increased activity of cytokines, catecholamines, and angiotensin signaling that initiate raised protein kinase activity in structural heart disease [2].

### 2.1 Cell Signaling in Hypertension

***Systemic Hypertension*** There is evidence that ROCKs in the brain stem contribute to regulate blood pressure

through the sympathetic nervous system [9]. Also in vascular smooth muscle cells, different stimuli, such as growth factors, cytokines, vasoconstrictors, and mechanical stretch, trigger intracellular signal transduction involving the epidermal growth factor receptor (EGFR), small G-proteins, and MAPKs [7]. Rho and ROCK have been implicated as mediators of enhanced contraction, hypertrophy, and migration of vascular smooth muscle cells and inhibition of endothelial function [7].

PKC antagonists could be valuable in the treatment of hypertension, especially when combined with ROCK inhibitors, since both enzymes have been described to play a relevant role in vasoconstriction at the arterial level. Interestingly, in small arteries PKC appears to regulate vasoconstriction without evident contribution of ROCK, and the nonspecific PKC inhibitor Ro 31-8220 seems to potentiate the vascular effects of calcium channel blockers [1].

There is data suggesting that pseudohypoaldosteronism type II (PHAII) might be related to mutations in two members of the WNK family of serine–threonine kinases, with evidence supporting a genetic gain of function mechanism. The kinases WNK1 and WNK4 are localized in distal nephron segments and are known to play a role in the homeostasis of electrolytes that are altered in PHAII. In fact, certain individuals with essential hypertension that have low plasma renin activity and respond well to thiazide diuretics might have WNK1 and WNK4 variants that underlie blood pressure variations. Therefore, WNK kinases and/or other components of their signaling pathway in the kidney may be potential new targets for the development of antihypertensive diuretic drugs [15–17].

***Pulmonary Arterial Hypertension*** Primary pulmonary hypertension is a fatal disease associated with endothelial dysfunction, increased contraction and proliferation of VSMCs, and inflammatory changes for which there is no definitive treatment available. Most therapeutic approaches for pulmonary hypertension including oral anticoagulants, vasodilators (prostaglandins, sildenafil, bosentan), and lung transplant have mixed clinical results leaving space for more effective therapeutic options to be developed. ROCK has been associated with vascular constriction and remodeling through myosin phosphatase inhibition and activation of downstream activators. Furthermore ROCK has been suggested to be an important mediator of arteriosclerosis and a couple of animal models of pulmonary hypertension have been reported to present ROCK activation. In fact, in these studies ROCK has been associated with enhanced pulmonary vasoconstriction and proliferating responses, endothelial dysfunction, and pulmonary remodeling [18].

### 2.2 Acute Coronary Syndromes and Stroke: Molecular Changes During Ischemia

There is evidence showing that myocardial ischemia induces activation of AMPK, although the consequences are still unclear. It is thought that AMPK is necessary for adiponectin's protective effect against I/R injury, and AMPK may activate downstream pathways. In fact, AMPK activates p38-MAPK in the heart, which may contribute to increased glucose uptake, and increased glucose utilization mediated by AMPK seems to be beneficial during and following ischemia. In transgenic mouse lines, targeted deletion of AMPK does not induce an obvious cardiac phenotype under normal workload, but glucose uptake is not increased in response to ischemia-reperfusion, which is probably related to insufficient GLUT4 translocation to the cell membrane. However, AMPK activation could be disadvantageous for the postischemic recovery. During ischemic episodes, oxidative metabolism is restrained while glucose uptake is amplified and glycolysis is increased. During the reperfusion period following mild ischemia or short intervals of severe ischemia, contractile function recovers when energy production has been reinstated. Even though there are many mechanisms that contribute to ischemic injury, recovery depends partly on the kind of substrate metabolized by the heart upon reperfusion. During ischemia fatty acid levels are elevated, and high fatty acid plasma levels are detrimental to both the ischemic and the reperfused heart. Furthermore, AMPK is activated during myocardial ischemia, which leads to phosphorylation and inhibition of ACC, increased GLUT4 levels in the cell membrane, and activation of PFK-2. Thus, activated AMPK increases ATP supplies overall through fatty acid oxidation and glycolysis. However, high rates of β-oxidation inhibit glucose oxidation rates via inhibition of the pyruvate dehydrogenase complex, and this eventually causes an imbalance between glucose oxidation and glycolysis with lactate and proton accumulation. It was originally proposed that activation of AMPK induced by ischemia favors a series of metabolic events that might be unfavorable to recovery of contractile function during reperfusion.

Although evidence from transgenic animals suggests that enhanced glucose uptake by AMPK may protect against ischemia/reperfusion injury, especially if the potentially negative consequences of increased fatty acid oxidation are not present. Therefore, further research in the area is necessary in order to understand whether AMPK activation is beneficial or detrimental to the I/R heart [13]. There is evidence using bovine brain microvessels that shows that endothelial cells subjected to stroke conditions have increased PKC-α, -β, and -ε levels and decreased PKC-γ after 6 h of hypoxia/

aglycemia. In this experimental model, PKC-α, -β, and -ε inhibition appears to downregulate the activity of the Na,K,2Cl-cotransporter (NKCC), and the use of a conventional PKC activator seems to upregulate it. These results imply that specific PKC inhibition or activation might impact outcome in stroke by brain ion transport [1].

A myocardial infarct model in rats provided evidence showing that treatment with LY333531 prevents reduced fractional shortening, increased matrix deposition and expression of TGF-β, which could be beneficial in preventing heart failure after infarction [1].

MAPKs are activated during ischemia/reperfusion injury and appear to contribute to structural as well as functional changes associated with myocardial ischemia, and evidence suggests that ERK1/ERK2 mediate survival pathway while p38 and JNK regulate a death pathway. It appears that cardiac protection dependent on ischemic preconditioning is associated with p38 activation, and direct p38 and JNK stimulation with anisomycin has also a protective effect. These contrasting effects may be explained by the activation of different p38 isoforms since p38α acts as proapoptotic, whereas p38β acts antiapoptotic in rat neonatal ventricular myocytes [19].

### 2.3 Atherosclerosis and Vascular Remodeling

The ROCK signaling pathway is known as a major modulator of vascular smooth muscle cell (VSMC) contraction, cell migration, proliferation, differentiation, as well as apoptosis, cell survival, and gene transcription [9]. This pathway seems to be involved in the regulation of endothelial barrier function, transendothelial leukocyte migration and inflammation, platelet activation and thrombosis, and oxidative stress. Activation of ROCK decreases eNOS expression by favoring eNOS messenger ribonucleic acid (mRNA) instability, since direct inhibition of Rho or inhibition of ROCK by hydroxyfasudil or overexpression of dominant-negative ROCK prevents eNOS downregulation [9]. It has been suggested that JNK could be a key mediator of atherosclerosis, since it has been shown that transfection with a dominant-negative JNK gene blocks neointimal formation in vivo and deletion of JNK2 in macrophages reduces atherosclerosis in mice. Furthermore, there is evidence showing that JNK2 promotes foam cell formation mediated by scavenger receptor A, which is associated with early atherogenesis. There is evidence that suggests that intimal hyperplasia after balloon injury in rat carotid artery can be inhibited by adenoviral gene delivery of dominant-negative ERK. In addition, the compound PD0185625, an oral selective ERK1/2 (p42/p44MAPK) inhibitor, may represent a novel therapeutic approach for vascular remodeling, since it appears to reduce neointimal formation after arterial injury [7].

### 2.4 Regulation of Platelet Function by Protein Kinases

After tissue injury, some effects of PI3Kγ seen in platelet aggregation and thrombus formation are apparently independent of chemokine receptors. It is well known that platelets respond to a wide range of GPCR agonists that activate PI3K signaling, such as thrombin, thromboxane, and ADP. However, PI3Kγ-deficient thrombocytes apparently have a normal response to both thrombin and thromboxane but show decreased aggregation when stimulated with ADP. This might be related to the fact that in platelets PI3Kγ is associated with the ADP receptor P2Y12, but not with the thrombin receptor, even if all three molecules are co-expressed in the same cell. Furthermore, PI3Kγ-deficient mice show normal bleeding time but have a decreased mortality rate in a model of ADP-induced thromboembolism [5].

Human platelets express all 4 p38 MAPK isoforms, and p38 α and β are the most abundant, although the exact role of these kinases in platelet function remains unclear. Different platelet agonists have been described to activate p38 in platelets. For example, in vitro stimulation of normal platelets with thrombin, collagen, and U46619, but not ristocetin, has been reported to result in strong p38 activation. However, it seems that platelet aggregation in this setting cannot be prevented by pretreatment with the p38 inhibitor VX-702. Even when it has been suggested that a potential role of p38 is to activate phospholipase A2, which triggers arachidonic acid metabolism and thus production of thromboxane, there is evidence that shows VX-702 has no effect on collagen-induced platelet aggregation (i.e., dependent on thromboxane production). Therefore, p38 block does not significantly affect platelet function and is not expected to be associated with elevated risk of bleeding [20].

### 2.5 Protein Phosphorylation and Cell Growth Regulation in Cardiac Hypertrophy and Heart Failure

Heart failure is a clinical syndrome initiated by a primary myocardial injury with subsequent activation of response mechanisms in the heart, vasculature, and neural and hormonal systems. Even though most of these mechanisms are initially compensatory, eventually many contribute to further worsening of heart function and progression of symptomatology. Currently, it has been shown that inhibition of the increased sympathetic nervous system and the renin–angiotensin system is a crucial element in the treatment of heart failure with a

positive effect on survival [21]. Therefore, it may be attractive to explore alternative antagonism of downstream elements of these pathways as therapeutic options in the management of these patients. However, it is difficult to single out the effects of specific signaling pathways, their intensity and duration of activation, and the potential adverse synergies when a given stimulus is considered as adaptive or maladaptive [22]. Furthermore, it is known that multiple pathways are involved in hypertrophic signaling, and these signaling networks are highly interconnected. It is common that alterations in one signaling pathway have consequences in other signaling cascades.

ERK1/2 appears to favor compensatory hypertrophy. JNK and p38MAPK activation is linked with both cardiac hypertrophy and apoptosis [1, 22, 23]. In fact, p38 activation in the heart has been observed in ischemia or pressure-overload-induced hypertrophy and heart failure. Furthermore, activation of p38 appears to decrease cell contractility by depressing myofilament response to intracellular calcium [23]. In fact, all MAPK-dependent signaling pathways appear to trigger a final common hypertrophic molecular program where the transcription factor myocyte-enhancing factor 2 (MEF2) participates [21].

Rho-kinase, which has been associated with the hypertrophic process in the heart since inhibition with fasudil, inhibits left-ventricular remodeling in experimental myocardial infarction [21]. It has been shown that ROCK is involved in left-ventricular remodeling after myocardial infarction (MI) by favoring the presence of pro-inflammatory cytokines, suggesting the therapeutic relevance of ROCK in preventing post-MI heart failure [9]. Furthermore, in response to pressure overload ROCK1-deficient (ROCK1–/–) mice had preserved compensatory hypertrophy but presented lower levels of perivascular and interstitial fibrosis. In addition, ROCK1-haploinsufficient (ROCK1+/–) mice showed in response to pressure overload cardiac fibrosis without hypertrophy [9]. ROCK's contribution to the development of cardiac hypertrophy is associated with enhanced oxidative stress by upregulation of endothelial reduced nicotinamide adenine dinucleotide phosphate (NADPH) oxidase and endothelial dysfunction [9].

CaMKII is upregulated in different models of cardiac hypertrophy and contributes to the loss of calcium homeostasis, typically seen during cardiac hypertrophy [24]. Transgenic mice overexpressing CaMK in the heart have severe cardiomyopathy [11]. Furthermore, CaMKIIδ overexpression causes cardiomyocyte hypertrophy, electrical remodeling, and arrhythmias [24]. In fact, when CaMKII is overexpressed in transgenic mouse hearts, it is associated with a prolonged calcium transient, enhanced L-type calcium currents, ryanodine receptor calcium leak, and loss of contractile function. The augmented ryanodine receptor leak due to PKA hyperphosphorylation is linked to structural heart disease in animal models and in patients [24]. In addition, the α-adrenergic signaling cascade activates CaMKII by a mechanism requiring PKC activation and turns on genes involved in cardiac hypertrophy, which is prevented by CaMKII inhibition [11].

In cardiac hypertrophy induced by either pressure or volume overload, energy metabolism is usually changed in order to reduce overall oxidative metabolism and increase glycolysis. Some of these alterations have been attributed to changes in expression of certain genes, and AMPK has been proposed to be directly responsible for increasing glycolysis via PFK-2 regulation. Apparently, switching to a more glycolytic state seems to be an adaptive response to chronic reduction in myocardial perfusion, which results from increased myocyte cell size. However, as the hypertrophic process advances, reductions in fatty acid oxidation may eventually deprive the heart of energy, and thus be considered as maladaptive. Even though, it is said that AMPK may also regulate other mediators of hypertrophy such as protein synthesis, its role in regulating cardiac myocyte growth remains controversial. In fact, while some studies show a correlation between AMPK activation and development of cardiac hypertrophy, other studies show hypertrophic growth inhibition mediated by AMPK activation [13, 14]. Mutations in human AMPK (γ subunit), which inhibit activation of AMPK by AMP, have been linked to hypertrophic cardiomyopathy. The AMPK mutation could lead to increased glycogen synthase activity, increased glycogen production, and the observed accumulation of glycogen in the heart [4, 14].

The role of mTOR in cardiac hypertrophy has been established after using its antagonist rapamycin to reduce this condition [21]. However, even when mTOR inhibition could be therapeutically attractive for the treatment of pathological hypertrophy, blocking other members of this pathway could have contrasting effects. For example, inhibition of AKT or PI3K (phosphatidyl inositol 3-kinase) under certain settings may not be advantageous because they modulate physiological and adaptive hypertrophy, which may be beneficial [22]. Transgenic mice that have chronic activation of Akt develop cardiac hypertrophy, which could be mediated through several mechanisms, including activation of molecules involved in protein synthesis, such as mTOR. Since Akt activation negatively regulates AMPK activity in the heart and pharmacological activation of AMPK attenuates cardiac hypertrophic growth and protein synthesis induced by constitutively active Akt, it has been proposed that inactivation of AMPK by Akt promotes protein synthesis in the hypertrophic myocardium. However, there are mouse models with impaired

AMPK activity that have normal or smaller heart weights compared to control hearts, which rejects AMPK inactivation as being essential for hypertrophic growth. Thus, despite the fact that AMPK activation prevents cardiac hypertrophy, further proof is needed to define whether AMPK is a necessary component of the hypertrophic signaling cascade. In fact, it is not known whether decreased AMPK activity renders the heart more susceptible to hypertrophic growth stimuli. On the other hand, adiponectin-deficient mice have decreased AMPK signaling and are more susceptible to pressure-overload-induced cardiac hypertrophy. Furthermore, adiponectin reversed the severity of cardiac hypertrophy in these animals and restored AMPK signaling, whereas expression of an AMPK-dominant negative mutant blocked the beneficial effects. However, it is not known whether AMPK's direct role in the hypertrophic process seen in adiponectin-deficient mice is relevant in other models of cardiac hypertrophy [13, 14]. Indeed, it has been suggested that activators of AMPK such as adiponectin, metformin, or thiazolidinediones may be used to prevent hypertrophic growth, even though the doses of metformin needed to activate AMPK in vitro are considerably higher than those seen clinically. Thus, it is thought that more specific compounds are needed in order to explore the value of AMPK activators for the prevention of cardiac hypertrophy [13].

Several animal models have linked PI3K with cardiac hypertrophy and ischemia and imply a specific role for the different isoforms. For example, it appears that PI3K is activated during ischemia and pressure overload. Furthermore, overexpression of PI3Kα, one of the isoforms activated by receptor tyrosine kinases, produces compensatory hypertrophy without cardiac dysfunction, while activation of PI3Kγ induces hypertrophy followed by heart failure. In fact, activation of receptor tyrosine kinases results in adaptive hypertrophy via PI3Kα, while stimulation of GPCRs and PI3Kγ causes maladaptive cardiac hypertrophy. In vivo PI3Kα inhibition seems to negatively modulate heart size while PI3Kγ inhibition appears to negatively modulate cardiac contractility [1].

The progress in basic science has clearly expanded the range of possible therapeutic targets for heart failure [21]. The use of mouse models of cardiomyopathy has provided new knowledge about the cellular and molecular basis of heart failure, which exposes a large number of potential targets for pharmacotherapy or gene therapy. Nonetheless, most studies and therapies used clinically to treat heart failure have mainly targeted cell surface receptors or the intracellular mineralocorticoid receptor. The specific targeting of intracellular signaling pathways is a promising new approach for drug development. Even when small-molecule inhibitors are available for pathways that contribute to myocyte hypertrophy in the heart, the use of these compounds is limited by their poor cardiac selectivity and associated toxicity [21].

### 2.6 Metabolic Syndrome, Diabetes, and Hyperlipidemia

Activators of AMPK could be useful to treat metabolic syndrome, diabetes, or hyperlipidemia [4] since there is evidence in vitro and in vivo showing that inactivation of hepatic AMPK is a key event in the pathogenesis of hyperlipidemia in diabetes [25]. In fact, AMPK inhibition has been observed in pancreatic cell lines exposed to elevated glucose [26], in the hearts of obese Zucker diabetic fatty rats and ob/ob mice [27], and in cultured human skeletal muscle of obese type II diabetic patients [28]. There is also evidence suggesting that class IA PI3K signaling is necessary to mediate insulin functions in vivo, and the loss of activity in this kinase could cause complex metabolic defects that may be linked to type II diabetes [5]. Furthermore, insulin favors glucose uptake into muscle by glucose transporter GLUT4 translocation to the plasma membrane, which is mediated by a PI3K-dependent pathway that requires AKT (i.e., the β isoform of protein kinase B) [12].

Overexpression of PKC-β (but not PKC-α, -δ, or -ζ) in rat endothelial cells was shown to decrease insulin and VEGF-induced Akt activation and eNOS expression. Studies in transgenic mice suggest that PKC-β activation in endothelial cells and vascular tissue appears to inhibit Akt-mediated eNOS regulation by insulin, and causes endothelial dysfunction associated with insulin resistance and obesity. Furthermore, in healthy humans PKC-β inhibition with LY333531 seems to prevent the decrease in endothelium-mediated vasodilation associated with acute hyperglycemia. This data suggests that hyperglycemia may impair endothelial function by mechanisms that involve PKC-β [1].

Studies in streptozotocin-induced type I diabetes suggest that oxidative stress and PKCβ2 related to chronic hyperglycemia may participate in the development of cardiomyopathy, since diabetic rat hearts appear to have increased expression of PKCb2 with concomitant myocyte hypertrophy and cardiac fibrosis. In addition, in vitro results show that nonspecific activation of PKC is enough to induce the same degree of hypertrophy as hyperglycemia. In fact, hypertrophy induced by a PKC activator (phorbol 12-myristate 13-acetate [PMA]) in cultured neonatal cardiomyocytes appears to be prevented by specific PKCβ2 inhibition, but not by PKCd antagonism; thus PKCb2 may be a prohypertrophic signal. However, specific PKCβ2 inhibition attenuates but does not completely prevent cardiomyocyte hypertrophy induced by hyperglycemia [29].

JNK activity is also elevated in obesity and diabetes. In two different models of murine obesity, the absence of JNK1 appears to result in decreased adiposity with improved insulin sensitivity and enhanced signaling capacity of the insulin receptor. Moreover, JNK2 has been linked to type I diabetes since deletion of JNK2 (knockout) in nonobese diabetic mice seems to reduce disease progression and development of type I diabetes. It has been suggested that JNK2 activity participates in controlling the immune response balance and represents a protecting mechanism against autoimmune diabetes [7].

### 2.7 Arrhythmias

There is evidence that increased and disordered intracellular calcium plays a role in arrhythmogenesis. In fact, loss of normal intracellular calcium homeostasis occurs in almost all forms of cardiac disease associated with fatal arrhythmias [24, 30]. For example, arrhythmias following drug-related excessive prolongation of repolarization (long QT syndromes) or electrical remodeling in cardiomyopathy are linked to disordered intracellular calcium. Moreover, ventricular fibrillation triggered by ischemia is also coupled with deregulation of intracellular calcium. However, even when intracellular calcium is consistently linked to arrhythmias, increases in calcium levels alone may not be sufficient to set off arrhythmias since it appears that a further step is required, where a calcium-sensing protein translates it into distorted cellular function [30].

The ubiquitous serine–threonine protein kinase CaMKII has been implicated in arrhythmogenesis and structural heart disease since CaMKII appears to be involved in regulating atrioventricular (AV) nodal conduction, and CaMKII activity and expression are higher in hearts prone to arrhythmia. Additionally, overexpression of CaMKII in the heart favors the development of cardiac hypertrophy, dysfunction, and proarrhythmogenic remodeling, while CaMKII inhibition decreases arrhythmias [1]. CaMKII has been recognized as a proarrhythmic signaling molecule since it can orchestrate arrhythmia signaling through ion channel proteins, intracellular calcium uptake and release, regulation of cell death, and activation of prohypertrophic pathways. CaMKII shares many of the same downstream protein targets as PKA, which is the kinase most associated with β-adrenergic receptor stimulation [24]. It has been described that calcium, calmodulin, and CaMKII interact in the feedback mechanisms for LTCC regulation, cardiac contraction, and arrhythmias in model systems involving smooth muscle as well as cardiac myocytes. Modulation of LTCC is critical for calcium homeostasis and signaling; thus defective LTCC regulation contributes to disease states associated with abnormal levels of calcium influx. In addition, CaMKII-induced phosphorylation favors channels to be in a high-activity gating mode, which leads to prolonged opening. Besides, CaMKII modulates calcium release via RyR since a CaMK inhibitor (peptide AC3I) enhances RyR calcium release while blocking calcium currents in acutely isolated rabbit cardiac myocytes, thus enhancing the gain of excitation-contraction coupling. In fact, addition of a constitutively active CaMKII yields the opposite result, suggesting that this kinase could be a functional link between LTCC and RyR during excitation-contraction coupling. Cardiac overexpression of CaMKIIδ in transgenic mice is associated with myocardial hypertrophy, left-ventricular chamber dilation, loss of normal contractile function, proarrhythmic electrical remodeling, and sudden death [11, 24].

Chronic overexpression of CaMKIIδ generates altered calcium handling with enhanced calcium sparks from the RyR, reduced SR content, and severe cardiac hypertrophy. This could be a direct result or an indirect result through changes in the protein level of other key calcium-cycling proteins associated with CaMKII overexpression, such as a decrease in SERCA and phospholamban (PLB) and an increase in the Na/Ca exchanger. Thus, even though CaMKII is an important regulator of intracellular calcium release, details of this process remain to be elucidated. It has been suggested that enhanced RyR calcium leak or proarrhythmic inward calcium current due to altered CaMKII activity can lead to increased Na/Ca exchanger activity, thus favoring the probability of producing delayed afterdepolarization. Another trigger for arrhythmias is stimulation by the β-adrenergic pathway, which leads to increased heart rate and which recruits CaMKII and enhances calcium cycling. These findings suggest that the efficacy of beta-blockers in preventing sudden cardiac death may be in part related to their inhibition of CaMKII activity [11].

It has been proposed that CaMKII directly regulates SERCA activity and phosphorylates PLB, and physiological increases in phosphorylation by CaMKII probably follow catecholamine-stimulated chronotropic and inotropic changes. Consequently, this could correlate with an adaptive response to sudden modifications in heart rate. In fact, chronic CaMKII inhibition apparently reduces the variability of calcium-induced calcium release in cardiac myocytes by regulating calcium release from SR stores and calcium reuptake by SERCA [11]. CaMKII upregulation may be associated with increased activity of calcium channels and arrhythmias in heart failure [24].

Pharmacologic probes, such as W-7, block calmodulin action and appear to suppress torsades de pointes in a rabbit model without shortening the QT interval.

Furthermore, in patients taking amiodarone, which is a calmodulin antagonist, there is a relative lack of proarrhythmia that follows QT interval prolongation suggesting that calcium-bound calmodulin is involved in triggering torsades de pointes [30].

Interestingly, mutations in human AMPK (γ subunit) have been linked to ventricular pre-excitation [4, 13]. These mutations appear to be linked to the development of hypertrophic cardiomyopathy associated with excessive glycogen storage and the Wolff–Parkinson–White syndrome. In fact, patients with mutations in the AMPK γ2 gene show a familial form of the Wolff–Parkinson–White syndrome with characteristic conduction system abnormalities that might be related to altered AMPK activity. Altered AMPK activity seems to be accountable for excessive glycogen deposition within the cardiac myocytes, which favors bypass tracts that lead to the development of the arrhythmia. However, altered AMPK activity might also have an effect on other targets, such as ion channels, since AMPK activation augments hH1 sodium channel activity in cardiac myocytes. It has also been determined that several of these naturally occurring mutations in the human γ2 subunit do not lead to constitutively active AMPK enzymes but alter AMP's ability to stimulate AMPK. Furthermore, transgenic mouse lines expressing these mutations in a heart-specific manner have consistent phenotypes with those of humans that have mutations in the γ2 subunit. However, the presence of higher levels of the γ2 subunit alone appear to have effects on the cardiac myocyte since control mice expressing the nonmutated subunit had no changes in AMPK activity but also developed increased glycogen deposition [13]. Contrasting results make it difficult to conclude whether deregulation of AMPK activity contributes directly to the development of Wolff–Parkinson–White syndrome. Transgenic mice with two different mutations in the γ2 subunit appear to have opposing effects on AMPK activity in vivo with almost identical phenotypes. It is proposed that AMPK-activating mutations promote glycogen accumulation by increasing glucose uptake surpassing glucose utilization, whereas AMPK-inactivating mutations enhance glycogen deposition by increasing glycogen synthesis. Furthermore, transgenic mice with the R531G mutation develop a progressive cardiac phenotype with normal function and AMPK activity in early life but with cardiac hypertrophy, significant glycogen accumulation, and reduced AMPK activity at 4 weeks of age. Therefore, simple loss or gain of function in AMPK does not explain the pathogenesis of the Wolff–Parkinson–White syndrome since the altered AMPK activity in these mice develops gradually and likely following glycogen buildup. As a result, deregulation of AMPK activity appears to contribute to the development of disease but further studies are needed to clarify this [13].

Several in vivo studies using transgenic mice suggest that CaMKII targets the LTCC and that CaMKII inhibition might reduce arrhythmias related to increased probability of LTCC opening. However, increases in LTCC opening do not appear to be the only proarrhythmogenic mechanism associated with CaMKII, and compensatory responses to chronic CaMKII inhibition are not predictable, and patients placed on chronic CaMKII blockade might exhibit compensatory responses, such as reverse electrical remodeling [24].

Finally, atrial tissue from patients with atrial fibrillation has been shown to present CaMKII upregulation with CaMKII-induced hyperphosphorylation of phospholamban when compared to control patients in sinus rhythm. In addition, it has been described that CaMKII activity seems to be important for stimulating the atrioventricular node impulse transmission at baseline and in response to catecholamine stimulation. Therefore, CaMKII inhibition could be used for ventricular rate control in the setting of atrial fibrillation. However, it is not known if CaMKII antagonism can affect atrial fibrillation [24].

## 3 STRUCTURAL AND FUNCTIONAL BIOLOGY OF PROTEIN KINASES IN THE HEART AND VASCULATURE: MODULATION IN CARDIOVASCULAR DISEASE

Protein kinases play a central role in signal transduction, and most aspects of cell life are controlled by the reversible phosphorylation of proteins. The development of specific, cell-permeable protein kinase antagonists has been crucial in the study of signal transduction processes so far, and they have been extensively used to establish the relevance and involvement of a particular protein kinase in health and disease. The main challenges in designing protein kinase inhibitors are (1) to understand what determines its selectivity and (2) being able to select the appropriate clinical setting where its use could potentially be beneficial with an adequate safety and efficacy profile. The great number of protein kinases, their close structural and sequence similarity, and their involvement in normal cell function complicate the design and use of compounds that target these proteins as therapeutic options in cardiovascular disease. Even though the availability of many kinase crystal structures has improved our knowledge of factors that determine selectivity at the molecular level, many small-molecule kinase inhibitors do not seem to reach optimal selectivity. Most clinical trials with protein kinase inhibitors are mainly in oncology; however, there

is a trend toward targeting protein kinases for the treatment of a number of chronic conditions other than cancer, including inflammatory and cardiovascular diseases [4].

### 3.1 Mitogen-Activated Protein Kinase (MAPK) Pathways

It is known that p38 inhibitors have anti-inflammatory effects in different animal models primarily by decreasing the expression of inflammatory mediators. Consequently, several p38 inhibitors have been evaluated in preclinical or clinical studies. Nevertheless, the recurring failure of p38 inhibitors in the clinic suggests that new approaches are needed for designing p38 antagonists. Pyridinyl imidazole inhibitors have been important to identify and characterize the p38 pathway, and with subsequent variation of the original structures several drug candidates such as SB203580 emerged. These first-generation antagonists, though, have been found to be hepatotoxic and potential carcinogens. There are other inhibitors that are structurally diverse (e.g., arylamides and diarylureas, indolylamides, and ketone-derived inhibitors). The p38 inhibitors differ in chemical structure and binding mode, and there are potent and selective inhibitors that have been included in clinical trials. However, only a few lead compounds (BIRB796, SCIO469, and VX702) have progressed to phase II trials and BIRB796 has been withdrawn. One of the main causes for stopping development of these inhibitors is related to an undesirable safety profile, with hepatic and neurological adverse effects and other safety concerns such as infection, skin, and renal disorders. It has not been determined whether the secondary effects are associated directly with p38 inhibition or to other effects of the drugs, but evidence showing different toxicity profiles for each structural class of inhibitors suggests that selectivity plays a role in the unsuccessful use of p38 inhibitors in the clinical setting. In fact, certain p38 inhibitors such as SB203580, SB202190, VX745, and BIRB796, which were thought to be highly selective, have been shown to affect other kinases and certain cross-reactivity with other cellular proteins might also occur [31].

***Acute Coronary Syndromes (ACS)*** It is well known that two kinds of stress-activated MAPKs (JNK and p38) are activated by ischemia/reperfusion (I/R), and there is some indication that inhibition of either might prove helpful for treating ACS. However, it has been difficult to validate whether activation of these kinases is beneficial or detrimental [4]. Some p38 inhibitors, such as SB203580 and SB242710, have shown to reduce I/R-induced apoptosis and preserve cardiac function in a rabbit heart model. Since the heart in this model is perfused and therefore there are no leukocytes, the findings suggest that p38 inhibition has additional reperfusion directly on the myocardium aside from its known effects on leukocyte recruitment and activation. It is thought that this leukocyte-independent protective effect of p38 inhibition on the heart could be associated to a decrease of I/R-induced production of cytotoxic cytokines and blocking of p38-dependent proapoptotic pathways in cardiomyocytes [4]. A p38 inhibitor that can be used in vivo models (SB239063) has beneficial effects in the intact rat model of I/R injury. The antagonist directly protects cardiomyocyte survival and produces a dramatic reduction of the inflammatory response in the heart, as evidenced by decreased upregulation of P-selectin and intercellular adhesion molecules and lower neutrophil accumulation within the ischemic zone. Also, some p38 inhibitors help preserve mechanical function of cold-stored hearts before transplantation, which may be related to augmented contractility due to enhanced calcium responsiveness of myofilaments [4].

Blocking the release of inflammatory cytokines and chemokines could be a promising approach against ACS and might slow the development and progression of atherosclerotic plaque. Indeed, VX702, a p38 inhibitor, is currently being tested in a phase II clinical trial with patients that have ACS with unstable angina [4, 20]. Preliminary safety data showed no clinically significant adverse events, such as bleeding and arrhythmias. Additionally, VX-702 was reported to reduce serum levels of C-reactive protein in patients with ACS undergoing percutaneous coronary intervention (PCI), and these remained significantly lower up to 4 weeks after a 5-day dosing period [8, 20].

Evidence suggests that p38a may play a key role in the pathogenesis of cardiac dysfunction, and its inhibition may have cardioprotective effects associated with attenuation of the inflammatory response and the functional deterioration present in acute myocardial infarction. In a rat model of L-NAME/ANG-II-induced myocardial injury, inhibition of p38a with the selective oral antagonist, SD-282, has been shown to be related to improvement of cardiac function (i.e., increasing stroke volume, cardiac output, ejection fraction and stroke work, and decreasing arterial elastance), reduction of inflammatory cell infiltration (i.e., macrophages), and abolishment of myocyte apoptosis [32]. On the other hand, p38 and ERK1/2 activation may play a cardioprotective role. There are data suggesting that cannabinoids can protect against ischemia in animal models, and the use of a selective CB2 receptor agonist JWH133 appears to reduce infarct size. Studies with pharmacological inhibitors suggest that p38 ERK1/2 and PKC contribute to the protective effect of cannabinoids.

Furthermore, JWH133 in healthy hearts seems to increase p38 phosphorylation and activity, and the cannabinoid receptor 1 (CB1) receptor agonist arachidonyl-2-chloroethylamide (ACEA) appears to increase ERK1/2 phosphorylation without any change in activity. During ischemia, both agonists increased p38 activity, and arachidonyl-2-chloroethylamide (ACEA) favored ERK1/2 phosphorylation and activity during reperfusion [33].

***Heart Failure*** Inhibition of p38 by SB203580 may enhance cell contractility in a dose-dependent manner and reduce cardiac hypertrophy [23]. It is thought that p38 is involved in mediating the action of pro-inflammatory cytokines. In fact, the p38 inhibitor RWJ-67657 appears to reduce remodeling and fibrosis in rats with myocardial infarction secondary to permanent coronary ligation. In fact, some of these compounds seem to be potent antagonists of the tumor necrosis factor (TNF) and other pro-inflammatory cytokines in humans. Clinical trials are currently underway for non-cardiovascular indications (e.g., rheumatoid arthritis and Crohn's disease), and cardiovascular disease (e.g., atherosclerosis) [21].

It could be possible that a broad anticytokine therapy in heart failure using p38 inhibitors could be promising; however, the RENEWAL (Randomized Etanercept Worldwide Evaluation) and ATTACH (Anti-TNF Therapy Against Congestive Heart Failure) trials, targeting TNF-α in heart failure with a monoclonal antibody or a soluble receptor, had negative results and raised concerns [4].

It remains to be determined how the knowledge of MAPK signaling in the heart could be used for treating human pathology since pharmacological ERK1/2 inhibition could decrease the beneficial aspects associated with compensated hypertrophy, whereas p38 or JNK inhibition may lead to a maladaptive hypertrophic response in the long term [22].

***Stroke*** Antagonism of several MAPKs, such as extracellular signal-regulated kinases (ERKs), JNKs, and p38, has been shown to be beneficial in animal models of stroke [4]. It appears that the MEK (mitogen-activated protein kinase kinase)-ERK pathway is turned on during focal cerebral ischemia and participates in the induction of damage in mice since treatment with the MEK1-specific inhibitor PD98059 before ischemia reduces focal infarct volume up to 72 h after transient occlusion of the midcerebral artery in a dose-dependent manner and with attenuation of neurological deficit, thus conferring a potent neuroprotective effect. Furthermore, PD98059 does not seem to alter respiratory and cardiovascular parameters nor produce hypothermia in this setting. On the other hand, in this model of focal cerebral ischemia, p38 and c-Jun phosphorylation do not appear to be increased, and, thus, no protection against brain injury is apparent with the p38 and JNKs inhibitor SB203580 [34].

In rat models of focal ischemic stroke there seems to be p38 activation in the brain, and oral or intravenous administration of the p38 antagonist SB239063, which has high selectivity and improved cellular and in vivo activity profiles, has been shown to reduce infarct size and neurological deficit. In addition to reducing p38 activity, SB239063 also seems to decrease stroke-induced expression of cytokines associated with brain injury [e.g., interleukin-1β (IL-1β) and TNF-α] [35].

Even though ERKs have generally been thought to be antiapoptotic in most settings, including I/R injury in the heart, neuroprotection in vivo has been achieved by direct injection into the cerebral ventricles of a MEK1/2 inhibitor (PD98059), which prevents ERK activation. In addition, studies showed that intravenous administration of another MEK1/2 inhibitor (U0126) was also protective against forebrain and focal cerebral ischemia. It has been suggested that in the brain, the mechanism of protection by ERK inhibition may involve prevention of excitotoxicity (i.e., neuronal death caused by release of excitatory amino acids that activate glutamate receptors), which plays a critical role in I/R injury. However, both inhibitors, PD98059 and U0126, also block MEK5, the kinase that activates ERK5, and it is not known whether this is relevant for the described effect [4]. In a phase I dose-ranging trial in cancer patients, another MEK1/2 inhibitor (PD184352) has been well tolerated with only fatigue, rash, and diarrhea when administered orally twice daily, for periods of 3 weeks and repeating every 4 weeks [4].

Some other possible targets for neuroprotection include the inactivation of JNK3 in the central nervous system via gene deletion in knockout mice, and inhibition of p38 by SB239063 is also protective since it reduced stroke-induced expression of TNF-α and IL-1β, which seem to enhance neuronal loss after I/R [4]. Another potent p38 inhibitor (SCIO 323) entered phase I clinical trials and showed potential therapeutic value against cerebral ischemia and diabetes mellitus [8].

### 3.2 G-Protein-Coupled Receptor Kinases (GRKs)

β-ARK1 inhibition may be a therapeutic option in human heart failure. Inhibition of cardiac β-ARK1 (beta-adrenergic receptor kinase 1) activity with a Gβγ sequestering peptide (β-ARKct) in transgenic models of heart failure appears to restore the ventricular contractile response to β-adrenergic stimulation. In fact, β-ARKct has been described to prolong survival

and improve cardiac function in a murine model of severe cardiomyopathy due to calsequestrin overexpression. Expression of β-ARKct in failing human cardiomyocytes seems to improve β-adrenergic responsiveness and contractile function. There is data in transgenic mice with pressure overload and varying degrees of cardiac-specific expression of β-ARKct peptide that suggest a correlation between the level of β-ARK1 inhibition and the degree of heart failure. Moreover, there seems to be a positive linear relationship between β-ARKct expression and fractional shortening (i.e., animals with high β-ARKct expression show less cardiac deterioration, preserved adenylyl cyclase activity with isoproterenol stimulation, and normal β-AR density), suggesting that the degree of β-ARK1 inhibition could have relevant therapeutic repercussions. Other signaling proteins that interact with β-ARK1 might be attractive therapeutic targets since there is no pharmacological inhibitor currently available [1].

### 3.3 Phosphatidylinositol 3-Kinase (PI3K) Signaling

Since PI3Kγ has a relevant role in regulating vascular smooth muscle tone, cardiac cell contractility, and inflammatory responses, it is an attractive target for cardioprotective therapies. There is interest in finding novel agents with dual effects on both the mechanical (pressure and contractility) and the inflammatory components of cardiovascular disease. In fact, this would match the expected profile from an ideal PI3Kγ inhibitor, based on data obtained from experimental animal models of cardiac overload and inflammation [5]. There is data suggesting that PI3Kγ expressed in smooth-muscle cells participates in regulating the balance between vasodilation and vasoconstriction, which is mediated by vasoactive GPCR agonists, such as angiotensin (Ang) II. It has been described that Ang-II-induced vasoconstriction in smooth-muscle cells is impaired by nonselective PI3K inhibitors, but only selective antagonism of PI3Kγ inhibits the Ang-II-induced phosphatidil inositol cascade and calcium influx. In fact, transgenic mice deficient in PI3Kγ are resistant to the hypertensive effect of chronic Ang II exposure. Consequently, PI3Kγ inhibition may potentially be useful to treat high blood pressure [5]. Moreover, PI3Kγ in cardiac myocytes regulates contractility without participating in growth control under normal conditions, and the absence of PI3Kγ function favors increased contractility in cardiomyocytes [5].

Recently, intravenous PI3Kγ inhibitors, such as TG100-115 (Targegen), are being evaluated for the treatment of acute myocardial infarction. Preclinical data show that TG100-115 lowers vascular permeability and reduces myocardial infarct development with improved myocardial contractility in animal models. The compound has been described to be selective for PI3K compared to other kinases and has limited selectivity within the class I PI3Ks. Apparently, no PI3Kγ inhibitor has progressed to the clinical setting yet since further research is needed in this area, and the initial patent applications are still very recent. There is a need to identify potent and selective inhibitors with acceptable pharmacokinetic properties that are effective and safe. When developing a new drug for the treatment of a given disease, the right balance must be achieved with reasonable isoform selectivity and potency combined with maximum efficacy and minimal unwanted adverse effects. Since there appears to be interdependence and cooperation between class IA and IB PI3K isoforms in the same cell type, dual inhibition of PI3Kγ and of a class IA PI3K (depending on the cell type, e.g., PI3Kβ in platelet aggregation and PI3Kα in cardiac myocyte function) might be desirable to reach a therapeutically useful effect [5]. Animal studies suggest that targeting PI3K might have similar results to those associated with βARK1 blocking since PI3Kγ interacts directly with βARK1 and, thus impacts cell contractility through β-adrenergic desensitization. Given the individual roles of the different PI3K isoforms, a PI3Kγ-specific inhibitor could attenuate its unwanted effects on calcium handling, β-adrenergic signaling, cardiac contractility, and hypertrophy [1].

The experimental evidence available validates PI3Kγ as a high potential target for the treatment of various cardiovascular disorders, such as thrombosis, atherosclerosis, and cardiac hypertrophy. PI3Kγ is involved in several different processes relevant to human disease but controls only a fraction of these. Thus, even potent and selective PI3Kγ antagonism would be expected to just reduce rather than eliminate any given biochemical event, and this subtle form of inhibition would affect several processes implicated in the multifactorial pathogenesis simultaneously. This would result in a concurrent, multibranched mode of action in different cell types and tissues with a beneficial overall therapeutic effect in a specific disease setting. Based on the restricted expression pattern of PI3Kγ, its limited functional association to a couple of GPCRs, and the benign phenotypes of PI3Kγ knockout and knock-in mice, the initial concerns about safety of selective PI3Kγ inhibitors have largely decreased. However, potential toxicity associated with PI3Kγ antagonists, including interference with innate immunity, blood pressure, or wound healing, cannot be ruled out [5]. It has been suggested that in vitro inhibition of PI3K using wortmannin or LY294002 in cultured cardiac myocytes decreases basal protein synthesis in resting cells and after stimulation

with phenylephrine or insulin. In addition, PI3K inhibition modulates gene expression, apparently through GSK3β [1].

Available PI3K inhibitors, such as wortmannin and LY294002, have been used to define PI3K-mediated signaling, but they have poor selectivity for different isoforms [1]. In fact, wortmannin can also inhibit other PI3K-related enzymes such as mammalian target of rapamycin (mTOR), DNA-dependent protein kinase (DNA-PK), some phosphatidylinositol 4-kinases, myosin light-chain kinase (MLCK), and mitogen-activated protein kinase (MAPK) at high concentrations [36, 37]. Other agents, such as AS-604850 and AS-605240, have been shown in vitro and in vivo to selectively inhibit PI3Kγ enzymatic activity, signaling, and chemotaxis. There is evidence to suggest that oral AS-605240 prevents progression of joint inflammation and damage in mouse models of rheumatoid arthritis as well as glomerulonephritis in murine models of systemic lupus. The efficacy of this inhibitor in cardiovascular disease is not known [1].

### 3.4 Rho-Dependent Serine–Threonine Kinases (ROCKs)

In several animal models and human trials, potentially beneficial effects of ROCK inhibitors in cardiovascular disease related to VSMC hyperconstriction (such as coronary and cerebral vasospasm, pulmonary hypertension, and systemic hypertension) have been identified. In addition, these compounds might also show benefits if used to treat stroke, coronary restenosis, I/R injury, heart failure, myocardial hypertrophy, cardiac allograft vasculopathy, and vein graft disease [9]. Fasudil has shown that ROCK mediates increased vascular resistance and impaired vasodilation in patients with heart failure patients or hypercholesterolemia. In fact, ROCK may participate in hemodynamic abnormalities seen in metabolic syndrome and obesity [7]. However, further studies are needed to prove the potential usefulness of oral ROCK inhibitors for the treatment of myocardial infarct, unstable angina, pulmonary hypertension, hypertensive vascular disease, and cardiac hypertrophy in humans [9].

***Current Status of the Development of Selective ROCK Inhibitors (Table 3)*** [9] Fasudil is the only ROCK inhibitor available for clinical use and is accessible only in Japan to treat cerebral vasospasm after subarachnoid hemorrhage. Clinical trials to test its antianginal effects in patients with stable angina have demonstrated that long-term oral treatment with this drug is safe and effective, improving the ST-segment depression in patients with stable angina; the drug was well tolerated and did not affect heart rate or blood pressure [9]. Fasudil selectively competes with ATP for binding to ROCK, and its

**TABLE 3 ROCK Inhibitors under Current Development to Treat Cardiovascular Disease and Other Pathologies**

| Compound | Stage of Testing | Indication |
|---|---|---|
| Fasudil (Asahi Kasei Pharma—Japan) | Launched | Cerebral vasospasm |
| | Registered | Acute cerebral ischemia |
| | Phase II | Angina |
| | Phase I | Pulmonary hypertension |
| GSK269962A (GlaxoSmithKline—UK) | Preclinical | Anti-inflammatory and vasodilator activity |
| SB772077B | Preclinical | Vasodilator activity |
| SLx2119 (Surface Logix—USA) | Preclinical | Atherosclerosis |
| NA (Mitsubishi Pharma—Japan) | Patented—Preclinical | Reperfusion injury, renal disease, ischemia |
| Wf536 (Mitsubishi Pharma—Japan) | Preclinical | Metastasis |
| Y39983 (Mitsubishi Pharma—Japan) | Phase I | Glaucoma |
| H1152P (Kowa—Japan) | Preclinical | Glaucoma |
| New Compounds | Stage of testing | Potential Indications |
| **Over 11 companies in 6 countries have potential agents** | Patented | **Cardiovascular disease:** angina, atherosclerosis, hypertension, cerebrovascular ischemia, congestive heart failure, reperfusion injury, renal disease<br>**Inflammatory disease:** asthma, rheumatoid arthritis<br>**CNS disease:** Alzheimer's disease, Huntington's chorea, spinal cord injury<br>**Ophthalmic disease:** corneal disease, glaucoma, diabetic retinopathy, lacrimal gland disease<br>**Others:** erectile dysfunction, hearing disorder, osteoporosis, fibrosis, metabolic disorder, cancer, HIV infection |

*Source*: Adapted from ref. 9.

major active metabolite after oral administration, hydroxyfasudil, has a more selective inhibitory effect on this kinase than the parent drug. However, both compounds exert nonspecific inhibitory effects on other serine/threonine kinases such as PKA and PKC [9]. At micromolar concentrations, fasudil inhibits several protein kinases, including ROCKs, but it is unclear whether its clinical efficacy results from inhibition of this or other protein kinases, or whether it is due to a non-kinase effect [6]. Certain fasudil analogs have proven to be highly specific inhibitors of ROCK, and compounds with a better ROCK2 inhibitory profile have been developed. These antagonists have been used to reduce intraocular pressure in rabbit ocular hypertensive models, which indicates that they might be useful for glaucoma, in addition to other ROCK-related diseases. Recently, GSK269962A and SB772077B have been described as members of a novel class of compounds that have similar inhibitory effects on ROCK1 and ROCK2 as Y27632 or fasudil but with higher potency [9]. Furthermore, an orally bioavailable, potent, and highly selective ROCK2 inhibitor, SLx2119, has been developed recently [9], but clinical trials are still pending.

***Hypertension*** Selective ROCK inhibitors might potentially be a novel approach to the treatment of hypertension in the future [4]. Several studies in spontaneously hypertensive rats, deoxycorticosterone acetate (DOCA)-salt-treated hypertensive rats, and renal hypertensive rats have shown that specific ROCK inhibitors, such as Y-27632, GSK269962A, and SB772077B, have vasodilator activity and lower systolic blood pressure, not affecting blood pressure in normal rats. However, this specific compound apparently affects contraction of smooth muscle cells outside the vasculature, as well as metastasis and neurite outgrowth; therefore the safety of Y-27632 and related agents remains to be evaluated in clinical trials [4, 9].

Intra-arterial infusion of fasudil enhances the vasodilator responses of forearm circulation in hypertensive patients; in fact, it has been reported to be significantly greater in hypertensive patients than in normal subjects. Thus, Rho-kinase may be involved in the pathogenesis of the increased peripheral vascular resistance in human hypertension [9, 38].

***Coronary Vasospasm and Angina*** ROCKs phosphorylate the myosin-binding subunit of myosin light-chain phosphatase and LIM kinase, which ultimately regulates myosin light-chain phosphorylation and thus vascular smooth-muscle cell contraction. Therefore, ROCK inhibition may regulate the calcium sensitivity of smooth-muscle cells in coronary arteries and thus favor coronary vasodilatation. In fact, it has been shown in a pig model that hydroxyfasudil blocks myosin light-chain phosphorylation and significantly inhibits coronary spasm [4]. Hydroxyfasudil has also a cardioprotective effect on coronary I/R injury in canine coronary microcirculation since it preserves endothelium-dependent vasodilator responses after I/R and decreases MI size for which nitric-oxide-mediated mechanisms may be involved [39, 40].

In vivo evidence shows that directed ROCK inhibition by adenovirus-mediated transfer of dominant negative Rho-kinase (DNRhoK) induces a marked regression of coronary constrictive remodeling and abolishes coronary vasospastic activity in 3 weeks in porcine coronaries chronically exposed to IL-1β [41].

A couple of small clinical trials propose that fasudil lessens myocardial ischemia in subjects with stable or vasospastic angina. It appears that this compound considerably increases the ischemic threshold during exercise in angina patients, with a trend toward increased exercise duration [1]. In one study that evaluated the involvement of ROCK in coronary microvascular dysfunction associated with effort angina, intracoronary administration of fasudil increased the oxygen saturation in coronary sinus vein and ameliorated pacing-induced myocardial ischemia in patients with stable effort angina without significant hemodynamic changes. These results are interpreted as suggestive that the ROCK pathway in the coronary microcirculation is not functional under normal conditions but appears to be upregulated under pathological conditions such as stable effort angina. It is not known whether long-term treatment with fasudil improves prognosis of patients with effort angina [42].

***Cerebral Vasospasm and Stroke*** ROCKs may be considered as a therapeutic target for ischemic stroke since fasudil had a positive impact on cerebral blood flow and infarct size in a murine model with cerebral artery occlusion. The neuroprotective effects have been linked to increases in eNOS expression and activity [1, 9].

***Pulmonary Hypertension*** In a rat model of monocrotaline-induced pulmonary hypertension, overall survival has been reported to improve, and the development of severe pulmonary hypertension, right-ventricular hypertrophy, and pulmonary vascular lesions to be decreased after long-term inhibition of Rho-kinase with fasudil administered orally. The beneficial effects appear to be present both when the drug is started concomitantly with monocrotaline and also when started after development of pulmonary hypertension. Apparently, fasudil's related improvement in this model is linked to suppression of VSMC proliferation and macrophage infiltration, enhanced VSMC apoptosis, and amelioration of endothelial dysfunction and VSMC contraction. However, it is important to

**TABLE 4 Cardiovascular Toxicity of Tyrosine Kinase Targeting Drugs Used in Oncology[a]**

| Agent | Molecular Target | Oncologic Target | Cardiovascular Toxicity[b] | Other Toxicity |
|---|---|---|---|---|
| Trastuzumab (Herceptin) | ERBB2 | Breast cancer | Known or likely cardiotoxicity | Infusion reactions, neutropenia |
| Imatinib (Gleevec) | ABL1/2 PDGFRα/β KIT | CML, Ph+ B-ALL, CMML, CEL, GIST | Known or likely cardiotoxicity Edema | Nausea, myelosuppression, immunosuppression |
| Dasatinib (Sprycel) | ABL1/2, PDGFRα/β, KIT, Src family | CML | Known or likely cardiotoxicity Edema, pericardial effusion, QT prolongation, hemorrhage | Myelosuppression, pleural effusion, panniculitis |
| Nilotibib (Tasigna) | ABL1/2, PDGFRα/β, KIT | CML | Known or likely cardiotoxicity QT prolongation | Myelosuppression, hyperbilirubinemia, rash |
| Sunitinib (Sutent) | VEGFR1–3, KIT, PDGFRα/ β, RET, CSF1R, FLT3 | Renal cell carcinoma, GIST | Known or likely cardiotoxicity Hemorrhage, hypertension | Adrenal dysfunction, hypothyroidism |
| Sorafenib (Nexavar) | VEGFR2, PDGFRβ, KIT, FLT3, RAF1, BRAF | Renal cell carcinoma, melanoma | Known or likely cardiotoxicity Acute coronary syndromes Hemorrhage, hypertension | Skin rash |
| Bevacizumab (Avastin) | VEGFR | Colorectal cancer, NSCLC | Known or likely cardiotoxicity Arterial thromboembolism Hemorrhage, hypertension | Gastrointestinal perforation, poor wound healing, neutropenia |
| Lapatinib (Tykerb) | EGFR | Breast cancer | Low cardiotoxicity | Skin rash, diarrhea |
| Gefitinib (Iressa) | EGFR | NSCLC | Low cardiotoxicity | Skin rash, diarrhea, nausea, interstitial lung disease |
| Erlotinib (Tarceva) | EGFR | NSCLC, pancreatic cancer | Low cardiotoxicity | Skin rash, diarrhea, nausea, interstitial lung disease |
| Cetuximab (Erbitux) | EGFR | Colorectal cancer, squamous cell carcinoma of head/neck | Low cardiotoxicity | Skin rash, infusion reactions, interstitial lung disease, hypomagnesemia |
| Panitumumab | EGFR | Colorectal cancer | Low cardiotoxicity | Skin rash |

[a]ERBB2, receptor tyrosine kinase member of the EGFR family; EGFR, epidermal growth factor receptor; ABL, Abelson tyrosine kinase; PDGFR, platelet-derived growth factor receptor; KIT, receptor for stem cell factor; CML, chronic myeloid leukemia; Ph+, Philadelphia chromosome positive; B-ALL, B-cell acute lymphoblastic leukemia; CMML, chronic myelomonocytic leukemia; CEL, chronic eosinophilic leukemia; GIST, gastrointestinal stromal tumor; QT, QT interval on electrocardiogram; VEGFR, vascular endothelial growth factor receptor; RET, rearranged during transfection; CSF1R, colony-stimulating factor 1 receptor;
FLT3, FMS-related tyrosine kinase 3; NSCLC, non-small-cell lung cancer,
[b]Rate of cardiotoxicity is known only for trastuzumab and lapatinib.
*Source*: Adapted from ref. 51.

mention that some animals in the fasudil groups died because of right-ventricular failure, thus suggesting that treatment with this compound was not effective in all animals [43].

In a pulmonary hypertension model induced by hypoxia using wild-type and eNOS-deficient mice, long-term inhibition of Rho-kinase with fasudil improved the hypertension and right-ventricular hypertrophy in wild-type mice and to a lesser extent in eNOS knockout mice. However, the therapeutic effects of fasudil after the development of hypoxia-induced pulmonary hypertension need to be evaluated [44].

Pulmonary vascular resistance in patients with pulmonary hypertension is significantly reduced by intravenous infusion of fasudil; in fact, the intravenous administration of fasudil appears to have an acute pulmonary vasodilator effect. There are current plans to perform clinical trials with long-term oral treatment with fasudil in patients with pulmonary hypertension in Japan [9, 18].

***Heart Failure*** It has been suggested that activation of Rho-kinase is involved in the increased peripheral vascular resistance seen in patients with congestive heart failure (CHF) [9, 45]. Furthermore, fasudil-mediated vasodilation correlates to plasma brain natriuretic peptide (BNP) levels, suggesting an activation of the Rho/Rhokinase pathway directly related to the severity of CHF [45].

***Atherosclerosis*** Selective ROCK2 inhibition with SLx2119 seems to attenuate arterial plaque formation in apolipoprotein-E-deficient mice, indicating that this selective antagonism could limit atherosclerosis without the unwanted hemodynamic side effects seen with nonselective ROCK inhibitors [9]. It has been established that long-term treatment with fasudil blocks macrophage accumulation in the adventitia and decreased migration into the media with subsequent inhibition of coronary lesion formation in a porcine model of coronary arteriosclerosis. In this experimental model, the atherosclerotic lesions where established through monocyte chemoattractant protein-1 and oxidized low-density lipoprotein. Furthermore, this drug suppressed in-stent neointimal formation in porcine coronary arteries through various mechanisms, including reduced vascular inflammation, enhanced apoptosis, and decreased collagen deposition. Thus, long-term ROCK inhibition induces a considerable regression of coronary remodeling and vasospasm in a porcine model with adventitial inflammation [9].

***Metabolic Syndrome*** ROCKs may interfere with insulin signaling by favoring insulin receptor substrate-1 (IRS-1) phosphorylation in cultured VSMC. ROCK activation has been found in skeletal muscle and aortic tissue of Zucker obese rats. In fact, long-term ROCK inhibition appears to decrease blood pressure, correct glucose and lipid metabolism, recover insulin signaling, and improve endothelial dysfunction [1].

### 3.5 Protein Kinase C (PKC) and Protein Kinase A (PKA)

In a clinical trial, the PKCβ inhibitor LY333531 has been shown to be safe and effective in treating diabetes-induced retinal hemodynamic abnormalities [1].

There is evidence that shows cAMP accumulation and PKA activation with increased opening of L-type calcium channels in myocardial ischemia. This is associated with reperfusion dysrhythmias and necrosis (i.e., reperfusion injury). In fact, the use of H89 in isolated rat heart appears to have a cardioprotective effect associated with PKA inhibition, improved contractile recovery, and reduction of infarct size, both when given as pretreatment or preconditioning [10].

Since heart failure is associated with activation of the β-adrenergic signal transduction pathway and the beneficial action of β-adrenergic blockade in this setting, PKA inhibition may be an additional target in the treatment. However, it appears that H89 is used only in preclinical research, and it remains to be established whether inhibition of PKA could have additional advantages above conventional β-adrenergic blockade [10]. PKCβ may be an attractive therapeutic target in patients with heart failure since signaling of many circulating neurohormones (e.g., α-adrenergic system, endothelin, and angiotensin II) is associated with this enzyme. The PKCβ inhibitor ruboxistaurin seems to exert antifibrotic effects after experimental myocardial infarction [21].

Other PKA antagonists have been developed, such as the cAMP analog Rp-cAMPS, whose effect is strongly dependent on kinase concentration (i.e., the higher the kinase concentration, the more efficient), and the protein kinase inhibitor (PKI) peptide, which is an endogenous peptide that modulates PKA function. In fact, Rp-cAMPS and synthetic PKI analogs seem to be more specific inhibitors of PKA than H89 [10].

### 3.6 Calcium/Calmodulin-Dependent Protein Kinase II (CaMKII)

It is known that structural heart disease is associated with electrical instability and arrhythmias, myocardial dysfunction, myocardial hypertrophy, and chamber dilation. Targeted CaMKII inhibition could improve some of these fundamental changes since there is evidence of altered CaMKII expression or activity in structural heart disease and some studies show that introducing an excess or constitutively active mutant leads to the disease in a previously normal tissue, whereas knocking out the protein or inhibiting its activity could be protective [11]. Genetic inhibition of CaMKII confers resistance to development of cardiac hypertrophy associated with excessive and sustained catecholamine stimulation. In contrast, some results suggest that CaMKII inhibition restores myocardial function in calcineurin cardiomyopathy with a minimum effect on reducing hypertrophy. Thus, CaMKII overactivity may have deleterious effects on cardiomyocyte function independent of cardiac hypertrophy. In vivo studies suggest that transgenic mice with CaMKII inhibition are resistant to apoptosis related to ischemia and high-dose isoproterenol. Furthermore, CaMKII inhibition seems to protect from developing apoptosis in phospholamban knockout mice. Therefore, it is thought that CaMKII antagonism can modulate cell death associated with ischemia and catecholamine toxicity [24].

Protein kinases involved in calcium-signaling pathways, such as Rho kinases and PKC, represent a poten-

tial therapeutic target in cardiovascular disease. ROCKs and PKC are upstream modulators of calcium homeostasis, while CaMKII is a downstream target. CaMKII is activated by β-adrenergic stimulation, and CaMKII antagonism appears to be beneficial without affecting the normal sympathetic responsiveness as with beta-blockers. It has been described that CaMKII antagonism in cardiac myocytes using the small peptide inhibitor of CaMKII AC3-I or the inhibitor KN-93 reduces cardiac hypertrophy and contractile dysfunction secondary to MI and chronic β-adrenergic stimulation. Nevertheless, AC3-I transgenic mice seem to have normal basal function and preserved acute β-adrenergic response. Furthermore, genetic inhibition provides evidence that CaMKII activity is needed for normal AV nodal conduction since antagonism slows conduction through this node, which suggests a potential therapeutic use of CaMKII [1].

It is well known that the β-adrenergic system is a prominent regulator of cardiac function, and blocking it is nowadays the major therapy for treating patients with structural heart disease and myocardial dysfunction. The traditional β-adrenergic signaling pathway involves activation of PKA, which targets many of the same proteins in the calcium cycle as CaMKII, leading to increases in calcium mobilization. In fact, there is evidence that CaMKII is an important downstream effector of the β-adrenergic receptor signaling cascade. Transgenic overexpression of a CaMKII inhibitory peptide (AC3I) ameliorates the deleterious impact of MI on left-ventricular function since the CaMKII block seems to prevent maladaptive remodeling associated with excessive β-adrenergic stimulation and induces balanced changes in excitation–contraction coupling needed for cardiac function [11, 46]. In addition, chronic inhibition of CaMKII in AC3I mice and acute inhibition with KN93 protects from cardiomyopathic responses seen with chronic β-adrenergic stimulation. It has been shown that CaMKII activity directly contributes to the loss of calcium homeostasis in two cardiac disease models associated with β-adrenergic activation [11]. There is data showing that CaMKII antagonism protects against apoptosis in vivo after myocardial infarction or excessive stimulation with isoproterenol. In fact, transgenic mice that express CaMK inhibitory peptide present a cardioprotective effect after MI and chronic β-adrenergic stimulation [11]. It appears that CaMKII is an attractive drug target in structural heart disease and arrhythmias. However, this multifunctional kinase has many known roles in cell functioning (e.g., gene regulation and molecular memory), and its complete block could be associated with unwanted adverse effects. Thus, it could be attractive to selectively target CaMKII substrates that are critical for calcium cycling [11]. The current knowledge of the structure of the regulatory and catalytic domains of CaMKII could facilitate the development of specific inhibitory agents. For example, inhibitors of CaMKII, such as KN62 and KN93, bind to the calcium/CaM domain and block CaMKII activation in cardiac myocytes. However, the results obtained with these drugs need to be interpreted with care because these drugs have effects on multiple ion channels, and more selective CaMKII inhibitors are needed [11].

### 3.7 Mammalian Target of Rapamycin (mTOR)

Rapamycin inhibits proliferation and migration of rodent and human VSMCs in vitro and porcine VSMCs in vivo by blocking cell cycle progression at the G1/S transition. In fact, a rapamycin-induced effect on VSMC appears to be related with a marked reduction in cyclin-dependent kinase (CDK) activity and retinoblastoma protein phosphorylation [47]. The use of stents covered with rapamycin/sirolimus, and thus local mTOR inhibition, prevents growth and proliferation of VSMC and coronary restenosis [1, 4]. The implantation of rapamycin-coated stents in lesions appears to be safe and effective in inhibiting neointimal formation. However, rapamycin seems to increase tissue factor expression associated with potential subacute and late stent thrombosis. Nevertheless, other studies suggest that regardless of enhanced tissue factor expression secondary to mTOR inhibition, there is no increased tissue factor activity.

It has been suggested that rapamycin has a preconditioning-like effect against myocardial infarction. In fact, rapamycin and mTOR blockage appears to have a protective effect in I/R injury since pretreatment of murine hearts reduced infarct size and decreased cardiac myocyte cell death [1].

### 3.8 Glycogen Synthase Kinase-3β (GSK-3β)

Even when GSK was identified in the 1970s as a kinase that blocks glycogen synthesis, it was not considered a potential drug target until the 1990s, when its function in insulin signal transduction was elucidated and its probable value in Alzheimer's disease emerged. The use of several potent GSK3 inhibitors in cell-based assays and animal models of type II diabetes suggest they are effective in lowering blood glucose levels, and thus they may have therapeutic potential for the management of diabetes [12]. It has been proposed that GSK3 inhibitors might be advantageous for treating diabetes because (1) insulin favors activation of glycogen synthase by restraining GSK3 activity, (2) there is elevated GSK3 expression in diabetic muscle, and (3) GSK3 appears to

attenuate insulin signaling by phosphorylating and downregulating IRS-1 in overexpression studies. Consequently, GSK3 antagonists could mimic insulin's ability to favor glucose conversion to glycogen, and thus overcome the insulin resistance associated with inhibition of the pathway at the insulin receptor/IRS level. In fact, the use of lithium, a relatively specifically in vitro GSK3 inhibitor, appears to activate glycogen synthase and to stimulate glycogen synthesis in muscle. Furthermore, in the late 1990s, a class of maleimides that blocked GSK3 more potently than lithium was identified, and these compounds (SB 216763 and SB 415286) were shown to activate glycogen synthase and speed up glucose conversion to glycogen in human liver cells. In addition, there is data reporting that other compounds in this class normalize blood glucose levels in vivo. On the other hand, these new compounds and lithium appear to share insulin's ability to suppress phosphoenolpyruvate carboxykinase and glucose-6-phosphatase transcription, the key rate-limiting enzymes of gluconeogenesis. These observations suggest that GSK3 antagonists might suppress glucose production by the liver and enhance its conversion into glycogen. However, SB 216763 and SB 415286 inhibit other protein kinases (e.g., cyclin-dependent protein kinases or CDKs) and lithium affects other distinct protein kinases and also blocks nonkinase targets (e.g., inositol monophosphatase and histone deacetylase). Thus, it is important to show that structurally diverse GSK3 inhibitors act similarly, before concluding that these effects are actually mediated via the inhibition of GSK3. Other compounds are being studied. For example, azakenpaullone seems to have greater selectivity for GSK3 than CDK1 and CDK5, and thienyl and phenyl α-halomethyl ketones are said to inhibit GSK3 without inhibiting PKA; but further information is needed regarding its selectivity against a large panel of other protein kinases. Although it has been established that a concentration in the micromolar range is needed for inhibition in vitro, these compounds and lithium are noncompetitive inhibitors, while SB 216763 and SB 415286 are ATP competitive. There are advances in the development of more potent and specific GSK3 antagonists that do not affect CDKs, such as CHIR 99021, AR-A014418, and bis-7-azaindolylmaleimides. A couple of studies report that CHIR 99021 and closely related compounds (CHIR 98023 and CHIR 98014) activate glycogen synthase and stimulate glycogen deposition in the liver. Furthermore, these drugs appear to be effective in lowering glucose and insulin concentrations in oral glucose tolerance tests in Zucker Diabetic Fatty (ZDF) rats, which reflects improved sensitivity to insulin. In studies lasting up to 24 h, CHIR 99021 seems to lower the fasting blood glucose level, probably associated to inhibition of gluconeogenesis. In fact, these compounds have been reported to be effective in three different rodent models of type II diabetes. In order to determine the real therapeutic value of GSK inhibitors in diabetes, even when these compounds appear to effectively lower blood glucose in rodent models of type II diabetes, there is still a need to extend these studies to non-leptin-deficient models of type II diabetes and rodent models of type I diabetes, and further studies are required to evaluate these drugs in humans [12]. It has been noted that GSK3 inhibitors might mimic the Wnt signaling pathway and therefore be potentially oncogenic. In fact, there is evidence that shows SB 216763, SB 415286, and lithium increase β-catenin levels in transformed cell lines. Furthermore, since GSK3 phosphorylates the transcription factors c-jun and c-myc, the inhibition of GSK3 could promote dephosphorylation of proto-oncogenes and thus lead to their activation. However, lithium has been used to treat bipolar disorder for many years, and long-term use does not seem to be associated with increased risk of developing cancer. Furthermore, infusion of CHIR99021 for up to 20 h in ZDF rats was not associated with an increase of β-catenin or cyclin D1 in brain, liver, lung, colon, or adipose tissue. Therefore, it is thought that GSK3 inhibition by itself might not be enough to elevate the levels of β-catenin in normal tissues, as it might only occur in transformed cell lines. Regardless of these assumptions, long-term treatment of cells and animals with GSK3 antagonists are required to fully evaluate this, since these compounds would need to be administered for many years to treat chronic disease states such as diabetes [12].

### 3.9 5′-AMP-Activated Protein Kinase (AMPK)

Another approach to target protein kinase activity is to identify compounds that activate rather than inhibit the enzyme, especially for protein kinases that are beneficially activated by allosteric mechanisms [4]. AMPK activation elevates fatty acid oxidation because it inhibits ACC, the rate-limiting enzyme in fatty acid synthesis that inhibits the catalytic process. In addition, AMPK activation seems to favor lower levels of hepatic sterol response element binding protein 1 and consequently decreases lipogenic gene expression; therefore, serum triglycerides could be lowered by therapeutic activators of AMPK. It has also been shown that AMPK inhibits HMG-CoA reductase, the rate-limiting enzyme in cholesterol biosynthesis, thus blocking cholesterol production and making therapeutic AMPK activators attractive pharmaceutical targets. Furthermore, it seems that AMPK is activated during exercise and is linked to favoring glucose uptake by skeletal muscle in an insulin-independent manner. In fact, there is data showing that AMPK activation with 5-aminoimidazole-4-carboxamide-1-β-D-ribofuranoside (AICAR) triggers

a phenomenon [4]. Moreover, the antidiabetic drug metformin appears to exert its effects in part by activating AMPK [4, 48]. Consequently, activators of AMPK may improve glucose intolerance. In fact in animal models AMPK activation is needed to achieve the inhibitory effect exerted by metformin on glucose production by hepatocytes, and stimulation of glucose uptake in skeletal muscle [49]. Alternative ways of modulating AMPK could be useful for the treatment of metabolic disorders [4, 49].

There is evidence that suggests polyphenols, such as resveratrol (found in red wine), apigenin, and S17834 increase AMPK and its downstream target ACC in hepatocytes, and they appear to be more potent than metformin. These compounds and genetic overexpression of AMPK also seem to prevent lipid accumulation related to high glucose exposure [25].

AMPK stimulation in type I diabetic low-density lipoprotein (LDL) receptor-deficient mice (DMLDLR–/–) with S17834 has been described to revert the decrease in AMPK and ACC activation and to prevent lipid accumulation in the liver associated with hyperlipidemia and aortic lesion development [25].

There has been speculation on the multiple advantages and disadvantages of pharmacological AMPK manipulation for treatment of cardiovascular disease. However, some experts say that currently there is no sufficient understanding of the role of AMPK in cardiac health and disease to direct pharmacological treatment for patients. It is recognized that AMPK may eventually be considered as a major regulator of muscle and heart glycogen metabolism, but the current knowledge is based on studies with constitutive mutations, which may not reflect normal cardiac physiology, and further research is required to understand AMPK's contribution to normal glycogen metabolism [14].

### 3.10 Other Protein Kinases

There is evidence that links mutations of WNK (with no lysine family of kinases) with pseudohypoaldosteronism type II, a rare hereditary form of hypertension. In fact, less severe mutations and/or polymorphisms of WNK may play a role in hypertension since there is evidence in the Framingham Heart Study that the *WNK4* gene lies close to a locus showing the strongest linkage to blood pressure variation. Thus WNK could be a potential target in hypertensive patients [4].

The p90 ribosomal S6 kinase (RSK) has been reported to regulate cardiac sodium/hydrogen exchanger isoform 1 (NHE1) activity and thus may be involved in the mediation of myocardial ischemia and I/R injury in the heart. In fact, cardiac-specific overexpression of dominant negative RSK in transgenic mice (DN-RSK-TG) appears to be associated with normal basal cardiac function and morphology, and DN-RSK-TG hearts seem to have reduced myocardial infarction after temporary left coronary artery occlusion. Furthermore, myocyte apoptosis is also shown to be reduced after I/R in DN-RSK compared with nontransgenic littermate controls. There is data that shows these effects to be related to inhibition of NHE1 activity by RSK blockade during I/R. However, large clinical trials to evaluate the NHE1 inhibitors cariporide and eniporide (GUARDIAN trial [Guard during ischemia against necrosis trial] and ESCAMI trial [evaluation of the safety and cardioprotective effects of eniporide in acute myocardial infarction]) in patients with myocardial infarction, show no reduction in mortality, which may be related to the fact that these antagonists block homeostatic functions of NHE1. Thus, targeted inhibition of RSK and reduction of NHE1 activity may be useful to treat cardiac I/R injury [50].

There are some studies that suggest that imatinib inhibits receptor tyrosine kinase signaling and, therefore, might be effective for the treatment of heart failure because it inhibits TGF-β and connective tissue growth factor signaling, which are activated in the ventricular remodeling process. However, it has been shown that imatinib may cause cardiomyopathy through inhibition of c-ABL [21].

## 4 OBSTACLES AND OPPORTUNITIES IN THE CLINICAL DEVELOPMENT OF TARGETED THERAPEUTICS

A critical concern in drug development is the identification of appropriate targets for therapeutic intervention. Since deregulation of a specific protein kinase can be an irrelevant consequence of the disease rather than a key contributing factor to the pathology, when this kinase is being considered as a potential target, it is not enough justification to know whether it is activated or inhibited in a specific disease state. Further evidence, such as clear genetic or physiological data is required that implicates a protein kinase as an attractive target [4]. Once a kinase is validated as a target for drug development, it is crucial to identify a series of promising compounds to be tested. In order to make the identification process efficient, high-throughput screening is needed. Based on the actual crystal structure of the kinase (structure-based library design) or on the structure of compounds already known to bind to the target (ligand-based library design), the virtual screening or molecular modeling turns out to be more proficient [4]. Computer-aided drug design can be an extremely valuable guide to choose the most promising compound since the structure–activity relationship of any candidate drug can be correlated with specific molecular interactions of

the compound with the kinase active site, and in this way the structure of the inhibitor can be optimized. When structure data do not exist for a specific kinase, knowledge of the structure of another member of the family can often be used to create binding models from which optimized compounds can be synthesized. However, it is possible that a compound with apparently high activity and specificity in vitro may display markedly different and even nonspecific activity in vivo [4]. Therefore, it is critical to test the promising drug candidates for safety and efficacy in relevant experimental animal models of cardiovascular disease and, furthermore, test the compounds in well designed clinical trials that will facilitate evidence-based decisions. The ultimate validation can only be reached when a compound shows efficacy and safety in advanced human clinical trials, and the new inhibitors need to have adequate potency, selectivity, and pharmacokinetic profile (absorption, distribution, metabolism, elimination, and toxicity) [5].

### 4.1 Biological Barriers for Drug Delivery

Crystallographic data reveals that kinases have a sufficient degree of uniqueness, which allows developing specific inhibitors. However, antagonism of kinase activity in a diseased tissue could also inhibit desired activity of the same kinase in other tissues, which could be overcome by localized drug delivery [1]. The specific function of a kinase inhibitor needs to be evaluated with caution since there might be nonspecific actions of some compounds at certain concentrations. The drug's cellular uptake, its half-life, diffusion rate, and/or multiple intracellular receptors are also important determinants of the final effect seen in experimental disease models and the clinical setting [3].

The ATP binding site has been more aggressively targeted for drug development than the protein substrate binding site, the interference with anchoring proteins to which kinases bind for proper cellular localization, or with the signaling proteins they recruit to mediate some of their effects. Directing development to the ATP binding site presents certain challenges since all protein kinases bind ATP and have a homologous protein fold, where ATP binds between the small and large lobe of the kinase catalytic domain. Distinct features at these sites have allowed the development of more selective ATP-based inhibitors [2]. However, other intracellular enzymes and proteins bind ATP or other nucleotides and may generate the problem of specificity. Furthermore, a successful antagonist drug needs to have a binding affinity many times higher than ATP since the competitive inhibition of ATP binding happens against millimolar levels of intracellular ATP. The ATP binding pocket of active kinases is homologous while its conformation is distorted and, therefore, more diverse in the autoinhibited state of the kinase. Since some kinases require phosphorylation by other kinases to become active, inhibitors that bind the inactive form could block phosphorylation and activation of the kinase by its protein kinase [2].

### 4.2 Protein Kinase-Modulating Compounds—Assessing and Managing Potential Toxicity

Toxicity remains a major concern because many of these kinases not only play roles in the pathogenesis of disease but also function in pathways that regulate basic cellular processes [4]. Clinical oncological trials have shed some light regarding toxic side effects, but it is important to remember that these studies are designed to identify or use the maximum tolerated dose, which may be considerably higher than the doses needed to treat cardiovascular diseases. Furthermore, the use of combination therapy to target two or more kinases on the same or different pathways may allow the administration of lower and less toxic doses [4].

The first report of cardiac toxicity of a targeted agent was for trastuzumab, cases of CHF have been described after treatment with imatinib, and undesirable cardiac effects are referred to in the prescribing information for dasatinib, sunitinib, sorafenib, and bevacizumab. However, cardiotoxicity seems not to be a tyrosine kinase inhibitor class effect because it appears to be uncommon with other compounds, such as those that target the epidermal growth factor receptor (ERBB1). As a result, toxicity for each agent needs to be determined on a case-by-case basis [51]. Recent evaluation of cardiotoxicity [i.e., drug-related direct injury to cardiac myocytes resulting in compromised left-ventricular (LV) function] in oncology has been focused on drugs that target tyrosine kinases. It has become evident that some of these agents can cause symptomatic CHF in some patients and asymptomatic LV dysfunction in others. Thus, tyrosine-kinase-targeted therapies can either have known or likely toxicity to the heart or present a very low cardiotoxicity risk. However, exact rates of cardiotoxicity associated with tyrosine kinase inhibitors are not known since most clinical trials have not looked at predefined cardiac endpoints (i.e., LV function before and during treatment has not been measured prospectively). In addition, diagnosis of cardiotoxicity and CHF has been based mainly on medical history and physical examination, which can be considered unreliable in patients with mild to moderate grades of CHF or with asymptomatic LV dysfunction. Furthermore, it might be difficult to clinically diagnose CHF in oncologic patients who have numerous reasons other than LV dysfunction to develop shortness of

breath, fatigue, and edema. Although some trials have assessed LV function, the way the results are reported makes it complicated to place the findings in a clinically meaningful context. Moreover, some clinical trials that determine rates of CHF before the approval of a drug often exclude patients with certain comorbidities (i.e., cardiovascular disease), thus underestimating cardiac toxicity, which might be seen after approval when patients with multiple risk factors and/or cardiovascular disease are likely to be treated. Finally, most clinical trials in this setting are typically of short duration (i.e., months), whereas treatment of specific conditions with these agents might be lifelong. Currently, some kinase targets in cancer are not expressed in cardiomyocytes and, therefore, have no direct role in their survival or function [51], thus making it hard to translate successful results in oncology to potential benefits when used in cardiovascular pathology.

Existing generation of tyrosine kinase inhibitors is nonselective and the focused design of multitargeted antagonists might allow a single agent to be more effective in more types of cancer, but with an increased risk of cardiotoxicity related to inhibition of targets that are not crucial for the killing of cancer cells but that are involved in cardiomyocyte survival and function [51].

### 4.3 Promiscuous Drugs Compared to Selective Drugs

From growing experience with protein kinase inhibitors, it has been established that detailed knowledge of the drug target predicts clinical success in cancer treatment [52]. The ideal of achieving absolute specificity has been a constant aim in drug development, and the switch from mechanism to target-oriented drug discovery associated with molecular genetics and genomics faces this ideal in a strict molecular context. It is thought that a compound should preferably interact with the selected target, trigger the intended response, disengage, be metabolized, and excreted without any further effects. However, it has proven difficult to find drugs that interact specifically with one target molecule or several desired targets. [53]. There is agreement that potency and selectivity are key factors for the effectiveness and safety of any drug [4]. When designing kinase inhibitors, selectivity is a critical consideration. In fact, compounds selectivity is screened against panels of other kinases (often over 30 kinases at the same time) to determine whether there are other targets affected aside from the intended one. These panels often include specific kinases to screen for directed inhibition and/or a broad selection of kinases with structural diversity at the active site to screen for nonspecific inhibition. Once selectivity is determined in vitro, it is then determined in cellular and whole-animal systems since sometimes there are paradoxical findings [4]. The specificity of protein kinase signaling is improved by its subcellular localization since many interact with adapter proteins to direct their activity (e.g., receptors for activated C kinase for PKC and A kinase anchoring proteins for PKA). Since interaction of kinases with specific proteins is crucial for normal function, the targeted disruption of this by peptides or small-molecule inhibitors is a possible strategy for decreasing protein kinase activity. Furthermore, signaling networks can be cell-type specific, increasing the selectivity of the therapeutic agent [2]. In view of the fact that the acceptable selectivity threshold for a given inhibitor may be different for a particular disease, it may not be evident at the outset of a drug discovery program but rather after extensive experimental pharmacology and toxicology studies. Compounds without an optimal selectivity profile in vitro should in fact be evaluated under certain circumstances to understand the complex relationships between chemical structure, isoform selectivity, efficacy, and toxicity [5].

Since ATP binding is essential for the kinase's activity, the ATP binding pocket is the target of most inhibitor screens. It appears that there is enough structural diversity at these sites to predict whether selective inhibitors can be developed [4]. The binding of the drug to the substrate interaction site offers another route to increase a compound's selectivity because these sites are more diverse than the ATP binding site. This, however, may involve protein–protein interactions driven by multiple low-affinity interactions, which may be difficult to interrupt with a small molecule. Furthermore, certain peptides have been shown to have high affinity and specificity but are not orally bioavailable and cannot penetrate cell membranes. On the other hand, higher specificity might be reached by using a bisubstrate inhibitor that binds elements of both the protein substrate and ATP binding sites. Another approach is to interrupt kinase localization or its response to second messengers, and inhibition of kinase anchoring might antagonize only a subset of kinase function [2].

In order to determine the mechanisms underlying the cardiac toxicity of some tyrosine kinase inhibitors, the target and the signaling pathways mediating the adverse cardiac effects need to be identified. None of the small-molecule tyrosine kinase inhibitors are completely selective, and some (e.g., sunitinib) even seem to be rather nonselective. In some cases, kinases with no previously known role at cardiac level have been identified as key targets of these compounds (e.g., ABL with imatinib) and once a kinase is identified as critical for normal function in the heart, all agents that target that

kinase could be considered potentially toxic (i.e., as "likely" to have cardiotoxicity) [51].

Since therapeutic compounds typically present intracellular concentrations in the nanomolar range and ATP reaches levels in the millimolar range, a clinical effective protein kinase inhibitor needs to have high affinity for the target kinase several orders of magnitude higher than that of ATP. Actually, since there is this need for extremely high binding affinity to the ATP pocket and there are relative similarities of ATP pockets across protein kinase families, it may be advantageous to find other determinants in protein kinases that may confer additional specificity [4].

An important lesson was learned when selective PI3Kγ inhibitors were developed. A crucial concern was its selectivity and the effect on other PI3K isoforms. Based on viability data obtained from mouse models with genetically manipulated PI3K isoforms, it seems that small-molecule PI3K inhibitors specific for PI3Kγ should have no impact on the vital class IA PI3Ks (i.e., PI3Kα and PI3Kβ) responsible for metabolic and other "house-keeping" functions (e.g., cell growth, maintenance, and survival). Achieving this degree of selectivity would minimize unwanted side effects, but it might also decrease the therapeutic efficacy of the drug, especially in clinical settings where simultaneous modulation of different signaling pathways in addition to PI3Kγ-dependent ones might be desirable [5].

An important problem that can be encountered when considering a specific intracellular signaling pathway for therapeutic purposes is not due to drug or tissue selectivity but simply because each effector can have both adaptive and maladaptive influences on the myocardium. Therefore, a future challenge will be to translate current knowledge of the signaling mechanisms into novel pharmacological or gene therapeutic treatments for human disease such as cardiac hypertrophy, heart failure, and ACS, among others [22].

### 4.4 From Animal Models to Clinical Trials

There is a need to define the specific individual molecular phenotype and/or kinase activity profile for each patient because the clinical diagnosis in cardiovascular medicine includes a diverse spectrum of molecular profiles, and failing to do so may discard agents that could be effective in specific subsets of patients. Interestingly, the majority of early trials seem to focus on individual kinases and their role in diseases that need only short-term therapy (e.g., ACS or stroke) or where targeted local delivery is achievable. However, it is important to recognize that these may not necessarily be the disease states most likely to benefit from targeted protein kinase therapy [4].

Protein kinase signaling in the cardiovascular system is mainly studied in cultured cells (e.g., neonatal cardiomyocytes), and the results may not reflect what happens in fully differentiated adult cells or in vivo. However, it has been established that studies performed in cultured cells are valuable because these cells naturally grow and divide in a stable manner, they form intercellular connections, and are readily available. Then again, there are noteworthy differences between cultured cells and fully differentiated adult cells, such as expression of protein kinases, their subcellular localization, their protein substrates, and the interaction between GPCRs and protein kinases that rule the responses to their activation/inhibition. In fact, experimental models that combine genetic and pharmacologic manipulation of protein kinase activity are attractive to determine in vivo the role of the different protein kinases in disease mechanisms. Many of the early in vivo validation studies are commonly performed in murine models, and further validation of the findings in larger animal models is important since there are differences in heart rate, metabolism, cellular calcium handling, and action potential durations between mice and humans [2].

In the complex process of drug development, there are numerous steps that lie between target selection and the initiation of phase II or proof-of-concept human clinical trials. The screening and hit identification of an agent is followed by several rounds of compound optimization focused on pharmacological and toxicological testing and, eventually, pharmaceutical development and production. Drug development requires the correct expertise, appropriate time, and resources to move from the initial concept to the every day practice. The crucial questions concern target validation (to define whether modulating the target activity will alter the course of human disease without causing any significant adverse effects) and practical development (to define whether compounds that will modulate the target's activity in the living organism with appropriate potency and specificity can be found) [5]. There are some reservations to the use of genetically modified animals in terms of their relevance to human disease. There are certain concerns regarding compensation during early development, intracellular changes (absence/excess of protein compared to absent/increased catalytic activity, complete and permanent versus inducible and dose-dependent loss/gain of activity) and inherent differences of animal disease models compared to human disease [5].

The effort to identify new pharmacological therapies for cardiovascular disease continues; however, for certain pathologies (e.g., heart failure) it appears that the pace is slowing considerably. This may be associated with the high cost of drug development, the perceived risk–return ratio, and the complexity of conducting nec-

essary clinical trials to obtain registration. For example, it has been estimated that the cost of a large-scale heart failure trial ranges between $102 and 207 million, and recent negative or adverse trials (e.g., OVERTURE with omapatrilat and MOXCON with moxonidine) may trigger certain reservations. This may be combined with the fact that drug development may be affected or limited by the need of highly selective agonists or antagonists that target cardiac isoforms exclusively or, instead, will depend on the concomitant use of specific targeted therapeutic delivery to the heart [21].

## 5 CONCLUSIONS AND FUTURE PERSPECTIVES

Agents that target protein kinases have been used successfully in oncology, and several have been tested in cardiology with mixed results, suggesting that in the future selective protein kinase inhibitors may be used to treat cardiovascular disease [4].

The intracellular signal transduction network is highly complex and redundant; hence, the challenge of developing a successful targeted therapy is not small. Nevertheless, the signaling pathways associated with disease are being gradually unraveled, encouraging the development of safe and effective drugs that specifically and selectively target protein kinases [5]. Refinement of the selectivity of protein kinases inhibitors may decrease its toxicity while producing more biological effects compared with drugs that inhibit upstream activation [2].

Dormant problems of modulating protein kinases are not small; its use may dysregulate a completely different organ (e.g., inducing apoptosis could reduce ischemia-induced cell death in cardiomyocytes but might favor tumor promotion in other organs). Notwithstanding, protein kinases are unraveling a new and exciting area of knowledge in cardiovascular disease. Developing new agents that selectively and safely inhibit protein kinases may provide a new therapeutic line of agents that may be used in ischemic heart disease, congestive heart failure, pulmonary and systemic hypertension, and arrhythmias.

## REFERENCES


1. Kumar R, Singh VP, Baker KM (2007). Kinase inhibitors for cardiovascular disease. *J Mol Cell Cardiol 42*(1), 1–11.
2. Anderson ME, Higgins LS, Schulman H (2006). Disease mechanisms and emerging therapies: Protein kinases and their inhibitors in myocardial disease. *Nat Clin Pract Cardiovasc Med 3*(8), 437–445.
3. Sridhar R, Hanson-Painton O, Cooper DR (2000). Protein kinases as therapeutic targets. *Pharm Res 17*(11), 1345–1353.
4. Force T, et al. (2004). Inhibitors of protein kinase signaling pathways: Emerging therapies for cardiovascular disease. *Circulation 109*(10), 1196–1205.
5. Ruckle T, Schwarz MK, Rommel C (2006). PI3Kgamma inhibition: Towards an "aspirin of the 21st century"? *Nat Rev Drug Discov 5*(11), 903–918.
6. Cohen P (2002). Protein kinases—The major drug targets of the twenty-first century? *Nat Rev Drug Discov 1*(4), 309–315.
7. Shirai H, Autieri M, Eguchi S (2007). Small GTP-binding proteins and mitogen-activated protein kinases as promising therapeutic targets of vascular remodeling. *Curr Opin Nephrol Hypertens 16*(2), 111–115.
8. Lee MR, Dominguez C (2005). MAP kinase p38 inhibitors: Clinical results and an intimate look at their interactions with p38alpha protein. *Curr Med Chem 12*(25), 2979–2994.
9. Shimokawa H, Rashid M (2007). Development of Rho-kinase inhibitors for cardiovascular medicine. *Trends Pharmacol Sci 28*(6), 296–302.
10. Lochner A, Moolman JA (2006). The many faces of H89: A review. *Cardiovasc Drug Rev 24*(3–4), 261–274.
11. Grueter CE, Colbran RJ, Anderson ME (2007). CaMKII, an emerging molecular driver for calcium homeostasis, arrhythmias, and cardiac dysfunction. *J Mol Med 85*(1), 5–14.
12. Cohen P, Goedert M (2004). GSK3 inhibitors: Development and therapeutic potential. *Nat Rev Drug Discov 3*(6), 479–487.
13. Dolinsky VW, Dyck JR (2006). Role of AMP-activated protein kinase in healthy and diseased hearts. *Am J Physiol Heart Circ Physiol 291*(6), H2557–2569.
14. Arad M, Seidman CE, Seidman JG (2007). AMP-activated protein kinase in the heart: Role during health and disease. *Circ Res 100*(4), 474–488.
15. Kahle KT, et al. (2004). WNK kinases: Molecular regulators of integrated epithelial ion transport. *Curr Opin Nephrol Hypertens 13*(5), 557–562.
16. Kahle KT, et al. (2004). Paracellular Cl-permeability is regulated by WNK4 kinase: Insight into normal physiology and hypertension. *Proc Natl Acad Sci USA 101*(41), 14877–14882.
17. Wilson FH, et al. (2001). Human hypertension caused by mutations in WNK kinases. *Science 293*(5532), 1107–1112.
18. Fukumoto Y, Tawara S, Shimokawa H (2007). Recent progress in the treatment of pulmonary arterial hypertension: Expectation for rho-kinase inhibitors. *Tohoku J Exp Med 211*(4), 309–320.
19. Abe J, Baines CP, Berk BC (2000). Role of mitogen-activated protein kinases in ischemia and reperfusion injury: The good and the bad. *Circ Res 86*(6), 607–609.

20. Kuliopulos A, Mohanlal R, Covic L (2004). Effect of selective inhibition of the p38 MAP kinase pathway on platelet aggregation. *Thromb Haemost 92*(6), 1387–1393.
21. Kaye DM, Krum H (2007). Drug discovery for heart failure: A new era or the end of the pipeline? *Nat Rev Drug Discov 6*(2), 127–139.
22. Heineke J, Molkentin JD (2006). Regulation of cardiac hypertrophy by intracellular signalling pathways. *Nat Rev Mol Cell Biol 7*(8), 589–600.
23. Pandya N, Santani D, Jain S (2005). Role of mitogen-activated protein (MAP) kinases in cardiovascular diseases. *Cardiovasc Drug Rev 23*(3), 247–254.
24. Anderson ME (2007). Multiple downstream proarrhythmic targets for calmodulin kinase II: Moving beyond an ion channel-centric focus. *Cardiovasc Res 73*(4), 657–666.
25. Zang M, et al. (2006). Polyphenols stimulate AMP-activated protein kinase, lower lipids, and inhibit accelerated atherosclerosis in diabetic LDL receptor-deficient mice. *Diabetes 55*(8), 2180–2191.
26. da Silva Xavier G, et al. (2003). Role for AMP-activated protein kinase in glucose-stimulated insulin secretion and preproinsulin gene expression. *Biochem J 371*(Pt 3), 761–774.
27. Wang MY, Unger RH (2005). Role of PP2C in cardiac lipid accumulation in obese rodents and its prevention by troglitazone. *Am J Physiol Endocrinol Metab 288*(1), E216–221.
28. Chen MB, et al. (2005). Impaired activation of AMP-kinase and fatty acid oxidation by globular adiponectin in cultured human skeletal muscle of obese type 2 diabetics. *J Clin Endocrinol Metab 90*(6), 3665–3672.
29. Xia Z, et al. (2007). *N*-Acetylcysteine attenuates PKCbeta2 overexpression and myocardial hypertrophy in streptozotocin-induced diabetic rats. *Cardiovasc Res 73*(4), 770–782.
30. Anderson ME (2002). Calmodulin and the philosopher's stone: Changing $Ca^{2+}$ into arrhythmias. *J Cardiovasc Electrophysiol 13*(2), 195–197.
31. Zhang J, Shen B, Lin A (2007). Novel strategies for inhibition of the p38 MAPK pathway. *Trends Pharmacol Sci 28*(6), 286–295.
32. Li Z, et al. (2006). Selective inhibition of p38alpha MAPK improves cardiac function and reduces myocardial apoptosis in rat model of myocardial injury. *Am J Physiol Heart Circ Physiol 291*(4), H1972–1977.
33. Lepicier P, et al. (2006). Signaling pathways involved in the cardioprotective effects of cannabinoids. *J Pharmacol Sci 102*(2), 155–166.
34. Alessandrini A, et al. (1999). MEK1 protein kinase inhibition protects against damage resulting from focal cerebral ischemia. *Proc Natl Acad Sci USA 96*(22), 12866–12869.
35. Barone FC, et al. (2001). Inhibition of p38 mitogen-activated protein kinase provides neuroprotection in cerebral focal ischemia. *Med Res Rev 21*(2), 129–145.
36. Vanhaesebroeck B, et al. (2001). Synthesis and function of 3-phosphorylated inositol lipids. *Annu Rev Biochem 70*, 535–602.
37. Ferby I, et al. (1996). PAF-induced MAPK activation is inhibited by wortmannin in neutrophils and macrophages. *Adv Exp Med Biol 416*, 321–326.
38. Masumoto A, et al. (2001). Possible involvement of Rho-kinase in the pathogenesis of hypertension in humans. *Hypertension 38*(6), 1307–1310.
39. Yada T, et al. (2005). Beneficial effect of hydroxyfasudil, a specific Rho-kinase inhibitor, on ischemia/reperfusion injury in canine coronary microcirculation in vivo. *J Am Coll Cardiol 45*(4), 599–607.
40. Yada T, Shimokawa H, Kajiya F (2006). Cardioprotective effect of hydroxyfasudil as a specific Rho-kinase inhibitor, on ischemia-reperfusion injury in canine coronary microvessels in vivo. *Clin Hemorheol Microcirc 34*(1–2), 177–183.
41. Morishige K, et al. (2001). Adenovirus-mediated transfer of dominant-negative rho-kinase induces a regression of coronary arteriosclerosis in pigs in vivo. *Arterioscler Thromb Vasc Biol 21*(4), 548–554.
42. Fukumoto Y, et al. (2007). Anti-ischemic effects of fasudil, a specific Rho-kinase inhibitor, in patients with stable effort angina. *J Cardiovasc Pharmacol 49*(3), 117–121.
43. Abe K, et al. (2004). Long-term treatment with a Rho-kinase inhibitor improves monocrotaline-induced fatal pulmonary hypertension in rats. *Circ Res 94*(3), 385–393.
44. Abe K, et al. (2006). Long-term inhibition of Rho-kinase ameliorates hypoxia-induced pulmonary hypertension in mice. *J Cardiovasc Pharmacol 48*(6), 280–285.
45. Kishi T, et al. (2005). Rho-kinase inhibitor improves increased vascular resistance and impaired vasodilation of the forearm in patients with heart failure. *Circulation 111*(21), 2741–2747.
46. Zhang R, et al. (2005). Calmodulin kinase II inhibition protects against structural heart disease. *Nat Med 11*(4), 409–417.
47. Marx SO, Marks AR (2001). Bench to bedside: The development of rapamycin and its application to stent restenosis. *Circulation*, *104*(8), 852–855.
48. Zang M, et al. (2004). AMP-activated protein kinase is required for the lipid-lowering effect of metformin in insulin-resistant human HepG2 cells. *J Biol Chem 279*(46), 47898–47905.
49. Zhou G, et al. (2001). Role of AMP-activated protein kinase in mechanism of metformin action. *J Clin Invest 108*(8), 1167–1174.
50. Maekawa N, et al. (2006). Inhibiting p90 ribosomal S6 kinase prevents $(Na^{+})$-$H^{+}$ exchanger-mediated cardiac ischemia-reperfusion injury. *Circulation 113*(21), 2516–2523.
51. Force T, Krause DS, Van Etten RA (2007). Molecular mechanisms of cardiotoxicity of tyrosine kinase inhibition. *Nat Rev Cancer 7*(5), 332–344.
52. Sawyers CL (2006). Will kinase inhibitors have a dark side? *N Engl J Med 355*(3), 313–315.
53. Drews J (2006). Case histories, magic bullets and the state of drug discovery. *Nat Rev Drug Discov 5*(8), 635–640.

# STATINS IN THE REDUCTION OF CARDIOVASCULAR EVENTS

Rajesh Sachdeva, Bradley Hughes, and Jawahar L. Mehta
*Division of Cardiovascular Medicine, Central Arkansas Veterans Healthcare System and The University of Arkansas for Medical Sciences, Little Rock, Arkansas*

## 1 DISCOVERY OF STATINS

In the 1960s, physicians became aware of cholesterol's role in the development of atherosclerotic heart disease. Several agents, such as nicotinic acid, cholestyramine, and clofibrate were developed, but these agents failed to substantially lower low-density lipoprotein (LDL)-cholesterol levels. Akira Endo, a Japanese scientist who had done years of research on fungi, realized that fungi could hold a solution to the cholesterol problem because bacteria—like humans—required cholesterol to maintain cell wall integrity. He reasoned that certain fungi produce a substance that destroys cholesterol in bacterial cell wall as an evolutionary adaptation. With this fundamental understanding, he set about testing more than 6000 fungal broths. In 1973, he identified the mold *Penicillium citrinum* that produced a substance that appeared to inhibit HMG-CoA reductase enzyme, a key step in LDL-cholesterol synthesis. This substance, named compactin, was the first known statin. It successfully lowered LDL-cholesterol levels in animal studies, and similar success was seen in early human trials. Interestingly enough, the first patient treated with Dr. Endo's new drug developed severe muscle cramps and weakness, now a well-known side effect of this class of drugs, and had to discontinue therapy. However, interest in this agent continued to grow as its success in lowering LDL-cholesterol became public knowledge, and several pharmaceutical companies developed several variations of this formulation. Merck was first to market lovastatin (Mevacor) in the United States in 1987. Today the world's most widely selling drug is a statin (atorvastatin, Lipitor; Pfizer), with annual worldwide sales of $12 billion [1].

## 2 STATINS: BIOCHEMISTRY AND MECHANISMS OF ACTION

Acetyl-CoA is produced from glucose, fatty acids, or essential amino acids, with three molecules of acetyl-CoA combining to form 3-hydroxy-3-methylglutaryl coenzyme A (HMG-CoA). HMG-CoA reductase catalyzes the reduction of HMG-CoA to produce mevalonate. Mevalonate produces isoprene units that condense to form squalene, and the cyclization of squalene produces the steroid ring system, with subsequent reactions generating cholesterol. HMG-CoA reductase is the rate-limiting step in the formation of cholesterol in the liver and intestine, though cholesterol formation does occur in most cells of the body (Fig. 1).

Statins compete with normal substrate for the enzyme's active binding site and once bound alter the conformation of the enzyme. This binding is reversible, but once statin binds to HMG-CoA reductase, the enzyme is prevented from attaining a functional conformation. The inhibition of HMG-CoA reductase by statins reduces the hepatocyte cholesterol content,

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

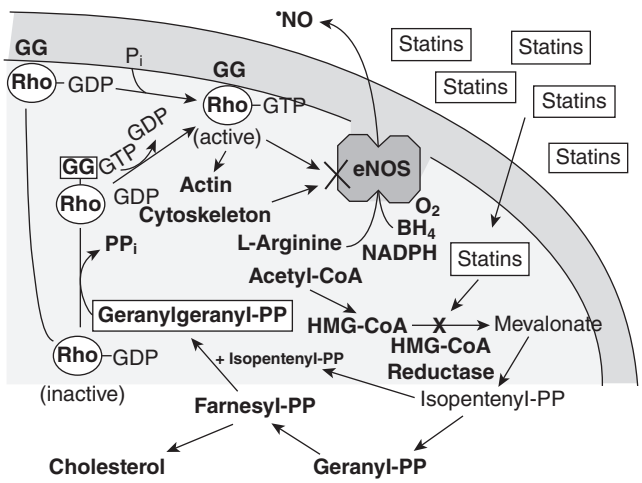


**Figure 1** Inhibition of HMG-CoA reductase by statins decreases the synthesis of isoprenoids and cholesterol. The isoprenoid, geranylgeranyl (GG), is an important lipid attachment for Rho, which permits the subsequent membrane translocation and activation of Rho and Rho kinase. Inhibition of Rho/Rho-kinase activity by statins leads to increase in eNOS expression and activity.

which subsequently upregulates the expression of LDL receptors, resulting in increased removal of LDL-cholesterol from the circulation. Statins also inhibit hepatic synthesis of apolipoprotein B-100, which results in a reduction in the synthesis and secretion of triglyceride-rich lipoproteins [2].

Statins are widely used for their LDL-cholesterol-lowering properties. Initial LDL-lowering ability varies among different statins, but with doubling of dose of any given statin, an additional 7% reduction in LDL-cholesterol can be expected. Statins also have effects on other components of the lipid profile. A reduction of triglyceride level by as much as 45% can be seen in patients with hypertriglyceridemia, and this reduction occurs via a reduction in hepatic secretion of very low density lipoprotein (VLDL). Statins also have a minor impact on high-density lipoprotein (HDL)-cholesterol levels increasing it by 5–10% [3].

## 3 STATINS: EXPERIMENTAL AND CLINICAL STUDIES

Long- and short-term trials with statins have demonstrated significant reductions in cardiovascular events in patients with and without a history of coronary heart disease (CHD). Statins are well-established LDL-cholesterol-lowering agents, but the benefit of statins exceeds that seen at a particularly low LDL-cholesterol level, which occurs naturally. This raises the issue that besides improving the lipid profile, statins have nonlipid properties that ultimately result in reduction in atherothrombosis. These effects arise through the inhibition of several mevalonate-derived metabolites other than cholesterol itself, which are involved in the control of different cellular functions. Although statins represent the reference in the prevention and treatment of CHD, combination therapy with other lipid-lowering drugs may increase their therapeutic potential.

### 3.1 Animal Studies

Initial trials of the first statin (compactin) in 1973 were disheartening. Dr. Endo initially tested his discovery in rats with disappointing results. Continued research led to the understanding that rats differ in the mechanisms by which they carry out cholesterol synthesis. Dr. Endo then tested statins in chickens with much more promising results.

With greater understanding of how animals metabolize cholesterol, selection of animal models for use in studies has improved. We now know that rats are generally resistant to atherosclerosis without genetic manipulation. Monkeys, baboons, and pigs have proven to be very good animal models, as they develop atherosclerosis in a manner similar to humans. In recent studies, animals have been genetically modified (apo-E knockout and LDL-receptor knockout mice) that develop atherosclerosis rapidly when fed diets rich in cholesterol. In these animals, statin therapy decreases aortic cholesterol accumulation, reduces the extent of atherosclerosis, and prevents diet-induced increase in LDL-cholesterol levels. These studies have shown preserved tissue superoxide dismutase activity (protective against free radicals), inhibition of LDL-cholesterol oxidation by activated macrophages, and interference with atherosclerotic lesion development by reducing intimal thickening and the expression of the cytotoxic inducible nitric oxidize synthase (iNOS). Statins have also been shown to inhibit smooth muscle cell proliferation. It is unclear if similar results are reproduced in humans at clinically relevant doses.

### 3.2 Human Studies

***Primary Prevention Trials*** Initial animal studies illustrated a connection between atherosclerotic CHD and serum lipid levels. This understanding led to the logical next step of evaluating lesion progression or regression with statin therapy. These trials were initially called "regression trials," due to the supposition that lipid-lowering therapies would cause plaque regression. Multiple large trials have proven that atherosclerotic lesions do not actually regress with statin therapy, but

progression of lesions may be slowed with a reduction in serum LDL-cholesterol levels [4]. A large number of clinical studies have shown that the statin group of drugs are effective in primary and secondary prevention of cardiovascular events. Details of these trials are given elsewhere. Here we discuss results of some large trials.

The Scandinavian Simvastatin Survival Study (4S) included 4444 Scandinavian patients and clearly showed the life-saving potential of statins [5]. However, many surprising benefits in addition to slowed atherosclerosis-related events became evident in this study. For example, it became clear that simvastatin can lower cardiovascular events regardless of baseline LDL-cholesterol, HDL-cholesterol, and triglyceride levels. The West of Scotland Coronary Prevention Study (WOSCOPS) [6] evaluated the benefit of another statin, pravastatin, on the rate of nonfatal myocardial infarction (MI) and CHD-related deaths in 6595 men with hypercholesterolemia and no prior evidence of MI or revascularization. At 5 years of follow-up, there was an absolute reduction in nonfatal MI or CHD death of 3%, and the reduction in clinical events began within 6 months of randomization. This study also showed that there was no difference in noncardiovascular death or incidence of cancer among participants.

The Air Force/Texas Coronary Atherosclerosis Prevention Study (AFCAPS/Tex CAPS) [7] was a randomized, placebo-controlled trial of lovastatin in 6605 patients without CHD. After a 5-year follow-up, for every 1000 men and women treated with lovastatin, 19 coronary events, 12 MIs, and 17 coronary revascularizations could be prevented, and this benefit was independent of changes in LDL-cholesterol. This study extended the results seen in WOSCOPS in two ways: It demonstrated the efficacy of primary prevention in women, and it demonstrated the efficacy of primary prevention in patients with "normal" serum LDL-cholesterol levels at the time of enrollment. The results of this study led to major revisions in the National Cholesterol Education Program (NCEP) Adult Treatment Panel III guidelines.

The Medical Research Council/British Heart Foundation (MRC/BHF) Heart Protection Study (HPS) assigned 20,536 subjects to treatment with either simvastatin (40 mg/day) or placebo. This study included patients with known CHD, occlusive arterial disease, and diabetes, making this study relevant in primary as well as secondary prevention. After 5 years of follow-up, there was no effect of statin therapy on the incidence of noncardiovascular deaths, but there was a significant reduction in ischemic strokes without an increase in hemorrhagic strokes [8]. This study showed benefit in men and women, older adults, and patients at risk for cardiovascular events regardless of baseline LDL-cholesterol level.

The recently completed Anglo-Scandinavian Cardiac Outcomes Trial—Lipid Lowering Arm (ASCOT-LLA) included 19,342 subjects with hypertension and at least 3 other cardiovascular risk factors [9]. Participants in this study were randomized to therapy with atorvastatin or placebo, in addition to standard antihypertensive therapy. This study was halted prematurely after 3 years when the lipid-lowering arm of ASCOT showed a 36% reduction in fatal CHD and nonfatal MI compared with the placebo group.

The Antihypertensive and Lipid-Lowering Treatment to Prevent Heart Attack Trial (ALLHAT) included 10,355 hypertensive patients who were randomized to therapy with either pravastatin or standard antihypertensive therapy. No significant benefits in terms of all-cause mortality or coronary and stroke events were apparent with statin use in this study [10]. This may be explained by the use of statins by a large number of subjects in the usual care group, leading to a smaller difference in total and LDL-cholesterol levels.

The Prospective Study of Pravastatin in the Elderly at Risk (PROSPER) included 5804 patients aged 70–82 years with a history of, or risk factors for, vascular disease [11]. Again, this study was designed as a primary and secondary prevention trial. Patients were randomized to pravastatin (40 mg) or placebo, and the follow-up period was 3.2 years. Primary endpoints (a composite of CHD death, nonfatal MI, and fatal or nonfatal stroke) were reduced by 15%. The CHD death rate was reduced by 24%, and although the incidence of transient ischemic attack was reduced, there was no reduction in stroke rate.

***Secondary Prevention Trials*** With the promising data available from these and other primary prevention trials, and the fundamental understanding that statin therapy clearly reduced serum LDL-cholesterol levels, the application of statin therapy in high-risk groups seemed promising. The absolute risk for an event is clearly highest in secondary prevention or high-risk groups, and lowest in primary or low-risk groups. Most studies were based on Framingham criteria (age, fasting LDL-cholesterol level, HDL-cholesterol level, blood pressure, diabetes, and smoking status).

Since 1994, there have been a total of six published prospective, double-blind, placebo-controlled, randomized trials with statin treatment in high-risk patient populations. These are the 4S, the Cholesterol and Recurrent Events (CARE) trial, Long-Term Intervention with Pravastatin in Ischemic Disease (LIPID), HPS, the Prospective Study of Pravastatin in the Elderly at Risk (PROSPER), and the ASCOT-LLA. All six of these

studies showed significant reduction in total cholesterol, LDL-cholesterol, and triglycerides, with an increase (or at least no reduction) in HDL-cholesterol. These studies also showed absolute relative risk reduction in all-cause mortality, fatal cardiac heart disease, and nonfatal MI, fatal and nonfatal stroke, need for revascularization, and any major vascular event.

This vast experience with both primary and secondary prevention trials led to rapid advances in current guidelines for the management of patients with cardiovascular disease. Further examination of such trial data has led to the recommendation that treatment goals for LDL-cholesterol should be the reduction of LDL-cholesterol levels to less than 100 mg/dL in moderate-risk patients and less than 70 mg/dL for high-risk patients. The beneficial effects of statins in addition to their LDL-cholesterol-lowering effects, include a reduction in inflammation, improvement in vascular function, and other pleiotropic effects, but it is not yet clear how much of the "statin effect" is related to these additional nonspecific properties [4].

## 4 STATINS: NEW ISSUES

### 4.1 Pleiotropic Effects

The pleiotropic effect of a drug refers to actions in addition to those for which the drug was specifically developed. These effects may or may not be related to the primary mechanism of the drug, and they may or may not be desirable. Such effects are, by definition, unanticipated. In addition to the well-documented beneficial effects of statins on lipid profiles, a number of studies have identified potentially other beneficial effects of statins (Table 1). Specific examples of such unanticipated effects of statins currently under investigation include the improvement in endothelial function, increased bioavailability of nitric oxide (NO), vasomotion, antioxidant and anti-inflammatory effects, and stimulation of endothelial progenitor cell recruitment.

**TABLE 1 Non-Lipid-Lowering Effects of Statins**

| |
|---|
| **Anti-Thrombotic Effects** |
| Reduction in PAI-1 |
| Increase in t-PA |
| Increase in NOS expression and activity |
| Reduction in tissue factor expression |
| Reduction in fibrinogen |
| Decreased platelet aggregation |
| Increase in thrombomodulin |
| **Effects on Endothelial Cell Biology** |
| Enhanced eNOS expression and activity |
| Reduction in genes for leukocyte adhesion and inflammatory signals |
| Reduction in endothelial permeability |
| Increased endothelial growth |
| Increase in endothelial progenitor cells/angiogenesis |
| Reduction in pre – pro ET-1 expression |
| **Effects on Vascular Smooth Muscle Cells** |
| Reduction in VSMC growth and migration |
| Decrease in expression of AT1 receptor in response to injurious stimuli |
| **Miscellaneous Effects** |
| Reduction in generation of oxidant species |
| Increase in gene expression of antioxidants |
| Decrease in fibroblast growth and collagen formation |
| Inhibition of LDL oxidation and ox-LDL uptake |
| Inhibition of mitogen-stimulated T and B-lymphocyte proliferation |
| Inhibition of expression of class II MHC |
| Blockade of the effects of angiotensin II |
| Inhibitor of natural killer cell cytotoxicity |
| Stimulation of bone formation |

PAI-1: plasminogen activator inhibitor type 1; t-PA: tissue plasminogen activator; ET-1: endothelin-1; VSMC: vascular smooth muscle cell; AT1: angiotensin II type 1; ox-LDL: oxidized low density lipoprotein; MHC: major histocompatibility complex.

*Source*: This table has previously been published and is reproduced with the permission of the publisher.

***Improvement in Endothelial Function/Increased Bioavailability of Nitric Oxide*** Long recognized as an inciting event in atherogenesis, endothelial dysfunction is an early but as-yet poorly understood process. The vascular endothelium closely regulates the contractile state of the vascular wall. Hypercholesterolemia impairs endothelial function, and this is often characterized by a paradoxical vasoconstriction caused by acetylcholine in response to an impaired synthesis, release, and activity of endothelium-derived NO. Abnormalities in endothelium-dependent vasomotor responses predict future progression of atherosclerosis and coronary events. Statins have been shown to enhance the expression of endothelial NO synthase (eNOS), the enzyme that catalyzes the formation of NO from L-arginine eNOS is generally downregulated in atherosclerosis, and the salutary effect of statins on eNOS may be mediated by the ability of LDL-cholesterol to increase the level of caveolin-1, an inhibitor of eNOS activity. Statins also enhance eNOS activity, thereby increasing the bioavailability of NO [12, 13]. Several mechanisms are likely involved, ranging from the reduction in the production of reactive oxygen species, which inactivate NO, and to the stabilization of eNOS messenger ribonucleic acid (RNA) [12].

As discussed above, treatment with statins to improve endothelial function, though the exact mechanism by which this occurs is as yet not clear. In a study of hypercholesterolemic patients, treatment with fluvastatin for 6–12 weeks improved myocardial perfusion in previously ischemic segments, and the improvement in perfu-

sion was greater than that seen in normal segments [14]. In another study of patients with moderately elevated cholesterol levels, treatment with simvastatin increased vasodilatory response to acetylcholine as determined by forearm blood flow. This improvement in blood flow continued to increase over the duration of treatment [15]. Other studies have compared changes in vasomotion after statin therapy to either dietary modifications or placebo, with significant improvement seen only in the statin group. These improvements appear to occur regardless of serum cholesterol levels or C-reactive protein (CRP) levels, and indeed they appear within hours of initiation of statin therapy—long before any lipid-lowering effect is appreciated. The normalized vasomotion achieved with statins has also been shown to occur in high-risk groups, such as diabetics.

***Antioxidant Effects*** Statins inhibit oxidation of LDL-cholesterol both in vitro and ex vivo, in addition to their ability to reverse the inhibitory effect of oxidized LDL-cholesterol on eNOS expression and activity. Statins have also been shown to inhibit the oxidation of LDL-cholesterol as well as HDL-cholesterol. In addition, statins have been shown to exert free-radical scavenging effect that may contribute to the inhibition of lipoprotein oxidation [17]. Statins also indirectly inhibit normal oxidative mechanisms by decreasing the ability of macrophages to oxidize lipoproteins, although the exact mechanism of this effect is still under investigation. In studies on the antioxidant effects of statins, the benefits were seen within days of initiation of treatment with statins, long before serum LDL-cholesterol levels were altered.

***Anti-Inflammatory Effects*** In the recent past, a large number of studies have revealed the importance of inflammation in the development of atherosclerosis. Elevated serum markers of inflammation such as (CRP), interleukin-6, intercellular adhesion molecule-1, and serum amyloid A have all been shown to be associated with increased risk for cardiovascular events. While the precise cause–effect relationship between these markers and overt cardiovascular disease remains elusive, statin therapy has been shown to lower the serum levels of these markers in humans. In the CARE trial, long-term follow-up showed that subjects with elevated CRP levels benefited most from provastatin [16]. In this study, reduction of CRP levels was not correlated with a reduction of serum LDL-cholesterol levels. Other studies have shown a reduction in CRP levels shortly after initiation of statin therapy and long before serum LDL-cholesterol levels decline. It has been postulated that this rapid reduction in CRP could potentially explain the early benefits of these drugs seen in acute coronary syndromes. However, no study has yet shown that statins can decrease CRP levels independent of LDL-cholesterol levels and reduce cardiovascular events.

***Plaque Regression/Stabilization*** Elegant studies in animal models have shown the plaque-stabilizing effects of statins. While a portion of this effect may be due to reduction in serum LDL-cholesterol levels, statins have also been shown to inhibit uptake of oxidized LDL-cholesterol by cluster of differentiation 36 scavenger receptor (CD-36), scavenger receptor A, and lectin-like oxidized LDL (LOX-1) receptors as well as inhibit macrophage oxidative properties. As mentioned above, elevated serum levels of inflammatory markers have been shown to portend poor outcomes. There is emerging data to suggest that some of these poor outcomes may be mediated by plaque rupture. Weakening of the fibrous cap in such unstable plaques is associated with increased macrophage production of matrix metalloproteinases. In human studies, statins have been shown to change the composition of such plaques in a way that favors stabilization. After statin treatment, such plaques are composed of less lipid and oxidized LDL-cholesterol, fewer macrophages, and fewer T cells. In addition, such plaques have higher collagen content and less matrix metalloproteinase immunoreactivity, as well as reduced rates of apoptosis [12]. These promising plaque-stabilizing effects of statins may also potentially explain the usefulness of this class of drugs in acute coronary syndromes.

***Stimulation of Endothelial Progenitor Cells*** Statins have shown promise in the prevention of vascular ischemic injury. Statins also play an important role in repair of vascular injury by augmenting endothelial progenitor cell differentiation. Work in humans has shown that statins are at least as effective as vascular endothelial growth factor, a key cytokine in the regulation of neovascularization, in stimulating this endothelial progenitor cell differentiation. In addition to this augmentation, statins have also been shown to increase the number of endothelial progenitor cells, which subsequently improve ischemia-induced neovascularization, accelerates reendothelialization after injury, and improves postischemic cardiac function [4]. As with so many of the pleiotropic effects of statins, these effects were seen at time intervals much shorter than those required for modulation of serum LDL-cholesterol levels.

***Immunomodulation*** Increasing evidence suggests that statins may act as immunomodulators. With this realization, these agents have been used in solid organ transplant recipients. Statins have been shown to decrease interferon-induced major histocompatibility complex-II (MHC-II) in human endothelial cells and macrohpages. This effect is reversed by mevalonate and

has been attributed to the inhibitory effect of statins on promoter IV of MHC-II transactivating factor, which leads to suppression of T-lymphocyte activation. This suppression of T-helper-cell-I immune response is the proposed mechanism for the benefits seen in solid organ transplant recipients.

### 4.2 Combination with Other Drugs

***Ezetimibe*** Ezetimibe, a potent intestinal cholesterol absorption inhibitor, offers the potential address the question of whether all the effects of statins can be attributed to their cholesterol-lowering benefits. Ezetimibe significantly lowers LDL-cholesterol levels when used as monotherapy or in combination with statins. Indeed, ezetimibe is often used in combination with statins in high-risk patients who require aggressive lipid-lowering therapy to achieve an appropriate serum LDL-cholesterol level. In a recent study, forearm blood flow responses to acetylcholine and sodium nitroprusside were measured in patients with stable coronary artery disease (CAD) after 4 weeks of therapy. After 4 weeks of therapy, LDL-cholesterol levels were reduced in all patients; however, in the ezetimibe group, there was no increase in acetylcholine-mediated forearm blood flow response. In contrast, dose escalation of atorvastatin or de novo atorvastatin monotherapy was associated with a significant improvement in acetylcholine-mediated forearm blood flow response [17]. This study shows that while ezetimibe effectively lowers LDL-cholesterol levels, only statin therapy was associated with improved vasodilator function. However, whether LDL-cholesterol reduction with ezetamide offers the same benefits as that with statins is not yet known.

***Niacin/Fibrates*** Reduction of LDL-cholesterol with statin therapy is currently identified in treatment guidelines as the primary focus for patients with or at risk of CHD. Yet despite effective statin therapy, there is still an unacceptably high residual coronary risk. A substantial proportion of patients with CHD have mixed dyslipidemia, including low levels of HDL-cholesterol, an independent and predictive risk factor for CHD. Although effective in reducing LDL-cholesterol, statin therapy has only modest effects in raising HDL-cholesterol. Fibrate therapy is an alternative lipid-modifying strategy, and is effective in reducing CHD mortality and morbidity, with the magnitude of clinical benefit similar to statin therapy.

Multidrug therapy with complementary mechanisms of action has been proposed as a means of improving lipid-modifying efficacy. Nicotinic acid is the most potent agent for increasing HDL-cholesterol and also substantially reduces LDL-cholesterol and triglycerides. Addition of nicotinic acid to statin therapy would be a logical management approach, given the potential for complementary therapeutic benefit. The clinical benefits of this combination are supported by the results of the HDL Atherosclerosis Treatment Study (HATS), which showed reduction of 60–90% in the incidence of major CAD events when both agents were administered. In addition, combination treatment led to angiographic regression of coronary stenosis, compared with placebo, rather than slowed progression as previously reported with statin monotherapy. Given that the prevalence of low HDL-cholesterol, particularly among individuals with CHD, is higher than previously anticipated, combining nicotinic acid and a statin represents an innovative approach to further reducing CHD risk.

To date, the most compelling evidence that the concentration of HDL-cholesterol is a strong, independent, inverse predictor of coronary heart disease risk comes from the Veterans Affairs High-Density Lipoprotein Cholesterol Intervention Trial (VA-HIT) and the Helsinki Heart Study (HHS). In VA-HIT, CHD patients randomized to gemfibrozil experienced a 6% increase in HDL-cholesterol, a 31% decrease in plasma triglycerides, and no change in LDL-cholesterol. Major coronary events were reduced by 22% compared with placebo. The change in HDL-cholesterol was significantly associated with reductions in coronary events, whereas the change in triglycerides was not. Statins when combined with fibrates have potentially beneficial effects on HDL-cholesterol in addition to LDL-cholesterol. However, fibrate addition to statin therapy increases the risk of hepatic injury.

***Cholestery Transfer Protein (CETP) Inhibitors*** Some years ago, a mutation in the CETP gene was discovered in a select Japanese population. This mutation was accompanied by a substantial increase in HDL-cholestrol. Because HDL-cholesterol regulates reverse cholesterol transport and has antioxidant, antithrombotic, and anti-inflammatory properties, the prevalence of atherosclerotic disease is low in Japanese. Epidemiological studies were undertaken to define the prevalence of the mutation. These studies suggested that the mutation occurred in 11% of the Japanese population and that affected individuals were resistant to atherosclerosis [18, 19]. These epidemiological data triggered interest in the biochemical characteristics of CETP.

High CETP concentration has been associated with more rapid progression of CAD, and statin therapy is more effective in this population, independent of baseline or on-treatment lipid levels, suggesting that the plasma CETP level may be an important determinant

of the response to statins [20]. Statins reduce plasma CETP activity by 5–10% and cholesteryl ester transfer by 35% [20–28]. For instance, atorvastatin reduces CETP activity by 7%, and cholesterol transfer from HDL to non-HDL-cholesterol particles is reduced by 37% [27, 28]. The magnitude of these effects is modified by the CETP genotype [29, 30]. Because statins alone cause a substantial decrease in triglyceride level and a small increase in HDL-cholesterol, one might speculate that the combination of a statin with a CETP inhibitor would be more effective than either therapy alone.

Recently, a new class of drugs, CETP inhibitors, has been developed. Torcetrapib and JTT-705 are currently in the preliminary stages of clinical development. JTT-705 in combination with pravastatin has been assessed in a randomized, double-blind, placebo-controlled trial conducted in 152 individuals with elevated LDL-cholesterol. This agent increased HDL-cholesterol by 30% [31]. Torcetracip has been studied alone and with atrovastatin [32]. Torcetrapib increases HDL-cholesterol from 60 to 120%, and a number of clinical trials are underway to examine if torcetrapib with statins reduces cardiovascular events more than statins, and if yes, are there associated side effects?

***Combination with Fish Oil*** There is clear evidence that statins are effective agents for lowering LDL-cholesterol and reducing mortality in patients with CHD. Omega-3 polyunsaturated fatty acids (PUFA) have been shown to reduce serum triglyceride levels, a condition that often persists after treatment with statins. In many early small clinical trials, there was evidence that omega-3 PUFA supplementation lowered CHD mortality; however, in all of these studies, patients were not receiving doses that would be expected to significantly lower serum triglyceride levels. Early enthusiasm over fish oil supplementation was also limited by data suggesting that daily dietary supplementation with omega-3 PUFA resulted in an undesirable increase in serum LDL-cholesterol.

Subsequent studies have shown that treatment with omega-3 PUFA in combination with statin therapy, and at doses sufficient to lower serum triglyceride levels, was generally well tolerated and safe. Other research has shown no adverse effects on LDL-cholesterol and no adverse interactions with statin treatment [33]. This new understanding has led to more widespread use of combined therapy with statins and omega-3 PUFA in patients with both elevated serum LDL-cholesterol levels and elevated triglyceride levels [34].

***Combination with Aspirin*** Aspirin has long been recognized for its immediate antiplatelet effect, and this has led to its widespread use in both acute coronary syndromes and as a preventative therapeutic measure. Statins are now widely used for their delayed antiatherogenic effect. Recent studies have examined the possibility of additive clinical benefits with combination (aspirin+statin) therapy. Recently published guidelines for statin and aspirin use indicate a favorable risk–benefit ratio for all patients who have survived prior cardiovascular events, and indeed, the U.S. Food and Drug Administration (FDA) has approved the use of combination therapy with aspirin and statin in the secondary prevention of cardiovascular disease.

Meta-analyses have shown relative risk reductions for fatal or nonfatal MI of 31% for pravastatin plus aspirin versus aspirin alone, and 26% for pravastatin plus aspirin versus pravastatin alone. For the composite endpoint of CHD death, nonfatal MI, coronary artery bypass graft, percutaneous transluminal coronary angioplasty, or ischemic stroke, similar relative risk reductions were seen. Despite these findings, there remains widespread underutilization of this effective drug combination [35].

***Combination with Peroxisome Proliferator-activated Receptor (PPAR) Activators*** Inhibition of cardiac remodeling has rapidly come to the forefront in long-term management of postinfarction patients, and statins have recently been recognized for their beneficial effects in prevention of adverse cardiac remodeling in the post-MI patient. Angiotensin II is known to enhance collagen synthesis, and it is this enhanced collagen synthesis that is felt to play a key role in cardiac remodeling following injury. Increased levels of angiotensin II have been shown in a mouse model to induce oxidative stress and procollagen-I expression in cardiac fibroblasts. In recently published studies, it has been shown that both pravastatin and pioglitazone, a member of the PPAR class of agents, both individually block the effect of angiotensin II, and both agents affect differing pathways. The combination of these agents appears to have an additive effect in terms of reduced myocyte oxidative stress and increased inhibition of procollagen I formation. While trials in humans have yet to be reported, these promising results in animal models point to another potential powerful use of statins for their cardioprotective effects [36].

***Combination with Antiplatelet Agents*** The early realization that clopidogrel and atorvastatin are competitive substrates for the cytochrome P450 3A4 (CYP3A4) hepatic enzyme system caused much concern. Because clopidogrel has low bioavailability, because atorvastatin has a higher affinity for the substrate binding site than does clopidogrel, and because atorvastatin has a long half-life, atorvastatin inhibits the ability of clopidogrel

to be metabolized into its active metabolite in a dose-dependent manner. A 600-mg loading dose of clopidogrel will overcome this competitive binding, but neither a 300-mg loading dose or the daily 75-mg dose of clopidogrel are sufficient to prevent atorvastatin from decreasing the metabolism of clopidogrel. The clinical significance of this drug–drug interaction remains, as yet, unclear. It is clear that other statins, namely pravastatin and fluvastatin, which are not metabolized via the hepatic CYP3A4 system, have little to no effect on metabolism of clopidogrel. However, it should be pointed out that it remains unclear today whether this potential interaction between atorvastatin and clopidogrel is clinically significant [37, 38].

## 5 WITHDRAWAL PHENOMENON

Despite the well-recognized beneficial effects of statins, there have been patients who have abruptly discontinued their use due to financial constraints, intolerable side effects, or mere forgetfulness. Several studies have shown an increased incidence of cardiac events in patients with atherosclerotic heart disease after abrupt cessation of statin therapy. In fact, one study showed a rapid increase in CRP and interleukin-6 levels in patients with hypercholesterolemia after abrupt cessation of statin therapy [39]. This study has led many to propose that the increased inflammatory response seen after such abrupt cessation of statin therapy may explain the increased incidence of cardiovascular events. Regardless of the mechanism of this phenomenon, statin therapy should not be abruptly discontinued.

## 6 OTHER DISEASE STATES

### 6.1 Aortic Stenosis

Once thought to be simply a "wear-and-tear" phenomenon, it is now thought that aortic stenosis has some pathological similarities with atherosclerotic disease. Lipid deposition and macrophage infiltration, with subsequent inflammation, now appears to be the common pathophysiologic process between the two conditions. There are studies that show accelerated progression of aortic stenosis in patients with hyperlipidemia, and the similarities among risk factors for aortic stenosis and atherosclerotic disease has led to attempts to slow the progression of aortic stenosis with statin therapy. Several small studies have reported reductions in aortic valve calcium accumulation, smaller increases in mean gradients across stenotic aortic valves, and reductions in progression of aortic stenosis in patients treated with statins (Table 2) [40–46]. This new understanding of the pathophysiology of aortic stenosis and the promising early data indicate that statins may have an increasing role to play in the management of aortic stenosis. However, a recent large study showed that intensive statin therapy does not halt the progression of calcific aortic stenosis or induce its regression [47].

### 6.2 Alzheimer's Disease

The pathological hallmark of Alzheimer's disease is the accumulation of β-amyloid neurofibrillatory plaques in the central nervous system tissue. Accumulation of β-amyloid is a result of abnormal proteolytic processing of amyloid precursor protein (APP), and there is evidence that this abnormal processing of APP is mediated by cholesterol [48]. Indeed, in transgenic mice overexpressing APP, a high-cholesterol diet significantly increases the level of β-amyloid plaques in the brain. When these mice were treated with cholesterol-lowering drugs, the β-amyloid plaque burden in the central nervous system (CNS) was decreased. This finding led to interest in the use of cholesterol-lowering agents as a preventative approach to Alzheimer's dementia. However, studies in humans have failed to substantiate this theory. In one recent study, after adjusting for age, gender, education, comorbid vascular disease, body mass index, and history of lipid-lowering drug therapy, there was no association between total LDL- or HDL-cholesterol and risk of dementia or Alzheimer's disease [49].

### 6.3 Bone Degeneration

Estrogen replacement therapy, selective estrogen receptor modulators, and bisphosphonates—all current therapies for treatment of osteoporosis—are focused on blunting the resorption phase of bone remodeling. Bone morphogenic protein-2 (BMP-2), a central growth factor in osteoblast differentiation, is incorporated into bone matrix during bone formation. It is later released, in active form, during bone resorption. BMPs are responsible for enhancing osteoclast differentiation, including stimulation of expression of the structural proteins of bone. Statins have been shown to stimulate transcription of BMP-2, and, therefore, enhance the process of osteoblast differentiation and new bone formation. Recent studies have indeed shown increases in bone mineral density and reduced fracture rates in patients treated with statins. One caveat in this arena is that all of the currently available statins are metabolized in the liver; therefore, their delivery to bone is greatly reduced. Further research into delivery of statins to bone, perhaps through skin patches, is needed [4].

**TABLE 2 Nonrandomized Retrospective Studies of Medical Therapy to Prevent Progression of Calcific Aortic Stenosis**

| First Author, Year | No. of Subjects | Mean Age (year) (% Male) | Mean F/U Interval (year) | Study Groups | Parameter Reported | *p-value* |
|---|---|---|---|---|---|---|
| EBT[a] calcium | | | | | | |
| Pohle 2001 [41] | 104 | 65 (86) | 1.25 | | Increase in valve calcium (per year) | |
| | | | | LDL ≤ 130 mg/dL | 9 ± 22% | ≤0.001 |
| | | | | LDL ≥ 130 mg/dL | 43 ± 44% | |
| Shavelle 2002 [42] | 65 | 67 | 2.5 | | Increase in valve calcium (per year) | |
| | | | | Statin therapy (43%) | 12.1% | 0.006 |
| | | | | No statin therapy | 32% | |
| Echocardiography | | | | | | |
| Aronow 2001 [43] | 180 | 82 (31) | 2.8 | | Increase in peak gradient (mmHg/year) | |
| | | | | LDL ≥ 125 mg/dL, not on statin | 6.3 ± 1.4 | <0.0001 |
| | | | | LDL ≥ 125 mg/dL, on statin | 3.4 ± 1.0 | |
| | | | | LDL ≤ 125 mg/dL, not on statin | 3.1 ± 1.1 | |
| Novaro 2001 [44] | 174 | 68 (44) | 1.8 | | Decrease in valve area ($cm^2$/year) | 0.03 |
| | | | | Statin therapy (33%) | 0.06 ± 0.016 | |
| | | | | No statin therapy | 0.11 ± 0.18 | |
| Bellamy 2002 [45] | 156 | 77 (58) | 3.7 | | Decrease in valve area ($cm^2$/year) | 0.04 |
| | | | | Statin therapy (24%) | 0.04 ± 0.15 | |
| | | | | No statin therapy | 0.09 ± 0.17 | |
| Rosenhek 2004 [46] | 211 | 70 (51) | 2.0 | | Increase in aortic velocity ($m.s^{-1}.y^{-1}$) | |
| | | | | Statin therapy | 0.10 ± 0.41 | 0.0001 |
| | | | | No statin therapy | 0.39 ± 0.42 | |
| | | | | ACE inhibitor | 0.29 ± 0.44 | 0.29 |
| | | | | No ACE inhibitor | 0.35 ± 0.44 | |

[a]EBT indicates electron-beam tomography. ACE indicates angiotensin-converting enzyme.

*Source*: This table has previously been published and is reproduced with the permission of the publisher.

### 6.4 Immunosuppression/Malignancy

As discussed previously, use of statins in recipients of solid organ transplants has shown promise in terms of reduced rates of rejection. Two major mechanisms for these beneficial effects have been proposed. First, statins inhibit interferon-δ-inducible MHC-II expression on macrophages and endothelial cells. Second, statins directly interfere with lymphocyte function-associated antigen-1 (LFA-1) binding to intracellular adhesion molecule-1. Both of these events impair activation of host T cells, and it is thought that this explains the reduced rates of rejection in transplant recipients. There have also been recent reports of promising results of statin use in patients with autoimmune diseases, such as rheumatologic conditions and multiple sclerosis. However, more data is needed before statins can be recommended as standard therapy in such conditions [50]. Statins have recently proven of use in modulation of inflammation-induced angiogenesis. This has spurred interest in their possible use in the treatment of malignancy. There have been some recent observational studies published that have shown reduced rates of diagnosis of cancer and fewer cancer-related deaths in statin users, but again, more data is needed.

### 6.5 Atrial Fibrillation

In the postoperative period, atrial fibrillation (AF) is thought to be initiated by triggered electrophysiologic abnormalities, which mostly occur in patients who have an abnormal atrial substrate [51]. High catecholamine

state [52] and inflammation [53] seen postoperatively is thought to play a key role. Increase in markers of inflammation, such as CRP and interleukin-6 after coronary artery bypass surgery, indicates an active inflammatory process after the procedure. Inflammatory changes have also been detected in the atrial biopsy specimens of patients with AF [54]. These findings suggest that AF may be induced by structural changes in the atria that are promoted by inflammation. Statins have modest anti-inflammatory actions, as indicated by decreased CRP levels with their administration [53, 55].

An observational study in 362 patients [56] showed postoperative AF was less frequent and its duration was shorter in statin group compared to nonstatin group ($p = 0.03$ and 0.0001, respectively). Another observational study of 25,268 patients from a registry of AF [57], 71.3% of patients had hyperlipidemia, and 66.8% were prescribed lipid-lowering drugs. Atrial fibrillation prevalence was 25.1% in patients taking lipid-lowering drugs compared with 32.6% in untreated hyperlipidemic patients and 32.8% in patients without hyperlipidemia ($p < 0.001$ for both comparisons). Lipid-lowering drug use was associated with reduced odds of AF, and this effect was larger than that of angiotensin-converting enzyme inhibitors/angiotensin receptor blockers or $\beta$-blockers. The beneficial impact of lipid-lowering drugs on atrial fibrillation risk was independent of their effects on the lipid profile.

## 7 CONCLUSION

The major advances in CHD diagnosis and treatment in the past half-century include the detection of thrombosis or a mechanism of acute coronary syndromes that leads to heart attacks and discovery of major risk factors that leads to progression of atherosclerosis. Next was the role of aspirin in prevention. The most eradicable risk factor is smoking cessation in the prevention of CHD. Most modifiable risk factor is LDL-cholesterol reduction that has been shown to cause regression of atherosclerosis and reduction in cardiac events. Statins play a major role in reducing LDL-cholesterol, hence reducing cardiac events in every study. Statins not only have their effects through LDL-cholesterol reduction, but have LDL-cholesterol-independent effects. The significance of the pleitropic effects is not well defined. Statins are, nonetheless, the only class of drugs that has been shown to consistently reduce CHD risk.

## REFERENCES


1. Landers P (2006, Jan. 9). Stalking cholesterol: How one scientist intrigued by molds found first statin. *Wall Street Journal*, A1.
2. Stancu C, Sima A (2001). Statins: Mechanism of action and effects. *J Cell Mol Med 4*, 378–387.
3. Vaughan C, Gotto A (2004). Update on statins: 2003. *Circulation 110*, 886–892.
4. Mehta J (2004). *Statins: Understanding Clinical Use*. Saunders Publishing.
5. Pederson TR, Kjekshus J, Pyorala K, et al. (1998). Effect of simvastatin on ischemic signs and symptoms in the Scandinavian Simvastatin Survival Study (4S). *Am J Cardiol 81*, 333–335.
6. West of Scotland Coronary Prevention Study Group (1997). Baseline risk factors and their association with outcome in the West of Scotland Coronary Prevention Study. *Am J Cardiol 79*, 756–762.
7. Downs JR, Clearfield M, Weis S, et al., for the AFCAPS/TexCAPS Research Group (1998). Primary prevention of acute coronary events with lovastatin in men and women with average cholesterol levels: Results of AFCAPS/TexCAPS. *JAMA 279*, 1615–1622.
8. Heart Protection Study Collaborative Group (2002). MRC/BHF Heart Protection Study of cholesterol lowering with simvastatin in 20536 high-risk individuals: A randomized controlled trial *Lancet 360*, 1623–1630.
9. ASCOTT-LLA (2003). Prevention of coronary and stroke events with atorvastatin in hypertensive patients who have average or lower-than-average cholesterol concentrations, in the Anglo-Scandinavian Cardiac Outcome Trial—Lipid Lowering Arm (ASCOTT-LLA): A multicentre randomized controlled trial. *Lancet 361*, 1149–1158.
10. The ALLHAT Officers and Coordinators for the ALLHAT Collaborative Research Group (2002). Major outcomes in moderately hypercholesterolemic, hypertensive patients randomized to pravastatin vs. usual care. *JAMA 288*, 2998–3007.
11. PROSPER Study Group (2002). Pravastatin in elderly individuals at risk of vascular disease (PROSPER): A randomized controlled trial. *Lancet 360*, 1623–1630.
12. Davignon J (2004). Beneficial cardiovascular pleiotropic effects of statins. *Circulation 109*(suppl III), III-39–III-43.
13. Calabro P, Yeh ET (2005). The pleiotropic effects of statins. *Curr Opin Cardiol 20*, 541–546.
14. Eichstadt HW, Eskotter H, Hoffman I, et al. (1995). Improvement of myocardial perfusion by short-term fluvastatin therapy in coronary artery disease. *Am J Cardiol 76*, 122A–125A.
15. O'Driscoll G, Green D, Taylor RR (1997). Simvastatin, an HMG-coenzyme A reductase inhibitor, improves endothelial function within 1 month. *Circulation 95*, 1126–1131.
16. Ridker PM, Rifai N, Pfeffer MA, et al. (1999). Long-term effects of pravastatin on plasma concentration of C-reactive protein. The cholesterol and Recurrent Events (CARE) Investigators. *Circulation 100*, 230–235.
17. Fichtischerer S, Schmidt-Lucke C, Bojunga S, Rossig L, Heeschen C, Dimmeler S, Zeiher A (2006). Differential effects of short-term lipid lowering with ezetimibe and

statins on endothelial function in patients with CAD: Clinical evidence for "pleiotropic" functions of statin therapy. *Eur Heart J 27*(10), 1182–1190.

18. Koizumi J, Mabuchi H, Yoshimura A, Michishita I, Takeda M, Itoh H, Sakai Y, Sakai T, Ueda K, Takeda R (1985). Deficiency of serum cholesteryl-ester transfer activity in patients with familial hyperalphalipoproteinemia. *Atherosclerosis 58*, 175–186.
19. Moriyama Y, Okamura T, Inazu A, Doi M, Iso H, Mouri Y, Ishikawa Y, Suzuki H, Iida M, Koizumi J, Mabuchi H, Komachi Y (1998). A low prevalence of coronary heart disease among subjects with increased high-density lipoprotein cholesterol levels, including those with plasma cholesteryl ester transfer protein deficiency. *Prev Med 27* (5, Pt 1), 659–667.
20. Klerkx AH, de Grooth GJ, Zwinderman AH, Jukema JW, Kuivenhoven JA, Kastelein JJ (2004). Cholesteryl ester transfer protein concentration is associated with progression of atherosclerosis and response to pravastatin in men with coronary artery disease (REGRESS). *Eur J Clin Invest 34*, 21–28.
21. Kwon MJ, Song YS, Choi MS, Park SJ, Jeong KS, Song YO (2003). Cholesteryl ester transfer protein activity and atherogenic parameters in rabbits supplemented with cholesterol and garlic powder. *Life Sci 72*, 2953–2964.
22. Kwon MJ, Song YS, Choi MS, Song YO (2003). Red pepper attenuates cholesteryl ester transfer protein activity and atherosclerosis in cholesterol-fed rabbits. *Clin Chim Acta 332*, 37–44.
23. Corbex M, Poirier O, Fumeron F, Betoulle D, Evans A, Ruidavets JB, Arveiler D, Luc G, Tiret L, Cambien F (2000). Extensive association analysis between the CETP gene and coronary heart disease phenotypes reveals several putative functional polymorphisms and gene-environment interaction. *Genet Epidemiol 19*, 64–80.
24. Guerin M, Dolphin PJ, Talussot C, Gardette J, Berthezene F, Chapman MJ (1995). Pravastatin modulates cholesteryl ester transfer from HDL to apoB-containing lipoproteins and lipoprotein subspecies profile in familial hypercholesterolemia. *Arterioscler Thromb Vasc Biol 15*, 1359–1368.
25. Contacos C, Barter PJ, Vrga L, Sullivan DR (1998). Cholesteryl ester transfer in hypercholesterolemia: fasting and postprandial studies with and without pravastatin. *Atherosclerosis 141*, 87–98.
26. Inazu A, Koizumi J, Kajinami K, Kiyohar T, Chichibu K, Mabuchi H (1999). Opposite effects on serum cholesteryl ester transfer protein levels between long-term treatments with pravastatin and probucol in patients with primary hypercholesterolemia and xanthoma. *Atherosclerosis 145*, 405–413.
27. Guerin M, Lassel TS, Le Goff W, Farnier M, Chapman MJ (2000). Action of atorvastatin in combined hyperlipidemia: Preferential reduction of cholesteryl ester transfer from HDL to VLDL1 particles. *Arterioscler Thromb Vasc Biol 20*, 189–197.
28. Mann CJ, Yen FT, Grant AM, Bihain BE (1991). Mechanism of plasma cholesteryl ester transfer in hypertriglyceridemia. *J Clin Invest 88*, 2059–2066.
29. Kotake H, Sekikawa A, Tokita Y, Ishigaki Y, Oikawa S (2002). Effect of HMG-CoA reductase inhibitor on plasma cholesteryl ester transfer protein activity in primary hypercholesterolemia: comparison among CETP/TaqIB genotype subgroups. *J Atheroscler Thromb 9*, 207–212.
30. Cilingiroglu M, Ballantyne C (2004). Endothelial lipase and cholesterol metabolism. *Curr Atheroscler Rep 6*, 126–130.
31. Kuivenhoven JA, deGrooth GJ, Kawamura H, Klerkx AH, Wilhelm F, Kastelein JJP (2005). Inhibition of cholesteryl ester transfer protein by JTT-705 in combination with pravastatin in type II dyslipidemia. *Am J Cardiol 95*, 1085–1088.
32. Brousseau ME, Schaefer EJ, Wolfe ML, Bloedon LT, Digenio AG, Clark RW, Mancuso JP, Rader DJ (2004). Effects of an inhibitor of cholesteryl ester transfer protein on HDL cholesterol. *N Engl J Med 350*, 1505–1515.
33. Durrington PN, Bhatnagar D, Mackness MI, Morgan J, Julier K, Khan MA, France M (2001). An omega-3 polyunsaturated fatty acid concentrate administered for one year decreased triglycerides in simvastatin treated patients with coronary heart disease and persisting hypertriglyceridaemia. *Heart 85*, 544–548.
34. Szapary PO, Rader DJ (2001). Pharmacological management of high triglycerides and low high-density lipoprotein cholesterol. *Curr Opin Pharmacol 1*, 113–120.
35. Hennekens CH, Sacks FM, Tonkin A, Jukema W, Byington RP, Pitt B, Berry DA, Berry SM, Ford NF, Walker AJ, Natarajan K, Sheng-Lin C, Fiedorek FT, Belder R (2004). Additive benefits of pravastatin and aspirin to decrease risks of cardiovascular disease. Randomized and observational comparisons of secondary prevention trials and their meta-analyses. *Arch Intern Med 164*, 40–44.
36. Chen J, Mehta JL (2006, Oct). Angiotensin II-mediated oxidative stress and pro-collagen-1 expression in cardiac fibroblasts: Blockade by pravastatin and pioglitazone. *Am J Physiol Heart Circ Physiol 291*, H1738–H1745.
37. Bates ER, Lau WC (2005). Controversies in antiplatelet therapy for patients with cardiovascular disease. *Circulation 111*, 267–271.
38. Tafreshi MJ, Zagnoni LG, Gentry EJ (2006). Combination of clopidogrel and statins: a hypothetical interaction or therapeutic dilemma? *Pharmacotherapy 26*, 388–394.
39. Li JJ, Li YS, Chen J, Yang JQ (2006). Rebound phenomenon of inflammatory response may be a major mechanism responsible for increased cardiovascular events after abrupt cessation of statin therapy. *Med Hypothesis 66*, 1199–1204.
40. Hughes BR, Chahoud G, Mehta JL (2005). Aortic stenosis: Is it simply a degenerative process or an active atherosclerotic process? *Clin Cardiol 28*, 111–114.
41. Pohle K, Maffert R, Ropers D, et al. (2001) Progression of aortic valve calcification: association with coronary

atherosclerosis and cardiovascular risk factors. *Circulation 104*, 1927–1932.

42. Shavelle DM, Takasu J, Budoff MJ, et al. (2002) HMG CoA reductase inhibitor (statin) and aortic valve calcium. *Lancet 359*, 1125–1126.
43. Aronow WS, Ahn C, Kronzon I, et al. (2001). Association of coronary risk factors and use of statins with progression of mild valvular aortic stenosis in older persons. *Am J Cardiol 88*, 693–695.
44. Novaro GM, Tiong IY, Pearce GL, et al. (2001). Effect of hydroxymethylglutaryl coenzyme A reductase inhibitors on the progression of calcific aortic stenosis. *Circulation 104*, 2205–2209.
45. Bellamy MF, Pellikka PA, Klarich KW, et al. (2002). Association of cholesterol levels, hydroxymethylglutaryl coenzyme-A reductase inhibitor treatment, and progression of aortic stenosis in the community. *J Am Coll Cardiol 40*, 1723–1730.
46. Rosenhek R, Rader F, Loho N, et al. (2004). Statins but not angiotensin-converting enzyme inhibitors delay progression of aortic stenosis. *Circulation 110*, 1291–1295.
47. Cowell SJ, Newby DE, Prescott RJ, et al. (2005). A randomized trial of intensive lipid lowering therapy in calcific aortic stenosis. *N England Med 352*, 2389–2397.
48. Li G, Shofer JB, Kukull WA, Peskind ER, Tsuang DW, Breitner JCS, McCormick W, Bowen JD, Teri L, Schellenbert GD, Larson EB (2005). Serum cholesterol and risk of Alzheimer disease: a community-based cohort study. *Neurology 65*, 1045–1050.
49. Li G, Higdon R, Kukull WA (2004). Statin therapy and risk of dementia in the elderly: a community-based prospective cohort study. *Neurology 63*, 1624–1628.
50. Schonbeck U, Libby P (2004). Inflammation, immunity, and HMG-CoA reductase inhibitors: Statins as anti-inflammatory agents? *Circulation 109*(Suppl II), II-18–II-26.
51. Cox JL (1993). A perspective of postoperative atrial fibrillation in cardiac operations. *Ann Thorac Surg 56*, 405.
52. Thompson AE, Hirsch GM, Pearson GJ (2002). Assessment of new onset postcoronary artery bypass surgery atrial fibrillation: Current practice pattern review and the development of treatment guidelines. *J Clin Pharm Ther 27*, 21–37.
53. Kumagai K, Nakashima H, Saku K (2004). The HMG-CoA reductase inhibitor atorvastatin prevents atrial fibrillation by inhibiting inflammation in a canine sterile pericarditis model. *Cardiovasc Res 62*, 105–111.
54. Frustaci A, Chimenti C, Bellocci F, Morgante E, Russo MA, Maseri A (1997). Histological substrate of atrial biopsies in patients with lone atrial fibrillation. *Circulation 96*, 1180–1184.
55. Plenge JK, Hernandez TL, Weil KM, Poirier P, Grunwald GK, Marcovina SM, Eckel RH (2002). Simvastatin lowers C-reactive protein within 14 days: An effect independent of low-density lipoprotein cholesterol reduction. *Circulation 106*, 1447–1452.
56. Ozaydin M, Dogan A, Varol E, Kapan S, Tuzun N, Peker O, Aslan SM, Altinbas A, Ocal A, Ibrisim E (2006). Statin use before by-pass surgery decreases the incidence and shortens the duration of postoperative atrial fibrillation. *Cardiology 107*, 117–121.
57. Hanna IR, Heeke B, Bush H, Brosius L, King-Hageman D, Dudley SC Jr, Beshai JF, Langberg JJ (2006). Lipid-lowering drug use is associated with reduced prevalence of atrial fibrillation in patients with left ventricular systolic dysfunction. *Heart Rhythm 3*, 881–886.

# CARDIOLOGY—IN-DEPTH LOOK: ADVANCES IN ANTIPLATELET THERAPY

Karin Mauer and Jorge F. Saucedo
*Department of Medicine, Cardiovascular Section, University of Oklahoma Health Sciences Center, Oklahoma City, Oklahoma*

## 1 PLATELET PHYSIOLOGY IN CARDIOVASCULAR DISEASE

Platelets have a central role in the pathophysiology of atherosclerotic vascular disease. The interaction of platelets, atherosclerotic lesions, and the coagulation system is crucial in the transition from stable coronary artery disease (CAD) to unstable coronary syndromes such as unstable angina, non-ST-elevation myocardial infarction (NSTEMI), and ST-elevation infarction (STEMI) [1]. During acute coronary syndromes (ACS) and percutaneous coronary intervention (PCI), there is disruption of the natural protection provided by the endothelium with exposure to the subendothelial matrix. The hemodynamic and biochemical environment of circulating platelets changes, inducing the aggregation process. Binding of platelet surface receptors to von Willebrand's factor and collagen promotes platelet adhesion to the vessel wall, with subsequent platelet activation, granule secretion, and activation of the glycoprotein (GP) IIb/IIIa receptors. Irreversible platelet aggregation and additional recruitment of platelets eventually lead to occlusive thrombus generation. In addition to factors in the subendothelial matrix, thrombin is a primary agonist responsible for platelet activation. Following activation, arachidonic acid (AA) is released from membrane phospholipids and converted to thromboxane (Tx)A2 by cyclooxygenase (COX)-1 and Tx synthase. Secreted TxA2 binds to specific receptors on the platelet membrane surface activating surrounding platelets as a positive feedback mechanism. Once activated, adenosine diphosphate (ADP) is released from dense granules in platelets, which amplifies the platelet activation and recruitment phenomenon. The binding of ADP to the P2Y12 receptor triggers intracellular signaling pathways that result in GPIIb/IIIa receptor activation. In addition, platelet activation is characterized by the expression of surface adhesion molecules, such as P-selectin and CD40 ligand, and aggregation of platelets with leukocytes, which contributes to inflammation and amplification of thrombin generation [2, 3].

The binding of fibrinogen to the platelet surface via the calcium-dependent GPIIb/IIIa complex on the platelet surface is necessary for platelet aggregation induced by any agonist. The aggregation mechanism reflects the importance of a rapid and effective response to hemorrhage, which explains the high density of GPIIb/IIIa on the surface of platelets and the high concentration of circulating fibrinogen. The conformational change in GPIIb/IIIa results in its high affinity for the adhesive glycoprotein ligands and is the final common effector for platelet aggregation [4].

There is a well-known interindividual variation in platelet reactivity, which may contribute to the risk for thrombosis in certain individuals but protection in others. Enhanced platelet reactivity (i.e., platelet hyperreactivity) can prospectively identify subjects who are at risk for thrombotic events. The mechanisms responsible for the hyperreactivity are not well

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

understood, and it is not clear whether the underlying platelet hyperreactivity contributes to the thrombotic event or whether the stress of the acute event induces platelet activation [5]. Patients with hypertension, hyperlipidemia, diabetes, ACS, and during PCI have increased expression of platelet activation markers and greater platelet reactivity to agonists. Platelet hyperreactivity despite combined antiplatelet therapy has been linked to stent thrombosis, restenosis, inflammation, myocardial infarction (MI), and other ischemic events after PCI [2].

## 2 THERAPEUTIC TARGETS FOR PLATELET INHIBITION

Platelets have multiple functions that are complex and not limited to hemostasis. Understanding of platelet pathophysiology continues to grow with the recognition of the role platelets play in acute inflammation and in chronic inflammatory processes, such as atherosclerosis [6]. Platelets are primarily involved in thrombus formation, and the rationale for using antiplatelet agents is to decrease platelet activation, detain platelet procoagulant effect, block platelet-mediated inflammation, and favor platelet disaggregation to overall facilitate reperfusion of occluded arteries [2, 4].

Antagonism of the platelet activation/aggregation cascade is a crucial target in the management of cardiovascular patients, and the goal of antiplatelet therapy is to prevent or treat the formation of platelet-dependent thrombus [7]. No single signaling pathway involved in platelet activation is accountable for all thrombotic complications. Therefore, a single agent directed against a specific receptor or pathway is not expected to overcome all thrombotic complications [2].

Both the physiological and the pathological aggregation mechanism is the same, making it difficult to separate the therapeutic from the harmful effects of antiplatelet drugs. In clinical practice, a favorable balance between the beneficial and detrimental effects of antiplatelet therapy needs to be achieved by using these agents in those patients whose thrombotic risk outweighs their risk of bleeding complications and by administration of the minimal therapeutic dose possible [4].

## 3 THROMBOXANE A2 (TXA2) INHIBITORS

### 3.1 Aspirin

Aspirin is a synthetic compound with antipyretic, analgesic, anti-inflammatory, and antiplatelet properties that binds irreversibly to the active site of COX in platelets through acetylation. COX inactivation interferes with prostaglandin biosynthesis and results in reduced production of TxA2. Regular low doses of aspirin (e.g., 81 mg/day) will usually result in greater than 95% inhibition of TxA2 generation [8].

The use of aspirin is well established as a strategy for primary and secondary prevention of cardiovascular events in different populations [2] (see Table 1). The first trial to report a benefit in using aspirin among healthy males over 40 years of age was the Physicians' Health Study with a 44% decrease in first MI during 5 years of follow-up [9, 10]. A meta-analysis that included the six primary prevention trials using aspirin showed that this agent decreases the risk of total coronary heart disease, nonfatal MI, and total cardiovascular events, without affecting the incidence of stroke or cardiovascular mortality [11]. A gender-specific meta-analysis of randomized controlled trials confirmed that for women and men, aspirin therapy reduces the risk of cardiovascular events. However, aspirin significantly increased the risk of bleeding to a similar degree among both genders [12]. Therefore, the use of aspirin for primary prevention is controversial, but since the net benefit of aspirin is greater with increasing cardiovascular risk, the U.S. Preventive Services Task Force recommends aspirin for primary prevention in individuals who are at increased risk for coronary heart disease [13].

In healthy population, gender may play an important role in the response to aspirin. For both genders aspirin therapy reduces the risk of overall cardiovascular events with a similar increase in bleeding. However, the beneficial effect of aspirin is by reducing the risk of ischemic stroke in women and the risk of MI in men [12]. The Women's Health Study showed that among healthy women aspirin lowers the risk of stroke without affecting the risk of MI or death from cardiovascular causes [14].

Secondary prevention of cardiovascular events with chronic aspirin therapy has been demonstrated in patients with established atherosclerotic disease. The benefits of antiplatelet therapy were first demonstrated in studies designed to determine the effects of aspirin on cardiovascular risk in various high-risk patient groups. One example is ISIS 2 (International Study of Infarct Survival 2), which reported a reduction in vascular mortality in the setting of acute MI similar to the benefit achieved with fibrinolytic therapy using streptokinase [15], and the early survival advantages produced by fibrinolytic therapy and aspirin started in acute MI were maintained for at least 10 years [16]. This and other studies have set the foundation for recommending aspirin as the basis of antiplatelet therapy in all high cardiovascular risk patients unless contraindicated [17].

Laboratory aspirin resistance refers to less suppression of TxA2 production by aspirin than expected,

**TABLE 1 Aspirin and Clopidogrel in Clinical Trials**[a]

| Trial | Description | *n* | Main Conclusions | Reference |
|---|---|---|---|---|
| Women's Health Study | ASA every other day vs. placebo in primary prevention of CVD | 39876 | No reduction of primary endpoint (CV death, MI, or stroke) in overall population<br>Reduction of CV events in women older than 65 years<br>No benefit in CV death or MI<br>Reduction in risk of ischemic stroke<br>Nonsignificant increase in hemorrhagic stroke | 14 |
| STIMS | Long-term ticlopidine in patients with intermittent claudication | 687 | Reduction of MI, stroke, and transitory ischemic attacks<br>Reduction in mortality | 28 |
| STARS | Ticlopidine and ASA vs. ASA alone in patients with coronary stenting | 1965 | Reduction of stent thrombosis<br>Increased bleeding | 30, 31 |
| CLASSICS | Ticlopidine vs. clopidogrel (plus ASA) in patients after coronary stenting | 1020 | Clopidogrel group has reduced major peripheral or bleeding complications, neutropenia, thrombocytopenia, or early discontinuation of drug<br>Comparable efficacy with regard to cardiac events | 33 |
| CAPRIE | Clopidogrel vs. ASA in secondary prevention of CVD | 19185 | Clopidogrel group has reduced vascular death, MI, or stroke<br>No difference in safety (bleeding)<br>Clopidogrel has even greater efficacy in diabetic subgroup, prior CABG, prior MI, or ischemic stroke | 35–37, 173–175 |
| CURE | Clopidogrel and ASA vs. ASA for secondary prevention of CVD in patients with NSTEMI | 12562 | Dual therapy further decreases coronary events compared with ASA alone at 1 year | 39, 41, 176 |
| PCI-CURE and CREDO | Pretreatment with clopidogrel and ASA vs. ASA followed by long-term therapy in patients undergoing PCI | 4774 | Long-term dual-antiplatelet therapy beneficial in reducing ischemic events at 1 year<br>Reduced ischemic events at 28 days if 300 mg LD received more than 6 h before PCI | 42–44 |
| COMMIT | Clopidogrel in addition to ASA in patients with STEMI | 45852 | Adding clopidogrel to aspirin and other standard treatments (such as fibrinolytic therapy) reduces mortality and major vascular events | 45 |
| CLARITY-TIMI 28 | Clopidogrel vs. placebo in patients with STEMI and fibrinolytic therapy | 3491 | Addition of clopidogrel improves the patency rate of the infarct-related artery and reduces ischemic complications<br>Clopidogrel pretreatment reduces the incidence of cardiovascular death or ischemic complications both before and after PCI<br>Benefit seen in the overall trial population in the risk of cardiovascular death, recurrent MI, or recurrent ischemia requiring urgent revascularization at 30 days—among patients undergoing CABG, there is a trend toward reduction of risk<br>No excess bleeding in the clopidogrel group<br>Clopidogrel appears to improve late coronary patency and clinical outcomes by preventing reocclusion of open arteries rather than by facilitating early reperfusion | 46, 47, 177, 178 |

(*Continued*)

**TABLE 1** ***Continued***

| Trial | Description | *n* | Main Conclusions | Reference |
|---|---|---|---|---|
| CHARISMA | Clopidogrel and ASA combination therapy in symptomatic (prior CVD) and asymptomatic (no prior CVD) populations | 15603 | Excess of all-cause mortality, and early and sustained risk of severe bleeding within the asymptomatic group<br>No significant effect on cardiovascular mortality in the symptomatic group<br>Do not support use of dual-antiplatelet therapy with clopidogrel and ASA in primary prevention population<br>Patients with documented prior MI, ischemic stroke, or symptomatic PAD benefit from dual antiplatelet therapy | 48, 179, 180 |
| ALBION | Clopidogrel LD 300, 600, or 900 mg plus ASA in low-to-moderate risk patients with NSTEMI | 103 | Clopidogrel LD greater than 300 mg provide a faster onset of action and greater reductions in platelet activation during the first 24 h<br>900-mg LD may induce greater antiplatelet effect than 600 mg, when compared with the standard 300-mg regimen | 51 |
| ISAR-CHOICE | Clopidogrel 300, 600, or 900 mg | 60 | Greater platelet inhibition with 600 and 900 mg LD<br>Similar platelet inhibition and active metabolite levels with 600 and 900 mg LD (there may be saturable absorption and/or metabolism) | 181 |
| ARMYDA 2 | Clopidogrel LD 300 or 600 mg in patients before PCI | 255 | Lower rates of periprocedural MI with 600 mg LD<br>No excess bleeding in the higher LD group | 49 |
| CURRENT OASIS 7 | Clopidogrel higher LD and early MD in patients with unstable angina and NSTEMI with planned early invasive therapy | NA | Ongoing (no results available) | NA |
| OPTIMUS | Clopidogrel 150 mg MD compared with 75 mg MD in patients with DM with suboptimal response to standard dose | 40 | Greater platelet inhibition with 150 mg MD<br>Persistent variability of response and presence of hyporesponsiveness<br>No clinical outcomes | 167 |
| ISAR-CHOICE 2 | Clopidogrel 150 mg MD compared with 75 mg MD in patients after PCI | 60 | Greater platelet inhibition with 150 mg MD<br>Persistent variability of response and presence of hyporesponsiveness<br>No clinical outcomes | 182 |

[a] *n*, number of subjects in trial; ASA, aspirin; CV, cardiovascular; CVD, cardiovascular disease; STEMI, ST-elevation myocardial infarction; LD, loading dose; MD, maintenance dose; DM, diabetes mellitus; MI, myocardial infarct; CABG, coronary artery bypass graft; NA, not available.

which is associated to the lack of an anticipated effect of the drug on a laboratory measure of platelet function [18]. Several COX-specific and nonspecific assays have shown a prevalence of aspirin nonresponse between less than 1% and up to 54% in different populations with cardiovascular disease [2]. There appears to be a clinical relevance of response variability to antiplatelet treatment since recent meta-analysis show that patients who are classified as nonresponsive to aspirin are at greater long-term risk of clinically important cardiovascular morbidity when compared to individuals who are sensitive to aspirin [19, 20]. Many tests have been used to assess aspirin resistance (e.g., arachidonic-acid- and ADP-induced light transmission aggregometry, VerifyNow aspirin assay, thromboelastography with platelet mapping, flow cytometry using arachidonic acid

stimulation, and serum and urinary thromboxane among others). Tests are ideally performed before and after drug administration. Even when some of these tests have been reported to be predictive of adverse clinical events, the majority of these studies are small and statistically underpowered to determine if each test can reliably predict the small number of adverse outcomes that were observed [8]. The assessment of aspirin nonresponse is highly dependent on the assay used and higher estimates of nonresponse are observed with methods that do not use arachidonic acid as the stimulus [21].

Several potential mechanisms could be considered responsible for the variability in aspirin response, such as decreased bioavailability, interference with other drugs, accelerated platelet turnover, and COX-1 variants that may be less responsive to aspirin inhibition [22]. There are also several potential reasons why aspirin may not prevent recurrent vascular events. Noncompliance has been reported in up to 60% of patients with cardiovascular disease; comorbidities like hypercholesterolemia, diabetes mellitus, or unstable angina and smoking may stimulate the production of isoprostanes, which may enhance the response of platelets to other agonists through COX-independent TxA2 biosynthesis; common risk factors such as hypercholesterolemia may affect modulation of thrombin in response to aspirin and increased oxidative stress associated with hyperglycemia may limit the effect of antiplatelet therapy [23].

There is controversy about whether higher doses of aspirin reduce hyporesponsiveness since higher doses of aspirin (650–1500 mg/day) are not superior compared to lower doses (50–100 mg/day) in the reduction in the risk of vascular death, MI, or stroke [23, 24].

### 3.2 Other TxA2 Inhibitors

There are other drugs that also target the TxA2 pathway in platelets, but its use has been more limited. For example, the TxA2 receptor and prostaglandin endoperoxide receptor antagonist ridogrel in RAPT (Ridogrel versus Aspirin Patency Trial) did not have any apparent advantages over aspirin in patients with acute MI [25].

There are other compounds under investigation. The nitric-oxide-releasing aspirin, NCX-4016, has shown beneficial effects in experimental models of restenosis and diabetic vein grafts. The TxA2 receptor antagonist S18886 reversibly inhibits both TxA2 and other eicosanoids not affected by aspirin. In animal models, this compound prevented atherogenesis and caused plaque regression, and single dose in patients with CAD improved endothelial function [26].

## 4 ADP RECEPTOR ANTAGONISTS: THIENOPYRIDINES

Platelet function depends on a complex network of signaling pathways that lead to the common goal of platelet activation and aggregation. Aspirin acts by blocking COX and thereby reducing levels of TxA2, but it does not prevent platelet activation by TxA2-independent pathways. This sets the rationale for combining aspirin with other antiplatelet agents that act by blocking complementary mechanisms [17]. There are studies performed prior to the appearance of clopidogrel that demonstrate a synergistic benefit of thienopyridine and aspirin therapy. These studies confirm the superiority of dual antiplatelet therapy over aspirin alone or aspirin with warfarin in preventing stent thrombosis.

### 4.1 Ticlopidine

Ticlopidine is a prodrug and needs to be metabolized through cytochrome P450 in the liver to an active metabolite that specifically and irreversibly blocks the platelet ADP receptor. It is the first P2Y12 antagonist approved by the Food and Drug Administration (FDA) and is given orally twice a day.

Results from numerous large studies showed that patients with prior transient ischemic attacks or stroke, and those with peripheral arterial disease or ischemic heart disease, benefit from using ticlopidine. Ticlopidine reduced the incidence of stroke, MI, or vascular death and was superior to placebo and aspirin in patients with recent stroke, transient ischemic attacks, or intermittent claudication. Ticlopidine also appeared to reduce anginal pain [27]. For example, the Swedish Ticlopidine Multicentre Study (STIMS) reported that morbidity and mortality from cardio- and cerebrovascular disease in patients with intermittent claudication could be reduced by long-term treatment with ticlopidine [28], and the risk of nonfatal myocardial infarction was reduced in patients with unstable angina treated with ticlopidine [29]. The STARS (Stent Anti-thrombotic Regimen Study) study showed that treatment with aspirin and ticlopidine compared with aspirin alone and a combination of aspirin and warfarin resulted in a lower rate of stent thrombosis, although there were more hemorrhagic complications than with aspirin alone [30]. The STARS registry showed that the acute clinical results of suboptimal stent deployment was acceptable when treatment with ticlopidine was used, even when it was not quite as good as with optimal stenting and similar drug therapy [31].

In large clinical trials, patients treated with ticlopidine had a higher incidence of adverse effects than placebo, but most did not require withdrawal of

treatment. Gastrointestinal symptoms (particularly diarrhea) were common, and other adverse effects were skin rash, hemorrhagic disorders, and hematological effects. These required careful monitoring of patients during the initial weeks of therapy [27]. Clopidogrel is preferred because of the adverse event profile for ticlopidine [2]. In the United States and other countries, ticlopidine has been largely replaced in clinical practice by clopidogrel, which is given in a more convenient daily dose and has a better safety and tolerability profile (e.g., less neutropenia and lower incidence of thrombotic thrombocytopenic purpura) [7].

### 4.2 Clopidogrel

Clopidogrel is also a prodrug that is metabolized by cytochrome P450 in the liver to an active metabolite that specifically and irreversibly blocks the platelet ADP receptor P2Y12 [8]. When no loading dose is administered, 75 mg of daily clopidogrel causes some level of platelet inhibition as early as 2 h after treatment initiation, but maximal inhibition is not achieved for 3–7 days. This delay can be considerably decreased by using a standard 300-mg loading dose. The time to achieve the maximal effect is between 2 and 24 h after administration of the loading dose. With a 600-mg loading dose, full antiplatelet effect takes 2–4 h to build up. A higher dose of 900 mg causes marginal increases in the magnitude and speed of platelet inhibition with no significant differences when compared to 600 mg [17].

In patients who have undergone PCI, a maintenance dose of 75 mg daily clopidogrel should be given for at least 1 month after bare metal stent implantation unless the patient is at increased risk for bleeding, or 3–6 months after drug eluting stent implantation, and ideally up to 12 months in patients who are not at high risk of bleeding [32]. The active clopidogrel metabolite is pharmacodynamically similar to aspirin, and it causes inhibition of platelet function by repeated daily administration of low doses. Platelet function returns to normal 7 days after the last dose. The cumulative inhibitory effect and the slow rate of recovery of platelet activity are consistent with permanent defects in platelet proteins, which can only be replaced as a function of platelet turnover. Therefore, only once daily administration is needed despite its short half-life of 8 h in human circulation [4].

There is abundant clinical evidence that supports the use of clopidogrel in the treatment of several cardiovascular conditions, especially acute coronary syndromes [26] (see Table 1).

CLASSICS (Clopidogrel Aspirin Stent International Cooperative Study) showed that the tolerability of clopidogrel was superior to that of ticlopidine, while having comparable efficacy with regard to cardiac events after successful coronary stenting [33]. A meta-analysis suggested that using clopidogrel decreases major cardiac events and mortality 30 days after coronary stenting as compared to ticlopidine [34]. Therefore, clopidogrel therapy has mainly replaced ticlopidine in PCI [26].

Based on results from CAPRIE (Clopidogrel versus Aspirin in Patients at Risk of Ischemic Events) showing that clopidogrel was marginally superior to aspirin for secondary prevention in patients with recent MI, ischemic stroke, or peripheral arterial disease (PAD) [35–37], clopidogrel has become the alternative antiplatelet agent for patients with aspirin intolerance [26]. However, because there was a small effect size and the statistical significance achieved in a large sample was marginal, the U.S. FDA did not grant a superiority claim over aspirin but assigned a claim of efficacy relative to placebo. The overall safety and tolerability profile of clopidogrel appeared to be as good as that of medium-dose aspirin [38].

CURE (Clopidogrel in Unstable Angina to Prevent Recurrent Events) confirmed that clopidogrel was effective in the acute management of NSTEMI [39] regardless of patient risk profile [40], and the therapeutic benefit was sustained after 1 year [41]. However, an increased bleeding risk was observed in this study in patients receiving clopidogrel. PCI-CURE analyzed the data from patients with non-ST-elevation acute coronary syndrome undergoing PCI in the CURE study and showed that pretreatment and long-term administration of clopidogrel in patients that undergo PCI was advantageous [42]. CREDO (Clopidogrel for the Reduction of Events During Observation) suggested that long-term clopidogrel therapy following PCI significantly reduced the risk of adverse ischemic events. A loading dose of clopidogrel given at least 3 h before the procedure did not reduce events at 28 days, but subgroup analyses suggested that longer intervals between the loading dose and the PCI may reduce events. The risk of major bleeding increased but not significantly with the use of clopidogrel [43, 44].

Clopidogrel in addition to standard therapy for the management of acute MI provided further reduction in mortality and ischemic events, as shown in COMMIT (Clopidogrel and Metoprolol in Myocardial Infarction Trial) [45] and CLARITY-TIMI 28 (Clopidogrel as Adjunctive Reperfusion Therapy—Thrombolysis in Myocardial Infarction 28), supporting the early use of clopidogrel in STEMI and the strategy of routine clopidogrel pretreatment in patients undergoing PCI [46, 47]. The synergistic benefits of combination therapy seen with clopidogrel and aspirin in acute coronary syndromes and after PCI led to the evaluation of long-term dual antiplatelet treatment in high-risk patients for primary and secondary prevention. The CHARISMA

(Clopidogrel for High Atherothrombotic Risk and Ischemic Stabilization Management and Avoidance) trial showed no benefit with an increase in moderate bleeding in the primary prevention subgroup, but there was reduction of death, MI, and stroke, the primary endpoint of the study, in patients with a prior MI, stroke, or symptomatic PAD [48]. Results suggested there was a relationship between treatment benefit and underlying cardiovascular risk. In high-risk patients with ACS or in those undergoing PCI (both conditions characterized by platelet hyperreactivity), the efficacy benefit outweighed the bleeding risk. On the contrary, bleeding risk exceeded benefit in low-risk patients who were not undergoing PCI [38].

Results from ARMYDA 2 (Anti-platelet Therapy for Reduction of Myocardial Damage during Angioplasty) favored the use of a higher loading dose of 600 mg of clopidogrel in patients undergoing PCI with a benefit in clinical outcomes [49]. A recent meta-analysis showed that clopidogrel loading dose greater than 300 mg reduced early ischemic events in patients scheduled for PCI with a possibility that the greater the underlying risk, the greater the favorable impact of a high loading dose [50]. The ongoing CURRENT-OASIS 7 (Clopidogrel Optimal Loading Dose Usage to Reduce Recurrent Events/Optimal Anti-platelet Strategy for Interventions 7) trial will evaluate the clinical effect of higher loading and maintenance dose of clopidogrel in patients with unstable angina and NSTEMI [26].

Variability of response to clopidogrel has led to changes in clinical practice and use of higher doses of clopidogrel, especially in patients undergoing PCI. The evidence that higher dose of clopidogrel maintenance therapy provides more effective platelet inhibition is supported by the ALBION (Assessment of the Best Loading Dose of Clopidogrel to Blunt Platelet Activation, Inflammation and Ongoing Necrosis) study. In this trial, greater loading doses were associated with significantly greater platelet inhibition when compared with the standard 300-mg loading dose. There was a dose–effect relationship for onset of action, maximal plateau, platelet inhibition at 24 h, and rates of low inhibition of platelet aggregation (less than 10% at 6 h). Similar but nonsignificant trends were observed for troponin release and major adverse cardiac events. Bleeding rates were similar among groups. The investigators concluded that in low- to moderate-risk patients with NSTEMI, clopidogrel loading doses greater than 300 mg provide a faster onset of action, a higher inhibition of platelet aggregation, and greater reductions in platelet activation during the first 24 h [51].

Current PCI guidelines provide a class IIb recommendation for patients in whom subacute stent thrombosis may be catastrophic or lethal to consider platelet aggregation studies and increasing the daily maintenance dose of clopidogrel from 75 to 150 mg if less than 50% inhibition of platelet aggregation is demonstrated [52].

All these studies indicate that clopidogrel plus aspirin is better in high-risk patients, such as those with NSTEMI, STEMI, and stenting, than aspirin alone. Long-term benefit appears to be associated with reductions in nonfatal MI, with little impact on death or stroke, while major bleeding complications are increased.

In patients with stable cardiovascular disease there seems to be no long-term benefit but an increased bleeding risk associated with chronic use of clopidogrel, and therefore, it is not recommended in such patients. However, in high-risk patients with NSTEMI and at low risk for bleeding, treatment with clopidogrel should be given for at least 1 month and up to 9 months. In patients with STEMI and fibrinolytic treatment at least 1 month of treatment is recommended, and there are no direct studies in patients with primary PCI [38].

The optimal timing and dose of clopidogrel pretreatment before PCI has not been established. Even though a nonrandomized subgroup analysis of CURE reflected that pretreatment was associated with favorable outcomes [38], in ISAR-REACT (Intracoronary Stenting and Antithrombotic Regimen—Rapid Early Action for Coronary Treatment), there were no treatment differences in low- to intermediate-risk patients receiving a 600-mg clopidogrel loading dose 2–3 h before PCI when compared to those receiving the same loading dose 12 h before the procedure [53]. The American College of Cardiology (ACC)/American Heart Association (AHA) guidelines recommend administration of 300 mg of clopidogrel at least 6 h before elective PCI, while the European Society of Cardiology guidelines recommend 600 mg clopidogrel at least 2–6 h before PCI. Further clinical trials are needed regarding dose, time course of pretreatment, and associated benefit of clopidogrel in the management of ACS patients and those undergoing elective PCI.

Despite the clinical benefit achieved with the adjunct of clopidogrel in patients with ACS and undergoing PCI, a considerable number of patients continue to have cardiovascular events. This has been attributed in part to the fact that some individuals may have poor response to clopidogrel treatment and have enhanced platelet reactivity, which is pivotal for the development of atherothrombotic complications [54]. Currently, no definition of true clopidogrel nonresponse has been universally accepted and different researchers have used various platelet function tests and assigned a variety of empiric cut-off values in the laboratory to define it [55].

Laboratory responses to P2Y12 inhibitors are largely based upon monitoring ADP-stimulated responses,

ideally before and after drug administration [8]. Maximum platelet aggregation induced by ADP and measured by light transmission aggregometry (LTA) is the most commonly used laboratory method to evaluate responsiveness to clopidogrel and is considered the gold standard [2]. LTA using 5 or 20 μmol/L ADP has been used to arbitrarily classify patients based upon the change in aggregation (delta aggregation) at baseline and after drug administration. Nonresponders have been defined with a delta aggregation of less than 10%. Several studies show that there is substantial variation in patient response to clopidogrel and up to 30% of patients may be nonresponders [8]. The phosphorylation state of vasodilator-stimulated phosphoprotein (VASP) can be measured by flow cytometry and is a specific marker of residual P2Y12 receptor activity in patients treated with a P2Y12 inhibitor, such as clopidogrel [8].

Some studies in patients undergoing percutaneous revascularization have suggested a relation between high posttreatment platelet reactivity, clopidogrel nonresponsiveness, and increased ischemic events, including stent thrombosis [56]. However, justification of clopidogrel nonresponse based on clinical outcomes can be misleading. A number of studies have shown that patients that do not respond to clopidogrel are at an increased risk of vascular events, especially subacute stent thrombosis (SAT), but the risk of SAT is 1–2%, while between 5 and 60% of patients are found to be clopidogrel nonresponders [55]. There are limited data that link clopidogrel nonresponsiveness to subsequent thrombotic events. Several small studies have suggested that patients with higher residual platelet reactivity after the initiation of clopidogrel therapy have a greater frequency of subsequent major adverse cardiovascular events. Higher doses of clopidogrel or other antiplatelet agents may, therefore, be beneficial in these patients [24, 57]. The PREPARE (Platelet Reactivity in Patients and Recurrent Events) study showed that patients with recurrent ischemic events despite dual antiplatelet therapy after elective stenting had higher platelet reactivity than patients without ischemic events [58]. Higher levels of platelet aggregation, active GPIIb/IIIa expression, and the P2Y12 reactivity ratio measured by VASP phosphorylation were reported in patients with stent thrombosis in the CREST (Clopidogrel Effect on Platelet Reactivity in Patients with Stent Thrombosis) study, suggesting inadequate P2Y12 receptor inhibition [59]. In CLEAR PLATELETS (Clopidogrel Loading with Eptifibatide to Arrest PLATELET Reactivity Study) a higher loading dose of clopidogrel in patients undergoing elective PCI achieved greater early platelet inhibition compared to the standard dose, and this seemed to be related with a decrease in myocardial necrosis and inflammation markers [60, 61]. Those patients who were less responsive to aspirin seemed also more frequently hyporesponsive to clopidogrel [62, 63], but further evidence is needed since other studies do not report an overlap in hyporesponsiveness with these agents [24].

Clopidogrel nonresponsiveness seems to depend on dose, and higher loading dose is associated with a narrower response profile. A higher loading dose with 600 mg of clopidogrel accelerates the onset of the antiplatelet effect and enhances the magnitude of platelet inhibition when compared to either no loading dose or the standard 300-mg dose. The higher dose also reduces the prevalence of nonresponse to clopidogrel, but there is limited clinical-trial-based evidence to support its safety and greater efficacy [57]. For example, in ARMYDA 2 it was not determined if subjects treated with 600 mg had a lower event rate because the higher dose was more rapidly acting or because it caused greater platelet inhibition and less hyporesponsiveness compared with the standard loading dose [24]. A small, prospective randomized trial presented at the American College of Cardiology Scientific Sessions 2008/i2 Summit-SCAI Annual Meeting suggests that titration of the clopidogrel dose upward based on the results of a platelet reactivity test that measures VASP can reduce the risk of clinical events over the short term.

In contrast, MATCH (Management of Atherothrombosis with Clopidogrel in High-Risk Patients) and CHARISMA showed that the lack of efficacy in patients after stroke and for primary prevention was not caused by low response to clopidogrel, and significant increases in bleeding events suggested adequate or even excessive platelet inhibition with clopidogrel [64].

The mechanisms leading to variability in clopidogrel responsiveness are not fully elucidated and, similarly to aspirin, are likely multifactorial. High baseline platelet reactivity, poor compliance, inappropriate dosing, upregulation of the P2Y12 pathway, and upregulation of P2Y12-independent pathways, differences in individual absorption and/or metabolism of clopidogrel as well as different levels of its active metabolite may all contribute to variations in the level of antiplatelet effects of clopidogrel seen in different patients. For example, drugs that are substrates or inhibit the cytochrome P450 (CYP) isoenzyme 3A4 can potentially interfere with the conversion of clopidogrel into its active metabolite, leading to reduced antiplatelet effects, and genetic polymorphisms of this and other CYP enzymes have been shown to modulate individual response to clopidogrel [54].

Before a patient is considered as a clopidogrel nonresponder, nonplatelet factors should be evaluated, such

as treatment compliance (about 20% of patients are noncompliant with treatment), suboptimal stent deployment (which can increase the risk of SAT despite adequate anticoagulation), drug–drug interactions, or other prothrombotic states [55]. The current therapeutic alternative for treatment of patients with poor clopidogrel response remains limited. Response variability to clopidogrel has served as the basis for the development of new P2Y12 inhibitors with superior pharmacodynamic profiles (e.g., prasugrel and AZD 6140) [56]. Only a well-designed, large, randomized study with a hard outcome or survival endpoint, supported by serial platelet assessment and measurement of clopidogrel metabolites, will determine if clopidogrel nonresponse is a real hazard as suggested by platelet biomarkers and function tests [64].

## 5 GLYCOPROTEIN IIB/IIIA INHIBITORS

Given the redundancy of signaling pathways that lead to platelet aggregation, the clinical efficacy of aspirin, ticlopidine, and clopidogrel is only partial. These drugs inhibit TXA2-mediated or ADP-mediated platelet aggregation but leave the activity induced by other platelet agonists (e.g., thrombin) mostly unchanged. The expression of active GPIIb/IIIa on the platelet surface is the final common pathway of platelet aggregation regardless of the initiating stimulus. Therefore, this molecule became a main target of antiplatelet drugs. GPIIb/IIIa receptors blockade appears as a particularly desirable therapeutic strategy because: (1) monoclonal antibodies to GPIIb/IIIa are more potent inhibitors of platelet function than aspirin; (2) GPIIb/IIIa is platelet specific; (3) GPIIb/IIIa antagonism leaves platelet adhesion largely intact, contributing to hemostasis without causing thrombotic damage; and (4) the hemorrhagic diathesis produced by the inherited deficiency of GPIIb/IIIa receptors (Glanzmann thrombasthenia) only rarely produces spontaneous brain hemorrhage, one of the most feared complications of antiplatelet therapy [4].

Abciximab is a monoclonal antibody that has high affinity and a slow rate of dissociation from GPIIb/IIIa, and the small molecules eptifibatide and tirofiban have a much more rapid rate of dissociation, with an off time of 10–15 s. The circulating pool of abciximab is predominantly associated with platelets, while maintenance of a constant concentration of tirofiban and eptifibatide in the blood is needed to achieve and sustain their inhibitory effects. The GPIIb/IIIa affinity of abciximab and tirofiban is greater than that of eptifibatide, and greater concentration of eptifibatide in blood is required for effective inhibition of the binding of fibrinogen to the activated conformer of GPIIb/IIIa [65]. There is clinical evidence that supports the use of GPIIb/IIIa inhibitors in the treatment of several cardiovascular conditions (see Table 2).

### 5.1 Abciximab

Abciximab is a Fab fragment of the 7E3 antibody that has high affinity and a slow rate of dissociation from GPIIb/IIIa. Abciximab has a plasma half-life of about 10 min, with a second-phase half-life of about 30 min. However, its effects on platelet function can be seen for up to 48 h after the infusion has been stopped [65].

In the 1990s, the placebo-controlled trials EPIC (Evaluation of c7E3 for Prevention of Ischemic Complications) [66, 67], EPILOG (Evaluation of PTCA to Improve Long-Term Outcome with Abciximab GPIIb-IIIa Blockade) [68–71], and EPISTENT (Evaluation of Platelet Inhibition in Stenting) [72] included over 7000 subjects from a broad group of PCI patients and consistently documented the efficacy of abciximab in reducing periprocedural ischemic complications of death, MI, and the need for repeat revascularization, with sustained benefits for at least 3 years of follow-up.

The CAPTURE (C7E3 Anti-Platelet Therapy in Unstable Refractory Angina) trial confirmed that in patients with refractory angina undergoing PCI, abciximab reduced the risk of death, MI, or repeat revascularization at 30 days. This benefit was absent at 6 months, probably due to differences in dosing regimes when compared to previous trials with abciximab where the drug was continued for 12 h rather than only 1 h after the procedure. On the other hand, CAPTURE showed that use of abciximab for 18–24 h prior to PCI decreased the incidence of MI before the procedure, suggesting that patients with unstable angina and non-Q-wave MI can benefit from GPIIb/IIIa antagonism regardless of PCI [73, 74]. This was addressed in GUSTO IV (Global Use of Strategies to Open Occluded Coronary Arteries IV), which evaluated abciximab used in medical management of ACS (excluding patients with ST-segment elevation MI) [75, 76]. The absence of any efficacy was not anticipated since previous trials using small molecules, such as PURSUIT (Platelet IIb-IIIa in Unstable Angina: Receptor Suppression Using Integrilin Therapy) [77, 78], PRISM (Platelet Receptor Inhibition in Ischemic Syndrome Management) [79], and PRISM-Plus (Platelet Receptor Inhibition in Ischemic Syndrome Management in Patients Limited by Unstable Angina) [80] showed benefit in this setting. The practical outcome of these trials was a more limited use of abciximab in patients scheduled for PCI and a restricted application in the medical management of ACS.

**TABLE 2 GPIIb/IIIa Inhibitors in Clinical Trials[a]**

| Trial | Description | *n* | Main Conclusions | Reference |
|---|---|---|---|---|
| EPIC | Abciximab vs. placebo for prevention of ischemic complications in high-risk PCI (with bailout stenting) | 2099 | Reduced risk of death, MI, or recurrent ischemia at 30 days, 6 months, and 3 years, but associated with more major and minor bleeding<br>Dose-dependent effect<br>Reduced incidence of clinical restenosis at 6 months | 66, 67 |
| EPILOG | Abciximab vs. placebo for prevention of ischemic complications in PCI (all patients, regardless of risk, with bailout stenting) | 2792 | Reduced risk of death, MI, or recurrent ischemia at 30 days and 6 months<br>No increase in bleeding risk with more conservative use of heparin during procedure, no postprocedure heparin and early sheath removal<br>No reduced incidence of clinical restenosis at 6 months | 68–71 |
| EPISTENT | Abciximab vs. placebo for prevention of ischemic complications in PCI (all patients, regardless of risk, with routine stenting) | 2399 | Periprocedural benefits with abciximab (death and MI) are independent of stenting<br>Abciximab + stent complimentary long-term clinical benefits<br>Overall stenting offers best protection against clinical restenosis<br>In diabetics addition of abciximab further reduces risk of restenosis | 72 |
| IMPACT II | Eptifibatide vs. placebo for prevention of ischemic complications in PCI | 4010 | Reduced risk of death, MI, or recurrent ischemia at 30 days but not 6 months<br>No dose-dependent effect<br>Only 30–50% inhibition of platelet aggregation (i.e., suboptimal dosing) | 99 |
| ESPRIT | Higher dose eptifibatide vs. placebo for prevention of ischemic complications in PCI (all patients, regardless of risk, with routine stenting) | 2604 | Reduced risk of death, MI, or recurrent ischemia at 30 days and 6 months<br>Increase in major bleeding risk despite conservative use of heparin<br>Greater than 90% inhibition of platelet aggregation | 100–103 |
| RESTORE | Tirofiban vs. placebo for prevention of ischemic complications in patients with ACS and elective PCI | 2139 | Reduced risk of death, MI, or recurrent ischemia at 30 days but not 6 months<br>No increase in bleeding risk<br>No reduced incidence of angiographic restenosis at 6 months | 104 |
| TARGET | Tirofiban vs. abciximab for prevention of ischemic complications in PCI | 4809 | Superiority of abciximab over tirofiban at 30 days with no difference in rates of major bleeding<br>Similar level of protection against the composite of death, myocardial infarction, and any target-vessel revascularization at 6 months<br>Similar level of survival benefit at 1 year | 105–107 |
| CAPTURE | Abciximab vs. placebo for prevention of ischemic complications in PCI to treat unstable refractory angina | 1265 | Reduced risk of death, MI, or repeat revascularization at 30 days but not 6 months<br>Reduced risk of MI prior to PCI with pretreatment | 73, 74 |

**TABLE 2** ***Continued***

| Trial | Description | *n* | Main Conclusions | Reference |
|---|---|---|---|---|
| PURSUIT | Eptifibatide vs. placebo for prevention of ischemic complications in medical management of ACS | 10498 | Reduced risk of death and MI at 30 days an 6 months<br>Most of benefit is derived from a decrease in non-Q-wave MI<br>Patients who subsequently underwent PCI were subgroup with most pronounced benefit | 77, 78 |
| PRISM | Tirofiban vs. placebo for prevention of ischemic complications in medical management of ACS | 3232 | Reduced risk of death, MI, or refractory ischemia at 48 h but not 30 days | 79 |
| PRISM-Plus | Tirofiban vs. placebo for prevention of ischemic complications in medical management of ACS | 1915 | Reduced risk of death, MI, or refractory ischemia in tirofiban with heparin treatment arm at 7 days, 30 days, and 6 months<br>Patients who subsequently underwent PCI were subgroup with most pronounced benefit<br>Increased risk of death in tirofiban only treatment arm at 7 days | 80 |
| GUSTO IV | Abciximab vs. placebo for prevention of ischemic complications in medical management of ACS | 7800 | Lack of efficacy<br>No survival benefit at 1 year<br>In patients with low cardiac troponin or elevated C-reactive protein, abciximab was associated with excess mortality | 75, 76 |
| RAPPORT | Adjunctive therapy with abciximab in primary PCI | 483 | Reduced risk of death, reinfarction, or target vessel revascularization at 30 days | 81 |
| ISAR-2, ACE, and ADMIRAL | Adjunctive therapy with abciximab in primary stenting | | Reduced risk of death, reinfarction, or target vessel revascularization at 30 days, 6 months, 1 year and up to 3 years | 82–87 |
| CADILLAC | Adjunctive therapy with abciximab in primary PCI or stenting | 2082 | In primary PCI reduced risk of death, reinfarction, target vessel revascularization, or stroke at 30 days<br>In primary stenting no benefit | 88, 89 |
| ASSENT III and GUSTO V | Abciximab and half-dose fibrinolytics vs. full-dose fibrinolytics | Over 22,000 | Decreased reinfarction rate<br>Increased major bleeding<br>No overall difference in mortality at 1 year | 90–93 |
| ISAR-REACT | Pretreatment with clopidogrel (LD 600 mg) in PCI with routine use of abciximab | 2159 | No additional benefit in reduction of death, MI, or urgent target vessel revascularization at 30 days | 53 |
| FINESSE | Half-dose reteplase/ abciximab vs. abciximab alone in STEMI patients undergoing PCI | 2453 | Improved coronary patency before PCI<br>No significant benefit in short-term outcomes: no differences between<br>treatment arms for the primary composite endpoint at 90 days of all-cause mortality, readmission for heart failure, ventricular fibrillation, or cardiogenic<br>shock<br>No differences in all-cause mortality, complications of MI, or any of the independent components<br>Major and minor bleedings more common in the combination strategy as compared with primary PCI and as compared with the abciximab-only group | 94, 95 |

(*Continued*)

**TABLE 2** ***Continued***

| Trial | Description | *n* | Main Conclusions | Reference |
|---|---|---|---|---|
| EXCITE, SYMPHONY, OPUS-TIMI 16, FROST | Oral IIb/IIIa (xemilofiban, sibrafiban, orbofiban, or lefradiban) in patients with ACS or undergoing PCI | Over 27,000 | No increased efficacy<br>Increased bleeding events<br>Increased mortality or MI | 108, 110 |
| 2nd SYMPHONY | Combination therapy with low-dose sibrafiban and aspirin vs. aspirin or high-dose sibrafiban alone in patients with ACS | 6637 | Terminated early because of results from SYMPHONY<br>Long-term use of sibrafiban associated with increased bleeding complications | 110 |
| BRAVO | Lotrafiban in patients with recent MI, unstable angina, transient ischemic attack, or ischemic stroke, or peripheral vascular disease combined with either cardiovascular or cerebrovascular disease | NA | Terminated early because excess mortality at interim analysis | 110, 183 |

[a]*n*, number of subjects in trial; ASA, aspirin; CV, cardiovascular; CVD, cardiovascular disease; STEMI, ST-elevation myocardial infarction; LD, loading dose; MD, maintenance dose; DM, diabetes mellitus; MI, myocardial infarct; CABG, coronary artery bypass graft; NA, not available.

The benefits of abciximab in primary PCI and primary stenting were confirmed in RAPPORT (ReoPro and Primary PTCA Organization and Randomized Trial) [81] and ISAR 2 (Intracoronary Stenting and Antithrombotic Regimen 2), ACE (Abciximab and Carbostent Evaluation), and ADMIRAL (Abciximab before Direct Angioplasty and Stenting in Myocardial Infarction Regarding Acute and Long-term Follow-up) [82–87]. In contrast, CADILLAC (Controlled Abciximab and Device Investigation to Lower Late Angiographic Complications) showed a benefit of adjunctive abciximab therapy with primary PCI, but none with primary stenting, which may have been associated with thienopyridine use prior to catheterization [88, 89].

ASSENT 3 (Assessment of the Safety of a New Thrombolytic 3) and GUSTO V (Global Use of Strategies to Open Occluded Coronary Arteries V) showed that abciximab used with fibrinolytics did not add benefit to pharmacological reperfusion therapy [90–93]. FINESSE (Facilitated Intervention with Enhanced Reperfusion Speed to Stop Events) showed there was improved coronary patency before PCI with half-dose reteplase/abciximab in STEMI patients, however, without significantly improving short-term outcome [94, 95]. This provided further insight into the role of combination of GPIIb/IIIa inhibitor and thrombolytics in acute MI.

In patients at low to intermediate risk who underwent elective PCI after pretreatment with a high loading dose of clopidogrel, abciximab was not associated with a clinical benefit within the first 30 days [96]. However, ISAR-REACT 2 showed in high-risk patients with NSTEMI undergoing a PCI after pretreatment with 600 mg clopidogrel that adverse events occured less frequently with abciximab, the early benefit was maintained at 1 year after administration, but the benefits provided by abciximab appeared to be confined to patients presenting with an elevated troponin level [97, 98]. Therefore, adjunctive use of clopidogrel and a GPIIb/IIIa inhibitor was effective and safe in high-risk patients with ACS, while pretreatment with a higher loading dose of clopidogrel may be an acceptable alternative to GPIIb/IIIa inhibition in low-risk patients [38]. However, it has been suggested that the use of the GPIIb/IIIa inhibitor abciximab in acute STEMI patients undergoing PCI after pretreatment with a 600-mg loading dose of clopidogrel does not reduce infarct size, according to the Bavarian Reperfusion Alternatives Evaluation-3 (BRAVE-3), study presented at the

American College of Cardiology 2008 Scientific Sessions/i2 Summit-SCAI Annual Meeting.

### 5.2 Eptifibatide

Eptifibatide is a cyclic heptapeptide. The small molecule is a highly specific, reversible, intravenously administered GPIIb/IIIa receptor antagonist that acts at the final common step of the platelet aggregation pathway. The plasma elimination half-life is approximately 2.5 h [65].

IMPACT II (Integrilin to Minimize Platelet Aggregation and Coronary Thrombosis II) showed clinical benefit of eptifibatide when compared to placebo in lowering clinical endpoints at 30 days in patients undergoing PCI. In this study there was a lack of dose–response relationship between the treatment regimen and the clinical outcome, which was likely due to suboptimal inhibition of platelet aggregation [99]. A new dosing regimen for eptifibatide was established to achieve 90% platelet inhibition in more than 90% of the patients (PRIDE study). The new dosing regimen was tested in the ESPRIT (Enhanced Suppression of the Platelet IIb-IIIa Receptor with Integrilin Therapy) trial, which enrolled patients receiving coronary stents. The trial showed that routine use of eptifibatide was associated with a robust reduction of ischemic endpoints. The magnitude of these risk reduction with eptifibatide was similar to that reported in previous trials with abciximab [100–103].

### 5.3 Tirofiban

Tirofiban is a nonpeptide molecule that blocks the GPIIb/IIIa receptor reversibly. Platelet inhibition greater than 90% is attained by the end of the 30-min infusion and tirofiban has a half-life of approximately 2 h [65].

The results of RESTORE (Randomized Efficacy Study of Tirofiban for Outcomes and Restenosis) with tirofiban broadened the spectrum of GPIIb/IIIa antagonists that can achieve reduction of clinical ischemic events after PCI [104]. Even when there are differences in the reported reduction of relative risk with all three GPIIb/IIIa inhibitors, variation in study design, patient population and definitions of clinical endpoints make it difficult to compare these compounds using the information derived from these trials. TARGET (Do Tirofiban and ReoPro Give Similar Efficacy Trial), the first head-to-head comparison in patients undergoing coronary stenting, showed abciximab was superior to tirofiban in reducing ischemic events with no difference in bleeding. These results may be associated to differences in pharmacokinetics, as well as in the binding site and selectivity of the agent [105–107]. The results of TARGET cannot be extrapolated to eptifibatide, and there is not a specific trial comparing this compound to abciximab.

### 5.4 Oral GPIIb/IIIa Inhibitors

In contrast to the efficacy of intravenous IIb/IIIa inhibitors in ACS and as adjunctive therapy in PCI, the studies with oral IIb/IIIa inhibitors have not shown any benefit, but there are increased bleeding events and mortality with long-term use of these agents [108–110].

The decrease of adverse clinical outcomes with xemilofiban in the ORBIT (oral glycoprotein IIb/IIIa receptor blockade to inhibit thrombosis) study was not seen in the larger EXCITE (Evaluation of Xemilofiban in Controlling Thrombotic Events) trial [111]. Comparable negative results were evident with sibrafiban in the SYMPHONY (Sibrafiban versus Aspirin to Yield Maximum Protection from Ischemic Heart Events Postacute Coronary Syndromes) trial [112] and with orbofiban in the prematurely canceled OPUS-TIMI 16 (Orbofiban in Patients with Unstable Syndromes—Thrombolysis in Myocardial Infarction 16) trial [113, 114]. The 2nd SYMPHONY study and BRAVO (Blockage of the GPIIb/IIIa Receptor to Avoid Vascular Occlusion) studies were terminated early because of negative results [110]. A number of potential mechanisms have been suggested for the failure of oral IIb/IIIa inhibitors. Potential explanations may include the presence of partial agonist activity of these compounds, phospholipase A2 polymorphism and other genetic factors, augmented cardiomyocyte apoptosis via caspase-3 activation, or suboptimal dosing with inadequate effects on platelet aggregation [26]. High levels of inhibition are not tolerable for chronic use because of excess bleeding, whereas low levels of inhibition may not be enough to reduce recurrent thrombotic events and in fact increase the risk of thrombotic events by inducing platelet activation and thrombin generation, and differences in the pharmacokinetic profiles of the different agents may be related to outcomes [110].

## 6 NOVEL ADP RECEPTOR ANTAGONISTS

New ADP receptor antagonists are being developed and are currently in advanced stage testing for use in patients with coronary artery disease (see Table 3). Antiplatelet agents with a faster onset of action could be valuable in the treatment of patients destined for PCI [17]. These newer agents appear to have characteristics that may solve the potential problems derived from limitations associated with clopidogrel

**TABLE 3 New ADP Receptor Antagonists in Clinical Trials**

| Trial | Description | *n* | Main Conclusions | Reference |
|---|---|---|---|---|
| JUMBO TIMI 26 | Prasugrel vs. clopidogrel in patients undergoing elective or urgent PCI—dose-ranging safety trial | 904 | Low rates of bleeding in both groups with no difference between groups<br>Favorable trend in incidence of death, MI, stroke, or target vessel thrombosis at 30 days (study not designed to assess efficacy) | 115 |
| TRITON TIMI38 | Prasugrel (60 mg LD and 10 mg MD) compared with standard dose clopidogrel (300 mg LD and 75 mg MD) in patients with ACS undergoing PCI | 13608 | Greater reduction in ischemic events with prasugrel<br>Greater bleeding risk with prasugrel<br>Populations at greater risk of bleeding: history of stroke or TIA, age greater than 75 years, weight less than 60 kg | 116 |
| PRINCIPLE TIMI44 | Prasugrel (60 mg LD and 10 mg MD) compared with high-dose clopidogrel (600 mg LD and 150 mg MD) in patients with elective PCI | 201 | Greater platelet inhibition with prasugrel after LD and MD<br>More rapid onset of antiplatelet effects with prasugrel<br>Less interindividual variance in response and less hyporesponsiveness with prasugrel<br>Bleeding events tended to be higher with prasugrel (not statistically significant)<br>Trial not powered to detect clinical efficacy differences | 117 |
| DISPERSE 2 | AZD6140 (90 or 180 mg) vs. clopidogrel (300 mg LD and 75 mg MD) in patients with NSTEMI | 990 | No difference in major bleeding but an increase in minor bleeding at the higher dose<br>Favorable trend in incidence of MI (study not designed to assess efficacy) | 118 |
| PLATO | AZD6140 vs. clopidogrel in patients with NSTEMI or STEMI | 18000 | Ongoing (no results available) | NA |
| CHAMPION-PCI and -PLATFORM | Cangrelor vs. clopidogrel in patients undergoing PCI | NA | Ongoing (no results available) | NA |

*n*, number of subjects in trial; ASA, aspirin; CV, cardiovascular; CVD, cardiovascular disease; STEMI, ST-elevation myocardial infarction; LD, loading dose; MD, maintenance dose; DM, diabetes mellitus; MI, myocardial infarct; CABG, coronary artery bypass graft; NA, not available.

use, such as the considerable interindividual variability in response, the slow onset of action, and the irreversible nature of the inhibitory effect [24]. The novel compounds include an oral, irreversible thienopyridine agent (prasugrel) and two reversible nonthienopyridine agents (AZD6140 and cangrelor).

### 6.1 Prasugrel

The oral thienopyridine prasugrel (CS-747 or LY640315) is a prodrug, and the active metabolite binds irreversibly to the P2Y12 receptor, resulting in a faster and more potent antiplatelet effect than clopidogrel. Prasugrel must be metabolized by cytochrome-dependent pathways to its active form. However, unlike clopidogrel, prasugrel requires only one cytochrome P450-dependent oxidative step to generate the active metabolite, and this may underlie the faster metabolic conversion and onset of action of prasugrel compared with clopidogrel [17]. Inhibition of ADP-induced platelet aggregation is achieved more rapidly and consistently (with less interindividual variation) than with clopidogrel. Poor responders to clopidogrel are responsive to prasugrel. Several days off therapy are required for full reversal of antiplatelet effects because prasugrel binds irreversibly to the P2Y12 receptor [24].

Phase I and II trials have demonstrated the safety of prasugrel. JUMBO-TIMI 26 (Joint Utilization of Medications to Block Platelets Optimally—Thrombolysis in Myocardial Infarction 26) showed that prasugrel was as safe as clopidogrel in patients undergoing PCI with no differences in the incidence of bleeding [115]. The safe use of prasugrel and a nonsignificant decrease in incidence of ischemic events seen in this safety study has led to other trials designed to assess both efficacy and safety. TRITON-TIMI 38 (Trial to Assess Improvement in Therapeutic Outcomes by Optimizing Platelet Inhibition with Prasugrel—Thrombolysis in Myocardial

Infarction 38) confirmed that prasugrel therapy in patients with moderate- to high-risk acute coronary syndromes scheduled for PCI was associated with a reduced rate of ischemic events, including stent thrombosis, but with an increased risk of major bleeding, including fatal bleeding without affecting overall mortality when compared to the standard dose of clopidogrel [116]. A new analysis from the TRITON-TIMI 38 trial presented at the American College of Cardiology 2008 Scientific Sessions/i2 Summit, SCAI Annual Meeting, showed a consistent reduction in stent thrombosis with prasugrel when compared to standard dose of clopidogrel regardless of stent type or stent thrombosis definition used. This reduction was seen early and late after stent implantation. Although relative reductions in stent thrombosis were overall consistent, the absolute benefit in specific groups of patients at higher risk for stent thrombosis was notable. These groups included people with diabetes, stents of more than 20 mm in length, bifurcation stents, and those with MI before presentation. It is not known if this benefit of prasugrel persists when compared to a high loading dose of clopidogrel. Overall, the results of this trial seem to validate the concept that greater platelet inhibition improves clinical outcomes, by reducing ischemic events. Post hoc analyses identified three subgroups with less clinical efficacy and greater absolute levels of bleeding than the overall cohort. These include patients with a history of stroke or transient ischemic attack before enrollment, elderly patients (age greater than 75 years), and those with a body weight of less than 60 kg. These risk factors have been previously associated with increased risk of adverse outcomes while using antiplatelet or antithrombotic agents. The TIMI 38 study was not powered to detect a reduction in the rate of death from cardiovascular causes, with no significant benefit for prasugrel over clopidogrel [116].

The PRINCIPLE-TIMI 44 (Prasugrel in Comparison to Clopidogrel for Inhibition of Platelet Activation and Aggregation—Thrombolysis in Myocardial Infarction 44) study showed in patients undergoing planned PCI that prasugrel (60-mg loading dose and 10-mg maintenance dose) achieves greater platelet inhibition than high dose of clopidogrel (600-mg loading dose and 150-mg maintenance dose) with a more rapid onset. There appeared to be greater levels of inhibition and fewer patients with poor response with standard dose of prasugrel compared to a higher than standard dose of clopidogrel. Even if the use of higher doses of clopidogrel is adopted in the future, greater inhibition of platelet function can be achieved with prasugrel. TIMI 44 was not powered to detect clinical efficacy differences between treatment groups, and few major adverse cardiovascular events occurred [117].

### 6.2 AZD6140

The nonthienopyridine AZD6140 is an orally active, reversible inhibitor of the P2Y12 receptor with distinct pharmacodynamic and pharmacokinetic properties than clopidogrel. Unique characteristics of this agent are that the parent molecule is active, and, therefore, liver metabolism is not needed for antiplatelet effect, and that binding to the receptor is reversible. AZD6140 has a faster onset of action and achieves greater and more consistent inhibition of ADP-induced platelet aggregation than standard doses of clopidogrel. Its short half-life requires dosing twice a day, but offset of effect is not expected to be immediate because of the persistence of AZD6140 in plasma [24].

The DISPERSE 2 (Dose Confirmation Study Assessing Anti-Platelet Effects of AZD6140 versus Clopidogrel in NSTEMI 2) safety trial compared different dose regimens of AZD6140 with clopidogrel. AZD6140 proved to have similar bleeding rates than clopidogrel in patients with NSTEMI [118]. A lower rate of bleeding in AZD6140-treated patients undergoing coronary artery bypass graft (CABG) 1–5 days after stopping study drug indicated a faster recovery of platelet function when compared to clopidogrel-treated patients. However, further studies with larger number of patients are needed to confirm such a benefit with this agent. The use of the reversible agent AZD6140, with a half-life of 12 h, would allow greater flexibility for use of P2Y12 blockers, in managing all patients at the time of presentation with ACS, and stopping treatment for those who need CABG and other types of elective surgery. In DISPERSE 2 overall tolerability of AZD6140 was good, with a higher incidence of dyspnea, hypotension, nausea, and a mild increase in uric acid levels among AZD6140 recipients. In this study, there were not enough clinical events to determine the efficacy of AZD6140 compared to clopidogrel [118]. Results of a large phase III trial are required to evaluate the true efficacy of this agent, which is currently being studied in the large outcomes trial PLATO (Platelet Inhibition and Patient Outcomes) in patients with ACS [26].

In a substudy of DISPERSE 2 AZD6140 inhibited platelet aggregation in a dose-dependent fashion and achieved greater and more consistent levels of inhibition than clopidogrel. AZD6140 further suppressed platelet aggregation in patients previously treated with clopidogrel [119].

### 6.3 Cangrelor

The nonthienopyridine cangrelor (AR-C69931MX) is an intravenous reversible P2Y12 antagonist, and its short half-life could offer a potential safety advantage

over clopidogrel [120]. Near total inhibition of ADP-induced platelet aggregation is achieved within minutes of initiation of intravenous infusion, and restoration of normal platelet aggregation is reached within 1 h after discontinuation [24]. Furthermore, greater P2Y12 receptor blockade can be achieved when compared to clopidogrel [121]. In a phase II study in patients undergoing PCI, there was an acceptable risk of bleeding and adverse cardiac events while achieving rapid, reversible inhibition of platelet aggregation with less prolongation of bleeding time than the GPIIb/IIIa inhibitor abciximab [122]. Patients treated with cangrelor and abciximab showed no significant differences in the incidence of combined major and minor bleeding after 7 days or in the composite incidence of adverse cardiac events after 30 days [17].

In a study in patients with unstable angina or non-Q-wave MI, cangrelor was well tolerated, with no difference in the incidence of serious adverse events compared with placebo [123]. Results from a small study to assess the safety and coronary artery patency of cangrelor as an adjunct to tissue plasminogen activator suggested the potential of this compound as an adjunct to fibrinolysis during treatment of acute MI [124]. The phase III trials CHAMPION-PCI and CHAMPION-PLATFORM (Cangrelor versus Standard Therapy to Achieve Optimal Management of Platelet Inhibition) will assess cangrelor's efficacy when compared to clopidogrel in patients undergoing PCI and will determine the effects of adding cangrelor to usual care in patients requiring PCI [17, 26].

## 7 OTHER ANTIPLATELET AGENTS

### 7.1 Dipyridamole

Dipyridamole was introduced clinically in the 1960s as a coronary vasodilator. Its antiplatelet effect is mediated through different biochemical mechanisms, such as selective inhibition of cyclic guanosine monophosphate (cGMP) phosphodiesterase type V enzyme with enhanced nitric oxide/cGMP signaling, inhibition of adenosine deaminase, reduction in adenosine uptake, and increase of prostanoid release from the endothelium [4, 26]. Dipyridamole is a less commonly used antiplatelet agent that has several effects, including decreased platelet activation, antiatherogenesis through antioxidant properties, and inhibition of restenosis by suppression of smooth muscle proliferation [6].

Addition of dipyridamole to aspirin in the ESPS 2 (European Stroke Prevention Study 2) and the ESPRIT (European/Australian Stroke Prevention in Reversible Ischemia Trial) reduced the relative risk of stroke by about 20% compared with aspirin alone, without causing excess bleeding. The mechanism by which it could reduce the risk for secondary strokes without excess bleeding may be associated to some of its effects beyond platelet inhibition on the vascular wall [125]. After the drug was reformulated to improve its bioavailability, positive results with the new preparation have been reported in ESPS, where dipyridamole appeared to have beneficial effects on stroke prevention without increasing bleeding complications in patients with cerebrovascular disease [4].

Even when this agent has been mainly studied in cerebrovascular disease, there are trials that evaluate its use for secondary prevention of MI. The PARIS (Persantine-Aspirin Reinfarction Study) trial showed no further benefit in death or MI prevention when added to aspirin [126], but dual therapy with dipyridamole and aspirin in PARIS II revealed reduction in the composite endpoint of MI or death after 12 months when compared to placebo [127].

### 7.2 Cilostazol

Cilostazol reversibly inhibits cyclic nucleotide phosphodiesterase type 3 enzyme and adenosine uptake in platelets [128], promotes arterial vasodilation, and suppresses vascular smooth muscle proliferation and hyperplasia of the intima [26]. In CREST (Cilostazol for Restenosis Trial) addition of cilostazol to aspirin and clopidogrel in patients undergoing PCI reduced the rate of restenosis [129]. Cilostazol therapy was effective for prevention of stent thrombosis, restenosis rate, and target lesion revascularization [130]. A small study showed that combination therapy with aspirin and cilostazol for the prevention of stent restenosis was comparable or superior to that of aspirin and clopidogrel in diabetic patients who underwent drug-eluting stent implantation [131]. Cilostazol in addition to an aspirin plus clopidogrel lead to additional inhibition of ADP-induced platelet activation in the primary PCI setting. This suggested that cilostazol may improve low responsiveness to clopidogrel in patients with acute MI who are undergoing stenting [132].

### 7.3 Protease-Activated Receptor Antagonists

Thrombin is a multifunctional serine protease produced at the site of vascular injury that transforms fibrinogen into fibrin, activates platelets, and affects multiple cell types such as endothelial cells, vascular smooth muscle cells, immune cells, and fibroblasts [133]. Platelet activation by thrombin occurs by binding to and cleavage of the extracellular domains of protease-activated receptors 1 and 4 (PAR1 and PAR4) [26, 134]. PAR1 seems to mediate platelet activation at low thrombin concen-

trations, whereas PAR4 is only able to mediate platelet activation at high thrombin concentrations. Thrombin induces platelet activation with shape change, mobilizes P-selectin and CD40 ligand to the platelet membrane, and favors the release of platelet agonists (e.g., ADP, serotonin, TxA2), chemokines, and growth factors. Thrombin inhibition is an emerging target for antiplatelet therapy. Thrombin antagonists, both parenteral (e.g., desirudin, lepidurin, bivalirudin, and argatroban) and oral (e.g., ximelagatran, dabigatran, and odiparcil) have shown efficacy in clinical trials for specific indications [133].

Several peptide and nonpeptide PAR antagonists have been developed. PAR1 antagonists seem to have an advantage over direct thrombin inhibitors since the PAR1 block does not inhibit enzymatic action of thrombin in the coagulation cascade with the consequent minimal bleeding side effects [135].

Thrombin is activated when thrombosis, hemorrhage, inflammation, or tissue damage occurs, and changes in the expression pattern of PARs with an upregulation of PAR1 have been associated with atherosclerosis and hypertension. Direct inhibition of PARs to modulate the effects of thrombin has been assessed. Anti-PAR1 antibodies directed against the thrombin cleavage site have been effective in vitro and in vivo to a lower extent. Peptide-competitive PAR antagonists are currently under development. The PAR1 antagonist RWJ-58259 prevents thrombus formation in nonhuman primates and PAR1 and PAR4 blocking with pepducins (i.e., cell-penetrating lipopeptides) inhibits platelet aggregation [133].

A safety study TRANSCENDENCE (Thrombin Receptor Antagonist for Clinical Event Reduction Over Standard Concomitant Therapies) with the oral PAR1 antagonist SCH 530348 (Schering-Plough) in subjects undergoing nonemergent PCI has been completed and ongoing studies with this compound are TRA 2°P—TIMI 50 (Trial to Assess the Effects of SCH 530348 in Preventing Heart Attack and Stroke in Patients with Atherosclerosis) and TRACER (Trial to Assess the Effects of SCH 530348 in Preventing Heart Attack and Stroke in Patients with Acute Coronary Syndrome) [26]. A phase II study that evaluated SCH 530348 for reducing PCI complications when added to standard antiplatelet therapy with aspirin and clopidogrel showed no increase in major and minor bleeding with a nonstatistically significant reduction in cardiovascular events at the highest dose tested. However, this study was not powered to establish efficacy. A 40-mg loading dose achieved greater than 80% inhibition of platelet aggregation in 2 h in up to 96% of the subjects, while maintenance doses (1 and 2.5 mg) sustained this level of inhibition at 30 and 60 days in all subjects [133]. In addition, there are a number of ongoing studies to evaluate the safety and tolerability of another PAR1 antagonist, E5555 (Eisai Medical Research Inc.) and its effects on markers of intravascular inflammation in subjects with CAD and ACS [26].

Simultaneous block of PAR1 and PAR4 seems to cause more effective inhibition of thrombin-induced platelet activation than antagonism of either receptor alone [136]; however, further investigation is needed to evaluate the effect of this synergic block on clinical outcomes.

## 8 EFFICACY VS. SAFETY

Optimal platelet inhibition requires a balance between minimizing the degree of ischemic risk and the possibility of bleeding.

### 8.1 Variable Response to Antiplatelet Therapy

Biologic processes are variable, and, accordingly, there is a variability of response to medications. The effect of any antiplatelet agent can be quantitated as the change in platelet activation and aggregation induced by the pharmacological intervention. This response to the therapeutic agent may fluctuate in different individuals. Response variability and nonresponsiveness have been demonstrated with both aspirin and clopidogrel. However, a specific and reliable laboratory method to identify nonresponse has not been accepted uniformly by different groups. It is not known if a laboratory test that detects nonresponse predicts future clinical events, and whether changing therapy based on the test result is beneficial to the patient. Until these links are evaluated within large trials, nonresponse in the laboratory cannot necessarily be ascribed as a cause of thrombosis. Therefore, it is still not yet clinically useful to test for nonresponse and change a patient's therapy on the basis of a laboratory test except in research trials [8].

### 8.2 Stent Thrombosis

Drug-eluting stents (DES) have been shown to reduce restenosis more effectively compared with bare metal stents (BMS) with similar risk of mortality [137–139]; however, DES may be associated with late stent thrombosis at discontinuation of clopidogrel and/or aspirin [140]. Premature discontinuation of clopidogrel therapy has been correlated with stent thrombosis leading to an increased risk for death. However, the wide window of thrombotic risk, which is the median time to late stent thrombosis after withdrawal of clopidogrel, and the fact that most cases of late stent thrombosis are not related to nonadherence to clopidogrel dispute the critical role

of compliance and suggests there may be additional mechanisms involved in this setting [38]. A prospective observational cohort study showed that nonresponsiveness to clopidogrel seems to be a strong independent predictor of stent thrombosis after 6 months of follow-up in patients receiving sirolimus or paclitaxel eluting stents. Posttreatment platelet aggregation after a 600-mg loading dose of clopidogrel may be a marker of the risk of stent thrombosis since patients with platelet aggregation greater than 70% after treatment had nearly fourfold increase in definite or probable stent thrombosis as compared with clopidogrel responders [141]. As DES thrombosis is associated with MI and mortality rates, there is a need for further studies to identify predictors of stent thrombosis and clarify the appropriate duration of antiplatelet therapy [142].

### 8.3 Bleeding

The major side effect of antiplatelet therapy is hemorrhage, and the risk of bleeding must be balanced against the risk of an atherothrombotic event. For example, a meta-analysis in patients with vascular disease showed that daily low-dose aspirin (75–150 mg) reduced the risk of major cardiovascular events (nonfatal MI, nonfatal ischemic stroke, or vascular death) by 32% and the risk of major bleeding was increased by around 50%; since these patients are at high risk of atherothrombotic events, a significant benefit of aspirin therapy persisted [6]. Overall, the risk–benefit ratio favors the use of dual antiplatelet therapy in acute coronary syndromes or after PCI, whereas it does not support its use for primary prevention [26].

### 8.4 Adverse Drug Reactions

Aspirin use is associated with symptomatic gastrointestinal disturbances in 2–10% of individuals, and hypersensitivity in 0.3% of those who use it. Clopidogrel tolerability is similar to aspirin. Ticlopidine is linked to severe neutropenia in 0.8–1% of individuals. In the CAPRIE study, clopidogrel is associated with less gastrointestinal bleeding than aspirin, but there is a higher bleeding risk in the elderly population. Clopidogrel has a lower risk of neutropenia than ticlopidine and thrombocytopenic purpura is reported rarely [2]. The major complication with GPIIb/IIIa inhibitors is bleeding and its use may be associated with thrombocytopenia [143].

### 8.5 Combined Use of Anticoagulants and Antiplatelet Drugs

The combination of anticoagulants and antiplatelet agents is indicated in the treatment of acute arterial thrombosis, and when a patient has multiple disorders the use of both anticoagulant and antiplatelet therapy may be justified or required (e.g., coexistence of atrial fibrillation and intracoronary stent, concomitant prothrombotic venous and arterial disease, valvular heart disease coupled with coronary artery disease, heart failure, and acute coronary syndrome) [144].

For the initial and long-term management of acute coronary syndromes combination therapy with anticoagulant and antiplatelet agents is more effective than antiplatelet therapy alone but increases the risk of bleeding. No benefits have been shown for continuing low-molecular-weight heparin beyond the first week in aspirin-treated patients with NSTEMI [145]. Because of the risk of bleeding, determination of the intensity and duration of combined treatment requires careful assessment of the specific patient.

A large number of new oral direct thrombin inhibitors and oral factor Xa inhibitors are being directly compared with existing anticoagulants for the prevention and treatment of arterial thrombosis [145].

## 9 SPECIAL CONDITIONS

### 9.1 Peripheral Artery Disease

Peripheral arterial disease (PAD) is associated with atherosclerosis in the coronary and carotid arteries, leading to increased incidence of MI, ischemic stroke, and cardiovascular death [146]. Patients with PAD manifest increased platelet aggregation, higher levels of soluble platelet activation markers, greater thrombin generation, and altered fibrinolytic potential. These changes are consistent across the different stages of PAD and justify considering antithrombotic prophylaxis in all such patients [147].

PAD is characterized by a very high rate of mortality due mainly to stroke and MI. Antiplatelet therapy decreases mortality and cardiovascular events, particularly MI and ischemic stroke, among high-risk patients with PAD. Chronic antiplatelet treatment improves patency rates among patients with peripheral arterial bypass grafting or angioplasty and is the standard of care for subjects undergoing arterial revascularization [148, 149]. Antiplatelet treatment is indicated in virtually all patients with PAD. Aspirin, 75–325 mg per day, is considered as first-line treatment, and clopidogrel, 75 mg per day, is an effective alternative [146]. The U.S. Physicians' Health Study shows that chronic administration of low-dose aspirin to apparently healthy men reduces the need for peripheral arterial surgery [150]. Low-dose aspirin has been reported to reduce the incidence of vascular events in PAD patients, including

those with concomitant type 2 diabetes [151]. However, the study was underpowered to reach definitive conclusions on the clinical efficacy of aspirin in the PAD population [152]. Aspirin does not improve claudication, but it delays the progression rate, and reduces the need for intervention and graft failure in patients with revascularization procedures [148]. The Antithrombotic Trialists' Collaboration report analyzed 42 randomized trials that evaluate antiplatelet therapy in over 9200 patients with PAD and supports the use of these agents [147]. CAPRIE included more than 6400 patients with symptomatic PAD, and after 2 years of follow-up clopidogrel further reduced the relative risk of MI, ischemic stroke, or vascular death when compared to aspirin, especially in subjects with PAD [35–37]. Even though the CHARISMA trial showed no overall advantage of clopidogrel addition to low-dose aspirin, there was a significant benefit from dual antiplatelet therapy with clopidogrel plus aspirin in patients with symptomatic PAD [48]. Such patients may benefit from antithrombotic therapy beyond aspirin alone, but this needs to be validated in future trials. Currently, antiplatelet therapy with aspirin or clopidogrel is recommended in all patients with PAD, including asymptomatic individuals with an abnormal ankle-brachial index (ABI). The optimal aspirin dose for the prevention of cardiovascular and limb-related events has not been established and a daily dose between 75 and 325 mg should be used. Even when efficacy data favor clopidogrel for patients with PAD, other factors such as comorbid conditions, tolerability, and cost must be considered when choosing the antiplatelet agent, and routine use of dual antiplatelet therapy with clopidogrel and aspirin is not recommended unless needed for another indication [149].

In the CASTLE (Cilostazol: A Study in Long-term Effects Long-Term Safety of Cilostazol in Patients with Peripheral Artery Disease) study all-cause or cardiovascular mortality was similar for cilostazol when compared to placebo, and serious bleeding events appeared not to be increased by cilostazol. However, the study was underpowered to detect a small adverse impact of cilostazol on mortality, and adherence to study medication was poor with over 60% of participants discontinuing therapy by 36 months [153]. Several studies have compared cilostazol with placebo in patients with intermittent claudication (IC), with improvement in maximal walking distance, and pain-free walking distance with cilostazol [154, 155]. Cilostazol has vasodilatory effects, antiplatelet properties, and vascular antiproliferative effects. The mechanism by which it increases walking distance in patients with IC has not been completely defined, but it appears to relate to its cyclic adenosine monophosphate (cAMP)-mediated vascular properties [156].

### 9.2 Ischemic Stroke and Transient Ischemic Attack

Antiplatelet therapy is indicated for the treatment of ischemic stroke or transient ischemic attack (TIA). It has been seen that aspirin reduces subsequent occlusive vascular events, including recurrent stroke, by about 25% [157]. Low-dose aspirin alone or sustained-release dipyridamole alone were equally effective in stroke and TIA prevention, and combination therapy showed additive effects being more effective than the single agents [158]. The Second European Stroke Prevention Study (ESPS2) reported that antiplatelet therapy did not influence the severity of recurrent stroke but seemed to lengthen the time the patient remained free from a recurrent stroke [159, 160]. However, combination therapy with aspirin and clopidogrel provided little benefit compared with either drug alone but increased the risk of serious bleeding [157]. MATCH showed that adding aspirin to clopidogrel in high-risk patients with recent ischemic stroke or TIA did not reduce major vascular events but increased the risk of life-threatening or major bleeding [161]. Clopidogrel is an effective alternative for long-term treatment of survivors of ischemic stroke or TIA that are intolerant to aspirin [157].

Recent evidence suggests that the risk of stroke in the first few months after TIA is higher than previously realized. In general, aspirin (50–325 mg/day) is the first line of treatment to prevent future stroke. Other antiplatelet agents such as clopidogrel and dipyridamole may be given. There is no evidence to prove that using a 300-mg loading dose of clopidogrel is superior (or inferior) to aspirin in the acute setting. However, among patients with TIA secondary to an acutely symptomatic carotid stenosis, the CARESS (The Canadian Registry of Palvizumab) study suggested that dual antiplatelet therapy with clopidogrel and aspirin was more effective than aspirin alone in reducing the risk of asymptomatic embolization and stroke [162].

### 9.3 Diabetes Mellitus

Atherosclerotic disease, including ischemic heart disease, stroke, and peripheral disease, accounts for the majority of morbidity and mortality associated with diabetes mellitus (DM). Several mechanisms may explain the increased atherothrombotic risk in diabetic patients, including other cardiovascular risk factors frequently present in this population (e.g., hypertension, dyslipidemia, and obesity), and other factors specific to the diabetic population (e.g., hyperglycemia, insulin resistance, and pro-inflammatory and prothrombotic status). The prothrombotic status is related to endothelial dysfunction, impaired fibrinolysis, increased coagulation factors,

increased platelet reactivity, and reduced responsiveness to oral antiplatelet agents [163].

Platelet dysfunction has been reported in patients with type 2 DM treated with standard dose of clopidogrel and aspirin, when compared to nondiabetic patients. Increased platelet aggregation and activation was observed acutely after a 300-mg loading dose of clopidogrel and persisted after long-term therapy [164, 165], with a high proportion of clopidogrel nonresponders in the diabetic population [165]. Diabetic individuals with ACS had a lower response to a 600-mg clopidogrel loading dose as measured by platelet aggregation, and suggested that diabetes has a substantial influence on residual platelet activity [166].

OPTIMUS (Optimizing Anti-Platelet Therapy In Diabetes Mellitus) evaluated the impact of a high clopidogrel maintenance dose in diabetic patients with previously established suboptimal clopidogrel response to standard doses. The study showed that a high 150-mg maintenance dose of clopidogrel was associated with enhanced antiplatelet effects compared with a standard 75-mg dose in high risk diabetic patients. However, despite the overall improvement in platelet function profiles in patients treated with the high maintenance dose, response to antiplatelet therapy remained variable, and a significant proportion of patients showed persistent platelet hyperreactivity. The clinical implications of this are not known [167]. High platelet reactivity determined in diabetic patients with coronary artery disease while on chronic dual antiplatelet therapy has been associated with a higher risk of long-term adverse cardiovascular events [168]. These findings draw attention to the potential need for more effective antithrombotic strategies and individualized therapy may be required, especially in high-risk populations, such as diabetic patients.

Availability of new and more potent P2Y12 receptor antagonists, currently under development, and tailored antithrombotic approaches with currently approved medications will represent an important approach to treat diabetic patients and help reduce their thrombotic risk [163]. In TIMI 38 prasugrel showed superior efficacy in specific subgroups. In regard to the primary endpoint, the benefit of prasugrel tended to be greater among patients with diabetes than among those without diabetes [116].

## 10 MONITORING PLATELET FUNCTION AND ANTITHROMBOTIC THERAPY

Even when monitoring anticoagulation is standard practice, monitoring antiplatelet therapy has not been fully adopted as a way to assess antithrombotic response. Variability in platelet inhibition by standard doses of aspirin and/or clopidogrel in patients with coronary artery disease may suggest that using fixed doses of these agents for all patients may lead to suboptimal clinical outcomes [169]. For patients with coronary artery disease, the potential utility of platelet function tests and point-of-care (POC) instruments is to monitor the efficacy of antiplatelet therapy and predict clinical outcomes [56]. No single test can capture all aspects of platelet function and no conclusive link has been shown between clinical outcomes and results from platelet function tests [170].

Light transmission aggregometry (LTA) is still considered as the gold standard of platelet function testing. This method has been the most widely used technique for monitoring the effects of antiplatelet agents (e.g., aspirin, clopidogrel, other P2Y12 inhibitors, and GPIIb/IIIa inhibitors) and to predict clinical outcomes. In this procedure, separated platelet-rich plasma is stirred under low-shear conditions, and platelets form aggregates after addition of various agonists. Conventional LTA requires large sample volumes, immediate processing and lengthy processing time, specialized equipment, and expertise to perform the tests [8, 56].

The VerifyNow instrument (Accumetrics) is a fully automated POC test. Blood sample tubes are simply mixed prior to insertion onto a specialized cartridge placed in the instrument. Aggregation is monitored by light transmission through two duplicate reaction chambers in each cartridge. There are specialized cartridges available for measuring platelet response to either aspirin or clopidogrel and other P2Y12 antagonists. The VerifyNow Aspirin cartridge contains fibrinogen-coated beads and a platelet activator (metallic cations and arachidonic acid) to stimulate the COX-1 pathway, and the test should give similar results to those obtained by arachidonic-acid-induced LTA. The VerifyNow P2Y12 cartridge uses prostaglandin (PG) E1 in addition to ADP to increase intracellular cAMP, enhancing the sensitivity and specificity of the test for ADP-induced activation of platelets via P2Y12 [8]. This method has been used to predict outcomes in patients undergoing PCI, and it is easy and fast to use and correlates with turbidimetric aggregometry [56].

Thromboelastography (TEG) has been traditionally used to determine the risk of bleeding and as a guide to transfusion requirements [8]. TEG analyzes the viscoelastic changes that occur during coagulation, providing a graphical representation of the fibrin polymerization process, which allows a complete evaluation of the process of clot initiation, formation, and stability, using whole blood or plasma. This technology has been applied to the monitoring of antiplatelet drugs [171]. The platelet mapping technique quantifies the contribution of arachidonic-acid-induced platelet aggregation and/or

ADP-induced aggregation to the overall tensile strength of a platelet-fibrin clot and correlates with turbidimetric aggregometry [56].

The expression of activation-dependent receptors on platelets (e.g., P-selectin and GPIIb/IIIa) at baseline and after stimulation can be quantified by flow cytometry with monoclonal antibodies. This technique has been useful to evaluate pharmacologic effects, but less information is available correlating receptor expression and clinical outcomes. Platelet–leukocyte aggregates have been measured as a marker of platelet activation and may be a more stable measure of acute coronary syndromes compared with P-selectin [56]. The phosphorylation of VASP, an intraplatelet actin regulatory protein, is dependent on the level of activation of the P2Y12 receptor, which is targeted by clopidogrel and other agents in this class. In the flow cytometric-based VASP assay ADP and prostaglandin E1 (PGE1) are used and the degree of VASP phosphorylation is measured, which is proportional to the level of inhibition of the P2Y12 receptor [8]. The platelet reactivity index (PRI) is the difference in VASP fluorescence intensity between resting (+PGE1) and activated (+ADP) platelets [172]. Advantages of this test to measure the effect of antiplatelet drugs include specificity for the P2Y12 signaling pathway and the stability of the method compared with aggregometry [56].

There is significant variation in the response of individuals (patients and normal controls) to aspirin, clopidogrel, and GPIIb/IIIa antagonists as measured by various platelet function tests [8]. Platelet activation and the occurrence of thrombotic events are mediated by multiple signaling pathways, and a compound directed against a single pathway cannot be expected to prevent the occurrence of all events. Treatment failure or nonresponsiveness to an antiplatelet agent may indicate persistent activity of the specific target of the antiplatelet agent and is not sufficient evidence of drug "resistance." Since no single test encompasses the complexity of platelet biology and function, it may be problematic to correlate platelet function or drug-related inhibition with clinical outcomes [56].

The optimal level of platelet inhibition has not been established, but it is thought that higher levels of inhibition and greater consistency of response will improve outcomes as long as bleeding events can be controlled or prevented [2, 3]. For example, in TRITON-TIMI 38 prasugrel achieved greater inhibition of platelet aggregation than clopidogrel, which was associated to better clinical outcomes [116].

The clinical usefulness of platelet function tests for monitoring different patient populations using antiplatelet therapy is not clear. There is lack of studies to address the clinical usefulness of modulating therapy based on laboratory findings [24], and further studies are needed to establish a link between inadequate platelet inhibition and cardiovascular thrombotic risk. Before antiplatelet monitoring can be recommended in clinical practice, more trials need to be carried out to answer basic questions on its clinical utility and cost-effectiveness. Until then, monitoring of antiplatelet therapy should be performed for investigational purposes only [22]. Currently, there are no uniformly accepted methods to measure platelet reactivity or to quantify the extent of platelet inhibition after antiplatelet therapy. Therefore, specific treatment guidelines for patients who exhibit high platelet reactivity during clopidogrel or aspirin therapy are not available [2].

## 11 CONCLUSIONS AND FUTURE PERSPECTIVES

Aspirin, ticlopidine, clopidogrel, and IIb/IIIa inhibitors have been widely used in ACS and following bare metal and drug-eluting stent placement. Disadvantages of these agents include a potential increase in bleeding events, individual variability in response, and other clinical side effects. Clopidogrel is superior to aspirin in reducing ischemic vascular events in high-risk patients and has improved gastrointestinal tolerance but is associated with rash, diarrhea, and adverse hematologic outcomes. Combination antiplatelet therapy with aspirin and a thienopyridine is the optimal currently available combination following PCI and in patients with an ACS. There is individual variability in platelet inhibitory response to both aspirin and clopidogrel, but the prevalence depends on the definition used (clinical events vs. pharmacodynamic testing) as well as the specific laboratory assays used. Even when clopidogrel nonresponsiveness can be decreased by higher doses, interindividual variability persists and nonresponse cannot be eliminated. Although there is evidence that monitoring the effects of antiplatelet therapy may help identify patients at an increased risk of developing ischemic events, and multiple studies suggest that platelet reactivity before and after PCI correlates directly with adverse clinical outcomes, current clinical guidelines do not support routine screening for antiplatelet drug response. Large randomized clinical trials will better define the link between the level of platelet inhibition in vitro and the clinical benefit derived from antiplatelet therapy.

Current advances in the development of new antiplatelet agents will provide a wider selection of therapeutic options for the management of cardiovascular patients. Some of these new compounds may be more effective and safer to use in subjects at risk of ischemic cardiovascular events.

## REFERENCES


1. Wiviott SD (2006). Clopidogrel response variability, resistance, or both? *Am J Cardiol 98*(10A), 18N–24N.
2. Wiviott SD, Tantry US, Gurbel PA (2006). Clinical applications of antiplatelet therapy. *Rev Cardiovasc Med 7*(3), 130–146; quiz 148–149.
3. Kereiakes DJ, Michelson AD (2006). Platelet activation and progression to complications. *Rev Cardiovasc Med 7*(2), 75–81.
4. Born G, Patrono C (2006). Antiplatelet drugs. *Br J Pharmacol 147*(Suppl 1), S241–251.
5. Bray PF (2007). Platelet hyperreactivity: Predictive and intrinsic properties. *Hematol Oncol Clin North Am 21*(4), 633–645, v–vi.
6. Troxler M, Dickinson K, Homer-Vanniasinkam S (2007). Platelet function and antiplatelet therapy. *Br J Surg 94*(6), 674–682.
7. Michelson AD (2008). P2Y12 antagonism: Promises and challenges. *Arterioscler Thromb Vasc Biol 28*(3), s33–38.
8. Harrison P, et al. (2007). Measuring antiplatelet drug effects in the laboratory. *Thromb Res 120*(3), 323–336.
9. Steering Committee of the Physicians' Health Study Research Group (1989). Final report on the aspirin component of the ongoing Physicians' Health Study. *N Engl J Med 321*(3), 129–135.
10. Ridker PM, et al. (1991). The effect of chronic platelet inhibition with low-dose aspirin on atherosclerotic progression and acute thrombosis: Clinical evidence from the Physicians' Health Study. *Am Heart J 122*(6), 1588–1592.
11. Bartolucci AA, Howard G (2006). Meta-analysis of data from the six primary prevention trials of cardiovascular events using aspirin. *Am J Cardiol 98*(6), 746–750.
12. Berger JS, et al. (2006). Aspirin for the primary prevention of cardiovascular events in women and men: A sex-specific meta-analysis of randomized controlled trials. *JAMA 295*(3), 306–313.
13. Hayden M, et al. (2002). Aspirin for the primary prevention of cardiovascular events: A summary of the evidence for the U.S. Preventive Services Task Force. *Ann Intern Med 136*(2), 161–172.
14. Ridker PM, et al. (2005). A randomized trial of low-dose aspirin in the primary prevention of cardiovascular disease in women. *N Engl J Med 352*(13), 1293–1304.
15. ISIS-2 (Second International Study of Infarct Survival) Collaborative Group (1988). Randomised trial of intravenous streptokinase, oral aspirin, both, or neither among 17,187 cases of suspected acute myocardial infarction: ISIS-2. *Lancet 2*(8607), 349–360.
16. Baigent C, et al. (1998). ISIS-2: 10 year survival among patients with suspected acute myocardial infarction in randomised comparison of intravenous streptokinase, oral aspirin, both, or neither. The ISIS-2 (Second International Study of Infarct Survival) Collaborative Group. *BMJ 316*(7141), 1337–1343.
17. Steinhubl S, Roe MT (2007). Optimizing platelet P2Y12 inhibition for patients undergoing PCI. *Cardiovasc Drug Rev 25*(2), 188–203.
18. Tran HA, et al. (2007). Aspirin resistance. *Thromb Res 120*(3), 337–346.
19. Snoep JD, et al. (2007). Association of laboratory-defined aspirin resistance with a higher risk of recurrent cardiovascular events: A systematic review and meta-analysis. *Arch Intern Med 167*(15), 1593–1599.
20. Krasopoulos G, et al. (2008). Aspirin "resistance" and risk of cardiovascular morbidity: Systematic review and meta-analysis. *BMJ 336*(7637), 195–198.
21. Gurbel PA, et al. (2007). Evaluation of dose-related effects of aspirin on platelet function: Results from the Aspirin-Induced Platelet Effect (ASPECT) study. *Circulation 115*(25), 3156–3164.
22. Cattaneo M (2007). Resistance to antiplatelet drugs: Molecular mechanisms and laboratory detection. *J Thromb Haemost 5*(Suppl 1), 230–237.
23. Shantsila E, Watson T, Lip GY (2007). Aspirin resistance: What, why and when? *Thromb Res 119*(5), 551–554.
24. Wiviott SD, et al. (2006). Therapeutic goals for effective platelet inhibition: A consensus document. *Rev Cardiovasc Med 7*(4), 214–225.
25. Randomized trial of ridogrel, a combined thromboxane A2 synthase inhibitor and thromboxane A2/prostaglandin endoperoxide receptor antagonist, versus aspirin as adjunct to thrombolysis in patients with acute myocardial infarction. The Ridogrel Versus Aspirin Patency Trial (RAPT). *Circulation 89*(2), 588–595, (1994).
26. Meadows TA, Bhatt DL (2007). Clinical aspects of platelet inhibitors and thrombus formation. *Circ Res 100*(9), 1261–1275.
27. McTavish D, Faulds D, Goa KL (1990). Ticlopidine. An updated review of its pharmacology and therapeutic use in platelet-dependent disorders. *Drugs 40*(2), 238–259.
28. Janzon L, et al. (1990). Prevention of myocardial infarction and stroke in patients with intermittent claudication; effects of ticlopidine. Results from STIMS, the Swedish Ticlopidine Multicentre Study. *J Intern Med 227*(5), 301–308.
29. Balsano F, et al. (1990). Antiplatelet treatment with ticlopidine in unstable angina. A controlled multicenter clinical trial. The Studio della Ticlopidina nell'Angina Instabile Group. *Circulation 82*(1), 17–26.
30. Leon MB, et al. (1998). A clinical trial comparing three antithrombotic-drug regimens after coronary-artery stenting. Stent Anticoagulation Restenosis Study Investigators. *N Engl J Med 339*(23), 1665–1671.
31. Cutlip DE, et al. (1999). Acute and nine-month clinical outcomes after "suboptimal" coronary stenting: Results from the STent Anti-thrombotic Regimen Study (STARS) registry. *J Am Coll Cardiol 34*(3), 698–706.
32. Smith SC Jr, et al. (2006). ACC/AHA/SCAI 2005 guideline update for percutaneous coronary intervention: A report of the American College of Cardiology/American

Heart Association Task Force on Practice Guidelines (ACC/AHA/SCAI Writing Committee to Update the 2001 Guidelines for Percutaneous Coronary Intervention). *J Am Coll Cardiol 47*(1), e1–121.

33. Bertrand ME, et al. (2000). Double-blind study of the safety of clopidogrel with and without a loading dose in combination with aspirin compared with ticlopidine in combination with aspirin after coronary stenting: The clopidogrel aspirin stent international cooperative study (CLASSICS). *Circulation 102*(6), 624–629.
34. Bhatt DL, et al. (2002). Meta-analysis of randomized and registry comparisons of ticlopidine with clopidogrel after stenting. *J Am Coll Cardiol 39*(1), 9–14.
35. CAPRIE Steering Committee (1996). A randomised, blinded, trial of clopidogrel versus aspirin in patients at risk of ischaemic events (CAPRIE). *Lancet 348*(9038), 1329–1339.
36. Harker LA, et al. (1999). Comparative safety and tolerability of clopidogrel and aspirin: Results from CAPRIE. CAPRIE Steering Committee and Investigators. Clopidogrel versus aspirin in patients at risk of ischaemic events. *Drug Saf 21*(4), 325–335.
37. Bhatt DL, et al. (2000). Reduction in the need for hospitalization for recurrent ischemic events and bleeding with clopidogrel instead of aspirin. CAPRIE investigators. *Am Heart J 140*(1), 67–73.
38. Eshaghian S, et al. (2007). Role of clopidogrel in managing atherothrombotic cardiovascular disease. *Ann Intern Med 146*(6), 434–441.
39. Yusuf S, et al. (2001). Effects of clopidogrel in addition to aspirin in patients with acute coronary syndromes without ST-segment elevation. *N Engl J Med 345*(7), 494–502.
40. Budaj A, et al. (2002). Benefit of clopidogrel in patients with acute coronary syndromes without ST-segment elevation in various risk groups. *Circulation 106*(13), 1622–1626.
41. Yusuf S, et al. (2003). Early and late effects of clopidogrel in patients with acute coronary syndromes. *Circulation 107*(7), 966–972.
42. Mehta SR, et al. (2001). Effects of pretreatment with clopidogrel and aspirin followed by long-term therapy in patients undergoing percutaneous coronary intervention: The PCI-CURE study. *Lancet 358*(9281), 527–533.
43. Brener SJ, et al. (2007). Prolonged dual antiplatelet therapy after percutaneous coronary intervention reduces ischemic events without affecting the need for repeat revascularization: Insights from the CREDO trial. *J Invas Cardiol 19*(7), 287–290.
44. Steinhubl SR, et al. (2002). Early and sustained dual oral antiplatelet therapy following percutaneous coronary intervention: A randomized controlled trial. *JAMA 288*(19), 2411–2420.
45. Chen ZM, et al. (2005). Addition of clopidogrel to aspirin in 45,852 patients with acute myocardial infarction: Randomised placebo-controlled trial. *Lancet 366*(9497), 1607–1621.
46. Sabatine MS, et al. (2005). Effect of clopidogrel pretreatment before percutaneous coronary intervention in patients with ST-elevation myocardial infarction treated with fibrinolytics: The PCI-CLARITY study. *JAMA 294*(10), 1224–1232.
47. Sabatine MS, et al. (2005). Addition of clopidogrel to aspirin and fibrinolytic therapy for myocardial infarction with ST-segment elevation. *N Engl J Med 352*(12), 1179–1189.
48. Bhatt DL, et al. (2007). Patients with prior myocardial infarction, stroke, or symptomatic peripheral arterial disease in the CHARISMA trial. *J Am Coll Cardiol 49*(19), 1982–1988.
49. Patti G, et al. (2005). Randomized trial of high loading dose of clopidogrel for reduction of periprocedural myocardial infarction in patients undergoing coronary intervention: Results from the ARMYDA-2 (Antiplatelet therapy for Reduction of MYocardial Damage during Angioplasty) study. *Circulation 111*(16), 2099–2106.
50. Lotrionte M, et al. (2007). Meta-analysis appraising high clopidogrel loading in patients undergoing percutaneous coronary intervention. *Am J Cardiol 100*(8), 1199–1206.
51. Montalescot G, et al. (2006). A randomized comparison of high clopidogrel loading doses in patients with non-ST-segment elevation acute coronary syndromes: The ALBION (Assessment of the Best Loading Dose of Clopidogrel to Blunt Platelet Activation, Inflammation and Ongoing Necrosis) trial. *J Am Coll Cardiol 48*(5), 931–938.
52. Smith SC Jr, et al. (2006). ACC/AHA/SCAI 2005 Guideline Update for Percutaneous Coronary Intervention—Summary Article: A Report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines (ACC/AHA/SCAI Writing Committee to Update the 2001 Guidelines for Percutaneous Coronary Intervention). *J Am Coll Cardiol 47*(1), 216–235.
53. Kandzari DE, et al. (2004). Influence of treatment duration with a 600-mg dose of clopidogrel before percutaneous coronary revascularization. *J Am Coll Cardiol 44*(11), 2133–2136.
54. Angiolillo DJ, et al. (2007). Variability in individual responsiveness to clopidogrel: Clinical implications, management, and future perspectives. *J Am Coll Cardiol 49*(14), 1505–1516.
55. Barsky AA, Arora RR (2006). Clopidogrel resistance: Myth or reality? *J Cardiovasc Pharmacol Ther 11*(1), 47–53.
56. Gurbel PA, et al. (2007). Platelet function monitoring in patients with coronary artery disease. *J Am Coll Cardiol 50*(19), 1822–1834.
57. Gurbel PA, Tantry US (2007). Clopidogrel resistance? *Thromb Res 120*(3), 311–321.
58. Gurbel PA, et al. (2005). Platelet reactivity in patients and recurrent events post-stenting: Results of the PREPARE POST-STENTING Study. *J Am Coll Cardiol 46*(10), 1820–1826.

59. Gurbel PA, et al. (2005). Clopidogrel effect on platelet reactivity in patients with stent thrombosis: Results of the CREST Study. *J Am Coll Cardiol 46*(10), 1827–1832.
60. Gurbel PA, Bliden KP, Tantry US (2006). Effect of clopidogrel with and without eptifibatide on tumor necrosis factor-alpha and C-reactive protein release after elective stenting: Results from the CLEAR PLATELETS 1b study. *J Am Coll Cardiol 48*(11), 2186–2191.
61. Gurbel PA, et al. (2005). Clopidogrel loading with eptifibatide to arrest the reactivity of platelets: Results of the Clopidogrel Loading with Eptifibatide to Arrest the Reactivity of Platelets (CLEAR PLATELETS) study. *Circulation 111*(9), 1153–1159.
62. Lev EI, et al. (2006). Aspirin and clopidogrel drug response in patients undergoing percutaneous coronary intervention: The role of dual drug resistance. *J Am Coll Cardiol 47*(1), 27–33.
63. Angiolillo DJ, et al. (2006). Influence of aspirin resistance on platelet function profiles in patients on long-term aspirin and clopidogrel after percutaneous coronary intervention. *Am J Cardiol 97*(1), 38–43.
64. Serebruany VL (2007). The "clopidogrel resistance" trap. *Am J Cardiol 100*(6), 1044–1046.
65. Schneider DJ, Aggarwal A (2004). Development of glycoprotein IIb-IIIa antagonists: Translation of pharmacodynamic effects into clinical benefit. *Expert Rev Cardiovasc Ther 2*(6), 903–913.
66. Topol EJ, et al. (1994). Randomised trial of coronary intervention with antibody against platelet IIb/IIIa integrin for reduction of clinical restenosis: Results at six months. The EPIC Investigators. *Lancet 343*(8902), 881–886.
67. Topol EJ, et al. (1997). Long-term protection from myocardial ischemic events in a randomized trial of brief integrin beta3 blockade with percutaneous coronary intervention. EPIC Investigator Group. Evaluation of Platelet IIb/IIIa Inhibition for Prevention of Ischemic Complication. *JAMA 278*(6), 479–484.
68. Platelet glycoprotein IIb/IIIa receptor blockade and low-dose heparin during percutaneous coronary revascularization. The EPILOG Investigators. *N Engl J Med 336*(24), 1689–1696 (1997).
69. Lincoff AM, et al. (2000). Abciximab and bleeding during coronary surgery: Results from the EPILOG and EPISTENT trials. Improve long-term outcome with abciximab GP IIb/IIIa blockade. Evaluation of Platelet IIb/IIIa Inhibition in STENTing. *Ann Thorac Surg 70*(2), 516–526.
70. Lincoff AM, et al. (1999). Sustained suppression of ischemic complications of coronary intervention by platelet GP IIb/IIIa blockade with abciximab: One-year outcome in the EPILOG trial. Evaluation in PTCA to improve long-term outcome with abciximab GP IIb/IIIa blockade. *Circulation 99*(15), 1951–1958.
71. Lincoff AM, et al. (1997). Standard versus low-dose weight-adjusted heparin in patients treated with the platelet glycoprotein IIb/IIIa receptor antibody fragment abciximab (c7E3 Fab) during percutaneous coronary revascularization. PROLOG Investigators. *Am J Cardiol 79*(3), 286–291.
72. Lincoff AM, et al. (1999). Complementary clinical benefits of coronary-artery stenting and blockade of platelet glycoprotein IIb/IIIa receptors. Evaluation of Platelet IIb/IIIa Inhibition in Stenting Investigators. *N Engl J Med 341*(5), 319–327.
73. Randomised placebo-controlled trial of abciximab before and during coronary intervention in refractory unstable angina: The CAPTURE Study. *Lancet 349*(9063), 1429–1435 (1997).
74. Klootwijk P, et al. (1998). Reduction of recurrent ischemia with abciximab during continuous ECG-ischemia monitoring in patients with unstable angina refractory to standard treatment (CAPTURE). *Circulation 98*(14), 1358–1364.
75. Simoons ML (2001). Effect of glycoprotein IIb/IIIa receptor blocker abciximab on outcome in patients with acute coronary syndromes without early coronary revascularisation: The GUSTO IV-ACS randomised trial. *Lancet 357*(9272), 1915–1924.
76. Ottervanger JP, et al. (2003). Long-term results after the glycoprotein IIb/IIIa inhibitor abciximab in unstable angina: One-year survival in the GUSTO IV-ACS (Global Use of Strategies to Open Occluded Coronary Arteries IV—Acute Coronary Syndrome) Trial. *Circulation 107*(3), 437–442.
77. Inhibition of platelet glycoprotein IIb/IIIa with eptifibatide in patients with acute coronary syndromes. The PURSUIT Trial Investigators. Platelet Glycoprotein IIb/IIIa in Unstable Angina: Receptor Suppression Using Integrilin Therapy. *N Engl J Med 339*(7), 436–443 (1998).
78. Harrington RA (1997). Design and methodology of the PURSUIT trial: Evaluating eptifibatide for acute ischemic coronary syndromes. Platelet Glycoprotein IIb-IIIa in Unstable Angina: Receptor Suppression Using Integrilin Therapy. *Am J Cardiol 80*(4A), 34B–38B.
79. A comparison of aspirin plus tirofiban with aspirin plus heparin for unstable angina. Platelet Receptor Inhibition in Ischemic Syndrome Management (PRISM) Study Investigators. *N Engl J Med 338*(21), 1498–1505 (1998).
80. Inhibition of the platelet glycoprotein IIb/IIIa receptor with tirofiban in unstable angina and non-Q-wave myocardial infarction. Platelet Receptor Inhibition in Ischemic Syndrome Management in Patients Limited by Unstable Signs and Symptoms (PRISM-PLUS) Study Investigators. *N Engl J Med 338*(21), 1488–1497 (1998).
81. Brener SJ, et al. (1998). Randomized, placebo-controlled trial of platelet glycoprotein IIb/IIIa blockade with primary angioplasty for acute myocardial infarction. ReoPro and Primary PTCA Organization and Randomized Trial (RAPPORT) Investigators. *Circulation 98*(8), 734–741.
82. Neumann FJ, et al. (2000). Effect of glycoprotein IIb/IIIa receptor blockade with abciximab on clinical and angiographic restenosis rate after the placement of coronary

stents following acute myocardial infarction. *J Am Coll Cardiol 35*(4), 915–921.

83. Giri S, et al. (2000). Synergy between intracoronary stenting and abciximab in improving angiographic and clinical outcomes of primary angioplasty in acute myocardial infarction. *Am J Cardiol 86*(3), 269–274.
84. Antoniucci D, et al. (2003). A randomized trial comparing primary infarct artery stenting with or without abciximab in acute myocardial infarction. *J Am Coll Cardiol 42*(11), 1879–1885.
85. Montalescot G, et al. (2001). Platelet glycoprotein IIb/IIIa inhibition with coronary stenting for acute myocardial infarction. *N Engl J Med 344*(25), 1895–1903.
86. Three-year duration of benefit from abciximab in patients receiving stents for acute myocardial infarction in the randomized double-blind ADMIRAL study. *Eur Heart J 26*(23), 2520–2523 (2005).
87. Antoniucci D, et al. (2004). Abciximab-supported infarct artery stent implantation for acute myocardial infarction and long-term survival: A prospective, multicenter, randomized trial comparing infarct artery stenting plus abciximab with stenting alone. *Circulation 109*(14), 1704–1706.
88. Tcheng JE, et al. (2003). Benefits and risks of abciximab use in primary angioplasty for acute myocardial infarction: The Controlled Abciximab and Device Investigation to Lower Late Angioplasty Complications (CADILLAC) trial. *Circulation 108*(11), 1316–1323.
89. Stone GW, et al. (2002). Comparison of angioplasty with stenting, with or without abciximab, in acute myocardial infarction. *N Engl J Med 346*(13), 957–966.
90. Efficacy and safety of tenecteplase in combination with enoxaparin, abciximab, or unfractionated heparin: The ASSENT-3 randomised trial in acute myocardial infarction. *Lancet 358*(9282), 605–613 (2001).
91. Topol EJ (2001). Reperfusion therapy for acute myocardial infarction with fibrinolytic therapy or combination reduced fibrinolytic therapy and platelet glycoprotein IIb/IIIa inhibition: The GUSTO V randomised trial. *Lancet 357*(9272), 1905–1914.
92. Lincoff AM, et al. (2002). Mortality at 1 year with combination platelet glycoprotein IIb/IIIa inhibition and reduced-dose fibrinolytic therapy vs conventional fibrinolytic therapy for acute myocardial infarction: GUSTO V randomized trial. *JAMA 288*(17), 2130–2135.
93. Sinnaeve PR, et al. (2004). Efficacy of tenecteplase in combination with enoxaparin, abciximab, or unfractionated heparin: one-year follow-up results of the Assessment of the Safety of a New Thrombolytic-3 (ASSENT-3) randomized trial in acute myocardial infarction. *Am Heart J 147*(6), 993–998.
94. Recio-Mayoral A, et al. (2007). Clinical Trials Update from the European Society of Cardiology Congress in Vienna, 2007: PROSPECT, EVEREST, ARISE, ALOFT, FINESSE, Prague-8, CARESS in MI and ACUITY. *Cardiovasc Drugs Ther 21*(6), 459–465.
95. Kindermann M, et al. (2007). Clinical Trial Updates and Hotline Sessions presented at the European Society of Cardiology Congress 2007: (FINESSE, CARESS, OASIS 5, PRAGUE-8, OPTIMIST, GRACE, STEEPLE, SCAAR, STRATEGY, DANAMI-2, ExTRACT-TIMI-25, ISAR-REACT 2, ACUITY, ALOFT, 3CPO, PROSPECT, EVEREST, COACH, BENEFiT, MERLIN-TIMI 36, SEARCH-MI, ADVANCE, WENBIT, EUROASPIRE I-III, ARISE, getABI, RIO). *Clin Res Cardiol 96*(11), 767–786.
96. Kastrati A, et al. (2004). A clinical trial of abciximab in elective percutaneous coronary intervention after pretreatment with clopidogrel. *N Engl J Med 350*(3), 232–238.
97. Kastrati A, et al. (2006). Abciximab in patients with acute coronary syndromes undergoing percutaneous coronary intervention after clopidogrel pretreatment: The ISAR-REACT 2 randomized trial. *JAMA 295*(13), 1531–1538.
98. Ndrepepa G, et al. (2008). One-year clinical outcomes with abciximab vs. placebo in patients with non-ST-segment elevation acute coronary syndromes undergoing percutaneous coronary intervention after pre-treatment with clopidogrel: Results of the ISAR-REACT 2 randomized trial. *Eur Heart J 29*(4), 455–461.
99. Randomised placebo-controlled trial of effect of eptifibatide on complications of percutaneous coronary intervention: IMPACT-II. Integrilin to Minimise Platelet Aggregation and Coronary Thrombosis-II. *Lancet 349*(9063), 1422–1428 (1997).
100. Novel dosing regimen of eptifibatide in planned coronary stent implantation (ESPRIT): a randomised, placebo-controlled trial. *Lancet 356*(9247), 2037–2044 (2000).
101. Gibson CM, et al. (2001). Effect of eptifibatide on coronary flow reserve following coronary stent implantation (an ESPRIT substudy). Enhanced Suppression of the Platelet IIb/IIIa Receptor with Integrilin Therapy. *Am J Cardiol 87*(11), 1293–1295.
102. O'Shea JC, et al. (2001). Platelet glycoprotein IIb/IIIa integrin blockade with eptifibatide in coronary stent intervention: The ESPRIT trial: A randomized controlled trial. *JAMA 285*(19), 2468–2473.
103. O'Shea JC, Tcheng JE (2001). Eptifibatide in Percutaneous Coronary Intervention: The ESPRIT Trial Results. *Curr Interv Cardiol Rep 3*(1), 62–68.
104. Effects of platelet glycoprotein IIb/IIIa blockade with tirofiban on adverse cardiac events in patients with unstable angina or acute myocardial infarction undergoing coronary angioplasty. The RESTORE Investigators. Randomized Efficacy Study of Tirofiban for Outcomes and REstenosis. *Circulation 96*(5), 1445–1453 (1997).
105. Moliterno DJ, et al. (2002). Outcomes at 6 months for the direct comparison of tirofiban and abciximab during percutaneous coronary revascularisation with stent placement: The TARGET follow-up study. *Lancet 360*(9330), 355–360.
106. Topol EJ, et al. (2001). Comparison of two platelet glycoprotein IIb/IIIa inhibitors, tirofiban and abciximab, for

the prevention of ischemic events with percutaneous coronary revascularization. *N Engl J Med 344*(25), 1888–1894.

107. Mukherjee D, et al. (2005). Mortality at 1 year for the direct comparison of tirofiban and abciximab during percutaneous coronary revascularization: Do tirofiban and ReoPro give similar efficacy outcomes at trial 1-year follow-up. *Eur Heart J 26*(23), 2524–2528.
108. Gurbel PA, Serebruany VL (2000). Oral platelet IIb/IIIa inhibitors: From attractive theory to clinical failures. *J Thromb Thrombolysis 10*(3), 217–220.
109. Kereiakes DJ (1999). Oral platelet glycoprotein IIb/IIIa inhibitors. *Coron Artery Dis 10*(8), 581–594.
110. Leebeek FW, et al. (2002). Oral glycoprotein IIb/IIIa receptor inhibitors in patients with cardiovascular disease: Why were the results so unfavourable. *Eur Heart J 23*(6), 444–457.
111. O'Neill WW, et al. (2000). Long-term treatment with a platelet glycoprotein-receptor antagonist after percutaneous coronary revascularization. EXCITE Trial Investigators. Evaluation of Oral Xemilofiban in Controlling Thrombotic Events. *N Engl J Med 342*(18), 1316–1324.
112. Comparison of sibrafiban with aspirin for prevention of cardiovascular events after acute coronary syndromes: A randomised trial. The SYMPHONY Investigators. Sibrafiban versus Aspirin to Yield Maximum Protection from Ischemic Heart Events Post-Acute Coronary Syndromes. *Lancet 355*(9201), 337–345 (2000).
113. Cannon CP, et al. (2000). Oral glycoprotein IIb/IIIa inhibition with orbofiban in patients with unstable coronary syndromes (OPUS-TIMI 16) trial. *Circulation 102*(2), 149–156.
114. Holmes MB, et al. (2000). Increased platelet reactivity in patients given orbofiban after an acute coronary syndrome: An OPUS-TIMI 16 substudy. Orbofiban in patients with unstable coronary syndromes. Thrombolysis in myocardial infarction. *Am J Cardiol 85*(4), 491–493, A10.
115. Wiviott SD, et al. (2005). Randomized comparison of prasugrel (CS-747, LY640315), a novel thienopyridine P2Y12 antagonist, with clopidogrel in percutaneous coronary intervention: Results of the Joint Utilization of Medications to Block Platelets Optimally (JUMBO)-TIMI 26 Trial. *Circulation 111*(25), 3366–3373.
116. Wiviott SD, et al. (2007). Prasugrel versus clopidogrel in patients with acute coronary syndromes. *N Engl J Med 357*(20), 2001–2015.
117. Wiviott SD, et al. (2007). Prasugrel compared with high loading- and maintenance-dose clopidogrel in patients with planned percutaneous coronary intervention: The Prasugrel in Comparison to Clopidogrel for Inhibition of Platelet Activation and Aggregation-Thrombolysis in Myocardial Infarction 44 trial. *Circulation 116*(25), 2923–2932.
118. Cannon CP, et al. (2007). Safety, tolerability, and initial efficacy of AZD6140, the first reversible oral adenosine diphosphate receptor antagonist, compared with clopidogrel, in patients with non-ST-segment elevation acute coronary syndrome: Primary results of the DISPERSE-2 trial. *J Am Coll Cardiol 50*(19), 1844–1851.
119. Storey RF, et al. (2007). Inhibition of platelet aggregation by AZD6140, a reversible oral P2Y12 receptor antagonist, compared with clopidogrel in patients with acute coronary syndromes. *J Am Coll Cardiol 50*(19), 1852–1856.
120. Fugate SE, Cudd LA (2006). Cangrelor for treatment of coronary thrombosis. *Ann Pharmacother 40*(5), 925–930.
121. Storey RF, Wilcox RG, Heptinstall S (2002). Comparison of the pharmacodynamic effects of the platelet ADP receptor antagonists clopidogrel and AR-C69931MX in patients with ischaemic heart disease. *Platelets 13*(7), 407–413.
122. Greenbaum AB, et al. (2006). Initial experience with an intravenous P2Y12 platelet receptor antagonist in patients undergoing percutaneous coronary intervention: Results from a 2-part, phase II, multicenter, randomized, placebo- and active-controlled trial. *Am Heart J 151*(3), 689 e1–689 e10.
123. Jacobsson F, et al. (2002). Safety profile and tolerability of intravenous AR-C69931MX, a new antiplatelet drug, in unstable angina pectoris and non-Q-wave myocardial infarction. *Clin Ther 24*(5), 752–765.
124. Greenbaum AB, et al. (2007). Preliminary experience with intravenous P2Y12 platelet receptor inhibition as an adjunct to reduced-dose alteplase during acute myocardial infarction: Results of the Safety, Tolerability and Effect on Patency in Acute Myocardial Infarction (STEP-AMI) Angiographic Trial. *Am Heart J 154*(4), 702–709.
125. Kim HH, Liao JK (2008). Translational therapeutics of dipyridamole. *Arterioscler Thromb Vasc Biol 28*(3), s39–42.
126. The Persantine-Aspirin Reinfarction Study Research Group (1980). Persantine and aspirin in coronary heart disease. *Circulation 62*(3), 449–461.
127. Klimt CR, et al. (1986). Persantine-Aspirin Reinfarction Study. Part II. Secondary coronary prevention with persantine and aspirin. *J Am Coll Cardiol 7*(2), 251–269.
128. Schror K (2002). The pharmacology of cilostazol. *Diabetes Obes Metab 4*(Suppl 2), S14–19.
129. Douglas JS Jr, et al. (2005). Coronary stent restenosis in patients treated with cilostazol. *Circulation 112*(18), 2826–2832.
130. Chen YD, et al. (2006). A prospective randomized antiplatelet trial of cilostazol versus clopidogrel in patients with bare metal stent. *Chin Med J (Engl) 119*(5), 360–366.
131. Ahn Y, et al. (2008). Randomized comparison of cilostazol vs clopidogrel after drug-eluting stenting in diabetic patients. *Circ J 72*(1), 35–39.
132. Kim JY, et al. (2007). Cilostazol could ameliorate platelet responsiveness to clopidogrel in patients undergoing

primary percutaneous coronary intervention. *Circ J 71*(12), 1867–1872.

133. Martorell L, et al. (2008). Thrombin and protease-activated receptors (PARs) in atherothrombosis. *Thromb Haemost 99*(2), 305–315.
134. Burke FM, et al. (2006). Synthesis of novel peptide inhibitors of thrombin-induced platelet activation. *Chem Biol Drug Des 68*(5), 235–238.
135. Ahn HS, et al. (2003). Development of proteinase-activated receptor 1 antagonists as therapeutic agents for thrombosis, restenosis and inflammatory diseases. *Curr Pharm Des 9*(28), 2349–2365.
136. Wu CC, Teng CM (2006). Comparison of the effects of PAR1 antagonists, PAR4 antagonists, and their combinations on thrombin-induced human platelet activation. *Eur J Pharmacol 546*(1–3), 142–147.
137. Stettler C, et al. (2007). Outcomes associated with drug-eluting and bare-metal stents: A collaborative network meta-analysis. *Lancet 370*(9591), 937–948.
138. Kastrati A, et al. (2007). Meta-analysis of randomized trials on drug-eluting stents vs. bare-metal stents in patients with acute myocardial infarction. *Eur Heart J 28*(22), 2706–2713.
139. Moreno R, et al. (2007). Meta-analysis comparing the effect of drug-eluting versus bare metal stents on risk of acute myocardial infarction during follow-up. *Am J Cardiol 99*(5), 621–625.
140. Bavry AA, et al. (2006). Late thrombosis of drug-eluting stents: A meta-analysis of randomized clinical trials. *Am J Med 119*(12), 1056–1061.
141. Buonamici P, et al. (2007). Impact of platelet reactivity after clopidogrel administration on drug-eluting stent thrombosis. *J Am Coll Cardiol 49*(24), 2312–2317.
142. Gurbel PA, Kandzari DE (2007). Stent thrombosis associated with first-generation drug-eluting stents: Issues with antiplatelet therapy. *Neth Heart J 15*(4), 148–150.
143. Huxtable LM, Tafreshi MJ, Rakkar AN (2006). Frequency and management of thrombocytopenia with the glycoprotein IIb/IIIa receptor antagonists. *Am J Cardiol 97*(3), 426–429.
144. Schneider DJ, Sobel BE (2007). Conundrums in the combined use of anticoagulants and antiplatelet drugs. *Circulation 116*(3), 305–315.
145. Eikelboom JW, Hirsh J (2007). Combined antiplatelet and anticoagulant therapy: Clinical benefits and risks. *J Thromb Haemost 5*(Suppl 1), 255–263.
146. Blinc A, Poredos P (2007). Pharmacological prevention of atherothrombotic events in patients with peripheral arterial disease. *Eur J Clin Invest 37*(3), 157–164.
147. Hackam DG, Eikelboom JW (2007) Antithrombotic treatment for peripheral arterial disease. *Heart 93*(3), 303–308.
148. Hankey GJ, Norman PE, Eikelboom JW (2006). Medical treatment of peripheral arterial disease. *JAMA 295*(5), 547–553.
149. Gornik HL, Creager MA (2006). Contemporary management of peripheral arterial disease: I. Cardiovascular risk-factor modification. *Cleve Clin J Med 73*(Suppl 4), S30–37.
150. Goldhaber SZ, et al. (1992). Low-dose aspirin and subsequent peripheral arterial surgery in the Physicians' Health Study. *Lancet 340*(8812), 143–145.
151. Catalano M, Born G, Peto R (2007). Prevention of serious vascular events by aspirin amongst patients with peripheral arterial disease: Randomized, double-blind trial. *J Intern Med 261*(3), 276–284.
152. Violi F, Hiatt W (2007) A critical review of antiplatelet treatment in peripheral arterial disease. *Intern Emerg Med 2*(2), 84–87.
153. Hiatt WR, Money SR, Brass EP (2008). Long-term safety of cilostazol in patients with peripheral artery disease: The CASTLE Study (Cilostazol: A Study in Long-term Effects). *J Vasc Surg 47*(2), 330–336.
154. Regensteiner JG, et al. (2002). Effect of cilostazol on treadmill walking, community-based walking ability, and health-related quality of life in patients with intermittent claudication due to peripheral arterial disease: Meta-analysis of six randomized controlled trials. *J Am Geriatr Soc 50*(12), 1939–1946.
155. Thompson PD, et al. (2002). Meta-analysis of results from eight randomized, placebo-controlled trials on the effect of cilostazol on patients with intermittent claudication. *Am J Cardiol 90*(12), 1314–1319.
156. Carman TL, Fernandez BB Jr (2006). Contemporary management of peripheral arterial disease: II. Improving walking distance and quality of life. *Cleve Clin J Med 73*(Suppl 4), S38–44.
157. Gorelick P, Sechenova O, Hennekens CH (2006). Evolving perspectives on clopidogrel in the treatment of ischemic stroke. *J Cardiovasc Pharmacol Ther 11*(4), 245–248.
158. Forbes CD (1998). Secondary stroke prevention with low-dose aspirin, sustained release dipyridamole alone and in combination. ESPS Investigators. European Stroke Prevention Study. *Thromb Res 92*(1 Suppl 1), S1–6.
159. European Stroke Prevention Study 2. Efficacy and safety data. *J Neurol Sci 151*(Suppl), S1–77 (1997).
160. Sivenius J, et al. (1999). Antiplatelet treatment does not reduce the severity of subsequent stroke. European Stroke Prevention Study 2 Working Group. *Neurology 53*(4), 825–829.
161. Diener HC, et al. (2004). Aspirin and clopidogrel compared with clopidogrel alone after recent ischaemic stroke or transient ischaemic attack in high-risk patients (MATCH): Randomised, double-blind, placebo-controlled trial. *Lancet 364*(9431), 331–337.
162. Sylaja PN, Hill MD (2007). Current management of transient ischemic attack. *Am J Cardiovasc Drugs 7*(1), 67–74.
163. Angiolillo DJ (2006). Tackling the diabetic platelet: Is high clopidogrel dosing the answer? *J Thromb Haemost 4*(12), 2563–2565.

164. Angiolillo DJ, et al. (2006). Insulin therapy is associated with platelet dysfunction in patients with type 2 diabetes mellitus on dual oral antiplatelet treatment. *J Am Coll Cardiol 48*(2), 298–304.
165. Angiolillo DJ, et al. (2005). Platelet function profiles in patients with type 2 diabetes and coronary artery disease on combined aspirin and clopidogrel treatment. *Diabetes 54*(8), 2430–2435.
166. Geisler T, et al. (2007). Platelet response to clopidogrel is attenuated in diabetic patients undergoing coronary stent implantation. *Diabetes Care 30*(2), 372–374.
167. Angiolillo DJ, et al. (2007). Randomized comparison of a high clopidogrel maintenance dose in patients with diabetes mellitus and coronary artery disease: results of the Optimizing Antiplatelet Therapy in Diabetes Mellitus (OPTIMUS) Study. *Circulation 115*(6), 708–716.
168. Angiolillo DJ, et al. (2007). Impact of platelet reactivity on cardiovascular outcomes in patients with type 2 diabetes mellitus and coronary artery disease. *J Am Coll Cardiol 50*(16), 1541–1547.
169. Cannon CP, McLean DS (2006) Critical pathways using platelet testing to potentially optimize the use of oral antiplatelet therapy. *Am J Cardiol 98*(10A), 33N–38N.
170. Steinhubl SR, et al. (2008). Determining the efficacy of antiplatelet therapies for the individual: Lessons from clinical trials. *J Thromb Thrombolysis 26*(1), 8–13.
171. Luddington RJ (2005). Thrombelastography/thromboelastometry. *Clin Lab Haematol 27*(2), 81–90.
172. Aleil B, et al. (2005). Flow cytometric analysis of intraplatelet VASP phosphorylation for the detection of clopidogrel resistance in patients with ischemic cardiovascular diseases. *J Thromb Haemost 3*(1), 85–92.
173. Bhatt DL, et al. (2002). Amplified benefit of clopidogrel versus aspirin in patients with diabetes mellitus. *Am J Cardiol 90*(6), 625–628.
174. Bhatt DL, et al. (2001). Superiority of clopidogrel versus aspirin in patients with prior cardiac surgery. *Circulation 103*(3), 363–368.
175. Ringleb PA, et al. (2004). Benefit of clopidogrel over aspirin is amplified in patients with a history of ischemic events. *Stroke 35*(2), 528–532.
176. Mehta SR, Yusuf S (2000) The Clopidogrel in Unstable Angina to Prevent Recurrent Events (CURE) Trial Programme; rationale, design and baseline characteristics including a meta-analysis of the effects of thienopyridines in vascular disease. *Eur Heart J 21*(24), 2033–2041.
177. McLean DS, et al. (2007). Benefits and risks of clopidogrel pretreatment before coronary artery bypass grafting in patients with ST-elevation myocardial infarction treated with fibrinolytics in CLARITY-TIMI 28. *J Thromb Thrombol 24*(2), 85–91.
178. Scirica BM, et al. (2006). The role of clopidogrel in early and sustained arterial patency after fibrinolysis for ST-segment elevation myocardial infarction: The ECG CLARITY-TIMI 28 Study. *J Am Coll Cardiol 48*(1), 37–42.
179. Wang TH, et al. (2007). An analysis of mortality rates with dual-antiplatelet therapy in the primary prevention population of the CHARISMA trial. *Eur Heart J 28*(18), 2200–2207.
180. Bhatt DL, et al. (2006). Clopidogrel and aspirin versus aspirin alone for the prevention of atherothrombotic events. *N Engl J Med 354*(16), 1706–1717.
181. von Beckerath N, et al. (2005). Absorption, metabolization, and antiplatelet effects of 300-, 600-, and 900-mg loading doses of clopidogrel: Results of the ISAR-CHOICE (Intracoronary Stenting and Antithrombotic Regimen: Choose Between 3 High Oral Doses for Immediate Clopidogrel Effect) Trial. *Circulation 112*(19), 2946–2950.
182. von Beckerath N, et al. (2007). A double-blind, randomized study on platelet aggregation in patients treated with a daily dose of 150 or 75 mg of clopidogrel for 30 days. *Eur Heart J 28*(15), 1814–1819.
183. Topol EJ, et al. (2000). Design of the blockade of the glycoprotein IIb/IIIa receptor to avoid vascular occlusion (BRAVO) trial. *Am Heart J 139*(6), 927–933.

# GASTROINTESTINAL DISORDERS: A FOCUS ON ACID-SUPPRESSIVE AGENTS

Brian Hemstreet
*University of Colorado School of Pharmacy, Anschutz Medical Campus, Aurora, Colorado*

## 1 INTRODUCTION

During the late nineteenth century many theories were conceived within the medical community that development of some gastric disorders, particularly gastric ulceration, was somehow related to the secretion of gastric acid [1]. Hypersecretion of gastric acid was thought to be the main underlying defect leading to not only gastric ulceration but also in the development of erosive lesions within the esophagus. Subsequent research into the physiology of gastric acid secretion helped to delineate the main stimuli for acid secretion, and much interest was focused on gastrin as the main target for attenuation of acid secretion, while others touted histamine as the major contributor and next potential drug target [2]. An infectious etiology of peptic ulcer disease was also proposed at this time but received little interest from a research standpoint until later in the twentieth century.

Early attempts at treating acid-related disorders focused on neutralization of gastric acid with alkaline substances, such as milk, sodium bicarbonate, calcium carbonate, or bismuth. Dietary interventions were implemented as additional adjunctive treatments and often included ingestion of bland foods and use of smaller more frequent meals [1–3]. Alternatively, surgical gastrectomy or vagotomies were also less favorable options for patients unresponsive to these interventions. As more research into the physiology of gastric secretion accumulated, the focus of drug development turned more toward inhibition of the production and secretions of gastric acid.

The production and secretion of gastric acid is facilitated by the presence of greater than one billion parietal cells located in the mucosa of the stomach. These parietal cells secrete hydrochloric acid in response to various neuronal and hormonal stimuli [4]. The main external stimulus for gastric acid secretion is oral intake of food. Following stimulation, gastric acid secretion occurs in three phases. The cephalic phase is triggered by sight and smell of food, while the gastric and intestinal phases are regulated mainly by distention of the stomach, and breakdown products of protein and other ingested nutrients. At the level of the parietal cell, the three primary molecular mediators of acid secretion are acetylcholine, histamine, and gastrin [5, 6]. Gastrin, stored in antral G cells, is released into the bloodstream in response to food intake or increases in gastric pH [4]. As mentioned previously, cholinergic input resulting from innervation of the stomach by the vagus nerve leads to local release of acetylcholine and stimulation of muscarinic $M_3$ receptors on the basolateral membrane of the parietal cell. Stimulation of the cholycystokinin$_B$ ($CCK_B$) receptor by gastrin, or the $M_3$ receptors by acetylcholine receptor on the parietal cell, ultimately results in increased cytosolic concentrations of calcium. This in turn leads to stimulation of histamine release from enterochromaffin-like (ECL) cells. Histamine stimulates the $H_2$ receptor on the parietal cell, leading to increases in cytosolic cyclic

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

adenosine monophosphate (cAMP). Ultimately, this initiates the production of gastric acid.

The final step in acid production following stimulation of the parietal cell is the secretion of $H^+$ from the secretory cannaliculus, located on the apical membrane of the parietal cell. This secretion is accomplished by migration of an enzyme, $H^+$, $K^+$-adenosine triphosphatase ($H^+$, $K^+$-ATPase), from the cytosol to the membrane of the cannaliculus into which it becomes integrated [3–5]. The $H^+$, $K^+$-ATPase, also known as the proton pump, releases a hydrogen ion into the gastric lumen in exchange for a potassium ion. The concentration of $H^+$ in the gastric lumen results in lowering of gastric pH to less than 1. In contrast, both prostaglandins and somatostatin serve as inhibitory mediators of gastric acid secretion and thus serve as necessary counterregulatory measures [4].

## 2 NEED FOR ACID-SUPPRESSIVE AGENTS: AN OVERVIEW OF ACID-RELATED DISEASES

The need for drug therapy that targets gastric acid secretion stems from the wide variety of clinical conditions in which gastric acid plays a significant role. In North America alone, upper gastrointestinal complaints related to diseases potentially responsive to acid suppression such as gastroesophageal reflux disease (GERD), peptic ulcer disease, and dyspepsia are very prevalent and rank among the most common reasons for visits to outpatient clinics and to gastroenterologist offices [7, 8]. Likewise, these diseases significantly impact the health care system due extensive direct and indirect costs incurred with their treatment and consequences, as well as the potential for substantial reductions in quality of life [9]. Several studies have demonstrated that acid-suppressive agents are commonly used by a substantial portion of the patient population for various gastrointestinal conditions [10–12].

Dyspepsia, defined as chronic recurrent pain or discomfort in the upper abdomen, has been estimated to affect up to 25% of the U.S. population [13, 14]. While prevalent, further investigation of dyspepsia may lead to identification of gastric or esophageal abnormalities, peptic ulcer disease, or entities that are distinguished by the absence of esophageal injury and variable responsiveness to acid suppression, such as nonerosive reflux disease (NERD) or functional dyspepsia [15]. In contrast, several factors such as increases in obesity and sedentary behavior have largely led to GERD, becoming one of the most prevalent gastrointestinal disorders in the United States [2, 7, 8]. GERD is defined as the symptoms or mucosal damage resulting from the abnormal reflux of contents from the stomach into the esophagus [16]. Typical symptoms may include heartburn or regurgitation, often in relation to meals, and have been reported in up to 19% of patients reporting upper gastrointestinal symptoms [8]. A severe consequence of GERD includes erosive esophagitis, which may be evident in anywhere from 10 to 60% of patients undergoing upper endoscopy for GERD symptoms [15, 17]. Erosive disease may precipitate severe symptoms and possibly lead to upper gastrointestinal bleeding. Other complications of long-standing untreated GERD include esophageal strictures and the precancerous condition known as Barrett's esophagus.

Acid-suppressive agents also play a significant role in the management and prevention of nonvariceal upper gastrointestinal ulceration and bleeding, the majority of which is caused by peptic ulcer disease (PUD) [18, 19]. Peptic ulcer disease encompasses both duodenal and gastric ulcers, and major causes include both nonsteroidal anti-inflammatory drug (NSAID) use and *Helicobacter pylori* (*H. pylori*) infection. Potent suppression of gastric acid, particularly with proton pump inhibitors (PPIs), can both facilitate healing of established ulcers, as well as prevent ulceration in NSAID users who have established risk factors for this complication of NSAID use [20]. Likewise, acid suppression is a key component in drug regimens used to eradicate *H. pylori* infection, thus facilitating healing of established PUD, and in patients with excessive acid secretion due to hypersecretion of gastrin, as seen in Zollinger–Ellison syndrome [21, 22]. Lastly, critically ill patients often exhibit an elevated risk for stress-related mucosal disease (SRMD) often due to the need for mechanical ventilation or presence of coagulopathies. Acid suppression is the mainstay of prevention of SRMD [23]. Based on the wide variety of disease states and conditions that gastric acid is thought to play a major role, it is clear why the search for both safe and effective acid suppressant agents has been the focus of much research over the last half century.

## 3 FROM BENCH TO BEDSIDE: THE RISE OF HISTAMINE-2 RECEPTOR ANTAGONISTS (H2RAS)

During the continued research into the physiology of gastric acid production, it soon became evident that of the three major chemical inputs for stimulating production of gastric acid, antagonizing histamine at the level of the parietal cell represented the most promising for pursuit of a viable drug option. Anticholinergic medications were initially tried but were found to be neither very effective nor well tolerated, and attempts to isolate a molecule that would effectively inhibit the function of gastrin were met with little success [1, 2, 24]. Histamine, a biologic amine that is synthesized from the amino

acid histidine, is a neurotransmitter that is classified as autocoid. It is found in mast cells, ECLs, basophils, and also in the human central nervous system and bone marrow [25–27]. Histamine's pharmacologic actions vary depending on the area it is released. In addition to its gastric effects, it has appreciable effects on the cardiovascular system, smooth muscle, nervous system, and hemapoetic system [26, 27].

As now known, histamine's effects on gastric acid secretion are mediated by its actions on the histamine receptor, located on the basolateral membrane of the parietal cell. In 1966, a report detailing the presence of different subtypes of histamine receptors has ultimately led to the discovery of four histamine receptor subtypes [1, 28, 29]. The histamine-2 (H2) receptor was subsequently found to function as a stimulus for acid secretion through its associated guanosine triphosphate (GTP)-binding protein, resulting in activation of intracellular adenylate-cyclase [30, 31]. Several attempts to synthesize agents that were based on the structure of histamine and that would ultimately antagonize the H2 receptor were unsuccessful. Through rigorous modeling of gastric acid secretion and study of the H2 receptor, Black and colleagues utilized histamine as a chemical starting point and synthesized an imidazole compound called burimamide that functioned as a viable H2 receptor antagonist and would subsequently be testable in human subjects [26, 27, 29, 32]. This was a major discovery that significantly impacted the treatment of acid-related diseases and ushered in the era of clinical development of various agents within the class of H2RAs, with cimetidine subsequently being released for use in clinical practice in the late 1970s [3, 33, 34].

As a class the H2RAs function as selective, competitive, and reversible antagonists of histamine at the H2 receptor [26, 30, 35]. Structurally, these compounds are weak bases consisting of an aromatic ring joined to a polar group by a flexible four-atom chain (Fig. 1) [26]. Initially, it was thought that the aromatic ring needed to be an imidazole derivative in order to effectively function as an H2RA. Despite the prototype H2RA cimetidine containing the imidazole ring, the subsequent H2RAs developed contained both furan (ranitidine) and thiazole (famotidine and nizatidine) derived ring structures [27, 36]. The length of the side chain determines the degree of binding to the H2 receptor [26]. There is a 20- to 50-fold difference in potency between the four existing agents used in clinical practice, with cimetidine being the least potent, and famotidine the most potent [27, 35, 36]. Ranitidine and nizatidine both have similar potencies. Potency of these agents also depends on the local concentration of histamine, with decreases in the reduction of acid secretion seen in conditions where local concentrations of histamine are highly elevated, such as in the Zollinger–Ellison syndrome [5].

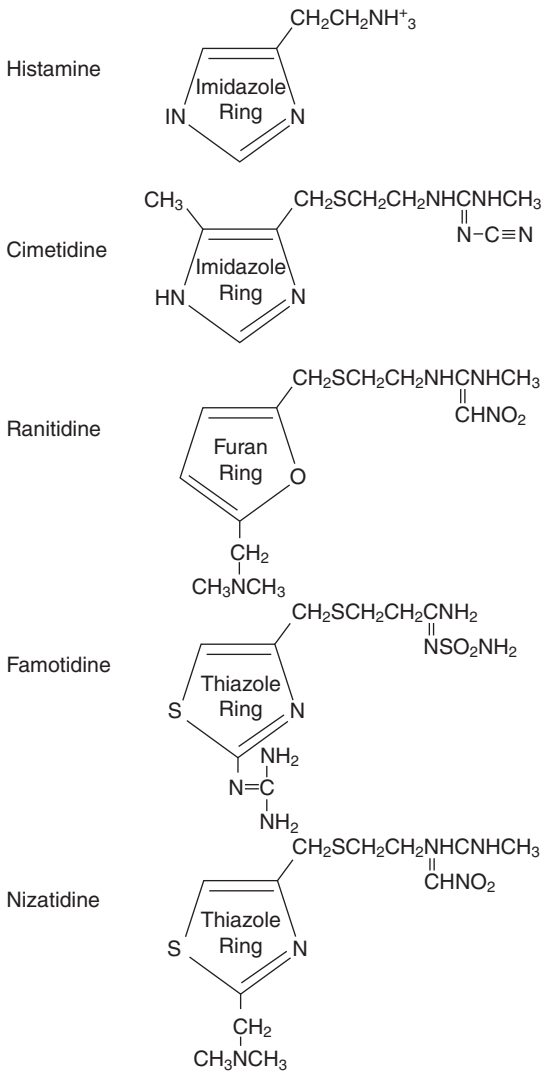


**Figure 1** Structure of the histamine-2 receptor antagonists.

All H2RAs are rapidly absorbed following oral ingestion, reaching peaking serum concentrations in 1–3 h, with bioavailabilities ranging from 40 to 100% [27, 35]. They are widely distributed throughout the body, with mean volumes of distribution of 1.5 l/kg and also exhibit a low degree of plasma protein binding (15–30%) [31]. Cimetidine, ranitidine, and famotidine are primarily metabolized by the liver and then renally excreted, while nizatidine is eliminated primarily via renal excretion [35]. Renal secretion also occurs as a minor route of elimination for both cimetidine and famotidine. All agents are rapidly eliminated via the kidney, with terminal half-lives of 1–4 h [27]. The pharmacokinetic profile of the H2RAs necessitates dose alteration in patients with renal disease. Likewise, both

cimetidine and ranitidine bind to the hepatic cytochrome P-450 enzyme system (CYP450). Cimetidine is a much more potent inhibitor than ranitidine, and use of these agents may result in significant drug interactions between other pharmacologic agents that share this route of metabolism [35]. Following the release of the oral H2RAs, intravenous (IV) dosage forms of cimetidine, ranitidine, and famotidine were also developed for clinical use. The availability of IV dosage forms expanded the role of H2RAs to use in patients who are unable to swallow or in those who require rapid acid suppression [37, 38].

Clinically, all four H2RAs inhibit both basal and meal-stimulated gastric acid secretion [3]. The effects on acid secretion typically occur within after 1–2 h following oral absorption, and in general there does appear to be a direct correlation between plasma concentration of the drug and degree of acid suppression [27, 39]. Evaluating the acid-suppressive effects of H2RAs using pentagastrin stimulation of gastric acid secretion and normal therapeutic doses, the duration above which the serum concentration of the H2RAs is at least able to inhibit 50% of acid secretion ($IC_{50}$) in vivo is roughly 6 h for cimetidine and 10 h for the other agents [27, 35]. However, given the reversible nature of their mechanism, use of H2RAs is met with the rapid development of tolerance to their antisecretory effects within a few days of treatment [5, 40–42]. This has been identified as a significant limitation to the long-term use of H2RAs for the maintenance treatment of acid-related diseases. Mechanisms by which tolerance to H2RAs is thought to occur include hyperstimulation of ECLs with subsequent histamine release as a direct consequence of the increases in gastrin, which accompanies blockade of the H2 receptor, possible upregulation of H2 receptors over time, or increases in parietal cell mass [3, 5, 40, 43]. Interestingly, *H. pylori* infection appears to possibly attenuate the development of tolerance to the acid-suppressive effects of the H2RAs in some patients, while concomitant use of PPIs has been shown to have variable affects on the development of H2RA tolerance [44–46].

As a drug class the H2RAs are generally well tolerated. Short-term transient adverse effects include both diarrhea and headache, while mental status changes and thrombocytopenia have also been reported, particularly in critically ill patients [26, 43]. Cimetidine does exhibit some antiandrogenic activity, leading to the development of loss of libido, impotence, and gynecomastia in some male patients [26].

Given that H2RAs have been used in clinical practice for more than 25 years, they have established efficacy in a wide variety of acid-related diseases. This includes healing of gastric and duodenal ulcers, treatment of GERD and esophagitis, prevention of SRMD, and treatment of *H. pylori* infection when prescribed with appropriate concomitant antibiotics [33, 47–49]. Release of these agents also lead to the reduction in the number of gastric surgeries performed for acid-related diseases [1]. Based on their rapid onset of action, the H2RAs have been useful as "on-demand" therapies for patient-directed treatment of heartburn and regurgitation [16, 50, 51]. Also, based on their favorable adverse effect profiles, these agents have also been approved for use on a nonprescription basis at doses only previously available with a prescription [52, 53]. Based on the increased prevalence of GERD and dyspepsia in the United States, this has allowed easy access to these agents, especially for effective treatment of symptomatic episodes of heartburn [54]. Another potential role for H2RAs is use in patients who experience nocturnal acid breakthrough (NAB) while receiving PPI therapy. This phenomenon is well described in users of PPIs, and use of H2RAs is based on the fact that they inhibit basal nocturnal acid secretion. Several studies have demonstrated that this may be an effective and lower cost strategy for managing NAB compared using twice daily PPI dosing [55–62].

## 4 UNMET NEEDS IN ACID SUPPRESSION: THE DEVELOPMENT OF PROTON PUMP INHIBITORS

The development and release of the H2RAs represented a major breakthrough in the management of acid-related diseases. However, there were drawbacks to their use in clinical practice. These included the ability to target only one of the three stimuli for acid secretion, the development of tolerance with prolonged use, the need for multiple daily dosing, the potential for drug–drug interactions and adverse effects, and the inability to provide sustained acid suppression leading to the relapse of PUD of erosive disease upon withdrawal of the drug in a significant portion of patients [1, 34]. Thus a search for more potent acid-suppressive agents soon was underway. Initially, attempts to synthesize more potent noncompetitive H2RAs with longer durations and of action were unsuccessful. This was based on the fact that the pronounced hypersecretion of gastrin seen with the use of noncompetitive agents led to the development of carcinoid and squamous tumors in early animal models [34]. Early research into the oxyntic cells of frogs revealed an enzyme on the parietal cell that exchanged the potassium for the hydrogen ion, later to be identified in humans as the $H^+$, $K^+$-ATPase, and subsequently referred to as the proton pump. As this enzyme represented the final common pathway for acid

secretion, this was considered a viable area for further pursuit, and in the mid-1970s prototype drugs, the PPIs, which could inhibit the proton pump, were first developed [1, 63].

The PPIs target the human $H^+$, $K^+$-ATPase, which is located in the parietal cell and is a heterodimer consisting of a 1034-amino-acid α subunit, and a smaller 291-amino-acid β subunit [4, 63]. In the resting state this enzyme resides in the cytoplasmic tubulovesicles, while upon stimulation the tubulovesicular membrane fuses into the apical secretory cannaliculus [4, 31]. The $H^+$, $K^+$-ATPase then exchanges $H^+$ for $K^+$ present in the extracellular cannalicular space, resulting in up to a 4-million-fold increase in $H^+$ concentration in this area. This constitutes the final pathway in acid secretion.

There are currently five PPIs marketed for use in the United States: omeprazole, lansoprazole, esomeprazole, pantoprazole, and rabeprazole. The synthesis of the PPIs is based on a common 2-pyridal-methyl-sulphinyl-benzimidazole structural nucleus (Fig. 2) [4, 31, 63, 64]. All of these agents are weak protonatable weak bases with $pK_a$'s in the range of 4–5 [65]. The PPIs are prodrugs, and at physiologic pH the unprotonated forms they are able to cross cell membranes, and thus when administered orally they are able to diffuse into the parietal cell from the bloodstream [64]. These agents then diffuse into the secretory cannaliculus, and upon protonation to their active sulfonamide forms in this acid environment, are able to bind covalently to cysteine residues on the α subunit of the $H^+$, $K^+$-ATPase [4, 64]. This binding is irreversible, resulting in permanent inactivation of the active $H^+$, $K^+$-ATPase. Likewise, the protonated forms of the PPIs are no longer able to cross cell membranes, leading to their further concentration in the secretory cannaliculus and inactivation of newly presented proton pumps [3, 63, 65].

The PPIs not only inhibit basal acid stimulation but also secretion from any source of stimulation, including food intake. Likewise, they inhibit the secretion of pepsinogen, as well as the activity of pepsin [66]. One substantial advantage of the PPIs compared to the H2RAs is the greater degree of acid suppression achieved. In general, potent acid suppression is defined as the ability of a drug to sustain an intragastric pH of greater than 4 for at least 16 h out of a 24-h day [66, 67]. Maintenance above this pH cutoff is associated with improvement in healing of esophagitis and ulceration, reduction in episodes of heartburn, and prevention of SRMD [43, 66–68]. Likewise, the area under the plasma concentration time curve (AUC) is indicative of the total amount of drug reaching the systemic circulation and also correlates well with the degree of gastric acid suppression [69]. The acid-suppressive effects of PPIs peak within the first 4–6 h and last much longer in duration

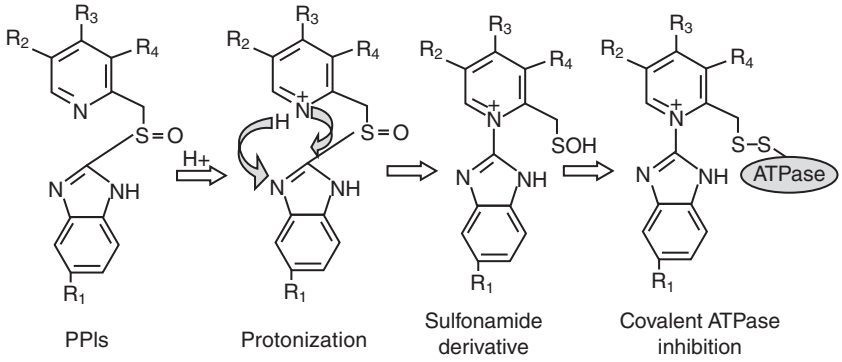


| PPIs | $R_1$ | $R_2$ | $R_3$ | $R_4$ |
|---|---|---|---|---|
| Omeprazol [(R)-Isomer] | $–OCH_3$ | $–CH_3$ | $–CH_3$ | $–CH_3$ |
| Lansoprazole | — | — | $–OCH_2CF_3$ | $–CH_3$ |
| Pantoprazole | $–OCF_2H$ | — | $–OCH_3$ | $–OCH_3$ |
| Rabeprazole | — | — | $–O(CH_2)_3–OCH_3$ | $–CH_3$ |
| Esomeprazole [(S)-Isomer] | $–OCH_3$ | $–CH_3$ | $–OCH_3$ | $–CH_3$ |

**Figure 2** Structure of the proton pump inhibitors.

compared to H2RAs [4, 34]. The PPIs are able to sustain intragastric pH above 3 for approximately 17 h per day and above a pH of 5 for roughly 9 h per day [3]. The rapidity in conversion to the sulfonamide species is affected by differences in $pK_a$ and the structural instability of the PPIs and does affect how quickly inactivation of the proton pump occurs. Rabeprazole and esomeprazole appear to be converted most quickly, followed by omeprazole and lansoprazole, and lastly pantoprazole [4, 65]. This appears to impact how rapid the acid-suppressive effects of these agents are achieved on the first day of dosing [70]. However, since PPIs only inhibit active proton pumps, their maximal effects on acid suppression is not achieved for several days [65]. Lastly, another major advantage over H2RAs is that tolerance to the acid-suppressive effects of PPIs does not develop over time. This is again related to their irreversible inhibition of the proton pump, thus making these drugs great choices for maintenance of acid-related disorders [66].

An interesting aspect of development of the PPIs has dealt with their unique pharmaceutical and pharmacokinetic properties. All PPIs are acid-labile drugs and are readily inactivated by gastric acid when given orally. This has necessitated the production of oral formulations that utilize enteric coatings or alkaline vehicles to facilitate administration and protect the parent drug [71]. All agents are well absorbed orally, and the actual plasma half-lives of all PPIs are extremely short, ranging from 0.7 to 1.2 h [72]. However, the irreversible nature of their mechanism leads to the prolonged duration of their pharmacologic effects. The oral bioavailabilities do differ among the available agents. Omeprazole and esomeprazole exhibit reduced bioavailability with initial dosing, with increases seen in bioavailability with repeated dosing. This is thought to be secondary to reductions in acid degradation of the drug as gastric pH increases over time [67]. Pantoprazole and lansoprazole have highest overall oral bioavailability [72, 73].

The PPIs should be taken during a fasting state, and prior to a meal in order to ensure that a maximal amount of proton pumps are active at the time of the peak serum concentration in the bloodstream [66]. All the PPIs are highly protein bound and undergo extensive hepatic metabolism by the CYP450 system, with the exception of rabeprazole, which undergoes nonenzymatic reduction to a thioether compound [72, 73]. The major CYP450 isoenzymes involved in the hepatic metabolism of the PPIs are CYP2C19 and to a lesser extent CYP3A4. Despite this, there are few major hepatic drug–drug interactions that occur due to the coadministration of PPIs with substrates of the cytochrome P-450 system. In fact, polymorphisms of the 2C19 enzymes are probably a more significant factor affecting the acid-suppressive profile of these agents [72, 73].

Once omeprazole was developed and marketed for use in the United States, this was followed by the development of lansoprazole, pantoprazole, and rabeprazole. In the search for more efficacious agents, an interesting approach was taken by revisiting the metabolic pathway of PPIs. By exploiting the differing pharmacologic properties of the structural isomers of PPIs, esomeprazole, the (S)-isomer of omeprazole, was subsequently developed for use in clinical practice. This was based on the fact that the metabolism of omeprazole by the CYP2C19 enzyme is stereoselective, with the metabolic rate of the (S)-isomer of omeprazole being lower and less variable than the (R)-isomer [74, 75]. The CYP2C19 enzyme is responsible for the formation of the hydroxy and 5-*O*-desmethyl metabolites of omeprazole, while the CYP450 enzyme is responsible for the development of the sulfone metabolite [74]. Interestingly, both isomers are protonated and converted to the same sulfenamide species, which inhibit the proton pump equally as effective [69, 75]. The differences in the pharmacologic profiles of omeprazole and esomeprazole do, therefore, lie in the differences in their metabolism. Overall the sum of the intrinsic clearances of all three metabolites of esomeprazole is approximately one-third that of the (R)-isomer, indicating a much slower clearance [74, 75]. This translates into a roughly 180% increase in plasma AUC for a 20 mg oral dose of esomeprazole compared to a 20 mg dose of omeprazole, leading to increases in the acid-suppressive effects of esomeprazole compared to the other PPIs [69, 75, 76].

Much like the H2RAs, once the oral PPIs were established as effective agents, research into the development of alternative dosage forms soon commenced. As mentioned previously, all the PPIs are acid-labile agents, and thus require formulation to protect them from gastric acid. This in essence makes them delayed-release products. Some formulations utilize capsules containing enteric-coated beads or spheres (omeprazole, lansoprazole, esomeprazole), while others utilize enteric-coated tablets (pantoprazole, rabeprazole) [71]. The advantage of some capsule formulations is that they can be opened and the contents sprinkled on various soft foods to ease administration. Likewise some products allow the contents to also be mixed with water or other fluids to facilitate administration through nasogastric tubes. Preparations of both lansoprazole and omeprazole suspension, utilizing sodium bicarbonate as a vehicle, were also formulated extemporaneously at many institutions for administration via nasogastric tubes [77]. Lansoprazole has most recently been formulated into an oral suspension, as well as an orally disintegrating tablet that can be dissolved orally or administered via a

nasogastric tube. The newest formulation of PPI on the market is a nonenteric-coated immediate-release omeprazole powder for oral suspension [78, 79]. This formulation results in rapid oral absorption of omeprazole and may have a role in on-demand therapy in those patients who have difficulty swallowing, patients with NAB, or in patients who have delayed gastric emptying [78]. Lastly, lansoprazole, esomeprazole, and pantoprazole have all been formulated into intravenous preparations, which can be given by intermittent bolus injection or by continuous infusion. This allows administration to hospitalized patients with GERD or gastrointestinal bleeding, those patients who are unable to swallow, or in those requiring pharmacologic prevention of SRMD [38, 71, 77, 80].

The superior acid-suppressive effects of the PPIs have resulted in their preferential use in many acid-related disorders compared to H2RAs. This has also been influenced by their favorable adverse effects profile. The most frequently reported adverse events are usually mild in nature and include diarrhea, nausea, abdominal pain, headache, and skin rashes [4]. The PPIs are associated with rebound hypersecretion of gastric acid upon withdrawal of their use. This is presumably due to increases in ECL mass as a result of their stimulation by the increased concentration of circulating gastrin that accompanies PPI use [5, 40]. This hypersecretion of gastrin seen with PPI has not been associated with significant increases in the incidence of gastric carcinoid or carcinomas [4, 66]. It is not totally clear what the consequences of this rebound increase in gastric acid are; however, some studies have documented increased in dyspeptic symptoms upon stoppage of PPIs [5]. Other potential effects of prolonged potent acid suppression include vitamin $B_{12}$ deficiency secondary to impaired release of dietary vitamin $B_{12}$ [66, 81]. A recent trial has also suggested that PPIs may impair dietary calcium absorption, leading to increases in the risk of hip fractures [82]. Lastly, given that gastric acid is thought to sterilize orally ingested nutrients, thus the potential for intestinal overgrowth of bacterial microorganisms with subsequent increases in rates of infection has been another concern with the use of PPIs. This includes increases in the risk of *Clostridium difficile* associated diarrhea, community-acquired pneumonia, and *Campylobacter jejuni* infection [63, 82–85].

Despite these potential adverse events, the PPIs have proven to be the best available agents for the treatment of most acid-related disorders [86]. In general, the PPIs all perform similarly when dosed appropriately for most conditions [87]. Use of PPIs can significantly reduce symptoms related to GERD and facilitate rapid healing of reflux esophagitis [16, 88–90]. They are also effective in preventing rebleeding in patients with nonvariceal upper gastrointestinal bleeding, as well as facilitating healing of gastric and duodenal ulcers and preventing recurrence of PUD [19, 91, 92]. The PPIs are also recommended as first-line agents for the prevention of NSAID-associated ulceration and treatment of *H. pylori* infection when prescribed with appropriate antibiotics [20, 21, 49]. Likewise, they are also very effective in the prevention of SRMD in critically ill patients and are considered the drugs of choice for the Zollinger–Ellison syndrome [22, 43]. While effective and safe, many studies have documented that the PPIs tend to be overused in clinical practice and are often prescribed for patients who do not meet criteria for their use [10, 93–97].

## 5 NEXT STEPS IN THE USE AND DEVELOPMENT OF ACID-SUPPRESSIVE AGENTS

Clearly, the development of H2RAs and PPIs revolutionized the treatment of acid-related disorders in the twentieth century. However, there are still several aspects of these drugs in the treatment of these disorders that are lacking. First and foremost, despite the potent acid suppression achieved with PPIs, there remains a percentage of the population, particularly those with GERD or erosive esophagitis, whose symptoms or disease are nonresponsive to this class of drugs [98–100]. There may also be substantial interpatient variability in response to PPIs, and the currently available agents do not provide 24-h suppression of gastric acid, with some patients experiencing significant NAB [99, 101]. Additionally, acid suppression with PPIs does not totally abolish the gastrointestinal toxicity secondary to NSAID therapy [102]. The use of a PPI that could be given without regards to meals and that may be used for on-demand therapy would also be favorable but has yet to be developed. This has lead to the search for newer agents to meet these needs.

Given that the PPIs have short half-lives, and thus are rapidly eliminated, research has focused on the development of agents with more favorable pharmacokinetic profiles. One such agent is the imidazopyridine PPI tenatoprazole. This agent not only differs in structure but has a much longer 7-h half-life. This offers the opportunity for a longer duration of action and possibly better control of nocturnal acid secretion [87, 102]. Much like the derivation of esomeprazole from omeprazole, the (S)-isomer of tenatoprazole is also being evaluated in human trials to assess its safety and efficacy [101]. Interestingly, some research into newer H2RAs, such as lafutidine, which exhibit more rapid and steady acid suppression in comparison to PPIs, may have a role

in on-demand treatment of acid-related disorders [103, 104].

Other novel approaches to acid suppression are also being researched. This includes agents such as the potassium competitive acid blockers (P-CABs), which bind to the $K^+$ binding region of the proton pump [102]. These agents act in a competitive and reversible manner and do not covalently bind to the $H^+ K^+$ ATPase as PPIs do. This results in a very rapid onset of action [99]. Preliminary results of trials show similar efficacy to esomeprazole [101]. Another class of agents, the $CCK_2$ receptor antagonists, appear to also suffer from the development of tolerance to their effects but may have a role as and adjunctive therapy [101]. Another potential target for reducing acid secretion is the H3 receptor, which is located in presynaptic brain neurons, as well as in the myenteric plexus in the gastrointestinal tract. Agonists of the H3 receptor have shown some promise as potential acid-suppressive agents in animal studies [99].

## 6 CONCLUSION

Acid-related disorders are prevalent in many countries. Research over the last century into the physiology of gastric acid secretion has yielded valuable insight into pathophysiology of many of these disorders. The discovery that histamine, gastrin, and acetylcholine all effect the production of gastric acid paved the way for the development of the H2RAs and PPIs. These agents have subsequently revolutionized the treatment of many acid-related disorders in clinical practice. The H2RAs provide both safe and effective acid suppression but are limited by the development of tolerance and the inability to provide sustained potent acid suppression over the course of a 24-h period. They remain useful for on-demand treatment of symptomatic diseases, such as GERD or dyspepsia, however, their use has been surpassed by superior acid suppression profile of the PPIs.

The PPIs represent the most potent acid-suppressant agents available today. Their irreversible potent acid-suppressant effects and favorable adverse effect profiles have made them the drugs of choice for most acid-related disorder. Tolerance does not develop to these agents, but they are associated with significant rebound acid secretion upon withdrawal, the clinical consequence of which is unknown. Despite their effectiveness, some patients remain unresponsive to the PPIs, and their short pharmacologic half-lives lead to the development of NAB in some patients. The search for longer acting PPIs is underway, with some agents already in clinical trials. The efficacy of other novel approaches to acid suppression, such as H3 agonists and P-CABS, wait to be seen. Thus targeting acid secretion remains a viable area of research for development of new pharmacologic agents.

## REFERENCES


1. Diaz-Rubio M (2005). Historical perspective of gastric acid inhibition. *Drugs 65* (Suppl 1), 1–6.
2. Malfertheiner P, et al. (2006). Review article: From gastrin to gastro-oesophageal reflux disease—A century of acid suppression. *Aliment Pharmacol Ther 23*(6), 683–690.
3. Garnett WR (2003). History of acid suppression: Focus on the hospital setting. *Pharmacotherapy 23*(10, Pt 2), 56S–60S.
4. Esplugues JV (2005). A pharmacological approach to gastric acid inhibition. *Drugs 65* (Suppl 1), 7–12.
5. Qvigstad G, Waldum H (2004). Rebound hypersecretion after inhibition of gastric acid secretion. *Basic Clin Pharmacol Toxicol 94*(5), 202–208.
6. Boyle JT (2003). Acid secretion from birth to adulthood. *J Pediatr Gastroenterol Nutr 37* (Suppl 1), S12–16.
7. Russo MW, et al. (2004). Digestive and liver diseases statistics, 2004. *Gastroenterology 126*(5), 1448–1453.
8. Frank L, et al. (2000). Upper gastrointestinal symptoms in North America: Prevalence and relationship to health-care utilization and quality of life. *Dig Dis Sci 45*(4), 809–818.
9. Sandler RS, et al. (2002). The burden of selected digestive diseases in the United States. *Gastroenterology*, *122*(5), 1500–1511.
10. Jacobson BC, et al. (2003). Who is using chronic acid suppression therapy and why? *Am J Gastroenterol 98*(1), 51–58.
11. Lassen A, Hallas J, Schaffalitzky De Muckadell OB (2004). Use of anti-secretory medication: A population-based cohort study. *Aliment Pharmacol Ther 20*(5), 577–583.
12. Martin RM, et al. (1998). Trends in prescribing H2-receptor antagonists and proton pump inhibitors in primary care. *Aliment Pharmacol Ther 12*(8), 797–805.
13. Talley NJ (2005). American Gastroenterological Association medical position statement: Evaluation of dyspepsia. *Gastroenterology 129*(5), 1753–1755.
14. Talley NJ, Vakil N (2005). Guidelines for the management of dyspepsia. *Am J Gastroenterol 100*(10), 2324–2337.
15. Fass R. (2007). Erosive esophagitis and nonerosive reflux disease (NERD): Comparison of epidemiologic, physiologic, and therapeutic characteristics. *J Clin Gastroenterol 41*(2), 131–137.
16. DeVault KR, Castell DO (2005). Updated guidelines for the diagnosis and treatment of gastroesophageal reflux disease. *Am J Gastroenterol 100*(1), 190–200.

17. Shaheen N, Provenzale D (2003). The epidemiology of gastroesophageal reflux disease. *Am J Med Sci 326*(5), 264–273.
18. Triadafilopoulos G (2005). Review article: The role of antisecretory therapy in the management of non-variceal upper gastrointestinal bleeding. *Aliment Pharmacol Ther 22*(Suppl 3), 53–58.
19. Barkun A, Bardou M, Marshall JK (2003). Consensus recommendations for managing patients with nonvariceal upper gastrointestinal bleeding. *Ann Intern Med 139*(10), 843–857.
20. Wilcox CM, et al. (2006). Consensus development conference on the use of nonsteroidal anti-inflammatory agents, including cyclooxygenase-2 enzyme inhibitors and aspirin. *Clin Gastroenterol Hepatol 4*(9), 1082–1089.
21. Chey WD, Wong BC (2007). American College of Gastroenterology guideline on the management of *Helicobacter pylori* infection. *Am J Gastroenterol 102*(8), 1808–1825.
22. Roy PK, et al. (2001). Gastric secretion in Zollinger-Ellison syndrome. Correlation with clinical expression, tumor extent and role in diagnosis—A prospective NIH study of 235 patients and a review of 984 cases in the literature. *Med. (Balt) 80*(3), 189–222.
23. Steinberg KP (2002). Stress-related mucosal disease in the critically ill patient: Risk factors and strategies to prevent stress-related bleeding in the intensive care unit. *Crit Care Med 30*(6, Suppl), S362–364.
24. Modlin IM, et al. (1997). The pivotal role of John S. Edkins in the discovery of Gastrin. *World J Surg 21*(2), 226–234.
25. Dove S. et al. (2004). Structure-activity relationships of histamine H2 receptor ligands. *Mini Rev Med Chem 4*(9), p. 941–954.
26. Pattichis K, Louca LL (1995). Histamine, histamine H2-receptor antagonists, gastric acid secretion and ulcers: An overview. *Drug Metabol Drug Interact 12*(1), 1–36.
27. Lin JH (1991). Pharmacokinetic and pharmacodynamic properties of histamine H2-receptor antagonists. Relationship between intrinsic potency and effective plasma concentrations. *Clin Pharmacokinet 20*(3), 218–236.
28. Ash AS, Schild HO (1966). Receptors mediating some actions of histamine. *Br J Pharmacol Chemother 27*(2), 427–439.
29. Black JW, et al. (1972). Definition and antagonism of histamine H2-receptors. *Nature 236*(5347), 385–390.
30. Sewing KF, Beil W, Hannemann H (1988). Comparative pharmacology of histamine H2-receptor antagonists. *Drugs 35*(Suppl 3), 25–29.
31. Shamburek RD, Schubert ML (1992). Control of gastric acid secretion. Histamine H2-receptor antagonists and H+K(+)-ATPase inhibitors. *Gastroenterol Clin North Am 21*(3), 527–550.
32. Black J (1993). Reflections on the analytical pharmacology of histamine h2-receptor antagonists. *Gastroenterology 105*(4), 963–968.
33. Bardhan KD, Nayyar AK, Royston C (2003). History in our lifetime: The changing nature of refractory duodenal ulcer in the era of histamine H2 receptor antagonists. *Dig Liver Dis 35*(8), 529–536.
34. Garner A (1994). Development of modern anti-ulcer drugs. *Br J Clin Pract Suppl 75*, 1–5; discussion 6–8.
35. Feldman M, Burton ME (1990). Histamine2-receptor antagonists. Standard therapy for acid-peptic diseases. 1. *N Engl J Med 323*(24), 1672–1680.
36. Sewing KF (1988). Chemical and biologic differences between various H2-receptor antagonists. *Scand J Gastroenterol Suppl 146*, 73–77.
37. Abe Y, et al. (2004). The comparative effects of single intravenous doses of omeprazole and famotidine on intragastric pH. *J Gastroenterol 39*(1), 21–25.
38. Kinoshita Y, Ishihara S (2004). Is an intravenous histamine H2 receptor antagonist superior to an intravenous proton pump inhibitor in acid suppression? *J Gastroenterol 39*(1), 92–93.
39. Adachi K, et al. (2005). Comparative study of the speed of acid-suppressing effects of oral administration of cimetidine and famotidine. *J Gastroenterol Hepatol 20*(7), 1012–1015.
40. Sandvik AK, Brenna E, Waldum HL (1997). Review article: The pharmacological inhibition of gastric acid secretion—Tolerance and rebound. *Aliment Pharmacol Ther 11*(6), 1013–1018.
41. Furuta K, et al. (2006). Tolerance to H2 receptor antagonist correlates well with the decline in efficacy against gastroesophageal reflux in patients with gastroesophageal reflux disease. *J Gastroenterol Hepatol 21*(10), 1581–1585.
42. Komazawa Y. et al. (2003). Tolerance to famotidine and ranitidine treatment after 14 days of administration in healthy subjects without *Helicobacter pylori* infection. *J Gastroenterol Hepatol 18*(6), 678–682.
43. Welage LS (2005). Overview of pharmacologic agents for acid suppression in critically ill patients. *Am J Health Syst Pharm 62*(10, Suppl 2), S4–S10.
44. Fujisawa T, et al. (2004). *Helicobacter pylori* infection prevents the occurrence of the tolerance phenomenon of histamine H2 receptor antagonists. *Aliment Pharmacol Ther 20*(5), 559–565.
45. Zentilin P, Savarino E, Savarino V (2005). *Helicobacter pylori* and tolerance to H2-blockers. *Aliment Pharmacol Ther 21*(3), 289–290; author reply 290–291.
46. Hashimoto H. et al. (2007). Does long-term medication with a proton pump inhibitor induce a tolerance to H2 receptor antagonist? *J Gastroenterol 42*(4), 275–278.
47. Feldman M, Burton ME (1990). Histamine2-receptor antagonists. Standard therapy for acid-peptic diseases (2). *N Engl J Med 323*(25), 1749–1755.

48. Graham DY, et al. (2003). Meta-analysis: Proton pump inhibitor or H2-receptor antagonist for *Helicobacter pylori* eradication. *Aliment Pharmacol Ther 17*(10), 1229–1236.
49. Gisbert JP, et al. (2003). Meta-analysis: Proton pump inhibitors vs. H2-receptor antagonists—Their efficacy with antibiotics in *Helicobacter pylori* eradication. *Aliment Pharmacol Ther 18*(8), 757–766.
50. Zacny J. et al. (2005). Systematic review: The efficacy of intermittent and on-demand therapy with histamine H2-receptor antagonists or proton pump inhibitors for gastro-oesophageal reflux disease patients. *Aliment Pharmacol Ther 21*(11), 1299–1312.
51. Kaplan-Machlis B, et al. (2000). Effectiveness and costs of omeprazole vs ranitidine for treatment of symptomatic gastroesophageal reflux disease in primary care clinics in West Virginia. *Arch Fam Med 9*(7), 624–630.
52. Furler MD, et al. (2002). Cost impact of switching histamine(2)-receptor antagonists to nonprescription status. *Ann Pharmacother 36*(7–8), 1135–1141.
53. Shi CW, et al. (2004). Consumer usage patterns of non-prescription histamine2-receptor antagonists. *Am J Gastroenterol 99*(4), 606–610.
54. Tran T, Lowry AM, El-Serag HB (2007). Meta-analysis: The efficacy of over-the-counter gastro-oesophageal reflux disease therapies. *Aliment Pharmacol Ther 25*(2), 143–153.
55. Fackler WK, et al. (2002). Long-term effect of H2RA therapy on nocturnal gastric acid breakthrough. *Gastroenterology 122*(3), 625–632.
56. Hatlebakk JG, et al. (1998). Nocturnal gastric acidity and acid breakthrough on different regimens of omeprazole 40 mg daily. *Aliment Pharmacol Ther 12*(12), 1235–1240.
57. Katz PO, et al. (1998). Gastro-oesophageal reflux associated with nocturnal gastric acid breakthrough on proton pump inhibitors. *Aliment Pharmacol Ther 12*(12), 1231–1234.
58. Leite LP, et al. (1996). Persistent acid secretion during omeprazole therapy: A study of gastric acid profiles in patients demonstrating failure of omeprazole therapy. *Am J Gastroenterol 91*(8), 1527–1531.
59. Pfefferkorn MD, et al. (2006). Nocturnal acid breakthrough in children with reflux esophagitis taking proton pump inhibitors. *J Pediatr Gastroenterol Nutr 42*(2), 160–165.
60. Tutuian R, et al. (2002). Over-the-counter H(2)-receptor antagonists do not compromise intragastric pH control with proton pump inhibitors. *Aliment Pharmacol Ther 16*(3), 473–477.
61. Vakil N, Guda N, Partington S (2006). The effect of over-the-counter ranitidine 75 mg on night-time heartburn in patients with erosive oesophagitis on daily proton pump inhibitor maintenance therapy. *Aliment Pharmacol Ther 23*(5), 649–653.
62. Vakil N. et al. (2006). Patterns of gastritis and the effect of eradicating *Helicobacter pylori* on gastro-oesophageal reflux disease in Western patients with non-ulcer dyspepsia. *Aliment Pharmacol Ther 24*(1), 55–63.
63. Richardson P. Hawkey CJ, Stack WA (1998). Proton pump inhibitors. Pharmacology and rationale for use in gastrointestinal disorders. *Drugs 56*(3), 307–335.
64. Pisegna JR (2002). Pharmacology of acid suppression in the hospital setting: Focus on proton pump inhibition. *Crit Care Med 30*(6, Suppl), S356–361.
65. Wolfe MM, Sachs G (2000). Acid suppression: Optimizing therapy for gastroduodenal ulcer healing, gastroesophageal reflux disease, and stress-related erosive syndrome. *Gastroenterology 118*(2, Suppl 1), S9–31.
66. Mearin F, Ponce J (2005). Potent acid inhibition: Summary of the evidence and clinical application. *Drugs 65*(Suppl 1), 113–126.
67. Freston JW (2004). Therapeutic choices in reflux disease: Defining the criteria for selecting a proton pump inhibitor. *Am J Med 117*(Suppl 5A), 14S–22S.
68. Hatlebakk JG (2003). Review article: Gastric acidity—Comparison of esomeprazole with other proton pump inhibitors. *Aliment Pharmacol Ther 17*(Suppl 1), 10–15; discussion 16–17.
69. Dent J (2003). Review article: Pharmacology of esomeprazole and comparisons with omeprazole. *Aliment Pharmacol Ther 17*(Suppl 1), 5–9.
70. Pantoflickova D, et al. (2003). Acid inhibition on the first day of dosing: Comparison of four proton pump inhibitors. *Aliment Pharmacol Ther 17*(12), 1507–1514.
71. Horn JR, Howden CW (2005). Review article: Similarities and differences among delayed-release proton-pump inhibitor formulations. *Aliment Pharmacol Ther 22*(Suppl 3), 20–24.
72. Yacyshyn BR, Thomson AB (2002). The clinical importance of proton pump inhibitor pharmacokinetics. *Digestion 66*(2), 67–78.
73. Ishizaki T, Horai Y (1999). Review article: Cytochrome P450 and the metabolism of proton pump inhibitors—Emphasis on rabeprazole. *Aliment Pharmacol Ther 13*(Suppl 3), 27–36.
74. Andersson T, et al. (2001). Pharmacokinetic studies with esomeprazole, the (S)-isomer of omeprazole. *Clin Pharmacokinet 40*(6), 411–426.
75. Kendall MJ (2003). Review article: Esomeprazole—The first proton pump inhibitor to be developed as an isomer. *Aliment Pharmacol Ther 17*(Suppl 1), 1–4.
76. Miner P Jr, et al. (2003). Gastric acid control with esomeprazole, lansoprazole, omeprazole, pantoprazole, and rabeprazole: A five-way crossover study. *Am J Gastroenterol 98*(12), 2616–2620.
77. Devlin JW (2005). Proton pump inhibitors for acid suppression in the intensive care unit: Formulary considerations. *Am J Health Syst Pharm 62*(10, Suppl 2), S24–30.
78. Howden CW (2005). Review article: Immediate-release proton-pump inhibitor therapy—Potential advantages. *Aliment Pharmacol Ther 22*(Suppl 3), 25–30.

79. Katz PO (2005). Review article: Putting immediate-release proton-pump inhibitors into clinical practice—Improving nocturnal acid control and avoiding the possible complications of excessive acid exposure. *Aliment Pharmacol Ther 22*(Suppl 3), 31–38.
80. Morgan D (2002). Intravenous proton pump inhibitors in the critical care setting. *Crit Care Med 30*(6, Suppl), S369–372.
81. Valuck RJ, Ruscin JM (2004). A case-control study on adverse effects: H2 blocker or proton pump inhibitor use and risk of vitamin B12 deficiency in older adults. *J Clin Epidemiol 57*(4), 422–428.
82. Yang YX, et al. (2006). Long-term proton pump inhibitor therapy and risk of hip fracture. *JAMA 296*(24), 2947–2953.
83. Dial S, et al. (2004). Risk of *Clostridium difficile* diarrhea among hospital inpatients prescribed proton pump inhibitors: Cohort and case-control studies. *CMAJ 171*(1), 33–38.
84. Cunningham R, et al. (2003). Proton pump inhibitors as a risk factor for *Clostridium difficile* diarrhoea. *J Hosp Infect 54*(3), 243–245.
85. Laheij RJ, et al. (2004). Risk of community-acquired pneumonia and use of gastric acid-suppressive drugs. *JAMA 292*(16), 1955–1960.
86. Gomollon F, Calvet X (2005). Optimising acid inhibition treatment. *Drugs 65*(Suppl 1), 25–33.
87. Sachs G, Shin JM, Howden CW (2006). Review article: The clinical pharmacology of proton pump inhibitors. *Aliment Pharmacol Ther 23*(Suppl 2), 2–8.
88. Vakil N, Fennerty MB (2003). Direct comparative trials of the efficacy of proton pump inhibitors in the management of gastro-oesophageal reflux disease and peptic ulcer disease. *Aliment Pharmacol Ther 18*(6), 559–568.
89. Klok RM, et al. (2003). Meta-analysis: Comparing the efficacy of proton pump inhibitors in short-term use. *Aliment Pharmacol Ther 17*(10), 1237–1245.
90. Edwards SJ, Lind T, Lundell L (2001). Systematic review of proton pump inhibitors for the acute treatment of reflux oesophagitis. *Aliment Pharmacol Ther 15*(11), 1729–1736.
91. Zed PJ, et al. (2001). Meta-analysis of proton pump inhibitors in treatment of bleeding peptic ulcers. *Ann Pharmacother 35*(12), 1528–1534.
92. Leontiadis GI, Sharma VK, Howden CW (2005). Systematic review and meta-analysis of proton pump inhibitor therapy in peptic ulcer bleeding. *BMJ 330*(7491), 568.
93. Nardino RJ, Vender RJ, Herbert PN (2000). Overuse of acid-suppressive therapy in hospitalized patients. *Am J Gastroenterol 95*(11), 3118–3122.
94. Naunton M, Peterson GM, Bleasel MD (2000). Overuse of proton pump inhibitors. *J Clin Pharm Ther 25*(5), 333–340.
95. Niklasson A, et al. (2003). Overuse of acid suppressive therapy in hospitalised patients with pulmonary diseases. *Respir Med 97*(10), 1143–1450.
96. Parente F. et al. (2003). Hospital use of acid-suppressive medications and its fall-out on prescribing in general practice: A 1-month survey. *Aliment Pharmacol Ther 17*(12), 1503–1506.
97. Guda NM, et al. (2004). Use of intravenous proton pump inhibitors in community practice: An explanation for the shortage? *Am J Gastroenterol 99*(7), 1233–1237.
98. Hunt RH (2005). Review article: The unmet needs in delayed-release proton-pump inhibitor therapy in 2005. *Aliment Pharmacol Ther 22*(Suppl 3), 10–19.
99. Vakil N (2004). Review article: New pharmacological agents for the treatment of gastro-oesophageal reflux disease. *Aliment Pharmacol Ther 19*(10), 1041–1049.
100. Katz PO, Scheiman JM, Barkun AN (2006). Review article: Acid-related disease—What are the unmet clinical needs? *Aliment Pharmacol Ther 23*(Suppl 2), 9–22.
101. Scarpignato C, Pelosini I, Di Mario F (2006). Acid suppression therapy: Where do we go from here? *Dig Dis 24*(1–2), 11–46.
102. Scarpignato C, Pelosini I (2006). Review article: The opportunities and benefits of extended acid suppression. *Aliment Pharmacol Ther 23*(Suppl 2), 23–34.
103. Inamori M, et al. (2005). Early effects of lafutidine or rabeprazole on intragastric acidity: Which drug is more suitable for on-demand use? *J Gastroenterol 40*(5), 453–458.
104. Tsuji S, Kawano S (2005). A new-generation H2 receptor antagonist: Quicker and stronger acid inhibition than proton pump inhibitors in the clinical setting? *J Gastroenterol 40*(5), 549–550.

# PHARMACOLOGY OF VASOPRESSIN AND THE RENIN–ANGIOTENSIN–ALDOSTERONE SYSTEM

Elizabeth J. Brant, Albert Q. Lam, and Julian L. Seifter
*Renal Division, Brigham and Women's Hospital, Boston, Massachusetts*

Pharmacological agents targeted toward renal pathophysiology have had a profound effect in medicine, serving to advance the understanding of both renal disease and normal physiology, and significantly improve disease outcomes in many circumstances. A review of the mechanisms of action of several of these agents will be reviewed here. In addition, new agents active in major hormonal pathways, including antidiuretic hormone (ADH) and the renin–angiotensin system will be introduced.

## 1 ADH ANTAGONISTS

The antidiuretic hormone (ADH), or vasopressin, is a nonapeptide hormone synthesized in the hypothalamus and secreted by the posterior pituitary [1]. It initiates several physiological mechanisms by acting at any of several receptor types, depending on tissue localization [1–4]. ADH stimulation of V1a receptors results predominantly in cardiovascular changes, including vasoconstriction, platelet aggregation, increased force of cardiac contraction, and increased protein synthesis in the myocardium. V1b receptor stimulation increases adrenocorticotropic hormone (ACTH) secretion by the hypothalamus. The primary function of ADH is maintenance of water balance, which is achieved by stimulation of V2 receptors. The net effect is uptake of free water in the collecting duct of the kidney. ADH receptors are found in numerous other tissues, and the role of ADH continues to be elucidated. However, the important role of ADH in water regulation will be the focus of this discussion.

The system by which ADH facilitates water balance is an elegant one. The distal collecting duct of the kidney is, at baseline, impermeable to water [5]. In the setting of increased plasma osmolality or decreased effective circulating volume, ADH levels rise, with subsequent binding of ADH to V2 receptors on the basolateral membrane of principal cells of the collecting duct [4, 5]. The result of binding is the G-coupled protein-mediated synthesis of cyclic adenosine monophosphate (cAMP), with subsequent activation of protein kinase A. Protein kinase A in turn promotes fusion with the apical membrane of vesicles containing preformed aquaporin-2, rendering the heretofore impermeable collecting duct membrane permeable to water [6]. In addition, the rise in cAMP promotes increased transcription of aquaporin-2 [4]. Simultaneous insertion of aquaporin-3 and aquaporin-4 channels into the basolateral membrane allows for the transcellular movement of water, a process driven by the sodium osmotic gradient. As ADH levels decrease in response to restoration of plasma osmolality and/or volume, the aquaporins are endocytosed, and the collecting duct membrane again becomes impermeable to water.

Disturbance of the ADH-mediated pathway that increases activity of ADH leads to a disruption of water balance, as reflected in the development of hyponatremia, and can be seen in such conditions as the syndrome of inappropriate antidiuretic hormone secretion (SIADH), heart failure, and liver disease. The

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

sequelae of hyponatremia can include mild neurological deficits, such as impaired balance and attention, or even increased morbidity and mortality in the context of cardiac, hepatic, or neurologic disease. Since correction of hyponatremia can ameliorate or even reverse some of these symptoms, medications that can interrupt this pathway are potentially beneficial not only in treating hyponatremia but also in improving survival.

Historically, demeclocycline, a tetracycline antibiotic derivative, has found utility in some cases of SIADH, as well as in cases of paraneoplastic production of ADH-like peptides [7, 8]. Sodium normalization is due to the development of a reversible nephrogenic diabetes insipidus but does not become apparent until 1–2 weeks after initiation of treatment. Lithium, which can also antagonize ADH activity, is not typically used clinically for this purpose because of an unpredictable response and the risk of adverse outcomes, including tremor, cognitive impairment, cardiotoxicity, thyroid dysfunction, and leukocytosis. Like demeclocycline, it causes a reversible nephrogenic diabetes insipidus. Demeclocycline is more predictable than lithium but still requires regular monitoring of plasma levels in order to maintain a level of 2 μg/mL. The most common and important adverse effects of both these medications are severe hypernatremia and nephrogenic diabetes insipidus, the latter being the effect that permits sodium normalization. In addition, there have been reports of acute kidney injury with both drugs, as well as cases of chronic interstitial nephritis with lithium administration. Because of the difficulties in correcting hyponatremia using these medications, typical management involves fluid restriction, liberal salt intake, and hypertonic saline in severe cases.

Drugs are currently being developed that act selectively as antagonists in the ADH pathway, the rationale being that such drugs offer the prospect of decreasing fluid overload and correcting sodium balance without causing or exacerbating other electrolyte disturbances, such as hypokalemia, or inducing renal dysfunction [9], as is seen with loop diuretics, a mainstay of treatment for heart failure and previously the only therapy available to decrease congestion-inducing fluid overload in patients with heart failure [10, 11].To briefly review, loop diuretics, such as furosemide, block sodium uptake in the thick ascending limb of the loop of Henle. The consequent increased sodium delivery to the distal nephron activates tubuloglomerular feedback and the renin–angiotensin–aldosterone axis, with resultant constriction of the afferent arteriole. Renal blood flow can thus be reduced to the point of compromising renal function, which is in itself a significant independent risk factor for mortality in patients with chronic heart failure. Furthermore, decreased renal function can limit the use of angiotensin-converting enzyme (ACE) inhibitors, known to be effective in increasing survival in patients with heart failure. Finally, loop diuretics can lead to potassium wasting, a risk factor for arrhythmias. The hypothesis was that a compound that could inhibit free water reabsorption at the level of the V2 receptor could effect the desired change of reduction of total body water without affecting sodium or potassium balance and without activating tubuloglomerular feedback.

Research has been conducted for more than 20 years in an attempt to develop effective ADH antagonists but with little success until recently [3]. While some formulations worked well in certain animal models, they did not work as well in others and sometimes not at all in human trials [7, 12]. In addition, the response to individual preparations could change with prolonged administration, notably converting from antagonists to agonists. The vaptans, a new class of agents, are proving to be selective and effective ADH antagonists in human trials, and, to date, two of them have been approved for use in the United States: conivaptan (intravenous) and tolvaptan (oral). Other agents (e.g., lixivaptan, satavaptan) are still undergoing investigation for potentially similar applications.

The vaptans are nonapeptide V2 receptor antagonists, and all are benzazepine derivatives [13]. As the name suggests, benzazepines are related to a common family of compounds that includes benzazepinones and benzodiazepines. Overall, this family of molecules affords a number of pharmacological properties, including antitumor activity, analgesia, and, as will be discussed here, vasopressin V2 receptor antagonism. ADH itself binds the transmembrane V2 receptor at the extracellular surface. Even though there is overlap in native ADH and antagonist binding sites, the ADH antagonists do not function by changing the affinity of ADH for its receptors [14]. Rather, the nonpeptide antagonists binding at a different site deep within the transmembrane portion of the receptor. Specificity of receptor binding by the antagonists is primarily related to differences in the shape of the binding pocket between receptor types. Affinity of the individual antagonists is further manipulated by individual amino acid substitutions [15].

Both oral and intravenous conivaptan (YM-087), a combination V1/V2 receptor antagonist, are able to effect an approximately sevenfold increase in urine flow rate, with concomitant decrease in urine osmolality from 600 mosm/L to less than 100 mosm/L, in healthy, normotensive male subjects [16]. There is an associated significant increase in plasma osmolality, with similar degrees of normalization using either formulation. As would be expected, serum ADH levels also increase significantly with conivaptan administration. Peak effect

occurs approximately 2 h after drug administration, with a return to baseline at 12 h. These findings suggest effective inhibition of renal V2 receptors. V1 receptor antagonism, as suggested by the inhibition of vasopressin-induced skin vasoconstriction, was at the limit of significance with intravenous administration of conivaptan, and was transient and statistically nonsignificant with oral conivaptan.

Intravenous conivaptan (50 mg) achieves a mean maximum serum concentration of 2681 ng/mL. The mean apparent volume of distribution is 34 L, with a mean distribution half-life of 0.15 h. The mean half-life of the terminal disposition phase is 3.1 h [16]. A multicenter, randomized, double-blind, placebo-controlled study of 84 patients with euvolemic or hypervolemic hyponatremia (serum sodium 115 to <130 mEq/L) investigated the efficacy and tolerability of intravenous conivaptan in these two groups of patients [17]. Patients assigned to receive conivaptan were administered a 20-mg loading dose administered over 30 min, followed by either 40 or 80 mg per day via a 96-h infusion. The primary efficacy endpoint of change in serum sodium concentration, as assessed by the baseline-adjusted area under the sodium–time curve (AUC), was significantly greater in both groups of patients receiving conivaptan ($6.3 \pm 0.7$ mEq/L with 40 mg/day and $9.4 \pm 0.8$ mEq/L with 80 mg/day) as compared with patients receiving placebo ($0.8 \pm 0.8$ mEq/L). The difference between the two conivaptan groups was not statistically significant, suggesting that the lower dose of intravenous conivaptan is reasonably effective in increasing serum sodium in most patients. Conivaptan also proved to be significantly superior to placebo in all secondary endpoints, including time from initial dose to achieve a 4-mEq/L increase in serum sodium (median 24 h with either dose of conivaptan; achieved at end of study with placebo in a few patients, therefore not estimable); total time a serum sodium 4 mEq/L above baseline was maintained (53 h with 40 mg and 73 h with 80 mg vs. 14 h with placebo); change in serum sodium from baseline to end of treatment; and number of patients with normalized (135 mEq/L) sodium or minimum 6 mEq/L increase in serum sodium concentration (20 patients receiving 40 mg/day, 23 patients receiving 80 mg/day, and 6 patients receiving placebo). These findings, in conjunction with generally good tolerability, suggest that intravenous conivaptan is an effective means of increasing serum sodium in patients with euvolemic or hypervolemic hyponatremia.

Oral conivaptan (60 mg) reaches a mean maximum serum concentration of 664 ng/mL within approximately 1 h after administration [16]. The mean distribution half-life is 1.8 h, with a mean terminal disposition phase half-life of 5.6 h. Bioavailability of oral conivaptan is approximately 44%. A separate 5-day randomized, double-blind, placebo-controlled study was also conducted to assess the efficacy and safety of 40 or 80 mg of oral conivaptan administered in two divided doses versus placebo in patients with euvolemic or hypervolemic hyponatremia [18]. Patients in this study also had a baseline serum sodium concentration of 115 to <130 mEq/L, and the primary endpoint was again change in baseline serum sodium concentration AUC. As occurred with intravenous conivaptan, the increase in AUC was significantly greater with 40 mg (2-fold) or 80 mg (2.5-fold) of conivaptan than with placebo. Secondary endpoints were the same as in the prior study and also demonstrated similar results. Specifically, the median time to achieve a 4 mEq/L or greater increase from baseline in serum sodium concentration was significantly less with conivaptan (27 h with 40 mg and 12 h with 80 mg) than with placebo (72 h). Second, this 4 mEq/L increase in sodium concentration was maintained significantly longer with 80 mg of conivaptan (88.8 h) than with placebo (46.5 h), although not with 40 mg of conivaptan (69.8 h) versus placebo. A third measure, mean change from baseline to end of treatment of serum sodium concentration was also statistically significant for 80 mg of conivaptan (8.2 mEq/L) versus placebo (3.4 mEq/L), but again not for 40 mg of conivaptan (6.4 mEq/L) versus placebo. Finally, normalization of serum sodium concentration (≥135 mEq/L) or a minimum increase of 6 mEq/L was achieved in 82% of patients receiving 80 mg of conivaptan as compared with 48% of patients receiving placebo, a difference that was statistically significant. The comparison of 40 mg of conivaptan (71% of patients meeting criteria) with placebo again did not meet statistical significance. The above findings suggest that oral conivaptan is also effective in achieving normalization or improvement of serum sodium concentration but may require a higher dose, a finding that correlates with the known bioavailability of the drug. The most common adverse events in patients receiving conivaptan included headache, hypotension, postural hypotension, nausea, and constipation and occurred similarly in the placebo and conivaptan groups. Conivaptan was nevertheless generally well tolerated.

Conivaptan received Food and Drug Administration (FDA) approval for short-term (not to exceed 4 days) in-hospital parenteral use in 2005 [4, 11]. It was originally approved for treatment of hyponatremia in euvolemic patients, but hypervolemic hyponatremia was subsequently added to its indications. Oral conivaptan, as with all of the vaptans, is a substrate of cytochrome (CYP) P450 3A4. Conivaptan appears to be the most potent inhibitor of this microsomal enzyme among this class of drugs. Therefore, there is

potential for serious interactions with a number of other drugs, including some chemotherapeutic agents, calcium channel blockers, 3-hydroxy-3-methyl-glutaryl-coenzyme A reductase inhibitors, benzodiazepines, and immunosuppressants, as well as amiodarone, digoxin, and warfarin. Therefore, conivaptan, in spite of oral efficacy, has been limited to parenteral use to date. Some of the other vaptans appear to have fewer CYP P450 3A4 interactions and still offer the prospect of oral use. Trials are currently underway to develop oral formulations of these agents.

Tolvaptan (OPC-41061), another member os this class of drugs, represents an N-substituted member of the benzazepine family [13, 15]. Tolvaptan has 30 times greater selectivity for the V2 receptor than for the V1a receptor and displays linear pharmacokinetics. The oral formulation is rapidly absorbed, reaching peak plasma concentrations in 2–3 h. It has a half-life of 6–8 h, with clearance decreased by 87% when administered with ketoconazole. Maximum serum concentrations are observed to be significantly higher in patients with heart failure after a single dose (average maximum concentration of 374 and 418 ng/mL after a single dose of 30 and 90 mg, respectively, in healthy subjects; 460 and 723 ng/mL in subjects with heart failure at the same doses). There are no significant interactions with amiodarone or warfarin.

Tolvaptan was considered as a possible agent for correction of water and electrolyte balance, particularly in patients with heart failure because of its V2 receptor antagonism [19]. The first prospective, randomized, double-blind, placebo-controlled trial to look at the effects of tolvaptan in this patient population was the ACTIV (Acute and Chronic Therapeutic Impact of a Vasopressin Antagonist in Congestive Heart Failure) trial [20]. Effects attributable to activation at the V2 receptor had previously been noted to occur at lower concentrations of ADH than those potentiated by the V1 receptors [21]. Inhibition of the ADH pathway at the level of the V2 receptor was hypothesized to result in a pure water diuresis, without associated decrease in electrolytes and without triggering of neurohumoral cascades seen with loop diuretics. The primary focus of the study was the evaluation of the diuretic potential of tolvaptan, as assessed by body weight (as a marker of fluid volume), edema, serum sodium, serum potassium, and renal function [20]. Patients with New York Heart Association (NYHA) class III or IV heart failure, left ventricular ejection fraction of less than 40% within the preceding 12 months, and at least two signs of heart failure were randomly assigned to receive either tolvaptan or placebo, in addition to standard heart failure regimens, including diuretics, for up to 10 days of hospitalization, then for 7 weeks following discharge. The endpoints were a change in body weight at 24 h of hospitalization for inpatients and clinically worsened heart failure for outpatients.

Patients receiving tolvaptan had a significant decrease in body weight at 24 h compared with patients receiving placebo, a response that was independent of both dose and left ventricular ejection fraction. Although there was no further decrease in body weight throughout the study period, the initial loss was maintained in the majority of patients. Tolvaptan also provoked dose-dependent increased production of dilute urine, an effect only assessed on the first day of therapy, and edema was improved. Serum potassium levels and renal function were unchanged. The effect of tolvaptan on serum sodium was interesting and somewhat surprising. Tolvaptan was expected to produce a pure water diuresis, but an increase in total sodium excretion was demonstrated, likely attributable to the increase in urine volume and flow. Patients were also noted to increase their fluid intake during administration of tolvaptan. Nevertheless, there was a net decrease in fluid balance at 24 h and a decrease in urine osmolality. In the final analysis, serum sodium normalized in patients with hyponatremia, and even though the effect on sodium waned over the course of the study, levels remained in the normal range for the majority of patients. In comparison, hyponatremia persisted or worsened in patients receiving placebo. Patients who had normal serum sodium levels at the start of the study and who received tolvaptan were observed to have a transient elevation in sodium levels that resolved within 3 weeks.

The ACTIV trial offered some optimistic findings for the treatment of heart failure using tolvaptan. However, a number of questions remained. Patients in the trial were noted to have a decreased response to therapy in terms of body weight and sodium normalization. Because monitoring only continued for 7 weeks after hospital discharge, it is unknown what will be the response to treatment for a more extended period. This question was of particular importance, given early studies of these compounds that demonstrated agonist activity with prolonged administration in animal models. Also, no neurohumoral activation was observed in studies of tolvaptan in animals and healthy subjects, but it was unknown if these same results would be observed in patients with heart failure or in other pathological states. In the relatively short time period of this trial, quality of life was not determined to be different in tolvaptan-treated patients versus those receiving placebo. Ultimately, given the significance of hyponatremia and neurohumoral activation as risk factors for mortality in patients with heart failure, longer studies assessing outcomes in patients receiving tolvaptan were warranted.

The EVEREST (Efficacy of Vasopressin Antagonism in Heart Failure Outcome Study with Tolvaptan) trial sought to answer some of these questions raised by the ACTIV trial [22]. EVEREST, a large international prospective, randomized, double-blind, placebo-controlled trial, looked at longer-term outcomes in patients with heart failure in whom tolvaptan administration was initiated within 48 h of onset of symptoms of decompensation. The study enrolled 4133 patients who were randomized to receive either oral tolvaptan 30 mg/day (2072 patients) or placebo (2061 patients) for a minimum of 60 days, along with a standard regimen of heart failure medications, including diuretics, as deemed appropriate by their treating physician. The trial overall comprised two short-term studies that looked at the effects of tolvaptan on clinical status and symptoms, as well as an outcome study that evaluated all patients randomized to the short-term studies with regard to both short- and long-term outcomes. The primary endpoints for the outcome study were all-cause mortality and the composite of cardiovascular death or hospitalization for heart failure. Consistent with the findings of the ACTIV trial, patients receiving tolvaptan were observed to have significant improvement in signs and symptoms of heart failure compared with placebo in the early phase of the study. Dyspnea and body weight improved at day 1, and the latter was maintained well beyond hospitalization. Likewise, edema improved from day 1 throughout postdischarge week 4. Serum sodium concentration increased in patients with baseline hyponatremia (sodium less than 134 mEq/L), an effect that was also maintained throughout 40 weeks of treatment. Serum urea nitrogen levels decreased in patients receiving tolvaptan as compared with placebo, with little effect on serum creatinine levels, thus supporting the previously observed finding of stable renal hemodynamics in patients receiving tolvaptan. Although several domains of quality of life as measured by the Kansas City Cardiomyopathy Questionnaire (KCCQ) improved significantly, a number of them (clinical summary, physical limitation, total symptoms, symptom frequency, symptom burden, symptom stability, and self-efficacy) did not. Furthermore, the overall quality-of-life score did not show a statistically significant difference compared with placebo. Of the adverse events reported, only thirst occurred with significantly increased frequency in the tolvaptan group. Dry mouth was also more common in the tolvaptan group, as was hypernatremia, but the occurrence of renal failure and hypotension was similar between the tolvaptan and placebo groups.

The scope of this study extended beyond that of the ACTIV trial with the analysis of long-term outcomes in patients receiving tolvaptan. While long-term administration of tolvaptan did not prove to be inferior to placebo with respect to mortality, it also did not demonstrate any favorable effect on all-cause mortality, cardiovascular mortality, or hospitalization for decompensated heart failure. Therefore, tolvaptan, via V2 receptor antagonism, may have utility in the acute setting of worsening heart failure for safely promoting fluid removal, without renal or electrolyte compromise, but does not thus far demonstrate efficacy in the reduction of mortality. There was a relatively high rate of discontinuance of the study drug, which may have diluted the findings of a positive effect on outcome. Also, the question of benefit in a subgroup of patients with hyponatremia is still not clearly answered since in this study, as with the ACTIV trial, few patients met the criterion for severe hyponatremia. This specific question has been addressed in two recent studies, discussed below.

The Study of Ascending Levels of Tolvaptan in Hyponatremia 1 and 2 (SALT-1 and SALT-2) were identical (for purposes of reproducibility and comparability), randomized, double-blind, and placebo-controlled studies looking at the effects of tolvaptan in the treatment of hypervolemic or euvolemic hyponatremia secondary to chronic heart failure, cirrhosis, or SIADH [23]. Unlike previous studies, while initiated in a hospital setting, the bulk of the data was collected on an outpatient basis. All patients included had a serum sodium concentration of less than 135 mmol/L and were stratified as having either mild (serum sodium of 130–135 mmol/L) or marked (serum sodium <130 mmol/L) hyponatremia. Patients with a serum sodium concentration of less than 120 mmol/L with associated neurologic impairment were excluded from the study. A distinction was also made between patients who had heart-failure-associated hyponatremia and those who did not. The two primary endpoints of the study were the change in average area under the curve for the serum sodium concentration from baseline to day 4 and from baseline to day 30.

At the conclusion of the study, patients receiving tolvaptan demonstrated a significantly increased area under the curve for serum sodium concentration as compared with placebo. This result was observed within 8 h of administration of the first dose of study drugs, as well as at all subsequent visits throughout the study. Furthermore, these results were achieved more quickly in the tolvaptan-treated group and were maintained across subgroups independent of whether the hyponatremia was mild or marked at baseline. Thus, a significantly greater number of patients receiving tolvaptan had normal serum sodium levels when assessed on day 4 and on day 30 than did patients receiving a placebo. Only four patients given tolvaptan achieved undesirably

rapid rates of sodium correction, and only four patients exceeded 146 mmol/L of sodium, the predefined criterion for a potentially serious elevation of sodium concentration. Less fluid restriction appeared to be required in patients receiving tolvaptan than in those receiving placebo, particularly in cases of marked hyponatremia. Finally, patients receiving tolvaptan scored higher than patients receiving placebo on the Mental Component Summary of the SF-12 Health Survey, reaching statistical significance in the combined analysis and in the SALT-1 analysis, although not in isolated SALT-2 analysis. Again, patients in the combined subgroup with marked hyponatremia showed improved scores of statistical significance. As had been seen in other studies of tolvaptan, thirst and dry mouth were the most common adverse events, while other adverse events, including death, had a similar incidence in both the tolvaptan and placebo groups. In the 7 days of follow-up, during which time the study drug was discontinued, hyponatremia recurred in parallel between the tolvaptan- and placebo-treated groups. These studies suggest that antagonism of the V2 receptor with tolvaptan is effective in the short term for treating hyponatremia independent of etiology. Tolvaptan was officially approved for use by the FDA in 2009.

As discussed previously, hyponatremia has been demonstrated to be an independent risk factor for mortality in patients with heart failure, cirrhoses, and neurologic disease. In addition, hyponatremia is itself a cause of subtle neurologic impairment, which confers a subsequent fall risk. Thus, correction of hyponatremia might plausibly reduce associated morbidity and offer a survival benefit. Studies to determine that possibility remain to be done. Recent trials, in addition to those discussed here, offer cause for optimism in this regard.

## 2 RENIN–ANGIOTENSIN SYSTEM

The classical model of the renin–angiotensin system (RAS), more specifically called the systemic RAS, comprises several peptides that work in concert to maintain arterial blood pressure and blood volume [1]. The systemic RAS, particularly angiotensin II, is rapidly acting and serves more in a compensatory fashion in the acute setting, such as hemorrhage, than in response to moment-by-moment fluctuations of hemodynamics. Renin, a proteolytic enzyme produced by the cells of the juxtaglomerular apparatus, is released in response to decreased pressure and stretch at renal baroreceptors, and in response to direct sympathetic outflow to the juxtaglomerular apparatus, such as would occur in the setting of low blood pressure or blood volume. Renin has an indirect influence on hemodynamics, acting to set into motion a cascade of events that begins with renin's cleavage of angiotensinogen to angiotensin I. Angiotensinogen is a proenzyme synthesized in the liver and readily available for conversion once renin enters the circulation. Angiotensin I is a decapeptide (angiotensin 1–10) and, like renin, has minimal direct biological activity, but is the precursor of angiotensin II, a key mediator in the regulation of cardiac and renal hemodynamics.

Angiotensin II is an octapeptide. Again referring to the classical model, the conversion of angiotensin I to angiotensin II is effected by the angiotensin-converting enzyme, which is produced predominantly by vascular endothelium of the pulmonary vasculature. Like angiotensin I, ACE is readily available once renin release has been stimulated. Angiotensin II functions directly as a potent vasoconstrictor in renal, cutaneous, and splanchnic vascular beds by activating angiotensin II type 1 (AT1) receptors, leading to an increase in systemic vascular resistance and thus raising blood pressure. (Importantly, AT1 receptors are absent or sparse in the coronary, cerebral, and pulmonary vasculature, thus protecting these systems from the potentially deleterious effects of increased vascular resistance and elevated blood pressure.) ACE also plays a role in bradykinin inactivation. Bradykinin is itself a vasodilator and also promotes synthesis of prostaglandins, another vasodilator. Therefore, in the setting of ACE activity, two pathways converge to produce increased systemic vascular resistance and thus increase blood pressure. First, ACE catalyzes the conversion of angiotensin I to angiotensin II, a potent vasoconstrictor. Second, by inactivating bradykinin, ACE interrupts a pathway that would otherwise decrease peripheral vascular resistance and thereby reduce blood pressure. Angiotensin II acts directly and indirectly to increase intravascular volume. Directly, it leads to enhanced sodium reabsorption in the renal proximal tubule via activation of AT1 receptors. It functions indirectly through two mechanisms. The first of these is stimulation of hypothalamic thirst centers, with consequent increased fluid intake and thus increased blood volume. Second, it promotes production of aldosterone, a mineralocorticoid hormone, in the adrenal cortex. Aldosterone in turn leads to enhanced sodium reabsorption in the renal collecting duct, with associated increased reabsorption of water.

It is tempting to present only this classical model of the RAS, but to do so would forfeit the opportunity to review newer data regarding additional mediators, as well as novel actions of the well-known players, of this system. This new knowledge permits an occasion to discuss the evolving understanding of the mechanisms of action of commonly used inhibitors of the RAS, to present new drugs that are proving beneficial in thwart-

ing the harmful effects of the system, and to consider tantalizing prospects regarding potential therapeutic targets.

Newer and rapidly evolving data clearly indicate that the systemic RAS is much more complex and exerts far greater influence than previously realized. Data from human and animal studies demonstrate that not only does each component in the system—renin, angiotensin, and aldosterone—have more diverse actions than described in the classical model of the system, but there are in fact many more players, some of which are proving to have relevance in the pathologies previously attributed to the classical model alone. Indeed, their role in pathology can be attributed in large part to the newfound realization that there are tissue-specific and intracellular renin–angiotensin systems (RAS) at work, and the interrelationship of the systemic and local systems, as well as of the individual elements of these systems, are far more complex and integrated than the classical model would suggest [24–27]. Angiotensin I, despite its lack of biological activity, can be considered the hub of the axis. It is the actions of various enzymes, not just ACE, on angiotensin I that directly or ultimately lead to the several known biologically active constituents of the RAS [24–26, 28–30]. To begin, as described above, cleavage by ACE of two amino acids from angiotensin I yields angiotensin II. However, it is now known that angiotensin II is the product not only of ACE activity but of at least one additional enzyme, chymase. The resultant peptide, angiotensin II, is identical to that produced by ACE. In addition, angiotensin I is not limited to conversion to angiotensin II. Rather, a third enzyme, ACE-2, cleaves a single amino acid from angiotensin I to produce angiotensin 1-9. To better understand the newer paradigm of the RAS as represented in Figure 1, the fate and function of each of these and succeeding components will be considered briefly.

The vasoconstrictor and salt- and water-retaining functions of systemic angiotensin II are well known and have been described previously. Intracellular and local renin–angiotensin systems have been identified in a multitude of tissues, including reproductive, hematologic, cardiac, renal, hepatic, enteric, pancreatic, and brain tissue. The effects of local angiotensin II have best

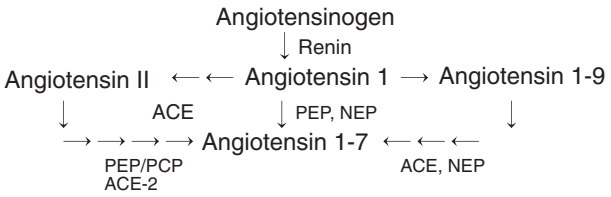


**Figure 1** New paradigm for the renin–angiotensin system.

been described in the vasculature, kidneys, and heart. Angiotensin II is now identified as a significant mediator of inflammation, stimulating the synthesis and activity of such inflammatory modulators as transforming growth factor-ß (TGF-ß) and nuclear factor-κB (NF-κB) [24, 29, 31, 32]. It both stimulates and interacts with reactive oxygen species (ROS), further fueling the production of NF-κB. Angiotensin II's relationship with ROS is important, as some studies demonstrate that even relatively brief exposure to angiotensin II decreases cell viability, likely secondary to oxidative stress. Additionally, angiotensin II can stimulate cellular hypertrophy, proliferation, or apoptosis, depending on the chronicity and level of angiotensin II, as well as the environmental milieu [33]. Acting principally as a cytokine, angiotensin II induces multiple mediators of fibrosis via the AT1 receptor. Further, in a single study, angiotensin II was shown to upregulate 48 stress-related genes [32]. Many of these are genes for heat-shock proteins, but genes for transcription factors and cell-signaling peptides are also represented, as are genes that participate in deoxyribonucleic (DNA) repair and/or replication. While many of these gene products when chronically stimulated have injurious effects, some of them confer a degree of protection by acting as chaperones or antioxidants, as well as by inhibiting apoptosis.

Angiotensin II production by chymase appears to have an effect at the local level [25, 26, 28, 29, 34, 35]. The question of an alternative pathway of angiotensin II production was raised when it was noted that ACE inhibitors (ACEIs) do not completely block the production of angiotensin II at the tissue level [34]. For example, in one study using ACE knockout mice, the level of angiotensin II in the kidneys remained stable in spite of the absence of ACE [25]. This finding was attributable to significant (14-fold) increase in chymase. Local production of angiotensin II in the heart and the vasculature is also thought to be due largely to chymase activity. Furthermore, chymase and ACE have been demonstrated to be differentially expressed in specific cell types and cellular compartments, as well as in the setting of specific pathological conditions. For example, in the normal human kidney, ACE is expressed in mesangial cells, vascular smooth muscle cells, and tubular epithelial cells, whereas chymase, while also expressed in mesangial cells and vascular smooth muscle cells, is not expressed at all in tubular epithelial cells [34].

Angiotensin II, whether produced by ACE or chymase, can be further metabolized to angiotensin III or even angiotensin IV [28, 35]. Angiotensin III functions similarly to angiotensin II, causing vasoconstriction and stimulating aldosterone secretion [28]. However, angiotensin III is less potent than angiotensin II, and it may participate in blood pressure regulation

by acting centrally. Also like angiotensin II, angiotensin III may stimulate protection of TGF-ß and other growth factors and may also participate in inflammatory processes by acting as a chemoattractant for polymorphonuclear cells. Some studies suggest that angiotensin IV has vasodilatory properties. Also, like angiotensin II, angiotensin IV appears to stimulate cell proliferation, as demonstrated by studies looking at effects in rat pituitary.

ACE-2 is an 805-amino-acid ACE homolog that is distributed throughout a number of disparate tissues, including the heart, lung, vasculature, kidneys, and gastrointestinal tract [30]. The function of ACE-2 most commonly described is cleavage of angiotensin I to form angiotensin 1-9 [36]. Although angiotensin 1-9 is not well characterized and may have limited biological activity, it can be further metabolized by ACE to form angiotensin 1-7, a highly active product that will be discussed in greater detail. (Note that this conversion can also be accomplished by a neutral endopeptidase) [30]. In addition to these actions, ACE-2 can cleave angiotensin II, again forming angiotensin 1-7. This latter process is believed to be the primary role of ACE-2, based on the very high affinity of ACE-2 for angiotensin II. ACE-2 also has demonstrated interactions with multiple other vasoactive peptides, the most notable of which is the interaction with apelin-13, a powerful inotrope. The products of ACE-2 activity are observed to act primarily as vasodilators. Thus, ACE-2 appears to represent the junction of a counterregulatory arm of the RAS [37].

Angiotensin 1-7, as described above, is the direct product of ACE-2-mediated cleavage of angiotensin II. It can also result from conversion of angiotensin 1-9 by ACE and by direct alteration of angiotensin I by prolylendopeptidase and neutral endopeptidase 24.11 [30]. The first known action of angiotensin 1-7 was stimulation of vasopressin release [38]. More recently, angiotensin 1-7 has been demonstrated to be a highly active peptide with diverse and significant functions, most notably in the cardiovascular system but possibly in the kidney, as well. In direct opposition to the actions of angiotensin II, angiotensin 1-7 has been shown to have both antitrophic and antifibrotic properties [30]. Furthermore, it has been shown to improve cardiac function in both normal and heart failure animal models. Angiotensin 1-7 was also demonstrated in animal models to improve contractile function in postischemic reperfused hearts [39]. It has been shown to have antiarrhythmic properties, as well as to diminish cardiac remodeling, in part by decreasing myocyte hypertrophy and limiting collagen deposition [30]. In the kidney, animal models suggest that angiotensin 1-7 is protective in the setting of concomitant hypertension and diabetes [40]. Benefits of angiotensin 1-7 in the heart and kidney may be effected through modulation of oxidative stress.

Thus, the landscape of the RAS has changed dramatically. There are several new players in the RAS, and likely more to follow, whose functions are just now being elucidated. Furthermore, the commonly known peptides have newly discovered activity that may hold relevance in understanding the pathogenesis and pathophysiology of some diseases. There are already drugs available that interrupt the system at every step of the cascade, including ACEIs, angiotensin receptor blockers (ARBs) and, more recently, renin inhibitors. Furthermore, many of these drugs are proving to have benefits well beyond the reduction of blood pressure and blood volume. Having a more complete understanding of the RAS offers not only some possible explanations for these observed benefits but also may provide further inroads for treatment and prevention of disease.

## 3 ACE INHIBITORS

Development of ACE inhibitors (ACEIs) began in the 1960s with the observation that the venom of the Brazilian pit viper, *Bothrops jararaca*, appeared to exert its toxic effect by causing a massive fall in blood pressure [7, 41]. Research demonstrated that the venom comprises, in particular, proline-rich peptides originally thought to work by enhancing bradykinin-dependent effects, particularly vasodilation [42]. In fact, these elements of the venom were referred to as bradykinin-potentiating peptides. Subsequently, kininase II, a bradykinin inactivator, and ACE were determined to be the same enzyme [7]. The first attempt at developing an antagonist to the converting enzyme, a substance that could mimic the effects of the venom in a controlled manner, yielded teprotide. This nonapeptide, when tested in humans, proved to effectively lower blood pressure in patients with essential hypertension and also displayed favorable effects in heart failure patients. ACE was also noted to cleave a dipeptide from the teprotide precursor molecule. By considering this latter action, as well as the inhibitory effects of teprotide, investigators in the mid- to late 1970s were able to develop the first orally active ACEIs, of which captopril, a metabolite of teprotide, demonstrated the most activity and is still used today.

Twelve of the many ACEIs that have been developed are currently available for use in the United States. They vary in terms of several features, including chemical structure, potency, whether or not the original compound is the active drug or a prodrug, and their pharmacokinetic properties [7]. Structurally, an ACEI can be considered one of three basic types, depending upon

which of the following it contains: (1) a sulfhydryl (captopril and related compounds; only captopril is available in the United States), (2) a dicarboxyl group (enalapril, lisinopril, benazepril, quinapril, ramipril, perindopril, and others), or (3) a phosphorous (fosinopril and related compounds; only fosinopril is available in the United States). While there are no studies clarifying the issue, the chemical differences in these compounds suggest the possibility that patients with a relatively mild allergy to a single ACEI might be expected to have a favorable response to a different agent of the same class. Given the enormous benefits of these agents in appropriately selected patients, a trial of a second agent in allergic patients is worth consideration.

In spite of their differences, all of the ACEIs are comparable in terms of clinical effectiveness and utility [7, 8], with the exception that fosinopril and spirapril are eliminated equally by the kidney and liver, whereas the remainder of the ACEIs is eliminated principally by the kidney [7]. Therefore, in the setting of renal impairment, levels of ACEIs are increased, and dosage should be adjusted accordingly. Another important consideration in administering these drugs is in patients with heart failure or who are sodium depleted. These patients would be expected to have elevated levels of renin, making them hyperresponsive to the hypotensive effects of ACEIs. Therefore, these patients should also be started on a reduced dose of medication and titrated up as needed. Table 1 provides an overview of the characteristics of the commonly used ACEIs.

ACEIs have found utility well beyond their function as antihypertensives, including demonstrated profound benefits in the setting of heart disease and chronic renal disease. As there is such an intricate relationship between heart disease, cardiovascular disease, and renal disease, ACEIs have proven to be an extraordinary tool for managing all of these diseases on many levels. The studies of ACEIs were initiated based on knowledge of the RAS and the observation that this system is aberrantly upregulated in many of these conditions. With knowledge of a new paradigm of the RAS, including its actions at the systemic, local, and intracellular level, some of the effects of ACEIs can be explored in greater detail.

There have been numerous trials studying the effects of ACEIs in heart disease [43–45]. Only a few of them will be discussed here, but, to summarize, almost every study looking at the use of ACEIs in the setting of chronic heart failure, myocardial infarction, coronary artery disease, and prophylaxis in patients at high risk for cardiovascular disease have shown a survival benefit. In one of the earliest trials to study the effects of ACEIs on mortality in patients with heart failure (CONSENSUS—Cooperative North Scandinavian Enalapril Survival Study), patients with NYHA class IV congestive heart failure who received enalapril showed a significant decrease in mortality [40, 31, and 27% at 6 months, 1 year, and the end of the study (20 months), respectively] compared with patients who received placebo [46]. Furthermore, due to improvement in symptoms, a significant number of patients were reclassified as having a less severe degree of heart failure in accordance with NYHA standards, suggesting a regression of disease. Consistent with that impression, heart size was reduced in the patients receiving enalapril, and less of the other standard heart failure medications were required in the setting of enalapril administration.

Another study (SOLVD—Studies of Left Ventricular Dysfunction) demonstrated a decreased incidence of heart failure in patients with left ventricular dysfunction when treated with enalapril versus placebo [47]. Additionally, patients with left ventricular systolic dysfunction and symptomatic heart failure (NYHA class II and III) receiving enalapril demonstrated a significant difference in cardiac function from those receiving placebo at one year. Specifically, while patients receiving placebo experienced worsening of cardiac function, with increased end-systolic and end-diastolic volumes, those receiving enalapril experienced a decrease in both parameters, as well as in left ventricular end-diastolic pressure. When enalapril was continued, end-systolic and end-diastolic volumes returned to baseline but still did not progress to the degree of the placebo group. Importantly, left ventricular volume increased in the placebo group but not in the enalapril group. Treatment effects were attributed to altered cardiac remodeling, as well as sustained decrease in preload and afterload. As discussed in the previous section, the ability of ACEIs to interfere with angiotensin II generation, as well as the consequent inflammation and fibrosis, at the tissue level support the idea of ACEIs preventing the deleterious remodeling that occurs in the setting of cardiac dysfunction.

ACEIs have been proven to be beneficial not only in patients with ventricular dysfunction and heart failure but also in patients after acute myocardial infarction [47, 48]. SOLVD comprised two parallel studies: Patients with evident symptoms of heart failure who were already being treated with standard heart failure drugs were enrolled in the treatment arm of the study, while patients with minimal or no symptoms and not on any drug regimen for heart failure were enrolled in the prevention arm. Myocardial infarction was a secondary endpoint of the studies (the primary endpoints were mortality and hospital admission for heart failure). In both the treatment and prevention trials, the overall risk of myocardial infarction was reduced similarly (23 and 24%, respectively), with data suggesting that the effect

**TABLE 1 Features of Some Commonly Used ACEIs**

| Drug | Absorption | Bioavailability | Time to Peak Plasma Concentration | Clearance | Half-life | Oral Dose | Comments |
|---|---|---|---|---|---|---|---|
| Captopril | Rapid | 75%; ↓ by 25–35% by food | ≤1 h | Urine; 40–50% intact | 2 h | 6.25–150 mg, 2–3x daily | |
| Enalapril (prodrug) | Rapid | 60%; not ↓ by food | Prodrug: ≤1 h; metabolite: 3–4 h | Urine; intact and metabolite | Prodrug: 1.3 h; metabolite: 11 h | 2.5–40 mg daily | |
| Lisinopril | Slow, variable | 30%; not ↓ by food | 7 h | Urine; intact | 12 h | 5–40 mg daily | Does not accumulate in tissues. |
| Benazepril (prodrug) | Rapid | 37%; slightly ↓ by food | Prodrug: 0.5–1 h; metabolite: 1–2 h | Urine and bile | 10–11 h | 5–80 mg daily | Accumulates in lung tissue |
| Fosinopril (prodrug) | Slow | 36%; rate, not extent, ↓ by food | 3 h | Urine and bile | 11.5 h | 10–80 mg daily; 5 mg daily if $Na^+$ or water depletion, or renal failure | Clearance not significantly altered by renal impairment. |
| Quinapril (prodrug) | Rapid | 60%; rate, not extent, ↓ by food | Prodrug: 1 h, may be delayed after food; metabolite: 2 h | Urine 61%; feces 37% | Initial = 2 h; terminal half-life of 25 h due to high-affinity binding to ACE in tissues | 5–80 mg daily | Conversion from prodrug to metabolite decreased in patients with impaired liver function. |
| Ramipril (prodrug) | Rapid | 50–60%; rate, not extent, ↓ by food | Prodrug: 1 h; metabolite: 3 h | Urine | Initial: 2–4h; intermediate: 9–18 h; terminal: >50 h | 1.25–20 mg daily | Triphasic elimination kinetics; initial—distribution to all tissues; intermediate—clearance of free metabolite from plasma; terminal—dissociation from tissue. |

*Source*: Adapted from information in ref. 7.

regarding nonfatal myocardial infarction was even greater (29%). The difference in effect between enalapril and placebo was not statistically significant at 6 months, but the difference between the two groups (enalapril versus placebo) increased significantly from that point on.

Very positive effects of ACEIs on the kidney have also been reported [49]. Hypertension is one of the two most common causes of chronic kidney disease [50]. Conversely, kidney disease can cause hypertension, so blood pressure control is an important objective in preventing the onset and progression of renal disease. Given the mechanism of action of ACEIs, as well as their initial indication as antihypertensives, they have an obvious role in preventing or ameliorating the effects of chronic kidney disease by lowering blood pressure.

Diabetes mellitus is the other of the two most common causes of chronic kidney disease [50]. Patients with diabetes who develop renal impairment typically have an insidious decrement of renal function over time. Some studies suggest that this progression is slowed with administration of antihypertensives of any kind. Other studies found that use of ACEIs specifically slowed the rate of progression of nephropathy in diabetic patients, even if the patients were normotensive at baseline [51]. Equipped with this knowledge, Lewis et al. carried out a landmark trial studying the beneficial effects of ACEIs (captopril, specifically) on renal function, apart from their antihypertensive properties, in patients with diabetes [52]. The primary endpoint of this randomized controlled trial (RTC) was doubling of baseline serum creatinine in patients with insulin-dependent diabetes mellitus, and the major secondary analyses were the length of time to the combined endpoints of death, dialysis, and transplantation, as well as of changes in renal function. Patients receiving captopril had an overall 48% decrease of doubling of serum creatinine as compared with patients receiving placebo. Subgroup analyses reflected a greater reduction in risk correlating with a higher baseline serum creatinine. Similarly, the rate of decline of renal function was significantly slower in the captopril group than in the placebo group. Finally, patients receiving captopril had a 50% decreased risk of the aforementioned secondary combined endpoints of death, dialysis, and transplantation. These results were independent of blood pressure between groups, thus suggesting that the benefits of ACEIs in diabetic nephropathy extend beyond blood pressure control.

A subsequent study looked at the effects of ACEIs in normotensive, type II diabetic patients with microalbuminuria [51]. The purpose of the study was to examine how ACEIs influenced proteinuria and the rate of decline of kidney function in this patient population. Albuminuria, blood pressure, serum creatinine, fasting blood glucose, and glycosylated hemoglobin levels were measured every 3–4 months. Renal function remained stable throughout the 5-year study period in patients receiving enalapril, compared with a 13% decline in renal function in patients receiving placebo. The increase in albuminuria was significantly less in the enalapril group than in the placebo group at the end of 5 years. Furthermore, patients in the albuminuria group experienced a decline in albuminuria in the first year of the study prior to the subsequent increase. Finally, mean blood pressure continued to be stable in the enalapril group for the duration of the study, while blood pressure in the placebo group increased by an average of 5 mmHg by the end of the study. This study was small, with only 94 patients. However, it extended the idea that ACEIs are effective in the growing population of patients with non-insulin-dependent diabetes mellitus (NIDDM), as well as in those with insulin-dependent diabetes mellitus (IDDH). In addition, it suggested the importance of early initiation of ACEIs, even in the absence of hypertension, in patients with even minimal proteinuria. These concepts have been supported by a number of larger studies.

Eventually, clinical trials were conducted to attempt to determine the effectiveness of ACEIs in slowing the rate of progression of nondiabetic renal disease. The REIN (Ramipril Efficacy in Nephropathy) study was one of the first to consider this question, looking specifically at whether an ACEI was more effective in slowing the rate of decline of the glomerular filtration rate (GFR) (the primary endpoint), decreasing proteinuria, and preventing end-stage renal disease in nondiabetic patients when compared with standard treatment regimens for the same population [53]. Patients in the study were randomly assigned to receive either ramipril or placebo after stratification according to baseline proteinuria (<3 g vs. ≥3 g). Standard antihypertensive medications were used to achieve a target diastolic blood pressure <90 mmHg. Not surprisingly, patients with a baseline proteinuria ≥3 g/24 h had significantly greater monthly rate of GFR decline than did those with baseline proteinuria <3 g/24 h. Additionally, the mean monthly rate of GFR decline in the patients with higher baseline proteinuria receiving ramipril was 0.39 mL/min as compared with 0.89 mL/min in patients receiving placebo. These findings from the second interim analysis were so significant that placebo treatment was terminated at that time and ramipril was made available to the patients who had been receiving placebo. These latter patients were subsequently found to have a significant decrease in the rate of decline of renal failure, and those patients continuing on ramipril experienced continued decrease in the rate of decline of

renal function such that it correlated with what was considered essentially normal. Further analyses demonstrated that patients begun on ramipril at the beginning of the study were significantly less likely to progress to end-stage renal disease (ESRD) than those switched to ramipril at the time of the interim analysis. The study continued for the group of patients with lower baseline proteinuria, and patients from this stratum, as with those presenting with higher baseline proteinuria, were significantly less likely to advance to ESRD than were those receiving placebo. This study, like those in patients with diabetes mellitus, strongly suggested that ACEIs are effective in decreasing the rate of decline of renal function in patients with chronic renal insufficiency to an extent not explained simply by blood pressure reduction.

In addition to the benefits enumerated above, ACEIs have been demonstrated to affect a significant reduction in mortality in patients with scleroderma renal crisis [54, 55]. Scleroderma renal crisis (SRC), defined as new-onset severe hypertension with rapidly progressive renal failure, is an uncommon but severe complication of systemic sclerosis, occurring in approximately 5% of patients. At one time, the only definitive treatment available was bilateral nephrectomy. Absent treatment, SRC was once uniformly fatal, usually within months. Prior to the availability of ACEIs specifically, the one-year mortality associated with SRC was 76%, but the administration of high-dose ACEIs has reduced the mortality to less than 15%. SRC appears to occur most commonly in patients with diffuse system sclerosis, and most patients experience SRC within 4 years (median approximately 2.4 years) of onset of symptoms of systemic sclerosis. Patients who are diagnosed early and treated aggressively with ACEIs consistently have the best outcomes, with many requiring no dialysis within the first year, or requiring only temporary dialysis [56]. Furthermore, patients who were continued on ACEIs after resolution of the renal crisis often showed continued improvement in renal function, whereas no patients who discontinued ACEIs demonstrated spontaneous recovery of renal function. Thus, continuation of ACEIs after blood pressure has been controlled, even in the setting of elevation in serum creatinine, is typically warranted. The pathology of SRC is well characterized, with the hallmark being concentric intimal proliferation in small arteries and arterioles, known colloquially as an "onion-skin" lesion. While the pathogenesis is unclear, most patients with SRC are hyperreninemic and display associated hypertrophy of the juxtaglomerular apparatus, and this hyperactivity of the renin–angiotensin–aldosterone axis appears to play a role. Thus, by interrupting this pathway, ACEIs offer an effective means of treating SRC in many patients.

## 4 ANGIOTENSIN II RECEPTOR BLOCKERS

Development of potential inhibitors of the angiotensin II AT1 receptor began in the early 1970s with elucidation of the functional components of the native angiotensin II octapeptide [7, 57]. Early investigations into the effects of angiotensin II on sympathetic nerves and smooth muscle cells revealed that, while the free C-terminal $CO_2H$ was necessary for both uptake into nerve terminals as well as the pressor response in muscle, an aromatic ring on the phenylalanine at position 8 was required only for the pressor response but did not affect uptake [58]. Therefore, a number of 8-substituted peptide analogs of angiotensin II were created in an attempt to find one or more that could antagonize the pressor effects of angiotensin II in vivo [59]. While a number of these analogs, including saralasin [60–62] and 1-sarcosine, 8-isoleucine angiotensin II [63, 64], and others, proved to be powerful inhibitors of angiotensin II, they were not clinically useful. One drawback was their peptide nature, which limited bioavailability and necessitated intravenous administration. More importantly, they exhibited considerable agonist activity in some patients, making them nonviable for clinical application.

In order to circumvent the issue of bioavailability, the next phase involved development of nonpeptide angiotensin II analogs [65]. The first of these to demonstrate therapeutic promise were S8307 and S8308 and related compounds, all of which were benzylimidazoles [7, 57, 66]. Unlike the previously developed peptides, these compounds were both orally bioavailable and selective for the angiotensin II receptor. However, they displayed only weak receptor antagonism and so were not an option for clinical use. Nevertheless, the structure of these compounds provided a chemical lead for further attempts at development of an effective angiotensin inhibitor.

Researchers first noted that the carboxylic acid moiety at the 2′ position of the biphenyl in the newer compounds was required for binding affinity to the angiotensin II receptor but conferred only limited bioavailability [67]. A number of modified compounds were created with substitution of acidic isosteres with varying degrees of lipophilicity and metabolic stability. The result of these efforts was the development of the first orally bioavailable, selective, and potent angiotensin II receptor blocker (ARB), DuP753, now called losartan. Losartan is a biphenylmethyl derivative, wherein a tetrazole ring is substituted for the original carboxylic acid at the 2′ position. Losartan received approval from the U.S. FDA in 1995, and a number of related compounds quickly followed [7]. The others are also biphenylmethyl derivatives, with the

exception of eprosartan, a thienylmethylacrylic acid derivative.

In contrast to the original peptide analogs, these compounds have no agonist activity, regardless of dose [7]. All of the clinically available ARBs exhibit high-affinity binding of the angiotensin receptor, with selectivity for the AT1 receptor being approximately 10,000 times greater than for the AT2 receptor. Among the biphenylmethyl derivatives, candesartan is reported to have the highest affinity of this class, followed by irbesartan, then telmisartan [7, 68]. Valsartan and the active metabolite of losartan (EXP3174) are essentially equal in affinity to telmisartan. Losartan has the lowest affinity of the group. A separate study demonstrated telmisartan to have the highest binding affinity, followed by olmesartan, then candesartan, and finally EXP3174, valsartan, and losartan [69]. Further details of the pharmacological characteristics of some of the commonly used ARBs are discussed below. All of the ARBs are competitive inhibitors of angiotensin II at the AT1 receptor, but based on binding affinity they can be further categorized as being surmountable or insurmountable [7, 69, 70]. That is, the dose–response curve of angiotensin II, while shifted to the right in the presence of any ARB, is not shifted in parallel when they are compared. Thus, in the setting of surmountable antagonism, as occurs with losartan, the maximal effect of the native compound (angiotensin II, in this case) can overcome the antagonism of the drug. Conversely, telmisartan and candesartan exhibit insurmountable antagonism. After efficient and tight binding to the AT1 receptor, telmisartan dissociates very slowly, and it is this slow dissociation that likely largely accounts for the insurmountable antagonism of telmisartan, as well as its long-lasting therapeutic effects.

Losartan, the first of the ARBs, has a bioavailability of less than 50%, as is generally the case with members of this class [7]. Losartan is partially metabolized to EXP 3174, a 5-carboxylic acid, and other metabolites by CYP2C9 and 3A4. EXP 3174 is the only known active metabolite and is a more potent antagonist of the angiotensin II AT1 receptor than is losartan itself. Peak plasma levels of losartan are reached within approximately 1 h, while those of EXP 3174 are reached in 3 h. The half-lives of losartan and EXP 3174 are 2.4 h and 6–9 h, respectively. Even though losartan and EXP 3174 are both hepatically and renally cleared, clearance is affected only by hepatic, but not renal, insufficiency. The standard dose of losartan is 25–100 mg, given as a single or divided dose.

Candesartan cilexetil is a prodrug [7]. The active metabolite, candesartan, is created by hydrolysis in the gastrointestinal tract, with subsequent absorption. The half-life of candesartan is 9 h, with peak plasma levels achieved within 3–4 h. Candesartan undergoes both renal (33%) and biliary (66%) clearance, but conversely to losartan, clearance is affected by renal, but not mild-to-moderate hepatic, insufficiency. Candesartan is dosed between 4 and 32 mg daily, given as a single or divided dose.

Eprosartan is distinguished pharmacologically by being a thienylmethylacrylic acid derivative [7]. The bioavailability is approximately 13% [71]. Eprosartan reaches peak plasma levels in 1–2 h, with a terminal half-life of 5–9 h. It undergoes partial metabolism to its glucuronide conjugate. Both eprosartan and its metabolite undergo renal and biliary clearance, and both renal and hepatic insufficiency alter the rate of clearance. The total daily dose of eprosartan is 400–800 mg, given in one or two doses. Eprosartan appears to have a unique characteristic in that, in animal models, it inhibits sympathetic outflow-induced pressor responses [72]. Thus, eprosartan acts both directly and indirectly to inhibit the actions of angiotensin II, perhaps affording it an advantage in the antihypertensive arena.

Irbesartan, like eprosartan, is partially metabolized to its glucuronide conjugate by cytochrome P450 2C9 [7]. Oral bioavailability is approximately 60–80% and is unaffected by food [73, 74]. The half-life is 11–15 h, although peak plasma levels are achieved within 1.5–2 h. The steady-state volume of distribution of irbesartan is 53–93 L. Twenty percent of the clearance of irbesartan is renal, while the remaining 80% is biliary, with less than 10% of intact drug is excreted in the urine and less than 30% in the feces. Neither renal nor mild-to-moderate hepatic insufficiency alters clearance. Irbesartan is administered as a single daily dose of 150–300 mg.

Telmisartan reaches peak plasma concentrations in 0.5–1 h, and its half-life of 24 h is the longest of any of the ARBs to date [7]. As touched on previously, telmisartan has an extended dissociation half-life of 75 min, which is nearly five times longer than that of angiotensin II, and which largely accounts for its long duration of action [70]. Unlike the other ARBs discussed, telmisartan is excreted primarily as the intact drug via biliary secretion. Clearance is affected by any degree of hepatic insufficiency. The dose of telmisartan is 40–80 mg, given once a day.

Valsartan has a half-life of 9 h, reaching peak plasma concentrations in 2–4 h [7]. Valsartan should be taken on an empty stomach, as food significantly decreases the absorption of the drug. Seventy percent of the clearance of valsartan is via a hepatic mechanism, and clearance is affected by hepatic insufficiency. The remainder is renally cleared, but no dosage adjustment is required for renal insufficiency. Valsartan is given as a once-daily dose of 80–320 mg.

There are numerous studies investigating the potential benefits of the ARBs in comparison with placebo,

as well as with other antihypertensives. Still others consider ARBs in combination with other agents. While space does not permit a detailed discussion of all of these studies or even of the questions they address, several of them will be presented in order to highlight a few of the relevant findings.

Until the development of ACEIs and ARBs, the most effective intervention for reducing the incidence of cardiovascular morbidity and mortality in patients with hypertension was a regimen of ß-blockers and diuretics. Furthermore, left ventricular hypertrophy (LVH) is a known independent risk factor for cardiovascular morbidity and mortality. The LIFE (Losartan Intervention for Endpoint reduction in hypertension) study sought to determine if ARBs could improve left ventricular hypertrophy and thereby reduce the incidence of cardiovascular morbidity and death compared with atenolol, a ß-blocker [75, 76]. In this randomized, double-blind, parallel-group trial, over 9000 patients with seated systolic blood pressures of 160–220 mmHg and diastolic blood pressures of 95–115 mmHg, as well as LVH, were assigned to receive either losartan or atenolol and were followed for a minimum of 4 years, or until 1040 patients had a primary cardiovascular event, defined as death, myocardial infarction, or stroke. Although blood pressure reduction was significant and similar in the two groups, significantly fewer patients in the losartan group experienced an event defined by the primary endpoint (508 patients in the losartan group versus 588 in the atenolol group, relative risk 0.87). This difference is largely attributable to a significantly lower incidence of fatal and nonfatal stroke—an impressive 25% adjusted risk reduction—in the losartan group. The incidence of myocardial infarction did not differ significantly between the two groups. Notably, LVH-defining voltage indices as defined by both the Cornell and Sokolow-Lyon criteria were significantly reduced in the losartan group as compared with the atenolol group. Another interesting discovery was a 25% reduction in the risk of new-onset diabetes in the patients receiving losartan. This is similar to findings in studies of ACEIs and is a theme that recurs in studies of others of the ARBs. The effect may be attributable to modulation of insulin resistance.

Another arm of the LIFE study considered the role of losartan in the prevention of cardiovascular morbidity and mortality in patients with diabetes [77]. Of the original study population, 1195 patients with diabetes in addition to hypertension and electrocardiographic evidence of LVH were randomly assigned to receive either losartan or atenolol after a placebo run-in period. As in the main study, the primary composite endpoint was cardiovascular morbidity and mortality, comprising cardiovascular death, stroke, or myocardial infarction. Not surprisingly, blood pressure reduction and control were comparable in the two groups. The relative risk of the composite endpoint in the losartan group was only 76% of that in the atenolol group, a significant difference, and all-cause mortality was also significantly lower in the losartan group (relative risk 0.61). In this subgroup study, the reduction in risk of composite endpoints was predominantly attributable to a reduction in cardiovascular mortality, as the difference in incidence of stroke and myocardial infarction between the two groups was nonsignificant. Also, with regard to secondary endpoints, the incidence of hospital admissions for heart failure was significantly lower in the losartan group (adjusted hazard ration 0.59). In both the principal study and in this subgroup, the benefits in the losartan group were evidently greater than those attributable to blood pressure control alone, as both groups had similar reductions in and maintenance of blood pressure throughout the study.

Another study to evaluate the potential benefits of ARBs with respect to cardiovascular disease was the Valsartan Antihypertensive Long-term Use Evaluation (VALUE) study [78, 79]. Investigators hypothesized that, given the known multiple deleterious effects of angiotensin II, an ARB would provide greater overall cardiovascular benefit than a calcium-channel blocker in patients with equivalent blood pressure control and high cardiovascular risk. The primary endpoint of the VALUE study was composite morbidity and mortality secondary to cardiac events. The VALUE study was an international, double-blind, prospective, parallel-group trial that randomized 15,314 patients 50 years of age or older with hypertension (treated or untreated) and a high risk of cardiovascular events to begin treatment with either valsartan or amlodipine. Of the patients randomized, 15,245 were considered in the final analysis. The study followed a carefully proscribed algorithm of medication selection and administration, wherein patients were started on a low-dose of either valsartan or amlodipine, and any necessary increase of study drug or addition of medications was done at monthly intervals in a specified sequence [i.e., begin with low-dose study drug, increase the study drug, add a diuretic (hydrochlorothiazide), increase the diuretic, add any antihypertensive other than ACEI, another ARB, or calcium antagonist]. A larger proportion of patients taking valsartan required the highest dose of study drug plus hydrochlorothiazide (HCTZ) plus other antihypertensives than did patients receiving amlodipine. The vast majority of patients were on antihypertensives at the start of the study and were rolled over into the study without a washout period. Interestingly, the decrease in blood pressure in the amlodipine group was significantly greater than in the valsartan group at the doses used. This effect was more pronounced in the early part of

the study. There were also fewer events in terms of the primary outcome measure in the amlodipine group, although this finding was statistically nonsignificant. This result may be attributable to the difference in blood pressure lowering between the two study drugs. Although the rates of heart failure admissions and both fatal and nonfatal stroke were similar in the two groups, the other secondary outcomes of myocardial infarction (fatal and nonfatal) and total cardiovascular events were greater with valsartan than with amlodipine. Nevertheless, the rates of all-cause death were not significantly different in the two groups. Importantly, the incidence of new-onset diabetes was significantly less in patients receiving valsartan than in those receiving amlodipine, 13.1 versus 16.4%, respectively.

Like ACEIs, ARBs were considered a promising alternative for treatment of patients with renal disease precisely because of their inhibition of the effects of angiotensin II. Once again, the high incidence of renal disease, as well as the significance of hypertension and cardiovascular risk in patients with diabetes, drove the investigation of the utility of ARBs in thwarting some of the many adverse consequences of diabetic and nondiabetic renal disease [80–84]. The Irbesartan Diabetic Nephropathy Trial (IDNT) sought to determine if an ARB (irbesartan) could prevent the progression of nephropathy independently of blood-pressure-lowering effects to a greater extent than either a calcium-channel blocker (amlodipine) or placebo [85–87]. In the IDNT, 1715 patients with type II diabetes mellitus, hypertension, and at least 900 mg proteinuria were randomized to receive irbesartan, amlodipine, or placebo. Other antihypertensive medications were added as necessary to achieve blood pressures of 135/85 or less. The primary endpoint was a composite of doubling of baseline serum creatinine, end-stage renal disease, and death due to any cause. The findings in the study were striking. Even though blood pressures were comparable in the three groups, and there was no difference in the rate of death due to all causes, the risk of reaching the primary combined endpoint in the irbesartan group was 20% lower than in the placebo group and 23% less than in the amlodipine group. In terms of the individual parameters, doubling of baseline serum creatinine was 33% lower with irbesartan than with placebo, and 37% lower with irbesartan than with amlodipine. Proteinuria was reduced by 33% in the irbesartan group, 6% in the amlodipine group, and 10% in the placebo group, with the relative risk of end-stage renal disease being 23% lower with irbesartan than with either placebo or amlodipine. There was, therefore, demonstrated renoprotection in the setting of type II diabetes-induced nephropathy with use of irbesartan, an effect apparently independent of the antihypertensive actions of this ARB.

The results of another study, the Reduction of Endpoints in NIDDM with the Angiotensin II Receptor Antagonist Losartan (RENAAL) study, very closely mirrored those of the IDNT [88]. This multinational, double-blind trial randomized 1513 patients with type II diabetes mellitus and at least 500 mg of proteinuria to receive either losartan or placebo, as well as additional antihypertensives as necessary to achieve a blood pressure of less than 140/90 mmHg. As with the IDNT, the primary endpoint was a composite of doubling of baseline serum creatinine, progression to end-stage renal disease, or death due to any cause. Patients receiving losartan had a 16% decreased risk of the primary composite endpoints compared with patients receiving placebo. With respect to renal parameters, the risk of doubling of serum creatinine was decreased by 25%, and the risk of end-stage renal disease was reduced by 28%, in patients receiving losartan versus those receiving placebo. The rate of all-cause mortality was not different between the two groups. RENAAL, like IDNT, clearly demonstrated beneficial renoprotective effects with use of an ARB and also showed that these effects extend beyond those attributable to blood pressure reduction alone.

Given the findings described above, ARBs appear to have renoprotective effects beyond those of placebo or calcium-channel blockers. Recall, too, the significant benefits of ACEIs in preventing the progression of renal disease in both diabetic and nondiabetic patients, which have been well documented, as described earlier. However, some studies have suggested an angiotensin II escape mechanism that blunts the long-term efficacy of the ACEIs [89]. Thus, the question was raised whether an ARB might offer greater long-term benefits in the treatment of renal disease. The obvious next step was a head-to-head comparison of these two classes of medications. In a small prospective, open-blind study, 68 Japanese patients with renal insufficiency (creatinine clearance 25–50 mL/min) and hypertension were randomized to receive either an ACEI (benazepril or trandolapril) or an ARB (candesartan or losartan) and were then monitored for changes from baseline GFR over a 5-year treatment period [90]. There was no significant difference between the two groups with respect to this primary endpoint. However, the study also investigated several secondary endpoints, including annual changes in GFR, serum creatinine levels, urinary protein excretion, blood pressure, and the rate of development of end-stage renal disease, and there were several outcomes of note. In particular, absolute decline in GFR in the group treated with ARBs was significantly greater than in patients treated with ACEIs, a difference that did not become apparent until year 3 of the study. In addition, progression to end-stage renal disease with

initiation of dialysis was significantly greater in the ARB group (81.2%) than in the ACEI group (52.7%), a difference that also did not develop until year 3 of the study. Throughout the study, there was no statistically significant difference in blood pressures between the two groups.

The results of the above study were not corroborated in a larger trial, Diabetics Exposed to Telmisartan and Enalapril (DETAIL) trial [22, 91]. This is perhaps due to the small patient population of the former study, as well as bias introduced by single-physician participation and the lack of double blinding. The DETAIL trial was a 5-year, double-blind, double-dummy, randomized study of 250 patients with hypertension and early-stage diabetic renal disease assigned to receive either telmisartan or enalapril and subsequently monitored for changes from baseline GFR (primary endpoint). The study secondarily considered progression to end-stage renal disease, as well as all-cause mortality. Contrary to what was found in the prior study, there was no significant difference in annual decline of GFR between the telmisartan group and the enalapril group. Additionally, all patients in both groups maintained a serum creatinine less than or equal to 200 μmol/L, and none of the patients in either group required dialysis throughout the entire 5-year study period. There were six deaths in each study group, and half of these deaths were cardiovascular in nature. These studies raise an interesting question: Would combination therapy with an ACEI and an ARB be more beneficial than monotherapy for treatment of chronic renal disease? Early studies suggested that this might be the case in select patients, particularly those in whom blood pressure is not controlled and in those in whom urinary protein excretion remains high despite monotherapy [92, 93]. The COOPERATE trial [103] was a randomized controlled trial comparing the effects of losartan, trandolapril, or the combination of the two drugs at equivalent doses in patients with nondiabetic chronic kidney disease. The primary endpoint was a composite of time to doubling of serum creatinine or end-stage renal disease. While this study suggested a benefit with combination therapy, the validity of the results was later questioned and the paper was subsequently retracted. Since then, studies have failed to demonstrate a significant synergistic benefit to using combination ACEI/ARB therapy, and in fact, multiple studies have shown an increased risk of adverse effects associated with combination therapy compared to monotherapy [104, 105]. The ONTARGET study was a large, randomized controlled trial comparing the effects of telmisartan, ramipril, and the combination of both drugs in 25,620 patients with vascular disease or high-risk diabetes [104]. Treatment with the combination of telmisartan and ramipril did not significantly reduce the incidence of cardiovascular disease or death compared to monotherapy and was associated with a significantly increased risk of hypotensive symptoms, syncope, and renal dysfunction. A subsequent report from the same investigators studied the effects of combination therapy versus monotherapy on renal outcomes in a subset of the patients enrolled in the original trial who who had impaired renal function and/or proteinuria [105]. Combination therapy resulted in a greater reduction in proteinuria when compared to monotherapy but was associated with an increased risk of the primary composite outcome of dialysis, doubling of serum creatinine, and death. Given the results of these studies, combination therapy with ACEI and ARB cannot be recommended at this time.

## 5 RENIN INHIBITORS

Thus far in this review, drugs that block the production of angiotensin II and that interfere with the actions of angiotensin II have been discussed. These drugs reflect a logical approach in disease modulation, based on an understanding of the RAS that was described previously. However, some interesting override mechanisms occur with the use of these medications that may explain to a great extent why they do not demonstrate an even greater effect than they do [89]. Recall from the previous discussion that the new understanding is of a local as well as systemic RAS. Tissues express various components of the system, including ACE and AT1 and AT2 receptors, that permit local production and activation of angiotensin II. Notably, however, renin and its precursor molecule, prorenin, are not present in these local systems. Rather, systemic renin produced in the juxtaglomerular apparatus is taken up by tissues and utilized to initiate the ultimate production of angiotensin II. Angiotensin II itself exerts negative feedback on the juxtaglomerular cells via the AT1 receptor, thereby bringing about a reduction in renin and consequently of angiotensin II. In the setting of ACEI or ARB, there is a reactive increase in renin, and to a somewhat lesser extent of prorenin, due to the lack of angiotensin-II-mediated negative feedback. With the rise in plasma renin activity (PRA), there is a resultant increase in the production of angiotensin I. In patients receiving ARBs, there is a corresponding increased production of angiotensin II. In addition, even though ACEIs are expected to prevent the production of angiotensin II, studies have demonstrated that renin levels can rise far enough above baseline during ACEI administration that over time normal or even supranormal levels of angiotensin II are restored, a mechanism known as *angiotensin escape*. Ultimately, then, the beneficial effects of both of these classes of drugs might be blunted or possibly even reversed as a result of misguided compensatory mechanisms.

A search has been underway to find a drug that could overcome these shortcomings of ACEIs and ARBs. The logical point of intervention is at the level of renin production, the rate-limiting compound of the RAS, which would suppress angiotensin I generation and angiotensin II production by ACE, chymase, or other enzymes, and would largely obviate the need for AT1 receptor antagonism. Indeed, the latest weapon in the arsenal of drugs that counter the RAS is a renin inhibitor, aliskiren.

Aliskiren is a low-molecular-weight, nonpeptide compound that exhibits limited (2.6%) oral bioavailability [94]. It is highly selective for the proteolytic binding site of renin, and binding of aliskiren to renin results in a decrease of plasma levels of angiotensin I and angiotensin II. There is a reactive rise in renin levels, but PRA is nevertheless effectively reduced with once-daily dosing. Peak plasma concentrations of aliskiren are reached 2–4 h after oral administration, and a single dose of 160 mg results in plasma concentrations that are 50-fold that of the 50% inhibitory concentration of 0.6 nmol/L. Although peak concentration is significantly reduced when aliskiren is administered with food, the effect on PRA is nevertheless unaffected, so timing of dosing relative to food intake is unnecessary. Clearance of aliskiren is predominantly via the gastrointestinal tract (91%, unchanged), with a mere 0.6% being excreted in the urine, and 1.4% undergoing oxidation. Metabolism does not appear to significantly involve the cytochrome P450 system. Aliskiren has a terminal half-life of 24–40 h due to accumulation in plasma, therefore permitting a dosing interval of 24 h. Studies to date suggest that 150–300 mg administered orally once a day is effective in lowering blood pressure.

In one study, 672 patients with elevated diastolic blood pressures were randomized to receive either aliskiren (150, 300, or 600 mg) or placebo for 8 weeks, followed by a 2-week withdrawal period [95]. Aliskiren demonstrated superiority in lowering both systolic and diastolic blood pressure at all doses in comparison with placebo. The antihypertensive effect was maintained throughout the 2-week withdrawal period. High trough-to-peak ratios permitted smooth and sustained blood pressure lowering. Aliskiren was well tolerated, having adverse event rates comparable to placebo, and there was little discontinuance due to adverse events in the aliskiren group.

A very similar short, multicenter, randomized, double-blind, placebo-controlled, active-comparator trial investigated the efficacy of aliskiren versus irbesartan and placebo for blood pressure reduction in patients with mild-to-moderate hypertension [96]. In the study, 652 patients were randomized to receive aliskiren, irbesartan, or placebo after a 2-week run-in period. Aliskiren at a 150-mg dose was comparable to irbesartan 150 mg with respect to blood pressure lowering, and at doses of 300 and 600 mg was significantly better. Aliskiren at all doses was significantly better than placebo in effecting a reduction in blood pressure. Aliskiren was generally well tolerated.

In a very small double-blind, parallel-group, three-way crossover trial, 18 healthy volunteers on a sodium-restricted (100 mmol/day) diet were randomized to receive two doses of aliskiren (40 mg and then 80 mg, or 160 mg and then 640 mg) plus either placebo or enalapril (20 mg) [97]. The trial was divided into three treatment periods of 8 days each. The treatment periods were separated by 6-day washout periods. Blood pressure was unaffected in this study population. Angiotensin I and angiotensin II were significantly reduced at all levels of aliskiren when compared with placebo. On day 1 and day 8, aliskiren significantly reduced angiotensin II levels within 1 h of oral administration, and this effect was maintained for up to 6 h at all doses except 40 mg. At the highest dose (640 mg) of aliskiren, angiotensin levels were decreased by as much as 89% on day 1. Aliskiren at 160 mg reduced baseline angiotensin II to a degree (56%) comparable to 20 mg of enalapril (57%), whereas 640 mg reduced angiotensin II levels by 76%, a significant difference. At the three lower doses, reduction in PRA and angiotensin I correlated with the reduction in angiotensin II. However, enalapril evoked a 15-fold increase in PRA and angiotensin on days 1 and 8. Note that PRA was increased by as much as 34-fold with the maximum dose of aliskiren. Aldosterone levels in both plasma and urine were decreased with aliskiren (≥80 mg) and enalapril administration. The duration of effect was 6 h with enalapril and as long as 24 h with 640 mg of aliskiren. Sodium excretion was significantly enhanced with aliskiren and enalapril versus placebo, whereas potassium was unaffected. Note that aliskiren was detected in the urine of some subjects throughout the 6-day washout period, suggesting that results were skewed by the prolonged duration of action of aliskiren. As in other studies, aliskiren was well tolerated.

In addition to these promising studies regarding the use of aliskiren as monotherapy in the treatment of hypertension, other studies show beneficial effects of aliskiren used in combination with other antihypertensives [98, 99]. In a study of aliskiren and valsartan, the two drugs used concomitantly afforded significantly greater blood pressure reduction than either drug used alone, but without any increase in adverse events [100]. Likewise, aliskiren plus HCTZ was significantly more effective in reducing blood pressure than either drug used as monotherapy [101]. Furthermore, more patients responded to the dual therapy than to single-drug therapy, and rates of blood pressure control were better in patients receiving combination therapy. An added benefit was the suppression of the rise in PRA that

occurs with HCTZ. Thus aliskiren appears to offer a new alternative in the treatment of hypertension as either monotherapy or in combination with other antihypertensives, with a good safety and tolerability profile [102].

## REFERENCES


1. Seifter J, Ratner A, Sloane D (2005). *Concepts in Medical Physiology*. Baltimore: Lippincott Williams & Wilkins.
2. Ali F, Guglin M, Vaitkevicius P, Ghali JK (2007). Therapeutic potential of vasopressin receptor antagonists. *Drugs 67*, 847–868.
3. Hays RM (2006). Vasopressin antagonists—progress and promise. *N Engl J Med 355*, 2146–2148.
4. Verbalis JG (2007). Vaptans for the treatment of hyponatremia: How, who, when, and why. *Nephrol Self-Assess Prog 6*, 199–209.
5. Knoers NV (2005). Hyperactive vasopressin receptors and disturbed water homeostasis. *N Engl J Med 352*, 1847–1850.
6. Nielsen S, Frør J, Knepper MA (1998). Renal aquaporins: Key roles in water balance and water balance disorders. *Curr Opin Nephrol Hypertens 7*, 509–516.
7. Hardman JG, Limbird LE, eds. (2001). *Goodman & Gilman's The Pharmacological Basis of Therapeutics*, 10th ed. New York: McGraw-Hill.
8. Katzung BG (2004). *Basic and Clinical Pharmacology*. New York: Lange Medical Books/McGraw-Hill.
9. Kasper DL, Braunwald E, Fauci AS, Hauser SL, Longo DL, Jameson JL, eds. (2005). *Harrison's Principles of Internal Medicine*, 16th ed. New York: McGraw-Hill.
10. Costello-Boerrigter LC, Smith WB, Boerrigter G, Ouyang J, Zimmer CA, Orlandi C, Burnett JC Jr (2006). Vasopressin-2-receptor antagonism augments water excretion without changes in renal hemodynamics or sodium and potassium excretion in human heart failure. *Am J Renal Physiol 290*, F273–F278.
11. Greenberg A, Verbalis JG (2006). Vasopressin receptor antagonists. *Kidney Int 69*, 2124–2130.
12. Kinter LB, Huffman WF, Stassen FL (1988). Antagonists of the antidiuretic activity of vasopressin. *Am J Physiol 254*, F165–F177.
13. Cordero-Vargas A, Quiclet-Sire B, Zard SZ (2006). A flexible approach for the preparation of substituted benzazepines: Application to the synthesis of tolvaptan. *Bioorg Med Chem 14*, 6165–6173.
14. Macion-Dazard Callahan N, Xu Z, Wu N, Thibonnier M, Shoham M (2006). Mapping the binding site of six nonpeptide antagonists to the human V2-renal vasopressin receptor. *J Pharmacol Exp Ther 316*, 564–671.
15. Kondo K, Ogawa H, Yamashita H, Miyamoto H, Tanaka M, Nakaya K, Kitano K, Yamamura Y, Nakamura S, Onogawa T, Mori T, Tominaga M (1999). 7-Chloro-5-hydroxy-1-[2-methyl-4-(2-methylenzoylamino)benzoyl]-2,3,4,5-tetrahydro-1*H*-1-benzazepine (OPC-41061): A potent, orally active nonpeptide arginine vasopressin V2 receptor antagonist. *Bioorg Med Chem 7*, 1743–1754.
16. Burnier M, Fricker AF, Hayoz D, Nussberger J, Brunner HR (1999). Pharmacokinetic and pharmacodynamic effects of YM087, a combined $V_1/V_2$ vasopressin receptor antagonist in normal subjects. *Eur J Clin Pharmacol 55*, 633–637.
17. Zeltser D, Rosansky S, van Rensburg H, Verbalis JG, Smith N. for the Conivaptan Study Group. (2007). Assessment of the efficacy and safety of intravenous conivaptan in euvolemic and hypervolemic hyponatremia. *Am J Nephrol 27*, 447–457.
18. Ghali JK, Koren MJ, Taylor JR, Brooks-Asplund E, Fan K, Long WA, Smith N. for the Conivaptan Study Group. (2006). Efficacy and safety of oral conivaptan: A V1A/V2 vasopressin receptor antagonist, assessed in a randomized, placebo-controlled trial in patients with euvolemic or hypervolemic hyponatremia. *J Clin Endocrinol Metab 91*, 2145–2152.
19. Gheorghiade M, Teerlink JR, Mebazaa A (2005). Pharmacology of new agents for acute heart failure syndromes. *Am J Cardiol 96*, 68G-73G.
20. Gheorghiade M, Niazi I, Ouyang J, Czerwiec F, Kambayashi J, Zampino M, Orlandi C (2003). Vasopressin $V_2$-receptor blockade with tolvaptan in patients with chronic heart failure: Results from a double-blind, randomized trial. *Circulation 107*, 2690–2696.
21. Gheorghiade M, Gattis WA, O'Connor CM, Adams KF Jr, Elkayam U, Barbagelata A, Ghali JK, Benza RL, McGrew FA, Klapholz M, Ouyang J, Orlandi C. for the Acute and Chronic Therapeutic Impact of a Vasopressin Antagonist in Congestive Heart Failure (ACTIV in CHF) investigators. (2004). Effects of tolvaptan, a vasopressin antagonist, in patients hospitalized with worsening heart failure. *JAMA 291*, 1963–1971.
22. Barnett A (2006). Preventing renal complications in type 2 diabetes: Results of the diabetics exposed to telmisartan and enalapril trial. *J A S N 17*, S132–S135.
23. Schrier RW, Gross P, Gheorghiade M, Berl T, Verbalis JG, Czerwiec FS, Orlandi C (2006). Tolvaptan, a selective oral vasopressin V2-receptor antagonist, for hyponatremia. *N Engl J Med 355*, 2099–2112.
24. Campbell RC (2004). The renin-angiotensin system: A 21st century perspective. *J Am Soc Nephrol 15*, 1963–1964.
25. Kobori H, Nangaku M, Navar LG, Nishiyama A (2007). The intrarenal renin-angiotensin system: From physiology to the pathobiology of hypertension and kidney disease. *Pharmacol Rev 59*, 251–287.
26. Kumar R, Singh VP, Baker KM (2007). The intracellular renin-angiotensin system: A new paradigm. *Trends Endocrinol Metabo 18*, 208–214.
27. Paul M, Mehr AP, Kreutz R (2006). Physiology of local renin-angiotensin systems. *Physiol Rev 86*, 747–803.
28. Kramkowski K, Mogielnicki A, Buczko W (2006). The physiological significance of the alternative pathways of

angiotensin II production. *J Physiol Pharmacol 57*, 529–539.

29. Miyazaki M, Takai S (2006). Tissue angiotensin II generating system by angiotensin-converting enzyme and chymase. *J Pharmacol Sci 100*, 391–397.
30. Raizada MK, Ferreira AJ (2007). ACE2: A new target for cardiovascular disease therapeutics. *J Cardiovasc Pharmacol 50*, 112–119.
31. Remuzzi G, Perico N, Macia M, Ruggenenti P (2005). The role of renin-angiotensin-aldosterone system in the progression of chronic kidney disease. *Kidney Int 68*, S57–S65.
32. Shalamanova L, McArdle F, Amara AB, Jackson MJ, Rustom R (2007). Albumin overload induces adaptive responses in human proximal tubular cells through oxidative stress but not via angiotensin II type 1 receptor. *Am J Physiol Renal Physiol 292*, F1846–F1857.
33. Cao Z, Kelly DJ, Cox A, Casley D, Forbes JM, Martinello P, Dean R, Gilbert RE, Cooper ME (2000). Angiotensin type 2 receptor is expressed in the adult rat kidney and promotes cellular proliferation and apoptosis. *Kidney Int 58*, 2437–2451.
34. Huang XR, Chen WY, Truong LD, Lan HY (2003). Chymase is upregulated in diabetic nephropathy: Implications for an alternative pathway of angiotensin II-mediated diabetic renal and vascular disease. *J Am Soc Nephrol 14*, 1738–1747.
35. Rüster C, Wolf G (2006). Renin-angiotensin-aldosterone system and progression of renal disease. *J Am Soc Nephrol 17*, 2985–2991.
36. Donoghue M, Hsieh F, Baronas E, Godbout K, Gosselin M, Stagliano N, Donovan M, Woolf B, Robison K, Jeyaseelan R, Reitbart RE, Acton S (2000). A novel angiotensin-converting enzyme-related carboxypeptidase (ACE2) converts angiotensin I to angiotensin 1–9. *Circ Res 87*, e1–e9.
37. Tikellis C, Cooper MF, Bialkowski K, Johnson CI, Burns WC, Lew RA, Smith A, Thomas MC (2006). *Developmental expression of ACE2 in the SHR kidney: A role in hypertension*? *Kidney Int 70*, 34–41.
38. Schiavone MT, Santos RAS, Brosnihan KB, Khosla MC, Ferrario CM (1988). Release of vasopressin from the rat hypothalamo-neurohypophysial system by angiotensin-(1-7) heptapeptide. *Proc Nat Acad Sci 85*, 4095–4098.
39. Ferreira AJ, Santo RAS, Almeida AP (2002). Angiotensin-(1-7) improves the post-ischemic function in isolated perfused rat hearts. *Brazil J Med Biol Res 35*, 1083–1090.
40. Benter IF, Yousif MHM, Dhaunsi GS, Kaur J, Chappell MC, Diz DI (2008). Angiotensin-(1-7) prevents activation of NADPH oxidase and renal vascular dysfunction in diabetic hypertensive rats. *Am J Nephrol 28*, 25–33.
41. Patlak M (2004). From viper's venom to drug design: Treating hypertension. *FASEB J 18*, 421e.
42. Hayashi, MAF, Camargo ACM (2005). The bradykinin-potentiating peptides from venom gland and brain of *Bothrops jararaca* contain highly site specific inhibitors of the somatic angiotensin-converting enzyme. *Toxicon 45*, 1163–1170.
43. Ahimastos AA, Aggarwal A, D'Orsa K, Formosa M, White AJ, Savarirayan R, Dart AM, Kingwell BA (2007). Effect of perindopril on large artery stiffness and aortic root diameter in patients with Marfan syndrome. *JAMA 298*, 1539–1547.
44. Goldbarg SH, Elmariah S, Miller MA, Fuster V (2007). Insights into degenerative aortic valve disease. *J Am Coll Cardiol 50*, 1205–1213.
45. Konstam MA, Rousseau MF, Kronenberg MW, Udelson JE, Melin J, Stewart D, Dolan N, Edens TR, Ahn S, Kinan D, *et al.* (1992). Effects of the angiotensin converting enzyme inhibitor enalapril on the long-term progression of left ventricular dysfunction in patients with heart failure. *Circulation 86*, 431–438.
46. Swedberg K, Kjekshus J (1987). Effects of enalapril on mortality in severe congestive heart failure. Results of the Cooperative North Scandinavian Enalapril Survival Study (CONSENSUS). The CONSENSUS Trial Study Group. *N Engl J Med 316*, 1429–1435.
47. SOLVD Investigators. (1992). Effect of enalapril on mortality and the development of heart failure in asymptomatic patients with reduced left ventricular ejection fractions. *N Engl J Med 327*, 685–691.
48. Yusef S, Pepine CJ, Garces C, Pouleur H, Salem D, Kostis J, Benedict C, Rousseau M, Bourassa M, Pitt B (1992). Effect of enalapril on myocardial infarction and unstable angina in patients with low ejection fractions. *Lancet 340*, 1173–1178.
49. Abdi R, Brenner BM (2004). Impact of renin angiotensin system blockade on renal function in health and disease: An end or a beginning? *Sem Nephrol 24*, 141–146.
50. Brenner BM, Zagrobelny J (2003). Clinical renoprotection trials involving angiotensin II-receptor antagonists and angiotensin-converting-enzyme inhibitors. *Kidney Int 63*, S77–S85.
51. Ravid M, Savin H, Jutrin I, Mental T, Katz B, Lishner M (1993). Long-term stabilizing effect of angiotensin-converting enzyme inhibition on plasma creatinine and on proteinuria in normotensive type II diabetic patients. *Ann Intern Med 118*, 577–581.
52. Lewis EJ, Hunsicker LG, Bain RP, Rohde RD. for The Collaborative Study Group. (1993). The effect of angiotensin-converting-enzyme inhibition on diabetic nephropathy. *N Engl J Med 329*, 1456–1462.
53. GISEN Group. (1997). Randomised placebo-controlled trial of effect of ramipril on decline in glomerular filtration rate and risk of terminal renal failure in proteinuric, non-diabetic nephropathy. *Lancet 349*, 1857–1863.
54. Steen VD, Medsger TA Jr (2000). Long-term outcomes of scleroderma renal crisis. *Ann Intern Med 133*, 600–603.

55. Steen V (2007). Scleroderma renal crisis. *Indian J Med Sci 61*, 71–72.

56. Teixeira L, Mahr A, Berezné A, Noël LH, Guillevin L, Mouthon L (2007). Scleroderma renal crisis, still a life-threatening complication. *Ann N Y Acad Sci 1108*, 249–258.

57. Aulakh GK, Sodhi RK, Singh M (2007). An update on non-peptide angiotensin receptor antagonists and related RAAS modulators. *Life Sci 81*, 615–639.

58. Khairallah PA, Toth A, Bumpus FM (1970). Analogs of angiotensin II. Mechanism of receptor interaction. *J Med Chem 13*, 181–184.

59. Marshall GR, Vine W, Needleman P (1970). A specific competitive inhibitor of angiotensin II. *Proc Nat Acad Sci 67*, 1624–1630.

60. Brown JJ, Brown WC, Fraser R, Lever AF, Morton JJ, Robertson JI, Rosei EA, Trust PM (1976). The effects of saralasin, an angiotensin II antagonist, on blood pressure and the renin-angiotensin-aldosterone system in normal and hypertensive subjects. *Aust N Z J Med 6*, 48–52.

61. Marks LS, Maxwell MH, Kaufman JJ (1975). Saralasin bolus test. Rapid screening procedure for renin-mediated hypertension. *Lancet 2*, 784–787.

62. Pettinger WA, Keeton K, Tanaka K (1975). Radioimmunoassay and pharmacokinetics of saralasin in the rat and hypertensive patients. *Clin Pharmacol Ther 17*, 146–158.

63. Fagard R, Amery A, Lijnen P, Reybrouck T (1977). Effects of the angiotensin II antagonist 1-sar-8-ala-angiotensin II in hypertension in man. *Eur J Clin Inves 7*, 473–479.

64. Hollenberg NK, Williams GH, Burger B, Ishikawa I, Adams DF (1976). Blockade and stimulation of renal, adrenal, and vascular angiotensin II receptors with 1-sar, 8-ala angiotensin II in normal man. *J Clin Invest 57*, 39–46.

65. Siegl PK (1993). Discovery of losartan, the first specific non-peptide angiotensin II receptor antagonist. *J Hypertens Suppl 11*, S19–S22.

66. Wexler RR, Carini DJ, Duncia JV, Johnson AL, Wells GJ, Chiut AT, Wong PC, Timmermans PB (1992). Rationale for the chemical development of angiotensin II receptor antagonists. *Am J Hypertens 5*, 209S-220S.

67. Carini DJ, Duncia JV, Aldrich PE, Chiu AT, Johnson AL, Pierce ME, Price WA, Santella III JB, Wells GJ, Wexler RR, Wong PC, Yoo S, Timmermans PBMWM (1991). Nonpeptide angiotensin II receptor antagonists: The discovery of a series of N-(biphenylylmethyl)-imidazoles as potent, orally active antihypertensives. *J Med Chem 34*, 2525–2547.

68. Unger T (2001). Pharmacology of AT1-receptor blockers. *Blood Press Suppl 3*, 5–10.

69. Unger T (1999). Significance of angiotensin type 1 receptor blockade: Why are angiotensin II receptor blockers different? *Am J Cardiol 84*, 9S-15S.

70. Maillard MP, Perregaux C, Centeno C, Stangier J, Wienen W, Brunner HR, Burnier M (2002). In vitro and in vivo characterization of the activity of telmisartan: An insurmountable angiotensin II receptor antagonis. *J Pharmacol Exp Ther 302*, 1089–1095.

71. Shusterman NH (1999). Safety and efficacy of eprosartan, a new angiotensin II receptor blocker. *Am Heart J 138*, S238–S245.

72. Brookes DP, Ohlstein EH, Ruffolo RR Jr (1999). Pharmacology of eprosartan, an angiotensin II receptor antagonist: Exploring hypotheses from clinical data. *Am Heart J 138*, S246–S251.

73. Croom KF, Curran MP, Goa KL, Perry CM (2004). Irbesartan: A review of its use in hypertension and in the management of diabetic nephropathy. *Drugs 64*, 999–1028.

74. Johnston CI (1999). Pharmacology of irbesartan. *Expert Opin Invest Drugs 8*, 655–670.

75. Dahlöf B, Devereux RB, Kjeldsen SE, Julius S, Beevers G, de Faire U, Fyhrquist F, Ibsen H, Kristiansson K, Lederballe-Pedersen O, Lindholm LH, Nieminen MS, Omvik P, Oparil S, Wedel H. for the LIFE study group (2002). Cardiovascular morbidity and mortality in the losartan intervention for endpoint reduction in hypertension study (LIFE): A randomized trial against atenolol. *Lancet 359*, 995–1003.

76. Gerdts E, Oikarinen L, Palmieri V, Otterstad JE, Wachtell K, Boman K, Dahlöf B, Devereux RB (2002). Correlates of left atrial size in hypertensive patients with left ventricular hypertrophy: The losartan intervention for endpoint reduction in hypertension (LIFE) study. *Hypertension 39*, 739–743.

77. Lindholm LH, Ibsen H, Dahlöf B, Devereux RB, Beevers G, de Faire U, Fyhrquist F, Julius S, Kjeldsen SE, Kristiansson K, Lederballe-Pedersen O, Nieminen MS, Omvik P, Oparil S, Wedel H, Aurup P, Edelman J, Snapinn S. for the LIFE study group (2002). Cardiovascular morbidity and mortality in patients with diabetes in the losartan intervention for endpoint reduction in hypertension study (LIFE): A randomized trial against atenolol. *Lancet 359*, 1004–1010.

78. Julius S, Kjeldsen SE, Weber M, Brunner HR, Ekman S, Hansson L, Hua T, Laragh J, McInnes GT, Mitchell L, Plat F, Schork A, Smith B, Zanchetti A. for the VALUE trial group (2004). Outcomes in hypertensive patients at high cardiovascular risk treated with regimens based on valsartan or amlodipine: The VALUE randomised trial. *Lancet 363*, 2022–2031.

79. Mann J, Julius S (1998). The valsartan antihypertensive long-term use evaluation (VALUE) trial of cardiovascular events in hypertension. Rationale and design. *Blood Press 7*, 176–183.

80. Estacio RO, Coll JR, Tran ZV, Schrier RW (2006). Effect of intensive blood pressure control with valsartan on urinary albumin excretion in normotensive patients with type 2 diabetes. *Am J Hypertens 19*, 1241–1248.

81. Haller H, Viberti GC, Mimran A, Remuzzi G, Rabelink AJ, Ritz E, Rump LC, Ruilope LM, Katayama S, Ito S, Izzo JL Jr, Januszewicz A (2006). Preventing microalbuminuria in patients with diabetes: Rationale and design of the randomised olmesartan and diabetes microalbuminuria prevention (ROADMAP) study. *J Hypertens 24*, 403–408.
82. Imai E, Ito S, Haneda M, Chan JCN, Makino H. for the ORIENT investigators (2006). Olmesartan reducing incindence of endstage renal disease in diabetic nephropathy trial (ORIENT): Rationale and study design. *Hypertens Res 29*, 703–709.
83. Ishimitsu T, Kameda T, Akashiba A, Takahashi T, Ando N, Ohta S, Yoshii M, Inada H, Tsukada K, Minami J, Ono H, Matsuoka H (2005). Effects of valsartan on the progression of chronic renal insufficiency in patients with nondiabetic renal diseases. *Hypertens Res 28*, 865–870.
84. Keane WF, Lyle PA (2003). Recent advances in management of type 2 diabetes and nephropathy: Lessons from the RENAAL study. *Am J Kidney Dis 41*, S22–S25.
85. Lewis EJ, Lewis JB (2003). Treatment of diabetic nephropathy with angiotensin II receptor antagonist. *Clini Exp Nephrol 7*, 1–8.
86. Lewis EJ, Hunsicker LG, Clarke WR, Berl T, Phol MA, Lewis JB, Ritz E, Atkins RC, Rohde R, Raz I. for the Collaborative Study Group. (2001). Renoprotective effect of the angiotensin-receptor antagonist irbesartan in patients with nephropathy due to type 2 diabetes. *N Engl J Med 345*, 851–860.
87. Rodby RA, Rohde RD, Clarke WR, Hunsicker LG, Anzalone DA, Atkins RC, Ritz E, Lewis EJ. for the Collaborative Study Group. (2000). The irbesartan type II diabetic nephropathy trial: Study design and baseline patient characteristics. *Nephrol Dial Transplant 15*, 487–497.
88. Brenner BM, Cooper ME, de Zeeuw D, Keane WF, Mitch WE, Parving H, Remuzzi G, Snapinn SM, Zhang Z, Shahinfar S. for the RENAAL study investigators. (2001). Effects of losartan on renal and cardiovascular outcomes in patients with type 2 diabetes and nephropathy. *N Engl J Med 345*, 861–869.
89. Danser AHJ (2007). Novel drugs targeting hypertension: Renin inhibitors. *J Cardiovasc Pharmacol 50*, 105–111.
90. Shoda J, Kanno Y, Suzuki H (2006). A five-year comparison of the renal protective effects of angiotensin-converting enzyme inhibitors and angiotensin receptor blockers in patients with non-diabetic nephropathy. *Intern Med 45*, 193–198.
91. Barnett AH, Bain SC, Bouter P, Karlerg B, Madsbad S, Jervell J, Mustonen J. for the Diabetics Exposed to Telmisartan and Enalapril Study Group. (2004). Angiotensin-receptor blockade versus converting-enzyme inhibition in type 2 diabetes and nephropathy. *N Engl J Med 351*, 1952–1961.
92. Kanno Y, Takenaka T, Nakamura T, Suzuki H (2006). Add-on angiotensin receptor blocker in patients who have proteinuric chronic kidney diseases and are treated with angiotensin-converting enzyme inhibitors. *Clin J Am Soc Nephrol 1*, 730–737.
93. Katayama S, Yagi S., Yamamoto H, Yamaguchi M, Izumida T, Noguchi Y, Inaba M, Inukai K (2007). Is renoprotection by angiotensin receptor blocker dependent on blood pressure?: The Saitama Medical School, albuminuria reduction in diabetics with valsartan (STAR) study. *Hypertens Res 30*, 529–533.
94. Van Tassell BW, Munger MA (2007). Aliskiren for renin inhibition: A new class of antihypertensives. *Ann Pharmacother 41*, 456–464.
95. Oh B, Mitchell J, Herron JR, Chung J, Khan M, Keefe DL (2007). Aliskiren, an oral renin inhibitor, provides dose-dependent efficacy and sustained 24-hour blood pressure control in patients with hypertension. *J Am Coll Cardiol 49*, 1157–1163.
96. Gradman AH, Schmieder RE, Lins RL, Nussberger J, Chiang Y, Bedigian MP (2005). Aliskiren, a novel orally effective renin inhibitor, provides dose-dependent antihypertensive efficacy and placebo-lie tolerability in hypertensive patients. *Circulation 111*, 1012–1018.
97. Nussberger J, Wuerzner G, Jensen C, Brunner HR (2002). Angiotensin II suppression in humans by the orally active renin inhibitor aliskiren (SPP100). Comparison with enalapril. *Hypertension 39*, e1–e8.
98. O'Brien E, Barton J, Nussberger J, Mulcahy D, Jensen C, Dicker P, Stanton A (2007). Aliskiren reduces blood pressure and suppresses plasma renin activity in combination with a thiazide diuretic, an angiotensin-converting enzyme inhibitor, or an angiotensin receptor blocker. *Hypertension 49*, 276–284.
99. Pool JL, Schmieder RE, Azizi M, Aldigier J, Januszewicz A, Zidek W, Chiang Y, Satlin A (2007). Aliskiren, an orally effective renin inhibitor, provides antihypertensive efficacy alone and in combination with valsartan. *Am J Hypertens 20*, 11–20.
100. Oparil S, Yarows SA, Patel S, Fang H, Zhang J, Satlin A (2007). Efficacy and safety of combined use of aliskiren and valsartan in patients with hypertension: A randomised, double-blind trial. *Lancet 370*, 221–229.
101. Villamil A, Chrysant SG, Calhoun D, Schober B, Hsu H, Matrisciano-Dimichino L, Zhang J (2007). Renin inhibition with aliskiren provides additive antihypertensive efficacy when used in combination with hydrochlorothiazide. *J Hypertens 25*, 217–226.
102. Stanton A, Jensen C, Nussberger J, O'Brien E (2003). Blood pressure lowering in essential hypertension with an oral renin inhibitor, aliskiren. *Hypertension 42*, 1137–1143.
103. Nakao N, Yoshimura A, Morita H, Takada M, Kayano T, Ideura T (2003). Combination treatment of angiotensin-II receptor blocker and angiotensin-converting-enzyme inhibitor in non-diabetic renal disease (COOPERATE): A randomised controlled trial. *Lancet 361*, 117–124.
104. Yusuf S, Teo KK, Pogue J, Dyal L, Copland I, Schumacher H, Dagenais G, Sleight P, Anderson C, ONTARGET Investigators (2008). Telmisartan, ramipril, or both in

patients at high risk for vascular events. *N Engl J Med 258*, 1547–1559.

105. Mann JF, Schmieder RE, McQueen M, Dyal L, Schumacher H, Pogue J, Wang X, Maggioni A, Budaj A, Chaithiraphan S, Dickstein K, Keltai M, Metsärinne K, Oto A, Parkhomenko A, Piegas LS, Svendsen TL, Teo TK, Yusuf S, ONTARGET Investigators (2008). Renal outcomes with telmisartan, ramipril, or both, in people at high vascular risk (the ONTARGET study): a multicentre, randomised, double-blind, controlled trial. *Lancet 372*, 547–553.

## BIBLIOGRAPHY

Gheorghiade M, Orlandi C, Burnett JC, Demets D, Grinfeld L, Maggioni A, Swedberg K, Udelson JE, Zannad F, Zimmer C, Konstam MA (2005). Rationale and design of the multicenter, randomized, double-blind, placebo-controlled study to evaluate the efficacy of vasopressin antagonism in heart failure: Outcome study with Tolvaptan (EVEREST). *J Card Failure 11*, 260–269.

Gheorghiade M, Gattis WA, Barbagelata A, Adams KF, Elkayam U, Orlandi C, O'Connor CM (2003). Rationale and study design for a multicenter, randomized, double-blind, placebo-controlled study of the effects of tolvaptan on the acute and chronic outcomes of patients hospitalized with worsening congestive heart failure. *Am Heart J 145*, S51–S54.

Ginès P (2007). Vaptans: A promising therapy in the management of advanced cirrhosis. *J Hepatol 46*, 1150–1152.

Konstam MA, Gheorghiade M, Burnett JC Jr, Grinfeld L, Maggioni AP, Swedberg K, Udelson JE, Zannad F, Cook T, Ouyang J, Zimmer C, Orlandi C. for the Efficacy of Vasopressin Antagonism in Heart Failure Outcome Study with Tolvaptan (EVEREST) investigators (2007). Effects of oral tolvaptan in patients hospitalized for worsening heart failure: The EVEREST Outcome Trial. *JAMA 297*, 1319–1331.

Tikellis C, Cooper MF, Bialkowski K, Johnson CI, Burns WC, Lew RA, Smith A, Thomas MC (2006). *Developmental expression of ACE2 in the SHR kidney: A role in hypertension*? *Kidney Int 70*, 34–41.

Wachtell K, Bella JN, Rokkedal J, Palmieri V, Papademetriou V, Dahlöf B, Aalto T, Gardts E, Devereus RB (2002). Change in diastolic left ventricular filling after one year of antihypertensive treatment: The losartan intervention for endpoint reduction in hypertension (LIFE) study. *Circulation 105*, 1071–1076.

# NEW DRUGS AND NUTRACEUTICALS IN THE TREATMENT OF OSTEOARTHRITIS

Y. Henrotin and J.-Y. Reginster
*Bone and Cartilage Research Unit, Institute of Pathology, University of Liège, Liège, Belgium*

## 1 INTRODUCTION

Osteoarthritis (OA) is one of the most prevalent and chronic diseases affecting the elderly. Its most prominent feature is the progressive destruction of articular cartilage that results in impaired joint motion, severe pain, and, ultimately, disability. Its prevalence and its moderate to severe impact on daily life pose a significant public health problem [1].

Today, a cure for OA remains elusive and the management of OA is largely palliative focusing on the alleviation of symptoms. Current recommendations for the management of OA include a combination of nonpharmacological interventions and pharmacological treatments (Table 1) [2, 3]. Until now, the pharmacological management of OA is dominated by nonsteroidal anti-inflammatory drugs (NSAIDs) and analgesics (mainly paracetamol). Beside these conventional drugs, some new compounds classified as symptomatic slow-acting disease-modifying drugs (SYSADOA) have shown symptomatic efficacy in treating human hip and knee OA. The term SYSADOA covers a range of agents, including glucosamine sulfate (GS), chondroitin sulfate (CS), diacerhein, hyaluronic acid (HA, intra-articularly), and avocado soybean unsaponifiables (ASU). This group of compounds is mainly composed by nutraceuticals, which have been investigated in clinical trials but commonly used as a health food supplement rather than prescribed drugs in Europe and United States. This chapter reviews the available scientific evidence supporting the efficacy, and explaining the mechanism of action, of drugs and nutraceuticals targeting OA.

## 2 OA PATHOPHYSIOLOGY: IDENTIFICATION OF NEW BIOCHEMICAL TARGETS

Osteoarthritis is a heterogeneous group of skeletal disorders characterized by common structural and functional changes in overall joint tissues, including cartilage loss, synovium inflammation, bone sclerosis, cysts, and osteophytes.

The chondrocytes play a critical role in the structural changes in OA cartilage. When chondrocytes experience modification of their environment, they undergo different "phenotype modulations" that result in the expression of a different pattern of genes. The main phenotypic modulations, comprise "dedifferentiation" (characterized by the overexpression of type I collagen), hypertrophy (with the expression of type X collagen, tidemark doubling, and apoptosis as a consequence), recapitulation of development (with the re-expression of type IIa and III collagens), as well as regeneration of mature cartilage (with an overproduction of cartilage-specific matrix macromolecules and cell proliferation) (see a review in [4]). It is unclear if the same chondrocyte may express successively different phenotypes or if chondrocytes from different areas of cartilage express different phenotypes. Nevertheless, it is clear that depending on

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

**TABLE 1 Treatment Modalities Identified for the Treatment of OA[a]**

| Nonpharmacological | Pharmacological | Intra-articular | Surgical |
|---|---|---|---|
| Education | Paracetamol | Corticosteroids | Arthroscopy |
| Exercise | NSAIDs | Hyaluronic acid | Osteotomy |
| Insoles | Opioid analgesic | Tidal irrigation | UKR |
| Orthotic devices | Sex hormones | | TKR |
| Weight loss | SYSADOA | | |
| Laser | Psychotropic drugs | | |
| Spa | Topical NSAID | | |
| Telephone | Topical capsaicin | | |
| Vitamins/minerals | | | |
| Pulsed EMF | | | |
| Ultrasound | | | |
| TENS | | | |
| Acupuncture | | | |
| Nutrients | | | |
| Herbal remedies | | | |

[a]EMF, electromagnitic field therapy; TENS, transcutaneous electrical nerve stimulation, NSAID, nonsteroidal anti-inflammatory drugs; SYSADOA, symptomatic slow-acting drugs for OA [includes avocado/soybean unsaponifiables (ASU), chondroitin, diacerhein and glucosamine]; UKR, unicompartmental knee replacement; TKR, total knee replacement.

*Source*: Modified from [3].

their position in relation to the lesion, chondrocytes show different patterns of gene expression. These phenotypic modulations of chondrocytes are usually associated with structural changes observed in OA cartilage, including matrix mineralization, cell clusters, cartilage thickening, and degradation. They may be initiated by mechanical and biochemical factors.

Environmental influences known to affect the regulation of biosynthetic and catabolic activity of chondrocytes have been shown to include soluble factors (oxygen, ions, vitamins, cytokines, hormones, growth, and differenciation factors), extracellular matrix (for a review see [5]), and mechanical stress [6–9]. Among these environmental factors, cytokines play a critical role in OA pathogenesis and constitute a common target for drugs used to treat OA. The inflammatory cytokines, mainly interleukin (IL)-1β, -6, -17, and -18 and the tumor necrosis factor-α (TNF-α) act in synergy to increase synthesis of metalloproteinases (MPs) and to decrease extracellular matrix components synthesis (see also [10–12]). In opposite, insulin-like growth factor (IGF)-1 and transforming growth factor (TGF)-β1, -2, and -3 stimulate extracellular matrix synthesis and counteract the inhibitory effect of IL-1β on extracellular matrix synthesis [13, 14]. Therefore, it is conceivable that, depending on the relative local concentration of IL-1/TNF and IGF/TGF, the net effect on cartilage would be either a reduced or enhanced synthesis of the cartilage matrix components.

### 2.1 Cartilage Repair Reaction

Chondrocytes respond to injuries that disrupt articular cartilage, but that do not extend into subchondral bone, by cell proliferation and an increased synthesis of matrix macromolecules near the injury [15, 16]. This reaction is considered a repair attempt, histologically characterized by cell clusters surrounded by large acellular area but that do not fill the tissue defect [17, 18]. Clustered chondrocytes produce not only large amounts of aggrecan and type II collagen but also molecules such as type I, IIA, III, or X collagens and tenascin, which are usually absent from normal cartilage [19]. This increased metabolic activity of the chondrocytes might be due to better access of chondrocytes to growth factors from the synovial fluid due to fissuring or loosening of the collagen network or due to the damage of the matrix itself.

However, despite this repair attempt, a net loss of cartilage is observed later in the disease process, suggesting that the cartilage repair reaction is transient and inefficient [20]. The explanation for this failure is uncertain and probably involves a network of biochemical and mechanical factors, including the lower resistance of regenerated tissue to mechanical stress, the overproduction of MPs, and reactive oxygen species (ROS) and cell death. Chondrocyte hypertrophy can contribute to the failure of cartilage repair through reduced expression of type II collagen and aggrecan, increased expression of type X collagen, upregulation of matrix MP-13

(MMP-13), and promotion of pathologic calcification [21–23]. It has recently been shown that OA chondrocytes overexpress IL-1 and TNF receptors, whereas transforming growth factor (TGF)β-RII receptor expression is decreased [20, 24, 25]. The modulation of cytokine and growth factor receptor expression focally modifies cell sensitivity and responses to these factors. As a consequence, chondrocytes are more sensitive to IL-1β and TNF-α, two cytokines, stimulating matrix resorption, but less sensitive to TGFβs, factors stimulating matrix formation. It results with an excess of catabolic events and then cartilage loss.

### 2.2 Cartilage Degradation

Cartilage matrix degradation is mainly conducted by mechanical stress, MPs and ROS (see also the reviews in [26, 27]. The mechanical aspect of OA is note discussed in from this review, which is focused on the biochemical aspect of cartilage degradation.

***Proteolytic Cleavage*** Current knowledge indicates an important involvement of MPs, in OA, mainly MMPs and a disintegrin and metalloproteases with thrombospondin motifs (ADAMTS) family [28–31]. In OA, MMPs are secreted by chondrocytes, as well as synovium cells, including fibroblasts and macrophages. MMPs are synthesized as inactive pro-enzymes that are activated in situ by other MMPs, serine and cysteine proteinases [32–34]. Under physiologic conditions, the activity of MMPs is balanced by natural inhibitors that act at different levels of the proteolytic cascade of MMP activation. In contrast, in OA, a shift of this equilibrium leads to excessive matrix degradation [35].

The initial degradation of collagen fibrils depends on the cleavage at the collagenase site located in triple-helical part of the molecule (between residues $Gly^{775}$ and $Leu^{776}$), most likely by collagenase-1 (MMP-1) and collagenase-3 (MMP-13) [36]. Subsequently, denatured collagen fibrils are thought to be further degraded by gelatinases, with gelatinase A (MMP-2) being the major player in OA [37]. A collagenase that may be of particular importance in the degradation of cartilage in pathologic condition is MMP-13, because this MMP is 5- to 10-fold more active on type II collagen than other collagenases [38, 39]. Studies of gene expression in normal and OA cartilage have shown upregulation of MMP-13 in late-stage OA [40].

One of the earliest changes observed in OA is the depletion of aggrecan due to increased proteolytic cleavage of its core protein, mainly by aggrecanases (members of the ADAMTS family) and stromelysin-1 (MMP-3). The aggrecanases (ADAMTS-1, -4, -5, -8, -9, -15) are glutamyl endopeptidases that cleave to the aggrecan core protein at the $Glu^{373}$-$Ala^{374}$ peptide [41], a cleavage site different from the MMP-3 site ($Asn^{341}$-$Ala^{374}$). Shortly after mechanical injury of cartilage explants, ADAMTS-5 and MMP-3 gene expression highly increases, whereas MMP-1 and MMP-13 show little change in expression [42]. In a study comparing gene expression in chondrocytes from patients with early and late-stage OA, Bau et al. [40] have observed that MMP-13 and ADAM-TS4 expression was very low in normal and early degenerative cartilage but increased in late-stage OA, whereas MMP-3 expression was the highest of the gene levels tested and was downregulated in late-stage OA specimens. Altogether these results indicate that, beside aggrecanases, stromelysin-1 plays an important role in the degradation of aggrecan and that collagenases mainly act in the late stage of OA.

***Oxidative Degradation*** Beside the role played by MMPs, there is an expanded body of evidence indicating that ROS are intimately involved in the pathology of OA [27]. The ROS family includes molecules such as hydrogen peroxide ($H_2O_2$), ions such as the hypochlorite ion ($OCl^-$), radicals such as hydroxyl radical ($^{\bullet}OH$), nitric oxide ($^{\bullet}NO$), or superoxide anion ($O_2^-$), which is an ion and a radical at the same time. Chondrocytes are basically able to produce $^{\bullet}NO$ and $O_2^-$, which subsequently generate derivative radicals including peroxynitrite ($ONOO^-$) and $H_2O_2$ [43–46]. $H_2O_2$ can be transformed into $^{\bullet}OH$ through an iron-ion-catalyzed reaction, named the Fenton reaction ($Fe^{2+} + H_2O_2 \rightarrow Fe^{3+} + {}^{\bullet}OH + OH$-). Both ONOO- and $^{\bullet}OH$ may initiate lipid peroxidation, and the formation of additional long-lived lipidic radicals ($RO^{\bullet}$, $ROO^{\bullet}$). Collectively, chondrocyte lipid peroxidation appears to play a major role in cartilage matrix oxidation and degradation [47]. The overproduction of ROS in pathological cartilage has been indirectly evidenced by the accumulation of lipid peroxidation products (i.e., lipid hydroperoxides and aldehydes) and nitrotyrosine in situ [48]. Recently, we have found elevated nitrated type II collagen peptides in sera of patients with OA, suggesting the formation of peroxynitrite ($ONOO^-$) in OA cartilage [49].

Cellular responses to ROS generation are dependent on the cellular redox status. When the oxidant does not exceed the reducing capacities of cells, ROS are strongly involved in the control of cellular functions including signal transduction. In contrast, in some pathological circumstances when the cellular antioxidant capacity is insufficient to detoxify ROS, oxidative stress may occur that damages not only cellular membranes and nucleic acids but also the extracellular matrix [50–52]. Oxidative stress may also cause cell death and release of cellular content into the extracellular environment [53, 54]. The degradation products and cellular contents including

oxidized molecules may trigger synovial macrophages and fibroblasts, resulting in the generation of a broad range of "inflammatory" mediators including cytokines (i.e., IL-1β and TNF-α), MMPs, and ROS. Thus, a focal "oxidative stress" occurring in cartilage may initiate a vicious circle, constituted by ROS and further inflammatory mediators, leading to chronic synovium inflammation.

### 2.3 Endochondral Ossification

During vertebrate development, cartilage forms a model for many of the bones of the skeleton. This cartilage is replaced by bone (a process termed endochondral ossification) with accompanying long-bone growth until puberty is reached. Several studies have shown that endochondral ossification is reinitiated during osteoarthritis, leading to thickening of the calcified layer, doubling of the tidemark, and, consequently, thinning of the overlying cartilage. Some mechanisms have been advanced to explain endochondral ossification reactivation in OA cartilage. Some authors have suggested that cartilage neovascularization is associated with cartilage mineralization [55, 56], and that the loss the ability of cartilage to remain avascular could trigger endochondral ossification [57]. Although vascular endothelium growth factor (VEGF) expression in cartilage vanishes in normal adults, VEGF is known to be expressed in chondrocytes of OA cartilage [58, 59]. In parallel, the expression of chondromodulin-1, a cartilage-derived antiangiogenic factor, is reduced in OA cartilage [60], leading to an imbalance between pro- and anti-angiogenic factors. The excess of proangiogenic factors promotes vessel in-growth through the microcracks and calcified cartilage vascularization [61]. Another group of researchers have implicated inflammatory cytokines [21], including IL-1β and chemokines (IL-8/CXCL8) in endochondral ossification, suggesting a new mechanistic linkage between joint inflammation and dysregulated chondrocyte differentiation within OA cartilage. In vitro, these cytolines stimulate chondrocyte hypertrophic differentiation [21]. Recently, Cecil et al. [60] have proposed an additional mechanism to explain inflammation-induced chondrocyte hypertrophy. This mechanism involveds the upregulation of RAGE (receptor for advanced glycation end products) and ligand expression (including S100/calgranulins) by human chondrocytes in response to chemokines.

### 2.4 Synovium Inflammation

During OA pathology, the synovial membrane undergoes a moderate inflammation, which is commonly opposite the cartilage lesion, indicating a link between cartilage lesion and synovium inflammation. The inflammatory reaction is triggered by mediators released from cartilage lesion, including inflammatory cytokines, matrix peptides, and microcrystals. In response to these stimuli, synovial cells, including macrophages, T cells, and fibroblasts, produce high levels of inflammatory cytokines, ROS, and MMPs, which, in turn, directly degrade cartilage or stimulate chondrocytes to secrete more MMPs and ROS [62–65], resulting in an amplifying cycle and probably explaining the more rapid progression of chondropathy associated with a flare-up episode [66]. The link between synovitis and chondrolysis has been evidenced by a number of studies. A one year longitudinal study has demonstrated that a larger volume of effusion aspirations for painful OA knee predicted OA progression as defined by an increased narrowing of the joint space [67]. Arthroscopic studies have reported that a baseline synovitis score predicts subsequent worsening of chondropathy in post-traumatic patellofemoral damage [68] and in a large cohort of patients with knee OA [69]. Magnetic resonance imaging and ultrasonography have identified synovitis in early OA, even in the absence of clinical signs of synoviotis, indicating that synovitis is not a feature occurring only in severe and late OA [70]. Immunohistochemical studies have confirmed that synovial tissue from patients with early OA is characterized by mononuclear cell infiltration and the production of pro-inflammatory cytokines and mediators of joint damage, including ROS and MMPs. More recently, it was reported that the presence of inflammatory signs observed at ultrasonography was statistically associated with advanced radiographic disease as expressed by high radiological scores (Kellgren and Lawrence grades 3–4), and at an earlier stage (Kellgren and Lawrence grades 1–2) with clinical signs or symptoms suggestive of inflammatory flare-up [71]. Recently, Sturmer et al. [72] have found an association between pain intensity as measured by a visual analog scale and the serum level of the *C*-reactive protein (CRP) in patients with advanced OA.

### 2.5 Subchondral Bone Sclerosis

One of the mechanisms of initiation and/or progression of OA may be a steep stiffness gradient in the underlying subchondral bone, as the integrity of the overlying cartilage depends on the mechanical properties of its bony bed. Evidence from a primate animal model (*Macaca fascicularis*) of OA indicated that alterations of the bony bed may precede the changes in cartilage [73, 74]. Evidence for and against this hypothesis has recently emerged from both animal models studies [75] and clinical trials [76–78]. As articular cartilage slowly erodes, sclerosis of the subchondral bone also pro-

gresses, and subchondral bone stiffness increases in this tissue, possibly contributing to further mechanical disturbances of the cartilage. Trabecular thickening is characterized by an increase of osteoid substance deposition and by its abnormally low mineralization pattern. Further, it is now well established that OA osteoblasts show an altered phenotype and secrete more alkaline phosphatase (ALP), osteoclacine, interleukin (IL)-1, -6, TGF-β, IGF-1, and urokinase plasminogen activator (uPA) than normal osteoblasts. In contrast, they produce less parathyroid-related peptide (PTHrp) and parathormone receptor [79–81]. Based on their production of leukotriene (LT) $B_4$, prosteglandin (PG) $E_2$, and IL-6, two populations of osteoblasts have been identified. The first group produces high levels of $PGE_2$ and IL-6 but a low level of $LTB_4$, while the second group shows a production of $PGE_2$ and IL-6 similar to that of normal osteoblasts but a high production of $LTB_4$ [82]. These local metabolic changes are favorable to osteoid deposition and sclerosis formation. Furthermore, OA osteoblasts are resistant to parathormone (PTH) stimulation, a situation that can also contribute to abnormal bone remodeling [83]. Beside these quantitative modifications, the OA osteoblast phenotype is characterized by the production of abnormalty high levels of homotremetric type I collagen, which shows a low affinity for calcium [84]. These findings suggest that abnormal osteoblasts play a critical role in subchondral bone sclerosis.

### 2.6 Microcracks and Neovascularization

Osteoblasts secrete a number of biochemical factors that are involved in the remodeling of bone tissue, and these could also contribute to the remodeling of the overlying cartilage in weight-bearing joints after seeping through microcracks in the osteochondral junction. Three elements support this hypothesis: (1) microcracks have been identified at the osteochondral junction, allowing an exchange between the two tissues [85, 86]; (2) bone cell products (HGF, Osteocalcin) are detected in the synovial fluid and in cartilage [87–89]; and (3) primary osteoblasts of OA patients significantly induced glycosaminoglycans release from normal cartilage in vitro [89].

Taking all these observations into consideration, our laboratory has proposed the following hypothesis: OA-altered subchondral osteoblasts may contribute to the degradation of overlying cartilage by seeping of biochemical factors through the microcracks. To verify this hypothesis, we have developed an original co-culture model in which subchondral osteoblasts in the monolayer are cultured together with OA chondrocytes in alginate beads. Thus, we have demonstrated that osteoblasts from the sclerotic (SC) zone of subchondral OA, but not nonsclerotic (N) osteoblasts, downregulated agrecan synthesis and upregulated MMP-3 and MMP-13 expression by chondrocytes [79]. These findings suggest that osteoblasts of the sclerotic subchondral bone produce a pattern of biochemical factors that promote cartilage degradation.

## 3 REQUIREMENT FOR REGISTRATION OF DRUGS IN OA

Both the European Agency for the Evaluation of Medicinal Products (EMEA) and the U.S. Food and Drug Administration (FDA) have issued a guidance document to assist sponsors who are developing drugs, devices, or biological products to treat human OA [90, 91].

The U.S. draft guidance document [91] was released in July 1999, whereas the European "Points to Consider" document was issued in July 1998. Since both of these are slightly outdated, the Group for the Respect of Ethics and Excellence in Science (GREES), a European nonprofit organization seeking to prepare draft documents to be submitted, as working documents, to European regulators, performed an extensive search of the literature and, after setting up experts as consultants, produced new suggestions for the evaluation of guidelines for registration of agents for therapy of structure modification in OA [92, 93].

The current EMEA document offers a classification of antiosteoarthritis therapies dividing medications as symptom-modifying drugs and structure-modifying drugs. Symptom-modifying drugs act on symptoms with no detectable effect on the structural changes of the disease. Registration of such drugs would require demonstration of a favorable effect on symptoms with no clinically significant adverse effects or structural changes of the disease.

Structure-modifying drugs are expected to have an effect on the progression of the pathological changes of OA. These drugs may or may not have independent effects on symptoms.

### 3.1 Symptom-Modifying Drugs

Pain attributable to the target joint is recommended as the primary endpoint for symptom-modifying drugs for OA. Functional disability is an important additional primary endpoint for symptom-modifying drugs. Studies should be powered to demonstrate a significant effect both on pain and on functional disability. Pain should be measured by self-assessment with validated methods. Use and at-rest pain should be assessed separately. A disease-specific and joint-specific instrument such as the

Western Ontario MacMaster University Osteoarthritis Index (WOMAC) or the Lequesne index is recommended to assess disability arising from OA of the hip or the knee.

### 3.2 Structure-Modifying Drugs

Epidemiological data support a relation between structural changes and a long-term clinical outcome. However, the nature and the magnitude of the structural changes that are likely to be clinically relevant in the long-term remain uncertain. Therefore, clinical endpoints such as necessity of joint replacement, a time frame for the need for surgery, and long term clinical evolution are preferable in the assessment of efficacy of such drugs. The radiographic measurement of joint space width or osteophytes was, in 1998, considered as a promising tool to assess the progression of OA, although, at that time, its validity was not fully demonstrated. The EMEA considered that by providing the applicant with data supporting this surrogacy, these changes could be considered as alternative primary endpoints. In any case, the clinical endpoint had to be assessed during the study. Emphasis was put on the reproducibility of X rays on successive visits, which was considered as a prerequisite for a reliable assessment of the progression of OA. The agency requires that the method defines the radio anatomic position of the joint, the beam alignment, and the anatomic landmarks for measurements.

Due to the differences in OA of the knee and OA of the hand, extrapolation of results obtained on the lower limbs to the hand or alternatively extrapolations of data obtained on the hands to lower limbs was considered highly questionable. Compounds having demonstrated efficacy either at the hip or at the knee level would be registered for "treatment of OA of the knee and the hip," whereas a compound having shown efficacy at the level of the hands would be registered for "treatment of OA of the hands". In order to obtain the indication "treatment of OA," a compound should, in the opinion of the EMEA, demonstrate efficacy at the level of the hands and at the level of the knee or the hip.

It was suggested that clinical trials aimed at evaluating the effect of drugs in OA of the hand should be better focused on assessing progression of the disease in proximal and distal interphalangeal joints than in the trapezo-metacarpal joint. The EMEA felt that, although OA of the hand was a potential target for assessing evolution of disease in trials, it is less important clinically than hip or knee disease. OA of the hip was considered as a common and disabling disease, as was OA of the knee. Outcome measures for both symptoms and structure were better validated for medial tibio-femoral disease than for lateral or patello-femoral disease. If the spine was used as a target joint, it was the responsibility of the applicant to demonstrate the validity of the endpoints chosen and their clinical relevance. The agency suggested that, to improve the homogeneity of the patient groups studied, inclusion criteria should limit the target joint to a single site. However, simultaneous assessment of other joints was always considered as possible and such results might generate supportive evidence for "general" efficacy.

To be enrolled in a study, patients should have both symptomatic and structural changes of OA in the target joint. This means pain related to use with radiological evidence of joint space narrowing for OA of the hip and knee and the diagnostic criteria of the *American College of Rheumatology* for hand OA.

For studies of structure-modifying drugs, the inclusion in phase II studies of special subpopulations of subjects who are at high risk for development of OA or rapidly progressive OA may be advantageous (e.g., obese women with unilateral radiographical OA). However, inclusions of a specific risk group in phase III studies might decrease the potential for generalization of the results.

Pivotal study design and duration may differ according to the properties of the drugs.

***1. Modification of Symptoms*** Studies were requested to have a randomized, double-blind, parallel group design. Primary endpoints for efficacy should be evaluated after at least 6 months. To establish that the symptom-modifying drug does not have deleterious effects on the joint, structural changes should be monitored for at least one year. Controlled studies are necessary with the most favorable comparator. Three-arm, placebo, and active controlled studies are recommended by the EMEA.

***2. Modification of Structure*** Studies should have also a randomized, double-blind, parallel group design. Clinical variables or alternatively structural changes when their surrogacy value is proven are required as primary endpoints. When structural changes are chosen as primary endpoint, the magnitude of a clinically relevant effect of a drug on such variables over a specified time should be predetermined, based on previously established findings. Due to the expected mechanism of action of these drugs, long-term studies, no shorter than 2 years, were required both for efficacy and safety assessment. Studies should have a placebo-controlled, parallel group design.

When the FDA issued its draft guidance document in July 1999, the U.S. agency discussed a number of potential claims, including an *improvement in symptoms* claim and a *delay in structural progression* claim. The

guidance also proposes in principle, a *prevention of OA* claim.

The draft guidance document suggested to simultaneously assess pain and function and, as in the EMEA document, recommended the use of the Lequesne index and the WOMAC OA index. This document considered that patient global assessments (visual analogue scale or the likert scale) were overall measurements of whatever the patient feels is most important. They were considered as simple, unidimensional measurements that permit individualization of content and have been extensively validated in OA. The question as to whether patient global measurements should have a primary endpoint status similar to pain and function measurements remained open.

As with the EMEA, the FDA stated that structure was a critical component of OA assessment but that the relationships between structure and/or function and between structure and future outcomes were not well developed. The agency suggested that because OA is a disease of all the tissues in the joint, not just the cartilage, measurement of structure had to be seen broadly and capture important anatomic features, such as osteophytes or ligamentous instability, in addition to cartilage loss. However, since most likely, improvement in joint space narrowing implies cartilage preservation and would thus be reflected in clinical benefit, measurement of the joint space narrowing using an X ray was considered as an appropriate structure measurement. FDA requests that determination of what change in joint space narrowing of the knee or hip is clinically relevant to the patient should be established.

For symptom-modifying drugs, whereas efficacy endpoints include a measurement of pain, a global patient assessment and a self-administered questionnaire, a measurement of structure if the trial lasts a year or more is essential for risk–benefit assessment, even if no structure claim is sought. Trial duration for demonstration of symptom improvement should be at least 3 months.

For structural management, symptom endpoints should be collected regardless of the outcome anticipated. Trials to demonstrate structure improvement should last at least one year. The agency estimates that the claim of slowing joint space narrowing might plausibly be dissociated from other claims when the mechanism of action of the product and/or the size of the effect on slowing of joint space narrowing are suggestive of future clinical benefits. However, in general, the draft guidelines state that products will not be considered for approval for separate claims if (1) they are not anticipated to have different effects on these parameters or (2) they show only small improvements in joint space narrowing without demonstrated effects of symptoms.

The FDA offers a hierarchy of claims for structural outcomes, for example, *normalize the X ray, improve the X ray, and slow joint space narrowing by at least a prespecified amount*.

The draft guidance document introduces the claim *prevention of the occurrence of OA*, using symptomatic and radiographic criteria in patients with OA or in individuals at risk to develop OA in the future. However, the document states that the practicality of this outcome is challenged and that any trial used to demonstrate this outcome should first define the term *new OA*.

Since these two official regulatory documents are somewhat dated, European experts felt it appropriate to make a critical analysis of the available science and to provide European regulatory authorities with a working document that would update, if needed, the guidelines for registration of drugs for OA [92, 93]. The first objective of this working growp was to assess whether new technologies may be used for the assessment of disease-modifying drugs for OA.

Magnetic resonance imaging (MRI) is now used to measure parameters of cartilage morphology and integrity in OA patients. While some data are encouraging, correlation between short-term changes in cartilage structure observed with MRI and long-term radiographic or clinical changes are needed. Hence, the GREES suggests that $MRI_5$ be used as an outcome in phase II studies, but that further data is needed before accepting MRI as a primary endpoint in phase III clinical trials. Biochemical markers of bone and cartilage remodeling are being tested to predict OA and measure disease progression. Recently published data are promising but validation as surrogate endpoints for OA disease progression requires additional study. The GREES suggests that biochemical markers remain limited to "proof of concept" studies or as secondary endpoints in phase II and III clinical trials. However, the GREES emphasizes the importance of acquiring additional information on biochemical markers in order to help better understand the mode of action of drugs to be used in OA.

Regulatory agencies consider that evidence of improvement in clinical outcomes is critical for approval of structure-modifying drugs. Time to total joint replacement surgery is probably the most relevant clinical endpoint for the evaluation of efficacy of a structure-modifying drug. However, at this time, time to surgery cannot be used in clinical trials because of bias by non-disease-related factors such as patient willingness for surgery or economic factors. At this stage, it appears that structure-modifying drugs should demonstrate a significant difference compared to placebo. Benefit should be measured by three primary endpoints: joint space narrowing, pain, and function. Secondary endpoints should

include the percentage of patients who are "responder" (or "failure") patients. The definition of a "failure" patient would be someone with progression of joint space narrowing >0.5 mm over a period of 2–3 years or who has a significant worsening in pain and/or function, based on validated cut-off values. The definition clinically relevant cut-off points for pain and function must be based on data evaluating the natural history of the disease (epidemiological cohorts or placebo groups from long-term studies). These cut-off points should reflect a high propensity, for an individual patient, to later require joint replacement.

This working group organized then a second working group to assess the value of time to joint surgery as a potential therapeutic failure outcome criterion for OA of the hip or knee in the assessment of potential structure-modifying agents.

There are no accepted guidelines for "time to" and "indications for" joint replacement surgery. A limited number of trials have examined joint replacement surgery within the study population. Several parameters, particularly joint space narrowing (interbone distance), correlate with surgical intervention. However, at the level of the knee, none of the parameters have positive predictive value for joint replacement surgery better than 30%. In contrast, lack of significant joint space narrowing has a strong negative predictive value for joint replacement surgery (>90%), which remains after controlling for OA pain severity. At that time, GREES could not recommend time to joint surgery as a primary endpoint of failure for structure-modifying trials of hip or knee OA as the parameter has sensitivity but lacks specificity. In contrast, in existing trials, a lack of progression of joint space narrowing has predictive value of >90% for not having surgery. GREES suggests utilizing joint space narrowing (e.g., >0.3–0.7 mm) combined with a lack of clinically relevant improvement in symptoms (e.g., ≥20–25%) for "failure" as a secondary outcome in structure-modifying trials of the hip and knee.

## 4 WHAT'S A NUTRACEUTICALS?

Nutraceuticals are natural, bioactive chemical compounds that have health-promoting, disease-preventing, or medicinal properties (Table 2). They can be provided as dietary supplements delivered over the counter, as prescribed drugs (in a limited number of countries), or as functional food (see also [94]). As functional food, nutraceuticals are incorporated into food and drink products consumed as part of daily diet. A functional food can be defined as any food or food ingredient that may provide a health benefit beyond the traditional nutrients it contains [the Institute of Medecine's Food and Nutrition Board (IOM/FNB, 1994]. A functional food can be distinguished from a traditional food "if it is satisfactorily demonstrated to affect beneficially one or more target functions in the body, beyond adequate nutritional effects in a way which is relevant to either the state of well-being and health or the reduction of the risk of a disease" [95, p. 56]. A food product can be made functional by eliminating a deleterious ingredient, by adding a beneficial ingredient, by increasing the concentration of an ingredient known to have beneficial effects, or by increasing the bioavaibility or stability of a beneficial ingredient [96, 97].

**TABLE 2 Nutraceuticals in OA**

| | |
|---|---|
| Ascorbic acid | Hyaluronic acid |
| Avocado/soybean unsaponifiables | Hydrolysate collagen |
| *Boswellia serrata* | Methylsulfonylmethane |
| Bromelain | Milk and "hyperimmune milk" |
| Cat's claw | n-3 PUFAs |
| Chondroitin sulfate | *Ribes nigrum* |
| Cetyl myristoleic oil | *Rosa canina* |
| Collagen hydrolysate | Selenium |
| Glucosamine $SO_4$/Acetyl/HCL | Strontium |
| Chitosan | Silicium |
| Devil's claw | Turmeric |
| Flavonoids | Vitamin D |
| Green lipid mussel | Vitamin E |
| Ginger | Willow |

## 5 GLUCOSAMINE SULFATE (GS) AND CHONDROITIN SULFATE (CS)

CS and GS are aminosaccharides that act as preferred substrates for the biosynthesis of glycosaminoglycan chains and subsequently for the production of aggrecan. Glucosamine may be administrated as a prescription drug or as a dietary supplement over the counter. In fact, the patented crystalline GS 1500-mg once-a-day soluble powder preparation used in the most relevant clinical trials is a prescription drug in most European and some non-European countries, whereas the Dietary Supplement Health and Education Act of 1994 favored the appearance on the dietary supplement market in the United States and other countries of several poorly documented glucosamine salts (e.g., hydrochloride), derivatives (e.g., N-acetyl-glucosamine), or dosage forms and regimens.

Recently, the oral bioavailability and pharmacokinetic ability of the patented crystalline GS has been determined in humans, with a newly developed and

validated liquid chromatography method with mass spectrometry detection (LC-MS/MS), having a low limit of quantitation (LOQ = 6.25 ng/mL = 0.03 μ*M*) [98]. Endogenous plasma levels of glucosamine were detected (10.4–204 ng/mL). After oral administration of the standard therapeutic once-daily dosage of 1500 mg crystalline GS, glucosamine plasma levels increased over 30-fold from baseline and peak at about 10 μ*M* (9.92 μ*M* or 1777.6 ng/mL) after 3–4 h, indicating that the drug is orally bioavailable at concentrations that are in line with those found to be effective in in vitro models. After 14-day consecutive once-daily administrations of 1500 mg GS, glucosamine maximum plasma concentrations in healthy subjects (8.9 ± 0.7 μ*M* or 1602 ± 123 ng/mL) are similar to those reached at corresponding sampling times in both plasma (7.9 ± 1.9 μ*M* or 1419 ± 312 ng/mL) and synovial fluid (7.2 ± 1.4 μ*M* or 1291 ± 257 ng/mL) of knee OA patients [99]. The compound was eliminated with $t_{1/2}$ estimated to be 15 h, supporting once-daily dosing. Conversely, other glucosamine formulation (i.e., glucosamine hydrochloride 500 mg three times a day) provide peak plasma levels in OA patients that are below 3 μ*M* [98, 99].

As common in mechanistic studies, most of the in vitro studies performed so far to characterize GS mechanism of action have used higher concentrations than those found after therapeutic doses in biological fluids in humans (that were previously unknown). In vitro, CS (100–1000 μg/mL) and GS (50–100 μg/mL) increased the production of proteoglycans by chondrocytes, with no detectable effects on collagens, or cell proliferation [100]. They also partially counteracted the effect of IL-1β on proteoglycan and MMP synthesis, suggesting that these compounds may interfere with the deleterious effects of one of the most active mediators of OA pathology [101–103]. Both CS and GS inhibit IL-1-induced COX-2 and $PGE_2$ synthesis, with no effect on the constitutively expressed COX-1, indicating that these compounds may have anti-inflammatory properties. GS may also have cyclo-oxygenase-independent anti-inflammatory properties, by inhibiting nuclear factor-κB (NF-κB) binding activity, and subsequently decreasing anti-inflammatory cytokines (i.e., IL-6, IL-1β, TNFα), metalloproteases (MMP-3 and ADAMTS5), and ROS production ($^{\bullet}NO$) [104, 105]. Effects connected with NF-κB decreased activation are already evident at concentrations of 10 μ*M*. Finally, a strong ROS scavenging effect was also observed in PMA-activated leukocytes by GS ($IC_{50}$ 0.16 ± 0.02 m*M*) [106]. At this time, the role played by sulfate in the biological activities of these compounds remains controversial. Sulfate-containing chondroprotective agents selectively increase serum concentrations of inorganic sulfate, suggesting that sulfate may mediate the therapeutic effect of GS or other sulfur-containing agents [107, 108]. However, until now, there is only one in vitro study comparing in similar cultural conditions the effects of acetylated glucosamine, sulfated glucosamine, and galactosamine. This study demonstrated that GS, but not acetyl glucosamine or galactosamine, inhibited in a dose-dependent manner (10–1000 μg/mL) the synthesis of pro-inflammatory mediators by IL-1β-stimulated human osteoarthritic chondrocytes through an NF-κB dependent mechanism [104]. This study underlined the role played by sulfate substitution in the biological activity of aminosaccharides. In contrast, Neil et al. [109] have reported that glucosamine-HCL at concentration of 10 μg/mL significantly reduced IL-1β-induced messenger ribonucleic acid (mRNA) expression of MMP-13, aggrecanase-1, COX-2, iNOS, and c-Jun-N-terminal kinase (JNK), while CS had no effect on gene expression at the concentrations tested. In the same way, glucosamine-HCL also reduces $NO_2^-$ and $PGE_2$ production induced by IL-1β and is able to fully prevent the upregulation of stromelysin (MMP-3) mRNA expression by IL-1β. Furthermore, the addition of glucosamine to rat chondrocytes treated with IL-1β decreased the activation of NF-κB, but not activator protein-1 and increased the expression of mRNA-encoding type IL-1β receptor [110]. Finally, Chan et al. [111] reported that a physiologically relevant concentration (1–20 μg/mL) of glucosamine-HCL downregulated IL-1β-induced gene expression of inducible nitric oxide synthase (iNOS) and cyclo-oxygenase-2 (COX-2) and repressed the cytokine-stimulated microsomal prostaglandin E synthase-1 (mPGEs1) transcript leading to the reduced synthesis of $^{\bullet}NO$ and $PGE_2$ in cartilage [112]. These data suggest that sulfate is probably not the only active group involved in the effects of GS in a medium supplemented in sulfur amino acids.

Crystalline GS was found to be effective in controlling OA symptoms in several clinical trials [113–115]. In particular, two randomized, placebo-controlled, double-blind trials of 3-year duration in knee OA patients showed that this symptom-modifying effect is sustained over long-term courses [116, 117]. Recently, two multicenter, randomized, placebo-controlled, double-blind clinical studies have been assessing the efficacy of oral glucosamine over a medium-term—6 months—treatment course in knee OA, using a standard symptomatic medication as a reference. These studies were named the Glucosamine/Chondroitin Arthritis Intervention Trial (GAIT) [118] and Glucosamine Unum in Die Efficacy (GUIDE) studies [119]. GAIT was performed by the National Institutes of Health (NIH) in the United States with nutraceutical glucosamine-HCL 500 mg t.i.d., while GUIDE has been running in Europe with the prescription GS formulation

1500 mg once a day. In the GAIT study, 1587 patients with symptomatic knee OA were randomly assigned to receive 500 mg glucosamine-HCL three times daily, 400 mg of sodium chondroitin sulfate three times daily, 500 mg of glucosamine-HCL plus 400 mg of chondroitin sulfate three times daily, 200 mg of celecoxib daily, or a placebo. The primary outcome measure was a response treatment, defined a priori by expert consensus as 20% decrease in the summed score for the WOMAC pain subscale from baseline to week 24. Analysis of the primary outcome measure revealed that the rate of response to glucosamine-HCL and chondroitin sulfate, either alone or in combination, was not significantly higher than the rate of response to placebo, while the rate of response to the celecoxib control was significantly higher than that for the placebo control. Nevertheless, the treatment effects in the moderate-to-severe pain stratum (WOMAC pain score, 301–400) were more substantial. Results for the primary outcome in this stratum, which included 22% of the patients in the trial, indicated that combined treatment was significantly more effective than placebo. Nevertheless, these results must be interpreted with caution and in regard to the following limitation. First, the high rate of response to placebo (60.1%) and the relatively mild degree of pain from OA among the participants may have limited the ability to detect benefits of the treatment. Second, It is confirmed that glucosamine-HCL is, at best, only partially effective, probably because of the absence of sulfates and, especially, because of the low glucosamine peaks achieved with the T.I.D. administration.

In the GUIDE study, 318 subjects with knee OA were randomly assigned to receive 1000 mg acetaminophen three times a day (total 3 g acetaminophen/day, i.e., the approved daily dose in Europe), 1500 mg crystalline GS once a day, or the appropriated placebo. The Lequesne index (change in score at 6 months) was designated as the primary efficacy outcome in the protocol, since it was used to calculate the study sample size. GS, but not acetaminophen, significantly reduced the Lequesne score, indicating that crystalline GS at the oral once-daily dose of 1500 mg should be the preferred symptomatic medication for medium- and long-term treatment in patients with knee OA.

## 6 HYALURONIC ACID

Several hyaluronic acid (HA) preparations are currently available in the United State, Canada, and Europe and are used to treat knee, hip, carpometacarpal joint of the thumb, or ankle OA (Table 3). These preparations require three (Synvisc, high-molecular-weight cross-linked HA, Orthovisc) to five (Hyalgan, Supartz) intra-articular (IA) injections (1–2 mL containing 10–20 mg/mL HA per injection depending on the joint). Significant adverse effects are limited to rare acute local reactions, such as transient pain or warm joint lasting for one night. Recently, four systematic review and meta-analysis [3, 120–122] including randomized controlled trials published until 2004 have reached different conclusion about the effectiveness of HA on knee OA symptoms. Arrich et al. [121] concluded that HA has not been proven clinically and may be associated with a greater risk of adverse events. Lo et al. [120] concluded that at best there is a small effect compared to placebo injection. Maheu and collaborators [122] have reviewed 24 clinical trials on hyalgan. They concluded that 3–5 weekly injections of Hyalgan significantly improved the pain and functional status of patients with OA and that, although the onset of improvement was delayed by 3–4 weeks, the effect can last at least 6 months and more after treatment cessation. Finally, the fourth systematic review [3], which is an update of the recommendations of the European League Against Rheumatism (EULAR) for the management of knee OA, concluded that there is evidence to support the efficacy of HA in the management of knee OA both for pain reduction and functional improvement.

**TABLE 3 Hyaluronic Acid Preparations**

| Trade Name | Corporation | Average MW (kDa) |
|---|---|---|
| Synvisc (Hylan G-20) | Biomatrix (Canada) | 6000–7000 |
| Healon | Pharmacia/Upjohn (Sweden) | 1900–3900 |
| Orthovisc | Anika (USA) | 1700–2900 |
| Arthrum | LCA (France) | 2000 |
| Adant | Meiji Seika (Japan) | 900–1200 |
| SupArtz, Artz, Artzal | Seikagaku (Japan) | 600–1200 |
| Ostenil | Chemedica (France) | 1200 |
| Hyalgan | Fidia (Italy) | 500–730 |

At least, 10 additional clinical trials have been published since 2004 [123–131]. A large majority of these clinical trials reported beneficial effects of HA on OA symptoms, not only for knee OA but also for ankle or carpometacarpal joint of the thumb OA [123–128, 131–133]. Recently, Fuchs et al. [126] reported that a single course of three injections of sodium hyaluronate (Ostenil mini) was effective in relieving pain and improving joint function in patients with OA of the carpometacarpal joint of the thumbs. Although in comparison with triamcinolone its effects are achieved more slowly, the results indicate a superior long-lasting effect

of hyaluronan at 6 months after end of treatment period. The effects of intra-articular hyaluronic acid (Artz) have been investigated in a prospective study including 93 patients with ankle OA. The authors concluded that five weekly IA injections of Artz provided pain relief and functional improvement in patients with radiological Kellgren-Lawrence grades I and II ankle OA. The clinical effect was rapid at 1 week and may last for 6 months or more [131]. Similarly, Hyalgan provided sustained relief of pain and improved function in patients with ankle OA compared to placebo [133]. Two recent studies [127, 128] have compared the efficacy of hylan G-F 20 (Synvics) and sodium hyaluronan (Orthovics) in the treatment of knee OA. With a molecular weight of 6000 kDa, the elastoviscous properties of hylan G-F20 (Synvics) are similar to those of the synovial fluid of healthy young adults [134]. The viscoelastic properties of hyaluronan (Orthovics) is lower than that of the hyaluronan in normal healthy synovial fluid, with a molecular weight of 1550 kDa. Both hyaluronan preparations induced a marked reduction of pain and improved function of patients with knee OA, but there was no significant difference between the two preparations.

Aside from the symptomatic efficacy of intra-articularly injected HA in OA, growing preclinical and clinical evidence indicates that HA may modify the progression of OA. In animal OA induced experimentally by menisectomy or anterior cruciate ligament transection, Artz, Hyalgan and Synvisc have been shown to reduce OA lesion progression (evaluated macroscopically or histologically) compared to saline injection (see details in [135]). In humans, clinical evidence is limited to short-term studies (≤1 year) evaluating disease-modifying effects of Hyalgan. The majority of these studies used arthroscopy or morphology as the outcome parameter [136–138]. Until now, only one large multicenter, double-blind, randomized study involving 408 patients with OA of the knee was carried out using joint space narrowing over 1 year as primary outcome measure [139]. In this study, patients received 3 courses of 3 weekly injections of Hyalgan or placebo (saline every 4 months). In comparison with placebo, Hyalgan significantly reduced the progression of joint space narrowing in the subset of patients who had higher joint space area at study entry (>4.6 mm). In contrast, there was no statistical difference between treatment groups when comparing the change in joint space width (JSW) in the patients with more severe OA at entry (<4.6 mm). It appears that HA acts on cartilage through various mechanisms of action, including the inhibition of catabolic cytokine (IL-1β, TNF-α) activities and synthesis [140, 141], the reduction of ROS production [142], and the stimulation of matrix synthesis [141, 143]. These effects are conducted by binding cluster of differentiation 44 (CD44) receptors [144, 145] and intracellular HA binding proteins (receptors) called receptors for hyaluronan-mediated mobility (RHAMMs) (see details in [146]).

Based on this global overview of the HA clinical trials, HA is particularly indicated—In painful OA despite other pharmacological and nonpharmacological therapies, radiologically moderate, with no or mild effusion, in the absence of mechanical symptoms—as the second-line treatment after a period of acute radiological chondrolysis when knee "dryness" has been obtained—in painful "endstage" OA when total joint replacement is contraindicated or refused by the patient and—as an alternative to NSAIDs when contraindicated, not tolerated, or inefficient.

## 7 LIPIDS

### 7.1 Avocado–Soybean Unsaponifiables

Avocado–soybean unsaponifiables, or ASUs are derived entirely from vegetable sources. ASU have been used in France as prescription treatment for OA, subsidized by the government, and meticulously tracked for safety for over 15 years. Piascledine300 (Laboratoires Expansciences, Courbevoie, France) is the name of the French prescription form of ASU. A Piascledine300 pill is composed by 100 mg of avocado oil and 200 mg of soybean oil. In other countries, ASU mixtures are available as a supplement and their compositions are different to that of Piascledine300. Indeed, ASU contained in Piascledine300 result from a specific extraction procedure that confers chemical particularities to this drug. Therefore, the extrapolation of the therapeutic effects of Piascledine300 to over-the-counter preparations is hazardous. Finally, eating avocado and soy, even in large amounts, does not provide enough of the unsaponifiable to have a therapeutic effect. In fact, less than one percent of the avocado and soy oils make up the unsaponifiable fraction of these oils.

In vitro, Piascledine300 displayed anabolic, anticatabolic, and anti-inflammatory effects on chondrocytes. Piascledine300 increased collagen synthesis [147] and inhibited the spontaneous and IL-1β-induced collagenase activity [148, 149]. This drug increased the basal synthesis of aggrecan and reversed the IL-1β-induced reduction of aggrecan synthesis [150]. It reduced the spontaneous and IL-1β-induced production of MMP-3, IL-6, IL-8, and $PGE_2$ while weakly reversing the IL-1β-induced decrease in tissue inhibitor of metalloproteinases-1 (TIMP-1) production [148–150]. It decreased the spontaneous production of ˙NO and macrophage inflammatory protein-1β (MIP-1β) [27] but

stimulated the expression of TGF-β1 and plasminogen activator inhibitor-1 [151]. Recently, it was reported that Piascledine300 inhibited P42/44 mitogen-activated protein kinases (MAPK) activity and NF-κB binding activity induced by IL-1β or a mechanical stress, suggesting a new mechanism of action for this drug [152]. Furthermore, we have demonstrated that Piascledine300 fully prevented the inhibitory effect of OA osteoblasts on cartilage matrix protein production by human OA chondrocytes [153]. This finding suggests that Piascledine300 could act on subchondral bone and thus have a beneficial effect on the overlying cartilage. Furthermore, this finding contributes to explaining the effect of Piascledine300 on subchondral bone sclerosis observed in sheep with lateral menisectomy [154]. In this animal model, Piascledine300 (900 mg/5 days per week) for 6 months reduced the toluidine blue stain in cartilage and prevented subchondral bone sclerosis in the inner zone of the lateral tibial plateau but not focal cartilage lesion.

Four double-blind placebo-controlled randomized controlled triads (RCTs) and one systematic review evaluated Piascledine300 on hip and knee OA [155–158]. In two 3-month RCTs, one on knee and hip OA [156] and one solely on knee OA [155], 300 mg of Piascledine300 once a day decreased NSAID intake. No statistical difference in any primary or secondary endpoints was detected between 300 and 600 mg once a day [155]. In a 6-month RCT on knee and hip OA, 300 mg once a day resulted in an improved lequesne functional index compared to placebo [157]. Only one study has evaluated the structure-modifying effect of Piascledine300 on OA. In this 2-year RCT on hip OA, Piascledine300 once a day did not slow down joint space width narrowing [158]. In addition, none of the secondary endpoints (Lequesne functional index, VAS pain, NSAID intake, and patients' and investigators' global assessments) were statistically different from placebo after one year. A post hoc analysis suggested that ASU might decrease joint space width narrowing in patients with severe OA (JSW ≤ median at baseline). In summary, while Piascledine300 might display medium-term (several months) symptom-modifying effects on knee and hip OA, their symptom-modifying effects at long term (over a year) have not been confirmed. Furthermore, the structure-modifying effect of ASU needs to be further investigated. Finally, its safety record is good with no significant differences in adverse effects between Piascledine300 and placebo.

### 7.2 Omega-3 Polyunsaturated Fatty Acids (n-3 PUFAs)

PUFAs are classified into n-3, n-6, and n-9 depending on the position of the last double bond along the fatty acid chain. The main dietary PUFAs are n-3 (such as linoleic acid and eicopentenoic acid) found in vegetable oils (soybean and canola oils), flaxseeds, walnuts, or fish oils, and n-6 (such as linoleic acid and arachidonic acid) found in vegetable oils (corn, soybean, sunflower) and in meat. Modern western diet is relatively low in n-3 PUFAs and relatively high in n-6 compared with eastern diets or diets in preindustrialized societies. A high n-3 intake correlates with low incidence of cardiovascular and inflammatory diseases [159, 160].

The articular cartilage content of arachidonic acid, an n-6 precursor of pro-inflammatory prostaglandins $PGE_2$, correlates with severe OA [161]. Dietary lipid interventions in animals modified the PUFA composition of articular cartilage [162]. The in vitro effects of 10–100 μg/mL of n-3 fatty acid (linoleic, eicopentaenoic, and docosahexaenoic acids) on chondrocytes have been investigated [163, 164]. The n-3 fatty acid did not affect the spontaneous GAG synthesis or its IL-1-induced decrease, but dose-dependently inhibited the IL-1-induced GAG release. They dose-dependently inhibited the IL-1-induced aggrecanase activity and basal and collagenase activity, whereas, in contrast, n-6 stimulated the basal aggrecanase and collagenase activity. They also decreased the IL-1-induced mRNA expression of ADAMTS-4 (aggrecanase), COX-3, 5-LOX, FLAP (5-LOX-activating protein), IL-1α and TNF-α, and the basal mRNA levels of these genes [165]. Finally, they decreased the basal and IL-1β-induced mRNA and protein levels of MMP-3 and -13. All these parameters were unaffected by n-6 PUFAs [163, 164]. Taken together, these results indicate that n-3 PUFAs have anti-inflammatory and anticatabolic properties. Nevertheless, a too low n-6 : n-3 ratio can be detrimental. A diet with very low levels of n-6 PUFAs induced occasional surface irregularities and localized proteoglycan depletion in cartilage in rats [162].

### 7.3 Lipid Extract from New Zealand Green-Lipped Mussel (Perna canaliculus)

The reason people first marketed p. *canaliculus* stems from the observation that native New Zealanders (Maoris) who lived near the sea had a much lower incidence of arthritis than their relatives who lived inland. Their diets were similar with the exception that the Maoris on the coast consumed large amounts of the green-lipped mussel.

Green-lipped mussel *(P. canaliculus)* is a New Zealand shellfish from which an extract has been shown to be useful in the treatment of rheumatoid arthritis (RA) and OA. The patented and stabilized lipid extract, with the trademark name of Lyprinol, contains a variety of fatty acids, including eicosatetraenoic acid (ETA), which is an omega-3 fatty acid.

Lyprinol acts as a dual inhibitor of cyclo-oxygenase and lipoxygenase pathways, limiting the production of protaglandins and leukotrienes, two mediators of joint inflammation and pain [166]. Lyprinol (50 mg) is administrated orally in capsule form (~260 mg, 2–4 capsules per day according to the manufacturer recommendations) containing natural monosaturated olive oil (100 mg) and vitamin E (D-alpha-tocopherol—0.225 mg) as an antioxidant.

In a 6-week double-blind placebo-controlled RCT in dogs, 1 g per day of Lyprinol directly incorporated as 0.3% in a dry diet significantly improved total arthritic score, joint pain, and joint swelling [167]. More than 50% of the dogs demonstrated a 30% or greater reduction in total arthritic score. However, it is unclear whether these dogs suffered from OA specifically and the disease severity was not described.

In a multicenter 2-month clinical trial, 60 patients with symptomatic knee and hip OA, receiving Lyprinol at a dose of 2 capsules twice a day, showed a significant reduction of the Lequesne functional index, suggesting that Lyprinol might improve the algofunctional status of OA patients without adverse effects [168].

## 8 VITAMINS AND MINERALS

The effects of vitamins C, E, and B on OA have been formally investigated. One obvious candidate not yet evaluated is vitamin D (vit D). However, structural and metabolic changes in OA subchondral bone suggest that vit D could play a role in this process. In addition, cartilage mineralization is an important feature of severe OA, and the expression of vit D receptors is upregulated in human OA chondrocytes [30]. Hence, suboptimal levels of vit D may impair bone and cartilage metabolism and predispose one to OA. The Framingham study found a threefold increase in risk of OA progression for patients in the middle and lowest tertile of serum levels of 25 vitD [169]. Low serum levels of vit D also predicted loss of joint space and osteophyte growth. The women's study of osteoporotic fractures found that the risk of incident hip OA defined by joint space narrowing was increased for patients in the middle and lowest tertiles of serum levels of 25 vit$D_3$ [170]. Based on these data, an RCT testing the efficacy of vit D seems to be required.

### 8.1 Vitamin C

In vitro, vitamin C (vit C) increases protein and proteoglycan synthesis by articular chondrocytes [171–173] and increases the mRNA levels of type I and II collagens [171, 174], aggrecan, and aprolyl 4-hydroxylase [171]. It decreases the LPS-induced GAG release [175]. It also affects the activities of lysosomal enzymes decreasing the activities of arylsulfatase A and B, but increasing the activities of acid phosphatase in normal and OA chondrocytes [172].

Vit C can cross-link collagen and other proteins, by nonenzymatic glycation leading to the formation of advanced glycation end (AGE) products. Threose, a metabolite of ascorbic acid, increases the AGE content of articular cartilage in vitro [176]. These cross-links increase the stiffness of the collagen network, which is hypothesized to increase cartilage susceptibility to OA. High in vitro levels of vit C (756 m*M*) also increase protein carbonylation, one type of oxidative damage [171]. However, when guinea pigs were fed with diets containing different levels of vit C, no modification in the AGE content of articular cartilage was detected [177].

In guinea pigs, 8 months of exposure to a very high dose (150 mg per day) of vit C decreased the severity of surgically induced knee OA [178] but increased severity of spontaneous OA [177], despite the ability of ascorbic acid to increase cartilage collagen content. A third guinea pig trial stated, on the contrary, without providing any details, that a fivefold increase of ascorbic acid to the drinking water (equivalent of 1 g per liter) had a slight chondroprotective effect on the development of spontaneous lesions but not on surgically induced OA [179].

The Framingham epidemiological study found a threefold reduction in risk of OA progression for both the middle and highest tertile of vit C intake (an average of 152 and 430 mg per day, respectively) and an inverse association between vit C intake and cartilage loss [169]. No association between vit C intake and osteophyte growth or disease incidence was found. A protective effect was also demonstrated for vit E and β-carotene, but this effect was less evident and consistent than in the case of vit C. The favorable effect of vit C was probably related to the fact that vit C is not only an antioxidant but is also an important cofactor in type II collagen synthesis. In a 14-day double-blind cross-over RCT on knee and hip OA, 1 twice a day of calcium ascorbate was more efficient than placebo in decreasing pain VAS [180]. However, the measured effect was small (a 4.6-mm decrease from a starting basal level of 50 mm) and obtained with a dose equal to the upper tolerable intake level for adults, well above the current 60–200 mg per day recommended daily allowances (RDAs). The long-term safety of such high doses of vit C in elderly OA patients needs to be evaluated while efficacy needs to be confirmed by longer RCTs. In view of these conflicting results and in the absence of strong evidence of efficacy in humans, it was recommended that vit C intake for OA should not be higher than the currently RDAs (90 mg per day for men and 75 mg per day for women) [177].

### 8.2 Vitamin E

Natural vitamin E (vit E) comprises eight different forms, α-, β-, γ-, and δ-tocopherol and α-, β-, γ, and δ-tocotreinol, found solely in plants. One of the most important food sources of vit E is edible plant oils. Synthetic α-tocopherols (the eight other possible side-chain stereisomers besides the natural one) and their esters (α-tocopherylsuccinate, α-tocopherylacetate) also exist. α-Tocopherylacetate is often used commercially because vit E esterification protects it from oxidation. In the human body, the ester is rapidly cleaved by cellular esterases making natural vit E available.

Only a few studies have investigated the in vitro effects of vit E on chondrocytes. Tiku et al. [47] showed that when chondrocytes were submitted to an oxidative stress, vit E reduced the catabolism of collagen by preventing the protein oxidation mediated by aldehydic downproducts of lipid peroxidation. Vit E strongly increased the sulfate incorporation while slightly reducing glycoprotein and glycolipid synthesis. Like vit C, vit E affected the activities of lysosomial enzymes: It decreased the activities of arylsulfatase A and of acid phosphatase in cultures of human articular chondrocytes [181]. However, vit E did not affect the LPS-induced catabolism of GAGs and did not prevent synoviocyte apoptosis induced by superoxide anions [182].

Five RCTs have tested the natural form of vit E or α-tocopherylacetate. Two trials concluded that vit E was more efficient than placebo in decreasing pain. In a small 10-day single-blind cross-over RCT on mainly spondylosis, 600 mg of vit E per day was superior to placebo as assessed by a patient's questionnaire [183] while in a 6-week double-blind RCT on OA, 400 IU of α-tocopherylacetate once a day was superior to placebo as assessed by a joint patient/doctor, global assessment of pain [184]. One trial suggested that vit E was no less efficient than diclofenac in decreasing pain. In a 3-week double-blind RCT on OA, no significant difference was found between 544 mg of α-tocopherylacetate 3 times a day or 50 mg diclofenac 3 times a day on VAS of pain [185]. However, the two most recent trials failed to show any benefit over placebo on knee OA. In a 6-month double-blind RCT, 500 IU of vit E a day showed no symptomatic benefit over placebo as assessed by WOMAC [186], while in a 2-year double-blind RCT, 500 IU of vit E a day showed no symptomatic or structure-modifying benefit over placebo as assessed by WOMAC or MRI, respectively [187]. Although 3 out of the 5 RCTs concluded that vit E decreased pain, the two longest, largest, and highest quality trials failed to detect any symptomatic or structural effects in knee OA, indicating that, at least for knee OA, vit E has no medium-term beneficial effect. No significant adverse event was reported.

### 8.3 Vitamins B

Vitamin $B_6$ deprivation induced disorders in avian articular cartilage, including a decreased content and increased extractability of total sulfated PGs from articular cartilage with guanidine HCl. These findings suggest that vitamin $B_6$ could be a necessary diet compound to maintain cartilage homeostasis [188].

Until now, only one double-blind study has demonstrated the efficacy of niacinamide (vitamin $B_3$) in OA. In this 3-month double-blind RCT, very high doses of niacinamide (500 mg 6 times per day) were more efficient than placebo in reducing drug intake and OA symptoms at the exception of pain [189]. A 2-month double-blind cross-over RCT in hand OA comparing 6400 μg of folate with or without 20 μg of cobalamin (vitamin $B_{12}$) daily to placebo failed to show significant effects on mean hand grip values [190]. Nevertheless, among patients with low hand grip values, the acetaminophen consumption was fewer in the cobalamin/folate group than in the placebo group. Recently, it was reported that the chondroprotective effect of glucosamine-HCL/CS mixture was enhanced by the addition of fursultiamine, a vitamin $B_1$ derivative, in rabbit OA induced surgically by menisectomy [191]. This effect was associated with a reduction in the level of MMP-1, which is known to play an important role in the pathogenesis of OA lesions. In these studies, no severe adverse effects have been reported. Altogether, these findings suggest that vit B supplementation may be of value in the treatment of OA; however, even larger studies are needed to confirm the beneficial effect of vit B on the natural course of OA.

### 8.4 Cocktail of Vitamins and Selenium

Because vitamins are rarely consumed separately, it makes sense to study the effects of a cocktail of vitamins. Two cocktails of vitamins containing selenium and a cofactor of the anti-oxidative enzyme glutathione peroxidase have been investigated. In a small pilot 6-month double-blind placebo-controlled RCT, a mixture of vitamins A, C, and E (in undisclosed amount) and 144 mg of selenium per day had no effect on VAS of pain or stiffness [192]. Vitamins A, C, E, $B_2$, and $B_6$ and selenium decreased OA incidence and severity in STR/1N mice, possibly through an antioxidant effect as suggested by the increased expression of two antioxidative enzymes, glutathione peroxidase and superoxide dismutase in cartilage [193].

## 9 PHYTOCHEMICALS AND PLANT EXTRACTS

### 9.1 Bromelain

Bromelain is present in all parts of the pineapple plant (*Anana cosmosus* Merr, mainly var Cayenne from the family of Bromeliaceae), but the stem is the most common commercial source. Bromelain is marketed as a crude, aqueous extract that contains a mixture of sulfur-containing protein-digesting enzymes and several other substances in smaller quantities, including peroxidase, acid phosphatase, protease inhibitors, and calcium. Bromelain was suggested to have anti-inflammatory, analgesic, antiedematous, antithrombotic, and fibrinolytic effects (see [194, 195] for a review). Three different preparations containing bromelain mixed with varied enzymes have been tested in nine trials on knee OA (see [194, 196] for a review and references). Bromelain was taken in tablets coated to resist stomach digestion at a daily dose ranging from 270 to 1890 mg. All nine trials might have been insufficiently powered, were of short duration (3–6 weeks), and enrolled OA patients with flare-up episodes. Most used the Lequesne functional index as a primary endpoint. Bromelain was shown to be no less or even more effective than 100–150 mg of diclofenac per day in seven trials, it was also not more efficient than placebo in two other trials. This absence of efficacy over placebo, coupled with the fact that in some diclofenac-controlled trials the Lequesne functional index continued to decrease in both groups even 4 weeks after the end of treatment [197], suggested a possible spontaneous resolution of the flare-up episode rather than a real efficacy of bromelain [194]. Longer trials of higher quality were advocated to confirm the efficacy of bromelain [194]. In one trial, a high dose of bromelain (945 mg/day) induced a higher incidence of adverse effects and drop-outs compared to diclofenac (questioning its safety and tolerability) [198].

### 9.2 Rosa canina

A standardized rose-hip powder made from the seeds and husks of the fruits from a subtype of *R. canina* (Hyben Vital), the common wild briar rose of English hedgerows, was evaluated in three RCTs. This extract contains a large amount of flavonoids and vit C and possess powerful antioxidant properties [199]. In a 4-month double-blind RCT on hip and knee OA, 2500 mg of this powder twice a day did not improve active or passive mobility (joint rotation, flexion, and extension) compared to placebo, except passive hip flexion [200]. In a 3-month cross-over double-blind RCT, 2500 mg of *Rosa* powder twice a day decreased pain (as measured by a categorical scale) more efficiently than placebo when placebo was administrated first but not when it was administrated second [201]. No pain difference was found when the two groups were compared before cross-over. This RCT, which enrolled patients with OA in various joints, did not include a washout period. In a 3-month cross-over double-blind RCT on knee and hip OA, which included a 3-week washout period, 2500 mg *Rosa* powder twice a day was more efficient than placebo in decreasing WOMAC pain after 3 weeks of treatment but not after 3 months [202]. At 3 months, the *Rosa* treatment decreased the WOMAC function and stiffness subscales (secondary endpoints) more efficiently than placebo. No major side effects were reported in the three RCTs. Hence, further research on this extract is required before making any conclusion about its efficacy or lack of efficacy.

### 9.3 Harpagophytum procumbens (Devil's Claw)

*H. procumbens,* also called devil's claw, is a South African plant that grows in regions bordering the Kalahari. Secondary tuberous roots are used to prepare powders or extracts, which were tested in several RCTs (see [203, 204] for reviews). Product standardization is based on the content of harpagoside, the principal compound found in the raw material. In all RCTs, the harpagoside content was similar and above 50 mg/day.

Recent in vitro studies indicate that preparations from *H. procumbens* may interact with the inflammatory cascade, including the inhibition of cytokines, COX and ROS production, through the suppression of NF-κB activation [205, 206]. An extract of the secondary root of *H. procumbens* reduced the IL-1β-induced production of MMP-1, MMP-3, and MMP-9 proteins by human chondrocytes [207].

A systematic review of the efficacy of *Harpagophytum* for OA concluded that evidence of the effectiveness of *Harpagophytum* products is not transferable from product to product. There is limited evidence of the efficacy for ethanolic extract providing less than 30 mg of harpagoside, one of the major compounds of *H. procumbens,* per day in the treatment of knee or hip OA and moderate evidence of efficacy for the use of powder providing 60 mg of harpagoside daily in the treatment of spine, hip, and knee OA [208]. In a 2-month double-blind RCT on spine and knee OA, 670 mg of *Harpagophytum* powder 3 times a day were more efficient than placebo in reducing VAS of pain [209]. In a 4-month double-blind diacerhein-controlled RCT on hip and knee OA with flare-up episodes at inclusion, 2.6 g of *Harpagophytum* powder per day were no less efficient than 200 mg of diacerhein per day in improving VAS of pain [210, 211]. No safety concerns appeared

from the 4300 patients that received *Harpagophytum* products. In a uncontrolled observational study, 0.8 g of an aqueous extract 3 times a day reduced blood sedimentation time and C-reactive protein levels [212].

### 9.4 Uncaria tomentosa and Uncaria guianensis (Cat's Claw)

Cat's claw is a vine from the basin of the Amazon River. There are two species, *Uncaria tomentosa* and *Uncaria guianensis*, that are traditionally used in South America for their anti-inflammatory properties. The bark and the root are prepared as an extract in hot water. Vincaria, a standardized alkaloid-depleted aqueous extract of *U. guianensis*, administrated in combination with SierraSil, a mineral-rich clay product, inhibited both basal and IL-1β-induced $^{\bullet}$NO production and glycosaminoglycans release from human cartilage explants [213].

In a small 4-week double-blind RCT on knee OA, 100 mg of Vincaria once a day was more effective than placebo in reducing VAS of pain associated with activity but not with at-rest pain or at night [214]. There was no more adverse effect in the cat's claw group than in placebo, although cat's claw has been reported as nephrotoxic [215].

### 9.5 *Salix sp* (Willow Bark)

The anti-inflammatory, antipyretic, and analgesic effects of willow bark are known since antiquity. Willow bark contains salicin derivatives, which are rapidly metabolized in vivo to salicylic acid. The acetylated derivative of salicylic acid is known as aspirin. Willow bark extracts are usually standardized based on salicin content even if salicin might not be the major active compound. Indeed, a pharmacokinetic trial demonstrated that serum salicylate concentrations in human volunteers after oral administration of current therapeutic doses of willow bark extracts (corresponding to 240 mg salicin/day) were too low to explain a relevant analgesic effect [216], indicating that probably other active compounds are present in the extract. In fact, willow bark extract has been demonstrated to inhibit the COX-2-mediated release of $PGE_2$ as well as the release of TNF-α and IL-1β by primary human monocytes, through compounds other than salicin or salicylate [206].

In a 2-week double-blind RCT on knee and hip OA, an amount of willow bark extract (*Salix purpurea x daphnoides*) corresponding to 240 mg of salicin a day was more effective than placebo in reducing the WOMAC pain subscore but the effect size was small [216]. Another 6-week double-blind RCT comparing the effects of another willow bark extract (*S. daphnoides*) at the same dose with placebo and diclofenac 50 mg twice a day on knee and hip OA confirmed the efficacy of diclofenac but failed to detect any significant difference between willow bark and placebo on the WOMAC pain subscore [217]. In another 2-month RCT, a cocktail of 5 plants, which included 100 mg of willow bark, failed to conclusively demonstrate an analgesic effect [218]. Skin allergic reaction have been linked to *Salix* ingestion in 3–11% of the patients [219].

### 9.6 Ginger and Tumeric

The Zingiberaceae family includes ginger and tumerics. Ginger is a very popular spice with a world production of 100,000 tons a year. It is used in traditional Japanese Kampo, Ayurvedic, and Chinese medicine as an anti-inflammatory agent for musculoskeletal diseases.

In vitro, ginger extract decreased the IL-1β and LPS-induced production of $^{\bullet}$NO and $PGE_2$ by OA cartilage [220]. In synoviocyte, it decreased the IL-1β and TNF-α-induced expression of TNF-α mRNA and protein, the TNF-α-induced production of COX-2 as well as the TNF-α-induced activation of NF-κB by reducing the protein level of the NF-κB inhibitor IκB [221].

Three RCTs evaluated ginger extracts prepared from the rhizomes of *Zingiber officinale* and *Alpinia galangal*. A 3-month 3-way cross-over double-blind RCT compared the efficacy of Chinese ginger extract (EV.ext-33; 170 mg) to ibuprofen (400 mg) and placebo in knee and hip OA but failed to demonstrate a difference between placebo and ginger extract as assessed by VAS of pain and Lequesne functional index [222]. In a 6-week double-blind RCT on knee OA, 225 mg of an extract of *Z. officinale* and *A. galangal* twice a day was more effective than placebo in reducing pain on standing based on the percentage of responders [223]. Finally, in a double-blind RCT on knee OA, after cross-over at 6 months, but not at 3 months before cross-over, 250 mg of an extract of *Z. officinale* 4 times a day reduced more effectively VAS of pain and handicap than placebo [224]. In two RCTs, a higher number of adverse effects and higher drop-out rates related to gastrointestinal adverse effects (heartburn) in ginger groups have been observed [223, 224], questioning the safety of these extracts.

Curcumin (diferuloylmethane), a dietary pigment of the root of *Curcuma longa*, is a potent inhibitor (~15 μ*M*) of JNK, c-Jun, NF-κB, and JAK/STAT activation by inflammatory cytokines (oncostatin M, IL-1β, TNF-α) in primary human and bovine chondrocytes in monolayer [225, 226]. In this way, curcumin (15 μ*M*) suppressed IL-1β, OSM and TNF-α-induced MMP-3 and MMP-13 synthesis [226]. Furthermore, curcumin (50 μ*M*) counteracted the IL-1-induced inhibition of type II collagen and activation of caspase-3 synthesis in a human chondrocytes monolayer culture [227, 228].

Finally, curcumin inhibited $PGE_2$ synthesis, by down-regulating COX-2 mRNA and protein levels [229]. It was also demonstrated that, in vitro, curcumin acts synergically with celecoxib, a COX-2 activity inhibitor, to inhibit $PGE_2$ synthesis, suggesting that curcumin may enable the celecoxib at lower and safer concentrations and may pave the way for a novel combination therapy for arthritis. Its anti-inflammatory efficacy has been demonstrated in several animal models. In a rat model of adjuvant-induced arthritis, curcumin notably lowered paw inflammation in arthritic rats when administrated orally (30 mg/kg) for a period of 15 days by gavage [230].

Altogether, these preclinical results indicate that curcumin may prevent inflammation and protect cartilage from cytokine-induced cartilage matrix degradation and then might have a therapeutic potential for treating OA and related osteoarticular disorders. However, these promising preclinical results have been observed at doses 5–25 times superior to the peak concentration (~1.75 μ*M*) found in serum after oral intake of 8 g/day [231]. Curcumin is not toxic even at significantly higher doses (up to 8 g a day) [231], suggesting a possible long-term administration of this active compound.

### 9.7 Flavonoids

Flavonoids are polyphenolic compounds that are ubiquitous in nature and are categorized according to chemical structure into flavonols (quercetin, kaempferol, myricetin, Isorhamnetin), flavones (luteolin, apigenin), flavanones (hesperetin, naringenin, eriodictyol), isoflavones (ipriflavone, genistein, daidzein, glycitein, etc.), anthocyanidins (cyanidin, delphinidin, malvidin, etc.), and chalcones [(+)-catechin, (−)-epicatechin, (−)-epicatechin 3-gallate, etc.). Over 4000 flavonoids have been identified, many of which occur in fruits, vegetables, and beverages (tea, coffee, beer, wine, and fruit drinks). The flavonoids have aroused considerable interest recently because of their potential beneficial effects on human health. They were reported to exhibit estrogenic or anti-estrogenic effects, antiviral, antimicrobial, spasmolytic, antiallergic, antiplatelet [232], anti-inflammatory [233], antitumor [234], and antioxidant activities [235].

Flavonoids were shown to be potent inhibitors of several enzymes, including 5-lipoxygenase and cyclo-oxygenases [236], cyclic adenosine monophosphate (cAMP) phosphodiesterase [237], hyaluronidase [238], tyrosine protein kinase [239], deoxyribonucleic acid (DNA) topoisomerases I and II [240], and viral reverse transcriptase [241]. Our knowledge on the effects of flavonoids on chondrocyte metabolism is limited to the effect of ipriflavone. In a three-dimensional human chondrocyte culture model, 1–100 μg/mL ipriflavone did not affect chondrocyte proliferation but significantly increased proteoglycan and type II collagen synthesis [242].

Regarding isoflavones, a 3-year treatment with 129 mg/day soy phytoestrogen (SUPRO 670-HG soy protein isolate, containing 1.105 mg of genistein, 0.365 mg of daidzein, and 0.08 mg of glycitein per gm of soy protein isolate) did not significantly modify OA lesion severity and IGFBP, collagen, and proteoglycans cartilage contents in ovariectomized cynomolgus monkeys [243, 244]. A 3-month double-blind RCT on knee OA failed to show that 40 g daily of soy protein, containing a total of 88 mg of soy ipriflavones, was more effective than a milk-based protein placebo in reducing symptoms as assessed by questionnaires on pain and quality of life [245]. Soy, but not milk, protein increased serum levels of IGF-1, an anabolic factor for chondrocytes.

Catechins are found predominantly in green tea (*Camellia sinensis*) leaves [246–249]. The most abundant catechin in green tea is epigallocatechin-3-gallate (EGCG) with others such as (−)-epicatechin (EC), (−)-epigallocatechin (EGC), and (−)epicatechin gallate (ECG). Anti-inflammatory and antimitotic properties have been attributed to these compounds [250–252].

EGCG also inhibits IL-1β, TNF-α-induced cartilage matrix degradation [253] via the inhibition of IL-1-activated JNK and AP-1 pathways in human chondrocytes [254, 255], and, in consequence, the expression of MMP-1, MMP-13, but also via the inhibition of aggrecan-degrading activity ADAMTS-1, -4, and -5 [256, 257]. The inhibitory effects of EGCG on MMP expression was observed at concentrations (20–100 μ*M*) that were beyond the plasma concentrations that can be attained simply by drinking green tea (0.1–5 μ*M*) [258, 259], whereas the inhibition of ADAMTS activity was effective at concentrations (0.1–1 μ*M*) below the plasmatic concentration. Beside these anticatabolic effects, EGCG also inhibits the IL-1β-induced activity and expression of COX-2 and NOS-2 in human chondrocytes, indicating that this compound may have anti-inflammatory effects [254, 260]. EGCG also reduced the IL-1β inhibitory effect on type II collagen, aggrecan core protein, and TGF-β receptor II (TGF-bRII) mRNA levels [261].

However, until now, no study has investigated the efficacy of green tea polyphenols on OA symptoms and structural changes in human and animal studies. One animal study on collagen-induced arthritis in mice has reported that mice receiving green tea polyphenolic fraction in the drinking water before collagen injection and during the course of the study (up to 85 days) were significantly less susceptible to the development of arthritis, and, if the mice developed arthritis, the disease was late in onset and mild in comparison to mice not given green tea polyphenolic fraction in drinking water. The arthritic joint had significantly lower concentrations

of inflammatory cytokines (IFNγ, TNF-α) and other mediators of inflammation such as COX-2 [262].

### 9.8 *Boswellia serrata*

The gummy oleoresin from the bark of *Boswellia serrata*, a tree from northwest India, is used to treat inflammatory diseases in Ayurvedic medicine. This compound is known as a potent inhibitor of 5-lipoxygenase, but its effects on chondrocytes remain unknown. The clinical efficacy of a herbomineral formulation containing roots of *Withania somnifera*, the stem of *B. serrata*, rhizomes of *Curcuma longa*, and a zinc complex (Articulin-F) was evaluated in a 6-month double-blind cross-over RCT in 42 patients with OA. A cocktail of 650 mg twice a day, called Articulin-F, was more effective than placebo in reducing pain and disability scores. Radiological assessment, however, did not show any significant changes in both groups. Side effects observed with this formulation did not necessitate withdrawal of treatment [263]. In an 8-week double-blind cross-over RCT including 30 patients with knee OA, 333 mg of the gum 3 times a day was more effective than placebo in reducing knee pain, loss of movement, and swelling score and in increasing walking distance [264].

### 9.9 Cocktails of Plant Extracts

SKI306X is an herbal mixture prepared from three plants (dried roots from *Clematis mandshurica*, *Trichosantes kirilowii*, and *Prunella vulgaris*) used for the treatment of inflammatory diseases in Far East Asia. It is clinically approved for the treatment of OA in Korea.

In vitro, SKI306X and *T. kirilowii* extract, but not the other two plant extracts, reduced the IL-1α-induced GAG release from rabbit articular cartilage explant [265]. Synergistic effects between the three extracts present in SKI306X were suspected since the proportion of *T. kirilowii* was insufficient to fully explain the product potency.

In rats, SKI306X did not cause significant gastric damage up to an oral dose of 2 g/kg and suppressed the diclofenac-induced gastric damage [266]. In rabbits, 200 mg/kg/day reduced the OA-like histological changes in collagenase-injected knees [265].

In a 4-week double-blind RCT on knee OA, 200, 400, and 600 mg 3 times a day were more effective than placebo in reducing VAS of pain, with no significant difference between the three doses [267]. In another 4-week double-blind RCT on knee OA, 200 mg three times a day was not less effective than 100 mg of diclofenac sustained release once a day in reducing VAS of pain [268]. During the study, there was no statistically significant difference between the two treatment groups in the incidence of any adverse effects. However, the number of patients who had adverse events, particularly heartburn, during the study was slightly greater in the diclofenac group than in the SKI306X group.

The effects of Gitadyl, a herbal formulation containing extracts from feverfew, American aspen, and milfoil, at a dose of 260 mg three times a day were compared to the effects of 400 mg of ibuprofen three times a day in a 42-day double-blind cross-over RCT. Both treatments failed to significantly change the pain and patients' ability to walk as assessed by four-point scales [269].

In a 4-week double-lind RCT on knee OA, 3 g, 3 times a day, of Duhuo Jisheng wan, a traditional Chinese cocktail of 15 plants improved the Lequesne functional index and multiple VAS of pain and stiffness as effectively as 25 mg of diclofenac, three times a day [270]. Duhuo Jisheng wan had a slower onset of action than diclofenac but an equal rate of adverse events, an observation that questions its safety.

## 10 METHYLSULFONYLMETHANE

Methylsulfonylmethane (MSM) is the first oxidized metabolite of dimethylsulfoxide (DMSO) and a putative methyl donor. DMSO is found in fruit, corn, tomatoes, tea, coffee, and milk. MSM has been isolated from plants such as *Equisetum arvense*, also known as horsetail. In commercial production, MSM is synthesized by reacting DMSO and hydrogen peroxide, which yields MSM and water. MSM is currently sold in over 52 different products as a single agent in capsule, caplet, lotion, and cream forms and in more than 30 different products in combination with other dietary supplements, mainly GS and CS. There is little information on the mechanism of action of MSM. MSM has been demonstrated to be a strong scavenger in numerous free-radical-generating systems. Clinical evidence for the usefulness of MSM in OA is limited to two short pilot studies. In a 12-week double-blind placebo-controlled RCT on knee OA, 500 mg of MSM, 3 times a day, combined or not with 500 mg of glucosamine-HCL, 3 times a day, significantly improved symptoms (VAS of pain and Lequesne functional index) [271]. In this study, both glucosamine-HCL and MSM improved OA symptoms, but when taken together the effect was greater than either alone. This beneficial effect has been recently confirmed in a double-blinded RCT including 50 patients with knee OA [272]. Compared to placebo, MSM (3 g twice a day for 12 weeks) produced modest, but significant, decreases in WOMAC pain and physical function impairment, while no changes were found in WOMAC stiffness and aggregated total symptoms

scores. MSM also produced improvement in performing activities of daily living when compared to placebo on the SF-36 evaluation. No evidence of a more general anti-inflammatory effect was found, as there were no significant changes in two measures of systemic inflammation: C-reactive protein level and erythrocyte sedimentation rate. The authors cautioned that this short pilot study did not address the long-term safety and usefulness of MSM but suggested that physicians should consider its use for certain OA patients, and that long-term studies should be conducted [272]. MSM is generally considered safe, even though there are no clinical studies on adverse effects. Nevertheless, in the two published RCT, no severe adverse effects have been reported.

## 11 MISCELLANEOUS

### 11.1 Biphosphonates and Strontium Ranelate

The rationale to use biphosphonates to treat OA was provided by the recent findings that subchondral bone remodeling plays a role in the progression of OA pathogenesis. Biphosphonates, mainly the nitrogen-containing biphosphonate, inhibit bone resorption by increasing osteoclast apoptosis [273] and by inhibiting their resorptive function (for more details see [274]). Furthermore, biphophonates inhibit calcification by binding to the surface of solid calcium phosphate crystals and acting as crystal poisons affecting both crystal growth and dissolution [275]. These findings suggest that biphosphonates could be beneficial in OA treatment, by restoring subchondral bone homeostasis and by preventing cartilage mineralization. In addition, biphosphonates induce macrophage apoptosis, suggesting that they could have an anti-inflammatory effect. In animal models, encapsulated chlodronate was found to reduce the number of macrophages and to reduce inflammation [276]. When a single intra-articular injection of encapsulated chlodronate was given to patients with RA, a depletion of synovial macrophages was observed [277].

Numerous animal studies of arthritis have suggested that biphosphonates may have chondroprotective abilities [278–281]. The most recent study [280] showed a reduction of cartilage degeneration, as determined by histologic criteria, and prevention of osteophyte formation by alendronate in a rat anterior cruciate ligament transection model. Alendronate suppressed subchondral bone resorption, inhibited vascular invasion into the calcified cartilage, and blocked osteoclast recruitment to subchondral bone and osteophytes. Two recent studies have demonstrated that treatment with biphosphonates (alendronate and zoledronate) decreased the urinary concentration of CTX-II, a marker of cartilage degradation, suggesting that biphosphonates may have chondroprotective effects in human [282, 283]. However, these studies suffer from some limitations, including short follow-up period, the lack of arthritis assessment procedures, or the absence of correlation between CTX-II levels and cartilage structural changes. More recently, a 1-year prospective, double-blind, placebo-controlled study, including mild to moderate OA of the medial compartment of the knee, has been designed to determine the efficacy of risedronate (5–15 mg/day) in structure and symptoms of knee OA [284]. The study showed that patients receiving 15 mg of risedronate had an improvement of the WOMAC index, particularly of physical function, significantly improved the total patient assessment, and decreased use of walking aids relative to patients receiving a placebo. While the differences in joint space width among the groups were not significant, there was a trend for less loss in the risedronate (15 mg) than in the placebo group. When data were analyzed in a post hoc manner to identify patients with detectable progression (i.e., loss of JSW ≥25% or ≥0.75 mm), 8% of the patients receiving placebo and 1% of patients receiving 15 mg risedronate were found to exhibit this degree of progression. Additionally, risedronate significantly reduced levels of CTX-II, but this was not associated with an attenuation of joint space loss.

Strontium ranelate (S12911), an uncoupling agent developed as a preventive and curative treatment of postmenopausal osteoporosis [285], prevents bone loss induced by estrogen deficiency by inhibiting bone resorption and stimulating bone formation [286]. In vitro, strontium ranelate increases collagen and noncollagen protein synthesis by mature osteoblast-enriched cells. Its effects on bone formation were confirmed as the drug-enhanced preosteoblastic cell replication. In the isolated osteoclasts, preincubation of bone slices with strontium ranelate induced dose-dependent inhibition of the bone resorbing activity of rat osteoclasts [287]. Strontium ranelate dose-dependently inhibited preosteoclast differentiation. Recently, we have studied the effects of strontium ranelate, calcium ranelate, and strontium chloride on the metabolism of primary human OA chondrocytes in a suspension culture [288]. We have observed that strontium ranelate, calcium ranelate, and $SrCl_2$ did not modify stromelysin synthesis. Calcium ranelate induced stromelysin activation, whereas strontium compounds were ineffective. Strontium ranelate, $SrCl_2$, and calcium ranelate strongly stimulated aggrecan production suggesting an ionic effect of strontium independent of the organic moiety. Morever, strontium ranelate increased the stimulatory effect of insulin-like growth factor (IGF)-1 on aggrecan synthesis but did not reverse the inhibitory effect of IL-1β. In summary,

strontium ranelate strongly stimulates human cartilage matrix formation in vitro by a direct ionic effect without stimulating the chondroresorption processes. This finding provides a preclinical basis for in vivo testing of strontium ranelate.

### 11.2 Diacerhein

Diacerhein is an anthraquinone that is entirely transformed by deacetylation during absorption into rhein, the active metabolite found in plasma and synovial fluid [289, 290]. In vitro, rhein inhibits the production of MMPs and stimulates the production of aggrecan by human OA chondrocytes [291, 292]. It also counteracts the effects of IL-1β on MMPs and aggrecan. These positive effects of rhein on cartilage matrix formation are explained by its stimulating effect on TGF-β1 and -β2 gene expression and by its inhibiting action on IL-1β secretion [293, 294]. Furthermore, rhein was demonstrated to downregulate $^{\bullet}$NO production by human OA chondrocytes in alginate beads [291], in monolayer [295], and in cartilage explants [296]. The effect of rhein appeared to result from iNOS transcriptional and/or posttranscriptional events, as indicated by a decrease in the levels of mRNA and protein for this enzyme in human OA chondrocytes [296]. Rhein was also demonstrated to decrease arthritis activity and $NO_2^-$ level in the plasma of rats with adjuvant-induced arthritis. These data suggest that the inhibitory effect of diacerein on rat-adjuvant-induced arthritis is partly related to its reduction of the $^{\bullet}$NO production induced by adjuvant-induced arthritis [292]. However, in some in vitro systems, rhein may also exhibit a pro-oxidant effect that could be responsible for its cellular toxicity [297, 298]. This problem was underestimated in some in vitro studies reporting the effect of rhein on chondrocytes. At concentrations over $10^{-5}$ *M*, rhein induced cytotoxicity in human chondrocyte [291]. One possible mechanism underlying this toxic effect is that rhein undergoes redox cycling that gives rise to oxygen metabolites that initiate peroxidation of cellular membrane [297].

In well-known and validated animal models of OA induced by vitamin A [299], postcontusion and postpatellectomy OA [300], diacerhein was demonstrated to prevent the development of OA. In these models, diacerhein was shown to be active in reducing macroscopic and microscopic lesions, in preventive and therapeutic ways. However, the exact mechanisms of the therapeutic action of diacerhein on OA are not clear. In an accelerated canine model of OA in which sensory input from the ipsilateral hind limb was interrupted by L4-S1 dorsal root ganglionectomy prior to cruciate ligament (ACL) transaction, diacerhein (15–20 mg/kg once daily for 8 weeks) did not significantly modify the severity of cartilage lesions, the collagenolytic activity and the proteoglycan level contained in OA cartilage [301]. In a less rapidly progressive canine model, in which OA was induced by ACL only, diacerhein (40 mg/kg for 8 months) has been shown to slow down the progression of cartilage lesions as estimated by an arthroscopic grading scale [301]. However, 32 weeks after ACL transection, the mean proteoglycan concentration of the OA cartilage and the level of collagenolytic activity in cartilage extracts were not significantly different in the diacerhein group compared to placebo. In an ovine model of OA induced by menisectomy, diacerhein (20–50 mg/kg daily for 9 months) failed to significantly improve the gross morphological and histological severity of cartilage lesions [302].

In several randomized, double-blind, placebo-controlled clinical trials of 2–6 months duration, diacerhein significantly reduced the pain and functional impairment in patients with hip or knee OA [299, 303, 304]. Recently, a 3-year, placebo-controlled trial (ECHODIAH) including 507 patients with primary OA of hip has demonstrated that the occurrence of radiographic progression (i.e., joint space narrowing measured by a graduated magnifying lens) of at least 0.5 mm during the study was significantly lower and occurred later in the diacerhein (100 mg daily) as compared with the placebo group [305].

### 11.3 Doxycycline

Doxycycline is a member of the tetracycline family of broad-spectrum antibiotics. During the last decade, it has been established that tetracyclines, which are rapidly absorbed and have a prolonged half-life, exert biological effects independent of their antimicrobial activity, which include inhibition of MMPs [306], NOS expression [307], tumor progression [308], bone resorption [309], angiogenesis [310], and inflammation [311]. Among these effects, some could have a beneficial effect on OA pathophysiology. In human OA cartilage, doxycycline (10–40 μg/mL) inhibited $^{\bullet}$NO production and augmented COX-2 and $PGE_2$ production (one- to two-fold) in the absence or in the presence of IL-1β and TNF-α or LPS. Thus, the effects of docycycline on $^{\bullet}$NO and $PGE_2$ production were independent of each other [307, 312, 313]. Doxycycline (1–10 μg/mL) also decreased levels of mRNA and protein of MMP-1, MMP-8, and MMP-13 in TNF-α-stimulated human normal and OA chondrocytes [314]. Beside this anticatabolic effect, doxycycline ($IC_{50} = 25\,\mu M$) inhibited collagens, including type II collagen, and the synthesis by bovine and human chondrocytes [315]. Beekman et al. also reported that the concentration of doxycycline needed to down-

regulate collagen synthesis was >10-fold lower than that needed to inhibit most of MMPs.

In dogs, that underwent L4-S1 dorsal root ganglionectomy followed by ipsilateral anterior cruciate ligament transection, a procedure that markedly accelerates development of OA in the unstable joint, prophylactic and therapeutic oral administration of doxycycline (roughly 3.5 mg/kg/day) significantly retarded the appearance of cartilage lesions [316] via inhibition of MMPs [317] and probably through an effect on subchondral bone [318]. Recently, it was demonstrated that 36 weeks of treatment with doxycycline (4 mg/kg/day) had beneficial effects on structural and mechanical changes observed on the superficial (0–3 mm) but not the deeper zones (4–5 mm) of the subchondral bone in late OA induced surgically by ACLT [319].

Five days of doxycycline treatment, in a dose of 100 mg bid, reduced collagenase and gelatinase activities in human OA cartilage ex vivo, while a single 200-mg dose, given 3 days before taking cartilage sample, was ineffective in inhibiting MMP activity [320]. Brandt and his colleagues [321] have shown that treatment with oral doxycycline can slow the rate of joint space narrowing in the index knee OA of obese women over a 30-month period. There were no significant effects on pain, but there was a small group difference in other clinical outcomes that favored those treated with doxycycline over those treated with placebo [321].

## 12 CONCLUDING REMARKS

Osteoarthritis is a disease with a multifactorial etiology affecting all joint tissues and involving both biochemical and mechanical factors that act in synergy to degrade cartilage. Until now, OA treatments are limited to pain relief, even if some new drugs of the SYSADOA class have shown potential structure-modifying effects. These drugs target cartilage as the endpoint of joint failure, underestimating the role played by subchondral bone and synovium in this process. However, recent studies emphasized the key role played by subchondral bone sclerosis and synovium inflammation in OA pathophysiology. Unfortunately, until now, there is no validated method for a reliable assessment of structural changes in these tissues, explaining that guidance documents edited by the official agencies and scientific societies are strictly focused on the assessment of cartilage changes. Actually, pharmacological management of OA is dominated by analgesic, NSAIDS, and SYSADOA, which constitute a new class of drugs characterized by slow action on OA symptoms and good tolerability. Some SYSADOAs are nutraceuticals (e.g., glucosamine or chondroitin sulfate) and are provided as dietary supplements or prescribed drugs according to the country and the product specifications. The absence of a clear definition and a requirement for registration of nutraceuticals throws the scientists and practitioners into confusion. Often data coming from preclinical and clinical studies of prescribed drugs are extrapolated to over-the-counter products that differ considerably in source, formula, purity, rhythm, and dose administrated. Actually, more than 30 nutraceutical products are marketed as potentially active in OA but few have demonstrated their efficacy and safety in human clinical trials. Clearly, in terms of regulations, there is a need to resolve the general confusion about whether nutraceuticals are food supplements or medicines.

## REFERENCES


1. Yelin E (2003). The economics of osteoarthritis. In *Osteoarthritis*. Brandt KD, Doherty M, Lohmander LS, Eds. Oxford: Oxford University Press, pp. 17–21.
2. Zhang W, Doherty M, Arden N, Bannwarth B, Bijlsma J, Gunther KP, Hauselmann HJ, Herrero-Beaumont G, Jordan K, Kaklamanis P, Leeb B, Lequesne M, Lohmander S, Mazieres B, Martin-Mola E, Pavelka K, Pendleton A, Punzi L, Swoboda B, Varatojo R, Verbruggen G, Zimmermann-Gorska I, Dougados M (2005). EULAR evidence based recommendations for the management of hip osteoarthritis: Report of a task force of the EULAR Standing Committee for International Clinical Studies Including Therapeutics (ESCISIT). *Ann Rheum Dis 64*(5), 669–681.
3. Jordan KM, Arden NK, Doherty M, Bannwarth B, Bijlsma JW, Dieppe P, Gunther K, Hauselmann H, Herrero-Beaumont G, Kaklamanis P, Lohmander S, Leeb B, Lequesne M, Mazieres B, Martin-Mola E, Pavelka K, Pendleton A, Punzi L, Serni U, Swoboda B, Verbruggen G, Zimmerman-Gorska I, Dougados M (2003). EULAR Recommendations 2003: An evidence based approach to the management of knee osteoarthritis: Report of a Task Force of the Standing Committee for International Clinical Studies Including Therapeutic Trials (ESCISIT). *Ann Rheum Dis 62*(12), 1145–1155.
4. Sandell LJ, Aigner T (2001). Articular cartilage and changes in arthritis. An introduction: Cell biology of osteoarthritis. *Arthritis Res 3*(2), 107–113.
5. Lee GM, Loeser RF (1998). Interactions of the chondrocyte with its pericellular matrix. *Cells Materials 8*, 135–149.
6. Gray ML, Pizzanelli AM, Grodzinsky AJ, Lee RC (1988). Mechanical and physiochemical determinants of the chondrocyte biosynthetic response. *J Orthop Res 6*(6), 777–792.
7. Sah RL, Kim YJ, Doong JY, Grodzinsky AJ, Plaas AH, Sandy JD (1989). Biosynthetic response of cartilage

explants to dynamic compression. *J Orthop Res 7*(5), 619–636.

8. Veldhuijzen JP, Bourret LA, Rodan GA (1979). In vitro studies of the effect of intermittent compressive forces on cartilage cell proliferation. *J Cell Physiol 98*(2), 299–306.
9. Schneiderman R, Keret D, Maroudas A (1986). Effects of mechanical and osmotic pressure on the rate of glycosaminoglycan synthesis in the human adult femoral head cartilage: An in vitro study. *J Orthop Res 4*(4), 393–408.
10. Martel-Pelletier J, Di Battista J, Lajeunesse D (1999). Biochemical factors in joint articular tissue degradation in osteoarthritis. In Reginster J-Y, Pelletier J-P, Martel-Pelletier J, Henrotin Y, Eds. *Osteoarthritis: Clinical and Experimental Aspects*. Oxford: Oxford University Press, pp. 156–187.
11. van den Berg P, van der Kraan P, van Beuningen H (1999). Role of growth factors and cartilage repair. In Reginster J-Y, Martel-Pelletier J, Pelletier J-P, Henrotin Y, Eds. *Osteoarthritis*. Berlin: Springer, pp. 188–209.
12. Goldring SR, Goldring MB (2004). The role of cytokines in cartilage matrix degeneration in osteoarthritis. *Clin Orthop Relat Res 427*(Suppl), S27–36.
13. van Beuningen HM, van der Kraan PM, Arntz OJ, van den Berg WB (1994). In vivo protection against interleukin-1-induced articular cartilage damage by transforming growth factor-beta 1: Age-related differences. *Ann Rheum Dis 53*(9), 593–600.
14. Glansbeek HL, van der Kraan PM, Lafeber FP, Vitters EL, van den Berg WB (1997). Species-specific expression of type II TGF-beta receptor isoforms by articular chondrocytes: Effect of proteoglycan depletion and aging. *Cytokine 9*(5), 347–351.
15. Braunstein EM, Brandt KD, Albrecht M (1990). MRI demonstration of hypertrophic articular cartilage repair in osteoarthritis. *Skeletal Radiol 19*(5), 335–339.
16. Kusuzaki K, Sugimoto S, Takeshita H, Murata H, Hashiguchi S, Nozaki T, Emoto K, Ashihara T, Hirasawa Y (2001). DNA cytofluorometric analysis of chondrocytes in human articular cartilages under normal aging or arthritic conditions. *Osteoarthritis Cartilage 9*(7), 664–670.
17. Buckwalter JA (2002). Articular cartilage injuries. *Clin Orthop Relat Res 402*, 21–37.
18. Buckwalter JA, Mankin HJ (1997). Articular cartilage II. Degeneration and osteoarthrosis, repair, regeneration and transplantation. *J Bone Joint Surg 79A*(4), 612–632.
19. Poole AR, Kojima T, Yasuda T, Mwale F, Kobayashi M, Laverty S (2001). Composition and structure of articular cartilage: A template for tissue repair. *Clin Orthop Relat Res 391*(Suppl), S26–33.
20. Aigner T, Zien A, Gehrsitz A, Gebhard PM, McKenna L (2001). Anabolic and catabolic gene expression pattern analysis in normal versus osteoarthritic cartilage using complementary DNA-array technology. *Arthritis Rheum 44*(12), 2777–2789.
21. Merz D, Liu R, Johnson K, Terkeltaub R (2003). IL-8/CXCL8 and growth-related oncogene alpha/CXCL1 induce chondrocyte hypertrophic differentiation. *J Immunol 171*(8), 4406–4415.
22. Johnson KA, van Etten D, Nanda N, Graham RM, Terkeltaub RA (2003). Distinct transglutaminase 2-independent and transglutaminase 2-dependent pathways mediate articular chondrocyte hypertrophy. *J Biol Chem 278*(21), 18824–18832.
23. Kirsch T, Swoboda B, Nah H (2000). Activation of annexin II and V expression, terminal differentiation, mineralization and apoptosis in human osteoarthritic cartilage. *Osteoarthritis Cartilage 8*(4), 294–302.
24. Attur MG, Dave MN, Leung MY, Cipolletta C, Meseck M, Woo SL, Amin AR (2002). Functional genomic analysis of type II IL-1beta decoy receptor: Potential for gene therapy in human arthritis and inflammation. *J Immunol 168*(4), 2001–2010.
25. Boumediene K, Conrozier T, Mathieu P, Richard M, Marcelli C, Vignon E, Pujol JP (1998). Decrease of cartilage transforming growth factor-beta receptor II expression in the rabbit experimental osteoarthritis—potential role in cartilage breakdown. *Osteoarthritis Cartilage 6*(2), 146–149.
26. Henrotin Y, Kurz B, Aigner T (2005). Oxygen and reactive oxygen species in cartilage degradation: Friends or foes? *Osteoarthritis Cartilage 13*(8), 643–654.
27. Henrotin YE, Bruckner P, Pujol JP (2003). The role of reactive oxygen species in homeostasis and degradation of cartilage. *Osteoarthritis Cartilage 11*(10), 747–755.
28. Naito K, Takahashi M, Kushida K, Suzuki M, Ohishi T, Miura M, Inoue T, Nagano A (1999). Measurement of matrix metalloproteinases (MMPs) and tissue inhibitor of metalloproteinases-1 (TIMP-1) in patients with knee osteoarthritis: Comparison with generalized osteoarthritis. *Rheumatol (Oxford) 38*(6), 510–515.
29. Tetlow LC, Adlam DJ, Woolley DE (2001). Matrix metalloproteinase and proinflammatory cytokine production by chondrocytes of human osteoarthritic cartilage: Associations with degenerative changes. *Arthritis Rheum 44*(3), 585–594.
30. Tetlow LC, Woolley DE (2001). Expression of vitamin D receptors and matrix metalloproteinases in osteoarthritic cartilage and human articular chondrocytes in vitro. *Osteoarthritis Cartilage 9*(5), 423–431.
31. Bluteau G, Conrozier T, Mathieu P, Vignon E, Herbage D, Mallein-Gerin F (2001). Matrix metalloproteinase-1, -3, -13 and aggrecanase-1 and -2 are differentially expressed in experimental osteoarthritis. *Biochim Biophys Acta 1526*(2), 147–158.
32. Milner JM, Elliott SF, Cawston TE (2001). Activation of procollagenases is a key control point in cartilage collagen degradation: Interaction of serine and metalloproteinase pathways. *Arthritis Rheum 44*(9), 2084–2096.
33. Schwab W, Gavlik JM, Beichler T, Funk RH, Albrecht S, Magdolen V, Luther T, Kasper M, Shakibaei M (2001). Expression of the urokinase-type plasminogen activator

receptor in human articular chondrocytes: Association with caveolin and beta 1-integrin. *Histochem Cell Biol 115*(4), 317–323.

34. Lang A, Horler D, Baici A (2000). The relative importance of cysteine peptidases in osteoarthritis. *J Rheumatol 27*(8), 1970–1979.
35. Treadwell BV, Pavia M, Towle CA, Cooley VJ, Mankin HJ (1991). Cartilage synthesizes the serine protease inhibitor PAI-1: Support for the involvement of serine proteases in cartilage remodeling. *J Orthop Res 9*(3), 309–316.
36. Tchetina EV, Squires G, Poole AR (2005). Increased type II collagen degradation and very early focal cartilage degeneration is associated with upregulation of chondrocyte differentiation related genes in early human articular cartilage lesions. *J Rheumatol 32*(5), 876–886.
37. Duerr S, Stremme S, Soeder S, Bau B, Aigner T (2004). MMP-2/gelatinase A is a gene product of human adult articular chondrocytes and is increased in osteoarthritic cartilage. *Clin Exp Rheumatol 22*(5), 603–608.
38. Mitchell PG, Magna HA, Reeves LM, Lopresti-Morrow LL, Yocum SA, Rosner PJ, Geoghegan KF, Hambor JE (1996). Cloning, expression, and type II collagenolytic activity of matrix metalloproteinase-13 from human osteoarthritic cartilage. *J Clin Invest 97*(3), 761–768.
39. Reboul P, Pelletier JP, Tardif G, Cloutier JM, Martel-Pelletier J (1996). The new collagenase, collagenase-3, is expressed and synthesized by human chondrocytes but not by synoviocytes. A role in osteoarthritis. *J Clin Invest 97*(9), 2011–2019.
40. Bau B, Gebhard PM, Haag J, Knorr T, Bartnik E, Aigner T (2002). Relative messenger RNA expression profiling of collagenases and aggrecanases in human articular chondrocytes in vivo and in vitro. *Arthritis Rheum 46*(10), 2648–2657.
41. Sandy JD, Flannery CR, Neame PJ, Lohmander LS (1992). The structure of aggrecan fragments in human synovial fluid. Evidence for the involvement in osteoarthritis of a novel proteinase which cleaves the Glu 373-Ala 374 bond of the interglobular domain. *J Clin Invest 89*(5), 1512–1516.
42. Lee JH, Fitzgerald JB, Dimicco MA, Grodzinsky AJ (2005). Mechanical injury of cartilage explants causes specific time-dependent changes in chondrocyte gene expression. *Arthritis Rheum 52*(8), 2386–2395.
43. Fermor B, Weinberg JB, Pisetsky DS, Misukonis MA, Banes AJ, Guilak F (2001). The effects of static and intermittent compression on nitric oxide production in articular cartilage explants. *J Orthop Res 19*(4), 729–737.
44. Henrotin Y, Deby-Dupont G, Deby C, De Bruyn M, Lamy M, Franchimont P (1993). Production of active oxygen species by isolated human chondrocytes. *Br J Rheumatol 32*(7), 562–567.
45. Hiran TS, Moulton PJ, Hancock JT (1997). Detection of superoxide and NADPH oxidase in porcine articular chondrocytes. *Free Radic Biol Med 23*(5), 736–743.
46. Moulton PJ, Goldring MB, Hancock JT (1998). NADPH oxidase of chondrocytes contains an isoform of the gp91phox subunit. *Biochem J 329* (Pt 3), 449–451.
47. Tiku ML, Shah R, Allison GT (2000). Evidence linking chondrocyte lipid peroxidation to cartilage matrix protein degradation. Possible role in cartilage aging and the pathogenesis of osteoarthritis. *J Biol Chem 275*(26), 20069–20076.
48. Loeser RF, Carlson CS, Del Carlo M, Cole A (2002). Detection of nitrotyrosine in aging and osteoarthritic cartilage: Correlation of oxidative damage with the presence of interleukin-1beta and with chondrocyte resistance to insulin-like growth factor 1. *Arthritis Rheum 46*(9), 2349–2357.
49. Deberg M, Labasse A, Christgau S, Cloos P, Bang Henriksen D, Chapelle JP, Zegels B, Reginster JY, Henrotin Y (2005). New serum biochemical markers (Coll 2–1 and Coll 2–1 NO(2)) for studying oxidative-related type II collagen network degradation in patients with osteoarthritis and rheumatoid arthritis. *Osteoarthritis Cartilage 13*(3), 258–265.
50. Burkhardt H, Swingel M, Menninger H, Macartney HW (1986). Oxygen radicals as effectors of cartilage destruction. direct degradative effect on matrix components and indirect action via activation of latent collagenase from polymorphonuclear leukocytes. *Arthritis Rheum 29*, 379–387.
51. Pelletier JP, Jovanovic D, Fernandes JC, Manning P, Connor JR, Currie MG, Di Battista JA, Martel-Pelletier J (1998). Reduced progression of experimental osteoarthritis in vivo by selective inhibition of inducible nitric oxide synthase. *Arthritis Rheum 41*(7), 1275–1286.
52. Pelletier JP, Jovanovic DV, Lascau-Coman V, Fernandes JC, Manning PT, Connor JR, Currie MG, Martel-Pelletier J (2000). Selective inhibition of inducible nitric oxide synthase reduces progression of experimental osteoarthritis in vivo: Possible link with the reduction in chondrocyte apoptosis and caspase 3 level. *Arthritis Rheum 43*(6), 1290–1299.
53. Del Carlo M Jr, Loeser RF (2002). Nitric oxide-mediated chondrocyte cell death requires the generation of additional reactive oxygen species. *Arthritis Rheum 46*(2), 394–403.
54. Blanco FJ, Ochs RL, Schwarz H, Lotz M (1995). Chondrocyte apoptosis induced by nitric oxide. *Am J Pathol 146*(1), 75–85.
55. Burr DB, Radin EL (2003). Microfractures and microcracks in subchondral bone: Are they relevant to osteoarthrosis? *Rheum Dis Clin North Am 29*(4), 675–685.
56. Muir P, McCarthy J, Radtke CL, Markel MD, Santschi EM, Scollay MC, Kalscheur VL (2005). Role of endochondral ossification of articular cartilage and functional adaptation of the subchondral plate in the development of fatigue microcracking of joints. *Bone*.
57. Fenwick SA, Gregg PJ, Rooney P (1999). Osteoarthritic cartilage loses its ability to remain avascular. *Osteoarthritis Cartilage 7*(5), 441–452.

58. Tanaka E, Aoyama J, Miyauchi M, Takata T, Hanaoka K, Iwabe T, Tanne K (2005). Vascular endothelial growth factor plays an important autocrine/paracrine role in the progression of osteoarthritis. *Histochem Cell Biol 123*(3), 275–281.
59. Pfander D, Kortje D, Zimmermann R, Weseloh G, Kirsch T, Gesslein M, Cramer T, Swoboda B (2001). Vascular endothelial growth factor in articular cartilage of healthy and osteoarthritic human knee joints. *Ann Rheum Dis 60*(11), 1070–1073.
60. Hayami T, Funaki H, Yaoeda K, Mitui K, Yamagiwa H, Tokunaga K, Hatano H, Kondo J, Hiraki Y, Yamamoto T, Duong le T, Endo N (2003). Expression of the cartilage derived anti-angiogenic factor chondromodulin-I decreases in the early stage of experimental osteoarthritis. *J Rheumatol 30*(10), 2207–2217.
61. Green WT Jr, Martin GN, Eanes ED, Sokoloff L (1970). Microradiographic study of the calcified layer of articular cartilage. *Arch Pathol 90*(2), 151–158.
62. Benito MJ, Veale DJ, FitzGerald O, van den Berg WB, Bresnihan B (2006). Synovial tissue inflammation in early and late osteoarthritis. *Ann Rheum Dis 64*, 1263–1267.
63. Myers SL, Brandt KD, Ehlich JW, Braunstein EM, Shelbourne KD, Heck DA, Kalasinski LA (1990). Synovial inflammation in patients with early osteoarthritis of the knee. *J Rheumatol 17*(12), 1662–1669.
64. Smith MD, Triantafillou S, Parker A, Youssef PP, Coleman M (1997). Synovial membrane inflammation and cytokine production in patients with early osteoarthritis. *J Rheumatol 24*(2), 365–371.
65. Vergunst CE, van de Sande MG, Lebre MC, Tak PP (2005). The role of chemokines in rheumatoid arthritis and osteoarthritis. *Scand J Rheumatol 34*(6), 415–425.
66. Pelletier JP, Martel-Pelletier J, Abramson SB (2001). Osteoarthritis, an inflammatory disease: Potential implication for the selection of new therapeutic targets. *Arthritis Rheum 44*(6), 1237–1247.
67. Ayral X (1999). [Efficacy and role of topical treatment of gonarthrosis]. *Presse Med 28*(22), 1195–1200.
68. Ayral X, Ravaud P, Bonvarlet JP, Simonnet J, Lecurieux R, Nguyen M, Sauvage E, Dougados M (1999). Arthroscopic evaluation of post-traumatic patellofemoral chondropathy. *J Rheumatol 26*(5), 1140–1147.
69. Ayral X, Mackillop N, Genant HK, Kirkpatrick J, Beaulieu A, Pippingskiold P, Will RK, Alava S, Dougados M (2003). Arthroscopic evaluation of potential structure-modifying drug in osteoarthritis of the knee. A multicenter, randomized, double-blind comparison of tenidap sodium vs piroxicam. *Osteoarthritis Cartilage 11*(3), 198–207.
70. Hill CL, Gale DG, Chaisson CE, Skinner K, Kazis L, Gale ME, Felson DT (2001). Knee effusions, popliteal cysts, and synovial thickening: Association with knee pain in osteoarthritis. *J Rheumatol 28*(6), 1330–1337.
71. D'Agostino MA, Conaghan P, Le Bars M, Baron G, Grassi W, Martin-Mola E, Wakefield R, Brasseur J-L, So A, Backhaus M, Malaise M, Burmester G, Schmidely N, Ravaud P, Dougados M, Emery P (2006). EULAR report on the use of ultrasonographyin painful knee osteoarthritis. Part 1: Prevalence of inflammation in osteoarthritis. *Ann Rheum Dis 64*, 1703–1709.
72. Stümer T, Brenner H, Koenig W, Günther K-P (2006). Severity and extent of osteoarthritis and low grade systemic inflammation as assessed by high sensitivity C reactive protein. *Ann Rheum Dis 63*, 200–205.
73. Carlson CS, Loeser RF, Purser CB, Gardin JF, Jerome CP (1996). Osteoarthritis in cynomolgus macaques. III: Effects of age, gender, and subchondral bone thickness on the severity of disease. *J Bone Miner Res 11*(9), 1209–1217.
74. Carlson CS, Loeser RF, Jayo MJ, Weaver DS, Adams MR, Jerome CP (1994). Osteoarthritis in cynomolgus macaques: A primate model of naturally occurring disease. *J Orthop Res 12*(3), 331–339.
75. Lahm A, Kreuz PC, Oberst M, Haberstroh J, Uhl M, Maier D (2005). Subchondral and trabecular bone remodeling in canine experimental osteoarthritis. *Arch Orthop Trauma Surg* 1–6.
76. Boyd SK, Muller R, Zernicke RF (2002). Mechanical and architectural bone adaptation in early stage experimental osteoarthritis. *J Bone Miner Res 17*(4), 687–694.
77. Chalmers HJ, Dykes NL, Lust G, Farese JP, Burton-Wurster NI, Williams AJ, Todhunter RJ (2006). Assessment of bone mineral density of the femoral head in dogs with early osteoarthritis. *Am J Vet Res 67*(5), 796–800.
78. Hunter DJ, Zhang Y, Niu J, Goggins J, Amin S, LaValley MP, Guermazi A, Genant H, Gale D, Felson DT (2006). Increase in bone marrow lesions associated with cartilage loss: A longitudinal magnetic resonance imaging study of knee osteoarthritis. *Arthritis Rheum 54*(5), 1529–1535.
79. Sanchez C, Deberg MA, Piccardi N, Msika P, Reginster JY, Henrotin YE (2005). Osteoblasts from the sclerotic subchondral bone downregulate aggrecan but upregulate metalloproteinases expression by chondrocytes. This effect is mimicked by interleukin-6, -1beta and oncostatin M pre-treated non-sclerotic osteoblasts. *Osteoarthritis Cartilage 13*(11), 979–987.
80. Sanchez C, Deberg MA, Piccardi N, Msika P, Reginster JY, Henrotin YE (2005). Subchondral bone osteoblasts induce phenotypic changes in human osteoarthritic chondrocytes. *Osteoarthritis Cartilage 13*(11), 988–997.
81. Hilal G, Martel-Pelletier J, Pelletier JP, Duval N, Lajeunesse D (1999). Abnormal regulation of urokinase plasminogen activator by insulin-like growth factor 1 in human osteoarthritic subchondral osteoblasts. *Arthritis Rheum 42*(10), 2112–2122.
82. Massicotte F, Lajeunesse D, Benderdour M, Pelletier JP, Hilal G, Duval N, Martel-Pelletier J (2002). Can altered production of interleukin-1beta, interleukin-6, transforming growth factor-beta and prostaglandin E(2) by isolated human subchondral osteoblasts identify two subgroups of osteoarthritic patients. *Osteoarthritis Cartilage 10*(6), 491–500.

83. Hilal G, Massicotte F, Martel-Pelletier J, Fernandes JC, Pelletier JP, Lajeunesse D (2001). Endogenous prostaglandin E2 and insulin-like growth factor 1 can modulate the levels of parathyroid hormone receptor in human osteoarthritic osteoblasts. *J Bone Miner Res 16*(4), 713–721.
84. Bailey AJ, Sims TJ, Knott L (2002). Phenotypic expression of osteoblast collagen in osteoarthritic bone: Production of type I homotrimer. *Int J Biochem Cell Biol 34*(2), 176–182.
85. Radin EL, Paul IL, Tolkoff MJ (1970). Subchondral bone changes in patients with early degenerative joint disease. *Arthritis Rheum 13*(4), 400–405.
86. Sokoloff L (1993). Microcracks in the calcified layer of articular cartilage. *Arch Pathol Lab Med 117*(2), 191–195.
87. Guevremont M, Martel-Pelletier J, Massicotte F, Tardif G, Pelletier JP, Ranger P, Lajeunesse D, Reboul P (2003). Human adult chondrocytes express hepatocyte growth factor (HGF) isoforms but not HgF: Potential implication of osteoblasts on the presence of HGF in cartilage. *J Bone Miner Res 18*(6), 1073–1081.
88. Sharif M, George E, Dieppe PA (1995). Correlation between synovial fluid markers of cartilage and bone turnover and scintigraphic scan abnormalities in osteoarthritis of the knee. *Arthritis Rheum 38*(1), 78–81.
89. Westacott CI, Webb GR, Warnock MG, Sims JV, Elson CJ (1997). Alteration of cartilage metabolism by cells from osteoarthritic bone. *Arthritis Rheum 40*(7), 1282–1291.
90. EMEA TEA.f.t.E.o.M.P. Points to consider on clinical investigation of medicinal products used in the treatment of osteoarthritis.
91. Services USD.o.H.a.H., Food and Drug Administration CfD.E.a.R.C., CBER CfB.E.a.R. and Health. CfD.a.R., Guidance for industry. Clinical development programs for drugs, and biological products intended for the treatment of osteoarthritis (OA).
92. Abadie E, Ethgen D, Avouac B, Bouvenot G, Branco J, Bruyere O, Calvo G, Devogelaer JP, Dreiser RL, Herrero-Beaumont G, Kahan A, Kreutz G, Laslop A, Lemmel EM, Nuki G, Van De Putte L, Vanhaelst L, Reginster JY (2004). Recommendations for the use of new methods to assess the efficacy of disease-modifying drugs in the treatment of osteoarthritis. *Osteoarthritis Cartilage 12*(4), 263–268.
93. Altman RD, Abadie E, Avouac B, Bouvenot G, Branco J, Bruyere O, Calvo G, Devogelaer JP, Dreiser RL, Herrero-Beaumont G, Kahan A, Kreutz G, Laslop A, Lemmel EM, Menkes CJ, Pavelka K, Van De Putte L, Vanhaelst L, Reginster JY (2005). Total joint replacement of hip or knee as an outcome measure for structure modifying trials in osteoarthritis. *Osteoarthritis Cartilage 13*(1), 13–19.
94. Ameye LG, Chee WS (2006). Osteoarthritis and nutrition. From nutraceuticals to functional foods: A systematic review of the scientific evidence. *Arthritis Res Ther 8*(4), R127.
95. Diplock AT, Agget PJ, Ashwell M, Bornet F, Fern FB, Roberfroid MB (1999). Scientific concepts of functional foods in Europe: Consensus document. *Br J Nutr 81*(Suppl 1), S1–S28.
96. Roberfroid MB (2000). A European consensus of scientific concepts of functional foods. *Nutrition 16*(7–8), 689–691.
97. Roberfroid MB (2000). Concepts and strategy of functional food science: The European perspective. *Am J Clin Nutr 71*(6 Suppl), 1660S–16604S; discussion 1674S–1675S.
98. Persiani S, Roda E, Rovati LC, Locatelli M, Giacovelli G, Roda A (2005). Glucosamine oral bioavailability and plasma pharmacokinetics after increasing doses of crystalline glucosamine sulfate in man. *Osteoarthritis Cartilage 13*(12), 1041–1049.
99. Persiani S, Rotini R, Trisolino G, Rovati L, Locatelli M, Roda A (2005). Glucosamine plasma and synovial fluid concentrations before and after administration of crystalline glucosamine sulfate in knee osteoarthritis patients. *Arthritis Rheum 9*(Suppl), 1342.
100. Bassleer C, Rovati L, Franchimont P (1998). Stimulation of proteoglycan production by glucosamine sulfate in chondrocytes isolated from human osteoarthritic articular cartilage in vitro. *Osteoarthritis Cartilage 6*(6), 427–434.
101. Dodge GR, Jimenez SA (2003). Glucosamine sulfate modulates the levels of aggrecan and matrix metalloproteinase-3 synthesized by cultured human osteoarthritis articular chondrocytes. *Osteoarthritis Cartilage 11*(6), 424–432.
102. Bassleer CT, Combal JP, Bougaret S, Malaise M (1998). Effects of chondroitin sulfate and interleukin-1 beta on human articular chondrocytes cultivated in clusters. *Osteoarthritis Cartilage 6*(3), 196–204.
103. Piperno M, Reboul P, Hellio Le Graverand MP, Peschard MJ, Annefeld M, Richard M, Vignon E (2000). Glucosamine sulfate modulates dysregulated activities of human osteoarthritic chondrocytes in vitro. *Osteoarthritis Cartilage 8*(3), 207–212.
104. Largo R, Alvarez-Soria MA, Diez-Ortego I, Calvo E, Sanchez-Pernaute O, Egido J, Herrero-Beaumont G (2003). Glucosamine inhibits IL-1beta-induced NFkappaB activation in human osteoarthritic chondrocytes. *Osteoarthritis Cartilage 11*(4), 290–298.
105. Piepoli T, Zanelli T, Letari O, Persiani S, Rovati L, Caselli G (2005). Glucosamine sulfate inhibits IL-1b-stimulated gene expression at concentration found in human after oral intake. *Arthritis Rheum 19*(Suppl), 1343.
106. Letari O, Colombo S, Zanelli T, Piperno M, Makovec F, Rovati L (2003). The mechanism of action of glucosamine sulfate in osteoarthritis: IL-1β mediated reactive oxygen species and modulation of NF-kB. *Arthrtis Rheum 48*(8), S286.

107. Hoffer LJ, Kaplan LN, Hamadeh MJ, Grigoriu AC, Baron M (2001). Sulfate could mediate the therapeutic effect of glucosamine sulfate. *Metabolism 50*(7), 767–770.
108. Cordoba F, Nimni ME (2003). Chondroitin sulfate and other sulfate containing chondroprotective agents may exhibit their effects by overcoming a deficiency of sulfur amino acids. *Osteoarthritis Cartilage 11*(3), 228–230.
109. Neil KM, Orth MW, Coussens PM, Chan PS, Caron JP (2005). Effects of glucosamine and chondroitin sulfate on mediators of osteoarthritis in cultured equine chondrocytes stimulated by use of recombinant equine interleukin-1beta. *Am J Vet Res 66*(11), 1861–1869.
110. Gouze JN, Bianchi A, Becuwe P, Dauca M, Netter P, Magdalou J, Terlain B, Bordji K (2002). Glucosamine modulates IL-1-induced activation of rat chondrocytes at a receptor level, and by inhibiting the NF-kappa B pathway. *FEBS Lett 510*(3), 166–170.
111. Chan PS, Caron JP, Orth MW (2005). Effect of glucosamine and chondroitin sulfate on regulation of gene expression of proteolytic enzymes and their inhibitors in interleukin-1-challenged bovine articular cartilage explants. *Am J Vet Res 66*(11), 1870–1876.
112. Chan PS, Caron JP, Rosa GJ, Orth MW (2005). Glucosamine and chondroitin sulfate regulate gene expression and synthesis of nitric oxide and prostaglandin E(2) in articular cartilage explants. *Osteoarthritis Cartilage 13*(5), 387–394.
113. McAlindon TE, LaValley MP, Gulin JP, Felson DT (2000). Glucosamine and chondroitin for treatment of osteoarthritis: A systematic quality assessment and meta-analysis. *JAMA 283*(11), 1469–1475.
114. Richy F, Bruyere O, Ethgen O, Cucherat M, Henrotin Y, Reginster JY (2003). Structural and symptomatic efficacy of glucosamine and chondroitin in knee osteoarthritis: A comprehensive meta-analysis. *Arch Intern Med 163*(13), 1514–1522.
115. Towheed TE, Anastassiades TP (1999). Glucosamine therapy for osteoarthritis. *J Rheumatol 26*(11), 2294–2297.
116. Reginster JY, Deroisy R, Rovati LC, Lee RL, Lejeune E, Bruyere O, Giacovelli G, Henrotin Y, Dacre JE, Gossett C (2001). Long-term effects of glucosamine sulphate on osteoarthritis progression: A randomised, placebo-controlled clinical trial. *Lancet 357*(9252), 251–256.
117. Pavelka K, Gatterova J, Olejarova M, Machacek S, Giacovelli G, Rovati LC (2002). Glucosamine sulfate use and delay of progression of knee osteoarthritis: A 3-year, randomized, placebo-controlled, double-blind study. *Arch Intern Med 162*(18), 2113–2123.
118. Clegg DO, Reda DJ, Harris CL, Klein MA, O'Dell JR, Hooper MM, Bradley JD, Bingham CO, 3rd, Weisman MH, Jackson CG, Lane NE, Cush JJ, Moreland LW, Schumacher HR Jr, Oddis CV, Wolfe F, Molitor JA, Yocum DE, Schnitzer TJ, Furst DE, Sawitzke AD, Shi H, Brandt KD, Moskowitz RW, Williams HJ (2006). Glucosamine, chondroitin sulfate, and the two in combination for painful knee osteoarthritis. *N Engl J Med 354*(8), 795–808.
119. Herrero-Beaumont G, Roman JA, Trabado MC, Blanco FJ, Benito P, Martin-Mola E, Paulino J, Marenco JL, Porto A, Laffon A, Araujo D, Figueroa M, Branco J (2005). The Glucosamine Unum in Die Efficacy (GUIDE) trial. *Arthritis Rheum 9*(Suppl), 1230.
120. Lo GH, LaValley M, McAlindon T, Felson DT (2003). Intra-articular hyaluronic acid in treatment of knee osteoarthritis: A meta-analysis. *JAMA 290*(23), 3115–3121.
121. Arrich J, Piribauer F, Mad P, Schmid D, Klaushofer K, Mullner M (2005). Intra-articular hyaluronic acid for the treatment of osteoarthritis of the knee: Systematic review and meta-analysis. *CMAJ 172*(8), 1039–1043.
122. Maheu E, Ayral X, Dougados M (2002). A hyaluronan preparation (500–730 kDa) in the treatment of osteoarthritis: A review of clinical trials with Hyalgan. *Int J Clin Pract 56*(10), 804–813.
123. Qvistgaard E, Christensen R, Torp-Pedersen S, Bliddal H (2006). Intra-articular treatment of hip osteoarthritis: A randomized trial of hyaluronic acid, corticosteroid, and isotonic saline. *Osteoarthritis Cartilage 14*(2), 163–170.
124. Neustadt D, Caldwell J, Bell M, Wade J, Gimbel J (2005). Clinical effects of intraarticular injection of high molecular weight hyaluronan (Orthovisc) in osteoarthritis of the knee: A randomized, controlled, multicenter trial. *J Rheumatol 32*(10), 1928–1936.
125. Bellamy N, Bell MJ, Goldsmith CH, Pericak D, Walker V, Raynauld JP, Torrance GW, Tugwell P, Polisson R (2005). The effectiveness of hylan G-F 20 in patients with knee osteoarthritis: An application of two sets of response criteria developed by the OARSI and one set developed by OMERACT-OARSI. *Osteoarthritis Cartilage 13*(2), 104–110.
126. Fuchs S, Erbe T, Fischer HL, Tibesku CO (2005). Intraarticular hyaluronic acid versus glucocorticoid injections for nonradicular pain in the lumbar spine. *J Vasc Interv Radiol 16*(11), 1493–1498.
127. Karatosun V, Unver B, Gocen Z, Sen A (2005). Comparison of two hyaluronan drugs in patients with advanced osteoarthritis of the knee. A prospective, randomized, double-blind study with long term follow-up. *Clin Exp Rheumatol 23*(2), 213–218.
128. Kotevoglu N, Iyibozkurt PC, Hiz O, Toktas H, Kuran B (2006). A prospective randomised controlled clinical trial comparing the efficacy of different molecular weight hyaluronan solutions in the treatment of knee osteoarthritis. *Rheumatol Int 26*(4), 325–330.
129. Raynauld JP, Goldsmith CH, Bellamy N, Torrance GW, Polisson R, Belovich D, Pericak D, Tugwell P (2005). Effectiveness and safety of repeat courses of hylan G-F 20 in patients with knee osteoarthritis. *Osteoarthritis Cartilage 13*(2), 111–119.
130. Petrella RJ, Petrella M (2006). A prospective, randomized, double-blind, placebo controlled study to evaluate

the efficacy of intraarticular hyaluronic acid for osteoarthritis of the knee. *J Rheumatol 33*(5), 951–956.

131. Sun SF, Chou YJ, Hsu CW, Hwang CW, Hsu PT, Wang JL, Hsu YW, Chou MC (2006). Efficacy of intra-articular hyaluronic acid in patients with osteoarthritis of the ankle: A prospective study. *Osteoarthritis Cartilage 14*(9), 867–874.
132. Sun SF, Hsu CW, Hwang CW, Hsu PT, Wang JL, Tsai SL, Chou YJ, Hsu YW, Huang CM, Wang YL (2006). Hyaluronate improves pain, physical function and balance in the geriatric osteoarthritic knee: A 6-month follow-up study using clinical tests. *Osteoarthritis Cartilage*.
133. Salk RS, Chang TJ, D'Costa WF, Soomekh DJ, Grogan KA (2006). Sodium hyaluronate in the treatment of osteoarthritis of the ankle: A controlled, randomized, double-blind pilot study. *J Bone Joint Surg Am 88*(2), 295–302.
134. Balazs EA, Watson D, Duff IF, Roseman S (1967). Hyaluronic acid in synovial fluid. I. Molecular parameters of hyaluronic acid in normal and arthritis human fluids. *Arthritis Rheum 10*(4), 357–376.
135. Goldberg VM, Buckwalter JA (2005). Hyaluronans in the treatment of osteoarthritis of the knee: Evidence for disease-modifying activity. *Osteoarthritis Cartilage 13*(3), 216–224.
136. Pasquali Ronchetti I, Guerra D, Taparelli F, Boraldi F, Bergamini G, Mori G, Zizzi F, Frizziero L (2001). Morphological analysis of knee synovial membrane biopsies from a randomized controlled clinical study comparing the effects of sodium hyaluronate (Hyalgan) and methylprednisolone acetate (Depomedrol) in osteoarthritis. *Rheumatol (Oxford) 40*(2), 158–169.
137. Guidolin DD, Ronchetti IP, Lini E, Guerra D, Frizziero L (2001). Morphological analysis of articular cartilage biopsies from a randomized, clinical study comparing the effects of 500–730 kDa sodium hyaluronate (Hyalgan) and methylprednisolone acetate on primary osteoarthritis of the knee. *Osteoarthritis Cartilage 9*(4), 371–381.
138. Frizziero L, Govoni E, Bacchini P (1998). Intra-articular hyaluronic acid in the treatment of osteoarthritis of the knee: Clinical and morphological study. *Clin Exp Rheumatol 16*(4), 441–449.
139. Jubb RW, Piva S, Beinat L, Dacre J, Gishen P (2003). A one-year, randomised, placebo (saline) controlled clinical trial of 500–730 kDa sodium hyaluronate (Hyalgan) on the radiological change in osteoarthritis of the knee. *Int J Clin Pract 57*(6), 467–474.
140. Julovi SM, Yasuda T, Shimizu M, Hiramitsu T, Nakamura T (2004). Inhibition of interleukin-1beta-stimulated production of matrix metalloproteinases by hyaluronan via CD44 in human articular cartilage. *Arthritis Rheum 50*(2), 516–525.
141. Fukuda K, Dan H, Takayama M, Kumano F, Saitoh M, Tanaka S (1996). Hyaluronic acid increases proteoglycan synthesis in bovine articular cartilage in the presence of interleukin-1. *J Pharmacol Exp Ther 277*(3), 1672–1675.
142. Fukuda K, Takayama M, Ueno M, Oh M, Asada S, Kumano F, Tanaka S (1997). Hyaluronic acid inhibits interleukin-1-induced superoxide anion in bovine chondrocytes. *Inflamm Res 46*(3), 114–117.
143. Kawasaki K, Ochi M, Uchio Y, Adachi N, Matsusaki M (1999). Hyaluronic acid enhances proliferation and chondroitin sulfate synthesis in cultured chondrocytes embedded in collagen gels. *J Cell Physiol 179*(2), 142–148.
144. Aruffo A, Stamenkovic I, Melnick M, Underhill CB, Seed B (1990). CD44 is the principal cell surface receptor for hyaluronate. *Cell 61*(7), 1303–1313.
145. Knudson CB, Knudson W (2004). Hyaluronan and CD44: Modulators of chondrocyte metabolism. *Clin Orthop Relat Res 427*(Suppl), S152–S162.
146. Ghosh P, Guidolin D (2002). Potential mechanism of action of intra-articular hyaluronan therapy in osteoarthritis: Are the effects molecular weight dependent? *Semin Arthritis Rheum 32*(1), 10–37.
147. Mauviel A, Daireaux M, Hartmann DJ, Galera P, Loyau G, Pujol JP (1989). [Effects of unsaponifiable extracts of avocado/soy beans (PIAS) on the production of collagen by cultures of synoviocytes, articular chondrocytes and skin fibroblasts]. *Rev Rhum Mal Osteoartic 56*(2), 207–211.
148. Mauviel A, Loyau G, Pujol JP (1991). [Effect of unsaponifiable extracts of avocado and soybean (Piascledine) on the collagenolytic action of cultures of human rheumatoid synoviocytes and rabbit articular chondrocytes treated with interleukin-1]. *Rev Rhum Mal Osteoartic 58*(4), 241–245.
149. Henrotin YE, Labasse AH, Jaspar JM, De Groote DD, Zheng SX, Guillou GB, Reginster JY (1998). Effects of three avocado/soybean unsaponifiable mixtures on metalloproteinases, cytokines and prostaglandin E2 production by human articular chondrocytes. *Clin Rheumatol 17*(1), 31–39.
150. Henrotin YE, Sanchez C, Deberg MA, Piccardi N, Guillou GB, Msika P, Reginster JY (2003). Avocado/soybean unsaponifiables increase aggrecan synthesis and reduce catabolic and proinflammatory mediator production by human osteoarthritic chondrocytes. *J Rheumatol 30*(8), 1825–1834.
151. Boumediene K, Felisaz N, Bogdanowicz P, Galera P, Guillou GB, Pujol JP (1999). Avocado/soya unsaponifiables enhance the expression of transforming growth factor beta1 and beta2 in cultured articular chondrocytes. *Arthritis Rheum 42*(1), 148–156.
152. Henrotin YE, Deberg MA, Crielaard JM, Piccardi N, Msika P, Sanchez C (2006). Avocado/soybean unsaponifiables prevent the inhibitory effect of osteoarthritic subchondral osteoblasts on aggrecan and type II collagen synthesis by chondrocytes. *J Rheumatol 33*(8), 1668–1678.
153. Henrotin Y, Deberg M, Crielaard JM, Piccardi N, Msika P, Sanchez C (2006). Avocado/soybean unsaponifiables prevent the inhibitory effect of osteoarthritic

subchondral osteoblasts on aggrecan and type II collagen synthesis by chondrocytes. *J Rheumatol 33*(8), 1168–1178.

154. Cake MA, Read RA, Guillou B, Ghosh P (2000). Modification of articular cartilage and subchondral bone pathology in an ovine meniscectomy model of osteoarthritis by avocado and soya unsaponifiables (ASU). *Osteoarthritis Cartilage 8*(6), 404–411.
155. Appelboom T, Schuermans J, Verbruggen G, Henrotin Y, Reginster JY (2001). Symptoms modifying effect of avocado/soybean unsaponifiables (ASU) in knee osteoarthritis. A double blind, prospective, placebo-controlled study. *Scand J Rheumatol 30*(4), 242–247.
156. Blotman F, Maheu E, Wulwik A, Caspard H, Lopez A (1997). Efficacy and safety of avocado/soybean unsaponifiables in the treatment of symptomatic osteoarthritis of the knee and hip. A prospective, multicenter, three-month, randomized, double-blind, placebo-controlled trial. *Rev Rhum Engl Ed 64*(12), 825–834.
157. Maheu E, Mazieres B, Valat JP, Loyau G, Le Loet X, Bourgeois P, Grouin JM, Rozenberg S (1998). Symptomatic efficacy of avocado/soybean unsaponifiables in the treatment of osteoarthritis of the knee and hip: A prospective, randomized, double-blind, placebo-controlled, multicenter clinical trial with a six-month treatment period and a two-month followup demonstrating a persistent effect. *Arthritis Rheum 41*(1), 81–91.
158. Lequesne M, Maheu E, Cadet C, Dreiser RL (2002). Structural effect of avocado/soybean unsaponifiables on joint space loss in osteoarthritis of the hip. *Arthritis Rheum 47*(1), 50–58.
159. Darlington LG, Stone TW (2001). Antioxidants and fatty acids in the amelioration of rheumatoid arthritis and related disorders. *Br J Nutr 85*(3), 251–269.
160. Calder PC (2004). n-3 Fatty acids and cardiovascular disease: Evidence explained and mechanisms explored. *Clin Sci (Lond) 107*(1), 1–11.
161. Lippiello L, Walsh T, Fienhold M (1991). The association of lipid abnormalities with tissue pathology in human osteoarthritic articular cartilage. *Metabolism 40*(6), 571–576.
162. Lippiello L (1990). Lipid and cell metabolic changes associated with essential fatty acid enrichment of articular chondrocytes. *Proc Soc Exp Biol Med 195*(2), 282–287.
163. Curtis CL, Hughes CE, Flannery CR, Little CB, Harwood JL, Caterson B (2000). n-3 fatty acids specifically modulate catabolic factors involved in articular cartilage degradation. *J Biol Chem 275*(2), 721–724.
164. Curtis CL, Rees SG, Little CB, Flannery CR, Hughes CE, Wilson C, Dent CM, Otterness IG, Harwood JL, Caterson B (2002). Pathologic indicators of degradation and inflammation in human osteoarthritic cartilage are abrogated by exposure to n-3 fatty acids. *Arthritis Rheum 46*(6), 1544–1553.
165. Curtis CL, Rees SG, Cramp J, Flannery CR, Hughes CE, Little CB, Williams R, Wilson C, Dent CM, Harwood JL, Caterson B (2002). Effects of n-3 fatty acids on cartilage metabolism. *Proc Nutr Soc 61*(3), 381–389.
166. Rainsford KD, Whitehouse MW (1980). Gastroprotective and anti-inflammatory properties of green lipped mussel (*Perna canaliculus*) preparation. *Arzneimittelforschung 30*(12), 2128–2132.
167. Bui LM, Bierer TL (2003). Influence of green lipped mussels (*Perna canaliculus*) in alleviating signs of arthritis in dogs. *Vet Ther 4*(4), 397–407.
168. Cho SH, Jung YB, Seong SC, Park HB, Byun KY, Lee DC, Song EK, Son JH (2003). Clinical efficacy and safety of Lyprinol, a patented extract from New Zealand green-lipped mussel (*Perna canaliculus*) in patients with osteoarthritis of the hip and knee: A multicenter 2-month clinical trial. *Allerg Immunol (Paris) 35*(6), 212–216.
169. McAlindon TE, Felson DT, Zhang Y, Hannan MT, Aliabadi P, Weissman B, Rush D, Wilson PW, Jacques P (1996). Relation of dietary intake and serum levels of vitamin D to progression of osteoarthritis of the knee among participants in the Framingham Study. *Ann Intern Med 125*(5), 353–359.
170. Lane NE, Gore LR, Cummings SR, Hochberg MC, Scott JC, Williams EN, Nevitt MC (1999). Serum vitamin D levels and incident changes of radiographic hip osteoarthritis: A longitudinal study. Study of Osteoporotic Fractures Research Group. *Arthritis Rheum 42*(5), 854–860.
171. Clark AG, Rohrbaugh AL, Otterness I, Kraus VB (2002). The effects of ascorbic acid on cartilage metabolism in guinea pig articular cartilage explants. *Matrix Biol 21*(2), 175–184.
172. Schwartz ER, Adamy L (1977). Effect of ascorbic acid on arylsulfatase activities and sulfated proteoglycan metabolism in chondrocyte cultures. *J Clin Invest 60*(1), 96–106.
173. Daniel JC, Pauli BU, Kuettner KE (1984). Synthesis of cartilage matrix by mammalian chondrocytes in vitro. III. Effects of ascorbate. *J Cell Biol 99*(6), 1960–1969.
174. Sandell LJ, Daniel JC (1988). Effects of ascorbic acid on collagen mRNA levels in short term chondrocyte cultures. *Connect Tissue Res 17*(1), 11–22.
175. Tiku ML, Gupta S, Deshmukh DR (1999). Aggrecan degradation in chondrocytes is mediated by reactive oxygen species and protected by antioxidants. *Free Radic Res 30*(5), 395–405.
176. Verzijl N, DeGroot J, Ben ZC, Brau-Benjamin O, Maroudas A, Bank RA, Mizrahi J, Schalkwijk CG, Thorpe SR, Baynes JW, Bijlsma JW, Lafeber FP, TeKoppele JM (2002). Crosslinking by advanced glycation end products increases the stiffness of the collagen network in human articular cartilage: A possible mechanism through which age is a risk factor for osteoarthritis. *Arthritis Rheum 46*(1), 114–123.
177. Kraus VB, Huebner JL, Stabler T, Flahiff CM, Setton LA, Fink C, Vilim V, Clark AG (2004). Ascorbic acid increases the severity of spontaneous knee osteoarthritis in a guinea pig model. *Arthritis Rheum 50*(6), 1822–1831.

178. Schwartz ER (1984). The modulation of osteoarthritic development by vitamins C and E. *Int J Vitam Nutr Res Suppl 26*, 141–146.
179. Meacock SC, Bodmer JL, Billingham ME (1990). Experimental osteoarthritis in guinea-pigs. *J Exp Pathol (Oxford) 71*(2), 279–293.
180. Jensen NH (2003). [Reduced pain from osteoarthritis in hip joint or knee joint during treatment with calcium ascorbate. A randomized, placebo-controlled cross-over trial in general practice]. *Ugeskr Laeger 165*(25), 2563–2566.
181. Schwartz ER (1979). Effect of vitamins C and E on sulfated proteoglycan metabolism and sulfatase and phosphatase activities in organ cultures of human cartilage. *Calcif Tissue Int 28*(3), 201–208.
182. Galleron S, Borderie D, Ponteziere C, Lemarechal H, Jambou M, Roch-Arveiller M, Ekindjian OG, Cals MJ (1999). Reactive oxygen species induce apoptosis of synoviocytes in vitro. Alpha-tocopherol provides no protection. *Cell Biol Int 23*(9), 637–642.
183. Machtey I, Ouaknine L (1978). Tocopherol in osteoarthritis: A controlled pilot study. *J Am Geriatr Soc 26*(7), 328–330.
184. Blankenhorn G (1986). [Clinical effectiveness of Spondyvit (vitamin E) in activated arthroses. A multicenter placebo-controlled double-blind study]. *Z Orthop Ihre Grenzgeb 124*(3), 340–343.
185. Scherak O, Kolarz G, Schodl C, Blankenhorn G (1990). High dosage vitamin E therapy in patients with activated arthrosis. *Z Rheumatol 49*, 369–373.
186. Brand C, Snaddon J, Bailey M, Cicuttini F (2001). Vitamin E is ineffective for symptomatic relief of knee osteoarthritis: A six month double blind, randomised, placebo controlled study. *Ann Rheum Dis 60*(10), 946–949.
187. Wluka AE, Stuckey S, Brand C, Cicuttini FM (2002). Supplementary vitamin E does not affect the loss of cartilage volume in knee osteoarthritis: A 2 year double blind randomized placebo controlled study. *J Rheumatol 29*(12), 2585–2591.
188. Masse PG, Ziv I, Cole DE, Mahuren JD, Donovan SM, Yamauchi M, Howell DS (1998). A cartilage matrix deficiency experimentally induced by vitamin B6 deficiency. *Proc Soc Exp Biol Med 217*(1), 97–103.
189. Jonas WB, Rapoza CP, Blair WF (1996). The effect of niacinamide on osteoarthritis: A pilot study. *Inflamm Res 45*(7), 330–334.
190. Flynn MA, Irvin W, Krause G (1994). The effect of folate and cobalamin on osteoarthritic hands. *J Am Coll Nutr 13*(4), 351–356.
191. Kobayashi T, Notoya K, Nakamura A, Akimoto K (2005). Fursultiamine, a vitamin B1 derivative, enhances chondroprotective effects of glucosamine hydrochloride and chondroitin sulfate in rabbit experimental osteoarthritis. *Inflamm Res 54*(6), 249–255.
192. Hill J, Bird HA (1990). Failure of selenium-ace to improve osteoarthritis. *Br J Rheumatol 29*(3), 211–213.
193. Kurz B, Jost B, Schunke M (2002). Dietary vitamins and selenium diminish the development of mechanically induced osteoarthritis and increase the expression of antioxidative enzymes in the knee joint of STR/1N mice. *Osteoarthritis Cartilage 10*(2), 119–126.
194. Brien S, Lewith G, Walker A, Hicks SM, Middleton D (2004). Bromelain as a treatment for osteoarthritis: A review of clinical studies. *Evid Based Complement Alternat Med 1*(3), 251–257.
195. Maurer HR (2001). Bromelain: Biochemistry, pharmacology and medical use. *Cell Mol Life Sci 58*(9), 1234–1245.
196. Akhtar NM, Naseer R, Farooqi AZ, Aziz W, Nazir M (2004). Oral enzyme combination versus diclofenac in the treatment of osteoarthritis of the knee–A double-blind prospective randomized study. *Clin Rheumatol 23*(5), 410–415.
197. Klein G, Kullich W, Schnitker J, Schwann H (2006). Efficacy and tolerance of an oral enzyme combination in painful osteoarthritis of the hip. A double-blind, randomised study comparing oral enzymes with non-steroidal anti-inflammatory drugs. *Clin Exp Rheumatol 24*(1), 25–30.
198. Singer F, Oberleitner H (1996). [Drug therapy of activated arthrosis. On the effectiveness of an enzyme mixture versus diclofenac]. *Wien Med Wochenschr 146*(3), 55–58.
199. Daels-Rakotoarison DA, Gressier B, Trotin F, Brunet C, Luyckx M, Dine T, Bailleul F, Cazin M, Cazin JC (2002). Effects of *Rosa canina* fruit extract on neutrophil respiratory burst. *Phytother Res 16*(2), 157–161.
200. Warholm O, Skaar S, Hedman E, Molmen HM, Eik L (2005). The effects of a standardized herbal remedy made from a subtype of *Rosa canina* in patients with osteoarthritis: A double-blind, randomized, placebo-controlled clinical trial. *Curr Ther res 64*, 21–31.
201. Rein MJ, Heinonen M (2004). Stability and enhancement of berry juice color. *J Agric Food Chem 52*(10), 3106–3114.
202. Winther K, Apel K, Thamsborg G (2005). A powder made from seeds and shells of a rose-hip subspecies (*Rosa canina*) reduces symptoms of knee and hip osteoarthritis: A randomized, double-blind, placebo-controlled clinical trial. *Scand J Rheumatol 34*(4), 302–308.
203. Chrubasik S, Conradt C, Roufogalis BD (2004). Effectiveness of *Harpagophytum* extracts and clinical efficacy. *Phytother Res 18*(2), 187–189.
204. Chrubasik JE, Lindhorst E, Neumann E, Gerlach U, Faller-Marquardt M, Torda T, Muller-Ladner U, Chrubasik S (2006). Potential molecular basis of the chondroprotective effect of *Harpagophytum procumbens*. *Phytomedicine 13*(8), 598–600.
205. Huang TH, Tran VH, Duke RK, Tan S, Chrubasik S, Roufogalis BD, Duke CC (2006). Harpagoside suppresses lipopolysaccharide-induced iNOS and COX-2 expression through inhibition of NF-kappaB activation. *J Ethnopharmacol 104*(1–2), 149–155.

206. Fiebich BL, Heinrich M, Hiller KO, Kammerer N (2001). Inhibition of TNF-alpha synthesis in LPS-stimulated primary human monocytes by *Harpagophytum* extract SteiHap 69. *Phytomedicine 8*(1), 28–30.

207. Schulze-Tanzil G, Hansen C, Shakibaei M (2004). [Effect of a *Harpagophytum procumbens* DC extract on matrix metalloproteinases in human chondrocytes in vitro]. *Arzneimittelforschung 54*(4), 213–220.

208. Gagnier JJ, Chrubasik S, Manheimer E (2004). *Harpagophytum procumbens* for osteoarthritis and low back pain: A systematic review. *BMC Complement Altern Med 4*, 13.

209. Lecomte A, Costa JP (1992). Harpagophytum dans l'arthrose: Étude en double insu contre placebo. *Le Magazine 15*, 27–30.

210. Leblan D, Chantre P, Fournie B (2000). *Harpagophytum procumbens* in the treatment of knee and hip osteoarthritis. Four-month results of a prospective, multicenter, double-blind trial versus diacerhein. *Joint Bone Spine 67*(5), 462–467.

211. Chantre P, Cappelaere A, Leblan D, Guedon D, Vandermander J, Fournie B (2000). Efficacy and tolerance of *Harpagophytum procumbens* versus diacerhein in treatment of osteoarthritis.*Phytomedicine 7*(3), 177–183.

212. Wegener T, Lupke NP (2003). Treatment of patients with arthrosis of hip or knee with an aqueous extract of devil's claw (*Harpagophytum procumbens* DC). *Phytother Res 17*(10), 1165–1172.

213. Sandoval M, Okuhama NN, Zhang XJ, Condezo LA, Lao J, Angeles FM, Musah RA, Bobrowski P, Miller MJ (2002). Anti-inflammatory and antioxidant activities of cat's claw (*Uncaria tomentosa* and *Uncaria guianensis*) are independent of their alkaloid content. *Phytomedicine 9*(4), 325–337.

214. Piscoya J, Rodriguez Z, Bustamante SA, Okuhama NN, Miller MJ, Sandoval M (2001). Efficacy and safety of freeze-dried cat's claw in osteoarthritis of the knee: Mechanisms of action of the species *Uncaria guianensis*. *Inflamm Res 50*(9), 442–448.

215. Jha V, Chugh KS (2003). Nephropathy associated with animal, plant, and chemical toxins in the tropics. *Semin Nephrol 23*(1), 49–65.

216. Schmid B, Kotter I, Heide L (2001). Pharmacokinetics of salicin after oral administration of a standardised willow bark extract. *Eur J Clin Pharmacol 57*(5), 387–391.

217. Biegert C, Wagner I, Ludtke R, Kotter I, Lohmuller C, Gunaydin I, Taxis K, Heide L (2004). Efficacy and safety of willow bark extract in the treatment of osteoarthritis and rheumatoid arthritis: Results of 2 randomized double-blind controlled trials. *J Rheumatol 31*(11), 2121–2130.

218. Mills SY, Jacoby RK, Chacksfield M, Willoughby M (1996). Effect of a proprietary herbal medicine on the relief of chronic arthritic pain: A double-blind study. *Br J Rheumatol 35*(9), 874–878.

219. Chrubasik S, Kunzel O, Model A, Conradt C, Black A (2001). Treatment of low back pain with a herbal or synthetic anti-rheumatic: A randomized controlled study. Willow bark extract for low back pain. *Rheumatology (Oxford) 40*(12), 1388–1393.

220. Shen CL, Hong KJ, Kim SW (2003). Effects of ginger (*Zingiber officinale Rosc.*) on decreasing the production of inflammatory mediators in sow osteoarthrotic cartilage explants. *J Med Food 6*(4), 323–328.

221. Frondoza CG, Sohrabi A, Polotsky A, Phan PV, Hungerford DS, Lindmark L (2004). An in vitro screening assay for inhibitors of proinflammatory mediators in herbal extracts using human synoviocyte cultures. *In Vitro Cell Dev Biol Anim 40*(3–4), 95–101.

222. Bliddal H, Rosetzsky A, Schlichting P, Weidner MS, Andersen LA, Ibfelt HH, Christensen K, Jensen ON, Barslev J (2000). A randomized, placebo-controlled, cross-over study of ginger extracts and ibuprofen in osteoarthritis. *Osteoarthritis Cartilage 8*(1), 9–12.

223. Altman RD, Marcussen KC (2001). Effects of a ginger extract on knee pain in patients with osteoarthritis. *Arthritis Rheum 44*(11), 2531–2538.

224. Wigler I, Grotto I, Caspi D, Yaron M (2003). The effects of Zintona EC (a ginger extract) on symptomatic gonarthritis. *Osteoarthritis Cartilage 11*(11), 783–789.

225. Li WQ, Dehnade F, Zafarullah M (2001). Oncostatin M-induced matrix metalloproteinase and tissue inhibitor of metalloproteinase-3 genes expression in chondrocytes requires Janus kinase/STAT signaling pathway.*J Immunol 166*(5), 3491–3498.

226. Liacini A, Sylvester J, Li WQ, Huang W, Dehnade F, Ahmad M, Zafarullah M (2003). Induction of matrix metalloproteinase-13 gene expression by TNF-alpha is mediated by MAP kinases, AP-1, and NF-kappaB transcription factors in articular chondrocytes. *Exp Cell Res 288*(1), 208–217.

227. Schulze-Tanzil G, Mobasheri A, Sendzik J, John T, Shakibaei M (2004). Effects of curcumin (diferuloylmethane) on nuclear factor kappaB signaling in interleukin-1beta-stimulated chondrocytes. *Ann N Y Acad Sci 1030*, 578–586.

228. Shakibaei M, Schulze-Tanzil G, John T, Mobasheri A (2005). Curcumin protects human chondrocytes from IL-l1beta-induced inhibition of collagen type II and beta1-integrin expression and activation of caspase-3: An immunomorphological study. *Ann Anat 187*(5–6), 487–497.

229. Lev-Ari S, Maimon Y, Strier L, Kazanov D, Arber N (2006). Down-regulation of prostaglandin E2 by curcumin is correlated with inhibition of cell growth and induction of apoptosis in human colon carcinoma cell lines. *J Soc Integr Oncol 4*(1), 21–26.

230. Joe B, Lokesh BR (1997). Effect of curcumin and capsaicin on arachidonic acid metabolism and lysosomal enzyme secretion by rat peritoneal macrophages. *Lipids 32*(11), 1173–1180.

231. Cheng AL, Hsu CH, Lin JK, Hsu MM, Ho YF, Shen TS, Ko JY, Lin JT, Lin BR, Ming-Shiang W, Yu HS, Jee SH, Chen GS, Chen TM, Chen CA, Lai MK, Pu YS, Pan MH, Wang YJ, Tsai CC, Hsieh CY (2001). Phase I clinical trial of curcumin, a chemopreventive agent, in patients with high-risk or pre-malignant lesions. *Anticancer Res 21*(4B), 2895–2900.
232. Tzeng SH, Ko WC, Ko FN, Teng CM (1991). Inhibition of platelet aggregation by some flavonoids. *Thromb Res 64*(1), 91–100.
233. Selmi C, Mao TK, Keen CL, Schmitz HH, Eric Gershwin M (2006). The anti-inflammatory properties of cocoa flavanols. *J Cardiovasc Pharmacol 47*(Suppl 2), S163–S171.
234. Adlercreutz H (2002). Phytoestrogens and breast cancer. *J Steroid Biochem Mol Biol 83*(1–5), 113–118.
235. Pietta PG (2000). Flavonoids as antioxidants. *J Nat Prod 63*(7), 1035–1042.
236. Moroney MA, Alcaraz MJ, Forder RA, Carey F, Hoult JR (1988). Selectivity of neutrophil 5-lipoxygenase and cyclo-oxygenase inhibition by an anti-inflammatory flavonoid glycoside and related aglycone flavonoids. *J Pharm Pharmacol 40*(11), 787–792.
237. Beretz A, Anton R, Stoclet JC (1978). Flavonoid compounds are potent inhibitors of cyclic AMP phosphodiesterase. *Experientia 34*(8), 1054–1055.
238. Kuppusamy UR, Khoo HE, Das NP (1990). Structure-activity studies of flavonoids as inhibitors of hyaluronidase. *Biochem Pharmacol 40*(2), 397–401.
239. Ogawara H, Akiyama T, Watanabe S, Ito N, Kobori M, Seoda Y (1989). Inhibition of tyrosine protein kinase activity by synthetic isoflavones and flavones. *J Antibiot (Tokyo) 42*(2), 340–343.
240. Yamashita Y, Kawada S, Nakano H (1990). Induction of mammalian topoisomerase II dependent DNA cleavage by nonintercalative flavonoids, genistein and orobol. *Biochem Pharmacol 39*(4), 737–744.
241. Inouye Y, Yamaguchi K, Take Y, Nakamura S (1989). Inhibition of avian myeloblastosis virus reverse transcriptase by flavones and isoflavones. *J Antibiot (Tokyo) 42*(10), 1523–1525.
242. Bassleer CT, Franchimont PP, Henrotin YE, Franchimont NM, Geenen VG, Reginster JY (1996). Effects of ipriflavone and its metabolites on human articular chondrocytes cultivated in clusters. *Osteoarthritis Cartilage 4*(1), 1–8.
243. Ham KD, Loeser RF, Lindgren BR, Carlson CS (2002). Effects of long-term estrogen replacement therapy on osteoarthritis severity in cynomolgus monkeys. *Arthritis Rheum 46*(7), 1956–1964.
244. Ham KD, Oegema TR, Loeser RF, Carlson CS (2004). Effects of long-term estrogen replacement therapy on articular cartilage IGFBP-2, IGFBP-3, collagen and proteoglycan levels in ovariectomized cynomolgus monkeys. *Osteoarthritis Cartilage 12*(2), 160–168.
245. Arjmandi BH, Khalil DA, Lucas EA, Smith BJ, Sinichi N, Hodges SB, Juma S, Munson ME, Payton ME, Tivis RD, Svanborg A (2004). Soy protein may alleviate osteoarthritis symptoms. *Phytomedicine 11*(7–8), 567–575.
246. Katiyar SK, Agarwal R, Mukhtar H (1992). Green tea in chemoprevention of cancer. *Compr Ther 18*(10), 3–8.
247. Katiyar SK, Agarwal R, Mukhtar H (1993). Protection against malignant conversion of chemically induced benign skin papillomas to squamous cell carcinomas in SENCAR mice by a polyphenolic fraction isolated from green tea. *Cancer Res 53*(22), 5409–5412.
248. Katiyar SK, Rupp CO, Korman NJ, Agarwal R, Mukhtar H (1995). Inhibition of 12-*O*-tetradecanoylphorbol-13-acetate and other skin tumor-promoter-caused induction of epidermal interleukin-1 alpha mRNA and protein expression in SENCAR mice by green tea polyphenols. *J Invest Dermatol 105*(3), 394–398.
249. Mukhtar H, Katiyar SK, Agarwal R (1994). Green tea and skin—Anticarcinogenic effects. *J Invest Dermatol 102*(1), 3–7.
250. Mukhtar H, Ahmad N (1999). Mechanism of cancer chemopreventive activity of green tea. *Proc Soc Exp Biol Med 220*(4), 234–238.
251. Chen ZP, Schell JB, Ho CT, Chen KY (1998). Green tea epigallocatechin gallate shows a pronounced growth inhibitory effect on cancerous cells but not on their normal counterparts. *Cancer Lett 129*(2), 173–179.
252. Yang F, de Villiers WJ, McClain CJ, Varilek GW (1998). Green tea polyphenols block endotoxin-induced tumor necrosis factor—Production and lethality in a murine model. *J Nutr 128*(12), 2334–2340.
253. Adcocks C, Collin P, Buttle DJ (2002). Catechins from green tea (*Camellia sinensis*) inhibit bovine and human cartilage proteoglycan and type II collagen degradation in vitro. *J Nutr 132*(3), 341–346.
254. Singh R, Ahmed S, Islam N, Goldberg VM, Haqqi TM (2002). Epigallocatechin-3-gallate inhibits interleukin-1beta-induced expression of nitric oxide synthase and production of nitric oxide in human chondrocytes: Suppression of nuclear factor kappaB activation by degradation of the inhibitor of nuclear factor kappaB. *Arthritis Rheum 46*(8), 2079–2086.
255. Singh R, Ahmed S, Malemud CJ, Goldberg VM, Haqqi TM (2003). Epigallocatechin-3-gallate selectively inhibits interleukin-1beta-induced activation of mitogen activated protein kinase subgroup c-Jun N-terminal kinase in human osteoarthritis chondrocytes. *J Orthop Res 21*(1), 102–109.
256. Ahmed S, Wang N, Lalonde M, Goldberg VM, Haqqi TM (2004). Green tea polyphenol epigallocatechin-3-gallate (EGCG) differentially inhibits interleukin-1 beta-induced expression of matrix metalloproteinase-1 and -13 in human chondrocytes. *J Pharmacol Exp Ther 308*(2), 767–773.
257. Vankemmelbeke MN, Jones GC, Fowles C, Ilic MZ, Handley CJ, Day AJ, Knight CG, Mort JS, Buttle DJ (2003). Selective inhibition of ADAMTS-1, -4 and -5 by catechin gallate esters. *Eur J Biochem 270*(11), 2394–2403.

258. Yang CS, Chen L, Lee MJ, Balentine D, Kuo MC, Schantz SP (1998). Blood and urine levels of tea catechins after ingestion of different amounts of green tea by human volunteers. *Cancer Epidemiol Biomarkers Prev 7*(4), 351–354.

259. Pietta PG, Simonetti P, Gardana C, Brusamolino A, Morazzoni P, Bombardelli E (1998). Catechin metabolites after intake of green tea infusions. *Biofactors 8*(1–2), 111–118.

260. Ahmed S, Rahman A, Hasnain A, Lalonde M, Goldberg VM, Haqqi TM (2002). Green tea polyphenol epigallocatechin-3-gallate inhibits the IL-1 beta-induced activity and expression of cyclooxygenase-2 and nitric oxide synthase-2 in human chondrocytes. *Free Radic Biol Med 33*(8), 1097–1105.

261. Andriamanalijaona R, Kypriotou M, Bauge C, Renard E, Legendre F, Raoudi M, Boumediene K, Gatto H, Monginoux P, Pujol JP (2005). Comparative effects of 2 antioxidants, selenomethionine and epigallocatechin-gallate, on catabolic and anabolic gene expression of articular chondrocytes. *J Rheumatol 32*(10), 1958–1967.

262. Haqqi TM, Anthony DD, Gupta S, Ahmad N, Lee MS, Kumar GK, Mukhtar H (1999). Prevention of collagen-induced arthritis in mice by a polyphenolic fraction from green tea. *Proc Natl Acad Sci U S A 96*(8), 4524–4529.

263. Kulkarni RR, Patki PS, Jog VP, Gandage SG, Patwardhan B (1991). Treatment of osteoarthritis with a herbomineral formulation: A double-blind, placebo-controlled, cross-over study. *J Ethnopharmacol 33*(1–2), 91–95.

264. Kimmatkar N, Thawani V, Hingorani L, Khiyani R (2003). Efficacy and tolerability of *Boswellia serrata* extract in treatment of osteoarthritis of knee—A randomized double blind placebo controlled trial. *Phytomedicine 10*(1), 3–7.

265. Choi JH, Kim DY, Yoon JH, Youn HY, Yi JB, Rhee HI, Ryu KH, Jung K, Han CK, Kwak WJ, Cho YB (2002). Effects of SKI 306X, a new herbal agent, on proteoglycan degradation in cartilage explant culture and collagenase-induced rabbit osteoarthritis model. *Osteoarthritis Cartilage 10*(6), 471–478.

266. Kim JH, Rhee HI, Jung IH, Ryu K, Jung K, Han CK, Kwak WJ, Cho YB, Joo HJ (2005). SKI306X, an oriental herbal mixture, suppresses gastric leukotriene B4 synthesis without causing mucosal injury and the diclofenac-induced gastric lesions. *Life Sci 77*(11), 1181–1193.

267. Jung YB, Roh KJ, Jung JA, Jung K, Yoo H, Cho YB, Kwak WJ, Kim DK, Kim KH, Han CK (2001). Effect of SKI 306X, a new herbal anti-arthritic agent, in patients with osteoarthritis of the knee: A double-blind placebo controlled study. *Am J Chin Med 29*(3–4), 485–491.

268. Lung YB, Seong SC, Lee MC, Shin YU, Kim DH, Kim JM, Jung YK, Ahn JH, Seo JG, Park YS, Lee CS, Roh KJ, Han CK, Cho YB, Chang DY, Kwak WJ, Jung KO, Park BJ (2004). A four-week, randomized, double-blind trial of the efficacy and safety of SKI306X: A herbal anti-arthritic agent versus diclofenac in osteoarthritis of the knee. *Am J Chin Med 32*(2), 291–301.

269. Ryttig K, Schlamowitz PV, Warnoe O, Wilstrup F (1991). [Gitadyl versus ibuprofen in patients with osteoarthrosis. The result of a double-blind, randomized cross-over study]. *Ugeskr Laeger 153*(33), 2298–2299.

270. Teekachunhatean S, Kunanusorn P, Rojanasthien N, Sananpanich K, Pojchamarnwiputh S, Lhieochaiphunt S, Pruksakorn S (2004). Chinese herbal recipe versus diclofenac in symptomatic treatment of osteoarthritis of the knee: A randomized controlled trial. *BMC Complement Altern Med 4*(1), 19.

271. Usha PR, Naidu MUR (2004). Randomized, double-blind, parallel, placebo-controlled study of oral glucosamine, methylsulfonylmethane and their combination in osteoarthritis. *Clin Drug Invest 24*(6), 353–363.

272. Kim LS, Axelrod LJ, Howard P, Buratovich N, Waters RF (2006). Efficacy of methylsulfonylmethane (MSM) in osteoarthritis pain of the knee: A pilot clinical trial. *Osteoarthritis Cartilage 14*(3), 286–294.

273. Hughes DE, Wright KR, Uy HL, Sasaki A, Yoneda T, Roodman GD, Mundy GR, Boyce BF (1995). Bisphosphonates promote apoptosis in murine osteoclasts in vitro and in vivo. *J Bone Miner Res 10*(10), 1478–1487.

274. Catterall JB, Cawston TE (2003). Drugs in development: Bisphosphonates and metalloproteinase inhibitors. *Arthritis Res Ther 5*(1), 12–24.

275. Jung A, Bisaz S, Bartholdi P, Fleisch H (1973). Influence of pyrophosphate on the exchange of calcium and phosphate ions on hydroxyapatite. *Calcif Tissue Res 13*(1), 27–40.

276. Kinne RW, Schmidt CB, Buchner E, Hoppe R, Nurnberg E, Emmrich F (1995). Treatment of rat arthritides with clodronate-containing liposomes. *Scand J Rheumatol Suppl 101*, 91–97.

277. Barrera P, Blom A, van Lent PL, van Bloois L, Beijnen JH, van Rooijen N, de Waal Malefijt MC, van de Putte LB, Storm G, van den Berg WB (2000). Synovial macrophage depletion with clodronate-containing liposomes in rheumatoid arthritis. *Arthritis Rheum 43*(9), 1951–1959.

278. Tanahashi M, Funaba Y, Itoh M, Kawabe N, Nakadate-Matsushita T (1998). Inhibitory effects of TRK-530 on rat adjuvant arthritis. *Pharmacology 56*(5), 242–251.

279. Podworny NV, Kandel RA, Renlund RC, Grynpas MD (1999). Partial chondroprotective effect of zoledronate in a rabbit model of inflammatory arthritis. *J Rheumatol 26*(9), 1972–1982.

280. Hayami T, Pickarski M, Wesolowski GA, McLane J, Bone A, Destefano J, Rodan GA, Duong le T (2004). The role of subchondral bone remodeling in osteoarthritis: Reduction of cartilage degeneration and prevention of osteophyte formation by alendronate in the rat anterior cruciate ligament transection model. *Arthritis Rheum 50*(4), 1193–1206.

281. Myers SL, Brandt KD, Burr DB, O'Connor BL, Albrecht M (1999). Effects of a bisphosphonate on bone histomorphometry and dynamics in the canine cruciate deficiency model of osteoarthritis. *J Rheumatol 26*(12), 2645–2653.

282. Lehmann HJ, Mouritzen U, Christgau S, Cloos PA, Christiansen C (2002). Effect of bisphosphonates on cartilage turnover assessed with a newly developed assay for collagen type II degradation products. *Ann Rheum Dis 61*(6), 530–533.
283. Garnero P, Christgau S, Delmas PD (2001). The bisphosphonate zoledronate decreases type II collagen breakdown in patients with Paget's disease of bone. *Bone 28*(5), 461–464.
284. Spector TD, Conaghan PG, Buckland-Wright JC, Garnero P, Cline GA, Beary JF, Valent DJ, Meyer JM (2005). Effect of risedronate on joint structure and symptoms of knee osteoarthritis: Results of the BRISK randomized, controlled trial. *Arthritis Res Ther 7*(3), R625–633.
285. Meunier PJ (2004). [Strontium ranelate: New therapeutic agent for postmenopausal osteoporosis]. *Med Sci (Paris) 20*(6–7), 631–633.
286. Marie PJ (2006). Strontium ranelate: A physiological approach for optimizing bone formation and resorption. *Bone 38*(2, Suppl 1), S10–14.
287. Marie PJ, Hott M, Modrowski D, De Pollak C, Guillemain J, Deloffre P, Tsouderos Y (1993). An uncoupling agent containing strontium prevents bone loss by depressing bone resorption and maintaining bone formation in estrogen-deficient rats. *J Bone Miner Res 8*(5), 607–615.
288. Henrotin Y, Labasse A, Zheng SX, Galais P, Tsouderos Y, Crielaard JM, Reginster JY (2001). Strontium ranelate increases cartilage matrix formation. *J Bone Miner Res 16*(2), 299–308.
289. Spencer CM, Wilde MI (1997). Diacerein. *Drugs 53*(1), 98–106; discussion 107–108.
290. Nicolas P, Tod M, Padoin C, Petitjean O (1998). Clinical pharmacokinetics of diacerein. *Clin Pharmacokinet 35*(5), 347–359.
291. Sanchez C, Mathy-Hartert M, Deberg MA, Ficheux H, Reginster JY, Henrotin YE (2003). Effects of rhein on human articular chondrocytes in alginate beads. *Biochem Pharmacol 65*(3), 377–388.
292. Tamura T, Kosaka N, Ishiwa J, Sato T, Nagase H, Ito A (2001). Rhein, an active metabolite of diacerein, down-regulates the production of pro-matrix metalloproteinases-1, -3, -9 and -13 and up-regulates the production of tissue inhibitor of metalloproteinase-1 in cultured rabbit articular chondrocytes. *Osteoarthritis Cartilage 9*(3), 257–263.
293. Felisaz N, Boumediene K, Ghayor C, Herrouin JF, Bogdanowicz P, Galerra P, Pujol JP (1999). Stimulating effect of diacerein on TGF-beta1 and beta2 expression in articular chondrocytes cultured with and without interleukin-1. *Osteoarthritis Cartilage 7*(3), 255–264.
294. Yaron M, Shirazi I, Yaron I (1999). Anti-interleukin-1 effects of diacerein and rhein in human osteoarthritic synovial tissue and cartilage cultures. *Osteoarthritis Cartilage 7*(3), 272–280.
295. Borderie D, Hernvann A, Lemarechal H, Menkes CJ, Ekindjian O (2001). Inhibition of the nitrosothiol production of cultured osteoarthritic chondrocytes by rhein, cortisol and diclofenac. *Osteoarthritis Cartilage 9*(1), 1–6.
296. Pelletier JP, Mineau F, Fernandes JC, Duval N, Martel-Pelletier J (1998). Diacerhein and rhein reduce the interleukin 1beta stimulated inducible nitric oxide synthesis level and activity while stimulating cyclooxygenase-2 synthesis in human osteoarthritic chondrocytes. *J Rheumatol 25*(12), 2417–2424.
297. Bironaite D, Ollinger K (1997). The hepatotoxicity of rhein involves impairment of mitochondrial functions. *Chem Biol Interact 103*(1), 35–50.
298. Malterud KE, Farbrot TL, Huse AE, Sund RB (1993). Antioxidant and radical scavenging effects of anthraquinones and anthrones. *Pharmacology 47*(Suppl 1), 77–85.
299. Kay AG, Griffiths LG, Volans GN, Grahame R (1980). Preliminary experience with diacetylrhein in the treatment of osteoarthritis. *Curr Med Res Opin 6*(8), 548–551.
300. Bendele AM, Bendele RA, Hulman JF, Swann BP (1996). [Beneficial effects of treatment with diacerhein in guinea pigs with osteoarthritis]. *Rev Prat 46*(6, Spec No), S35–S39.
301. Brandt KD, Smith G, Kang SY, Myers S, O'Connor B, Albrecht M (1997). Effects of diacerhein in an accelerated canine model of osteoarthritis. *Osteoarthritis Cartilage 5*(6), 438–449.
302. Hwa SY, Burkhardt D, Little C, Ghosh P (2001). The effects of orally administered diacerein on cartilage and subchondral bone in an ovine model of osteoarthritis. *J Rheumatol 28*(4), 825–834.
303. Nguyen M, Dougados M, Berdah L, Amor B (1994). Diacerhein in the treatment of osteoarthritis of the hip. *Arthritis Rheum 37*(4), 529–536.
304. Pelletier JP, Yaron M, Haraoui B, Cohen P, Nahir MA, Choquette D, Wigler I, Rosner IA, Beaulieu AD (2000). Efficacy and safety of diacerein in osteoarthritis of the knee: A double-blind, placebo-controlled trial. The Diacerein Study Group. *Arthritis Rheum 43*(10), 2339–2348.
305. Dougados M, Nguyen M, Berdah L, Mazieres B, Vignon E, Lequesne M (2001). Evaluation of the structure-modifying effects of diacerein in hip osteoarthritis: ECHODIAH, a three-year, placebo-controlled trial. Evaluation of the Chondromodulating Effect of Diacerein in OA of the Hip. *Arthritis Rheum 44*(11), 2539–2547.
306. Greenwald RA, Moak SA, Ramamurthy NS, Golub LM (1992). Tetracyclines suppress matrix metalloproteinase activity in adjuvant arthritis and in combination with flurbiprofen, ameliorate bone damage. *J Rheumatol 19*(6), 927–938.
307. Amin AR, Attur MG, Thakker GD, Patel PD, Vyas PR, Patel RN, Patel IR, Abramson SB (1996). A novel mechanism of action of tetracyclines: Effects on nitric oxide synthases. *Proc Natl Acad Sci USA 93*(24), 14014–14019.

308. DeClerck YA, Imren S (1994). Protease inhibitors: Role and potential therapeutic use in human cancer. *Eur J Cancer 30A*(14), 2170–2180.
309. Rifkin BR, Vernillo AT, Golub LM, Ramamurthy NS (1994). Modulation of bone resorption by tetracyclines. *Ann NY Acad Sci 732*, 165–180.
310. Maragoudakis ME, Peristeris P, Missirlis E, Aletras A, Andriopoulou P, Haralabopoulos G (1994). Inhibition of angiogenesis by anthracyclines and titanocene dichloride. *Ann NY Acad Sci 732*, 280–293.
311. Ramamurthy N, Greenwald R, Moak S, Scuibba J, Goren A, Turner G, Rifkin B, Golub L (1994). CMT/Tenidap treatment inhibits temporomandibular joint destruction in adjuvant arthritic rats. *Ann NY Acad Sci 732*, 427–430.
312. Patel RN, Attur MG, Dave MN, Patel IV, Stuchin SA, Abramson SB, Amin AR (1999). A novel mechanism of action of chemically modified tetracyclines: Inhibition of COX-2-mediated prostaglandin E2 production. *J Immunol 163*(6), 3459–3467.
313. Attur MG, Patel RN, Patel PD, Abramson SB, Amin AR (1999). Tetracycline up-regulates COX-2 expression and prostaglandin E2 production independent of its effect on nitric oxide. *J Immunol 162*(6), 3160–3167.
314. Shlopov BV, Smith GN Jr, Cole AA, Hasty KA (1999). Differential patterns of response to doxycycline and transforming growth factor beta1 in the down-regulation of collagenases in osteoarthritic and normal human chondrocytes. *Arthritis Rheum 42*(4), 719–727.
315. Beekman B, Verzijl N, de Roos JA, Koopman JL, TeKoppele JM (1997). Doxyclycline inhibits collagen synthesis by bovine chondrocytes cultured in alginate. *Biochem Biophys Res Commun 237*(1), 107–110.
316. Yu LP Jr, Smith GN Jr, Brandt KD, Myers SL, O'Connor BL, Brandt DA (1992). Reduction of the severity of canine osteoarthritis by prophylactic treatment with oral doxycycline. *Arthritis Rheum 35*(10), 1150–1159.
317. Yu LP Jr, Smith GN Jr, Hasty KA, Brandt KD (1991). Doxycycline inhibits type XI collagenolytic activity of extracts from human osteoarthritic cartilage and of gelatinase. *J Rheumatol 18*(10), 1450–1452.
318. Yu LP Jr, Burr DB, Brandt KD, O'Connor BL, Rubinow A, Albrecht M (1996). Effects of oral doxycycline administration on histomorphometry and dynamics of subchondral bone in a canine model of osteoarthritis. *J Rheumatol 23*(1), 137–142.
319. Pardy CK, Matyas JR, Zernicke RF (2004). Doxycycline effects on mechanical and morphometrical properties of early- and late-stage osteoarthritic bone following anterior cruciate ligament injury. *J Appl Physiol 97*(4), 1254–1260.
320. Smith GN Jr, Yu LP Jr, Brandt KD, Capello WN (1998). Oral administration of doxycycline reduces collagenase and gelatinase activities in extracts of human osteoarthritic cartilage. *J Rheumatol 25*(3), 532–535.
321. Brandt KD, Mazzuca SA, Katz BP, Lane KA, Buckwalter KA, Yocum DE, Wolfe F, Schnitzer TJ, Moreland LW, Manzi S, Bradley JD, Sharma L, Oddis CV, Hugenberg ST, Heck LW (2005). Effects of doxycycline on progression of osteoarthritis: Results of a randomized, placebo-controlled, double-blind trial. *Arthritis Rheum 52*(7), 2015–2025.

# DEVELOPMENT OF ANTICYTOKINE AGENTS FOR RHEUMATOID ARTHRITIS

Jonathan Sherlock, Christopher D. Buckley, and Karim Raza

*Rheumatology Research Group, MRC Centre for Immune Regulation, University of Birmingham, Birmingham, United Kingdom*

## 1 INTRODUCTION

Rheumatoid arthritis (RA) is a common systemic inflammatory disease [1] characterized by joint destruction and extra-articular inflammation, resulting in severe morbidity and increased cardiovascular mortality [2]. The past decades have witnessed the gradual development of orally administered immunomodulatory treatments with methotrexate and sulfasalazine being the most frequently used nonbiological agents in the United Kingdom. Experience with these drugs has resulted in the concept of "disease-modifying" anti-rheumatic drugs (DMARDs), which can be introduced early after a diagnosis of RA has been made to reduce the rate of progression of joint damage.

Despite the success of methotrexate and other disease-modifying agents, many patients with RA fail to respond adequately to these drugs. This might be because these drugs have been used too late, at the wrong dose, or more likely because they do not adequately target the right disease processes. Recent years have witnessed the exciting development of biological therapeutics that use antibody or protein technology to target, with high specificity, particular molecules involved in the inflammatory pathway, thus disrupting the cellular communication necessary for the orchestration of inflammation. Their use in combination with traditional DMARDs has led to highly effective gold standard treatments in which clinical remission has at long last become possible [3–5].

## 2 CASE STUDY: ANTI-TNF AGENTS

Neutralization of tumor necrosis factor (TNF) was the first example of specific targeting of a cytokine in RA [6], and its enormous success has meant it is now available as part of routine clinical practice. The identification of TNF as a clinically relevant target highlights several concepts central to any discussion of the development of anticytokine agents. Though the subject of several reviews by Feldmann and Maini who pioneered this field [6, 7], the key points are summarized below.

While work by Feldmann's group demonstrated that multiple cytokines were present in the rheumatoid synovium, it was initially suspected that their number was so great and function so redundant that the neutralization of only one cytokine would have a minimal effect on inflammation [6, 7]. Interleukin-1 (IL-1) was viewed as a key molecule of interest, since it exhibited prolonged expression in cultures of rheumatoid synovial cells, in contrast to its transient expression in models of infection where the inflammation resolved [6]. The concept thus emerged that the dysregulated, chronic production of cytokines was responsible for the persistent inflammation seen in RA [6, 8]. The key observation came when an anti-TNF antibody was added to rheumatoid synovial cell cultures resulting in cessation of IL-1 production [6, 9]. When it was later demonstrated that the neutralization of TNF also led to reductions in other inflammatory molecules such as IL-6 and IL-8, the concept emerged that TNF was a master

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

regulator sitting at the apex of a hierarchy of inflammatory cytokines [6, 7]. Further evidence for the importance of TNF came from the observation that both TNF and its receptor were upregulated in biopsies of rheumatoid tissue [10], and that neutralization of TNF in mice resulted in amelioration of disease in an inflammatory arthritis model [11], whereas transgenic expression of TNF resulted in a destructive polyarthritis [12].

Subsequent clinical trials demonstrated that the chimeric anti-TNF antibody infliximab resulted in a dramatic reduction in disease activity, improvement in physical function, reduction in inflammatory markers such as C reactive protein (CRP) and cytokines such as IL-1 and IL-6, as well as an inhibition of angiogenesis and a slowing of the rate of joint erosion [13–16]. Following treatment with infliximab, synovial membrane biopsies demonstrated reduced cellularity [6], and the observation that the drug results in reduced articular retention of radio-labeled leukocytes demonstrated that neutralization of TNF-modified cell trafficking [17].

Despite the enormous success of infliximab, and other agents such as etanercept and adalimumab, which also target TNF, the use of this class of drugs is associated with an increased risk from specific infectious agents, in particular *Mycobacterium tuberculosis* [18], and is rarely associated with development of other autoimmune diseases [19]. Furthermore, a significant proportion of patients do not respond to TNF-neutralizing agents. These problems have fueled the search for a better understanding of the cellular and molecular pathogenesis of RA in order to identify alternative targets for neutralization.

## 3 NOVEL THERAPIES TARGETING CYTOKINES

Study of the rheumatoid synovium reveals it to be a highly activated tissue in which multiple cell types are present and multiple cytokines are expressed. This tissue contains activated macrophages, mast cells, fibroblasts, endothelial cells, and lymphocytes bathed in a rich cocktail of inflammatory cytokines. It is apposed to bone on which sit activated osteoclasts that actively resorb the bony surface. The expansion of these synovial cell populations is facilitated by the neovascularization that is a hallmark of the rheumatoid synovium.

Recent years have seen the identification of several new cytokines within the inflamed rheumatoid joint. We have chosen to focus our review on several representative agents that demonstrate therapeutic promise and shed particular light on pathogenic mechanisms operating in chronic inflammatory disease. For example, we will demonstrate how IL-15 has emerged as a key cytokine involved in chronicity of inflammation through the promotion of memory T-cell survival, how IL-18 allows T cells to produce cytokines independently of stimulation via the T cell receptor (TCR) and thus fosters antigen-independent inflammation, and how the IL-23/IL-17 pathway demonstrates that autoimmune inflammation proceeds by a mechanism different from that of the IL-12/IFNγ (interferon gamma) pathway classically associated with resistance to common infections.

## 4 INTERLEUKIN-15

IL-15 is a 14- to 15-kDa member of the four-α-helix bundle cytokine family [20], initially discovered as a molecule with T lymphocyte and natural killer (NK) cell activating properties and structural homology to IL-2 [21, 22]. It is produced by monocytes, macrophages, dendritic cells, epithelial cells, and fibroblasts [23–29], and its receptor shares the common $\gamma_c$ chain with the receptors for IL-2, IL-4, IL-7, IL-9, and IL-21. However, it is the unique IL-15Rα chain, to which IL-15 binds with high affinity, that is critical for its effect in vivo. Interestingly, this receptor has an unusual mechanism of action—instead of IL-15 being simply secreted from cells, it is expressed on the cell membrane bound to IL-15Rα, which presents the cytokine in *trans* to neighboring cells such as $CD8^+$ T cells and natural killer (NK) cells, which express only the IL-2/IL-15Rβ and $\gamma_c$ chain [30]. While the IL-15/IL-15Rα complex undergoes transendosomal recycling, its stability and constant reappearance on the cell membrane allows persistent activity. In keeping with this mechanism of action, coordinate expression of IL-15 and IL-15Rα on the surface of monocytes and dendritic cells is facilitated by shared promoter elements in these two genes conferring response to interferons and signals from Toll-like receptors (TLRs) [31–33]. Consistently, in vivo mouse models had demonstrated a requirement for IL-15Rα on antigen presenting cells (APCs), rather than on responding T or NK cells, before elucidation of this model [34].

IL-15 plays a key role in the promotion of NK cell [35], natural killer T (NKT) cell, and CD8 memory lymphocyte homeostasis. Thus transgenic mice overexpressing IL-15 have raised numbers of these cells, and their survival is augmented by the ability of IL-15 to inhibit IL-2-induced activation induced cell death (AICD) [36]. Conversely, IL-15- or IL-15Rα-deficient mice are lymphopaenic with reduced mature lymphocytes—particularly the CD8 subset—secondary to defective homing and proliferation [34, 37]. Despite being in the same cytokine family, IL-15 and IL-2 have

opposite effects on memory T cells, survival of which is supported by IL-15 but suppressed by IL-2 [38]. The importance of T cells in the rheumatoid joint is demonstrated by the therapeutic benefits of drugs targeting T-cell activation, such as ciclosporin, and the association of RA with certain human leukocyte antigen (HLA) alleles. Given the ability of IL-15 to maintain memory T cells and thus its potential to maintain chronic inflammation, this cytokine represents a key target for investigation in RA.

Indeed, despite being a predominantly membrane-bound cytokine, raised levels of IL-15 have been detected in serum and synovial fluid in RA patients compared to individuals with osteoarthritis (OA) [27, 39, 40] and levels have been found to rise with disease duration [41]. Moreover, IL-15 is also present in rheumatoid nodules [42]. While expression in fibroblast-like synoviocytes and endothelial cells has been observed, lining layer macrophages are the main cellular source of IL-15 in RA [27–29]. IL-15 is chemotactic to T cells and increases their transendothelial migration [27, 29, 43], thus acting as a potent agent recruiting cells of the $CD45RO^+$ memory group [44]. The importance of this function in RA in vivo has been emphasized by experiments showing that IL-15 leads to accumulation of T cells in human RA synovium engrafted into severe combined immunodeficient (SCID) mice [29], and the inflammatory infiltrate induced by injection of IL-15 into the hind foot pads of mice primed with *Corynebacterium parvum* consists of cells just over half of which express CD3 [27]. The stimulatory effects of IL-15 on memory cell numbers, promotion of their proliferation [38, 45] and inhibition of AICD [36], as well as *Bcl-2*-mediated inhibition of T-cell apoptosis [25], suggests that the high expression of this cytokine in RA could contribute to loss of tolerance and chronicity of inflammation by promoting survival of autoreactive T cells, as well as stimulating their homing to the synovium.

Moreover, as well as its effects on memory T lymphocytes, IL-15 has strong pro-inflammatory actions on cells of the innate immune system, stimulating cytokine production and the oxidative burst by macrophages and neutrophils, and granule release from the latter [46] as well as activating dendritic cells [47]. Furthermore, it activates fibroblasts to proliferate and upregulate Bcl-2 and Bcl-$X_L$ [48], and secrete cytokines and matrix metalloproteinases (MMPs) [49]. IL-15 critically links T cells with the innate immune system, since T cells activated with this cytokine induce TNF production by blood- or synovium-derived macrophages in a cell contact-dependent manner [39], and this may contribute to a positive feedback loop in which activated macrophages express IL-15, contributing to further T-cell chemoattraction, resulting in further stimulation of macrophages to produce TNF. Given the recent identification of the critical role played by IL-17 in autoimmune joint inflammation [50], it is noteworthy that IL-15 stimulates IL-17 production in a cyclosporine–sensitive way [51], and IL-17 in turn induces stromal cells [52] and macrophages [53] to produce inflammatory cytokines.

Given the newly emphasized role of B lymphocytes in RA, following from the therapeutic success of the B-cell-depleting antibody rituximab, the effects of IL-15 on B lymphocytes are of great interest. The cytokine induces their proliferation and isotype switching [54], and in animal models where IL-15 is targeted, the production of collagen-specific $IgG_{2a}$ antibodies is delayed [55]. This aspect of IL-15 biology requires further elucidation.

Animal models have confirmed the importance of IL-15 in autoimmune inflammation in vivo. In particular, in the collagen induced arthritis (CIA) model, administration of IL-15 with incomplete Freund's adjuvant during priming with type II collagen resulted in an inflammatory and erosive arthritis in DBA/1 mice, which closely resembled that in mice primed with complete Freund's adjuvant [56]. Binding of IL-15 by administration of a soluble IL-15Rα suppressed CIA, with reduced TNF production and delayed formation of collagen-specific $IgG_{2a}$ antibodies [55]. Moreover, using an antagonistic mutant IL-15/Fc fusion protein, Ferrari-Lacraz et al. have shown that this agent not only inhibits the onset of CIA but also blocks disease progression in established arthritis, reduces synovitis, bone erosion, and cartilage destruction, as well as decreasing intra-articular expression of TNF, IL-1β, IL-6, and IL-17 [57]. Of particular note, the responder frequency of autoreactive T cells was reduced not only in vitro but in vivo as assessed by carboxyfluorescein succinimidyl ester (CFSE) staining of splenocytes. Remarkably, after withdrawal of the agent, sustained clinical remission was obtained. This suggests that inhibition of the IL-15 pathway, coupled with possible FcR-mediated destruction of IL-15R expressing cells by the Fc component of the reagent, hinders the perpetuation of autoreactive T cells. The cytocidal properties of the reagent are critical, since remission was not observed with IL-15 blockade using the soluble IL-15Rα.

A recent phase I–II clinical trial of the human $IgG_1$ anti-IL-15 antibody HuMax-IL-15, demonstrated that the drug was well tolerated and had no significant effect on NK or T-lymphocyte subset numbers. Substantial clinical improvements were seen, as measured by the American College of Rheumatology criteria for 20% improvement (63% of patients), 50% improvement (38% of patients), and 70% improvement (25% of patients) [58].

Therapeutics targeting IL-15 thus show promise, and it will be of great interest to establish if subgroups of patients who are unresponsive to currently used TNF-based drugs might benefit from neutralization of IL-15.

## 5 INTERLEUKIN-18

IL-18 was discovered as a molecule that stimulates IFNγ production from splenocytes and $T_h1$ cells and is produced in the livers of mice treated with *Propionibacterium acnes* and lipopolysaccharide (LPS) [59–62]. Although its biology is strongly reminiscent of that of IL-12, IL-18 has no protein or nucleotide sequence similarity to the IL-12 family. Initial structural-based comparisons, which suggested that it is part of the IL-1 cytokine family [63], have been confirmed, and indeed IL-18 processing as well as its receptor and signal transduction system are analogous to the IL-1/IL-1R and related TLR system [64]. While this may suggest its place in an evolutionarily ancient system capable of activating innate immunity, most attention has focused on the ability of IL-18 to enhance IL-12-mediated $T_h1$ cell activation and IFNγ production.

IL-18 is produced by macrophages [62, 65, 66] and dendritic cells [67] and stimulates IFNγ production by T cells and NK cells as well as enhancing their cytotoxicity. Indeed, for maximal production of IFNγ, human T cells require both IL-12 and IL-18 [64], and in mice each cytokine upregulates the other's receptor [68, 69]. While IL-12 stimulated T-cell production of IFNγ, which usually requires TCR-mediated T-cell activation, the addition of IL-18 not only causes more IFNγ production than addition of excess IL-12 [61] but allows its production in a manner independent of TCR stimulation [70–72]. This is confirmed by the observation that cyclosporine, an inhibitor of TCR-derived signaling, does not inhibit this effect, nor does immobilized anti-CD3 enhance it [72, 73]. This may contribute to chronicity of inflammation by licensing T cells to produce pro-inflammatory cytokines independent of stimulation by their cognate antigen. Additionally, IL-18 activates a distinct "helper" subset of NK cells that produce high levels of IFNγ [74]. Moreover, the important role played by antigen-presenting cells themselves in production of IFNγ has been increasingly appreciated [75], and IL-18 is important in this process. Bone-marrow-derived macrophages produce IFNγ in response to IL-12 and IL-18 stimulation [76], as do dendritic cells [77] and B lymphocytes [78].

IL-18 is found at raised levels in the serum and synovial fluid of RA patients compared with those with OA, and ex vivo synovial fluid samples from RA patients stimulate IFNγ production in an IL-18-dependent manner [79, 80]. Macrophages purified from rheumatoid synovium are able to release IL-18 protein [80], and IL-18 expression is observed in $CD68^+$ macrophages and fibroblast-like synoviocytes, while the IL-18Rα has been demonstrated on synovial T cells and macrophages. IL-18 expression correlates with macrophage infiltration, microscopic inflammation score, erythrocyte sedimentation rate (ESR), and local production of IL-1β and TNF in the synovium [81]. IL-18 upregulates production of the inflammatory molecules granulocyte monocyte colony stimulating factor (GM-CSF) and nitric oxide and operates in positive feedback with TNF, since both IL-18 and TNF upregulate each other's production [79]. This is clinically important, since treatment with infliximab has been shown to reduce the serum concentration of IL-18 as early as 2 weeks after treatment, and the change in IL-18 correlates with the clinical response, suggesting that reduction in IL-18 may be an important component of the action of anti-TNF agents [82, 83]. IL-18 has additional prodegradative effects on articular cartilage—it is induced in chondrocytes by IL-1β and upregulates the destructive enzymes cyclooxygenase 2 (COX2) and inducible nitric oxide synthase (iNOS), and stromelysin in an IL-1β-independent manner [84], while reducing the proliferation of chondrocytes. Indeed, application of IL-18 to normal human cartilage results in the release of glycosaminoglycans.

The pro-inflammatory role of IL-18 in autoimmune joint inflammation has been confirmed in vivo by the reduced incidence and severity of CIA in *IL-18* knockout (KO) mice [85]. Moreover, whereas immunization of DBA/1 mice with collagen and incomplete Freund's adjuvant (IFA) is usually unable to induce CIA, concomitant administration of IL-18 induces CIA, which is histologically indistinguishable from that induced with complete Freund's adjuvant (CFA), and this treatment with IL-18 augments expression of TNF and IL-6 [86]. These studies have been extended by the development or reagents to block the in vivo function of IL-18. Single administration of an anti-IL-18 antibody or daily injections of IL-18BP both resulted in a delay in arthritis and prevention of cartilage destruction; however, only the antibody caused significant reduction in joint inflammation and cell infiltration, emphasizing the importance of the mechanism of action of IL-18 neutralizing reagents in the outcome [87]. A different approach has used intra-articular adenoviral-mediated gene expression of murine IL-18BP, which reduces both severity and incidence of arthritis and synovial inflammation and prevents bone erosion and cartilage destruction [88].

Given the recent realization that CIA is critically dependant on IL-23-induced $T_h17$ cells, rather than IL-12-induced $T_h1$ cells between which there is a mutual antagonism [50], it is important to re-evaluate the con-

tribution of the direct pro-inflammatory effects of IL-18 and those mediated via IFNγ, since the latter might be expected to be less important, at least in the mouse system. Indeed, it has been observed that IL-18 treatment of mice, while inducing TNF and IL-6, caused less IFNγ production than did treatment with IL-12 [86], emphasizing the dominant role of IL-12 in the induction of $T_h1$ cytokine responses. The importance of the IFNγ pathway in joint inflammation is again questioned by the observation that despite the high level of expression of IL-18 in RA synovial membranes, there is a relative lack of IFNγ and IL-12 expression [89]. This may be secondary to antagonism by a dominant IL-17 pathway in the joint. Moreover, it has been emphasized that IL-18 is not an exclusive stimulator of $T_h1$ cells, as in the absence of IL-12, it can promote $T_h2$ differentiation and allergic inflammation [64]. It will, therefore, be of great interest to determine any role it plays in $T_h17$ lymphocyte biology.

Thus the importance of IL-18 in RA may result from its direct IFNγ-independent effects, such as its ability to increase chemokine expression [90], upregulate Fas ligand-mediated cytotoxicity of NK [91] and T cells [92], and induce nuclear translocation of nuclear factor of kappa light polypeptide gene enhancer in B-cells (NF-κB) [93]. IL-18 directly upregulates TNF production by splenic macrophages [86] and stimulates the release of IL-1β and IL-8 from $CD14^+$ cells, in a process dependent on its ability to stimulate TNF production from T and NK cells [90].

IL-18 plays two roles in vascular biology of relevance to its therapeutic targeting in rheumatoid arthritis. Given the importance of angiogenesis in RA [94–96], the characterization of IL-18 as an angiogenic mediator [97] that can stimulate synthesis of MMPs and promote matrix breakdown [84] suggests that this is another mechanism whereby its inhibition may ameliorate joint inflammation. Furthermore, IL-18 also has been shown to be a major mediator of atherosclerosis and to contribute to plaque instability [98–101]. In view of the increased atherosclerotic risk associated with RA and other autoimmune diseases [102], it will be of great interest to determine if this is reduced by targeting of the IL-18 pathway.

One significant concern with the use of therapies targeting IL-18 is the potential to increase susceptibility to infection in the recipient. In particular, IL-18 plays a key role in immunity to multiple pathogens, including bacteria, viruses, fungi, and protozoa [64].

## 6 INTERLEUKIN-6

IL-6 was initially discovered as a B-cell differentiation factor [103, 104], but it is now appreciated to be a pleiotropic cytokine that plays key roles in hematopoiesis, the acute phase response and inflammation [105, 106]. It induces T-cell proliferation, macrophage and cytotoxic T-cell differentiation, and immunoglobulin production by B lymphocytes. Administration of IL-6 to humans causes fever and leukocytosis [107], and transgenic mice overexpressing human IL-6 develop plasmacytosis, hypergammaglobulinaemia, an increase in acute phase reactants and low albumin, and increased bone marrow megakaryocytes [108], all of which are very reminiscent of the findings in patients with RA. Moreover, IL-6 in the presence of soluble IL-6R triggers differentiation of osteoclasts, and can be responsible for the joint erosion and osteoporosis associated with RA [109, 110]. IL-6 is also a pivotal cytokine in triggering increased vascular endothelial growth factor (VEGF) production, in synergy with TNF or IL-1β [111], thus supporting the angiogenesis that is crucial for the development of synovial hyperplasia.

The multiple roles of IL-6 in the pathogenesis of RA suggested by all of these observations is supported by the demonstration that IL-6 promotor polymorphisms are associated with disease activity. In particular, a certain allele predisposes to worse disease as measured by an increased number of swollen and tender joints, raised disease activity score (DAS28), and a high ESR [112]. Furthermore, serum levels of IL-6 are closely correlated with clinical measures of RA disease activity [113] and radiological joint damage [113–116], and a correlation of synovial fluid IL-6 level with the histological characteristics of chronic synovitis suggests an important local role for IL-6 [115].

Strong support for the importance of IL-6 in vivo comes from the demonstration that IL-6 KO mice are resistant to CIA [117, 118], and the absence of IL-6 prevents the spontaneous autoimmune joint inflammation normally observed in SKG mice. Moreover, the resistance of the latter mice to joint inflammation is greater than in TNF or IL-1 KO mice [119]. Conversely, a mutation in mice of the cytoplasmic tail of the gp130 component of the IL-6 receptor complex, which interfers with binding of a negative regulator of signaling Src homology domain protein tyrosine phosphatase 2 (SHP2) and thus allows increased signaling, is associated with a spontaneous arthritis that displays great similarity to human RA on histological and radiological grounds [120]. It has been noted that antibodies to IL-6 result in the formation of monomeric immune complexes [121], and indeed efforts to develop therapeutics have concentrated on agents targeting the IL-6R. Treatment of mice with an anti-IL-6R antibody reduced the incidence and severity of CIA, and inhibited both the proliferative response of lymphocytes to collagen and the development of collagen-specific antibodies [122].

The humanized antibody tocilizumab has been developed to target the IL-6R in RA [123]. A phase II multicenter, placebo-controlled trial has demonstrated a dose-dependent improvement in disease activity, with 57% of patients receiving 4 mg/kg and 78% of those receiving 8 mg/kg showing an improvement greater than or equal to 20% according to the American College of Rheumatology (ACR) criteria [124]. Furthermore, serum markers of bone formation were increased, and urinary markers of bone absorption were decreased, suggesting that IL-6 blockade resulted in improved bone metabolism, which may prevent the osteoporosis associated with RA. A subsequent study has demonstrated that tocilizumab delays radiographic progression of joint disease in RA patients in a manner superior to DMARD treatment [125].

The benefits of blockade of IL-6 are much more widespread than those accrued at the level of the joint, with multiple improvements in systemic features of RA. As expected from the known contribution of IL-6 to the acute-phase response, treatment with tocilizumab is associated with improvement in the platelet count and in hemoglobin, fibrinogen, and albumin levels. Of interest is the reduction in serum amyloid A levels following treatment, since this therapy may induce regression of amyloid deposits, a serious complication of RA that can be difficult to treat [126]. The improvement in hemoglobin level following IL-6 neutralization may result from a reduced expression of the suppressor of cytokine signalling (SOCS) molecule, which is induced by chronic stimulation of the IL-6R to dampen signaling, but that also dampens signaling via the erythropoietin receptor, which shares the janus kinase (Jak)–signal transducer and activator of transcription (STAT) pathway with the IL-6R [127]. Another mechanism responsible for the improvement in anemia following treatment with tocilizumab may be reduction in hepcidin, a molecule produced by the liver upon stimulation by IL-6 and that downregulates iron absorption in the gut [128]. Moreover, serum levels of VEGF return to normal in patients treated with tocilizumab [111], and given the importance of angiogenesis in synovial inflammation, this may play a role in the efficacy of the drug.

All therapeutics neutralizing cytokines carry the potential to increase the risk of infection secondary to immunosuppression. In particular, infliximab has been associated with tuberculosis [18]. In this regard, animal models suggest that therapeutics targeting the IL-6 pathway may have an advantage, since in contrast to mice inoculated with *M. tuberculosis* and treated with anti-TNF antibodies, those treated with antibodies to IL-6R survived. This is in keeping with the key role played by TNF in granuloma formation, and the increased risk of tuberculosis with infliximab may be related to reduced IFNγ [129, 130].

## 7 INTERLEUKIN-23/INTERLEUKIN-17 PATHWAY

The discovery of IL-23 at the turn of the millennium has heralded extraordinarily rapid advances in the understanding of tissue-specific autoimmune disease. IL-23 was identified by structure-based searching of the genome for IL-6 cytokine family members and is a heterodimeric cytokine consisting of a p40 subunit, which it shares with IL-12, and a unique p19 subunit [131, 132]. Like IL-12, IL-23 is produced by activated macrophages and dendritic cells, but its main targets are activated rather than naïve T cells [131, 132] as well as myeloid cells [133]. Its receptor consists of IL-12Rβ1, which it shares with IL-12, and a unique component IL-23R [134], which confers specificity to IL-23.

Classical investigation of the IL-12 cytokine pathway in inflammatory arthritis and other autoimmune diseases used genetic or serological targeting of the p40 subunit of IL-12. Indeed anti-p40 antibodies reached clinical trials to test the efficacy of IL-12 blockade before it was discovered that they also can result in neutralization of IL-23. With the discovery of IL-23, it became obvious that the importance attributed to IL-12 with anti-p40 blockade had been greatly overstated Use of mice deficient in the unique subunits of each cytokine, rather than the common p40 subunit, has shown that IL-23, rather than IL-12, is the critical cytokine in CIA [50]. Furthermore, its role as a master regulator of autoimmune inflammation beyond the joint has been demonstrated by its critical position in the control of autoimmunity in the central nervous system (CNS) [133, 135], bowel [136–140], and skin [141, 142].

Traditionally, experiments targeting IL-12p40 led to the paradigm that tissue-specific autoimmune disease was dependent on IL-12-driven Th1 lymphocyte polarization and production of IFNγ. However, many observations, particularly in humans, were not consistent with this model, and in particular the role of IFNγ in autoimmunity was much more complex than the model predicted [143]. Indeed, when administered 24 h before inducing arthritis with adjuvant, IFNγ enhanced arthritis. However, when given after immunization, arthritis was actually suppressed [144]. Daily administration of IFNγ to Lewis rats after induction of adjuvant arthritis resulted in suppression of paw swelling, regression of bone lesions, reduction in splenomegaly, and reduced systemic inflammatory response as measured by ESR [145]. Conversely, loss of IFNγ [146] or its receptor [147] resulted in increased severity

of CIA as well as other tissue-specific autoimmune diseases.

Writing only a few years after the discovery of Th1 and Th2 lymphocyte subsets by Mosmann et al. [148], Jacob's group had already highlighted conflicting reports addressing the importance of IFNγ in autoimmunity and suggested that this might reflect the activity of additional distinct T-cell subsets [144]. While the existence of further subsets had also been predicted in Mosmann and co-workers' seminal paper [148], and suggested following characterization of IL-17 producing human T cells [149], it is only recently that the details of these subsets, the transcriptional master regulators that regulate their production and the cytokines that they produce, has come to light.

Although the predominant action of IL-23 on mouse memory T cells had been appreciated since its discovery [132], a key breakthrough came with the finding that it stimulated memory T cells to produce IL-17 [150]. It is now clear that IL-23 drives autoimmune pathology by stimulation of a unique T-lymphocyte subset now termed $T_h17$ [151] and has a particular role in the maintenance of this subset [131, 152, 153]. In contrast to $T_h1$ cells, which express genes important in cytotoxicity, such as *FasL*, TNF-related apoptosis-inducing ligand (TRAIL), interferon-γ (IFNγ) and granzyme, $T_h17$ cells regulate local tissue inflammation by producing IL-6, TNF, IL-22, and IL-17 [135]. In keeping with the broad expression of its receptor, IL-17 itself acts on multiple cell types to cause the release of a diverse range of pro-inflammatory cytokines [154]. Notably, it induces stromal, epithelial, and endothelial cells and monocytes [154] to express TNF and IL-1β [53] and fibroblasts to secrete IL-6 and IL-8 [155, 156]. It enhances recruitment of neutrophils by inducing secretion of the neutrophil chemokine IL-8 from epithelial and endothelial cells and rheumatoid synovium [154] and stimulates neutrophil production through release of granulocyte colony stimulating factor (G-CSF) [157]. Thus IL-17 emerges as a critical molecule in orchestrating autoimmune inflammation. Its action on cells usually forming compartmental boundaries, such as endothelial or epithelial cells, places it in a prime position to mediate the breakdown of these barriers and stimulate the inflammatory infiltration of joints and other organs. While research on IL-17 has focused predominantly on its pathogenic effects, the beneficial roles played by IL-17 include the ability to respond to serious injury or infection [158] by rapidly mounting a massive inflammatory response through the recruitment of neutrophils.

Raised IL-17 levels have been detected in CIA [50], and adenoviral-mediated overexpression of IL-17 in joints exacerbates CIA, causing increased joint destruction as assessed histologically and radiographically [159]. While this is associated with upregulation of IL-1β, neutralization of IL-1β is without effect. IL-17 also exacerbates streptococcal cell-wall-induced arthritis, converting it into a severe chronic arthritis with increased damage to cartilage, and notably can substitute for the usually obligate role of IL-1 in proteoglycan depletion [160]. Conversely, genetic ablation of IL-17 ameliorates CIA [161], as does neutralization of IL-17 [160, 162, 163], which is associated with reduced inflammation, decreased receptor activator of nuclear factor kappa-B ligand (RANKL) expression, and bone erosion [162]. Mice lacking IL-12p35 have increased $T_h17$ cells and exacerbated CIA associated with reduced IFNγ, whereas mice lacking IL-23p19 are completely resistant to CIA and have raised IFNγ expressing Th1 cells [50]. This not only demonstrates the key role of IL-23 and IL-17 in joint pathology and the reciprocal nature of the Th1 and Th17 pathways but also demonstrates the paradoxical protection from autoimmune inflammation mediated by IFNγ, which is consistent with the historical observations detailed above. The opposing roles of IFNγ and IL-17 extend beyond simply *inflammation* in the joint, since IFNγ itself inhibits osteoclastogenesis and hence bone erosion, whereas $T_h17$ cells are directly osteoclastogenic through expression of IL-17 and trigger bone loss even in the absence of autoimmunity [164]. Furthermore, $T_h17$ cells express RANKL [164] and IL-17 induces expression of RANKL in osteoclastogenesis-supporting cells [165], demonstrating multiple routes by which IL-17 can cause the erosive pathology associated with RA.

The importance of IL-17 in human pathology has been confirmed by human studies that have shown raised levels of IL-17 in synovial fluid, serum, and synovial biopsies of RA patients compared with OA patients [165–168]. Remarkably, some time before the current interest in $T_h17$ cells, immunohistochemistry of synovial biopsies from RA patients demonstrated memory $CD4^+$ T cells ($CD45RO^+$), which co-stained for IL-17 [165], suggesting that $T_h17$ cells are present in the rheumatoid synovium. Joosten et al. have also demonstrated IL-$17^+$ T cells in the synovium [81], and cloning of T cells from the rheumatoid synovium demonstrates the presence of IL-$17^+$ IFN$\gamma^+$ double-positive T cells [149]. It will be very informative to define the identity of T lymphocytes associated with destruction in RA.

Human studies demonstrate multiple aggressive functions of IL-17 in the joint. IL-17 promotes formation of osteoclasts in vitro, as does the culture media of RA synovial tissues [165]. The ability of the latter to stimulate differentiation of osteoclasts is reduced by 80% when IL-17 is neutralized, emphasizing the key role played by IL-17 in osteoclastogenesis. IL-17 also causes increased IL-6 production and collagen

destruction by RA synovial explants, increased bone resorption, and decreased formation by RA bone explants, effects that are exacerbated by the addition of IL-1 [172]. Blocking endogenous IL-17 in these models reduces bone destruction. Of significance, IL-17 simulates a positive feedback loop that may perpetuate autoimmune inflammatory joint disease. Thus RA T cells induce fibroblasts to produce the major inflammatory cytokines TNF, IL-15 and IL-18, which in turn further stimulate the T cells to release IL-17 and IFNγ to further stimulate the fibroblasts [168]. IL-17 then synergies with TNF to cause cartilage destruction [173]. The relevance of these findings is emphasized by the observation that messenger ribonucleic acid (mRNA) levels of IL-1β, TNF, IL-10, and IL-17 are predictive of the rate of progression of joint damage in RA, and that IL-17 and TNF are also synergistic in this model [174]. As previously noted [175], many of these experiments have unfortunately not included measurements of IL-23; however, one report demonstrates raised IL-23p19 in the sera and synovial fluid of RA patients, compared with OA or healthy controls [176]. Since IL-17 was found to induce IL-23p19 expression in fibroblasts [176], this again highlights the effects of positive feedback in Th17 mediated joint inflammation.

Further potential cytokine targets have emerged from studying the control of T cells producing IL-17, and key differences have emerged between murine and human $T_h17$ cells. Whereas, in the mouse, IL-23 is important for IL-17 cell expansion and commitment in keeping with the early observations by Oppmann et al. that IL-23 acts on murine memory cells [132] and those of Aggarwal showing that IL-23 augments IL-17 production from memory but not naïve T cells [150], Th17 cells are produced from murine naïve cells not by the action of IL-23 itself but by transforming growth factor β (TGFβ) and IL-6 [153, 177, 178]. While TGFβ has often been considered an anti-inflammatory cytokine, and is pivotal in polarization of $T_{reg}$ lymphocytes, it is now suggested that in the context of inflammatory stimuli, and thus IL-6 production, T cells may be diverted from a $T_{reg}$ fate into Th17 cells and thus that the two populations have reciprocal developmental pathways. Similarly, while IL-2 is vital for $T_{reg}$ cells, it potently impairs generation of Th17 cells [179], again suggesting dichotomous polarization. Recently, the action of IL-6 has been shown to be mediated by IL-21, which is produced by Th17 cells and acts in an autocrine feedback mechanism to stabilize Th17 polarization [180]. The combination of IL-6 and TGFβ does not polarize *human* $T_h17$ cells, which instead undergo differentiation in response to IL-23 itself in combination with IL-1β [171], in keeping with the original observations of Oppmann and coworkers that IL-23 can act on human naïve T cells [132]. While several papers have confirmed the suppressive effects on TGFβ on $T_h17$ development in humans [181, 182], IL-6 has emerged as a strong stimulator of $T_h17$ cells in the absence of TGFβ [181] and in combination with IL-1β. This raises the question as to whether the success of anti-IL-6R antibodies in RA is contributed to by inhibition of $T_h17$ cell polarization. Very recent experiments have identified key elements of the innate immune system responsible for polarization of Th17 cells, including the intracellular molecule nucleotide oligomerization domain 2 (NOD2) [183] and Toll-like receptors [182]. The field of human $T_h17$ cell biology is developing rapidly, and it is likely that work over the next few years will clarify the conditions under which these cells develop in humans.

A key question is which of the cytokines in this new pathway represent the best therapeutic target for RA, an issue that has not yet been directly addressed in animal models. IL-17 is promising, particularly given the fact that neutralization of this cytokine will directly target the T-cell subset responsible for autoimmune inflammation. This view is supported by the success of therapeutics targeting IL-17 in animal models. Yet, while IL-17 is important, the upstream cytokine IL-23 has additional actions beyond induction of IL-17. Indeed early work concentrated on its ability to directly activate myeloid cells; IL-23 directly induces peritoneal macrophages to upregulate TNF and IL-1β and acts via IFNγ to stimulate expression of a further set of pro-inflammatory molecules [133], and injection of IL-23 into the skin results in what is essentially a wound-healing response [141]. Furthermore, a T-cell-independent model of inflammatory colitis induced with anti-CD40 antibodies in recombinase activating gene (RAG) deficient mice demonstrates that IL-23 is critical for local tissue pathology despite the absence of T cells and, interestingly, shows local production of IL-17 in the colon—most likely by innate immune cells—that is dependent on IL-23 as it is abrogated in p19 and p40 KO mice [138]. Several observations also underline the vital role of IL-23 in the continued operation of $T_h17$ lymphocytes [140]. For example, while $T_h17$ cells can develop in the absence of IL-23 upon infection with the intestinal pathogen *Citrobacter rodentium*, full protection of the host requires IL-23 [178]. Similarly, neutralization of IL-23p19 prevents the inflammatory bowel disease (IBD) induced by transfer for $T_h17$ cells and can also treat established disease. Thus IL-23 plays a key role in the optimal expansion of $T_h17$ cells but has independent pro-inflammatory effects on the innate immune system, thus representing a critical target for therapeutic use. Moreover, in combination with IL-18, IL-23 can induce TCR-independent production of IL-17 from T cells [184], emphasizing the important role therapeutics targeting IL-18 may have in the regulation of this pathway. Cytokines in addition to IL-23 may play impor-

tant roles in maintaining the function of $T_h17$ cells. Indeed it has been noted that IL-1R1 KO mice have a defective Th17 response to immunization, and T cells deficient in IL-1R1 do not produce IL-17 when cultured with IL-23 [152].

It is hoped that targeting the IL-23 pathway will not be accompanied by substantial infective complications. Indeed studies from France have demonstrated that patients with combined defects in the IL-12 and IL-23 pathway due to mutations in the shared p40 ligand or IL-12Rβ1 receptor subunits have susceptibility to only two classes of pathogens—salmonella and mycobacteria, similar to those with mutations in the IFNγ pathway [185, 186]. While p40-deficient mice are susceptible to many pathogens, use of specific IL-12- and IL-23-deficient mice has shown that this is largely due to the defect in the IL-12/IFNγ pathway, with a contribution from IL-23 only being detected in the absence of IL-12. Additionally, infections that appear to require the presence of the IL-23/IL-17 pathway in mice, such as *Klebsiella* [187] and systemic *Candida albicans* [188], are not prevalent in humans with mutations in p40, again providing an encouraging prediction of low infective risk with IL-23 blockade and emphasizing the difference between mouse and human IL-23/Th17 biology.

## 8 OTHER THERAPEUTIC STRATEGIES

While we have focused on cytokine-based therapeutics, multiple other targets exist against which biological agents can be developed and that we have not been able to cover in this review. For example, great success has recently been attained by reducing numbers of B cells with the anti-CD20 agent rituximab [189], promoting an increased appreciation of the role of humoral immunity in RA. This also raises the question of whether the best strategies will target cytokines themselves or their cellular source. However, it is notable that so far other cellular targets have not had promising results [6]. Increased understanding of the signaling cascades inside cells has given rise to additional therapeutic targets, as has increased mechanistic knowledge of the chemokines responsible for the regulation of cellular trafficking in autoimmune inflammation [190, 191]. The critical role of angiogenesis [190] in promoting a rich vascular supply for inflammatory cells and molecules to reach the normally minimally vascular synovium is frequently emphasized; however, to date this process has not been specifically targeted. Furthermore, greater knowledge of osteoimmunology [192, 193] has allowed the key interactions between inflammatory cells and osteoclasts to be understood and has identified critical molecules, such as RANKL, necessary for osteoclast activation, many of which represent promising therapeutic targets.

A theme that emerges is the enormous complexity and multilevel interaction between all of these systems—cytokines, chemokines, co-stimulatory molecule expression, angiogenesis, and osteoclast activation. Thus neutralization of many of the cytokines we have discussed has affected, for example, the expression of co-stimulatory molecules or angiogenesis as well as expression of other cytokines. Clearly, the development of optimal clinical therapeutics must take into account these complex networks of interactions. It may be that certain patients require more than one biological therapeutic, and thus greater understanding of the use of anticytokine reagents in combination is needed. Furthermore, it must be highlighted that the cytokine patterns of very early rheumatoid arthritis are distinct from later cytokine expression profiles [194], emphasising that therapeutic approaches may need to differ depending of the phase of the patient's disease during which treatment is being administered.

Given the success of hematologists in using combinations of cellular markers, cytokine levels and genetic profiles to predict responses to treatment and to clarify prognosis [195–201], it would be rewarding if a greater understanding of cytokine biology by rheumatologists not only lead to the development of further therapeutics but also to algorithms that would predict responses to specific drugs.

Not only are developments in the identification and targeting of cytokines vital, but the detailed pharmacology of the therapeutics themselves is of critical importance. Already it is striking to note that patients not responding to one TNF-neutralizing drug may respond well to another, emphasizing that details of the manner of targeting (by antibody or decoy receptor), and route of administration (subcutaneous or intravenous) is as important as the molecule that is targeted. Additionally, the use of large molecules, such as antibodies and decoy receptors, to target cytokines necessitates parenteral administration. Knowledge of pathways downstream of the relevant receptors may allow development of small-molecule agents that will inhibit the ability of cytokines to trigger cellular responses, with the advantage that these agents could be orally active. While novel methods of neutralizing TNF have included liposomal delivery of small interfering RNA in animal models [202], complete cure of CIA necessitates ongoing administration, making an orally administered drug an attractive option.

The study of cytokine biology has proved to be highly rewarding from a therapeutic perspective. The proof of concept demonstrated by the success of anti-TNF agents, combined with the experimental ease with which cytokines can be investigated and manipulated, has opened the future of rheumatology to exciting new conceptual and therapeutic prospects.

## REFERENCES


1. Ollier WE, Harrison B, Symmons D (2001). What is the natural history of rheumatoid arthritis? *Best Pract Res Clin Rheumatol 15*, 27–48.
2. Stevens RJ, Douglas KM, Saratzis AN, Kitas GD (2005). Inflammation and atherosclerosis in rheumatoid arthritis. *Expert Rev Mol Med 7*, 1–24.
3. Goekoop-Ruiterman YP, de Vries-Bouwstra JK, Allaart CF, et al. (2005). Clinical and radiographic outcomes of four different treatment strategies in patients with early rheumatoid arthritis (the best study): A randomized, controlled trial. *Arthritis Rheum 52*, 3381–3390.
4. Quinn MA, Emery P (2005). Potential for altering rheumatoid arthritis outcome. *Rheum Dis Clin North Am 31*, 763–772.
5. Quinn MA, Conaghan PG, O'Connor PJ, et al. (2005). Very early treatment with infliximab in addition to methotrexate in early, poor-prognosis rheumatoid arthritis reduces magnetic resonance imaging evidence of synovitis and damage, with sustained benefit after infliximab withdrawal: Results from a twelve-month randomized, double-blind, placebo-controlled trial. *Arthritis Rheum 52*, 27–35.
6. Feldmann M, Maini RN (2003). Lasker clinical medical research award. TNF defined as a therapeutic target for rheumatoid arthritis and other autoimmune diseases. *Nat Med 9*, 1245–1250.
7. Feldmann M, Brennan FM, Maini RN (1996). Role of cytokines in rheumatoid arthritis. *Annu Rev Immunol 14*, 397–440.
8. Buchan G, Barrett K, Turner M, Chantry D, Maini RN, Feldmann M (1988). Interleukin-1 and tumour necrosis factor MRNA expression in rheumatoid arthritis: Prolonged production of IL-1 alpha. *Clin Exp Immunol 73*, 449–455.
9. Brennan FM, Chantry D, Jackson A, Maini R, Feldmann M (1989). Inhibitory effect of TNF alpha antibodies on synovial cell interleukin-1 production in rheumatoid arthritis. *Lancet 2*, 244–247.
10. Chu CQ, Field M, Feldmann M, Maini RN (1991). Localization of tumor necrosis factor alpha in synovial tissues and at the cartilage-pannus junction in patients with rheumatoid arthritis. *Arthritis Rheum 34*, 1125–1132.
11. Williams RO, Feldmann M, Maini RN (1992). Anti-tumor necrosis factor ameliorates joint disease in murine collagen-induced arthritis. *Proc Natl Acad Sci U S A 89*, 9784–9788.
12. Keffer J, Probert L, Cazlaris H, et al. (1991). Transgenic mice expressing human tumour necrosis factor: A predictive genetic model of arthritis. *Embo J 10*, 4025–4031.
13. Charles P, Elliott MJ, Davis D, et al. (1999). Regulation of cytokines, cytokine inhibitors, and acute-phase proteins following anti-TNF-alpha therapy in rheumatoid arthritis. *J Immunol 163*, 1521–1528.
14. Lipsky PE, van der Heijde DM, St Clair EW, et al. (2000). Infliximab and methotrexate in the treatment of rheumatoid arthritis. Anti-tumor necrosis factor trial in rheumatoid arthritis with concomitant therapy study group. *N Engl J Med 343*, 1594–1602.
15. Maini RN, Breedveld FC, Kalden JR, et al. (2004). Sustained improvement over two years in physical function, structural damage, and signs and symptoms among patients with rheumatoid arthritis treated with infliximab and methotrexate. *Arthritis Rheum 50*, 1051–1065.
16. Paleolog EM, Young S, Stark AC, McCloskey RV, Feldmann M, Maini RN (1998). Modulation of angiogenic vascular endothelial growth factor by tumor necrosis factor alpha and interleukin-1 in rheumatoid arthritis. *Arthritis Rheum 41*, 1258–1265.
17. Taylor PC, Peters AM, Paleolog E, et al. (2000). Reduction of chemokine levels and leukocyte traffic to joints by tumor necrosis factor alpha blockade in patients with rheumatoid arthritis. *Arthritis Rheum 43*, 38–47.
18. Keane J, Gershon S, Wise RP, et al. (2001). Tuberculosis associated with infliximab, a tumor necrosis factor alpha-neutralizing agent. *N Engl J Med 345*, 1098–1104.
19. Ramos-Casals M, Brito-Zeron P, Munoz S, et al. (2007). Autoimmune diseases induced by TNF-targeted therapies: Analysis of 233 cases. *Medicine (Baltimore) 86*, 242–251.
20. Waldmann TA, Tagaya Y (1999). The multifaceted regulation of interleukin-15 expression and the role of this cytokine in NK cell differentiation and host response to intracellular pathogens. *Annu Rev Immunol 17*, 19–49.
21. Burton JD, Bamford RN, Peters C, et al. (1994). A lymphokine, provisionally designated interleukin T and produced by a human adult T-cell leukemia line, stimulates T-cell proliferation and the induction of lymphokine-activated killer cells. *Proc Natl Acad Sci U S A 91*, 4935–4939.
22. Grabstein KH, Eisenman J, Shanebeck K, et al. (1994). Cloning of a T cell growth factor that interacts with the beta chain of the interleukin-2 receptor. *Science 264*, 965–968.
23. Bamford RN, Battiata AP, Burton JD, Sharma H, Waldmann TA (1996). Interleukin (IL) 15/IL-T Production by the adult T-cell leukemia cell line hut-102 is associated with a human T-cell lymphotrophic virus type I region/IL-15 fusion message that lacks many upstream augs that normally attenuates IL-15 MRNA translation. *Proc Natl Acad Sci U S A 93*, 2897–2902.
24. Carson WE, Giri JG, Lindemann MJ, et al. (1994). Interleukin (IL) 15 is a novel cytokine that activates human natural killer cells via components of the IL-2 receptor. *J Exp Med 180*, 1395–1403.
25. Waldmann TA (2005). Targeting the interleukin-15 system in rheumatoid arthritis. *Arthritis Rheum 52*, 2585–2588.
26. Doherty TM, Seder RA, Sher A (1996). Induction and regulation of IL-15 expression in murine macrophages. *J Immunol 156*, 735–741.

27. McInnes IB, al-Mughales J, Field M, et al. (1996). The role of interleukin-15 in T-cell migration and activation in rheumatoid arthritis. *Nat Med 2*, 175–182.
28. Thurkow EW, van der Heijden IM, Breedveld FC, et al. (1997). Increased expression of IL-15 in the synovium of patients with rheumatoid arthritis compared with patients with *yersinia*-induced arthritis and osteoarthritis. *J Pathol 181*, 444–450.
29. Oppenheimer-Marks N, Brezinschek RI, Mohamadzadeh M, Vita R, Lipsky PE (1998). Interleukin 15 is produced by endothelial cells and increases the transendothelial migration of T cells in vitro and in the SCID mouse-human rheumatoid arthritis model in vivo. *J Clin Invest 101*, 1261–1272.
30. Dubois S, Mariner J, Waldmann TA, Tagaya Y (2002). IL-15ralpha recycles and presents IL-15 in trans to neighboring cells. *Immunity 17*, 537–547.
31. Azimi N, Nagai M, Jacobson S, Waldmann TA (2001). IL-15 plays a major role in the persistence of tax-specific CD8 cells in ham/tsp patients. *Proc Natl Acad Sci U S A 98*, 14559–14564.
32. Mariner JM, Mamane Y, Hiscott J, Waldmann TA, Azimi N (2002). IFN regulatory factor 4 participates in the human T cell lymphotropic virus type I-mediated activation of the IL-15 receptor alpha promoter. *J Immunol 168*, 5667–5674.
33. Mariner JM, Lantz V, Waldmann TA, Azimi N (2001). Human T cell lymphotropic virus type I tax activates IL-15r alpha gene expression through an NF-kappa *b* site. *J Immunol 166*, 2602–2609.
34. Lodolce JP, Boone DL, Chai S, et al. (1998). Il-15 receptor maintains lymphoid homeostasis by supporting lymphocyte homing and proliferation. *Immunity 9*, 669–676.
35. Becknell B, Caligiuri MA (2005). Interleukin-2, interleukin-15, and their roles in human natural killer cells. *Adv Immunol 86*, 209–239.
36. Marks-Konczalik J, Dubois S, Losi JM, et al. *(*2000). Il-2-induced activation-induced cell death is inhibited in IL-15 transgenic mice. *Proc Natl Acad Sci U S A 97*, 11445–11450.
37. Kennedy MK, Glaccum M, Brown SN, et al. (2000). Reversible defects in natural killer and memory CD8 T cell lineages in interleukin 15-deficient mice. *J Exp Med 191*, 771–780.
38. Ku CC, Murakami M, Sakamoto A, Kappler J, Marrack P (2000). Control of homeostasis of CD8+ memory T cells by opposing cytokines. *Science 288*, 675–678.
39. McInnes IB, Leung BP, Sturrock RD, Field M, Liew FY (1997). Interleukin-15 mediates T cell-dependent regulation of tumor necrosis factor-alpha production in rheumatoid arthritis. *Nat Med 3*, 189–195.
40. Petrovic-Rackov L, Pejnovic N (2006). Clinical significance of IL-18, IL-15, IL-12 and TNF-alpha measurement in rheumatoid arthritis. *Clin Rheumatol 25*, 448–452.
41. Gonzalez-Alvaro I, Ortiz AM, Garcia-Vicuna R, Balsa A, Pascual-Salcedo D, Laffon AI (2003). ncreased serum levels of interleukin-15 in rheumatoid arthritis with long-term disease. *Clin Exp Rheumatol 21*, 639–642.
42. Hessian PA, Highton J, Kean A, Sun CK, Chin M (2003). Cytokine profile of the rheumatoid nodule suggests that it is a Th1 granuloma. *Arthritis Rheum 48*, 334–338.
43. Wilkinson PC, Liew FY (1995). Chemoattraction of human blood T lymphocytes by interleukin-15. *J Exp Med 181*, 1255–1259.
44. al-Mughales J, Blyth TH, Hunter JA, Wilkinson PC (1996). The chemoattractant activity of rheumatoid synovial fluid for human lymphocytes is due to multiple cytokines. *Clin Exp Immunol 106*, 230–236.
45. Zhang X, Sun S, Hwang I, Tough DF, Sprent J (1998). Potent and selective stimulation of memory-phenotype CD8+ T cells in vivo by IL-15. *Immunity 8*, 591–599.
46. Connell L, McInnes IB (2006). New cytokine targets in inflammatory rheumatic diseases. *Best Pract Res Clin Rheumatol 20*, 865–878.
47. Ohteki T, Suzue K, Maki C, Ota T, Koyasu S (2001). Critical role of IL-15-IL-15r for antigen-presenting cell functions in the innate immune response. *Nat Immunol 2*, 1138–1143.
48. Kurowska M, Rudnicka W, Kontny E, et al. (2002). Fibroblast-like synoviocytes from rheumatoid arthritis patients express functional IL-15 receptor complex: Endogenous IL-15 in autocrine fashion enhances cell proliferation and expression of bcl-x(l) and bcl-2. *J Immunol 169*, 1760–1767.
49. McInnes IB, Gracie JA (2004). Interleukin-15: A new cytokine target for the treatment of inflammatory diseases. *Curr Opin Pharmacol 4*, 392–397.
50. Murphy CA, Langrish CL, Chen Y, et al. (2003). Divergent pro- and antiinflammatory roles for IL-23 and IL-12 in joint autoimmune inflammation. *J Exp Med 198*, 1951–1957.
51. Ziolkowska M, Koc A, Luszczykiewicz G, et al. (2000). High levels of IL-17 in rheumatoid arthritis patients: IL-15 triggers in vitro IL-17 production via cyclosporin a-sensitive mechanism. *J Immunol 164*, 2832–2838.
52. Fossiez F, Djossou O, Chomarat P, et al. (1996). T cell interleukin-17 induces stromal cells to produce proinflammatory and hematopoietic cytokines. *J Exp Med 183*, 2593–2603.
53. Jovanovic DV, Di Battista JA, Martel-Pelletier J, et al. (1998). IL-17 stimulates the production and expression of proinflammatory cytokines, IL-beta and TNF-alpha, by human macrophages. *J Immunol 160*, 3513–3521.
54. Armitage RJ, Macduff BM, Eisenman J, Paxton R, Grabstein KH (1995). IL-15 has stimulatory activity for the induction of B cell proliferation and differentiation. *J Immunol 154*, 483–490.
55. Ruchatz H, Leung BP, Wei XQ, McInnes IB, Liew FY (1998). Soluble IL-15 receptor alpha-chain

administration prevents murine collagen-induced arthritis: A role for IL-15 in development of antigen-induced immunopathology. *J Immunol 160*, 5654–5660.

56. Liew FY, McInnes IB (2002). The role of innate mediators in inflammatory response. *Mol Immunol 38*, 887–890.
57. Ferrari-Lacraz S, Zanelli E, Neuberg M, et al. (2004). Targeting IL-15 receptor-bearing cells with an antagonist mutant IL-15/fc protein prevents disease development and progression in murine collagen-induced arthritis. *J Immunol 173*, 5818–5826.
58. Baslund B, Tvede N, Danneskiold-Samsoe B, et al. (2005). Targeting interleukin-15 in patients with rheumatoid arthritis: A proof-of-concept study. *Arthritis Rheum 52*, 2686–2692.
59. Kohno K, Kataoka J, Ohtsuki T, et al. (1997). IFN-gamma-inducing factor (IGIF) is a costimulatory factor on the activation of Th1 but not Th2 cells and exerts its effect independently of IL-12. *J Immunol 158*, 1541–1550.
60. Matsui K, Yoshimoto T, Tsutsui H, et al. (1997). Propionibacterium acnes treatment diminishes CD4+ NK1.1+ T cells but induces type I T cells in the liver by induction of IL-12 and IL-18 production from kupffer cells. *J Immunol 159*, 97–106.
61. Okamura H, Nagata K, Komatsu T, et al. (1995). A novel costimulatory factor for gamma interferon induction found in the livers of mice causes endotoxic shock. *Infect Immun 63*, 3966–3972.
62. Okamura H, Tsutsi H, Komatsu T, et al. (1996). Cloning of a new cytokine that induces IFN-gamma production by T cells. *Nature 378*, 88–91.
63. Bazan JF, Timans JC, Kastelein RA (1996). A newly defined interleukin-1? *Nature 379*, 591.
64. Nakanishi K, Yoshimoto T, Tsutsui H, Okamura H (2001). Interleukin-18 regulates both Th1 and Th2 responses. *Annu Rev Immunol 19*, 423–474.
65. Sanchez-Bueno A, Verkhusha V, Tanaka Y, Takikawa O, Yoshida R (1996). Interferon-gamma-dependent expression of inducible nitric oxide synthase, interleukin-12, and interferon-gamma-inducing factor in macrophages elicited by allografted tumor cells. *Biochem Biophys Res Commun 224*, 555–563.
66. Shibata Y, Foster LA, Kurimoto M, et al. (1998). Immunoregulatory roles of IL-10 in innate immunity: IL-10 inhibits macrophage production of IFN-gamma-inducing factors but enhances NK cell production of IFN-gamma. *J Immunol 161*, 4283–4288.
67. Stoll S, Jonuleit H, Schmitt E, et al. (1998). Production of functional IL-18 by different subtypes of murine and human dendritic cells (DC): DC-derived IL-18 enhances IL-12-dependent Th1 development. *Eur J Immunol 28*, 3231–3239.
68. Okamura H, Kashiwamura S, Tsutsui H, Yoshimoto T, Nakanishi K (1998). Regulation of interferon-gamma production by IL-12 and IL-18. *Curr Opin Immunol 10*, 259–264.
69. Xu D, Chan WL, Leung BP, et al. (1998). Selective expression and functions of interleukin 18 receptor on T helper (th) type 1 but not Th2 cells. *J Exp Med 188*, 1485–1492.
70. Ahn HJ, Maruo S, Tomura M, et al. (1997). A mechanism underlying synergy between IL-12 and IFN-gamma-inducing factor in enhanced production of IFN-gamma. *J Immunol 159*, 2125–2131.
71. Tominaga K, Yoshimoto T, Torigoe K, et al. (2000). IL-12 synergizes with IL-18 or IL-1beta for IFN-gamma production from human T cells. *Int Immunol 12*, 151–160.
72. Yang J, Murphy TL, Ouyang W, Murphy KM (1999). Induction of interferon-gamma production in Th1 CD4+ T cells: Evidence for two distinct pathways for promoter activation. *Eur J Immunol 29*, 548–555.
73. Yoshimoto T, Takeda K, Tanaka T, et al. (1998). IL-12 up-regulates IL-18 receptor expression on T cells, Th1 cells, and B cells: Synergism with IL-18 for IFN-gamma production. *J Immunol 161*, 3400–3407.
74. Mailliard RB, Alber SM, Shen H, et al. (2005). IL-18-induced CD83+CCR7+ NK helper cells. *J Exp Med 202*, 941–953.
75. Frucht DM, Fukao T, Bogdan C, Schindler H, O'Shea JJ, Koyasu S (2001). IFN-gamma production by antigen-presenting cells: Mechanisms emerge. *Trends Immunol 22*, 556–560.
76. Munder M, Mallo M, Eichmann K, Modolell M (1998). Murine macrophages secrete interferon gamma upon combined stimulation with interleukin (IL)-12 and IL-18: A novel pathway of autocrine macrophage activation. *J Exp Med 187*, 2103–2108.
77. Fukao T, Matsuda S, Koyasu S (2000). Synergistic effects of IL-4 and IL-18 on IL-12-dependent IFN-gamma production by dendritic cells. *J Immunol 164*, 64–71.
78. Yoshimoto T, Okamura H, Tagawa YI, Iwakura Y, Nakanishi K (1997). Interleukin 18 together with interleukin 12 inhibits IgE production by induction of interferon-gamma production from activated B cells. *Proc Natl Acad Sci U S A 94*, 3948–3953.
79. Gracie JA, Forsey RJ, Chan WL, et al. (1999). A proinflammatory role for IL-18 in rheumatoid arthritis. *J Clin Invest 104*, 1393–1401.
80. Yamamura M, Kawashima M, Taniai M, et al. (2001). Interferon-gamma-inducing activity of interleukin-18 in the joint with rheumatoid arthritis. *Arthritis Rheum 44*, 275–285.
81. Joosten LA, Radstake TR, Lubberts E, et al. (2003). Association of interleukin-18 expression with enhanced levels of both interleukin-1beta and tumor necrosis factor alpha in knee synovial tissue of patients with rheumatoid arthritis. *Arthritis Rheum 48*, 339–347.
82. van Oosterhout M, Levarht EW, Sont JK, Huizinga TW, Toes RE, van Laar JM (2005). Clinical efficacy of infliximab plus methotrexate in DMARD naive and DMARD refractory rheumatoid arthritis is associated with decreased synovial expression of TNF alpha and IL18 but not cxcl12. *Ann Rheum Dis 64*, 537–543.
83. Pittoni V, Bombardieri M, Spinelli FR, et al. (2002). Anti-tumour necrosis factor (TNF) alpha treatment of rheu-

matoid arthritis (infliximab) selectively down regulates the production of interleukin (IL) 18 but not of IL12 and IL13. *Ann Rheum Dis 61*, 723–725.

84. Olee T, Hashimoto S, Quach J, Lotz M (1999). Il-18 is produced by articular chondrocytes and induces proinflammatory and catabolic responses. *J Immunol 162*, 1096–1100.
85. Wei XQ, Leung BP, Arthur HM, McInnes IB, Liew FY (2001). Reduced incidence and severity of collagen-induced arthritis in mice lacking IL-18. *J Immunol 166*, 517–521.
86. Leung BP, McInnes IB, Esfandiari E, Wei XQ, Liew FY (2000). Combined effects of IL-12 and IL-18 on the induction of collagen-induced arthritis. *J Immunol 164*, 6495–6502.
87. Plater-Zyberk C, Joosten LA, Helsen MM, et al. (2001). Therapeutic effect of neutralizing endogenous IL-18 activity in the collagen-induced model of arthritis. *J Clin Invest 108*, 1825–1832.
88. Smeets RL, van de Loo FA, Arntz OJ, Bennink MB, Joosten LA, van den Berg WB (2003). Adenoviral delivery of IL-18 binding protein C ameliorates collagen-induced arthritis in mice. *Gene Ther 10*, 1004–1011.
89. McInnes IB, Gracie JA, Leung BP, Wei XQ, Liew FY (2000). Interleukin 18: A pleiotropic participant in chronic inflammation. *Immunol Today 21*, 312–315.
90. Puren AJ, Fantuzzi G, Gu Y, Su MS, Dinarello CA (1998). Interleukin-18 (IFNgamma-inducing factor) induces IL-8 and IL-1beta via TNFalpha production from non-CD14+ human blood mononuclear cells. *J Clin Invest 101*, 711–721.
91. Tsutsui H, Nakanishi K, Matsui K, et al. (1996). IFN-gamma-inducing factor up-regulates Fas ligand-mediated cytotoxic activity of murine natural killer cell clones. *J Immunol 157*, 3967–3973.
92. Dao T, Ohashi K, Kayano T, Kurimoto M, Okamura H (1996). Interferon-gamma-inducing factor, a novel cytokine, enhances Fas ligand-mediated cytotoxicity of murine T helper 1 cells. *Cell Immunol 173*, 230–235.
93. Matsumoto S, Tsuji-Takayama K, Aizawa Y, et al. (1997). Interleukin-18 activates NF-kappaB in murine T helper type 1 cells. *Biochem Biophys Res Commun 234*, 454–457.
94. Colville-Nash PR, Scott DL (1992). Angiogenesis and rheumatoid arthritis: Pathogenic and therapeutic implications. *Ann Rheum Dis 51*, 919–925.
95. Paleolog EM, Miotla JM (1998). Angiogenesis in arthritis: Role in disease pathogenesis and as a potential therapeutic target. *Angiogenesis 2*, 295–307.
96. Szekanecz Z, Szegedi G, Koch AE (1998). Angiogenesis in rheumatoid arthritis: Pathogenic and clinical significance. *J Investig Med 46*, 27–41.
97. Park CC, Morel JC, Amin MA, Connors MA, Harlow LA, Koch AE (2001). Evidence of IL-18 as a novel angiogenic mediator. *J Immunol 167*, 1644–1653.
98. Mallat Z, Corbaz A, Scoazec A, et al. (2001). Expression of interleukin-18 in human atherosclerotic plaques and relation to plaque instability. *Circulation 104*, 1598–1603.
99. Mallat Z, Corbaz A, Scoazec A, et al. (2001). Interleukin-18/interleukin-18 binding protein signaling modulates atherosclerotic lesion development and stability. *Circ Res 89*, E41–45.
100. Tiret L, Godefroy T, Lubos E, et al. (2005). Genetic analysis of the interleukin-18 system highlights the role of the interleukin-18 gene in cardiovascular disease. *Circulation 112*, 643–650.
101. Whitman SC, Ravisankar P, Daugherty A (2002). Interleukin-18 enhances atherosclerosis in apolipoprotein e(-/-) mice through release of interferon-gamma. *Circ Res 90*, E34–38.
102. Sherer Y, Shoenfeld Y (2006). Mechanisms of disease: Atherosclerosis in autoimmune diseases. *Nat Clin Pract Rheumatol 2*, 99–106.
103. Hirano T, Yasukawa K, Harada H, et al. (1986). Complementary DNA for a novel human interleukin (bsf-2) that induces B lymphocytes to produce immunoglobulin. *Nature 324*, 73–76.
104. Teranishi T, Hirano T, Arima N, Onoue K (1982). Human helper T cell factor(s) (THF). II. Induction of IgG production in B lymphoblastoid cell lines and identification of T cell-replacing factor- (TRF) like factor(s). *J Immunol 128*, 1903–1908.
105. Hirano T (1998). Interleukin 6 and its receptor: Ten years later. *Int Rev Immunol 16*, 249–284.
106. Van Snick J (1990). Interleukin-6: An overview. *Annu Rev Immunol 8*, 253–278.
107. Weber J, Yang JC, Topalian SL, et al. (1993). Phase I trial of subcutaneous interleukin-6 in patients with advanced malignancies. *J Clin Oncol 11*, 499–506.
108. Suematsu S, Matsuda T, Aozasa K, et al. (1989). IgG1 plasmacytosis in interleukin 6 transgenic mice. *Proc Natl Acad Sci U S A 86*, 7547–7551.
109. Kotake S, Sato K, Kim KJ, et al. (1996). Interleukin-6 and soluble interleukin-6 receptors in the synovial fluids from rheumatoid arthritis patients are responsible for osteoclast-like cell formation. *J Bone Miner Res 11*, 88–95.
110. Tamura T, Udagawa N, Takahashi N, et al. (1993). Soluble interleukin-6 receptor triggers osteoclast formation by interleukin 6. *Proc Natl Acad Sci U S A 90*, 11924–11928.
111. Nakahara H, Song J, Sugimoto M, et al. (2003). Anti-interleukin-6 receptor antibody therapy reduces vascular endothelial growth factor production in rheumatoid arthritis. *Arthritis Rheum 48*, 1521–1529.
112. Pawlik A, Wrzesniewska J, Florczak M, Gawronska-Szklarz B, Herczynska M (2005). Il-6 promoter polymorphism in patients with rheumatoid arthritis. *Scand J Rheumatol 34*, 109–113.
113. Madhok R, Crilly A, Watson J, Capell HA (1993). Serum interleukin 6 levels in rheumatoid arthritis: Correlations with clinical and laboratory indices of disease activity. *Ann Rheum Dis 52*, 232–234.
114. Dasgupta B, Corkill M, Kirkham B, Gibson T, Panayi G (1992). Serial estimation of interleukin 6 as a measure of

systemic disease in rheumatoid arthritis. *J Rheumatol 19*, 22–25.

115. Sack U, Kinne RW, Marx T, Heppt P, Bender S, Emmrich F (1993). Interleukin-6 in synovial fluid is closely associated with chronic synovitis in rheumatoid arthritis. *Rheumatol Int 13*, 45–51.
116. Uson J, Balsa A, Pascual-Salcedo D, et al. (1997). Soluble interleukin 6 (IL-6) receptor and IL-6 levels in serum and synovial fluid of patients with different arthropathies. *J Rheumatol 24*, 2069–2075.
117. Alonzi T, Fattori E, Lazzaro D, et al. (1998). Interleukin 6 is required for the development of collagen-induced arthritis. *J Exp Med 187*, 461–468.
118. Sasai M, Saeki Y, Ohshima S, et al. (1999). Delayed onset and reduced severity of collagen-induced arthritis in interleukin-6-deficient mice. *Arthritis Rheum 42*, 1635–1643.
119. Hata H, Sakaguchi N, Yoshitomi H, et al. (2004). Distinct contribution of IL-6, TNF-alpha, IL-1, and IL-10 to T cell-mediated spontaneous autoimmune arthritis in mice. *J Clin Invest 114*, 582–588.
120. Atsumi T, Ishihara K, Kamimura D, et al. (2002). A point mutation of tyr-759 in interleukin 6 family cytokine receptor subunit gp130 causes autoimmune arthritis. *J Exp Med 196*, 979–990.
121. Montero-Julian FA, Klein B, Gautherot E, Brailly H (1995). Pharmacokinetic study of anti-interleukin-6 (IL-6) therapy with monoclonal antibodies: Enhancement of IL-6 clearance by cocktails of anti-IL-6 antibodies. *Blood 85*, 917–924.
122. Takagi N, Mihara M, Moriya Y, et al. (1998). Blockage of interleukin-6 receptor ameliorates joint disease in murine collagen-induced arthritis. *Arthritis Rheum 41*, 2117–2121.
123. Sato K, Tsuchiya M, Saldanha J, et al. (1993). Reshaping a human antibody to inhibit the interleukin 6-dependent tumor cell growth. *Cancer Res 53*, 851–856.
124. Nishimoto N, Yoshizaki K, Miyasaka N, et al. (2004). Treatment of rheumatoid arthritis with humanized anti-interleukin-6 receptor antibody: A multicenter, double-blind, placebo-controlled trial. *Arthritis Rheum 50*, 1761–1769.
125. Nishimoto N, Hashimoto J, Miyasaka N, et al. (2007). Study of active controlled monotherapy used for rheumatoid arthritis, an IL-6 inhibitor (samurai): Evidence of clinical and radiographic benefit from an X ray reader-blinded randomised controlled trial of tocilizumab. *Ann Rheum Dis 66*, 1162–1167.
126. Okuda Y, Takasugi K (2006). Successful use of a humanized anti-interleukin-6 receptor antibody, tocilizumab, to treat amyloid A amyloidosis complicating juvenile idiopathic arthritis. *Arthritis Rheum 54*, 2997–3000.
127. Sasaki A, Yasukawa H, Shouda T, Kitamura T, Dikic I, Yoshimura A (2000). Cis3/SOCS-3 suppresses erythropoietin (epo) signaling by binding the epo receptor and jak2. *J Biol Chem 275*, 29338–29347.
128. Nemeth E, Rivera S, Gabayan V, et al. (2004). IL-6 mediates hypoferremia of inflammation by inducing the synthesis of the iron regulatory hormone hepcidin. *J Clin Invest 113*, 1271–1276.
129. Nishimoto N, Kishimoto T (2006). Interleukin 6: From bench to bedside. *Nat Clin Pract Rheumatol 2*, 619–626.
130. Wallis RS (2007). Reactivation of latent tuberculosis by TNF blockade: The role of interferon gamma. *J Investig Dermatol Symp Proc 12*, 16–21.
131. Kastelein RA, Hunter CA, Cua DJ (2007). Discovery and biology of IL-23 and IL-27: Related but functionally distinct regulators of inflammation. *Annu Rev Immunol 25*, 221–242.
132. Oppmann B, Lesley R, Blom B, et al. (2000). Novel p19 protein engages IL-12p40 to form a cytokine, IL-23, with biological activities similar as well as distinct from IL-12. *Immunity 13*, 715–725.
133. Cua DJ, Sherlock J, Chen Y, et al. (2003). Interleukin-23 rather than interleukin-12 is the critical cytokine for autoimmune inflammation of the brain. *Nature 421*, 744–748.
134. Parham C, Chirica M, Timans J, et al. (2002). A receptor for the heterodimeric cytokine IL-23 is composed of IL-12rbeta1 and a novel cytokine receptor subunit, IL-23r. *J Immunol 168*, 5699–5708.
135. Langrish CL, Chen Y, Blumenschein WM, et al. (2005). IL-23 drives a pathogenic T cell population that induces autoimmune inflammation. *J Exp Med 201*, 233–240.
136. Hue S, Ahern P, Buonocore S, et al. (2006). Interleukin-23 drives innate and T cell-mediated intestinal inflammation. *J Exp Med 203*, 2473–2483.
137. Kullberg MC, Jankovic D, Feng CG, et al. (2006). IL-23 plays a key role in helicobacter hepaticus-induced T cell-dependent colitis. *J Exp Med 203*, 2485–2494.
138. Uhlig HH, McKenzie BS, Hue S, et al. (2006). Differential activity of IL-12 and IL-23 in mucosal and systemic innate immune pathology. *Immunity 25*, 309–318.
139. Yen D, Cheung J, Scheerens H, et al. (2006). IL-23 is essential for T cell-mediated colitis and promotes inflammation via IL-17 and IL-6. *J Clin Invest 116*, 1310–1316.
140. Elson CO, Cong Y, Weaver CT, et al. (2007). Monoclonal anti-interleukin 23 reverses active colitis in a T cell-mediated model in mice. *Gastroenterology 132*, 2359–2370.
141. Chan JR, Blumenschein W, Murphy E, et al. (2006). IL-23 stimulates epidermal hyperplasia via TNF and IL-20R2-dependent mechanisms with implications for psoriasis pathogenesis. *J Exp Med 203*, 2577–2587.
142. Zheng Y, Danilenko DM, Valdez P, et al. (2007). Interleukin-22, a T(h)17 cytokine, mediates IL-23-induced dermal inflammation and acanthosis. *Nature 445*, 648–651.
143. Steinman L (2007). A brief history of T(h)17, the first major revision in the T(h)1/T(h)2 hypothesis of T cell-mediated tissue damage. *Nat Med 13*, 139–145.
144. Jacob CO, Holoshitz J, Van der Meide P, Strober S, McDevitt HO (1989). Heterogeneous effects of

IFN-gamma in adjuvant arthritis. *J Immunol 142*, 1500–1505.

145. Zamvil S, Nelson P, Trotter J, et al. (1985). T-cell clones specific for myelin basic protein induce chronic relapsing paralysis and demyelination. *Nature 317*, 355–358.
146. Guedez YB, Whittington KB, Clayton JL, et al. (2001). Genetic ablation of interferon-gamma up-regulates interleukin-1beta expression and enables the elicitation of collagen-induced arthritis in a nonsusceptible mouse strain. *Arthritis Rheum 44*, 2413–2424.
147. Vermeire K, Heremans H, Vandeputte M, Huang S, Billiau A, Matthys P (1997). Accelerated collagen-induced arthritis in ifn-gamma receptor-deficient mice. *J Immunol 158*, 5507–5513.
148. Mosmann TR, Cherwinski H, Bond MW, Giedlin MA, Coffman RL (1986). *Two types of murine helper T cell clone*. I. Definition according to profiles of lymphokine activities and secreted proteins. *J Immunol 136*, 2348–2357.
149. Aarvak T, Chabaud M, Miossec P, Natvig JB (1999). IL-17 is produced by some proinflammatory Th1/Th0 cells but not by Th2 cells. *J Immunol 162*, 1246–1251.
150. Aggarwal S, Ghilardi N, Xie MH, de Sauvage FJ, Gurney AL (2003). Interleukin-23 promotes a distinct CD4 T cell activation state characterized by the production of interleukin-17. *J Biol Chem 278*, 1910–1914.
151. Furuzawa-Carballeda J, Vargas-Rojas MI, Cabral AR (2007). Autoimmune inflammation from the Th17 perspective. *Autoimmun Rev 6*, 169–175.
152. Sutton C, Brereton C, Keogh B, Mills KH, Lavelle EC (2006). A crucial role for interleukin (IL)-1 in the induction of IL-17-producing T cells that mediate autoimmune encephalomyelitis. *J Exp Med 203*, 1685–1691.
153. Veldhoen M, Hocking RJ, Atkins CJ, Locksley RM, Stockinger B (2006). TGFbeta in the context of an inflammatory cytokine milieu supports de novo differentiation of IL-17-producing T cells. *Immunity 24*, 179–189.
154. Kolls JK, Linden A (2004). Interleukin-17 family members and inflammation. *Immunity 21*, 467–476.
155. Yao Z, Fanslow WC, Seldin MF, et al. (1995). Herpesvirus saimiri encodes a new cytokine, IL-17, which binds to a novel cytokine receptor. *Immunity 3*, 811–821.
156. Yao Z, Painter SL, Fanslow WC, et al. (1995). Human IL-17: A novel cytokine derived from T cells. *J Immunol 155*, 5483–5486.
157. Jones CE, Chan K (2002). Interleukin-17 stimulates the expression of interleukin-8, growth-related oncogene-alpha, and granulocyte-colony-stimulating factor by human airway epithelial cells. *Am J Respir Cell Mol Biol 26*, 748–753.
158. Cua DJ, Kastelein RA (2006). TGF-beta, a "double agent" in the immune pathology war. *Nat Immunol 7*, 557–559.
159. Lubberts E, Joosten LA, Oppers B, et al. (2001). IL-1-independent role of IL-17 in synovial inflammation and joint destruction during collagen-induced arthritis. *J Immunol 167*, 1004–1013.
160. Koenders MI, Lubberts E, Oppers-Walgreen B, et al. (2005). Induction of cartilage damage by overexpression of T cell interleukin-17a in experimental arthritis in mice deficient in interleukin-1. *Arthritis Rheum 52*, 975–983.
161. Nakae S, Nambu A, Sudo K, Iwakura Y (2003). Suppression of immune induction of collagen-induced arthritis in IL-17-deficient mice. *J Immunol 171*, 6173–6177.
162. Koenders MI, Lubberts E, Oppers-Walgreen B, et al. (2005). Blocking of interleukin-17 during reactivation of experimental arthritis prevents joint inflammation and bone erosion by decreasing rankl and interleukin-1. *Am J Pathol 167*, 141–149.
163. Lubberts E, Koenders MI, Oppers-Walgreen B, et al. (2004). Treatment with a neutralizing anti-murine interleukin-17 antibody after the onset of collagen-induced arthritis reduces joint inflammation, cartilage destruction, and bone erosion. *Arthritis Rheum 50*, 650–659.
164. Sato K, Suematsu A, Okamoto K, et al. (2006). Th17 functions as an osteoclastogenic helper T cell subset that links T cell activation and bone destruction. *J Exp Med 203*, 2673–2682.
165. Kotake S, Udagawa N, Takahashi N, et al. (1999). IL-17 in synovial fluids from patients with rheumatoid arthritis is a potent stimulator of osteoclastogenesis. *J Clin Invest 103*, 1345–1352.
166. Honorati MC, Meliconi R, Pulsatelli L, Cane S, Frizziero L, Facchini A (2001). High in vivo expression of interleukin-17 receptor in synovial endothelial cells and chondrocytes from arthritis patients. *Rheumatology (Oxford) 40*, 522–527.
167. Hwang SY, Kim HY (2005). Expression of IL-17 homologs and their receptors in the synovial cells of rheumatoid arthritis patients. *Mol Cells 19*, 180–184.
168. Cho ML, Yoon CH, Hwang SY, et al. (2004). Effector function of type II collagen-stimulated T cells from rheumatoid arthritis patients: Cross-talk between T cells and synovial fibroblasts. *Arthritis Rheum 50*, 776–784.
169. Acosta-Rodriguez EV, Rivino L, Geginat J, et al. (2007). Surface phenotype and antigenic specificity of human interleukin 17-producing T helper memory cells. *Nat Immunol 8*, 639–646.
170. Annunziato F, Cosmi L, Santarlasci V, et al. (2007). Phenotypic and functional features of human Th17 cells. *J Exp Med 204*, 1849–1861.
171. Wilson NJ, Boniface K, Chan JR, et al. (2007). Development, cytokine profile and function of human interleukin 17-producing helper T cells. *Nat Immunol 8*, 950–957.
172. Chabaud M, Lubberts E, Joosten L, van Den Berg W, Miossec P (2001). IL-17 derived from juxta-articular bone and synovium contributes to joint degradation in rheumatoid arthritis. *Arthritis Res 3*, 168–177.

173. Van Bezooijen RL, Van Der Wee-Pals L, Papapoulos SE, Lowik CW (2002). Interleukin 17 synergises with tumour necrosis factor alpha to induce cartilage destruction in vitro. *Ann Rheum Dis 61*, 870–876.
174. Kirkham BW, Lassere MN, Edmonds JP, et al. (2006). Synovial membrane cytokine expression is predictive of joint damage progression in rheumatoid arthritis: A two-year prospective study (the damage study cohort). *Arthritis Rheum 54*, 1122–1131.
175. Kikly K, Liu L, Na S, Sedgwick JD (2006). The IL-23/Th(17) axis: Therapeutic targets for autoimmune inflammation. *Curr Opin Immunol 18*, 670–675.
176. Kim HR, Cho ML, Kim KW, et al. (2007). Up-regulation of IL-23p19 expression in rheumatoid arthritis synovial fibroblasts by IL-17 through pi3-kinase-, NF-kappaB- and p38 MAPK-dependent signalling pathways. *Rheumatology (Oxford) 46*, 57–64.
177. Bettelli E, Carrier Y, Gao W, et al. (2006). Reciprocal developmental pathways for the generation of pathogenic effector Th17 and regulatory T cells. *Nature 441*, 235–238.
178. Mangan PR, Harrington LE, O'Quinn DB, et al. (2006). Transforming growth factor-beta induces development of the T(h)17 lineage. *Nature 441*, 231–234.
179. Laurence A, Tato CM, Davidson TS, et al. (2007). Interleukin-2 signaling via stat5 constrains T helper 17 cell generation. *Immunity 26*, 371–381.
180. Nurieva R, Yang XO, Martinez G, et al. (2007). Essential autocrine regulation by IL-21 in the generation of inflammatory T cells. *Nature 448*, 480–483.
181. Acosta-Rodriguez EV, Napolitani G, Lanzavecchia A, Sallusto F (2007). Interleukins 1beta and 6 but not transforming growth factor-beta are essential for the differentiation of interleukin 17-producing human T helper cells. *Nat Immunol 8*, 942–949.
182. Evans HG, Suddason T, Jackson I, Taams LS, Lord GM (2007). Optimal induction of T helper 17 cells in humans requires T cell receptor ligation in the context of Toll-like receptor-activated monocytes. *Proc Natl Acad Sci U S A 104*, 17034–17039.
183. van Beelen AJ, Zelinkova Z, Taanman-Kueter EW, et al. (2007). Stimulation of the intracellular bacterial sensor NOD2 programs dendritic cells to promote interleukin-17 production in human memory T cells. *Immunity 27*, 660–669.
184. Weaver CT, Hatton RD, Mangan PR, Harrington LE (2007). IL-17 family cytokines and the expanding diversity of effector t cell lineages. *Annu Rev Immunol 25*, 821–852.
185. Fieschi C, Casanova JL (2003). The role of interleukin-12 in human infectious diseases: Only a faint signature. *Eur J Immunol 33*, 1461–1464.
186. Novelli F, Casanova JL (2004). The role of IL-12, IL-23 and IFN-gamma in immunity to viruses. *Cytokine Growth Factor Rev 15*, 367–377.
187. Happel KI, Dubin PJ, Zheng M, et al. (2005). Divergent roles of IL-23 and IL-12 in host defense against *Klebsiella pneumoniae*. *J Exp Med 202*, 761–769.
188. Huang W, Na L, Fidel PL, Schwarzenberger P (2004). Requirement of interleukin-17a for systemic anti-*Candida albicans* host defense in mice. *J Infect Dis 190*, 624–631.
189. Atzeni F, Doria A, Maurizio T, Sarzi-Puttini P (2007). What is the role of rituximab in the treatment of rheumatoid arthritis? *Autoimmun Rev 6*, 553–558.
190. McInnes IB, Schett G (2007). Cytokines in the pathogenesis of rheumatoid arthritis. *Nat Rev Immunol 7*, 429–442.
191. Tarner IH, Muller-Ladner U, Gay S (2007). Emerging targets of biologic therapies for rheumatoid arthritis. *Nat Clin Pract Rheumatol 3*, 336–345.
192. Sato K, Takayanagi H (2006). Osteoclasts, rheumatoid arthritis, and osteoimmunology. *Curr Opin Rheumatol 18*, 419–426.
193. Takayanagi H (2007). Osteoimmunology: Shared mechanisms and crosstalk between the immune and bone systems. *Nat Rev Immunol 7*, 292–304.
194. Raza K, Falciani F, Curnow SJ, et al. (2005). Early rheumatoid arthritis is characterized by a distinct and transient synovial fluid cytokine profile of T cell and stromal cell origin. *Arthritis Res Ther 7*, R784–795.
195. Dave SS, Wright G, Tan B, et al. (2004). Prediction of survival in follicular lymphoma based on molecular features of tumor-infiltrating immune cells. *N Engl J Med 351*, 2159–2169.
196. Harjunpaa A, Taskinen M, Nykter M, et al. (2006). Differential gene expression in non-malignant tumour microenvironment is associated with outcome in follicular lymphoma patients treated with rituximab and chop. *Br J Haematol 135*, 33–42.
197. Montserrat E (2006). New prognostic markers in cll. *Hematol Am Soc Hematol Educ Prog* 279–284.
198. Morabito F, Damle RN, Deaglio S, Keating M, Ferrarini M, Chiorazzi N (2006). The CD38 ectoenzyme family: Advances in basic science and clinical practice. *Mol Med 12*, 342–344.
199. Saito B, Shiozawa E, Yamochi-Onizuka T, et al. (2004). Efficacy of rituximab plus chemotherapy in follicular lymphoma depends on ki-67 expression. *Pathol Int 54*, 667–674.
200. Seliem RM, Freeman JK, Steingart RH, Hasserjian RP (2006). Immunophenotypic changes and clinical outcome in B-cell lymphomas treated with rituximab. *Appl Immunohistochem Mol Morphol 14*, 18–23.
201. Sfiridaki A, Miyakis S, Tsirakis G, et al. (2005). Systemic levels of interleukin-6 and matrix metalloproteinase-9 in patients with multiple myeloma may be useful as prognostic indexes of bone disease. *Clin Chem Lab Med 43*, 934–938.
202. Khoury M, Louis-Plence P, Escriou V, et al. (2006). Efficient new cationic liposome formulation for systemic delivery of small interfering rna silencing tumor necrosis factor alpha in experimental arthritis. *Arthritis Rheum 54*, 1867–1877.

# OVERVIEW OF CNS NEUROPHARMACEUTICALS

Tauser Roxana-Georgiana[1] and Antohe Dan-Stefan[2]
[1]*Medicinal Chemistry Department, Faculty of Pharmacy,* [2]*Department of Neuroanatomy, Faculty of Medicine*
*University of Medicine and Pharmacy "Gr.T.Popa", Iasi, Romania*

## 1 INTRODUCTION

The advances in the molecular neurosciences during "the decade of the brain" in conjunction with new genomic and proteomic sciences have radically changed the paradigm of central nervous system (CNS) drug design and development. The classical, chemistry-based small-molecule drug discovery has been broadened by the twenty-first century biology-driven paradigm of large-molecule neuropharmaceuticals: peptides, recombinant proteins, monoclonal antibodies, antisense oligonucleotides, and gene therapy [1].

An interdisciplinary approach focusing on the chemistry–biology interface would facilitate a more rapid and efficient discovery process of novel lead compounds for CNS diseases. Moreover, combinatorial chemistry and high-throughput screening techniques, combined with structure-based or ligand-based de novo design algorithms, have resulted in libraries containing hundreds of thousands of drug candidates with diverse chemical bases and have consequently enormously increased the possibility of finding new lead structures. Synthetic chemistry platforms paralleled by high-throughput screening (HTS) serve as decision aids in planning neuropharmaceutical development projects. A computational prediction strategy that classifies the compounds by their high-affinity to the target and adequate biopharmaceutical properties must be incorporated early in the discovery process because it would enable rapid selection of promising new chemical entities and also would save both costs and time. [2, 3].

Despite the progress in the field of neuroproteomics and rational drug design strategies, there is still a high failure rate of the candidates in later stages of projects (over 98% of novel compounds do not cross the blood–brain barrier) and consequently an unsatisfied demand for novel neurotherapeutic and neurodiagnostic medicines. Therefore, CNS drug discovery and CNS drug targeting must be merged early in the CNS drug development process in order to obtain neuropharmaceuticals active in the brain following systemic administration [1, 4].

## 2 NEUROPROTECTION AND CNS DISORDERS

The enormous social and economic impact of CNS disorders is highlighted by the incidence of 80 million cases in the United States and the annual cost exceeding $90 billion per year [1]. For example, the financial cost to society of Alzheimer's disease (AD) is estimated at $20–$40 billion per year in the United States [5].

*Amyotrophic lateral sclerosis (Lou Gehrig's disease, ALS)* is a prototypic age-related neurodegenerative disease, characterized by the progressive and specific loss of motor neurons in the brain and spinal cord, which begins in midadult life (the average age at onset is 55 years), life expectancy after the clinical onset varying between 3 and 5 years [5, 7, 8]. Two clinically indistinguishable forms of ALS have been identified: the sporadic form accounting for 90% of cases and the

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

familial form accounting for 10% of cases. In 10–20% of patients with dominant familial ALS mutations of the superoxide dismutase *SOD1* genes have been detected. Neuronal caspase activation (especially caspase 1, 3, and 9), cytochrome c release, and proapoptotic changes in the *Bcl-2* family have also been revealed in spinal cords of animal models of ALS [7].

Neuropathology of ALS comprises the marked reduction in synaptophysin expression that parallels the severity of the neuronal loss and the aberrant accumulation of neurofilaments in the anterior horns of the spinal cord; degeneration of lower motor neurons from the nuclei of cranial nerves and from the ventral horn of the spinal cord; atrophy of striated muscles; and ubiquitin-labeled skeinlike structures in anterior horn nerve cells. The current pathogenetical hypotheses in ALS neuronal death include complex interaction between altered functions of mitochondria and cytoskeleton, glutamate excitotoxicity, and oxidative stress. To date, no therapeutic approaches have been able to significantly improve the course of the disease and riluzole, the only recognized treatment, prolongs survival by 3 months [5].

*Huntington's disease (HD)* is an autosomal-dominant with complete penetrance neurodegenerative disorder, characterized by selective destruction and gliosis of GABAergic, enkephalin- and substance P-containing medium-size spiny neurons within the caudate nucleus and putamen. The loss of inhibitory innervation to the globus pallidus and substantia nigra pars reticulata leads to hyperactivity of the dopaminergic system. The prevalence of HD is between 5 and 10 per 100,000; its onset usually occurs during the fourth or fifth decade of life and is invariably fatal in 10–20 years after the onset of clinical symptoms. HD is characterized by restless, progressive involuntary choreiform movements with psychiatric and cognitive deterioration and dementia [5, 7].

The disease is caused by a mutation of a gene located on the short arm of chromosome 4 (the mutation responsible for the abnormal expansion of CAG-encoded polyglutamine repeats). It has been suggested that the gene product, huntingtin, might be related to the caspase mechanism of apoptotic cell death. One of the earliest events seems to be transcriptional upregulation of the caspase 1 gene, possibly resulting from nuclear translocation of N-terminal fragments of mutant huntingtin. As the disease progresses, there the caspases 3, 8, and 9 are activated and the caspase-mediated cleavage of huntingtin is triggered with an increased generation of huntingtin fragments and depletion of wild-type huntingtin. The caspase-mediated cascade determines the downregulation of the receptors that bind other important neurotransmitters, leading to extensive neuronal dysfunction [8]. Recent studies provide strong evidence for transcriptional dysregulation mediated by aberrant histone acetyl transferase activity as a mechanism of HD pathogenesis [9].

*Parkinson's disease (PD)* is a progressive degenerative disease whose cardinal features are resting tremor, bradykinesia, rigidity, cognitive impairment, and dementia. PD results from degeneration of pigmented cells in the substantia nigra and dopamine depletion of the nigrostriatal system, as well as perturbations of other neurotransmitter systems [10–12].

Pharmacological neuroprotection is largely substitutive and aims to compensate the loss of dopaminergic neurons. Oral L-DOPA associated with carbidopa as peripheral L-amino-acid aromatic decarboxylase inhibitor, the mainstay of current therapy, ameliorates clinical symptoms for a limited period (5–10 years) but is unable to stop nigral dopaminergic neuron death. Other disadvantages are the adverse reactions such as dyskinesia and dystonia, due to the lack of site-specific delivery, and the impossibility to maintain sustained drug levels in the CNS resulting in unpredictable "on/off" motor fluctuation [2].

During *stroke* or *traumatic brain injury*, neuroprotection fails in the core of the lesion, characterized by severe hypoxia and necrotic cell death. Neuroprotective pharmacological intervention is confined only to neurons belonging to penumbra where moderate hypoxia, energy deprivation, and relatively delayed apoptosis create a defined therapeutic time window opportune to interfere with the apoptotic cascade [13–16].

Neuroprotection in stroke and acute neurodegenerative diseases must be targeted at the periphery of the focal lesions where it is possible to intervene on the *contagious apoptosis phenomenon* ("the kindergarten effect") by the inhibition of apoptotic cascade not only in the dying neuron but also in the neighboring nerve cells affected by diffusible proapoptotic factors [interleukin-1 (IL-1), tumor necrotic factor-α (TNF-α), free radicals] [8].

There is no effective treatment able to stop the progression of chronic CNS diseases. The clinical efficacy of current pharmacological neuroprotective approaches decreases as the neurodegeneration progresses and is also overwhelmed by adverse side effects due to site-specific delivery inability. In order to restore the morphological and functional integrity of the affected neurons, two complementary strategies have been considered. The former, a "neuroprotective strategy," aims to prevent continuous neurodegeneration by neurotrophic factor gene therapy [5, 8]. The latter, a "substitution strategy," is based on replacement of the lost neurons by fetal neural tissue transplantation and is now undergoing testing in clinical trials. For example, in PD, transplantation of genetically engineered dopamine-

secreting or neurotrophins-secreting autologous grafts provides an alternative for long-term symptomatic relief and restoration of neurological function [17].

## 3 CRUCIAL ROLE OF THE BLOOD–BRAIN BARRIER (BBB)

### 3.1 Functional Organization of the BBB

The BBB represents the interface between blood and brain and has a vascular aspect represented by capillary endothelium and collagen basement membrane with embedded microvascular pericytes and a brain aspect constituted by the continuous sheath of astrocyte foot processes. The endothelial cells of brain capillaries are linked by complex intercellular tight junctions, without fenestrations or pinocytotic vesicles and form together with the thin pericytes layer and astrocyte foot processes the functional neurovascular unit [18, 19].

The BBB is a physical, metabolic, and pharmacological barrier. The BBB is selectively permeable, playing an important role in regulating the ions influx and efflux, nutrient uptake, and metabolite and xenobiotic clearance between plasma and brain parenchyma. The tight junctions of brain capillary endothelial cells are missing in some areas, for example, circumventricular organs, area postrema, median eminence of the hypothalamus, pineal gland, and the choroid plexuses, where the BBB is considered to be "interrupted," allowing the passage of proteins and small organic molecules [7].

### 3.2 Prediction of Drug Permeability through the BBB

Prediction of BBB permeation has become an essential issue in project development in the neuropharmaceutical industry and must be assessed early, as well as at different phases in the discovery process, in order to achieve a high rate of success and time and cost benefits. BBB penetration prediction in industry is often based on different available models (in vitro, in silico, in vivo) applied in parallel [20]. The animal experimental determination of blood–brain partitioning with radiolabeled and nonradiolabeled compounds is characterized by the highest level of predictability, but also by the highest cost and time requirements, and lowest throughput; various in vitro models of the BBB require interlaboratory validation [20, 21].

Consequently, in silico approaches, mainly based on the correlation between physicochemical properties (solubility, lipophilicity, molecular size, hydrogen-bonding capacity, charge, polar surface area, and solvent-free energies) and the ability of molecules to cross the BBB, either by transcellular passive diffusion and active transport (e.g., receptor-mediated or carrier-mediated transcytosis) have been developed [21, 22]. Rational neuroactive drug design involves computational approaches for prediction of drug permeability through the BBB [23, 24]. Molecular descriptors, together with Bayesian-regularized neural nets, were employed to model the partitioning of a diverse range of drugs and other small molecules across the BBB [25].

### 3.3 BBB Genomics and Proteomics

The application of genomic and functional proteomic methods in BBB research has revealed that, compared to brain or peripheral tissues, its unique molecular characteristics are derived from the brain capillaries genes' differential expression and specific protein profile expression. The selectively expressed genes mediate both BBB functions and the complex involvement of the BBB in neuropathology.

These new methodologies, considered the most powerful ever applied to brain drug development, aim to identify novel, selectively expressed, endogenous transporters as new targets for innovative drug delivery through the BBB in vivo [1, 4, 18, 26].

### 3.4 Strategies for Brain Drug Targeting

The invasive neurosurgical-based strategies either deliver drug *behind* the BBB (intracerebroventricular infusion or intracerebral implants) or physically *disrupt* the BBB following the intracarotid infusion of noxious agents [1].

Strategies for the *noninvasive BBB drug targeting* focus on the endogenous BBB transport systems [receptor-mediated transcytosis (RMT) or carrier-mediated transport across the brain capillary endothelium), which are revealed by BBB genomics and proteomics programs and allow drugs to be transported *through the BBB*.

*Chimeric peptide technology* involves the conjugation of the neuropharmaceutical agents (small molecules or peptidomimetic compounds) to a brain-targeting vector that binds with high affinity and specificity to a BBB-expressed transport system and triggers the transcytosis through the BBB in vivo. The cited brain drug-targeting *vectors* include monoclonal antibodies or "endocytosing MAb" [human insulin receptor (HIR) monoclonal antibody (HIR MAb) or the transferrin receptor (TfR MAb)], endogenous peptides, or plasma proteins (ligands for the RMT or absorbtive-mediated systems), highly cationized proteins (cationized humanized MAb) [1, 27]. The *linker strategy* (chemical conjugates, genetically engineered fusion proteins, avidin-biotin, or

pegylated immunoliposome technology) must be chosen to assure the bifunctionality of drug–vector conjugate and to optimize its plasma pharmacokinetics in order to improve its brain uptake [1].

## 4 DESIGN OF ANTISENSE NEUROPHARMACEUTICALS

Antisense oligonucleotides (antisense OGN) are short, monocatenar, synthetic sequences of chemically modified nucleic acids, which specifically hybridize, according to the base-pairing rules (adenine–tymine/uracil, guanine–cytosine), with the complementary nucleotide sequences from target deoxyribonucleic acid (DNA) or target ribonucleic acid (RNA), thus blocking the protein synthesis encoded by these targets in certain cells/tissues. The antisense mechanisms of action include activation of ribonuclease H (RNase H, the enzyme that cleaves the messenger RNA (mRNA) strand from the mRNA/antisense OGN hybrid), translation arrest, inhibition of polyadenilation or 5′-capping, splicing inhibition, and target mRNA degradation due to intrinsic catalytic properties of the ribozymes [28].

There are several chemical classes with distinct antisense properties, their medicinal chemistry being described in Crooke [29] and Gualtieri [30]. Peptide nucleic acids (PNA), consisting of repetitive *N*-aminoethylglycine units, represent *very attractive lead compounds* for the development of antisense drugs because of their (1) neutral, achiralic, hydrophilic, and low-molecular-weight structure; (2) intense hybridization properties; (3) high chemical and biological stability; (4) negligible interactions with proteins and thus low toxic effects; (5) ease of synthesis; (6) absence of neurotoxicity; and (7) design as neurodiagnostic and neurotherapeutic agents by conjugation with BBB delivery systems.

The antisense oligonucleotides as neuropharmaceuticals have the following *advantages* over the classical medicines: (1) remarkable selectivity because antisense OGN discriminate between closely related proteins derived from homologous genes, polymorphic alleles, or alternative splicing products; (2) highly specific regional targeting in the brain; (3) facile design and shorter period required to develop an antisense drug due to powerful in vitro and in vivo screening methods; (4) the possibility to select the most susceptible situs of the targeted mRNA for antisense hybridization; (5) the opportunity to choose the most specific, selective, and stable antisense oligonucleotide from a wide range of chemical modifications; and (6) simple solid synthesis (phosphoramidite method) at low cost [28, 30].

The *antisense neurotherapeutic candidates* might be indicated in neurodegenerative disorders, brain viral disease, and cerebral tumors. For instance, in HD, the molecular target for antisense strategy is either huntingtin or glyceraldehyd-phosphate-3-dehydrogenase (GAPDH) mRNA; in cerebral acquired immure deficiency syndrome (AIDS), human immunodeficiency virus (HIV) mRNA; in malignant gliomas, antisense OGN are directed against aberrant transcripts responsible for the uncontrolled growth and proliferation. The neuroprotective role of the antisense OGN anti-NMDA-R1 receptor transcript was demonstrated in rat models of cerebral ischemia [29].

In contrast to classical AD therapy, very low doses of antisense OGN targeted to specific acetylcholinesterase (AChE) isoform transcripts appear to be the only solution to block de novo synthesis of AChE protein, thereby inhibiting both its catalytic and noncatalytic activities. In familial AD, phosphorothioate antisense directed to presenilin-1 mRNA proved to be a neuroprotective activity against NMDA-induced neuronal cell death [28, 31–33].

The antisense oligonucleotide AS3 specific to AChE, in trauma-sensitive AChE transgenic mice, minimized mortality, facilitated neurological recovery, and protected CA3 hippocampal neurons, suggesting its possible role in the treatment of acute head trauma. AS3, delivered in nanomolar quantities by surgically implanted cannulae to the cerebrospinal fluid (CSF) of cognitively impaired transgenic mice, has shown promising results [28].

Antisense radiopharmaceuticals as *neurodiagnostic tools* are useful for direct and early detection of the exact moment of expression of any pathologic gene in vivo [1].

Antisense oligonucleotides have been proved as important *tools in validating* different isoforms of CNS proteins as *potential targets* for therapeutic intervention, such as β-secretase and caspase-2 in AD [34], 5-$HT_{2A}$ receptors, and corticotropin-releasing factor (CRF2) receptors in anxiety and depression, galanin (GAL1) receptors in galanin antinociceptive effects, neuropeptide Y role in food intake regulation, adenosine triphosphate (ATP)-dependent A-type potassium channel Kv1.4 in the long-term potentiation and spatial memory, the opioid receptors mu, delta, kappa characterization, and nNOS-1 and nNOS-2 role on morphine analgesia [28].

## 5 NEUROTROPHIC FACTORS—A PROMISING THERAPEUTIC STRATEGY

### 5.1 Classical Neurotrophic Factors

Neurotrophic factors belong to neuronal intrinsic homeostasy. They are proteins that act on neurotrophin

receptors and regulate the neuronal survival/death ratio by stimulating neuroblastic growth and differentiation, axonal sprouting to final anatomical targets, hierarchical distribution of synapses on somatodendritic surfaces, dendritic spines proliferation, synaptic plasticity, and normal brain chemical architecture.

Different neurotrophic factors have distinct roles in modulating the function of a neuron or specific neuronal populations. In the living brain, neurotrophic factors are associated in various combinations that act synergistically to promote neuronal survival and regeneration in CNS diseases.

The neurotrophin receptors are classified as (a) the *nonspecific* low-affinity p75 receptor, able to bind all neurotrophins and modulate their interactions with trk receptors and (b) *the family of tyrosine kinase-linked (trk) receptors*, which bind specific neurotrophins.

The neurotrophic factors are divided into (a) *the classic neurotrophin family* including nerve growth factor (NGF), brain-derived neurotrophic factor (BDNF), neurotrophin-3 (NT-3), neurotrophin-4/5 (NT-4/5), and the nonmammalian growth factors neurotrophin-6 (NT-6) and neurotrophin-7 (NT-7); (b) *the cytokine family*, consisting of leukemia inhibitory factor (LIF or cholinergic differentiation factor), ciliary neurotrophic factor (CNTF) and cardiotrophin-1; (c) *the fibroblast growth factors family*, with FGF-1 and FGF-2 (also known as acidic FGF and basic FGF); (d) *insulin and insulin-like growth factors* (IGF-I and IGF-II); (e) *the transforming growth factor-β (TGF-β) superfamily* containing: glial cell-line-derived neurotrophic factor (GDNF), neurturin, persephin, and the bone morphogenic proteins (BMPs); (f) *the epidermal growth factor family* including epidermal growth factor (EGF), transforming growth factor-α (TGF-α), neuregulins, and NTAK (Neural- and Thymus-derived Activator for ErbB Kinases); and (g) *additional molecules with growth-promoting activities*: and hepatocyte growth factor (HGF), immunophilins, platelet-derived growth factor (PDGF), interleukins, heparin-binding neurotrophic factors (pleiotrophin) [5].

The mature forms of neurotrophins are released by proteolytic cleavage of their precursor forms, proneurotrophins, which might have distinct and sometimes contrasting effects in comparison with the mature neurotrophins on neuronal survival and degeneration and on regulating synaptic plasticity in neuropathological conditions [35, 36].

The success of exogenously neurotrophic factors as neuroprotective and regenerative therapy of neurological disorders is highly dependent on the following aspects: bioavailability after various routes of delivery, developing effective brain-targeting strategies across the BBB, and targeted and safe delivery restricted to specific neuronal populations.

The neurotrophic factors may be delivered by gene therapy to precise brain neuronal population targets, in a potentially safe and long-lasting manner. The in vivo gene therapy approach is based on viral-derived vectors able to directly insert the desired neurotrophic factor gene into targeted brain cell populations. Site-specific transplants of ex vivo genetically modified single cell or small clusters of cells (autologous host cells or xenogeneic grafts from different species) express and secrete human neurotrophins. The polymer-encapsulated genetically engineered cell implants provide the following advantages: characterization (e.g., parameters like neurotrophic factor secretion and stability, homogeneity, and reproducibility) and amplification to generate working cell storage banks, as well as constant delivery to the CNS; prevention of graft rejection and alleviation of the necessity of pharmacological chronic immunosuppression of the host; and the possibility of rapidly retrieving it in the case of undesired side effects [5].

The neurotrophic factor strategy has been validated in small and large animal models of PD, HD, ALS, and AD, using regionally restricted transplants of encapsulated genetically modified cells to secrete diverse neurotrophins. However, some issues remain to be solved in the future: new optimal methods and site of gene delivery, number of neurotrophins-secreting grafts in the brain, the stability and the duration of gene expression achievement, the regulation of gene expression, the duration of neuroprotection after transgene expression ceases, the feasibility of repeating the grafting procedure in the case of downregulation of gene expression, and better methods of testing (innovative neuroimaging techniques) on genetic models of neurodegenerative diseases [5, 37].

Delivery of the neurotrophic growth factors across the BBB, after their peripheral administration, might also be achieved by chimeric peptide technology [1].

Neurotrophic factor therapy is particularly relevant in HD because it may be applied before the onset of striatal neurodegeneration to virtually all individuals of genetically identified families. NGF is the first successful therapeutic undertaking for early phases of AD and HD and is now in a phase III of clinical trial for diabetic neuropathy. BDNF is subject to clinical trials for ALS, preventing the degeneration of basal forebrain cholinergic and GABAergic neurons, and also of the brain stem neurons after axotomy in spinal cord injury models [5].

In the early stages, PD also benefits from dopaminergic growth factor therapy. GDNF is undergoing clinical trials for PD, thanks to its proven remarkable and long-lasting functional stimulatory effects, being the most potent and the most selective neurotrophin on nigral dopaminergic neurons [5]. NT-3 is undergoing clinical trials in neuropathy, spinal cord injury with motor

deficits, and demyelinating conditions. In vivo FGF-1 reduces the neuronal loss after hippocampal ischemia, promotes axonal growth after spinal cord injury, and has entered clinical trials for stroke.

Experimental and clinical data have demonstrated the additive and, in some cases, synergistic neurotrophic effects of combined neurotrophin administration. For instance, BNDF in conjunction with CNTF, or cardiotrophin-1 (CT-1) associated with GDNF, have shown a synergistic neuroprotective effect in mouse models or clinical trials for ALS [5].

### 5.2 Small-Molecule Neurotrophin Mimetics

Design and development of synthetic, small-molecule, neurotrophin mimetics (neurotrophin receptor agonists, partial agonists, inverse agonists or antagonists, modulators of neurotrophin-related signal transduction) that preserve the putative active sites of their larger parent factors and are endowed with favorable profiles of stability, BBB penetration, and targeted biological actions would be useful therapeutic candidates for neurological disorders [38].

Small-molecule TGF-β mimetics might be developed into candidates for the treatment of neurodegeneration because TGF-β is an organizer of the brain's response to injury and is known to be neuroprotective, reducing the accumulation of Aβ peptide [39, 40].

ADNF-9 and nucleosome assembly protein (NAP) are neurotrophic short peptide derivatives of the activity-dependent neuroprotective protein (ADNP) and activity-dependent neurotrophic factor (ADNF) that promote neurite outgrowth, NAP being currently in clinical trials [41, 42].

Pan-neurotrophins and neurotrophic factor "fusion" proteins are synthetic peptides that contain active sites of at least two neurotrophic factors within a single molecule and prove additive neurotrophic properties. For example, pan-neurotrophin-1 is a fusion of NGF, BDNF, and NT-3 active sites. Pan-neurotrophins are designed as potential therapeutic strategy for neurological disorders characterized by the degeneration of multiple types of neurons, including stroke, trauma, and AD [5].

## 6 NEUROTRANSMITTER SYSTEMS AS THERAPEUTIC TARGETS

### 6.1 Chemical Central Synapse

Chemical transmission is the most important mechanism by which the neurons communicate with one another. Morphologically, the interneuronal contact called synapse is formed by the presynaptic membrane (axonal termination), the synaptic cleft, and the postsynaptic membrane (somatodendritic) encapsulated by glial cells. Chemical transmission of information requires successively: (1) the presence of the neurotransmitter in the presynaptic terminal stored in secretory vesicles, (2) quantal release in the synaptic cleft by exocytosis, (3) the presence of specific receptors on the postsynaptic membrane, and (4) the mechanism of neurotransmitter inactivation. The iontophoretic application of neurotransmitter on postsynaptic receptors should mimic the electrical stimulation of presynaptic terminal.

*The life cycle of a neurotransmitter (Fig. 1)* is a sequential process that starts with *precursor uptake* into the neuron from the glial cells. *Enzymatic processing* of precursor and *neurotransmitter synthesis* take place in the endosomes in the presence of mitochondria. By *budding*, the endosome gives rise to the neurotransmitter vesicle that is progressively *translocated* to the active zone of the presynaptic membrane. Here, in the presence of $Ca^{2+}$, selected specific proteins (synapsin, syntaxin, synaptobrevin) realize the attachment (*docking* and *priming*) and the *fusion* of the synaptic vesicle to the active zone, the opening of the vesicle and neurotransmitter release in the synaptic cleft (*exocytosis*).

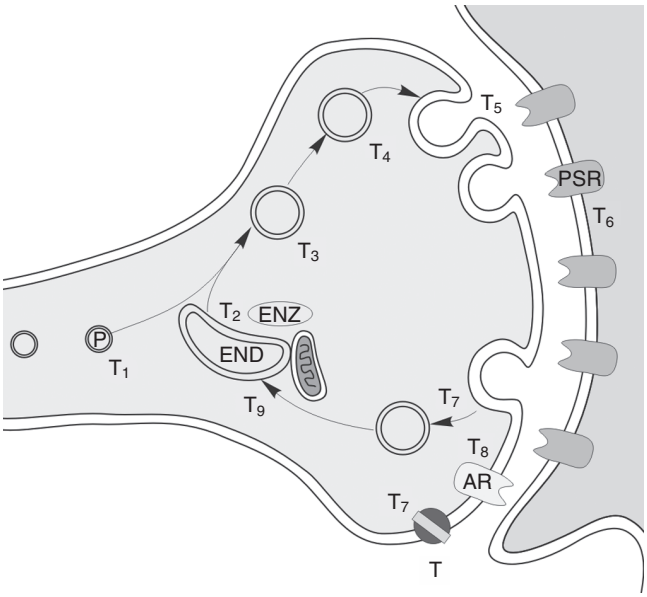


**Figure 1** Potential targets of neuroprotective strategies (diagram of a chemical central synapse): T1 = precursor inactivation, T2 = enzymatic synthese inhibition, T3 = storage interferation, T4 = release blocking, T5 = intrasynaptic cleft inactivation, T6 = postsynaptic receptor agonist/antagonist interaction, T7 = reuptake (vesicular/transporter) inhibition, T8 = autoreceptor interaction (modulatory), T9 = catabolism inhibition, END = endosome, ENZ = enzyme, P = precursor, PSR = postsynaptic receptor, AR = autoreceptor, T = transportor.

The botulinic toxins are able to disrupt this process between priming and fusion.

In the synaptic cleft the neurotransmitter interacts with the receptors of the postsynaptic membrane, directly coupled with ion channels or with receptors that act through second messengers as G-proteins. In the presynaptic membrane there are neurotransmitter autoreceptors that modulate exocytosis or neurotransmitter reuptake.

From the synaptic cleft the released neurotransmitter is inactivated by *enzymatic degradation*, *reuptake* into presynaptic terminal by endocytosis, by means of a protein transporter coupled with a $Na^+$ channel or by uptake in the glial cells. The reuptaken neurotransmitter is recycled or metabolized by specific enzymes.

*The neurotransmitters* may be classified as excitatory or inhibitory, based on receptor response, but often the quality of response depends upon receptor specialization.

*The neuromodulators* are compounds with both neuronal and nonneuronal origin that act hormone-like and influence the synaptic events, reducing, enhancing, or prolonging the action of neurotransmitters.

The human CNS has two kinds of neurons:

*Long axon Golgi 1* are relay neurons and receive on the somatodendritic membrane more than 90,000 synaptic contacts of which the excitatory ones are located on the peripheral, spiny part of the dendritic tree, while the inhibitory ones on the basal, smooth dendrites and soma. The modulatory synapses are situated presynaptically on the axonal hillock or terminal bouton. The long axons of Golgi 1 neurons project in various brain regions and release a dominant, usually excitatory neurotransmitter in the innervated territory. The therapeutic interference of the neurotransmitter system will be expected to have localized effects related to the territory's function.

*The short axon Golgi 2* are intrarelay interneurons organized as local microcircuits whose terminals secrete inhibitory neurotransmitters. Their therapeutic interference will be expected to have rather generalized diffuse effects. The most important classes of neurotransmitters as targets of neuroprotective strategies will be discussed below.

### 6.2 Serotonin (5-Hydroxytryptamine)

The 5-hydroxytryptamine (5-HT) has been known since the mid-nineteenth century primarily as a cause of hypertension. After its isolation within the CNS, extensive studies demonstrated that the serotoninergic neurons are located in the raphe nuclei of the brain stem reticular formation. According to Dahlstrőm and Fuxe, they are classified into caudal medullary groups (B1–B3), which project into the spinal cord, and pontine and mesencephalic groups (B4–B9), which practically innervate the whole forebrain [6]. Subsequent to these connections, serotonin is involved in every type of behavior, including eating and sleep, in circadian rhythm regulation, neuroendocrine function, and pain perception [7].

Due to the serotonin indolalkyl amine structure and its resemblance to the psychedelic drug lysergic acid diethylamide (LSD), the abnormalities of serotonin synthesis were thought to be involved in various forms of psychic disorders, emesis, migraines, and sexual, sleep, eating, and stress disorders.

Acording to Figure 1, the life cycle of serotonin may be pharmacologically interferred with as follows [2, 7, 43–53]

***T1.*** Diet without tryptophan.

***T2.*** The synthetic pathway of serotonin has two stages: In the first stage the precursor tryptophan is converted to 5-hydroxytryptophan by tryptophan hydroxylase, and in the second, 5-hydroxytryptophan is decarboxylated to form serotonin. During the first stage of synthesis, tryptophan hydroxylase is inhibited by *p*-chlorophenylalanine and α-propyldopacetamide.

***T3.*** The uptake and storage into synaptic vesicles is strongly inhibited, with severe serotonin depletion, by reserpine and tetrabenazine.

***T4.*** The release of serotonin into the synaptic cleft is indirectly inhibited by LSD, which decreases the firing rate of serotoninergic neurons. Until today, there are no therapeutic candidates to interfere with serotonin release.

***T6.*** At least 18 serotonin receptor subtypes are known nowadays. Based on their coupling to second messengers and amino acid sequence homology, they may be divided into: $5\text{-HT}_{1\ (1A,\ 1B,\ 1D,\ 1E,\ 1F)}$, $5\text{-HT}_{2\ (2A,\ 2B,\ 2C)}$, $5\text{-HT}_3$, $5\text{-HT}_4$, $5\text{-HT}_{5(5A,\ 5B)}$, $5\text{-HT}_6$, $5\text{-HT}_7$ families. The opportunities of their pharmacological interference as well as the effector pathways are presented in Table 1.

***T7.*** The presynaptic autoreceptors $5\text{-HT}_{1A}$ (agonist buspirone) are located on the somatodendritic membrane and the $5\text{-HT}_{1B,1D}$ (agonist sumatriptan) on the terminal bouton membrane.

***T8.*** The reuptake into the presynaptical terminal seems to be the main way of serotonin inactivation in the synaptic cleft. The process is inhibited by tricyclic antidepressants (TCA) and selective serotonin reuptake inhibitors (SSRI).

**TABLE 1 Ligands of Serotonin Receptors**

| 5-HT Receptor Subtype | Effector | Agonist | Antagonist |
|---|---|---|---|
| 1A | cAMP, $K^+$channel | 8-OH-DPAT, ipsapirone, piclozotan<br>5-CAT | SDZ 216-525, WAY-100135, spiperone, S(2)-pindolol |
| 1B | cAMP | CP-93,129<br>5-CAT | Isamoltane |
| 1D | cAMP | Sumatriptan, 5-CAT<br>5-(nonyloxy)tryptamine | Isamoltane<br>GR 127935 |
| 1E | cAMP | — | — |
| 1F | cAMP | — | — |
| 2A | $IP_3$/DAG | α-Methyl-5-HT<br>DOI, DOB | Ketanserin, ritanserin, spiperone, LY-53857<br>MDL 100,907 |
| 2B | $IP_3$/DAG | α-Methyl-5-HT, DOI, MK212 | SB 200646A, LY-53857 ketanserin |
| 2C | $IP_3$/DAG | α-Methyl-5-HT, MK212, DOB | SB 200646A, meuslergine, LY-53857, ketanserin, spiperone, metergoline |
| 3 | Cation channel | 2-Methyl-5-HT, *m*-CPBG | Tropisetron, ondansetron, granisetron, zacopride |
| 4 | cAMP | 5-Methoxytryptamine, SB 205149, BIMU8, SC-53116 | GR113808<br>RS 23597-190 |
| 6 | AC | 5-CAT, LSD | Lisuride, clozapine |
| 7 | AC | LSD | Clozapine, amitryptiline |

***T9.*** Serotonin is catabolized by monoamine oxidase (MAO), which is inhibited by MAO inhibitors (MAOIs).

In conclusion, serotonin neurons and receptors are potential targets for a wide class of compounds expected to have antidepressant, anxiolytic, antipsychotic, and antiemetic properties.

### 6.3 Acetylcholine

In 1907, Hunt proposed acetylcholine (ACh) as a mediator of intercellular signaling, in 1914 Dale evidenced its parasympathomimetic action, and Loewi (1921) in his famous "dream experience" demonstrated ACh liberation at nerve terminals after vagal stimulation. After these early discoveries, cholinergic neurons were found in numerous regions of both peripheral (PNS) and central nervous systems (CNS). The cholinergic brain is built up of cholinergic cell groups (Ch1–Ch6) and cholinergic pathways [7, 54].

The Ch1–Ch3 neurons are located in the rostral basal forebrain (medial septal nucleus, nucleus of the diagonal band of Broca, and preoptic area) and project to the hippocampus, the amygdalae, the olfactory bulb, and the limbic cortex. The Ch4 cell group is located in the caudal part of basal forebrain (nucleus basalis Meynert, nucleus ansae reticularis, and substantia innominata) and project to the entire neocortex (pyramidal neurons of layer 5). The Ch5 and Ch6 cell groups are located in the pontomesencephalic region (pedunculopontine and dorsolateral tegmental nuclei) and project to the tectum, thalamus, striatum, substantia nigra, locus coeruleus, and reticular formation. ACh is also synthesized in spinal and cranial nerves motoneurons, transported along the axons, and released at the level of neuromuscular junctions of the skeletal muscles [54].

The ACh is involved in the regulation of biorhythms, of aggressive and sexual behavior, of sleep and arousal, as the neurotransmitter of ascending reticular activating system, as well as in learning, attention, memory, and other cognitive processes that progressively deteriorate in AD.

The life cycle of ACh may be pharmacologically interferred as follows [2, 7, 44, 47, 52, 55–57]:

***T2.*** Styryl pyridine derivatives block the ACh synthesis from acetylCoA and choline.

***T3.*** ACh storage into synaptic vesicles is blocked by AH5183 (vesamicol).

***T4.*** Release is blocked by $Mg^{2+}$, botulinum toxin, cytochalasin B, and is promoted by $La^{3+}$ and β-bungarotoxin.

***T5.*** In the synaptic cleft ACh may be inactivated by stimulation of acetylcholinesterases reactivators. The acetylcholinesterase is inhibited by physostigmine (reversible), by DFP (irreversible), or by donepezil-tacrine hybrid-related derivatives.

***T6.*** The ACh receptors, classically divided into muscarinic and nicotinic, may be classified after resolution of receptor subtypes with chemically synthesized agonist and antagonists. Selective agonists and antagonists of both nicotinic

**TABLE 2 Agonists and Antagonists of ACh Receptor**

| ACh Receptor Subtype | Alcaloid Response | Agonist | Antagonist |
|---|---|---|---|
| Muscular $(\alpha1)_2\beta\varepsilon\delta$, $(\alpha1)_2\beta\gamma\delta$ | Nicotinic | Acetylcholine, nicotine, phenyltrimethyl ammonium, methylcarbamylcholine, dimethylphenylpiperazinium, cytisine, methylisoarecolone | *d*-Tubocurarine, elapid α-toxins, lophotoxin, dihydro-bethaerythroidine, α-bungarotoxin, conotoxin MI |
| Neuronal (α2-α9, β2-β5) | Nicotinic | Acetylcholine, nicotine, methylcarbamylcholine, dimethylphenylpiperazinium, cytisine, methylisoarecolone | *d*-Tubocurarine, trimethaphan, bungarotoxin, lophotoxin, dihydro-bethaerythroidine, κ-bungarotoxin neosurugatoxin |
| $m_1$-$M_1$ | Muscarinic | Acetylcholine, muscarine, pilocarpine, oxotremorine, McN-$A_{343}$ | Atropine, propylbenzilylcholine mustard, quinuclidinyl benzilate, pirenzepine, telezepine, 4-DAMP |
| $m_2$-$M_2$ | Muscarinic | Acetylcholine, muscarine, pilocarpine, oxotremorine, bethanechol | Atropine, propylbenzilylcholine mustard, quinuclidinyl benzilate, methoctramine, gallamine himbacine, AF-DX-116 |
| $m_3$-$M_3$ | Muscarinic | Acetylcholine, muscarine, pilocarpine, oxotremorine, L-689,660 | Atropine, propylbenzilylcholine mustard, quinuclidinyl benzilate, hexahydroxiladifenidol *p*-fluorohexahydroxiladifenidol, 4-DAMP |
| $m_4$-$M_4$ | Muscarinic | Acetylcholine, muscarine, pilocarpine, oxotremorine, McN-$A_{343}$ | Atropine, propylbenzilylcholine mustard, quinuclidinyl benzilate, himbacine, tropicamid, 4-DAMP |
| $m_5$-$M_5$ | Muscarinic | Acetylcholine, muscarine, pilocarpine, oxotremorine | Atropine, propylbenzilylcholine mustard, quinuclidinyl benzilate, 4-DAMP |

and muscarinic receptors are shown in the Table 2.

***T8.*** Choline competitive reuptake is blocked by hemicholinium and troxypirrolinium tosylate.

### 6.4 Inhibitory and Excitatory Amino Acids

***γ-Aminobutyric Acid (GABA)*** GABA is the dominant inhibitory neurotransmitter in the CNS. The GABAergic cell groups are located in various regions of the brain such as the cerebral cortex, hippocampus, striatum, substantia nigra, cerebellar cortex, and spinal cord. The neurons within these areas form long GABAergic pathways (cerebellar Purkinje cells output to deep nuclei and lateral vestibular nucleus and the striatonigral connections) or local microcircuits and realize both post- and presynaptic inhibition.

GABA plays an important role in various physiological and behavioral mechanisms, including locomotory activity, thermoregulation, cardiovascular activity, pain perception, feeding, aggressive and sexual behavior, and is involved in the pathogenesis of epilepsy, tardive dyskinesias, schizophrenia, PD and HD, alcoholism, sleep disorders, and senile dementia [7, 54].

The life cycle of GABA features some attractive targets for therapeutic interference [2, 47, 52, 58–60].

***T2.*** GABA is synthesized from glutamate in a single step, catalyzed by glutamate acid decarboxylase (GAD), and is catabolyzed by the action of GABA aminotransferase (GABA-T) and succinic semialdehyde dehydrogenase (SSADH). This closed-loop process, called GABA shunt, has a double goal, to produce and to conserve the neurotransmitter. Enzymatic synthesis of GABA is inhibited by different hydrazines, but L-glutamate hydrazide and allylglycine seems to inhibit selectively the GAD.

***T4.*** The release of GABA into the synaptic cleft is a calcium-dependent process, but the information concerning selective inhibition of GABA release is still missing.

***T6.*** In mammalian CNS three main subtypes of receptors ($GABA_A$, $GABA_B$, $GABA_C$) have been identified [60–64]. The $GABA_A$ receptors are ligand-gated $Cl^-$ channels that mediate fast inhibitory postsynaptic potentials (IPSPs) in the postsynaptic neuron. The glycoprotein of this receptor is made up of at least three subunits, α, β, and γ, each of them with three to six isoforms. The receptor complex exhibits a GABA binding site that mediates the effects of neurotransmitter agonists and antagonists, and four other binding sites for benzodiazepines, barbiturates, picrotoxine, and neurosteroids. The $GABA_B$ receptors are G-protein-coupled receptors, which are present also in the presynaptical membrane

**TABLE 3 Ligands for Different GABA Receptors Subtypes**

| GABA Receptors Subtype | Agonists | Antagonists | Modulators |
|---|---|---|---|
| $GABA_A$ | GABA, muscimol, isoguvacine | Bicuculline, GABAzine, picrotoxin, TBPS | Benzodiazepines, barbiturates, neurosteroids |
| $GABA_B$ | GABA, baclofen | Phaclophen, 2-hydroxysaclofen, CGP-36742, 3-APPA | — |
| $GABA_C$ | CACA, GABA, muscimol | — | — |

where they inhibit the neurotransmitter release. The $GABA_C$ receptors are also associated to the $Cl^-$ channel, but they show no modulation by benzodiazepines or barbiturates (Table 3).

***T7.*** Presynaptic autoreceptors are probably involved in the regulation of GABA release.

***T8.*** From the synaptic cleft GABA is actively taken up into presynaptic terminals by the sodium-dependent mechanism, which is inhibited by some compounds such as 4-methyl-GABA, 2-hydroxi-GABA, tiagabine, and *N*-methyl-*exo*-THPO [44, 47].

**T9.** The reuptaken GABA is catabolyzed by mitochondrial GABA-T, and this mechanism is inhibited by gabacuculline, nipeconic acid, amino-oxyacetic acid, and acetylenic GABA.

***Glycine*** Glycine is the simplest amino acid that is present in all tissues, participates in different metabolic pathways, and is found especially in the spinal cord and hindbrain where it acts as specific neurotransmitter of Renshaw cells mediating recurrent inhibition. In the brain, glycine is synthesized from serine by the action of serine hydroxymethyltransferase and is released into the synaptic cleft by a calcium-dependent mechanism [44].

The glycine receptors, made up of two subunits, α ($\alpha_1$–$\alpha_4$) and β, are inhibitory ligand-gated $Cl^-$ channels that produce fast IPSPs in the postsynaptic neuron. The major antagonist of the glycine receptor is strychnine, which produces global excitation and convulsant effects. The role of glycine in sensorimotor functions and behavior is confirmed by its involvement in the "startle disease," an inherited disorder characterized by sudden muscular rigidity as a response to an unexpected stimulus (loud noise). Although it is an inhibitory neurotransmitter, glycine also participates as cofactor in the activation of NMDA excitatory receptors [7, 54, 65].

***Glutamate and Aspartate*** Glutamic and aspartic acids act as neurotransmitters in an ionized form called glutamate and aspartate. They have high brain concentration and potent excitatory effects on postsynaptic neurons. Glutamate is a ubiquitous neurotransmitter of the long descending pathways of the motor system, of the limbic projection, including the circuit of Papez and of the parallel fibers of the cerebellar cortex. Aspartate was found as a neurotransmitter of the climbing olivo-cerebellar fibers and of some cortical and hippocampal connection [7, 54].

Glutamate is a nonessential amino acid that does not cross the BBB. Therefore, it is synthesized in the neurons starting from glucose, by transamination of α-oxoglutarate acid and in the glial cells from glutamine by enzymatic action of glutaminase within mitochondria. Glutamate is taken up and stored into synaptic vesicles through $Mg^{2+}$/ATP-dependent mechanism and is released by exocytosis [44].

On the postsynaptic membrane there are the NMDA, AMPA, and kainate ionotropic receptors and G-protein-coupled metabotropic receptors. Activation of NMDA receptors gives rise to slow excitatory postsynaptic potentials (EPSPs) and stimulates some $Ca^{2+}$-dependent enzymes, such as calcium/calmodulin-dependent protein kinase II (CaM-KII), protein kinase C (PKC), phospholipase $A_2$, and nitric oxide synthase (NOS). AMPA and kainate receptors mediate fast EPSPs in the postsynaptic neuron [66, 67]. The metabotropic receptors are classified into three groups based on second-messenger coupling and pharmacological properties and are located on both presynaptic and postsynaptic membranes (Table 4) [65, 68–72].

The rapid removal of glutamate from the synaptic cleft serves both to interrupt the excitatory signal and to prevent the excitotoxic damage. It is taken up into the presynaptical terminal by a specific neuronal $Na^+$-dependent and enantioselective transporter and into glial cells by an other two different glutamate transporters.

*Glutamate Receptors and LTP* NMDA, AMPA, and metabotropic receptors play an important role in the synaptic response enhancement called long-term potentiation (LTP). It was demonstrated that LTP involves increased presynaptic glutamate release and also an increased sensitivity of the postsynaptic membrane. These processes are dependent on the activity of some $Ca^{2+}$-dependent enzymes (Ca M-KII, PKC) and require

**TABLE 4 Excitatory Amino Acid Receptor Ligands**

| Receptor | | Agonist | | Antagonist | |
|---|---|---|---|---|---|
| | | RSSA[a] | RSA[b] | Rssa[c] | Rsa[d] |
| NMDA | Glutamate site | NMDA | NMDA | *D*(-)-AP-5, *D*(-)-AP-7, CPP | *D*(-)-AP-5 |
| | Glycine site | Glycine, D-serine, *R*(+) HA-966, L-687, 414 | | 7-Chlorokynurenic, MNQX | |
| AMPA | | | AMPA | | CNQX, DNQX |
| Kainate | | | Kainic acid | | CNQX, DNQX |
| Metabotropic | | | L-$AP_4$ | | MCPG |

[a]RSSA: receptor subtype selective agonist.
[b]RSA: receptor selective agonist.
[c]rssa: receptor subtype selective antagonist.
[d]rsa: receptor selective antagonist.

the retrograde second-messenger action of NO and arachidonic acid. Experimental evidence that NMDA receptor blockers (*D*-AP-5) interrupt the acquisition of information, and that AMPA receptor agonists, such as 1-(1,3-benzodioxol-5-ylcarbonyl)-piperidine, facilitate memory retention, confirms the interrelation between LTP, learning, and memory and shows promising perspectives for "memory" drugs design [7, 54].

*Excitotoxicity* appears when energy metabolism is compromised and leads to acute neurodegeneration. Depletion of energy stores affects the ability of energy-dependent ionic pumps to maintain resting membrane potential and, subsequently, the depolarization of presynaptic neuron triggers action potentials, release of glutamate in the synaptic cleft, and postsynaptic AMPA and NMDA receptor activation. The increased influx of $Ca^{2+}$ through glutamate receptors triggers cytosolic calcium dyshomeostasis and the cascade of neuronal death. Some neuroprotective drugs, such as ifenprodil, appear to be NMDA receptor blockers. Chronic glutamate receptor activation has been implicated in some neurodegenerative disorders, such as ALS, Huntington's chorea, and olivopontocerebellar atrophy. Recent demonstration of the topology of receptor subunits and the identification of the major structural domains responsible for ion permeability suggests that new therapeutic strategies must be targeted at the molecular diversity of these receptors [44, 54, 73–75].

### 6.5 Catecholamines

Catecholamines are a group of important neurotransmitters or hormones that contain a catechol nucleus and an amine group. They consist of norepinephrine (NE), epinephrine (EPI), and dopamine (DA).

***Norepinephrine*** NE is a neurotransmitter in both CNS and PNS, EPI is a hormone released from the adrenal gland, and DA serves as neurotransmitter in some important pathways that form together the dopaminergic brain.

The noradrenergic neurons, clustered in the brain stem reticular formation (FR), are organized into three different nervous cell populations. The dorsal one (A6 and A4) located in the upper lateral aspect of the rhomboid fossa is represented by locus coeruleus, considered the main brain catecholamine source. Their long axons project to the spinal cord, the cerebellum, the diencephalon, and the cerebral cortex and belong to the activatory system of FR, which is involved in attention, arousal, and vigilance. The lateral tegmental noradrenergic neurons (A1, A5, and A7) and the dorsal medullary group (A2) project to the brain stem nuclei and hypothalamus and are involved in feeding behavior and hunger regulation. Peripheral noradrenergic cells are located in sympathetic ganglia and in the adrenal medulla [54].

The precursor of catecholamine biosynthesis is L-tyrosine. Through the action of tyrosine hydroxylase results L-DOPA, which is decarboxylated to dopamine by DOPA decarboxylase. For all neurons that synthesize EPI or NE, DA serves as an essential precursor. DA by action of dopamine β-hydroxylase forms NE, which is converted to EPI by phenylethanolamine *N*-methyltransferase. The first step of the biosynthetic pathway, prior to dopamine formation, occurs in the cytosol. DA is taken up into storage vesicles and the final β-hydroxylation takes place within vesicles in the noradrenergic neurons. NE is released into the synaptic cleft by calcium-dependent exocytosis and is taken up into the axon terminal by an $Na^+/Cl^-$-dependent membrane transporter [7].

The life cycle of NE may be interfered as follows [2, 7, 28, 76]:

***T2.*** Enzymatic synthesis may be blocked by α-methyltyrosine, a competitive inhibitor of tyrosine hydroxylase. The dopamine β-

hydroxylase is blocked by Fla-63, a dithiocarbamate derivative.

***T3.*** The mechanism of uptake and storage of the catecholamines into synaptic vesicles is interfered by reserpine and tetrabenasine.

***T4.*** The release of NE is increased by amphetamine that blocks the reuptake mechanism.

***T7.*** The reuptake of NE is inhibited by the tricyclic antidepressive drug desipramine.

***T9.*** In the presynaptic boutons NE is degraded by MAO, which is inhibited by the antidepressive drug pargyline.

The opportunities to interfere with the life cycle of NE, inside the CNS, are restricted by the abundant distribution of the receptors in various peripheral territories, by the low ability of known adrenergic agonists and antagonists to cross the BBB, and by the lack of specificity of various compounds for postsynaptic central neuron receptors.

***Dopamine*** For a long time, dopamine was considered to be an intermediate in the biosynthesis of NE. The discovery of significant levels of dopamine inside the brain, especially confined to the basal ganglia, and thereafter important striatal dopamine depletion in the PD, led to the hypothesis that dopamine might be involved in the control of the motor system.

Dopamine is estimated to form over 80% of the total brain catecholamine content. The dopaminergic nerve cells are classified into A8–A15 neuronal populations and are located in the brain stem and the diencephalon. Functionally, they are organized into *short systems* of amacrine-like interneurons of the retina and olfactory bulb, *intrahypothalamic system* (tuberohypofisial, incertohypothalamic, medullary periventricular projections) and *long systems* that link substantia nigra and ventral tegmental area (the main dopamine sources of the brain) to the neostriatum, the limbic cortex, and other limbic structures (olfactory tubercle, nucleus accumbens septi, amygdalae, and piriform cortex) [7, 54].

The life cycle of dopamine may be interfered with as follows: the DOPA decarboxylase is inhibited by NSD-1015; the DA uptake and storage is inhibited by reserpine, tetrabenazine, and meperidine analog. Amphetamine increases dopamine release and blocks DA reuptake. Cocaine also blocks DA reuptake and stimulates the DA release. Nomifensine and glutamate-binding protein (GBP) are DA reuptake blockers only [7, 44, 51].

DA receptors may be classified into $D_1$ and $D_2$ subtypes. The $D_1$-like receptor family includes $D_1$ and $D_5$, which stimulate the activity of adenylyl cyclase (AC) and have high affinity for benzazepines (antagonist SCH 23390) and low affinity for benzamides. Of the two types, $D_5$ receptors have higher affinity for DA. The $D_2$-like receptor family comprises $D_2$, $D_3$, and $D_4$ receptor and inhibits AC. Compared with $D_2$, the $D_3$ receptor has a higher affinity for the antiparkinsonian drugs quinpirole and pergolide, as well as for DA autoreceptor inhibitors (+)-UH232 and (+)-AJ76. The $D_4$ receptor has a particularly high affinity for clozapine and is considered a rational target for antipsychotic drug design (Table 5) [2, 7, 44, 77–79].

The action of some drugs on dopamine turnover [2, 7, 44]:

1. *Major tranquilizers.* The phenothiazine and butyrophenone antipsychotic classes increase the brain dopamine turnover, blocking both dopamine autoreceptors and postsynaptic dopamine receptors, and it has been suggested that excessive mesolimbic dopaminergic activity might be involved in the pathogenesis of schizophrenia.
2. *Minor tranquilizers.* The anxiolytic drugs interact preferentially with mesoprefrontal dopaminergic neurons and their effect might be attributed to GABAergic facilitation.
3. *Antidepressants.* The MAO inhibitors increase the endogenous levels of both brain DA and NE. The tricyclic antidepressants interact with noradrenergic neurons, while the atypical antidepressants do not have a selective interaction with the dopaminergic system.

**TABLE 5 Dopamine Receptors Ligands**

| Receptor Family | Selective Agonist | Selective Antagonist | CNS Location |
|---|---|---|---|
| $D_1$ | SKF 38393, fenoldopam, dihydrexidine | SCH 23390, NNC-112 | Striatum, NAc, OT, SN |
| $D_2$ | Quinpirole, RU 24926, EMD 23448 | Haloperidol, pimozide, (−)-sulpiride, domperidone, spiperone, raclopride | Striatum, NAc |
| Autoreceptor | 3PP, EMD 23-448 | (+)-UH 232, (+)-AJ 76 | Somatodendritic membrane and terminal boutons of dopaminergic neurons |

### 6.6 Histamine

Histamine is a biogenic amine involved in various allergic and inflammatory reaction but also in behavior regulation and in the action of some psychotropic drugs. The histaminergic neurons are located in the posterior, tuberomamillary part of the hypothalamus and are divided into five subgroups designated $E_1$–$E_5$ [80]. These small cell clusters are connected afferently with the prefrontal cortex, the medial preoptic area, and the septum and send diffuse projections to other hypothalamic nuclei, the cerebral cortex, the hippocampus, the olfactory bulb, the amygdalae, the medial septum, the cerebellum, and the brain stem. The co-localization of GABA, substance P, galanin, met-enkephalin, and brain natriuretic peptide in the histaminergic neurons of tuberomamillary nucleus has also been demonstrated [81, 82].

In the brain, histamine is synthesized from histidine by the action of histidine decarboxylase and catabolized by histamine methylltransferase, and monoamine oxidase The histamine receptors are divided into three classes: $H_1$, $H_2$, and $H_3$ (Table 6).

The significance of brain histamine receptors for neuroprotective drug design includes the sedative properties of some $H_1$-blocking antihistaminic agents, the opioid-mediated analgesia related to $H_2$ receptor antagonists, and agonist/antagonist $H_3$ receptors manipulation in order to modulate arousal, appetite, seizure threshold, and some cognitive processes [54, 81].

## 7 NEW TARGETS AND APPROACHES FOR CNS THERAPY

### 7.1 Novel Insights into the Mechanism of Action of Known Drugs

Selegiline (*N*-propanyl-*l*-amphetamine; *l*-deprenyl) is a selective irreversible inhibitor of MAO-B used to treat PD. It has been demonstrated that its neuroprotective properties are also mediated by the suppression of free radical formation elicited by glutamate, upregulation of the antioxidative enzymes (SOD and catalase), induction of proteins interfering with the apoptotic pathway, and the modulation of the expression of neurotrophic factors and of GAPDH and other redox active molecules (thioredoxin). From this point of view, selegiline might represent a model for the development of new neuroprotective *N*-propargyl analog [83].

Minocycline, a second-generation, orally bioavailable and BBB-permeable tetracycline, exhibited remarkable neuroprotective properties due to direct inhibition of cytochrome c release and of the downstream events related to cell death (in particular, the activation of caspase 3 and caspase 1); inhibition of ischemia-induced nitric oxide synthase, of matrix metalloproteinase MMP9 during acute ischemic stroke, and of reactive microgliosis. These effects, demonstrated in mouse models of neurodegenerative disorders and ischemic and traumatic brain injury, are currently investigated in clinical trials for HD and ALS [8, 84]. Tetracycline-based antioxidants, such as 12*S*-hydroxy-1,12-pyrazolinominocycline, devoid of antibiotic activity, have also shown neuroprotective properties in oxidative stress-mediated complications [85].

### 7.2 NOS Inhibitors

In the acute ischemic brain, nitric oxide synthase (NOS) activity is greatly increased, and nitric oxide (NO) rapidly reacts with superoxide to form peroxynitrite, which is a powerful oxidant and highly cytotoxic [16]. It is assumed that NO released by neuronal isoform nNOS would play a major role in ischemic neurotoxicity. The inducible enzyme iNOS could be the major source of delayed NO release in the ischemic brain contributing to further neuronal injury, whereas endothelial eNOS would have a neuroprotective effect. Therefore, nNOS and iNOS seem to be the proper therapeutic targets for pharmacological manipulation of NO activity in acute cerebral ischemia in a defined time window [86].

The dual role of brain nitric oxide in neurodegeneration/neuroprotection and its complex involvement in both brain physiology (brain development, synaptic plasticity, learning, memory) and pathology, as well as the potential role of therapeutic target in stroke, neuroinflammation, and neurodegenerative disorders, are illustrated in Contestabile et al. [87], Hobbs et al. [88], and Mena et al. [89].

**TABLE 6 Histamine Receptors Agonists and Antagonists**

| Receptors | Agonist | Antagonist |
|---|---|---|
| $H_1$ | 2-(*m*-Bromophenyl)-histamine | Mepyramine, *trans*-triprolidine |
| $H_2$ | Amthamine, impromidine | Ranitidine, tiotidine |
| $H_3$ | (R)-α-methylhistamine, (R)-α,(S)-β-dimethylhistamine, imetit | Thioperamide, clobenpropit |

The rational design of isozyme-specific NOS inhibitors for ischemic stroke in humans is based upon a combined crystallographic and computational docking approach [90–92]. The neuroprotective effects of nitroindazole derivatives, especially 3-bromo-7-nitroindazole and 7-nitroindazole, are mainly due to selective inhibition of nNOS but also to MAO-B inhibition [93, 94] and have been demonstrated in animal models of stroke, PD, and brain injury [16, 95, 96]. Their neuroprotective efficacy is dependent on the dose, phase, and severity of cerebral ischemia-reperfusion injury and on ischemic brain acidosis [96–101]. For example, 7-nitroindazole has provided significant neuroprotection for hippocampal CA1 neurons during early reperfusion after transient forebrain ischemia [102] and has potentiated the anticonvulsant activity of phenobarbital, diazepam, and valproate [103].

The heterocyclic substituted amidine ARL 17477, as well as delta-(*S*-methylisothioureido)-L-norvaline (L-MIN), have been proven potent and selective inhibitors of nNOS, able to penetrate the BBB and to decrease the lesion volume after transient cerebral ischemia [104, 105]. Aminoguanidine derivatives are relatively selective iNOS inhibitors that have shown time- and dose-dependent neuroprotective effects, safety, and good tolerability after oral and intravenous administration in clinical trials [86].

### 7.3 Caspases Cascade Inhibition

Caspases are involved in the pathogenetic mechanism of neurodegenerative and neurological disorders as triggers of apoptotic neuronal death. Pharmacological inhibition of caspases activity, either by a broad caspase inhibitor or with semiselective inhibitors of caspase 1 and caspase 3, has proven neuroprotective effects only for neuronal cells from the penumbra of the ischemic region afflicted by apoptosis [8].

### 7.4 Calpains Inhibitors

Increased proteolytic activity of the classical calpains (μ-calpain or calpain-1 and *m*-calpain or calpain-2) has also been involved in neurodegenerative diseases and ischemic and traumatic brain injuries. Among the substrates intensively cleaved by classical calpains are (a) tau protein and other neurofibrillary tangles' proteins abundant in the brain of AD patients; (b) huntingtin protein, generating toxic fragments responsible for neuronal loss; (c) protein p35, important in neural tissue development, converted to p25, that is responsible for the hyperphosphorylation of the tau protein, the latter becoming highly resistant to μ-calpain degradation; (d) α-synuclein, the major component of Lewy bodies in PD; and (e) αII-spectrin, myelin, and other cytoskeletal proteins [106–108].

Synthetic peptidic and nonpeptidic calpain inhibitors have shown promising neuroprotective effects in animal models. In order to improve potency, cell permeability, and selectivity, the third generation of calpain inhibitors, such as peptidyl ketoamides, morpholine derivatives, benzodioxothiazines, and chromones, were developed. The α-mercaptoacrylic acids are nonpeptidic, reversible noncompetitive inhibitors from the fourth generation, with improved stability, selectivity, and pharmacological profile compared with the peptidic inhibitors. To overcome their modest ability to cross the BBB, calpain inhibitors were coupled with molecules (such as taurine, L-aminocarnitine) that have specific carriers in the brain [106, 109–111].

### 7.5 Other Targets and Neuroprotective Agents

The significance of other potential targets in CNS diseases and the related pharmacological neuroprotective attempts are exemplified in Table 7.

## 8 DEVELOPMENT OF NEUROPHARMACEUTICALS FOR ALZHEIMER'S DISEASE

### 8.1 Introduction

AD, the most common neurodegenerative disorder, currently afflicts over 8 million people worldwide, 4–5 million Americans, 5–10% of the population over the age of 65, and 30–50% over the age of 85 [5, 130]. Each year 360,000 new cases of AD are developing, and the incidence of AD will double in the next 25 years as the proportion of the elderly population increases. It is estimated that all over the world more than 20 million people will be affected and the annual economic expenses will exceed $100 billion [131]. In the highly developed industrial countries Alzheimer dementia is the fourth most common cause of death after heart disease, cancer, and stroke [132].

The clinical debut of the disease is insidious, with impairment of recent memory and attention, progressive deterioration of language, abstract thinking, judgment, and alteration of personality. In the final stage of illness the patients, severely demented, lose all psychomotor abilities including grasping and suckling reflexes.

AD is a global subcortico-cortical disconnection with the hippofunction of the basal forebrain cholinergic system due to neuronal loss in the basal nucleus of Meynert, the medial septal nucleus, and the nucleus of diagonal band Broca. Nerve cell death in these areas

**TABLE 7 Other Potential Targets and Therapeutic Approaches for CNS Disorders**

| Targets | Pathological Significance | Therapeutic Approaches | References |
|---|---|---|---|
| SOD | Oxidative stress-induced neurotoxicity | Extracellular-SOD mimetics; Cu/Zn SOD gene delivery | 112, 113 |
| JNK, p38MAPK, MLKs | Stress-activated protein kinases in neuronal death, upstream regulatory kinases of JNK | CEP-1347—the most advanced MLK inhibitor in clinical trials for PD | 114, 115 |
| CDK5 | Apoptosis and neurodegeneration | CDK5 inhibitors: flavopiridol roscovitine | 116 |
| GSK-3 | Increased activity in neuronal apoptosis, amyloid metabolism, tau protein hyperphosphorylation | GSK-3 inhibitors: lithium | 117, 118 |
| JAK2-STAT5 axis | Antiapoptotic, antioxidant effects | Neuroprotective recombinant human EPO (rh-EPO) | 119 |
| Cannabinoid system | Neurodegeneration | Cannabinoid receptors agonists; Potent and selective inhibitors of the endocannabinoid transporter | 120–123 |
| Neurosteroid nuclear receptor in Purkinje neurons | Promotion of dendritic growth, spinogenesis, synaptogenesis, and neurogenesis | Neurosteroid replacement therapy: allopregnanolone (3α,5α-tetrahydroprogesterone) | 124–126 |
| Adenosine receptors | Striato-pallidal neurons degeneration | Neuroprotective adenosine $A_{2A}$ antagonists | 127 |
| Ubiquitin-proteasome system | Neuronal development, synaptic plasticity | Inhibition of ubiquitin-rich protein aggregates | 128 |
| GLP-1 receptors | Neurotrophic properties on cholinergic neurons in the basal forebrain of the rat against Aβ and glutamate-induced apoptosis | GLP-1 agonists | 129 |

determines cholinergic denervation of amygdala, entorhinal, dentate, and hippocampic cortices and triggers a *domino effect* with important neuronal loss, anterograde in frontal and temporal neocortex, and retrograde in locus coeruleus, raphe nuclei, and other brainstem nuclei.

The diagnosis of AD is based on characteristic clinical features associated with distinctive histopathological lesions of neuronal cytoskeleton, the neurofibrilary tangles (NFT), and abnormal deposits of amyloid into extracellular matrix, senile plaques (SP), neuropil threads (NTh), Hirano bodies, and granulovacuolar bodies.

### 8.2 Pathogenesis

Despite its ethiological heterogeneity, therapeutic strategies arise from understanding the molecular bases of pathogenesis. It is difficult to identify an outstanding *primum movens,* precocious and singular, that could mark the disease onset.

***Genetic Defect*** For decades it has been known that 10–15% of cases of AD occur in an autosomal-dominant pattern and also that typical histopathological lesions AD-like were found in virtually all subjects with trisomy 21 (Down) who survived until their forties. Correlating this information, research focused on supposed genetic factors and demonstrated point mutation in the gene that encodes β-amyloid precursor protein (βAPP). This disease-causing gene is located on the long arm of chromosome 21 and induces the increased production of Aβ42 peptides or total Aβ peptides, thus triggering the amyloid cascade [133]. This hypothesis was supported by the discovery in the mid-1980s of the diffuse Aβ deposits in the hippocampus, other limbic and association cortices, without neuronal or glial damage, in the brains of patients who had died of Down syndrome aged 10–20 years [54].

Shortly after the discovery of βAPP gene mutation there were identified *presenilin 1 (PS1)* and the highly homologous gene *presenilin 2 (PS2)*, located on the chromosomes 14 and 1, respectively. Missens mutations of these genes increase Aβ42 peptide production and characterize the early onset of familial AD [134, 135].

Further biochemical studies identified apolipoprotein E (ApoE), a cerebrospinal fluid protein capable of binding Aβ amyloid. The ε4 polymorphism of the ApoE

gene significantly increased the risk of late-onset familial AD development [136].

The candidate genes in AD are located near a well-established linkage peak on chromosome 9q22 and encode βAPP, presenilin 1 (PS1) and PS2, *tau* protein, and ubiquilin UBQLN1, respectively. Genetic polymorphism of APP, PS1, and PS2 is correlated to the early onset of familial AD, and the APOE4 genotype is associated with a shift to an earlier age at onset [12, 137]. Tau gene mutation causes frontotemporal dementia, progressive loss of neuronal cells, atrophy of the forebrain, and memory impairment [138]. There is a significant association between AD and various single-nucleotide polymorphisms in UBQLN1, a gene that interacts with PS1 and PS2 and is involved in proteasome degradation of proteins [12, 139, 140].

***Hypocholinergic Cascade*** The role of ACh in learning, memory, and attention was demonstrated both experimentally and clinically [140a, 141]. The key role in the molecular aspects of ACh action on the postsynaptic membrane is played by the muscarine receptors that act via G proteins and also participate in the βAPP processing by the following mechanism:

The $M_1$, $M_3$, and $M_5$ receptors activate the phospholipids that catalyze the hydrolysis of phosphatidylinositol with diacylglycerol (DAG) formation. DAG activates the PKC responsible for APP phosphorylation, thus triggering the secretion of the physiological, trophic form of Aβ.

Cholinergic synapses depletion determines hypofunction of muscarinic mChR receptors, perturbation of the mChR/G protein couple function, abnormal APP phosphorylation, increased secretion of Aβ toxic form, and its deposition as senile plaques. The restoring of the mChR function could prevent the abnormal phosphorylation and seems to be a positive neuroprotective perspective.

Emerging evidence suggests that AChE may play a role in abnormal neurite projection growth and also mediates the in vitro aggregation of β-amyloid protein, the major component of AD senile plaques [142].

***Amyloid Cascade*** Amyloid in senile plaque (SP) is composed by a portion of an integral membrane glycoprotein, the βAPP, that exists in three isoforms containing 695 ($APP_{695}$), 751 ($APP_{751}$), or 770 ($APP_{770}$) amino acids. The APP is involved in cell development, growth, and adhesion. It follows normal membrane dynamics and is processed by proteolysis into different fragments of the initial molecule [7]. The Aβ peptide is released by the action of γ secretase. The Aβ is a 40- to 42-amino-acids peptide released in a soluble form by normal neurons and glial cells and is present in the CSF and plasma, which suggests the existence of a brain-blood clearance [143].

The amyloid cascade is triggered by various genetic, endogen, and exogen factors that influence APP synthesis and processing, enhance the Aβ secretion, and reduce its clearance. Dysfunction of the Aβ clearance mechanism (receptor-mediated Aβ transport across the BBB and enzyme-mediated degradation) leads to Aβ accumulation in the brain of patients with sporadic or late-onset AD. Efflux of Aβ from brain to blood is mediated by lipoprotein receptor-related protein (LRP) and P-glycoprotein [144, 145]. Excess amounts of Aβ down-regulate LRP receptors and, on the other hand, significantly upregulate by a positive-feedback mechanism the expression of the receptors for advanced glycation end products (RAGE), which mediates Aβ influx from blood to brain in AD patients and in APP transgenic models of AD. Experimental and epidemiological studies clearly show the correlation between neprilysin and sequence variants of the insulin-degrading enzyme (IDE), the principal zinc metalloendopeptidases degrading Aβ, with the extent of the amyloid deposits in late-onset AD [145].

These dysfunctions determine the insolubilization of AD, β-folded oligomeric structure fibrillization (β sheet), and deposition as SP, with neurotoxic effects, neurotransmitter deficits, neuronal death, and all clinical signs of dementia [146, 147].

***Tau Protein Cascade*** Tau is a soluble microtubule binding protein that, together with tubulin, promotes the stability and assures the normal cycle of microtubule assembly and disassembly. Abnormal phosphorylation of tau protein results in insoluble paired helical filaments (PHFs) that are made up of hyperphosphorylated isoforms [148, 149]. The PHFs conglomerate as neurofibrillary tangles (NFT) in the neuronal soma, as neuropil threads (NTh) in the dendrites, and the dystrophic axons develop bulbous terminals [150]. Thus, the axonal transport is impaired and the synaptic output is diminished, compromising the viability of the neuron. As neurons die, the remaining NFTs mark like gravestones the site of lost nervous cells. The externalization of the NTFs after the neuronal death significantly increases the insolubility of SP.

***Excitotoxic Cascade*** The neurodegenerative effect of Aβ is exerted through the increasing of excitatory amino acid (EAA) amount as a result of glutamate receptors (GluR) decrease in the hippocampus [151]. This activates the acute excitotoxicity mechanisms that start with an initial calcium-independent process when magnesium ions block NMDA receptor, and continues with cytosolic calcium ion increase, hyperactivation of enzy-

matic systems (proteases, lipases, and nucleases), neuronal lysis, endogenous glutamate release, and the additional hyperactivation of glutamate receptors that amplify the excitotoxic phenomenon. The magnesium ions dissociation from NMDA receptors and important influx of calcium ions might be due to the disruption of the mitochondrial function and low production of ATP, typical mechanism for long-term neurodegeneration [152].

AD evolution and progressive degradation of cholinergic and glutamatergic systems may be partially compensated by activating other neurotransmitter systems as GABA and monoaminergic systems.

***Oxidative Cascade*** The disruption of cytoskeleton determines dyshomeostasis of cytosolic calcium, initiation of lipid peroxidation (LP), generation of reactive oxygen species (ROS) [153, 154], mitochondrial destruction, and energy depletion of the neurons. Numerous studies demonstrate that ROS are generated directly from the amyloid plaques and NTF or by activation of surrounding microglial cells [155]. Intracellular calcium released from the destroyed cytoskeleton increases ROS production and activates NOS. The role of NO is ambiguous because by formation of peroxynitrite it has a cytotoxic effect acting synergically with ROS, yet its neuroprotective effect has also been demonstrated [156]. It is also known that copper and iron dyshomeostasis have a substantial contribution to the oxidative neuronal damage [157].

***Apoptotic Cascade*** Apoptosis is a crucial physiological mechanism, tightly regulated, that maintains the equilibrium between cell division and cell death [158]. The molecular biochemical phenomena of apoptosis recognize triggers, activators, and amplifiers of endogenous regulation. It passes through premitochondrial, mitochondrial, and postmitochondrial phases. The last one is triggered by caspases and by calpains, distinct families of cysteine proteases. Caspases are cell death executors that inactivate the vital cell proteins and cleave the cytoskeletal proteins, stop DNA replication, destroy the structure of nucleus, and disintegrate the cell into apoptotic bodies. The other family of cysteine proteases, the calpains, naturally inhibited by calpastatine, increase proteolysis of neurofilament proteins during neuronal apoptosis [8, 108]. The interruption of apoptotic cascade and restoration of the mitochondrial function are future targets for anti-AD drugs design.

### 8.3 Therapeutic Strategies

The novel strategies in the design of new therapeutic agents for treatment and prevention of AD are based on molecular neurochemical dysfunctions, and they must be focused on the substitution of the hypofunctional cholinergic system, the activation of other neuroprotective neurotransmitter systems indirectly affected, the interference of Aβ metabolism and of the NFT formation, and the antitoxic, antioxidative, and anti-inflammatory neuroprotection [159, 160].

***Substitutive Cholinergic Therapy*** The first target of anti-AD therapeutic strategy is the primary affected cholinergic system. The compensation of ACh deficit was approached as follows:

**The increasing of ACh synthesis** is based on the *therapeutic augmentation of the ACh precursor choline*. Exogenous choline, phosphatidylcholine, and lecithine have stimulating memory properties and from this group today are in use Gliatilin (Italfarmaco) and Brezal (Sandoz) [161]. The other agent, now in clinical trials, acetyl-L-carnitine has the advantage of *stabilizing the mitochondrial function and lipid metabolism* [162].

**The stimulation of ACh release** may be realized by very different mechanisms. Thus linopirdine (DuP 996), now in clinical trials, acts through the *selective inhibition of presynaptic potassium channels* [163]. From a series of 4-substituted imidazole derivatives that demonstrated antidementia properties in animal models [164, 165], clobenpropit is a H3 *histamine receptor (HR) antagonist,* while Thioperamide (Glaxo) and GT-2016 (Gliatech) act as *H3-HR inverse agonists*. Iodoproxyfan, UCL-1390, and AQ-0145 agents, as well as imidazole derivatives, have similar effects [82, 166]. *Presynaptic choline reuptake* and *ACh resynthesis* is stimulated by the Mitsubishi Chemical Product MKC-231, which also has a neuroprotective effect by suppressing NO formation triggered by calcium influx.

**Acetylcholinesterase inhibitors** (AChEI) are, historically, the first and most widely experimented therapeutic attempt. Theoretically, the inhibition of AChE in the synaptic cleft results in a *potentiation of cholinergic receptor* activity, but it is dependent on the integrity of presynaptic neurons, and it will be proportionally reduced with cholinergic neuron loss in the late stages of AD.

Acetylcholinesterase inhibitor therapy in AD is characterized by the exclusive inhibition of the catalytic activity without influence on profound morphogenic effects on neuronal and synaptic architecture, lack of selectivity on AChE isoforms, the potential to aggravate certain conditions by elevating the levels of catalytically inactivated AChE (the rare isoform AchE-R "read-through") via the feedback loop, and individual differences in sensitivity and therapeutic response [28].

The first generation of AChEI Tacrine (Cognex), Amiridine (NIK-247), and physostigmine lack brain

specificity inhibiting another enzyme (butyrylcholinesterase) and have important side effects, such as hepatotoxicity and sedation. Their clinical efficacy (slowing decline in executive functions more than memory and cognition) is about 20–30% [167]. (−)-Phenserine tartrate, the phenylcarbamate derivative of (−)-physostigmine, is a selective, noncompetitive, potent, and long-lasting inhibitor of brain AChE, which reduced the secretion of Aβ in cell lines by translational regulation of βAPP mRNA via a noncholinergic mechanism. (−)-Phenserine tartrate has produced significant improvements in cognitive function and is well tolerated in patients with AD, currently being investigated in phase III clinical trials [168].

The second generation [169] of AChEI, designed to reduce the side effects and to increase brain specificity, includes rivastigmine, ENA 713 [170] (Novartis), Aricept/Donepezil [171] (Pfizer), galanthamine [172], and eptastigmine [173] (Jenssen and Mediolanum). These drugs are especially effective in the incipient stages of AD. A comparison of the above-mentioned drugs with more potent and efficacious therapeutic candidates that are currently in clinical trials or are ready to go into clinical trials, and the medicinal chemistry approaches toward designing better pharmaceuticals, is reviewed in Suh et al. [174].

Bayer realized a proform of dimethyl-2,2-dichlorvynilphosphate called metrifonate with slow release of the active component and long-term inhibition of AChE. Molecular modeling, docking, and ADMET studies were applied to the design of novel molecular hybrids of tacrine fused with donepezil as AChEI for future treatment of AD [175]. The design of more potent and selective inhibitors of choline acetyltransferase might be guided by the 3D quantitative structure–activity relationship (3D-QSAR) studies carried out using the comparative molecular field analysis (CoMFA), molecular docking scores (FlexiDock, FlexX), and comparative molecular similarity indices analysis (CoMSIA) methods [176, 177]. Recently, it was demonstrated that AChE accelerates fibrillization of the soluble Aβ and forms with Aβ a neurotoxic complex [178, 179]. These findings reveal a new mechanism of AChEI therapeutic activity [180].

**The ligands of muscarinic cholinergic receptors (mChR)** might be the source of anti-AD medicaments due to mChR involvement in the APP processing. The literature reports as potential antidementia drugs some quinuclidine derivatives [181]. These are either *agonists of $M_1$ subtype of mChR* such as AF-102-B and SB 202026 agents (Sabcomeline/Memric) or *$M_1$ agonist/$M_2$ antagonist* such as agent Lu 25-109. All these ligands showed therapeutic effects only in animal models of AD. Xanomeline (LY 246708), selective agonist of $M_1$ and $M_4$ mChR, is a more promising agent that has now entered clinical trials.

**The ligands of nicotinic cholinergic receptors (nChR)** are very important for cognitive processes of the limbic system because they are involved in the formation of working memory [182]. They are significantly decreased in AD and strategies using nChR ligands as potential antidementia drugs are now developing. A *selective agonist of α4β2 of nChR*, ABT-418, produced by Abbott Laboratories [183], has entered clinical trials and its analogs A-92824, A-84543, and A-85380 are also being investigated. Other nChR agonists, nefiracetam [184] and SIB 1553A [185], are also proposed as potential anti-AD drugs.

***Neuroprotective Properties of Other Neurotransmitter Systems*** Activation of these systems is an alternative therapeutic strategy that may act *synergistically with the direct cholinergic compensation* described above. The literature describes various compounds with significant antidementia effects that act particularly on GABAergic, monoaminergic, and glutamatergic systems.

Pronounced cognitive function improvement have been shown: valproic acid, *GABA enhancing* anticonvulsant agent that prevents calcium dyshomeostasis and cytoskeletal lesions, CGP-36742 (Ciba) *GABA-B receptor antagonist*, and Suritozol (Hoechst) *inverse agonist of GABA-A receptors* at benzodiazepine site [186–189].

The *inhibitors of monoamine oxidase B* (MAO-B) deprenyl and lazabemide, used in clinical practice as antiparkinsonian agents, have also demonstrated antitoxic action against βAPP, possible inhibition of NOS and antioxidant effects [189a, 190, 191].

In experimental animal models 8-OH-DPAT, a *selective agonist of 5-$HT_{1A}$ serotonin receptors*, and ondansetron, a *selective antagonist* of the same receptors, demonstrated antidementia protection via antiexcitotoxic effects and increasing of brain extracellular ACh [54]. Recent in vitro and in vivo pharmacological and biochemical studies have proven the role of 5-$HT_4$ receptors in cognitive processes and the APP processing, thus *selective 5-$HT_4$ agonists* may represent a new approach for the treatment of AD [192].

The involvement of the *glutamatergic system* in AD pathogenesis is suggested by the decreasing of hippocampal glutamate receptors (GluR) and by the ability of endogenous glutamate to induce the excitotoxicity resulting in neuronal death.

Memantine [194], *antagonist of NMDA receptors* and *agonist of AMPA* receptors, presented moderate anti-AD efficacy, but its specificity for phencyclidine site of receptors associated with psychomimetic effects restricts its use. In animal models, the agent CX516 belonging to

the ampakines group displayed cognition stimulating properties. D-Cycloserine, agonist of the glycine B site of the NMDA receptor, has the same promising properties on animal models [193]. Among the agents targeted at the glutamatergic system with favorable results in experimental AD models are sabelusole, which blocks glutamate-induced calcium influx and stabilizes the neuronal cytoskeleton, the natural glycoside gastrodin extracted from *Gastrodia elate*, which is a memory enhancer used in Chinese medicine, and taurine, which also blocks glutamate-induced calcium influx.

Small molecules that mimic the biological activity of NO, including NO donors, may ameliorate cognition and provide neuroprotection. GT 1061 is an NO chimera, an NO mimetic compound that contains an ancillary, synergistic pharmacophore currently in clinical trials. NO chimeras and hybrid nitrates hold promise as therapeutics for AD with multiple sites of action [195, 196].

***Interference of Amyloidic Cascade*** **Interference of Aβ formation** is the most rational therapeutic target to block the cascade of the neurodegenerative process. As the processing of APP regulated by mChR involves the proteolytic activity of β- and γ-secretases, the search of direct specific blockers seems the most attractive, but information is still missing [197–199]. The normalization of APP cleavage by stimulation of metabotropic glutamate receptors may be also used to prevent the development of Aβ [200].

Recently, monensin and brefeldin A were reported to decrease Aβ formation by destabilization of pH gradient and of vesicular transport. Bafilomycin A blocks the lysosomal acetyltransferase and indirectly inhibits the β-secretase activity, and leupeptine, novel isocoumarin JLK inhibitors, and compound E-64 are acting by inhibition of γ-secretase [201, 202]. The synthetic derivative DAPT reduces in vivo the brain level of Aβ.

Reduction of serum cholesterol level and normalization of cholesterol metabolism via the inhibition of *3-hydroxy-3-methyl-glutaryl-coenzymeA (HMG-CoA) reductase* was demonstrated to decrease the production of Aβ, and the statin agents (lovastatin, pravastatin, simvastatin, atorvastatin) are today used to decrease the prevalence of AD [203–207]. It was also attempted to delay or reverse the fibrillization of Aβ by interfering with the formation of β-folded aggregates [147, 208]. The functional role of antiparallel β sheets in the insoluble Aβ42 structure and the application of β-turn mimicking strategies in the design of novel therapeutic agents are discussed in Brakch et al. [209].

The aggregation and deposition of Aβ fibrils in the brain may be decreased by various compounds such as the glycoprotein laminin and laminin-related peptides [210, 211], Paroxetine [212], the antibiotics rifampicin, erythromycin, tetracycline, or daunomycin, the last binding to Aβ and blocking the amyloid fibrils formation [213, 214], or the antioxidant *N*-acetyl cysteine. Their chemical structures might be employed as starting points for AD drug design and development [212, 215]. Inhibition of Aβ fibril formation was also demonstrated for a newly synthesized benzofuran series, particularly for the compound SKF 74652 [214]. Recent studies have shown that short peptides like $A\beta_{15–22}$ prevent Aβ self-aggregation and may be useful in AD therapy [216].

Other targets of special interest emerge from the intervention of ApoE and ApoJ in the endogenous mechanism of amyloid fibril formation. Involved in the metabolism of lipids and lipoproteins, ApoE has several isoforms (E2, E3, and E4) and the allele *ε4* of the *APOE* gene is known as a genetic risk factor for AD. The neurodegenerative mechanism triggered by *ε4* involves the diminution of the Aβ efflux from neuron to extracellular space [217, 218]. Since the ApoJ protein stimulates the accumulation of the trophic form of amyloid, it is very attractive to influence the balance of ApoE/ApoJ in order to normalize the Aβ metabolism.

*Enhancing Aβ Clearance* Potential therapeutic strategies developed for enhancing the Aβ clearance from AD brain comprise: (a) promoting receptor-mediated Aβ efflux either by RAGE-specific IgG or statins (simvastatin, lovastatin) [219]; (b) upregulating enzyme-mediated Aβ degradation through neprilysin or IDE gene therapy and pharmacological activation (somatostatin and somatostatin receptor agonists) [220]; (c) immunotherapy [145]; and (d) peripheral treatment with Aβ-bindable substances (gelsolin, curcumin, enoxaparin), which favor soluble Aβ species much more easily eliminated [221].

*Immunotherapy* is another approach of AD therapy because it suppresses the formation of β-amyloid, dissolves already formed fibrillar amyloid, and delays learning and memory deficits in mice [222–224]. Active immunization with $A\beta_{1–42}$ in humans was stopped in phase II clinical trials due to frequently developed adverse reactions, such as autoimmune meningoencephalitis and cerebral hemorrhage [145, 222].

New $A\beta_{1–15}$ or $A\beta_{42}$ peptides or Aβ DNA vaccines and antibodies toward the EFRH sequence (glutamate-phenylalanine-arginine-histidine sequence) improve memory and cognition in AD animal models and in phase IIa clinical trials, and alleviate neuroinflammatory adverse reactions of classical immunotherapy [145, 225–227]. Passive immunotherapy with antibodies against Aβ, also endowed with α-secretase-like activity, reduces Aβ concentrations in CSF, improves cognitive function, and has proven to be much safer than the

active immunization and is in a phase I clinical trial in the United States [228, 229].

*Inhibition of NFT Formation* Since NFT, the other prominent AD morphopathological event, involves the abnormal hyperphosphorylation of microtubule-associated tau protein (MAP), the logical neuroprotective strategy seems to be either the inhibition of tau-phosphorylase/proteinkinase or the activation of tau-phosphatase [149]. The information concerning such specific compounds is missing in the literature, lithium salts only being mentioned as inhibitor of glycogen synthase kinase-3, which catalyses tau protein phosphorylation [230]. Small-molecule inhibitors of tau aggregation, able to cross the BBB and to improve cytoskeletal integrity and neuron viability, such as a family of heat shock protein HSP90 inhibitors, might be identified by HTS [80, 150].

*Interference of Aβ Neurotoxic Effect* As has been shown, the intimate mechanisms of Aβ neurotoxicity are closely related to calcium dyshomeostasis, and the attempts to stabilize the calcium homeostasis of neuron would have beneficial effects on neuronal survival. The anticalcium strategy is thus justified and agents such as nimodipine, blocker of L-type calcium channels and the antagonists of ryanodine receptors that regulate the levels of cytosolic calcium, are waiting for clinical trials [231].

***Antioxidant Therapy*** For a long time the theory of oxidative stress has occupied a central place in the ethiology of AD and offers strong arguments for antioxidant therapy for this type of dementia [232]. The morphologic intact neuron has *endogenous systems* that protect it from higher than normal NOS levels and stimulation of these systems is an opportune target for AD treatment. Recently, it has been shown [233] that the nuclear transcription factor NF-κB, which increases during oxidative stress, is able to induce the transcription of genes encoding catalase and glutathion peroxidase, enzymes that protect the cell against Aβ neurotoxicity.

An alternative neuroprotective strategy should be the application of external antioxidants that block ROS action. The presence of functional groups able to accept oxygen radicals or free electrons and the penetrance through the BBB are the most important conditions for a molecule to be an antioxidant. Vitamin E and its synthetic analog idebenone have been demonstrated to delay the involution of AD. Other synthetic analogs of α-tocopherol, IRF 10016 and MDL 74180DA, also have antioxidant properties now being tested in preclinical studies [232a].

Another antioxidant described in the literature as a promising agent for dementia therapy is the extract of *Ginkgo biloba* ECB761, which neutralizes NO radicals and exhibits neuroprotective properties in animal models. Celastrol, herbal tripterine, has strong antioxidant and anti-inflammatory activity and significantly stimulates learning and memory in animals with experimentally induced dementia [234].

Antioxidant properties are displayed by steroid hormones and their derivatives, which activate specific hormonal receptors, modulate neurotransmission, and may be synthesized in the neurons. The most widely studied are estradiol and its derivatives [235, 236], Estraderm (Novartis) and J811 and J861 (JenaPharm), which have pronounced neuroprotective properties and enhance cognition [237–239].

Melatonin, the hormone of the pineal gland, interacts with hydroxyl radicals and through antioxidant effects regulates the processing of APP, inhibits the formation of amyloid fibrils, and protects the neurons against excitotoxicity. *N*-acetylserotonin (NAS), the melatonin precursor, displays higher antiamyloid activity and radical-scavenger properties [240–242].

***Interference of Apoptotic Cascade*** As described above, the associated mitochondria peptides regulatory factor controls apoptosis by acting on mitochondrial calcium. The increased calcium influx into mitochondria during apoptosis may be blocked by the phenylhydrasone derivative p-trifluoromethoxy carbonyl cyanide phenyl hydrazone (FCPP), which acts as protonophor [243]. The muscle relaxant agent dantrolen demonstrated antiapoptotic activity modulating the interaction mitochondria/antiapoptotic protein Bcl-2 [244]. Bongkrekate and cyclosporine A inhibit the proapoptotic peptide Bax and the compound ALCAR (acetyl-L-carnitine) stabilizes the cell membranes and mitochondrial energy production [191, 245].

***Anti-inflammatory Therapy*** Anti-inflammatory therapy is targeted at the late phase of AD pathogenesis when cytokines IL-1 and IL-6 activate the complement cascade, and anaphylatoxin activates microglia. The cyclo-oxigenases (COXs), prostaglandin endoperoxide H synthase, and lipoxygenases act on arahidonic acid and generate eicosanoids (prostaglandins and leucotrienes), the very active mediators of inflammation. All these processes may potentiate the neurotoxic effect of Aβ, trigger ROS formation, and activate microglia.

It has been demonstrated that the nonsteroidal anti-inflammatory drugs ibuprofen and naproxen [246] decrease the risk of AD by nonselective inhibition of COXs. Rofecoxib (Merck), selective inhibitor of COX-2, is also investigated as a potential anti-AD drug.

Ibuprofen, indomethacin, and flurbiprofen, by combining anti-inflammatory and antiamyloidogenic properties, have proven to have direct Aβ-lowering effects in cell cultures as well as transgenic models of AD-like amyloidosis [247]. A novel, integrative chemical/biological approach for the rapid discovery of potential lead compounds targeting neuroinflammation is described for synthetic libraries containing pyridazine molecular fragment, allowing the identification of orally bioavailable compounds with the desired physical properties and in vivo activity [248]. Among other anti-inflammatory agents with promising anti-AD effects, prednisone, hydroxychloroquine, and colchicine are mentioned in the literature [249–251].

## Abbreviations

| | |
|---|---|
| Aβ | β-amyloid |
| AC | adenylyl cyclase |
| ACh | acetylcholine |
| AChE | acetylcholinesterase |
| AChEI | acetylcholinesterase inhibitors |
| AD | Alzheimer disease |
| ADMET | absorption, distribution, metabolism, toxicity |
| ADNF | activity-dependent neurotrophic factor |
| ADNP | activity-dependent neuroprotective protein |
| ALS | amyotrophic lateral sclerosis (Lou Gehrig's disease) |
| AMPA | α-amino-3-hydroxy-5-methylisoxazole-4-propionic acid |
| AP-5 | 2-amino-5-phosphonopentanoic acid |
| AP-7 | 7-amino-phosphonoheptanoic acid |
| ApoE | apolipoprotein E |
| ApoJ | apolipoprotein J |
| βAPP | β-amyloid precursor protein |
| 3-APPA | 3-aminopropylphosphinic acid |
| BBB | blood–brain barrier |
| BDNF | brain-derived neurotrophic factor |
| CaM-KII | calcium/calmodulin-dependent protein kinase II |
| 5-CAT | 5-carboxamidotryptamine |
| CDK5 | cyclin-dependent kinase 5 |
| CNQX | 6-cyano-7-nitroquinoxaline-2,3-dione |
| CNS | central nervous system |
| CNTF | ciliary neurotrophic factor |
| CoMFA | comparative molecular field analysis |
| CoMSIA | comparative molecular similarity indices analysis |
| COXs | cyclo-oxigenases |
| CPP | 3-(2-carboxypiperazin-4-yl)-propyl-1-phosphonic acid |
| CSF | cerebrospinal fluid |
| DA | dopamine |
| DAG | diacylglycerol |
| 4-DAMP | 4-diphenylacetoxy-*N*-methylpiperidine |
| DATP | *N*-[*N*-(3,5-diflorophenacetyl)-L-alanyl]-*S*-phenylglycine, *t*-butyl ester |
| DNA | deoxyribonucleic acid |
| DNQX | 6,7-dinitroquinoxaline-2,3-dione |
| DOB | 2,5-dimethoxy-4-bromoamphetamine |
| DOI | 2,5-dimethoxy-4-iodoamphetamine |
| 3D-QSAR | 3D quantitative structure–activity relationship |
| EAA | excitatory amino acids |
| ECB | extract of *Ginkgo biloba* |
| ECF | interstitial or extracellular fluid |
| EFRH | sequence-glutamate-phenylalanine-arginine-histidine sequence |
| EGF | epidermal growth factor |
| EPSPs | excitatory postsynaptic potentials |
| FCCP | p-trifluoromethoxy carbonyl cyanide phenyl hydrazone |
| FGF-1, -2 | fibroblast growth factors |
| FR | reticular formation |
| GABA | γ-aminobutyric acid |
| GABA-T | GABA aminotransferase |
| GAD | glutamate acid decarboxylase |
| GAPDH | glyceraldehyd-3-phosphate dehydrogenase |
| GBP | glutamate-binding protein |
| GDNF | glial cell-line-derived neurotrophic factor |
| GLP-1 | glucagon-like peptide-1 |
| GluR | glutamate receptors |
| GSK-3 | glycogen synthase kinase-3 |
| HA-966 | 3-amino-1-hydroxypyryolid-2-one |
| HAT | histone acetyl transferase |
| HD | Huntington's disease |
| HGF | hepatocyte growth factor |
| HIR | human insulin receptor |
| HMG-CoA | 3-hydroxy-3-methyl-glutaryl-coenzyme A |
| 5-HT | 5-hydroxytryptamine |
| HTS | high-throughput screening |
| IDE | insulin-degrading enzyme |
| IFN-β | interferon-β |
| IGF-I, -II | insulin-like growth factors |
| IL-1 | interleukin-1 |
| $IP_3$ | phosphatidylinositol triphosphate |
| IPSPs | inhibitory postsynaptic potentials |
| JNK | c-Jun N-terminal kinases |
| L-687,414 | R(+)-cis-β-methyl-3-amino-I-hydroxypyrrolid-2-one |

| | |
|---|---|
| L-AP4 | L-2-amino-4-phosphonobutanoic acid |
| LIF | leukemia inhibitory factor (or cholinergic differentiation factor) |
| LP | lipid peroxidation |
| LRP | lipoprotein receptor-related protein |
| LSD | lysergic acid diethylamide |
| LTP | long-term potentiation |
| MAb | monoclonal antibody |
| MAO-B | monoamine oxidase B |
| MAOIs | monoamine oxidase inhibitors |
| MAP | microtubule-associated tau protein |
| mChR | muscarinic cholinergic receptors |
| *m*-CPBG | *m*-chlorophenylbiguanide |
| MCPG | α-methyl-4-carboxyphenylglycine |
| MLKs | mixed lineage kinases |
| MMP 2, 9 | matrix metalloproteinases |
| MNQX | 5,7-dinitroquinoxaline-2,3-dione |
| mRNA | messenger RNA |
| MS | multiple sclerosis |
| NAc | nucleus accumbens septi |
| NAP | nucleosome assembly protein |
| NAS | *N*-acetylserotonin |
| nChR | nicotinic cholinergic receptors |
| NFT | neurofibrillary tangles |
| NF-κB | nuclear transcription factor |
| NGF | nerve growth factor |
| NMDA | *N*-methyl-D-aspartic acid |
| NNC-112 | 3-methyl-1-(7-benzofuranyl)-2,3,4,5-tetrahydro-8-chloro-7-hydroxy-(1H)-3-benzazepine |
| NO | nitric oxide |
| NOS | nitric oxide synthase (endothelial eNOS, neuronal nNOS, inducible iNOS) |
| NT | neurotrophin |
| NTAK | Neural- and Thymus-derived Activator for ErbB Kinases |
| NTh | neuropil threads |
| 8-OH-DPAT | 8-hydroxy-dipropylaminotetraline |
| OGN | (antisense) oligonucleotides |
| OT | olfactory tubercle |
| p38MAPK | p38 mitogen-activated protein kinase |
| PD | Parkinson's disease |
| PDGF | platelet-derived growth factor |
| PHFs | paired helical filaments |
| PINK1 | PTEN-induced putative kinase-1 |
| PKC | protein kinase C |
| PNA | peptide nucleic acids |
| PNS | peripheral nervous system |
| PS1, PS2 | presenilin 1, presenilin 2 |
| RAGE | receptors for advanced glycation end products |
| RMT | receptor-mediated transcytosis |
| RNA | ribonucleic acid |
| RNase H | ribonuclease H |
| ROS | reactive oxygen species |
| SCH 23390 | 3-methyl-1-phenyl-2,3,4,5-tetrahydro-7-chloro-8-hydroxy-(1H)-3-benzazepine |
| SKF 38393 | 1-phenyl-2,3,4,5-tetrahydro-7,8-dihydroxy-(1H)-3-benzazepine |
| SN | substantia nigra |
| SOD | superoxide dismutase |
| SP | senile plaques |
| SSADH | succinic semialdehyde dehydrogenase |
| SSRI | selective serotonin reuptake inhibitors |
| TBPS | *t*-butylbicyclophosphorothionate |
| TCA | tricyclic antidepressants |
| TfR | transferrin receptor |
| TGF-α, -β | transforming growth factor-α, -β |
| TNF-α | tumor necrotic factor-α |
| UBQLN1 | ubiquilin |
| UCH-L1 | ubiquitin C-terminal hydrolase L1 |

## REFERENCES


1. Pardridge MW (2001). *Brain Drug Targeting. The Future of Brain Drug Development*. Cambridge University Press. Cambridge, MA.
2. Block HJ, Beale JM (2004). *Wilson and Gisvold's Textbook of Organic Medicinal and Pharmaceutical Chemistry*. Philadelphia: Lippincott Williams & Wilkins.
3. Leach AR, Brycel RA, Robinson AJ (2000). Synergy between combinatorial chemistry and de novo design. *J Mol Graphics Model 18*(4–5), 358–367.
4. Pardridge WM (2005). The blood-brain barrier: Bottleneck in brain drug development. *NeuroRx 2*, 3–14.
5. Tuszynski HM, Kordower JH (1999). *CNS Regeneration. Basic Science and Clinical Advances*. San Diego and London: Academic.
6. Dahlström A, Fuxe K, (1964). *Acta Physiol Scand 62*, Suppl. 232, 1.
7. Feldman RS, Meyer JS, Quenzer LF (1997). *Principles of Neuropsychopharmacology*. Sunderland, MA: Sinauer Associates.
8. Friedlander RM (2003). Apoptosis and caspases in neurodegenerative diseases. *New Engl J Med 348*(14), 1365–1375.
9. Sadri-Vakili G, Jang-Ho J (2006). Histone deacetylase inhibitors: A novel therapeutic approach to Huntington's disease (complex mechanism of neuronal death). *Curr Alzheimer Res 3*(4), 403–408.
10. Feany MB (2004). New genetic insights into Parkinson's disease. *New Engl J Med 351*(19), 1937–1940.
11. Lücking CB, et al. (2000). Association between early-onset Parkinson's disease and mutations in the Parkin gene. *New Engl J Med 342*(21), 1560–1567.
12. Nussbaum RL, Ellis EC (2003). Genomic medicine: Alzheimer's disease and Parkinson's disease. *New Engl J Med 348*(14), 1356–1364.

13. Brott T, Bogousslavsky J (2000). Treatment of acute ischemic stroke. *New Engl J Med 343*(10), 710–722.
14. Nadareishvili Z, Hallenbeck J (2003). Neuronal regeneration after stroke. *New Engl J Med 348*(23), 2355–2356.
15. O'Neill MJ, Astles PC, Allan SM, Anthony DC (2004). Anti-inflammatory modulators in stroke. *Drug Discov Today Ther Strategies 1*(1), 59–67.
16. Schulz JB, Matthews RT, Muqit MM, Browne SE, Beal MF (1995). Inhibition of neuronal nitric oxide synthase by 7-nitroindazole protects against MPTP-induced neurotoxicity in mice. *J Neurochem 64*(2), 936–939.
17. Foster GA, Stringer BMJ (2004). Cell-based therapies for Parkinson's disease. *Drug Discov Today Ther Strategies 1*(1), 43–49.
18. Licinio J, Ma-Li Wong (2002). *Pharmacogenomics—The Search for Individualized Therapies*. Weinheim, Germany: Wiley-VCH Verlag, pp. 311–332, 347–363.
19. Zoppo G (2006). Stroke and neurovascular protection. *New Engl J Med 354*(6), 553–555.
20. Mertsch K, Maas J (2002). Blood-brain barrier penetration and drug development from an industrial point of view. *Curr Med Chem Central Nervous Syst Agents 2*(3), 187–201.
21. Boer A, Gaillard PJ (2002). In vitro models of the blood-brain barrier: When to use which? *Curr Med Chem Central Nervous Syst Agents 2*(3), 203–209.
22. Atkinson F, Cole S, Green C, van de Waterbeemd H (2002). Lipophilicity and other parameters affecting brain penetration. *Curr Med Chem Central Nervous Syst Agents 2*(3), 229–240.
23. Kaznessis YN (2005). A review of methods for computational prediction of blood-brain partitioning. *Central Nervous Syst Agents Med Chem 5*(3), 185–191.
24. Sippl W (2002). Computational approaches for the prediction of blood-brain barrier permeation. *Curr Med Chem Central Nervous Syst Agents 2*(3), 211–227.
25. Winkler DA, Burden RF (2004). Modeling blood-brain barrier partitioning using Bayesian neural nets. *J Mol Graphics Model 22*(6), 499–505.
26. Calabria AR, Shusta EV (2006). Blood-brain barrier genomics and proteomics: Elucidating phenotype, identifying disease targets and enabling brain drug delivery. *Drug Discov Today 11*(17/18), 792–799.
27. Ng PP, Morrison SL, Penichet ML (2002). Antibody fusion proteins: Applications in brain targeting. *Curr Med Chem Central Nervous Syst Agents 2*(4), 275–285.
28. Crooke TS (2001). *Antisense Drug Technology. Principles, Strategies and Applications*. New York: Marcel Dekker.
29. Crooke TS (1998). *Antisense Research and Application*. Berlin: Springer-Verlag, pp. 545–557.
30. Gualtieri F (2000). *New Trends in Synthetic Medicinal Chemistry*. Weinheim: Wiley-VCH Verlag GmbH, pp. 261–335.
31. Grisaru D, Sternfeld M, Eldor A, Glick D, Soreq H (1999). Structural roles of acetylcholinesterase variants in biology and pathology. *Eur J Biochem 264*, 672–686.
32. Soreq H, Glick D (2000). Novel roles for cholinesterases in stress and inhibitor responses. In E. Giacobini, Ed. *Cholinesterases and Cholinesterases Inhibitors*. London: Martin Dunitz, pp. 47–61.
33. Sternfeld M, Shoham S, et al. (2000). Excess "readthrough" acetylcholinesterase attenuates but the "synaptic" variants intensifies neurodeterioration correlates. *Proc Natl Acad Sci USA 97*, 8647–8652.
34. Troy CM, et al. (2000). Caspase-2 mediates neuronal cell death induced by β-amyloid. *J Neurosci 20*, 1386–1392.
35. Hempstead BL (2006). Dissecting the diverse actions of pro- and mature neurotrophins. *Curr Alzheimer Res 3*(1), 19–24.
36. Longo FM, Massa SM (2005). Neurotrophin receptor-based strategies for Alzheimer's disease. *Curr Alzheimer Res 2*(2), 167–169.
37. Yébenes JG, Sánchez MP (2001). Neurotrophic factors for the investigation and treatment of movement disorders. *Curr Med Chem Central Nervous Syst Agents 1*(3), 225–238.
38. Longo FM, Xie Y, Massa SM (2005). Neurotrophin small molecule mimetics: Candidate therapeutic agents for neurological disorders. *Curr Med Chem Central Nervous Syst Agents 5*(1), 29–41.
39. Wyss-Coray T (2006). TGF-β pathway as a potential target in neurodegeneration and Alzheimer's. *Curr Alzheimer Res 3*(3), 191–195.
40. Zhang H, Zou K, Tesseur I, Wyss-Coray T (2005). Small molecule TGF-beta mimetics as potential neuroprotective factors. *Curr Alzheimer Res 2*(2), 183–186.
41. Gozes I, Spivak-Pohis I (2006). Neurotrophic effects of the peptide NAP: A novel neuroprotective drug candidate. *Curr Alzheimer Res 3*(3), 197–199.
42. Gozes I, Zaltzman R, Hauser J, Brenneman DE, Shohami E, Hill JM (2005). The expression of activity-dependent neuroprotective protein (ADNP) is regulated by brain damage and treatment of mice with the ADNP derived peptide, NAP, reduces the severity of traumatic head injury. *Curr Alzheimer Res 2*(2), 149–153.
43. Blackburn TP (2004). Serotonergic agents and Parkinson's disease. *Drug Discov Today Ther Strategies 1*(1), 35–41.
44. Cooper RJ, Bloom EF, Roth HR (1996). *The Biochemical Basis of Neuropharmacology*. New York: Oxford University Press.
45. Dukat M (2004). 5-$HT_3$ serotonin receptor agonists: A pharmacophoric journey. *Curr Med Chem Central Nervous Syst Agents 4*(2), 77–94.
46. Kamei K, et al. (2006). Synthesis, SAR studies, and evaluation of 1,4-benzoxazepine derivatives as selective 5-$HT_{1A}$ receptor agonists with neuroprotective effects: Discovery of Piclozotan. *Bioorg Med Chem 14*(6), 1978–1992.

47. Krosgaard-Larsen P, Liljefors T, Madsen U (2002). *Textbook of Drug Design and Discovery*. New York: Taylor & Francis.
48. López-Rodriguez ML, Benhamu B, Morcillo MJ, Porras E, Lavandera JL, Pardo L (2004). Serotonin 5-$HT_7$ receptor antagonists. *Curr Med Chem Central Nervous Syst Agents 4*(3), 203–214.
49. Mattes H, Pfannkuche H-J (2003). Serotonin type-4 (5-$HT_4$) receptors as therapeutic targets: Past and future roles of 5-$HT_4$ receptor agonists. *Curr Med Chem Central Nervous Syst Agents 3*(1), 27–36.
50. Mattes H, Pfannkuche H-J (2003). Serotonin type-4 (5-$HT_4$) receptors as therapeutic targets: 5-$HT_4$ receptor antagonists. *Curr Med Chem Central Nervous Syst Agents 3*(1), 37–42.
51. Rhoden BJ, Bouvet M, Izenwasser S, Wade D, Lomenzo AS, Trudell L (2005). Structure-activity studies of 3′,4′-dichloro-meperidine analogues at dopamine and serotonin transporters. *Bioorg Med Chem 13*(19), 5623–5634.
52. Smith HJ, Williams H (1998). *Introduction to the Principles of Drug Design and Action*. Amsterdam: Harwood Academic.
53. Wood M (2005). Role of the 5-$HT_{2C}$ receptor in atypical antipsychotics: Hero or villain? *Curr Med Chem Central Nervous Syst Agents 5*(1), 63–66.
54. Siegel GJ, Agranof RW, Albers RW, Fisher SK, Ulher MD (1998). *Basic Neurochemistry—Molecular, Cellular and Medical Aspects*, 6th ed. Philadelphia and New York: Lippincott-Raven.
55. Alonso D, et al. (2005). Donepezil-tacrine hybrid related derivatives as new dual binding site inhibitors of AChE. *Bioorg Med Chem 13*(24), 6588–6597.
56. Schott Y, Decker M, Rommelspacher H, Lehmann J (2006). 6-Hydroxy- and 6-methoxy-β-carbolines as acetyl- and butyrylcholinesterase inhibitors. *Bioorg Med Chem Lett 16*(22), 5840–5843.
57. Yuana H, Petukhov AP (2006). Computational evidence for the ligand selectivity to the α4β2 and α3β4 nicotinic acetylcholine receptors. *Bioorg Med Chem 14*(23), 7936–7942.
58. Crittenden LD, Park A, Qiu J, Silverman BR, Duke KR, Johnston GAR, Jordan M, Chebib M (2006). Enantiomers of cis-constrained and flexible 2-substituted GABA analogues exert opposite effects at recombinant $GABA_C$ receptors. *Bioorg Med Chem 14*(2), 447–455.
59. Guerrini G, et al. (2006). Benzodiazepine receptor ligands. 8: Synthesis and pharmacological evaluation of new pyrazolo[5,1-*c*] [1,2,4]benzotriazine 5-oxide 3- and 8-disubsituted: High affinity ligands endowed with inverse agonist pharmacological efficacy. *Bioorg Med Chem 14*(3), 758–775.
60. Sieghart W, Ernst M (2005). Heterogeneity of $GABA_A$ receptors: Revived interest in the development of subtype-selective drugs. *Central Nervous Syst Agents Med Chem 5*(3), 217–242.
61. Davies PA, Hanna MC, Hales TG, Kirkness EF (1997). Intensitivity to anaesthetic agents conferred by a class of $GABA_A$ receptor subunit. *Nature 385*, 820–823.
62. Luddens H, Korpi E, Seeburg PH (1995). $GABA_A$benzodiazepine receptor heterogeneity: Neurophysiological implications. *Neuropharmacology 34*, 245–254.
63. Macdonald RL, Olsen RW (1994). $GABA_A$ receptor channels. *Annu Rev Neurosci 17*, 569–602.
64. McKernan RM, Whiting PJ (1996). Which $GABA_A$-receptor subtypes really occur in the brain? *Trends Neurosci 19*, 139–143.
65. Varano F, et al. (2005). 1-Substituted pyrazolo[1,5-c] quinazolines as novel Gly/NMDA receptor antagonists: Synthesis, biological evaluation and molecular modeling study. *Bioorg Med Chem 13*(19), 5536–5549.
66. Bachurin SO, Grigoriev VV, Drany OA, Proshin AN, Tkachenko SE (1999). Isothiouronium derivative potentiates Ampa/kainate—Induced currents in rat cerebellar Purkinje neurons. *J Neurochem 73*(Suppl), S143D.
67. Frølund B, et al. (2005). Tetrazolyl isoxazole amino acids as ionotropic glutamate receptor antagonists: Synthesis, modeling and molecular pharmacology. *Bioorg Med Chem 13*(18), 5391–5398.
68. Conti P, et al. (2005). Synthesis, binding affinity at glutamic acid receptors, neuroprotective effects, and molecular modeling investigation of novel dihydroisoxazole amino acids. *J Med Chem 48*(20), 6315–6325.
69. Graindorge E, Francotte P, Boverie S, de Tullio P, Pirotte B (2004). New trends in the development of positive allosteric modulators of AMPA receptors. *Curr Med Chem Central Nervous Syst Agents 4*(2), 95–103.
70. Szénási G, Hársing LG Jr (2004). Pharmacology and prospective therapeutic usefulness of negative allosteric modulators of AMPA receptors. *Drug Discov Today Ther Strategies 1*(1), 69–76.
71. Takano Y, Shiga F, Asano J, Ando N, Uchiki H, Fukuchi K, Anraku T (2005). Design, synthesis and AMPA receptor antagonistic activity of novel 6-nitro-3-oxoquinoxaline-2-carboxylic acid with a substituted phenyl group at the 7 position. *Bioorg Med Chem 13*(20), 5841–5863.
72. Takano Y, Shiga F, Asano J, Hori W, Fukuchi K, Anraku T, Uno T (2006). Design and synthesis of novel 7-heterocycle-6-trifluoromethyl-3-oxoquinoxaline-2-carboxylic acids bearing a substituted phenyl group as superior AMPA receptor antagonists with good physicochemical properties. *Bioorg Med Chem 14*(3), 776–792.
73. Campiani G, et al. (2001). A rational approach to the design of selective substrates and potent nontransportable inhibitors of the excitatory amino acid transporter EAAC1 (EAAT3). New glutamate and aspartate analogues as potential neuroprotective agents. *J Med Chem 44*(16), 2507–2510.

74. Hashimoto K, Okamura N, Shimizu E, Iyo M (2004). Glutamate hypothesis of schizophrenia and approach for possible therapeutic drugs. *Curr Med Chem Central Nervous Syst Agents 4*(2), 147–154.
75. Sun MK (2003). NMDA receptor antagonists as antidepressant and antidementia drugs: Recent developments and future prospects. *Curr Med Chem Central Nervous Syst Agents 3*(4), 323–334.
76. Thorsten K, Weikop P (2004). Monoamine transporters as continuing targets for drug discovery in depression. *Drug Discov Today Ther Strategies 1*(1), 111–116.
77. Kim H-J, Cho YS, Koh HY, Kong JY, No KT, Pae AN (2006). Classification of dopamine antagonists using functional feature hypothesis and topological descriptors. *Bioorg Med Chem 14*(5), 1454–1461.
78. Vangveravong S, McElveen E, Taylor M, Xu J, Tu Z, Luedtke RR, Mach HR (2006). Synthesis and characterization of selective dopamine D2 receptor antagonists. *Bioorg Med Che 14*(3), 815–825.
79. Zhang K, Baldessarini RJ, Tarazi IF, Neumeyer LN (2002). Selective dopamine $D_4$ receptor antagonists: Review of structure-activity relationships. *Curr Med Chem Central Nervous Syst Agents 2*(4), 259–274.
80. Michaelis ML, Seyb KI, Ansar S (2005). Cytoskeletal integrity as a drug target. *Curr Alzheimer Res 2*(2), 227–229.
81. Bachurin S, et al. (2001). Antihistamine agent Dimebon as novel neuroprotector and cognition enhancer. *Ann NY Acad Sci 939*, 425–435.
82. Murakami K, Yokoyama H, Onodera K, Iinuma K, Watanabe T (1995). AQ-0145, a newly developed histamine H3 antagonist, decreased seizure susceptibility of electrically induced convulsions in mice. *Methods Find Exp Clin Pharmacol 17*(Suppl C), 70–73.
83. Wu RM, Murphy DL, Chiueh CC (2004). Selegiline (*l*-deprenyl) as a unique neuroprotective agent for chronic neurodegenerative disorders—A lesson from MAO inhibition. *Curr Med Chem Central Nervous Syst Agents 4*(4), 255–267.
84. Machado LS, Kozak A, Ergul A, Hess DC, Borlongan CV, Fagan SC (2006). Delayed minocycline inhibits ischemia-activated matrix metalloproteinases 2 and 9 after experimental stroke. *Neuroscience 7*, 56–65.
85. Lertvorachona J, et al. (2005). 1,12-Substituted tetracyclines as antioxidant agents. *Bioorg Med Chem 13*(15), 4627–4637.
86. Vaughan CJ, Delanty N, O'Mahony D (2000). Nitric oxide in acute ischaemic stroke. *J Neurol Neurosurg Psychiatry 68*(1), 123–129.
87. Contestabile A, Monti B, Contestabile A, Ciani E (2005). Updating the dual role of brain nitric oxide in neurodegeneration/neuroprotection: Understanding molecular mechanisms to devise drug approaches. *Front Med Chem 2*(1), 63–109.
88. Hobbs A, Higgs A, Moncada S (1999). Inhibition of nitric oxide synthase as a potential therapeutic target. *Annu Rev Pharmacol Toxicol 39*, 191–220.
89. Mena MA, Casarejos MJ, Canals S (2005). Nitric oxide and dopamine neurons. Implications for Parkinson's disease. *Central Nervous Syst Agents Med Chem 5*(3), 193–205.
90. Litzinger AE, Martasek P, Roman JL, Silverman BR (2006). Design, synthesis and biological testing of potential heme-coordinating nitric oxide synthase inhibitors. *Bioorg Med Chem 14*(9), 3185–3198.
91. Raman CS, Li H, Martasek P, Southan G, Masters BS, Poulos TL (2001). Crystal structure of nitric oxide synthase bound to nitro indazole reveals a novel inactivation mechanism. *Biochemistry 40*(45), 13448–13455.
92. Rosenfeld RJ, et al. (1997). Conformational changes in nitric oxide synthases induced by chlorzoxazone and nitroindazoles: Crystallographic and computational analyses of inhibitor potency. *Chem Res Toxicol 10*(4), 364–368.
93. Castagnoli K, Palmer S, Anderson A, Bueters T, Castagnoli N (2002). The neuronal nitric oxide synthase inhibitor 7-nitroindazole also inhibits the monoamine oxidase-B-catalyzed oxidation of 1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine. *Biochemistry 41*(47), 13915–13925.
94. Muramatsu Y, et al. (2002). Therapeutic effect of neuronal nitric oxide synthase inhibitor (7-nitroindazole) against MPTP neurotoxicity in mice. *Metab Brain Dis 17*(3), 169–182.
95. Gorlach C, Hortobagyi T, Hortobagyi S, Benyo Z (2000). Neuronal nitric oxide synthase inhibitor has a neuroprotective effect in a rat model of brain injury. *Restor Neurol Neurosci 17*(2–3), 71–76.
96. Nanri K, Montecot C, Springhetti V (1998). The selective inhibitor of neuronal nitric oxide synthase, 7-nitroindazole, reduces the delayed neuronal damage due to forebrain ischemia in rats. *Stroke 29*, 1248–1254.
97. Chalimoniuk M, Strosznajder J (1998). NMDA receptor-dependent nitric oxide and cGMP synthesis in brain hemispheres and cerebellum during reperfusion after transient forebrain ischemia in gerbils: Effect of 7-nitroindazole. *J Neurosci Res 54*(5), 681–690.
98. Coert BA, Anderson RE, Meyer FB (2003). Is neuroprotective efficacy of nNOS inhibitor 7-NI dependent on ischemic intracellular pH? *Am J Physiol Heart Circ Physiol 284*(1), 151–159.
99. Hirabayashi H, Takizawa S, Fukuyama N, Nakazawa H, Shinohara Y (1999). 7-Nitroindazole attenuates nitrotyrosine formation in the early phase of cerebral ischemia-reperfusion in mice. *Neurosci Lett 268*(2), 111–113.
100. Ishida A, Trescher WH, Lange MS, Johnston MV (2001). Prolonged suppression of brain nitric oxide synthase activity by 7-nitroindazole protects against cerebral hypoxic-ischemic injury in neonatal rat. *Brain Dev 23*(5), 349–354.
101. Lei B, Adachi N, Nagaro T, Arai T, Koehler RC (1999). Nitric oxide production in the CA1 field of the gerbil hippocampus after transient forebrain ischemia: Effects

of 7-nitroindazole and NG-nitro-L-arginine methyl ester. *Stroke 30*(3), 669–677.

102. Jiang HM, Kaku T, Hada J, Hayashi Y (1999). 7-Nitroindazole reduces nitric oxide concentration in rat hippocampus after transient forebrain ischemia. *Eur J Pharmacol 380*(2–3), 117–121.
103. De Sarro G, Gareri P, Falconi U, De Sarro A (2000). 7-Nitroindazole potentiates the antiseizure activity of some anticonvulsants in DBA/2 mice. *Eur J Pharmacol 394*(2–3), 275–288.
104. Nagafuji T, Sugiyama M, Muto A, Makino T, Miyauchi T, Na H (1995). The neuroprotective effect of a potent and selective inhibitor of type I NOS (L-MIN) in a rat model of focal cerebral ischaemia. *Neuroreport 6*(11), 1541–1545.
105. Zhang GZ, et al. (1996). ARL 17477, a potent and selective neuronal NOS inhibitor decreases infarct volume after transient middle cerebral artery occlusion in rats. *J Cerebral Blood Flow Metab 16*, 599–604.
106. Saez EM, Ramirez-Lorca R, Moron JF, Ruiz A (2006). The therapeutic potential of the calpain family: New aspects. *Drug Discov Today 11*(19/20), 917–923.
107. Wu HI, Matsui H, Tomizawa K (2005). Crosstalk between calpain and calcineurin in excitotoxic neurodegeneration; therapeutic targets for the treatment of excitotoxic neurodegeneration. *Central Nervous Syst Agents Med Chem 5*(3), 207–216.
108. Zatz M, Starling A (2005). Calpains and disease. *New Engl J Med 352*(23), 2413–2423.
109. Carragher NO (2006). Calpain inhibition: A therapeutic strategy targeting multiple disease states. *Curr Pharm Des 12*, 615–638.
110. Gafni J, Ellerby LM (2002). Calpain activation in Huntington's disease. *J Neurosci 22*, 4842–4849.
111. Ray SK, et al. (2003). Calpain in the pathophysiology of spinal cord injury: Neuroprotection with calpain inhibitors. *Brain Res Rev 42*, 169–185.
112. Levin ED (2005). Extracellular superoxide dismutase (EC-SOD) quenches free radicals and attenuates age-related cognitive decline: Opportunities for novel drug development in aging. *Curr Alzheimer Res 2*(2), 191–196.
113. Peluffo H, Acarin L, Arís A, Pau G, Villaverde A, Castellano B, Berta G (2006). Neuroprotection from NMDA excitotoxic lesion by Cu/Zn superoxide dismutase gene delivery to the postnatal rat brain by a modular protein vector. *Neuroscience 7*, 35–51.
114. Bendotti C, Tortarolo M, Borsello T (2006). Targeting stress activated protein kinases, JNK and p38, as new therapeutic approach for neurodegenerative diseases. *Central Nervous Syst Agents Med Chem 6*(2), 109–117.
115. Maroney AC, Saporito MS, Hudkins RL (2002). Mixed lineage kinase family, potential targets for preventing neurodegeneration. *Curr Med Chem Central Nervous Syst Agents 2*(2), 143–155.
116. Pallas M, Canudas AM, Verdaguer E, Allgaier C, de Arriba SG, Alvira D, Sureda FX, Camins A (2005). Inhibitors of cyclin-dependent kinases: Potential drugs for the treatment of neurodegenerative disorders? *Central Nervous Syst Agents Med Chem 5*(2), 101–109.
117. Mora A, Sabio G, Centeno F (2003). Neuroprotective effects of lithium—Pointing out protein phosphatases as drug targets? *Curr Med Chem Central Nervous Syst Agents 3*(4), 335–339.
118. Strunecká A, Grof P (2006). The role of glucose in the pathogenesis of Alzheimer's disease revisited: What does it tell us about the therapeutic use of lithium? *Central Nervous Syst Agents Med Chem 6*(3), 175–192.
119. Karagiannis A, Tziomalos K (2006). Neuroprotective properties of erythropoietin in cerebral ischemia. *Central Nervous Syst Agents Med Chem 6*(3), 153–161.
120. Baker D, Pryce G (2004). The potential role of the endocannabinoid system in the control of multiple sclerosis. *Curr Med Chem Central Nervous Syst Agents 4*(3), 195–202.
121. Fernandez-Ruiz JJ, Lastres-Becker I, Cabranes A, de Lago E, Ramos JA (2004). Involvement of the endocannabinoid system in motor disorders. *Curr Med Chem Central Nervous Syst Agents 4*(3), 175–182.
122. Fowler JC (2004). Metabolism of the endocannabinoids anandamide and 2-arachidonoyl glycerol, a review, with emphasis on the pharmacology of fatty acid amide hydrolase, a possible target for the treatment of neurodegenerative diseases and pain. *Curr Med Chem Central Nervous Syst Agents 4*(3), 161–174.
123. López-Rodríguez ML, Viso A, Ortega-Gutiérrez S, Fernández-Ruiz J, Ramos AJ (2002). Endocannabinoid transporter inhibitors. *Curr Med Chem Central Nervous Syst Agents 2*(2), 129–141.
124. Diaz Brinton R, Wang JM (2006). Preclinical analyses of the therapeutic potential of allopregnanolone to promote neurogenesis *in vitro* and *in vivo* in transgenic mouse model of Alzheimer's disease. *Curr Alzheimer Res 3*(1), 11–17.
125. Diaz Brinton R, Wang JM (2006). Therapeutic potential of neurogenesis for prevention and recovery from Alzheimer's disease: Allopregnanolone as a proof of concept neurogenic agent. *Curr Alzheimer Res 3*(3), 185–190.
126. Tsutsui K, Mellon SH (2006). Neurosteroids in the brain neuron: Biosynthesis, action and medicinal impact on neurodegenerative disease. *Central Nervous Syst Agents Med Chem 6*(1), 73–82.
127. Popoli P, et al. (2004). The controversial role of adenosine $A_{2A}$ receptor antagonists as neuroprotective agents. *Curr Med Chem Central Nervous Syst Agents 4*(1), 35–45.
128. Yi JJ, Ehlers MD (2005). Targeting protein degradation in the nervous system. *Central Nervous Syst Agents Med Chem 5*(4), 285–306.
129. Perry TA, Nigel HG (2005). Enhancing central nervous system endogenous GLP-1 receptor pathways for inter-

vention in Alzheimer's disease. *Curr Alzheimer Res 2*(3), 377–385.

130. Ebly EM, Parhad IM, Hogan DB, Fung TS (1994). Prevalence and types of dementia in the very old: Results from the Canadian study of health and aging. *Neurology 44*, 1593–1600.
131. Sykes CM, Marks DF, McKinley JM (2001). Alzheimer disease and associated disorders: Project funding opportunities within the European community. *Alzheimer Dis Assoc Disord 15*, 102–105.
132. Ruther E, Ritter R, Apecechea M, Freytag S, Gmeinbauer R, Windisch M (2000). Sustained improvements in patients with dementia of Alzheimer's type (DAT) 6 months after termination of Cerebrolysin therapy. *J Neural Transm 107*, 815–829.
133. Goate, A, et al. (1991). Segregation of a missense mutation in the amyloid precursor protein gene with familial Alzheimer's disease. *Nature 349*, 704–706.
134. Borchelt DR, et al. (1996). Familial Alzheimer's disease-linked presenilin 1 variants elevate Aβ1-42/1-40 ratio *in vitro* and *in vivo*. *Neuron 17*, 1005–1013.
135. Sherrington R, et al. (1997). Cloning of a novel gene bearing missense mutations in early onset familial Alzheimer's disease. *Trends Neurosci 20*, 154–159.
136. Schmechel DE, et al. (1993). Increased amyloid β-peptide deposition in cerebral cortex as a consequence of apolipoprotein E genotype in late-onset Alzheimer disease. *Proc Natl Acad Sci USA 90*, 9649–9653.
137. Bird TD (2005). Genetic factors in Alzheimer's disease. *New Engl J Med 352*(9), 862–864.
138. Tanzi RE (2005). Tangles and neurodegenerative disease—A surprising twist. *New Engl J Med 353*(17), 1853–1855.
139. Bertram L (2005). The ubiquilin 1 gene and Alzheimer's disease. *New Engl J Med 352*(26), 2752–2753.
140. Bertram L, et al. (2005). Family-based association between Alzheimer's disease and variants in UBQLN1. *New Engl J Med 352*(9), 884–894.

140a. Bartus RT, Dean RL, Beer B, Lippa AS (1982). The cholinergic hypothesis of geriatric memory dysfunction. *Science 217*, 408–417.

141. Davies P, Maloney AJF (1976). Selective loss of cholinergic neurons in Alzheimer's disease. *Lancet 2*, 1403.
142. Campos EO, Alvarez A, Nibaldo CI (1998). Brain acetylcholinesterase promotes amyloid-beta-peptide aggregation but does not hydrolyze amyloid precursor protein peptides. *Neurochem Res 23*, 135–140.
143. Del Toro D, Coma M, Uribesalgo I, Guix FX, Muñoz FJ (2005). The amyloid β-protein precursor and Alzheimer's disease. Therapeutic approaches. *Central Nervous Syst Agents Med Chem 5*(4), 271–283.
144. Vogelgesang S, et al. (2004). The role of P-glycoprotein in cerebral amyloid angiopathy; Implications for the early pathogenesis of Alzheimer's disease. *Curr Alzheimer Res 1*(2), 121–125.
145. Wang YJ, Zhou HD, Zhou XF (2006). Clearance of amyloid-beta in Alzheimer's disease: Progress, problems and perspectives. *Drug Discov Today 11*(19/20), 931–938.
146. Behl C, Davis J, Cole GM, Schubert D (1992). Vitamin E protects nerve cells from amyloid-β protein toxicity. *Biochem Biophys Res Commun 186*, 944–950.
147. Soto C, Sigurdsson EM, Morelli L, Kumar RA, Castano EM, Frangione B (1998). β-Sheet breaker peptides inhibit fibrilogenesis in a rat brain model of amyloidosis: Implication to Alzheimer's therapy. *Nature Med 4*, 822–826.
148. Chad AD, Eriksen J, Kamal A, Burrows F, Kasibhatla S, Eckman CB, Hutton M, Petrucelli L (2005). Development of a high throughput drug screening assay for the detection of changes in Tau levels—Proof of concept with HSP90 inhibitors. *Curr Alzheimer Res 2*(2), 231–238.
149. Iqbal K, Grundke-Iqbal I (2005). Pharmacological approaches of neurofibrillary degeneration. *Curr Alzheimer Res 2*(3), 335–341.
150. Pickhardt M, von Bergen M, Gazova Y, Hascher A, Bierna J, Mandelkow EM, Mandelkow E (2005). Screening for inhibitors of tau polymerization. *Curr Alzheimer Res 2*(2), 219–226.
151. Doble A (1999). The role of excitotoxicity in neurodegenerative disease: Implication for therapy. *Pharmacol Ther 81*, 163–221.
152. Zeevalk GD, Nicklas WJ (1991). Mechanisms underlying initiation of excitotoxicity associated with metabolic inhibition. *J Pharmacol Exp Ther 253*, 1285–1292.
153. Bell C (1999). Alzheimer's disease and oxidative stress: Implication for novel therapeutic approaches. *Progr Neurol 57*, 301–323.
154. Madrigal LMJ, Garcia-Bueno B, Cardenas A, Lizasoain I, Moro AM, Lorenzo P, Leza JC (2004). Oxidative/nitrosative brain damage in stress: Possible target for neuropsychopharmacological drugs. *Curr Med Chem Central Nervous Syst Agents 4*(4), 235–242.
155. Tabner BJ, Turnbull S, El-Agnaf OMA, Allsop D (2001). Production of reactive oxygen species from aggregating proteins implicated in Alzheimer's disease, Parkinson's disease and other neurodegenerative diseases. *Curr Topics Med Chem 1*, 507–517.
156. Wink DA, Hanbauer I, Krishna MC, De Graff W, Gamson J, Mitchell JB (1993). *Nitric oxide* protects against cellular damage and cytotoxicity from reactive oxygen species. *Prod Natl Acad Sci USA 90*, 9813–9817.
157. Moreira PI, Siedlak SL, Aliev G, Zhu X, Cash AD, Smith MA, Perry G (2005). Oxidative stress mechanisms and potential therapeutics in Alzheimer disease. *J Neural Transm 112*, 921–932.
158. Behl C (2000). Apoptosis and Alzheimer's disease. *J Neural Transm 107*, 1325–1344.

159. Higuchi M, et al. (2005). Understanding molecular mechanisms of proteolysis in Alzheimer's disease: Progress toward therapeutic interventions. *Biochim Biophys Acta 1751*, 60–67.

160. Schorderet M (1995). Alzheimer's disease: Fundamental and therapeutic aspects. *Experientia 51*, 99–105.

161. Braginskaya FI, et al. (2001). Some blood biochemistry parameters during the cholinergic treatment of Alzheimer's disease. *Neurosci Behav Physiol 31*, 457–461.

162. Virmani MA, et al. (1995). Protective actions of L-carnitine and acetyl-L-carnitine on the neurotoxicity evoked by mitochondrial uncoupling or inhibitors. *Pharmacol Res 32*, 383–389.

163. Schnee ME, Brown BS (1998). Selectivity of linopirdine (DuP 996), a neurotransmitter release enhancer, in blocking voltage-dependent and calcium-activated potassium currents in hippocampal neurons. *J Pharmacol Exp Ther 286*, 709–717.

164. Miyazaki S, Onodera K, Imaizumi M, Timmerman H (1997). Effects of clobenpropit (VUF-9153), a histamine H3-receptor antagonist, on learning and memory, and on cholinergic and monoaminergic systems in mice. *Life Sci 61*, 355–361.

165. Stark H, et al. (1996). Novel carbamates as potent histamine H3 receptor antagonists with high *in vitro* and oral *in vivo* activity. *J Med Chem 39*, 1157–1163.

166. Ganellin CR, et al. (1996). A novel series of (phenoxyalkyl)imidazoles as potent H3-receptor histamine antagonists. *J Med Chem 39*, 3806–3813.

167. Behl P, Lanctôt KL, Streiner DL, Guimont I, Black SE (2006). Cholinesterase inhibitors slow decline in executive functions, rather than memory, in Alzheimer's disease: A 1-year observational study in the Sunnybrook dementia cohort. *Curr Alzheimer Res 3*(2), 147–156.

168. Nigel GH, Sambamurti K, Qian-Sheng Yu, Brossi A, Bruinsma GB, Debomoy KL (2005). An overview of phenserine tartrate, a novel acetylcholinesterase inhibitor for the treatment of Alzheimer's disease. *Curr Alzheimer Res 2*(3), 281–290.

169. Snape MF, Misra A, Murray TK, De Souza RJ, Williams JL, Cross AJ, Green AR (1999). A comparative study in rats of the *in vitro* and *in vivo* pharmacology of the acetylcholinesterase inhibitors tacrine, donepezil and NXX-066. *Neuropharmacology 38*, 181–193.

170. Enz A, Amstutz R, Boddeke H, Gmelin G, Malanowski J (1993). Brain selective inhibition of acetylcholinesterase: A novel approach to therapy for Alzheimer's disease. *Prog Brain Res 98*, 431–438.

171. Bryson HM, Benfield P (1997). Donepezil. *Drugs Aging 10*, 234–239.

172. Scott LJ, Goa KL (2000). Galantamine: A review of its use in Alzheimer's disease. *Drugs 60*, 1095–1122.

173. Imbimbo BP, Martelli P, Troetel WM, Lucchelli F, Lucca U, Thal LJ (1999). Efficacy and safety of eptastigmine for the treatment of patients with Alzheimer's disease. *Neurology 52*, 700–708.

174. Suh WH, Suslick KS, Suh YH (2005). Therapeutic agents for Alzheimer's disease. *Central Nervous Syst Agents Med Chem 5*(4), 259–269.

175. Silvaa CH, Camposa VL, Carvalhoa I, Taftb CA (2006). Molecular modeling, docking and ADMET studies applied to the design of a novel hybrid for treatment of Alzheimer's disease. *J Mol Graphics Model 25*(2), 169–175.

176. Akula N, Lecanu L, Greeson J, Papadopoulos V (2006). 3D QSAR studies of AChE inhibitors based on molecular docking scores and CoMFA. *Bioorg Med Chem Lett 16*, 6277–6280.

177. Chandrasekarana V, McGaugheyb BG, Cavallitoc JC, Bowena PJ (2004). Three-dimensional quantitative structure-activity relationship (3D-QSAR) analyses of choline acetyltransferase inhibitors. *J Mol Graphics Model 23*(1), 69–76.

178. Nibaldo CI, Soledad U, Colombres M (2004). Acetylcholinesterase (AChE)–amyloid-β-peptide complexes in Alzheimer's disease. The Wnt signaling pathway. *Curr Alzheimer Res 1*(4), 249–254.

179. Zhen Y, Feng J (2004). Alzheimer's disease: Interactions between cholinergic functions and β-amyloid. *Curr Alzheimer Res 1*(4), 241–248.

180. Munoz FJ, Inestrosa NC (1999). Neurotoxicity of acetylcholinesterase amyloid beta-peptide aggregates is dependent on the type of A-beta peptide and the AChE concentration present in the complexes. *FEBS Lett 450*, 205–209.

181. Fisher A, Michaelson DM, Brandeis R, Haring R, Chapman S, Pittel Z (2000). M1 muscarinic agonists as potential disease-modifying agents in Alzheimer's disease. Rationale and perspectives. *Ann NY Acad Sci 920*, 315–320.

182. Nordberg A (2001). Nicotinic receptor abnormalities of Alzheimer's disease: Therapeutic implications. *Biol Psychiatry 49*(3), 200–210.

183. Potter A, Corwin J, Lang J, Piasecki M, Lenox R, Newhouse PA (1999). Acute effects of the selective cholinergic channel activator (nicotinic agonist) ABT-418 in Alzheimer's disease. *Psychopharmacol (Berl) 142*, 334–442.

184. Nishizaki T, Matsuoka T, Nomura T, Sumikawa K, Shiotani T, Watabe S, Yoshii M (1998). Nefiracetam modulates acetylcholine receptor currents via two different signal transduction pathways. *Mol Pharmacol 53*, 1–5.

185. Vernier JM, et al. (1999). 4-[[2-(1-Methyl-2-pyrrolidinyl) ethyl]thio]phenol hydrochloride (SIB-1553A): A novel cognitive enhancer with selectivity for neuronal nicotinic acetylcholine receptors. *J Med Chem 42*, 1684–1686.

186. Mark RJ, Ashford JW, Goodman Y, Mattson MP (1995). Anticonvulsants attenuate amyloid beta-peptide neuro-

toxicity, $Ca^{2+}$ deregulation, and cytoskeletal pathology. *Neurobiol Aging 16*, 187–198.

187. Pittaluga A, Feligioni M, Ghersi C, Gemignani A, Raiteri M (2001). Potentiation of NMDA receptor function through somatostatin release: A possible mechanism for the cognition-enhancing activity of GABA(B) receptor antagonists. *Neuropharmacology 41*, 301–310.
188. Porsteinsson AP, et al. (2001). Placebo-controlled study of divalproex sodium for agitation in dementia. *Am J Geriatr Psychiatry 9*, 58–66.
189. Robbins DK, Hutcheson SJ, Miller TD, Green VI, Bhargava VO, Weir SJ (1997). Pharmacokinetics of MDL 26479, a novel benzodiazepine inverse agonist, in normal volunteers. *Biopharm Drug Dispos 18*, 325–334.

189a. Thomas T (2000). Monoamine oxidase-B inhibitors in the treatment of Alzheimer's disease. *Neurobiol Aging 21*, 343–348.

190. Cesura AM, Borroni E, Gottowik J, Kuhn C, Malherbe P, Martin J, Richards JG (1999). Lazabemide for the treatment of Alzheimer's disease: Rationale and therapeutic perspectives. *Adv Neurol 80*, 521–528.
191. Riekkinen PJ (1998). Review on the long-term efficacy and safety of selegeline in the treatment of Alzheimer's disease. *Neurobiol Aging (Abstract) 19*(Suppl. 4S), 1263.
192. Maillet M, Robert SJ, Lezoualc'h F (2004). New insights into serotonin 5-$HT_4$ receptors: A novel therapeutic target for Alzheimer's disease? *Curr Alzheimer Res 1*(2), 79–85.
193. Land C, Riccio DC (1999). D-Cycloserine: Effects on long-term retention of a conditioned response and on memory for contextual attributes. *Neurobiol Learn Mem 72*, 158–168.
194. Winblad B, Poritis N (1999). Memantine in severe dementia: Results of the 9M-Best Study (Benefit and efficacy in severely demented patients during treatment with memantine). *Int J Geriatr Psychiatry 14*, 135–146.
195. Thatcher GRJ, Bennett BM, Reynolds JN (2005). Nitric oxide mimetic molecules as therapeutic agents in Alzheimer's disease. *Curr Alzheimer Res 2*(2), 171–182.
196. Thatcher GRJ, Bennett BM, Reynolds JN (2006). NO chimeras as therapeutic agents in Alzheimer's disease. *Curr Alzheimer Res 3*(3), 237–245.
197. Chakrabarti E, Smith JD (2005). Drug library screen to identify compounds that decrease secreted Aβ from a human cell line. *Curr Alzheimer Res 2*(2), 255–259.
198. Marjaux E, De Strooper B (2004). γ-Secretase inhibitors: Still in the running as Alzheimer's. *Drug Discov Today Ther Strategies 1*(1), 1–6.
199. Talaga P (2004). Inhibitors of beta-amyloid aggregation: Still an issue of structure and function? *Drug Discov Today Ther Strategies 1*(1), 7–12.
200. Nitsch RM, Deng A, Wurtman RJ, Growdon JH (1997). Metabotropic glutamate receptor subtype mGluR 1a stimulates the secretion of the amyloid β-protein precursor ectodomaim. *J Neurochem 69*, 704–712.
201. Checler F, et al. (2005). JLK inhibitors: Isocoumarin compounds as putative probes to selectively target the γ-secretase pathway. *Curr Alzheimer Res 2*(3), 327–334.
202. Dovey HF, et al. (2001). Functional gamma-secretase inhibitors reduce beta-amyloid peptide levels in brain. *J Neurochem 76*, 173–181.
203. Fassbender K, et al. (2001). Simvastatin strongly reduces levels of Alzheimer's disease beta-amyloid peptides Abeta 42 and Abeta 40 *in vitro* and *in vivo*. *Proc Natl Acad Sci 98*, 5856–5861.
204. Jick H, Zornberg GL, Jick SS, Seshadri S, Drachman DA (2000). Statins and the risk of dementia. *Lancet 356*, 1627–1631.
205. Kwak B, Mulhaupt F, Myit S, Mach F (2000). Statins as a newly recognized type of immunomodulator. *Nat Med 6*, 1399–1402.
206. Refolo LM, et al. (2000). Hypercholesterolemia accelerates the Alzheimer's amyloid pathology in a transgenic mouse model. *Neurobiol Dis 7*, 321–331.
207. Sparks DL, Martin TA, Gross DR, Hunsaker JC (2000). Link between heart disease, cholesterol, and Alzheimer's disease: A review. *Microsc Res Tech 50*, 287–290.
208. Tjernberg LO, et al. (1996). Arrest of beta-amyloid fibril formation by a pentapeptide ligand. *J Biol Chem 271*, 8545–8548.
209. Brakch N, El Abida B, Rholam M (2006). Functional role of β-turn in polypeptide structure and its use as template to design therapeutic agents. *Central Nervous Syst Agents Med Chem 6*(3), 163–173.
210. Drouet B, Pincon-Raymond M, Chambaz J, Pillot T (1999). Laminin 1 attenuates β-amyloid peptide Ab(1-40) neurotoxicity of cultured fetal rat cortical neurons. *J Neurochem 73*, 742–749.
211. Monji A, Tashiro K, Yoshida I, Hashioka S, Kato T, Kanba S (2005). Laminin and its related peptides for the treatment of Alzheimer's disease. *Central Nervous Syst Agents Med Chem 5*(4), 243–247.
212. Tucker S, et al. (2006). RNA therapeutics directed to the non coding regions of APP mRNA, *in vivo* anti-amyloid efficacy of paroxetine, erythromycin, and *N*-acetyl cysteine. *Curr Alzheimer Res 3*(3), 221–227.
213. Howlett DR, et al. (1999). Inhibition of fibril formation in beta-amyloid peptide by a novel series of benzofurans. *Biochem J 340*, 283–289.
214. Howlett DR, George AR, Owen DE, Ward RV, Markwell RE (1999). Common structural features determine the effectiveness of carvedilol, daunomycin and rolitetracycline as inhibitors of Alzheimer beta-amyloid fibril formation. *Biochem J 343*, 419–423.
215. Yuanbin L, Schubert D (2006). Treating Alzheimer's disease by inactivating bioactive amyloid β peptide. *Curr Alzheimer Res 3*(2), 129–135.
216. Matsunaga Y, Yamada T (2005). Hydrolytic activity of amyloid-beta and its inhibition with short peptides. *Central Nervous Syst Agents Med Chem 5*(3), 165–170.

217. Soto C, Golabek A, Wisniewski T, Castano EM (1996). Alzheimer's beta-amyloid peptide is conformationally modified by apolipoprotein E in vitro. *NeuroReport 7*, 721–725.
218. Yang DS, Small DH, Seydel U, Smith JD, Hallmayer J, Gandy SE, Martins RN (1999). Apolipoprotein E promotes the binding and uptake of beta-amyloid into Chinese hamster ovary cells in an isoform-specific manner. *Neuroscience 90*, 1217–1226.
219. Deane R, et al. (2004). RAGE (yin) versus LRP (yang) balance regulates Alzheimer amyloid beta-peptide clearance through transport across the blood-brain barrier. *Stroke 35*(11, Suppl. 1), 2628–2631.
220. Eckman EA, Eckman CB (2005). Abeta-degrading enzymes: Modulators of Alzheimer's disease pathogenesis and targets for therapeutic intervention. *Biochem Soc Trans 33*, 1101–1105.
221. Bergamaschini L, et al. (2004). Peripheral treatment with enoxaparin, a low molecular weight heparin, reduces plaques and beta-amyloid accumulation in a mouse model of Alzheimer's disease. *J Neurosci 24*, 4181–4186.
222. Beka S (2004). Alzheimer's disease and immunotherapy. *Curr Alzheimer Res 1*(3), 149–163.
223. Frenkel D, Solomon B, Benhar I (2000). Modulation of Alzheimer's beta-amyloid neurotoxicity by site-directed single chain antibody. *J Neuroimmunol 106*, 23–31.
224. Schenk D, et al. (1999). Immunization with amyloid-β attenuates Alzheimer-disease-like pathology in the PDAPP mouse. *Nature 400*, 173–177.
225. Agadjanyan MG, et al. (2005). Prototype Alzheimer's disease vaccine using the immunodominant B cell epitope from beta-amyloid and promiscuous T cell epitope pan HLA DR-binding peptide. *J Immunol 174*, 1580–1586.
226. Janus C, et al. (2000). A beta peptide immunization reduces behavioural impairment and plaques in a model of Alzheimer's disease. *Nature 408*, 979–982.
227. Morgan D, et al. (2000). A beta peptide vaccination prevents memory loss in an animal model of Alzheimer's disease. *Nature 408*, 982–985.
228. Dodel RC, et al. (2004). Intravenous immunoglobulins containing antibodies against beta-amyloid for the treatment of Alzheimer's disease. *J Neurol Neurosurg Psychiatry 75*, 1472–1474.
229. Fukuchi K, et al. (2006). Amelioration of amyloid load by anti-Abeta single-chain antibody in Alzheimer mouse model. *Biochem Biophys Res Commun 344*, 79–86.
230. Lovestone S, Davis DR, Webster MT, Kaech S, Brion JP, Matus A, Anderton BH (1999). Lithium reduces tau phosphorylation: Effects in living cells and in neurons at therapeutic concentrations. *Biol Psychiatry 45*, 995–1003.
231. Bachurin SO (2003). Medicinal chemistry approaches for the treatment and prevention of Alzheimer's disease. *Med Res Rev 23*(1), 48–88.
232. Aslan M, Ozben T (2004). Reactive oxygen and nitrogen species in Alzheimer's disease. *Curr Alzheimer Res 1*(2), 111–119.
232a. Bolkenius FN, Verne-Mismer J, Wagner J, Grisar JM (1996). Amphiphilic alpha-tocopherol analogues as inhibitors of brain lipid peroxidation. *Eur J Pharmacol 298*, 37–43.
233. Lipton SA (1997). Janus faces of NF-kappa B: Neurodestruction versus neuroprotection. *Nature Med 3*, 20–22.
234. Allison AC, Cacabelos R, Lombardi VR, Alvarez XA, Vigo C (2001). Celastrol, a potent antioxidant and anti-inflammatory drug, as a possible treatment for Alzheimer's disease. *Prog Neuropsychopharmacol Biol Psychiatry 25*, 1341–1357.
235. Dubal D, Wilson ME, Wise PM (1999). Estradiol: A protective and trophic factor in the brain. *Alzheimer Dis Rev 4*, 1–9.
236. Xu H, et al. (1998). Estrogen reduces neuronal generation of Alzheimer beta-amyloid peptides. *NatMed 4*, 447–451.
237. Behl C, Lezoualc'h F (1998). Estrogens with intact phenolic group prevent death of neuronal cells following glutathione depletion. *Restorative Neurol Neurosci 12*, 127–134.
238. Blum-Degen D, Gotz ME (1999). Novel 17α-Estradiol analogues as potent radical scavengers. In Iqbal K, Swaab DF, Winblad B, Wisniweski HM, Eds. *Alzheimer's Disease and Related Disorders*. Wiley: Chichester, pp. 671–677.
239. Lermontova NN, et al. (2000). Effects of 17beta-estradiol and its isomer 17alpha-estradiol on learning in rats with chronic cholinergic deficiency in the brain. *Bull Exp Biol Med 129*, 525–527.
240. Bachurin S, et al. (1999). *N*-Acetylserotonin, melatonin and their derivatives improve cognition and protect against b-amyloid neurotoxicity. *Ann NY Acad Sci 890*, 155–166.
241. Longoni B, Pryor WA, Marchiafava P (1997). Inhibition of lipid peroxidation by *N*-acetylserotonin and its role in retinal physiology. *Biochem Biophys Res Commun 233*, 778–780.
242. Wolfler A, Abuja PM, Schauenstein K, Liebmann PM (1999). *N*-Acetylserotonin is a better extra- and intracellular antioxidant than melatonin. *FEBS Lett 449*, 206–210.
243. Stout AK, Raphael HM, Kanterewicz BI, Klann E, Reynolds IJ (1998). Glutamate-induced neuron death requires mitochondrial calcium uptake. *Nature Neurosci 1*, 366–373.
244. Wei H, Leeds P, Chen RW, Wei W, Leng Y, Bredesen DE, Chuang DM (2000). Neuronal apoptosis induced by pharmacological concentrations of 3-hydroxykynurenine: Characterization and protection by dantrolene and Bcl-2 overexpression. *J Neurochem 75*, 81–90.

245. Mather M, Rottenberg H (2000). Aging enhances the activation of the permeability transition pore in mitochondria. *Biochem Biophys Acta 273*, 603–608.

246. Lim GP, et al. (2000). Ibuprofen suppresses plaque pathology and inflammation in a mouse model for Alzheimer's disease. *J Neurosci 20*, 5709–5714.

247. Pratico D (2005). Alzheimer's disease and non-steroidal anti-inflammatory drugs: Old therapeutic tools with novel mechanisms of action? *Central Nervous Syst Agents Med Chem 5*(2), 111–117.

248. Wenhui H, Hantamalala RR, Craft JM, Van Eldik LJ, Watterson DM (2005). Validation of the neuroinflammation cycle as a drug discovery target using integrative chemical biology and lead compound development with an Alzheimer's disease-related mouse model. *Curr Alzheimer Res 2*(2), 197–205.

249. Aisen PS, et al. (2000). A randomized controlled trial of prednisone in Alzheimer's disease. Alzheimer's Disease Cooperative Study. *Neurology 54*, 588–593.

250. Aisen PS, Davis KL (1997). Anti-inflammatory therapy for Alzheimer's disease: A status report. *Int J Geriatric Psychopharm 1*, 2–5.

251. Aisen PS, Marin DB, Brickman AM, Santoro J, Fusco M (2001). Pilot tolerability studies of hydroxychloroquine and colchicine in Alzheimer disease. *Alzheimer Dis Assoc Disord 15*, 96–101.

# THINKING OUTSIDE THE BOX IN ALZHEIMER DISEASE TREATMENT

Rudy J. Castellani,[1] Michael W. Marlatt,[2] Akihiko Nunomura,[3] Paula I. Moreira,[4] Hyoung-Gon Lee,[5] Gemma Casadesus,[6] Xiongwei Zhu,[5] George Perry,[5,7] and Mark A. Smith[5]

*[1]Department of Pathology, University of Maryland, Baltimore, Maryland, [2]Merck Research Laboratories, Merck & Company, Inc., West Point, Pennsylvania, [3]Department of Psychiatry and Neurology, Asahikawa Medical College, Asahikawa, Japan, [4]Center for Neuroscience and Cell Biology of Coimbra, University of Coimbra, Coimbra, Portugal, [5]Department of Pathology, Case Western Reserve University, Cleveland, Ohio, [6]Department of Neuroscience, Case Western Reserve University, Cleveland, Ohio, [7]College of Sciences, University of Texas at San Antonio, San Antonio, Texas*

## 1 INTRODUCTION

Alzheimer disease (AD) is the most prevalent neurodegenerative disease worldwide and affects over 4 million Americans [1]. As the average life expectancy increases in the United States and throughout the world, so does the number of elderly men and women who become increasingly susceptible to AD as they age. In fact, it is estimated that by 2050, 14 million Americans will have AD if preventative treatments do not become available [2]. The cost of treatment is considerable: National direct and indirect cost of caring for AD patients have been estimated to be upward of $100 billion per year, a figure that will rise as the prevalence increases [3].

AD is characterized by two basic neuropathological lesions, the senile plaque and the neurofibrillary tangle. The major protein component of the senile plaque is amyloid-β (Aβ), while the major protein component of the neurofibrillary tangle is hyperphosphorylated tau [1]. It is important to point out, however, that our understanding of the neuropathology of AD is based upon autopsy examination of end-stage disease. Thus, plaques and tangles are not "diagnostic" of AD; indeed, they may be seen in variable, and often large, numbers in cognitively intact elderly individuals. Rather, they are quantitative phenomena, being relatively more numerous in AD compared to controls. Application of standard criteria, therefore, yields a spectrum of diagnostic certainty in the case of Consortium to Establish a Registry for Alzheimer's Disease (CERAD) or National Institute on Aging (NIA)-Reagan criteria [4, 5], or a spectrum of stages in the case of Braak criteria [6]. It is also interesting to note that among the various neurodegenerative diseases, the one with the greatest overlap with "normal" aging, and the one with the most diagnostic uncertainty, is AD. Degenerative tauopathies such as corticobasal degeneration and progressive supranuclear palsy, for example, are essentially impossible to confuse with normal aging pathologically; this is also the case for synucleinopathies such as multiple system atrophy. Even Lewy body dementia, while nosologically cumbersome, is much more qualitative than classic AD. While Lewy bodies are described as age-related lesions, they are uncommonly seen in aged brains (i.e., are more "pathological"), whereas neurofibrillary tangles are essentially invariable in the medial temporal lobe of the elderly, as are amyloid deposits in the cerebral cortex.

In short, the diagnosis of AD on the basis of plaques and tangles at autopsy overlaps substantially with normal aging, and it follows that accumulation of plaques and tangles is as much an age-related process

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

as it is strictly pathological [7]. From the larger perspective, it makes much more mechanistic sense that accumulation of these proteins is more consequence than cause [8, 9]. Therapeutic targeting of earlier pathological processes is warranted for more effective therapy.

Ongoing investigations are aimed at targeting multiple metabolic abnormalities in AD. Such combinational drug therapies will hopefully reveal the effectiveness of drugs acting at multiple targets in the treatment of AD, rather than an isolated process, or an isolated protein that is more relevant to transgenic models and Mendelian conditions than it is to sporadic AD. In this review, we discuss a number of targets separately, namely Aβ, oxidative stress, lipid metabolism, inflammation, hormonal alterations with advanced age, phosphorylated tau, excitotoxicity, and cholinergic failure, bearing in mind that no one construct will likely represent the magic bullet for sporadic AD, and that multiple targets may give the greatest yield in the short term [10].

## 2 AMYLOID-β

Significant evidence in favor of the amyloid cascade hypothesis is the existence of familial "AD" in which germ line mutations in specific proteins lead to autosomal-dominant, early-onset AD. The specific proteins include amyloid-β protein precursor (AβPP) and presenilin-1 and -2. All three mutations affect processing of the Aβ peptides, and all are associated with prodigious accumulations of Aβ as well as neurofibrillary tangles with disease. Down syndrome subjects, who have an extra copy of chromosome 21 and, therefore, constitutively increased synthesis of AβPP, are predisposed to AD changes at a relatively young age. Down syndrome, therefore, represents another critical component of the amyloid cascade hypothesis.

Not surprisingly, a large body of knowledge regarding Aβ processing has accumulated. The Aβ peptide is a proteolytic cleavage product of AβPP, a type-1 transmembrane protein of unknown function. The two aspartyl proteases responsible for cleavage and conversion of AβPP to Aβ are referred to as β- and γ-secretases [11]. However, most AβPP molecules undergo cleavage by α-secretase, which cleaves this protein near the middle of the Aβ domain [12], resulting in a cleavage product with a large soluble ectodomain ($AβPP_s$-α) that is released into the extracellular space. The remaining C-terminal fragment can then be cleaved by γ-secretase to create a smaller fragment known as the p3 fragment. The function of such proteolytic processing in normal neurons has yet to be defined, although research has indicated that cleavage of AβPP by γ-secretase allows the release of the AβPP intracellular domain (AICD) to the nucleus where it is thought to participate in transcriptional signaling [11]. Cleavage of AβPP by β-secretase leaves a longer C-terminal fragment that is retained in the cellular membrane and is subjected to further cleavage by γ-secretase, which finally results in the Aβ peptide cleavage product.

β-Secretase has become an attractive therapeutic target because it initiates and is the rate-limiting step in the formation of Aβ (it is interesting to note that Aβ is constitutively secreted by mammalian cells and normally occurs in plasma and cerebrospinal fluid (CSF) [13, 14] ). Interestingly, inhibitors of β-secretase have similar chemical structures as AβPP. Sinha and Lieberburg developed StatV, a transition-state analog that spans the cleavage site of Swedish mutant AβPP for use as a β-secretase inhibitor [15]. OM99-2, a similar AβPP analog, was subsequently developed and is an even more potent β-secretase inhibitor. To date, however, available inhibitors are too large and lack sufficient specificity to be suitable for therapeutic use [16]. Since the discovery of presenilin-linked familial AD, the relationship between mutation and mechanism of AβPP processing has been the study of intensive investigation [17]. Early studies suggested that presenilin is a cofactor for γ-secretase [18] or that presenilin is itself a γ-secretase [19]. Later interpretations of experimental data indicated that presenilin is an unusual and unprecedented intramembrane-cleaving aspartyl protease [19]. Most recently, investigations utilizing genetic analysis of Notch signaling in *Caenorhabditis elegans* have revealed that γ-secretase is a multiprotein complex that has multiple regulatory subunits and that presenilin, nicastrin, aph-1, and pen-2 proteins are all required for γ-secretase activity [20, 21]. Nevertheless, regulation of presenilin and γ-secretase is still under investigation, and whether the extensive knowledge of the γ-secretase complex and its relationship with AβPP will translate into productive therapy remains to be determined.

As discussed, Aβ can normally be found in the CSF as a self-antigen, and cell-mediated immune responses would not be traditionally predicted to this peptide. Self-reactive T cells mediate a number of autoimmune diseases; however, no evidence exists showing deleterious effects of T-cell reactivity against Aβ. It has been established that activated T cells penetrate the Central Nervous System (CNS) [22] and that healthy elderly individuals and patients with AD have higher Aβ-reactive T-cell responses than middle-aged adults [23]. This suggests that the normal immune response includes an active response to remove excess Aβ from the brain. In light of this phenomenon, immunological approaches to clearing Aβ from the brain have been developed in recent years. Active immunization and passive immunization strategies have been studied, although only active

immunization has reached clinical testing. Briefly, findings regarding active immunization with synthetic intact Aβ1-42 were excellent in transgenic mouse models producing human Aβ. A significant reduction of Aβ plaques were reported in mice immunized before and after plaque development [24]. Additionally, immunization allowed these animals to regain functional effects [25, 26]. The underlying mechanism behind these beneficial reports has not been reported, however, one possibility is that reduction in Aβ deposition is due to Fc-receptor-mediated microglial phagocytosis [24]. Passive immunization strategy with antibodies raised against Aβ have suggested that microglia are responsible for clearing Aβ plaques through Fc-receptor-mediated phagocytosis and subsequent degradation [27]. Clinical trials were initiated to test tolerance to treatment in humans shortly after animal trials were completed. Synthetic peptide Aβ1-42 was combined with adjuvant (QS-21) to create the immunogen AN-1792 (Elan Pharmaceuticals/Wyeth). AN-1792 reached phase IIa clinical trials in the United States and Europe; however, the program was suspended in January 2002 when approximately 5% of the patients in the active treatment group developed meningoencephalitis [28].

While the Aβ peptide continues to be a popular target for new therapies, it is relevant that Aβ plaques have not proven to directly correlate with dementia and that plaque-producing transgenic models are not directly associated with neurological deficits. These inconsistencies have led to a modification of the amyloid cascade hypothesis that seeks to establish the pathogenecity of Aβ. Current research indicates that Aβ assembly of soluble oligomers has been implicated in the physical degeneration of synapses [29], and in vivo experiments have established that oligomers cause a failure in long-term potentiation (LTP) of neurons in the hippocampus. Exciting data has also come from human amyoid precursor protein (hAPP) transgenic mice and the ability of antibodies directed against pathogenic oligomers to reverse memory failure; this reversal of memory failure occurred without a reduction in plaques [30]. This phenomenon supported the concept that plaques are not responsible for neurological deficits, but rather oligomers are responsible for synaptic failure [31].

The Aβ oligomers shown to have the properties of disrupting LTP and synaptic degeneration are often referred to as oligomeric ligands or amyloid beta-derived diffusible ligands (ADDLs). Given the nature of ADDLs, they appear to be an excellent therapeutic target for the prevention and treatment of AD. These ADDLs uniquely influence short-term synaptic plasticity and long-term neuronal survival. They are also considered excellent therapeutic targets due to their regional specificity, especially since ADDLs target hippocampal but not cerebellar neurons [32]. In this regard, it should be noted that the vaccination approach may lead to *increased* levels of ADDLs [33].

## 3 OXIDATIVE STRESS

Multiple lines of evidence link oxidative stress and AD. Reduced glucose metabolism and mitochondrial abnormalities [34] are associated with AD as mitochondria are considered the source of oxidative stress. Oxidative phosphorylation produces superoxide radicals as a by-product associated with the electron transport. While most of these free radicals are sequestered in the mitochondria, oxidative insult is exacerbated by age, metabolic demand, and disease [35]. Heightened superoxide radical formation in AD also correlates with heightened superoxide dismutase (SOD) levels that may allow the release of $H_2O_2$ from the mitochondria to the cytoplasm. Mitochondrial abnormalities have also been associated with deficiencies in enzyme activities, specifically α-ketoglutarate dehydrogenase complex, pyruvate dehydrogenase complex, and cytochrome oxidase in AD neurons [36–38]. Increased $H_2O_2$ in the cytoplasm may cause a localized increase in the concentration of reactive oxygen species, especially in the presence of redox-active metals [39]. It has been shown that metals are dysregulated in AD [40] and that redox-active transition metals are aberrantly accumulated in the cytoplasm of AD-susceptible neurons [41].

A potential role for heavy metals in AD pathogenesis has been suggested in a number of studies; indeed, there is a close association between redox-active iron and pathological lesions in AD [42]. It has also been shown that Aβ and AβPP exhibit physicochemical interactions with micromolar and submicromolar concentrations of metal ions [43–46]. The pharmacological basis of chelation therapy and concomitant side effects, however, are complex and multifaceted; to date, reproducible therapeutic efficacy has yet to be shown [47]. Alternate approaches, using nanoparticle technology, may circumvent such aspects [48]. Also interesting with respect to metal ions is that the overall concentration of zinc is decreased in AD brain [49]. This metal has antioxidant activity and has been evaluated as a therapeutic agent to prevent production of reactive oxygen species.

It is also interesting to note that the Aβ peptide is a strong redox-active agent capable of reducing transition metals in the cytoplasm and allowing for the conversion of molecular oxygen ($O_2$) to $H_2O_2$ [50–52].

Antioxidant projection via vitamin supplementation is another avenue for long-term and safe treatments of AD. The most notable antioxidant treatment that is currently available is vitamin E supplementation.

Chemically *RRR*-α-tocopherol is the naturally occurring stereoisomer of vitamin E and is secreted by the liver due to specific binding by liver α-tocopherol transfer protein (α-TTP). Vitamin E has been demonstrated to be a lipid-soluble, chain-breaking antioxidant [53]. Vitamin regimens that include vitamin E and ascorbic acid (vitamin C) have indicated that consuming organic antioxidants may be beneficial. Some studies have found that antioxidants blunt the cognitive decline in AD or slow disease progression [54, 55]. A large epidemiological study showed that use of vitamins E and C in combination was associated with reduced AD prevalence and incidence [56].

## 4 LIPID METABOLISM

Recent studies suggest that it is possible to reduce cerebral Aβ levels in vivo with the Food and Drug Adminstration (FDA)-approved drug simvastatin [57]. The mechanism, however, is unclear, although the possibility that it may be related to apolipoprotein E is intriguing. Unlike molecular gene mutations, *Apo ε4* genotypes are not implicated in the generation of Aβ peptide; however, the *ε4* allele is associated with lowering the age of AD onset. Midlife high blood pressure and high cholesterol levels are associated with increased risk of developing AD later in life. While the data and mechanisms implicating cholesterol and lipid transport with development of AD are not well established, a number of studies indicate that the use of statins significantly reduces the risk AD [58, 59].

Experimentally, primary neurons treated with statins halt production of Aβ [60] and reduce the level of Aβ1-42. However, the mechanism underlying the action of statins remains unclear. Statins also act to decrease cellular cholesterol by inhibition of methylglutaryl coenzyme A (HMG-CoA). Given these simultaneous actions, it cannot be strictly stated whether statins or the reduction in membrane cholesterol are responsible for reduced Aβ levels.

## 5 INFLAMMATION

Exciting findings regarding the use of nonsteroidal anti-inflammatory drugs (NSAIDs) were originally reported in epidemiological data showing that classical NSAIDs could prevent or retard AD [61, 62]. Data based on positron emission tomography (PET) has shown that inflammation and activation of microglia are early pathogenic events that precede cerebral atrophy [63]. Additional evidence that inflammation mediates development of AD comes from studies demonstrating that inflammation from head trauma plays a role in initiating AD [64]. It is hypothesized that Aβ peptide is responsible for the innate immune response that occurs in AD patients. Additional findings suggest that Aβ deposition and plaque formation are associated with an innate immune response that includes activation of complement [65], secretion of pro-inflammatory cytokines, expression of chemokines, and excretion of nitric oxide, which mediates apoptosis [66, 67]. Experiments with transgenic mice show that the activation of the complement cascade may be beneficial for the brain [67]. However, a number of studies have shown that attenuating this inflammatory innate immune response may delay the onset and slow the progress of AD [68]. It appears that Aβ is responsible for activation of microglia and astrocytes since toxic forms of Aβ that are not completely removed induce chronic immune responses.

NSAIDs have been used extensively in medical practice and are currently prescribed for a number of inflammation-related ailments. NSAIDs downregulate the pro-inflammatory signals sent out by glia and astrocytes. In addition to anti-inflammatory action, NSAIDs have been shown to alter the production of Aβ1-42. It has been shown that some NSAIDs may reduce the risk of AD by lowering Aβ1-42. Production of Aβ1-42 is mediated through a pathway that involves a small G protein call Rho. NSAIDs have been shown to inhibit the activity of Rho and decrease Aβ1-42 production in transgenic mice. It has also been shown that NSAIDs that do not block Rho activity do not lower Aβ1-42 levels. The downstream effector proteins of Rho, Rock 1 and 2, are also pharmacological targets that could prevent Aβ1-42 production.

Epidemiologically, a number of case-control and population-based cross-sectional studies have been conducted to evaluate the relationship between the use of NSAIDs and the risk of AD. Data from a co-twin control study found an inverse relationship between anti-inflammatory treatments and AD [69]. The follow-up was conducted with siblings with a high level of familial AD and found that NSAIDs have protective effects [70]. Population-based studies have also suggested that anti-inflammatory drugs are protective against AD [71–73]. Whether these findings will translate into a change in the indications for NSAID therapy remains to be seen. Moreover, a thorough analysis of this data was undertaken by Launer, who concludes that no recommendation can be made concerning when or how a person can take NSAIDs to reduce risk of developing AD [74].

## 6 HORMONAL CHANGES WITH AGE

Some epidemiological gender studies concerning the prevalence and incidence of AD have indicated that

women have an almost twofold higher risk of developing AD [75]. Postmenopausal estrogen deficiency is hypothesized as a basis for the effect. This in turn has prompted the idea of hormone replacement therapy (HRT) as a possible treatment.

Early studies concerning the use of HRT as a treatment for AD came to divergent conclusions concerning HRT-based therapies. These findings were reported for therapy that included estrogen in addition to progestin regardless of time-course regimens. Evidence supporting the protective effects of HRT in postmenopausal women was reported for a large epidemiological study [76] and confirmed further by studies indicating that women with high levels of endogenous estrogen were less prone to develop AD [77]. However, HRT was found to be ineffective in preventing AD when administered in postmenopausal women aged over 65 years [78].

The mechanistic basis for HRT in AD remains to be resolved, although progress is being made. Notable in this regard, a novel therapy that uses leuprolide acetate, a gonadotropin-releasing hormone (GnRH) agonist that suppresses production of luteinizing hormone by downregulating pituitary GnRH receptors, is currently being studied with encouraging results [79–81].

## 7 CELL CYCLE

Primary neurons in the normal brain are viewed as being quiescent and in $G_0$. However, in AD, multiple lines of evidence suggest that neurons vulnerable to degeneration emerge from this postmitotic state—phenotypically suggestive of cells that are cycling, rather than the normal terminally differentiated nondividing state [10, 82–101]. The successful duplication of deoxyribonucleic acid (DNA) [102] shows that neurons successfully complete S phase and precludes the possibility that the re-expression of various cell cycle markers is an epiphenomena. However, while various mitotic markers are upregulated in vulnerable neurons in AD, no evidence of actual mitosis has ever been found. Given this lack of evidence for successful completion of the cell cycle, it is likely that the reactivation of cell cycle machinery in postmitotic neurons leads to their death. The cell cycle may, therefore, represent a novel approach to the treatment of AD [103, 104].

## 8 PHOSPHORYLATED TAU

Abnormally phosphorylated tau is the major protein component of neurofibrillary tangles, the first lesion to be characterized and associated with AD. Hyperphosphorylation causes disengagement of tau from microtubules and aggregation of the filamentous protein. [105]. Importantly, hyperphosphorylated tau (phospho-tau), which accumulates into paired helical filaments/neurofibrillary tangles, is not toxic to the neuron and does not induce the apoptotic cascade, and as much as 40% of phospho-tau is not aggregated into neurofibrillary tangles [106]. Pathologically active phospho-tau does not bind tubulin but instead sequesters normal tau in addition to other microtubule-associated proteins, which interrupts assembly and disassembly of normal microtubules [107, 108]. Tau is the substrate for a number of kinases, such as glycogen synthase kinase 3 (GSK3), cyclin-dependent protein kinase 5, and protein kinase A [109]. Dephosphorylation of phospho-tau protein has been shown in vivo by alkaline phosphatase, protein phosphatase 2A (PP-2A), PP-2B, and PP-1, all of which convert it into a normal state lacking toxic properties [110, 111]. PP-2A and PP-1 are responsible for 90% of the serine/threonine protein phosphatase activity in mammalian cells [112]. Activities of PP-2A and PP-1 have been shown to be compromised in AD [110, 113], indicating that insufficient dephosphorylation could be responsible for the appearance of phospho-tau. These mechanisms for aberrant accumulation of phospho-tau generate a number of novel therapeutic targets, currently in the initial stages of design and testing. For instance, restoring PP-2A activity by repressing PP-2A inhibitors, or inhibiting the activity of the kinases that phosphorylate tau, will prevent abnormal phosphorylation from occurring. A second strategy involves preventing the ability of phospho-tau to sequester normal tau, allowing normal function of other microtubule-associated proteins, thus preserving cellular integrity. Drug designs capable of achieving these goals are still in the early stages of development. High-throughput assays that are capable of measuring antisequestering effects are being addressed by researchers in the field [114].

## 9 EXCITOTOXICITY

Almost all neurons of the CNS carry the *N*-methyl-D-aspartate subtype of ionotrophic L-glutamate receptors. These receptors mediate postsynaptic influx of $Ca^{2+}$ after neurotransmission. It is theorized that excitotoxicity, resulting from excessive activation of these receptors, increases the vulnerability of CNS neurons, leading to neuronal degeneration characteristic of the development of AD.

In one school of thought, clinical manifestations early in the disease are the result of synaptic failure [31], while disease severity is the result of synaptic loss [115]. Memantine, the most recently approved drug for AD treatment, prevents pathological excitotoxicity at the

synapse and is believed to preserve physiological activation [116–118].

Early pathological changes in AD apparently occur in the medial temporal lobe, including the transentorhinal allocortex [119] and hippocampal formation. Neocortical glutamatergic neurons in layers III and IV are also vulnerable [120]. Excitatory amino acids, such as L-gluamate (L-Glu), are among the neurotransmitters used by these cells, although L-Glu distribution within the CNS is extensive. At basal conditions, the concentration of L-Glu in the synaptic cleft is around 0.6 $\mu M$ [121], while synaptic transmission allows for a localized rise to 10 $\mu M$ within the synaptic cleft [122]. Termination of excitation occurs by high-affinity reuptake on both pre- and postsynaptic neuronal cell membranes and the membranes of adjacent glial cells [123]. When uptake mechanisms are impaired, excitatory amino acids cause selective neuronal toxicity, through N-methyl-D-aspartic acid (NMDA)-mediated $Ca^{2+}$ influx [124].

A potentially important role of astrocytes is apparent in their intrinsic function to convert glutamate to glutamine by glutamine synthetase. Glutamine is then released by astrocytes and taken up by neurons where it is converted to glutamate in the mitochondria by the mitochondrial glutaminase. Recent studies have shown that glutamine synthetase is oxidized in the brains of individuals with AD [125]. This may explain the reported decrease in the activity of glutamine synthetase in AD patients [126]. While antioxidants may prove to be effective at preventing pathological oxidation, one therapy thought to block excitotoxicity through interaction with membrane receptors was also developed. Memantine is a blocker of glutamate-gated NMDA channels [127] and has become a popular therapy in recent years.

Memantine is thought to prevent excitotoxicity mediated by glutamate-gated NMDA receptors by blocking pathological activation. This property has been shown to occur due to the rapid, voltage-dependent interaction between memantine and the NMDA receptor channel [128]. Memantine blocks the receptor channel under normal physiological conditions and leaves the receptor channel upon activation. Additionally, memantine remains in the channel during pathological activation, preventing excitotoxicity and allowing for normal activation and memory formation. In animal models, memantine has been shown to prevent excitotoxic activation of glutamate receptors and preserve their physiological activation [116, 117, 129]. Clinical safety has been thoroughly studied in subjects diagnosed with mild to moderate dementia [130, 131], demonstrating significant improvements in psychopathological performance and behavior. Additional studies in a severely demented population [132] found that memantine treatment significantly improved the functional capabilities of these patients.

## 10 CHOLINERGIC DEFICITS

AD can be characterized as a deficit in cholinergic neurotransmission due to loss of cholinergic neurons of the basal nucleus of Meynert [133]. Evidence has demonstrated that the enzymes involved in synthesis [choline acetyltransferase (ChAT)] and degradation [acetylcholinesterase (AChE)] have reduced activity and may be responsible for this deficit [134, 135]. In light of this neurotransmitter deficiency, strategies to boost the level of acetylcholine have been developed for the treatment of AD. This has lead to the development of AChE inhibitors (AChEIs) that prevent the breakdown of the neurotransmitter inside the synaptic cleft. The first drug approved for the treatment of AD in 1993 was 1,2,3,4-tetrahydro-9-aminoacridine (tacrine) [136]. Currently, three other AChEIs are available on the market from different manufacturers. The second generation of drugs has proven to be more clinically effective and produce less severe side effects than the first generation. These drugs are currently prescribed for patients with mild to moderate progression of AD under the following names: Donepezil (Eisai Company and Pfizer Inc.), ENA-713 (Novartis Pharmaceuticals), and Galantamine (Hoechst Marion Roussel Inc., Shire Pharmaceutical Group, and Janssen Pharmaceutical).

Cholinergic therapy is considered a short-term intervention for the symptomatic treatment of AD. The effect of these inhibitors is to cognitively stabilize 50% of patients in treatment for 1 year. Clinical studies have indicated that beneficial effects can be maintained up to 36 months [137–139]. The stabilization of cognitive decline has lead researchers to investigate whether AChEIs can affect the amyloid cascade or the neurotoxicity associated with Aβ [140].

Cholinesterase inhibitors are currently being investigated to find additional benefits such as delay of onset and time of efficacy. To test these criteria, research has moved from AD patients to patients with mild cognitive impairment (MCI) [141]. MCI patients are identified due to their high risk of developing dementia. Two major studies are underway in the United States and Europe to address cholinesterase inhibitor use in populations of MCI patients [142].

Long-term stabilization that can occur when AD patients are treated with AChEI therapy has been associated with the ability of AChEI to affect AβPP metabolism [137]. Evidence has been published showing that AChEIs increase secretion of AβPP, thus slowing down the production of Aβ peptide and amyloidogenic com-

pounds [143, 144]. This suggests that AChEI could be administered at much earlier preclinical stages in at-risk individuals, and that combinatorial therapies represent future alternatives to AChEI monotreatment [137].

Neurotropins are currently being investigated as treatments to supplement the other therapeutic targets addressed. Neurotropins are of particular interest due to their ability to upregulate functions of cholinergic neurons [145]. Evidence has shown that neurotropins share the same signaling pathways as Aβ pathways mediating toxicity in neurons [146, 147]. Aβ has been found to bind to the p75 neurotropin receptor ($p75^{NTR}$), which causes $p75^{NTR}$-mediated cell death [148]. A precursor of nerve growth factor, proNGF, has been associated with signaling through $p75^{NTR}$ [149]. Research concerning proNGF in AD patients has shown the levels are increased in the brain [150].

Neurotropins are difficult to deliver effectively due to the short half-lives, on the order of minutes, significantly hampering the ability of these compounds to reach the central nervous system [151]. Additionally, neurotropins affect a broad range of biological signals, including sprouting of sympathetic nerves and upregulation of pain mechanisms. The development of neurotropin mimetics, altering only a subset of signaling cascades, is a partial answer to these side effects.

Investigations of NGF interaction with $p75^{NTR}$ have shown that regulation of different intracellular effector molecules occurs upon binding [147, 152]. Studies concerning peptidomimetics corresponding to specific regions of NGF have shown that their effects can occur via $p75^{NTR}$-dependent mechanisms to prevent neuronal death. These studies have included NGF peptidomimetics corresponding to loops 1 and 4 of NGF globular protein, both preventing neuronal death mechanisms via $p75^{NTR}$ receptor [153–155]. Recent findings have shown that a number of adaptor proteins are involved in the survival and death signaling mediated via $p75^{NTR}$ [156–158]. An interesting finding concerning these adaptors is that the selection of an adaptor is regulated by the nature of the ligand, which suggests that peptidomimetics may be able to regulate adaptor proteins [159]. These findings provide further support that peptidomimetics could possibly trigger $p75^{NTR}$ signaling in a manner to prevent cell death in AD patients.

## 11 CONCLUSION

Sporadic AD is a complex process that develops over decades and involves biochemical dysfunction in multiple arenas. While Aβ and tau are attractive targets for therapy, both accumulate as late phenomena in the disease process, particularly in sporadic disease. The lack of specificity of plaques and tangles for AD, particularly in early and intermediate stages of the disease, and with advanced age, highlight the need to look beyond Aβ and tau for early pathogenic factors and treatment. Indeed, one school of thought is that Aβ and tau are protective adaptations to disease. Oxidative stress and cell cycle alteration, for example, are known to predate clinical symptoms and pathological accumulations of Aβ and tau. Moreover, the strides that have been made in lipid metabolism with statins, and in neuroprotection from anti-inflammatory treatment, as well as antiexcitotoxic and cholinergic treatment, emphasize the point that AD is complex and that any one strategy or construct may not be sufficient [10]. As knowledge expands, the list of avenues for treatment will only expand, and an open-minded approach to treatment will be all the more necessary.

## ACKNOWLEDGMENTS

We thank Margaret Hampton for title inspiration. Margaret, and her husband Al, caregiver and patient, respectively, represent two out of the many millions touched by this devastating disease and to whom this work is dedicated.

Work in the authors' laboratories is supported by the National Institutes of Health and the Alzheimer's Association.

## REFERENCES

1. Smith MA (1998). Alzheimer disease. *Int Rev Neurobiol 42*, 1–54.
2. Hebert LE, Scherr PA, Bienias JL, Bennett DA, Evans DA (2003). Alzheimer disease in the U.S. population: Prevalence estimates using the 2000 census. *Arch Neurol 60*, 1119–1122.
3. Ernst RL, Hay JW (1994). The U.S. economic and social costs of Alzheimer's disease revisited. *Am J Public Health 84*, 1261–1264.
4. Mirra SS, Heyman A, McKeel D, Sumi SM, Crain BJ, Brownlee LM, Vogel FS, Hughes JP, van Belle G, Berg L (1991). The Consortium to Establish a Registry for Alzheimer's Disease (CERAD). Part II. Standardization of the neuropathologic assessment of Alzheimer's disease. *Neurol 41*, 479–486.
5. Khachaturian ZS (1985). Diagnosis of Alzheimer's disease. *Arch Neurol 42*, 1097–1105.
6. Evans DA, Funkenstein HH, Albert MS, Scherr PA, Cook NR, Chown MJ, Hebert LE, Hennekens CH, Taylor JO (1989). Prevalence of Alzheimer's disease in a community population of older persons. Higher than previously reported. *JAMA 262*, 2551–2556.

7. Castellani RJ, Lee HG, Zhu X, Nunomura A, Perry G, Smith MA (2006). Neuropathology of Alzheimer disease: Pathognomonic but not pathogenic. *Acta Neuropathol 111*, 503–509.
8. Rottkamp CA, Atwood CS, Joseph JA, Nunomura A, Perry G, Smith MA (2002). The state versus amyloid-beta: The trial of the most wanted criminal in Alzheimer disease. *Peptides 23*, 1333–1341.
9. Lee HG, Perry G, Moreira PI, Garrett MR, Liu Q, Zhu X, Takeda A, Nunomura A, Smith MA (2005). Tau phosphorylation in Alzheimer's disease: Pathogen or protector? *Trends Mol Med 11*, 164–169.
10. Zhu X, Raina AK, Perry G, Smith MA (2004). Alzheimer's disease: The two-hit hypothesis. *Lancet Neurol 3*, 219–226.
11. Kimberly WT, Zheng JB, Guenette SY, Selkoe DJ (2001). The intracellular domain of the beta-amyloid precursor protein is stabilized by Fe65 and translocates to the nucleus in a notch-like manner. *J Biol Chem 276*, 40288–40292.
12. Sisodia SS, Koo EH, Beyreuther K, Unterbeck A, Price DL (1990). Evidence that beta-amyloid protein in Alzheimer's disease is not derived by normal processing. *Science 248*, 492–495.
13. Shoji M, Golde TE, Ghiso J, Cheung TT, Estus S, Shaffer LM, Cai XD, McKay DM, Tintner R, Frangione B, Younkin S (1992). Production of the Alzheimer amyloid beta protein by normal proteolytic processing. *Science 258*, 126–129.
14. Haass C, Schlossmacher MG, Hung AY, Vigo-Pelfrey C, Mellon A, Ostaszewski BL, Lieberburg I, Koo EH, Schenk D, Teplow DB, Selkoe DJ (1992). Amyloid beta-peptide is produced by cultured cells during normal metabolism. *Nature 359*, 322–325.
15. Sinha S, Lieberburg I (1999). Cell mechanisms of beta-amyloid production and secretion. *Proc Nat Acad Sci U S A 96*, 11049–11053.
16. Gruninger-Leitch F, Schlatter D, Kung E, Nelbock P, Dobeli H (2002). Substrate and inhibitor profile of BACE (beta-secretase) and comparison with other mammalian aspartic proteases. *J Biol Chem 277*, 4687–4693.
17. Scheuner D, Eckman C, Jensen M, Song X, Citron M, Suzuki N, Bird TD, Hardy J, Hutton M, Kukull W, Larson E, Levy-Lahad E, Viitanen M, Peskind E, Poorkaj P, Schellenberg G, Tanzi R, Wasco W, Lannfelt L, Selkoe D, Younkin S (1996). Secreted amyloid beta-protein similar to that in the senile plaques of Alzheimer's disease is increased in vivo by the presenilin 1 and 2 and APP mutations linked to familial Alzheimer's disease. *Nat Med 2*, 864–870.
18. De Strooper B, Saftig P, Craessaerts K, Vanderstichele H, Guhde G, Annaert W, Von Figura K, Van Leuven F (1998). Deficiency of presenilin-1 inhibits the normal cleavage of amyloid precursor protein. *Nature 391*, 387–390.
19. Wolfe MS, Xia W, Moore CL, Leatherwood DD, Ostaszewski B, Rahmati T, Donkor IO, Selkoe DJ (1999). Peptidomimetic probes and molecular modeling suggest that Alzheimer's gamma-secretase is an intramembrane-cleaving aspartyl protease. *Biochemistry 38*, 4720–4727.
20. Goutte C, Tsunozaki M, Hale VA, Priess JR (2002). APH-1 is a multipass membrane protein essential for the Notch signaling pathway in *Caenorhabditis elegans* embryos. *Proc Nat Acad Sci U S A 99*, 775–779.
21. Francis R, McGrath G, Zhang J, Ruddy DA, Sym M, Apfeld J, Nicoll M, Maxwell M, Hai B, Ellis MC, Parks AL, Xu W, Li J, Gurney M, Myers RL, Himes CS, Hiebsch R, Ruble C, Nye JS, Curtis D (2002). aph-1 and pen-2 are required for Notch pathway signaling, gamma-secretase cleavage of betaAPP, and presenilin protein accumulation. *Dev Cell 3*, 85–97.
22. Becher B, Prat A, Antel JP (2000). Brain-immune connection: Immuno-regulatory properties of CNS-resident cells. *Glia 29*, 293–304.
23. Monsonego A, Zota V, Karni A, Krieger JI, Bar-Or A, Bitan G, Budson AE, Sperling R, Selkoe DJ, Weiner HL (2003). Increased T cell reactivity to amyloid beta protein in older humans and patients with Alzheimer disease. *J Clin Invest 112*, 415–422.
24. Schenk D, Barbour R, Dunn W, Gordon G, Grajeda H, Guido T, Hu K, Huang J, Johnson-Wood K, Khan K, Kholodenko D, Lee M, Liao Z, Lieberburg I, Motter R, Mutter L, Soriano F, Shopp G, Vasquez N, Vandevert C, Walker S, Wogulis M, Yednock T, Games D, Seubert P (1999). Immunization with amyloid-beta attenuates Alzheimer-disease-like pathology in the PDAPP mouse. *Nature 400*, 173–177.
25. Morgan D, Diamond DM, Gottschall PE, Ugen KE, Dickey C, Hardy J, Duff K, Jantzen P, DiCarlo G, Wilcock D, Connor K, Hatcher J, Hope C, Gordon M, Arendash GW (2000). A beta peptide vaccination prevents memory loss in an animal model of Alzheimer's disease. *Nature 408*, 982–985.
26. Janus C, Pearson J, McLaurin J, Mathews PM, Jiang Y, Schmidt SD, Chishti MA, Horne P, Heslin D, French J, Mount HT, Nixon RA, Mercken M, Bergeron C, Fraser PE, St George-Hyslop P, Westaway D (2000). A beta peptide immunization reduces behavioural impairment and plaques in a model of Alzheimer's disease. *Nature 408*, 979–982.
27. Bard F, Cannon C, Barbour R, Burke RL, Games D, Grajeda H, Guido T, Hu K, Huang J, Johnson-Wood K, Khan K, Kholodenko D, Lee M, Lieberburg I, Motter R, Nguyen M, Soriano F, Vasquez N, Weiss K, Welch B, Seubert P, Schenk D, Yednock T (2000). Peripherally administered antibodies against amyloid beta-peptide enter the central nervous system and reduce pathology in a mouse model of Alzheimer disease. *Nat Med 6*, 916–919.
28. Nicoll JA, Wilkinson D, Holmes C, Steart P, Markham H, Weller RO (2003). Neuropathology of human Alzheimer

disease after immunization with amyloid-beta peptide: a case report. *Nat Med 9*, 448–452.

29. Mucke L, Masliah E, Yu GQ, Mallory M, Rockenstein EM, Tatsuno G, Hu K, Kholodenko D, Johnson-Wood K, McConlogue L (2000). High-level neuronal expression of abeta 1-42 in wild-type human amyloid protein precursor transgenic mice: Synaptotoxicity without plaque formation. *J Neurosci 20*, 4050–4058.
30. Dodart JC, Bales KR, Gannon KS, Greene SJ, DeMattos RB, Mathis C, DeLong CA, Wu S, Wu X, Holtzman DM, Paul SM (2002). Immunization reverses memory deficits without reducing brain Abeta burden in Alzheimer's disease model. *Nat Neurosci 5*, 452–457.
31. Selkoe DJ (2002). Alzheimer's disease is a synaptic failure. *Science 298*, 789–791.
32. Klein WL, Krafft GA, Finch CE (2001). Targeting small Abeta oligomers: The solution to an Alzheimer's disease conundrum? *Trends Neurosci 24*, 219–224.
33. Lee HG, Zhu X, Nunomura A, Perry G, Smith MA (2006). Amyloid-beta vaccination: Testing the amyloid hypothesis?: Heads we win, tails you lose! *Am J Pathol 169*, 738–739.
34. Reiman EM, Caselli RJ, Yun LS, Chen K, Bandy D, Minoshima S, Thibodeau SN, Osborne D (1996). Preclinical evidence of Alzheimer's disease in persons homozygous for the epsilon 4 allele for apolipoprotein E. *N Engl J Med 334*, 752–758.
35. Smith CD, Carney JM, Tatsumo T, Stadtman ER, Floyd RA, Markesbery WR (1992). Protein oxidation in aging brain. *Ann N Y Acad Sci 663*, 110–119.
36. Yates CM, Butterworth J, Tennant MC, Gordon A (1990). Enzyme activities in relation to pH and lactate in postmortem brain in Alzheimer-type and other dementias. *J Neurochem 55*, 1624–1630.
37. Mastrogiacomo F, Bergeron C, Kish SJ (1993). Brain alpha-ketoglutarate dehydrogenase complex activity in Alzheimer's disease. *J Neurochem 61*, 2007–2014.
38. Simonian NA, Hyman BT (1994). Functional alterations in Alzheimer's disease: Selective loss of mitochondrial-encoded cytochrome oxidase mRNA in the hippocampal formation. *J Neuropathol Exp Neurol 53*, 508–512.
39. Lee HG, Zhu X, O'Neill MJ, Webber K, Casadesus G, Marlatt M, Raina AK, Perry G, Smith MA (2004). The role of metabotropic glutamate receptors in Alzheimer's disease. *Acta Neurobiol Exp (Wars) 64*, 89–98.
40. Smith MA, Wehr K, Harris PL, Siedlak SL, Connor JR, Perry G (1998). Abnormal localization of iron regulatory protein in Alzheimer's disease. *Brain Res 788*, 232–236.
41. Smith MA, Richey Harris PL, Sayre LM, Beckman JS, Perry G (1997). Widespread peroxynitrite-mediated damage in Alzheimer's disease. *J Neurosci 17*, 2653–2657.
42. Smith MA, Harris PL, Sayre LM, Perry G (1997). Iron accumulation in Alzheimer disease is a source of redox-generated free radicals. *Proc Nat Acad Sci U S A 94*, 9866–9868.
43. Bush AI, Multhaup G, Moir RD, Williamson TG, Small DH, Rumble B, Pollwein P, Beyreuther K, Masters CL (1993). A novel zinc(II) binding site modulates the function of the beta A4 amyloid protein precursor of Alzheimer's disease. *J Biol Chem 268*, 16109–16112.
44. Bush AI, Pettingell WH, Multhaup G, Paradis M, Vonsattel JP, Gusella JF, Beyreuther K, Masters CL, Tanzi RE (1994). Rapid induction of Alzheimer A beta amyloid formation by zinc. *Science 265*, 1464–1467.
45. Multhaup G, Bush AI, Pollwein P, Masters CL (1994). Interaction between the zinc (II) and the heparin binding site of the Alzheimer's disease beta A4 amyloid precursor protein (APP). *FEBS Lett 355*, 151–154.
46. Huang X, Atwood CS, Moir RD, Hartshorn MA, Vonsattel JP, Tanzi RE, Bush AI (1997). Zinc-induced Alzheimer's Abeta1-40 aggregation is mediated by conformational factors. *J Biol Chem 272*, 26464–26470.
47. Cuajungco MP, Faget KY, Huang X, Tanzi RE, Bush AI (2000). Metal chelation as a potential therapy for Alzheimer's disease. *Ann N Y Acad Sci 920*, 292–304.
48. Liu G, Garrett MR, Men P, Zhu X, Perry G, Smith MA (2005). Nanoparticle and other metal chelation therapeutics in Alzheimer disease. *Biochim Biophy Acta 1741*, 246–252.
49. Panayi AE, Spyrou NM, Iversen BS, White MA, Part P (2002). Determination of cadmium and zinc in Alzheimer's brain tissue using inductively coupled plasma mass spectrometry. *J Neurol Sci 195*, 1–10.
50. Bondy SC, Guo-Ross SX, Truong AT (1998). Promotion of transition metal-induced reactive oxygen species formation by beta-amyloid. *Brain Res 799*, 91–96.
51. Huang X, Atwood CS, Hartshorn MA, Multhaup G, Goldstein LE, Scarpa RC, Cuajungco MP, Gray DN, Lim J, Moir RD, Tanzi RE, Bush AI (1999). The A beta peptide of Alzheimer's disease directly produces hydrogen peroxide through metal ion reduction. *Biochemistry 38*, 7609–7616.
52. Sayre LM, Perry G, Harris PL, Liu Y, Schubert KA, Smith MA (2000). In situ oxidative catalysis by neurofibrillary tangles and senile plaques in Alzheimer's disease: A central role for bound transition metals. *J Neurochem 74*, 270–279.
53. Burton GW, Joyce A, Ingold KU (1982). First proof that vitamin E is major lipid-soluble, chain-breaking antioxidant in human blood plasma. *Lancet 2*, 327.
54. Jama JW, Launer LJ, Witteman JC, den Breeijen JH, Breteler MM, Grobbee DE, Hofman A (1996). Dietary antioxidants and cognitive function in a population-based sample of older persons. The Rotterdam Study. *Am J Epidemiol 144*, 275–280.
55. Perrig WJ, Perrig P, Stahelin HB (1997). The relation between antioxidants and memory performance in the old and very old. *J Am Geriatr Soc 45*, 718–724.

56. Zandi PP, Anthony JC, Khachaturian AS, Stone SV, Gustafson D, Tschanz JT, Norton MC, Welsh-Bohmer KA, Breitner JC (2004). Reduced risk of Alzheimer disease in users of antioxidant vitamin supplements: The Cache County Study. *Arch Neurol 61*, 82–88.
57. Fassbender K, Simons M, Bergmann C, Stroick M, Lutjohann D, Keller P, Runz H, Kuhl S, Bertsch T, von Bergmann K, Hennerici M, Beyreuther K, Hartmann T (2001). Simvastatin strongly reduces levels of Alzheimer's disease beta-amyloid peptides Abeta 42 and Abeta 40 in vitro and in vivo. *Proc Nat Acad Sci U S A 98*, 5856–5861.
58. Wolozin B, Kellman W, Ruosseau P, Celesia GG, Siegel G (2000). Decreased prevalence of Alzheimer disease associated with 3-hydroxy-3-methyglutaryl coenzyme A reductase inhibitors. *Arch Neurol 57*, 1439–1443.
59. Jick H, Zornberg GL, Jick SS, Seshadri S, Drachman DA (2000). Statins and the risk of dementia. *Lancet 356*, 1627–1631.
60. Simons M, Keller P, De Strooper B, Beyreuther K, Dotti CG, Simons K (1998). Cholesterol depletion inhibits the generation of beta-amyloid in hippocampal neurons. *Proc Nat Acad Sci U S A 95*, 6460–6464.
61. McGeer PL, Schulzer M, McGeer EG (1996). Arthritis and anti-inflammatory agents as possible protective factors for Alzheimer's disease: A review of 17 epidemiologic studies. *Neurol 47*, 425–432.
62. in t' Veld BA, Ruitenberg A, Hofman A, Launer LJ, van Duijn CM, Stijnen T, Breteler MM, Stricker BH (2001). Nonsteroidal antiinflammatory drugs and the risk of Alzheimer's disease. *N Engl J Med 345*, 1515–1521.
63. Cagnin A, Brooks DJ, Kennedy AM, Gunn RN, Myers R, Turkheimer FE, Jones T, Banati RB (2001). In-vivo measurement of activated microglia in dementia. *Lancet 358*, 461–467.
64. Nandoe RD, Scheltens P, Eikelenboom P (2002). Head trauma and Alzheimer's disease. *J Alzheimers Dis 4*, 303–308.
65. Rogers J, Cooper NR, Webster S, Schultz J, McGeer PL, Styren SD, Civin WH, Brachova L, Bradt B, Ward P, Lieberburg I (1992). Complement activation by beta-amyloid in Alzheimer disease. *Proc Nat Acad Sci U S A 89*, 10016–10020.
66. Griffin WS, Stanley LC, Ling C, White L, MacLeod V, Perrot LJ, White CL 3rd, Araoz C (1989). Brain interleukin 1 and S-100 immunoreactivity are elevated in Down syndrome and Alzheimer disease. *Proc Nat Acad Sci U S A 86*, 7611–7615.
67. Dickson DW, Lee SC, Mattiace LA, Yen SH, Brosnan C (1993). Microglia and cytokines in neurological disease, with special reference to AIDS and Alzheimer's disease. *Glia 7*, 75–83.
68. Akiyama H, Barger S, Barnum S, Bradt B, Bauer J, Cole GM, Cooper NR, Eikelenboom P, Emmerling M, Fiebich BL, Finch CE, Frautschy S, Griffin WS, Hampel H, Hull M, Landreth G, Lue L, Mrak R, Mackenzie IR, McGeer PL, O'Banion MK, Pachter J, Pasinetti G, Plata-Salaman C, Rogers J, Rydel R, Shen Y, Streit W, Strohmeyer R, Tooyoma I, Van Muiswinkel FL, Veerhuis R, Walker D, Webster S, Wegrzyniak B, Wenk G, Wyss-Coray T (2000). Inflammation and Alzheimer's disease. *Neurobiol Aging 21*, 383–421.
69. Breitner JC, Gau BA, Welsh KA, Plassman BL, McDonald WM, Helms MJ, Anthony JC (1994). Inverse association of anti-inflammatory treatments and Alzheimer's disease: Initial results of a co-twin control study. *Neurol 44*, 227–232.
70. Breitner JC, Welsh KA, Helms MJ, Gaskell PC, Gau BA, Roses AD, Pericak-Vance MA, Saunders AM (1995). Delayed onset of Alzheimer's disease with nonsteroidal anti-inflammatory and histamine H2 blocking drugs. *Neurobiol Aging 16*, 523–530.
71. Anthony JC, Breitner JC, Zandi PP, Meyer MR, Jurasova I, Norton MC, Stone SV (2000). Reduced prevalence of AD in users of NSAIDs and H2 receptor antagonists: The Cache County study. *Neurol 54*, 2066–2071.
72. The Canadian Study of Health and Aging (1994). Risk factors for Alzheimer's disease in Canada. *Neurol 44*, 2073–2080.
73. Beard CM, Waring SC, O'Brien PC, Kurland LT, Kokmen E (1998). Nonsteroidal anti-inflammatory drug use and Alzheimer's disease: A case-control study in Rochester, Minnesota, 1980 through 1984. *Mayo Clinic Proc 73*, 951–955.
74. Launer L (2003). Nonsteroidal anti-inflammatory drug use and the risk for Alzheimer's disease: Dissecting the epidemiological evidence. *Drugs 63*, 731–739.
75. Jorm AF, Jolley D (1998). The incidence of dementia: A meta-analysis. *Neurol 51*, 728–733.
76. Tang MX, Jacobs D, Stern Y, Marder K, Schofield P, Gurland B, Andrews H, Mayeux R (1996). Effect of oestrogen during menopause on risk and age at onset of Alzheimer's disease. *Lancet 348*, 429–432.
77. Manly JJ, Merchant CA, Jacobs DM, Small SA, Bell K, Ferin M, Mayeux R (2000). Endogenous estrogen levels and Alzheimer's disease among postmenopausal women. *Neurol 54*, 833–837.
78. Shumaker SA, Legault C, Rapp SR, Thal L, Wallace RB, Ockene JK, Hendrix SL, Jones BN 3rd, Assaf AR, Jackson RD, Kotchen JM, Wassertheil-Smoller S, Wactawski-Wende J (2003). Estrogen plus progestin and the incidence of dementia and mild cognitive impairment in postmenopausal women: The Women's Health Initiative Memory Study: A randomized controlled trial. *JAMA 289*, 2651–2662.
79. Casadesus G, Zhu X, Atwood CS, Webber KM, Perry G, Bowen RL, Smith MA (2004). Beyond estrogen: Targeting gonadotropin hormones in the treatment of Alzheimer's disease. *Curr Drug Targets CNS Neurol Disord 3*, 281–285.
80. Casadesus G, Webber KM, Atwood CS, Pappolla MA, Perry G, Bowen RL, Smith MA (2006). Luteinizing hormone modulates cognition and amyloid-beta deposi-

tion in Alzheimer APP transgenic mice. *Biochim Biophys Acta 1762*, 447–452.

81. Casadesus G, Milliken EL, Webber KM, Bowen RL, Lei Z, Rao CV, Atwood CS, Perry G, Keri RA, Smith MA (2007). Increases in luteinizing hormone are associated with declines in cognitive performance. *Mol Cell Endocrinol 269*, 107–111.
82. Smith TW, Lippa CF (1995). Ki-67 immunoreactivity in Alzheimer's disease and other neurodegenerative disorders. *J Neuropathol Exp Neurol 54*, 297–303.
83. Arendt T, Holzer M, Gartner U (1998). Neuronal expression of cycline dependent kinase inhibitors of the INK4 family in Alzheimer's disease. *J Neural Transm 105*, 949–960.
84. Arendt T, Holzer M, Grossmann A, Zedlick D, Bruckner MK (1995). Increased expression and subcellular translocation of the mitogen activated protein kinase kinase and mitogen-activated protein kinase in Alzheimer's disease. *Neurosci 68*, 5–18.
85. Arendt T, Rodel L, Gartner U, Holzer M (1996). Expression of the cyclin-dependent kinase inhibitor p16 in Alzheimer's disease. *Neuroreport 7*, 3047–3049.
86. Vincent I, Rosado M, Davies P (1996). Mitotic mechanisms in Alzheimer's disease? *J Cell Biol 132*, 413–425.
87. Vincent I, Jicha G, Rosado M, Dickson DW (1997). Aberrant expression of mitotic cdc2/cyclin B1 kinase in degenerating neurons of Alzheimer's disease brain. *J Neurosci 17*, 3588–3598.
88. Vincent I, Zheng JH, Dickson DW, Kress Y, Davies P (1998). Mitotic phosphoepitopes precede paired helical filaments in Alzheimer's disease. *Neurobiol Aging 19*, 287–296.
89. Nagy Z, Esiri MM, Smith AD (1997). Expression of cell division markers in the hippocampus in Alzheimer's disease and other neurodegenerative conditions. *Acta Neuropathol 93*, 294–300.
90. Nagy Z, Esiri MM, Cato AM, Smith AD (1997). Cell cycle markers in the hippocampus in Alzheimer's disease. *Acta Neuropathol 94*, 6–15.
91. McShea A, Harris PL, Webster KR, Wahl AF, Smith MA (1997). Abnormal expression of the cell cycle regulators P16 and CDK4 in Alzheimer's disease. *Am J Pathol 150*, 1933–1939.
92. McShea A, Wahl AF, Smith MA (1999). Re-entry into the cell cycle: A mechanism for neurodegeneration in Alzheimer disease. *Med Hypotheses 52*, 525–527.
93. McShea A, Zelasko DA, Gerst JL, Smith MA (1999). Signal transduction abnormalities in Alzheimer's disease: Evidence of a pathogenic stimuli. *Brain Res 815*, 237–242.
94. Raina AK, Monteiro MJ, McShea A, Smith MA (1999). The role of cell cycle-mediated events in Alzheimer's disease. *Int J Exp Pathol 80*, 71–76.
95. Raina AK, Takeda A, Smith MA (1999). Mitotic neurons: A dogma succumbs. *Exp Neurol 159*, 248–249.
96. Raina AK, Zhu X, Monteiro M, Takeda A, Smith MA (2000). Abortive oncogeny and cell cycle-mediated events in Alzheimer disease. *Prog Cell Cycle Res 4*, 235–242.
97. Raina AK, Zhu X, Rottkamp CA, Monteiro M, Takeda A, Smith MA (2000). Cyclin' toward dementia: Cell cycle abnormalities and abortive oncogenesis in Alzheimer disease. *J Neurosci Res 61*, 128–133.
98. Zhu X, Raina AK, Smith MA (1999). Cell cycle events in neurons. Proliferation or death? *Am J Pathol 155*, 327–329.
99. Zhu X, Raina AK, Boux H, Simmons ZL, Takeda A, Smith MA (2000). Activation of oncogenic pathways in degenerating neurons in Alzheimer disease. *Int J Develop Neurosci 18*, 433–437.
100. Zhu X, Rottkamp CA, Boux H, Takeda A, Perry G, Smith MA (2000). Activation of p38 kinase links tau phosphorylation, oxidative stress, and cell cycle-related events in Alzheimer disease. *J Neuropathol Exp Neurol 59*, 880–888.
101. Zhu X, Rottkamp CA, Raina AK, Brewer GJ, Ghanbari HA, Boux H, Smith MA (2000). Neuronal CDK7 in hippocampus is related to aging and Alzheimer disease. *Neurobiol Aging 21*, 807–813.
102. Yang Y, Geldmacher DS, Herrup K (2001). DNA replication precedes neuronal cell death in Alzheimer's disease. *J Neurosci 21*, 2661–2668.
103. Woods J, Snape M, Smith MA (2007). The cell cycle hypothesis of Alzheimer's disease: Suggestions for drug development. *Biochim Biophy Acta 1772*, 503–508.
104. Webber KM, Raina AK, Marlatt MW, Zhu X, Prat MI, Morelli L, Casadesus G, Perry G, Smith MA (2005). The cell cycle in Alzheimer disease: A unique target for neuropharmacology. *Mech Ageing Dev 126*, 1019–1025.
105. Grundke-Iqbal I, Iqbal K, Quinlan M, Tung YC, Zaidi MS, Wisniewski HM (1986). Microtubule-associated protein tau. A component of Alzheimer paired helical filaments. *J Biol Chem 261*, 6084–6089.
106. Kopke E, Tung YC, Shaikh S, Alonso AC, Iqbal K, Grundke-Iqbal I (1993). Microtubule-associated protein tau. Abnormal phosphorylation of a non-paired helical filament pool in Alzheimer disease. *J Biol Chem 268*, 24374–24384.
107. Alonso AC, Grundke-Iqbal I, Iqbal K (1996). Alzheimer's disease hyperphosphorylated tau sequesters normal tau into tangles of filaments and disassembles microtubules. *Nat Med 2*, 783–787.
108. Alonso AD, Grundke-Iqbal I, Barra HS, Iqbal K (1997). Abnormal phosphorylation of tau and the mechanism of Alzheimer neurofibrillary degeneration: Sequestration of microtubule-associated proteins 1 and 2 and the disassembly of microtubules by the abnormal tau. *Proc Nat Acad Sci U S A 94*, 298–303.
109. Iqbal K, Alonso AD, Gondal JA, Gong CX, Haque N, Khatoon S, Sengupta A, Wang JZ, Grundke-Iqbal I (2000). Mechanism of neurofibrillary degeneration and

pharmacologic therapeutic approach. *J Neural Transm Suppl 59*, 213–222.

110. Gong CX, Shaikh S, Wang JZ, Zaidi T, Grundke-Iqbal I, Iqbal K (1995). Phosphatase activity toward abnormally phosphorylated tau: Decrease in Alzheimer disease brain. *J Neurochem 65*, 732–738.
111. Wang JZ, Grundke-Iqbal I, Iqbal K (1996). Restoration of biological activity of Alzheimer abnormally phosphorylated tau by dephosphorylation with protein phosphatase-2A, -2B and -1. *Brain Res Mol Brain Res 38*, 200–208.
112. Oliver CJ, Shenolikar S (1998). Physiologic importance of protein phosphatase inhibitors. *Front Biosci 3*, D961–972.
113. Gong CX, Singh TJ, Grundke-Iqbal I, Iqbal K (1993). Phosphoprotein phosphatase activities in Alzheimer disease brain. *J Neurochem 61*, 921–927.
114. Iqbal K, Alonso Adel C, El-Akkad E, Gong CX, Haque N, Khatoon S, Pei JJ, Tanimukai H, Tsujio I, Wang JZ, Grundke-Iqba I (2003). Alzheimer neurofibrillary degeneration: Therapeutic targets and high-throughput assays. *J Mol Neurosci 20*, 425–429.
115. Terry RD, Masliah E, Salmon DP, Butters N, DeTeresa R, Hill R, Hansen LA, Katzman R (1991). Physical basis of cognitive alterations in Alzheimer's disease: Synapse loss is the major correlate of cognitive impairment. *Ann Neurol 30*, 572–580.
116. Wenk GL, Danysz W, Roice DD (1996). The effects of mitochondrial failure upon cholinergic toxicity in the nucleus basalis. *Neuroreport 7*, 1453–1456.
117. Zajaczkowski W, Quack G, Danysz W (1996). Infusion of (+) -MK-801 and memantine—contrasting effects on radial maze learning in rats with entorhinal cortex lesion. *Eur J Pharmacol 296*, 239–246.
118. Zajaczkowski W, Frankiewicz T, Parsons CG, Danysz W (1997). Uncompetitive NMDA receptor antagonists attenuate NMDA-induced impairment of passive avoidance learning and LTP. *Neuropharmacology 36*, 961–971.
119. Braak H, Braak E (1995). Staging of Alzheimer's disease-related neurofibrillary changes. *Neurobiol Aging 16*, 271–278; discussion 278–284.
120. Bussiere T, Giannakopoulos P, Bouras C, Perl DP, Morrison JH, Hof PR (2003). Progressive degeneration of nonphosphorylated neurofilament protein-enriched pyramidal neurons predicts cognitive impairment in Alzheimer's disease: Stereologic analysis of prefrontal cortex area 9. *J Comp Neurol 463*, 281–302.
121. Bouvier M, Szatkowski M, Amato A, Attwell D (1992). The glial cell glutamate uptake carrier countertransports pH-changing anions. *Nature 360*, 471–474.
122. Clements JD, Lester RA, Tong G, Jahr CE, Westbrook GL (1992). The time course of glutamate in the synaptic cleft. *Science 258*, 1498–1501.
123. Danbolt NC (2001). Glutamate uptake. *Prog Neurobiol 65*, 1–105.
124. Koh JY, Choi DW (1991). Selective blockade of non-NMDA receptors does not block rapidly triggered glutamate-induced neuronal death. *Brain Res 548*, 318–321.
125. Castegna A, Aksenov M, Aksenova M, Thongboonkerd V, Klein JB, Pierce WM, Booze R, Markesbery WR, Butterfield DA (2002). Proteomic identification of oxidatively modified proteins in Alzheimer's disease brain. Part I: Creatine kinase BB, glutamine synthase, and ubiquitin carboxy-terminal hydrolase L-1. *Free Radical Biol and Med 33*, 562–571.
126. Hensley K, Hall N, Subramaniam R, Cole P, Harris M, Aksenov M, Aksenova M, Gabbita SP, Wu JF, Carney JM, Lovell M, Markesbery WR, Butterfield DA. (1995). Brain regional correspondence between Alzheimer's disease histopathology and biomarkers of protein oxidation. *J Neurochem 65*, 2146–2156.
127. Kornhuber J, Bormann J, Retz W, Hubers M, Riederer P (1989). Memantine displaces [3H]MK-801 at therapeutic concentrations in postmortem human frontal cortex. *Eur J Pharmacol 166*, 589–590.
128. Parsons CG, Gruner R, Rozental J, Millar J, Lodge D (1993). Patch clamp studies on the kinetics and selectivity of N-methyl-D-aspartate receptor antagonism by memantine (1-amino-3,5-dimethyladamantan). *Neuropharmacology 32*, 1337–1350.
129. Misztal M, Frankiewicz T, Parsons CG, Danysz W (1996). Learning deficits induced by chronic intraventricular infusion of quinolinic acid—protection by MK-801 and memantine. *Eur J Pharmacol 296*, 1–8.
130. Ditzler K (1991). Efficacy and tolerability of memantine in patients with dementia syndrome. A double-blind, placebo controlled trial. *Arzneimittel-Forschung 41*, 773–780.
131. Gortelmeyer R, Erbler H (1992). Memantine in the treatment of mild to moderate dementia syndrome. A double-blind placebo-controlled study. *Arzneimittel-Forschung 42*, 904–913.
132. Winblad B, Poritis N (1999). Memantine in severe dementia: Results of the 9M-Best Study (Benefit and efficacy in severely demented patients during treatment with memantine). *Int J Geriatr Psychiatry 14*, 135–146.
133. Kasa P, Rakonczay Z, Gulya K (1997). The cholinergic system in Alzheimer's disease. *Prog Neurobiol 52*, 511–535.
134. Sims NR, Bowen DM, Allen SJ, Smith CC, Neary D, Thomas DJ, Davison AN (1983). Presynaptic cholinergic dysfunction in patients with dementia. *J Neurochem 40*, 503–509.
135. DeKosky ST, Harbaugh RE, Schmitt FA, Bakay RA, Chui HC, Knopman DS, Reeder TM, Shetter AG, Senter HJ, Markesbery WR (1992). Cortical biopsy in Alzheimer's disease: Diagnostic accuracy and neurochemical, neuropathological, and cognitive correlations. Intraventricular Bethanecol Study Group. *Ann Neurol 32*, 625–632.
136. Whitehouse PJ (1993). Cholinergic therapy in dementia. *Acta Neurol Scand Suppl 149*, 42–45.

137. Giacobini E (2000). Cholinesterase inhibitors stabilize Alzheimer's disease. *Ann N Y Acad Sci 920*, 321–327.
138. Giacobini E (2001). Do cholinesterase inhibitors have disease-modifying effects in Alzheimer's disease? *CNS Drugs 15*, 85–91.
139. Giacobini E (2002). Long-term stabilizing effect of cholinesterase inhibitors in the therapy of Alzheimer' disease. *J Neural Transm Suppl* 181–187.
140. Racchi M, Mazzucchelli M, Porrello E, Lanni C, Govoni S (2004). Acetylcholinesterase inhibitors: Novel activities of old molecules. *Pharmacol Res 50*, 441–451.
141. Meyer JS, Li Y, Xu G, Thornby J, Chowdhury M, Quach M (2002). Feasibility of treating mild cognitive impairment with cholinesterase inhibitors. *Int J Geriatr Psychiatry 17*, 586–588.
142. Giacobini E (2004). Cholinesterase inhibitors: New roles and therapeutic alternatives. *Pharmacol Res 50*, 433–440.
143. Mori F, Lai CC, Fusi F, Giacobini E (1995). Cholinesterase inhibitors increase secretion of APPs in rat brain cortex. *Neuroreport 6*, 633–636.
144. Giacobini E, Mori F, Lai CC (1996). The effect of cholinesterase inhibitors on the secretion of APPS from rat brain cortex. *Ann N Y Acad Sci 777*, 393–398.
145. Tuszynski MH (2002). Growth-factor gene therapy for neurodegenerative disorders. *Lancet Neurol 1*, 51–57.
146. Savage MJ, Lin YG, Ciallella JR, Flood DG, Scott RW (2002). Activation of c-Jun N-terminal kinase and p38 in an Alzheimer's disease model is associated with amyloid deposition. *J Neurosci 22*, 3376–3385.
147. Rabizadeh S, Bredesen DE (2003). Ten years on: Mediation of cell death by the common neurotrophin receptor p75(NTR). *Cytokine Growth Factor Rev 14*, 225–239.
148. Yaar M, Zhai S, Fine RE, Eisenhauer PB, Arble BL, Stewart KB, Gilchrest BA (2002). Amyloid beta binds trimers as well as monomers of the 75-kDa neurotrophin receptor and activates receptor signaling. *J Biol Chem 277*, 7720–7725.
149. Lee R, Kermani P, Teng KK, Hempstead BL (2001). Regulation of cell survival by secreted proneurotrophins. *Science 294*, 1945–1948.
150. Fahnestock M, Michalski B, Xu B, Coughlin MD (2001). The precursor pro-nerve growth factor is the predominant form of nerve growth factor in brain and is increased in Alzheimer's disease. *Mol Cell Neurosci 18*, 210–220.
151. Poduslo JF, Curran GL (1996). Permeability at the blood-brain and blood-nerve barriers of the neurotrophic factors: NGF, CNTF, NT-3, BDNF. *Brain Res Mol Brain Res 36*, 280–286.
152. Hempstead BL (2002). The many faces of p75NTR. *Current Opin Neurobiol 12*, 260–267.
153. Longo FM, Manthorpe M, Xie YM, Varon S (1997). Synthetic NGF peptide derivatives prevent neuronal death via a p75 receptor-dependent mechanism. *J Neurosci Res 48*, 1–17.
154. Xie Y, Tisi MA, Yeo TT, Longo FM (2000). Nerve growth factor (NGF) loop 4 dimeric mimetics activate ERK and AKT and promote NGF-like neurotrophic effects. *J Biol Chem 275*, 29868–29874.
155. Maliartchouk S, Debeir T, Beglova N, Cuello AC, Gehring K, Saragovi HU (2000). Genuine monovalent ligands of TrkA nerve growth factor receptors reveal a novel pharmacological mechanism of action. *J Biol Chem 275*, 9946–9956.
156. DeFreitas MF, McQuillen PS, Shatz CJ (2001). A novel p75NTR signaling pathway promotes survival, not death, of immunopurified neocortical subplate neurons. *J Neurosci 21*, 5121–5129.
157. Khursigara G, Bertin J, Yano H, Moffett H, DiStefano PS, Chao MV (2001). A prosurvival function for the p75 receptor death domain mediated via the caspase recruitment domain receptor-interacting protein 2. *J Neurosci 21*, 5854–5863.
158. Mamidipudi V, Wooten MW (2002). Dual role for p75(NTR) signaling in survival and cell death: Can intracellular mediators provide an explanation? *J Neurosci Res 68*, 373–384.
159. Dobrowsky RT, Carter BD (2000). p75 neurotrophin receptor signaling: Mechanisms for neurotrophic modulation of cell stress? *J Neurosci Res 61*, 237–243.

# NEUROTROPHIC FACTORS AS NOVEL THERAPEUTIC TARGETS

Stephen D. Skaper

*Neurosciences Centre of Excellence for Drug Discovery, GlaxoSmithKline Research and Development Limited, Essex, United Kingdom*

## 1 INTRODUCTION

In mammals and other vertebrates, soluble peptide growth factors play essential roles in intercellular communication. They exert their effects by signaling through surface membrane receptors that interact with diverse types of intracellular second-messenger systems. In a sometimes surprising manner, many growth factors have been found to subserve a wide variety of functions by acting on many cell types at different stages of development or in adult life. One of the major advances of cellular neuroscience has been to recognize that much of the cellular damage resulting from such central nervous system (CNS) insults as stroke, trauma, and neurodegenerative disease may be caused by a limited number of endogenously generated molecules with neurotoxic activities. Less well developed is the idea that endogenous mechanisms exist to provide neuroprotection and that endogenous molecules may be produced specifically to service neuroprotective signaling functions.

Neurotrophic factors are secreted proteins that promote neurite outgrowth, neuronal cell differentiation, and survival both in vivo and in vitro. Since their discovery in the 1950s, neurotrophic factors have raised expectations that their clinical application to neurodegenerative diseases might provide an effective therapy for what are now untreatable conditions. There is a large volume of evidence for neuroprotective effects of neurotrophic factors in animal models of neurodegerative diseases. These applications include abolishing cell loss in the septum following fimbria-fornix transection, a cholinergic model for Alzheimer's disease [1, 2], recovery from chemically induced parkinsonism [3, 4], motor neuron disease [5, 6], and spinal cord injury [7–9]. While showing great promise for the treatment of such dirorders, the poor pharmacokinetics and bioavailability of these protein factors have limited the use of polypeptides as drugs. The development of nonpeptidyl small-molecule neurotrophic factor mimetics would obviate the need for protein factors and the potential problems associated with them. This chapter will discuss various strategies whereby a small molecule can influence neurotrophic signaling.

## 2 NEUROTROPHINS AND THEIR RECEPTORS

Neurotrophins were identified as promoters of neuronal survival, but it is appreciated that they regulate many aspects of neuronal development and function, including synapse formation and synaptic plasticity [10–13]. The first neurotrophin, nerve growth factor (NGF), was discovered during a search for survival factors that could explain the deleterious effects of deletion of target tissues on the subsequent survival of motor and sensory neurons [14]. NGF is part of the neurotrophin

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

family of approximately 25-kDa polypeptides that function as homodimers, and that share a high degree of structural homology and includes brain-derived neurotrophic factor (BDNF), neurotrophin-3 (NT-3), and neurotrophin-4 (NT-4) [15]. Neurotrophins are found in both mammals and lower vertebrates, while the neurotrophin homologs NT-6 and NT-7 were cloned in fish [16, 17].

The neurotrophins interact with two entirely distinct classes of receptors, Trks (tropomyosin receptor kinases) and $p75^{NTR}$ [neurotrophin receptor (NTR)], the first discovered member of the tumor necrosis factor receptor superfamily. The former was initially identified as a low-affinity receptor for NGF but was subsequently shown to bind each of the neurotrophins with approximately equal nanomolar affinity [18, 19] but lacks a catalytic motif. $p75^{NTR}$ binds NGF along the interface between two NGF monomers, and binding results in a conformational change in NGF that alters the monomeric interface on the opposite side of the NGF dimer, eliminating the potential for binding of one NGF dimer to two $p75^{NTR}$ monomers. In contrast to interactions with $p75^{NTR}$, the neurotrophins dimerise the Trk receptors, resulting in activation through transphosphorylation of the kinases present in their cytoplasmic domains. The four neurotrophins exhibit specificity in their interactions with the three members of this receptor family with NGF activating TrkA, BDNF and NT-4 activating TrkB, and NT-3 activating TrkC. In addition, NT-3 can activate the other Trk receptors with less efficiency (Fig. 1). In vitro studies indicate that the presence of $p75^{NTR}$ potentiates activation of TrkA by suboptimal concentrations of NGF, although it does not appear to potentiate the activation of other Trk receptors similarly by their ligands [20, 21], and that $p75^{NTR}$ collaborates with TrkA to form high-affinity ($10^{-11}$ *M*) binding sites for NGF [22]. The major site at which neurotrophins interact with the Trk receptors is in the membrane-proximal immunoglobulin (Ig)-like domain. The three-dimensional structures of the domain in each of the Trk receptors have been solved [23]. In addition, the structure of NGF bound to the TrkA Ig domain has also been determined [24]. These results have made it possible to identify residues in the neurotrophins and the Trk receptors that account for the specificity observed in their interactions [25, 26]. Such information may prove useful in the design of small-molecule mimetic ligands for neurotrophin receptors, as will be discussed in a later section.

Through Trk receptors, neurotrophins activate Ras, phosphatidyl inositol-3 kinase, phospholipase C-γ1, and signaling pathways controlled through these proteins, such as mitogen-activated protein kinases. Activation of $p75^{NTR}$ results in activation of the nuclear factor-κB and Jun kinase as well as other signaling pathways. Several excellent reviews describe recent progress in understanding the signaling pathways stimulated by the neurotrophins that affect the survival, differentiation, and function of cells within the nervous system [10–13].

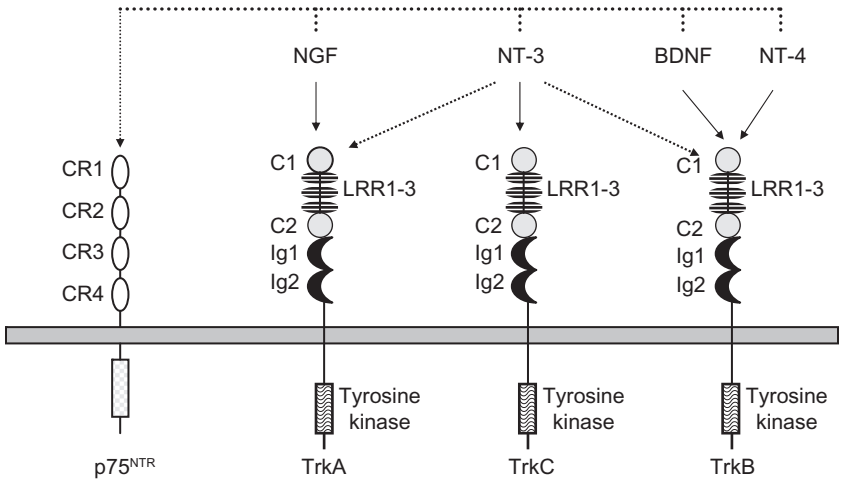


**Figure 1** Neurotrophins and their receptors. The neurotrophins display specific interactions with the three Trk receptors: NGF binds TrkA, BDNF and NT-4 bind TrkB, and NT-3 binds TrkC. In some cellular contexts, NT-3 can also activate TrkA and TrkB albeit with less efficiency. All neurotrophins bind to and activate $p75^{NTR}$. CR1–CR4, cysteine-rich motifs; C1/C2, cysteine-rich clusters; LRR1–3, leucine-rich repeats; Ig1/Ig2, immunoglobulin-like domains.

## 3 NEUROTROPHIC PROTEINS AS THERAPEUTIC AGENTS

The power of neurotrophic factors to regulate neuronal cell survival in the developing nervous system and to promote survival after injury or protect neurons in toxin-mediated disease models in animals has spurred the idea that such proteins could be harnessed for the treatment of neurodegenerative disease. For example, the traditional perspective of applying neurotrophins in the context of Alzheimer's disease is based on the premise that neurotrophins are capable of upregulating cholinergic function and of rendering neurons less vulnerable to certain processes causing degeneration [1, 2]. Another case is that of glial cell-line-derived neurotrophic factor (GDNF), which acts on dopaminergic neurons. Administered intracerebrally, GNDF has neuroprotective and neurorestorative effects in toxin-induced rodent and nonhuman primate models of Parkinson's disease [3, 27, 28]. In spite of such encouraging preclinical data, clinical studies have proven largely disappointing. Neurotrophic proteins have failed to show a beneficial effect in human trials of amytrophic lateral sclerosis [29–33], and in some cases the side effects were severe [34, 35]. A negative outcome was also reported for randomized, controlled trials of GDNF in Parkinson's disease patients [36, 37]. Neurotrophic factor treatment of CNS diseases presents an especially complex problem since these polypeptides have poor pharmacokinetics and bioavailability and are not able to cross the blood–brain barrier. Neurotrophic proteins can be modified to increase blood–brain barrier transport [38], and recent studies have shown that intravenous administration of BDNF, conjugated to an antitransferrin antibody, is able to traverse the blood–brain barrier in rats and provide neuroprotection in focal transient brain ischemia [39, 40]. Toxicology studies in primates showing cerebellar degeneration following the administration and subsequent withdrawal of high doses of GDNF raise concerns about dosing and targeting of delivery. Viral vector or cell-based gene therapy approaches may prove more advantageous for the targeted delivery of lower doses of neurotrophic factors. Dopaminergic neuroprotection by viral-mediated delivery or GDNF or the GDNF-related factor neurturin [41] has been demonstrated in animals [3, 42, 43]. In the case of spinal cord injury, the transplantation of genetically modified cells (e.g., viral expression of neurotrophins either in vivo or ex vivo in stem cells) to lesion sites has been investigated as a means to augment axonal regeneration [44–46]. A cell-based approach for NGF delivery has shown promise in a phase I trial in Alzheimer's disease [47].

Peripheral nervous system (PNS) disorders might be more amenable to a neurotrophic protein therapeutic. NGF, which is selectively trophic for small fiber sensory and sympathetic neurons, was selected as a potential therapy for diabetic polyneuropathy because of the serious consequences associated with degeneration of those neuronal populations in this condition. In addition, a reduced NGF availability may contribute to the pathogenesis of diabetic neuropathy [48, 49], and animal models of neuropathy respond to the exogenous administration of NGF [50]. Two sets of phase II clinical trials suggested that recombinant human NGF (rhNGF) administration was effective in ameliorating the symptoms associated with both diabetic polyneuropathy [51] and human immunodeficiency virus (HIV)-related neuropathy [52]. These early studies, however, revealed that painful side effects were dose limiting for NGF. A large-scale phase III clinical trial of 1019 patients randomized to receive either rhNGF or placebo for 48 weeks failed to confirm the earlier indications of efficacy [53]. Such side effects are not surprising, as increased NGF content in peripheral tissues can cause inappropriate axonal sprouting [54] and hyperalgesia [55].Other protein factors with peripheral nerve regeneration activity are efficacious in experimental diabetic neuropathy and include sonic hedgehog protein [56] and ciliary neurotrophic factor [57]. Whether or not the latter would produce the desired trophic effects of NGF in humans without causing hyperalgesia remains to be investigated.

## 4 FUNCTIONAL MIMETICS OF NEUROTROPHINS ACTING AT TRK RECEPTORS

Despite the therapeutic potential of neurotrophic proteins, there are inherent drawbacks with the use of polypeptides applied as drugs: in vivo instability; unsuccessful tissue targeting; pleiotrophic effects due to binding multiple receptors and co-receptors, and activation of multiple pathways that were not intentionally targeted; and relatively high production cost for medicinal applications. Such problems may be averted by developing highly selective receptor ligands, with chemical and pharmacological properties such as proteolytic stability, small size, ability to penetrate tissue barriers, and ability to agonize or antagonize the receptors. The pharmacological advantages of mimetics over protein therapeutics are well documented and include lack of immunogenicity, low molecular mass, low manufacturing costs, and good pharmacokinetic profiles [58].

The first rationally designed TrkA agonists were based on the use of ligand mimicry and antibody

mimicry strategies [58, 59] to generate peptide analogs of two agonists directed to the extracellular domain of TrkA: the peptide sequence of NGF itself [60–62] and the anti-TrkA monoclonal antibody (mAb) 5C3 [63]. This antibody most likely acts through its two variable domains to physically join together two Trk monomers. TrkA binding is mediated by a discrete β-turn region (amino acid residues 92–97) of these ligands, since crystallized peptides and peptidomimetic compounds incorporating these residues were found to activate TrkA and demonstrate neurotrophic effects in cell-based models [64–66]. Such cyclic peptide dimeric ligands may act as TrkA partial agonists, inducing receptor dimerization, phosphorylation, and and internalization [64]. Other conformers or linear peptides were inactive. In fact, most monomer cyclic peptide β-turn mimetics of NGF and BDNF are reported to have only antagonistic properties, inhibiting neurotrophin binding and activity [60, 67]. Similar findings show that a dimeric (but not monomeric) cyclized version of amino acids 92–96 of NGF also acts as a partial TrkA agonist, with neurotrophic effects being blocked by TrkA inhibitors [65].

Based on the pharmacophores of the mAb 5C3 and NGF peptide analogs described above, a focused β-turn peptidomimetic library of ~60 compounds was synthesized, leading to the identification of D3 [66], a small, selective, and proteolytically stable agonist of the TrkA receptor (Fig. 2). A similar compound, C59, lacked neurotrophic effects. These results indicate that a small β-turn peptidomimetic can activate a tyrosine kinase neurotrophin receptor that normally binds a relatively large protein ligand. Although D3 is still far short of an NGF mimetic in terms of its potency, treatment of cognitively impaired aged rats with D3 has been shown to act as a neuroprotective agent. Aged rats showed improved memory and improved cholinergic phenotype. The effect was long-lived, for 3 months after drug delivery. D3 performed equally or better than NGF [68].

Crystallographic and molecular modeling studies of NGF/TrkA complexes [24, 25, 69, 70], site-directed mutagenesis [71–75], and results from chemical modification of amino acid residues [70, 76] have identified loops 1 and 4 and the N-terminal region of NGF as the most relevant for its biological activity. Among various peptides reported in the literature, linear peptides showed inhibitory, but no agonist, activity [77], and, as mentioned above, cyclic peptides corresponding to the loop 1 sequence were shown later to have a 10-fold higher inhibitory activity [60]. The first NGF loop domain peptides described to have neurotrophic activity were the loop 1 homodimer mimetics reported by Longo et al. [78]; loop 4 mimetics with agonist activity appeared several years later [64–66]. The low efficacy of these compounds led Colangelo and colleagues [79] to use a different approach based on the combination of different NGF regions in the same molecule. Linking both the L1 and L4 loops together with an appropriate spacer, with or without the N-terminal region, produced peptides (NL1L4 and L1L4) that induced differentiation of dorsal root ganglion neurons and stimulated tyrosine phosphorylation of TrkA but not TrkB. In addition, L1L4 was able to induce differentiation of PC12 cells, with an $EC_{50}$ (1 μ*M*) well below the range of concentrations reported for the more active homopeptides [65, 78]. More interestingly, the L1L4 peptide reduced neuropathic behavior and restored neuronal function in a rat model of peripheral neuropathic pain, thereby suggesting a potential therapeutic role for this NGF-mimetic peptide [79].

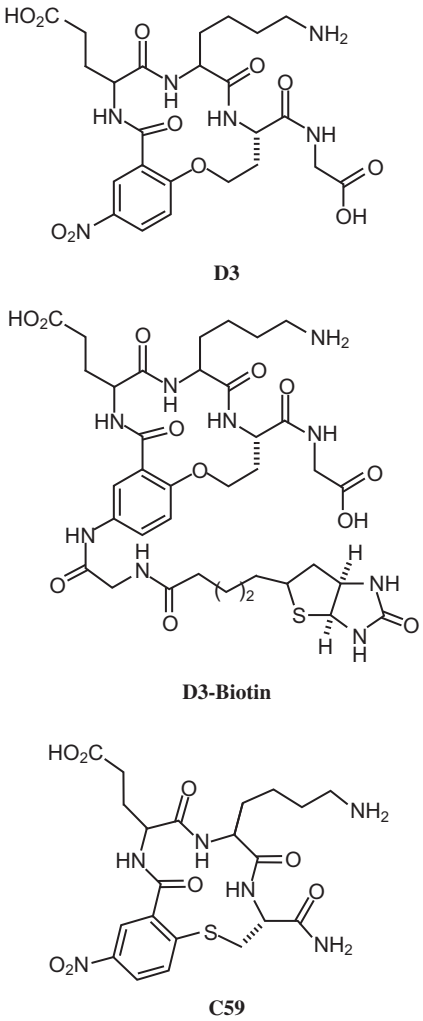


**Figure 2** Structures of D3, D3-biotin, and C59. (Reproduced from ref. 66, Fig. 1. Copyright 2000, with permission from the American Society for Pharmacology and Experimental Therapeutics.)

Small, dimeric peptides designed to mimic a pair of solvent-exposed loops important for the binding and activation of the BDNF receptor, TrkB, have been described [80]. The monomer components that make the dimers were based on a monocyclic monomeric peptide mimic of a single loop of BDNF (loop 2) shown previously to be an inhibitor of BDNF-mediated neuronal cell survival [81]. These bicyclic dimeric peptides behaved as partial agonists with respect to BDNF, promoting sensory neuron survival in vitro. A highly conformationally constrained tricyclic dimeric peptide was then synthesized in which the monomeric subunits were linked by two dimerizing linkages. Although the tricyclic dimer was still a partial BDNF agonist, it was 10-fold more potent in its neuronal survival effect than the best of the bicyclic dimers, making it ~2-fold less potent than BDNF itself. Partial agonistic behavior could result from the compound being less efficient than BDNF in bringing about dimerization of BDNF or because of inappropriate conformational flexibility of the dimerizing linkages (although the retention of partial agonism by the highly constrained tricyclic compound would argue against this). A consequence of reduced dimerization efficiency could be a reduced ability to cause the autophosphorylation of the five tyrosine residues in TrkB (resulting in reduced activation of downstream signaling pathways). An alternative mechanism for the partial agonist behavior of the dimeric peptides is that they act through TrkB only, even as full TrkB agonists, and not $p75^{NTR}$. A series of dimeric β-turn mimetics that act as agonists of TrkC have recently been described [82] (Fig. 3), as well as a tandem repeat cyclic peptide agonist of TrkB [83].

Additional examples of dimeric peptides able to mimic the effects of larger protein agonists, presumably by being able to bring about receptor dimerization include agonists of the cell adhesion molecule N-cadherin. In the nervous system, N-cadherin function has been implicated in cell migration, axonal growth and guidance, and synapse formation and synaptic plasticity [84]. Peptide mimetics of two linear sequences from extracellular domain 1 (HAVDI and INPISGQ) of N-cadherin function as highly specific antagonists [85], while cyclic peptides containing a tandem repeat of the individual motifs function as N-cadherin agonists that stimulate neurite outgrowth in a similar manner to native N-cadherin [86]. The cyclic peptide N-Ac-*CHAVDINGHAVDIC*-$NH_2$ (SW4) containing the HAVDI sequence in tandem was efficacious also in promoting the in vitro survival of several populations of CNS neurons in paradigms where fibroblast growth factor-2 (FGF-2) is effective [87] (Fig. 4). The neuroprotective effects of SW4 displayed a concentration dependence similar to those inducing neuritogenesis and were inhibited by a monomeric version of the same motif and by a specific fibroblast growth factor (FGF) receptor antagonist. These data suggest that SW4 functions by binding to and clustering N-cadherin in neurons and thereby activating an N-cadherin/FGF signaling cascade,

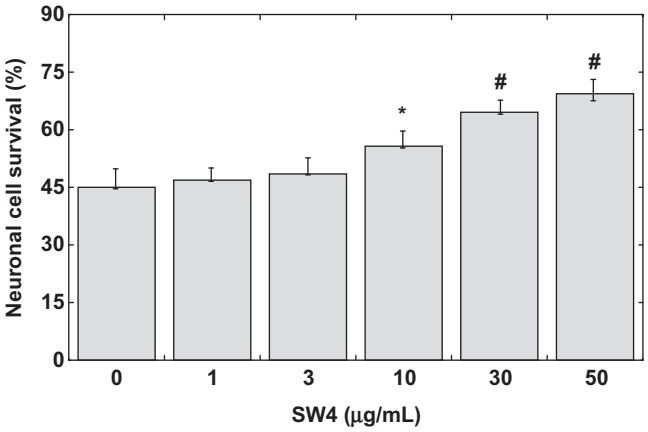


**Figure 4** The Dimeric HAVDI peptide SW4 attenuates glutamate-induced toxicity in culture hippocampal neurons. Cultures (days in vitro 7–9) were pretreated with the indicated concentrations of SW4 peptide for 48 h and then exposed to 50 μ*M* glutamate, and neuronal survival was quantified 24 h later by a colourimetric reaction (MTT). Data are means ± SD (three experiments) expressed relative to no glutamate (100%). $^{*}P < 0.05$ or $^{\#}P < 0.01$ vs. glutamate only. (Reprinted from ref. 87, Fig. 3. Copyright 2004, with permission from Elsevier.)

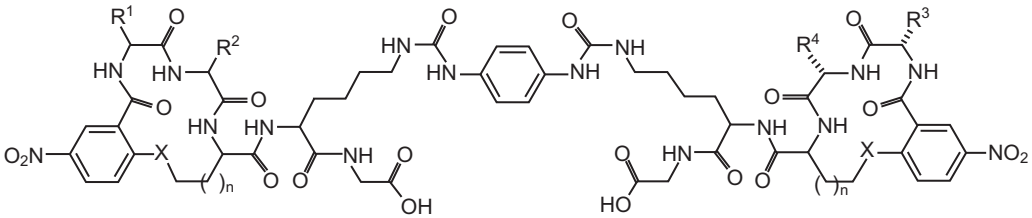

**Figure 3** TrkC agonist structure. (Reprinted from ref. 82, Scheme 1, structure 1. Copyright 2001, with permission from Elsevier.)

consistent with the ability of cell adhesion molecules to activate and FGF receptor-signaling cascade in growth cones [88].

## 5 TARGETING THE P75$^{NTR}$ AS A STRATEGY TO PROMOTE NEURON SURVIVAL

Specific effects of neurotrophins depend on whether signaling is activated via a member of the Trk receptor tyrosine kinase family or the p75 neurotrophin receptor p75$^{NTR}$. In contrast to the Trk receptors, the function of p75$^{NTR}$ remains enigmatic, although it participates in multiple cellular activities apparently by functioning as a co-receptor with different partners [11, 89]. In the absence of a co-expressed Trk receptor, p75$^{NTR}$ can mediate death in neurons [90, 91] and oligodendrocytes [92]. p75$^{NTR}$ can be activated selectively by proneurotrophins [93], which are potent inducers of p75$^{NTR}$-mediated cell death [94, 95] attributable to concomitant interaction with the sorting receptor sortilin [96]. Pro-NGF induces death of oligodendrocytes after spinal cord injury [94] and cortical neurons after injury [95]. Pro-NGF is also abundant in the Alzheimer brain [97], raising the possibility that this form of NGF may contribute to death of septal cholinergic neurons via activation of p75$^{NTR}$-mediated apoptosis. Indeed, a recent study has shown that pro-NGF produced in the brain under pathological conditions can elicit apoptosis in basal forebrain neurons in vitro [98].

Studies showing that neurotrophin binding to p75$^{NTR}$ can promote cell survival in the absence of Trk receptors [99, 100], together with recent structural data indicating that NGF may bind to p75$^{NTR}$ in a monovalent manner [101], raise the possibility that small molecules mimicking a single p75$^{NTR}$-interacting domain of NGF might be capable of competing with NGF or pro-NGF binding and/or modifying p75$^{NTR}$ signaling and positively regulate cell survival. A pharmacophore designed to capture selected structural and physicochemical features of a neurotrophin domain (β-hairpin loop 1) known to interact with p75$^{NTR}$ was applied to in silico screening of small-molecule libraries. Four of 23 compounds selected showed neurotrophic activity for cultured sensory neurons (Fig. 5) [102] and promoted survival of CNS neurons while inhibiting pro-NGF-mediated death of mature oligodendrocytes [103]. Compound LM11A-24 prevented p75$^{NTR}$-dependent motor neuron death induced by either exogenous addition of NGF or by NGF-producing reactive astrocytes in co-culture conditions [104]. Increased levels of NGF and p75$^{NTR}$ re-expression have been implicated in motor neuron death following axotomy [105] and in motor neuron disease models [106, 107]. Targeted modulation of

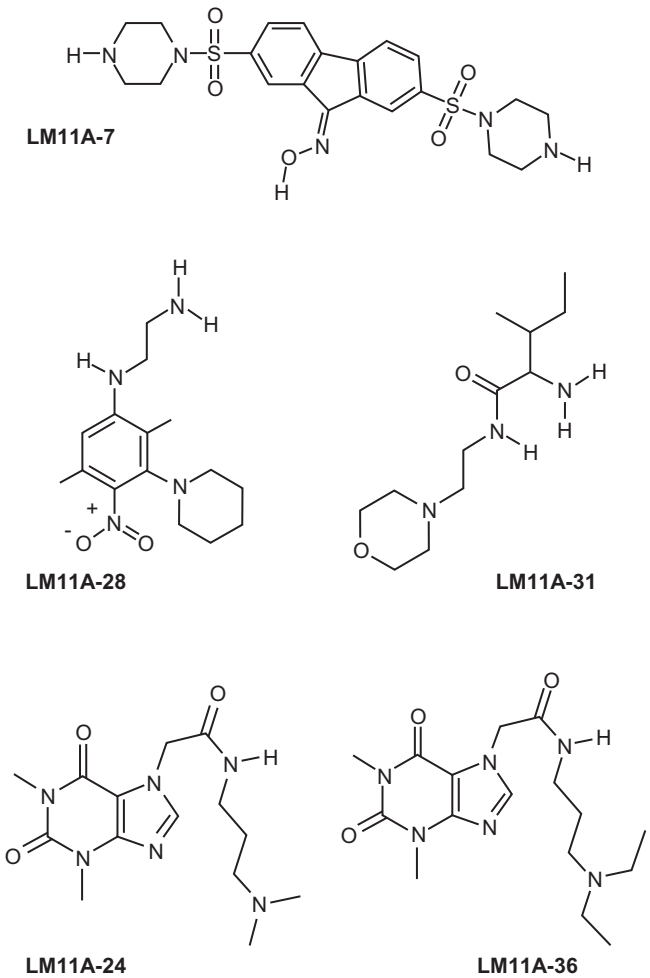


**Figure 5** LM11A compound structures. (Reproduced from ref. 103, Fig. 2. Copyright 2006, with permission from The Society for Neuroscience.)

p75$^{NTR}$ signaling may constitute a strategy for therapeutic development.

Alzheimer's disease is characterized by the accumulation of neurotoxic amyloidogenic peptide Aβ and progressive impairment of cognitive function, particularly memory [108]. As mentioned above, it has been proposed that increased levels of proneurotrophins, such as occur in Alzheimer's disease, may underlie degeneration of the cholinergic innervation to the hippocampus (the septohippocampal pathway) via p75$^{NTR}$ activation. However, it is also possible that Aβ could be directly responsible for activating p75$^{NTR}$-mediated cell death in Alzheimer's disease, given that it is a ligand for the receptor [109] and can stimulate p75$^{NTR}$-mediated cell death signaling cascades in cell lines [110, 111]. Like the amyloid precursor protein, the extracellular domain of p75$^{NTR}$ is cleaved by a metalloprotease, generating a transmembrane-linked C-terminal fragment. This is

then cleaved by a γ-secretase, liberating a soluble intracellular domain [112]. Ligand binding can initiate this process, and both fragments of $p75^{NTR}$ have been shown to mediate death signaling [113–116]. A new study has now provided evidence of a direct link between $p75^{NTR}$ signaling and Aβ-induced toxicity in hippocampal neurons in vitro and in cholinergic basal forebrain neurons in vivo [117]. Neuronal death correlated with $A\beta_{1-42}$ peptide-stimulated accumulation of the constitutively active death-signaling $p75^{NTR}$ C-terminal fragment generated by inhibition of γ-secretase activity [117]. These results provide a novel mechanism to explain the early and characteristic degeneration of the septohippocampal pathway that occurs in Alzheimer's disease.

Agonist monoclonal antibodies [118] or peptide mimetics [83], which activate TrkB without binding to $p75^{NTR}$ at all, represent a potential alternative strategy to the deleterious consequences of $p75^{NTR}$ activation. These molecules demonstrate neurotrophic activity [83, 118] and promote neurite growth in the presence of a myelin inhibitor [118], for which $p75^{NTR}$ serves as the signaling component [119]. Thus, agents that activate Trk receptors in the absence of $p75^{NTR}$ binding, or that inhibit neurotrophin/$p75^{NTR}$ binding, might be more effective than neurotrophins at promoting repair.

## 6 SMALL-MOLECULE NONPEPTIDE NEUROTROPHIN RECEPTOR AGONISTS

Asterriquinones are naturally occurring bis-indolyl-dihydroxyquinones that activate tyrosine kinase receptors. They were originally identified as activators of the insulin receptor [120, 121]. The molecules are small and readily cell permeable, and they act directly on the receptor tyrosine kinase domain, although the mechanism is not known [120]. Utilizing a natural product screen, a small-molecule nonpeptidyl fungal metabolite isolated from *Pseudomassaria*, L-783,281 was initially identified as an activator of the insulin receptor [120] (Fig. 6). The original compound demethylasterriquinone-B1 (DAQ-B1, or L-783,281) crossed the blood–brain barrier and, in a separate study, was shown to be efficacious at low micromolar concentrations in stimulating TrkA, B, and C phosphorylation to various extents in CHO cell lines stably expressing human Trk receptors, although it was toxic to neurons [122]. L-783,281 also stimulated Trk phosphorylation in a number of rat and human primary neuronal cell cultures, whereas the highly similar compound, L-767,827 was without effect [122]. Mechanistic studies suggested that L-783,281 probably causes receptor dimerization and transphosphorylation by binding in a noncovalent

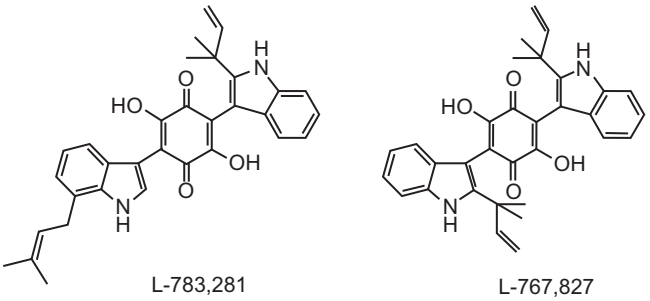


**Figure 6** The structures of L-783,281 and L-767,827. (Reproduced from ref. 122, Fig. 1. Copyright 2001, with permission of Blackwell Publishing.)

manner. Unfortunately, cytotoxicity of L-783,281 precluded evaluating effects on neuronal cell survival and neurite outgrowth. In a more recent study, combinatorial libraries of asterriquinones (>300) and mono-indolyl-quinones (>60) were screened using a 96-well enzyme-linked immunosorbent assay that detects phosphorylated TrkA [123]. The libraries were also screened for cytotoxicity. From these screens, the authors generated quantitative structure–activity relationship models for activity and toxicity, culminating in the selection of two compounds, 2-(6-chloro-1*H*-indol-3-yl)-5-(2-cyclopropyl-1*H*-indol-3-yl)-3,6-dihydroxy-[1,4]-benzoquinone (1H5) and 2,5-dimethoxy-3-(7-fluoro-1*H*-indol-3-yl)-[1,4]-benzoquinone (5E5), for further study based on high activity and low cytotoxicity. Compound 1H5 (30 μ*M*) is an asterriquinone that is a moderate TrkA activator (50% of the activity of 100 ng/mL NGF), and it shows little toxicity at concentrations up to 100 μ*M*. 1H5 can protect terminally differentiated PC12 cells from apoptotic cell death induced by NGF withdrawal [123]. Compound 5E5 (30 μ*M*) is a mono-indolyl-quinone that is a very strong activator of TrkA (>200% the activity of 100 ng/mL NGF), and it is non-toxic up to 10 μ*M*. Activation of TrkA could be detected at 1 μ*M* 5E5, and 3–10 μ*M* 5E5-activated TrkA and extracellular signal-regulated kinase as strongly as a maximal concentration of NGF (100 ng/mL) [123]. Moreover, a low concentration of 5E5 (1 μ*M*) combined with a submaximal concentration of NGF (10 ng/mL) promoted neuronal-like differentiation of PC12 cells. As the mitogen-activated protein kinase (MAPK) pathway is essential for NGF-induced neuritogenesis [124, 125], this may explain why 5E5 was better at stimulating differentiation than promoting survival, because 5E5 is a potent activator of extracellular signal-regulated kinase but a weaker activator of Akt (protein kinase B). The phosphatidylinositol 3-kinase/Akt signaling pathway contributes more than the MAPK pathway in neuronal survival [126, 127]. Although 1H5 and 5E5

themselves are unsuitable as drug candidates, they provide lead compounds that can be applied to preclinical models of neurodegeneration for proof-of-principle experiments.

Another approach to target Trk receptor activation may be that of identifying agents that increase the activation of Trk receptors in the co-presence of (sub) threshold concentrations of neurotrophin. Historical evidence suggests that the monosialic acid ganglioside GM1, like the neurotrophins, has the potential to protect injured and aged CNS neurons, enhance neuronal phenotypic expression, induce neuronal sprouting, and enhance neuronal function and metabolism [128]. The pleiotropic neurotrophic activity of GM1 extends to a wide range of neuronal cell populations, suggesting an interaction with a wide spectrum of neurotrophic factors or common signaling pathways. Indeed, GM1 was shown to enhance NGF-induced TrkA activation and dimerization [126–132] and to potentiate the neurotrophic effect of NGF in vivo [133, 134]. GM1 added to brain slices activated extracellular-regulated protein kinases (Erks) in a Trk-sensitive manner, while intracerebroventricular administration of GM1 induced a transient phosphorylation of TrkA and Erk1/2 in hippocampus and striatum [135]. Both GM1 and its semisynthetic derivative LIGA20 activated TrkB and prevented glutamate-mediated apoptosis of cerebellar granule neurons, similar to BDNF [136]. The mechanism underlying the ganglioside actions have not been elucidated, and such compounds are not very brain penetrant.

K-252a and K-252b belong to a class of related indolocarbazoles of microbial origin that inhibit a range of kinases, including Trk [137, 138]. At concentrations below those that inhibit Trk activation, K-252b potentiates the stimulatory effect of NT-3 on TrkA phosphorylation and promotes peripheral sensory neuron survival and cholinergic neuron differentiation [138]. However, these compounds are not pharmacologically clean (e.g., non-Trk-related effects). Examination of a series of staurosporines led to the identification of a compound, L-753,000 (NB-506), that potentiated the action of NT-3 on TrkA (but not other Trks) and produced a range of neurotrophic effects at nanomolar concentrations without the kinase inhibitory response of K-252b [139] (Fig. 7). A compound structure similarity search of chemical libraries conducted based on the structures of L-753,000 and K-252b yielded several molecules that potentiated NT-3-mediated TrkB activation, exhibited no activity in the absence of neurotrophin, and did not potentiate receptor activation by NT-4 or BDNF. One such compound, BMS355249, was maximally effective at 100 n*M* in enhancing the effect of NT-3 at TrkA and TrkB and in potentiating NT-3-mediated cell survival [140]. Experiments using chimeric Trk receptors demonstrated that the extracellular domain was essential for compound potentiation and ruled out interaction with intracellular signaling molecules as a mechanism of compound activity [140].

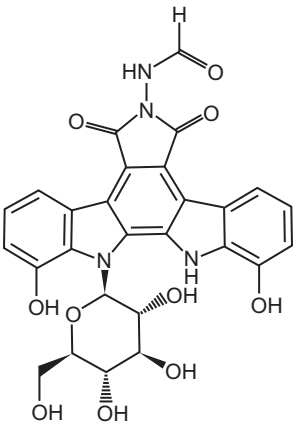


**Figure 7** The Structure of L-753,000.

## 7 BDNF–SEROTONIN INTERACTIONS

BDNF and serotonin (5-hydroxytryptamine, 5-HT) are two seemingly distinct signaling systems that play regulatory roles in many neuronal functions, including survival, neurogenesis, and synaptic plasticity. A common feature of the two systems is their ability to regulate the development and plasticity of neural circuits involved in mood disorders such as depression and anxiety [141–143]. BDNF promotes the survival and morphological differentiation of 5-HT neurons both in culture and in vivo [144, 145]. Conversely, administration of antidepressant selective serotonin reuptake inhibitors enhances BDNF gene expression [146, 147]. The effect of antidepressants appears to be a general phenomenon, as pharmacologically diverse compounds, including inhibitors of monoamine transporters and metabolism, activate TrkB rapidly in the rodent anterior cingulate cortex and hippocampus [148]. Furthermore, acute and long-term antidepressant treatments induced TrkB-mediated activation of phospholipase-Cγ1 and increased the phosphorylation of cyclic adenosine monophosphate (CAMP)-related element binding protein, a major transcription factor mediating neuronal plasticity. Modulation of the phosphorylation of TrkB Shc binding site, phosphorylation of mitogen-activated protein kinase, or Akt by antidepressants was not observed [148]. Brain monoamines are crucial mediators of antidepressant-induced TrkB activation, as antidepressants did not produce TrkB activation in the brains of

serotonin- or norepinephrine-depleted mice. Rapid activation of the TrkB neurotrophin receptor and phospholipase-Cγ1 signaling thus seems to be a common mechanism for all antidepressant drugs and suggests that such molecules may serve as chemical templates for the design of TrkB modulators.

## 8 NEUROIMMUNOPHILINS

Immunophilins are a class of proteins best known for their key role in mediating the immunosuppressive actions of FK506 and cyclosporin A [149]. FK506 and FK506-derived inhibitors of the FK506-binding protein (FKBP)-type peptidylprolyl *cis/trans*-isomerases (PPIase) binds to an intracellular protein designated FKBP12. This FK506–FKBP12 complex binds to and inhibits the phosphatase calcineurin [150]. FKBP12 is highly expressed in the brain, and FK506 was found to promote neurite outgrowth in cell culture and to stimulate nerve regeneration following peripheral nerve injury [151], and to be neuroprotective in brain ischemia [152, 153]. The concentrations of FK506 required for neurotrophic activity also inhibit calcineurin, which can lead to suppression of immune function.

The adverse effects of FK506 have stimulated efforts to identify compounds with neurotrophic activity but lacking calcineurin inhibitory activity. The immunophilin ligand GPI1046 (Fig. 8) was reported to display neurotrophic actions and to stimulate structural and functional recovery in in vivo models of neurodegeneration [154–157] and to protect cultured dopaminergic neurons from toxicity by 1-methyl-4-phenylpyridine ($MPP^+$) and 6-hydroxydopamine (6-OHDA) [156–158] as well as promote dopaminergic function in 6-OHDA-lesioned rats [159]. More recent studies have shown GPI1046 to protect neurons from the lethal effects of human immunodeficiency virus-1 (HIV-1) proteins gp120 and Tat, by modulating endoplasmic reticulum calcium overload [160] and to induce selective expression of the astroglial glutamate transporter GLT1 in vitro and in vivo [161]. In the last study, treatment with GPI1046 protected motor neurons in an in vitro model of chronic excitotoxicity and prolonged the survival of transgenic mice with motor neuron disease [161]. V-10,367, an immunophilin ligand that binds to FKBP12 ($K_i = 0.5\,\mathrm{n}M$) (Fig. 8), and V-13,661, an immunophilin ligand that does not bind to FKBP12 ($K_i > 25\,\mathrm{m}M$), both protected against loss of dopaminergic terminals in the striatum following treatment with MPTP (*N*-methyl-4-phenyl-1,2,3,6-tetrahydropyridine), which induces Parkinson-like symptoms in mice and primates [162]. V-10,367 was also protective in animals treated with a 6-OHDA lesion in the striatum. Both compounds promoted sprouting of intact neurons but did not increase survival of neurons in the substantia nigra. Other studies, however, have failed to demonstrate morphological protection of dopaminergic neurons in the substantia nigra by GPI1046 in MPTP-treated mice [163] or after 6-OHDA lesions [164]. One study testing GPI-1046 for behavioral and neuropathological changes following MPTP administration to monkeys found no evidence for a neuroprotective effect [165]. Taken together, the data on GPI1046 and V-10,367 suggest that these immunophilin ligands are able to elicit modest sprouting effects in vivo and in vitro, under certain conditions, but have no effect on neuronal cell survival.

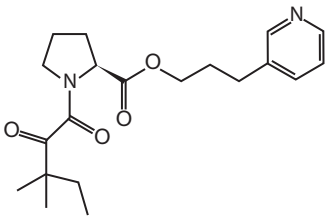


**Figure 8** The Structures of GPI1046 and V-10,367.

The nature of the FKBP representing the primary target of these drugs remains enigmatic. Until now, 16 FKBPs have been identified in the human genome. There is increasing evidence to suggest that FKBP12 does not affect neuronal cell protection and regeneration. The peptidylprolyl isomerase (PPIase) activity targeted by active site-directed FKBP ligands, such as GPI1046, inhibit the calmodulin/$Ca^{2+}$-regulated FKBP38 with up to 80-fold higher affinity than FKBP12. The FKBP38-specific cycloheximide derivative, *N-(N′,N′*-dimethylcarboxamidomethyl)cycloheximide caused neuronal cell protection as well as neural stem cell proliferation in a rat model of transient focal cerebral ischemia [166]. These effects were still dominant when compound was applied 2–6 h postinsult. In parallel, sustained motor behavior deficits of injured rats were inproved by drug administration [166].

## 9 INHIBITION OF INTRACELLULAR SIGNALING PATHWAYS

Apoptosis has been proposed as a mechanism of cell death in Alzheimer's, Parkinson's, and Huntington's diseases and in amyotrophic lateral sclerosis. The stress-activated protein kinase or c-Jun N-terminal kinase (JNK) is a MAPK pathway activated in response to many extracellular stimuli and different forms of environmental stress (e.g., metabolic stress, radiation), including neurotrophic factor deprivation [167]. Staining with antibodies that recognize phosphorylated JNK

proteins reveals JNK activation in brain regions (cerebral cortex, hippocampus) affected by Alzheimer's disease pathology but not in unaffected regions or control brains [168, 169]. JNK activation has been observed also in brain specimens from patients with Parkinson's disease [170] and in midbrain dopaminergic neurons from MPTP-treated rodents [171]. MPTP-induced neurodegeneration and motor deficits were significantly reduced in JNK2/JNK3 knockout mice [172].

The MAPK pathway is organized as a cascade of at least three kinases: an MAPK kinase kinase (MAPKKK) first phosphorylates a dual-specificity protein kinase (MAPKK), which in turn phosphorylates the MAPK. The JNK pathway can be activated by small G-proteins, such as Cdc42 and Rac, members of the Rho family guanosine triphosphates (GTPases). In neurons, the mixed lineage kinase (MLK) family of kinases serves as the major MAPKKKs and phosphorylates the MAPK kinases 4 and 7 (MKK4 and MKK7), which in turn phosphorylate the JNKs [167]. Activation of the JNK pathway induces the expression of cell death genes (Fig. 9). Components of this pathway provide opportunities for pharmacological intervention and neuroprotection.

As discussed earlier, K-252a at high concentrations inhibits the survival-promoting effects of neurotrophins. In contrast, at low concentrations, K-252a on its own promotes survival and differentiation similar to the effects of neurotrophins [173, 174]. Compounds based on K-252a were synthesized to enhance the neurotrophic effects of K-252a while "dialing out" Trk inhibitory activity. A series of 3,9-bis[(alkylthio)methyl]- and bis(alkoxymethyl)-K-252a derivatives with reduced Trk inhibitory activity were identified, with 3,9-bis-[(ethylthio)methyl]-(8*R**,9*S**,11*S**)-(–)-9-hydroxy-9-methoxycarbonyl-8-methyl-2,3,9,10-tetrahydro-8,11-epoxy-1*H*,8*H*,11*H*,-2,7*b*,11*a*-triazadibenzo(*a,g*)cycloocta(*cde*)trinden-1-one] (CEP-1347 or KT-7515) (Fig. 10) having the most potent neurotrophic effects [175]. Upstream regulators of the JNK pathway were co-transfected into cells with JNK1 in the presence or absence of CEP-1347 and JNK1 activity measured by an in vitro kinase assay. This led to the identification of MLKs as a target for CEP-1347 [176]. CEP-1347 directly inhibits the kinase activities of the MLK family members in vitro with nanomolar $IC_{50}$ values [177] that are in the

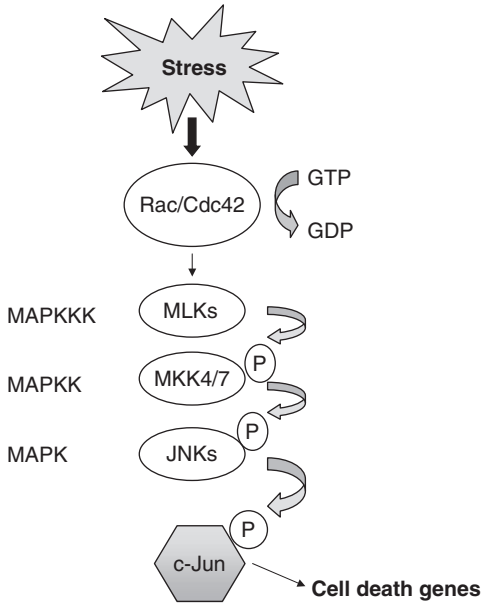


**Figure 9** The c-Jun N-terminal kinase (JNK) pathway in cell death. See text for details.

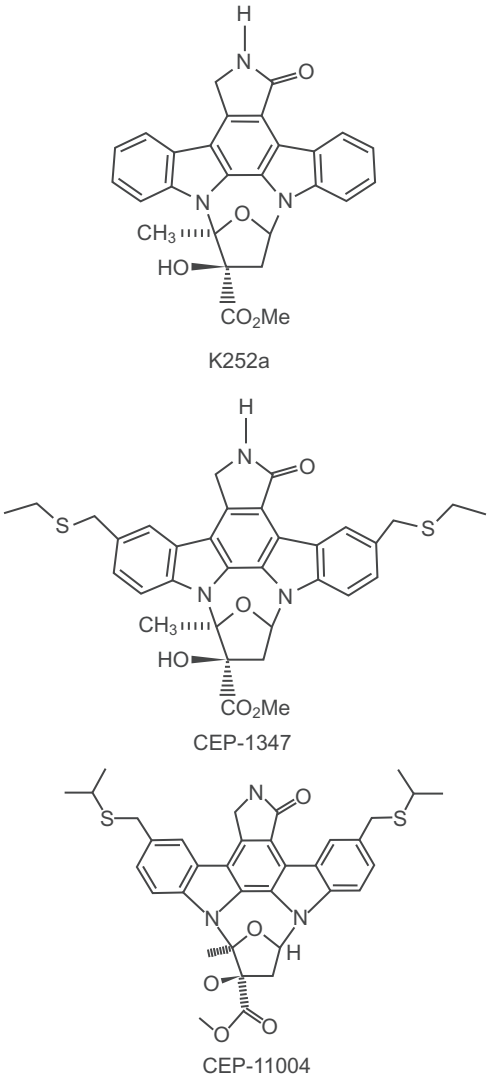


**Figure 10** The Structures of K-252a, CEP-1347, and CEP-11004.

same range as $EC_{50}$ values for survival promotion in intact neurons [178, 179]. The related indolocarbazole analog 3,9-bis-[(isopropylthio)methyl]-(8*R**,9*S**,11*S**)-(−)-9-hydroxy-9-methoxycarbonyl-8-methyl-2,3,9,10-tetrahydro-8,11-epoxy-1*H*,8*H*,11*H*,-2,7*b*,11*a*-triazadibenzo(*a,g*)cycloocta(*cde*)trinden-1-one] (CEP-11004) (Fig. 10) also inhibits MLK1, MLK2, and MLK3 with nanomolar affinity [177].

CEP-1347 and CEP11004 have broad neuroprotective effects in vitro, inhibiting NGF withdrawal-induced cell death [177, 178], trophic factor-deprived spinal cord motoneuron death [179], $MPP^+$-induced neurotoxicity [177, 180], and β-amyloid-induced cortical neuron death [177, 181]. Inhibition of JNK pathway signaling by CEP-1347 also preserves metabolism and growth of neurotrophic factor-deprived neurons, thereby maintaining not only survival but the "quality of life" of neurons [182]. More recent findings suggest that in addition to the direct inhibition of the JNK pathway, the indirect activation of the phosphatidylinositol 3-kinase/Akt pathway via Trk activation is important for MLK inhibitor-mediated long-term neuronal survival and trophism [183, 184].

Neuroprotective effects of CEP-1347 in vivo include prevention of developmental programmed and axotomy-induced motoneuron cell death [185], basal forebrain cholinergic neuron death caused by fimbria-fornix transaction [186] or excitotoxins [187], and MPTP-induced substantia nigra dopaminergic cell loss in rodents [188] and nonhuman primates [189]. Additionally, systemic administration of CEP-11004 to mice is reported to inhibit the MPTP-mediated increase in phosphor-MKK4 in the substantia nigra and attenuated the MPTP-mediated loss of striatal dopaminergic terminals [177].

There is increasing evidence that neuroinflammation driven by abnormal or prolonged glial cell activation contributes to the pathogenesis and progression of Alzheimer's disease. Nornally, glia become transiently activated in the healthy brain in response to a number of stresses and generally act to return the CNS to homeostasis. However, the neuroinflammation hypothesis of Alzheimer's disease progression proposes that stresses such as amyloid peptide result in increased levels of inflammatory molecules produced by chronically activated glia, leading to neuron damage, inducing further glial activation, thereby resulting in a vicious cycle of neuroinflammation and neurodegeneration [190]. Indeed, activated microglia can cause neuronal cell death in vitro [191–193]. Interestingly, pharmacological blockade of MLK has been described to inhibit pro-inflammatory cytokine release from lipopolysaccharide- and amyloid β-peptide-stimulated microglia in vitro, and brain tumor necrosis factor-α (TNF-α) production induced by intracerebroventricular injection of lipopolysaccharide in mice [194, 195]. Modulation of the inflammatory response with MLK inhibitors conceivably may be of therapeutic benefit in retarding the progression of Alzheimer's disease pathology, as well as neuronal cell loss and neurodegeneration in other CNS disorders resulting from trauma, ischemia, or inflammation.

Amyotrophic lateral sclerosis is a fatal neurodegenerative disorder characterized by progressive death of cortical and spinal motor neurons. As mentioned earlier, when administered to animal models, by cellular or gene therapy, neurotrophic factors were effective in slowing motor neuron degeneration [196–199]. However, clinical trials failed to demonstrate efficacy, probably due to the poor absorption, distribution, and metabolic properties of these polypeptides [29–35, 200]. Interest has, therefore, turned toward identifying small molecules that can mimic trophic factors and support motor neuron survival and thereby treat motor neuron diseases. Using a phenotypic cell-based screening approach, a collection of approximately 40,000 low-molecular-weight compounds was assessed to identify small-molecule therapeutics. This led to the discovery of cholest-4-en-one, oxime (TRO19622) (Fig. 11), as a potential drug candidate [201]. In vitro, TRO19622 promoted motor neuron survival in the absence of trophic support, and in vivo the compound rescued motor neurons from axotomy-induced cell death in neonatal rats and promoted nerve regeneration following sciatic nerve crush in mice [201]. In $SOD^{G93A}$ transgenic mice, a model of familial amytrophic lateral sclerosis, TRO19622 treatment improved motor performance, delayed the onset of the clinical disease, and extended survival. TRO19622 bound directly to two components of the mitochondrial permeability transition pore: the voltage-dependent ion channel and the translocator protein 18 kDa (or peripheral benzodiazepine receptor), defining these as potential new therapeutic targets for the diverse traumatic and genetic lesions of the motor system that involve mitochondrial dysfunction [201].

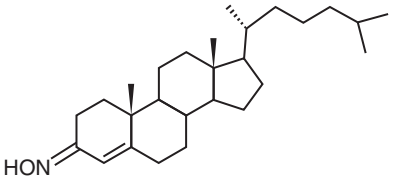


**Figure 11** The Structure of TRO19622. TRO19622 exists as a stable mixture of syn- and anti-isomers of cholest-4-en-3-one, oxime.

## 10 CONCLUDING REMARKS

It has long been hoped that the power of neurotrophic factors to regulate neuronal cell survival in the developing nervous system could be harnessed for the treatment of neurodegenerative disease. Preclinical data suggest that subcutaneous or intravenous administration of neurotrophic factors may be effective for the treatment of peripheral and central nervous system diseases. In spite of such encouraging experimental findings, clinical studies have proven largely disappointing [202]. Not only have neurotrophic proteins failed to show a beneficial effect, but in some cases adverse side effects were noteworthy. Therapeutic benefit may be dependent on the extent of neurodegeneration before therapy initiation, and, although studies will be difficult, assessment of the effects of neurotrophic factors on the rate of progression over sufficiently long time frames may well be critical in determining their ultimate efficacy.

A detailed understanding of neurotrophic factor-receptor interactions and intracellular signaling [203] has been crucial in the design of peptidyl and nonpeptidyl small-molecule neurotrophin mimetics that interact directly with the receptor, or which potentiate neurotrophin activation of its cognate receptor [204, 205]. Another example of how such knowledge has been exploited is that of the $p75^{NTR}$, which can promote cell survival in the absence of Trk receptors. A pharmacophore designed to capture selected structural and physicochemical features of a neurotrophin domain known to interact with $p75^{NTR}$ was applied to in silico screening of small-molecule libraries to select neurotrophic compounds. In spite of such successes, several chemical entities directed against key targets involved in the apoptotic process (mixed linkage kinases [MLKs]; glyceraldehyde 3-phosphate dehydrogenase [GAPDH]) have proved ineffective in Parkinson's disease patients, although robust neuroprotective activity was demonstrated in multiple in vitro and in vivo models of neuronal cell death [206]. Reasons for failure are potentially many, including possible redundancy in death pathways, neurons being "resuced" but not functional, pharmacokinetic behaviors between patients, and the relevance of current animal models to the disease in humans.

In spite of these challenges, the development of neurotrophic compounds is potentially very rewarding and appears to offer real promise for disease modification. Neurotrophic drug development, however, has historically been a high-risk approach. Nonetheless, novel emerging targets and technological improvements, as well as the development of biomarkers that can act as surrogates to assess drug activity at the defined molecular target together with a more focused effort on translational medicine approaches [207], may facilitate the development of neurotrophic small molecules with lower associated attrition rates.

## REFERENCES


1. Hefi F (1986). Nerve growth factor promotes survival of septal cholinergic neurons after fimbrial transactions. *J Neurosci 6*, 2155–2162.
2. Williams LR, Varon S, Peterson GM, Wictorin K, Fischer W, Bjorklund A, Gage FH (1986). Continuous infusion of nerve growth factor prevents basal forebrain neuronal death after fimbria fornix transaction. *Proc Nat Acad Sci USA 83*, 9231–9235.
3. Kordower JH, Emborg ME, Bloch J, Ma SY, Chu Y, Leventhal L, McBride J, Chen EY, Palfi S, Roitberg BZ, Brown WD, Holden JE, Pyzalski R, Taylor MD, Carvey P, Ling Z, Trono D, Hantraye P, Deglon N, Aebischer P (2000). Neurodegeneration prevented by lentiviral vector delivery of GDNF in primate models of Parkinson's disease. *Science 290*, 767–773.
4. Grondin R, Zhang Z, Ai Y, Gash DM, Gerhardt GA (2003). Intracranial delivery of proteins and peptides as a therapy for neurodegenerative diseases. *Prog Drug Res 61*, 101–123.
5. Ishii DN, Pu S-F (1999). Neuropharmacology of insulin-like growth factors. In Hefti F, Ed. *Neurotrophic Factors*. Berlin: Springer, pp. 119–146.
6. Sendtner M (1999). Neurotrophic factors and amyotrophic lateral sclerosis. In Hefti F, Ed. *Neurotrophic Factors*. Berlin: Springer, pp. 81–117.
7. Schnell L, Schneider R, Kolbeck R, Barde Y-A (1994). Neurotrophin-3 enhances sprouting of corticospinal tract during development and after spinal cord lesion. *Nature 367*, 170–173.
8. Kobayashi NR, Fan DP, Giehl KM, Bedard AM, Wiegand SJ, Tetzlaff W (1997). BDNF and NT-4/5 prevent atrophy of rat rubrospinal neurons after cervical axotomy, stimulate GAP-43 and Tα1-tubulin mRNA expression, and promote axonal regeneration. *J Neurosci 17*, 9583–9595.
9. Ramer MS, Priestley JV, McMahon SB (2000). Functional regeneration of sensory axons into the adult spinal cord. *Nature 403*, 312–316.
10. Sofroniew MV, Howe CL, Mobley WC (2001). Nerve growth factor signaling, neuroprotection, and neural repair. *Annu Rev Neurosci 24*, 1217–1281.
11. Chao MV (2003). Neurotrophins and their receptors: A convergence point for many signalling pathways. *Nat Rev Neurosci 4*, 299–309.
12. Lu B, Pang PT, Woo NH (2005). The yin and yang of neurotrophin action. *Nat Rev Neurosci 6*, 603–614.
13. Reichardt LF (2006). Neurotrophin-regulated signalling pathways. *Philoso Trans R Soci B 361*, 1545–1564.
14. Levi-Montalcini R (1987). The nerve growth factor 35 years later. *Science 237*, 1154–1162.

15. Ibáñez CF (1994). Structure-function relationships in the neurotrophin family. *J Neurobiol 25*, 1349–1361.
16. Götz R, Köster R, Winkler C, Raulf F, Lottspeich F, Schartl M, Thoenen H (1994). Neurotrophin-6 is a new member of the nerve growth factor family. *Nature 372*, 266–269.
17. Lai K-O, Fu W-Y, Ip FCF, Ip NY (1998). Cloning and expression of a novel neurotrophin, NT-7, from carp. *Mol Cell Neurosci 11*, 64–76.
18. Rodriguez-Tebar A, Dechant G, Barde YA (1990). Binding of brain-derived neurotrophic factor to the nerve growth factor receptor. *Neuron 4*, 487–492.
19. Frade JM, Barde Y-A (1998). Nerve growth factor: Two receptors, multiple functions. *BioEssays 20*, 137–145.
20. Davies AM, Lee K-F, Jaenisch R (1993). p75-Deficient trigeminal sensory neurons have an altered response to NGF but not to other neurotrophins. *Neuron 11*, 565–574.
21. Mahadeo D, Kaplan L, Chao MV, Hempstead BL (1994). High affinity nerve growth factor binding displays a faster rate of association than $p140^{trk}$ binding. *J Biol Chem 269*, 6884–6891.
22. Esposito D, Patel P, Stephens RM, Perez P, Chao MV, Kaplan DR, Hempstead BL (2001). The cytoplasmic and transmembrane domains of the p75 and Trk A receptors regulate high affinity binding to nerve growth factor. *J Biol Chem 276*, 32687–32695.
23. Ultsch MH, Wiesmann C, Simmons LC, Henrich J, Yang M, Reilly D, Bass SH, de Vos AM (1999). Crystal structures of the neurotrophin-binding domain of TrkA, TrkB and TrkC. *J Mol Biol 290*, 149–159.
24. Wiesmann C, Ultsch MH, Bass SH, de Vos AM (1999). Crystal structure of nerve growth factor in complex with the ligand-binding domain of the TrkA receptor. *Nature 401*, 184–188.
25. Urfer R, Tsoulfas P, O'Connell L, Hongo J-A, Zhao W, Presta LG (1998). High resolution mapping of the binding site of TrkA for nerve growth factor and TrkC for neurotrophin-3 on the second immunoglobulin-like domain of the Trk receptors. *J Biol Chem 273*, 5829–5840.
26. Wiesmann C, de Vos AM (2001). Nerve growth factor: Structure and function. *Cell Mol Life Sci 58*, 748–759.
27. Björklund A, Rosenblad C, Winkler C, Kirik D (1997). Studies on neuroprotective and regenerative effects of GDNF in a partial lesion model of Parkinson's disease. *Neurobiol Dis 4*, 186–200.
28. Gash DM, Zhang Z, Ovadia A, Cass WA, Yi A, Simmerman L, Russell D, Martin D, Lapchak PA, Collins F, Hoffer BJ, Gerhardt GA (1996). Functional recovery in parkinsonian monkeys treated with GDNF. *Nature 380*, 252–255.
29. Lange DJ, Felice KJ, Festoff BW, Gawel MJ, Gelinas DF, Kratz R, Lai EC, Murphy MF, Natter HM, Norris FH, Rudnicki S (1996). Recombinant human insulin-like growth factor-1 in ALS: Description of a double-blind, placebo-controlled study. North American ALS/IGF-1 Study Group. *Neurology 47*, 93–94.
30. Borasio GD, Robberecht W, Leigh PN, Emile J, Guiloff RJ, Jerusalem F, Silani V, Vos PE, Wokke JH, Dobbins T (1999). A placebo-controlled trial of insulin-like growth factor-1 in amytrophic lateral sclerosis. European ALS/IGF-1 Study Group. *Neurology 51*, 583–586.
31. The BDNF Study Group (Phase III) (1999). A controlled trial of recombinant methionyl human BDNF in ALS. *Neurology 52*, 1427–1433.
32. Ochs G, Penn RD, York M, Tonn J, Haigh J, Malta E, Traub M, Sendtner M, Toyka KV (2000). A phase I/II trial of recombinant methionyl human brain derived neurotrophic factor administered by intrathecal infusion to patients with amyotrophic lateral sclerosis. *Amyotrophic Lateral Sclerosis and Other Motor Neuron Disorders 1*, 201–206.
33. Beck M, Flachenecker P, Magnus T, Giess R, Reiners K, Toyka KV, Naumann M (2005). Autonomic dysfunction in ALS: A preliminary study on the effects of intrathecal BDNF. *Amyotrophic Lateral Sclerosis and Other Motor Neuron Disorders 6*, 100–103.
34. ALS CNTF Treatment Study Group (1996). A double-blind placebo-controlled clinical trial of subcutaneous recombinant human ciliary neurotrophic factor (rHCNTF) in amyotrophic lateral sclerosis. *Neurology 46*, 1244–1249.
35. Turner MR, Parton MJ, Leigh PN (2001). Clinical trials in ALS: An overview. *Semin Neurol 21*, 167–175.
36. Nutt JG, Burchiel KD, Comella CL, Jankovic J, Land AE, Laws ER Jr, Lozano AM, Penn RD, Simpson RK Jr, Stacy M, Wooten GF, ICV GDNF Study Group (2003). Randomized, double-blind trial of glial cell line-derived neurotrophic factor (GDNF) in PD. *Neurology 60*, 69–73.
37. Lang AE, Gill S, Patel NK, Lozano A, Nutt JG, Penn R, Brooks DJ, Hotton G, Moro E, Heywood P, Brodsky MA, Burchiel K, Kelly P, Dalvi A, Scott B, Stacy M, Turner D, Wooten VGF, Elias WJ, Laws ER, Dhawan V, Stoessl AJ, Matcham J, Coffey RJ, Traub M (2006). Randomized controlled trial of intraputamenal glial cell line-derived neurotrophic factor infusion in Parkinson disease. *Ann Neurol 59*, 459–466.
38. Podulso JF, Curran GL, Gill JS (1998). Putrescine-modified nerve growth factor: Bioactivity, plasma pharmacokinetics, blood-brain/nerve barrier permeability, and nervous system biodistribution. *J Neurochem 71*, 1651–1660.
39. Wu D, Pardridge WM (1999). Neuroprotection with non-invasive neurotrophin delivery to the brain. *Proc Nat Acad Sci USA 96*, 254–259.
40. Zhang Y, Pardridge WM (2001). Neuroprotection in transient focal brain ischemia after delayed intravenous administration of brain-derived neurotrophic factor conjugated to a blood-brain barrier drug targeting system. *Stroke 32*, 1378–1384.

41. Kotzbauer PT, Lampe PA, Heuckeroth RO, Golden JP, Creedon DJ, Johnson EM Jr, Milbrandt J (1996). Neurturin, a relative of glial-cell-line-derived neurotrophic factor. *Nature 384*, 467–470.
42. Choi-Lundberg DL, Lin Q, Chang Y-N, Chiang YL, Hay CM, Mohajeri H, Davidson BL, Bohn MC (1997). Dopaminergic neurons protected from degeneration by GDNF gene therapy. *Science 275*, 838–841.
43. Fjord-Larsen L, Johansen JL, Kusk P, Tornøe, Grønborg M, Rosenblad C, Wahlberg LU (2005). Efficient in vivo protection of nigral dopaminergic neurons by lentiviral gene transfer of a modified Neurturin construct. *Exp Neurol 195*, 49–60.
44. Zhou L, Baumgartner BJ, Hill-Felberg SJ, McGowen LR, Shine HD (2003). Neurotrophin-3 expressed in situ induces axonal plasticity in the adult injured spinal cord. *J Neurosci 23*, 1424–1431.
45. Blesch A, Yang H, Weidner N, Hoang A, Otero D (2004). Axonal responses to cellularly delivered NT-4/5 after spinal cord injury. *Mol Cell Neurosci 27*, 190–201.
46. Taylor L, Jones L, Tuszynski MH, Blesch A (2006). Neurotrophin-3 gradients established by lentiviral gene delivery promote short-distance axonal bridging beyond cellular grafts in the injured spinal cord. *J Neurosci 26*, 9713–9721.
47. Tuszynski MH, Thal L, Pay M, Salmon DP, U HS, Bakay R, Patel P, Blesch A, Vahlsing HL, Ho G, Tong G, Potkin SG, Fallon J, Hansen L, Mufson EJ, Kordower JH, Gall C, Conner J (2005). A phase 1 clinical trial of nerve growth factor gene therapy for Alzheimer disease. *Nat Med 11*, 551–555.
48. Brewster WJ, Fernyhough P, Diemel, LT, Mohiuddin L, Tomlinson RR (1994). Diabetic neuropathy, nerve growth factor and other neurotrophic factors. *Trends Neurosci 17*, 321–325.
49. Anand P, Terenghi G, Warner G, Kopelman P, Williams-Chestnut RE, Sinicropi DV (1996). The role of endogenous nerve growth factor in human diabetic neuropathy. *Nat Med 2*, 703–707.
50. Apfel SC (2002). Nerve growth factor for the treatment of diabetic neuropathy: What went wrong, what went right, and what does the future hold? *Intl Rev Neurobiol 50*, 393–413.
51. Apfel SC, Kessler JA, Adornato BT, Litchy WJ, Sanders C, Rask CA (1998). Recombinant human nerve growth factor in the treatment of diabetic polyneuropathy. NGD Study Group. *Neurology 51*, 695–702.
52. McArthur JC, Yiannoutsos C, Simpson DM, Adornato BT, Singer EJ, Hollander H, Marra C, Rubin M, Cohen BA, Tucker T, Navia BA, Schifitto G, Katzenstein D, Rask C, Zaborski L, Smith ME, Shriver S, Millar M, Clifford DB (2000). A phase II trial of nerve growth factor for sensory neuropathy associated with HIV infection. *Neurology 54*, 1080–1088.
53. Apfel SC, Schwartz S, Adornato BT, Freeman R, Biton V, Rendell M, Vinik A, Giuliani M, Stevens JC, Barbano R, Dyck PJ (2000). Efficacy and safety of human nerve growth factor in patients with diabetic polyneuropathy: A randomized controlled trial. *JAMA 284*, 2215–2221.
54. Schmidt RE, Dorsey DA, Beaudet LN, Parvin CA, Escandon E (2001). Effect of NGF and neurotrophin-3 treatment on experimental diabetic autonomic neuropathy. *J Neuropathol Exp Neurol 60*, 263–273.
55. Lewin GR, Mendell LM (1993). Nerve growth factor and nociception. *Trends Neurosci 16*, 353–359.
56. Calcutt NA, Allendoerfer KL, Mizisin AP, Middlemas A, Freshwater JD, Burgers M, Ranciato R, Delcroix J-D, Taylor FR, Shapiro R, Strauch K, Dudek H, Engber TM, Galdes A, Rubin LL, Tomlinson DR (2003). Therapeutic efficacy of sonic hedgehog protein in experimental diabetic neuropathy. *J Clin Investig 111*, 507–514.
57. Mizisin AP, Vu Y, Shuff M, Calcutt NA (2004). Ciliary neurotrophic factor improves nerve conduction and ameliorates regeneration deficits in diabetic rats. *Diabetes 53*, 1807–1812.
58. Saragovi HU, Greene MI, Chrusciel RA, Kahn M (1992). Loops and secondary structure mimetics: Development and applications in basic science and rational drug design. *Biotechnology 10*, 773–778.
59. Saragovi HU, Fitzpatrick D, Raktabuhr A, Nakanishi H, Kahn M, Green MI (1991). Design and synthesis of a mimetic of an antibody complementarity region. *Science 253*, 792–795.
60. LeSauteur L, Wei L, Gibbs BF, Saragovi HU (1995). Small peptide mimics of nerve growth factor bind TrkA receptors and affect biological responses. *J Biol Chem 270*, 6564–6569.
61. LeSauteur L, Cheung N-KV, Lisbona R, Saragovi HU (1996). Small molecule nerve growth factor analogs image receptors in vivo. *Nat Biotechnol 14*, 1120–1122.
62. Debeir T, Saragovi HU, Cuello AC (1999). A nerve growth factor mimetic TrkA antagonist causes withdrawal of cortical cholinergic boutons in the adult rat. *Proc Nat Acad Sci USA 96*, 4067–4072.
63. LeSauteur L, Maliartchouk S, Le Jeune H, Quirion R, Saragovi HU (1996). Potent human p140-TrkA agonists derived from an anti-receptor momoclonal antibody. *J Neurosci 16*, 1308–1316.
64. Maliartchouk S, Debeir T, Beglova N, Cuello AC, Gehring K, Saragovi HU (2000). Genuine monovalent ligands of TrkA nerve growth factor receptors reveal a novel pharmacological mechanism of action. *J Biol Chem 275*, 9946–9956.
65. Xie Y, Tisi MA, Yeo TT, Longo FM (2000). Nerve growth factor (NGF) loop 4 dimeric mimetics activate ERK and AKT and promote NGF-like neurotrophic effects. *J Biol Chem 275*, 29868–29874.
66. Maliartchouk S, Feng Y, Ivanisevic L, Debeir T, Cuello AC, Burgess K, Saragovi HU (2000). A designed peptidomimetic agonistic ligand of TrkA nerve growth factor receptors. *Mol Pharmacol 57*, 385–391.
67. Rashid K, Van der Zee CEEM, Ross GM, Chapman CA, Stanisz J, Riopelle RJ, Racine RJ, Fahnestock M (1995).

A nerve growth factor peptide retards seizure development and inhibits neuronal sprouting in a rat model of epilepsy. *Proc Nat Acad Sci USA 92*, 9495–9499.

68. Bruno MA, Clarke PBS, Seltzer A, Quirion R, Burgess K, Cuello AC, Saragovi HU (2004). Long-lasting rescue of age-associated deficits in cognition and the CNS cholinergic phenotype by a partial agonist peptidomimetic ligand of TrkA. *J Neurosci 24*, 8009–8018.
69. McDonald NQ, Lapatto R, Murray-Rust J, Gunning J, Wlodawer A, Blundell TL (1991). New protein fold revealed by a 2.3-Å resolution crystal structure of nerve growth factor. *Nature 354*, 411–414.
70. McDonald NQ, Chao MV (1995). Structural determinants of neurotrophin action. *J Biol Chem 270*, 19669–19672.
71. Kahle P, Burton LE, Schmelzer CH, Hertel C (1992). The amino terminus of nerve growth factor is involved in the interaction with the receptor tyrosine kinase $p140^{trkA}$. *J Biol Chem 267*, 22707–22710.
72. Shih A, Laramee GR, Schmelzer CH, Burton LE, Winslow JW (1994). Mutagenesis identifies amino-terminal residues of nerve growth factor necessary for Trk receptor binding and biological activity. *J Biol Chem 269*, 27679–27686.
73. Woo SB, Timm DE, Neet KE (1995). Alteration of $NH_2$-terminal residues of nerve growth factor affects activity and Trk binding without affecting stability or conformation. *J Biol Chem 270*, 6278–6285.
74. Krüttgen A, Heymach JV Jr, Kahle PJ, Shooter EM (1997). The role of the nerve growth factor carboxyl terminus in receptor binding and conformational stability. *J Biol Chem 272*, 29222–29228.
75. Rydén M, Ibàñez CF (1997). A second determinant of binding to the p75 neurotrophin receptor revealed by alanine-scanning mutagenesis of a conserved loop in nerve growth factor. *J Biol Chem 272*, 33085–33091.
76. Bradshaw RA, Murray-Rust J, Ibàñez CF, McDonald NQ, Lapatto R, Blundell TL (1994). Nerve growth factor: Structure/function relationship. *Prot Sci 3*, 1901–1913.
77. Longo FM, Vu TK, Mobley WC (1990). The in vitro biological effect of nerve growth factor is inhibited by synthetic peptides. *Cell Regul 1*, 189–195.
78. Longo FM, Manthorpe M, Xie YM, Varon S (1997). Synthetic NGF peptide derivatives prevent neuronal death via a p75 receptor-dependent mechanism. *J Neurosci Res 48*, 1–17.
79. Colangelo AM, Bianco MR, Vitagliano L, Cavaliere C, Cirillo G, De Gioia L, Diana D, Colombo D, Redaelli C, Zaccaro L, Morelli G, Papa M, Sarmientos P, Alberghina L, Martegani E (2008). A new nerve growth factor-mimetic peptide active on neuropathic pain in rats. *J Neurosci 28*, 2698–2709.
80. O'Leary PD, Hughes RA (2003). Design of potent peptide mimetics of brain-derived neurotrophic factor. *J Biol Chem 278*, 25738–25744.
81. O'Leary PD, Hughes RA (1998). Structure-activity relationships of conformationally constrained peptide analogues of loop 2 of brain-derived neurotrophic factor. *J Neurochem 70*, 1712–1721.
82. Zhang AJ, Khare S, Gokulan K, Linthicum DS, Burgess K (2003). Dimeric β-turn peptidomimetics as ligands for the neurotrophin receptor TrkC. *Bioorg Med Chem Lett 11*, 207–210.
83. Williams G, Williams E-J, Maison P, Pangalos MN, Walsh FS, Doherty P (2005). Overcoming the inhibitors of myelin with a novel neurotrophin strategy. *J Biol Chem 280*, 5862–5869.
84. Skaper SD (2005). Neuronal growth-promoting and inhibitory cues in neuroprotection and neuroregeneration. *Ann NY Acad Sci 1053*, 376–385.
85. Williams E, Williams G, Gour BJ, Blaschuk OW, Doherty P (2000). A novel family of cyclic peptide antagonists suggests that N-cadherin specificity is determined by amino acids that flank the HAV motif. *J Biol Chem 275*, 4007–4012.
86. Williams G, Williams E-J, Doherty P (2002). Dimeric versions of two short N-cadherin binding motifs (HAVDI and INPISG) function as N-cadherin agonists. *J Biol Chem 277*, 4361–4367.
87. Skaper SD, Facci L, Williams G, Williams E-J, Walsh FS, Doherty P (2004). A dimeric version of the short N-cadherin binding motif HAVDI promotes neuronal cell survival by activating an N-cadherin/fibroblast growth factor receptor signalling cascade. *Mol Cell Neur 26*, 17–23.
88. Williams EJ, Furness J, Walsh FS, Doherty P (1994). Activation of the FGF receptor underlies neurite outgrowth stimulated by L1, N-CAM, and N-cadherin. *Neuron 13*, 583–594.
89. Roux P, Barker P (2002). Neurotrophin signaling through the p75 neurotrophin receptor. *Prog Neurobiol 67*, 203–233.
90. Rabizadeh S, Oh J, Zhong LT, Yang J, Bitler CM, Butcher LL, Bredesen DE (1993). Induction of apoptosis by the low-affinity NGF receptor. *Science 261*, 345–358.
91. Friedman WJ (2000). Neurotrophins induce death of hippocampal neurons via the p75 receptor. *J Neurosci 20*, 6340–6346.
92. Casaccia-Bonnefil P, Carter BD, Dobrowsky RT, Chao MV (1996). Death of oligodendrocytes mediated by the interaction of nerve growth factor with its receptor p75. *Nature 383*, 716–719.
93. Lee R, Kermani P, Teng KK, Hempstead BL (2001). Regulation of cell survival by secreted proneurotrophins. *Science 294*, 1945–1948.
94. Beattie MS, Harrington AW, Lee R, Kim JY, Boyce SL, Longo FM, Bresnahan JC, Hempstead BL, Yoon SO (2002). ProNGF induces p75-mediated death of oligodendrocytes following spinal cord injury. *Neuron 36*, 375–386.

95. Harrington AW, Leiner B, Blechschmitt C, Arevalo JC, Lee R, Mörl K, Meyer M, Hempstead BL, Yoon SO, Giehl KM (2004). Secreted proNGF is a pathophysiological death-inducing ligand after adult CNS injury. *Proc Nat Acad Sci USA 101*, 6226–6230.
96. Nykjaer A, Lee R, Teng KK, Jansen P, Madsen P, Nielsen MS,, Jacobsen C, Kliemannel M, Schwarz E, Willnow TE, Hempstead BL, Petersen CM (2004). Sortilin is essential for proNGF-induced neuronal cell death. *Nature 427*, 843–848.
97. Fahnestock M, Michalski B, Xu B, Coughlin MD (2001). The precursor pro-nerve growth factor is the predominant form of nerve growth factor in brain and is increased in Alzheimer's disease. *Mol Cell Neurosci 18*, 210–220.
98. Volosin M, Song W, Almeida RD, Kaplan DR, Hempstead BL, Friedman WJ (2006). Interaction of survival and death signaling in basal forebrain neurons: Roles of neurotrophins and proneurotrophins. *J Neurosci 26*, 7756–7766.
99. Carter BD, Kaltschmidt C, Kaltschmidt B, Offenhäuser N, Böhm-Matthaei R, Baeuerle PA, Barde Y-A (1996). Selective activation of NF-κB by nerve growth factor through the neurotrophin receptor p75. *Science 272*, 542–545.
100. Reddypalli S, Roll K, Lee H-K, Lundell M, Barea-Rodriguez E, Wheeler EF (2005). $p75^{NTR}$–Mediated signaling promotes the survival of myoblasts and influences muscle strength. *J Cell Physiol 204*, 819–829.
101. He XL, Garcia KC (2004). Structure of nerve growth factor complexed with the shared neurotrophin receptor p75. *Science 304*, 870–875.
102. Massa SM, Xie Y, Longo FM (2003). Alzheimer's therapeutics: neurotrophin domain small molecule mimetics. *J Mol Neurosci 20*, 323–326.
103. Massa SM, Xie Y, Yang T, Harrington AW, Kim MY, Yoon SO, Kraemer R, Moore LA, Hempstead BL, Longo FM (2006). Small, nonpeptide $p75^{NTR}$ ligands induce survival signaling and inhibit proNGF-induced death. *J Neurosci 26*, 5288–5300.
104. Pehar M, Cassina P, Vargas MR, Xie Y, Beckman JS, Massa SM, Longo FM, Barbeito L (2006). Modulation of $p75^{NTR}$ –dependent motor neuron death by a small non-peptidyl mimetic of the neurotrophin loop 1 domain. *Eur J Neurosci 24*, 1575–1580.
105. Ferri CC, Moore FA, Bisby MA (1998). Effects of facial nerve injury on mouse motoneurons lacking the p75 low-affinity neurotrophin receptor. *J Neurobiol 34*, 1–9.
106. Lowry KS, Murray SS, McLean CA, Talman P, Mathers S, Lopes EC, Cheema SS (2001). A potential role for the p75 low-affinity neurotrophin receptor in spinal motor neuron degeneration in murine and human amyotrophic lateral sclerosis. *Amyotrophic Lateral Sclerosis and Other Motor Neuron Disorders 2*, 127–134.
107. Wiese S, Metzger F, Holtmann B, Sendtner M (1999). The role of $p75^{NTR}$ in modulating neurotrophin survival effects in developing motoneurons. *Eur J Neurosci 11*, 1668–1676.
108. Hardy J, Selkoe DJ (2002). The amyloid hypothesis of Alzheimer's disease: Progress and problems on the road to therapeutics. *Science 297*, 353–356.
109. Yaar M, Zhai S, Pilch PF, Doyle SM, Eisenhauer PB, Fine RE, Gilchrest BA (1997). Binding of β-amyloid to the p75 neurotrophin receptor induces apoptosis. A possible mechanism for Alzheimer's disease. *J Clin Investig 100*, 2333–2340.
110. Costantini C, Rossi F, Formaggio E, Bernardoni R, Cecconi D, Della-Bianca V (2005). Characterization of the signaling pathway downstream p75 neurotrophin receptor involved in β-amyloid peptide-dependent cell death. *J Mol Neurosci 25*, 141–156.
111. Coulson EJ (2006). Does the p75 neurotrophin receptor mediate Aβ-induced toxicity in Alzheimer's disease? *J Neurochem 98*, 654–660.
112. Jung K-M, Tan S, Landman N, Petrova K, Murray S, Lewis R, Kim PK, Kim DS, Ryu SH, Chao MV, Kim T-W (2003). Regulated intramembrane proteolysis of the p75 neurotrophin receptor modulates its association with the TrkA receptor. *J Biol Chem 278*, 42161–42169.
113. Kenchappa RS, Zampieri N, Chao MV, Barker PA, Teng HK, Hempstead BL, Carter BD (2006). Ligand-dependent cleavage of the p75 neurotrophin receptor is necessary for NRIF nuclear translocation and apoptosis in sympathetic neurons. *Neuron 50*, 219–232.
114. Podlesniy P, Kichev A, Pedraza C, Saurat J, Encinas M, Perez B, Ferrer I, Espinet C (2006). Pro-NGF from Alzheimer's disease and normal human brain displays distinctive abilities to induce processing and nuclear translocation of intracellular domain of $p75^{NTR}$ and apoptosis. *Am J Pathol 169*, 119–131.
115. Coulson EJ, May LM, Osborne SI, Reid K, Underwood CK, Meunier FA, Bartlett PF, Sah P (2008). p75 Neurotrophin receptor mediates neuronal cell death by activating GIRK channels through phosphatidylinositol 4,5-bisphosphate. *J Neurosci 28*, 315–324.
116. Underwood CK, Reid K, May LM, Bartlett PF, Coulson EJ (2008). Palmitoylation of the C-terminal fragment of $p75^{NTR}$ promotes death signaling and is required for subsequent cleavage by γ-secretase. *Mol Cell Neurosci 37*, 346–358.
117. Sotthibundhu A, Sykes AM, Fox B, Underwood CK, Thangnipon W, Coulson EJ (2008). β-$Amyloid_{1-42}$ induces neuronal death through the p75 neurotrophin receptor. *J Neurosci 28*, 3941–3946.
118. Qian MD, Zhang J, Tan X-Y, Wood A, Gill D, Cho S (2006). Novel agonist monoclonal antibodies activate TrkB receptors and demonstrate potent neurotrophic activities. *J Neurosci 26*, 9394–9403.
119. Wong ST, Henley JR, Kanning KC, Huang KH, Bothwell M, Poo MM (2002). A $p75^{NTR}$ and Nogo receptor complex mediates repulsive signaling by myelin-associated glycoprotein. *Nat Neurosci 5*, 1302–1308.
120. Zhang B, Salituro G, Szalkowski D, Li Z, Zhang Y, Royo I, Vilella D, Diez MT, Pelaez F, Ruby C, Kendall RL, Mao X, Griffin P, Calaycay Zierath JR, Heck JV, Smith RG,

Moller D (1999). Discovery of a small molecule insulin mimetic with antidiabetic activity in mice. *Science 284*, 974–977.

121. Liu K, Xu L, Szalkowski D, Li Z, Ding V, Kwei G, Huskey S, Moller DE, Heck JV, Zhang BB, Jones AB (2000). Discovery of a potent, highly selective, and orally efficacious small-molecule activator of the insulin receptor. *J Med Chem 43*, 3487–3494.
122. Wilkie N, Wingrove PB, Bilsland JG, Young L, Harper SJ, Hefti F, Ellis S, Pollack SJ (2001). The non-peptidyl fungal metabolite L-783,281 activates TRK neurotrophin receptors. *J Neurochem 78*, 1135–1145.
123. Lin B, Pittung MC, Deng L, Li Z, Liu Y, Webster JG (2007). Neuroprotection by small molecule activators of the nerve growth factor receptor. *J Pharmacol Exp Ther 322*, 59–69.
124. Skaper SD, Walsh FS (1998). Neurotrophic molecules: Strategies for designing effective therapeutic molecules in neurodegeneration. *Mol Cell Neurosci 12*, 179–193.
125. Xiao J, Liu Y (2003). Differential roles of ERK and JNK in early and late stages of neuritogenesis: A study of a novel PC12 model system. *J Neurochem 86*, 1516–1523.
126. Kaplan DR, Miller FD (2000). Neurotrophin signal transduction in the nervous system. *Curr Opin Neurobiol 10*, 381–391.
127. Huang EJ, Reichardt LF (2003). Trk receptors: Roles in neuronal signal transduction. *Annu Rev Biochem 72*, 609–642.
128. Skaper SD, Leon A, Toffano G (1989). Ganglioside function in the development and repair of the mammalian nervous system: From basic science to clinical application. *Mol Neurobiol 3*, 173–199.
129. Ferrari G, Anderson BL, Stephens RM, Kaplan DR, Greene LA (1995). Prevention of apoptotic neuronal death by GM1 ganglioside. Involvement of Trk neurotrophin receptors. *J Biol Chem 270*, 3074–3080.
130. Mutoh T, Tokuda A, Miyadai T, Hamaguchi M, Fujiki N (1995). Ganglioside GM1 binds to the Trk protein and regulates receptor function. *Proc Nat Acad Sci USA 92*, 5087–5091.
131. Rabin SJ, Mocchetti I (1995). GM1 ganglioside activates the high-affinity nerve growth factor receptor *trk*A. *J Neurochem 65*, 347–354.
132. Farooqui T, Franklin T, Pearl DK, Yates AJ (1997). Ganglioside GM1 enhances induction by nerve growth factor of a putative dimmer of TrkA. *J Neurochem 68*, 2348–2355.
133. Cuello AC, Garofalo L, Kenigsberg RL, Maysinger D (1989). Gangliosides potentiate in vivo and in vitro effects of nerve growth factor on central cholinergic neurons. *Proc Nat Acad Sci USA 86*, 2056–2060.
134. Fong TG, Vogelsberg V, Neff NH, Hadjiconstantinou M (1995). GM1 and NGF synergism on choline acetyltransferase and choline uptake in aged brain. *Neurobiol Aging 16*, 917–923.
135. Duchemin A-M, Ren Q, Mo L, Neff NH, Hadjiconstantinou M (2002). GM1 ganglioside induces phosphorylation and activation of Trk and Erk in brain. *J Neurochem 81*, 696–707.
136. Bachis A, Rabin SJ, Del Fiacco M, Mocchetti I (2002). Gangliosides prevent excitotoxicity through activation of TrkB receptor. *Neurotoxicol Res 4*, 225–234.
137. Berg MM, Sternberg DW, Parada LF, Chao MV (1992). K-252a inhibits nerve growth factor-induced Trk proto-oncogene tyrosine phosphorylation and kinase activity. *J Biol Chem 267*, 13–16.
138. Knüsel B, Kaplan DR, Winslow JW, Rosenthal A, Burton LE, Beck KD, Rabin S, Nikolics K, Hefti F (1992). K-252b selectively potentiates cellular actions and trk tyrosine phosphorylation mediated by NT-3. *J Neurochem 59*, 715–722.
139. Pollack S, Young L, Bilsland J, Wilkie N, Ellis S, Hefti F, Broughton H, Harper S (1999). The staurosporine-like compound L-753,000 (NB-506) potentiates the neurotrophic effects of neurotrophin-3 by acting selectively at the TrkA receptor. *Mol Pharmacol 56*, 185–195.
140. Lewis MA, Hunihan L, Franco D, Robertson B, Palmer J, St. Laurent DR, Balasubramanian BN, Li Y, Westphal RS (2006). Identification and characterization of compounds that potentiate NT-3-mediated Trk receptor activity. *Mol Pharmacol 69*, 1396–1404.
141. Manji HK, Drevets WC, Charney DS (2001). The cellular neurobiology of depression. *Nat Med 7*, 541–547.
142. Nestler EJ, Barrot M, DiLeone RJ, Eisch AJ, Gold SJ, Monteggia LM (2002). Neurobiology of depression. *Neuron 34*, 13–25.
143. Castrén E, Võikar V, Rantamäki T (2007). Role of neurotrophic factors in depression. *Curr Opin Pharmacol 7*, 18–21.
144. Eaton MJ, Whittemore SR (1996). Autocrine BDNF secretion enhances the survival and serotonergic differentiation of raphe neuronal precursor cells grafted into the adult rat CNS. *Exp Neurol 140*, 105–114.
145. Mamounas LA, Blue ME, Siuciak JA, Altar CA (1995). Brain-derived neurotrophic factor promotes the survival and sprouting of serotonergic axons in rat brain. *J Neurosci 15*, 7929–7939.
146. Nibuya M, Morinobu S, Duman RS (1995). Regulation of BDNF and trkB mRNA in rat brain by chronic electroconvulsive seizure and antidepressant drug treatments. *J Neurosci 15*, 7539–7547.
147. Russo-Neustadt A, Beard RC, Cotman CW (1999). Exercise, antidepressant medications, and enhanced brain derived neurotrophic factor expression. *Neuropsychopharmacology 21*, 679–682.
148. Rantamäki T, Hendolin P, Kankaanpää A, Mijatovic J, Piepponen P, Domenici E, Chao MV, Männistö PT, Castrén E (2007). Pharmacologically diverse antidepressants rapidly activate brain-derived neurotrophic factor receptor TrkB and induce phospholipase-Cγ signaling pathways in mouse brain. *Neuropsychopharmacology 32*, 2152–2162.

149. Marks AR (1996). Cellular functions of immunophilins. *Physiol Rev 76*, 631–649.
150. Liu J, Farmer JD Jr, Lane WS, Friedman J, Weissman I, Schreiber SL (1991). Calcineurin is a common target of cyclophil-cyclosporin A and FKBP-FK506 complexes. *Cell 66*, 807–815.
151. Gold BG, Katoh K, Storm-Dickerson T (1995). The immunosuppressant FK506 increases the rate of axonal regeneration in rat sciatic nerve. *J Neurosci 15*, 7509–7516.
152. Butcher SP, Henshall DC, Teramura Y, Iwasaki K, Sharkey J (1997). Neuroprotective actions of FK506 in experimental stroke: In vivo evidence against an antiexcitotoxic mechanism. *J Neurosci 17*, 6939–6946.
153. Toung TJ, Bhardwaj A, Dawson VL, Dawson TM, Traystman RJ, Hurn PD (1999). Neuroprotective FK506 does not alter in vivo nitric oxide production during ischemia and early reperfusion in rats. *Stroke 30*, 1279–1285.
154. Steiner JP, Connolly MA, Valentine HL, Hamilton GS, Dawson TM, Hester L, Snyder SH (1997). Neurotrophic actions of nonimmunosuppressive analogues of immunosuppressive drugs FK506, rapamycin and cyclosporine A. *Nat Med 3*, 421–428.
155. Steiner JP, Hamilton GS, Ross DT, Valentine HL, Guo H, Connolly MA, Liang S, Ramsey C, Li JH, Huang W, Howorth P, Soni R, Fuller M, Sauer H, Nowotnik AC, Suzdak PD (1997). Neurotrophic immunophilin ligands stimulate structural and functional recovery in neurodegenerative animal models. *Proc Nat Acad Sci USA 94*, 2019–2024.
156. Poulter MO, Payne KB, Steiner JP (2004). Neuroimmunophilins: A novel drug therapy for the reversal of neurodegenerative disease? *Neuroscience 128*, 1–6.
157. Guo X, Dawson VL, Dawson TM (2001). Neuroimmunophilin ligands exert neuroregeneration and neuroprotection in midbrain dopaminergic neurons. *Eur J Neurosci 13*, 1683–1693.
158. Guo X, Dillman JF, Dawson VL, Dawson TM (2001). Neuroimmunophilins: Novel neuroprotective and neuroregenerative targets. *Ann Neurol 50*, 6–16.
159. Zhang C, Steiner JP, Hamilton GS, Hicks TP, Poulter MO (2001). Regeneration of dopaminergic function in 6-hydroxydopamine-lesioned rats by neuroimmunophilin ligand treatment. *J Neurosci 21*, RC156 (1–6).
160. Caporello E, Nath A, Slevin J, Galey D, Hamilton G, Williams LR, Steiner JP, Haughey NJ (2006). The immunophilin ligand GPI1046 protects neurons from the lethal effects of the HIV-1 proteins gp120 and Tat by modulating endoplasmic reticulum calcium load. *J Neurochem 98*, 146–155.
161. Ganel R, Ho T, Maragakis NJ, Jackson M, Steiner JP, Rothstein JD (2006). Selective up-regulation of the glial $Na^+$-dependent glutamate transporter GLT1 by a neuroimmunophilin ligand results in neuroprotection. *Neurobiol Dis 21*, 556–567.
162. Costantini LC, Cole D, Chaturvedi P, Isacson O (2001). Immunophilin ligands can prevent progressive dopaminergic degeneration in animal models of Parkinson's disease. *Eur J Neurosci 13*, 1085–1092.
163. Bocquet A, Lorent G, Fuks B, Grimée R, Talaga P, Daliers J, Klitgaard H (2001). Failure of GPI compounds to display neurotrophic activity in vitro and in vivo. *Eur J Pharmacol 415*, 173–180.
164. Harper S, Bilsland J, Young L, Bristow L, Boyce S, Mason G, Rigby M, Hewson L, Smith D, O'Donnell R, O'Connor D, Hill RG, Evans D, Swain C, Williams B, Hefti F (1999). Analysis of the neurotrophic effects of GPI-1046 on neuronal survival and regeneration in culture and *in vivo*. *Neuroscience 88*, 257–267.
165. Emborg ME, Shin P, Roitberg B, Sramek JG, Chu Y, Stebbins GT, Hamilton JS, Suzdak PD, Steiner JP, Kordower JH (2001). Systemic administration of the immunophilin ligand GPI 1046 in MPTP-treated monkeys. *Exp Neurol 168*, 171–182.
166. Edlich F, Weiwad M, Wildemann D, Jarczowski F, Kilka S, Moutty M-C, Jahreis G, Lücke C, Schmidt W, Striggow F, Fischer G (2006). The specific FKBP38 inhibitor *N-(N′,N′*-dimethylcarboxamidomethyl)cycloheximide has potent neuroprotective and neurotrophic properties in brain ischemia. *J Biol Chem 281*, 14961–14970.
167. Staton PC, Skaper SD (2003). The mitogen-activated protein kinase superfamily: Signal transducers of neuronal survival and death. *Recent Res Dev Neurochem 6*, 17–41.
168. Shoji M, Iwakami N, Takeuchi S, Waragai M, Suzuki M, Kanazawa I, Lippa CF, Ono S, Okazawa H (2000). JNK activation is associated with intracellular β-amyloid accumulation. *Mol Brain Res 85*, 221–233.
169. Zhu X, Raina AK, Rottkamp CA, Aliev G, Perry G, Boux H, Smith MA (2001). Activation and redistribution of c-Jun N-terminal kinase/stress activated protein kinase in degenerating neurons in Alzheimer's disease. *J Neurochem 76*, 435–441.
170. Ferrer I, Blanco R, Carmona M, Puig B, Barrachina M, Gómez C, Ambrosio S (2001). Active, phosphorylation-dependent mitogen-activated protein kinase (MAPK/ERK), stress-activated protein kinase/c-Jun N-terminal kinase (SAPK/JNK), and p38 kinase expression in Parkinson's disease and dementia with Lewy bodies. *J Neural Transm 108*, 1383–1396.
171. Saporito MS, Thomas BA, Scott RW (2000). MPTP activates c-Jun $NH_2$-terminal kinase (JNK) and its upstream regulatory MKK4 in nigrostriatal neurons in vivo. *J Neurochem 75*, 1200–1208.
172. Hunot S, Vila M, Teismann P, Davis RJ, Hirsch EC, Przedborski S, Rakic P, Flavell RA (2004). JNK-mediated induction of cyclooxygenase 2 is required for neurodegeneration in a mouse model of Parkinson's disease. *Proc Nat Acad Sci USA*, *101*, 665–670.
173. Borasio GD (1990). Differential effects of the protein kinase inhibitor K-252a on the in vitro survival of chick embryonic neurons. *Neurosci Lett 108*, 207–212.

174. Glicksman MA, Forbes ME, Prantner JE, Neff NT (1995). K-252a promotes survival and choline acetyltransferase activity in striatal and basal forebrain neuronal cultures. *J Neurochem 64*, 1502–1512.

175. Kaneko M, Saito Y, Saito H, Matsumoto T, Matsuda Y, Vaught JL, Dionne CA, Angeles TS, Glicksman MA, Neff NT, Rotella DP, Kauer JC, Mallamo JP, Hudkins RL, Murakata C (1997). Neurotrophic 3,9-bis[(alkylthio) methyl]- and-bis(alkoxymethyl)-K-252a derivatives. *J Med Chem 40*, 1863–1869.

176. Maroney AC, Finn JP, Connors TJ, Durkin JT, Angeles T, Gessner G, Xu Z, Meyer SL, Savage MJ, Greene LA, Scott RW, Vaught JL (2001). CEP-1347 (KT7515), a semisynthetic inhibitor of the mixed lineage kinase family. *J Biol Chem 276*, 25302–25308.

177. Murakata C, Kaneko M, Gessner G, Angeles TS, Ator MA, O'Kane TM, McKenna BAW, Thomas BA, Mathiasen JR, Saporito MS, Bozyczko-Coyne D, Hudkins RL (2002). Mixed lineage kinase activity of indolocarbazole analogues. *Bioorg Med Chem Lett 12*, 147–150.

178. Maroney AC, Glicksman MA, Basma AN, Walton KM, Knight E Jr, Murphy CA, Bartlett BA, Finn JP, Angeles T, Matsuda Y, Neff NT, Dionne CA (1998). Motoneuron apoptosis is blocked by CEP-1347 (KT 7515), a novel inhibitor of the JNK signaling pathway. *J Neurosci 18*, 104–111.

179. Maroney AC, Finn JP, Bozyczko-Coyne D, O'Kane TM, Neff NT, Tolkovsky AM, Park DS, Yan CYI, Troy CM, Greene LA (1999). CEP-1347 (KT7515), an inhibitor of JNK activation, rescues sympathetic neurons and neuronally differentiated PC12 cells from death evoked by three distinct insults. *J Neurochem 73*, 1901–1912.

180. Mathiasen JR, McKenna BAW, Saporito MS, Ghadge GD, Roos RP, Holskin BP, Wu Z-L, Trusko SP, Connors TC, Maroney AC, Thomas BA, Thomas JC, Bozyczko-Coyne D (2004). Inhibition of mixed lineage kinase 3 attenuates $MPP^+$-induced neurotoxicity in SH-SY5Y cells. *Brain Res 1003*, 68–97.

181. Bozyczko-Coyne D, O'Kane TM, Wu Z-L, Dobrzanski P, Murthy S, Vaught JL, Scott RW (2001). CEP-1347/KT-7515, an inhibitor of SAPK/JNK pathway activation, promotes survival and blocks multiple events associated with Aβ-induced cortical neuron apoptosis. *J Neurochem 77*, 849–863.

182. Harris CA, Deshmukh M, Tsui-Pierchala B, Maroney AC, Johnson EM Jr (2002). Inhibition of the c-Jun N-terminal kinase signaling pathway by the mixed lineage kinase inhibitor CEP-1347 (KT7515) preserves metabolism and growth of trophic factor-deprived neurons. *J Neurosci 22*, 103–113.

183. Roux PP, Dorval G, Boudreau M, Angers-Loustau A, Morris SJ, Makkerh J, Barker PA (2002). K252a and CEP1347 are neuroprotective compounds that inhibit mixed-lineage kinase-3 and induce activation of Akt and ERK. *J Biol Chem 277*, 49473–49480.

184. Wang LH, Paden AJ, Johnson EM Jr (2005). Mixed-lineage kinase inhibitors require the activation of Trk receptors to maintain long-term neuronal trophism and survival. *J Pharmacol Exp Ther 312*, 1007–1019.

185. Glicksman MA, Chiu AY, Dionne CA, Harty M, Kaneko M, Murakata C, Oppenheim RW, Prevette D, Sengelaub DR, Vaught JL, Neff NT (1998). CEP-1347/KT7515 prevents motor neuronal programmed cell death and injury-induced dedifferentiation *in vivo*. *J Neurobiol 35*, 361–370.

186. Harper SJ, Saporito MS, Hewson L, Young L, Smith D, Rigby M, Jackson P, Curtis N, Swain C, Hefti F, Vaught JL, Sirinathsinghji D (2000). CEP-1347 increases ChAT activity in culture and promotes cholinergic neurone survival following fimbria-fornix lesion. *NeuroReport 11*, 2271–2276.

187. Saporito MS, Brown ER, Carswell S, DiCamillo AM, Miller MS, Murakata C, Neff NT, Vaught JL, Haun FA (1998). Preservation of cholinergic activity and prevention of neuron death by CEP-1347/KT-7515 following excitotoxic injury of the nucleus basalis magnocellularis. *Neuroscience 86*, 461–472.

188. Saporito MS, Brown EM, Miller MS, Carswell S (1999). CEP-1347/KT-7515, an inhibitor of *c-jun* N-terminal kinase activation, attenuates the 1-methyl-4-phenyl tetrahydropyridine-mediated loss of nigrostriatal dopaminergic neurons in vivo. *J Pharmacol Exp Ther 288*, 421–427.

189. Saporito MS, Hudkins RL, Maroney AC (2002). Discovery of CEP-1347/KT-7515, an inhibitor of the JNK/SAPK pathway for the treatment of neurodegenerative diseases. *Prog Med Chem 40*, 23–62.

190. Griffin WS, Sheng JG, Royston MC, Gentleman SM, McKenzie JE, Graham DI, Roberts GW, Mrak RE (1998). Glial-neuronal interactions in Alzheimer's disease: The potential role of a "cytokine cycle 2" in disease progression. *Brain Pathol 8*, 65–72.

191. Golde S, Chandran S, Brown GC, Compston A (2002). Different pathways for iNOS-mediated toxicity in vitro dependent on maturation and NMDA receptor expression. *J Neurochem 82*, 269–282.

192. Skaper SD, Facci L, Culbert A, Evans NA, Chessell I, Davis JB, Richardson JC (2006). $P2X_7$ receptors on microglial cells mediate injury to cortical neurons in vitro. *Glia 54*, 234–242.

193. Culbert AA, Skaper SD, Howlett DR, Evans NA, Facci L, Soden PE, Seymour ZM, Guillot F, Gaestel M, Richardson JC (2006). MAPKAP kinase 2 deficiency in microglia inhibits pro-inflammatory mediator release and resultant neurotoxicity: Relevance to neuroinflammation in a transgenic mouse model of Alzheimer's disease. *J Biol Chem 281*, 23658–23667.

194. Lund S, Porzgen P, Mortensen AL, Hasseldam H, Bozyczko-Coyne D, Morath S, Hartung T, Bianchi M, Ghezzi P, Bsibsi M, Dijkstra S, Leist M (2005). Inhibition of microglial inflammation by the MLK inhibitor CEP-1347. *J Neurochem 92*, 1439–1451.

195. Ciallella JR, Saporito M, Lund S, Leist M, Hasseldam H, McGann N, Smith CS, Bozyczko-Coyne D, Flood DG

(2005). CEP-11004, an inhibitor of the SAPK/JNK pathway, reduces TNF-α release from lipopolysaccharide-treated cells and mice. *Eur J Pharmacol 515*, 179–187.

196. Sendtner M, Schmalbruch H, Stöckli KA, Carroll P, Kreutzberg GW, Thoenen H (1992). Ciliary neurotrophic factor prevents degeneration of motor neurons in mouse mutant progressive motor neuronopathy. *Nature 358*, 502–504.
197. Haase G, Pettmann B, Bordet T, Villa P, Vigne E, Schmalbruch H, Kahn A (1999). Therapeutic benefit of ciliary neurotrophic factor in progressive motor neuronopathy depends on the route of delivery. *Ann Neurol 45*, 296–304.
198. Bordet T, Lesbordes J-C, Rouhani S, Castelnau-Ptakhine L, Schmalbruch H, Haase G, Kahn A (2001). Protective effects of cardiotrophin-1 adenoviral gene transfer on neuromuscular degeneration in transgenic ALS mice. *Hum Mol Genet 10*, 1925–1933.
199. Lesbordes J-C, Cifuentes-Diaz C, Miroglio A, Joshi V, Bordet T, Kahn A, Melki J (2003). Therapeutic benefits of cardiotrophin-1 gene transfer in a mouse model of spinal muscular atrophy. *Hum Mol Genet 12*, 1233–1239.
200. Apfel SC (2001). Neurotrophic factor therapy—Prospects and problems. *Clin Chem Lab Med 39*, 351–355.
201. Bordet T, Buisson B, Michaud M, Drouot C, Galéa P, Delaage P, Akentieva NP, Evers AS, Covey DF, Ostuni MA, Lacapère J-J, Massaad C, Schumacher M, Steidl E-M, Maux D, Delaage M, Henderson CE, Pruss RM (2007). Identification and characterization of cholest-4-en-3-one, oxime (TRO19622), a novel drug candidate for amyotrophic lateral sclerosis. *J Pharmacol Exp Ther 322*, 709–720.
202. Kotzbauer PT, Holtzman DM (2006). Expectations and challenges in the therapeutic use of neurotrophic factors. *Ann Neurol 59*, 444–447.
203. Zampieri N, Chao MV (2006). Mechanisms of neurotrophin receptor signalling. *Biochem Soc Trans 34*, 607–611.
204. Peleshok J, Saragovi HU (2006). Functional mimetics of neurotrophins and their receptors. *Biochem Soc Trans 34*, 612–617.
205. Pollack SJ, Harper SJ (2002). Small molecule Trk receptor agonists and other neurotrophic factor mimetics. *Curr Drug Targets CNS Neurol Disord 1*, 59–80.
206. Waldmeier P, Bozyczko-Coyne D, Williams M, Vaught JL (2006). Recent clinical failures in Parkinson's disease with apoptosis inhibitors underline the need for a paradigm shift in drug discovery for neurodegenerative diseases. *Biochem Pharmacol 72*, 1197–1206.
207. Fitzgerald GA (2005). Opinion: Anticipating change in drug development: The emerging era of translational medicine and therapeutics. *Nat Rev Drug Discov 4*, 815–818.

# ONCOLOGY

Arkadiusz Z. Dudek and Steven F. Powell
*Division of Hematology, Oncology and Transplantation, University of Minnesota, Minneapolis, Minnesota*

## 1 INTRODUCTION

Cancer is a devastating disease that is often fatal. Even with therapy, many patients will die. According to the World Health Organization (WHO), cancer was responsible for 7.8 million (13%) of 58 million deaths worldwide [1]. Due to this, it is no surprise that anticancer drug research is one of the most funded and vigorously pursued fields of research in the world today. Cancer remains a leading cause of morbidity and mortality throughout the world. In the United States, the pharmaceutical industry has exploded with the production of many new therapeutic agents. Anticancer drugs make up a great deal of these agents.

The quest for anticancer agents is fraught with difficulty. Anticancer drugs have typically been limited due to the severity and frequency of adverse effects seen in patients. *Chemotherapy*, as anticancer drugs are commonly termed, is often feared by patients. As research has progressed, we have been able to find more agents that are active against a variety of cancers. With improvement in our understanding of cancer, these agents have been designed to be less toxic to patients. Despite this, toxicity remains a major concern with anticancer agents.

The toxicity of anticancer agents can be attributed to the subtle differences in cancerous cells when compared to normal healthy cells. A cancerous or malignant cell is essentially a cell that has gone out of control. These cells grow, divide, and spread relentlessly. Their life spans are often infinite. In addition to this, they have the ability to invade other tissues and cause organ damage. Cancerous cells do not typically provide any beneficial biologic function and will replace normal healthy tissue. They aggregate as abnormal tissue, often called tumors. The spread of these tumors result in severe morbidity and mortality. The mechanisms behind the malignant cells erroneous processes often protect them from normal host defenses and chemotherapy. Anticancer agents are imperfect and cannot recognize all the subtle differences seen in malignant cells. They often can only target rapidly dividing cells. The result is damage to normal, healthy cells that reproduce more often. These are often the cells of the gastrointestinal tract, the bone marrow, and the sex organs. Additionally, some of these agents can have drastic effects on other normal healthy tissues, including the cardiovascular and neurological systems. Damaging these organ systems with drugs can cause further medical problems in already ill patients.

In addition to their ability to "hide" from the host's immune system, malignant cells have the ability to rapidly adapt to toxic conditions. For instance, a chemotherapy agent that might show activity in malignant cells at one point in time may eventually lose its ability to damage that cell type in the future. This is due to the resistance mechanisms present in malignant cells. These resistance mechanisms can be either innate or acquired. Alterations in drug transport mechanisms, drug targets, drug metabolism, DNA (deoxyribonucleic acid) repair, DNA synthesis, and apoptosis mechanisms can all contribute to resistance. A recurring problem in anticancer research is to find more agents to treat these resistant cells.

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

Thankfully, the rise of technology, specifically mathematical and chemical modeling, has been of great benefit to anticancer drug research. The progression of new technologies such as high-throughput screening, bioinformatics, genomics, and proteomics have created a surge in the understanding of cancer biology and therapy. These technologies are already being credited for helping to create a number of new promising anticancer agents. We are only at the forefront of understanding their utility in anticancer research. It can be presumed that these technologies will continue to help us produce amazing breakthroughs in anticancer drug discovery.

In addition to the newer technologies, we must not forget the standard approaches for drug discovery. This is where we find our roots in anticancer drug discovery. Areas such as natural products research have led to the development of further research avenues, such as combinatorial chemistry. Many of the drugs created from these older research methods are still quite useful today. A majority of them remain first-line agents for the treatment of a variety of cancers. These areas of research are still actively producing new compounds.

The process of anticancer drug development is quite lengthy. It can take well over 10 years for an anticancer drug to go from preclinical development to its final use in a clinical setting [2]. It is therefore important to have a basic understanding of the preclinical and clinical testing that is performed to develop an anticancer agent. There are over 100 agents approved for the treatment of various cancers worldwide [3]. Hundreds more are in the stages of clinical and preclinical testing and may never make it to a clinical setting.

This chapter is meant to give a basic understanding of the methods of anticancer drug research. This will be achieved by giving a background of where anticancer drug discovery has come from to where it is going. From natural products research to DNA microarrays, all facets of anticancer drug research will be explored. Additionally, the current classes of anticancer agents will be explained. Significant detail will be given to the newer agents that are being developed. Future directions will also be highlighted throughout the chapter. In the end, this chapter is designed to stimulate thought among researchers interested in anticancer drug discovery. With more insight into the treatment of such a devastating disease, we can hope to improve the quality of life for millions.

## 2 NATURAL PRODUCTS

### 2.1 Introduction

The development of drugs from sources in nature is the oldest method of drug discovery. Drugs such as digoxin, penicillin, and even aspirin trace their roots to natural sources. It is no surprise that many chemotherapeutic drugs have been derived from naturally existing sources. It is estimated that at least two-thirds of the anticancer agents currently in use are either directly isolated from natural sources or are a semisythetic version of natural compounds [4]. The main question pertaining to these compounds is aimed at why anticancer compounds exist in nature. Perhaps the simplest explanation to this is evolution.

Over the course of evolution many organisms develop adaptations that give them advantages over their competitors over different stresses in the ecosystem. These adaptations are usually the development of chemical and biochemical compounds. On a cellular and molecular level these compounds can often include mechanisms that cause cell cycle arrest, programmed cell death, and even direct toxicity. These intracellular systems may allow an organism to have a survival advantage against predators or make them a more effective predator.

By understanding these basic principles, we have seen the benefit in the field of oncology. The ability to isolate active compounds has improved as our skills in organic chemistry and biochemistry have evolved. Our understanding of the cell cycle further enhances our ability to understand the utility of natural products. Despite the promising findings in natural product chemotherapeutics, there is a great deal of difficulty and complexity to this field.

### 2.2 Methods of Discovery

The easiest way to think of the natural product drug discovery process is to view the process from start to finish. The first step in natural product research is taken by finding an organism of interest. This organism is then harvested in fairly large amounts. Once an adequate amount of material is harvested, a crude large-scale extraction is performed on the organism. This extraction is performed by using various different organic solvents to extract hundreds to thousands of compounds from the organism. These crude extractions are then tested for activity on cancer cell lines. If promising activity is noted with one particular extract, further studies will be done on this. Using modern organic chemistry techniques such as high-performance liquid chromatography (HPLC), nuclear magnetic resonance (NMR), mass spectroscopy, and infrared spectroscopy the exact chemical composition of the active compound or compounds from a given extract can be ascertained. Once the structure and chemical composition of a possible drug is known, steps toward synthetic or semisynthetic development of a drug can be pursued.

One factor that helps determine which organism to use is in regards to supply. Due to the large amount of biologic material that is needed for experimentation and eventually clinical use, one must be selective in choosing an organism that is easily accessible and is in adequate supply. An excellent example of this is the anticancer agent paclitaxel (Taxol). Paclitaxel was initially found from the bark of the Pacific yew tree. The drug was initially found in 1967 [5]. From its discovery in 1967 until 1993 almost all paclitaxel used in research and treatment was derived from the bark of the Pacific yew tree (*Taxus brevifolia*). To create a single dose of paclitaxel, a proportionally large amount of bark was needed due to very poor yield. Due to the large amount of bark needed, there was a great deal of concern about the number of yew trees being harvested [6]. The tree, of course, took many years to reach a point at which it could be utilized to harvest the drug. Even with planting new trees, the demand was starting to outweigh the supply. It was feared that the yew tree would eventually become extinct. Due to this concern, a great deal of work was done to try and develop a way to synthetically or semisynthetically make paclitaxel. Initial efforts were thwarted by inefficiency and impractical methods; however, in 1992 a feasible method was found. A group led by Robert Holten eventually found that by using the needles of the *Taxus baccata*, a yew tree related to *T. breviolia*, one could semisynthetically make paclitaxel with an 80% yield [7].

Ethics also plays a role in the field of natural product drug discovery. The more intelligent a life-form is, the less likely it is that its use will be accepted. For instance, the majority of drugs are found from plant, simple marine, and microbial sources. Concerns may arise when a more complex life-form is investigated for drug discovery. Destroying intelligent life for the pursuit of drug development creates an ethical dilemma. At what point do we determine an organism should not be utilized? This is a question that researchers should be pondering when pursuing a possible new source for a natural product. Other ethical dilemmas lie in areas such as habitat destruction and annihilation of certain species.

The number of organisms on this planet is vast, therefore making it difficult to understand which organisms will produce active compounds. Due to the vast array of organisms on this planet, searching for anticancer drugs can sometimes seem like "looking for a needle in a haystack." In order to decide what organisms to pursue for studies, the general approach has been to look at organisms that have historically had activity. For this reason, the majority of anticancer agents have been found from marine, plant, and microbial sources.

### 2.3 Marine Sources

Ocean's cover about 70% of the Earth's surface, which makes it obvious that a large majority of life on this planet can be found under the sea. The biodiversity of these life-forms creates great potential for finding anticancer agents. As we have become more adept at deep-sea diving and ocean exploration, finding these organisms has become much easier. Additionally, advances in aquaculture technology has helped us to cultivate the amount of biomass needed for research.

The first marine anticancer agents discovered were spongouridine and spongothymidine. These were derived from the sea sponge *Cryptothethya crypta* [8, 9]. From these agents the commonly used drug, cytosine arabinoside, was synthesized. Further expanding on sea sponges and other related sea life, many exciting drugs have been found. The bryostatins are one class of drug that were isolated from the marine bryozoan *Bugula neritina*. This group of drug is noted as having protein kinase C activity [10]. They have shown promise as single agents but are most notably showing a great deal of activity as synergistic agents. At this point, bryostatins are in phase II clinical trials [11, 12].

Other agents derived from sea life include the ecteinascidins isolated from the Caribbean sea squirt, dolastatin [13] from the Indian Ocean sea hare *Dolabella auricularia,* didemnin B [14] from the tunicate *Trididemnum solidum*, aplidine [15] from the Mediterranean tunicate *Aplidium albicans,* and kahalalide F [16]from the Hawaiian mollusk *Elysia rufescens*. All of these agents have shown activity in preclinical studies and are now in various stages of clinical trials. Hundreds of other agents are now being studied in preclinical trials [17–20].

### 2.4 Plant Sources

As previously mentioned, paclitaxel is a famous example of an anticancer drug developed from a plant source. Interestingly enough, various plants have been known to play roles in cultural and tribal remedies. Many of these plants that have been used in traditional remedies have actually been shown to have pharmacologically active agents when they are purified. An example of this is the Madagascar periwinkle *Cathranthus roseus*. This plant was used as a traditional remedy for diabetes in several cultures. When the plant was researched to find out if there were any agents that could be utilized for diabetes, an interesting discovery was made. The vinca alkaloids, vinblastine and vincristine, were found instead. These drugs play a vital role in the field of oncology to this day [21].

Some drugs from plant and other sources have only weak activity or may have severe side effects that

limit their use clinically. This is why semisynthetic derivatives are sometimes pursued. The camptothecins are an excellent example of this. The Chinese ornamental tree *Camptotheca acuminate* produced the camptothecins [22]. They were noted to have good activity in cancer cell lines; unfortunately, they had a great deal of systemic toxicity. Semisynthetic derivatives were created and the two agents topotecan and irinotecan were born. These two agents are strong inhibitors of DNA topoisomerase I. They have been shown to have exceptional activity in a wide variety of tumor types. Due to this, further work is currently being done to find more semisynthetic derivatives of this class.

Multiple other plant-derived products are currently being investigated. Homoharringtonine is a plant alkaloid that is isolated from the Chinese tree *Cephalotaxus harringtonia*. This agent has good activity in a variety of hematologic malignancies and is in preclinical studies [23]. Elliptinium is a DNA topoisomerase II inhibitor and DNA intercalater found from the Fijian medicinal plant *Bleekeria vitensis*. This agent has been shown to have preclinical activity in metastatic breast cancer. Unfortunately, systemic toxicities are limiting its use in clinical trials [24]. The combretastatins are a group of drugs isolated from the African bush willow *Combretum caffrum*. They are thought to work by inhibiting tubulin polymerization, which causes rapid vascular shutdown vial endothelial cell apoptosis. The combretastatins are currently in clinical trials [25].

### 2.5 Microbial Sources

Microbial organisms have provided us with a great deal of drugs. Perhaps penicillin is the most well-known example throughout medicine. Oncology too has benefited from microbial sources of drugs. Many chemotherapeutics including the anthracyclines (doxorubicine, daunorubicin, epirubicin, and idarubicin), bleomycin, actinomycin, and mitomycin are from microbial sources. Microbial organisms are an excellent source for many reasons. One of the major reasons is that they reproduce rather rapidly. This allows for a great deal of drug to be available for research and clinical use. The need for synthetic drugs from these sources is much less than with slower growing organisms in the plant and animal kingdoms.

Epothilone A and B are two drugs from the myxobacterium *Sorangium cellulosssum*, which were initially investigated as antifungals but now are shown to have activity in human cancer cell lines. This class of drug works in a similar manner to paclitaxel and is in preclinical and clinical studies at this time [26].

### 2.6 Conclusion

Natural products have been present in anticancer drug research since the beginning. As we enter the era of rationally designed drugs, there is a question as to what role natural products will play. It is obvious that there could still be thousands of drugs out there in nature. As long as researchers continue to pursue this area of drug discovery, there will be new drugs developed from natural sources. Perhaps the most exciting concept to evolve in natural product drug discovery is the seemingly unending ability of chemists and biochemists to create semisythetic and synthetic versions of these drugs, which will eventually eliminate the need to destroy a large amount of biologic material to treat cancer.

## 3 COMBINATORIAL CHEMISTRY

Combinatorial chemistry is often referred to as one of the greatest advances in medicinal chemistry. Since its development, combinatorial chemistry has shown promise in the field of oncology. It has drastically altered the method in which anticancer compounds are synthesized. Combinatorial chemistry essentially involves the rapid synthesis or computer simulation of a large number of different but structurally related molecules.

The traditional approach in medicinal chemistry involved creating structurally related compounds individually, and then eventually sets of analogs would build up serially. When the analogs were developed in this way, they would require individualized reaction sequences and purification methods. This was very labor intensive and costly. Combinatorial chemistry on the other hand, works by using a defined pathway of bond-forming reactions to combine sets of simple molecules or building blocks to give complex products in a small number of steps. This can be easily repeated over and over again, thus allowing for libraries or large numbers of analogs that can be created in an organized and efficient manner. This repetitiveness allows for automation, which in turn allows for a format that is amenable to high-throughput screening.

Combinatorial chemistry was initially developed in the 1960s by R. B. Merrifield. He initiated the concept when he was studying the synthesis of peptides. Merrifield developed a concept called solid-phase synthesis (SPS) [27]. SPS is essentially a method in which molecules are bound on a bead and synthesized step-by-step in a reactant solution. As technology advanced and computers became more readily available, this process has become "digitalized."

The pharmaceutical industry has utilized combinatorial chemistry to develop hundreds of thousands of potential compounds per year. When an active compound is found, combinatorial chemistry can be used to optimize the compound, thus developing many different, but related compounds. These digital compounds are placed in a *virtual library*, which can contain thousands to millions of different compounds. Using other technologies such as quantitative structure-activity relationship (QSAR), these compounds are screened and the most promising subsets will be synthesized. Currently, there are five types of combinatorial chemistry libraries, which include biological libraries, spatially addressable parallel solid-phase or solution-phase libraries, synthetic library methods requiring deconvolution, the "one-bead one-compound" library method, and synthetic library methods using affinity chromatography selection [28].

### 3.1 Combinatorial Chemistry in Anticancer Drug Discovery

Combinatorial chemistry is currently being used in the discovery of many different anticancer drug classes. They have been shown to have utility in expanding upon drugs derived from natural products [29]. More recently, combinatorial chemistry has shown its effectiveness in investigating small-molecule protein kinase inhibitors. This includes the cyclin-dependent kinase inhibitors, the protein kinase C inhibitors, matrix metalloproteinase inhibitors, protein phophatase inhibitors, and tyrosine kinase inhibitors.

The cyclin-dependent kinase (CDK) family of enzymes are integral in controlling the cell cycle progression. They work primarily at the cell cycle checkpoints. In cancerous cells they result in aberrant cell cycle control. This may stem from a tumor suppressor gene malfunction or from the activation of oncogenes. Tumors with abnormal CDKs will have increased CDK activity. The majority of CDK inhibitors currently used inhibit the adenosine triphosphate (ATP)-binding domain of the CDKs. The active component of small-molecule CDK inhibitors usually involves the purine pharmacopore. Combinatorial chemistry is currently being used to create libaries of compounds with variations on the purine pharmacopore [30].

Protein kinase C (PKC) is another kinase involved in the regulation of cell growth and signaling. It involves a family of at least 12 closely related serine–threonine kinase isoforms. They have been a quite interesting target for targeted anticancer agents. The development of clinically useful PKC inhibitors has proven to be more difficult than expected due to the complex nature of the PKC signal transduction pathways and the lack of isoform-specific modulators of PKC activity. One example of this is diacylglycerol (DAG). DAG is a second messenger that activates and recruits some PKC family members to the plasma membrane. DAG has been exploited using combinatorial chemistry to create a library of 16 members that are all derived from the active DAG enantiomer [31].

Inhibitors of DNA-dependent protein kinases (DNA-PK) are yet another group of kinases that play a role in cell development. They are serine–threonine protein kinases that work by detecting and repairing DNA double-strand breaks. By inhibiting these kinases, one could allow for radiosensitization and chemosensitization. Combinatorial chemistry is currently being used on one class of these inhibitors, the thiopyranones. Work with this class using combinatorial chemistry has been able to produce an analog that is one of the most potent DNA-PK inhibitors to date [32].

Tyrosine kinases are extremely popular targets for inhibition in anticancer therapy. A large number of small molecules are currently in development in this class. The major target of inhibition is the ATP-binding site of the tyrosine kinase. Combinatorial chemistry is credited as helping develop the multikinase inhibitor, sorafenib [33]. Further work is being done using combinatorial chemistry to find analogs of various known active tyrosine kinase inhibitors.

With the rise of computerized technologies, combinatorial chemistry has been born. It is now a firmly established component of drug discovery, including the field of anticancer drug discovery. Combinatorial chemistry allows us to expand further on what is known. This, in turn, affords the opportunity to expand one drug into many, some of which could be more effective than the initial compound. There is no doubt that combinatorial chemistry will continue to play a vital role in anticancer drug discovery.

## 4 HIGH-THROUGHPUT SCREENING

As anticancer drug discovery has progressed, the number of new drugs available to the research pipeline is enormous. Due to the volume of new drugs, there has been a push to increase the efficiency of the drug development process so that only drugs with the best efficacy with fewer side effects are ultimately approved for use in patients. High-throughput screening (HTS) is one technology that has aided in this quest.

### 4.1 High-Throughput Screening

High-throughput screening refers to the automated method for rapidly evaluating thousands of chemical

compounds. This method has evolved a great deal as technology has progressed. HTS essentially started out as a more efficient and rapid method of doing regular laboratory assays in bulk. The assays available at that time were limited in their throughput and output. This meant that it could take months to complete the assays. Eventually, major changes in assay design, data capture, data analysis, and quality control were discovered. As this occurred, HTS also advanced by integrating these technologies with robotics to allow for improved automation. As a result, HTS became more systematic and industrialized. The current HTS systems involve similar concepts including robotics, incubators, liquid dispensing mechanisms, sensors, data collection software, and data analysis software [34].

The basic principle of HTS is to start with a library of compounds and systematically test a large number of these on specific assays to determine if they have a desirable effect. Once a compound is noted as having the desired effect (a *hit*), it will move on through a battery of more testing. If a hit passes these tests, it will be a *lead compound*, which will go on for more detailed experiments.

The easiest way to understand the current methods of HTS is to think about each crucial component of the process. These components include the compounds being tested, the drug targets, assay design, reagent design, the screening method, quality control methods, data analysis, and the process from which hit compounds are progressed to lead compounds.

### 4.2 Compounds

As previously mentioned in other sections, there is a vast supply of potential anticancer agents available in various stages of development. HTS is currently being used to evaluate compounds derived from natural products, medicinal chemistry, and combinatorial chemistry.

Natural products have long been researched as potential anticancer agents. There are already a large number of anticancer agents that have been derived from natural products that are currently being used clinically. Since more technologically sophisticated methods of developing drugs have evolved, the use of HTS in this field has been dwindling. HTS is still currently being used in this field to some extent. The best example of this is the process of prefractionation of extracts [35]. As mentioned previously, extracts from natural products are obtained to be tested on cancer cell lines.

These extracts have been quite crude and contain hundreds to thousands of potentially active compounds. Prefractionation is an HTS method being used to reduce the time taken after initial screening to help identify the active components in a crude extract. Several procedures have already been established and range from solvent/solvent extraction to chromatographic procedures (adsorption chromatography, high-performance liquid chromatography [HPLC], medium pressure liquid chromatography [MPLC]). These typically have had a low level of automation. One of the first commercially available automated methods for the extraction and separation of plant material is the HPLC-based workstation, Sepbox. This automated method incorporated efficient fractionation of crude extracts to work toward obtaining pure compounds. Another newer method of prefractionation that has been developed works by using solid-phase extraction (SPE). This works by mixing the organic extract with an initial solvent and running it through solid-phase extraction using adsorptive resins. The product of this is then fractionated into several samples, and they are subsequently mixed with an organic solvent and ran through another SPE step. The samples are again fractionated and the process can be repeated as often as desired. Ultimately, the final compound is a more pure concentration of the desired fraction of the initial crude extract.

Combinatorial chemistry is a relatively new concept that has used technology to create vast libraries of potential anticancer agents. The vast number of these drugs are often called *libraries*. These libraries are essentially compound collections developed by researchers that may be potentially tested to assess their anticancer properties. Combinatorial chemistry has exponentially increased the number of potential anticancer agents in preclinical research. This has proven to be both a boon and a bane, however.

The vast number of compounds make it difficult to decide which agents should be tested first and what agents may be the most promising. We have been very limited in testing these compounds in the past due to our somewhat slow standard preclinical testing techniques. Perhaps this is why HTS has proven to be so helpful. HTS allows for rapid testing of thousands of compounds at once. The most common approach, sometimes called *full-deck screening*, has been to test an entire library of compounds at once using automated methods. This method is good as all potential agents are tested, however, it is both costly and more time consuming. An alternative to this has been to use *diversity decks*, which are more focused libraries that reorganize the full deck to format the compounds into chemically related groups [36]. These groups are then tested separately from the entire library. The more active groups are then tested further, rather than doing further testing using the entire library. These methods differ and have varying advantages and disadvantages. Several factors have to be considered when utilizing these methods of HTS and include cost, time, and the variability of the

compound library. Regardless of this, it is obvious that HTS has proven to be very helpful in deriving compounds through combinatorial chemistry.

The final group of compounds that utilize HTS are compounds present in medicinal chemistry collections [37]. These collections are essentially libraries of compounds that have been collected over the years from pharmaceutical research. When a promising anticancer agent is discovered, there are often a number of other, similar compounds that are discovered along the way. Because time and money do not allow for testing of all of these compounds, they are often stored in medicinal chemistry libraries. By using HTS methods similar to those used in combinatorial chemistry, new lead compounds are being developed from these libraries.

### 4.3 Target Sources

As our understanding of molecular biology and biochemistry has evolved, a number of cellular targets have been identified as targets for anticancer agents [38]. Using this knowledge, we can apply HTS techniques to help screen agents that will focus on these targets. These targets include tyrosine kinases, matrix metalloproteinases, DNA repair enzymes, proteases, G-coupled proteins, and many others. By using HTS methods, one can essentially evaluate libraries of compounds in regards to their activity on one of these specific targets. Using automation, one can screen in a relatively rapid manner. This method of study is quite different from traditional techniques in which a potential compound was evaluated to see if it had activity in cancer cell lines and the mechanism of action was determined later. Using target sources, one can find drugs that have the desired mechanism of action first, then these agents can be tested on cell lines for overall activity later on.

### 4.4 Assay Design

HTS incorporates the use of specific assays. These assays essentially are the backbone of the experiments. The assays are the environment and conditions in which the reagents will be tested. In HTS, these assays must be small, efficient, and easily reproducible. The easiest way to think of these assays is in regards of their two major components, screen construction and screen validation [39]. Screen construction creates the basis of the experiment. It sets up the optimal conditions and the constraints of the experiment being done in the assay. These components all pertain to the ultimate goal of the assay. Kinetic parameters are also set during screen construction. More importantly in HTS, screen construction must incorporate the concept of automation to help create a more streamlined and efficient backbone for experimentation.

### 4.5 Screen Validation

HTS essentially involves the "amplification" of lower throughput methods. In essence this means if one was to use a small number of plates to run experiments using low-throughput methods, in high-throughput methods, this number is amplified to thousands of plates. With this amplification, the margin of error can also increase a great deal. When designing HTS systems, screen validation is therefore an integral component. Screen validation ensures that the assay or assays will work within the constraints of the experiment. It also ensures that the experiment will be easily reproducible and predictable. This step is often done prior to initial experimentation and will also be ongoing throughout the experiment. If a large amount of variability or "noise" is noted in the experimentation, the assay may need to be redesigned.

### 4.6 Reagent Design

The reagents are the integral ingredients of the experiment. In HTS it is important to consider the types of reagents being used when designing the HTS system. Solubility, compatibility with the assays, wear on the machinery, use of radiolabeled compounds, and storage are all important factors to consider when using an HTS system. For instance, if a specific compound is relatively unstable and requires special storage, this may prove to be difficult with some HTS systems and may increase the cost of experimentation. Another factor to consider is the ease of creating large amounts of the reagent. If it takes a great deal of time to create the reagents, it may not be worthwhile to use an HTS system, as the reagents themselves may be the rate-limiting step. Likewise, one must consider the amount of reagent needed for the HTS system. If it is not feasible to have the amount needed for a specific HTS system, it may be useful to scale down assays to accommodate the reagents. Although HTS has made screening methods much easier with automation, it can be difficult if the wrong types of reagents are used.

### 4.7 Quality Control

As mentioned before, HTS systems can increase the amount of variability and error in experimentation. This is sometimes referred to as noise. Many different factors contribute to this noise. These factors include compound variability, kinetics, stability of the compound and the assay, and assay variability. When designing HTS systems, assay design, screen validation, and reagent design are helpful in minimizing this noise; however, there is always a need for quality control. Quality control is often done throughout the experimentation process. It provides a

method to quantify and record the variations in the experiment. This can be done through a variety of methods and is not always entirely computerized. For example, an experimenter may analyze the results of daily experimentation, and, if there is a degree of variability that is greater than desired, the experimentation may be halted and steps to decrease the variability may be taken.

### 4.8 Data Analysis

As HTS is becoming more efficient and productive, the volume of data produced is becoming astronomical. It is, therefore, very important to have a similarly efficient and capable data analysis system. As computer hardware and software have improved, this step has become much easier [40]. The hardware must be capable of performing processes at high speeds and must have adequate memory to store the data. The software must be able to perform all the data analysis operations, including recording, basic calculations, statistics calculations, data visualization, and error. When these calculations are complete, one can surmise the activity of the various compounds being tested and compare their activity to other compounds. The ultimate goal of data analysis in HTS is to be streamlined and efficient, while at the same time allowing the researchers to determine which of the screened compounds should move on for further testing. An error in data analysis can easily make the rest of the HTS system invalid.

### 4.9 Hits to Leads

As the HTS system comes to an end, the desired result is the production of hit compounds. These compounds have essentially fit the desired parameters in the screen and will need to be tested further. After further testing, they will be termed lead compounds [41]. In order to become a lead compound multiple batteries of tests will be performed. These can be automated in a manner similar to the prior HTS system or they can be more manual. The tests can also be hypothetical and will not require further bench research at all. Nonetheless, several different factors must be evaluated for a hit compound to become a lead compound. These factors include things such as cost of production, availability of the drug, solubility, the delivery vehicle, patentability, and other special considerations. The ultimate goal of this further testing is to find the most promising overall compounds in a given library of compounds.

HTS is a very complex and extensive process. The design of the system is perhaps the most important aspect of HTS. As technology continues to progress, HTS systems will evolve into systems that are more efficient, more complex, and will have decreased variability. They are an integral component of anticancer drug discovery and will continue to be important as potential drug libraries continue to expand.

## 5 BIOINFORMATICS

Computer technology is playing a vital role in anticancer drug development. Perhaps the use of computers is the most obvious in the fields of bioinformatics and computational biology. Bioinformatics is a field that melds biological sciences (genetics, cellular biology, systems biology, chemistry, and biochemistry) with computer sciences (applied mathematics, statistics, artificial intelligence, modeling, and informatics). This field essentially started as a basic adjunct to standard chemical and biochemical experiments. It rapidly evolved when it was utilized in the areas of combinatorial chemistry and high-throughput screening. By applying the expansive computer technologies of today, the utility of bioinformatics has become substantial. The basic premise of bioinformatics is to use computer technology to solve chemical and biochemical problems. Examples of this include using computers in analyzing and predicting RNA (ribonucleic acid) and DNA sequences [42], analyzing and predicting gene expression, identifying possible new drug targets, and analyzing protein structure and intracellular location [43]. Its utility in solving these types of problems has led to bioinformatics to be used in the drug discovery field.

Bioinformatics essentially allows for virtual experimentation, sometimes called in silico experimentation. This type of experimentation is beneficial over standard "bench" research as it does not require the materials and time necessary in bench research. By applying known scientific principles together in a virtual environment, one can predict possible outcomes. The first step in using bioinformatics in anticancer drug discovery is to utilize computers to sort through the enormous amount of cancer research data. This concept is termed *data mining* [44]. The data of interest includes information about malignant and normal cellular genetics, protein makeup, signaling pathways, response to chemotherapeutic agents, and other biochemical characteristics.

Bioinformatics can be divided into two different categories, floating-point problems and integer problems [45]. Floating-point problems usually involve using complex algorithms that utilize chemistry and advanced physics techniques. Kinetics, protein structure and formation, metabolism, and molecular dynamics are common problems solved utilizing floating-point techniques. They can be focused on optimizing high-

throughput screening as they replace the actual "physical" chemistry steps that are needed to develop lead compounds. These methods are constantly evolving. The current hope is that in the near future these techniques will be able to perfectly model the processes in malignant cells down to the molecular level. By modeling these processes, in silico experimentation can be performed and the activity of different compounds can be predicted. Anticancer drug discovery could essentially be virtualized.

Integer problems are quite different than floating-point problems. This technology is typically algorithm based. The algorithms compare matching phrases and terms or characters in sequences. This makes it an obvious choice for solving genome sequencing problems as well as a host of other molecular biology problems. These processes can typically be done over high speeds, which is quite desirable in the field of genomics. Integer problem solving using bioinformatic technology was very useful in aiding in the discovery of the human genome.

Data mining is another technique included in the repertoire of bioinformatics. This works by combining statistical analysis with pattern discovery techniques. Data mining uses these methods to investigate vast libraries of data. It allows for the detection of intricate and subtle similarities between items that were previously thought to be unrelated [46]. This can be quite useful, particularly in clinical research. Factors such as patient demographics, treatment response to different drugs, genetic abnormalities of their cancer, and other items from the clinical record can be evaluated and similarities or differences can be unveiled and exploited. This is proving to be most useful in evaluating patients with specific gene expression profiles and their response to different treatments. By finding similar gene expression profiles and similarities in response to treatments, further research can be done to see why this is so on a molecular level. The best example of the utility of data mining can be seen in patients with HER2/neu positive breast cancer [47]. Trastuzumab is an agent that is used in patients with HER2/neu overexpression. HER2/neu is an oncogene that is overexpressed in some patients with breast cancer. Trazstuzumab targets this abnormal oncogene and has provided to be a therapeutic asset in this subgroup of patients. Unfortunately, HER2/neu overexpression is not that straightforward. Different groups of patients may have a greater or lesser degree of overexpression. Overexpression of HER2/neu requires a number of intracellular molecular processes. By using data mining to understand these processes better at a genetic and molecular basis, patient response can be better predicted and dosing of the medication can be better approximated. Additionally, clinicians may be able to decide whether or not the patient will need an additional chemotherapeutic agent based on data mining information.

There is another useful bioinformatic concept that involves the systematic study of various interactions in a biologic system, called systems biology [48]. Systems biology can be utilized in molecular biology to read and record the mechanisms of gene expression and regulation in order to model phenotypic expression. This modeling can essentially be used to predict phenotypes of certain genes. By utilizing this technology, it is possible to model the expression of oncogenes in malignant cells. By modeling these genes and understanding their phenotypes, new drugs can be developed.

Ultimately it is important to remember that anticancer drug discovery is based on the principles of finding drugs that target malignant cells. These cells have subtle differences when compared to normal healthy cells. By exploiting these differences, better drugs can be developed. Bioinformatics is useful in this aspect as one can simulate these differences in a computerized environment through in silico experimentation. It should also be known that many of the genetic and molecular differences in various cancers can vary in each individual. This can mean that the same treatment in two patients with the same type of cancer can have two very different effects. The ultimate goal of bioinformatics is to incoporate the principles of in silico experimentation with data mining capabilities and standard experimentation to essentially tailor treatments to an individual, based on their genetic and molecular uniqueness.

## 6 GENOMICS AND PROTEOMICS

Our understanding of the molecular biology of cancer has progressed dramatically over the past decade. As we grow to understand more about the genetics of malignant cells, we are now learning how to apply this to create new chemotherapeutic agents that target these genetic differences. Being able to target specific genetic abnormalities in cancer cells allows us the ability to tailor therapies that minimize systemic toxicity. The fields of genomics and proteomics are what allow us to realize these therapies.

Genomics is essentially the study of the function and interactions of genes in an organism's genome [49]. For centuries, it has been understood that there are diseases that seem to run in families. This knowledge has progressed over the years and in the 1980s we began to understand that mutated genes were responsible for some disease processes. In the field of oncology, a number of such genetic mutations have been identified [50]. As we identify these abnormalities, it is human

nature to investigate methods to more easily identify these mutations and ultimately use this knowledge for treatment. Perhaps this is why genomics has provided researchers with a plethora of information for the development of new drugs.

Proteomics is similar to genomics and is essentially the study of proteins [51]. This is based on better understanding of protein structure, pathways, and function. In cells, genes code for specific proteins that will often be used in certain intra- and extracellular processes. Proteomics provides additional understanding on how cellular processes occur. When combined with genomics, proteomics can provide detailed information on how cells grow, divide, differentiate, and die.

As our genomic and proteomic knowledge has progressed, researchers have seen the enormous potential in the field of anticancer drug discovery [52, 53]. The goal of anticancer research has been to develop treatments that are toxic to malignant cells, while sparing normal cells. By understanding the genetic basis of cancer, we are able to pinpoint the problems in malignant cells. After pinpointing the abnormal genes and subsequent proteins that they code for, one can try to tailor therapies to target these unique proteins. This method allows for the selective eradication of malignant cells.

Better understanding of cancer genomics and proteomics can give researchers and clinicians more information on the somatic genetic mutations present in specific types of cancer. Additionally, the mutant proteins produced from these gene mutations can be identified and their aberrant processes understood. Understanding these changes allows for better understanding of molecular abnormalities in these cells. This information can provide data by explaining the changes in growth, invasiveness, metastatic potential, drug resistance, and drug responsiveness in different tumor types.

### 6.1 DNA Microarrays

DNA is often termed the "building block" of life. This is evident in the growth and development of cancerous cells. Cancerous cells often come into being due to some error in cellular DNA. Whether this is an error in DNA repair or a de novo mutation, the end result can be abnormal cell growth and proliferation. As scientists have begun to understand the human genome and the proteins created from normal genes, they have also begun to understand abnormal genes and proteins.

A number of different genes and proteins can cause a cell to become cancerous. For example, proto-oncogenes are normal genes that can end up becoming oncogenes. This can occur through a variety of methods but is often due to mutations or abnormal expression of proto-oncogenes. The proto-oncogene then codes for proteins that regulate cell growth, differentiation, proliferation, and apoptosis. When these proto-oncogenes and their corresponding oncogenes are abnormally expressed, cell growth becomes abnormal. This abnormal growth leads to malformed tissues that do not provide the desired biologic function. These cells can in turn have an innate "immortality" and do not usually die like regular cells. This ends up in the development of cancerous cells.

Understanding these abnormal processes is just one step in the quest for anticancer agents. The next step is to determine how to stop these processes. As protein kinase inhibitors are becoming more popular in the treatment of cancer, we are finally able to see the benefits of this approach. The major problem remains. How do we know which types of cancerous cells will respond to what targeted agents?

DNA microarrays or "gene chips" have arisen as a means to try to solve this problem. This technology has been developed as a product of HTS and bioinformatics combined with the concepts of genomics and proteomics. These microarrays are a relatively inexpensive and flexible way to evaluate the gene expression in different cell types. Shortly after microarrays were developed, their use in anticancer research was realized [54].

A DNA microarray essentially consists of a solid surface with multiple nucleic acid polymers embedded on the surface. These nucleic acid polymers are essentially single-stranded gene fragments, which may include known or predicted variants and possible genetic polymorphisms [55]. The solid surface serves as a backbone and is usually made of silicone or glass, similar to microchips. This solid surface holds the nucleic acid polymers with a chemical matrix.

There are a number of different types of microarrays available. It is too lengthy to explain all of them in detail, but the same basic idea reigns supreme in each of these microarrays. These systems essentially work to utilize their nucleic acid polymers as probes to bind labeled free targets derived from a biologic sample. In cancer research, this sample is typically the complementary DNA (cDNA) of a malignant cell. The cDNA is essentially the synthesized DNA from the cell's messenger RNA via the action of the reverse transcriptase enzyme. The resulting cDNA is labeled so that it can be followed throughout the microarray process. This cDNA selectively binds to its corresponding nucleic acid polymer. The resulting product then gives off a signal, which is usually fluorescent in nature. This allows the product to be read by the microarray's sensing mechanism. This is then analyzed and recorded by the microarray's corresponding software. Further statistical analysis

can then be done and the final readout of the genetic expression of these cells can be realized.

The microarrays are designed to be simple to use. They can facilitate gene expression measurements for thousands of different genes at the same time. They are easily reproducible and can measure a number of genes expressed in parallel. Their corresponding software can be used to make a multitude of complex calculations and solve multiple statistical questions. The result of a DNA microanalysis is data regarding the genetic expression of specific cell types. When this is applied to malignant cells, the resulting data gives insight on the genetic expression of that type of malignant cells.

Years ago, it may have been useless to understand the genetic expression of malignant cells. However, with our knowledge of the normal human genome, these differences can be exploited. For instance, if a specific type of cancer cell overexpresses a certain genetic polymorphism, the resulting protein that is coded for can be further evaluated. This abnormal protein may not cause any specific problems, but it may show significant differences from normal, healthy cells. On the other hand, the abnormal products of this overexpressed gene can lead to the major driving force for the cancerous cell's ability to grow, divide, invade, and live on relentlessly. By looking at the abnormal products of the overexpressed genes, one can find potential targets for anticancer therapies.

In addition to finding new targets for anticancer therapies, DNA microarrays can help researchers to better understand drug resistance [56]. Drug resistance is quite complex and can be present innately or acquired. Either way, drug resistance poses a huge problem in the treatment of cancer. A tumor that may be sensitive to a specific agent may eventually become resistant to that agent and many other agents. DNA microarrays are used to evaluate for genetic mechanisms in this resistance. By understanding the genetic mechanisms in resistance patterns, different treatment plans can be developed. Some hope to evaluate these resistance patterns by using DNA microarrays to compare the genetic expressions in a set of chemosensitive cells versus chemoresistant cells. By analyzing the genetic differences in these cells, one can understand the proteins created from these genes and hopefully the resistance mechanism.

Drug metabolism and toxicity is yet another area of great interest pertaining to DNA microarrays [57, 58]. Anticancer agents, just like other drugs, are metabolized by various enzymes in the human body. Additionally, toxic by-products from drugs can cause various degrees of end-organ damage. Metabolism and toxicity can vary from person to person. Utilizing DNA microarrays to understand the genetic variance in drug metabolism and toxicity has been a recent area of research. This area of research is commonly referred to as pharmacogenomics. The definition of pharmacogenomics is literally the understanding of genetic variance on drug pharmacokinetics and pharmacodynamics. By understanding these principles in individuals, the hope is to create individualized medicines for the treatment of their cancers. For instance, two individuals with the same type of cancer may need a different regimen or dose to treat their cancer based on their genetic differences in drug metabolism. Knowledge of this prior to dosing the medication may be able to reduce the amount of toxicity and increase the efficacy of the drug.

DNA microarrays are often touted as the future of anticancer research. Utilizing them to perform high-throughput processes allows for the rapid understanding of the genetic basis of tumor cells. Furthermore, they will help scientists to understand the subtle differences in various types of cancers. This combined with information on an individual's ability to metabolize a specific drug will assist in the development of personalized therapies [59]. The ultimate goal of this technology is to aid in "tailoring" therapy to an individual. This approach will lead to more effective treatments, with markedly reduced toxicity. These principles are the ultimate goals of anticancer therapy.

## 7 CURRENT CLASSES OF ANTICANCER AGENTS

Anticancer drug research has been going on for decades. Over the years, a vast array of anticancer drugs has been found. Many of these drugs work in a similar manner and belong to specific drug classes. The activities of specific drug classes will often stimulate further research to develop similar drugs. Knowledge of the mechanisms of action of current anticancer agents is vital to the creation of new agents. By evaluating the advantages and disadvantages of various available agents, work can be done to make more effective drugs.

Anticancer drugs are essentially drugs that work by killing cancerous cells. These agents can often destroy or damage normal healthy cells. The basic goal in anticancer drug research is to develop agents that are highly active in tumor cells and less active in healthy cells. Perhaps the best method to understand the scope of anticancer treatment is to look at the current classes of anticancer agents [60].

### 7.1 Alkylating Agents

Perhaps the first anticancer agents used were alkylating agents. Alkylating agents can be further divided into

**TABLE 1 Nitrogen Mustard Agents**

| Drug | Uses | Major Toxicities |
|---|---|---|
| Chlorambucil | Chronic lymphocytic leukemia (CLL), non-Hodgkin lymphoma (NHL), waldenstrom's macroglobulinemia, polycythemia vera, ovarian cancer | Bone marrow suppression, gastrointestinal (GI) distress, central nervous system (CNS), hepatotoxicity |
| Cyclophosphamide | Leukemia, lymphoma, breast cancer, ovarian cancer | Bone marrow suppression, GI distress, alopecia, hemorrhagic cystitis |
| Ifosfamide | Lymphoma, breast cancer, sarcoma, lung cancer, ovarian cancer | Bone marrow suppression, GI distress, alopecia, hemorrhagic cystitis |
| Melphalan | Multiple myeloma, ovarian cancer, mesothelioma | Bone marrow suppression, GI distress, alopecia |

**TABLE 2 Nitrosureas**

| Drug | Uses | Major Toxicities |
|---|---|---|
| Carmustine bis-chloroethylnitrosourea (BCNU) | Brain cancers, multiple myeloma, lymphoma | Bone marrow suppression, pulmonary toxicity, GI distress, hepatotoxicity |
| Streptozotocin | Islet cell cancer | Bone marrow suppression, GI distress |
| Lomustine *N*-(2-chloroethyl)-*N*′-cyclohexyl-*N*-nitrosourea (CCNU) | Brain cancers | Bone marrow suppression |

nitrogen mustard agents, nitrosureas, alkyl sulfonates, hydrazines, and aziridines. Platinum-containing agents are often included as alkylating agents as well. The alkylating agents are cytotoxic agents that work by alkylating DNA as their principle mechanism of action. Alkylating agents can react with various chemical groups on other cellular nucleophiles. By alkylating the cellular DNA, it becomes damaged. The DNA will subsequently become uncoiled and separates. This halts the DNA replication process and the cell can no longer divide.

Because malignant cells divide much more rapidly than normal cells, they undergo DNA replication more often than normal cells. Therefore, alkylating agents are useful as they damage the rapidly dividing cells. Unfortunately, the alkylating agents do not exclude all healthy cells. They will work on any cell that is undergoing DNA replication. Toxicities of the alkylating agents are typically seen in rapidly dividing cells such as the gastrointestinal tract, the reproductive system, and the bone marrow. Additionally, malignant cells can develop resistance to these drugs due to increased capability to repair DNA lesions. They can also develop resistance by selecting for cells that have decreased membrane permeability to alkylating agents. Resistance can also arise from the development of intracellular compounds that inactivate the alkylating agent before it can bind with the DNA.

Nitrogen mustard agents are often credited as being the first anticancer agents. During World War I, mustard gas was used as a chemical weapon. In the 1920s, mustard gas was dissolved in water and applied to superficial skin tumors. These tumors were noted to regress. This finding was quite astonishing initially but was abandoned initially. Then during World War II, the thought of using nitrogen mustard was revisited when an explosion of a ship containing mustard gas was shown to cause marrow and lymphoid hypoplasia in soldiers exposed to the blast. In the 1940s, Gilman and Philips performed the first anticancer clinical trial using nitrogen mustard in patients with lymphomas. The study had promising results, showing dramatic regression of the lymphomas, but these cancers were noted to grow back in many cases. Despite this, nitrogen mustard agents were discovered and the field of anticancer drug research was born.

Nitrogen mustard agents are the oldest modern anticancer agents. They are a class of alkylating agents. There are a number of nitrogen mustard agents that have been developed. Notable agents include cyclophosphamide, ifosfamide, and melphalan. A list of common nitrogen mustard agents is shown in Table 1. Although these agents are quite aged, they still are widely used today.

Nitrosureas are another important class of alklating agents. These agents do not typically have the same resistance patterns as other alkylating agents. Additionally, they are highly lipid soluble, which facilitates their transfer across the blood–brain barrier. Important nitrosureas include carmustine and streptozocin. A more extensive list is given in Table 2.

The platinum analogs are yet another important class of alkylating agents. These agents contain the inorganic metal platinum. Originally, these agents were found to

be alkylating agents when it was noted that they inhibited division of *Escherichia coli.* This group of drugs is not always considered as alkylating agents as their exact mechanism of action is still undefined. It is hypothesized that these agents work in the same way as alkylating agents by binding DNA by making various cross-linkages. This in effect inhibits DNA synthesis and causes cell death. These agents have also been seen to work synergistically with other chemotherapy agents (chemosensitizer) and with radiation therapy (radiosensitizer). Important agents included in this group are cisplatin, carboplatin, and oxaliplatin. These are explained in more detail in Table 3.

There are a number of other alkylating agents that have been discovered. These classes include the alkyl sulfonates, the hydrazines, and the aziridines. Please see Table 4 for more information on the drugs in these classes.

### 7.2 Antimetabolites

Cellular life and growth depends on various cellular metabolites. These metabolites include a variety of chemical compounds. Perhaps the most important of these compounds are the nucleic acids. As mentioned previously, cancerous cells grow and divide much more rapidly than normal cells. This frequent growth and division requires frequent synthesis of DNA. Nucleic acids and other cellular metabolites are required for DNA synthesis. By exploiting these links in the synthesis of DNA, anticancer agents can be developed. The class of antimetabolites does just this.

Most antimetabolites are compounds that show similarities to nucleotides or precursors to nucleic acids. They work by inhibiting normal DNA synthesis either by the production of abnormal nucleotides or by erroneous integration into the cellular DNA during replication. These agents can also work on the enzymes that are integral to nucleic acid synthesis. By inhibiting these enzymes, appropriate DNA synthesis cannot be achieved and further division cannot be achieved.

There are three major classes of antimetabolites. These include the folic acid antagonists, the purine analogs, and the pyrimidine analogs. The agents included in these classes are shown in Tables 5–7. The more widely used agents will be discussed below.

Methotrexate is a commonly used folic acid antagonist. For years, it has been known that folic acid is a necessary compound in cell growth and synthesis. Due to this, the idea of using an "antifolate" agent was pursued. Methotrexate was developed based on this concept. It works by binding to the active catalytic site of the dihydrofolate reductase (DHFR) enzyme. Inhibition of the DHFR enzyme inhibits the production

**TABLE 3 Platinum Agents**

| Drug | Uses | Major Toxicities |
|---|---|---|
| Cisplatin | Lung cancer (small-cell and non-small-cell), ovarian cancer, lymphomas, sarcomas, germ cell tumors | Nephrotoxicity, peripheral neuropathy, nausea, ototoxicity, bone marrow suppression |
| Carboplatin | Ovarian cancer, lung cancer, head and neck cancer | Bone marrow suppression, GI distress, nephrotoxicity |
| Oxaliplatin | Colorectal cancer | Neuropathy, GI distress, bone marrow suppression, ototoxicity |

**TABLE 5 Folic Acid Antagonists**

| Drug | Uses | Major Toxicities |
|---|---|---|
| Methotrexate | Leukemia, lymphoma, breast cancer, head and neck cancer, lung cancer | Bone marrow suppression, pulmonary toxicity, hepatotoxicity, nephrotoxicity, mucositis |
| Pemetrexed | Mesothelioma, non-small-cell lung cancer (NSCLC) | Bone marrow suppression, GI distress, stomatitis, anorexia, rash |

**TABLE 4 Other Alkylating Agents**

| Drug | Uses | Major Toxicities |
|---|---|---|
| Busulfan | Chronic myelogenous leukemia (CML), CLL | Pulmonary toxicity, neurotoxicity, marrow toxicity, hepatotoxicity |
| Procarbazine | Hodgkin's lymphoma, brain cancers | Bone marrow suppression, GI distress, neurotoxicity |
| Temozolomide | Brain cancer, melanoma | Bone marrow suppression, GI distress |
| N,N′N′-triethylenethiophosphoramide (ThioTEPA) | Breast cancer, ovarian cancer, bladder cancer | Bone marrow suppression |

**TABLE 6 Purine Analogs**

| Drug | Uses | Major Toxicities |
|---|---|---|
| Cladribine | Hairy cell leukemia | Fevers, bone marrow suppression |
| Mercaptopurine | Leukemia, lymphoma | Marrow suppression, hepatotoxicity, GI distress |
| Thioguanine | Leukemia, CML | Marrow suppression, hepatoxicity, GI distress |
| Fludarabine | CLL, acute leukemias, lymphomas | Marrow suppression, neurotoxicity, pulmonary toxicity |

**TABLE 7 Pyrimidine Analogs**

| Drug | Uses | Major Toxicities |
|---|---|---|
| Cytarabine | Acute myelogenous leukemia (AML), lymphomas | Marrow suppression, mucositis, neurotoxicity (high dose), conjunctivitis (high dose), pulmonary edema |
| 5-Flurouracil (5-FU) | Colorectal cancer, pancreatic cancer | Marrow suppression, mucositis, neurotoxicity, rash |
| Capecitabine | Colorectal cancer, breast cancer | Hand and foot syndrome, marrow suppression, GI distress |
| Gemcitabine | Pancreatic cancer, NSCLC, bladder cancer, breast cancer | Marrow suppression, GI distress, fevers, hepatotoxicity |

of thymidylate, the purine nucleotides, serine, and methionine. These intracellular compounds are all integral to the formation of DNA, RNA, and critical cellular proteins. Decreased production of these compounds thus results in inhibited cellular division and growth. This effect is seen much more predominantly in rapidly dividing cancer cells.

Because the effects of methotrexate are seen in rapidly dividing cells, the most common adverse effects are related to bone marrow toxicity, gastrointestinal tract irritation, and mucositis. Cancer cells can become resistant to methotrexate through a variety of processes. Cases of decreased drug transport across the cellular membrane have been seen. Additionally, some cancer cells will be able to overcome methotrexate due to increased levels of the DHFR enzyme. Furthermore, cancer cells can sometimes develop a mutated form of the DHFR enzyme that has reduced affinity to the drug.

Purine and pyrimidine antagonists are also important antimetabolite drugs. Normal purines and pyrimidines are necessary for the production of DNA and RNA. By utilizing antagonists of these building blocks, DNA and RNA synthesis can be inhibited. Important pyrimidine antagonists include 5-fluorouracil (5-FU), capecitabine, cytarabine, and gemcitabine. Some purine antagonists include 6-mercaptopurine, fludarabine, and cladarabine. Because of its wide use and interesting mechanism of action, 5-FU warrants further discussion.

5-Flourouracil, or 5-FU, is a commonly used pyrimidine antagonist. Its utility is evident in its multiple mechanisms of action. The drug's main mechanism of action pertains to thymidylate synthase inhibition. The inhibition of this enzyme results in the decreased production of thymidine. Thymidine is a pyrmidine that is necessary for DNA replication. Furthermore, the drug is converted intracellularly to 5-flurouridine-5′-triphosphate (FUTP). During RNA synthesis, FUTP is incorperated into the RNA. This faulty integration causes problems with RNA processing and translation. In addition to this, the drug is also converted to 5-fluorodeoxyuridine-5′-triphosphate (FdUTP). This product is incorporated into DNA, causing inhibition of DNA synthesis. These multiple inhibitory effects on the thymidylate synthase enzyme, DNA, and RNA work together to cause cytotoxicity. Adverse effects seen with 5-FU and most purine and pyrimidine analogs include bone marrow toxicity, gastrointestinal tract irritation, and mucositis.

### 7.3 Plant Alkaloids

Natural products have held their place in drug discovery for years. Their contribution to the anticancer drug library is quite substantial. The vinca alkaloids, taxanes, epipodophyllotoxins, and comptothecins are all widely used, highly active classes of anticancer agents. Each class warrants further discussion.

The vinca alkaloids are drugs derived from the periwinkle family of plants. The vinca alkaloids include vinblastine, vincristine, vinorelbine, and vindesine. Vinblastine and vincristine are direct derivatives from the periwinkle plant *Vinca rosea.* Vinorelbine and vindesine are semisynthetic vinca alkaloids. These agents are displayed in Table 8. Their mechanisms of action of these agents are all quite similar. Their activity in various tumor types can differ, however.

The vinca alkaloids work by inhibition of tubulin polymerization. Inhibition of tubulin polymerization in turn inhibits the assembly of microtubules. Microtubules play an integral role in the maintenance of the cellular cytoskeleton. Furthermore, they are necessary elements in the development of the mitotic spindle during cellular division. Disruption of this system causes disruption of cell division during mitosis. This ultimately results in cell death. Of course, in malignant cells cell division is much

**TABLE 8 Vinca Alkaloids**

| Drug | Uses | Major Toxicities |
|---|---|---|
| Vincristine | Lymphomas, acute leukemias | Vesicant, marrow suppression, neurotoxicity, GI distress, cardiovascular toxicity |
| Vinblastine | Hodgkin's lymphoma, NSCLC, breast cancer, testicular cancer | Vesicant, marrow suppression, hypertension, Raynaud's phenomenon |
| Vinorelbine | Breast cancer, NSCLC | Vesicant, marrow suppression, bronchospasm, cough |

more frequent. Therefore, these agents work much more effectively in cancerous cells.

Because of their effects on rapidly dividing cells, these agents are noted to cause bone marrow toxicity, alopecia, and gastrointestinal toxicity. The agent vincristine has also been shown to cause severe peripheral neuropathy and syndrome of inappropriate antidiuretic hormone hypersecretion (SIADH). Despite these adverse effects, these agents are commonly used in a variety of cancers, including non-small-cell lung cancer, breast cancer, leukemias, lymphomas, germ cell cancers, and multiple myeloma.

The taxanes are another important class of plant alkaloids. They include the agents paclitaxel and docetaxel. Both are products of the yew tree (genus *Taxus)*. They act in a similar manner to vinca alkaloids in that they are mitotic spindle inhibitors. Interestingly, the drugs do not work to inhibit the formation of the mitotic spindle. The drugs bind to microtubules and actually enhance tubulin polymerization. The enhanced polymerization occurs in the absence of compounds that are needed to make functional microtubule structures. The enhanced polymerization also inhibits normal microtubule breakdown, effectively locking the microtubules in place. The drug essentially causes erroneous formation of nonfunctional microtubule structures. These abnormal structures inhibit mitosis and ultimately cause cell death.

Because of their mitotic inhibition, the taxanes have substantial activity in a variety of cancers. Paclitaxel is the most widely used and is utilized in the treatment of a number of different solid tumors. Both agents do have substantial dose-limiting side effects. The side effects from paclitaxel are mostly related to its delivery vehicle, Cremaphor EL. Due to this a large amount of research is being done to find an alternative vehicle. This is further discussed in the drug formulation and delivery chapter. Please see Table 9 for more information on adverse effects and the tumor types treated with the taxanes.

**TABLE 9 Taxanes**

| Drug | Uses | Major Toxicities |
|---|---|---|
| Paclitaxel | NSCLC, breast cancer, ovarian cancer, head and neck cancer | Hypersensitivity reactions, marrow suppression, mucositis, alopecia, peripheral neuropathy |
| Docetaxel | Breast cancer, ovarian cancer, NSCLC | Hypersensitivity, fluid retention, marrow suppression, peripheral neuropathy |

**TABLE 10 Camptothecins**

| Drug | Uses | Major Toxicities |
|---|---|---|
| Topotecan | Ovarian cancer, lung cancer | Marrow suppression, mucositis, GI distress, alopecia |
| Irinotecan | Colorectal cancer | Diarrhea, marrow suppression, alopecia, GI distress, pulmonary toxicity |

The camptothecins are yet another class of plant alkaloids in use today. They are derived from the *C. acuminate* tree. Agents include topotecan, irinotecan, camptothecin, and rubitecan. The drugs have a similar mechanism of action that involves the inhibition of the enzyme topoisomerase I. This enzyme is integral in the normal breakdown and religation of the phosphodiester backbone of DNA strands. Inhibition of this enzyme, therefore, inhibits DNA synthesis and damages DNA. The damaged DNA causes apoptosis and cell death. These effects are especially noted in rapidly dividing cells such as cancer but can also be noted to have effects in the bone marrow and the gastrointestinal tract. The adverse effects and cancers treated with these agents are noted in Table 10.

The epipodophyllotoxins are anticancer agents derived from the root of the mayapple plant (*Podophyllum peltatum*). Included in this class are the compounds etoposide and teniposide. These agents are similar in their mechanism of action. They work by inhibiting topoisomerase II. Similar to topoisomerase I inhibition, the inhibition of topoisomerase II causes DNA damage. This damage is mediated by the strand breakage caused by complexes of the drug, the DNA, and the topoisomerase enzyme. This arrests cell division and causes cell death. This effect is once again enhanced in rapidly dividing cells, including cancer cells. Please see Table 11 for more information regarding the side effects and cancers treated with these drugs.

**TABLE 11 Epipodophyllotoxins**

| Drug | Uses | Major Toxicities |
|---|---|---|
| Etoposide | Ewing's sarcoma, lung cancer, brain cancer, lymphoma, testicular cancer | Marrow suppression, alopecia, hypersensitivity, GI distress, mucositis, hypotension |
| Teniposide | Acute lymphocytic leukemia (ALL) | Marrow suppression, alopecia |

### 7.4 Anticancer Antibiotics

Many drug discoveries occur when researching other diseases. The search for antimicrobials has yielded several useful anticancer agents. As microbes are often the source for many antibiotics, they often serve as natural sources for anticancer agents. Many microbes have defense mechanisms that work in a manner similar to the desired effects of some anticancer agents. These defense mechanisms have been researched and can be successfully exploited to destroy cancerous cells.

Perhaps the most famous class of anticancer antibiotics is the anthracyclines. The anthracycline antibiotics are compounds derived from the microbe *Streptomyces peucetius*. Agents derived from this microbe include doxorubicin, daunorubicin, idarubicin, epirubicin, and mitoxantrone. The agents included in this group have been utilized as anticancer agents for many years. They are used in a very wide variety of cancers. The reason for this is likely due to their array of mechanisms of action.

The anthracyclines have a number of mechanisms to destroy cancerous cells. First of all, these compounds have a high affinity to DNA/RNA and will intercalate between base pairs. This intercalation causes disruption of DNA and RNA synthesis. Like many other drugs that interrupt DNA and RNA synthesis, this action causes cell death. In addition to this primary mechanism, anthracyclines generate iron-dependent, enzyme-mediated oxygen free radicals. These free radicals wreak havoc on the cell by damaging DNA and cell membranes. The anthracyclines will also damage the cellular membrane by binding to it and altering fluidity and ion transport. Finally, anthracyclines have the ability to inhibit topoisomerase II. Their inhibition of this enzyme causes halting of the enzymes normal progression. This mechanism essentially causes the complex to stabilize before it can reseal the DNA double helix during replication. This effectively terminates DNA synthesis and thus causes cell death.

By utilizing several different cytotoxic mechanisms, the anthracyclines hold their place as some of the most useful chemotherapeutics to date. Doxorubicin is perhaps the most widely used, with proven clinical activity in breast, endometrial, testicular, ovarian, lung, and stomach cancer, just to name a few. The drug is also quite useful in a number of hematologic malignancies. The side effect profile of these agents can vary slightly with each drug, but they all include some element of bone marrow toxicity, mucositis, and cardiotoxicity. Detailed information on the individual side effect profiles and clinical uses for the anthracyclines is shown in Table 12.

**TABLE 12 Anthracyclines**

| Drug | Uses | Major Toxicities |
|---|---|---|
| Daunorubicin | AML, ALL | Marrow suppression, mucositis, alopecia, cardiovascular toxicity, vesicant |
| Doxorubicin | Leukemia, lymphoma, bladder cancer, breast cancer, stomach cancer, lung cancer, ovarian cancer, sarcomas, multiple myeloma, | Marrow suppression, mucositis, alopecia, cardiovascular toxicity, vesicant |
| Idarubicin | AML | Marrow suppression, minimal cardiotoxicity |
| Epirubicin | Breast cancer, ovarian cancer, lung cancer, gastric cancer, lymphomas | Marrow suppression, cardiotoxic |
| Mitoxantrone | Breast cancer, AML, lymphoma | Marrow suppression, cardiotoxic, blue-colored urine, sclera, nails |

There are a number of other antitumor antibiotics that are used in clinical practice today. These agents are similar to the anthracyclines in that they work by inhibiting normal DNA and RNA synthesis. They include mitoxantrone, dactinomycin, mitomycin, and bleomycin. The clinical uses and side effects of these agents are shown in Table 13.

### 7.5 Hormonal Agents

Breast and prostate cancer are two very common and lethal cancers. As these cancers are typically found in a specific sex, it is no surprise that the sex hormones play a role in their proliferation. If the hormones play a role in proliferation, they can also be a key to the cancer's

**TABLE 13 Other Antitumor Antibiotics**

| Drug | Uses | Major Toxicities |
|---|---|---|
| Bleomycin | Hodgkin's lymphoma, testicular cancer | Pulmonary toxicity, Raynaud's syndrome, skin toxicity |
| Mitomycin | Esophageal cancer, breast cancer, bladder cancer | Marrow suppression, vesicant, pulmonary toxicity, cardiac toxicity, hemolytic uremic syndrome |
| Actinomycin | Wilm's tumor, rhabdomyosarcoma | Marrow suppression, GI distress, mucositis, vesicant, alopecia |
| Mitoxantrone | Breast cancer, AML, lymphoma | Marrow suppression, cardiotoxic, blue-colored urine, sclera, nails |

demise. At the present time, there has been a great deal of research in "antihormone" agents that can be utilized to treat breast and prostate cancer. The plan of blocking the hormonal growth stimulus has proven to be a useful weapon in the anticancer arsenal. The agents currently used can be divided into the classes of selective estrogen receptor modulators, aromatase inhibitors, androgen antagonists, and gonadotropin-releasing hormone agonists.

### 7.6 Selective Estrogen Receptor Modulators

The selective estrogen receptor modulators (SERMs) are a class of drugs that exhibit effects on human estrogen receptors [61]. These agents are quite interesting in regards to their ability to have both agonistic and antagonistic effects on estrogen receptors depending on the tissue affected. Estrogen has a number of effects on a number of normal tissues. The fact that breast cancer is often noted to have its growth modulated by the presence of estrogen makes this class of drug particularly interesting. The SERM of most interest and utility as an anticancer agent today is tamoxifen.

Tamoxifen is a SERM with mixed estrogen receptor agonism and antagonism in various tissues. Most importantly tamoxifen has profound estrogen receptor antagonism in breast tissue. The drug is taken orally and is converted in the liver to its active form. The active form of the agent then competitively binds to estrogen receptors in the target tissues and, more importantly, estrogen-receptor-positive tumors. Once it has bound to the receptor, the drug produces a nuclear complex that inhibits the effects of estrogen on the tissues. One of these inhibitory effects is to decrease DNA synthesis. The decreased DNA synthesis results in the halt of the cell cycle. The drug is often termed cytostatic rather than cytotoxic due to these effects.

Tamoxifen is generally well tolerated due to its "estrogen"-limited mechanism of action. In contrast to its estrogen antagonism tumors, it works as a weak agonist in bone. This makes it a useful agent in the treatment of osteoporosis. However, the drug does have some notable side effects. The major concern when using the drug is related to its partial agonism in endometrial tissue. This partial agonism leads to an increased incidence of abnormal uterine bleeding and in some cases an increased risk for endometrial cancer. Tamoxifen also increases the risk for venous thromboembolism and hot flashes.

In spite of its side effects, tamoxifen remains one of the most widely used agents for estrogen-receptor-positive tumors. Its use has stimulated a great deal of research in the area of antihormonal therapies. Other SERMs continue to be investigated and compared to tamoxifen to determine if they are superior. For the time being, tamoxifen remains the most useful SERM in regards to breast cancer treatment. Development of tissue-specific SERMs that exhibit beneficial effects in bone and the cardiovascular system, while continuing their antagonism in tumor tissue, is the focus of current research.

### 7.7 Aromatase Inhibitors

Other methods of blocking the effects of estrogen on tumor cells have given birth to the aromatase inhibitors. The production of estrogen in the human body is mediated by the enzyme aromatase. This enzyme works to convert androgens into estrogens. As noted earlier, estrogen enhances DNA synthesis in many tissues, including estrogen-receptor-positive tumors. By inhibiting the production of estrogen in the body, this DNA synthesis will also be decreased. In tumors, this is particularly useful as it causes the tumor cells to have arrested development. The aromatase inhibitors work in various ways to inhibit the aromatase enzyme. The agents included in this class are anastrozole, aminoglutethimide, letrozole, and exemestane. Anastrozole and letrozole act reversibly on the aromatase enzyme while exemestane binds irreversibly.

When compared to tamoxifen in clinical trials, aromatase inhibitors have been shown to be more effective. Perhaps this is because estradiol has a higher affinity to estrogen receptors than tamoxifen. This means that any amount of estradiol in the body can limit the effectiveness of tamoxifen. From this information it appears that it is quite important to decrease the amount of

estrogens circulating in the body as much as possible. The aromatase inhibitors allow for this to occur. The aromatase inhibitors have been shown to be slightly superior to tamoxifen in trials involving postmenopausal patients.

The aromatase inhibitors are also well tolerated. Their side effect profile is similar to that of the symptoms of menopause due to the loss of estrogen. Some common side effects seen with all aromatase inhibitors include hot flushes, fatigue, arthralgias, and headaches. Despite this side effect profile, these agents will continue to be important agents in estrogen-receptor-positive cancers.

### 7.8 Androgen Antagonists

Just as the estrogen plays an important role in cell growth in estrogen-receptor-positive cancers, androgens play an important role in androgen-receptor-positive tumors. Prostate cancer is a type of cancer that exhibits androgen-related growth. Inhibition of the androgen receptor in prostate cancer results in slowed cell growth. The agents flutamide and bicalutamide are both antiandrogen agents that are utilized in prostate cancer due to this effect. They are not cytotoxic, but they are useful in slowing the growth of the disease. The drugs are fairly well tolerated but commonly cause hot flushes, nausea, and mild elevations in transaminases.

### 7.9 Gonadotropin-Releasing Hormone Agonists

Gonadotropin-releasing hormone agonists (GnRH agonists) serve as another method to disrupt androgen synthesis. Gosrelin and leuprolide are both synthetically produced analogs of GnRHs. These drugs work by causing a "superagonist" effect on the pituitary. This in turn causes downregulation of the production of lutenizing hormone (LH) and follicle-stimulating hormone (FSH). These hormones are responsible for the production of the androgens. Once again, these agents are cytostatic agents. They are useful in slowing the progression of prostate cancers. Due to this they are widely used in advanced prostate cancer. Their side effect profile includes hot flushes, gynecomastia, and impotence.

### 7.10 Molecular Targeted Therapies

The evolution of biomedical research has expanded our understanding of cellular biology. As we learn about normal cells, we are able to see the differences present in malignant cells. In particular, recent research has produced knowledge of the intricate cellular signaling pathway in a number of tissue types. The signaling pathways in many tumors have also been analyzed. These signaling pathways are vital for the control of cell growth, division, differentiation, metabolism, and apoptosis. By targeting these pathways, these cellular processes can be modulated or even stopped. The era of molecular medicine has risen from this concept. Oncology has been a significant beneficiary from this area of research.

The protein kinases are a class of enzymes that play an integral role in cellular differentiation, proliferation, and regulation [62]. These enzymes can be abnormally regulated in some forms of cancer. Included in the class of protein kinases are the growth-related kinases. Perhaps the most interesting growth-related kinases are the tyrosine kinases. These kinases include epidermal growth factor receptor (EGFR), vascular endothelial growth factor (VEGF), and platelet-derived growth factor (PDGF). Inhibition of these kinases can lead to suppressed cell growth. It is no surprise that tyrosine kinase inhibitors have been heavily researched in recent years as potential anticancer agents.

There are a number of receptor and nonreceptor tyrosine kinase inhibitors that have been developed and that are in current development. All of these drugs work through inhibitory mechanisms on the tyrosine kinase enzyme. Some may target only one tyrosine kinase, while others will target several [63]. Perhaps the best way to understand these molecules is to discuss a few of them.

### 7.11 Imatinib

One of the most well-known tyrosine kinase inhibitors is imatinib (Gleevec). This tyrosine kinase inhibitor works by inhibiting the tyrosine kinase domain of the Bcr-Abl oncoprotein, thus preventing phosphorylation of the kinase substrate by ATP. The Bcr-Abl oncoprotein is a product of the Philadelphia chromosome translocation, t(9:22). This translocation is commonly seen in chronic myelogenous leukemia (CML) and other leukemias. The Bcr-Abl oncoprotein activates mitogenic and antiapoptotic pathways, thus causing abnormal cell growth and virtual immortality. Inhibition of this oncoprotein has proven to be a successful way to halt these processes.

Soon after the Bcr-Abl product was found, the search for an inhibitor was started. Imatinib was developed by chemists by utilizing high-throughput screening of chemical libraries. Ultimately, a lead compound was found and imatinib was derived from this. The drug is credited for drastically changing the therapy for CML.

Interestingly, imatinib has been shown to have activity in gastrointestinal stromal tumors (GISTs) as well. This is due to the presence of the *c-kit* tyrosine kinase mutation in these tumors. In addition to inhibition of

the Bcr-Abl product, imatinib has been found to be a selective mast/stem cell growth factor receptor (KIT) tyrosine kinase inhibitor. This multikinase inhibition allows for the agent to be useful in multiple types of cancer. This quality is seen in several protein kinase inhibitors.

Molecular targeted agents are typically well tolerated due to their specific mechanism of action. Imatinib is no exception to this rule and is usually well tolerated. Side effects include as nausea, myalgias, edema, and a rash.

### 7.12 Gefitinib and Erlotinib

Growth factor receptors are a group of tyrosine kinases that regulate various cellular growth mechanisms. There are a number of different types of this receptor. One of these is the epidermal growth factor receptor (EGFR). This receptor is commonly overexpressed in a variety of cancers, notably breast, colon, and lung cancer. Activation of EGFR causes the Ras signal transduction pathway, which causes an antiapoptotic response. This leads to uninhibited cell growth, unregulated proliferation, increased angiogenesis, stimulation of cell invasion, and metastasis.

Gefitinib (Iressa) is a small-molecule inhibitor of the tyrosine kinase portion of EGFR. It works by binding to the ATP-binding site of the tyrosine kinase enzyme. The resulting uninhibited Ras signal transduction cascade is effectively halted by this molecule. The ultimate effect is inhibition of cell growth.

Due to its effects, gefitinib has been evaluated in a number of cancers. It is now approved for use in refractory non-small-cell lung cancer. Its side effect profile is well tolerated. Side effects include rash, nausea, stomatitis, and fatigue.

Erlotinib (Tarceva) is a small molecule that is quite similar to gefitinib. It works like gefitinib to inhibit the tyrosine kinase domain of the EGFR. It displays a similar side effect profile to gefitinib and is also indicated for use in refractory non-small-cell lung cancer. Additionally, erlotinib has been recently approved for use in combination with gemcitabine for the treatment of advanced pancreatic cancer.

### 7.13 Sorafenib and Sunitinib

Multitarget tyrosine kinase inhibitors are an interesting group of protein kinase inhibitors. These agents allow for the inhibition of several different tyrosine kinase enzymes to work synergistically to halt abnormal cell growth. Sorafenib (Nexavar) and sunitinib (Sutent) are examples of two multitarget tyrosine kinase inhibitors.

Sorafenib targets the tyrosine kinases c-KIT, PDGF, VEGF, Raf, B-Raf, and extracellular regulated kinase (ERK). Explanation of the actions of each of these kinases is quite detailed. Inhibition of these tyrosine kinases allows for the simultaneous inhibition of cellular growth, proliferation, division, and angiogenesis. Inhibition of multiple pathways allows for a more effective agent.

Sorafenib has been approved for use in advanced renal cell carcinoma and has been shown to improve progression-free survival [64]. It has also been shown to have activity in other cancers, including hepatocellular carcinoma [65]. The drug is fairly well tolerated with rashes, diarrhea, and hypertension as main side effects.

Sunitinib is another important multitarget tyrosine kinase inhibitor. It works on the tyrosine kinases c-KIT, PDGF, VEGF, RET, and fms-like tyrosine kinase receptor-3 (FLT3). Similarly to sorafinib, this agent's multikinase inhibition causes inhibition of multiple processes integral to the proliferation of malignant cells. The drug is currently approved for use in advanced renal cell cancer. Due to its c-KIT activity, sunitinib has been approved for imatinib-resistant GIST. The drug has also been shown to have activity in a number of other tumors including breast cancer, non-small-cell lung cancer, and neuroendocrine tumors [66]. Like sorafinib, the drug is very well tolerated with skin changes, hypertension, nausea, fatigue, and diarrhea as major side effects.

There are a number of approved tyrosine kinase inhibitors available and in research. Please see Table 14 for more information on some of these agents.

**TABLE 14 Protein Kinase Inhibitors**

| Drug | Approved Uses | Major Side Effects |
|---|---|---|
| Dasatinib | CML | Marrow suppression, edema, headaches, diarrhea |
| Erlotinib | NSCLC, pancreatic cancer | Rash, diarrhea, anorexia |
| Gefitinib | NSCLC | Acne, diarrhea, GI distress, hepatitis |
| Imatinib | CML, gastrointestinal stromal tumors (GISTs) | Edema, nausea, rash, myalgias |
| Lapatinib | Breast cancer | GI distress, anorexia, rash |
| Nilotinib | CML | Cardiotoxicity |
| Sorafinib | Renal cell cancer, hepatocellular carcinoma | Rash, diarrhea, hypertension |
| Sunitinib | Renal cell cancer, GIST | Fatigue, GI distress, hypertension, skin changes, stomatitis |

### 47.7.14 Monoclonal Antibodies

Monoclonal antibodies are essentially antibodies created to target a specific antigen. By doing so, drugs can be made to bind cell receptors and enzymes that mediate growth. Monoclonal antibodies can also be created to target abnormal oncoproteins that are commonly found in malignant cells. Several different groups of these agents have been developed. Some work by targeting and inhibiting tyrosine kinases. Others work by binding and inactivating other integral cellular receptors and enzymes.

Trastuzumab (Herceptin) is one example of a monoclonal antibody that is used for cancer treatment. This drug works by binding to the extracellular HER2/neu receptor. HER2/neu is a tyrosine kinase mediated receptor that is overexpressed in some types of breast cancer. Overexpression of the receptor causes overactivation of a mitogenic signaling pathway involving mitogen-activated protein kinase (MAP) kinases. The resulting effect is promotion of cellular invasion, increased angiogenesis, and prolonged survival. Tumors expressing this abnormality are typically thought to have a worse prognosis due to resistance to standard chemotherapy. By binding and inhibiting the HER2/neu receptor, trastuzumab downregulates effects of the receptor. This results in arresting the cell cycle and suppression of angiogenesis.

Trastuzumab has been noted to drastically change the treatment of HER2/neu-positive breast cancer and is indicated for therapy. This agent shows significant activity in tumors that previously showed poor response to standard chemotherapy. Additionally, this drug is well tolerated when compared to standard chemotherapy. Major side effects include cardiac toxicity, rash, nausea, diarrhea, and myalgias.

Bevacizumab (Avastin) is another monoclonal antibody that inhibits tyrosine kinases. This drug works to target the VEGF. Targeted inhibition of VEGF results in inhibition of angiogenesis. This process is normally vital to tumor growth due to the need for blood supply. Inhibition of angiogenesis essentially causes the tumor to "starve" and it will not be able to grow further or metastasize. Inhibition of this growth factor receptor has shown promise in a number of tumor types. Bevacizumab is currently indicated for use in colon cancer and most recently in breast cancer [67]. It has shown promise in a variety of other cancers, including lung cancer, ovarian cancer, and renal cell cancer. Due to its inhibition of angiogenesis, bevacizumab has been noted to cause problems with wound healing. It has also been noted to cause hypertension, rash, and increased risk of bleeding.

A number of other monoclonal antibodies are used as anticancer agents. Their uses and side effects are noted in Table 15.

**TABLE 15 Monoclonal Antibodies**

| Drug | Approved Uses | Major Side Effects |
|---|---|---|
| Cetuximab | Colorectal cancer, head and neck cancer | Severe rash |
| Trastuzumab | *Her2/neu*-positive breast cancer | Cardiotoxicity |
| Rituximab | B-cell lymphoma and leukemia | Infusion reactions, pulmonary toxicity, fever |
| Bevacizumab | Colorectal cancer | Impaired healing, hypertension, bleeding |
| Gemtuzumab ozogamicin | AML | Marrow suppression, venoocclusive disease, fevers |

### 7.15 mTOR Inhibitors

Mammalian target of rapamycin (mTOR) is a serine–threonine protein kinase that regulates cell growth, proliferation, protein synthesis, and survival [68]. Inhibition of this protein kinase has been a popular mechanism for immunosuppressants. Recently, inhibition of this protein kinase has shown some utility for its use as an anticancer agent.

Temsirolimus (Torisel) is an mTOR inhibitor that has shown its ability as an anticancer agent [69]. By inhibiting mTOR, temsirolimus effectively hinders cell proliferation and angiogenesis. It has shown significant activity in several different tumor types.

Recently, temsirolimus was approved for the treatment of renal cell carcinoma. The side effect profile of this drug is mild and includes bone marrow suppression, rash, weakness, nausea, and diarrhea. The mTOR pathway is integral for growth and angiogenesis in a number of different tumor types. There is hope that temsirolimus will find utility in multiple tumor types due to these effects.

### 7.16 Cancer Vaccines and Immunotherapy

There is existing evidence that suggests that the immune system plays an important role in controlling malignancy. This has been supported by the observation of the regression of cancer in some patients as well as increased cancer incidence in patients who are immunosupressed. In addition, in numbers of tumors including melanoma there is evidence of tumor infiltration by immune cells. Those observations led to the development of several strategies including immunotherapy with cytokines, two of which, interferon and interleukin-2, are available for treatment of renal cell carcinoma and

malignant melanoma. In addition, several vaccine strategies have been developed for the treatment of cancer, however, to date none has been effective enough to be part of cancer patient management outside of a clinical trial. Cells of an innate immune system respond to signals provided by growing tumors that cause induced inflammation and activate innate effector cells with antitumor activity, and stimulate antigen presenting cells and dendritic cells. This leads to migration of the immune cells to the draining lymph nodes and results in triggering adaptive response by B and T lymphocytes. Natural killer cells and T cells produce interferon after activation, which on its own has a tumoricidal activity and stimulates macrophages and dendritic cells to produce interleukin-12. The expansion and activation of natural killer cells can also be accomplished by increased levels of interleukin-2. The use of high-dose interleukin-2 therapy has been shown to induce responses in 16% of patients with metastatic malignant melanoma and 21% of patients with metastatic renal cell carcinoma. Six percent of patients with metastatic kidney cancer or melanoma can be cured with that strategy, even at the metastatic stage of the disease.

A different strategy to enhance immune responses against cancer was the administration of adjuvant agents. Those agents may contain bacterial products such as lipopolysaccharide or CpG oligodeoxynucleotide. (C stands for cytosine, p for phosphodiester, and G for guanine.) containing oligodeoxynucleotides, recognized by Toll-like receptors. In recent years, development of small molecules directly stimulating Toll-like receptors has been also seen. The activation of Toll-like receptors leads to production of pro-inflammatory cytokines and productive interaction between the innate and adaptive immune responses. Those strategies were recently tested in patients with non-small-cell lung carcinoma as was malignant melanoma.

Another very important recent strategy in developing increased immunity against cancer is to suppress activity of regulatory T cells. This strategy had been developed with developing antibodies against the CTLA4 molecule. This strategy has been tested in malignant melanoma and lung cancer with mixed results.

In order to improve visibility of the cancer antigens to immune system, a number of different strategies have been developed, including identification of tumor-specific proteins and the introduction of those proteins either alone or with adjuvant therapies.

In addition, a number of therapies has been developed to stimulate dendritic cells or to "teach" dendritic cells to recognize tumor antigens ex vivo. One of these strategies is cancer cell fusion with autologous dendritic cells, which would allow the enhanced presentation of tumor antigens to other elements of the immune system. One early strategies was isolation from tumor-infiltrating lymphocytes and after ex vivo expansion re-infusion to patients with cancer. Similarly ex vivo expansion of natural killer cells and infusion to patients with leukemia or breast cancer resulted in promising activity.

Genetically modified T cells expressing cytokines have also been used as a strategy to enhance tumor recognition. Recently, a renaissance in interest in developing clinical vaccines has been seen even against solid tumors that were previously thought to have little immunogenic potential, for example, lung cancer, prostate cancer, or brain tumors. In conclusion, there is a multitude of different immunomodulatory and vaccine strategies in development. So far, the proven clinical value has only been provided by interferon and interleukin-2 in patients with renal cell carcinoma and melanoma.

Vaccine strategies are very appealing because they present a possibility of tumor-restricted therapy with limited toxicity suffered by normal organs. As the current knowledge of immune responses to the foreign antigen or mutated self-antigen is being developed, new clinical studies are being designed. This hopefully will lead to clinical introduction of anticancer vaccine strategies in the next several years.

### 7.17 Conclusion

There is a multitude of different agents currently available for cancer treatment. Some agents have quite broad mechanisms of action while others are more specific. In general, it is noted that the agents that have a more specific mechanism of action are typically tolerated better. This is why the molecularly targeted agents have gained so much esteem recently. Despite this, standard cytotoxic chemotherapy does still play a role in the treatment of cancer. In the future we can expect to see more molecularly targeted therapies being developed. These drugs will likely become more tumor specific, thus improving their side effect profile. Standard cytotoxic agents will also continue to be useful. The development of these agents will continue to increase our arsenal of therapies for cancer. The ultimate goal is to develop a treatment regimen that has virtually no side effects, while still effectively irradicating tumor cells. As our knowledge of cancer cell biology increases, we will only find more targets for our drugs. Utilizing the vast array of research tools will facilitate the development of these agents.

## 8 DRUG FORMULATION AND DELIVERY SYSTEMS

The process of finding potential anticancer drugs is often extensive and difficult. When promising agents are eventually found, they must overcome more hurdles. One such hurdle is in regards to how they will be

delivered. Drug solubility, stability, bioavailability, concentration, and local toxicity all play a role in developing appropriate delivery systems for anticancer drugs. In this current day and age, there is a multitude of ways to deliver drugs. Intravenous, oral, sublingual, intranasal, subcutaneous, and even transdermal routes are used to deliver a number of drugs. Anticancer agents are notably cytotoxic agents and often require specific delivery systems. At this point, most anticancer agents are administered intravenously. There is, however, a number of newer agents that are being developed that utilize an oral route. The advancement of drug delivery systems is to thank for these advances. The easiest way to think of anticancer drug delivery is to break it up into several categories: drug formulation, drug delivery systems, drug conjugates, and prodrugs.

### 8.1 Drug Formulation

Drug formulation is the process by which several different chemical compounds are combined with a drug to create the final medicinal compound. The formulation of a medicinal compound can be achieved with several different goals in mind. Currently, the most widely utilized approach in anticancer research is to develop a formulation that keeps the integrity of the drug as "pure" as possible [70]. This essentially means that care is taken to make sure that the mechanism of action of the drug is not changed by the formulation. Another popular approach is to develop a formulation that enhances the drug's effectiveness, either by increasing its effect on the target cells or by minimizing the adverse effect of the drug so that a higher dose may be given.

The first step in formulation is actually preformulation techniques that are used to analyze the chemical structure of the drug. This step is called analytical characterization. Analytical characterization utilizes approaches such as nuclear magnetic resonance (NMR) imaging, mass spectroscopy, differential scanning calorimetry (DSC), X-ray diffraction analysis, and infared spectroscopy [71]. These steps are performed to find information regarding the chemical structure of the compound, as well as to evaluate some of the properties of the compound. Properties such as solubility, drug stability, and molecular size can be determined using these technologies. Once this data is ascertained, further work on formulation of the final drug compound can be more easily achieved.

### 8.2 Solubility

Solubility is a property that is integral in drug formulation. A potential anticancer agent may be useful in vitro, however, if it is insoluble in vivo, it can be useless. There are multiple acceptable solvents available for the delivery of anticancer agents. However, these solvents can come with their own adverse effects. A prime example of this is Cremophor EL [72]. This is a nonionic surfactant that is used to deliver paclitaxel. When it is given intravenously, it can cause severe anaphylactic reactions and over time can cause peripheral neuropathies. When receiving paclitaxel infusions, patients typically have to have the infusion run over a long period of time to prevent anaphylaxis. They often have to be premedicated with antihistamines or corticosteroids to prevent reactions. For this reason, solvents must be carefully selected. Optimally, an anticancer agent will be soluble in a well-tolerated medium, however, events do occur in which an agent is entirely insoluble. In this case further techniques must be employed to aid in solubility. Other solubilization approaches include adjusting the pH, utilizing co-solvent or surfactant systems, creating emulsions, and developing colloidal systems.

If a compound is insoluble using standard solvents, altering the pH is often the first approach to improve solubility. By adjusting the pH, one can change the ionized form of the compound to allow it to have better solubility in the corresponding solvent. When drastically altering the pH, it is important to evaluate the changes in the compound's stability and possible changes in efficacy. It is also important to ensure that the altered pH will not cause the drug formulation to become harmful to the recipient.

Co-solvent and surfactant systems are also commonly used to improve solubility. Common cosolvents used today include dimethyl sulfoxide (DMSO), ethanol, polyethylene glycol (PEG), sorbitol, glycerin, propylene glycol, and *N*-methyl-2-pyrrolidone [73]. An appropriate co-solvent will be miscible with an aqueous solution. This will allow the drug to be dissolved in the co-solvent and then this solution can be dissolved into a water-based solution. Unfortunately, these systems can be difficult to use as minor changes in temperature, pH, or further dilution can cause the compounds to precipitate out of the solution. This can be remedied by the use of surfactant systems. Surfactant systems work by using particles that have polar and nonpolar regions that will essentially reorganize themselves to create micelles in aqueous solutions. These systems are quite beneficial in more nonpolar compounds as the micelles will surround the nonpolar particles with their nonpolar regions, keeping the polar regions on the outer surface of the micelle, thus allowing the micelle to be soluble in water. There are a number of surfactant systems used today including Cremophor EL, pluronic copolymers, and polysorbate polymeric systems.

Emulsions can be utilized to create soluble drug formulations when the compound of interest is highly lipo-

philic. An emulsion is essentially a mixture of two immiscible substances. There are many common emulsions we see in everyday life. These include butter, mild, ream, and mayonnaise. Emulsions work by mixing one substance (sometimes called the dispersed phase) into another substance (the continuous aqueous phase). This "oil-and-water" mixture is usually stabilized by an emulsifier. An example of an emulsion system being researched currently is the paclitaxel vitamin E emulsion (TOCOSOL) [74]. This formulation uses vitamin E, a tocol, which is a large lipophilic molecule. The formulation involved dissolving paclitaxel in vitamin E initially. This lipophilic solution is then homogenized with water, utilizing surfactants as the emulsifying agents. The surfactants utilized are d-alpha tocopheryl polyethylene glycol 1000 succinate (TPGS) and Poloxamer 407. The final product is an emulsion system that consists of three principal components, an inner core consisting of the lipophilic tocol (vitamin E), the paclitaxel particles residing inside the core, and the surfactants surrounding the emulsion particle. This emulsion system is still being researched and is in clinical trials [75].

Colloidal systems are yet another method to dissolve a previously insoluble drug compound in a delivery medium. Liposomes are an excellent example of the use of colloidal systems in drug formulation in recent times. These are explained in detail later on.

### 8.3 Stability

The next important property that must be considered in drug formulation is the compound's stability. Preformulation studies often will evaluate the chemical stability of the drug compound. They will also provide vital information on compound degradation and its breakdown products. Other factors that can be evaluated include the compound's response to various temperatures, pH, wavelengths of light, and even radiation. Once these factors are determined, appropriate delivery vehicles can be chosen more effectively. Lyophilization is one method that is commonly used in drug formulation today [76]. This method essentially is freeze-drying the compound so that it will remain stable for longer periods. Once the lyophilized compound is needed, it can be dissolved back into an appropriate solvent. Ultimately, the formulation of the drug should maintain the stability of the compound so that it can be easily stored, transported, and administered to patients in its most active form.

### 8.4 Liposomal Delivery

Liposomes are essentially spherical vesicles that consist of a phospholipid bilayer. This concept is based on basic cellular biology and the phospholipids bilayer that creates the cellular membrane. A phopholipid is an amphiphilic lipid that consists of a hydrophilic phosphate head and a hydrophobic lipid tail. Their amphiphilic nature allows them to self-assemble into spherical bilayers when introduced into an aqueous solution. These liposomes can be utilized to encapsulate various types of drug compounds. Lipophilic drugs can be captured within the lipophilic bilayer of the liposome. Water-soluble drugs can be encapsulated within the core of the liposome. Liposomal delivery can be used for improving solubility, changing biodistribution, allowing for controlled release, improving tissue penetration, overcoming multidrug resistance, and for facilitating targeted drug delivery [77].

As mentioned previously, liposomes are of significant interest in solubility due to their ability to deliver hydrophobic molecules by incorporating these molecules in the lipophilic bilayer. Paclitaxel is one poorly soluble agent that is currently being investigated for use with liposomes [72]. The liposomal paclitaxel showed comparable activity when compared with standard paclitaxel in mouse models. Additionally the in vivo half-life of liposomal paclitaxel was noted to be longer than standard paclitaxel. The drug distribution was enhanced as well. Because of these findings, liposomes continue to be a popular choice for the delivery of poorly water-soluble drugs.

Improved biodistribution to certain tissues has been another reason for the interest in liposomal preparations of drugs [78]. Liposomes favor distribution in regions that are lipophilic or that have improved vascular permeability. Tumor cells are noted to have the latter finding due to increased angiogenesis. Utilizing liposomes allows the systemic toxicities of the drug to be diminished due to less penetration in healthy organs. By improving biodistribution in this manner, liposomes allow for an improved therapeutic index. Additionally, the liposomal preparations can allow for locoregional therapy. This therapy works by injected the drug in certain region with the hope that the highest concentrations of the drug will remain in that area. Due to their lipophilic nature, liposomes will tend to remain in a specific tissue compartment longer than water-soluble compounds. A liposomal preparation of doxorubicin (Doxil) has been developed due to these concepts [79]. The liposomal formulation of doxorubicin was developed to reduce the number of systemic side effects sustained from standard administration of free doxorubicin. Using the liposomal delivery system allowed for better tumor absorption of the drug and changed the pharmacodistribution of the drug so that there were reduced concentrations in commonly damaged organs (e.g., cardiac tissue) [80].

As noted previously, liposomes can change the biodistribution of a drug. Along with this concept is the finding that liposomes can actually be used to overcome some biological barriers. By overcoming these barriers, liposomes are of special interest in helping to overcome cancers with multidrug resistance (MDR). The size of liposomes can be manipulated to make either small or large vesicles. By using smaller liposomes, one can deliver drugs into cells and potentially overcome the efflux mechanisms found in MDR cancer cells [77]. They can also be manipulated to help deliver the cells to specific compartments of the cell. This finding is quite intriguing and continues to be a significant area of research.

Controlled release continues to be an important concept in anticancer drug delivery. Liposomes are also great candidates for this technology. As mentioned above, liposomes can alter the biodistribution and in vivo half-life of a drug. This allows for controlled release of agents. Many factors come into play in controlled release, but it is important to know that due to their ability to be easily modified, liposomes continue to play a role in controlled release of drugs. An easy way to think of this is that in standard chemotherapy, the pharmacokinetics of the drug may allow for toxic concentrations of the drug to be reached immediately after infusion. Liposomes can delay this toxicity due to the entrapment of the drug in the liposomal system. Additionally, the drug will not be metabolized as fast when it is in the liposomal system. This can allow for a longer half-life after administration. In addition to these concepts, liposomes can also be formulated to be "sensitive" to specific environmental triggers. Changes in pH, temperature, and even the chemical environment in the target tissue can be "programmed" into the liposome to stimulate release [81]. These ideas in combination can allow for longer exposure to the drug after infusion and focused release in the target tissues. Utilizing these concepts can allow for higher concentrations of the drug to sometimes be used without increasing systemic side effects.

Targeted drug delivery is currently the area of most interest in anticancer drug development [82]. It is no surprise that liposomes have been an area of investigation regarding this field. Liposomes are essentially biological membranes that contain phospholipids and potentially other cellular components. In addition to phospholipids, the human cellular phospholipid bilayer contains various proteins and carbohydrate compounds. In human cells, these proteins and carbohydrates are used for cellular signaling and messaging. One can apply these findings to liposomes as well. Just as human cellular membranes can contain signaling proteins and carbohydrates, so can liposomes. A number of different ligands are under investigation for their use in liposomes. Antibodies, cytokines, hormones, growth factors, and even toxins can be incorporated into liposomal membranes. This allows for targeted delivery of the lipsomally delivered drug to specific cells. One interesting utilization of this concept is the development of immunoliposomes. An immunoliposome targeted against the *HER2/ErbB2/neu* oncogene (anti-HER2 immunoliposome) is currently being evaluated [83]. Researches have incorporated doxorubicin in this targeted liposome. Studies of both in vitro and in vitro use of this immunoliposome showed appropriate binding to cells overexpressing this mutation. Furthermore the liposomes were selectively internalized into these cells. These preclinical studies utilizing the anti-HER2 immunoliposome with encapsulated doxorubicin showed improvement of therapeutic efficacy compared to treatment with anti-HER2 antibodies alone or in combination with standard chemotherapy. These promising results have provoked further research and the doxorubicin immunoliposome is currently in clinical trials. Successful findings such as this further stimulate the interest in using liposomes for targeted delivery agents.

Although liposomes are of great clinical and intellectual interest, there are problems that must be evaluated and overcome before they are utilized more extensively. First of all, size is a factor in liposomal delivery [84]. Larger sized liposomes (larger than 200 nm) are usually very short lived due to rapid host response and destruction of the liposome when it reaches the mononuclear phagocytic system (MPS). Smaller sized liposomes (<70 nm) are shorter lived after infusion but have improved penetrance into more tissues due to their size. Intermediate sized liposomes (70–200 nm) tend to last longer after infusion due to their ability to avoid the MPS. Depending on the specific need for the liposomal system, an appropriate size must be determined and evaluated. Following this, storability is also an issue. Liposomes may have an insufficient chemical stability to maintain long-term shelf life. Various manipulations can be made to the liposome to improve its stability. Nevertheless, this remains an important pitfall to overcome. Another significant problem is the host immune response to liposomes. Liposomes are a biological component and can evoke immune responses. When signaling proteins or antibodies are added to the membrane, the chance for immunologic side effects is increased. The benefit versus the risk of adding these targeting mechanisms must be heavily weighed when developing this technology.

### 8.5 Polymeric Drug Delivery Systems

Liposomes continue to be a biologically derived technology that has been of interest for the delivery of anti-

cancer agents. Along with this idea comes the concept of utilizing polymers, both biologic and synthetic, to deliver anticancer agents. The utility of polymers is similar to that of liposomes in that the improvement of solubility, change in biodistribution, increased half-life, ability to overcome MDR, and improvement in therapeutic index can be achieved. Additionally, polymers can be easily manipulated to encorporate targeting moieties to allow for targeted drug delivery. Finally, polymers have the ability to be manipulated into many forms. This allows for the development of implants, microspheres, biogels, and even tablet formulations of polymer–drug conjugates.

In 1975, the idea of creating polymer–drug conjugates was first published by Ringsdorf [85]. His model hypothesized that the polymeric–drug conjugate systems could be used to improve pharmacokinetics of the drug attached to the polymer. Additionally, his hypothesis included the concept that a targeted delivery system could be developed using this system by including a homing moiety on the polymeric system. With the rise in interest in targeted delivery systems, this idea has been advanced further. As work has progressed utilizing polymeric–drug conjugates, studies have been showing that therapeutic efficacy can be improved by penetrance into target tissues and extended half-life of the compound [86]. Additionally, the number of side effects can be decreased. The stability of the drug compound can be improved utilizing polymeric–drug conjugate systems as well. Of course, these polymeric systems also allow for the addition of targeting moieties to the polymeric backbone. It is no surprise that a number of these agents have been developed encorporating a number of chemotherapeutic agents [87].

Perhaps the best example of the polymer–drug conjugate systems is the synthetic polymer–drug conjugate poly(L-glutamic acid) with paclitaxel (PG-TXL, CT-2103; Xyotax) [88]. PG-TXL was initially created to overcome the poor solubility of paclitaxel and the hypersensitivity reaction of the standard administration of paclitaxel in Cremaphor EL. The PG-TXL polymer–drug conjugate involves the conjugation of paclitaxel to the synthetic polyamino acid, poly(L-glutamic acid) through the 2′-hydroxyl group via an ester linkage. This conjugation increases the water solubility of paclitaxel significantly. In preclinical studies this polymer–drug conjugate displayed improved antitumor activity and decreased side effects compared to standard paclitaxel delivered in Cremaphor EL [89]. PG-TXL has been evaluated in mouse models of human breast cacners, ovarian cancers, and lung cancers and has shown significant activity. Biodistribution was also noted to be different with PG-TXL with up to five times greater tumor exposure when compared to Cremophor-EL-delivered paclitaxel. The amount of time that the tumor was exposed to the drug was increased almost six times more than the standard preparation when the PG-TXL compound was injected into mouse tumors. Overall, the improved therapeutic index of PG-TXL when compared to the standard preparation was determined to be due to enhanced permeability and retention (EPR).

Furthermore, PG-TXL is of special interest due to its use in several phase I, II, and III clinical trials. Phase I and II trials in a broad range of tumor types including gastric, esophageal, breast, and ovarian cancer showed improved side effect profiles in regards to alopecia, nausea, vomiting, and especially hypersensitivity reactions [88]. The STELLAR 2, 3, and 4 phase III clinical trials compared efficacy of PG-TXL compared to several standard chemotherapy regimens for non-small-cell lung cancer (NSCLC) [90]. In these studies, the PG-TXL showed no statistical improvement in overall survival when compared to the standard treatments. Currently, there is a trial underway that is comparing the use of PG-TXL in combination with carboplatin and comparing this to standard paclitaxel and carboplatin.

Although PG-TXL has been the most successful polymer–drug conjugate to date, there are a number of other potential candidates. Many of these conjugates are in preclinical studies, but a number have made it to clinical trials. The polymers *N*-(2-hydroxypropyl) methacrylamide (HPMA), polyelthyelene glycol (PEG), and cyclodextrin-based polymers are currently being used to deliver a number of standard chemotherapy drugs in clinical trials. Some polymer–drug conjugates and references to their clinical studies are outlined in Table 16.

Polymer–drug conjugates continue to be very interesting candidates for improving anticancer drug delivery. Their ability to be modified to target specific tumor types will solidify their acceptance into the anticancer drug delivery field. Furthermore, their uncanny ability to be formulated into various release systems will allow more opportunities to be able to utilize these

**TABLE 16 Polymer–Drug Conjugates in Clinical Trials**

| Polymer–Drug Conjugate |
|---|
| HPMA–Doxorubicin |
| HPMA–Doxorubicin–galactosamine |
| HPMA–camptothecin |
| HPMA–TXL |
| Polymeric micelles |
| PEG–camptothecin |
| PEG–SN38 |
| Cyclodextrin-based polymer–camptothecin (CPT) |

drugs beyond the standard intravenous approach. Future directions that will continue to be explored will be the use of conjugates that can be used in combination with radiotherapy and ultrasound for controlled-release, polymer–prodrug conjugates in conjunction with polymer-directed enzymes, and the use of polymer–multidrug conjugates to deliver combination chemotherapies.

## 8.6 Prodrugs

Another aspect of drug delivery research has been focusing on the development of prodrugs. Prodrugs are essentially an inactive or less active form of a drug compound. The idea behind prodrugs is that they will become the active drug under the right circumstances. By being inactive or less active initially, the pharmacokinetics and biologic activity can be altered. The thought processes that led to the development of prodrugs focused on changing biodistribution, solubility, absorption, metabolism, and excretion. Specifically, many prodrugs were created so that a drug could be administered via an oral route versus the traditional intravenous approach.

Prodrugs typically require one to many metabolic steps to become active. This can involve spontaneous changes or, more commonly, enzyme-mediated changes. Due to this, the prodrug must remain similar in structure to the final drug compound. In order to reach the final drug's structure, the prodrug must contain chemical bonds that can be cleaved or changed to eventually change the prodrug into the drug. These reactions are typically done by enzymes present in the host but can occur spontaneously. Once the drug is converted, it must be able to be absorbed by the target cells. Additionally, it is ideal if the prodrug is not excreted prior to conversion to the drug. Furthermore, one must be sure that during the reactions to make the prodrug into the drug compound that minimal or no toxic metabolites are created as side products of the reactions.

There are two classifications of prodrugs, types I and II [91]. Type I prodrugs are converted to the final active drug intracellularly. Type II prodrugs are converted extracellularly. These classifications are further categorized by the subtypes A and B. The subtypes IA and IB determine whether or not the prodrug is converted at the site of therapeutic action. Type IA prodrugs are converted at the therapeutic target (tumor cells), while type IB prodrugs are converted in the metabolic tissues (liver, lung, kidney, etc.). A type IA and IB category can both apply to some prodrugs when the site of the therapeutic target and conversion is the same. Because type II drugs are converted extracellularly, the subtypes A and B have different meanings. Type IIA prodrugs are typically converted when they are in the gastrointestinal (GI) tract after they have been taken orally. Type IIB prodrugs are absorbed through the GI tract and converted into the drug when they reach the systemic circulation.

Because tumor cells have altered metabolic activities when compared to normal, healthy cells, the use of prodrugs has become of great interest. As the fields of genomics, proteomics, data mining, and HTS have evolved, we are seeing the potential benefits of prodrugs. Our knowledge of the intracellular enzymatic process has grown exponentially. These intracellular processes can be exploited by the use of appropriately created prodrugs. By creating a prodrug that will be selectively activated by tumor cells, tumor-targeted drug delivery can be achieved.

One successful example of a prodrug is the agent capecitabine. Capecitabine is a prodrug of 5-fluorouracil (5-FU) [92]. It is currently registered for use as a monotherapy in colorectal cancers and taxane- and anthracycline-resistant breast cancers. It has also been approved for use in combination with docetaxel in locally advanced or metastatic breast cancer. Capecitabine is a type IA prodrug. The drug is an oral formulation that can be easily administered in an outpatient setting. After oral administration, capecitabine is taken up in the GI tract and converted into two major metabolites, 5′-deoxyfluorocytidine (5′-DFCR) and then 5′-deoxy-5-fluorouridine (5′-DFUR). The 5′-DFCR is created in the liver by carboxylesterase enzymes. The 5′-DFUR is created from the 5′-DFCR by cytidine deaminase, which is present in the liver and tumor tissue. Once the 5′-DFUR has been created, it is then converted into 5-FU intracellularly by thymidine phosphorylase. Thymidine phosphorylase is often overexpressed in tumor tissue. There is a small amount of the prodrug that is converted to 5-FU directly after absorption. However, the amount of 5-FU in the systemic circulation is minimal and can be decreased by concomitant food intake with drug administration.

Capecitabine has been quite successful due to its tumor selectivity and reduced systemic side effects when compared to standard 5-FU. It has begun to replace 5-FU in many chemotherapy regimens and may someday replace it entirely. Due to its success, more prodrugs are being developed. With expanding knowledge of cancer genomics and proteomics in combination with data mining, HTS, and bioinformatics, we can expect to see many more prodrugs like capecitabine coming through the research pipeline.

## 8.7 Nanoparticles

Nanotechnology and more specifically nanomedicine have been important areas of medical research in current years. In regards to anticancer drug research

perhaps the most important area has been the area of nanoparticles. There are a number of synthetic and biologic nanoparticles available today. A nanoparticle is typically defined as a particle with its dimensions between 1 and 100 nm in any dimension [93]. This definition has changed over the years and the dimensions have been expanded somewhat. Examples of nanoparticle systems being investigated currently include liposomes, dendrimers, micelles, polymer–drug conjugates, silica nanoparticles, quantum dots, and even cellular proteins. The utility of nanoparticles is similar to other areas of drug delivery and includes improved therapeutic index, decreased systemic side effects, altered biodistribution, and the ability to be modified for targeted drug delivery. Perhaps the best way to show the utility of nanoparticles is to give an example.

Abraxane is a novel Cremophor-EL-free, albumin-bound, 130-nm form of the chemotherapeutic agent paclitaxel [94]. This agent is created by the high-pressure homogenization of paclitaxel in the presence of human albumin. It results in a nanoparticle colloidal suspension with paclitaxel–albumin nanoparticles ranging from 130 to 150 nm in mean diameter. Due to their size and water solubility, the albumin-based nanoparticles are easily soluble in an aqueous solution and can be administered intravenously without significant Cremaphor-EL-related side effects. In addition to this, the nanomolecules exploit the gp60 receptor. The gp60 receptors are specific for albumin as a ligand [95]. They allow for the intracellular transport of albumin. Because these receptors are commonly found in endothelial cell walls, they are often present in larger concentrations in the area of a tumor. This is due to increased angiogenesis to feed tumor growth. As a result, the albumin-bound paclitaxel has greater intratumor concentrations than an equivalent dose of Cremaphor-EL-delivered paclitaxel. Due to this finding, Abraxane requires less volume and administration time than standard paclitaxel. Additionally, the lack of Cremaphor EL eliminates the need for anaphylaxis prevention with steroid pretreatment.

Abraxane displayed the utility of nanoparticles in clinical oncology when it became approved by the U.S. Food and Drug Administration (FDA) in 2005 for the treatment of relapsed or refractory breast cancer. Its success has stimulated further work in the area of nanoparticle anticancer drug delivery systems. Albumin-bound formulations of docetaxel and rapamycin are currently being researched. Further work is currently being done with a variety of chemotherapeutics using a number of nanoparticles.

### 8.8 Conclusions

As if the discovery of the drug compounds was not vast enough, drug delivery remains a vital part of developing a potential anticancer drug. Drug formulation is an initial hurdle that must be overcome to ensure that the drug will maintain its integrity, activity, and effectiveness. When drug formulation proves to be difficult, other drug delivery options must be pursued. Advances such as liposomal delivery, polymer–drug conjugates, prodrugs, and nanoparticles are providing new techniques to facilitate drug delivery. In addition to facilitation of drug delivery, these options are giving researchers the ability to adapt tumor-specific targeting mechanisms to cytotoxic agents. The field of drug delivery will likely thrive along with the fields of drug discovery in the years to come.

## 9 IN VITRO CELL LINES

As with many other fields of drug discovery, cancer drug discovery begins in the laboratory. Before a potential chemotherapeutic agent is even considered for use in animal research, it must first be tested in vitro. By studying the effects of potential anticancer agents in vitro, researchers can determine if further work should be done in in vivo studies.

Cancer cell lines provide the first step in the evaluation of potential anticancer agents. This process is essentially a screening procedure in which a potential drug is put through several different steps to determine if it is worthy of further development. A good example of this type of screening can be explained by the National Cancer Institutes (NCI) Development Therapeutic Program [96]. In this screen, a potential cancer agent is first put through a prescreen in which it is exposed at a fixed concentration to three different cancer cell lines for a 48-h period. If there is any activity in any of these prescreen cell lines, the compound is then moved on to be tested in the more diverse NCI 60 cell lines. These lines consist of lung, colon, melanoma, prostate, ovarian, breast, and kidney cancer cell lines. The composition and methods involved in the NCI 60 cell lines is explained later. Depending on the distribution of activity in each tumor type, the potential drug may be tested in further related cell lines or it may progress on to in vivo experimentation.

In both in vitro and in vivo testing the main outcome being evaluated is drug activity on the tumor cells. Drug activity is assessed in in vitro studies by evaluating growth inhibition or cytotoxicity. Assessing for this involves the use of a quantified technique in which to exactly measure and compare the amount of inhibition or cytotoxicity. Colorimetric assays have grown to become the standard for these measurements.

There are a multitude of colorimetric assays currently available. They all utilize the same basic idea however. With the advent of fiber-optic, multiwell

sensors integrated with computer interfaces, these assays can be rather efficient and precise. This coupled with the popularization of multiwell microtiter plates allows for rapid analysis of multiple drug activity assays at once. The basic premise of colorimetric assays starts with culturing cells in multiwell microtiter plates. Now 96-well microtiter plates are often used for culturing. In each well there will be a specified number of cells that are allowed to be grown. This way each plate has the same or very close to the same number of cells present to be tested. Once the cells are allowed to grow to the specified number, various concentrations of the anticancer agent is applied to each microtiter plate. One row or group of cell cultures is kept as a control and is not exposed to the drug. After a specified time period has elapsed, the cell cultures are ready to be evaluated for drug activity. At this point the colorimetric assay is utilized. Typically this works by adding a chemical or chemicals to each well that is absorbed by the remaining, living cells. This chemical will then cause the living cells to exhibit a certain color. The wells containing the cells are then placed into a computer interfaced, fiber-optic spectrophotometer, which will send a specified wavelength of light through the wells. The sensor then evaluates how much of the light is absorbed in each well. This data is stored in the computer and can be extrapolated to determine the percentage of living cells in each well. This data can thus show the effects of varying concentrations of the potential anticancer agent on specific cancer cell lines.

An excellent example of this method can be shown by MTT assays currently used in our lab. MTT [3-(4,5 dimethylthiazol-2-Yl)-2,5-diphenyltetrazolium bromide] assays work through the intracellular reduction of tetrazolium salts to purple-colored formazans [97]. This reaction occurs in living cells in culture and allows for easy spectrophotometer analysis at a wavelength of 500–600 nm. MTT assays are quite beneficial as they can be easily automated and are quite precise. An example of the utility of these assays can be shown by its use in the evaluation of MT-103, an experimental compound derived from a topological computer model, currently being evaluated in our lab. Using the MTT assay, we evaluated varying concentrations of this drug on H226 lung cancer cell lines over a 48-h period. The type of data obtained from the assay can be seen in Figure 1.

As indicated from the example above, one can compare the activity with varying concentrations of the drug. This example indicates that as the concentration of the drug was increased, more activity was noted. Most often a trend of more inhibition or cytotoxicity can be noted with increasing concentration of the drug. However, in some cell lines, there may be no discernable activity of the agent. Other factors such as drug solubility, bacterial or viral infection of the cell lines, cell mutation, faulty culture media, inappropriate measurement of drug, and stability of the colormetric agent can play a role in causing error in colormetric assays. Regardless of this, they remain standard in evaluating drug activity in in vitro studies.

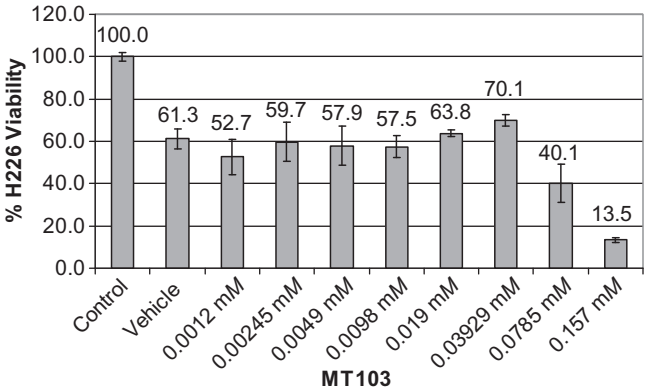


**Figure 1** H226 (5000 cells/well) + HT103 in DMSO alone after 24 h.

As mentioned previously, the first step in anticancer drug testing involves the use of in vitro testing. This utilizes various cancer cell lines. The current trend is to utilize modern technologies to allow for rapid and standardized methods to run and evaluate these experiments. Colormetric assays allow for this rapid and standardized approach. The NCI 60 cancer cell lines are the best example of this and warrant further discussion.

## 10 NCI 60 CANCER CELL LINES

Up until the late 1980s, drug discovery was mostly based on evaluating drug activity on transplantable murine tumor models. This had been proving useful for finding drugs to treat leukemia. However, the more common human solid tumors, breast, lung, colon, and prostate were proving difficult targets for anticancer agents. Due to this problem, the need for a standardized and effective method to test the activity of drugs on these cell types was present. Specifically, interest in the treatment of lung cancer initiated the development of human lung cancer cell lines that were expanded to include other tumor types, which now compose the present-day NCI 60 cell line [96].

The NCI 60 cell line has proven to be a useful research tool in the past and is currently still vital today. As previously mentioned, the cell line was created as a means to find more agents for the treatment of human solid malig-

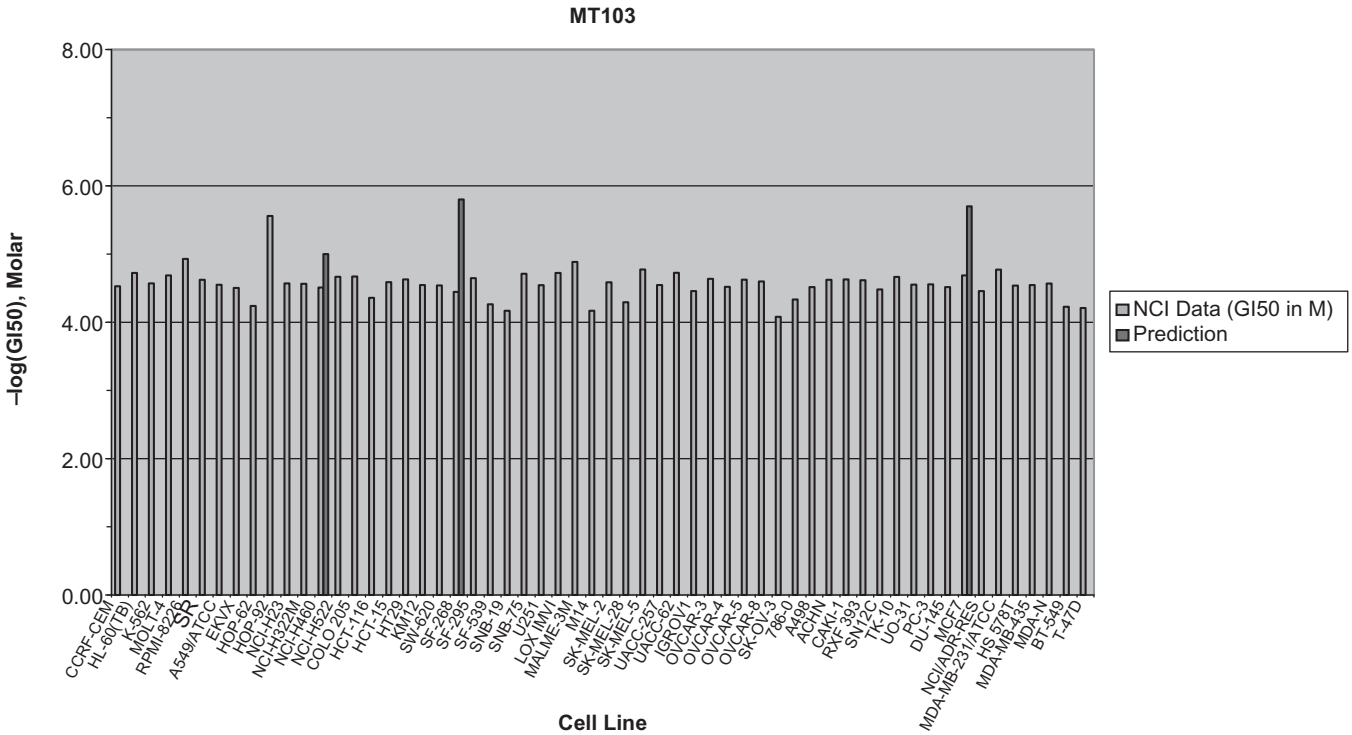


**Figure 2** Example evaluation of compound activity (here MT103) against NCI 60 cell lines.

nancies. It initially found its roots in lung cancer cell lines. Once these lines were created, the need for control groups was understood. Fibroblasts are normal human cells that are resistant to almost all types of chemotherapeutic agents. On the other hand renal epithelial cells are overly sensitive to most anticancer agents. After these controls were added, more human solid tumor types were added. Finally the NCI 60 cell line consisted of nine separate groups of human solid tumors: breast, lung, colon, central nervous system (CNS), prostate, renal, melanoma, ovarian, and leukemia.

The basic principle of the cell line was to have a standardized and efficient way to evaluate the growth inhibitory or anticancer effects of different compounds on the cell lines. These cell lines have also proven to be useful in evaluating drug resistance patterns and molecular activity of drugs. The NCI 60 cell line has seen a great deal of change since its inception, but the basic principle of evaluating potential drug activity remains paramount.

Various different cancer cell lines have commonly been used in cancer research. When the NCI 60 cell line was initially developed, there were more than 1000 different cell lines investigated. This large number of cell lines, of course, defeated the purpose of the screen, so the most promising lines were chosen. Favorable properties of the cell lines chosen included acceptable growth rates, adaptability, resistance to infection, and variable drug sensitivity. Ultimately 60 different cell lines were chosen.

It is important to understand how the NCI 60 cell lines are used. The model is an in vitro screening system. The NCI 60 cell lines are plated in multiwell tissue culture dishes. The cells are allowed to grow to reach a specified number of cells per well. When this number has been reached, the experimental drug is added at five different doses for 48 h. Once the 48 h has passed, a colorimetric assay is used to determine the growth inhibitory or cytotoxic effects of the agent. An example of a reading from an NCI 60 cell line is shown in Figure 2.

The drug may show activity in one specific cell line over another, which can give details about its possible mechanism of action. It may also show broad activity over a number of different types of cell lines. Regardless of this, if the agent shows promising activity in any cancer cell lines, it will likely be moved on for further testing. This further testing may be further more detailed in in vitro studies to determine the exact mechanism of action of the drug or it may jump to the next stage, in vivo testing. Approximately 2500 compounds a year are tested using the NCI 60 cell lines. Of these compounds, about 2% will move on to in vivo studies [96]. Due to the large number of anticancer agents being evaluated every year, it is important to have solid screening tools. The NCI 60 cancer cell lines remain an invaluable screening tool.

## 11 IN VIVO STUDIES

After a compound is shown to have activity in an in vitro setting, testing in an in vivo setting is desired. Anticancer drugs cannot move directly into human research until appropriate information is obtained regarding toxicity, dosage, and administration. In vivo studies provide the arena to find this data. These studies can be done using a variety of animal subjects; however, the mouse is most commonly used. Due to their size, durability, and easiness to handle, mice have been used in in vivo studies for years.

### 11.1 Mouse Models

Mouse models can be developed through a variety of methods. Mouse studies initially were done using mice with spontaneous arising tumors or by chemically inducing a malignancy. Fortunately, the ability to create immunodeficient, nude mice has come to fruition [98]. The use of these mice allows the ability to transplant tumors to normal, healthy mice. These mouse tumor models are either murine allograft models or human xenograft models. Murine allograft models are essentially tumor models that involve the transplantation of a murine or mouse tumor to a nude, immunodeficient mouse. Human xenograft models involve transplanting human tumor cells to a nude, immunodeficient mouse.

These mouse tumor models typically involve the transplantation of a predetermined number of tumor cells to a site other than the primary site of the malignancy (i.e., the flank). Sometimes the cells will be implanted orthotopically at the site of the primary malignancy. However, this is not a commonly used approach due to difficulties in implantation and tumor size measurement. Because the tumor cells are implanted in an immunodeficient environment, they are able to proliferate without any host response. The size of the tumor is monitored closely until it reaches a predetermined size. Once the tumor reaches the necessary size, the mice are divided into control and treatment groups. The treatment group is given the drug or combination of drugs as determined by the study design. The control group is given a placebo, usually normal saline. Tumor growth is monitored closely throughout the study. Additionally, the mice are monitored for any adverse effects. The adverse effects seen in mice give important information for later studies. They can often provide some preliminary information on drug toxicity. Throughout the study, the tumor response to therapy and placebo is recorded. At the end of the study, the effect of treatment is compared to the effects of placebo. Another parameter that is evaluated at the end of the study is animal survival. This is often a measure of how toxic the drug is in an in vivo setting. Ultimately, the goals are to see that the drug had significant activity compared to the control in vivo and does not display significant toxicity at the utilized doses.

Mouse models also require a great deal of supervision. In the United States there are a number of advisory and regulatory committees that monitor the treatment of animals used in studies. These groups are present to ensure that the animals are not inappropriately harmed or mistreated in any way. They typically review studies before they can be initiated and will monitor the research process to ensure that appropriate care is given to the test animals.

Mouse models provide an additional step to evaluate the efficacy of an anticancer agent on various types of cancer. They also allow researchers to see how a drug will work in an environment with active organ systems and metabolism. Additionally, in vivo models will provide preliminary toxicity data. Unfortunately, these models are not perfect. Transplanted human tumors do not grow and proliferate as they do in their primary site in humans. Differences in anatomy and physiology at these sites are part of the reason. Interestingly, orthotopic human tumor models do not show similar growth either [99]. Due to this, it is important for promising agents to move on to clinical studies in humans before being finally approved for use. Despite this, mouse and other animal models will continue to play an integral role in the preclinical development of anticancer agents.

## 12 CLINICAL TRIALS

Clinical trials are major hurdles for a potential drug to overcome before it can be used clinically. These studies are not designed to make it difficult for an experimental agent to become an approved drug. They are designed to make sure that the drug is safe, effective, and appropriate for its intended use. Because they use human subjects, clinical trials are very heavily regulated. Institutional review boards (IRBs) are integral in the control of these studies. These committees work to approve, monitor, and review clinical trials to ensure that the subjects are not harmed or abused. Clinical trials are vital in anticancer research. They involve three major steps and are divided into phase I, phase II, and phase III clinical trials.

### 12.1 Phase I Clinical Trials

Anticancer agents go through a rigorous preclinical research process. If they show promising activity in in vitro and in vivo preclinical studies, they will often

move on to clinical studies. The first step is the phase I clinical trial.

Phase I clinical trials are created with the goal of finding the appropriate dosing of a drug or combinations of drugs [100]. They also help to evaluate initial pharmacologic and pharmacokinetic data. These trials are designed as dose escalation studies. In cancer clinical trials, only patient's with a known malignancy are allowed in the study. These patient's typically have end-stage disease and lack further treatment options. The major goal of these studies is to determine the maximum tolerated dose (MTD). The MTD is the maximum dose that can be achieved while still having an acceptable level of dose-limiting toxicity (DLT). When evaluating the MTD, one must account for patients that may have organ damage from their disease. Some drugs are metabolized and excreted via certain organs. For example, patients with kidney or liver damage from their disease may notice more toxicity from certain drugs. This must be accounted for prior to the study. These patients are often excluded from these studies. Otherwise, a number of different patients with different tumor types will be included in these studies.

The basis of designing these studies is to determine the correct dose escalation of the drug. By escalating the dose too rapidly, patients can be exposed to excessive toxicities. Escalating the dose too slowly can expose patients to ineffective levels of the drug and deny them adequate therapy for their cancer. The first step in designing these trials is to decide upon an appropriate starting dose. This dose should be relatively safe overall but must not be too low as to make it ineffective. The starting dose is typically extrapolated from doses used in previous animal studies. Toxicity is usually thought to be a function of body weight. This is relatively constant throughout species. The starting dose can be determined using the concept of allometric scaling. Allometric scaling allows for the comparison of two different species of different sizes. This concept has been used to predict the rough equivalent MTD in humans when compared to mice. The murine equivalent $LD_{10}$ ($MELD_{10}$) is the dose associated with 10% lethality in mice. This dose is approximately the MTD in humans. When a second species is needed to determine the MTD, the canine is commonly used. In this situation the lethal dose ($LD_{10}$) is used rather than the $MELD_{10}$. The typical starting dose for phase I trials is usually one-tenth the estimated MTD based on the $MELD_{10}$ or one-third the $LD_{10}$ [101].

The difference that sets phase I clinical trials apart from one another is the rate of dose escalation. There are a number of different models for dose escalation, but they can be easily categorized as single ascending dose (SAD) studies and multiple ascending dose (MAD) studies. The SAD studies involve starting at a single starting dose. The patients are closely monitored after getting the dose for side effects. Pharmacokinetic data is also evaluated after the dose. If there are no adverse effects and there is adequate pharmacokinetic data, the dose will then be escalated in a new group of patients. This will be continued until the MTD is reached. MAD studies involve giving multiple doses of the drug. Patients are given multiple low doses of the drug. Pharmacokinetic data is taken through these doses at predetermined points. These types of studies are useful for understanding pharmacokinetics and pharmacodynamics.

### 12.2 Phase II Clinical Trials

Once the MTD has been determined from phase I studies, the next step is to evaluate its activity. Phase II trials are the next evaluation process. These studies usually have larger patient groups than phase I trials. Phase II trials are designed to evaluate the efficacy of the drug and to further evaluate its safety [102]. Tumor response to the drug is usually the major clinical endpoint in these studies. Ultimately, these studies are used to discern between potentially active agents and agents with minimal or undesired activity. Drugs are essentially "weeded out" during these studies. Successful drugs may then be further evaluated in follow-up phase II studies or they may move on to phase III studies.

### 12.3 Phase III Clinical Trials

The final step before being an approved agent is the phase III clinical trial. These studies are the largest and most time-consuming studies. They are often randomized, controlled, multicenter trials [102]. These studies are also usually very expensive and difficult to run. They require a great deal of participation from both the patients and the researchers. These studies are utilized to determine just how effective the drug is compared to the gold standard. Along with testing the drug's effectiveness, phase III trials provide a larger group to investigate adverse effects. Less common adverse effects are often picked up during phase III trials due to the large number of patients.

Phase III studies will often test a drug compared to a control. In clinical trials in general, this control can either be a placebo or the standard therapy used in practice. In cancer clinical trials, it is unethical to deny one group a potentially beneficial therapy, so the use of placebo is unethical. For this reason, another drug, often the standard therapy, will be used. If the test drug shows adequate efficacy compared with the control and has tolerable side effects, it will often move on to the drug approval phase. The drug approval phase involves the

submission of a detailed document explaining the methods and results of all research utilizing the drug. Additionally, information must be given regarding the mechanism of action, proper formulation, dosing, and correct administration. If the drug shows significant efficacy based on well-designed and well-controlled studies, a reasonable side effect profile, and a feasible route of administration, it will likely be approved.

The path to drug approval is a very long process, especially in anticancer drug research. Preclinical studies can take around 6–8 years before the drug is ready to enter clinical studies. Clinical studies can take a similar amount of time before adequate data is obtained. For this reason, there are a number of policies and procedures that have been developed to ensure that a drug can be available to the market as rapidly as possible. These steps include: the Fast Track Program and Rolling NDA Submission, the Accelerated Approval, Special Protocol Assessment, and Priority Review [102]. These procedures are typically reserved for agents that show significant benefit over standard therapy in initial studies. By using these pathways, a drug can move through clinical studies more rapidly. To better understand each of these procedures, please see their corresponding references.

The clinical trial process is lengthy and information intensive. These studies provide the information needed to get a drug to move on to clinical practice. Vital information pertaining to efficacy, adverse effects, dosing, and improvement of survival can be obtained from clinical studies. It is important that these studies continue to be heavily regulated and monitored to ensure that patient's are getting the best care possible and not being harmed. Clinical trials will continue to play a pivotal role in anticancer drug discovery, therefore understanding the processes and procedures involved is important for all researchers.

## 3 FUTURE DIRECTIONS

Anticancer drug research has a vast array of technologies at its disposal. Many promising agents have been found over a relatively short period of time. By utilizing new technologies, we are able to realize many different ways to treat cancer. The protein kinase inhibitors are perhaps the newest class of drugs that illustrate our enhanced understanding of cancer biology. Fortunately, this is not the end of the road for anticancer research. Several new avenues are just beginning to be explored.

Antisense drugs are one class of compounds that are being explored. By targeting specific RNA sequences with designed complementary oligonucleotides, one can target specific genetic expressions in cancer cells [3]. This "gene therapy" can be used to inhibit production of abnormal proteins by the erroneous RNA in cancer cells. Additionally, gene therapies may be able to be used to target cancer cells and create proteins that induce apoptosis. This is an exciting area of research that will continue to be explored as our knowledge of genetics continues to expand.

In addition to this we will continue to see more sophisticated methods of drug delivery. The advancement of nanomedicine is perhaps one of the reasons why drug delivery is becoming such an interesting area. For instance, the development of protein cages, based on the idea of viral capsids, are currently being evaluated for their use in drug delivery. These delivery mechanisms are similar to liposomal delivery but may have the ability to be more easily manipulated or altered for targeted delivery. Concepts such as this will continue to stimulate further research in the area of drug delivery.

## 14 CONCLUSION

The field of anticancer drug research is vast. As our understanding of cancer biology grows, so will our ability to create new cancer drugs. By keeping the interests of the patient in mind, we must remember that the most effective agents are those that have maximal efficacy in addition to limited side effects. We may never be able to eradicate cancer, but we can hope to improve the quality of life of those who have it. As long as cancer remains a major cause of morbidity and mortality, anticancer drug research will continue to grow.

## REFERENCES

1. WHO (2006). World Health Organization Cancer Factsheet. February. Factsheet No. 297.
2. Dimasi JA (2001). New drug development in the United States from 1963 to 1999. *Clin Pharmacol Ther 69*, 286–296.
3. Adjei AA, Rowinsky EK (2003). Novel anticancer agents in clinical development. *Cancer Biol Ther 2*(4:Suppl. 1), S5–S15.
4. Cragg GM, Newman DJ, Weiss RB (1997). Coral reefs, forests, and thermal vents: The worldwide exploration of nature for novel antitumor agents. *Semin Oncol 24*, 156–163.
5. Gelmon K (1994). The taxoids: Paclitaxel and docetaxel. *Lancet 344*, 1267–1272.
6. Goodman J, Jordan V (2001). *The Story of Taxol: Nature and Politics in the Pursuit of an Anti-Cancer Drug*. Cambridge: Cambridge University Press, pp. 81–97.
7. Holton RA, Liu JH, Gentile LN (1992). Semisynthesis of taxol. Second National Cancer Institute Workshop on Taxol and Taxus. September 1992 Alexandria, Virginia.

8. Bergmann W, Feeny RJ (1951). Contributions to the study of marine products. XXXII. The nucleosides of sponges. *J Org Chem 16*, 981–987.
9. Bergmann W, Burke DC (1955). Contributions to the study of marine products. XXXIX. The nucleosides of sponges. III. Spongothymidine and spongouridine. *J Org Chem 20*, 1501–1507.
10. Mohammad RM, Beck FW, et al. (1998). Potentiation of 2-chlorodeoxyadenosine activity by bryostatin 1 in the resistant chronic lymphocytic leukemia cell line (WSU-CLL): Association with increased ratios of dCK/5′-NT and Bax/Bcl-2. *J Biol Chem 379*, 1253–1261.
11. Pagliaro L, Daliani D, Amato R, et al. (2000). A phase II trial of bryostatin-1 for patient with metastatic renal cell carcinoma. *Cancer 89*, 615–618.
12. Zonder JA, Shields AF, Zalupski M, et al. (2001). A phase II trial of bryostatin 1 in the treatment of metastatic colorectal cancer. *J Clin Cancer Res 7*, 38–42.
13. Bai R, Pettit GR, Hamel E. Dolastatin 10, a powerful cytostatic peptide derived from a marine animal. Inhibition of tubulin polymerization mediated through the vinca alkaloid binding domain. *Biochem Pharmacol*. 1990 Jun 15; *39*(12), 1941–1949.
14. Rinehart KLJ, Gloer JB, Hughes RGJ, et al. (1981). Didemnins: antiviral and antitumor depsipeptides from a Caribbean tunicate. *Science 212*, 933–935.
15. Garcia-Fernandez LF, Losada A, Alcaide V, et al Aplidin induces the mitochondrial apoptotic pathway via oxidative stress-mediated JNK and p38 activation and protein kinase C delta. *Oncogene*. 2002 Oct 24; *21*(49), 7533–7544.
16. Lopez-Macia A, Jimenez JC, Royo M, et al. (2001). Synthesis and structure determination of kahalalide F (1,2). *J Am Chem Soc 23*, 11398–11401.
17. Krug LM, Miller VA, Kalemkerian GP, et al. (2000). Phase II study of dolastatin-10 in patients with advanced non-small cell lung cancer. *Ann Oncol 11*, 227–228.
18. Maroun JA, Stewart D, Verma S, et al. (1998). Phase I clinical study of didemnin B. A National Cancer Institute of Canada Clincal Trials Group study. *Invest New Drugs 16*, 51–56.
19. Maroun JA, Belanger K, Seymour L, et al. (2006). Phase I study of Aplidine in a dailyx5 one-hour infusion every 3 weeks in patients with solid tumors refractory to standard therapy. A National Cancer Institute of Canada Clinical Trials Group study: NCIC CTG IND 115. *Ann Oncol 17*(9), 1371–1378.
20. Pardo B, Paz-Ares L, Tabernero J, et al. (2008). Phase I clinical and pharmacokinetic study of kahalalide F administered weekly as a 1-hour infusion to patients with advanced solid tumors. *Clin Cancer Res 14*(4), 1116–1123.
21. Johnson IS, Armstrong JG, Gorman M, Burnett JP (1963). The vinca alkaloids: A new class of oncolytic agents. *Cancer Res 23*, 1390–1427.
22. Wall ME, Wani MC, Cook CE, et al. (1966). Plant antitumor agents. I. The isolation and structure of camptothecin, a novel alkaloidal leukemia and tumor inhibitor from camptotheca acuminate. *J Am Chem Soc 88*, 3888–3890.
23. Zhou DC, Zittoun R, Marie JP (1995). Homoharringtonine: an effective new natural product in cancer chemotherapy. *Bull Cancer 82*, 987–995.
24. Kayitalire L, Thomas F, Le Chevalier T, et al. (1992). Phase II study of a combination of elliptinium and vinblastine in metastatic breast cancer. *Invest New Drugs 1*, 303–307.
25. Dowlati A, Robertson K, Cooney M, et al. (2002). A phase I pharmacokinetic and translational study of the novel vascular targeting agent combretastatin a-4 phosphate on a single-dose intravenous schedule in patients with advanced cancer. *Cancer Res 62*, 3408–3416.
26. Schwartzmann G, Ratain MJ, Cragg GM, et al. Anticancer drug discovery and development throughout the world. *J Clin Oncol* 2002 Sep 15; *20*(18 Suppl), 47S–59S.
27. Merrifield RB (1963). Solid phase peptide synthesis I. The synthesis of a tetrapeptide. *J Am Chem Soc 85*, 2149–2154.
28. Lam KS (1997). Application of combinatorial library methods in cancer research and drug discovery. *Anticancer Drug Design 12*, 145–167.
29. Newman DJ, Cragg GM (2007). Natural products as sources of new drugs over the last 25 years. *J Nat Prod 70*, 461–477.
30. Gray NS, Wodicka L, Thunnissen AWH, et al. (1998). Exploiting chemical libraries, structure, genomics in the search for kinase inhibitors. *Science 281*, 533–537.
31. Nacro K, Sigano DM, Yan S, et al. (2001). AN optimized protein kinase C activationg diacylglycerol combining high binding affinity (Ki) with reduced lipophilicity (logP). *J Med Chem 44*, 1892–1904.
32. Smith GCM, Jackson SP (1999). The DNA-dependent kinase. *Genes Dev 13*, 916–934.
33. Wilhelm S, Carter C, Lynch M, Lowinger T, et al. (2006). Discovery and development of sorafenib: A multikinase inhibitor for treating cancer. *Nat Rev Drug Discov 5*(10), 835–844.
34. Fox S, Farr-Jones S, Yund M (1999). High throughput screening for drug discovery: Continually transitioning into new technology. *J Biomol Screen 4*, 183–186.
35. Schmid I, Sattler I, Grabley S, Thiericke R (1999). Natural products in high throughput screening: Automated high-quality sample preparation. *J Biomol Screen 4*, 15–25.
36. Hodgkin E, Andrews-Cramer K (2000). Compound collections get focused: Smaller, well-designed compound collections make better use of HTS resources. *J Am Chem Soc 122*, 55–60.
37. Spencer R (1998). High-throughput screening of historic collections: Observations on file size, biological targets, and file diversity. *Biotechnol Bioeng 61*, 61–67.
38. Lenz G, Nas H, Jindal S (2000). Chemical ligands, genomics, and drug discovery. *Drug Disocv Today 5*, 145–156.

39. Lutz M, Menius A, Choi T, et al. (1996). Experimental design for high-throughput screening. *Drug Discov Today 1*, 277–286.
40. Zhung J, Chung T, Oldenburg K (1999). A simple statistical parameter for use in evaluation and validation of high-throughput screening assays. *J Biomol Screen 4*, 67–73.
41. Oprea T, Davis A, Teague S, et al. Is there a difference between leads and drugs? A historical perspective. *J Chem Inf Comput Sci* 2001 Sep-Oct; *41*(5), 1308–1315.
42. Stormo GD (2006). An overview of RNA structure prediction and applications to RNA gene prediction and RNAi design. *Current Protocols in Bioinformatics*. Chapter 12, Unit 12.1.
43. Gago F (2004). Modeling and simulation: A computational perspective in anticancer drug discovery. *Curr Med Chem Anti-Cancer Agents 4*, 401–403.
44. Rabow AA, Shoemaker RH, Sausville EA, Covell DG (2002). Mining the National Cancer Institute's tumor-screening database: Identification of compounds with similar cellular activities. *J Med Chem 45*(4), 818–840.
45. Budman DR, Calvert AH, Rowinsky EK (2003). *Handbook of Anticancer Drug Development*. Philadelphia: Lippincott Williams & Wilkins USA, p. 80.
46. Chen YP, Chen F (2008). Identifying targets for drug discovery using bioinformatics. *Expert Opin Ther Targets 12*(4), 383–389.
47. Slamon DJ, Godolphin W, Jones LA, et al. (1989). Studies of the HER-2/neu proto-oncogene in human breast and ovarian cancer. *Science 244*, 707–712.
48. Academy of Medical Sciences and The Royal Engineering. Systems Biology: A Vision for Engineering and Medicine. Available: http://www.raeng.org.uk/policy/engagement/pdf/Systems_Biology_Report.pdf. Accessed February 2007.
49. Guttmacher AE, Collins FS (2002). Genomic medicine—a primer. *N Engl J Med. 347*(19), 1512–1520.
50. Grogan L, Kirsch IR (1997). Genetic testing for cancer risk assessment: A review. *Oncologist 2*(4), 208–222.
51. Sjøholt G, Bedringaas SL, Døskeland AP, Gjertsen BT (2006). Proteomic strategies for individualizing therapy of acute myeloid leukemia (AML). *Curr Pharm Biotechnol 7*(3), 159–170.
52. Emilien G, Ponchon M, Caldas C, Isacson O, Maloteaux JM (2000). Impact of genomics on drug discovery and clinical medicine. *QJM 93*(7), 391–423.
53. Chung CH, Levy S, Chaurand P, Carbone DP (2007). Genomics and proteomics: Emerging technologies in clinical cancer research. *Crit Rev Oncol Hematol 61*(1), 1–25.
54. Quackenbush J (2006). Microarray analysis and tumor classification. *N Engl J Med 354*(23), 2463–2472.
55. Ramaswamy S (2004). Translating cancer genomics into clinical oncology. *N Engl J Med 350*(18), 1814–1816.
56. Lee CH, Macgregor PF (2004). Using microarrays to predict resistance to chemotherapy in cancer patients. *Pharmacogenomics 5*(6), 611–625.
57. Mancinelli L, Cronin M, Sadée W (2000). Pharmacogenomics: The promise of personalized medicine. *AAPS PharmSci 2*(1), E4.
58. Lee W, Lockhart AC, Kim RB, Rothenberg ML (2005). Cancer pharmacogenomics: Powerful tools in cancer chemotherapy and drug development. *Oncologist 10*(2), 104–111.
59. Langreth R, Waldholz M (1999). New era of personalized medicine: Targeting drugs for each unique genetic profile. *Oncologist 4*(5), 426–427.
60. Katzung BG (2007). *Basic & Clinical Pharmacology, 10th Ed*. New York: McGraw-Hill Medical, pp. 878–907.
61. Riggs BL, Hartmann LC (2003). Selective estrogen-receptor modulators—mechanisms of action and application to clinical practice. *N Engl J Med 348*(7), 618–629.
62. Fabbro D, Ruetz S, Buchdunger E, Cowan-Jacob SW, et al. (2002). Protein kinases as targets for anticancer agents: From inhibitors to useful drugs. *Pharmacol Ther 93*(2–3), 79–98.
63. Verweij J, de Jonge M (2007). Multitarget tyrosine kinase inhibition: [and the winner is . . . ]. *J Clin Oncol 25*(17), 2340–2342.
64. Costa LJ, Drabkin HA (2007). Renal cell carcinoma: New developments in molecular biology and potential for targeted therapies. *Oncologist 12*(12), 1404–1415.
65. Pang RW, Poon RT (2007). From molecular biology to targeted therapies for hepatocellular carcinoma: The future is now. *Oncolog 72*(Suppl 1), 30–44.
66. Deeks ED, Keating GM (2006). Sunitinib. *Drugs 66*(17), 2255–2266.
67. Scott LJ (2007). Bevacizumab: In first-line treatment of metastatic breast cancer. *Drugs 67*(12), 1793–1799.
68. Rao RD, Buckner JC, Sarkaria JN (2004). Mammalian target of rapamycin (mTOR) inhibitors as anti-cancer agents. *Curr Cancer Drug Targets 4*(8), 621–635.
69. Rini BI (2008). Temsirolimus, an inhibitor of mammalian target of rapamycin. *Clin Cancer Res 14*(5), 1286–1290.
70. Jonkman-deVries JD, Flora KP, Bult A, et al. (1996). Pharmaceutical development of investigational anticancer agents for parenteral use. *Drug Dev Ind Pharm 22*, 475.
71. Budman DR, Calvert AH, Rowinsky EK (2003). *Handbook of Anticancer Drug Development*. Philadelphia: Lippencott Williams and Wilkins, pp. 205–210.
72. Hennenfent KL, Govindan R (2006). Novel formulations of taxanes: A review. Old wine in a new bottle? *Ann Oncol 17*(5), 735–779.
73. Budman DR, Calvert AH, Rowinsky EK (2003). *Handbook of Anticancer Drug Development*. Philadelphia: Lippencott Williams and Wilkins, p. 213.

74. Spigel SC, Jones SF, Greco FA (2002). S-8184 vitamin E paclitaxel emulsion: Preclinical and phase I data. *Proc Am Soc Clin Oncol 20*, 406.
75. Bogdanova N, Lissianskaya A, Gorelov A (2003). Paclitaxel vitamin E emulsion: Phase I/II study of weekly administration in patients with non-small cell lung cancer, transitional cell, ovarian, or colorectal cancer. *Proc Am Soc Clin Oncol 21*, 988.
76. Carpenter JF, Pikal MJ, Chang BS, et al. (1997). Rational design of stable lyophilized protein formulations: Some practical advice. *Pharm Res 14*, 969–975.
77. Lian T, Ho RJ (2001). Trends and developments in liposome drug delivery systems. *J Pharm Sci 90*(6), 667–680.
78. Goyal P, Goyal K, Vijaya Kumar SG, et al. (2005). Liposomal drug delivery systems—clinical applications. *Acta Pharm 55*(1), 1–25.
79. Abraham SA, Waterhouse DN, Mayer LD (2005). The liposomal formulation of doxorubicin. *Methods Enzymol 391*, 71–97.
80. Waterhouse DN, Tardi PG, Mayer LD, Bally MB (2001). A comparison of liposomal formulations of doxorubicin with drug administered in free form: Changing toxicity profiles. *Drug Saf 24*(12), 903–920.
81. Voinea M, Simionescu M (2002). Designing of "intelligent" liposomes for efficient delivery of drugs. *J Cell Mol Med 6*(4), 465–474.
82. Medina OP, Zhu Y, Kairemo K (2004). Targeted liposomal drug delivery in cancer. *Curr Pharm Des 10*(24), 2981–2989.
83. Park JW, Kirpotin DB, Hong K, Shalaby R (2001). Tumor targeting using anti-her2 immunoliposomes. *J Control Release 74*(1–3), 95–113.
84. Abra RM, Hunt CA (1981). Liposome disposition in vivo. III. Dose and vesicle-size effects. *Biochim Biophys Acta 666*, 493–503.
85. Ringsdorf H (1975). Sturcture and properties of pharmacologically active polymers. *J Polym Sci Symp 51*, 135–153.
86. Matsumura Y, Maeda H (1986). A new concept for macromolecular therapies in cancer chemotherapy: Mechanism of tumoritropic accumulation of proteins and the antitumor agent SMANCS. *Cancer Res 6*, 6387–6392.
87. Li C, Wallace S (2008). Polymer-drug conjugates: Recent development in clinical oncology. *Adv Drug Deliv Rev 60*(8), 886–898.
88. Singer JW (2005). Paclitaxel poliglumex (XYOTAX, CT-2103): A macromolecular taxane. *J Control Release 109*(1–3), 120–126.
89. Li C, Yu DF, Newman RA, et al. (1998). Complete regression of well-established tumors using a novel water-soluble poly(lglutamic acid)-paclitaxel conjugate. *Cancer Res 58*, 2404–2409.
90. Paz-Ares L, Ross H, O'Brien M, Riviere A, et al. (2008). Phase III trial comparing paclitaxel poliglumex vs docetaxel in the second-line treatment of non-small-cell lung cancer. *Br J Cancer 98*(10), 1608–1613.
91. Wu KM, Farrelly JG (2007). Regulatory perspectives of Type II prodrug development and time-dependent toxicity management: Nonclinical Pharm/Tox analysis and the role of comparative toxicology. *Toxicology 236*, 1–6.
92. Schellens JH (2007). Capecitabine. *Oncologist 12*(2), 152–155.
93. Liu Y, Miyoshi H, Nakamura M (2007). Nanomedicine for drug delivery and imaging: A promising avenue for cancer therapy and diagnosis using targeted functional nanoparticles. *Int J Cancer 120*(12), 2527–2537.
94. Green MR, Manikhas GM, Orlov S, et al (2006). Abraxane®, a novel Cremophor ®-free, albumin-bound particle form of paclitaxel for the treatment of advanced non-small-cell lung cancer. *Ann Oncol 17*, 1263–1268.
95. Hawkins MJ, Soon-Shiong P, Desai N (2008). Protein nanoparticles as drug carriers in clinical medicine. *Adv Drug Deliv Rev 60*(8), 876–885.
96. Shoemaker RH (2006). The NCI60 human tumour cell line anticancer drug screen. *Nat Rev Cancer 6*(10), 813–823.
97. Hayon T, Dvilansky A, Shpilberg O, Nathan I (2003). Appraisal of the MTT-based assay as a useful tool for predicting drug chemosensitivity in leukemia. *Leuk Lymphoma 44*(11), 1957–1962.
98. Ovejera AA, Houchens DP (1981). Human tumor xenografts in athymic nude mice as a preclinical screen for anticancer agents. *Semin Oncol 8*, 386–393.
99. Wilmanns C, Fan D, O'Brian CA, et al. (1992). Orthotopic and ectopic organ environments differentially influence the sensitivity of murine colon carcinoma cells to doxorubicin and 5-fluorouracil. *Int J Cancer 52*, 98–104.
100. Eisenhauer EA, O'Dwyer PJ, Christian M, Humphrey JS (2000). Phase I clinical trial design in cancer drug development. *J Clin Oncol 18*(3), 684–692.
101. Fuse E, Kobayashi T, Inaba M, et al. (1995). Prediction of the maximal tolerated dose (MTD) and therapeutic effect of anticancer drugs in humans: Integration of pharmacokinetics with pharmacodynamics and toxicodynamics. *Cancer Treat Rev 21*, 133–157.
102. Budman DR, Calvert AH, Rowinsky EK (2003). *Handbook of Anticancer Drug Development*. Philadelphia: Lippencott Williams and Wilkins, p. 8–9.

# TYROSINE KINASE INHIBITORS IN CHRONIC MYELOID LEUKEMIA

Muhammad Mansoor Alam and Moshe Talpaz
*University of Michigan School of Medicine, Ann Arbor, Michigan*

## 1 IMATINIB

### 1.1 Introduction

The BCR-ABL protein tyrosine kinase is the product of a chimeric gene produced by the Philadephia (Ph) chromosome. Initially identified in 1960, the Ph chromosome results from a reciprocal translocation between the long arms of chromosomes 9 and 22. Chronic myeloid leukemia (CML) is a myeloproliferaive disorder that results from this acquired mutation affecting hemapoietic stem cells. Imatinib, also known as Gleevac, developed by Novartis, is a specific inhibitor of BCR-ABL tyrosine kinase. If untreated CML inevitably progresses to more aggressive accelerated and blast phases. Once in blast phase the median survival of patients is less than 6 months. Imatinib has substantial activity in the chronic phase, as well as significant activity in the accelerated phase and the blast crises phase of CML [1, 2].

### 1.2 Mechanism of Action and Development

There is a family of enzymes called protein tyrosine kinases that bind adenosine triphosphate (ATP) and catalyse the transfer of the γ-phosphate to the hydroxyl group of a tyrosine residue on a protein. In a signal transduction cascade these phosphorylated sites can then serve as binding sites for phosphorylation of other substrates. The kinases themselves are usually activated either by interaction of a ligand with a receptor located within the extracellular domain of the kinase or, alternatively, as part of an intracellular signaling pathway. The human genome encodes upto 800 serine/threonine and tyrosine protein kinases, all of which bind ATP in a highly conserved fashion within their catalytic domains. Molecules collectively referred to as kinase inhibitors that are currently available possess an inhibitory profile whereby they inhibit the target kinase with relative selectivity toward other often closely related kinases to give an acceptable side effect profile [3].

In vitro and animal studies have established that BCR-ABL alone is sufficient to cause CML, and mutational analysis has established that tyrosine kinase activity of the protein is required for its oncogenic activity [4]. The ABL gene encodes the tightly regulated nonreceptor protein tyrosine kinase cytosolic ABL (cAbl), which plays a fundamental role in regulating cell proliferation, adherence, and apoptosis [5]. In contrast, the BCR-ABL fusion gene encodes a constitutively activated kinase. This transformation results in hemopoeitic stem cells to exhibit deregulated clonal proliferation, reduced adherence to bone marrow stroma, and reduced apoptotic response to disruption in cellular genomic integrity. As a result immature cells of the myeloid lineage are released into the blood in various stages of maturation as well as populate the reticuloendothelial system organs such as the liver and the spleen [6, 7].

This has rendered BCR-ABL an attractive target for drug intervention in CML. Unlike membrane-bound receptor tyrosine kinases, BCR-ABL is a cytoplasmic kinase. It is constitutively active and does not depend

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

on ligand binding and subsequent receptor dimerization or activation. Because of this reason BCR-ABL activity cannot be modulated by means of receptor-based inhibitions. The most obvious approach in this situation would be based on an ATP-competitive inhibitor of the BCR-ABL, which will result in blocking the kinase activity, as a result stopping this enzyme cascade [8, 9].

After the identification of BCR-ABL as a suitable target, Novartis (based in Basel, Switzerland) screened its large and diverse compound archive, which led to the identification of a pyrimidine structure as an attractive lead molecule for activity against cABL. Further development of the molecule occurred after structure–activity studies led to *N*-acyl and *N*-aroyl amide derivatives. Subsequently, methyl substituent ortho to pyrimidinylamino group controlled activity against other kinases and at the same time enhanced potency against cABL. The resulting molecules were noted to be highly insoluble. This was overcome by introducing suitably placed hydrophilic moieties. This resulted in the development of imatinib.

Imatinib binds only to the inactive conformation of BCR-ABL and is a potent, competitive inhibitor of cytosolic BCR-ABL. Imatinib also inhibits cKit and platelet-derived growth factor (PDGFR)-β, but is devoid of activity against other kinases including the Src family of kinases. The chemical designation of imatinib mesylate is 4-[(4-methyl-1-piperazinyl)methyl]-*N*-[4-methyl-3-[[4-(3-pyridinyl)-2-pyrimidinyl]amino]-phenyl]benzamide methanesulfonate. (See Fig. 1.)

### 1.3 Pharmacokinetics

Imatinib is well absorbed orally. It is available in a formulation as a methane sulfonate salt, which has good water solubility. It has an excellent pharmacokinetic profile. Maximum plasma concentration is achieved within 2–4 h after ingestion. Imatinib is 95% bound to plasma proteins, mainly albumin and α-1-acid glycoprotein with a mean bioavailability of 98%. A 400-mg/day dose of imatinib achieves fairly predictable plasma concentrations. Peak levels are noted to be 2.6 ± 0.8 mcg/mL and trough levels 1.2 ± mcg/mL, which easily exceed 0.5 mcg/mL concentration that is needed for tyrosine kinase inhibition in vitro [11]. (See Fig. 2.)

Imatinib is metabolized in the liver mainly by the CYP3A4 isoenzyme of the cytochrome P450 system. The major circulating active metabolite is the N-demethylated piperazine derivative, formed predominantly by CYP3A4. This active metabolite demonstrates in vitro potency similar to the parent drug and has a plasma area under the curve that is approximately 15% that of the parent drug. The mean elimination half-lives of imatinib and its active metabolite are 18 and 40 h, respectively. Elimination is predominantly through

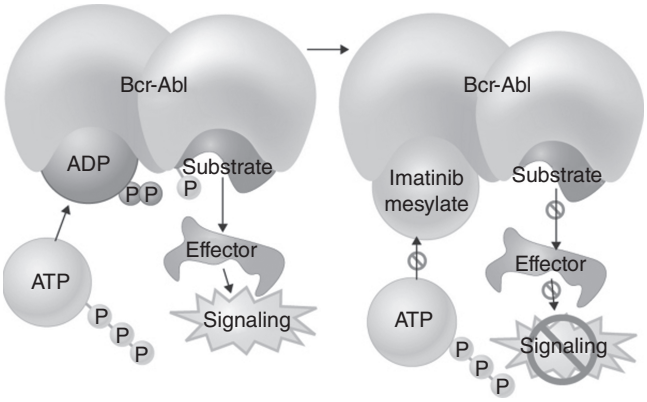


**Figure 2** Mechanism of action of imatinib. By occupying the ATP-binding pocket of the ABL kinase domain, imatinib prevents substrate phosphorylation and downstream activation of signals, thus inhibiting the leukemogenic effects of BCR-ABL on cells in chronic myelogenous leukemia. ADP = adenosine diphosphate; ATP = adenosine triphosphate; P = phosphate group [10].

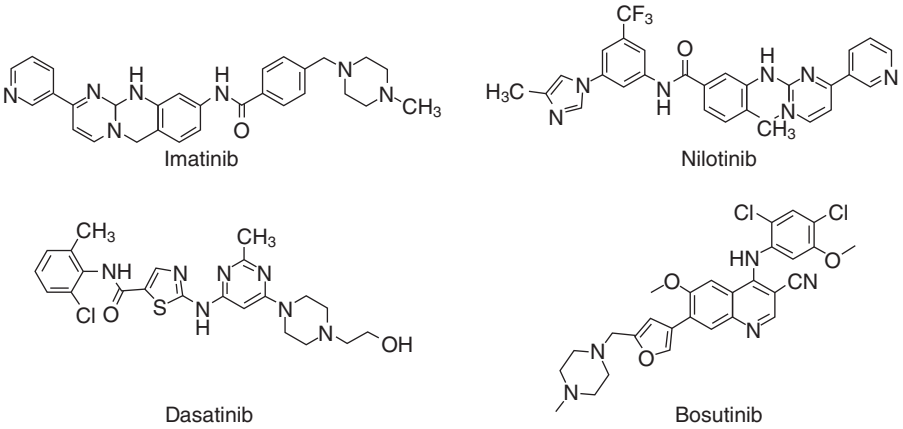


**Figure 1** Chemical structures of imatinib, nilotinib, dasatinib, and bosutinib.

the gastrointestinal tract in feces [11]. Imatinib or its metabolites are not significantly excreted through the kidneys. There is a potential for drug-to-drug interactions, especially with either inducers or inhibitors of the CYP3A4 isoenzyme. Therefore, drugs that induce imatinib metabolism may decrease the clinical activity of imatinib, while drugs that inhibit imatinib metabolism may lead to toxic side effects [11].

### 1.4 Clinical Efficacy

Imatinib is currently the standard first-line therapy for all patients with CML. In a landmark phase 1 study 98% of subjects in the chronic phase of CML that were treated with imatinib demonstrated a complete hematologic response, with 54% showing a cytogenetic response [4]. Response rate in myeloid blast crises of CML was 55%, and the rate of complete remission was 11%. Response rate has been noted to be 70% and complete remission 20% in patients with lymphoid blast crises or Ph + chromosome ALL [1]. Subsequent phase II studies confirmed these findings. Based on remarkable remission rates with imatinib in patients that were previously treated with interferon-α, the U.S. Food and Drug Administration (FDA) approved the drug for patients with CML [12–15]. There has been a dramatic impact of imatinib on survival in patient with CML. In a long-term study with 5 years follow-up of patients, with continuous treatment of chronic-phase CML with imatinib as initial therapy was found to induce durable responses in a high proportion of patients. With a median follow-up of 60 months, Kaplan–Meier estimates of cumulative best rate of complete cytogenetic response was 69% by 12 months and 87% by 60 months. Estimated overall survival was noted to be 89% at 60 months [16]. (See Fig. 3.)

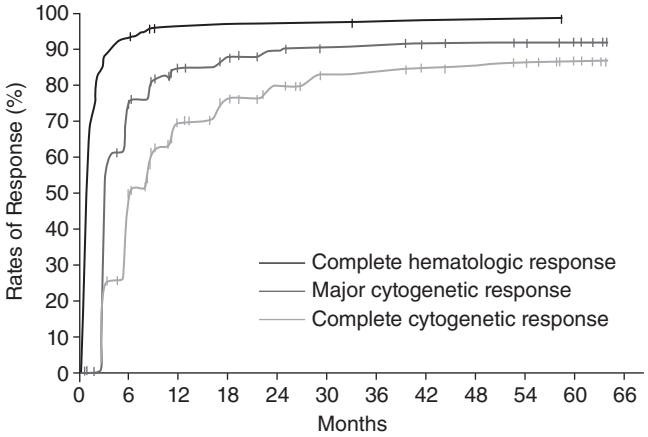


**Figure 3** Kaplan–Meier estimates of the cumulative best response to initial imatinib therapy [16].

When a person is on imatinib, the annual risk of disease progression decreases with time. The overall risk of progression is dependent on degree of response regardless of when it is achieved. At 1 year the annual rate of progression with partial cytogenetic response is 1.4%, with complete cytogenetic response it is 0.6%, and with major molecular response it is 0%. Patients with major molecular response have been noted to have 100% survival at 2 years [17]. In case of annual event rates in patients after achievement of complete cytogenetic response on imatinib as the first-line treatment, 2.1% progressed to accelerated phase or blast crises in the first year after achievement of complete cytogenetic response, whereas in the fourth year none of the patients progressed to either accelerated phase or blast crises phase of CML. Late responses to imatinib occur and responses are durable.

### 1.5 Dosage

Imatinib is available in 400-mg tablets. There is no weight adjustment required, and the starting dose of imatinib mesylate in the chronic phase of CML is 400 mg per day. There is some greater benefit reported with doses greater than or equal to 600 mg/day during different stages of CML. This approach has been tried versus starting patients on newer medications [18]. In a phase II study patients who were previously treated with 400 mg/day of imatinib were given doses of 600 mg/day during accelerated phase of CML. Doses of 600 mg/day led to a higher complete molecular response rate (28% compared to 16%), longer duration of response (79% compared to 57% at 12 months), time to disease progression (67% compared to 44% at 12 months), and overall survival (78% compared to 65% at 12 months) with no clinically relevant increase in toxicity (Table 1). With regard to patients in the chronic phase of CML, a further higher dose of 400 mg twice daily has led to a 38% complete molecular response rate, which repre-

**TABLE 1 Annual Event Rates in Patients after Achievement of Complete Cytogenetic Response on First-Line Imatinib**

| Years after Complete Cytogenetic Response | All Events Inclusive of Deaths and Loss of Response (%) | Progression to Accelerated or Blast Phase (%) |
|---|---|---|
| 1 | 5.5 | 2.1 |
| 2 | 2.3 | 0.8 |
| 3 | 1.1 | 0.3 |
| 4 | 0.4 | 0 |

*Source*: From ref. 16.

sents a sixfold increase over initial studies [19]. Patients that are started on 400 mg twice daily that are diagnosed to be in the chronic phase of CML with an estimated 2-year survival rate is 94%. During a median follow-up of 15 months no patient in the high-dose arm progressed to the accelerated or the blast phase [19]. However, there are some studies that indicate that the benefit obtained from increasing the dose of imatinib is suboptimal. The gains need to be weighed against a higher toxicity profile and patient compliance [20]. Plasma trough levels of imatinib at steady state may be a prognostic factor for clinical responses in patients with CML as they have shown to correlate with both cytogenetic and molecular responses [21, 22].

### 1.6 Side Effects

Side effects that have been noted with imatinib include nausea (50–55%), vomiting (17–41%), edema and fluid retention (41–59%), neutropenia (17–40%), thrombocytopenia (9–33%), elevated liver enzymes (5–14%), diarrhea (15–21%) and fatigue (17–20%) [1, 16]. There is evidence of the increased incidence of toxic effects with doses of imatinib greater than 800 mg per day. Most grades 3–4 toxicities have been noted to occur early in treatment with imatinib (Table 2).

### 1.7 Resistance

Imatinib has shown excellent responses in terms of symptom control and hematologic parameters [2]. However, a significant proportion of patients are resistant to imatinib. Patients are either primarily resistant or develop resistance later on after initially showing a response to imatinib. This has been attributed to development of cytogenetic resistance in leukemic cells [23]. Overall relapse rate for patients with newly diagnosed CML are higher in patients with accelerated and blast-phase disease versus chronic-phase disease [24]. Relapse rate is lower for patients achieving complete cytogenetic response. In patients started on imatinib during the early chronic phase of CML, resistance was noted to emerge only in 16% of the patients at 42 months of follow-up. However, 73 and 95% of the patients develop resistance at 48 months of follow-up, initially in the accelerated and blast phases of CML, respectively. Patients in relapse need new therapies, which has pushed development for new agents to circumvent imatinib resistance [25].

Different mechanisms for resistance have been identified [26]. Mutations in the kinase domain of BCR-ABL are the predominant mechanisms underlying drug resistance. Another mechanism is the amplification of the BCR-ABL fusion gene. A third mechanism of resistance are possibly BCR-ABL-independent mechanisms, possibly related to the activation of the SRC family of kinases or additional molecular events. Point mutations in BCR-ABL kinase domain restrict its ability to adopt an inactive conformation. Mutations in BCR-ABL kinase domain confer varying degrees of resistance to imatinib. In select cases, dose escalation of imatinib may overcome mutation-based resistance. But this has not been seen with mutation T3151 of the BCR-ABL kinase domain [25]. (See Fig. 4.)

Many avenues in overcoming resistance to imatinib have been pursued. One avenue has been to modify imatinib in such a way that the drug binds more tightly, which has been accomplished by Novartis in designing

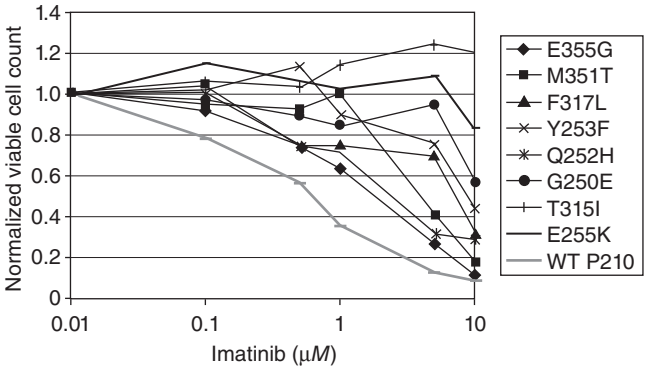


**Figure 4** Varying degrees of resistance to imatinib from mutations in BCR-ABL kinase domain [27].

**TABLE 2 Grade 3–4 Toxicity to First-Line Imatinib**

| | Overall Incidence | Onset after 2 years | Onset after 4 years |
|---|---|---|---|
| Toxicity | Percentage of Patients | | |
| Neutropenia | 16.7 | 3.7 | 1.0 |
| Thrombocytopenia | 8.9 | 1.5 | 0.2 |
| Anemia | 4.4 | 1.8 | 0.5 |
| Elevated liver Enzymes | 5.3 | 0.4 | 0 |
| Other side effects | 16.9 | 4.8 | 1.7 |

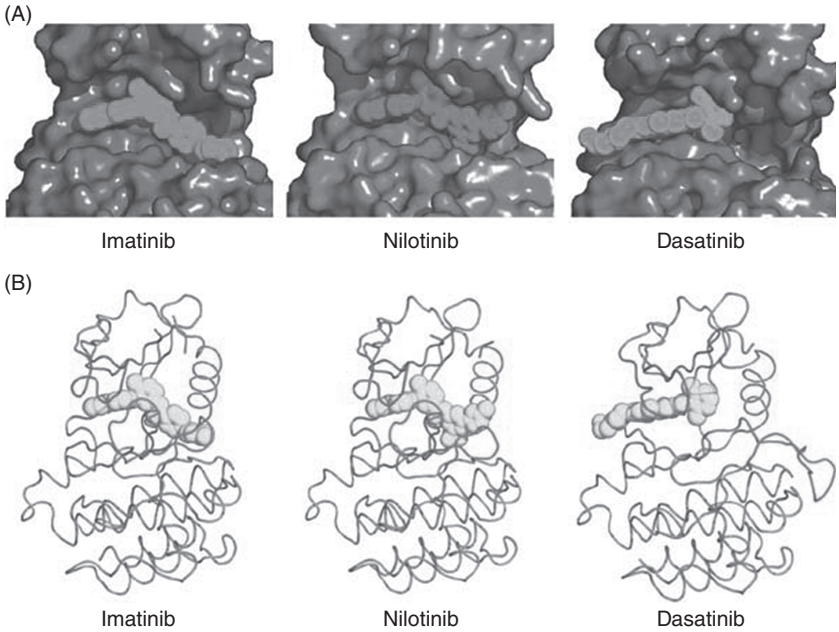


**Figure 5** (A) Surface representations of crystal structures of ABL kinase in complex with imatinib, nilotinib, and dasatinib. (B) Comparison of the different binding modes of three ABL inhibitors: imatinib (left), nilotinib (middle), and dasatinib (right) [28].

nilotinib. Another approach has been the development of entirely different compounds, which was done by Bristol Myers Squibb with dasatinib, which was based on finding that *SRC* inhibitors also inhibit ABL kinase and most imatinib-resistant mutations. Both nilotinib and dasatinib are more potent inhibitors than imatinib, however, they are not active against T3151 mutation. Both dasatanib and nilotinib constitute second-generation drugs. Furthermore, there is a third generation of drugs designed to overcome resistance such as conferred by the T3151 mutation.

## 2 SECOND-GENERATION TYROSINE KINASE INHIBITORS

### 2.1 Nilotinib

Nilotinib (Tasigna) is a second-generation tyrosine kinase inhibitor developed with the intention of treating patients with CML who have developed resistance to imatinib. The chemical structure of nilotinib is 4-methyl-*N*-[3-(4-methylimidazol-1-yl)-5-(trifluoromethyl)phenyl]-3-[(4-pyridin-3-ylpyrimidin-2-yl)amino]benzamide. It is a high-affinity aminopyrimidine-based ATP-competitive inhibitor that decreases proliferation and viability of wild-type BCR-ABL and also cells that express BCR-ABL, which is resistant to imatinib in vitro. (See Fig. 5.) Imatinib-resistant Bcr-Abl mutant-expressing cells in vitro selectively inhibit BCR-ABL autophosphorylation. In addition nilotinib is shown to be a greater than 30-fold more potent inhibitor of wild-type BCR-ABL compared to imatinib in a wide range of CML-derived and transfected cell lines [28, 29].

Not only does nilotinib bind more tightly to BCR-ABL, it has been designed to override resistance caused by mutations. Like imatinib, nilotinib binds to the inactive conformation of the ABL tyrosine kinase, with P-loop folding over the ATP-binding site, and the activation loop blocking the substrate binding site, to disrupt the ATP-binding site and inhibit the catalytic activity of the enzyme. Although nilotinib maintains activity against a majority of imatinib-resistant mutants, it has no significant activity against the T315I mutant BCR-ABL-expressing cell because nilotinib binds closely to the T315 residue inside the BCR-ABL structure similar to imatinib. There is also resistance noted in mutations affecting the P-loop region of BCR-ABL kinase [30].

***Clinical Efficacy*** In a dose-escalating phase I study, a total of 109 patients with imatinib-resistant chronic myelogenous leukemia (CML) in different phases of the disease and Ph+ ALL were treated with nilotinib for up to 385 days [31]. The maximum tolerated dose was determined to be 600 mg twice a day. Side effects noted were myelosuppression, mild–moderate skin rash, and

transient indirect hyperbilirubinemia. In this study, nilotinib was not associated with the edema frequently associated with imatinib. Among patients with chronic-, accelerated-, and blast-phase CML, hematological responses were achieved in 92, 72, and 39%, respectively. Cytogenetic responses were noted in 53, 48, and 27% of patients. The best responses were seen at doses greater than 400 mg once a day and with 400 mg twice a day. In a phase II trial that was designed to characterize the efficacy and safety of nilotinib (400 mg twice daily) with hematologic response as primary efficacy endpoint in patients with accelerated-phase CML who are noted to be resistant and/or intolerant to imatinib, 119 patients were enrolled and had a median duration of treatment of 202 days (range 2–611). A hematologic response was observed in 47% of the patients. Major cytogenetic response was observed in 29% of the patients. Overall survival rate among the 119 patients after 12 months of follow-up was 79%. Nonhematologic adverse events were noted to be mild to moderate. Severe peripheral edema and pleural effusions were not observed. The most common severe hematologic adverse events were thrombocytopenia (35%) and neutropenia (21%) [32]. In one phase II open-label study of patients in chromic-phase CML, 400 mg nilotinib was administered orally twice daily to 280 patients after imatinib failure or intolerance. Patients had at least 6 months of follow-up. At 6 months, the rate of major cytogenetic response was 48%; complete cytogenetic response was noted in 31% and partial response in 16% of the patients. The estimated survival at 12 months was calculated to be 95%. Nilotinib was effective in patients harboring BCR-ABL mutations associated with imatinib resistance (except T315I) and also in patients with a resistance mechanism independent of BCR-ABL mutations. Adverse events were mostly mild to moderate, and there was minimal cross intolerance with imatinib. Grades 3 to 4 neutropenia and thrombocytopenia were observed in 29% of patients; pleural or pericardial effusions were observed in 1% [31].

Nilotinib has substantial clinical utility in rescuing patients who develop imatinib resistance due to point mutations and could potentially be used as a single agent in patients at risk for progression. There are newer studies that suggest both nilotinib and dasatinib can safely be used as first-line medications in patients with chronic-phase CML [33, 34].

Additionally, there is growing interest in testing the hypothesis that administration of multiple ABL kinase inhibitors in early-phase patients, such as nilotinib, dasatinib, and imatinib, could be used to delay or prevent the emergence of drug-resistant clones. However, since the T315I mutation of BCR-ABL is highly resistant to imatinib, nilotinib, and dasatinib, this approach needs to be extended to include inhibitors of T315I BCR-ABL to prevent this mutation from becoming more prevalent. The failure of some patients to respond to nilotinib, especially those with more advanced disease, might arise due to the development of new mutations that impede the interaction between nilotinib and BCR-ABL. Thus, the identification and characterization of BCR-ABL point mutants conferring resistance to nilotinib will assist in the prediction of patient responses to nilotinib, identifying combination partners, as well as in the design and development of novel inhibitors of BCR-ABL that can override resistance to such mutants. The U.S. FDA recently approved nilotinib (Tasigna) for treatment of Philadelphia chromosome-positive chronic myeloid leukemia in adult patients whose disease has progressed or who cannot tolerate other therapies, which included imatinib.

### 2.2 Dasatinib

Dasatinib (Sprycel) is considered a second-generation tyrosine kinase inhibitor used in the treatment of CML. The chemical structure of dasatinib is *N*-(2-chloro-6-methylphenyl)-2-[[6-[4-(2-hydroxyethyl)-1-piperazinyl]-2-methyl-4-pyrimidinyl]amino]-5-thiazole carboxamide monohydrate. In in vitro studies it is active against all tested ABL mutations, except for T315I like both imatinib and nilotinib. Dasatinib has demonstrated high efficacy in phase I and II studies in patients with CML following the failure of imatinib therapy. In addition to binding ABL kinase with less stringent conformational requirements than imatinib, dasatinib has the ability to be a potent SRC kinase inhibitor as well. Therefore, dasatinib is considered a dual BCR-ABL and SRC family tyrosine kinases inhibitor. SRC kinase activation has been linked to BCR-ABL-mediated leukemogenesis and CML progression. It has been approved by the FDA for use in patients with CML after imatinib treatment and Ph+ ALL [35].

The drug is named after the inventor chemist, Jagabandhu Das, who co-discovered it while working at Bristol Myers Squibb [36]. It binds to both inactive and active conformations of BCR-ABL [37]. It is 325-fold more potent at inhibiting BCR-ABL kinase activity than imatinib [38].

It is an oral medication. Its oral bioavailability varies from 14 to 34%, probably as a result of a combination of incomplete absorption and high first-pass metabolism. Pharmacokinetic data suggest that dasatinib is mainly cleared by oxidative metabolism, which involves the CYP3A4 enzyme system in the liver [35, 39]. The terminal half-life is noted to be 3–5 h. Pharmacodynamic data indicates there is inhibition of CrkL which is a downstream substrate of BCR-ABL kinase, which is

more prolonged compared to the relatively short half-life of dasatinib, which indicates that more frequent dosing with dasatinib may not be necessary [39].

***Clinical Efficacy*** In one phase I, open-label, dose escalation study in patients with CML or Ph+ ALL who could not tolerate or were resistant to imatinib, dasatinib was evaluated. Complete hematologic response was achieved in 92%, and major cytogenetic response was seen in 45% of the patients. Most of the complete hematologic responses occurred at doses of 50 mg/day, whereas most of the cytogenetic responses required higher doses. Responses were maintained in 95% of patients with chronic-phase disease and in 82% of patients with accelerated-phase disease. In the case of CML responses were seen to occur at all genotypes except for T315I mutation. Myelosuppression was seen as a major side effect, which was reversible and not dose limiting. Grades 3–4 neutropenia occurred in 45% of the patients [35]. A maximum tolerated dose was not reached, leading to the adoption of a schedule of 70 mg of dasatinib twice a day in some phase II studies.

In a phase II study, 387 patients with chronic-phase CML and imatinib resistance or intolerance were enrolled. At a median follow-up of 15 months, 91% of patients were noted to have achieved complete hematologic responses that were previously resistant to imatinib. Of patients resistant to imatinib, 52 and 40 had achieved a major cytogenetic response and complete cytogenetic remission, respectively, compared with 80 and 75% of patients who were intolerant of imatinib [40]. The occurrence of a major cytogenetic response was similar in patients with and without mutations and in the 72% of patients who had received imatinib at doses of 600 mg or higher before starting dasatinib. In phase II clinical trials at 8-month follow-up, dasatinib induced major hematologic responses in 34 and 31% of myeloid blast crises and lymphoid blast crises CML patients and major cytogenetic responses in 31 and 50% of these patients, respectively. Most (86%) of these major cytogenetic responses were complete cytogenetic responses. Responses were rapid and durable: 88 and 46%, respectively, of myeloid blast crises and lymphoid blast crises CML patients achieving major hematologic response had not experienced disease progression at the 8-month follow-up. Response rates were similar in patients with and without BCR-ABL mutations known to confer resistance to imatinib. Recent phase III data has demonstrated that treatment with 100 mg per day of dasatinib is associated with a most favorable risk–benefit ratio compared to other regimens including 70 mg twice a day and 140 mg once a day [41].

Dasatinib overall has been noted to be well tolerated. Nonhematologic adverse events were mild to moderate. Cytopenias were common and could be managed by dose modification [42]. Fluid retention seems to be the most troublesome nonhematogical toxic effect of dasatinib, which includes pleural effusions. In one study 35% of patients treated with dasatinib developed pleural effusions that were usually exudative, some reaching grade 3 or 4 toxicity, and were mainly in participants who received daily doses of 140 mg of dasatinib or greater [40, 43]. Dasatinib may also be active after failure with treatment with both imatinib and nilotinib [44]. In one study patients who had demonstrated resistance to standard dose of imatinib 400–600 mg/day in chronic-phase CML were treated with either a maximum dose of dasatinib of 90 mg twice daily or a high dose of imatinib of 800 mg/day. Patients were followed for a median of 15 months. Patients taking dasatinib versus imatinib had a higher incidence of complete hematologic response of 93 versus 82% and a higher major cytogenetic response of 52 versus 33%, and complete cytogenetic response of 40 versus 16% [15]. This indicates that it is preferable to switch to dasatinib right away with patients with imatinib resistance or failure rather than increasing the dosage of imatinib to evaluate for response. Although both nilotinib and dasatinib show encouraging preliminary findings, the concern exists that their use will result in the identification of new mutations in the ABL-kinase domain.

Several groups have now reported clinical resistance to dasatinib in association with persisting or emerging mutations. Recent studies have identified F317L/I and T315I as the most frequent mutations. In addition BCR-ABL with V299L and V304D are also shown to be resistant to dasatinib [35]. As mentioned earlier, in many patients mechanisms other than kinase domain mutations are likely responsible for disease resistance.

### 2.3 Bosutinib

Bosutinib, previously known as SKI-606, is a 4-anilino-3-quinoline carbonitrile dual inhibitor of SRC and ABL, which is orally active and has antiproliferative activity against CML [45]. It is currently being developed by Wyeth. This drug has no activity against the tyrosine kinase receptors for platelet-derived growth factor [the inhibition of which is thought to account for the fluid retention noted with other tyrosine kinase inhibitors (TKIs)] and cKit. Its inhibitory activity against BCR-ABL is 200 times more potent than imatinib. It inhibits a wide variety of BCR-ABL kinase mutants, except T315I [45, 46].

Bosutinib is known to reduce vascular endothelial growth factor-mediated vascular permeability and tumor-cell extravasation, and its effect on cell-to-cell interactions between tumor cells and host cells might be

crucial for its in vivo efficacy. Phase I trials have shown the dose to be 500 mg once a day for testing in phase II trials. Recent preliminary data from a phase II study for assessable patients with chronic-phase CML, who were resistant to imatinib, showed that 74% achieved complete hematological response, 42% achieved major cytogenetic response with 33% of patients receiving complete cytogenetic response on bosutinib with a median duration of treatment of 5.1 months. Among patients who had previously received nilotinib and dasatinib, 38% achieved complete hematologic response, whereas 25% were able to achieve major cytogenetic response. Major hematologic toxicities noted were thrombocytopenia in 9% of the patients, neutropenia in 8%, and anemia in 1% of the patients. Most common side effects noted are gastrointestinal but are mild and transient [47]. In patients with accelerated-phase and blast-phase CML, complete hematologic response has been noticed in 29 and 25% of the patients, respectively, with median duration of treatment with bosutinib for 2.7 months [48]. There is a potential for synergistic effect with imatinib based on in vitro studies [46].

## 3 THIRD-GENERATION TYROSINE KINASE INHIBITORS

### 3.1 MK0457

The aurora family of serine–threonine kinases is essential for mitotic progression. All members of this family have identifiable ATP-binding sites and have been linked to tumor genesis via induction of genetic instability. MK0457, previously known as VX680, is an example of an aurora kinase inhibitor and exerts antiproliferative and proapoptotic activity in several tumor cell lines. For example, cells treated with MK0457 enter mitosis and proceed through the S-phase but do not divide. Furthermore, MK0457 binds to the active conformation of BCR-ABL and is noted to have activity against the T315I mutation. It is also an inhibitor of Janus kinase 2. This inhibitor has entered clinical studies of imatinib-resistant CML and Philadelphia-chromosome-positive acute lymphoblastic leukemia. Initial reports show that patients with resistant disease associated with the T315I mutation had hematological responses, and in some of these patients, the concentration of the mutant clone decreased. These reported cases are the first clinically observed activity of a kinase inhibitor against the T315I phenotype of CLL [49].

### 3.2 Adaphostin

Adaphostin (also known as NSC680410) is a tyrophostin inhibitor that was designed to compete for the ATP substrate rather than ATP with BCR-ABL. Adaphostin has the ability to select CML cells over healthy cells in clonogenic assays and induces apoptosis in other leukemic cells, which suggests that its effect might not be entirely BCR-ABL specific. Another way adaphostin may be effective may be related to an increase in reactive oxygen species, which amplifies the effect of adaphostin. Adaphostin also downregulates the BCR-ABL protein. This effect is not prevented by antioxidants, thereby implying that the downregulation of BCR-ABL precedes or parallels the production of reactive oxygen species. It has been found effective against wild-type BCR-ABL and all known mutations including T315I.

### 3.3 XL228

XL228 is a protein kinase inhibitor. It has potent activity against different isoforms of BCR-ABL including T315I. It also has activity against IGF1R and SRC. It has shown to modulate CrkL activity, indicating downstream inhibition of BCR-ABL signaling in T315I isoforms. Preliminary phase I results show that XL228 is well tolerated. It is given as an intravenous infusion. It has a mean terminal half-life ranging from 15 to 40 h. XL228 is a promising agent for the treatment of CML and Ph+ ALL that is resistant to earlier generation tyrosine kinase inhibitors [50].

### 3.4 AP24534

AP24534 is an orally active BCR-ABL inhibitor. It is shown to inhibit unmutated and mutated isoforms, particularly T315I of BCR-ABL. It is inactive against aurora kinases. It also inhibits the Flt3, SRC, VEGFR, and FGFR family of kinases, which may contribute to its potency toward treating CML. The in vitro kinase profile suggests that the inhibition of BCR-ABL is the main mechanism for activity against BCR-ABL models. Based on promising preliminary studies, phase I studies are currently planned to assess the safety and tolerability of this agent [51].

### 3.5 INNO406

INNO406, also known as CNS9 or NS187, is 25–55 times more potent than imatinib against the ABL kinase and also inhibits LYN, a protein that has been implicated in the progression to blast phase [23]. However, this drug is also not active against the T315I mutation. INNO406 is orally bioavailable in a murine model and seems to be a competitive inhibitor of ATP [52]. Initial phase I trials have shown that it is well tolerated. It has been tested in patients who have been treated with imatinib and the second-generation agents nilotinib and dasat-

inib. The recommended phase II dose is likely to be 240 mg twice daily. Up to now, both hematological and cytogenetic responses have been shown. This drug is likely to be a competitor for first-line or second-line treatment of CML, rather than third-line treatment.

## 4 SUMMARY

The discovery of imatinib has revolutionized the treatment of CML. It is a classic example of drug development based on the identification of a common molecular target involved in a signaling pathway. It was the first medication that inhibited a particular molecular target in the cell to treat a cancer type instead of as in earlier chemotherapy drugs that nonspecifically target rapid dividing cells. It is considered such a breakthrough that it has been featured on the cover of *Time* magazine after approval by the FDA [53]. Imatinib has demonstrated significantly lower toxicity while having significant effectiveness compared to earlier therapies; however, it is not considered curative. Furthermore, issues with imatinib resistance have emerged [54]. This has led to the development of more targeted therapies to overcome resistance. Dasatinib and nilotinib have received FDA approval, both of which are effective in treating patients who have demonstrated resistance or intolerance to imatinib. Further drugs are in development that show promise to be even effective in patients resistant to dasatinib and nilotininb as well. This will provide further options and even improved outcomes in patients being treated for CML.

## REFERENCES


1. Druker BJ, et al. (2001). Activity of a specific inhibitor of the BCR-ABL tyrosine kinase in the blast crisis of chronic myeloid leukemia and acute lymphoblastic leukemia with the Philadelphia chromosome. *N Engl J Med 344*, 1038–1042.
2. Talpaz M, et al. (2002). Imatinib induces durable hematologic and cytogenetic responses in patients with accelerated phase chronic myeloid leukemia: Results of a phase 2 study. *Blood 99*, 1928–1937.
3. Manley PW, et al. (2002). Imatinib: A selective tyrosine kinase inhibitor. *Eur J Cancer 38*(Suppl 5), S19–S27.
4. Druker BJ, et al. (2007). Efficacy and safety of a specific inhibitor of the BCR-ABL tyrosine kinase in chronic myeloid leukemia. *N Engl J Med 344*, 1031–1037.
5. Schiffer CA (2007). BCR-ABL tyrosine kinase inhibitors for chronic myelogenous leukemia. *N Engl J Med 357*, 258–265.
6. Talpaz M, et al. (1995). Percentage of Philadelphia chromosome (Ph)-negative and Ph-positive cells found after autologous transplantation for chronic myelogenous leukemia depends on percentage of diploid cells induced by conventional-dose chemotherapy before collection of autologous cells. *Blood 85*, 3257–3263.
7. Talpaz M, Qiu X, Cheng K, Cortes JE, Kantarjian H, Kurzrock R (2000). Autoantibodies to Abl and Bcr proteins. *Leukemia 14*, 1661–1666.
8. Druker BJ, Lydon NB (2000). Lessons learned from the development of an abl tyrosine kinase inhibitor for chronic myelogenous leukemia. *J Clin Invest 105*, 3–7.
9. Talpaz M (2002). Treatment of CML: Thoughts after ASCO 2002. *MedGenMed 4*, 11.
10. Goldman L (2007). *Cecil Medicine*, 23 ed. New York: Elsevier, pp. 1397–1402.
11. Peng B, Lloyd P, Schran H (2005). Clinical pharmacokinetics of imatinib. *Clin Pharmacokinet 44*, 879–894.
12. Braziel RM, et al. (2002). Hematopathologic and cytogenetic findings in imatinib mesylate-treated chronic myelogenous leukemia patients: 14 months' experience. *Blood 100*, 435–441.
13. Ottmann OG, et al. (2002). A phase 2 study of imatinib in patients with relapsed or refractory Philadelphia chromosome-positive acute lymphoid leukemias. *Blood 100*, 1965–1971.
14. Sawyers CL, et al. (2002). Imatinib induces hematologic and cytogenetic responses in patients with chronic myelogenous leukemia in myeloid blast crisis: Results of a phase II study. *Blood 99*, 3530–3539.
15. Shah N (2006). Dasatinib (D) vs high dose imatinib (IM) in patients (pts) with chronic phase chronic myeloid leukemia (CP-CML) resistant to imatinib. Results of CA180017 START-R randomized trial. 2006 ASCO Annual Meeting Proceedings Part I. *J Clin Oncol 24*(18S, (June 20 Suppl), 6507.
16. Druker BJ, et al. (2006). Five-year follow-up of patients receiving imatinib for chronic myeloid leukemia. *N Engl J Med 355*, 2408–2417.
17. Hughes TP, et al. (2003). Frequency of major molecular responses to imatinib or interferon alfa plus cytarabine in newly diagnosed chronic myeloid leukemia. *N Engl J Med 349*, 1423–1432.
18. Kantarjian H, et al. (2007). Dasatinib or high-dose imatinib for chronic-phase chronic myeloid leukemia after failure of first-line imatinib: A randomized phase 2 trial. *Blood 109*, 5143–5150.
19. Kantarjian H, et al. (2004). High-dose imatinib mesylate therapy in newly diagnosed Philadelphia chromosome-positive chronic phase chronic myeloid leukemia. *Blood 103*, 2873–2878.
20. Schiffer C (2007). What if conventional-dose imatinib fails? *Blood 109*, 5068–5069.
21. Larson RA, et al. (2008). Imatinib pharmacokinetics and its correlation with response and safety in chronic-phase chronic myeloid leukemia: A subanalysis of the IRIS study. *Blood 111*, 4022–4028.

22. Picard S, et al. (2007). Trough imatinib plasma levels are associated with both cytogenetic and molecular responses to standard-dose imatinib in chronic myeloid leukemia. *Blood 109*, 3496–3499.
23. Apperley JF (2007). Part II: Management of resistance to imatinib in chronic myeloid leukaemia. *Lancet Oncol 8*, 1116–1128.
24. Druker BJ (2006). Circumventing resistance to kinase-inhibitor therapy. *N Engl J Med 354*, 2594–2596.
25. Shah NP (2005). Loss of response to imatinib: Mechanisms and management. *Hematol Am Soc Hematol Ed Program* 183–187.
26. Kujawski L, Talpaz M (2007). Strategies for overcoming imatinib resistance in chronic myeloid leukemia. *Leuk Lymphoma 48*, 2310–2322.
27. Shah NP, et al. (2002). Multiple BCR-ABL kinase domain mutations confer polyclonal resistance to the tyrosine kinase inhibitor imatinib (STI571) in chronic phase and blast crisis chronic myeloid leukemia. *Cancer Cell 2*, 117–125.
28. Weisberg E, et al. (2005). Characterization of AMN107, a selective inhibitor of native and mutant Bcr-Abl. *Cancer Cell 7*, 129–141.
29. Golemovic M, et al. (2005). AMN107, a novel aminopyrimidine inhibitor of Bcr-Abl, has in vitro activity against imatinib-resistant chronic myeloid leukemia. *Clin Cancer Res 11*, 4941–4947.
30. VonBubnoff N, Manley PW, Mestan J, Sanger J, Peschel C, Duyster J (2006). Bcr-Abl resistance screening predicts a limited spectrum of point mutations to be associated with clinical resistance to the Abl kinase inhibitor nilotinib (AMN107). *Blood 108*, 1328–1333.
31. Kantarjian HM, et al. (2007). Nilotinib (formerly AMN107), a highly selective BCR-ABL tyrosine kinase inhibitor, is effective in patients with Philadelphia chromosome-positive chronic myelogenous leukemia in chronic phase following imatinib resistance and intolerance. *Blood 110*, 3540–3546.
32. leCoutre P, et al. (2007). Nilotinib (formerly AMN107), a highly selective BCR-ABL tyrosine kinase inhibitor, is active in patients with imatinib-resistant or -intolerant accelerated phase chronic myelogenous leukemia. *Blood 110*, 3540–3546.
33. Cortes J (2007). Efficacy of nilotinib (AMN107) in patients (Pts) with newly diagnosed, previously untreated Philadelphia chromosome (Ph)-positive chronic myelogenous leukemia in early chronic phase (CML-CP). *Blood 110*(11), 29.
34. Cortes J (2007). Efficacy of dasatinib in patients (pts) with previously untreated chronic myelogenous leukemia (CML) in early chronic phase (CML-CP). *Blood 110*(11), 30.
35. Talpaz M, et al. (2006). Dasatinib in imatinib-resistant Philadelphia chromosome-positive leukemias. *N Engl J Med 354*, 2531–2541.
36. Das J, et al. (2006). 2-Aminothiazole as a novel kinase inhibitor template. Structure-activity relationship studies toward the discovery of *N*-(2-chloro-6-methylphenyl)-2-[[6-[4-(2-hydroxyethyl)-1-piperazinyl)]-2-methyl-4-pyrimidinyl] amino)]-1,3-thiazole-5-carboxamide (dasatinib, BMS-354825) as a potent pan-Src kinase inhibitor. *J Med Chem 49*, 6819–6832.
37. Shah NP, Tran C, Lee FY, Chen P, Norris D, Sawyers CL (2004). Overriding imatinib resistance with a novel ABL kinase inhibitor. *Science 305*, 399–401.
38. O'Hare T, et al. (2005). Combined Abl inhibitor therapy for minimizing drug resistance in chronic myeloid leukemia: Src/Abl inhibitors are compatible with imatinib. *Clin Cancer Res 11*, 6987–6993.
39. Luo FR, et al. (2006). Dasatinib (BMS-354825) pharmacokinetics and pharmacodynamic biomarkers in animal models predict optimal clinical exposure. *Clin Cancer Res 12*, 7180–7186.
40. Hochhaus A (2007). Dasatinib for the treatment of Philadelphia chromosome-positive chronic myelogenous leukaemia after imatinib failure. *Expert Opin Pharmacother 8*, 3257–3264.
41. Hochhaus A, et al. (2008). Favorable long-term follow-up results over 6 years for response, survival, and safety with imatinib mesylate therapy in chronic-phase chronic myeloid leukemia after failure of interferon-{alpha} treatment. *Blood 111*, 1039–1043.
42. Cortes J, et al. (2007). Dasatinib induces complete hematologic and cytogenetic responses in patients with imatinib-resistant or -intolerant chronic myeloid leukemia in blast crisis. *Blood 109*, 3207–3213.
43. Quintas-Cardama A, Kantarjian H, Cortes J (2006). Targeting ABL and SRC kinases in chronic myeloid leukemia: Experience with dasatinib. *Future Oncol 2*, 655–665.
44. Quintas-Cardama A, et al. (2007). Dasatinib (BMS-354825) is active in Philadelphia chromosome-positive chronic myelogenous leukemia after imatinib and nilotinib (AMN107) therapy failure. *Blood 109*, 497–499.
45. Golas JM, et al. (2003). SKI-606, a 4-anilino-3-quinolinecarbonitrile dual inhibitor of Src and Abl kinases, is a potent antiproliferative agent against chronic myelogenous leukemia cells in culture and causes regression of K562 xenografts in nude mice. *Cancer Res 63*, 375–381.
46. Redaelli S (2007). Synergistic activity of the novel Src/abl inhibitor bosutinib in combination with Imatinib. *Blood 110*, 4064–4072.
47. Cortes J (2007). Efficacy and safety of bosutinib (SKI-606) among patients with chronic phase Ph+ CML. *Blood 110*, 3540–3546.
48. Gambacorti-Passerini C (2007). Bosutinib (SKI606) demonstrated clinical activity and is well tolerated among patients with AP and BP CML and Ph+ ALL. *Blood 110*, 2600–2609.
49. Giles FJ, Cortes J, Jones D, Bergstrom D, Kantarjian H, Freedman SJ (2007). MK-0457, a novel kinase inhibitor, is active in patients with chronic myeloid leukemia or acute lymphocytic leukemia with the T315I BCR-ABL mutation. *Blood 109*, 500–502.

50. Shah NP (2008). Interim results from a phase 1 clinical trial of the bcr-abl inhibitor XL228 in drug resistant Ph+ leukemias. Paper presented at the 13th Congress of the EHA, June 15, 2008.

51. Rivera V (2007). Potent anti-tumor activity of AP24534, an orally active Inhibitor of BCR-ABL variants including T315I, in *in vitro* and *in vivo* models of CML. *Blood 110*(11), 1032–1089.

52. Naito H, et al. (2006). In vivo antiproliferative effect of NS-187, a dual Bcr-Abl/Lyn tyrosine kinase inhibitor, on leukemic cells harbouring Abl kinase domain mutations. *Leuk Res 30*, 1443–1446.

53. There is new ammunition against war against cancer. These are the bullets. *Time*, May 28, 2008.

54. Savona M, Talpaz M (2008). Getting to the stem of chronic myeloid leukaemia. *Nat Rev Cancer 8*, 341–350.

# AROMATASE INHIBITORS FOR BREAST CANCER

Akm Hossain[1] and Jame Abraham[2]

[1]*Section of Hematology/Oncology, West Virginia University School of Medicine, Morgantown, West Virginia*
[2]*Breast Cancer Program, Section of Hematology/Oncology, West Virginia University School of Medicine, Morgantown, West Virginia*

## 1 INTRODUCTION

The continued evolution of endocrine therapy has contributed to many therapeutic options for patients with breast cancer. In 1998, Oxford overview demonstrated that the selective estrogen receptor modulator (SERM) tamoxifen (TM) was equally effective in pre- and postmenopausal breast cancer patients. Development of aromatase inhibitors (AIs) added a new option in addition to SERMs for the treatment of breast cancer.

The landmark study of anastrozole (Arimidex), tamoxifen alone, or TM in a combination (ATAC) trial by Michael Baum presented at the San Antonio Breast Cancer Symposium in 2001 marked a turning point in breast cancer therapy. In the ATAC trial anastrozole had a better disease-free survival rate compared to TM when used in postmenopausal patients. Since the publication of the ATAC trial [1], several other trials have looked at the role of aromatase inhibitors such as letrozole [Breast Internation Group (BIG 98)] and exemestane [Intergroup Exemestane Study (IES)] in adjuvant therapy for postmenopausal breast cancer patients. All these studies have confirmed that AIs are superior to tamoxifen and have fewer side effects.

## 2 BIOCHEMISTRY

The two main circulating estrogens in the body are estrone and estradiol, derived from cholesterol. By inhibiting the enzymes, which act on earlier steps in the branching pathway of steroidogenesis, it is possible to have consequences of undesirable impact on the production of other physiologically necessary and important hormones such as aldosterone and cortisol. The conversion of androgens to estrogens occurs at peripheral sites such as in muscle, fat, liver, and also in breast tumor cells. Aromatase is a cytochrome P450 enzyme that has both a steroid-binding and iron-containing site.

Androstenedione is the substrate for aromatase and it is linked to the enzyme's steroid-binding site. Based on this structural relationship, there are two reasonable and effective ways to inhibit aromatase, either by occupying the steroid-binding site of the enzyme with a compound or by binding the iron with nitrogen-containing compounds such as aminoglutethimide. Since aromatase catalyzes the last step in estrogen production, it makes an ideal target for the development of selective and potent inhibitors.

The comparative efficacy of AIs may be assessed in two ways. The conversion measurement of androgens into estrogens provides direct evidence of enzymatic activity. On the other hand, measurement of suppression of plasma estrogens provides indirect evidence of enzymatic aromatase activity. Drug metabolism (half-life) and tissue distribution are the two main factors for in vivo measurements of the biochemical efficacy of a compound. Intergeneration and intrageneration significant differences have been reported in the time to reach steady-state plasma levels, half-life, elimination, and in the interaction with several cytochromes (CXP) of the P450 superfamily. Aromatase inhibition and estradiol

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

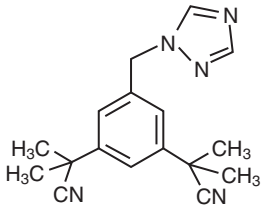


**Figure 1** Chemical structure of anastrozole.

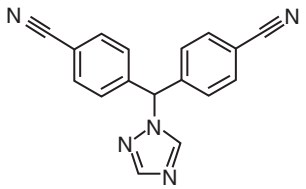


**Figure 2** Chemical structure of letrozole.

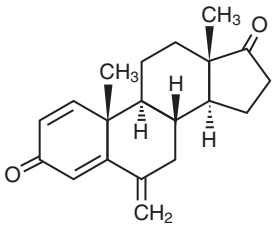


**Figure 3** Chemical structure of exemestane.

**TABLE 1 Classification of Aromatase Inhibitors**

| Generation | Type I (Steroidal Inactivator) | Type II (Nonsteroidal Inhibitor) |
|---|---|---|
| First | None | Aminogluthethimide (Cytadren) |
| Second | Formestane (Lentaron) | Fadrozole<br>Rogletimide |
| Third | Exemestane (Aromasin) | Anastrozole (Arimidex)<br>Letrozole (Femara)<br>Vorozole |

suppression show good correlation and are consistently higher for third-generation AIs when compared with first- and second-generation AIs. (See Figs. 1–3.)

## 3 CLASSIFICATION

Based on the mechanism of action and chronologic order of development along with relative potency, AIs can be divided into two types and three generations, respectively (Table 1). AIs represent several generations of development, and in each step the evolution presented a compound with increase potency and specificity. The third-generation AIs are drugs of immense potential and undoubtedly play a major role in the management of postmenopausal women with hormone-dependent breast cancer.

## 4 CLINICAL PHARMACOLOGY

The type I or steroidal inactivators are derivatives of androstenedione, which is the aromatase physiological substrate. Type II or nonsteroidal inhibitor compounds are derived by phenobarbitone (e.g., aminoglutethimide) or may be characterized by the imidazole/triazole structure. Type I AIs bind covalently to the enzyme, causing its irreversible inactivation. But the nonsteroidal compounds provide reversible binding. First- and second-generation AIs are characterized by lower selectivity and potency in comparison to newer AIs.

Aromatase, a cytochrome P450 enzyme, is involved in catalyzing three consecutive hydroxylation reactions converting C19 androgens to aromatic C18 estrogens. After receiving electrons from nicotinamide adenine dinucleotide phosphate (NADPH)–cytochrome P450 reductase, aromatase converts androstenedione and testosterone, respectively, to estrone and estradiol. Aromatization of androgen is the terminal and rate-limiting step in estrogen production. Abnormal overexpression of aromatase in breast tissue plays a significant part in breast cancer pathogenesis. Under the above circumstances, inhibition of aromatase is a promising and new strategy for reducing growth-stimulatory effects of estrogen in breast cancer pathology. Multiple studies suggest potential endocrine manipulation in the treatment of breast cancer and support adjuvant and sequential systemic treatment including AIs.

Anastrozole provides 70% estradiol reduction after 24 h and 80% after 2 weeks of therapy. Duration of estradiol reduction is about 6 days. It is well absorbed and not affected by food. Plasma protein binding is about 40%. It is metabolized extensively by liver (~85%) via *N*-dealkylation, hydroxylation, and glucuronidation, but the primary metabolite is inactive. Anastrozole is excreted in the urine (10% as unchanged drug; 60% as metabolites). Letrozole is also well absorbed and the absorption is not affected by food. Its plasma protein binding is weak. It is metabolized through liver by CYP3A4 and 2A6 to an inactive carbinol metabolite. The drug is excreted in the urine (90%; 6% as unchanged drug, 75% as glucuronide carbinol metabolite, 9% as unidentified metabolites).

Exemestane is rapidly but moderately (~42%) absorbed following oral administration. Absorption may increase by another ~40% following a high-fat meal. Plasma protein binding is extensive (90%), pri-

marily to albumin and $\alpha_1$-acid glycoprotein. It is extensively metabolized by liver [oxidation (CYP3A4) of methylene group], reduction of 17-keto group with the formation of many secondary metabolites, but the metabolites are inactive. The drug is excreted in the urine (39–45% as metabolites, <1% as unchanged drug) and feces (36–48%).

## 5 DESCRIPTION OF AI IN BRIEF

The first-generation inhibitor refers to the nonsteroidal inhibitor aminoglutethamide (AG). The reports of adrenal insufficiency led to withdrawal from clinical use. AG is less specific and inhibits other CYP450 enzymes involved in cortisol and aldosterone biosynthesis, which results in toxicity. Its efficacy in inhibiting aromatase activity stimulated the development of various new inhibitors during the 1980s and 1990s. The second-generation inhibitors are the imidazole derivative fadrozole and the steroid analog formestane. Fadrozole is more selective and potent than AG, but it still has inhibitory effects on progesterone, aldosterone, and corticosterone production. Formestane was the first selective AI to be used clinically and was effective and well tolerated, but its requirement of intramuscular administration limited its clinical use.

The third-generation inhibitors were developed in the early 1990s. Anastrozole and letrozole are nonsteroidal derivatives characterized by a triazole functional group that interacts with the heme prosthetic group of aromatase and act as competitive inhibitors with respect to the androgen substrates. Exemestane is a steroidal and mechanism-based inhibitor that is catalytically converted into a chemically reactive species, leading to irreversible inactivation of aromatase. These are widely used as the first-line drugs in the endocrine treatment of hormone-dependent postmenopausal breast cancer in patients. Anastrozole, letrozole, and exemestane are administered orally at 1, 2.5, and 25 mg once daily, respectively. Compared to the first- and second-generation inhibitors, the third-generation inhibitors produce greater clinical benefit with near-complete specificity at clinical use.

## 6 AROMATASE INHIBITORS IN POSTMENOPAUSAL PATIENTS

### 6.1 Anastrozole in Metastatic Breast Cancer

The Tamoxifen or Anastrozole Randomized Group Efficacy and Tolerability (TARGET) study was a randomized, double-blind trial that compared TM to anastrozole as front-line treatment in advanced breast cancer. All the available data from this study and a North American study conducted at the same time were combined and analyzed. At a median follow-up of 43.7 months, mortality was 56% in the anastrozole group and 56.1% in the TM group. Both treatment arms were well tolerated, however, there were less gynecological complications (1% vs. 2.5%) and thromboembolic phenomena (5.3% vs. 9%) in the anastrozole arm. It is to be concluded that anastrozole provided similar efficacy as tamoxifen but appeared to be better tolerated.

### 6.2 Anastrozole in Adjuvant Therapy

The ATAC trial compared 5 years of anastrozole versus 5 years of TM versus the combination of anastrozole and TM in postmenopausal women as an adjuvant therapy. After 68 months of follow-up, anastrozole showed significantly improved disease-free survival (hazard ratios (HR) 0.87, 95% CI 0.78–0.97, $P = 0.01$). A significant reduction in contralateral breast cancer (42% reduction, 95% CI 12–62, $P = 0.01$) and distant metastases (HR 0.86, CI 0.74–0.99, $P = 0.04$) were observed.

### 6.3 Anastrozole in Switching Adjuvant Therapy

The Italian Tamoxifen Anastrozole (ITA) trial was a prospective randomized study evaluating the benefit of switching from TM to anastrozole in adjuvant scenarios. After 2–3 years of TM therapy, women were randomly assigned to receive an additional 2–3 years of TM or anastrozole. At median follow-up of 36 months, event-free and recurrence-free survival was significantly better in the anastrozole group (HR 0.35, 95% CI 0.2–0.63, $P = 0.0002$; HR 0.35, 95% CI 0.18–0.68, $P = 0.001$) and local disease control was also improved in anastrozole group (HR 0.15, 95% CI 0.03–0.65, $P = 0.003$).

The Arimidex–Nolvadex (ARNO) and the Austrian Breast and Colorectal Cancer Study Group (ABCSG) trials both provided the data on switching to anastrozole after 2 years of TM. After the combined analysis, at median follow-up of 28 months, event-free survival was better in the anastrozole group in comparison to patients who stayed on TM (HR 0.60, 95% CI 0.44–0.81, $P = 0.0009$).

The above-mentioned three trials clearly showed superiority of anastrozole over tamoxifen in postmenopausal women with breast cancer. Based on these findings, the Food and Drug Administration (FDA) approved the use of anastrozole 1 mg p.o. daily in metastatic breast cancer in January 1996 and as an adjuvant therapy for early breast cancer in September 2002.

### 6.4 Letrozole in Metastatic Breast Cancer

The International Letrozole Breast Cancer Group conducted a trial that compared letrozole to TM as first-line therapy in postmenopausal women with advanced breast cancer. Time to progression and time to treatment failure were significantly better in the letrozole group (9.4 vs. 6 months, $P < 0.0001$; 9.0 vs. 5.7 months, $P < 0.0001$, respectively). Objective response rate were also significantly higher in the letrozole group (32% vs. 21%, $P = 0.0002$). A trend toward improved overall survival for women in the letrozole group was observed, but this was not statistically significant (34 vs. 30 months, $P = 0.53$).

### 6.5 Letrozole in Adjuvant Therapy

The BIG 1-98 Collaborative Group was a phase III, randomized, double-blind study comparing letrozole and TM each as monotherapy versus sequential administration. Disease-free survival was significantly improved in the letrozole arm compared to TM (HR 0.81, 95% CI 0.7–0.93, $P = 0.003$) along with distant recurrence (HR 0.73, 95% CI 0.6–0.88, $P = 0.001$). Five-year estimates of disease-free survival were 81.4% for the TM arm and 84% for the letrozole arm. Subgroup analyses showed better response in patients on letrozole with node-positive disease (HR 0.71, 95% CI 0.59–0.85, $P < 0.001$) and patients who received letrozole and chemotherapy (HR 0.70, 95% CI 0.54–0.92, $P = 0.01$).

### 6.6 Letrozole in Extended Adjuvant Therapy

The National Cancer Institute of Canada MA-17 study is a phase III randomized, double-blind, placebo-controlled trial that randomized postmenopausal women to receive an additional 5 years of letrozole versus placebo who completed about 5 years of TM. The first interim analysis showed the estimated 4-year disease-free survival to favor letrozole (93% vs. 87%, HR 0.57, 95% CI 0.43–0.75, $P = 0.00008$). Subgroup analyses demonstrated that letrozole was as effective in node-negative disease (HR 0.47, $P = 0.005$) as it was in node-positive disease (HR 0.60, $P = 0.003$). Based on the above data, FDA approved letrozole 2.5 mg for extended adjuvant therapy in breast cancer treatment.

### 6.7 Exemestane in Switching Adjuvant Therapy

The IES is an international, double-blind, phase III randomized switching trial that involves TM for 5 years versus exemestane after at least 2 years, and no more than 3 years and 1 month of TM. After a median follow-up of 37.4 months, an interim analysis showed improved disease-free survival in women who switched to exemestane (HR 0.73, 95% CI 0.62–0.86, $P = 0.0001$). Exemestane 25 mg daily is approved by the FDA for use in the adjuvant setting after 2–3 years of tamoxifen.

## 7 CONCLUSION ON AIS IN BREAST CANCER THERAPY

The American Society of Clinical Oncology recommends that an AI should be included in the adjuvant treatment of postmenopausal women with hormone receptor-positive breast cancer. Treatment options include 5 years of an AI versus sequential therapy with TM (combination of 2–3 years or 5 years) followed by an AI for 2–3 or 5 years. Patients who cannot tolerate TM should be offered an AI. After progression on TM or AI, consideration could be given to fulvestrant or megestrol acetate. (See Table 2.)

## 8 LACK OF CROSS RESISTANCE

The first published data on lack of cross resistance between different AIs was reported in advanced breast cancer patients who showed clinical response to formestane after progression on AG. Lonning et al. showed that after 8 weeks of exemestane, the overall response rate was 8.1% in patients previously treated with AG and 4.8% in those treated with other nonsteroidal AIs. Exemestane following AG was associated with enhanced aromatase inhibition. There was no difference in any type of serum estrogen with the switch from other nonsteriodal AIs to exemestane. Several other studies have addressed this issue, providing the evidence that significant benefit can be achieved with exemestane after second- and third-generation nonsteroidal AIs.

## 9 SIDE EFFECTS

It is obvious that AIs have an adverse reaction profile like other endocrine therapies. It can mimic menopausal symptoms due to depletion of estrogen and are the same, such as hot flashes and sweating. Adverse reactions related to AIs include stomach upsets (nausea, vomiting, diarrhea), rash, and arthralgia. First-generation AI was poorly tolerated and could cause drowsiness, fever, and inhibition of cortisol synthesis. Formestane is well-tolerated as a treatment but can cause a local reaction around the injection site. These drugs were also found to be better tolerated than TM and were associated with lower incidences of endometrial cancer, vaginal bleeding and discharge, cerebrovascular events,

**TABLE 2 FDA-Approved Indications, Doses, and Duration of Commonly Used AIs**

| AIs | FDA-Approved Indications | Doses/Duration |
|---|---|---|
| Anastrezole | 1. Postmenopausal with hormone receptor-positive early breast cancer<br>2. Postmenopausal with hormone receptor-positive early breast cancer after 2–3 years of TM therapy<br>3. Locally advanced or metastatic breast cancer (hormone receptor-positive or status unknown)<br>4. Postmenopausal women with metastatic disease progression after TM therapy | 1 mg orally daily; maximum duration up to 5 years |
| Letrozole | 1. Postmenopausal with hormone receptor-positive early breast cancer<br>2. Postmenopausal with hormone receptor-positive early breast cancer after 5 years of TM therapy<br>3. Locally advanced or metastatic breast cancer (hormone receptor-positive or status unknown)<br>4. Postmenopausal women with metastatic disease progression after antiestrogen therapy | 2.5 mg orally daily; maximum duration up to 5 years |
| Exemestane | 1. Postmenopausal women with disease progression after TM therapy (orphan drug)<br>2. Postmenopausal with hormone receptor-positive early breast cancer after 2–3 years of TM therapy<br>3. Postmenopausal with hormone receptor-positive early breast cancer who are switched to exemestane for completion of total of 5 consecutive years of adjuvant hormone treatment | 25 mg orally after meal daily; maximum duration up to 5 years |

venous thromboembolic events, and hot flashes. (Se Table 3.)

There were significantly less endometrial malignancies (0.2% vs. 0.8%, $P = 0.02$), ischemic cerebrovascular events (2% vs. 2.8%, $P = 0.03$), vaginal bleeding/discharge (8.9% vs. 23.4%, $P < 0.0001$), venous thromboembolic events (2.8% vs. 4.5%, $P = 0.0004$), and hot flashes with anastrozole compared to tamoxifen. There was a nonsignificant trend toward more ischemic cardiovascular disease in the anastrozole group (4.1% vs. 3.4%, $P = 0.1$). Women who switched to anastrozole had significantly more gastrointestinal side effects and lipid metabolism disorders than those on tamoxifen (7.9% vs. 2.7%, $P = 0.3$ and 9.3% vs. 4%, $P = 0.04$, respectively). Patients who remained on tamoxifen had more significant gynecological abnormalities, including endometrial carcinoma (11.3% vs. 2%, $P = 0.0002$). Letrazole showed a significantly increased risk of thromboembolic events (3.5% vs. 1.5%, $P < 0.001$), hot flashes, and vaginal discharge in comparison to tamoxifen. Women on letrozole also developed fewer invasive endometrial cancers (0.1% vs. 0.3%, $P = 0.18$).

**TABLE 3 Common Side Effects with Commonly Used AIs**

| Side Effects | Anastrozole (%) | Letrozole | Exemestane |
|---|---|---|---|
| Hot flashes | 12–36 | 5–50 | 13–21 |
| Nausea/vomiting | 11–19 | 9–17 | 6–15 |
| Fatigue/weakness | 7–18 | 4–34 | 8–22 |
| Bone pain | 6–20 | 8–22 | 7–21 |
| Arthalgia | 2–15 | 8–22 | 15–29 |
| Headache | 7–19 | 4–20 | 5–18 |
| Hypertension | 2–13 | 3–11 | 5–15 |
| Mood disturbance | 11–19 | 2–12 | 4–10 |

## 10 SKELETAL EVENTS

In clinical practice, the long-term side effects on the skeletal system should be monitored closely. Multiple trials have included bone side effect profile in the studies (Table 4). In ATAC trial after a median follow-up of 68 months, it has been reported that the clinical fracture rate in the AIs group is significantly higher than the TM group (11% vs. 7.7%, $p < 0.0001$). There was no difference in the rate of hip fractures (1.2% vs. 1.0%, $p = 0.5$) and wrist fractures (2.3% vs. 2.0%, $p = 0.4$). But spine fractures (1.5% vs. 0.9%, $p = 0.03$ and other site fractures (7.1% vs. 4.6%, $p < 0.0001$) rate were higher in the AIs group in comparison to the TM group. After a median follow-up of 60 months, bone mineral density (BMD) changes were adversely affected in the AIs group both in the spine (−6.1 vs. +2.8) and in hip (−7.2 vs. +0.7) (Table 5). Patients on AIs had both

**TABLE 4 Clinical Fracture Rates in Different Trials Using AIs**

| Trial | Follow-up in Months | Primary Agent | Comparator Agent | AIs (%) | Comparator (%) | *P* Value |
|---|---|---|---|---|---|---|
| ATAC | 68 | Anastrozole | Tamoxifen (TM) | 340 (11) | 237 (7.7) | <0.0001 |
| BIG 1-98 | 25.8 | Letrozole | Tamoxifen | 225 (5.7) | 159 (4) | <0.001 |
| ITA | 36 | Anastrozole | Tamoxifen | 2 (1) | 2 (1.3) | 0.60 |
| ABCSG8/ARNO 95 | 28 | Anastrozole | Tamoxifen | 34 (2) | 16 (1) | 0.015 |
| IES | 30.5 | Exemestane | Tamoxifen | 72 (3.1) | 53 (2.3) | 0.08 |
| MA 17 | 31 | Letrozole | Placebo | 137 (5.3) | 119 (4.6) | 0.25 |

**TABLE 5 Rate of Bone Mineral Density Changes in Different Trials Using AIs**

| Trial | Follow-up in Months | Primary Agent | Comparator Agent | AIs | Comparator | *P* Value |
|---|---|---|---|---|---|---|
| ATAC | 60 | Anastrozole | Tamoxifen | S –6.1 | S 2.8 | — |
| | | | | H –7.2 | H 0.7 | |
| TEAM | 12 | Exemestane | Tamoxifen | S –2.75 | S 1.0 | 0.027 |
| | | | | H –2.2 | H 3.3 | 0.012 |
| IES | 12 | Exemestane | Tamoxifen | S –2.97 | S –0.02 | — |
| | | | | H –1.57 | H –0.05 | |
| Lonning | 24 | Exemestane | Placebo | S –4.34 | S –3.68 | — |
| | | | | H –5.44 | S –2.96 | |
| MA.17 | 19 | Letrozole | Placebo | S –5.35 | S –0.71 | 0.008 |
| | | | | H –3.6 | S –0.7 | 0.044 |
| ABCSG | 36 | Anastrozole | Tamoxifen | S –17.4 | S –11.6 | <0.0001 |
| IBIS—II | 12 | Anastrozole | Placebo | S –2.5 | S –0.97 | 0.002 |
| | | | | H –1.34 | H –0.37 | 0.02 |

S = spine fractures
H = hip fractures

**TABLE 6 Osteoporosis Rates in Different Trials Using AIs**

| Trial | Follow-up in Months | Primary Agent | Comparator Agent | AIs (%) | Comparator Agent (%) | *P* Value |
|---|---|---|---|---|---|---|
| IES | 31 | Exemestane | Tamoxifen | 7.4 | 5.7 | 0.05 |
| MA 17 | 30 | Letrozole | Placebo | 8.1 | 6.0 | 0.003 |
| Goss et al (2011) [2] | 29 | Letrazole | Placebo | 5.8 | 4.5 | 0.07 |

increased bone resorption (15.5%, 95% CI = 3–25%) and bone formation (20%, 95% CI = 14–25%), whereas patient on TM had both decreased bone formation and resorption.

In the BIG I-98 trial, after a median follow-up of 25.8 months, clinical fracture rate was higher in the AIs group than the TM group (5.7% vs. 4.0%, $p < 0.001$). In the TEAM trial, the AI group had negative changes in BMD both in the spine (–2.75 vs. 1.0, $p = 0.0257$) and in the hip (-2.2 vs. 3.3, $p = 0.012$) in comparison to TM. In all the above-mentioned trials TM was found to prevent bone health. The ABCSG8/ARNO 95 (two different trials with one combined analysis for possible statistical power) found higher clinical fracture rates in the AIs group (2% vs. 1%, $p = 0.015$) and negative BMD change in the spine (–17.4 vs. 11.6, $p < 0.0001$).

Studies have shown that most of the recurrence in hormone receptor-positive breast cancer will occur after 5 years of TM therapy. The MA 17 clinical trial addressed this issue. Letrozole was compared with placebo after 5 years of TM therapy. There was a nonsignificant but higher rate of clinical fracture in the AI group (5.3% vs. 4.3%, $p = 0.25$). But changes in BMD were significantly higher in the letrozole group both in the spine (–5.35% vs. –0.71%, $p = 0.008$) and the hip (–3.6% vs. –0.70%, $p = 0.044$). (See Table 6.)

## 11 TREATMENT OF SKELETAL-RELATED EVENTS WITH AIS

Life-style modification (e.g., smoking cessation, balanced diet, physical activity, regular physical exercise, maintaining ideal body weight), supplementation with calcium (1200 mg/day), and vitamin D (800 IU) are recommended for patients at risk for osteoporosis because these levels prevent decreases in femoral neck BMD and reduce the risk for hip fractures in elderly women. The American Society of Clinical Oncology (ASCO) expert panel recommended breast cancer patients who have evidence of bone destruction on plain radiographs be treated with intravenous pamidronate 90 mg delivered over 2 h or zoledronic acid 4 mg over 15 min every 3–4 weeks. There is insufficient evidence supporting the efficacy of one bisphosphonate over the other, although zoledronic acid is used more widely. If there is an abnormal bone scan and an abnormal computed tomography (CT) or magnetic resonance imaging (MRI) scan showing bone destruction, but normal plain radiographs, starting bisphosphonates is considered appropriate based on the findings in women with lytic or mixed lytic/blastic changes on plain radiographs. It is also not recommended by the ASCO panel to start bisphosphonates in women with only an abnormal bone scan but without evidence of bone destruction on radiographs, CT scans, or MRI. Integrating bisphosphonates in the treatment without evidence of bone metastases even in the presence of other extra-skeletal metastases is not recommended. There should be no change in dosage, infusion time, or interval in patients with a serum creatinine less than 3.0 mg/dL (265 μmol/L). Infusion times less than 15 min with zoledronic acid or less than 2 h with pamidronate should be avoided to preserve renal function. Creatinine should be monitored before each dose of either agent as per FDA labeling instructions. It is not recommended to use any biochemical markers to monitor bisphosphonate response, and therapy should be continued until there are subjective and objective evidences of substantial decline in patient's general performance status.

## 12 CONTRAINDICATION OF AI

Hypersensitivity to AIs or any component of the formulation and pregnancy are the contraindications for AIs use.

## 13 DRUG INTERACTIONS

*CYP3A4 inducers*: May decrease the levels/effects of AIs. Examples of inducers include but are not limited to aminoglutethimide, carbamazepine, efavirenz, fosphenytoin, nafcillin, nevirapine, oxcarbazepine, pentobarbital, phenobarbital, phenytoin, primidone, rifabutin, rifampin, and rifapentine. Dosage adjustment required with potent inducers. Rifampin may decrease exemestane levels.

## 14 NUTRITION AND HERB INTERACTIONS

Plasma levels may be increased by 40% if AIs are taken with a fatty meal. St. John's wort may decrease AI levels. Patient should avoid black cohosh, dong quai in estrogen-dependent tumors.

## 15 PREGNANCY IMPLICATIONS

AIs have been associated with prolonged gestation, abnormal or difficult labor, and increased resorption, reduced number of live fetuses, decreased fetal weight, and retarded ossification in rats. AIs are not indicated for premenopausal women, but if exposure occurred during pregnancy, risk to the fetus and potential risk for loss of the pregnancy should be discussed. Excretion of AIs in breast milk is unknown.

## 16 FUTURE DIRECTIONS

In clinical practice, third-generation AIs are firmly established as effective agents in the therapy for postmenopausal women with advanced and early-stage breast cancer. Some of the new data might also indicate a greater role in preventive settings. The available data, especially information from bone subprotocols have shown an adverse effect of AIs in bone health. But at this time we need more studies to define more clearly the absolute and relative risk of osteoporosis, fractures, and endpoint relation of BMD in this population. Even given all these limitations, oncologist should use AIs where indicated with appropriate and expanded knowledge of the effect of AIs on bones and their timely management. At the same time, bone health should be evaluated and monitored accordingly. Appropriate treatment with diet modification, calcium, vitamin D, and bisphosphonate should be initiated when indicated.

## REFERENCES

1. Howell A, Cuzick J, Daum M, Buzdar A, Dowsett M, Forbes JF, Hoctin-Does G, Houghton J, Locker GY, Tobias JS.

(2005). Results of the ATAC (Arimidex, Tamoxifen, alone or in combination) trial after completion of 5 years' adjuvant treatment for breast cancer. *Lancet 365*(9453), 60–62.

2. Goss PE, Ingle JN, Ales-Martinez JE, et al. (2011). Exemestane for breast-cancer prevention in postmenopausal women. *N Engl J Med 364*, 2381–2391, PMID: 21639806.

## SUGGESTED READING

Barnadas A, Gil M, Sánchez-Rovira P, Llombart A, Adrover E, Estevez LG, de la Haba J, Calvo L. (2008). Neoadjuvant endocrine therapy for breast cancer: Past, present and future. *Anticancer Drugs 19*(4), 339–347.

Bjarnason NH, Hitz M, Jorgensen NR, Vestergaard P. (2008). Adverse bone effects during pharmacological breast cancer therapy. *Acta Oncol 47*(4), 747–754.

Brufsky AM. (2008). Bone health issues in women with early-stage breast cancer receiving aromatase inhibitors. *Curr Oncol Rep 10*(1), 18–26.

Cuzick J. (2008). Aromatase inhibitors in early breast-cancer treatment: The story so far. *Breast 17*(Suppl 3), S2–8.

Gibson LJ, Dawson CK, Lawrence DH, Bliss JM. (2007). Aromatase inhibitors for treatment of advanced breast cancer in postmenopausal women. *Cochrane Database Syst Rev* Jan *24*(1), CD003370.

Herold CI, Blackwell KL. (2008). Aromatase inhibitors for breast cancer: Proven efficacy across the spectrum of disease. *Clin Breast Cancer 8*(1), 50–64.

Howell A. (2006). The "Arimidex", Tamoxifen, Alone or in Combination (ATAC) Trial: A step forward in the treatment of early breast cancer. *Rev Recent Clin Trials 1*(3), 207–215.

Lonning PE, Geisler J. (2008). Indications and limitations of third-generation aromatase inhibitors. *Expert Opin Investig Drugs 17*(5), 723–739.

Nabholtz JM, Gligorov J. (2006). The emerging role of aromatase inhibitors in the adjuvant management of breast cancer. *Rev Recent Clin Trials 1*(3), 237–249.

# EVOLUTION OF SELECTIVE ESTROGEN AND ANDROGEN RECEPTOR MODULATORS: STATUS OF CURRENT THERAPY AND NEW DRUG DEVELOPMENT

Gregory T. Wurz and Michael W. DeGregorio

*Department of Internal Medicine, Division of Hematology and Oncology, School of Medicine, University of California, Davis, Sacramento, California*

## 1 BACKGROUND

Selective estrogen and androgen receptor modulators (SERMs and SARMs) are compounds that can act, respectively, as either estrogen receptor (ER) or androgen receptor (AR) agonists or antagonists on a tissue-selective basis. For example, the SERM tamoxifen, which is widely used in the treatment of breast cancer, is an ER antagonist in breast tissue but has agonistic effects in the endometrium. SERMs and SARMs are diverse groups of molecules comprised of structurally unrelated compounds from several different chemical classes. The chemical structures of selected SERMs and SARMs are shown in Figures 1 and 2. Included in Tables 1 and 2 are some of the various chemical classes of SERMs and SARMs as well as some clinically approved and investigational compounds from each class.

### 1.1 How SERMs and SARMs Became a Focus of Drug Development

The search for new contraceptive agents in the 1950s and 1960s led to the development of the first clinically useful SERM, tamoxifen [1]. Rather than acting as a contraceptive, tamoxifen, then known as a nonsteroidal antiestrogen, was found to actually induce ovulation [2], an effect that led to the approval of tamoxifen and a related drug, clomiphene, for use as fertility agents [3–5]. Tamoxifen was later developed as a treatment for breast cancer, and it is currently approved for the treatment of all stages of breast cancer and for reducing the risk of breast cancer in high-risk women. While tamoxifen is a very effective breast cancer treatment due in part to its antiestrogenic effects in breast tissue, it was soon recognized that tamoxifen was having estrogenic effects in the endometrium, which is associated with an increased risk of endometrial cancer [6–9], on serum lipids [10, 11], and on bone [12, 13]. These observations led to efforts to develop new compounds, that is, SERMs, with the beneficial effects of estrogen on bone density and serum cholesterol [14–16] and without the negative side effects of tamoxifen and estradiol on the endometrium [6–9] and breast [17], respectively. Early SERM development was focused on breast cancer applications in an attempt to improve upon tamoxifen, but the success of raloxifene as an osteoporosis treatment [18], and the potential risks posed by long-term use of hormone replacement therapy (HRT) [17] in postmenopausal women, have generated an interest in developing new SERM compounds for the treatment of postmenopausal osteoporosis, menopausal symptoms,

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

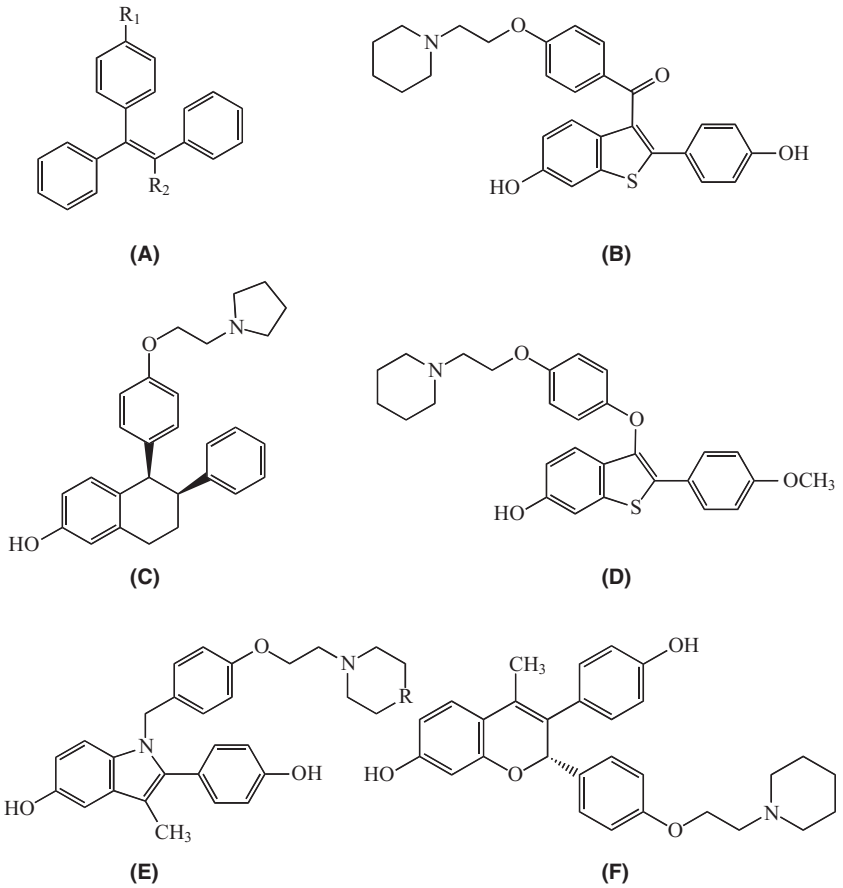


**Figure 1** Chemical Structures of Approved and Investigational SERMs. (A) Generic triphenylethylene structure; for **Tamoxifen**, $R_1 = OCH_2CH_2N(CH_3)_2$, $R_2 = CH_2CH_3$; for **Toremifene**, $R_1 = OCH_2CH_2N(CH_3)_2$, $R_2 = CH_2CH_2Cl$; for **Clomiphene**, $R_1 = OCH_2CH_2N(CH_2CH_3)_2$, $R_2 = Cl$; for **Ospemifene**, $R_1 = OCH_2CH_2OH$, $R_2 = CH_2CH_2Cl$; for **GW5638**, $R_1 = HC = CHCOOH$, $R_2 = CH_2CH_3$. (B) **Raloxifene**. (C) **Lasofoxifene**. (D) **Arzoxifene**. (E) Generic indole structure; for **Bazedoxifene**, $R = CH_2CH_2$ (seven-membered ring); for **ERA-923**, $R = CH_2$ (six-membered ring). (F) **Acolbifene**.

sexual dysfunction, and for cardiovascular and cognitive benefits.

There are currently five major chemical classes of SERMs (see Table 1); however, only four SERMs have been approved for clinical use: tamoxifen, toremifene, clomiphene, and raloxifene. The "ideal" SERM would be a compound with beneficial estrogen-like effects on bone, the cardiovascular system, and cognition, while having antiestrogenic effects in the breast and endometrium. This compound would be devoid of side effects that impact a woman's quality of life, such as hot flashes and vaginal dryness, and life-threatening side effects, such as thromboembolic complications and breast or endometrial cancer [19]. None of the currently approved SERMs meet this definition, but several new agents are currently in development and will be discussed later in this chapter.

The need for an ideal SARM is also recognized since androgen replacement therapy employing steroidal androgens, for example, testosterone, in men is not without its drawbacks [20]. Testosterone, with the exception of testosterone undecanoate, has poor oral bioavailability due to extensive first-pass metabolism, and large variations in serum testosterone concentrations can be observed with intramuscularly injected testosterone preparations. Transdermal delivery of testosterone via scrotal and nonscrotal skin patches is also used, but these have been associated with localized skin irritation in up to 10% of patients using the nonscrotal patch [20]. The high doses of androgen used for hormonal male contraception can lead to weight gain, reductions in high-density lipoprotein (HDL) cholesterol, and acne [21]. Long-term effects on the prostate are another potential concern [22]. Thus, there is an interest in devel-

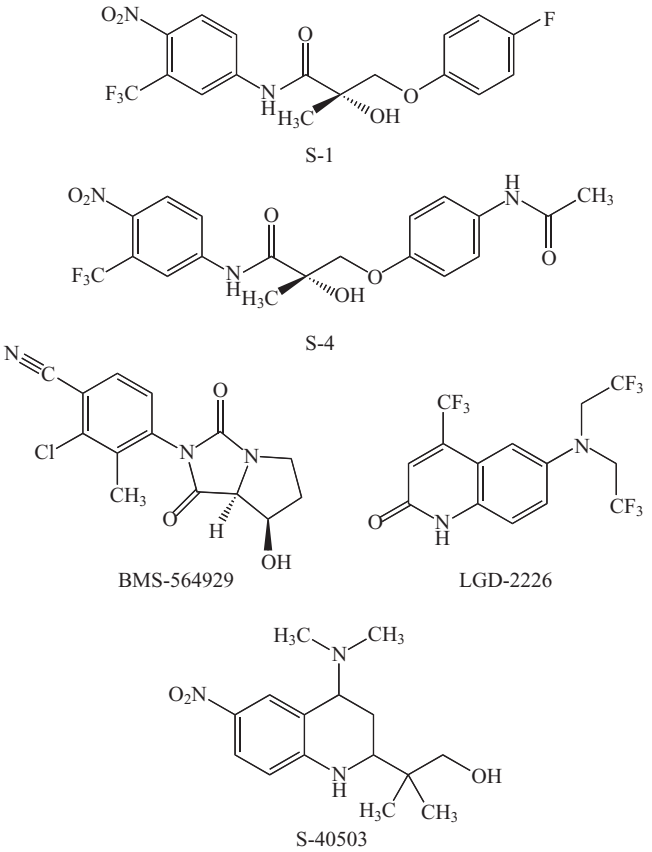


**Figure 2** Chemical Structures of Investigational SARMs.

oping orally bioavailable, nonsteroidal androgens that could provide the benefits of steroidal androgen therapy while avoiding steroid-related side effects. Compared to SERMs, the discovery of nonsteroidal androgens, now termed SARMs, was relatively recent, with the first such molecules being described in 1998 [23]. There are currently four major chemical classes of SARMs described in the medical literature (see Table 2). Although none of these new agents have been approved by the Food and Drug Administration (FDA), clinical trials are well underway.

Another factor driving the development of SARMs is our aging population. It is known that as men age, their serum testosterone levels gradually decline, potentially resulting in a condition known as androgen deficiency in the aging male (ADAM), also known as andropause [24]. Associated with the decline in serum testosterone is an increased estrogen-to-androgen concentration ratio in the prostate, which may factor in the development of benign prostatic hyperplasia (BPH) and potentially prostate cancer [25, 26]. Androgen deficiency can eventually lead to loss of libido, sexual dysfunction, and osteoporosis, among other symptoms, in older men [27]. The use of steroidal androgens in this population raises fears of increased prostate cancer risk, and therefore the potential risks of current treatment options must be weighed carefully. Thus, although the benefits of androgen replacement have long been known, more selective and efficacious treatment options, especially those with prostate-sparing effects, for example, SARMs, are clearly needed. Three excellent reviews [27–29] discuss the discovery and potential therapeutic applications of SARMs in more detail.

**TABLE 1 Chemical Classes and Representative Agents of Approved and Investigational SERMs**

| Triphenylethylenes | Benzothiophenes | Naphthalenes | Benzopyrans | Indoles |
|---|---|---|---|---|
| Tamoxifen[a] | Raloxifene[a] | Lasofoxifene | Acolbifene | Bazedoxifene |
| Toremifene[a] | Arzoxifene | Nafoxidine[b] | Levormeloxifene[b] | ERA-923 |
| Clomiphene[a] | | | CHF-4056 | |
| Ospemifene | | | | |
| Droloxifene[b] | | | | |
| Idoxifene[b] | | | | |
| GW5638 | | | | |

[a]FDA approved.
[b]No longer in clinical development.

**TABLE 2 Chemical Classes and Representative Agents of Investigational SARMs**

| Aryl Propionamides | Bicyclic Hydantoins | Quinolines | Tetrahydroquinolines |
|---|---|---|---|
| S-1 | BMS-564929 | LGD-2226 | S-40503 |
| S-4 | | | |
| C-6 | | | |

### 1.2 SERM and SARM Mechanisms of Action

Although the precise mechanisms of the tissue-selective actions of SERMs and SARMs remain incompletely understood, the principal effects of these agents are mediated via the ER and AR, respectively. Both the ER and AR are members of the nuclear receptor superfamily of ligand-activated transcription factors, which includes the progesterone, glucocorticoid, vitamin D, thyroid, and mineralocorticoid receptors. The ER and AR have a similar domain structure, with an amino-terminal domain containing the ligand-independent transactivation function 1 (AF-1), a centrally located deoxyribonucleic acid (DNA)-binding domain, and a carboxyl-terminal ligand-binding domain (LBD) containing a second, ligand-dependent transactivation function (AF-2). There are two known isoforms of the ER, designated α and β, while only one isoform of the AR has been observed. Upon binding their cognate ligands [estradiol for ER and testosterone or dihydrotestosterone (DHT) for AR], the receptors undergo a conformational change that dissociates them from heat shock proteins that stabilize them in an inactive conformation within the cytoplasm. The receptors then dimerize, translocate to the nucleus, and bind their respective DNA consensus sequences in the estrogen (ERE) and androgen response elements (ARE) within the promoter regions of estrogen- and androgen-regulated genes, respectively.

Whether a particular SERM or SARM acts as an agonist or antagonist is thought to depend on the specific conformational change of the receptor following binding of the ligands within the LBD, which is composed of 12 α-helices in both the ER and AR. The resulting position of helix 12 appears to be particularly important for transcriptional activation. When the ER and AR are bound to agonists such as estradiol or DHT, respectively, helix 12 folds to enclose the ligand within the LBD, allowing the receptor to interact with the coactivators and other elements of the general transcription machinery necessary for target gene expression [30–32]. If the receptors are bound to antagonists, helix 12 adopts a different, inhibitory conformation that promotes interaction with co-repressor proteins [33]. What is unique about the AR compared to the ER is that AF-1 appears to be the predominant transactivation function [34, 35]. In the ER, the ligand-dependent AF-2 is the predominant transactivation function [36]. The main function of AF-2 in the AR appears to be the promotion of an amino-terminal (N) to carboxyl-terminal (C) interaction (N/C interaction), which is dependent upon ligand binding [37, 38]. Whether this interaction is intramolecular or intermolecular is still a matter of debate; however, the most recent data suggests that the N/C interaction occurs intramolecularly [39]. There are conflicting data regarding the importance of the N/C interaction, with some studies showing that it is required for normal androgen target gene expression [40, 41], and other studies showing that AR target gene transcription can occur through AF-2 in the absence of an N/C interaction [42, 43]. Although the AR N/C interaction is known to decrease the ligand dissociation rate [44], a more recent study has shown that the N/C interaction may additionally play an important role in chromatin binding in vivo where remodeling of chromatin is necessary for target gene expression [41]. The N/C interaction is apparently not unique to the AR, as a similar type of interaction has been described for the ER [45].

## 2 CLINICAL TARGETS OF SERMS AND SARMS

There are numerous potential clinical targets for SERMs and SARMs in the genitourinary, musculoskeletal, and cardiovascular systems. Outlined in Table 3 are the primary clinical targets associated with these multifunctional agents. At present, there are four SERMs that have been FDA approved for clinical use: tamoxifen (for breast cancer treatment and prevention), toremifene (breast cancer treatment), raloxifene (treatment and prevention of postmenopausal osteoporosis), and clomiphene (induction of ovulation in anovulatory women). Although no SARMs have been approved for clinical use, the most mature of these agents have reached the phase II stage of clinical evaluation. Briefly discussed below are the current status of endocrine therapies, the current status of HRT, and the clinical use of tamoxifen, toremifene, and raloxifene.

**TABLE 3 Major Clinical Targets of Approved and Investigational SERMs and SARMs[a]**

| Clinical Target | SERM | SARM |
|---|---|---|
| Osteoporosis treatment | × | × |
| Osteoporosis prevention | × | × |
| Breast cancer treatment | × | |
| Breast cancer prevention | × | |
| Prostate cancer treatment | × | × |
| Prostate cancer prevention | × | × |
| Menopausal symptoms | × | |
| Urogenital atrophy | × | |
| Muscle wasting | | × |
| Male contraception | | × |
| BPH | | × |
| Hypogonadism/ADAM | | × |

[a]Currently, there are no drugs classified as SARMs that have been FDA approved for clinical use.

### 2.1 Current Status of Endocrine Therapy

Hormonal, or endocrine-based, therapies exist for a variety of maladies including prostate cancer (flutamide, nilutamide, and bicalutamide), breast cancer (tamoxifen, toremifene, and aromatase inhibitors), osteoporosis (raloxifene), menopausal symptoms (HRT), and androgen insufficiency or male hypogonadism (testosterone). Flutamide, nilutamide, and bicalutamide are nonsteroidal antiandrogens (not SARMs) used in androgen ablation therapy of prostate cancer [46]. As will be discussed below, the first SARMs were based on the structure of the nonsteroidal antiandrogens. For over 25 years, the triphenylethylene tamoxifen has been the endocrine treatment of choice for both early- and late-stage ER-positive breast cancer in postmenopausal women. In the last several years, however, third-generation aromatase inhibitors such as letrozole and anastrozole have proven to be more efficacious compared to tamoxifen in both advanced and early-stage breast cancer, which has led to aromatase inhibitors supplanting tamoxifen as the treatment of choice [47–50]. Aromatase inhibitors are currently being evaluated as potential breast cancer chemopreventive agents in postmenopausal women [50]. The triphenylethylene toremifene, a SERM related to tamoxifen, is approved for the treatment of metastatic breast cancer in postmenopausal women, and it is also currently being evaluated in a phase III clinical trial for prostate cancer chemoprevention [51]. The benzothiophene SERM raloxifene is currently approved for the treatment and prevention of postmenopausal osteoporosis. Raloxifene was also evaluated as a breast cancer chemopreventive agent in the recently completed Study of Tamoxifen and Raloxifene (STAR) [52]. Hormone replacement therapy (estrogen alone or estrogen plus progestin) is currently used for the alleviation of menopausal symptoms in postmenopausal women. Once thought to impart cardioprotective benefits in addition to its beneficial effects on bone mineral density (BMD) [16], the use of HRT is now largely restricted to short-term use for the alleviation of hot flashes due to the potential for life-threatening side effects associated with long-term use of HRT [17, 53]. Androgen insufficiency, also known as hypogonadism, ADAM, or andropause, is commonly treated with androgen supplementation (testosterone). As discussed earlier, this therapy is far from ideal [20–22], and it is hoped that a more efficacious therapy will result from current SARM development efforts.

### 2.2 Current Status of Hormone Replacement Therapy

The benefits of HRT in postmenopausal women include the prevention of osteoporosis [16] and the alleviation of menopausal symptoms [54, 55]. Prior to the Women's Health Initiative (WHI) study of HRT in healthy, postmenopausal women, it was also believed that HRT use was associated with cardiovascular benefits [56–58] and cognitive benefits such as the prevention of Alzheimer's disease [59–62], in addition to the positive effects of HRT on osteoporosis and menopausal symptoms. However, results from the WHI estrogen plus progestin study, a prospective, randomized, placebo-controlled clinical trial in 16,608 postmenopausal women without hysterectomy, have shown that HRT, given as 0.625 mg/day conjugated equine estrogens plus 2.5 mg/day of medroxyprogesterone acetate, provided no benefit in the primary prevention of coronary heart disease (CHD) and may actually increase the risk of CHD [17, 63]. Also, the incidence of invasive breast cancer, which was the primary adverse outcome measured in the trial, was found to be elevated in the HRT group [17]. Interestingly, in the WHI estrogen alone trial in hysterectomized postmenopausal women, a significant difference in breast cancer incidence was not observed between the estrogen and placebo groups, although an increased risk of stroke was seen [64, 65].

Furthermore, contrary to the earlier reported cognitive benefits of HRT, results from the WHI Memory Study (WHIMS) showed no improvement in cognitive function with HRT compared to placebo, and a small but significantly increased risk of cognitive decline in the HRT arm [66]. The risk of dementia was also significantly increased in postmenopausal women in the HRT group [67]. Similar declines in cognitive function and an increased risk of dementia were observed among those postmenopausal women who received estrogen alone in the WHIMS [68, 69]. In regard to the benefits of HRT in bone, the positive effects of HRT were confirmed in the WHI estrogen plus progestin study, with significant increases in BMD and a significantly reduced incidence of fractures in the HRT group [70]. Similar, positive bone effects were observed in the estrogen alone trial [64]. Ultimately, however, the increase in risks compared to benefits led to the early termination of both the WHI estrogen plus progestin trial and the estrogen alone trial [17, 64].

Although there has been some controversy surrounding the results from the WHI estrogen plus progestin trial because of the average age of the participants and the fact that it was stopped 3 years short of its planned 8.5 years [55, 58], the use of HRT declined following the announcement of the trial results [71]. Because of adverse effects, the long-term use of both HRT and unopposed estrogen is no longer recommended for the treatment or prevention of cardiovascular disease or the prevention of osteoporosis and dementia [72]. Consequently, there has been renewed

interest in SERMs that may potentially have many of the benefits of estrogen but with a lower incidence of its associated side effects.

### 2.3 Tamoxifen for Reducing the Risk of Breast Cancer

The benefits of using tamoxifen in the adjuvant setting for the treatment of postmenopausal ER-positive breast cancer, where it has been shown to significantly increase 10-year survival [73], led to the hypothesis that it may be useful for the primary prevention of breast cancer as well. Following the results of the National Surgical Adjuvant Breast and Bowel Project's (NSABP) Breast Cancer Prevention Trial (BCPT) [74], tamoxifen was approved for reducing the risk of breast cancer in women at high risk. Tamoxifen thus became the first agent, and first SERM, to demonstrate efficacy in the primary prevention of breast cancer in women at high risk. In that study, which included 13,388 women deemed to be at high risk of developing breast cancer, tamoxifen was shown to reduce the incidence of invasive and noninvasive breast cancers by 50% compared to placebo [74]. Two earlier studies failed to demonstrate this benefit, but they may have been confounded by the inclusion of hysterectomized women, who are at a lower risk for breast cancer, in one study [75, 76], and by a different basis of assessing risk in the other study [77]. These studies also had far fewer participants than the BCPT as well as variable use of HRT among the subjects [75–77]. The recently completed STAR trial, in which the efficacy of tamoxifen in preventing breast cancer was compared with that of raloxifene, showed that these two SERMs were equally efficacious in preventing invasive breast cancers, while the incidence of noninvasive breast cancers was higher in the raloxifene group [52]. However, because tamoxifen is also associated with an increased risk of endometrial carcinoma, thromboembolic complications, and vasomotor side effects [78], it is unlikely that tamoxifen will ever be used extensively as a breast cancer chemopreventive agent.

### 2.4 Raloxifene for Osteoporosis

Raloxifene is a benzothiophene derivative currently approved for the treatment and prevention of postmenopausal osteoporosis. Osteoporosis is a skeletal disease characterized by a progressive deterioration of the bone tissue resulting in an increased risk of fractures [79]. Following menopause, circulating estrogen levels decline, resulting in an imbalance between bone formation and bone resorption, which may lead to rapid bone loss and the development of osteoporosis [80, 81]. At one time known as keoxifene, raloxifene was originally developed as a breast cancer treatment, but after disappointing clinical trial results, raloxifene's development as a breast cancer agent was discontinued in the late 1980s [1]. However, the beneficial effects of SERMs on the maintenance of BMD [82], the possibility that they may not carry the increased risk of breast cancer associated with HRT, and clinical data showing the positive effects of raloxifene on BMD in postmenopausal women [18] led to the successful development of raloxifene as an osteoporosis drug. Furthermore, unlike tamoxifen, raloxifene lacks ER agonistic effects in the endometrium, and thus it has not been associated with an increased risk of endometrial cancer [18, 83]. Given its more favorable side effect profile compared to tamoxifen, and clinical trials data suggesting that it may be useful in breast cancer prevention [84, 85], raloxifene was compared to tamoxifen for efficacy in the primary prevention of breast cancer in the STAR trial [52].

### 2.5 Raloxifene for Reducing the Risk of Breast Cancer

Results from the Multiple Outcomes of Raloxifene Evaluation (MORE) trial, suggesting that raloxifene may be useful for breast cancer chemoprevention [85], and its more favorable side effect profile in regard to endometrial cancer risk [18, 83] led to the genesis of the NSABP P-2 trial, also known as the STAR trial. The STAR trial, a prospective, double-blind, randomized phase III trial comparing the effects of tamoxifen (20 mg/day) and raloxifene (60 mg/day) treatment on breast cancer incidence, was designed to measure the incidence of invasive and noninvasive breast cancer, uterine cancer, and thromboembolic events, among other endpoints, following 5 years of treatment with either raloxifene or tamoxifen in postmenopausal women at increased risk for breast cancer [52, 86, 87]. The results showed no significant difference between the raloxifene and tamoxifen treatment groups with respect to invasive breast cancer, with 168 and 163 cases observed, respectively [52]. For noninvasive breast cancer, however, the number of cases found in the raloxifene group (80) was higher than in the tamoxifen group (57), which was very close to being statistically significant [52]. These results would seem to indicate that raloxifene has little if any effect in preventing noninvasive breast cancer. Raloxifene was superior with respect to the incidence of uterine cancer, with 23 cases found in the raloxifene group compared to 36 cases in the tamoxifen group. This difference was, however, not significant [52]. The incidence of endometrial hyperplasia, on the other hand, was significantly lower in the

raloxifene group (14 cases) compared to the tamoxifen group (84 cases) [52]. When the incidence of thromboembolic events was considered, raloxifene also proved superior to tamoxifen, with 100 events observed in the raloxifene group and 141 events observed in the tamoxifen group [52].

### 2.6 Raloxifene for the Prevention of Cardiovascular Disease

Cardiovascular disease is the most common chronic disease in postmenopausal women in the United States, and it is the leading cause of death in American women. Raloxifene has been shown to reduce low-density lipoprotein (LDL) cholesterol and homocysteine levels in postmenopausal women, although levels of HDL cholesterol and triglycerides are not significantly affected by raloxifene treatment [88, 89]. Clinical data from women taking raloxifene for osteoporosis suggests that raloxifene may provide some cardiovascular benefits [90]. In order to determine whether raloxifene imparts any cardioprotective effects in postmenopausal women at risk of developing cardiovascular disease, the Raloxifene Use for the Heart (RUTH) trial was conducted [91]. The RUTH study was a double-blind, placebo-controlled clinical trial designed to determine the incidence of coronary events and breast cancer as primary outcomes [92]. The results showed that raloxifene (60 mg/day) had no significant effects on the risk of CHD compared to placebo. Moreover, raloxifene treatment was associated with a significantly increased risk of venous thromboembolic events and an increased risk of fatal stroke [91]. The RUTH trial did, however, confirm the benefit of raloxifene in preventing invasive breast cancer. Compared to placebo, raloxifene treatment (60 mg/day) resulted in a significant 44% reduction in the incidence of invasive breast cancer. As in the STAR trial, raloxifene had no significant effect on the incidence of noninvasive breast cancer [91].

Overall, raloxifene appears to be only marginally better than tamoxifen, and, as with tamoxifen, the risks associated with raloxifene treatment must be seriously considered before this drug is used in healthy women. Perhaps more importantly, the results of the STAR and RUTH trials pave the way for newer SERMs, with fewer associated side effects, in breast cancer prevention.

### 2.7 Toremifene for Breast Cancer

Toremifene is currently approved for the treatment of metastatic breast cancer in postmenopausal women. The pharmacological properties of toremifene are very similar to tamoxifen [93–96], which is not surprising considering that toremifene differs from tamoxifen by a single chlorine atom substitution. Multiple phase III trials have demonstrated that toremifene and tamoxifen have essentially equivalent efficacies and side effect profiles in the treatment of advanced breast cancer [97–101]. Two additional clinical trials have evaluated the efficacy of toremifene compared to tamoxifen in the adjuvant setting. As in advanced disease, there was no significant difference between toremifene and tamoxifen for adjuvant use in breast cancer [102, 103]. However, toremifene has not been approved for this indication in the United States. A notable finding in these adjuvant trials was that the incidence of secondary cancers was also similar [102, 103]. Although it has long been thought that toremifene has an advantage over tamoxifen in regard to the incidence of endometrial cancer because of its reduced genotoxic potential [104–106], it is still unclear whether toremifene offers any distinct benefit in this respect [107]. The most recent review of clinical data does, however, suggest that toremifene treatment may be associated with a lower incidence of endometrial carcinomas compared to tamoxifen [108]. In spite of its potential endometrial cancer benefit compared to tamoxifen, toremifene has many of the same side effects as tamoxifen [99] and is thus not being considered as a breast cancer chemopreventive.

### 2.8 Toremifene for the Prevention of Prostate Cancer

Following promising preclinical results seen with toremifene in the Transgenic Adenocarcinoma of Mouse Prostate (TRAMP) model [109], toremifene is currently being evaluated in clinical trials for prostate cancer chemoprevention. In a recently completed phase IIB study involving 514 patients with high-grade prostatic intraepithelial neoplasia (PIN), considered to be an indicator for the future development of prostate cancer, toremifene at a dose of 20 mg/day for 12 months was found to significantly reduce the incidence of prostate cancer compared to placebo [110]. In a previous exploratory phase IIA study in 21 men with high-grade PIN, 13 of the 18 subjects who completed toremifene treatment (60 mg/day for 4 months) were found to be free of high-grade PIN [111]. These promising results have led to the design of a phase III clinical trial of toremifene 20 mg/day compared to placebo for the prevention of prostate cancer in men with high-grade PIN. This trial is currently ongoing and involves 1260 men with high-grade PIN [112]. Toremifene has also been evaluated for efficacy in the treatment of prostate cancer in a phase II trial, but it was found to be ineffective [113]. As an agent for the prevention of prostate cancer, toremifene appears to be very promising.

## 3 AGENTS UNDER DEVELOPMENT AND THEIR POTENTIAL CLINICAL TARGETS

Discussed below are five different investigational SERMs (ospemifene, arzoxifene, bazedoxifene, acolbifene, and lasofoxifene) and SARMs (S-1, S-4, BMS-564929, LGD-2226, and S-40503) and their potential clinical targets.

### 3.1 SERMs

***Ospemifene*** Ospemifene is a triphenylethylene similar in structure to toremifene (Fig. 1). Originally developed as an osteoporosis agent [114, 115], ospemifene has now completed phase III clinical evaluation for the treatment of vulvovaginal atrophy (VVA) due to its estrogenic effects in the vaginal epithelium [116–119], a property that appears to be unique to ospemifene among the current investigational and approved SERMs. In preclinical animal models, ospemifene has also been shown to have breast cancer chemopreventive effects [114, 120, 195]. Phase I studies have indicated that ospemifene is generally safe and well tolerated, with pharmacokinetics suitable for once-a-day dosing [116, 121]. One of the main concerns with investigational SERMs is their potential for estrogenic activity in the uterus. Tamoxifen's estrogenic effect on the endometrium and its association with endometrial cancer are well known, and, in order to be successful, any new SERM must be largely devoid of endometrial side effects. The effects of ospemifene on the uterus have been studied both preclinically and clinically, and the data show that ospemifene has neutral to slight, yet insignificant, estrogen agonist activity in the uterus [114, 116–119, 196], comparable to raloxifene [118].

*Ospemifene for the Treatment of Vulvovaginal Atrophy* Perhaps the most interesting observation from the early clinical studies of ospemifene is that it has a strong estrogenic effect on the vaginal epithelium [116–119], an effect that has not been observed with other SERMs, with the possible exception of lasofoxifene [122]. Raloxifene does not have this effect and may in fact aggravate the condition [123]. Thus, ospemifene may be a novel treatment for urogenital atrophy, also known as vulvovaginal atrophy (VVA), a condition that affects approximately 40% of postmenopausal women [124]. In a placebo-controlled phase II study, the effects of ospemifene (30, 60, and 90 mg/day for 12 weeks) on the vaginal epithelium were assessed using Pap smears. Specifically, the percentages of cells in the parabasal, intermediate, and superficial layers were determined and the changes were compared to placebo. Significant changes were seen at all dose levels for all cell types compared to placebo, except for superficial cells at the 30-mg dose level [117]. Similar effects were observed in another phase II study as well as in a repeated-dose phase I study [116, 118]. Ospemifene has been evaluated in three phase III clinical trials for efficacy in treating VVA, the symptoms of which include vaginal dryness, dyspareunia, and incontinence in postmenopausal women [197]. The first randomized, double-blind phase III trial included 826 postmenopausal women who were randomly assigned to receive either placebo or ospemifene at 30 or 60 mg/day for 12 weeks [196, 198, 199]. All of the subjects in this trial had at least one moderate or severe symptom of VVA in addition to having a vaginal pH greater than 5.0 and less than 5% superficial cells in the vaginal smear. Efficacy of treatment was assessed by measuring four co-primary endpoints: changes from baseline to week 12 in the percentages of superficial and parabasal cells in the vaginal smear (vaginal maturation index), vaginal pH, and self-assessed most bothersome symptom (MBS) of vaginal dryness or dyspareunia compared to placebo. Except for dyspareunia at the 30-mg dose, both doses of ospemifene showed statistically significant improvements from baseline to week 12 in all four co-primary endpoints compared to placebo, with the 60-mg dose being more efficacious than the 30-mg dose. Ospemifene was well tolerated, with no significant differences in adverse events between the treatment groups. Only minor, clinically insignificant changes were seen in endometrial thickness with ospemifene treatment [196], as seen in phase II [117].

These positive results were reinforced in the second randomized, double-blind, placebo-controlled phase III clinical trial in 919 postmenopausal women suffering from VVA. Based on their MBS of either vaginal dryness or dyspareunia, these women were stratified into two cohorts [198]. This trial compared ospemifene 60 mg/day for 12 weeks to placebo. The same co-primary endpoints as in the first phase III trial were used in this trial. In the cohort including 314 women reporting vaginal dryness as their MBS, highly significant changes were seen from baseline to week 12 compared to placebo for percentages of superficial and parabasal cells in the vaginal maturation index as well as vaginal pH. Ospemifene also significantly improved the MBS of vaginal dryness. In the second cohort of 605 women identifying dyspareunia as their MBS, very highly significant changes from baseline to week 12 were observed for all four co-primary endpoints compared to placebo [198]. The phase III long-term safety study showed that the positive effects of ospemifene are maintained over a one-year period. This study included 118 women who continued from the second phase III study of ospemifene 60 mg/day versus placebo [198].

The phase III clinical data clearly show that ospemifene is a safe and effective treatment for VVA, and now that the phase III development program for ospemifene has been completed, a New Drug Application (NDA) is expected to be filed with the FDA by the end of 2012.

*Ospemifene for Osteoporosis Prevention and Treatment* Results from preclinical studies [114, 125] and two phase II clinical trials have shown that ospemifene has potential for use in the treatment and prevention of osteoporosis. The first phase II study compared ospemifene at 30, 60, and 90 mg/day to raloxifene 60 mg/day for 12 weeks in postmenopausal women. Levels of biochemical markers of bone resorption (urinary outputs of N- and C-terminal cross-linking telopeptides of type I collagen, NTX and CTX, respectively) and formation [serum osteocalcin, bone-specific alkaline phosphatase (ALP), and procollagen type I N and C propeptide, PINP and PICP, respectively] following ospemifene treatment were comparable to the levels seen following raloxifene treatment, which demonstrates that the bone-sparing effects of ospemifene are equivalent to those of raloxifene [126]. Only minor changes in endometrial thickness were observed in all study groups, and no significant differences in endometrial thickness were seen between groups [118]. In the second phase II study, ospemifene treatment at 30, 60, and 90 mg/day for 3 months was compared to placebo in healthy postmenopausal women [115]. Urinary concentrations of NTX and CTX and serum levels of PINP, PICP, and ALP were measured to assess bone resorption and formation, respectively. A significant, dose-dependent decrease was seen in the levels of NTX at all dose levels compared to placebo, and CTX was significantly decreased at the 90-mg dose level. Levels of PINP, PICP, and bone ALP were significantly decreased at the 60- and 90-mg ospemifene dose levels compared to placebo [115]. Furthermore, ospemifene was not associated with endometrial stimulation, nor did it worsen climacteric symptoms [117].

Together, the preclinical and clinical data on ospemifene's effects in bone indicate that it is effective in reducing bone turnover, similar to raloxifene. However, it should be emphasized that the clinical trial study periods were only 12 weeks, and longer-term studies in more subjects are needed to fully assess the potential of ospemifene to prevent and treat osteoporosis.

*Ospemifene for Breast Cancer Prevention* In vitro and in vivo preclinical data suggest that ospemifene may potentially help prevent the development of breast cancer. In a rat model of dimethylbenzanthracene (DMBA)-induced mammary carcinoma, treatment with ospemifene reduced the incidence of mammary tumors dose dependently [114]. In nude mice bearing MCF-7 human breast cancer xenografts, ospemifene treatment following estrogen withdrawal did not support tumor growth or the expression of pS2 [114]. Ospemifene was also shown to inhibit the growth of MCF-7 xenografts in nude mice compared to control, indicating that ospemifene acts as an antagonist in breast tissue [127]. In a mouse model of DMBA-induced mammary carcinoma, ospemifene was shown to significantly reduce the incidence of breast tumors, similar to tamoxifen, in separate 37- and 52-week studies [120]. Most recently, ospemifene was shown to both prevent and treat breast cancer in an immune-intact transgenic mouse model of spontaneous breast cancer [195]. Although these preclinical results look promising, clinical trials are needed to assess the breast cancer chemopreventive and therapeutic potentials of ospemifene in humans.

***Arzoxifene*** A benzothiophene related to raloxifene, arzoxifene (Fig. 1) has been developed primarily for the prevention and treatment of breast cancer, while also having been studied for efficacy in preventing osteoporosis and the treatment of endometrial cancer. Arzoxifene, currently in phase III development, was first described in 1997 as a new SERM more potent than raloxifene in antagonizing the effects of estradiol in breast and uterine tissue, as well as having bone-sparing and cholesterol-lowering effects in ovariectomized rats [128].

*Arzoxifene for Breast Cancer Prevention and Treatment* In preclinical studies, arzoxifene has been shown to be effective in the treatment and prevention of breast cancer [129–134]. Clinical trials of arzoxifene in breast cancer that have been completed and published thus far include three phase I and two phase II studies [135–138]. In a phase I study of arzoxifene in 32 patients with recurrent or metastatic breast cancer, subjects received 10, 20, 50, or 100 mg/day arzoxifene. Although no clinical responses were observed, six patients had their disease stabilized for a median duration of 7.7 months. Following 12 weeks of therapy, no evidence of endometrial thickening was found. Overall, the treatment was well tolerated, with mild to moderate hot flashes being the most frequently encountered side effect. Given that all study subjects had previously failed tamoxifen therapy, the lack of responses in this trial was not surprising [135].

Two additional phase I trials were conducted to determine whether arzoxifene has any potential as a breast cancer chemopreventive agent. The first study, a phase IA, included 50 pre- or postmenopausal women randomized to receive no treatment or 10, 20, or 50 mg/day arzoxifene. The other study, a phase IB, included 76 postmenopausal women randomized to receive either

placebo or 20 mg/day arzoxifene. In both trials, sex hormone binding globulin levels were increased, insulin-like growth factor 1 (IGF-1) levels decreased, and the ratio of IGF-1 to IGF binding protein 3 was significantly decreased compared to placebo in the phase IB. Decreases in biomarkers of proliferation were more prevalent for arzoxifene 20 mg in the phase IA study, and in the phase IB study, ER expression was significantly decreased by arzoxifene, although changes in proliferation biomarkers were not significantly different from placebo. The most common side effects were hot flashes, headaches, nausea, leg cramps, and breast pain; however, none of these were significantly increased by arzoxifene compared to placebo in the phase IB study [136]. Further study is needed to determine whether arzoxifene could potentially be useful for breast cancer chemoprevention.

The efficacy of arzoxifene in treating locally advanced or metastatic breast cancer was examined in two phase II studies, the first of which included 112 women receiving either 20 or 50 mg/day arzoxifene. Forty-nine of these patients were considered tamoxifen sensitive and 63 were tamoxifen refractory. Interestingly, the response rate among tamoxifen-sensitive patients receiving 20 mg arzoxifene was more than triple that seen at 50 mg (26.1 vs. 8.0%). Time to progression was also more than double at the lower dose. In tamoxifen-refractory patients, no differences in response rate or time to progression were noted between the two doses. The most common side effects were hot flashes, nausea, cutaneous toxicity, neuromotor toxicity, weight changes, alopecia, headache, and vomiting [137]. The other phase II study involved 92 patients given either 20 or 50 mg/day arzoxifene for 12 weeks. The results of this trial were very similar to the first trial; however, a response rate of 42.9% was observed at the 20-mg dose level compared to 27.3% at the 50-mg dose [138].

In a phase III trial of arzoxifene compared to tamoxifen as first-line therapy for breast cancer, women with ER-positive breast cancer were randomized to receive either 20 mg tamoxifen or 20 mg arzoxifene. During the first interim analysis, time to progression was determined and arzoxifene was found to be inferior to tamoxifen, which led to this trial being halted early [139]. The future of arzoxifene in cancer is now uncertain.

*Arzoxifene for the Prevention of Osteoporosis* Arzoxifene's effects on preventing bone loss have been studied in ovariectomized (OVX) rats [128, 140–142]. In a 5-week study in OVX rats administered 0.1 μg–10 mg/kg/day arzoxifene, OVX-induced bone loss was significantly prevented by arzoxifene, with maximal effects observed at the 0.1- to 1.0-mg/kg dose level. Arzoxifene also maintained bone strength in the femoral neck and midshaft. Compared to OVX control, arzoxifene had only slight estrogenic effects on uterine weight, similar to raloxifene; however, no stimulatory effect on the endometrium was observed. The efficacy of arzoxifene in this study was found to be equivalent to both estradiol and raloxifene; however, arzoxifene was more potent than raloxifene [141]. When administered long term (one year), similar effects were observed with arzoxifene dosed orally at 0.1 or 0.5 mg/kg/day in OVX rats. Arzoxifene prevented OVX-induced increases in body weight, serum cholesterol, and prevented bone loss associated with OVX. No significant uterine histological changes were seen, and arzoxifene had only a slight estrogenic effect on uterine weight compared to OVX controls [142]. Although these preclinical data appear promising, no clinical data regarding the effects of arzoxifene on bone health have been published. There are, however, phase III studies of arzoxifene underway that should help determine the potential of arzoxifene in postmenopausal osteoporosis [139].

***Bazedoxifene*** An indole derivative that structurally resembles raloxifene, bazedoxifene (Fig. 1) is currently in phase III clinical trials for osteoporosis treatment and prevention, as well as treatment of menopausal symptoms (in combination with conjugated estrogens). Although phase I–II trials of bazedoxifene have been completed, few clinical data have been published. A review summarizing this data was published in 2004 [143]. Published results of pertinent preclinical studies will be discussed, followed by an overview of the available clinical trials data.

*Bazedoxifene for Osteoporosis Prevention and Treatment* The potential of bazedoxifene to prevent bone loss was evaluated using the OVX rat model. Bazedoxifene was administered orally at doses ranging from 0.1 to 3.0 mg/kg/day for 6 weeks. Starting at 0.3 mg/kg, bazedoxifene significantly prevented the loss of BMD in the proximal tibia compared to OVX control. At the lowest dose, 0.1 mg/kg, bazedoxifene reduced total serum cholesterol by 50% compared to OVX control, while uterine wet weight was not significantly affected in any dose group [144]. Similar results were seen in a newer study [145]. When evaluated in immature rats with 3 days of treatment, bazedoxifene at doses of 10 or 100 μg had no significant effects on uterine wet weight and completely antagonized the stimulatory effects of estradiol [144]. In another study in the immature rat model, bazedoxifene produced a small increase in uterine wet weight at 0.5 mg/kg and had no significant effects at 5.0 mg/kg. Interestingly, bazedoxifene antagonized the uterine stimulatory effects of raloxifene in this study [145].

In a phase II trial in 494 healthy postmenopausal women, bazedoxifene was dosed at 5, 10, or 20 mg/day for 90 days in combination with a 600-mg/day calcium supplement. Markers of bone turnover were reduced dose dependently compared to placebo, with statistical significance already observed at 5 mg/day [143]. Similarly suggestive results were observed in a previous phase I study in 107 healthy postmenopausal women. In that study, bazedoxifene was administered orally for 30 days at doses of 1–80 mg/day. Urinary concentrations of NTX were reduced by approximately 20%, with no significant adverse effects [143]. Bazedoxifene's efficacy in the treatment and prevention of osteoporosis is currently being examined in two phase III clinical trials [143, 146].

*Bazedoxifene for the Treatment of Menopausal Symptoms* Preclinically, the effects of bazedoxifene on the relief of menopausal symptoms were assessed using the rat hot flash model. When mature, OVX rats that have developed a morphine dependency are given the opioid antagonist naloxone, an intense vasomotor response can be detected in the tail, where a 4–5°C increase in temperature can be observed within 30 minutes of injection. This response can be inhibited by estradiol. Bazedoxifene antagonized estradiol's effects at doses above 1.0 mg/kg, but not at the effective bone-sparing dose of bazedoxifene (0.3 mg/kg). Thus, bazedoxifene, but not raloxifene, could be given in combination with estrogen to treat osteoporosis without antagonizing estrogen's effects on hot flash alleviation. Bazedoxifene would, however, antagonize the effects of estradiol in the breast and endometrium [145, 147].

In a phase II trial of 10 and 20 mg bazedoxifene in combination with conjugated estrogens, bazedoxifene significantly inhibited the stimulatory effects of the conjugated estrogens on the endometrium at both doses compared to conjugated estrogens alone. Menopausal symptoms were also significantly reduced with the combination compared to placebo [143]. The uterine safety and the efficacy of bazedoxifene in treating menopausal symptoms are currently being examined in a phase III clinical trial of bazedoxifene in combination with conjugated estrogens versus HRT [143]. An additional phase III trial testing bazedoxifene plus conjugated estrogens versus raloxifene for uterine safety is also under way [143]. In regard to uterine safety, results from a 6-month study of bazedoxifene in 497 healthy postmenopausal women revealed no significant changes in endometrial thickness at doses of 2.5–20 mg/day compared to placebo; however, significant decreases in endometrial thickness were observed at the 30- and 40-mg doses. This finding is notable because uterine antagonism has not previously been reported for any of the other SERMs [148].

***Acolbifene*** Acolbifene (Fig. 1) is a benzopyran derivative currently in phase II development for the treatment of breast cancer [149]. This compound is the active metabolite of the precursor molecule EM-800, and among the currently known SERMs, it is the most potent antiestrogen [150]. Numerous preclinical studies have examined the effects of acolbifene and its precursor in preventing bone loss as well as preventing and treating breast cancer. Based on the results of the only currently published clinical trial of acolbifene, the intended target for this new SERM appears to be breast cancer treatment. A summary of the preclinical data and a discussion of the phase II trial results follow.

*Acolbifene for the Treatment and Prevention of Breast Cancer* Preclinically, the potential of acolbifene to prevent and treat breast cancer was studied in the rat model of DMBA-induced mammary carcinoma [151] and in a mouse breast cancer xenograft model [152]. In the prevention study, rats were dosed orally with vehicle control, or 25, 75, or 250 μg/day acolbifene for 9 months. Mammary tumor incidence in all dose groups was significantly reduced compared to control, although tumor incidence was still 28% at the highest dose, compared to an incidence of 95% in the control. Total serum cholesterol, triglycerides, and uterine wet weight were significantly reduced at all doses, while no significant effects were observed on bone mineral content or BMD [151]. When acolbifene (500 μg/day for 287 days) was orally administered to OVX nude mice bearing ZR-75-1 human breast tumor xenografts, the tumors regressed to 20% of pretreatment size and became undetectable in 65% of the treated mice. Interestingly, 93% of the tumors that became undetectable did not recur upon treatment with estradiol for 12 weeks [152].

In a phase II clinical trial in 43 women with tamoxifen-refractory breast cancer, acolbifene, in the form of the EM-800 prodrug, was administered orally at 20 or 40 mg/day for 6 months. An objective response rate of approximately 12% was observed, slightly better than that observed with arzoxifene [138] in this patient population [149]. Stabilization of disease was observed in 10 patients for at least 3 months and in 7 patients for at least 6 months. Median response duration was approximately 8 months. Overall, the treatment was well tolerated, and the most frequently encountered side effects included bone/muscle pain, nausea, fatigue, vomiting, headache, and diarrhea [149, 153]. This study demonstrated that acolbifene is incompletely cross-resistant with tamoxifen and may potentially be of some value for salvage therapy in patients that have failed on tamoxifen. A phase III trial of acolbifene versus the aromatase inhibitor anastrozole in patients who had

failed to respond to tamoxifen was conducted, but acolbifene proved to be far less efficacious than anastrozole, leading to early termination of the trial [154].

*Acolbifene for the Prevention of Osteoporosis* The effects of acolbifene on BMD have been evaluated in the OVX rat model [155–157]. Goss et al. [155] looked at the effects of acolbifene (0.01, 1.0, and 2.5 mg/kg/day for 12 weeks) alone and in combination with estradiol on bone, serum lipids, and the uterus. Compared to OVX controls, femoral and lumbar vertebral BMD was significantly higher in the treated groups after 12 weeks of treatment. Furthermore, the combination of estradiol and acolbifene resulted in higher BMD levels than estradiol treatment alone. Acolbifene treatment was also associated with a significant reduction in the OVX-induced increase in serum markers of bone turnover, and mechanical bone strength was increased. At the two higher dose levels, acolbifene antagonized the stimulatory effects of estradiol on the endometrium, while showing no stimulatory effects as a single agent [155]. A similar study showed that acolbifene, given in combination with estradiol, blocked estradiol's stimulation of uterine and breast tissue, while still maintaining BMD [156]. Thus, the potential exists for acolbifene to be used in combination with estradiol, similar to bazedoxifene. In a study by Martel et al. [157], the effects of acolbifene and raloxifene on the prevention of bone loss in OVX rats were compared. Acolbifene and raloxifene were administered orally at doses of 0.01–1.0 mg/kg/day for 37 weeks. Lumbar spine and femoral BMD were maintained at 90–93% and 85–90% of intact control by acolbifene and raloxifene, respectively. At the lowest dose, acolbifene significantly maintained trabecular bone volume, while raloxifene had no discernable effect. At the highest dose, the two SERMs were equivalent. With respect to endometrial effects, raloxifene, but not acolbifene, was associated with hypertrophy of the uterine endometrium at the highest dose [157]. This study demonstrated that acolbifene is more potent than raloxifene without having stimulatory effects on the endometrium.

The most recent preclinical data suggests that acolbifene also has potential as a cholesterol-lowering agent [158] as well as a treatment for vaginal atrophy in combination with conjugated estrogens and dehydroepiandrosterone [159]. Overall, acolbifene appears to be a very promising new SERM; however, very little clinical data exist for this compound, and thus no firm conclusions can be drawn regarding its clinical potential. One potentially troubling finding is that reactive metabolites of acolbifene have been shown to form DNA adducts and produce DNA damage [160]. The significance of this finding is unknown.

***Lasofoxifene*** Lasofoxifene is another interesting new SERM that has reached phase III development. Like nafoxidine, lasofoxifene (Fig. 1) is a naphthalene derivative; however, lasofoxifene is not a planar molecule, a property that gives it a high degree of potency and much better oral bioavailability than raloxifene [161]. Pharmacokinetic data from phase I studies show that lasofoxifene has a very long terminal half-life of elimination ranging from 165 to 193 h in healthy subjects [162, 163]. Lasofoxifene has been developed primarily for the treatment and prevention of osteoporosis [164, 165]. Preclinical data show that lasofoxifene also has potential utility in the treatment and prevention of breast cancer [166] and vaginal atrophy [122].

*Lasofoxifene for the Treatment and Prevention of Breast Cancer* The potential for lasofoxifene to prevent and treat breast cancer has been studied in the rat model of *N*-nitroso-*N*-methylurea-induced mammary carcinoma [166]. For breast cancer treatment, tamoxifen or lasofoxifene was orally administered to groups of 12 induced rats at doses of 0.1–10 mg/kg/day. Treatments began upon the appearance of the first control tumor with a diameter >1 cm and continued for up to 20 weeks postinduction. Compared to control, lasofoxifene significantly inhibited tumor growth, similar to tamoxifen. For breast cancer prevention, groups of 20 induced rats were treated with tamoxifen (10 mg/kg) or lasofoxifene (0.1–10 mg/kg) starting at 2 weeks postinduction and continuing for 8 weeks. Treatments were then stopped and the rats monitored for 8 additional weeks. At the highest dose level, lasofoxifene significantly increased tumor latency and significantly decreased tumor incidence to 30%, compared to 75% in the control. Tamoxifen at the same dose level decreased tumor incidence to 15% [166]. These preliminary data look promising, but more studies are needed to determine the potential of lasofoxifene to prevent or treat breast cancer.

*Lasofoxifene for the Treatment and Prevention of Osteoporosis* Numerous preclinical studies looking at the ability of lasofoxifene to prevent osteoporosis in rats and monkeys have been performed [122, 161, 166–170]. These studies indicate that lasofoxifene effectively and potently prevents bone loss and reduces bone turnover in OVX rats, and in orchidectomized (ORX) and aged male rats, lasofoxifene has shown an ability to prevent bone loss without affecting the prostate [168, 169]. In OVX cynomolgus monkeys, lasofoxifene (1.0 and 5.0 mg/kg) prevented the loss of BMD and inhibited bone turnover over a 2-year period [170].

Clinically, lasofoxifene has been evaluated in several phase II and phase III clinical trials [165]. In a 2-year phase II study, 410 postmenopausal women were ran-

domized to receive placebo, 0.25 mg/day lasofoxifene, 1.0 mg/day lasofoxifene, or 60 mg/day raloxifene. Changes in lumbar spine and total hip BMD were assessed. Lasofoxifene treatment was associated with significant increases in lumbar spine BMD at both doses compared to placebo and to raloxifene. Lasofoxifene and raloxifene had similar effects on total hip BMD. With respect to biochemical markers of bone turnover, both raloxifene and lasofoxifene treatment produced significant reductions compared to placebo, with the reductions following lasofoxifene treatment generally greater than those seen following raloxifene treatment. While the adverse events seen with lasofoxifene and raloxifene were generally similar, there was one important difference. Both doses of lasofoxifene were associated with significant increases in endometrial thickness compared to placebo, while raloxifene did not significantly affect the endometrium [171]. Other phase II studies of lasofoxifene have also demonstrated the positive, bone-sparing effects of this novel SERM [165]. Preliminary results from a completed 2-year phase III trial in 1,907 healthy postmenopausal women show significant increases in lumbar spine and femoral BMD at all dose levels of lasofoxifene (0.025, 0.25, and 0.5 mg/day) compared to placebo. The most frequently reported adverse events were hot flashes and leg cramps [165]. Two additional phase III trials of lasofoxifene are still ongoing, one assessing the efficacy of lasofoxifene in reducing spinal fractures and another comparing the efficacy of lasofoxifene to raloxifene [165].

As alluded to above, lasofoxifene may have beneficial, estrogen-like effects on the vaginal epithelium, similar to ospemifene. In a long-term study of lasofoxifene in rats, all lasofoxifene-treated animals that were examined had vaginal mucification, compared to about 46% of OVX controls that showed vaginal atrophy [122]. In a phase II study, lasofoxifene, dosed at 0.025, 0.25, and 0.5 mg/day, was associated with significant declines in vaginal pH and parabasal cells and a significant increase in superficial cells compared to placebo following 3 months of treatment at all dose levels [165]. Similar improvements in the same measures of vaginal atrophy were observed in a phase III trial following 1 and 2 years of treatment with lasofoxifene at the same doses [165]. Results from a short-term study using immature OVX rats suggest that increased levels of ERβ and AR may be the mechanism through which lasofoxifene exerts its effects on the vaginal epithelium [172]. Finally, in what could be a potential safety concern, lasofoxifene was reported to stimulate uterine matrix metalloprotease 2 activity in rats to the same degree as estradiol [173].

Although lasofoxifene appears to be a very promising new SERM, the manufacturer received a nonapprovable letter from the FDA in late 2005 following submission of a new drug application seeking approval for lasofoxifene as a treatment for osteoporosis. The future of this compound is thus uncertain.

### 3.2 SARMs

***3-(4-Fluorophenoxy)-2-Hydroxy-2-Methyl-N-[4-Nitro-3-(Trifluoromethyl)Phenyl]-Propanamide (S-1)*** This compound (Fig. 2), referred to in the literature as S-1, is an aryl propionamide derivative, a chemical class of SARMs based on the structure of the antiandrogen bicalutamide. The aryl propionamide SARMs were the first to be described [23]. S-1 is currently in phase I clinical development [28], but no clinical data are available in the literature. In a preclinical study, the tissue-selective pharmacological activity of S-1 was demonstrated using castrated rats. Following 14 days of treatment with S-1 at 0.1–1.0 mg/day, only minor stimulation of the prostate and seminal vesicles was observed compared to castrated controls, but the weight of the levator ani muscle was restored to approximately 74% of intact control. These data show that the activity of this compound is selective for anabolic organs compared to androgenic organs [174]. Pharmacokinetic studies of S-1 in rats at doses ranging from 0.1 to 30 mg/kg show that this compound is rapidly absorbed, and it has an oral bioavailability of approximately 55–60%. S-1 is extensively metabolized and eliminated both in the urine and feces, with a terminal half-life of 3.6–5.2 h [175]. Preclinical data indicate that this compound may potentially be useful in the treatment of BPH [176].

*S-1 and the Treatment of BPH* As a follow-up to the earlier study where S-1 was shown to selectively maintain levator ani muscle mass in castrated male rats without substantially affecting the prostate [174], the effects of S-1 compared to finasteride, a 5α-reductase inhibitor, and hydroxyflutamide, an antiandrogen, were determined in intact male rats to assess the potential of S-1 as a treatment for BPH [176]. Benign prostatic hyperplasia is a common ailment in men over the age of 50, and it is characterized by an enlarged prostate gland, which leads to difficulty in urination [177]. Intact male rats were dosed with S-1 (5, 10, and 25 mg/kg), hydroxyflutamide (0.5, 1, 5, 10, and 25 mg/kg), or finasteride (5 mg/kg) for 9 days. All three compounds caused significant reductions in prostate weight, but only hydroxyflutamide caused a significant decrease in levator ani muscle mass. Hydroxyflutamide treatment also resulted in dose-dependent increases in serum testosterone, luteinizing hormone (LH), and follicle stimulating hormone (FSH) levels, while S-1 did not increase the serum hormone levels. Finasteride, due to

its inhibition of 5α-reductase, caused a significant increase in testosterone levels by 9 days of treatment [176]. This study further demonstrated the tissue-selective effects of S-1 and its potential to be used as a treatment for BPH.

***S-3-(4-Acetylaminophenoxy)-2-Hydroxy-2-Methyl-N-(4-Nitro-3-Trifluoromethylphenyl)-Propionamide (S-4)*** Another aryl propionamide SARM referred to in the literature as S-4 (Fig. 2) is also currently in phase I clinical development [28], although published clinical data are lacking. The synthesis of this compound was described in 2004 [178]. Pharmacokinetic studies in rats using orally (1, 10, and 30 mg/kg) and intravenously (0.5, 1, 10, and 30 mg/kg) administered doses of S-4 revealed that the oral bioavailability of S-4 is essentially complete at the lower doses, and S-4's terminal half-life of elimination is very similar to that of the S-1 sister molecule (2.6–5.3 h) [179]. In vitro metabolism studies showed that deacetylation catalyzed by cytochrome P450 3A4 is the major phase I reaction [180]. Like S-1, S-4 was shown to have tissue-selective anabolic effects in castrated and intact male rats; however, the effects of S-4 on the levator ani muscle were greater than the effects of both S-1 and testosterone [174, 176], suggesting that S-4 may be useful in treating muscle wasting. The most recent preclinical data show that S-4 could potentially be useful in preventing and/or treating osteoporosis in both men and women [181, 182].

*S-4 for the Prevention of Osteoporosis* The abilities of S-4 and DHT to restore ORX-induced bone loss were compared in a 20-week study in ORX male rats. Twelve weeks after being orchidectomized, the rats were treated with S-4 (3 and 10 mg/kg), DHT (3 mg/kg), or vehicle control for 8 weeks. A group of intact, vehicle-treated control rats was included. Serum osteocalcin levels in the S-4 and DHT-treated ORX rats were significantly lower compared to the intact controls, suggesting reduced bone turnover. Compared to the ORX controls, treatment with S-4 increased total body BMD significantly more than DHT. The 10-mg/kg dose of S-4 also produced a significant increase in total body bone mineral content (BMC) compared to ORX controls. In contrast, no significant changes in either BMD or BMC were observed with DHT treatment [181].

In the same study, the effects of S-4 treatment on the soleus and levator ani muscles, the prostate and seminal vesicles, and plasma gonadotropin levels were assessed. Interestingly, soleus muscle mass was not significantly different in the S-4 and DHT-treated rats compared to the ORX controls after 8 weeks of treatment. Soleus muscle strength, however, was significantly increased by S-4 and DHT. As in previous studies, levator ani muscle mass was increased by S-4 to intact control levels. Although their anabolic effects were similar, the effects of S-4 and DHT on the prostate and seminal vesicles were strikingly different. Dihydrotestosterone treatment caused the prostate and seminal vesicle weights to increase to more than 200% of the weights observed in the intact controls, while treatment with S-4 at 3 mg/kg had only minimal effects on the weights of these organs (16 and 17% of intact control, respectively). Both doses of S-4 produced significant decreases in plasma LH and FSH compared to ORX control, suggesting agonist action in the pituitary; however; LH and FSH levels were still significantly higher in the S-4 3-mg/kg group compared to intact controls. Treatment with DHT reduced LH and FSH to intact control levels, as expected [181].

The bone-sparing effects of S-4 have also been shown in a female rat model of osteoporosis. One day after ovariectomy, rats were administered S-4 at doses ranging from 0.1 to 3.0 mg/day by subcutaneous injection. After approximately 4 months of treatment, whole-body BMD, L5–L6 vertebral BMD, cortical and trabecular BMD, and femoral bone strength were assessed. At S-4 doses >0.1 mg/day, whole-body BMD was significantly greater than OVX control. At the 3-mg/day dose, S-4 fully preserved whole-body BMD. The coadministration of bicalutamide partially antagonized the effects of S-4, suggesting that S-4's effects were mediated through the AR. Vertebral L5–L6 BMD in OVX rats increased dose dependently with S-4 treatment. At the highest S-4 dose (3 mg/day), vertebral BMD was fully maintained. Treatment with DHT failed to prevent bone loss in the L5–L6 region. Again, coadministration of bicalutamide partially opposed the bone-sparing effects of S-4 alone. In cortical bone, the two highest doses of S-4 (1.0 and 3.0 mg/day) fully maintained cortical BMC and BMD at the femoral midshaft, while DHT only partially prevented bone loss in this region. With respect to trabecular BMD in the distal femur, S-4 only partially prevented the loss of bone in OVX animals. In S-4-treated intact rats, however, trabecular BMD was significantly increased compared to vehicle-treated intact rats. Finally, femoral bone strength was fully maintained by S-4, similar to DHT [182]. This was the first study to demonstrate the bone-sparing effects of a SARM in a rat model of postmenopausal osteoporosis.

***(7R,7aS)-2-(3-Chloro-4-Cyano-2-Methylphenyl)-7-Hydroxytetrahydro-2H-Pyrrolo[1,2-e]Imidazole-1,3-Dione (BMS-564929)*** BMS-564929 (Fig. 2) is a bicyclic hydantoin derivative, a chemical class of SARMs based on the structure of the antiandrogen nilutamide

[27]. Based on the latest published data, this potent, muscle-selective compound appears to be targeted toward the treatment of age-related musculoskeletal decline, or muscle wasting [183], and has entered phase I clinical development [27, 28, 183], although there are currently no clinical data available in the literature. The muscle versus prostate selectivity of BMS-564929 was studied by measuring AR luciferase reporter gene expression in vitro, and this compound was found to be approximately 20 times as potent in muscle cells compared to prostate cells [183].

*BMS-564929 for the Treatment of Muscle Wasting* The anabolic and androgenic effects of BMS-564929 compared to testosterone were studied in an ORX mature rat model [183]. Starting 14 days after castration, rats were treated orally with BMS-564929 (0.00001–10 mg/kg) or subcutaneously with testosterone propionate (0.03–10 mg/kg). Following a 14-day treatment period, levator ani muscle weights, prostate weights, and serum LH levels were assessed. The $ED_{50}$ (effective dose 50%) of BMS-564929 in levator ani muscle was 0.0009 mg/kg compared to an $ED_{50}$ of 0.14 mg/kg in the prostate, indicating that BMS-564929 is over 150-fold more selective for muscle. Testosterone propionate, in contrast, was only about twice as selective for muscle ($ED_{50}$ 0.21 mg/kg) compared to prostate ($ED_{50}$ 0.42 mg/kg). Thus, BMS-564929 was approximately 200 times as potent as testosterone in stimulating the levator ani muscle. Similar results were seen following 8 weeks of treatment in a separate experiment. One potentially troublesome finding in this study was that BMS-564929 potently suppressed serum LH levels with an $ED_{50}$ of 0.008 mg/kg, or 8 μg/kg [183]. This could have implications for the chronic use of this compound.

***6-N,N-Bis(2,2,2-Trifluoroethyl)Amino-4-Trifluoromethylquinolin-2(1H)-one (LGD-2226)*** Based on available information, potential clinical targets of LGD-2226 (Fig. 2), a quinoline analog, include muscle and bone loss due to male hypogonadism [184, 185]. The quinoline analogs are a novel class of SARMs first described by Ligand Pharmaceuticals [27]. The synthesis of LGD-2226 was recently described [186]. In vitro, the potency of LGD-2226 was similar to that of DHT in an androgen-responsive luciferase reporter assay. In prostate cancer cells, LGD-2226 showed weak partial agonism in activation of the prostate-specific antigen (PSA) promoter [185]. LGD-2226 is reportedly in phase I clinical development [28], but a recent publication suggests that the development of this compound has been abandoned because of preclinical toxicity issues [29]. Discussed below are the results of the only published study of LGD-2226.

*LGD-2226 for the Treatment of Male Hypogonadism* A series of studies was conducted to assess the effects of LGD-2226 on the levator ani muscle, the prostate, and on bone in ORX rats. In a 2-week study comparing LGD-2226 (1–100 mg/kg) to testosterone (1–100 mg/kg), both compounds stimulated the levator ani muscle dose dependently, with sham control levels achieved at 3 mg/kg. Testosterone stimulated prostate growth more potently than LGD-2226, as observed with other SARMs, although LGD-2226 fully restored prostate weight at 100 mg/kg. Likewise, serum LH suppression by LGD-2226 was less potent than that of testosterone [185]. To assess the bone effects of LGD-2226, a 6-week study in ORX rats was performed. Doses of 1, 3, and 10 mg/kg were compared to 100 mg/kg of the synthetic androgen fluoxymesterone. The muscle selectivity displayed by LGD-2226 in the 2-week study was also observed following 6 weeks of treatment. At 3 mg/kg, LGD-2226 fully inhibited the increase in LH caused by ORX. Urinary levels of deoxypyridinoline, a marker of bone degradation, were significantly reduced by LGD-2226 and fluoxymesterone. In cancellous bone, the rates of mineral apposition and bone formation were significantly reduced by LGD-2226, suggesting reduced bone turnover [185]. A 16-week study in ORX rats was also performed to determine the effects of LGD-2226 (1, 3, and 10 mg/kg) in cortical bone and total BMD. Total BMD was significantly increased above sham levels at the two highest doses of LGD-2226. The periosteal mineral apposition and bone formation rates were also significantly increased by LGD-2226. Finally, bone strength was significantly enhanced by LGD-2226 compared to sham control [185]. The results of these studies show that LGD-2226 simultaneously decreased bone turnover in cancellous bone and increased bone formation in cortical bone.

***2-(4-Dimethylamino-6-Nitro-1,2,3,4-Tetrahydroquinolin-2-yl)-2-Methylpropan-1-ol (S-40503)*** This SARM is a tetrahydroquinoline analog (Fig. 2), a chemical class based on the structures of the antiandrogens bicalutamide and hydroxyflutamide and steroidal androgens [27]. The most recent published information indicates that S-40503 is still in preclinical development [28], and thus there are few data on this compound available in the literature. The only peer-reviewed paper that has been published shows that S-40503 could potentially be useful in the treatment of male hypogonadism and possibly postmenopausal osteoporosis [187], and this work is discussed below.

*S-40503 for Male Hypogonadism and Osteoporosis* A series of studies were conducted to examine the effects of S-40503 on the levator ani muscle and prostate in ORX and intact male rats, and on bone in both ORX male and OVX female rats. In ORX rats, S-40503 was injected subcutaneously at doses of 1–30 mg/kg/day for 4 weeks beginning one day after castration. For comparison, groups of rats were administered DHT at doses of 0.01–10 mg/kg. Both S-40503 and DHT increased levator ani muscle mass dose dependently compared to ORX control. At the highest dose (30 mg/kg), S-40503 significantly increased levator ani muscle weight compared to sham control, similar to DHT. Femoral BMD was also increased significantly compared to ORX control by S-40503 at 30 mg/kg. However, unlike the other SARMs discussed, S-40503 had a substantially greater stimulatory effect on the prostate, with the 30-mg/kg dose restoring the prostate to approximately 75% of sham control. Dihydrotestosterone at 10 mg/kg increased prostate weight to nearly 1.5 times that of the sham control. Similar effects were observed on the seminal vesicles [187]. In intact rats, S-40503 at 30 mg/kg/day for 4 weeks did not stimulate the prostate while significantly increasing levator ani muscle mass [187]. To assess the effects of S-40503 in an OVX rat model of osteoporosis, S-40503 (30 mg/kg) was administered for 8 weeks to OVX rats starting 4 weeks after ovariectomy. Dihydrotestosterone (10 mg/kg) and estradiol (20 μg/kg) were included for comparison. Both S-40503 and DHT treatment significantly increased cortical bone BMD and periosteal mineral apposition rate above sham control level, while estradiol had no effect on cortical bone BMD and significantly reduced the periosteal mineral apposition rate below sham control level. Estradiol did, however, cause a significant increase in cancellous bone BMD compared to OVX control, while S-40503 and DHT failed to produce any significant changes in cancellous bone BMD. Biomechanical bone strength was significantly enhanced by S-40503, but not DHT, compared to OVX control [187]. These results showed that S-40503 and DHT primarily stimulated bone formation while estradiol prevented bone resorption.

## 4 FUTURE OF SERMS AND SARMS

The drawbacks of steroidal therapies, for example, HRT and androgen replacement, the side effects of earlier generation SERMs, and the emerging needs of our aging population have led to intensive research and development efforts to discover newer and more efficacious nonsteroidal SERM and SARM therapies that offer the benefits of steroid-based therapies while avoiding their associated side effects. Probably the biggest obstacle in new SERM development so far has been uterine side effects, which has led to a number of compounds being dropped from clinical development, including droloxifene, idoxifene, and levormeloxifene [188]. While raloxifene is relatively devoid of uterine side effects, its other side effects, for example, hot flashes, stroke, and thromboembolism, will likely restrict its use in healthy women. Among the SERMs in late-stage clinical development, ospemifene appears to be one of the more promising agents. In addition to the investigational SERMs already described, there are several new compounds still in early development such as GW5638 (a triphenylethylene derivative), HMR-3339 (a steroid-based SERM), ERA-923 (an indole derivative), and CHF-4056 and CHF-4227.01 (both benzopyran derivatives). The clinical potential of these compounds is still largely unknown, but the future of SERM development looks very promising.

In contrast to SERMs, SARMs are relatively new, and thus all of these agents are still in preclinical or early clinical development. The nonsteroidal SARMs offer a distinct advantage over steroidal androgens in that they are not substrates for aromatase or 5α-reductase and have tissue-selective activity, similar to SERMs. Of the SARMs that have entered clinical trials, S-1, S-4, and BMS-564929 appear to be the most promising based on their preclinical data, and clinical trial results for these compounds are eagerly awaited. Besides the SARMs discussed above, there are several new agents in the very early stages of development described in the recent literature. In preclinical studies, another aryl propionamide SARM, C-6, has shown promise as a potential male contraceptive agent [189]. Two new chemical classes of SARMs, arylisothiocyanates [190] and benzimidazoles, for example, JNJ-26146900 [191], have shown antiprostate cancer effects in vitro and in vivo. A third new SARM chemical class based on the structure of mifepristone (RU486) has also been described [192]. Finally, an interesting new SARM-like compound, fispemifene, has been shown in phase I–II trials to significantly increase testosterone levels in men with secondary hypogonadism [193]. Fispemifene was originally developed as a SERM; however, it is currently being developed as a SARM. Its mechanism may differ from other SARMs in that it functions through the ER to increase testosterone levels in hypogonadal men by inhibiting estrogen feedback inhibition in the pituitary [194]. In summary, SERMs and SARMs research has uncovered a number of promising new ER- and AR-targeted agents for the treatment and/or prevention of breast cancer, osteoporosis, vaginal atrophy, male hypogonadism, and BPH, and we can look forward to more advances in this exciting field in the future.

## 5 CONCLUSIONS

Prior to the discovery of nonsteroidal SERMs and SARMs, the only available treatment modalities for conditions such as menopausal symptoms and postmenopausal osteoporosis in women, and the symptoms associated with hypogonadism in men, were steroid based. Recent clinical trial data have shown that the long-term use of HRT in postmenopausal women is associated with an increased risk of breast cancer and may increase the risk of CHD and dementia, contrary to the results of previous observational studies. There are also concerns that androgen replacement therapy in hypogonadal men may pose a risk to prostate health, in addition to the difficulties inherent with the delivery of steroidal androgens. Thus, there is now great interest in developing nonsteroidal, orally bioavailable therapies, that is, SERMs and SARMs, that can provide the respective benefits of estrogens and androgens while minimizing their associated side effects. Side effects associated with the currently approved SERMs—tamoxifen, toremifene, and raloxifene—preclude their widespread use in a healthy population; however, newer generation SERMs currently in development, such as ospemifene, offer great promise for the treatment of urogenital atrophy, osteoporosis, and, potentially, the chemoprevention of breast cancer. While SARMs are relatively new on the scene, the first such agents being described in 1998, available preclinical data for a number of investigational agents show that these compounds hold tremendous potential for the treatment of male hypogonadism, BPH, prostate cancer, and osteoporosis in both men and women. However, due to the lack of published clinical data on any of these compounds, the positive preclinical observations remain to be proven clinically. Nevertheless, the ongoing development of next-generation SERMs and SARMs offers great promise to meet the emerging and unmet medical needs of our aging population.

## REFERENCES


1. Jordan VC (2003). Antiestrogens and selective estrogen receptor modulators as multifunctional medicines. 1. Receptor interactions. *J Med Chem 46*(6), 883–908.
2. Greenblatt RB, Barfield WE, Jungck EC, Ray AW (1961). Induction of ovulation with MRL-41. Preliminary report. *JAMA 178*, 101–104.
3. Klopper A, Hall M (1971). New synthetic agent for the induction of ovulation: Preliminary trials in women. *Br Med J 1*(5741), 152–154.
4. Williamson JG, Ellis JD (1973). The induction of ovulation by tamoxifen. *J Obstet Gynaecol Br Commonwealth 80*(9), 844–847.
5. Huppert LC (1979). Induction of ovulation with clomiphene citrate. *Fertil Steril 31*(1), 1–8.
6. Goldstein SR (1994). Unusual ultrasonographic appearance of the uterus in patients receiving tamoxifen. *Am J Obstet Gynecol 170*(2), 447–451.
7. Decensi A, Fontana V, Bruno S, Gustavino C, Gatteschi B, Costa A (1996). Effect of tamoxifen on endometrial proliferation. *J Clin Oncol 14*(2), 434–440.
8. Fornander T, Rutqvist LE, Cedermark B, Glas U, Mattsson A, Silfversward C, Skoog L, Somell A, Theve T, Wilking N, et al. (1989). Adjuvant tamoxifen in early breast cancer: Occurrence of new primary cancers. *Lancet 1*(8630), 117–120.
9. Fisher B, Costantino JP, Redmond CK, Fisher ER, Wickerham DL, Cronin WM (1994). Endometrial cancer in tamoxifen-treated breast cancer patients: Findings from the National Surgical Adjuvant Breast and Bowel Project (NSABP) B-14. *J Natl Cancer Inst 86*(7), 527–537.
10. Chang J, Powles TJ, Ashley SE, Gregory RK, Tidy VA, Treleaven JG, Singh R (1996). The effect of tamoxifen and hormone replacement therapy on serum cholesterol, bone mineral density and coagulation factors in healthy postmenopausal women participating in a randomized, controlled tamoxifen prevention study. *Ann Oncol 7*(7), 671–675.
11. Love RR, Wiebe DA, Newcomb PA, Cameron L, Leventhal H, Jordan VC, Feyzi J, DeMets DL (1991). Effects of tamoxifen on cardiovascular risk factors in postmenopausal women. *Ann Intern Med 115*(11), 860–864.
12. Powles TJ, Hickish T, Kanis JA, Tidy A, Ashley S (1996). Effect of tamoxifen on bone mineral density measured by dual-energy X-ray absorptiometry in healthy premenopausal and postmenopausal women. *J Clin Oncol 14*(1), 78–84.
13. Love RR, Mazess RB, Barden HS, Epstein S, Newcomb PA, Jordan VC, Carbone PP, DeMets DL (1992). Effects of tamoxifen on bone mineral density in postmenopausal women with breast cancer. *N Engl J Med 326*(13), 852–856.
14. The Writing Group for the PEPI Trial (1995). Effects of estrogen or estrogen/progestin regimens on heart disease risk factors in postmenopausal women: The Postmenopausal Estrogen/Progestin Interventions (PEPI) Trial. *JAMA 273*(3), 199–208.
15. Riggs BL, Khosla S, Melton LJ III (2002). Sex steroids and the construction and conservation of the adult skeleton. *Endocrinol Rev 23*(3), 279–302.
16. The Writing Group for the PEPI (1996). Effects of hormone therapy on bone mineral density: Results from the postmenopausal estrogen/progestin interventions (PEPI) trial. *JAMA 276*(17), 1389–1396.
17. Rossouw JE, Anderson GL, Prentice RL, LaCroix AZ, Kooperberg C, Stefanick ML, Jackson RD, Beresford SA, Howard BV, Johnson KC, Kotchen JM, Ockene J (2002). Risks and benefits of estrogen plus progestin in healthy

postmenopausal women: Principal results from the Women's Health Initiative randomized controlled trial. *JAMA 288*(3), 321–333.

18. Delmas PD, Bjarnason NH, Mitlak BH, Ravoux AC, Shah AS, Huster WJ, Draper M, Christiansen C (1997). Effects of raloxifene on bone mineral density, serum cholesterol concentrations, and uterine endometrium in postmenopausal women. *N Engl J Med 337*(23), 1641–1647.
19. Anthony M, Williams JK, Dunn BK (2001). What would be the properties of an ideal SERM? *Ann NY Acad Sci 949*, 261–278.
20. Bhasin S, Bremner WJ (1997). Clinical review 85: Emerging issues in androgen replacement therapy. *J Clin Endocrinol Metab 82*(1), 3–8.
21. Kamischke A, Nieschlag E (2004). Progress towards hormonal male contraception. *Trends Pharmacol Sci 25*(1), 49–57.
22. Brawer MK (2003). Androgen supplementation and prostate cancer risk: Strategies for pretherapy assessment and monitoring. *Rev Urol 5*(Suppl 1), S29–S33.
23. Dalton JT, Mukherjee A, Zhu Z, Kirkovsky L, Miller DD (1998). Discovery of nonsteroidal androgens. *Biochem Biophys Res Commun 244*(1), 1–4.
24. Morales A, Tenover JL (2002). Androgen deficiency in the aging male: When, who, and how to investigate and treat. *Urol Clin N Am 29*(4), 975–982.
25. Shibata Y, Ito K, Suzuki K, Nakano K, Fukabori Y, Suzuki R, Kawabe Y, Honma S, Yamanaka H (2000). Changes in the endocrine environment of the human prostate transition zone with aging: Simultaneous quantitative analysis of prostatic sex steroids and comparison with human prostatic histological composition. *Prostate 42*(1), 45–55.
26. Griffiths K (2000). Estrogens and prostatic disease. International Prostate Health Council Study Group. *Prostate 45*(2), 87–100.
27. Cadilla R, Turnbull P (2006). Selective androgen receptor modulators in drug discovery: Medicinal chemistry and therapeutic potential. *Curr Topics Med Chem 6*(3), 245–270.
28. Chen J, Kim J, Dalton JT (2005). Discovery and therapeutic promise of selective androgen receptor modulators. *Mol Interventions 5*(3), 173–188.
29. Segal S, Narayanan R, Dalton JT (2006). Therapeutic potential of the SARMs: Revisiting the androgen receptor for drug discovery. *Expert Opin Investig Drugs 15*(4), 377–387.
30. Feng W, Ribeiro RCJ, Wagner RL, Nguyen H, Apriletti JW, Fletterick RJ, Baxter JD, Kushner PJ, West BL (1998). Hormone-dependent coactivator binding to a hydrophobic cleft on nuclear receptors. *Science 280*(5370), 1747–1749.
31. Brzozowski AM, Pike ACW, Dauter Z, Hubbard RE, Bonn T, Engstrom O, Ohman L, Greene GL, Gustafsson JA, Carlquist M (1997). Molecular basis of agonism and antagonism in the oestrogen receptor. *Nature 389*(6652), 753–758.
32. Moras D, Gronemeyer H (1998). The nuclear receptor ligand-binding domain: Structure and function. *Curr Opin Cell Biol 10*(3), 384–391.
33. Shang Y, Brown M (2002). Molecular determinants for the tissue specificity of SERMs. *Science 295*(5564), 2465–2468.
34. Alen P, Claessens F, Verhoeven G, Rombauts W, Peeters B (1999). The androgen receptor amino-terminal domain plays a key role in p160 coactivator-stimulated gene transcription. *Mol Cell Biol 19*(9), 6085–6097.
35. He B, Gampe RT Jr, Kole AJ, Hnat AT, Stanley TB, An G, Stewart EL, Kalman RI, Minges JT, Wilson EM (2004). Structural basis for androgen receptor interdomain and coactivator interactions suggests a transition in nuclear receptor activation function dominance. *Mol Cell 16*(3), 425–438.
36. Henttu PMA, Kalkhoven E, Parker MG (1997). AF-2 activity and recruitment of steroid receptor coactivator 1 to the estrogen receptor depend on a lysine residue conserved in nuclear receptors. *Mol Cell Biol 17*(4), 1832–1839.
37. He B, Kemppainen JA, Voegel JJ, Gronemeyer H, Wilson EM (1999). Activation function 2 in the human androgen receptor ligand binding domain mediates interdomain communication with the $NH_2$-terminal domain. *J Biol Chem 274*(52), 37219–37225.
38. He B, Wilson EM (2002). The $NH_2$-terminal and carboxyl-terminal interaction in the human androgen receptor. *Mol Genet Metab 75*(4), 293–298.
39. Schaufele F, Carbonell X, Guerbadot M, Borngraeber S, Chapman MS, Ma AA, Miner JN, Diamond MI (2005). The structural basis of androgen receptor activation: Intramolecular and intermolecular amino-carboxy interactions. *Proc Natl Acad Sci USA 102*(28), 9802–9807.
40. He B, Lee LW, Minges JT, Wilson EM (2002). Dependence of selective gene activation on the androgen receptor $NH_2$- and COOH-terminal interaction. *J Biol Chem 277*(28), 25631–25639.
41. Li J, Fu J, Toumazou C, Yoon H-G, Wong J (2006). A role of the amino-terminal (N) and carboxyl terminal (C) interaction in binding of androgen receptor to chromatin. *Mol Endocrinol 20*(4), 776–785.
42. Sathya G, Chang C, Kazmin D, Cook CE, McDonnell DP (2003). Pharmacological uncoupling of androgen receptor-mediated prostate cancer cell proliferation and prostate-specific antigen secretion. *Cancer Res 63*(22), 8029–8036.
43. Estebanez-Perpina E, Moore JMR, Mar E, Delgado-Rodrigues E, Nguyen P, Baxter JD, Buehrer BM, Webb P, Fletterick RJ, Guy RK (2005). The molecular mechanisms of coactivator utilization in ligand-dependent transactivation by the androgen receptor. *J Biol Chem 280*(9), 8060–8068.
44. Langley E, Kemppainen JA, Wilson EM (1998). Intermolecular $NH_2$-/carboxy-terminal interactions in

androgen receptor dimerization revealed by mutations that cause androgen insensitivity. *J Biol Chem 273*(1), 92–101.

45. Kraus WE, McInerney EM, Katzenellenbogen BS (1995). Ligand-dependent, transcriptionally productive association of the amino- and carboxyl-terminal regions of a steroid hormone nuclear receptor. *Proc Natl Acad Sci USA 92*(26), 12314–12318.
46. Reid P, Kantoff P, Oh W (1999). Antiandrogens in prostate cancer. *Investig New Drugs 17*(3), 271–284.
47. Mouridsen H, Gershanovich M, Sun Y, Perez-Carrion R, Boni C, Monnier A, Apffelstaedt J, Smith R, Sleeboom HP, Janicke F, Pluzanska A, Dank M, Becquart D, Bapsy PP, Salminen E, Snyder R, Lassus M, Verbeek JA, Staffler B, Chaudi-Ross HA, DuganM (2001). Superior efficacy of letrozole versus tamoxifen as first-line therapy for postmenopausal women with advanced breast cancer: Results of a phase III study of the International Letrozole Breast Cancer Group. *J Clin Oncol 19*(10), 2596–2606.
48. Paridaens R, Dirix L, Lohrisch C, Beex L, Nooij M, Cameron D, Biganzoli L, Cufer T, Duchateau L, Hamilton A, Lobelle JP, Piccart M, European Organization for the Research and Treatment of Cancer (EORTC), Investigational Drug Branch for Breast Cancer (IDBBC) (2003). Mature results of a randomized phase II multicenter study of exemestane versus tamoxifen as first-line hormone therapy for postmenopausal women with metastatic breast cancer. *Ann Oncol 14*(9), 1391–1398.
49. Mauriac L (2006). Aromatase inhibitors: Effective endocrine therapy in the early adjuvant setting for postmenopausal women with hormone-responsive breast cancer. *Best Practice Res Clin Endocrinol Metab 20*(Suppl 1), S15–S29.
50. Jordan VC, Brodie AMH (2007). Development and evolution of therapies targeted to the estrogen receptor for the treatment and prevention of breast cancer. *Steroids 72*(1), 7–25.
51. Taneja SS (2005). Drug therapies for eradicating high-grade prostatic intraepithelial neoplasia in the prevention of prostate cancer. *Rev Urol 7*(Suppl 3), S19–S29.
52. Vogel VG, Costantino JP, Wickerham DL, Cronin WM, Cecchini RS, Atkins JN, Bevers TB, Fehrenbacher L, Pajon ER Jr, Wade JL III, Robidoux A, Margolese RG, James J, Lippman SM, Runowicz CD, Ganz PA, Reis SE, McCaskill-Stevens W, Ford LG, Jordan VC, Wolmark N (2006). Effects of tamoxifen vs. raloxifene on the risk of developing invasive breast cancer and other disease outcomes. *JAMA 295*(23), 2727–2741.
53. Writing Group for the Women's Health Initiative Investigators (2003). Estrogen plus progestin and the risk of coronary heart disease. *N Engl J Med 349*(6), 523–534.
54. Utian WH, Shoupe D, Bachmann G, Pinkerton JV, Pickar JH (2001). Relief of vasomotor symptoms and vaginal atrophy with lower doses of conjugated equine estrogens and medroxyprogesterone acetate. *Fertil Steril 75*(6), 1065–1079.
55. Warren MP, Halpert S (2004). Hormone replacement therapy: Controversies, pros and cons. *Best Practice Res Clin Endocrinol Metab 18*(3), 317–332.
56. Psaty BM, Heckbert SR, Atkins D, Lemaitre R, Koepsell TD, Wahl PW, Siscovick DS, Wagner EH (1994). The risk of myocardial infarction associated with the combined use of estrogens and progestins in postmenopausal women. *Arch Intern Med 154*(12), 1333–1339.
57. Sidney S, Petitti DB, Quesenberry CP Jr (1997). Myocardial infarction and the use of estrogen and estrogen-progestogen in postmenopausal women. *Ann Intern Med 127*(7), 501–508.
58. Rosano GMC, Vitale C, Silvestri A, Fini M (2003). Hormone replacement therapy and cardioprotection. The end of the tale? *Ann NY Acad Sci 997*, 351–357.
59. Paganini-Hill A, Henderson VW (1996). Estrogen replacement therapy and risk of Alzheimer disease. *Arch Intern Med 156*(19), 2213–2217.
60. Carlson MC, Zandi PP, Plassman BL, Tschanz JT, Welsh-Bohmer KA, Steffens DC, Bastian LA, Mehta KM, Breitner JC, Cache County Study Group (2001). Hormone replacement therapy and reduced cognitive decline in older women: The Cache County Study. *Neurology 57*(12), 2210–2216.
61. Henderson VW (1997). Estrogen, cognition, and a woman's risk of Alzheimer's disease. *Am J Med 103*(3A), 11S–18S.
62. Kawas C, Resnick S, Morrison A, Brookmeyer R, Corrada M, Zonderman A, Bacal C, Lingle DD, Metter E (1997). A prospective study of estrogen replacement therapy and the risk of developing Alzheimer's disease: The Baltimore longitudinal study of aging. *Neurology 48*(6), 1517–1521.
63. Manson JE, Hsia J, Johnson KC, Rossouw JE, Assaf AR, Lasser NL, Trevisan M, Black HR, Heckbert SR, Detrano R, Strickland OL, Wong ND, Crouse JR, Stein E, Cushman M (2003). Estrogen plus progestin and the risk of coronary heart disease. *N Engl J Med 349*(6), 523–534.
64. The Women's Health Initiative Steering Committee (2004). Effects of conjugated equine estrogen in postmenopausal women with hysterectomy: The Women's Health Initiative randomized controlled trial. *JAMA 291*(14), 1701–1712.
65. Stefanick ML, Anderson GL, Margolis KL, Hendrix SL, Rodabough RJ, Paskett ED, Lane DS, Hubbell FA, Assaf AR, Sarto GE, Schenken RS, Yasmeen S, Lessin L, Chlebowski RT (2006). Effects of conjugated equine estrogens on breast cancer and mammography screening in postmenopausal women with hysterectomy. *JAMA 295*(14), 1647–1657.
66. Rapp SR, Espeland MA, Shumaker SA, Henderson VW, Brunner RL, Manson JE, Gass MLS, Stefanick ML, Lane DS, Hays J, Johnson KC, Coker LH, Dailey M, Bowen D (2003). Effect of estrogen plus progestin on global cognitive function in postmenopausal women: The Women's Health Initiative Memory Study: A randomized controlled trial. *JAMA 289*(20), 2663–2672.

67. Shumaker SA, Legault C, Rapp SR, Thal L, Wallace RB, Ockene JK, Hendrix SL, Jones BN III, Assaf AR, Jackson RD, Kotchen JM, Wassertheil-Smoller S, Wactawski-Wende J (2003). Estrogen plus progestin and the incidence of dementia and mild cognitive impairment in postmenopausal women: The Women's Health Initiative Memory Study: A randomized controlled trial. *JAMA 289*(20), 2651–2662.
68. Shumaker SA, Legault C, Kuller L, Rapp SR, Thal L, Lane DS, Fillit H, Stefanick ML, Hendrix SL, Lewis CE, Masaki K, Coker LH (2004). Conjugated equine estrogens and the incidence of probable dementia and mild cognitive impairment in postmenopausal women: Women's Health Initiative Memory Study. *JAMA 291*(24), 2947–2958.
69. Espeland MA, Rapp SR, Shumaker SA, Brunner R, Manson JE, Sherwin BB, Hsia J, Margolis KL, Hogan PE, Wallace R, Dailey M, Freeman R, Hays J (2004). Conjugated equine estrogens and global cognitive function in postmenopausal women: Women's Health Initiative Memory Study. *JAMA 291*(24), 2959–2968.
70. Cauley JA, Robbins J, Chen Z, Cummings SR, Jackson RD, LaCroix AZ, LeBoff M, Lewis CE, McGowan J, Neuner J, Pettinger M, Stefanick ML, Wactawski-Wende J, Watts NB (2003). Effects of estrogen plus progestin on risk of fracture and bone mineral density: The Women's Health Initiative randomized trial. *JAMA 290*(13), 1729–1738.
71. Riggs BL, Hartmann LC (2003). Selective estrogen receptor modulators—mechanisms of action and application to clinical practice. *N Engl J Med 348*(7), 618–629.
72. Tormey SM, Malone CM, McDermott EW, O'Higgins NJ, Hill ADK (2006). Current status of combined hormone replacement therapy in clinical practice. *Clin Breast Cancer 6*(Suppl 2), S51–S57.
73. Early Breast Cancer Trialists' Collaborative Group (1998). Tamoxifen for early breast cancer: An overview of the randomized trials. *Lancet 351*(9114), 1451–1467.
74. Fisher B, Costantino JP, Wickerham DL, Redmond CK, Kavanah M, Cronin WM, Vogel V, Robidoux A, Dimitrov N, Atkins J, Daly M, Wieand S, Tan-Chiu E, Ford L, Wolmark N (1998). Tamoxifen for prevention of breast cancer: Report of the National Surgical Adjuvant Breast and Bowel Project P-1 Study. *J Natl Cancer Inst 90*(18), 1371–1388.
75. Veronesi U, Maisonneuve P, Costa A, Sacchini V, Maltoni C, Robertson C, Rotmensz N, Boyle P (1998). Prevention of breast cancer with tamoxifen: Preliminary findings from the Italian randomized trial among hysterectomized women. Italian Tamoxifen Prevention Study. *Lancet 352*(9122), 93–97.
76. Guerrieri-Gonzaga A, Galli A, Rotmensz N, Decensi A (2001). The Italian breast cancer prevention trial with tamoxifen: Findings and new perspectives. *Ann NY Acad Sci 949*, 113–122.
77. Powles T, Eeles R, Ashley S, Easton D, Chang J, Dowsett M, Tidy A, Viggers J, Davey J (1998). Interim analysis of the incidence of breast cancer in the Royal Marsden Hospital tamoxifen randomized chemoprevention trial. *Lancet 352*(9122), 98–101.
78. Morello KC, Wurz GT, DeGregorio MW (2002). SERMs: Current status and future trends. *Crit Rev Oncol/Hematol 43*(1), 63–76.
79. Consensus Development Conference (1991). Prophylaxis and treatment of osteoporosis. *Am J Med 90*(1), 107–110.
80. Fiorelli G, Brandi ML (1999). Skeletal effects of estrogens. *J Endocrinol Investig 22*(8), 589–593.
81. Kuiper GGJM, van den Bemd GJCM, van Leeuwen JPTM (1999). Estrogen receptor and the SERM concept. *J Endocrinol Investig 22*(8), 594–603.
82. Jordan VC, Phelps E, Lindgren JU (1987). Effects of antiestrogens on bone in castrated and intact female rats. *Breast Cancer Res Treatment 10*(1), 31–35.
83. Jordan VC (2003). Antiestrogens and selective estrogen receptor modulators as multifunctional medicines. 2. Clinical considerations and new agents. *J Med Chem 46*(7), 1081–1111.
84. Grady D, Ettinger B, Moscarelli E, Ploufe L Jr, Sarkar S, Ciaccia A, Cummings S, for the Multiple Outcomes of Raloxifene Evaluation Investigators (2004). Safety and adverse effects associated with raloxifene: Multiple outcomes of raloxifene evaluation. *Obstet Gynecol 104*(4), 837–844.
85. Cummings SR, Eckert S, Krueger KA, Grady D, Powles TJ, Cauley JA, Norton L, Nickelsen T, Bjarnason NH, Morrow M, Lippman ME, Black D, Glusman JE, Costa A, Jordan VC (1999). The effect of raloxifene on risk of breast cancer in postmenopausal women: Results from the MORE randomized trial. Multiple Outcomes of Raloxifene Evaluation. *JAMA 281*(23), 2189–2197.
86. Vastag B (2006). Raloxifene prevails in STAR trial, may face easier road to acceptance than previous drugs (news). *J Natl Cancer Inst 98*(11), 733–735.
87. Land SR, Wickerham DL, Constantino JP, Ritter MW, Vogel VG, Lee M, Pajon ER, Wade JL III, Dakhil S, Lockhart JB, Wolmark N, Ganz PA (2006). Patient-reported symptoms and quality of life during treatment with tamoxifen or raloxifene for breast cancer prevention. *JAMA 295*(23), 2742–2751.
88. Walsh BW, Kuller LH, Wild RA, Paul S, Farmer M, Lawrence JB, Shah AS, Anderson PW (1998). Effects of raloxifene on serum lipids and coagulation factors in healthy postmenopausal women. *JAMA 279*(18), 1445–1451.
89. De Leo V, la Marca A, Morgante G, Lanzetta D, Setacci C, Petraglia F (2001). Randomized control study of the effects of raloxifene on serum lipids and homocysteine in older women. *Am J Obstet Gynecol 184*(3), 350–353.
90. Barrett-Connor E, Grady D, Sashegy A, Anderson PW, Cox DA, Hoszowski K, Rautaharju P, Harper KD (2002). Raloxifene and cardiovascular events in osteoporotic

postmenopausal women: Four-year results from the MORE (Multiple Outcomes of Raloxifene Evaluation) randomized trial. *JAMA 287*(7), 847–857.

91. Barrett-Connor E, Mosca L, Collins P, Geiger MJ, Grady D, Kornitzer M, McNabb MA, Wenger NK (2006). Effects of raloxifene on cardiovascular events and breast cancer in postmenopausal women. *N Engl J Med 355*(2), 125–137.
92. Mosca L, Barrett-Connor E, Wenger NK, Collins P, Grady D, Kornitzer M, Moscarelli E, Paul S, Wright TJ, Helterbrand JD, Anderson PW (2001). Design and methods of the raloxifene use for the heart (RUTH) study. *Am J Cardiol 88*(4), 392–395.
93. Robinson SP, Parker CJ, Jordan VC (1990). Preclinical studies with toremiphene as an antitumor agent. *Breast Cancer Res Treatment 16*(Suppl), S9–17.
94. di Salle E, Zaccheo T, Ornati G (1990). Antiestrogenic and antitumour properties of the new triphenylethylene derivative toremifene in the rat. *J Steroid Biochem 36*(3), 203–206.
95. Kangas L (1990). Review of the pharmacological properties of toremifene. *J Steroid Biochem 36*(3), 191–195.
96. Howell SJ, Johnston SRD, Howell A (2004). The use of selective estrogen receptor modulators and selective estrogen receptor down-regulators in breast cancer. *Best Practice Res Clin Endocrinol Metab 18*(1), 47–66.
97. Pyrhonen S, Valavaara R, Modig H, Pawlicki M, Pienkowski T, Gundersen S, Bauer J, Westman G, Lundgren S, Blanco G, Mella O, Nilsson I, Hietanen T, Hindy I, Vuorinen J, Hajba A (1997). Comparison of toremifene and tamoxifen in postmenopausal patients with advanced breast cancer: A randomized double-blind, the "nordic" phase III study. *Br J Cancer 76*(2), 270–277.
98. Hayes DF, Van Zyl JA, Hacking A, Goedhals L, Bezwoda WR, Mailliard JA, Jones SE, . . . Shemano I (1995). Randomized comparison of tamoxifen and two separate doses of toremifene in postmenopausal patients with metastatic breast cancer. *J Clin Oncol 13*(10), 2556–2566.
99. Gershanovich M, Garin A, Baltina D, Kurvet A, Kangas L, Ellmen J (1997). A phase III comparison of two toremifene doses to tamoxifen in postmenopausal women with advanced breast cancer. Eastern European Study Group. *Breast Cancer Res Treatment 45*(3), 251–262.
100. Milla-Santos A, Milla L, Rallo L, Solano V (2001). Phase III randomized trial of toremifene vs. tamoxifen in hormonodependent advanced breast cancer. *Breast Cancer Res Treatment 65*(2), 119–124.
101. Pyrhonen S, Ellmen J, Vuorinen J, Gershanovich M, Tominaga T, Kaufmann M, Hayes DF (1999). Meta-analysis of trials comparing toremifene with tamoxifen and factors predicting outcome of antiestrogen therapy in postmenopausal women with breast cancer. *Breast Cancer Res Treatment 56*(2), 133–143.
102. Holli K, Valavaara R, Blanco G, Kataja V, Hietanen P, Flander M, Pukkala E, Joensuu H (2000). Safety and efficacy results of a randomized trial comparing adjuvant toremifene and tamoxifen in postmenopausal patients with node-positive breast cancer. Finnish Breast Cancer Group. *J Clin Oncol 18*(20), 3487–3494.
103. Holli K (2002). Tamoxifen versus toremifene in the adjuvant treatment of breast cancer. *Eur J Cancer 38*(Suppl 6), S37–S38.
104. Hard GC, Iatropoulos MJ, Jordan K, Radi L, Kaltenberg OP, Imondi AR, Williams GM (1993). Major differences in the hepatocarcinogenicity and DNA adduct forming ability between toremiphene and tamoxifen in female Crl;cd (BR) rats. *Cancer Res 53*(19), 4534–4541.
105. Karlsson S, Hirsimaki Y, Mantyla E, Nieminen L, Kangas L, Hirsimaki P, Perry CJ, Mulhern M, Millar P, Handa J, Williams GM (1996). A two-year dietary carcinogenicity study of the antiestrogen toremifene in Sprague-Dawley rats. *Drug Chem Toxicol 19*(4), 245–266.
106. White IN, de Matteis F, Davies A, Smith LL, Crofton-Sleigh C, Venitt S, Hewer A, Phillips DH (1992). Genotoxic potential of tamoxifen and analogues in female Fischer F344/n rats, DBA/2 and C57BL/6 mice and in human MCL-5 cells. *Carcinogenesis 13*(12), 2197–2203.
107. Pagani O, Gelber S, Price K, Zahrieh D, Gelber R, Simoncini E, Castiglione-Gertsch M, Coates AS, Goldhirsch A (2004). Toremifene and tamoxifen are equally effective for early-stage breast cancer: First results of International Breast Cancer Study Group Trials 12-93 and 14-93. *Ann Oncol 15*(12), 1749–1759.
108. Harvey HA, Kimura M, Hajba A (2006). Toremifene: An evaluation of its safety profile. *Breast 15*(2), 142–157.
109. Raghow S, Hooshdaran MZ, Katiyar S, Steiner MS (2002). Toremifene prevents prostate cancer in the transgenic adenocarcinoma of mouse prostate model. *Cancer Res 62*(5), 1370–1376.
110. Price D, Stein B, Sieber P, Tutrone R, Bailen J, Goluboff E, Burzon D, Bostwick D, Steiner M (2006). Toremifene for the prevention of prostate cancer in men with high grade prostatic intraepithelial neoplasia: Results of double-blind, placebo controlled, phase IIB clinical trial. *J Urol 176*(3), 965–970.
111. Steiner MS, Pound CR (2003). Phase IIa clinical trial to test the efficacy and safety of toremifene in men with high-grade prostatic intraepithelial neoplasia. *Clin Prostate Cancer 2*(1), 24–31.
112. Taneja SS, Smith MR, Dalton JT, Raghow S, Barnette G, Steiner M, Veverka KA (2006). Toremifene—A promising therapy for the prevention of prostate cancer and complications of androgen deprivation therapy. *Expert Opin Investig Drugs 15*(3), 293–305.
113. Stein S, Zoltick B, Peacock T, Holroyde C, Haller D, Armstead B, Malkowicz SB, Vaughn DJ (2001). Phase II trial of toremifene in androgen-independent prostate cancer. *Am J Clin Oncol 24*(3), 283–285.
114. Qu Q, Zheng H, Dahllund J, Laine A, Cockcroft N, Peng Z, Koskinen M, Hemminki K, Kangas L, Vaananen K, Harkonen P (2000). Selective estrogenic effects of a novel triphenylethylene compound, FC1271a, on bone,

cholesterol level and reproductive tissues in intact and ovariectomized rats. *Endocrinology 141*(2), 809–820.

115. Komi J, Heikkinen J, Rutanen E-M, Halonen K, Lammintausta R, Ylikorkala O (2004). Effects of ospemifene, a novel SERM, on biochemical markers of bone turnover in healthy postmenopausal women. *Gynecol Endocrinol 18*(3), 152–158.
116. Voipio SK, Komi J, Kangas L, Halonen K, DeGregorio MW, Erkkola RU (2002). Effects of ospemifene (FC-1271a) on uterine endometrium, vaginal maturation index, and hormonal status in healthy postmenopausal women. *Maturitas 43*(3), 207–214.
117. Rutanen E-M, Heikkinen J, Halonen K, Komi J, Lammintausta R, Ylikorkala O (2003). Effects of ospemifene, a novel SERM, on hormones, genital tract, climacteric symptoms, and quality of life in postmenopausal women: A double-blind, randomized trial. *Menopause 10*(5), 433–439.
118. Komi J, Lankinen KS, Harkonen P, DeGregorio MW, Voipio S, Kivinen S, Tuimala R, Vihtamaki T, Vihko K, Ylikorkala O, Erkkola R (2005). Effects of ospemifene and raloxifene on hormonal status, lipids, genital tract, and tolerability in postmenopausal women. *Menopause 12*(2), 202–209.
119. Gennari L (2004). Ospemifene (Hormos). *Curr Opin Investig Drugs 5*(4), 448–455.
120. Wurz GT, Read KC, Marchisano-Karpman C, Gregg JP, Beckett LA, Yu Q, DeGregorio MW (2005). Ospemifene inhibits the growth of dimethylbenzanthracene-induced mammary tumors in Sencar mice. *J Steroid Biochem Mol Biol 97*(3), 230–240.
121. DeGregorio MW, Wurz GT, Taras TL, Erkkola RU, Halonen KH, Huupponen RK (2000). Pharmacokinetics of (deaminohydroxy)toremifene in humans: A new, selective estrogen-receptor modulator. *Eur J Clin Pharmacol 56*(6–7), 469–475.
122. Ke HZ, Foley GL, Simmons HA, Shen V, Thompson DD (2004). Long-term treatment of lasofoxifene preserves bone mass and bone strength and does not adversely affect the uterus in ovariectomized rats. *Endocrinology 145*(4), 1996–2005.
123. Davies GC, Huster WJ, Lu Y, Plouffe L Jr, Lakshmanan M (1999). Adverse events reported by postmenopausal women in controlled trials with raloxifene. *Obstet Gynecol 93*(4), 558–565.
124. Bachmann GA, Nevadunsky NS (2000). Diagnosis and treatment of atrophic vaginitis. *Am Family Phys 61*(10), 3090–3096.
125. Qu Q, Harkonen PL, Vaananen HK (1999). Comparative effects of estrogen and antiestrogens on differentiation of osteoblasts in mouse bone marrow culture. *J Cell Biochem 73*(4), 500–507.
126. Komi J, Lankinen KS, DeGregorio M, Heikkinen J, Saarikoski S, Tuppurainen M, Halonen K, Lammintausta R, Vaananen K, Ylikorkala O, Erkkola R (2006). Effects of ospemifene and raloxifene on biochemical markers of bone turnover in postmenopausal women. *J Bone Mineral Metab 24*(4), 314–318.
127. Taras TL, Wurz GT, DeGregorio MW (2001). In vitro and in vivo biologic effects of ospemifene (FC-1271a) in breast cancer. *J Steroid Biochem Mol Biol 77*(4–5), 271–279.
128. Palkowitz AD, Glasebrook AL, Thrasher KJ, Hauser KL, Short LL, Phillips DL, Muehl BS, Sato M, Shetler PK, Cullinan GJ, Pell TR, Bryant HU (1997). Discovery and synthesis of [6-hydroxy-3-[4-[2-(1-piperidinyl) ethoxy] phenoxy]-2-(4-hydroxyphenyl)]benzo[*b*] thiophene: A novel, highly potent, selective estrogen receptor modulator. *J Med Chem 40*(10), 1407–1416.
129. Freddie CT, Larsen SS, Bartholomaeussen M, Lykkesfeldt AE (2004). The effect of the new SERM arzoxifene on growth and gene expression in MCF-7 breast cancer cells. *Mol Cell Endocrinol 219*(1–2), 27–36.
130. Detre S, Riddler S, Salter J, A'Hern R, Dowsett M, Johnston SRD (2003). Comparison of the selective estrogen receptor modulator arzoxifene (LY353381) with tamoxifen on tumor growth and biomarker expression in an MCF-7 human breast cancer xenograft model. *Cancer Res 63*(19), 6516–6522.
131. Schafer JM, Lee E-S, Dardes RC, Bentrem D, O'Regan RM, De Los Reyes A, Jordan VC (2001). Analysis of cross-resistance of the selective estrogen receptor modulators arzoxifene (LY353381) and LY117018 in tamoxifen-stimulated breast cancer xenografts. *Clin Cancer Res 7*(8), 2505–2512.
132. Suh N, Glasebrook AL, Palkowitz AD, Bryant HU, Burris LL, Starling JJ, Pearce HL, Williams C, Peer C, Wang Y, Sporn MB (2001). Arzoxifene, a new selective estrogen receptor modulator for chemoprevention of experimental breast cancer. *Cancer Res 61*(23), 8412–8415.
133. Wu L, Tannock IF (2005). Effect of the selective estrogen receptor modulator arzoxifene on repopulation of hormone-responsive breast cancer xenografts between courses of chemotherapy. *Clin Cancer Res 11*(22), 8195–8200.
134. Suh N, Lamph WW, Glasebrook AL, Grese TA, Palkowitz AD, Williams CR, Risingsong R, Farris MR, Heyman RA, Sporn MB (2002). Prevention and treatment of experimental breast cancer with the combination of a new selective estrogen receptor modulator, arzoxifene, and a new rexinoid, LG 100268. *Clin Cancer Res 8*(10), 3270–3275.
135. Munster PN, Buzdar A, Dhingra K, Enas N, Ni L, Major M, Melemed A, Seidman A, Booser D, Theriault R, Norton L, Hudis C (2001). Phase I study of a third-generation selective estrogen receptor modulator, LY353381.HCl, in metastatic breast cancer. *J Clin Oncol 19*(7), 2002–2009.
136. Fabian CJ, Kimler BF, Anderson J, Tawfik OW, Mayo MS, Burak WE Jr, O'Shaughnessy JA, Albain KS, Hyams DM, Budd GT, Ganz PA, Sauter ER, Beenken SW, Grizzle WE, Fruehauf JP, Arneson DW, Bacus JW, Lagios MD, Johnson KA, Browne D (2004). Breast cancer che-

moprevention phase I evaluation of biomarker modulation by arzoxifene, a third generation selective estrogen receptor modulator. *Clin Cancer Res 10*(16), 5403–5417.

137. Buzdar A, O'Shaughnessy JA, Booser DJ, Pippen JE Jr, Jones SE, Munster PN, Peterson P, Melemed AS, Winer E, Hudis C (2003). Phase II, randomized, double-blind study of two dose levels of arzoxifene in patients with locally advanced or metastatic breast cancer. *J Clin Oncol 21*(6), 1007–1014.
138. Baselga J, Llombart-Cussac A, Bellet M, Guillem-Porta B, Enas N, Krejcy K, Carrasco E, Kayitalire L, Kuta M, Lluch A, Vodvarka P, Kerbrat P, Namer M, Petruzelka L (2003). Randomized, double-blind, multicenter trial comparing two doses of arzoxifene (LY353381) in hormone-sensitive advanced or metastatic breast cancer patients. *Ann Oncol 14*(9), 1383–1390.
139. Munster PN (2006). Arzoxifene: The development and clinical outcome of an ideal SERM. *Expert Opin Investig Drugs 15*(3), 317–326.
140. Marchisano-Karpman C, DeGregorio MW (2003). Arzoxifene (Eli Lilly). *IDrugs 6*(9), 880–885.
141. Sato M, Turner CH, Wang T, Adrian MD, Rowley E, Bryant HU (1998). LY353381.HCl: A novel raloxifene analog with improved SERM potency and efficacy *in vivo*. *J Pharmacol Exp Ther 287*(1), 1–7.
142. Ma YL, Bryant HU, Zeng Q, Palkowitz A, Jee WSS, Turner CH, Sato M (2002). Long-term dosing of arzoxifene lowers cholesterol, reduces bone turnover, and preserves bone quality in ovariectomized rats. *J Bone Mineral Res 17*(12), 2256–2264.
143. Gruber C, Gruber D (2004). Bazedoxifene (Wyeth). *Curr Opin Investig Drugs 5*(10), 1086–1093.
144. Miller CP, Collini MD, Tran BD, Harris HA, Kharode YP, Marzolf JT, Moran RA, Henderson RA, Bender RHW, Unwalla RJ, Greenberger LM, Yardley JP, Abou-Gharbia MA, Lyttle CR, Komm BS (2001). Design, synthesis, and preclinical characterization of novel, highly selective indole estrogens. *J Med Chem 44*(11), 1654–1657.
145. Komm BS, Kharode YP, Bodine PVN, Harris HA, Miller CP, Lyttle CR (2005). Bazedoxifene acetate: A selective estrogen receptor modulator with improved selectivity. *Endocrinology 146*(9), 3999–4008.
146. Close P, Neuprez A, Reginster J-Y (2006). Developments in the pharmacotherapeutic management of osteoporosis. *Expert Opin Pharmacother 7*(12), 1603–1615.
147. Komm BS, Lyttle CR (2001). Developing a SERM: Stringent preclinical selection criteria leading to an acceptable candidate (WAY-140424) for clinical evaluation. *Ann NY Acad Sci 949*, 317–326.
148. Ronkin S, Northington R, Baracat E, Nunes MG, Archer DF, Constantine G, Pickar JH (2005). Endometrial effects of bazedoxifene acetate, a novel selective estrogen receptor modulator, in postmenopausal women. *Obstet Gynecol 105*(6), 1397–1404.
149. Labrie F, Champagne P, Labrie C, Roy J, Laverdiere J, Provencher L, Potvin M, Drolet Y, Pollak M, Panasci L, L'Esperance B, Dufresne J, Latreille J, Robert J, Samson B, Jolivet J, Yelle L, Cusan L, Diamond P, Candas B (2004). Activity and safety of the antiestrogen EM-800, the orally active precursor of acolbifene, in tamoxifen-resistant breast cancer. *J Clin Oncol 22*(5), 864–871.
150. Gauthier S, Caron B, Cloutier J, Dory YL, Favre A, Larouche D, Mailhot J, Ouellet C, Schwerdtfeger A, Leblanc G, Martel C, Simard J, Merand Y, Belanger A, Labrie C, Labrie F (1997). (*S*)-(+)-4-[7-(2,2-dimethyl-1-oxopropoxy)-4-methyl-2-[4-[2-(1-piperidinyl)-ethoxy] phenyl]-2*H*-1-benzopyran-3-yl]-phenyl 2,2-dimethylpro panoate (EM-800): A highly potent, specific, and orally active nonsteroidal antiestrogen. *J Med Chem 40*(14), 2117–2122.
151. Luo S, Labrie C, Belanger A, Candas B, Labrie F (1998). Prevention of development of dimethylbenz(*a*)anthracene (DMBA)-induced mammary tumors in the rat by the new nonsteroidal antiestrogen EM-800 (SCH57050). *Breast Cancer Res Treatment 49*(1), 1–11.
152. Roy J, Couillard S, Gutman M, Labrie F (2003). A novel pure SERM achieves complete regression of the majority of human breast cancer tumors in nude mice. *Breast Cancer Res Treatment 81*(3), 223–229.
153. Labrie F, Labrie C, Belanger A, Simard J, Gauthier S, Luu-The V, Merand Y, Giguere V, Candas B, Luo S, Martel C, Singh SM, Fournier M, Coquet A, Richard V, Charbonneau R, Charpenet G, Tremblay A, Tremblay G, Cusan L, Veilleux R (1999). EM-652 (SCH 57068), a third generation SERM acting as pure antiestrogen in the mammary gland and endometrium. *J Steroid Biochem Mol Biol 69*(1–6), 51–84.
154. Vogelvang TE, van der Mooren MJ, Mijatovic V, Kenemans P (2006). Emerging selective estrogen receptor modulators: Special focus on effects on coronary heart disease in postmenopausal women. *Drugs 66*(2), 191–221.
155. Goss PE, Qi S, Cheung AM, Hu H, Mendes M, Pritzker KPH (2004). The selective estrogen receptor modulator SCH 57068 prevents bone loss, reduces serum cholesterol, and blocks estrogen-induced uterine hypertrophy in ovariectomized rats. *J Steroid Biochem Mol Biol 92*(1–2), 79–87.
156. Labrie F, El-Alfy M, Berger L, Labrie C, Martel C, Belanger A, Candas B, Pelletier G (2003). The combination of a novel selective estrogen receptor modulator with an estrogen protects the mammary gland and uterus in a rodent model: The future of postmenopausal women's health? *Endocrinology 144*(11), 4700–4706.
157. Martel C, Picard S, Richard V, Belanger A, Labrie C, Labrie F (2000). Prevention of bone loss by EM-800 and raloxifene in the ovariectomized rat. *J Steroid Biochem Mol Biol 74*(1–2), 45–56.
158. Lemieux C, Gelinas Y, Lalonde J, Labrie F, Richard D, Deshaies Y (2006). Hypocholesterolemic action of the selective estrogen receptor modulator acolbifene in intact and ovariectomized rats with diet-induced hypercholesterolemia. *Metabolism Clin Exp 55*(5), 605–613.

159. Berger L, El-Alfy M, Martel C, Labrie F (2005). Effects of dehydroepiandrosterone, Premarin and acolbifene on histomorphology and sex steroid receptors in the rat vagina. *J Steroid Biochem Mol Biol 96*(2), 201–215.

160. Liu J, Liu H, van Breemen RB, Thatcher GRJ, Bolton JL (2005). Bioactivation of the selective estrogen receptor modulator acolbifene to quinone methides. *Chem Res Toxicol 18*(2), 174–182.

161. Rosati RL, Da Silva Jardine P, Cameron KO, Thompson DD, Ke HZ, Toler SM, Brown TA, Pan LC, Ebbinghaus CF, Reinhold AR, Elliot NC, Newhouse BN, Tjoa CM, Sweetnam PM, Cole MJ, Arriola MW, Gauthier JW, Crawford DT, Nickerson DF, Pirie CM, Qi H, Simmons HA, Tkalcevic GT (1998). Discovery and preclinical pharmacology of a novel, potent, nonsteroidal estrogen receptor agonist/antagonist, CP-336156, a diaryltetrahydronaphthalene. *J Med Chem 41*(16), 2928–2931.

162. Bramson C, Ouellet D, Roman D, Randinitis E, Gardner MJ (2006). A single-dose pharmacokinetic study of lasofoxifene in healthy volunteers and subjects with mild and moderate hepatic impairment. *J Clin Pharmacol 46*(1), 29–36.

163. Gardner M, Taylor A, Wei G, Calcagni A Jr, Duncan B, Milton A (2006). Clinical pharmacology of multiple doses of lasofoxifene in postmenopausal women. *J Clin Pharmacol 46*(1), 52–58.

164. Gennari L (2005). Lasofoxifene (Pfizer). *Curr Opin Investig Drugs 6*(10), 1067–1078.

165. Gennari L, Merlotti D, Martini G, Nuti R (2006). Lasofoxifene: A third-generation selective estrogen receptor modulator for the prevention and treatment of osteoporosis. *Expert Opin Investig Drugs 15*(9), 1091–1103.

166. Cohen LA, Pittman B, Wang C-X, Aliaga C, Yu L, Moyer JD (2001). LAS, a novel selective estrogen receptor modulator with chemopreventive and therapeutic activity in the *N*-nitroso-*N*-methylurea-induced rat mammary tumor model. *Cancer Res 61*(24), 8683–8688.

167. Ke HZ, Paralkar VM, Grasser WA, Crawford DT, Qi H, Simmons HA, Pirie CM, Chidsey-Frink KL, Owen TA, Smock SL, Chen HK, Jee WSS, Cameron KO, Rosati RL, Brown TA, Dasilva-Jardine P, Thompson DD (1998). Effects of CP-336,156, a new, nonsteroidal estrogen agonist/antagonist, on bone, serum cholesterol, uterus, and body composition in rat models. *Endocrinology 139*(4), 2068–2076.

168. Ke HZ, Qi H, Crawford DT, Chidsey-Frink KL, Simmons HA, Thompson DD (2000). Lasofoxifene (CP-336,156), a selective estrogen receptor modulator, prevents bone loss induced by aging and orchidectomy in the adult rat. *Endocrinology 141*(4), 1338–1344.

169. Ke HZ, Qi H, Chidsey-Frink KL, Crawford DT, Thompson DD (2001). Lasofoxifene (CP-336,156) protects against the age-related changes in bone mass, bone strength, and total serum cholesterol in intact aged male rats. *J Bone Mineral Res 16*(4), 765–773.

170. Lees C, Shen V, Brommage R (2007). Effects of lasofoxifene on bone in surgically postmenopausal cynomolgus monkeys. *Menopause 14*(1), 97–105.

171. McClung MR, Siris E, Cummings S, Bolognese M, Ettinger M, Moffett A, Emkey R, Day W, Somayaji V, Lee A (2006). Prevention of bone loss in postmenopausal women treated with lasofoxifene compared with raloxifene. *Menopause 13*(3), 377–386.

172. Wang X-N, Simmons HA, Salatto CT, Cosgrove PG, Thompson DD (2006). Lasofoxifene enhances vaginal mucus formation without causing hypertrophy and increases estrogen receptor β and androgen receptor in rats. *Menopause 13*(4), 609–620.

173. Helvering AM, Adrian MD, Geiser AG, Estrem ST, Wei T, Huang S, Chen P, Dow ER, Calley JN, Dodge JA, Grese TA, Jones SA, Halladay DL, Miles RR, Onyia JE, Ma YL, Sato M, Bryant HU (2005). Differential effects of estrogen and raloxifene on messenger RNA and matrix metalloproteinase 2 activity in the rat uterus. *Biol Reprod 72*(4), 830–841.

174. Yin D, Gao W, Kearbey JD, Xu H, Chung K, He Y, Marhefka CA, Veverka KA, Miller DD, Dalton JT (2003). Pharmacodynamics of selective androgen receptor modulators. *J Pharmacol Exp Ther 304*(3), 1334–1340.

175. Wu D, Wu Z, Yang J, Nair VA, Miller DD, Dalton JT (2006). Pharmacokinetics and metabolism of a selective androgen receptor modulator in rats: Implication of molecular properties and intensive metabolic profile to investigate ideal pharmacokinetic characteristics of a propanamide in preclinical study. *Drug Metab Dispos 34*(3), 483–494.

176. Gao W, Kearbey JD, Nair VA, Chung K, Parlow AF, Miller DD, Dalton JT (2004). Comparison of the pharmacological effects of a novel selective androgen receptor modulator, the 5α-reductase inhibitor finasteride, and the antiandrogen hydroxyflutamide in intact rats: New approach for benign prostate hyperplasia. *Endocrinology 145*(12), 5420–5428.

177. Chacon A, Monga M (1999). Medical management of benign prostatic hyperplasia. *Geriat Nephrol Urol 9*(1), 39–48.

178. Marhefka CA, Gao W, Chung K, Kim J, He Y, Yin D, Bohl C, Dalton JT, Miller DD (2004). Design, synthesis, and biological characterization of metabolically stable selective androgen receptor modulators. *J Med Chem 47*(4), 993–998.

179. Kearbey JD, Wu D, Gao W, Miller DD, Dalton JT (2004). Pharmacokinetics of *S*-3-(4-acetylaminophenoxy)-2-hydroxy-2-methyl-*N*-(4-nitro-3-trifluoromethylphenyl)-propionamide in rats, a non-steroidal selective androgen receptor modulator. *Xenobiotica 34*(3), 273–280.

180. Gao W, Wu Z, Bohl CE, Yang J, Miller DD, Dalton JT (2006). Characterization of the in vitro metabolism of selective androgen receptor modulator using human, rat, and dog liver enzyme preparations. *Drug Metab Dispos 34*(2), 243–253.

181. Gao W, Reiser PJ, Coss CC, Phelps MA, Kearbey JD, Miller DD, and Dalton JT (2005). Selective androgen receptor modulator treatment improves muscle strength and body composition and prevents bone loss in orchidectomized rats. *Endocrinology 146*(11), 4887–4897.
182. Kearbey JD, Gao W, Narayanan R, Fisher SJ, Wu D, Miller DD, Dalton JT (2007). Selective androgen receptor modulator (SARM) treatment prevents bone loss and reduces body fat in ovariectomized rats. *Pharm Res 24*(2), 328–335.
183. Ostrowski J, Kuhns JE, Lupisella JA, Manfredi MC, Beehler BC, Krystek SR Jr, Bi Y, Sun C, Seethala R, Golla R, Sleph PG, Fura A, An Y, Kish KF, Sack JS, Mookhtiar KA, Grover GJ, Hamann LG (2007). Pharmacological and x-ray structural characterization of a novel selective androgen receptor modulator: Potent hyperanabolic stimulation of skeletal muscle with hypostimulation of prostate in rats. *Endocrinology 148*(1), 4–12.
184. Rosen J, Negro-Vilar A (2002). Novel, non-steroidal, selective androgen receptor modulators (SARMs) with anabolic activity in bone and muscle and improved safety profile. *J Musculoskel Neuron Interact 2*(3), 222–224.
185. Miner JN, Chang W, Chapman MS, Finn PD, Hong MH, Lopez FJ, Marschke KB, Rosen J, Schrader W, Turner R, van Oeveren A, Viveros H, Zhi L, Negro-Vilar A (2007). An orally active selective androgen receptor modulator is efficacious on bone, muscle, and sex function with reduced impact on prostate. *Endocrinology 148*(1), 363–373.
186. van Oeveren A, Motamedi M, Mani NS, Marschke KB, Lopez FJ, Schrader WT, Negro-Vilar A, Zhi L (2006). Discovery of 6-*N*,*N*-bis(2,2,2-trifluoroethyl)amino-4-trifluoromethylquinolin-2(1*H*)-one as a novel selective androgen receptor modulator. *J Med Chem 49*(21), 6143–6146.
187. Hanada K, Furuya K, Yamamoto N, Nejishima H, Ichikawa K, Nakamura T, Miyakawa M, Amano S, Sumita Y, Oguro N (2003). Bone anabolic effects of S-40503, a novel nonsteroidal selective androgen receptor modulator (SARM), in rat models of osteoporosis. *Biol Pharm Bull 26*(11), 1563–1569.
188. Albertazzi P, Sharma S (2005). Urogenital effects of selective estrogen receptor modulators: A systematic review. *Climacteric 8*(3), 214–220.
189. Chen J, Hwang DJ, Bohl CE, Miller DD, Dalton JT (2005). A selective androgen receptor modulator for hormonal male contraception. *J Pharmacol Exp Ther 312*(2), 546–553.
190. Hwang DJ, Yang J, Xu H, Rakov IM, Mohler ML, Dalton JT, Miller DD (2006). Arylisothiocyanato selective androgen receptor modulators (SARMs) for prostate cancer. *Bioorg Med Chem 14*(19), 6525–6538.
191. Allan G, Lai M-T, Sbriscia T, Linton O, Haynes-Johnson D, Bhattacharjee S, Dodds R, Fiordeliso J, Lanter J, Sui Z, Lundeen S (2007). A selective androgen receptor modulator that reduces prostate tumor size and prevents orchidectomy-induced bone loss in rats. *J Steroid Biochem Mol Biol 103*(1), 76–83.
192. Kazmin D, Prytkova T, Cook CE, Wolfinger R, Chu T-M, Beratan D, Norris JD, Chang C, McDonnell DP (2006). Linking ligand-induced alterations in androgen receptor structure to differential gene expression: A first step in the rational design of selective androgen receptor modulators. *Mol Endocrinol 20*(6), 1201–1217.
193. QuatRx Pharmaceuticals Press Releases, QuatRx Pharmaceuticals, Ann Arbor, MI. Available: http://www.quatrx.com/news/01_08_07.htm, accessed January 30, 2007.
194. QuatRx Pharmaceuticals Fispemifene Clinical Portfolio, QuatRx Pharmaceuticals, Ann Arbor, MI. Available: http://www.quatrx.com/clinical/fisp.htm, accessed January 30, 2007.
195. Burich RA, McCall JL, Mehta NR, Greenberg BE, Bell KE, Griffey SM, DeGregorio MW, Wurz GT (2011). Ospemifene and 4-hydroxyospemifene effectively prevent and treat breast cancer in the MTag.Tg transgenic mouse model. *Menopause* (IN PRESS).
196. Bachmann GA, Komi JO, and the Ospemifene Study Group (2010). Ospemifene effectively treats vulvovaginal atrophy in postmenopausal women: Results from a pivotal phase 3 study. *Menopause 17*(3), 480–486.
197. Burich RA, DeGregorio MW (2011). Current treatment options for vulvovaginal atrophy. *Expert Rev Obstet Gynecol 6*(2), 141–151.
198. McCall JL, DeGregorio MW (2010). Pharmacologic evaluation of ospemifene. *Expert Opin Drug Metab Toxicol 6*(6), 773–779.
199. Gennari L, Merlotti D, Valleggi F, Nuti R (2009). Ospemifene use in postmenopausal women. *Expert Opin Investig Drugs 18*(6), 839–849.

# IN-DEPTH LOOK: ANTI-TNF-α THERAPIES

Satadal Chatterjee and Shobhan Gaddameedhi

*Department of Pharmaceutical Sciences, College of Pharmacy, Nursing, and Allied Sciences, North Dakota State University, Fargo, North Dakota*

## 1. INTRODUCTION

### 1.1 Tumor Necrosis Factor-α Discovery

Long before the discovery of *tumor necrosis factor-α* (TNF-α), a German physician named P. Bruns reported the regression of tumors after bacterial infection. This led an American oncologist, William Coley, to use bacterial extracts (referred to as Coley's toxins) for the treatment of human cancers [1]. William Coley first observed the phenomenon of tumor necrosis when he saw that the tumors of some of his cancer patients shrank after they had acquired streptococcal infections. Subsequently, in 1975, Lloyd Old and his colleagues, at Memorial Sloan Kettering Cancer Center, identified the TNF-α in the serum of bacterium-primed, endotoxin-challenged mice [2]. It was classified along with other cytokines, such as interferon-γ (IFN-γ) and interleukin-2 (IL-2). Due to its ability to cause tumor necrosis TNF-α soon became the next best hope in cancer treatment. It was tested on cancer patients all over the world, but the results were not as promising as expected.

At the same time, another group at Rockefeller University, headed by Anthony Cerami, was trying to unravel the basis of cachexia—a wasting syndrome associated with many chronic diseases such as cancer, human immunodeficiency virus–acquired immunodeficiency syndrome (HIV–AIDS), tuberculosis (TB), and the like. He found that cachexia was associated with lipemia, the accumulation of lipids in the blood due to suppression of the fat-metabolizing enzyme lipoprotein lipase (LPL) [3], and conjectured that the same factor, which is causing cachexia, is also suppressing LPL, thus triggering lipemia. Furthermore, he had surmised that cachexia and associated lipemia might be due to a protective factor produced by the host in order to fight the infection, but its chronic production became a threat to the body. They tested and validated this hypothesis in a mouse model of endotoxin-induced cachexia [4]. Later, one of Cerami's postdoctoral associates, Bruce Beutler, successfully isolated this factor and their team named this factor as "cachetin" [5, 6]. It was thought to be a unique protein until Beutler sequenced it and found that its sequence was similar to that of human TNF-α discovered earlier by Old's group [7]. The biological activities of TNF-α and cachetin were also found to be alike [8].

The rediscovery of TNF-α as cachetin and elucidation of its role in cachexia explained the side effects seen in cancer patients treated with TNF-α [9]. Beutler's team later found that blocking TNF-α activity using antibodies reduced infection-induced inflammation and endotoxin-induced lethality [10]. Cloning of TNF-α by Goeddel and colleagues in 1984 made sufficient recombinant materials available for study [11]. The studies demonstrated that TNF-α was a multifunctional cytokine that mediates key roles in acute and chronic inflammation and infection. Today, anti-TNF-α drugs are being successfully used in the treatment of various inflammatory diseases such as rheumatoid arthritis (RA) and Crohn's disease. In later years, research showed that TNF-α is involved in the progression of cancer in many

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

other ways besides causing cancer-associated cachexia. Anti-TNF-α therapy aiming to block TNF-α, directly or indirectly, is now being investigated for treatment as well as for supportive care in cancer patients. Currently, it appears to be one of the promising approaches in resistant and nonresistant tumors as well as for cancer-associated cachexia. This chapter will review some of the approaches being exploited to block TNF-α and the anti-TNF-α drugs that have made it to clinical trials or have been approved by U.S. Food and Drug Administration (FDA) for anticancer therapy.

### 1.2 Structure–Function Relationships in the TNF and TNFR Superfamily

TNF-α and lymphotoxin-alpha (LT-α, also called TNF-β) were first identified as products of macrophages and lymphocytes that caused the lysis of certain types of cells, especially tumor cells [2, 12–14]. When the complementary deoxyribonucleic acids (cDNAs) encoding LT-α and TNF-α were cloned [11, 15], they were similar to one another [8], and, eventually, it became clear that they were members of a gene superfamily. Even the receptors for these proteins constitute a TNF-α receptor (TNFR)-related gene superfamily. It is now established that the TNF superfamily is a growing family of cytokines with 19 peptides signaling through 29 receptors [16]. A nomenclature system adopted for these proteins and their receptors is comprised of the term TNFSF for TNF superfamily and TNFRSF for TNFR superfamily and a number ranging from 1 to 19 [17]. The members of the TNFSF along with their receptors as well as their extracellular and cytoplasmic domains are shown in Figure 1 [16]. The functions of all these proteins are yet to be completely elucidated.

Almost all of the ligands of TNFSF are expressed by cells of the immune system, such as B cells, T cells, natural killer (NK) cells, monocytes, and dendritic cells. The only exception to this is the vascular endothelial cell growth inhibitor (VEGI), which is expressed by endothelial cells. In contrast, receptors of TNFRSF are expressed by a wide variety of cells, accounting for the diversity in biological effects observed with these cytokines. The activities mediated by TNFSF ligands and their receptors range from tissue damage to organogenesis. Most of the members are involved in the regulation of cell proliferation and apoptosis and some of them, namely, TNF-α, LT-α, Fas ligand (FasL or CD95L), CD30 ligand (CD30L), and CD40 ligand (CD40L) have pro-inflammatory properties. TNF-related apoptosis-inducing ligand (TRAIL) appears to be unique since it antagonizes autoimmune inflammation, in part, by blocking cell cycle progression [17]. In physiological conditions, members of TNF/TNFRSF are essential for hematopoiesis, protection from bacterial infection, immune surveillance, and tumor regression. Dysregulation of these processes leads to various diseases including AIDS, Crohn's disease, RA, heart failure, asthma, Systemic lupus eythematosus (SLE), as well as cancer [16].

All ligands in TNFSF, except LT-α and VEGI, which are secreted, are type 2 transmembrane proteins (single-pass transmembrane with a cytoplasmic N-terminus and an extracellular or luminal C-terminus). They can have both membrane-bound "pro" and cleaved, soluble "mature" forms [18]. Most members are released from the cell surface by proteolysis through distinct proteases. For example, TNF and receptor activator of nuclear factor-κB (NF-κB) ligand (RANKL) are cleaved by metalloproteases of the ADAM (a disintegrin and metalloprotease domain) family [19, 20], CD95L by matrilysin [21], and B-cell-activating factor (BAFF), ectodermal dysplasin (EDA), TNF-like weak inducer of apoptosis (TWEAK), and a proliferation-inducing ligand (APRIL) by members of the furin family [22, 23]. Both the forms—membrane bound as well as soluble ligand—are active and, infact, in certain cases the soluble ligand can compete with the membrane-bound ligand for the receptor. When bound to the receptor, some soluble ligands act as agonists while others act as antagonists. For instance, CD95L kills human peripheral blood T cells while soluble CD95L blocks this killing [24]. Furthermore, the transmembrane expression of these ligands indicates that they are meant to act locally. However, only under nonphysiological conditions, when they are cleaved, they act on different cell types and prove harmful. This seems to be true for TNF-α, which when released into the circulation, binds to various cell types and mediates various diseases, including cancer [16].

Ligand protomers are composed of antiparallel β-pleated sheets with antiparallel β strands that form a "jelly role" β-structure, which is typical of TNFSF [25]. To be biologically active, ligands have to self-assemble as noncovalent trimers and they do so by interacting at hydrophobic interfaces [26]. The TNFSF ligands share 25–30% amino acid similarity, and this mainly involves the internal aromatic residues responsible for trimer assembly. The external surfaces of ligand trimers show little sequence similarity, and this account for their receptor selectivity [27]. As shown in Figure 1, most of the ligands bind to a single receptor, whereas some can bind to more than one.

A threefold stoichiometry defines the entire signaling cascade mediated by TNF/TNFR SF proteins. The trimeric TNFSF ligand forms a 3:3 symmetric complex with the receptors of TNFRSF. The receptors adopt an elongated structure on the cell membrane, and these

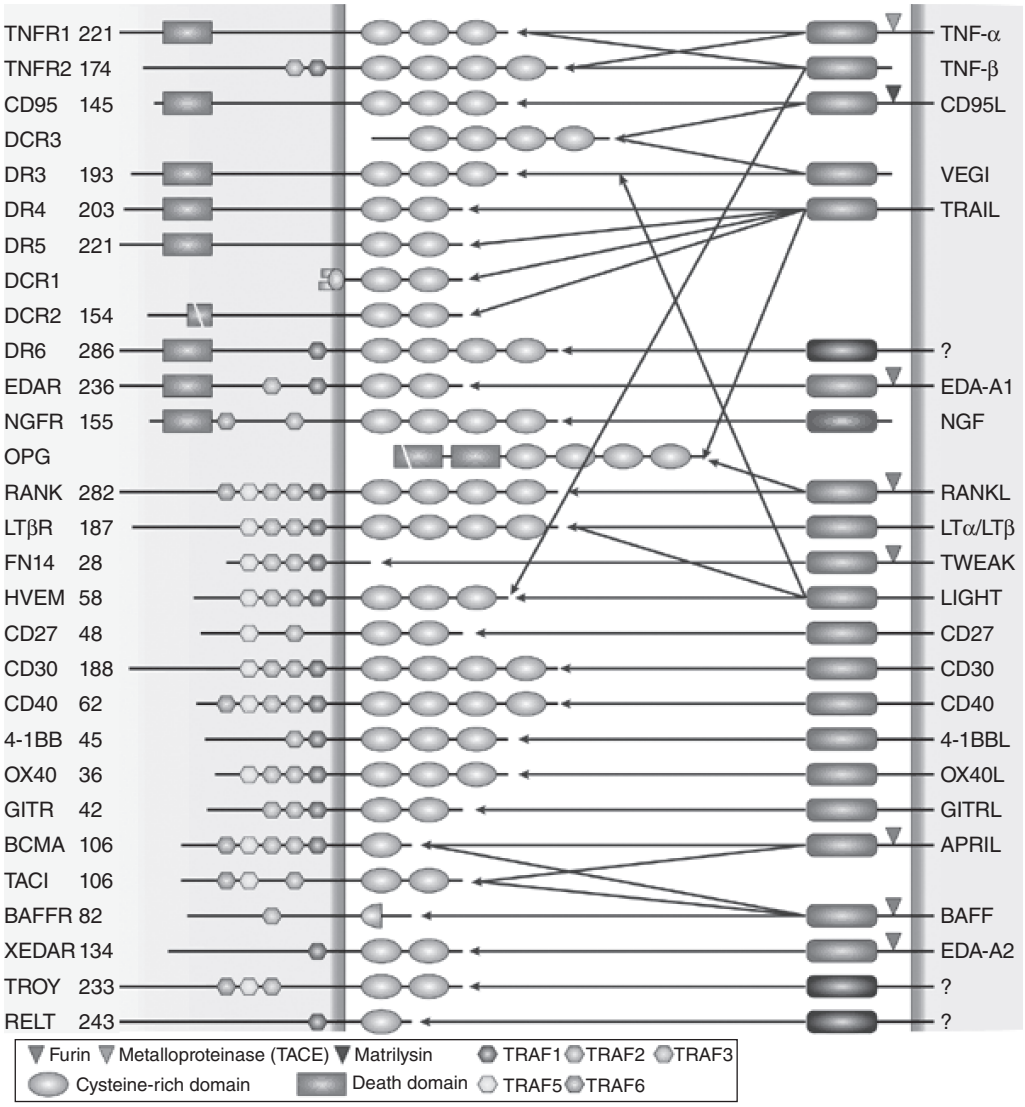


**Figure 1** Description of the interaction of members of the TNFSF with their receptors as well as their cytoplasmic and extracellular domains. (Adapted by permission from Nature Publishing Group [16].)

elongated receptor chains fit into the grooves between protomers of the ligand trimer [27]. Interaction of TNF-α with its receptors is shown in Figure 2 [28].

Receptors of TNFRSF, on the surface of the cell membrane, function as pre-formed complexes rather than as individual receptor subunits. They tend to oligomerize once the ligand binds to it. After ligand binding, the receptor cytoplasmic tails form a 3:3 internal complex with signaling proteins such as TNFR-associated factors (TRAF) and Fas-associated death domain (FADD). Hence, the cytoplasmic domains of these receptors serve as docking sites for various signaling molecules.

Unlike the ligands of TNFSF, receptors in TNFRSF are type 1 transmembrane proteins (single-pass transmembrane with a cytoplasmic C-terminus and an extracellular or luminal N-terminus). The extracellular region of these receptors has a characteristic "cysteine-rich domain" (CRDs), and this is the hallmark of receptors of TNFRSF. CRDs are typically 40 amino acid pseudorepeats and are defined by 3 intrachain disufides generated by 6 highly conserved cysteines [29]. The receptors in TNFRSF can be classified as signaling receptors or nonsignaling "decoy" receptors (DCRs). Based on the type of cytoplasmic region, signaling members of TNFRSF can be further classified as those that contain

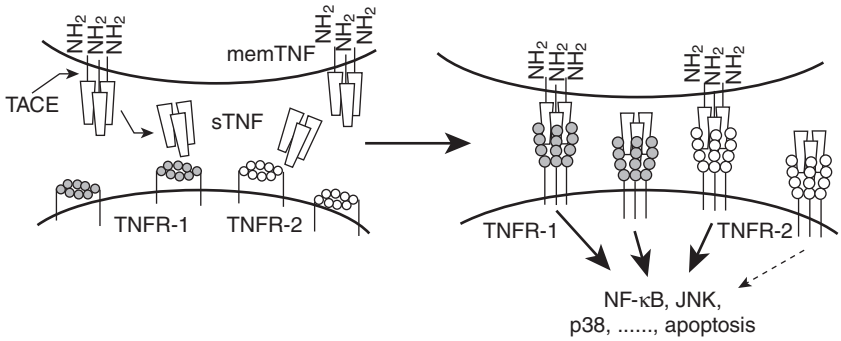


**Figure 2** Interaction of TNF-α with its receptors. (Adapted by permission from Nature Publishing Group [28].)

a cytoplasmic death domain (DD), called death receptors (DRs), and those that are devoid of a cytoplasmic death domain. A DD is a region of approximately 80 amino acid residues and is responsible for death ligand-induced apoptosis that constitutes the cell-extrinsic pathway of apoptosis [30]. The second subgroup of signaling members of TNFRSF lacks a cytoplasmic death domain and instead displays a homologous sequence motifs that mediates interaction with specific adaptors from the TRAF family [31]. In case of TRAF1, TRAF2, and TRAF 5, the interaction requires a Pro-Xaa-Gln-Xaa-Thr motif, while in the case of TRAF6, this interaction is mediated through a distinct motif: Gln-Xaa-Pro-Xaa-Glu [32]. TRAFs link these receptors to serine/threonine protein kinase cascades that regulate gene transcription, such as the inhibitor of κB kinase (IKK), JUN N-terminal kinase (JNK)/p38 and extracellular-signal-regulated kinase/mitogen-activated protein kinase (ERK/MAPK) pathways. In contrast, DCRs are the nonsignaling subset of the TNFRSF that compete with the signaling receptors for binding to the ligand and, hence, attenuate the ligand–receptor signaling [33]. TNFRSF includes 4 decoy receptors (DCRs), namely, DCR1, DCR2, DCR3, and osteoprotegerin (OPG). Hence, the unique structural attributes of the ligands and receptors in the TNF/TNFR superfamily couple them directly to signaling pathways for cell proliferation, survival, and differentiation [27].

### 1.3 TNF-α Signaling through Its Receptors

TNF-α signaling is mediated through two distinct transmembrane, glycoprotein receptors of the TNFRSF: TNFR-1 (also known as p55/60 or CD120a receptor) and TNFR-2 (also known as p75/80 receptor or CD120b receptor). Like other members of the TNFRSF family, TNFR-1 and TNFR-2 carry four CRDs in the extracellular region [34] and assemble as homotrimers on the cell surface, in response to ligand binding. However, there have been some reports suggesting that TNFR chains preassemble into complexes on the cell surface prior to ligand binding, utilizing a conserved domain in the extracellular region called "preligand assembly domain" (PLAD) [35]. Both the receptors of TNF-α are present on all the cell types except endothelial cells. However, their distribution varies with the cell types. While TNFR-1 is ubiquitously expressed on mammalian cells, TNFR-2 expression is mainly restricted to endothelial and immune cells [16, 36]. Thus, in a vast majority of cells, TNFR-1 appears to be the key mediator of TNF-α signaling, whereas in the lymphoid system TNFR-2 seems to play a major role. Also, TNFR-1 expression is constitutive, while that of TNFR-2 is inducible.

In response to ligand binding, both the receptors shed their extracellular domains from the surface of the cells. This proteolytic cleavage is carried out by enzymes of the matrix metalloprotease (MMPs) family, which are zinc-dependent endopeptidases. In case of TNFR-2, this MMP is a TNF-α converting enzyme (TACE), while in case of TNFR-1, the enzyme responsible for cleavage is still undefined. However, TNFR-1 cleavage is biologically significant since cleavage resistant TNFR-1 mutations are linked with dominantly inherited autoinflammatory syndromes (TNFR-1-associated periodic syndromes; TRAPS) [37] Interestingly, the shed extracellular domains retain their ability to bind TNF-α. Therefore, one of the hypothesis says that the shed domains serve as natural antagonists to TNF-α binding to the membrane-bound receptors [38]. Based on the TNF-α neutralizing ability of soluble receptors, dimeric IgG fusion proteins, for example, etanercept, have been synthesized that mimic the soluble receptors and prevent TNF-α binding to the membrane-bound receptor to initiate the signaling cascade. Alternatively, the second hypothesis says that the shed domains serve as ligand

passers (delivery peptides) for the circulating TNF-α, that is, they deliver TNF-α to TNFR for signaling when concentrations of TNF-α are low [17, 39]. The binding affinities for TNF-α differ among the two receptors. TNFR-1 has been shown to have higher affinity for soluble (17-kDa) TNF-α (sTNF) [40], while TNFR-2 can only be fully activated with membrane-bound TNF-α (memTNF) and not sTNF-α [41]. It has been proposed that under some experimental conditions, TNFR-2 may also serve as ligand passer for the circulating TNF-α [42].

TNFRs themselves have no metabolic capability and must bind intracellular proteins in order to regulate the signal transduction events. The major difference between the two receptors lies in the presence of cytoplasmic DD in TNFR-1 and its absence in TNFR-2. Therefore, TNFR-1 is a DR and has the capability of carrying out ligand-induced caspase activation and apoptotic cell death [43]. In contrast, TNFR-2 is a TRAF-interacting receptor. Due to the presence of different cytoplasmic domains, both receptors also differ in their intracellular signaling pathway. Therefore, depending on the receptor type activated by TNF-α and the associated downstream signaling molecules, TNF-α has an optional capacity to induce apoptosis or promote cell survival. The TNF-α signaling pathways are illustrated in Figure 3 [36].

TNF-α is synthesized as a membrane-bound propeptide (pro-TNF-α of 26 kDa). It is secreted as soluble, 51 kDa, homotrimeric cytokine (sTNF) after cleavage by the metalloprotease enzyme (TACE). The sTNF tends to dissociate into 17-kDa protomers, at concentrations below nanomolar range, thereby loosing its bioactivity [28]. The pro-TNF-α is also functional and binds to TNFR via direct cell-to-cell contact, as shown in Figure 2 [28].

***TNFR-1 Signaling*** As explained previously, subsequent to the binding of homotrimeric ligand TNF-α to the TNFR-1, trimerization of TNFR-1 takes place. In the absence of ligand, a protein called silencer of death domain (SODD) functions to mask the intrinsic property of the DD of TNFR-1 to self-aggregate and signal the induction of apoptosis [44]. But as soon as the ligand binds to the receptor, SODD dissociates from TNFR-1 complexes [45] and the DD of TNFR-1 complexes can now interact with an adaptor protein such as TNFR-associated death domain (TRADD) [46]. TNFR-1-bound TRADD serves as an assembly platform for binding of various other adaptor proteins such as FADD and DD containing serine–threonine kinase receptor interacting protein (RIP) or TRAF-2 [47]. RIP is also capable of interacting with TRAF-2 via its amino-terminal kinase domain and its central intermediate domain [48]. These adaptor proteins then recruit key proteins that are responsible for further signaling events.

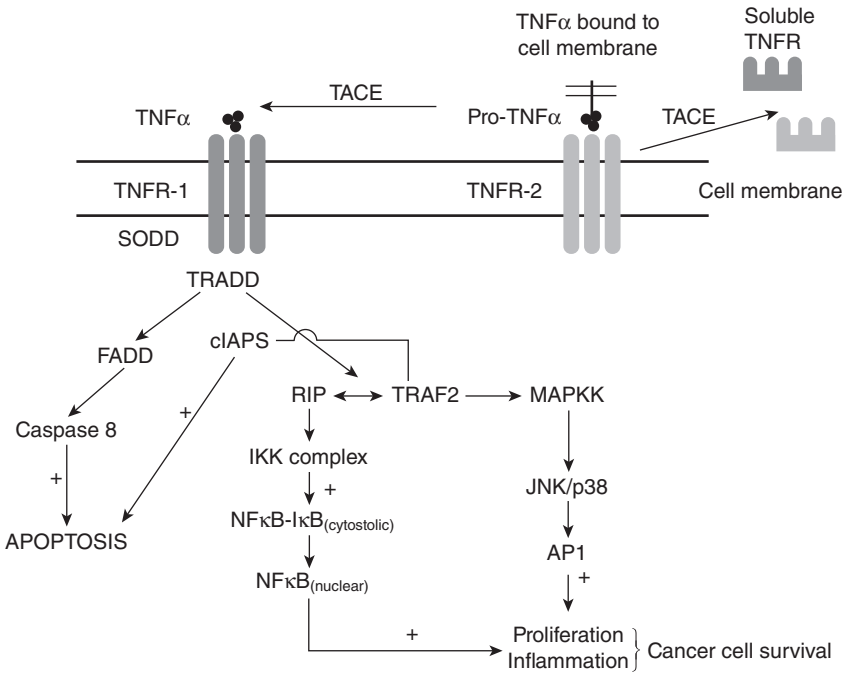


**Figure 3** TNF-α signaling pathways. (Adapted by permission from Elsevier Ltd. [36].)

TNFR-1 is indirectly linked to FADD by the adaptor protein TRADD, which is also responsible for bridging TNFR-1 to TRAF-2 and the IKK complex [49]. Receptor-bound FADD in turn interacts with caspase-8/10 via the death effector domains contained in the amino-terminal parts of both molecules [50, 51]. FADD-mediated recruitment brings several pro-caspase-8/10 molecules in close proximity and thereby facilitates autoproteolytic activation of these proteases. Thus, when TNFR-1 signals apoptosis, FADD binds pro-caspase-8/10, which is subsequently activated to active caspases and results in apoptosis [52]. On the other hand, when TNFR-1 signals cell survival, TRAF-2 is recruited to the complex, which inhibits apoptosis by recruiting the cytoplasmic inhibitor of apoptosis protein (cIAP) [53, 54]. Hence, depletion of TRAF-2, which is a major part of TNFR-1 signaling complex, sensitizes cells for apoptotic action of TNF-α [53, 55, 56].

Subsequent to binding of TRAF-2, transcription factor NF-κB gets activated via RIP. NF-κB comprises a group of transcription factors with five members, namely, c-Rel, Rel A, Rel B, NF-κB1, NF-κB2, and they are characterized by the presence of Rel homology domain (RHD), which serves as their dimerization, DNA binding, and principal regulatory domain [57]. At the C-terminus of RHD, there is a nuclear localization signal (NLS), which is rendered inactive by the bound inhibitors of the NF-κB and IκB proteins. IκB proteins are the specific inhibitors of NF-κB and bind to the RHD through their core domain composed of six to seven ankyrin repeats and thereby mask NLS. Activation of NF-κB, in response TNF-α binding to TNFR-1 or in response to various other cytokines, like IL-1, mostly depends on the phosphorylation-dependent ubiquitination of the IκB proteins. This involves two protein complexes the IKK complex (for phosphorylation of IκB proteins) and the $E3^{I\kappa B}$ ubiquitin ligase complex (for poly-ubiquitination of IκB proteins) [58]. Hence, subsequent to phosphorylation and poly-ubiquitination, the IκB proteins undergo degradation through the 26 S proteosome. The liberated NF-κB dimers have a free NLS to migrate to the nucleus where they can carry out the transcription of the target genes [59]. The target genes for the NF-κB group of transcription factors include a huge number of inflammation-related genes that get activated in response to cytokines such as TNF-α, IL-1, and the like [60]. Besides, the target genes also include a variety of antiapoptotic factors, and in this way NF-κB regulates TNFR-1-mediated triggering of cell apoptosis [61]. Thus, the net effect of TNF-α-mediated activation of NF-κB, through TNFR-1, is cell proliferation and inflammation.

Binding of TRAF-2 also initiates a cascade of phosphorylation events resulting in the activation of transcription factors such as c-Fos/c-Jun and activation transcription factor-2 (ATF-2) via the MAPK and JNK/p38 pathway [62]. It has been shown that the TRAF-2 adaptor is necessary for coupling the JNK pathway to TNFR-1 [63, 64]. c-Jun belongs to a group of basic region–leucine zipper proteins that dimerize to form the transcription factor commonly known as activator protein-1 (AP-1) [65]. The AP-1 proteins play an important role in cell proliferation, differentiation, and induction, as well as prevention of apoptosis [65]. TNF-α-induced activation of JNK has also been implicated in upregulation of collagenases [66], the chemoattractant monocyte chemotactic protein-1 (MCP-1) [67] and E-selectin [68].

***TNFR-2 Signaling*** TNFR-2 is a non-DD-containing receptor, but signaling through this receptor has also been associated with induction of apoptosis. With the use of agonistic TNFR-2-specific antibodies, it has been demonstrated that exclusive triggering of this receptor is sufficient to induce cell death in some cells [69–71]. However, unlike TNFR-1 it cannot directly engage an apoptotic program but relies on the induction of endogenously produced memTNF-α, which subsequently activates TNFR-1 [72, 73]. The action of endogenously produced memTNF-α on TNFR-1 is drastically enhanced by the existing stimulation of TNFR-2 [28]. Additionally, a TNFR-2-dependent depletion of TRAF-2 and antiapoptotic cIAP proteins also leads to induction of apoptosis [28, 74].

### 1.4 Biology of TNF-α

TNF-α is produced mainly by macrophages but also by many other cell types, such as lymphoid cells, mast cells, endothelial cells, fibroblasts, and neuronal cells. It was first characterized as a cytokine that caused tumor necrosis in vivo. Contrastingly, it also exhibited antitumor activity by inducing apoptosis. Later, it was identified that TNF-α is one of the pleiotropic cytokines with significant biological functions. This includes (a) modulation of growth, differentiation, and metabolism in various cell types; (b) mediation of cachexia by stimulating lipolysis, inhibiting LPL activity in adipocytes, and by stimulating hepatic lipogenesis; (c) initiation of apoptosis in cancerous or transformed cells, epithelial cells, virally infected cells, and T lymphocytes; and (d) inflammation [75]. sTNF is produced in large amounts in response to lipopolysaccharide (LPS) and other bacterial products. In concert with other cytokines, it also mediates septic shock [76].

Like IL-1, TNF-α is also a powerful inducer of the inflammatory response and is the most important regulator of innate immune response. Inflammatory

responses to TNF-α are mediated both directly and through stimulation of the expression of IL-1 and also via more distal pro-inflammatory cytokines. Secondary mediators that are known to be induced by systemically administered TNF-α include cytokines [IL-1, IL-2, IL-4, IL-6, IL-10, IL-12, IL-18], IFN-γ, transforming growth factor-β (TGF-β), leukemia inhibitory factor (LIF), and macrophage migration inhibitory factor (MIF), hormones (cortisol, epinephrine, glucagon, insulin, and norepinephrine), and many other molecules such as acute phase proteins (APP), leukotrienes, oxygen free radicals, platelet-activating factor (PAF), nitric oxide (NO), and prostaglandins (PGs) [17]. TNF-α also plays an important role in regulation of the T-helper type 1 cells (TH-1) immune response, in turn resulting in increased levels of $CD4^+$ T-cell activation and IFN-γ production. This is particularly important in the pathogenesis of a number of chronic inflammatory diseases such as inflammatory bowel disease and RA, in which a positive feedback loop linking the autoimmune and the inflammatory or innate immune responses has been established [77]. In response to various stimuli and physiological conditions, TNF-α mediates diverse effects in different organs and tissues as shown in Figure 4 [78, 79]. Under various pathological conditions, such as trauma, ischemia, and inflammatory diseases, the expression and release of TNF-α can be rapidly increased in the brain [80]. Clinical studies have found increased expression of TNF-α in cerebrospinal fluid (CSF), plasma, or postmortem brain tissue in patients following stroke, brain injury, HIV dementia, multiple sclerosis, alcoholism, and Alzheimer's disease [80]. It has been shown that central nervous system (CNS) damage can also increase TNF-α active transport into the brain [81]. Accumulating evidence indicates that TNF-α has been implicated as contributing to both neuroprotection and neurodegeneration [82]. Acute inhibition of endogenous TNF-α reduces ischemic and traumatic brain damage in rodent models, suggesting that TNF-α contributes to the brain damage [83, 84].

In vitro studies have demonstrated that TNF-α increases lipolysis in both rodents and human adipocytes [85, 86]. It is not clear how TNF-α activates lipolysis, but studies have shown that TNF-α downregulates the expression of the lipid droplet-associated protein perilipin (PLIN), which is thought to modulate the accession of hormone-sensitive lipase (HSL) to the surface of the fat droplet [87]. A recent clinical trial

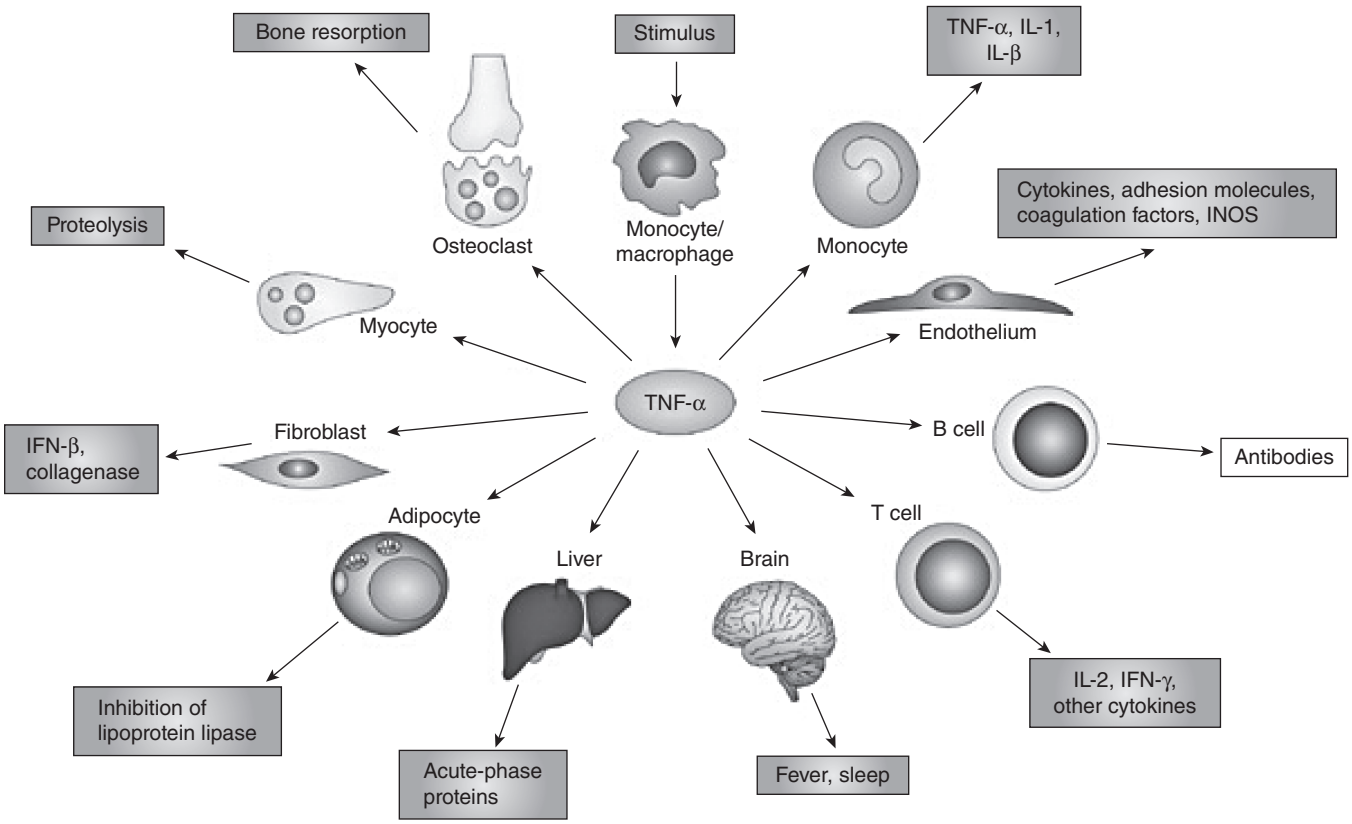


**Figure 4** Various targets of TNF-α action. (Adapted by permission from Nature Publishing Group [78] and Elsevier Ltd. [79].)

in humans demonstrated that anti-TNF-α treatment increased high-density lipoprotein (HDL) without influencing low-density lipoprotein (LDL), which suggests that TNF-α is involved in the development of a high-risk lipid profile [88].

As discussed previously, TNF-α is also a potent activator of NF-κB, and this has been implicated in decreased muscle protein synthesis [89–91]. TNF-α activation is also responsible for the reduction of MyoD protein, a transcription factor that modulates signaling pathways involved in muscle development [92]. It has been shown that TNF-α-mediated skeletal muscle degeneration is associated with cachexia, a debilitating syndrome characterized by extreme weight loss and whole-body wasting [93–95]. Pathophysiology of cachexia and role of TNF-α in mediating its pathophysiology has been discussed in detail in a later section.

### 1.5 Regulation of TNF-α Expression

Macrophages/monocytes form the first line of defense against invading pathogens and also for chemoattraction of other components of the immune system to the site of entry, to effectively prevent invasion by the pathogen. One of the very early responses of macrophages to the invading pathogens is the accumulation of messenger RNA (mRNA) coding for inflammatory and chemotactic polypeptides, including TNF-α, IL-1α and β, and members of the chemokine family (IL-8, MIF), subsequently followed by the secretion of the respective protein. TNF-α is one of the "immediate early" effector proteins induced by a variety of stimuli. Thus, in blood monocytes and tissue macrophages, levels of TNF-α mRNA sharply increase 5- to 50-fold, within 15–30 min after exposure to an inciting agent [96]. Furthermore, the efficiency of mRNA translation can increase over 100-fold [4]. Therefore, TNF-α is one of the major factors secreted by activated macrophages [97], but the quiescent cells produce it only in barely detectable quantities. To regulate the TNF-α expression both transcriptional as well as posttranscriptional mechanisms operate in the body.

Human TNF-α gene locus lies within the highly polymorphic major histocompatibility complex (MHC) class III region located on the shorter arm of chromosome 6 (6p21.3) [98, 99]. The two genes, TNF-α and TNF-β (or LT-α), probably deriving from a duplication of an ancestral gene, are arranged in tandem within a 7-kb stretch of DNA in MHC class III [100]. The TNF-α gene is approximately 3 kbp in length and has four exons and three introns [98].

Most important functional promoter elements of the TNF-α gene have been found in the proximal region up to –200 bp with respect to the transcription start site [101]. TNF-α promoter contains a TNF-responsive element (between –101 and –125 bp) [102]. The regulation by TNF-α of its own promoter (autocrine stimulation) through the TNF-responsive element has been demonstrated in erythroleukemia cells [102]. TNF produced an 8- to 10-fold activation of the human TNF promoter. The TNF-responsive element contains a palindrome, 5′ TGAGCTCA 3′, that resembles the consensus binding sequences for the transcription factors, AP-1, adenosine monophosphate (CAMP)-responsive element binding protein (CREB), and ATF, but the involvement of these transcription factors in the autocrine stimulation is still unproven.

There are a number of biological and chemical substances that trigger the expression of TNF-α. This includes LPS, phorbol esters, mutagens such as dimethylnitrosamine, toxins such as staphylococcal exotoxin B, various vitamins, and cytokines [103, 104]. Many of these have been shown to induce activation of NF-κB transcription factor, thereby leading to NF-κB-induced transcription of the TNF-α gene. Besides this, many tumor cell lines have been found to constitutively express TNF-α mRNA, or can be induced to accumulate the mRNA on induction with any of the agents mentioned above [105, 106]. Many inhibitors of TNF-α mRNA expression have also been reported, and they find clinical application as immunosuppressants, anti-inflammatory, or anticancer agents. This includes dexamethasone, pentoxifylline, protein kinase C inhibitors, cyclosporin C, prostoglandin E2 (PGE2), thalidomide, and its analogs [103, 104]. TNF-α shows tissue-specific expression. Most TNF-α mRNA containing cells do not release TNF-α protein, suggesting that TNF-α in these cells is predominantly regulated by a posttranscriptional mechanism [103].

Posttranscriptional regulation constitutes one of the major and most well-studied means of regulation of TNF-α gene expression [107]. It is generally mediated by the 5′ or 3′ untranslated regions (UTRs) of the mature mRNA, which flank the protein coding sequence. Sequences rich in adenine and uridine (AU-rich elements; ARE) are present in the 3′ UTRs of mRNAs and are recognized by a specific subset of RNA-binding proteins known as ARE-binding proteins (ARE-BPs). AREs are among the predominant *cis*-acting factors that regulate mRNA stability [108]. AREs are composed of a variable number of copies of the AUUUA or UUAUUUAUU nonamers. Mammalian TNFα mRNAs possess a 3′ UTR of almost 800 nucleotides, within which is a well-conserved ARE containing 7 or 8 copies of the AUUUA sequence. The presence of these 3′ AREs was responsible for both the suppression of TNF-α mRNA expression in the unstimulated state as well as the derepression that occurs following

stimulation of macrophages with bacterial endotoxin [109, 110].

The AREs are known to be recognition sequences for several RNA-binding proteins (ARE-BPs) of which only a few have been characterized to date, for example, tristetrapolin (TTP) and Tpl2 [111–114]. Some of these proteins appear to be involved in determining TNF-α mRNA stability, while others are involved in translational silencing. The exact mechanism by which AREs mediate the destabilization of mRNAs is still under intense investigation. Although the current understanding favors that AREs lead to shortening of the poly-A tail (deadenylation) [115, 116] and/or removal of the 5′ cap (decapping) [116]. Targeted mRNA may then be degraded from the 3′ to 5′ end either through the recruitment of 5′–3′ exonucleases or through the sequestration of a complex of 3′–5′ exonucleases called exosome [109, 117–119].

Recently, involvement of micro-RNAs (miRNAs) in ARE-mediated mRNA instability has been reported [120]. miRNAs are genomically encoded short noncoding single-strand RNA species found in a wide variety of organisms [121]. Based on the degree of complementarity to the target mRNA sequence, binding of the miRNA to the mRNA results in translational repression of the protein-coding genes or cleavage of mRNA [122]. In regard to the ARE-mediated mRNA instability of TNF-α, it was found that a miRNA, miR16, is involved in mRNA instability of TNF-α in Drosophila S2 and HeLa cells [120]. By analyzing the sequence of miRNA, it was found that miR16 contains eight bases that can pair with ARE in TNF-α mRNA and has no fortuitous match with TNF-α mRNA in other regions. miR16 also interacts with TTP, and it has been speculated that TTP is a factor that assists miR16 targeting to ARE sequences due to its interaction with eiF2C/argonaute (AGO) family members, which are the components of RNA induced silencing complex (RISC). Also, using an RNAi screen it has been shown that Dicer and argonaute, the key components of the miRNA/RNAi system, are required for the ARE-mediated mRNA decay in Drosophila S2 and HeLa cells. This study suggests that both ARE-BPs and miRNA are involved in recognition of ARE sequence and are essential for ARE-mediated mRNA degradation.

Several studies have suggested the role of microribonucleoproteins (microRNP)—the slicer in miRNA-mediated decay, as regulators of ARE-binding miRNAs. First, RISC, the functional microRNP, includes two ARE-associated proteins: plasminogen activator inhibitor-1 RNA binding protein (PAI-RBP1) and fragile-Xmental-retardation–related protein 1 (FXR1) [123]; FXR1 interacts via AGO 2 [124, 125], which is an essential functional effector of the microRNP [126, 127] or the repressor in miRNA-mediated translational repression [128]. Second, miRNAs have been localized to the same cytoplasmic bodies [129–131] as ARE-BPs [131, 132]. It has now been proposed that translation regulation by microRNPs oscillates between repression and activation during the cell cycle [133]. Upon cell-cycle arrest, the ARE in TNF-α mRNA is transformed into a translation activation signal, recruiting microRNP-associated factors, AGO and FXR1 [134, 135]. Recently, it has been shown that human miRNA miR369-3 directs association of these proteins with the ARE to activate translation. Two well-studied miRNAs—Let-7 and the synthetic miRNA cxcr4—likewise induce translation upregulation of target mRNAs upon cell-cycle arrest; yet, they repress translation in proliferating cells. Thus, activation is a common function of microRNPs upon cell-cycle arrest.

AREs and miRNAs target sites have been found to be conserved sequences in mRNA 3′-untranslated regions (3′ UTRs) that post transcriptionally control gene expression [136]. Now, the involvement of miRNAs in ARE-mediated mRNA instability has been well documented and intense investigation is going on to gain a deeper insight into this area [137].

Besides ARE, a novel constitutive decay element (CDE) has also been detected in the 3′ UTR, which carries out mRNA decay in an ARE-independent mechanism [138, 139]. CDE is also located in 3′ UTR of the TNF-α gene, immediately downstream of ARE (80 nt downstream) and is unresponsive to signals that stabilize ARE-containing mRNAs. It also has distinct protein binding specificities and targets the mRNA transcript for rapid decay.

Highly regulated control of biosynthesis of TNF-α ensures that the TNF-α protein is not produced if there is no exogenous stimulation or in tissues that are not destined to synthesize the protein. Localized and transient secretion of TNF-α is physiological, whereas persistent production occurs in certain autoimmune, inflammatory, and malignant diseases and can lead to severe tissue damage. Thus, an understanding of regulators of TNF-α expression can offer novel means of intervention to control its undesirable production.

## 2 INFLAMMATION AND CANCER—THE TNF-α LINK

### 2.1 Inflammation and Neoplastic Progression

Inflammation includes a series of protective responses initiated whenever the host is under a stressor by internal or external environmental stimuli. Acute inflammation, manifested for a short time, is thought to be protective and therapeutic. However, chronic

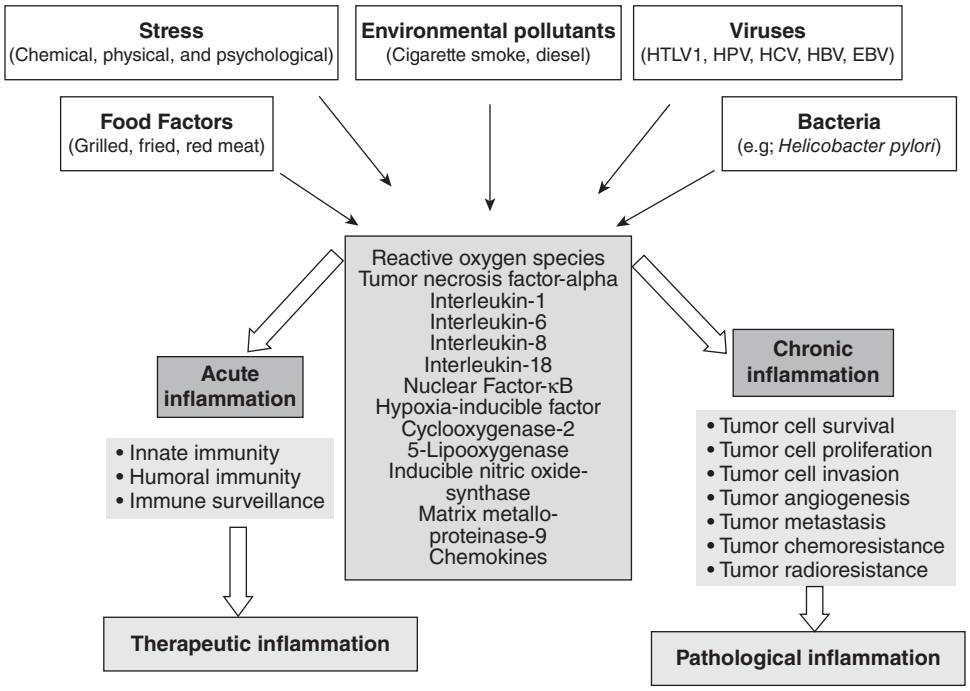


**Figure 5** Initiators and mediators of inflammation and its role in tumorigenesis. (Adapted by permission from Elsevier Ltd. [140].)

inflammation, which lasts for a very long time, may prove harmful and has been thought to be an underlying cause of many diseases, including cardiovascular diseases, diabetes, arthritis, Alzheimer's disease, pulmonary disease, autoimmune disease, as well as cancer [140]. As shown in Figure 5, release of mediators of inflammation can be elicited by various factors, which in turn can lead to acute or chronic inflammation [140]. The role of chronic inflammation has been implicated in various steps in tumorigenesis, including tumor cell survival, proliferation, invasion, tumor angiogenesis, metastasis, chemoresistance, and radioresistance. As shown in Table 1, chronic inflammation, attributable to infectious, chemical, and physical agents, is thought to be a major risk factor for many types of cancers [140]. The link between inflammation and cancer is further strengthened by the observation that long-term use of non-steroidal anti-inflammatory drugs (NSAIDs) decreases cancer risk by almost 40–50% in certain cases [141].

The concept of inflammation promoting cancer is not new and was first proposed by Rudolf Virchow in 1863, but only in recent years has this became a generally accepted phenomenon [142–146]. In fact, chronic inflammation associated with neoplastic progression is visualized as analogous to inflammation associated with wound healing. The distinguishing feature between the two is that inflammation in wound healing is self-

**TABLE 1 Chronic Inflammation and Associated Cancers**

| Inducer | Inflammation | Cancers |
|---|---|---|
| Tobacco smoke | Bronchitis | Lung cancer |
| *Helicobacter pylori* | Gastritis | Gastric cancer |
| Human papillomavirus | Cervicitis | Cervical cancer |
| Hepatitis B and C virus | Hepatitis | Hepatocellular carcinoma |
| Bacteria, gall bladder stones | Cholecystitis | Gall bladder cancer |
| Gram-uropathogens | Cystitis | Bladder cancer |
| Tobacco, genetics | Pancreatitis | Pancreatic cancer |
| Gastric acid, alcohol, tobacco | Esophagitis | Esophageal cancer |
| Asbestos fibres | Asbestosis | Mesothelioma |
| Epstein–Barr virus | Mononucleosis | Burkitt's lymphoma |
| | Hodgkin's disease | |
| Gut pathogens | Inflammatory bowel disease | Colorectal cancer |
| Ultraviolet light | Sunburn | Melanoma |
| Infections, sexually transmitted diseases | Prostate inflammatory atrophy | Prostate cancer |

Adapted by permission from Elsevier Ltd.: *Biochemical Pharmacology* [140] © (2006).

limiting, since secretion of anti-inflammatory cytokines follows the pro-inflammatory cytokines, while tumors act as "wounds that never heal" [147]. The chronic inflammation smoldering beneath cancers continues for a long time either due to persistence of initiating factors or a failure of mechanisms required for terminating the inflammatory responses [144].

Tumor microenvironment, also known as tumor stroma, is rich in inflammatory mediators, and the interplay between them plays a major role in fostering carcinogenesis [143, 148]. There is a continuous battle between these inflammatory mediators to strengthen the immune response against tumors on one hand and promote the growth of tumor cells on the other hand. However, neoplastic progression indicates that inflammatory mediators tend to loose the battle as immune system effector molecules and in fact assist the growth and spread of tumors. The specialized leucocytes present in tumor stroma are tumor-associated dendritic cells (TADCs), tumor-associated macrophages (TAMs), and tumor-infiltrating leucocytes (TILs). TADCs normally have an immature phenotype and lack the ability to stimulate T cells, which in turn reflects an ineffective immune response against the tumor growth [149]. TAMs, once activated, can kill the tumor cells. On the other hand, TAMs can also produce growth and angiogenic factors, which promote tumor growth, as well as proteases, which can degrade the extracellular matrix (ECM) and hence stimulate tumor invasion and metastasis [150, 151]. Signaling via T-cell receptor (TCR) is also defective in TILs, which further contributes to an ineffective immune response against the tumor growth [152]. NO and reactive oxygen species (ROS) released from inflammatory cells may induce DNA damage, which further predisposes the cells to a high risk of malignant transformation [153]. In addition to these, tumor cells as well as inflammatory cells present in tumor stroma release cytokines. Some of the cytokines are classified as pro-inflammatory cytokines, for example, IFN-γ, IL-1 and 6, MIF, and TNF-α, while others are growth factors, for example, macrophage-colony stimulating factor (M-CSF) and TGF-β. Yet others serve to regulate the immune cells, for example, IL-4, 5, 10, or promote angiogenesis or vascular permeability, for example, vascular endothelial growth factor (VEGF). Besides these, there are a number of chemokines, for example, MCP-1 and macrophage-derived chemokine (MDC), which are secreted by tumor cells or immune cells. There is strong evidence supporting the direct role of these cytokines and chemokines in malignant progression [154–157]. Moreover, many cytokines and chemokines, including TNF-α, are inducible by hypoxia, which is a major physiological difference between normal and tumor tissue [158]. Furthermore, expression of all these genes is mainly regulated by the transcription factor NF-κB, which is constitutively active in most tumors and is induced by many carcinogens and tumor-promoting factors as well [159–163]. Thus, it is clear that ongoing inflammatory processes within the tumor microenvironment have the capacity to not only produce tumor-promoting factors but to also profoundly suppress the ability of the host antitumor immune responses. A greater understanding of how chronic inflammation leads to cancer, and how inflammatory mediators maintain the balance between pro-inflammatory and anti-inflammatory functions, will help us to design effective anticancer therapy to counter these effects. Also, the information may aid in the development of vaccines and other strategies to augment immune attacks on tumors.

### 2.2 A Paradox—Tumor Necrosis Factor or Tumor-Promoting Factor

TNF-α is one of the key chemical mediators implicated in inflammation-associated cancer. It is secreted by the tumor cells as well as the stromal cells in the tumor microenvironment. As mentioned previously, depending on the receptor type activated by TNF-α and the associated downstream signaling molecules, TNF-α has an optional capacity to induce apoptosis or promote cell survival. Hence, this section will provide some evidence for the paradoxical role of pro-inflammatory cytokine TNF-α as a tumor necrosis factor as well as a tumor-promoting factor.

***Tumor Necrosis Factor—Evidence*** TNF-α, ideal to its name, was found to induce necrosis in a number of animal models, human tumor xenografts, and experimental syngeneic models when administered at high doses [164, 165]. In human tumor xenograft models, it was seen that TNF-α was most effective in its tumoricidal action when it was administered locally at the tumor bearing site [164, 166]. Even greater tumoricidal effect was observed when along with local TNF-α, systemic IFN-γ was also administered. Following administration of both these cytokines, platelet adherence to tumor cells was observed. There was an upregulation of adhesion molecules and other cytokines in the tumor stroma, and macroscopic necrosis as well as apoptosis of tumor cells was witnessed [167].

The antitumor effects of TNF-α were found to be due to direct effects on tumor cells and/or indirect effects on cells of tumor stroma [168]. One of the proposed mechanisms for the tumoricidal action of TNF-α includes direct cytotoxic effect on vascular endothelium [169] and modulation of endothelial procoagulant properties [170]. Tumor-associated vasculature (TAV), which develops due to tumor angiogenesis, has a characteristic

phenotype that distinguishes it from normal vasculature. TAV has a noncontinuous endothelium, an enlarged basal membrane, and aberrant pericyte coverage [171]. Moreover, the vascular hierarchy of arterioles, capillaries, and venules is also rarely found in tumors, resulting in loosely associated pericytes [172]. All this results in a leaky vasculature in the tumor [173] through which blood cells such as lymphocytes and monocytes easily extravasate into the tumor. TNF-α further induces a rapid and prolonged extravasation of red cells into the tumors [174]. The strong extravastion of blood cells results in massive hemorrhagic necrosis of the tumor. This is consistent with an experimental finding in mice bearing methylcholanthrene-A (Meth-A) induced sarcoma tumor cells, where systemic administration of TNF-α consistently caused hemorrhagic necrosis of vascular tumors but not of avascular tumors [175]. Further, in combination with IFN-γ, TNF-α induced downregulation of αVβ3 integrin, which, in turn, resulted in detachment of endothelial cells from the basement membrane and subsequent apoptosis [176]. High-dose TNF-α also has direct cytotoxic effects on endothelial cells, through TNFR-1 signaling pathways, thereby leading to apoptosis. Loss of endothelial cells further enhances the hyperpermeability. Besides the destruction of tumor vasculature, an immediate effect of enhanced hyperpermeability subsequent to treatment with TNF-α lies in augmented accumulation and distribution of the drug in the tumor, which results in better exposure of the tumor cells to the cytostatic agents such as doxorubicin and melphalan [177–179]. This provides an explanation for better response rates seen in patients treated with a combination of TNF-α and chemotherapeutic drugs.

T-cell-mediated immunity has also been proposed to be an important constituent of TNF-α's antitumor response. This is substantiated by the observation that T-cell deficient mice exhibited diminished antitumor effects when treated with TNF-α [175]. Also, development of T-cell-mediated immunity was observed in tumor-bearing mice treated with TNF-α [180]. Besides, in experimental animal models, TNF-α has been implicated to be an important effector molecule required for the cytotoxic ($CD8^+$) T-cell and NK cell-mediated killing of tumor cells [181–184].

In conclusion, either of the above-mentioned mechanisms, alone or in combination, might be responsible for antitumor effects seen with local administration of high doses of TNF-α. All this evidence for antitumor effects of TNF-α provided a rationale for the use of recombinant TNF-α in the treatment of cancers. In 1998, recombinant TNF-α was approved by the European Agency for the Evaluation of Medicinal Products (EMEA) for the regional treatment of limb-threatening, locally advanced soft tissue sarcoma (STS) and metastatic melanomas in an isolated limb perfusion (ILP) setting, to avoid amputation of the limb [185, 186]. The success of TNF-α therapy in these patients is based on the antitumor effects observed due to high concentrations of TNF-α reached locally during ILP. TNF-α, given in combination with melphalan, showed 70–90% complete response rates [187]. It became the first agent to be registered for the treatment of cancer that improved drug penetration in tumors and selectively destroyed angiogenic vessels as well [188]. However, with systemic administration of TNF-α, severe toxicity of this cytokine became the major impediment for its use in other types of tumors. At the same time, there came very strong support for the tumor-promoting effects of chronically produced endogenous TNF-α, which has been discussed in detail in the following sections, and laid a strong foundation for the use of anti-TNF-α therapy in cancer patients.

***Tumor-Promoting Factor—Evidence*** TNF-α is expressed in a range of human tumors. Both mRNA and protein levels are found to be upregulated in tumor and/or stromal cells in human ovarian, breast, prostate, bladder, colorectal cancer, leukemias, and lymphomas [189–193]. In ovarian cancer, TNFR-1 was found on tumor and stromal cells while TNFR-2 was found in the leukocyte infiltrate, suggesting a potential for both paracrine and autocrine signaling [189]. In many reports, high circulating levels of TNF-α and its soluble receptors are associated with poor prognosis and resistance to therapy [193–197]. Some of the experimental evidence for procancer activity of tumors in animal models has been listed in Table 2 [168].

TNF-α cluster contains many polymorphisms including microsatellites and single nucleotide polymorphisms (SNPs). To date, 14 SNPs, mostly in the 5′ promoter regulatory region of the TNF-α gene, and 6 microsatellites have been characterized [198]. Eleven of these SNPs are located within this regulatory region and might therefore influence the transcription rate of this gene. Since an endogenous TNF-α level is being directly correlated with tumor promotion, people naturally producing high levels of TNF-α, due to SNP affecting TNF-α synthesis, have been found to exhibit higher susceptibility as well as severity of certain malignancies. For instance, in a study of 273 lymphoma patients, TNF-308 polymorphism has been linked to high plasma levels of the cytokine at disease presentation [199]. Furthermore, people possessing at least two TNF/LT high producer alleles were found to be at a high risk factor for first-line treatment failure, shorter progression-free survival, and over-all survival [199] as well as increased risk of developing myeloma [200]. TNF-α 308.2 polymorphism has been found to be associated

**TABLE 2 Experimental Evidence for pro-cancer actions of TNF-α**

| |
|---|
| TNF–/– mice are resistant to skin carcinogenesis |
| TNFR–1–/– mice are resistant to liver damage and liver cancer |
| Increasing TNF-α production by tumor cells increases metastatic potential |
| Pretreatment of mice with TNF-α enhances experimental metastasis |
| TNF-α is part of a cascade of cytokines that control tumor cell adhesion to endothelium in experimental metastasis models |
| Secretion of soluble TNFR-1 enhances tumor formation and persistence in vivo |
| TNF-α is directly transforming from 3T3 fibroblasts in vitro |
| TNF-α may drive lymphadenopathy and plasmacytoid tumors development in FasL–/– mice |
| TNF-α induces peritoneal adhesion and stromal formation in ascitic ovarian cancer xenografts |
| TNF-α is implicated in the action of nongenotoxic liver carcinogens |

Adapted by permission from Elsevier Ltd.: *Cytokine and Growth Factor Reviews* [168] © (2002).

**TABLE 3 Actions of TNF-α That May Mediate Tumor Growth, Invasion and Metastasis**

| |
|---|
| DNA damage via ROS |
| Inhibition of DNA repair via ROS |
| Autocrine growth and survival factor for malignant cells |
| Tissue remodeling via induction of MMPs |
| Control of leukocyte infiltration in tumors via modulation of chemokines and their receptors |
| Induction of angiogenic factors |
| Loss of androgen responsiveness |
| Resistance to cytototoxic drugs |

Adapted by permission from Elsevier Ltd.: *Cytokine and Growth Factor Reviews* [168] © (2002).

with advanced hepatic fibrosis and higher risk for hepatocellular carcinoma [201]. SNPs in the promoter region of TNF-α have been linked to susceptibility to prostate cancer [202]. TNF-α microsatellite polymorphisms have also been associated with increased susceptibility to human cancers [203, 204].

### 2.3 Mechanisms of TNF-α in Cancer Promotion/Progression

Pro-inflammatory cytokine TNF-α plays a direct role in survival, growth, proliferation, differentiation, and movement of not only the tumor cells but also of other cells found in the tumor microenvironment. Additionally, it may influence the interaction of tumor cells with stromal cells as well as their interaction with the extracellular matrix. Potential mechanisms of the procancer actions of TNF-α are summarized in Table 3 [168].

TNF-α may have direct effects on tumor cells. Experimentally, it has been shown that exposure to TNF-α in vitro is capable of transforming the carcinogen-treated fibroblasts, which can then form tumors in nude mice [205]. The molecular mechanisms for this effect have not been clearly understood but might involve damage to the DNA via formation of ROS [206]. Another plausible explanation is that via inhibition of cytochrome p450 or glutathione *S*-transferase isoenzymes, oxidative stress can be generated, which may induce DNA damage tolerance, inhibition of protein synthesis, and high rates of carcinogen-induced mutation [207]. At the same time, TNF-α may also have effects on tumor stroma and orchestrate changes in tissue architecture, typical to that seen in inflammatory process associated with wound healing. For instance, they have been found to induce angiogenic factors [e.g., VEGF, basic fibroblast growth factor (bFGF), and IL-8] [208–211], key angiogenic enzyme-thymidine phosphorylase [212], endothelial integrin [176] MMPs (e.g., MMP-9, involved in metastasis) [213], and various chemokines (e.g., MCP-1) [214]. It can also stimulate collagen production in mouse and human fibroblasts [215, 216] and stimulate their growth [217]. All these changes further promote tumor cell growth, proliferation, invasion, and metastasis. Moreover, through activation of transcription factor NF-κB, TNF-α can lead to increased receptor expression of other cytokines that induce proliferation, such as IL-3, granulocyte-macrophage colony-stimulating factor (GM-CSF) and IL-1β. Yet another report suggests that in renal cell cancer, von Hippel–Lindau (pVHL) tumor suppressor gene can lead to translational repression of TNF-α, thereby establishing a new genetic link between TNF-α and malignancy [218]. Thus, the mechanism of TNF-α in the promotion of cancer is still a hot area of research and is witnessing new developments every year.

### 2.4 TNF-α in Pathophysiology of Cachexia and Cancer

TNF-α has been implicated in the pathophysiology of cancer-associated cachexia and anorexia syndrome as well as many solid tumors and hematological malignancies. In this section, the role of TNF-α in the pathophysiology of cancer-associated cachexia, multiple myeloma (MM), myelodysplastic syndromes (MDS), and other malignancies has been elaborated and hence explains the use of anti-TNF-α in these conditions.

***Cancer-Associated Cachexia*** Patients with advanced cancer show progressive weight loss (up to 80%), which is mainly due to loss of fat and skeletal muscle. This complex metabolic state characterized by fatigue, weight loss, and loss of appetite (anorexia) is known as cachexia. Cachexia and associated anorexia is one of the significant factors in poor performance status and high mortality rates of cancer patients. It accounts for 20% of cancer deaths [219, 220].

TNF-α, along with other cytokines and growth-promoting factors released by tumor cells, is involved in the pathophysiology of cancer-associated cachexia and anorexia syndrome, and both peripheral and central mechanisms have been implicated [221]. The role of TNF-α and other cytokines in the progression of cachexia is summarized below:

1. Cancer patients undergo a characteristic acute phase response (APR)—a series of physiological and metabolic changes associated with elevated levels of acute phase proteins (APPs) such as fibrinogen, C-reactive protein (CRP), serum amyloid peptide (SAM), and the like. This response is similar to that seen in tissue injury, infection, or inflammation [222]. The presence of an APP clearly distinguishes tumor-related weight loss from simple starvation [223]. The magnitude of the APR and the proportionate rate of body-mass loss in patients with lung and gastrointestinal cancers has been associated with shortened survival in these patients [224]. It has been seen that the APR is not only activated but also modulated by cytokine, including TNF-α production [219].
2. Uncoupling proteins (UCPs)—1, 2, and 3—are a family of mitochondrial proteins that mediate proton leakage and decrease the coupling of respiration to ADP phosphorylation, resulting in the generation of heat instead of ATP. UCP expression levels have shown to be increased in cancer patients, especially UCP3 found on skeletal muscle [225, 226]. This might account for increased energy expenditure seen in cancer-associated cachexia. It has been seen in animal models that changes in UCP expression might be induced by cytokines. For instance, a single injection of TNF-α in rats resulted in a significant increase in UCP2 and UCP3 mRNA levels in skeletal muscle [227].
3. Cytokine-induced activation of NF-κB results in decreased muscle protein synthesis [90, 91]. Cytokines are also responsible for the reduction of MyoD protein, a transcription factor, which modulates signaling pathways involved in muscle development. TNF-α and IFN-γ act together to specifically inhibit the activation of mRNA for myosin heavy-chain synthesis [92].
4. Cancer patients with cachexia experience progressive erosion of skeletal muscle, resulting in asthenia. Ubiquitin-dependent proteolysis in skeletal muscle is considered to be the most important pathway for protein degradation in a catabolic disease such as cachexia [228–230]. TNF-α is thought to be involved in induction of proteolysis through the ubiqutin–proteosome proteolytic pathway [231, 232]. Muscle proteolysis yields amino acids and oligopeptides that are utilized for hepatic synthesis of APPs. The ubiquitin–proteosome system also indirectly modulates protein synthesis through degradation of IκB protein and subsequently activating NF-κB. Cytokine-mediated release of cortisol from the adrenal gland further propagates the activity of the ubiquitin–proteosome system [233].
5. Cancer-associated cachexia is also characterized by loss of body fat, which seems to arise from an increase in lipolysis and $\beta$ oxidation [234]. TNF-α has been associated with increased lipolysis, mainly through suppression of the lipoprotein cleaving enzyme LPL [235, 236]. Subsequent increase in very low-density lipoprotein (VLDL) synthesis and decreased LPL activity hinder triacylglycerol clearance and result in hypertriglyceridemia [237]. All these processes result in progressive weight loss [238].
6. Under normal conditions energy intake is controlled primarily in the hypothalamus by specific neuronal populations, which integrate peripheral blood-borne signals conveying information on energy and adiposity status [239]. However, consistent data seem to suggest that cancer anorexia is mediated by the inability of the hypothalamus to respond appropriately to peripheral signals, indicating an energy deficit. Cytokines, including TNF-α and IL-1, appear to mediate this "hypothalamic resistance," by hyperactivating anorexogenic neurons and suppressing prophagic neurons [240, 241]. There is an increased production of corticotrophin-releasing factor, a potent anorectic agent, and in concert with PGs, suppress the production of the orexigenic agent Neuropeptide Y (NPY) [242]. Cytokine-driven deregulation of the monoamine system contributes significantly to the modulation of these effects [220, 240].

Although TNF-α is definitely important for tissue catabolism, its importance in cancer-associated cachexia has not been conclusively established. This is because

the serum levels of these cytokines in humans have not been correlated with development of cachexia, and there have been some conflicting reports [243–245]. Nonetheless, many clinical trials are underway to treat the cancer-associated cachexia with anti-TNF-α therapy, and their positive outcomes will be able to provide irrefutable evidence regarding their importance in supportive care for cancer. The use of this therapy is also likely to yield improved response rates and prolonged survival of cancer patients.

***Myelodysplastic Syndromes (MDS)*** MDS is one of the most prevalent hematological malignancies, which include a heterogenous group of bone marrow disorders with a common denominator being the ineffective production of normal mature blood cells [246]. The patients diagnosed with MDS are at a high risk to develop acute myelogenous leukemia (AML) [247]. One of the paradoxical clinical settings seen in most patients with MDS is peripheral cytopenia with a cellular or hypercellular bone marrow [248]. However, the cells in bone marrow fall short of developing into terminally differentiated mature blood cells due to a failure of differentiation and increased cell death or apoptosis [249]. Although there is no direct evidence for intramedullary apoptosis, the foremost hypothesis today is that MDS results from accelerated apoptosis [250]. This increased apoptosis of hematopoietic cells seems to be due to increased signaling through the DR and mitochondrial pathway of caspase activation [251–253].

Proapoptotic cytokines like TNF-α are reported to be present at elevated levels in MDS bone marrow [254, 255] and can serve to trigger the apoptosis through the DR pathway as shown in Figure 3 [256–258]. This forms the rationale for the use of anti-TNF-α therapy in MDS patients with an ultimate goal to restore ineffective hematopoiesis caused by the accelerated apoptosis [259, 260]. However, a word of caution is always associated with antiapoptotic therapy in that it may also immortalize a malignant clone if apoptosis is a protective mechanism.

***Multiple Myeloma (MM)*** In MM, a tumor of postgerminal center bone marrow plasmablasts/plasma cells (normal mature B cells) forms and these malignant cells are almost exclusively confined to the bone marrow. The MM cells interact with the bone marrow microenvironment, and these interactions are mediated through direct cell–cell interactions via adhesion molecules and cytokines secreted by bone marrow and MM cells. MM is accompanied by the abnormal presence of paraprotein in serum and/or urine, decreased normal immunoglobulin (Ig) levels and lytic bone disease [261, 262]. Increased bone marrow microvessel density (MVD) and increased levels of proangiogenic molecules such as bFGF and VEGF, mediate the enhanced angiogenesis for tumor cell survival, and this provides the rationale for treatment with antiangiogenic drugs such as thalidomide.

Serum concentrations of TNF-α was found to be higher in patients with MM disease [263]. Especially, it was higher in patients with progressive MM disease compared to those with stable disease [264]. TNF-α plays a complex role in the pathogenesis of MM. TNF-α, along with other cytokines such as IL-6, insulin-like growth factor-1 (IGF-1), and VEGF, is found in the bone marrow microenvironment and mediates the growth and survival of MM cells [265]. Their growth-promoting effects are mediated through the RAS/RAF/MAPK kinase (MEK)/MAPK (mitogen-activated protein kinase) pathway while their survival effects or resistance to apoptosis is mediated through the phosphatidylinositol 3-kinase (PI3K)/Akt pathway [266]. The cytokine-mediated activation of antiapoptotic PI3K/Akt pathway is also responsible for resistance to conventional and novel therapies for MM, such as dexamethasone and thalidomide, which executes their pharmacological action by targeting the mechanisms leading to enhanced apoptosis of the tumor cells [262].

Through activation of NF-κB, TNF-α upregulates cell surface adhesion molecules such as intercellular adhesion molecule-1 (ICAM-1) and vascular cell adhesion molecule-1 (VCAM-1), on both MM cells and bone marrow stromal cells (BMSCs), which thereby increase the tumor cell BMSC binding [266]. The adhesion of MM cells to BMSCs further upregulates the transcription of cytokines (e.g., IL-6, VEGF) in BMSCs, which thereby promote growth, survival, and migration of tumor cells through autocrine and paracrine signaling [267–270]. It has also been shown that adhesion of MM cells to BMSCs confers protection against drug-induced apoptosis [271].

Thus, anti-TNF-α therapy, which will target not only the tumor cells but also its microenvironment, can prove to be an effective way to overcome the resistance to current therapies and, hence, serve as a promising treatment model to improve patient outcome in MM patients.

***Other Malignancies*** TNF-α has been demonstrated to be involved in the pathophysiology of many other hematologic malignancies, such as AML [272, 273], hairy cell leukemia (HCL) [274], chronic lymphocytic leukemia (CLL) [275, 276], and agnogenic myeloid metaplasia (AMM) [277–280]. Besides, drugs modulating TNF-α levels have proven effectiveness in a range of other tumors such as renal cell and prostate carcinoma, melanoma, glioma, Kaposi's sarcoma, and the like. Approaches inactivating or blocking TNF-α may

increase response rates and/or survival in patients with these malignancies.

## 3 ANTI-TNF-α THERAPY FOR CANCER

All the anti-TNF-α drugs that have been approved or are being currently investigated for use in cancer treatment are listed in Table 4, along with their actual or proposed structures and their mechanism of action.

### 3.1 Protein-Based Inhibitors

Protein-based inhibitors of TNF-α function include anti-TNF-α monoclonal antibodies (e.g., infliximab, CDP571, D2E7) and soluble TNF-α receptors (e.g., etanercept and lenercept). Infliximab and etanercept are the ones that are being intensively investigated for their potential in oncology. Currently these two drugs have an approval from the FDA for use in RA, Crohn's disease, ankylosing spondylitis, adult patients with chronic plaque arthritis, and ulcerative colitis. However, the potential for adverse effects associated with the use of these proteins-based inhibitors must be addressed carefully before they are approved for use in oncology. Patients taking infliximab have been reported to be at a high risk for opportunistic infections that may progress to serious diseases and even death. The most prevalent concern has been an increased incidence of re-exacerbation of latent TB [281, 282]. TB testing has been recommended for patients before and during the anti-TNF-α therapies. Studies with infliximab have also shown a potential risk for worsening congestive heart failure, and a dose >5 mg/kg is contraindicated in patients with moderate to severe heart failure. In addition, a potentially increased incidence of lymphoma has been observed in patients treated with the injectable TNF-α inhibitors, and this risk is thought to increase with disease severity [283–286]. Issues regarding their immunogenecity also need to be taken care of. Besides, there are problems regarding the use of antibodies as pharmaceuticals, as they need to be injected and are relatively expensive. Nonetheless, if they are effective enough, these therapies can be quite lucrative.

***Infliximab*** Infliximab (Remicade; Centocor/Schering-Plough/Tanabe Sieyaku) is a recombinant chimeric (human–mouse) IgG1κ monoclonal antibody comprising of a constant region from human and a variable region from mouse. It binds with high affinity (association constant $10^{10} M^{-1}$) and specificity, to both soluble and membrane-bound form of TNF-α and, hence, prevents it from binding to its receptors [287, 288]. In turn, all the biological responses mediated by TNF-α through its receptors are neutralized. Due to its ability to bind to transmembrane TNF-α, infliximab may cause lysis of the cells that produce TNF-α, such as macrophages and monocytes, and this depends on antibodies and complement [288].

Due to its ability to neutralize the actions of TNF-α and ameliorate the cachectic symptoms in tumor-bearing mice [289, 290], infliximab's potential is now being explored for treatment and prevention of cancer-associated cachexia in patients with advanced cancer disease. Preliminary studies evaluating the safety and pharmacokinetic profile of infliximab for the cancer-associated cachexia and anorexia have reported improved survival and prevention of any weight loss in cancer patients [291]. Currently, infliximab is in clinical trials for the treatment of cachectic symptoms and fatigue associated with cancer.

Anti-TNF-α therapy with infliximab has also been attempted in MDS [292–294]. Currently, infliximab is in phase II clinical trials for the treatment of MDS. In clinical studies, treatment with infliximab has shown a limited but significant activity and a favorable side effect profile. Patients taking infliximab showed improved hematopoietic response associated with decreased apoptosis and 50% reduction in abnormal metaphase [293]. A multilineage response was observed with sustained increase in neutrophil and platelet count and rise in hemoglobin concentration [293]. The most common side effects seen were thrombocytopenia (19%), neutropenia (27%), rash (5%), and facial swelling (5%) [294].

***Etanercept*** Etanercept (Enbrel; Amgen/Wyeth) is a soluble dimeric form of the TNFR-2, consisting of the extracelluar ligand-binding domain of the human TNFR-2 linked to the Fc portion of human IgG1. The Fc component of etanercept contains the CH2 domain, the CH3 domain, and hinge region but excludes the CH1 domain of IgG1. It is produced by recombinant DNA technology in a Chinese hamster ovary (CHO) mammalian cell expression system. It consists of 934 amino acids and has an apparent molecular weight of approximately 150 kDa [295]. Being dimeric, etanercept has a stronger affinity for TNF-α than the monomeric forms of the receptor. It acts as a competitive inhibitor and binds specifically to two molecules of free TNF-α in the circulation, consequently inhibiting the signaling cascades as well as biological effects of TNF-α mediated through its cell surface receptors. Unlike infliximab, etanercept forms less stable complexes with memTNF-α and monomeric TNF-α [296, 297]. As a result, cells expressing transmembrane TNF-α that bind etanercept are not lysed in vitro in the presence or absence of complement. Also, etanercept can bind to TNF-β as well

**TABLE 4 Anti-TNF-α Drugs in Cancer Treatment**

| Class | Drugs | Structure | Status (for Cancer Treatment) |
|---|---|---|---|
| Protein based inhibitors | | | In clinical trials<br>In clinical trials |
| | Infliximab | 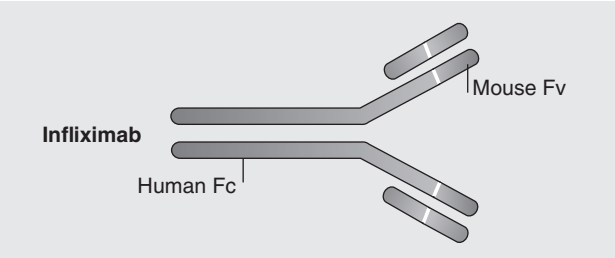 | |
| | Etanercept | 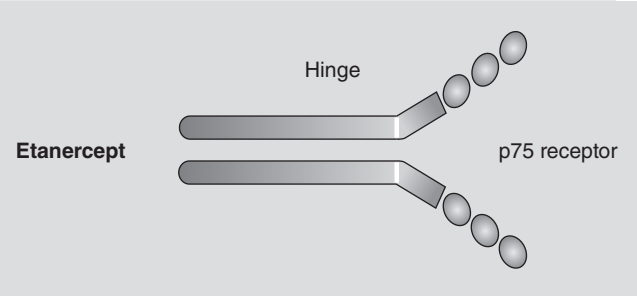 | |
| Immunomodulatory drugs | | | Approved<br>Approved |
| | Thalidomide | 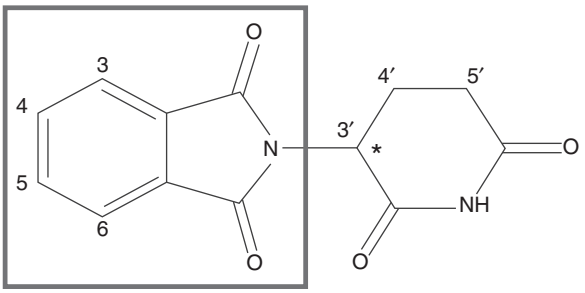 <br>Thalidomide | |
| | | **CC-5013** | |
| | Lenalidomide (CC-5013) | 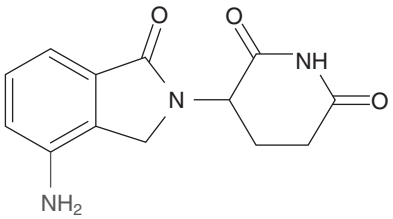 | |
| | | **CC-4047** | |
| | Pomalidomide (CC-4047) | 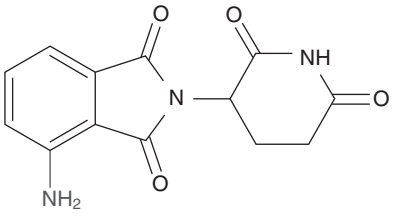 | |
| Selective cytokine inhibitory drugs | CC-8490 | N.A. | In clinical trials |
| | CC-1088 | N.A. | In clinical trials |
| p38 MAPK inhibitors | | | In clinical trials |
| | SCIO-469 | 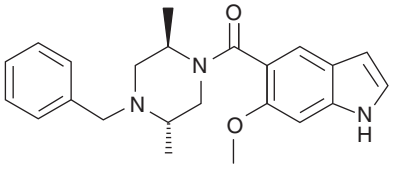 | |
| TACE inhibitors | Marimastat | | In clinical trials |

N.A. structures not available.
The boxed area shows the pthaloyl ring.

and prevents it from binding to cell surface TNFRs, rendering it biologically inactive [295].

Due to its ability to block TNF-α, efficacy and safety of etanercept in recurrent ovarian cancer [298], metastatic breast cancer [299], pancreatic adenocarcinoma (in combination with gemcitabine) [300], and various other solid tumors is being investigated. Etanercept has also been evaluated in the setting of hematological malignancies, such as MM [301], MDS [302], myelofibrosis [303], and refractory hematologic malignancies [304]. The use of etanercept as a single agent has shown improvement in the constitutional symptoms, but only in a limited number of patients. Infact, an increase in TNF-α plasma levels was seen in some of the MM patients during etanercept treatment, which was coupled with poor prognosis in these patients [301, 305]. However, etanercept administration was found to be safe and the proportion of patients who experienced toxicities was quite low. The most common adverse events seen were fatigue, fever, flulike syndrome, myalgia, injection site reactions, low titer antinuclear antibodies formation, minor infections, and abdominal cramps [304]. Moreover, in combination with other active biologic or cytotoxic therapeutic agents, such as thalidomide and azacitidine, it may improve the clinical outcome of patients with these diseases, and this approach is currently in clinical trials.

### 3.2 TNF-α Synthesis Inhibitors

#### *Immunomodulatory Drugs*

##### *Thalidomide*

HISTORY AND CURRENT STATUS Thalidomide (THALOMID; Celgene Corporation) was first introduced in 1956 as a drug to cure morning sickness in pregnant women. It was reported to have sedative, tranquilizing, and antiemetic effects. However, this culprit drug soon became a bane of human life when its teratogenic effects came light. Pregnant women taking this drug gave birth to babies with severe developmental deformities that included stunted limb development and other congenital defects of ear, heart, and internal organs [306]. Around 8000–12,000 "thalidomide babies" were born at that time, and this is one of the major drug-associated tragedies to date [307]. It was withdrawn in 1961, just to be reinvented in 1965, this time as an immunonodulatory and an anti-inflammatory drug for erythema nodosum leporosum (ENL)—a potential life-threatening inflammatory complication of lepromatous leprosy [308, 309]. Thalidomide was given FDA approval for the treatment of acute ENL in 1998, only after thorough investigations of its immunological effects [310]. Meanwhile, many studies demonstrated the efficacy of thalidomide in the treatment of autoimmune and inflammatory diseases. As the potent antiangiogenic effects of thalidomide were discovered [311], a huge number of studies began to investigate its role as an anticancer drug. Finally, in 2006, thalidomide in combination with dexamethasone got approved by the FDA for the treatment of patients with newly diagnosed MM. However, due to the potential risks of teratogenecity associated with thalidomide, it is approved for marketing only under a special restricted distribution program called "System for Thalidomide Education and Prescribing Safety (S.T.E.P.S.) initiated by the manufacturer of thalidomide (Celgene, Warren, NJ, USA). Under this program, only prescribers and pharmacists registered with the program are allowed to prescribe the product. The patients receiving the drug should also comply with the requirements of the S.T.E.P.S.

CLINICAL USES Thalidomide possesses immunomodulatory, anti-inflammatory and antiangiogenic properties. Currently, thalidomide has FDA approval for MM and ENL. However, a number of clinical interventions for other autoimmune and inflammatory disorders as well as some cancers are in progress, which will validate the use of thalidomide in a range of other diseases.

Thalidomide in combination with dexamethasone is indicated for the treatment of patients with newly diagnosed MM. Thalidomide treatment in MM patients leads to decreased MVD [312]. Treatment with thalidomide leads to a 25–33% response rate in patients with refractory MM [313] and significant response rates in patients at other stages of disease as well [314, 315]. Thalidomide given in combination with dexamethasone acts synergistically and gives even higher response rate of 60–70%, even when patients have been unresponsive to either agent alone [316, 317]. Thalidomide is capable of overcoming drug resistance of human MM cells to conventional therapy [318]. Efficacy of thalidomide, with or without a combination with steroids or other chemotherapeutic drugs, has been shown in patients with many other hematological and solid tumors, as summarized in Table 5 [319], and confirmatory studies for some of these cancers are in progress. In some patients with myelofibrosis or MDS the need for transfusions may reduce after thalidomide treatment.

Due to its ability to reduce TNF-α levels, thalidomide has been shown to relieve clinical symptoms of cachexia associated with cancer [320], TB [321], HIV [322], and ENL [310]. In HIV patients thalidomide can also relieve the cachexia symptoms, apthous ulcerations seen in advanced HIV infection, as well as Kaposi's sarcoma associated with HIV-AIDS [322–325].

**TABLE 5 Application of Thalidomide in Different Types of Cancers**

| Type of Cancer | References |
|---|---|
| Multiple myeloma | 15–17 |
| Myelodysplastic syndromes | 18 |
| Waldenstrom macroglobulinemia | 19 |
| Acute myeloid leukemia | 20, 21 |
| Myelofibrosis | 22 |
| Kaposi sarcoma | 23 |
| Renal cell carcinoma | 24, 25 |
| Prostate cancer | 26–28 |
| Glioblastoma | 29 |
| Metastatic melanoma | 30 |
| Breast cancer | 30, 31 |
| Cancers of ovary, head and neck | 32 |

Adapted by permission from Elsevier Ltd.: *American Journal of Medicine* [319] © (2004).

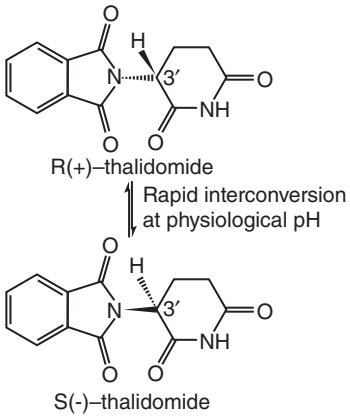


**Figure 6** Structure of thalidomide enantiomers. (Adapted by permission from Elsevier Ltd. [331].)

In leprosy patients, thalidomide is indicated for the acute treatment as well as maintenance therapy for prevention and suppression of the cutaneous manifestations of moderate to severe ENL. But it cannot be given as monotherapy for such ENL treatment in the presence of moderate to severe neuritis.

Studies have demonstrated thalidomide's efficacy in many autoimmune and inflammatory disorders such as RA [326], cutaneous lesions of SLE [327], Bahcet's disease [328], chronic graft-versus-host disease associated with allogeneic bone marrow transplantation [329], and also in congestive heat failure [330].

CHEMISTRY AND STRUCTURE–ACTIVITY RELATIONSHIPS

Chemically, thalidomide is [2-(2, 6-dioxo-piperidine-3-yl)-iso-indole-1, 3-dione] or α-(*N*-phthalimido) glutarimide and has the empirical formula $C_{13}H_{10}N_2O_4$. It has a gram molecular weight of 258.2. The glutarimide moiety contains a single asymmetric center, as marked by a star in its structure shown in Table 4 and, therefore, may exist in either of two optically active forms designated *S*-(–) or *R*-(+) in Figure 6 [331]. Thalidomide is a racemic mixture containing equal amounts of the *S*-(–) and *R*-(+) enantiomers and, therefore, has a net optical rotation of zero. Segregation of activities occurs between the two enantiomeric forms of thalidomide. While, the *S*-(–) enantiomer is associated with the teratogenic effects, and potently inhibits the release of TNF-α from monocytes [332], the *R*-(+) enantiomer seems to be responsible for sedation and acts through sleep receptors in the forebrain [333]. The two isoforms rapidly interconvert under physiological conditions [334] and, hence, the purification of a safer *R*-(+) enantiomer did not prove quite useful to get rid of the teratogenic effects of thalidomide. Thalidomide is an off-white to white, odorless, crystalline powder that is soluble at 25°C in dimethyl sulfoxide. Because of its lipophilicity, it is only sparingly soluble in water and ethanol.

Thalidomide intrinsically possesses pharmacophores with a wide range of activities within its small molecular skeleton. This includes sedative/hypnotic, antimalarial, and immunomodulatory activity. It is assumed that glutarimide moiety is related to its sedative/hypnotic and antimalarial effects while the N-substituted phthalimide structure is related to its immunomodulaory activity [335]. Thalidomide exhibits moderate antiangiogenic activity (ca. 26% inhibition at 100 μ*M*). The structure of some of the thalidomide analogs is shown in Figure 7 [336]. The major metabolite: 5′-OH-thalidomide (5-HT) shows moderate activity compared to thalidomide. Pomalidomide (CC-4047), the 4-amino analog of thalidomide, is less active than thalidomide (ca. 21% inhibition at 100 μ*M*). However, replacing the electron-donating amino group at the 4-position with an electron-withdrawing nitro group in 4-NT, resulted in a more potent antiangiogenic activity than that of thalidomide (ca. 32% inhibition at 100 μ*M*). The removal of the decarbonyl group of pomalidomide resulted in a newer, more potent analog of thalidomide, lenalidomide (CC-5013) (ca. 32% inhibition at 100 μ*M*). Replacing the electron-donating amino group at the 4-position of CC-5013 in 4-NHT, with an electron-withdrawing nitro group, shows just the opposite effect of CC-4047 and results in decreased antiangiogenic activity as compared to CC-5013. Substitution of bulky alkyl groups at the 2- and 6-position of the phenylphthalimide ring (structures 7–17) results in a much more potent antiangiogenic activity. The specific feature required for the enhanced antiangiogenic activity of these analogs is 2,

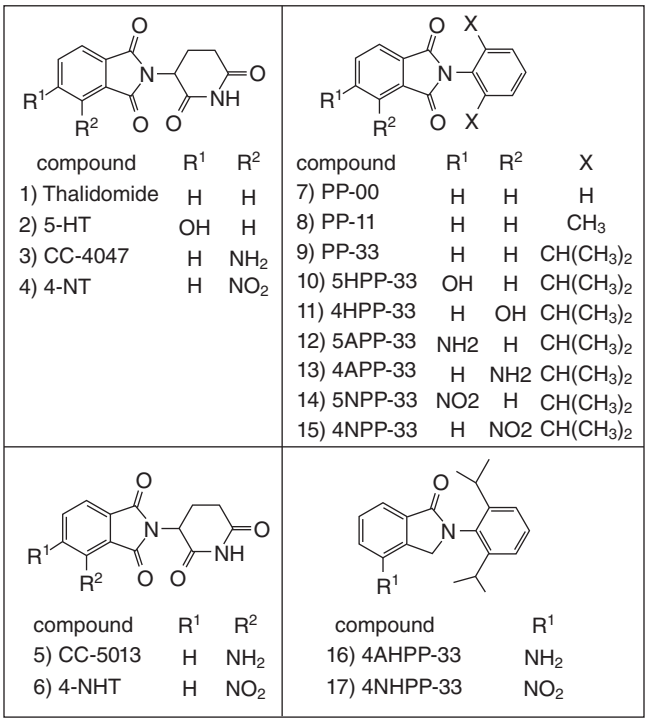


**Figure 7** Structure of thalidomide analogues. (Adapted by permission from Elsevier Ltd. [336].)

6-diisopropylphenylphthalimide structure, and the removal of this bulky alkyl group results in no or little activity [336].

MECHANISM OF ACTION OF THALIDOMIDE AND IMMUNOMODULATORY DRUGS (IMiDs) IN MM PATIENTS Available data from in vitro studies and clinical trials suggest that the immunomodulatory effects of thalidomide can vary substantially under different conditions. The mechanism of action of thalidomide and IMiDs is not fully understood. Also, in vivo, thalidomide gets metabolized rapidly to produce a number of metabolites. It is not very clear whether thalidomide or one of its metabolite contributes specifically to its immunomodulatory activity. Potential mechanism of actions of thalidomide and IMiDs in MM patients are depicted in Figure 8 [337].

INHIBITION OF PRO-INFLAMMATORY CYTOKINES In 1991, it was discovered that thalidomide inhibits the synthesis of TNF-α by activated monocytes, and it is mainly due to destabilization of TNF-α's mRNA. Thalidomide may also inhibit TNF-α released by other cellular sources, such as microglia and Langerhans cells, activated by inflammatory stimuli [338, 339]. Additionally, it may also inhibit the synthesis of other pro-inflammatory cytokines or growth factors such as IL-6 [340], IL-12 [341], IFN-γ [342], and IGF-1 [343]. Since TNF-α level is found to be elevated in many types of cancers, anti-TNF-α activity of thalidomide is likely to be beneficial in the treatment of cancer patients.

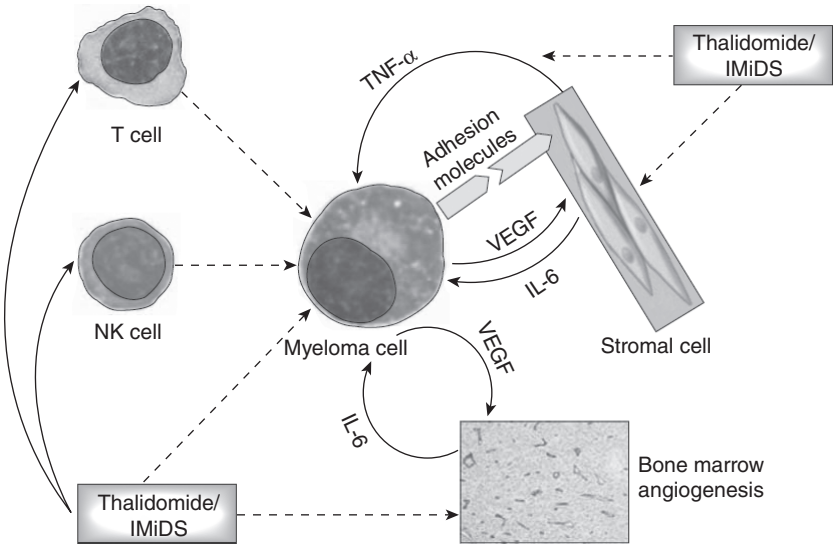


**Figure 8** Potential mechanism of actions of thalidomide and IMiDs in MM patients. Solid arrows indicate proliferation and survival of MM cells or proliferation of endothelial cells and vessel formation or activation of immune cells, respectively. Dashed arrows indicate inhibition of these effects. (Adapted by permission from Nature Publishing Group [337].)

INHIBITION OF NF-κB ACTIVATION It has been proposed that thalidomide's inhibitory activity can be explained on account of TNF-α induced activation of NF-κB, as seen in endothelial, epithelial, and myeloid cells, as well as, T lymphocytes [344]. As mentioned previously, NF-κB activation mediates many downstream biological effects of TNF-α. Thalidomide prevents the degradation of IκB-α by inhibiting IKK-β [345]. IMiDs are capable of triggering caspase-8-mediated apoptosis [346], and this IMiD-induced apoptosis in MM cells is associated with downregulation of NF-κB's DNA-binding activity, as well as the reduced expression of NF-κB-dependent apoptosis inhibitory proteins, such as the cIAP-2 and FADD-like interleukin-1 beta-converting enzyme (FLICE) inhibitory protein (c-FLIP) [346]. Thus, the immunomodulatory activity of thalidomide and IMiDs is mediated in part by inhibition of NF-κB [347]. In a nut shell, all these effects leading to increased susceptibility to apoptosis and reduction of growth-promoting factors result in inhibition of tumor growth.

INHIBITION OF ANGIOGENESIS Inhibition of angiogenesis (formation of new blood vessels) is an important process for growth and metastasis of solid tumors and forms the major basis for thalidomide's use in cancer. Thalidomide was first shown to inhibit bFGF-mediated angiogenesis by D'Amato et al. in 1994 [311]. Since then a number of studies have tried to shed light on the mechanism by which thalidomide exerts its antiangiogenic activity, though none of them has been conclusively established. In general, the angiogenic activity of thalidomide has been explained based on its ability to downregulate the proangiogenic genes, including VEGF and bFGF in the bone marrow endothelial cells. Also, in one study it has been found that thalidomide decreases expression of VEGF receptors in vitro through modulation of ceramide (a family of membrane lipid molecules) [348]. An inverse correlation has been established between TNF-α concentration and antiangiogenic effect, and it has been postulated that suppression of TNF-α is the mechanism by which thalidomide exerts its antiangiogenic activity. Thalidomide significantly decreases the levels of other proangiogenic factors as well, such as IL-6.

Cyclooxyegenase (COX)-2 is highly expressed in various human cancers and is needed for angiogenesis in a rat corneal model [349]. Thalidomide has been found to inhibit LPS-mediated induction of COX-2 by destabilizing its mRNA [350]. Subsequently, COX-2-mediated biosynthesis of PGE2 also gets inhibited, which further contributes to thalidomide's antiangiogenic activity [331, 351]. Thalidomide can also decrease the density of TNF-α-induced ICAM-1, VCAM-1, and E-selectin on the endothelial cells. Direct inhibitory effect on proliferation of endothelial cells for angiogenesis has also been reported.

In the early 1970s, using a rabbit cornea micropocket assay, it was proposed that classic congenital defects (abnormal limb development) that are caused by thalidomide treatment were due to inhibition of blood vessel growth in the developing fetal limb [311]. One of the models proposes that teratogenic effects of thalidomide are observed due to its ability to intercalate into guanine-rich promoter regions of DNA [343]. Thalidomide, or a breakdown product of thalidomide, specifically binds to these glycoprotein (GC) promoter sites, decreasing transcription efficiency of the associated genes. A cumulative decrease interferes with normal angiogenesis, which results in truncation of the limb. Intercalation into G-rich promoter regions of DNA may explain why certain thalidomide analogs are not teratogenic while retaining their TNF-α activity. On the other hand, the tissue specificity of thalidomide and its antiangiogenic effect against only certain neoplasias may be explained by the fact that various developing tissues and neoplasias depend on different angiogenesis or vasculogenesis pathways, only some of which are thalidomide-sensitive [343]. However, this link between inhibition of angiogenesis and thalidomide's teratogenicity remains largely unproven. There are other groups that have subsequently demonstrated that thalidomide inhibits mesenchymal proliferation in the limb bud [352] and induces embryonic oxidative stress [353]. Also, there is a strong evidence suggesting that, in vivo, thalidomide-induced apoptosis causes tissue degradation during early embryonic development resulting in limb truncations and microphthalmia [354, 355].

T-CELL CO-STIMULATORY ACTION Immune response begins when an antigen, that is, an infectious particle or abnormal cell, is engulfed by specialized phagocytic cells, such as macrophages or dendritic cells. These phagocytic cells are more appropriately termed as antigen-presenting cells (APCs). Following uptake of antigen, the APCs migrate to the lymph node and oligopeptides from the digested antigen are loaded onto MHC class II molecules present on the surface of APCs. These oligopeptides, in association with MHC-II are then recognized by specific TCRs present on the surface of TH cells [356]. A second co-stimulatory signal is also provided to the naïve T cells, which facilitates their activation and the subsequent generation of antigen-specific effector response. As shown in Figure 9A, this co-stimulatory interaction occurs between members of the B7 family of proteins on antigen-presenting cells and the CD28 molecules expressed on the surface of T cells [78]. In 1998 it was shown that thalidomide and IMiDs are able to co-stimulate T cells that have been partially

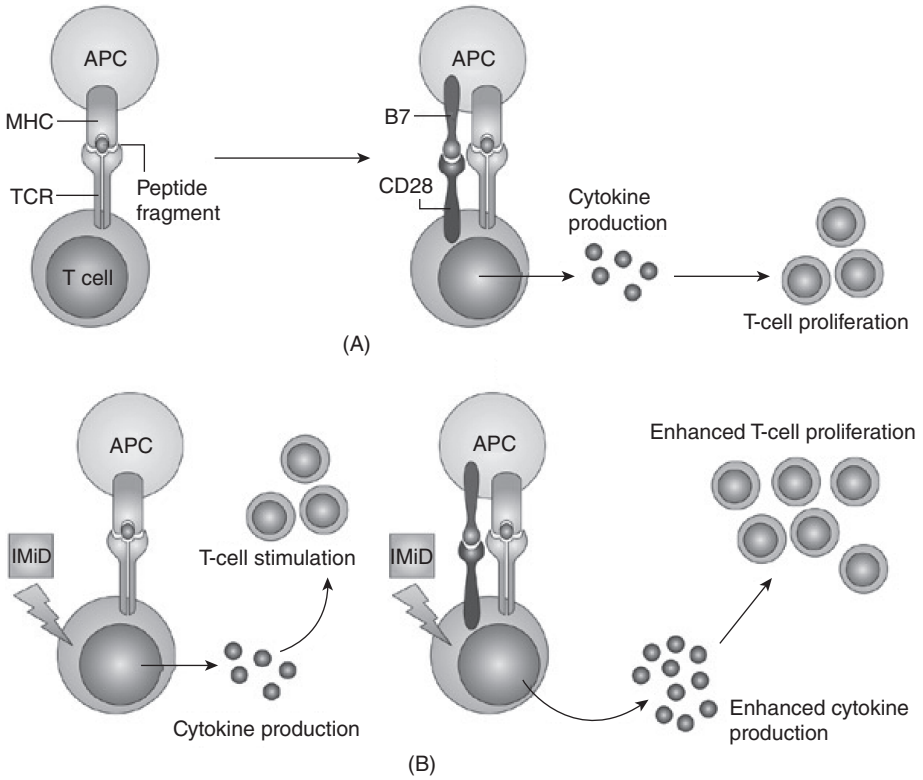


**Figure 9** (A) Activation of naïve t-cells requires co-stimulation. (B) Thalidomide and IMiDs overcome the requirement for co-stimulation. (Adapted by permission from Nature Publishing Group [78].)

activated by the TCR and hence overcome the requirement for co-stimulatory interaction between B7 and CD28 proteins [357]. Its co-stimulatory activity is depicted in Figure 9B. Thalidomide seems to enhance TCR-mediated signaling both in the absence and presence of secondary signals mediated through B7 and CD28 molecules. Due to its co-stimulatory activity, it is proposed that thalidomide might act as an immunological adjuvant to enhance an otherwise ineffective immune response against tumor antigens [358, 359]. Additionally, thalidomide stimulates cytotoxic T-cell proliferation and induces secretion of IFN-γ and IL-2 (cytokines involved in cytotoxic immune response). It also triggers TH-2 mediated cytokine production while concomitantly inhibiting TH-1 mediated cytokine production, the net effect being greater stimulation of cellular immune response against tumor cells. Thalidomide and IMiDs also modulate NK cell activity against tumor cells.

Interestingly, T-cell co-stimulatory activity of thalidomide is in contrast to the TNF-α synthesis inhibitor activity that leads to decreased TNF-α level. This differential activity might help to explain the clinically diverse effects of thalidomide, which include beneficial activity in some autoimmune diseases that are associated with increased TH-1 cells and some cancers that are associated with lack of tumor-specific TH-1 cells [78]. An increase in TNF-α level is seen in HIV-seropositive individuals [360]. Therefore, the immunomodulatory effects of thalidomide cannot be simply classified as TNF-α inhibitor or T-cell co-stimulator, and the clinical response is likely to depend on the patient's disease state and immunological status.

PHARMACOKINETICS Thalidomide is currently available only as a 50-, 100-, and 200-mg capsule formulation (Thalomid, Celgene, Summit, NJ, USA) for oral administration. The mean time to peak plasma concentrations ($T_{max}$) of thalidomide ranges from 2.9 to 5.7 h, indicating that thalidomide is slowly absorbed from the gastrointestinal tract. Poor solubility of thalidomide in aqueous media is likely to hinder its rate of absorption. In human blood plasma, the geometric mean plasma protein binding was 55 and 66%, respectively, for (+)-(*R*)- and (−)-(*S*)-thalidomide. Oral administration of thalidomide also leads to high concentrations of the drug in cerebro-

spinal fluid (about 50% of the plasma concentrations) [361]. In HIV-seropositive adult male subjects receiving thalidomide 100 mg/day, thalidomide is detectable in the semen as well.

Presently, the exact metabolic route and fate of thalidomide is not known in humans. Thalidomide undergoes both, nonenzymatic (spontaneous) and enzymatic metabolism mediated by cytochrome P450 enzymes in the liver, though the extent of enzymatic metabolism is much less than the nonenzymatic metabolism [362].

Thalidomide is rapidly degraded by spontaneous hydrolysis in an aqueous medium at physiological pH [363]. The four amide bonds of thalidomide are targets for hydrolytic cleavage and at pH 7.4, 12 hydrolysis products are formed by splitting these bonds [364]. Between pH 6 to pH 7, only the phthalimide ring undergoes cleavage, whereas at pH 7 and above, the glutarimide ring also undergoes hydrolysis [365]. The rates of inversion and hydrolysis increases with pH in the interval 7.0–7.5 [366]. All of the known hydrolytic products of thalidomide have been shown to be largely inactive. Some of the hydrolytic products with intact phthalimido and glutarimide rings may further undergo enzymatic metabolism by CYP40 enzymes, though this has still not been confirmed [362].

The enzymatic hydrolysis of thalidomide has been a subject of extensive investigation over several years. Oxidized products of thalidomide have been identified and are a focus of much attention. Glucuronide conjugates of thalidomide were also recently identified in mouse urine, but not in humans so far [367]. Thalidomide metabolism by cytochrome P450 enzymes is depicted in Figure 10 [331].

Thalidomide shows species-specific metabolism that may account for its species-specific action [368]. This is substantiated by the observation that thalidomide's teratogenecity is seen in rabbits and other species, but not in mice [361]. An active metabolite, 5′-OH-thalidomide, generated by cytochrome P450 2C19 (CYP2C19) isozyme-mediated oxidation of thalidomide seems to be responsible for its antiangiogenic effect [369, 370] . However, whether this is true in humans as well is still unclear. Furthermore, CYP2C19, responsible for the generation of 5′-OH-thalidomide, shows genetic polymorphism and eight variant alleles that predict the poor metabolizer phenotype (*CYP2C19*2* to *CYP2C19*8*) [371]. Thus, it has been hypothesized that interindividual variation in CYP2C19 activity may affect the activity and toxicity profile of thalidomide and may also lead to potential interactions with other drugs metabolized by the same enzyme [372].

The mean half-life of elimination for thalidomide ranges from approximately 5 to 7 h following a single dose and is not altered upon multiple dosing. Thalidomide itself has a renal clearance of 1.15 mL/min with less than 0.7% of the dose excreted in the urine as unchanged drug.

DOSING REGIMEN IN MM PATIENTS Thalidomide is administered in combination with dexamethasone in 28-day treatment cycles. The dose of thalidomide is 200 mg administered orally once daily with water, preferably at bedtime and at least 1 h after the evening meal. The dose of dexamethasone is 40 mg daily administered orally on days 1–4, 9–12, and 17–20 every 28 days.

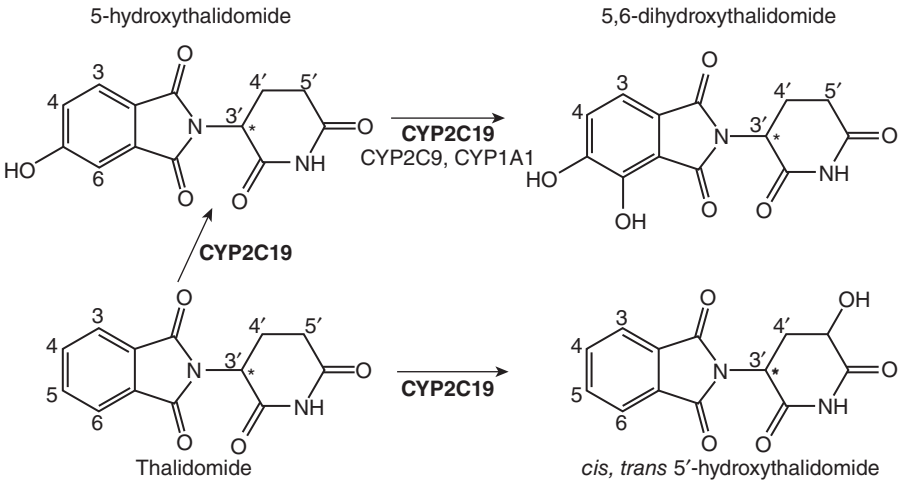


**Figure 10** Thalidomide metabolism by cytochrome P450 enzymes. (Adapted by permission from Elsevier Ltd. [331].)

**TABLE 6 Adverse Effects Seen with Thalidomide and IMiDs**

| Frequent Side Effects | Less Common Side Effects |
|---|---|
| **Thalidomide** | |
| Birth defects | Headache |
| Drowsiness and somnolence | Confusion |
| Skin rash | Malaise |
| Constipation | Hyperglycemia or hypoglycemia |
| Peripheral neuropathy | Tremor |
| Xerostomia | Pruritis |
| Orthostatis hypotension and dizziness | Peripheral edema |
| Bradycardia | Hepatitis (elevated serum transaminases) |
| | Stevens–Johnson syndrome |
| | Impotence |
| | Deep vein thrombosis |
| | Nausea |
| | Hypothyroidism |
| | Neutropenia |
| **Lenalidomide** | |
| Neutropenia | Skin rash |
| Thrombocytopenia | Light-headedness |
| | Fatigue |
| | Leg cramps |
| **Pomalidomide** | |
| Neutropenia | Skin rash |
| Thrombocytopenia | Neuropathy |
| Deep vein thrombosis | Constipation |
| | Nausea |
| | Edema |
| | Hypertension |

Adapted by permission from Nature Publishing Group: *Nature Clinical Practice Oncology* [337] © (2005).

ADVERSE EFFECTS Thalidomide dose of 200 mg or less is generally well tolerated in patients and fewer side effects are observed. However, all patients taking more than 400 mg/day experience thalidomide-related toxicity. The adverse effects seen with thalidomide and IMiDs are listed in Table 6 [337]. The most common side effects are somnolence and constipation, and most serious ones are venous thromboembolic events, such as deep venous thrombosis and pulmonary embolus, and peripheral neuropathy. Thalidomide is usually administered at bedtime to minimize somnolence and fatigue [319]. The incidence of thrombosis after treatment with thalidomide ranges between 2 and 23% but is higher among patients who receive chemotherapy [373–376]. The mechanism of thrombosis is still unknown, though there is speculation that thalidomide might alter the interaction between cancer cells and coagulation factors, induce prothrombotic events, and activate platelets and vascular endothelial cells [377, 378].

Teratogenecity is one of the most serious side effects seen with even a single dose of thalidomide taken by pregnant women. Major human fetal abnormalities that have been documented with thalidomide administration include amelia (absence of limbs), phocomelia (short limbs), hypoplasticity of the bones, absence of bones, external ear abnormalities (including anotia, micro pinna, small or absent external auditory canals), facial palsy, eye abnormalities (anophthalmos, microphthalmos), and congenital heart defects. Alimentary tract, urinary tract, and genital malformations have also been observed. Death of the thalidomide baby occurs at or shortly after birth. To avoid any teratogenic effects of thalidomide, the use of thalidomide is absolutely contraindicated in pregnant females. Since thalidomide can be found in the semen of males, unprotected sex with male patients taking thalidomide is also forbidden. Breast feeding and blood donation are prohibited for patients taking thalidomide.

Thalidomide treatment has led to an unexpected rise of HIV viral load in HIV-seropositive patients. Plasma levels of HIV mRNA were found to be elevated [379]. Though the clinical significance of this finding is not very clear, it indeed calls for caution.

Management of thalidomide-related adverse effects is based on four principles: (i) dose adjustment or temporary interruption of the treatment; (ii) avoidance of agents that enhance the thalidomide-associated adverse effects; (iii) administration of agents or measures that relieve the adverse effects; and (iv) regular clinical and laboratory monitoring for prevention of toxicity. Frequent laboratory tests recommended before the start of and during thalidomide treatment include tests for pregnancy, to avoid any teratogenic effects, neutropenia, and HIV viral load.

DRUG INTERACTIONS Thalidomide has been reported to enhance the sedative activity of barbiturates, alcohol, chlorpromazine, and reserpine. Also, the medications known to be associated with peripheral neuropathy should be used with caution in patients receiving thalidomide.

CONTRAINDICATIONS Thalidomide is contraindicated in women of childbearing potential unless alternative therapies are considered inappropriate. The patient should accept the condition that she cannot become pregnant while on thalidomide therapy. Because thalidomide is present in the semen of patients receiving the drug, males receiving thalidomide must never have unprotected sex with women of childbearing potential, even if, he has undergone a successful vasectomy. Thalidomide is also contraindicated in patients who have demonstrated hypersensitivity to the drug and its components.

*Thalidomide Analogs* Due to the severe toxicity and risks associated with thalidomide, there was a need to synthesize compounds with increased immunological and anticancer properties, but devoid of its toxic effects. Thus, two major classes of compounds have been synthesized using the first-generation IMiD thalidomide as the lead compound [78]. These include class I or immunomodulatory drugs (IMiDs) and class II or selective cytokine inhibitory drug (SelCIDs) [380]. While the IMiDs® were more potent TNF-α inhibitors and modulate the immune system by different mechanisms, SelCIDs selectively inhibit phosphodiestrase 4 (PDE4) and indirectly modulate TNF-α levels.

IMMUNOMODULATORY DRUGS (IMIDS) The second-generation IMiDs were developed in the mid-1990s with a rationale to improve the inhibition of TNF-α [381, 382]. Accordingly, a series of 4-amino-phthaloyl thalidomide analogs were developed [383–385], which were found to be 50, 000 times more potent at inhibiting TNF-α than the parent compound in vitro. Out of the many second-generation IMiDs, lenalidomide (CC-5013, Revimid, Celgene, Warren, NJ, USA) and pomalidomide (CC-4047, Actimid, Celgene, Warren, NJ, USA) have taken a lead. While lenalidomide already has FDA approval, pomalidomide is in phase II clinical trials for solid tumors. Third-generation IMiDs are also being developed now and are in preclinical testing. The focus for the development of these newer generation compounds is to enhance their antiangiogenic and immunomodulatory activities [336, 386] and, thus, could potentially have greater anticancer activity.

Like thalidomide, the mechanism of action of IMiDs is unknowm and appears to be mediated through the similar molecular targets as mentioned for thalidomide. However, IMiDs have been reported to be more potent and less toxic analogs of thalidomide [78]. They do not have any effect on PDE4 levels. Most significantly, they are far more potent than thalidomide at co-stimulating T cells that have been partially activated via the TCR [387]. The potential of IMiDs in enhancing protective antitumor immunity might prove to be an effective adjuvant in anticancer vaccination regimen. This is supported by preclinical in vivo data since pomalidomide was demonstrated to enhance a partially effective cancer vaccine and enable the generation of a long-term protective antitumor responses [388].

*Lenalidomide* Lenalidomide (REVLIMID; Celgene Corporation) is a second-generation IMiD. It is found to be more potent than the parent compound thalidomide [386, 389]. Its chemical name is 3-(4-amino-1-oxo 1,3-dihydro-2*H*-isoindol-2-yl) piperidine-2,6-dione. It has a mechanism of anticancer action that is similar to thalidomide and other IMiDs. In 2005, it was approved by the FDA for the treatment of patients with transfusion-dependent anemia due to low- or intermediate-risk 5q MDS. 5q MDS is a type of MDS where a part of the longer arm (q) of chromosome 5 has been deleted. Patients with this type of MDS may have low red blood cell counts that require treatment with blood transfusions. In 2006, it also got approved, in combination with dexamethasone, for treatment of multiple myeloma patients who have received at least one prior therapy. Clinical trials for many other types of cancers and in combination with other anticancer drugs are in progress. Being a thalidomide analog, lenalidomide also carries the risk of teratogenicity and is accordingly contraindicated in female patients of childbearing potential. Like thalidomide, lenalidomide is also available only under a special restricted distribution program called REVASSIST. It has been shown in human in vitro metabolism studies and nonclinical studies that lenalidomide is neither metabolized by nor inhibits or induces the cytochrome P450 pathway, suggesting that lenalidomide is not likely to be the cause or subject to P450-based metabolic drug interactions in humans. The major adverse effects associated with lenalidomide treatment are myelosuppression, leading to neutropenia and thrombocytopenia, and deep vein thrombosis and pulmonary embolism.

SELECTIVE CYTOKINE INHIBITORY DRUGS (SELCIDS) SelCIDs selectively inhibit phosphodiesterase 4 (PDE4) and indirectly inhibit TNF-α. Normally, cyclic adenosine monophosphate (cAMP) and cyclic guanosine monophosphate (cGMP) are converted to AMP and GMP, respectively, through the action of PDE4. Inhibition of PDE4 allows cAMP/cGMP to be elevated within the cells and this result in activation of protein kinase A (PKA). Activation of PKA, through a phosphorylation cascade, prevents activation of transcription factor NF-κB and in turn prevents transcription of the gene-encoding TNF-α, thereby resulting in a decrease in TNF-α synthesis [382, 390].

PDE4 isoenzymes play a particularly important role in inflammatory and immunomodulatory cells and are the predominant PDE in inflammatory cells, including mast cells, neutrophils, basophils, eosinophils, T lymphocytes, macrophages, and structural cells such as sensory nerves and epithelial cells [391]. Accordingly, the most promising role of SelCIDs compounds appears to be as anti-inflammatory agents, and many PDE4-selective inhibitors have shown efficacy in a number of inflammatory models, including asthma and RA [392–395]. However, a number of these compounds have exhibited antiangiogenic activity independent of their immunomodulatory activity [386, 396]. Moreover, SelCIDs have exhibited a potent activity against a range of solid tumors [397]. Their antisolid tumor activity is shown to

be caspase-3-dependent with an associated downregulation of antiapoptotic bcl-2 protein [397]. In yet another study, their antimyeloma activity is thought to be mediated through two mechanisms, that is, inhibition of VEGF and IL-6 production by interacting myeloma and endothelium and induction of myeloma cell apoptosis [398].

A number of SelCIDs have been developed by Celgene Corporation and investigated for different types of cancers. One of the compounds, CC-8490, a benzopyranone, was found to inhibit the growth of glioma cells and induce glioma cell death in preclinical studies [399]. Recently, it has completed phase I clinical trials for progressive or recurrent malignant gliomas where the maximum dose of CC-8490 that can be given to patients and the safety and tolerability of this drug were assessed. Another compound, CC-1088, is also being studied as a potential treatment for MDS and leukemia and is in phaseI/II clinical trials. Thus, SelCIDs are yet another type of thalidomide analogs with a novel mechanism of action for anticancer therapy.

***p38 MAP Kinase Inhibitors*** Kinases are the phosphorylating enzymes that carry out ATP-dependent phosphorylation of serine, threonine, or tyrosine residues on regulatory and structural proteins and hence modulate various signal transduction cascades involving these proteins. The p38 MAPK is a serine/threonine kinase having four isoforms (α, β, γ, δ) and plays an important role in the cytokine production as well as in cytokine action. The production of various cytokines is regulated by p38 MAPK at both trancriptional and translational level. mRNAs for the inducible cytokines are short lived and, as mentioned previously, they contain ARE in their 3′ UTR, which is one of the major point of posttranscriptional regulation. Under normal conditions these AREs are occupied by ARE-BPs, which prevents their translation. In response to various stimuli such as LPS or environmental stress such as osmotic shock, UV light or certain growth factors, p38 MAPK gets activated through a series of phosphorylation events, as mentioned in Figure 11 [400]. As shown in Figure 12, once activated, p38 MAPK, via MAPK-activated protein kinase 2 (MAPKAP K2) phosphorylates the androgen binding proteins (ABPs), thereby releasing them from ARE on mRNAs [400]. The free mRNAs can now undergo translation. Furthermore, p38 MAPK, along with other MAPKs and Janus kinase (JAK)/signal transducers and activators of transcription (STATs) pathways, also phosphorylates a variety of transcription factors, such as NF-κB, which is responsible for transcriptional expression of pro-inflammatory cytokines such as TNF-α [400]. In addition to that, as shown in Figure 12, p38 MAPK pathway also gets activated in response to TNF-α binding to TNFR-2 and thereby mediates various downstream biological actions of TNF-α, which promote cancer cell survival.

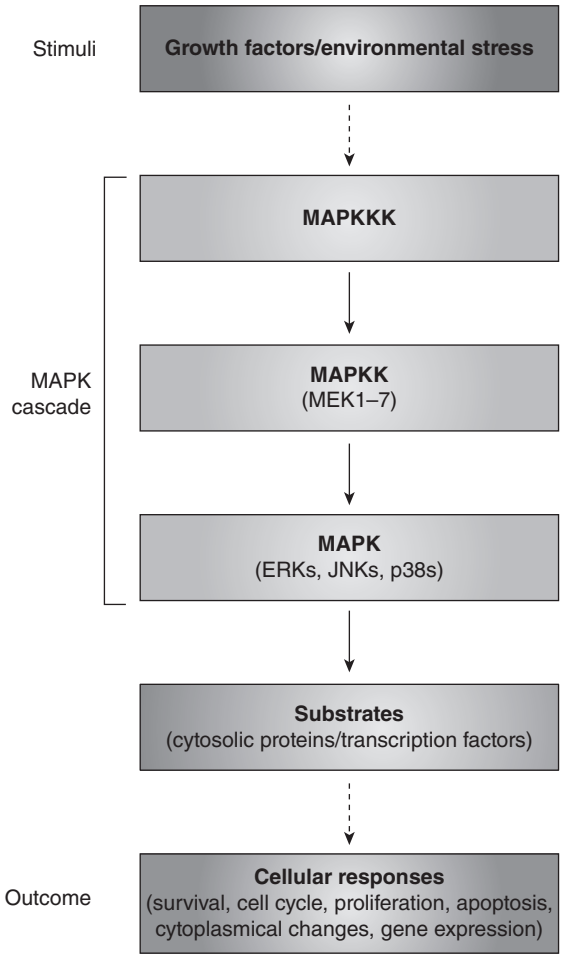


**Figure 11** MAPK pathway. Mitogen-activated protein kinases (MAPKs), which integrate and process various extracellular signals. MAPK kinase (MAPKK or MEK), MAPKK kinase (MAPKKK), extracellular signal-regulated kinase (ERK), c-Jun N-terminal kinase (JNK). (Adapted by permission from Nature Publishing Group [400].)

p38 MAPK inhibitors block the action of p38 MAPK and, in turn, inhibit the action and biosynthesis of various pro-inflammatory cytokines such as IL-1, IL-6, IL-8, TNF-α, TGF-β, and IFN-γ. This forms the rationale for their use in anticancer therapy. These inhibitors are selective, competitive inhibitor of p38 MAPK and compete for the binding of ATP within a central binding pocket in the kinase enzyme. This central binding pocket is formed between α and β domains and is structurally conserved in all catalytically active kinases to

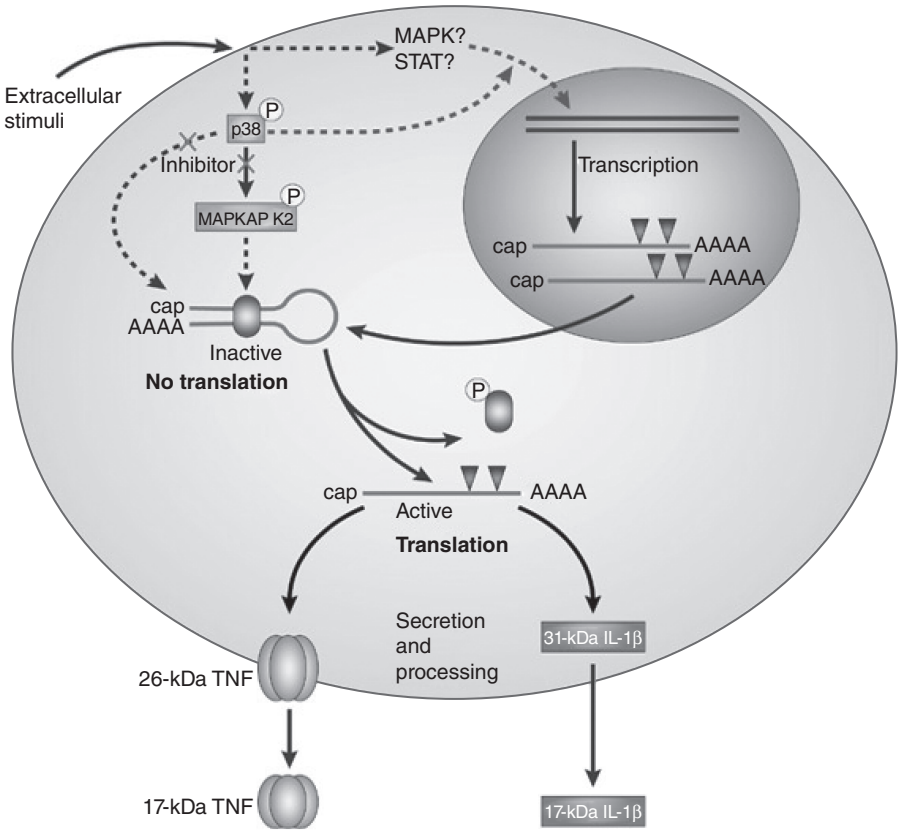


**Figure 12** Inhibition of TNF-α mRNA translation and transcription. Triangles indicate AREs, purple circles indicate ARE-BPs, red crosses indicate point of action of p38 MAPK inhibitor. (Adapted by permission from Nature Publishing Group [400].)

facilitate the transfer of γ phosphate from ATP to the substrate [401].

Various p38 MAPK inhibitors have been developed, such as AMG 548, BIRB 796, VX 702, VX 745, SCIO 469, SCIO 323, 681323, and RWJ67657, and many of them are being investigated for effectiveness in inflammatory, autoimmune, and malignant conditions [402]. SCIO-469, belonging to the Scios group of inhibitors, is of particular interest since it is at the most advanced stage in clinical testing and has shown efficacy in cancer. It was synthesized by Scios but now has been purchased by Johnson & Johnson, Inc. Its structure has not been released, but it is assumed that it is a piperazine indole oxalic acid analog as shown in Table 4 [403, 404]. Currently, it is in phase II clinical testing for RA [405], MM [406], and MDS [407]. In MM patients, SCIO-469 has been found to decrease proliferation of MM cells and block the IL-6 and VEGF secretion from BMSCs [408]. It also prevents TNF-α-induced adhesion of MM cells through ICAM-1- and VCAM-1-independent mechanisms [409]. Given in combination with a proteosome inhibitor, bortezomib, SCIO-469 has been shown to augment the cytotoxicity of bortezomib, in part due to downregulation of heat shock protein 27 (Hsp27), a molecule whose overexpression has been associated with resistance to dexamethasone [410]. It also enhances bortezomib-induced MM cell apoptosis by upregulating proapoptotic p53 and downregulating antiapoptotic Bcl-$X_L$ and Mcl-1. In murine models of MM, SCIO-469 has been shown to prevent the development of osteolytic bone disease, reduce tumor burden, and increase survival [411]. Besides, there is evidence that p38 MAPK is constitutively activated or phosphorylated in MDS bone marrows. Inhibition of p38 MAPK by SCIO-469 has been reported to decrease apoptosis in MDS $CD34^+$ progenitors and lead to dose-dependent increase in hematopoiesis [412]. Use of p38 MAPK inhibitors is advantageous since the p38 MAPK pathway is used by many cytokines that are implicated in the pathophysiology of MDS and MM. Thus, targeting this pathway may have wider clinical benefits than treatments targeting individual cytokines alone, such as anti-TNF-α

monoclonal antibodies. Although in developmental stages, p38 MAPK inhibition appears to be a promising approach for effective anticancer therapy.

***TACE Inhibitors*** TNF-α-converting enzyme (TACE/ADAM17/CD156q) is the enzyme responsible for cleavage of membrane-bound propeptide (pro-TNF-α of 26 kDa) to form a secreted, soluble, 51 kDa, homotrimeric cytokine (sTNF). It is also required for shedding of the extracellular domains of TNFR-2 from the surface of the cells [413]. TACE is a zinc-dependent metalloprotease enzyme and is a member of the a distintegrin and metalloprotease (ADAM) family. Along with many other MMPs, the ADAM family belongs to the metzincin superfamily [414]. MMPs are a group of 24 enzymes that degrade the extracellular matrix (ECM) and basement membrane and aid in tumor invasion and metastasis [415, 416].

TACE is a multidomain, type I transmembrane protein and its expression is largely constitutive [417]. Besides shedding of TNF-α and TNFR-2 extracellular domains, it is required for generating soluble forms of TGF-α [413], L-selectin [413], amyloid precursor protein [418], and many other proteins [419]. The domain structure of TACE, as show in Figure 13A, has a zinc-dependent catalytic domain and a disintegrin-cysteine-rich (Dis Cys-rich) sequence [417]. The prodomain contains a cysteine that interacts with the active-site zinc and gets displaced for the generation of an active enzyme [417]. The role of the disintegrin domain is still unknown. Only processed TACE appears on the cell surface [417, 420]. In the absence of cell activators, the surface pool of the enzyme is long-lived. However, prolonged exposure of cells to cell activators, such as phorbol esters, causes the internalization and degradation of TACE [420]. Cleavage action of TACE has been depicted in Figure 13B [417].

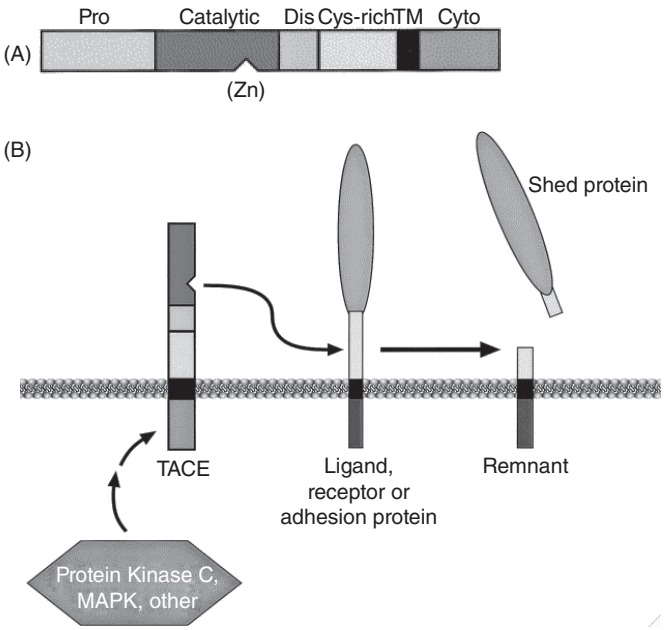


**Figure 13** (A) Domain structure of TACE. The notch signifies the active site. TM; transmembrane, cyto; cytoplasmic. (B) Cleavage action by TACE. (Adapted by permission from Elsevier Ltd. [417].)

A number of peptide-hydroxamates, containing a hydroxamic acid group that chelates zinc, are found to be potent inhibitors of TACE. However, these inhibitors are not specific to TACE inhibition and can inhibit other MMPs as well. One such synthetic broad-spectrum MMP inhibitor (MMPI), which hit the market, is Marimastat (BB-2516) developed by British Biotech Pharmaceuticals Ltd (Oxford, United Kingdom). It was the first MMPI to have entered the clinical trials in the field of oncology and since then has been widely tested in many cancer models [421–428]. The principal effect seen with Marimastat was to retard tumor growth and metastatic spread by inhibiting the degradation of the extracellular matrix by MMPs. However, Marimastat failed drastically in clinical trials for glioblastoma, pancreatic, gastric, breast, and ovarian cancers [429] either due to major side effects and/or no clinical improvement [430]. The main toxicity observed with Marimastat in clinical trials was dose-dependent musculoskeletal pain that could lead to contractures, particularly of the hand, if the treatment was continued [431].

It was initially thought that soluble TNF-α generated by cleavage of membrane-bound TNF-α is the only active form of this cytokine capable of eliciting biological responses. However, it later became clear that membrane-bound TNF-α could also assemble as homotrimeric ligand and produce biological effects through both TNFR-1 and TNFR-2, via direct cell-to-cell contact. In this current scenario, TACE inhibitors, which inhibit the production of sTNF-α by inhibiting TACE, and have no effect on memTNF-α action seems to be of limited therapeutic benefits. Moreover, all the TACE inhibitors that have been synthesized so far not only inhibit TACE but also other MMPs. MMPIs are likely to prove beneficial in cancer therapy but no major success has been achieved so far [432–434]. This is because of incomplete understanding of the role of MMPs in cancer. Also, the complex interactions between MMPIs, natural tissue inhibitors of MMPs, and other key activating factors have not been unraveled. MMPs such as ADAMs have been shown to have other important cellular functions, and inhibiting them might lead to unforeseen side effects with MMPIs. Further complication arises because 20–30 different MMP subtypes have now been identified and the function of one is

likely to be compensated by others [429]. Lastly, MMPIs are cytostatic drugs and were found to be ineffective against advanced or late-stage tumors, which were present in most patients. In conclusion, success of TACE inhibitors is quite skeptical at this point, but further understanding into it roles and mechanisms is likely to give some useful inights for its use in oncology.

### 3.3 Miscellaneous Approach

***NF-κB Inhibitors*** As mentioned in Section 1, the target genes for the NF-κB group of transcription factors include a huge number of inflammation-related genes that get activated in response to cytokines such as TNF-α and IL-1 [60]. Genes dependent on the activation of NF-κB include those encoding TNF-α, IL-1β, IL-6, IL-8, and IFN-γ, the adhesion molecules E-selectin, ICAM-1, and VCAM-1, the enzymes nitric oxide synthase and COX-2, MHC proteins, and viral proteins. NF-κB also induces the transcription of antiapoptotic proteins called the inhibitors of apoptosis [435]. Besides various cytokines, a number of chemotherapeutic drugs, such as irinotecan and doxorubicin, can also activate NF-κB. Due to its ability to activate the transcription of prosurvival genes, activation of NF-κB in tumors leads to a resistant phenotype of tumors to both TNF-α-mediated cytotoxicity and chemotherapies. Thus, inhibition of TNF-α-induced activation of NF-κB offers an attractive adjunct therapy for treating many types of cancers. NF-κB can be inhibited by many different approaches [347]. This includes:

1. Small molecules that can interfere with the binding of NF-κB to DNA and inhibit NF-κB-induced activation of transcription of several target genes.
2. Small molecules that can interfere with dimerization of NF-κB proteins.
3. Inhibitors of IκB degradation and NF-κB nuclear translocation [436]. This includes inhibitors of E3 ubiquitin ligases and the E2 ubiquitin-conjugating enzymes responsible for phosphorylation-dependent poly-ubiquitination of the IκBs and inhibitors of proteosomes responsible for degradation of phosphorylated, polyubiqutinated IκB. One such inhibitor, bortezomib (Velcade, Millenium), is a 26S proteosome inhibitor and prevents IκB degradation. It has been approved by the FDA for treatment in multiple myeloma and mantle cell lymphoma [437, 438].
4. Inhibitors of IKK activity. Several NSAIDs, such as aspirin and salicylates [439–441], sulindac and its analogs [442], and sulphasalazine and its metabolites [443] are capable of interfering with IKK-NF-κB system. Thalidomide and its analogs, IMiDs, are also capable of inhibiting IKK activity and hence interfering with NF-κB activation [345]. Besides these, development of several selective inhibitors of IKK is in progress and is likely to show exciting results in the coming years [347].

Besides synthetic compounds, a number of plant derivatives such as sesquiterpene lactones, capsaicin, and resiniferatoxin have also been shown to interfere with NF-κB activation [444, 445].

## 4 CONCLUSIONS AND FUTURE PROSPECTS

At present, the advent of multifaceted compounds stemming from a variety of painstaking mechanistic approaches clearly demonstrate that anti-TNF therapy for the treatment of cancer is definitely on the horizon. Current understanding of cancer biology makes it worthwhile to picture tumor growth with a much bigger perspective. Culprits involved in tumor progression are not only the aberrantly growing tumor cells but also many other normally growing cells of the body, which contribute in forming an environment conducive to tumor growth. Thus, tumor growth depends on intercommunication between various cell types present in the tumor stroma, and at biochemical level this interdependence is manifested by the exchange of various types of mitogenic and trophic factors. These factors are secreted not only by the tumor cells but also by inflammatory cells and other cells present in the tumor stroma. One such factor present at elevated levels in tumor stroma is TNF-α, which is not only the key mediator of inflammation but many strong evidences now implicate its role in tumor promotion as well.

TNF-α is a unique molecule with various diverse actions on its target cells. Depending on the receptor type activated in response to its secretion, as well as site and duration of secretion, TNF-α can lead to cell proliferation in some instances and cell death in others. Likewise, its role in cancer is also paradoxical, inducing destruction of blood vessels and cell-mediated killing of certain tumors, as well as acting as tumor promoter. However, strong evidence now conclusively implicates the role of TNF-α in tumor progression, besides being a key mediator of inflammation. Due to the presence of a huge number of inflammatory cells in the tumor stroma, inflammation and tumor progression are viewed as two intertwined processes, further supporting the role of pro-inflammatory cytokine, TNF-α, in tumor progression. Thus, TNF-α serves as molecular bridge between inflammation and cancer, and targeting TNF-α with

anti-TNF-α therapy, with the aim of inhibiting tumor growth, would indeed be a lucrative approach.

Direct involvement of TNF-α has been shown in the pathophysiology of cancer-associated cachexia and anorexia. Thus, anti-TNF-α therapy with protein-based inhibitors such as infliximab and etanercept, which selectively block TNF-α action, is likely to prove beneficial in the supportive care for cancer by improving appetite, slowing down weight loss, and increasing general well being of cancer patients. However, interactions of tumor cells with other cells types in the tumor stroma are increasingly heterotypic, involving a huge variety of cells, cytokines, growth factors, and a number of different MMPs. Targeting these different allies of tumor progression with nonselective inhibitors of TNF-α such as immunomodulatory drugs (thalidomide and its analogs), p38 MAPK inhibitors, TACE inhibitors, and NF-κB inhibitors, and combining them with other chemotherapeutic regimens, would be a preferential approach to inhibit tumor progression. This is all the more important in light of the fact that the function of one of these allies can be well compensated by yet another type of cytokine, growth factor, or MMP. In this situation, a multitargeted approach with chemotherapeutic drugs possessing several modes of action or combination of several chemotherapeutic drugs with varying modes of action is likely to prove highly effective in deterring tumor growth.

Cancer is a disease comprising heterogenous pathophysiologies and etiologies that differs with every cancer type. Anti-TNF-α therapy mainly targets the tumor stroma. However, tumor stroma of a cancer patient also differs with each cancer type and its respective pathophysiology. Thus, variations in response to anti-TNF-α therapy in cancer patients are likely to occur due to different tumor stroma and inflammatory profile of each patient. Furthermore, as we aim to target tumor inflammation mediators for antitumor therapies, the potential consequences of inhibiting these key players of immune system, which also participate in antitumor immune surveillance, should not be disregarded. Due to their ability to mediate immune system development as well as cytotoxicity of immune cells against tumor cells, they are indeed the integral components of the body's first-line defense against tumor development and form the basis for anticancer immunotherapy. Moreover, due to their pleiotropic nature, many other biological processes, not involving immune system but mediated by these inflammatory mediators, are also likely to be affected and thus accounting for a huge number of side effects seen with chemotherapeutic drugs inhibiting their action. Nonetheless, it will be an analysis of loss-to-benefit ratio for each cancer patient that will determine the success of anti-TNF-α therapy in the coming years.

A myriad of anti-TNF-α drugs are in developmental stages and in different stages of clinical trials. In the coming years, a better understanding of the tumor immunology and cell-signaling pathways is likely to precisely define the role of anti-TNF-α therapy in oncology. Also, the development of high-throughput screening (HTS) methods is likely to aid in the development of many more anti-TNF-α drugs for oncology. Thus, anti-TNF-α drugs are yet another addition to the mission for the achievement of the "complete cure of cancer."

List of Abbreviations

| | |
|---|---|
| ADAM | a distintegrin and metalloprotease |
| AGO | Argonaute |
| AIDS | acguired immuno deficiency syndrome |
| AML | Acute myelogenous leukemia |
| AMM | Agnogenic myeloid metaplasia |
| AP-1 | Activator protein-1 |
| APC | Antigen presenting cell |
| APP | Acute phase proteins |
| APR | Acute phase response |
| APRIL | A proliferation-inducing ligand |
| ARE | Adenine and uridine-rich elements |
| ARE-BP | ARE-binding protein |
| ATF | Activation transcription factor |
| BAFF | B-cell-activating factor |
| BAFFR | BAFF receptor |
| BCMA | B-cell maturation antigen |
| bFGF | Basic fibroblast growth factor |
| BMSC | Bone marrow stromal cell |
| cAMP | Cyclic adenosine monophosphate |
| CDE | Constitutive decay element |
| cDNA | complementary deoxyribonucleic acid |
| CD30L | CD30 ligand |
| CD40L | CD40 ligand |
| cGMP | Cyclic guanosine monophosphate |
| c-FLIP | Cellular-FLICE inhibitory protein |
| CHO | Chinese hamster ovary |
| cIAP | Cytoplasmic inhibitor of apoptosis protein |
| CLL | Chronic lymphocytic leukemia |
| CNS | Central nervous system |
| COX | Cyclooxyegenase |
| CRD | Cysteine-rich domain |

| | |
|---|---|
| CREB | Cyclic AMP-responsive element-binding protein |
| CRP | C-reactive protein |
| CSF | Cerebrospinal fluid |
| DD | Death domain |
| DR | Death receptor |
| DCR | Decoy receptor |
| EBV | Epstein–Barr virus |
| ECM | Extracellular matrix |
| EDA | Ectodermal dysplasin |
| EDAR | Ectodermal dysplasin receptor |
| EMEA | European Agency for the Evaluation of Medicinal Products |
| ENL | Erythema nodosum leporosum |
| ERK | Extracellular-signal-regulated kinase |
| FADD | Fas-associated death domain |
| FGF | Fibroblast growth factor |
| FXR1 | Fragile-Xmental-retardation–related protein 1 |
| GITRL | Glucocorticoid-induced TNFR family ligand |
| GITR | Glucocorticoid-induced TNFR family receptor |
| GM-CSF | Granulocyte-macrophage colony-stimulating factor |
| HBV | Hepatitis B virus |
| HCL | Hairy cell leukemia |
| HCV | Hepatitis C virus |
| HDL | High-density lipoprotein |
| HIV | human immuno deficiency virus |
| HPV | Human papillomavirus |
| HSL | Hormone-sensitive lipase |
| HTLV-1 | Human T-cell lymphotropic virus type 1 |
| HVEM | Herpes-virus entry mediator |
| ICAM-1 | Intercellular adhesion molecule-1 |
| Ig | Immunoglobulin |
| IGF-1 | Insulin-like growth factor-1 |
| IκB | Inhibitors of the NF-κB |
| IKK | Inhibitor of κB kinase |
| IFN-γ | Interferon-γ |
| IL | Interleukin |
| ILP | Isolated limb perfusion |
| IMiD | Immunomodulatory drug |
| iNOS | Inducible nitric-oxide synthase |
| JNK | c-JUN N-terminal kinase |
| LDL | Low-density lipoprotein |
| LIF | Leukemia inhibitory factor |
| LPL | Lipoprotein lipase |
| LPS | Lipopolysaccharide |
| LT-α | Lymphotoxin-alpha |
| MAPK | Mitogen-activated protein kinase |
| MAPKAP K2 | MAPK-activated protein kinase 2 |
| MAPKK/MEK | MAPK kinase |
| MAPKKK | MAPKK kinase |
| MCP-1 | Monocyte chemotactic protein-1 |
| M-CSF | Macrophage-colony stimulating factor |
| MDC | Macrophage-derived chemokine |
| MDS | Myelodysplastic syndromes |
| memTNF | Membrane-bound TNF |
| Meth-A | Methylcholanthrene-A |
| MHC | Major histocompatibility complex |
| MIF | Migration inhibitory factor |
| miRNA | micro-RNA |
| microRNP | Microribonucleoproteins |
| MM | Multiple myeloma |
| MMP | Matrix metalloprotease |
| mRNA | Messenger RNA |
| MVD | Microvessel density |
| NGF | Nerve growth factor |
| NGFR | Nerve growth factor receptor |
| NF-κB | Nuclear factor-κB |
| NK | Natural killer cells |
| NLS | Nuclear localization signal |
| NO | Nitric oxide |
| NPY | Neuropeptide Y |
| NSAID | Nonsteroidal anti-inflammatory drugs |
| OPG | Osteoprotegerin |
| PAF | Platelet-activating factor |
| PAI-RBP1 | Plasminogen activator inhibitor-1 RNA binding protein |
| PDE4 | Phosphodiestrase 4 |
| PG | Prostaglandin |
| PI3K | Phosphatidylinositol 3-kinase |
| PLAD | Preligand assembly domain |
| PLIN | Perilipin |
| PKA | Protein kinase A |
| RA | Rheumatoid arthritis |
| RANKL | TNF and receptor activator of NF-κB ligand |
| RELT | Receptor expressed in lymphoid tissues |
| RHD | Rel homology domain |
| RIP | Receptor interacting protein |
| RISC | RNA-induced silencing complex |
| ROS | Reactive oxygen species |
| SAM | Serum amyloid peptide |
| SelCID | Selective cytokine inhibitory drug |
| SLE | Systemic lupus eythematosus |
| SNP | Single nucleotide polymorphisms |

| | |
|---|---|
| SODD | Silencer of death domain |
| sTNF | Soluble TNF |
| S.T.E.P.S. | System for Thalidomide Education and Prescribing Safety |
| STS | Soft tissue sarcoma |
| $T_{max}$ | Mean time to peak plasma concentrations |
| TACE | TNF-α converting enzyme |
| TACI | Transmembrane activator and cyclophilin ligand interactor |
| TADC | Tumor-associated dendritic cell |
| TAM | Tumor-associated macrophages |
| TAV | Tumor-associated vasculature |
| TB | Tuberculosis |
| TCR | T-cell receptor |
| TGF-β | Transforming growth factor-β |
| TH | T helper cells |
| TIL | Tumor-infiltrating leucocytes |
| TM | Transmembrane |
| TNF-α | Tumor-necrosis factor-α |
| TNFR | TNF-α receptor |
| TNFSF | TNF superfamily |
| TNFRSF | TNFR superfamily |
| TRADD | TNFR-associated death domain |
| TRAF | TNFR-associated factors |
| TRAIL | TNF-related apoptosis-inducing ligand |
| TRAPS | TNFR-1 associated periodic syndromes |
| TTP | Tristetrapolin |
| TWEAK | TNF-like weak inducer of apoptosis |
| UCP | Uncoupling proteins |
| USFDA | U.S. Food and Drug Administration |
| UTR | Untranslated region |
| VCAM-1 | Vascular cell adhesion molecule-1 |
| VEGF | Vascular endothelial growth factor |
| VEGI | Vascular endothelial cell growth inhibitor |
| VLDL | Very low density lipoprotein |
| von Hippel–Lindau | pVHL |
| XEDAR | X-linked EDA receptor |

## ACKNOWLEDGMENTS

These studies were supported, in part, by grants RO1CA65920 from the NCI/NIH, 001 from the American Institute for Cancer Research, and EPSCoR from NDSU.

## REFERENCES

1. Coley WB (1891). II. Contribution to the knowledge of sarcoma. *Ann Surg 14*, 199–220.
2. Carswell EA, Old LJ, Kassel RL, Green S, Fiore N, Williamson B (1975). An endotoxin-induced serum factor that causes necrosis of tumors. *Proc Nat Acad Sci U S A 72*, 3666–3670.
3. Rouzer CA, Cerami A (1980). Hypertriglyceridemia associated with Trypanosoma brucei brucei infection in rabbits: Role of defective triglyceride removal. *Mol Biochem Parasitol 2*, 31–38.
4. Kawakami M, Cerami A (1981). Studies of endotoxin-induced decrease in lipoprotein lipase activity. *J Exp Med 154*, 631–639.
5. Beutler B, Greenwald D, Hulmes JD, Chang M, Pan YC, Mathison J, Ulevitch R, Cerami A (1985). Identity of tumour necrosis factor and the macrophage-secreted factor cachectin. *Nature 316*, 552–554.
6. Beutler B, Mahoney J, Le Trang N, Pekala P, Cerami A (1985). Purification of cachectin, a lipoprotein lipase-suppressing hormone secreted by endotoxin-induced RAW 264.7 cells. *J Exp Med 161*, 984–995.
7. Aggarwal BB, Kohr WJ, Hass PE, Moffat B, Spencer SA, Henzel WJ, Bringman TS, Nedwin GE, Goeddel DV, Harkins RN (1985). Human tumor necrosis factor. Production, purification, and characterization. *J Biol Chem 260*, 2345–2354.
8. Pennica D, Nedwin GE, Hayflick JS, Seeburg PH, Derynck R, Palladino MA, Kohr WJ, Aggarwal BB, Goeddel DV (1984). Human tumour necrosis factor: Precursor structure, expression and homology to lymphotoxin. *Nature 312*, 724–729.
9. Bashyam H (2007). Uncovering the dark side of TNF. *J Exp Med 204*, 2013.
10. Beutler B, Cerami A (1989). The biology of cachectin/TNF—a primary mediator of the host response. *Ann Rev Immunol 7*, 625–655.
11. Pennica D, Hayflick JS, Bringman TS, Palladino MA, Goeddel DV (1985). Cloning and expression in *Escherichia coli* of the cDNA for murine tumor necrosis factor. *Proc Nat Acad Sci U S A 82*, 6060–6064.
12. Granger GA, Shacks SJ, Williams TW, Kolb WP (1969). Lymphocyte in vitro cytotoxicity: Specific release of lymphotoxin-like materials from tuberculin-sensitive lymphoid cells. *Nature 221*, 1155–1157.
13. Aggarwal BB, Moffat B, Harkins RN (1984). Human lymphotoxin. Production by a lymphoblastoid cell line, purification, and initial characterization. *J Biol Chem 259*, 686–691.
14. Aggarwal BB, Henzel WJ, Moffat B, Kohr WJ, Harkins RN (1985). Primary structure of human lymphotoxin derived from 1788 lymphoblastoid cell line. *J Biol Chem 260*, 2334–2344.
15. Gray PW, Aggarwal BB, Benton CV, Bringman TS, Henzel WJ, Jarrett JA, Leung DW, Moffat B, Ng P,

Svedersky LP, et al. (1984). Cloning and expression of cDNA for human lymphotoxin, a lymphokine with tumour necrosis activity. *Nature 312*, 721–724.

16. Aggarwal BB (2003). Signalling pathways of the TNF superfamily: A double-edged sword. *Nat Rev Immunol 3*, 745–756.
17. Schottelius AJ, Moldawer LL, Dinarello CA, Asadullah K, Sterry W, Edwards CK 3rd. (2004). Biology of tumor necrosis factor-alpha—implications for psoriasis. *Exp Dermatol 13*, 193–222.
18. Idriss HT, Naismith JH (2000). TNF alpha and the TNF receptor superfamily: Structure-function relationship(s). *Microsc Res Tech 50*, 184–195.
19. Black RA, Rauch CT, Kozlosky CJ, Peschon JJ, Slack JL, Wolfson MF, Castner BJ, Stocking KL, Reddy P, Srinivasan S, Nelson N, Boiani N, Schooley KA, Gerhart M, Davis R, Fitzner JN, Johnson RS, Paxton RJ, March CJ, Cerretti DP (1997). A metalloproteinase disintegrin that releases tumour-necrosis factor-alpha from cells. *Nature 385*, 729–733.
20. Lum L, Wong BR, Josien R, Becherer JD, Erdjument-Bromage H, Schlondorff J, Tempst P, Choi Y, Blobel CP (1999). Evidence for a role of a tumor necrosis factor-alpha (TNF-alpha)-converting enzyme-like protease in shedding of TRANCE, a TNF family member involved in osteoclastogenesis and dendritic cell survival. *J Biol Chem 274*, 13613–13618.
21. Powell WC, Fingleton B, Wilson CL, Boothby M, Matrisian LM (1999). The metalloproteinase matrilysin proteolytically generates active soluble Fas ligand and potentiates epithelial cell apoptosis. *Curr Biol 9*, 1441–1447.
22. Chen Y, Molloy SS, Thomas L, Gambee J, Bachinger HP, Ferguson B, Zonana J, Thomas G, Morris NP (2001). Mutations within a furin consensus sequence block proteolytic release of ectodysplasin-A and cause X-linked hypohidrotic ectodermal dysplasia. *Proc Nat Acad Sci U S A 98*, 7218–7223.
23. Schneider P, MacKay F, Steiner V, Hofmann K, Bodmer JL, Holler N, Ambrose C, Lawton P, Bixler S, Acha-Orbea H, Valmori D, Romero P, Werner-Favre C, Zubler RH, Browning JL, Tschopp J (1999). BAFF, a novel ligand of the tumor necrosis factor family, stimulates B cell growth. *J Exp Med 189*, 1747–1756.
24. Suda T, Hashimoto H, Tanaka M, Ochi T, Nagata S (1997). Membrane Fas ligand kills human peripheral blood T lymphocytes, and soluble Fas ligand blocks the killing. *J Exp Med 186*, 2045–2050.
25. Bazan JF (1993). Emerging families of cytokines and receptors. *Curr Biol 3*, 603–606.
26. Fesik SW (2000). Insights into programmed cell death through structural biology. *Cell 103*, 273–282.
27. Locksley RM, Killeen N, Lenardo MJ (2001). The TNF and TNF receptor superfamilies: Integrating mammalian biology. *Cell 104*, 487–501.
28. Wajant H, Pfizenmaier K, Scheurich P (2003). Tumor necrosis factor signaling. *Cell Death Differ 10*, 45–65.
29. Smith CA, Farrah T, Goodwin RG (1994). The TNF receptor superfamily of cellular and viral proteins: Activation, costimulation, and death. *Cell 76*, 959–962.
30. Itoh N, Yonehara S, Ishii A, Yonehara M, Mizushima S, Sameshima M, Hase A, Seto Y, Nagata S (1991). The polypeptide encoded by the cDNA for human cell surface antigen Fas can mediate apoptosis. *Cell 66*, 233–243.
31. Arch RH, Gedrich RW, Thompson CB (1998). Tumor necrosis factor receptor-associated factors (TRAFs)—a family of adapter proteins that regulates life and death. *Genes Dev 12*, 2821–2830.
32. McWhirter SM, Pullen SS, Werneburg BG, Labadia ME, Ingraham RH, Crute JJ, Kehry MR, Alber T (1999). Structural and biochemical analysis of signal transduction by the TRAF family of adapter proteins. *Cold Spring Harb Symp Quant Biol 64*, 551–562.
33. Ashkenazi A (2002). Targeting death and decoy receptors of the tumour-necrosis factor superfamily. *Nat Rev Cancer 2*, 420–430.
34. Naismith JH, Sprang SR (1998). Modularity in the TNF-receptor family. *Trends Biochem Sci 23*, 74–79.
35. Chan FK, Chun HJ, Zheng L, Siegel RM, Bui KL, Lenardo MJ (2000). A domain in TNF receptors that mediates ligand-independent receptor assembly and signaling. *Science 288*, 2351–2354.
36. Szlosarek PW, Balkwill FR (2003). Tumour necrosis factor alpha: A potential target for the therapy of solid tumours. *Lancet Oncol 4*, 565–573.
37. McDermott MF, Aksentijevich I, Galon J, McDermott EM, Ogunkolade BW, Centola M, Mansfield E, Gadina M, Karenko L, Pettersson T, McCarthy J, Frucht DM, Aringer M, Torosyan Y, Teppo AM, Wilson M, Karaarslan HM, Wan Y, Todd I, Wood G, Schlimgen R, Kumarajeewa TR, Cooper SM, Vella JP, Amos CI, Mulley J, Quane KA, Molloy MG, Ranki A, Powell RJ, Hitman GA, O'Shea JJ, Kastner DL (1999). Germline mutations in the extracellular domains of the 55 kDa TNF receptor, TNFR1, define a family of dominantly inherited autoinflammatory syndromes. *Cell 97*, 133–144.
38. Edwards CK 3rd. (1999). PEGylated recombinant human soluble tumour necrosis factor receptor type I (r-HusTNF-RI): Novel high affinity TNF receptor designed for chronic inflammatory diseases. *Ann Rheum Dis 58* (Suppl 1), I73–I81.
39. Frishman JI, Edwards CK 3rd, Sonnenberg MG, Kohno T, Cohen AM, Dinarello CA (2000). Tumor necrosis factor (TNF)-alpha-induced interleukin-8 in human blood cultures discriminates neutralization by the p55 and p75 TNF soluble receptors. *J Infect Dis 182*, 1722–1730.
40. Grell M, Wajant H, Zimmermann G, Scheurich P (1998). The type 1 receptor (CD120a) is the high-affinity receptor for soluble tumor necrosis factor. *Proc Nat Acad Sci U S A 95*, 570–575.
41. Grell M, Douni E, Wajant H, Lohden M, Clauss M, Maxeiner B, Georgopoulos S, Lesslauer W, Kollias G, Pfizenmaier K, Scheurich P (1995). The transmembrane

form of tumor necrosis factor is the prime activating ligand of the 80 kDa tumor necrosis factor receptor. *Cell 83*, 793–802.

42. Tartaglia LA, Pennica D, Goeddel DV (1993). Ligand passing: The 75-kDa tumor necrosis factor (TNF) receptor recruits TNF for signaling by the 55-kDa TNF receptor. *J Biol Chem 268*, 18542–18548.
43. Ashkenazi A, Dixit VM (1998). Death receptors: Signaling and modulation. *Science 281*, 1305–1308.
44. Jiang Y, Woronicz JD, Liu W, Goeddel DV (1999). Prevention of constitutive TNF receptor 1 signaling by silencer of death domains. *Science 283*, 543–546.
45. Takada H, Chen NJ, Mirtsos C, Suzuki S, Suzuki N, Wakeham A, Mak TW, Yeh WC (2003). Role of SODD in regulation of tumor necrosis factor responses. *Mol Cell Biol 23*, 4026–4033.
46. Hsu H, Xiong J, Goeddel DV (1995). The TNF receptor 1-associated protein TRADD signals cell death and NF-kappa B activation. *Cell 81*, 495–504.
47. Wullaert A, Heyninck K, Beyaert R (2006). Mechanisms of crosstalk between TNF-induced NF-kappaB and JNK activation in hepatocytes. *Biochem Pharmacol 72*, 1090–1101.
48. Hsu H, Huang J, Shu HB, Baichwal V, Goeddel DV (1996). TNF-dependent recruitment of the protein kinase RIP to the TNF receptor-1 signaling complex. *Immunity 4*, 387–396.
49. Hsu H, Shu HB, Pan MG, Goeddel DV (1996). TRADD-TRAF2 and TRADD-FADD interactions define two distinct TNF receptor 1 signal transduction pathways. *Cell 84*, 299–308.
50. Boldin MP, Goncharov TM, Goltsev YV, Wallach D (1996). Involvement of MACH, a novel MORT1/FADD-interacting protease, in Fas/APO-1- and TNF receptor-induced cell death. *Cell 85*, 803–815.
51. Muzio M, Chinnaiyan AM, Kischkel FC, O'Rourke K, Shevchenko A, Ni J, Scaffidi C, Bretz JD, Zhang M, Gentz R, Mann M, Krammer PH, Peter ME, Dixit VM (1996). FLICE, a novel FADD-homologous ICE/CED-3-like protease, is recruited to the CD95 (Fas/APO-1) death-inducing signaling complex. *Cell 85*, 817–827.
52. Degterev A, Boyce M, Yuan J (2003). A decade of caspases. *Oncogene 22*, 8543–8567.
53. Shu HB, Takeuchi M, Goeddel DV (1996). The tumor necrosis factor receptor 2 signal transducers TRAF2 and c-IAP1 are components of the tumor necrosis factor receptor 1 signaling complex. *Proc Nat Acad Sci U S A 93*, 13973–13978.
54. Fotin-Mleczek M, Henkler F, Samel D, Reichwein M, Hausser A, Parmryd I, Scheurich P, Schmid JA, Wajant H (2002). Apoptotic crosstalk of TNF receptors: TNFR-2-induces depletion of TRAF2 and IAP proteins and accelerates TNFR-1-dependent activation of caspase-8. *J Cell Sci 115*, 2757–2770.
55. Weiss T, Grell M, Hessabi B, Bourteele S, Muller G, Scheurich P, Wajant H (1997). Enhancement of TNF receptor p60-mediated cytotoxicity by TNF receptor p80: Requirement of the TNF receptor-associated factor-2 binding site. *J Immunol 158*, 2398–2404.
56. Li X, Yang Y, Ashwell JD (2002). TNF-RII and c-IAP1 mediate ubiquitination and degradation of TRAF2. *Nature 416*, 345–347.
57. Ghosh S, May MJ, Kopp EB (1998). NF-kappa B and Rel proteins: Evolutionarily conserved mediators of immune responses. *Ann Rev Immunol 16*, 225–260.
58. Karin M, Ben-Neriah Y (2000). Phosphorylation meets ubiquitination: The control of NF-[kappa]B activity. *Ann Rev Immunol 18*, 621–663.
59. Ghosh S, Karin M (2002). Missing pieces in the NF-kappaB puzzle. *Cell 109* (Suppl), S81–96.
60. Baud V, Karin M (2001). Signal transduction by tumor necrosis factor and its relatives. *Trends Cell Biol 11*, 372–377.
61. Beg AA, Baltimore D (1996). An essential role for NF-kappaB in preventing TNF-alpha-induced cell death. *Science 274*, 782–784.
62. Chen G, Goeddel DV (2002). TNFR-1 signaling: A beautiful pathway. *Science 296*, 1634–1635.
63. Lee SY, Reichlin A, Santana A, Sokol KA, Nussenzweig MC, Choi Y (1997). TRAF2 is essential for JNK but not NF-kappaB activation and regulates lymphocyte proliferation and survival. *Immunity 7*, 703–713.
64. Yeh WC, Shahinian A, Speiser D, Kraunus J, Billia F, Wakeham A, de la Pompa JL, Ferrick D, Hum B, Iscove N, Ohashi P, Rothe M, Goeddel DV, Mak TW (1997). Early lethality, functional NF-kappaB activation, and increased sensitivity to TNF-induced cell death in TRAF2-deficient mice. *Immunity 7*, 715–725.
65. Chang L, Karin M (2001). Mammalian MAP kinase signalling cascades. *Nature 410*, 37–40.
66. Brenner DA, O'Hara M, Angel P, Chojkier M, Karin M (1989). Prolonged activation of jun and collagenase genes by tumour necrosis factor-alpha. *Nature 337*, 661–663.
67. Hanazawa S, Takeshita A, Amano S, Semba T, Nirazuka T, Katoh H, Kitano S (1993). Tumor necrosis factor-alpha induces expression of monocyte chemoattractant JE via fos and jun genes in clonal osteoblastic MC3T3-E1 cells. *J Biol Chem 268*, 9526–9532.
68. Min W, Pober JS (1997). TNF initiates E-selectin transcription in human endothelial cells through parallel TRAF-NF-kappa B and TRAF-RAC/CDC42-JNK-c-Jun/ATF2 pathways. *J Immunol 159*, 3508–3518.
69. Bigda J, Beletsky I, Brakebusch C, Varfolomeev Y, Engelmann H, Bigda J, Holtmann H, Wallach D (1994). Dual role of the p75 tumor necrosis factor (TNF) receptor in TNF cytotoxicity. *J Exp Med 180*, 445–460.
70. Grell M, Scheurich P, Meager A, Pfizenmaier K (1993). TR60 and TR80 tumor necrosis factor (TNF)-receptors can independently mediate cytolysis. *Lymphokine Cytokine Res 12*, 143–148.

71. Medvedev AE, Sundan A, Espevik T (1994). Involvement of the tumor necrosis factor receptor p75 in mediating cytotoxicity and gene regulating activities. *Eur J Immunol 24*, 2842–2849.
72. Grell M, Zimmermann G, Gottfried E, Chen CM, Grunwald U, Huang DC, Wu Lee YH, Durkop H, Engelmann H, Scheurich P, Wajant H, Strasser A (1999). Induction of cell death by tumour necrosis factor (TNF) receptor 2, CD40 and CD30: A role for TNFR-1 activation by endogenous membrane-anchored TNF. *EMBO J 18*, 3034–3043.
73. Vercammen D, Vandenabeele P, Declercq W, Van de Craen M, Grooten J, Fiers W (1995). Cytotoxicity in L929 murine fibrosarcoma cells after triggering of transfected human p75 tumour necrosis factor (TNF) receptor is mediated by endogenous murine TNF. *Cytokine 7*, 463–470.
74. Brown KD, Hostager BS, Bishop GA (2002). Regulation of TRAF2 signaling by self-induced degradation. *J Biol Chem 277*, 19433–19438.
75. Ksontini R, MacKay SL, Moldawer LL (1998). Revisiting the role of tumor necrosis factor alpha and the response to surgical injury and inflammation. *Arch Surg 133*, 558–567.
76. Mannel DN, Echtenacher B (2000). TNF in the inflammatory response. *Chem Immunol 74*, 141–161.
77. Feldmann M, Brennan FM, Elliott MJ, Williams RO, Maini RN (1995). TNF alpha is an effective therapeutic target for rheumatoid arthritis. *Ann N Y Acad Sci 766*, 272–278.
78. Bartlett JB, Dredge K, Dalgleish AG (2004). The evolution of thalidomide and its IMiD derivatives as anticancer agents. *Nat Rev Cancer 4*, 314–322.
79. Eigler A, Sinha B, Hartmann G, Endres S (1997). Taming TNF: Strategies to restrain this proinflammatory cytokine. *Immunol Today 18*, 487–492.
80. Allan SM, Rothwell NJ (2001). Cytokines and acute neurodegeneration. *Nat Rev Neurosci 2*, 734–744.
81. Pan W, Kastin AJ (2001). Increase in TNFalpha transport after SCI is specific for time, region, and type of lesion. *Exp Neurol 170*, 357–363.
82. Mattson MP, Goodman Y, Luo H, Fu W, Furukawa K (1997). Activation of NF-kappaB protects hippocampal neurons against oxidative stress-induced apoptosis: Evidence for induction of manganese superoxide dismutase and suppression of peroxynitrite production and protein tyrosine nitration. *J Neurosci Res 49*, 681–697.
83. Barone FC, Arvin B, White RF, Miller A, Webb CL, Willette RN, Lysko PG, Feuerstein GZ (1997). Tumor necrosis factor-alpha. A mediator of focal ischemic brain injury. *Stroke 28*, 1233–1244.
84. Mayne M, Ni W, Yan HJ, Xue M, Johnston JB, Del Bigio MR, Peeling J, Power C (2001). Antisense oligodeoxynucleotide inhibition of tumor necrosis factor-alpha expression is neuroprotective after intracerebral hemorrhage. *Stroke 32*, 240–248.
85. Green A, Dobias SB, Walters DJ, Brasier AR (1994). Tumor necrosis factor increases the rate of lipolysis in primary cultures of adipocytes without altering levels of hormone-sensitive lipase. *Endocrinology 134*, 2581–2588.
86. Ryden M, Dicker A, van Harmelen V, Hauner H, Brunnberg M, Perbeck L, Lonnqvist F, Arner P (2002). Mapping of early signaling events in tumor necrosis factor-alpha-mediated lipolysis in human fat cells. *J Biol Chem 277*, 1085–1091.
87. Souza SC, de Vargas LM, Yamamoto MT, Lien P, Franciosa MD, Moss LG, Greenberg AS (1998). Overexpression of perilipin A and B blocks the ability of tumor necrosis factor alpha to increase lipolysis in 3T3-L1 adipocytes. *J Biol Chem 273*, 24665–24669.
88. Popa C, Netea MG, Radstake T, Van der Meer JW, Stalenhoef AF, van Riel PL, Barerra P (2005). Influence of anti-tumour necrosis factor therapy on cardiovascular risk factors in patients with active rheumatoid arthritis. *Ann Rheum Dis 64*, 303–305.
89. DiDonato J, Mercurio F, Rosette C, Wu-Li J, Suyang H, Ghosh S, Karin M (1996). Mapping of the inducible IkappaB phosphorylation sites that signal its ubiquitination and degradation. *Mol Cell Biol 16*, 1295–1304.
90. Guttridge DC, Mayo MW, Madrid LV, Wang CY, Baldwin AS, Jr (2000). NF-kappaB-induced loss of MyoD messenger RNA: Possible role in muscle decay and cachexia. *Science 289*, 2363–2366.
91. Wheeler MT, Snyder EC, Patterson MN, Swoap SJ (1999). An E-box within the MHC IIB gene is bound by MyoD and is required for gene expression in fast muscle. *Am J Physiol 276*, C1069–1078.
92. Acharyya S, Ladner KJ, Nelsen LL, Damrauer J, Reiser PJ, Swoap S, Guttridge DC (2004). Cancer cachexia is regulated by selective targeting of skeletal muscle gene products. *J Clin Invest 114*, 370–378.
93. Tracey KJ, Cerami A (1994). Tumor necrosis factor: A pleiotropic cytokine and therapeutic target. *Annu Rev Med 45*, 491–503.
94. Argiles JM, Lopez-Soriano FJ (1999). The role of cytokines in cancer cachexia. *Med Res Rev 19*, 223–248.
95. Oliff A, Defeo-Jones D, Boyer M, Martinez D, Kiefer D, Vuocolo G, Wolfe A, Socher SH (1987). Tumors secreting human TNF/cachectin induce cachexia in mice. *Cell 50*, 555–563.
96. Strieter RM, Remick DG, Lynch JP 3rd, Genord M, Raiford C, Spengler R, Kunkel SL (1989). Differential regulation of tumor necrosis factor-alpha in human alveolar macrophages and peripheral blood monocytes: A cellular and molecular analysis. *Am J Respir Cell Mol Biol 1*, 57–63.
97. Beutler B, Krochin N, Milsark IW, Luedke C, Cerami A (1986). Control of cachectin (tumor necrosis factor) synthesis: Mechanisms of endotoxin resistance. *Science 232*, 977–980.
98. Nedwin GE, Naylor SL, Sakaguchi AY, Smith D, Jarrett-Nedwin J, Pennica D, Goeddel DV, Gray PW (1985).

Human lymphotoxin and tumor necrosis factor genes: Structure, homology and chromosomal localization. *Nucleic Acids Res 13*, 6361–6373.

99. Spies T, Morton CC, Nedospasov SA, Fiers W, Pious D, Strominger JL (1986). Genes for the tumor necrosis factors alpha and beta are linked to the human major histocompatibility complex. *Proc Natl Acad Sci U S A 83*, 8699–8702.
100. Nedospasov SA, Shakhov AN, Turetskaya RL, Mett VA, Azizov MM, Georgiev GP, Korobko VG, Dobrynin VN, Filippov SA, Bystrov NS, et al. (1986). Tandem arrangement of genes coding for tumor necrosis factor (TNF-alpha) and lymphotoxin (TNF-beta) in the human genome. *Cold Spring Harb Symp Quant Biol 51* (Pt 1), 611–624.
101. Economou JS, Rhoades K, Essner R, McBride WH, Gasson JC, Morton DL (1989). Genetic analysis of the human tumor necrosis factor alpha/cachectin promoter region in a macrophage cell line. *J Exp Med 170*, 321–326.
102. Leitman DC, Ribeiro RC, Mackow ER, Baxter JD, West BL (1991). Identification of a tumor necrosis factor-responsive element in the tumor necrosis factor alpha gene. *J Biol Chem 266*, 9343–9346.
103. Jongeneel CV, Shakhov AN, Nedospasov SA, Cerottini JC (1989). Molecular control of tissue-specific expression at the mouse TNF locus. *Eur J Immunol 19*, 549–552.
104. Pauli U (1994). Control of tumor necrosis factor gene expression. *Crit Rev Eukaryot Gene Expr 4*, 323–344.
105. Kronke M, Hensel G, Schluter C, Scheurich P, Schutze S, Pfizenmaier K (1988). Tumor necrosis factor and lymphotoxin gene expression in human tumor cell lines. *Cancer Res 48*, 5417–5421.
106. Sung SJ, Walters JA, Hudson J, Gimble JM (1991). Tumor necrosis factor-alpha mRNA accumulation in human myelomonocytic cell lines. Role of transcriptional regulation by DNA sequence motifs and mRNA stabilization. *J Immunol 147*, 2047–2054.
107. Clark A (2000). Post-transcriptional regulation of pro-inflammatory gene expression. *Arthritis Res 2*, 172–174.
108. Wilusz CJ, Wormington M, Peltz SW (2001). The cap-to-tail guide to mRNA turnover. *Nat Rev Mol Cell Biol 2*, 237–246.
109. Chen CY, Shyu AB (1995). AU-rich elements: Characterization and importance in mRNA degradation. *Trends Biochem Sci 20*, 465–470.
110. Fialcowitz EJ, Brewer BY, Keenan BP, Wilson GM (2005). A hairpin-like structure within an AU-rich mRNA-destabilizing element regulates trans-factor binding selectivity and mRNA decay kinetics. *J Biol Chem 280*, 22406–22417.
111. Dumitru CD, Ceci JD, Tsatsanis C, Kontoyiannis D, Stamatakis K, Lin JH, Patriotis C, Jenkins NA, Copeland NG, Kollias G, Tsichlis PN (2000). TNF-alpha induction by LPS is regulated posttranscriptionally via a Tpl2/ERK-dependent pathway. *Cell 103*, 1071–1083.
112. Zhang T, Kruys V, Huez G, Gueydan C (2002). AU-rich element-mediated translational control: Complexity and multiple activities of trans-activating factors. *Biochem Soc Trans 30*, 952–958.
113. Baseggio L, Charlot C, Bienvenu J, Felman P, Salles G (2002). Tumor necrosis factor-alpha mRNA stability in human peripheral blood cells after lipopolysaccharide stimulation. *Eur Cytokine Netw 13*, 92–98.
114. Garnon J, Lachance C, Di Marco S, Hel Z, Marion D, Ruiz MC, Newkirk MM, Khandjian EW, Radzioch D (2005). Fragile X-related protein FXR1P regulates pro-inflammatory cytokine tumor necrosis factor expression at the post-transcriptional level. *J Biol Chem 280*, 5750–5763.
115. Xu N, Chen CY, Shyu AB (1997). Modulation of the fate of cytoplasmic mRNA by AU-rich elements: Key sequence features controlling mRNA deadenylation and decay. *Mol Cell Biol 17*, 4611–4621.
116. Wilusz CJ, Gao M, Jones CL, Wilusz J, Peltz SW (2001). Poly(A)-binding proteins regulate both mRNA deadenylation and decapping in yeast cytoplasmic extracts. *RNA 7*, 1416–1424.
117. Mukherjee D, Gao M, O'Connor JP, Raijmakers R, Pruijn G, Lutz CS, Wilusz J (2002). The mammalian exosome mediates the efficient degradation of mRNAs that contain AU-rich elements. *EMBO J 21*, 165–174.
118. Stoecklin G, Mayo T, Anderson P (2006). ARE-mRNA degradation requires the 5'-3' decay pathway. *EMBO Rep 7*, 72–77.
119. Lieberman AP, Pitha PM, Shin ML (1992). Poly(A) removal is the kinase-regulated step in tumor necrosis factor mRNA decay. *J Biol Chem 267*, 2123–2126.
120. Jing Q, Huang S, Guth S, Zarubin T, Motoyama A, Chen J, Di Padova F, Lin SC, Gram H, Han J (2005). Involvement of microRNA in AU-rich element-mediated mRNA instability. *Cell 120*, 623–634.
121. Zhao Y, Srivastava D (2007). A developmental view of microRNA function. *Trends Biochem Sci 32*, 189–197.
122. Saxena S, Jonsson ZO, Dutta A (2003). Small RNAs with imperfect match to endogenous mRNA repress translation. Implications for off-target activity of small inhibitory RNA in mammalian cells. *J Biol Chem 278*, 44312–44319.
123. Caudy AA, Myers M, Hannon GJ, Hammond SM (2002). Fragile X-related protein and VIG associate with the RNA interference machinery. *Genes Dev 16*, 2491–2496.
124. Jin P, Zarnescu DC, Ceman S, Nakamoto M, Mowrey J, Jongens TA, Nelson DL, Moses K, Warren ST (2004). Biochemical and genetic interaction between the fragile X mental retardation protein and the microRNA pathway. *Nat Neurosci 7*, 113–117.
125. Xu K, Bogert BA, Li W, Su K, Lee A, Gao FB (2004). The fragile X-related gene affects the crawling behavior of Drosophila larvae by regulating the mRNA level of the DEG/ENaC protein pickpocket1. *Curr Biol 14*, 1025–1034.

126. Meister G, Landthaler M, Patkaniowska A, Dorsett Y, Teng G, Tuschl T (2004). Human Argonaute2 mediates RNA cleavage targeted by miRNAs and siRNAs. *Mol Cell 15*, 185–197.
127. Liu J, Carmell MA, Rivas FV, Marsden CG, Thomson JM, Song JJ, Hammond SM, Joshua-Tor L, Hannon GJ (2004). Argonaute2 is the catalytic engine of mammalian RNAi. *Science 305*, 1437–1441.
128. Pillai RS, Artus CG, Filipowicz W (2004). Tethering of human Ago proteins to mRNA mimics the miRNA-mediated repression of protein synthesis. *RNA 10*, 1518–1525.
129. Jakymiw A, Lian S, Eystathioy T, Li S, Satoh M, Hamel JC, Fritzler MJ, Chan EK (2005). Disruption of GW bodies impairs mammalian RNA interference. *Nat Cell Biol 7*, 1267–1274.
130. Liu J, Valencia-Sanchez MA, Hannon GJ, Parker R (2005). MicroRNA-dependent localization of targeted mRNAs to mammalian P-bodies. *Nat Cell Biol 7*, 719–723.
131. Pillai RS, Bhattacharyya SN, Artus CG, Zoller T, Cougot N, Basyuk E, Bertrand E, Filipowicz W (2005). Inhibition of translational initiation by Let-7 MicroRNA in human cells. *Science 309*, 1573–1576.
132. Stoecklin G, Stubbs T, Kedersha N, Wax S, Rigby WF, Blackwell TK, Anderson P (2004). MK2-induced tristetraprolin:14-3-3 complexes prevent stress granule association and ARE-mRNA decay. *EMBO J 23*, 1313–1324.
133. Vasudevan S, Tong Y, Steitz JA (2007). Switching from repression to activation: MicroRNAs can up-regulate translation. *Science 318*, 1931–1934.
134. Carmell MA, Xuan Z, Zhang MQ, Hannon GJ (2002). The Argonaute family: Tentacles that reach into RNAi, developmental control, stem cell maintenance, and tumorigenesis. *Genes Dev 16*, 2733–2742.
135. Vasudevan S, Steitz JA (2007). AU-rich-element-mediated upregulation of translation by FXR1 and Argonaute 2. *Cell 128*, 1105–1118.
136. Valencia-Sanchez MA, Liu J, Hannon GJ, Parker R (2006). Control of translation and mRNA degradation by miRNAs and siRNAs. *Genes Dev 20*, 515–524.
137. Tili E, Michaille JJ, Cimino A, Costinean S, Dumitru CD, Adair B, Fabbri M, Alder H, Liu CG, Calin GA, Croce CM (2007). Modulation of miR-155 and miR-125b levels following lipopolysaccharide/TNF-alpha stimulation and their possible roles in regulating the response to endotoxin shock. *J Immunol 179*, 5082–5089.
138. Stoecklin G, Lu M, Rattenbacher B, Moroni C (2003). A constitutive decay element promotes tumor necrosis factor alpha mRNA degradation via an AU-rich element-independent pathway. *Mol Cell Biol 23*, 3506–3515.
139. Stoecklin G, Ming XF, Looser R, Moroni C (2000). Somatic mRNA turnover mutants implicate tristetraprolin in the interleukin-3 mRNA degradation pathway. *Mol Cell Biol 20*, 3753–3763.
140. Aggarwal BB, Shishodia S, Sandur SK, Pandey MK, Sethi G (2006). Inflammation and cancer: How hot is the link? *Biochem Pharmacol 72*, 1605–1621.
141. Thun MJ, Henley SJ, Patrono C (2002). Nonsteroidal anti-inflammatory drugs as anticancer agents: Mechanistic, pharmacologic, and clinical issues. *J Nat Cancer Inst 94*, 252–266.
142. Mantovani A (2005). Cancer: Inflammation by remote control. *Nature 435*, 752–753.
143. Balkwill F, Mantovani A (2001). Inflammation and cancer: Back to Virchow? *Lancet 357*, 539–545.
144. Coussens LM, Werb Z (2002). Inflammation and cancer. *Nature 420*, 860–867.
145. Balkwill F, Charles KA, Mantovani A (2005). Smoldering and polarized inflammation in the initiation and promotion of malignant disease. *Cancer Cell 7*, 211–217.
146. Marx J (2004). Cancer research. Inflammation and cancer: The link grows stronger. *Science 306*, 966–968.
147. Dvorak HF (1986). Tumors: Wounds that do not heal. Similarities between tumor stroma generation and wound healing. *N Engl J Med 315*, 1650–1659.
148. Perwez Hussain S, Harris CC (2007). Inflammation and cancer: An ancient link with novel potentials. *Int J Cancer 121*, 2373–2380.
149. Allavena P, Sica A, Vecchi A, Locati M, Sozzani S, Mantovani A (2000). The chemokine receptor switch paradigm and dendritic cell migration: Its significance in tumor tissues. *Immunol Rev 177*, 141–149.
150. Mantovani A, Bussolino F, Dejana E (1992). Cytokine regulation of endothelial cell function. *FASEB J 6*, 2591–2599.
151. Coussens LM, Tinkle CL, Hanahan D, Werb Z (2000). MMP-9 supplied by bone marrow-derived cells contributes to skin carcinogenesis. *Cell 103*, 481–490.
152. Mizoguchi H, O'Shea JJ, Longo DL, Loeffler CM, McVicar DW, Ochoa AC (1992). Alterations in signal transduction molecules in T lymphocytes from tumor-bearing mice. *Science 258*, 1795–1798.
153. Yoshimura A (2006). Signal transduction of inflammatory cytokines and tumor development. *Cancer Sci 97*, 439–447.
154. Dranoff G (2004). Cytokines in cancer pathogenesis and cancer therapy. *Nat Rev Cancer 4*, 11–22.
155. Dalgleish AG (2007). Cytokines in the genesis and treatment of cancer. *Br J Cancer 97*, 1598–1599.
156. Balkwill F (2004). Cancer and the chemokine network. *Nat Rev Cancer 4*, 540–550.
157. Yan L, Anderson GM, DeWitte M, Nakada MT (2006). Therapeutic potential of cytokine and chemokine antagonists in cancer therapy. *Eur J Cancer 42*, 793–802.
158. Koong AC, Denko NC, Hudson KM, Schindler C, Swiersz L, Koch C, Evans S, Ibrahim H, Le QT, Terris DJ, Giaccia AJ (2000). Candidate genes for the hypoxic tumor phenotype. *Cancer Res 60*, 883–887.

159. Aggarwal BB (2004). Nuclear factor-kappaB: The enemy within. *Cancer Cell 6*, 203–208.
160. Pikarsky E, Porat RM, Stein I, Abramovitch R, Amit S, Kasem S, Gutkovich-Pyest E, Urieli-Shoval S, Galun E, Ben-Neriah Y (2004). NF-kappaB functions as a tumour promoter in inflammation-associated cancer. *Nature 431*, 461–466.
161. Mann DA, Oakley F (2005). NF-kappaB: A signal for cancer. *J Hepatol 42*, 610–611.
162. Greten FR, Eckmann L, Greten TF, Park JM, Li ZW, Egan LJ, Kagnoff MF, Karin M (2004). IKKbeta links inflammation and tumorigenesis in a mouse model of colitis-associated cancer. *Cell 118*, 285–296.
163. Balkwill F, Coussens LM (2004). Cancer: An inflammatory link. *Nature 431*, 405–406.
164. Balkwill FR, Lee A, Aldam G, Moodie E, Thomas JA, Tavernier J, Fiers W (1986). Human tumor xenografts treated with recombinant human tumor necrosis factor alone or in combination with interferons. *Cancer Res 46*, 3990–3993.
165. Talmadge JE, Phillips H, Schneider M, Rowe T, Pennington R, Bowersox O, Lenz B (1988). Immunomodulatory properties of recombinant murine and human tumor necrosis factor. *Cancer Res 48*, 544–550.
166. Creasey AA, Reynolds MT, Laird W (1986). Cures and partial regression of murine and human tumors by recombinant human tumor necrosis factor. *Cancer Res 46*, 5687–5690.
167. de Kossodo S, Moore R, Gschmeissner S, East N, Upton C, Balkwill FR (1995). Changes in endogenous cytokines, adhesion molecules and platelets during cytokine-induced tumour necrosis. *Br J Cancer 72*, 1165–1172.
168. Balkwill F (2002). Tumor necrosis factor or tumor promoting factor? *Cytokine Growth Factor Rev 13*, 135–141.
169. Watanabe N, Niitsu Y, Umeno H, Sone H, Neda H, Yamauchi N, Maeda M, Urushizaki I (1988). Synergistic cytotoxic and antitumor effects of recombinant human tumor necrosis factor and hyperthermia. *Cancer Res 48*, 650–653.
170. Nawroth P, Handley D, Matsueda G, De Waal R, Gerlach H, Blohm D, Stern D (1988). Tumor necrosis factor/cachectin-induced intravascular fibrin formation in meth A fibrosarcomas. *J Exp Med 168*, 637–647.
171. Carmeliet P, Jain RK (2000). Angiogenesis in cancer and other diseases. *Nature 407*, 249–257.
172. Baluk P, Hashizume H, McDonald DM (2005). Cellular abnormalities of blood vessels as targets in cancer. *Curr Opin Genet Dev 15*, 102–111.
173. Hashizume H, Baluk P, Morikawa S, McLean JW, Thurston G, Roberge S, Jain RK, McDonald DM (2000). Openings between defective endothelial cells explain tumor vessel leakiness. *Am J Pathol 156*, 1363–1380.
174. North RJ, Havell EA (1988). The antitumor function of tumor necrosis factor (TNF) II. Analysis of the role of endogenous TNF in endotoxin-induced hemorrhagic necrosis and regression of an established sarcoma. *J Exp Med 167*, 1086–1099.
175. Havell EA, Fiers W, North RJ (1988). The antitumor function of tumor necrosis factor (TNF), I. Therapeutic action of TNF against an established murine sarcoma is indirect, immunologically dependent, and limited by severe toxicity. *J Exp Med 167*, 1067–1085.
176. Ruegg C, Yilmaz A, Bieler G, Bamat J, Chaubert P, Lejeune FJ (1998). Evidence for the involvement of endothelial cell integrin alphaVbeta3 in the disruption of the tumor vasculature induced by TNF and IFN-gamma. *Nature Med 4*, 408–414.
177. van der Veen AH, de Wilt JH, Eggermont AM, van Tiel ST, Seynhaeve AL, ten Hagen TL (2000). TNF-alpha augments intratumoural concentrations of doxorubicin in TNF-alpha-based isolated limb perfusion in rat sarcoma models and enhances anti-tumour effects. *Br J Cancer 82*, 973–980.
178. Nooijen PT, Manusama ER, Eggermont AM, Schalkwijk L, Stavast J, Marquet RL, de Waal RM, Ruiter DJ (1996). Synergistic effects of TNF-alpha and melphalan in an isolated limb perfusion model of rat sarcoma: A histopathological, immunohistochemical and electron microscopical study. *Br J Cancer 74*, 1908–1915.
179. de Wilt JH, Manusama ER, van Tiel ST, van Ijken MG, ten Hagen TL, Eggermont AM (1999). Prerequisites for effective isolated limb perfusion using tumour necrosis factor alpha and melphalan in rats. *Br J Cancer 80*, 161–166.
180. Palladino MA, Jr, Patton JS, Figari IS, Shalaby MR (1987). Possible relationships between in vivo antitumour activity and toxicity of tumour necrosis factor-alpha. *Ciba Found Symp 131*, 21–38.
181. Kashii Y, Giorda R, Herberman RB, Whiteside TL, Vujanovic NL (1999). Constitutive expression and role of the TNF family ligands in apoptotic killing of tumor cells by human NK cells. *J Immunol 163*, 5358–5366.
182. Hill GR, Teshima T, Rebel VI, Krijanovski OI, Cooke KR, Brinson YS, Ferrara JL (2000). The p55 TNF-alpha receptor plays a critical role in T cell alloreactivity. *J Immunol 164*, 656–663.
183. Prevost-Blondel A, Roth E, Rosenthal FM, Pircher H (2000). Crucial role of TNF-alpha in CD8 T cell-mediated elimination of 3LL-A9 Lewis lung carcinoma cells in vivo. *J Immunol 164*, 3645–3651.
184. Baxevanis CN, Voutsas IF, Tsitsilonis OE, Tsiatas ML, Gritzapis AD, Papamichail M (2000). Compromised antitumor responses in tumor necrosis factor-alpha knockout mice. *Eur J Immunol 30*, 1957–1966.
185. Eggermont AM, de Wilt JH, ten Hagen TL (2003). Current uses of isolated limb perfusion in the clinic and a model system for new strategies. *Lancet Oncol 4*, 429–437.
186. Lejeune FJ, Ruegg C, Lienard D (1998). Clinical applications of TNF-alpha in cancer. *Curr Opin Immunol 10*, 573–580.

187. van Horssen R, Ten Hagen TL, Eggermont AM (2006). TNF-alpha in cancer treatment: Molecular insights, antitumor effects, and clinical utility. *Oncologist 11*, 397–408.
188. Lejeune FJ, Ruegg C (2006). Recombinant human tumor necrosis factor: An efficient agent for cancer treatment. *Bull Cancer 93*, E90–100.
189. Naylor MS, Stamp GW, Foulkes WD, Eccles D, Balkwill FR (1993). Tumor necrosis factor and its receptors in human ovarian cancer. Potential role in disease progression. *J Clin Invest 91*, 2194–2206.
190. Miles DW, Happerfield LC, Naylor MS, Bobrow LG, Rubens RD, Balkwill FR (1994). Expression of tumour necrosis factor (TNF alpha) and its receptors in benign and malignant breast tissue. *Int J Cancer 56*, 777–782.
191. Nakashima J, Tachibana M, Ueno M, Miyajima A, Baba S, Murai M (1998). Association between tumor necrosis factor in serum and cachexia in patients with prostate cancer. *Clin Cancer Res 4*, 1743–1748.
192. Sati HI, Greaves M, Apperley JF, Russell RG, Croucher PI (1999). Expression of interleukin-1beta and tumour necrosis factor-alpha in plasma cells from patients with multiple myeloma. *Br J Haematol 104*, 350–357.
193. Foa R, Massaia M, Cardona S, Tos AG, Bianchi A, Attisano C, Guarini A, di Celle PF, Fierro MT (1990). Production of tumor necrosis factor-alpha by B-cell chronic lymphocytic leukemia cells: A possible regulatory role of TNF in the progression of the disease. *Blood 76*, 393–400.
194. Warzocha K, Salles G, Bienvenu J, Bastion Y, Dumontet C, Renard N, Neidhardt-Berard EM, Coiffier B (1997). Tumor necrosis factor ligand-receptor system can predict treatment outcome in lymphoma patients. *J Clin Oncol 15*, 499–508.
195. Waage A, Espevik T (1994). TNF receptors in chronic lymphocytic leukemia. *Leuk Lymphoma 13*, 41–46.
196. Vinante F, Rigo A, Tecchio C, Morosato L, Nadali G, Ricetti MM, Krampera M, Zanolin E, Locatelli F, Gallati H, Chilosi M, Pizzolo G (1998). Serum levels of p55 and p75 soluble TNF receptors in adult acute leukaemia at diagnosis: Correlation with clinical and biological features and outcome. *Br J Haematol 102*, 1025–1034.
197. Herrmann F, Helfrich SG, Lindemann A, Schleiermacher E, Huber C, Mertelsmann R (1992). Elevated circulating levels of tumor necrosis factor predict unresponsiveness to treatment with interferon alfa-2b in chronic myelogenous leukemia. *J Clin Oncol 10*, 631–634.
198. Hajeer AH, Hutchinson IV (2001). Influence of TNFalpha gene polymorphisms on TNFalpha production and disease. *Hum Immunol 62*, 1191–1199.
199. Warzocha K, Ribeiro P, Bienvenu J, Roy P, Charlot C, Rigal D, Coiffier B, Salles G (1998). Genetic polymorphisms in the tumor necrosis factor locus influence non-Hodgkin's lymphoma outcome. *Blood 91*, 3574–3581.
200. Davies FE, Rollinson SJ, Rawstron AC, Roman E, Richards S, Drayson M, Child JA, Morgan GJ (2000). High-producer haplotypes of tumor necrosis factor alpha and lymphotoxin alpha are associated with an increased risk of myeloma and have an improved progression-free survival after treatment. *J Clin Oncol 18*, 2843–2851.
201. Jeng JE, Tsai JF, Chuang LY, Ho MS, Lin ZY, Hsieh MY, Chen SC, Chuang WL, Wang LY, Yu ML, Dai CY, Chang JG (2007). Tumor necrosis factor-alpha 308.2 polymorphism is associated with advanced hepatic fibrosis and higher risk for hepatocellular carcinoma. *Neoplasia 9*, 987–992.
202. Oh BR, Sasaki M, Perinchery G, Ryu SB, Park YI, Carroll P, Dahiya R (2000). Frequent genotype changes at –308, and 488 regions of the tumor necrosis factor-alpha (TNF-alpha) gene in patients with prostate cancer. *J Urol 163*, 1584–1587.
203. Hajeer AH, Lear JT, Ollier WE, Naves M, Worthington J, Bell DA, Smith AG, Bowers WP, Jones PW, Strange RC, Fryer AA (2000). Preliminary evidence of an association of tumour necrosis factor microsatellites with increased risk of multiple basal cell carcinomas. *Br J Dermatol 142*, 441–445.
204. Maini RN, Taylor PC (2000). Anti-cytokine therapy for rheumatoid arthritis. *Annu Rev Med 51*, 207–229.
205. Komori A, Yatsunami J, Suganuma M, Okabe S, Abe S, Sakai A, Sasaki K, Fujiki H (1993). Tumor necrosis factor acts as a tumor promoter in BALB/3T3 cell transformation. *Cancer Res 53*, 1982–1985.
206. Jaiswal M, LaRusso NF, Burgart LJ, Gores GJ (2000). Inflammatory cytokines induce DNA damage and inhibit DNA repair in cholangiocarcinoma cells by a nitric oxide-dependent mechanism. *Cancer Res 60*, 184–190.
207. Patel HR, Hewer A, Phillips DH, Hayes JD, Wolf CR, Campbell FC (1997). Metabolic competence and susceptibility of intestinal epithelium to genotoxic injury during regeneration. *Carcinogenesis 18*, 2171–2177.
208. Torisu H, Ono M, Kiryu H, Furue M, Ohmoto Y, Nakayama J, Nishioka Y, Sone S, Kuwano M (2000). Macrophage infiltration correlates with tumor stage and angiogenesis in human malignant melanoma: Possible involvement of TNFalpha and IL-1alpha. *Int J Cancer 85*, 182–188.
209. Ryuto M, Ono M, Izumi H, Yoshida S, Weich HA, Kohno K, Kuwano M (1996). Induction of vascular endothelial growth factor by tumor necrosis factor alpha in human glioma cells. Possible roles of SP-1. *J Biol Chem 271*, 28220–28228.
210. Samaniego F, Markham PD, Gendelman R, Gallo RC, Ensoli B (1997). Inflammatory cytokines induce endothelial cells to produce and release basic fibroblast growth factor and to promote Kaposi's sarcoma-like lesions in nude mice. *J Immunol 158*, 1887–1894.
211. Yoshida S, Ono M, Shono T, Izumi H, Ishibashi T, Suzuki H, Kuwano M (1997). Involvement of interleukin-8, vascular endothelial growth factor, and basic fibroblast growth factor in tumor necrosis factor alpha-dependent angiogenesis. *Mol Cell Biol 17*, 4015–4023.
212. Leek RD, Landers R, Fox SB, Ng F, Harris AL, Lewis CE (1998). Association of tumour necrosis factor alpha and

its receptors with thymidine phosphorylase expression in invasive breast carcinoma. *Br J Cancer 77*, 2246–2251.

213. Leber TM, Balkwill FR (1998). Regulation of monocyte MMP-9 production by TNF-alpha and a tumour-derived soluble factor (MMPSF). *Br J Cancer 78*, 724–732.
214. Negus RP, Turner L, Burke F, Balkwill FR (1998). Hypoxia downregulates MCP-1 expression: Implications for macrophage distribution in tumors. *J Leukocyte Biol 63*, 758–765.
215. Gordon HM, Kucera G, Salvo R, Boss JM (1992). Tumor necrosis factor induces genes involved in inflammation, cellular and tissue repair, and metabolism in murine fibroblasts. *J Immunol 148*, 4021–4027.
216. Mauviel A, Heino J, Kahari VM, Hartmann DJ, Loyau G, Pujol JP, Vuorio E (1991). Comparative effects of interleukin-1 and tumor necrosis factor-alpha on collagen production and corresponding procollagen mRNA levels in human dermal fibroblasts. *J Invest Dermatol 96*, 243–249.
217. Battegay EJ, Raines EW, Colbert T, Ross R (1995). TNF-alpha stimulation of fibroblast proliferation. Dependence on platelet-derived growth factor (PDGF) secretion and alteration of PDGF receptor expression. *J Immunol 154*, 6040–6047.
218. Galban S, Fan J, Martindale JL, Cheadle C, Hoffman B, Woods MP, Temeles G, Brieger J, Decker J, Gorospe M (2003). von Hippel-Lindau protein-mediated repression of tumor necrosis factor alpha translation revealed through use of cDNA arrays. *Mol Cell Biol 23*, 2316–2328.
219. Tisdale MJ (2002). Cachexia in cancer patients. *Nat Rev Cancer 2*, 862–871.
220. Laviano A, Meguid MM, Inui A, Muscaritoli M, Rossi-Fanelli F (2005). Therapy insight: Cancer anorexia-cachexia syndrome—when all you can eat is yourself. *Nat Clin Pract Oncol 2*, 158–165.
221. Morley JE, Thomas DR, Wilson MM (2006). Cachexia: Pathophysiology and clinical relevance. *Am J Clin Nutr 83*, 735–743.
222. Falconer JS, Fearon KC, Plester CE, Ross JA, Carter DC (1994). Cytokines, the acute-phase response, and resting energy expenditure in cachectic patients with pancreatic cancer. *Ann Surg 219*, 325–331.
223. McMillan DC, Scott HR, Watson WS, Preston T, Milroy R, McArdle CS (1998). Longitudinal study of body cell mass depletion and the inflammatory response in cancer patients. *Nutr Cancer 31*, 101–105.
224. Falconer JS, Fearon KC, Ross JA, Elton R, Wigmore SJ, Garden OJ, Carter DC (1995). Acute-phase protein response and survival duration of patients with pancreatic cancer. *Cancer 75*, 2077–2082.
225. Collins P, Bing C, McCulloch P, Williams G (2002). Muscle UCP-3 mRNA levels are elevated in weight loss associated with gastrointestinal adenocarcinoma in humans. *Br J Cancer 86*, 372–375.
226. Clapham JC, Arch JR, Chapman H, Haynes A, Lister C, Moore GB, Piercy V, Carter SA, Lehner I, Smith SA, Beeley LJ, Godden RJ, Herrity N, Skehel M, Changani KK, Hockings PD, Reid DG, Squires SM, Hatcher J, Trail B, Latcham J, Rastan S, Harper AJ, Cadenas S, Buckingham JA, Brand MD, Abuin A (2000). Mice overexpressing human uncoupling protein-3 in skeletal muscle are hyperphagic and lean. *Nature 406*, 415–418.
227. Busquets S, Sanchis D, Alvarez B, Ricquier D, Lopez-Soriano FJ, Argiles JM (1998). In the rat, tumor necrosis factor alpha administration results in an increase in both UCP2 and UCP3 mRNAs in skeletal muscle: A possible mechanism for cytokine-induced thermogenesis? *FEBS Lett 440*, 348–350.
228. Lecker SH, Solomon V, Mitch WE, Goldberg AL (1999). Muscle protein breakdown and the critical role of the ubiquitin-proteasome pathway in normal and disease states. *J Nutr 129*, 227S-237S.
229. Camps C, Iranzo V, Bremnes RM, Sirera R (2006). Anorexia-cachexia syndrome in cancer: Implications of the ubiquitin-proteasome pathway. *Support Care Cancer 14*, 1173–1183.
230. Temparis S, Asensi M, Taillandier D, Aurousseau E, Larbaud D, Obled A, Bechet D, Ferrara M, Estrela JM, Attaix D (1994). Increased ATP-ubiquitin-dependent proteolysis in skeletal muscles of tumor-bearing rats. *Cancer Res 54*, 5568–5573.
231. Melstrom LG, Melstrom KA, Jr, Ding XZ, Adrian TE (2007). Mechanisms of skeletal muscle degradation and its therapy in cancer cachexia. *Histol Histopathol 22*, 805–814.
232. Llovera M, Carbo N, Garcia-Martinez C, Costelli P, Tessitore L, Baccino FM, Agell N, Bagby GJ, Lopez-Soriano FJ, Argiles JM (1996). Anti-TNF treatment reverts increased muscle ubiquitin gene expression in tumour-bearing rats. *Biochem Biophys Res Commun 221*, 653–655.
233. Pende A, Musso NR, Vergassola C, Puppo F, Ioverno A, Criscuolo D, Indiveri F, Lotti G (1990). Neuroendocrine effects of interferon alpha 2-a in healthy human subjects. *J Biol Regu Homeost Agents 4*, 67–72.
234. Ryden M, Arvidsson E, Blomqvist L, Perbeck L, Dicker A, Arner P (2004). Targets for TNF-alpha-induced lipolysis in human adipocytes. *Biochem Biophys Res Commun 318*, 168–175.
235. Tsukaguchi K, Yoneda T, Yoshikawa M, Fu A, Tokuyama T, Okamoto Y, Yamamoto C, Takenaka H, Okamura H, Narita N (1996). [Interaction between nutrition and production of IL-1 beta, TNF alpha, and IL-6 by peripheral blood monocytes in patients with lung cancer]. *Nihon Kyobu Shikkan Gakkai Zasshi 34*, 778–784.
236. Kawakami M, Murase T, Ogawa H, Ishibashi S, Mori N, Takaku F, Shibata S (1987). Human recombinant TNF suppresses lipoprotein lipase activity and stimulates lipolysis in 3T3-L1 cells. *J Biochem 101*, 331–338.

237. Grunfeld C, Feingold KR (1996). Regulation of lipid metabolism by cytokines during host defense. *Nutrition 12*, S24–26.
238. Kotler DP (2000). Cachexia. *Ann Intern Med 133*, 622–634.
239. Schwartz MW, Woods SC, Porte D, Jr, Seeley RJ, Baskin DG (2000). Central nervous system control of food intake. *Nature 404*, 661–671.
240. Laviano A, Meguid MM, Rossi-Fanelli F (2003). Cancer anorexia: Clinical implications, pathogenesis, and therapeutic strategies. *Lancet Oncol 4*, 686–694.
241. Torelli GF, Meguid MM, Moldawer LL, Edwards CK 3rd, Kim HJ, Carter JL, Laviano A, Rossi Fanelli F (1999). Use of recombinant human soluble TNF receptor in anorectic tumor-bearing rats. *Am J Physiolo 277*, R850–855.
242. Behan DP, Grigoriadis DE, Lovenberg T, Chalmers D, Heinrichs S, Liaw C, De Souza EB (1996). Neurobiology of corticotropin releasing factor (CRF) receptors and CRF-binding protein: Implications for the treatment of CNS disorders. *Mol Psychiatry 1*, 265–277.
243. Maltoni M, Fabbri L, Nanni O, Scarpi E, Pezzi L, Flamini E, Riccobon A, Derni S, Pallotti G, Amadori D (1997). Serum levels of tumour necrosis factor alpha and other cytokines do not correlate with weight loss and anorexia in cancer patients. *Support Care Cancer 5*, 130–135.
244. Karayiannakis AJ, Syrigos KN, Polychronidis A, Pitiakoudis M, Bounovas A, Simopoulos K (2001). Serum levels of tumor necrosis factor-alpha and nutritional status in pancreatic cancer patients. *Anticancer Res 21*, 1355–1358.
245. Moradi MM, Carson LF, Weinberg B, Haney AF, Twiggs LB, Ramakrishnan S (1993). Serum and ascitic fluid levels of interleukin-1, interleukin-6, and tumor necrosis factor-alpha in patients with ovarian epithelial cancer. *Cancer 72*, 2433–2440.
246. Corey SJ, Minden MD, Barber DL, Kantarjian H, Wang JC, Schimmer AD (2007). Myelodysplastic syndromes: The complexity of stem-cell diseases. *Nat Rev Cancer 7*, 118–129.
247. Disperati P, Ichim CV, Tkachuk D, Chun K, Schuh AC, Wells RA (2006). Progression of myelodysplasia to acute lymphoblastic leukaemia: Implications for disease biology. *Leuk Res 30*, 233–239.
248. Harris NL, Jaffe ES, Diebold J, Flandrin G, Muller-Hermelink HK, Vardiman J, Lister TA, Bloomfield CD (1999). World Health Organization classification of neoplastic diseases of the hematopoietic and lymphoid tissues: report of the Clinical Advisory Committee meeting—Airlie House, Virginia, November 1997. *J Clin Oncol 17*, 3835–3849.
249. Rosenfeld C, List A (2000). A hypothesis for the pathogenesis of myelodysplastic syndromes: Implications for new therapies. *Leukemia 14*, 2–8.
250. Parker JE, Mufti GJ, Rasool F, Mijovic A, Devereux S, Pagliuca A (2000). The role of apoptosis, proliferation, and the Bcl-2-related proteins in the myelodysplastic syndromes and acute myeloid leukemia secondary to MDS. *Blood 96*, 3932–3938.
251. Greenberg PL (1998). Apoptosis and its role in the myelodysplastic syndromes: Implications for disease natural history and treatment. *Leuk Res 22*, 1123–1136.
252. Shimazaki K, Ohshima K, Suzumiya J, Kawasaki C, Kikuchi M (2000). Evaluation of apoptosis as a prognostic factor in myelodysplastic syndromes. *Br J Haematol 110*, 584–590.
253. Shetty V, Hussaini S, Broady-Robinson L, Allampallam K, Mundle S, Borok R, Broderick E, Mazzoran L, Zorat F, Raza A (2000). Intramedullary apoptosis of hematopoietic cells in myelodysplastic syndrome patients can be massive: Apoptotic cells recovered from high-density fraction of bone marrow aspirates. *Blood 96*, 1388–1392.
254. Allampallam K, Shetty V, Mundle S, Dutt D, Kravitz H, Reddy PL, Alvi S, Galili N, Saberwal GS, Anthwal S, Shaikh MW, York A, Raza A (2002). Biological significance of proliferation, apoptosis, cytokines, and monocyte/macrophage cells in bone marrow biopsies of 145 patients with myelodysplastic syndrome. *Int J Hematol 75*, 289–297.
255. Kitagawa M, Saito I, Kuwata T, Yoshida S, Yamaguchi S, Takahashi M, Tanizawa T, Kamiyama R, Hirokawa K (1997). Overexpression of tumor necrosis factor (TNF)-alpha and interferon (IFN)-gamma by bone marrow cells from patients with myelodysplastic syndromes. *Leukemia 11*, 2049–2054.
256. Gersuk GM, Beckham C, Loken MR, Kiener P, Anderson JE, Farrand A, Troutt AB, Ledbetter JA, Deeg HJ (1998). A role for tumour necrosis factor-alpha, Fas and Fas-Ligand in marrow failure associated with myelodysplastic syndrome. *Br J Haematol 103*, 176–188.
257. Mundle SD, Mativi BY, Bagai K, Feldman G, Cheema P, Gautam U, Reza S, Cartlidge JD, Venugopal P, Shetty V, Gregory SA, Robin E, Rifkin S, Shah R, Raza A (1999). Spontaneous down-regulation of Fas-associated phosphatase-1 may contribute to excessive apoptosis in myelodysplastic marrows. *Int J Hematol 70*, 83–90.
258. Mundle SD, Reza S, Ali A, Mativi Y, Shetty V, Venugopal P, Gregory SA, Raza A (1999). Correlation of tumor necrosis factor alpha (TNF alpha) with high Caspase 3-like activity in myelodysplastic syndromes. *Cancer Lett 140*, 201–207.
259. Steensma DP, Bennett JM (2006). The myelodysplastic syndromes: Diagnosis and treatment. *Mayo Clinic Proc 81*, 104–130.
260. Faderl S, Kantarjian HM (2005). Drug insight: Emerging new drugs in the treatment of myelodysplastic syndromes. *Nat Clin Pract Oncol 2*, 348–355.
261. Kuehl WM, Bergsagel PL (2002). Multiple myeloma: Evolving genetic events and host interactions. *Nat Rev Cancer 2*, 175–187.
262. Hideshima T, Anderson KC (2002). Molecular mechanisms of novel therapeutic approaches for multiple myeloma. *Nat Rev Cancer 2*, 927–937.

263. Lichtenstein A, Berenson J, Norman D, Chang MP, Carlile A (1989). Production of cytokines by bone marrow cells obtained from patients with multiple myeloma. *Blood 74*, 1266–1273.

264. Filella X, Blade J, Guillermo AL, Molina R, Rozman C, Ballesta AM (1996). Cytokines (IL-6, TNF-alpha, IL-1alpha) and soluble interleukin-2 receptor as serum tumor markers in multiple myeloma. *Cancer Detect Prev 20*, 52–56.

265. Jourdan M, Tarte K, Legouffe E, Brochier J, Rossi JF, Klein B (1999). Tumor necrosis factor is a survival and proliferation factor for human myeloma cells. *Eur Cytokine Netw 10*, 65–70.

266. Hideshima T, Chauhan D, Schlossman R, Richardson P, Anderson KC (2001). The role of tumor necrosis factor alpha in the pathophysiology of human multiple myeloma: Therapeutic applications. *Oncogene 20*, 4519–4527.

267. Uchiyama H, Barut BA, Mohrbacher AF, Chauhan D, Anderson KC (1993). Adhesion of human myeloma-derived cell lines to bone marrow stromal cells stimulates interleukin-6 secretion. *Blood 82*, 3712–3720.

268. Chauhan D, Uchiyama H, Akbarali Y, Urashima M, Yamamoto K, Libermann TA, Anderson KC (1996). Multiple myeloma cell adhesion-induced interleukin-6 expression in bone marrow stromal cells involves activation of NF-kappa B. *Blood 87*, 1104–1112.

269. Dankbar B, Padro T, Leo R, Feldmann B, Kropff M, Mesters RM, Serve H, Berdel WE, Kienast J (2000). Vascular endothelial growth factor and interleukin-6 in paracrine tumor-stromal cell interactions in multiple myeloma. *Blood 95*, 2630–2636.

270. Gupta D, Treon SP, Shima Y, Hideshima T, Podar K, Tai YT, Lin B, Lentzsch S, Davies FE, Chauhan D, Schlossman RL, Richardson P, Ralph P, Wu L, Payvandi F, Muller G, Stirling DI, Anderson KC (2001). Adherence of multiple myeloma cells to bone marrow stromal cells upregulates vascular endothelial growth factor secretion: Therapeutic applications. *Leukemia 15*, 1950–1961.

271. Damiano JS, Cress AE, Hazlehurst LA, Shtil AA, Dalton WS (1999). Cell adhesion mediated drug resistance (CAM-DR): Role of integrins and resistance to apoptosis in human myeloma cell lines. *Blood 93*, 1658–1667.

272. Gao XZ, Bi S, Copra H, Devemy E, Venugopal P, Li B, Hsu WT, Loew J, Galvez A, Gregory S, Yang J, Horvath E, Preisler HD (1998). Cytokine gene activity in AML cells in vivo in patients. *Leuk Res 22*, 429–438.

273. Kurzrock R, Kantarjian H, Wetzler M, Estrov Z, Estey E, Troutman-Worden K, Gutterman JU, Talpaz M (1993). Ubiquitous expression of cytokines in diverse leukemias of lymphoid and myeloid lineage. *Exp Hematol 21*, 80–85.

274. Barak V, Nisman B, Polliack A (1999). The tumor necrosis factor family and and correlation with disease activity and response to treatment in hairy cell leukemia. *Eur J Haematol 62*, 71–75.

275. Sivaraman S, Deshpande CG, Ranganathan R, Huang X, Jajeh A, O'Brien T, Huang RW, Gregory SA, Venugopal P, Preisler HD (2000). Tumor necrosis factor modulates CD 20 expression on cells from chronic lymphocytic leukemia: A new role for TNF alpha? *Microsc Res Tech 50*, 251–257.

276. Zapata JM, Krajewska M, Krajewski S, Kitada S, Welsh K, Monks A, McCloskey N, Gordon J, Kipps TJ, Gascoyne RD, Shabaik A, Reed JC (2000). TNFR-associated factor family protein expression in normal tissues and lymphoid malignancies. *J Immunol 165*, 5084–5096.

277. Ulich TR, Guo KZ, Remick D, del Castillo J, Yin SM (1991). Endotoxin-induced cytokine gene expression in vivo. III. IL-6 mRNA and serum protein expression and the in vivo hematologic effects of IL-6. *J Immunol 146*, 2316–2323.

278. Sime PJ, Marr RA, Gauldie D, Xing Z, Hewlett BR, Graham FL, Gauldie J (1998). Transfer of tumor necrosis factor-alpha to rat lung induces severe pulmonary inflammation and patchy interstitial fibrogenesis with induction of transforming growth factor-beta1 and myofibroblasts. *Am J Pathol 153*, 825–832.

279. Warshamana GS, Corti M, Brody AR (2001). TNF-alpha, PDGF, and TGF-beta(1) expression by primary mouse bronchiolar-alveolar epithelial and mesenchymal cells: Tnf-alpha induces TGF-beta(1). *Exp Mol Pathol 71*, 13–33.

280. Sappino AP, Seelentag W, Pelte MF, Alberto P, Vassalli P (1990). Tumor necrosis factor/cachectin and lymphotoxin gene expression in lymph nodes from lymphoma patients. *Blood 75*, 958–962.

281. Keane J, Gershon S, Wise RP, Mirabile-Levens E, Kasznica J, Schwieterman WD, Siegel JN, Braun MM (2001). Tuberculosis associated with infliximab, a tumor necrosis factor alpha-neutralizing agent. *N Engl J Med 345*, 1098–1104.

282. Wolfe F, Michaud K, Anderson J, Urbansky K (2004). Tuberculosis infection in patients with rheumatoid arthritis and the effect of infliximab therapy. *Arthritis Rheum 50*, 372–379.

283. Baecklund E, Ekbom A, Sparen P, Feltelius N, Klareskog L (1998). Disease activity and risk of lymphoma in patients with rheumatoid arthritis: Nested case-control study. *BMJ 317*, 180–181.

284. Ekstrom K, Hjalgrim H, Brandt L, Baecklund E, Klareskog L, Ekbom A, Askling J (2003). Risk of malignant lymphomas in patients with rheumatoid arthritis and in their first-degree relatives. *Arthritis Rheum 48*, 963–970.

285. Baecklund E, Sundstrom C, Ekbom A, Catrina AI, Biberfeld P, Feltelius N, Klareskog L (2003). Lymphoma subtypes in patients with rheumatoid arthritis: Increased proportion of diffuse large B cell lymphoma. *Arthritis Rheum 48*, 1543–1550.

286. Wolfe F, Michaud K (2004). Lymphoma in rheumatoid arthritis: The effect of methotrexate and anti-tumor

necrosis factor therapy in 18,572 patients. *Arthritis Rheum 50*, 1740–1751.

287. Knight DM, Trinh H, Le J, Siegel S, Shealy D, McDonough M, Scallon B, Moore MA, Vilcek J, Daddona P, et al. (1993). Construction and initial characterization of a mouse-human chimeric anti-TNF antibody. *Mol Immunol 30*, 1443–1453.
288. Scallon BJ, Moore MA, Trinh H, Knight DM, Ghrayeb J (1995). Chimeric anti-TNF-alpha monoclonal antibody cA2 binds recombinant transmembrane TNF-alpha and activates immune effector functions. *Cytokine 7*, 251–259.
289. Sherry BA, Gelin J, Fong Y, Marano M, Wei H, Cerami A, Lowry SF, Lundholm KG, Moldawer LL (1989). Anticachectin/tumor necrosis factor-alpha antibodies attenuate development of cachexia in tumor models. *FASEB J 3*, 1956–1962.
290. Siegel SA, Shealy DJ, Nakada MT, Le J, Woulfe DS, Probert L, Kollias G, Ghrayeb J, Vilcek J, Daddona PE (1995). The mouse/human chimeric monoclonal antibody cA2 neutralizes TNF in vitro and protects transgenic mice from cachexia and TNF lethality in vivo. *Cytokine 7*, 15–25.
291. Jatoi A, Jett JR, Sloan J, Novotny P, Ford J, Prabhakar U, Loprinzi CL (2004). A pilot study on safety and pharmacokinetics of infliximab for the cancer anorexia/weight loss syndrome in non-small-cell lung cancer patients. *Support Care Cancer 12*, 859–863.
292. Raza A (2000). Anti-TNF therapies in rheumatoid arthritis, Crohn's disease, sepsis, and myelodysplastic syndromes. *Microsc Res Tech 50*, 229–235.
293. Stasi R, Amadori S (2002). Infliximab chimaeric anti-tumour necrosis factor alpha monoclonal antibody treatment for patients with myelodysplastic syndromes. *Br J Haematol 116*, 334–337.
294. Raza A, Candoni A, Khan U, Lisak L, Tahir S, Silvestri F, Billmeier J, Alvi MI, Mumtaz M, Gezer S, Venugopal P, Reddy P, Galili N (2004). Remicade as TNF suppressor in patients with myelodysplastic syndromes. *Leuk Lymphoma 45*, 2099–2104.
295. Mohler KM, Torrance DS, Smith CA, Goodwin RG, Stremler KE, Fung VP, Madani H, Widmer MB (1993). Soluble tumor necrosis factor (TNF) receptors are effective therapeutic agents in lethal endotoxemia and function simultaneously as both TNF carriers and TNF antagonists. *J Immunol 151*, 1548–1561.
296. Gardam MA, Keystone EC, Menzies R, Manners S, Skamene E, Long R, Vinh DC (2003). Anti-tumour necrosis factor agents and tuberculosis risk: Mechanisms of action and clinical management. *Lancet Infect Dis 3*, 148–155.
297. Scallon BJ, Trinh H, Nedelman M, Brennan FM, Feldmann M, Ghrayeb J (1995). Functional comparisons of different tumour necrosis factor receptor/IgG fusion proteins. *Cytokine 7*, 759–770.
298. Madhusudan S, Muthuramalingam SR, Braybrooke JP, Wilner S, Kaur K, Han C, Hoare S, Balkwill F, Ganesan TS (2005). Study of etanercept, a tumor necrosis factor-alpha inhibitor, in recurrent ovarian cancer. *J Clin Oncol 23*, 5950–5959.
299. Madhusudan S, Foster M, Muthuramalingam SR, Braybrooke JP, Wilner S, Kaur K, Han C, Hoare S, Balkwill F, Talbot DC, Ganesan TS, Harris AL (2004). A phase II study of etanercept (Enbrel), a tumor necrosis factor alpha inhibitor in patients with metastatic breast cancer. *Clin Cancer Res 10*, 6528–6534.
300. Eckel F, Schneider G, Schmid RM (2006). Pancreatic cancer: A review of recent advances. *Expert Opin Investig Drugs 15*, 1395–1410.
301. Tsimberidou AM, Waddelow T, Kantarjian HM, Albitar M, Giles FJ (2003). Pilot study of recombinant human soluble tumor necrosis factor (TNF) receptor (p75) fusion protein (TNFR:Fc; Enbrel) in patients with refractory multiple myeloma: Increase in plasma TNF alpha levels during treatment. *Leuk Res 27*, 375–380.
302. Deeg HJ, Gotlib J, Beckham C, Dugan K, Holmberg L, Schubert M, Appelbaum F, Greenberg P (2002). Soluble TNF receptor fusion protein (etanercept) for the treatment of myelodysplastic syndrome: A pilot study. *Leukemia 16*, 162–164.
303. Steensma DP, Mesa RA, Li CY, Gray L, Tefferi A (2002). Etanercept, a soluble tumor necrosis factor receptor, palliates constitutional symptoms in patients with myelofibrosis with myeloid metaplasia: Results of a pilot study. *Blood 99*, 2252–2254.
304. Tsimberidou AM, Giles FJ (2002). TNF-alpha targeted therapeutic approaches in patients with hematologic malignancies. *Expert Rev Anticancer Ther 2*, 277–286.
305. Kast RE (2005). Evidence of a mechanism by which etanercept increased TNF-alpha in multiple myeloma: New insights into the biology of TNF-alpha giving new treatment opportunities—the role of bupropion. *Leuk Res 29*, 1459–1463.
306. McBride WG (1977). Thalidomide embryopathy. *Teratology 16*, 79–82.
307. Lenz W (1988). A short history of thalidomide embryopathy. *Teratology 38*, 203–215.
308. Sheskin J (1965). Thalidomide in the treatment of lepra reactions. *Clin Pharmacol Ther 6*, 303–306.
309. Hales BF (1999). Thalidomide on the comeback trail. *Nat Med 5*, 489–490.
310. Sampaio EP, Kaplan G, Miranda A, Nery JA, Miguel CP, Viana SM, Sarno EN (1993). The influence of thalidomide on the clinical and immunologic manifestation of erythema nodosum leprosum. *J Infect Dis 168*, 408–414.
311. D'Amato RJ, Loughnan MS, Flynn E, Folkman J (1994). Thalidomide is an inhibitor of angiogenesis. *Proc Nat Acad Sci U S A 91*, 4082–4085.
312. Kumar S, Witzig TE, Dispenzieri A, Lacy MQ, Wellik LE, Fonseca R, Lust JA, Gertz MA, Kyle RA, Greipp PR, Rajkumar SV (2004). Effect of thalidomide therapy on bone marrow angiogenesis in multiple myeloma. *Leukemia 18*, 624–627.

313. Singhal S, Mehta J, Desikan R, Ayers D, Roberson P, Eddlemon P, Munshi N, Anaissie E, Wilson C, Dhodapkar M, Zeddis J, Barlogie B (1999). Antitumor activity of thalidomide in refractory multiple myeloma. *N Engl J Med 341*, 1565–1571.
314. Juliusson G, Celsing F, Turesson I, Lenhoff S, Adriansson M, Malm C (2000). Frequent good partial remissions from thalidomide including best response ever in patients with advanced refractory and relapsed myeloma. *Br J Haematol 109*, 89–96.
315. Rajkumar SV, Gertz MA, Lacy MQ, Dispenzieri A, Fonseca R, Geyer SM, Iturria N, Kumar S, Lust JA, Kyle RA, Greipp PR, Witzig TE (2003). Thalidomide as initial therapy for early-stage myeloma. *Leukemia 17*, 775–779.
316. Weber D (2003). Thalidomide and its derivatives: New promise for multiple myeloma. *Cancer Control 10*, 375–383.
317. Weber D, Rankin K, Gavino M, Delasalle K, Alexanian R (2003). Thalidomide alone or with dexamethasone for previously untreated multiple myeloma. *J Clin Oncol 21*, 16–19.
318. Hideshima T, Chauhan D, Shima Y, Raje N, Davies FE, Tai YT, Treon SP, Lin B, Schlossman RL, Richardson P, Muller G, Stirling DI, Anderson KC (2000). Thalidomide and its analogs overcome drug resistance of human multiple myeloma cells to conventional therapy. *Blood 96*, 2943–2950.
319. Dimopoulos MA, Eleutherakis-Papaiakovou V (2004). Adverse effects of thalidomide administration in patients with neoplastic diseases. *Am J Med 117*, 508–515.
320. Gordon JN, Trebble TM, Ellis RD, Duncan HD, Johns T, Goggin PM (2005). Thalidomide in the treatment of cancer cachexia: A randomised placebo controlled trial. *Gut 54*, 540–545.
321. Tramontana JM, Utaipat U, Molloy A, Akarasewi P, Burroughs M, Makonkawkeyoon S, Johnson B, Klausner JD, Rom W, Kaplan G (1995). Thalidomide treatment reduces tumor necrosis factor alpha production and enhances weight gain in patients with pulmonary tuberculosis. *Mol Med 1*, 384–397.
322. Kaplan G, Thomas S, Fierer DS, Mulligan K, Haslett PA, Fessel WJ, Smith LG, Kook KA, Stirling D, Schambelan M (2000). Thalidomide for the treatment of AIDS-associated wasting. *AIDS Res Hum Retroviruses 16*, 1345–1355.
323. Klausner JD, Makonkawkeyoon S, Akarasewi P, Nakata K, Kasinrerk W, Corral L, Dewar RL, Lane HC, Freedman VH, Kaplan G (1996). The effect of thalidomide on the pathogenesis of human immunodeficiency virus type 1 and M. tuberculosis infection. *J Acquir Immune Defic Syndr Hum Retrovirol 11*, 247–257.
324. Little RF, Wyvill KM, Pluda JM, Welles L, Marshall V, Figg WD, Newcomb FM, Tosato G, Feigal E, Steinberg SM, Whitby D, Goedert JJ, Yarchoan R (2000). Activity of thalidomide in AIDS-related Kaposi's sarcoma. *J Clin Oncol 18*, 2593–2602.
325. Ramirez-Amador VA, Esquivel-Pedraza L, Ponce-de-Leon S, Reyes-Teran G, Gonzalez-Guevara M, Ponce-de-Leon S, Sierra-Madero JG (1999). Thalidomide as therapy for human immunodeficiency virus-related oral ulcers: A double-blind placebo-controlled clinical trial. *Clin Infect Dis 28*, 892–894.
326. Gutierrez-Rodriguez O (1984). Thalidomide. A promising new treatment for rheumatoid arthritis. *Arthritis Rheum 27*, 1118–1121.
327. Atra E, Sato EI (1993). Treatment of the cutaneous lesions of systemic lupus erythematosus with thalidomide. *Clin Exp Rheumatol 11*, 487–493.
328. Hamza MH (1986). Treatment of Behcet's disease with thalidomide. *Clin Rheumatol 5*, 365–371.
329. McCarthy DM, Kanfer EJ, Barrett AJ (1989). Thalidomide for the therapy of graft-versus-host disease following allogeneic bone marrow transplantation. *Biomed Pharmacother 43*, 693–697.
330. Agoston I, Dibbs ZI, Wang F, Muller G, Zeldis JB, Mann DL, Bozkurt B (2002). Preclinical and clinical assessment of the safety and potential efficacy of thalidomide in heart failure. *J Card Fail 8*, 306–314.
331. Franks ME, Macpherson GR, Figg WD (2004). Thalidomide. *Lancet 363*, 1802–1811.
332. Wnendt S, Finkam M, Winter W, Ossig J, Raabe G, Zwingenberger K (1996). Enantioselective inhibition of TNF-alpha release by thalidomide and thalidomide-analogues. *Chirality 8*, 390–396.
333. Frederickson RC, Slater IH, Dusenberry WE, Hewes CR, Jones GT, Moore RA (1977). A comparison of thalidomide and pentobarbital—new methods for identifying novel hypnotic drugs. *J Pharmacol Exp Ther 203*, 240–251.
334. Kenyon BM, Browne F, D'Amato RJ (1997). Effects of thalidomide and related metabolites in a mouse corneal model of neovascularization. *Exp Eye Res 64*, 971–978.
335. Hashimoto Y (2002). Structural development of biological response modifiers based on thalidomide. *Bioorg Med Chem 10*, 461–479.
336. Noguchi T, Fujimoto H, Sano H, Miyajima A, Miyachi H, Hashimoto Y (2005). Angiogenesis inhibitors derived from thalidomide. *Bioorg Med Chem Lett 15*, 5509–5513.
337. Kumar S, Anderson KC (2005). Drug insight: Thalidomide as a treatment for multiple myeloma. *Nat Clin Pract Oncol 2*, 262–270.
338. Lokensgard JR, Hu S, van Fenema EM, Sheng WS, Peterson PK (2000). Effect of thalidomide on chemokine production by human microglia. *J Inf Dis 182*, 983–987.
339. Deng L, Ding W, Granstein RD (2003). Thalidomide inhibits tumor necrosis factor-alpha production and antigen presentation by Langerhans cells. *J Invest Dermatol 121*, 1060–1065.
340. Singhal S, Mehta J (2001). Thalidomide in cancer: Potential uses and limitations. *BioDrugs 15*, 163–172.

341. Moller DR, Wysocka M, Greenlee BM, Ma X, Wahl L, Flockhart DA, Trinchieri G, Karp CL (1997). Inhibition of IL-12 production by thalidomide. *J Immunol 159*, 5157–5161.
342. Rowland TL, McHugh SM, Deighton J, Dearman RJ, Ewan PW, Kimber I (1998). Differential regulation by thalidomide and dexamethasone of cytokine expression in human peripheral blood mononuclear cells. *Immunopharmacology 40*, 11–20.
343. Stephens TD, Bunde CJ, Fillmore BJ (2000). Mechanism of action in thalidomide teratogenesis. *Biochem Pharmacol 59*, 1489–1499.
344. Majumdar S, Lamothe B, Aggarwal BB (2002). Thalidomide suppresses NF-kappa B activation induced by TNF and H2O2, but not that activated by ceramide, lipopolysaccharides, or phorbol ester. *J Immunol 168*, 2644–2651.
345. Keifer JA, Guttridge DC, Ashburner BP, Baldwin AS, Jr (2001). Inhibition of NF-kappa B activity by thalidomide through suppression of IkappaB kinase activity. *J Biol Chem 276*, 22382–22387.
346. Mitsiades N, Mitsiades CS, Poulaki V, Chauhan D, Richardson PG, Hideshima T, Munshi NC, Treon SP, Anderson KC (2002). Apoptotic signaling induced by immunomodulatory thalidomide analogs in human multiple myeloma cells: Therapeutic implications. *Blood 99*, 4525–4530.
347. Karin M, Yamamoto Y, Wang QM (2004). The IKK NF-kappa B system: A treasure trove for drug development. *Nat Rev Drug Discov 3*, 17–26.
348. Yabu T, Tomimoto H, Taguchi Y, Yamaoka S, Igarashi Y, Okazaki T (2005). Thalidomide-induced antiangiogenic action is mediated by ceramide through depletion of VEGF receptors, and is antagonized by sphingosine-1-phosphate. *Blood 106*, 125–134.
349. Yamada M, Kawai M, Kawai Y, Mashima Y (1999). The effect of selective cyclooxygenase-2 inhibitor on corneal angiogenesis in the rat. *Curr Eye Res 19*, 300–304.
350. Fujita J, Mestre JR, Zeldis JB, Subbaramaiah K, Dannenberg AJ (2001). Thalidomide and its analogues inhibit lipopolysaccharide-mediated Iinduction of cyclooxygenase-2. *Clin Cancer Res 7*, 3349–3355.
351. Daniel TO, Liu H, Morrow JD, Crews BC, Marnett LJ (1999). Thromboxane A2 is a mediator of cyclooxygenase-2-dependent endothelial migration and angiogenesis. *Cancer Res 59*, 4574–4577.
352. Tabin CJ (1998). A developmental model for thalidomide defects. *Nature 396*, 322–323.
353. Parman T, Wiley MJ, Wells PG (1999). Free radical-mediated oxidative DNA damage in the mechanism of thalidomide teratogenicity. *Nat Med 5*, 582–585.
354. Knobloch J, Schmitz I, Gotz K, Schulze-Osthoff K, Ruther U (2007). Thalidomide induces limb anomalies by PTEN stabilization, Akt suppression and stimulation of caspase-dependent cell death. *Mol Cell Biol 16*, 2089–2096.
355. Knobloch J, Shaughnessy JD, Jr, Ruther U (2007). Thalidomide induces limb deformities by perturbing the Bmp/Dkk1/Wnt signaling pathway. *FASEB J 21*, 1410–1421.
356. Weinberg RA (2007). *The Biol of Cancer*. City: Garland Science, Taylor & Francis Group, Boca Raton, FL.
357. Haslett PA, Corral LG, Albert M, Kaplan G (1998). Thalidomide costimulates primary human T lymphocytes, preferentially inducing proliferation, cytokine production, and cytotoxic responses in the CD8+ subset. *J Exp Med 187*, 1885–1892.
358. Wolkenstein P, Latarjet J, Roujeau JC, Duguet C, Boudeau S, Vaillant L, Maignan M, Schuhmacher MH, Milpied B, Pilorget A, Bocquet H, Brun-Buisson C, Revuz J (1998). Randomised comparison of thalidomide versus placebo in toxic epidermal necrolysis. *Lancet 352*, 1586–1589.
359. Teo SK, Stirling DI, Zeldis JB (2005). Thalidomide as a novel therapeutic agent: New uses for an old product. *Drug Discov Today 10*, 107–114.
360. Jacobson JM, Greenspan JS, Spritzler J, Ketter N, Fahey JL, Jackson JB, Fox L, Chernoff M, Wu AW, MacPhail LA, Vasquez GJ, Wohl DA (1997). Thalidomide for the treatment of oral aphthous ulcers in patients with human immunodeficiency virus infection. National Institute of Allergy and Infectious Diseases AIDS Clinical Trials Group. *N Engl J Med 336*, 1487–1493.
361. Meierhofer C, Dunzendorfer S, Wiedermann CJ (2001). Theoretical basis for the activity of thalidomide. *BioDrugs 15*, 681–703.
362. Lepper ER, Smith NF, Cox MC, Scripture CD, Figg WD (2006). Thalidomide metabolism and hydrolysis: Mechanisms and implications. *Curr Drug Metab 7*, 677–685.
363. Huupponen R, Pyykko K (1995). Stability of thalidomide in human plasma. *Clin Chem 41*, 1199.
364. Eriksson T, Bjorkman S, Hoglund P (2001). Clinical pharmacology of thalidomide. *Eur J Clin Pharmacol 57*, 365–376.
365. Schumacher H, Smith RL, Williams RT (1965). The metabolism of thalidomide: The spontaneous hydrolysis of thalidomide in solution. *Br J Pharmacol Chemother 25*, 324–337.
366. Eriksson T, Bjorkman S, Roth B, Fyge A, Hoglund P (1998). Enantiomers of thalidomide: Blood distribution and the influence of serum albumin on chiral inversion and hydrolysis. *Chirality 10*, 223–228.
367. Lu J, Palmer BD, Kestell P, Browett P, Baguley BC, Muller G, Ching LM (2003). Thalidomide metabolites in mice and patients with multiple myeloma. *Clin Cancer Res 9*, 1680–1688.
368. Bauer KS, Dixon SC, Figg WD (1998). Inhibition of angiogenesis by thalidomide requires metabolic activation, which is species-dependent. *Biochem Pharmacol 55*, 1827–1834.
369. Ando Y, Fuse E, Figg WD (2002). Thalidomide metabolism by the CYP2C subfamily. *Clin Cancer Res 8*, 1964–1973.

370. Price DK, Ando Y, Kruger EA, Weiss M, Figg WD (2002). 5'-OH-thalidomide, a metabolite of thalidomide, inhibits angiogenesis. *Ther Drug Monit 24*, 104–110.
371. Ingelman-Sundberg M (2005). The human genome project and novel aspects of cytochrome P450 research. *Toxicol Appl Pharmacol 207*, 52–56.
372. Ando Y, Price DK, Dahut WL, Cox MC, Reed E, Figg WD (2002). Pharmacogenetic associations of CYP2C19 genotype with in vivo metabolisms and pharmacological effects of thalidomide. *Cancer Biol Ther 1*, 669–673.
373. Bennett CL, Schumock GT, Desai AA, Kwaan HC, Raisch DW, Newlin R, Stadler W (2002). Thalidomide-associated deep vein thrombosis and pulmonary embolism. *Am J Med 113*, 603–606.
374. Cavo M, Zamagni E, Cellini C, Tosi P, Cangini D, Cini M, Valdre L, Palareti G, Masini L, Tura S, Baccarani M (2002). Deep-vein thrombosis in patients with multiple myeloma receiving first-line thalidomide-dexamethasone therapy. *Blood 100*, 2272–2273.
375. Osman K, Comenzo R, Rajkumar SV (2001). Deep venous thrombosis and thalidomide therapy for multiple myeloma. *N Engl J Med 344*, 1951–1952.
376. Desai AA, Vogelzang NJ, Rini BI, Ansari R, Krauss S, Stadler WM (2002). A high rate of venous thromboembolism in a multi-institutional phase II trial of weekly intravenous gemcitabine with continuous infusion fluorouracil and daily thalidomide in patients with metastatic renal cell carcinoma. *Cancer 95*, 1629–1636.
377. Wang J, Weiss I, Svoboda K, Kwaan HC (2001). Thrombogenic role of cells undergoing apoptosis. *Br J Haematol 115*, 382–391.
378. Kaushal V, Kohli M, Zangari M, Fink L, Mehta P (2002). Endothelial dysfunction in antiangiogenesis-associated thrombosis. *J Clin Oncol 20*, 3042; author reply 3042–3043.
379. Teo S, Noormohamed F, Youle M, Johnson M, Peters B, Stirling D, Thomas S (2002). Transient increase in plasma HIV-1 viral load and associated weight gain after thalidomide dosing. *AIDS 16*, 2355–2356.
380. Aragon-Ching JB, Li H, Gardner ER, Figg WD (2007). Thalidomide analogues as anticancer drugs. *Recent Patents Anticancer Drug Discov 2*, 167–174.
381. Muller GW, Corral LG, Shire MG, Wang H, Moreira A, Kaplan G, Stirling DI (1996). Structural modifications of thalidomide produce analogs with enhanced tumor necrosis factor inhibitory activity. *J Med Chem 39*, 3238–3240.
382. Marriott JB, Westby M, Cookson S, Guckian M, Goodbourn S, Muller G, Shire MG, Stirling D, Dalgleish AG (1998). CC-3052: A water-soluble analog of thalidomide and potent inhibitor of activation-induced TNF-alpha production. *J Immunol 161*, 4236–4243.
383. Muller GW, Chen R, Huang SY, Corral LG, Wong LM, Patterson RT, Chen Y, Kaplan G, Stirling DI (1999). Amino-substituted thalidomide analogs: Potent inhibitors of TNF-alpha production. *Bioorg Med Chem Lett 9*, 1625–1630.
384. de Almeida MV, Teixeira FM, de Souza MV, Amarante GW, Alves CC, Cardoso SH, Mattos AM, Ferreira AP, Teixeira HC (2007). Thalidomide analogs from diamines: Synthesis and evaluation as inhibitors of TNF-alpha production. *Chem Pharm Bull 55*, 223–226.
385. Zhu X, Giordano T, Yu QS, Holloway HW, Perry TA, Lahiri DK, Brossi A, Greig NH (2003). Thiothalidomides: Novel isosteric analogues of thalidomide with enhanced TNF-alpha inhibitory activity. *J Med Chem 46*, 5222–5229.
386. Dredge K, Marriott JB, Macdonald CD, Man HW, Chen R, Muller GW, Stirling D, Dalgleish AG (2002). Novel thalidomide analogues display anti-angiogenic activity independently of immunomodulatory effects. *Br J Cancer 87*, 1166–1172.
387. Marriott JB, Clarke IA, Dredge K, Muller G, Stirling D, Dalgleish AG (2002). Thalidomide and its analogues have distinct and opposing effects on TNF-alpha and TNFR2 during co-stimulation of both CD4(+) and CD8(+) T cells. *Clin Exp Immunol 130*, 75–84.
388. Dredge K, Marriott JB, Todryk SM, Muller GW, Chen R, Stirling DI, Dalgleish AG (2002). Protective antitumor immunity induced by a costimulatory thalidomide analog in conjunction with whole tumor cell vaccination is mediated by increased Th1-type immunity. *J Immunol 168*, 4914–4919.
389. Lentzsch S, Rogers MS, LeBlanc R, Birsner AE, Shah JH, Treston AM, Anderson KC, D'Amato RJ (2002). S-3-Amino-phthalimido-glutarimide inhibits angiogenesis and growth of B-cell neoplasias in mice. *Cancer Res 62*, 2300–2305.
390. Muller GW, Shire MG, Wong LM, Corral LG, Patterson RT, Chen Y, Stirling DI (1998). Thalidomide analogs and PDE4 inhibition. *Bioorg Med Chem Lett 8*, 2669–2674.
391. Torphy TJ (1998). Phosphodiesterase isozymes: Molecular targets for novel antiasthma agents. *Am J Respir Criti Care Med 157*, 351–370.
392. Burnouf C, Pruniaux MP (2002). Recent advances in PDE4 inhibitors as immunoregulators and anti-inflammatory drugs. *Curr Pharm Des 8*, 1255–1296.
393. Barnette MS (1999). Phosphodiesterase 4 (PDE4) inhibitors in asthma and chronic obstructive pulmonary disease (COPD). *Prog Drug Res 53*, 193–229.
394. Giembycz MA (2002). Development status of second generation PDE4 inhibitors for asthma and COPD: The story so far. *Monaldi Arch Chest Dis 57*, 48–64.
395. Spina D (2003). Theophylline and PDE4 inhibitors in asthma. *Curr Opin Pulm Med 9*, 57–64.
396. Gee MS, Makonnen S, al-Kofahi K, Roysam B, Payvandi F, Man HW, Muller GW, Lee WM (2003). Selective cytokine inhibitory drugs with enhanced antiangiogenic activity control tumor growth through vascular inhibition. *Cancer Res 63*, 8073–8078.
397. Marriott JB, Clarke IA, Czajka A, Dredge K, Childs K, Man HW, Schafer P, Govinda S, Muller GW, Stirling DI, Dalgleish AG (2003). A novel subclass of thalidomide analogue with anti-solid tumor activity in which caspase-

dependent apoptosis is associated with altered expression of bcl-2 family proteins. *Cancer Res 63*, 593–599.

398. Molostvov G, Morris A, Rose P, Basu S, Muller G (2004). The effects of selective cytokine inhibitory drugs (CC-10004 and CC-1088) on VEGF and IL-6 expression and apoptosis in myeloma and endothelial cell co-cultures. *Br J Haematol 124*, 366–375.
399. Hui AM, Zhang W, Chen W, Xi D, Purow B, Friedman GC, Fine HA (2004). Agents with selective estrogen receptor (ER) modulator activity induce apoptosis in vitro and in vivo in ER-negative glioma cells. *Cancer Res 64*, 9115–9123.
400. Kumar S, Boehm J, Lee JC (2003). p38 MAP kinases: Key signalling molecules as therapeutic targets for inflammatory diseases. *Nat Rev Drug Discov 2*, 717–726.
401. Lee MR, Dominguez C (2005). MAP kinase p38 inhibitors: Clinical results and an intimate look at their interactions with p38alpha protein. *Curr Med Chem 12*, 2979–2994.
402. Sweeney SE, Firestein GS (2006). Mitogen activated protein kinase inhibitors: Where are we now and where are we going? *Ann Rheum Dis 65* (Suppl 3), iii83–88.
403. Mavunkel BJ, Chakravarty S, Perumattam JJ, Luedtke GR, Liang X, Lim D, Xu YJ, Laney M, Liu DY, Schreiner GF, Lewicki JA, Dugar S (2003). Indole-based heterocyclic inhibitors of p38alpha MAP kinase: Designing a conformationally restricted analogue. *Bioorg Med Chem Lett 13*, 3087–3090.
404. Nikas SN, Drosos AA (2004). SCIO-469 Scios Inc. *Curr Opin Investig Drugs 5*, 1205–1212.
405. Westra J, Limburg PC (2006). p38 mitogen-activated protein kinase (MAPK) in rheumatoid arthritis. *Mini Rev Med Chem 6*, 867–874.
406. Hwang JJ, Ghobrial IM, Anderson KC (2006). New frontiers in the treatment of multiple myeloma. *ScientificWorldJournal 6*, 1475–1503.
407. Katsoulidis E, Li Y, Yoon P, Sassano A, Altman J, Kannan-Thulasiraman P, Balasubramanian L, Parmar S, Varga J, Tallman MS, Verma A, Platanias LC (2005). Role of the p38 mitogen-activated protein kinase pathway in cytokine-mediated hematopoietic suppression in myelodysplastic syndromes. *Cancer Res 65*, 9029–9037.
408. Hideshima T, Akiyama M, Hayashi T, Richardson P, Schlossman R, Chauhan D, Anderson KC (2003). Targeting p38 MAPK inhibits multiple myeloma cell growth in the bone marrow milieu. *Blood 101*, 703–705.
409. Nguyen AN, Stebbins EG, Henson M, O'Young G, Choi SJ, Quon D, Damm D, Reddy M, Ma JY, Haghnazari E, Kapoun AM, Medicherla S, Protter A, Schreiner GF, Kurihara N, Anderson J, Roodman GD, Navas TA, Higgins LS (2006). Normalizing the bone marrow microenvironment with p38 inhibitor reduces multiple myeloma cell proliferation and adhesion and suppresses osteoclast formation. *Exp Cell Res 312*, 1909–1923.
410. Hideshima T, Podar K, Chauhan D, Ishitsuka K, Mitsiades C, Tai YT, Hamasaki M, Raje N, Hideshima H, Schreiner G, Nguyen AN, Navas T, Munshi NC, Richardson PG, Higgins LS, Anderson KC (2004). p38 MAPK inhibition enhances PS-341 (bortezomib)-induced cytotoxicity against multiple myeloma cells. *Oncogene 23*, 8766–8776.
411. Vanderkerken K, Medicherla S, Coulton L, De Raeve H, Willems A, Lawson M, Van Camp B, Protter AA, Higgins LS, Menu E, Croucher PI (2007). Inhibition of p38alpha mitogen-activated protein kinase prevents the development of osteolytic bone disease, reduces tumor burden, and increases survival in murine models of multiple myeloma. *Cancer Res 67*, 4572–4577.
412. Navas TA, Mohindru M, Estes M, Ma JY, Sokol L, Pahanish P, Parmar S, Haghnazari E, Zhou L, Collins R, Kerr I, Nguyen AN, Xu Y, Platanias LC, List AA, Higgins LS, Verma A (2006). Inhibition of overactivated p38 MAPK can restore hematopoiesis in myelodysplastic syndrome progenitors. *Blood 108*, 4170–4177.
413. Peschon JJ, Slack JL, Reddy P, Stocking KL, Sunnarborg SW, Lee DC, Russell WE, Castner BJ, Johnson RS, Fitzner JN, Boyce RW, Nelson N, Kozlosky CJ, Wolfson MF, Rauch CT, Cerretti DP, Paxton RJ, March CJ, Black RA (1998). An essential role for ectodomain shedding in mammalian development. *Science 282*, 1281–1284.
414. Stocker W, Grams F, Baumann U, Reinemer P, Gomis-Ruth FX, McKay DB, Bode W (1995). The metzincins—topological and sequential relations between the astacins, adamalysins, serralysins, and matrixins (collagenases) define a superfamily of zinc-peptidases. *Protein Sci 4*, 823–840.
415. Brinckerhoff CE, Matrisian LM (2002). Matrix metalloproteinases: A tail of a frog that became a prince. *Nat Rev Mol Cell Biol 3*, 207–214.
416. Egeblad M, Werb Z (2002). New functions for the matrix metalloproteinases in cancer progression. *Nat Rev Cancer 2*, 161–174.
417. Black RA (2002). Tumor necrosis factor-alpha converting enzyme. *Int J Biochem Cell Biol 34*, 1–5.
418. Buxbaum JD, Liu KN, Luo Y, Slack JL, Stocking KL, Peschon JJ, Johnson RS, Castner BJ, Cerretti DP, Black RA (1998). Evidence that tumor necrosis factor alpha converting enzyme is involved in regulated alpha-secretase cleavage of the Alzheimer amyloid protein precursor. *J Biol Chem 273*, 27765–27767.
419. Reddy P, Slack JL, Davis R, Cerretti DP, Kozlosky CJ, Blanton RA, Shows D, Peschon JJ, Black RA (2000). Functional analysis of the domain structure of tumor necrosis factor-alpha converting enzyme. *J Biol Chem 275*, 14608–14614.
420. Doedens JR, Black RA (2000). Stimulation-induced down-regulation of tumor necrosis factor-alpha converting enzyme. *J Biol Chem 275*, 14598–14607.
421. Chirivi RG, Garofalo A, Crimmin MJ, Bawden LJ, Stoppacciaro A, Brown PD, Giavazzi R (1994). Inhibition of the metastatic spread and growth of B16-BL6 murine melanoma by a synthetic matrix metalloproteinase inhibitor. *Int J Cancer 58*, 460–464.

422. Davies B, Brown PD, East N, Crimmin MJ, Balkwill FR (1993). A synthetic matrix metalloproteinase inhibitor decreases tumor burden and prolongs survival of mice bearing human ovarian carcinoma xenografts. *Cancer Res 53*, 2087–2091.

423. Watson SA, Morris TM, Robinson G, Crimmin MJ, Brown PD, Hardcastle JD (1995). Inhibition of organ invasion by the matrix metalloproteinase inhibitor batimastat (BB-94) in two human colon carcinoma metastasis models. *Cancer Res 55*, 3629–3633.

424. Watson SA, Morris TM, Collins HM, Bawden LJ, Hawkins K, Bone EA (1999). Inhibition of tumour growth by marimastat in a human xenograft model of gastric cancer: Relationship with levels of circulating CEA. *Br J Cancer 81*, 19–23.

425. Bramhall SR, Rosemurgy A, Brown PD, Bowry C, Buckels JA (2001). Marimastat as first-line therapy for patients with unresectable pancreatic cancer: A randomized trial. *J Clin Oncol 19*, 3447–3455.

426. Steward WP (1999). Marimastat (BB2516): Current status of development. *Cancer Chemother Pharmacol 43* (Suppl), S56–60.

427. Bramhall SR, Hallissey MT, Whiting J, Scholefield J, Tierney G, Stuart RC, Hawkins RE, McCulloch P, Maughan T, Brown PD, Baillet M, Fielding JW (2002). Marimastat as maintenance therapy for patients with advanced gastric cancer: A randomised trial. *Br J Cancer 86*, 1864–1870.

428. Groves MD, Puduvalli VK, Hess KR, Jaeckle KA, Peterson P, Yung WK, Levin VA (2002). Phase II trial of temozolomide plus the matrix metalloproteinase inhibitor, marimastat, in recurrent and progressive glioblastoma multiforme. *J Clin Oncol 20*, 1383–1388.

429. Fletcher L (2000). MMPI demise spotlights target choice. *Nature Biotech 18*, 1138–1139.

430. Coussens LM, Fingleton B, Matrisian LM (2002). Matrix metalloproteinase inhibitors and cancer: Trials and tribulations. *Science 295*, 2387–2392.

431. Nemunaitis J, Poole C, Primrose J, Rosemurgy A, Malfetano J, Brown P, Berrington A, Cornish A, Lynch K, Rasmussen H, Kerr D, Cox D, Millar A (1998). Combined analysis of studies of the effects of the matrix metalloproteinase inhibitor marimastat on serum tumor markers in advanced cancer: Selection of a biologically active and tolerable dose for longer-term studies. *Clin Cancer Res 4*, 1101–1109.

432. Overall CM, Lopez-Otin C (2002). Strategies for MMP inhibition in cancer: Innovations for the post-trial era. *Nat Rev Cancer 2*, 657–672.

433. Vihinen P, Ala-aho R, Kahari VM (2005). Matrix metalloproteinases as therapeutic targets in cancer. *Curr Cancer Drug Targets 5*, 203–220.

434. Turk B (2006). Targeting proteases: Successes, failures and future prospects. *Nat Rev Drug Discov 5*, 785–799.

435. Lin H, Chen C, Chen BD (2001). Resistance of bone marrow-derived macrophages to apoptosis is associated with the expression of X-linked inhibitor of apoptosis protein in primary cultures of bone marrow cells. *Biochem J 353*, 299–306.

436. Alkalay I, Yaron A, Hatzubai A, Orian A, Ciechanover A, Ben-Neriah Y (1995). Stimulation-dependent I kappa B alpha phosphorylation marks the NF-kappa B inhibitor for degradation via the ubiquitin-proteasome pathway. *Proc Nat Acad Sci U S A 92*, 10599–10603.

437. Barr P, Fisher R, Friedberg J (2007). The role of bortezomib in the treatment of lymphoma. *Cancer Invest 25*, 766–775.

438. Manochakian R, Miller KC, Chanan-Khan AA (2007). Clinical impact of bortezomib in frontline regimens for patients with multiple myeloma. *Oncologist 12*, 978–990.

439. McCarty MF, Block KI (2006). Preadministration of high-dose salicylates, suppressors of NF-kappaB activation, may increase the chemosensitivity of many cancers: An example of proapoptotic signal modulation therapy. *Integr Cancer Ther 5*, 252–268.

440. Yin MJ, Yamamoto Y, Gaynor RB (1998). The anti-inflammatory agents aspirin and salicylate inhibit the activity of I(kappa)B kinase-beta. *Nature 396*, 77–80.

441. Kopp E, Ghosh S (1994). Inhibition of NF-kappa B by sodium salicylate and aspirin. *Science 265*, 956–959.

442. Berman KS, Verma UN, Harburg G, Minna JD, Cobb MH, Gaynor RB (2002). Sulindac enhances tumor necrosis factor-alpha-mediated apoptosis of lung cancer cell lines by inhibition of nuclear factor-kappaB. *Clin Cancer Res 8*, 354–360.

443. Wahl C, Liptay S, Adler G, Schmid RM (1998). Sulfasalazine: A potent and specific inhibitor of nuclear factor kappa B. *J Clin Invest 101*, 1163–1174.

444. Hehner SP, Heinrich M, Bork PM, Vogt M, Ratter F, Lehmann V, Schulze-Osthoff K, Droge W, Schmitz ML (1998). Sesquiterpene lactones specifically inhibit activation of NF-kappa B by preventing the degradation of I kappa B-alpha and I kappa B-beta. *J Biol Chem 273*, 1288–1297.

445. Singh S, Natarajan K, Aggarwal BB (1996). Capsaicin (8-methyl-*N*-vanillyl-6-nonenamide) is a potent inhibitor of nuclear transcription factor-kappa B activation by diverse agents. *J Immunol 157*, 4412–4420.

# DEVELOPMENT OF THERAPEUTIC AGENTS: THE MATERNAL–FETAL PERSPECTIVE

C. Gedeon,[1,3] F. Garcia Bournissen,[1] and G. Koren*[1,2,3]

[1]*Hospital for Sick Children, Toronto, Ontario, Canada*
[2]*Department of Molecular Toxicology, University of Western Ontario, London, Ontario, Canada*
[3]*Faculty of Medicine, Department of Clinical Pharmacology, University of Toronto, Toronto, Ontario, Canada*

## 1 INTRODUCTION

The use of medicinal drugs in pregnancy is not uncommon. Women during pregnancy are affected by the same illnesses as the nonpregnant female population of comparable age and socioeconomical status, as well as from multiple conditions that can arise from, or are worsened by, pregnancy. For example, some pregnant women suffer from medical conditions that require ongoing or episodic drug treatment such as asthma, epilepsy, or hypertension. Many of these conditions may require drug therapy, where withdrawal or a decrease of treatment can lead to a worsening of the underlying condition and, in other cases, to long-term complications or even death of either the mother or her fetus. Moreover, pregnancy can induce conditions such as nausea and vomiting that may need to be treated pharmacologically. Despite the dearth of data on safety and efficacy of most medications during this period, it has been estimated that as much as 60–80% of pregnant women in developed countries receive medications and most of them on more than one occasion [1].

While the woman's well-being is at the heart of any medical treatment, placental transfer of drugs leading to potential toxicity to the fetus is a major concern in the pharmacological management of the pregnant patient. This poses a dilemma to the health provider, who needs to weigh risks and benefits of drug treatment for the pregnant patient, including evaluation of the uncertainties produced by the little available data on fetal safety for most drugs. Hence, when managing a pregnant patient with medication, the exposure of two individual patients, mother and fetus, should be considered independently, and the decision must be based on the risk–benefit assessment of both.

While the use of prescription drugs can sometimes increase the risk of teratogenicity, some medical conditions such as gestational diabetes, hyperthyroidism, or hypertension may require drug therapy in order to ensure optimal health of the mother and fetus. In this setting, a small risk of fetotoxicity (or even a moderate one in certain circumstances such as previously transplanted patients who become pregnant) is offset by the large maternal benefits provided by the drug treatment.

Of the thousands of available drugs, relatively few have been clearly shown to have the potential to adversely affect the fetus when used at the recommended doses. For example, the use of most traditional anticonvulsants, while necessary for the treatment of the mother, may cause structural defects as well as impaired neurocognitive development in a small proportion of exposed babies [2, 3]. Thus, a potential maternal–fetal conflict brings to light the need to identify drugs that

*Supported by a grant by the Canadian Institute for Health Research.

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

**TABLE 1 Changes in Selected Physiological Parameters in Pregnancy**

| Physiology | Change in Pregnancy |
|---|---|
| Cardiac output | Increased |
| Tidal volume | Increased |
| Pulmonary blood flow | Increased |
| Gastric pH | Increased |
| Glomerular filtration rate | Increased |
| Renal drug elimination | Increased |
| Protein binding capacity | Decreased |
| Clearance | Increased |
| Total body water | Increased |
| Volume of distribution | Increased |
| Steady-state plasma concentration | Decreased |
| Peak serum concentration | Decreased |
| Intestinal motility | Decreased |

can effectively treat the mother without adversely affecting the fetus.

Placental passage of a drug, leading to fetal exposure, is a function of multiple factors such as protein binding, lipid solubility, pH difference between maternal and fetal fluids, ionization constant ($pK_a$), and the effect of hemodynamic changes in the mother during pregnancy, which govern maternal and fetal pharmacokinetics such as volume of distribution, rate of metabolism, and clearance by the placenta (Table 1). Ideally, one may wish to identify drugs that will not cross the placental barrier. However, with the exception of drugs with large molecular weights, such as heparin or insulin, most medications appear to cross the placenta and are associated with varying degrees of fetal exposure. Our objective is to present the state of relevant research and to stimulate interest in research efforts aimed at developing selective drugs for use in pregnancy. The hope is to explore therapeutic strategies that can either reduce unnecessary fetal exposure to drugs required by the mother or, in some instances, reduce maternal exposure to drugs destined for fetal treatment [4].

## 2 PLACENTAL PHYSIOLOGY

Typically discoid, the human placenta consists of the chorionic villi with attendant blood vessels and connective tissues [5]. Viewed from the maternal surface, the placenta has slightly raised areas denoted as maternal lobes or cotyledons. Each maternal cotyledon is associated with several fetal cotyledons, allowing for maternal circulation to be nearly superimposed on to fetal circulation. Considering the movement of materials from the mother to the fetus, molecules are carried by the maternal bloodstream in one of three ways: in free form dissolved in plasma, bound to carrier proteins, or bound to red blood cells. As a molecule crosses from the maternal to the fetal bloodstream, the solutes must cross the syncytiotrophoblast either by passing through the plasma membrane of the trophoblast or via a network of transporters (Fig. 1). Thus, transport across the placenta involves the movement of molecules among three compartments: the maternal blood, the cytoplasm of the syncytiotrophoblast, and the fetal blood. Solute levels in each of these compartments will play a key role in controlling the rate by which substances cross the placenta and are largely dependent on the initial absorption of the drug into the maternal bloodstream.

There is evidence for the presence of different cytochromes (CYPs), such as CYP1A1, CYP2E1, CYP3A4, CYP3A5, CYP3A7 and CYP4B1 at the messenger ribonucleic acid (mRNA) level, and protein and enzyme activity early in the first trimester of pregnancy [6]. Hokkola et al. [7] demonstrated that the mRNA but not the protein of several of the CYPs can be detected in term placenta, but levels were negligible in comparison to maternal liver. Furthermore, uridine glucuronosyltransferases (UGTs) (UGT1A, UG2B7, UGT2B11, UGT2B17, UGT2B11) increase the intrinsic clearance of several nonspecific substrates from 7.5 to 43% in full-term placenta as compared to liver. However, it is unlikely that the fetal or placental metabolism contribute significantly to the changes in CYP-catalyzed drug metabolism in pregnancy. For a listing of CYP expression in the placenta see Table 2 [8].

Furthermore, the trophoblast is the interface of exchange between the maternal and fetal circulations. As a barrier, the trophoblast has abundant expression of ABC transporters such as P-glycoprotein (P-GP) and multidrug resistance proteins 1, 2, and 3 (MRP1, MRP2, MRP3, respectively) (Table 3). Powered by adenosine triphosphate (ATP), these transporters actively extrude substrates from the placenta [9] (Table 3). Specifically, P-GP and MRP2 have been shown to be expressed on the brush border (maternal side) of the human placental trophoblast [10], while MRP1 and MRP3 are on the basal membrane (fetal side) [11,12]. Multidrug resistance transporters may result in lower cellular concentration of a drug via an efflux mechanism, thus creating pharmacological sanctuaries. For example, MRP1 and MRP3 preferentially transport organic anions, promote the excretion of glutathione and glucoronide metabolites, and thus attenuate their entry into fetal blood. P-GP is also an active drug transporter of the ATP-binding cassette transporter family with a wide range of substrates [11]. Abundant in the apical membrane of the placental trophoblast, P-GP transports its substrates in an outward (extracellular) direction. Since it can be detected in placental tropho-

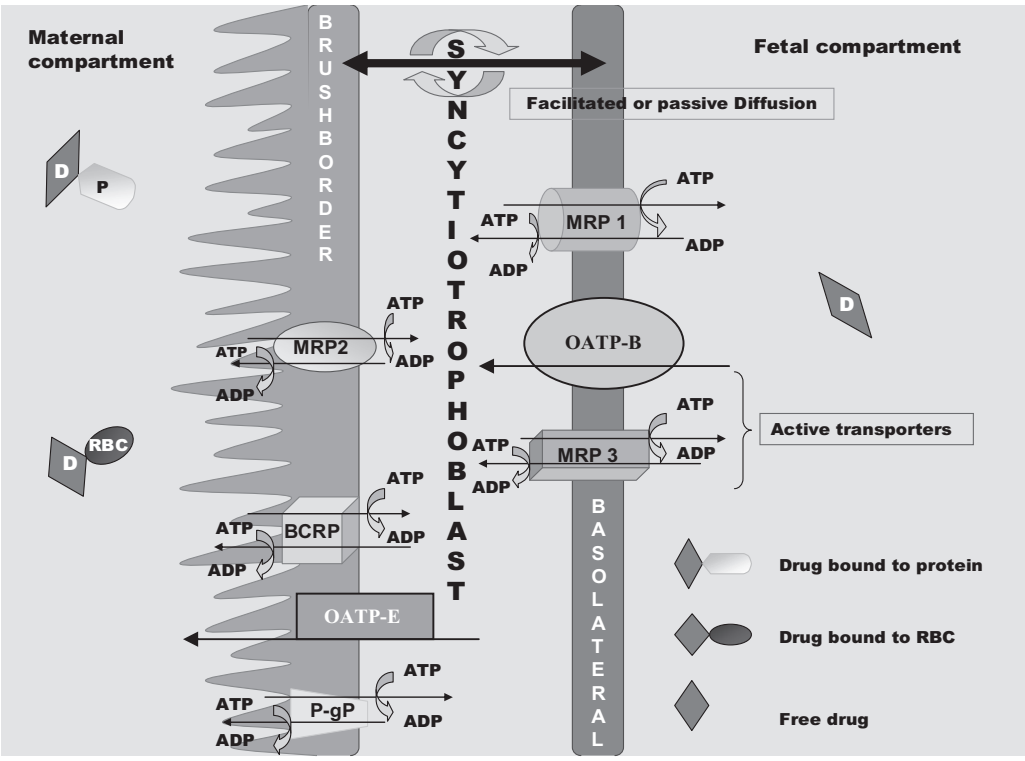


**Figure 1** Schematic of mechanisms of drug transport across the human placental barrier. (Adapted from [49].)

**TABLE 2 Comparison of Cytochrome P 450 Expression in Human Placenta between First Trimesters and at Term**

| CYP Form | First Trimester | Term |
|---|---|---|
| 1A1 | Present | Present |
| 1A2 | Present | Absent |
| 1B1 | Present | Present |
| 2C | Present | Present |
| 2D6 | Present | Absent |
| 2E1 | Present | Present |
| 2F1 | Present | Present |
| 3A4/3A7 | Present | Expression unclear |
| 4B1 | Present | Present |
| CYP 19 | Present | Present |

*Source*: From [6, 7].

blast from the first trimester of pregnancy [13], it is likely that it protects the fetus from amphipathic xenotoxins. Thus, fetal tolerance to maternal concentrations of drugs that are considered to be good P-GP substrates will be higher than that for drugs considered to be poor P-GP substrates. Then, it may be preferable to treat pregnant women with drugs that are good P-GP substrates, such as paclitaxel during antineoplastic chemotherapy in pregnancy, which may be preferred over other antineoplastic agents that cross the placenta more easily [14].

## 3 PHYSIOLOGICAL CHANGES IN PREGNANCY

During pregnancy, the maternal gastrointestinal absorption of drugs may be altered because of changes in gastric secretion and motility. Changes in gastric pH influence the degree of ionization and solubility of many drugs, modifying absorption rate and thereby altering drug bioavailability. Maternal drug metabolism may also be altered due to elevation of endogenous hormones such as progesterone. This can stimulate the hepatic microsomal oxidase system where elevated rates of hepatic metabolism may result in increased transformation of drugs such as phenytoin [2, 3]. Conversely, theophyline and caffeine exihibit a reduced hepatic elimination due to elevated levels of estradiol. In turn, estradiol inhibits microsomal drug metabolizing enzymes such as CYP1A2. Drug concentrations may also be affected by hemodynamic changes that take place throughout pregnancy. Cardiac output and blood volume increase up to 40%, primarily due to an increase in plasma volume [15]. Total body water also increases by 5–8%, due to the expansion of the extracellular fluid space and the growth of new tissue (placenta, mammary gland, the fetus) [2, 3]. Water also accumulates in the fetus, placenta and amniotic fluid, which collectively

**TABLE 3 Selected ABC Transporters and Substrates**

| Substrates | | Transporter |
|---|---|---|
| Actinomycine D | etoposide | P-glycoprotein(MDR1) |
| Doxorubincin | daunorubcin | |
| Irinotecan | mitomycin C | |
| Mitroxantrone | paclitaxel | |
| Teniposide | topotecan | |
| Vinblastine | vincristine | |
| Morphine | loperamide | |
| Dexamethasone | ketoconazole | |
| β-Acetyldigoxin | α-methyldigoxin | |
| Digitosin | quinidine | |
| Amprenavir | Indinavir | |
| Nelfinavir | saquinavir | |
| Ritonavir | Cyclosporine A | |
| Tacrolimus | Erythromycin | |
| Levofloxacin | sparfloxacin | |
| Atorvastatin | lovastatin | |
| Cerivastatin | simvastatin | |
| Diltiazem | mibefradil | |
| Verapamil | Loperamide | |
| Aldosterone | bisanterine | |
| Calcein-M | citalopram | |
| Colchicin | corticosterone | |
| Ondasetron | Paclitaxel | |
| Phenytoin | prednisolone | |
| Quinidine | Rifampin | |
| Terfenadine | (99m)-Tc-tetrofosfine | |
| Adiamycin | Etoposide | Multidrug resistance protein (MRP) 1 |
| Glutathione | mithoxantrone | |
| Methotrexate | organic anions | |
| Aflatoxin B1 | cisplatinum | |
| Paclitaxel | topotecan | |
| Vincristine | colchicine | |
| Dehydroepiandrosterone | | |
| Etoposide | fluorscein | |
| 17-β-glucoronyl estradiol | | |
| *S*-gluthathionyl prostaglandin A2 | | |
| Methotrexate | | |
| Methotrezate | Cisplatin | Multidrug resistance protein (MRP) 2 |
| Vinblastine | Vincristine | |
| Ceftriaxone | grepafloxacin | |
| Rifampicin | saquinavir | |
| Ritonavir | indinavir | |
| Estradiol | Octreotide | |
| Temocaprilate | Furosemide | |
| Pravastatin | Flavonoids | |
| Heavy metals | Arsenite | |
| Epirubicin | etoposide | |
| Doxorubicin | Sulphinpyrazone | |
| Acetaminophen | glucuronide | Multidrug resistance protein (MRP) 3 |
| Etoposide | leukotriene C4 | |
| Methotrexate | teniposide | |
| Topotecan | 9 aminocamptotecin | Breast cancer resistance protein (BCRP) |
| Doxorubicin | irinotecan | |
| Danurubicin | epirubicin | |
| Zidovudine | idrubicinol | |
| Lamivudine | flavopiridol | |
| Mitoxantrone | Prazocin | |
| Methotrezate | Indolocarbazol | |
| Adriamycin | etoposide. | |
| Pheophorbide | | |

*Source*: Adapted from [49].

**TABLE 4 Pregnancy-Related Pharmacokinetic Changes in Pregnancy for Selected Drugs**

| | *T*½ | | *Vd*(L)[a] | | CL (mL/min) | |
|---|---|---|---|---|---|---|
| Drugs | Pregnant | Nonpregnant | Pregnant | Nonpregnant | Pregnant | Nonpregnant |
| **Ampicillin** | 54.2 ± 3.9 | 69.6 ± 6.1 | 32.8 ± 2.5 | 34.5 ± 2.7 | 450 ± 31 | 370 ± 30 |
| **Cefuroxime** | 44 ± 5 | 58 ± 8 | 17.8 ± 1.9 | 16.3 ± 2.1 | 282 ± 34 | 198 ± 27 |
| **Imipenem** | 36 ± 8 | 41 ± 16 | 47.1 ± 14.8 | 18.9 ± 5.8 | 973 ± 47 | 338 ± 85 |
| **Piperacillin** | 46.5 ± 10 | 53.7 ± 4.6 | 67.6 ± 11.8 | 41.9 ± 6.2 | 1538 ± 362 | 540 ± 75 |
| **Azlocillin** | 65.4 | 72 | 15.4 | 24.7 | 126.1 | 195.7 |
| **Nifedipine** | 81 ± 18 | 360 | n/a | n/a | 266 ± 105 | 27 ± 15 |
| **Labetolol** | 102 ± 16 | 160 | n/a | n/a | 1704 ± 531 | 1430 |
| **Sotalol** | 396 ± 36 | 558 ± 42 | 106.4 ± 8.1 | 87.3 ± 7.2 | 196 ± 24 | 109 ± 7 |

n/a = not available.
Vd = volume of distribution.
*Source*: Adapted from [49].

contributes to an increase in volume of distribution and may lower the concentration of drugs and increase their elimination half-life (Table 4).

### 3.1 Protein Binding in Pregnancy

Protein binding alterations occur in pregnancy as a result of changes in the concentrations of specific proteins as well as changes in protein binding affinity for certain drugs. Decreases in maternal serum albumin may lead to corresponding increases in the free fraction of drugs [16, 17]. Increased levels of free fatty acids and total lipids together with gestational hormonal changes can also decrease the binding capacity of drugs to protein. This may have implications to maternal–fetal drug transfer, since binding changes may influence the maternal plasma concentration of free drug available to partition across the placental barrier. Furthermore, in late pregnancy there is a lower concentration of drug binding proteins and changes in binding affinity of drugs in the fetus: albumin concentration in the cord blood of neonates is markedly lower than the maternal levels [18]. In addition, there is a threefold lower level of $\alpha_1$-acid glycoprotein in the fetus compared to the mother [15]. Should free drug cross the placenta, the fetus will exhibit larger concentrations of free drug, due to diminished protein binding capacity.

### 3.2 pH Changes in Pregancy

Lipid solubility, the pH of the maternal and fetal fluids, the ionization constant of the drug ($pK_a$), and molecular weight all influence the passage of the drug across the placenta. In general, uncharged, un-ionized molecules with high lipid solubility and low molecular weight penetrate the cell membranes more readily than hydrophillic ionized drug molecules [19].

The transfer of weak acids and bases is influenced by the pH gradient between the maternal and fetal circulations and by the $pK_a$ of the drug. Since fetal plasma is slightly more acidic, 0.1 lower than maternal plasma pH, for weak bases, the un-ionized free drug crossing the placenta becomes ionized and is "trapped" in the more acidic fetal circulation [2, 18]. The pH of the amniotic fluid is also lower than that of maternal plasma. Therefore, the more acidic environment of both fetal blood and amniotic fluid favor ionization and accumulation of a basic drug due to "ion trapping," often resulting in fetal drug concentrations that may exceed maternal plasma concentrations and lead to toxicity.

Finally, at the placental cytoplasmic barrier, free drug can cross either by simple or facilitated diffusion or by active transport. Diffusion, simple or facilitated, depends upon the transmembrane concentration gradient, where it moves from a higher concentration compartment to an area of lower concentration. In contrast, active transport depends not only on solute concentration but also on a steady supply of ATP and progresses against a concentration gradient, potentially accumulating drugs in fetal circulation or, more commonly, pumping drugs back into the maternal circulation, against a higher drug concentration in maternal plasma. In addition, the placenta itself contributes to maternal–fetal drug transfer by possessing metabolizing enzymes capable of oxidation, reduction, hydrolysis, and conjugation, which can potentially convert an inactive drug into an active metabolite and vice versa. Clearly, the transfer of drugs across the placental membrane will vary, depending on the apparent volume of distribution, the degree of protein binding, acid–base equilibrium, metabolic and excretory mechanisms in the placenta and fetus, as well as hemodynamic alterations.

## 4 EXPERIMENTAL TECHNIQUES

Species differences in placental anatomy and physiology have precluded the generalizability of animal

**TABLE 5 Species Variation in Placental Physiology**

| Order | Placental Shape | Placental Type | Feto-maternal Integration | F/M Blood flow Interactions |
|---|---|---|---|---|
| Human | Discoidal | Hemomonochorial | Villous | Multivillous |
| Rabbit | Discoidal | Hemodichorial | Labyrinthine | Countercurrent |
| Mouse/rat | Discoidal | Hemotrichorial | Labyrinthine | Countercurrent |
| Goat/sheep | Cotyledonary | Epithelichorial | Villous | Multivillous |
| Guinea pig | Discoidal | hemomonochorial | Labryrinthine | Countercurrent |
| Great apes | Discoidal | hemomonochorial | Villous | Multivillous |

*Source*: From [5].

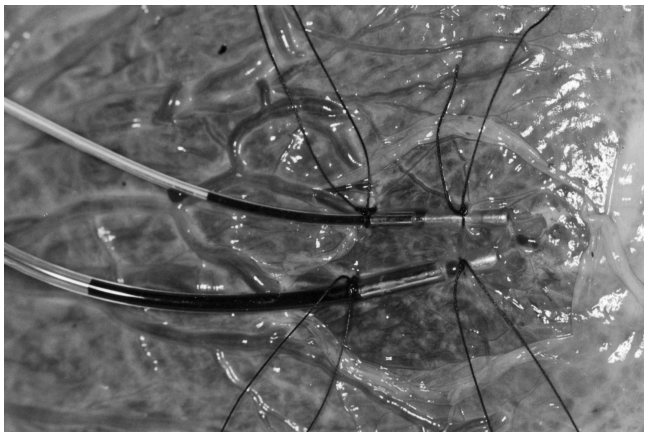

**Figure 2** Photograph of canulas inserted into a vein and artery in a dually perfused isolated human placental cotyledon.

studies as they relate to humans [5] (Table 5). To circumvent these issues, techniques such as the dual perfusion of isolated human placental cotyledon have been well validated through transfer studies [8, 12, 20–26]. This technique is useful in the measurement of: (1) placental transfer of substances, (2) the effects of endogenous and exogenous substances on fetal pressure and transport, and (3) the release of endogenous substances into maternal and fetal compartments. The technique of in vitro perfusion of human placenta has been reviewed by Sastry [27] and Bourget [28].

### 4.1 Dually Perfused Isolated Human Placental Cotyledon

The technique consists of collecting human term placenta after delivery and immersing it in heparinized Krebs–Ringer bicarbonate buffer within 10 min of collection to remove any fetal blood components [29]. A single cotyledon, nourished by a single vein–artery pair without leaks, is isolated for perfusion studies. See Figure 2.

The placental cotyledon is perfused with 21% $O_2$, 5% $CO_2$. If necessary, in both the maternal and fetal circulating fluids (perfusate), albumin can also be added to mimic plasma. In this case, protein binding of the drug must be considered while measuring placental transfer [30]. Antipyrine is generally used as a freely diffusible marker to evaluate the placental transfer of drugs. Experiments can be performed using either the closed (recirculating) method or the open (single-pass or non-circulating) method [8] (Fig. 3).

In the closed model, both the maternal and fetal perfusates are recirculated. After a 1-h equilibration period, the drug and antipyrine (10 μg/mL, 53 μ*M*) are added to either the maternal and/or fetal reservoir and allowed to equilibrate for 3 h. Samples (1 mL) are taken from the maternal and fetal circuits at regular intervals (30 min) for the measurement of the drug, antipyrine, lactate production, and glucose consumption. The integrity and viability of placental tissue is ensured by monitoring the lack of fluid shifts, changes in partial pressures in both maternal and fetal circuits, lactate output, lactic dehydrogenase release, and glucose consumption [8].

Recirculating studies are not useful for calculating clearance data because these data change over time as more drug is transferred and accumulated toward equilibrium between the circuits [31]. Rather, we can calculate the percent transfer of drug, expressing the data as a ratio:

$$\%\ \text{transfer} = \frac{\text{Fetal concentration (at time } t)}{\text{Maternal concentration (at time } t)} \times 100$$

In open perfusion experiments, the fetal, the maternal, or both perfusates are not recirculated. In this context, a steady state can be attained. Maternal-to-fetal or fetal-to-maternal drug clearance (Cl) can be calculated by a standard formula [32], that is, maternal clearance:

$$\frac{[\ ]_{\text{fetal vein}} - [\ ]_{\text{fetal artery}}}{[\ ]_{\text{maternal artery}}} \times \text{fetal perfusion rate}$$

where [] is concentration.

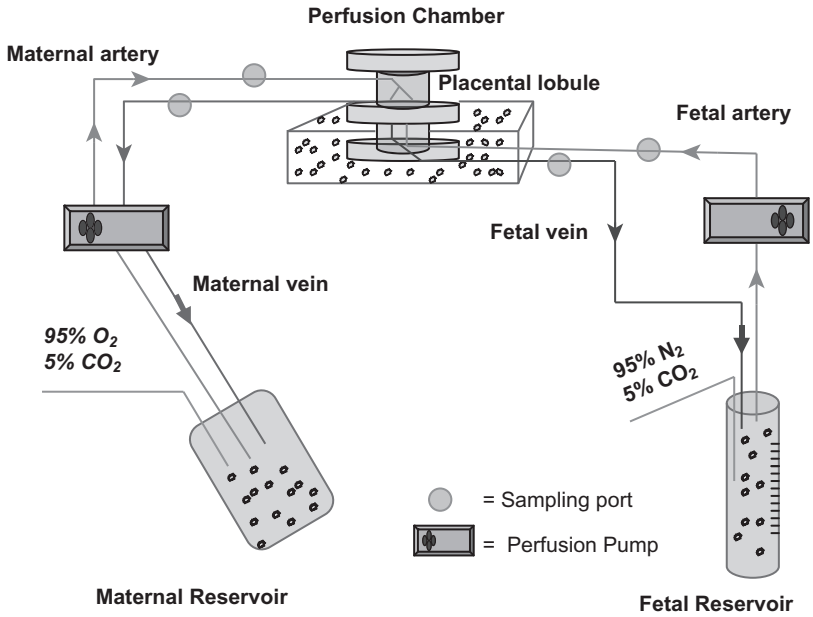


**Figure 3** Schematic of a closed-circuit system used in the dually perfused isolated human placental cotyledon.

### 4.2 Trophoblastic Tissue

Human trophoblastic tissue preparations have been used to study the uptake into placental tissues, most classically used for amino acids and drugs uptake [33]. The assumption follows that what is taken up by the trophoblast would be transferred into fetal circulation by diffusion in vivo. The uptake is assumed to represent the transfer of substances from the maternal circulation into syncytiotrophoblast cells since there is no intact fetal placental circulation where the capillaries are usually collapsed [34]. Various types of preparations have been used for this type of pharmacological experiments; for example, placental slices, dissected syncytiotrophoblast tissue, microvillus membrane vesicles, and subcellular fractions such as microsomes have been shown to be good for drug metabolism studies [35].

### 4.3 Placental Tissue Slices

Placental slices have been found to be suitable for studying the effects of induction of transport systems relevant to fetal growth retardation [36]. Most often, they have been used to study amino acid uptake or transport using nonmetabolizable amino acids, such as α-aminobutyric acid (AIB) or as drugs of addiction. The sensitivity of this technique has been suggested to be the expression of compensatory responses such as the synthesis of new transporters and their incorporations into the plasma membrane. Amino acid transport has been suggested to be induced by preincubation [37]. This effect is seen in both term and preterm placenta. This behavior is characterized by the overall activity of these transporters being diminished in the early portion of the preincubation period due to an increase in membrane fluidity [38]. Practically, such compensatory mechanisms represent the possibility that the placenta may compensate in times of maternal starvation to synthesize new transporters or to increase their availability to spare the fetus from amino acid deprivation [39].

### 4.4 Syncytiotrophoblast Tissue

It is also possible to dissect only placental villi and to suspend them in physiological buffers, which also allows for the measurement of uptake of substances added to the medium [40]. In this preparation, the trophoblast surface is exposed to the medium. This preparation is used (1) to measure the uptake by the trophoblast from the suspension buffer of nutrients and drugs, (2) to measure the release of hormones, peptides, and biogenic amines from the trophoblast, and (3) to measure the effect of drugs and chemicals on these processes. Often, as with radiolabeled drugs, the uptake is measured as the inside/outside ratio [41]. This ratio assumes that all drug disappearing from the medium is taken up by the tissue. This assumption is valid only if uptake calculated by the above method is compared with the direct measurement of labeled drug (e.g., radiolabeled) in the placental tissue. The total drug measured directly in the placental tissue at the end of the experiment should be

the same as that calculated by the calculated method. During a 90-min incubation period, ratios of 7–10 are typically obtained and uptake often remains linear for up to 2 h, depending on the buffer medium [42].

This method has advantages compared to direct measurement of intracellular drug concentration in tissue. Because there is a significant variability among different pieces of the same placenta, by eliminating the removal of placental tissue to sample uptake, the same tissue remains in the medium throughout the experiment [43]. This technique allows for the same tissue to serve as its own control when uptake during the first half of the experiment is compared with uptake during the second half (i.e., should a second drug be added). This method is sufficiently sensitive to detect small drug-induced differences in uptake that cannot be observed by sampling the tissue directly [44]. Since the incubation medium is sampled instead of the tissue, more samples can be collected, more frequently.

### 4.5 Plasma Membrane Vesicles

Plasma membrane vesicles consist of exfoliated microvilli from the synciotrophoblast cells of human placenta [45]. The exfoliated microvilli are purified by differential centrifugation and washing. Essentially, plasma membrane forms microvesicles whose orientation is tested by two markers: alkaline phosphatase to test for brush border vesicles enrichment and $Na^2+/k$ + ATPase for basal lateral membrane enrichment [46]. This preparation is a more specific way of studying the effects of drugs on transporters or endogenous substances across specific placental plasma membranes [47].

### 4.6 Cultured Human Placental Villi and Cells

The human placental membrane contains several receptors as well as transporters (Fig. 1) [48, 49]. It is difficult to study receptor up/downregulation, sensitization/desensitization, and tolerance development in whole placenta. The availability of in vitro purified trophoblast systems allows for the study of several aspects of trophoblast functions, such as transport, secretion of steroids or hormones, and other systems in human placenta. Danels-McQueen et al. [50] have developed a method for isolating trophoblast from first-trimester placenta as well as recipes for culture media for their growth. Large multinucleated islands that represent differentiated forms of trophoblast are marked by their difference in human chorionic gonadotropin (hCG) and protein production. First-trimester placentas are favored over term placentas since they yield 40% more cytotrophoblast as compared to only 11% cytotrophoblast from term placentas.

Furthermore Hall et al. [51] have described a procedure for short-term tissue culture of human midterm and term placentas. From term placentas, monolayer cell cultures can be obtained that contain viable trophoblast cells. The optimal culture conditions occur with media containing epidermal growth factor (EGF), high-density lipoprotein (HDL), insulin, and transferrin. Hormone secretion is promoted in media containing 10% fetal calf serum (Dubelco's minimal essential medium, DMEM). Kliman et al. [52] improved on Hall's procedure by introducing a discontinuous Percoll density gradient to purify trophoblast cells. The cells prepared by this method become cultured aggregates and a "synctium" is formed that secretes hCG and human placental lactogeu (hPL) within 18 h and by 24 h.

### 4.7 Malignant Trophoblast Cell Culture

Three pure trophoblast cell lines derived from human gestational choriocarcinoma are presently available.

***BeWo Cells*** BeWo cells are a parabiotic combination of malignant gestation choriocarcinoma of the fetal placenta and human decidual tissue, which secrete a spectrum of placental hormones such as hCG, hPL, polypeptide hormone, estrogenic and progestational steroids, estrone, estradiol, and progesterone. Nevertheless, Takeuchi et al. [53] have shown that the molecular characteristics of BeWo hCG and its subunits, such as molecular weight, possible structure of sugar chains, are different from those found in normal placental cells.

These cells exist as two phenotypes in culture: the predominant form, 96–99% of the total population is cytotrophoblast-like (CTL), are proliferative and mononucleated; the remaining cells are syncytiotrophoblast-like (STL), nonproliferative and multinucleated. Furthermore, these cells are methotrexate (MX) [a deoxyribonucleic acid (DNA) synthesis inhibitor] resistant. This property is useful for the differentiation of BeWo cells. Methoxytrexate treatment causes inhibition of cell division but does not stop nuclear division of DNA or protein synthesis, yielding a population of giant multinucleated, nonproliferative cells that are morphologically indistinguishable from the naturally occurring STL cells. This effect is reversible upon withdrawal of methoxytrexate treatment. As such, the BeWo cell line is a valuable model culture system for investigating mammalian trophoblast development.

***JAr Cells*** JAr cells were established by R.A Pattillo [54] directly from a trophoblastic tumor of the placenta and is considered the preferred placental malignant cells model. This cell line produces estrogen, progester-

one, gonadotropin, and lactogen in culture. For a complete characterization of cellular morphology, hormone production, protein production, and enzyme compliment see [41].

Techniques have been developed to grow JAr cells as muticellular spheroids [50], which renders them the most suitable model for studying normal placental cell functions. A spheroid is a cultured suspension, and their growth is not limited by the size of their culturing environment. A spheroid culture system allows for the formation of a three-dimensional arrangement of cells that can be kept in culture for at least 3 weeks. These cells can grow as large as 1 mm in diameter, which can allow for the assessment of the contribution of cell density on treatment effect. The long culture period and free-floating three-dimensional nature of the spheroids is what mimics in vivo conditions. This is more advantageous than traditional monolayer cultures for the study of placental functions, which change with time, acute and chronic effects of drugs on trophoblast function, or for the study of higher placental functions such as implantation or differentiation, where the three-dimensional arrangement of the cells is desirable to observe.

***JEG-3 Cells*** JEB-3 cells are one of six clonally derived cell lines isolated and described by Kohler and Bridson [55]. They have been shown to release human chorionic gonadotropin, human chorionic somatomammotropin, and progesterone and transform steroid precursors to estrone and estradiol. Ringer et al. [56] have suggested that cyclic adenosine monophosphate (cAMP) regulates the expression of endrocrine functions of normal trophoblast cells in part by controlling the abundance of mRNAs encoding steroidogenic enzymes and hCG subunits. Several studies have used JEG-3 cells to investigate aspects of trophoblastic functions including the differential regulation of hormones, transcriptional regulation, stimulation of synthesis, and the release of hormones. JEG-3 is slowly proving to be an important model system for investigating the effects of drug and toxicants on cellular differentiation.

### 4.8 Animal Models

Although placentas of all species are similar in that the distance between maternal blood and fetal blood is decreased in later stages of gestation, the manner in which they do so is variable [57]. With the exception of the primate, the features of the intervening cells with regard to permeability, active transport, and even metabolic activity vary sufficiently that implication from one species should be extended only cautiously to another.

On the other hand, the macaque and human hemomonochorial chorioallantoic placentas are quite similar, with the principal variations originating in the mechanism of blastocyst implantation [58]. While the macaque blastocyst invades the luminal epithelium of the endometrium on day 9 of a 168-day pregnancy, the human blastocyst is believed to implant about 7–8 days after spontaneous ovulations. The process of trophoblast development is described in several reports [59, 60]. Overall, the placentas of the macaque and human are very similar physiologically except that the anchoring villi have longer cytotrophoblastic cell columns in the macqaque. In addition, the junction of the endometrium with fetal tissue at the basal plate is much more regular in the macaque than in the human. In reviewing comparative implantation, it should be noted that caution should be exercised in extrapolation of information about blastocyst implantation from one species to another. This is equally applicable to extrapolating information concerning placental transfer characteristic between different species.

## 5 DRUG DEVELOPMENT PROCESS

The drug development process includes two main stages: preclinical and clinical development. Preclinical testing is mainly concerned with the identification of a molecule as a potentially useful drug and its evaluation in different experimental settings, including animals. The clinical stage encompasses all the steps necessary to prove safety and efficacy of a drug in humans.

### 5.1 Preclinical Testing

After a potential new drug is identified, it must undergo a myriad of experimental tests to define its mechanism of action, receptors, binding characteristics, organ distribution in an animal model. Once detailed information for the molecule is obtained, further studies are usually conducted in animals to evaluate pharmacokinetics (absorption, distribution, metabolism, and excretion) of the drug, toxicity, efficacy, and effective drug doses that will serve as a basis for studies in humans. Toxicity evaluation is an important part of animal testing and provides important information on what should be expected in humans, although not all human side effects are predicted by animal studies, neither will all toxic effects observed in animals occur in humans.

In North America, the complete data from preclinical studies has to be submitted to the Food and Drug Administration (FDA) in an investigational new drug (IND) application in order to obtain authorization to proceed to clinical studies. Similar procedures are

followed in other regions such as the European Union, Japan, and Canada, with variations according to the particular legislation of each country.

### 5.2 Clinical Testing

Testing of new drugs in human subjects in the context of controlled experimental trials is an essential requisite before the widespread use of the drug for a certain disease in people. As a matter of fact, many drugs never succeed in passing the clinical testing stage, which can reveal aspects of the drug that were previously undisclosed by preclinical studies (e.g., a drug may produce toxicities that were not observed in animals).

Most drugs are studied in humans following a defined sequence, commonly referred to as phase I, phase II, and phase III clinical trials. Each phase addresses slightly different questions regarding safety and efficacy of the drug in humans, with subsequent phases enrolling progressively more subjects, from roughly a dozen in phase I trials to hundreds, or even thousands, in phase III. All human clinical studies are expected to be conducted taking in consideration the ethics principles outlined in the Declaration of Helsinki (www.wma.net), and following good clinical practices as outlined by the ICH (International Conference on Harmonization) guidelines (www.ich.org).

### 5.3 Phase I Studies

The initial studies in humans of a new drug are known as phase I studies. Most of these studies are carried out in a few (usually no more than a dozen) healthy volunteers in order to obtain information on how the drug behaves in humans, including pharmacokinetics (what the body does to the drug, that is, absorption, distribution, metabolism, and excretion) and pharmacodynamics (what the drug does to the body, that is, drug effects) data, and basic toxicology information (i.e., common side effects). In some particular scenarios, drugs are tested on patients from the start, usually due to the expected toxicity of these medications that would only be offset by the expected benefits that they would provide to patients. Examples of this situation can be found in phase I studies of drugs for cancer treatment, human immunodeficiency virus (HIV), and some neurological diseases.

### 5.4 Phase II Studies

Once, and if, a drug has successfully passed phase I, it is subsequently tested in phase II clinical trials. It is worth noting that most (9 out of 10) drugs never reach this stage, as most drugs tested in phase I trials are found to have unexpected toxicities, little absorption, interactions with other drugs, or other metabolism issues or other problems that would make the drug potentially dangerous, useless in humans, too difficult to develop or administer (or, sometimes, even difficult to commercialize).

In phase II, drugs are tested in selected groups of patients that have the condition for which the drug has been developed. These studies enroll more patients than the previous phase (usually more than a dozen, but hardly ever more than a hundred), and they are designed to provide information that will allow optimal selection of dosage for the next phase. Also, data on therapeutic efficacy of the medication is obtained, although efficacy can not be proven with such small numbers of patients. This phase also provides information on the common toxicity spectrum in the population of patients where the drug is going to be used, which may differ from the one observed in healthy subjects.

By the time a drug enters phase II studies, it is required (by the FDA) that animal studies are done on three different species (usually rats, mice, and rabbits) to evaluate potential reproductive toxicity and teratogenesis. These are the only tests throughout the development process that attempt to address the effects of the drug during pregnancy. However, as previously stated, not all human adverse effects can be inferred from animals studies. These studies have to be completed before authorization is given to market the drug.

### 5.5 Phase III Clinical Trials

In this last phase of clinical testing, drugs are studied in large clinical trials in populations of patients with the target condition. These trials need to prove efficacy of the drug in the target population before a drug can be marketed. Most of these trials are double-blind, randomized, placebo-controlled clinical trials enrolling hundreds to thousands of patients (sometimes more). These studies most commonly compare the efficacy of the drug against a placebo (placebo controlled) or sometimes the standard of treatment, if a treatment that is partially effective already exists for the condition. The drug is usually given to the patients without the patient or the doctor administering it knowing whether medication or placebo is being taken (double blind).

If phase III clinical trials successfully demonstrate that a new drug is effective and safe, the drug is poised for marketing. Only then can the regulatory agency grant permission to sell the drug to the general population for the indication studied in the phase III clinical trials.

### 5.6 Postmarketing Studies (Phase IV)

After a drug is on the market, pharmaceutical companies are required to collect data on safety. These monitoring efforts should continue while the drug is on the market and include surveillance of reproductive effects if the drug has the potential to be used in pregnant women. Many registries collect information on pregnancy outcomes in women exposed to medications. This is the main way that information on effects of drugs in pregnancy is obtained, as formal clinical trials rarely include pregnant women (or even women in reproductive age unless they provide negative pregnancy tests and commit to use adequate contraception during the study). Clinical trials carried out exclusively in pregnant women are almost nonexistent (rare exemptions will be provided below) and are frequently small and largely funded by and conducted in academic centers, instead of the pharmaceutical companies as is the rule in most other situations [61]

## 6 SELECTED CASES OF DRUGS DESIGNED FOR USE PARTICULARLY DURING PREGNANCY

Few drugs have been developed specifically for pregnancy. Nevertheless, there are certain pathologies that develop only during the course of gestation, usually as a consequence of the interplay between the mother and the fetus. In this case, drug development in pregnant patients is the only option, where ethical and logistical issues render the drug impossible to test in a nonpregnant population.

The following examples are used to exemplify the particular obstacles faced in their development.

### 6.1 Atosiban

Spontaneous preterm labor and delivery accounts for approximately one-third of all preterm births and is the predominant cause of perinatal mortality and morbidity [62]. Oxytocin acting via uterine oxytocin receptors plays a central role in mechanisms of uterine contraction in preterm labor [63]. Research on the pathophysiological mechanisms of preterm labor and the local production of oxytocin in fetal membranes and deciduas lead to the idea of blocking oxytocin receptors to prevent premature uterine contractility. Oxytocin stimulates both oxytocin and vasopressin V1a receptors in preterm pregnancy. There is a dramatic increase of oxytocin receptor in the pregnancy myometrium as gestation progresses [64]. In this regard, the increase in myometrial oxytocin receptor density without a corresponding increase in oxytocin plasma levels during the early stages of pregnancy allows for the assumption that the uterine hemostatic imbalance is the result of an increased affinity of oxytocin to its receptor [65]. As such, oxytocin is believed to be the main culprit in inducing contractility of myometrium in both term and preterm births. Consequently, the mainstay of treatment is to employ oxytocin receptor antagonist to delay preterm labor.

Atosiban is an oxytocin receptor antagonist that binds to vasopressin V1a as well as to the oxytocin receptors. Atosiban has been shown to be equally as potent as previously used therapies, such as β-2 receptor agonists, and safer for both mother and fetus. An in vitro study of the myometrium of 27 women was treated with oxytocin, then washed in atosiban, and then rechallenged with oxytocin [65]. This study clearly showed an antagonist action of atosiban on oxytocin-induced myometrial activity in a concentration-dependent manner.

Following in vitro evidence of pharmacological activity of atosiban against oxytocin-induced myometrial contractility, safety studies were carried out in animals and later in healthy (male) volunteers. These steps were necessary in order to obtain authorization to proceed to testing atosiban in other human populations. A case had to be made to the regulatory agencies such as the European Medicines Authority (EMA) (Europe) and the FDA (United States) that this drug could only be tested in phase II clinical trials in pregnant women undergoing premature delivery, given that there is no other population where this drug would be effective. This posed a dilemma, given that drugs are only tested on a limited number of subjects before reaching phase II, and it was impossible to rule out uncommon and serious adverse drug reactions at this stage. To make matters worse, the particular population of pregnant women implied a double risk for adverse effects, as not only the mother but also the fetus might be exposed to the medication, albeit for a short period, and at a less sensitive developmental stage than the first trimester.

Finally, a placebo-controlled trial with 501 patients was conducted and showed that atosiban provides a significantly better tocolytic effect at or beyond 28 weeks of pregnancy when compared to placebo [63]. Due to the adverse effect profile of other tocolytic agents employed to date, such as pulmonary edema, particularly with the β-sympathomimetics, the use of tocolytics in extreme premature labor is still not justified due to the severe neonatal consequences such as respiratory distress syndrome, fetal tachycardia, and hypoglycemia. Furthermore, previous tocolytics have been shown to cause increases in maternal pulse and blood pressure, lipolysis, and severe metabolic acidosis. Nevertheless, atosiban treatment seems warranted since it provides both the maximum therapeutic effectiveness

and safety for both the mother and her unborn child as compared to previously used nonspecific oxytocin receptor antagonists [63–65].

### 6.2 Glibenclamide (Glyburide)

In some cases, drugs already developed for indications in the nonpregnant population find an application in pregnancy-specific conditions. Diabetes mellitus is a disorder of glucose metabolism, generally classified as type 1 (insulin dependent) or type 2 (noninsulin dependent) diabetes, with a special class denoted gestational diabetes, defined as diabetes mellitus that is identified during pregnancy [65]. Gestational diabetes usually develops in or after the 24th week of gestation, mainly due to the effects of placental hormones on glucose metabolism in the mother. Failure to recognize and treat this condition may lead to unnecessary morbidity in some pregnancies (i.e., macrosomia, shoulder dystocia, neonatal complications) [66]. Standard management for gestational diabetes includes a diabetic diet and exercise; however, some cases require further medical treatment in order to achieve and maintain euglycemia. Perinatal morbidity has been shown to correlate to the mother's level of glycemic control [67], and several clinical trials have shown that strict glycemic control can reduce the rate of malformations and perinatal mortality in the pregnant diabetic population [68].

Drugs usually used for treatment of type II diabetes have potential in the treatment for gestational diabetes, but their use in pregnancy is accompanied by risks, uncertainties, and/or significant difficulties. Insulin, the most accepted pharmacological treatment to date [69–70], still presents a risk of transplacental transfer, and, more importantly, noncompliance due to the need of frequent injections in an otherwise normal healthy population of pregnant women who need to acquire the skills required for correct compliance and control of the treatment. On the other hand, sulfonylureas, oral alternatives to insulin for the treatment of type II diabetes mellitus, are a class of drugs that appear to act by inhibiting potassium efflux via ATP-dependent potassium channels in pancreatic β cells. This action leads to cellular depolarization and calcium-stimulated release of insulin in the pancreas. In other words, they enhance the release of endogenous insulin. However, sulfonylureas have a teratogenic potential, and it is recognized that they can cross the placental barrier and appear in the fetal circulation, increasing the risk of neonatal hyperinsulinemia/hypoglycemia. This led to the virtual banning of their use in the pregnant population [71].

Still, Glyburide (glibenclamide) is a second-generation sulfonylurea that has been shown not to cross the human placenta both in vitro in perfused human placenta studies, as well as in vivo, where it is likely effluxed from the fetal circulation to the mother [70]. On the basis of this data, and given the substantial discomfort associated with insulin use in the population of pregnant women with gestational diabetes, a randomized control trial of 404 pregnant women suffering from gestational diabetes was conducted by Langer et al. [2]. No detectable levels of glibenclamide were found in cord serum of infants following maternal administration. Many factors seem to play a role in the failure of glibenclamide to cross the human placenta, including its exceptionally high protein binding (99.8%), which allows for less than 0.2% of free drug to circulate and cross the placenta. Additionally, its short elimination half-life, made possible by its low volume of distribution (0.2 L/kg) and rapid clearance (1.3 mL/kg/min), may limit the amount of Glyburide available for exchange through the placental barrier before it is eliminated from systemic circulation. Simply put, glibenclamide has only a brief opportunity to cross the human placenta [69]. Moreover, factors such as transport of Glyburide by drug transporters, such as the ABC transporter family, may be facilitating its efflux from the fetal side back into the maternal circulation. The unique interplay of pharmacokinetic parameters, such as high protein binding and short elimination half-life, in combination with its affinity for the ABC transporter family, is suspected to prevent Glyburide from crossing the placentae, making it suitable for treatment in pregnancy.

### 6.3 Doxylamine (Diclectin)

Diclectin is an antiemetic medication that contains 10 mg doxylamine succinate, a common antihistamine with intrinsic antiemetic properties found in over-the-counter sleeping medication and 10 mg pyridoxine hydrochloride (vitamin $B_6$) [72]. Pyridoxine may have intrinsic antiemetic properties and also may be synergistic with the antinauseate property of antihistamines. Diclectin has been used since 1950 and is considered to be a safe treatment for nausea and vomiting associated with pregnancy. However, because of its removal from the market in 1983, there is a 20-year gap in clinical pharmacological and toxicological research both in studying pharmacokinetics/dynamics as well as maternal and fetal safety. A large number of cohorts and case control studies as well as two meta-analyses have clearly shown that pregnant women exposed to the doxylamine–pyridoxine combination do not have an increased risk of delivering infants with birth defects. Furthermore, the international Cochrane Collaboration has systematically reviewed randomized trials of Diclectin and concluded that it safely provides considerable relief for nausea and vomiting associated with pregnancy.

**TABLE 6 Pharmacokinetic Behavior of Selected anti-HIV Drugs in Pregnancy**

| Drug Class | Drug | Half-Life | Clearance | Protein Binding | Bioavailability | Placental Transfer |
|---|---|---|---|---|---|---|
| Nucleoside reverse transcriptase inhibitors | Zidovudine | No change (1.1 h) | Increases | <25% | 63% | Cross by diffusion |
| | Lamivudine | No change (6 h) | No change | 10–50% | 55% | Cross by diffusion |
| | Didanosine | No change (1.5 h) | Increases | | 45% | Cross by diffusion |
| | Stavudine | No change | | | 86% | Cross by diffusion |
| Nonnucleoside reverse transcriptase inhibitors | Nevirapine | Unchanged | Increased | 60% | 90% | Cross by diffusion |
| | Efavirenz | Neural tube defect : not recommended for use in pregnancy | | | | |
| | Delavirdire | | | | | |
| Protease inhibitors | Nelfinavir | Unchanged (3–5 h) | | 98% | | Not detectable in cord blood |
| | Lopinavir | Unchanged (5–6 h) | | 99% | | |
| | Amprenavir | Unchanged (6 h) | | 98% | | 6–8% of maternal [concentration] in cord blood |
| | Saquinavir | | | 98% | | Not detectable in cord blood |
| | Hydroxicarbamide: Not recommended in pregnancy | | | | | |

*Source*: Adapted from [49].

Although there are newer antiemetics with potentially better profiles of efficacy and maternal safety ($5H_3$ antagonist), the issue of fetal safety dictates the use of older drugs in which more data exist [73].

### 6.4 Selected Anti-HIV Drugs

In certain clinical situations, fetal exposure to a drug may be desirable in order to treat the unborn child. An important example is antiretroviral treatment of HIV-infected pregnant women. Data suggest that nucleoside reverse transcriptase inhibitors rapidly cross the placentae (Table 6). Cord blood concentrations of Zidovudine and Lamivudine tend to equal maternal concentrations at the time of delivery, whereas cord blood concentrations of didanosine and zalcitabine are approximately 50% that of maternal concentrations [74, 75]. Nonnucleotide reverse transcriptase inhibitors, such as nevirapine, have also been shown to cross the placenta [74, 76]. Several factors may account for their passage, including the drug's lower protein binding (60%), low molecular weights, and favorable degree of lipophilicity (log octanol/water coefficient 1.81) [77]. Also, the majority are placental transporter substrates. Hence, these drugs passively diffuse across the placenta and are administered in sufficient doses to cross the placenta and prevent maternal–fetal transmission of HIV during labor.

Conversely, protease inhibitors (PI) do not cross the human placenta to a clinically appreciable extent [78]. Nelfinavir, ritonavir, saquinavir, and lopinavir undergo incomplete transplacental transfer [79]. Low PI placental transfer can be attributed to high protein binding (98%), and several are substrates for placental P-GP transporter [9]. For example, saquinavir is a P-GP substrate with a high molecular weight (767 g/mol), high protein binding (98%), and a low partition coefficient (octanol/water 4.1 $\log_{10}$) that may contribute to the small amount that crosses the placenta [80]. Other protease inhibitors that follow similar placental kinetics include indinavir, ritonavir, lopinavir, and nelfinavir. P-GP's relative affinity to different PI drugs varies three- to fourfold [81]. Nelfinavir has become a commonly used protease inhibitor during pregnancy because of its tolerability and potency; however, it has also been demonstrated to exhibit a more rapid development of viral resistance and less durable suppression of viral replication leading to a greater likelihood of HIV transmission to the fetus. Saquinavir monotherapy is not adequate and often requires "boosting" with low-dose ritonavir. However, ritonavir is rarely used in pregnancy owing to its frequent and severe gastrointestinal adverse effects [78]. Only 5.3% of maternal plasma concentrations of ritonavir are detected in fetal cord blood. Ritonavir's well-documented interaction and perhaps inhibition-selected members of the ABC transporter family may account for its exclusion from the placenta [82]. However, ritonavir's inhibition of these transporters is disadvantageous since it gives rise to drug exclusions, altered bioavailability, and changes in drug

distribution resulting in a decreased efficacy of treatment with time. If certain PIs are effluxed by the placenta, and not reaching the fetus, future use of appropriate drug transporter inhibitors, such as valspodar for p-glycoprotein, may allow for greater concentrations in the fetal compartment and the prevention of maternal–fetal HIV transmission.

As the use of combination antiretroviral regimens becomes increasingly common among HIV-infected pregnant women and the rate of transmission decreases, the safety, toxicity, and teratogenicity of these agents becomes paramount. For example, a common toxicity with zidovudine is bone marrow suppression and decreases in hemoglobin in infants exposed to zidovudine at 3 weeks of age [83]. As a class, nucleoside reverse transcriptase inhibitors have been suggested to cause fetal hepatic microsomal damage in infants exposed perinatally [84]. Additionally, combination therapy has been suggested to trigger glucose intolerance. Zidovudine in combination with lamivudine and nelfinavir fosters a reduction in neonatal insulin similar to that observed in type I diabetes [83]. These data may indicate an early damaging effect on fetal pancreas such as inhibition of proinsulin conversion to insulin due to the activity of these protease inhibitors. Cumulative damaging effects on pancreatic β cells may culminate in the death of these cells and consequent insulinopenia followed by clinical diabetes.

## 7 FUTURE DIRECTION: ORPHAN DRUG STATUS

Because of the disincentive of small markets and high liability, pregnancy-specific problems that would benefit from pharmacologic intervention such as preterm birth, induction of labor, nausea, vomiting, and cholestasis during pregnancy are removed from the natural connection between basic science and industry that has resulted in the majority of the progress from which we benefit in modern medicine [85].

However, many of the factors limiting drug development for pregnant women could be addressed by a "pregnancy drug act" modeled on the Orphan Drug Act or Best Pharmaceuticals for Children Act [86].

The Orphan Drug Act was the first initiative to specifically target improvements in the number of treatments for rare diseases. Here, patient advocacy groups put pressure on the government to promulgate the Orphan Drug Act in 1983. This was the first significant step worldwide as it provided a range of benefits that would encourage manufacturers to develop drugs for populations suffering from rare diseases. It was, indeed, the Orphan Drug Act that facilitated the rapid introduction of several drugs HIV infection to the U.S. market. The act defined a rare disease as one that is present in less than 7.5 per 10,000 of the population. Some of the benefits offered to the pharmaceutical industry included a 50% federal tax credit for clinical research costs, exemption from the FDA registration fee, 7 years market exclusivity, protocol assistance, and congressional grants of approximately $20 million [87]. Data from the first 11 years of the act shows that over 600 products have been given orphan drug status. Of these, 110 were licensed and are currently used to treat 8 million patients. In contrast, only 10 orphan products were developed in the previous decade [88]. Beneficial treatments over recent years have been developed for rare diseases such as Gaucher's disease, cystic fibrosis, hemophilia, HIV/AIDS, multiple sclerosis, and Tourette's syndrome.

Pregnancy is not rare but, sadly, today, pregnant women are a therapeutically orphaned population. The study of drugs in pregnancy is critical for several reasons. First, pregnant women often need drugs to treat conditions that can endanger the life of the fetus or their own. These drugs cannot be used optimally without proper clinical studies that define the pharmacokinetic and pharmacodynamic characteristics in this population and allow physicians to minimize toxicity and maximize efficacy. Second, this is a special population exhibiting unique characteristics not shared by any other, namely, she who hosts a second human being that can be adversely affected by drugs she takes [89].

Hence, clinicians are in a bind. There is a clear need for drugs properly tested in pregnant women in order to use them safely, but they cannot be tested unless there is reassuring data about its safety during pregnancy. How can this cycle be broken?

Considering that more than half of all pregnancies are unplanned, capturing these pregnancies and possible inadvertent exposures to drugs and following them up to term can create large cohorts that teach us about fetal safety. Case control and cohort studies can be gathered to address fetal safety. This can provide a first review of the basic characteristics of drugs in the pregnant population. Useful information about teratogenicity, adverse effects in the mother, and perinatal risk can be obtained and weighted in order to decide whether it might be safe to conduct trials with a given drug. This can be partially addressed by pregnancy registries such as Organization of Teratology Information Specialists (OTIS) and Motherisk. Nevertheless, the numbers of pregnancies captured by these registries are a minute fraction of the actual amount of pregnancies exposed to medications.

Second, preliminary pharmacokinetic information can be extracted from population studies that can be useful as a foundation for further studies. For most drugs, there is a wealth of pharmacokinetic data from the nonpregnant population that can serve as a guide

for the design of future trials, especially if the different physiologic variables that can alter the pharmacokinetics of the drug in pregnancy are taken into account. Cord blood sampling can be performed on neonates whose mothers received the agent to be studied. Both cord blood and maternal levels can be established and collected in a centralized manner in order to derive population pharmacokinetics data [85–88]. Moreover, many women are prescribed by their physicians drugs they need in pregnancy, and there is no ethical hurdle in studying their pharmacokinetics. Indeed, it might be even said that not studying these cases would be unethical, as an experiment is actually taking place, albeit informally, when a patient is prescribed an inadequately studied drug. It is widely held as unethical to perform such experiments and not to record the information they yield. This information would be of enormous benefit to future users of the same drug.

Third, in vitro studies, such as placental perfusion studies, can provide very useful insights into the pharmacological properties of the drug, some aspects of its pharmacokinetics, and characteristics of its placental passage. As shown above, this has proven to be of great value in the case of Glyburide in gestational diabetes [3].

Fourth, ideally, drugs used during the second and third trimester should be studied with a design that allows the subject to serve as her own control (i.e., obtaining data serially during the second and third trimester and then during the postpartum period) [90]. The studies should take into account the physiological variations that occur in pregnant women in order to measure the differences in pharmacokinetic variables (i.e., changes in protein binding, distribution volume, metabolism, half-life, etc.) [90]. New drugs should not be commonly prescribed during the first trimester until proven safe to minimize teratogenic risks.

Finally, a comprehensive list of drugs potentially useful for pregnant women with parallel medical conditions that need to be better tested should be established in order to prioritize those that are potentially most effective and safe. This should emphasize pregnant population needs over industry needs to approve a particular drug. For example, in a perfect situation, antibiotics or antihypertensives would be a much higher priority to test in well-controlled large clinical trials than other medications that address more infrequent problems but are more commercially interesting.

As such, a pregnancy drug act mirroring the already approved Best Pharmaceuticals for Children Act that pushes clinical drug trials in children, or the more recently approved European legislation Medicines for Children, may encourage the pharmaceutical companies to develop drugs should they be awarded patent extensions and liability protection. Furthermore, industry would benefit from a complimentary approach that would have regulatory agencies working hand in hand with them on pregnancy-centered study designs. Without progress on the regulatory front, drugs such as atosiban and Glyburide will unlikely have a chance to fully prove themselves as effective treatments in pregnancy in North America.

## 8 CONCLUSION

Presently, despite a very short list of proven human teratogens, pregnant women are orphaned from the tremendous advances in therapeutics, putting at risk their health as well as the health of their unborn children. Unless society takes a rationale approach, by employing breakthrough methods and models in the study and development of drugs in pregnancy, some of which have been described here, women and their unborn children will continue to be the object of major injustice and exposed to unjustified risks.

## REFERENCES


1. Bloomgardern A (1997). New therapeutic approaches to non-insulin dependent diabetes mellitus. *Endocr Pract 3*(5), 307–312.
2. Langer O, Conway D, Berkus M, Xenakis E, Gonzales O (2000). A comparison of glyburide in insulin in women with gestational diabetes mellitus. *N Engl J Med 343*(16), 1134–1138.
3. Lobstein R, Lalkin A, Koren G (2001). *Pharmacokinetic Changes During Pregnancy and Their Clinical Relevance. Maternal Fetal Toxicology a Clinician Guide*, 3rd ed. Marcel Dekker, New York, pp. 1–21.
4. Agarwal M, Punnose J (2004). Recent advances in the treatment of gestational diabetes. *Expert Opin Invest Drugs 13*(9), 1103–1111.
5. Ender AC, Blakenship TN (1999). Comparative placental structure. *Adv Drg Deliv Rev 38*, 3–15.
6. Pasanen M (1999). The expression and regulation of drug metabolism in human placenta. *Adv Drug Deliv Rev 38*, 81–97.
7. Hokkola J, Pasanen M, Pelkonen O, Huikkanen J, Evisalmi S, Anttila A, Kane M, Mantyla K, Purkunen S, Saarikoski M, Tooming K, Raunio K (1997). Expression of CYP1B1 in human adult and fetal tissues and differential inducibility of CYP1B1 and CYP1A1 by ah-receptor ligands in human placenta and cultured cells. *Carcinogenesis 18*, 391–397.
8. Sastry BVR (1999). Techniques to study human placental transport. *Adv Dr Del Rev 38*, 17–39.
9. Huisman M, Smit J, Schinkel A (2000). Significance of P-glycoprotein for the pharmacology and clinical use of protease inhibitors. *AIDS 14*, 237–242.

10. St-Pierre M, Serrano M, Marcias R (2000). Expression of members of the multidrug resistance protein family in human placenta. *Am J Physiol Regul Integr Comp Physiol 27*, R1495–1503.
11. Gottesman MM, Pastan I, Ambudkar S (1996). P-glycoprotein and multidrug resistance. *Curr Opin Genet Dev 6*, 610–617.
12. Brandes JM, Tavolini N, Potter B, Sarkozi L, Shepard P, Berk D (1983). A new recycling technique for the human placental cotyledon perfusion: Application to the studies of the fetomaternal transfer of glucose, insulin and antipyrine. *Am J Obstet Gynecol 146*, 800–806.
13. Faber J, Thornburg K (1983). *Placental Physiology*, 1st ed. New York: Raven.
14. Smit JW, Huisman MT, Cantellingen O, Wiltshire HR, Schinkel AH (1999). Absence or pharmacological blocking of placental P-glycoprotein profoundly increases fetal drug exposure. *J Clin Invest 104*, 1441–1447.
15. Bournissen G, Feig D, Koren G (2003). Maternalefetal transport of hypoglycemic drugs. *Clin Pharmaco-kinet 42*(4), 303–313.
16. Beck F (1981). Comparative placental morphology and function. *Dev Toxicol 2*, 35–54.
17. Boyd J, Hamilton W (1970). *The Human Placenta*, 1st ed. Cambridge: Heffer.
18. Holcberg G, Tsadkin Tamir M, Sapir O, Huleihel M, Mazor M, Ben Zvi Z (2003). New aspects in placental drug transfer. *Placental Drug Transfer 5*, 873–876.
19. Heikkila A, Renkonen OV, Erkkola R (1992). Pharmacokinetics and placental passage of imipenem during pregnancy. *Antimicrob Agents Chemother 36*, 2652–2655.
20. Schenker S, Dicke J, Johnson R, Mor L, Henderson G (1987). Human placental transport of cimetidine. *J Clin Invest 80*, 1428–1434.
21. Contractor S, Stannard P (1983). The use of AIB transport of assess the suitability of a system of human placental perfusion for drug transport studies. *Placenta 4*, 19–29.
22. Schenker S, Yang Y, Johnson R, Downing J, Shenken S, Henderson G, King T (1993). The transfer of cocaine and its metabolites across the term human placenta. *Clin Pharmacol Ther 53*, 329–339.
23. Krishna R, Levitz M, Dancis J (1993). Transfer of cocaine by the perfused human placenta: The effect of binding to serum proteins. *Am J Obstet Gynecol 169*, 1418–1423.
24. Herman N, Van Decar T, Jones K, Knepe G, Downing J (1993). Transfer of methohexital across the perfused human placenta. *Anesthesiology 79*, A1008 (Abstract).
25. Ala Kokko T, Pienimaki P, Hollmen A, Vahakangas (1993). Placental transfer of lidocaine and bupivacaine in human placental perfusion model. *Anesthesiology 79*, A1007 (Abstract).
26. Sakowski M, Schlessinger J, Dumbroff S, Grant G, Turndorf H (1993). In vitro human placental uptake and transfer of fentanyl. *Anesthesiology 79*, A1006 (Abstract).
27. Sastry B, Janson V, Clark C, Owens L (1997). Cellular toxicity of 2:4:5-trichlorophenoxyacetic acid. *Cell Mol Biol 43*, 559–566.
28. Bourget P, Roulot C, Fernandez H (1995). Models for placental transfer studies of drugs. *Clin Pharmacokinet 28*, 161–180.
29. Sastry B (1995). Opiod addiction, placental function and fetal growth in BCR. In Sastry B, Ed. *Placental Toxicology*. Boca Raton, FL: CRC Press, pp. 83–106.
30. Sastry B (1995). Cacaine addiction, placental function and fetal growth in BCR. In Sastry B, Ed. *Placental Toxicology*. Boca Raton, FL: CRC Press pp. 133–160.
31. Miller RK, Berndt WO (1973). Evidence for a Mg dependent Na K activated ATPase and Ca ATPase in human term placenta. *Proc Soc Exp Biol Med 143*, 118–122.
32. Miller RK, Berndt WO (1974). Characterization of neutral amino acid accumulation in human term placenta slices. *Am Physiol 277*, 1236–1242.
33. Longo LD, Yuen P, Gussek D (1973). Anaerobic glycogen dependent transport of amino acids by the placenta. *Nature (London) 243*, 531–533.
34. Scheider H, Dancis J (1974). Amino acid transport in human placental slices. *Am J Obstet Gynecol 120*, 1092–1098.
35. Enders R, Judd R, Donohue T, Smith C (1976). Placental amino acid uptake. Transport systems for neutral amino acides. *Am J Physiol 230*, 706–710.
36. Miller RK, Koszalka TR, Brent RL (1975). The transport of molecules across placental membranes. In Poste G, Nicholson GL, Eds. *The Cell Surface in Animal Embryogenesis and Development*. Amsterdam: Elsevier, pp. 125–223.
37. Smith G, Depper R (1974). Placental amino acid uptake II (1974). Tissue preincubation fluid distribution and mechanisms of regulation. *Pediatr Res 8*, 679–703.
38. Sastry B, Olubadewo R, Harbison D, Schmidt DE (1976). Human placental cholinergic system: Occurrence, distrubution and variation with gestational age of acetylcholine in human placenta. *Biochem Pharmacol 25*, 425–431.
39. Olubadewo JO, Sastry B (1978). Human placental cholinergic system: Stimulation secretion coupling for release of acetylcholine from isolated placental villus. *J Pharmacol Exp Ther 204*, 433–445.
40. Rowell PP, Sastry B (1978). The influence of cholinergic blockade on the uptake of alpha-aminoisobutyric acid by isolated human placental villi. *Toxicol Appl Pharmacol 45*, 79–93.
41. Sastry B (1984). Amino acid uptake by human placenta: Alternations by nicotine and tabacco smoke components and their implications on fetal growth. In Caciagli R, Biacobini E, Paoletti R, Eds. *Developmental Neuroscience, Physiological, Pharmacological and Clinical Aspects*. New York: Elsevier, pp. 137–140.
42. Barnwell S, Sastry B (1987). *S*-Adenosyl-L-methionine mediated enzymatic methylations in the plasma membrane of the human trophoblast. *Trophoblast Res 2*, 95–120.

43. Smith NC, Brush M, Luckette S (1974). Preparation of human placental villus surface membranes of the human trophoblast. *Trophoblast Res 2*, 95–120.
44. Ruzyski M, Kelley C, Smith C (1978). Placental amino acid uptake. Transport by microvillus membrane vessicles. *Am J Physiol 234*, C27–C35.
45. Smith NC, Brush M, Juckette S (1974). Preparation of human villus surface membrane. *Nature (London) 252*, 302–303.
46. Fant M, Harbison R (1981). Syncytiotrophoblast membrane vesicles: A model for examining the human placental cholinergic system. *Teratology 24*, 187–199.
47. Cemerikie B, Zamah R, Ahmed M (1994). The effect of in vitro methadone administration on trophoblast tissue opioid receptors and mediated responses. *Trophoblast Res 8*, 405–419.
48. Cemerikie B, Zamah R, Ahmed M (1995). Opioid tolerance in human placenta due to in vitro methadone administration. *J Pharmacol Exp Ther 273*, 987–994.
49. Gedeon C, Koren G (2006). Designing pregnancy centered medications: Drug which do not cross the human placenta. *Placenta 27*, 861–868.
50. Daniels-McQueen S, Krichevsky A, Boime I (1987). Isolation and characterization of human cytotrophoblast cells. *Trophoblast Res 2*, 423–445.
51. Hall CS, James TE, Goodyear C, Branchard C, Guyda H, Giroud C (1977). Short term tissue culture of human midterm and term placenta: Parameters of hormonogenesis. *Steroids 30*, 569–580.
52. Kliman HJ, Feinman JF, Strauss III (1977). *Differentiation* of human cytotrophoblast in syncytiotrophoblast in culture. *Trophoblast Res 2*, 407–421.
53. Takeuchi Y, Sakakibara R, Ishiguro M (1990). Synthesis and secretion of human chronic gonadotropin and its subunits in choriocarcinoma cells: A comparative study with normal placental cells. *Mol Cell Endocrinol* 5, *69*(2–3), 145–156.
54. Pattillo, RA (1971). The JAR cell line—continuous human multihormone production and controls. *In Vitro 6*, 398–399.
55. Serrano MA, Macias RI, Briz O, Monte MJ, Blazquez AG, Williamson C, Kubitz R, Marin JJ (2007). Expression in human trophoblast and choriocarcinoma cell lines, BeWo, Jeg-3 and JAr of genes involved in the hepatobiliary-like excretory function of the placenta. *Placenta 28*, 107–117.
56. Yasuda S, Itagaki S, Hirano T, Iseki K (2006). Effects of sex hormones on regulation of ABCG2 expression in the placental cell line BeWo. *J Pharm Pharm Sci 9*(1), 133–139.
57. Jollie WP (1990). Development and morphology and function of the yolk sac placenta of laboratory rodents. *Teratology 41*, 361–381.
58. Enders AC, Blankenship TN, Lantz KC, Enders SS (1998). Morphological variation in the interhemal areas of chorioallentoic placentas. *Troph Res 12*, 1–19.
59. Welsh AO, Enders AC (1987). Trophoblast—Decidual cell interactions and establishment of maternal blood circulation in the parietal yolk sac placenta of the rat. *Anat Rec 217*, 203–219.
60. Welsh AO, Enders AC (1986). Chorioallantoic placenta formation in the rat II. Angiogenesis and maternal blood circulation in the mesometrial region of the implantation chamber prior to placenta formation. *Am J Anat 192*, 263–281.
61. Thom EA, Rouse DJ; National Institute of Child Health and Human Development Maternal-Fetal Medicine Units Network (2003). What we have learned about conducting randomized controlled trials in the NICHD MFMU network. *Semin Perinatol 27*(3), 253–260.
62. Haram K, Mortesen JH, Wollen AL (2003) Preterm delivery: An overview. *Acta Obstet Gynecol Scand 82*, 687–704.
63. Devendra T, Tan LK, Tan HK (2006). Tocolytic treatment for the management of preterm labour: A systematic review. *Singapore Med J 47*(5), 361–366.
64. Richter O, Dorn C, van de Vondel P, Ulrich U, Schmolling J (2005). Tocolysis with atosiban: Experience in the management of premature labor before 24 weeks of pregnancy. *Arch Gynecol Obstet 272*, 26–30.
65. Pierzynski P, Lemancewicz A, Reinheimer T, Akerlund M, Laudanski T (2004). Inhibitory effect of Barusiban and Atosiban on oxytocin-induced contractions of myometrium from preterm and term pregnant women. *J Soc Gynecol Invest 11*, 384–387.
66. Kjos SL, Buchanan TA (1999). Gestational diabetes mellitus. *N Engl J Med 341*, 1749–1756.
67. Adams KM, Li H, Nelson RL, Ogburn PL Jr, Danilenko-Dixon DR (1998). Sequelae of unrecognized gestational diabetes. *Am J Obstet Gynecol 178*(6), 1321–1332.
68. Reece EA, Homko CJ (1998). Diabetes mellitus in pregnancy. What are the best treatment options? *Drug Saf 18*(3), 209–220.
69. Koren G (2001). Glyburide and fetal safety; trans-placental pharmacokinetic considerations. *Reprod Toxicol 15*, 227–229.
70. Kramer J, Feig D, Klein J, Koren G (2006). The transport of glyburide in the human placenta: Implication for treatment of gestational diabetes. *Am J Obstet Gynecol 29*, 1692–1699.
71. Kitzmiller JL, Gavin LA, Gin GD, Jovanovic-Peterson L, Main EK, Zigrang WD (1991). Preconception care of diabetes. Glycemic control prevents congenital anomalies. *JAMA 265*(6), 731–736.
72. Kemball ML, McIver C, Milner RD, Nourse CH, Schiff D, Tiernan JR (1970). Neonatal hypo-glycaemia in infants of diabetic mothers given sulphonylurea drugs in pregnancy. *Arch Dis Child 45*(243), 696–701.
73. Reeve B, Cook B, Babineau D, Scholes C, Buckley DN (2005). Impact of perinatal dialysis dose guidelines on clinical outcomes. *Can J Anesth 52*(1), 55–61.

74. Rogers RC, Sibai BM, Whybrew WD (1990). Labetolol pharmacokinetics in pregnancy induced hypertension. *Am J Obstet Gynecol 162*(2): 362–366.
75. Sanberg JA, Slikker W (1995). Developmental pharmacology and toxicology of anti-HIV therapeutic agents: Dideozynuclosides. *FASEB J 9*, 1157–1163.
76. Mirochnick M (2000). Antiretroviral pharmacology in pregnant women and their newborns. *Ann NY Acad Sci 918*, 287–297.
77. Voit R, Schroder S, Peiker G (1985). Pharmacokinetics studies of azlocillin and piperacillin during late pregnancy. *Chemotherapy 31*, 417–424.
78. Maarten T, et al. (2001). P-glycoprotein limits oral available brain and fetal penetration of saquinavir even with high doses of ritonavir. *Mol Pharmacol 59*(4), 806–813.
79. Owen A, Chandler B, Back DJ (2005). The implication of P-glycoprotein in HIV: Friend or foe? *Fund Clin Pharmacol 19*, 283–296.
80. Forestier F, De Renty P, Peytavin G, Dohin E, Farinotti R, Mandelbrot L (2001). Maternal fetal transfer of saquinavir studied in the ex vivo placental perfusion model. *Am J Obstet Gynecol 185*(1), 178–181.
81. Jones K, Hoggard PG, Sales SK, Khoo S, Davey R, Back DJ (2001). Differences in the intracellular accumulation of HIV protease inhibitors in vitro and the effect of active transport. *AIDS 15*, 675–681.
82. Bawdon RE (1998). The ex vivo human placental transfer of the anti-HIV antinucleoside inhibitor abacabir and the protease inhibitor amprenavir. *Infect Dis Obstet Gynecol 6*(6), 244–246.
83. Connor E, Sperling R, Gelber R (1994). Reduction of maternal-infant transmission of human immunodeficiency virus type 1 with zidovudine treatment. *N Engl J Med 331*, 1173–1180.
84. Blanche S, Tardieu M, Rustin P, Slama A, Barret B, Firtion G (1999). Persistent mitochondrial dysfunction and perinatal exposure to antiretroviral nucleoside analogues. *Lancet 354*, 1084–1099.
85. Atanackovic G, Navioz Y, Moretti M, Koren G (2001). The safety of higher than standard dose of doxylamine-pyridoxine (Diclectin) for nausea and vomiting of pregnancy. *J Clin Pharmacol 41*, 842–845.
86. Maeder T (May 2003). The orphan drug backlash. *Sci Am*, pp. 81–87.
87. Orphan Drug STOA Publication (1999). Available: http://www.europarl.eu.int/stoa/publi/167780/chap5_en.htm.
88. European Organization of Rare Disorders. Eurodis survey of orphan drugs: Availability and pricing in EU member states. Available: http://www.Euroordis.org/article.php3?id_article=487.
89. Health Canada Drugs Directorate (January 16, 1997). Orphan drug policy. Available: http://www.hc-sc.gc.ca/hpfb-dgpsa/tpd-dpt/orphdrug_e.pdf.
90. Elliott B, Schenker S, Langer O, Jonhson R, Prihoda T (1994). Comparative placental transport of oral hypoglycemic agents in humans: A model of human placental drug transfer. *Am J Obstet Gynecol 171*, 653–660.

## SUGGESTED READINGS

Barton JR, Prevost RR, Wilson DA (1991). Nefedipine pharmacokinetics and pharmacodynamics during immediate postpartum period in patients with pre-eclampsia. *Am J Obstet Gynecol 165*, 951–954.

Chamberlain A, White S, Bawdon R (1993). Pharmacokinetics of ampicillin and salmactam in pregnancy. *Am J Obstet Gynecol 168*, 667–673.

Conway D, Gonzales O, Skiver D (2004). Use of glibenclamide for the treatment of gestational diabetes: San Antonio experience. *J Matern Fetal Neonatal Med 15*, 51–55.

De Lange E (2004). Potential role of ABC transporter as a detoxification system at the blood CSF barrier. *Adv Drug Deliv Rev 56*, 1793–1809.

Elliott B, Langer O, Schenker S, Johnson RF (1991). Insignificant transfer of glyburide occurs across the human placenta. *Am J Obstet Gynecol 165*(4, Pt 1), 807–812.

Feig D, Kraemer J, Klein J, Koren G (2004). Transfer of glyburide into breast milk [Abstract]. *Clin Pharmacol Ther 75*, 28.

Frederiksen MC (2002). Drug approval process in the USA. The drug development process and the pregnant woman. *J Midwifery Womens Health 47*(6), 422–425. Available: www.fda.gov, www.emea.eu.int.

Gluek C, Goldenberg N, Streicher P, Wang P (2002). The contentious nature of gestational diabetes: diet, insulin, glyburide and metformin. *Expert Opin Pharmacother 3*(11), 1557–1568.

Hay W (1989). Placental control of fetal metabolism. *R Coll Obstet Gynecol 20*, 33–52.

Marsolini C, et al. (2002). Swiss mother þ child HIV cohort study. Transplacental passage of protease inhibitors at delivery. *AIDS 16*(6), 889–893.

O'Hara MF, Leahey W, Murraghan GA (1983). Pharmacokinetics of sotalol during pregnancy. *Eur J Clin Pharmacol 24*(4), 521–524.

Pacifi GM, Nottoli R (1995). Placental transfer of drugs administrated to the mother. *Clin Pharmacokinet 28*, 235–269.

Page K (1993). *The physiology of the Human Placenta*. London: University College.

Perucca E, Richens A, Ruprah M (1981). Serum protein binding of phenytoin in pregnant women. *Proc Br Pharmacol Soc 11*, 409–410.

Philipson A, Stiernstede G (1982). Pharmacokinetics of cefuroxime in pregnancy. *Am J Obstet Gynecol 142*, 823–828.

Shelley H (1979). *Transfer of Carbohydrate: Placental Transfer*. Tunbridge Wells: Pitman Medical, pp.118–141.

Sibley C, Boyd R (1988). *Control of Transfer Across the Mature Placenta. Oxford Reviews of Reproductive Biology*. Oxford: Oxford University Press, pp. 382–435.

Tamas G, Kerenyi Z (2002). Current controversies in the mechanisms and treatment of gestational diabetes. *Curr Diabetes Rep 2*(4), 337–346.

Yazdanian M (1999). Blood brain barrier properties of human immunodeficiency virus antiretrovirals. *J Pharm Sci 88*, 950–954.

# HISTORY OF ANTIPSYCHOTIC DRUG DEVELOPMENT

Shankar S. Gollapudi[1] and Rajan Radhakrishnan[2]

[1]*College of Pharmacy, Western University of Health Sciences, Pomona, California*
[2]*College of Pharmacy, Roseman University of Health Sciences—Utah Campus, South Jordan, Utah*

## 1 HISTORICAL PERSPECTIVE

Schizophrenia is a devastating mental illness that affects approximately 1% of the world population. The disease is characterized by disordered thinking and behavior, with an equal prevalence in men and women. The symptoms of psychiatric disorder in patients with schizophrenia can be divided into two types, positive and negative. Positive symptoms include hallucinations, delusions, and thought disorder. Negative symptoms represent the loss of qualities normally present in healthy individuals, including diminished emotion, lack of interest, and depressive signs and symptoms. The course of schizophrenia is variable but is usually recurrent and chronic, often causing severe disability. Prior to 1923 only physical methods of treatment were available for schizophrenia in the United States, and in 1923 psychotherapy was introduced in the United States by a European psychiatrist named Adolph Meyers. Psychotherapy was the mainstay of treatment available to manage schizophrenia until the 1950s when chlorpromazine came into the picture serendipitously. The revelation that pharmacological agents can treat schizophrenia led to the explosive growth of psychopharmacology and transformed the "art" of treating mental illnesses to the "science" of modern psychiatry. Currently, there are several classes of antipsychotic drugs available to psychiatrists to treat schizophrenia. Recent studies have shown that taking antipsychotic medications consistently is crucial and necessary to manage the symptoms and to control the risks of relapse. This chapter reviews the history of antipsychotics starting with the birth of typical antipsychotics developed over a half century ago, the current newer safer atypical agents for treatment of schizophrenia, and novel emerging therapies being developed in the new millennium.

## 2 DEVELOPMENT OF PHARMACOTHERAPY OF SCHIZOPHRENIA

### 2.1 Birth of (Typical) Antipsychotics

The discovery in the 1950s that chlorpromazine could effectively be used to treat schizophrenia was serendipitous. The drug, 4560 RP, later known as chlorpromazine (CPZ), was synthesized in 1950 by Paul Charpentier, a chemist at Laboratoires Rhône-Poulenc in France. Chlorpromazine is chemically a phenothiazine derivative with the IUPAC name 3-(2-chloro-10*H*-phenothiazin-10-yl)-*N*,*N*-dimethyl-propan-1-amine (Fig. 1).

It was designed and synthesized as a medication to reduce agitation or anxiety in patients before surgery. It was the observational ability and curiosity of a French surgeon Henri Labroit that brought to light the magical powers of chlorpromazine in treating schizophrenia. In 1952, with the help of some skeptical psychiatrist colleagues, Labroit tested chlorpromazine in a 24-year-old man who had mania. The medication made the patient calm for several hours with his eyes shut (although his facial expression still looked rather maniacal). Within 3

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

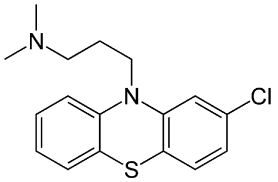


**Figure 1** Structure of chlorpromazine.

weeks he had largely calm and was even able to play bridge [1]. This was followed by several clinical trials in larger patient populations in several countries. Clinically, chlorpromazine significantly decreased positive symptoms in psychotic patients. By 1954 chlorpromazine became popular in Canada, the United Kingdom, and the United States. By 1956, the reduction in the number of inmates in asylums in the United Kingdom fell significantly, and pharmaceutical companies started coming up with several antidepressants and antipsychotics in the next few years. The realization of the clinical efficacy of chlorpromazine in schizophrenia marked the beginning of the golden age of psychopharmacology and psychiatry as a medical science. The treatment of mental disorders started to be based on symptoms, and the science of modern psychiatry began to evolve. The discovery of antipsychotics is sometimes compared to the discovery of penicillin for infectious diseases, anticonvulsants for epilepsy, or antihypertensive drugs for cardiovascular disease. The antipsychotic drugs soon became the cornerstone in the pharmacopoeia for the treatment of psychiatric illnesses. The antipsychotic drugs of this era are called "typical" antipsychotic drugs. However, because of their side effect profiles and substantial limitations, the optimism gradually faded about these drugs—typified by chlorpromazine, a low-potency medication; perphenazine, a drug of intermediate potency; and haloperidol and fluphenazine, high-potency agents. Following chlorpromazine, a number of different compounds were developed called the neuroleptics or typical antipsychotics. The term *neuroleptics* refers to the neurological, motor, or extrapyramidal side effects of these initial antipsychotic drugs [2]. Before the euphoria of the new discovery subsided, clinicians began to realize the troubling side effects of first-generation or typical antipsychotics exemplified by CPZ (Table 1).

Typical antipsychotics have been widely used and shown to be effective in the treatment of positive symptoms of schizophrenia and related psychoses, as well as in preventing psychotic relapses [3]. Typical antipsychotics are broken down into low-potency and high-potency classifications. Low-potency typical antipsychotics, such as chlorpromazine, thioridazine, and mesoridazine, have less extrapyramidal syndrome (EPS) side effects and more muscarinic (dry mouth), α-adrenergic (orthostasis), and histaminic (sedation, weight gain) side effects. In contrast, high-potency typical antipsychotics, including fluphenazine, haloperidol, molindone, thiothixene, trifluoperazine, loxapine, perphenazine, and pimozide, have more EPS side effects and less histaminic, α-adrenergic, or muscarinic side effects.

**TABLE 1 Currently Used Typical (First-Generation) Antipsychotic Agents**

| Low Potency Generic (Brand) | High Potency Generic (Brand) |
|---|---|
| Chlorpromazine (Thorazine) | Perphenazine (Trilafon) |
| Thioridazine (Mellaril) | Trifluoperazine (Stelazine) |
| Mesoridazine (Serentil) | Fluphenazine (Prolixin) |
| | Thiothixene (Navane) |
| | Haloperidol (Haldol) |
| | Molindone (Moban) |
| | Loxapine (Loxitane) |
| | Pimozide (Orap®) |

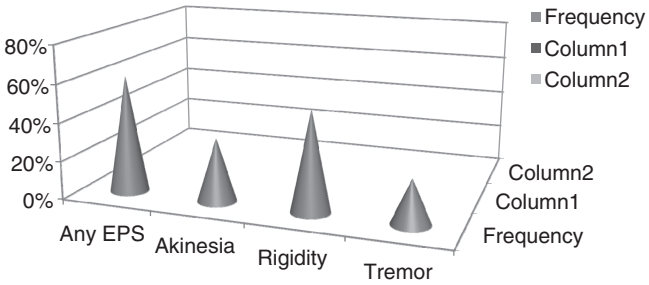


**Figure 2** Frequency of antipsychotic-induced parkinsonism with typical antipsychotics. (Data from PJ Weiden. *DSM_IV*, Vol. 1. APA Press; 1994.)

After the serendipitous discovery of the antipsychotic effects of CPZ, it was found that CPZ, as well as the other typical antipsychotics, improved the positive symptoms in psychotic patients by blocking dopamine receptors in the brain. Another observation that supports the dopamine hypothesis is the development of paranoid schizophrenia in patients with high doses of amphetamine, a drug that causes excess release of dopamine in the brain. This led to the concept that psychosis is caused by increased levels of dopamine and its activity, and most antipsychotic drugs block postsynaptic dopamine receptors [4].

The two most common unfavorable side effects precipitated from the treatments with typical antipsychotics are EPS and prolactemia [2, 5]. Extrapyramidal symptoms include pseudoparkinsonism, dystonia, akathisia, and tardive dyskinesia (Fig. 2). Dopamine is

**TABLE 2 Atypical Antipsychotics (Second-Generation Antipsychotics)**

| Generic (Brand) |
|---|
| Aripiprazole (Abilify) |
| Clozapine (Clozaril) |
| Olanzapine (Zyprexa) |
| Quetiapine (Seroquel) |
| Risperidone (Risperdal) |
| Ziprasidone (Geodon) |
| Paliperidone (Invega) |

involved in several functions (emotion, motor activity, cognition) and also helps to regulate the release of prolactin. Thus blockage of $D_2$ receptors may treat positive symptoms (mesolimbic pathway) but may cause EPS (nigrostraital pathway) and increase of prolactin level (tuberoinfundibular pathway) [5].

The discovery of conventional antipsychotics has marked a major turning point in the history of psychiatry. While drugs in this class were effective in treating and preventing the positive symptoms in individuals with schizophrenia or related psychosis, the negative symptoms and cognitive deficits of the disease in fact worsened.

### 2.2 Development of Atypical Antipsychotics

The absence of an effect of typical antipsychotics on the negative symptoms and the severity of their adverse effects prompted the development of second generation antipsychotics that became known as "atypical" antipsychotics (Table 2). The introduction of atypical antipsychotics (AAs) marked a dramatic change and novel approaches in psychiatric pharmacotherapy. Clozapine was the first medication in this class, which was introduced and approved in the early 1990s for the treatment of psychotic conditions in individuals with schizophrenia. The term *atypical* describes their different dominant characteristics compared to "typical" or "conventional" antipsychotics. The AAs appear to cause far fewer EPS (pseudoparkinsonism and dystonia) and the absence or minimum effect of prolactinemia after being administered chronically, and yet they have the efficient in controlling positive symptoms [6].

Clozapine was developed in 1959 by a Swiss drug company and introduced in some European countries in 1971 and 1972 [6]. It took over 10 years for clozapine to gain its clinical reputation in the European market. The drug was later halted abruptly in 1974 when a rare but potentially fatal hematological (agranulocytosis) side effect was detected in treated patients. Fifteen years later, clozapine was approved by the Food and Drug Administration (FDA) in 1990 for use in the United States, after a large controlled clinical study had demonstrated the superiority of clozapine over conventional antipsychotic drugs. Clozapine has a number of characteristics that make it unique. It is significantly more effective in both positive and negative symptoms of schizophrenia. These days clozapine is only used for the treatment of resistant schizophrenia when at least two atypical antipsychotics failed to act on the patient [3]. The binding of clozapine at $D_2$ receptors is weak compared with conventional agents, and this characteristic may explain the absence of EPS. Other common side effects of clozapine are weight gain and strong sedation, which probably related to the blockage of 5-$HT_{2c}$ (serotonin), adrenergic $\alpha_1$, and $H_1$ receptors. Another common serious side effect is seizures, which occur in 5–10% of patients treated with doses above 500 mg [7].

Since the introduction of clozapine, several other antipsychotics have been approved by the FDA, carrying the descriptive label of "atypical antipsychotic." During the 1990s, olanzapine, risperidone, and quetiapine were introduced, with ziprasidone and aripiprazole following in the early 2000s. The newest atypical antipsychotic, paliperidone, was approved by the FDA in late 2006 [8].

Besides minimal or negligible EPS, all AAs have also been found to be effective in improving negative symptoms and cognitive dysfunction of schizophrenia. Studies showed that patients have experienced an "awakening" state, characterized by a return to a near normal level of cognitive, interpersonal, and vocational functioning and not just significant improvement in positive symptoms as observed with typical antipsychotic agents.

Serotonin-dopamine antagonism, selective $D_2/D_3$, or transient $D_2$ antagonism are key features for clinical actions of atypical agents [9, 10]. More details about the receptor's mechanism of action of atypical agents will be discussed throughout the chapter.

Atypical antipsychotics demonstrate an important advance in the treatment of schizophrenia [9]. Although associated with significantly less extrapyramidal motor side effects (EPS), negligible elevation of prolactin, and efficacy in both negative and positive symptoms [11], the metabolic side effects that have been of most concern to clinicians, patients, and the FDA are excessive weight gain, hyperglycemia, hyperlipidemia, as they may induce type II diabetes and heart disease in susceptible individuals [12, 13]. However, there are differences in their individual side effect profiles that should be taken into account when selecting a treatment. For instance, studies have shown that clozapine and olanzapine can cause significant weight gain and that risperidone may cause reproductive or hormonal side effects (prolactinemia) (Table 3).

Based on clinical experience, ziprasidone and aripiprazole produced less or no effects on metabolism compared to other AAs. This may be explained through partial agonism of $D_{2/3}$ and 5-$HT_{1A}$ receptors of aripiprazole, and high 5-$HT_{2A}/D_2$ receptor binding ratio of ziprasidone [15, 16].

## 3 NEUROPHARMACOLOGY OF ANTIPSYCHOTIC DRUGS

### 3.1 Dopamine Pathways and Dopamine Hypothesis of Schizophrenia

Although the hypotheses on schizophrenia have progressed beyond the dopamine hypothesis, the dopamine hypothesis is still prominent when describing the etiology of schizophrenia. There is a lot of anecdotal evidence for the involvement of brain dopamine in schizophrenia such as manic behavior in people consuming drugs that elevate brain dopamine levels (amphetamine, cocaine). Importantly, the majority of antipsychotic drugs are antagonists at dopamine receptors. Thus, there is no doubt that dopamine and dopaminergic pathways in the brain are involved in the pathophysiology of schizophrenia.

**TABLE 3 Comparison of Weight Gain after 10 Weeks of Antipsychotic Treatment**

| Drug | Mean Weight Gain (lb) |
|---|---|
| Ziprasidone | 0.09 |
| Haloperidol | 1.1 |
| Aripiprazole | 1.6 |
| Risperidone | 4.4 |
| Chlorpromazine | 4.7 |
| Olanzapine | 7.8 |
| Thioridazine | 7.8 |
| Clozapine | 8.9 |

*Source*: Adapted from [14].

The major dopamine pathways in the brain are the nigrostriatal pathway (pathway 1 in Fig. 3), which extends between the substantia nigra and the striatum; the mesolimbic pathway (pathway 2 in Fig. 3), which connects the ventral tegmentum in the midbrain and nucleus accumbens; the mesocortical pathway (pathway 3 in Fig. 3), which extends from the ventral tegmentum to the cortex; and the tuberoinfundibular pathway (pathway 4 in Fig. 3), which originates in the hypothalamus and projects into the pituitary gland.

As discussed before, the most prominent hypothesis regarding the etiology of schizophrenia is the dopamine theory. The ability of typical antipsychotic drugs to diminish positive psychotic symptoms was shown to be initiated by the blockade of the dopamine $D_2$ receptors in the mesolimbic nuclei (second pathway in the diagram), especially the stria terminalis and the extended amygdale.

### 3.2 Dopamine and Serotonin Receptors in Schizophrenia

Most drugs used in schizophrenia have been shown to have affinities for several dopamine receptor subtypes. By antagonizing dopamine receptors, it is possible to block hallucinations during the active phase of schizophrenia [17]. This deactivates the biochemical processes

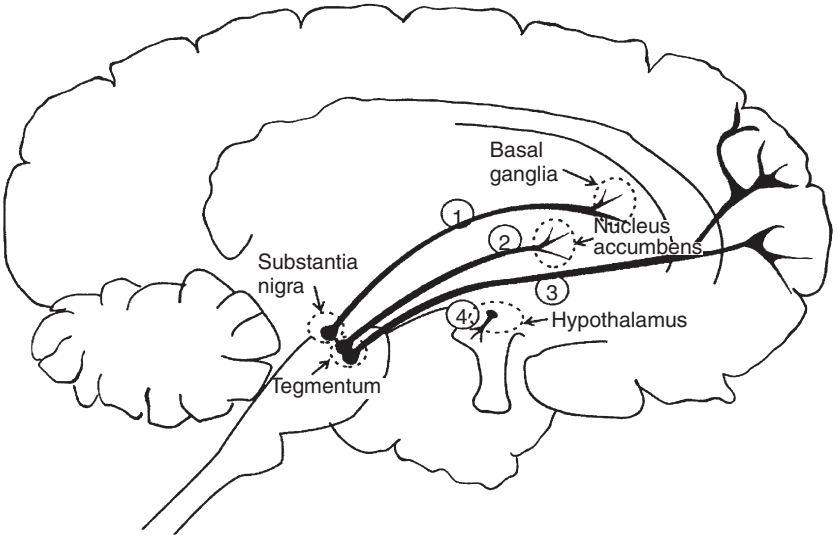


**Figure 3** Dopamine pathways.

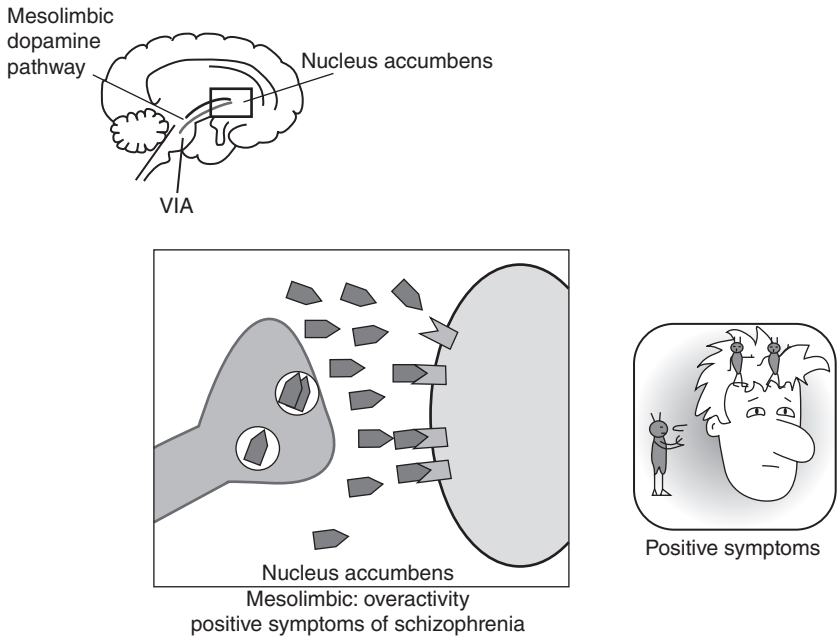


**Figure 4** Dopamine receptors and positive symptoms. (Reproduced with permission from SM Stahl. *Essential Psychopharmacology*, 3rd ed. Cambridge University Press, New York, 2008. © Copyright by the Neuroscience Education Istitute.)

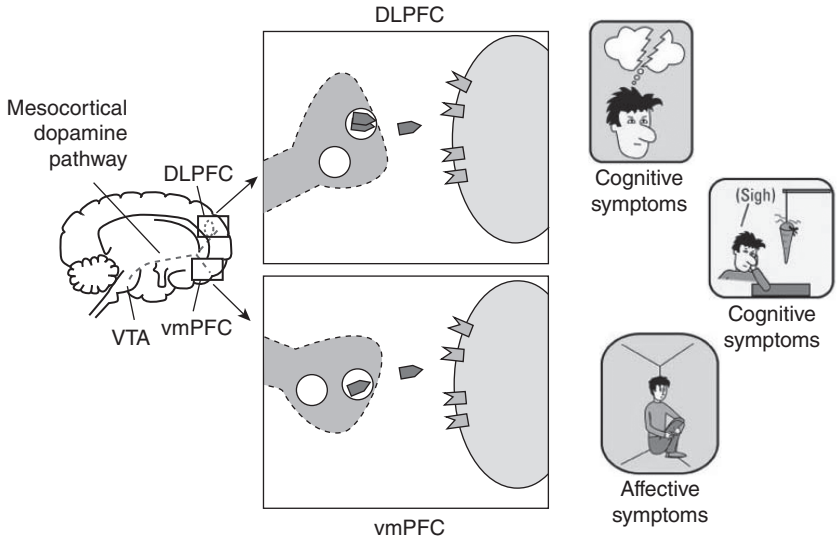


**Figure 5** Dopamine receptors and negative symptoms.

normally initiated by dopamine binding. The brain responds to this receptor blockade by making extra dopamine receptors (Figs. 4 and 5). These extra receptors restore the cell's sensitivity to dopamine. The brain also compensates by increasing dopamine synthesis. The increase in dopamine synthesis lasts 1–2 weeks from the start of therapy, which is the same time required for the medication to become effective. Most effective antipsychiatrics, typical as well as atypical, have affinity for dopamine receptors. When drugs block dopamine receptors in the basal ganglia, the symptoms of schizophrenia are reduced. Five dopamine receptor subtypes ($D_{1-5}$) have been discovered so far. Their function is to bind to dopamine secreted by presynaptic nerve cells. The $D_1$ and $D_5$ receptors are involved in episodic memory, emotion, and cognition, which are disturbed in schizophrenic patients. There is less $D_1$ binding of dopamine in schizophrenic patients as compared to healthy sub-

jects as a result of fewer $D_1$ receptors. Certain antipsychotic drugs stimulate $D_1$-regulated pathways, which increases the $D_1$ to $D_2$ activity balance in the brain. This balance can also be regained by the release of dopamine. The $D_2$ is the second most abundant dopamine receptor in the brain. $D_2$ receptor blockade is the main target for antipsychotic drugs because there is a higher density of $D_2$ in schizophrenic brains. According to studies, there is a loss of $D_3$ expression in the parietal and motor cortices of postmortem, schizophrenic brains. This may be due to either the course of the disease or the therapy given to the patient during the course of the disease. Not much is known about $D_5$ due to the lack of drugs that are selective for it [18, 19].

Haloperidol was synthesized in 1956. Haloperidol is the most widely used typical antipsychotic that binds nonselectively to a broad range of receptors [18]. It can bind to dopamine $D_1$ and $D_2$, 5-$HT_2$, histamine $H_1$ and $\alpha_2$-adrenergic receptors in the brain; however, its major site of action is the $D_2$ receptor. Long-term use of this drug results in an irreversible motor disorder involving the orofacial muscles and extremities. Haloperidol produces more EPS but less histaminic, orthostatic, and anticholinergic side effects. On the other hand, chlorpromazine produces less EPS but more histaminic (sedation), $\alpha_1$-adrenergic (orthostatic hypotension), muscarinic (anticholinergic side effects). Typical drugs have stronger affinity for the $D_2$ receptors rather than the 5-$HT_{2A}$ receptors and do not decrease negative symptoms of schizophrenia. First-generation antipsychotic (FGA) drugs reduce the positive symptoms of schizophrenia. The first-generation antipsychotic drugs cause severe extrapyramidal side effects such as acute dystonic reaction, subacute parkinsonism, akathisia, and tardive kinesia or dystonia, due to strong $D_2$ blockade. These EPS side effects have been and still are major concerns in the therapeutic community. [19].

More recently (1990s), it has become widely accepted that antipsychotic drugs, the second-generation antipsychotics (atypicals), that block both serotonin and dopamine, improve therapeutic outcomes compared with those that mainly block $D_2$ receptors. Serotonin is a neurotransmitter and neuromodulator. Therefore, it can control dopamine levels in the brain. Higher affinity for the 5-$HT_{2A}$ receptor compared to the $D_2$ receptor was the basis for the difference between typical and atypical agents. Another major revolution of receptor bindings between typical and atypical antipsychotic drugs is that atypical drugs with low affinity for $D_2$ receptors (e.g., clozapine and quetiapine) have higher dissociation rates from $D_2$ receptors. Their ability to bind shortly and loosely explains their less EPS side effects. The atypical drugs arranged in order of descending dopamine affinity are as following: risperidone > ziprasidone > olanzapine > clozapine > quetiapine. EPS, which requires very high levels of receptor occupancy, occurs occasionally with risperidone, which is a strong of second-generation dopamine antagonist, but occurs rarely with the other second-generation antipsychotic drugs.

Elevated levels of serotonin (5-HT) are found in the schizophrenia patient. Serotonin has been implicated in a variety of behaviors and somatic functions that are disturbed in schizophrenia: cognition, including memory, perception, and attention, sensory gating, mood, aggression, sexual drive, appetite, energy level, pain sensitivity, endocrine function, and sleep. Many of these functions are clearly relevant to the etiology of positive and negative symptoms and the cognitive impairments that constitute the core abnormalities of schizophrenia. Serotonin receptors include 5-$HT_{1A}$, 5-$HT_{2A}$, 5-$HT_{2C}$, 5-$HT_3$, 5-$HT_6$, and 5-$HT_7$, but most of the antipsychotic drugs are developed on the basis of 5-$HT_{2A}$. Some medications such as clozapine and olanzapine have equivalent affinity for 5-$HT_{2A}$ and 5-$HT_{2C}$, whereas others such as risperidone and ziprasidone have more selectivity for 5-$HT_{2A}$. Based on some evidence, 5-$HT_{A1}$ antagonists improve learning and memory by blocking the inhibitory effects of 5-$HT_{A1}$ receptors on the firing of hippocampal neurons. This suggests that a partial agonist acting as an antagonist may be beneficial for schizophrenia. Atypical drugs also have some antagonist effect on $\alpha_1$ and $\alpha_2$ adrenoceptors. Risperidone, clozapine, olanzepine, zotepine, quetiapine, sertindole, and ziprasidone are potent $\alpha_1$ antagonist. The $\alpha_1$ receptor antagonists have shown to increase dopamine efflux in the shell but not the core of the nucleus in the limbic system. Clozapine and other atypical drugs have shown cardiovascular side effects because of their $\alpha$-adrenergic activities.

### 3.3 Pharmacotherapy of Schizophrenia

The term *atypical antipsychotics* refers to drugs that do not cause extrapyramidal symptoms or tardive dyskinesia in humans. Unlike typical antipsychotics, which may increase levels of prolactin, the atypicals have less or no effect on prolactin levels, and they decrease the negative effects of schizophrenia. These drugs are stronger antagonist of 5-$HT_{2A}$ with lower affinity for $D_2$. Since the newer drugs are more selective on limbic system, they improve cognition as well as positive and negative symptoms. As far as psychosis, both typical and atypical drugs work the same. It appears that the atypical drugs as a class produce substantial weight gain compared with conventional drugs. Histaminic activity correlates with weight gain as well as incidence of diabetes. Atypical antipsychotic drugs have also been associated with alterations in glucose metabolism. In addition, they

cause increased rates of cardiovascular disease, such as hypertension and coronary artery disease [19].

In order to understand how the mechanism of action of typical and atypical antipsychotic drugs differ, a study analyzed the affinities, the occupancies, and the dissociation time-course of various antipsychotics at dopamine $D_2$ receptors and at serotonin (5-HT) receptors, both in the test tube and in live patients [20]. According to this study, the older antipsychotics such as trifluperazine, chlorpromazine, fluphenazine, haloperidol, and flupenthixol bind more tightly than dopamine itself to the dopamine $D_2$ receptor. On the other hand, the newer antipsychotics such as quetiapine, remoxipride, clozapine, olanzapine, sertindole, ziprasidone, and amisulpride all bind more loosely than dopamine to the dopamine $D_2$ receptor. These tight and loose binding data agree with the rates of antipsychotic dissociation from the human-cloned $D_2$ receptor. In this study they measured the time of the dissociation of medication from the receptors. The results for older drugs showed very slow dissociation of about 30 min, while radioactive quetiapine, clozapine, remoxipride, and amisulpride dissociate rapidly, in less than 60 s. This study suggests that atypical drugs work by occupying $D_2$ receptors and then rapidly dissociating to allow normal dopamine neurotransmission. This keeps prolactin levels normal, spares cognition, and lowers EPS side effect. This study also confirms that the newer drugs block 5-$HT_{2A}$ receptors at the same time as they block dopamine receptors. According to the study, the antipsychotic threshold occupancy of $D_2$ for antipsychotic action remains at about 65% for both typical and atypical antipsychotic drugs, regardless of whether 5-$HT_{2A}$ receptors are blocked or not. At the same time, the antipsychotic threshold occupancy of $D_2$ for eliciting EPS remains at about 80% for both typical and atypical antipsychotic drugs, regardless of the occupancy of 5-$HT_{2A}$ receptors.

Abilify (aripiprazole) is an antipsychotic drug with a new mechanism of action. It works as a partial agonist against dopamine $D_2$ receptors and differs from other antipsychotics in this respect. Pharmacological studies demonstrate that Abilify can be a high-affinity partial agonist, a full agonist, or an antagonist at the dopamine receptor. It shows inhibitory action when large amounts of dopamine are released in the brain and stimulatory action when only small amounts of dopamine are released. Abilify is known as a dopamine system stabilizer (DSS). It also improves both positive and negative symptoms of schizophrenia considered to be caused by dopamine abnormalities [21]. It is expected that this drug can be used for long periods since it is less likely to cause sleepiness or weight gain. There are several other important components, such as glutamatergic neurotransmission, and an array of receptor subtypes for both dopamine and serotonin. Although antipsychotics can also have antimuscarinic and antihistaminic properties and cause $\alpha_1$-adrenergic and $\alpha_2$-adrenergic blockade, these properties are thought to be related to different side effect profiles rather than to antipsychotic efficacy. The pharmacological profile of Abilify differs from typical and atypical drugs in being a partial agonist as opposed to an antagonist at the dopamine receptor. Abilify has a low serotonin/dopamine affinity ratio and a high affinity for dopamine receptor. Second-generation antipsychotic drugs have been distinguished from typical agents by a high serotonin/dopamine affinity ratio and a low affinity for dopamine receptor.

In conclusion, typical antipsychotic drugs such as haloperidol had shown their effectiveness mainly by blocking dopamine ($D_2$) receptors in the mesolimbic system. Atypical drugs have promising results due to both their serotonin and dopamine antagonist activities. Clozapine was the first atypical antipsychotic drug that introduced the important role of serotonin receptors in the effectiveness of antipsychotic medication. Typical antipsychotic drugs have a stronger blocking property for the serotonin receptor relative to dopamine blocking property. Controlling dopamine and dopamine receptors is essential for the treatment of schizophrenia. Partial dopamine receptor agonists, which may act as an agonist at presynaptic dopamine receptors and antagonist at postsynaptic dopamine receptors, are a new class of drugs that have promise in the future. Partial dopamine agonists may modulate dopamine neurotransmission more adequately than pure antagonists by producing a mix of direct agonist and antagonistic effects. Multireceptor agents appear to be more promising as antipsychotic agents for the majority of psychiatric patients. Atypical compounds are in early stages of development. Among them is iloperidone (Zomaril), by Novartis. Zomaril is one of a new class of drugs known as serotonin/dopamine receptor antagonists, which inhibit serotonin and dopamine activity by binding to these receptors in the brain. Specifically, it is a mixed 5-$HT_{2a}$ /dopamine $D_2$ antagonist. This drug is in phase III clinical trial. Solvay Pharmaceuticals is in phase III testing of Bifeprunox (DU127090), a mixed dopamine $D_2$ partial agonist and 5-$HT_{1a}$ partial agonist. This atypical antipsychotic targets both the positive and negative symptoms of schizophrenia. ACP-103, a potent and selective 5-$HT_{2A}$ inverse agonist, is in phase II testing by ACADIA Pharmaceuticals as adjunctive therapy. The hope is that adding ACP-103 to currently available antipsychotic agents may achieve the best combination of dopamine receptor blockade and 5-$HT_{2A}$ inverse agonism. In theory, this should reduce adverse effects while expanding the range of efficacy [17, 19].

Research into glutamate drugs for schizophrenia has focused on compounds that can enhance *N*-methyl-D-aspartate (NMDA) receptor function as well as those that can reduce glutamate release. Based on a study, inhibition of glycine reuptake by sarcosine, an antagonist of the glycine transporter 1, was more effective at reducing both positive and negative symptoms than was direct activation of the glycine/D-serine site by D-serine [22]. Together with direct activators of the NMDA receptor, future treatments are likely to include drugs that enhance the production of endogenous modulators of the NMDA receptor, such as serine racemase, an astrocytic enzyme synthesizing D-serine. Interestingly, expression of this enzyme was recently reported to be altered in schizophrenia. The potential to target activity and expression of a multitude of synaptic and extrasynaptic molecules involved in synthesis and control of NMDA receptor function could produce new drugs with enhanced efficacy in patients with schizophrenia [23]. A member of a new class of antipsychotic drugs, LY2140023 is still in early stage clinical development. Lilly's LY2140023 differs from established antipsychotic drugs by targeting glutamate rather than dopamine or serotonin receptors. The result of studies suggested that LY2140023 can be a safe and well-tolerated drug with fewer side effects of hyperprolactinaemia, extrapyramidal symptoms, and weight gain (see the Eli Lily website) [24].

In order to improve treatment outcome, it is important for clinicians to be familiar with different receptor activity of old and new generation of antipsychotic drugs. By knowing the receptor activity, it is easy to predict the side effect profile of medication, which would increase the chances of achieving the best possible treatment outcome.

## 4 ADVERSE EFFECTS OF ANTIPSYCHOTICS

Antipsychotic medications are associated with a variety of adverse effects in addition to their therapeutic effects in controlling symptoms of mental illness. Recently, much of the discussion has been focused on the association of psychotropic medications, specifically second-generation antipsychotics (SGAs) and metabolic effects such as glucose dysregulation (Table 4) and lipid disorders. Due to the impact of diabetes mellitus and hyperlipidemia on coronary heart disease, clinicians need to be aggressive in monitoring the emergence of these adverse events. Here we discuss the effects of psychotropic medications on metabolic function, sexual function, prolactin levels, and cardiovascular health.

**TABLE 4 Prevalence of Type 2 Diabetes Mellitus in Patients Receiving Antipsychotic Agents**

| Drug | Patients with Type 2 DM (%) |
|---|---|
| Fluphenazine | 4.5 |
| Risperidone | 6 |
| Haloperidol | 6.6 |
| Olanzapine | 11 |
| Clozapine | 15.5 |

### 4.1 Metabolism and Weight Gain

Patients with mental illness have a higher prevalence of weight gain and obesity compared to patients without mental illness. Major complications associated with obesity are cancer, congestive heart failure, coronary heart disease, diabetes mellitus (DM), dyslipidemia, gallstones, genitourinary problems, hypertension, menstrual problems, sleep apnea, and stroke. Obesity-related illnesses cost the nation $117 billion in 2000 alone and are responsible for as many as 300,000 premature deaths each year. Obesity can be 1.5–2 times more prevalent in people with schizophrenia and affective disorders than in the general population. Recent findings suggest that psychotropic medications used to treat this population also contribute to weight gain.

There are several hypotheses on the metabolic effect of antipsychotics. They are believed to act on the hypothalamic feeding center and directly stimulate appetite. Recent evidence suggests that antagonism of serotonin 2C (5-$HT_{2C}$), dopamine (DA), adrenergic ($\alpha$), histamine type 1 ($H_1$), and glutamine receptors may play a role in antipsychotic-related weight gain. Other neurotransmitters such as acetylcholine as well as metabolic and biologic mechanisms may also be involved. Apart from these, environmental, psychological, and lifestyle factors are definitely key contributors to the development of weight gain in patients treated with antipsychotics.

Agents that stimulate 5-$HT_{2C}$ receptors—such as fenfluramine and *m*-chlorophenyl-piperazine (mCPP) (a metabolite of trazodone and nefazodone)—have been associated with weight loss. Conversely, agents that antagonize 5-$HT_{2C}$ and $H_1$ receptors have been hypothesized to increase appetite and cause weight gain. Studies demonstrate an exponential relationship between antipsychotic $H_1$ receptor affinities and maximum weight gain. Clozapine and olanzapine have very strong binding affinities to 5-$HT_{2A}$, 5-$HT_{2C}$, and $H_1$ and muscarinic ($M_1$) receptors and less affinity for the $D_2$ receptors. These neurotransmitter systems and receptor subtypes appear to help regulate food intake and energy homeostasis and may play a role in weight gain associated with antipsychotic use. The preferential antagonism of 5-$HT_{2A}$ versus $D_2$ receptors may also

explain the advantage of SGAs versus FGAs in alleviating negative symptoms of schizophrenia and better side effect profile. There is also recent evidence that polymorphism of the 5-$HT_{2C}$ receptor gene may be involved in protection against antipsychotic-induced weight gain [25].

Three peptides are implicated in the regulation of food intake, energy expenditure, and body energy balance: galanin, neuropeptide Y, and leptin. Among the three, leptin has been more extensively investigated. Leptin, a member of the interleukin-6 cytokine family, is found in multiple tissues and secreted by adipose cells. Leptin plays an important role in the regulation of insulin secretion and energy metabolism in fat cells and skeletal muscle. Since leptin levels are chronically increased in obese individuals, obesity may be associated with malfunctioning of leptin receptors, a condition called leptin resistance. In patients with new-onset DM induced by antipsychotic agents, leptin levels are elevated. There is evidence with clozapine that the increase in leptin may be directly due to the drug, not just due to weight gain. A small study challenged this correlation and found no difference in leptin levels between patients with chronic schizophrenia and matched controls [26]. Improving responsivity to leptin in obese patients may lead to weight loss, although the efficacy of exogenous leptin administration in decreasing weight is yet unknown. In addition, there are numerous other neurotransmitters and peptides that have been implicated in the regulation of appetite, eating behavior, and control.

Although weight gain is associated with both FGAs and SGAs, studies have shown that there are differences in weight gain potential among antipsychotic agents [27]. Long-term treatment with SGAs has demonstrated that aripiprazole and ziprasidone are associated with less weight gain, while olanzapine and clozapine are associated with more weight gain. Patients with lowest baseline body mass index (BMI) have the greatest weight gain, independent of dose. Clozapine is associated with greater mean weight gain than olanzapine, risperidone, or ziprasidone. Risperidone—with a lower weight gain potential than most atypicals—displays high affinity for 5-$HT_{2A}$ and $\beta_1$ receptors and a lower affinity for 5-$HT_{2C}$ receptors than for $D_2$ receptors. The weight gain potential of ziprasidone is even less than that of risperidone and the exact mechanism is unknown.

### 4.2 Lipid Abnormalities

The association between SGAs and hypertriglyceridemia has been established in a wide range of studies. In addition to age, sex, and low- (LDL) and high-density lipoproteins (HDL), plasma triglycerides (TG) is an independent risk factor for the development of coronary atherosclerosis and coronary heart disease (CHD). Due to the increased prevalence of heart disease and associated mortality in the schizophrenia population, the potential for SGAs to increase TG is an important concern.

Both hyperlipidemia and hypertriglyceridemia are thought to be associated with insulin resistance. In extreme cases of hypertriglyceridemia (i.e., development of pancreatitis), the expected physiologic response is the decreased production of islet cells and decreased glucose utilization. Although all atypicals (with the exception of ziprasidone and aripiprazole) increase serum TG to some degree, severe hypertriglyceridemia occurs predominantly with clozapine and olanzapine. Although both drugs have favorable efficacy profiles, the mechanism of their antipsychotic activity may include altering the various lipid pools. Ziprasidone and aripiprazole usually do not elevate lipid levels.

In a small clinical study, researchers evaluated the effects of olanzapine on weight, fasting total cholesterol (TC), and TG in 25 patients with schizophrenia and related psychoses. After 12 weeks of olanzapine treatment, the group mean body weight increased 5.4 kg ($P < 0.02$) [28]. Although there was no significant change in cholesterol levels, the mean TG levels rose by 37%. In a 5-year naturalistic study of patients treated with clozapine, researchers found that weight gain increased until 46 weeks after drug initiation [29]. There was no significant increase in total cholesterol and significant increase in serum TG levels. A univariate analysis also correlated elevated serum TG levels with increased risk for coronary artery disease. Clozapine was also associated with the greatest increases in weight, lipid levels, and blood glucose levels. Triglyceride levels have been reported to be increased in numerous studies involving clozapine and olanzapine. In one study of four patients with psychotic disorders, serum TG levels were decreased after switching therapy from clozapine to risperidone [30]. In two of the four patients, clozapine was reinstated after risperidone was discontinued and serum TG levels increased. This finding is consistent with previous reports involving clozapine rechallenge. In a study with olanzapine and risperidone, olanzapine was found to have 4.1 kg weight gain compared to risperidone (2.3 kg) over 7 months [31]. Long-term data (greater than 8 weeks) for quetiapine and weight gain is not available.

### 4.3 Diabetes Mellitus

Diabetes mellitus is a serious, life-threatening illness that can lead to heart failure, stroke, renal failure, and blindness. Recent data collected by the Schizophrenia Patient Outcomes Research Team (SPORT) suggest that people with schizophrenia may be at increased risk

for type II DM due to side effects of antipsychotic medications (especially SGAs), poor overall physical health, sedentary lifestyles, poor diet, and inadequate health care [32]. Schizophrenia and bipolar disorders have been associated with an elevated risk of DM (primarily type II) and disturbances in glucose regulation. A recent Veterans Affairs study found that the rate of diabetes (diagnosed) in men with schizophrenia was six to eight times higher than men without schizophrenia as estimated by the National Health and Nutrition Examination Survey (NHANES) study in 1998 [33, 34]. Type II DM as well as impaired glucose tolerance tests have been associated with some conventional as well as some atypical antipsychotic agents. Individuals with schizophrenia are also at higher risk for developing diabetic ketoacidosis due to the lack of consistent medical care or due to drugs that have a propensity to increase weight.

Several mechanisms may explain the association between SGAs and DM. Both olanzapine and clozapine show great affinities for 5-$HT_{1C}$ and 5-$HT_{2C}$ receptors. One study postulated that clozapine and olanzapine may damage pancreatic islet cells and cause dysregulation of the sympathetic nervous system. In addition, decreased serum insulin by way of depressed pancreatic β-cell responsiveness to serum glucose may be related to 5-$HT_{1A}$ antagonism seen with these newer antipsychotic agents. In a recent study, 15% of patients with a first episode of schizophrenia (medication naïve) had impaired fasting glucose and insulin resistance compared to matched controls [35].

The major problem for patients taking antipsychotic medications is that antipsychotic medications can increase weight and increase levels of adiposity. In patients with schizophrenia, intra-abdominal fat (observed by computed tomography) was significantly elevated versus matched controls [36]. The acute stress of psychosis may also affect glucose and insulin levels. In the same study, cortisol levels were significantly elevated in patients with schizophrenia compared to controls. Increasing adiposity increases the risk of type II DM and hyperglycemia, and this occurs via increases in insulin resistance. The main abnormality of DM is hyperglycemia, but this can occur due to defects in insulin secretion, insulin action, or both.

When antipsychotic medications were first introduced with chlorpromazine in the 1950s, several reports were published regarding phenothiazines and their potential to aggravate existing DM and/or induce new-onset type II DM. For example, the introduction of chlorpromazine increased the prevalence of DM in one clinical setting from approximately 4% to approximately 17% [37]. Various groups reported these abnormalities in glucose control, including the National Diabetes Data Group, which added phenothiazines to the list of medications that could disturb glucose metabolism in the 1970s. Among atypical antipsychotics, there are also differences in the risk of causing glucose dysregulation. Dibenzodiazepines (clozapine, olanzapine, quetiapine) are associated with DM occurring early in the course of therapy, even without significant weight gain (0–10 lb) (Table 4).

With the emergence of SGAs, there have been increased reports of new-onset DM, exacerbations of existing DM, impaired glucose tolerance, and complications such as diabetic ketoacidosis (DKA) associated with these medications. Although there are reports of new-onset or exacerbation of DM related to many of the atypical agents, clozapine and olanzapine are the ones most frequently cited (Table 4). Clozapine and olanzapine cause an increase in body weight more than other agents, and this in itself increases the risk for diabetes. However, there appear to be other mechanisms involved that are independent of weight. In a study of patients with schizophrenia treated with clozapine, olanzapine, risperidone, or FGAs, patients on olanzapine had significantly elevated glucose levels at all time points compared with those on typical agents and untreated controls [38]. There was no significant increase in risperidone-treated patients compared with patients on FGAs; however, fasting and postloading glucose levels were greater in patients on risperidone when compared with normal controls. There were treatment-related increases in insulin levels, suggesting that the primary defect may be due to tissue sensitivity rather than a secretory defect.

In an epidemiologic study of 19,637 patients with schizophrenia in the United Kingdom, the odds ratio (OR) for the development of DM while taking olanzapine was nearly 6 times the rate in those not receiving any antipsychotics [39]. The OR for those on risperidone and FGAs were approximately 2 and 1.4, respectively. In a retrospective review of patients treated with an SGA who have also been treated for DM, 11 out of 14 patients had acute elevations in glucose levels (fasting blood glucose > 145 mg/dL) after being treated with clozapine, olanzapine, or quetiapine [40]. Five patients developed DKA and were referred to a tertiary care institution for treatment. The median time to onset of DKA was 33 days.

Researchers have noted that treatment with clozapine results in a cumulative risk of diabetes. In a 5-year study of clozapine-treated patients, new cases of DM continued to rise throughout the observation period [29]. By the end of the study, 36.6% of the sample had been diagnosed with DM. There were also significant increases in weight and lipid levels in patients receiving clozapine. Federal regulations now require manufacturers of six SGAs to include labeling language

warning about an increased risk of patients developing diabetes.

### 4.3 Monitoring of Weight Gain, Lipids, and Diabetes and Therapeutic Considerations

Since most of the weight gain with antipsychotic agents occur within the first year, monitoring of patients during this period is crucial. Clinicians need to monitor weight by including a BMI at baseline and periodically thereafter. In addition to baseline BMI, fasting glucose, waist circumference, lipids, and blood pressure need to be monitored. Patients also need to be educated in terms of lifestyle patterns and monitoring of their own weight. Signs and symptoms of DM and DKA are also important patient education points. Clinicians also need to assess risk factors for DM and coronary artery disease as outlined by national guidelines (Table 5).

The level of monitoring for glucose abnormalities also needs to be heightened among clinicians. This will be especially important in patients with risk factors such as increased BMI, family and personal history of glucose intolerance, smoking, and other cardiovascular (CV) risk factors. Beyond mere monitoring of weight, clinicians who understand the neuropharmacology of obesity may be able to recommend alternative drugs or facilitate drug switches to alleviate or reverse psychotropic drug-induced weight gain. To manage psychotropic drug-induced weight gain, one can either treat the weight gain or switch medications. Switching to an agent that does not lead to weight gain is conceivable but not always realistic. All available substitutes may share the same undesirable qualities. Some patients can, however, have very different responses to drugs that share the same mechanism, so switching to a drug within the same class may be an option. This may be particularly true for the typical antipsychotics loxapine and molindone, which seem to generate the least amount of weight gain among the antipsychotics. A weight loss program that produces a loss of 0.5–1% body weight per week is considered safe and acceptable. Moderate obesity can be managed by reducing food intake by 500 calories with dietary consultation and exercising 30 min each day. For patients with metabolic syndrome, correcting the underlying causes of the features, such as obesity, physical inactivity, and lipid dysfunction is recommended.

**TABLE 5 Monitoring and Assessment Parameters, Education, and Management**

*Monitoring and Assessment Parameters*

- Weight and Height (BMI) → BL, 4, 8, 12 weeks, quarterly
- Waist circumference → BL, annually
- Fasting glucose → BL, 12 weeks, annually
- Fasting lipid panel → BL, 12 weeks, every 5 years
- Blood pressure → BL, 12 weeks, annually
- Assess DM risk factors
- Assess CV risk factors

*Education*

- Nutrition and weight management
- Physical activity
- Signs/symptoms of DM
- Signs/symptoms of DKA

*Management*

- If on an FGA, consider a SGA
- Avoid clozapine and olanzapine in patients with risk factors for CHD, DM, or in obese patients or in patients receiving drugs that may increase weight
- If ≥ 5% weight gain, consider switching (by cross-titration) to another antipsychotic; continue with weight loss education and nonpharmacologic management
- If develop severe hyperglycemia and/or signs/symptoms of DKA, refer immediately for medical management
- If BMI ≥ 30 kg/m$^2$ or BMI ≥ 27 kg/ m$^2$ + cardiovascular risk factors, consider pharmacotherapy for weight loss (e.g., amantadine, nizatidine, orlistat, sibutramine, topiramate)

***Hyperprolactinemia*** Hyperprolactinemia is a concern with antipsychotic medications, and compliance may often be adversely affected with this negative association. The link between antipsychotics and hyperprolactinemia has significant implications for the short- and long-term outcomes in the treatment of schizophrenia. Compared to conventional antipsychotics, SGAs are much less likely to increase prolactin levels (Table 6).

**TABLE 6 Effects of Antipsychotics on Prolactin Elevation**

| | $D_2$ Binding Affinities | Prolactin Elevation |
|---|---|---|
| *First-Generation Antipsychotics* | | |
| Chlorpromazine | High | Moderate |
| Fluphenazine | High | Moderate |
| Haloperidol | High | Moderate |
| Perphenazine | High | Moderate |
| Thioridazine | High | Moderate |
| *Second-Generation Antipsychotics* | | |
| Clozapine | Moderate | 0 |
| Olanzapine | Moderate | 0 |
| Quetiapine | Minimal | 0 |
| Risperidone | High | 0 to moderate (dose-related) |

[a]$D_2$—dopamine receptor type 2.

Prolactin is a polypeptide hormone secreted by the lactotroph cells of the anterior pituitary gland. The primary physiologic role of prolactin is the induction of lactation. However, prolactin interacts with other central nervous system (CNS) and peripheral processes. The tuberoinfundibular pathway is involved in the regulation of prolactin secretion through DA neurons. Blockade of DA receptors in the tuberoinfundibular pathway would reverse prolactin inhibitory effects and lead to hyperprolactinemia.

The central blockade of the $D_2$ receptor by antipsychotics is responsible for the pathogenesis of antipsychotic-induced prolactin release. The main physiological effects of prolactin are the suppression of gonadal activity and induction of the synthesis and secretion of milk. Hyperprolactinemia is seldom detected but can have serious consequences. The clinical effects usually occur when levels are between 30 and 60 mcg/L or higher. Some of the manifestations of hyperprolactinemia include decreased bone mineral density and osteoporosis, galactorrhea, breast enlargement, anovulation, infertility, amenorrhea, increased risk of cardiovascular disorders, increased risk of tardive dyskinesia, increased risk of breast cancer, mood disorders (anxiety, depression, hostility), decreased testosterone levels and sperm mobility, sexual dysfunction, decreased libido and arousal, and an- or hypo-orgasmia.

Premenopausal women may present with menstrual disturbances, galactorrhea, breast engorgement, sexual dysfunction, and infertility. Men may experience gynecomastia, priapism, loss of libido, erectile/ejaculatory dysfunction, and azoospermia. Galactorrhea often goes unnoticed and is therefore underreported. Not all women with hyperprolactinemia suffer from galactorrhea, and, conversely, the symptoms may be present in women with normal levels of prolactin. Galactorrhea may also occur in men although to a lesser degree. Decreased bone density is a long-term consequence of hyperprolactinemia mediated by relative or absolute deficiency of estrogen and testosterone. Prolactin levels inversely affect estrogen and testosterone levels in both men and women. If prolactin is increased, testosterone and estrogen are decreased. This in turn is associated with decreased bone density, which may predispose patients to osteoporosis. Estrogen may have protective effects against arteriosclerosis, hypertension, and cholesterol. The low estrogen levels that commonly occur with hyperprolactinemia may increase the rate of cardiovascular disorders in this group.

The normal range of prolactin levels in nonlactating subjects is between 1 and 25 mcg/L. One study established that prolactin levels in patients with schizophrenia are normal prior to receiving treatment for psychosis, alluding to the fact that the schizophrenia itself does not appear to affect prolactin [41]. Risperidone and FGAs raise prolactin levels, but clozapine, quetiapine, and olanzapine do not due to their lack of dopamine blockade within the tuberoinfundibular tract. Second-generation antipsychotics as a group cause less hyperprolactinemia than conventional antipsychotics with the exception of risperidone. The decreased risk for hyperprolactinemia may be due to weaker dopamine $D_2$ antagonist effects or higher ratios of 5-$HT_2$ to $D_2$ antagonist effects of SGAs. Risperidone at higher doses has been shown to produce increases in prolactin similar to conventional antipsychotics, although at lower doses, risperidone may have higher elevations compared to haloperidol. At the opposite end of the spectrum, clozapine and quetiapine produce minimal sustained increases in prolactin that are similar to placebo. Correlations between prolactin elevations and clinical symptoms have not been established. High-potency antipsychotics have the most impact on serum prolactin level since they are the most potent $D_2$ receptor blockers. Studies comparing the effects of haloperidol and quetiapine concluded that quetiapine was not associated with increases in prolactin secretion [42]. Clozapine, the prototypical atypical antipsychotic, is not associated with hyperprolactinemia. One trial found that male patients with elevated prolactin levels and treatment-refractory schizophrenia had a decrease in prolactin levels after being switched to clozapine [43]. In one study, 20 female patients with schizophrenia on risperidone were switched to olanzapine [44]. After the switch, prolactin levels were significantly reduced, demonstrating the reversal of hyperprolactinemia in patients previously treated with risperidone. Sustained elevations in prolactin levels were not observed with quetiapine or ziprasidone. In addition, aripiprazole has not been associated with an increase in prolactin levels.

*Monitoring* Hyperprolactinemia can be defined as plasma prolactin levels of >20 mcg/L for men and >25 mcg/L for women. Prolactin levels should be measured at baseline when initiating agents that have the potential to increase prolactin. If clinical symptoms emerge, prolactin levels may be useful as a monitoring parameter. Only the presence of clinical symptoms, not an elevated prolactin level alone, should direct changing a patient's drug treatment strategy.

### *Cardiovascular Disturbances*

*Epidemiology* The prevalence of cardiovascular disease is high in schizophrenic patients. This increased risk has been attributed to multiple causes ranging from platelet aggregation to poor living conditions and life-

style differences such as diet. Antipsychotic medications are associated with cardiovascular side effects that can have serious consequences to the patient. Risk factors for cardiovascular adverse effects with the atypical antipsychotics include known cardiovascular disease, electrolyte disturbances (particularly hypokalemia and hypomagnesemia), genetic characteristics, increasing age, female sex (for QTc lengthening and torsade de pointes), autonomic dysfunction, high doses, the use of interacting drugs, and psychiatric illness itself.

*Clinical Evidence*

BLOOD PRESSURE Hypotension is a significant side effect encountered with SGAs. Adrenergic blockade may result in a number of risk factors, such as syncope, fractures, increased anginal episodes, and orthostatic hypotension. The agents that most commonly cause hypotension include clozapine, quetiapine, risperidone, and olanzapine. Olanzapine does not block $\alpha$-adrenergic receptors and has not been linked to orthostatic hypotension, although dizziness has been reported in some patients. Risperidone can occasionally cause orthostatic dizziness, hypotension including orthostatic hypotension, tachycardia including reflex tachycardia, and hypertension. The least hypotensive effects are reported with ziprasidone and haloperidol. Antipsychotic drugs in combination with cardiovascular medications such as methyldopa, diuretics, adrenergic blockers, calcium antagonists, angiotensin-converting enzyme inhibitors, angiotensin-II receptor blockers, nitrates, and others may aggravate the hypotensive effects.

CONGESTIVE HEART FAILURE AND MYOCARDITIS Some of the SGAs report congestive heart failure as a direct cardiac effect. This effect has been reported infrequently (1/100–1/1000) with olanzapine and rarely (<1/1000) with clozapine and quetiapine. The vasodilator effects of the atypicals may counteract any major hemodynamic deterioration. Reports of myocarditis and cardiomyopathy associated with clozapine have raised concerns that this agent may be associated with other forms of cardiac toxicity. In January 2002, Novartis Pharmaceuticals Corp. reported 213 cases of myocarditis, 85% of which occurred while patients were taking recommended doses of clozapine within the first 2 months of therapy. Monitoring for cardiac toxicity is particularly challenging because its manifestations (e.g., tachycardia, fatigue, orthostatic hypotension) are frequently observed in clozapine-treated patients. This is especially true during dosage titrations. Clozapine received a black-box warning in February 2002 for the increased risk of fatal myocarditis, especially but not limited to the first month of treatment. Clozapine should be discontinued if a patient presents with new-onset fatigue, dyspnea, tachypnea, fever, chest pain, palpitations, and symptoms of heart failure or electrocardiograph findings of ST-T wave abnormalities or arrhythmias. Clozapine is also contraindicated in patients with cardiac failure. In patients suffering from cardiovascular disorders, the initial dose should be 12.5 mg given once on the first day, and dosage increase should be done slowly and in small increments.

QTc INTERVAL PROLONGATION QT interval is an electrocardiograph (ECG) measure that describes both depolarization and repolarization of the ventricles. QT interval is usually corrected for heart rate (QTc) since the interval shortens with increasing heart rates. QTc intervals less than 440 are considered normal, and QTc intervals greater than 500 ms place the individual at a significantly increased risk for serious arrhythmias called torsade de pointes (TdP). The arrhythmia can arise from myocardial ischemia, myocarditis, long QT syndrome, and a variety of other causes. They are often self-terminating but can be rapidly fatal if they persist. It would be safest to find a drug that has no effect on the QTc interval, but that alternative does not always exist. It is important to realize that QTc interval prolongation is a warning, not the risk itself. Not all drugs that prolong the QTc interval produce TdP and sudden death. All SGAs can prolong QT intervals to some degree by reducing the flow of repolarizing $K^+$ currents, ultimately making the myocardium more excitable.

QTc prolongation has been reported for all atypical antipsychotics with varying incidences and degrees of severity. Atypical antipsychotics should be used with caution in patients who are being treated with drugs that prolong the QTc interval (e.g., erythromycin, clarithromycin, antiarrhythmic drugs, etc.) During clinical trials, ziprasidone was noted to increase the QTc more than other atypical agents. The best available data come from a Pfizer study upon the FDA's request comparing the effect of ziprasidone, risperidone, olanzapine, quetiapine, thioridazine, and haloperidol on QT interval [45]. After the initial washout period and baseline ECGs, patients were then randomly assigned to receive one of six antipsychotics. For each drug, except thioridazine, the highest dose recommended in the package insert was administered. Blood levels were obtained and repeat ECGs were recorded at the time of maximum drug blood level. In the final phase of the trial, a metabolic pathway inhibitor of that drug was coadministered and repeat ECGs were obtained. Although the dose of thioridazine was only half of the recommended dose, the drug produced the most prolongation. It is interesting to note that risperidone prolonged the QTc interval by almost 12 ms, even though there is no convincing

evidence in the literature that it has ever caused torsade de pointes or sudden death. The FDA has used QTc intervals greater than 500 ms as a clinically significant cutoff.

It is important to complete a cardiovascular assessment before prescribing an antipsychotic and to consider ordering an ECG for patients with personal or a family history of syncope, personal history of angina pectoris, myocardial infarction, congestive heart failure, cardiac arrhythmia, hypomagnesemia, or significant risk factors. An ECG also is required or strongly recommended before prescribing the antipsychotic drugs most commonly associated with QT prolongation and torsade de pointes [e.g., droperidol, haloperidol in large doses intravenously (IV) such as 100 mg/day, mesoridazine, pimozide, and thioridazine).

The FDA has strengthened the warning labels required for these agents, adding black-box warnings about the risks of prolonged QTc intervals, torsade de pointes, and sudden death for droperidol, mesoridazine, and thioridazine. Thioridazine, for example, is indicated only for patients with schizophrenia who fail to show an acceptable response to other antipsychotic drugs. Its use is contraindicated in patients who take fluvoxamine, propranolol, and pindolol, any drug that inhibits the cytochrome P450 2D6 isoenzyme (e.g., fluoxetine, paroxetine), and agents known to prolong the QTc interval. Use of thioridazine also is contraindicated in patients known to have reduced levels of the cytochrome P450 2D6 isozyme, as well as in patients with congenital LQTS or a history of cardiac arrhythmias.

Of the SGAs, ziprasidone is frequently cited to have increased risk for prolonged QTc interval, although there is no black-box warning for this drug. Although the FDA does not require a baseline ECG monitoring for patients initiated on ziprasidone, clinicians are urged to look for warning signs of prolonged QTc. In addition, electrolyte disturbances (e.g., $K^+$ depletion) that may increase the risk of TdP should be aggressively treated to prevent morbidity and mortality. Psychiatrists are advised to read the warnings and prescribing information in the labeling of all antipsychotics for potential cardiovascular side effects.

CEREBROVASCULAR EVENTS Although SGAs are not approved for the treatment of dementia, these agents at low doses are often used to control the behavioral disturbances associated with dementia. Recently, manufacturers of risperidone and olanzapine have announced that these drugs may be associated with increased rates of cerebrovascular events including stroke and transient ischemic attacks.

The data for risperidone comes from four placebo-controlled trials in 1230 patients [46]. Cerebrovascular adverse events were twice as frequent in the risperidone group (4%) compared to the placebo group (2%). A Dear Healthcare Provider letter was sent in April 2003 by Janssen Pharmaceuticals regarding this finding. Eli Lilly and Company also sent a letter to U.S. health care providers in January 2004 due to study findings that the drug was associated with a 3.5% occurrence of death in elderly patients with dementia compared to 1.5% death in the placebo group [47]. Currently, the package labeling of risperidone and olanzapine include the risk of cerebrovascular events, including stroke in elderly patients with dementia. Risperidone and olanzapine have not been shown to be safe or effective in patients with dementia-related psychosis or behavioral disorders. Physicians should be aware of this potential adverse event and advise patients regarding this potential risk.

Mental illnesses are devastating chronic conditions and usually require long-term treatment to not only treat the initial presentation but to prevent recurrent episodes. Although psychotropic medications have many adverse effects, recent evidence suggests that they also have serious long-term complications that can increase rates of morbidity and mortality in patients with psychiatric disorders. Metabolic disturbances, especially weight gain and glucose abnormalities among atypical antipsychotics, are particularly concerning since patients with schizophrenia and affective disorders are at increased risk for developing obesity and diabetes mellitus. Cardiac disturbances, with the introduction of ziprasidone, cannot be ignored, although typical antipsychotics have been known to carry this risk. Sexual dysfunction and hyperprolactinemia are by no means less concerning than other adverse effects, and they can often lead to treatment nonadherence, relapse, and possibly increased morbidity and/or mortality. Clinicians should be aware of potential adverse effects of psychotropic medications and patient risk factors that may increase their likelihood of developing or exacerbating that condition. Primary prevention is usually the best management option for these serious adverse effects.

## 5 CURRENT DEVELOPMENTS AND FUTURE PROSPECTS

### 5.1 Focus on Neurodegeneration in Schizophrenia

Schizophrenia is associated with abnormal brain development. It is a potentially disabling disorder because of its capacity to destroy our normal capacities for perception, reasoning, and interpersonal relationships. Brain morphological alterations in schizophrenic patients have led to the neurodevelopmental hypothesis; there

is a growing consensus that the disorder may also involve neurodegeneration [48]. Many studies have shown progressive morphological changes in schizophrenic patients. Also, neuroanatomical abnormalities have been observed in patients with schizophrenia. Neurodegenerative process in schizophrenia has been found in some studies. The etiology of neurodegeneration in schizophrenia has not been clearly understood. It must be determined if neurodegeneration is associated to schizophrenia itself or it is caused by antipsychotic medications or genetic factors. Also, there is an intense discussion about possible mechanisms of neurodegeneration in schizophrenia.

Antipsychotic drugs are widely used in schizophrenic patients with the purpose to suppress the negative and positive symptoms. However, there are many adverse effects that have been observed in patients using antipsychotic medications. Many studies have found that longer periods of using antipsychotic medication could be associated with poorer outcomes. Metabolic effects of atypical antipsychotics have resulted in calls to consider the public health implications of new-onset diabetes mellitus, weight gain, and hyperlipidemia. Antipsychotic drugs have been associated with structural brain changes in the development of tardive dyskinesia and with cognitive dysfunction. Moreover, antipsychotic drugs could cause neurotoxicity, decreased neuroplasticity, apoptosis as well as the more specific cellular changes in the brain that could lead to neurodegeneration in schizophrenic patients [49].

Antipsychotic medications can affect the structure and function of neurons via a number of processes. These include neuroplasticity in which gradual cellular changes develop in response to the environment. Several investigators have reviewed how the brains of schizophrenic patients react to the presence of an antipsychotic. They found the loss of striatal neurons and increasing in neuronal size and gliosis, and neurodegeneration in the substantia nigra, 10–20% decreasing in neuronal density in striatum, loss of dendritic spines, and increasing numbers of axon terminals [49]. Neurotoxicity is cellular damage, with hydrolysis of cell membranes and an inflammatory response. For example, haloperidol has led to decreased cell viability, membrane blebs, and retraction of cell extensions in primary hippocampal neurons. Also, it has led to acute reversible injury to dendritic microtubes in the anterior parietal region, the central gray matter, and the hippocampus [49].

Gene expression is also known to be involved in neurodegeneration in schizophrenia [48, 50]. Some genes and chromosomal regions have already been associated with a greater risk for the illness, although no single gene or genes have been definitively identified as causal for schizophrenia. The gene for brain-derived neurotrophic factor (BDNF) has been implicated in the risk for schizophrenia. Each human gene has one or more different structural variants (alleles) that are determined in part by past mutations and their permanent incorporation into the human genome. This is also true for the BDNF gene. Schizophrenic patients with one or two *met* BDNF alleles (instead of the more common *val* allele) had measurably lower abilities to perform tasks of learning and memory, and had abnormal rCBF in medial temporal regions (including the hippocampus) during the performance of tasks of learning and memory. Also, they have lower levels of *N*-acetylaspartate. The *met* BDNF allele's extra load of diminished performance on functions of learning and memory may contribute to the cognitive impairments seen in subjects with schizophrenia [50]. Other recent studies have found that allelic variation in the *Notch4* gene decreases in subjects with schizophrenia [51].

Morphological alterations in the brains of schizophrenia patients have been found mainly in the cerebellum, frontal and temporal lobes, nucleus accumbens, mediodorsal thalamus, and other brain regions [52]. There have been several reports of reductions in the gray matter volume of cortical structures [51, 53]. Reports of decreases in the volume of the entire temporal lobe [52] have been followed by reports of hippocampal volume reductions [51, 54–56] in patients with schizophrenia. Similarly, reductions in the volume of the parahippocampal gyrus were reported [51]. A general decrease of brain mass from the onset of the disorder has also been reported [52]. These pathological findings may be observed since the earlier psychotic episodes, and they are more evident in patients who have suffered multiple episodes of the disease [56]. Thus, they are not likely to be due to chronic antipsychotic medication.

The cellular basis of gray matter volume losses in schizophrenia is not yet known, it appears that it is more likely due to the loss or disorganization of neuronal processes. Several postmortem studies have reported abnormal neuronal organization within corticolimbic structures in schizophrenic patients [51, 53]. Reduced number of neurons has been reported in the anterior cingulated cortex, nucleus accumbens, hippocampus, and thalamus from people who died with schizophrenia (6). Also, the gray matter losses in schizophrenic patients have been hypothesized to involve abnormal increases in apoptosis.

Recent research using noninvasive neuroimaging techniques, such as magnetic resonance imaging (MRI), suggests that some patients with schizophrenia show progressive losses of gray matter in the frontal and temporal lobes of the brains. In addition to this, neuroimaging studies have demonstrated anatomical abnormalities

in a network of neocortical limbic regions and interconnecting white matter tracts. Bilateral ventriculomegaly and decrease brain volume exist in medial temporal areas such as the hippocampus and amygdale [51, 55].

Some follow-up neuroimaging studies have shown a progressive ventricular enlargement in schizophrenic patients [51, 54]. The neurobiologic process of ventricular enlargement is difficult to explain, but posited mechanisms include response to medication or chronic stress, increased vascular perfusion and metabolism, increased size or numbers of neurons and glia, increased connective tissue, increased intercellular fluid, and increased dopaminergic neurotransmission.

Several investigators using MR spectroscopy (MRS) have found that schizophrenic patients have differences in the concentrations of *N*-acetyl aspartate (NAA) and various phoshomonoester (PME) and phosphodiester (PDE) moieties in specific brain regions including the temporal and prefrontal cortex. NAA is an abundant amino acid located in neuronal cell bodies and processes. PMEs are phospholipids involved in membrane synthesis, and PDEs are by-products of lipid metabolism that are believed to reflect membrane turnover. Investigators found NAA reduced in the temporal cortex, PME concentrations reduced in the prefrontal cortex, and PDEs elevated in the prefrontal cortex primarily [57]. These results are interesting and may reflect pathophysiologic processes in schizophrenia.

Accumulating evidence suggests that BDNF is involved in the pathophysiology of schizophrenia. BDNF belongs to the neurotrophic factor family and is highly concentrated in the nervous system and the serum. It plays a critical role in the development, regeneration, survival, and maintenance of the function of neurons and plasticity of multiple neurotransmitter systems. Reduced neurotrophic factor signaling could lead to both an altered development of the central nervous system and neuronal death [52, 58].

Recent studies have revealed a strong link between BDNF signals and dopaminergic (DA) function in the brain [49, 58]. Specially, BDNF is involved in the neurodevelopment of dopaminergic-related systems, interacts with the mesolimbic DA systems, and also involved in the therapeutic response to antipsychotic medications. Recent postmortem studies have shown that BDNF is reduced in the hippocampus and prefrontal cortex of patients with schizophrenia [58]. Recently, two studies revealed decreased BDNF levels in the serum of chronic schizophrenic patients [58]. Some studies have found that both atypical and typical antipsychotic drugs such as haloperidol, clozapine, and risperidone decreased BDNF concentrations in frontal, occipital cortex, and hippocampus of patients with schizophrenia [49]. Tan et al. [58] have reported that BDNF were significantly lower in medicated patients with chronic schizophrenia than in healthy control subjects. Antipsychotic treatment might have affected levels of BDNF. So, decreased BDNF in patients with schizophrenia might be responsible for the neurodegeneration due to antipsychotic drugs.

Reduced expression of BDNF and its receptor could also contribute to an enhanced susceptibility for apoptotic death in schizophrenia since this neurotrophin possesses neuroprotective properties [52]. Reduced BDNF has been reported in the prefrontal cortex from schizophrenic brains and in serum of schizophrenic patients [58]. Glutamatergic hypofunction could lead to reduced BDNF expression since activation of non-NMDA receptor (NMDAR) leads to a 10-fold increase in BDNF messenger ribonucleic acid (mRNA) content in hippocampal neurons. Also, kynurenic acid (an endogenous antagonist of glutamatergic receptors) blocks the increased BDNF mRNA content induced by the activation of glutamatergic receptors in hippocampal neurons, and thus glutamatergic hypofunction could also be related to the decreased BDNF expression. Furthermore, glutamatergic hypofunction could lead to decrease antiapoptotic signaling mediated by the Akt protein [52].

Gliosis has been reported in schizophrenia for at least 20 years [51, 52]. Many studies have tried to detect gliosis in the brains of patients with schizophrenia. However, little evidence of a gliotic reaction typical of a neurodegenerative process has been found [51]. Glial proliferation has been found in the subiculum and orbitofrontal cortex of the brains of people with schizophrenia and dementia compared to those of people with schizophrenia without dementia, independently of neuroleptic exposure. The result of this study was not significantly different from control and study samples. It suggested that a neurodegenerative process leading to gliosis could be present in patients with chronic schizophrenia [52].

Several possible mechanisms of neuronal injury are now being investigated in patients with schizophrenia [51, 52]. Excitotoxicity, for example, neurodegeneration via the overactivity of excitatory neurotransmission, is an attractive mechanism to explain neuronal injury. Since it could be initiated and/or maintained through the action of neurotransmitter systems, such as the monoamines, that have long been implicated in schizophrenia [51]. Excitotoxic process is induced by the production of dopamine-generated reactive oxygen species, including superoxide, hydroxyl radicals, and oxidized glutathione. All of which can cause pathological changes in neurons via oxidation of deoxyribonucleic acid (DNA), proteins, and lipids [49].

The release of dopamine is involved with glutamatergic neurons and NMDA receptor (NMDAR) activation. Since glutamatergic neurons in the medial prefrontal cortex project to the nucleus accumbens and reduce

dopamine release through action on metabotropic glutamate receptors (mGluR). Glutamatergic neurons also stimulate GABAergic interneurons through NMDAR. Some studies found that dopamine in striatal decreased secondary to an increased release of glutamate, which led to excitotoxic damage in the globus pallidus and substantia nigra [49].

NMDAR antagonism could also lead to neurotoxicity. Some investigators have suggested that NMDAR blockage reduces cortical inhibitory tone, thus increasing the firing rate of glutamatergic neurons leading to excitotoxic death at postsynaptic neurons through non-NMDAR. Overstimulation or blockade of glutamatergic neurotransmission in different metabolic pathways could lead to neurodegeneration in schizophrenia [52].

Another possible mechanism leading to neurodegeneration is apoptosis [49, 51] or programmed cell death, which can occur rapidly and under a number of circumstances including dopamine depletion or in the absence or decrease of BDNF. The process to apoptosis involves activation of intracellular proteases via mitochondria or tumor necrosis factor. Some studies have been hypothesized that cellular mechanisms responsible for such gray matter losses may be involved in apoptosis [51]. Apoptosis may also relate to gene expression. For example, significantly reduced content of the antiapoptotic protein Bcl-2 and increased Bax/Bcl-2 ratio (associated to susceptibility for apoptotic death) in Brodmann's area (BA) and ultrastructural changes in oligodendroglial cells suggestive of apoptotic death have been found in the brains of people who died with schizophrenia [49, 52]. Moreover, apoptosis may be induced by antipsychotic medications. For example, a single dose of haloperidol given to rats resulted in a substantial increase in apoptotic GABAergic cells in the substantia nigra and striatum within 48 h. Also, DNA fragmentation, cell shrinkage and nuclear membrane disintegration were observed in patients taking haloperidol [49].

Mitochondrial dysfunction [49, 52] may be one mechanism underlying apoptosis induced by antipsychotic drugs. A single dose of haloperidol in mice has led to inhibit complex I of the mitochondrial respiratory chain and led to oxidative stress in a number of brain areas [49]. Some investigators have suggested mitochondrial dysfunction is associated to excitotoxicity. Mitochondrial dysfunction reduces Na/K ATPase activity, increasing membrane depolarization. NMDA receptor becomes voltage activated and magnesium is extruded from the ion channel permitting calcium influx, which in excess may lead to excitotoxic death. Mitochondrial dysfunction leads in turn to dopaminergic overactivity by decreasing dopamine uptake [52].

Dopamine depletion may be another mechanism underlying apoptosis since dopamine receptor stimulation increases NMDAR activity and excitotoxicity. This mechanism could be involved in dopamine-induced nitric oxide (NO) synthesis. NO is synthesized by the enzyme NO synthase (NOS) after NMDAR activation. Some studies have found evidence for a significant reduction in NO biosynthesis in patients with schizophrenia. This result may not be caused by antipsychotic medications. Reduced NO biosynthesis may lead to some susceptibility for neuronal death because NO is associated to both neurotoxic and neuroprotective effects [52].

The effects of stress may provide a possible explanation for the neuronal structure change [52, 55]. Oxidative stress may play a role in the pathophysiology of most neurodegenerative disorders. Some studies have suggested that oxidative stress could be associated with the development of tardive dyskinesia and might be caused by antipsychotic treatment [49]. Other studies have found a reduced total glutathione content in the cerebrospinal fluid (CSF) and prefrontal cortex of neuroleptic-treated schizophrenic patients, which may reflect reduced antioxidant defense and oxidative stress [52]. It suggested the presence of oxidative stress may lead to cell death. On the other hand, oxidative stress has been associated with impaired learning and might be involved in the cognitive deficit of patients with schizophrenia.

Neurocognition is severely impaired in schizophrenia. Patients with schizophrenia appear to perform less well on most cognitive tasks and show greater levels of impairment in selected functions such as attention, working memory, verbal memory, executive functions [48, 59, 60]. Some studies have found that schizophrenic patients, who have more episodes, more severe psychotic symptoms over sustained periods, experience severe cognitive decline. Also, patients with longer duration of untreated psychosis have greater cognitive dysfunction than first-episode patients [60]. Another study found that T cell deficiency might lead to cognitive dysfunction in schizophrenic patients [61].

The evidence of neural degeneration in the pathophysiology of schizophrenic illness suggests that there may be treatment opportunities through neural protection. Neuroprotection, which refers to treatment that helps maintain central nervous system functionality in response to neurobiologic stress, may be responsible for prevention of disease progression and deterioration. Based on the neurodegeneration in schizophrenia and evidence from in vivo neuroimaging studies, it is reported that haloperidol treatment was associated with a greater rate of cortical gray matter loss than olanzapine. It suggested that olanzapine has neuroprotective effect since it might slow the progressive loss of cortical gray matter in patients with schizophrenia [51]. Also, other studies stated that haloperidol has been

associated with reactive oxygen species (ROS) generation and other neurotoxic effects [49, 52]. It leads to an increased basal glutamate concentration in the striatum and nucleus accumbens [49, 52]. It seems that typical drugs do not have good neuroprotective effects for schizophrenic patients.

In contrast to typical antipsychotics, atypical antipsychotics such as olanzapine and clozapine have shown neuroprotective properties and cognitive improvement in patients with schizophrenia [62]. Those drugs have been shown to reduce apoptosis, cell death, and to increase BDNF levels and protein levels [49]. Recent studies reported that olanzapine and clozapine have increased levels of BDNF in the brain of schizophrenic patients [62]. Olanzapine protects against methamphetamine-induced neurotoxicity and reduces apoptotic cell death induced by trophic factor withdrawal. Both olanzapine and clozapine increase Bcl-2 protein and mRNA levels in the frontal cortex and hippocampus. Bcl-2 protein is protective against neuronal death induced by several mechanisms and has been shown to decrease in the temporal cortex of brains from people who died with schizophrenia. Neuroprotective effects of atypical antipsychotics may be associated with the inhibitory effect of 5-$HI_{1A}$ receptors. Also, atypical antipsychotics such as olanzapine and clozapine appear to improve cognitive performance in schizophrenic patients [52]. Keefe et al. [63] reported that olanzapine has a beneficial effect on neurocognitive function in patients with a first episode of psychosis.

Many recent studies have yielded valuable data indicating the possibility of significant side effects of antipsychotic medications on the cellular function and structure of the brain in schizophrenia. The neurodevelopmental hypothesis has emerged over the past few years as a popular and useful approach to understanding and studying schizophrenia. Hypothesis of neurodegeneration in schizophrenia should be studied further since we only know possible mechanisms of neurodegeneration in schizophrenia, yet we do not know the etiology of neurodegeneration. Whether there are progressive brain changes over time, or there is the possibility of long-term antipsychotic drugs that induce the neurodegeneration, or it is due to genetic, environmental, and developmental factors needs to be further explored.

### 5.2 Neuroprotection as Future Therapeutic Goal in Schizophrenia

According to Lieberman, schizophrenia is a continuous neurodegenerative process as patients degenerate slowly from early "prodromal symptoms" to "psychosis" where a person is unable to function in society. Neuronal degeneration usually occurs after the onset of psychosis. It is also more difficult to treat patients who have relapsed from previous therapy. It is believed that this is due to degeneration of the neurons. Thus, it is important to have a neuroprotective agent as part of the treatment therapy at an early stage of the disease in order to slow down or prevent the neurodegeneration process. By doing so, it will improve patient outcomes and minimize the chance of relapse [64]. This opens the door to a novel mode of treatment for schizophrenia called neuroprotection. Ehrenreich defined neuroprotection as "the effort to maintain the highest possible intactness of cellular interactions/intercellular communication in the brain resulting in an overall undisturbed function [65]." The mechanism of neuroprotection is basically to prevent the deterioration of neuronal function in patients with schizophrenia. Researchers are looking at the role of atypical antipsychotics, glutamate, metabotropic glutamate receptor agonist, erythropoietin (EPO), and estrogen as future antipsychotic therapies for schizophrenia with neuroprotective effect.

As mentioned before, people with schizophrenia suffer from loss of gray matter. When compared with normal individuals, schizophrenics have more gray matter loss. Brain images showed that people with schizophrenia at a later stage have more loss in gray matter than people at the early stage [66]. According to Lieberman et al., atypical antipsychotics, such as olanzapine, have been proven to have the ability to not only slow down the degeneration of the gray matter, but they also increase gray matter in the temporal cortical after 3 months. The study also found that patients receiving atypicals has greater improvement in cognition than patients on typical antipsychotics such as haloperidol. Olanzapine has also been demonstrated to reduce neuronal loss in the hippocampus in rats treated with kainic acid [67]. In his study, Bian et al. [67a] reported that perospirone, quetiapine, and ziprasidone also protect the neurons by inhibiting the production of toxic NO generated by microglial cells. Another study conducted by Dazzan et al. [68] showed that patients taking typical antipsychotics have gray matter loss in the temporal gyrus, while those on atypicals have significant enlargement of the thalamus. Thus, atypical antipsychotics seem to have an important and promising role in neuroprotection in patients with schizophrenia.

The pathology of schizophrenia involves many neurotransmitters such as dopamine, serotonin, and glutamate. The first class of medication in treating schizophrenia targeted the dopamine pathway. This was followed by atypical antipsychotics, which targeted mainly serotonin and dopamine pathways. Recent studies are looking at the role of glutamate in the pathology of schizophrenia. It is the major neurotrans-

mitter of the cortex, limbic system, thalamus, hippocampus, and amygdala. It has been shown that schizophrenia can be caused by hypofunction of the glutamate receptors in the cortical area, specifically the NMDA receptors [69]. In the normal individual, NMDA receptors have inhibitory GABAergic neurons that normally inhibit the overproduction of dopamine in the mesolimbic dopamine pathway. With schizophrenic individuals, the GABAergic neurons are inhibited by the hypofunction of glutamate receptors, thus leading to disinheriting of glutamate. As a result of the disinhibition, glutamate receptors are stimulated to the point that it can result in glutamate excitoxicity and hypersecretion of dopamine. This theory was observed in patients abusing phencyclidine (PCP), which is an NMDA antagonist. In healthy individuals PCP induces schizophrenic signs and symptoms such as psychosis. Excessive firing of the glutamate neurotransmitter can lead to detrimental effects on neurons because too much glutamate is released. Glutamate activation is followed by the release of large quantities of sodium and calcium, allowing them to enter the neuron and initiate the chemical reactions that ultimately result in death of the neurons. Thus, glutamate receptor agonists, glutamate release inhibitors, AMPA agonists and antagonists, and ampakines are beneficial for schizophrenia and are good candidates for potential antipsychotics.

Glycine is an example of glutamate receptor agonists. The binding of glycine to its receptors located on the NMDA receptors is necessary in order for glutamate neurotransmission to occur. Therefore, increasing glycine concentration could potentially activate more NMDA receptors and reverse the hypofunction of the glutamate receptors. Studies have shown that at doses of 30–60 g/day, glycine can significantly improve negative symptoms [69, 70]. Its limitation is that glycine has poor blood–brain barrier penetration. D-serine is another glycine agonist, but it is more permeable at the blood–brain barrier than glycine. At a dose of 30 mg/kg/day in addition to neuroleptics, the combination significantly increased negative, positive, and cognitive functions. Thus, D-serine seems to be the most promising [69]. Another mechanism to increase glycine levels is by inhibiting the glycine transporter (GLY-T1). By inhibiting the GLY-T1 reuptake pump more glycine is available to bind to the NMDA receptors. Sacrcosine is a GLY-T1 inhibitor that has been shown to have positive impact on negative symptoms as well as cognition [71]. However, one study in mice showed that mice without GLY-T1 died due to respiratory depression [72].

Metabotropic glutamate (mGlu) receptors are also potential targets for the development of neuroprotective drugs. mGlu2/mGlu3 (found in the cortex, hippocampus, cerebellar cortex) and mGlu5 (found in the hippocampus, cerebellar cortex, thalamus nuclei, and spinal cord) receptors have been examined and the result seems promising [73]. Activation of these receptors results in the inhibition of glutamate release, inhibition of calcium and potassium channels, and stimulation of the formation of astrocytes. The limitations are overstimulation of mGlu5R can cause epilepsy, convulsion, and neurotoxicity, and activation of mGlu2/mGlu3 can result in the inactivation of the CREB cycles [74].

Yale University is currently recruiting patients for a randomized, double-blind, placebo-controlled study to look at the efficacy of riluzole in the treatment of refractory schizophrenia. Riluzole is currently indicated for the treatment of Lou Gehrig's disease. It is a severe and rapidly progressing neurodegenerative disease that affects neurons in the brain and spinal cord. Riluzole exerts its neuroprotective effects by inhibiting the release of glutamate, inhibiting the voltage-gated sodium channels, and inhibiting protein kinase C. Riluzole could be a promising new neuroprotective drug for patients suffering from schizophrenia as schizophrenia is also a progressing neurodegenerative disease [75]. Another class of drug that is in the early stage of clinical trials is the ampakines and CX516 belongs to this class of drug. CX516 enhances NMDA receptors indirectly through the AMPA receptors. Significant improvements in cognition, attention, and memory were seen in schizophrenic patients in two clinical studies when CX516 was combined with clozapine [69].

Erythropoietin has been widely used in the hospital setting for the treatment of anemia for many years. Not until recently has it been shown to have neuroprotective effects. Under normal conditions, its presence is low; however, when neurons are damaged or under stress, its presence is upregulated. It was hypothesized that the upregulation of EPO helps to protect the degeneration of neurons. Clinical studies have shown that erythropoietin has antiapoptotic, antioxidative, anti-inflammatory, neurotrophic, angiogenic, synaptogenic, and glutamate inhibitory properties [76]. With such promising mechanism of action, EPO could be a potential "add-on" agent for neuroprotection in patients with schizophrenia. EPO receptors are overly expressed in the hippocampus and cortex of schizophrenic patients [77]. EPO can cross the blood–brain barrier and exerts its action locally in the brain. A pilot study called Gottingen EPO-Stroke showed that patients treated with EPO have better outcomes such as less neurological loss, suffer from smaller brain lesion, and less glial damage (low S-100B levels) [78]. More importantly patients on EPO also recover to baseline faster. In a 2-year multicentered double-blind, randomized placebo-controlled study, also conducted by Ehrenriech et al. [79] demonstrated that patients on EPO as add-on have significantly better

cognitive performance than placebo. This study also showed a significant reduction in glial damage (low level of S100B) [79].

Epidemiologically, the prevalence of schizophrenia is equal for males and females. Males have earlier onset of schizophrenia than females. However, as women go through menopause, more females are diagnosed with schizophrenia than males. During menstruation or pregnancy, women experience symptoms. This is thought to be associated with a decrease in estrogen and that estrogen acts as a protective agent. Like the atypical antipsychotics, estrogen exerts its action on both dopamine and serotonin receptors. Estrogen has been shown to reduce dopamine concentration in the striatum and modulate the serotonin system by increasing the serotonin receptor transporter. In addition, it also increased NMDA receptors [80] and has antioxidant properties that prevent oxidative stress on the neurons. The overall mechanism of neuroprotection of estrogen is described in the flowchart shown in Figure 6.

The concept of neuroprotection for the treatment of schizophrenia seems to be promising. Those agents such as glycine and D-serine, which target the NMDA receptors, have the most potential in the development of a new class of antipsychotics. Atypical antipsychotics have been shown to be an effective treatment for schizophrenia; however, there are people who relapse from this therapy alone. An "add-on" agent such as EPO might be beneficial in preventing neuron damage due to uncontrolled schizophrenia. Estrogen also has promising results in the neuroprotection of schizophrenia. It acts on all three main neurotransmitters that are involved in the pathology of schizophrenia. Since most of the data are derived from animal studies, more studies in humans are necessary in order to determine the safety and efficacy. This promising novel class of medication will hopefully help schizophrenic patients from deteriorating neurologically as well as symptomatically.

### 5.3 Newer Hypotheses and Future Drugs

As discussed before, hyperdopaminergic brain function is currently the most prominent hypothesis used in explaining the etiology of schizophrenia. Dopamine hypothesis has several lacunae and fails to explain the lack of effect of dopamine antagonists in treating the negative symptoms of schizophrenia. Current research

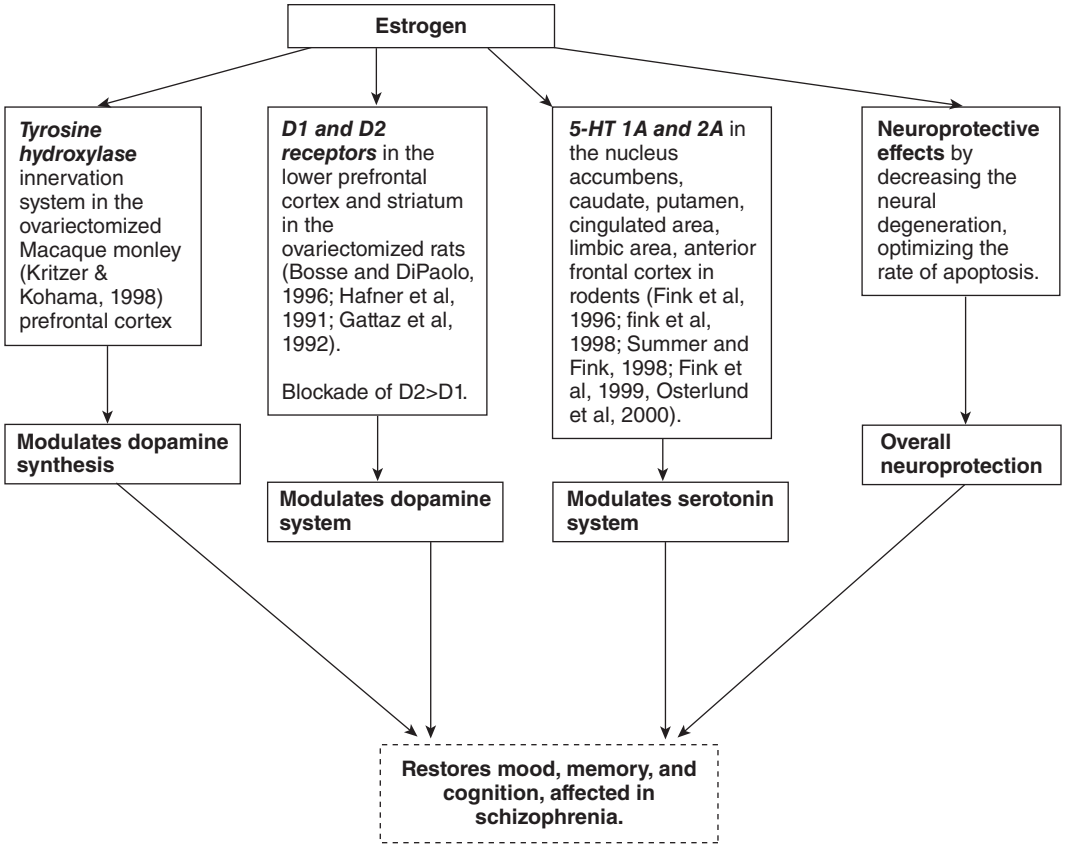


**Figure 6** Neuroprotection by estrogen. (Reproduced with permission from [81].)

has several explanations for the lack of effect of dopamine antagonists and has put forward newer hypotheses to better explain the diseases and the effects of drugs. Although some of these have been referred to in previous sections, to put in perspective, we are summarizing them below.

***Glutamate Hypothesis*** One of the most notable alternative hypotheses is the glutamate hypothesis, which states that schizophrenia is produced by the hypogultaminergic activity in the brain. Glutamate is an excitatory neurotransmitter in the CNS. The glutamate hypothesis took roots from the observation that the glutamate antagonist PCP produces symptoms, especially hallucinations, of schizophrenia. One interesting thing about glutamate antagonism is that it produces both positive and negative symptoms as opposed to just positive symptoms produced by dopamine overactivity. PCP acts somewhat selectively on a glutamate receptor subtype known as NMDA (*N*-methyl-D-aspartate) receptors. These receptors are excitatory in function and PCP probably inhibits the release of neurotransmitters mediated by NMDA receptors. Other antagonists at NMDA receptors such as ketamine, dexoxadrol, and MK-801 have also been reported to produce psychosis-like symptoms [82–84]. Another supporting finding for glutamate theory is that glycine, an NMDA agonist, reverses PCP-induced symptoms in rodents and makes patients with schizophrenia better [85]. However, caution has to be exercised in interpreting these results since there are a few findings in the literature that contradict the glutamate hypothesis.

***Cholinergic Role in Schizophrenia*** It is known that many people with schizophrenia smoke heavily, indicating that nicotine might relieve some of the symptoms of this condition. This observation also implies the dysfunction of the cholinergic system in schizophrenia. One of the earliest publications relating cholinergic system and mania came out in 1972 [86]. Cholinergic muscarinic receptors were found to be mediating the extrapyramidal effects of several first-generation antipsychotics [87]. In a postmortem study in humans, a higher cholinergic activity, indicated by choineacetyltransferase (ChAT) enzyme, was found in the hippocampus, caudate, putamen, and nucleus accumbens of schizophrenia patients compared to controls [88]. In another clinical study, therapeutic antagonism was shown between anticholinergic agents and neuroleptics [89]. The negative symptoms of schizophrenia have been attributed to the overactivity of the cholinergic system [90]. The same group showed a differential effect of anticholinergic drugs on schizophrenia, where these drugs increased the positive symptoms significantly, while reducing negative symptoms [91]. Another interesting observation related to cholinergic involvement is that chronic treatment in rats with haloperidol resulted in reduction of ChAT activity and a visual decrease in the immunoreactivity in neurons as well as in their processes in the striatum and the hippocampus [92]. This might also indicate the cholinergic mediation in the therapeutic effect of haloperidol. Hippocampal nicotinic receptors have been shown to be decreased in schizophrenia in a postmortem study [93]. Several recent studies involving humans and animals have indicated that altered expression and function of the $\alpha_7$-nicotinic cholinergic receptor may be responsible for the auditory sensory gating deficit characterized in schizophrenia patients and their relatives as diminished suppression of an auditory-evoked response to repeated stimuli [94]. Atypical, but not typical, antipsychotic drugs have been shown to increase cortical acetylcholine release without an effect in the nucleus accumbens or striatum [95]. An agonist at the $\alpha_7$-nicotinic receptor, 3-2,4 dimethoxybenzylidene anabaseine (DMXBA), has been shown to enhance auditory sensory gating in animal models [96]. Therapeutic doses of clozapine, olanzapine, and, to a lesser extent, quetiapine are associated with clinically relevant anticholinergic activity indicating a role in their therapeutic efficacy. All these findings point to a definite involvement of cholinergic system in the pathophysiology of schizophrenia. However, there are no drugs available to treat schizophrenia that act by modulating cholinergic system.

***New Drugs in the Pipeline*** Although we have come a long way from chlorpromazine type of drugs, there still need to be several refinements to be done to develop an ideal antipsychotic drug without/minimal significant side effects (which may never be possible practically). There were recent efforts to modify the drug delivery of existing drugs to improve the therapeutic index. Risperidone's long-acting formulation was shown to be well tolerated [97] and, of course, improved therapeutic levels in schizophrenic patients who are generally poorly compliant. Sustained release quetiapine from AstraZeneca (Seroquel) is undergoing clinical studies currently and preliminary data indicates significant improvement in improved symptoms associated with schizophrenia (measured by positive and negative syndrome scale [PANSS]) and an increase in time to psychiatric relapse, when administered once a day (http://www.astrazeneca.com/pressrelease/5310.aspx).

Apart from delivery modifications, there are a few targets that have shown promise in preclinical studies and concept-of-proof studies. Prominent among them is the $\alpha_7$ neuronal nicotinic receptor. Data from several studies suggest a link between the $\alpha_7$ neuronal nicotinic acetylcholine receptor (nAChR) and brain disorders,

including schizophrenia. A novel molecule (PNU-120596), developed by Pfizer, works as a powerful positive allosteric modulator of $\alpha_7$-nicotinic receptors. Systemic administration of PNU-120596 in animals improved the auditory gating deficit caused by amphetamines, a model that mimics circuit level disturbance associated with schizophrenia [98]. Another selective partial $\alpha_7$-nicotinic agonist, DMXBA, has been shown to improve these learning and memory deficits and deficits in sensory inhibition in rodent models. Clinical trials with DMXBA also demonstrated improved sensory inhibition in 12 schizophrenia patients and showed improvement in several subtests of the RBANS learning and memory assessment instrument. Thus these data strongly suggest that $\alpha_7$-nicotinic activators may prove useful in treating some of the symptoms of schizophrenia [99]. Glutamate receptor (NMDA) modulators also have shown some promise in preclinical trials, but there is no clinical data yet to prove their usefulness in humans.

## 6 CONCLUSION

The pharmacotherapy of schizophrenia has evolved significantly from chlorpromazine to second-generation antipsychotics. Although scientific research could get rid of the extrapyramidal side effects of first-generation antipsychotics, there are newer side effects that came with the atypical antipsychotics. With the advancement of molecular biology and biomedical sciences, we could hope that more molecular targets for treating schizophrenia with fewer side effects will be available in the near future.

Abbreviations

| | |
|---|---|
| AA | Atypical antipsychotic |
| ACh | Acetylcholine |
| BMI | Body mass index |
| CHD | Coronary heart disease |
| CNS | Central nervous system |
| CPZ | Chlorpromazine |
| CV | Cardiovascular |
| DM | Diabetes mellitus |
| DA | Dopamine |
| DKA | Diabetic ketoacidosis |
| $D_1$ | Dopamine receptor type 1 |
| $D_2$ | Dopamine receptor type 2 |
| FDA | Food and Drug Administration |
| FGA | First-generation antipsychotic |
| GI | Gastrointestinal |
| HDL | High-density lipoprotein |
| $H_1$ | Histamine receptor type 1 |
| $M_1$ | Muscarinic receptor type 1 |
| SGA | Second-generation antipsychotic |
| TA | Typical antipsychotic |
| TG | Triglyceride |
| 5-HT | Serotonin |
| $5\text{-HT}_{1A}$ | Serotonin receptor type 1A |
| $5\text{-HT}_{2A}$ | Serotonin receptor type 2A |
| $5\text{-HT}_{2C}$ | Serotonin receptor type 2C |

## REFERENCES


1. Turner TI (2007). Chlorpromazine: unlocking psychosis. *BMJ 334* (Suppl 1), s7.
2. Caroff SN, Mann SC, Campbell EC, Sullivan KA (2002). Movement disorders associated with atypical antipsychotic drugs. *J Clin Psychiatry 63*(4), 12–19.
3. Kane J, et al. (1988). Clozapine for the treatment-resistant schizophrenic: A double-blind comparison with chlorpromazine. *Arch Gen Psychiatry 45*, 789–796.
4. Kapur S (1998). A new framework for investigating antipsychotic action in humans: Lessons from PET imaging. *Mol Psychiatry 3*, 101–102.
5. Wieck A, Haddad PM (2003). Antipsychotic-induced hyperprolactinaemia in women: Pathophysiology, severity and consequences. Selective literature review. *Br J Psychiatry 182*, 199–204.
6. Wander AG (1989). The history of clozapine. *Psychopharmacology 99*(1), S3–S5.
7. Devinsky O, Pacia SV (1994). Seizures during clozapine therapy. *J Clin Psychiatry 55*(B), S153–S156.
8. U.S. Food and Drug Administration (2007). *Electronic Orange Book*. Accessed October 25, 2007. *http://www.accessdata.fda.gov/scripts/cder/ob/default.cfm*.
9. Kapur S, Remington G (2001). Atypical antipsychotics: New directions and new challenges in the treatment of schizophrenia. *Ann Rev Med 52*, 503–517.
10. Stahl SM (2001). "Hit-and-run" actions at dopamine receptors, part 1: Mechanism of action of atypical antipsychotics. *J Clin Psychiatry 62*, 670–671.
11. Geddes J, Freemantle N, Harrison P, Bebbington P (2000). Atypical antipsychotics in the treatment of schizophrenia: Systematic overview and meta-regression analysis. *Br Med J 321*, 1371–1376.
12. Melkersson KI, Hulting AL (2001). Insulin and leptin levels in patients with schizophrenia or related psychoses—a comparison between different antipsychotic agents. *Psychopharmacology 154*, 205–212.
13. Lteif AA, Mather KJ, Clark CM (2003). Diabetes and heart disease an evidence-driven guide to risk factors management in diabetes. *Cardiol Rev 11*, 262–274.
14. Allison DB, et al. (1999). Antipsychotic-induced weight gain: A comprehensive research synthesis. *Am J Psychiatry 156*(11), 1686–1696.

15. Schoemaker H, et al. (1997). Neurochemical characteristics of amisulpride, an atypical dopamine D2/D3 receptor antagonist with both presynaptic and limbic selectivity. *J Pharmacol Exp Ther 280*, 83–97.
16. Meltzer HY, Matsubara S, Lee JC (1989). The ratios of serotonin2 and dopamine2 affinities differentiate atypical and typical antipsychotic drugs. *Psychopharmacol Bull 25*, 390–392.
17. Lieberman JA, Golden R, Stroup S, McEvoy J (2000). Drugs of the psychopharmacological revolution in clinical psychiatry. *Psychiatr Serv 51*, 1254–1258. Available: www.psychservices.psychiatryonline.org/cgi/content/full/51/10/1254. Accessed October 9, 2000.
18. Rupinder Mann. The role of dopamine receptors in schizophrenia. Available: http://wwwchem.csustan.edu/chem44x0/SJBR/Mann.htm. Accessed August 2011.
19. Meltzer HY (2007). Mechanism of action of atypical antipsychotic drugs chapter 58. Vandebilt university school of medicine, Nashville, Tennessee. Available: www.scholar.lib.vt.edu/theses/available/etd-02052001-120403/unrestricted/finalthesis.pd; accessed on Oct 7.
20. Seeman, P (2002). Atypical antipsychotics: Mechanism of action. *Can J Psychiatry 47*(1), 27–38.
21. Kessler RM (2007). Aripiprazole: What is the role of dopamine D(2) receptor partial agonism? *Am J Psychiatry 164*(9), 1310–1312.
22. Hashimoto K. (2010). Glycine transport inhibitors for the treatment of schizophrenia. *Open Med Chem J 4*, 10–19.
23. Kristiansen LV, Huerta I, Beneyto M, Meador-Woodruff JH (2007). NMDA receptors and schizophrenia. *Curr Opin Pharma*col *7*(1), 48–55.
24. Sodhi M, Wood KH, Meador-Woodruff J (2008). Role of glutamate in schizophrenia: integrating excitatory avenues of research. *Expert Rev Neurother 8*(9), 1389–1406. Review.
25. Reynolds GP, Zhang ZJ, Zhang XB (2002). Association of antipsychotic drug-induced weight gain with a 5-HT2C receptor gene polymorphism. *Lancet 359*(9323), 2086–2087.
26. Hertzberg MA, et al. (1999). A preliminary study of lamotrigine for the treatment of posttraumatic stress disorder. *Biol Psychiatry 45*(9), 1226–1229.
27. Allison DB, et al. (1999). Antipsychotic-induced weight gain: A comprehensive research synthesis. *Am J Psychiatry 156*(11), 1686–1696.
28. Osser DN, Najarian DM, Dufresne RL (1999). Olanzapine increases weight and serum triglyceride levels. *J Clin Psychiatry 60*(11), 767–770.
29. Henderson DC, et al. (2000). Clozapine, diabetes mellitus, weight gain, and lipid abnormalities: A five-year naturalistic study. *Am J Psychiatry 157*(6), 975–981.
30. Ghaeli P, Dufresne RL (1999). Elevated serum triglycerides with clozapine resolved with risperidone in four patients. *Pharmacotherapy 19*(9), 1099–1101.
31. Tran PV, et al. (1997). Double-blind comparison of olanzapine versus risperidone in the treatment of schizophrenia and other psychotic disorders. *J Clin Psychopharmacol 17*(5), 407–418.
32. Dixon L, et al. (2000). Prevalence and correlates of diabetes in national schizophrenia samples. *Schizophr Bull 26*(4), 903–912.
33. Sernyak MJ, Leslie DL, Alarcon RD, Losonczy MF, Rosenheck R (2002). Association of diabetes mellitus with use of atypical neuroleptics in the treatment of schizophrenia. *Am J Psychiatry 159*(4), 561–566.
34. Harris MI, et al. (1998). Prevalence of diabetes, impaired fasting glucose, and impaired glucose tolerance in U.S. adults. The Third National Health and Nutrition Examination Survey, 1988–1994. *Diabetes Care 21*(4), 518–524.
35. Ryan MC, Collins P, Thakore JH (2003). Impaired fasting glucose tolerance in first-episode, drug-naive patients with schizophrenia. *Am J Psychiatry 160*(2), 284–289.
36. Thakore JH, Mann JN, Vlahos I, Martin A, Reznek R (2002). Increased visceral fat distribution in drug-naive and drug-free patients with schizophrenia. *Int J Obes Relat Metab Disord 26*(1), 137–141.
37. Thonnard-Neumann E (1968). Phenothiazines and diabetes in hospitalized women. *Am J Psychiatry 124*(7), 978–982.
38. Lindenmayer JP, Czobor P, Volavka J, Citrome L, Sheitman B, McEvoy JP . . . Lieberman JA (2003). Changes in glucose and cholesterol levels in patients with schizophrenia treated with typical or atypical antipsychotics. *Am J Psychiatry 160*(2), 290–296.
39. Koro CE, et al. (2002). Assessment of independent effect of olanzapine and risperidone on risk of diabetes among patients with schizophrenia: Population based nested case-control study. *BMJ 325*(7358), 243.
40. Wilson DR, D'Souza L, Sarkar N, Newton M, Hammond C (2003). New-onset diabetes and ketoacidosis with atypical antipsychotics. *Schizophr Res 59*(1), 1–6.
41. Kuruvilla A, Srikrishna G, Peedicayil J, Kuruvilla K, Kanagasabapathy AS (1993). A study on serum prolactin levels in schizophrenia: correlation with positive and negative symptoms. *Int Clin Psychopharmacol 8*(3), 177–179.
42. Atmaca M, Kuloglu M, Tezcan E, Canatan H, Gecici O. (2002). Quetiapine is not associated with increase in prolactin secretion in contrast to haloperidol. *Arch Med Res 33*(6), 562–565.
43. Markianos M, Hatzimanolis J, Lykouras L (1999). Switch from neuroleptics to clozapine does not influence pituitary-gonadal axis hormone levels in male schizophrenic patients. *Eur Neuropsychopharmacol 9*(6), 533–536.
44. Kim KS, et al. (2002). Effects of olanzapine on prolactin levels of female patients with schizophrenia treated with risperidone. *J Clin Psychiatry 63*(5), 408–413.
45. Pfizer (2000). *FDA briefing document for Zeldox capsules (Ziprasidone)*. New York: Pfizer Inc., July 18.
46. Janssen-Ortho (2002). *Risperidal (risperidone) and cerebrovascular adverse events in placebo-controlled dementia trials*. Toronto: Janssen-Ortho, Inc., October 11.

47. Lee PE, Gill SS, Freedman M, Bronskill SE, Hillmer MP, Rochon PA (2004). Atypical antipsychotic drugs in the treatment of behavioural and psychological symptoms of dementia: Systematic review. *BMJ 329*(7457), 75.
48. Daniel WR (1996). On the plausibility of "The neurodevelopmental hypothesis" of schizophrenia. *Neuropsychopharmacology 14*, 1S–11S.
49. Charles DE (2006). Antipsychotic-associated neuronal changes in the brain: Toxic, therapeutic, or irrelevant to the long-term outcome of schizophrenia? *Prog Neuro-Psychopharmacol Biol Psychiatry 30*, 174–189.
50. Michael EF, Weinberger R, Daniel, Lu Bai (2003). Brain-derived neurotropic factor and genetic risk. *Am J Psychiatry 160*, 1242.
51. John CG (2007). Neurodegeneration in schizophrenia: Evidence from in vivo neuroimaging studies. *Sci World J 7*, 135–143.
52. Ivan P, et al. (2006). Possible mechanisms of neurodegeneration in schizophrenia. *Neurochem Res 31*, 1279–1294.
53. Lynn SD, et al. (2002). Smaller frontal gray matter volume in postmortem schizophrenic brains. *Am J Psychiatry 159*, 1983–1991.
54. Vita A, de Peri L (2007). Hippocampal and amygdala volume reductions in first-episode schizophrenia. *Br J Psychiatr 190*, 271.
55. Dennis V, et al. (2006). Hippocampal and amygdala volumes according to psychosis stage and diagnosis. *Arch Gen Psychiatry 63*, 139–149.
56. Dennis V, et al. (1999). Hippocampal volume in first-episode psychoses and chronic schizophrenia. *Arch Gen Psychiatry 56*, 133–141.
57. Jeffrey LA (1999). Is schizophrenia a neurodegenerative disorder? A clinical and neurobiological perspective. *Biol Psychiatry 46*, 729–739.
58. Tan LY, et al. (2005). Decreased BDNF in serum of patients with chronic schizophrenia on long-term treatment with antipsychotics. *Neurosci Lett 382*, 27–32.
59. Aaron BT, Neil RA (2005). Cognitive approaches to schizophrenia: Theory and therapy. *Annu Rev Clin Psycho 1*, 577–606.
60. Michael GF (2007). Stimulating the development of drug treatments to improve cognition in schizophrenia. *Annu Rev Clin Psychol 3*, 159–180.
61. Jonathan K, Cohen H, Cardon M, Ziv Y, Schwartz M (2004). T cell deficiency leads to cognitive dysfunction: Implications for therapeutic vaccination for schizophrenia and other psychiatric conditions. *PNAS 101*, 8180–8185.
62. Jann MW (2004). HYPERLINK "http://www.ncbi.nlm.nih.gov/pubmed/15585443" Implications for atypical antipsychotics in the treatment of schizophrenia: Neurocognition effects and a neuroprotective hypothesis. *Pharmacotherapy 24*(12), 1759–1783. Review.
63. Keefe SER, et al. (2004). Comparative effect of atypical and conventional antipsychotic drugs on neurocognition in first-episode psychosis: A randomized, double-blind trial of olanzapine versus low doses of haloperidol. *Am J Psychiatry 161*, 985–995.
64. Jarksog LF, Lieberman JA (2006). Neuroprotection in schizophrenia. *J Clin Psychiatry 67*(9), e09.
65. Ehrenreich H, Siren AL (2001). Neuroprotection—What does it mean? What means do we have? *Eur Arch Psychiatry Clin Neurosci 251*, 149–151.
66. Lieberman JA, et al. (2005). Antipsychotic drug effects on brain morphology in first-episode psychosis. *Arch Gen Psychiatry 62*, 361–370.
67. Csernansky JG, Martin MV, Czeisler B, Meltzer MA, Ali Z, Dong H (2008). Neuroprotective effects of olanzapine in a rate model of neurodevelopmental injury. *Pharmacol Biochem Behav 83*, 208–213.
67a. Bian Q, et al. (2008). The effect of atypical antipsychotics, perospirone, ziprasidone and quetiapine on microglial activation induced by interferon-γ. *Prog Neuro-Psychopharmacol Biol Psychiatry 32*(1), 42–48.
68. Dazzan P, et al. (2005). Different effects of typical and atypical antipsychotics on grey matter in first episode psychosis: The AESOP study. *Neuropsychopharmacology 30*, 765–774.
69. Goff DC, Coyle JT (2001). The emerging role of glutamate in the pathophysiology and treatment of schizophrenia. *Am J Psychiatry 158*, 1367–1377.
70. Miyamoto S, Duncan GE, Goff DC, Lieberman JA. Therapeutics of schizophrenia. In: Davis KL, Charney D, Coyle JT, Nemeroff C, eds. (2002). *Neuropsychopharmacology: The fifth generation of progress*. Philadelphia: Lippincott Williams & Wilkins.
71. Lieberman JA, Jarskog LF, Malaspina D (2006). Preventing clinical deterioration in the course of schizophrenia: The potential for neuroprotection. *J Clin Psychiatry 67*(6), 983–990.
72. Burton S (2006). Symptom domains of schizophrenia: The role of atypical antipsychotic agents. *J Psychopharmacol 20*(6), 6–19.
73. Chavez-Noriega LA, Marino MJ, Schaffhauser H, Campbell UC, Conn PJ (2005). Novel potential therapeutics for schizophrenia: Focus on the modulation of metabotropic glutamate receptor function. *Curr Neuropharmacol 3*, 9–34.
74. Corti C, et al. (2007). The use of knock-out mice unravels distinct roles for mGlu2 and mGlu3 metabotropic glutamate receptors in mechanisms of neurodegeneration/neuroprotection. *J Neurosci 27*(31), 8297–8308.
75. Zolan Z. Treatment of refractory schizophrenia with Riluzole. Clinical trial.gov. Available: http://clinicaltrials.gov/ct/show/NCT00375050. Accessed October 24 2007.
76. Ehrenreich H, et al. (2004). Erythropoietin: A candidate compound for neuroprotection in schizophrenia. *Mol Psychiatry 9*, 42–54.
77. Anonymous. Neuroprotection of schizophrenia. Where we stand and where we are going. Available: http://www.

neuroprotection-schizophrenia.de/. Accessed October 20, 2007.

78. Ehrenreich H, Timner W, Sirén AL (2004). A novel role for an established player: Anemia drug erythropoietin for the treatment of cerebral hypoxia/ischemia. *Transfus Apher Sci 31*, 39–44.
79. Ehrenreich H, Aust C, Krampe H, Jahn H, Jacob S, Herrmann M, Siren AL (2004). Erythropoietin: Novel approaches to neuroprotection in human brain disease. *Metabol Brain Dis 19*(3/4), 195–206.
80. Syr M, Calon F, Morissette M, Paolo TD (2002). Estrogenic modulation of brain activity: Implications for schizophrenia and Parkinson's disease. *J Psychiatry Neurosci 27*(1), 12–27.
81. Chattopadhyay S (2004). Estrogen and schizophrenia: Any link? *Internet J Mental Health 2*(1).
82. Gouzoulis-Mayfrank E, Heekeren K, Neukirch A, Stoll M, Stock C, Obradovic M, Kovar KA (2005). Psychological effects of (*S*)-ketamine and *N*,*N*-dimethyltryptamine (DMT): A double-blind, cross-over study in healthy volunteers. *Pharmacopsychiatry 38*(6), 301–311.
83. Browne RG, Welch WM (1982). Stereoselective antagonism of phencyclidine's discriminative properties by adenosine receptor agonists. *Science 217*(4565), 1157–1159.
84. Hitri A, O'Connor DA, Cohen JM, Keuler DJ, Deutsch SI (1993). Differentiation between MK-801- and apomorphine-induced stereotyped behaviors in mice. *Clin Neuropharmacol 16*(3), 220–236.
85. Javitt DC, Sershen H, Hashim A, Lajtha A (2000). Inhibition of striatal dopamine release by glycine and glycyldodecylamide. *Brain Res Bull 52*(3), 213–216.
86. Janowsky DS, el-Yousef MK, Davis JM, Sekerke HJ (1972). A cholinergic-adrenergic hypothesis of mania and depression. *Lancet 2*(7778), 632–635.
87. Snyder SH, Greenberg D, Yamumura HI (1974). Antischizophrenic drugs: Affinity for muscarinic cholinergic receptor sites in the brain predicts extrapyramidal effects. *J Psychiatr Res 11*, 91–95.
88. McGeer PL, McGeer EG (1977). Possible changes in striatal and limbic cholinergic systems in schizophrenia. *Arch Gen Psychiatry 34*(11), 1319–1323.
89. Mohan M, Kay SR (1978). Therapeutic antagonism between anticholinergics and neuroleptics: possible involvement of cholinergic mechanisms in schizophrenia. *Schizophr Bull 4*(1), 3–6.
90. Tandon R, Greden JF (1989). Cholinergic hyperactivity and negative schizophrenic symptoms. A model of cholinergic/dopaminergic interactions in schizophrenia. *Arch Gen Psychiatry 46*(8), 745–753.
91. Tandon R, Mann NA, Eisner WH, Coppard N (1990). Effect of anticholinergic medication on positive and negative symptoms in medication-free schizophrenic patients. *Psychiatry Res 31*(3), 235–241.
92. Mahadik SP, Laev H, Korenovsky A, Karpiak S (1988). Haloperidol alters rat CNS cholinergic system: Enzymatic and morphological analyses. *Biol Psychiatry 24*(2), 199–217.
93. Freedman R, Hall M, Adler LE, Leonard S (1995). Evidence in postmortem brain tissue for decreased numbers of hippocampal nicotinic receptors in schizophrenia. *Biol Psychiatry 38*(1), 22–33.
94. Adler LE, Olincy A, Waldo M, Harris JG, Griffith J, Stevens K, Flach K, Nagamoto H, Bickford P, Leonard S, Freedman R (1998). Schizophrenia, sensory gating, and nicotinic receptors. *Schizophr Bull 24*(2), 189–202.
95. Ichikawa J, Dai J, O'Laughlin IA, Fowler WL, Meltzer HY (2002). Atypical, but not typical, antipsychotic drugs increase cortical acetylcholine release without an effect in the nucleus accumbens or striatum. *Neuropsychopharmacology 26*(3), 325–339.
96. Martin LF, Kem WR, Freedman R (2004). Alpha-7 nicotinic receptor agonists: Potential new candidates for the treatment of schizophrenia. *Psychopharmacology (Berl) 174*(1), 54–64.
97. Harrison TS, Goa KL (2004). Long-acting risperidone: A review of its use in schizophrenia. *CNS Drugs 18*(2), 113–132.
98. Hurst RS, Hajós M, Raggenbass M, Wall TM, Higdon NR, Lawson JA, Rutherford-Root KL, Berkenpas MB, Hoffmann WE, Piotrowski DW, Groppi VE, Allaman G, Ogier R, Bertrand S, Bertrand D, Arneric SP (2005). A novel positive allosteric modulator of the alpha7 neuronal nicotinic acetylcholine receptor: in vitro and in vivo characterization. *J Neurosci 25*(17), 4396–4405.
99. Olincy A, Stevens KE (2007). Treating schizophrenia symptoms with an alpha7 nicotinic agonist, from mice to men. *Biochem Pharmacol 74*(8), 1192–1201.

# RADIOPHARMACEUTICALS AND MEDICAL IMAGING: DEVELOPMENT OF RADIOIMMUMOCONJUGATES FOR RADIOIMMUNOIMAGING AND RADIOIMMUNOTHERAPY OF MALIGNANT DISEASES

Gang Niu and Xiaoyuan Chen

*Molecular Imaging Program at Stanford (MIPS), Department of Radiology and Bio-X Program, Stanford University School of Medicine, Stanford, California*

## 1 INTRODUCTION

Radioisotopes have been applied to detect and treat malignant diseases for a long time. Radiopharmaceuticals are generally classified as either being diagnostic or therapeutic according to their application. As shown in Figure 1, a radiopharmaceutical can be considered to have three components: (i) the targeting carrier, (ii) the radionuclide, and (iii) a linker moiety [1]. A perfect radiopharmaceutical would, after intravenous administration, travel via the bloodstream to the target cells, where it would efficiently bind only with cell surface or intracellular counterparts or interact solely with the desired molecular pathway. Any radioligand that did not reach the desired target would be rapidly excreted from the body so that the only radioactivity remaining in the body would be that localized to the target site. The development of radiopharmaceuticals has followed closely behind the advances made by the relevant disciplines such as particle physics, chemistry, radiochemistry, radiopharmacy, and biology. Radioimmunopharmaceuticals or radioimmunoconjugates (RICs) are using antibody (Ab) or antibody derivatives as targeting carriers, taking advantage of the specific binding between antibody and antigen. In this chapter, we will discuss several aspects of RIC development and their application in radioimmunoimaging (RII) and radioimmunotherapy (RIT) for malignant diseases, including lymphoma, prostate cancer, colon cancer, and breast and ovarian cancers.

## 2 RADIONUCLIDES

Radionuclides function as beacons in RII and killing agents in RIT. For imaging purpose, γ-ray and positron emitters will be used, while β-particle emitters, α-particle emitters, and Auger and Coster-Kronig electron emitters will be considered for RIT. A number of factors need to be considered when selecting appropriate radionuclides for RII and RIT, including emission characteristics, physical half-life, cost, availability, and linkage chemistry. The radionuclides evaluated in preclinical and clinical RII and RIT studies have been summarized in several reviews [1, 2].

As one of the predominant factors for determining a radionuclide's suitability for RII and RIT, the physical half-life should be comparable to the time needed for

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

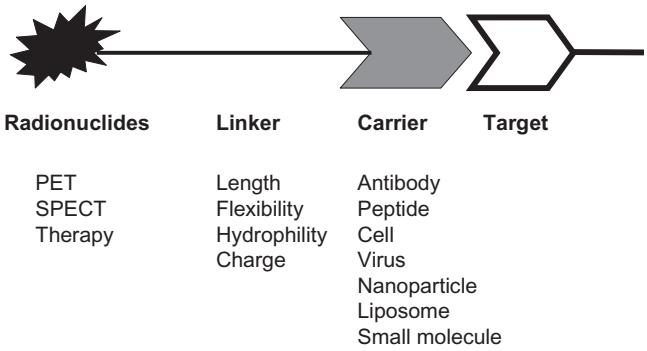


**Figure 1** Radiopharmaceutical is typically composed of a radionuclide, a carrier that is/contains a targeting ligand, and a linker between the carrier and the radiolabel.

the radiolabeled antibody or fragments targeting the tumor (typically 2–4 days for intact Abs and 2–6 h for Ab fragments). In general, radiolabeled intact Abs have been used in tumor detection and treatment planning, as well as in RIT, whereas the application of Ab fragments has been restricted to tumor detection only. Although the use of shorter-lived radionuclides (with half-lives of hours) for RII with Ab fragments is an option, the kinetics of intact Abs demand the use of long-lived radionuclides (half-lives of days) to allow imaging at later time points for obtaining maximum information. Based on their half-lives, radionuclides such as $^{94m}$Tc ($t_{1/2} = 0.87$ h), $^{68}$Ga ($t_{1/2} = 1.13$ h), $^{18}$F ($t_{1/2} = 1.83$ h), and $^{99m}$Tc ($t_{1/2} = 6.02$ h) are better suited for use in combination with Ab fragments. Radionuclides with longer half-lives including $^{64}$Cu ($t_{1/2} = 12.7$ h), $^{86}$Y ($t_{1/2} = 14.7$ h), $^{76}$Br($t_{1/2} = 16.2$ h), $^{124}$I ($t_{1/2} = 78.4$ h), and $^{89}$Zr ($t_{1/2} = 100.3$ h) are suitable for imaging of Ab fragments as well as intact Abs. The half-lives of these radionuclides offer an advantage with respect to the logistics of radiolabeling and transportation but a potential disadvantage with respect to the radiation burden to the patient, especially when coupled to compounds with long biologic half-lives, such as intact Abs. For RIT purposes, Abs labeled with shorter-lived radionuclides will have suffered significant decay of activity before maximum uptake in the tumor.

The applicability of a radionuclide is also dependent on the availability of methods to obtain it in pure and sufficient amounts, the time period available for transportation and labeling, the availability of procedures for its stable coupling to an Ab with maintenance of the Abs in vivo biodistribution characteristics, and identification of undesirable components in its decay scheme that could compromise image quality and quantification accuracy. For example, the presence of prompt single γ-photons will cause spurious true coincidences during positron emission tomography (PET) study, such as $^{86}$Y, $^{76}$Br, and $^{124}$I. High positron energy will result in an intrinsic resolution loss ($^{94m}$Tc, $^{68}$Ga, $^{76}$Br, and $^{124}$I), and the presence of radionuclidic impurities or radioisotopes of the same element generated during cyclotron production such as $^{94m}$Tc will disturb the labeling process and reduce labeling yield [2].

Radionuclides usually used for therapy produce intermediate and long-range β emissions, which have a traveling range of several millimeters to a centimeter in tissue. The range in tissue of these β emissions allows for a potential "cross-fire" effect, where despite heterogeneous distribution of the radiolabeled antibody, radiation can reach not only the antibody-targeted cell but also neighboring nontargeted cells. Some therapeutic radionuclides such as $^{131}$I also have γ-ray emission, which may be useful for scintigraphic imaging and for determinations of in vivo localization of the radiopharmaceutical and evaluation of dosimetry. Besides $^{131}$I ($t_{1/2} = 8.4$ days), $^{186}$Re ($t_{1/2} = 3.8$ days), and $^{90}$Y ($t_{1/2} = 2.67$ days) are two commonly used therapeutic radionuclides. $^{90}$Y is a useful radionuclide for RIT due to its energetic β emissions ($E_{max} = 2.3$ MeV), ready availability at moderate cost, and absence of γ emissions, allowing outpatient RIT. Its half-life of 64 h is comparable to the uptake and residence time of many antibodies on the tumor [3]. Other radiometals, such as $^{67}$Cu, $^{188}$Re, and $^{177}$Lu also have suitable β emissions for therapy, but availability may limit their widespread use in clinical trials.

Compared with β particles, the high linear energy transfer (LET) characteristics of α particles make it possible to perform much localized irradiation while preserving the surrounding tissues. Moreover, cell toxicity can be obtained with only a few disintegrations at the cell surface. On the other hand, the cross-fire effect in α-RIT is limited, owing to the short path length of these particles. These physical characteristics appear to be very suitable for targeting isolated cells of hematological malignancies, such as multiple myeloma and leukemia. For instance, the targeting of interleukin-2 (IL-2) receptors using specific CD25 antibodies coupled with $^{213}$Bi ($t_{1/2} = 45.6$ min) have proved to be more efficient than the use of $^{90}$Y in the treatment of T-cell leukemia [4]. Multiple myeloma can be efficiently targeted by α-RIT using an anti-CD138 antibody (BB4 MAb) [5, 6]. A synergistic effect has been documented between α-RIT and chemotherapy using paclitaxel in multiple myeloma cell lines. Owing to the short half-life of $^{213}$Bi, $^{212}$Bi ($t_{1/2} = 60.6$ min), or $^{211}$At ($t_{1/2} = 7.0$ h), the number of injected radioactive atoms is extremely limited [7]. Such a limit imposed by short half-life of radionuclides has prompted using antibody fragments such as $F(ab')_2$ or diabodies to improve the biodistribution [8].

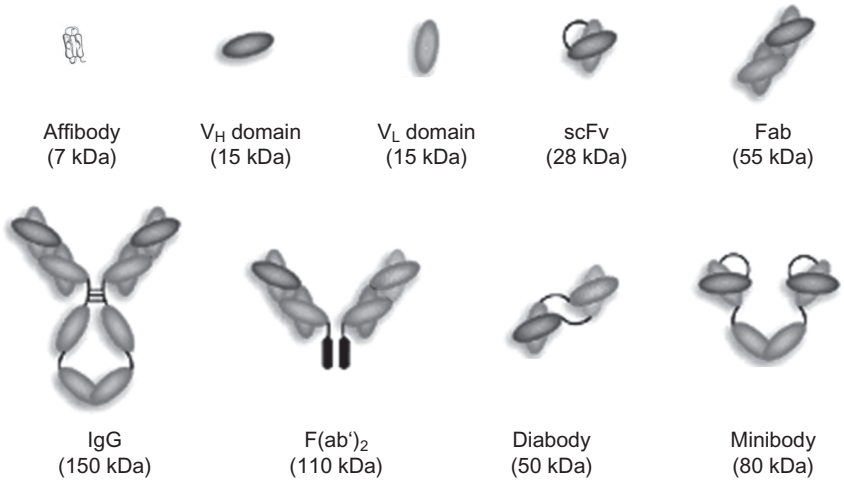


**Figure 2** Engineered antibody fragments and Affibody used for imaging HER-2 expression. Reproduced with permission from ref. 29 [9].

## 3 ANTIBODY AND ANTIBODY FRAGMENTS

### 3.1 History of Antibody Development

The majority of Abs produced in mammals has a molecular weight of approximately 150 kDa. The Ab molecule is composed of pairs of polypeptide chains, light and heavy, that are linked by disulfide bridges. Specifically, the $IgG_1$ molecule consists of two variable light chains ($V_L$) and two variable heavy chains ($V_H$), two constant light chains ($C_L$), and six constant heavy chains (Fig. 2). The antigen-binding region of the Ab consists of the $V_L$, $V_H$, $C_L$, and $C_{H1}$ domains, also known as the Fab portion of the antibody. The Fc portion consists of the $C_{H2}$ and the $C_{H3}$ domains and is involved in a number of immune response effector functions and involved in antibody clearance from circulation. Effector functions that have been associated with the Fc portion include complement-mediated cytotoxicity (CMC) and antibody-dependent cellular cytotoxicity (ADCC). The whole immunoglobulin is bivalent because it has two antigen-binding sites, one on each "arm" of the Y-shaped molecule.

Over the last several decades, there have been continued refinements in the production and the design of antibodies used for RII and RIT. Initial efforts used polyclonal antibodies, which were produced by a collection of antiserum after immunization of a given species, such as rabbit, to a specific antigen. In 1975, Kohler and Milstein [10] developed the hybridoma technique, which allowed for the production of monoclonal antibodies (mAbs). The selected hybridoma would produce large quantities of a specific mAb with single specificity allowing for the rapid expansion of clinical trials evaluating radiolabeled mAbs to a variety of tumor-associated antigens. These antibodies provided new approaches for the RII and RIT of cancers and metastases [11].

However, the clinical utility of rodent mAbs is limited by their immunogenicity and often inefficient secondary immune functions. These issues also hinder the application of murine antibodies in repetitive imaging. Humanized antibodies are commonly created by transplanting the antigen-binding segments, known as complementarity determining regions (CDRs), from rodent antibodies into human antibodies [12]. The localization of antibody in tumors following passive transfer of antibodies is governed by a complex series of events involving extravasation of antibody and its diffusion into the tumor tissue. The deposition of an antibody is also dependent on the tumor mass and its location, the binding affinity to tumor-associated antigens, the degree of expression of the relevant antigen, and its stability on tumor cells [13]. The large molecular size of intact antibodies leads to prolonged serum half-life. Intact antibodies also take a long time for maximal dose deposition in the target tissue, due to the slow diffusion from the vasculature into the tumor [14]. Moreover, Ab intratumoral diffusion is limited by its size to a penetration rate of about 1 mm every 2 days, potentially resulting in heterogeneous deposition in the tumor [15]. Tailoring the antibody molecules by enzymatic digestion, antibody fragments $F(ab)'_2$ and Fab′ may allow more rapid elimination from the blood and normal tissues (except kidneys), due to the smaller size than intact antibodies [16]. However, these antibody fragments are tedious to prepare and not readily reproducible [17]. Antibody engineering has made it possible to tailor antigen-binding domains into a single polypeptide with

a much smaller size than intact immunoglobulin. A single-chain Fv (scFv) recombinant protein for a given monoclonal antibody can be prepared by connecting genes encoding for heavy-chain and light-chain variable regions at the deoxyribonucleic acid (DNA) level by an appropriate oligonucleotide linker. The resulting translation product forms a single polypeptide chain with a linker bridging the two variable domains [18]. Compared with intact parental antibody (150 kDa), scFv (28 kDa) for a given immunoglobulin exhibits faster clearance kinetics and deeper tumor penetration; however, the absolute dose deposition of scFv is much lower [19]. This is primarily due to its monovalent binding nature, which results in lower functional avidity. The degree and specificity of tumor localization of scFv can be improved by increasing the affinity for its corresponding antigen. Divalent and multivalent scFvs (diabodies or tribodies) can result either from the spontaneous dimerization or covalent association of two or more monovalent scFvs [20, 21]. With their bivalent or multivalent binding, they exhibit better tumor deposition than their scFv counterparts, yet clear faster from the circulation than intact antibodies because of their smaller molecular size (55 kDa for diabody and 80 kDa for tribody) [22]. These recombinant antibody fragments with di- or multivalent binding with intermediate molecular weight (60–120 kDa) are a compromise between slow-clearing high-localizing immunoglobulins and fast-clearing low-localizing monovalent scFvs [23]. The fusion proteins of scFv with the hinge region and CH3 domain of immunoglobulin molecule are called minibodies. The hinge region facilitates dimerization mediated by disulfide bond formation and imparts flexibility to the antigen-binding domain. With the molecular weight of 80 kDa, minibodies persist longer than diabodies in the serum, hence allowing higher dose deposition in the tumors [24]. The structure and approximate molecular weight of IgG antibody, engineered antibody fragments, and Affibody is shown in Figure 2.

### 3.2 Affibody and Small Antibody Mimics

Affibody molecules are small nonimmunoglobulin affinity ligands based on a 58-amino-acid Z-domain scaffold, derived from one of the IgG-binding domains of *staphylococcal* protein A [25]. Randomization of 13 amino acid positions in the binding surface of this domain scaffold has been used to construct combinatorial phagemid libraries, from which affibody molecules bound to desired target molecules can be selected by phage display [26, 27]. The affibody molecule $His_6$-$Z_{HER\text{-}2/neu:4}$ was demonstrated to bind with nanomolar affinity (approximately 50 n$M$) to the HER-2-ECD molecule at a different site than the humanized antibody trastuzumab. $^{125}$I-labeled $His_6$-$Z_{HER\text{-}2/neu:4}$ affibody made by indirect radioiodination using *N*-succinimidyl-*p*-(trimethylstannyl) benzoate (SPMB) showed specific binding to native HER-2 overexpressed on the SKBR-3 tumor cell line [28, 29]. The bivalent form of the affibody ligand, $(Z_{HER\text{-}2:4})_2$ has a molecular weight of 15.6 kDa and an apparent affinity ($K_D$) against HER-2 of 3 n$M$, which is approximately the same as trastuzumab. Due to its small size, $^{125}$I-$(Z_{HER\text{-}2:4})_2$ affibody was primarily excreted through the kidneys, which caused extremely high kidney accumulation. The second-generation HER-2-specific affibody molecule ($His_6$-$Z_{HER\text{-}2:342}$), obtained by a single-library affinity maturation, bound HER-2 with a $K_D$ of 22 p$M$. Gamma camera imaging with $^{125}$I-$His_6$-$Z_{HER\text{-}2:342}$ showed clear, high-contrast visualization of HER-2-expressing xenografts in mice as early as 6 h postinjection. Due to increased affinity, the uptake for $Z_{HER2:342,}$ compared to the parental $Z_{HER2:4}$ monomer, improved by 2.03, 3.94, and 24.29 times at 1, 4, and 24 h, respectively, with a tumor uptake of ~ 9% injected dose (ID)/g at 4 h [30]. The relatively small size of affibody molecules also enables a direct chemical peptide synthesis to link chelators directly and site specifically at high yields [31].

Two CDRs, $V_H$CDR1 and $V_L$CDR3, through a cognate framework region ($V_H$FR2), yield small anibody mimetics. These mimetics retain the antigen recognition of their parent molecules but have a superior capacity to penetrate tumors. The antigen recognition abilities of these approximately 3-kDa mimetics surpass those of comparable fragments lacking the framework region. In vivo activities of the mimetics suggests that the structural orientation of their CDRs approximates the conformation of the CDRs in the complex of the parent antibody with antigen [32]. Currently, there is no RIC available based on this small antibody mimetics.

### 3.3 Antibody Affinity

Antibody affinity is defined as the strength with which an antibody molecule binds an epitope, that is, antigenic determinant. Mainly, the nocovalent interactions between antibody and its epitopes, including ionic interactions such as attraction of opposite charges on amino acids, hydrogen bonds and hydrophobic interactions, contribute to affinity, which can be quantified by determining an association constant, $K$. Abs with binding constants less than $10^8\,M^{-1}$ are not likely to be useful for drug targeting or tumor imaging [33]. Affinity maturation of antibodies is often associated with increased in vitro biological potency [34]. Evidence for increased affinity prolonging the tumor retention of radiolabeled antibody comes from studies using B72.3, an mAb with specificity for the pancarcinoma antigen TAG-72. The tumor targeting of B72.3 was compared to a series of

higher affinity second-generation mAbs, each reactive with a different epitope of TAG-72. The second-generation mAb CC49 ($K_d = 6 \times 10^{-11}$) localized human tumor xenografts in athymic nude mice better than did the lower affinity molecule B72.3 ($K_d = 2 \times 10^{-9}$) [35]. Through recombining, engineering, and phage display screening, recombining antibodies, Fab fragments, and mimetics with enhanced affinity could be obtained [9]. In a striking example, the cumulative effect of many small structural alterations resulted in a single-chain antibody fragment (scFv) with 1800-fold increase in fluorescein-binding affinity compared with its high affinity wild type [36]. By creating mutant scFv gene repertoires and displaying their encoded proteins on the surface of filamentous bacteriophage, affinity mutants of a human scFv (C6.5) that bind to an epitope on the extracellular domain of HER-2 has been created [37]. The scFv mutants span a 320-fold range of affinities. After iodination, biodistribution studies were performed in severely compromised immune deficient (SCID) mice bearing human ovarian SKOV-3 tumors. At 24 h after injection, tumor retention of the highest affinity scFv was 7-fold greater than that of a mutant with 320-fold lower affinity [37].

However, there is another contradictory theory about the potential impact of affinity on the successful targeting and retention of mAb in tumors. Fujimori et al. [38] has postulated that high-affinity mAb will not successfully penetrate deeply into tumors due to a binding site barrier effect in which interaction with the first antigen encountered at the periphery of the tumor will block further diffusion of the mAb into the tumor. Support for this hypothesis comes from the observation that intravenous (IV) administration of the mAb $^{131}$I-D3, specific for the chemically induced L10 carcinoma line ($K_a = 1.6 \times 10^{10}\,M^{-1}$), to guinea pigs bearing intradermal L10 tumors, resulted in peripheral binding to antigenic patches, whereas $^{125}$I-labeled control mAb was uniformly distributed throughout the tumor [39]. Moreover, even though localized better in animal models, in subsequent clinical trials, the second-generation mAb CC49 failed to display improved tumor retention over that observed for B72.3 [40]. Blindly increased affinity, however, might approach an upper limit of antigen-binding affinity beyond which further increases in affinity do not enhance tumor localization [41].

The relationship between the antigen-binding affinity of an antibody and its in vivo properties, with tumor targeting being the most widely studied property, is still unclear. With a panel of HER-2-specific scFvs, it has been demonstrated that the antigen-binding affinity of an antibody influences tumor localization. A threshold affinity ($K_d$) of $10^{-7}$–$10^{-8}\,M$ was required for specific tumor localization of scFvs, with uptake reaching a plateau at a $K_d$ of $10^{-9}$–$10^{-11}\,M$. The lowest affinity scFv had the most uniform distribution throughout the tumor, whereas the highest affinity scFv was found mainly in the perivascular region of the tumor [42]. There may exit an upper limit of antigen-binding affinity beyond which further increases in affinity do not seem to be beneficial for tumor localization. Conversely, overly high binding affinity will limit tumor penetration, which is consistent with the "binding site barrier," as described by Weinstein and colleagues [38]. In addition, simple dimerization of a low-affinity scFv can lead to considerably greater tumor localization than can be achieved by the more time-consuming process of scFv affinity maturation [43]. It has also been reported that no significant difference in tumor uptake was found between both anti–carcinogenic antigen (CEA) antibodies, despite their 10-fold difference in affinity [39]. The marked effect of antigen-binding affinity on the in vitro and in vivo activities of antibodies supports the idea that affinity also has a considerable impact on the clinical utility of antibodies. Whether there is an optimal antigen-binding affinity for antibodies remains unexplored. Antibody internalization may also affect the tumor uptake and retention of RIC. In a human colon cancer LS174T mouse tumor model, biodistribtution study showed that the internalizing antibody was accumulated more rapidly compared with the noninternalizing antibody. However, the maximum uptake showed no statistical difference [44].

## 4 ANTIBODY LABELING

### 4.1 Halogens

Initially, halogen radionuclides, especially iodine, were used in RII ($^{123}$I) and clinical RIT ($^{131}$I) trials because it was readily available, and labeling methods were well established, with iodine usually covalently linked to the antibody through tyrosine residues [45]. For direct iodination, it has been noted that even under very mild oxidizing conditions, some proteins lose a portion of their biologic activity, most likely from oxidation of thiols. For Abs, the loss of activity might mainly result from iodination of a tyrosine that was involved in the binding. Moreover, iodination of tyrosine residues or agents with a similar structure tends to be recognized and dehalogenized in vivo by deiodinases. In addition, if bound to tumor antigen and internalized into the cell, as in some tumor systems, further dehalogenation is possible. Free iodine usually deposits in the thyroid, salivary glands, and stomach [46]. Radiohalogenation of proteins using indirect methods has been investigated, including optimization of labeling chemistry and biological characterization of some labeled conjugates. Two indirect radiohalogena-

tion methods were used, representing two different labeling principles: activated ester of benzoic acid [47] and the derivative of *closo*-dodecaborate anion [48] (Fig. 3). The nonphenolic linker as well as the borate-halogen moiety probably prevent dehalogenation. The negative charge of the potential catabolic products of *closo*-dedecaborate anion might trap radiohalgodens intracellularly [48]. Antibodies can be conjugated with various *closo*-dodecaborate(2-) containing ligands. DABI, a derivative of *closo*-dodecaborate, [(4-isot hiocyanatobenzylammonio)undecahydro-*closo*-dodecaborate], has been successfully radioiodinated with a high yield (93–95%) and coupled to an antibody. In vivo data revealed that $^{125}$I-DABI-mAb conjugates yielded a low thyroid uptake with no accumulation compared with a direct labeling antibody [48]. Compared with direct labeling, indirect labeling usually generates lower yields and is more time-consuming and complicated.

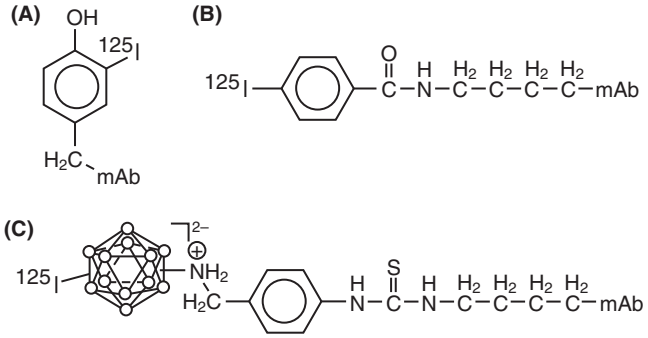


**Figure 3** Structures of the three different $^{125}$I-antibody conjugates used. (A) Directly labeled mAb, (B) indirectly PIB-labeled mAb, and (C) indirectly DABI-labeled mAb [48].

### 4.2 Coppers

Most radionuclides, especially radiometals, need to be labeled to Abs by indirect labeling methods with bifunctional chelates (Fig. 4). Bifunctional chelators contain, in addition to the metal chelating unit, a linkage and conjugation group for attachment to the tumor-seeking moiety, that is, Abs in RICs. Conjugation of chelates to antibodies is performed via electrophilic conjugation groups (activated esters), attached to a linker, which can be a phenyl group or a short peptide in addition to the phenyl group (Fig. 4). For copper ($^{67}$Cu and $^{64}$Cu) labeling, several chelating agents are available. The prototype is the macrocyclic chelator 1,4,8,11-tetraazacyclotetradecane-*N*,*N*′,*N*″,*N*‴-tetraacetic acid (TETA) developed for copper labeling by the Meares group [49]. TETA is a bifunctional carboxylated cyclam-14 derivative that complexes copper rapidly at ambient temperature to maintain protein function. The TETA ligand, as well as a number of other 4N macrocycles, forms copper(II) complexes of high thermodynamic stability, which also exhibit high kinetic stability under physiological conditions. Despite the high kinetic stability of the TETA copper complex, transchelation of Cu from the complex to these proteins in the liver has been observed in vivo with $^{67}$Cu-TETA-labeled antibodies and $^{64}$Cu-TETA-labeled octreotide [50]. Cross-bridged cyclam ligands were synthesized with increased reduction potential in order to lessen the electrochemical reduction of Cu(II) to Cu(I), which is believed to affect the dissociation of copper from the complex [51]. However, complexation of copper with these ligands requires elevated temperatures (75°C), which is not optimal for postconjugation labeling of proteins, and these ligands, therefore, may be more suitable for peptide labeling. We have labeled with various

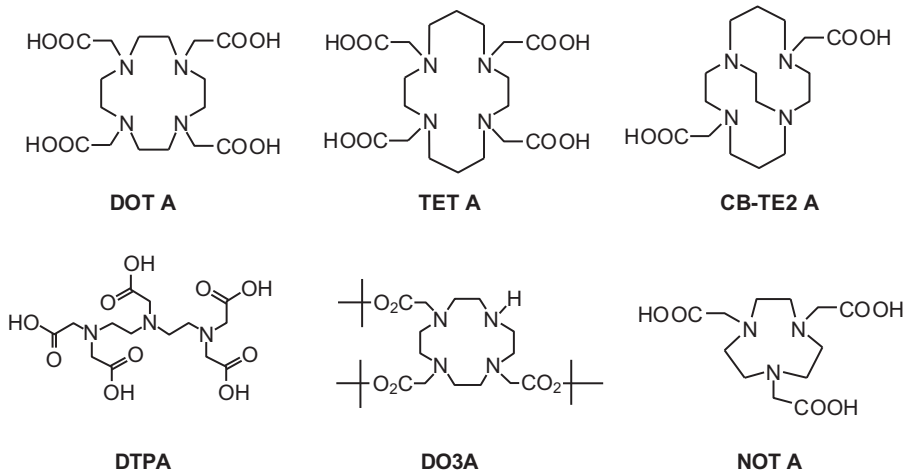


**Figure 4** Representative chelators for radiometal labeling.

peptides and antibodies with $^{64}$Cu using bifunctional 1,4,7,10-tetraazacyclododecane-N,N′,N″,N‴-tetraacetic acid (DOTA) as chelating agent and performed PET imaging in animal models [52–54]. We have found that the $^{64}$Cu-DOTA complex is very stable when attached to macromolecules or nanoparticles, with the majority of the $^{64}$Cu label remaining intact after incubation in full mouse serum at 37°C for 24 h [55].

Besides the in vivo stability of the used copper complexes, their charge and lipophilicity also have an important impact on the biodistribution of Cu-labeled immunoconjugates. Macrocyclic copper complexes bearing different charges and exhibiting different degrees of lipophilicity, such as $^{67}$Cu-TETA or -DO3A complexes, lead to distinct differences in biodistribution when used for labeling of antibody fragments. The chelates (TETA, DOTA, DO3A) with carboxylic groups (p$K_a$ values in the range of 2–5) attached to the nitrogen donor atoms are very hydrophilic. Indeed, clearance of radioactivity from the kidneys was found to be much faster in the case of antibody fragments substituted with negatively charged hydrophilic copper complexes [56]. The extent of conjugation of chelates to the antibody has to be optimized in every case in order to ensure that immunoreactivity is retained after conjugation. It also has to be considered that even when no effect on immunoreactivity is observed, substituting the protein with too large a number of ligands can lead to significant changes in biodistribution and increased activity in the liver [57].

### 4.3 Yttrium-90

Numerous chelating agents for $^{90}$Y have been investigated for RIT, primarily derivatives of the acyclic chelator diethylenetriaminepentaacetic acid (DTPA) [58] and the macrocyclic chelator DOTA [59]. Yttrium complexes of the acyclic chelators are not as kinetically stable as those of DOTA [60]. The extraordinary stability of the yttrium-DOTA complex minimizes the loss of $^{90}$Y to bone, a target organ of the radionuclide itself. However, the radiolabeling yields of $^{90}$Y-DOTA immunoconjugates are 50% typically, which is much less than that of $^{90}$Y-DTPA immunoconjugates (often in yields of >90%). Improved yields are needed to reduce the cost of therapy with $^{90}$Y-DOTA immunoconjugates and to minimize the exposure of radiolabeling personnel to radiation [59].

### 4.4 Chelator Antibody Ratio

Most indirect labeling of antibody is through the usage of a bifunctional chelator. The ideal condition is that after chelator conjugation, the antibody still keeps its immunoreactivity and kinetics in vivo. Several studies aimed at illustrating the effect of chelator/antibody ratio (c/a) on immunoreactivity and biodistribution have been performed. With mAbs labeled with $^{153}$Sm through the bifunctional DTPA molecule, rapid blood clearance and liver accumulation was observed in rats when 20 chelate groups were coupled per mAb [61]. A similar phenomenon was observed for mAbs labeled with $^{67}$Cu through the 14N4 macrocycle in mice [62]. To determine the maximum chelator: antibody ratio (c/a) of 2IT-BAT-Lym-1 at which functional integrity is preserved, immunoconjugates with a c/a of 1.3–23 were prepared and examined by radioimmunoassay and competitive antigen binding versus lightly iodinated Lym-1. Conjugates with a c/a up to 5 exhibited no loss of immunoreactivity, and conjugates with a c/a up to 11 retained 75% or greater immunoreactivity relative to unmodified Lym-1. Tumor uptake declined with increasing c/a, but the effect was insignificant at c/a 2.1 and 4.3. The biodistribution data suggested a negative correlation between c/a of 2IT-BAT-Lym-1 and uptake in tumor and a positive correlation between c/a and uptake in liver. Conjugates of c/a 4–5 were found to be optimal for the preparation of RIC of high specific activity with minimal, if any, loss of functional integrity [57]. In another study, magnesium based chalators (MAGs) were used as chelators to label mAb with $^{99m}$Tc and $^{186}$Re for RII/RIT. Immunoreactivity of mAbs remained nearly intact upon coupling of up to 20 $^{186}$Re-MAG3 groups. However, there was a serious alteration of the pharmacokinetic behavior of conjugates containing more than eight $^{186}$Re-MAG3 groups on average. There was a strong correlation between the $^{186}$Re-MAG3:MAb E48 molar ratio and blood clearance [63]. The reason for higher liver uptake may be due to nonspecific binding of the radiolabel to the mAb and the presence of aggregates [64]. The mechanism behind the rapid blood clearance of mAbs containing too many metal-chelator groups is not yet known. Some authors addressed that both the increased fraction of immunoconjugate with both binding sites inactivated and strong anodal shifts in isoelectric point (*p*I) of immunoconjugates contributed to the decline in tumor uptake and increased liver uptake [65]. In addition, the tolerance of metal-chelator conjugation seemed to be varying with different antibodies [57].

### 4.5 Biodistribution and Dosimetry Evaluation

Besides lesion detection, another main application of RII is to localize the biodistribution of RIC and calculate the dosimetry for effective RIT. Some therapeutic radionuclides, such as $^{131}$I, have γ-ray emission, which will facilitate imaging and following dosimetry

evaluation. However, some therapeutic radionuclides are pure β or α emitters, including $^{90}$Y, need another γ emitter to perform RII.

Biodistribution of $^{90}$Y for dosimetry calculations is typically obtained by imaging, using the surrogate radiometal $^{111}$In because $^{90}$Y does not emit photons. $^{111}$In has an almost identical half-life as $^{90}$Y, emits 2 γ rays of 171 and 245 keV, and can be readily incorporated into the same metal chelating agents as $^{90}$Y. For these reasons, $^{111}$In has been considered an excellent analog for $^{90}$Y. For example, the Food and Drug Administration (FDA) approved RIC ibritumomab tiuxetan (Zevalin, BiogenIDEC) labeled with $^{90}$Y to treat non-Hodgkin's lymphoma (NHL), and it is required that $^{111}$In-ibritumomab tiuxetan be used to produce images of the tumor and normal organs for dosimetry and biodistribution studies [66]. However, it has been reported that $^{111}$In-trastuzumab did not parallel the uptake of $^{86}$Y-trastuzumab in the bone and thus may not accurately predict the level of $^{90}$Y accumulation in the bone for clinical RIT applications [67, 68]. Subtle differences in radionuclide retention of the metal chelating agents used for $^{111}$In and $^{90}$Y mAb labeling can result in a differential release between the two radiometals. With less stability than $^{111}$In-DTPA, transchelation of $^{90}$Y from the chelating agent can result in a higher dose to the bone and bone marrow, which would not be anticipated from biodistribution data derived from $^{111}$In, possibly leading to underestimation of the radiation dose to bone marrow. Thus, quantitative information offered by PET through $^{86}$Y radiolabels could enable more accurate absorbed dose estimation for $^{90}$Y RIT [68]. A HER-2-specific antibody, trastuzumab, has been labeled with $^{86}$Y using 2-(4-isothiocyanatobenzyl) DTPA (SCN-CHX-A-DTPA) as chelator to assess the therapeutic potential of $^{90}$Y-trastuzumab in a human ovarian carcinoma (SKOV-3) model [69]. It has also been reported that $^{99m}$Tc-mAb conjugates showed a similar pharmacokinetic behavior as $^{186}$Re-mAb conjugates and can thus be used to predict the localization of $^{186}$Re-labeled mAbs and make dosimetric predictions in individual patients [63].

## 5 PRETARGETING

All pretargeting procedures use a modified mAb that not only recognizes a target antigen but also permits a second component to bind to it specifically. By separating the targeting steps, the mAb does not need to be treated with reagents used for conjugation to a suitable radionuclide, which ensures its immunoreactivity is uncompromised. The unlabeled modified mAb is administered first, distributes throughout the body to bind to the cells expressing the target antigen, eventually clearing from other tissues. The second component, which is radiolabeled, is then administered and, ideally, localizes at sites where the modified mAb has accumulated. If the second component has higher permeation, clearance, and diffusion rates than those of an mAb, the radionuclide will localize more rapidly to the tumor and higher tumor selectivity can be achieved compared to a directly radiolabeled IgG [70]. Pretargeting methods can be divided into several types such as biotin–avidin (or streptavidin) involved or a bispecific monoclonal antibody (bsmAb) involved pretargeting (Fig. 5) [71]. All radionuclide-pretargeting methods require the modified mAb to be cleared adequately from the blood before the injection of the radiolabeled effector; otherwise, a sizable portion of the effector would bind to it in the blood, taking on its pharmacokinetic and distribution properties. Regardless of the pretargeting method used, each has been shown in animal models and in patients to be highly efficient targeting procedures, with improved therapeutic properties as a result of the enhanced tumor/blood ratios [71, 72].

Two distinct methods for pretargeting using the avidin/streptavidin–biotin method have evolved. In one, StAv is conjugated to an antitumor IgG, making a conjugate with a molecular size of approximately 170–220 kD [73]. Because of their large size, a clearing agent is needed to remove excess StAv–IgG from the blood before the radiolabeled biotin injection; otherwise, the radiolabeled biotin, with its ultra-high affinity for StAv, would be bound tightly to the StAv conjugate in the circulation [74]. Though low concentration of biotin is present naturally in serum and tissues, this system has been used effectively in patients without efforts to reduce endogenous biotin levels. However, the presence of biotinidases in the serum may cleave the biotin–radionuclide conjugate [75]. The second StAv–biotin pretargeting configuration uses biotin-conjugated antitumor IgG as the modified mAb [76]. To make use of radiolabeled biotin instead of radiolabeled StAv or avidin, glycosylated avidin is injected first to remove the biotin–IgG conjugate from the blood. Within 30 min, nonglyscosylated StAv is given. As StAv enters the tumor, it binds tightly to any biotin–IgG it encounters. After allowing time for the avidin and StAv to clear from the body, radiolabeled biotin is given. With four biotin-binding arms, ample sites are available on the StAv within the tumor to foster the bridging of the radiolabeled biotin to the tumor-bound biotin–IgG [71].

A bispecific antibody (bsmAb) binds both to a target antigen as well as the radiometal–chelate complex of the effector molecule [77]. Unlike approaches based on avidin/streptavidin that rely on a clearing step, bsmAb pretargeting simply delays giving the radionuclide until the bsmAb has cleared sufficiently from the blood. Due to its smaller size, bsmAb clears more quickly than the

larger IgG-based conjugates. BsmAb pretargeting relies on a radiolabeled effector that bears two haptens capable of binding to the antihapten binding arm of the bsmAb. This technique, called the affinity enhancement system (AES), improves the uptake and retention of the radiolabeled hapten-peptide [78]. The major advantage for a bsmAb approach is the ability to use humanized antibodies that will have a lower immunogenicity than the foreign streptavidin or avidin proteins. In contrast to the monovalent chelate–radiometal complex, a bivalent structure would be bound more stably at the tumor site, presumably because the divalent hapten-peptide could be cross-linked by two adjacent bsmAbs [79]. The hapten portion of the bsmAb pretargeting system is responsible for binding to the antihapten arm of the bsmAb. Initially, the hapten was a radiometal-binding chelate, such as DTPA, which could also be loaded with a radiometal, such as $^{111}$In. Two DTPAs could be joined together synthetically by a short (e.g., usually 2–4 amino acids) peptide. The uptake of the radiolabeled hapten-peptide appears to be highly efficient from a number of perspectives. Because the hapten-peptide uptake achieves it maximum accretion within minutes [80], not only is the total radiation dose delivered to the tumor higher, but it is delivered at a much higher dose rate [81]. For imaging, this has resulted in tumor/blood ratios that are 40-fold higher than a directly radiolabeled Fab fragment [80]. In addition, the hapten-peptide can be radiolabeled at higher specific activities than a directly radiolabeled antibody, which would bring more radionuclide to the tumor per unit mass of hapten-peptide delivered.

The antigen marker for a pretargeting system being targeted must be accessible and should stay accessible in sufficient concentration to allow time for the untargeted bsmAb to clear from the blood. A targeted substance that internalizes quickly after being bound by the bsmAb would not be a good candidate for a pretargeting application, and, therefore, direct targeting would be preferred. With antigen being present in the blood, a directly radiolabeled antibody would likely result in higher accretion of radioactivity in the liver or spleen after forming complexes with the circulating antigen. In contrast, as long as the unconjugated bsmAb–antigen complexes are effectively removed before the administration of the radiolabeled hapten-peptide, efficient targeting of the tumor would be obtained with pretargeting strategy [82].

## 6 RII AND RIT OF MALIGNANT DISEASES

### 6.1 Lymphoma

The T101 monoclonal antibody recognizes a pan-T-cell antigen present on normal T cells and is also found in high concentrations in cutaneous T-cell lymphoma. Gamma-camera imaging using $^{111}$In-T101 has been performed in 11 patients with advanced disease to detect sites of metastatic cutaneous T-cell lymphoma [83, 84]. $^{111}$In-labeled MoAb directed against eosinophil peroxidase could localize lymphomas infiltrated by eosinophils [85, 86]. With $^{99m}$Tc-labeled a murine monoclonal antibody LL2 (EPB-2) Fab′, a number of previously unknown tumor sites were revealed by LL2-radioimmunodetection despite normal morphologic imaging results [87]. However, it has also been reported that radioimmunodetection with $^{99m}$Tc-labeled CD19 mAb is inferior to computed tomography (CT) and gallium-67 scanning in the diagnosis of patients with B-cell non-Hodgkin's lymphoma [88].

The CD20 antigen is an excellent target because it is expressed on pre-B cells and mature B cells, including up to 95% of B-cell NHL. Moreover, CD20 does not shed, modulate, or internalize. More importantly, it is absent from stem cells and plasma cells [89]. Rituximab, a chimeric monoclonal antibody directed against the CD20 antigen, was the first monoclonal antibody approved by the FDA for the treatment of NHL [90]. Rituximab immunotherapy is now routinely used in the treatment of low-grade NHL as an up-front single agent [91] or as part of combination chemotherapy [92]. By further improving the efficacy, anti-CD20 RIT has enhanced the survival not only of patients with large-cell NHL but also of those with follicular NHL.

Ibritumomab tiuxetan (Zevalin, BiogenIDEC) and tositumomab (Bexxar, GlaxoSmithKline) are both FDA-approved RICs for the treatment of patients with CD20 positive, follicular, non-Hodgkin's lymphoma. Zevalin contains ibritumomab, a murine anti-CD20 antibody, and tiuxetan, an MX-DTPA linker–chelator. Ibritumomab tiuxetan can be coupled with either $^{111}$In for tumor imaging and dosimetry or with $^{90}$Y for therapy. Bexxar is a murine monoclonal antibody that binds to CD20 and is radiolabeled with $^{131}$I for RIT. In contrast to $^{90}$Y-ibritumomab tiuxetan, the therapeutic dose of tositumomab is based on dosimetry. Patient-specific dosing is necessary because of the interpatient differences in spleen size, tumor burden, and the metabolism and renal excretion of $^{131}$I [93].

To address whether the anti-CD20 RICs really offered any advantages over the respective unlabeled anti-CD20 antibodies, $^{90}$Y-ibritumomab tiuxetan was compared to rituximab in patients with relapsed CD20$^{+}$ NHL who had never received rituximab [94]. The complete response (CR) rate was 30% in the $^{90}$Y-ibritumomab tiuxetan group and 16% in the rituximab ($P = 0.04$) group. The estimated median time to progression (TTP) was similar in both groups. However, the estimated time to next therapy for patients with nontransformed histology was significantly longer for $^{90}$Y-ibritumomab

tiuxetan patients (17.8+ months; range: 2.1–21.7+) than for rituximab patients (11.2 months; range: 1.3–19.0+) ($P = 0.040$). RIT for relapsed indolent NHL produces an overall response rate (ORR) of approximately 60–80% and a CR in 15–30% of patients. The RICs produce a higher response rate than their respective cold unlabeled antibody. RIT has a high response rate (approximately 75–80%) in rituximab-refractory patients.

### 6.2 Breast and Ovarian Cancer

As early as 1967, $^{131}$I-labeled antibodies to human fibrinogen have been used to localize mammary carcinoma and 12 of 17 lesions were visualized [95]. Consequently, various antibodies have been labeled with $^{123}$I, $^{131}$I, or $^{111}$In for breast or ovarian cancer localization or lymph node metastases [96–98]. Following IV administration, $F(ab')_2$ fragments cleared from the blood more rapidly than the intact IgG [98]. The mean tumor uptake of $^{111}$In-labeled HMFG1 $F(ab')_2$ fragments at 24 h was significantly higher than the intact IgG, but at 48 h there was no difference. Lymph node metastases showed higher antibody uptake than the corresponding primary tumors. Successful localization was observed in approximately 50% of the primary and metastatic lesions with no false-positive results. A monoclonal antibody MA5 reacted with greater than 95% of the primary and metastatic lesions and recognized a cell surface antigen. In 17 patients with breast cancer, $^{111}$In-labeled MA5 monoclonal antibody visualized bone lesions, primary tumors, soft tissue recurrences, and lung metastases larger than 3 cm diameter [99]. RII has been performed to image bone or lymph node metastases of breast cancer [100, 101]. $^{99m}$Tc-labeled antigranulocyte monoclonal antibody BW 250/183 (AGMoAb) scans in 32 patients showed bone marrow defects in 25/32 (78%) patients; bone invasion was subsequently confirmed in 23 (72%) patients [101]. Two of antiaberrant MUC-1 mucin epitopes mAbs, BrE3 and m-170, which show high levels of staining on most breast cancer biopsy specimens, have provided excellent tumor targeting in imaging studies [102].

Human epidermal growth factor receptor 2 (HER-2) is a cell membrane surface-bound receptor tyrosine kinase. Overexpression of HER-2 has been detected in up to 30% of breast and ovarian cancers [103]. The first HER-2-specific imaging study was performed in 1991 with a class-switched murine monoclonal antibody SV2-61r that recognizes the extracellular domain of HER-2 [104]. Monoclonal antibody ICR12 is a rat IgG2a with high affinity for an external epitope on the protein core of HER-2 (human epidermal growth factor) [105, 106]. In patients selected by IHC staining, planar and single photon emission computer tomography (SPECT) imaging with $^{99m}$Tc-ICR12 showed good tumor localization in both the primary lesion and regional node metastases [107]. Scintigram with $^{99m}$Tc-labeled another murine monoclonal antibody CIBCgp185 against HER-2 also showed high radiolabel uptake by the human breast cancer BT474 tumor tissue in mice [108]. PET imaging with $^{124}$I-labeled ICR 12 has been performed, and MDA-MB-361 tumor xenografts in nude mice of 6 mm diameter could be visualized at 24, 48, and 120 h after probe injection [109].

The murine mAb 4D5, directed against the extracellular domain of HER-2, is a potent inhibitor of growth of human breast cancer cells that overexpress HER-2 [110]. By inserting the CDRs of mAb 4D5 into the framework of a consensus human IgG, a recombinant humanized anti-HER-2 monoclonal antibody called trastuzumab has been developed. Trastuzumab is the first monoclonal antibody approved by the FDA for use in women with breast cancer who have tumors that overexpress HER-2 protein. Previously unidentified lesions were visualized in 13 of the 15 patients on the $^{111}$In-DTPA-trastuzumab scintigraphs, and the overall detection rate of tumor lesions was 45% at the single-lesion level [111]. The trastuzumab Fab was radioloabeled with $^{111}$In after DTPA derivatization, and the binding affinity was not significantly decreased with a dissociation constant ($K_d$) that changed from 14–36 to 47. $^{111}$In-trastuzumab Fab localized specifically in HER-2-positive BT-474 human breast cancer xenografts in athymic mice (compared with $^{111}$In-HuM195 anti-CD33 Fab). BT-474 tumors as small as 3–5 mm in diameter could also be imaged as early as 24 h postinjection [112]. A direct comparison between Fab and full-length antibody has been reported with $^{111}$In-labeled trastuzumab and a trastuzumab-derived Fab (Fab4D5) through DOTA conjugation [113]. In a spontaneous mammary tumor model in MMTV/HER-2 transgenic mice, $^{111}$In-DOTA-Fab4D5 showed early tumor localization at 2 h with little retention followed by rapid accumulation in the kidneys by 6 h, which indicated a rapid washout of Fab from the tumor after initial accumulation. $^{111}$In-DOTA-trastuzumab was slow to accumulate in tumors and slow to clear from normal tissues, and significant tumor deposition was achieved by 24 h. The peak tumor uptake for Fab was around 12% ID/g at 16 h postinjection and around 45% ID/g at 24 h postinjection for trastuzumab. By introducing an albumin binding sequence to Fab4D5, a bifunctional molecule AB.Fab4D5 capable of binding albumin and tumor antigen HER-2 simultaneously has been developed [114]. Vallabhajosula et al. [115] labeled $F(ab')_2$ fragment of trastuzumab with a positron emitter $^{68}$Ga ($t_{1/2} = 68$ min) to assess the degradation of HER-2 by 17-allylaminogeldanamycin (17-AAG), a heat shock

protein 90 (Hsp90) inhibitor. Taking advantage of the fast blood clearance of $F(ab')_2$ and quick decay of $^{68}Ga$, they repetitively imaged HER-2 expression at 24-h intervals. Based on the microPET quantification, HER-2 expression was reduced by almost 80% in the animals 24 h after 17-AAG treatment [115]. In a follow-up study, tumor response to 17-AAG treatment were assessed by $^{68}Ga$-DOTA-$F(ab')_2$-trastuzumab and $^{18}F$-FDG PET. Within 24 h after treatment, a significant decrease in HER-2 was measured by HER-2 PET, whereas $^{18}F$-fludeoxyglucose (FDG) PET uptake was virtually unchanged. The 17-AAG-treated animals had a reduced uptake of $^{68}Ga$-$F(ab')_2$-trastuzumab that lasted until 5 days after treatment, although notable growth inhibition occurred only by 11 days after treatment. The $^{18}F$-FDG PET imaging data revealed no significant differences in $^{18}F$-FDG metabolism between treated and control group at any of the time points examined. This indicates that HER-2-PET with $^{68}Ga$-DOTA-$F(ab')_2$-herceptin could provide accurate information for tumor early response to 17-AAG treatment [116].

By creating mutant scFv gene repertoires and displaying their encoded proteins on the surface of filamentous bacteriophage, affinity mutants of a human scFv (C6.5) that bind to an epitope on the extracellular domain of HER-2 has been created. Tumor retention of the highest affinity scFv was 7-fold greater than that of a mutant with 320-fold lower affinity at 24 h after injection [37]. Divalent and multivalent scFvs (diabodies or tribodies) can result either from the spontaneous dimerization or covalent association of 2 or more monovalent scFvs [19, 20]. Compared to a monovalent 741F8 Fab fragment of approximately the same size, the divalent bismalimidohexane-linked 741F8 $(sFv')_2$ showed significantly greater tumor localization. This suggests that the increased avidity of the $(sFv')_2$ was a factor in its improved tumor retention [17, 117]. Instead of direct iodination by the chloramine-T method, 741F8-1 $(sFv')_2$ has also been radioiodinated by using *N*-succinimidyl *p*-iodobenzoate (PIB), a protein radioiodination acylating agent. The PIB method involves the coupling of the radioiodine to a nonphenolic aromatic ring of the PIB reagent that is not susceptible to in vivo dehalogenase activity. The $^{125}I$-PIB complex is then conjugated to lysine residues on the scFv through the formation of an amide bond [118]. This method leads to high retention of the radiolabel on the protein and results in prominent tumor localization in tumor-bearing SCID mice [119]. However, dehalogenation of the radiotracer resulted in significant uptake of iodine by the thyroid and stomach. C6.5 diabody has also been labeled through a water-soluble form of Bolton–Hunter reagent sulfosuccinimidyl-3-(4-hydroxyphenyl) propionate (SHPP) to avoid dehalogenaton. Animals receiving SHPP-labeled diabody contained lower levels of free iodine in their blood than animals receiving iodogen-labeled diabody. However, the SHPP methodology decreased the immunoreactive fraction of the diabody about 2-fold compared with the iodogen-labeling methodology. Moreover, PET images of animals injected with $^{124}I$-SHPP-C6.5 diabody exhibited an overall increase in background signal due to residualization of the SHPP moiety [120].

Breast-cancer-targeted RIT has been performed in animal models first and $^{131}I$-labeled monoclonal antibody BW 495/36 showed a high tumor dose accumulation and significant reduction in tumor diameter with 42 days postinjection [121]. RIT with $^{131}I$-labeled ICR12 in athymic mice bearing MDA-MB-361 human breast cancer xenografts showed growth inhibition for >24 days at a dose of 600 μCi/mouse [105, 122]. RICs targeting to HER-2 are attractive for RIT because of the relationship of this target to very aggressive breast cancers. Trastuzumab has been used in RIT for tumor control after being labeled with β-particle emitters such as $^{131}I$ and $^{186}Re$ and α-emitter $^{90}Y$ [123, 124].

Besides HER-2, several antigen targets have proven effective for targeting metastatic breast cancer [125]. The L6 cell surface antigen, which is highly expressed on lung, breast, colon, and ovarian carcinomas, is a 24-kDa surface protein containing 3 hydrophobic transmembrane regions [126]. RIT with $^{131}I$-ChL6 provided biologic enhancement. By binding vascular endothelium, the resulting complement activation by L6 MAb increased tumor uptake of the subsequently delivered RIT [127]. An infusion of 100–200 mg of nonradioactive ChL6-saturated endothelium antigens allowed subsequently administered $^{131}I$-ChL6 to reach extravascular tumor. Activation of complement on the endothelium was associated with increased vessel permeability and uptake of $^{131}I$-ChL6 by tumor. Multiple infusions of $^{131}I$-ChL6 to 13 women with metastatic breast cancer showed that 6 of these patients had clinically measurable tumor responses; 4 had partial response [128].

RIT has been performed in breast cancer patients using T84.66, a murine $IgG_1$ with high specificity and affinity for CEA [129]. Using PBSC after radiological imaging technology (RIT) with $^{90}Y$-DTPAcT84.66, the 6 patients treated (15–22.5 mCi/m$^2$) had temporary but measurable therapeutic responses, although the maximum tolerated dose was not reached [130]. Several mAbs react with aberrant MUC-1 mucin epitopes upregulated on breast cancer. MAb to both the MUC-1 peptide core and immature sugar residues have been studied in RIT trials of breast cancer [99, 131]. Pharmacokinetics, dosimetry, and therapy trials of RIC to these epitopes in metastatic breast cancer have been well described [129]. BrE-3 to MUC-1 core peptide

epitope, in a dose escalation study as $^{90}$Y-MX-DTPA BrE-3, decreased measurable disease in 50% of patients after a single $^{90}$Y-MAb dose [132]. A single, higher dose of $^{90}$Y BrE-3 and PBSC also resulted in objective partial responses in 50% of the patients, but Abs against the mouse mAb, human antimouse antibody (HAMA), developed rapidly in most patients, so that additional therapy doses were not administered [132]. Humanized BrE-3 as $^{90}$Y MX-DTPA-hBrE-3 has been used in a dose escalation protocol with PBSC. Of 4 patients with measurable disease, 2 partial response, 1 mixed response, and 1 stable disease were reported. Anti–MUC-1 MoAb 170H.82 (m-170), against a synthetic asialo GM1 terminal disaccharide related to the Thonsen–Friedereich disaccharide, demonstrated 90% staining of 99 metastatic breast cancer biopsy specimens [125]. $^{111}$In/$^{90}$Y-DOTA-peptide-m170 has been studied in breast cancer patients and compared with $^{111}$In/$^{90}$Y-2ITBAD- m170 to evaluate the effect of a radiochelate linkage that can be catabolized in the liver. Both RICs showed excellent tumor targeting, and the new cathepsin degradable peptide linkage resulted in an improved tumor to liver radiation dose by 30%. However, PBSC support was again needed to deliver effective tumor therapy [61]. Two mAbs, B72.3 and CC49, to the cancer-related glycoprotein TAG-72, have been extensively studied for RIT in breast cancer and ovarian cancer [133]. Recently, combined-modality radioimmunotherapy (CMRIT) has been performed in human breast animal model with $^{90}$Y-radiolabeled humanized anti-Lewis Y hu3S193 monoclonal antibody and paclitaxel. CMRIT displayed significant improved efficacy over single-modality therapies [134]. In another CMRIT study, paclitaxel was combined with $^{90}$Y-DOTA-776.1, a monoclonal antibody prepared against the human ovarian cancer antigen CA125. Results indicated a synergistic antitumor effect on the growth of OVCAR-3 xenografts [135]. The heterogeneity of tumor antigens expression may be an obstacle of RIT and therapeutic efficacy for ovarian cancer treatment will be enhanced by the combined targeting of TAG-72, MUC1, and CA125 [136]. However, It was found that in patients with recurrence in epithelial ovarian cancer receiving intraperitoneal consolidation treatment with yttrium-90-labeled murine HMFG1, there is no improvement in overall survival [137].

### 6.3 Prostate Cancer

In 1983, Goldenberg et al. [138] first demonstrated that a prostate-associated marker could be targeted and imaged by Abs was the use of rabbit Abs against prostatic acid phosphatase (PAP) labeled with $^{131}$I. PAP is an enzyme produced by the prostate. It has been found in increased amounts in men who have prostate cancer [139]. With $^{111}$In labeled an anti-PAP MoAb (PAY-276), RII resulted in a tumor detection rate of 15% in 10 patients with prostate cancer [140]. Besides PAP, several other prostate-cancer-specific antigens have been used as target for detection of prostate cancer lesions and metastasis. These antigens include prostate-specific antigen (PSA) [141] and prostate-specific membrane antigen (PSMA) [142].

PSMA is a surface protein highly expressed on prostate cancer cells compared with other tissues [143]. 7E11.C5 is a murine monoclonal antibody specific for PSMA. An $^{111}$In-7E11.C5 (Prostascint, Cyt-356, capromab pendetide; Cytogen Corp., Princeton, NJ) has been approved by the FDA as a diagnostic agent for prostate cancer [144]. In a patient population of 40, $^{111}$In-CYT-356 immunoscintigraphy detected bony metastases in 21 of 38 patients (55%) [143]. It also has been applied to detect sites of occult prostate cancer and assess the pelvic lymph nodes [145, 146] and helpful in the detection of occult recurrent carcinoma after radical prostatectomy [147]. In another study, ProstaScint imaging result was found to be not as good as the rate of increase of PSA in assessing the response to salvage radiation therapy [148]. Though the epitope on PSMA recognized by 7E11.C5 has been determined to be the first 6 amino acids from the N-terminal, expressed on the cytoplasmic side of the plasma membrane, uptake and labeling of PSMA by fluorescient isocyanate (FITC)-labeled 7E11. C5 in viable cells in vitro has been observed. This is a major basis for the effectiveness of the 7E11.C5 antibody during in vivo imaging applications with $^{111}$In-labeled antibody (ProstaScint, Cyt-356) [149].

The treatment of metastatic prostate cancer with RIC was initiated by Meredith et al. [150] with $^{131}$I-labeled CC49 mAb to TAG 72.3. The majority of these patients experienced pain relief at radiation doses calculated to be greater than 25 Gy, representing a 20-fold increase over the whole-body radiation dose. The CYT-356 mAb has also been labeled with $^{90}$Y for systemic RIT of disseminated prostate cancer following failed radical prostatectomy. However, there was no therapeutic effects observed in a dose escalation trial [151]. The J591 monoclonal antibody is one of the newer generation of antibodies that have been identified for therapeutic purposes based on their recognition of and avidity for extracellular epitopes [152]. Milowsky et al. [153] initially reported a phase I trial using $^{90}$Y-labeled humanized J591 mAb in which 29 patients with androgen-independent prostate cancer were treated and 2 patients experienced objective disease responses. An α-particle therapeutic, $^{312}$Bi conjugated to this mAb through a chelating agent, *N*-[2-amino-3-(*p*-isothiocyanatophen-yl)propyl]-trans-cyclohexane-1, 2-diamine-*N*,*N*′,*N*′,*N*″,*N*‴-pentaacetic acid, reduced PSA levels and increased tumor-free survival in mice with human prostate cancer xenografts [154].

Although MUC-1 targets in breast cancer have been extensively studied over the last two decades, immunohistopathologic studies in prostate cancer have only recently been performed and have demonstrated upregulated and hypoglycosylated MUC-1, providing epitopes for targeted therapy [155]. The implications of these observations have been corroborated by excellent targeting of $^{111}$In/$^{90}$Y-2IT-AD-m170 mAb in prostate cancer patients. Calculated tumor doses ranged from 2 to 20 Gy with PSA responses and bone pain relief [156]. Initial CMRT with this same RIC and paclitaxel in prostate cancer has also shown PSA responses along with measurable tumor response but with the need for PBSC [155].

### 6.4 Colon Cancer

The first reported radioimmunoscintigraphy was performed in 1974 with human colon carcinoma grafted into nude mice using $^{131}$I-labeled antibodies to CEA [157]. A following study in 52 colorectal carcinoma patients and 15 patients with other types of cancer showed that 34 of 63 colorectal carcinoma tumor sites had significant accumulation of antibody, whereas none of the 20 sites of other cancer types gave positive results. $F(ab')_2$ fragments of the mAb gave a slightly higher percentage (61%) of positive results than did intact mAbs (51%) [158]. Gamma camera, SPECT, and PET imaging have been performed with anti-CEA antibodies and antibody fragments conjugated with corresponding radionuclides, namely $^{131}$I, $^{111}$In, and $^{124}$I, respectively [159–161]. RII using anibodis against Tag-72 [162], fibroblast activation protein (FAP) [163] and sialosyl-Tn antigen in mucin [164], and the tumor-associated antigen Lewis Y oligosaccharide [68] have been performed in human colon cancer animal models or patients with colon cancer.

A clinical testing of a bsmAb pretargeting method for cancer imaging reported successful localization of colon cancer metastases in the liver using an $^{111}$In-chelate complex given 5 days after the administration of an unlabeled anti-CEA antichelate bsmAb [165]. A bsmAb reacting with both the adhesion molecule CD44V6, a tumor-associated antigen upregulated in the colon and a low-molecular-weight Ga chelater labeled with positron emitter $^{68}$Ga (hapten). In human colon carcinoma-bearing nude mice, multistep immunoscintigraphy with bsmAbs showed an initial tumor uptake approximately of 15.5% ID/g [166]. In another study, the first antibody arm of the bsmAb recognized CEA and the secondary arm againsts a histaminyl-succinyl-glycine (HSG) structural subunit. The bsAbs were prepared as Fab′ × Fab′ conjugates using chemical cross-linking methods and as bispecific diabodies using recombinant DNA technologies. An HSG-bivalent hapten conjugate bearing the macrocyclic ring chelating agent 1,4,7,10-tetraazacyclododecane-*N*,*N*′,*N*″,*N*‴-tetraacetic acid (DOTA) was radiolabeled with $^{68}$Ga or $^{67}$Ga. High tumor-to-normal tissue ratios of this system were found at 1–6 h after injection of the $^{67}$Ga-labeled bivalent hapten peptide [167]. With the similar construct labeled with $^{124}$I, tumor uptake measured at necropsy was 3- to 15-fold higher and tumor-to-blood ratios were 10- to 20-fold higher than those for $^{124}$I-Fab′ at 1 and 24 h, respectively. Tumor visualization with $^{18}$F-FDG at approximately 1.5 h was also good but showed substantially more uptake in several normal tissues [168].

Response rates from clinical trials of RIT in colon cancer have been pretty low. Therapy with the human immunoglobulin G4 chimeric version of the mAb $^{131}$I-ch-B72.3, an $IgG_1$ that recognizes a mucinlike colon cancer associated antigen, in 12 patients resulted in a high incidence of HAMA [169]. Using another target for colorectal cancer RIT, the A33 antigen, a transmembrane glycoprotein of the immunoglobulin superfamily [170], 23 patients were treated with escalating doses of $^{131}$I-A33 MAb, which demonstrated variable uptake in normal bowel and no objective responses [171]. The NR-LU-10 pancarcinoma mAb (NeoRx Corp., Seattle, WA) was also studied in colorectal cancer patients using a pretargeting scheme and a streptavidin conjugate of the Ab. Twenty-five patients treated with a single dose (110 mCi/m$^2$) of $^{90}$Y-DOTA–biotin had a response rate of 8%, and 4 patients had stable disease with freedom from progression for 10–20 weeks [8].

Among 119 tumors in 93 patients with CEA-expressing cancers that underwent radioimmunotherapy with the $^{131}$I-labeled IgG, anti-CEA antibodies NP-4 ($K_a = 10^8 M^{-1}$) or MN-14 ($K_a = 10^9 M^{-1}$), its humanized form hMN-14, as well as the anticolon-specific antigen-p (CSAp) antibody, Mu-9, tumor uptake appeared to be the most important dose-determining factor. No significant difference in tumor uptake was found between both anti-CEA antibodies, despite their 10-fold difference in affinity. Since an inverse logarithmic relationship between tumor size and antibody uptake was found for both anti-CEA antibodies, anti-CEA antibodies seemed to prevail in colorectal cancer patients having minimal disease [39]. This has also been demonstrated in animal models [172]. Behr et al. [173] reported responses in patients refractory to 5-fluorouracil but with small-volume metastatic disease (<3 cm) or patients in an adjuvant setting after metastatic liver resection. Treatment consisted of a single injection of 60 mCi/m$^2$ of humanized anti-CEA mAb, $^{131}$I-hMN-14. The objective response rate in 19 assessable patients was 16% and an overall response rate of 58% (partial, minor responses) with a mean duration of 9 months. Seven of the 9 patients given adjuvant RIT remained disease free

after follow-up of up to 36 months. Five patients with stabilized disease were retreated at the same dose and 2 of 4 assessable patients obtained a partial response. In moving toward CMRIT, studies by Buchegger et al. [174] showed that RIT with external-beam radiation was efficient as an adjuvant after surgery. Clinically, 6 patients with limited liver metastases from colorectal cancer were treated with RIT using 20 mCi $^{131}$I-labeled anti-CEA F(ab′)$_2$ combined with external-beam radiation of 20 Gy to the liver. Although reversible bone marrow and liver toxicity were observed, 3 patients had stable disease and 1 had a partial response.

## 7 OUTLOOK AND PERSPECTIVES

RII and RIT have played important roles in lesion localization and treatment, especially those with refractory malignant diseases. To further improve the efficacy of diagnosis and therapy, the challenges such as the selection of antibody ligand and radioisotope, bioconjugation and immunoreactivity, metabolic stability, and in vivo kinetics of the RICs all need to be optimized. All the improvements and refinery of RICs are centered on the three important components, which are radionuclides, chelators, and antibodies. Besides emission characteristics and half-life, several practical concerns should be considered when selecting the appropriate radionuclides, including their availability within an academic or commercial setting, the chemical processes required for their coupling to the targeting carrier, and the biodistribution of the radionuclide in the biological tissues and its decay chain.

For clinical application, site-specific, homogenous, reproducible, and easy radiolabeling is desirable. However, the most commonly used targeting agents are usually modified at multiple and randomly distributed sites by using modification chemistries based on amine, thiol, or tyrosine-reactive reagents. Thus, the resulting targeting molecules are heterogeneous with various degrees of modification. By using synthetic chemistry of HER-2-specific affibody molecules, the chelators can be introduced site specifically [29, 30, 175]. Site-specific radiolabeling and imaging for relatively small biomolecules such as VEGF [176] and scFv [177] have been reported. The relative big molecules such as anti-CEA diabody (55 kDa) also has been reported to be site specific and labeled after cysteine modification [177]. However, the optimal strategy for intact antibody site-specific labeling remains to be worked out. ROI analysis techniques currently used for imaging quantification need to outline the tumor region on images. The continuing development of SPECT/CT, PET/CT and PET/MRI will provide more accurate tumor outlining based on the anatomical images acquired by CT or MRI [178].

RII may be applicable to therapy response monitoring in addition to lesion localization and provide guidance to RIT. For example, HER-2 is one of the critical client proteins for Hsp90, and HER-2 level inside tumors declines dramatically upon heat shock protein 90 inhibitor, 17-AAG treatment. HER-2 degradation during 17-AAG treatment could be visualized and quantified by $^{68}$Ga-HER-2-PET [115]. More importantly, HER-2 PET is an earlier predictor of tumor response to 17-AAG therapy than $^{18}$F-FDG PET [116]. Recently, our group labeled anti-EGFR mAb with $^{64}$Cu to noninvasively monitor the EGFR degradation, upon

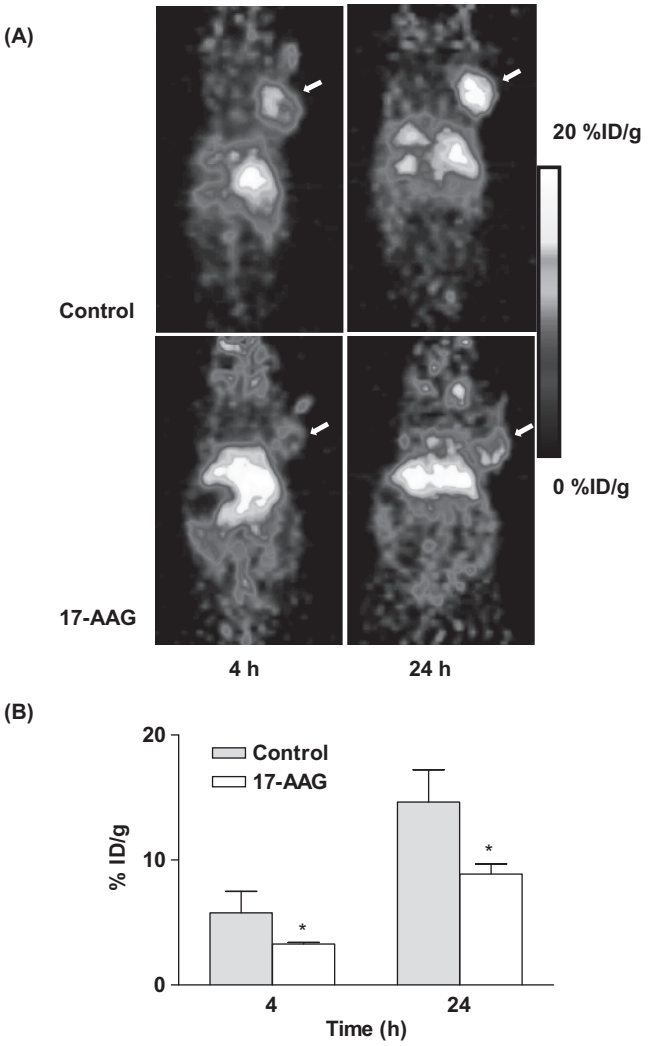


**Figure 5** (**A**) MicroPET images of PC-3 tumor-bearing nude mice at 4 and 24 h after intravenous injection of $^{64}$Cu-DOTA-cetuximab with or without 17-AAG treatment, and (**B**) PC-3 tumor uptake of $^{64}$Cu-DOTA-cetuximab as quantified from microPET scans.

17-AAG treatment through microPET imaging. The results showed that quantitative PET imaging of EGFR expression can detect the early response to 17-AAG treatment in prostate tumors (Fig. 5) [179]. Histone deacetylase (HDAC) is capable of regulating gene expression through diverse mechanisms, including modulation of the histone code and acetylation of many nonhistone proteins [180, 181]. HDAC inhibitors are a structurally diverse group of drugs that show preclinical and clinical promise as cancer therapeutics [182]. HDAC inhibitors, for instance, can selectively reduce HER-2 or EGFR expression by repressing the promoter and accelerating the decay of cytoplasmic transcripts [183]. RII for HER-2 or EGFR may also be useful to monitor the therapeutic response of HDAC inhibitors.

In conclusion, RII and RIT have made a great contribution to both preclinical studies and clinical applications. Along with the development of more specific antibodies, more suitable radionuclides, and more advanced imaging and therapeutic apparatus, better tumor lesions detection and treatment will be accomplished. At the same time, broader application of this technology to the management of cancer can be anticipated.

## REFERENCES


1. Mather SJ (2007). Design of radiolabelled ligands for the imaging and treatment of cancer. *Mol Biosyst 3*, 30–5.
2. Verel I, Visser GWM, van Dongen GA (2005). The promise of immuno-PET in radioimmunotherapy. *J Nucl Med 46*(Suppl 1), 164S–171S.
3. Remmenga SW, Colcher D, Gansow O, Pippen CG, Raubitschek A (1994). Continuous infusion chemotherapy as a radiation-enhancing agent for yttrium-90-radiolabeled monoclonal antibody therapy of a human tumor xenograft. *Gynecol Oncol 55*, 115–122.
4. Zhang M, et al. (2003). Pretarget radiotherapy with an anti-CD25 antibody-streptavidin fusion protein was effective in therapy of leukemia/lymphoma xenografts. *Proc Natl Acad Sci USA 100*, 1891–1895.
5. Couturier O, et al. (1999). Validation of 213Bi-alpha radioimmunotherapy for multiple myeloma. *Clin Cancer Res 5*, 3165s–3170s.
6. Kraeber-Bodere F, et al. (2002). Enhanced antitumor activity of combined pretargeted radioimmunotherapy and paclitaxel in medullary thyroid cancer xenograft. *Mol Cancer Ther 1*, 267–274.
7. Jurcic JG, et al. (2002). Targeted alpha particle immunotherapy for myeloid leukemia. *Blood 100*, 1233–1239.
8. Knox SJ, et al. (2000). Phase II trial of yttrium-90-DOTA-biotin pretargeted by NR-LU-10 antibody/streptavidin in patients with metastatic colon cancer. *Clin Cancer Res 6*, 406–414.
9. Holliger P, Hudson PJ (2005). Engineered antibody fragments and the rise of single domains. *Nat Biotech 23*, 1126–1136.
10. Kohler G, Milstein C (1975). Continuous cultures of fused cells secreting antibody of predefined specificity. *Nature 256*, 495–497.
11. Baldwin RW, Pimm MV (1983). Antitumor monoclonal antibodies for radioimmunodetection of tumors and drug targeting. *Cancer Metastasis Rev 2*, 89–106.
12. Sandhu JS (1992). Protein engineering of antibodies. *Crit Rev Biotechnol 12*, 437–462.
13. Goldenberg DM (1987). Current status of cancer imaging with radiolabeled antibodies. *J Cancer Res Clin Oncol 113*, 203–208.
14. Esteban JM, et al. (1987). Quantitative and qualitative aspects of radiolocalization in colon cancer patients of intravenously administered MAb B72.3. *Int J Cancer 39*, 50–59.
15. Clauss MA, Jain RK (1990). Interstitial transport of rabbit and sheep antibodies in normal and neoplastic tissues. *Cancer Res 50*, 3487–3492.
16. Wu AM, Senter PD (2005). Arming antibodies: Prospects and challenges for immunoconjugates. *Nat Biotechnol 23*, 1137–1146.
17. Jain M, Batra SK (2003). Genetically engineered antibody fragments and PET imaging: A new era of radioimmunodiagnosis. *J Nucl Med 44*, 1970–1972.
18. Adams GP, et al. (1993). Highly specific in vivo tumor targeting by monovalent and divalent forms of 741F8 anti-c-erbB-2 single-chain Fv. *Cancer Res 53*, 4026–4034.
19. Batra SK, Jain M, Wittel UA, Chauhan SC, Colcher D (2002). Pharmacokinetics and biodistribution of genetically engineered antibodies. *Curr Opin Biotechnol 13*, 603–608.
20. Todorovska A, Roovers RC, Dolezal O, Kortt AA, Hoogenboom HR, Hudson PJ (2001). Design and application of diabodies, triabodies and tetrabodies for cancer targeting. *J Immunol Methods 248*, 47–66.
21. Wu AM, Yazaki PJ (2000). Designer genes: Recombinant antibody fragments for biological imaging. *Q J Nucl Med 44*, 268–283.
22. Adams GP, et al. (2006). Avidity-mediated enhancement of in vivo tumor targeting by single-chain Fv dimers. *Clin Cancer Res 12*, 1599–1605.
23. Goel A, et al. (2001). $^{99m}$Tc-labeled divalent and tetravalent CC49 single-chain Fv's: Novel imaging agents for rapid in vivo localization of human colon carcinoma. *J Nucl Med 42*, 1519–1527.
24. Ugur O, et al. (1996). Comparison of the targeting characteristics of various radioimmunoconjugates for radioimmunotherapy of neuroblastoma: Dosimetry calculations incorporating cross-organ beta doses. *Nucl Med Biol 23*, 1–8.

25. Nilsson B, et al. (1987). A synthetic IgG-binding domain based on staphylococcal protein A. *Protein Eng 1*, 107–113.
26. Nord K, Nilsson J, Nilsson B, Uhlen M, Nygren P-A (1995). A combinatorial library of an {alpha}-helical bacterial receptor domain. *Protein Eng 8*, 601–608.
27. Nord K, Gunneriusson E, Ringdahl J, Ståhl S, Uhlén M, Nygren PA (1997). Binding proteins selected from combinatorial libraries of an alpha-helical bacterial receptor domain. *Nat Biotechnol 15*, 772–777.
28. Wikman M, et al. (2004). Selection and characterization of HER2/neu-binding affibody ligands. *Protein Eng Design Select 17*, 455–462.
29. Tran T, Orlova A, Sivaev I, Sandstrom M, Tolmachev V (2007). Comparison of benzoate- and dodecaborate-based linkers for attachment of radioiodine to HER2-targeting Affibody ligand. *Int J Mol Med 19*, 485–493.
30. Orlova A, et al. (2006). Comparative in vivo evaluation of technetium and iodine labels on an anti-HER2 affibody for single-photon imaging of HER2 expression in tumors. *J Nucl Med 47*, 512–519.
31. Engfeldt T, Renberg B, Brumer H, Nygren PA, Karlström AE (2005). Chemical synthesis of triple-labelled three-helix bundle binding proteins for specific fluorescent detection of unlabelled protein. *Chembiochem 6*, 1043–1050.
32. Qiu X-Q, Wang H, Cai B, Wang L-L, Yue S-T (2007). Small antibody mimetics comprising two complementarity-determining regions and a framework region for tumor targeting. *Nat Biotechnol 25*, 921–929.
33. DeNardo GL, Kennel SJ, Siegel JA, Denardo SJ (2004). Radiometals as payloads for radioimmunotherapy for lymphoma. *Clin Lymphoma 5*(Suppl 1), S5–10.
34. Chowdhury PS, Wu H (2005). Tailor-made antibody therapeutics. *Methods 36*, 11–24.
35. Colcher D, Minelli MF, Roselli M, Muraro R, Simpson-Milenic D, Schlom J (1988). Radioimmunolocalization of human carcinoma xenografts with B72.3 second generation monoclonal antibodies. *Cancer Res 48*, 4597–4603.
36. Midelfort KS, Hernandez HH, Lippow SM, Tidor B, Drennan CL, Wittrup KD (2004). Substantial energetic improvement with minimal structural perturbation in a high affinity mutant antibody. *J Mol Biol 343*, 685–701.
37. Adams GP, et al. (1998). Increased affinity leads to improved selective tumor delivery of single-chain Fv antibodies. *Cancer Res 58*, 485–490.
38. Fujimori K, Covell DG, Fletcher JE, Weinstein JN (1990). A modeling analysis of monoclonal antibody percolation through tumors: A binding-site barrier. *J Nucl Med 31*, 1191–1198.
39. Behr TM, et al. (1997). Variables Influencing tumor dosimetry in radioimmunotherapy of CEA-expressing cancers with anti-CEA and antimucin monoclonal antibodies. *J Nucl Med 38*, 409–418.
40. Gallinger S, et al. (1993). Comparative dual label study of first and second generation antitumor-associated glycoprotein-72 monoclonal antibodies in colorectal cancer patients. *Cancer Res 53*, 271–278.
41. Carter PJ (2006). Potent antibody therapeutics by design. *Nat Rev Immunol 6*, 343–357.
42. Adams GP, et al. (2001). High affinity restricts the localization and tumor penetration of single-chain fv antibody molecules. *Cancer Res 61*, 4750–4755.
43. Nielsen JT, Poulsen J, Flo C, Marqversen J, Rehling M (1995). Pharmacokinetics and dosimetry of [111In] F(ab′)2 fragments against prostatic acid phosphatase after intraprostatic injection for immunoscintigraphy in prostate cancer. *Clin Physiol 15*, 467–481.
44. Bryan JN, et al. (2005). Comparative uptakes and biodistributions of internalizing vs. noninternalizing copper-64 radioimmunoconjugates in cell and animal models of colon cancer. *Nucl Med Biol 32*, 851–858.
45. Zalutsky MR, Garg PK, Narula AS (1990). Labeling monoclonal antibodies with halogen nuclides. *Acta Radiol 374*, 141–145.
46. Hnatowich DJ (1990). Recent developments in the radiolabeling of antibodies with iodine, indium, and technetium. *Semin Nucl Med 20*, 80–91.
47. Vaidyanathan G, Affleck DJ, Li J, Welsh P, Zalutsky MR (2001). A polar substituent-containing acylation agent for the radioiodination of internalizing monoclonal antibodies: *N*-Succinimidyl 4-guanidinomethyl-3-[131I]iodobenzoate ([131I]SGMIB). *Bioconjugate Chem 12*, 428–438.
48. Nestor M, et al. (2003). Biodistribution of the chimeric monoclonal antibody U36 radioiodinated with a closo-dodecaborate-containing linker. Comparison with other radioiodination methods. *Bioconjugate Chem 14*, 805–810.
49. Meares CF, McCall MJ, Reardan DT, Goodwin DA, Diamanti CI, McTigue M (1984). Conjugation of antibodies with bifunctional chelating agents: Isothiocyanate and bromoacetamide reagents, methods of analysis, and subsequent addition of metal ions. *Anal Biochem 142*, 68–78.
50. Mirick GR, O'Donnell RT, DeNardo SJ, Shen S, Meares CF, DeNardo GL (1999). Transfer of copper from a chelated 67Cu-antibody conjugate to ceruloplasmin in lymphoma patients. *Nucl Med Biol 26*, 841–845.
51. Sun X, et al. (2002). Radiolabeling and in vivo behavior of copper-64-labeled cross-bridged cyclam ligands. *J Med Chem 45*, 469–477.
52. Chen X, et al. (2004). Pegylated Arg-Gly-Asp peptide: 64Cu labeling and PET imaging of brain tumor alphavbeta3-integrin expression. *J Nucl Med 45*, 1776–1783.
53. Cai W, Chen K, He L, Cao Q, Koong A, Chen X (2007). Quantitative PET of EGFR expression in xenograft-bearing mice using (64)Cu-labeled cetuximab, a chimeric anti-EGFR monoclonal antibody. *Eur J Nucl Med Mol Imag 34*, 850–858.

54. Cai W, et al. (2007). Quantitative radioimmunoPET imaging of EphA2 in tumor-bearing mice. *Eur J Nucl Med Mol Imag 3*, 3.
55. Liu Z, et al. (2007). In vivo biodistribution and highly efficient tumour targeting of carbon nanotubes in mice. *Nat Nano 2*, 47–52.
56. Novak-Hofer I, Zimmermann K, Maecke HR, Amstutz HP, Carrel F, Schubiger PA (1997). Tumor uptake and metabolism of copper-67-labeled monoclonal antibody chCE7 in nude mice bearing neuroblastoma xenografts. *J Nucl Med 38*, 536–544.
57. Kukis DL, et al. (1995). Effect of the extent of chelate substitution on the immunoreactivity and biodistribution of 2IT-BAT-Lym-1 immunoconjugates. *Cancer Res 55*, 878–884.
58. Ali MS, Quadri SM (1996). Maleimido derivatives of diethylenetriaminepentaacetic acid and triethylenetetraaminehexaacetic acid: Their synthesis and potential for specific conjugation with biomolecules. *Bioconjugate Chem 7*, 576–583.
59. Stimmel JB, Stockstill ME, Kull FC (1995). Yttrium-90 chelation properties of tetraazatetraacetic acid macrocycles, diethylenetriaminepentaacetic acid analogues, and a novel terpyridine acyclic chelator. *Bioconjugate Chem 6*, 219–225.
60. Harrison A, et al. (1991). The in vivo release of 90Y from cyclic and acyclic ligand-antibody conjugates. *Int J Radiat Appl Instrument B Nucl Med Biol 18*, 469–476.
61. Richman CM, et al. (1999). Dosimetry-based therapy in metastatic breast cancer patients using 90Y monoclonal antibody 170H.82 with autologous stem cell support and cyclosporin A. *Clin Cancer Res 5*, 3243s–3248s.
62. Scott AM, et al. (1994). Clinical applications of fusion imaging in oncology. *Nucl Med Biol 21*, 775–784.
63. van Gog FB, Visser GW, Klok R, van der Schors R, Snow GB, van Dongen GA (1996). Monoclonal antibodies labeled with rhenium-186 using the MAG3 chelate: Relationship between the number of chelated groups and biodistribution characteristics. *J Nucl Med 37*, 352–362.
64. John E, Thakur ML, Wilder S, Alauddin MM, Epstein AL (1994). Technetium-99m-labeled monoclonal antibodies: Influence of technetium-99m binding sites. *J Nucl Med 35*, 876–881.
65. Slinkin MA, et al. (1993). Biodistribution of anti-CEA F(ab′)2 fragments conjugated with chelating polymers: Influence of conjugate electron charge on tumor uptake and blood clearance. *Nucl Med Biol 20*, 443–452.
66. Wiseman GA, et al. (2003). Radiation dosimetry results and safety correlations from 90Y-ibritumomab tiuxetan radioimmunotherapy for relapsed or refractory non-Hodgkin's lymphoma: Combined data from 4 clinical trials. *J Nucl Med 44*, 465–474.
67. Garmestani K, Milenic DE, Plascjak PS, Brechbiel MW (2002). A new and convenient method for purification of $^{86}$Y using a Sr(II) selective resin and comparison of biodistribution of $^{86}$Y and $^{111}$In labeled herceptin. *Nucl Med Biol 29*, 599–606.
68. Lovqvist A, et al. (2001). PET imaging of (86)Y-labeled anti-Lewis Y monoclonal antibodies in a nude mouse model: Comparison between (86)Y and (111)In radiolabels. *J Nucl Med 42*, 1281–1287.
69. Palm S, et al. (2003). Pharmacokinetics and biodistribution of $^{86}$Y-Trastuzumab for $^{90}$Y dosimetry in an ovarian carcinoma model: Correlative MicroPET and MRI. *J Nucl Med 44*, 1148–1155.
70. Goodwin DA, Meares CF (2001). Advances in pretargeting biotechnology. *Biotechnol Adv 19*, 435–450.
71. Goldenberg DM, Sharkey RM, Paganelli G, Barbet J, Chatal J-F (2006). Antibody pretargeting advances cancer radioimmunodetection and radioimmunotherapy. *J Clin Oncol 24*, 823–834.
72. Boerman OC, van Schaijk FG, Oyen WJG, Corstens FHM (2003). Pretargeted radioimmunotherapy of cancer: Progress step by step. *J Nucl Med 44*, 400–411.
73. Hnatowich DJ, Virzi F, Rusckowski M (1987). Investigations of avidin and biotin for imaging applications. *J Nucl Med 28*, 1294–1302.
74. Axworthy DB, et al. (2000). Cure of human carcinoma xenografts by a single dose of pretargeted yttrium-90 with negligible toxicity. *PNAS 97*, 1802–1807.
75. Chauhan J, Dakshinamurti K (1986). Purification and characterization of human serum biotinidase. *J Biol Chem 261*, 4268–4275.
76. Paganelli G, et al. (1991). Three-step monoclonal antibody tumor targeting in carcinoembryonic antigen-positive patients. *Cancer Res 51*, 5960–5966.
77. Reardan DT, et al. (1985). Antibodies against metal chelates. *Nature 316*, 265–268.
78. Boerman OC, et al. (1999). Pretargeting of renal cell carcinoma: Improved tumor targeting with a bivalent chelate. *Cancer Res 59*, 4400–4405.
79. Karacay H, et al. (2002). Pretargeting for cancer radioimmunotherapy with bispecific antibodies: Role of the bispecific antibody's valency for the tumor target antigen. *Bioconjugate Chem 13*, 1054–1070.
80. Sharkey RM, Goldenberg DM (2005). Perspectives on cancer therapy with radiolabeled monoclonal antibodies. *J Nucl Med 46*(Suppl 1), 115S–127S.
81. Karacay H, et al. (2005). Therapeutic advantage of pretargeted radioimmunotherapy using a recombinant bispecific antibody in a human colon cancer xenograft. *Clin Cancer Res 11*, 7879–7885.
82. Goldenberg DM, Sharkey RM (2007). Novel radiolabeled antibody conjugates. *Oncogene 26*, 3734–3744.
83. Carrasquillo JA, et al. (1986). Radioimmunodetection of cutaneous T-cell lymphoma with 111In-labeled T101 monoclonal antibody. *N Engl J Med 315*, 673–680.
84. Keenan AM, et al. (1987). Immunolymphoscintigraphy and the dose dependence of 111In-labeled T101 monoclonal antibody in patients with cutaneous T-cell lymphoma. *Cancer Res 47*, 6093–6099.

85. Samoszuk MK, et al. (1993). Radioimmunodetection of Hodgkin's disease and non-Hodgkin's lymphomas with monoclonal antibody to eosinophil peroxidase. *J Nucl Med 34*, 1246–1253.
86. Samoszuk M, et al. (1992). Phase I imaging study of Hodgkin's disease and non-Hodgkin's lymphomas with radiolabeled antibody to eosinophil peroxidase. *Cell Biophys 21*, 53–60.
87. Baum RP, et al. (1994). Initial clinical results with technetium-99m-labeled LL2 monoclonal antibody fragment in the radioimmunodetection of B-cell lymphomas. *Cancer 73*, 896–899.
88. Vervoordeldonk SF, et al. (1996). 99mTc-CD19 monoclonal antibody is not useful for imaging of B cell non-Hodgkin's lymphoma. *Cancer Immunol Immunother 42*, 291–296.
89. Almasri NM, Duque RE, Iturraspe J, Everett E, Braylan RC (1992). Reduced expression of CD20 antigen as a characteristic marker for chronic lymphocytic leukemia. *Am J Hematol 40*, 259–263.
90. McLaughlin P, et al. (1998). Rituximab chimeric anti-CD20 monoclonal antibody therapy for relapsed indolent lymphoma: Half of patients respond to a four-dose treatment program. *J Clin Oncol 16*, 2825–2833.
91. Colombat P, et al. (2001). Rituximab (anti-CD20 monoclonal antibody) as single first-line therapy for patients with follicular lymphoma with a low tumor burden: Clinical and molecular evaluation. *Blood 97*, 101–106.
92. Marcus R, et al. (2005). CVP chemotherapy plus rituximab compared with CVP as first-line treatment for advanced follicular lymphoma. *Blood 105*, 1417–1423.
93. Wahl RL (2003). The clinical importance of dosimetry in radioimmunotherapy with tositumomab and iodine I 131 tositumomab. *Semin Oncol 30*, 31–38.
94. Cheson BD, et al. (1999). Report of an international workshop to standardize response criteria for non-Hodgkin's lymphomas. NCI Sponsored International Working Group. *J Clin Oncol 17*, 1244–1262.
95. Spar IL, Bale WF, Marrack D, Dewey WC, McCardle RJ, Harper PV (1967). 131-I-labeled antibodies to human fibrinogen. Diagnostic studies and therapeutic trials. *Cancer 20*, 865–870.
96. Harrington KJ, Epenetos AA (1994). Recent developments in radioimmunotherapy. *Clin Oncol (R Coll Radiol) 6*, 391–398.
97. Haisma H, et al. (1984). Specific localization of In-111-labeled monoclonal antibody versus 67-Ga-labeled immunoglobulin in mice bearing human breast carcinoma xenografts. *Cancer Immunol Immunother 17*, 62–65.
98. Kalofonos HP, et al. (1989). Kinetics, quantitative analysis and radioimmunolocalization using indium-111-HMFG1 monoclonal antibody in patients with breast cancer. *Br J Cancer 59*, 939–942.
99. Major PP, Dion AS, Williams CJ, Mattes MJ, Wang T, Rosenthall L (1990). Breast tumor radioimmunodetection with a 111In-labeled monoclonal antibody (MA5) against a mucin-like antigen. *Cancer Res 50*, 927s–931s.
100. Tjandra JJ, et al. (1989). The detection of axillary lymph node metastases from breast cancer by radiolabelled monoclonal antibodies: A prospective study. *Br J Cancer 59*, 296–302.
101. Duncker CM, et al. (1990). Radioimmune imaging of bone marrow in patients with suspected bone metastases from primary breast cancer. *J Nucl Med 31*, 1450–1455.
102. DeNardo SJ, et al. (1997). Radioimmunotherapy for breast cancer using indium-111/yttrium-90 BrE-3: Results of a phase I clinical trial. *J Nucl Med 38*, 1180–1185.
103. Traish AM, Wotiz HH (1987). Prostatic epidermal growth factor receptors and their regulation by androgens. *Endocrinology 121*, 1461–1467.
104. Saga T, et al. (1991). Scintigraphic detection of overexpressed c-erbB-2 protooncogene products by a class-switched murine anti-c-erbB-2 protein monoclonal antibody. *Cancer Res 51*, 990–994.
105. Styles JM, Harrison S, Gusterson BA, Dean CJ (1990). Rat monoclonal antibodies to the external domain of the product of the C-erbB-2 proto-oncogene. *Int J Cancer 45*, 320–324.
106. Dean CJ, et al. (1993). Rat MAbs to the product of the c-erbB-2 proto-oncogene for diagnosis and therapy in breast cancer. *Cell Biophys 22*, 111–127.
107. Allan SM, Dean CJ, Eccles S, Sacks NP (1994). Clinical radioimmunolocalization with a rat monoclonal antibody directed against c-erbB-2. *Cell Biophys 24–25*, 93–98.
108. Meenakshi A, Kumar RS, Ganesh V, Kumar NS (2002). Preliminary study on radioimmunodiagnosis of experimental tumor models using technetium-99m-labeled anti-C-erbB-2 monoclonal antibody. *Tumori 88*, 507–512.
109. Bakir MA, et al. (1992). c-erbB2 protein overexpression in breast cancer as a target for PET using iodine-124-labeled monoclonal antibodies. *J Nucl Med 33*, 2154–2160.
110. Shepard HM, et al. (1991). Monoclonal antibody therapy of human cancer: Taking the HER2 protooncogene to the clinic. *J Clin Immunol 11*, 117–127.
111. Perik PJ, et al. (2006). Indium-111-labeled trastuzumab scintigraphy in patients with human epidermal growth factor receptor 2-positive metastatic breast cancer. *J Clin Oncol 24*, 2276–2282.
112. Quadri SM, Malik AB, Tang XZ, Patenia R, Freedman RS, Vriesendorp HM (1995). Preclinical analysis of intraperitoneal administration of 111In-labeled human tumor reactive monoclonal IgM AC6C3-2B12. *Cancer Res 55*, 5736s–5742s.
113. Dennis MS, et al. (2007). Imaging tumors with an albumin-binding Fab, a novel tumor-targeting agent. *Cancer Res 67*, 254–261.
114. Nguyen A, et al. (2006). The pharmacokinetics of an albumin-binding Fab (AB.Fab) can be modulated as a

function of affinity for albumin. *Protein Eng Des Select 19*, 291–297.

115. Vallabhajosula S, Smith-Jones PM, Navarro V, Goldsmith SJ, Bander NH (2004). Radioimmunotherapy of prostate cancer in human xenografts using monoclonal antibodies specific to prostate specific membrane antigen (PSMA): Studies in nude mice. *Prostate 58*, 145–155.
116. Smith-Jones PM, Solit D, Afroze F, Rosen N, Larson SM (2006). Early tumor response to Hsp90 therapy using HER2 PET: Comparison with $^{18}$F-FDG PET. *J Nucl Med 47*, 793–796.
117. George AJ, et al. (1995). Radiometal labeling of recombinant proteins by a genetically engineered minimal chelation site: Technetium-99m coordination by single-chain Fv antibody fusion proteins through a C-terminal cysteinyl peptide. *Proc Nat Acad Sci USA 92*, 8358–8362.
118. Wilbur DS, et al. (1989). Development of a stable radioiodinating reagent to label monoclonal antibodies for radiotherapy of cancer. *J Nucl Med 30*, 216–226.
119. Adams GP, et al. (1995). Enhanced tumor specificity of 741F8-1 (sFv′)2, an anti-c-erbB-2 single-chain Fv dimer, mediated by stable radioiodine conjugation. *J Nucl Med 36*, 2276–2281.
120. Yuan QA, Simmons HH, Robinson MK, Russeva M, Marasco WA, Adams GP (2006). Development of engineered antibodies specific for the Mullerian inhibiting substance type II receptor: A promising candidate for targeted therapy of ovarian cancer. *Mol Cancer Ther 5*, 2096–2105.
121. Senekowitsch R, Reidel G, Mollenstadt S, Kriegel H, Pabst HW (1989). Curative radioimmunotherapy of human mammary carcinoma xenografts with iodine-131-labeled monoclonal antibodies. *J Nucl Med 30*, 531–537.
122. Smellie WJ, et al. (1995). Radioimmunotherapy of breast cancer xenografts with monoclonal antibody ICR12 against c-erbB2 p185: Comparison of iodogen and *N*-succinimidyl 4-methyl-3-(tri-*n*-butylstannyl)benzoate radioiodination methods. *Cancer Res 55*, 5842s–5846s.
123. Kotts CE, et al. (1996). $^{186}$Re-labeled antibodies to p185HER2 as HER2-targeted radioimmunopharmaceutical agents: Comparison of physical and biological characteristics with $^{125}$I and $^{131}$I-labeled counterparts. *Cancer Biother Radiopharm 11*, 133–144.
124. Wiercioch R, Balcerczak E, Byszewska E, Mirowski M (2003). Uptake of radiolabelled herceptin by experimental mammary adenocarcinoma. *Nucl Med Rev Cent East Eur 6*, 99–103.
125. Howell LP, DeNardo SJ, Levy NB, Lund J, DeNardo GL (1995). Immunohistochemical staining of metastatic ductal carcinomas of the breast by monoclonal antibodies used in imaging and therapy: A comparative study. *Int J Biol Markers 10*, 129–135.
126. Marken JS, Schieven GL, Hellström I, Hellström KE, Aruffo A (1992). Cloning and expression of the tumor-associated antigen L6. *Proc Nat Acad Sci USA 89*, 3503–3507.
127. DeNardo SJ, et al. (1994). The biologic window for chimeric L6 radioimmunotherapy. *Cancer 73*, 1023–1032.
128. Denardo SJ, et al. (1997). Radioimmunotherapy for advanced breast cancer using I-131-ChL6 antibody. *Anticancer Res 17*, 1745–1751.
129. Esteban JM, Felder B, Ahn C, Simpson JF, Battifora H, Shively JE (1994). Prognostic relevance of carcinoembryonic antigen and estrogen receptor status in breast cancer patients. *Cancer 74*, 1575–1583.
130. Wong JY, et al. (1999). Initial clinical experience evaluating yttrium-90-chimeric T84.66 anticarcinoembryonic antigen antibody and autologous hematopoietic stem cell support in patients with carcinoembryonic antigen-producing metastatic breast cancer. *Clin Cancer Res 5*, 3224s–3231s.
131. Kramer EL, et al. (1994). Radioimmunolocalization of breast cancer using BrE-3 monoclonal antibody. *Adv Exp Med Biol 353*, 181–192.
132. Schrier DM, et al. (1995). High-dose 90Y Mx-diethylenetriaminepentaacetic acid (DTPA)-BrE-3 and autologous hematopoietic stem cell support (AHSCS) for the treatment of advanced breast cancer: A phase I trial. *Cancer Res 55*, 5921s–5924s.
133. Schott ME, et al. (1994). Biodistribution and preclinical radioimmunotherapy studies using radiolanthanide-labeled immunoconjugates. *Cancer 73*, 993–998.
134. Kelly MP, Lee FT, Smyth FE, Brechbiel MW, Scott AM (2006). Enhanced efficacy of 90Y-radiolabeled anti-Lewis Y humanized monoclonal antibody hu3S193 and paclitaxel combined-modality radioimmunotherapy in a breast cancer model. *J Nucl Med 47*, 716–725.
135. Masters GR, Berger MA, Albone EF (2006). Synergistic effects of combined therapy using paclitaxel and [90Y-DOTA]776.1 on growth of OVCAR-3 ovarian carcinoma xenografts. *Gynecol Oncol 102*, 462–467.
136. Chauhan SC, et al. (2007). Combined staining of TAG-72, MUC1, and CA125 improves labeling sensitivity in ovarian cancer: Antigens for multi-targeted antibody-guided therapy. *J Histochem Cytochem 55*, 867–875.
137. Oei AL, et al. (2007). Decreased intraperitoneal disease recurrence in epithelial ovarian cancer patients receiving intraperitoneal consolidation treatment with yttrium-90-labeled murine HMFG1 without improvement in overall survival. *Int J Cancer 120*, 2710–2714.
138. Goldenberg DM, et al. (1983). Radioimmunodetection of prostatic cancer. In vivo use of radioactive antibodies against prostatic acid phosphatase for diagnosis and detection of prostatic cancer by nuclear imaging. *JAMA 250*, 630–635.
139. Ostrowski WS, Kuciel R (1994). Human prostatic acid phosphatase: Selected properties and practical applications. *Clin Chim Acta 226*, 121–129.
140. Halpern SE, et al. (1988). Scintigraphy with In-111-labeled monoclonal antitumor antibodies: Kinetics, biodistribution, and tumor detection. *Radiology 168*, 529–536.

141. Larson A, Arnberg H, Maripuu E, Nilsson S (1988). Radioimmunodetection of prostatic cancer with 125I-labelled antibody against prostatic specific antigen. *Scand J Urol Nephrol Suppl 110*, 149–153.
142. Lopes AD, Davis WL, Rosenstraus MJ, Uveges AJ, Gilman SC (1990). Immunohistochemical and pharmacokinetic characterization of the site-specific immunoconjugate CYT-356 derived from antiprostate monoclonal antibody 7E11-C5. *Cancer Res 50*, 6423–6429.
143. Wynant GE, et al. (1991). Immunoscintigraphy of prostatic cancer: Preliminary results with 111In-labeled monoclonal antibody 7E11-C5.3 (CYT-356). *Prostate 18*, 229–241.
144. Israeli RS, Powell CT, Corr JG, Fair WR, Heston WD (1994). Expression of the prostate-specific membrane antigen. *Cancer Res 54*, 1807–1811.
145. Kahn D, et al. (1994). Radioimmunoscintigraphy with 111indium labeled CYT-356 for the detection of occult prostate cancer recurrence. *J Urol 152*, 1490–1495.
146. Babaian RJ, Sayer J, Podoloff DA, Steelhammer LC, Bhadkamkar VA, Gulfo JV (1994). Radioimmunoscintigraphy of pelvic lymph nodes with 111indium-labeled monoclonal antibody CYT-356. *J Urol 152*, 1952–1955.
147. Fang DX, Stock RG, Stone NN, Krynyckyi BR, Kim CK, Machac J (2000). Use of radioimmunoscintigraphy with indium-111-labeled CYT-356 (ProstaScint) scan for evaluation of patients for salvage brachytherapy. *Tech Urol 6*, 146–150.
148. Wilkinson S, Chodak G (2004). The role of 111indium-capromab pendetide imaging for assessing biochemical failure after radical prostatectomy. *J Urol 172*, 133–136.
149. Barren RJ, 3rd, Holmes EH, Boynton AL, Misrock SL, Murphy GP (1997). Monoclonal antibody 7E11.C5 staining of viable LNCaP cells. *Prostate 30*, 65–68.
150. Meredith RF, et al. (1994). Treatment of metastatic prostate carcinoma with radiolabeled antibody CC49. *J Nucl Med 35*, 1017–1022.
151. Kahn D, Austin JC, Maguire RT, Miller SJ, Gerstbrein J, Williams RD (1999). A phase II study of [90Y] yttrium-capromab pendetide in the treatment of men with prostate cancer recurrence following radical prostatectomy. *Cancer Biother Radiopharm 14*, 99–111.
152. Gong MC, Chang SS, Sadelain M, Bander NH, Heston WD (1999). Prostate-specific membrane antigen (PSMA)-specific monoclonal antibodies in the treatment of prostate and other cancers. *Cancer Metastasis Rev 18*, 483–490.
153. Milowsky MI, Nanus DM, Kostakoglu L, Vallabhajosula S, Goldsmith SJ, Bander NH (2004). Phase I trial of yttrium-90-labeled anti-prostate-specific membrane antigen monoclonal antibody J591 for androgen-independent prostate cancer. *J Clin Oncol 22*, 2522–2531.
154. McDevitt MR, et al. (2000). An alpha-particle emitting antibody ([213Bi]J591) for radioimmunotherapy of prostate cancer. *Cancer Res 60*, 6095–6100.
155. DeNardo SJ, et al. (2005). Enhancement of the therapeutic index: From nonmyeloablative and myeloablative toward pretargeted radioimmunotherapy for metastatic prostate cancer. *Clin Cancer Res 11*, 7187s–7194s.
156. O'Donnell RT, et al. (2001). Radioimmunotherapy with (111)In/(90)Y-2IT-BAD-m170 for metastatic prostate cancer. *Clin Cancer Res 7*, 1561–1568.
157. Mach JP, Carrel S, Merenda C, Sordat B, Cerottini JC (1974). In vivo localisation of radiolabelled antibodies to carcinoembryonic antigen in human colon carcinoma grafted into nude mice. *Nature 248*, 704–706.
158. Mach JP, et al. (1983). Tumor localization in patients by radiolabeled monoclonal antibodies against colon carcinoma. *Cancer Res 43*, 5593–5600.
159. Henze E, Kubel R, Waitzinger J, Buchler M, Adam WE, Beger HG (1987). Tumor scintigraphy with 131I anti CEA monoclonal antibodies and F(ab′)2 in colorectal cancer. *Eur J Nucl Med 13*, 125–129.
160. Perkins AC, Whalley DR, Ballantyne KC, Pimm MV (1988). Gamma camera emission tomography using radiolabelled antibodies. *Eur J Nucl Med 14*, 45–49.
161. Westera G, et al. (1991). Radioimmuno positron emission tomography with monoclonal antibodies: A new approach to quantifying in vivo tumour concentration and biodistribution for radioimmunotherapy. *Nucl Med Commun 12*, 429–437.
162. Petersen BM, Jr., Bass BL, Bates HR, Chandeysson PL, Harmon JW (1993). Use of the radiolabeled murine monoclonal antibody, 111In-CYT-103, in the management of colon cancer. *Am J Surg 165*, 137–142; discussion 142–143.
163. Welt S, et al. (1994). Antibody targeting in metastatic colon cancer: A phase I study of monoclonal antibody F19 against a cell-surface protein of reactive tumor stromal fibroblasts. *J Clin Oncol 12*, 1193–1203.
164. Sakahara H, et al. (1995). Immunoscintigraphy of colorectal cancer using 111In-labeled monoclonal antibody to mucin. *Cancer Immunol Immunother 41*, 157–161.
165. Gridley DS, Ewart KL, Cao JD, Stickney DR (1991). Hyperthermia enhances localization of 111In-labeled hapten to bifunctional antibody in human colon tumor xenografts. *Cancer Res 51*, 1515–1520.
166. Klivenyi G, et al. (1998). Gallium-68 chelate imaging of human colon carcinoma xenografts pretargeted with bispecific anti-CD44V6/anti-gallium chelate antibodies. *J Nucl Med 39*, 1769–1776.
167. Griffiths GL, et al. (2004). Reagents and methods for PET using bispecific antibody pretargeting and 68Ga-radiolabeled bivalent hapten-peptide-chelate conjugates. *J Nucl Med 45*, 30–39.
168. McBride WJ, et al. (2006). Bispecific antibody pretargeting PET (immunoPET) with an 124I-labeled hapten-peptide. *J Nucl Med 47*, 1678–1688.

169. Meredith RF, et al. (1992). Phase I trial of iodine-131-chimeric B72.3 (human IgG4) in metastatic colorectal cancer. *J Nucl Med 33*, 23–29.
170. Heath JK, et al. (1997). The human A33 antigen is a transmembrane glycoprotein and a novel member of the immunoglobulin superfamily. *Proc Nat Acad Sci USA 94*, 469–474.
171. Welt S, et al. (1994). Phase I/II study of iodine 131-labeled monoclonal antibody A33 in patients with advanced colon cancer. *J Clin Oncol 12*, 1561–1571.
172. Mayer A, et al. (2003). Higher dose and dose-rate in smaller tumors result in improved tumor control. *Cancer Invest 21*, 382–388.
173. Behr TM, et al. (2002). Radioimmunotherapy of small-volume disease of metastatic colorectal cancer. *Cancer 94*, 1373–1381.
174. Buchegger F, et al. (2000). Combined radioimmunotherapy and radiotherapy of liver metastases from colorectal cancer: A feasibility study. *Anticancer Res 20*, 1889–1896.
175. Engfeldt T, et al. (2006). Imaging of HER2-expressing tumours using a synthetic Affibody molecule containing the $^{99m}$Tc-chelating mercaptoacetyl-glycyl-glycyl-glycyl (MAG3) sequence. *Eur J Nucl Med Mol Imaging 33* (suppl 14), 49–62.
176. Backer MV, et al. (2007). Molecular imaging of VEGF receptors in angiogenic vasculature with single-chain VEGF-based probes. *Nat Med 13*, 504–509.
177. Olafsen T, et al. (2004). Covalent disulfide-linked anti-CEA diabody allows site-specific conjugation and radiolabeling for tumor targeting applications. *Protein Eng Des Sel 17*, 21–27.
178. Visvikis D, Turzo A, Bizais Y, Cheze-Le Rest C (2004). Technology related parameters affecting quantification in positron emission tomography imaging. *Nucl Med Commun 25*, 637–641.
179. Niu G, Cai W, Chen K, Chen X (2008). Non-invasive PET imaging of EGFR degradation induced by a heat shock protein 90 inhibitor. *Mol Imaging Biol 10*, 99–106.
180. Strahl BD, Allis CD (2000). The language of covalent histone modifications. *Nature 403*, 41–45.
181. Stoeber K, et al. (2002). Diagnosis of genito-urinary tract cancer by detection of minichromosome maintenance 5 protein in urine sediments. *J Natl Cancer Inst 94*, 1071–1079.
182. Drummond DC, Noble CO, Kirpotin DB, Guo Z, Scott GK, Benz CC (2005). Clinical development of histone deacetylase inhibitors as anticancer agents. *Annu Rev Pharmacol Toxicol 45*, 495–528.
183. Drummond DC, et al. (2005). Enhanced pharmacodynamic and antitumor properties of a histone deacetylase inhibitor encapsulated in liposomes or ErbB2-targeted immunoliposomes. *Clin Cancer Res 11*, 3392–3401.

# THERAPEUTIC DRUG DEVELOPMENT FOR KIDNEY DISEASES

Mitchell H. Rosner, Joshua King, and Peter Monteleone
*Department of Medicine, University of Virginia Health System, Charlottesville, Virginia*

## 1 INTRODUCTION

Diseases that affect the kidney can be thought of in terms of the functional compartment that is affected by the pathogenic cause. These compartments include the renal vasculature, glomeruli, renal interstitium, renal tubules, and urinary collecting system. In each of these compartments, processes can be acute or chronic in nature and can lead to severe disease that would require some form of renal replacement therapy (hemodialysis, peritoneal dialysis, or kidney transplantation). Current therapeutic agents for kidney disease are limited and for the most part are not specifically targeted to pathogenic pathways. Given the epidemic of chronic kidney disease (CKD) (estimates are that over 8 million people in the United States have an estimated glomerular filtration rate less than 59 mL/min) [1], there is an urgent need for therapeutic agents that target this problem. Furthermore, in those patients with CKD, management of the complications associated with this disease state such as anemia and metabolic bone disease need to be effectively managed to avoid serious complications. Finally, in those patients who advance to end-stage renal disease, therapies to improve the outcomes of renal transplantation are needed.

This chapter will explore drug development across the spectrum of kidney disease. This ranges from acute kidney injury (AKI) to specific causes of chronic kidney disease such as polycystic kidney disease. Drugs to effectively manage the complications of CKD are also discussed. Finally, in each section, common obstacles to drug development for this disease state are discussed.

## 2 DRUG DEVELOPMENT FOR ACUTE KIDNEY INJURY

Acute kidney injury is due to a variety of conditions and has serious consequences. There are numerous definitions of AKI, but in general this term refers to the acute deterioration in renal function as manifested by rises in serum creatinine and fall in urine output. Clinically, this leads to impairment in the excretion of waste products and fluid and in the extreme the need for dialysis. Using large databases of U.S. hospitalizations over the past 10–15 years, there is evidence for a marked increase in the incidence of AKI [2, 3]. This, in part, reflects the increasing comorbidity and age of patients suffering AKI. It is widely recognized that AKI leads to high morbidity and mortality in hospitalized patients. While there may be evidence that morality rates are decreasing, [2, 3] they remain unacceptably high, and there is an urgent need for effective therapy [4].

Except for a few isolated studies, the vast majority of animal and clinical studies have yet to conclusively demonstrate the benefit of pharmacological treatment for AKI. Table 1 summarizes potential factors that have contributed to the lack of success of human studies. Table 2 describes novel pharmacological therapies on the horizon for the treatment.

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

**TABLE 1 Barriers to Development of Therapeutic Agents for AKI**

| |
|---|
| Patient and comorbid factors |
| Complexity of AKI—multiple interacting pathways lead to injury |
| AKI is a multisystem disease |
| Design issues for clinical trials |
| Need for better diagnostic biomarkers to detect early injury |
| Definition of AKI—need for standardization |
| Determination of proper endpoints in clinical trials |

**TABLE 2 Emerging Pharmacological Agents for Treatment of AKI**

| |
|---|
| **Antiapoptosis/necrosis agents** |
| Caspase inhibitors |
| Nonselective caspase inhibitors |
| Selective caspase 3 and 7 inhibitors |
| Selective caspase 1 inhibitors |
| Minocycline |
| Guanosine |
| Pifithrin-α (p53 inhibitor) |
| Poly ADP-ribose polymerase (PARP) inhibitor |
| **Free-radical scavengers** |
| Deferoxamine |
| **Antisepsis** |
| Ethyl pyruvate |
| Activated protein C |
| Insulin |
| **Growth factors** |
| Recombinant erythropoietin |
| Hepatocyte growth factor (HGF) |
| **Vasodilators** |
| Carbon monoxide release compounds |
| Bilirubin |
| Endothelin antagonist |
| **Anti-inflammatory drugs** |
| Sphingosine 1 phosphate analogs |
| Adenosine 2A agonists |
| Adenosine analogs |
| Inducible nitic oxide synthase (iNOS) inhibitors |
| Fibrates |
| **Other agents** |
| Neutrophil gelatinase-associated lipocalin |
| IL-6 and C5a antagonists |
| IL-10 |
| α-Melanocyte stimulating hormone |

### 2.1 Barriers to Successful Clinical Trials and Drug Development in Acute Kidney Injury

***Patients and Comorbid Factors*** The rising incidence of AKI over the decades is associated with a changing spectrum of illnesses. In particular, there is evidence for a heavy burden of patients with significant comorbid and extrarenal complications [2, 5]. Supporting this, using the Deyo-Charlson comorbidity index, those patients with higher comorbidity were associated with a higher incidence of AKI especially if they were on mechanical ventilation [2]. In a multicenter study of 618 patients with AKI in the intensive care unit [PICARD (Program to Improve Care in Acute Renal Disease) network], the incidence of comorbid conditions was high, including 30% with chronic kidney disease, 37% with coronary artery disease, 29% with diabetes mellitus, and 21% with chronic liver disease [5]. AKI was accompanied by extrarenal organ system failure in most patients. These comorbid conditions are likely contributors to failed treatment regimens, especially if mortality is used as an endpoint for the clinical trial. Furthermore, drug development focused on isolated medical problems may fail due to the complex interplay of multiple comorbidities in a single patient.

***Pathogenesis of AKI Is Complex*** The pathogenesis of AKI is complex. Ischemia or toxins are major factors that precipitate injury, and, although the initiating events may be dissimilar, subsequent injury responses likely involve similar pathways. As an example, AKI associated with ischemia is due to a reduction of renal blood flow below the limits of blood flow autoregulation. A variety of molecular responses then occur that are "maladaptive" and stereotypical. These responses lead to endothelial and epithelial cell injury following the onset of reperfusion [6]. Pathogenetic factors, such as vasoconstriction, leukostasis, vascular congestion, apoptosis, and abnormalities, in immune modulators and growth factors have formed the basis of logical therapeutic interventions. However, many of these targeted therapies have failed, are inconclusive, or have yet to be performed [7, 8]. Given the complexity of the pathogenesis of AKI, it may be naïve to expect that one therapeutic intervention would have success unless that intervention focuses on prevention of AKI and targets a specific initiating etiology. This is true in that the greatest successes for therapeutic interventions have come in the prevention of contrast nephropathy: saline [9], sodium bicarbonate [10], low and isosmolar contrast [11, 12], and theophylline [13]. Given the multiple overlapping pathways involved in AKI, therapies may need to simultaneously target multiple pathways in order to achieve success [14]. However, the evolving and deepening understanding of the pathogenesis of AKI allows investigators to specifically target pathogenic pathways (e.g., the Toll-like receptors) that are logical and based upon sound reasoning.

***Acute Kidney Injury (AKI) Is a MultiSystem Disease*** While AKI is an independent risk factor for mortality from cardiopulmonary dysfunction [15], in most studies

of AKI, renal failure per se is usually not the cause of death [14]. In several studies, either congestive heart failure [16, 17] or noncardiogenic acute respiratory distress syndrome (ARDS) [18] is associated with AKI. Thus, the potential systemic effects of AKI involve multiple organs leading to high mortality. Animal studies have confirmed that isolated AKI may lead to distant cardiopulmonary dysfunction [18, 19]. Thus, the complexity created by the systemic effects of isolated AKI may contribute to ineffective treatments in the past. These observations also suggest that potential therapeutic strategies should not be limited to treatment of kidney injury alone but should be broad based to treat systemic effects of AKI.

### 2.2 Design Issues in Clinical Trials

Common design reasons for the inability to demonstrate that a particular therapeutic agent is beneficial include low statistical power, lack of a consensus definition of AKI in previous trials, improper endpoints, difficulty in timing administration of the drug, side effects of the drug, and patient heterogeneity [20]. Furthermore, difficulties in patient recruitment and randomization that controls for illness severity have also proven to be barriers to successful clinical trials. Three clinical trials in AKI serve as examples of these difficulties.

Recombinant human insulin-like growth factor I (rhIGF-I) reduced kidney injury when administered 30 min after reperfusion in experimental ischemia–reperfusion injury [21] In addition, rhIGF-I is known to decrease apoptosis and inflammation in experimental acute renal failure [22, 23]. Thus, it was surprising when, in human trials, rhIGF-1 failed to protect patients from the development of AKI. This may be, in part, due to hypotension associated with rhIGF-1 (42 vs. 27%), but in addition administration could have been delayed by as much as 6 days following diagnosis [24]. Because the therapeutic window for prevention of AKI is likely to be narrow, delayed treatment is likely to be unhelpful.

Similar design issues were seen in the trial of atrial natriuretic peptide (ANP). ANP is known to dilate afferent arterioles, constrict efferent arterioles, induce natriuresis, [25] and reduce ischemia–reperfusion injury in animal models. However, when infused into humans, anaritide (human atrial natriuretic peptide; 200 ng/kg/min) did not reduce 21-day dialysis-free survival [26]. A subgroup with oliguria seemed to fare better but was not confirmed in a subsequent study [27]. A confounding variable was the low blood pressure that occurred in the h-ANP group. In a subsequent study using a lower dose of h-ANP (50 ng/kg/min), significantly less hypotension was observed and 21-day dialysis survival and estimated glomerular filtration rate (GFR) were improved [28].

Lastly, dopamine has been shown to be ineffective in the treatment of human AKI [29–31] despite an increase in natriuresis. In animal studies it has been shown that, although dopamine increases outer medullary blood flow by 35%, it does not increase medullary $pO_2$, an important measure of tissue oxygen delivery during ischemia–reperfusion injury. Furthermore although dopamine reduced renal vascular resistance in patients without AKI, it paradoxically increased resistance in patients with AKI [32]. Thus, "low-dose" dopamine worsens renal perfusion in patients with AKI, which may in part explain the lack of success with this agent.

One of the greatest impediments to trial design is the fact that serum creatinine is a poor biomarker of AKI, yet it is commonly used to time interventions [4]. Many factors regulate the generation, volume of distribution, and excretion of creatinine. Net excretion of creatinine is due both to filtration and, to a variable degree, proximal tubule secretion. Thus, organic compounds and drugs may block creatinine secretion. Most importantly, the rise in serum creatinine is slow following AKI (Fig. 1). By the time a change is observed in serum creatinine, a critical therapeutic window may have been missed. The earliest evidence of injury precedes frank acute tubular necrosis and rises in serum creatinine and is likely the point at which interventions need to be introduced. Thus, the detection of the earliest evidence of AKI will ultimately necessitate the use of novel serum or urinary biomarkers. For example, patients with prerenal AKI could represent a group that demonstrates a graded increase in tubular enzymuria (biomarker) and other biomarkers that suggest the earliest evidence of epithelial cell injury yet insufficient

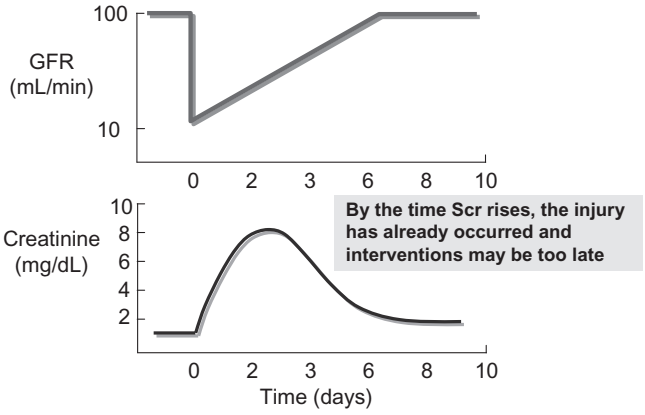


**Figure 1** Biomarkers for the early diagnosis of AKI are needed as increases in serum creatinine occur late in the course of the disease process.

to cause frank necrosis and demonstrable rises in serum creatinine. The use and validation of sensitive biomarkers may permit identification of individuals with early tubular injury and identify a subgroup of patients that might be the target for early intervention. Currently, several urinary biomarkers are being validated [33–37].

The majority of clinically significant AKI occurs in the intensive case unit (ICU). Nephrologists in this setting serve as consultants and typically care for patients only when significant rises in serum creatinine or oliguria develop. As stated above, at this point, significant injury has occurred and therapeutic interventions may be unsuccessful. Thus, to ensure success of any clinical trial, close collaboration between intensivists and nephrologists is required to identify appropriate patients at the earliest stages of injury. Unfortunately, this has not always occurred in the design and implementation of clinical trials.

Defining clinically significant AKI is critical. In the past, clinical trials have used widely varying definitions ranging from a 20–30% rise in serum creatinine to the need for dialysis, which has led to reported incidence of AKI of 1–25% [15, 38] and mortality rates varying from 28 to 90% [39, 40]. The absence of consensus on a definition for AKI led to the Acute Dialysis Quality Initiative (ADQI) and the development of the Acute Kidney Injury Network (AKIN), which represents the efforts of work groups that seek to develop consensus and evidence-based statements in the field of AKI. ADQI used a set of criteria called the RIFLE (Risk, Injury, Failure, Loss, and End stage) criteria [41] that has recently been modified by the AKIN working group. Validation of the classification and staging system of AKI will be required in future clinical studies and holds promise that this classification scheme can improve the design of trials.

Finally, appropriate endpoints for clinical trials in AKI need to be defined [42]. The endpoints can range from the need for dialysis at a specified time point to the development of renal recovery or mortality at a specified endpoint. These specific endpoint goals have important implications for the appropriate size and power of trials as well as the potential for finding a positive outcome. For instance, an intervention that may have a small effect on renal function would not be discovered in a trial that focused on total mortality and was powered for that outcome.

## 2.3 Drugs on the Horizon

Given the failure of multiple pharmaceutical agents in the therapy of AKI, novel agents are needed. A number of drugs and investigational compounds are currently undergoing evaluation for the treatment of AKI (Table 2).

### *Antiapoptosis/Necrosis Agents*

*Caspase Inhibitors* Caspases are a family of proteases involved in the initiation and execution phase of apoptosis. Nonselective and selective caspase inhibitors are effective in attenuating renal injury in ischemia- or endotoxemia-induced AKI when administered before or at the time of injury [23, 43, 44].

*Minocycline* Minocyclines are second-generation tetracycline antibiotics with proven human safety data. Minocycline is known to have antiapoptotic and anti-inflammatory effects. When administered 36 h before renal ischemia, minocycline reduced tubular cell apoptosis and decreased mitochondrial release of cytochrome c and p53 and upregulated bax [6]. Furthermore minocycline reduced kidney inflammation and also microvascular permeability [45, 46]. Currently, minocycline is undergoing testing in phase III clinical trials for renal protection.

*Guanosine and Pifithrin-a (p53 Inhibitor)* Guanosine triphosphate (GTP) salvage by exogenous administration of guanosine reduced renal tubular cell apoptosis, an effect associated with inhibition of p53 expression [47]. Pifithrin-α, a novel p53 inhibitor, also led to decreased tubule cell apoptosis and preserved renal function [48].

*Poly ADP-Ribose Polymerase (PARP) Inhibitor* PARP is a ubiquitous nuclear enzyme that participates in DNA (deoxyribonucleic acid) repair [49, 50]. Paradoxically, excessive activation of PARP from cellular injury leads to intracellular nicotinamide adenine dinucleotide $(NAD)^+$, and adenosine triphosphate (ATP) depletion, ultimately resulting in cell death. PARP overactivation has been known to play a role in the pathogenesis of ischemia–reperfusion injury to kidney, heart, and brain [51–53]. Inhibition of PARP immediately at reperfusion reduced injury. PARP inhibitors are currently in phase II clinical trials of reperfusion injury.

### *Free-Radical Scavengers*

*Deferoxamine* A key early feature of AKI is the generation of reactive oxygen species. The iron chelator deferoxamine is a well-known free-radical scavenger. In several models of AKI, deferoxamine has demonstrated a marked protective effect [54–57]. The protective effect of deferoxamine in different models suggests the central role of free radicals in acute renal failure. Studies in AKI are currently underway to test the efficacy of iron chelation.

### *Antisepsis*

*Ethyl Pyruvate* Pyruvate has been known as a potent endogenous antioxidant and free-radical scavenger, and its derivative, ethyl pyruvate, proved to be effective in reducing mortality in animal models of lethal hemorrhagic shock and systemic inflammation caused by endotoxemia or sepsis [58]. In addition to an effect on mortality, ethyl pyruvate also reduced kidney injury in a cecal ligation puncture (CLP) model of sepsis [59]. Ethyl pyruvate is a widely used food additive and has shown to be safe in phase I clinical trials. It is now being tested in a phase II trial in patients undergoing cardiopulmonary bypass (CBP).

*Activated Protein C* Activated protein C (APC) is a physiologic anticoagulant generated by the thrombin–thrombomodulin complex in endothelial cells. However, in addition to its effect on coagulation, APC has been shown to have anti-inflammatory, antiapoptotic effects [60, 61]. APC also attenuated renal ischemia reperfusion injury (IRI) by inhibiting leukocyte activation [62]. APC is currently approved by the Food and Drug Administration (FDA) in treating patients with an severe sepsis with an Acute Physiology and Chronic Health Evaluation (APACHE) score of 25 or higher.

*Insulin* Insulin resistance and hyperglycemia are common in critically ill patients and intensive insulin therapy targeting blood glucose level between 80–110 mg/dL reduced the incidence of ARF requiring dialysis or hemofiltration [63]. The relationship of hyperglycemia and adverse outcome in critically ill patients with AKI was recently also observed in subgroup analysis of Program to Improve Care in Acute Renal Failure (PICARD) study [64]. The mechanism for clinical benefit may relate the direct metabolic and nonmetabolic effects of hyperglycemia. Endothelial dysfunction and subsequent hypercoagulation, and dyslipidemia, commonly observed in critically ill patients can also be partially corrected by insulin independent of its blood glucose lowering effect [65, 66]. However, despite these promising results, the effect of intensive insulin treatment in the setting of AKI has not been tested and needs to be confirmed in appropriate powered randomized clinical trials.

### *Growth Factors*

*Recombinant Erythropoietin* Erythropoietin has been shown to have anti-inflammatory, antiapoptotic effects in ischemic brain damage, spinal cord injury, and retinal damage [67]. Exogenously admininistered erythropoietin before or at the time of reperfusion reduced kidney injury by reducing tubular necrosis and apoptosis [68–70]. It enhanced tubular proliferation in cisplatin-induced ARF [71] and also mediated mobilization and proliferation of endothelial progenitor cells (EPC) from the bone marrow that has been shown to participate tissue repair [72, 73]. Clinical trials of AKI using recombinant erythropoietin are currently in progress.

*Hepatocyte Growth Factor (HGF)* HGF can promote cell growth, motility, and morphogenesis of various types of cells [74, 75]. Renal expression of HGF and its receptor, c-met, increases after IRI, and exogenous administration of HGF reduces renal injury and accelerates renal regeneration in a murine model of AKI [76–78]. The mechanism of protection is thought to involve a decrease in leukocyte–endothelial interaction with reduced inflammation and also a decrease in tubular cell apoptosis [79]. Currently, phase I/II study of recombinant human HGF in fulminant hepatic failure patients and another phase II study of HGF via plasmid vector in critical limb ischemia patients with peripheral ischemic ulcer are underway. Experience in these clinical trials may shed light on human AKI.

### *Vasodilators*

*Carbon Monoxide Release Compounds and Bilirubin* Heme oxygenase (HO) induction plays a central role in limiting the extent of tissue injury [80, 81]. HO activity leads to the production of carbon monoxide (CO) and a potent antioxidant bilirubin, and it is thought that the protective effect of HO activation is through these factors. In renal IRI administration of CO donor compounds tricarbonyldichlororuthenium(II) dimer, ([Ru(CO)(3)Cl(2)](2), or tricarbonylchloro(glycinato) ruthenium(II) ([Ru(CO)(3)Cl(glycinate)] (CORM-3) 1 h before the onset of ischemia significantly decreased the levels of plasma creatinine 24 h after reperfusion as compared with vehicle-treated mice. This suggests that CO itself may be protective and limit renal damage in ischemia-induced ARF [82]. Bilirubin has also been shown to reduce kidney injury from IRI [83], and, when biliverdin and CO are used in combination, they are syngeristic in improving heart allograft survival [84].

*Endothelin Antagonist* A potent vasoconstrictor, endothelin-1 (ET-1) has been implicated to play important roles in animal models of AKI or radiocontrast nephropathy [85, 86]. ET-1 mediates its biological effects by binding to $ET_A$ or $ET_B$ receptors. In rat kidney, $ET_A$ receptor stimulation is known to mediate vasoconstriction, while $ET_B$ receptor activation can also mediate vasodilation by the generation of nitric oxide and prostacyclin [87, 88]. In addition, ET-1 can stimulate the expression of adhesion molecules and the production

cytokines from monocytes and neutrophils, suggesting the possible role of ET-1 in inflammation in AKI [89]. Several studies demonstrated the beneficial effect of selective $ET_A$ or nonselective endothelin receptor antagonist (ERA) in ischemic ARF, but the major limitation of those studies are the fact that ERA was administered before injury. Administration of the drug at a later time point during the reperfusion was ineffective. However, Wilhelm et al. recently showed that tezasartan, a dual ET1 receptor antagonist, attenuated renal injury even when administered after ischemia [90]. Currently, no clinical trial testing ERA in AKI is undergoing.

***Anti-inflammatory Drugs*** Inflammatory cells, including polymorphoneutrophil, monocytes, macrophages, and T cells, have received considerable attention as important contributors to ischemic acute renal failure. Several new compounds appear to be effective in reducing injury for ischemia–reperfusion through direct action on leukocytes.

*Spingosine 1 Phosphate (S1P) Analogs* S1P is a specific ligand for a family of G-protein-coupled endothelial differentiation gene (*Edg*) receptors (also referred to as S1PRs 1–5) that evoke diverse cellular signaling responses. S1PRs regulate different biological processes depending on their pattern of expression and the diverse G proteins present. S1P binds to receptors or acts as a second messenger to stimulate cell survival, inhibit cell apoptosis, and inhibit cell adhesion and movement [91]. An S1P analog, FTY720, acts as an agonist at four sphingosine 1 phosphate receptors, which lead to sequestration of lymphocytes in secondary lymphatic tissue [92]. In studies of ischemia–reperfusion kidney IRI, FTY720 or similar compounds produced lymphopenia and renal tissue protection [93–95].

*$A_{2A}$ Agonists and Other Adenosine Analogs* Adenosine binds to receptors that are members of the G-protein-coupled receptor family, which includes four subtypes: $A_1$-, $A_{2A}$-, $A_{2B}$-, and $A_3$-ARs (adenosine receptor subtypes) [96]. Accumulating data demonstrate that selective activation of $A_{2A}$-ARs reduces parenchymal injury in nonrenal tissue including heart, liver, spinal cord, lung, and brain [97–99]. The selective $A_{2A}$ agonist, ATL146e, is highly protective against ischemia–reperfusion (IR) injury of kidney and reduces injury by 70–80% [100, 101]. Following administration either before or immediately at the onset of reperfusion, ATL146e alone or in combination with a phosphodiesterase inhibitor reduced renal injury [102]. ATL146e is in human clinical studies for cardiac imaging, and current efforts are directed toward human clinical studies in AKI. Additional studies demonstrate that strategies using A1 agonists or A3 blockers may be effective in AKI [103, 104].

*Inducible Nitic Oxide Synthase Inhibitors* The role of nitric oxide and nitric oxide synthases has been extensively studied. Both in vivo and in vitro studies point toward the important role of inducible nitric oxide sythase (iNOS) in mediating injury to proximal tubules [105]. Selective iNOS inhibitors are currently used in human investigation.

FIBRATSE Peroxisome proliferator-activated receptors (PPARs) are transcription factors that regulate glucose, lipid metabolism. Recent studies have indicated PPARs play an important role in inflammation and immunity [106]. Pretreatment of animals with fibtrates (PPARa ligand) ameliorated cisplatin-induced renal dysfunction, and this was accompanied by suppression of nuclear factor (NF)-κB activation, cytokine/chemokine expression, and neutrophil infiltration, suggesting that the protective effect of fibrates are mediated through its anti-inflammatory effect [107].

*Other Compounds* Neutrophil gelatinase-associated lipocalin [108], interlewkin-6 (IL-6) and C5a antagonists [109]}, IL-10 [110], α-melanocyte-stimulating hormone [111] are other potential compounds that have multiple mechanisms of tissue protection and may be beneficial in human AKI.

***Conclusions*** Barriers to successful clinical trials have thwarted progress in developing effective therapeutics for AKI. Implementation of strategies to prevent or treat AKI will depend upon the coordinated efforts of academic institutions, private industry, and the federal government in developing and testing novel therapeutic agents in well-designed clinical trials.

Furthermore, the complexity of AKI is due in part to activation of multiple overlapping and distinct temporal pathways. Inflammation (cellular and humoral) is a key mediator of AKI. It is unlikely that targeting events that occur late in AKI or even a single pathway will be effective. It might be that new strategies to treat or prevent AKI will require compounds that target pathways more proximal to onset. Such strategies could include the use of compounds that affect multiple pathways or combination therapy that targets several areas, rather than one.

## 3 DRUG DEVELOPMENT FOR POLYCYSTIC KIDNEY DISEASE

Autosomal-dominant polycystic kidney disease (ADPKD) is typified by the progressive development

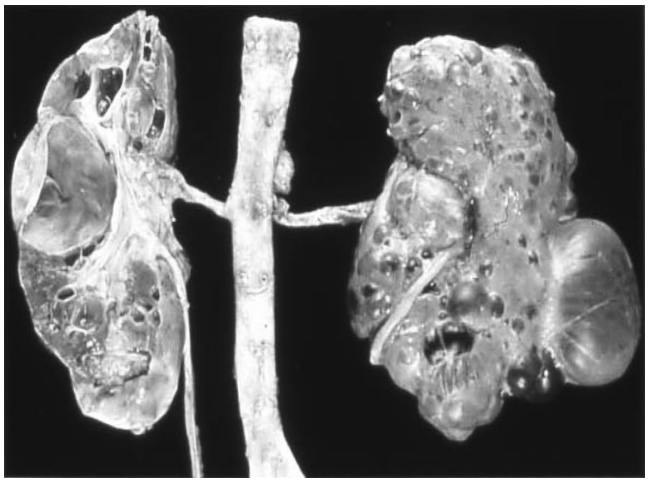

**Figure 2** Gross appearance of kidneys from a patient with autosomal-dominant polycystic kidney disease.

**TABLE 3 Potential Experimental Targets for Drug Development in ADPKD**

| Target System | Potential Drugs |
|---|---|
| Renin–angiotensin system[a] | ACE inhibitors, angiotension receptor blockers |
| Vasopressin[a] | V2-receptor antagonists |
| Endothelin | Bosentan, darusentan |
| mTOR[a] | Sirolimus, temsirolimus, everolimus |
| Caspase system | IDN-8050 (pan-caspase inhibitor) |
| Somatostatin[a] | Octreotide |
| Epidermal growth factor receptor | EKI-785, pelitinib |

[a]In phase III clinical trial development.

of bilateral kidney cysts (Fig. 2) that ultimately lead to deterioration of kidney function and the need for renal replacement therapy (dialysis or transplantation) in 50% of affected individuals by the fifth decade of life. It is the most common life-threatening hereditary disease. ADPKD has an estimated prevalence of between 1 in 400 and 1 in 1000 [112]. In the United States, approximately 2144 patients every year start renal replacement therapy due to ADPKD [113], with an annual incidence rate of end-stage renal disease in men and women of 8.7 and 6.9 per million, respectively [113]. The higher incidence in men suggests a more progressive disease in this population.

ADPKD is a genetically heterogeneous disorder with two genes identified: *PKD1* (chromosome region 16p13.3, accounting for around 85% of cases) and *PKD2* (chromosome region 4q21, around 15% of cases) [114, 115]. Polycystin (PC)-1 encoded by *PKD1* has a large extracellular region, 7–11 transmembrane domains and a C-terminal cytoplasmic tail [114]. Based upon its structure and homology to other proteins, it has been suggested that PC-1 may play a role in cell–cell or cell–matrix interactions [116]. It is believed that PC-1 is a receptor for an unidentified ligand and may function as a mechanoreceptor. PC-1 is localized to the primary cilium, a single hairlike organelle projecting from the surface of most mammalian cells [116]. The protein has also been localized to tight junctions, adherens junctions, desmosomes, focal adhesions, and apical vesicles [117]. The C-terminal tail of PC-1 contains tyrosine and serine sites that can be specifically phosphorylated by c-Src, focal adhesion kinase, protein kinase A, and protein kinase X as well as heterotrimeric G-protein activation [118]. PC-1 may also be involved in intracellular signaling resulting in modulation of gene transcription [118].

Polycystin (PC)-2 is encoded by *PKD2* and contains an N-terminal cytoplasmic region, six transmembrane domains, and a C-terminal cytoplasmic tail [119]. The transmembrane domain has homology to PC-1 and transient receptor potential (TRP) channel subunits [119]. This domain acts as a calcium-permeable cation channel [120]. Studies have demonstrated that PC-1 and PC-2 interact to form a functional complex that acts as a mechanoreceptor on cilia to detect changes in flow [121]. These changes lead to calcium influx through the PC-2 channel, which is further amplified by intracellular calcium release [121]. Furthermore, these rises in intracellular calcium limit cyclic adenosine monophosphate (cAMP) accumulation [122]. Overall, these complexes are essential for maintenance of the differentiated, polarized, predominantly reabsorptive phenotype in tubular epithelial cells, which normally have low rates of proliferation and apoptosis [118]. Mutations in either PC-1 or PC-2 lead to polarization defects, cellular differentiation, proliferation, development of a secretory (rather than resorptive) phenotype, and a high rate of apoptosis that culminate in the development of renal cysts and progressive kidney failure.

The genetic and molecular biological advances in the understanding of the pathogenesis of ADPKD have led to the identification of several key pathways for drug development (Table 3). These strategies are now entering clinical trials to determine possible efficacy in slowing or halting the progression of ADPKD.

### 3.1 Barriers to Successful Clinical Trials in ADPKD

One of the greatest barriers to designing a clinical trial that would test the efficacy of a therapeutic agent in slowing or halting the progression of kidney disease in

ADPKD is that the natural history of progression occurs over many decades. This makes any trial not feasible if the endpoint is change in renal function over time. This same argument holds for many of the progressive renal diseases that slowly lead to deterioration of renal function. To circumvent this critical problem, robust surrogate markers are required that allow for the determination of changes on a shorter time scale but are highly correlated with long-term outcomes. For ADPKD, recent studies have determined that both baseline kidney volume and changes in kidney volume that occur over relatively short time periods (1–5 years) highly correlate with long-term changes in GFR [123]. Thus, higher rates of kidney volume increase are associated with progressive decline in GFR, and this surrogate marker can be used for design and implementation of clinical trials.

### 3.2 Molecular Targets for Therapeutic Drugs in ADPKD

***Vasopressin Receptor Antagonists*** As mentioned above, calcium homeostasis is abnormal in ADPKD. This may lead to increased renal concentrations of cAMP by stimulation of $Ca^{2+}$-inhibitable adenylyl cyclase 6 or inhibition of $Ca^{2+}$-dependent phosphodiesterase 1 [124]. cAMP stimulates mitogen-activated protein kinase (MAPK) and extracellularly regulated kinase (ERK) signaling and cell proliferation in renal epithelial cells of patients with ADPKD, whereas it has an inhibitory effect in wild-type cells [125]. The abnormal proliferative response to cAMP is directly linked to alterations in intracellular $Ca^{2+}$ concentration since it can be reproduced in wild-type cells by lowering the intracellular concentration of $Ca^{2+}$ [126]. Furthermore, the increases in intracellular cAMP may be related to both upregulation of the vasopressin V2 receptor (VPV2R) and high circulating vasopressin concentrations [127].

On the basis of the observed VPV2R activation of adenylyl cyclase and cAMP in cysts from cells from renal cysts, recent studies examining the effect of a nonpeptide vasopressin receptor antagonist (OPC-31260) have been conducted in several murine and rat models of PKD (Fig. 3). The drug prevented renal cAMP accumulation, inhibited renal cystogenesis and increase in renal cyst size, and decreased blood urea nitrogen levels [128]. Administration of the drug, even after the cystic disease was already established, halted the progression or caused regression of the renal cysts [128, 129]. The decrease in cAMP is associated with inhibition of Ras and ERK activity [130].

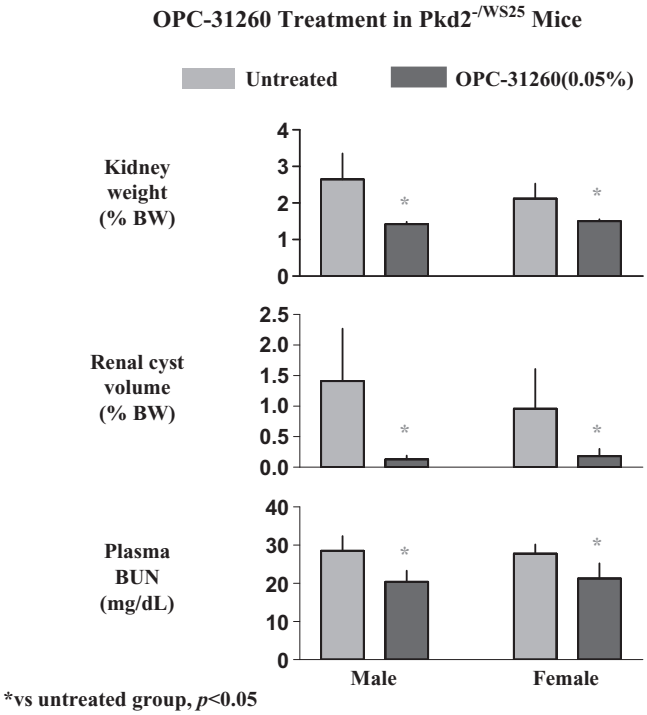


**Figure 3** Effects of the vasopressin receptor antagonist OPC-31269 on kidney weight, cyst volume, and blood urea nitrogen (BUN) in a murine model of ADPKD. (Adapted from ref. 127.)

Several VPV2R antagonists are currently in phase III clinical testing for the therapy of hyponatremia in patients with the syndrome of inappropriate antidiuretic hormone secretion and for patients with disorders of water retention such as congestive heart failure (Table 4). These drugs are attractive since they are relatively selective for the renal VPV2R and side effects have mostly included moderate thirst that is well tolerated.

An antagonist with high potency and selectivity for the human VPV2R, tolvaptan, has completed phase II testing in patients with ADPKD and a phase III clinical trial is now well underway.

***Somatostatin Antagonists*** As discussed above, cyst proliferation and fluid secretion into the cysts is under the control of cAMP. There are five receptors for somatostatin that bind to somatostatin or somatostatin analogs with various affinities [131]. The somatostatin type 2 receptor has been detected in the kidney and its stimulation results in inhibition of cAMP and stimulatory action on phosphatase and phospholipase C [132]. Mechanistically, this would seem an ideal target for therapy. Supporting this was the finding of the inhibition

**TABLE 4 Vasopressin Receptor Antagonists in Clinical Development**

| Name | Tolvaptan | Lixivaptan | Satavaptan | Mozavaptan | Conivaptan |
|---|---|---|---|---|---|
| Dose | 15–60 mg once daily | 50–100 mg twice daily | 5–25 mg once daily | 30–60 mg twice daily | 40–80 mg Infusion |
| Administration | Oral | Oral | Oral/IV | Oral | IV |
| SelectivityV1a/V2[a] | 29 | 100 | 112 | 10 | 0.15 |
| Half-life (h) | 6–8 | 7–10 | 14–17 | n/a | 3–8 |
| Metabolism[b] | Hepatic (CYP3A4) | Hepatic (CYP3A4) | Hepatic (90%CYP3A4) (10%CYP2D6) | Hepatic (CYP3A4) | Hepatic (CYP3A4) |
| Elimination | Stool | Stool | Stool | Stool | Stool |
| Manufacturer | Otsuka | CardiokineWyeth-Ayerst | Sanofi-Aventis | Otsuka | Astellas |

[a]V1a: vasopressin 1a receptor found on vasculature.
[b]CYP: cytochrome P450.

of cyst growth in an ADPKD patient with a pituitary adenoma treated with octreotide (a somastatin analog) for 2 years [133].

Recently, a randomized, cross-over, placebo-controlled trial has been conducted to evaluate the benefits and adverse effects of 6 months' treatment with slow-release octreotide or placebo in patients with ADPKD [133]. In the group of patients receiving octreotide, there was a 60% reduction in cyst volume, whereas there was no change in the placebo group. There was no change in GFR noticed in the octreotide group. These encouraging findings will need to be verified in larger and longer trials.

***Mammalian Target of Rapamycin (mTOR)*** Polycystin 1 has been demonstrated to interact with the tuberous sclerosis gene product tuberin [134]. There is a convergence of signaling pathway activation downstream of these proteins that leads to activation of the target of rapamycin (mTOR) [135]. mTOR is a serine threonine kinase that plays an important role in protein translation, cell growth, and proliferation [136]. Consistent with this finding, patients with ADPKD have shown increased mTOR activity [137].

On the basis of the hypothesis that mTOR in renal cystic cells leads to abnormal cyst formation, several studies have been conducted to evaluate the effects of mTOR inhibitors. Sirolimus and everolimus are FDA-approved specific inbitors of mTOR [138]. These drugs reduce cell growth, the rate of cell proliferation, and the rate of cell cycle progression [138]. In an rat model of PKD, sirolimus decreased tubular cell proliferation and inhibited cyst formation and renal enlargement [139]. Most recently, a retrospective study of ADPKD patients who had received a kidney transplant demonstrated that sirolimus treatment (as part of the immunosuppressant regimen) significantly decreased renal volumes of the cystic kidneys [140]. Phase III clinical trials of mTOR inhibitors are now underway to test this hypothesis in a prospective manner.

***Epidermal Growth Factor Receptor*** There are numerous studies that support a strong role of the epidermal growth factor (EGF)/transforming growth factor (TGF)–α/EGF receptor axis in promoting tubular epithelial cell proliferation, and cyst formation and enlargement [141]. Furthermore, studies have shown that renal cyst fluid from patients with ADPKD, as well as animal models of PKD, contain EGF-related peptides, and there is overexpression of the EGF receptor [142]. These receptors have a high affinity for these cyst-derived EGF-related peptides and result in autoactivation (through tyrosine kinase activity) of signaling cascades resulting in increased cellular proliferation [143].

Inhibition of EGF receptor tyrosine kinase activity significantly inhibits cyst development in animal models of PKD, suggesting that this may be a viable therapeutic option [144, 145]. Several selective therapeutic inhibitors with limited adverse reaction are in phase I or II clinical trials.

### 3.3 Conclusions

The rapid breakthroughs in the molecular understanding of pathogenesis of ADPKD had made several biochemical pathways logical avenues for potential therapeutic agents (Table 3). These therapeutics share a common theme: namely, intervening on growth promoting pathways to halt cyst formation and growth. Several

clinical trials are now ongoing, and it is hoped that a potential therapy for a disease that up until now has had no effective therapy will be available soon.

## 4 THERAPEUTIC AGENTS TO SLOW THE PROGRESSION OF CHRONIC KIDNEY DISEASE

Chronic kidney disease is growing in epidemic proportions. Within the United States both the incidence and prevalence of CKD are climbing annually. Estimates project that by 2014, 800,000 individuals will require either dialysis treatment or transplantation for CKD, a 60% increase from 2001. From 1991 to 2000 the prevalence of patients requiring some form of therapy for CKD within the U.S. population increased by 97% and the incidence by 57% [146, 147]. Obviously, this makes this area a rich opportunity for the development of novel therapeutic agents to both halt the progression of CKD to end-stage renal disease as well as treat the complications associated with the loss of kidney function.

### 4.1 Natural History of Progressive Chronic Renal Disease

The nephron is the fundamental structural and filtration unit of the kidney. The human kidney is composed of approximately 1–1.5 million nephrons. Each nephron is composed of multiple parts that provide for filtration of the blood and processing of the filtrate via resorption and excretion of both water and solute to create urine, which is excreted from the body. Chronic kidney disease progresses via a vicious cycle of nephron damage followed by an adaptive response that in many circumstances is critical in leading to CKD [148]. After an initial insult to the kidney (such as that due to hypertension, diabetes mellitus, or immune-complex-mediated glomerular disease of lupus nephritis), the kidney responds via a process called adaptive hyperfiltration in which the glomerular filtration rate (the volume of fluid filtered from the glomerular capillaries into Bowman's capsule per unit time) of the remaining normal nephrons is increased. The increased filtration rate of the surviving nephrons allows the remnant nephrons to do the filtration work previously performed by the entire kidney, allowing the serum creatinine in mild or early kidney disease to remain at or near the normal range. Adaptation within the tubules of surviving nephrons allows serum electrolytes to be maintained as well [149].

Although initially protective, adaptive hyperfiltration results in progressive damage to the surviving nephrons via elevated intraglomerular pressure and cellular damage. This resulting nephron damage due to hyperfiltration induces further adaptive hyperfiltration in a viscous cycle that contributes to the progression of kidney disease. The combination of this adaptive cycle, continuation/worsening of the initial disease insult, and multiple other pathophysiologic factors contribute to the chronic and often devastating progression of CKD.

### 4.2 Obstacles to the Development of Novel Therapeutic Agents to Prevent Progression of CKD

Just as with ADPKD, the time course for the progression of most kidney diseases occurs over a time span of decades. Thus, novel biomarkers that predict the risk of progression must be identified as surrogate markers for CKD progression. The biomarker most often studied is urine microalbumin or urine protein. Multiple studies have demonstrated that the level of urine protein correlates with the risk for progression and that treatment strategies that lower proteinuria lead to renal protection [147]. However, other biomarkers are needed, and their robustness in predicting CKD must be so great that the FDA would approve them as surrogate outcome markers for worsening renal function in clinical trials. Thus far, no such biomarkers exist.

### 4.3 Treatment Strategies

Treatment strategies for CKD can be multifaceted. There are wider population goals, which include improving initial identification of CKD patients so as to begin treatment earlier (with the hope that earlier therapy will prevent later progression). There are also cause-specific strategies that rely on appropriate and early therapy of underlying etiologies. For example, these would include aggressive glycemic control for diabetics, and immunosuppression for inflammatory diseases such as systemic lupus erythematosis. However, some responses of the kidney to injury are relatively stereotyped and allow for the development of more global preventative strategies that can be applied across the spectrum of causes [150]. Some of those will be discussed here and are listed in Table 5. In general, strategies to protect the kidney have thus far focused on agents that lower blood pressure as well as lower urine protein excretion.

***Angiotensin-Converting Enzyme Inhibitors (ACEi)*** The rennin–angiotensin–aldosterone system (RAAS) is activated by multiple stimuli. The kidneys respond to hypotension, decreased sodium concentration in the distal tubule, and decreased blood volume by releasing renin from the juxtaglomerular apparatus (Fig. 4). Renin, a proteolytic enzyme, cleaves angiotensinogen

(produced in the liver and released into the circulation) into angiotensin I. The angiotensin-converting-enzyme (ACE) present mainly in the pulmonary capillaries cleaves angiotensin I into angiotensin II. Angiotensin II has effects on many organs, but, most importantly, it acts as a potent vasoconstrictor and leads to the release of aldosterone from the adrenal gland. As shown in Figure 4 there are several potential steps in the pathway that can be blocked, and therapeutic drug development has focused on these various steps.

In the kidney, angiotensin II stimulates smooth muscle contraction; and, specifically at the glomerular level, angiotensin II causes asymmetric constriction of the efferent >> afferent arterioles of the glomerular apparatus. The increased efferent arteriole pressure results in increased hydrostatic pressure in the glomerular capillaries and increased GFR (responsible for the hyperfiltration phase of kidney injury). This effect of the RAAS has made the system a target to break the cycle of adaptive hyperfiltration that characterizes the progression of chronic kidney disease.

**TABLE 5 Therapeutic Strategies for the Treatment of Chronic Kidney Disease**

1. Optimal blood pressure control to <125/75 mmHg
   a. Utilization of drugs that preferentially block the renin–angiotensin system
      i. ACE inhibitors
      ii. Angiotensin receptor blockers
      iii. Direct renin inhibitors
      iv. Combination therapy
2. Use of statin agents
3. Use of drugs to limit oxidative damage
4. Nutritional management
5. Anti-inflammatory drugs
6. Treatment of specific etiologies
   a. Diabetes mellitus
   b. Systemic lupus eryuthematosis
   c. Other inflammatory diseases

Initial work upon a component of the venom of the *Bothrops jararaca* pit viper led to the discovery of a peptidic fraction referred to as bradykinin potentiating factor (BPF) that potentiated the effects of bradykinin in animal models by inhibition of bradykinin breakdown in the lung [151–154]. Further work demonstrated that BPF also inhibited the conversion of angiotensin I to angiotensin II. The component that was subsequently isolated was a nonapeptide, teprotide, which had the greatest ACE inhibition potency both in vitro and in vivo [151–156].

Subsequently, synthetic compounds were produced mimicking the structure of teprotide that were capable of ACE inhibiton and blood pressure lowering, utlimately leading to the discovery of captopril, the first orally bioavailable ACEi [151]. Furthermore, this discovery was one of the first examples of specific structurally defined pharmaceutical development.

Early animal experimentation evaluated the efficacy of ACEi as a means of slowing the progression of CKD [156]. Much of this experimentation utilized a model system whereby the animals underwent a four-fifths nephrectomy to lead to significant renal mass reduction.

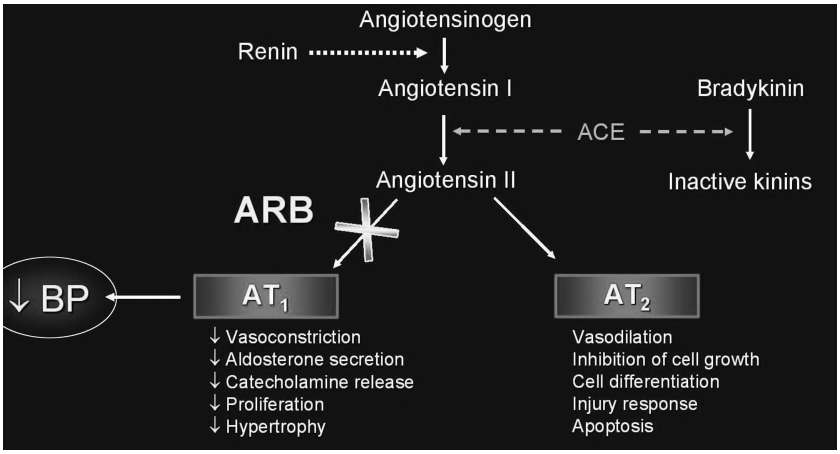


**Figure 4** Renin–angiotensin system. Renin released by the kidney acts as a proteolytic enzyme to cleave angiotensinogen to angiotensin I, which is cleaved to the biologically active angiotensin II. Angiotensin II can bind to either the type 1 or 2 angiotensin receptor to achieve their biological effects. Drugs that act on the renin–angiotensin system include: (1) direct renin inhibitors, which block the renin-mediated conversion of angiotensinogen to angiotensin I; (2) ACEi, which block the conversion of angiotensin I to angiotensin II; and (3) angiotensin receptor blockers, which block the AT1 receptor.

This leads to a marked increase in both perfusion and filtration of the remaining nephrons. These changes result in development of progressive proteinuria and eventual glomerulosclerosis, thus serving as a model of progressive kidney disease [156].

Importantly, ACEi protected these experimental rats from CKD progression. In these animal models, ACEi resulted in reduced capillary pressure within the glomerulus, inhibition of renal cellular growth, and reduced permeability of the glomerular capillary to protein, which reduced proteinuria and prevented the development of glomerulosclerosis [156]. Other animal research demonstrated that rennin–angiotensin system blockade exerts potent antiatherosclerotic effects, which are mediated by their antihypertensive, anti-inflammatory, antiproliferative, and oxidative stress-lowering properties [157].

Based upon these encouraging preclinical studies, many large clinical trials have been performed demonstrating the effectiveness of ACE inhibition in a CKD preventative strategy. In an early trial, patients were randomized to treatment with captopril or nifedipine for 3 years [158]. CKD progression was measured every 2 months. A greater decrease in urine protein excretion was demonstrated in ACEi-treated patients and in those patients with a mean blood pressure (BP) < 100 mmHg, a significantly slower decrease in GFR was demonstrated in patients treated with ACEi [158]. The seminal trial supporting the use of ACEi in a renal protective strategy occurred in 1993, when is was shown that captopril when given to type I diabetes with significant diabetic nephropathy led to a dramatic decrease in the rate of progressive renal disease (Fig. 5) [159]. Several other important studies have confirmed the significant benefits of ACEi (independent of blood pressure lowering) in patients with CKD across diverse causes [160, 161]. Overall, in a meta-analysis that compared blood pressure regimens containing an ACEi with regimens not including ACEi the risk of either doubling serum creatinine or reaching end-stage renal disease was reduced by 36% in ACEi-treated patients [162]. Furthermore, a reduction in urine protein excretion occurred only in patients treated with ACEi [162].

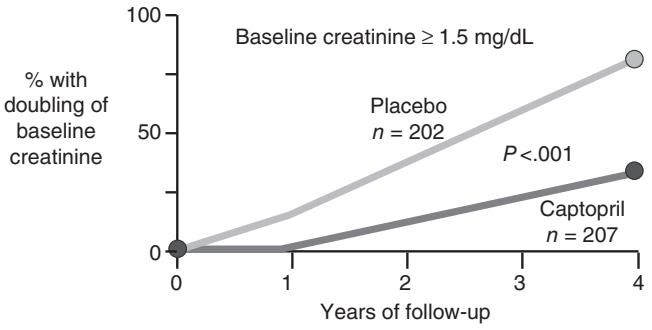


**Figure 5** Effects of the ACEi captopril on the progression of CKD in type I diabetics with existing nephropathy. (Adapted from ref. 158.)

As described above, targeting and inhibiting the angiotensin-converting enzyme also potentiates the effects of bradykinin via inhibition of bradykinin-destroying enzymes in the lung. Potentiation of bradykinin activity leads to two of the most frequently cited side effects of ACE inhibitors. The first is angioedema. Bradykinin is a potent vasodilator. Vasodilation can lead to rapid accumulation of interstitial fluid and swelling. When this swelling occurs around the mouth or the mucosa of the mouth or throat, stridor and respiratory compromise can quickly occur. Angioedema of differing severities and locations can occur in 0.1–0.2% of ACEi-treated patients and can be life threatening. ACE inhibitors can also result in a chronic cough thought to occur as a result of sensitization of airway sensory nerves by potentiated bradykinin [162]. Although less life threatening than angioedema, the cough can result in decreased patient comfort and quality of life and resulting decrease in medication compliance. The chronic, usually dry and nonproductive cough can occur in up to 20% of patients taking ACEi [163].

***Angiotensin Receptor Blockers*** Alternative methods of RAAS blockade have been developed. The angiotensin II type 1 receptor (AT1) receptor is thought to be the mediator of nearly all of the deleterious effects of angiotensin II signaling [164]. Through AT1 signaling, angiotensin II induces vasoconstriction, decreased renal blood flow, vascular smooth muscle cell growth, and proliferation, cardiac left ventricular hypertrophy, and sympathetic nervous system activation [164]. Agents that target the AT1 receptor (angiotensin receptor blockers, or ARBs) were, therefore, developed to attack the RAAS without targeting the ACE [164]. These agents should be devoid of the side effects described above for ACEi. Currently, seven branded ARBs are available in the United States.

Blockade of the AT1 receptor is often discussed as physiologically providing more complete termination of the negative signals of the RAAS compared to ACE inhibitors [165]. This is due to the fact that enzymes other than ACE, including trypsin, cathepsin G, chymostatin-sensitive angiotensin II generating enzyme (or CAGE), and chymase, all catalyze the formation of angiotensin II from angiotensin I [165]. Therefore, ARBs could theoretically achieve a higher degree of RAAS blockade.

Furthermore, by avoiding inhibition of ACE, AT1 receptor antagonists do not potentiate bradykinin. This

avoids triggering the mechanism of the chronic cough and angioedema seen with ACEi.

Finally, by blocking the AT1 receptor, increased levels of angiotensin II are potentially available to activate the angiotensin type 2, or AT2, receptor (Fig. 4). At the AT2 receptor, angiotensin II signals largely favorable effects, regulating differentiation and proliferation of cells, controlling tissue regeneration and apoptosis, as well as vasodilatation [162, 165]. Thus, ARBs have several potential advantages over ACEi: (1) better side effect profile, (2) more complete RAAS blockade, and (3) stimulation of the AT2 receptor.

The AT1 antagonists have now been well studied in clinical trials and demonstrated clearly to prevent the progression of CKD (diabetic or nondiabetic). For example, in the Irbesartan Diabetic Nephropathy Trial (IDNT), patients with diabetes mellitus type II (DMII) were randomized to daily irbesartan, amlodipine or placebo [166]. The primary combined endpoint was doubling the baseline serum creatinine, development of end-stage renal disease (ESRD), or death. While amlodopine demonstrated no benefit, the irbesartan group had a 20–23% lower incidence of achieving the primary endpoints. This suggests that in patients with type II diabetes, irbesartan helps to halt CKD progression beyond its blood pressure effect. [167]. Similar results have been seen in other trials of ARB [168, 169].

***Direct Renin Inhibitors*** In viewing the RAAS pathway, a logical step for blockade is at the level of renin. Blockade of this critical initial step would allow downregulation of the entire RAAS, including allowing the production of substrates for alternative non-RAAS pathways. However, the development of an orally active, bioavailable compound was difficult. In 1995, alliskiren was patented and was the first viable direct renin inhibitor. The drug is similar in structure to renin's natural substrate, angiotensinogen, but is a nonpeptide [170]. However, since renin is a species-specific enzyme, further preclinical testing required the development of a novel transgenic rat model outfitted with the human renin and angiotensinogen genes that allowed the testing of aliskiren in an animal model [170]. Preclinical data demonstrated that aliskiren prolonged survival, decreased cardiac hypertrophy, and the inducibility of arrhythmias, proteinuria, and attenuated inflammation in this model [170]. All these features might result in improved target-organ damage. Studies in humans attest to an effective BP-lowering action, a largely side-effect-free profile, and the option of several combination therapies [171]. Furthermore, recent data has now demonstrated significant falls in urine protein excretion in patients with nephropathy who receive a combination of aliskiren and losartan [172]. Aliskiren is the first of a novel antihypertensive drug class.

From a drug development strategy the RAAS pathway has provided a logical and rich source of targets. Thus far, all agents that effectively downregulate this system (acting at several different sites) have demonstrated clinical benefit. The next step in development will be clinical trials to test these various agents in combinations to more fully suppress the RAAS in the hope of better organ protection.

***Statins*** Multiple large clinical trials have shown statins to be the most effective pharmaceutical means of lowering low-density lipoprotein (LDL) cholesterol values and of significant value in preventing both primary and secondary cardiovascular outcomes. Cardiovascular events are one of the main causes of mortality and morbidity in patients with CKD [173]. Perhaps the best quantitation of this relationship is from analysis of the National Health and Nutrition Examination Survey population [174]. This work demonstrated increased cardiovascular disease mortality in patients with both reduced GFR and increased proteinuria. Mortality rates related to cardiovascular disease in this population were 4.1, 8.6, and 20.5 deaths/1000 person-years among participants with an estimated GFR of ≥90, 70–89, and <70 mL/min, respectively.

Implementation of statin therapy in CKD patients would thus be logical to treat a major cause of mortality in this population. However, there are also other means by which statin therapy could theoretically be beneficial in slowing the progression of CKD. Statin therapy is highly anti-inflammatory and markers of inflammation such as C-reactive protein fall significantly once statins are initiated, and this effect is independent of its effect on cholesterol levels [175]. By decreasing serum levels of inflammatory cytokines, it is possible that statins could benefically impact one of the pathways of CKD progression [176].

A second mechanism by which statin therapy could potentially be useful in treatment of the progression of chronic kidney disease is via the effects these drugs have on the vasculature. Statin therapy has been proven to have beneficial effects on both endothelial function as well as vascular tone [177–179]. Theoretically, these improvements in vascular function could lead to improved renal blood flow and function. Interestingly, statin use for organ protection outside of the cardiovascular system has also been proposed for the central nervous system [180].

In terms of clinical outcomes, a post hoc analysis of the Heart Protection Study demonstrated that progressive creatinine elevation was slowed in the simvastatin-treated group [181]. Pravastatin in the Cholesterol and

Recurrent Events (CARE) trial demonstrated an overall reduction in the rate of CKD progression in patients treated with pravastatin of 34% [182]. This equates to an absolute reduction in the rate of GFR decrease by 0.2 mL/min per year in patients treated with pravastatin versus placebo. These encouraging results require validation in long-term, large multicenter clinical trials.

### 4.4 Drugs Specifically Targeted for Diabetic Kidney Disease

Diabetic nephropathy is the most common cause of ESRD in the United States and represents 44 of all new cases of ESRD in the United States. Diabetic nephropathy is distinct from other forms of CKD in that it is characterized by particular pathological abnormalities such as glomerular basement membranes (GBM) thickening and mesangial expansion that alters the configuration of the normal glomerulus structure [183].

The primary goals toward treatment of diabetic nephropathy are strict blood sugar and blood pressure control as well as treatment with ACEi, ARBs, and statins (as described above). In addition to these treatment goals, other new modalities are being developed to specifically target the glomerular abnormalities of diabetic nephropathy. This effort is driven by an extensive knowledge of the pathogenesis of diabetic nephropathy (Fig. 6).

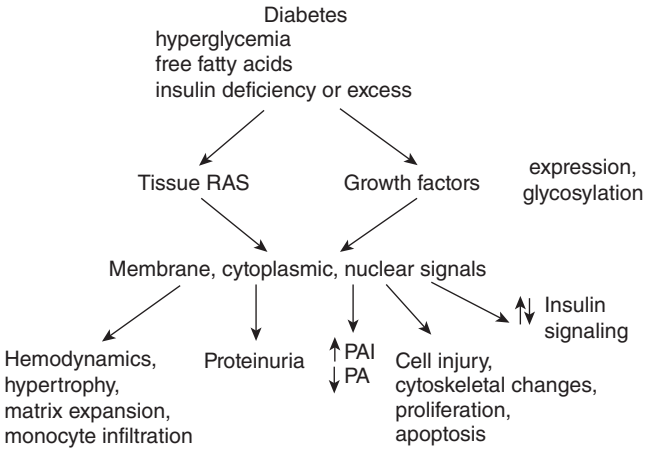


**Figure 6** Pathogenetic pathways for diabetic nephropathy. Under the influence of hyperglycemia, free fatty acids, and insulin deficiency or excess, there is progressive activation of growth factors, cytokines, and the tissue renin–angiotensin system (RAS). This ultimately leads to damage to cellular membranes, activation of plasminogen activator inhibitor (PAI), and decreases in plasminogen activator (PA) and the pathological hallmarks of diabetic nephropathy (matrix expansion and glomerular basement membrane thickening).

***Sulodexide*** This agent is comprised of four linked glycosaminoglycan (GAG) polysaccharides that are isolated from porcine intestinal mucosa [184]. Although the exact mechanism of GAGs in diabetic nephropathy patients is unknown, preclinical studies and functional theory suggest that these compounds can prevent or reverse GBM thickening, prevent and correct GBM anionic charge depletion via both direct deposition as well as induction of heparin sulfate synthesis and sulfation, as well as inhibit heparinase-1 activity [185]. In preclinical work, sulodexide also suppressed early pathological mesangial cell proliferation [186]. The clinical development of sulodexide, which has occurred over the past decade, has included eight small exploratory European trials, a phase II European dose-ranging study, and a phase III trial initiated in the United States in 2004. No serious adverse events have been noted [187]. Encouragingly, there have been significant reductions in urine protein excretion in those patients receiving sulodexide [188]. However, definitive results will require the completion of the U.S. study.

***Pyridoxamine*** This agent is an effective inhibitor of the formation of advanced glycation end products (AGEs) that arise from the breakdown of glycated proteins (Amadori products) [189]. AGEs are thought to be intimately involved in the pathogenesis of diabetic complications [189]. Pyridoxamine is a derivative of vitamin $B_6$, and several studies have now demonstrated that oral pyridoxamine is highly effective in maintaining glomerular function in diabetic animal models [190]. The drug has shown a favorable safety profile in phase II testing [191]. In a single study, the rate of increase in serum creatinine in the pyridoxamine group tended to be lower compared with placebo [191]. Further studies are ongoing.

***Ruboxistaurin*** Protein kinase C (PKC) is a group of multifunctional isoenzymes involved in signal transduction that leads to cell growth, fibrosis, and tissue injury [192]. PKC activity is increased in diabetics, and it is believed that inhibition of PKC activity could be beneficial in preventing both micro- and macrovascular complications of diabetes [193]. Ruboxistaurin is a specific inhibitor of the PKC-β isoform [194]. It has been extensively evaluated in animal models and shown to normalize glomerular hyperfiltration (an early sign of diabetic nephropathy) and attenuate proteinuria [195]. In a pilot trial, 123 patients with diabetic nephropathy were randomized to placebo or ruboxistaurin. Ruboxistaurin led to significant reductions in proteinuria and maintained renal function over one year [196]. The drug has been more extensively studied for the prevention of retinopathy and is now under FDA review for this indication

[197]. Larger scale trials investigating renal protection are underway.

## 5 TREATING THE COMPLICATIONS OF CKD: VITAMIN D ANALOGS

One of the most common complications of advanced CKD is the development of metabolic bone disease due in part to a combination of a deficiency of active vitamin D as well as impairments in renal phosphorus excretion. The precursor of active vitamin D can be acquired from the diet from plant sources in the form of vitamin $D_2$ or from animal sources as the fat-soluble steroid vitamin $D_3$ (cholecalciferol). If not acquired in the diet, vitamin $D_3$ can be synthesized in the human skin where ultraviolet radiation results in cleavage of 7-dehydrocholesterol and formation of vitamin $D_3$. In the microsomes and mitochondria of the liver, hydroxylation by cytochrome-P450-like enzymes results in formation of 25-hydroxycholecalciferol or calcifediol. This is then transferred to the proximal tubule cells of the kidney where a second hydroxylation is performed by 25(OH)D-1-α-hydroxylase to create the mature hormone 1,25-dihydroxyvitamin D or calcitriol [198, 199] (Fig. 7).

Calcitriol alters gene expression by binding with high affinity to its intracellular receptor, the vitamin D receptor (VDR), which acts as a nuclear transcription factor. Liganded VDR undergoes conformational change and forms a heterodimer with a second protein, the retinoid X receptor (RXR). This, in turn, binds to DNA elements in the promoter regions of target genes described as vitamin D response elements (VDREs). Binding to the VDREs may promote transcription, as is the case with osteocalcin in the osteoblast, or inhibit transcription, as for the parathyroid hormone (PTH) in the parathyroids, by either enhancing or repressing the activity of transcription machinery [198, 200].

The development of advanced CKD (usually when the GFR is less than 60 mL/min) results in progressive loss of nephron mass, including the loss of proximal tubule cells capable of producing mature, active vitamin D, and thus vitamin D deficiency ensues [198]. Deficiency of mature vitamin D promotes the hyperplasia of parathyroid gland tissue and increased synthesis of PTH. Vitamin D deficiency also decreases serum calcium levels via decreasing gut absorption of calcium. The resulting serum hypocalcemia triggers further increased secretion of PTH. Thus, the deficiency of vitamin D leads to a characteristic biochemical abnormality seen in patients with CKD: secondary hyperparathyroidism. The consequences of secondary hyperparathyroidism include: (1) enhanced bone resorption triggered by elevated PTH levels (osteitis fibrosa cystica) leading to spontaneous fractures and (2) in the setting of an elevated calcium–phosphate product (approximately 30 $mg^2/dL^2$ in normal patients and up to 60–70 $mg^2/dL^2$ in patients with advanced CKD) metastatic calcifications develop in arteries, soft tissues, and organs with effects ranging from ischemia to organ failure [201, 202].

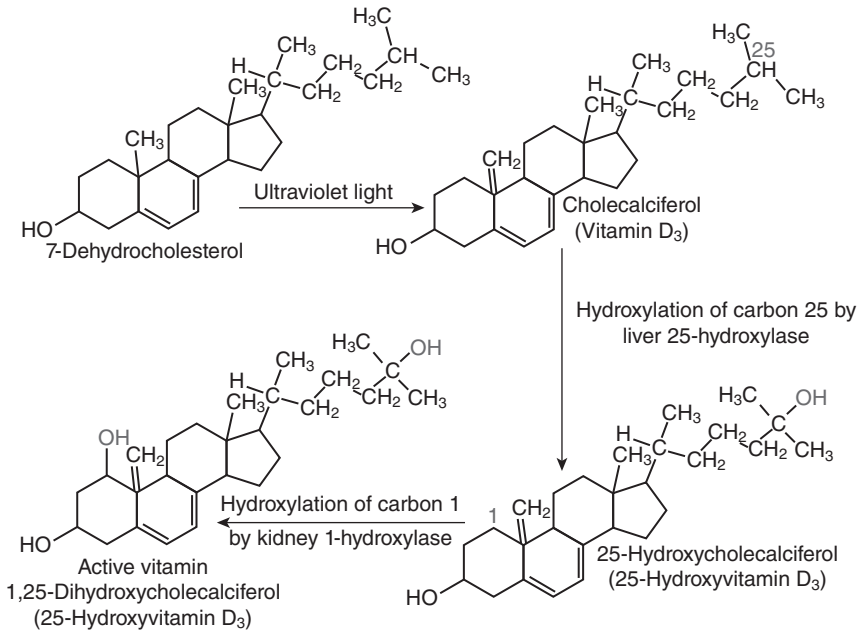


**Figure 7** Biosynthetic pathway for active vitamin D production.

Given that vitamin D deficiency is a primary cause of these abnormalities in CKD patients and that 1-α-hydroxylation is impaired in CKD, vitamin D analogs not requiring hydroxylation in the kidney have been developed to treat this condition. These include calcitriol, doxercalciferol, and paricalcitol in the United States [203].

### 5.1 Multiple Vitamin D Analogs

The major goal of the treatment of the vitamin D deficiency associated with CKD is to achieve suppression of PTH secretion, while limiting the actions of vitamin D on intestinal calcium and phosphorus absorption. To achieve this, drug development has utilized the variable affinity of vitamin D receptors in different tissues to develop an analog that can signal the parathyroid tissues to inhibit their hyperactivity and yet avoid the intestinal effects that would drive up the serum calcium–phosphorus product and lead to an increased risk for metastasic calcification. The efficacy of these medications have been individually studied in CKD patients [203].

Calcitriol is the natural form of active vitamin D and is available in an oral and intravenous form. It was the first agent administered to treat the metabolic bone disease of CKD and is the basic preparation to which others are compared for efficacy and safety profiles.

Paricalcitol (19-nor-1-α,25-dihydroxyvitamin $D_2$) is the classic "second-generation" vitamin D analog. Initial analysis of paricalcitol suggested that this medication was as efficacious as calcitriol with regards to lowering serum PTH levels [203]. It also suggested that the paricalcitol was less likely to result in hypercalcemia and that, therefore, the medication likely had a wider therapeutic window to suppress PTH without the hypercalcemia and hyperphosphatemic effects [204, 205]. The mechanism that explains this selectivity is not understood. Paricalcitol and calcitriol have been compared to each other in clinical trials. In a seminal study, paricalcitol was compared with calcitriol in 263 hemodialysis patients with plasma PTH levels greater than 300 pg/mL and a serum calcium phosphate product less than 75 $mg^2/dL^2$ [206]. Primary endpoints of this study included the ability of each drug to reduce the plasma PTH to 50% of baseline values; the incidence of hypercalcemia and/or a single occurrence of an elevated calcium–phosphorus product. In this study, no statistically significant difference was found in the ability of the two drugs to lower plasma PTH levels. Approximately 60% of each group met the target reduction of serum PTH. The study also failed to demonstrate a statistically significant difference between the two drugs with regards to either the incidence of a single occurrence of hypercalcemia and/or a single occurrence of elevated calcium–phosphorus product. However, when the groups were compared with regards to a secondary endpoint measuring incidence of hypercalcemia for two consecutive laboratory draws and/or calcium–phosphorus product >75 $mg^2/dL^2$ for four consecutive laboratory draws, the paricalcitol group performed better with only 18% of patients attaining these elevations versus 33% of the calcitriol patients. This suggested that paricalcitol was as effective as calcitriol in suppressing PTH, and, most importantly, it also suggested that paricalcitol was perhaps safer with regard to protecting patients from the dangers of prolonged elevated calcium–phosphorus products.

This study was followed by a landmark mortality study involving a historical cohort that compared 36-month survival among patients on long-term hemodialysis between 1999 and 2001 between patients receiving calcitriol or paricalcitol [207]. The paricalcitol-treated group had a lower mortality rate of 0.180 per person-year versus calcitriol with a mortality rate of 0.223 per person-year. Paricalcitol-treated patients also had less elevation in levels of both calcium and phosphorus at all times measured. The results of this retrospective analysis suggest a survival advantage with the second-generation paricalcitol versus its predecessor calcitriol. However, a randomized prospective study is required to confirm these important results [208].

Also approved for vitamin D supplementation in CKD for use in the United States is doxercalciferol. Limited comparative analysis has been done between this medication and its predecessors. In small trials, it has demonstrated the ability to decrease PTH levels in CKD patients with secondary hyperparathyroidism [209]. Analogs used outside of the United States include alfacalcidol, falecalcitriol, and 22-oxacalcitriol [210]. However, as with doxercalciferol, there is limited clinical data for these compounds.

## 6 TREATING THE COMPLICATIONS OF CKD: CALCIMIMETICS

As described above, the majority of patients with chronic kidney disease stages III and IV, and virtually all patients with end-stage renal disease, develop secondary hyperparathyroidism due to impaired phosphorus excretion, PTH resistance, and vitamin D deficiency. Secondary hyperparathyroidism as defined by an elevated serum PTH has been associated with numerous adverse events such as an increased risk of fractures, an increased risk of vascular calcifications resulting in calcemic uremic arteriolopathy, and calcification of soft tissues including the myocardium resulting in significantly increased cardiovascular morbidity and mortality [201]. The National

Kidney Foundation (NKF) quality guidelines aim to achieve a calcium–phosphorus product of less than 55 $mg^2/dL^2$ and a PTH less than 35–70 pg/mL, 70–110 pg/mL, or 150–300 pg/mL for CKD stages III, IV, and V, respectively [211].

Multiple studies, most notably the Dialysis Outcomes and Practice Patterns Study (DOPPS), have shown that failure of patients with end-stage renal disease to achieve the above practice guideline target ranges for calcium–phosphorus product and serum PTH are associated with significantly increased mortality [212, 213]. Additionally, there is a body of evidence suggesting that not only is failure to meet PTH goals associated with increased mortality but that the degree of increased mortality climbs with increasing levels of serum PTH [212, 213]. Unfortunately, multiple studies have shown that the majority of patients on hemodialysis fail to meet the NKF guidelines for PTH, and a substantial number fail to meet the calcium–phosphorus product guidelines as well, thus representing an area where potential treatments are urgently needed [212].

Calcimimetics are a novel class of agents that act on the calcium-sensing receptor (CaSR) on parathyroid gland cells to treat secondary hyperparathyroidism by directly reducing the secretion of PTH [214]. Cincacalcet is currently the only FDA-approved agent in this class of drugs. Multiple preclinical and clinical trials have demonstrated cinacalcet's efficacy in achieving the surrogate clinical endpoint of reduced PTH, with a currently small but growing body of evidence that its effects on secondary hyperparathyroidism correlate with reductions in fractures and cardiovascular morbidity [215]. Interestingly, the development of calcimimetics occurred concurrently with the discovery of the CaSR.

### 6.1 Discovery and Characterization of the Calcium-Sensing Receptor and Agonist Development and Mechanism of Action

The development of cinacalcet is unique in that the agent is both the first approved drug to act on a G-coupled protein receptor (GPCR) and the first drug aimed specifically at lowering PTH directly, rather than acting to lower serum phosphorus to achieve a secondary reduction in PTH [215]. Shortly after the development of the first calcimimetic agents, the receptor on which they act upon was discovered. The CaSR, a member of one of the less common GPCR families, is unique among all discovered GPCRs in that it is the only known receptor in the superfamily to bind caclium and magnesium as a first messenger. Hebert and Brown first isolated, cloned, and characterized the CaSR protein from bovine parathyroid tissue in 1993 [214]. Calcimimetics were being developed concomitantly, having been identified through molecular screening assays performed in parathyroid tissue before the discovery of the CaSR [215].

The first calcimimetics, R-467 and R-568, were identified through screening of small-molecule componds (analogs of fendiline, a calcium-channel blocker marketed as an antianginal drug) in assays of bovine parathyroid tissue [215, 216]. Several compounds, R-467, R-568, and cinacalcet, were shown to have a dose-dependent increase in cytosolic calcium, which was used as a marker of intracellular G-protein activation [215, 217]. Also, R-568 showed promise in early animal trials, mediating both a dose-dependent decrease in serum PTH and serum calcium. However, its development for humans was abandoned due both to poor oral bioavailability and extensive metabolism by CYP2D6, leading to a very high risk for drug–drug interactions [217]. Cinacalcet, developed several years later, caused a similar level of decrease in serum PTH but had greater than 70% oral bioavailability and was metabolized through multiple CYP isoforms with a lower risk for drug–drug interactions [218].

Preclinical trials have demonstrated that cinacalcet is able to reduce serum PTH in both rat models of uremia and cultured parathyroid gland cells derived from patients with severe secondary hyperparathyroidism [219, 220]. Additonally, and perhaps more of a clinically relevant endpoint, animal studies exist that suggest cinacalcet and other calcimimetics may aid in preventing extraosseous calcification (as assessed by biopsy) and renal osteodystrophy (assessed by dual-energy X-ray absorptiometry) [221, 222]. As control of serum PTH is used as a surrogate endpoint for, among other objectives, reduction in metabolic bone disease due to renal osteodystrophy, these studies are suggestive that cinacalcet may play a role in reducing fractures and calcemic uremic arteriolopathy. However, this has yet to be proven in clinical trials.

Cinacalcet's mechanism of action is unusual in several regards. Not only is it the first FDA-approved drug to act on the CaSR, it is the first approved allosteric inhibitor of any GPCR. In addition to cinacalcet's effects on the CaSR in the cells of the parathyroid gland, there is some evidence to suggest that calcimimetic agents may act on the CaSR in thyroid parafollicular C cells to release calcitonin, which would in turn lower calcium in a manner independent of PTH [223]. A documented effect on gastrin-secreting cells of the antrum may also play a role in hypocalcemia through somewhat unclear mechanisms [224]. The exact mechanism of cinacalcet's interaction with the CaSR has not been determined, largely due to the inability to date to generate crystals of the CaSR suitable for analysis. However, computerized models of CaSR based upon homology with bovine

rhodopsin have been developed, and R-568 and another calcimimetic, NPS-2143, have been hypothesized to bind residues on the second, third, and seventh transmembrane domains of CaSR [225]. Two site-directed mutagenesis experiments that show decreased affinity for calcimimetic with mutations of Glu-837 in the seventh transmembrane domain support this hypothesis [225, 226]. Future basic science research related to development of new calcimimetics may serve to further elucidate cinacalcet's mechanism of action.

### 6.2 Effects of Cinacalcet on the Calcium–Phosphorus Product and PTH

There is strong clinical trial evidence that cinacalcet is effective at reducing serum PTH and phosphorus. An analysis of pooled phase III trial data showed that 307 out of 547 patients on hemodialysis (56%) treated with cinacalcet and standard care met NKF guidelines for serum PTH versus 10% of standard care and placebo-treated patients, and 65% of cinacalcet-treated patients in the same analysis met NKF calcium–phosphorus product guidelines versus 36% with placebo [227]. The end result of these trials was that 41% of cinacalcet-treated patients met both calcium–phosphorus product and PTH guidelines versus 6% of placebo-treated patients. As of yet, open-label clinical data evaluating cinacalcet on PTH control has only been extended to 3 years, with the majority of patients who achieve NKF goals for PTH with cinacalcet able to maintain these goals through this time period [228].

### 6.3 Effects of Cincalcet on Morbidity and Mortality and Adverse Effects

While cinacalcet has repeatedly demonstrated substantial effects upon biochemical parameters, evidence for its clinical outcomes has been somewhat limited to date. An elevated calcium–phosphorus product has been repeatedly shown to correlate with increased mortality among dialysis patients, but a systematic review examining eight randomized trials of cinacalcet for patients with secondary hyperparathyroidism, five of which evaluated mortality as an endpoint, confirmed improvement in patients' PTH and calcium–phosphorus levels, but did not detect a significant improvement in any patient-level outcome, including all-cause mortality [229]. A combined analysis of four randomized, double-blinded trials evaluating cinacalcet for secondary hyperparathyroidism in patients with end-stage renal disease demonstrated a 14-fold reduction in parathyroidectomy, a 2-fold reduction in fracture, and a 40% reduction in hospitalizations for cardiovascular disease [230].

Data from phase III clinical trials suggests that cinacalcet has relatively few serious adverse effects. Gastrointestinal toxicities such as nausea, vomiting, and diarrhea were most commonly reported with rates of approximately 2–3%, with largely asymptomatic hypocalcemia observed in 1–5% of patients [231, 232]. Preliminary evidence from a phase IIIB trial suggests that taking cinacalcet with meals may ameliorate this nausea somewhat [233].

While more clinical trials with a longer period of observation are needed to evaluate the clinical endpoints of therapy with cinacalcet, particularly in regard to mortality, current data suggests that calcimimetics have potential to become a routine therapy in patients with advanced kidney disease.

## 7 SUMMARY

Renal disease, especially CKD, is increasing at an exponential pace. This makes this area of medicine attractive for the development of therapeutic agents. Specific disease states such as acute kidney injury and CKD are especially attractive given their known pathogenesis and logical targets for therapeutic agent design. The next decade will likely see a large increase in the number of therapeutic agents in this arena. One difficulty that will remain will be the design of clinical trials that can demonstrate therapeutic benefit on a feasible time scale. The development of surrogate and informative biomarkers will greatly aid this endeavor.

## REFERENCES

1. Coresh J, Astor BC, Greene T, Eknoyan G, Levey AS (2003). Prevalence of chronic kidney disease and decreased kidney function in the adult US population: Third National Health and Nutrition Examination Survey. *Am J Kidney Dis 41*, 1–12.
2. Waikar SS, Curhan GC, Wald R, McCarthy EP, Chertow GM (2006). Declining mortality in patients with acute renal failure, 1988 to 2002. *J Am Soc Nephrol 17*, 1143–1150.
3. Xue JL, Daniels F, Star RA, Kimmel PL, Eggers PW, Molitoris BA, Himmelfarb J, Collins AJ (2006). Incidence and mortality of acute renal failure in Medicare beneficiaries, 1992 to 2001. *J Am Soc Nephrol 17*, 1135–1142.
4. Star RA (1998). Treatment of acute renal failure. *Kidney Int 54*, 1817–1831.
5. Mehta RL, Pascual MT, Soroko S, Savage BR, Himmelfarb J, Ikizler TA, Paganini EP, Chertow GM (2004). Spectrum of acute renal failure in the intensive care unit: The PICARD experience. *Kidney Int 66*, 1613–1621.

6. Sutton TA, Fisher CJ, Molitoris BA (2002). Microvascular endothelial injury and dysfunction during ischemic acute renal failure. *Kidney Int 62*, 1539–1549.
7. Conger JD (1995). Interventions in clinical acute renal failure: What are the data? *Am J Kidney Dis 26*, 565–576.
8. Ronco C, Bellomo R (2003). Prevention of acute renal failure in the critically ill. *Nephron 93*, C13–C20.
9. Solomon R, Werner C, Mann D, D'Elia J, Silva P (1994). Effects of saline, mannitol, and furosemide on acute decreases in renal function induced by radiocontrast agents. *N Engl J Med 151*, 208–260.
10. Merten GJ, Burgess WP, Gray LV, Holleman JH, Roush TS, Kowalchuk GJ, Bersin RM, Van Moore A, Simonton CA 3rd, Rittase RA, Norton HJ, Kennedy TP (2004). Prevention of contrast-induced nephropathy with sodium bicarbonate: A randomized controlled trial. *JAMA 291*, 2328–2334.
11. Rudnick MR, Goldfarb S, Wexler L, Ludbrook PA, Murphy MJ, Halpern EK, Hill JA, Winniford M, Cohen MB, VanFossen DB (1995). Nephrotoxicity of ionic and nonionic contrast media in 1196 patients: A randomized trial. The Iohexol Cooperative Study. *Kidney Int 47*, 254–261.
12. Aspelin P, Aubry P, Fransson SG, Strasser R, Willenbrock R, Berg KJ (2003). Nephrotoxicity in high-risk patients study of iso-osmolar and low-osmolar non-ionic contrast media study investigators: Nephrotoxic effects in high-risk patients undergoing angiography. *N Engl J Med 348*, 491–499.
13. Ix JH, McCulloch CE, Chertow GM (2004). Theophylline for the prevention of radiocontrast nephropathy: A meta-analysis. *Nephrol Dial Transplant 19*, 2747–2753.
14 Kelly KJ, Molitoris BA (2000). Acute renal failure in the new millennium: Time to consider combination therapy. *Semin Nephrol 20*, 4–19.
15. Chertow GM, Levy EM, Hammermeister KE, Grover F, Daley J (1998). Independent association between acute renal failure and mortality following cardiac surgery. *Am J Med 104*, 343–348.
16. Lien J, Chan V (1985). Risk factors influencing survival in acute renal failure treated by hemodialysis. *Arch Intern Med 145*, 2067–2069.
17. Groeneveld AB, Tran DD, van der Meulen J, Nauta JJ, Thijs LG (1991). Acute renal failure in the medical intensive care unit: Predisposing, complicating factors and outcome. *Nephron 59*, 602–610.
18. Rabb H, Chamoun F, Hotchkiss J (2001). Molecular mechanisms underlying combined kidney-lung dysfunction during acute renal failure. *Contrib Nephrol 132*, 41–52.
19. Kelly KJ (2003). Distant effects of experimental renal ischemia/reperfusion injury. *J Am Soc Nephrol 14*, 1549–1558.
20. Star R (2001). Design issues for clinical trials in acute renal failure. *Blood Purif 19*, 233–237.
21. Miller SB, Martin DR, Kissane J, Hammerman MR (1992). Insulin-like growth factor I accelerates recovery from ischemic acute tubular necrosis in the rat. *Proc Natl Acad Sci USA 89*, 11876–11880.
22. Goes N, Urmson J, Vincent D, Ramassar V, Halloran PF (1996). Effect of recombinant human insulin-like growth factor-1 on the inflammatory response to acute renal injury. *J Am Soc Nephrol 7*, 710–720.
23. Daemen MA, van 't Veer C, Denecker G, Heemskerk VH, Wolfs TG, Clauss M, Vandenabeele P, Buurman WA (1999). Inhibition of apoptosis induced by ischemia-reperfusion prevents inflammation. *J Clin Invest 104*, 541–549.
24. Hirschberg R, et al. (1999). Multicenter clinical trial of recombinant human insulin-like growth factor I in patients with acute renal failure. *Kidney Int 55*, 2423–2432.
25. Knepper MA, Lankford SP, Terada Y (1991). Renal tubular actions of ANF. *Can J Physiol Pharmacol 69*, 1537–1545.
26. Allgren RL, Marbury TC, Rahman SN, Weisberg LS, Fenves AZ, Lafayette RA, Sweet RM, Genter FC, Kurnik BRC, Conger JD, Sayegh MH (1997). Anaritide in acute tubular necrosis. Auriculin Anaritide Acute Renal Failure Study Group. *N Engl J Med 336*, 828–834.
27. Lewis J, Salem MM, Chertow GM, Weisberg LS, McGrew F, Marbury TC, Allgren RL (2000). Atrial natriuretic factor in oliguric acute renal failure. Anaritide Acute Renal Failure Study Group. *Am J Kidney Dis 36*, 767–774.
28. Sward K, Valsson F, Odencrants P, Samuelsson O, Ricksten SE (2004). Recombinant human atrial natriuretic peptide in ischemic acute renal failure: A randomized placebo-controlled trial. *Crit Care Med 32*, 1310–1315.
29. Bellomo R, Chapman M, Finfer S, Hickling K, Myburgh J (2000). Low-dose dopamine in patients with early renal dysfunction: A placebo-controlled randomised trial. Australian and New Zealand Intensive Care Society (ANZICS) Clinical Trials Group. *Lancet 356*, 2139–2143.
30. Friedrich JO, Adhikari N, Herridge MS, Beyene J (2005). Meta-analysis: Low-dose dopamine increases urine output but does not prevent renal dysfunction or death. *Ann Intern Med 142*, 510–524.
31. Denton MD, Chertow GM, Brady HR (1996). "Renal-dose" dopamine for the treatment of acute renal failure: Scientific rationale, experimental studies and clinical trials. *Kidney Int 50*, 4–14.
32. Lauschke A, Teichgraber UK, Frei U, Eckardt KU (2006). "Low-dose" dopamine worsens renal perfusion in patients with acute renal failure. *Kidney Int 69*, 1669–1674.
33. Han WK, Bonventre JV (2004). Biologic markers for the early detection of acute kidney injury. *Curr Opin Crit Care 10*, 476–482.

34. Ichimura T, Hung CC, Yang SA, Stevens JL, Bonventre JV (2004). Kidney injury molecule-1: A tissue and urinary biomarker for nephrotoxicant-induced renal injury. *Am J Physiol Renal Physiol 286*, F552–F563.
35. Mishra J, Mori K, Ma Q, Kelly C, Barasch J, Devarajan P (2004). Neutrophil gelatinase-associated lipocalin: A novel early urinary biomarker for cisplatin nephrotoxicity. *Am J Nephrol 24*, 307–315.
36. Parikh CR, Mishra J, Thiessen-Philbrook H, Dursun B, Ma Q, Kelly C, Dent C, Devarajan P, Edelstein CL (2006). Urinary IL-18 is an early predictive biomarker of acute kidney injury after cardiac surgery. *Kidney Int 70*, 199–203.
37. Esson ML, Schrier RW (2002). Diagnosis and treatment of acute tubular necrosis. *Ann Intern Med 137*, 744–752.
38. de Mendonca A, Vincent JL, Suter PM, Moreno R, Dearden NM, Antonelli M, Takala J, Sprung C, Cantraine F (2000). Acute renal failure in the ICU: Risk factors and outcome evaluated by the SOFA score. *Intens Care Med 26*, 915–921.
39. Metnitz PG, Krenn CG, Steltzer H, Lang T, Ploder J, Lenz K, Le Gall JR, Druml W (2002). Effect of acute renal failure requiring renal replacement therapy on outcome in critically ill patients. *Crit Care Med 30*, 2051–2058.
40. Cosentino F, Chaff C, Piedmonte M (1994). Risk factors influencing survival in ICU acute renal failure. *Nephrol Dial Transplant 9*(Suppl 4), 179–182.
41. Bellomo R, Ronco C, Kellum JA, Mehta RL, Palevsky P (2004). Acute renal failure: Definition, outcome measures, animal models, fluid therapy and information technology needs—The Second International Consensus Conference of the Acute Dialysis Quality Initiative (ADQI) Group. *Crit Care 8*, R204–R212.
42. Palevsky PM, Metnitz PG, Piccinni P, Vinsonneau C (2002). Selection of endpoints for clinical trials of acute renal failure in critically ill patients. *Curr Opin Crit Care 8*, 515–518.
43. Tiwari MM, Brock RW, Megyesi JK, Kaushal GP, Mayeux PR (2005). Disruption of renal peritubular blood flow in lipopolysaccharide-induced renal failure: Role of nitric oxide and caspases. *Am J Physiol Renal Physiol 289*, F1324–F1332.
44. Melnikov VY, Faubel S, Siegmund B, Lucia MS, Ljubanovic D, Edelstein CL (2002). Neutrophil-independent mechanisms of caspase-1- and IL-18-mediated ischemic acute tubular necrosis in mice. *J Clin Invest 110*, 1083–1091.
45. Kelly KJ, Sutton TA, Weathered N, Ray N, Caldwell EJ, Plotkin Z, Dagher PC (2004). Minocycline inhibits apoptosis and inflammation in a rat model of ischemic renal injury. *Am J Physiol Renal Physiol 287*, F760–F766.
46. Sutton TA, Kelly KJ, Mang HE, Plotkin Z, Sandoval RM, Dagher PC (2004). Minocycline reduces renal microvascular leakage in a rat model of ischemic renal injury. *Am J Physiol Renal Physiol 288*, F91–F97.
47. Kelly KJ, Plotkin Z, Dagher PC (2001). Guanosine supplementation reduces apoptosis and protects renal function in the setting of ischemic injury. *J Clin Invest 108*, 1291–1298.
48. Kelly KJ, Plotkin Z, Vulgamott SL, Dagher PC (2003). P53 mediates the apoptotic response to GTP depletion after renal ischemia-reperfusion: Protective role of a p53 inhibitor. *J Am Soc Nephrol 14*, 128–138.
49. Padanilam BJ (2003). Cell death induced by acute renal injury: A perspective on the contributions of apoptosis and necrosis. *Am J Physiol Renal Physiol 284*, F608–F627.
50. Chiarugi A (2002). Poly(ADP-ribose) polymerase: Killer or conspirator? The "suicide hypothesis" revisited. *Trends Pharmacol Sci 23*, 122–129.
51. Chatterjee PK, Chatterjee BE, Pedersen H, Sivarajah A, McDonald MC, Mota-Filipe H, Brown PA, Stewart KN, Cuzzocrea S, Threadgill MD, Thiemermann C (2004). 5-Aminoisoquinolinone reduces renal injury and dysfunction caused by experimental ischemia/reperfusion. *Kidney Int 65*, 499–509.
52. Thiemermann C, Bowes J, Myint FP, Vane JR (1997). Inhibition of the activity of poly(ADP ribose) synthetase reduces ischemia-reperfusion injury in the heart and skeletal muscle. *Proc Natl Acad Sci USA 94*, 679–683.
53. Zingarelli B, Cuzzocrea S, Zsengeller Z, Salzman AL, Szabo C (1997). Protection against myocardial ischemia and reperfusion injury by 3-aminobenzamide, an inhibitor of poly (ADP-ribose) synthetase. *Cardiovasc Res 36*, 205–215.
54. Walker PD, Shah SV (1988). Evidence suggesting a role for hydroxyl radical in gentamicin-induced acute renal failure in rats. *J Clin Invest 81*, 334–341.
55. Walker PD, Shah SV (1990). Reactive oxygen metabolites in endotoxin-induced acute renal failure in rats. *Kidney Int 38*, 1125–1132.
56. Baliga R, Zhang Z, Baliga M, Ueda N, Shah SV (1998). In vitro and in vivo evidence suggesting a role for iron in cisplatin-induced nephrotoxicity. *Kidney Int 53*, 394–401.
57. Baliga R, Ueda N, Walker PD, Shah SV (1999). Oxidant mechanisms in toxic acute renal failure. *Drug Metab Rev 31*, 971–997.
58. Ulloa L, Ochani M, Yang H, Tanovic M, Halperin D, Yang R, Czura CJ, Fink MP, Tracey KJ (2002). Ethyl pyruvate prevents lethality in mice with established lethal sepsis and systemic inflammation. *Proc Natl Acad Sci USA 99*, 12351–12356.
59. Miyaji T, Hu X, Yuen PS, Muramatsu Y, Iyer S, Hewitt SM, Star RA (2003). Ethyl pyruvate decreases sepsis-induced acute renal failure and multiple organ damage in aged mice. *Kidney Int 64*, 1620–1631.
60. Grey ST, Tsuchida A, Hau H, Orthner CL, Salem HH, Hancock WW (1994). Selective inhibitory effects of the anticoagulant activated protein C on the responses of human mononuclear phagocytes to LPS, IFN-gamma, or phorbol ester. *J Immunol 153*, 3664–3672.

61. Grey S, Hau H, Salem HH, Hancock WW (1993). Selective effects of protein C on activation of human monocytes by lipopolysaccharide, interferon-gamma, or PMA: Modulation of effects on CD11b and CD14 but not CD25 or CD54 induction. *Transplant Proc 25*, 2913–2914.
62. Mizutani A, Okajima K, Uchiba M, Noguchi T (2000). Activated protein C reduces ischemia/reperfusion-induced renal injury in rats by inhibiting leukocyte activation. *Blood 95*, 3781–3787.
63. van den Berghe G, Wouters P, Weekers F, Verwaest C, Bruyninckx F, Schetz M, Vlasselaers D, Ferdinande P, Lauwers P, Bouillon R (2001). Intensive insulin therapy in the critically ill patients. *N Engl J Med 345*, 1359–1367.
64. Basi S, Pupim LB, Simmons EM, Sezer MT, Shyr Y, Freedman S, Chertow GM, Mehta RL, Paganini E, Himmelfarb J, Ikizler TA (2005). Insulin resistance in critically ill patients with acute renal failure. *Am J Physiol Renal Physiol 289*, F259–F264.
65. Langouche L, Vanhorebeek I, Vlasselaers D, Vander Perre S, Wouters PJ, Skogstrand K, Hansen TK, Van den Berghe G (2005). Intensive insulin therapy protects the endothelium of critically ill patients. *J Clin Invest 115*, 2277–2286.
66. Mesotten D, Swinnen JV, Vanderhoydonc F, Wouters PJ, Van den Berghe G (2004). Contribution of circulating lipids to the improved outcome of critical illness by glycemic control with intensive insulin therapy. *J Clin Endocrinol Metab 89*, 219–226.
67. Celik M, Gokmen N, Erbayraktar S, Akhisaroglu M, Konakc S, Ulukus C, Genc S, Genc K, Sagiroglu E, Cerami A, Brines M (2002). Erythropoietin prevents motor neuron apoptosis and neurologic disability in experimental spinal cord ischemic injury. *Proc Natl Acad Sci USA 99*, 2258–2263.
68. Yang CW, Li C, Jung JY, Shin SJ, Choi BS, Lim SW, Sun BK, Kim YS, Kim J, Chang YS, Bang BK (2003). Preconditioning with erythropoietin protects against subsequent ischemia-reperfusion injury in rat kidney. *FASEB J 17*, 1754–1755.
69. Vesey DA, Cheung C, Pat B, Endre Z, Gobe G, Johnson DW (2004). Erythropoietin protects against ischaemic acute renal injury. *Nephrol Dial Transplant 19*, 348–355.
70. Sharples EJ, Patel N, Brown P, Stewart K, Mota-Philipe H, Sheaff M, Kieswich J, Allen D, Harwood S, Raftery M, Thiemermann C, Yaqoob MM (2004). Erythropoietin protects the kidney against the injury and dysfunction caused by ischemia-reperfusion. *J Am Soc Nephrol 15*, 2115–2124.
71. Vaziri ND, Zhou XJ, Liao SY (1994). Erythropoietin enhances recovery from cisplatin-induced acute renal failure. *Am J Physiol 266*, F360–F366.
72. Patschan D, Krupincza K, Patschan S, Zhang Z, Hamby C, Goligorsky MS (2006). Dynamics of mobilization and homing of endothelial progenitor cells after acute renal ischemia: Modulation by ischemic preconditioning. *Am J Physiol Renal Physiol 291*, F176–F185.
73. Bahlmann FH, DeGroot K, Duckert T, Niemczyk E, Bahlmann E, Boehm SM, Haller H, Fliser D (2003). Endothelial progenitor cell proliferation and differentiation is regulated by erythropoietin. *Kidney Int 64*, 1648–1652.
74. Nakamura T, Nishizawa T, Hagiya M, Seki T, Shimnishi M, Sugimura A, Tashiro K, Shimizu S (1989). Molecular cloning and expression of human hepatocyte growth factor. *Nature 342*, 440–443.
75. Gohda E, Tsubouchi H, Nakayama H, Hirono S, Sakiyama O, Takahashi K, Miyazaki H, Hashimoto S, Daikuhara Y (1988). Purification and partial characterization of hepatocyte growth factor from plasma of a patient with fulminant hepatic failure. *J Clin Invest 81*, 414–419.
76. Liu Y, Tolbert EM, Lin L, Thursby MA, Sun AM, Nakamura T, Dworkin LD (1999). Up-regulation of hepatocyte growth factor receptor: An amplification and targeting mechanism for hepatocyte growth factor action in acute renal failure. *Kidney Int 55*, 442–453.
77. Rabkin R, Fervenza F, Tsao T, Sibley R, Friedlaender M, Hsu F, Lassman C, Hausmann M, Huie P, Schwall RH (2001). Hepatocyte growth factor receptor in acute tubular necrosis. *J Am Soc Nephrol 12*, 531–540.
78. Kawaida K, Matsumoto K, Shimazu H, Nakamura T (1994). Hepatocyte growth factor prevents acute renal failure and accelerates renal regeneration in mice. *Proc Natl Acad Sci USA 91*, 4357–4361.
79. Mizuno S, Nakamura T (2005). Prevention of neutrophil extravasation by hepatocyte growth factor leads to attenuations of tubular apoptosis and renal dysfunction in mouse ischemic kidneys. *Am J Pathol 166*, 1895–1905.
80. Nath KA, Balla G, Vercellotti GM, Balla J, Jacob HS, Levitt MD, Rosenberg ME (1992). Induction of heme oxygenase is a rapid, protective response in rhabdomyolysis in the rat. *J Clin Invest 90*, 267–270.
81. Sikorski E, HT, Hill-Kapturczak N, Agarwal A (2004). The story so far: Molecular regulation of the heme oxygenase-1 gene in renal injury. *Am J Physiol Renal Physiol 286*, F425–F441.
82. Vera T, Henegar JR, Drummond HA, Rimoldi JM, Stec DE (2005). Protective effect of carbon monoxide-releasing compounds in ischemia-induced acute renal failure. *J Am Soc Nephrol 16*, 950–958.
83. Adin CA, Croker BP, Agarwal A (2005). Protective effects of exogenous bilirubin on ischemia-reperfusion injury in the isolated, perfused rat kidney. *Am J Physiol Renal Physiol 288*, F778–F784.
84. Nakao A, Neto JS, Kanno S, Stolz DB, Kimizuka K, Liu F, Bach FH, Billiar TR, Choi AM, Otterbein LE, Murase N (2005). Protection against ischemia/reperfusion injury in cardiac and renal transplantation with carbon monoxide, biliverdin and both. *Am J Transplant 5*, 282–291.
85. Birck R, Knoll T, Braun C, Kirchengast M, Munter K, van der Woude FJ, Rohmeiss P (1998). Improvement of

postischemic acute renal failure with the novel orally active endothelin-A receptor antagonist LU 135252 in the rat. *J Cardiovasc Pharmacol 32*, 80–86.

86. Heyman SN, Clark BA, Kaiser N, Spokes K, Rosen S, Brezis M, Epstein FH (1992). Radiocontrast agents induce endothelin release in vivo and in vitro. *J Am Soc Nephrol 3*, 58–65.
87. Goto K, Hama H, Kasuya Y (1996). Molecular pharmacology and pathophysiological significance of endothelin. *Jpn J Pharmacol 72*, 261–290.
88. Edwards RM, Stack EJ, Pullen M, Nambi P (1993). Endothelin inhibits vasopressin action in rat inner medullary collecting duct via the ETB receptor. *J Pharmacol Exp Ther 267*, 1028–1033.
89. Hayasaki Y, Nakajima M, Kitano Y, Iwasaki T, Shimamura T, Iwaki K (1996). ICAM-1 expression on cardiac myocytes and aortic endothelial cells via their specific endothelin receptor subtype. *Biochem Biophys Res Commun 229*, 817–824.
90. Wilhelm SM, Stowe NT, Robinson AV, Schulak JA (2001). The use of the endothelin receptor antagonist, tezosentan, before or after renal ischemia protects renal function. *Transplantation 71*, 211–216.
91. Spiegel S, Milstien S (2003). Sphingosine-1-phosphate: An enigmatic signalling lipid. *Nat Rev Mol Cell Biol 4*, 397–407.
92. Brinkmann V, Davis MD, Heise CE, Albert R, Cottens S, Hof R, Bruns C, Prieschl E, Baumruker T, Hiestand P, Foster CA, Zollinger M, Lynch KR (2002). The immune modulator FTY720 targets sphingosine 1-phosphate receptors. *J Biol Chem 277*, 21453–21457.
93. Awad AS, Ye H, Huang L, Li L, Foss FW Jr, Macdonald TL, Lynch KR, Okusa MD (2006). Selective sphingosine 1-phosphate 1 receptor activation reduces ischemia-reperfusion injury in mouse kidney. *Am J Physiol Renal Physiol 290*, F1516–F1524.
94. Lien YH, Yong KC, Cho C, Igarashi S, Lai LW (2006). S1P(1)-selective agonist, SEW2871, ameliorates ischemic acute renal failure. *Kidney Int 69*, 1601–1608.
95. Foss FW Jr, Clemens JJ, Davis MD, Snyder AH, Zigler MA, Lynch KR, Macdonald TL (2005). Synthesis, stability, and implications of phosphothioate agonists of sphingosine-1-phosphate receptors. *Bioorg Med Chem Lett 15*, 4470–4474.
96. Linden J (2001). Molecular approach to adenosine receptors: Receptor mediated mechanisms of tissue protection. *Annu Rev Pharmacol Toxicol 41*, 775–787.
97. Jordan JE, Zhao Z, Sato H, Taft S, Vinten-Johansen J (1997). Adenosine A2 receptor activation attenuates reperfusion injury by inhibiting neutrophil accumulation, superoxide generation and coronary endothelial adherence. *J Pharmacol Exp Ther 280*, 301–309.
98. Lasley RD, JM, Mentzer RM Jr (2001). Beneficial effects of adenosine (2a) agonist CGS-21680 in infarcted and stunned porcine myocardium. *Am J Physiol Heart Circ Physiol 280*, H1660–H1666.
99. Day YJ, Marshall MA, Huang L, McDuffie MJ, Okusa MD, Linden J (2004). Protection from ischemic liver injury by activation of A2A adenosine receptors during reperfusion: Inhibition of chemokine induction. *Am J Physiol Gastrointest Liver Physiol 286*, G285–G293.
100. Day YJ, HL, McDuffie MJ, Rosin DL, Ye H, Chen JF, Schwarzschild MA, Fink JS, Linden J, Okusa MD (2003). Renal protection from ischemia mediated by A2A adenosine receptors on bone marrow-derived cells. *J Clin Invest 112*, 883–891.
101. Okusa MD, Linden J, Macdonald T, Huang L (1999). Selective A2A-adenosine receptor activation during reperfusion reduces ischemia-reperfusion injury in rat kidney. *Am J Physiol 277*, F404–F412.
102. Okusa MD, Linden J, Huang L, Rosin DL, Smith DF, Sullivan G (2001). Enhanced protection from renal ischemia-reperfusion injury with A2A-adenosine receptor activation and PDE 4 inhibition. *Kidney Int 59*, 2114–2125.
103. Lee HT, Ota-Setlik A, Xu H, D'Agati VD, Jacobson MA, Emala CW (2003). A3 adenosine receptor knockout mice are protected against ischemia- and myoglobinuria-induced renal failure. *Am J Physiol Renal Physiol 284*, F267–F273.
104. Lee HT, Gallos G, Nasr SH, Emala CW (2004). A1 adenosine receptor activation inhibits inflammation, necrosis, and apoptosis after renal ischemia-reperfusion injury in mice. *J Am Soc Nephrol 15*, 102–111.
105. Noiri E, Peresleni T, Miller F, Goligorsky MS (1996). In vivo targeting of inducible NO synthase wtih oligodeoxynucleotides protects rat kidneys against ischemia. *J Clin Invest 97*, 2377–2383.
106. Daynes RA, Jones DC (2002). Emerging roles of PPARs in inflammation and immunity. *Nat Rev Immunol 2*, 748–759.
107. Li S, Gokden N, Okusa MD, Bhatt R, Portilla D (2005). Anti-inflammatory effect of fibrate protects from cisplatin-induced ARF. *Am J Physiol Renal Physiol 289*, F469–F480.
108. Mishra J, Mori K, Ma Q, Kelly C, Yang J, Mitsnefes M, Barasch J, Devarajan P (2004). Amelioration of ischemic acute renal injury by neutrophil gelatinase-associated lipocalin. *J Am Soc Nephrol 15*, 3073–3082.
109. Riedemann NC, Neff TA, Guo RF, Bernacki KD, Laudes IJ, Sarma JV, Lambris JD, Ward PA (2003). Protective effects of IL-6 blockade in sepsis are linked to reduced C5a receptor expression. *J Immunol 170*, 503–507.
110. Deng J, Kohda Y, Chiao H, Wang Y, Hu X, Hewitt SM, Miyaji T, McLeroy P, Nibhanupudy B, Li S, Star RA (2001). Interleukin-10 inhibits ischemic and cisplatin-induced acute renal injury. *Kidney Int 60*, 2118–2128.
111. Chiao H, Kohda Y, McLeroy P, Craig L, Housini I, Star RA (1997). Alpha-melanocyte-stimulating hormone protects against renal injury after ischemia in mice and rats. *J Clin Invest 99*, 1165–1172.

112. Iglesias CG, et al. (1983). Epidemiology of adult polycystic kidney disease, Olmstead County, Minnesota. *Am J Kidney Dis 2*, 630–639.
113. Renal Data System US (USRDS). (1999). USRDS 1999 annual data report. Bethesda, MD: National Institutes of Health.
114. Hughes J, et al. (1995). The polycystic kidney disease 1 (PKD1) gene encodes a novel protein with multiple cell recognition domains. *Nat Genet 10*, 151–160.
115. Mochizuki T, et al. (1996). PKD2, a gene for polycystic kidney disease that encodes an integral membrane protein. *Science 272*, 1339–1342.
116. Ong AC, Harris PC (2005). Molecular pathogenesis of ADPKD: The polycystin complex gets complex. *Kidney Int 67*, 1234–1247.
117. Yoder BK, Hou X, Guay-Woodford LM (2002). The polycystic kidney disease proteins: Polycystin-1, polycystin-2, polaris, and cystin and co-localized in renal cilia. *J Am Soc Nephrol 13*, 2508–2516.
118. Masoumi A, Reed-Gitomer B, Kelleher C, Schrier RW (2007). Potential pharmacological interventions in polycystic kidney disease. *Drugs 67*, 2495–2501.
119. Cai Y, et al. (1999). Identification and characterization of polycystin-2, the PKD2 gene product. *J Biol Chem 274*, 28557–28565.
120. Gonzalez-Perrett S, et al. (2001). Polycystin-2, the protein mutated in autosomal dominant polycystic kidney disease, is a $Ca^{2+}$-permeable non-selective cation channel. *Proc Natl Acad Sci USA 98*, 1182–1187.
121. Igarashi P, Somlo S (2002). Genetics and pathogenesis of polycystic kidney disease. *J Am Soc Nephrol 13*, 2384–2398.
122. Dousa TP (1999). Cyclic-3′5′-nucleotide phosphodiesterase isozymes in cell biology and pathophysiology of the kidney. *Kidney Int 55*, 29–62.
123. Grantham JJ, Torres V, Chapman AB, for the CRISP Investigators (2006). Volume progression in polycystic kidney disease. *N Engl J Med 354*, 2122–2130.
124. Chabardes D, Imbert-Teboul M, Elalouf JM (1999). Functional properties of $Ca^{2+}$-inhibitable type 5 and type 6 adenylyl cyclases and the role of $Ca^{2+}$ increase in the inhibition of intracellular cAMP content. *Cell Signal 11*, 651–663.
125. Yamaguchi T, et al. (2004). Calcium restriction allows cAMP activation of the B-Raf/ERK pathway, switching cells to a cAMP-dependent growth stimulated phenotype. *J Biol Chem 279*, 40419–40430.
126. Hanaoka K, Guggino W (2000). cAMP regulates cell proliferation and cyst formation in autosomal polycystic kidney disease cells. *J Am Soc Nephrol 11*, 1179–1187.
127. Michalski A, Grzesczak W (1996). The effect of hypervolemia on electrolyte level and level of volume regulating hormones in patients with autosomal dominant polycystic kidney disease. *Pol Arch Med Wewn 96*, 329–343.
128. Torres VE, et al. (2004). Effective treatment of an orthologous model of autosomal dominant polycystic kidney disease. *Nat Med 10*, 363–364.
129. Gattone VH, et al. (2003). Inhibition of renal cystic disease development and progression by a vasopressin V2 receptor antagonist. *Nat Med 9*, 1323–1326.
130. Wang X, et al. (2005). Effectiveness of vasopressin V2 receptor antagonists OPC-31260 and OPC 41061 on polycystic kidney disease development in the PCK rat. *J Am Soc Nephrol 16*, 846–851.
131. Reubi JC, et al. (1993). Human kidney as target for somatostatin. High affinity receptors in tubules and vasa recta. *J Clin Endocrinol Metab 17*, 1323–1328.
132. Reichlin S (1983). Somatostatin. *N Engl J Med 309*, 1495–1501.
133. Ruggenenti P, et al. (2005). Safety and efficacy of long-acting somatostatin treatment in autosomal dominant polycystic kidney disease. *Kidney Int 68*, 206–216.
134. Kleymenova E, et al. (2001). Tuberin-dependent membrane localization of polycystin-1: A functional link between polycystic kidney disease and the TSC2 tumor suppressor gene. *Mol Cell 4*, 823–832.
135. Li Y, et al. (2004). TSC2: Filling the GAP in the mTOR signaling pathway. *Trends Biochem Sci 29*, 32–38.
136. Schmelzle T, Hall MN (2000). TOR, a central controller of cell growth. *Cell 103*, 253–262.
137. Inoki K, Corradetti MN, Guan KI (2005). Dysregulation of the TSC-mTOR pathway in human disease. *Nat Genet 37*, 19–24.
138. Huang S, Houghton PJ (2003). Targeting mTOR signaling for cancer therapy. *Curr Opin Pharmacol 3*, 371–377.
139. Tao Y, et al. (2005). Rapamycin markedly slows disease progression in a rat model of polycystic kidney disease. *J Am Soc Nephrol 16*, 46–51.
140. Shillingford J, Murcia NS, Larson CH (2006). The mTOR pathway is regulated by polycystin-1 and its inhibition reverses renal cystogenesis in polycystic kidney disease. *Proc Natl Acad Sci USA 103*, 5247–5251.
141. Lee D, Chan K, Chan S (1998). Expression of transforming growth factor α and epidermal growth factor receptor in adult polycystic kidney disease. *J Urol 159*, 291–296.
142. Richards W, et al. (2000). Epidermal growth factor receptor activity mediates renal cyst formation in polycystic kidney disease. *J Clin Invest 101*, 935–939.
143. Nakanishi K, Sweeney W Jr, Avner ED (2001). Segment-specific c-ErbB2 expression in human autosomal recessive polycystic kidney disease. *J Am Soc Nephrol 12*, 379–384.
144. Sweeney WE Jr, et al. (2000). Treatment of polycystic kidney disease with a novel tyrosine kinase inhibitor. *Kidney Int 57*, 33–40.
145. Torres VE, et al. (2003). EGF receptor tyrosine kinase inhibition attenuates the development of PKD in Han:SPRD rats. *Kidney Int 64*, 1573–1579.

146. Coresh J, et al. (2005). Chronic kidney disease awareness, prevalence, and trends among U.S. adults, 1999 to 2000. *J Am Soc Nephrol 16*, 180–188.
147. Xue JL, Ma JZ, Louis TA, Collins AJ (2001). Forecast of the number of patients with end-stage renal disease in the United States to the year 2010. *J Am Soc Nephrol 12*, 2753–2758.
148. Stevens LA, Coresh J, Greene T, Levey AS (2006). Assessing kidney function—measured and estimated glomerular filtration rate. *N Engl J Med 354*, 2473–2483.
149. Remuzzi G, Bertani T (1998). Pathophysiology of progressive nephropathies. *N Engl J Med 339*, 1448–1456.
150. Levey AS, et al. (2003). National kidney foundation practice guidelines for chronic kidney disease: Evaluation, classification, and stratification. *Ann Intern Med 139*, 137–147.
151. Bakhle YS (1968). Conversion of angiotensin I to angiotensin II by cell-free extracts of dog lung. *Nature 220*, 919–921.
152. Ng KK, Vane JR (1968). Fate of angiotensin I in the circulation. *Nature 218*, 144–150.
153. Ferreira SH, Greene LH, Alabaster VA, Bakhle YS, Vane JR (1970). Activity of various fractions of bradykinin potentiating factor against angiotensin I converting enzyme. *Nature 225*, 379–380.
154. Ferreira SH, Vane JR (1967). The detection and estimation of bradykinin in the circulating blood. *Br J Pharmacol Chemother 29*, 367–377.
155. Patlak M (2004). From viper's venom to drug design: Treating hypertension. *FASEB J 18*, 421.
156. Anderson S, Rennke HG, Brenner BM (1986). Therapeutic advantage of converting enzyme inhibitors in arresting progressive renal disease associated with systemic hypertension in the rat. *J Clin Invest 77*, 1993–2000.
157. Schmieder RE, Higers KF, Schlaich MP, Schmidt BM (2007). Renin-angiotensin system and cardiovascular risk, *Lancet 369*, 1208–1219.
158. Zucchelli P, et al. (1992). Long-term comparison between captopril and nifedipine in the progression of renal insufficiency. *Kidney Int 42*, 452–458.
159. Lewis EJ, Hunsicker LG, Bain RP, Rohde RD (1993). The effect of angiotensin-converting-enzyme inhibition on diabetic nephropathy. The Collaborative Study Group. *N Engl J Med 329*, 1456–1462.
160. Maschio G, et al. (1999). Angiotensin-converting enzyme inhibitors and kidney protection: The AIPRI trial. The ACE inhibition in progressive renal insufficiency (AIPRI) study group. *J Cardiovasc Pharmacol 33*(Suppl 1), S16–S20.
161. Klahr S, et al. (1994). The effects of dietary protein restriction and blood-pressure control on the progression of chronic renal disease. Modification of diet in renal disease study group. *N Engl J Med 330*, 877–884.
162. Jafar TH, et al. (2001). Angiotensin-converting enzyme inhibitors and progression of nondiabetic renal disease. A meta-analysis of patient-level data. *Ann Intern Med 135*, 73–87.
163. Fox AJ, Lalloo UG, Belvisi MG, Bernareggi M, Chung KF, Barnes PJ (1996). Bradykinin-evoked sensitization of airway sensory nerves: A mechanism for ACE-inhibitor cough. *Nat Med 2*, 814–817.
164. Kaschina E, Unger T (2003). Angiotensin AT1/AT2 receptors: Regulation, signalling and function. *Blood Press 12*, 70–88.
165. Unger T (2002). The role of the renin-angiotensin system in the development of cardiovascular disease. *Am J Cardiol 89*, 3A–9A.
166. Parving HH, et al. (2001). The effect of irbesartan on the development of diabetic nephropathy in patients with type 2 diabetes. *N Engl J Med 345*, 870–878.
167. Kramkowski K, Mogielnicki A, Buczko W (2006). The physiological significance of the alternative pathways of angiotensin II production. *J Physio Pharmacol 57*, 529–539.
168. Brenner BM, et al. (2000). The losartan renal protection study—Rationale, study design and baseline characteristics of RENAAL (reduction of endpoints in NIDDM with the angiotensin II antagonist losartan). *J Renin-Angiotensin-Aldosterone Syst 1*, 328–335.
169. Rodby RA, et al. (2000). The irbesartan type II diabetic nephropathy trial: Study design and baseline patient characteristics. For the collaborative study group. *Nephrol Dial Transplant 15*, 487–497.
170. Muller DN, Derer W, Dechend R (2008). Aliskiren-mode of action and preclinical data. *J Mol Med 86*, 659–662.
171. Shafiq MM, Menon DV, Victor RG (2008). Oral direct renin inhibition: Premise, promise, and potential limitations of a new antihypertensive drug. *Am J Med 121*, 265–271.
172. Parving H-H, et al. (2008). Aliskiren combined with losartan in type 2 diabetes and nephropathy. *N Engl J Med 358*, 2433–2446.
173. Muntner P, He J, Hamm L, Loria C, Whelton PK (2002). Renal insufficiency and subsequent death resulting from cardiovascular disease in the united states. *J Am Soc Nephrol 13*, 745–753.
174. Garg AX, Clark WF, Haynes RB, House AA (2002). Moderate renal insufficiency and the risk of cardiovascular mortality: Results from the NHANES I. *Kidney Int 61*, 1486–1494.
175. Zimmermann J, Herrlinger S, Pruy A, Metzger T, Wanner C (1999). Inflammation enhances cardiovascular risk and mortality in hemodialysis patients. *Kidney Int 55*, 648–658.
176. Chang JW, Yang WS, Min WK, Lee SK, Park JS, Kim SB (2002). Effects of simvastatin on high-sensitivity C-reactive protein and serum albumin in hemodialysis patients. *Am J Kidney Dis 39*, 1213–1217.
177. Blacher J, Guerin AP, Pannier B, Marchais SJ, Safar ME, London GM (1999). Impact of aortic stiffness on survival in end-stage renal disease. *Circulation 99*, 2434–2439.

178. Ichihara A, Hayashi M, Ryuzaki M, Handa M, Furukawa T, Saruta T (2002). Fluvastatin prevents development of arterial stiffness in haemodialysis patients with type 2 diabetes mellitus. *Nephrol Dial Transpl 17*, 1513–1517.
179. Kosch M, Barenbrock M, Suwelack B, Schaefer RM, Rahn KH, Hausberg M (2003). Effect of a 3-year therapy with the 3-hydroxy-3-methylglutaryl coenzyme a reductase-inhibitor fluvastatin on endothelial function and distensibility of large arteries in hypercholesterolemic renal transplant recipient. *Am J Kidney Dis 41*, 1088–1096.
180. Orr JD (2008). Statins in the spectrum of neurologic disease. *Curr Atheroscler Rep 10*, 11–18.
181. Tonelli M, et al. (2005). Effect of pravastatin on rate of kidney function loss in people with or at risk for coronary disease. *Circulation 112*, 171–178.
182. Navaneethan SD, Pansini F, Strippoli GF (2006). Statins in patients with chronic kidney disease: Evidence from systematic reviews and randomized clinical trials. *PLoS Med/Publ Lib Sci 3*(5), e123.
183. Kanwar YS, et al. (2008). Diabetic nephropathy: Mechanisms of renal disease progression. *Exp Biol Med 233*, 4–11.
184. Ofosu FA (1998). Pharmacological actions of sulodexide. *Semin Thromb Hemost 24*, 127–138.
185. Gambaro G, et al. (1992). Glycosaminoglycans prevent morphological renal alterations and albuminuria in diabetic rats. *Kidney Int 42*, 285–291.
186. Gambaro G, Baggio B (1996). Glycosaminoglycans: A new paradigm in the prevention of proteinuria and progression of glomerular disease. *Nephrol Dial Transplant 11*, 762–764.
187. Solini A, et al. (1997). Glycosaminoglycans delay the progression of nephropathy in NIDDM. *Diabetes Care 20*, 819–823.
188. Poplawska A, et al. (1997). Effect of glycosaminoglycans on urinary albumin excretion in insulin-dependent diabetic patients with micro- or macroalbuminuria. *Diabetes Res Clin Pract 38*, 109–114.
189. Edelstein D, Brownlee M (1992). Mechanistic studies of advanced glycation end product inhibition by aminoguanidine. *Diabetes 41*, 26–29.
190. Williams ME, et al. (2003). A phase 2 clinical trial of pyridoxamine (Pyridorin) in type 1 and type 2 diabetic patients with overt nephropathy (PYR-206). *J Am Soc Nephrol 14*, 7A.
191. McGill JB, et al. (2004). A phase 2 clinical investigation of pyridoxamine (Pyridorin) in type 1 and type 2 diabetic patients with overt nephropathy (PYR 205-207). *J Am Soc Nephrol 15*, 565A.
192. Goekijian PG, Iirousek MR (1999). Protein kinase C in the treatment of disease: Signal transduction pathways, inhibitors and agents in development. *Curr Med Chem 6*, 877–903.
193. DeRubertis FR, Craven PA (1994). Activation of protein kinase C in glomerular cells in diabetes: Mechanisms and potential links to the pathogenesis of diabetic glomerulopathy. *Diabetes 43*, 1–8.
194. Anderson PW, McGill JB, Tuttle KR (2007). Protein kinase C-beta inhibition: The promise for treatment of diabetic nephropathy. *Curr Opin Nephrol 16*, 397–402.
195. Kelly DJ, et al. (2003). Protein kinase C-beta inhibition attenuates the progression of experimental diabetic nephropathy in the presence of continued hypertension. *Diabetes 52*, 512–518.
196. Tuttle KR, et al. (2005). The effect of ruboxistaurin on nephropathy in type 2 diabetes. *Diabetes Care 28*, 2686–2690.
197. The PKC-DRS2 Study Group (2006). Effect of ruboxistaurin on visual loss in patients with diabetic retinopathy. *Ophthalmology 113*, 2221–2230.
198. Holick MF (2007). Vitamin D deficiency. *N Engl J Med 357*, 266–281.
199. Rose BD, Rennke HG (1994). *Renal Pathophysiology—The Essentials*. Baltimore, MD: Lippincott Williams and Wilkins.
200. Levin A, et al. (2007). Prevalence of abnormal serum vitamin D, PTH, calcium, and phosphorus in patients with chronic kidney disease: Results of the study to evaluate early kidney disease. *Kidney Int 71*, 31–38.
201. Goldsmith D, Ritz E, Covic A (2004). Vascular calcification: A stiff challenge for the nephrologist: Does preventing bone disease cause arterial disease? *Kidney Int 66*, 1315–1333.
202. Davies MR, Hruska KA (2001). Pathophysiological mechanisms of vascular calcification in end-stage renal disease. *Kidney Int 60*, 472–479.
203. Palmer SC, McGregor DO, Macaskill P, Craig JC, Elder GJ, Strippoli GF (2007). Meta-analysis: Vitamin D compounds in chronic kidney disease. *Ann Intern Med 147*, 840–853.
204. Martin KJ, Gonzalez EA, Gellens M, Hamm LL, Abboud H, Lindberg J (1998). 19-*nor*-1-Alpha-25-dihydroxyvitamin D2 (paricalcitol) safely and effectively reduces the levels of intact parathyroid hormone in patients on hemodialysis. *J Am Soc Nephrol 9*, 1427–1432.
205. Llach F, et al. (1998). Suppression of parathyroid hormone secretion in hemodialysis patients by a novel vitamin D analogue: 19-*nor*-1,25-Dihydroxyvitamin D2. *Am J Kidney Dis 32*(2, Suppl 2), S48–S54.
206. Sprague SM, Llach F, Amdahl M, Taccetta C, Batlle D (2003). Paricalcitol versus calcitriol in the treatment of secondary hyperparathyroidism. *Kidney Int 63*, 1483–1490.
207. Teng M, Wolf M, Lowrie E, Ofsthun N, Lazarus JM, Thadhani R (2003). Survival of patients undergoing hemodialysis with paricalcitol or calcitriol therapy. *N Engl J Med 349*, 446–456.
208. Tentori F, et al. (2006). Mortality risk among hemodialysis patients receiving different vitamin D analogs. *Kidney Int 70*, 1858–1865.

209. Ritzerfeld M, Klasser M, Mann H (2001). Alfacalcidol in the therapy of renal bone disease. *Int J Clin Pharmacol Ther 39*, 546–550.
210. Martin KJ, Gonzalez EA (2001). Vitamin D analogues for the management of secondary hyperparathyroidism. *Am J Kidney Dis 38*(5, Suppl 5), S34–S40.
211. (2003). K/DOQI clinical practice guidelines for bone metabolism and disease in chronic kidney disease. *Am J Kidney Dis 42*(Suppl 3), 1–201.
212. Belozeroff V, Klassen P, Wentworth C, Ofsthun N, Arab L, Rothman KJ (2005). Effects of uncontrolled biochemical parameters on cardiovascular mortality in dialysis patients. *Nephrol Dial Transpl 20*(Suppl 5), v103–v104.
213. Block GA, Hulbert-Shearon T, Levin NW, Port FK (1998). Association of serum phosphorus and calcium × phosphate product with mortality risk in chronic hemodialysis patients: A national study. *Am J Kidney Dis 31*, 607–617.
214. Brown EM, Gamba G, Riccardi D, Lombardi M, Butters R, Kifor O, Sun A, Hediger MA, Lytton J, Hebert SC (1993). Cloning and characterization of an extracellular $Ca^{(2+)}$-sensing receptor from bovine parathyroid. *Nature 66*, 575–580.
215. Nemeth EF (2006). Misconceptions about calcimimetics. *Ann NY Acad Sci 1068*, 471–476.
216. Tripathi O, Schreibmayer W, Tritthart HA (1993). Fendiline inhibits L-type calcium channels in guinea-pig ventricular myocytes: A whole-cell patch-clamp study. *Br J Pharmacol 108*, 865–869.
217. Coburn JW, Maung HM (2000). Calcimimetic agents and the calcium-sensing receptor. *Curr Opin Nephrol Hypertens 9*, 123–132.
218. Nemeth EF, et al. (2004). Pharmacodynamics of the type II calcimimetic compound cinacalcet HCl. *J Pharmacol Exp Ther 308*, 627–635.
219. Levi R, et al. (2006). Increased parathyroid hormone gene expression in secondary hyperparathyroidism of experimental uremia is reversed by calcimimetics: Correlation with posttranslational modifi cation of the trans acting factor AUF1. *J Am Soc Nephrol 17*, 107–112.
220. Kawata T, et al. (2006). Direct in vitro evidence of the suppressive effect of cinacalcet HCl on parathyroid hormone secretion in human parathyroid cells with pathologically reduced calcium-sensing receptor levels. *J Bone Min Metab 24*, 300–306.
221. Nemeth EF, Steffey ME, Hammerland LG, Hung BCP, Van Wagenen BC, Delmar EG, Balandrin MF (1998). Calcimimetic compound NPS R-568 stimulates calcitonin secretion but selectively targets parathyroid gland $Ca^{2+}$ receptor in rats. *Proc Natl Acad Sci 95*, 4040–4045.
222. Wada M, et al. (1998). NPS R-568 halts or reverses osteitis fibrosa in uremic rats. *Kidney Int 53*, 448–453.
223. Fox J, Lowe SH, Conklin RL, Petty BA, Nemeth EF (1999). Calcimimetic compound NPS R-568 stimulates calcitonin secretion but selectively targets parathyroid gland Ca21 receptor in rats. *J Pharmacol Exp Ther 290*, 480–486.
224. Igarashi T, Ogata E, Maruyama K, Fukuda T, Azuma J (2000). Effect of calcimimetic agent, KRN568, on gastrin secretion in healthy subjects. *Endocrine J 47*, 517–523.
225. Miedlich SU, Gama L, Seuwen K, Wolf RM, Breitwieser GE (2004). Homology modeling of the transmembrane domain of the human calcium sensing receptor and localization of an allosteric binding site. *J Biol Chem 279*, 7254–7263.
226. Hu J, Reyes-Cruz G, Chen W, Jacobson KA, Spiegel AM (2002). Identification of acidic residues in the extracellular loops of the seven-transmembrane domain of the human $Ca^{2+}$ receptor critical for response to $Ca^{2+}$ and a positive allosteric modulator. *J Biol Chem 277*, 46622–46631.
227. Moe SM, et al. (2005). Achieving NKF-K/DOQI bone metabolism and disease treatment goals with cinacalcet HCl. *Kidney Int 67*, 760–771.
228. Moe SM, et al. (2005). Long-term treatment of secondary hyperparathyroidism with the calcimimetic cinacalcet HCl. *Nephrol Dial Transpl 20*, 2186–2193.
229. Strippoli G, Palmer S, Tong A, Elder G, Messa P, Craig J (2006). Meta-analysis of biochemical and patient-level effects of calcimimetic therapy. *Am J Kidney Dis 47*, 715–726.
230. Cunningham J (2006). Effects of the calcimimetic cinacalcet HCl on cardiovascular disease, fracture, and health-related quality of life in secondary hyperparathyroidism. *Kidney Int 68*, 1793–1800.
231. Block GA, et al. (2004). Cinacalcet for secondary hyperparathyroidism in patients receiving hemodialysis. *New Engl J Med 350*, 1516–1525.
232. Lindberg JS, et al. (2005). Cinacalcet HCl, an oral calcimimetic agent for the treatment of secondary hyperparathyroidism in hemodialysis and peritoneal dialysis: A randomized, double-blind, multicenter study. *J Am Soc Nephrol 16*, 800–807.
233. Schaefer RM, et al. (2006). The SENSOR Study: A study to evaluate the efficacy of administering cinacalcet with the first meal after dialysis. *Nephrol Dial Transplant 21*, IV. 228.

# TARGETING CHEMOKINES AND ANGIOGENESIS IN RHEUMATOID ARTHRITIS*

Zoltán Szekanecz[1] and Alisa E. Koch[2,3]

[1]*Division of Rheumatology, Third Department of Medicine, University of Debrecen Medical and Health Sciences Center, Debrecen, Hungary*
[2]*Veterans' Administration, Ann Arbor Healthcare System, Ann Arbor, Michigan*
[3]*Department of Internal Medicine, Division of Rheumatology, University of Michigan Health System, Ann Arbor, Michigan*

## 1 INTRODUCTION

Inflammatory rheumatic diseases, such as rheumatoid arthritis (RA), are associated with leukocyte ingress into the synovial tissue. In arthritis, leukocytes extravasate through the vascular endothelium. The transendothelial migration of white blood cells involve numerous inflammatory chemokines and chemokine receptors [1–14]. Angiogenesis, the growth of new capillaries from pre-existing vessels, is perpetuated during the progression of inflammatory synovitis. A number of cell-surface-bound and soluble mediators may promote neovascularization [3, 6–9]. In addition, some chemokines are also involved in synovial angiogenesis [3, 10, 15]. On the other hand, some naturally released mediators, including chemokines as well as synthetic compounds, may suppress RA-associated blood vessel formation [3, 4, 6, 8, 9, 13, 16]. There is an imbalance between pro- and anti-inflammatory chemokines, as well as angiogenic and angiostatic mediators in favor of accelerated inflammation and neovascularization in RA [1–15]. Strategies of chemokine and/or angiogenesis blockade or the administration of angiostatic agents may control synovial leukocytic inflammation and thus the perpetuation of arthritis [1–16].

In this chapter, we will briefly review the role of chemokines, chemokine receptors, and angiogenesis in the pathogenesis of arthritis. We will draw special attention to chemokines involved in neovascularization. Finally, we will describe studies on chemokine and angiogenesis targeting obtained in animal models of arthritis as well as in humans.

## 2 CHEMOKINES AND CHEMOKINE RECEPTORS IN ARTHRITIS

Chemokines drive inflammatory leukocytes into the inflamed synovial tissue [1–5, 17–19]. Chemokines have been classified into the CXC, CC, C, and $CX_3C$ supergene families. Chemokines bind to their respective receptors, which have been assigned a designation of CXCR, CCR, CR, and $CX_3CR$ [1–3, 17, 20–22]. Each chemokine has its own traditional name (see later), however, they are also considered as CXCL, CCL, XCL, and $CX_3CL$ chemokine ligands. Both designations will

*This work was supported by NIH grants AR-048267 and AI-40987 (A.E.K.), the William D. Robinson, M.D., and G. L. Frederick. Huetwell Endowed Professorship (A.E.K.), funds from the Veterans' Administration (A.E.K.); and grant No T048541 from the National Scientific Research Fund (OTKA) (Z.S.).

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

**TABLE 1 Chemokines and Chemokine Receptors Relevant for Inflammatory Synovitis[a]**

| Chemokine Receptor | Chemokine Ligands |
|---|---|
| CXC chemokine receptors | |
| CXCR1 | IL-8/CXCL8, GCP-2/CXCL6 |
| CXCR2 | IL-8/CXCL8, ENA-78/CXCL5, groα/CXCL1, CTAP-III/CXCL7, GCP-2/CXCL6 |
| CXCR3 | IP-10/CXCL10, PF4/CXCL4, Mig/CXCL9 |
| CXCR4 | SDF-1/CXCL12 |
| CXCR5 | BCA-1/CXCL13 |
| CXCR6 | CXCL16 |
| CXCR7 | I-TAC/CXCL11, SDF-1/CXCL12 |
| C-C chemokine receptors | |
| CCR1 | MIP-1α/CCL3, RANTES/CCL5, MCP-3/CCL7, HCC-1/CCL14, HCC-2/CCL15, HCC-4/CCL16 |
| CCR2 | MCP-1/CCL2, MCP-3/CCL7, HCC-4/CCL16 |
| CCR3 | RANTES/CCL5, MCP-2/CCL8, MCP-3/CCL7, HCC-2/CCL15 |
| CCR4 | TARC/CCL17, CKLF1 |
| CCR5 | MIP-1α/CCL3, RANTES/CCL5, MCP-2/CCL8, HCC-1/CCL14 |
| CCR6 | MIP-3α/CCL20 |
| CCR7 | SLC/CCL21 |
| C chemokine receptors | |
| XCR1 | Lymphotactin/XCL1 |
| C-X3-C chemokine receptors | |
| CX3CR1 | Fractalkine/CX3CL1 |
| Other | |
| DARC | Duffy antigen, some CC and CXC chemokines |

[a]See text for abbreviations.

be used in this chapter. Apart from these structural differences, chemokines have also been functionally classified into homeostatic and inflammatory subclasses [5]. Although these functions may sometimes overlap, some chemokines are primarily involved in lymphocyte migration and lymphoid tissue organization, while others rather play a role in the pathogenesis of inflammatory diseases [4, 5] (Table 1).

### 2.1 CXC, CC and Other Chemokines in Arthritis

***CXC Chemokines*** CXC chemokines drive neutrophils as well as mononuclear cells into the synovial tissue [5, 17]. These inflammatory mediators also enhance leukocyte adhesion to other cells, cytoskeletal reorganization, as well as neutrophil functions [21]. Some CXC chemokines promote, while others rather suppress, synovial angiogenesis [3, 23]. As discussed later, the ELR amino acid sequence found in angiogenic chemokines may be important in synovial neovascularization [23].

Among CXC chemokines, IL-8/CXCL8, ENA-78/CXCL5, groα/CXCL1, CTAP-III/CXCL7, granulocyte chemotactic protein 2 (GCP-2)/CXCL6, IP-10/CXCL10, Mig/CXCL9, PF4/CXCL4, SDF-1/CXCL12, B-cell activating chemokine 1 (BCA-1)/CXCL13, and CXCL16 have been implicated in RA [1–5, 24–30] (Table 1).

Large amounts of IL-8/CXCL8, ENA-78/CXCL5, and groα/CXCL1 have been detected in the sera, synovial fluids, and synovial tissues of RA patients [24, 26, 27, 31–33]. Synovial macrophages are major producers of these chemokines [24, 26, 27, 33, 34]. All three CXC chemokines are inflammatory, express the ELR motif, and are proangiogenic [24, 26, 35]. Vascular endothelial cells express CXCR2, a major receptor for these chemokines [36].

CTAP-III/CXCL7 is produced by platelets, and it has been detected in RA sera and synovial tissue samples [37]. This inflammatory chemokine stimulates synovial matrix deposition, fibrosis, and angiogenesis [1, 37]. There is also abundant production of GCP-2/CXCL6 in RA [38].

IP-10/CXCL10, Mig/CXCL9, and PF4/CXCL4 exert inflammatory, but antiangiogenic, effects in RA [1, 23, 39–41]. These chemokines have been detected in the sera. Synovial macrophages and fibroblasts release these chemokines [39–41]. All three chemokines lack the ELR amino acid sequence in their structure, and they inhibit blood vessel formation in the synovium [23, 39, 41].

SDF-1/CXCL12 is in many ways a unique chemokine in RA. First, while many of other chemokines have common receptors, SDF-1/CXCL12 is a specific ligand for CXCR4. Moreover, SDF-1/CXCL12 is primarily a homeostatic chemokine involved in lymphoid organization. However, it is also involved in the development of synovial inflammation [28, 42, 43]. Synovial T cell–fibroblast interactions result in the upregulation of SDF-1/CXCL12 production. This mechanism involves IL-17 and CD40-CD40 ligand costimulation [44]. Finally, although this chemokine lacks the ELR motif, it still promotes synovial capillary formation [45]. SDF-1/CXCL12 binds to endothelial proteoglycans, and it is involved in integrin-mediated leukocyte-endothelial adhesion, osteoclastogenesis, and periarticular bone resorption [28, 46–48].

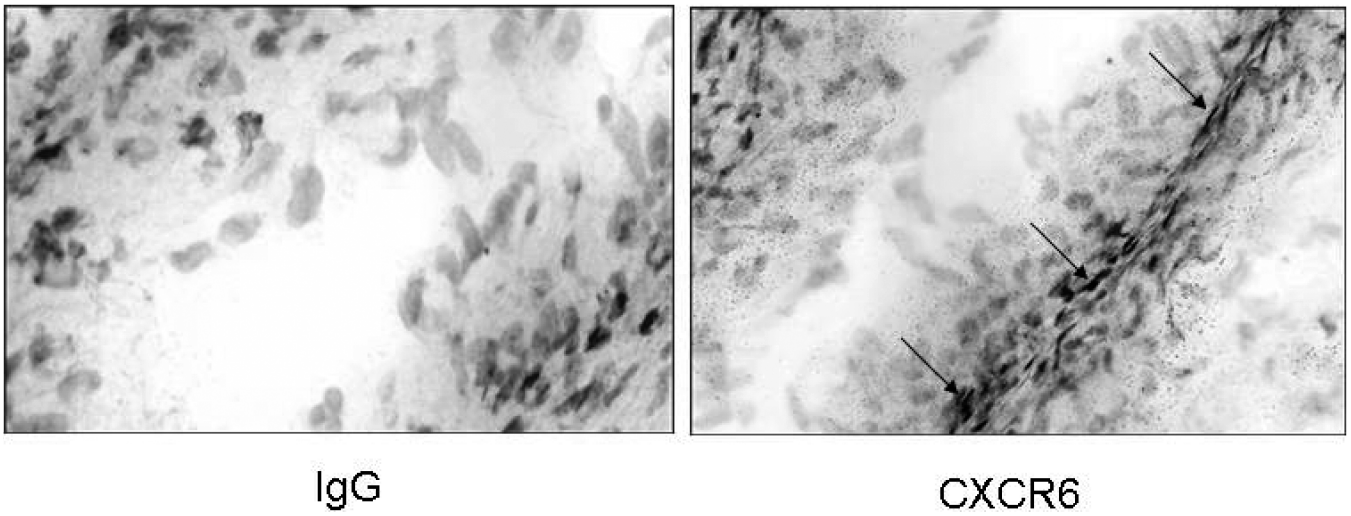


**Figure 1** Synovial endothelial cells show prominent expression of CXCR6. (Immunoperoxidase histochemistry, magnification 400×).

The crucial role of B cells in the pathogenesis of RA has been acknowledged, and numerous biologics targeting B cells have been developed recently [49–51]. Among homeostatic chemokines involved in B-cell migration and lymphoid tissue organization, some are also involved in inflammation, such as in the pathogenesis of RA. For example, BCA-1/CXCL13, the specific ligand for CXCR5, is involved in the formation of germinal centers [1, 52, 53]. On the other hand, this chemokine is also expressed by synovial fibroblasts, endothelial cells, and follicular dendritic cells within the RA synovium [54].

CXCL16, the specific ligand for CXCR6, has also been considered a homeostatic chemokine as it mediates lymphocyte recruitment and lymph node organization. In addition, large amounts of CXCL16 were detected in RA synovial tissues and fluids [29, 30]. Synovial macrophages and fibroblasts secrete CXCL16 [29, 30]. CXCL16-mediated leukocyte ingress into the synovium is mediated by the mitogen activated protein (MAP) kinase (MAPK) pathway [29].

***CC Chemokines*** Monocyte chemoattractant protein-1 (MCP-1)/CCL2, macrophage inflammatory protein 1α (MIP-1α)/CCL3, MIP-3α/CCL20, MCP-4/CCL13, regulated upon activation, normal T-cell expressed and secreted (RANTES)/CCL5, CCL18, Epstein–Barr virus-induced gene 1 ligand chemokine (ELC)/CCL19, secondary lymphoid tissue chemokine (SLC)/CCL21, thymus and activation regulated chemokine (TARC)/CCL17, pulmonary and activation regulated chemokine (PARC)/CCL18, and chemokine-like factor 1 (CKLF1) have been implicated in the pathogenesis of RA-associated inflammation [1–4, 54–74]. There is also some data available that MCP-2/CCL8, MCP-3/CCL7, hemofiltrate CC chemokine-1 (HCC-1)/CCL14, HCC-2/CCL15, and HCC-4/CCL16 may also be involved in the development of synovitis [57] (Table 1).

MCP-1/CCL2, MIP-1α/CCL3, and RANTES/CCL5 are chemotactic for monocytes, T cells, and natural killer (NK) cells [2, 58–60]. These chemokines have been detected in RA sera, synovial fluid, and tissue samples [33, 39, 58–60]. Tumor necrosis factor-α (TNF-α) and interleukin-1 (IL-1) further stimulate the release of MCP-1/CCL2 and MIP-1α/CCL3 [33, 59, 61]. MCP-1/CCL2 has been implicated in TNF-mediated osteoclast differentiation of peripheral blood monocytes. Thus, this chemokine may be involved in the development of bone erosions in arthritis [62]. A single nucleotide polymorphism in the RANTES promoter gene has been associated with disease susceptibility in Chinese RA patients [63]. Chemokine ligand 3-like 1 (CCL3L1) is encoded by a gene with variable copy numbers. Higher gene copy number of this chemokine has been associated with increased susceptibility to RA [64].

MIP-3α/CCL20, the specific ligand for CCR6, is chemotactic for mononuclear cells and has been detected in the RA synovium [1, 65]. This chemokine has been implicated in the recruitment of IL-17-producing $CCR6^+$

Th17 cells into the synovium [66]. MIP-3α/CCL2 may induce both osteoblast proliferation and osteoclast differentiation. Furthermore, increased expression of this chemokine was detected in the subchondral bone of RA patients. Thus, MIP-3α/CCL20 may exert a special role in RA-associated bone metabolism as this chemokine and the receptor activated nuclear factor kappa (RANK) ligand system collaborate in the uncoupling between new bone formation and bone resorption [67].

CCL18 facilitates T-cell attraction by antigen-presenting cells. Circulating levels of CCL18 is increased in RA patients in comparison to controls and correlate with disease activity in RA [56]. Synovial fluid neutrophils release CCL18 during their recruitment into the joints [68].

MCP-4/CCL13 is expressed in the cartilage of the RA joint. IFN-γ (interferon gamma) in combination with TNF-α and IL-1 stimulate the release of this chemokine from arthritic chondrocytes. MCP-4/CCL13 enhances the proliferation of synovial fibroblasts [69].

Some primarily homeostatic CC chemokines have also been implicated in the pathogenesis of synovial inflammation. ELC/CCL19 has been detected in RA synovial tissues [70]. The release of SLC/CCL21 is associated with the formation of germinal centerlike structures during lymphoid neogenesis in the RA synovium [54, 71]. TARC/CCL17 and PARC/CCL18 have also been implicated in T-cell recruitment into the RA synovium [72–74]. These chemokines have been detected in RA sera and synovia [72–74].

In one study, MCP-2/CCL8, MCP-3/CCL7, HCC-1/CCL14, HCC-2/CCL15, and HCC-4/CCL16 were detected in the RA synovium [57]. The function of these chemokines in RA needs to be confirmed by further studies. CKLF1 is a functional ligand of CCR4. Its expression is upregulated on activated $CD4^+$ and $CD8^+$ T cells but not on B cells in RA [55].

***C and $CX_3C$ Chemokines*** The C chemokine family contains lymphotactin/XCL1 and single C motif 1β (SCM-1β)/XCL2. The $CX_3C$ subset contains a single member, fractalkine/$CX_3CL1$ [2,3,75,76]. Lymphotactin/XCL1 involved in T-cell chemotaxis has been detected on $CD8^+$ and $CD4^+/CD28^-$ T cells in RA [77].

Fractalkine/$CX_3CL1$ is chemotactic for monocytes and lymphocytes, and it also serves as a $CD4^+$ T-cell adhesion molecule [75, 78, 79]. This chemokine induces T-cell-dependent proliferation of synovial fibroblasts [80]. It regulates RA synovial fibroblast cytoskeletal structure and migration. These processes involve JUN NH2-terminated kinase (JNK), ERK-1/2, and Akt activation [81]. In the RA synovium, macrophages, fibroblasts, and endothelial and dendritic cells produce fractalkine/$CX_3CL1$ [78, 79]. Fractalkine/$CX_3CL1$ has been associated with disease activity in rheumatoid vasculitis [82]. Interestingly, fractalkine/$CX_3CL1$ may also be involved in accelerated atherosclerosis associated with RA [83–85].

### 2.2 Chemokine Receptors in Rheumatoid Arthritis

These 7-transmembrane domain receptors are expressed on the target cells (PPT/8). Some chemokine receptors, such as CXCR2, CCR1, or CCR3, have numerous chemokine ligands, while others including CXCR4, CXCR5, CXCR6, CCR8, or CCR9 are specific receptors for a single ligand [1, 2, 17] (Table 1).

Certain chemokine receptor expression patterns have been correlated with various functions in autoimmune inflammation. For example, Th1-type inflammation in RA may be primarily associated with CXCR3 and CCR5. In contrast, asthma, a known Th2-type disease, rather involves CCR3, CCR4, and CCR8 [17, 78, 86, 87]. Regarding chemokine-induced angiogenesis, CXCR2, a receptor recognizing most $ELR^+$ CXC chemokines, is involved in inflammatory angiogenesis, while CXCR3, a receptor for most $ELR^-$ CXC chemokines may rather mediate angiostasis [1, 2, 17, 23].

Among CXC chemokine receptors, CXCR1 and CXCR2 are abundantly expressed in the RA synovium, where they recognize the most relevant inflammatory and angiogenic CXC chemokines described above [1, 2]. CXCR3 may be the most important receptor in leukocyte homing into the RA synovium [78, 86, 88]. CXCR4 is involved in SDF-1/CXCL12-derived retention of lymphocytes within the RA synovial tissue [28]. CXCR5 is expressed by T and B cells, macrophages, and endothelial cells in the RA synovium [89]. CXCR5 is involved in lymphoid neogenesis associated with antigen-induced arthritis in mice [90]. CXCR6 mediates CXCL16-induced inflammation [29]. CXCR4, CXCR5, and CXCR6 bind their respective homeostatic chemokine ligands, SDF-1/CXCL12, BCA-1/CXCL13, and CXCL16. Thus, these CXC chemokine receptors are involved in both homeostatic and inflammatory processes [1, 5, 28–30, 52].

CCR1 and CCR5 are abundantly expressed in the RA synovium [57, 78, 86, 91]. CCR1, CCR2, CCR3, and CCR5 are also expressed by articular chondrocytes [92]. CCR4 has also been implicated in leukocyte ingress into the RA joint [93]. CCR5 density on the surface of T cells may determine the intensity of synovitis in RA [94]. In some studies, a genetic polymorphism leading to the production of the truncated Δ32-CCR5 nonfunctional receptor allele was found to be protective against RA [95]. However, this protective role could not be confirmed in national cohorts [96, 97], suggesting geograph-

ical differences. CCR6, the single receptor for MIP-3α/CCL20, has also been detected on RA synovial leukocytes [65]. As mentioned above, CCR6 expressed by Th17 cells is involved in the ingress of these arthritogenic cells into the joint [66]. CCR7 has been associated with synovial lymphoid neogenesis in mice [90]. A putative chemokine receptor, CCR-like receptor 2 (CCRL2), has been identified on RA synovial fluid neutrophils and macrophages [98].

Regarding the C and $CX_3C$ chemokine receptors, XCR1 is expressed on RA synovial lymphocytes, macrophages, and fibroblasts [1, 2], while $CX_3CR1$ has been detected on macrophages and dendritic cells [76]. The expression of $CX_3CR1$ on $CD4^+/CD28^-$ T cells correlated with premature vascular damage in RA [85]. The Duffy blood group antigen receptor (DARC), which is expressed by RA synovial endothelial cells, also binds some CXC and CC chemokines [99].

### 2.3 Inflammatory and Homeostatic Chemokines in RA

*Homeostatic chemokines* are constitutively produced in lymphoid or nonlymphoid tissues. The RA synovium also contains lymphoid aggregates and germinal center-like structures. Thus, the synovium, in many ways, is similar to the skin- and mucosa-associated lymphoid tissues [5, 52–54]. Some otherwise homeostatic chemokines may also be involved in synovial inflammation [5, 52–54]. These chemokines with overlapping functions include, as discussed above, SDF-1/CXCL12, BCA-1/CXCL13, CXCL16, TARC/CCL17, PARC/CCL18, ELC/CCL19, SLC/CCL21, and maybe others [1, 5, 25, 28–30, 45, 54, 70].

Yet, most CXC and CC chemokines, as well as all C and $CX_3C$ chemokines, implicated in the pathogenesis of RA are primarily *inflammatory chemokines* [1, 2, 52]. Among these mediators, IL-8/CXCL8, ENA-78/CXCL5, groα/CXCL1, CTAP-III/CXCL7, IP-10/CXCL10, Mig/CXCL9, PF4/CXCL4, GCP-2/CXCL6, MCP-1/CCL2, MIP-1α/CCL3, MIP-3α/CCL20, RANTES/CCL5, lymphotactin/XCL1, and fractalkine/$CX_3CL1$, as well as their respective receptors, have been implicated in leukocyte recruitment during the evolution of synovitis [1–4, 54, 55].

## 3 ANGIOGENESIS IN ARTHRITIS

### 3.1 Process of Neovascularization

RA is considered an "angiogenic" disease state as there is a perpetuation of neovascularization in the synovium associated with synovial inflammation [3, 6–10]. The angiogenic process itself, its mediators and inhibitors, cellular and molecular interactions underlying neovascularization, as well as the role of angiogenesis have been extensively discussed in recent reviews [3, 6–10]. The RA synovial tissue is rich in newly formed vessels. The high turnover of capillary formation leading to increased vascular endothelial surface may enable the extravasation of inflammatory leukocytes into the synovium [3, 6–8, 10].

During the neovascularization process, angiogenic mediators activate endothelial cells in pre-existing vessels. Activated endothelial cells then release matrix-degrading enzymes, which digest the underlying basal membrane and the extracellular matrix. Matrix degradation enables the emigration, proliferation, migration, and sprouting of endothelial cells. Lumen formation within primary and further generation endothelial sprouts will lead to the development and outgrowth of new capillaries [3, 10, 11, 100, 101].

### 3.2 Angiogenic Mediators

Angiogenic mediators also involved in the pathogenesis of RA include numerous growth factors, cytokines, chemokines, extracellular matrix components, cell adhesion receptors, proteolytic enzymes, and other molecules. These mediators are released by synovial endothelial cells and macrophages [6, 7, 11, 34] (Table 2).

***Hypoxia–VEGF–Angiopoietin Pathway*** *Vascular endothelial growth factor* (VEGF) binds to heparin in the synovial matrix and plays an essential role in the induction, maintenance, and regulation of angiogenesis [15]. *Hypoxia* has been intra-articularly detected within RA joints [102], and it stimulates VEGF production in arthritis [6, 15, 103]. Hypoxia acts through the induction of the hypoxia-inducible factor heterodimer, HIF-1α/HIF-1β [12, 104]. Among other inflammatory mediators, hepatocyte growth factor (HGF), epidermal growth factor (EGF), prostaglandin $E_2$ ($PGE_2$), and nitric oxide (NO) may also stimulate VEGF release [6, 15, 16].

*Angiopoietins*, including Ang1 and Ang2, regulate endothelial functions upon stimulation by VEGF. Both Ang1 and Ang2 interact with the endothelial tyrosine kinase receptor Tie2 [14, 15]. Interactions between Ang1 and Tie2 lead to the stabilization of blood vessels [105]. Ang2 antagonizes the effects of Ang1, promotes endothelial cell motility, while inhibiting the stabilization and maturation of new capillaries [13, 15]. Survivin is an inhibitor of apoptosis, has also been implicated in VEGF-dependent neovascularization, and endothelial cell survival [15, 106]. VEGF, HIF-1, Ang1, Tie2, and

**TABLE 2 Most Important Angiogenic and Angiostatic Molecules in Arthritis[a]**

| Class | Mediators | Inhibitors |
|---|---|---|
| Growth factors | VEGF, HGF, EGF, HIF-1, PDGF, FGF-1, FGF-2, IGF-I, TGF-β | — |
| Pro-inflammatory cytokines | TNF-α, IL-1, IL-6, IL-15, IL-17, IL-18, G-CSF, GM-CSF, oncostatin M, MIF | IFN-α, IFN-γ, IL-4, IL-12, IL-13, LIF |
| Chemokines | $ELR^+$ CXC chemokines, SDF-1/CXCL12, MCP-1/CCL2, fractalkine/CXC3CL1 | $ELR^-$ CXC chemokines |
| Extracellular matrix components | Collagen, laminin, fibronectin, vitronectin tenascin | Thrombospondin 1, thrombospondin 2 |
| Cell adhesion molecules | $\alpha_V\beta_3$ integrin, E-selectin, VCAM-1, endoglin, MUC18, $Lewis^y$/H, PECAM-1 | — |
| Proteases | MMPs, plasminogen activators | TIMPs, PAIs |
| Environmental factors | Hypoxia | — |
| Vessel stabilizers, apoptosis inhibitors | Ang1-Tie2, survivin | — |
| Others | Prostaglandin E2, angiogenin, angiotropin, PAF, prolactin, substance P, endothelin-1, serum amyloid A, histamine, adenosine, erythropoietin, thrombin, etc. | DMARDs, antibiotics, thalidomide, 2-methoxyestradiol, angiostatin, endostatin, paclitaxel, troponin, osteonectin, chondromodulin, etc. |

[a]See text for abbreviations.

survivin have been detected in the RA synovial tissue [15, 107–110].

***Other Angiogenic Mediators*** Other growth factors involved in angiogenesis include fibroblast growth factor (FGF)-1, FGF-2, human growth factor (HGF), platelet-derived growth factor (PDGF), EGF, insulin-like growth factor-I (IGF-I), and TGF-β [3, 6, 7, 10]. Among pro-inflammatory cytokines, TNF-α, IL-1, IL-6, IL-15, IL-17, IL-18, granulocyte (G-CSF), granulocyte-macrophage colony-stimulating factor (GM-CSF), oncostatin M, and macrophage migration inhibitory factor (MIF) also promote synovial neovascularization [3, 6, 7, 111–119]. TNF-α may also act via the VEGF–Ang1–Tie2 system [120], while IL-18 rather stimulates the release of angiogenic SDF-1/CXCL12, MCP-1/CCL2, and VEGF by synovial fibroblasts [116]. MIF acts via the upregulation of VEGF and IL-8/CXCL8 production [118].

Some *extracellular matrix constituents*, *cell adhesion molecules*, and matrix-degrading *proteolytic enzymes* are also involved in endothelial cell adhesion, migration, and angiogenesis. These angiogenic factors include type I collagen, fibronectin, laminin, tenascin, $\beta_1$ and $\alpha_V\beta_3$ integrins, E-selectin, selectin-related glycoconjugates including $Lewis^y$/H and melanoma cell adhesion molecule (MUC18), vascular cell adhesion molecule 1 (VCAM-1), platelet-endothelial adhesion molecule 1 (PECAM-1; CD31), endoglin, as well as some matrix metalloproteinases (MMPs), and plasminogen activators [6, 10, 34, 100, 121–123]. The $\alpha_V\beta_3$ integrin has recently drawn a lot of attention in relation to capillary formation [123]. The $\alpha_V$ subunit of this integrin is encoded by the *ITGAV* (integrin alpha-V) gene. A recent genome-wide study confirmed the association of ITGAV with susceptibility to RA [124]. The focal adhesion kinases (FAK), which are involved in signaling through the $\alpha_V\beta_3$ integrin, have been implicated in synovitis and RA-associated angiogenesis [125].

Among other important angiogenic mediators, endothelin-1, also produced by endothelial cells, induces VEGF production, endothelial proliferation, and angiogenesis [126, 127]. Increased levels of endothelin-1 have been detected in RA sera and synovia [126, 127]. Serum amyloid A (SAA), a major acute-phase reactant, stimulates synovial hyperplasia, endothelial cell migration, and capillary formation [128]. Other angiogenic factors not mentioned above in detail include $PGE_2$, angiogenin, angiotropin, platelet-activating factor (PAF), histamine, substance P, adenosine, erythropoietin, prolactin, thrombin, and many others [6, 7, 10].

### 3.3 Angiogenesis Inhibitors

Endogenous inhibitors of neovascularization suppress the effects of angiogenic mediators discussed above (Table 2). These inhibitors include cytokines, such

as IFN-α, IFN-γ, IL-4, IL-12, IL-13, and leukemia inhibitory factor (LIF) [6, 103, 129]. Tissue inhibitors of metalloproteinases (TIMP) and plasminogen activator inhibitors (PAI) antagonize the effects of angiogenic MMPs and plasminogen activators, respectively [3, 6, 7, 15, 34]. ELR-lacking angiostatic chemokines will be discussed later. Thrombospondin-1 (TSP1) blocks the activity of heparin-binding growth factors [3, 6, 7, 15]. 2-Methoxyestradiol (2ME2), a natural estrogen metabolite, inhibits neovascularization by disrupting microtubules and by suppressing HIF-1α activity [130]. 2ME2 has anti-arthritic properties in the rat adjuvant induced arthritis (AIA) model [131].

Other angiostatic compounds not yet mentioned include angiostatin (a fragment of plasminogen), endostatin (a fragment of type XIII collagen), paclitaxel, osteonectin, opioids, troponin I, and chondromodulin-1 [3, 6, 7, 103]. As angiostatin and endostatin have already been tested in therapeutic studies, they will be discussed in more detail later.

## 4 CHEMOKINES AND CHEMOKINE RECEPTORS IN ANGIOGENESIS

### 4.1 Angiogenic Chemokines and Chemokine Receptors

In general, CXC chemokines containing the ELR amino acid residue, such as IL-8/CXCL8, ENA-78/CXCL5, groα/CXCL1, and CTAP-III/CXCL7, stimulate, while $ELR^-$ CXC chemokines, such as PF4/CXCL4, IP-10/CXCL10, and Mig/CXCL9, rather inhibit neovascularization [3, 23]. As discussed above, SDF-1/CXCL12 lacks the ELR sequence; nevertheless, it still stimulates neovascularization [3, 23].

IL-8/CXCL8 is chemotactic and mitogenic for vascular endothelial cells [3, 21, 23]. ENA-78/CXCL5, CTAP-III/CXCL7, and groα/CXCL1 have also been shown to promote neovascularization [3, 6, 21, 37, 132]. The angiogenic effects of prostaglandin $E_2$ are also mediated, in part, by groα/CXCL1 [133].

SDF-1/CXCL12, despite lacking the ELR sequence, induces endothelial cell chemotaxis and angiogenesis [45, 134, 135]. SDF-1/CXCL12 acts in close cooperation with the hypoxia-hypoxia-inducible factor (HIF)-VEGF pathway. Hypoxia induces the release of this chemokine by RA synovial fibroblasts [45]. SDF-1/CXCL12 may induce angiogenesis indirectly via VEGF as this chemokine stimulates VEGF release via the phosphatidyl inositol 3 kinase (PI3K)/Akt pathway [136]. SDF-1/CXCL12 expression has also been associated with tumor neovascularization, and its expression may serve as a prognostic marker in gliomas [137].

Little is known of the angiogenic nature of CC chemokines. MCP-1/CCL2 induces endothelial cell chemotaxis and vessel formation via the CCR2 endothelial receptor [36, 138]. The angiogenic effects of MCP-1/CCL2 involve the upregulation of the Ets-1 transcription factor, as well as integrin and ERK-1/2 activation [138]. Myeloid progenitor inhibitory factor 1 (MPIF-1)/CCL23 has also been implicated in endothelial cell migration and metalloproteinase secretion [139].

As discussed above, fractalkine/$CX_3CL1$ has been implicated in angiogenesis and atherosclerosis associated with the progression of RA [76, 83, 84, 140]. $CX_3CR1$-deficient mice develop a lesser degree of atherosclerosis than do wild-type animals [140]. Moreover, an M280/I249 polymorphism in the $CX_3CR1$ gene has been associated with decreased cardiovascular risk in humans [83]. These observations may not only be important for the pathogenesis of RA, but also for understanding the mechanisms of accelerated atherosclerosis and vascular disease, leading mortality factors in RA and other inflammatory diseases.

*Chemokine receptors*, recognizing the most relevant angiogenic chemokine ligands described above, exert inflammatory and angiogenic properties in RA. For example, CXCR2, a receptor for angiogenic, $ELR^+$ CXC chemokines, is expressed on endothelial cells as well as macrophages and neutrophils in the RA synovial tissue [3, 36, 92]. As discussed above, CXCR4 is involved in SDF-1/CXCL12-mediated synovial neovascularization [28, 141]. Hypoxia induces CXCR4 expression via HIF-1 and VEGF production [142]. CXCR7, recognizing I-TAC/CXCL11 and SDF-1/CXCL12, may also be involved in synovial angiogenesis [143].

CCR2 is a receptor for the angiogenic MCP-1/CCL2 [3, 36]. In a murine model of skeletal muscle injury, CCR2-deficient animals exerted delayed muscular angiogenesis and decreased VEGF production [144]. DARC, a chemokine-binding receptor of the Duffy antigen, has been implicated in breast-cancer-associated neovascularization [145].

### 4.2 Angiostatic Chemokines and Chemokine Receptors

IP-10/CXCL10 inhibits VEGF-dependent endothelial cell migration, as well as angiogenesis [3, 21, 23, 146, 147]. On the other hand, the angiogenic VEGF induces endothelial expression of IP-10/CXCL10 [147]. Thus, IP-10/CXCL10 and VEGF may form an autocrine loop, which is involved in the up- and downregulation of capillary formation [146, 147]. Mig/CXCL9 and PF4/CXCL4 are also angiostatic [3, 23, 148]. The PF4var/CXCL4L1 nonallelic variant of the corresponding chemokine also inhibits neovascularization, as well as melanoma and

lung carcinoma proliferation [149, 150]. Among CC chemokines, SLC/CCL21 exerts angiostatic effects and inhibits tumor progression [151].

In contrast to chemokine receptors that promote chemokine-medited neovascularization, CXCR3, which binds the angiostatic chemokines IP-10/CXCL10 and Mig/CXCL9, may rather be involved in chemokine-mediated angiogenesis inhibition [3].

## 5 TARGETING OF CHEMOKINES AND CHEMOKINE RECEPTORS

Chemokines and chemokine receptors may be targeted by nonspecific strategies and by chemokine-specific approaches. Here, in vitro and animal studies, as well as a limited number of published human trials will be presented [1–4, 152] (Table 3).

**TABLE 3 Chemokine and Chemokine Receptor Targeting Strategies in RA[a]**

| Compound | Chemokine/Chemokine Receptor | Animal (A)/ Human (H) Study[b] | Reference |
|---|---|---|---|
| | *Nonspecific Blockade* | | |
| Dexamethasone | IL-8/CXCL8, MCP-1/CCL2 | H | 153 |
| NSAIDs | IL-8/CXCL8 | A | 154 |
| Sulfasalazine | IL-8/CXCL8, groα/CXCL1, MCP-1/CCL2 | H | 155 |
| Methotrexate | CCR2 | H | 157 |
| Methotrexate + leflunomide | MCP-1/CCL2 | H | 156 |
| Infliximab | IL-8/CXCL8, RANTES/CCL5, MCP-1/CCL2, groα/CXCL1, CXCL16 | H | 158–165 |
| Etanercept | CXCR3 | H | 163 |
| Antioxidants | IL-8/CXCL8, MCP-1/CCL2 | H | 166 |
| Simvastatin | IL-8/CXCL8 | H | 167 |
| Triptolide | MCP-1/CCL2, RANTES/CCL5 | A | 168 |
| Epigallocatechin-3-gallate | ENA-78/CXCL5, groα/CXCL1, RANTES/CCL5 | H | 169 |
| ML3000 | Míg/CXCL9, IP-10/CXCL10, ITAC/CXCL11 | H | 170 |
| PPARγ agonists | MCP-1/CCL2 | H | 62 |
| | *Specific Blockade* | | |
| Anti-ENA-78/CXCL5 (preventative) | ENA-78/CXCL5 | A | 172 |
| Anti-groα/CXCL1 (preventative) | groα/CXCL1 | A | 173 |
| PF4/CXCL4 peptide | PF4/CXCL4 | A | 25 |
| Anti-CXCL16 | CXCL16 | A | 30 |
| Anti-IL-8/CXCL8 | IL-8/CXCL8 | A | 171 |
| Anti-MIP-1α/CCL3 | MIP-1 α/CCL3 | A | 173 |
| Anti-MCP-1/CCL2 | MCP-1/CCL2 | A | 174–176 |
| Anti-RANTES/CCL5 | RANTES/CCL5 | A | 177 |
| KE-298 | MCP-1/CCL2 + RANTES/CCL5 | A | 178 |
| ABN912 | MCP-1/CCL2 | H | 200 |
| Antifractalkine/CX3CL1 | Fractalkine/CX3CL1 | A | 179 |
| Combined blockade | Various chemokines together | A | 180, 181 |
| CXCR2 antagonist | CXCR2 | A | 182 |
| DF2162 | CXCR1/2 | A | 183, 184 |
| Anti-CXCR3 | CXCR3 | A | 185 |
| AMD3100 | CXCR4 | A | 187 |
| FC131, other T140 analogs | CXCR4 | A | 188 |
| J-113863, other CCR1 antagonists | CCR1 | A | 189–194 |
| MetRANTES | CCR1 + CCR5 | A | 195, 196 |
| Maraviroc | CCR5 | H | 202 |
| CCR5 antagonists | CCR5 | A | 197 |
| Anti-CCR2 antibody | CCR2 | A | 199 |
| CP-481,715 | CCR1 | H | 190, 201 |

[a]See text for abbreviations.
[b]Human studies include both in vitro studies with isolated cells or histological samples, as well as in vivo studies.

### 5.1 Nonspecific Approaches

A number of antirheumatic drugs, traditional disease-modifying agents (DMARD) and biologics may influence chemokine production. Corticosteroids, such as dexamethasone, inhibit IL-8/CXCL8 and MCP-1/CCL2 production in RA patients [153]. Diclofenac and meloxicam attenuated IL-8/CXCL8 production in the rat antigen-induced arthritis (AgIA) model [154]. Sulfasalazine inhibited the release of IL-8/CXCL8, groα/CXCL1, and MCP-1/CCL2 in RA synovial explant cultures [155]. Combination of methotrexate and leflunomide decreased MCP-1/CCL2 expression in the RA synovium [156]. Methotrexate also suppressed the expression of CCR2 on RA peripheral blood monocytes, which was associated with lower disease activity [157]. Infliximab inhibited the expression of IL-8/CXCL8, RANTES/CCL5, and MCP-1/CCL2 in RA sera and synovial tissues [158, 159]. This anti-TNF agent also suppressed groα/CXCL1 and CXCL16 production in RA [160, 161]. Infliximab reduced CCR3 and CCR5 expression on RA T cells [162]. Treatment of RA patients with infliximab or etanercept resulted in the clearance of $CXCR3^+$ T cells from the synovium [163]. Anti-TNF biologics also suppressed CXCL16 expression on synovial macrophages [164]. Infliximab reduced the secretion of IL-8/CXCL8, MIP-1α/CCL3, and MCP-1/CCL2 in response to mycobacteria. Thus, the development of tuberculosis in infliximab-treated RA patients may be related, in part, to the inhibition of TNF-dependent chemokine gradients and impaired leukocyte migration [165].

Among other nonspecific non-DMARD inhibitors, antioxidants, such as *N*-acetyl-L-cysteine and 2-oxothiazolidine-4-carboxylate, inhibited the expression of IL-8/CXCL8 and MCP-1/CCL2 essenger ribonucleic acid (mRNA) by activated human synovial fibroblasts [166]. Simvastatin inhibited the release of IL-8/CXCL8 by TNF-α-stimulated RA synovial fibroblasts [167]. Triptolide, a diterpenoid triepoxide with potent anti-inflammatory effects, inhibited MCP-1/CCL2, MIP-1α/CCL3, and RANTES/CCL5 production in the rat adjuvant-induced arthritis (AIA) model [168]. Epigallocatechin-3-gallate (EGCG), an anti-inflammatory compound derived from green tea, suppressed IL-1-induced ENA-78/CXCL5, groα/CXCL1, and RANTES/CCL5 production by RA synovial fibroblasts [169]. The dual cyclooxygenase-lipoxygenase inhibitor, ML3000, downregulates the expression of CXCR3 ligands including Mig/CXCL9, IP-10/CXCL10, and ITAC/CXCL11 on RA synovial fibroblasts [170]. Activation of peroxisome proliferator-activated receptor γ (PPARγ) suppresses MCP-1/CCL2 expression in monocytes [62]. Thus, PPARγ agonists, such as glitazones, may inhibit chemokine production.

### 5.2 Specific Targeting

***Animal Models*** Neutralizing antibodies to IL-8/CXCL8 prevented arthritis in rabbits [171]. A neutralizing polyclonal anti-ENA-78/CXCL5 antibody was administered intravenously to rats using the AIA model. The antibody prevented the onset of the disease, however, it failed to control synovitis when administered therapeutically [172]. The preventative administration of an anti-groα/CXCL1 antibody delayed the onset and severity of collagen-induced arthritis (CIA) in mice [173]. A bioactive synthetic peptide derived from *PF4/CXCL4* inhibited the development of murine CIA [25]. An antibody to CXCL16 reduced the clinical arthritis score and suppressed joint destruction in murine CIA [30]. Passive immunization of mice with anti-MIP-1α/CCL3 decreased the severity of murine CIA [173]. A monoclonal antibody to MCP-1/CCL2 reduced ankle swelling and synovitis in rat CIA [174]. A novel inhibitor of endogenous MCP-1/CCL2, p8A-MCP-1, reduced cytokine expression, leukocyte infiltration, bone erosion, and improved clinical signs of arthritis in the rat AIA model [175]. A peptide inhibitor of MCP-1/CCL2 suppressed the development of arthritis in MRL-lpr mice [176]. An anti-RANTES/CCL5 antibody inhibited the progression of murine CIA [177]. KE-298, a combined MCP-1/CCL2 and RANTES/CCL5 inhibitor, attenuated the severity of rat AIA [178]. An antibody to fractalkine/$CX_3CL1$ decreased clinical severity and joint destruction in murine CIA [179].

Combined chemokine blockade has also been employed in some studies. A combination of MCP-1/CCL2 and groα/CXCL1 inhibition resulted in a greater extent of arthritis suppression than MCP-1/CCL2 blockade alone in murine AIA [180]. In the rabbit endotoxin-induced arthritis model, the combination of anti-IL-8/CXCL8 and anti-groα/CXCL1 antibodies resulted in a more pronounced inhibition of knee arthritis than did any of the two antibodies alone [181]. Certainly, an increased toxicity using combined strategies may be an important issue in the future [1].

Regarding chemokine receptor targeting, a nonpeptide oral antagonist of the CXCR2 receptor inhibited IL-8/CXCL-induced arthritis in rabbits [182]. DF2162, an allosteric non-competitive inhibitor of CXCR1/CXCR2 diminished neutrophil influx in mice and ameliorated AIA in rats [183, 184]. In the AIA model, an anti-CXCR3 antibody inhibited T-cell recruitment and disease progression [185]. Small-molecule CXCR4

antagonists gave promising results in arthritis studies [186]. Among these compounds, AMD3100 inhibited CIA in IFN-γ-deficient mice [187]. Analogs of the 14-mer peptide T140, acting as CXCR4 inhibitors, ameliorated CIA in mice [188]. Numerous CCR1 and CCR2 antagonists have been developed, and preclinical studies with chemokine receptor-deficient mice have been initiated in recent years [189–194]. For example, J-113863, a small-molecule CCR1 antagonist attenuated synovitis and joint destruction in murine CIA [194]. Met-RANTES, a CCR1/CCR5 antagonist, inhibited both murine CIA and rat AIA [195, 196]. A nonpeptide CCR5 antagonist preventatively inhibited mouse CIA [197].

Studies using chemokine receptor-deficient animals or inhibitory compounds may give different results, so early studies in this context may be difficult to interpret. For example, while a nonpeptide CCR1 antagonist improved murine CIA, CCR1-deficient mice produced more TNF-α than controls [194]. Anti-CCR2 antibody administered during the initiation of murine CIA improved the clinical symptoms [192], while CCR2-deficient animals developed more severe CIA than did wild-type mice [192, 193, 198]. The MC-21 monoclonal antibody raised against CCR2 efficiently depleted $CCR2^+$ monocytes from the blood and synovial tissue of mice with CIA. Yet, while low doses of MC-21 caused marked improvement of arthritis, high doses of this antibody rather had pro-inflammatory effects [199]. Yet, CCR blockade using antibodies or synthetic inhibitors may be promising for future therapies.

***Human Studies*** There is only a limited number of published human antichemokine or antichemokine receptor trials. ABN912, a monoclonal antibody to MCP-1/CCL2, has been tested in a randomized, controlled human RA trial. In this study, 33 patients received the active compound, and 12 received a placebo. Serial arthroscopic biopsies were performed. ABN912 treatment was well tolerated, but there was no detectable clinical benefit or significant change in synovial biomarkers [200].

A CCR1 antagonist has been tried in a 2-week phase Ib study. This inhibitor decreased the number of synovial macrophages. About one-third of the patients also fulfilled the ACR20% criteria for clinical improvement [191]. CP-481,715, another CCR1 antagonist, inhibited monocyte chemotactic activity present in human RA synovial fluid samples [190]. This compound has been evaluated in phase I for pharmacokinetics and safety [201]. Some CCR2 inhibitors have also entered clinical trials [192]. Among CCR5 inhibitors, maraviroc has been introduced to phase II–III trials in human immunodeficiency virus (HIV) infection and acquired immunodeficiency syndrome (AIDS), as well as to phase II trials in RA [202].

## 6 TARGETING OF ANGIOGENESIS

With respect to angiogenesis and its targeting, chronic synovitis and malignancies have numerous similarities. Therefore, angiostatic compounds that undergo cancer clinical trials may soon be also tried in RA [104]. There may be two major strategies to control the perpetuation of neovascularization in "angiogenic diseases," such as RA, as well as malignancies [6, 7, 10, 104]. First, endogenous angiogenesis inhibitors described above are naturally produced in the RA synovium; however, these inhibitors should be administered in excess in order to counterbalance intensive capillary formation observed in arthritis. On the other hand, numerous small molecular synthetic compounds, as well as biologics, currently used to treat RA may, among other effects, block neovascularization as well. For example, chemokine and chemokine receptor inhibitors discussed in the previous section may also inhibit chemokine-induced angiogenesis [3, 6, 7, 103] (Table 4).

### 6.1 Animal Models

***Endogenous Inhibitors*** There has been a lot of attention regarding angiostatin and endostatin. These extracellular matrix derivatives inhibit $\alpha_v\beta_3$ integrin-dependent angiogenesis [15, 16]. Both molecules suppressed the development of arthritis in various models [3, 6, 103, 104, 203]. Angiostatin gene transfer inhibited synovitis in the mouse CIA model [104, 203]. Protease-activated kringles 1–5 (K1–5), an angiostatin analog, suppressed CIA more potently than angiostatin itself [204]. Endostatin interferes with VEGF receptor 2 signaling [15, 205]. Endostatin suppressed pannus formation and joint destruction in various rodent models [15, 206]. Intra-articularly administered kallistatin gene suppressed arthritis of the ankles in rats [207]. Arresten, canstatin, and tumstatin are derived from type IV collagen. All these molecules inhibit neovascularization [208].

Thrombospondins (TSP1 and TSP2) are angiostatic extracellular matrix molecules also produced by synovial macrophages and fibroblasts. TSPs link connective tissue molecules to integrins. A TSP1-derived peptide ameliorated inflammation and angiogenesis in peptidoglycan-induced rat arthritis [209]. TSP2 blocked capillary formation in arthritic severe combined immunodeficiency (SCID) mice [210].

Among endogenously produced cytokines, chemokines, and growth factors, IL-4 and IL-13 gene transfer attenuated synovitis and angiogenesis in the rat AIA

**TABLE 4 Possibilities of Angiogenesis Targeting in Arthritis**

| Mode of Inhibition | Chemokine/Chemokine Receptor | Animal (A)/Human (H) study | Reference |
|---|---|---|---|
| Endogenous | Angiostatin | A | 104, 203, 204 |
| | Endostatin | A | 15, 205, 206 |
| | Kallistatin | A | 207 |
| | Arresten, canstatin, tumstatin | A | 208 |
| | Thrompospondin 1 and 2 | A | 209, 210 |
| | IL-4, IL-13 gene transfer | A | 129, 211 |
| | PF4/CXCL4 | A | 25 |
| | Fumagillin analogues | A | 104, 213–215 |
| Exogenous | Classical DMARDs | H | 3, 7, 15 |
| | Anti-TNF agents (infliximab) | H | 7, 15, 120, 224 |
| | Thalidomide | A+H | 104, 216–218 |
| | VEGF inhibitors | A | 6, 104, 220 |
| | Vatalanib | H | 104, 221 |
| | HIF inhibitors (e.g., YC-1) | A | 219 |
| | Ang1/Tie2 inhibitors | A | 223 |
| | αVβ3 inhibitors (e.g., Vitaxin) | A+H | 6, 7, 104 |
| | Paclitaxel | H | 104 |
| | PPARγ agonists (e.g., pioglitazone) | H | 225 |
| | MMP inhibitors | A | 122 |
| | Endothelin-1 antagonists | H | 126 |

system [129, 211]. The administration of angiostatic chemokines to arthritis models was discussed above.

TNP-470 and PPI2458 are two angiostatic derivatives of fumagillin, a naturally occurring product of *Aspergillus fumigatus*. These compounds inhibit methionine aminopeptidase-2, an enzyme involved in neovascularization [212]. Fumagillin analogs act by the inhibition of VEGF production [6, 104]. In rodent models of RA, TNP-470 prevented arthritis when administered before the onset of the disease [104, 213]. PPI2458 suppressed synovitis and the development of joint erosions in murine CIA [214, 215].

***Exogenous Strategies*** Among nonspecific strategies, thalidomide, currently used to treat multiple myeloma, has been introduced into the treatment of RA and lupus. Thalidomide antagonizes the effects of TNF-α and inhibits angiogenesis [104, 216]. Some studies on thalidomide-mediated angiostasis gave contradictory results, as in some studies thalidomide suppressed VEGF production, capillary formation, and synovitis [104, 217], while in another rat CIA study it did not affect the release of angiogenic VEGF or TNF-α [218]. Thus, the antiangiogenic potential of thalidomide in arthritis needs further clarification. Yet, CC1069, a thalidomide analog, potently inhibited rat AIA [104].

The hypoxia–HIF pathway may also be targeted using nonspecific inhibitor compounds. YC-1, a superoxide-sensitive stimulator of soluble guanylyl cyclase is also an inhibitor of HIF-1 [219].

Regarding specific blockade of major angiogenic mediators, VEGF could be a primary therapeutic target. VEGF has been targeted by using synthetic VEGF and VEGF receptor inhibitors, anti-VEGF antibodies, as well as inhibitors of VEGF and VEGF receptor signaling [6, 104, 220]. Among VEGF receptor tyrosine kinase inhibitors, vatalanib has been tried and attenuated knee arthritis in rabbits [104, 221]. Semaphorin-3A blocks the effects of the 165-amino-acid form of VEGF (VEGF165) and inhibits angiogenesis [222].

The Ang–Tie system may also be a suitable target. A soluble Tie2 receptor transcript was delivered via an adenoviral vector to mice with CIA. The delivery of this transcript postponed the onset and attenuated the severity of arthritis [223].

Vitaxin, a humanized antibody to the $\alpha_V\beta_3$ integrin, inhibited synovial angiogenesis in rodent models of arthritis [6, 7, 104]. Some specific MMP inhibitors have been tested in angiogenesis models [122].

### 6.2 Human in Vitro and in Vivo Studies

***Endogenous Inhibitors*** Among the endogenous angiostatic molecules described above, only angiostatin and endostatin have been tried to treat human cancer [7, 15, 222]. Reports on RA studies are still awaited.

***Exogenous Strategies*** Traditional DMARDs and biologics currently used to treat RA may suppress neovascularization, as well as chemokine release described above. These agents may inhibit synovial

capillary formation by nonspecifically blocking the effects of angiogenic mediators [3, 7, 15]. For example, infliximab + methotrexate treatment inhibited systemic and synovial VEGF release, resulting in decreased synovial neovascularization [7, 15, 224]. Anti-TNF therapy of RA patients also reduced Ang1–Tie2 and survivin but stimulated Ang2 production [120].

Among other small molecular compounds, the possible use of thalidomide in human RA was described above [104, 216]. In a phase I study, paclitaxel, currently used to treat cancer, was effective and safe in RA patients [104]. Pioglitazone, an antidiabetic PPARγ agonist, is also angiostatic. This agent effectively controlled psoriatic arthritis in 10 patients [225]. Endothelin-1 antagonists currently used in the therapy of primary pulmonary hypertension, as well as scleroderma, may also exert antiangiogenic effects [126].

Regarding specific approaches, numerous small molecular VEGF receptor tyrosine kinase inhibitors are under development for the treatment of malignancies. These compounds include vatalanib, sunitinib malate, sorafenib, vandetanib, and AG013736 [104, 220]. The anti-VEGF monoclonal antibody, bevacizumab, is currently used for the treatment of colon and lung cancer [104]. There have been no completed and published studies using these VEGF inhibitors in RA. Vitaxin has also been introduced to a phase II human RA trial, but it showed very limited efficacy [104].

## 7 SUMMARY

In this review, we have discussed the potential role of chemokines, chemokine receptors, and angiogenesis in the pathogenesis of RA. A number of CXC chemokines, some CC and $CX_3C$ chemokines, and chemokine receptors may be involved in the angiogenic, as well as inflammatory, events underlying RA. In addition, chemokine and angiogenesis research may have clinical relevance for RA as well. The determination of synovial vascularity and the expression of angiogenic mediators in the arthritic synovium may have some value for determining the progression of RA. Antichemokine and antiangiogenesis targeting using either endogenously produced or exogenously administered inhibitors may control synovial inflammation.

## REFERENCES


1. Koch AE (2005). Chemokines and their receptors in rheumatoid arthritis. *Arthritis Rheum 52*, 710–721.
2. Szekanecz Z, Kim J, Koch AE (2002). Chemokines and chemokine receptors in rheumatoid arthritis. *Sem Immunol 399*, 1–7.
3. Szekanecz Z, Koch AE (2001). Chemokines and angiogenesis. *Curr Opin Rheumatol 13*, 202–208.
4. Szekanecz Z, Szücs G, Szántó S, Koch AE (2006). Chemokines in rheumatic diseases. *Curr Drug Targ 7*, 91–102.
5. Vergunst CE, Tak PP (2005). Chemokines: Their role in rheumatoid arthritis. *Curr Rheumatol Rep 7*, 382–388.
6. Koch AE (1998). Angiogenesis: Implications for rheumatoid arthritis. *Arthritis Rheum 41*, 951–962.
7. Szekanecz Z, Gaspar L, Koch AE (2005). Angiogenesis in rheumatoid arthritis. *Front Biosci 10*, 1739–1753.
8. Paleolog EM, Fava RA (1998). Angiogenesis in rheumatoid arthritis: Implications for future therapeutic strategies. *Springer Semin Immunopathol 20*, 73–94.
9. Rudolph EH, Woods JM (2005). Chemokine expression and regulation of angiogenesis in rheumatoid arthritis. *Curr Pharm Des 11*, 613–631.
10. Walsh DA (1999). Angiogenesis and arthritis. *Rheumatology (Oxford) 38*, 103–114.
11. Szekanecz Z, Koch AE (2004). Vascular endothelium and immune responses: Implications for inflammation and angiogenesis. *Rheum Dis Clin North Am 30*, 97–114.
12. Liu LX, Lu H, Luo Y, Date T, Belanger AJ, Vincent KA, Akita GY, Goldberg M, Cheng SH, Gregory RJ, Jiang C (2002). Stabilization of vascular endothelial growth factor mRNA by hypoxia-inducible factor 1. *Biochem Biophys Res Commun 291*, 908–914.
13. Holash J, Maisonpierre PC, Compton D, Boland P, Alexander CR, Zagzag D, Yancopoulos GD, Wiegand SJ (1999). Vessel cooption, regression and growth in tumors mediated by angiopoietins and VEGF. *Science 284*, 1994–1998.
14. Suri C, Jones PF, Patan S, Bartunkova S, Maisonpierre PC, Davis S, Sato TN, Yancopoulos GD (1996). Requisite role of angiopoietin-1 a ligand for the TIE2 receptor, during embryonic angiogenesis. *Cell 87*, 1171–1180.
15. Veale DJ, Fearon U (2006). Inhibition of angiogenic pathways in rheumatoid arthritis: Potential for therapeutic targeting. *Best Pract Res Clin Rheumatol 20*, 941–947.
16. Milkiewicz M, Ispanovic E, Doyle JL, Haas TL (2006). Regulators of angiogenesis and strategies for their respective manipulation. *Int J Biochem Cell Biol 38*, 333–357.
17. Zlotnik A, Yoshie O (2000). Chemokines: A new classification system and their role in immunity. *Immunity 12*, 121–127.
18. Harris ED (1990). Rheumatoid arthritis: Pathophysiology and implications for therapy. *N Engl J Med 332*, 1277–1287.
19. Oppenheim JJ, Zachariae COC, Mukaida N, Matsushima K (1991). Properties of the novel proinflammatory supergene "intercrine" cytokine family. *Annu Rev Immunol 9*, 617–648.
20. Taub DD (1996). C–C chemokines: An overview. In Koch AE, Strieter RM, Eds. *Chemokines in Disease*. Austin, TX: RG Landes Company, pp. 27–54.

21. Walz A, Kunkel SL, Strieter RM (1996). C–X–C chemokines: An overview. In Koch AE, Strieter RM, Eds. *Chemokines in Disease*. Austin, RG Landes Company, pp. 1–25.
22. Bacon K, Baggiolini M, Broxmeyer H, Horuk R, Lindley I, Mantovani A, Maysushima K, Murphy P, Nomiyama H, Oppenheim J (2002). Chemokine/chemokine receptor nomenclature. *J Interferon Cytokine Res 22*, 1067–1068.
23. Strieter RM, Polverini PJ, Kunkel SL, Arenberg DA, Burdick MD, Kasper J, Dzuiba J, Van Damme J, Walz A, Marriott D (1995). The functional role of the ELR motif in CXC chemokine-mediated angiogenesis. *J Biol Chem 270*, 27348–27357.
24. Koch AE, Kunkel SL, Shah MR, Hosaka S, Halloran MM, Haines GK, Burdick MD, Pope RM, Strieter RM (1995). Growth related gene product alpha: A chemotactic cytokine for neutrophils in rheumatoid arthritis. *J Immunol 155*, 3660–3666.
25. Wooley PH, Schaefer C, Whalen JD. Dutcher JA, Counts JF (1997). A peptide sequence from platelet factor 4 (CT-112) is effective in the treatment of type II collagen induced arthritis in mice. *J Rheumatol 24*, 890–898.
26. Koch AE, Kunkel SL, Harlow LA, Johnson B, Evanoff HL, Haines GK, Burdick MD, Pope RM, Strieter RM (1994). Epithelial neutrophil activating peptide-78: A novel chemotactic cytokine for neutrophils in arthritis. *J Clin Invest 94*, 1012–1018.
27. Koch AE, Kunkel SL, Burrows JC, Evanoff HL, Haines GK, Pope RM, Strieter RM (1991). Synovial tissue macrophage as a source of the chemotactic cytokine IL-8. *J Immunol 147*, 2187–2195.
28. Nanki T, Hayashida K, El-Gabalawy HS, Suson S, Shi K, Girschick HJ, Yavuz S, Lipsky PE (2000). Stromal cell-derived factor-1-CXC chemokine receptor 4 interactions play a central role in CD4+ T-cell accumulation in rheumatoid arthritis synovium. *J Immunol 165*, 6590–6598.
29. Ruth JH, Haas CS, Park CC, Amin MA, Martinez RJ, Haines GK, Shahrara S, Campbell PL, Koch AE (2006). CXCL16-mediated cell recruitment to rheumatoid arthritis synovial tissue and murine lymph nodes is dependent upon the MAPK pathway. *Arthritis Rheum 54*, 765–778.
30. Nanki T, Shimaoka T, Hayashida K Taniguchi K, Yonehara S, Miyasaka N (2005). Pathogenic role of CXCL16-CXCR6 pathway in rheumatoid arthritis. *Arthritis Rheum 52*, 3004–3014.
31. Deleuran B, Lemche P, Kristensen M, Chu CQ, Field M, Jensen J, Matsushima K, Stengaard-Pedersen K (1994). Localisation of interleukin 8 in the synovial membrane, cartilage–pannus junction and chondrocytes in rheumatoid arthritis. *Scand J Rheumatol 23*, 2–7.
32. Koch AE, Volin MV, Woods JM, Kunkel SL, Connors MA, Harlow LA, Woodruff DC, Burdick MD, Strieter RM (2001). Regulation of angiogenesis by the C–X–C chemokines interleukin-8 and epithelial neutrophil activating peptide-78 in the rheumatoid joint. *Arthritis Rheum 44*, 31–40.
33. Hosaka S, Akahoshi T, Wada C, Kondo H (1994). Expression of the chemokine superfamily in rheumatoid arthritis. *Clin Exp Immunol 97*, 451–457.
34. Szekanecz Z, Koch AE (2007). Macrophages and their products in rheumatoid arthritis. *Curr Opin Rheumatol 19*, 289–295.
35. Koch AE, Polverini PJ, Kunkel SL, Harlow LA, DiPietro LA, Elner VM, Elner SG, Strieter RM (1992). Interleukin-8 as a macrophage-derived mediator of angiogenesis. *Science 258*, 1798–1801.
36. Salcedo R, Ponce ML, Young HA, Wasserman K, Ward JM, Kleinman HK, Oppenheim JJ, Murphy WJ (2000). Human endothelial cells express CCR2 and respond to MCP-1: Direct role of MCP-1 in angiogenesis and tumor progression. *Blood 96*, 34–40.
37. Castor CW, Smith EM, Hossler PA, Bignall MC, Aaron BP (1992). Detection of connective tissue activating peptide-III isoforms in synovium from osteoarthritis and rheumatoid arthritis patients: Patterns of interaction with other synovial cytokines in cell culture. *Arthritis Rheum 35*, 783–793.
38. Pierer M, Rethage J, Seibl R, Lauener R, Brentano F, Wagner U, Hantzschel H, Michel BA, Gay RE, Gay S, Kyburz D (2004). Chemokine secretion of rheumatoid arthritis synovial fibroblasts stimulated by Toll-like receptor 2 ligands. *J Immunol 172*, 1256–1265.
39. Hanaoka R, Kasama T, Muramatsu M Yajima N, Shiozawa F, Miwa Y, Negishi M, Ide H, Miyaoka H, Uchida H (2003). A novel mechanism for the regulation of IFN-γ inducible protein-10 expression in rheumatoid arthritis. *Arthritis Res Ther 5*, R74–R81.
40. Patel DD, Zachariah JP, Whichard LP (2001). CXCR3 and CCR5 ligands in the rheumatoid arthritis synovium. *Clin Immunol 98*, 39–45.
41. Blades MC, Ingegnoli F, Wheller SK, Manzo A, Wahid S, Panayi GS, Perretti M, Pitzalis C (2002). Stromal cell-derived factor 1 (CXCL12) induces monocyte migration into human synovium transplanted onto SCID mice. *Arthritis Rheum 46*, 824–836.
42. Bradfield PF, Amft N, Vernon-Wilson E Exley AE, Parsonage G, Rainger GE, Nash GB, Thomas AM, Simmons DL, Salmon M (2003). Rheumatoid fibroblast-like synoviocytes overexpress the chemokine stromal cell-derived factor 1 (CXCL12) which supports distinct patterns and rates of CD4+ and CD8+ T cell migration within synovial tissue. *Arthritis Rheum 48*, 2472–2482.
43. Hansen IB, Ellingsen T, Hornung N, Poulsen JH, Lottenburger T, Stengaard-Pedersen K (2006). Plasma level of CXC-chemokine CXCL12 is increased in rheumatoid arthritis and is independent of disease activity and methotrexate treatment. *J Rheumatol 33*, 1754–1759.
44. Kim KW, Cho ML, Kim HR, Ju JH, Park MK, Oh HJ, Kim JS, Park SH, Lee SH, Kim HY (2007). Up-regulation of stromal cell-derived factor 1 (CXCL12) production in rheumatoid synovial fibroblasts through interactions with T lymphocytes: Role of interleukin-17 and CD40L-CD40 interaction. *Arthritis Rheum 56*, 1076–1086.

45. Pablos JL, Santiago B, Galindo M Torres C, Brehmer MT, Blanco FJ, Garcia-Lazaro FJ (2003). Synoviocyte-derived CXCL12 is displayed on endothelium and induces angiogenesis in rheumatoid arthritis. *J Immunol 170*, 2147–2152.
46. Santiago B, Baleux F, Palao G, Gutierrez-Canas I, Ramirez JC, Arenzana-Seisdedos F, Pablos JL (2006). CXCL12 is displayed by rheumatoid endothelial cells through its basic amino-terminal motif on heparan sulfate proteoglycans. *Arthritis Res Ther 8*, R43.
47. Joven B, Gonzalez N, Aguilar F, Santiago B, Galindo M, Alcami J (2005). Association between stromal cell derived factor 1 chemokine gene variant and radiographic progression of rheumatoid arthritis. *Arthritis Rheum 52*, 354–356.
48. Gronthos S, Zannettino AC (2007). The role of the chemokine CXCL12 in osteoclastogenesis. *Trends Endocrinol Metab 18*, 108–113.
49. Silverman GJ, Carson DA (2003). Roles of B cells in rheumatoid arthritis. *Arthritis Res Ther 5*(Suppl 4), 1–6.
50. De Vita S, Zaja F, Sacco S, De Candia A, Fanin R, Ferraccioli G (2002). Efficacy of selective B cell blockade in the treatment of rheumatoid arthritis. *Arthritis Rheum 46*, 2029–2033.
51. Edwards JCW, Cambridge G, Abrahams VM (1999). Do self-perpetuating B lymphocytes drive human autoimmune disease? *Immunology 97*, 188–196.
52. Moser B, Loetscher P (2001). Lymphocyte traffic control by chemokines. *Nat Immunol 2*, 123–128.
53. Kunkel EJ, Butcher EC (2002). Chemokines and the tissue-specific migration of lymphocytes. *Immunity 16*, 1–4.
54. Manzo A, Paoletti S, Carulli M, Blades MC, Barone F, Yanni G, Fitzgerald O, Bresnihan B, Caporali R, Montecucco C (2005). Systematic microanatomical analysis of CXCL13 and CCL21 in situ production and progressive lymphoid organization in rheumatoid synovitis. *Eur J Immunol 35*, 1347–1359.
55. Li T, Zhong J, Chen Y, Qiu X, Zhang T, Ma D, Han W (2006). Expression of chemokine-like factor 1 is upregulated during T lymphocyte activation. *Life Sci 79*, 519–524.
56. Van Lieshout AW, Fransen J, Flendrie M, Eijsbouts AM, van den Hoogen FH, van Riel PL, Radstake TR (2007). Circulating levels of the chemokine CCL18 but not CXCL16 are elevated and correlate with disease activity in rheumatoid arthritis. *Ann Rheum Dis 66*, 1334–1338.
57. Haringman JJ, Smeets TJ, Reinders-Blankert P, Tak PP (2006). Chemokine and chemokine receptor expression in paired peripheral blood mononuclear cells and synovial tissue of patients with rheumatoid arthritis, osteoarthritis and reactive arthritis. *Ann Rheum Dis 65*, 294–300.
58. Koch AE, Kunkel SL, Harlow LA Johnson B, Evanoff HL, Haines GK, Burdick MD, Pope RM, Strieter RM (1992). Enhanced production of monocyte chemoattractant protein-1 in rheumatoid arthritis. *J Clin Invest 90*, 772–779.
59. Villiger PM, Terkeltaub R, Lotz M (1992). Production of monocyte chemoattractant protein-1 by inflamed synovial tissue and cultured synoviocytes. *J Immunol 149*, 722–727.
60. Volin MV, Shah MR, Tokuhira M, Haines GK, Woods JM, Koch AE (1998) RANTES expression and contribution to monocyte chemotaxis in arthritis. *Clin Immunol Immunopathol 89*(1), 44–53.
61. Rathanaswami P, Hachicha M, Sadick M, Scahll TJ, McColl SR (1993). Expression of the cytokine RANTES in human rheumatoid synovial fibroblasts. Differential regulation of RANTES and interleukin-8 genes by inflammatory cytokines. *J Biol Chem 268*, 5834–5839.
62. Hounoki H, Sugiyama E, Mohamed SG, Shinoda K, Taki H, Abdel-Aziz HO, Maruyama M, Kobayashi M, Miyahara T (2008). Activation of peroxisome proliferator-activated receptor gamma inhibits TNF-alpha mediated osteoclast differentiation in human peripheral monocytes in part via suppression of monocyte chemoattractant protein-1 expression. *Bone 42*, 765–774.
63. Wang CR, Guo HR, Liu MF (2005). RANTES promoter polymorphism as a genetic risk factor for rheumatoid arthritis in the Chinese. *Clin Exp Rheumatol 23*, 379–384.
64. McKinney C, Merriman ME, Chapman PT, Gow PJ, Harrison AA, Highton J, Jones PB, McLean L, O'Donnell JL, Pokorny V, Spellerberg M, Stamp LK, Willis J, Steer S, Merriman TR (2008). Evidence for an influence of chemokine ligand 3-like 1 (CCL3L1) gene copy number on susceptibility to rheumatoid arthritis. *Ann Rheum Dis 67*, 409–413.
65. Matsui T, Akahoshi T, Namai R, Hashimoto A, Kurihara Y, Rana M, Nishimura A, Endo H, Kitasato H, Kawai S (2001). Selective recruitment of CCR6-expressing cells by increased production of MIP-3 alpha in rheumatoid arthritis. *Clin Exp Immunol 125*, 155–161.
66. Hirota K, Yoshitomi H, Hashimoto M, Maeda S, Teradaira S, Sugimoto N, Yamaguchi T, Nomura T, Ito H, Nakamura T, Sakaguchi N, Sakaguchi S (2007). Preferential recruitment of CCR6-expressing Th17 cells to inflamed joints via CCL20 in rheumatoid arthritis and its animal model. *J Exp Med 204*, 2803–2812.
67. Lisignoli G, Piacentini A, Cristino S, Grassi F, Cavallo C, Cattini L, Tonnarelli B, Manferdini C, Facchini A (2007). CCL20 chemokine induces both osteoblast proliferation and osteoclast differentiation: Increased levels of CCL20 are expressed in subchondral bone tissue of rheumatoid arthritis patients. *J Cell Physiol 210*, 798–806.
68. Auer J, Blass M, Schulze-Koops H, Russwurm S, Nagel T, Kalden JR, Röllinghoff M, Beuscher HU (2007). Expression and regulation of CCL18 in synovial fluid neutrophils in patients with rheumatoid arthritis. *Arthritis Res Ther 9*, R94.
69. Iwamoto T, Okamoto H, Kobayashi S, Ikari K, Toyama Y, Tomatsu T, Kamatani N, Momohara S (2007). A role of

monocyte chemoattractant protein-4 (MCP-4)/CCL13 from chondrocytes in rheumatoid arthritis. *FEBS J 274*, 4904–4912.

70. Buckley CD (2003). Why do leucocytes accumulate within chronically inflamed joints? *Rheumatology (Oxford) 42*, 1433–1444.
71. Manzo A, Bugatti S, Caporali R, Prevo R, Jackson DG, Uguccioni M, Buckley CD, Montecucco C, Pitzalis C (2007). CCL21 expression pattern of human secondary lymphoid organ stroma is conserved in inflammatory lesions with lymphoid neogenesis. *Am J Pathol 171*, 1549–1562.
72. Okamoto H, Koizumi K, Yamanaka H, Saito T, Kamatani N (2003). A role for TARC/CCL17, a CC chemokine, in systemic lupus erythematosus. *J Rheumatol 30*, 2369–2373.
73. Momohara S, Okamoto H, Iwamoto T, Mizumura T, Ikari K, Kawaguchi Y, Takeuchi M, Kamatani N, Tomatsu T (2007). High CCL18/PARC expression in articular cartilage and synovial tissue of patients with rheumatoid arthritis. *J Rheumatol 34*, 266–271.
74. Van Lieshout AW, van der Voort R, le Blanc LM, Roelofs MF, Schreurs BW, van Riel PL, Adema GJ, Radstake TR (2006). Novel insights in the regulation of CCL18 secretion by monocytes and dendritic cells via cytokines, toll-like receptors and rheumatoid synovial fluid. *BMC Immunol 19*, 7–23.
75. Bazan JF, Bacon KB, Hardiman G, Wang W, Soo K, Rossi D, Greaves DR, Zlotnik A, Schall TJ (1997). A new class of membrane bound chemokine with a X3C motif. *Nature 385*, 640–644.
76. Ruth JH, Volin MV, Haines III GK, Woodruff DC, Katschke KJ Jr, Woods JM, Park CC, Morel JC, Koch AE (2001). Fractalkine, a novel chemokine in rheumatoid arthritis and rat adjuvant-induced arthritis. *Arthritis Rheum 44*, 1568–1581.
77. Blaschke S, Middel P, Dorner BG, Blaschke V, Hummel KM, Kroczek RA, Reich K, Benoehr P, Koziolek M, Muller GA (2003). Expression of activation-induced, T cell-derived, and chemokine related cytokine/lymphotactin and its functional role in rheumatoid arthritis. *Arthritis Rheum 48*, 1858–1872.
78. Loetscher P, Uguccioni M, Bordoli L, Baggiolini M, Moser B, Chizzolini C, Dayer JM (1998). CCR5 is characteristic of Th1 lymphocytes. *Nature 391*, 344–345.
79. Sawai H, Park YW, Robertson J, Iwai T, Goronzy JJ, Weyand CM (2005). T cell costimulation by fractalkine-expressing synoviocytes in rheumatoid arthritis. *Arthritis Rheum 52*, 1392–1401.
80. Sawai H, Park YW, He X, Goronzy JJ, Weyand CM (2007). Fractalkine mediates T cell-dependent proliferation of synovial fibroblasts in rheumatoid arthritis. *Arthritis Rheum 56*, 3215–3225.
81. Volin MV, Huynh N, Klosowska K, Chong KK, Woods JM (2007). Fractalkine is a novel chemoattractant for rheumatoid arthritis fibroblast-like synoviocytes signalling through MAP kinases and Akt. *Arthritis Rheum 56*, 2512–2522.
82. Matsunawa M, Isozaki T, Odai T, Yjima N, Takeuchi HT, Negishi M, Ide H, Adachi M, Kasama T (2006). Increased serum levels of soluble fractalkine (CX3CL1) correlate with disease activity in rheumatoid vasculitis. *Arthritis Rheum 54*, 3408–3416.
83. McDermott DH, Fong AM, Yang Q Sechler JM, Cupples LA, Merrell MN, Wilson PW, D'Agostino RB, O'Donnell CJ, Patel DD (2003). Chemokine receptor mutant CX3CR1-M280 has impaired adhesive function and correlates with protection from cardiovascular disease in humans. *J Clin Invest 111*, 1241–1250.
84. Volin MV, Woods JM, Amin MA, Connors MA, Harlow LA, Koch AE (2001). Fractalkine: A novel angiogenic chemokine in rheumatoid arthritis. *Am J Pathol 159*, 1521–1530.
85. Pingiotti E, Cipriani P, Marrelli A, Liakouli V, Fratini S, Penco M, Giacomelli R (2007). Surface expression of fractalkine receptor (CX3CR1) on CD4+/CD28– T cells in RA patients and correlation with atherosclerotic damage. *Ann NY Acad Sci 1107*, 32–41.
86. Qin S, Rottman JB, Myers P, Kassam N, Weinblatt M, Loetscher M, Koch AE, Moser B, Mackay CR (1998). The chemokine receptors CXCR3 and CCR5 mark subsets of T cells with a homing predilection for certain inflammatory sites. *J Clin Invest 101*, 746–750.
87. Norii M, Yamamura M, Iwahashi M, Ueno A, Yamana J, Makino H (2006). Selective recruitment of CXCR3+ and CCR5+ T cells into synovial tissue in patients with rheumatoid arthritis. *Acta Med Okayama 60*, 149–157.
88. Garcia-Lopez MA, Sanchez-Madrid F, Rodriguez-Frade JM (2001). CXCR3 chemokine receptor distribution in normal and inflamed tissues. *Lab Invest 81*, 409–418.
89. Schmutz C, Hulme A, Burman A Salmon M, Ashton B, Buckley C, Middleton J (2005). Chemokine receptors in the rheumatoid synovium: Upregulation of CXCR5. *Arthritis Res Ther 7*, R217–R229.
90. Wengner AM, Hopken UE, Petrow PK, Hartmann S, Schurigt U, Brauer R, Lipp M (2007). CXCR5- and CCR7-dependent lymphoid neogenesis in a murine model of chronic antigen-induced arthritis. *Arthritis Rheum 56*, 3271–3283.
91. Katschke KJ Jr, Rottman JB, Ruth JH, Qin S, Wu L, LaRosa G, Ponath P, Park CC, Pope RM, Koch AE (2001). Differential expression of chemokine receptors on peripheral blood, synovial fluid and synovial tissue monocytes/macrophages in rheumatoid arthritis. *Arthritis Rheum 44*, 1022–1032.
92. Borzi RM, Mazzetti I, Cattini L, Uguccioni M, Baggiolini M, Facchini A (2000). Human chondrocytes express functional chemokine receptors and release matrix-degrading enzymes in response to C–X–C and C–C chemokines. *Arthritis Rheum 43*, 1734–1741.
93. Ruth JH, Rottman JB, Katschke KJ Jr, Qin S, Wu L, LaRosa G, Ponath P, Pope RM, Koch AE (2001). Selective

lymphocyte chemokine receptor expression in the rheumatoid joint. *Arthritis Rheum 44*, 2750–2760.

94. Desmetz C, Lin YL, Mettling C, Portales P, Noel D, Clot J, Jorgensen C, Corbeau P (2007). Cell surface CCR5 density determines the intensity of T cell migration towards rheumatoid arthritis synoviocytes. *Clin Immunol 123*, 148–154.
95. Prahalad S (2006). Negative association between the chemokine receptor CCR5-Δ32 polymorphism and rheumatoid arthritis: A metaanalysis. *Genes Immun 7*, 264–268.
96. Kohem CL, Brenol JC, Xavier RM, Bredemeier M, Brenol CV, Dedavid e Silva TL, de Castilhos Mello A, Canedo AD, Neves AG, Chies JA (2007). The chemokine receptor CCR5 genetic polymorphism and expression in rheumatoid arthritis patients. *Scand J Rheumatol 36*, 359–364.
97. Lindner E, Nordang GB, Melum E, Flato B, Selvaag AM, Thorsby E, Kvien TK, Forre OT, Lie BA (2007). Lack of association between the chemokine receptor 5 polymorphism CCR5delta32 in rheumatoid arthritis and juvenile idiopathic arthritis. *BMC Med Genet 8*, 33.
98. Galligan CL, Matsuyama W, Matsukawa A, Mizuta H, Hodge DR, Yoshimura T (2004). Up-regulated expression and activation of the orphan chemokine receptor, CCRL2, in rheumatoid arthritis. *Arthritis Rheum 50*, 1806–1814.
99. Patterson AM, Siddall H, Chamberlain G, Gardner L, Middleton J (2002). Expression of the Duffy antigen/receptor for chemokines (DARC) by the inflamed synovial endothelium. *J Pathol 197*, 108–116.
100. Szekanecz Z, Szegedi G, Koch AE (1998). Angiogenesis in rheumatoid arthritis: Pathogenic and clinical significance. *J Invest Med 46*, 27–41.
101. Folkman J, Klagsbrun M (1987). Angiogenic factors. *Science 235*, 442–447.
102. Taylor PC, Sivakumar B (2005). Hypoxia and angiogenesis in rheumatoid arthritis. *Curr Opin Rheumatol 17*, 293–298.
103. Auerbach W, Auerbach R (1994). Angiogenesis inhibition: A review. *Pharmacol Ther 63*, 265–311.
104. Lainer-Carr D, Brahn E (2007). Angiogenesis inhibition as a therapeutic approach for inflammatory synovitis. *Nat Clin Pract Rheumatol 3*, 434–442.
105. Davis S, Aldrich TH, Jones PF, Acheson A, Compton DL, Jain V, Ryan TE, Bruno J, Radziejewski C, Maisonpierre PC, Yancopoulos GD (1996). Isolation of angiopoietin-1, a ligand for the TIE2 receptor, by secretion-trap expression cloning. *Cell 87*, 1161–1169.
106. Tran J, Master Z, Yu JL, Rak J, Dumont DJ, Kerbel RS (2002). A role for survivin in chemoresistance of endothelial cells mediated by VEGF. *Proc Natl Acad Sci USA 99*, 4349–4354.
107. Koch AE, Harlow LA, Haines GK, Amento EP, Unemori EN, Wong WL, Pope RM, Ferrara N (1994). Vascular endothelial growth factor. A cytokine modulating endothelial function in rheumatoid arthritis. *J Immunol 152*, 4149–4156.
108. Giatromanolaki A, Sivridis E, Maltezos E, Athanassou N, Papazoglou D, Gatter KC, Harris AL, Koukourakis MI (2003). Upregulated hypoxia inducible factor-1a and -2a pathway in rheumatoid arthritis and osteoarthritis. *Arthritis Res Ther 5*, R193–R201.
109. Gravallese EM, Pettit AR, Lee R, Madore R, Manning C, Tsay A, Gaspar J, Goldring MB, Goldring SR, Oettgen P (2003). Angiopoietin-1 is expressed in the synovium of patients with rheumatoid arthritis and is induced by tumour necrosis factor alpha. *Ann Rheum Dis 62*, 100–107.
110. Shahrara S, Volin MV, Connors MA, Haines GK, Koch AE (2002). Differential expression of the angiogenic Tie receptor family in arthritic and normal synovial tissue. *Arthritis Res 4*, 201–208.
111. Szekanecz Z, Strieter RM, Koch AE (1998). Cytokines in rheumatoid arthritis: Potential targets for pharmacological intervention. *Drugs Aging 12*, 377–390.
112. Brennan F, Beech J (2007). Update on cytokines in rheumatoid arthritis. *Curr Opin Rheumatol 19*, 296–301.
113. Park CC, Morel JC, Amin MA, Connors MA, Harlow LA, Koch AE (2001). Evidence of IL-18 as a novel angiogenic mediator. *J Immunol 167*, 1644–1653.
114. Angiolillo AL, Kanegane H, Sgadari C, Reaman GH, Tosato G (1997). Interleukin-15 promotes angiogenesis in vivo. *Biochem Biophys Res Commun 233*, 231–237.
115. Numasaki M, Watanabe M, Suzuki T, Takahashi H, Nakamura A, Goto J, Lotze MT, Sasaki H (2005). IL-17 enhances the net angiogenic activity and in vivo growth of human non-small cell lung cancer in SCID mice through promoting CXCR2-dependent angiogenesis. *J Immunol 175*, 6177–6189.
116. Amin MA, Mansfield PJ, Pakozdi A, Campbell PL, Ahmed S, Martinez RJ, Koch AE (2007). Interleukin-18 induces angiogenic factors in rheumatoid arthritis synovial tissue fibroblasts via distinct signaling pathways. *Arthritis Rheum 56*, 1787–1797.
117. Fearon U, Mullan R, Markham T, Connolly M, Sullivan S, Poole AR, FitzGerald O, Bresnihan B, Veale DJ (2006). Oncostatin M induces angiogenesis and cartilage degradation in rheumatoid arthritis synovial tissue and human cartilage cocultures. *Arthritis Rheum 54*, 3152–3162.
118. Amin MA, Volpert OV, Woods JM, Kumar P, Harlow LA, Koch AE (2003). Migration inhibitory factor mediates angiogenesis via mitogen-activated protein kinase and phosphatidylinositol kinase. *Circ Res 93*, 321–329.
119. Morand EF, Leech M, Bernhagen J (2006) MIF: A new cytokine link between rheumatoid arthritis and atherosclerosis. *Nat Rev Drug Discov 5*, 399–410.
120. Markham T, Mullan R, Golden-Mason L, Rogers S, Bresnihan B, Fitzgerald O, Fearon U, Veale DJ (2006). Resolution of endothelial activation and down-regulation of Tie2 receptor in psoriatic skin after infliximab therapy. *J Am Acad Dermatol 54*, 1003–1012.

121. Madri JA, Williams KS (1983). Capillary endothelial cell cultures: Phenotypic modulation by matrix components. *J Cell Biol 97*, 153–165.

122. Skotnicki JS, Zask A, Nelson FC, Albright JD, Levin JI (1999). Design and synthetic considerations of matrix metalloproteinase inhibitors. *Ann NY Acad Sci 878*, 61–72.

123. Johnson B, Haines GK, Harlow LA, Koch AE (1993). Adhesion molecule expression in human synovial tissues. *Arthritis Rheum 36*, 137–146.

124. Jacq L, Garnier S, Dieudé P, Michou L, Pierlot C, Migliorini P, Balsa A, Westhovens R, Barrera P, Alves H, Vaz C, Fernandes M, Pascual-Salcedo D, Bombardieri S, Dequeker J, Radstake TR, Van Riel P, van de Putte L, Lopes-Vaz A, Glikmans E, Barbet S, Lasbleiz S, Lemaire I, Quillet P, Hilliquin P, Teixeira VH, Petit-Teixeira E, Mbarek H, Prum B, Bardin T, Cornélis F; European Consortium on Rheumatoid Arthritis Families (2007). The ITGAV rs3738919-C allele is associated with rheumatoid arthritis in the European Caucasian population: A family-based study. *Arthritis Res Ther 9*, R63.

125. Shahrara S, Castro-Rueda HP, Haines GK, Koch AE (2007). Differential expression of the FAK family kinases in rheumatoid arthritis and osteoarthritis synovial tissues. *Arthritis Res Ther 9*, R112.

126. Koch AE, Distler O (2007). Vasculopathy and disordered angiogenesis in selected rheumatic diseases: Rheumatoid arthritis and systemic sclerosis. *Arthritis Res Ther 9*(Suppl 2), S3.

127. Haq A, El-Ramahi K, Al-Dalaan A (1999). Serum and synovial fluid concentrations of endothelin-1 in patients with inflammatory arthritides. *J Med 30*, 51–60.

128. Lee MS, Yoo SA, Cho CS, Suh PG, Kim WU, Ryu SH (2006). Serum amyloid A binding to formyl peptide receptor-like 1 induces synovial hyperplasia and angiogenesis. *J Immunol 177*, 5585–5594.

129. Haas CS, Amin MA, Ruth JH, Allen BL, Ahmed S, Pakozdi A, Woods JM, Shahrara S, Koch AE (2007). In vivo inhibition of angiogenesis by interleukin-13 gene therapy in a rat model of rheumatoid arthritis. *Arthritis Rheum 56*, 2535–2548.

130. Mabjeesh NJ, Escuin D, LaVallee TM, Pribluda VS, Swartz GM, Johnson MS, Willard MT, Zhong H, Simons JW, Giannakakou P (2003). 2ME2 inhibits tumor growth and angiogenesis by disrupting microtubules and dysregulating HIF. *Cancer Cell 3*, 363–375.

131. Issekutz AC, Sapru K (2008). Modulation of adjuvant arthritis in the rat by 2/methoxyestradiol. An effect independent of an anti-angiogenic action. *Int Immunopharmacol 8*, 708–716.

132. Caunt M, Hu L, Tang T, Brooks PC, Ibrahim S, Karpatkin S (2006). Growth-regulated oncogene is pivotal in thrombin-induced angiogenesis. *Cancer Res 66*, 4125–4132.

133. Wang D, Wang H, Brown J, Daikoku T, Ning W, Shi Q, Richmond A, Strieter R, Dey SK, DuBois RN (2006). CXCL1 induced by prostaglandin E2 promotes angiogenesis in colorectal cancer. *J Exp Med 203*, 941–951.

134. Salcedo R, Wasserman K, Young HA, Grimm MC, Howard OM, Anver MR, Kleinman HK, Murphy WJ, Oppenheim JJ (1999). Vascular endothelial growth factor and basic fibroblast growth factor induce expression of CXCR4 on human endothelial cells: In vivo neovascularization induced by stromal-derived factor-1alpha. *Am J Pathol 154*, 1125–1135.

135. Petit I, Jin D, Rafii S (2007). The SDF-1-CXCR4 signaling pathway: A molecular hub modulating neo-angiogenesis. *Trends Immunol 28*, 299–307.

136. Liang Z, Brooks J, Willard M, Liang K, Yoon Y, Kang S, Shim H (2007). CXCR4/CXCL12 axis promotes VEGF-mediated tumor angiogenesis through Akt signaling pathway. *Biochem Biophys Res Commun 359*, 716–722.

137. Calatozzolo C, Maderna E, Pollo B, Gelati M, Marras C, Silvani A, Croci D, Boiardi A, Salmaggi A (2006). Prognostic value of CXCL12 expression in 40 low-grade oligodendrogliomas and oligoastrocytomas. *Cancer Biol Ther 5*, 827–832.

138. Stamatovic SM, Keep RF, Mostarica-Stojkovic M, Andjelkovic AV (2006). CCL2 regulates angiogenesis via activation of Ets-1 transcription factor. *J Immunol 177*, 2651–2661.

139. Son KN, Hwang J, Kwon BS, Kim J (2006). Human CC chemokine CCL23 enhances expression of matrix metalloproteinase-2 and invasion of vascular endothelial cells. *Biochem Biophys Res Commun 340*, 498–504.

140. Lesnik P, Haskell CA, Charo IF (2003). Decreased atherosclerosis in CX3CR1 –/– mice reveals a role for fractalkine in atherogenesis. *J Clin Invest 111*, 333–340.

141. Buckley CD, Amft N, Bradfield PF, Pilling D, Ross E, Arenzana-Seisdedos F, Amara A, Curnow SJ, Lord RM, Scheel-Toellner D, Salmon M (2000). Persistent induction of the chemokine receptor CXCR4 by TGF-beta 1 on synovial T cells contributes to their accumulation within the rheumatoid synovium. *J Immunol 165*, 3423–3429.

142. Zagzag D, Lukyanov Y, Lan L, Ali MA, Esencay M, Mendez O, Yee H, Voura EB, Newcomb EW (2006). Hypoxia-inducible factor 1 and VEGF upregulate CXCR4 in glioblastoma: implications for angiogenesis and glioma cell invasion. *Lab Invest 86*, 1221–1232.

143. Burns JM, Summers BC, Wang Y, Melikian A, Miao Z, Kuo CJ, Wei K, Wright K, Howard MC, Schall TJ (2006). A novel chemokine receptor for SDF-1 and I-TAC involved in cell survival, cell adhesion and tumor development. *J Exp Med 203*, 2201–2213.

144. Ochoa O, Sun D, Reyes-Reyna SM, Waite LL, Michalek JE, McManus LM, Shireman PK (2007). Delayed angiogenesis and VEGF production in CCR2–/– mice during impaired skeletal muscle regeneration. *Am J Physiol Regul Integr Comp Physiol 293*, R651–R661.

145. Wang J, Ou ZL, Hou YF, Luo JM, Shen ZZ, Ding J, Shao ZM (2006). Enhanced expression of Duffy antigen receptor for chemokines by breast cancer cells

attenuates growth and metastasis potential. *Oncogene 25*, 7201–7211.

146. Boulday G, Haskova Z, Reinders ME, Pal S, Broiscoe DM (2006). Vascular endothelial growth factor-induced signaling pathways in endothelial cells that mediate overexpression of the chemokine IFN-gamma-inducible protein of 10 kDa in vitro and in vivo. *J Immunol 176*, 3098–3107.
147. Bodnar RJ, Yates CC, Wells A (2006). IP-10 blocks vascular endothelial growth factor-induced endothelial cell motility and tube formation via inhibition of calpain. *Circ Res 98*, 617–625.
148. Maurer AM, Zhou B, Han ZC (2006). Roles of platelet factor 4 in hematopoiesis and angiogenesis. *Growth Factors 24*, 242–252.
149. Vandercappellen J, Noppen S, Verbeke H, Put W, Conings R, Gouwy M, Schutyser E, Proost P, Sciot R, Geboes K, Opdenakker G, Van Damme J, Struyf S (2007). Stimulation of angiostatic platelet facor-4 variant (CXCL4L1/PF-4var) versus inhibition of angiogenic granulocyte chemotactic protein-2 (CXCL6/GCP-2) in normal and tumoral mesenchymal cells. *J Leukoc Biol 82*, 1519–1530.
150. Struyf S, Burdick MD, Peeters E, van den Broeck K, Dillen C, Proost P, van Damme J, Strieter RM (2007). Platelet factor-4 variant chemokine CXCL4L1 inhibits melanoma and lung carcinoma growth and metastsis by preventing angiogenesis. *Cancer Res 67*, 5940–5948.
151. Vicari AP, Ait-Yahia S, Chemin K, Mueller A, Zlotnik A, Caux C (2000). Antitumor effects of the mouse chemokine 6Ckine/SLC through angiostatic and immunological mechanisms. *J Immunol 165*, 1992–2000.
152. Tak PP (2006). Chemokine inhibition in inflammatory arthritis. *Best Pract Res Clin Rheumatol 20*, 929–939.
153. Loetscher P, Dewald B, Baggiolini M, Seitz M (1994). Monocyte chemoattractant protein 1 and interleukin 8 production by rheumatoid synoviocytes: Effects of antirheumatic drugs. *Cytokine 6*, 162–170.
154. Lopez-Armada MJ, Sanchez-Pernaute O, Largo R, Diez-Ortego I, Palacios I, Egido J, Herrero-Beaumont G (2002). Modulation of cell recruitment by anti-inflammatory agents in antigen-induced arthritis. *Ann Rheum Dis 61*, 1027–1030.
155. Volin MV, Campbell PL, Connors MA, Woodruff DC, Koch AE (2002). The effect of sulfasalazine on rheumatoid arthritis synovial tissue chemokine production. *Exp Mol Pathol 73*, 84–92.
156. Ho CY, Wong CK, Li EK, Tam LS, Lam CW (2003). Suppressive effect of combination treatment of leflunomide and methotrexate on chemokine expression in patients with rheumatoid arthritis. *Clin Exp Immunol 133*, 132–138.
157. Ellingsen T, Hornung N, Moller BK, Poulsen JH, Stengaard-Pedersen K (2007). Differential effect of methotrexate on the increased CCR2 density on circulating CD4 T lymphocytes and monocytes in active chronic rheumatoid arthritis, with down regulation only on monocytes in responders. *Ann Rheum Dis 66*, 151–157.
158. Taylor PC, Peters AM, Paleolog E, Chapman PT, Elliott MJ, McCloskey R, Feldmann M, Maini RN (2000). Reduction of chemokine levels and leukocyte traffic to joints by tumor necrosis factor alpha blockade in patients with rheumatoid arthritis. *Arthritis Rheum 43*, 38–47.
159. Klimiuk PA, Sierakowski S, Domyslawska I, Chwiecko J (2006). Regulation of serum chemokines following infliximab therapy in patients with rheumatoid arthritis. *Clin Exp Rheumatol 24*, 529–533.
160. Torikai E, Kageyama Y, Suzuki M, Ichikawa T, Nagano A (2006). The effect of infliximab in chemokines in patients with rheumatoid arthritis. *Clin Rheumatol 26*, 1088–1093.
161. Kageyama Y, Torikai E, Nagano A (2006). Anti-tumor necrosis factor.alpha antibody treatment reduces serum CXCL16 levels in patients with rheumatoid arthritis. *Rheumatol Int 27*, 467–472.
162. Nissinen R, Leirisalo-Repo M, Peltomaa R, Palosuo T, Vaarala A (2004). Cytokine and chemokine receptor profile of peripheral blood mononuclear cells during treatment with infliximab in patients with active rheumatoid arthritis. *Ann Rheum Dis 63*, 681–687.
163. Aeberli D, Seitz M, Juni P, Villiger PM (2005). Increase of peripheral CXCR3 positive T lymphocytes upon treatment of RA patients with TNF-alpha inhibitors. *Rheumatology (Oxford) 44*, 172–175.
164. van der Voort R, van Lieshout AW, Toonen LW Sloetjes AW, van den Berg WB, Figdor CG, Radstake TR, Adema GJ (2005). Elevated CXCL16 expression by synovial macrophages recruits memory T cells into rheumatoid joints. *Arthritis Rheum 52*, 1381–1391.
165. Newton SM, Mackie SL, Martineau AR, Wilkinson KA, Kampmann B, Fischer C, Dutta S, Levin M, Wilkinson RJ, Pasvol G (2008). Reduction of chemokine secretion in response to mycobacteria in infliximab-treated patients. *Clin Vaccine Immunol 15*, 506–512.
166. Sato M, Miyazaki T, Nagaya T, Murata Y, Ida N, Maeda K, Seo H (1996). Antioxidants inhibit tumor necrosis factor-alpha mediated stimulation of interleukin-8, monocyte chemoattractant protein-1, and collagenase expression in cultured human synovial cells. *J Rheumatol 23*, 432–438.
167. Yokota K, Miyazaki T, Hirano M, Akiyama Y, Mimura T (2006). Simvastatin inhibits production of interleukin-6 (IL-6) and IL-8 and cell proliferation induced by tumor necrosis factor-alpha in fibroblast-like synoviocytes from patients with rheumatoid arthritis. *J Rheumatol 33*, 463–471.
168. Wang Y, Wei D, Lai Z, Le Y (2006). Triptolide inhibits CC chemokines expressed in rat adjuvant-induced arthritis. *Int Immunopharmacol 6*, 1825–1832.
169. Ahmed S, Pakozdi A, Koch AE (2006). Regulation of interleukin-1β-induced chemokine production and matrix metalloproteinase 2 activation by epigallocatechin-3-gallate in rheumatoid arthritis synovial fibroblasts. *Arthritis Rheum 54*, 2393–2401.

170. Ospelt C, Kurowska/Stolarska M, Neidhart M, Michel BA, Gay RE, Laufer S, Gay S (2008). The dual inhibitor of lipoxygenase and cyclooxygenase ML3000 decreases the expression of CXCR3 ligands. *Ann Rheum Dis 67*, 524–529.
171. Akahoshi T, Endo H, Kondo H, Kashiwazaki S, Kasahara T, Mukaida N, Harada A, Matsushima K (1994). Essential involvement of interleukin-8 in neutrophil recruitment in rabbits with acute experimental arthritis induced by lipopolysaccharide and interleukin-1. *Lymphokine Cytokine Res 13*, 113–116.
172. Halloran MM, Woods JM, Strieter RM, Szekanecz Z, Volin MV, Hosaka S, Haines GK III, Kunkel SL, Burdick MD, Walz A, Koch AE (1999). The role of an epithelial neutrophil-activating peptide-78-like protein in rat adjuvant-induced arthritis. *J Immunol 162*, 7492–7500.
173. Kasama T, Strieter RM, Lukacs NW, Lincoln PL, Burdick MD, Kunkel SL (1995). Interleukin-10 expression and chemokine regulation during the evolution of murine type II collagen-induced arthritis. *J Clin Invest 95*, 2868–2876.
174. Ogata H, Takeya M, Yoshimura T, Takagi K, Takahashi K (1997). The role of monocyte chemoattractant protein-1 (MCP-1) in the pathogenesis of collagen-induced arthritis in rats. *J Pathol 182*, 106–114.
175. Shahrara S, Proudfoot AE, Park CC, Volin MV, Haines GK, Woods JM, Aikens CH, Handel TM, Pope RM (2008). Inhibition of monocyte chemoattractant protein-1 ameliorates rat adjuvant-induced arthritis. *J Immunol 180*, 3447–3456.
176. Gong JH, Ratkay LG, Waterfield JD, Clark-Lewis I (1997). An antagonist of monocyte chemoattractant protein 1 (MCP-1) inhibits arthritis in the MRL-lpr mouse model. *J Exp Med 186*, 131–137.
177. Barnes DA, Tse J, Kaufhold M, Owen M, Hesselgesser J, Strieter R, Horuk R, Perez HD (1998). Polyclonal antibody directed against human RANTES ameliorates disease in the Lewis rat adjuvant-induced arthritis model. *J Clin Invest 101*, 2910–2919.
178. Inoue T, Yamashita M, Higaki M (2001). The new antirheumatic drug KE-298 suppresses MCP-1 and RANTES production in rats with adjuvant-induced arthritis and in IL-1-stimulated synoviocytes of patients with rheumatoid arthritis. *Rheumatol Int 20*, 149–153.
179. Nanki T, Urasaki Y, Imai T, Nishimura M, Muramoto K, Kubota T, Miyasaka N (2004). Inhibition of fractalkine ameliorates murine collagen-induced arthritis. *J Immunol 173*, 7010–7016.
180. Gong JH, Yan R, Waterfield JD, Clark-Lewis I (2004). Post-onset inhibition of murine arthritis using combined chemokine antagonist therapy. *Rheumatology 43*, 39–42.
181. Matsukawa A, Yoshimura T, Fujiwara K, Maeda T, Ohkawara S, Yoshinaga M (1999). Involvement of growth-related protein in LPS-induced rabbit arthritis. *Lab Invest 79*, 591–600.
182. Podolin PL, Bolognese BJ, Foley JJ, Schmidt DB, Buckley PT, Widdowson KL, Jin Q, White JR, Lee JM, Goodman RB (2002). A potent and selective nonpeptide antagonist of CXCR2 inhibits acute and chronic models of arthritis in the rabbit. *J Immunol 169*, 6435–6444.
183. Cunha TM, Barsante MM, Guerrero AT, Verri WA Jr, Ferreira SH, Coelho FM, Bertini R, Di Giacinto C, Allegretti M, Cunha FQ, Teixeira MM (2008). Treatment with DF 2162, a non-competitive allosteric inhibitor of CXCR1/2, diminishes neutrophil influx and inflammatory hypernociception in mice. *Br J Pharmacol 154*(4), 460–470.
184. Barsante MM, Cunha TM, Allegretti M, Cattani F, Policani F, Bizzarri C, Tafuri WL, Poole S, Cunha FQ, Bertini R, Teixeira MM (2008). Blockade of the chemokine receptor CXCR2 ameliorates adjuvant-induced arthritis in rats. *Br J Pharmacol 153*, 992–1002.
185. Mohan K, Issekutz TB (2007). Blockade of the chemokine receptor CXCR3 inhibits T cell recruitment to inflamed joints and decreases the severity of adjuvant arthritis. *J Immunol 179*, 8463–8469.
186. Khan A, Greenman J, Archibald SJ (2007). Small molecule CXCR4 chemokine receptor antagonists: Developing drug candidates. *Curr Med Chem 14*, 2257–2277.
187. Hatse S, Princen K, Bridger G, DeClerq E, Schols D (2002). Chemokine receptor inhibition by AMD3100 is strictly confined to CXCR4. *FEBS Lett 527*, 255–262.
188. Tamamura H, Fujii N (2005). The therapeutic potential of CXCR4 antagonists in the treatment of HIV infection, cancer metastasis and rheumatoid arthritis. *Expert Opin Ther Targets 9*, 1267–1282.
189. Pease JR, Horuk R (2005). CCR1 antagonists in clinical development. *Expert Opin Invest Drugs 14*, 785–796.
190. Gladue RP, Tylaska LA, Brissette WH, Lira PD, Kath JC, Poss CS, Brown MF, Paradis TJ, Conklyn MJ, Ogborne KT (2003). CP-481,715, a potent and selective CCR1 antagonist with potential therapeutic implications for inflammatory diseases. *J Biol Chem 278*, 40473–40480.
191. Haringman JJ, Kraan MC, Smeets TJM, Zwinderman KH, Tak PP (2003). Chemokine blockade and chronic inflammatory disease: Proof of concept in patients with rheumatoid arthritis. *Ann Rheum Dis 62*, 715–721.
192. Quinones MP, Estrada CA, Kalkonde Y, Ahuja SK, Kuziel WA, Mack M, Ahuja SS (2005). The complex role of the chemokine receptor CCR2 in collagen-induced arthritis: Implications for therapeutic targeting of CCR2 in rheumatoid arthritis. *J Mol Med 83*, 672–681.
193. Quinones MP, Jimenez F, Martinez H, Estrada CA, Willmon O, Dudley M, Kuziel WA, Melby PC, Reddick RL, Ahuja SK (2006). CC chemokine receptor (CCR)-2 prevents arthritis development following infection by *Mycobacterium avium*. *J Mol Med 84*, 503–512.
194. Amat M, Benjamim CF, Williams LM, Prats N, Terricabras E, Beleta J, Kunkel SL, Godessart N (2006). Pharmacological blockade of CCR1 ameliorates murine arthritis and alters cytokine networks in vivo. *Br J Pharmacol 149*, 666–675.
195. Plater-Zyberk C, Hoogewerf AJ, Proudfoot AE, Power CA, Wells TN (1997). Effect of a CC chemokine receptor

antagonist on collagen induced arthritis in DBA/1 mice. *Immunol Lett 57*, 117–120.

196. Shahrara S, Proudfoot AE, Woods JM Ruth JH, Amin MA, Park CC, Haas CS, Pope RM, Haines GK, Zha YY (2005). Amelioration of rat adjuvant-induced arthritis by Met-RANTES. *Arthritis Rheum 52*, 1907–1919.
197. Yang YF, Mukai T, Gao P Yamaguchi N, Ono S, Iwaki H, Obika S, Imanishi T, Tsujimura T, Hamaoka T (2002). A non-peptide CCR5 antagonist inhibits collagen-induced arthritis by modulating T cell migration without affecting any collagen T cell responses. *Eur J Immunol 32*, 2124–2132.
198. Quinones MP, Ahuja SK, Jimenez F Schaefer J, Garavito E, Rao A, Chenaux G, Reddick RL, Kuziel WA, Ahuja SS (2004). Experimental arthritis in CC chemokine receptor 2-null mice closely mimics severe human rheumatoid arthritis. *J Clin Invest 113*, 856–866.
199. Brühl H, Cihak J, Plachy J, Kunz-Schughart L, Niedermeier M, Denzel A, Rodriguez-Gomez M, Talke Y, Luckow B, Stangassinger M, Mack M (2007). Targeting of Gr-1+, CCR2+ monocytes in collagen-induced arthritis. *Arthritis Rheum 56*, 2975–2985.
200. Haringman JJ, Gerlag DM, Smeets TJ, Baeten D, van den Bosch F, Bresnihan B, Breedveld FC, Dinant HJ, Legay F, Gram H (2006). A randomized, controlled trial with an anti-CCL2 (anti-monocyte chemotactic protein 1) monoclonal antibody in patients with rheumatoid arthritis. *Arthritis Rheum 54*, 2387–2392.
201. Clucas AT, Shah A, Zhang YD, Chow VF, Gladue RP (2007). Phase I evaluation of the safety, pharmacokinetics and pharmacodynamics of CP-481,715. *Clin Pharmacokinet 46*, 757–766.
202. Meanwell NA, Kadow JF (2007). Maraviroc, a chemokine CCR5 receptor antagonist for the treatment of HIV infection and AIDS. *Curr Opin Invest Drugs 8*, 669–681.
203. Takahashi H, Kato K, Miyake K, Hirai Y, Yoshino S, Shimada T (2005). Adeno-associated virus vector-mediated anti-angiogenic gene therapy for collagen-induced arthritis in mice. *Clin Exp Rheumatol 23*, 455–461.
204. Sumariwalla PF, Cao Y, Wu HL, Feldmann EM, Paleolog E (2003). The angiogenesis inhibitor protease-activated kringles 1-5 reduces the severity of murine collagen-induced arthritis. *Arthritis Res Ther 5*, R32–R39.
205. O'Reilly MS, Boehm T, Shing Y, Fukai N, Vasios G, Lane WS, Flynn E, Birkhead JR, Olsen BR, Folkman J (1997). Endostatin: An endogenous inhibitor of angiogenesis and tumor growth. *Cell 88*, 277–285.
206. Yue L, Shen YX, Feng LJ, Chen FH, Yao HW, Liu LH, Wu Q, Wang H (2007). Blockage of the formation of new blood vessels by recombinant human endostatin contributes to the regression of rat adjuvant arthritis. *Eur J Pharmacol 567*, 166–170.
207. Wang CR, Chen SY, Shiau AL, Wu CL, Jou IM, Chao L, Chao J (2007). Upregulation of kallistatin expression in rheumatoid joints. *J Rheumatol 34*, 2171–2176.
208. Mundel TM, Kalluri R (2007). Type IV collagen-derived angiogenesis inhibitors. *Microvasc Res 74*, 85–89.
209. Rico MC, Castaneda JL, Manns JM, Uknis AB, Sainz IM, Safadi FF, Popoff SN, Dela Cadena RA (2007). Amelioration of inflammation, angiogenesis and CTGF expression in an arthritis model by a TSP-1-derived peptide treatment. *J Cell Physiol 211*, 504–512.
210. Park YW, Kang YM, Butterfield J, Detmar M, Goronzy JJ, Weyand CM (2004). Thrombospondin 2 functions as an endogenous regulator of angiogenesis and inflammation in rheumatoid arthritis. *Am J Pathol 165*, 2087–2098.
211. Haas CS, Amin MA, Allen BB, Ruth JH, Haines GK 3rd, Woods JM, Koch AE (2006). Inhibition of angiogenesis by interleukin-4 gene therapy in rat adjuvant-induced arthritis. *Arthritis Rheum 54*, 2402–2414.
212. Ingber D, Fujita T, Kishimoto S, Sudo K, Kanamaru T, Brem H, Folkman J (1990). Synthetic analogues of fumagillin that inhibit angiogenesis and suppress tumour growth. *Nature 348*, 555–557.
213. Peacock DJ, Banquerigo ML, Brahn E (1992). Angiogenesis inhibition suppresses collagen arthritis. *J Exp Med 175*, 1135–1138.
214. Hannig G, Bernier SG, Hoyt JG, Doyle B, Clark E, Karp RM, Lorusso J, Westlin WF (2007). Suppression of inflammation and structural damage in experimental arthritis through molecular targeted therapy with PPI-2458. *Arthritis Rheum 56*, 850–860.
215. Bainbridge J, Madden L, Essex D, Binks M, Malhotra R, Paleolog EM (2007). Methionine aminopeptidase-2 blockade reduces chronic collagen-induced arthritis: Potential role for angiogenesis inhibition. *Arthritis Res Ther 9*, R127.
216. D'Amato RJ, Loughnan MS, Flynn E, Folkman J (1994). Thalidomide is an inhibitor of angiogenesis. *Proc Natl Acad Sci USA 91*, 4082–4085.
217. Komorowski J, Jerczyńska H, Siejka A, Barańska P, Ławnicka H, Pawłowska Z, Stepień H (2006). Effect of thalidomide affecting VEGF secretion, cell migration, adhesion and capillary tube formation of human endothelial EA.hy926 cells. *Life Sci 78*, 2558–2663.
218. Oliver SJ, Cheng TP, Banquerigo ML, Brahn E (1998). The effect of thalidomide and two analogs on collagen induced arthritis. *J Rheumatol 25*, 964–969.
219. Yeo EJ, Chun YS, Cho YS, Kim J, Lee JC, Kim MS, Park JW (2003). YC-1: A potential anticancer drug targeting hypoxia-inducible factor 1. *J Natl Cancer Inst 95*, 516–525.
220. Kiselyov A, Balakin KV, Tkachenko SE (2007). VEGF/VEGFR signaling as a target for inhibiting angiogenesis. *Expert Opin Invest Drugs 16*, 83–107.
221. Grosios K, Wood J, Esser R, Raychaudhuri A, Dawson J (2004). Angiogenesis inhibition by the novel VEGF tyrosine kinase inhibitor, PTK787/ZK222584, causes significant anti-arthritic effects in models of rheumatoid arthritis. *Inflamm Res 53*, 133–142.

222. Guttmann-Raviv N, Shraga-Heled N, Varshavsky A, Guimaraes-Sternberg C, Kessler O, Neufeld G (2007). Semaphorin-3A and semaphorin-3F work together to repel endothelial cells and to inhibit their survival by induction of apoptosis. *J Biol Chem 282*, 26294–26305.

223. Chen Y, Donnelly E, Kobayashi H, Debusk LM, Lin PC (2005). Gene therapy targeting the Tie2 function ameliorates collagen-induced arthritis and protects against bone destruction. *Arthritis Rheum 52*, 1585–1594.

224. Goedkoop AY, Kraan MC, Picavet DI, de Rie MA, Teunissen MB, Bos JD, Tak PP (2004). Deactivation of endothelium and reduction in angiogenesis in psoriatic skin and synovium by low dose infliximab therapy in combination with stable methotrexate therapy. *Arthritis Res Ther 6*, R326–R334.

225. Bongartz T, Coras B, Vogt T, Schölmerich J, Müller-Ladner U (2005). Treatment of active psoriatic arthritis with the PPARγ ligand pioglitazone: An open-label pilot study. *Rheumatology 44*, 126–129.

# DEVELOPMENT OF THERAPEUTIC AGENTS: METHODS AND APPROACHES IN THE DEVELOPMENT OF VACCINES AGAINST PROTOZOAN PARASITES

Catherine Ivory and Kris Chadee
*Faculty of Medicine, Department of Microbiology and Infectious Diseases, University of Calgary, Calgary, Alberta, Canada*

## 1 INTRODUCTION

Parasitic infections make up a large percent of diseases in tropical countries and pose a significant threat to human health worldwide. Parasites are complex pathogens associated with chronic disease, high reinfection rates, and host immune depression. Furthermore, parasitic infections, unlike most viral and bacterial infections, are more complex and chronic, requiring special consideration when developing therapeutic agents. Designing strategies against parasitic infections demands that scientists address the multifactorial aspects of special pathogens. The challenge in developing therapeutic agents against parasites lies in the nature of the parasites and their close relationship with both human immunity and anthropogenic activities.

## 2 PARASITE LIFE CYCLES AND VECTORS

The leading cause of death from parasitic disease, malaria, presents unique obstacles for scientists trying to develop effective therapies [1]. *Plasmodium falciparum,* the etiological agent of malaria, is transmitted to humans via the *Anopheles gambiae* mosquito. Sporozoites are introduced into the human host and enter a hepatic cycle before being released into the blood for the erythrocytic cycle (Fig. 1). After a blood meal, the mosquito vector will ingest parasite gametocytes, which undergo a sporogonic cycle (sexual cycle) giving rise to infective sporozoites that perpetuate the life cycle. Looking at the malaria life cycle, one can identify two main obstacles: (1) the parasite develops within two different hosts and different cells, and (2) there are multiple immunologically distinct parasite stages. For this reason, scientists have concentrated their efforts on formulating a vaccine targeted at multiple stages and multiple antigens to mimic the array of immunological responses activated during a natural infection [2]. The strategy here is to find optimal antigens from different parasite stages that could mount proper immune responses at various stages in parasite development. Another problem with malarial disease is that different stages require different immunological responses to effectively eradicate the infective parasite [3]. Strong antibody responses are required to fight blood stage sporozoites as well as gametocyte transmission to the mosquito. In the liver stage, however, a strong cytotoxic T-cell-mediated response has been associated with clearance of infected host cells. Combining the need for a multiantigen approach with the need to stimulate both humoral and cell-mediated

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

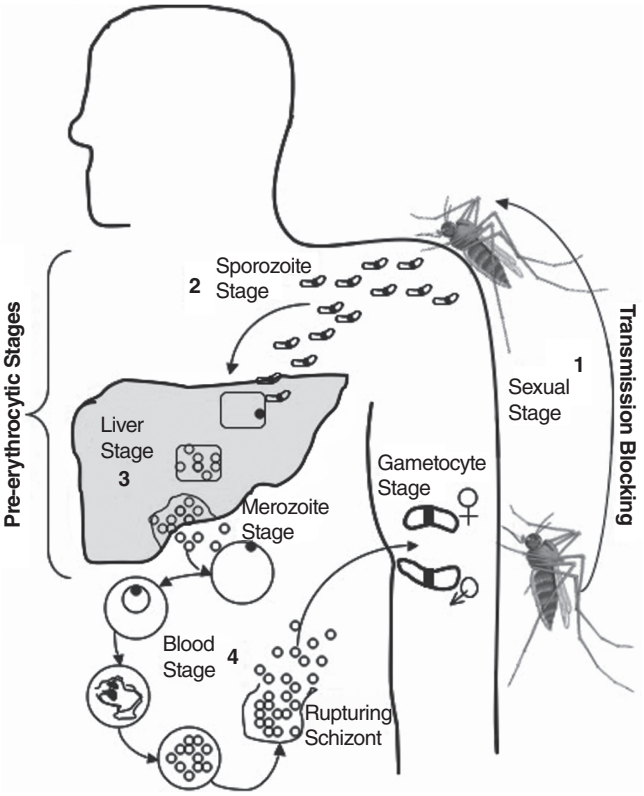


**Figure 1** Malaria life cycle. During a blood meal an infected *Anopheles* mosquito inoculates the human host with sporozoites (2), which infect liver cells and mature into schizonts (3). Merozoites are released from the liver cells and infect erythrocytes where they undergo asexual multiplication (4). Some parasites differentiate into male or female gametocytes that are then taken up by another mosquito vector. Within the mosquito, gametocytes undergo sexual reproduction (1) forming zygotes that become oocysts releasing sporozoites. The sporozoites make their way to the salivary glands and once inoculated into a new host, the cycle is perpetuated. Vaccines for disease prevention would target pre-erythrocytic stages (2, 3), while disease reduction could be achieved with a blood-stage vaccine (4). Transmission-blocking vaccines (1) would prevent infection of a new host.

immunity, vaccinologists have turned to modern molecular and proteomic tools to address this complicated infectious disease. Proteomics has allowed scientist to identify the full repertoire of parasite proteins and the changes in protein expression with specific parasite stages. This facilitates the selection of vaccine candidates, which should be expressed on the cell surface and also serve an essential biological function. Once candidates have been chosen, these genes can be incorporated into vaccine DNA (deoxyribonucleic acid) vectors or a combination of viral vectors to induce strong T- and B-cell responses.

## 3 HOST IMMUNITY AND PARASITE ANTIGENIC VARIATION

The malaria parasite has evolved clever ways to evade the immune system or change the immune response in favor of its survival. Antigenic variation is believed to be associated with the chronic infections seen in endemic malarial regions, due to a sequential variation in polymorphic antigens expressed on the surface of infected erythrocytes [4]. One such protein is the variant surface antigen, *P. falciparum* erythrocyte protein I (PfEM1), which has been the target of many vaccination studies [5]. PfEMP1 is a surface protein involved in parasite adhesion to an array of human receptors. This molecule helps sequester the parasite in postcapillary venules, avoiding the spleen-mediated filtration of infected red blood cells. Correspondingly, by continuously changing its surface expression, the parasite evades the attack from host immune cells. There are 59 PfEMP1 variants, and each parasite expresses one variant during each cycle to improve survival. Anti-PfEMP1 immunity is known to be protective against severe forms of the disease, including cerebral and placental malaria [6]. Although PfEMP1 is a protein, which undergoes antigen variation, it may be a principle target for developing natural immunity against the parasite. A blood-stage vaccine targeted at PfEMP1 would need to induce anti-adhesion antibodies to prevent placental malaria and would have to be panreactive to protect against a range of isolates in the parasite population.

## 4 MULTIPLE STRATEGIES, DIFFERENT OUTCOMES

The complexity of the malaria life cycle introduces many possible vaccine targets, and the effects of the immunization will depend on the antigen selected. The *Plasmodium* life cycle could be interrupted at various stages (Fig. 1), and scientists have focused on three major vaccines types: (1) transmission-blocking vaccines, (2) pre-erythrocytic vaccines, and (3) blood-stage vaccines.

Transmission-blocking vaccines aim to prevent the formation of oocysts in the mosquito, thereby blocking the formation of the infective sporozoites. The prime candidate proteins for this type of vaccine are Pfs25 and Pfs28 found on the *Plasmodium* zygote [7]. The strategy in the transmission-blocking vaccine is to ensure mosquitoes acquire the antibodies against Pfs25/Pfs28 during a blood meal on an immunized human. The antibodies would then mediate complement attack on the gametocytes or block fertilization by agglutination. In conjunction with conventional malaria control pro-

grams, the transmission-blocking vaccine would reduce the incidence of new infections.

Pre-erythrocytic vaccines fall into two categories, those that target the sporozoites stages and those that target the liver stages, each of which require different immune responses. The circumsporozoite protein-1 (CSP-1) is the target of vaccines designed to induce neutralizing antibodies that would prevent sporozoites from infecting liver cells and entering the hepatic cycle [8]. This step is crucial as one sporozoite can give rise to 40,000 merozoites from which symptoms of malaria occur. On the other hand, a vaccine targeted at the liver stage would require strong cell-mediated immune responses to effectively eliminate the intracellular parasite stage. The most promising antigen target for this stage is the liver-stage antigen-1 (LSA-1) expressed on the surface of the infected hepatocyte [9]. Unlike other vaccine candidates, the LSA-1 sequence seems to be conserved between different parasite strains. In order to induce strong T-cell-mediated immunity against liver-stage malaria, vaccinologists have engaged the use of heterologous prime–boost systems [10]. The strategy in this case is to first prime the immune system with a bacterial plasmid encoding the antigen of choice and later boost the patient with a viral vector also encoding the same protein. In this way, the immune response is targeted to the encoded protein and not to the vector components. Modified viral vectors, which are nonreplicating and safe, are excellent for boosting as they induce high protein expression and presentation via major histocompatability complex (MHC) class I, thus amplifying the antigen-specific T-cell response. The principles of the strategy are to enhance T-cell-mediated responses and to generate memory T cells to the vaccine antigen. Clinical trials with RTS,S/ASO2D, encoding the CSP-1, have demonstrated that the prime–boost method was safe, tolerated, and immunogenic in young children [11].

The last type of malaria vaccine is a blood-stage vaccine, targeting the intracellular erythrocytic stage. The parasite resides inside the red blood cell and is hidden from the immune system. Since erythrocytes do not express MHC class I or II molecules, the desired immune response at this stage is antibody mediated. Vaccine trials for blood-stage malaria target the apical membrane antigen (AMA-1) or the merozoite surface protein (MSP-1). Antigen-specific antibodies mediate antibody-dependent cytotoxicity and complement activation [12]. Immune responses at this stage of the parasite cycle are designed to reduce symptoms of disease, while pre-erythrocytic vaccines are aimed at disease prevention. It is important to consider the interaction of the human immune system with the malaria parasite, as different stages induce different responses and require special attention while developing a vaccination strategy. Perhaps the ultimate vaccine would incorporate antigens to all these stages and induce strong antibody as well as cell-mediated immunity to the parasite.

## 5 *ENTAMOEBA HISTOLYTICA*: THE IDEAL PARASITE?

Parasites include a diverse array of organisms, ranging from multicellular worms to microscopic single-celled amebas. In fact, parasite biology can range from relatively simple demands to complex multiple-host life cycles. Unlike the complexity seen with the malaria parasite, *E. histolytica* presents a simple survival scheme. *Entamoeba histolytica* is an enteric protozoan parasite responsible for the human disease known as amebiasis. The disease is acquired through the ingestion of an infective cyst stage found in contaminated food or water. The cyst migrates to the large intestine where it gives rise to the mobile trophozoite stage responsible for the onset of amebic colitis (Fig. 2). In some cases of amebic infection the ameba can breach the intestinal lining and disseminate via portal circulation to the liver,

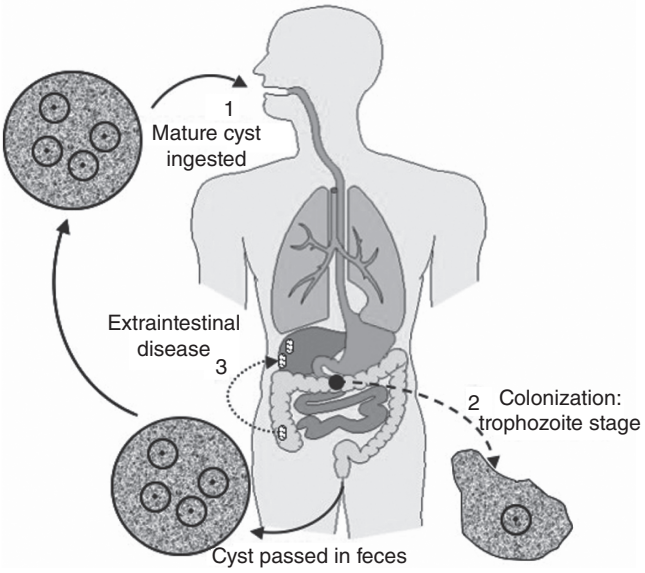


**Figure 2** The *E. histolytica* life cycle. The infective cyst stage is ingested by a human via contaminated food or water (1). The cysts in the gut give rise to mobile trophozoites, which colonizes the large intestine (2). In the event of invasive amebiasis, the parasite breaches the gut lining and is disseminated to the liver, causing amebic liver abscess (3). New cysts are formed in the colon and passed into the environment with feces, thus perpetuating the life cycle. An amebiasis vaccine would target the trophozoite stage (2), inhibiting parasite colonization with adherence blocking antibodies.

where they can cause contained amebic liver abscesses. Humans are the only relevant hosts for this parasite, and, unlike malaria, this organism cycles between two stages, with only the trophozoites being pathogenic. *Plasmodium* is an intracellular parasite, which has an array of immune evasion mechanisms, while *Entamoeba* is an extracellular pathogen, with no known antigenic variation or obvious immune evasion methods. Consequently, the focus of therapies against amebiasis is aimed at the trophozoite stage, using chemotherapy that can effectively kill the parasite and the development of vaccines to block trophozoite colonization.

Amebiasis is a cosmopolitan disease; however, it is endemic in developing countries and afflicts an estimated 50 million people [13]. Although chemotherapy (metronidazole) against amebiasis is effective, access to treatment is limited and infections rates remain high in these endemic regions. There is a need to develop preventative or prophylactic therapies to decrease the incidence of disease and mortality. Amebiasis could be eradicated with the implementation of proper sanitation in these endemic areas; however, the social changes and public health interventions required for such a task are not foreseeable in the near future. Based on these limitations, vaccination is a more plausible and cost-effective way of preventing amebiasis. When addressing the possibility of a vaccine against an infectious agent such as *E. histolytica,* it is important to consider several key features of the disease: (1) the existence of natural acquired immunity, (2) the pathogenesis of disease, and (3) the prevalence of amebiasis and the target population.

## 6 IMMUNITY TO *E. HISTOLYTICA*

It is necessary to determine the hallmarks of natural immunity to *E. histolytica* in humans to design a vaccination strategy with the desired immune indicators. The distinct clinical manifestations of disease with *E. histolytica*, mainly amebic colitis and amebic liver abscess, also have distinct immune responses. Amebic colitis is characterized by a prominent pro-inflammatory response, whereas amebas in the liver do not cause hepatitis but an abscess surrounded by connective tissue to segregate healthy liver parenchyma from diseased areas. Despite this pathophysiological difference, there are many common immune components active during amebic colitis and amebic liver abscess.

The parasite first encounters intestinal epithelial cells, and, therefore, gut inflammation in amebic colitis is the result of epithelial cell release of pro-inflammatory cytokines interleukin-8 (IL-8) and IL-1β [14]. These mediators can attract neutrophils and macrophages to the site of infection. Neutrophils are also the first host defense cell to contact amebas in the liver, indicating recruitment of these cells to the site of infection [15]. In the case of amebic colitis, however, it would appear that neutrophil recruitment exacerbates inflammation through the release of their toxic contents [16]. Although inflammation in amebic colitis definitely contributes to tissue damage, it may also be required to limit infection.

Macrophages have been associated with cell-mediated immunity against *E. histolytica*. The activation of these effector cells depends on environmental cues and the inflammatory stimuli. The major priming signal for macrophage activation is interferon-γ (IFN-γ). Naïve bone-marrow-derived macrophages exposed to *E. histolytica* proteins demonstrate increased expression of *c-fos*, which is involved in activation, and increased tumor necrosis factor-α (TNF-α) messenger ribonucleic acid (mRNA) [17]. Macrophages can effectively kill amebas in vitro in a nitric oxide (NO)-dependent manner [18]. TNF-α production by activated macrophages can enhance inducible nitric oxide synthase (iNOS) mRNA for increased NO production [19]. The ability of *E. histolytica* proteins to stimulate TNF-α could be central in activating primed macrophages for amebicidal activity. In the case of amebic liver abscess formation, IFN-γ is important to prime macrophages for activation, and iNOS is essential to mount a proper amebicidal response. In fact, it has recently been reported that higher IFN-γ production is correlated to a reduced risk of amebic disease in humans [20]. IFN-γ is a key pro-inflammatory cytokine regulating macrophage activity against the invading parasite.

Most importantly, in terms of human immunity to *E. histolytica*, the presence of antilectin mucosal IgA has been associated with a reduced risk of infection in endemic areas [21]. Recent epidemiological studies have illustrated the importance of mucosal immunity in resistance to *E. histolytica* infection. These studies were conducted in preschool Bangladeshi children where there is a 55% *E. histolytica* infection rate and colitis is diagnosed in 4% of these cases of infection [22]. These studies determined that children with antiamebic stool IgA had 64% fewer infections, and the occurrence of these antibodies was thought to prevent parasite colonization [21]. In most cases the antiamebic stool IgA was detectable near the time of resolution. While the presence of stool IgA was associated with protection from reinfection, the absence of serum IgG against the parasite was also associated with innate resistance to *E. histolytica* infections [22]. Upon closer examination it was determined that the presence of stool IgA specifically against the Gal-lectin was protective, associated with 86% less new infections over the following year.

Protective immunity after amebic infection appears to be mediated by mucosal IgA against *E. histolytica*. It is not surprising that mucosal IgA antibodies directed against the parasite provide protection since the disease is acquired in the gastrointestinal tract. These findings are essential and determined that acquired immunity in humans is mucosal. A vaccine for amebiasis should logically try to stimulate the gut-associated lymphoid tissue and induce sufficient mucosal responses, specifically secretory IgA against the parasite. It is also important to consider the role of the inflammatory response in amebiasis and design a vaccine that could provide both the mucosal antibodies and protective Th1 cytokines associated with reduction of disease.

## 7 PATHOGENESIS OF AMEBIASIS

The pathogenesis of amebiasis is a complicated process involving three distinct stages and multiple parasite virulence factors. The parasite is acquired by the ingestion of the cyst stage in contaminated food or water. However, infection and disease progression requires these three stages: colonization, mucus depletion, and damage of host cells. The parasite must overcome the innate mucosal defenses of the host, including both physical barriers and immune mechanisms.

The first step in pathogenesis of amebiasis is the colonization of the host's colon by the mobile amebic trophozoites. The human gastrointestinal tract is protected by a gelatinous mucus layer, which is composed of carbohydrate-rich mucin molecules. A single molecule, the Gal-lectin, mediates colonization by the parasite to the mucin layer. The Gal-lectin binds to galactose and *N*-acetyl-D-galactosamine residues of colonic mucins [23]. The Gal-lectin is a major surface protein that contains the carbohydrate recognition domain involved in adherence to the terminal sugar moieties of colonic mucins and host cells [24]. Adherence to colonic mucins allows the parasite to multiply and prevents it from being carried away, thus establishing infection. At the same time, however, mucins provide a protective role, preventing parasite contact with underlying epithelial cells.

In the second stage, the parasite overcomes this physical barrier by both depletion and destruction of mucins. Colonic mucins are produced by goblet cells in the intestine, and it is believed that *E. histolytica* can induce hypersecretion in these cells. The parasite can induce significant mucous secretion from goblet cells. In fact, studies in gerbils suggested that the trophozoites release a mucous secretagogue, which would lead to the eventual depletion of mucin secretion from goblet cells [25]. In conjunction with mucous depletion is the process of mucous destruction by the parasite, which occurs through the activity of cysteine proteases. It has specifically been shown that *E. histolytica* cysteine protease 5 (EhCP5) can directly cleave the C-terminus of the MUC2 polymer, thus destroying the gel-forming properties of mucins [26].

Upon removal of this protective mucous barrier, the parasite comes in contact with host intestinal epithelial cells. As mentioned above, parasite adherence is mediated by the Gal-lectin, and so this molecule is also required for binding to host cells. Adherence to host cells is followed by cytolysis, thus leading to tissue destruction from which the parasite gets its name. The importance of Gal-lectin-mediated cell contact has been illustrated in vitro with a Chinese hamster ovary (CHO) cells deficient in terminal galactose/*N*-acetyl galactosamine residues, which were resistant to amebic cytolytic activity compared to normal cells [27]. It has also been demonstrated in vivo that Gal-lectin-deficient trophozoites are less virulent, causing very little pathology [28]. Cytotoxicty is observed within 5–15 min after parasite contact and involves both apoptosis and necrotic cell death [29, 30]. This cell death is thought to be associated with the action of amebapores [31], pore-forming proteins that are inserted by the trophozoites into the host cell membrane upon contact and cause cell lysis.

Understanding the events in the pathogenesis of amebiasis helps in vaccine design by identifying target antigen and target areas to interrupt parasite life cycle. In the events described above it is made apparent that the Gal-lectin molecule mediates all three major steps in disease progression. The important role of the Gal-lectin in *E. histolytica* biology makes this antigen a prime candidate for an amebiasis vaccine. Furthermore, amebic infection is determined by the parasite's ability to first colonize the human intestine, and blocking this step with a vaccine that induces Gal-lectin-specific mucosal antibodies would drastically reduce the incidence of invasive disease.

## 8 PREVALENCE OF DISEASE AND TARGET POPULATION

The World Health Organization estimates that *E. histolytica* infects 50 million people worldwide and leads up to 100,000 deaths a year [13]. *Entamoeba histolytica* is most prevalent in subtropical areas such as Mexico, Africa, India, and South America. Higher prevalence in those areas is due to lower standards of sanitation leading to greater contamination of food and water. All races are equally susceptible to infection, but high-risk groups include travelers to endemic areas, homosexuals [32], and immunocompromised individuals. Following

this last point, severe cases of amebiasis are more common in young children, pregnant women, and elderly individuals.

The majority of infected individuals remain asymptomatic, with the parasite residing in the large intestine as a commensal. For yet undetermined reasons, an estimated 10% of infected humans will develop symptoms of invasive disease. The factors that determine whether an infected individual will develop invasive disease are not known. A recent study conducted in an endemic area of Bangladesh suggests that there may be an association between the parasite genotype and risk of complication [33]. There may also be host factors that could predict the outcome of disease; however, these remain unidentified to date. Considering these yet unsolved mysteries regarding the onset of invasive disease, it would be most useful to test a potential vaccine candidate in the population for which it is intended. Several factors could influence the immune response to a prospective vaccine, including gut flora, environmental factors, nutrition, and concurrent or past infections with *E. histolytica* or other gut pathogens.

## 9 VACCINE DESIGN

### 9.1 Choosing a Candidate Vaccine Antigen: Gal-lectin

One of the most well-studied molecules of *E. histolytica* is the galactose/*N*-acetyl-D-galactosamine-specific lectin (Gal-lectin). Gal-lectin is a transmembrane 260-kDa heterodimer composed of a 170-kDa heavy subunit linked by disulfide bonds to a 35-kDa light subunit [34]. This molecule is a prime candidate for an amebiasis vaccine. Gal-lectin meets many criteria for a vaccine antigen. First, it is a cell surface molecule highly exposed to host immune cells and is known to have strong immunogenic properties. Second, Gal-lectin is a parasite virulence factor without which the parasite loses its pathogenicity [28]. Third, Gal-lectin is a highly conserved antigen and shows very little genetic variability. *Entamoeba histolytica* does not have a sexual cycle and therefore appears to be clonal with closely related organisms found throughout the world [35].

### 9.2 Gal-lectin Immunogenicity

Gal-lectin, specifically the heavy subunit, is a highly conserved antigen and is recognized by human immune sera from patients throughout the world [36]. In fact, 95% of sera from amebiasis patients have antibodies against Gal-lectin [37], and it is the major antigen recognized by pooled human immune serum. Most importantly, the presence of antilectin mucosal IgA has been associated with a reduced risk of infection in endemic areas [21]. Acquired resistance to *E. histolytica* infection was linked to these protective intestinal IgA antibodies that are directed against the carbohydrate recognition domain of the Gal-lectin.

In addition to humoral responses, Gal-lectin also demonstrates immunogenicity in cell-mediated responses. Peripheral blood monocytes isolated from individuals previously exposed to *E. histolytica* respond to Gal-lectin, demonstrating the potency of this antigen [38]. It was also observed that T cells treated with Gal-lectin for 5 days in vitro became amebicidal, while cells from control individuals did not. Another study looked at the specific Gal-lectin region involved in this T-cell response. These studies illustrated that there may be important T-cell epitopes on the Gal-lectin antigen responsible for the induction of cell-mediated immunity against *E. histolytica*.

Gal-lectin-mediated cell immune responses have extensively been studied in macrophages. These specialized antigen-presenting cells play an important role in innate and adaptive immune responses, specifically in pathogen elimination. Activated macrophages are able to kill *E. histolytica* trophozoites in vitro [39]. Stimulation of murine bone-marrow-derived macrophages with Gal-lectin induces the production of TNF-α and IL-12 [19, 40]. This TNF-α production is mediated by the cysteine-rich region as monoclonal antibodies (mAbs) against amino acids 596–1082 could effectively block this macrophage stimulation. In another study it was demonstrated that Gal-lectin could also induce NO production in IFN-γ-primed murine macrophages. Macrophages were amebicidal when treated simultaneously with IFN-γ and Gal-lectin, and this cytotoxic effect could be inhibited by mAbs against amino acids 596–818 of the cysteine-rich region.

### 9.3 Targeting the Right Epitopes

Gal-lectin's primary role is to mediate parasite binding to host glycoconjugates with high affinity. By this mechanism Gal-lectin mediates colonization of the human colon by binding to colonic mucin. Gal-lectin also mediates binding to target host cells and confers complement resistance via its CD59-like domain [41]. Carbohydrate recognition has been mapped to the heavy subunit of Gal-lectin. The carbohydrate recognition domain (CRD) has been determined with the use of mAbs that map to epitopes on the cysteine-rich region of the extracellular domain on the heavy subunit. These mAbs could block parasite-binding activity and correspond to amino acids 895–998 [42]. This sequence of 104 amino acids is highly conserved among products of the *hgl*

**TABLE 1 Gal-lectin Vaccine Trials and Efficacies**

| Antigen | Form[a] | Delivery | Route[b] | % Eff[c] | Reference |
|---|---|---|---|---|---|
| 260 kDa | Native | Freund's | i.p. | 67[d,e] | 53 |
| aa 649–1201 | Recomb. Gst[f] Fusion | Freund's | i.p. | 81 | 54 |
| aa 767–1138 | Recomb. His-tagged | Titermax | i.p. | 71[d] | 55 |
| aa 482–1138 | Recomb. His-tagged | Freund's | i.p. | 45[d] | 56 |
| aa 482–1138 | Recomb. GST fusion | *S. dublin* | Orally | 15.3[d] | 56 |
| aa 1–436 | Recomb. | Freund's | i.p. | 6.7[e] | 57 |
| aa 436–624 | Recomb. | Freund's | i.p. | 37.5[d] | 57 |
| aa 799–939 | Recomb. | Freund's | i.p. | 11.1 | 57 |
| aa 939–1053 | Recomb. | Freund's | i.p. | 62.5[d] | 57 |
| aa 895–998 | Recomb. His-tagged | Freund's | i.p. | 0[d] | 58 |
| aa 1005–1029 | Synthetic, KLH[f] fusion | Freund's | i.p. | 33.3[d] | 59 |
| aa 1005–1029 | Recomb. CtxB[f] fusion | None | Orally | 0–30[d] | 59 |
| aa 1005–1029 | Recomb. CtxB fusion | Freund's | i.p. | 0–55 | 59 |
| 260 kDa | Native | Ctx | i.n. | 84–100 | 60 |
| aa 578–1154 | Recomb. His-tagged | Ctx | i.n. | 34–91 | 60 |
| 260 kDa | YopE[f] fusion | *Yersinia* | Orally | NA | 61 |

[a]All recombinant proteins were produced in *Escherichia coli*.
[b]Routes i.p. = intraperitoneally, i.n. = intranasally, Orally = by oral gavage.
[c]Vaccine efficacy = (number of immunized animals without disease ÷ total number of immunized animals) − (number of control animals without disease ÷ total number of control animals) × 100.
[d]The lesions in nonprotected animals were significantly smaller than control animals.
[e]The lesions in nonprotected animals were significantly larger than control animals.
[f]GST = glutathione S-transferase; KLH = keyhole limpet hemocyanin; CtxB = cholera toxin B sub-unit; YopE = Yersinia outer protein enterocolitica.

gene family. It is obvious from these studies that carbohydrate recognition occurs within the cysteine-rich region of the heavy subunit, and this should be considered in vaccine design. During the mapping of the CRD, it was observed that certain mAbs (3F4 and 8A3) could enhance Gal-lectin-mediated binding to CHO cells [43]. The Fab fragments alone could induce this effect; therefore, it was not due to antibody clustering of the lectin. In fact, the mAb binding seemed to be changing Gal-lectin to a more active conformation. The regulation of Gal-lectin conformation is thought to be mediated by inside-out signaling. The Gal-lectin cytoplasmic tail contains regions of identity with the cytoplasmic tail of β2 intergrin-like domains, and this could mediate the inside-out signaling [44].

Conformational changes in Gal-lectin are important in designing an amebiasis vaccine, since some epitopes may be hidden while others may actually exacerbate disease by enhancing parasite adherence. Parasitologists are able to grow *E. histolytica* trophozoites in the laboratory, and isolation of the native Gal-lectin protein is possible; however, on large scale purifying native amebic proteins would not be cost effective. Expression of amebic antigens in prokaryotic or plant systems has proven to be successful and offer an alternative for large-scale vaccine production. The advent of recombinant vaccines such as hepatitis B has established large-scale production methods that could be adapted for new developing vaccines. Native Gal-lectin and recombinant forms containing the CRD have been tested as amebiasis vaccines (Table 1).

### 9.4 Animal Models of Amebiasis and Gal-Lectin Vaccine Trials

*Entamoeba histolytica* is acquired through the ingestion of the infective cyst stage from which arise the mobile trophozoites that cause disease in humans. Unfortunately, to date there is no laboratory method to cultivate this infective cyst stage, and, therefore, natural infection cannot be recreated in animal models. All experimentally induced amebiasis methods currently available require administration of live *E. histolytica* trophozoites into the target tissue. Although this procedure does not represent the natural course of infection, experimental amebiasis models are useful to test vaccine formulations and provide evidence of the efficacy of the immune responses produced.

One of the first and most frequently used models is the gerbil (*Meriones unguiculatus*) model for intestinal amebiasis and amebic liver abscess. Male gerbils are anesthetized for laparotomy, and *E. histolytica* trophozoites are injected directly into the left liver lobe to induce abscess. The gerbil model of amebic liver abscess is highly reproducible and the main model to study extraintestinal amebiasis. The main disadvantage of this

model is the lack of immunological reagents to characterize the host response to *E. histolytica* infection. Mice disease models are advantageous because of the wide range of reagents available to dissect the immune response as well as various gene knockout strains that can allow the study of specific immune components in disease pathogenesis. In the case of amebiasis, mice have demonstrated strain-dependent resistance to *E. histolytica* and have made the development of reliable mice models difficult [45]. Nonetheless, it has been shown that C3H/Hej and CBA/J mouse strains are susceptible to intestinal amebiasis, while C57BL/6 and BALB/c mice are naturally resistant. It is also important to note that mice deficient in IL-12, IFN-γ, and NO are resistant to intestinal amebiasis [46]. The nature of resistance in these mice strains is unclear and complicated. Neutrophil depletion leads to increased susceptibility to intestinal amebiasis in a normally partially resistant CBA mouse strain but had no effect on the resistant C57BL/6 mice [47]. A recent study has examined the factors involved in mouse resistance using bone marrow chimeras from resistant and susceptible mice. Shinjiro Hamano et al. [48] put forth a two-tier resistance model from their chimera studies, suggesting that the first line of resistance is conferred by the intestinal epithelial lining and the second by the underlying immune cells. The authors also reported the contribution of IL-10 to innate resistance, suggesting that the cytokine had physiological effects on intestinal epithelial cells inducing a protective barrier to infection.

The susceptible C3H/Hej mice have been used as model for intestinal amebiasis [46]. These mice are anesthetized and a laparotomy is performed to expose the cecum in which the live *E. histolytica* trophozoites are injected. These mice develop chronic nonhealing lesions in the cecum characterized by mast cell infiltration and cecitis was marked by IL-4 and IL-13 cytokine production. Depletion of CD4+ T cells decreased parasite burden and greatly reduced inflammation associated with mast cell infiltration. In this model CD4+ cells exacerbated diseases, demonstrating the importance of host immune cells in disease progression. The disadvantage of this model is that it has a 60% infection rate and therefore requires a larger number of animals for reproducibility. In order to examine the role of T and B cells in immunity to amebiasis, severe combined immunodefient mice (SCID) have been used as a model for amebic liver abscess [49]. SCID mice develop abscesses characterized by neutrophil infiltration that surround necrotic areas containing trophozoites. The abscess is clearly separated from healthy hepatic tissue by a ring of inflammatory cells and fibrosis.

Although the C3H/Hej and SCID mouse models can illustrate immune parameters because of the availability of immunological reagents, these are limited in their ability to truly discern the interaction with a human parasite. The SCID-HU-INT model, using a human intestinal xenograft, was developed to address this issue [50]. A piece of human intestinal tissue is grafted onto the subcapsular region of a SCID mouse and allowed to integrate for 8 weeks. The graft area is then infected with *E. histolytica* and develops severe inflammation characterized by neutrophil infiltration and IL-8 as well as IL-1β secretion from epithelial cells. This model is useful to study the human immune responses to *E. histolytica*. However, it is not a tool that every laboratory can use because it requires access to human tissue and also requires 2 months before experiments can take place.

### 9.5 Vaccine Trials

Acquired immunity to invasive amebiasis has been achieved repeatedly either using total amebic antigen immunizations or drug-cured animals resistant to secondary infection. Based on the animal and epidemiological evidence that acquire immunity occurs against *E. histolytica*, the feasibility of an amebiasis vaccine has become more evident. To date only a few amebic proteins have been considered as vaccine candidates, including the serine-rich *E. histolytica* protein (SREHP), the 29-kDa cysteine-rich protein, and Gal-lectin [51]. SREHP is a surface membrane protein whose function remains unclear but that demonstrates immunogenicity both in humans and in animal models. The 29-kDa cysteine-rich protein is believed to be involved in parasite detoxification of hydrogen peroxide and also demonstrates high immunogenicity with 80% of patients with amebic abscess having antibodies against this antigen [52]. Gal-lectin, however, is the most promising vaccine candidate because it is an essential parasite virulence factor and extremely immunogenic. The central role of Gal-lectin in *E. histolytica* pathogenesis makes it a prime target for the host's immune system. Several immunization trials have used the native Gal-lectin or recombinant portions of the molecule and exhibited some degree of protection against amebic abscess or intestinal disease (Table 1).

The protective effects of Gal-lectin were first reported [53] where the native molecule was used with Freund's adjuvant to immunize gerbils. All immunized gerbils developed high titers of anti-Gal-lectin antibody capable of blocking parasite adherence in vitro. Combining the intraperitoneal and subcutaneous immunization, Gal-lectin demonstrated 67% vaccine efficacy, compared to sham immunized controls. The immunized gerbils that were not protected developed significantly larger abscesses than controls; however, there was no correla-

tion between antibody titers and protection. The formation of larger abscesses in these animals was believed to be due to some immunosuppressive effects of Gal-lectin. This initial vaccination study revealed that Gal-lectin was a potent vaccine antigen but that it could also exacerbate disease in some instances. This brought about the notion that different regions of Gal-lectin may be responsible for the protective and the deleterious effects observed after challenge infections. Many future vaccination trials examined the protective effects of subunits of Gal-lectin [54–56]. A particularly interesting study by Lotter et al. determined the protective and exacerbating regions of Gal-lectin using four recombinant proteins corresponding to four main regions of the heavy subunit [57]. The four proteins included the cysteine-poor region [r170CP 1–143 amino acids (aa)], the pseudorepeat region (r170PR 437–624 aa), and two cysteine-rich regions (r170CR1 799–939 aa) and (r170CR2 940–1053). Gerbils were immunized and the formation of amebic liver abscess compared for each of the four recombinant proteins. Immunization with the r170CP was found to exacerbate disease and this due to the antibody binding the amino terminal of the molecule. The r170CR1 had no effect, while both r170PR and r170CR2 demonstrated protection against amebic liver abscess. Passive transfer of rabbit immune serum for r170CR2 and r170PR into SCID mice also protected against abscess formation. It was also observed that 100% of serum samples from human patients with amebic disease reacted with the exacerbating r170CP protein, whereas 78% asymptomatic patient serum recognized the protective r170CR2 protein. This study identified the region of Gal-lectin involved in protection from disease.

Although the cysteine-rich region of Gal-lectin was protective against disease progression, at this point it was unclear whether the same effects would be observed at the intestinal level. Houpt et al. [60] showed that both purified native Gal-lectin and a recombinant portion (LecA 578–1154 aa) could protect against intestinal amebiasis in the C3H/Hej model. The mice were immunized intranasally with either native Gal-lectin or LecA adjuvated with cholera toxin and boosted by intraperitoneal injection with Freund's adjuvant and the vaccine. The mice were then challenged by intracecal injection of live *E. histolytica* trophozoites and assessed for intestinal pathology by histological damage scores. The native Gal-lectin immunizations conferred protection with a vaccine efficacy of 84–100% while the recombinant LecA had a vaccine efficacy of 91 and 34% in the first and second trial, respectively. Protection in this study was associated with the presence of prechallenge fecal anti-Gal-lectin IgA. The vaccination trial demonstrated that Gal-lectin and LecA could prevent *E. histolytica* colonization but did not seem to reduce disease severity in those unprotected animals.

Another revolutionary vaccination strategy, DNA vaccines, can induce both humoral and systemic immune responses by directly transfecting host cells with plasmid encoding the vaccine antigen. Recently, in our laboratory a codon-optimized DNA vaccine encoding the cysteine-rich region (894–1081 aa) of Gal-lectin was developed and administered to BALB/c mice by intradermal immunization [62]. This DNA vaccine induced Gal-lectin-specific IgG2a Th1-type antibodies that could reduce parasite adherence to CHO cells. Cell-mediated responses were also induced by Gal-lectin DNA vaccination, characterized by T-cell proliferation to Gal-lectin.

### 9.6 Improving Vaccine Delivery and Efficacy

Amebiasis is a gastrointestinal disease, and the strategy for an effective vaccine is to stimulate the gut immune system and block parasite colonization. Several attempts have been made to improve vaccine delivery and efficacy with Gal-lectin. *Salmonella* strains have been used as a means of delivering antigens to the gut-associated lymphoid tissue as these bacteria effectively invade Peyer's patches. This strategy could effectively induce both mucosal and systemic immunity, which could prevent parasite colonization and reduce disease severity. *Salmonella dublin* expressing the LecA fragment of Gal-lectin was used as an oral vaccine vector [56]. Oral immunization of gerbils with $10^8$–$10^{10}$ bacteria three times provided some protection against amebic liver challenge infection, reducing abscess size compared to controls. This study demonstrated that oral immunization conferred systemic immunity against *E. histolytica*. The bacterial vector strategy was repeated recently with *Yersinia enterocolitica* expressing Gal-lectin fused to the outer protein [61, 63]. Live attenuated *Yersinia* was administered orally to gerbils that were then challenged for amebic liver abscess. Oral immunization with this vector also protected gerbils against amebic liver abscess formation.

Recently, as our knowledge about innate immunity increases, the choice of adjuvant in vaccine development has become as important as the choice in vaccine antigen. It is imperative that the vaccine formulations induce the correct response to ensure protection. Dendritic cells have a crucial role in this vaccine response as they direct the capture of antigens and the ensuing responses. In fact, antigen capture in the absence of concurrent dendritic cell maturation leads to immune tolerance and not immunity. An emerging field in vaccination research is the use of toll-like receptor (TLR) ligands in immunizations to ensure the proper

adjuvant response. CpG-oligodeoxynucleotides (CpG-ODNs) are recognized by TLR9, expressed on subsets of dendritic cells and induce maturation as well as potent Th1 cytokine production [64]. CpG-ODNs have been used in animal models of various diseases and shown consistent adjuvant properties both with parenteral and mucosal vaccines [65]. Ongoing clinical trials with CpG-ODN have investigated the use of these TLR ligands as adjuvants and demonstrated that they are well tolerated by humans and produce the desired improved immune responses. Gal-lectin is a prime candidate for an amebiasis vaccine, and given the immunogenic Th1 response and protection observed to date with Gal-lectin, efforts should be made to improve vaccine efficacy.

The development of mucosal vaccines in humans has been limited by the lack of safe yet effective mucosal adjuvants. The gold standard mucosal adjuvant in animals is the cholera toxin; however, it is too toxic for use in humans. In order to address the feasibility of a mucosal amebiasis vaccine for potential use in humans, we tested an intranasal Gal-lectin vaccine using CpG-ODN as adjuvant [66]. We chose Gal-lectin as a protective amebic antigen and chose CpG-ODN as a potent Th1-inducing nontoxic adjuvant. CpG-ODNs containing nuclease-resistant phosphorothioate backbones are coadministered with the vaccine antigen to induce specific immunity. We tested the native Gal-lectin and CpG-ODN vaccine in the gerbil model of amebic liver abscess and in the C3H/Hej mouse model of amebic colitis. Repeated intranasal immunization followed by one booster intraperitoneal injection resulted in the production of both anti-Gal-lectin IgG and IgA antibodies capable of blocking parasite adherence in vitro. The immunized animals produced more IFN-γ than sham immunized controls and had Gal-lectin-specific lymphocytes capable of proliferating in response to Gal-lectin exposure. More importantly, immunized gerbils were protected from subsequent challenge infection and did not develop amebic liver abscesses. Several properties of Gal-lectin and CpG-ODN contributed to the efficacy of this vaccine. CpG-ODN adjuvant primed the immune response to the coadministered Gal-lectin, and both components are strong Th1 inducers activating immune cells in a pro-inflammatory fashion. The intranasal administration proved to be successful in inducing the desired mucosal IgA response, with the production of Gal-lectin-specific fecal antibodies. Although we did not correlate these anti-Gal-lectin IgA with protection from intestinal disease in the mouse model, it is interesting to note the presence of these specific antibodies after immunization and the epidemiological evidence that in humans IgA is protective. Protection could be correlated, however, with elevated levels of IFN-γ in the gerbil model of amebiasis. These findings underline the importance of vaccine design and the need to optimize the desired responses to provide both mucosal and cell-mediated immunity to *E. histolytica*. In this way a vaccine capable of inducing both mucosal and systemic immunity could prevent colonization in an antibody-mediated fashion and prevent invasive disease mediated by a strong Th1 response to Gal-lectin.

## 10 CONTINUING RESEARCH EFFORTS

In order to address the challenges in vaccine development continued research about the nature of protective immunity in amebiasis and the factors that determine amebic invasion are required. A new bioinformatics resource center, *Pathema,* created for category A–C pathogens described by the National Institute of Allergy and Infectious Diseases, includes *E. histolytica*. This service provides scientists with the tools to study *E. histolytica*, including protein cataloging, functional genomics, and comparative analyses. The completion of the *Entamoeba* genome was a keystone event in amebiasis research, allowing researchers to probe the genome to identify differences in virulence and immune targets [67]. The *Entamoeba* genome is readily available on *Pathema* and researchers can exchange protocols and new findings. This informatics center encourages the type of collaborative work that could help establish the groundwork for future amebiasis therapies.

Continued vaccine research in rodent models is essential to help elucidate the immune mechanism involved in protective immunity; however, these studies cannot predict the efficacy of a vaccine in humans. Nonhuman primate models are a good starting point for testing vaccines for human use. The specificity of *E. histolytica* to humans makes experimental amebiasis difficult to study; fortunately, it has been demonstrated that baboons can be naturally infected with *E. histolytica*. A recent study in baboons tested a Gal-lectin peptide vaccine containing amino acids 758–1134 of the heavy subunit with cholera toxin adjuvant [68]. Repeated intranasal immunizations in uninfected seronegative baboons, induced antigen-specific serum IgA and IgG as well as intestinal IgA antibodies capable of neutralizing amebic activity *in vitro*. Continued research in this area is crucial to the advancement of vaccine development for amebiasis.

The strategy for a protective vaccine against amebiasis is to formulate a vaccine that will be administered orally to stimulate gut mucosal immunity. Oral vaccines are easy to administer and with the right formulation could stimulate both mucosal and systemic responses. Recently, efforts have been made to address the feasibility of an oral amebiasis vaccine. The advent of genetic engineering has led to the development of transgenic

plants capable of expressing vaccine antigens without the common contaminants found in bacterial or yeast expression systems. The Gal-lectin heavy subunit gene was inserted into the chloroplast genome of the tobacco plant to evaluate its potential use as an oral vaccine [69]. The plants expressed the Gal-lectin protein, and crude plant extracts were able to induce Gal-lectin-specific antibodies when administered subcutaneously. Although the plants were not ingested in this study, it was the first report of Gal-lectin expression in a plant system that could facilitate oral vaccine delivery. Another recent study examined the use of live attenuated *Yersinia* to deliver the Gal-lectin protein to the gut mucosa [63]. *Yersinia* can colonize the Peyer's patches, and through its type III secretion system present the vaccine antigen to lamina propria immune cells. Using *Yersinia* as a delivery system protected gerbils from amebic liver abscess formation and was associated with an elevated IFN-γ production. Other studies have also tested the use of attenuated *Salmonella* strains containing plasmids encoding amebic proteins. These studies have established the feasibility of using various expression systems to deliver *E. histolytica* antigens to the gut associated lymphoid tissues (GALT).

Finally, perhaps the most important factor in developing an amebiasis vaccine is to continue to investigate the natural immunity mechanisms in the human population as well as to find innovative ways to mimic that immunity in a vaccine. Epidemiological studies and experimental vaccine studies are powerful tools that are helping elucidate the path to an effective vaccine.

## REFERENCES


1. World Health Organization (WHO) (1998). *World health Report—Life in the 21st century: A vision for all*. Geneva: WHO.
2. Jiang G, et al. (2007). Induction of multi-antigen multistage immune responses against *Plasmodium falciparum* in rhesus monkeys, in the absence of antigen interference, with heterologous DNA prime/poxvirus boost immunization. *Malaria J 6*, 135.
3. Wipasa J, Riley EM (2007). The immunological challenges of malaria vaccine development. *Expert Opin Biol Ther 7*, 1841–1852.
4. Horrocks P, Pinches R, Christodoulou Z, Kyes SA, Newbold CI (2004). Variable var transition rates underlie antigenic variation in malaria. *Proc Natl Acad Sci USA 101*, 11129–11134.
5. Flick K, Chen Q (2004). Var genes, PfEMP1 and the human host. *Mol Biochem Parasitol 134*, 3–9.
6. Chen Q, Schlichtherle M, Wahlgren M (2000). Molecular aspects of severe malaria. *Clin Microbiol Rev 13*, 439–450.
7. Duffy PE, Kaslow DC (1997). A novel malaria protein, Pfs28, and Pfs25 are genetically linked and synergistic as falciparum malaria transmission-blocking vaccines. *Infect Immun 65*, 1109–1113.
8. Mikolajczak SA, Aly AS, Kappe SH (2007). Preerythrocytic malaria vaccine development. *Curr Opin Infect Dis 20*, 461–466.
9. Kurtis JD, Hollingdale MR, Luty AJ, Lanar DE, Krzych U, Duffy PE (2001). Pre-erythrocytic immunity to *Plasmodium falciparum*: The case for an LSA-1 vaccine. *Trends Parasitol 17*, 219–223.
10. Heppner DG, Schwenk RJ, Arnot D, Sauerwein RW, Luty AJ (2007). The dog that did not bark: Malaria vaccines without antibodies. *Trends Parasitol 23*, 293–296.
11. Aponte JJ, et al. (2007). Safety of the RTS,S/AS02D candidate malaria vaccine in infants living in a highly endemic area of Mozambique: A double blind randomised controlled phase I/IIb trial. *Lancet 370*, 1543–1551.
12. Polhemus ME, et al. (2007). Phase I dose escalation safety and immunogenicity trial of *Plasmodium falciparum* apical membrane protein (AMA-1) FMP2.1, adjuvanted with AS02A, in malaria-naive adults at the Walter Reed Army Institute of Research. *Vaccine 25*, 4203–4212.
13. World Health Organization (1997). Ameobiasis. *Weekly Epidemiol Rec 72*, 97–100.
14. Eckmann L, Reed SL, Smith JR, Kagnoff MF (1995). *Entamoeba histolytica* trophozoites induce an inflammatory cytokine response by cultured human cells through the paracrine action of cytolytically released interleukin-1 alpha. *J Clin Invest 96*, 1269–1279.
15. Tsutsumi V, Mena-Lopez R, Anaya-Velazquez F, Martinez-Palomo A (1984). Cellular bases of experimental amebic liver abscess formation. *Am J Pathol 117*, 81–91.
16. Seydel KB, Li E, Zhang Z, Stanley SL Jr (1998). Epithelial cell-initiated inflammation plays a crucial role in early tissue damage in amebic infection of human intestine. *Gastroenterology 115*, 1446–1453.
17. Seguin R, Keller K, Chadee K (1995). *Entamoeba histolytica* stimulates the unstable transcription of c-fos and tumour necrosis factor-alpha mRNA by protein kinase C signal transduction in macrophages. *Immunology 86*, 49–57.
18. Denis M, Chadee K (1989). Cytokine activation of murine macrophages for in vitro killing of *Entamoeba histolytica* trophozoites. *Infect Immun 57*, 1750–1756.
19. Seguin R, Mann BJ, Keller K, Chadee K (1997). The tumor necrosis factor alpha-stimulating region of galactose-inhibitable lectin of *Entamoeba histolytica* activates gamma interferon-primed macrophages for amebicidal activity mediated by nitric oxide. *Infect Immun 65*, 2522–2527.
20. Haque R, Mondal D, Shu J, Roy S, Kabir M, Davis AN, Duggal P, Petri WA Jr (2007). Correlation of interferon-gamma production by peripheral blood mononuclear cells with childhood malnutrition and susceptibility to amebiasis. *Am J Trop Med Hyg 76*, 340–344.
21. Haque R, Ali IM, Sack RB, Farr BM, Ramakrishnan G, Petri WA Jr (2001). Amebiasis and mucosal IgA antibody

against the *Entamoeba histolytica* adherence lectin in Bangladeshi children. *J Infect Dis 183*, 1787–1793.

22. Haque R, Duggal P, Ali IM, Hossain MB, Mondal D, Sack RB, Farr BM, Beaty TH, Petri WA Jr (2002). Innate and acquired resistance to amebiasis in bangladeshi children. *J Infect Dis 186*, 547–552.
23. Chadee K, Petri WA Jr, Innes DJ, Ravdin JI (1987). Rat and human colonic mucins bind to and inhibit adherence lectin of *Entamoeba histolytica*. *J Clin Invest 80*, 1245–1254.
24. Chadee K, Ndarathi C, Keller K (1990). Binding of proteolytically-degraded human colonic mucin glycoproteins to the Gal/GalNAc adherence lectin of *Entamoeba histolytica*. *Gut 31*, 890–895.
25. Chadee K, Meerovitch E (1985). *Entamoeba histolytica*: Early progressive pathology in the cecum of the gerbil (*Meriones unguiculatus*). *Am J Trop Med Hyg 34*, 283–291.
26. Moncada D, Keller K, Chadee K (2003). *Entamoeba histolytica* cysteine proteinases disrupt the polymeric structure of colonic mucin and alter its protective function. *Infect Immun 71*, 838–844.
27. Li E, Becker A, Stanley SL Jr (1989). Chinese hamster ovary cells deficient in *N*-acetylglucosaminyltransferase I activity are resistant to *Entamoeba histolytica*-mediated cytotoxicity. *Infect Immun 57*, 8–12.
28. Blazquez S, Rigothier MC, Huerre M, Guillen N (2007). Initiation of inflammation and cell death during liver abscess formation by *Entamoeba histolytica* depends on activity of the galactose/*N*-acetyl-D-galactosamine lectin. *Int J Parasitol 37*, 425–433.
29. Huston CD, Boettner DR, Miller-Sims V, Petri WA Jr (2003). Apoptotic killing and phagocytosis of host cells by the parasite *Entamoeba histolytica*. *Infect Immun 71*, 964–972.
30. Berninghausen O, Leippe, M (1997). Necrosis versus apoptosis as the mechanism of target cell death induced by *Entamoeba histolytica*. *Infect Immun 65*, 3615–3621.
31. Young JD, Young TM, Lu LP, Unkeless JC, Cohn ZA (1982). Characterization of a membrane pore-forming protein from *Entamoeba histolytica*. *J Exp Med 156*, 1677–1690.
32. Weinke T, Friedrich-Janicke B, Hopp P, Janitschke K (1990). Prevalence and clinical importance of *Entamoeba histolytica* in two high-risk groups: Travelers returning from the tropics and male homosexuals. *J Infect Dis 161*, 1029–1031.
33. Ali IK, Mondal U, Roy S, Haque R, Petri WA Jr, Clark CG (2007). Evidence for a link between parasite genotype and outcome of infection with *Entamoeba histolytica*. *J Clin Microbiol 45*, 285–289.
34. Petri WA Jr, Chapman MD, Snodgrass T, Mann BJ, Broman J, Ravdin JI (1989). Subunit structure of the galactose and *N*-acetyl-D-galactosamine-inhibitable adherence lectin of *Entamoeba histolytica*. *J Biol Chem 264*, 3007–3012.
35. Beck DL, Tanyuksel M, Mackey AJ, Haque R, Trapaidze N, Pearson WR, Loftus B, Petri WA (2002). *Entamoeba histolytica*: Sequence conservation of the Gal/GalNAc lectin from clinical isolates. *Exp Parasitol 101*, 157–163.
36. Petri WA Jr, Joyce MP, Broman J, Smith RD, Murphy CF, Ravdin JI (1987). Recognition of the galactose- or *N*-acetylgalactosamine-binding lectin of *Entamoeba histolytica* by human immune sera. *Infect Immun 55*, 2327–2331.
37. Ravdin JI, Jackson TF, Petri WA Jr, Murphy CF, Ungar BL, Gathiram V, Skilogiannis J, Simjee AE (1990). Association of serum antibodies to adherence lectin with invasive amebiasis and asymptomatic infection with pathogenic *Entamoeba histolytica*. *J Infect Dis 162*, 768–772.
38. Schain DC, Salata RA, Ravdin JI (1992). Human T-lymphocyte proliferation, lymphokine production, and amebicidal activity elicited by the galactose-inhibitable adherence protein of *Entamoeba histolytica*. *Infect Immun 60*, 2143–2146.
39. Lin JY, Seguin R, Keller K, Chadee K (1995). Transforming growth factor-beta 1 primes macrophages for enhanced expression of the nitric oxide synthase gene for nitric oxide-dependent cytotoxicity against *Entamoeba histolytica*. *Immunology 85*, 400–407.
40. Campbell D, Mann BJ, Chadee K (2000). A subunit vaccine candidate region of the *Entamoeba histolytica* galactose-adherence lectin promotes interleukin-12 gene transcription and protein production in human macrophages. *Eur J Immunol 30*, 423–430.
41. Braga LL, Ninomiya H, McCoy JJ, Eacker S, Wiedmer T, Pham C, Wood S, Sims PJ, Petri WA Jr (1992). Inhibition of the complement membrane attack complex by the galactose-specific adhesion of *Entamoeba histolytica*. *J Clin Invest 90*, 1131–1137.
42. Ravdin JI, Petri WA, Murphy CF, Smith RD (1986). Production of mouse monoclonal antibodies which inhibit in vitro adherence of *Entamoeba histolytica* trophozoites. *Infect Immun 53*, 1–5.
43. Petri WA Jr, Snodgrass TL, Jackson TF, Gathiram V, Simjee AE, Chadee K, Chapman MD (1990). Monoclonal antibodies directed against the galactose-binding lectin of *Entamoeba histolytica* enhance adherence. *J Immunol 144*, 4803–4809.
44. Vines RR, Ramakrishnan G, Rogers JB, Lockhart LA, Mann BJ, Petri WA Jr (1998). Regulation of adherence and virulence by the *Entamoeba histolytica* lectin cytoplasmic domain, which contains a beta2 integrin motif. *Mol Biol Cell 9*, 2069–2079.
45. Ghadirian E, Kongshavn PA (1984). Genetic control of susceptibility of mice to infection with *E. histolytica*. *Parasite Immunol 6*, 349–360.
46. Houpt ER, Glembocki DJ, Obrig TG, Moskaluk CA, Lockhart LA, Wright RL, Seaner RM, Keepers TR, Wilkins TD, Petri WA Jr (2002). The mouse model of amebic colitis reveals mouse strain susceptibility to infec-

tion and exacerbation of disease by CD4+ T cells. *J Immunol 169*, 4496–4503.

47. Asgharpour A, Gilchrist C, Baba D, Hamano S, Houpt E (2005). Resistance to intestinal *Entamoeba histolytica* infection is conferred by innate immunity and Gr-1+ cells. *Infect Immun 73*, 4522–4529.
48. Hamano S, Asgharpour A, Stroup SE, Wynn TA, Leiter EH, Houpt E (2006). Resistance of C57BL/6 mice to amoebiasis is mediated by nonhemopoietic cells but requires hemopoietic IL-10 production. *J Immunol 177*, 1208–1213.
49. Cieslak PR, Virgin HW, Stanley SL Jr (1992). A severe combined immunodeficient (SCID) mouse model for infection with *Entamoeba histolytica*. *J Exp Med 176*, 1605–1609.
50. Seydel KB, Li E, Swanson PE, Stanley SL Jr (1997). Human intestinal epithelial cells produce proinflammatory cytokines in response to infection in a SCID mouse-human intestinal xenograft model of amebiasis. *Infect Immun 65*, 1631–1639.
51. Stanley SL Jr, Jackson TF, Reed SL, Calderon J, Kunz-Jenkins C, Gathiram V, Li E (1991). Serodiagnosis of invasive amebiasis using a recombinant *Entamoeba histolytica* protein. *JAMA 266*, 1984–1986.
52. Flores BM, Reed SL, Ravdin JI, Torian BE (1993). Serologic reactivity to purified recombinant and native 29-kilodalton peripheral membrane protein of pathogenic *Entamoeba histolytica*. *J Clin Microbiol 31*, 1403–1407.
53. Petri WA Jr, Ravdin JI (1991). Protection of gerbils from amebic liver abscess by immunization with the galactose-specific adherence lectin of *Entamoeba histolytica*. *Infect Immun 59*, 97–101.
54. Zhang T, Stanley SL Jr (1994). Protection of gerbils from amebic liver abscess by immunization with a recombinant protein derived from the 170-kilodalton surface adhesin of *Entamoeba histolytica*. *Infect Immun 62*, 2605–2608.
55. Soong CJ, Kain KC, Abd-Alla M, Jackson TF, Ravdin JI (1995). A recombinant cysteine-rich section of the *Entamoeba histolytica* galactose-inhibitable lectin is efficacious as a subunit vaccine in the gerbil model of amebic liver abscess. *J Infect Dis 171*, 645–651.
56. Mann BJ, Burkholder BV, Lockhart LA (1997). Protection in a gerbil model of amebiasis by oral immunization with Salmonella expressing the galactose/*N*-acetyl D-galactosamine inhibitable lectin of *Entamoeba histolytica*. *Vaccine 15*, 659–663.
57. Lotter H, Zhang T, Seydel KB, Stanley SL Jr, Tannich E (1997). Identification of an epitope on the *Entamoeba histolytica* 170-kD lectin conferring antibody-mediated protection against invasive amebiasis. *J Exp Med 185*, 1793–1801.
58. Dodson JM, Lenkowski PW Jr, Eubanks AC, Jackson TF, Napodano J, Lyerly DM, Lockhart LA, Mann BJ, Petri WA Jr (1999). Infection and immunity mediated by the carbohydrate recognition domain of the *Entamoeba histolytica* Gal/GalNAc lectin. *J Infect Dis 179*, 460–466.
59. Lotter H, Khajawa F, Stanley SL Jr, Tannich E (2000). Protection of gerbils from amebic liver abscess by vaccination with a 25-mer peptide derived from the cysteine-rich region of *Entamoeba histolytica* galactose-specific adherence lectin. *Infect Immun 68*, 4416–4421.
60. Houpt E, Barroso L, Lockhart L, Wright R, Cramer C, Lyerly D, Petri WA (2004). Prevention of intestinal amebiasis by vaccination with the *Entamoeba histolytica* Gal/GalNac lectin. *Vaccine 22*, 611–617.
61. Lotter H, Russmann H, Heesemann J, Tannich E (2004). Oral vaccination with recombinant *Yersinia enterocolitica* expressing hybrid type III proteins protects gerbils from amebic liver abscess. *Infect Immun 72*, 7318–7321.
62. Gaucher D, Chadee K (2002). Construction and immunogenicity of a codon-optimized *Entamoeba histolytica* Gal-lectin-based DNA vaccine. *Vaccine 20*, 3244–3253.
63. Lotter H, Russmann H, Heesemann J, Tannich E (2008). Attenuated recombinant *Yersinia* as live oral vaccine carrier to protect against amoebiasis. *Int J Med Microbiol 298*, 79–86.
64. Behboudi S, Chao D, Klenerman P, Austyn J (2000). The effects of DNA containing CpG motif on dendritic cells. *Immunology 99*, 361–366.
65. Krieg AM (2000). Immune effects and mechanisms of action of CpG motifs. *Vaccine 19*, 618–622.
66. Ivory CP, Chadee K (2007). Intranasal immunization with Gal-inhibitable lectin plus an adjuvant of CpG oligodeoxynucleotides protects against *Entamoeba histolytica* challenge. *Infect Immun 75*, 4917–4922.
67. Loftus B, et al. (2005). The genome of the protist parasite *Entamoeba histolytica*. *Nature 433*, 865–868.
68. Abd Alla MD, White GL, Rogers TB, Cary ME, Carey DW, Ravdin JI (2007). Adherence-inhibitory intestinal immunoglobulin a antibody response in baboons elicited by use of a synthetic intranasal lectin-based amebiasis subunit vaccine. *Infect Immun 75*, 3812–3822.
69. Chebolu S, Daniell H (2007). Stable expression of Gal/GalNAc lectin of *Entamoeba histolytica* in transgenic chloroplasts and immunogenicity in mice towards vaccine development for amoebiasis. *Plant Biotechnol J 5*, 230–239.

# TOXICOLOGY

John H. Duffus
*Edinburgh Centre for Toxicology, Edinburgh, Scotland, United Kingdom*

## 1 INTRODUCTION

Toxicology is most simply defined as the science of poisons. Using naturally occurring substances—first as foodstuffs and subsequently as drugs—must inevitably have led to the recognition of poisoning as a result of excessive intake or overdose at a very early stage in human evolution. There must also have been recognition that death or other ill effect sometimes followed the intake of very small amounts of a foodstuff or other substance. This would have led to the deliberate use of such substances as poisons and their use on weapons, such as arrows, to paralyze or kill. The use on arrows led to the science of toxicology, derived from the Greek word τοξικον (toxicon) for an arrow poison. Today, the main emphasis of toxicology is on detecting and preventing the unwanted effects of chemical agents, although their use as poisons, especially as pesticides and in chemical warfare, continues.

Toxicology is not a "pure science." It occupies an interface between chemistry, medicine, and biology and is embedded in legislation that governs what substances may be manufactured and how they are used. Thus, it influences, and is itself influenced by, laws and regulations introduced to protect people from the toxic effects of substances that, like pharmaceuticals, may have been marketed because of their beneficial effects when used properly. As a result, the findings of toxicology can have profound economic effects and whole industries, such as asbestos manufacturing, may collapse if their products are condemned as excessively toxic in normal use.

The main objective of toxicology is to prevent harmful effects of useful substances, including therapeutic agents. For new pharmaceutical agents, extensive toxicity and efficacy data are required, including data from volunteer studies in humans. Pharmaceutical agents, at appropriate dose levels, are expected to have a beneficial effect on patients. Other substances, in normal use, whatever their beneficial effects, should have no biological effects on people. This includes pesticides, which, when used as intended, should produce a biological effect in pests without any effect on humans. Absence of effect of pesticides on people depends partly on differences in biochemistry and physiology between pests and humans and partly on careful application of the pesticides to minimize human exposure.

This chapter on toxicology can give you only a limited idea of all that is involved on the subject and you are strongly advised to refer to one or more of the good general textbooks that are available [1–3].

## 2 BASIC PRINCIPLES

Three concepts underlie modern toxicology. The most fundamental may be paraphrased as "the dose makes the cure or the poison." This concept was first recorded by Paracelsus (Theophrastus Phillippus Aureolus Bombastus von Hohenheim, 1493–1541), a Swiss physician and alchemist. In other words, all substances, including pharmaceuticals, are potentially toxic. Defining precisely the therapeutic dose and the toxic dose is, thus,

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

essential if pharmaceuticals are to be used successfully in treating disease. For other substances, regulation must ensure that "safe" exposure levels are not exceeded. Determining safe exposure levels is thus a major activity of regulatory toxicologists. This requires establishing a permissible exposure level (PEL) for all substances that may reasonably be considered to pose a risk to humans or other organisms of concern.

The second essential concept is that the biological effects of different substances depend upon their chemical structure. This concept has been associated with Ambroise Paré, a sixteenth-century French surgeon who recognized that toxic agents have different effects depending on their inherent character. Studying the relationship of chemical structure to biological effect (structure–activity relationship, SAR) is a central activity of both pharmacology and toxicology and has been of great importance in the development of more effective and safer drugs and pesticides.

The third essential concept is that humans are mammals and thus effects on other mammals are likely to apply also to humans. This concept has enabled toxicologists to derive an understanding of the nature of toxicity and the mechanisms that produce it by extrapolating from experimental studies of effects on other mammals such as rats and mice exposed to substances of concern. Because mammalian species are not identical, such extrapolations are done with care to allow for known differences in physiology and metabolism. As our understanding of toxicity to mammals, and especially to humans, has developed, it has been possible to improve our protection of humans from toxicity and to provide effective treatment of poisoning when it occurs.

## 3 DOSE–EFFECT AND DOSE–RESPONSE RELATIONSHIPS

There are two types of dose–effect relationships. One has a threshold below which no effect occurs (Fig. 1). Although it is almost impossible to determine a threshold value precisely by experiment, doses just above or just below the threshold can be measured and used as approximations for the calculation of safe exposure levels. This will be described below.

A threshold can be identified for all toxic effects except for mutagenic effects, including those that lead to the subsequent development of cancer, and for some, if not all, teratogenic effects. A mutation can theoretically occur through a single chemical molecule producing a change in the chemical structure of a single DNA (Deoxyribonucleic acid) molecule, altering the nucleotide sequence and its transcription product. Two molecules have twice as much chance as causing this effect and so on. This leads to a linear relationship between the dose of a mutagenic substance and the probability that a mutation will occur (Fig. 2). The true risk of a single molecule causing a mutation is infinitely small and there are also many defense mechanisms within cells, including DNA repair mechanisms, which can prevent mutation from occurring. Thus, it is very likely that thresholds exist even for mutagenic effects, but so far no test method exists to establish such thresholds, and thus determination of permissible exposure levels for such effects must be based on calculations of their probability. It is generally agreed that a significant risk is reached with a high enough exposure to produce a

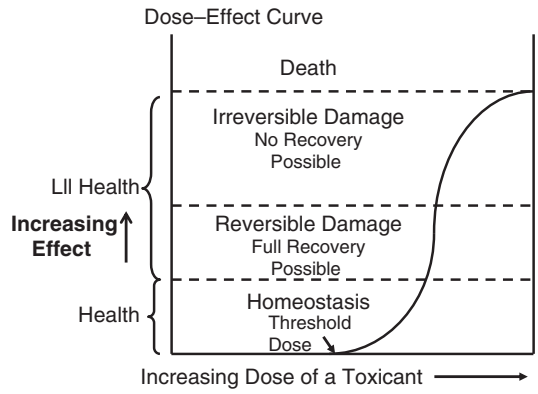


**Figure 1** Dose–effect relationship for a nonnutrient substance with a threshold dose below which no effect occurs (a nonstochastic relationhip).

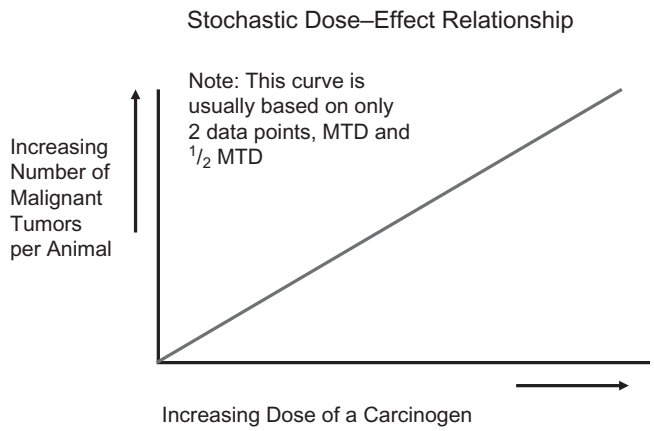


**Figure 2** Dose–effect relationship for a nonnutrient carcinogenic substance assuming there is no threshold dose. Thus increasing the dose is associated with increasing probability of a toxic effect occurring (a stochastic relationship). MTD = maximum tolerated dose—the highest *dose* that is expected on the basis of an adequate *subchronic* study to produce only limited *toxicity* when administered for the duration of the test period.

probable excess incidence of disease of above one in a million in the exposed population.

## 4 CHEMICAL INTERACTIONS

A major problem in assessing the likely effect of exposure to a chemical is that of assessing possible interactions [4]. The simplest interaction is an additive effect: This is an effect that is the result of two or more chemicals acting together, which is the simple sum of their effects when acting independently. In mathematical terms: $1 + 1 = 2$, $1 + 5 = 6$, and so on. The effects of organochlorine pesticides are usually additive reflecting similar structures and mode of action.

More complex is a synergistic (multiplicative) effect: This is an effect of two chemicals acting together that is greater than the simple sum of their effects when acting alone; it may be called synergism. In mathematical terms: $1 + 1 = 4$, $1 + 5 = 10$, and so on. Asbestos fibers and cigarette smoking act together to increase the risk of lung cancer by a factor of 40, taking it well beyond the risk associated with independent exposure to either of these agents.

Another possible form of interaction is potentiation. In potentiation, a substance that on its own causes no harm makes the effects of another chemical much worse. This may be considered to be a form of synergism. In mathematical terms: $0 + 1 = 5$, $0 + 5 = 20$, and so on. For example, isopropanol, at concentrations that are not themselves harmful to the liver, increases (potentiates) the liver damage caused by a given concentration of carbon tetrachloride.

The opposite of synergism is antagonism: An antagonistic effect is the result of a chemical counteracting the adverse effect of another; in other words, the situation where exposure to two chemicals together has less effect than the simple sum of their independent effects. Such chemicals are said to show antagonism. In mathematical terms: $1 + 1 = 0$, $1 + 5 = 2$, and so on. An example of this is the antagonism of compounds containing SeH groups for compounds containing reactive mercury species.

While we recognize that such interactions occur, we are still at a very early stage in our ability to predict when they are likely. It may be that a number of the adverse reactions that have been attributed to specific drugs are, in fact, the results of their interactions with other drugs, or environmental chemicals, or even with substances that are otherwise nutrients. Probably the drug interaction with adverse effects that has the most recognition is that with alcoholic drinks, which potentiate the toxic effects of a number of drugs.

## 5 SAFETY ASSESSMENT

Safety assessment is the name given to the process that results in establishing for humans an "acceptable daily intake" (ADI) for specific chemicals or a "therapeutic dose" for pharmaceuticals. Safety assessment is normally based on data from mammalian toxicity tests made with the underlying assumption that humans are sufficiently similar to the mammals used in tests to permit extrapolation of observed effects and dose–effect and dose–response relationships. However, it is also recognized that humans may be more sensitive than the most sensitive mammalian species tested and calculation of safe exposure levels makes allowance for this [5].

## 6 CLASSIFICATION AND LABELING

In practical terms, the first step in safety assessment for the user of any substance is to look at the information given on the label of the chemical container. Many countries have now passed laws or regulations that legally require specific toxicity information to be given on labels or on safety data sheets (SDS), allowing appropriate protective measures to be taken. However, these laws and regulations vary, and this can cause confusion, particularly when a chemical has wide international usage. Thus, a Global Harmonization System (GHS) has been developed and is expected to be finally agreed upon soon [6].

The toxicity information is derived from data from appropriate tests that are discussed further below. For information on how toxicity data are interpreted in the GHS for the purposes of labeling, see the document prepared by the Environmental Protection Agency U.S. (EPA) Office of Pesticide Programs [7].

## 7 TOXICITY TESTING

Safety assessment depends mostly upon in vivo testing programs using mammals, usually rats and mice, although teratogenicity tests must also be carried out on nonrodents such as rabbits. This follows the failure to detect the teratogenic effect of thalidomide in rodents and the resulting phocomelia that occurred in the children of women who had been treated with this drug.

In vivo tests must be carried out in at least two mammalian species and may be for short periods or for a complete lifetime, depending upon the actual or suspected adverse effects. For example, a suspected human carcinogen may require lifetime exposure of a rodent (2

years in the rat) before malignancy can be detected. Lifetime tests are very expensive, and often a suspected carcinogen will be banned from use without such tests unless it has a unique and important use for which no substitute is readily available.

All toxicity tests providing data for regulatory purposes must conform to Good Laboratory Practice (GLP) guidelines. These guidelines originate with the Organization for Economic Cooperation and Development (OECD) [8]. GLP requires that all staff involved must be appropriately trained, follow standard operating practice (SOP), keep meticulous records, and archive all records and samples that contribute significantly to the data. The need for GLP guidelines is an acknowledgment of the serious consequences that may follow toxicity classification, whether low, permitting marketing and use of a substance, or high, preventing marketing and use.

## 8 SHORT-TERM (ACUTE) TESTS

Short-term (acute) tests are conducted in both sexes of one or more rodent species with the purpose of determining the dose range for adverse effects of a substance and to determine the exact shape of the dose–response curve. In addition to the determination of the no-observed-adverse-effect level (NOAEL), the benchmark dose (BMD), the $LD_{50}$ (the median dose lethal to half the test organisms), and/or the lowest-observed-adverse-effect level (LOAEL), or the limit dose (a legally defined no-effect dose used for toxicity classification), the results of acute studies may be used to establish the appropriate dose range for longer-term experiments. At the end of a study all animals are autopsied, with complete gross and microscopic examination of tissues. Blood chemistry is evaluated and an overall clinical assessment is made on each animal.

As well as establishing dose dependency of effects, acute studies can determine the nature of toxicity, time of onset of toxic signs, and the capacity for recovery in the surviving animals. This information is essential in any emergency situation where humans or domestic animals have suffered from poisoning. Acute toxicity studies may also help in identifying antidotes to a given toxicant.

## 9 MEDIUM-TERM (SUBCHRONIC) TESTS

Subchronic tests are generally conducted in both sexes of two mammalian species, one of which is a nonrodent species, often the rabbit. Such studies generally last 3–6 months, covering a range of doses. The main aims of a subchronic study are to determine target-organ toxicity, to determine the effects of prolonged dosing, and to provide more evidence on which to base regulation of exposure levels for food additives and drugs. As with acute tests, at the end of a study all animals are autopsied, with complete gross and microscopic examination of tissues, evaluation of blood chemistry, and a complete clinical assessment of each animal.

## 10 LONG-TERM (CHRONIC) TESTS

If the safety of use of a potentially valuable substance is still unclear after subchronic tests, chronic tests may be required. Such studies are carried out for a time that approaches the life span of the animal. Again they are applied to at least two mammalian species, usually a rodent and a nonrodent, and in both sexes. Chronic studies need careful management with multiple clinical evaluations, daily observations, weekly food and water consumption records, and body weight determinations. At the end of the study, on death, and (or) at intermediate times, if thought necessary, all animals are autopsied.

Chronic toxicity studies provide the most complete information on the effects of a given substance on homeostasis, bodily functions, induced diseases, and total life span. For drugs, they provide the most valuable preclinical information for assessment of safety and risk.

Where there is a potential for a substance to cause cancer, a specific cancer bioassay may be required. Such a bioassay is conducted in rats and mice and lasts for the natural life span of the species. The test is usually carried out at two dose levels, the maximum tolerated dose (MTD) and the half maximum tolerated dose, preferably mimicking the route of human exposure. Sometimes nonphysiological routes of exposure, such as intraperitoneal injection or intramuscular injection, have been used but the data from these must be viewed with some scepticism. In particular, the process of intramuscular injection may trigger neoplasia even with otherwise relatively harmless substances.

The possibility of reproductive toxicity must be tested in several species because there is no species that provides a perfect model to predict possible effects in humans. Rats, mice, rabbits, and sometimes dogs may be used for tests. The main objective is to find out whether the test substance alters the reproductive cycle of the female or spermatogenesis in the male. However, the tests are also designed to examine the effects of the given substance during the first, second, and third trimester of pregnancy and during parturition and lactation. Multigenerational studies are conducted to determine the overall effects of a given chemical on the

parent generation, the offspring, and on the ability of the offspring to reproduce normally.

## 11 SCREENING FOR MUTAGENICITY AND POSSIBLE CARCINOGENICITY—THE AMES TEST

The Ames test is named after its creator, Dr. Bruce Ames. The test is one of a number [9, 10] used as a screening test to identify mutagens and hence for mutagenic carcinogens. It is universally required in any battery of tests for potential toxicity. It is, therefore, important to understand it so that its results are properly interpreted.

The test uses strains of *Salmonella typhimurium* that have been altered by genetic engineering to make them unable to synthesize histidine. They are, therefore, dependent on an exogenous supply of this amino acid unless a mutation restores the histidine biosynthesis pathway. They are also engineered to be much more susceptible to mutation than the normal bacterium as a result of loss of DNA repair systems. Mutations occur when the bacterial DNA is altered by interaction with the test substance or with a reactive metabolite produced from the test substance.

To perform the test, selected *Salmonella* strains are mixed in a test tube with a solution of the substance of interest. Because the bacteria lack the cytochrome P450 enzymes that metabolize many foreign chemicals in mammals, rat liver enzymes (the so-called S9 microsome preparation) are added to the test system in a second test tube to permit metabolic products to be assessed. After exposure, the bacteria are transferred to histidine-deficient nutrient agar in a Petri dish for one or two days to see if they can grow. Without mutation, the bacteria used for the test require the amino acid histidine to grow. Thus, only mutated bacteria that have recovered the ability to synthesize histidine can grow on the histidine-deficient agar, and the number of colonies is a measure of the mutagenic potency of the substance tested.

A positive result in an Ames test does not prove that a particular chemical is a mutagen for normal cells nor that it is capable of causing cancer. Essentially, the Ames test simply shows a capacity to alter DNA in cells that are extremely sensitive to DNA damage. Thus, a number of chemicals that test positive in the Ames test do not cause mutations in mammalian cells or subsequent cancer. Further, some chemicals that are known to cause cancer do not test positive in the Ames test. More extensive testing is, therefore, needed to determine whether a chemical is truly carcinogenic in humans. The test is useful as a screening tool because it is inexpensive and quick. It is used to eliminate potential mutagens and carcinogens early in the process of developing new drugs and other substances.

## 12 OTHER IN VITRO TESTS

Other in vitro tests have been developed to assess potential immunotoxicity, hormone action, eye irritation, and other changes that are correlated with disease. However, only a few of these have been validated. The advantages of in vitro tests are that they are quick, relatively inexpensive, and that specific mechanisms of action can be tested. The major disadvantage of these tests is that the homeostatic mechanisms, repair pathways, and hormonal regulatory systems found in mammals are not present. This makes it very difficult to extrapolate the data from most of these tests to assess risk to the human species.

## 13 TIERED TOXICITY TESTING

The strategy now used for toxicity evaluation of substances is described as a tiered strategy. The first tier includes in vitro and acute studies. From this, further tiers proceed to a final tier of multigenerational studies and carcinogenesis bioassays. Special tests may be needed for special toxicity problems, for example, endocrine disruption or modification, now a matter of great concern.

## 14 ESTABLISHING SAFE EXPOSURE LEVELS

### 14.1 For Occupational Health and Safety [11]

Safety factors, better called uncertainty factors, are used in establishing public health and regulatory guidelines and standards based on the available toxicological data. For substances producing toxic effects with a threshold dose, the regulatory values are usually based on surrogate values for the threshold. These values may be the NOAEL, the BMD, the LOAEL, the limit dose, or, in the absence of any better datum, the $LD_{50}$ (the median dose lethal to half the test organisms). The surrogate value is reduced by dividing it by an uncertainty factor of 10 to allow for the differences between mammals and humans and the possibility that humans are more sensitive. Another 10-fold factor is applied to allow for extremely sensitive groups that may occur in the human population. The resultant 100-fold ($10 \times 10$) safety level is the minimum normally used for establishing regulatory standards. An additional uncertainty factor of 10 may be applied where there is reason to suspect that the available data are poor or where there may be special sensitivity considerations, thus dividing the

surrogate value by 1000 (10 × 10 × 10) in order to derive a permissible exposure level. Ultimately, scientific judgment must be applied in the use of uncertainty factors and robust data from epidemiological studies on humans may permit a minimum uncertainty factor of 10 to be applied in a very small number of cases.

### 14.2 For Food Additives

Safety assessment for food additives, such as colorants or flavors, is normally calculated as an acceptable daily intake (ADI) for the relevant foodstuff containing the additive [12]. As usual, the calculation is based on the NOAEL (or other threshold surrogate) in laboratory mammals, with the application of appropriate uncertainty factors. The ADI is defined as an estimate by the World Health Organization (WHO) Joint Expert Committee on Food Additives (JECFA) of the amount of a food additive, expressed on a body weight basis, that can be ingested daily over a lifetime without appreciable health risk. For the calculation of the ADI, a standard body mass of 60 kg is assumed. Tolerable daily intake (TDI) is the analogous term used for contaminants.

### 14.3 For Therapeutic Agents

Safety assessment for therapeutic agents is also based on mammalian toxicity data and is a function of the dose to be prescribed, the diseases to be treated, and the nature of the population at risk [13, 14]. The safety assessment for an over-the-counter (OTC) drug is more stringent than for a drug to be used for the treatment of a life-threatening situation.

A therapeutic index (TI) can be developed for most drugs. The TI is the ratio of the toxic dose to the therapeutically effective dose. The greater TI, the greater is the margin of safety in use of the drug. A drug used for cancer chemotherapy may have a greater toxicity and a much smaller TI than a drug used to treat the common cold. A cancer patient is in a life-threatening situation and during treatment is constantly monitored by a doctor or nurse for signs and symptoms of toxicity. Thus, the drug treatment can be suspended quickly if the toxic effects become a problem. The situation is different for a person taking uncontrolled amounts of an OTC product. Such a person is not subject to regular monitoring by a health professional. Furthermore, there is a very large consumer population for OTC products and a significant part of this population is unaware of the risks of toxicity associated with these products.

### 14.4 For Household Products

Household products present particularly difficult problems for toxicity assessment and regulation. Anything can happen to such products, and it is possible that the consumer may be exposed to them by any route. Because of this, packaging and labeling must be considered as part of the safety assessment and regulation. Thus, a household cleanser that is caustic and highly oxidizing must be given strong, vivid, unequivocal labeling and childproof packaging.

## 15 TOXICOKINETICS

Whatever tests are done, there is always the possibility of failing to identify a potentially toxic substance. This possibility may be reduced if the way in which the mammalian body handles the substance is known. The physiological handling of substances by the body is usually considered in terms of absorption, distribution, metabolism, and excretion, the relevant science being described by the term toxicokinetics.

Absorption, the process that links chemical exposure to internal dose, normally occurs by ingestion, inhalation, or through the skin, although therapeutically it may involve intravenous, intramuscular, or other forms of injection or implantation into the body. Processes of distribution, metabolism, and excretion vary depending upon the route of absorption. Toxicokinetic consequences then depend upon the site of injection or implantation and must be assessed specifically for each case. Rectal or vaginal suppositories require special consideration.

Distribution from the site of absorption may involve both blood circulation and lymphatic circulation. Metabolism of the absorbed substance may occur at any stage of distribution and in any cell, tissue, or organ. The liver is a major metabolic organ, and it is particularly important for substances absorbed from the small intestine following ingestion. Some substances, such as benzene, are harmless until they are metabolized by the body to form toxic derivatives. Other substances pass through the body unchanged, while others react directly with receptor molecules.

The chief excretory pathways are urination, defecation, and exhalation. Lactation is also a means of excretion, particularly of fat-soluble chemicals, to the potential detriment of the infant. Under hot conditions, sweating may be more important than urination for the excretion of water-soluble substances. Under these circumstances, volatiles may be lost to the air while some substances may be reabsorbed through the skin. Some substances such as mercury or inorganic arsenicals may be excreted in hair and nails. A minor but important route of excretion is into genital secretions. Substances excreted by this route may adversely affect fertility.

Sensitivity to toxicity varies from person to person, depending on factors such as genetics, age, sex, and environmental conditions. It is generally thought that children, the elderly, and those with pre-existing disease may be more susceptible to toxicity than healthy adults but this may not always be true.

## 16 INTERACTION OF GENES AND TOXICITY [15]

Genes act by programming nucleic acids and proteins that support the formation of a molecular system in cells and organisms, including people, upon which environmental factors act. Among these environmental factors are chemicals, drugs, and infectious agents in air, water, food, and consumer products. The environmental factors are in addition to other physical and psychosocial stresses, and all these stresses may interfere with normal cellular functioning of the system and cause disease. For example, polycyclic aromatic hydrocarbons are produced by combustion of oil, coal, cigarettes, and various industrial processes; these chemicals are breathed in through the lungs, enter the circulating blood, and are converted in the liver and other organs into highly reactive intermediates that form chemical adducts with any DNA they contact. Adducts cause the DNA genetic code to be misread, leading to mutations that may be harmful in themselves and may ultimately cause cancers.

Many pharmaceutical agents have variable therapeutic effects and toxic effects in different patients. Drugs may be metabolized (changed by enzyme action) into more active, or less active, molecules, depending on the inherited form of the gene coding for a relevant enzyme. The same applies to nutrients. Discovery of this gave rise to the concept of "inborn errors of metabolism," a phrase introduced by Sir Archibald Garrod in 1908. It was discovered in the case of the amino acid phenylalanine that specific mutations, usually involving both the maternal and the paternal forms of the gene (an autosomal recessive pattern of inheritance, with 25% risk for each offspring), caused deficiency of a key enzyme necessary for its metabolism. Because of this, in those with recessive genes from both parents, phenylalanine became toxic to the developing brain at the doses obtained from a normal diet. The effects of the resultant disorder (phenylketonuria, or PKU) can be prevented by diagnosing the condition at birth through newborn screening of a heel-stick blood sample and putting the child on a diet low in phenylalanine for the first 5 years of life, while the brain is growing rapidly. The special diet can be less stringent during childhood and adolescence. For women, it is crucial that they are put on a stringent low-phenylalanine diet during pregnancy; otherwise, the high circulating levels will damage the fetus with a 100% risk of mental retardation.

A classic genetic disorder affecting tissue response to xenobiotics is glucose-6-phosphate dehydrogenase (G6PD) deficiency. G6PD is the first rate-limiting enzyme in the pentose phosphate shunt pathway, essential to maintenance of the integrity of the red blood cell membrane. Mutant types of G6PD associated with enzyme deficiency are found at significant frequencies in black and Mediterranean populations. Since the gene for G6PD is carried on the X chromosome, males are most likely to be affected. Many drugs with oxidizing properties can cause acute hemolysis in otherwise healthy but genetically predisposed people by overwhelming the capacity of G6PD-deficient red blood cells to counteract oxidative stress. Such stress may also be caused by exposure to a component of fava beans, the pyrimidine aglycone divicine, or to naphthoquinone in mothballs. Hepatitis and diabetic acidosis are also associated with red blood cell damage resulting from G6PD deficiency.

Other examples of well-established examples of genetic disorders affecting drug metabolism are:

1. Deficiency in succinylcholinesterase activity: Succinylcholine causes prolonged and life-threatening paralysis of respiratory muscles in the 1 in every 2000 people who inherit two copies of the recessive gene for an atypical form of the enzyme, plasma pseudocholinesterase; otherwise, such people are entirely normal.
2. Deficiency in acetylation: A recessive mutation in hepatic *N*-acetyltransferase reduces the activity of this important acetylating enzyme; slow acetylators are liable to side effects from hydralazine, isoniazid, dapsone, and some sulfa drugs; about 50–55% of Caucasian and black populations are slow acetylators, whereas only 10–15% of Orientals are slow acetylators. According to Calabrese [15], the hypothesis that slow acetylators are at increased risk of arylamine-induced bladder cancer has good theoretical foundations in animal model studies; epidemiological evidence is consistent with the animal data.
3. The 4% of the general population who are carriers of the serum alpha-antitrypsin (SAT) deficiency are at enhanced risk (2 times normal) of developing pollutant-induced emphysema.

More knowledge is needed about the heterogeneity of genetic predispositions, environmental exposures, and disease risks. Unfortunately, most public health research on infectious disease and environmental chemical risks has paid little attention to inherited

susceptibility in people, focusing only on the environmental disease agents. Heterogeneity of study populations has often been neglected in epidemiological studies in order to generate sufficient numbers to justify the analysis statistically. For quantitative traits, pharmacologists, toxicologists, and psychologists have generally emphasized means and standard errors of the means, and neglected potentially informative people with values outside two standard deviations from the mean.

Nutrition and genetics interact extensively. A fairly common genetic disorder, hereditary hemochromatosis, results in sensitivity to iron in the diet at relatively low amounts, leading to damage from iron deposition in the heart, liver, pancreas (diabetes), testes (infertility), skin, and joints (arthritis).

## 17 ECOTOXICOLOGY [16]

For some time, there has been a requirement in many countries for premanufacturing notification of new chemicals and their properties, including their toxicity, both to humans and to other organisms that may be at risk. Recognition of the dangers inherent in compounds, and especially those that bioaccumulate or otherwise persist in the environment, has led to further tests to identify and to exclude from commerce such compounds, including pharmaceuticals since they are designed to be biologically active and must ultimately reach the natural environment in some form. This requires making due allowance for the chemical transformations that occur in treated humans and animals and under environmental conditions.

Toxic effects of substances on ecosystems may have an adverse effect on human health. Historically, ecotoxicology (in relation to human health) has concentrated on contamination of the food chain, especially the bioaccumulation of toxic agents in food. Now there is an understanding that there are effects of substances on climate, desertification, and crop yield, all of which impact upon human health. The natural world contributes enormously to human well-being, and an appreciation of this has led to a body of legislation designed to prevent or minimize further environmental degradation.

## 18 ECOSYSTEMS

The health of humans, like that of all living organisms, is dependent on living in a healthy ecosystem. Ecosystem degradation can lead to human health risks from unsafe drinking water, polluted air, climate change, emerging new diseases, and the resurgence of old diseases resulting from ecological imbalances.

An ecosystem may be defined as a community of plants and animals interacting with each other and with their environment. Ecosystems may be differentiated on the basis of the dominant vegetation, the topography, the climate, or other criteria. For example, prairies are characterized by the predominance of grasses. The Arctic tundra is determined partly by its harsh climatic zone. In most parts of the world, the human community is the dominant component of the ecosystem. Ecosystems include not only natural areas (e.g., forests, lakes, or marine coastal systems) but also manmade systems (e.g., urban ecosystems, agro-ecosystems, or impoundments). Human populations are increasingly concentrated in urban ecosystems, and it is estimated that already 50% of the world's population is living in urban areas.

Any given area comprises a mosaic of ecosystems and includes towns, rivers, lakes, agricultural systems, and so on. Precise boundaries between ecosystems are often difficult to establish. Regions may overlap gradually, with a fairly wide "transition" zone.

## 19 ECOSYSTEM HEALTH AND TOXICITY

*Ecosystem health* is a term that is difficult to define precisely. It is easier to define examples of ecosystems that have become "sick," particularly as a result of human activity. For example, the discharge of nutrients from sewage, industrial waste, or agricultural runoff into lakes or rivers adversely affects the normal functioning of their ecosystems. This is an example of ecosystem toxicity. Excessive nutrient inputs from human activity caused considerable ecological damage to Lake Erie, Lake Ontario, and Lake Michigan. This effect, called *eutrophication* has affected many natural waters around the world.

Resilience is the capacity of an ecosystem to maintain its structure and functions in the face of disturbances. Systems that have been pushed by environmental stresses to the extreme limits of their resilience are less likely to recover from further stress of human pollutants than are those systems that have been relatively less stressed.

Healthy ecosystems can be characterized in economic, social, and human health terms. Healthy ecosystems can support a certain level of human economic activity. This potential should always be given due consideration before decisions on further development are made. If human developments exceed their environmental sustainability, ecosystem degradation will ensue and undermine the long-term sustainability of the human economy that is associated with it. In the short

term, ecological damage may not be evident, as natural capital (e.g., soils and otherwise renewable resources) may tolerate overuse temporarily and permit enhanced economic returns.

## 20 ECOSYSTEM PATHOLOGY

Human activities paralleling human population increases are now the main cause of healthy ecosystems becoming sick. Natural disturbances (fires, floods, drought, periodic insect infestations, or infectious disease) are built in to the dynamics of most ecosystems. These processes help to "reset" ecosystems by recycling nutrients and clearing space for recolonization by biota that may be better adapted to changing environments. On the other hand, stresses imposed on ecosystems as a result of human activity, such as the release of nutrients and toxic substances into the air, water, and land, often result in chronic ecological dysfunction.

Waste chemical discharges from municipal, industrial, and agricultural sources into air, water, and land have severely damaged many of the Earth's ecosystems. The effects are particularly apparent in aquatic ecosystems, for example, the effects of atmospheric acid precipitation or other acidic discharges. In some lakes that lack a natural buffering capacity, acid precipitation has eliminated most of the fish and other organisms as a result of the pH of the water being lowered to a level that made aluminum ions available in solution at a toxic concentration. Systems that once contained a variety of organisms and were highly biologically productive have become devoid of most life-forms except for a few acid-tolerant bacteria and sediment-dwelling organisms.

## 21 ECOSYSTEM HEALTH AND HUMAN HEALTH

An important aspect of ecosystem degradation is the associated increased risk to human health. Traditionally, the concern has been with contaminants, particularly industrial chemicals that can have adverse impacts on human development, neurological functions, reproductive functions, and that appear to be causative agents in a variety of carcinomas. In addition to these environmental concerns (where the remedy is often reduction in the release of contaminants), there are a large number of other risks to human health stemming from ecological imbalance.

Ecosystem pathology results in the loss of valued ecosystem services, including flood control, water quality, air quality, fish and wildlife diversity, and recreation. One of the major signs of ecosystem pathology is increased disease incidence, both in humans and other species. Human population health can thus be viewed in an ecological context as an expression of the integrity and health of the life-supporting capacity of the environment.

Ecological imbalances triggered by global climate change, itself believed to be a product of human action releasing "greenhouse gases" into the atmosphere, lead to increased human health risks. Global climate change, particularly as it impacts both temperatures and precipitation, is highly correlated with the prevalence of vector-borne diseases.

The increase in prevalence of *Vibrio cholerae* has been strongly linked to degraded coastal marine environments [17]. Nutrient-enriched warmer coastal waters, resulting from a combination of climate change and the use of fertilizers, provide an ideal environment for reproduction and dissemination of *V. cholerae*. Nutrient enrichment and warmer temperatures give rise to algal blooms and an abundance of macrophytes. The algal blooms provide abundant food for copepods, and the increasing copepod and macrophyte populations provide *V. cholerae* with a habitat. Subsequent dispersal of *V. cholerae* into estuaries and freshwater bodies allows contact with humans who use these waters for drinking and bathing. Global distribution of marine pathogens such as *V. cholerae* is further facilitated by ballast water discharge from vessels. Ballast water is transported from its source over long distances to the discharge point and may contain a cocktail of pathogens, including *V. cholerae*.

## 22 ANTIBIOTIC RESISTANCE AND AGRICULTURAL PRACTICE

Antibiotic resistance is a growing threat to public health [18]. Antibiotic resistant strains of *Streptococcus pneumoniae*, a common bacterial pathogen in humans and a leading cause of many infections, including chronic bronchitis, pneumonia, and meningitis, have greatly increased in prevalence since the mid-1970s. In some regions of the world, up to 70% of bacterial isolates taken from patients proved resistant to penicillin and other β-lactam antibiotics. The use of large quantities of antibiotics in agriculture and aquaculture appears to have been a key factor in the development of antibiotic resistance by farm animal pathogens that may subsequently infect humans. One of the most serious risks to human health from such practices is vancomycin-resistant enterococci. The use of avoparcin, an animal growth promoter, appears to have compromised the utility of vancomycin, the last antibiotic effective against multi-drug-resistant bacteria. In areas where avoparcin

has been used, such as on farms in Denmark and Germany, vancomycin-resistant bacteria have been detected in meat sold in supermarkets. Avoparcin was subsequently banned by the European Union. Another example is the use of ofloxacin to protect chickens from infection and thereby enhance their growth. This drug is closely related to ciprofloxacin, which was one of the most widely used antibiotics in the year 2000. There have been cases of resistance to ciprofloxacin directly related to its veterinary use. In the United Kingdom, ciprofloxacin resistance developed in strains of *Campylobacter*, a common cause of diarrhea. Multidrug-resistant strains of *Salmonella* have been detected in European egg production facilities.

## 23 ECOSYSTEM RESTORATION

Ecosystem pathology can be reversed in some cases simply by removing the source of stress. In cases, for example, where ecosystem degradation is the result of point-source additions of nutrients or toxic chemicals, removal of these stresses may result in recovery of ecosystem health. A classic case is Lake Washington (near Seattle, Washington) [19]. This lake had become highly anoxic (oxygen depleted) owing to a sewage outfall entering the lake. Redirecting the sewage outfall away from the lake reversed many of the signs of ecopathology.

In cases where it is not feasible to remove the source of stress, it may be possible to minimize the stress. For example, in the Kyrönjoki and Lestijoki Rivers in western Finland, spring and autumn runoff leads to sharp pulses of acidity because of release of acid from tilled or dug soils. This runoff has been particularly damaging to fish during the critical time of year for spawning. Fish reproduction is severely curtailed, if not all together eliminated in highly acidic water. There have also been massive fish kills resulting from the highly acidic waters. Replacement of the original drains that take runoff to the rivers with new drains containing lime to neutralize the acidity appears to provide at least a partial solution to the problem [20].

## 24 PREVENTION OF ECOSYSTEM DAMAGE

Because it is very difficult to reverse ecosystem degradation, and because of the many human health problems associated with the loss of ecosystem health, it is clear that every effort must be made to prevent ecosystem disruption. In both developed and developing countries there is strong pressure for economic growth in spite of accompanying environmental damage. The justification for this is that economic growth has many obvious health benefits, such as providing more plentiful food, and funding for better health services as well as supporting research to improve standards of living.

However, the human population is at an all-time high, and the associated human activities have led to more and more degradation of the Earth's ecosystems. In this respect, there is growing concern about the contamination of the environment by therapeutic agents and their derivatives as components of human waste. These substances have been detected in rivers and associated with effects on wildlife, including feminization of male fish by hormonal therapeutics and their excretory products. Thus, a new area of toxicology is developing as more and more regulators require assessment of possible environmental effects of human and veterinary therapeutics at all stages from production to final disposal by whatever means.

## REFERENCES


1. Klaasen CD (2008). *Casarett and Doull's Toxicology—The Basic Science of Poisons*. New York: McGraw-Hill Medical.
2. Duffus JH, Worth HGJ (2006). *Fundamental Toxicology*. Cambridge: Royal Society of Chemistry.
3. Timbrell J (2002). *Introduction to Toxicology*. Boca Raton, FL: CRC Press.
4. Hertzberg R, Teuschler L (2002). Evaluating quantitative formulas for dose-response assessment of chemical mixtures. *Environ Health Perspectives 110* (Suppl 6), 965–970.
5. Hayes AW (2008). *Principles and Methods in Toxicology*. Boca Raton, FL: CRC Press.
6. United Nations Economic Commission for Europe (UNECE) (2010) GHS—The Globally Harmonized System of Classification and Labelling of Chemicals, Revision 4. Available at: http://live.unece.org/trans/danger/publi/ghs/ghs_welcome_e.html
7. Office of Pesticide Programs (OPP) (2004). Chemical hazard classification and labeling: Comparison of OPP requirements and the GHS. Available: http://www.epa.gov/oppfead1/international/globalharmon.htm.
8. Organisation for Economic Co-operation and Development (OECD) (1981–present). Guidelines for the testing of chemicals. Available: http://www.oecd.org/document/40/0,3343,en_2649_34377_37051368_1_1_1_1,00.html.
9. Kirkland DJ (2005). *Basic Mutagenicity Tests*. Cambridge: Cambridge University Press.
10. Dahlstrom DL, Buckalew M (2007). In Hayes AW, Ed. *Principles and Methods in Toxicology*, 5th ed. Boca Raton, FL: CRC Press, pp. 633–692.
11. Stellman JM (1998). *ILO Encyclopaedia of Occupational Health and Safety*. Geneva: International Labour Office.

12. Reddy CS, Hayes AW (2007). In Hayes AW, Ed. *Principles and Methods in Toxicology*, 5th ed. Boca Raton, FL: CRC Press, pp. 633–692.
13. Dorato MA, McMillan CL, Vodicnic MJ (2001). In Hayes AW, Ed. *Principles and Methods in Toxicology*. Philadlphia: Taylor and Francis, pp. 325–368.
14. Hulot, J-S. (2010). Pharmacogenomics and personalized medicine: Lost in translation? Genome Medicine, 2–13, doi: 10.1186/gm13.
15. Calabrese EJ (1988). Comparative biology of test species. *Environ Health Perspect 77*, 55–62.
16. Walker CH, Hopkin SP, Sibly RM, Peakall DB (2005). *Principles of Ecotoxicology*. Boca Raton, FL: CRC Press.
17. Epstein PR (1993). Algal blooms in the spread and persistence of cholera. *Biosystems 31*(2–3), 209–221.
18. Lewis R (1995). The rise of antibiotic-resistant infections. Available: http://www.fda.gov/Fdac/features/795_antibio.html.
19. King County WA (2005). The Lake Washington story. Available: http://dnr.metrokc.gov/wlr/waterres/lakes/biolake.htm.
20. Bärlund I, Tattari S, Yli-Halla M, Åström M (1999). Effects of lime filter drains and controlled groundwater level on drainage water quality on acid sulphate soils in Finland. Paper presented at the ICID 21st European Regional Conference. May 15–19, 2005. Available: www.zalf.de/icid/ICID_ERC2005/HTML/ERC2005PDF/Topic_3/ExtAbstr/Baerlund.pdf.

# DEVELOPMENT OF PHOSPHODIESTERASE 4 INHIBITORS FOR ALLERGIC AND NONALLERGIC INFLAMMATORY DISEASES OF THE AIRWAYS

Mark A. Giembycz

*Department of Physiology & Pharmacology Institute of Infection, Immunity and Inflammation, Faculty of Medicine, University of Calgary, Calgary, Alberta, Canada*

## 1 INTRODUCTION

Asthma and chronic obstructive pulmonary disease (COPD) are the most common respiratory disorders and together, in the global population, currently afflict >1.5 billion individuals. The prevalence of asthma is rising worldwide, and this has become the most common chronic disease in industrialized nations with 5–10% of the population being affected [1]. For example, in the United States, asthma attacks are a primary cause of hospitalization—497,000 discharges due to asthma were recorded in 2004—and, in the preceding year, were responsible for 4099 deaths [2] (see www.lungusa.org). Asthma is a chronic inflammatory disease characterized by *reversible* airway obstruction, airway remodeling, and nonspecific airway hyperresponsiveness (AHR) [3, 4]. The inflammation in asthma is multifaceted, often allergic and involves both infiltrating leukocytes and structural cells. $CD4^+$ T lymphocytes are believed to play a pivotal pathogenic role; they elicit a Th2-driven inflammatory response and, through orchestrating pulmonary eosinophil recruitment, may contribute to airway remodeling, which is a cardinal feature of this disease [5–7].

COPD is a nonspecific term that embraces several debilitating inflammatory pathologies that often coexist and is characterized by a slowly progressive and largely *irreversible* decrement in lung function [8–10]. Persistent chronic airflow limitation, usually associated with airway collapse, edema, mucus hypersecretion, and fibrosis, is present to a greater or lesser extent and accounts for the wide spectrum of the disease [8–10]. COPD afflicts middle-aged and elderly people [11] and is caused, in approximately 95% of all cases, by chronic cigarette smoking, although other environmental insults such as the burning of biomass fuels are also major risk factors [12, 13]. COPD has a long latency before symptoms become evident, afflicts 9–10% of the population globally over the age of 40 years [14], and *is the only cause of mortality that is increasing worldwide* [15]. Indeed, by 2020 the Global Burden of Disease Study predicts that COPD will be the third leading cause of death, driven by the expanding epidemic of cigarette smoking and changing demographics, with more of the global population living longer [16, 17]. Even in 2007, COPD was underrecognized, underdiagnosed, and underreported [18]. Accordingly, COPD is undertreated [13].

A recent systematic review and meta-analysis of studies performed in 28 countries between 1994 and 2004 provides compelling evidence that the prevalence of COPD is appreciably higher in individuals who smoke or who are ex-smokers when compared to subjects who have never smoked [14]. Accordingly, COPD is largely preventable, and, in the early stages, its progression can be arrested by smoking cessation. Unfortunately, the

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

highly addictive properties of nicotine in tobacco render most individuals unable to renounce cigarette smoking despite intensive interventions.

## 2 GLUCOCORTICOIDS AND THE TREATMENT OF ASTHMA AND COPD

Inhaled corticosteroids (ICSs) or, more correctly, glucocorticoids are the most efficacious treatment currently available for asthma (see www.ginasthma.org). These drugs control symptoms, reduce exacerbations, and improve health status in most patients irrespective of disease severity [19]. Accordingly, ICSs are the first-line therapy for all patients who use a $\beta_2$-adrenoceptor agonist inhaler more than once a day, and this is reflected in all national and international guidelines for asthma management (see www.ginasthma.org) [20–23]. In contrast, unlike in individuals with asthma, ICSs appear to be of *relatively* limited benefit in subjects with COPD. Although there is some evidence that certain indices of pulmonary inflammation are suppressed by ICSs (reviewed in ref. 24), the effect on mortality is unclear (see [25]) [26–31]. Currently, there is only one published *prospective* trial of ICSs in COPD that has been powered on mortality and, contrary to *retrospective* analyses [32, 33], that study demonstrated that the rate of death in subjects taking the ICS, fluticasone propionate, alone did not differ from placebo [34]. In contrast, ICSs have unequivocal and clinically meaningful anti-inflammatory efficacy in subjects with asthma and significantly reduce mortality [35].

## 3 PHOSPHODIESTERASE INHIBITORS

Although glucocorticoids can effectively alleviate the inflammation that characterizes asthma, they are not without adverse effects [36, 37]. Accordingly, new drugs with enhanced selectivity and improved side effect profiles are clearly required that also have efficacy in subjects with COPD in whom glucocorticoids are relatively refractory. One group of drugs that, from a theoretical perspective, may exhibit powerful anti-inflammatory and immunomodulatory activity are inhibitors of a large family of enzymes, generically known as cyclic nucleotide phosphodiesterases (PDEs). Eleven structurally and immunologically distinct families have, thus far, been identified that are differentially expressed between different cell types and can also be distinguished from each other in terms of their substrate specificity, dependency on cofactors, and sensitivity to inhibitors (see Table 1 for details) [39–44]. With respect to treating airway inflammatory diseases, it is those enzymes that comprise the PDE4 isoenzyme family—which degrade cyclic adenosine-3′,5′-monophosphate (cAMP)—that are the target for small-molecule inhibitors. The rationale for developing compounds that attenuate PDE4 activity is based on three critical findings: (i) PDE4 is the major regulator of cAMP metabolism in almost every pro-inflammatory and immune cell, and agents the elevate cAMP have a pharmacology consistent with anti-inflammatory activity; (ii) PDE4 inhibitors, of varied structural classes, suppress a myriad of in vitro responses such as cytokine generation, nicotinamide adenine dinucleotide phosphate (NADPH) oxidase activity, degranulation, proliferation, mediator generation, and chemotaxis; and (iii) PDE4 inhibitors are efficacious in animal models of inflammation [42]. If these observations hold in humans, then, conceptually, PDE4 inhibitors should show a pleiotropic inhibitory profile of activity on pro-inflammatory and immune cell function. Despite the potential for adverse events due to the ubiquitous distribution of PDE4 in "nontarget" cells (see below), almost all of the main pharmaceutical companies in the world have synthesized and evaluated in the clinic PDE4 inhibitors of varied structural classes.

It is the purpose of this chapter to describe the evolution of PDE4 inhibitors as novel anti-inflammatory drugs, the distribution of PDE4 across cells that participate in allergic and nonallergic inflammatory reactions, and potential sites where selective PDE4 inhibitors act to alleviate the inflammation associated with asthma and COPD. Clinical data are also provided for one PDE4 inhibitor, roflumilast (Daxas), which is in late-stage clinical development for both of these diseases.

## 4 STANDARDIZED NOMENCLATURE

A standardized GenBank nomenclature for existing and newly discovered PDEs was introduced in 1994 [45]. This taxonomy is based on traditional biochemical criteria and on protein sequencing and analyses of partial and full-length complementary deoxyribonucleic acid (cDNA) clones. Consider two hypothetical PDEs, CPPDE4D4B and HSPDE11A1A (Fig. 1). The first two letters refer to the species (CP for *Cavia porcellus*, HS for *Homo sapiens*). The next three letters plus an Arabic numeral correspond to the PDE gene family (i.e., PDE4, PDE11), while the next letter and Arabic numeral (i.e., D4, A1) denote the gene product and spliced variant (if appropriate). This nomenclature will be used throughout this chapter.

## 5 ROLIPRAM: A PROTOTYPE PDE4 INHIBITOR

One of the first selective inhibitors of PDE4 to be reported was rolipram (**1**) [46] together with the less

**TABLE 1 Properties and Selective Inhibitors of PDEs**

| PDE | Family | $K_m$ (μM) cAMP | $K_m$ (μM) cGMP | Selective Inhibitors |
|---|---|---|---|---|
| 1 | $Ca^{2+}$/Calmodulin stimulated | 0.3–124 | 0.6–6 | SCH 57801<br>KS-505a<br>IC-224 |
| 2 | cGMP stimulated | 28–100 | 10–28 | ENHA<br>Oxindole<br>BAY 60-7550<br>IC933<br>PDP |
| 3 | cGMP inhibited | 0.02–0.3 | 0.1–0.5 | Siguazodan,<br>SK&F 95654<br>Cilostamide<br>Cilostazol |
| 4 | cAMP specific | 1–10 | >50 | Rolipram<br>Ro 20-1724<br>Cilomilast<br>Roflumilast<br>Piclamilast |
| 5 | cGMP binding | 280 | 3–7 | Zaprinast<br>Sildenafil<br>Vardenafil<br>Tadalafil |
| 6 | Photoreceptor | 600–700 | 15–17 | None reported |
| 7 | Rolipram insensitive, cAMP specific | 0.01–0.5 | NA[a] | BRL 50481<br>PF 332040<br>IC242 |
| 8 | Rolipram insensitive, cAMP specific | 0.06 | NA | PF04957325 |
| 9 | Zaprinast insensitive, cGMP specific | 228 | 0.7–0.17 | BAY 73-6691 |
| 10 | cAMP inhibited (?) | 0.22–1.1 | 13–14 | None reported |
| 11 | Dual selective | 1–2 | 2–3 | None reported |

[a]NA = Reliable data not available.

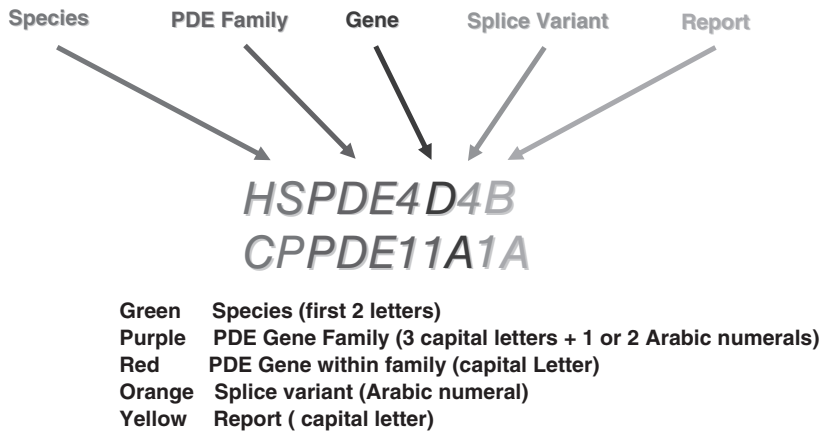


**Figure 1** GenBank nomenclature for the classification of PDEs. The first two letters (green) represent the species, the next three letters plus an Arabic numeral denote the PDE family (purple), the next letter and numeral refer to the gene product within the PDE family (red) and the spliced variant (if appropriate; orange), respectively, and the last letter indicates the report (yellow). See text for further details.

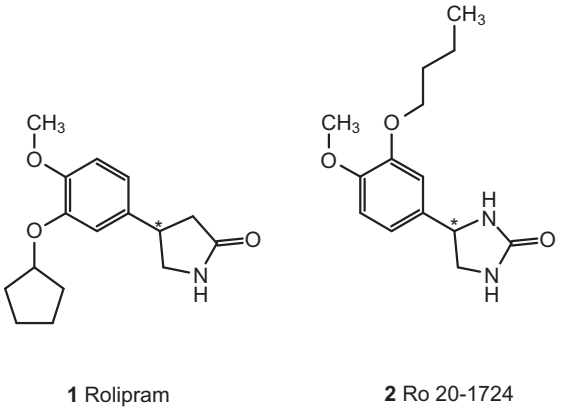


**Figure 2** Chemical structures of rolipram [**1**; (±) 4-[3-(cyclopentyloxy)-4-methoxyphenyl]pyrrolidin-2-one) and Ro 20,1724 (**2**; (+/–) 4-(3-butoxy-4-methoxyphenyl) imidazolidin-2-one]. Asterisk (*) denotes chiral center.

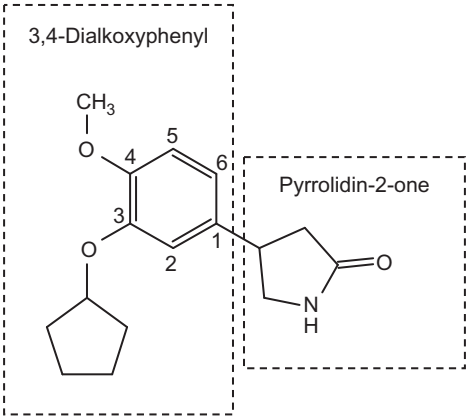


**Figure 3** Two pharmacophores of rolipram.

potent, structural analog, Ro-20,1724 (**2**) [47] (Fig. 2). Rolipram is composed of two pharmacophores: a 3-cyclopentoxy-4-methyl group and a pyrrolidine-2-one, which is also known as a γ-lactam (Fig. 3). Depression was the first indication in which rolipram was evaluated clinically [48, 49], and initial studies reported good efficacy [50]. However, the optimism that resulted from these early data was short lived as subsequent trials were much less convincing and associated with side effects, of which nausea was particularly problematic [51]. A similar lack of efficacy was found in Parkinson's disease and, again, dosing was associated with episodes of nausea [52, 53]. Thus, nausea together with other gastrointestinal events of concern resulted in the development of rolipram being discontinued.

## 6 PDE4 INHIBITORS BEYOND ROLIPRAM

Despite the failure of rolipram, a large number of PDE4 inhibitor research programs have been initiated in the hope that emetic liability can be reduced or even eliminated (reviewed in [54]). Currently, several PDE4 inhibitors of varied structural class are in clinical development for asthma, COPD, as well as other disorders including Alzheimer's disease, memory impairment, and multiple sclerosis (Table 2; Fig. 4). However, none have yet reached the market, and many have been discontinued presumably due to lack of efficacy and/or dose-limited side effects*. Nevertheless, optimism for this class of compounds remains high, and a discussion of their potential therapeutic utility still remains topical.

## 7 PHOSPHODIESTERASE 4

A cAMP-specific PDE inhibited by Ro 20-1724, but largely unaffected by cyclic guanosine monophosphate (cGMP), was originally purified from canine kidney [55, 56] and was the first report of a member of a large family of similar proteins collectively known as PDE4 isoenzymes. Since that original description, PDE4s have been purified and partially characterized from a number of sources, including human monocytes [57] and human leukocytes [58]. Human recombinant proteins have also been heterologously expressed and purified to facilitate investigations of structure–function relationships and to aid in the identification of selective inhibitors by high-throughput screening [59]. Without apparent exception, PDE4 isoenzymes are acid proteins ($p$I = 4–6) and exclusively hydrolyse cAMP.

In 1989 Davies and colleagues isolated and cloned the *Drosophila melanogaster dunce* gene and found that it encoded a PDE of the PDE4 family. In the same year four rat cDNA homologs [60–62] of the *dunce* cAMP PDE were identified, establishing a molecular basis for the observed heterogeneity of gene products within the PDE4 family. These clones represent transcripts of four different genes and are now named *RNPDE4A*, *RNPDE4B*, *RNPDE4C*, and *RNPDE4D* according to the nomenclature proposed by Beavo and colleagues [45] (see Fig. 1 for details). Four human genes (*A* to *D*) have also been identified and many of their products studied in some detail (see ref. 38) [39]. An astonishing finding that emerged from the molecular cloning of PDE4 isoenzymes is the presence of (messenger ribonuclerc acid (mRNA) transcripts of different sizes for

*The PDE4 inhibitor roflumilast (Daxas/Daliresp) was recently approved in the European Union and in the USA for the treatment of a subpopulation of patients with severe COPD.

**TABLE 2 Development State of PDE4 Inhibitors**

| Company | Drug | Indication | Development Stage |
|---|---|---|---|
| Almirall | Arofylline | Bronchitis | Phase II/III |
| Glenmark | Oglemilast (GRC 3886) | Asthma, COPD | Phase II |
| Pfizer | Tofimilast (CP-325366) | Asthma | Discontinued |
| ONO | ONO 6126 | Asthma, COPD | Phase II |
| Otsuka | Tetomilast (OPC 6535) | COPD | Phase III |
| Nycomed | Roflumilast | Asthma, COPD | Phase III |
| GlaxoSmithKline | Cilomilast (SB 207499) | COPD | Discontinued |
| GlaxoSmithKline | AWD-12-281 (GSK 842470) | Asthma | Discontinued |
| Celgene | Apremilast (CC-10004) | Asthma | Phase II |
| ICOS | IC-485 | COPD | Discontinued |
| MediciNova | Ibudilast | Multiple sclerosis | Phase II |
| Helican Therapeutics[a] | IPL-455903 (HT-0712) | Memory impairment | Phase II |
| Pfizer | UK-500,001 | COPD | Phase II[b] |
| Merck | MK-0952 | Alzheimer's disease | Phase II |

[a]Licensed from Inflazyme.
[b]Development may have been discontinued.
*Source*: Data collated in part from Mealy and Bayes [160].

each of the four variants (greater than 20 have been unequivocally identified at the time of this writing) that are differentially expressed between tissues [38, 39, 63]. The basis for this profound heterogeneity of PDE4 isoenzymes is attributable to both alternative mRNA splicing and to the fact that the *PDE4* genes express multiple promoter regions and, therefore, provide several potential start codons for the translation of the protein.

Proteins encoded by each *PDE4* gene may be either "long," "short," or "supershort" (see ref. 38) [64]. Structurally, long PDE4 isoforms have an N-terminal domain, two, so-called, upstream conserved regions (UCR1 and UCR2; thought to mediate dimerization through ionic interactions [65, 66]), a highly conserved catalytic center and a unique C-terminus; in contrast, short PDE4 variants lack UCR1, whereas supershort isoenzymes also have a truncated UCR2 [67]. Long isoenzymes are encoded by all *PDE4* genes, whereas, apparently, only *PDE4A*, *PDE4B*, and *PDE4D* encode short forms [64].

### 7.1 Regulation of PDE4 by Protein Kinesis A (PKA)

PDE4 is subject to both short and long-term regulation by agents that elevate cAMP. This finding has led to the proposal (see ref. 68) that upregulation of PDE4 in allergic asthma could *contribute* to the tolerance of airway smooth muscle cells to the bronchodilator and protective effects of $\beta_2$-adrenoceptor agonists [69–78]. Indeed, a meta-analysis of 22 independent randomized, placebo-controlled studies conducted between 1989 and 2001 lead the authors to conclude that regular use of $\beta_2$-adrenoceptor agonists for at least one week in patients with asthma results in tolerance to the drug's bronchodilator activity and that this may be associated with poorer asthma control [72]. It is pertinent to mention that the mechanism of tolerance is not restricted to $\beta_2$-adrenoceptor desensitization. It is likely that in many subjects poor asthma control reduces the efficacy of $\beta_2$-adrenoceptor agonists. For example, mucus impaction may limit the delivery of drug to the desired sites of action and so compromise efficacy by physical means.

Upregulation of PDE4 can occur through either posttranslational modification (e.g., phosphorylation) of existing enzyme or gene induction. PKA was originally shown to phosphorylate the long PDE4 isoform, PDE4D3, at two serine resides: $S^{13}$, located within the unique N-terminus of the protein, and $S^{54}$, located at the beginning of UCR1 [79–85]. Phosphorylation at this second serine residue activates the enzyme (6- to 8-fold) and also increases its sensitivity (100- to 300-fold) to inhibition by certain PDE4 inhibitors including RS 25344 and RS 33793 [80]. The effect on catalytic activity of PKA-mediated phosphorylation is not restricted to PDE4D3. A similar enhancement (~50%) of cAMP degradation occurs upon phosphorylation of the serine residue that corresponds to $S^{54}$ in PDE4D3 in many long isoforms, including PDE4A8, PDE4B1, PDE4C2, PDE4D5, PDE4D7, PDE4D8, and PDE4D9, although, for reasons not yet explained, sensitivity to PDE4 inhibitors is unchanged [85–87].

Similarly, prolonged activation of PKA can lead to the accumulation of certain short PDE4 isoforms. This effect is believed to be mediated by the phosphorylation of the transcription factor, cAMP response element bing protein (CREB) which leads to the generation of PDE4 mRNA from an intronic promoter [62, 87–89].

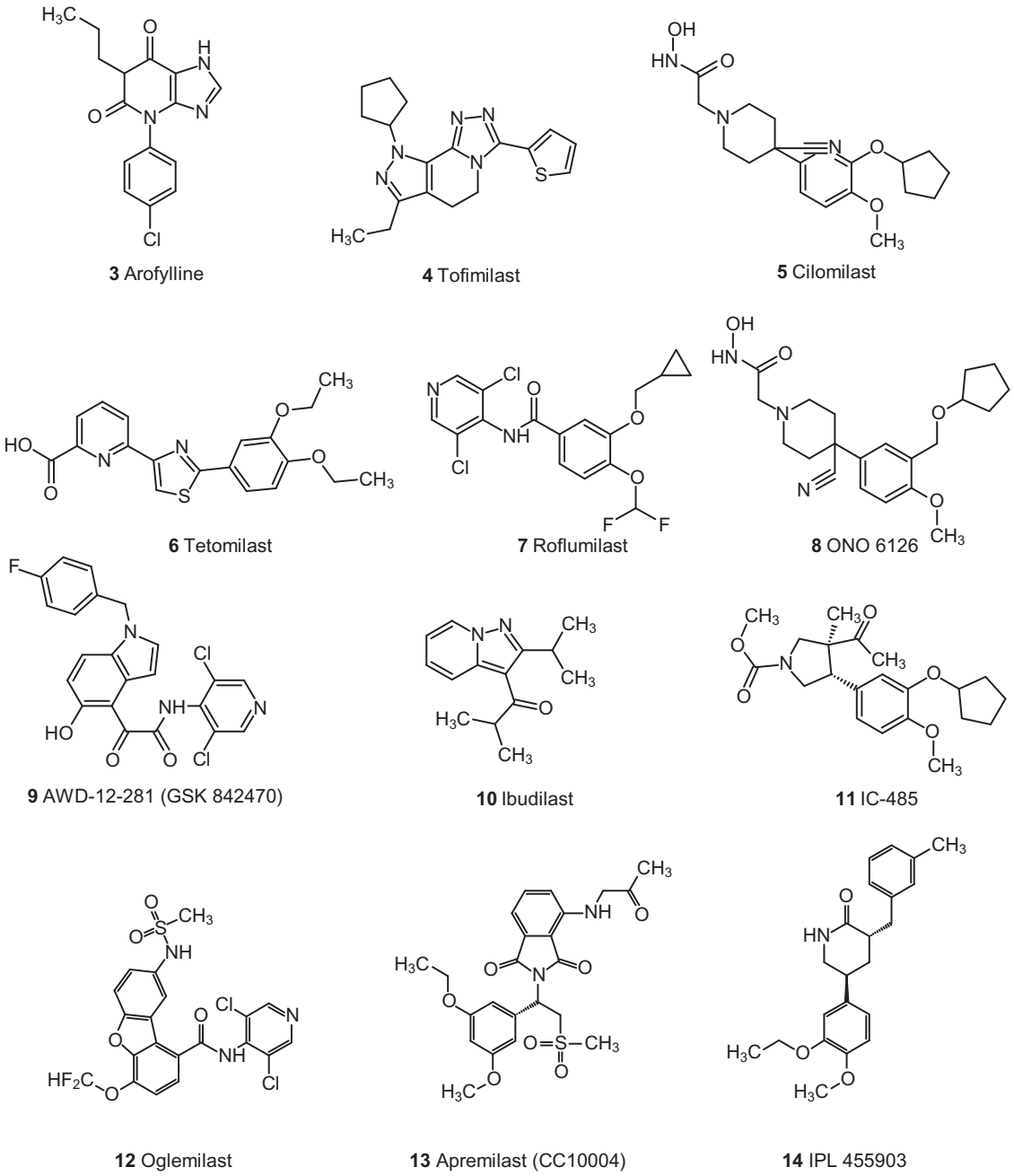


**Figure 4** PDE4 inhibitors that are in, or have been in, development for asthma and/or COPD.

With respect to $\beta_2$-adrenoceptors in the lung, it is the PDE4 isoenzyme family that is a primary regulator of cAMP metabolism [42]. In this paradigm, tolerance to $\beta_2$-adrenoceptor agonists is directly related to an increase in PDE activity. This effect would theoretically compromise cell signaling through all $G_s$-coupled receptors leading to heterologous desensitization of susceptible cells to cAMP-dependent events. It is hypothesized that this could occur as a direct consequence of regular treatment of asthmatic subjects with $\beta_2$-adrenoceptor agonists [68].

While generally ignored, the concept of increased cAMP PDE activity as a mechanism of reducing the sensitivity of cells to hormones and other agonists that interact with Gs-coupled receptors is not new. In fact, evidence that this phenomenon accounts for much of the reduced responsiveness that cells exhibit to chronic hormone exposure was provided in 1978 [90] and has since been documented in vitro in many cells implicated in the pathogenesis of asthma such as T lymphocytes, neutrophils, monocytes, macrophages, platelets, and airways smooth muscle (ASM) [90–102].

An important issue that arises from the aforementioned discussion is whether induction and/or phosphorylation of PDE4 can be demonstrated in immune/pro-inflammatory cells and in vivo in response to $\beta_2$-adrenoceptor agonists. Although limited data are available, the answer to both parts of this question is "Yes." Torphy and colleagues [101] demonstrated that the $\beta_2$-adrenoceptor agonist, salbutamol, and the selective PDE4 inhibitor, rolipram, when given in combination to the human monocytic cell line, U-937, increased PDE4 activity in a time-dependent manner. Significantly, this effect required new protein synthesis, indicating that the increase in enzyme activity was attributable to the induction of one or more PDE4 isogenes. Reverse transcription-PCR (polymerase chain reaction) and western analyses performed by the same authors demonstrated subsequently that salbutamol and rolipram increased the expression of PDE4A and PDE4B at the mRNA and protein levels. A similar investigation by Verghese et al. essentially confirmed these observations [102]. Thus, exposure of human peripheral blood monocytes and Mono Mac 6 cells to cAMP-elevating agents promoted the transcription of the PDE4A, B, and D isogenes with the generation of at least three distinct mRNA transcripts and proteins. Induction of PDE4 isogenes has also been reported in U-937 and Jurkat T cells in response to prolonged exposure to dibutyryl cAMP [103] and, more recently, the same phenomenon was documented in guinea pig macrophages [104], human T lymphocytes [91], human neutrophils [95], and human ASM cells [105]. In the latter study, upregulation of the PDE4D5 splice variant was described, and this may have a particular significance given that this isoform interacts preferentially with β-arrestins and may play a role in $\beta_2$-adrenoceptor desensitization [106, 107].

A consistent and highly significant finding is that the responsiveness of many cells in which PDE4 is induced to cAMP-generating agonists is restored, at least in part, by the addition of a PDE inhibitor providing compelling evidence that upregulation of PDE is a significant contributing factor in the development of tolerance. In 2000, Finney et al. [108] reported the upregulation of PDE4 in the lungs of rats treated chronically for 7 days with salbutamol. Thus, this phenomenon can be reproduced in vivo and may have clinical relevance in the development of tolerance following chronic use of $\beta_2$-adrenoceptor agonists.

### 7.2 Regulation of PDE4 by ERK

The catalytic subunit of all PDE4 isoenzymes, accept PDE4A, features an ERK consensus site that, when phosphorylated, results in either an increase or decrease in activity that is determined by the upstream conserved regions (UCRs). Thus, the ERK-mediated phosphorylation of long forms, short forms, and supershort forms is associated with an inhibition, activation, and weak inhibition of catalytic activity, respectively [109–111]. Thus, PDE4 activity can be either augmented or inhibited following activation of ERK that will be determined, to a large extent, by the complement of long *versus* short isoforms present in the cell of interest. With respect to airway inflammation such as seen in asthma and COPD, regulation of PDE4 activity by ERK-mediated phosphorylation may be highly relevant given that many pro-inflammatory cytokines and chemokines signal via activation of mitogen-activated protein kinase cascades.

## 8 ANTI-INFLAMMATORY EFFECTS OF PDE4 INHIBITORS IN VIVO

The actions of PDE inhibitors upon the acute (IgE-mediated) or chronic (pro-inflammatory/immunocompetent cell-mediated) consequences of allergen provocation in vivo have, in general, been documented in some detail and include studies of passive cutaneous anaphylaxis (PCA), cell infiltration into sites of inflammation, and microvascular leakage.

### 8.1 IgE-Mediated Processes

It seems unlikely that PDE inhibitors of any isoenzyme family reduce IgE levels in vivo. In humans, it is reported that ibudilast, which is an essentially nonselective PDE inhibitor [112], is without effect on IgE levels in asthmatic individuals [113]. Similarly, the elevated level of total serum IgE in a murine model of dermatitis caused by the repeated application of 2,4,6-trinitro-1-chlorobenzene is unaffected by oral administration of the PDE4 inhibitor rolipram [114].

In contrast, PDE4 inhibitors exhibit efficacy at reducing PCA reactions in rats, mice, and guinea pigs [115, 116]. Furthermore, rolipram is effective at reducing the infiltration of indium-labeled eosinophils in to the skin of guinea pigs following a PCA reaction [117]. Collectively, these data imply that PDE4 inhibitors can suppress the degranulation of IgE-bearing leukocytes and, therefore, allergen-induced mediator release. Further support for this proposal derives from studies in sensitized guinea pigs where rolipram, administered intravenously, inhibits antigen- but not leukotriene $D_4$ ($LTD_4$)-induced bronchoconstriction [118]. Thus, rolipram preferentially exerts an inhibitory influence at the level of mast cells and basophils rather than exerting a direct antispasmogenic action at the level of airway smooth muscle.

### 8.2 Pro-inflammatory Cell Infiltration

A large number of investigations (too many to describe individually herein) have evaluated the effect of PDE inhibitors upon the infiltration of pro-inflammatory cells into the airway lumen, skin, and eye of guinea pigs, rats, and, monkeys in responses to a variety of dissimilar stimuli including allergen. For example, pretreatment of sensitized guinea pigs with zardaverine, a mixed PDE3/4 inhibitor, markedly suppressed allergen-induced infiltration of eosinophils, macrophages, and neutrophils into the bronchoalveolar lavage (BAL) fluid to a level achieved with dexamethasone [119]. Qualitatively identical data have been reported for the PDE3/4 inhibitor, benafentrine, on platelet-activating factor (PAF)—and allergen-induced pulmonary eosinophil recruitment in guinea pigs after chronic (6 days) dosing [120, 121]. Studies conducted more recently with rolipram and second-generation PDE4 inhibitors including cilomilast, roflumilast, and AWD 12-281 have corroborated these findings both in the airways an in allergic cutaneous reactions [122–131].

Similarly, treatment of sensitized Cynomolgus monkeys with rolipram abrogates the pulmonary neutrophilia and eosinophilia, and airway hyperresponsiveness after multiple exposures to the antigen and is without effect on the immediate increase in airway resistance that follows acute antigen provocation [132]. Thus, these data are consistent with the findings that PDE4 inhibitors may be anti-inflammatory and act primarily to prevent the activation of immune cells in the lung rather than by exerting an antispasmogenic or spasmolytic effect the level of airway smooth muscle [118].

### 8.3 Microvascular Leakage and Edema

It is well recognized that microcirculations play an important role in inflammatory reactions. Under normal conditions, the endothelium lining the postcapillary microvenules in the skin and bronchial circulation (capillaries in the pulmonary vasculature) is largely impermeable to blood cells and to macromolecules; but following a pro-inflammatory insult, localized arteriolar vasodilatation occurs with a consequent increase in blood flow. This effect is induced by the liberation of pro-inflammatory mediators such as histamine and tachykinins. An increase in capillary and microvenular pressure then ensues, together with the liberation of other mediators such $LTD_4$, which contract directly the microvenular endothelial cells. Together, these effects, by increasing microvenular permeability, permit the loss of plasma proteins from the vascular compartment. Furthermore, the resulting increase in osmotic pressure due to loss of solute from the circulation leads to marked fluid exudation and to edema.

In guinea pigs, PDE4 inhibitors given by the intravenous, oral, or intratracheal route markedly attenuate PAF-induced microvascular leakage in both small and large airways and into the BAL fluid of anaesthetized guinea pigs [133–135].

The finding that rolipram is active when given directly into the airways indicates an important local action and highlights that systemic administration is not necessary for these compounds to exert an anti-inflammatory influence. This is an important observation since the administration of PDE4 inhibitors by the inhaled route could reduce untoward side effects while maintaining efficacy (see below).

Svensjo and colleagues [136] have reported that PDE4 and PDE3/4 inhibitors also attenuate the increase in microvascular permeability evoked by bradykinin in the hamster cheek pouch. Qualitatively identical results are obtained for the effect of rolipram and denbufylline upon arachidonate-induced edema in rodents [137].

## 9 CLINICAL PHARMACOLOGY OF PDE4 INHIBITORS

### 9.1 Roflumilast Clinical Development Program

Despite there having been many clinical trials of PDE4 inhibitors in asthma and COPD, there is a relative paucity of data in the public domain. Indeed, of those PDE4 inhibitors in phase II and phase III development for airway inflammatory diseases (see Table 2), information on only one of these—roflumilast (Daxas)—is available and these data are discussed here.

The results of two multicenter, double-blind, placebo-controlled phase III studies have been published following peer review. In one of those, which was conducted in 159 centers across Australia, Austria, Belgium, Canada, France, Germany, Hungary, Ireland, South Africa, Spain, and the United Kingdom, 1411 subjects were randomized to receive roflumilast orally (250 or 500 μg o.d. [once a day]) or placebo for 24 weeks [138]. The primary efficacy endpoint measures were change in postbronchodilator forced expiratory volume in 1 second ($FEV_1$) and health-related quality of life using the St. George's Respiratory Questionaire (SGRQ). Secondary measures included change from baseline in prebronchodilator $FEV_1$, FVC (forced vital capacity), $FEV_6$, $FEF_{25\text{-}75}$ (forced expiratory $flow_{25\text{-}75}$) and COPD exacerbations. Compared to placebo, roflumilast statistically improved lung function at both doses and at all time points over the 24 weeks of treatment at which measurements were made. At the end of the treatment

period postbronchodilator $FEV_1$ relative to placebo was increased by 74 and 97 mL for the 250 and 500-μg doses, respectively [138], and subjects had experienced fewer exacerbations in the active treatment groups (the mean exacerbation number per patient = 1.13, 1.03, and 0.75 for placebo, roflumilast 250 μg, and roflumilast 500 μg, respectively). In contrast, there was no significant difference in health-related quality of life as assessed by the SGRQ for any domain whether the effect of roflumilast was compared against baseline or placebo. With regard to secondary outcome measures, significant improvements in prebronchodilator $FEV_1$, postbronchodilator FVC, $FEV_6$, and $FEF_{25\text{-}75}$ were also reported [138].

Despite the apparently positive outcome of this roflumilast trial in COPD, certain claims made by Rabe and colleagues [138] in that publication have attracted criticisms, some of which merit highlighting here (see [139–143]). Notably, Bergmann [139] attests that if the *raw* data are analyzed (see [144]) there is, in fact, no significant difference at the end of the trial between the placebo and the roflumilast treatment groups. Thus, after 24 weeks of treatment $FEV_1$ relative to baseline was increased and decreased by 51 and 45 mL in those subjects that were given roflumilast (500 μg o.d.) and placebo, respectively [138]. If those data are corrected for the difference in mean baseline $FEV_1$ between the placebo and roflumilast treatment groups at the beginning of the trial (these were 1.57 and 1.50 L, respectively), then the absolute difference in $FEV_1$ at week 24 favors roflumilast by a mere 26 mL [i.e., (1.5 L + 0.05 L) – (1.57 L – 0.045 L) – (1.5 L + 0.05 L) = –0.25. This is neither statistically nor clinically meaningful. Bergmann [139] also challenges the statistical evaluation of exacerbation rate between the three treatment groups stating that a classical $\chi^2$ test returns a nonsignificant $P$ value of 0.48. Another valid criticism, leveled by Morice [140], is the assertion by Rabe and colleagues that the modest improvement in lung function was due to an anti-inflammatory effect of roflumilast. As indices of inflammation were not measured in this trial, it is remarkable that a statement to this effect is incorporated into the title of the paper. By their own admission [145], the authors of the study extrapolate this interpretation from experiments conducted with laboratory animals (which at best can be misleading) and from previously published data where cilomilast was found not to be a directly acting bronchodilator [146].

The results of the other phase III trial were published in July 2007 [147]. In this 159 multicenter, randomized, placebo-controlled, double-blind parallel group study, the effect of roflumilast (500 μg o.d.) was compared against placebo in 1513 subjects with severe COPD [Global Obstructive Lung Disease (GOLD) stages III and IV]. The efficacy endpoint measures were change in postbronchodilator $FEV_1$ and exacerbation rate per patient per year. The main secondary variable was health-related quality of life using the SGRQ. Prebronchodilator $FEV_1$ postbronchodilator $FEV_6$, FVC, and $FEF_{25\text{-}75}$ were also determined. Inclusion and exclusion criteria were typical of most studies of PDE4 inhibitors in COPD except that previously prescribed inhaled corticosteroid (ICS) use was permitted. Compared to placebo, roflumilast improved lung function at all time points over the 52 weeks of treatment at which measurements were made. At the end of the treatment period postbronchodilator $FEV_1$ relative to placebo was increased by 39 mL; similar modest improvements in the other indices of lung function relative to placebo were also recorded, demonstrating that these effects can be maintained over a period of 1 year. However, there was no associated difference in the exacerbation rate between the two treatment groups (0.92 and 0.86 exacerbations/subject/year for placebo and roflumilast, respectively) or in health status. Calverley et al., [147] also performed a retrospective analysis of exacerbation frequency in those subjects characterized with GOLD stage IV disease. In that subpopulation, the exacerbation rate was 36% lower in subjects treated with roflumilast when compared to those given placebo (1.01 and 1.59 exacerbations/patient/year, respectively).

### 9.2 Assessment of the Roflumilast Clinical Development Program

The improvement in lung function effected by roflumilast in both of the aforementioned phase III clinical trials was modest and of similar magnitude to that evoked by ICSs. While these data are probably of limited clinical significance, the improvement in lung function reported by Calverley et al., [147] was evoked in subjects in whom the use of ICSs was permitted. Thus, perhaps roflumilast (and other PDE4 inhibitors) can improve lung function beyond that achievable by ICSs alone, despite the limited responsiveness of these subjects to bronchodilators [147].

It is worth noting that a meta-analysis of 13 clinical studies involving 244 subjects found that oral theophylline, a nonselective PDE inhibitor, improves $FEV_1$ in subjects with stable COPD by approximately 100 mL [148], which is superior to that produced by roflumilast in the phase III efficacy trials described here. These data may suggest that PDE4 inhibitors could further improve lung function if given at higher doses. Regrettably, most of the second-generation PDE4 inhibitors evaluated in clinical trials thus far conducted have been given at the highest tolerated dose, so this possibility cannot be

tested empirically. Whether the therapeutic ratio of this class of compound can be improved is currently unclear. However, clinical trials are ongoing, with a variety of new PDE4 inhibitors including oglemilast and tetomilast (Table 2), and the result of those studies could be instructive.

### 9.3 Safety and Tolerability of PDE4 Inhibitors

Despite some encouraging data from the pivotal phase II and III studies of COPD, all PDE4 inhibitors thus far examined including cilomilast (now discontinued—see http://www.reuters.com/article/sphereNews/idUSL152723420071015?sp=true&view=sphere) and roflumilast are hampered by a low therapeutic ratio. This limitation became clear early on in the development of these compounds with nausea, diarrhea, abdominal pain, vomiting, and dyspepsia being the most common adverse events reported (reviewed in [149]) [150]. Indeed, the number of subjects failing to complete all controlled cilomilast trials conducted by GlaxoSmithKline (GSK) due to an adverse event was positively dose related across the 2.5-, 5-, 10-, 15-mg treatment groups (6.9, 12.8, 20.4, and 25.9% withdrawal, respectively) with gastrointestinal disturbances being the most prevalent [151]. Similarly, the most common adverse events related to roflumilast treatment are diarrhea, nausea, and headache [138, 147]. Unfortunately, the unwanted gastrointestinal actions, which are mediated both locally (i.e., in the gastrointestinal tract) and centrally, can be accounted for by the ubiquitous distribution of PDE4 isoforms across many tissues and represent an extension of the pharmacology of PDE4 inhibitors that is typical of first-generation compounds such as rolipram.

## 10 OPTIMIZING THE EFFICACY OF PDE4 INHIBITORS

### 10.1 Oral Dosing

In the short term a number of options are available that could result in PDE4 inhibitors with enhanced efficacy and an improved therapeutic ratio. Perhaps the most appealing is to synthesize compounds that are delivered topically (i.e., by inhalation) directly to the site of action. These compounds would not be bioavailable, and, thus, systemic consequences including emetic liability and a propensity to cause cardiovascular lesions should be minimized. This approach has been used in the past with the development of piclamilast (Rhone-Poulence Rorer—now Sanofi Aventis), AWD 12-241 (Elbion/GlaxoSmithKline), GSK 256066 (GlaxoSmithKline), and tofimilast (Pfizer) but, as yet, no PDE4 inhibitor of this type has reached the market.

### 10.2 Combination Therapies

Other possibilities of harnessing the therapeutic potential of PDE4 inhibitors are through the development of combination therapies. Indeed, clinical trials are underway with PDE4 inhibitor/muscarinic $M_3$-receptor antagonist combinations (see http://www.clinicaltrials.gov/ct/search;jsessionid=628D4D1FDE10D07C7D6651C13193C47E?term=roflumilast) and UCB Pharma have recently published on UCB 101333-3, a 4,6-diaminopyridine derivative, which is a potent dual $M_3$-receptor antagonist *and* PDE4 inhibitor [152]. Such an approach could be more effective than either therapy alone as these drugs should improve lung function by interacting with complementary but mechanistically distinct processes. Combining an inhaled PDE4 inhibitor with an ICS could also afford greater clinical benefit than either drug given as a monotherapy in the same way that long-acting $\beta_2$-adrenocepotor agonist (LABA)/ICS combinations deliver superior efficacy [153]. The prospect of triple combination therapies incorporating a (ultra)LABA, ICS, and inhaled PDE4 inhibitor is also an appealing pharmacotherapeutic strategy.

### 10.3 Nonselective PDE Inhibitors

Theophylline has been shown to improve $FEV_1$ in subjects with COPD by at least 100 mL [148]. Based on this improvement in lung function, it is tempting to speculate that nonselective PDE inhibitors not based on the xanthine nucleus, which evokes many of the serious side effects reported with theophylline use, may prove to be the best therapeutic strategy for enhancing efficacy. Indeed, concurrent inhibition of multiple PDEs could exert clinically relevant effects, not produced by selective PDE4 inhibition, on processes that contribute to the pathogenesis of chronic airflow limitation, including airway remodeling (PDE1-regulated), endothelial cell permeability (PDE2-regulated), mast cell stabilization (PDE3-mediated), airway smooth muscle tone (PDE3-regulated), and remodeling of the pulmonary vasculature (PDE5-regulated) [154].

### 10.4 Subtype-Selective Inhibitors

Studies in mice in which three of the four PDE4 genes have been deleted indicate that they encode proteins that generally regulate specific, nonredundant functions [155]. Accordingly, it is possible that selective inhibition, with small-molecule inhibitors, of proteins derived from the transcription of a specific gene may allow the benefi-

cial effects of nonselective compounds to be retained while, at the same time, reducing or eliminating adverse effects. Indeed, using a surrogate measure, mice deficient in PDE4D display enhanced emetic behavior, whereas experimentally induced inflammation is relatively unaffected. In contrast, neither deletion of *PDE4A* nor *PDE4B* in mice is emetic, yet airway inflammation in vivo and cytokine generation from leukocytes ex vivo is markedly suppressed [156–159].

## 11 CONCLUDING REMARKS

The decision by the pharmaceutical industry to develop second-generation PDE4 inhibitors for the treatment of allergic and nonallergic inflammatory diseases of the airways is based on a conceptually robust hypothesis that is supported by a wealth of preclinical data. Therefore, if shown to be potentially disease modifying in clinical trials, PDE4 inhibitors will offer physicians a novel class of drug to treat patients with inflammation that may be refractory to other drugs such as glucocorticoids. That said, it is the view of this author that further major refinements in the design and/or use of PDE4 inhibitors are still necessary to improve their therapeutic ratio if major clinical benefit is to be achieved. Several options are available, and some of these are currently being evaluated in clinical trials. Thus, it is likely that several more years will elapse before we can assess whether the initial optimism in PDE4 inhibitors will be realized.

## ACKNOWLEDGMENTS

The author is an Alberta Heritage Foundation of Medical Research Senior Scholar and is funded by the Canadian Institutes of Health Research.

## REFERENCES

1. Gendo K, Lodewick MJ (2005). Asthma economics: Focusing on therapies that improve costly outcomes. *Curr Opin Pulm Med 11*, 43–50.
2. Krishnan V, Diette GB, Rand CS, Bilderback AL, Merriman B, Hansel NN, Krishnan JA (2006). Mortality in patients hospitalized for asthma exacerbations in the United States. *Am J Respir Crit Care Med 174*, 633–638.
3. James A (2005). Airway remodeling in asthma. *Curr Opin Pulm Med 11*, 1–6.
4. Johnson PR, Roth M, Tamm M, Hughes M, Ge Q, King G, Burgess JK, Black JL (2001). Airway smooth muscle cell proliferation is increased in asthma. *Am J Respir Crit Care Med 164*, 474–477.
5. Humbles AA, Lloyd CM, McMillan SJ, Friend DS, Xanthou G, McKenna EE, Ghiran S, Gerard NP, Yu C, Orkin SH, Gerard C (2004). A critical role for eosinophils in allergic airways remodeling. *Science 305*, 1776–1779.
6. Cho JY, Miller M, Baek KJ, Han JW, Nayar J, Lee SY, McElwain K, McElwain S, Friedman S, Broide DH (2004). Inhibition of airway remodeling in IL-5-deficient mice. *J Clin Invest 113*, 551–560.
7. Williams TJ (2004). The eosinophil enigma. *J Clin Invest 113*, 507–509.
8. Barnes PJ (2000). Chronic obstructive pulmonary disease. *N Engl J Med 343*, 269–280.
9. Hogg JC (2004). Pathophysiology of airflow limitation in chronic obstructive pulmonary disease. *Lancet 364*, 709–721.
10. Hogg JC, Chu F, Utokaparch S, Woods R, Elliott WM, Buzatu L, Cherniack RM, Rogers RM, Sciurba FC, Coxson HO, Pare PD (2004). The nature of small-airway obstruction in chronic obstructive pulmonary disease. *N Engl J Med 350*, 2645–2653.
11. Briggs DD (2004). Chronic obstructive pulmonary disease overview: Prevalence, pathogenesis, and treatment. *J Manag Care Pharm 10*, S3–10.
12. Barnes PJ, Shapiro SD, Pauwels RA (2003). Chronic obstructive pulmonary disease: molecular and cellular mechanisms. *Eur Respir J 22*, 672–688.
13. Rabe KF, Hurd S, Anzueto A, Barnes PJ, Buist SA, Calverley P, Fukuchi Y, Jenkins C, Rodriguez-Roisin R, van Weel C, Zielinski J (2007). Global strategy for the diagnosis, management, and prevention of COPD—2006 update. *Am J Respir Crit Care Med 126*, 527–555.
14. Halbert RJ, Natoli JL, Gano A, Badamgarav E, Buist AS, Mannino DM (2006). Global burden of COPD: Systematic review and meta-analysis. *Eur Respir J 28*, 523–532.
15. Jemal A, Ward E, Hao Y, Thun M (2005). Trends in the leading causes of death in the United States, 1970–2002. *JAMA 294*, 1255–1259.
16. Lopez AD, Shibuya K, Rao C, Mathers CD, Hansell AL, Held LS, Schmid V, Buist S (2006). Chronic obstructive pulmonary disease: Current burden and future projections. *Eur Respir J 27*, 397–412.
17. Murray CJ, Lopez AD (1997). Mortality by cause for eight regions of the world: Global Burden of Disease Study. *Lancet 349*, 1269–1276.
18. Shahab L, Jarvis MJ, Britton J, West R (2006). Prevalence, diagnosis and relation to tobacco dependence of chronic obstructive pulmonary disease in a nationally representative population sample. *Thorax 61*, 1043–1047.
19. Barnes PJ (2006). Corticosteroids: The drugs to beat. *Eur J Pharmacol 533*, 2–14.
20. National Heart, Lung and Blood Institute (NHLBI) (1997). National asthma education and prevenation program. *Expert panel report 2: Guidelines for the diag-*

*nosis and management of asthma*. Washington, DC: NHLBI.

21. National Heart, Lung and Blood Institute (NHLBI) (2002). Global initiative for asthma. *Global strategy for asthma management and prevention*. Washington, DC: NHLBI.
22. British Thoracic Society/Scottish Intercollegiate Guidelines Network (BTS/SIGN) (2003). British guideline on the management of asthma. *Thorax 58*, (Suppl 1), i1–94.
23. Lemiere C, Bai T, Balter M, Bayliff C, Becker A, Boulet LP, Bowie D, Cartier A, Cave A, Chapman K, Cowie R, Coyle S, Cockcroft D, Ducharme FM, Ernst P, Finlayson S, FitzGerald JM, Hargreave FE, Hogg D, Kaplan A, Kim H, Kelm C, O'Byrne P, Sears M, Markham AW (2004). Adult Asthma Consensus Guidelines Update 2003. *Can Respir J 11*, 9A–18A.
24. Sin DD, Man SF (2007). Do chronic inhaled steroids alone or in combination with a bronchodilator prolong life in chronic obstructive pulmonary disease patients? *Curr Opin Pulm Med 13*, 90–97.
25. Fan VS, Bryson CL, Curtis JR, Fihn SD, Bridevaux PO, McDonell MB, Au DH (2003). Inhaled corticosteroids in chronic obstructive pulmonary disease and risk of death and hospitalization: Time-dependent analysis. *Am J Respir Crit Care Med 168*, 1488–1494.
26. Kiri VA, Pride NB, Soriano JB, Vestbo J (2005). Inhaled corticosteroids in chronic obstructive pulmonary disease: Results from two observational designs free of immortal time bias. *Am J Respir Crit Care Med 172*, 460–464.
27. Sin DD, Man SF (2003). Inhaled corticosteroids and survival in chronic obstructive pulmonary disease: Does the dose matter? *Eur Respir J 21*, 260–266.
28. Sin DD, Tu JV (2001). Inhaled corticosteroids and the risk of mortality and readmission in elderly patients with chronic obstructive pulmonary disease. *Am J Respir Crit Care Med 164*, 580–584.
29. Soriano JB, Vestbo J, Pride NB, Kiri V, Maden C, Maier WC (2002). Survival in COPD patients after regular use of fluticasone propionate and salmeterol in general practice. *Eur Respir J 20*, 819–825.
30. Suissa S (2003). Effectiveness of inhaled corticosteroids in chronic obstructive pulmonary disease: Immortal time bias in observational studies. *Am J Respir Crit Care Med 168*, 49–53.
31. Tkacova R, Toth S, Sin DD (2006). Inhaled corticosteroids and survival in COPD patients receiving long-term home oxygen therapy. *Respir Med 100*, 385–392.
32. Macie C, Wooldrage K, Manfreda J, Anthonisen NR (2006). Inhaled corticosteroids and mortality in COPD. *Chest 130*, 640–646.
33. Sin DD, Wu L, Anderson JA, Anthonisen NR, Buist AS, Burge PS, Calverley PM, Connett JE, Lindmark B, Pauwels RA, Postma DS, Soriano JB, Szafranski W, Vestbo J (2005). Inhaled corticosteroids and mortality in chronic obstructive pulmonary disease. *Thorax 60*, 992–997.
34. Calverley PM, Anderson JA, Celli B, Ferguson GT, Jenkins C, Jones PW, Yates JC, Vestbo J (2007). Salmeterol and fluticasone propionate and survival in chronic obstructive pulmonary disease. *N Engl J Med 356*, 775–789.
35. Panizza JA, James AL, Ryan G, de Klerk N, Finucane KE (2006). Mortality and airflow obstruction in asthma: A 17-year follow-up study. *Intern Med J 36*, 773–780.
36. Dahl R (2006). Systemic side effects of inhaled corticosteroids in patients with asthma. *Respir Med 100*, 1307–1317.
37. Toogood JH (1998). Side effects of inhaled corticosteroids. *J Aller Clin Immunol 102*, 705–713.
38. Bender AT, Beavo JA (2006). Cyclic nucleotide phosphodiesterases: Molecular regulation to clinical use. *Pharmacol Rev 58*, 488–520.
39. Lugnier C (2006). Cyclic nucleotide phosphodiesterase (PDE) superfamily: A new target for the development of specific therapeutic agents. *Pharmacol Ther 109*, 366–398.
40. Houslay MD (2005). The long and short of vascular smooth muscle phosphodiesterase-4 as a putative therapeutic target. *Mol Pharmacol 68*, 563–567.
41. Houslay MD, Schafer P, Zhang KY (2005). Phosphodiesterase-4 as a therapeutic target. *Drug Discov Today 10*, 1503–1519.
42. Torphy TJ (1998). Phosphodiesterase isozymes: Molecular targets for novel antiasthma agents. *Am J Respir Crit Care Med 157*, 351–370.
43. Giembycz MA (2005). Phosphodiesterase 4: Selective and dual-specificity inhibitors for the therapy of chronic obstructive pulmonary disease. *Proc Am Thoracia Soc 2*, 326–333.
44. Giembycz MA (2005). Life after PDE4: Overcoming adverse events with dual-specificity phosphodiesterase inhibitors. *Curr Opin Pharmacol 5*, 238–244.
45. Beavo JA, Conti M, Heaslip RJ (1994). Multiple cyclic nucleotide phosphodiesterases. *Mol Pharmacol 46*, 399–405.
46. Schwabe U, Miyake M, Ohga Y, Daly JW (1976). 4-(3-Cyclopentyloxy-4-methoxyphenyl)-2-pyrrolidone (ZK 62711): A potent inhibitor of adenosine cyclic 3′,5′-monophosphate phosphodiesterases in homogenates and tissue slices from rat brain. *Mol Pharmacol 12*, 900–910.
47. Sheppard H, Wiggan G (1971). Analogues of 4-(3,4-dimethoxybenzyl)-2-imidazolidinone as potent inhibitors of rat erythrocyte adenosine cyclic 3′,5′-phosphate phosphodiesterase. *Mol Pharmacol 7*, 111–115.
48. Bobon D, Breulet M, Gerard-Vandenhove MA, Guiot-Goffioul F, Plomteux G, Hernandez M, Schratzer M, Troisfontaines B, von Frenckell R, Wachtel H (1988). Is phosphodiesterase inhibition a new mechanism of anti-

depressant action? A double blind double-dummy study between rolipram and desipramine in hospitalized major and/or endogenous depressives. *Eur Arch Psychiatry Neurol Sci 238*, 2–6.

49. Zeller E, Stief HJ, Pflug B, Hernandez M (1984). Results of a phase II study of the antidepressant effect of rolipram. *Pharmacopsychiatry 17*, 188–190.
50. Horowski R, Sastre Y, Hernandez M (1985). Clinical effects of the neurotropic selective cAMP phosphodiesterase inhibitor rolipram in depressed patients: Global evaluation of the preliminary reports. *Curr Ther Res 38*, 23–29.
51. Scott AI, Perini AF, Shering PA, Whalley LJ (1991). Inpatient major depression: is rolipram as effective as amitriptyline? *Eur J Clin Pharmacol 40*, 127–129.
52. Parkes JD, Thompson C, Brennan L, Gajraj N, Howcroft B, Ruiz J (1984). Rolipram in Parkinson's disease. *Adv Neurol 40*, 563–565.
53. Casacchia M, Meco G, Castellana F, Bedini L, Cusimano G, Agnoli A (1983). Therapeutic use of a selective cAMP phosphodiesterase inhibitor (Rolipram) in Parkinson's disease. *Pharmacol Res Commun 15*, 329–334.
54. McKenna JM, Muller GW (2007). Medicinal chemistry of PDE4 Inhibitors. In Beavo JA, Francis SH, Houslay MD, Eds. *Cyclic Nucleotide Phosphodiesterases in Health and Disease*. Boca Raton, FL: CRC Press, pp. 667–699.
55. Thompson WJ, Epstein PM, Strada SJ (1979). Purification and characterization of high-affinity cyclic adenosine monophosphate phosphodiesterase from dog kidney. *Biochemistry 18*, 5228–5237.
56. Epstein PM, Strada SJ, Sarada K, Thompson WJ (1982). Catalytic and kinetic properties of purified high-affinity cyclic AMP phosphodiesterase from dog kidney. *Arch Biochem Biophys 218*, 119–133.
57. Torphy TJ, DeWolf WE Jr, Green DW, Livi GP (1993). Biochemical characteristics and cellular regulation of phosphodiesterase IV. *Agents Actions Suppl 43*, 51–71.
58. Truong VH, Muller T (1994). Isolation, biochemical characterization and N-terminal sequence of rolipram-sensitive cAMP phosphodiesterase from human mononuclear leukocytes. *FEBS Lett 353*, 113–118.
59. Salanova M, Jin SC, Conti M (1998). Heterologous expression and purification of recombinant rolipram-sensitive cyclic AMP-specific phosphodiesterases. *Methods 14*, 55–64.
60. Colicelli J, Birchmeier C, Michaeli T, O'Neill K, Riggs M, Wigler M (1989). Isolation and characterization of a mammalian gene encoding a high-affinity cAMP phosphodiesterase. *Proc Natl Acad Sci USA 86*, 3599–3603.
61. Swinnen JV, Joseph DR, Conti M (1989). Molecular cloning of rat homologues of the *Drosophila melanogaster* dunce cAMP phosphodiesterase: Evidence for a family of genes. *Proc Natl Acad Sci USA 86*, 5325–5329.
62. Swinnen JV, Joseph DR, Conti M (1989). The mRNA encoding a high-affinity cAMP phosphodiesterase is regulated by hormones and cAMP. *Proc Natl Acad Sci USA 86*, 8197–8201.
63. Houslay MD, Adams DR (2003). PDE4 cAMP phosphodiesterases: Modular enzymes that orchestrate signalling cross-talk, desensitization and compartmentalization. *Biochem J 370*, 1–18.
64. Houslay MD, Baillie GS, Maurice DH (2007). cAMP-Specific phosphodiesterase-4 enzymes in the cardiovascular system: A molecular toolbox for generating compartmentalized cAMP signaling. *Circ Res 100*, 950–966.
65. Beard MB, Olsen AE, Jones RE, Erdogan S, Houslay MD, Bolger GB (2000). UCR1 and UCR2 domains unique to the cAMP-specific phosphodiesterase family form a discrete module via electrostatic interactions. *J Biol Chem 275*, 10349–10358.
66. Richter W, Conti M (2002). Dimerization of the type 4 cAMP-specific phosphodiesterases is mediated by the upstream conserved regions (UCRs). *J Biol Chem 277*, 40212–40221.
67. Houslay MD, Sullivan M, Bolger GB (1998). The multienzyme PDE4 cyclic adenosine monophosphate-specific phosphodiesterase family: Intracellular targeting, regulation, and selective inhibition by compounds exerting anti-inflammatory and antidepressant actions. *Adv Pharmacol 44*, 225–342.
68. Giembycz MA (1996). Phosphodiesterase 4 and tolerance to $\beta_2$-adrenoceptor agonists in asthma. *Trends Pharmacol Sci 17*, 331–336.
69. O'Connor BJ, Aikman SL, Barnes PJ (1992). Tolerance to the nonbronchodilator effects of inhaled $\beta_2$-agonists in asthma. *N Engl J Med 327*, 1204–1208.
70. Cockcroft DW, McParland CP, Britto SA, Swystun VA, Rutherford BC (1993). Regular inhaled salbutamol and airway responsiveness to allergen. *Lancet 342*, 833–837.
71. Haney S, Hancox RJ (2005). Rapid onset of tolerance to beta-agonist bronchodilation. *Respir Med 99*, 566–571.
72. Salpeter SR, Ormiston TM, Salpeter EE (2004). Meta-analysis: Respiratory tolerance to regular $\beta_2$-agonist use in patients with asthma. *Ann Intern Med 140*, 802–813.
73. Wraight JM, Hancox RJ, Herbison GP, Cowan JO, Flannery EM, Taylor DR (2003). Bronchodilator tolerance: The impact of increasing bronchoconstriction. *Eur Respir J 21*, 810–815.
74. Hancox RJ, Aldridge RE, Cowan JO, Flannery EM, Herbison GP, McLachlan CR, Town GI, Taylor DR (1999). Tolerance to beta-agonists during acute bronchoconstriction. *Eur Respir J 14*, 283–287.
75. Hancox RJ, Subbarao P, Kamada D, Watson RM, Hargreave FE, Inman MD (2002). $\beta_2$-Agonist tolerance and exercise-induced bronchospasm. *Am J Respir Crit Care Med 165*, 1068–1070.
76. Bhagat R, Kalra S, Swystun VA, Cockcroft DW (1995). Rapid onset of tolerance to the bronchoprotective effect of salmeterol. *Chest 108*, 1235–1239.

77. Cheung D, Timmers MC, Zwinderman AH, Bel EH, Dijkman JH, Sterk PJ (1992). Long-term effects of a long-acting $\beta_2$-adrenoceptor agonist, salmeterol, on airway hyperresponsiveness in patients with mild asthma. *N Engl J Med 327*, 1198–1203.
78. Jones SL, Cowan JO, Flannery EM, Hancox RJ, Herbison GP, Taylor DR (2001). Reversing acute bronchoconstriction in asthma: The effect of bronchodilator tolerance after treatment with formoterol. *Eur Respir J 17*, 368–373.
79. Sette C, Vicini E, Conti M (1994). Modulation of cellular responses by hormones: Role of cAMP specific, rolipram-sensitive phosphodiesterases. *Mol Cell Endocrinol 100*, 75–79.
80. Alvarez R, Sette C, Yang D, Eglen RM, Wilhelm R, Shelton ER, Conti M (1995). Activation and selective inhibition of a cyclic AMP-specific phosphodiesterase, PDE-4D3. *Mol Pharmacol 48*, 616–622.
81. Sette C, Vicini E, Conti M (1994). The ratPDE3/IVd phosphodiesterase gene codes for multiple proteins differentially activated by cAMP-dependent protein kinase. *J Biol Chem 269*, 18271–18274.
82. Sette C, Conti M (1996). Phosphorylation and activation of a cAMP-specific phosphodiesterase by the cAMP-dependent protein kinase. Involvement of serine 54 in the enzyme activation. *J Biol Chem 271*, 16526–16534.
83. Sette C, Iona S, Conti M (1994). The short-term activation of a rolipram-sensitive, cAMP-specific phosphodiesterase by thyroid-stimulating hormone in thyroid FRTL-5 cells is mediated by a cAMP-dependent phosphorylation. *J Biol Chem 269*, 9245–9252.
84. Hoffmann R, Wilkinson IR, McCallum JF, Engels P, Houslay MD (1998). cAMP-specific phosphodiesterase HSPDE4D3 mutants which mimic activation and changes in rolipram inhibition triggered by protein kinase A phosphorylation of Ser-54: Generation of a molecular model. *Biochem J 333*, 139–149.
85. MacKenzie SJ, Baillie GS, McPhee I, MacKenzie C, Seamons R, McSorley T, Millen J, Beard MB, Van Heeke G, Houslay MD (2002). Long PDE4 cAMP specific phosphodiesterases are activated by protein kinase A-mediated phosphorylation of a single serine residue in Upstream Conserved Region 1 (UCR1). *Br J Pharmacol 136*, 421–433.
86. Richter W, Jin SL, Conti M (2005). Splice variants of the cyclic nucleotide phosphodiesterase PDE4D are differentially expressed and regulated in rat tissue. *Biochem J 388*, 803–811.
87. Wang D, Deng C, Bugaj-Gaweda B, Kwan M, Gunwaldsen C, Leonard C, Xin X, Hu Y, Unterbeck A, De Vivo M (2003). Cloning and characterization of novel PDE4D isoforms PDE4D6 and PDE4D7. *Cell Signal 15*, 883–891.
88. D'Sa C, Tolbert LM, Conti M, Duman RS (2002). Regulation of cAMP-specific phosphodiesterases type 4B and 4D (PDE4) splice variants by cAMP signaling in primary cortical neurons. *J Neurochem 81*, 745–757.
89. Vicini E, Conti M (1997). Characterization of an intronic promoter of a cyclic adenosine 3′,5′-monophosphate (cAMP)-specific phosphodiesterase gene that confers hormone and cAMP inducibility. *Mol Endocrinol 11*, 839–850.
90. Barber R, Clark RB, Kelly LA, Butcher RW (1978). A model of desensitization in intact cells. *Adv Cyclic Nucleotide Res 9*, 507–516.
91. Seybold J, Newton R, Wright L, Finney PA, Suttorp N, Barnes PJ, Adcock IM, Giembycz MA (1998). Induction of phosphodiesterases 3B, 4A4, 4D1, 4D2, and 4D3 in Jurkat T-cells and in human peripheral blood T-lymphocytes by 8-bromo-cAMP and Gs-coupled receptor agonists. Potential role in $\beta_2$-adrenoreceptor desensitization. *J Biol Chem 273*, 20575–20588.
92. Ashby B (1989). Model of prostaglandin-regulated cyclic AMP metabolism in intact platelets: Examination of time-dependent effects on adenylate cyclase and phosphodiesterase activities. *Mol Pharmacol 36*, 866–873.
93. Bousquet-Melou A, Galitzky J, Moreno CM, Berlan M, Lafontan M (1995). Desensitization of β-adrenergic responses in adipocytes involves receptor subtypes and cAMP phosphodiesterase. *Eur J Pharmacol 289*, 235–247.
94. Conti M, Monaco L, Geremia R, Stefanini M (1986). Effect of phosphodiesterase inhibitors on Sertoli cell refractoriness: Reversal of the impaired androgen aromatization. *Endocrinology 118*, 901–908.
95. Ortiz JL, Dasi FJ, Cortijo J, Morcillo EJ (2000). β-Adrenoceptor stimulation up-regulates phosphodiesterase 4 activity and reduces prostaglandin $E_2$-inhibitory effects in human neutrophils. *Naunyn Schmiedeberg's Arch Pharmacol 361*, 410–417.
96. Dasi FJ, Ortiz JL, Cortijo J, Morcillo EJ (2000). Histamine up-regulates phosphodiesterase 4 activity and reduces prostaglandin $E_2$-inhibitory effects in human neutrophils. *Inflamm Res 49*, 600–609.
97. Mehats C, Tanguy G, Dallot E, Cabrol D, Ferre F, Leroy MJ (2001). Is up-regulation of phosphodiesterase 4 activity by $PGE_2$ involved in the desensitization of β-mimetics in late pregnancy human myometrium? *J Clin Endocrinol Metab 86*, 5358–5365.
98. Mehats C, Tanguy G, Dallot E, Robert B, Rebourcet R, Ferre F, Leroy MJ (1999). Selective up-regulation of phosphodiesterase-4 cyclic adenosine 3′,5′-monophosphate (cAMP)-specific phosphodiesterase variants by elevated cAMP content in human myometrial cells in culture. *Endocrinology 140*, 3228–3237.
99. Persani L, Lania A, Alberti L, Romoli R, Mantovani G, Filetti S, Spada A, Conti M (2000). Induction of specific phosphodiesterase isoforms by constitutive activation of the cAMP pathway in autonomous thyroid adenomas. *J Clin Endocrinol Metab 85*, 2872–2878.
100. Torphy TJ, Zhou HL, Cieslinski LB (1992). Stimulation of β-adrenoceptors in a human monocyte cell line (U937)

up-regulates cyclic AMP-specific phosphodiesterase activity. *J Pharmacol Exp Ther 263*, 1195–1205.

101. Torphy TJ, Zhou HL, Foley JJ, Sarau HM, Manning CD, Barnette MS (1995). Salbutamol up-regulates PDE4 activity and induces a heterologous desensitization of U937 cells to prostaglandin $E_2$. Implications for the therapeutic use of β-adrenoceptor agonists. *J Biol Chem 270*, 23598–23604.
102. Verghese MW, McConnell RT, Lenhard JM, Hamacher L, Jin SL (1995). Regulation of distinct cyclic AMP-specific phosphodiesterase (phosphodiesterase type 4) isozymes in human monocytic cells. *Mol Pharmacol 47*, 1164–1171.
103. Engels P, Fichtel K, Lubbert H (1994). Expression and regulation of human and rat phosphodiesterase type IV isogenes. *FEBS Lett 305*, 291–295.
104. Kochetkova M, Burns FM, Souness JE (1995). Isoprenaline induction of cAMP-phosphodiesterase in guinea-pig macrophages occurs in the presence, but not in the absence, of the phosphodiesterase type IV inhibitor rolipram. *Biochem Pharmacol 50*, 2033–2038.
105. Le Jeune IR, Shepherd M, Van Heeke G, Houslay MD, Hall IP (2002). Cyclic AMP-dependent transcriptional up-regulation of phosphodiesterase 4D5 in human airway smooth muscle cells. Identification and characterization of a novel PDE4D5 promoter. *J Biol Chem 277*, 35980–35989.
106. Bolger GB, McCahill A, Huston E, Cheung YF, McSorley T, Baillie GS, Houslay MD (2003). The unique amino-terminal region of the PDE4D5 cAMP phosphodiesterase isoform confers preferential interaction with beta-arrestins. *J Biol Chem 278*, 49230–49238.
107. Perry SJ, Baillie GS, Kohout TA, McPhee I, Magiera MM, Ang KL, Miller WE, McLean AJ, Conti M, Houslay MD, Lefkowitz RJ (2002). Targeting of cyclic AMP degradation to $\beta_2$-adrenergic receptors by β-arrestins. *Science 298*, 834–836.
108. Finney PA, Belvisi MG, Donnelly LE, Chuang TT, Mak JC, Scorer C, Barnes PJ, Adcock M, Giembycz MA (2000). Albuterol-induced downregulation of Gsα accounts for pulmonary $\beta_2$-adrenoceptor desensitization *in vivo*. *J Clin Invest 106*, 125–135.
109. Baillie GS, MacKenzie SJ, McPhee I, Houslay MD (2000). Sub-family selective actions in the ability of Erk2 MAP kinase to phosphorylate and regulate the activity of PDE4 cyclic AMP-specific phosphodiesterases. *Br J Pharmacol 131*, 811–819.
110. Hoffmann R, Baillie GS, MacKenzie SJ, Yarwood SJ, Houslay MD (1999). The MAP kinase ERK2 inhibits the cyclic AMP-specific phosphodiesterase HSPDE4D3 by phosphorylating it at $Ser^{579}$. *EMBO J 18*, 893–903.
111. MacKenzie SJ, Baillie GS, McPhee I, Bolger GB, Houslay MD (2000). ERK2 mitogen-activated protein kinase binding, phosphorylation, and regulation of the PDE4D cAMP-specific phosphodiesterases. The involvement of COOH-terminal docking sites and NH2-terminal UCR regions. *J Biol Chem 275*, 16609–16617.
112. Souness JE, Villamil ME, Scott LC, Tomkinson A, Giembycz MA, Raeburn D (1994). Possible role of cyclic AMP phosphodiesterases in the actions of ibudilast on eosinophil thromboxane generation and airways smooth muscle tone. *Br J Pharmacol 111*, 1081–1088.
113. Kawasaki A, Hoshino K, Osaki R, Mizushima Y, Yano S (1992). Effect of ibudilast: A novel antiasthmatic agent, on airway hypersensitivity in bronchial asthma. *J Asthma 29*, 245–252.
114. Harada D, Tsukumo Y, Takashima Y, Manabe H (2006). Effect of orally administered rolipram, a phosphodiesterase 4 inhibitor, on a mouse model of the dermatitis caused by 2,4,6-trinitro-1-chlorobenzene (TNCB)-repeated application. *Eur J Pharmacol 532*, 128–137.
115. Davies GE, Evans DP (1973). Studies with two new phosphodiesterase inhibitors (ICI 58, 301 and ICI 63, 197) on anaphylaxis in guinea pigs, mice and rats. *Int Arch Allergy Appl Immunol 45*, 467–478.
116. Broughton BJ, Chaplen P, Knowles P, Lunt E, Marshall SM, Pain DL, Wooldridge KR (1975). Antiallergic activity of 2-phenyl-8-azapurin-6-ones. *J Med Chem 18*, 1117–1122.
117. Teixeira MM, Rossi AG, Williams TJ, Hellewell PG (1994). Effects of phosphodiesterase isoenzyme inhibitors on cutaneous inflammation in the guinea-pig. *Br J Pharmacol 112*, 332–340.
118. Howell RE, Sickels BD, Woeppel SL (1993). Pulmonary antiallergic and bronchodilator effects of isozyme-selective phosphodiesterase inhibitors in guinea pigs. *J Pharmacol Exp Ther 264*, 609–615.
119. Schudt C, Winder S, Eltze M, Kilian U, Beume R (1991). Zardaverine: A cyclic AMP specific PDE III/IV inhibitor. *Agents Actions Suppl 34*, 379–402.
120. Sanjar S, Aoki S, Boubekeur K, Chapman ID, Smith D, Kings MA, Morley J (1990). Eosinophil accumulation in pulmonary airways of guinea-pigs induced by exposure to an aerosol of platelet-activating factor: Effect of anti-asthma drugs. *Br J Pharmacol 99*, 267–272.
121. Sanjar S, Aoki S, Kristersson A, Smith D, Morley J (1990). Antigen challenge induces pulmonary airway eosinophil accumulation and airway hyperreactivity in sensitized guinea-pigs: The effect of anti-asthma drugs. *Br J Pharmacol 99*, 679–686.
122. Griswold DE, Webb EF, Badger AM, Gorycki PD, Levandoski PA, Barnette MA, Grous M, Christensen S, Torphy TJ (1998). SB 207499 (*Ariflo*), a second generation phosphodiesterase 4 inhibitor, reduces tumor necrosis factor alpha and interleukin-4 production in vivo. *J Pharmacol Exp Ther 287*, 705–711.
123. Underwood DC, Osborn RR, Novak LB, Matthews JK, Newsholme SJ, Undem BJ, Hand JM, Torphy TJ (1993). Inhibition of antigen-induced bronchoconstriction and eosinophil infiltration in the guinea pig by the cyclic AMP-specific phosphodiesterase inhibitor, rolipram. *J Pharmacol Exp Ther 266*, 306–313.
124. Underwood DC, Kotzer CJ, Bochnowicz S, Osborn RR, Luttmann MA, Hay DW, Torphy TJ (1994). Comparison

of phosphodiesterase III, IV and dual III/IV inhibitors on bronchospasm and pulmonary eosinophil influx in guinea pigs. *J Pharmacol Exp Ther 270*, 250–259.

125. Underwood DC, Bochnowicz S, Osborn RR, Kotzer CJ, Luttmann MA, Hay DW, Gorycki PD, Christensen SB, Torphy TJ (1998). Anti-asthmatic activity of the second-generation phosphodiesterase 4 (PDE4) inhibitor SB 207499 (*Ariflo*) in the guinea pig. *J Pharmacol Exp Ther 287*, 988–995.
126. Bundschuh DS, Eltze M, Barsig J, Wollin L, Hatzelmann A, Beume R (2001). *In vivo* efficacy in airway disease models of roflumilast, a novel orally active PDE4 inhibitor. *J Pharmacol Exp Ther 297*, 280–290.
127. Baumer W, Gorr G, Hoppmann J, Ehinger AM, Ehinger B, Kietzmann M (2002). Effects of the phosphodiesterase 4 inhibitors SB 207499 and AWD 12-281 on the inflammatory reaction in a model of allergic dermatitis. *Eur J Pharmacol 446*, 195–200.
128. Baumer W, Gorr G, Hoppmann J, Ehinger AM, Rundfeldt C, Kietzmann M (2003). AWD 12-281, a highly selective phosphodiesterase 4 inhibitor, is effective in the prevention and treatment of inflammatory reactions in a model of allergic dermatitis. *J Pharm Pharmacol 55*, 1107–1114.
129. Hoppmann J, Baumer W, Galetzka C, Hofgen N, Kietzmann M, Rundfeldt C (2005). The phosphodiesterase 4 inhibitor AWD 12-281 is active in a new guinea-pig model of allergic skin inflammation predictive of human skin penetration and suppresses both Th1 and Th2 cytokines in mice. *J Pharm Pharmacol 57*, 1609–1617.
130. Kuss H, Hoefgen N, Johanssen S, Kronbach T, Rundfeldt C (2003). *In vivo* efficacy in airway disease models of *N*-(3,5-dichloropyrid-4-yl)-[1-(4-fluorobenzyl)-5-hydroxy-indole-3-yl]-glyo xylic acid amide (AWD 12-281), a selective phosphodiesterase 4 inhibitor for inhaled administration. *J Pharmacol Exp Ther 307*, 373–385.
131. Draheim R, Egerland U, Rundfeldt C (2004). Anti-inflammatory potential of the selective phosphodiesterase 4 inhibitor *N*-(3,5-dichloro-pyrid-4-yl)-[1-(4-fluorobenzyl)-5-hydroxy-indole-3-yl]-gly oxylic acid amide (AWD 12-281), in human cell preparations. *J Pharmacol Exp Ther 308*, 555–563.
132. Turner CR, Andresen CJ, Smith WB, Watson JW (1994). Effects of rolipram on responses to acute and chronic antigen exposure in monkeys. *Am J Respir Crit Care Med 149*, 1153–1159.
133. Ortiz JL, Cortijo J, Valles JM, Morcillo EJ (1992). Rolipram inhibits PAF-induced airway microvascular leakage in guinea-pig: A comparison with milrinone and theophylline. *Fund Clin Pharmacol 6*, 247–249.
134. Raeburn D, Karlsson JA (1993). Effects of isoenzyme-selective inhibitors of cyclic nucleotide phosphodiesterase on microvascular leak in guinea pig airways *in vivo*. *J Pharmacol Exp Ther 267*, 1147–1152.
135. Raeburn D, Underwood SL, Lewis SA, Woodman VR, Battram CH, Tomkinson A, Sharma S, Jordan R, Souness JE, Webber SE (1994). Anti-inflammatory and bronchodilator properties of RP 73401, a novel and selective phosphodiesterase type IV inhibitor. *Br J Pharmacol 113*, 1423–1431.
136. Svensjo E, Andersson KE, Bouskela E, Cyrino FZ, Lindgren S (1993). Effects of two vasodilatory phosphodiesterase inhibitors on bradykinin-induced permeability increase in the hamster cheek pouch. *Agents Actions 39*, 35–41.
137. Crummey A, Harper GP, Boyle EA, Mangan FR (1987). Inhibition of arachidonic acid-induced ear oedema as a model for assessing topical anti-inflammatory compounds. *Agents Actions 20*, 69–76.
138. Rabe KF, Bateman ED, O'Donnell D, Witte S, Bredenbroker D, Bethke TD (2005). Roflumilast—An oral anti-inflammatory treatment for chronic obstructive pulmonary disease: A randomised controlled trial. *Lancet 366*, 563–571.
139. Bergmann JF (2005). Roflumilast for chronic obstructive pulmonary disease. *Lancet 366*, 1845–1847.
140. Morice AH (2005). Roflumilast for chronic obstructive pulmonary disease. *Lancet 366*, 1845–1847.
141. Eller P, Pechlaner C (2005). Roflumilast for chronic obstructive pulmonary disease. *Lancet 366*, 1845–1846.
142. Boyd AE, Bhowmik A, Rajakulasingam K (2005). Roflumilast for chronic obstructive pulmonary disease. *Lancet 366*, 1846–1847.
143. Vassiliou V (2005). Roflumilast for chronic obstructive pulmonary disease. *Lancet 366*, 1846–1847.
144. Hughes EG (2003). Randomized clinical trials: The meeting place of medical practice and clinical research. *Semin Reprod Med 21*, 55–64.
145. Rabe KF (2005). Roflumilast for chronic obstructive pulmonary disease—Author's reply. *Lancet 366*, 1846–1847.
146. Grootendorst DC, Gauw SA, Baan R, Kelly J, Murdoch RD, Sterk PJ, Rabe KF (2003). Does a single dose of the phosphodiesterase 4 inhibitor, cilomilast (15 mg), induce bronchodilation in patients with chronic obstructive pulmonary disease? *Pulm Pharmacol Ther 16*, 115–120.
147. Calverley PMA, Sanchez-Toril F, McIvor A, Teichman P, Bredenbroker D, Fabbri LM (2007). Effect of one year treatment with roflumilast in severe chronic obstructive lung disease. *Am J Respir Crit Care Med 176*, 151–161.
148. Ram FS, Jardin JR, Atallah A, Castro AA, Mazzini R, Goldstein R, Lacasse Y, Cendon S (2005). Efficacy of theophylline in people with stable chronic obstructive pulmonary disease: A systematic review and meta-analysis. *Respir Med 99*, 135–144.
149. Giembycz MA (2002). Development status of second generation PDE4 inhibitors for asthma and COPD: The story so far. *Monaldi Arch Chest Dis 57*, 48–64.
150. Giembycz MA (2001). Cilomilast: A second generation phosphodiesterase 4 inhibitor for asthma and chronic obstructive pulmonary disease. *Expert Opin Invest Drugs 10*, 1361–1379.

151. GlaxoSmithKline (2003). SB 207499 (Ariflo, Cilomilast)—New drugs application (21–573). Pulmonary and allergy drug products advisory committee briefing document. Available: http://www.fda.gov/ohrms/dockets/ac/03/briefing/3976B1_02_D–FDA–%20Tab%203.pdf.

152. Provins L, Christophe B, Danhaive P, Dulieu J, Durieu V, Gillard M, Lebon F, Lengele S, Quere L, van Keulen B (2006). First dual M3 antagonists-PDE4 inhibitors: Synthesis and SAR of 4,6-diaminopyrimidine derivatives. *Bioorg Med Chem Lett 16*, 1834–1839.

153. Giembycz MA, Kaur M, Leigh R, Newton R (2008). A Holy Grail of asthma management: Toward understanding how long-acting $\beta_2$-adrenoceptor agonists enhance the clinical efficacy of inhaled corticosteroids. *Br J Pharmacol 153*, 1090–1104.

1º Giembycz MA (2007). Re-inventing the wheel: Nonselective phosphodiesterase inhibitors for chronic inflammatory diseases. In Beavo JA, Francis SH, Houslay MD, Eds. *Cyclic Nucleotide Phosphodiesterases in Health and Disease*. Boca Raton, FL: CRC Press, pp. 649–665.

155. Jin SL, Richter W, Conti M (2007). Insights into the physiological functions of PDE4 from knockout mice. In Beavo JA, Francis SH, Houslay MD, Eds. *Cyclic Nucleotide Phosphodiesterases in Health and Disease*. Boca Raton, FL: CRC Press, pp. 323–346.

156. Hansen G, Jin S, Umetsu DT, Conti M (2000). Absence of muscarinic cholinergic airway responses in mice deficient in the cyclic nucleotide phosphodiesterase PDE4D. *Proc Natl Acad Sci USA 97*, 6751–6756.

157. Jin SL, Conti M (2002). Induction of the cyclic nucleotide phosphodiesterase PDE4B is essential for LPS-activated TNF-$\alpha$ responses. *Proc Natl Acad Sci USA 99*, 7628–7633.

158. Jin SL, Lan L, Zoudilova M, Conti M (2005). Specific role of phosphodiesterase 4B in lipopolysaccharide-induced signaling in mouse macrophages. *J Immunol 175*, 1523–1531.

159. Ariga M, Neitzert B, Nakae S, Mottin G, Bertrand C, Pruniaux MP, Jin SL, Conti M (2004). Nonredundant function of phosphodiesterases 4D and 4B in neutrophil recruitment to the site of inflammation. *J Immunol 173*, 7531–7538.

160. Mealy NE, Bayes M (2005). Treatment of respiratory disorders. *Drugs of the Future 30*, 59–112.

# DRUG INTERACTIONS FOR THERAPEUTIC PROTEINS: A JOURNEY JUST BEGINNING

HONGHUI ZHOU and BERND MEIBOHM

## 1.1 INTRODUCTION

Over the last three decades, therapeutic proteins, in particular, antibody-based biotherapeutics, have played an increasingly important role in pharmacotherapy, and in some therapeutic areas, such as immune-mediated inflammatory diseases (IMIDs) and oncology, therapeutic proteins have fundamentally changed the therapeutic paradigm. Therapeutic proteins have also presented enormous commercial potential. For example, the top 10 antibody-based biotherapeutics accounted for around $50 billion of worldwide sales in 2011.[1] The majority of these are either in IMID (adalimumab, etanercept, infliximab, rituximab, natalizumab, omalizumab) or in oncology (rituximab, bevacizumab, trastuzumab, cetuximab) therapeutic areas. Hundreds of investigational antibody-based and other protein therapeutics are currently under development at different stages, spanning discovery to phase III clinical investigations.

Owing to an expected increase in the coadministration of biotherapeutic agents with established pharmacotherapy regimens, there is an increasing likelihood for the occurrence of clinically relevant drug interactions. Therapeutic proteins, however, have long been perceived to have a very low propensity for drug–drug interactions because they are eliminated via catabolic routes, either nonspecific pathways or target-mediated pathways, that are independent from the elimination pathways of small molecules, which are usually eliminated by noncatabolic pathways such as hepatic metabolism via cytochrome P450 (CYP), renal excretion, and biliary excretion. Though it has been known for decades that some cytokines such as interferons, tumor necrosis factor α (TNF-α), and

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*,
Edited by Shayne C. Gad
Copyright © 2013 John Wiley & Sons, Inc.

interleukin 6 (IL-6) can down-regulate CYPs,[2] very few drug–drug interactions had been reported for biotherapeutics until 2007, when two review articles containing examples of drug interactions involving therapeutic proteins were published.[3,4] The majority of reported drug interactions associated with therapeutic proteins seem to be indirect; however, a mechanistic understanding for many of the observed interactions is still lacking.[5–7]

## 1.2 SCIENTIFIC/REGULATORY LANDSCAPE OF THERAPEUTIC PROTEIN–DRUG INTERACTIONS

To help assess the common practice of evaluating therapeutic protein–drug interactions across the biotech/pharma industry and to shed some light on how and when a sensible therapeutic protein–drug interaction assessment strategy should be incorporated into therapeutic protein drug development, a survey was conducted within the Biotechnology Industry Organization (BIO) member companies in 2010. It is not surprising that a majority of the responder companies did not have internal strategies for evaluating therapeutic protein–drug interactions at the time of the survey. Nevertheless, the most favored approach employed to address potential drug–drug interactions of therapeutic proteins at that time was a tailored and integrated (i.e., case-by-case) strategy that addressed the possibility of the therapeutic protein acting as either an initiator (perpetrator) or target (victim) of the interaction. Despite the fact that many of the companies responding to the survey reported drug–drug interactions involving therapeutic proteins, the majority of the clinical therapeutic protein–drug interactions studied did not warrant dose adjustment. In other words, most of the observed clinical therapeutic protein–drug interactions did not reach a clinically significant level. Routine *in vitro* screening and preclinical drug–drug interaction studies were not widely used for the evaluation of therapeutic proteins. For clinical development, dedicated clinical pharmacology drug–drug interaction studies were the most frequently used methodology, followed by population pharmacokinetics-based and clinical cocktail approaches.[8]

The BIO survey results indicated that there was a pressing need to have a science-driven and risk-based assessment strategy for therapeutic protein–drug interactions (TP-DIs). A closer collaboration among scientists from the biotech/pharma industry, regulatory agencies, and academia appeared to be essential in reaching that goal. As a result, a TP-DI steering committee from industry, the FDA, and academia was founded in 2009 to address this challenge. The initial scope of this committee was focused only on pharmacokinetics (PK) and metabolism-based drug–drug interactions for the major classes of therapeutic proteins, including monoclonal antibodies, fusion proteins, cytokines (excluding antibody–drug conjugates). The committee intended to investigate the potential for therapeutic proteins to interact, either as initiators or targets, with drugs that are metabolized via CYP enzyme pathways. Two major focus areas the committee concentrated on were (1) to critically assess standard *in vitro* screening techniques and methodologies

(e.g., for cytokine-related drug–drug and drug–disease interactions) and (2) to provide guidance for study designs with consideration of specific disease area (e.g., oncology) issues and timings.

Several scientific knowledge gaps were identified from a 2010 American Association of Pharmaceutical Scientists (AAPS) workshop on Strategies to Address Therapeutic Protein-Drug Interactions during Clinical Development.[9] One gap was associated with the relevance of *in vitro* systems to assess potential therapeutic protein–drug interactions, and another gap was a lack of best practices for using population PK-based approaches to assess potential therapeutic protein–drug interactions. The steering committee also identified similar gaps and consequently formed two working groups to specifically tackle them.

During the same time period, scientists from the FDA published two important review articles on TP-DI, but these were mostly from a regulatory perspective.[10,11] In 2012, a draft of a new drug–drug interaction guidance document was made available by the FDA for public comments.[12] That draft included a dedicated section on therapeutic protein–drug interaction to address specifically the newly emerging area of drug–drug interactions with therapeutic proteins.

The Workshop on Recent Advances in the Investigation of Therapeutic Protein Drug-Drug Interactions: Preclinical and Clinical Approaches was held on June 4–5, 2012. The workshop, co-sponsored by the FDA Office of Clinical Pharmacology and the Drug Metabolism and Clinical Pharmacology Leadership Group of the IQ Consortium, was intended to facilitate a better understanding of the current science, investigative approaches, knowledge gaps, and regulatory requirements related to the evaluation of therapeutic protein–drug interactions. The workshop also provided an opportunity to discuss the current views from the two (*in vitro* and population PK approaches) therapeutic protein–drug interaction working groups. The proceedings from this workshop are being compiled with the intent of issuing white papers in these subject areas. It is anticipated that the recommendations from both white papers will soon provide pharmaceutical scientists with sensible and scientifically sound best practices and an assessment framework for using *in vitro* and population PK-based approaches for evaluating therapeutic protein–drug interactions.

Our current understanding of the mechanisms of many therapeutic protein–drug interactions is still in its infancy. Much basic research needs to be conducted to verify several existing hypotheses related to therapeutic protein–drug interactions. Continued close collaborations among fellow scientists in industry, academia, and regulatory agencies will be vital to generate more plausible mechanistic hypotheses and collectively address the many challenges in this area. Through these collaborative efforts, the knowledgebase on therapeutic protein–drug interactions will likely be largely expanded in the near future, and it is hoped and anticipated that over the next decade a similar level of mechanistic understanding and systemic assessment methodology will be achieved and developed for drug interactions with protein therapeutics as it has been established in the last two decades for small molecule drugs. The journey toward that goal has just begun.

## REFERENCES


1. R&D Pipeline News, Top 30 Biologics 2011, April 25, 2012. Available at www.pipelinereview.com.
2. Morgan ET. Regulation of cytochrome P450 by inflammatory mediators: why and how? *Drug Metab Dispos* **29**, 207–12 (2001).
3. Seitz K, Zhou H. Pharmacokinetic drug-drug interaction potentials for therapeutic monoclonal antibodies: reality check. *J Clin Pharmacol* **47**, 1104–18 (2007).
4. Mahmood I, Green MD. Drug interaction studies of therapeutic proteins or monoclonal antibodies. *J Clin Pharmacol* **47**, 1540–54 (2007).
5. Zhou H, Mascelli MA. Mechanisms of monoclonal antibody-drug interactions. *Annu Rev Pharmacol Toxicol* **51**, 359–72 (2011).
6. Kraynov E, Martin SW, Hurst S, et al. How current understanding of clearance mechanisms and pharmacodynamics of therapeutic proteins can be applied for evaluation of their drug-drug interaction potential. *Drug Metab Dispos* **39**, 1779–83 (2011).
7. Meibohm B. Mechanistic basis for potential drug-drug interactions with therapeutic proteins. Paper presented at the Workshop on Recent Advances in the Investigation of Therapeutic Protein Drug-Drug Interactions: Preclinical and Clinical Approaches. Silver Spring, MD, June 4–5, 2012.
8. Lloyd P, Zhou H, Theil FP, et al. Highlights from a recent BIO survey on therapeutic protein-drug interactions. *J Clin Pharmacol* **52**, 1755–63 (2012).
9. Girish S, Martin SW, Peterson MC, et al. AAPS workshop report: strategies to address therapeutic protein-drug interactions during clinical development. *AAPS J* **13**, 405–16 (2011).
10. Huang SM, Zhao H, Lee JI, et al. Therapeutic protein-drug interactions and implications for drug development. *Clin Pharmacol Ther* **87**, 497–503 (2010).
11. Lee JI, Zhang L, Men AY, et al. CYP-mediated therapeutic protein-drug interactions: clinical findings, proposed mechanisms and regulatory implications. *Clin Pharmacokinet* **49**, 295–310 (2010).
12. U.S. Department of Health and Human Services, FDA, and Center for Drug Evaluation and Research. Guidance for industry: drug interaction studies–study design, data analysis, implications for dosing, and labeling recommendations. February 2012. Available at www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM292362.pdf.

# PHARMACOKINETIC AND PHARMACODYNAMIC-BASED DRUG INTERACTIONS FOR THERAPEUTIC PROTEINS

DAN LU, SANDHYA GIRISH, FRANK-PETER THEIL, and AMITA JOSHI

## 2.1 INTRODUCTION

Therapeutic proteins (TPs) are protein products manufactured for pharmaceutical use. They include monoclonal antibodies (mAbs), antigen-binding fragments, antibody–drug conjugates (ADCs), cytokines, enzymes, growth factors, and miscellaneous proteins (e.g., fusion proteins and recombinant proteins). The development of therapeutic biologics, including TPs, is increasingly important in the pharmaceutical industry.[1] To achieve greater clinical benefits, TPs are often being combined with other TPs and small molecule drugs (SMDs). Whether drug interactions (DIs) in combination therapy result in an undesirable impact on efficacy and safety needs evaluation. To date, for the observed therapeutic protein–drug interactions (TP-DIs) that affect the exposure of TPs, only a modest change in exposure is observed and no impact on safety or efficacy has been documented, suggesting a limited clinical relevance.[2] This might be because most TPs have a relatively large therapeutic range compared to the majority of traditional SMDs. However, TP-DIs that affect the exposure of some drugs with a narrow therapeutic range (NTR), such as some SMDs and ADCs, may have an impact on efficacy and safety. The TP-DIs that result in enhanced toxicity due to undesirable pharmacodynamic (PD) interactions without a direct impact on exposures may also be clinically relevant. Thus the evaluation of TP-DIs is an important and evolving topic for the development of TPs in combination with other drugs.

This chapter reviews the major absorption, distribution, metabolism, and excretion (ADME) pathways of TPs, summarizes the potential mechanisms of

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*,
Edited by Shayne C. Gad
Copyright © 2013 John Wiley & Sons, Inc.

pharmacokinetic (PK) and PD-based TP-DIs, and recommends a question-based TP-DI risk assessment strategy during clinical development. The DIs for some nonprotein biologics such as nucleic acid–based derivatives are reviewed in other chapters.

## 2.2 DISTRIBUTION, CATABOLISM/METABOLISM, AND EXCRETION MECHANISMS OF THERAPEUTIC PROTEINS

ADME processes determine the PK properties of SMDs and TPs. In drug combinations, one drug may impact the ADME processes of another drug, leading to a change in its exposure. For SMDs, absorption is mainly mediated by the solubility and permeability of a SMD and its interaction with transporters. Distribution of SMDs is mediated by several key processes, such as blood perfusion, permeability across membrane barriers, and nonspecific binding. Metabolism of SMDs is mainly mediated by cytochrome P450 (CYP) and non-CYP enzymes (such as *N*-acetyl and glucuronyl transferase). Excretion of SMDs mainly occurs via renal filtration or renal and biliary secretion mediated by transporters.[3] Figure 2-1*a* depicts the typical clearance pathways for SMDs.

For TPs, ADME processes are different from SMDs.[4–6] Owing to high gastrointestinal enzyme activity and low permeability through the gastrointestinal mucosa, most TPs are not therapeutically active on oral administration. Consequently other routes of administration, such as intravenous, subcutaneous, and intramuscular routes of injection are used for TPs.[6] For subcutaneous injections of TPs with large molecular weight, convective transport across local lymphatic vessels is the major mechanism of absorption from the injection site.[7] The processes of distribution, catabolism, and excretion of TPs are reviewed in detail in this chapter. As illustrated in Figure 2-1*b*, the catabolism of TPs are mainly mediated by nonspecific clearance pathways. Target-mediated drug disposition (TMDD) and immunogenicity-mediated pathways also play roles in the clearance of some TPs. ADCs belong to a more complex group of TPs, made up of both a mAb and a small molecule cytotoxic agent. Their PK properties are also reviewed here.

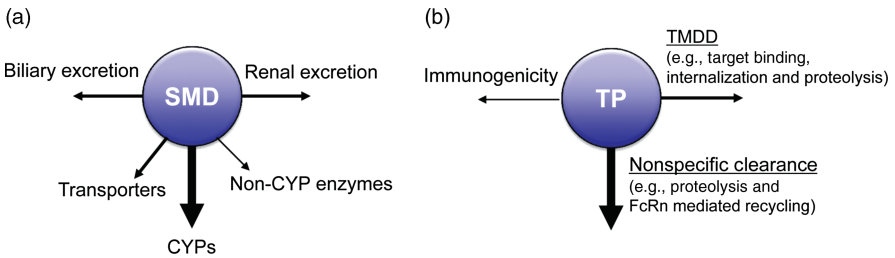


**FIGURE 2-1** Comparison of clearance mechanisms of (a) a SMD and (b) a TP. CYP: cytochrome P450; FcRn: neonatal Fc receptor; SMD: small molecule drug; TMDD: target-mediated drug disposition; TP: therapeutic protein.

### 2.2.1 Distribution of Therapeutic Proteins

Distinct from most SMDs that widely distribute to various tissues and organs after administration, distribution of mAbs and large TPs is usually confined by their large size; consequently the molecules have limited mobility through membranes. This often results in a relatively small volume of distribution. The volume of distribution of mAbs and ADCs at steady state is often a low multiple (1 to 2) of physiologic plasma volume (approximately 50 mL/kg). This is similar to the distribution characteristics for an endogenous immunoglobulin G (IgG). The distribution of TPs outside the systemic circulation is mediated by limited interstitial penetration in various organs, convection-dominated lymphatic drainage, specific and nonspecific binding to peripheral tissues, and target-mediated cellular uptake. For TPs with relatively low molecular mass, preclinical study results have demonstrated better tissue penetration.[6] Unlike SMDs, transporters usually do not play a role in the distribution of large TPs.

### 2.2.2 Catabolism of Therapeutic Proteins

Most TPs are mainly catabolized by proteolytic degradation in cellular lysosomes through nonspecific pathways, resulting in peptides and amino acids that are reutilized for protein synthesis.[4,6] It is generally believed that nonspecific catabolism of TPs may take place predominantly in the lysosomes of endothelial cells and the mononuclear phagocyte system (MPS). TPs, such as mAbs and some fusion proteins containing a fragment crystallizable region (Fc region), interact with neonatal Fc receptors (FcRn) similar to endogenous IgGs. In adults, FcRn is primarily expressed in the vascular endothelial cells. FcRn is also detectable on monocytes, tissue macrophages, and dendritic cells. The FcRn-mediated recycling protects IgG type of proteins (e.g., endogenous IgGs, mAbs, and Fc fusion proteins) from proteolytic degradation in lysosomes, consequently delaying their catabolism and prolonging their half-lives compared to other types of proteins that are not rescued by FcRn-mediated recycling.[4,5] As a result, endogenous IgGs, mAbs, and Fc fusion proteins usually have relatively long half-lives, ranging from several days to weeks. The pathways of nonspecific clearance and FcRn-mediated recycling are typically low-affinity and high-capacity pathways, which are usually nonsaturable at therapeutically relevant doses. For mAbs, relatively constant values of nonspecific clearance are found in each species. In humans, this value is 3–6 mL/day/kg and is affected by multiple pathophysiological and demographical covariates.[8]

In addition to the nonspecific clearance pathways, TMDD may also play a role in the clearance of target-binding proteins (e.g., mAbs, Fc fusion proteins, recombinant proteins). By this mechanism, a TP is cleared from the systemic circulation by binding to its target antigen followed by proteolytic degradation. The target antigens can be cell-surface receptors or soluble antigens. For targets that are cell-surface receptors, a TP is cleared after the TP–antigen complex is internalized and degraded in the lysosomes of target cells or when the TP-opsonized cell engages in immune effector function, which triggers apoptosis of the target cells by

complement-dependent cytotoxicity and antibody-dependent cellular cytotoxicity followed by degradation of the TP. For targets that are soluble antigens, a TP is cleared after the TP–antigen complex is eliminated via phagocytosis and proteolysis by endothelial cells and MPS. TMDD is typically a high affinity, low capacity and saturable process. When TMDD plays an important role in TP clearance, the PK parameters of the TP is concentration and dose dependent and may show a time-dependent decrease of clearance if receptor capacity is decreased after repeated treatment. For example, efalizumab[9] and panitumumab[10] show higher clearance at low concentrations and doses in clinical applications. The clearances of gemtuzumab and rituximab decrease after the second dose compared to the first dose, which may result from the decrease of target-mediated clearance after a reduction in target tumor cell number after the first dose of treatment.[11] For most TPs with TMDD involvement, the TMDD pathway is usually more dominant at low doses and low concentrations of the TPs when this pathway is not saturated. At therapeutic doses of these TPs, the therapeutic protein is often in great excess compared to the expression level of the respective target antigen available for binding under equilibrium conditions; consequently, the nonspecific clearance pathways play a dominant role. For these TPs at their prescribing doses (e.g., pertuzumab,[12] bevacizumab,[13] and trastuzumab[14]), changes of target antigen levels generally have a minimal impact on their clearance, and their PK parameters are concentration and dose independent.

The ability of TPs to elicit humoral responses, i.e., immunogenicity, can often modulate the clearance of TPs. The humoral response leads to the formation of antitherapeutic antibodies (ATAs), which may form immunocomplexes with TPs and consequently affect the clearance rates by affecting the binding of a TP to its target or affecting the nonspecific clearance pathways. For example, accelerated clearance of infliximab and adalimumab has been reported after development of ATA in rheumatoid arthritis (RA) patients.[15,16]

### 2.2.3 Excretion of Therapeutic Proteins

Excretion mechanisms for TPs also differ from those for SMDs. Renal clearance is generally negligible when the molecular size of a TP exceeds the cutoff size for renal filtration of approximately 45 kDa.[17] Tubular secretion does not occur to any significant extent for large TPs. The peptides resulting from TP catabolism may be partially reabsorbed in the proximal or distal tubule of the nephron or are further catabolized in kidney. Biliary excretion of TPs has been reported for only some fragment peptides and proteins such as immunoglobulin A and octreotide,[6,18] which are subsequently degraded in the gastrointestinal tract.

### 2.2.4 Pharmacokinetic Properties of Antibody–Drug Conjugates

ADCs, as an emerging class of TPs, have the PK properties of both mAbs and SMDs. ADCs are composed of a potent cytotoxic agent conjugated to a mAb via various types of linkers.[19,20] ADCs bind to their target receptors on the surface of tumor cells. The complexes are internalized and degraded and subsequently release

the cytotoxic agents to kill tumor cells. Usually the PK properties of multiple analytes, such as the conjugate and the unconjugated cytotoxic agent, are assessed after administration of an antibody–drug conjugate.

To date all ADCs are administered intravenously.[20] The distribution of ADCs is similar to their unconjugated mAbs. For example, in a preclinical *in vivo* study, it was found that trastuzumab emtansine (T-DM1), an ADC for the treatment of human epidermal growth receptor 2 (HER2) positive solid tumors, had similar tissue distribution to that of trastuzumab, the mAb component of T-DM1, indicating that conjugation does not impact the distribution of trastuzumab.[21] ADCs are catabolized by similar pathways as mAbs, including nonspecific proteolytic degradation and TMDD pathways. Immunogenicity may also play a role in ADC clearance.

In addition, the processes of linker chemistry-determined deconjugation in plasma and tissue are also involved in the catabolism and clearance of ADCs. The formation rate of the small molecule cytotoxic component by catabolism of the ADC is usually much slower than the elimination clearance of the small molecule cytotoxic component itself, resulting in formation rate-limited pharmacokinetics. Upon formation, these unconjugated cytotoxic molecules undergo typical clearance pathways of SMDs, such as hepatic metabolism and renal and biliary excretion, as mediated by CYP, non-CYP enzymes, and transporters.[21,22] The low dose of the SMD component of an ADC and relatively slow formation rate combined with a relatively fast elimination rate of the unconjugated SMD molecules may explain the observed relatively low systemic exposure of the unconjugated cytotoxic agent. For example, the average maximal concentration of the derivative of maytansine (DM1) is ~5 ng/mL after the administration of 3.6 mg/kg of T-DM1 every 3 weeks.[22] The average maximal free monomethyl auristatin E (MMAE) concentrations are 5–7 ng/mL after the every-3-week administration of 1.8–2.7 mg/kg of brentuximab vedotin,[23] a MMAE-containing ADC.[24]

## 2.3 MAJOR MECHANISMS OF THERAPEUTIC PROTEIN–DRUG INTERACTIONS

We are categorizing DIs as either PK based or PD based. PK-based DIs are those resulting from direct competition, inhibition, or induction of drug ADME mechanisms without involvement of the therapeutic targets. PD-based DIs are those resulting from modulation of the systems or target biology via the PD effects of drugs in combination. Both PK- and PD-based DIs may result in relevant changes in exposure and lead to a potential impact on safety and efficacy outcomes, especially for drugs with a NTR. PD-based DIs may also cause undesirable toxicity without an impact on exposure. Unlike SMDs, which are often susceptible to PK-based DIs due to an alteration in CYP and transporter-mediated ADME processes by drug combinations,[3,25] TP-DIs are often mechanistically different.

### 2.3.1 Impact of Pharmacokinetic-Based Therapeutic Protein–Drug Interactions on the Exposures of Therapeutic Proteins and Small Molecule Drugs

PK-based TP-DIs are not common because TPs and SMDs have distinct PK properties. The nonspecific clearance pathways for TPs are usually unsaturable at therapeutic concentrations. Likewise, these pathways are unlikely to be saturated by the combination of two TPs. For example, the clearances of trastuzumab and bevacizumab are dominated by nonspecific pathways at their clinically efficacious doses. No alteration in PK properties is observed when they are given in combination.[26]

When a TP is combined with a SMD, there is usually no direct overlap and competition in the metabolism and clearance pathways, thus PK-based DIs are unlikely. For example, chemotherapeutic agents such as irinotecan, 5-fluorouracil, and platinum-based therapy (i.e., cisplatin, carboplatin) do not affect the PK properties of cetuximab in cancer patients.[27–29] Similarly, no PK-based DIs are observed between bevacizumab and any of the following agents: capecitabine, cisplatin, 5-fluorouracil, irinotecan, oxaliplatin, or paclitaxel.[30] A dedicated study was conducted to evaluate potential TP-DIs for the combination of bevacizumab and irinotecan (as part of the FOLFIRI regimen containing irinotecan, fluorouracil, and leucovorin). This study demonstrated that the 90% confidence interval of geometric mean ratios for exposure of irinotecan and SN-38 (the active metabolite of irinotecan) in the absence of versus in the presence of bevacizumab were both within the prespecified no effect boundaries, indicating no clinically relevant TP-DIs for this combination.[31] Additional examples of no TP-DIs for combinations of anticancer mAbs with chemotherapeutic and antineoplastic SMDs have been reviewed in recent publications.[2,30,32–35]

PK-based TP-DIs involving ADCs are theoretically possible because the cytotoxic component of the ADC, once deconjugated, may elicit PK-based DIs when the ADC is combined with other SMDs. The cytotoxic agent, which is often a CYP substrate, is likely a victim of DIs when combined with SMDs that are CYP inhibitors or inducers. However, the cytotoxic agent has a relatively low systemic exposure. Thus it is not expected to have any impact on CYP and transporter activities in clinical settings and is unlikely to be a perpetrator. Data for ADC-related DIs are limited to assessments for T-DM1 and brentuximab vedotin (Adcetris). When T-DM1 is given in combination with taxanes (paclitaxel or docetaxel), the PK properties of taxanes and DM1 remain unchanged because taxanes and DM1 are not potent CYP inhibitors or inducers at clinically relevant concentrations.[36,37] A dedicated study of brentuximab vedotin found that it does not affect the PK parameters of midazolam, a CYP3A4 substrate. In the same study, the unconjugated MMAE exposure increased ~34% when brentuximab vedotin was combined with ketoconazole (a potent CYP3A4 inhibitor) and decreased ~46% when brentuximab vedotin was combined with rifampin (a potent CYP3A4 inducer). Therefore it is recommended that patients who are receiving strong CYP3A4 inhibitors concomitantly with brentuximab vedotin should be closely monitored for MMAE-related adverse reactions. These results are expected because MMAE is a substrate of CYP3A4 but not a CYP inhibitor or inducer at clinically relevant concentrations.[23]

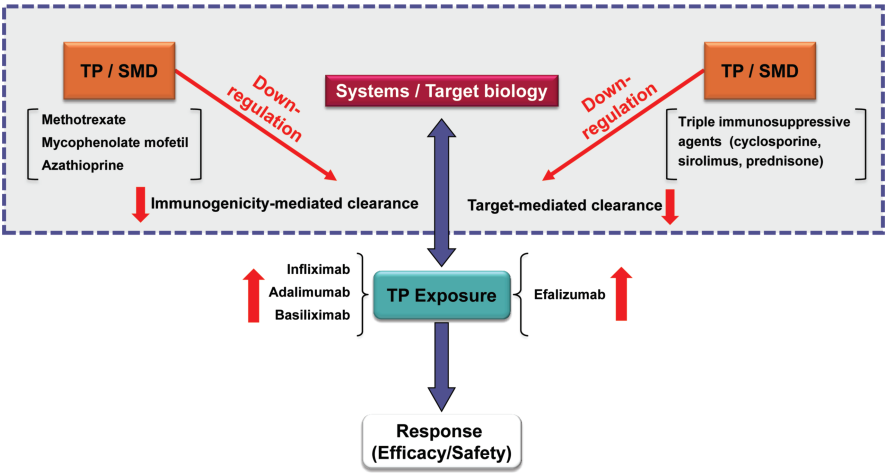


**FIGURE 2-2** Theoretical mechanisms of PD-based TP-DIs that change exposure of the TP. SMD: small molecule drug; TP: therapeutic protein.

### 2.3.2 Impact of Pharmacodynamic-Based Therapeutic Protein–Drug Interactions on the Exposure of Therapeutic Proteins

Distinct from the less common cases of PK-based TP-DIs, there are several plausible mechanisms for PD-based TP-DIs that change the exposure of TPs. As shown in Figure 2-2, interaction between the biological systems or target biology with a TP may affect the TP's exposure through immunogenicity-mediated clearance or target-mediated clearance pathways. The TPs or SMDs given in combination may modulate these clearance pathways by their PD effect, leading to DIs. For example, immunosuppressants such as methotrexate (MTX), mycophenolate mofetil, and azathioprine increase the exposures of infliximab, adalimumab, and basiliximab, possibly due to the effect of the immunosuppressants on decreasing the immunogenicity rate of these mAbs when they are given in combination. In another case, triple immunosuppressive agents may decrease target (CD11a+ T-cells) level, subsequently decreasing target-mediated clearance of efalizumab and thus increasing its exposure.

***2.3.2.1 Pharmacodynamic-Based Therapeutic Protein–Drug Interactions Owing to Changes in Immunogenicity-Mediated Clearance of Therapeutic Proteins*** Some immunosuppressive drugs may modulate the humoral immune response and decrease the immunogenicity of a TP, thus modulating its clearance. This is possible only for TPs that have a relatively high immunogenicity rate and when the clearances of TPs are impacted by immunogenicity. Examples include infliximab, adalimumab, and basiliximab, as listed in Table 2-1. Infliximab and adalimumab are both mAbs antagonizing tumor necrosis factor α (TNF-α)[44] and are often given in combination with immunosuppressive agents

**TABLE 2-1 Examples and Clinical Relevance of PD-Based TP-DIs That Have an Impact on PK Properties of TPs**

| Drug Combination | Type of TP and Indication | Possible Mechanisms for TP-DIs | Change in PK Properties | Clinical Relevance | References |
|---|---|---|---|---|---|
| Infliximab + MTX | Infliximab: anti-TNFα mAb<br>Indication: RA | MTX may inhibit humoral response; ↓ immunogenicity of infliximab. | MTX ↓ CL of infliximab. | No dose adjustment is recommended. | 38,39 |
| Adalimumab + MTX | Adalimumab: anti-TNFα mAb<br>Indication: RA | MTX may inhibit humoral response; ↓ immunogenicity of adalimumab. | MTX ↓ CL of adalimumab by 29–44%. | No dose adjustment is recommended. | 15,40 |
| Basiliximab + CsA + corticosteroids + mycophenolate mofetil or azathioprine | Basiliximab: chimeric mAb against IL-2Rα on T-cells<br>Indication: prevent transplant rejection | Mycophenolate mofetil or azathioprine may inhibit humoral response; ↓ immunogenicity of basiliximab. | Mycophenolate mofetil ↓ CL of basiliximab by 51%; azathioprine ↓ CL of basiliximab by 22%. | No dose adjustment is recommended. | 41,42 |
| Efalizumab + triple immunosuppressive agents (CsA, sirolimus, prednisone) | Efalizumab: humanized IgG1 anti-CD11a mAb<br>Indication: psoriasis, prevent transplant rejection | Triple immunosuppressive agents may ↓ TMDD of efalizumab by suppression of CD11a+ T-cells. | Triple immunosuppressive agents ↓ CL of efalizumab by 50% in transplant patients compared to psoriasis patients. | No dose adjustment is recommended. | 43 |

CL: clearance; CsA: cyclosporine; IL-2R: interleukin 2 receptor; mAb: monoclonal antibody; MTX: methotrexate; PD: pharmacodynamic; PK: pharmacokinetic; RA: rheumatoid arthritis; TMDD: target-mediated drug disposition; TNF: tumor necrosis factor; TPs: therapeutic proteins; TP-DIs: therapeutic protein–drug interactions.

such as MTX for the treatment of autoimmune diseases and prevention of rejection after organ transplantation. It was found that patients with positive ATA responses needed higher doses of infliximab to maintain the therapeutic concentration.[16] Similarly, serum trough adalimumab concentrations were dramatically lower in patients who developed antiadalimumab antibodies.[15] Concomitant dosing of MTX was associated with a decreased incidence of ATA formation against infliximab[38,45] and adalimumab.[15,45] As a result, when given in combination with MTX, the clearance of infliximab and adalimumab decreased and their serum concentrations increased.[15,38,40,44,45]

Basiliximab is a high-affinity chimeric IgG1-based mAb that binds to and blocks the interleukin 2 receptor α-chain (IL-2Rα) on the surface of activated T-lymphocytes and competitively inhibits IL-2-mediated activation of lymphocytes, a critical pathway in the cellular immune response involved in allograft rejection.[41] Basiliximab is often given in combination with triple immunosuppressive agents, including cyclosporine (CsA), corticosteroids, and either azathioprine or mycophenolate mofetil to prevent renal allograft rejection.[42] A cross-study comparison of basiliximab PK parameters in *de novo* renal allograft recipients showed that basiliximab clearance was $29 \pm 14$ mL/h when coadministered with azathioprine and $18 \pm 8$ mL/h with mycophenolate mofetil, and both were significantly lower compared to a clearance of $37 \pm 15$ mL/h from an earlier study of basiliximab with dual immunosuppressive agents of CsA and corticosteroids only.

Although the interstudy variability could potentially be a caveat of this observation, a plausible reason is that immunosuppressive agents such as azathioprine and mycophenolate mofetil may inhibit humoral immune responses and consequently decrease basiliximab clearance. Mycophenolate mofetil may have a stronger inhibition effect than azathioprine. The immunogenicity data in these studies further suggest that if the reduced clearance of basiliximab in the presence of azathioprine or mycophenolate mofetil is due to an inhibition of humoral immune responses, it appears that these responses are directed to a portion of basiliximab other than the IL-2Rα binding portion of basiliximab. Nonetheless, the PD effect of basiliximab as indicated by the range of the durations of IL-2Rα saturation when basiliximab was combined with triple immunosuppressive agents did not extend outside the range of the durations after the treatment of basiliximab in combination with dual immunosuppressive agents. Hence, regardless of the observed clearance difference, no dosing adjustment was necessary.[42]

***2.3.2.2 Pharmacodynamic-Based Therapeutic Protein–Drug Interactions Owing to Changes in Target-Mediated Drug Disposition of Therapeutic Proteins*** Combinations that affect the TMDD pathway of some TPs may result in PD-based TP-DIs, when TMDD plays an important role in the clearance processes of these TPs at their therapeutic doses. Based on a theoretical simulation of a typical mAb with strong TMDD (i.e., TMDD accounts for >50% of total clearance in the therapeutically relevant dose range), an increase of baseline receptor expression level (e.g., 3-fold or 10-fold of baseline) leads to a significant increase of the observed total clearance; a decrease of baseline receptor expression

level (e.g., 10%, 30% of baseline) leads to a significant decrease of the observed total clearance. The magnitudes of changes are higher at low dose levels, when TMDD accounts for higher percentage of total clearance (Figure 2-3*a*).

A clinical example is efalizumab in combination with immunosuppressive agents for preventing organ rejection after transplantation, is listed in Table 2-1. At clinical doses, TMDD of efalizumab after binding with CD11a receptors expressed on T-cells plays an important role in efalizumab clearance. Consequently, clearance of efalizumab is concentration and dose dependent and governed by CD11a expression level.[48] When efalizumab was given by subcutaneous administration in combination with oral administration of triple immunosuppressive agents (CsA, sirolimus, and prednisone) in transplant recipients, the clearance of efalizumab was 50% lower than the clearance observed in psoriatic patients. One of the reasons might be related to the reduction of CD11a+ T-cells in circulation by immunosuppressive agents given in combination in transplant recipients, resulting in reduced TMDD of efalizumab.[43,48]

However, in most clinical situations in which TPs are given at efficacious doses at which the TMDD pathway does not play a major role in their clearance, the nonspecific proteolytic degradation pathways, which are usually of high capacity and unsaturable, dominate the clearance of these TPs. They exhibit linear PK characteristics with relatively constant total clearance at therapeutic doses. For these TPs, the possibility of PD-based TP-DIs as a result of a target receptor expression level change by the combination drug is usually low. Based on a theoretical simulation of a typical mAb with negligible TMDD (i.e., TMDD accounts for $\leq$1% of total clearance in the therapeutically relevant dose range), an increase or a decrease of baseline receptor expression level (e.g., 3-fold, 10-fold, 10%, or 30% of baseline) does not have a significant impact on the observed total clearance (Figure 2-3*b*).

A clinical example is the combination of T-DM1 with pertuzumab, an ADC with a mAb.[49] Both the trastuzumab component of T-DM1 and pertuzumab bind to

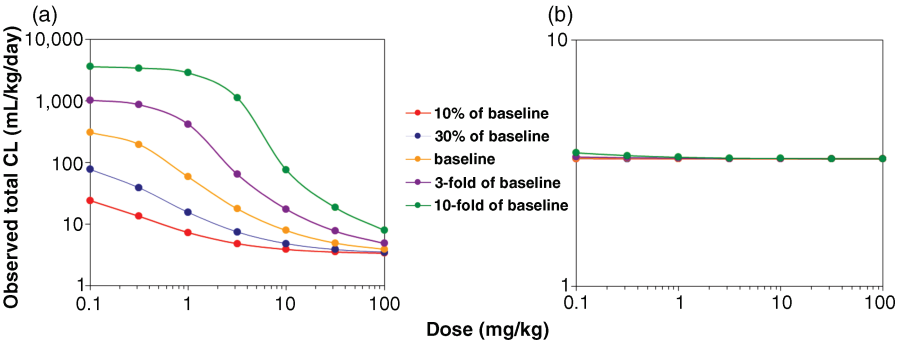


**FIGURE 2-3** Effect of receptor expression level on the dose-total CL relationship for a typical mAb. Observed total CL = dose/area under the concentration–time profile. Simulation is performed using a standard TMDD model after a single dose of a typical mAb [46,47]. The baseline free receptor expression levels was altered to 3-fold, 10-fold, 10%, or 30% of baseline, and the observed total CL values were plotted versus dose levels. (a) TMDD accounts for >50% of total clearance in the therapeutically relevant dose range. (b) TMDD accounts for $\leq$1% of total clearance in the therapeutically relevant dose range. CL: clearance.

HER2 receptors. Therefore a potential for PD-based DIs might exist for this combination. However, pertuzumab and T-DM1 bind to distinct epitopes of HER2 simultaneously without steric hindrance.[50,51] Furthermore, both T-DM1 and pertuzumab have linear PK characteristics at their clinically efficacious dose ranges when given in combination,[12,22] indicating that HER2-mediated clearance is unlikely to play a major role in their clearance in these dose ranges. Hence a potential impact on HER2 biology by T-DM1 is unlikely to affect the PK characteristics of pertuzumab, and vice versa. Cross-study comparisons of PK parameters of T-DM1 and pertuzumab in the combination study and the historical single-agent studies demonstrate that there is no DI observed for this combination.[49]

### 2.3.3 Impact of Pharmacodynamic-Based Therapeutic Protein–Drug Interactions on the Exposure of Small Molecule Drugs Owing to Therapeutic Protein–Cytokine–Cytochrome P450 Modulation Effects

Because TPs and SMDs do not share common ADME mechanisms, TPs are not predicted to affect the hepatic, renal, or biliary elimination of SMDs directly. However, some immunomodulatory TPs may exert an indirect effect on the hepatic clearance pathways of SMDs through the TP-cytokine-CYP modulation effect, resulting in PD-based TP-DIs. As summarized in Figure 2-4, infection and inflammatory diseases may cause an elevation of systemic cytokine level, such as interferons (IFNs), interleukins, and tumor necrosis factors, which in turn inhibit the CYP

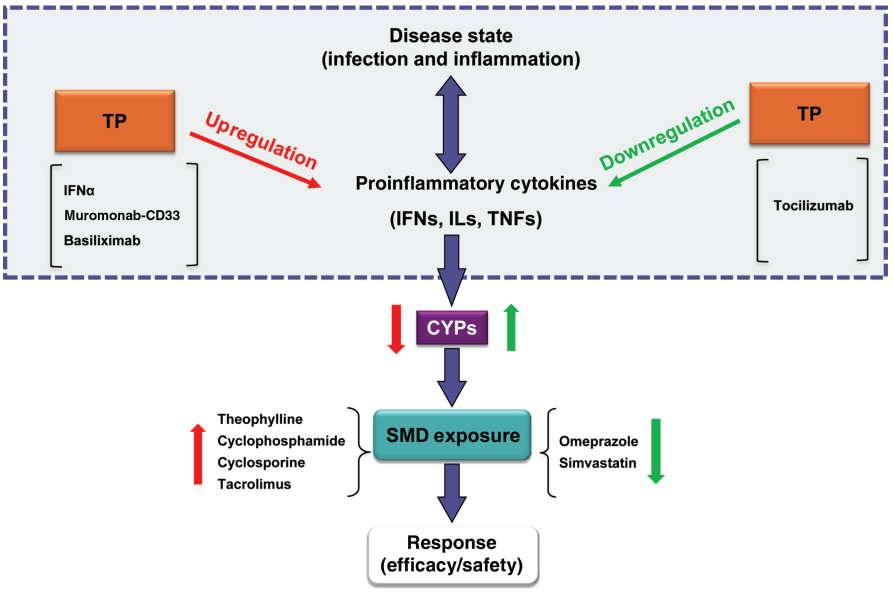


**FIGURE 2-4** Theoretical mechanisms of PD-based TP-DIs that have an impact on the PK properties of SMDs through TP-cytokine-CYP modulation effect. CYP: cytochrome P450; IFN: interferon; IL: interleukin; SMD: small molecule drug; TNF: tumor necrosis factor; TP: therapeutic protein.

activity and consequently increase the exposure of SMDs that are CYP substrates.[52] Some immunomodulatory TPs, such as exogenous IFNα, muromanab-CD3, and basiliximab may upregulate certain cytokine levels, inhibit CYP activities, and increase the exposure of CYP substrates (e.g., theophylline, cyclophosphamide, CsA, and tacrolimus). In contrast, some immunomodulatory TPs such as tocilizumab may downregulate the pathologically elevated cytokine levels in the inflammation disease state, normalize CYP activities to a higher level similar to healthy individuals, and decrease the exposure of CYP substrates (e.g., omeprazole, simvastatin). In addition to these clinical observations, there is strong *in vitro* evidence suggesting the differential effects of cytokines on various CYP enzymes.[1] The TP-DIs induced by the TP-cytokine-CYP modulation effect may be clinically relevant for SMDs with a NTR. The mechanisms and various scenarios of this type of TP-DI are reviewed in detail in this section.

Pathophysiological changes in patients and animal models in infectious or inflammatory conditions are often associated with decreased production and activities of hepatic and intestinal CYP enzymes, which are important for the first-pass extraction and metabolism of most SMDs. Based on *in vitro* and *in vivo* data, proinflammatory cytokines such as ILs (IL-1β, IL-6), TNFs (TNFα), and IFNs (IFNγ, IFNα-2b) may be potentially important mediators of this CYP modulation effect. These cytokines may induce a decrease in transcription factor activity for CYP enzyme expression or a decrease in CYP enzyme stability. As a result, CYP activities decrease in patients with infectious and inflammatory disease with elevated proinflammatory cytokine levels. The patients usually have an increased exposure of the SMDs metabolized by CYP enzymes because of decreased clearance and/or increased bioavailability compared with healthy individuals.[52]

Direct administration of exogenous cytokines may produce an immunopathological state similar to infection or inflammation, downregulate CYP activities, and potentially increase the exposure of some SMDs given in combination.[52] For instance, a clinical study using the "Pittsburgh cocktail" approach in high-risk melanoma patients found that the high-dose IFNα-2b[53] therapy differentially impairs CYP-mediated metabolism, causing no effect on some enzymes (e.g., CYP2E1) but substantial effects on others (e.g., CYP1A2: median 60% activity decrease; CYP2C19, median 40% activity decrease).[54] IFNα-2b has also been used in combination with theophylline for the treatment of various diseases (e.g., chronic hepatitis B). The clearance of theophylline, a CYP1A2 substrate drug with a NTR, decreased after IFNα-2b treatment, resulting in a 100% increase in serum theophylline levels.[53] In a clinical study in multiple myeloma patients, it was found that the administration of IFNα before cyclophosphamide (CP) caused a decrease in CP clearance to 63%, a 137% longer half-life and a 137% higher peak plasma concentration compared to the results obtained when IFNα was administered 24 h after CP. Consequently, the active metabolite 4-hydroxycyclophosphamide, which is formed by CYP-mediated metabolism from the prodrug CP, had 45% lower exposure when IFNα was given before CP compared with that observed when IFNα was administered 24 h after CP. As a result, a significant decrease in leukocyte count was observed when IFNα was given 24 h after CP.[55] These examples (some of them are listed in Table 2-2) demonstrate the

**TABLE 2-2 Examples and Clinical Relevance of TP-DIs Affecting the PK Properties of SMDs Caused by the TP-Cytokine-CYP Modulation Effect**

| Drug Combination | Type of TP and Indication | Possible Mechanisms for TP-DIs | Change in PK Properties | Clinical Relevance | References |
|---|---|---|---|---|---|
| IFNα-2b (Intron A) + theophylline | IFNα-2b: exogenous interferon<br>Indication: various diseases (e.g., chronic hepatitis B) | Theophylline is a CYP1A2 substrate<br>IFNα-2b inhibits CYP1A2. | IFNα-2b ↑ serum theophylline concentration by 100%. | Be cautious, considering theophylline is a NTR drug | 53 |
| IFNα + CP | IFNα: exogenous interferon<br>Indication: multiple myeloma | CP is converted to its active metabolite 4-OHCP by CYPs; IFNα inhibits CYP activity. | Administration of IFNα before CP caused ↓ CP clearance by 37%, ↑ half-life by 37%, and ↑ peak plasma concentration by 37%, compared to the results obtained when IFNα was administered 24 h after CP. | Sequence-dependent toxicity profile: a significant decrease in leukocyte count when IFNα was given 24 h after CP. | 55 |
| Muromonab-CD3 + CsA | Muromonab-CD3: murine IgG2a mAb, targets CD3 on T-cells<br>Indication: prevent acute transplant rejection | CsA is a CYP3A4 substrate.<br>Muromonab-CD3 results in T-cell activation and subsequent release of multiple | ↑ CsA trough by 100% at day 5 post-transplantation, compared to the control group | CsA doses were adjusted based on day 5 trough levels. | 56,57 |

(*continued*)

**TABLE 2.2** (*Continued*)

| Drug Combination | Type of TP and Indication | Possible Mechanisms for TP-DIs | Change in PK Properties | Clinical Relevance | References |
|---|---|---|---|---|---|
| | | cytokines immediately after injection, ↓ CYP3A4 activities. | | | |
| Basiliximab + tacrolimus<br>Basiliximab + CsA | Basiliximab: chimeric mAb antagonist of IL-2Rα on activated T-lymphocytes<br>Indication: prevent organ transplant rejection | Tacrolimus and CsA are CYP3A4 substrates<br>Binding of basiliximab to the IL-2R on activated T-cells may allow circulating IL-2 binding to IL-2Rs on hepatic and intestinal cells, resulting in ↓ CYP3A4 enzyme activity. | ↑ day 3 trough level of tacrolimus by 63%<br>↑ CsA levels in a pediatric patients receiving kidney allograft | Tacrolimus dose was decreased until the plasma concentration returned to desirable range.<br>Recommend that the initial CsA dose should be limited to 400 mg/m$^2$ if used in combination with basiliximab. | 41,58,59 |

| | | | | | |
|---|---|---|---|---|---|
| Tocilizumab + simvastatin or omeprazole | Tocilizumab: anti-IL-6R mAb that binds specifically to both sIL-6R and mIL-6R and inhibits IL-6-mediated signaling through blocking these receptors<br>Indication: RA | Simvastatin is a CYP3A4 substrate.<br>Omeprazole is a CYP2C19 and CYP3A4 substrate.<br>Tocilizumab may reverse IL-6-induced suppression of CYP activities in the disease state, leading to normalization of metabolism of drugs that are CYP substrates. | ↓ simvastatin exposure by 57%;<br>↓ omeprazole exposure by 12–28%. | Effect may be clinically relevant for CYP substrates with a NTR, for which dose is individually adjusted. | 60,61 |

CP: cyclophosphamide; CsA: cyclosporine; CYP: cytochrome P450; IFN: interferon; IL-2R: interleukin 2 receptor; IL-6: interleukin 6; IL-6R: IL-6 receptor; mAb: monoclonal antibody; mIL-6R: membrane-bound IL-6 receptor; NTR: narrow therapeutic range; 4-OHCP: 4-hydroxycyclophosphamide; PK: pharmacokinetic; RA: rheumatoid arthritis; sIL-6R: soluble IL-6 receptor; TP: therapeutic protein; TP-DIs: therapeutic protein-drug interactions.

effect of exogenous cytokines on the activity of CYPs in the clinical setting, which implies that endogenous cytokine levels may play an important role as modulators of hepatic metabolism.

Similarly, some mAbs may increase cytokine levels and activities by their PD effects and may lead to similar effects as the treatment by exogenous cytokines. As listed in Table 2-2, both muromonab-CD3 and basiliximab are mAbs for preventing rejection after organ transplantation.[41,56] Muromonab-CD3 is a murine IgG2a mAb targeting CD3 on T-cells to prevent acute organ rejection. Immediately after administration of muromonab-CD3, the activation of T-cells and release of numerous cytokines are observed followed by the blocking of all known T-cell functions. The initial T-cell activation and cytokine release may downregulate CYP enzyme activity and expression. CsA is a CYP3A4 substrate that may be combined with muromonab-CD3 in transplantation patients. It was found that muromonab-CD3 treatment caused a significant rise in CsA trough levels at day 5 post–renal transplantation compared to the group of patients receiving antilymphocyte globulin immunoprophylaxis therapy. As a result, CsA doses were adjusted based on trough levels obtained on day 5 post-transplantation.[57]

Clinically significant DIs were also reported for basiliximab in combination with CsA or tacrolimus. Both CsA and tacrolimus are metabolized by CYP3A4 and require therapeutic drug monitoring owing to their NTRs. In children who have undergone kidney transplant, basiliximab combination led to a higher trough concentration of CsA compared to the patients in the control group who were not on basiliximab. Consequently, it was recommended that the initial CsA dose should be limited to 400 $mg/m^2$ if used in combination with basiliximab in children with kidney transplants.[58]

Similarly, a 63% increase in tacrolimus trough levels was reported in basiliximab-treated adult patients on day 3 postrenal transplantation, when compared to the control group receiving antithymocyte globulin induction therapy without basiliximab treatment. As a result, the tacrolimus dose was decreased until the plasma concentration returned to a desirable range.[59] It was postulated that both of these clinically significant TP-DIs may be related to the effect of basiliximab on CYP activities via cytokine modulation. Interaction between IL-2 and its respective IL-2R on intestinal epithelial cells and hepatocytes may decrease the expression or activity of CYP3A4.[62] Binding of basiliximab to the IL-2R on activated T-cells may allow circulating IL-2 binding to IL-2Rs on hepatic and intestinal cells, resulting in a downregulation of CYP3A4 enzyme activity.

On the other hand, the treatment of some TPs results in a downregulation and normalization of elevated cytokine levels in the disease state through cytokine or cytokine receptor antagonism. Subsequent inhibition of cytokine-mediated downstream signaling transduction may restore the previously down-regulated CYP enzyme activity to a normal level and consequently normalize the exposure of SMDs that are CYP substrates.[52] As listed in Table 2-2, tocilizumab, an anti-IL-6R mAb for RA treatment, may have DIs based on this mechanism. In RA patients, IL-6 serum concentrations are elevated up to 50–60 pg/mL as compared to those in healthy subjects ($\sim$5 pg/mL). Tocilizumab binds specifically to both

soluble and membrane-bound IL-6 receptors (sIL-6R and mIL-6R) and inhibits IL-6-mediated signaling through blocking these receptors. This ameliorates the inflammation status.

*In vitro* data suggest that IL-6 inhibits CYP activity, and tocilizumab prevents CYP3A4 suppression by IL-6 in human hepatocytes, indicating that tocilizumab may reverse IL-6-induced suppression of CYP activities in the disease state.[60] This was demonstrated in a clinical study in RA patients in which the tocilizumab combination resulted in up to a 57% decrease in the exposure of simvastatin, a CYP3A4 probe substrate, in comparison to simvastatin alone in the same individuals. In another study, tocilizumab resulted in a 12–28% decrease in exposure of omeprazole, a CYP3A4 and CYP2C19 substrate, by comparing the exposure in the same RA patients before and after tocilizumab dosing. These observations may have clinical relevance for patients who take tocilizumab with NTR drugs for which the doses need to be individually adjusted.[60,61] Some other approved cytokine modulators, such as canakinumab (anti-IL-1β mAb),[63] golimumab (anti-TNFα mAb),[64] rilonacept (anti-IL-1R fusion protein),[65] and ustekinumab (anti-IL-12/23 mAb)[66] may also potentially normalize CYP enzymes in patients. Although no formal DI studies have been conducted for these TPs, the potential risk of TP-DIs and the need to monitor the therapeutic effects or concentrations of drugs with a NTR when given in combination with these TPs are addressed in their package inserts.

Transporters play important roles for the ADME process of SMDs. Inflammatory disease states may also affect several important drug transporters, such as *p*-glycoproteins, multidrug resistance-associated proteins, and organic anion-transporting peptides.[67] Theoretically, the risk of transporter-mediated DIs for the combinations of TPs and SMDs exists. However, to date no clinically relevant TP-DIs caused by this mechanism have been reported.

### 2.3.4 Pharmacodynamic-Based Therapeutic Protein–Drug Interactions That Lead to Toxicity without Affecting Exposures

The major objective for using combination therapy is to improve efficacy by desirable PD interactions without increasing safety concerns. As discussed earlier, PD-based TP-DIs may result in exposure changes in either the TP or the SMD given in combination, which may have clinical relevance for drugs with a NTR. In addition, PD-based TP-DIs can also lead to toxicity without affecting the exposure of the drugs in combination. PD-based TP-DIs that result in undesirable toxicity are usually documented in the labels of these TPs to guide the prescribers to avoid certain combinations. An example is the combination of anakinra and etanercept. Anakinra is a recombinant IL-1R antagonist protein that competitively inhibits the binding of IL-1 to IL-1Rs.[68] Anakinra has a molecular weight of 17.3 kDa. Etanercept is a dimeric fusion protein consisting of the extracellular ligand-binding portion of the human 75 kDa tumor necrosis factor receptor linked to the Fc portion of human IgG1.[69] Both TPs were approved for RA treatment. Because anakinra is eliminated by renal excretion and etanercept is eliminated by proteolysis with the salvage pathway of FcRn-mediated recycling, there are no overlapping clearance pathways for

them. In a phase III study in RA patients using concurrent therapy of anakinra with etanercept, the PK properties of both agents were unchanged by the combination. However, there was no significant improvement in efficacy for the combination therapy, and the risk of serious infection and neutropenia increased when compared to etanercept alone. PD-based DIs may potentially cause patients to be highly immunosuppressed and prone to infection, and an additive neutropenia effect also likely results from the use of this combination.[69,70]

Other examples are also recently reviewed.[30] For example, filgrastim, a human granulocyte colony–stimulating factor, during a period of 24 h before through 24 h after cytotoxic chemotherapy should be avoided because of the potential sensitivity of rapidly dividing myeloid cells to cytotoxic chemotherapy.[71] Aldesleukin,[72] a human recombinant IL-2 protein, may enhance the toxicity of drugs of nephrotoxic, myelotoxic, cardiotoxic, or hepatotoxic effects when given in combination. IFNα-2b[53] may cause increased myelotoxicity of myelosuppressive agents, such as zidovudine, when given in combination.

## 2.4 STRATEGIES TO ASSESS THE RISK OF THERAPEUTIC PROTEIN–DRUG INTERACTIONS IN CLINICAL DEVELOPMENT OF THERAPEUTIC PROTEINS

### 2.4.1 Challenges of Assessing Therapeutic Protein–Drug Interactions in Clinical Development

It is critical to understand the theoretical TP-DI mechanisms and build an individualized and iterative strategy to ensure the safe and effective use of TPs for combination treatment. The FDA guidance document for DI studies,[3] the European Medicines Agency (EMA) guideline on the clinical investigation of the PK characteristics of TPs,[73] and several recent review papers from the FDA highlighted the importance of evaluating TP-DIs.[1,2]

There are multiple challenges in assessing TP-DIs in clinical development of TPs, owing to the distinct PK and PD properties of TPs. Considering potential toxicity, most clinical trials involving TPs need to be conducted in the patient population instead of in healthy subjects. The PK characteristics of TPs may differ between patients and healthy subjects because of target expression levels. For immunomodulatory TPs, the CYP activity normalization by TP-cytokine-CYP interactions occurs only in patients with infection or inflammatory diseases. However, formal TP-DI studies in patients with life-threatening diseases (e.g., cancer, transplantation, severe autoimmune diseases) are rarely conducted owing to logistical constrains. Cross-over studies are difficult because of the long washout period required for TPs with long half-lives; a parallel design of two arms in a randomized study or a cross-over design for TP-DI assessment is not ethical if one of the drugs lacks efficacy when it is used alone. Concurrent medications also confound the DI evaluation of studied combinations[2,30] In addition to these challenges in clinical evaluation, *in vitro* and preclinical DI assessment strategies for TPs are very

different from SMDs and are still evolving owing to the inherent differences in metabolic pathways between TPs and SMDs.[30] Therefore, more creative and less disruptive strategies are needed to assess the risk of TP-DIs in clinical development.

### 2.4.2 Question-Based Strategy of Therapeutic Protein–Drug Interaction Assessment During Clinical Development

As summarized in Figure 2-5, a question-based TP-DI risk assessment strategy is recommended.[30] This strategy recommends both theoretical TP-DI risk assessments based on PK and PD properties of the drugs in combination (as described earlier in this chapter) and PK evaluation in phase Ib/II studies to support the theoretical assessment. If a clinically relevant TP-DI signal is detected, further assessment and confirmation might be considered in phase II/III studies. All TP-DI relevant information–including the theoretical risk of TP-DIs based on the PK and PD properties of the drugs in combination, the PK and ADME data collected at various stages in the development of combination therapy, the potential impact of TP-DIs on the safety and/or efficacy outcomes, and whether the impact warrants clinical dose adjustment and/or therapeutic drug monitoring—is documented in the package insert to provide physicians with data that allow an informed decision about whether to adjust a drug dose when administered in combination or whether to avoid a particular concomitant medication. The TP-DI risk assessments need an individualized strategy for each molecule and are iterative based on emerging data.

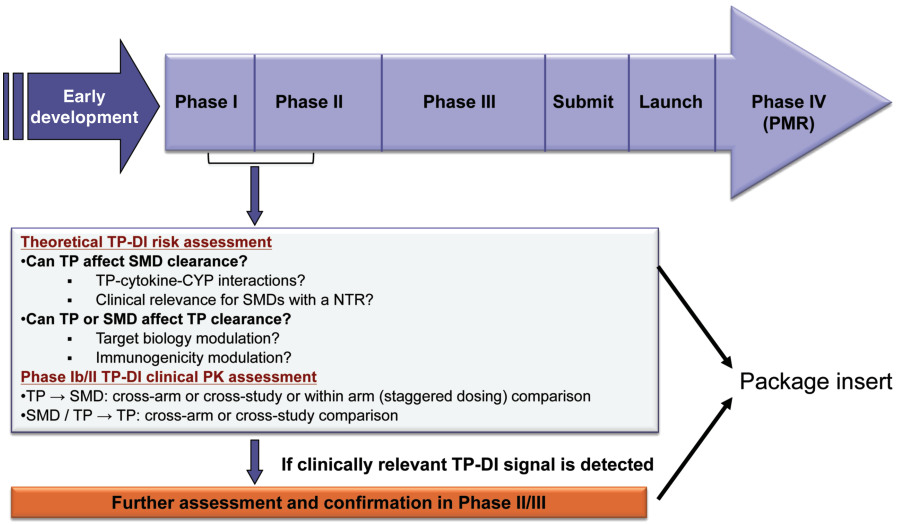


**FIGURE 2-5** Question-based TP-DI risk assessment strategy during TP development. CYP: cytochrome P450; NTR: narrow therapeutic range; PK: pharmacokinetic; PMR: post-marketing requirement; SMD: small molecule drug; TP: therapeutic protein; TP-DI: therapeutic protein–drug interaction.

***2.4.2.1 Theoretical Risk Assessments for Therapeutic Protein–Drug Interactions*** The PK and PD properties form the basis for the theoretical risk assessments of TP-DIs that may affect the exposure of combination drugs. To illustrate the general considerations for this theoretical assessment based on various mechanisms of TP-DIs, three tables provide examples for the combinations of a TP with a SMD or another TP in different therapeutic areas: a mAb combined with a SMD for cancer treatment (bevacizumab and irinotecan; Table 2-3), a mAb combined with a SMD for inflammatory disease treatment (tocilizumab and MTX or simvastatin; Table 2-4), and an ADC combined with a mAb for cancer treatment (T-DM1 and pertuzumab; Table 2-5).

***2.4.2.2 Evaluation of TP–DI Risk in Combination Studies by Clinical PK Assessment*** In addition to the theoretical TP-DI risk assessments, PK comparison with single-agent PK properties is recommended in phase Ib/II combination studies. These assessments are particularly important for drugs with a NTR, such as ADCs and some SMDs.

PK analysis for TP-DI assessment can be conducted by various approaches such as noncompartmental analysis (NCA) of individual patient PK profiles and population PK approaches using nonlinear mixed-effect modeling.[82,83] Among population PK approaches, the indicator approach and the simulation-based visual predictive check (VPC) approach are commonly used. With the indicator approach, a dichotomous indicator with the value of 1 or 0, corresponding to the presence or absence of the combination drug, is used as a covariate for PK parameters. The confidence intervals of the ratios from the PK parameters between the presence and absence of the interacting drug can be calculated. This approach was used to assess the impact of multiple combination agents on the PK properties of ustekinumab[84] and the impact of pertuzumab on the PK properties of T-DM1.[49] In addition, the VPC approach to PK data from combination studies based on simulations using a historical population PK model is also frequently used for cross-study PK comparisons.

Cross-arm and/or cross-study PK comparisons are often used for TP-DI assessment in phase Ib/II studies of combination therapy. For example, in a phase I combination study, MetMAb, a mAb directed against the tyrosine kinase Met, was given in combination with bevacizumab. Comparison of MetMAb in this combination with phase I single-agent MetMAb PK data using the NCA method and comparison of bevacizumab PK in this combination with historical bevacizumab PK data using the VPC method showed that no TP-DIs were observed. For a phase II study of MetMAb given in combination with erlotinib, the PK properties of MetMAb, erlotinib, and OSI-420 (a major metabolite of erlotinib) were also not altered in comparison to historical single-agent PK data using the VPC method and/or in comparison to PK parameters of the placebo plus the erlotinib arm in the same study.[85] In a phase II study of T-DM1 in combination with pertuzumab, the PK properties of pertuzumab and T-DM1-related analytes, including T-DM1, total trastuzumab, and DM1, were not changed compared to their historical single-agent data based on a cross-study comparison using both the NCA and population PK approach.[49]

**TABLE 2-3 Theoretical TP-DI Risk Assessments for the Combination of Bevacizumab and Irinotecan in Cancer Patients [13,31,74–76]**

| | | Bevacizumab | Irinotecan |
|---|---|---|---|
| PK characteristics | | Mainly cleared by nonspecific proteolytic degradation at clinically efficacious doses; linear PK property; $t_{½} = \sim 20$ days; CL = 0.2–0.3 L/day; interindividual CL variability = ~26%. | A prodrug metabolized to its active form, SN-38, by carboxyesterases; SN-38 is inactivated by UGT1A and CYP3A4; irinotecan $t_{½} = 6–12$ h; SN-38 $t_{½} = 10–20$ h. There is large interindividual PK variability for both irinotecan and its metabolite SN-38. UGT1A gene polymorphism may play a role in PK variability. |
| Risk of PK-based TP-DIs | | Low: No overlapping clearance pathways for bevacizumab and irinotecan | |
| PD characteristics | MOA | Binds to VEGF and prevents the interaction of VEGF to its receptors, leading to antiangiogenesis effect | SN-38 binds to topoisomerase I–DNA complex, preventing religation of a single-strand break induced by topoisomerase I, leading to DNA damage and cytotoxicity. |
| | TMDD | Not a major clearance route at clinically efficacious dose ranges | NA |
| | Immunogenicity | Low | NA |
| | Cytokine modulator | No | No |
| | Immuno suppressive agent | No | Yes, due to the neutropenia effect |
| | Therapeutic range | Moderate | Narrow |

(*continued*)

**TABLE 2.3** (*Continued*)

| | Bevacizumab | Irinotecan |
|---|---|---|
| Risk of PD-based TP-DIs | Risk of impact on bevacizumab exposure is low.<br>TMDD does not play a major role in total CL at clinically efficacious dose ranges.<br>Bevacizumab has relatively low immunogenicity; there is no documented impact of immunogenicity on bevacizumab CL. | Risk of impact on irinotecan exposure is low.<br>UGT1A and CYP activities are unlikely to be impacted by combination with bevacizumab |
| Risk of clinically relevant TP-DIs | Low | |

CL: clearance; CYP: cytochrome P450; MOA: mechanism of action; NA: not applicable; PD: pharmacodynamic; PK: pharmacokinetic; TMDD: target-mediated drug disposition; TP-DIs: therapeutic protein–drug interactions; UGT1A: uridine diphosphate glucuronosyltransferase 1A; VEGF: vascular endothelial growth factor.

For single-arm studies of TP-SMD combinations, a staggered dosing approach in the first treatment cycle may be applied to assess the impact of the TP on the PK properties of the SMD. In this approach, since the short half-lives of most SMDs determine a short wash-out period, a SMD may be administered alone in the first treatment cycle to allow PK assessment in the absence of the TP. In subsequent treatment cycles, the TP and the SMD are administered concurrently, and PK properties of the SMD in the presence of the TP can be assessed. This approach was used to evaluate the effect of T-DM1 on the PK properties of taxane,[36,37] the effect of bevacizumab on the PK of irinotecan,[31] and the effect of tocilizumab on the PK properties of simvastatin.[60]

A confirmatory TP-DI substudy in phase II/III might be needed when theoretical analysis and TP-DI assessment in phase Ib/II studies suggest a clinically relevant TP-DI potential. For example, it is hypothesized that tocilizumab may restore CYP activity in RA patients and therefore normalize the required doses of combination drugs of CYP substrates. To evaluate this hypothesis, a dedicated TP-DI study using simvastatin as the CYP3A4 probe substrate was conducted.[60]

In some situations, postmarketing studies to evaluate TP-DI might have to be conducted when the registration trial did not adequately address the potential for clinically relevant TP-DIs. For example, a dedicated TP-DI study for the combination of bevacizumab and irinotecan (as part of the FOLFIRI regimen combining irinotecan, fluorouracil, and leucovorin) was conducted as a postmarketing

**TABLE 2-4 Theoretical TP-DI Risk Assessments for the Combination of Tocilizumab and Other SMDs (e.g., Methotrexate or Simvastatin) in Rheumatoid Arthritis Patients [60,61,77,78]**

| | | Tocilizumab | MTX | Simvastatin |
|---|---|---|---|---|
| PK characteristics | | Nonlinear PK behavior at clinically efficacious doses; the total CL is the sum of linear and nonlinear CL; linear CL = 0.3 L/day; the concentration-dependent apparent $t_{½}$ = up to 11 days for 4 mg/kg and up to 13 days for 8 mg/kg at steady-state given on an every-4-weeks regimen. | Undergoes hepatic and intracellular metabolism to polyglutamated forms, which can be converted back to MTX by hydrolase enzymes; $t_{½}$ = 3–10 h; 80–90% of the administered dose is excreted unchanged in the urine within 24 h.<br>MTX clearance rates vary widely and are generally decreased at higher doses.<br>Not a CYP substrate | A lactone that is readily hydrolyzed *in vivo* to the corresponding β-hydroxyacid form, mediated mainly by CYP3A4, and extensively metabolized by the liver. |
| Risk of PK-based TP-DIs | | Low: No overlapping CL pathways for tocilizumab and MTX or tocilizumab and simvastatin | | |
| PD characteristics | MOA | Tocilizumab binds specifically to both soluble and membrane-bound IL-6R and inhibits IL-6-mediated signaling through these receptors; used for treatment of RA. | MTX inhibits dihydrofolic acid reductase, thus interferes with DNA synthesis, repair, and cellular replication; combination with tocilizumab is used for treatment of moderate to severe active RA in adult patients. | β-hydroxyacid is a potent inhibitor of HMG-CoA reductase, the enzyme that catalyzes the conversion of HMG-CoA to mevalonate, an early and rate-limiting step in the biosynthetic pathway for cholesterol. |
| | TMDD | Plays an important role in total CL at clinically efficacious dose ranges | NA | NA |

(*continued*)

**TABLE 2.4** (*Continued*)

| | Tocilizumab | MTX | Simvastatin |
|---|---|---|---|
| Immunogenicity | Low (~2%) | NA | NA |
| Cytokine modulator | Yes | No | No |
| Immuno-suppressive agent | Yes | Yes | No |
| Therapeutic range | Moderate | Narrow | Moderate |
| Risk of PD-based TP-DIs | Risk of impact on tocilizumab exposure is low.<br>MTX is unlikely to modulate IL-6R-mediated CL of tocilizumab.<br>Tocilizumab has low immunogenicity, and there is no documented impact of immunogenicity on tocilizumab CL. | Risk of impact on MTX exposure is low.<br>In RA patients, IL-6 concentrations are elevated, and CYP activities are suppressed.<br>Tocilizumab may restore the previously down-regulated CYP enzyme activity in RA patients to normal levels and consequently increase the metabolism of SMDs that are CYP substrates, through its inhibition of IL-6-mediated signaling. However, CYPs don't play a major role in MTX metabolism.<br>Risk of impact on simvastatin exposure is high: Simvastatin is mainly metabolized by CYP3A4. | |
| Risk of clinically relevant TP-DIs | Impact on exposure of CYP (e.g., CYP3A4) substrates: dose increase may be needed to maintain its therapeutic effect when given in combination with tocilizumab.<br>The effect of tocilizumab on CYPs may be persistent for several weeks after stopping treatment. | | |

CL: clearance; CYP: cytochrome P450; HMG-CoA: 3-hydroxy-3-methylglutaryl-coenzyme A; IL-6: interleukin 6; IL-6R: interleukin-6 receptors; MOA: mechanism of action; MTX: methotrexate; NA: not applicable; PD: pharmacodynamic; PK: pharmacokinetic; RA: rheumatoid arthritis; SMDs: small molecule drugs; TMDD: target-mediated drug disposition; TP-DIs: therapeutic protein–drug interactions.

**TABLE 2-5 Theoretical TP-DI Risk Assessments for the Combination of T-DM1 and Pertuzumab in HER2+ Metastatic Breast Cancer Patients [12,22,49,79–81]**

| | | T-DM1-Related Analytes | | |
|---|---|---|---|---|
| | | T-DM1 | DM1 | Pertuzumab |
| PK characteristics | | Mainly cleared by nonspecific proteolytic degradation at clinically efficacious doses; deconjugation may also clear T-DM1; linear PK property; CL = ~0.7 L/day; $t_{½}$ = 3–4 days; interindividual CL variability = ~21%. | Downstream catabolite of T-DM1, metabolized by CYP3A4 and CYP3A5; excreted mainly by biliary secretion; mean maximal concentration at clinically efficacious dose: 5–6 ng/mL | Mainly cleared by nonspecific proteolytic degradation at clinically efficacious doses; linear PK property; CL = ~0.2 L/day; terminal $t_{½}$ = 15–28 days; interindividual CL variability = ~31%. |
| Risk of PK-based TP-DIs | | Low: nonspecific proteolytic degradation pathways play a major role in the clearance of T-DM1and pertuzumab at clinically efficacious doses. The combination is unlikely to saturate these pathways.<br>No overlapping clearance pathway of DM1 and pertuzumab.<br>Low levels of unconjugated DM1 in systemic circulation and DM1 is not a CYP inhibitor or inducer and not a *p*-gp inhibitor at clinically relevant concentrations. So DM1 is not expected to be a perpetrator of DIs. | | |
| PD characteristics | MOA | T-DM1 binds to an epitope in subdomain IV of HER2, specifically delivers DM1 to HER2+ cancer cells, and induces cancer cell killing by antimitotic effect of DM1.<br>Trastuzumab component of T-DM1 inhibits HER2-mediated intracellular signaling and induces ADCC of HER2+ cancer cells. | | Pertuzumab binds to an epitope in subdomain II of HER2 and inhibits ligand-activated heterodimerization of HER2 with other HERs (e.g., HER1, HER3, and HER4) and subsequent intracellular signaling, leading to growth arrest and apoptosis of HER2+ cancer cells. |

(*continued*)

**TABLE 2.5** (*Continued*)

| | T-DM1-Related Analytes | | |
|---|---|---|---|
| | T-DM1 | DM1 | Pertuzumab |
| TMDD | Not a major CL route at clinically efficacious dose ranges | NA | Not a major CL route at clinically efficacious dose ranges |
| Immunogenicity | Low | NA | Low |
| Cytokine modulator | No | No | No |
| Immuno-suppressive agent | No | No | No |
| Therapeutic range | Narrow | Narrow | Wide |
| Risk of PD-based TP-DIs that impact on exposure | Low: T-DM1 and pertuzumab bind to distinct epitopes of HER2 simultaneously without steric hindrance.<br>TMDD does not play a major role in the CL of T-DM1and pertuzumab at clinically efficacious doses.<br>T-DM1and pertuzumab have low immunogenicity rates; there is no identified impact of immunogenicity on their CL; immunogenicity and CYP activities are unlikely to be affected by the combination. | | |
| Risk of clinically relevant TP-DIs | Low | | |

ADCC: antibody-dependent cellular cytotoxicity; CL: clearance; CYP: cytochrome P450; DIs: drug interactions; DM1: cytotoxic component of T-DM1; HER: human epidermal growth factor receptor; MOA: mechanism of action; NA: not available; PD: pharmacodynamic; *p*-gp: *p*-glycoprotein; PK: pharmacokinetic; T-DM1: trastuzumab emtansine; TMDD: target-mediated drug disposition; TP-DIs: therapeutic protein–drug interactions.

requirement, because a phase III study suggested a 33% increase of SN-38 (active metabolite of irinotecan) exposure and an increased incidence of grade 3 or 4 diarrhea and leukopenia in the bevacizumab and FOLFIRI combination arm compared to the FOLFIRI alone arm.[74] Whether there was a clinically relevant TP-DI potential was unknown. There are no theoretical mechanisms supporting this based on the PK and PD properties of bevacizumab and irinotecan, as assessed in Table 2-3. Unfortunately, the PK substudy in this phase III trial used limited sampling for irinotecan and thus was not definitive given irinotecan's large intersubject PK variability. The study also lacked an analysis of the uridine diphosphate glucuronosyltransferase 1A (UGT1A) genotype, which has been thought to affect the metabolic pathway of irinotecan. As a result, a controlled and fully powered clinical trial to fulfill the postmarketing requirement was conducted to investigate whether bevacizumab affected the PK properties of irinotecan and SN-38 using a staggered dosing design in cycle 1 and concurrent dosing in later cycles for bevacizumab and FOLFIRI. An exploratory analysis of UGT1A genotyping was also performed. This confirmatory study found that the PK characteristics of irinotecan and its active metabolite SN-38 were not altered by the bevacizumab combination. Pharmacogenetic analysis further confirmed that UGT1A1 gene polymorphisms were associated with variability in irinotecan and SN-38 exposure, which might explain the cross-arm exposure difference of SN-38 observed in the phase III study.[31]

## 2.5 SUMMARY

The PK properties of TPs are distinct from SMDs. TPs with a large molecular weight usually have a relatively small volume of distribution that is confined to plasma with limited penetration to the extravascular space. Unlike SMDs, most TPs are mainly catabolized via nonspecific proteolysis pathways with no involvement of CYP enzymes. FcRn-mediated recycling prolongs the half-lives of TPs that bind to FcRn, such as mAbs and some Fc fusion proteins. Some TPs may exhibit dose- and concentration-dependent PK due to a TMDD pathway. The clearance of TPs may also be affected by immunogenicity. Biliary secretion and renal excretion, the common routes of excretion for SMDs, are not common for TPs. The evolving class of ADCs belong to a more complex type of TPs made up of both a mAb and a small molecule cytotoxic agent and cleared by mechanisms relevant to both TPs and SMDs, as well as the deconjugation process determined by linker chemistry.

Owing to differences in the PK and PD properties of TPs and SMDs, the plausible mechanisms and relative risks for TP-DIs are distinct from SMD-DIs. PK-based DIs are not common for TPs, considering that the nonspecific proteolysis pathways are usually not saturable at the therapeutic concentrations of most TPs and there are usually no overlapping clearance pathways for TPs and SMDs. PD-based DIs induced by the PD effects of the combination drugs are possible for TPs. The major mechanisms of PD-based TP-DIs include modulation of immunogenicity-mediated clearance of TPs, modulation of TMDD, and modulation of CYP enzyme activity and expression through TP-cytokine-CYP interactions. To date, only modest

changes in exposure owing to TP-DIs have been observed with little relevance in the clinical setting. However, the risk of potential TP-DIs involving SMDs with a NTR may be clinically relevant. Because most drugs with a NTR are closely monitored by therapeutic drug monitoring, potential changes in exposure can typically be adjusted as needed to maintain efficacious concentrations and minimize the occurrence of adverse events. ADCs are likely to have similar mechanisms of PD-based TP-DIs, with the additional possibility of DIs between the small molecule cytotoxic component of the ADC and the combination drug.

The distinct PK and PD properties of TPs pose multiple challenges for assessing TP-DIs. Consequently, novel strategies that are different from the traditional DI assessment strategies for SMDs are required. A question-based TP-DI risk assessment approach is recommended during clinical development. This approach includes both theoretical risk assessments based on PK and PD properties of TPs and SMDs in combination and clinical PK evaluation in support of the theoretical assessment.

## ACKNOWLEDGMENTS

The authors acknowledge Dr. Yanan Zheng for her assistance with the preparation of Figure 2-3. The preparation of this manuscript was supported by Genentech Inc.

## REFERENCES

1. Lee JI, Zhang L, Men AY, et al. CYP-mediated therapeutic protein-drug interactions: clinical findings, proposed mechanisms and regulatory implications. *Clin Pharmacokinet* **49**(5), 295–310 (2010).
2. Huang SM, Zhao H, Lee JI, et al. Therapeutic protein-drug interactions and implications for drug development. *Clin Pharmacol Ther* **87**(4), 497–503 (2010).
3. FDA. Guidance for industry: drug interaction studies–study design, data analysis, and implications for dosing and labeling. 2006. Available at www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/ucm072101.pdf (retrieved October 2012).
4. Tabrizi MA, Tseng CM, Roskos, LK. Elimination mechanisms of therapeutic monoclonal antibodies. *Drug Discov Today* **11**(1–2), 81–88 (2006).
5. Mould DR, Green B. Pharmacokinetics and pharmacodynamics of monoclonal antibodies: concepts and lessons for drug development. *BioDrugs* **24**(1), 23–39 (2010).
6. Meibohm B. *Pharmacokinetics and pharmacodynamics of biotech drugs: principles and case studies in drug development*. Wiley-VCH, San Francisco, 2006.
7. Porter CJ, Edward, GA, Charman SA. Lymphatic transport of proteins after S.C. injection: implications of animal model selection. *Adv Drug Deliv Rev* **50**(1–2), 157–71 (2001).
8. Deng R, Iyer S, Theil FP, et al. Projecting human pharmacokinetics of therapeutic antibodies from nonclinical data: what have we learned? *MAbs* **3**(1), 61–66 (2010).

9. Ng CM, Joshi A, Dedrick RL, et al. Pharmacokinetic-pharmacodynamic-efficacy analysis of efalizumab in patients with moderate to severe psoriasis. *Pharm Res* **22**(7), 1088–100 (2005).
10. Ma P, Yang BB, Wang YM, et al. Population pharmacokinetic analysis of panitumumab in patients with advanced solid tumors. *J Clin Pharmacol* **49**(10), 1142–156 (2009).
11. Dowell JA, Korth-Bradley J, Liu H, et al. Pharmacokinetics of gemtuzumab ozogamicin, an antibody-targeted chemotherapy agent for the treatment of patients with acute myeloid leukemia in first relapse. *J Clin Pharmacol* **41**(11), 1206–14 (2001).
12. Ng CM, Lum BL, Gimenez V, et al. Rationale for fixed dosing of pertuzumab in cancer patients based on population pharmacokinetic analysis. *Pharm Res* **23**(6), 1275–84 (2006).
13. Lu JF, Bruno R, Eppler S, et al. Clinical pharmacokinetics of bevacizumab in patients with solid tumors. *Cancer Chemother Pharmacol* **62**(5), 779–86 (2008).
14. Bruno R, Washington CB, Lu JF, et al. Population pharmacokinetics of trastuzumab in patients with HER2+ metastatic breast cancer. *Cancer Chemother Pharmacol* **56**(4), 361–69 (2005).
15. Bartelds GM, Wijbrandts CA, Nurmohamed MT, et al. Clinical response to adalimumab: relationship to anti-adalimumab antibodies and serum adalimumab concentrations in rheumatoid arthritis. *Ann Rheum Dis* **66**(7), 921–26 (2007).
16. Haraoui B, Cameron L, Ouellet M, White B. Anti-infliximab antibodies in patients with rheumatoid arthritis who require higher doses of infliximab to achieve or maintain a clinical response. *J Rheumatol* **33**(1), 31–36 (2006).
17. Anderson NL, Anderson NG. The human plasma proteome: history, character, and diagnostic prospects. [Review]. *Mol Cell Proteomics* **1**(11), 845–67 (2002).
18. Chanson P, Timsit J, Harris AG. Clinical pharmacokinetics of octreotide. Therapeutic applications in patients with pituitary tumours. *Clin Pharmacokinet* **25**(5), 375–91 (1993).
19. Alley SC, Okeley NM, Senter PD. Antibody-drug conjugates: targeted drug delivery for cancer. *Curr Opin Chem Biol* **14**(4), 529–37 (2010).
20. Prabhu S, Boswell CA, Leipold D, et al. Antibody delivery of drugs and radionuclides: factors influencing clinical pharmacology. *Therapeutic Delivery* **2**(6), 769–91 (2011).
21. Shen B-Q, Bumbaca D, Saad O, et al. Metabolic fate and pharmacokinetic characterization of trastuzumab emtansine (T-DM1): an emphasis on preclinical and clinical catabolism *Clin Pharmacol Ther* **89**(S1), S90 (2011).
22. Girish S, Gupta M, Wang B, et al. Clinical pharmacology of trastuzumab emtansine (T-DM1): an antibody-drug conjugate in development for the treatment of HER2-positive cancer. *Cancer Chemother Pharmacol* **69**(5), 1229–40 (2012).
23. Seattle Genetics, Inc. Package insert for Adcetris (brentuximab vedotin, SGN-35). Bothell, WA, 2011.
24. Younes A, Bartlett NL, Leonard JP. Brentuximab vedotin (SGN-35) for relapsed CD30-positive lymphomas. *N Engl J Med* **363**(19), 1812–21 (2010).
25. European Medicines Agency. Guideline on the investigation of drug interactions. London, 2010. Available at www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2010/05/WC500090112.pdf.
26. Pegram M, Chan D, Dickmann R. Phase II combined biological therapy targeting the HER2 proto-oncogene and vascular endothelial growth factor using trastuzumab (T) and

bevacizumab (B) as first line therapy of HER2 amplified breast cancer. *Breast Cancer Res Treat* **100** (Suppl 1), S28 (2006).

27. Baselga J, Pfister D, Cooper MR, et al. Phase I studies of anti-epidermal growth factor receptor chimeric antibody C225 alone and in combination with cisplatin. *J Clin Oncol* **18**(4), 904–14 (2000).
28. Bourhis J, Rivera F, Mesia R, et al. Phase I/II study of cetuximab in combination with cisplatin or carboplatin and fluorouracil in patients with recurrent or metastatic squamous cell carcinoma of the head and neck. *J Clin Oncol* **24**(18), 2866–72 (2006).
29. Folprecht G, Lutz MP, Schoffski P, et al. Cetuximab and irinotecan/5-fluorouracil/folinic acid is a safe combination for the first-line treatment of patients with epidermal growth factor receptor expressing metastatic colorectal carcinoma. *Ann Oncol* **17**(3), 450–56 (2006).
30. Girish S, Martin SW, Peterson MC, et al. AAPS Workshop report: strategies to address therapeutic protein-drug interactions during clinical development. *AAPS J* **13**(2), 405–16 (2011).
31. Denlinger CS, Blanchard R, Xu L, et al. Pharmacokinetic analysis of irinotecan plus bevacizumab in patients with advanced solid tumors. *Cancer Chemother Pharmacol* **65**(1), 97–105 (2009).
32. Mahmood, I, Green MD. Drug interaction studies of therapeutic proteins or monoclonal antibodies. *J Clin Pharmacol* **47**(12), 1540–54 (2007).
33. Seitz K, Zhou H. Pharmacokinetic drug-drug interaction potentials for therapeutic monoclonal antibodies: reality check. *J Clin Pharmacol* **47**(9), 1104–18 (2007).
34. Zhou H, Davis HM. Risk-based strategy for the assessment of pharmacokinetic drug-drug interactions for therapeutic monoclonal antibodies. *Drug Discov Today* **14**(17–18), 891–98 (2009).
35. Zhou H, Mascelli MA. Mechanisms of monoclonal antibody-drug interactions. *Annu Rev Pharmacol Toxicol* **51**, 359–72 (2011).
36. Lu D, Krop I, Modi S, et al. Pharmacokinetics (PK) of trastuzumab-DM1 (T-DM1) and paclitaxel (T) in patients with HER2-positive locally advanced or metastatic breast cancer (MBC) Previously treated with a trastuzumab-containing regimen. Paper given at the San Antonio Breast Cancer Symposium, 2010, P3-14-22.
37. Lu D, Modi S, Elias A, et al. Pharmacokinetics (PK) of Trastuzumab emtansine and paclitaxel or docetaxel in patients with HER2-positive MBC previously treated with a trastuzumab-containing regimen. Paper given at the San Antonio Breast Cancer Symposium, 2011.
38. Maini RN, Breedveld FC, Kalden JR, et al. Therapeutic efficacy of multiple intravenous infusions of anti-tumor necrosis factor alpha monoclonal antibody combined with low-dose weekly methotrexate in rheumatoid arthritis. *Arthritis Rheum* **41**(9), 1552–63 (1998).
39. Janssen Biotech, Inc. Package insert for Remicade (infliximab). Horsham, PA, 2011.
40. Abbott, Laboratories. Package insert for Humira (adalimumab). North Chicago, 2002.
41. Novartis Pharmaceuticals, Corp. Package insert for Simulect (basiliximab). East Hanover, NJ, 2003.
42. Kovarik JM, Pescovitz MD, Sollinger HW, et al. Differential influence of azathioprine and mycophenolate mofetil on the disposition of basiliximab in renal transplant patients. *Clin Transplant* **15**(2), 123–30 (2001).

43. Vincenti F, Mendez R, Pescovitz M, et al. A phase I/II randomized open-label multicenter trial of efalizumab, a humanized anti-CD11a, anti-LFA-1 in renal transplantation. *Am J Transplant* **7**(7), 1770–77 (2007).

44. Emi Aikawa N, de Carvalho JF, Artur Almeida Silva C, Bonfa E. Immunogenicity of anti-TNF-alpha agents in autoimmune diseases. *Clin Rev Allergy Immunol* **38**(2–3) 82–89 (2010).

45. Vermeire S, Noman M, Van Assche G. et al. Effectiveness of concomitant immunosuppressive therapy in suppressing the formation of antibodies to infliximab in Crohn's disease. *Gut* **56**(9), 1226–31 (2007).

46. Ng CM. Stefanich E, Anand BS, et al. Pharmacokinetics/pharmacodynamics of nondepleting anti-CD4 monoclonal antibody (TRX1) in healthy human volunteers. *Pharm Res* **23**(1), 95–103 (2006).

47. Zheng Y, Scheerens H, Davis JC Jr, et al. Translational pharmacokinetics and pharmacodynamics of an FcRn-variant anti-CD4 monoclonal antibody from preclinical model to phase I study. [Clinical Trial, Phase I Research Support, Non-U. S. Gov't]. *Clin Pharmacol Ther* **89**(2), 283–90 (2011).

48. Joshi A, Bauer R, Kuebler P, et al. An overview of the pharmacokinetics and pharmacodynamics of efalizumab: a monoclonal antibody approved for use in psoriasis. *J Clin Pharmacol* **46**(1), 10–20 (2006).

49. Lu D, Burris HA 3rd, Wang B, et al. Drug interaction potential of trastuzumab emtansine (T-DM1) combined with pertuzumab in patients with HER2-positive metastatic breast cancer. *Curr Drug Metab* **13**(7), 911–22 (2012).

50. Franklin MC, Carey KD, Vajdos FF, et al. Insights into ErbB signaling from the structure of the ErbB2-pertuzumab complex. *Cancer Cell* **5**(4), 317–28 (2004).

51. Scheuer W, Friess T, Burtscher H, et al. Strongly enhanced antitumor activity of trastuzumab and pertuzumab combination treatment on HER2-positive human xenograft tumor models. *Cancer Res* **69**(4), 9330–36 (2009).

52. Morgan ET. Impact of infectious and inflammatory disease on cytochrome P450-mediated drug metabolism and pharmacokinetics. *Clin Pharmacol Ther* **85**(4), 434–38 (2009).

53. Schering, Corp. Package insert for Intron A (interferon-alfa-2b). Whitehouse Station, NJ, 2008.

54. Islam M, Frye RF, Richards TJ, et al. Differential effect of IFNalpha-2b on the cytochrome P450 enzyme system: a potential basis of IFN toxicity and its modulation by other drugs. *Clin Cancer Res* **8**(8), 2480–87 (2002).

55. Hassan M, Nilsson C, Olsson H, et al. The influence of interferon-alpha on the pharmacokinetics of cyclophosphamide and its 4-hydroxy metabolite in patients with multiple myeloma. *Eur J Haematol* **63**(3), 163–70 (1999).

56. Janssen, Pharmaceuticals. Package insert for Orthoclone (muromonab-CD3). Titusville, NJ, 1992.

57. Vasquez EM, Pollak R. OKT3 therapy increases cyclosporine blood levels. *Clin Transplant* **11**(1), 38–41 (1997).

58. Strehlau J, Pape L, Offner G, et al. Interleukin-2 receptor antibody-induced alterations of ciclosporin dose requirements in paediatric transplant recipients. *Lancet* **356**(9238), 1327–28 (2000).

59. Sifontis NM, Benedetti E, Vasquez, EM. Clinically significant drug interaction between basiliximab and tacrolimus in renal transplant recipients. *Transplant Proc* **34**(5), 1730–32 (2002).
60. Schmitt C, Kuhn B, Zhang X, et al. Disease-drug-drug interaction involving tocilizumab and simvastatin in patients with rheumatoid arthritis. *Clin Pharmacol Ther* **89**(5), 735–40 (2011).
61. Genentech, Inc. Package insert for Actemra (tocilizumab). South San Francisco, 2011.
62. Elkahwaji J, Robin MA, Berson A, et al. Decrease in hepatic cytochrome P450 after interleukin-2 immunotherapy. *Biochem Pharmacol* **57**(8), 951–54 (1999).
63. Novartis Europharm, Ltd. Package insert for Ilaris (canakinumab). West Sussex, UK, 2009.
64. Centocor Ortho Biotech Inc. Package insert for Simponi (golimumab). Horsham, PA, 2010.
65. Regeneron Pharmaceuticals, Inc. Package insert for Arcalyst (rilonacept) Tarrytown, NY, 2008.
66. Centocor Ortho Biotech, Inc. Package insert for Stelara (ustekinumab). Horsham, PA, 2009.
67. Petrovic V, Teng S, Piquette-Miller M. Regulation of drug transporters during infection and inflammation. *Mol Interv* **7**(2), 99–111 (2007).
68. Swedish Orphan Biovitrum AB. Package insert for Kineret (anakinra). Stockholm, 2011.
69. Immunex Corp. Package insert for Enbrel (etanercept). Thousand Oaks, CA, 2009.
70. Genovese MC, Cohen S, Moreland L, et al. Combination therapy with etanercept and anakinra in the treatment of patients with rheumatoid arthritis who have been treated unsuccessfully with methotrexate. *Arthritis Rheum* **50**(5), 1412–19 (2004).
71. Amgen Inc. Package insert for Neupogen (filgrastim). Thousand Oaks, CA, 2010.
72. Novartis Pharmaceuticals Corp. Package insert for Proleukin (aldesleukin) East Hanover, NJ, 2008.
73. European Medicines Agency. Guideline on the clinical investigation of the pharmacokinetics of therapeutic proteins. London, 2007. Available at www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC500003029.pdf.
74. Hurwitz H, Fehrenbacher L, Novotny W, et al. Bevacizumab plus irinotecan, fluorouracil, and leucovorin for metastatic colorectal cancer. *N Engl J Med* **350**(23), 2335–42 (2004).
75. Genentech, Inc. Package insert for Avastin (bevacizumab). South San Francisco, 2009.
76. Sagent Pharmaceuticals, Inc. Package insert for Camptosar (irinotecan). Schaumburg, IL, 2008.
77. Mayne Pharma (USA) Inc. Package insert for Trexall (methotrexate). Paramus, NJ, 2005.
78. Merck & Co, Inc. Package insert for Zocor (simvastatin). Whitehouse Station, NJ, 2011.
79. Gupta M, Lorusso PM, Wang B, et al. Clinical implications of pathophysiological and demographic covariates on the population pharmacokinetics of trastuzumab emtansine, a HER2-targeted antibody-drug conjugate, in patients with HER2-positive metastatic breast cancer. *J Clin Pharmacol* **52**, 691–703 (2012).
80. Krop IE, Beeram M, Modi S, et al. Phase I study of trastuzumab-DM1, an HER2 antibody-drug conjugate, given every 3 weeks to patients with HER2-positive metastatic breast cancer. *J Clin Oncol* **28**(16), 2698–704 (2010).

81. Lewis Phillips GD, Li G, Dugger DL, et al. Targeting HER2-positive breast cancer with trastuzumab-DM1, an antibody-cytotoxic drug conjugate. *Cancer Res* **68**(22), 9280–90 (2008).
82. Sheiner LB, Beal S, Rosenberg B, Marathe VV. Forecasting individual pharmacokinetics. *Clin Pharmacol Ther* **26**(3), 294–305 (1979).
83. Sheiner LB, Rosenberg B, Marathe VV. Estimation of population characteristics of pharmacokinetic parameters from routine clinical data. *J Pharmacokinet Biopharm* **5**(5), 445–79 (1977).
84. Zhu Y, Hu C, Lu M, et al. Population pharmacokinetic modeling of ustekinumab, a human monoclonal antibody targeting IL-12/23p40, in patients with moderate to severe plaque psoriasis. *J Clin Pharmacol* **49**(2), 162–75 (2009).
85. Bai S, Xin Y, Jin D, et al. Population pharmacokinetic analysis from phase I and phase II studies of the humanized monovalent antibody, MetMAb, in patients with advanced solid tumors. *J Clin Oncol* **29** (Suppl.) 2571 (2011).

# DRUG INTERACTION ASSESSMENT STRATEGIES: SMALL MOLECULES VERSUS THERAPEUTIC PROTEINS

SHANNON DALLAS, CARLO SENSENHAUSER,
SOUVIK CHATTOPADHYAY, and JOSE SILVA

## 3.1 INTRODUCTION

Paramount to every small molecule drug (SMD) development program is the assessment of drug interaction (DI) liabilities involving metabolic enzymes or drug transporters. Generally, these assessments are undertaken to understand how one SMD (perpetrator) may affect the absorption, distribution, metabolism, and excretion (ADME) of another SMD (victim) and therefore determine the potential for drug–drug interactions (DDIs). The importance of DDI assessment during SMD drug development (generally in the context of pharmacokinetics-based interactions) has been highlighted in recent guidance documents and papers published by regulatory agencies worldwide, including the FDA[1–3] and the European Medicines Agency (EMA).[4] Many preclinical studies (mostly *in vitro* but also *in vivo*) have been developed over the last few decades to understand how a particular SMD may have a potential DDI risk and whether such a risk will translate into the clinic. In contrast, for therapeutic proteins (TPs), or biologics, recent concern has centered more around the ability of a TP to affect the ADME of a coadministered SMD (i.e., as a perpetrator only), although exceptions do exist. These types of TP-DI studies historically were not part of standard Biologics License Application (BLA) submissions. Table 3-1 gives a few examples of approved TPs for which DI assessment was performed and reported in the compound label.[5–10]

DDI risk assessment for a SMD often starts with well-established *in vitro* studies using human-derived models (e.g., human hepatocytes and microsomes). Inhibition (reversible, time-dependent, or mechanism-based) and induction studies of the

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*,
Edited by Shayne C. Gad
Copyright © 2013 John Wiley & Sons, Inc.

**TABLE 3-1 Examples of TPs That Have Drug Interaction Information Listed on the Label**

| Drug | Drug Interactions Indicated on Label |
|---|---|
| Adalimumab (Humira); TNF blocker | MTX: Adalimumab has been studied in RA patients taking concomitant MTX. Although MTX reduced the apparent adalimumab clearance, the data do not suggest the need for dose adjustment of either drug. |
| Tocilizumab (Actemra); IL-6 receptor inhibitor | MTX: Population pharmacokinetic analyses of RA patients did not detect any effect of MTX, NSAIDs, or corticosteroids on tocilizumab clearance. Other drugs for treatment of concomitant administration of a single dose of 10 mg/kg tocilizumab with 10–25 mg MTX once weekly had no clinically significant effect on MTX exposure.<br>CYP450 enzymes: *In vitro* studies showed that tocilizumab has the potential to affect expression of multiple CYP enzymes, including CYP1A2, CYP2B6, CYP2C9, CYP2C19, CYP2D6, and CYP3A4. Its effects on CYP2C8 or transporters is unknown. *In vivo* studies with omeprazole (metabolized by CYP2C19 and CYP3A4) and simvastatin (metabolized by CYP3A4) showed up to a 28% and 57% decrease, respectively, in exposure 1 week after a single dose of tocilizumab. The effect of tocilizumab on CYP enzymes may be clinically relevant for CYP450 substrates with a narrow therapeutic index, for whichthe dose is individually adjusted. |
| Trastuzumab (Herceptin); HER2/neu receptor antagonist | Chemotherapeutic drugs: In a study involving metastatic breast cancer patients, the mean serum trough concentration of trastuzumab (4 mg/kg initial dose of trastuzumab followed by 2 mg/kg weekly) was consistently elevated approximately 1.5-fold when administered in combination with paclitaxel (175 mg/m$^2$ over 3 hours every 21 days for at least six cycles) as compared to trough concentrations of trastuzumab when administered in combination with an anthracycline and cyclophosphamide (doxorubicin 60 mg/m$^2$ or epirubicin 75 mg/m$^2$ plus 600 mg/m$^2$ cyclophosphamide every 21 days for six cycles). In other pharmacokinetic studies, where trastuzumab was administered in combination with paclitaxel, docetaxel, or doxorubicin, trastuzumab did not alter the plasma concentrations of these chemotherapeutic agents or the metabolites that were analyzed.<br>In another drug interaction study involving gastric cancer patients, the pharmacokinetics of cisplatin (80 mg/m$^2$ on day 1), capecitabine (1000 mg/m$^2$ orally twice a day on days 1–14), and their metabolites were not altered when administered in combination with trastuzumab (initial dose of 8 mg/kg administered on day 1 prior to chemotherapy, followed by 6 mg/kg every 21 days until disease progression). |

**TABLE 3.1** (*Continued*)

| Drug | Drug Interactions Indicated on Label |
|---|---|
| Belatacept (Nulojix); fusion protein (Fc portion of human IgG1 linked to CTLA-4 | MMF: In a pharmacokinetic study the plasma concentrations of MPA were measured in 41 patients who received fixed MMF doses of 500–1500 mg twice daily with either 5 mg/kg of belatacept or cyclosporine. The mean dose-normalized MPA $C_{max}$ and $AUC_{0-12}$ were approximately 20% and 40% higher, respectively, with belatacept coadministration than with cyclosporine coadministration. |
| Peginterferon α-2a (Pegasys) | Theophylline: Treatment with peginterferon α-2a once weekly for 4 weeks in healthy subjects was associated with an inhibition of CYP1A2 and a 25% increase in theophylline AUC.<br>Methadone: In a pharmacokinetic study of HCV patients concomitantly receiving methadone, treatment with peginterferon α-2a once weekly for 4 weeks was associated with methadone levels that were 10–15% higher than at baseline. The clinical significance of this finding is unknown. |
| Interferon γ-1b (Actimmune) | CYP450s: Preclinical studies in rodents using species-specific interferon-γ have demonstrated a decrease in hepatic microsomal CYP450 concentrations. This could potentially lead to a depression of the hepatic metabolism of certain drugs that use this degradative pathway. |

CYP450: cytochrome P450; HCV: hepatitis C virus; HER2: human epidermal growth receptor 2; IgG: immunoglobulin G; IL-6: interleukin 6; MPA: mycophenolic acid; MMF: mycophenolate mofetil; MTX: methotrexate; NSAID: nonsteroidal anti-inflammatory drug; RA: rheumatoid arthritis; TNF: tumor necrosis factor.

major cytochrome P450s, inhibition studies of non-CYP pathways (e.g., uridine 5′-diphospho-glucuronosyltransferases), and metabolic enzyme phenotyping are well established and run routinely. Although the liver is often the first organ examined, studies can be expanded to other organ systems (e.g., kidney or intestine) depending on a compound's characteristics (e.g., physiochemical properties or pharmacokinetics profile), mode and site of action, molecular target, clearance mechanisms, and clinical population. The potential role of drug transporters in compound disposition (as substrates, inhibitors, inducers, or suppressors) is assessed by numerous *in vitro* assays, which are also readily available. However, the relevance of individual transporters for potential DDIs is currently less well understood compared to metabolic enzymes. The *in vitro* kinetic data generated from these various assays (e.g., kinetic constants) can then be used to predict DDI risk clinically, using mechanistic modeling and simulation, such as physiologically based pharmacokinetic (PBPK) modeling. If a potential risk is identified, clinical DDI studies are conducted to characterize the magnitude of the interaction and predict how this might affect the safety and use of a potential new drug entity with comedications. PBPK modeling can provide simulations broadly across multiple populations, including special populations, such as pediatrics, geriatrics, and in renal or hepatic insufficiency. In addition, PBPK modeling is emerging as a viable alternative to actual

clinical DDI trials, in some cases. When submitting a new drug for approval, these studies would be a standard part of the drug application process. More recently, the possibility of some TPs, such as monoclonal antibodies and fusion proteins, to cause clinically relevant DIs (either directly or indirectly) has received renewed attention. The FDA and EMA have included language in their drug development guidance documents about how to assess the risk of a TP-DI.[3,11]

This chapter is not intended to give an expansive and detailed review of all of the methods and assays used for DDI assessment during SMD development because this has already been accomplished by others in the field.[12] Rather, general concepts and methods will be discussed for some of the typical assays undertaken in SMD development, including whether these methods can be, should be, or already have been applied in the context of TP-DI assessments. This chapter serves as a starting point for understanding the complexity and difficulties in assessing TP-DIs. The reader will be guided to other chapters within this volume that discuss these concepts more extensively and to specific case studies when appropriate.

## 3.2 DRUG-METABOLIZING ENZYMES

### 3.2.1 Cytochrome P450 Enzymes

Metabolic clearance generally involves biotransformation reactions catalyzed by a wide variety of drug-metabolizing enzymes (DME). Biotransformations are typically classified as either phase I or phase II reactions. Generally, phase I metabolic reactions aim at inactivating a drug through the introduction of a more polar functional group (—OH, —COOH, —$NH_2$, —SH, etc.) by oxidation, dealkylation, reduction, hydrolysis, cyclization, etc. Various classes of enzymes can be involved in phase I metabolism: cytochrome P450 (CYP), flavin-containing monooxygenases, amidases, carboxylesterases, and epoxide hydrolases. Because as much as 60% of currently marketed drugs are metabolized by one or several CYP isoforms,[13,14] focus has been directed at DDIs involving CYPs, although similar kinetic approaches can be applied to other classes of phase I enzymes.

Phase II metabolism, on the other hand, leads to formation of conjugates of the initial substrate—either the parent compound through direct conjugation or a phase I metabolite through indirect conjugation via uridine 5′-diphospho-glucuronosyltransferases (UGTs), glutathione-*S*-transferases, sulfotransferases, *N*-acetyltransferases, and methyltransferases—resulting in higher molecular weight compounds with improved water solubility, which increases renal or biliary clearance.[15,16] Phase II metabolism also provides a mechanism to inactivate potentially toxic electrophilic products of phase I metabolism. Although clinically relevant DDIs involving phase II metabolism have not been reported as frequently as with the CYPs, bioactivation through conjugation can occur, for instance through the formation of acyl glucuronides or thioesters resulting from conjugation of carboxylic acid moieties. Their inherent electrophilicity, deriving from the spontaneous tendency to undergo hydrolysis and intramolecular rearrangements (e.g., acyl migration) makes them

susceptible to protein adduct formation, and they have been implicated in observed adverse effects.[17] One notable example of phase II metabolism-derived DDI is the inactivation of CYP2C8 by gemfibrozil glucuronide, which will be briefly discussed later in the chapter.

The CYP superfamily of enzymes is widespread throughout nature and is present in all types of living organisms, ranging from bacteria to plants and animals. Of the more than 11,000 CYP proteins identified to date,[18] only a few dozen isoforms are of interest from a drug metabolism perspective and in terms of potential for adverse DDIs. Their nomenclature is based on the extent of amino acid sequence homology. Only the CYP1, CYP2, and CYP3 subfamilies are known to be primarily involved in the metabolism and bioactivation of pharmaceutical SMDs.[19] CYPs are proteins that contain a heme molecule that is noncovalently bound in the active site. The heme binds molecular oxygen and allows the catalytic cycle to carry out a biotransformation reaction by receiving protons and electrons from a cofactor, NADPH, and a closely associated oxidoreductase protein, producing 1 mol of water as a byproduct. CYP proteins, together with other phase I and some phase II enzymes are situated in the lipid bilayer of the endoplasmic reticulum, where they can efficiently catalyze the conversion of lipophilic xenobiotics to their oxidized metabolites and conjugates.

Other DMEs are found in the cytosolic region of the cell. Most DMEs share the characteristic of having broad substrate selectivity, being able to accommodate structurally diverse substrates in the active sites, albeit at the expense of turnover rates, which are typically much lower than those observed for enzymes involved in endogenous cell regulation and biosynthesis. This lack of substrate specificity, leading to the possibility of the same substrate being metabolized by different CYPs or different CYPs metabolizing the same substrate to produce different metabolites at different turnover rates, can ultimately lead to observed adverse DDIs. Characterization of the kinetic parameters of these reactions from *in vitro* assays can provide a valuable predictive tool during SMD development when incorporated with current modeling and simulation methods, such as PBPK.

### 3.2.2 Inhibition

A drug binding to a CYP enzyme typically displays hyperbolic saturation (Michaelis-Menten) kinetics, meaning that as the concentration of substrate is increased the first-order rate of product formation (often referred to as rate of substrate turnover or velocity), reaches a maximal rate of efficiency ($V_{max}$). Under these conditions, the enzyme present is completely in the enzyme–substrate (ES) complex form and is therefore saturated. Rates higher than $V_{max}$ are not possible; therefore, increasing substrate concentration beyond a certain point will not have any effect on substrate turnover, and zero-order kinetics are attained. In cases that deviate from classical Michaelis-Menten kinetics, such as substrate inhibition or autoactivation, changes in $V_{max}$ as substrate concentration increases can be observed). In addition, the substrate concentration that yields 50% of the maximal turnover rate, the Michaelis-Menten constant ($K_m$) represents the affinity of a substrate for the enzyme catalyzing its metabolic conversion. The ratio $V_{max}/K_m$, or

intrinsic clearance ($CL_{int}$), expressed in units of volume over time, provides a measure of the overall enzymatic capacity for a specific substrate. When this basic kinetic model is modified by the introduction of a second compound, which can either compete with the substrate for the active site of the same enzyme or act as an effector molecule, binding to an allosteric site of the enzyme, the rate of substrate turnover may be either inhibited or activated. In these cases, metabolic conversion of the substrate, the victim drug, will be driven not only by its concentration but also by the concentration of the perpetrator drug and the relative affinities of the two compounds to the enzyme. If a victim drug does not have alternate routes of elimination, inhibition of the enzyme will affect the exposure and pharmacokinetic profile more significantly. In addition to the *in vitro* determination of the kinetic parameters of inhibition, consideration of the fraction metabolized ($f_m$) by the inhibited enzyme and by other CYP and non-CYP enzymes, contributions by the gut to first-pass metabolism, the extent of biliary and renal clearance of the unchanged drug, and plasma protein binding all have to be taken into account as inputs in the prediction of changes in exposure (AUC) of a victim drug resulting from coadministration of a perpetrator drug.

In SMD development, several models, ranging from whole cell (hepatocytes) to subcellular fractions (microsomes; S9) and single enzyme expression systems, are routinely used in assays designed to evaluate the effect of a drug on CYP activity. For the determination of kinetic parameters, however, microsomes are the system of choice[20] because they provide kinetic data independent of other cellular and transport mechanisms present in whole cell systems. It is also widely recognized that Michaelis-Menten kinetic models require certain assumptions based on the steady-state, rapid equilibrium approach: (a) the amount of enzyme is negligible compared to the amount of substrate; (b) all of the enzyme can be accounted for as either free enzyme (E) or ES and enzyme-inhibitor (EI) complexes, all of which are in equilibrium for the duration of the assay; and (c) the amount of product formed is minimal compared to the substrate concentration, so the reaction remains within the linear range and substrate depletion is not rate limiting and does not affect assay kinetics. Consistent with these assumptions, microsomal assays should be run at low protein concentrations (typically $<0.5$ mg/mL, provided product formation is sufficiently detectable and linear within the time frame it is being measured), with constant conditions (cofactor, pH, temperature, and ionic strength) throughout the reaction.[20] Recent advances in methodologies, highlighted in the newest FDA DDI guidance[3] have included the identification and standardization of CYP-specific probe substrates and inhibitors, which specifically address the need to accurately monitor the activity of a CYP isoform without significant contributions from other isoforms. The widespread availability of metabolite reference standards and their deuterated analogs also allow for more accurate and sensitive quantification methods using current high-performance liquid chromatography tandem mass spectrometry (HPLC/MS/MS) analytical techniques.

***3.2.2.1 Reversible Inhibition*** Reversible inhibition results from noncovalent binding of an inhibitor to the enzyme in such a way that access of a substrate to the

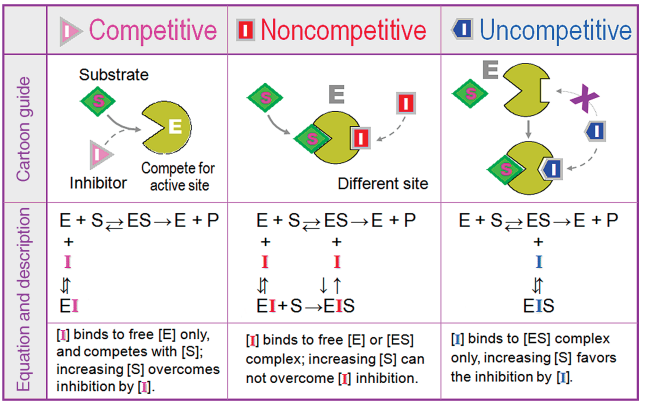


**FIGURE 3-1** Enzyme-substrate-inhibitor interactions and kinetic equilibria for the three major models of reversible inhibition. E: enzyme; ES: enzyme–substrate; I: inhibitor; P: product; S: substrate. (Reprinted with permission from Rong-Huay Juang, National Taiwan University.[70])

active site is prevented, either by directly competing for the active site, or by causing conformational changes to the enzyme (caused by allosteric effects) that affect the efficiency (turnover rate) of the reaction. The inhibitor could be a molecule that is structurally similar to the substrate but that is not necessarily chemically modified by the enzyme it inhibits or is modified at a substantially lower rate. It could also be a molecule that binds to a site different from the active site (allosteric or effector site), causing changes that prevent the substrate from binding. Therefore, from a kinetics standpoint, reversible inhibition models vary based on whether the inhibitor binds only to the free enzyme or also to the ES complex and whether product can be formed from the ES complex only or also from the enzyme-substrate-inhibitor (ESI) complex, albeit possibly at a reduced rate. Briefly, these models are summarized as follows (Figure 3-1):

- *Competitive inhibition.* Both substrate and inhibitor compete for the active site of the enzyme, and the equilibrium can be driven one way or the other by increasing the concentration of either compounds. As such, the $V_{max}$ is not affected, although more substrate is needed to compensate for the presence of the inhibitor and therefore apparent $K_m$ increases. Competitive inhibition can also be observed as a result of binding to an allosteric site on the enzyme, provided that it binds only to the free enzyme.
- *Mixed inhibition.* The inhibitor can bind to both E and the ES complex, to an allosteric site different from the binding site, causing conformational changes and thus reducing the efficiency of the enzyme and the affinity of the substrate for the enzyme. However, the equilibrium can be partially driven by increases in substrate concentration; therefore, apparent $K_m$ increases.

- *Noncompetitive inhibition.* A subset of mixed inhibition, the inhibitor binds to a different enzyme site but does not affect the affinity of the substrate binding to the active site, although it cannot be chemically modified by the enzyme. It reduces the efficiency of the enzyme, therefore affecting the maximal turnover rate ($V_{max}$ decreases), but because the affinity of the substrate for the enzyme does not change, apparent $K_m$ is not affected.
- *Uncompetitive inhibition.* The least common model with CYP enzymes, in which the inhibitor binds only to the ES complex, reducing its efficiency but increasing the apparent affinity of the substrate for the enzyme. In this case, both $V_{max}$ and $K_m$ decrease.

Experimental assessment of the potential to inhibit a specific enzyme typically begins by determining if incubation of the test compound in the presence of an established probe substrate causes loss of enzymatic activity as represented by a decrease in formation of the metabolite of the probe substrate. If the appropriate range of test compound concentrations is selected and, most important, if the probe substrate is incubated at a concentration that is equal or below its determined $K_m$ value, the $IC_{50}$ value of the inhibitor–that is, the concentration of inhibitor that lowers enzymatic basal activity levels by 50% can be determined. $IC_{50}$ values provide an approximate estimation of the potential of a compound to act as an inhibitor, but they are extrinsic constants and depend on experimental conditions and substrate concentrations.

The EI complex dissociation constant, $K_i$, is an intrinsic kinetic parameter and more appropriate in estimating potential DDIs. It can be estimated from determined $IC_{50}$ values, provided that substrate concentration is equal to its $K_m$ value and that the mode of inhibition is consistent with either the competitive model (in which case $IC_{50} = 2 \times K_i$) or the noncompetitive model ($IC_{50} = K_i$).[21] Experimentally, it is determined by incubating several substrate concentrations (preferably bracketing the substrate $K_m$ value) in the presence of a range of inhibitor concentrations (preferably bracketing 0.2 to 5 times the expected $K_i$ value)[22] and plotting the enzyme velocities ($v$, turnover rates) as functions of inhibitor ([I]) and substrate ([S]) concentrations. Graphical representations of the data, such as Michaelis-Menten ($v$ versus [S]), Lineweaver-Burk ($1/v$ versus 1/[S]), Dixon ($1/v$ versus [I]) or Eadie-Hofstee plots ($v$ versus $v$/[S]), can also provide insight on the appropriate inhibition mechanism (Figure 3-2).

More complex models have been extensively observed with enzymes exhibiting kinetic profiles that deviate from the traditional Michaelis-Menten model. CYP3A4, the most abundant and primary isoform involved in the metabolism of xenobiotics, has been shown to commonly exhibit non-Michaelis-Menten kinetics, due to homotropic cooperative effects resulting from the presence of multiple binding sites[23,24] or to flexible binding sites that can accommodate multiple substrates simultaneously.[25] These models can be represented as nonhyperbolic kinetic profiles—for example, in the case of substrate inhibition, as with midazolam, velocity decreases at higher substrate concentrations after an initial hyperbolic phase. In the case of autoactivation, as with testosterone, a sigmoidal profile of velocity as a function of

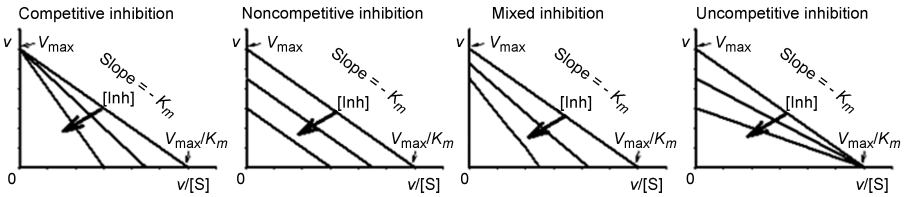


**FIGURE 3-2** Enzyme inhibition models. The different reversible inhibition models (competitive, noncompetitive, mixed, and uncompetitive) can be differentiated by using Eadie-Hofstee plots. The *y*-intercept corresponds to $V_{max}$, the slope corresponds to $-K_m$ and the *x*-intercept corresponds to $V_{max}/K_m$, or the intrinsic clearance ($CL_{int}$). Inh: inhibition.

substrate concentration is observed.[26,27] It is common practice, therefore, to screen interactions between a potential CYP3A4 inhibitor with more than one CYP3A4-specific substrate, to assess binding to the different domains (midazolam, testosterone, and nifedipine are used as the prototypical substrates).[28]

As inhibition constants provide an indication whether exposure of an enzyme to a perpetrator drug may have a concomitant effect on the pharmacokinetic profile of a victim drug, their value is essential in estimating the potential for *in vivo* DDI by a SMD. Therapeutic proteins, though unlikely to have similar kinetic intracellular interactions with CYP enzymes, could trigger different indirect mechanisms affecting the basal activity levels of an enzyme, prompting as a result, similar pharmacokinetic interactions (see section 3.2.4).

***3.2.2.2 Time-Dependent Inhibition*** When an inhibitor or substrate is converted enzymatically to a reactive species, it may inactivate the enzyme in a time- and concentration-dependent fashion.[29] Recently, emphasis has been placed on differentiating between time-dependent inhibition (TDI) and a particular subset defined as mechanism-based inactivation (MBI). Several mechanisms can trigger TDI of an enzyme. They can be typically identified after a preincubation period in the presence of a metabolizing system and are therefore NADPH dependent. Some of these mechanisms may not necessarily require bioactivation and may result, for instance, from the formation of species with a higher inhibitory affinity than the parent compound. A compound causing MBI, on the other hand, is one that binds catalytically to the target enzyme and is metabolically converted to a reactive species. This reactive metabolite can then bind covalently and irreversibly to a moiety of the active site of the apoprotein or in a quasi-irreversible fashion, if a reactive species forms a metabolic intermediate complex (MIC) with the heme group of the CYP enzyme. For example, alkyl and aromatic amines can be bioactivated to form a nitroso intermediate species, which coordinates tightly with the heme but in a noncovalent way. This interaction can be reversed almost completely by oxidizing agents, such as potassium ferricyanide, as in the case with troleandomycin.[30]

Greater concern from a DDI perspective is generated when these reactive intermediates interact covalently or quasi-irreversibly to CYP3A4. Often, DDIs triggered by these irreversible mechanisms can be more unfavorable compared to DDIs owing to reversible inhibition because the inactivated enzyme has to be replaced by

newly synthesized protein (*de novo* synthesis). As in the case with reversible inhibition, the potency of inactivation can be assessed by determining the kinetic parameters of inactivation, $k_{inact}$, defined as the maximal rate of inactivation and $K_I$, indicating the potency or affinity of the inactivator for the enzyme, defined as the concentration that causes 50% of the maximal rate of inactivation. An additional parameter provides a measure of efficiency of the inactivation and is defined as the number of inactivator (substrate) molecules that are turned over per molecule of enzyme that is inactivated; it is called the apparent partition ratio (APR).[31–33]

APR screens can be carried out easily and effectively by incubating the test compound at serially diluted concentrations, with and without NADPH, in a primary reaction mixture. After 60 min, an aliquot of the reaction mixture is transferred to a secondary reaction that contains an excess of CYP-specific probe substrate and additional NADPH, to assay for residual enzymatic activity. A dilution in the transfer step, typically 10- to 20-fold and the excess of probe substrate, ensure that any observed loss in residual activity is the result of inactivation of the enzyme during the primary incubation and not due to possible reversible interactions with the unchanged inactivator or any generated metabolite transferred to the secondary reaction mixture. The residual activity data are normalized to the maximal activity observed in the absence of inactivator and are plotted as a function of the molar ratio of inactivator to CYP enzyme. Any separation between the curves corresponding to the +NADPH and −NADPH primary incubation samples indicate that the observed loss of activity appears to be time, concentration, and NADPH dependent, identifying a TDI and possible MBI. The APR value can be determined from the linear range of the activity curve as a function of molar ratio, which can be extrapolated to the intercept of the $x$-axis, as shown by the interaction of mifepristone a known 3A4 MBI (Figure 3-3). In contrast, ketoconazole, a strong reversible inhibitor exhibits the opposite behavior (greater loss of activity without NADPH), as the potent reversible inhibition interacts with the probe substrate.

Although microsomes have traditionally been the system of choice for TDI screens and kinetic determinations, hepatocytes have recently gained increasing value as a better predictive model that incorporates mechanisms and transport not present with microsomes. A study of several known CYP3A4 inactivators resulted in generally lower potencies, either lower affinities or lower rates of inactivation than those typically calculated from microsomal data. However the hepatocyte-derived lower potencies appeared to correlate better with observed clinical data when simulated with PBPK.[34]

Determination of the kinetic parameters of inactivation, $k_{inact}$ and $K_I$, is an extension of the APR assay. Again, serially diluted test compound concentrations are incubated in a primary reaction containing NADPH, but activity is assayed at several time points. The apparent rates of inactivation generated from the semilog time course can be plotted as a function of the inactivator to yield the customary Michaelis-Menten-like hyperbolic plot, from which the kinetic parameters of inactivation can be determined by nonlinear regression (Figure 3-4). The concept of screening for potential TDIs or MBIs has become pivotal in SMD development, as evidenced also by the postmarket withdrawals of several previously approved drugs found to

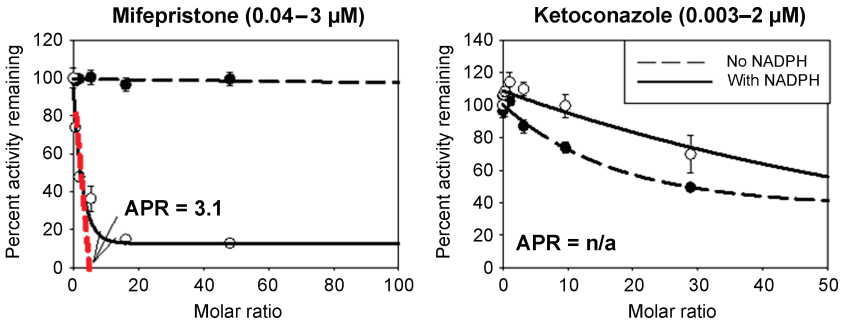


**FIGURE 3-3** APR curves for mifepristone, a known CYP3A4 TDI, and ketoconazole, a potent CYP3A4 reversible inhibitor. Serially diluted concentrations of either compound were incubated with an NADPH (solid lines) or without an NADPH (dashed lines) regenerating system for 60min. Residual CYP3A4 activity was assayed by transferring an aliquot of each initial incubation to secondary incubations containing midazolam (CYP3A4 probe substrate) and NADPH. Residual activity was normalized to the activity in a vehicle control and was graphed as a function of inactivator to CYP3A4 molar ratio. The concentration- and NADPH-dependent loss of activity was consistent with a TDI. Ketoconazole, as a potent but reversible inhibitor, does not cause NADPH-dependent loss of activity and exhibits stronger loss of activity in the absence of NADPH. APR: apparent partition ratio; TDI: time-dependent inhibitor.

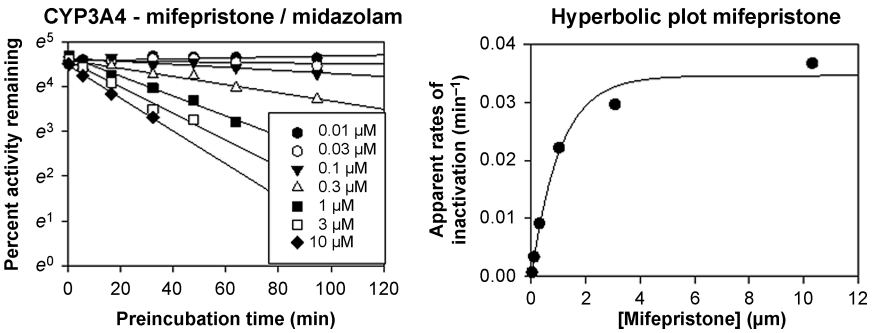


**FIGURE 3-4** Inactivation time course of CYP3A4 by mifepristone and hyperbolic: plot of apparent first-order rates of inactivation as a function of mifepristone concentrations. Mifepristone was serially diluted (0.01 to 30 μM) and incubated in the presence of an NADPH-regenerating system. Aliquots were removed for each concentration level at specific time points (0, 5, 15, 30, 45, 60 and 90 min) and transferred to a secondary reaction mixture containing 10 μM midazolam ($5\times K_m$) to assay for residual activity. Samples were normalized for each time point to vehicle incubates and plotted as the natural log of residual activity as a function of time. Apparent first-order rates of inactivation were derived from the slopes of the decay curve for each concentration and were plotted as a function of inactivator concentration to yield a hyperbolic inactivation plot. The maximal rate of inactivation, $k_{\text{inact}}$ ($=0.035\,\text{min}^{-1}$), was determined from the zero-order kinetic plateau of the hyperbolic curve and the inactivation constant, $K_{\text{I}}$ ($=0.74\,\mu\text{M}$) was calculated as the inactivator concentration yielding 50% of the maximal inactivation rate.

be TDIs and the cause of unexpected drug-induced liver injuries and idiosyncratic toxicities.[35,36] Because of the different nature of the interactions between TPs and DMEs, TDI in the classical sense is unlikely, although interactions affecting signaling pathways leading to time-dependent effects in the expression or activity levels of DMEs are certainly possible. In these cases, investigating the TP as a potential perpetrator would assess the likelihood of potential DDIs.

***3.2.2.3 Inhibition of Non-CYP Enzymes*** A number of additional classes of phase I enzymes play significant roles in drug metabolism. Recurrence of DDI is less frequent, and as a result, enzymes such as the flavin-containing monooxygenase (FMOs), epoxide hydrolases, and carboxylesterases have not been the source of consistent scrutiny. Their significant role in drug design is to provide alternate routes of elimination of drugs to mitigate the possibilities of CYP-related DDI. Phase II enzymes also do not tend to be involved in clinically relevant DDI. Glucuronidation and other phase II reactions tend to have higher capacity, faster turnover rates than phase I reactions and consequently do not contribute to marked changes in AUC and other PK parameters. Several significant exceptions should however be noted: the UGT1A1-catalyzed glucuronidation of bilirubin, the rate-limiting step in the clearance of bilirubin,[37] and the UGT2B7-catalyzed conversion of gemfibrozil to the acyl glucuronide conjugate, which was found to be a more potent reversible inhibitor of CYP2C8 than the parent drug.[38] Subsequently, the gemfibrozil acyl glucuronide conjugate was also identified as a strong time-dependent inhibitor of the same isoform through an uncommon CYP2C8-catalyzed oxidation of the conjugated form (phase I hydroxylation following a phase II conjugation), generating a reactive intermediate (possibly a substituted hydroquinone), which can then covalently bind to and inactivate the CYP enzyme.[39] As with the CYPs, DDIs arising from direct interactions of TPs with non-CYP enzymes such as UGTs are unlikely. However, indirect interactions via transcriptional or post-transcriptional regulation cannot be ruled excluded.

### 3.2.3 Reaction Phenotyping

Reaction phenotyping in SMD development is designed to establish if a drug candidate has the potential to be a victim drug and consequently have a higher *in vivo* DDI risk. Determining the CYP isoforms, non-CYP enzymes, and phase II reaction pathways that are involved as routes of metabolic clearance of a new molecular entity will allow the identification of potential metabolic liabilities. Phenotyping studies should therefore be designed to provide a complete scheme of the biotransformation landscape of a drug candidate. Commercially available single recombinant enzyme expression systems from bacterial, insect, or mammalian cells have facilitated determining these metabolic pathways. These assays can be routinely used to confirm involvement of specific enzymes in the catalytic conversion of a SMD. Studies are usually designed in a sequential approach as follows:

- General metabolic stability is initially determined in either microsomal or intact cell systems.

- An initial review of the major metabolites detected can subsequently trigger the selection of the possible biotransformation pathways to be investigated.
- Determination of the kinetic parameters ($K_m$, $V_{max}$) of the major metabolic pathways is carried out in microsomes by monitoring rates of metabolite formation at varying substrate concentrations. Establishing the rates of conversion and the affinity for a specific pathway will provide a better understanding of the potential for saturation of an enzyme and possible DDI consequences. This should be done in relation to the expected systemic concentration levels based on anticipated dosing regimens.
- Contributions of a specific pathway to the overall metabolic clearance are then investigated by simulated DDI scenarios; varying levels of known chemical inhibitors of the targeted pathway or inhibitory monoclonal antibodies of the specific enzyme are incubated with the substrate to monitor and quantify relative changes in the metabolic profile.

Reaction phenotyping as applied to enzyme substrates in SMD development is presumably not applicable to TP as the likelihood of similar enzyme–substrate interactions is remote. But it can be used as an important tool for defining the risk of TP-DI, with the protein acting as a potential perpetrator.

### 3.2.4 Induction and Suppression

Induction potential of compounds toward CYP1A2, 2B6, and 3A4 via the aryl hydrocarbon (AhR), constitutive androstane (CAR), and pregnane X (PXR) nuclear receptors (NRs), respectively, is routinely determined during SMD development. Significant CYP induction can result in therapeutic failure from clinical exposure of a compound outside the therapeutic window, if the respective enzymes are responsible for a significant portion of its overall metabolism and clearance. Similarly, exposure of a comedication that is also metabolized by the induced enzymes can be decreased significantly in a similar manner. Over repeated doses, autoinduction can occur by which the exposure of the drugs will decrease over time and result in nonlinear PK.[40] In this case, subtherapeutic concentrations of the drugs may result, which can lead to reduced efficacy at the site of target—that is, therapeutic failure. In addition, if metabolism via the induced CYP enzyme results in toxic or pharmacologically active compounds being produced, exaggerated pharmacological effects may be seen resulting in direct and/or indirect toxicity.

Multiple well-characterized models are available for examining CYP induction, either in a discovery screening setting (e.g., NR assays, hepatocyte cells lines) or later during drug development (e.g., human hepatocytes). A complete list of *in vitro* and *in vivo* methods used to study CYP induction, along with respective advantages and disadvantages is available.[41,42] Nuclear receptor luminescence assays such as PXR and AhR are well suited for a discovery setting because they can be easily automated and provide a first readout on potential enzyme induction. In general, there appears to be good agreement between $EC_{50}$ values generated in the nuclear receptor assays compared to those generated in human hepatocytes.[41] Use of cryopreserved or freshly

isolated human hepatocytes plated on simple multiwell plates (2D cultures) is considered the gold standard *in vitro* assay for examining potential induction of CYP enzymes via nuclear receptors.[3] It is well known that upon isolation and culture, basal levels of CYP enzymes decrease dramatically in hepatocytes owing to rapid cell dedifferentiation. Therefore these models are well suited for detecting a potential increase in enzyme expression after incubation for a predetermined time frame. Typically, hepatocytes are plated on collagen-coated microwell plates, allowed to attach, and then cultured for hours to several days before treatment with known inducers or test compounds for an additional 48–72 h. In this simple system, the cells cannot survive beyond 5 days. Use of a matrigel or collagen overlay to generate sandwich cultures has been used to extend the life of hepatocyte cultures for an additional 1–2 days. At the end of the dosing period, gene expression and enzyme activities can be assessed.[43] Kinetic constants (i.e., $EC_{50}$, $E_{max}$) generated from these studies can then be used for PBPK modeling to predict clinical effects. Although the *in vitro* to *in vivo* correlation (IVIVC) using primary hepatocytes has been relatively good, the absolute fold induction tends to be overpredicted by the *in vitro* assay.[41]

In addition to primary hepatocytes, hepatocyte-derived cell lines such as Fa2N-4, HepG2, BC2, and HepRG have been used previously for CYP induction or other mechanistic studies.[44,45] Ease of use, unlimited availability, and ability for manipulation via SiRNA or transfection has made these models attractive. However, at this time they are not considered a suitable replacement for primary human hepatocyte models because they either are not fully equivalent to primary hepatocytes in terms of enzyme expression or have not been fully validated.[46,47] For example, the Fa2N-4 cell line does not express functional levels of the CAR receptor and therefore cannot be used to assess CYP2B6.[48] Induction of other metabolic enzymes has been studied during SMD development, but more as a case-by-case basis. UGT1A1 induction can be examined in 2D hepatocytes alongside the CYPs. However, in this system induction of other UGT isoforms is generally modest.

Preclinical *in vivo* models (e.g., rodents, dogs, primates) to understand CYP induction risk are generally used in a more mechanistic manner during SMD development, for example, to understand nonlinear PK or toxicity observed during chronic dose studies. Even for SMD, extrapolating enzyme changes in preclinical species to humans is difficult due to species differences in enzyme expression and activity.[49] Recent advances in transgenic technologies have led to the availability of humanized liver mice, which may hold promise for bridging the gap between *in vitro* human and clinical studies, however this remains to be proven. Chapter 6 provides an overview of the potential use of animal models for predicting DDI by TP. Induction potential of metabolic enzymes in the liver by TP (*in vitro* or *in vivo*) has not been previously reported. It is interesting that exposure of nonhepatic cell types, including peripheral blood mononuclear and brain parenchymal cells to proinflammatory cytokines, does result in CYP induction[50,51] The relevance of induction of metabolic enzymes in these cells, which have little metabolic capability, is currently unknown.

Generally, suppression of CYP enzyme expression (gene or activity) has not been of particular importance during SMD development. Suppression of CYP enzymes by biologics such as immunomodulators (e.g., interferon-γ; IFN-γ)

appears to be a more likely liability. The concern for enzyme suppression in the context of TP is as follows: a patient with an inflammatory disease has increased circulatory cytokines, such as interleukin 6 (IL-6) and tumor necrosis factor α (TNF-α). These cytokines have the potential capacity to decrease CYP enzyme levels through global suppression of NRs via interaction with a shared heterodimer subunit RXR and the master regulator NF-κB.[52,53] In these patients doses of narrow therapeutic index compounds such as theophylline are titrated according to the patient's metabolic capacity to achieve effective and safe drug exposures. If a patient is then treated with an anti-IL-6 or anti-TNF-α antibody, a reversal of the reduced basal CYP activity levels might occur, resulting in therapeutic failure of the coadministered SMD, via increased drug metabolism and decreased exposure and/or increased toxicity (see Chapter 11 on Tocilizumab example). In contrast, for TPs that are direct immunomodulators (e.g., IFN-γ), suppression of CYP levels might result in decreased clearance of coadministered drugs, which could result in increased drug exposure and toxicity.

Because the mechanism of CYP suppression by cytokines is likely via NRs, 2D hepatocyte cultures have been used by several groups to understand how particular cytokines, either alone or in combination, may affect the expression and/or activity of the major CYPs.[54–57] However, studies on the suppression of non-CYP enzymes by cytokines in these same models are currently lacking. The advantages, disadvantages, and open questions regarding this methodology as well as other *in vitro* methods in the context of TP-DI are addressed in detail in Chapters 5 and 13. At this time, the predictability of simple hepatocyte assays has been questioned by multiple groups.[55,58] This lack of confidence in predictability likely explains the current regulatory opinion regarding *in vitro* assessments of TP-DI (see Chapter 10). Nonetheless, further understanding of the molecular mechanisms underlying cytokine-mediated CYP suppression may help bridge the gap of IVIVC for TP-DI. For example, the suppression of CYP3A4 by IL-6 is known to occur in a time-dependent manner based on both mRNA and activity measurements[57]. Furthermore, interaction of IL-6 with its cell surface receptor (a required step to start the cascade toward CYP suppression) results in what one could consider a quasi-irreversible activation of molecular pathways. Once activated, a subsequent invagination and recycling of the receptor occurs. Based on these observations, Kenny[59] has suggested modeling the suppression of CYP mRNA using kinetic methodology similar to SMD TDI assessments (also time dependent and irreversible). By measuring the suppression of mRNA over time and using multiple concentrations of IL-6, $K_I$, and $k_{inact}$ constants can be determined and used for subsequent PBPK modeling of TP-DI risk for comedications. General strategies around model-based methods and TP-DIs are described in Chapter 4. While still early, these types of analyses show promise.

## 3.3 TRANSPORTERS

In general, compounds can cross biological membranes in several different ways: through paracellular or transcellular routes or via transporter proteins embedded in

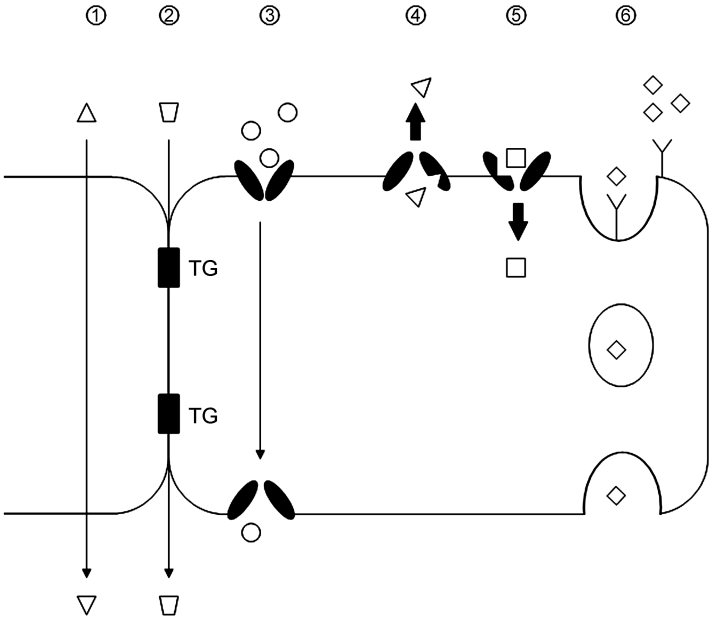


**FIGURE 3-5** Routes of translocation for pharmaceutical compounds through cellular membranes. 1. Lipid soluble compounds can permeate through the epithelium by simple passive diffusion (transcellular route). 2. Small molecular weight compounds such as $CO_2$ can permeate through the tight junctions (TG) between cells (paracellular route). 3. Facilitative transport (e.g., uptake of amino acids) occurs via a transport carrier down the concentration gradient of the solute. Transport can also occur via a carrier but against the concentration gradient of the solute. 4. In the case of efflux transporters such as P-glycoprotein, ATP-hydrolysis is the driving force for this transport. 5. Uptake carriers, such as the organic anion, and organic cation transporters can also transport substances against a concentration gradient. The driving force depends on the particular transporter but can be pH dependent or electrogenic, for example. 6. Some hormones are transported through the epithelium via receptor-mediated transcytosis. The ligand binds to a cell surface receptor, which causes an invagination in the membrane. This invagination is internalized and carried to the opposite membrane, where the ligand is then released (Reprinted with permission from Ref. 62.)

the membrane (Figure 3-5). From a pharmaceutical perspective, active efflux and uptake transporters (routes 4 and 5 in the figure) are considered most relevant with respect to SMD. Two superfamilies of transporters are particularly important: the solute carrier or SLC family and the ATP binding cassette or ABC superfamily. Within these families several dozen drug transporters have been identified, and their characteristics (substrates, inhibitors, clinical relevance) have been reviewed in detail elsewhere.[60] Recently seven of these drug transporters have been highlighted to have particular clinical relevance for DDI and have been described in publications by the FDA or consortia, such as the International Drug Transporter (ITC) consortium.[3,61] These transporters include P-glycoprotein (P-gp), breast cancer resistance protein (BCRP), organic anion transporter polypeptide 1B1 and 1B3

(OATP1B1/1B3), organic cation transporter 2 (OCT2), and organic anion transporters 1 and 3 (OAT1/3).

### 3.3.1 Substrate and Inhibitor Studies

Drug transporters are found in multiple organs throughout the body and play an important role in absorption, distribution, and overall drug clearance (metabolism and excretion).[62] Multiple organ-specific *in vitro* transporter models exist that are of varying value for understanding compound transport characteristics. These include 2D suspension hepatocytes (liver uptake), 2D hepatocyte overlay, or "sandwich" cultures (liver efflux), brain endothelial cell models (brain efflux/influx), and Caco-2 cells (intestinal permeability). These cellular systems are easy to use, readily available, contain multiple tissue-relevant transporters (along with metabolizing enzymes in some cases), and can be useful in determining whether a potential DDI liability may exist. For example, uptake of a compound into hepatocytes, which is both saturable and inhibited by statins (e.g., rosuvastatin), may indicate an OATP1B1 liability. This potential DDI risk can then be followed up mechanistically using single transporter transfected/transduced cell lines. Madin-Darby canine kidney (MDCKII), human embryonic kidney (HEK), and Chinese hamster ovary (CHO) cells are most commonly used. These polarized cells can be grown on permeable inserts to study compound flux (from the apical-to-basal and basal-to-apical directions), grown on impermeable plates to access apically located transporters only, or turned into inside-out vesicles. The latter model is useful for compounds that are poorly permeable, such as conjugated drug metabolites. As with the CYPs, these *in vitro* systems can be used to generate kinetic values for DDI risk assessment using PBPK modeling and simulation. Michaelis-Menten constants such as $K_m$ and $J_{\max}$ describe the affinity and maximal rate, respectively, of a compound for a particular transporter. Inhibition studies that generate $IC_{50}$ values can assess the risk of inhibiting transport resulting in a DDI, either as a victim or a perpetrator. Compared to the CYPs, the prediction of clinically relevant DDI involving transporters is still in its infancy. Generally, the ability of a compound to interact with a given drug transporter (as a substrate or inhibitor) is based largely on physiochemical properties (log P, MW, solubility, etc.). Therefore, it is unlikely that large TPs such as monocloncal antibodies or fusion proteins will be direct substrates or inhibitors of classical drug transporters. However, drug-transporter function can also be affected indirectly through perturbations of lipid bilayer integrity and composition, alterations in required transporter-driving forces (e.g. $Na^+/K^+$ ATPase, pH gradient) or alterations in proper membrane trafficking.[63,64] Whether these mechanisms will be of importance in the context of TP-DI is yet to be determined.

### 3.3.2 Induction and Suppression

The clinical significance of induction of transporters, in contrast to CYPs, is poorly understood, and difficult to replicate *in vitro*. Although regulation of transporters occurs via nuclear receptors such as PXR and CAR, the duplicative cross-talk that

occurs between these NRs and others such as the farnesoid X receptor and AhR is still yet to be fully explored. At present, there are no clinical examples of a TP that has caused an increase in transporter expression or function directly. However, similar to the CYPs, there is a theoretical concern that suppression of transporter expression may occur, particularly for immunosuppressant compounds. The consequence of this suppression depends largely on the characteristics of the victim SMD. For example, for an anionic drug cleared exclusively via the kidney, a suppression of OAT1 or OAT3 could decrease renal excretion and therefore increase exposure and potential drug toxicity.[65] On the other hand, for a cationic compound, this suppression may be of no consequence. Unlike the CYPs, for which the major focus of possible suppression has been centered mainly on the liver, in the case of transporters changes in function or expression in multiple organs may affect the overall ADME of a coadministered compound, including the gut (absorption), kidney (elimination), and brain (central nervous system distribution).

Thus far, work around drug transporter suppression in the liver via cytokines has been undertaken in preclinical species such as rodents[66,67] and, more recently, in primary human hepatocytes using methodologies similar to those used for CYPs.[68,69] Known species differences in transporter expression and affinities may complicate interpretation of rodent generated data. In addition, sandwich hepatocyte cultures will be more appropriate for examining canalicular transporters, such as P-gp and BCRP, because the correct orientation of those transporters is lost in simple 2D hepatocyte cultures. Because the underlying mechanism for transporter suppression may be similar to that of the CYPs, presumably one could apply similar $K_I/k_{inact}$-type PBPK analyses for TP-DI prediction. However, the inherent complexity of drug transporter regulation and the propensity for multifaceted interactions involving metabolism/transporter interplay might make these seemingly simplistic models difficult to interpret. A significant amount of research is certainly needed in this area.

## 3.4 CONCLUSION

In the SMD arena, a robust toolbox of *in vitro* and preclinical models have been developed to derisk many DDI liabilities and decrease the number of clinical DDI studies required for a given compound. Nonetheless, even for small molecules the predictability of some DDI assessments (e.g., transporter induction and transporter-metabolism interplay) remains difficult owing to a lack of well-validated models and a poor understanding of the molecular mechanisms. Although it seems logical to apply decades of knowledge and experience from the small molecule world to TP-DI assessment, it is not surprising that the inherent differences between these two types of compounds have made this challenging. As our understanding of the underlying mechanisms driving TP-mediated interactions increases and as technology platforms evolve and improve (both *in silico* and *in vitro*), accurate prediction of clinically relevant TP-DIs will, we hope, become possible. In the meantime, there is much work to do to make this a practical reality.

## REFERENCES


1. Huang SM, Zhao H, Lee JI, et al. Therapeutic protein-drug interactions and implications for drug development. *Clin Pharmacol Ther* **87**(4), 497–503 (2010).
2. Lee JI, Zhang L, Men AY, et al. CYP-mediated therapeutic protein-drug interactions: clinical findings, proposed mechanisms and regulatory implications. *Clin Pharmacokinet* **49**(5), 295–310 (2010).
3. US Food and Drug Administration, US Department of Health and Human Services, Center for Drug Evaluation and Research, Center for Biologics Evaluation and Research (CBER). Guidance for industry: drug interaction studies–study design, data analysis, and implications for dosing and labeling. Silver Spring, MD, 2012. Available at www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM292362.pdf.
4. European Medicines Agency. Guideline on the investigation of drug interactions. London, 2010. Available at www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2010/05/WC500090112.pdf.
5. Abbott Laboratories. *Packet insert for Humira (Adalimumab)*. North Chicago, 2012.
6. Genentech, Inc. *Packet insert for Actemra (Tocilizumab)*. South San Francisco, 2011.
7. Genentech, Inc. *Packet insert for Herceptin (Trastuzumab)*. South San Francisco, 2011.
8. Bristol-Myers Squibb. *Packet insert for Nulojix (Belatacept)*. Princeton, NJ, 2011.
9. Hoffmann-La Roche. *Packet insert for Pegasys (Peginterferon α-2a)*. Nutley, NJ, 2011.
10. InterMune, Inc. *Packet insert for Actimmune (Interferon γ-1b)*. Brisbane, CA, 2007.
11. European Medicines Agency. Guideline on the clinical investigation of the pharmacokinetics of therapeutic proteins. London, 2009. Available at www.emea.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC500003029.pdf.
12. Pang KS, Rodrigues AD, Peter RM, eds. *Enzyme- and transporter-based drug-drug interactions: progress and future challenges*. New York, Springer, 2010.
13. Venkatakrishnan K, Von Moltke LL, Greenblatt DJ. Human drug metabolism and the cytochromes P450: application and relevance of in vitro models. *J Clin Pharmacol* **41**(11), 1149–179 (2001).
14. Williams JA, Hyland R, Jones BC, et al. Drug-drug interactions for UDP-glucuronosyltransferase substrates: a pharmacokinetic explanation for typically observed low exposure (AUCi/AUC) ratios. *Drug Metab Dispos* **32**(11), 1201–08 (2004).
15. Smith DA, Jones BC, Walker DK. Design of drugs involving the concepts and theories of drug metabolism and pharmacokinetics. *Med Res Rev* **16**(3), 243–66 (1966).
16. Thompson TN. Optimization of metabolic stability as a goal of modern drug design. *Med Res Rev* **21**(5), 412–49 (2001).
17. Regan SL, Maggs JL, Hammond TG, et al. Acyl glucuronides: the good, the bad and the ugly. *Biopharm Drug Dispos* **31**(7), 367–95 (2010).
18. Nelson DR. The cytochrome P450 homepage. *Hum Genomics* **4**(1), 59–65 (2009).
19. Clarke SE, Jones BC. Human cytochrome P450 and their role in metabolism-based drug-drug interactions. In Rodrigues AD, ed. *Drug-drug interactions*. 2nd ed. New York, Informa Healthcare USA, 2008: 53–85.
20. Bjornsson TD, Callaghan JT, Einolf HJ, et al. The conduct of in vitro and in vivo drug-drug interaction studies: a Pharmaceutical Research and Manufacturers of America (PhRMA) perspective. *Drug Metab Dispos* **31**(7), 815–32 (2003).

21. Segel IH. Simple inhibition systems. In Segel IH, ed. *Enzyme kinetics: behavior and analysis of rapid equilibrium and steady-state enzyme systems*. New York, Wiley Interscience, 1975: 100–60.
22. Houston JB, Galetin A. In vitro techniques to study drug-drug interactions of drug metabolism: cytochrome P450. In Pang KS, Rodrigues AD, Peter RM, eds. *Enzyme- and transporter-based drug-drug interactions: progress and future challenges*. New York, Springer; 2010: 169–215.
23. Galetin A, Clarke SE, Houston JB. Quinidine and haloperidol as modifiers of CYP3A4 activity: multisite kinetic model approach. *Drug Metab Dispos* **30**(12), 1512–22 (2002).
24. Galetin A, Clarke SE, Houston JB. Multisite kinetic analysis of interactions between prototypical CYP3A4 subgroup substrates: midazolam, testosterone, and nifedipine. *Drug Metab Dispos* **31**(9), 1108–16 (2003).
25. Atkins WM. Non-Michaelis-Menten kinetics in cytochrome P450-catalyzed reactions. *Annu Rev Pharmacol Toxicol* **45**, 291–310 (2005).
26. Clarke SE. In vitro assessment of human cytochrome P450. *Xenobiotica* **28**(12), 1167–202 (1998).
27. Houston JB, Kenworthy KE. In vitro-in vivo scaling of CYP kinetic data not consistent with the classical Michaelis-Menten model. *Drug Metab Dispos* **28**(3), 246–54 (2000).
28. Houston JB, Kenworthy KE, Galetin A. Typical and atypical enzyme kinetics. In Fisher M, Lee J, Obach RS, eds. *Drug metabolizing enzymes: cytochrome P450 and other enzymes in drug discovery and development*. Lausanne, Switzerland, Fontis Media, 2003, 211–54.
29. Silverman RB. Mechanism-based enzyme inactivators. *Methods Enzymol* **249**, 240–83 (1995).
30. Masubuchi Y, Horie T. Toxicological significance of mechanism-based inactivation of cytochrome P450 enzymes by drugs. *Crit Rev Toxicol* **37**(5), 389–412 (2007).
31. O'Donnell JP, Dalvie DK, Kalgutkar AS, Obach RS. Mechanism-based inactivation of human recombinant P450 2C9 by the nonsteroidal anti-inflammatory drug suprofen. *Drug Metab Dispos* **31**(11), 1369–77 (2003).
32. Yang J, Jamei M, Yeo KR, et al. Kinetic values for mechanism-based enzyme inhibition: assessing the bias introduced by the conventional experimental protocol. *Eur J Pharm Sci* **26**(3–4), 334–40 (2005).
33. Lim HK, Duczak N Jr, Brougham L, et al. Automated screening with confirmation of mechanism-based inactivation of CYP3A4, CYP2C9, CYP2C19, CYP2D6, and CYP1A2 in pooled human liver microsomes. *Drug Metab Dispos* **33**(8), 1211–19 (2005).
34. Chen Y, Liu L, Monshouwer M, Fretland AJ. Determination of time-dependent inactivation of CYP3A4 in cryopreserved human hepatocytes and assessment of human drug-drug interactions. *Drug Metab Dispos* **39**(11), 2085–92 (2011)
35. Robin MA, Maratrat M, Le Roy M, et al. Antigenic targets in tienilic acid hepatitis. Both cytochrome P450 2C11 and 2C11-tienilic acid adducts are transported to the plasma membrane of rat hepatocytes and recognized by human sera. *J Clin Invest* **98**(6), 1471–80 (1996).

36. Kalgutkar AS, Obach RS, Maurer TS. Mechanism-based inactivation of cytochrome P450 enzymes: chemical mechanisms, structure-activity relationships and relationship to clinical drug-drug interactions and idiosyncratic adverse drug reactions. *Curr Drug Metab* **8**(5), 407–47 (2007).
37. Kang TW, Kim HJ, Ju H, et al. Genome-wide association of serum bilirubin levels in Korean population. *Hum Mol Genet* **9**(18), 3672–78 (2010).
38. Shitara Y, Hirano M, Sato H, Sugiyama Y. Gemfibrozil and its glucuronide inhibit the organic anion transporting polypeptide 2 (OATP2/OATP1B1:SLC21A6)-mediated hepatic uptake and CYP2C8-mediated metabolism of cerivastatin: analysis of the mechanism of the clinically relevant drug-drug interaction between cerivastatin and gemfibrozil. *J Pharmacol Exp Ther* **311**(1), 228–36 (2004).
39. Ogilvie BW, Zhang D, Li W, et al. Glucuronidation converts gemfibrozil to a potent, metabolism-dependent inhibitor of CYP2C8: implications for drug-drug interactions. *Drug Metab Dispos* **34**(1), 191–97 (2006).
40. Jana S, Paliwal J. Molecular mechanisms of cytochrome p450 induction: potential for drug-drug interactions. *Curr Protein Pept Sci* **8**(6), 619–28 (2007).
41. Hewitt NJ, Lecluyse EL, Ferguson SS. Induction of hepatic cytochrome P450 enzymes: methods, mechanisms, recommendations, and in vitro-in vivo correlations. *Xenobiotica* **37**(10–11), 1196–224 (2007).
42. McGinnity DF, Zhang G, Kenny JR, et al. Evaluation of multiple in vitro systems for assessment of CYP3A4 induction in drug discovery: human hepatocytes, pregnane X receptor reporter gene, and Fa2N-4 and HepaRG cells. *Drug Metab Dispos* **37**(6), 1259–68 (2009).
43. Roymans D, Annaert P, Van Houdt J, et al. Expression and induction potential of cytochromes P450 in human cryopreserved hepatocytes. *Drug Metab Dispos* **33**(7), 1004–16.
44. Gomez-Lechon MJ, Donato T, Jover R, et al. Expression and induction of a large set of drug-metabolizing enzymes by the highly differentiated human hepatoma cell line BC2. *Eur J Biochem* **268**(5), 1448–59 (2001).
45. Westerink WM, Schoonen WG. Cytochrome P450 enzyme levels in HepG2 cells and cryopreserved primary human hepatocytes and their induction in HepG2 cells. *Toxicol In Vitro* **21**(8), 1581–91 (2007).
46. Chu V, Einolf HJ, Evers R, et al. In vitro and in vivo induction of cytochrome p450: a survey of the current practices and recommendations: a Pharmaceutical Research and Manufacturers of America perspective. *Drug Metab Dispos* **37**(7), 1339–54 (2009).
47. Gerets HH, Tilmant K, Gerin B, et al. Characterization of primary human hepatocytes, HepG2 cells, and HepaRG cells at the mRNA level and CYP activity in response to inducers and their predictivity for the detection of human hepatotoxins. *Cell Biol Toxicol* **28**(2), 69–87 (2012).
48. Hariparsad N, Carr BA, Evers R, Chu X. Comparison of immortalized Fa2N-4 cells and human hepatocytes as in vitro models for cytochrome P450 induction. *Drug Metab Dispos* **36**(6), 1046–55 (2008).
49. Kraynov E, Dowty ME, Fahmi OA, Fields O. Animal models for evaluation of drug-drug interaction potential of biotherapeutics. *Curr Drug Metab* **13**(7), 947–950 (2012).

50. Tindberg N, Baldwin HA, Cross AJ, Ingelman-Sundberg M. Induction of cytochrome P450 2E1 expression in rat and gerbil astrocytes by inflammatory factors and ischemic injury. *Mol Pharmacol* **50**(5), 1065–72 (1996).
51. Liptrott NJ, Penny M, Bray PG, et al. The impact of cytokines on the expression of drug transporters, cytochrome P450 enzymes and chemokine receptors in human PBMC. *Br J Pharmacol* **156**(3), 497–508 (2009).
52. Zordoky BN, El-Kadi AO. Role of NF-kappaB in the regulation of cytochrome P450 enzymes. *Curr Drug Metab* **10**(2), 164–78 (2009).
53. Yang J, Hao C, Yang D, et al. Pregnane X receptor is required for interleukin-6-mediated down-regulation of cytochrome P450 3A4 in human hepatocytes. *Toxicol Lett* **197**(3), 219–26 (2010).
54. Aitken AE, Morgan ET. Gene-specific effects of inflammatory cytokines on cytochrome P450 2C, 2B6 and 3A4 mRNA levels in human hepatocytes. *Drug Metab Dispos* **35**(9), 1687–93 (2007).
55. Dickmann LJ, Patel SK, Rock DA, et al. Effects of interleukin-6 (IL-6) and an anti-IL-6 monoclonal antibody on drug-metabolizing enzymes in human hepatocyte culture. *Drug Metab Dispos* **39**(8), 1415–22 (2011).
56. Dallas S, Sensenhauser C, Batheja A, et al. De-risking bio-therapeutics for possible drug interactions using cryopreserved human hepatocytes. *Curr Drug Metab* **13**(7), 923–929 (2012).
57. Dickmann LJ, Patel SK, Wienkers LC, Slatter JG. Effects of Interleukin 1beta (IL-1beta) and IL-1beta/interleukin 6 (IL-6) Combinations on drug metabolizing enzymes in human hepatocyte culture. *Curr Drug Metab* **13**(7), 930–937 (2012).
58. Girish S, Martin SW, Peterson MC, et al. AAPS workshop report: strategies to address therapeutic protein-drug interactions during clinical development. *AAPS J* **13**(3), 405–16 (2011).
59. Kenny J.Cytokine mediated regulation of human cytochrome P450s—can we translate in vitro evaluation to clinical relevance? Paper presented at the AAPS/National Biotechnology Conference, San Francisco, CA, May 17, 2011.
60. You G, Morris ME, eds. *Drug transporters: molecular characterization and role in drug disposition*. Hoboken, NJ, Wiley, 2007.
61. Giacomini KM, Huang SM, Tweedie DJ, et al. Membrane transporters in drug development. *Nat Rev Drug Discov* **9**(3), 215–36 (2010).
62. Dallas S, Ward P. Drug Transporters and ADME. In Swarbrick J, ed. *Encyclopedia of Pharmaceutical Technology*. 4th ed. New York, Informa Healthcare, in press 2012.
63. Meyer dos Santos S, Weber CC, Franke C, et al. Cholesterol: coupling between membrane microenvironment and ABC transporter activity. *Biochem Biophys Res Commun* **354**(1), 216–21 (2007).
64. Akhtar N, Ahad A, Khar RK, et al. The emerging role of P-glycoprotein inhibitors in drug delivery: a patent review. *Expert Opin Ther Pat* **21**(4), 561–76 (2011).
65. Hagos Y, Wolff NA. Assessment of the role of renal organic anion transporters in drug-induced nephrotoxicity. *Toxins (Basel)* **2**(8), 2055–82 (2010).
66. Hartmann G, Cheung AK, Piquette-Miller M. Inflammatory cytokines, but not bile acids, regulate expression of murine hepatic anion transporters in endotoxemia. *J Pharmacol Exp Ther* **303**(1), 273–81 (2002).

67. Geier A, Dietrich CG, Voigt S, et al. Cytokine-dependent regulation of hepatic organic anion transporter gene transactivators in mouse liver. *Am J Physiol Gastrointest Liver Physiol* **289**(5), G831–41 (2005).
68. Diao L, Li N, Brayman TG, Hotz KJ, Lai Y. Regulation of MRP2/ABCC2 and BSEP/ABCB11 expression in sandwich cultured human and rat hepatocytes exposed to inflammatory cytokines TNF-α, IL-6, and IL-1β. *J Biol Chem* **285**(41), 31185–92 (2010).
69. Le Vee M, Jouan E, Stieger B, et al. Regulation of drug transporter expression by oncostatin M in human hepatocytes. *Biochem Pharmacol* **82**(3), 304–11 (2011).
70. Juang RH. Enz04 inhibition. Paper presented at the National Taiwan University, 2004. Available at http://juang.bst.ntu.edu.tw/BCbasics.

# MODEL-INDEPENDENT AND MODEL-BASED METHODS TO ASSESS DRUG–DRUG INTERACTIONS FOR THERAPEUTIC PROTEINS

DI WU and JEFFREY S. BARRETT

## 4.1 INTRODUCTION

Therapeutic proteins (TPs) are playing a greater role in pharmacotherapy due to their high potency and mechanism of action. The main therapeutic protein classes include monoclonal antibodies, interferons, interleukins, hormones such as growth hormone and insulin, colony-stimulating factors, and blood clotting factors. TP therapies have been applied in a variety of disease areas, such as oncology, autoimmune disorders, infectious diseases, and hormone deficiencies.

The value of the global market for TPs was estimated at $93 billion in 2010, with a compound annual growth rate (CAGR) of 16.4% between 2002 and 2010. The market is forecasted to grow at a CAGR of 6.2% between 2010 and 2017, to reach $141.5 billion in 2017.[1] Therefore, TPs are expected to occupy a significant portion of the commercial portfolios of many large pharmaceutical companies. Likewise, their presence on drug formularies and in the marketplace in general necessitates a greater understanding for both the lay and the scientific communities.

By the year 2050, 34.6% of the world population will be aged more than 65 years.[2] Polypharmacy is frequently encountered when treating senior patients, resulting in a high drug–drug interaction (DDI) potential. Unexpected DDIs can lead to reduced efficacy or toxicity. Because the approval of human recombinant insulin as the first biologic by the FDA in 1982,[3] TPs have been used clinically. With the increasing use of TPs, TP-drug interactions (TP-DIs) are more prevalent, and public awareness has increased.

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*,
Edited by Shayne C. Gad
Copyright © 2013 John Wiley & Sons, Inc.

## 4.2 TP-DIs

Because TP-DI research is still a new and emerging area, limited data are available from the drug development of TPs. The interactions of TPs with small molecule drugs (SMDs) and other TPs[4] are likewise based on sparse experience. The basic mechanisms behind TP-DIs certainly need to be further investigated.

### 4.2.1 Therapeutic Target and Window

Monoclonal antibodies (mAbs) are one of the most important classes of TPs. They have been approved in a variety of indications, but most are developed to treat cancer and autoimmune diseases. A distinguishing feature of the class is that mAbs can specifically bind to therapeutic targets (i.e., target cells or proteins). Binding may then stimulate a patient's immune system to attack those targets. The therapeutic targets can be identified as either soluble or cellular. The mAbs targeting cellular antigens exhibit complex and nonlinear PK, whereas mAbs that bind to soluble antigens usually have linear PK over a wide dose range with clearance usually assessed as dose independent. In most cases, the clearance of cellular antigen-binding mAbs decreases with increasing dose in a low dose range, stabilizing to a constant value after the target-mediated, nonlinear elimination pathway is saturated. Thus it is important to select appropriate dose(s) for a DDI study of a mAb that binds to cellular antigen.[3]

Similar to small molecules, TPs with a wide therapeutic range would not present a safety concern in regard to DDI. A safety margin of 10 or greater is empirically considered a wide therapeutic window for a TP.[3]

## 4.3 *IN VITRO* AND *IN VIVO* APPROACHES FOR EVALUATING TP-DI

*In vitro* and *in vivo* approaches for evaluating DDI for TPs can be grouped into model-independent and model-based methods. The strategy to design and conduct a TP-DI study should be based on the clearance mechanisms and biological mechanisms of action.[5] Multiple doses of small molecules are sometimes required for testing TP-DIs because small molecules usually have much shorter half-lives than those of TPs.[6] Compared with small molecule DDI evaluation, a scientifically based and operationally feasible strategy and design should be incorporated into TP-DI exploration and investigation. Selection of interacting drugs should be based on clinical probability of concomitant utilization, the therapeutic range of the small molecule drug, and the pharmacokinetic (PK) and pharmacodynamic (PD) characteristics of the drugs.

The FDA published draft guidance for DDIs for industry, including a section on TP-DI, in February 2012.[7] This draft guidance provides instructions for *in vitro* and *in vivo* studies on TP-DIs. However, case-by-case-based TP-DI studies are commonly designed and evaluated. Risk-based strategies for TP-DI evaluation are extensively executed in the pharmaceutical industry in the United States.[3,5]

Model-independent methods include *in vitro* screening, dedicated DDI studies, and "cocktail" DDI studies. Model-based methods include population PK/PD-based DDI studies.

### 4.3.1 *In Vitro* Screening

Currently, human-cultured hepatocyte assay is considered as the most commonly used preclinical approach to assess the effects of TPs on CYP enzymes and transporters using clinically relevant concentrations of TPs or physiologically relevant concentration of cytokine/cytokine modulators.[3,8,9] This method allows for direct assessment of the effect of TPs on CYP isozymes and drug transporters. It has been suggested that preclinical evaluations should not be required if a TP drug candidate is one of the cytokines or cytokine modulators with reported effects on cytochrome P450 (CYP) isozymes and drug transporters,[10] or if it is a mAb directed against one of those cytokines and/or their receptors. This class of TP drug candidates should be directly assessed in a clinical DDI study.[8,10] However, if the TP drug candidate is one of the cytokines or cytokine modulators with no reported effects on CYP isozymes and transporters,[10] or it is a mAb directed against one of those cytokines and/or their receptors, using a human cultured hepatocyte assay is recommended for testing the TP-DI propensity with a minimum of three different hepatocyte donors.[3,8] No preclinical studies are generally required if a TP drug candidate is not cytokine related.

An *in vitro* cytokine release assay has been used to evaluate cytokine release potentials from a TP drug candidate. Because healthy human blood is applied, the assay is limited to the safety assessment of TP immunomodulatory capacity and results from the assay are difficult to translate to the *in vivo* setting where TP targets are located in the body outside the circulation.[8]

The *in vitro* approach needs to be optimized further to ensure both qualitative and quantitative evaluation of TP-DIs. At this point, all results from *in vitro* screening should be viewed qualitatively. Furthermore, *in vitro–in vivo* correlation for TP-DIs has not been established and needs further investigation. In addition, *in silico* translational prediction built on a good understanding of the basic mechanisms underlying TP-DIs is still under development.[5]

A good attempt has been made to predict the human PK of a mAb exhibiting target-mediated drug disposition (TMDD).[11] System parameters were obtained from *in vitro* experiments or literature, such as binding affinity ($k_{on}$ and $k_{off}$), internalization of the drug–target complex ($k_{int}$), target degradation rate ($k_{deg}$), and target abundance ($R_o$), incorporated into a TMDD model and developed in a population-based approach to predict human PK using monkey PK data as well. The modeling and simulation results showed that prediction of human PK was within one- to two-fold of the observed data from a phase I study. This well-designed study illustrates the excellent combination of *in vitro* experimental methods with population-based modeling to predict clinical data in a clinically relevant range.

Animal models are rarely used in the assessment of TP-DIs at present. One limitation of animal models used in small molecule drug development is the fact of

highly species-specific immune tolerance. Because tolerance is highly species specific and good correlations between animals and humans are hardly established, animal models have so far been rarely used in drug development of TPs and evaluation of TP-DIs.[12]

### 4.3.2 Clinical DDI Investigations

Clinical DDI studies are recommended to be conducted in patients, not healthy subjects, for several reasons. Firstly, PK differences of TPs may exist between healthy subjects and patients, and secondly, there is the potential of toxicity of some TPs and small molecule drugs, making it unethical to conduct the studies in healthy subjects. More important, CYP isozyme levels may have been down-regulated in chronic inflammatory diseases, such as rheumatoid arthritis and some infectious diseases. If TP-DIs are mediated through CYP isozymes, anticytokine effects of TPs would normalize the down-regulated CYP enzyme activities or expressions and thereby alter the clearance of coadministered small molecule drugs.[5,10]

***4.3.2.1 Dedicated Clinical DDI Studies*** A dedicated clinical DDI study is often applied for the reliable evaluation of TP-DIs. Common study designs for dedicated clinical DDI studies include parallel, crossover (TPs with short half-lives), and single-sequence crossover (TPs with long half-lives; small molecule drug PK during run-in period compared with small molecule drug PK after single or multiple doses of TP).[8,13]

In a pharmacokinetic study of repeated subcutaneous and intravenous administration of golimumab in patients with rheumatoid arthritis, the effect of concomitant administration of methotrexate (MTX) was examined in an open-label, randomized study.[14] A parallel design was used in the clinical DDI study of golimumab. Golimumab pharmacokinetics was described using noncompartmental analysis (NCA). In the group receiving subcutaneous golimumab administration, the geometric means of golimumab CL/F in patients cotreated with MTX and patients not cotreated with MTX were 13.9 and 21.2 mL/d/kg, respectively. Compared with the patients not cotreated with MTX, the mean $C_{max}$ and $AUC_{0-28d}$ at steady state were 24% and 37% higher, respectively, in patients cotreated with MTX. In this example, NCA is used efficiently to evaluate TP-DI by comparing $C_{max}$, AUC, and CL/F in the two scenarios. NCA requires a sufficient number of samples collected especially in the terminal phase and around $C_{max}$.

Another dedicated DDI study investigated the effects of concomitant tocilizumab on MTX PK in patients with rheumatoid arthritis. The study showed that the 90% CIs for mean ratios in systemic exposure ($AUC_{last}$ and $C_{max}$) of a combination therapy of tocilizumab and MTX and a monotherapy of either MTX or its metabolite 7-hydroxymethotrexate were all within the 80–125% bounds. C-reactive protein, a PD marker, returned to baseline within a week after tocilizumab injection. Therefore, tocilizumab and MTX can be administered concurrently without any concern of DDI in this population.[15]

In addition, a cross-study comparison approach can be used in TP-DI evaluations involving PK data comparison between healthy subjects and patients or historical data. Basiliximab clearance was significantly lower when co-treated with azathioprine ($29 \pm 14$ mL/h) or mycophenolate mofetil ($18 \pm 8$ mL/h) when compared with its established clearance of $37 \pm 15$ mL/h. Population-based nonlinear mixed-effect models were used to derive all the PK parameters used in the comparison.[16] Another example of using historical data for TP-DI study comparisons is a study of trastuzumab emtansine and pertuzumab.[17] However, results generated from such comparisons should be confirmed in specific DDI studies in the same patient population.[4] As another approach, a crossover design can be used to evaluate the effect of a coadministered drug on a TP. TP PK after cotreatment with another TP or a small molecule drug in patients can be compared with TP PK obtained in a first-in-patient TP monotherapy study. The most rigorous approach for assessing TP-DI is to implement a within-patient comparison. Careful interpretation should be given on the possible interference of antidrug antibodies resulting from the crossover design.[8]

***4.3.2.2 Cocktail Clinical DDI Study*** The "cocktail" clinical DDI study is designed to screen the effect of a TP on multiple CYP isozymes via the use of corresponding prototypical CYP probes. Compared to administration of only a single probe in a dedicated DDI study, multiple CYP probes are simultaneously administered in a cocktail DDI study to identify several potential CYP isoenzymes involved in TP-DIs. The cocktail used should be safe and use validated probe drugs for specific enzymes. Ideally, only one enzyme should be involved in the elimination of the probe drug. But if the second enzyme is metabolizing the same probe drug, its contribution to the total clearance of the very probe drug should be less than 10%. The probe drugs used should be validated by the applicant or published in the literature. In cocktail DDI studies, doses used for probe drugs should be validated both individually and in combination to ensure no interference in the pharmacokinetic interaction(s) among them. In any DDI studies, dose and duration of the investigational drug should be sufficient to estimate the maximum inhibition and/or induction achieved at a clinically relevant dose.[7,18,19] Though cocktail DDI studies are sometimes perceived as less optimal than studies testing DDIs individually, they are an efficient approach to gather more information in a single study.[3] Principles and different designs of cocktail DDI studies have been reviewed.[20,21] (see also Chapter 7).

Few TP cocktail DDI studies have been published at present. It is still questionable if the design and experimental conditions used for TPs would be the same or similar as those used for a small-molecule cocktail study. One example in cocktail TP-DI studies has been reported in a study assessing the effect of IFNα-2b on CYP isozymes.[22] High-dose IFNα-2b (HDI) is used as a standard adjuvant therapy for the treatment of high-risk melanoma. The potential effect of HDI on CYP isozymes was evaluated in 17 patients with high-risk melanoma. A Pittsburgh cocktail was used in this study with a single-sequence crossover design. A CYP isozyme activity was measured using the index derived by quantifying a selectively isozyme probe drug (and/or its metabolite) which were given at certain points of time. The results

showed that HDI had substantial effects on some CYP isozymes (e.g., CYP1A2 by 60% decrease) and no effect on other isozymes (e.g., CYP2E1). A significant association has been found between the degree of CYP inhibition and the occurrence of side effects in HDI therapy, indicating a pathophysiological basis of the HDI-induced toxicity.

TP-DI studies are preferentially conducted early in the drug development process (e.g., during phase Ib/II) because of the relatively small study size needed. In addition, early assessment of DDIs by rigorous scientific design can provide safety information and development strategies to ensure the success of late-stage drug development.

### 4.3.3 Model-Based Method: Population PK/PD–based DDI Study

Population PK/PD modeling plays a significant role in drug development of TPs. Population PK/PD modeling can be a screening tool to assess and identify TP-DIs initially, which can then be confirmed in subsequent formal DDI studies if deemed necessary. It can also be used to explore whether a TP is a perpetrator or a victim for a TP-DI if both TP and small molecule drug are sampled simultaneously.[3]

The structure and mechanism of action of mAbs are closely linked to the systemic exposure and pharmacological response of mAbs. PK/PD modeling on mAbs has been used to elucidate the mechanisms of action and to optimize dose regimens.[23–25]

Both NCA and population PK approaches have their advantages and limitations. Population PK modeling aims to develop models to describe drug disposition and obtain insights on mechanisms of drug actions when linked with PD as well as assess covariate effects. NCA allows the estimation of terminal half-lives and exposure statistics. Population PK modeling can incorporate data more efficiently (e.g., sparse or rich data as well as data from different occasions) and provide uncertainty measures for the estimated PK parameters. Certainly NCA requires less effort to perform and generate results in a much faster manner. However, population PK modeling approach provides advantages over NCA when sparse samples are present and covariate effects or mechanisms are needed to explore in the target population.

In the DDI study of MTX and infliximab (IFX) in ankylosing spondylitis patients, population PK modeling was used to estimate population parameters and evaluate the MTX effect on final parameters.[26] A parallel design was used in this investigation. MTX cotreatment did not influence any population PK parameter when examined as a covariate in the established IFX population model. Furthermore, no significant difference of $AUC_{18}$ and AUE (area under the curve of PD marker) was observed between the IFX and the IFX + MTX groups.

## 4.4 BIOANALYTICAL CONSIDERATIONS

Most TP concentrations are quantified by ligand-binding assays (LBAs) in designated biological matrices. Specificity and selectivity issues are the bioanalytical

challenges when using LBA. Specificity and selectivity of an assay for a TP of interest should be assessed during method validation of LBA.[8,27] LBA capture and detection reagents are crucial to determine assay specificity for free, partially bound, fully bound, or total antibodies as well as specificity for free and total targets. Existing metabolites and inactive species in the biological matrices might interfere with the determination of the TP concentrations and lead to overestimation of the assay results. It is critical to understand the mechanisms of the assays for a better interpretation of study results in regard to data reliability and limitations. The design of LBA bioanalytical methods (free or total concentrations) should be conducted according to the aims of the applications.[28]

In a recent study of identifying the discrepancy between enzyme-linking immunosorbent assay (ELISA) and liquid chromatography tandem mass spectrometry (LC-MS/MS) assay results of a PEGylated scaffold protein in postdose monkey plasma samples, the presence of antidrug antibodies specific to the engineered antigen-binding region of this scaffold protein affected the ability of drug to bind to the target antigen used in the ELISA assay.[29] Thus the ELISA method measured only the active circulating PEGylated scaffold protein, whereas the LC-MS/MS method measured the total circulating drug. LC-MS/MS results exerted an approximately 1.53-fold higher total exposure ($AUC_{tot}$) than the ELISA results. The bioanalysis of TPs can be compromised by drug-binding endogenous components or ADAs in the incurred samples. Therefore, good understanding of the principles of different bioanalytical methods is essential to select the appropriate approaches for specific TPs.

## 4.5 CONCLUSION

Individualized and iterative approaches for evaluation of TP-DIs should clearly be implemented in the TP drug development program. With a thorough understanding of the mechanisms of actions of TPs, TP-DI studies can be scientifically assessed, designed, and conducted. Thus a risk-based strategy is expected to guide the appropriate design and execution of TP-DIs. The risk-based strategy takes in account an evaluation of therapeutic target (cellular or soluble antigens), therapeutic window (wide or narrow), and mechanisms of DDI (*in vitro* and *in vivo*) of a particular TP of interest. Once the risk factors are evaluated and identified, the appropriate approach can be selected to design the DDI study (parallel, single-sequence crossover, or crossover) and analyze the data (NCA and/or population PK-based modeling).[3]

Model-independent and model-dependent approaches are the two main methods used to evaluate DDI between a TP and a small-molecule drug as well as between two TPs, and to predict DDI in different scenarios within model defined populations and scopes. Further investigations into the likely mechanisms involved in TP-based DDIs are needed. A critical next step is the early identification of DDI potential for TPs as well as a more granular understanding of those likely to be clinically relevant.

## REFERENCES


1. GBI, Research. Therapeutic proteins market to 2017—high demand for monoclonal antibodies will drive the market. GBI Research 2011.
2. De Luca d'Alessandro E, Bonacci S, Giraldi G. Aging populations: the health and quality of life of the elderly. *Clin Ter* **162**, e13-8 (2011).
3. Zhou H, Davis HM. Risk-based strategy for the assessment of pharmacokinetic drug-drug interactions for therapeutic monoclonal antibodies. *Drug Discov Today* **14**, 891–98 (2009).
4. Pang KS, Peter RM, Rodrigues AD. Drug-drug interactions: what have we learn and where are we going? In Pang KS, Peter RM, Rodrigues AD, eds *Enzyme- and transporter-based drug-drug interaction*. Springer, New York, 2010: 701–22.
5. Girish S, Martin SW, Peterson MC, et al. AAPS workshop report: strategies to address therapeutic protein-drug interactions during clinical development. *AAPS J* **13**, 405–16 (2011).
6. Seitz K, Zhou H. Pharmacokinetic drug-drug interaction potentials for therapeutic monoclonal antibodies: reality check. *J Clin Pharmacol* **47**, 1104–18 (2007).
7. US Food and Drug Administration, US Department of Health and Human Services, Center for Drug Evaluation and Research, Center for Biologics Evaluation and Research (CBER). Draft guidance for industry: drug interaction studies—study design, data analysis, implications for dosing, and labeling. Silver Spring, MD, 2012. Available at www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM292362.pdf.
8. Kraynov E, Martin SW, Hurst S, et al. How current understanding of clearance mechanisms and pharmacodynamics of therapeutic proteins can be applied for evaluation of their drug-drug interaction potential. *Drug Metab Dispos* **39**, 1779–83 (2011).
9. Zhou H, Mascelli MA. Mechanisms of monoclonal antibody-drug interactions. *Annu Rev Pharmacol Toxicol* **51**, 359–72 (2011).
10. Huang SM, Zhao H, Lee JI, et al. Therapeutic protein-drug interactions and implications for drug development. *Clin Pharmacol Ther* **87**, 497–503 (2010).
11. Luu KT, Bergqvist S, Chen E, et al. A model-based approach to predicting the human pharmacokinetics of amonoclonal antibody exhibiting target-mediated drug dispostion. *J Pharmacol Exp Ther* **341**, 702–8 (2012).
12. Baumann A. Early development of therapeutic biologics-pharmacokinetics. *Current Drug Metab* **7**, 15–21 (2006).
13. Vincenti F, Mendez R, Pescovitz M, et al. A phase I/II randomized open-label multicenter trial of efalizumab, a humanized anti-CD11a, anti-LFA-1 in renal transplantation. *Am J Transplant* **7**, 1770–77 (2007).
14. Zhuang Y, Xu Z, Frederick B, et al. Golimumab pharmacokinetics after repeated subcutaneous and intravenous administration in patients with rheumatoid arthritis and effect of concomitant methotrexate: an open-label, randomized study. *Clin Ther* **34**, 77–90 (2012).
15. Schmitt C, Kuhn B, Zhang X, et al. Tocilixumab has no clinically relevant effects on methotrexate pharmacokinetics in patients with rheumatoid arthritis. *Int J Clin Pharmacol Ther* **50**, 218–23 (2012).

16. Kovarik JM, Pescovitz MD, Sollinger HW, et al. Differential influence of azathioprine and mycophenolate mofetil on the disposition of basiliximab in renal transplant patients. *Clin Transplant* **15**, 123–30 (2001).
17. Lu D, Burris HA, Wang B, et al. Drug interaction potential of trastuzumab emtansine combined with pertuzumab in patients with HER2-positivie metastatic breast cancer. *Curr Drug Metab* **13**, 911–22 (2012).
18. Huang S-M, Temple R, Throckmorton DC, et al. Drug interaction studies: study design, data analysis, and implications for dosing and labeling. *Clin Pharmacol Ther* **81**, 298–304 (2007).
19. Turpault S, Brian W, Van Horn R, et al. Pharmacokinetics assessment of a five-probe cocktail for CYPs 1A2, 2C9, 2C19, 2D6, and 3A. *Br J Clin Pharmacol* **68**, 928–35 (2009).
20. Zhou H, Tong Z, McLeod JF. "Cocktail" approaches and strategies in drug development: valuable tool or flawed science? *J Clin Pharmacol* **44**, 120–34 (2004).
21. Fuhr U, Jetter A, Kirchheiner J. Appropriate phenotyping procedures for drug metabolizing enzymes and transporters in humans and their simutaneous use in the "cocktail" approach. *Clin Pharmacol Ther* **81**, 270–83 (2007).
22. Islam M, Frye RF, Richards TJ, et al. Differential effect of IFNα-2b on the cytochrome P450 enzyme system: a potential basis of IFN toxicity and its modulation by other drugs. *Clin Cancer Res* **8**, 2480–87 (2002).
23. Mould DR, Sweeney KR. The pharmacokinetics and pharmacodynamics of monoclonal antibodies—mechanistic modeling applied to drug development. *Curr Opin Drug Discov Devel* **10**, 84–96 (2007).
24. Mould DR, Green B. Pharmacokinetics and pharmacodynamics of monoclonal antibodies: concepts and lessons for drug development. *BioDrugs* **24**, 23–39 (2010).
25. Mould D, Frame B. Population pharmacokinetic-pharmacodynamic modeling of biological agents: when modeling meets reality. *J Clin Pharmacol* **50**, 91S–100S (2010).
26. Ternant D, Mulleman D, Lauferon F, et al. Influence of methotrexate on infliximab pharmacokinetics and pharmacodynamics in ankylosing spondylitis. *Br J Clin Pharmacol* **73**, 55–65 (2011).
27. Hamuro LL, Kishnani NS. Metabolism of biologics: biotherapeutic proteins. *Bioanalysis* **4**, 189–95 (2012).
28. Garofolo F, Rocci Jr, ML Dumont I, et al. White paper on recent issues in bioanalysis and regulatory findings from audits and inspections. *Bioanalysis* **3**, 2081–96 (2011).
29. Wang SJ, Wu ST, Gokemeijer J, et al. Attribution of the discrepancy between ELISA and LC-MS/MS assay results of a PEGylated scaffold protein in post-dose monkey plasma samples due to the presence of anti-drug antibodies. *Anal Nioanal Chem* **402**, 1229–39 (2012).

# UTILITY OF *IN VITRO* METHODS IN DRUG–DRUG INTERACTION ASSESSMENT AND PREDICTION FOR THERAPEUTIC BIOLOGICS

THERESA NGUYEN, NARENDRA KISHNANI, and RAYMOND EVERS

## 5.1 INTRODUCTION

Therapeutic proteins (TPs) are increasingly being coadministered with small molecule (SM) drugs and/or other biologics such as monoclonal antibodies (mAbs). Historically, interactions between TPs and SMs have not been considered very likely because mechanisms involved in the disposition of both modalities are fundamentally different. In the case of SMs, various drug-metabolizing enzymes and transporters expressed primarily in the intestine, liver, and kidney are involved in absorption, distribution, and elimination. For mAb, non-ligand and ligand-mediated clearance pathways are involved. In addition, binding to the neonatal Fc receptor (FcRn) functions as a salvage pathway that protects IgG molecules from lysosome-mediated degradation. In general, these differences will make it less likely that drug–drug interactions (DDIs) will occur between TPs and SMs. However, the situation may be different for TPs involved in pathways that modulate the immune system because it has been demonstrated that increased cytokine exposure can result in reduced levels of cytochrome P450 (CYP) enzymes and drug transporters.

Theoretically, the effects of cytokines on CYPs can therefore mediate DDIs between TPs and SMs when immunomodulating biologics (e.g., cytokines and cytokine modulators) are used in the treatment of cancer, infections, or inflammation. The scenarios that could lead to suppression or desuppression of enzymes have been outlined by Morgan[1] (Figure 5-1): When an individual, previously stabilized on a SM drug regimen becomes afflicted with an infection or inflammatory

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*,
Edited by Shayne C. Gad
Copyright © 2013 John Wiley & Sons, Inc.

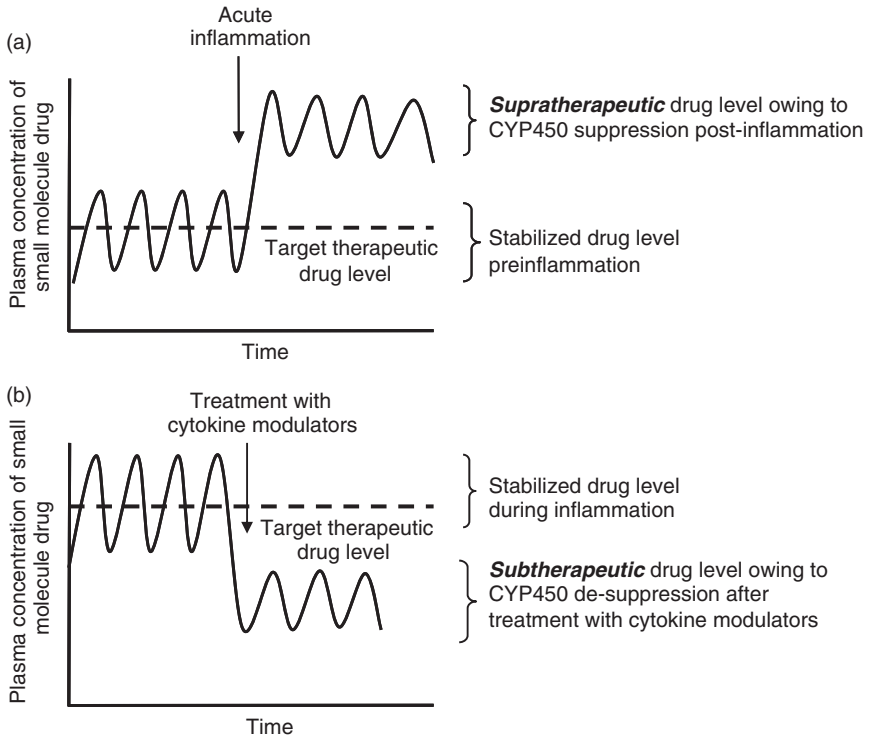


**FIGURE 5-1 Suppression and desuppression of CYP450 enzymes owing to cytokines and cytokine modulators.** (a) An individual previously stabilized on a small molecule drug regimen becomes afflicted with an infection or inflammatory response. As a result, the associated elevated levels of proinflammatory cytokines (e.g., tumor necrosis factor α, interleukin 1β, and interferon γ) may decrease levels of drug-metabolizing enzymes and potentially lead to a decline in drug clearance and supratherapeutic small molecule drug exposure. (b) In contrast, in patients with a preexisting inflammatory disease accompanied with increased levels of the proinflammatory cytokines, treatment with an anticytokine drug could result in normalization of previously suppressed drug-metabolizing enzymes levels and potentially increase drug clearance, leading to subtherapeutic exposure of small molecule drugs.

response, the associated elevated levels of proinflammatory cytokines (e.g., tumor necrosis factor α, interleukin 1β, interleukin 6, and interferon γ) may decrease the levels of drug-metabolizing enzymes and potentially lead to a decline in drug clearance and supratherapeutic SM drug exposure.[1] In contrast, in patients with a preexisting inflammatory disease accompanied with increased levels of the proinflammatory cytokines, treatment with an anticytokine drug could result in normalization of previously suppressed drug-metabolizing enzyme levels and potentially increase of drug clearance, leading to subtherapeutic exposure of SM drugs.

A classic example of metabolizing enzyme suppression owing to infection or inflammation is the administration of theophylline (a CYP1A2 substrate) to patients

with influenza, resulting in toxic accumulation of drug as a result of reduced levels of CYP1A2.[2] Reduced clearance (10–50%) of theophylline was also found in patients infected with the hepatitis B or C virus who were treated with interferon α (IFN-α) or IFN-β.[3] Recently, an example of desuppression of CYP3A4 activity was described in patients with rheumatoid arthritis (RA) treated with the anti-IL-6 receptor mAb tocilizumab.[4] It was observed that administration of simvastatin to RA patients resulted in higher plasma drug concentrations than those reported in healthy subjects. Conversely, simvastatin exposure approached normal levels after treatment with tocilizumab. Because clearance by CYP3A4 contributes substantially to the clearance of simvastatin, treatment with tocilizumab may have normalized the inflammation-mediated reduction in CYP3A4 activity.

Currently, drug–drug interactions reported for TPs as perpetrators and SM drugs as victims in general have been small, and a generalization of the extent of interactions is not feasible at this point owing to the limited availability of clinical data. Excellent reviews on the clinical examples available have been published[5–8] and therefore will not be discussed further in this chapter.

Due to the long half-life of commonly used TPs such as mAbs and the difficulty in many cases to recruit patients for trial enrollment, the conduct of well-designed clinical interaction studies is generally costly and difficult. Therefore, there is a strong desire to develop *in vitro* approaches allowing predictions of whether significant interactions between TPs and SMs are likely. Goals of this chapter are to provide an overview of the current literature on the effects of proinflammatory cytokines on absorption, distribution, metabolism, and excretion (ADME) genes *in vitro* and to provide a perspective on the current challenges and future opportunities to establish improved *in vitro* systems.

## 5.2 MECHANISMS INVOLVED IN SUPPRESSION OF DRUG-METABOLIZING ENZYMES

The data presented in Table 5-1 suggest variability in the effect of individual cytokines on CYP enzymes in hepatocytes. For example, interleukin 6 (IL-6) and IFN in general caused *pan*-suppression across most CYP isoforms analyzed. In contrast, IL-1 appears to affect only a subset of the analyzed CYP isoforms. This suggests that there may be more than one mechanism in hepatocytes that could contribute to cytokine-mediated modulation of CYP expression levels. Potentially, this could have implications for DDI potential because CYP regulation may be affected by different types of inflammatory disease.[1] The general mechanisms by which cytokines interact with metabolic enzymes and transporters are illustrated in Figure 5-2. Several reviews on the molecular mechanisms identified to date that could lead to suppression of drug-metabolizing enzymes and transporters are available.[1,24–27] The major mechanisms resulting in gene suppression are summarized below.

The transcriptional activation of CYP3A4 as illustrated in Figure 5-3 is known to be mediated by the nuclear receptor pregnane X receptor (PXR), a ligand-dependent transcription factor.[28] PXR forms a complex with its obligate partner retinoid X

**TABLE 5-1 Effects of Cytokines on Cytochrome P450 Enzymes**

| Cytokines | CYP450 Enzymes | | | | | | |
|---|---|---|---|---|---|---|---|
| | 1A1 | 1A2 | 2A6 | 2B6 | 2C8 | 2C9 | 2C11 |
| IL-1β | ↓r | ↓r | | | | | |
| | | ↓a,r | | | ↓r (2C) | | |
| | | | | ↓r | | ↓r | |
| | | | | ↔r↓p | ↓r | ↔r↓p | |
| IL-1 | | | | | | | |
| IL-2 | ↓a,r | | | ↓r(2B1/2) | | | ↓r |
| IL-4 | | ↔r | | | ↔r (2C) | | |
| IL-6 | ↔r | ↔r | | | | | |
| | | ↓a,r | | | ↓r (2C) | | |
| | | ↓a,p | | ↓a | | | |
| | | | | ↓r | ↓r | ↓r | |
| | | | | ↓p,r | ↓r | ↓p,r | |
| | ↓r | ↓a,r | | ↓r | ↓r | ↓r | |
| Oncostatin M | | ↓a,p,r | ↓a | ↓a | | | |
| TNF-α | | ↓a,r | | | ↓r (2C) | | |
| | | | | ↔r↓p | ↓r | ↔r↓p | |
| IFN-γ | | ↓a,r | | | ↔ r (2C) | | |
| | | ↓a,p | ↓a | ↓a | | | |
| | | ↓a,p | ↓a | ↓a | | | |
| | | | | ↓p,r | ↓r | ↔r↓p | |
| IFN-α | | | | | | | |
| IFN-α2b | | ↓p,r | | ↓r | | | |
| Cytokine mix[a] | | | | ↓p,r | | | |

Upregulation (↑), downregulation (↓), or no effect (↔) on activity (a), protein levels (p), and/or mRNA (r) expression; IL: interleukin; IFN: interferon; KC: Kupffer cell; NPC: nonparenchymal cells; TNF: tumor necrosis factor.

[a]TNFα, IL-1β, and IFNγ.

| 2C18 | 2C19 | 2D1 | 2E1 | 3A4 | Model | Reference |
|---|---|---|---|---|---|---|
| | | | | | Rat hepatocytes | 9 |
| | | | ↓r | ↓a,r (3A) | Human hepatocytes | 10 |
| | | | | ↓r | Human hepatocytes | 11 |
| ↔r | ↔r | | | ↓p,r | Human hepatocytes | 12 |
| | | | | ↓a | Human hepatocytes | 13 |
| | | ↓r | | ↓a,r (3A) | Rat hepatocytes | 14 |
| | | | | ↓a | Human hepatocytes–KC co-culture | 13 |
| | | | ↑r | ↔ a,r (3A) | Human hepatocytes | 10 |
| | | | | | Rat hepatocytes | 9 |
| | | | ↓r | ↓ a,r (3A) | Human hepatocytes | 10 |
| | | | | ↓a | Human hepatocytes | 15 |
| | | | | ↓r | Human hepatocytes | 16 |
| | | | | ↔ a,r | Human hepatocytes | 17 |
| | | | | ↓a (3A) | Human hepatocytes | 13 |
| ↔ r | ↓r | | | ↓p,r | Human hepatocytes | 12 |
| | ↓r | | | ↓a,r | Human hepatocytes | 18 |
| | | | | ↓a,p,r | Human hepatocytes | 15 |
| | | | ↓r | ↓a,r (3A) | Human hepatocytes | 10 |
| ↔ r | ↔ r | | | ↓r ↔p | Human hepatocytes | 12 |
| | | | | ↓r | Human hepatocytes | 19 |
| | | | ↓r | ↔r↓a (3A) | Human hepatocytes | 10 |
| | | | | ↓a,p,r | Human hepatocytes | 20 |
| | | | | ↓a,p | Human hepatocytes | 15 |
| ↔ r | ↔ r | | | ↓p,r | Human hepatocytes | 12 |
| | | | | ↓a,p,r | Human HepG2 | 21 |
| | | | ↓r | ↑p,r | Human hepatocytes–NPC co-culture | 22 |
| | | | ↓r ↔ p | ↓p,r | Human hepatocytes | 23 |

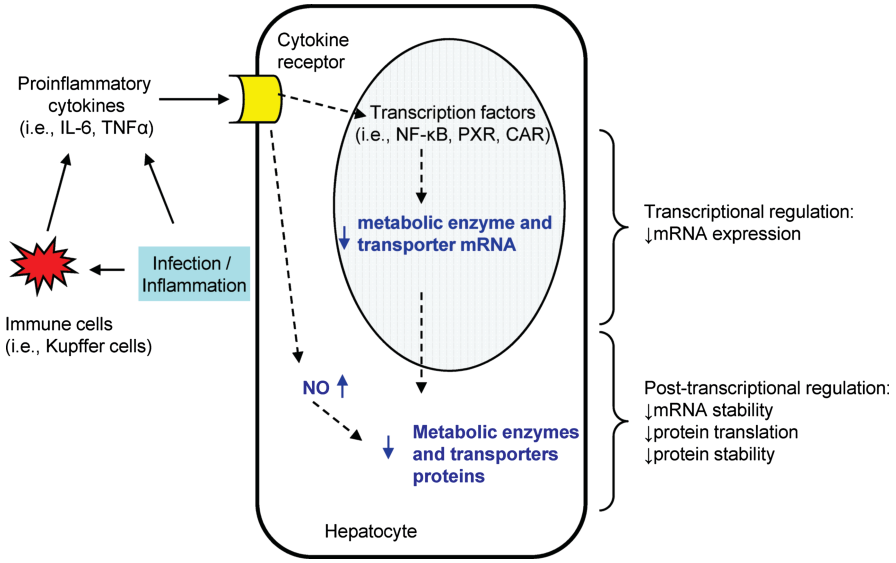


**FIGURE 5-2 Proposed general mechanisms of cytokine regulation of metabolic enzymes and transporters.** At the transcriptional level, cytokines can modulate gene expression of metabolic enzymes and transporters through interaction with various nuclear transcription factors—such as NF-kB, PXR, and CAR. At the post-transcriptional level, decreased enzyme and transporter proteins may be the result of reduction in mRNA stability owing to oxidative stress, suppression of protein translation, and/or increase in enzyme degradation due to elevated levels of NO associated with inflammation. Dotted lines: multiple intermediate steps may be involved; solid lines: a more direct effect. CAR: constitutive androstane receptor; IL: interleukin; NO: nitric oxide; PXR: pregnane X receptor; TNF: tumor necrosis factor.

receptor (RXR), and this complex in turn binds to the consensus DNA sequence elements in target genes (Figure 5-3a). PXR can be activated after binding to agonists such as rifampicin and play an important role in the induction of CYP gene transcription such as CYP3A4. In addition, it has been demonstrated that, at least in mice, PXR also plays a role in the basal transcription of Cyp3a.[29] Gu and coworkers[19] showed in human primary hepatocytes that TNF-α (2 ng/mL, 24 h) inhibited rifampicin-induced CYP3A4 mRNA levels by more than 50%. It was suggested that suppression of CYP3A4 may be mediated through the interference of the PXR-RXR-regulated pathway, and not by reduction in PXR gene transcription, because mRNA levels were unchanged in human primary hepatocytes. The researchers showed that activation of NF-κB in HepG2 cells was directly responsible for the suppression of the PXR-regulated gene expression and the presence of SRIκBα, an NF-κB super repressor, was able to reverse the suppression. In addition, it was demonstrated that NF-κB directly interacts with RXR to create a steric hindrance and disrupts the interaction of the RXR-PXR heterodimer to the PXR response element. The result is a suppression of gene transcription (Figure 5-3b).

The constitutive androstane receptor (CAR) is also involved in the transcriptional control of a number of drug-metabolizing enzymes and transporters, such as

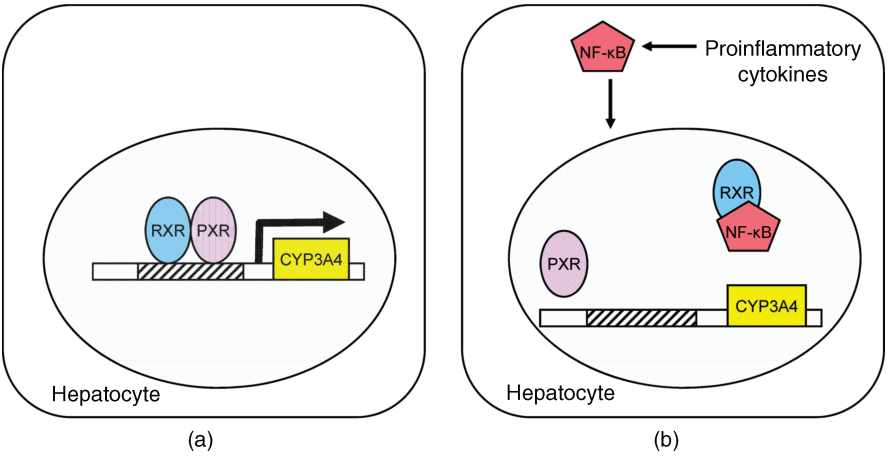


**FIGURE 5-3 Regulation of CYP3A4 by pregnane X receptor (PXR) and its obligate partner retinoid X receptor (RXR).** (a) Upon heterodimerization of PXR with RXR and co-activators, the complex binds to the consensus PXR response elements in target genes, leading to gene transcription. (b) The presence of proinflammatory cytokines will activate NF-κB, which in turn will bind to RXR and disrupt interaction of the RXR-PXR heterodimer to the PXR response element. The result is a suppression of gene transcription. It is hypothesized that the same mechanism could result in suppression of CAR-mediated transcription activity (CAR: constitutive androstane).

CYP3A4, CYP2B6, and UGT1A1.[30] Similar to PXR, CAR-mediated gene transcription also requires the formation of the CAR-RXR heterodimer complex. It is hypothesized that the same mechanism by which NF-κB interacts with the PXR-RXR complex (Figure 5-3b) may also explain the interaction of NF-κB with the CAR-RXR complex to suppress of CAR-mediated transcription activity. In addition, NF-κB may also regulate CAR-mediated transcription of drug-metabolizing enzymes and transporters through interaction with the glucocorticoid receptor (GR). Assenat and coworkers[11] demonstrated that IL-1β activation of NF-κB can inhibit the transcriptional activation of GR, resulting in diminished CAR mRNA expression and leading to reduced CAR-mediated gene transcription (i.e., CYP3A4). A similar mechanism has been proposed for the downregulation of the aryl hydrocarbon receptor (AhR), a nuclear receptor involved in the regulation of CYP1A transcription after treatment with lipopolysaccharide (LPS) or TNF-α.[31]

IFN-α activation of gene transcription is mediated through the Janus kinase–signal transducer and activator of transcription (JAK-STAT) pathway. In the hepatoma cell line HepG2, it was demonstrated that transfection with STAT1 downregulated transcription of a CYP3A4 promoter containing reporter construct, whereas overexpression of a dominant negative STAT1 partially reversed the down-regulated CYP3A4 expression.[21] Based on these results, it was concluded that components of the JAK-STAT pathway are either directly or indirectly involved in the downregulation of CYP3A4 by IFN-α. These results are in contrast to earlier data in rodents

which suggested that IFN regulation of CYP3A4 is through the induction of xanthine oxidase, which can lead to elevated reactive oxygen intermediates and result in CYP3A4 protein degradation.[32] Subsequent work, however, indicated that this would not be a plausible mechanism and that downregulation is more likely to occur at the transcriptional level.[33]

Reactive oxygen species can also mediate transcriptional regulation of drug transporters by activating NF-κB. Ros and coworkers[34] noted that TNF-α induction of Mdr1b can be facilitated through various pathways, including NF-κB and p53, transcription factors for which binding sites in the Mdr1b promoter have been identified.[35,36] The authors observed that in both rat primary hepatocytes and hepatoma H-4-II-E cells, a time-dependent stimulation of the Mdr1b expression by TNF-α was detectable. Furthermore, in rat hepatoma H-4-II-E cells expressing dominant negative p53 (rendering nonfunctional proteins), TNF-α upregulation of Mdr1b was unaffected, suggesting that induction was independent of functional p53. In addition, in transient transfection assays, Ros' group also showed that Mdr1b upregulation correlated with activation of the promoter and that mutation of the NF-κB site in the Mdr1b promoter prevented its induction by TNF-α.[34]

Collectively, these results from both rat primary hepatocytes and hepatoma H-4-II-E cells suggest that induction of the rat Mdr1b gene by TNF-α is a result of activated NF-κB signaling (by elevated reactive oxygen species) and independent of p53.

Although the majority of evidence suggests that cytokines exert their effect on drug-metabolizing enzymes at the transcriptional level, work by Morgan's group has provided evidence that post-transcriptional mechanisms can be involved as well.[1] For instance, in human and rat hepatocytes, the presence of either LPS or IL-1 has been shown to induce nitric oxide (NO) synthase, resulting in elevated NO levels, which can lead to CYP2B protein degradation.[23,37] Lee and coworkers demonstrated that in rat hepatocytes IL-1 can stimulate NO-dependent CYP2B1 degradation via increased ubiquitination and subsequent proteasome-mediated protein degradation.[37] Donato and coworkers demonstrated that in human primary hepatocytes within 24 h of IFN-γ treatment (300 U/mL), an increase in NO release (200% over untreated control) was observed in parallel with the inhibition of CYP1A2 activity and protein levels.[20] Similarly, after the addition of *S*-nitroso-*N*-acetylpenicillamine (SNAP), which spontaneously generates NO in the cell culture medium, the concentration of nitrite increased and a corresponding concentration-dependent inhibition of CYP1A2 activity was observed; the highest inhibition (70% compared to untreated cells) reached at 2 mM SNAP. Conversely, the addition of up to 1 mM of $N^G$-monomethyl-L-arginine, a competitive inhibitor of NO biosynthesis, lead to reduction in IFN-γ-induced NO formation, and partial reversal of IFN-γ inhibition of basal and 3-methylcholanthrene-induced CYP1A2 activity.

## 5.3 *IN VITRO* ASSAYS

The present landscape of available *in vitro* methodologies for examining the interactions between cytokines and cytokine modulators and the array of metabolic

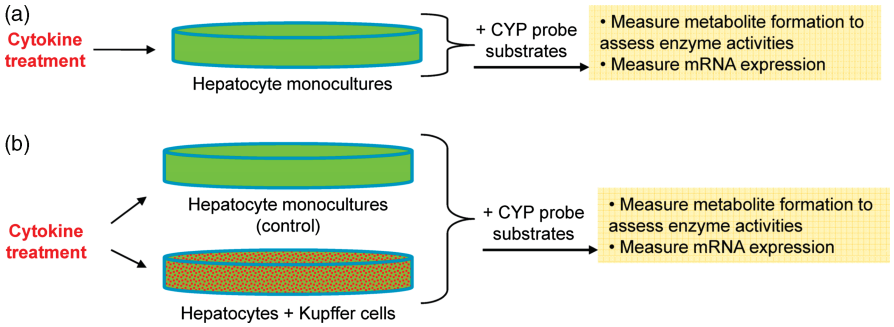


**FIGURE 5-4 General experimental design of *in vitro assays*.** Hepatocyte monocultures (a) and hepatocyte-Kupffer cell co-cultures (b) are exposed to cytokines of interest for 24 to 96 h. The cultures are then incubated with CYP probe substrates, followed by measurement of metabolite formation and mRNA levels to assess enzyme activities and gene expression.

enzymes and transporters is limited. Current practices predominantly involve the use of primary hepatocyte cultures because they are the most prevalent model employed owing to their relatively easy access and commercial availability from various preclinical species as well as from humans. In addition, plateable cryopreserved hepatocytes are commercially available, which have the advantage of reduced interassay variability. This allows an easier comparison of the effect of different test molecules than in hepatocytes freshly isolated from different donors. A general experimental design for assessing the effects of cytokines on CYP enzymes in a hepatocyte monoculture *in vitro* system is illustrated in Figure 5-4a. Hepatocyte monocultures are cultivated and then incubated with cytokines of interest for the desired period (~24–96 h). The monocultures are then exposed to CYP probe substrates, followed by measurement of metabolite formation and mRNA levels to assess enzyme activities and gene expression.

A limitation of the use of hepatocyte monocultures in studying the effect of cytokines is that typically one cytokine is studied at a time, whereas *in vivo* the liver is exposed to a wide range of immunomodulatory molecules. The question therefore is warranted whether data generated in hepatocytes are representative of *in vivo* conditions or, more important, of disease states in which the concentration of individual cytokines can be grossly altered. As a strategy to overcome this limitation, co-cultures of hepatocytes with either purified Kupffer cells[13] or nonparenchymal cell (NPC) mixtures containing immature hepatocytes, Kupffer cells, biliary cells, stellate cells, and hepatic stem cells are currently being explored.[22] These co-culture systems may offer an advantage over primary hepatocytes by providing a physiological environment that is more closely representative of an *in vivo* scenario.

A general experimental design for assessing the effects of cytokines on CYP enzymes in a hepatocyte–Kupffer cell co-culture *in vitro* system is illustrated in Figure 5-4b. Hepatocytes are co-cultured with Kupffer cells (or NPC) and then incubated with cytokines of interest for the desired period (~24–96 h).

The co-cultures are then exposed to CYP probe substrates followed by measurement of metabolite formation and mRNA levels to assess enzyme activities and gene expression.

The limited availability of *in vitro* tools for studying the impact of cytokines on liver enzymes and transporters highlights a need for more sophisticated assay systems and advocates for exploring new technologies.

## 5.4 EFFECTS OF CYTOKINES ON METABOLIZING ENZYMES AND TRANSPORTERS

### 5.4.1 IL-1 Investigations

***5.4.1.1 Effects of IL-1 on Cytochrome P450 Enzymes*** As summarized in Table 5-1, the interaction of IL-1β on CYP enzymes has been widely observed. Barker and coworkers showed in isolated rat hepatocytes that IL-1β (1–4 U/mL, 24 h) substantially reduced 2,3,7,8-tetrachlorodibenzo-*p*-dioxin-induced CYP1A1 and CYP1A2 mRNA expression by approximately 65% and 95%, respectively.[9] The inhibitory effects were dose dependent, with $IC_{50}$ at 2 and 0.5 U/mL for CYP1A1 and CYP1A2, respectively.

Similar observations of CYP enzyme suppression were made in human primary hepatocytes by numerous investigators. Abdel-Razzak and coworkers reported a reduction in activity of CYP1A2 (12–89%) by IL-1β (100 U/mL, 72 h).[10] The interaction was believed to be at the transcriptional level owing to a corresponding decrease in CYP1A2 mRNA (60–93%). IL-1β (100 U/mL, 24 h) also showed suppression of both basal levels and phenobarbital- and bilirubin-mediated induction of CYP2B6 mRNA.[11] Aitken and Morgan were unable to detect similar suppression of CYP2B6 mRNA in their experimental studies with IL-1β (5 ng/mL, 24 and 40 h) but did observe reduced CYP2B6 protein levels and a corresponding increased NO production.[12] Collectively, these results suggest that the IL-1β interaction with CYP2B6 may be mediated by various mechanisms, including alteration of gene transcription, inhibition of protein translation, and/or increased enzyme degradation associated with elevated NO production.

In terms of the CYP2C family, CYP2C8 and CYP2C9 mRNA expression was shown to be downregulated by IL-1β, and a significant reduction in CYP2C9 protein levels was also observed.[11,12] Similar to CYP2B6, a corresponding increase in NO production was detected in parallel with a reduction in CYP2C9 protein after IL-1β exposure.[12] This was not seen for CYP2C8 mRNA inhibition, however. These results suggest that IL-1β may interact with CYP2C8 at the level of gene transcription, whereas it may regulate CYP2C9 through both gene transcription and protein expression/degradation. In contrast to CYP2C8 and CYP2C9, CYP2C18 and CYP2C19 mRNA levels were unaffected by IL-1β.[12]

Analogous to the other CYP enzyme families, IL-1β also negatively affected CYP3A4 mRNA expression, protein levels, and enzyme activity. Abdel-Razzak's group reported that IL-1β (100 U/mL, 72 h) produced significant reduction in gene expression (60–93%) and CYP3A4 activity (30–90%).[10] Similarly, Assenat and

coworkers reported IL-1β suppression of both basal and phenobarbital- and bilirubin-mediated induction of CYP3A4 mRNA expression (100 U/mL, 24 h).[11] Aitken and Morgan also demonstrated significant downregulation of CYP3A4 mRNA expression (>95%) accompanied with a corresponding decrease in CYP3A4 protein after IL-1β exposure (5 ng/mL, 24 h and 40 h).[12] Thus IL-1β regulation of CYP3A4 appears to be most likely at the transcriptional level.

Assenat and coworkers hypothesized that the mechanism by which IL-1β interacts with CYP3A4 and other CYP enzymes at the level of gene transcription may potentially be through the downregulation of CAR expression.[11] This was supported by demonstrating that IL-1β activated NF-κB, which led to the downregulation of glucocorticoid receptor (GR) transactivation of the CAR promoter, resulting in suppressed CAR mRNA expression and decreased CAR-mediated gene transcription. Another potential regulatory mechanism associated with IL-1β may involve the downregulation of RXRα and PXR mRNA, leading to altered gene expression of CYP enzyme.[38]

Post-transcriptionally, NO-induced enzyme degradation may also be involved in the regulation of the CYP3A family.[37] Although mRNA levels were unaffected by IL-1β in primary rat hepatocytes, Lee and coworkers detected a 40% downregulation of CYP3A1 protein.[37] This inhibition in protein expression was reversed by NO synthase inhibitors.

***5.4.1.2 Effects of IL-1 on Phase II Metabolizing Enzymes*** Although not as extensively studied, various phase II metabolic enzymes have also been shown to also be affected by proinflammatory cytokines (Table 5-2). Abdel-Razzak and coworkers reported that in human primary hepatocytes IL-1β was a potent suppressor of various enzymes including epoxide hydrolase.[10] Assenat's group showed that IL-1β (100 U/mL, 24 h) suppressed both basal expressions and phenobarbital- and bilirubin-induced UGT1A1, GSTA1, and GSTA2 mRNA.[11] The authors suggest that similar to CYP3A4, IL-1β regulation of phase II metabolic enzymes may also involve the activation of NF-κB, leading to inhibition of GR-mediated CAR mRNA expression.

In rat primary hepatocytes, IL-1β (20–300 U/mL, 8–48 h) produces a marked decrease in basal mRNA levels of rGSTA2 and rGSTM1, with a corresponding decrease in rGSTM1 protein expression (~44%) and activity (47–78%).[39] In addition, IL-1β also suppressed 3-methylcholanthrene-, phenobarbital-, and/or oltipraz-induced expression of rGSTA2 and rGSTM1 mRNAs and proteins. Associated with these findings a shorter mRNA half-life was observed, suggesting the effects of IL-1β on GST were the result of mRNA destabilization rather than reduced gene transcription. In contrast to rGSTA2 and rGSTM1, IL-1β exposure led to increased rGSTP1 mRNA levels and had no effect on the chemically induced expression of both rGSTP1 mRNA and protein.

***5.4.1.3 Effects of IL-1 on Drug Transporters*** In addition to metabolizing enzymes, drug transporters are also affected by proinflammatory cytokines (Table 5-3). Sukhai and coworkers observed that IL-1β (0–50 ng/mL, 72 h) had no

**TABLE 5-2 Effects of Cytokines on Phase II Metabolic Enzymes**

| | Phase II Enzymes | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Cytokines | GSTA1 | GSTA2 | UGT1A1 | UGT2B7 | SULT1A1 | SULT2A1 | CES2 | Epoxide hydrolase | Model | Reference |
| IL-1β | | | | | | | | ↓r | Human hepatocytes | 10 |
| | | ↓a,p,r (rGSTA2, rGSTM1) | | | | | | | Rat hepatocytes | 39 |
| | ↓r | ↓r | ↓r | | | | | | Human hepatocytes | 11 |
| IL-2 | ↔ a | | | | | | | | Rat hepatocytes | 39 |
| IL-4 | | | | | | | | ↔ r | Human hepatocytes | 10 |
| | ↔ a | | | | | | | | Rat hepatocytes | 39 |
| | ↓r (GSTA1/A2) | | | | | | | | Human hepatocytes | 40 |
| IL-6 | | | | | | | | ↓r | Human hepatocytes | 10 |
| IL-6 | ↔ a | | | | | | | | Rat hepatocytes | 39 |
| | ↔ r (GSTA1/A2) | | | | | | | | Human hepatocytes | 40 |
| TNF-α | | | | | | | | ↓r | Human hepatocytes | 10 |
| IFN-γ | | | | | | | | ↔ r | Human hepatocytes | 10 |
| IFN-γ/α | ↔ a | | | | | | | | Rat hepatocytes | 39 |
| IFN-α2b | | | | ↓p,r | ↓ r | ↔ r | ↔ r | | Human hepatocytes-NPC co-culture | 22 |

Downregulation (↓) or no effect (↔) on activity (a), protein levels (p), and/or mRNA (r) expression; CES: carboxylesterase; GST: glutathione *s*-transferase; IL: interleukin; IFN: interferon; NPC: nonparenchymal cells; SULT: sulfotransferase; TNF: tumor necrosis factor; UGT: UDP-glucuronosyltransferase..

effect on Mdr1a or Mdr1b mRNA in primary rat hepatocytes, but P-glycoprotein (P-gp) protein expression was significantly reduced (32–75%), which also corresponded with a reduction in efflux activity.[42] Because no effects were observed on Mdr1 gene expression, the effect of IL-1β on P-gp is potentially mediated through either protein synthesis and/or protein stability in rat hepatocytes. Diao's group detected a significant decrease in levels of MRP2 protein in sandwich-cultured (SC) rat hepatocytes but not mRNA in the presence of IL-1β (1 ng/mL, 24 and 48 h).[44] They also saw downregulation of both Bsep gene expression (Abcb11 mRNA) and protein levels in the presence of IL-1β.

General IL-1β suppression of transporters was also observed in human primary hepatocytes and HepaRG cells by Le Vee and coworkers whereby mRNA expression, protein levels, and transporter activities were reduced for both organic anion transporting polypeptide (OATP) 1B1 and sodium-taurocholate co-transporting polypeptide (NTCP).[43] Analogous observations were also made by Assenat's group: IL-1β (100 U/mL, 24 h) suppressed both basal and phenobarbital- and bilirubin-induced expression of OATP1B1 mRNA expression.[11] Besides OATP1B1 and NTCP, downregulation of gene expression was also detected for other uptake transporters such as OATP1B3 and OATP2B1 and efflux pumps such as multidrug resistance protein (MRP) 2, MRP3, and MRP4 and breast cancer resistance protein (BCRP).[43] The reduction in MRP2 gene expression was similarly confirmed by Diao and coworkers as was reduced protein levels after IL-1β exposure in SC human hepatocytes.[44] It is interesting that IL-1β also diminished BSEP gene expression, but protein levels were significantly induced.[44] This finding was contradictory to the reduction in BSEP protein expression observed in HepaRG cells by Le Vee and colleagues.[43] The discrepancy highlights one of the challenges of the *in vitro* system; interlab variability can limit data reproducibility and limit result interpretation.

### 5.4.2 IL-2 Investigations

***5.4.2.1 Effects of IL-2 on Cytochrome P450 Enzymes*** Tinel and coworkers reported exposure to IL-2 in rat hepatocytes caused a decrease in mRNA levels of CYP3A2 and CYP2C11 and a corresponding reduction in protein expression of CYP1A1 (35%), CYP2B1/2 (35%), CYP2C11 (36%), CYP2D1 (26%), and CYP3A (56%).[14] The interaction of IL-2 on the various CYP enzymes may primarily be at the transcriptional level.

In the clinic, IL-2 has been observed to suppress both activity and protein levels of CYP3A.[51,52] *In vitro*, using cultured primary human hepatocytes, the inhibitory effect was transient with suppression observed at 48 h but returned to baseline at 72 h.[13] However, when Kupffer cells were co-cultured with hepatocytes at physiologically relevant ratios of 10:1 and 10:4 (hepatocytes to Kupffer cells), representing "normal" and "inflamed" states, respectively, IL-2 showed prolonged suppression of CYP3A activity (50–70%) out to 72 h in the co-cultures. Results suggest that the co-culture system may be more physiological representative and thus can offer an advantage for studying the effects of

**TABLE 5-3 Effects of Cytokines on Transporters**

| | Transporters | | | | | | |
|---|---|---|---|---|---|---|---|
| Cytokines | OAT2 | OCT1 | OATP1B1 | OATP1B3 | OATP2B1 | NTCP | MRP2/ Mrp2 |
| IL-1β | | | ↓r<br>↓a,p,r | ↓r | ↓r | ↓a,p,r | ↓r<br>↔r↓p<br>↓p,r |
| IL-2 | | | | | | | |
| IL-6 | ↓r | ↓a,r | ↓a,p,r (OATP) | ↓r | ↓r | ↓a,p,r | ↓p,r<br>↓p,r<br>↓p,r |
| Oncostatin | ↓r | ↓r | ↓a,p,r (OATP) | ↓r | ↓p,r | ↓a,p,r | ↓r↔p |
| TNF-α | ↓r | ↓a,r | ↓a,p,r | ↓a,r | ↓a,r | ↓a,p,r | ↔p,r<br>↓p,r<br>↓p,r |
| IFN-γ | ↔r | ↔r | ↓r | ↓r | ↓r | ↓r | ↓r |
| IFN-α2b | ↓p,r | ↔r | ↔r | ↔r | ↔r | | ↑r |

Upregulation (↑), downregulation (↓), or no effect (↔) on activity (a), protein levels (p), and/or mRNA (r) expression; BCRP: breast cancer resistance protein; BSEP: bile salt export pump; IL: interleukin; IFN: interferon; MDR1 P-gp: multi-drug resistant 1 P-glycoprotein; MRP: multidrug resistance protein; NPC: nonparenchymal cells; NTCP: sodium-taurocholate co-transporting polypeptide; OAT: organic anion transporter; OATP: organic anion transporting polypeptide; OCT: organic cation transporter; SC: sandwich cultured; TNF: tumor necrosis factor.

| MRP3 | MRP4 | MDR1 P-gp/ Mdr1a P-gp/ Mdr1b P-gp | BCRP | BSEP/ Bsep | Model | References |
|---|---|---|---|---|---|---|
| | | ↓r | | | Human Caco-2 | 41 |
| | | ↔r↓a,p | | | Rat hepatocytes | 42 |
| | | | | | Human hepatocytes | 11 |
| ↓r | ↓r | | ↓p | ↓p | Human hepatocytes and hepatoma HepaRG cells | 43 |
| | | | | ↓p,r | SC rat hepatocytes | 44 |
| | | | | ↓r↑p | SC human hepatocytes | 44 |
| | | ↓a,r | | | Caco-2 (human) | 45 |
| | | ↓a,p,r | | | Rat hepatocytes | 42 |
| ↔r↑p | ↓r | ↓r↔p | ↓p,r | ↔r | Human hepatocytes | 46 |
| | | | | ↓r↑p | SC human hepatocytes | 44 |
| | | | | ↑r↔p | SC rat hepatocytes | 44 |
| ↔p,r | ↔r | ↔p,r | ↓p,r | ↔r | Human hepatocytes and hepatoma HepaRG | 47 |
| | | ↑a,p,r | | | Rat hepatocytes | 48 |
| | | ↑r | | | Rat hepatocytes and hepatoma H-4-II-E cells | 34 |
| | | ↔r | | | Human Caco-2 | 41 |
| | | ↓a,r↔p | | | Human Caco-2 | 49 |
| ↔r↑p | ↔r | ↔p,r | ↔r↑p | ↓r | Human hepatocytes | 46 |
| | | | | ↓r↔p | SC human hepatocytes | 44 |
| | | | | ↓p,r | SC rat hepatocytes | 44 |
| | | ↓r | | | Human Caco-2 | 41 |
| | | ↑p,r↔a | | | Human Caco-2 | 49 |
| ↓r | | ↓r | ↓r | ↓r | Human hepatocytes | 50 |
| | | ↔r | ↔r | | Human hepatocytes-NPC co-culture | 22 |

cytokines, whereas hepatocytes alone are insufficient. The authors proposed that IL-2 effects on CYP3A may be through multiple mechanisms, including: direct binding to the hepatocytes IL-2R to produce a transient suppression of CYP3A; indirectly stimulating Kupffer cells to secret other proinflammatory cytokines, such as IL-1, IL-6, and TNF-α;[52] and indirectly stimulating Kupffer cells to release other proinflammatory mediators, such as prostaglandins, prostacyclins, and eicosanoids, which have been shown to regulate CYP activity. Further work is needed to investigate these hypotheses and the applicability of the hepatocyte-Kupffer cell co-culture system in general.

***5.4.2.2 Effects of IL-2 on Drug Transporters*** In Caco-2 cells, IL-2 significantly reduced P-gp transport of rhodamine-123 (21%) (100–5000 IU/mL, 24–72 h) and strongly suppressed MDR1 mRNA levels (1000 IU/mL, 72 h).[45] Currently, the mechanism by which IL-2 interacts with P-gp is unknown.

### 5.4.3 IL-4 Investigations

Limited ADME-related work has been reported for cytokine IL-4. Abdel-Razzak and colleagues noted variable effects of IL-4 (150 U/mL, 72 h) on CYP enzymes in human primary hepatocytes, in contrast to the usual suppressive effects observed with other proinflammatory cytokines. In their investigation, IL-4 appeared to have no effect on CYP1A2 and CYP2C gene expression, whereas CYP2E1 mRNA levels were significantly induced by five-fold.[10] Effects of IL-4 on CYP3A mRNA expression and activity were also not observed.

### 5.4.4 IL-6 Investigations

***5.4.4.1 Effects of IL-6 on Cytochrome P450 Enzymes*** As one of the major proinflammatory cytokines, IL-6 effects on CYP enzymes have been widely reported. Initially, in contrast to results from IL-1β exposure, early work by Barker and coworkers did not show an effect of IL-6 on 2,3,7,8-tetrachlorodibenzo-*p*-dioxin-induced CYP1A mRNA expression in rat hepatocytes.[9] However, later investigations in human primary hepatocytes produced consistent inhibitory effects by IL-6 on various CYP enzymes. The discrepancy highlights inherent interlab variability owing to a lack of standardization of experimental design and/or species evaluated, which is a limiting factor of the *in vitro* system. IL-6 suppression of CYP1A2 activity (18–90%) has been reported by Abdel-Razzak's group (50 U/mL, 72 h),[10] Guillen's group (100 U/ml, 48 h),[15] and Dickmann's group (0.005–50 ng/mL, 24–96 h).[18] In parallel with decreased enzyme activity, similar inhibition of CYP1A2 mRNA levels (25–%)[18] and protein expression were also detected.[15] It is interesting that Dickmann and coworkers demonstrated that the incorporation of an IL-6 neutralizing mAb can completely block IL-6 suppression of CYP1A2 activity.[18]

In regard to the mechanism of interaction, Guillen and coworkers noted that measured NO levels were unchanged in cytokine-treated and untreated hepatocytes, suggesting that IL-6 interaction with CYP1A2 may not be attributed to enzyme degradation associated with NO biosynthesis.[15] Abdel-Razzak and coworkers suggested that the presence of an IL-6-response element in the 5′ flanking regions of the CYP1A1/1A2 genes may play a role in the transcriptional regulation of CYP genes by IL-6.[10]

In addition to CYP1A2, IL-6 also had significant impact on CYP2B6 enzyme. Guillen and coworkers reported a 48% reduction in CYP2B6 activity by IL-6 (100 U/mL, 48 h),[15] which was consistent with results from Aitken and Morgan, who found IL-6 exposure (10 ng/mL, 24 h and 40 h) in human primary hepatocytes led to substantial suppression of both CYP2B6 mRNA (80%) and protein levels (25–50%).[12] Pascussi and colleagues also confirmed IL-6- mediated (100 U/mL, 24 h) inhibition of both CYP2B6 basal mRNA levels (60%) as well as phenobarbital- and rifampicin-induced mRNA levels (80–90%).[16] Guillen's group proposed that, similar to the interaction with CYP1A2, the effect of IL-6 on CYP2B6 may not be through protein degradation but most likely mediated by transcriptional regulation.[15] This was supported by Pascussi's group, who reported rapid and marked decrease in PXR (~80%) and CAR (~60%) mRNA expression in the presence of IL-6 (100 U/mL, 6 h).[16] These results suggest that the IL-6-mediated suppression of both PXR and CAR could potentially be the regulatory mechanism associated with the effects of IL-6 on rifampicin- and phenobarbital-induced CYP2B6 expression.

Consistent with the other CYP enzymes, IL-6 inhibition of the CYP2C family has also been observed. The mRNA levels of CYP2C8, CYP2C9, and CYP2C19 in general were decreased (25–80%) in the presence of IL-6.[10,12,16,18] The suppression of rifampicin- and phenobarbital-induced CYP2C8 and CYP2C9 mRNA expression (80–90%) was also observed after IL-6 exposure.[16] In addition, similar changes in CYP2C9 protein levels (25–50% suppression) were reported by Aitken and Morgan.[12] In contrast to the general inhibitory effects on the CYP2C family, IL-6 exposure did not affect CYP2C18, perhaps owing to the low basal levels of expression of this enzyme in the liver.[12]

The effects of IL-6 on the CYP3A family have also been extensively studied. Suppression of CYP3A4 activity (20–80%), mRNA expression (25–90%), and protein levels (25–50%) have been reported by various labs.[10,12,13,15,18] Dickmann and coworkers also observed that, similar to CYP1A2, the incubation with an IL-6 neutralizing mAb can partially prevent IL-6 suppression of CYP3A4 activity.[18] In addition to basal CYP3A4 levels, Pascussi's group[16] and Dickmann's group[18] also reported that IL-6 exposure can lead to decreased rifampicin- and phenobarbital-induced CYP3A4 gene expression. However, this was observed only at supraphysiological concentrations. Guillen and coworkers observed no changes in NO levels with IL-6 treatment that could cause enzyme degradation.[15] Pascussi and coworkers proposed that, similar to CYP2B6, IL-6-mediated suppression of both PXR and CAR genes could potentially be the regulatory mechanism by which IL-6 mediates the inhibition of rifampicin- and phenobarbital-induced CYP3A4 expression.[16]

***5.4.4.2 Effects of IL-6 on Phase II Metabolizing Enzymes*** For phase II metabolic enzymes, IL-6 had inhibitory effects on epoxide hydrolase mRNA expression (>40%)[10] but no interaction with glutathione *s*-transferase (GST) enzymes[40] when studied in human primary hepatocytes.

***5.4.4.3 Effects of IL-6 on Drug Transporters*** Unlike IL-1β, IL-6 (0–50 ng/mL, 24 h) treatment of rat hepatocytes significantly downregulated Mdr1b mRNA expression, suppressed P-gp protein levels (20–38%), and reduced P-gp-mediated rhodamine-123 transport.[42] It is believed that transcriptional modulation of the Mdr1 gene may be responsible for IL-6 effects on P-gp protein and activity. Mechanisms of interaction were not investigated, but the authors proposed that owing to similarities shared between CYP and Mdr1 gene families in regard to substrates, inducers, and inhibitors there could be parallelism in regulatory mechanisms of IL-6 on Mdr genes.[53]

In addition to P-gp, Le Vee and coworkers observed that IL-6 (0.1–10 ng/mL, 8–48 h) treatment of primary human hepatocytes suppressed mRNA levels of major influx and efflux transporters, including NTCP, OATP1B1, OATP1B3, OATP2B1, OCT1, OAT2, MRP2, MRP4, and BCRP.[46] The effects on these transporters were mostly concentration dependent (NTCP, OATP1B1, OATP1B3, OATP2B1, OCT1, and BCRP) and time dependent (MDR1). On the other hand, both BSEP and MRP3 gene expression were unaffected by IL-6. In regard to protein levels, IL-6 (10 ng/mL, 48 h) did not alter P-gp but did reduce NTCP, OATP1B1, BCRP, and MRP2, and induced MRP3. Consistent with results of mRNA and protein expression, NCTP, OATPs, and OCT1 activities were also significantly reduced in human primary hepatocytes. Although the mechanisms of the effects were unexplored, modulation of gene transcription leading to suppressed protein and activity levels may explain how IL-6 interacts with NCTP, OATP, and OCT1.

Similar to Le Vee and coworkers, Diao's group observed significant IL-6 (100 ng/mL, 24–48 h) downregulation of both mRNA and protein levels of MRP2/Mrp2 in SC human and rat hepatocytes.[44] On the other hand, variable results were reported with BSEP in SC human hepatocytes: IL-6 suppressed mRNA transcription while inducing protein levels. In SC rat hepatocytes, however, *Bsep* gene expression was induced but no effects were seen on protein levels. Collectively, results demonstrate a differential effect of BSEP/Bsep expression between SC human and rat hepatocytes most likely attributed to species-dependent signaling pathways.

The mechanism of IL-6 regulation was uninvestigated, but it is believed to involve both transcriptional mechanisms associated with the phosphorylation of extracellular-reactive kinase (ERK), Jun kinase (JNK), and p38 mitogen-activated protein kinase (MAPK) as well as post-transcriptional mechanisms.[26,43,54]

### 5.4.5 Oncostatin

***5.4.5.1 Effects of Oncostatin on Cytochrome P450 Enzymes*** Oncostatin M (OSM) is a multifunctional cytokine that is related to IL-6. Both have overlapping biological effects in many cellular systems owing to similarity in structure,

function,[55,56] and signal transduction pathway activation.[57–59] OSM acts on a wide variety of cells: It regulates growth of certain tumor and nontumor-derived cell lines,[60,61] induces cell differentiation,[55,62] and upregulates low-density lipoprotein in liver cells[63] and acute phase protein in hepatocytes (i.e., α1-antichymotrypsin, haptoglobin, and α1-acid glycoprotein).[64]

Guillen and coworkers demonstrated that in human primary hepatocytes exposure to OSM (25–50 U/mL) produced maximal reductions in CYP1A2 (~55%) and CYP3A4 (~30%) activity after 48 h of treatment.[15] Comparable decreases were also observed for CYP2A6 (30%) and CYP2B6 (62%). The inhibitory effects of OSM on CYP enzymes were also detected after induction of both CYP1A2 and CYP3A4 by 3-methylcholanthrene (MC) and dexamethasone (DEX), respectively. MC-induced CYP1A2 showed a sixfold increase in activity compared to an uninduced control. In the presence of OSM (25 U/mL), only 43% of the induced CYP1A2 activity was detected. Similarly, DEX-induced CYP3A4 activity was increased twofold, and in the presence of OSM 47% of induced CYP3A4 activity was observed.

Reported OSM suppression of CYP activities was consistent with observed protein levels. The presence of OSM in increasing concentration (up to 50 U/mL) resulted in decreased protein synthesis of both CYP1A2 and CYP3A4. After MC- and DEX-induction of CYP1A2 and CYP3A4, respectively, protein levels were increased by ~60% (CYP1A2) and ~100% (CYP3A4) of uninduced control. Introducing OSM after induction inhibited the synthesis of both CYP1A2 and CYP3A4 enzymes to levels that were similar to uninduced cells. OSM presence did not alter NO synthesis; therefore, its effect on CYP protein levels was not attributed to NO.

Parallel to activities and protein synthesis, significant reduction of CYP1A2 and CYP3A4 mRNA expression occurs in the presence of OSM (50 U/mL).[15] This suggests that OSM regulation of CYP enzymes may occur at the transcriptional level, perhaps by similar mechanisms as the related cytokine IL-6, whereby PXR and CAR mRNA levels were altered.[16]

***5.4.5.2 Effects of Oncostatin on Drug Transporters*** In addition to CYP enzymes, Le Vee and coworkers also evaluated the effects of OSM on sinusoidal and canalicular drug transporters in human primary hepatocytes.[47] Exposure to OSM (0.01–20 ng/mL, 8–48 h) produced no effects on P-gp, MRP3, MRP4, and BSEP mRNA expression. In contrast, OSM exposure led to the downregulation of mRNA levels of major sinusoidal solute carrier (SLC) influx transporters (NTCP, OATP1B1, OATP1B3, OATP2B1, OCT1, and OAT2) as well as ATP-binding cassette (ABC) efflux transporters (MRP2 and BCRP). The OSM $EC_{50}$ of mRNA expression ranged between 0.12 and 1.16 ng/mL in human primary hepatocytes. Correspondingly, the protein expression of NTCP, OATP1B1, OATP2B1, and BCRP along with transport activities of NTCP and OATP were also inhibited after OSM exposure.

Rank correlation analysis of suppressive effects of OSM toward drug transporters showed that OSM effects were highly correlated with those of IL-6 and, in a more moderate manner, with those of IL-1β. By contrast, they were not correlated with those of TNF-α and IFN-γ. Type II OSM receptor has been implicated as a

mediator of OSM transcriptional regulation of these various transporters. Knocked-down expression of the type II OSM receptor gp130 and OSMRβ subunits with siRNA transfection has counteracted the downregulation of NTCP, OATP1B1, and OATP2B1 mRNA expression. Le Vee and coworkers hypothesized that the signaling pathway that is activated by this receptor, especially by its signal-transducing component gp130 subunit, may play a major role in OSM-related changes in transporter expression.[47] This is further supported by the fact that altered transporter expression owing to OSM treatment was highly correlated with inhibitory effects triggered by IL-6 and that the IL-6 receptor also uses the gp130 subunit for signal transduction.[65] OSM and IL-6 are known to activate several distinct signaling pathways through the gp130 subunit, including the Janus kinase/signal transducer and activator of transcription cascade, the MAPK pathway, and the phosphatidylinositol 3-kinase pathway.[66] In agreement with these data, the MAPK ERK was found to be activated by OSM.[47]

### 5.4.6 TNF-α Investigations

***5.4.6.1 Effects of TNF-α on Cytochrome P450 Enzymes*** Abdel-Razzak and coworkers observed substantial inhibition of CYP1A2 (30–85%) and 3A activities (24–90%) by TNF-α (50 U/mL, 72 h) in human primary hepatocytes.[10] Analogous reduction in mRNA expression of these enzymes (30–80%) was also detected, indicating a probable transcriptional component to TNF-α effects on these enzymes.[10,12]

Similar to CYP1A2 and CYP3A, gene expression of CYP2C and CYP2E1 was also negatively affected by the presence of TNF-α, with 30–80% inhibition observed.[10] On the other hand, Aitken and Morgan observed only minimal interactions of TNF-α (10 ng/mL, 24 h and 40 h) with mRNA expression of CYP2B6, CYP2C9, CYP2C18, and CYP2C19 in human primary hepatocytes.[12] Despite minor effects on gene expression, CYP2B6 and CYP2C9 protein levels were inhibited, suggesting that TNF-α may interact with these enzymes through modification of protein translation and/or enzyme degradation. The differential effects of TNF-α on various CYP enzymes highlights potentially different regulatory mechanisms involved. TNF-α may regulate CYP3A4 at the transcriptional level, possibly via NF-κB interference with the PXR-RXR complex as suggested by Gu and coworkers (Figure 5-3b) and, in contrast, affect the translation and/or stability of either CYP2B6 or CYP2C9 enzymes.[19]

***5.4.6.2 Effects of TNF-α on Phase II Metabolizing Enzymes*** Limited studies have been dedicated to evaluating the effects of TNF-α on the phase II metabolic enzymes. Abdel-Razzak's group has observed, however, a decrease of epoxide hydrolase mRNA levels in human primary hepatocytes.[10]

***5.4.6.3 Effects of TNF-α on Drug Transporters*** Several studies have increased the understanding of the interaction between TNF-α and the various drug transporters. In rat primary hepatocytes, Hirsch-Ernst and coworkers observed that

TNF-α (4000 U/ml, 72 h) exposure led to a twofold induction of Mdr1b mRNA and P-gp protein and a corresponding decrease in intracellular accumulation of the P-gp substrate rhodamine-123.[48] It was postulated that the induction of Mdr1b mRNA may have been due to elevated oxygen species associated with the presence of TNF-α. Reactive oxygen species have been implicated as a mediator of various TNF-α effects, including activation of NF-κB,[67] stimulation of gene transcription,[68] and alteration of protein phosphorylation.[67] After treating rat primary hepatocytes with antioxidants, Hirsch-Ernst and coworkers detected marked suppression of TNF-α-induced Mdr1b Pgp mRNA and protein levels.[48]

Inhibition of Mdr1 mRNA (56%) was measured in human Caco-2 cells after treatment with TNF-α (1–10 ng/mL, 24–96 h) by Belliard's group.[49] In contrast to the results from rat hepatocytes, however, the suppressed gene expression did not translate to a decreased P-gp protein level but did decrease (20%) unidirectional transport of rhodamine-123. The discrepancy in TNF-α induction/suppression of Mdr1 may be attributed to differences in the experimental models (rat hepatocytes vs. human carcinoma cell lines) and/or differences in regulatory mechanisms in rat and human.

Besides P-gp, TNF-α interaction with MRP2 and BSEP were assessed by Diao and colleagues in both SC human and rat hepatocyte models.[44] After TNF-α treatment (100 ng/mL, 24 and 48 h), significant suppression of both mRNA and protein levels were measured for MRP2/Mrp2 in SC human and rat hepatocytes. On the other hand, variable results were detected with BSEP in SC human hepatocytes as TNF-α-suppressed mRNA transcription but showed no effects on protein. In SC rat hepatocytes, however, both Bsep mRNA and protein levels were reduced by TNF-α. Similar to IL-6, results demonstrate a differential effect of BSEP/Bsep expression between SC human and rat hepatocytes that may be attributed to species-dependent signal transduction pathways. Similar to IL-1β and IL-6, regulation by TNF-α may involve transcriptional mechanism associated with the activation of ERK, JNK, and p38 MAPK phosphorylation.[26]

Le Vee's group observed that in human primary hepatocytes incubation with TNF-α (0.01–100 ng/mL, 8–48 h) led to significant reduction of NTCP, OATP2B1, OATP1B1, OATP1B3, OCT1, OAT2, and BSEP mRNA, whereas mRNA levels of MDR1, MRP2, MRP3, MRP4, and BCRP were unchanged.[46] The effects were mostly concentration dependent (0.01–100 ng/mL), with marked inhibition of gene expression at ~1 ng/mL TNF-α. In regard to protein expression, TNF-α incubation did not alter MRP2 and P-gp protein levels (corresponding to unaltered gene expression), but NTCP and OATP1B1 levels were reduced and BCRP and MRP3 were induced. Consistent with mRNA and protein expression results, NCTP, OATP, and OCT1 activities were also significantly decreased in human primary hepatocytes. For these transporters, although the mechanism of interaction was unexplored, experimental outcomes suggest that regulation by TNF-α may be at the transcriptional level.

### 5.4.7 IFN Investigations

***5.4.7.1 Effects of IFN-γ on Cytochrome P450 Enzymes*** The interaction of IFN with CYP1A2 has been extensively observed in human primary hepatocytes.

In general, downregulation of CYP1A2 basal activity (29–53%) and 3-methylcholanthrene-induced activity (32%) was observed in the presence of IFN-γ (50–300 U/mL, 24–72 h).[10,15,20] The decrease in activity was consistent with lowered mRNA (up to ~40%)[10]and diminished basal and 3-methylcholanthrene-induced CYP1A2 protein expression (72%).[15,20] These results suggest that the interaction of IFN-γ with CYP1A2 is most likely mediated through gene transcription. However, post-transcriptional regulation may also play a contributing factor because IFN-γ has previously been shown to suppress CYP1A2 activity in parallel with markedly elevated NO production.[20] Besides IFN-γ, IFN-α2b has also been shown to produce analogous inhibition of CYP1A2. Similarly, IFN-α2b can also interact with the interferon receptor to influence the upregulation or downregulation of various genes in the liver.[69]

Similar to CYP1A2, suppression of CYP2A6 and CYP2B6 activities were also observed in the presence of IFN-γ (300 U/mL, 24 h).[15,20] This decrease was in parallel to elevated NO production, which was partially reversed by the presence of an inhibitor of NO biosynthesis ($N^G$-monomethyl-L-arginine). Associated with the activity suppression, Aitken and Morgan also saw significant inhibition of CYP2B6 mRNA (80%) and protein levels (~65%) after exposure to IFN (10 ng/mL, 24 h).[12] It appears that IFN may regulate CYP2A6 and CYP2B6 by both gene transcription (as evident by changes in mRNA levels) and enzyme stability (influenced by increased NO biosynthesis).

Like other CYP enzymes, IFN downregulated CYP3A activity (21–55%) but had no effect on gene expression after incubation with IFN-γ (50 U/mL, 72 h).[10] On the other hand, IFN-γ at 300 U/mL for 24–48 h elicited suppression of both basal and DEX-induced CYP3A4 activity (20–34%), along with reduced mRNA ($\geq$ 80%) and protein levels (~ 50%).[12,15,20] Discrepancy in gene expression results may be attributed to differences in donors and/or experimental conditions. The noted changes in mRNA expression suggest a transcriptional component in IFN regulation of CYP3A4. In addition, Donato and coworkers detected elevated NO biosynthesis in parallel with CYP3A4 inhibition and a partial reversal of the inhibition after incorporation of $N^G$-monomethyl-L-arginine to culture media, similar to CYP1A2.[20] These findings suggest that enzyme stability may have a minor role in IFN regulation of the CYP3A family.

In contrast to the other CYP enzymes, IFN in general does not appear to have much interaction with the CYP2C family. With the exception of CYP2C8, for which significant mRNA reduction was detected, no effects were observed on the gene expression of CYP2C9, CYP2C18, and CYP2C19 after IFN-γ (10 ng/mL, 24 and 40 h) incubation in primary human hepatocytes.[10,12]

Chen's group assessed the impact of IFN-α2b on the expression of various drug-metabolizing enzymes in freshly prepared human primary hepatocytes co-cultured with nonparenchymal cells (which included a mixture of immature hepatocytes, Kupffer cells, biliary cells, stellate cells, and hepatic stem cells).[22] At therapeutically relevant concentrations (1000–3000 U/mL), IFN-α2b was shown to downregulate mRNA expression of several metabolizing enzymes, including CYP1A2,

CYP2B6, and CYP2E1. The observed downregulation of mRNA was as high as 50%, with a corresponding decrease in protein expression. Conversely, significant upregulation of CYP3A4 mRNA and protein expression was detected. The observed induction of CYP3A4 in this experimental system was unexpected and in contrast to CYP3A4 suppression detected by Guillen's group[15] with IFN and Aitken and Morgan[12] with IFN-γ. Possible mechanisms mediating this effect are currently unknown and require further elucidation. In contrast to CYP3A4, IFN-α2b suppression of CYP1A2, CYP2B6, and CYP2E1 mRNA and protein expression could potentially be attributed to regulation at the transcriptional level owing to substantial inhibition of CYP gene expression.

***5.4.7.2 Effects of IFN-γ on Phase II Metabolizing Enzymes*** In human primary hepatocytes co-cultured with nonparenchymal cells, IFN-α2b (at therapeutically relevant concentrations of 1000–3000 IU/mL) was shown to have an impact on the expression of various phase II metabolizing enzymes, including the downregulation of UDP-glucuronosyltransferase (UGT2B7) and sulfotransferase (SULT1A1) mRNA expression.[22] Similar to gene expression, UGT2B7 protein levels were also inhibited. Although the regulatory mechanism remains to be elucidated, transcriptional modulation leading to altered protein expression is most likely a method by which IFN-α2b interacts with the various phase II metabolic enzymes. In contrast to UGT2B7 and SULT1A1, IFN-α2b showed no significant effect on the mRNA expression of SULT2A1, carboxylesterase 2 (CES2)[22] or epoxide hydrolase.[10]

***5.4.7.3 Effects of IFN-γ on Drug Transporters*** Belliard and coworkers observed induction of both MDR1 mRNA (2.5-fold) and P-gp protein levels when Caco-2 cells were incubated with IFN-γ (10 ng/mL, 24–72 h).[49] The increased protein expression, however, did not alter P-gp efflux activity. Confocal laser scanning microscopy showed altered distribution of P-gp with more co-localization with lateral F-actin, which may explain the absence in correlation between the upregulation of P-gp expression by IFN-γ and the absence of a change in P-gp activity.

In human primary hepatocytes, while gene expression of both OAT2 and OCT1 remained unchanged, a general downregulation of mRNA levels of both sinusoidal influx transporters (NTCP, OATP1B1, OATP1B3, and OATP2B1) and efflux transporters (MDR1, MRP2, MRP3, BCRP, and BSEP) was detected after treatment with IFN-γ (10 ng/mL, 24 h).[50] The opposite was observed when human hepatocytes were co-cultured with nonparenchymal cells and exposed to IFN-α2b.[22] In general, there was no significant effect on the mRNA expression of OATP1B1, OATP1B3, OATP2B1, OCT1, P-gp, and BCRP at therapeutically relevant IFN-α2b concentrations (1000–3000 IU/mL). The only exception was a significant decrease in both mRNA and protein expression of OAT2. The mechanism of interaction is most likely at the level of gene transcription, leading to altered protein expression.

## 5.5 SUMMARY AND CONCLUSION

In general, with further characterization and validation, the primary hepatocyte monoculture system does add value as an *in vitro* tool for studying the effects of proinflammatory cytokines on metabolizing enzymes and drug transporters. As an isolated system, hepatocyte monocultures are useful for exploring the mechanisms of interaction between cytokines and drug-metabolizing enzymes and transporters. Furthermore, the commercial availability of hepatocytes from numerous preclinical species and from humans makes them an easily accessible experimental tool to facilitate better understanding of the effects of proinflammatory cytokines in potential toxicological species as well as potential future clinical extrapolation. In addition, the commercial availability of plateable cryopreserved hepatocytes also offers the advantage of reducing interassay variability.

On the other hand, despite some advantages current utility of the hepatocyte monoculture system has limitations (Figure 5-5), including incomplete mechanistic understanding of CYP downregulation and lack of agreement on positive controls representing various pathways. It should also be recognized that mimicking the complex effects of disease state on ADME genes *in vitro* may not be feasible. Furthermore, currently only qualitative assessment of the effects of proinflammatory cytokines on CYP enzymes and transporters can be ascertained in this experimental model. The literature reviewed here illustrates that the overall downregulation of CYP enzymes and transporters can be detected when assessed with the major proinflammatory cytokines (IL-1, IL-6, TNF-α, and IFN-γ), however, the limitation of this *in vitro* model is more apparent when extrapolation of these data to clinical

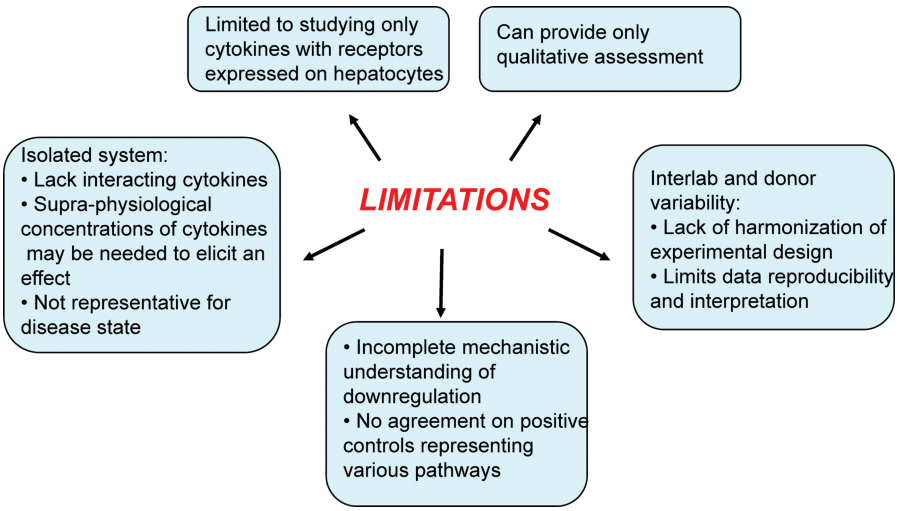


**FIGURE 5-5** Limitations of hepatocyte monocultures.

significance is needed. In addition, the lack of standardization of experimental design and study conditions across different labs also limits study reproducibility and data interpretation. Likely, the hepatocyte monoculture system can facilitate assessment only with cytokines that have receptors expressed on the hepatocytes. This limits the utility of the experimental system for cytokines that do not bind directly to receptors expressed on hepatocytes but can still render an effect on CYP enzymes and transporters indirectly through stimulation of nonparenchymal cells (i.e., Kupffer cells). These cells could release proinflammatory cytokines that can in turn act on the hepatocytes. Last, because the monoculture system is an isolated system with no interacting cytokines, supra-physiological concentrations of studied cytokines may be needed to elicit an effect. As a result of all the drawbacks from the hepatocyte monoculture experimental system, interest is gaining for a more inclusive and physiologically representative *in vitro* model for studying the effects of proinflammatory cytokines on the metabolic disposition of SM drugs. To this goal, work by Sunman and coworkers and Chen and coworkers attempted to implement nonparenchymal cells (including Kupffer cells) with hepatocytes in a co-culture system.[13,22] The paucity of literature on this co-culture system indicates that experience (or success) with this experimental model is limited across industry and academia. An additional complication associated with this model is the difficulty of obtaining relatively pure cellular fractions. Thus further characterization and validation of this *in vitro* co-culture model is necessary before its value can be determined.

It is well recognized across academia and industry that the limited amount of *in vitro* data that have been generated thus far limits the use of the hepatocyte monoculture model as a predictive tool for clinical DDIs. In general, these limitations are related to the lack of acceptable and uniformly available models that have been demonstrated to provide not only reasonable *in vitro–in vivo* correlations but also mechanistic rationale defining the limitations of such models. Hence,it is now widely recognized that there is a need to generate more *in vitro* data, especially for biologics that have been clinically shown to cause DDIs.

As discussed in this chapter, initial focus on changes in net CYP activities is a reasonable approach because these enzymes are involved in the metabolism of most small molecule therapeutics and can be well characterized by using established probe substrates. This exercise, however, should not preclude generation of gene expression data in parallel because this technique allows assessing gene expression of any gene of interest (particularly phase II enzymes and transporters). A primary challenge for *in vitro–in vivo* correlations will be to extend data generated *in vitro* to predict whether novel biologics will cause a net change in drug metabolism related to enzymes and transporters in patients. A particular challenge would be the difficulty in predicting DDIs that may be affected or may be variable when the DDI is evaluated relative to the stage of disease.

In conclusion, assessing changes of enzyme and transporter activity and gene expression in novel *in vitro* systems that could predict potential SM and biologic

DDIs is a valid and important effort, particularly because the clinical assessment of such DDIs is long, difficult, and expensive.

## REFERENCES


1. Morgan ET. Impact of infectious and inflammatory disease on cytochrome P450-mediated drug metabolism and pharmacokinetics. *Clin Pharmacol Ther* **85**(4), 434–38 (2009).
2. Petrovic V, Teng S, Piquette-Miller M. Regulation of drug transporters during infection and inflammation. *Mol Interv* **7**, 99–111 (2007).
3. Williams, SJ, Baird-Lambert JA, Farrell JC. Inhibition of theophylline metabolism by interferon. *Lancet* **330**, 939–41 (1987).
4. Schmitt C, Kuhn B, Zhang X, et al. Disease-drug-drug interaction involving tocilizumab and simvastatin in patients with rheumatoid arthritis. *Clin Pharmacol Ther* **89**(5), 735–40 (2011).
5. Mahmood I, Green MD. Drug interaction studies of therapeutic proteins or monoclonal antibodies. *J Clin Pharmacol* **47**, 1540–54 (2007).
6. Seitz K, Zhou H. Pharmacokinetic drug-drug interaction potentials for therapeutic monoclonal antibodies: reality check. *J Clin Pharmacol* **47**, 1104–18 (2007).
7. Lee JI, Zhang L, Men AY, et al. CYP-mediated therapeutic protein-drug interactions clinical findings, proposed mechanisms and regulatory implications. *Clin Pharmacokinet* **49**(5), 295–310 (2010).
8. Huang SM, Zhao H, Lee JI, et al. Therapeutic protein–drug interactions and implications for drug development. *Clin Pharmacol Ther* **87**(4), 497–503 (2010).
9. Barker CW, Fagan JB, Pasco DS. Interleukin-1β suppresses the induction of P4501A1 and P4501A2 mRNAs in isolated hepatocytes. *J Biol Chem* **267**(12), 8050–55 (1992).
10. Abdel-Razzak Z, Loyer P, Fautrel A. et al. Cytokines down-regulate expression of major cytochrome P-450 enzymes in adult human hepatocytes in primary culture. *Mol Pharmacol* **44**, 707–15 (1993).
11. Assenat E, Gerbal-Chaloin S, Larrey D, et al. Interleukin-1β inhibits CAR-induced expression of hepatic genes involved in drug and bilirubin clearance. *Hepatology* **40**, 951–60 (2004).
12. Aitken AE, Morgan ET. Gene-specific effects of inflammatory cytokines on cytochrome P450 2C, 2B6 and 3A4 mRNA levels in human hepatocytes. *Drug Metab Dispos* **35**, 1687–93 (2007).
13. Sunman JA, Hawke RL, LeCluyse EL, Kashuba ADM. Kupffer cell-mediated IL-2 suppression of CYP3A activity in hepatocytes. *Drug Metab Dispos* **32**, 359–63 (2004).
14. Tinel M, Robin MA, Doostzadeh J, et al. The interleukin-2 receptor down-regulates the expression of cytochrome P450 in cultured rat hepatocytes. *Gastroenterology* **109**, 1589–99 (1995).
15. Guillen MI, Donato MT, Jover R, et al. Oncostatin M down-regulates basal and induced cytochromes P450 in human hepatocytes. *J Pharmacol Exper Ther* **285**, 127–34 (1998).
16. Pascussi JM, Gerbal-Chaloin S, Pichard-Garcia L, et al. Interleukin-6 negatively regulates the expression of pregnane X receptor and constitutively activated receptor in primary human hepatocytes. *Biochem Biophys Res Commun* **274**, 707–13 (2000).

17. Gubbins PO, Melchert RB, McConnell SA, et al. Effect of interleukin-6 on the hepatic metabolism of itraconazole and its metabolite hydroxyitraconazole using primary human hepatocytes. *Pharmacology* **67**, 195–201 (2003).
18. Dickmann LJ, Patel SK, Rock DA, et al. Effects of interleukin-6 (IL-6) and an anti-IL-6 monoclonal antibody on drug-metabolizing enzymes in human hepatocyte culture. *Drug Metab Dispos* **39**, 1415–22 (2011).
19. Gu X, Ke S, Liu D, et al. Role of NF-κB in regulation of PXR-mediated gene expression: a mechanism for the suppression of cytochrome P450 3A4 by proinflammatory agents. *J Biol Chem* **281**(26), 17882–89 (2006).
20. Donato MT, Guillen MI, Jover R, et al. Nitric oxide-mediated inhibition of cytochrome P450 by interferon-γ in human hepatocytes. *J Pharmacol Exper Ther* **281**, 484–90 (1997).
21. Flaman AS, Gravel C, Hashem AM, et al. The effect of interferon-α on the expression of cytochrome P450 3A4 in human hepatoma cells. *Toxicol Appl Pharmacol* **253**, 130–36 (2011).
22. Chen C, Han YH, Yang Z, Rodrigues AD. Effect of interferon-α2b on the expression of various drug metabolizing enzymes and transporters in cocultures of freshly prepared human primary hepatocytes. *Xenobiotica* **41**(6), 476–85 (2011).
23. Aitken AE, Lee CM, Morgan ET. Roles of nitric oxide in inflammatory down-regulation of human cytochrome P450. *Free Radical Biol Med* **44**(6), 1161–68 (2008).
24. Morgan ET. Regulation of cytochrome P450 by inflammatory mediators: why and how? *Drug Metab Disp* **29**, 207–12 (2001).
25. Aitken AE, Richardson TA, Morgan ET. Regulation of drug-metabolizing enzymes and transporters in Inflammation. *Ann Rev Pharmacol Toxicol* **46**, 123–49 (2006).
26. Fardel O, Le Vee M. Regulation of human hepatic drug transporter expression by pro-inflammatory cytokines. *Expert Opin Drug Metab Toxicol* **5**, 1469–481 (2009).
27. Liptrott NJ, Owen A. The role of cytokines in the regulation of drug disposition: extended functional pleiotropism? *Expert Opin Drug Metab Toxicol* **7**(3), 341–52 (2011).
28. Na SY, Kang BY, Chung SW, et al. Retinoids inhibit interleukin-12 production in macrophages through physical associations of retinoid X receptor and NF-κB. *J Biol Chem* **274**, 7674–80 (1999).
29. Xie W, Barwick JL, Downes M, et al. Humanized xenobiotic response in mice expressing nuclear receptor SXR. *Nature* **406**, 435–39 (2000).
30. Urquhart BL, Tirona RG, Kim RB. Nuclear receptors and the regulation of drug-metabolizing enzymes and drug transporters: Implications for interindividual variability in response to drugs. *J Clin Pharmacol* **47**, 566–78 (2007).
31. Ke S, Rabson AB, Germino JF, et al. Mechanism of suppression of cytochrome P-450 1A1 expression by tumor necrosis factor-alpha and lipopolysaccharide. *J Biol Chem* **276**, 39638–644 (2001).
32. Ghezzi P, Bianchi M, Gianera L, et al. Role of reactive oxygen intermediates in the interferon-mediated depression of hepatic drug metabolism and protective effect of *N*-acetylcysteine in mice. *Cancer Res* **45**, 3444–47 (1985).
33. Cribb AE, Renton KW. Dissociation of xanthine oxidase induction and cytochrome P450 depression during interferon induction in the rat. *Biochem Pharmacol* **46**(11), 2114–17 (1993).

34. Ros JE, Schuetz JD, Geuken M, et al. Induction of *Mdr1b* expression by tumor necrosis factor-α in rat liver cells is independent of p53 but requires NF-κB signaling. *Hepatology* **33**, 1425–31 (2001).
35. Zhou G, Kuo MT. NF-kappaB-mediated induction of *Mdr1b* expression by insulin in rat hepatoma cells. *J Biol Chem* **272**, 15174–83 (1997).
36. Zhou G, Kuo MT. Wild-type p53-mediated induction of rat *Mdr1b* expression by the anticancer drug daunorubicin. *J Biol Chem* **273**, 15387–94 (1998).
37. Lee CM, Pohl J, Morgan ET. Dual mechanisms of CYP3A protein regulation by proinflammatory cytokine stimulation in primary hepatocyte cultures. *Drug Metab Dispos* **37**, 865–72 (2009).
38. Assenat E, Gerbal-Chaloin S, Maurel P, et al. Is nuclear factor kappa-B the missing link between inflammation, cancer and alteration in hepatic drug metabolism in patients with cancer? *Eur J Cancer* **42**, 785–92 (2006).
39. Maheo, K., Antras-Ferry, J., Morel, F., et al. Modulation of glutathione s-transferase subunits A2, M1, and P1 expression by interleukin-1beta in rat hepatocytes in primary culture. *J Biol Chem* **272**(26), 16125–32 (1997).
40. Langouet S, Corcos L, Abdel-Razzak Z, et al. Up-regulation of glutathione *s*-transferases alpha by interleukin 4 in human hepatocytes in primary culture. *Biochem Biophys Res Commun* **216**(3), 793–800 (1995).
41. Bertilsson PM, Olsson P, Magnusson KE. Cytokines influence mRNA expression of cytochrome P450 3A4 and MDR1 in intestinal cells. *J Pharm Sci* **90**(5), 638–46 (2001).
42. Sukhai M, Yong A, Pak A, Piquette-Miller M. Decreased expression of P-glycoprotein in interleukin-1β and interleukin-6 treated rat hepatocytes. *Inflamm Res* **50**, 362–70 (2001).
43. Le Vee M, Gripon P, Stieger B, Fardel O. Down-regulation of organic anion transporter expression in human hepatocytes exposed to the proinflammatory cytokine interleukin-1β. *Drug Metab Dispos* **36**, 217–22 (2008).
44. Diao L, Li N, Brayman TG, et al. Regulation of MRP2/ABCC2 and BSEP/ABCB11 expression in sandwich cultured human and rat hepatocytes exposed to inflammatory cytokines TNF-α, IL-6, and IL-1β. *J Biol Chem* **285**(41), 31185–92 (2010).
45. Belliard AM, Tardivel S, Farinotti R, et al. Effect of hr-IL-2 treatment on intestinal P-glycoprotein expression and activity in Caco-2 cells. *J Pharm Pharmacol* **54**(8), 1103–09 (2002).
46. Le Vee M, Lecureur V, Stieger B, Fardel O. Regulation of drug transporter expression in human hepatocytes exposed to the proinflammatory cytokines tumor necrosis factor-α or interleukin-6. *Drug Metab Dispos* **37**, 685–93 (2009).
47. Le Vee M, Jouan E, Stieger B, et al. Regulation of drug transporter expression by oncostatin M in human hepatocytes. *Biochem Pharmacol* **82**, 304–11 (2011).
48. Hirsch-Ernst KI, Ziemann C, Foth H, et al. Induction of mdr1b mRNA and P-glycoprotein expression by tumor necrosis factor alpha in primary rat hepatocyte cultures. *Journal of Cellular Physiology*, **176**, 506–515, (1998).
49. Belliard AM, Lacour B, Farinotti R, Leroy C. Effect of tumor necrosis factor-alpha and interferon-gamma on intestinal P-glycoprotein expression, activity, and localization in Caco-2 cells. *J Pharm Sci* **93**, 1524–36 (2004).
50. Le Vee M, Jouan E, Moreau A, Fardel O. Regulation of drug transporter mRNA expression by interferon-c in primary human hepatocytes. *Fundam Clin Pharmacol* **25**, 99–103 (2011).

51. Elkahwaji J, Robin MA, Berson A, et al. Decrease in hepatic cytochrome P450 after interleukin-2 immunotherapy. *Biochem Pharmacol* **57**, 951–54 (1999).
52. Piscitelli SC, Forrest A, Vogel S, et al. Pharmacokinetic modeling of recombinant interleukin-2 in patients with human immunodeficiency virus infection. *Clin Pharmacol Ther* **64**, 492–98 (1998).
53. Wacher VJ, Wu CY, Benet LZ. Overlapping substrate specificities and tissue distribution of CYP 3A and PGP: Implications for drug delivery and activity in cancer chemotherapy. *Mol Carcinog* **13**, 129–34 (1995).
54. Ogata S, Kubota Y, Yamashiro T, et al. Signaling pathways regulating IL-1alpha-induced COX-2 expression. *J Dental Res* **86**, 186–91 (2007).
55. Rose TM, Bruce AG. Oncostatin M is a member of a cytokine family that includes leukemia-inhibitory factor, granulocyte colony-stimulating factor, and interleukin 6. *Proc Natl Acad Sci USA* **88**, 8641–45 (1991).
56. Hibi M, Nakajima K, Hirano T. IL-6 cytokine family and signal transduction: a model of the cytokine system. *J Mol Med* **74**, 1–12 (1996).
57. Gearing DP, Comeau MR, Friend DJ, et al. The IL-6 signal transducer, pg130: an oncostatin M receptor and affinity converter for the LIF receptor. *Science* **255**, 1434–37 (1992).
58. Liu J, Modrell B, Aruffo A, et al. Interleukin-6 signal transducer gp130 mediates oncostatin M signaling. *J Biol Chem* **267**, 6763–66 (1992).
59. Kishimoto T, Taga T, Akira S. Cytokine signal transduction. *Cell* **76**, 253–62 (1994).
60. Zarling JM, Shoyab M, Marquardt H, et al. Oncostatin M: a growth regulator produced by differentiated histiocytic lymphoma cells. *Proc Natl Acad Sci U S A* **83**, 9739–43 (1986).
61. Horn D, Fitzpatrick M, Gompper PT, et al. Regulation of cell growth by recombinant oncostatin-M. *Growth Factors* **2**, 157–65 (1990).
62. Brown TJ, Rowe JM, Liu JW, Shoyab M. Regulation of IL-6 expression by oncostatin M. *J Immunol* **174**, 2175–80 (1991).
63. Grove RI, Mazzuco CE Radka SF, et al. Oncostatin M up-regulates low density lipoprotein receptors in HepG2 cells by a novel mechanism. *J Biol Chem* **266**, 194–99 (1991).
64. Richards CD, Brown TJ, Shoyab M, et al. Recombinant oncostatin M stimulates the production of acute phase proteins in hepG2 cells and rat primary hepatocytes in vitro. *J Immunol* **148**,1731–36 (1992)
65. Silver JS, Hunter CA. gp130 at the nexus of inflammation, autoimmunity, and cancer. *J Leukoc Biol* **88**(6), 1145–56 (2010).
66. Nakashima K, Taga T. gp130 and the IL-6 family of cytokines: signaling mechanisms and thrombopoietic activities. *Semin Hematol* **35**(3), 210–21 (1998).
67. Remacle J, Reas M, Toussaint O, et al. Low levels of reactive oxygen species as modulators of cell function. *Mutat Res* **316**, 103–22 (1995).
68. Lo YY, Cruz TT. Involvement of reactive oxygen species in cytokine and growth factor induction of c-*fos* expression in chondrocytes. *J Biol Chem* **270**, 11727–30 (1995).
69. Stark GR, Kerr IM, Williams BR, et al. How cells respond to interferons. *Annu Rev Biochem* **67**, 227–64 (1998).

# USE OF ANIMAL MODELS FOR PROJECTION OF CLINICAL DRUG–DRUG INTERACTIONS FOR THERAPEUTIC PROTEINS

EUGENIA KRAYNOV and MARTIN E. DOWTY

## 6.1 INTRODUCTION

Animal models play an important role in assessing various aspects of a drug candidate during its discovery and clinical development, including anticipated efficacy, pharmacokinetics, and safety. At the same time, there are limitations with animal models in that they do not always reproduce human pathophysiology or disease drivers, pharmacokinetic pathways, or safety mechanisms relevant to humans. Hence the application of animal models must be underpinned by an understanding of the translatability of relevant mechanisms to humans. The most valuable insight into drug–drug interaction (DDI) potential is predicated on a thorough comprehension of the pharmacokinetic (PK) and pharmacodynamic (PD) properties of both interacting drugs and an understanding of their potential to be a DDI victim as well as a perpetrator.

The use of *in vitro* human systems to predict clinical DDIs of small molecules mediated through cytochrome P450 (CYP) metabolism has seen significant value.[1] However, DDI mediated through non-CYP mechanisms requires greater effort to understand *in vitro–in vivo* correlations (IVIVCs). Although the preclinical DDI assessment for therapeutic proteins can theoretically be addressed by using *in vitro* human systems alone, the relevance of the *in vitro* data in the context of a more complex *in vivo* system requires greater validation. *In vivo* animal models can potentially help bridge this gap. However, additional work is needed to address the potential for animal models to provide insight into DDI predictions of therapeutic proteins.

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*,
Edited by Shayne C. Gad
Copyright © 2013 John Wiley & Sons, Inc.

## 6.2 SELECTION OF THE ANIMAL MODEL

In general, a valid animal model should at least have qualitative similarity to humans in drug-metabolizing enzymes and transporters, substrate specificity, response to the inhibitor or inducer (small molecule or therapeutic protein), and disposition mechanisms of both small molecule and therapeutic proteins. These considerations depend on the breadth of questions being asked.

When it comes to evaluation of therapeutic protein–small molecule DDIs, animal models may be used to qualitatively characterize an alteration in absorption, tissue distribution, excretion, or inhibition and induction of metabolism of the small molecule (SM) drug.[2] These processes reflect changes in uptake or efflux transporters and drug-metabolizing enzymes. However, all animal species differ from humans in amino acid sequence identity and substrate specificity of drug-metabolizing enzymes and transporters. The rat is a particularly common species used in preclinical development, although rats have a greater variability in expression and extent of activity in phase I and II drug-metabolizing enzymes and typically metabolize SM drugs considerably faster than do humans.[3–5]

More similarities are found in species more closely related to humans in an evolutionary sense, in particular monkeys.[2] Rhesus and cynomolgus monkeys are used routinely for preclinical pharmacokinetics, pharmacodynamics, and safety evaluations and more recently for the study of *in vivo* drug interactions. For example, cynomolgus monkey CYP3A8 shows 93% homology to human 3A4 (a significant metabolizing enzyme for small molecules), and rhesus monkey CYP3A64 has been shown to be 93% homologous to human CYP3A4 and 83% homologous to CYP3A5.[6]

Genetically modified mouse models in which a specific gene is removed or replaced are powerful tools for various mechanistic studies.[7,8] However, there are important considerations in their successful implementation. For example, modification of a given gene does not always result in the anticipated phenotype. Therefore, interpretation of the significance of the findings from studies using genetically modified mouse models is not always as straightforward as one would expect, especially when it is desired to extrapolate the findings to humans. Nevertheless, knockout and transgenic animal models can play an important role in understanding mechanisms and may become more useful in the future.

For monoclonal antibody (mAb) based therapeutic proteins, species selection depends on mAb cross-reactivity because target-mediated clearance often plays a role in their elimination. A general limitation of rodent models is that therapeutic proteins intended for human use may not cross-react with the rodent target. A rodent-specific surrogate therapeutic protein may provide value in this regard but requires additional effort to develop and assess. In many cases, the pharmacokinetics of therapeutic proteins and its extrapolation to humans is best understood in monkeys because the therapeutic proteins are typically cross-reactive to monkey targets.

Currently no preclinical *in vivo* DDI evaluation is typically being performed for therapeutic proteins. However, given the aforementioned, in most cases the monkey is probably the most appropriate species for preclinical *in vivo* characterization of

DDIs involving therapeutic proteins if such evaluation is attempted. Monkeys generally need to be naive—that is, not previously treated with biotherapeutics (BioTx) of a similar class, to avoid formation of antidrug antibodies (ADAs) confounding the interpretation of these studies.

## 6.3 STUDY DESIGN

When a therapeutic protein is designed to be coadministered with a narrow therapeutic index (TI) drug, clinical DDI studies may need to be conducted even when there is no known mechanism of interaction. A preliminary evaluation of this interaction in the appropriate animal DDI model can aid in the design of the clinical study. Preclinical *in vivo* evaluation can be conducted for any therapeutic protein intended to be used in combination (with small molecule or another protein therapeutic drug). Combination therapy is now a standard approach in oncology and frequently used in a number of inflammatory and autoimmune diseases.

For a therapeutic protein intended to be used as a single agent and for which no known DDI mechanisms exist, it would be difficult to design an appropriate preclinical DDI study, and in most cases such evaluation is not needed.

Multiple aspects of therapeutic protein PK need to be taken into account when designing an animal DDI study with a small molecule drug. In the case of mAbs, for example, the differences in half-life—typically days for mAbs vs. only hours for small molecules—can affect the time of assessment. An interacting small molecule drug could also alter the expression of the mAb target over time or affect the disposition of the mAb. For mAbs undergoing significant target-mediated elimination characterized by nonlinear pharmacokinetic profiles, the relevance of the animal species and the study design would need to account for this interaction.[9]

From a bioanalytical perspective, the potential interference of each concomitant drug on the quantification of the other should be investigated. Quantification of small molecule drugs typically uses liquid chromatography coupled with tandem mass spectrometry (LC-MS/MS) assays, whereas analytical methods for detection of therapeutic proteins are mostly based on ligand-binding assays. The specificity and selectivity of ligand-binding assays largely rely on the reagents used in the assays and the matrices in which the samples are collected. Because there is no extraction step included in this procedure, coadministered drugs could potentially interfere with the quantification of the therapeutic protein of interest.

## 6.4 DISEASE MODELS

The only preclinical DDI evaluation being currently performed for therapeutic proteins is for cytokine-mediated DDI using human hepatocytes.[10] However, even this approach is still being optimized, and its translation to the clinical situation is not well understood.

Cytokine-mediated changes in drug-metabolizing enzymes and resulting DDIs with therapeutic proteins can potentially be evaluated in the disease models of inflammation. The adjuvant-induced inflammatory rat model shares common features with human rheumatoid arthritis with respect to development of systemic inflammation and structural changes in bone and cartilage.[11] Adjuvant treatment causes cytokine activation, resulting in a number of hepatic events, such as defenestration of sinusoids, irregular-shaped lysosomes, increased level of α1-acid glycoprotein, and lower CYP levels compared to normal rats.[12–14]

Inflammation alters expression of hepatobiliary transporters: Expression of *Oatp1a1, Oatp1a4*, and *Bsep* is downregulated in adjuvant-treated rats, while the expression of *Mrp3* is upregulated nearly by twofold.[15] Inflammatory cytokines—including tumor necrosis factor α (TNFα), interleukin 1 (IL-1), and IL-6—secreted by macrophages, and Kupffer cells have been characterized as mediators that reduce bile flow and organic anion excretion.[15]

These effects are similar to cytokine effects observed with human hepatic enzymes and transporters *in vitro* [16,17] and may at least partly reproduce the systemic pathophysiology of human rheumatoid arthritis.

In human systemic inflammatory disease, the clearance of drugs that are extensively metabolized in the liver is decreased.[18,19] Anti-inflammatory drugs should restore or normalize the amount and activity of microsomal enzymes. This was nicely demonstrated in the preadjuvant arthritis rat model after treatment with anti-TNFα mAb, infliximab.[20] The same authors earlier reported that preadjuvant arthritis is associated with elevated serum nitrite, reduced hepatic CYP content, and reduced verapamil clearance.[18] The clearance of verapamil is also reduced in patients with rheumatoid arthritis.[18] Treatment with infliximab resulted in a significant increase in total CYP, CYP1A, and CYP3A protein content in preadjuvant arthritic rats. Interestingly verapamil clearance was not significantly affected, which the authors attributed to the lack of effect on plasma protein binding.

Cytokine-mediated DDIs are typically studied *in vitro* in human hepatocytes,[16,17] and clinical DDI studies are usually conducted for high-risk therapeutic proteins. However, the IVIVC in humans is not well understood. Conceptually, this IVIVC link could be initially established using rat models of inflammation. However, the interspecies relationship may be difficult to characterize, and this may therefore reduce the predictive power of animal models. Specifically, rat and human cytokines have sufficiently low homology (e.g., mature human IL-6 shares 41% amino acid sequence identity with mouse and rat IL-6),[21] and therefore antihuman cytokine therapeutic proteins may not recognize the homologous target in a rat model. In addition, human therapeutic are likely to be immunogenic in rodents.

At present it is fair to say that while inflammatory rodent models appear to be rather straightforward, further characterization is needed.

In cases when DDI occurs via more than one pathway, animal models may be better able to reflect a net outcome of the interplay between those human-relevant pathways that otherwise would be difficult to predict from *in vitro* assays by

themselves. Even if quantitative prediction from animal models is not understood, it may still be a better qualitative predictor of potential DDI than *in vitro* studies alone. Similar to small molecule DDI, linking *in vitro* data in human and animal models to *in vivo* animals systems, may allow for more predictive DDI characterization.

## 6.5 EMERGING CHALLENGES

The structural diversity of therapeutic proteins is constantly expanding. Drug modalities representing hybrids of small molecule and proteins pose unique challenges from the DDI perspective. Included in this class are antibody-drug conjugates (ADCs) and various nanoparticles. In the case of ADCs, a small molecule is linked to a carrier antibody that provides targeting specificity through recognition of a specific target on the cell surface. As the antibody–receptor complex gets internalized, the small molecule is delivered to the target tissue where it is released. Nanoparticles can have a small molecule drug encapsulated inside while on the outside they can have specific peptides or proteins, such as mAbs that specifically deliver the small molecule to the target tissue.

These newer therapeutic protein modalities containing small molecule drugs may have a higher probability of being involved in DDI with conventional small molecule drugs. This type of DDI is more challenging to characterize *in vitro* owing to the complex and targeted nature of these drug products. One option would be to evaluate the small molecule component's DDI potential separately from the carrier or targeting moiety. However, this would not truly represent the *in vivo* situation and therefore might not adequately predict clinical behavior. Preclinical animal models may be helpful in the initial characterization of these modalities.

Biodistribution of these compounds should probably be evaluated as part of their DDI evaluation. Factors that influence tissue distribution of therapeutic proteins include size, charge, target and off-target binding properties, and route of administration.[22] For proteins that are above the glomerular filtration cutoff (such as immunoglobulins), the major elimination organs are liver and other highly perfused organs. These proteins cannot easily pass through biological membranes and distribute primarily into the vascular space and extracellular fluids.

## 6.6 CONCLUSIONS

It is important to appreciate that findings from animal DDI models should initially be characterized for qualitative information only. An appropriately selected and used animal model can be a valuable tool for DDI assessments during the drug discovery and development processes. Although clinical DDI assessment remains a necessary part of the development of therapeutic proteins, additional insight and characterization of preclinical tools, including animal models, may provide greater value in the future.

## REFERENCES


1. Obach RS. Predicting drug-drug interactions from in vitro drug metabolism data: challenges and recent advances. *Curr Opin Drug Discov Devel* **12**, 81–89 (2009).
2. Tang C, Prueksaritanont T. Use of *in vivo* animal models to assess pharmacokinetic drug-drug interactions. *Pharm Res* **27**(9), 1772–87 (2010).
3. Nedelcheva V, Gut I. P450 in the rat and man: methods of investigation, substrate specificities and relevance to cancer. *Xenobiotica* **24**, 1151–75 (1994).
4. Smith DA. Species differences in metabolism and pharmacokinetics: are we close to an understanding? *Drug Metab Rev* **23**, 355–73 (1991).
5. Soucek P, Gut I. Cytochromes P450 in rats: structure, functions, properties and relevant human forms. *Xenobiotica* **22**, 83–103 (1992).
6. Carr B, Norcross R, Fang Y, et al. Characterization of the rhesus monkey CYP3A64 enzyme: species comparisons of CYP3A substrate specificity and kinetics using baculovirus-expressed recombinant enzymes. *Drug Metab Dispos* **34**,1703–12 (2006).
7. Ashino T, Arima T, Shioda S, et al. Effect of interleukin-6 neutralization on CYP3A11 and metallothionein-1/2 expressions in arthritic mouse liver. *Eur J Pharmacol* **558**, 199–207 (2007).
8. Lin JH. Applications and limitations of genetically modified mouse models in drug discovery and development. *Curr Drug Metab* **9**, 419–38 (2008).
9. Mager DE. Target-mediated drug disposition and dynamics. *Bioch Pharmacol* **72**, 1–10 (2006).
10. Lee J-I, Zhang L, Men AY, et al. CYP-mediated therapeutic protein-drug interactions: clinical findings, proposed mechanisms and regulatory implications. *Clin Pharmacol* **49**(5), 295–310 (2010).
11. Van Bilsen JH, Wagenaar-Hilbers JP, Boot EP, et al. Searching for the cartilage-associated mimicry epitope in adjuvant arthritis. *Autoimmunity* **35**(3), 201–10 (2002).
12. Monshouwer M, Hoebe KH. Hepatic (dys-)function during inflammation. *Toxicol In Vitro* **17**, 681–86 (2003).
13. Whitehouse MW, Beck FJ. Impaired drug metabolism in rats with adjuvant-induced arthritis: a brief review. *Drug Metab Dispos* **1**, 251–55 (1973).
14. Hung, DY, Siebert GA, Chang P, et al. Hepatic pharmacokinetics of propranolol in rats with adjuvant-induced systemic inflammation. *Am J Physiol Gastrointest Liver Physiol* **290**, G343–51 (2006).
15. Roberts MS, Liu X, Zou Y, et al. Effect of adjuvant-induced systemic inflammation in rats on hepatic disposition kinetics of taurocholate. *Am J Physiol Gastroint Liver Physiol* **300**, G130–36 (2011).
16. Zhang X, Schmitt C, Grange S, et al. Disease-drug interaction studies of tocilizumab with cytochrome p450 substrates *in vitro* and *in vivo*. *Clin Pharmacol Ther* **85**, S59 (2009).
17. Huang S-M, Zhao H, Lee J-I, et al. Therapeutic protein-drug interactions and implications for drug development. *Clin Pharmacol Ther* **87**(4), 497–503 (2010).
18. Mayo PR, Skeith K, Russell AS, Jamali F. Decreased dromotropic response to verapamil despite pronounced increased drug concentration in rheumatoid arthritis. *Br J Clin Pharmacol* **50**, 605–13 (2000).

19. Ling S, Jamali F. Effect of early phase adjuvant arthritis on hepatic P450 enzymes and pharmacokinetics of verapamil: an alternative approach to the use of an animal model of inflammation for pharmacokinetic studies. *Drug Metab Dispos* **33**, 579–86 (2005).
20. Ling S, Jamali F. The effect of infliximab on hepatic cytochrome P450 and pharmacokinetics of verapamil in rats with pre-adjuvant arthritis: a drug–disease and drug–drug interaction. *Basic Clin Pharmacol Toxicol* **105**, 24–29 (2009).
21. Chiu C-P, Moulds C, Coffman L, et al. Multiple biological activities are expressed by a mouse interleukin 6 cDNA clone isolated from bone marrow stromal cells. *Proc Natl Acad Sci USA* **85**, 7099–103 (1988).
22. Vugmeyster Y, Defranco D, Sckut P, et al. Biodistribution of-labeled therapeutic proteins: application in protein drug development beyond oncology. *J Pharm Sci* **99**(2), 1028–45 (2010).

# THE COCKTAIL APPROACH AND ITS UTILITY IN DRUG–DRUG INTERACTION ASSESSMENTS FOR THERAPEUTIC PROTEINS

ALEXANDER JETTER and UWE FUHR

## 7.1 ASSESSMENT OF ENZYME ACTIVITIES USING THE COCKTAIL APPROACH

The cocktail approach is used to assess the effect of drugs on the activities of the main cytochrome P450 (CYP) enyzmes *in vivo*. As for all other drugs, any effect of therapeutic proteins on the metabolism of other (victim) drugs needs to be tested and quantified. The main mechanisms of interaction for small molecules on CYPs are enzyme inhibition by direct interaction with the enzyme, increased synthesis of enzymes via aryl hydrocarbon (Ah)- or retinoid X receptor (RXR) related processes, or decreased turnover of the enzyme (in the case of the cytochrome P450 enzyme CYP2E1). These mechanisms of interaction are related to the structural similarity of victim substrates and other small molecule drugs. Appropriate *in vitro* tests have been established for enzyme inhibition and, with limitations in quantification of the effect, also for the described mechanisms of enzyme induction. In contrast, therapeutic proteins would rarely have a direct action on the enzyme or would trigger classical enzyme induction, which developed as specific defense mechanisms to protect an organism against exposure to toxic nutritional components. Thus *in vitro* or animal studies with therapeutic proteins have limited power to predict clinical interactions because the molecular mechanisms are only partially understood and/or often cannot be mimicked in nonclinical test systems.[1] Because not all potential combinations of a therapeutic protein with other drugs can be tested in a clinical study, the effect of such drugs on drug metabolizing enzymes, mainly

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*,
Edited by Shayne C. Gad
Copyright © 2013 John Wiley & Sons, Inc.

CYP enzymes, would usually be assessed using the phenotyping approach. Phenotyping for an enzyme—that is, quantification of its actual *in vivo* activity in an individual—is performed by administration of a selective substrate for the enzyme and subsequent determination of appropriate pharmacokinetic metrics closely reflecting enzyme activity.[2] Pharmacokinetic metrics reflecting individual CYP activity best, depending on the individual phenotyping substrate used, may be metabolic ratios of metabolite over parent drug in plasma or urine, clearance of the parent drug, or partial clearance via a specific pathway. It is desirable to use metrics based on a limited number of blood samples. The standard design of respective phenotyping interaction studies is a crossover design with administration of phenotyping agents in both periods, which is combined with coadministration of the therapeutic protein to be tested at its highest chronic therapeutic dose in one period.[2,3]

Concomitant administration of phenotyping substances in a cocktail to test the effects of drugs including therapeutic proteins on several CYPs at a single occasion has been reported extensively.[2,4,5] This approach offers the advantage that the burden for both investigators and volunteers is considerably decreased. In the case of therapeutic proteins, this is especially important because such studies often need to be conducted in patients (discussed later). Beyond the selection of appropriate phenotyping metrics within a phenotyping cocktail, it should be safeguarded that there is no relevant interaction between individual cocktail components.[6] If an interaction between cocktail components cannot be excluded (e.g., chlorzoxazone, a CYP2E1 substrate, is an inhibitor for CYP3A4),[7] it is also possible to split the cocktail, administering the interacting substance on a different day of the study. Tolerability of such drug cocktails is usually excellent because it is possible to select very low doses.[2] Thus the cocktail phenotyping approach is the method of choice for screening for the potential of therapeutic proteins to change CYP activities *in vivo*.

## 7.2 GUIDELINES APPLICABLE FOR COCKTAIL DRUG–DRUG INTERACTION STUDIES

Current regulatory guidelines provide recommendations on individual phenotyping agents.[8,9] Recommended CYP phenotyping agents include CYP1A2, theophylline, caffeine; CYP2B6, efavirenz, bupropion; CYP2C8, amodiaquine, cerivastatin, repaglinide, rosiglitazone; CYP2C9, warfarin, tolbutamide; CYP2C19, omeprazole, esomeprazole, lansoprazole, pantoprazole; CYP2D6, metoprolol, desipramine, dextromethorphan, atomoxetine; CYP2E1, chlorzoxazone; and CYP3A4, midazolam, buspirone, felodipine, lovastatin, eletriptan, sildenafil, simvastatin, triazolam.[8,10] The European Medicines Agency (EMA) list is shorter and more specific. The FDA published an additional list in 2011,[11] presenting substrates considered as sensitive and substrates with a narrow therapeutic range (according to the 1999 document "Guidance for Industry: In Vivo Drug Metabolism/Drug Interaction Studies—Study Design, Data Analysis, and Recommendations for Dosing and Labeling"), which is also incorporated in the 2012 draft guideline.[8,10]

In most cases, several different pharmacokinetic metrics may be used to assess activity of a given CYP. However, the 2010 EMA (draft) guideline on the investigation of drug interactions (CPMP/EWP/560/95/Rev. 1—Corr.[10]) and the updated (draft) FDA guidance[8] request the determination of complete area under the curve (AUC) values of phenotyping substrates. The reason for the reluctance to accept other metrics is that validation is in most cases limited to the quantification of baseline activity in healthy individuals and in the absence of factors with major impact on pharmacokinetics. However, any effects of therapeutic proteins on pharmacokinetic processes may not be limited to the activity of the CYP to be investigated. Additional mechanisms could then be identified by a change in the entire concentration vs. time profile of the phenotyping drug.[3]

Cocktail studies are addressed specifically by the EMA guidance; the FDA standpoint is also available.[8,12,13] The FDA clearly accepts data from cocktail studies. For both agencies, the appropriate composition of a cocktail is a prerequisite.[2] The FDA considers data from cocktail studies to supplement other *in vitro* or *in vivo* information, including physiology-based pharmacokinetic modeling and simulation data. Cocktail studies may also be used to determine the need for further *in vivo* studies. In its most recent (draft) guideline,[8] a further implementation of cocktail studies is supported, and it is mentioned that negative results from cocktail studies may eliminate the need to conduct additional studies. For an investigational therapeutic protein that is a cytokine or cytokine modulator, it is requested that studies should be conducted to determine the effects on cytochrome P450 enzymes (or transporters), and the use of the cocktail approach is mentioned to this end.[8]

## 7.3 COCKTAIL INTERACTION STUDIES WITH THERAPEUTIC PROTEINS: SPECIAL FEATURES

The metabolism and elimination of therapeutic proteins is distinct from that of small molecule drugs, which share metabolism and elimination pathways with most toxic compounds and components of food and environment. Because dietary proteins are predominantly digested in the gut and taken up as amino acids or dipeptides, no particular enzymes or transporters exist for therapeutic proteins. It is therefore unlikely that a therapeutic protein is a substrate of a drug-metabolizing enzyme of phases I and II or of a drug transporter, because the molecules are too large and their structure is not suitable for metabolism or transport by those enzymes and transporters usually encountered in pharmacokinetics. Thus, in general, therapeutic proteins have a lower risk of causing specific metabolic drug interactions mediated by changes in activities of individual enzymes. However, therapeutic proteins may act as modulators of function of these drug-metabolizing and -transporting proteins. Several cytokines, including interleukin 1 (IL-1), IL-6, IL-10, tumor necrosis factor α (TNF-α), and interferon (IFN), have been shown repeatedly to change (usually to downregulate) the expression of various cytochrome P450 enzymes in *vitro*[14–16] and *in vivo*.[17] Accordingly, therapeutic proteins that block the action of such cytokines in pathophysiological processes will also block their effects on drug-metabolizing

enzymes. These cytokine-mediated changes in cytochrome P450 enzyme activities take some time to develop because protein expression is affected. This is in contrast to most enzyme inhibitions by small molecule drugs, which are present almost immediately after the inhibitory drug has been administered, because the underlying mechanism is a direct competitive or a noncompetitive inhibitory interaction with the enzyme itself. Translated to the timing of cocktail dosing this would mean that the administration of the probe substrate cocktail should not take place immediately after the first dose, because the expected effect may require several days and/or repeated dosing until it is fully developed.

Another factor that has to be taken into account is the pharmacokinetics of the therapeutic protein. Because therapeutic proteins are often administered subcutaneously or are chemically modified for increased metabolic stability (e.g., by pegylation) their residence time in the body is often much longer than that of small molecules. Because the therapeutic protein is continuously present in the body for days or even weeks, the effect on drug-metabolizing enzymes and drug transporters may be fully present only after several days or weeks. It is therefore crucial to take the biologic half-life of the compound under investigation into account when the timing of the cocktail-phenotyping days is planned. In addition, it may be wise to consider more than one cocktail-phenotyping day during one treatment course with a therapeutic protein.

Because some therapeutic proteins may cause severe adverse effects, including infections and cancer, their use in healthy subjects is not always justifiable. Thus cocktail interaction studies may need to be conducted in patients. However, such studies in patients may pose major challenges in regard to recruitment of participants, further interactions by concomitant medication, and the variability caused by disease processes. This may also include the amelioration in organ function caused by the therapeutic protein. To give an example, interferons used in the treatment of hepatitis, or therapeutic proteins directed against hepatically metastasizing tumors, may lead to improvements in hepatic function during treatment. This may then also translate into increased activities of CYP enzymes, other drug-metabolizing enzymes, and transporters. Another example is the association between the stage of congestive heart failure, TNF-α and IL-6 concentrations, and their inverse relationship to the activity of the CYP enzyme CYP2C19.[17]

In conclusion, therapeutic proteins may lead to changes in the activities of CYP enzymes by mechanisms different from small molecule drugs. These interactions may take longer to develop (and presumably also longer to subside, but this has not yet been studied), and may also necessitate other designs in cocktail drug–drug interaction studies, particularly because of the long half-life of most therapeutic proteins.

## 7.4 PUBLISHED COCKTAIL INTERACTION STUDIES WITH THERAPEUTIC PROTEINS

In the literature, there are only few drug–drug interaction studies on therapeutic proteins that use the cocktail phenotyping approach, although such an approach has

been endorsed by authors of some review articles in recent years[1,18,19] and by regulatory guidelines on drug–drug interaction studies for drug approval.[8]

A very early example of a cocktail interaction study was carried out with IFN-α in 1993.[20] In that study, a drug cocktail consisting of 150 mg theophylline, 250 mg antipyrine, and 250 mg hexobarbitone was administered to seven cancer patients before, with the first dose, and after 2 weeks of IFN-α dosing. Because these drugs are either no longer in use at all, and/or are not established probe drugs for specific enzymes, the findings are difficult to interpret in terms of generalizability to other drug substrates. The 33% decrease in oral theophylline clearance in this study may be interpreted as a decrease in CYP1A2 activity, the enzyme mainly involved in theophylline clearance; but CYP2A6 may also contribute. Antipyrine is a nonspecific CYP substrate, and the nonsignificant 20–21% decrease in antipyrine clearance both after the first dose and after 14 days of treatment cannot be attributed to a change in a single enzyme activity. Hexobarbitone clearance was unchanged, butbecause hexobarbitone is probably stereoselectively metabolized by various CYP enzymes, the significance of this result remains unclear.

Another, more recent, study investigated the effects of multiple doses of pegylated IFN-α2b on the activities of CYP1A2, CYP2C8/2C9, CYP2D6, CYP3A4/3A5, and *N*-acetyltransferase 2 (NAT2) in 26 patients with chronic hepatitis C infection.[21] The cocktail used in the study consisted of 200 mg caffeine for CYP1A2 phenotyping, 500 mg tolbutamide for CYP2C8/2C9 phenotyping, 45 mg dextromethorphan for CYP2D6 activity assessment, and 100 mg dapsone for the assessment of NAT2 activity, all administered orally and simultaneously, while 4.0 mg midazolam for overall CYP3A phenotyping was administered as a syrup 24 h later. The phenotyping procedures were carried out twice, once before and once after 4 weekly doses of pegylated IFN-α2b adjusted to the body weight of the participants. This cocktail is, according to the authors of the paper, validated but unpublished as "data on file" of the pharmaceutical company that initiated the study. Its composition most resembles the Indianapolis cocktail, with the addition of dapsone as a NAT2 probe substrate, a higher dose of dextromethorphan, and the omission of intravenous midazolam, which is used as a probe for hepatic CYP3A activity.[2]

The authors used appropriate bioequivalence testing methods to calculate point estimates of the geometric means ratios for the selected phenotyping parameters and corresponding 90% confidence intervals. They concluded that multiple doses of pegylated IFN-α2b moderately increased CYP2D6 activity by 67%, but variability was huge (25–123% increase in activity), although genetically poor metabolizers of CYP2D6 had been excluded from study participation. In addition, the authors observed a minor increase in CYP2C8/2C9 activity by 28% (90% confidence interval 15–42%), which was statistically significant but of presumably limited clinical importance, because one border of the 90% confidence interval was within the bioequivalence zone. They did not observe statistically significant changes in overall CYP3A activity and NAT2 activity, but a minimal, but statistically significant decrease in CYP1A2 activity to 92% was noted (90% confidence interval 87–98%).

Since the 90% confidence interval is still entirely within the 80–125% bioequivalence zone, this finding is certainly not clinically relevant.

It is interesting that these findings do not entirely match with the information from studies with pegylated IFN-α2a or IFN-α2b. Namely, CYP2D6 activity is not influenced by pegylated IFN-α2a, while a larger effect on CYP1A2 activity (in the order of a 25% reduction in activity) is mentioned. Possible explanations include the differences in study populations (the present cocktail study was carried out in hepatitis C patients, whereas the studies with IFN-α2a were done with healthy subjects) and an indirect effect mediated by a general amelioration in hepatic function in the patients with chronic hepatitis C by IFN-α2b administration. This study is cited in the U.S. prescribing information, and it is the basis of the precautions for the use of CYP2C9 and CYP2D6 substrates together with pegylated IFN-α2b.

Although this example nicely shows that a cocktail phenotyping study is suitable for the clinical assessment of the relevance of interactions between a therapeutic protein and drug-metabolizing enzymes, there is no other published *in vivo* cocktail interaction study investigating the effect of a therapeutic protein on enzymes and transporters. A search of the U.S. National Institutes of Health clinical study register (www.clinicaltrials.gov) revealed one other cocktail study on the interaction potential of a therapeutic protein. However, that study, which aimed at investigating the effect of briakinumab (ABT-874), a IgG-based monoclonal antibody targeting IL-12/23, was stopped before enrollment, perhaps in the course of the withdrawal of filing of briakinumab at the U.S. and the European medicines agencies because of safety issues. The study was planned to be carried out in patients with plaque psoriasis, the target population for briakinumab, and the scheduled cocktail included caffeine, tolbutamide, omeprazole, metoprolol, and midazolam, for the assessment of the activities of CYP1A2, CYP2C9, CYP2C19, CYP2D6, and CYP3A4. The cocktail was scheduled to be administered twice, once before, and once 14 days after a single dose of briakinumab.[22]

## 7.5 CONCLUSIONS

In summary, it has been shown that the cocktail approach is suitable for the study of interactions between therapeutic proteins and drugs in humans, namely for the study of the effect of therapeutic proteins on CYP enzyme activities. This is the area in which the FDA's current draft guideline on drug interaction studies recommends the use of cocktail phenotyping studies as an alternative to individual therapeutic protein–drug interaction studies.[8]

## REFERENCES

1. Huang SM, Zhao H, Lee JI, et al. Therapeutic protein-drug interactions and implications for drug development. *Clin Pharmacol Ther* **87**, 497–503 (2010).

2. Fuhr U, Jetter A, Kirchheiner J. Appropriate phenotyping procedures for drug metabolizing enzymes and transporters in humans and their simultaneous use in the "cocktail" approach. *Clin Pharmacol Ther* **81**, 270–83 (2007).
3. Zadoyan G, Rokitta D, Klement S, et al. Effect of Ginkgo biloba special extract EGb 761(R) on human cytochrome P450 activity: a cocktail interaction study in healthy volunteers. *Eur J Clin Pharmacol* **68**, 553–60 (2012).
4. Tanaka E, Kurata N, Yasuhara H. How useful is the "cocktail approach" for evaluating human hepatic drug metabolizing capacity using cytochrome P450 phenotyping probes in vivo? *J Clin Pharm Ther* **28**, 157–65 (2003).
5. Zhou H, Tong Z, McLeod JF. "Cocktail" approaches and strategies in drug development: valuable tool or flawed science? *J Clin Pharmacol* **44**, 120–34 (2004).
6. Turpault S, Brian W, Van Horn R, et al. Pharmacokinetic assessment of a five-probe cocktail for CYPs 1A2, 2C9, 2C19, 2D6 and 3A. *Br J Clin Pharmacol* **68**, 928–35 (2009).
7. Palmer JL, Scott RJ, Gibson A, et al. An interaction between the cytochrome P450 probe substrates chlorzoxazone (CYP2E1) and midazolam (CYP3A). *Br J Clin Pharmacol* **52**, 555–61 (2001).
8. U.S. Department of Health and Human Services, U.S. Food and Drug administration, Center for Drug Evaluation and Research, (CDER). Draft guidance for industry: drug interaction studies—study design, data analysis, implications for dosing, and labeling recommendations. Silver Spring, MD, 2012. Available at www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM292362.pdf.
9. European Medicines Agency. Note for guidance on the investigation of drug interactions. London, 1997. Available at www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC500002966.pdf.
10. European Medicines Agency. Draft Guideline on the investigation of drug interactions. London, 2010. Available at www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2010/05/WC500090112.pdf.
11. U.S. Department of Health and Human Services, U.S. Food and Drug administration. Drug development and drug interactions: table of substrates, inhibitors and inducers. Silver Spring, MD, 2011. Available at www.fda.gov/Drugs/DevelopmentApprovalProcess/DevelopmentResources/DrugInteractionsLabeling/ucm093664.htm.
12. Zhang L. A regulatory perspective on the utilization of cocktail approach in assessing drug interactions. Paper presented at the FIP Pharmaceutical Sciences World Congress—American Association of Pharmaceutical Scientists Annual Meeting, New Orleans, LA, November 16, 2010. Available at www.fda.gov/downloads/Drugs/DevelopmentApprovalProcess/DevelopmentResources/DrugInteractionsLabeling/UCM237160.htm.
13. Zhang L, Reynolds KS, Zhao P, Huang SM. Drug interactions evaluation: an integrated part of risk assessment of therapeutics. *Toxicol Appl Pharmacol* **243**, 134–45 (2010).
14. Morgan ET. Regulation of cytochromes P450 during inflammation and infection. *Drug Metab Rev* **29**, 1129–88 (1997).
15. Dickmann LJ, Patel SK, Rock DA, et al. Effects of interleukin-6 (IL-6) and an anti-IL-6 monoclonal antibody on drug-metabolizing enzymes in human hepatocyte culture. *Drug Metab Dispos* **39**, 1415–22 (2011).

16. Morgan ET. Impact of infectious and inflammatory disease on cytochrome P450-mediated drug metabolism and pharmacokinetics. *Clin Pharmacol Ther* **85**, 434–38 (2009).
17. Frye RF, Schneider VM, Frye CS, Feldman AM. Plasma levels of TNF-alpha and IL-6 are inversely related to cytochrome P450-dependent drug metabolism in patients with congestive heart failure. *J Card Fail* **8**, 315–19 (2002).
18. Girish S, Martin SW, Peterson MC, et al. AAPS workshop report: strategies to address therapeutic protein-drug interactions during clinical development. *AAPS J* **13**, 405–16 (2011).
19. Lee JI, Zhang L, Men AY, et al. CYP-mediated therapeutic protein-drug interactions: clinical findings, proposed mechanisms and regulatory implications. *Clin Pharmacokinet* **49**, 295–310 (2010).
20. Israel BC, Blouin RA, McIntyre W, Shedlofsky SI. Effects of interferon-alpha monotherapy on hepatic drug metabolism in cancer patients. *Br J Clin Pharmacol* **36**, 229–35 (1993).
21. Gupta SK, Kolz K, Cutler DL. Effects of multiple-dose pegylated interferon alfa-2b on the activity of drug-metabolizing enzymes in persons with chronic hepatitis C. *Eur J Clin Pharmacol* **67**, 591–99 (2011).
22. Abbott. Open label, single dose, non-randomized study to assess the drug to drug interaction of briakinumab on CYP substrates. 2010. Available at http://clinicaltrials.gov/ct2/show/NCT01260844.

# LOGISTIC CONSIDERATIONS IN STUDY DESIGN FOR BIOLOGIC DRUG–DRUG INTERACTION ASSESSMENTS

MIN ZHU and YU-NIEN (TOM) SUN

## 8.1 INTRODUCTION

Evaluation of the drug interaction potential of biologics or therapeutic proteins (TPs) with other TPs or small molecule drugs is an area of emerging research. Different from the pharmacokinetics (PK) drug–drug interaction (DDI) among small molecule drugs in which there are clear mechanisms associated with metabolic pathways mediated by cytochrome P450 (CYP450) enzymes or transporters, there is limited understanding of the mechanistic basis for TP–drug interactions to date. Therefore the design of such interaction studies is challenging in many cases. As described in the literature,[1–3] the common TP elimination mechanisms may include filtration (e.g., into urine), secretion (e.g., into the bile), biotransformation (e.g., catabolism), and target-mediated disposition. Renal elimination, which is one of the primary pathways of small-molecule drugs, is relatively unimportant for large molecules, such as immunoglobin G (IgGs) with a molecular weight of ~150 kDa that prevents efficient filtration through the glomerulus. Secretion into the bile is an important pathway of elimination of IgA antibodies, but this route is not a significant contributor to the elimination of IgG antibodies. The majority of IgG elimination occurs via intracellular catabolism, following fluid-phase or receptor-mediated endocytosis. TP elimination mechanisms distinguish themselves from the hepatic metabolism that is the dominant mechanism for small molecules. In addition, as TPs may be viewed as foreign substances by the body, immune responses that lead to the generation of endogenous antibodies against the TPs may be triggered, which can increase the clearance of TPs.

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*,
Edited by Shayne C. Gad
Copyright © 2013 John Wiley & Sons, Inc.

According to the BIO Survey in 2010, potential TP-DDIs were most evaluated in the therapeutic areas of immunology and oncology.[4] More specifically, TPs as anti-inflammatory agents were often evaluated as potential victims for DDIs. The activities for selected cytochrome P450 enzymes may be altered in the presence of TPs that are cytokines, cytokine modulators, or hormones.[5] Drug- and disease-mediated TP-DDIs have been observed for tocilizumab, which modulates CYP enzymes in patients with rheumatoid arthritis via interleukin 6 (IL-6) specific blockage.[6] It has been demonstrated that methotrexate, an immunosuppressant, will decrease the formation of antidrug antibodies against infliximab and therefore increases its overall exposure.[7] Anticancer TPs were mostly evaluated as potential perpetrators to concomitant medications, particularly for small molecules that are intended to be used in combination therapy and have a narrow therapeutic window (e.g., chemotherapeutics with cytotoxic properties). For example, a DDI evaluation was performed for irinotecan plus bevacizumab in patients with solid tumors, and no significant effect of bevacizumab on the pharmacokinetics of irinotecan or its active metabolite SN38 was observed.[8]

Due to the unique characteristics of biologics in pharmacokinetics and immunogenicity, some design parameters of TP–drug interaction studies are different from those for small molecules. A risk-based drug–drug interaction assessment strategy for TPs should be implemented according to its risk factors and appropriate study designs should be selected.[5,9] The challenges and logistic considerations in study designs for TP–drug interactions are discussed in this chapter.

## 8.2 CHALLENGES IN THE CONDUCT OF A TP–DRUG INTERACTION STUDY

Therapeutic proteins can be roughly divided into two groups according to indications–namely, oncology TPs (for cancer treatment) and nononcology TPs (or rheumatoid arthritis, psoriasis, asthma, prevention of cardiac ischemic complication or other cardiovascular disease). The challenges in the assessment of TP–drug interactions for the two groups of TPs are discussed separately.

In oncology, drug combination (e.g., TP with chemotherapy) is commonly used for the improvement of anticancer activity. With respect to TP–drug interaction assessments, *in vitro* screening–guided drug interaction studies would have limited value because there is often no known underlying mechanism. Furthermore, the TP–drug interaction study is usually not a standalone but rather a substudy of Phase I–III clinical trials conducted in cancer patients in conjunction with efficacy and safety assessments. As a result, conventional well-controlled DDI study conditions (e.g., a controlled diet, comedications, disease status, dose, dosing time, and schedule) are rarely preserved. TP pharmacokinetics data collection is usually less intensive than that for small molecule drug interaction studies. The number of subjects may not be well balanced across treatment cycles or arms, and the number of qualifying subjects is also affected by deviations of scheduled dose and sampling times due to dose interruptions and reductions or patient discontinuations.[10] Challenges

exist in performing noncompartmental pharmacokinetics data analysis because there are often a significant number of data exclusions. Pharmacokinetics assessments are further complicated due to a number of oncology TPs or small molecule drugs exhibiting nonlinear pharmacokinetic properties. Moreover, because a TP-alone arm is uncommon in a drug combination oncology trial, historical data are usually referenced in the TP–drug interaction evaluation either by cross-study comparison or by population pharmacokinetic modeling. Thus a scientifically sound and technically feasible integrated TP–drug interaction evaluation strategy needs to be developed to address these challenges.[11]

In contrast, for nononcology therapeutics such as those in the area of immunology and anti-inflammation, TPs can be used as a monotherapy agent. The concomitant drugs can be randomly prescribed medications for symptom relief or for treatment of other comorbidity conditions according to individual patient's needs. The most documented TP–drug interaction involves an indirect mechanism of cytokine-mediated changes in drug metabolizing enzymes that is secondary to changes in inflammation status. Studies have demonstrated the effect of individual cytokines and their modulators on CYP enzymes and transporters.[6,12–14] In this regard, the TP plays the role of a perpetrator and the small molecule drugs are usually victims. This type of interaction is not solely drug related but also disease related. To date, inflammation- or cytokine-mediated inhibition of CYP activities has become an evolving research area for understanding the mechanism of disease–drug interactions and the associated effect on TP–drug interactions.[15] The challenge in the design of such TP–drug interaction studies is that both the inflammation stage and the concomitant drugs can be moving targets. *In vitro* experiments by using different cytokines and selected animal models at the early stage of drug development may provide some insight into the pharmacokinetic interaction potential for coadministered drugs that mainly undergo the CYP-mediated drug disposition, but the magnitude and clinical relevance of the interaction may still need to be evaluated in the target patient population. A strategy of early establishment of a therapeutic exposure range under TP monotherapy may be important for anti-inflammatory agents to assess extent and clinical significance of changes in drug exposures.

Additional logistic considerations of TP–drug interaction assessments are provided in Table 8-1. Clearly, because the assessment is related to the impact on drug efficacy and safety, early strategy planning is essential to reduce logistic burdens and meet regulatory requirements. The recently released "Draft Guidance for Industry: Drug Interaction Studies–Study Design, Data Analysis, Implications for Dosing, and Labeling" from the FDA, clearly indicates that (1) if an investigational TP is a cytokine or cytokine modulator, studies should be conducted to determine the TP's effect on CYP enzymes or transporters; (2) for TPs that will be used in combination with other drugs as a combination therapy, studies should evaluate the effect of each product on the other; and (3) when there are known mechanisms or prior experience with certain pharmacokinetic or pharmacodynamic interactions, appropriate *in vitro* and *in vivo* assessments for possible interactions should be conducted.[16]

**TABLE 8-1 Logistic Considerations of TP–Drug Interaction Assessment Approach**

| Approach of Assessing TP–Drug Interaction | Pros | Cons |
|---|---|---|
| No evaluation of TP–drug interaction potential | No burden on clinical trial operation and costs | Lack of data to assess potential TP–drug interaction risk and impact on safety and efficacy<br>Lack of data to fulfill regulatory requirements |
| Test of TP–drug interaction potential in healthy subjects | Traditional drug interaction designs and conditions can be applied<br>Low operational burden | Drug combination may not be safe to be tested in healthy subjects (e.g., anticancer drugs)<br>Results may not be applicable to target patient population owing to disease–drug interaction (e.g., anti-inflammatory agents) |
| Estimation of TP–drug interaction in early drug development stage (e.g., Phases I and II) | Data may be used to address potential DDI issue and assist in better design of later stage clinical trials<br>May be executable with limited number of patients and at low cost<br>May identify large signal with an estimation study | Adds some operational burden to clinical trials<br>All tested drugs, including marketed drugs, will be treated as IMPs in Europe and governed by associated regulations<br>Estimation study may not have enough power to identify a mild interaction signal |
| Estimation of TP–drug interaction in late drug development stage (e.g., Phase III) | "No DDI findings for the tested drugs" may be included on a product label | Adds significant operational burden and cost to clinical trials<br>All tested drugs will be treated as IMPs in Europe<br>Study conditions may not be controlled consistently across countries in a global trial<br>Study noise may be increased in a large global trial for detecting a mild interaction signal |
| Dedicated TP–drug interaction study as postmarketing | Study can be powered to address the TP–drug interaction question definitively | Needs an independent study<br>High operational burden and cost to clinical trials |

**TABLE 8-1** (*Continued*)

| Approach of Assessing TP–Drug Interaction | Pros | Cons |
|---|---|---|
| commitment or during drug development | A claim of "lack of TP–drug interaction" may be acceptable on a product label | Patient enrollment may be low owing to intensive PK sample collection and other logistic issues<br>Data evaluability may not be high because of patient dropout and change in dose and regimen owing to side effects (e.g., oncology drugs) |

DDI: drug–drug interaction; IMP: investigational medicinal product; PK: pharmacokinetics; TP: therapeutic protein.

## 8.3 TP–DRUG INTERACTION STUDY DESIGN

Unlike a small molecule drug–drug interaction study in which a two-way drug interaction (i.e., A → B *and* B → A) can be evaluated in the same study, a TP–drug interaction assessment may need to be evaluated as a one-way interaction (i.e., A → B *or* B → A), especially in oncology. Therefore, an important consideration in the design of a clinical TP–drug interaction study is to clearly define the aim of the study and then design the study components accordingly. As mentioned in the previous section, most of the TP–drug interaction studies for oncology and anti-inflammatory therapeutics need to be conducted in patient populations and strictly controlling the potential influential factors (e.g., other comedications, dose and schedule, and disease status) may be challenging. Thus conventional drug interaction study settings generally applicable to most small molecule drugs may not be practical for the TP–drug interaction assessment.

In the oncology research area, because there is no clear mechanism-based TP–drug interaction identified yet, the value of *in vitro* screening and preclinical assessments for guiding a clinical TP–drug interaction study planning has not been well established. To date, inclusion of clinical TP–drug interaction components in Phase I–III studies in patient populations for fixed drug combinations has become a common approach for screening purposes and for identifying if exposures of any drug in a fixed combination are considerably affected, especially for drugs with narrow therapeutic windows. The study is usually performed on a small scale for providing initial data for assessing the risk level of a PK interaction. If a high risk is identified, a confirmative study or formal TP–drug interaction study may be triggered, depending on the clinical relevance of the interaction. This approach appears to be accepted by the regulatory agencies.

In the nononcology research areas, *in vitro* screening and *in vivo* cocktail studies may be applicable for risk assessments at a case-by-case basis. With the *in vitro* evaluation, the impact of cytokines and cytokine modulators on CYP enzymes, mRNA, protein, and activities can be assessed in a concentration-dependent manner. Nevertheless, there are many uncontrolled factors in the experimental conditions, and the *in vitro* system could be donor dependent. Uncertainty of the translatability of nonclinical results to clinical impact may be high.[17] With respect to the *in vivo* cocktail approach, a number of probe drugs for enzymes and transporters can be administered together for simultaneous assessment of interaction between small molecule drugs and TPs.[15] The approach may be efficient for screening TP–drug interaction potential and may be used when no *in vitro* assays can be properly established or for verifying *in vitro* screening results or for planning further clinical assessments. The challenges of this approach may be related to the lack of ideal probe substrates, the validity of cocktail studies, the composition of the cocktails, the extrapolation of results to other drugs, and the differentiation of PK and pharmacodynamic (PD) interactions. Further TP–drug interaction studies may be needed for labeling purposes to provide the magnitude of change in exposure for a particular drug of interest in the cocktail.

With respect to clinical TP–drug interaction assessments, there are at least three types of studies commonly employed–namely, screening studies, estimation studies, and dedicated studies–depending on the purpose of the assessment, phase of clinical development, disease areas, TP–drug interaction risks, and sample size. The pros and cons of these types of study are discussed in the following sections.

### 8.3.1 Screening Study for TP–Drug Interactions

A screening study is usually conducted in patients with a relatively sparse PK sampling scheme to determine if any TP–drug interaction signals are apparent. The study is commonly designed for TPs for which the mechanism of interaction is unknown or an interaction is not expected based on the clearance mechanism of the drugs involved.

The screening study for oncology therapeutics aims at testing the effect of the coadministered drug in a fixed combination with the TP ($D \rightarrow TP$). Sparse TP PK samples are usually collected from all subjects in a trial for evaluating the safety or efficacy while the PK samples of concomitant drugs in the fixed combination are not collected. The TP concentration time data are analyzed preferably via population PK modeling with pooled historical data of monotherapy and the data from the combination study to assess covariate effects of coadministered drugs on the TP clearance. The assessment is feasible in most of clinical trials for screening signals and to check if the TP clearance was considerably affected. If any concerns are raised based on an observed change in drug exposure, additional data analysis may be required, such as including data from additional trials to examine confounding factors because the change in exposure may be the result of not only a TP–drug interaction but also of other factors (e.g., cancer type or disease stage). Besides population PK analysis, summary statistics for comparison of observed TP

concentrations in a drug combination study with those in monotherapy studies are also commonly performed to identify a TP–drug interaction signal. The screening study is usually conducted during Phases I and II trials. Further TP–drug interaction studies may be needed if clinically meaningful changes in TP exposure was to be identified. For nononcology TPs, cocktail studies can be used for screening in order to test the effect of the TP on small molecule drugs (TP → D) because the small molecule drugs are commonly the victims.

### 8.3.2 Estimation Study for TP–Drug Interaction

An estimation study is commonly used for a TP–drug interaction assessment when the mechanism of PK interaction is unknown or the interaction is not expected based on the clearance mechanisms of the drug involved. Unlike a screening study, an estimation study collects PK profiles over a specified interval for concomitantly administrated drugs used in combinations, either for small or large molecule drugs, and relative sparse PK samples (e.g., before and after infusion) for the tested TP. The aim of the study is to evaluate the TP–drug interaction potential with a predefined sample size, which is commonly used for the evaluation of the effect of a TP on coadministered drugs (TP → D). Nevertheless, the sample size for an estimation study is usually 6–15 evaluable subjects per arm, depending on the intersubject variability in drug exposure. This is not a powered study for hypothesis testing. The advantage of this kind of study is to test TP–drug interactions in a small group of patients so the operational challenge is relatively low while it still provides evidence for risk assessment. If remarkable changes in drug exposure are identified, the study data can be used for hypothesis generation and provide valuable information for the design of a formal study. The limitation of this type of study is that the effect of coadministered drugs on TPs (D → TP) may need to be assessed by cross-study comparison or population PK modeling if there is no TP alone arm or testing period in the trials.

### 8.3.3 Dedicated Study for TP–Drug Interaction

The dedicated study for the TP–drug interaction assessment is not frequently conducted for oncology TPs but may be conducted for nononcology TPs. In this type of study, intensive PK samples for both TP and comediations are collected. There is not much difference between the dedicated TP–drug interaction study and the traditional drug–drug interaction study for small molecule drugs except the TP–drug interaction studies are mostly conducted in patient populations to test the effect of drug–disease interactions (e.g., inflammation) or fixed drug combinations (e.g., oncology). The study aims are hypothesis testing with sufficient detection power calculated according to results obtained from a screening study or estimation study. Although this type of TP–drug interaction study can provide confirmative results, the logistic issues are apparent as it usually requires much larger sample size, depending on intersubject variability of TPs, the predefined no-effect boundary, and the rates of evaluable individuals after accounting for patient dropout, dose reduction, and dose interruptions.

## 8.4 TIMING OF TP–DRUG INTERACTION STUDY

The question when to conduct the TP–drug interaction study may be associated with therapeutic areas and the risk level of individual therapeutic proteins. For instance, in the oncology area, it is preferable to include TP–drug interaction assessments in the early Phases I and II trials because this can be the first time that a drug combination is tested, trials are relatively small in size, and operational issues can be addressed easier. It is important that obtaining the interaction data early can help establishing a development strategy and guide later phase study designs. If clinically meaningful interaction signals are identified in early trials, expanded assessments may be further considered. The limitation of early assessments is that the optimal drug combination may be modified in later stages of development or the tested drug combination may not proceed to later phase due to ending of the development program in certain indications. Although it is possible to test TP–drug interactions in late phase studies for product registration, it significantly increases the complexity of pivotal Phase III trials and is thus likely to have more confounding factors and causes difficulty in data interpretation.[10]

In contrast, for TPs for the treatment of immunology- and inflammation-related diseases, most clinical trials in the early phase of clinical development are monotherapy trials. The risk in a change in drug exposure may be from an inflammation-mediated disease–drug interaction rather than a direct TP-mediated PK interaction with randomly prescribed medications. However, because the treatment of inflammation with TPs may affect the magnitude of disease–drug interactions, the treatment-associated TP–drug interaction effect may still need to be monitored if patients are taking narrow therapeutic drugs along with the TPs. Owing to the complexity of the interaction, this type of TP–drug interaction study is not easy to design becuase the extent of interaction could change with treatment and disease status. Standalone trials can be conducted for nononcology TPs by using probe substrates (e.g., midazolam for CYP3A4) to examine the risk level. This type of study is commonly planned in the middle or late phase of development or the postmarketing stage based on clinical significance of effect.

## 8.5 STRATEGIC PLANNING OF TP–DRUG INTERACTION STUDIES

Owing to uncertainties in drug development, strategic planning of TP–drug interaction assessments at the nonclinical, clinical Phases I–III, and postmarketing stages is important for an increase of research effectiveness. A stepwise strategy may be needed for planning purposes, which should be focused on the therapeutic area and specific drug, according to the characteristics of the molecules. To elaborate this, examples for general TP–drug interaction assessment strategies in the oncology and anti-inflammation areas are further discussed.

### 8.5.1 Oncology

The TPs for cancer treatment are frequently used in fixed drug combinations (e.g., with chemotherapy) for enhancement of treatment effect and survival benefit. Because there is no known mechanism-based TP–drug interaction identified to date, either for TPs as victim or perpetrators, the study planning heavily relies on clinical screening with data from both monotherapy and fixed drug combination trials for initial risk assessment. There is no one size fits all strategy in such assessments. Although, there is no clear interaction mechanism, evidence of lack of interaction is still required by regulatory agencies.

The strategy of interaction assessment should be clearly described in a clinical pharmacology plan of an investigational TP at the early phase of drug development. The plan may include the purpose of assessment (e.g., evaluate the TP as a victim or a perpetrator, or both), method of assessment (e.g., population analysis approach, noncompartmental analysis, or a combined approach), criteria of analysis (e.g., what is the clinical significance with a defined magnitude change in exposure?), study planning (alignment in study conditions, sampling strategy, and study population across studies), data integration plan (e.g., identification of studies from which data will be used for the TP–drug interaction evaluation), covariate data collection plan (guarantee that common covariates are collected consistently across studies), communication strategy (alignment with clinical teams), and labeling strategy. It is preferable to collect the TP–drug interaction data in Phases I and II; avoiding Phase II, however, reduces operational challenges. Strategic and systemic planning of TP–drug interaction assessment is important in the oncology area because a cross-study comparison is hardly avoidable, especially for TPs with long half-lives.

For small molecule drugs (e.g., chemotherapy agents) that are commonly used in combination with TPs for cancer treatment, it would be useful to establish an integrated internal database across different oncology programs for effectively using all available data.[18] For instance, it is common that selected chemotherapy backbone treatments (e.g., paclitaxel, cisplatin) are tested repeatedly for the same or similar indications (e.g., lung and colon cancers) for different investigational TPs. However, the sample size of each trial might be small (e.g., $n = 5$–7 evaluable subjects) owing to operational difficulties, and missing samples are common because patients are unable to go back to clinics to have blood drawn and often have dose delays or reductions. Therefore, it would be useful to collect PK data of a specific chemotherapy agent with a standardized sample collection scheme for all oncology programs and the data from the control arms (e.g., paclitaxel backbone cohort) of each study in the same patient population may be combined for data analysis to increase detection power and establish a no-effect boundary for a test agent.[19,20]. The established exposure range can be used for all other relevant studies in which the agent was included in drug combinations.

For TPs, PK data will be collected from most subjects, thus sample size will not be an issue. As a result of rare standalone drug interaction trials, TP monotherapy data, usually collected in the first-in-human study, need to be included in the data analysis. It is important to predefine a no-effect boundary as a reference exposure range for

each drug in a drug combination based on pharmacokinetics, safety, and efficacy information. If the exposures in the testing cohort were within the ranges, lack of TP–drug interaction effect may be claimed. Otherwise, a detailed covariate analysis will be performed to identify if the change in exposure is owing to a TP–drug interaction or nondrug interaction factors (e.g., cancer type, disease stage, biomarker related subpopulation). If there is clear evidence that the change in exposure is due to nondrug interaction factors, lack of TP–drug interaction may be claimed. If the change in exposure could be the result of a TP–drug interaction and the magnitude of change may have a clinical consequence, further evaluation may be considered. All evaluation results of different TP–drug combinations in a TP development program can be summarized to provide an overall picture of interaction potential for an investigational TP. The results of the TP–drug interaction assessment via planned analysis should be provided on the drug product label with a detailed description of how data were collected and analyzed. This approach may be applicable not only to oncology TPs but also to nononcology TPs that have no clear mechanism of interaction.

### 8.5.2 Immunology and Inflammation

In contrast, in nononcology areas, such as immunology and anti-inflammation, the TPs are frequently used as monotherapy. Although *in vitro* screening and *in vivo* cocktail approaches may be useful for initial risk assessment, clinical studies under actual treatment conditions in target populations will provide more convincing evidence. Because drugs mainly metabolized by CYP enzymes may be affected by the disease–drug interaction, adequate recording of concomitantly used drugs over a sufficiently long period (e.g., >6 months) may provide some evidence for assessing risk of TP–drug interactions. Covariate analysis via population PK modeling is commonly used for identifying the interaction signals with the individually prescribed and long-term used comedications.[15] If a significant effect is identified and change in exposure may have a clinical consequence, further evaluation of a drug interaction may be considered.

## 8.6 CONSIDERATIONS IN STUDY DESIGN

### 8.6.1 Study Population

In drug–drug interaction studies for small molecules, it is acceptable to conduct the study in healthy subjects because the purpose of the study is to evaluate the direct effect of one drug on the disposition of another concomitantly administered drug for which drug metabolism and elimination pathways may be affected. However, the assessment of oncology TP–drug interaction is rarely conducted in healthy subjects owing to safety considerations. In the immunology therapeutic area, because inflammation-mediated disease–drug interaction governs the mechanism of TP–drug interaction, the target patient population has to be evaluated to assess the magnitude of disease–drug interaction over time. As a result, the conduct of TP–drug interaction studies in target patient populations is appropriate in most cases.

### 8.6.2 Selection of Interacting Drugs

The selection of interacting drugs for investigation is mainly based on the potential of concomitant usage, PK, and PD properties and the therapeutic window of the selected drugs. For instance, if an investigational TP will be used in a combination of two chemotherapy agents and one supportive hormonal agent, and the hormone drug has a wide safety window and is prescribed at a low dose, the selected testing drugs for the interaction study can be the TP and the two chemotherapy agents. In the inflammation area, because each patient may be taking his or her own long-term medications along with the TP treatment, the drugs with narrow therapeutic windows will be the choice of probe drugs for the assessment. The primary consideration in drug selection should be based on the drug elimination mechanism.

### 8.6.3 Type of Study Design

The parallel design and the one-sequence crossover are the commonly selected study designs in TP–drug interaction evaluations (Figure 8-1). The design should be chosen based on the PK properties of the TP and coadministered drugs as well as the potential interaction mechanism.

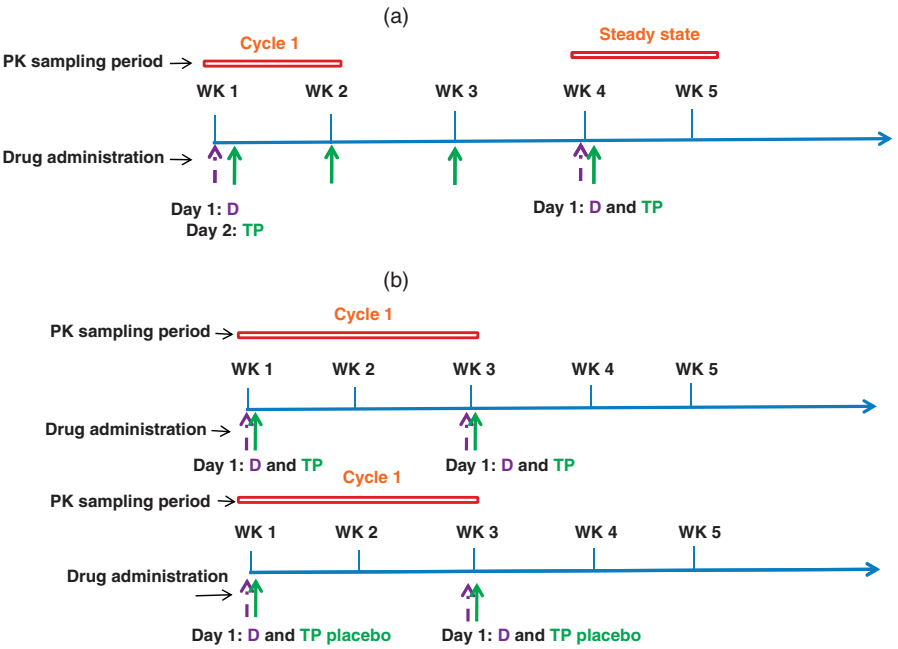


**FIGURE 8-1** Examples of (a) one-sequence crossover and (b) parallel designs. D: small molecule drug; PK: pharmacokinetics; TP: therapeutic protein; WK: week.

The one-sequence crossover (sequential) design is common for testing the effect of a TP on a small molecule drug (TP → D), such as administration of a chemotherapy agent (D) on day 1 and TP on day 2 in one cycle followed by administration of both D and TP on day 1 in a cycle at steady state (Figure 8-1a). Each subject serves as his or her own control to minimize variability and allow for the achievement of sufficient statistical power with a reasonable sample size.

Examples exist for the evaluation of two-way interactions of both TP and small molecule drugs with one sequence crossover design in a single study. In oncology, a study was conducted for the evaluation of the interaction between trastuzumab and paclitaxel in which paclitaxel was administered alone in cycle 1 to collect paclitaxel concentrations, followed by paclitaxel coadministration with trastuzumab with assessment of PK profiles of both drugs in cycle 4, and then trastuzumab monotherapy in cycle 10 or beyond for collection of trastuzumab-alone concentration data.[21] In the inflammation area, the interaction between digoxin and etanercept was assessed with a one-way, crossover, three-period design (digoxin → digoxin + etanercept → etanercept).[22]

The sequential one-way crossover study is ideal for short-half life ($t_{½}$) drugs (e.g., hours) and TPs (e.g., a few days). The advantage of this design is that the study requires fewer subjects when compared to that with parallel design. A limitation of this type of study is that it requires longer study duration so that it may be suitable only for patient populations that have long life expectancy. However, if patients drop out from the study for any reason, imbalanced patient numbers over the study period may occur, which will affect the study evaluation and power of assessment. Moreover, this type of study cannot avoid a sequence effect, especially when dose interruption occurs, and it cannot be used to evaluate the effect of D → TP if there is no TP-alone arm or testing period in the study. In this case, a cross-study comparison for TP exposure needs to be made.

Many TP products have long half-lives (weeks) as well as nonlinear PK owing to target-mediated clearance, or they form antiproduct antibodies that may alter their own clearance. These unique TP characteristics produce unavoidable sequence effects that preclude the use of the crossover designs for assessing TP–drug interaction effects. More complicated studies can be designed to assess the sequence effect. As an alternative for avoiding the sequence effect, parallel designs can be considered for the TP–drug interaction assessments.

A parallel design includes two arms (Figure 8-1b), one for examining drug exposures when administered alone and the other for examining drug exposures in the presence of TP. When a TP has a relative long half-life (10–20 days) and patients have relative short life expectancy (< 6 months, with end stage cancer), this type of design may become the only choice for the drug interaction assessment. Intensive PK samples are usually collected in the first cycle of treatment rather than at steady-state owing to the short life expectancy of the patients. Sparse PK samples can be collected continuously in the following cycles for examining drug accumulation and steady-state concentrations.[18] The study is more appropriate for Phase II oncology studies in which chemobackbone is included in a positive control arm. PK samples of chemotherapy agent from both arms can be collected to compare

exposures with and without the TP. The assessment of effect of chemotherapy on TP may need a cross-study comparison by either noncompartmental analysis (NCA) or population PK modeling owing to lack of a TP alone arm in the same study.

When the TP elimination half-life is relatively short, it is possible to evaluate TP → D and D → TP effects with a parallel design in a single study. This was demonstrated in an interaction study between cetuximab ($t_{½} = 5$ days) and irinotecan ($t_{½} < 1$ day). In this study, two test groups were designed in parallel. Group A was to test the effect of cetuximab on the PK of irinotecan, in which irinotecan was administered alone followed by the drug combination. Group B was to test the effect of irinotecan on the PK of cetuximab in which cetuximab was administered alone followed by the drug combination.[23]

The advantage of the parallel design is the study duration is relatively short when compared to the crossover design, and it is relatively easy to randomize patients in the parallel arms to minimize bias in patient selection owing to baseline characteristics. There is no sequence effect for this type of design, and the design is commonly used to examine TP–drug interaction effects for D + TP and TP + TP combinations in either open-label or double-blind studies. The disadvantage of the design is that much more patients are needed, especially for highly variable TPs, to achieve the same detection power as with the crossover design.

A hybrid approach with a one-way sequential design for a Phase Ib study (e.g., chemo followed by chemo + TP, open label) staggered with a double-blind parallel design in the Phase II portion of the study for testing the effect of TP on chemotherapy is not uncommon in oncology trials.[10] The study is more appropriate for Phase Ib/II trials in which insufficient data on chemotherapy are collected in the Phase Ib portion and the data will be continuously collected in the Phase II portion. The assessment of the effect of chemotherapy on TP may still need a cross-study comparison.

### 8.6.4 Regimen Selection

In oncology, because TPs are commonly used as add-on therapy, the regimen of the backbone treatment (e.g., chemotherapy or target therapy) becomes an important determining factor for the choice of regimen. For instance, the TP regimen will be once every 3 weeks (Q3W) for the treatment of lung cancer if coadministered paclitaxel is given Q3W. On the other hand, if a small molecule drug is given daily, then the regimen of TP can be flexible (e.g., QW, Q2W, Q3W), according to the half-life of the TP and the safety and efficacy profiles. In immunology, in which TPs are frequently used as monotherapy, the choice of regimen is mainly based on the PK, efficacy, and safety profiles of the TP.

### 8.6.5 Sampling Scheme

Depending on the PK characteristics of the tested drugs, intensive PK samples that enable assessment of the AUC of the tested drugs are usually collected for drugs

with short half-lives, and relatively sparse PK samples are collected for TPs owing to their much longer half-lives. Combined sampling schemes–for example, collection of intensive sampling in cycle 1 and sparse sampling at steady state–are common for TPs. Population PK modeling and simulation can be used to optimize sampling schemes. To evaluate the clinical relevance of change in drug exposure parameters, samples for safety, pharmacodynamics, and efficacy measures may be collected along with the concentration samples. If an interaction was observed, clinical relevance of the interaction can be evaluated in conjunction with safety and efficacy data for supporting data interpretation.

### 8.6.6 Endpoints for TP–Drug Interaction Evaluation

The peak ($C_{max}$) and trough ($C_{min}$) concentration are common PK endpoints for TPs in TP–drug interaction assessment when only sparse concentration data are collected. With intensive concentration sample collection and model-based data analysis, the AUC and clearance can be the estimated exposure and PK parameters. Because intensive PK samples can be collected for small molecules, $C_{max}$, $C_{min}$, and AUC are the commonly estimated exposure endpoints. The most relevant exposure parameters should be the ones associated with safety and efficacy events. If a TP–drug interaction signal is observed, an assessment based on a joint evaluation of PK, safety, and efficacy data should be the primary consideration in data interpretation.

### 8.6.7 Sample Size

For a screening TP–drug interaction study, sample size can be $< 10$ subjects per arm, with intensive sampling for small molecule drugs and several hundred subjects with sparse sampling for TPs. All subjects with relevant data can be included in the data analysis. For an estimation study, the TP–drug interaction evaluation is usually performed with 10–20 patients per arm in a PK subgroup assessment of clinical trials. The sample size of fully powered TP–drug interaction studies for hypothesis testing is much larger, depending on power of detection, intersubject variability in TP exposure (e.g., CV = 30–50%), and the preset criteria of no-effect boundary.

### 8.6.8 Criteria for TP–Drug Interaction Evaluation

One of the important aspects in strategic planning and perspective study design is to predefine acceptable criteria of no-effect boundary. Prior knowledge and literature information from molecules with similar structures and elimination mechanisms will be valuable for developing the acceptable clinical criteria. In fact, experience in development of such criteria is very limited for TPs so far. It requires collective knowledge of the disease, comorbidities, cytokine levels under physiological and disease conditions, immunogenicity, biological target, pharmacology of the molecules, PK and PD of involved drugs, and the various mechanisms involved in the

TP clearance. Accumulation of such knowledge on a continuous basis will help properly establish the criteria.[11]

For small molecule drugs, the nominal 80–125% criterion is commonly accepted for no-effect boundary of parameter estimates. However, if the same criterion is applied to TPs, a larger sample size is needed not only because of the large intersubject variability in TP exposures but also because of the studies needed to be conducted in patients in the outpatient setting. In oncology, for instance, patients participate in PK evaluation voluntarily. They may not come to clinics for a blood draw according to protocol-specified time as a reslut of individual difficulties. Data are commonly missed; changes of dosing schedule and reduction of dose levels may be needed for management of adverse events. These factors may lead to an increase in the number of unevaluable subjects so that an extra subjects are need to make up for unpredictable events.

In the TP–drug interaction assessment, the no-effect boundary may be considered to be slightly larger for TPs (e.g., 70–143%) than that for small molecule drugs if it is clinically justifiable. Eventually the determining factors for the acceptable criteria should be associated with the therapeutic window established for the drugs of interest.

## 8.7 DATA ANALYSIS

A data analysis approach is closely related to the study design. For dedicated TP-DI studies, intensive PK samples are collected in a sufficient number of subjects. Noncompartmental analysis (NCA) is the primary data analysis method, which is straightforward, makes few assumptions, and is accepted by regulatory agencies. For screening and estimation TP-DI studies, depending on the sampling scheme and number of subjects with evaluable samples, both NCA and population PK modeling approaches as well as a combination of the two approaches may be applied. For the prospectively designed Phase IIb or Phase III studies with DDI components, the primary analysis approach is population PK modeling. Each analysis approach has advantages and limitations, which are discussed in the following sections.

### 8.7.1 Noncompartment Analysis

As mentioned, the NCA method is simple, descriptive, less resource constrained, and widely accepted by regulatory agencies. Therefore it should be the first choice when intensive PK samples are collected. This is especially true when an interaction study is conducted in healthy subjects. However, when a TP–drug interaction study is conducted in patients, whether this method is still applicable depends on the number of evaluable samples collected in a dosing interval and whether the original sampling scheme is closely followed. If the preset dosing conditions have been greatly modified, it would be difficult to provide summary statistics and make reasonable comparisons to assess the interaction effect. In such cases, the population PK modeling approach will be more appropriate for further analyzing the data.

### 8.7.2 Population Pharmacokinetic Modeling

The great advantages of the population PK modeling approach may include the following:

- The method is less affected by sampling conditions and dose modifications and interruption.
- Data from different trials can be combined to increase detection power and accuracy of parameter estimation and provide a comprehensive evaluation of interaction potential for drugs of interest.
- Both sparse and intensive PK data can be used, and multiple covariates can be evaluated at the same time.
- The magnitude of multiple covariate effects on drug exposure or clearance can be ranked, thus the effect contributed by coadministered drugs can be quantified.
- The effect on drug metabolite(s), if any, in the TP–drug interaction can be further evaluated.

The effect of one drug on the other is usually assessed as a covariate of drug clearance. The method is able to differentiate the effect of a coadministered drug from other factors (e.g., baseline characteristics, demographics, lab values, biomarkers, disease status, etc.) if sufficient information is available. With combination of efficacy and safety data, a PK-PD model can be further established for the evaluation of the clinical relevance of such an effect. The model is used not only to describe observed data but also to simulate data under untested study conditions and for an improved design of the next interaction studies (e.g., to select an optimal sampling scheme and to overcome operational difficulties). The simulation may help evaluate logic of dosing schedules for coadministered drugs to minimize the interaction. Model-based simulations with a large number of subjects may increase the confidence of the analysis.

Concerns of using population PK analysis may include the following:

- The modeling approach makes assumptions and may not be accepted by all regulatory agencies around the world.
- It requires much longer time or more resources for data analysis when compared to NCA.
- When there are too few data and parameter estimates have a large uncertainty, the modeling result may be inconclusive.
- The interaction assessment may be affected if dosing records for tested TP and coadministered drugs are incomplete.
- The interaction effect may be underestimated or overestimated when using sparse data pooled from multiple studies and when certain study factors are imbalanced across studies.

- Studies generating PK data were not designed for the drug interaction assessment, thus the evaluation results may be considered as exploratory.

The lack of a signal can occur if the studies are underpowered to detect the interaction signal. Therefore, the modeling approach will be best used when the study is designed for an interaction evaluation, and potential confounding factors are controlled as much as possible.

## 8.8 PROSPECTIVELY DESIGN OF TP–DRUG INTERACTION STUDY

The screening and estimation types of TP–drug interaction studies, as mentioned in the early part of this chapter, are exploratory in nature for hypothesis generation. The findings may need to be confirmed, especially when changes in drug exposure have potential clinical impact. It is possible to prospectively design the interaction study so that the evaluation results can be considered as confirmatory from the statistical perspective.

The analysis approach requires (1) establishing a model structure for the drug of interest with historical monotherapy data with incorporation of prior knowledge, (2) simulation of concentration-time profiles based on conditions set in the TP interaction study protocol, (3) analysis of the simulated data with the established model and making adjustments in the model structure if needed (the sample size of the study needs to be reasonable, which will be estimated based on the number of subjects in each arm, number of samples per subject, and variability in drug exposure), (4) the selected model structure and no-effect boundary of drug exposure need to be prespecified in the data analysis section of the protocol or in the related statistical analysis plan, and (5) the analysis of the data from the prospectively designed interaction with the prespecified model and examination if the observed data fall within the expected concentration range. Additional sensitivity analysis and ad-hoc analysis may be performed based on the outcomes of analysis. If the observed data are within the specified concentration range, lack of interaction can be confirmed. If significant interaction is observed, the clinical relevance of the effect will be further evaluated in conjunction with assessments of efficacy and safety data. A publication on how to perform confirmatory analysis with population PK modeling is recommended for further reading.[24]

## 8.9 SUMMARY

Assessment of TP–drug interactions is an emerging research area of clinical pharmacology. The logistic considerations in conducting a TP–drug interaction study are being realistic and being practical. The FDA is currently revising its NDA and Biologics License Application (BLA) recommendations to industry sponsors to recognize the accumulation of TP–drug interaction experiences.[5,16] The current guidance does acknowledge the challenges associated with the use

of *in vitro* and *in vivo* data for assessing the risk of TP–drug interactions as well as the design and conduct of such studies. Scientific investigation in this emerging area is needed to improve study design and analysis methods. Prospectively designing TP–drug interaction studies with integrated data analysis may improve the power of detection and reduce certain operational challenges. A further understanding of the mechanisms of TP–drug interaction is needed to better design studies and to ensure patient safety.

## REFERENCES


1. Lobo ED, Hansen RJ, Balthasar JP. Antibody pharmacokinetics and pharmacodynamics. *J Pharm Sci* **93**, 2645–68 (2004).
2. Wang W, Wang EQ, Balthasar JP. Monoclonal antibody pharmacokinetics and pharmacodynamics. *Clin Pharmacol Ther* **84**, 548–58 (2008).
3. Kraynov E, Martin SW, Hurst S, et al. How current understanding of clearance mechanisms and pharmacodynamics of therapeutic proteins can be applied for evaluation of their drug-drug interaction potential. *Drug Metab Dispos* **39**, 1779–83 (2010).
4. Zhou H. A recent BIO survey of therapeutic protein-drug interaction evaluation. Paper presented at the AAPS Workshop on Strategies to Address TPDI During Clinical Development, San Francisco, May 15–16, 2010.
5. Huang SM, Zhao H, Lee JI, et al. Therapeutic protein-drug interactions and implications for drug development. *Clin Pharmacol Ther* **87**, 497–503 (2010).
6. Morgan ET. Impact of infectious and inflammatory disease on cytochrome P450-mediated drug metabolism and pharmacokinetics. *Clin Pharmacol Therap* **85**, 434–38 (2009).
7. Aikawa NE, de Carvalho JF, Silva CAA, Bonfá E. Immunogenicity of anti-TNF-α agents in autoimmune diseases. *Clin Rev Allergy Immunol* **38**, 82–89 (2010).
8. Denlinger CS, Blanchard R, Xu L, et al. Pharmacokinetic analysis of irinotecan plus bevacizumab in patients with advanced solid tumors. *Cancer Chemother Pharmacol* **65**, 97–105 (2009).
9. Sun Y-N. Risk-based drug-drug interaction assessment strategy for therapeutic proteins. Paper presented at the Symposium for Therapeutic Proteins: From Bench to Bedside; 39th American College of Clinical Pharmacology Annual Meeting, Baltimore, MD, September 12–14, 2010.
10. Zhu M, Sun Y-N. Drug-drug interaction of oncology biologics—study design and logistics in oncology setting. Paper presented at the AAPS National Biotech Conference, Seattle, June 24–27, 2009.
11. Zhu M. An integrated approach to assess therapeutic protein drug interaction in oncology setting—current strategy. Paper presented at the TPDI Workshop, AAPS National Biotechnology Conference, San Francisco, May 17–21, 2010.
12. Strehlau J, Pape L, Offner G, et al. Interleukin-2 receptor antibody induced alterations of ciclosporin dose requirements in paediatric transplant recipients. *Lancet* **356**, 1327–28 (2000).
13. Prandota J. Important role of proinflammatory cytokines/other endogenous substances in drug-induced hepatotoxicity: depression of drug metabolism during infections/

inflammation states, and genetic polymorphisms of drug-metabolizing enzymes/cytokines may markedly contribute to this pathology. *Am J Ther* **12**, 254–61 (2005).

14. Morgan ET, Goralski KB, Piquette-Miller M, Renton, et al. Regulation of drug-metabolizing enzymes and transporters in infection, inflammation, and cancer. *Drug Metab Dispos* **36**, 205–16 (2008).
15. Zhou H. Clinical strategies to assess therapeutic protein-drug interaction (TP-DIs) for therapeutic proteins in immunology. Paper presented at the at AAPS National Biotechnology Conference, San Franciso, May 16–18, 2011.
16. US Food and Drug Administration, US Department of Health and Human Services, Center for Drug Evaluation and Research, Center for Biologics Evaluation and Research (CBER). Draft guidance for industry: drug interaction studies—study design, data analysis, implications for dosing, and labeling. Silver Spring, MD, 2012. Available at www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM292362.pdf.
17. Kenny J. Cytokine mediated regulation of human cytochrome P450s—can we translate in vitro evaluation to clinical relevance? Paper presented at the AAPS National Biotechnology Conference, San Francisco, May 16–18, 2011.
18. Zhu M. Clinical strategies to assess TP-DDI in oncology: study design and data analysis. Paper presented at the AAPS National Biotechnology Conference, San Francisco, May 16–18, 2011.
19. Girish S, Martin SW, Peterson MC, et al. AAPS workshop report: strategies to address therapeutic protein–drug interactions during clinical development. *AAPS J* **13**, 405–16 (2011).
20. Zhu M, Jonsson EN, Hsu C-P, Sun Y-N. Establish a model system for evaluation of drug interaction in oncology: paclitaxel example. Paper presented at the annual American Conference of Pharmacometrics Meeting, San Diego, CA, April 3–6, 2011.
21. Leland-Jones B, Gelmon K, Ayoub JP, et al. Pharmacokinetics, safety and efficacy of trastuzumab administered every three weeks in combination with paclitaxel. *J Clin Oncol* **21**, 3965–71 (2003).
22. Zhou H, Parks V, Patat A, et al. Absence of a clinical relevant interaction between etanercept and digoxin. *J Clin Pharmcol* **44**, 1244–51 (2004).
23. Delbaldo C, Pierga JY, Dieras V, et al. Pharmacokinetic profile of cetuximab (Erbitux™) alone and in combination with irinotecan in patients with advanced EGFR-positive adenocarcinoma. *Eur J Can* **41**1739–45 (2005).
24. Hu C, Zhang, J, Zhou, H. Confirmatory analysis for phase III population pharmacokinetics. *Pharm Stat* **10**, 14–26 (2011).

# STATISTICAL CONSIDERATIONS IN ASSESSING DRUG–DRUG INTERACTIONS FOR THERAPEUTIC BIOLOGICS

CHUANPU HU and HONGHUI ZHOU

## 9.1 INTRODUCTION

Therapeutic biologics have different mechanisms of action from small molecules, usually making drug–drug interactions (DDIs) less likely, although this may depend on the specific circumstances. That is, the *a priori* expectation of specific DDIs is different with therapeutic biologics. In addition, therapeutic biologics usually have long half-lives, preventing the effective use of the usual crossover trials to assess drug–drug interactions. These issues influence the design and analysis of DDI assessments. Furthermore, population pharmacokinetics modeling has been used increasingly in drug development and has promising potential in DDI assessment. This chapter discusses the associated statistical issues, particularly to ensure appropriate power and type I error.

DDIs can adversely affect the safety or efficacy of a treatment. Therefore a new drug's DDI potential needs to be fully assessed. Namely, the objective is to obtain estimates of AUC and $C_{\max}$ ratios of interest and their corresponding 90% confidence intervals (CIs). The 90% CIs falling within a predetermined interval–for example, (0.80, 1.25)–indicates the absence of a pharmacokinetic DDI. Dedicated DDI study designs may be the conventional $2 \times 2$ crossover, single sequence crossover, or parallel designs. A specifically selected sequence of *in vitro* and *in vivo* studies can generally define a new experimental drug's clinically relevant DDI potential. The FDA has published a draft guidance on the design and analysis of DDI studies.[1] While AUC assessments are the mainstay of DDI studies, $C_{\max}$ is not assessed in all situations.

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*,
Edited by Shayne C. Gad
Copyright © 2013 John Wiley & Sons, Inc.

Therapeutic biologics often have long elimination half-lives of up to 3 weeks, making conventional crossover studies with adequate washout periods difficult to conduct.[2] A single sequence crossover study with the investigated therapeutic biologic added to the second period is still possible, but a potential confounding time effect cannot be excluded. Some other unique challenges of designing and conducting DDI studies of therapeutic biologics were acknowledged by the FDA draft guidance. For example, classical biotransformation studies are not required routinely, and *in vitro* metabolism-based interactions studies are not generally predictive. On the other hand, there is a concern for the appropriate assessment of metabolic DDIs involving certain biologics, such as type I interferons that inhibit CYP1A2.[3]

There is not yet a uniformly accepted approach regarding DDI assessments for therapeutic biologics. Ideally, an integrated risk assessment should be made in consideration of existing data and knowledge, and clinical relevance. Important factors include therapeutic target, therapeutic window, and mechanisms of drug action.[4]

Population pharmacokinetics (PK) modeling is a powerful tool to synthesize multiple sources of information, including occasional sparse data, to assess PK along with its between-subject variability and to determine the influence of covariates. Because concomitant medication is a type of covariate, population PK can also be used for DDI assessments. It has special appeal in light of the difficulties of conducting conventional DDI studies for therapeutic biologics and has been used in multiple drug development programs.[5–16]

The methodology of population PK is still evolving, along with its applications. The FDA guidance document acknowledged its benefits in assessing DDI, and emphasized the importance of the associated study design and analysis planning. Most applications adopt the exploratory approach where multiple forms of structural, between-subject and within-subject variability models are attempted and one final, best-fitting model is chosen. Conclusions are then based on the final model. In principle, the exploratory approach is hypothesis generating, and the findings need to be confirmed later. While $p$-values may be generated, they are for the purpose of selecting the final model and do not have the appropriate original statistical interpretation of type I error (i.e., false positive rates). On the other hand, it is often desirable for DDI assessment to be confirmatory, especially in regulatory settings. That is, evidence of DDI, or the lack thereof, needs to be quantifiable with controllable type I error. Using population PK modeling in this setting requires special considerations.[17] In the following sections, the necessary theory is introduced along with an example for its application.

## 9.2 METHODOLOGY FOR DRUG–DRUG INTERACTION ASSESSMENTS

### 9.2.1 Conventional DDI Assessment

The statistical methodology for assessing DDI is based on that for bioequivalence. The most common purpose of bioequivalence assessments is to evaluate the

comparability of bioavailability of several drug formulations. Standard bioequivalence studies are $2 \times 2$ crossover studies, and subjects are densely sampled so that individual PK exposure parameters, AUC and $C_{max}$, can be determined with precision. These individual PK parameters are then log-transformed to be used with standard models, adjusting for sequence and period effects, to obtain a 90% confidence interval of log(test/reference ratio). The confidence interval is then back-transformed to the original scale to obtain the confidence interval for the test/reference ratio of average PK parameters. Confidence intervals falling within (0.80, 1.25) indicate bioequivalence. For the statistical theory, see Ref. 18 for details.

DDI studies assess the influence on pharmacokinetics by other drugs instead of formulation effects. Such studies are usually single-sequence crossover or parallel studies, but the assessment method remains the same–that is, whether the confidence intervals of AUC and $C_{max}$ ratios fall within (0.80, 1.25). Of resemblance are PK similarity assessments between subject populations (e.g., healthy volunteers vs. patients). The single-sequence design is more convenient than the $2 \times 2$ crossover, although it faces the theoretical risk that a potential period effect may confound the DDI effect. The parallel design is the most convenient, but in addition to the risk of confounding period effect, it also requires larger sample sizes due to uncontrolled between-subject variability.

Testing whether the DDI effect is significant is not appropriate. The FDA draft guidance states that "because small . . . differences can be statistically significant ($p < 0.05$), but not clinically relevant." There is also a further reason to this–namely that inadequate sample sizes can result in statistically insignificance even though the real effect is clinically relevant.

Preliminary testing—that is, testing the statistical significance of some terms such as period or sequence effect to decide whether to include them in the analysis model—is generally not conducted. This would result in poor statistical behavior of the overall test owing to correlation between the preliminary test and the final test.[19]

### 9.2.2 Population Pharmacokinetics Based DDI Assessments

Population PK has been used for DDI assessment, typically by assessing a drug interaction effect on clearance (CL) or apparent CL.[6–15] Statistical significance of, for example, a DDI effect on CL, has been used on occasion as the deciding criterion; however, this is not appropriate for the same reason as with traditional DDI assessments. The current FDA draft guidance also discourages the use of this simplistic statistical significance criterion by stating: "Population analyses should focus on excluding a specific clinically meaningful PK change." In some other instances, empirical Bayesian individual parameter estimates (e.g., for CL) have been used; however, this is also inappropriate because the individual estimates suffer from shrinkage and are not independent.[20,21]

The appropriate population PK based DDI assessment should be performed as follows: Under the PK model, the effect of the interacting drug on any model parameter can be determined along with 90% CI. Specifically, let $\theta$ be the

parameter, CL, and $\theta_i$ be the parameter value of the $i$th subject, and $X_i = 0$ or 1 be the indicator of whether subject $i$ has the concomitant medication, and $\beta$ represents the DDI effect. The parameter estimate of $\beta$, along with its 90% CI, can be obtained. The DDI effect is most commonly modeled as

$$\theta_i = \beta^{X_i}\theta \tag{9-1}$$

with constraint $\beta > 0$, or

$$\theta_i = (1 + \beta X_i)\theta \tag{9-2}$$

with constraint $\beta > -1$.

Usually, a DDI effect on AUC can be considered the same as that on CL or apparent CL owing to the relationship AUC = Dose/CL.

A DDI effect on $C_{\max}$, if needed, can be determined from the DDI effect on model parameters, at least in theory. In practice, this may be most easily done by reparameterizing the model and using $C_{\max}$ as a parameter. For example, in an intravenous (IV) one-compartment model with a single dose, the DDI effect on $C_{\max}$ can be considered the same as that on volume $V$ owing to the relationship $C_{\max}$ = Dose/$V$. Then, the model can be reparameterized using ($K_a$, AUC, $C_{\max}$) instead of ($K_a$, $V$, CL). If reparameterization cannot be easily done (e.g., for two- or three-compartment models with first-order absorption), the assessment can be done using the bootstrap analysis outlined later in the chapter.

It is possible to obtain an approximate sense of the sample size needed.[22] Simulation is a general option that can be used to assess sample size and design characteristics.[23]

***9.2.2.1 Similarity and Difference between Traditional and Population PK Based Assessments*** In principle, the traditional method assesses the drug interaction effect on the geometric means of AUC and $C_{\max}$, whereas the population PK based model assesses the drug interaction effect on the population average effect. The relationship between the geometric means and the population average AUC and $C_{\max}$ therefore deserves elaboration.

Commonly used population PK models assume between-individual variability of CL as log-normally distributed, which implies that the geometric mean AUC in the population equals the AUC of the average individual.[17] Likewise, under the commonly used assumption of the central volume $V$ being lognormally distributed, the geometric mean $C_{\max}$ in the population (approximately) equals the $C_{\max}$ of the average individual for IV dosing regimens. However, no such general conclusions for $C_{\max}$ holds under oral or subcutaneous (SC) dosing regimens, regardless of distributional assumptions for the rest of the PK parameters.[20] In such situations, the computation of $C_{\max}$ depends on all other model parameters. Unlike for CL, in most situations no single model parameter would amount to the drug-interaction effect on $C_{\max}$.

Obtaining the average $C_{max}$ from population PK models with absorption terms is difficult, because the distribution of $C_{max}$ usually cannot be directly calculated from the between-subject variability distribution of model parameters. Despite the difference, there is no *a priori* reason to prefer the geometric means, and assessing the drug interaction effect on the $C_{max}$ of the average individual, as computed from the estimated population PK model parameters, is reasonable.[20]

***9.2.2.2 Confidence Interval Construction under Population Pharmacokinetic Modeling*** Given a population PK model and data, estimates of DDI effects and the corresponding 90% CIs can be computed.

A simple way to obtain the 90% CI of a DDI effect β is to use its estimate β′ and standard error SE(β′), as

$$(\beta' - \Phi(0.95) * \mathrm{SE}(\beta'),\ \beta' + \Phi(0.95) * \mathrm{SE}(\beta'))$$

where Φ is the cumulative standard normal distribution, and $\Phi(0.95) \approx 1.645$. This involves a normal approximation that, in theory, requires that β is parameterized on the scale $(-\infty, +\infty)$. Therefore, using the following parameterization of the DDI effect β

$$\theta_i = \exp(\beta)^{X_i}\theta \tag{9-3}$$

instead of Eq. 9-1 or 9-2 can make the normal approximation more tenable. If estimation is precise–that is, SE(β′) is small relative to β—the difference between using Eq. 9-3 vs. Eq. 9-1 or 9-2 will be small.

The standard errors of parameter estimates from nonlinear mixed-effect models are approximate, not exact. Bootstrap is an alternative method for generating standard errors and CIs of parameter estimates.[24] It is often believed to be more accurate, although some evidence suggests that it behaves similarly as the standard errors of parameter estimates.[25,26] In the case that the DDI effect on $C_{max}$ needs to be assessed but cannot be formulated as a model parameter, assessment based on standard error of parameter estimate is not possible. However bootstrap can still be used, because in each bootstrap run, the predicted $C_{max}$ with and without the interacting drug, and thus their ratio, can be calculated from the population model parameters.

*Bootstrap Confidence Interval Construction* Bootstrap can be used to generate 90% CIs by the percentile bootstrap method, which uses the 5% and 95% percentiles of the bootstrap runs. Specifically, in each bootstrap run, a bootstrap data set can be generated in which the subjects were resampled with replacement. Parameter estimates can be obtained for the bootstrap data set, and thus a ratio of AUC and $C_{max}$ can be determined. Results of all bootstrap runs can be assembled and the 5% and 95% percentiles used to construct the 90% bootstrap confidence intervals. In general, bootstrap results are also approximate, and the precision depends on whether the numbers of subjects in the resampled categories (i.e., studies) are large

enough to properly represent the entire population. In addition, the following issues are important for its implementation:

*Stratifications.* More stratification variables usually lead to more similar bootstrap datasets and thus narrower CIs. Therefore the proper level of stratification needs to be maintained to ensure correct statistical interpretations. The principle of bootstrap relies on the assumption that the resampling distribution correctly represents the true population distribution. This implies that, in general, design factors, and only design factors should be chosen as stratification factors. For example, *study* should always be a stratification factor–that is, the respective numbers of subjects in each study from a multistudy dataset should be kept unchanged. A more complex and subtle factor is the ratio of subjects with and without the interacting drug in each study. Often it is not a design variable and therefore should not be stratified. However if the study had been included in the DDI analysis because it has a particular ratio of subjects with and without concomitant medication, then the ratio should be a stratification variable. It may happen when all design factors are stratified, as they should be, the number of subjects within some strata becomes too small to accurately represent the population distribution. This is a potential limitation to bootstrap use.

*Number of bootstrap runs to conduct.* This may depend on the desired precision and the type of data, and thus is difficult to say in general. It has been suggested that at least 2000 samples may be needed.[24] In the authors' experience, 2000–5000 will generally suffice.

*Which bootstrap runs to include.* It may be argued that the runs that "failed" in certain ways (e.g., unable to converge or provided standard errors, parameter estimates near boundary) should be excluded. On the other hand, such failed runs correspond to data that are more likely to be atypical, such as at the tail ends of the distribution. In estimating the 90% confidence interval, these types of samples are of particular interest, and runs need to be restarted until convergence is achieved. A counterargument to restarting the failed runs is that the successful runs were not paid similar attention, thus the bias, if any, caused by ignoring failed runs may not be so severe. It may also be argued that when the software converges successfully but fails to produce standard error of parameter estimates, it presents only an algebraic problem unrelated to the quality of convergence and thus, such runs should be included. The problem can be alleviated, to a certain extent, by using multiple starting estimates for each run to improve the convergence rate. A conservative solution may be to include all runs. Future investigations are necessary to determine how to best deal with failed runs.

***9.2.2.3 Influence of Model Searching on DDI Effect Assessment*** The use of population PK requires the selection of a population PK model, which consists of three components: a structural model, a variability model (for both between-subject and within-subject variability), and a covariate model. Influences of these

components on DDI assessments are elaborated in the subsections below. Most applications of population PK were conducted using the exploratory approach, which selects the model that best fits the data among the many possible alternative models, either explicitly or implicitly. While this approach has intuitive appeal, the inferences drawn from the final selected model do not possess statistical validity.[20] Even though at times some hypothesis tests may be conducted, they are generally used as a tool for model selection. Because the model and hypothesis tests are not prespecified, the associated p-values and CIs are not statistically interpretable. The source of the problem is similar to that of the earlier mentioned preliminary testing in bioequivalence (BE) analyses. Further reasons that model exploration may adversely affect DDI analysis include overfitting and selection bias.[27–30]

In theory, inferences could appropriately be made by including the model search into the considerations. For example, bootstrap could be done by including the entire model search process instead of only the final selected model. This has not been described yet, but this argument shows that model search usually leads to increased uncertainty in the results.

*Influence of Structural Model on DDI Effect Assessment* The AUC (and thus the ratio of with to without interacting drug) estimate should be relatively robust because models usually represent the average concentration reasonably well. For $C_{max}$, however, the opposite is true, because fitting simpler structural models to the same data generally leads to smaller $C_{max}$ estimates. Therefore, a simpler model is more likely to bias the DDI assessment of $C_{max}$ toward lack of DDI. This argument could be extended to say that any population PK model will likely underestimate $C_{max}$ because the reality is always more complex. This may be true, but these findings should be placed into perspective. The traditional use of measured $C_{max}$ also underestimates the true $C_{max}$, because $T_{max}$ may not be among the sampling times. A practical prerequisite of using population PK is that sufficient data are available to describe the absorption process.

*Influence of Random Effect Model on DDI Effect Assessment* The within-subject random effect model, also called residual error model, has the effect of weighting concentration measurements differently during the model fitting. Commonly used models are the additive-plus-proportional, proportional, and lognormal model,[29] which take the form as following, respectively:

$$y = p + p * \varepsilon_1 + \varepsilon_2 \tag{9-4}$$

$$y = p + p * \varepsilon_1 \tag{9-5}$$

$$\log(y) = \log(p) + \varepsilon_1 \tag{9-6}$$

where $y$ is the observed concentration, $p$ is model prediction, and $\varepsilon_1$, $\varepsilon_2$ are independently normally distributed. These, in increasing order, weigh smaller concentrations more. Therefore model (9-4) may result in a $C_{max}$ estimation closer to

observed data than models (9-5) and (9-6). Choice among these models, especially between models (9-6) and (9-5), is still of much debate, and currently there is no consensus on which one to use. It may be prudent to base the choice on the type of assay and prior experience with the class of compounds.

Other models have occasionally appeared in literature, at times selected by goodness-of-fit criteria, but caution is necessary. For example, the log-likelihoods of model (9-5) and (9-6) have been compared, which is not statistically meaningful. Another example lies in the attempted approach to combine models (9-4) and (9-5), which takes the form of

$$\log(y) = \log(p) + \varepsilon_1 + \log(p) * \varepsilon_2.$$

This form has the variance of the total residual error on the log scale at the minimum when the prediction $p = 1$, which is difficult to justify. Yet another example has a type of hybrid residual error model as

$$\log(y) = \log(p) + \sqrt{\varepsilon^2 + \varepsilon^2/p^2}$$

This model has the property that, as the prediction approaches zero, the variance approaches infinity. This is also difficult to justify, especially when predose concentration measurements suggest limited variability. Inappropriate residual error models may lead to biased estimation of $C_{\max}$ and AUC.

Many between-subject random effect model parameters, such as correlations among random effects, are of limited interest in DDI assessment. In theory, there is no value to formally test them because there is no *a priori* reason that they should not exist, and any statistical insignificance would only be due to the lack of power to detect the effects. However, it seems plausible that including such effects could cost degrees of freedom and may dilute the power for assessing DDI effects. We suggest that such terms are not included, especially in situations in which power is lacking, pending further investigation.

*Influence of Covariate Models on DDI Effect Assessment* The influence of including additional covariate effects (e.g., gender) in a population PK model is discussed here. In traditional DDI assessments of standard crossover studies, adjusting for gender has no effect because each subject serves as his or her own comparison. Including gender in a population PK assessment for DDI is analogous to adjusting for gender in BE assessments of parallel studies, which adds one term to the linear model and costs only one degree of freedom but can potentially improve the error estimate and thus improve the power of the DDI assessment. However, in population PK modeling it is less clear where to add the gender effect, unless prior information is available. Adding gender influence to all model parameters not only reduces more degrees of freedom but also increases the complexity of parameter estimation. Bootstrap analyses could be particularly sensitive to model instability, because a large number of model runs are needed and convergence could become

problematic. Furthermore, casual correlations between gender and interacting drug usage, if present, can also confound DDI assessments.

In general, the statistical behavior of an analysis becomes worse the more exploration is allowed.[19,28] Intuitively, more model exploration would adversely affect the power of concluding the similarity of PK, because some degrees of freedom must be sacrificed for exploration. The issue would be much more apparent if an exploratory analysis had arrived at some exotic model, such as one that had ethnic origin influence absorption rate that is both weak from its pharmacological mechanism and numerically unstable.

***9.2.2.4 Confirmatory Population Pharmacokinetic Modeling for DDI Assessments*** A confirmatory population PK approach has been recently proposed that achieves the appropriate inferences by prespecifying the analysis model based on available information and the analysis objective.[30,31] The quality of the confirmatory analysis will clearly depend on the suitableness of the prespecified model for the DDI analysis. At most stages of the DDI analysis, data from prior Phase I and II studies are available and provide good information about the PK structural and variability models and even information for some relevant covariates, such as weight. Depending on the analysis objective, all components of the model do not have to be "correct" to yield appropriate results.[29] For example, assessing DDI effects on CL is often robust to model misspecification of between-subject random effect components on the absorption rate constant in a one-compartment model with first-order absorption.[30,32]

Typical steps of DDI assessments using confirmatory population PK in Phase III studies involve the following:

1. *Determine which DDI effects need to be assessed based on mechanistic rationale and whether the sample size is sufficient.* The sample size needed for a reasonably precise determination of the effect size depends on the number of subjects with and without the concomitant medication, along with the number of PK concentration measurements and PK between- and within-subject variability.
2. *Prespecify the population PK structural and variability models based on available information and current Phase III study design.* This can frequently be done by using a previously built population PK model using data from Phases I and II to simulate datasets under the current Phase III study design. Model exploration may be used at this time on the simulated data to see what choice of the structural model, variability model, or even limited covariate models, such as weight can be best supported from the current Phase III study design. The model most stable under the simulated data can be used as the population PK analysis model for the DDI assessment. Most often, a one-compartment model will be selected. As the observed PK concentration data of the current Phase III study have not been used to select the model, they are prespecified and thus the analysis is confirmatory.

3. *Prespecify the covariate model.* Most often, the concomitant medication effects are on clearance. Mechanistically meaningful covariates based on prior information should be retained, such as weight on clearance and volume when existing in the previously built population PK model.
4. *Prespecify sensitivity analyses.* Alternative competing assessments, if needed. Such analyses, if conducted at all, should be limited, given the confirmatory objective.
5. *Determine the magnitude of a relevant DDI effect.* For example, try 70–143%. Determine the power of the DDI assessment accordingly, likely through simulations, and assess only those DDI effects that have sufficient power. If a clear criterion cannot easily be determined, the rule of thumb of at least 20 subjects per effect estimated might suffice, and the resulting 90% CIs may be compared with the nominal 80–125% criterion.

Some of the above steps, especially 1–3, are interrelated and should perhaps be better viewed as one joint step with multiple components. Step 5, determining the criterion of DDI, affects the power assessment but not the assessment method, thus it may be less crucial in practical implementations For further details, see Ref. 29.

In rare situations unexpected complexities may arise, including substantial noncompliance or model nonconvergence. Deviations from the prespecified model would be necessary in these situations to deal with such irregularities but should be maintained at a minimal level and should stay true to the principle of the prespecified analysis.

Because of the many assumptions, a population PK model, whether prespecified or obtained by exploratory search, is unlikely to be "true." However, the relevant question for most applications of modeling is never whether the final model is right or wrong. As has been repeatedly quoted, "All models are wrong, but some are useful." The focus should be on the DDI assessment results instead of on the model–that is, the purpose of the analysis is to obtain not the best description of the data but the most credible estimates along with the associated standard errors for DDI effects.

### 9.2.3 Comparison between Traditional and Population PK Based DDI Assessments

The traditional and the population PK approaches have different advantages. The traditional assessment employs fewer assumptions, is robust to model misspecification, better protects against type I error, and is easy to interpret. The population PK approach can be used in situations not originally designed for DDI assessments, such as, Phase III studies, and to incorporate information from multiple sources by using meta-analyses. A population PK assessment based on large Phase III studies also may better reflect the DDI effect in actual clinical practice. However, population PK is more complex in methodology and thus requires more expertise in design, analysis, and evaluation stages. Data/design quality (e.g., variability in

actual dose/sampling times) may adversely affect study power. However this should not affect type I errors and thus presents only a sponsor risk.

## 9.3 POPULATION PHARMACOKINETICS FOR DRUG–DRUG INTERACTION ASSESSMENTS: USTEKINUMAB

An example of using confirmatory population PK for DDI assessments in phase III clinical studies is described in the following. Ustekinumab is a human monoclonal antibody that binds with high specificity and affinity to the shared p40 subunit of IL-12 and IL-23 and blocks interaction with the IL-12Rβ1 cell surface receptor. The pharmacokinetic properties of ustekinumab have been well characterized after single-dose SC administration (0.27–2.7 mg/kg) and single-dose intravenous administration (0.1–5.0 mg/kg) in patients with psoriasis. The PK of ustekinumab was linear over the wide dose range studied.

***9.3.1.1 Study Design*** Data from two large Phase III trials of ustekinumab in patients with moderate to severe psoriasis (PHOENIX 1 and PHOENIX 2) were used for the DDI assessment.[33,34] Both were multicenter, randomized, double-blind, placebo-controlled, parallel design studies. The study designs were complex, and consisted of placebo-controlled, placebo crossover, dose optimization, randomized withdrawal, and long-term extension periods. Briefly, in PHOENIX 1, a total of 766 patients were assigned to receive SC injections of ustekinumab 45 mg, ustekinumab 90 mg, or placebo at weeks 0 and 4, followed by active treatment every 12 weeks or placebo crossover to ustekinumab 45 mg or 90 mg starting at week 12 (weeks 12–40), followed by randomized withdrawal (weeks 40–76). In PHOENIX 2, a total of 1230 patients were assigned to the same treatment groups and design until week 28, followed by dose schedule optimization (weeks 28–52). Blood samples for the measurement of serum ustekinumab concentrations were collected at weeks 0, 4, 12, 16, 24, 28, 40, 44, 48, and 52 in PHOENIX 1 and at weeks 0, 4, 12, 16, 20, 24, and 28 in PHOENIX 2. At visits when patients received the study agent, blood samples were collected prior to study agent administration. A validated electrochemiluminescent immunoassay method, with a lower limit of quantification of 0.17 μg/mL at a minimum required 1 : 10 dilution, was used to measure serum ustekinumab concentrations. Multiple covariates were collected in the trials, including demographics and concomitant medications.

### 9.3.2 Confirmatory Population Pharmacokinetic Analysis Plan

Several key elements were considered in deciding the covariates to be included in the analysis. These include inherent pharmacokinetic behavior, intended indication and patient population, intended product label, and anticipated important/relevant covariates as learned from similar or related molecules. Before beginning the analysis, a list of the covariates to be included was developed based on physiological relevance, pharmacological importance, baseline disease severity and/or status,

geographic location, past use of relevant medications, relevant concurrent comorbidities, and most commonly used concurrent medications. The following covariates were included: weight, immune response (positive/nonpositive), age (<65 vs. ≥65 years), sex, race (Caucasian, Black, Asian, other), creatinine clearance, geographic region, past use of biologics (yes/no), past use of methotrexate (yes/no), past use of cyclosporine (yes/no), duration of disease, baseline disease scores (PASI), baseline physician global assessment score, baseline dermatology quality index, diabetes status (yes/no), hypertension history (yes/no), lipidemia history (yes/no), alcohol history (yes/no), smoking status (yes/no), and the following concomitant medicines: amlodipine, atenolol, atorvastatin, hydrochlorothiazide, lisinopril, metoprolol, ramipril, metformin, citalopram, and influenza vaccine. The list was developed with the assurance that at least 20 subjects were available for each covariate effect parameter estimated. The exact numbers of subjects who had the concomitant medications were 67, 69, 61, 140, 110, 74, 57, 76, 46, and 106, respectively.

The following analysis plan and rationale were specified before beginning the analysis on observed concentrations.

*Base structural, random-effect and covariate model.* A previous population PK model based on a combined prior Phase I/II dataset was available, with a structural model of a one-compartment model with extravascular absorption. It had between-subject variance on absorption rate constant ($K_a$), apparent clearance (CL/F) and apparent volume of distribution ($V/F$), and an additive-plus-proportional residual error model. In addition, CL/F was proportional to weight raised to the power of 0.75 and $V/F$ proportional to weight. This model was used to simulate data with the actual dosing information and sampling times in the Phase III studies. The best random-effect structure was developed from the simulated data, based on the usual fitting criteria. Base-model exploration on simulated data showed that, as expected, between-subject random-effect on $K_a$ could not be identified. Influence of weight on CL/$F$ and $V/F$ was deemed estimable. Therefore, the prespecified base model had between-subject random effects on CL/$F$ and $V/F$ and also weight effects on them, with an additive-plus-proportional residual error model.

*Covariate model and DDI assessments.* The full model consisting of the covariates listed was used on CL/$F$ only. That is, no covariate models were used on $V/F$ or $K_a$.

### 9.3.3 Covariate and Drug–Drug Interaction Effect Assessment

The prespecified model was implemented with NONMEM (version VI, level 1, ICON Development Solutions, Ellicott City, MD). Model estimation was stable. The parameter estimates and their associated standard errors were used to compute the 90% CIs for all covariate effects on CL/$F$, including the concomitant medications. The 90% CIs are plotted in Figure 9-1. All concomitant medication effects had 90% CIs within (0.8, 1.25), therefore the lack of DDI was concluded for them.

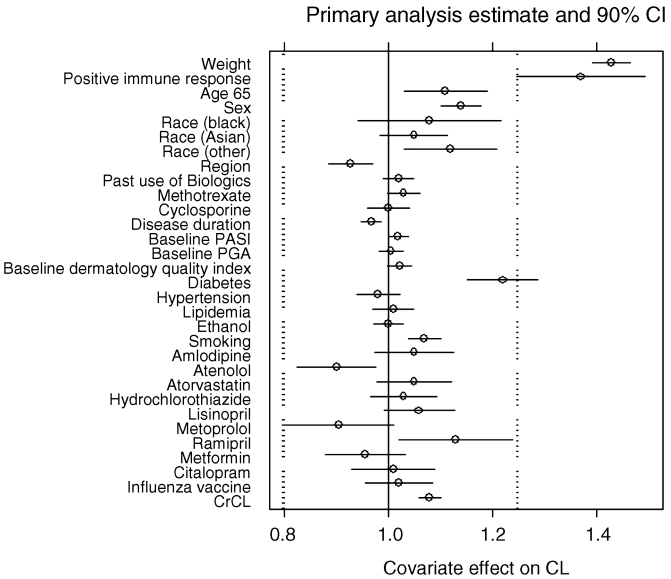


**FIGURE 9-1** The 90% confidence intervals (CI) of covariate and concomitant medication effects. CL: apparent clearance; CrCL: creatinine clearance; PASI: Psoriasis Area and Severity Index score; PGA: Physician's Global Assessment score.

The confirmatory population PK approach gave statistically appropriate 90% CIs for all covariate effects on CL/F, including the concomitant medications.

## 9.4 SUMMARY

DDI assessments for therapeutic biologics have unique challenges. Conventional DDI assessment is still the most robust and least controversial approach. Although it can be inconvenient to conduct, risk-based approaches can be used to guide the assessment. In contrast, the population PK approach has the ability to integrate information from multiple sources, and its potential in DDI assessment has yet to be fully utilized. Adopting the confirmatory approach ensures methodological appropriateness and increases the confidence in its conclusions. The population PK approach may be most useful when AUC is the main focus. The methodology of population PK is still evolving, and special attention needs to be paid to fulfill the objectives of DDI assessments and maintain statistical appropriateness. In principle, population PK assessment should fully parallel the traditional assessments, both in prespecifying the analysis model and ensuring sufficient study design and sample size.

## REFERENCES

1. US Food and Drug Administration, US Department of Health and Human Services, Center for Drug Evaluation and Research, Center for Biologics Evaluation and Research

(CBER). Draft guidance for industry: drug interaction studies–study design, data analysis, implications for dosing, and labeling. Silver Spring, MD, 2012. Available at www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM292362.pdf.

2. Zhou H, Mascelli MA. Mechanisms of monoclonal antibody-drug interactions. *Annu Rev Pharmacol Toxicol* **51**, 359–72 (2011).
3. Seitz K, Zhou H. Pharmacokinetic drug-drug interaction potentials for therapeutic monoclonal antibodies: reality check. *J Clin Pharm* **47**(9), 1104–18 (2007).
4. Zhou H, Davis HM. Risk-based strategy for the assessment of pharmacokinetic drug-drug interactions for therapeutic monoclonal antibodies. *Drug Discov Today* **14**, 891–98 (2009).
5. Zhou H. Population-based assessments of clinical drug-drug interactions: qualitative indices or quantitative measures? *J Clin Pharmacol* **46**, 1268–89 (2006).
6. Delavenne X, Laporte B, Demasles S, et al. Investigation of PK-PD drug-drug interaction between acenocoumarol and amoxicillin plus clavulanic acid. *Fundam Clin Pharmacol* **23**, 127–35 (2009).
7. de Jonge ME, Huitema ADR, Rodenhuis S, Beijnen JH. Integrated population pharmacokinetic model of both cyclophosphamide and thiotepa suggesting a mutual drug–drug interaction. *J Pharmacokinet and Pharmacodyn* **31**(2), 135–56 (2004).
8. Mirochnick M, Cooper E, Capparelli E, et al. Population pharmacokinetics of dapsone in children with human immunodeficiency virus infection. *Clin Pharmacol Ther* **70**(1), 24–32 (2011).
9. de Maat MMR, Huitema ADR, Mulder JW, et al. Drug interaction of fluvoxamine and fluoxetine with nevirapine in HIV-1-infected individuals. *Clin Drug Invest* **23**(10), 629–37 (2003).
10. Lehr T, Staab A, Trommeshauser D, et al. Semi-mechanistic population pharmacokinetic drug-drug interaction modelling of a long half-life substrate and itraconazole. *Clin Pharmacokinet* **49**(1), 53–66 (2010).
11. Sallas WM, Milosavljev S, D'Souza J, Hossain M. Pharmacokinetic drug interactions in children taking oxcarbazepine. *Clin Pharmacol Ther* **74**(2), 138–49 (2003).
12. Joerger M, Huitema ADR, Boogerd W, et al. Interactions of serum albumin, valproic acid and carbamazepine with the pharmacokinetics of phenytoin in cancer patients. *Basic Clin Pharmacol Toxicol* **99**, 133–40 (2006).
13. Bockbrader HN, Burger P, Knapp L. Pregabalin effect on steady-state pharmacokinetics of carbamazepine, lamotrigine, phenobarbital, phenytoin, topiramate, valproate, and tiagabine. *Epilepsia* **52**(2), 405–09 (2011).
14. Bouillon T, Bruhn J, Radu-Radulescu L, et al. Non–steady state analysis of the pharmacokinetic interaction between propofol and remifentanil. *Anesthesiology* **97**, 1350–62 (2002).
15. Vucicevic K, Miljkovic B, Pokrajac M, et al. The influence of drug–drug interaction and patients' characteristics on valproic acid's clearance in adults with epilepsy using nonlinear mixed effects modeling. *Eur J Pharm Sci* **38**, 512–18 (2009).
16. Duan JZ. Applications of population pharmacokinetics in current drug labelling. *J Clin Pharm Ther* **32**, 57–79 (2007).
17. Hu C, Moore K, Kim Y, Sale ME. Statistical issues in a modeling approach to assessing bioequivalence or PK similarity with presence of sparsely sampled subjects. *J Pharmacokinet Pharmacodyn* **33**(4), 321–39 (2004).

18. Schuirmann DJ. A comparison of the two one-sided tests procedure and the power approach for assessing the equivalence of average bioavailability. *J Pharmacokinet Biopharm* **15**(6), 657–80 (1987).
19. Senn S. *Statistical issues in drug development*. Wiley, Chichester, 1997.
20. Hu C, Dong Y. Estimating the predictive quality of dose-response after model selection. *Statis Med* **26**, 3114–39 (2007).
21. Dubois A, Gsteiger S, Pigeolet E, Mentré F. Bioequivalence tests based on individual estimates using non-compartmental or model-based analyses: evaluation of estimates of sample means and type I error for different designs. *Pharm Res* **27**(1), 92–104 (2010).
22. Kang D, Schwartz JB, Verotta D. A sample size computation method for non-linear mixed effects models with applications to pharmacokinetics models. *Statis Med* **23**(16), 2551–66 (2004).
23. Vong C, Bergstrand M, Nyberg J, Karlsson MO. Rapid sample size calculations for a defined likelihood ratio test-based power in mixed effects models. *AAPS J* **14**(2), 176–86 (2012).
24. Efron B, Tibshirani RJ. *An introduction to the bootstrap*. Chapman & Hall, London, 1993.
25. Hu C, Xu Z, Rahman MU, et al. A latent variable approach for modeling categorical endpoints among patients with rheumatoid arthritis treated with golimumab plus methotrexate. *J Pharmacokinet Pharmacodyn* **37**(4), 309–321 (2010).
26. Hu C, Xu Y, Rahman M, et al. A population approach for exposure-response modeling of golimumab in patients with rheumatoid arthritis. *J Clin Pharmacol* **51**(5), 639–48 (2011).
27. Faraway JJ. On the cost of data analysis. *J Comput Graph Stat* **1**, 213–29 (1992).
28. Harrell F. *Regression modeling strategies: with applications to linear models, logistic regression, and survival analysis*. Springer, New York, 2001.
29. Hu C, Zhang J, Zhou H. Confirmatory analysis for phase III population pharmacokinetics. *Pharm Stat* **10**(1), 14–26 (2011).
30. Ribbing J, Jonsson EN. Power, selection bias and predictive performance of the population pharmacokinetic covariate model. *J Pharmacokin Pharmacodyn* **31**(2), 109–34 (2004).
31. Hu C, Zhou H. An improved approach for confirmatory Phase III population pharmacokinetic analysis. *J Clin Pharmacol* **48**(7), 812–22 (2008).
32. Kowalski KG, Hutmacher MM. Design evaluation for a population pharmacokinetic study using clinical trial simulations: a case study. *Stat Med* **20**, 75–91 (2001).
33. Leonardi CL, Kimball AB, Papp KA, et al. Efficacy and safety of ustekinumab, a human interleukin-12/23 monoclonal antibody, in patients with psoriasis: 76-week results from a randomised, double-blind, placebo-controlled trial (PHOENIX 1). *Lancet* **371**(9625), 1665–74 (2008).
34. Papp KA, Langley RG, Lebwohl M, et al. Efficacy and safety of ustekinumab, a human interleukin-12/23 monoclonal antibody, in patients with psoriasis: 52-week results from a randomised, double-blind, placebo-controlled trial (PHOENIX 2). *Lancet* **371**(9625), 1675–84 (2008).

# SCIENTIFIC PERSPECTIVES ON THERAPEUTIC PROTEIN DRUG–DRUG INTERACTION ASSESSMENTS*

HONG ZHAO, LEI ZHANG, KELLIE REYNOLDS, and SHIEW-MEI HUANG

## 10.1 INTRODUCTION

Many intrinsic and extrinsic factors can affect an individual patient's drug exposure and response.[1] The FDA has published a number of guidance documents that recommend how and when to evaluate these factors during drug development.[2] The most recent FDA draft guidance on drug interactions[3] provides advice for *in vitro* and *in vivo* drug interaction studies, including suggestions for study design, dosing strategies, data analysis and interpretation, and the type of information that should be included in medical product labels; it updates the FDA's recommendations on the evaluation of important cytochrome P450 enzyme- and transporter-based drug interactions, and the use of physiologically based pharmacokinetic modeling approaches for drug interaction assessment; and it expands recommendations on therapeutic protein–drug interaction assessment. The recommendations are illustrated via a series of decision trees. This draft guidance and other FDA publications related to drug–drug interactions (DDIs),[4–8] including the website *Drug Development and Drug Interactions*,[9] provide recommendations on how to evaluate drug interaction potential for both small molecules (drugs or Ds) and therapeutic proteins (TPs).

Two review papers published in 2007 contain examples of drug interactions involving TPs.[10,11] Several recent publications summarize the TP-DDI evaluation included in Biologics License Application (BLA) submissions and current strategies for the evaluation of TP-DDI potential during drug development.[12–17]

*The views expressed in this book chapter are those of the authors and do not necessarily reflect the official views of the US Food and Drug Administration.

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*,
Edited by Shayne C. Gad
Copyright © 2013 John Wiley & Sons, Inc.

The purpose of this chapter is to describe our review of the literature on drug interactions involving TPs, the FDA's expanded recommendations on TP-DDIs in the updated draft drug interaction guidance, and FDA-approved product labels that include TP-DDI information. The information presented in this chapter includes the following:

- TPs with drug interaction studies conducted during product development
- Types of study designs used to evaluate TP-DDIs
- Impact of study results on product label

## 10.2 THERAPEUTIC PROTEIN–DRUG INTERACTION STUDIES

Therapeutic proteins are increasingly being combined with small-molecule drugs and/or with other TPs to gain greater clinical benefit. Our survey of 76 new molecular entity TPs that were approved by the FDA up to February 2010 examined drug interaction information in product labels to help determine the types of TPs for which drug interaction studies are informative (Table 10-1).[14] The survey showed that 44 TPs (58%) included some information on drug interactions in the label (including 9 with dedicated TP-DDI studies). The products most likely to have drug interaction information in the label were therapeutic cytokines (9/11, 81%), followed by therapeutic monoclonal antibodies (mAbs), some of which are cytokine modulators (19/29, 66%). Some cytokines are involved in the pathophysiology of various human diseases and are produced during infection and inflammation. Therapeutic cytokines are used in the treatment of certain diseases such as cancer. Various *in vitro* and limited *in vivo* human studies indicate that cytokines and cytokine modulators can differentially influence the expression or activities of specific cytochrome P450 (CYP) enzymes and/or drug transporters through multiple mechanisms[18–23] and therefore have the potential to alter the systemic exposure and/or clinical response of concomitantly administered drugs that are substrates for the

**TABLE 10-1 Summary of Drug Interaction Studies in the FDA-Approved Package Inserts for New Molecular Entity Therapeutic Proteins Approved by the End of February 2010[a]**

| Category | Dedicated Studies | Some Information | No Information | Total |
|---|---|---|---|---|
| Cytokines | 2 | 7 | 2 | 11 (14%) |
| Growth factors | 0 | 2 | 8 | 10 (13%) |
| Enzymes | 1 | 7 | 9 | 17 (22%) |
| Monoclonal antibodies | 6 | 13 | 10 | 29 (38%) |
| Miscellaneous | 0 | 6 | 3 | 9 (12%) |
| Total | 9 (12%) | 35 (46%) | 32 (42%) | 76 (100%) |

[a]Adapted from Ref. 14.

affected CYP enzymes or transporters. For example, therapeutic mAbs that modulate cytokine activities can indirectly influence the expression of specific CYP enzymes and drug transporters by affecting cytokine concentrations.[24,25] Examples of TPs that affect CYP enzymes are listed in Table 10-2. In addition, the exposure to therapeutic mAbs can be altered by concomitantly administered immunosuppressive drugs that affect the formation of antiproduct antibodies.[26,27]

TP-drug interactions (either pharmacokinetics [PK] or pharmacodynamic [PD] based) have been observed, and information about these interactions is included in product labeling (Tables 10-3 and 10-4).[26] The FDA's recommendations on the evaluation of DDIs involving small molecule drugs, including study design, data analysis, and dosing and labeling implications have been described in detail.[4–9] Many of the same principles apply to the evaluation of DDI involving TPs. The current thought process for evaluation of TP-DDIs and the types of studies that have been conducted during drug development to evaluate TP-DDI interactions are illustrated in Figure 10-1.[3] General considerations are discussed below:

- If an investigational TP is a cytokine or cytokine modulator that has known effects on CYP enzymes or transporters, studies should be conducted to determine the magnitude of the TP's effects on drugs that are substrates of the affected CYP enzymes or transporters[13,19]

  *In vitro* and *in vivo* studies have been conducted to determine the TP's effects on CYP enzymes or transporters. When the TP is a cytokine modulator, the effects of both the specific cytokine(s) being modulated and the cytokine modulator on enzymes and transporters may be assessed using *in vitro* models (such as human hepatocytes with the addition of a physiologically relevant

**TABLE 10-2 CYP Enzymes with Decreased Activities (Unless Noted) in the Presence of Specific Cytokines, Cytokine Modulators, and Human Growth Hormone Based on *In Vitro* and/or *In Vivo* Human Studies[a]**

| CYP Enzyme | Cytokines/Cytokine Modulators |
|---|---|
| CYP1A2 | IFNα, IFNα-2b, IFNβ, IL-2, IL-6, hGH[b] |
| CYP2C8 | IL-1 |
| CYP2C9 | IL-2, IL-10 |
| CYP2C19 | Tocilizumab,[c] IFNα-2b, IFNβ, IL-2, TNFα, IL-6, hGH |
| CYP2D6 | IFNα-2b |
| CYP2E1 | IL-2, IFNα-2b |
| CYP3A | Basiliximab, muromonab-CD3, tocilizumab,[c] IL-1, IL-2, IL-6, IL-10 |

[a]Modified from Ref. 13; data from Refs. 14 and 20. All have data from *in vitro* and *in vivo* studies in humans to support the changes in cytochrome P450 (CYP) enzyme activities and/or expression with the exception of human growth hormone (hGH), interleukin 10 (IL-10), muromonab-CD3, and basiliximab, for which only *in vivo* data are available.
[b]The effect of hGH is increased activity.
[c]The effect of tocilizumab, an IL-6 receptor inhibitor, is increased activity.
IFN: interferon; TNF: tumor necrosis factor.

**TABLE 10-3 Examples of Labeling Language for Therapeutic Protein–Drug Interactions (Pharmacokinetic-Based)**

| TP Name<br>(Brand Name)<br>Pharmacologic Class<br>Year of Labeling | Item | Labeling Language |
|---|---|---|
| 1. CYP enzyme modulation by cytokines and cytokine modulators | | |
| Interferon α-2a<br>(Roferon-A)<br>INF-α<br>2006 | Precautions: drug interactions | Roferon-A has been reported to reduce the clearance of theophylline. The clinical relevance of this interaction is presently unknown.<br>α-Interferons may affect the oxidative metabolic process by reducing the activity of hepatic microsomal cytochrome enzymes in the P450 group. Although the clinical relevance is still unclear, this should be taken into account when prescribing concomitant therapy with drugs metabolized by this route. |
| Peginterferon α-2a<br>(Pegasys)<br>INF-α<br>2011 | Highlights of prescription information: drug interactions | *Drugs metabolized by CYP1A2*: monitor for increased serum levels of theophylline and adjust dose accordingly. (§7.2)<br>*Methadone*: monitor for signs and symptoms of methadone toxicity. (§7.3) |
| | 4 Contraindications | Pegasys/Copegus combination therapy is additionally contraindicated in combination with didanosine. Reports of fatal hepatic failure and peripheral neuropathy, pancreatitis, and symptomatic hyperlactatemia/lactic acidosis have been reported in clinical trials. (§7.4) |
| | 7 Drug interactions | |
| | 7.1 Drugs metabolized by cytochrome P450 | There was no effect on the pharmacokinetics of representative drugs metabolized by CYP 2C9, CYP 2C19, CYP 2D6 or CYP 3A4.<br>Treatment with Pegasys once weekly for 4 weeks in healthy subjects was associated with an inhibition of P450 1A2 and a 25% increase in theophylline AUC. |
| | 7.2 Theophylline | Treatment with Pegasys once weekly for 4 weeks in healthy subjects was associated with an inhibition of P450 1A2 and a 25% increase in theophylline AUC. Theophylline serum levels should be monitored and appropriate dose adjustments considered for patients given both theophylline and Pegasys. |

| | | |
|---|---|---|
| | 7.3 Methadone | In a PK study of HCV subjects concomitantly receiving methadone, treatment with Pegasys once weekly for 4 weeks was associated with methadone levels that were 10–15% higher than at baseline. The clinical significance of this finding is unknown; however, patients should be monitored for the signs and symptoms of methadone toxicity.<br>The PK of concomitant administration of methadone and Pegasys were evaluated in 24 Pegasys naive cCHC subjects (15 male, 9 female) who received 180 μg Pegasys SC weekly. All subjects were on stable methadone maintenance therapy (median dose 95 mg, range 30 mg to 150 mg) before receiving Pegasys. Mean methadone PK parameters were 10–15% higher after 4 weeks of Pegasys treatment compared to baseline. Methadone did not significantly alter the PK of Pegasys compared to a PK study of 6 cCHC subjects not receiving methadone. |
| Interferon α-2b<br>(Intron A)<br>INF-α<br>2011 | Precautions: drug Interactions | Interactions between Intron A and other drugs have not been fully evaluated. Caution should be exercised when administering Intron A therapy in combination with other potentially myelosuppressive agents such as zidovudine. Concomitant use of INF-α and theophylline decreases theophylline clearance, resulting in a 100% increase in serum theophylline levels. |
| Peginterferon α-2b<br>(PEG-INTRON)<br>INF-α<br>2011 | Highlights of prescription information: drug interactions | *Drug metabolized by CYP450*: caution with drugs metabolized by CYP2C8/9 (e.g., warfarin, phenytoin) or CYP2D6 (e.g., flecainide). (§7.1)<br>*Methadone*: monitor for increased narcotic effect. (§7.2) |
| | 7 Drug Interactions | |
| | 7.1 Drugs metabolized by cytochrome P450 | When administering PEG-INTRON with medications metabolized by CYP2C8/9 (e.g., warfarin and phenytoin) or CYP2D6 (e.g., flecainide), the therapeutic effect of these substrates may be decreased. (§12.3). |
| | 7.2 Methadone | PEG-INTRON may increase methadone concentrations. (§12.3). The clinical significance of this finding is unknown; however, patients should be monitored for the signs and symptoms of increased narcotic effect. |

(*continued*)

**TABLE 10-3** (*Continued*)

| TP Name (Brand Name) Pharmacologic Class Year of Labeling | Item | Labeling Language |
|---|---|---|
| | 12 Clinical Pharmacology | |
| | 12.3 Pharmacokinetics: drug interactions | |
| | Drugs metabolized by cytochrome P450 | The pharmacokinetics of representative drugs metabolized by CYP1A2 (caffeine), CYP2C8/9 (tolbutamide), CYP2D6 (dextromethorphan), CYP3A4 (midazolam), and *N*-acetyltransferase (dapsone) were studied in 22 subjects with CHCwho received PEG-INTRON (1.5 μg/kg) once weekly for 4 weeks. PEG-INTRON treatment resulted in a 28% (mean) increase in a measure of CYP2C8/9 activity. PEG-INTRON treatment also resulted in a 66% (mean) increase in a measure of CYP2D6 activity; however, the effect was variable: 13 subjects had an increase, 5 subjects had a decrease, and 4 subjects had no significant change. (§7.1) |
| | | No significant effect was observed on the pharmacokinetics of representative drugs metabolized by CYP1A2, CYP3A4, or *N*-acetyltransferase. The effects of PEG-INTRON on CYP2C19 activity were not assessed. |
| | Methadone | The pharmacokinetics of concomitant administration of methadone and PEG-INTRON were evaluated in 18 PEG-INTRON-naive CHC subjects receiving 1.5 μg/kg PEG-INTRON SC weekly. All subjects were on stable methadone maintenance therapy receiving ≥40 mg/day before initiating PEG-INTRON. Mean methadone AUC was approximately 16% higher after 4 weeks of PEG-INTRON treatment compared to baseline. In 2 subjects, methadone AUC was approximately double after 4 weeks of PEG-INTRON treatment compared to baseline. (§7.2). |
| Tocilizumab (Actemra) Human IL-6 receptor blocker 2011 | 7 Drug interactions 7.2 Interactions with CYP450 substrates | Cytochrome P450s in the liver are downregulated by infection and inflammation stimuli, including cytokines, such as IL-6. Inhibition of IL-6 signaling in RA patients treated with tocilizumab may restore CYP450 activities to higher levels than those in the absence of tocilizumab, leading to increased metabolism of drugs that are CYP450 |

substrates. *In vitro* studies showed that tocilizumab has the potential to affect expression of multiple CYP enzymes, including CYPlA2, CYP2B6, CYP2C9, CYP2C19, CYP2D6 and CYP3A4. Its effect on CYP2C8 or transporters is unknown. *In vivo* studies with omeprazole, metabolized by CYP2C19 and CYP3A4, and simvastatin, metabolized by CYP3A4, showed up to a 28% and 57% decrease in exposure 1 week after a single dose of Actemra, respectively. The effect of tocilizumab on CYP enzymes may be clinically relevant for CYP450 substrates with narrow therapeutic index, where the dose is individually adjusted. On initiation or discontinuation of Actemra, in patients being treated with these types of medicinal products, therapeutic monitoring of effect (e.g., warfarin) or drug concentration (e.g., cyclosporine or theophylline) should be performed and the individual dose of the medicinal product adjusted as needed. Prescribers should exercise caution when Actemra is coadministered with CYP3A4 substrate drugs when a decrease in effectiveness is undesirable (e.g., oral contraceptives, lovastatin, atorvastatin.). The effect of tocilizumab on CYP450 enzyme activity may persist for several weeks after stopping therapy. (§12.3).

12 Clinical pharmacology
12.3 Pharmacokinetics: drug interactions

*In vitro* data suggested that IL-6 reduced mRNA expression for several CYP450 isoenzymes, including CYPlA2, CYP2B6, CYP2C9, CYP2C19, CYP2D6, and CYP3A4, and this reduced expression was reversed by co-incubation with tocilizumab at clinically relevant concentrations. Accordingly, inhibition of IL-6 signaling in RA patients treated with tocilizumab may restore CYP450 activities to higher levels than those in the absence of tocilizumab, leading to increased metabolism of drugs that are CYP450 substrates. Its effect on CYP2C8 or transporters (e.g., P-gp) is unknown. This is clinically relevant for CYP450 substrates with a narrow therapeutic index, where the dose is individually adjusted. On initiation of Actemra, in patients being treated with these types of medicinal products, therapeutic monitoring of the effect (e.g., warfarin) or drug concentration (e.g., cyclosporine or theophylline) should be performed and the individual dose of the medicinal product

(*continued*)

**TABLE 10-3** (*Continued*)

| TP Name (Brand Name) Pharmacologic Class Year of Labeling | Item | Labeling Language |
|---|---|---|
| | | adjusted as needed. Caution should be exercised when Actemra is coadministered with drugs where decrease in effectiveness is undesirable, such as oral contraceptives (CYP3A4 substrates). (§7.2). |
| | Simvastatin | Simvastatin is a CYP3A4 and OATPlBl substrate. In 12 RA patients not treated with Actemra and receiving 40 mg simvastatin, exposures of simvastatin and its metabolite, simvastatin acid, was 4- to 10-fold and 2-fold higher, respectively, than the exposures observed in healthy subjects. One week following administration of a single infusion of Actemra (10 mg per kg), exposure of simvastatin and simvastatin acid decreased by 57% and 39%, respectively, to exposures that were similar or slightly higher than those observed in healthy subjects. Exposures of simvastatin and simvastatin acid increased upon withdrawal of Actemra in RA patients. Selection of a particular dose of simvastatin in RA patients should take into account the potentially lower exposures that may result after initiation of Actemra (due to normalization of CYP3A4) or higher exposures after discontinuation of Actemra. |
| | Omeprazole | Omeprazole is a CYP2C19 and CYP3A4 substrate. In RA patients receiving 10 mg omeprazole, exposure to omeprazole was approximately twofold higher than that observed in healthy subjects. In RA patients receiving 10 mg omeprazole, before and 1 week after Actemra infusion (8 mg per kg), the omeprazole $AUC_{inf}$ decreased by 12% for poor ($n = 5$) and intermediate metabolizers ($n = 5$) and by 28% for extensive metabolizers ($n = 8$) and were slightly higher than those observed in healthy subjects. |

| | | |
|---|---|---|
| | Dextromethorphan | Dextromethorphan is a CYP2D6 and CYP3A4 substrate. In 13 RA patients receiving 30 mg dextromethorphan, exposure to dextromethorphan was comparable to that in healthy subjects. However, exposure to its metabolite, dextrorphan (a CYP3A4 substrate), was a fraction of that observed in healthy subjects. One week following administration of a single infusion of Actemra (8 mg per kg), dextromethorphan exposure was decreased by approximately 5%. However, a larger decrease (29%) in dextrorphan levels was noted after Actemra infusion. |
| Rilonacept (Arcalyst) IL-1β blocker 2008 | 7 Drug interactions<br>7.2 Cytochrome P450 substrates | The formation of CYP450 enzymes is suppressed by increased levels of cytokines (e.g., IL-1) during chronic inflammation. Thus it is expected that for a molecule that binds to IL-1, such as rilonacept, the formation of CYP450 enzymes could be normalized. This is clinically relevant for CYP450 substrates with a narrow therapeutic index, where the dose is individually adjusted (e.g., warfarin). On initiation of Arcalyst, in patients being treated with these types of medicinal products, therapeutic monitoring of the effect or drug concentration should be performed and the individual dose of the medicinal product may need to be adjusted as needed. |
| Canakinumab (Ilaris) IL-1β blocker 2012 | 7 Drug interactions<br>7.3 Cytochrome P450 substrates | Interactions between Ilaris and other medicinal products have not been investigated in formal studies.<br>The formation of CYP450 enzymes is suppressed by increased levels of cytokines (e.g., IL-1) during chronic inflammation. Thus it is expected that for a molecule that binds to IL-1, such as canakinumab, the formation of CYP450 enzymes could be normalized. This is clinically relevant for CYP450 substrates with a narrow therapeutic index, where the dose is individually adjusted (e.g., warfarin). On initiation of canakinumab, in patients being treated with these types of medicinal products, therapeutic monitoring of the effect or drug concentration should be performed and the individual dose of the medicinal product may need to be adjusted as needed. |
| Golimumab (Simponi) TNF blocker | 7 Drug interactions<br>7.4 Cytochrome P450 substrates | The formation of CYP450 enzymes may be suppressed by increased levels of cytokines (e.g., TNF-α) during chronic inflammation. Therefore, it is expected that for a |

(*continued*)

**TABLE 10-3** (*Continued*)

| TP Name (Brand Name) Pharmacologic Class Year of Labeling | Item | Labeling Language |
|---|---|---|
| 2011 | | molecule that antagonizes cytokine activity, such as golimumab, the formation of CYP450 enzymes could be normalized. On initiation or discontinuation of Simponi in patients being treated with CYP450 substrates with a narrow therapeutic index, monitoring of the effect (e.g., warfarin) or drug concentration (e.g., cyclosporine or theophylline) is recommended and the individual dose of the drug product may be adjusted as needed. |
| Infliximab (Remicade) TNF blocker 2011 | 7 Drug interactions 7.5 Cytochrome P450 substrates | The formation of CYP450 enzymes may be suppressed by increased levels of cytokines (e.g., TNF-α, IL-1, IL-6, IL-10, IFN) during chronic inflammation. Therefore, it is expected that for a molecule that antagonizes cytokine activity, such as infliximab, the formation of CYP450 enzymes could be normalized. On initiation or discontinuation of Remicade in patients being treated with CYP450 substrates with a narrow therapeutic index, monitoring of the effect (e.g., warfarin) or drug concentration (e.g., cyclosporine or theophylline) is recommended, and the individual dose of the drug product may be adjusted as needed. |
| Ustekinumab (Stelara) human IL12/IL23 antagonist 2012 | 7 Drug interactions 7.3 Cytochrome P450 substrates | Drug interaction studies have not been conducted with Stelara. The formation of CYP450 enzymes can be altered by increased levels of certain cytokines (e.g., IL-1, IL-6, IL-10, TNFα, IFN) during chronic inflammation. Thus, STELARA®, an antagonist of IL-12 and IL-23, could normalize the formation of CYP450 enzymes. Upon initiation of STELARA® in patients who are receiving concomitant CYP450 substrates, particularly those with a narrow therapeutic index, monitoring for therapeutic effect (e.g., for warfarin) or drug concentration (e.g., for cyclosporine) should be considered and the individual dose of the drug adjusted as needed [*see Clinical Pharmacology (12.3)*]. |

| | | |
|---|---|---|
| | 12 Clinical pharmacology | |
| | 12.3 Pharmacokinetics: drug–drug interactions | The effect of IL-12 or IL-23 on the regulation of CYP450 enzymes were evaluated in an *in vitro* study using human hepatocytes, which showed that IL-12 and /or IL-23 at levels of 10 ng/mL did not alter human CYP450 enzyme activities (CYP1A2, 2B6, 2C9, 2D6, or 3A4). However, the clinical relevance of *in vitro* data has not been established. (§7.3). |
| Belatacept (Nulojix) Selective T-cell costimulation blocker 2011 | 7 Drug interactions | |
| | 7.1 Cytochrome P450 Substrates | No formal drug interaction studies have been conducted with Nulojix. Other biologic therapies that are cytokines or cytokine modulators have been shown to affect the expression and/or functional activities of CYP450 enzymes *in vitro* and/or *in vivo*. *In vitro* studies have shown that Nulojix inhibits the production of certain cytokines during an alloimmune response. No studies in kidney transplant patients have been conducted to assess if Nulojix inhibits cytokine production *in vivo*. The potential for Nulojix to alter the systemic concentrations of drugs that are CYP450 substrates has not been studied; however, in the event that kidney transplant patients receiving Nulojix exhibit signs and symptoms of altered administered drugs that are known to be efficacy or adverse events associated with cometabolized by CYP450, the clinician should be aware of potentially altered CYP450 metabolism of these drugs. |
| | 7.2 Use with mycophenolate mofetil | In a pharmacokinetic substudy of Studies 1 and 2, the plasma concentrations of MPA were measured in 41 patients who received fixed MMF doses of 500–1500 mg twice daily with either 5 mg/kg of Nulojix or cyclosporine. The mean dose-normalized MPA $C_{max}$ and $AUC_{0-12}$ were approximately 20% and 40% higher, respectively, with Nulojix coadministration than with cyclosporine coadministration.<br>Clinicians should be aware that there is also a potential change of MPA exposure after crossover from cyclosporine to Nulojix or from Nulojix to cyclosporine in patients concomitantly receiving MMF. |
| Somatropin (rDNA origin) (Genotropin) human growth hormone | Highlights of prescription | *Inhibition of 11 β-hydroxysteroid dehydrogenase type 1*: may require the initiation of glucocorticoid replacement therapy. Patients treated with glucocorticoid replacement |

(*continued*)

**TABLE 10-3** (*Continued*)

| TP Name (Brand Name) Pharmacologic Class Year of Labeling | Item | Labeling Language |
|---|---|---|
| 2011 | information: drug interactions | therapy for previously diagnosed hypoadrenalism may require an increase in their maintenance doses (§7.1; §7.2).<br>*Glucocorticoid replacement*: should be carefully adjusted (§7.2)<br>*Cytochrome P450-metabolized drugs*: monitor carefully if used with somatropin (§7.3)<br>*Oral estrogen*: larger doses of somatropin may be required in women (§7.4)<br>Insulin and/or oral/injectable hypoglycemic agents: may require adjustment (§7.5) |
| | 7 Drug interactions | |
| | 7.1 11 β-hydroxysteroid dehydrogenase type 1 | The microsomal enzyme 11βHSD-1 is required for conversion of cortisone to its active metabolite, cortisol, in hepatic and adipose tissue. GH and somatropin inhibit 11βHSD-1. Consequently, individuals with untreated GH deficiency have relative increases in 11βHSD-1 and serum cortisol. Introduction of somatropin treatment may result in inhibition of 11βHSD-1 and reduced serum cortisol concentrations. As a consequence, previously undiagnosed central (secondary) hypoadrenalism may be unmasked and glucocorticoid replacement may be required in patients treated with somatropin. In addition, patients treated with glucocorticoid replacement for previously diagnosed hypoadrenalism may require an increase in their maintenance or stress doses after initiation of somatropin treatment; this may be especially true for patients treated with cortisone acetate and prednisone because conversion of these drugs to their biologically active metabolites depend on the activity of 11βHSD-1. |
| | 7.3 Cytochrome P450-Metabolized drugs | Limited published data indicate that somatropin treatment increases CYP450-mediated antipyrine clearance in humans. These data suggest that somatropin administration may alter the clearance of compounds known to be metabolized by CYP450 liver enzymes (e.g., corticosteroids, sex steroids, anticonvulsants, cyclosporine). Careful monitoring is advisable when somatropin is administered in combination with other |

| | | |
|---|---|---|
| | | drugs known to be metabolized by CYP450 liver enzymes. However, formal drug interaction studies have not been conducted. |
| | 7.4 Oral estrogen | In patients on oral estrogen replacement, a larger dose of somatropin may be required to achieve the defined treatment goal. (§2.2) |
| | 7.5 Insulin and/or oral/injectable hypoglycemic agents | In patients with diabetes mellitus requiring drug therapy, the dose of insulin and/or oral/injectable agent may require adjustment when somatropin therapy is initiated. (§5.4) |
| 2. Immunosuppressive effect | | |
| Infliximab (Remicade) | 7 Drug Interactions | |
| TNF blocker | 7.3 MTX and other concomitant medications | Specific drug interaction studies, including interactions with MTX, have not been conducted. Concomitant MTX use may decrease the incidence of antiinfliximab antibody production and increase infliximab concentrations. |
| 2011 | | |
| Golimumab (Simponi) | 7 Drug interactions | |
| TNF blocker | | |
| 2011 | 7.1 Methotrexate | For the treatment of RA, Simponi should be used with MTX. (§14.1) Because the presence or absence of concomitant MTX did not appear to influence the efficacy or safety of Simponi in the treatment of PsA or AS, Simponi can be used with or without MTX in their treatment. (§12.3) |
| | 12 Clinical pharmacology | |
| | 12.3 Pharmacokinetics | Patients with RA, PsA, and AS treated with Simponi, 50 mg and MTX had approximately 52%, 36%, and 21% higher mean steady-state trough concentrations of golimumab, respectively, compared with those treated with Simponi 50 mg without MTX. The presence of MTX also decreased antigolimumab antibody incidence from 7% to 2%. (§6.1) For RA, Simponi should be used with MTX. In the PsA and AS trials, the presence or absence of concomitant MTX did not appear to influence clinical efficacy and safety parameters. (§14.1) |
| | | Patients who developed anti-Simponi antibodies generally had lower steady-state serum trough concentrations of Simponi. |

(*continued*)

**TABLE 10-3** (*Continued*)

| TP Name<br>(Brand Name)<br>Pharmacologic Class<br>Year of Labeling | Item | Labeling Language |
|---|---|---|
| Adalimumab<br>(Humira)<br>TNF blocker<br>2012 | 7 Drug interactions<br>7.1 Methotrexate | Humira has been studied in RA patients taking concomitant MTX. Although MTX reduced the apparent adalimumab clearance, the data do not suggest the need for dose adjustment of either Humira or MTX (§12.3) |
| | 12 Clinical pharmacology<br>12.3 Pharmacokinetics | In RA patients receiving 40 mg Humira every other week, adalimumab mean steady-state trough concentrations of approximately 5 μg/mL and 8–9 μg/mL, were observed without and with MTX, respectively. MTX reduced adalimumab apparent clearance after single and multiple dosing by 29% and 44%, respectively, in patients with RA.<br>Adalimumab mean steady-state trough concentrations were slightly higher in psoriatic arthritis patients treated with 40 mg Humira every other week (6–10 μg/mL and 8.5–12 μg/mL, without and with MTX, respectively) compared to the concentrations in RA patients treated with the same dose. |
| Certolizumab pegol<br>(Cimzia)<br>TNF blocker<br>2012 | 6 Adverse reactions<br>Immunogenicity | Antibody formation was associated with lowered drug plasma concentration and reduced efficacy. In patients receiving the recommended Cimzia dosage of 200 mg every other week with concomitant MTX, the ACR20 response was lower among antibody positive patients than among antibody-negative patients (Study RA-I, 48% versus 60%; Study RA-II 35% versus 59%, respectively). In Study RA-III, too few patients developed antibodies to allow for meaningful analysis of ACR20 response by antibody status. In Study RA-IV (monotherapy), the ACR20 response was 33% versus 56%, antibody-positive versus antibody-negative status, respectively. (§12.3) |

| | | |
|---|---|---|
| | 12 Clinical Pharmacology | |
| | 12.3 Pharmacokinetics: drug interaction studies | Methotrexate pharmacokinetics is not altered by concomitant administration with Cimzia in patients with RA. The effect of methotrexate on Cimzia pharmacokinetics was not studied. However, methotrexate-treated patients have lower incidence of antibodies to Cimzia. Thus therapeutic plasma levels are more likely to be sustained when Cimzia is administered with methotrexate in patients with RA. |
| 3. Substrate of the therapeutic enzyme | | |
| Glucarpidase (Voraxaze) Carboxypeptidase 2012 | Highlights of prescription information: drug interactions | Leucovorin is a substrate for Voraxaze. Do not administer leucovorin within 2 h before or after a dose of Voraxaze. (§7.1) Other potential exogenous substrates of Voraxaze include reduced folates and folate antimetabolites. (§7.2) |
| | 7 Drug interactions | |
| | 7.1 Use of Voraxaze with Leucovorin | Leucovorin is a substrate for Voraxaze. Do not administer leucovorin within 2 h before or after a dose of Voraxaze. No dose adjustment is recommended for the continuing leucovorin regimen because the leucovorin dose is based on the patient's pre-Voraxaze methotrexate concentration. (§12.3) |
| | 7.2 Other substrate interference | Other potential exogenous substrates of Voraxaze may include reduced folates and folate antimetabolites. |
| | 12 Clinical pharmacology | |
| | 12.3 Pharmacokinetics: drug interactions | In a study of cancer patients receiving a high-dose methotrexate ($\geq$1 g/m$^2$) and leucovorin rescue regimen, intravenous administration of 50 Units/kg Voraxaze 2 h before leucovorin reduced (6S)-leucovorin $AUC_{0-3h}$ by 33% and $C_{max}$ by 52% and also reduced its active metabolite, (6S)-5-methyltetrahydrofolate, $AUC_{0-3h}$ by 92% and $C_{max}$ by 93%. (§7.1) |
| 4. Mechanisms to be elucidated | | |
| Trastuzumab (Herceptin) HER2/neu receptor antagonist 2010 | 7 Drug interactions | In Study 5, the mean serum trough concentration of trastuzumab was consistently elevated approximately 1.5-fold, when administered in combination with paclitaxel compared to trough concentrations of trastuzumab when administered in combination with an anthracycline and cyclophosphamide. |
| Palifermin | | Heparin may increase systemic exposure (§7) |

(*continued*)

**TABLE 10-3** (*Continued*)

| TP Name (Brand Name) Pharmacologic Class Year of Labeling | Item | Labeling Language |
|---|---|---|
| (Kepivance) Mucocutaneous epithelial human growth factor 2011 | Highlights of prescription information: drug interactions | |
| | 7 Drug interactions | *In vitro* and *in vivo* data showed that palifermin interacts with unfractionated as well as low molecular weight heparins. Heparin coadministration resulted in a fivefold increase in palifermin systemic exposure. Avoid coadministration of palifermin with heparin. If heparin is used to maintain an intravenous line, rinse the line with saline before and after Kepivance administration. (§12.3) |
| | 12 Clinical pharmacology | |
| | 12.3 Pharmacokinetic: drug interactions | The potential PK interaction between palifermin and heparin was evaluated in a single-dose study in 27 healthy subjects receiving palifermin (60 μg/kg) coadministered with and without therapeutic levels of unfractionated heparin. This coadministration resulted in a fivefold increase in palifermin AUC and an 80% decrease in the mean CL. There was no significant effect of palifermin on heparin activity with respect to aPTT. The clinical relevance of this observed increase in palifermin systemic exposure is unclear. (§7) |

[a]Data from Ref. 26.

ACR20: American College of Rheumatology 20 Response; ANC: Absolute Neutrophil Count; aPTT: activated Partial Thromboplastin Time; AS: Ankylosing Spondylitis; 11βHSD-1: 11β-hydroxysteroid dehydrogenase type 1; CHC: Chronic Hepatitis C; CL: Clearance; CYP: Cytochrome P450; GH: Growth Hormone; HCV: Hepatitis C Virus; HER2: Human Epidermal Growth Factor Receptor 2; IL: Interlukin; IMDH: Inosine Monophosphate Dehydrogenase; INF: Interferon; MMF: Mycophenolate Mofetil; MPA: Mycophenolic Acid; 6MTITP: 6-Methylthioionosine Monophosphate; MTX: Methotrexate; NRTI: Nucleoside/Nucleotide Reverse Transcriptase Inhibitor; P-gp: P-glycoprotein; PK: Pharmacokinetics; PsA: Psoriatic Arthritis; RA: Rheumatoid Arthritis; TFN: Tumor Necrosis Factor.

amount of cytokine and/or cytokine modulator).[20,21,28] However, complex signaling may be involved in TP-DDI and it is not currently possible to ensure that cell culture systems provide the appropriate mix of cell types required for meaningful evaluation *in vitro*. Presently, *in vitro* and animal studies have limited value in the qualitative and quantitative projection of clinical interactions because translation of results from *in vitro* to *in vivo* and from animal to human has been inconsistent, necessitating *in vivo* drug interaction studies. Research that increases our understanding of CYP enzyme or transporter regulation by cytokines may increase the utility of *in vitro* studies. Standardized and optimal assay conditions and the identification of suitable positive controls to monitor assay performance are needed. Until such information is available, it is important to carry out *in vivo* risk-based drug interaction studies for TP-DDI assessments. Many TPs cannot be administered to healthy subjects, and cytokines have a different expression levels between healthy subjects and patients, therefore clinical studies (including TP-DDI studies) are conducted in patients. One efficient study design is the "cocktail" approach. Other *in vivo* study designs allow evaluation of the effect of the TP on substrates for specific CYP enzymes and transporters in targeted patient populations, as described in the next section. Table 10-3 provides an extensive list of labeling examples on PK-based TP-DDI.

Recent BLA review examples include tocilizumab (labeling based on *in vitro* and *in vivo* TP-DDI data information) and rilonacept (labeling without TP-DDI studies).

- If a TP will be used with other drug products (D or TP) as a combination therapy, studies should evaluate the effect of each product on the other. The studies should assess effects on PK and, when appropriate, PD of either drug. This evaluation is particularly important when a drug used in combination has a narrow therapeutic range such as chemotherapeutic agents.

  Examples include cancer therapies in which multiple drugs are combined, such as bevacizumab in combination with intravenous 5-fluorouracil-based chemotherapy including an IV bolus of IFL (irinotecan, 5-FU, and leucovorin) or FOLFOX-4 (oxaliplatin, 5-FU, and leucovorin) for the treatment of first- or second-line metastatic colorectal cancer, and trastuzumab in combination with doxorubicin, cyclophosphamide, and either paclitaxel or docetaxel or in combination with docetaxel and carboplatin for adjuvant treatment of HER2 overexpressing node positive or node negative (estrogen receptor/progesterone receptor [ER/PR] negative or with one high risk feature) breast cancer.[26]

- If there are known mechanisms for interactions or prior experience suggests potential PK or PD interactions, appropriate evaluations for possible interactions should be conducted. Some interactions between drugs and TPs are based on mechanisms other than CYP enzyme or transporter modulation.

  For example, methotrexate, an immunosuppressive agent may alter the clearance of concomitantly administered TPs through the reduction of antibodies formed against TP. Methotrexate decreased the clearance of infliximab

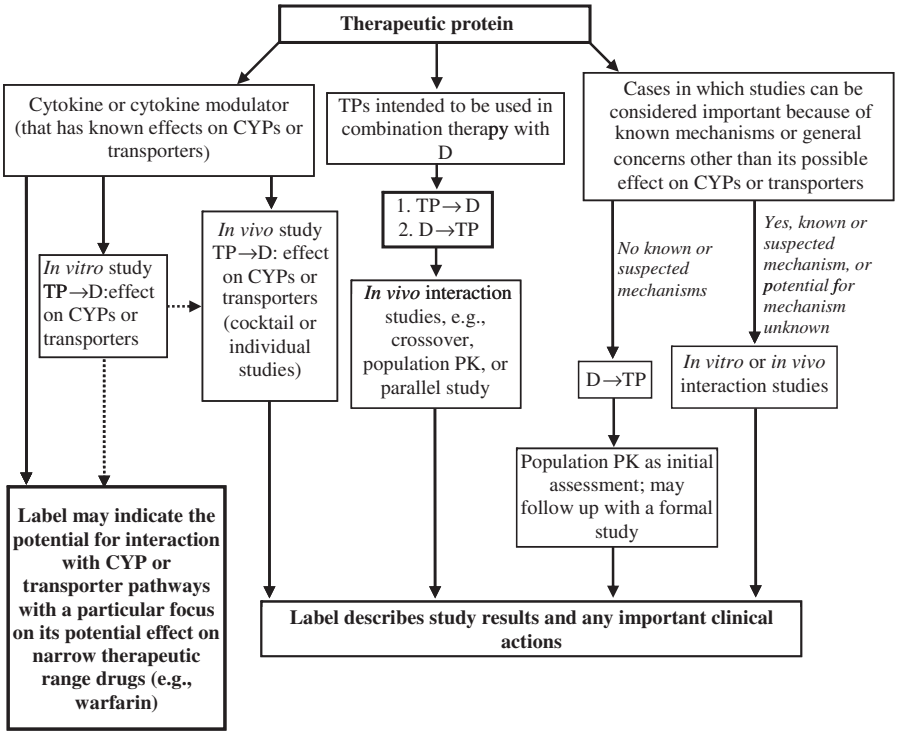


**FIGURE 10-1** Summary of the types of studies that have been used during drug development to evaluate therapeutic protein (TP)–small molecule drug (D) interactions. Evaluation includes the effect of TP on D (TP → D) and the effect of D on TP (D → TP). Broken lines: limited use of *in vitro* studies for informing *in vivo* study design or labeling; CYP: cytochrome P450; PK: pharmacokinetics. (Modified from Ref. 13.)

(a tumor necrosis factor α [TNF-α] blocker), which was associated with the decreased incidence of anti-infliximab antibody formation that affects infliximab clearance. At 8 weeks after a maintenance dose of 3–10 mg/kg of infliximab, median infliximab serum concentrations ranged from approximately 0.5 to 6 μg/mL; however, infliximab concentrations were not detectable (<0.1 μg/mL) in patients who tested positive for antibodies to infliximab.[26,27] The incidence of developing anti-infliximab antibodies was 7% and 21%, respectively, with or without concomitant administration of methotrexate at 12 weeks after last infliximab infusion of a 3 mg/kg dose.[26,27] Methotrexate may also affect the clearance of other TNF-α blocking agents. For example, it decreased the clearance of adalimumab by 29% after single dosing and by 44% after multiple dosing, which may result from reduced immunogenicity incidence; patients treated with concomitant methotrexate had a lower incidence of anti-adalimumab antibody formation than patients on adalimumab monotherapy (1% versus 12%).[26]

Other examples include heparin's effect on palifermin (increased exposure) and paclitaxel's effect on trastuzumab (increased exposure)[26] (Table 10-3), for which mechanisms have not been fully elucidated. In postmarketing studies, coadministration of palifermin with heparin resulted in a 5-fold increase in the systemic exposure of palifermin.[26,29] In clinical studies, administration of paclitaxel in combination with trastuzumab resulted in a 1.5-fold increase in trastuzumab serum levels.[26]

The recently approved glucarpidase label includes information about an interaction between methotrexate and leucovorin involving glucarpidase.[26] Glucarpidase is a carboxypeptidase enzyme indicated for the treatment of toxic plasma methotrexate concentrations ($>1$ μmole/L) in patients with delayed methotrexate clearance due to impaired renal function. In a study of cancer patients receiving a high-dose methotrexate ($\geq 1$ g per $m^2$) and leucovorin rescue regimen, intravenous administration of glucarpidase 2 h before leucovorin reduced (6S)-leucovorin $AUC_{0-3h}$ by 33% and $C_{max}$ by 52%, and also reduced its active metabolite, (6S)-5 methyltetrahydrofolate, $AUC_{0-3h}$ by 92% and $C_{max}$ by 93%.[26] This interaction is anticipated because both methotrexate and leucovorin are substrates of glucarpidase.

In addition to PK interactions, PD interactions can also occur between TP and drug or TP and TP. For example, an increased risk of serious infections has been observed in clinical studies of concurrent therapy for the treatment of rheumatoid arthritis, for eample, etanercept and anakinra, or etanercept and abatacept, without added benefit. Therefore, a warning is stated in the label of a newly approved TNF-α blocker, golimumab, about its risk in combination use with another biologics such as abatacept or anakinra. The labeling states that "use of Simponi (golimumab) with abatacept or anakinra is not recommended" (Table 10-4).[26]

Some PD interactions with TP-DDI have resulted in increased toxicity with combination chemotherapy. It is stated in the Vectibix labeling under Section 10.5 "Warnings and Precautions" that "Vectibix (panitumumab) is not indicated for use in combination with chemotherapy. In an interim analysis of Study 2, the addition of Vectibix to the combination of bevacizumab and chemotherapy resulted in decreased overall survival and increased incidence of NCI-CTC grade 3–5 (87% vs 72%) adverse reactions" (Table 10-4).[26]

As described in Figure 10-1, when there have been known mechanisms or prior experience suggesting potential PK or PD interactions, as illustrated by examples discussed earlier, appropriate *in vitro* and/or *in vivo* assessment for possible interactions have been requested and conducted.

## 10.3 TYPES OF STUDY DESIGNS

*In vitro* methods to evaluate effects of a drug on CYP enzymes and transporters are discussed in detail in other publications and on the FDA Drug Development and Drug Interaction website.[4–8] A white paper from the International Transporter Consortium provides recommendations on *in vitro* models to evaluate transporter-based interactions, including specific decision trees to evaluate the following transporters:

**TABLE 10-4 Examples of Labeling Language for Therapeutic Protein–Drug Interactions (Pharmacodynamics Based)**

| TP Names (Brand Names) Pharmacological Class; Year of Labeling | Item | Labeling Language |
|---|---|---|
| 1. Growth factors (Palifermin and G-CSF) and chemotherapy | | |
| Palifermin (Kepivance) Mucocutaneous Epithelial Human Growth Factor 2011 | Highlights of prescription information: drug interactions | Myelotoxic chemotherapy (§7; §14) |
| | 7. Drug interactions | Do not administer Kepivance within 24 h before, during infusion of, or within 24 h after administration of myelotoxic chemotherapy. (§2.1; §14) In a clinical trial, administration of Kepivance within 24 h of chemotherapy resulted in increased severity and duration of oral mucositis. |
| Filgrastim (Neupogen) Granulocyte colony-stimulating factor (G-CSF) 2012 | Precautions | The safety and efficacy of Neupogen given simultaneously with cytotoxic chemotherapy have not been established. Because of the potential sensitivity of rapidly dividing myeloid cells to cytotoxic chemotherapy, do not use Neupogen in the period 24 h before through 24 h after the administration of cytotoxic chemotherapy (see Dosage and Administration). |
| Pegfilgrastim (Neulasta) G-CSF 2007 | Precautions | Neulasta should not be administered in the period between 14 days before and 24 h after administration of cytotoxic chemotherapy (see Dosage and Administration) because of the potential for an increase in sensitivity of rapidly dividing myeloid cells to cytotoxic chemotherapy |
| 2. Growth factors and myeloablative therapy | | |
| Oprelvekin (Neumega) | Warnings | *Increased toxicity after myeloablative therapy*: Neumega is not indicated after myeloablative chemotherapy. In a randomized, placebo-controlled Phase II study, the |

| | | |
|---|---|---|
| Thrombopoietic growth factor (TGF, IL-11)<br>2009 | | effectiveness of Neumega was not demonstrated (see Clinical Studies, Study in Patients Following Myeloablative Chemotherapy). In this study, a statistically significant increased incidence in edema, conjunctival bleeding, hypotension, and tachycardia was observed in patients receiving Neumega as compared to placebo.<br>The following severe or fatal adverse reactions have been reported in postmarketing use in patients who received Neumega after bone marrow transplantation: fluid retention or overload (e.g., facial edema, pulmonary edema), capillary leak syndrome, pleural and pericardial effusion, papilledema and renal failure. |
| 3. Increased toxicities with concurrent therapy (psychotropic, nephrotoxic, myelotoxic, cardiotoxic or hepatotoxic drugs) | | |
| Aldesleukin (Proleukin)<br>Human recombinant IL-2<br>2011 | Drug interactions | Proleukin may affect central nervous function. Therefore, interactions could occur after concomitant administration of psychotropic drugs (e.g., narcotics, analgesics, antiemetics, sedatives, tranquilizers). Concurrent administration of drugs possessing nephrotoxic (e.g., aminoglycosides, indomethacin), myelotoxic (e.g., cytotoxic chemotherapy), cardiotoxic (e.g., doxorubicin) or hepatotoxic (e.g., methotrexate, asparaginase) effects with Proleukin may increase toxicity in these organ systems. The safety and efficacy of Proleukin in combination with any antineoplastic agents have not been established.<br>In addition, reduced kidney and liver function secondary to Proleukin treatment may delay elimination of concomitant medications and increase the risk of adverse events from those drugs. |
| Peginterferon α-2b (PegIntron)<br>INF-α<br>2011 | 7. Drug interactions<br>7.3 Use with ribavirin (nucleoside analogues) | Hepatic decompensation (some fatal) has occurred in cirrhotic HIV/HCV co-infected patients receiving combination antiretroviral therapy for HIV and INF-α and ribavirin. Adding treatment with INF-α alone or in combination with ribavirin may increase the risk in this patient subset. Patients receiving interferon with ribavirin and NRTIs should be closely monitored for treatment-associated toxicities, especially hepatic decompensation and anemia. Discontinuation of NRTIs should be considered as medically appropriate (see Individual NRTI Product Information). Dose reduction or discontinuation of interferon, ribavirin, or both should also be considered if worsening |

(*continued*)

**TABLE 10-4** (*Continued*)

| TP Names (Brand Names) Pharmacological Class; Year of Labeling | Item | Labeling Language |
|---|---|---|
| | | clinical toxicities are observed, including hepatic decompensation (e.g., Child-Pugh $< 6$). |
| 4. Increased infections with concurrent therapy | | |
| Etanercept (Enbrel) TNF blocker 2009 | Precaution: drug interactions | In a study in which patients with active RA were treated for $\leq 24$ weeks with concurrent Enbrel and anakinra therapy, a 7% rate of serious infections was observed, which was higher than that observed with Enbrel alone (0%). Two percent of patients treated concurrently with Enbrel and anakinra developed neutropenia (ANC $< 1 \times 10^9$/L |
| | | In a study of patients with Wegener's granulomatosis, the addition of Enbrel to standard therapy (including cyclophosphamide) was associated with a higher incidence of noncutaneous solid malignancies. The use of Enbrel in patients receiving concurrent cyclophosphamide therapy is not recommended (see Warnings: Malignancies and Adverse reactions: Malignancies). |
| | | Patients in a clinical study who were on established therapy with sulfasalazine, to which Enbrel was added, were noted to develop a mild decrease in mean neutrophil counts compared to groups treated with either Enbrel or sulfasalazine alone. The clinical significance of this observation is unknown. |
| Infliximab (Remicade) TNF blocker 2011 | Highlights of prescription information: drug interactions | Use with anakinra or abatacept increased risk of serious infections (§7.1) |
| | 7 Drug interactions | |
| | 7.1 Use with anakinra or abatacept | An increased risk of serious infections was seen in clinical studies of other TNF-α-blocking agents used in combination with anakinra or abatacept, with no added clinical benefit. Because of the nature of the adverse reactions seen with these |

| | | |
|---|---|---|
| | | combinations with TNF-blocker therapy, similar toxicities may also result from the combination of anakinra or abatacept with other TNF-α-blocking agents. Therefore, the combination of Remicade and anakinra or abatacept is not recommended. (§5.9; §5.10) |
| | 7.2 Use with tocilizumab | The use of tocilizumab in combination with biological DMARDs, such as TNF antagonists, including Remicade, should be avoided because of the possibility of increased immunosuppression and increased risk of infection. |
| Adalimumab (Humira) TNF blocker 2012 | Highlights of prescription information: drug interactions | Abatacept increased risk of serious infection (§5.1; §5.11; §7.2)<br>Anakinra increased risk of serious infections (§5.1; §5.7; §7.2) |
| | 7 Drug interactions | |
| | 7.2 Biologic products | In clinical studies in patients with RA, an increased risk of serious infections has been seen with the combination of TNF blockers with anakinra or abatacept, with no added benefit; therefore, use of Humira with abatacept or anakinra is not recommended in patients with RA. (§5.7; 5.11) A higher rate of serious infections has also been observed in patients with RA treated with rituximab who received subsequent treatment with a TNF blocker. There is insufficient information to provide recommendations regarding the concomitant use of Humira and other biologic products for the treatment of RA, PsA, AS, Crohn's disease, and plaque psoriasis. |
| Golimumab (Simponi) TNF blocker 2010 | Highlights of prescription information: drug interactions | Abatacept increased risk of serious infection (§5.1; §5.5; §7.2)<br>Anakinra increased risk of serious infections (§5.1; §5.6; §7.2) |
| | 7 Drug interactions | |
| | 7.2 Biologic products for RA, PsA, and/or AS | An increased risk of serious infections has been seen in clinical RA studies of other TNF blockers used in combination with anakinra or abatacept, with no added benefit; therefore, use of Simponi with abatacept or anakinra is not recommended. (§5.6) A higher rate of serious infections has also been observed in RA patients treated with rituximab who received subsequent treatment with a TNF blocker. There is insufficient |

(*continued*)

**TABLE 10-4** (*Continued*)

| TP Names (Brand Names) Pharmacological Class; Year of Labeling | Item | Labeling Language |
|---|---|---|
| | | information to provide recommendations regarding the concomitant use of Simponi and other biologic products approved to treat RA, PsA, or AS. |
| Certolizumab pegol (Cimzia) TNF blocker 2012 | Highlights of prescription information: drug interactions | Use with biological DMARDs increased risk of serious infections (§5.8; §7.1) |
| | 7 Drug interactions | |
| | 7.1 Use with anakinra, abatacept, rituximab, and natalizumab | An increased risk of serious infections has been seen in clinical studies of other TNF-blocking agents used in combination with anakinra or abatacept, with no added benefit. Formal drug interaction studies have not been performed with rituximab or natalizumab. Because of the nature of the adverse events seen with these combinations with TNF-blocker therapy, similar toxicities may also result from the use of Cimzia in these combinations. There is not enough information to assess the safety and efficacy of such combination therapy. Therefore, the use of Cimzia in combination with anakinra, abatacept, rituximab, or natalizumab is not recommended. (§5.8) |
| Canakinumab (Ilaris) IL-1β blocker 2009 | 7 Drug interactions | |
| | 7.1 TNF and IL-1 blockers | An increased incidence of serious infections and an increased risk of neutropenia have been associated with administration of another IL-1 blocker in combination with TNF inhibitors in another patient population. Use of Ilaris with TNF inhibitors may also result in similar toxicities and is not recommended because this may increase the risk of serious infections. (§5.1) |
| | | The concomitant administration of Ilaris with other drugs that block IL-1 has not been studied. Based on the potential for pharmacological interactions between Ilaris and a recombinant IL-1ra, concomitant administration of Ilaris and other agents that block IL-1 or its receptors is not recommended |

| | | |
|---|---|---|
| Rilonacept<br>(Arcalyst)<br>IL-1β blocker<br>2008 | 7 Drug interactions<br>7.1 TNF and IL-1 blockers | Specific drug interaction studies have not been conducted with Arcalyst. Concomitant administration of another drug that blocks IL-1 with a TNF-blocking agent in another patient population has been associated with an increased risk of serious infections and an increased risk of neutropenia. The concomitant administration of Arcalyst with TNF-blocking agents may also result in similar toxicities and is not recommended. (§5.1)<br>The concomitant administration of Arcalyst with other drugs that block IL-1 has not been studied. Based on the potential for pharmacologic interactions between rilonacept and a recombinant IL-1ra, concomitant administration of Arcalyst and other agents that block IL-1 or its receptors is not recommended. |
| 5. Increased myelotoxicity INF-α-2b and myelosuppressive agents) | | |
| INF-α-2b (Intron A)<br>INF-α<br>2011 | Warnings: drug interactions | Caution should be exercised when administering Intron A therapy in combination with other potentially myelosuppressive agents such as zidovudine. |
| Interferon alfa-2a<br>(Roferon-A)<br>Alpha interferon<br>2006 | Precautions: drug interactions | Caution should be exercised when administering Roferon-A in combination with other potentially myelosuppressive agents. Synergistic toxicity has been observed when Roferon-A is administered in combination with zidovudine (AZT) (see Warnings: Bone Marrow Toxicity).<br>The neurotoxic, hematotoxic or cardiotoxic effects of previously or concurrently administered drugs may be increased by INF. Interactions could occur following concurrent administration of centrally acting drugs. Use of Roferon-A in conjunction with IL-2 may potentiate risks of renal failure. |

(*continued*)

**TABLE 10-4** (*Continued*)

| TP Names (Brand Names) Pharmacological Class; Year of Labeling | Item | Labeling Language |
|---|---|---|
| 6. Increased toxicity with combination chemotherapy | | |
| Panitumumab (Vectibix) EGFR antagonist 2009 | Highlights of prescribing information: warnings and precautions | Increased toxicity with combination chemotherapy. Vectibix is not indicated for use in combination with chemotherapy (§5.3) |
| | 5 Warnings and precautions | |
| | 5.3 Increased toxicity with combination chemotherapy | Vectibix is not indicated for use in combination with chemotherapy. In an interim analysis of Study 2, the addition of Vectibix to the combination of bevacizumab and chemotherapy resulted in decreased overall survival and increased incidence of NCI-CTC grade 3–5 (87% vs 72%) adverse reactions. (§14) |
| 7. Others | | |
| Etanercept (Enbrel) TNF blocker 2009 | Precautions: drug interactions | In a study of patients with Wegener granulomatosis, the addition of Enbrel therapy to standard therapy (including cyclophosphamide) was associated with a higher incidence of noncutaneous solid malignancies. The use of Enbrel in patients receiving concurrent cyclophosphamide therapy is not recommended (see Warnings: Malignancies and Adverse Malignancies). |
| Somatropin | 7. Drug interactions | |

| | | |
|---|---|---|
| (Genotropin)<br>Human growth hormone<br>2011 | | |
| | 7.2 Pharmacologic glucocorticoid therapy and supraphysiologic glucocorticoid treatment | Pharmacologic glucocorticoid therapy and supraphysiologic glucocorticoid treatment may attenuate the growth-promoting effects of somatropin in children. Therefore, glucocorticoid replacement dosing should be carefully adjusted in children receiving concomitant somatropin and glucocorticoid treatments to avoid both hypoadrenalism and an inhibitory effect on growth. |
| Incobotulinumtoxi A<br>(Xeomin)<br>Acetylcholine release inhibitor and neuromuscular blocker<br>2011 | Highlights of prescribing information: drug interactions | Concomitant treatment of Xeomin and aminoglycoside antibiotics, spectinomycin, or other agents that interfere with neuromuscular transmission (e.g., tubocurarine-like agents), or muscle relaxants, should be observed closely because the effect of Xeomin may be potentiated. (§7) |
| | 7 Drug interactions | Coadministration of Xeomin and aminoglycoside antibiotics or other agents interfering with neuromuscular transmission (e.g., tubocurarine-type muscle relaxants) should be performed only with caution because these agents may potentiate the effect of the toxin.<br>Use of anticholinergic drugs after administration of Xeomin may potentiate systemic anticholinergic effects.<br>The effect of administering different botulinum toxin products at the same time or within several months of each other is unknown. Excessive neuromuscular weakness may be exacerbated by administration of another botulinum toxin before the resolution of the effects of a previously administered botulinum toxin. Excessive weakness may also be exaggerated by administration of a muscle relaxant before or after administration of Xeomin. |
| Pegloticase<br>(Krystexxa)<br>Recombinant uricase<br>2012 | 7 Drug interactions | Because antipegloticase antibodies appear to bind to the PEG portion of the drug, there may be potential for binding with other PEGylated products. The impact of anti-PEG antibodies on patients' responses to other PEG-containing therapeutics is unknown. |
| Abobotulinumtoxin A | 7 Drug interactions | |

(*continued*)

**TABLE 10-4** (*Continued*)

| TP Names (Brand Names) Pharmacological Class; Year of Labeling | Item | Labeling Language |
|---|---|---|
| (Dysport) Neuromuscular blocker 2012 | | Patients treated concomitantly with botulinum toxins and aminoglycosides or other agents interfering with neuromuscular transmission (e.g., curare-like agents) should be observed closely because the effect of the botulinum toxin may be potentiated. Use of anticholinergic drugs after administration of Dysport may potentiate systemic anticholinergic effects such as blurred vision.<br>The effect of administering different botulinum neurotoxin products at the same time or within several months of each other is unknown. Excessive weakness may be exacerbated by another administration of botulinum toxin before the resolution of the effects of a previously administered botulinum toxin.<br>Excessive weakness may also be exaggerated by administration of a muscle relaxant before or after administration of Dysport. |

[a]Data from Ref. 26.

ANC: Absolute Neutrophil Count; AS: Ankylosing Spondylitis; AZT: Azidothymidine; DMARD: Disease-Modifying Anti-Rheumatic Drugs; EGFR: Epidermal Growth Factor Receptor; G-CSF: Granulocyte colony-stimulating factor; HCV: Heptitis C Virus; IL: Interlukin; INF: Interferon; NCI-CTC: National Cancer Institute Common Toxicity Criteria; NRTI: Nucleoside/Nucleotide Reverse Transcriptase Inhibitor; PEG: Polyethylene Glycol; PsA: Psoriatic Arthritis; RA: Rheumatoid Arthritis; TGF: Thrombopoietic Growth Factor.

P-glycoprotein (P-gp), breast cancer resistance protein (BCRP), organic anion transporting polypeptide (OATP), organic anion transporter (OAT), and organic cation transporter (OCT). These models, when used with other *in vitro* and *in vivo* human data, can inform clinical study design and provide helpful information on drug labels.[30–32] However, at the present time, there are no well–validated *in vitro* models that can help predict TP-DDI. This knowledge gap was acknowledged at a AAPS/NBC TP-DDI workshop held in 2010.[12] Additional research is ongoing to further explore the utility of *in vitro* models for TP-DDI assessments.

An important consideration for most TP-DDI studies is the general need to conduct these studies in patient populations instead of healthy subjects. There are often concerns with the toxicity of the TP or drug in healthy subjects, and PK characteristics and/or target protein expression levels may differ between patients and healthy subjects. The need to conduct studies in patient populations has an impact on the feasibility of study designs.

Table 10-5 lists study designs that have been used in the *in vivo* evaluation of drug interactions involving TP products in human subjects. In studies with crossover designs, each subject serves as his or her own control, minimizing variability and allowing smaller sample size with increased statistical power. Many TP-DDI studies have been conducted using a one-way crossover

**TABLE 10-5 Examples of Study Designs Used to Assess Therapeutic Protein and Small Molecule Drug Interactions[a]**

| Study Designs[b] | Therapeutic Protein (TP) and Drug (D) Names | Study Purposes[c] |
|---|---|---|
| One-way crossover (sequential) | | |
| 1. D; D + TP (×8) | 1. Trastuzumab on paclitaxel | 1. TP → D |
| 2. D; TP; TP + D | 2. Bevacizumab on irinotecan (and SN38) | 2. TP → D |
| 3. D; D+TP (x63. | 3. Bevacizumab on carboplatin (or paclitaxel) | 3. TP → D |
| 4. D; D+TP (x4. | 4. Peginterferonα-2b on methadone | 4. TP → D |
| 5. TP + low D; TP + high D | 5. 5-FU on cetuximab | 5. D → TP |
| 6. D; TP + D | 6. Tocilizumab and probe drugs (simvastatin, omeprazole, or dextromethorphan) | 6. TP → D |
| Population pharmacokinetics | 1. Cisapride, paclitaxel, doxorubicin, gemcitabine, irinotecan on cetuximab | 1. D → TP |
| | 2. Bolus-IFL on bevacizumab | 2. D → TP |
| | 3. Comedications on tocilizumab | 3. D → TP |

(*continued*)

**TABLE 10-5** (*Continued*)

| Study Designs[b] | Therapeutic Protein (TP) and Drug (D) Names | Study Purposes[c] |
|---|---|---|
| Parallel group (within study) | | |
| 1. Group A: D; D + TP; D + TP; TP<br>Group B: TP; TP; TP; TP+D | 1. Cetuximab on irinotecan<br>Irinotecan on cetuximab | 1. TP → D[d,]<br>D → TP[e] |
| 2. Group A: D<br>Group B: TP<br>Group C: D1 + TP<br>Group D: D2 + TP | 2. Methotrexate, cyclophosphamide on rituximab | 2. D1 → TP<br>D2 → TP<br>TP → D1 |
| Cross-study comparison (historical control) | 1. Basiliximab or muromonab-CD3 and cyclosporine or tacrolimus | 1. TP → D |
| | 2. IFN-α and theophylline (patients vs. healthy) | 2. TP → D |
| Others | IL-6 or TNF-α and a cocktail (evaluation of the correlation of IL-6 level with enzyme activities) | TP → D |

[a]Data from Refs. 13,14,26,33–38.
[b]Study design: the administration of D or TP was listed sequentially; for example, "D; TP; TP + D" indicates that D alone was given followed by TP alone, which was followed by TP and D given together; "D; D + TP (×4)" indicates that D alone was given followed by D and TP given together for 4 cycles (days or weeks).
[c]Study purpose: TP → D denotes a study in which the effect of TP on D is examined; D → TP denotes a study in which the effect of D on TP is examined.
[d]Comparison of D exposure was made between Groups A (first D) and Group B (TP + D).
[e]This portion of study is a one-way crossover (sequential) design.
5-FU:5-Fluorouracil; IL-6: Interlukin-6; INF-α: Interferon-α; TNF-α: Tumor Necrosis Factor-α.

(sequential) design. This study design is often used for evaluation of the effect of TP on concomitant chemotherapy agents (TP → D) in the oncology clinical setting. In the evaluation of TP-DDI between bevacizumab and carboplatin, patients received carboplatin first and bevacizumab was then added to the regimen. PK information collected in both occasions was evaluated to determine the effect of bevacizumab on carboplatin (Table 10-5). This within-study crossover design may not be feasible, however, for the assessment of the effect of chemotherapy agents on TP (D → TP). Many TP products, such as mAbs, have long half-lives, nonlinear PK because of a component of saturable target-mediated clearance, and potential for formation of antiproduct antibodies (APAs) that may alter their own clearance. These unique characteristics of TP produce unavoidable sequence effects, which preclude the use of crossover designs for the assessment of the effect of chemotherapy agents on TP in the combination

therapy (D → TP). Alternative study designs, such as parallel group study or population PK analysis, have therefore been used for such assessments.

Population PK studies have been conducted to evaluate drug interactions, and sponsors have been encouraged to include population PK sample collection (and PD, if appropriate) and data analysis in late phase clinical trials to evaluate patient factors and concomitant medications that may contribute to the variability in systemic exposure and/or drug response for small molecular drugs.[2,4–8] Population PK analyses have also been used to determine the effect of concomitant medications on PK and PD of TPs. Table 10-5 includes examples of the use of a population PK approach for evaluating the effect of a drug on a TP (D → TP). For example, the tocilizumab label states that population PK analyses did not detect any effect of methotrexate, nonsteroidal anti-inflammatory drugs, or corticosteroids on tocilizumab clearance.[26] In most cases, the concentrations of concomitantly administered drugs are not determined in these clinical trials, and therefore the studies do not provide information on the effect of a TP on the exposure of concomitantly used drugs (TP → D).

Table 10-5 also shows examples of parallel study design used in the assessment of the TP → D or D → TP effects. In addition to parallel design studies, some assessments were made using a cross-study evaluation, which may involve comparisons of data between healthy subjects and patients or comparisons of data between combination therapy and monotherapy. Data from such analyses should be considered as "hypothesis-generating" and the results generally need to be confirmed.

Finally, Table 10-5 presents information collected in association studies, such as the evaluation of correlation between cytokine levels and enzyme activities. Association studies provide mechanistic understanding of the effects of cytokines on CYP enzyme or transporter activities and predict effects of various cytokine modulators on the activities of specific CYP enzymes or transporters.

## 10.4 LABELING IMPLICATIONS

When a drug product's optimal dose for an indication has been selected based on assessment of the overall relation of risk-to-benefit, it is relatively straightforward to determine dosing adjustments needed for patients with specific intrinsic and extrinsic factors that alter their PK profiles: dosing adjustments for individuals with altered PK profiles are generally made to achieve the same systemic exposure as has been shown to be safe and effective for patients without altered PK profiles.[4–8,12] The principles that guide recommendations for drug dosing adjustments have been used in the interpretation of the results from TP-DDI studies. Details of the drug interaction results have been included in the pharmacokinetics subsection (12.3) of the clinical pharmacology section of the labeling.[2,39]

Tables 10-3 and 10-4 provide examples of TP-DDI information (PK and PD based, respectively) in current product labels. Although some TP-DDI information is included in TP labels, unlike small molecule drugs, where some drug interactions have resulted in specific labeling language in the sections titled "Highlights,"

"Contraindications," and "Dosage and Administration," most TP labels include descriptive information about drug interactions in the sections called "Drug interactions," "Clinical Pharmacology," and "Warnings and Precautions," suggesting lesser perceived clinical significance of these interactions with TPs. In particular, the current labels do not recommended specific dosing adjustments for TP-DDI in the "Dosage and Administration" section.

The label for tocilizumab, an interleukin 6 (IL-6) receptor inhibitor approved in the United States for rheumatoid arthritis and systemic juvenile idiopathic arthritis, includes results of studies that demonstrate an effect of tocilizumab on the PK of CYP2C19 or CYP3A substrates.[24–26] CYP enzymes in the liver are downregulated by infection and inflammation stimuli, including cytokines such as IL-6. Inhibition of IL-6 signaling in rheumatoid arthritis patients treated with tocilizumab may restore CYP enzyme activities to higher levels than those not treated with tocilizumab leading to increased metabolism of drugs that are CYP enzyme substrates. The tocilizumab labeling states that the effect of tocilizumab on CYP enzymes may be clinically relevant for CYP450 substrates with a narrow therapeutic index, where the dose is individually adjusted. It further recommends that

> Upon initiation or discontinuation of tocilizumab, in patients being treated with these types of medicinal products, therapeutic monitoring of effect (e.g., warfarin) or drug concentration (e.g., cyclosporine or theophylline) should be performed and the individual dose of the medicinal product adjusted as needed.[26]

The development of physiologically based PK (PBPK) models linking the *in vitro* CYP suppression data for tocilizumab to the *in vivo* CYP alternation (measured via changes in systemic exposure of probe substrates, such as simvastatin) helps explain the association between IL-6 levels and CYP suppression.[40] PBPK models have been shown to be a useful tool in clinical pharmacology evaluations, including drug–drug interactions involving small molecule drugs.[41,42]

Several TP labels included language regarding the potential for drug interactions (either at the time of market approval or at their labeling updates). These examples include rilonacept (IL-1β blocker), golimumab (TNF-α blocker), balatacept (IL-2 modulator), infliximab (TNF-α blocker), and ustekinumab (anti-IL12/IL23 mAb) (Table 10-3). Labeling language for these cytokine modulators indicates that increased levels of cytokines (such as IL-1, IL-6, IL-10, TNF-α, IFN-α) during inflammation can alter the expression or activity of CYP enzymes, and that these cytokine modulators may normalize the activity of CYP enzymes, thus affecting systemic exposure of CYP enzyme substrates. The labels include a general recommendation to monitor the effect on drug concentration and to adjust the dose as needed for CYP enzyme substrates with a narrow therapeutic index (such as warfarin, cyclosporine, theophylline). However, no specific TP-DDI study results have been included in the labels of these TPs. In some cases, postmarketing drug interaction studies that provide useful information are added to the labels at a later date.[29]

## 10.5 CONCLUSION

Information about drug–drug interactions is critical to public health. The use of therapeutic proteins has increased in recent years, and they are often used in combination with small molecule drugs in critical therapeutic areas. It is anticipated that as more data are collected and the understanding of potential therapeutic protein–drug interactions increases, we will be better able to assess the potential for TP-DDIs during the development of therapeutic proteins.

## REFERENCES


1. Huang SM, Temple R. Is this the drug or dose for you? Impact and consideration of ethnic factors in global drug development, regulatory review, and clinical practice. *Clin Pharmacol Ther* **84**, 287–94 (2008).
2. US Food and Drug Administration. Clinical pharmacology guidances. Available at www.fda.gov/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/ucm064982.htm.
3. US Food and Drug Administration, US Department of Health and Human Services, Center for Drug Evaluation and Research, Center for Biologics Evaluation and Research (CBER). Draft guidance for industry: drug interaction studies–study design, data analysis, implications for dosing, and labeling. Silver Spring, MD, 2012. Available at http://www.fda.gov/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/ucm064982.htm
4. Huang SM, Temple R, Throckmorton DC, Lesko LJ. Drug interaction studies: study design, data analysis, and implications for dosing and labeling. *Clin Pharmacol Ther* **81**, 298–304 (2007).
5. Huang SM, Strong JM, Zhang L, et al. New era in drug interaction evaluation: US Food and Drug Administration update on CYP enzymes, transporters, and the guidance process. *J Clin Pharmacol* **48**, 662–70 (2008).
6. Zhang L, Zhang YD, Huang SM. Scientific and regulatory perspectives on metabolizing enzyme-transporter interplay and its role in drug interactions–challenges in predicting drug interactions. *Mol Pharm* **6**, 1766–74 (2009).
7. Zhang L, Zhang YD, Zhao P, Huang SM. Predicting drug-drug interactions: an FDA perspective. *AAPS J* **11**, 300–06 (2009).
8. Zhang L, Reynolds KS, Zhao P, Huang SM. Drug interactions evaluation: an integrated part of risk assessment of therapeutics. *Toxicol Appl Pharmacol* **243**, 134–45 (2010).
9. US Food and Drug, Administration. Drug Development and Drug Interaction. Available at www.fda.gov/Drugs/DevelopmentApprovalProcess/DevelopmentResources/DrugInteractionsLabeling/ucm080499.htm.
10. Mahmood I, Green MD. Drug interaction studies of therapeutic proteins or monoclonal antibodies. *J Clin Pharmacol* **47**, 1540–54 (2007).
11. Seitz K, Zhou H. Pharmacokinetic drug-drug interaction potentials for therapeutic monoclonal antibodies: reality check. *J Clin Pharmacol* **47**, 1104–18 (2007).
12. Girish S, Martin SW, Peterson MC, et al. AAPS workshop report: strategies to address therapeutic protein-drug interactions during clinical development. *AAPS J* **13**, 405–16 (2011).

13. Huang SM, Zhao H, Lee JI, et al. Therapeutic protein-drug interactions and implications for drug development. *Clin Pharmacol Ther* **87**, 497–503 (2010).
14. Lee JI, Zhang L, Men AY, et al. CYP-mediated therapeutic protein-drug interactions: clinical findings, proposed mechanisms and regulatory implications. *Clin Pharmacokinet* **49**, 295–310 (2010).
15. Lloyd P, Zhou H, Theil FP, et al. Highlights from a recent BIO survey on therapeutic protein-drug interactions. *J. Clin Pharm*. 52(11):1755–1763, 2012 (Epub Dec 13, 2011).
16. Zhou H, Mascelli MA. Mechanisms of monoclonal antibody-drug interactions. *Annu Rev Pharmacol Toxicol* **51**, 359–72 (2011).
17. Kraynov E, Martin SW, Hurst S, et al. How current understanding of clearance mechanisms and pharmacodynamics of therapeutic proteins can be applied for evaluation of their drug-drug interaction potential. *Drug Metab Dispos* **39**, 1779–83 (2011).
18. Le VM, Gripon P, Stieger B, Fardel O. Down-regulation of organic anion transporter expression in human hepatocytes exposed to the proinflammatory cytokine interleukin 1beta. *Drug Metab Dispos* **36**, 217–22 (2008).
19. Le VM, Lecureur V, Moreau A, et al. Differential regulation of drug transporter expression by hepatocyte growth factor in primary human hepatocytes. *Drug Metab Dispos* **37**, 2228–35 (2009).
20. Morgan ET, Goralski KB, Piquette-Miller M, et al. Regulation of drug-metabolizing enzymes and transporters in infection, inflammation, and cancer. *Drug Metab Dispos* **36**, 205–16 (2008).
21. Morgan ET. Impact of infectious and inflammatory disease on cytochrome P450-mediated drug metabolism and pharmacokinetics. *Clin Pharmacol Ther* **85**, 434–38 (2009).
22. Petrovic V, Teng S, Piquette–Miller M. Regulation of drug transporters during infection and inflammation. *Mol Interv* **7**, 99–111 (2007).
23. Vee ML, Lecureur V, Stieger B, Fardel O. Regulation of drug transporter expression in human hepatocytes exposed to the proinflammatory cytokines tumor necrosis factor-alpha or interleukin-6. *Drug Metab Dispos* **37**, 685–93 (2009).
24. Schmitt C, Kuhn B, Zhang X, et al. Disease-drug-drug interaction involving tocilizumab and simvastatin in patients with rheumatoid arthritis. *Clin Pharmacol Ther* **89**, 735–40 (2011).
25. Zhang X, Schmitt C, Grange S, K, et al. Disease-drug interaction studies of tocilizumab with cytochrome P450 substrates in vitro and in vivo. *Clin Pharm Ther* **85**(ser. 1), 59 (2009).
26. US Food and Drug Administration. Drugs@FDA. Available at www.accessdata.fda.gov/scripts/cder/drugsatfda/index.cfm.
27. Maini RN, Breedveld FC, Kalden JR, et al. Therapeutic efficacy of multiple intravenous infusions of anti-tumor necrosis factor alpha monoclonal antibody combined with low-dose weekly methotrexate in rheumatoid arthritis. *Arthritis Rheum* **41**, 1552–63 (1998).
28. Dickmann LJ, Patel SK, Rock DA, et al. Effects of interleukin-6 (IL-6) and an anti-IL-6 monoclonal antibody on drug-metabolizing enzymes in human hepatocyte culture. *Drug Metab Dispos* **39**, 1415–22 (2011).
29. US Food and Drug Administration. Postmarket Requirements and Commitments: Introduction. Available at www.fda.gov/Drugs/GuidanceComplianceRegulatoryInformation/Post-marketingPhaseIVCommitments/default.htm.

30. Giacomini KM, Huang SM, Tweedie DJ, et al. Membrane transporters in drug development. *Nat Rev Drug Discov* **9**, 215–36 (2010).
31. Huang SM, Zhang L, Giacomini KM. The International Transporter Consortium: a collaborative group of scientists from academia, industry, and the FDA. *Clin Pharmacol Ther* **87**, 32–36 (2010).
32. Huang SM, Woodcock J. Transporters in drug development: advancing on the critical path. *Nat Rev Drug Discov* **9**, 175–76 (2010).
33. Folprecht G, Lutz MP, Schoffski P, et al. Cetuximab and irinotecan/5-fluorouracil/folinic acid is a safe combination for the first-line treatment of patients with epidermal growth factor receptor expressing metastatic colorectal carcinoma. *Ann Oncol* **17**, 450–56 (2006).
34. Leyland-Jones B, Gelmon K, Ayoub JP, et al. Pharmacokinetics, safety, and efficacy of trastuzumab administered every three weeks in combination with paclitaxel. *J Clin Oncol* **21**, 3965–71 (2003).
35. Denlinger CS, Blanchard R, Xu L, et al. Pharmacokinetic analysis of irinotecan plus bevacizumab in patients with advanced solid tumors. *Cancer Chemother Pharmacol* **65**, 97–105 (2009).
36. Breedveld F, Agarwal S, Yin M, et al. Rituximab pharmacokinetics in patients with rheumatoid arthritis: B-cell levels do not correlate with clinical response. *J Clin Pharmacol* **47**, 1119–28 (2007).
37. Delbaldo C, Pierga JY, Dieras V, et al. Pharmacokinetic profile of cetuximab (Erbitux) alone and in combination with irinotecan in patients with advanced EGFR-positive adenocarcinoma. *Eur J Cancer* **41**1739–45 (2005).
38. Lu JF, Bruno R, Eppler S, et al. Clinical pharmacokinetics of bevacizumab in patients with solid tumors. *Cancer Chemother Pharmacol* **62**, 779–86 (2008).
39. US Food and Drug Administration, US Department of Health and Human Services, Center for Drug Evaluation and Research, Center for Biologics Evaluation and Research (CBER). Guidance for Industry: Labeling for Human Prescription Drug and Biological Products—Implementing the New Content and Format Requirements. Rockville, MD, 2006. Available at http://www.fda.gov/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/ucm065010.htm
40. Kenny J. Cytokine mediated regulation of human cytochrome P450—can we translate in vitro evaluation to clinical relevance? Paper presented at the AAPS/National Biotechnology Conference, San Francisco, May 17, 2011.
41. Huang SM, Rowland M. The role of physiologically based pharmacokinetic modeling in regulatory review. *Clin Pharmacol Ther* **91**, 542–49 (2012).
42. Zhao P, Zhang L, Grillo JA, et al. Applications of physiologically based pharmacokinetic (PBPK) modeling and simulation during regulatory review. *Clin Pharmacol Ther* **89**, 259–67 (2011).

# DISEASE–DRUG–DRUG INTERACTION ASSESSMENTS FOR TOCILIZUMAB—A MONOCLONAL ANTIBODY AGAINST INTERLEUKIN-6 RECEPTOR TO TREAT PATIENTS WITH RHEUMATOID ARTHRITIS

XIAOPING ZHANG and BARBARA J. BRENNAN

## 11.1 INTRODUCTION

Tocilizumab (TCZ, Actemra) is a recombinant humanized monoclonal antibody of the immunoglobulin G (IgG) subclass that selectively binds to human interleukin 6 (IL-6) receptors. TCZ binds to both membrane-bound and soluble forms of IL-6R, blocking signal transduction pathways through competitive inhibition of IL-6 binding.[1] TCZ 8 mg/kg given intravenously (IV) is currently approved for the treatment of rheumatoid arthritis (RA) in over 106 countries, including the United States,[2] Canada,[3] Australia,[4] European countries,[5] and Japan.[6] IL-6 receptor blockade was considered as a mechanism for pharmacologic treatment of RA due to the central role of IL-6 in inflammatory processes. IL-6 is a pleiotropic cytokine with numerous activities, including effects on the immune response and inflammation (e.g., interaction with neutrophils in the synovium), bone metabolism, and hematopoiesis.[7,8] Elevated IL-6 levels have been found in the serum and synovial fluid of RA patients, and inflammatory cell infiltration has been correlated with IL-6 levels in synovial tissues.[9] Furthermore, IL-6 levels correlate with disease activity in RA patients and improvement of the disease after treatment with disease-modifying antirheumatic drugs (DMARDs) is accompanied by a reduction in serum IL-6.[10]

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*,
Edited by Shayne C. Gad
Copyright © 2013 John Wiley & Sons, Inc.

A systematic review of TCZ for RA has been reported.[11] Ten published clinical trials (seven Phase III, three Phase II) for TCZ were retrieved. Compared with methotrexate (MTX) monotherapy, TCZ 8 mg/kg IV monotherapy had higher ACR20 ($p < 0.001$), ACR50 ($p < 0.002$), and ACR70 ($p < 0.001$) response rates at week 24. TCZ 8 mg/kg IV plus oral MTX had a higher ACR20 response rate than oral MTX plus placebo in patients with RA who failed to respond to MTX or anti–tumor necrosis factor therapy ($p < 0.001$). Patients receiving TCZ 8 mg/kg had less radiographic progression as assessed by Genant-modified Sharp score (85% had no progression progression) than the control group (67% had no progression) ($p < 0.001$). In the all exposure population composed of 4009 RA patients (3577 received treatment for at least 6 months, 3309 for at least one year; 2954 received treatment for at least 2 years and 2189 for 3 years), the rate of serious infections was 4.7 events/100 patient-years of exposure in all TCZ exposure groups containing 4009 patients (3577 received TCZ for at least 6 months, 3309 for at least one year, 2595 for at least 2 years and 2189 for 3 years).[2] A greater frequency of neutropenia, thrombocytopenia, hyperlipidemia, and transaminitis was observed with TCZ compared with placebo.

Following 4 and 8 mg/kg administration of TCZ to RA patients, serum TCZ concentration-time profiles were adequately described by a 2-compartment disposition pharmacokinetic model with parallel linear and nonlinear elimination kinetics. At low TCZ concentrations, the target-mediated elimination (e.g., via IL-6R) is the predominant elimination pathway. As the concentration of TCZ increases, the target is gradually saturated, and linear clearance by the nonspecific reticuloendothelial system becomes a predominant elimination pathway. Total TCZ clearance is therefore increased with time as TCZ concentration decreases after administration. As a result, the terminal half-life of TCZ increases as TCZ concentration decreases during a dosing interval. The half-life of TCZ was up to 11 to 13 days for therapeutic dose of 4 and 8 mg/kg, respectively, every 4 weeks at steady state.[12]

Several biomarkers of TCZ activity have been routinely measured in clinical pharmacology studies as well as in Phase III studies of TCZ, including IL-6 plasma levels, soluble IL-6 receptor plasma levels, and C-reactive protein (CRP). In RA patients, mean plasma IL-6 concentrations increase significantly after TCZ IV infusion and remain elevated for at least 4 weeks after a single 10 mg/kg dose and return to baseline after approximately 8 weeks.[13] This increase in IL-6 is thought to occur as a result of IL-6R blockade inhibiting clearance of IL-6 from the serum.[14,15] Similarly, increases in the soluble IL-6 receptor are also observed after administration of TCZ. CRP is an acute phase protein that is elevated in the serum of RA patients and used as a marker of inflammation. As a result of IL-6R blockade by TCZ, CRP serum levels are reduced to within the normal range by approximately 7 days postinfusion.[13]

The role of proinflammatory cytokines, such as interleukins, interferons, and tumor necrosis factor α (TNF-α), in regulation of cytochrome P450 (CYP) mediated drug metabolism has been well described.[16–18] There are several examples of altered drug metabolism in patients with acute or chronic inflammatory conditions, such as influenza and cancer. Furthermore, several reports have described the ability

of interferon to reduce the metabolic activity of several CYP enzymes in clinical drug–drug interactions studies.[19–21]

IL-6 has been reported to down-regulate several CYP isoforms *in vitro*, with CYP3A4 being the most sensitive.[22–24] RA patients have elevated mean IL-6 levels (~50–60 pg/mL) as compared to approximately 5 pg/mL or less in healthy subjects.[25] Treatment with TCZ blocks the IL-6 signaling pathway and therefore has the potential to reverse the downregulation of the CYP system and restore metabolic activity. In this situation, a decrease from pretreatment baseline concentrations would be expected for drugs that are substrates of CYP upon initiation of treatment with TCZ.

## 11.2 PRECLINICAL EVALUATION

As a first step to investigate the potential for disease–drug–drug interactions (DDDIs) resulting from reduced IL-6 signaling, an *in vitro* assessment in human hepatocytes was performed.[26] The effects of TCZ and IL-6 on hepatic drug metabolizing enzyme expression levels were investigated. The hepatocytes were obtained as cells seeded on 24 well plates. Expression levels of mRNA for CYP3A4, 2D6, 1A2, 2B6, 2C19, and 2C9 were evaluated in the presence and absence of IL-6 at a low concentration of 0.5 ng/mL and at a high concentration of 12.5 ng/mL. Expression levels were also assessed in the presence of 12.5 ng/mL IL-6 plus 250 μg/mL of TCZ. All incubations lasted 72 h.

Results of the hepatocyte experiment are provided in Table 11-1. IL-6 had a profound impact on CYP3A4 expression, particularly at the high concentration of 12.5 ng/mL. No changes were apparent in the other isoforms at the low IL-6 concentration of 0.5 ng/mL; however, there was suppression of all isoforms at the high concentration of 12.5 ng/mL. The addition of TCZ at a concentration of 250 μg/mL to the incubation was able to completely prevent the IL-6-mediated downregulation of CYP3A4. A similar effect was apparent for the other CYP isoforms to a lesser extent than for CYP3A4. These *in vitro* findings supported that a clinical

**TABLE 11-1 Meanm RNA Expression Levels Relative to Control in Human Hepatocytes[a,b]**

| IL-6 (ng/mL) | TCZ (μg/mL) | CYP Isozyme | | | | | |
|---|---|---|---|---|---|---|---|
| | | 3A4 | 2D6 | 1A2 | 2B6 | 2C19 | 2C9 |
| 0 | 0 | **1.00 (0)** | 1.00 (0) | 1.00 (0) | 1.00 (0) | 1.00 (0) | 1.00 (0) |
| 0.5 | 0 | **0.79 (0.14)** | 1.69 (0.29) | 1.03 (0.31) | 1.08 (0.36) | 3.76 (2.02) | 1.15 (0.49) |
| 12.5 | 0 | **0.06 (0.07)** | 0.67 (0.28) | 0.18 (0.13) | 0.28 (0.12) | 0.43 (0.29) | 0.58 (0.25) |
| 12.5 | 250 | **0.95 (0.82)** | 0.68 (0.42) | 0.76 (0.23) | 0.92 (0.34) | 0.96 (0.87) | 1.44 (0.63) |

[a]From Ref. 26.
[b]Standard deviations are in parentheses.
TCZ: tocilizumab.

disease–drug–drug interaction between TCZ and drugs that are metabolized via CYP enzymes should be evaluated in RA patients to assess the clinical relevance of the *in vitro* findings.

## 11.3 CLINICAL DDDI EVALUATIONS

Following the *in vitro* assessment, three clinical DDDI assessments were performed in patients with RA. Because the mechanism underlying the potential clinical DDDI is a function of the increased IL-6 levels in RA patients, it was essential that the clinical studies were performed in RA patients as opposed to healthy subjects. The effect of TCZ on the pharmacokinetics (PK) exposure of omeprazole (a sensitive CYP2C19 substrate with some CYP3A4 involvement),[26] dextromethorphan (a sensitive CYP2D6 substrate with some CYP3A4 involvement),[26] and simvastatin (a sensitive CYP3A4 substrate),[27] was therefore studied in RA patients. In addition to the studies that were planned with a reasonable mechanistic hypothesis, supported by *in vitro* data suggesting a clinical DDDI could be likely, the effect of TCZ on methotrexate (MTX) PK was also evaluated clinically because MTX is a commonly administered concomitant medication for the treatment of RA.[29] IL-6 has been reported *in vitro* to suppress mRNA of some anion transporters such as MRP2 that are involved in MTX elimination.[28] Elevated IL-6 levels present in RA patients could therefore be associated with downregulation of these transporters. If true, TCZ may reverse this IL-6-induced downregulation of transporters as it does for CYP3A4. Therefore DDDI for TCZ and MTX in RA patients was also evaluated.[29] The potential for other common concomitant medications in RA patients to alter the PK of TCZ was evaluated by population PK analysis.[12]

The first clinical DDDI study was a two-part study that evaluated the effect of TCZ on the PK of omeprazole (OME) and dextromethorphan (DEX) in 18 and 13 patients with RA, respectively. Patients diagnosed with RA based on the 1987 American College of Rheumatology (ACR) criteria[30] with CRP $> 1.5$ mg/dL, were enrolled. DMARDs or nonsteroidal anti-inflammatory drugs (NSAIDs) had been stabilized for at least 2 weeks before receiving the study drug on day 1. More criteria for inclusion and exclusion from the study can be found elsewhere.[26] All patients who participated in the OME evaluation were genotyped for CYP2C19 metabolizer status.

The second clinical DDDI study was a two-part study evaluated the effect of TCZ on the PK of simvastatin (SIM) in 12 patients in Group 1 and effect of TCZ on the PK of methotrexate (MTX) in 12 patients in Group 2. Patients diagnosed with RA based on the 1987 ACR criteria[30] with CRP $> 1.5$ mg/dL, were enrolled. MTX was given for at least 12 weeks immediately before day 1, of which the last 8 weeks were at a stable dose between 10 and 25 mg/week. Patients on other DMARDs were excluded. NSAIDs and oral corticosteroids (CSs) were stabilized for at least 2 weeks before receiving the study drug on day 1. More criteria for inclusion and exclusion from the study can be found elsewhere.[27]

### 11.3.1 Omeprazole DDDI Evaluation

The OME DDDI study was an open-label, two period, single-dose crossover study in 18 patients with RA.[13,26] Of the 18 patients, 5 were classified as CYP2C19 poor metabolizers (PMs), 8 were extensive metabolizers (EMs), and 5 were considered intermediate metabolizers (IMs). All patients received a single dose of OME 10 mg orally on day 1, followed by a single 1-h IV infusion of 8 mg/kg of TCZ on day 8 and then another single oral 10 mg dose of OME 7 days after the TCZ infusion. Seven days was selected as the appropriate time period necessary to evaluate changes in CYP-mediated metabolism as a result of inhibition of IL-6 signal transduction. Serial blood samples were collected after each OME dose for measurement of plasma OME concentrations.

The mean (%CV) OME PK parameters in the absence and presence of TCZ are provided in Table 11-2 for each of the CYP2C19-metabolizer groups.

As expected, plasma exposure of OME at baseline was lower in the EM group as compared to the IM or PM groups. One week after administration of TCZ, plasma OME concentrations were lower than those measured at baseline. The geometric mean ratio of $AUC_{0-\infty}$ and $C_{max}$ and associated 90% confidence intervals for each group are provided in Table 11-2. The decrease was more pronounced in the EM group as compared to the IM or PM group, as expected, because patients with the EM phenotype should be more sensitive to an inflammation-associated reduction in CYP2C19 activity. In the EM group, OME $AUC_{0-\infty}$ was reduced by approximately 28% as compared to an approximate 12% decrease in both the IM and PM groups.

### 11.3.2 Dextromethorphan DDDI Evaluation

The DDDI study with DEX was an open-label, single-dose cross-over study conducted in RA patients. Thirteen RA patients received a single 30-mg dose of DEX alone and a single 30-mg dose of DEX 1 week after receiving a single 1-h IV infusion of TCZ 8 mg/kg. Blood samples were collected for evaluation of DEX plasma exposure. DEX is metabolized primarily by CYP2D6 to dextrorphan, which then undergoes further metabolism via CYP3A4. Therefore, both DEX and dextrorphan plasma concentrations were measured to assess effects on both CYP2D6 and CYP3A4.

After administration of TCZ, plasma exposure to DEX was essentially unchanged, as shown in Table 11-3, with $AUC_{0-\infty}$ geometric mean ratio of approximately 1 and associated 90% confidence intervals for $AUC_{0-\infty}$ between the equivalence limits of 80% and 125%. However, exposure to the major metabolite, dextrorphan, a CYP3A4 substrate, was approximately 30% lower in the presence of TCZ, suggesting a potential effect of TCZ on CYP3A4 activity. Although this study provided some insight into the potential for alteration of CYP3A4-mediated metabolism by TCZ, DEX is not considered a sensitive CYP3A4 substrate. Therefore a subsequent DDDI study was planned with simvastatin to fully characterize the effect of TCZ on CYP3A4 activity in RA patients.

**TABLE 11-2 Mean Omeprazole (OME) Pharmacokinetics Parameters and Ratio of Geometric Least Squares Mean[a]**

| | OME (PM) | | OME (IM) | | OME (EM) | |
|---|---|---|---|---|---|---|
| | OME Day 1 | OME + TCZ Day 15 | OME Day 1 | OME + TCZ Day 15 | OME Day 1 | OME + TCZ Day 15 |
| $AUC_{0-\infty}$ (ng·h/mL) | 2723 (74) | 1904 (56) | 1923 (65) | 1207 (79) | 1099 (110) | 789 (124) |
| AUC ratio | – | 0.88 [0.78–0.99] | – | 0.88 [0.77–1] | – | 0.72 [0.66–0.79] |
| $C_{max}$ (ng/mL) | 612 (36) | 564 (30) | 586 (46) | 555 (58) | 417 (52) | 212 (97) |
| $C_{max}$ ratio | – | 0.98 [0.83–1.15] | – | 0.94 [0.82–1.08] | – | 0.69 [0.55–0.87] |

[a]Coefficient of variation ([standard deviation/mean]100%) is in parentheses; 90% confidence interval is in brackets.
$AUC_{0-\infty}$: area under plasma concentration-time curve from zero to infinity; $C_{max}$: maximum plasma concentration; EM: extensive metabolizer; IM: intermediate metabolizer; PM: poor metabolizer; TCZ: tocilizumab.

**TABLE 11-3 Mean Dextromethorphan (DEX) and Dextrorphan Pharmacokinetics Parameters and Ratio of Geometric Least Squares Mean[a]**

| | DEX PK | | Dextrorphan PK | |
|---|---|---|---|---|
| | DEX Day 1 | DEX + TCZ Day 15 | DEX Day 1 | DEX + TCZ Day 15 |
| $AUC_{0-\infty}$ (ng·h/mL) | 21.9 (136) | 21.3 (126) | 30.2 (55) | 20.8 (45) |
| AUC ratio | – | 0.95 [0.82–1.1] | – | 0.71 [0.59–0.84] |
| $C_{max}$ (ng/mL) | 2.76 (115) | 2.17 (109) | 7.45 (58) | 4.36 (46.7) |
| $C_{max}$ ratio | – | 0.79 [0.66–0.96] | – | 0.63 [0.53–0.75] |

[a]Coefficient of variation ([standard deviation/mean]100%) is in parentheses; 90% confidence interval is in brackets.
$AUC_{0-\infty}$: area under plasma concentration-time curve from zero to infinity; $C_{max}$: maximum plasma concentration; TCZ: tocilizumab.

### 11.3.3 Simvastatin DDDI Evaluation

To further elucidate and characterize the effect of IL-6 blockade with TCZ on the metabolic activity of CYP3A4, a second DDDI study was undertaken to examine the impact of TCZ on the pharmacokinetics of SIM, which is considered a sensitive substrate of CYP3A4. SIM is almost exclusively metabolized by CYP3A4 and is recommended as a sensitive probe substrate by the FDA.[31] This study was performed in 12 RA patients and he study design is illustrated in Figure 11-1.

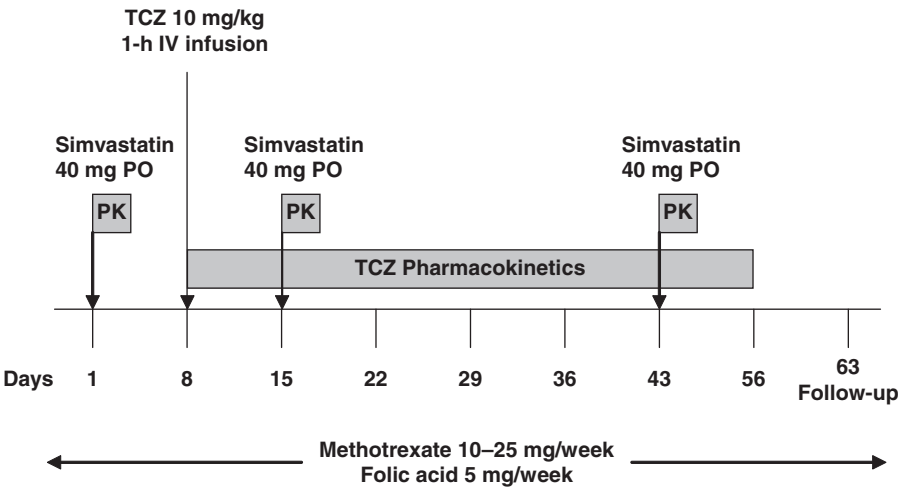


**FIGURE 11-1** Design of simvastatin disease–drug–drug interaction study. PK: pharmacokinetics; TCZ: tocilizumab. PO: per oral. IV: intravenous.

Briefly, a single 40 mg dose of SIM was administered alone, followed by a single 10 mg/kg infusion of TCZ after a 7-day washout period (day 8). In this study a 10-mg/kg dose of TCZ was used to represent steady-state exposure of the 8 mg/kg every 4 weeks regimen. One week (day 15) after the TCZ infusion, another single 40-mg dose of SIM was administered to evaluate the peak effect of TCZ on CYP3A4 activity. In this study, the time course of the potential effect on CYP3A4 activity was also evaluated by administration of an additional 40-mg dose of SIM 5 weeks after the TCZ infusion (day 43), to determine if the effect on CYP3A persisted. Blood samples were collected serially for pharmacokinetic assessment of SIM and its major metabolite, simvastatin acid, which is also a CYP3A4 substrate.

The study included male and female RA patients aged 18 to 75 years. Inclusion and exclusion criteria for the study were similar to the previous drug–drug interaction study.[26,27]

Mean $AUC_{0\text{–last}}$ and $C_{max}$ for SIM and simvastatin acid are provided in Table 11-4. The SIM $AUC_{0\text{–last}}$ in RA patients in this study was 102 ng·h/mL compared to values ranging from 11 to 25 ng·h/mL in studies in healthy subjects reported in the literature.[32–35] One week after administration of TCZ the $AUC_{0\text{–last}}$ and $C_{max}$ of SIM were significantly reduced by approximately 57%. This effect persisted even 5 weeks after the TCZ infusion, with an approximate reduction from baseline in $AUC_{0\text{–last}}$ of 40%. The findings were similar for simvastatin acid. Plasma concentrations were higher in RA patients at baseline as compared to literature reports in healthy adults ($AUC_{0\text{–last}}$ of 35 ng·h/mL in RA patients vs. 18 to 26 ng·h/mL in healthy adults). The $AUC_{0\text{–last}}$ of simvastatin acid was reduced by approximately 40% and 20%, 1 week and 5 weeks after TCZ administration, respectively. The effect of TCZ on CYP3A4 activity 5 weeks post dosing was less than 1 week post dosing because SIM $AUC_{0\text{–last}}$ and $C_{max}$ were higher on day 43 than on day 15 (Table 11-4).

These findings are consistent with the pattern of IL-6 receptor blockade by TCZ as evidenced by the time course of CRP (Figure 11-2). The maximum effect of TCZ in normalizing CRP levels was observed 1 week after infusion and corresponded to the maximum decrease in SIM levels. At 5 weeks after the infusion, serum concentrations of CRP represented only one-third of the baseline level, and the effect on SIM exposure was less pronounced.

### 11.3.4 Methotrexate DDDI Evaluation

The effect of TCZ on MTX PK was evaluated as a second part of the SIM DDDI study. Patients receiving MTX 10–25 mg/week as part of their RA therapy were enrolled in the study. All patients had serial PK samples collected for MTX, 1 week before TCZ administration and then again after the TCZ infusion, at similar times as for SIM. Results of the clinical DDDI evaluation confirmed that no significant changes were observed in MTX exposure in the presence of TCZ with geometric mean ratio of $C_{max}$ and $AUC_{0'\text{last}}$ for MTX with and without TCZ coadministration of 1.0 (0.92–1.09) and 1.10 (1.03–1.18), respectively.[29] Despite that IL-6 has been reported *in vitro* to suppress mRNA to downregulate the levels of MRP2,[28] which

**TABLE 11-4 Mean Simvastatin (SIM) and Simvastatin Acid Pharmacokinetics Parameters and Ratio of Geometric Least Squares Mean**

| | Simvastatin | | | Simvastatin Acid | | |
|---|---|---|---|---|---|---|
| | SIM Day 1 | SIM + TCZ Day 15 | SIM + TCZ Day 43 | SIM Day 1 | SIM + TCZ Day 15 | SIM + TCZ Day 43 |
| $AUC_{0-last}$ (ng·h/mL) | 102 (43) | 42.9 (41) | 63.2 (54) | 35.1 (72) | 22.4 (75) | 28 (63) |
| AUC ratio | – | 0.43 [0.34–0.55] | 0.6 [0.46–0.76] | – | 0.61 [0.49–0.76] | 0.8 [0.64–1] |
| $C_{max}$ (ng/mL) | 36.1 (60) | 14.0 (45) | 22.4 (82) | 5.37 (67) | 3.45 (81) | 3.7 (58) |
| $C_{max}$ ratio | – | 0.43 [0.33–0.55] | 0.61 [0.47–0.78] | – | 0.59 [0.47–0.74] | 0.7 [0.56–0.88] |

[a]Coefficient of variation ([standard deviation/mean]100%) is in parentheses; 90% confidence interval is in brackets.
$AUC_{0-last}$: area under plasma concentration-time curve from zero to the last measurable concentration; $C_{max}$: maximum plasma concentration; TCZ: tocilizumab.

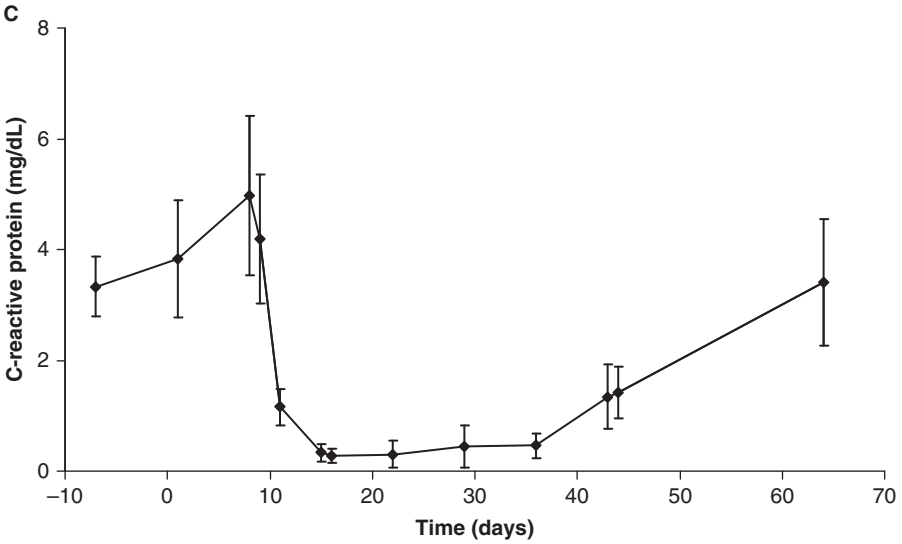


**FIGURE 11-2** Mean (±SE) C-reactive protein concentration vs. time.

is one of the main efflux transporters involved in the MTX elimination,[36] the absence of DDDI may suggest that elevated levels of IL-6, which as present in RA patients, have no effect on the expression and/or activity of MRP2 *in vivo* (expression of MRP2 was not measured in this study). Alternatively, other transporters might have compensated for MRP2 action.

### 11.3.5 Additional Drug–Drug Interaction Evaluations

The potential for drug–drug interactions with common agents coadministered for the treatment of RA was evaluated using population PK analysis with concentration-time data collected from RA patients enrolled in four Phase III clinical trials.[12] In this analysis, the concomitant use of RA-specific treatments–including MTX, chloroquine and derivatives, azathioprine, leflunomide, glucocorticoids, folic acid and derivatives, nonsteroidal anti-inflammatory agents, cyclooxygenase inhibitors, and analgesics (paracetamol, tramadol, codeine, and derivatives)—did not alter the PK of TCZ.

## 11.4 LABELING

The findings of these clinical DDDI evaluations have been incorporated into TCZ prescribing information globally to provide important guidance to clinicians caring for RA patients and managing comorbidities that require additional medications. As an example, we provide an excerpt from the US prescribing information:[2]

Interactions with CYP450 Substrates

Cytochrome P450s in the liver are down-regulated by infection and inflammation stimuli including cytokines such as IL-6. Inhibition of IL-6 signaling in RA patients treated with tocilizumab may restore CYP450 activities to higher levels than those in the absence of tocilizumab leading to increased metabolism of drugs that are CYP450 substrates. In vitro studies showed that tocilizumab has the potential to affect expression of multiple CYP enzymes including CYP1A2, CYP2B6, CYP2C9, CYP2C19, CYP2D6 and CYP3A4. Its effects on CYP2C8 or transporters is unknown. In vivo studies with omeprazole, metabolized by CYP2C19 and CYP3A4, and simvastatin, metabolized by CYP3A4, showed up to a 28% and 57% decrease in exposure one week following a single dose of ACTEMRA, respectively. The effect of tocilizumab on CYP enzymes may be clinically relevant for CYP450 substrates with narrow therapeutic index, where the dose is individually adjusted. Upon initiation or discontinuation of ACTEMRA, in patients being treated with these types of medicinal products, therapeutic monitoring of effect (e.g., warfarin) or drug concentration (e.g., cyclosporine or theophylline) should be performed and the individual dose of the medicinal product adjusted as needed. Prescribers should exercise caution when ACTEMRA is coadministered with CYP3A4 substrate drugs where decrease in effectiveness is undesirable, e.g., oral contraceptives, lovastatin, atorvastatin, etc. The effect of tocilizumab on CYP450 enzyme activity may persist for several weeks after stopping therapy.

## 11.5 DISCUSSION

The results of the clinical DDDI evaluations performed with TCZ confirm the longstanding hypothesis that clinically relevant drug interactions can occur as a result of reversal of an inflammatory process that suppresses CYP metabolic enzyme activity, hence "normalizing" activity. However, unlike direct drug–drug interactions, disease–drug–drug interactions are relevant only for CYP substrates with a narrow therapeutic index for which doses are individually adjusted. For those drugs, monitoring of drug effects or concentrations should be performed within 2 weeks upon initiation of TCZ therapy or several weeks after stopping the therapy. From these evaluations, CYP3A4 appears to be the most sensitive isozyme, based on *in vitro* and *in vivo* findings.

When considering the DDDI observed with simvastatin, $AUC_{0-\infty}$ was approximately sixfold higher in RA patients at baseline compared to healthy subjects (105 vs. 17.3 ng·h/mL). This observation confirmed that CYP3A4 activity is suppressed in RA patients by RA-associated inflammation. Upon initiation of TCZ (1 week postinfusion), SIM $AUC_{0-\infty}$ was reduced by approximately twofold (105 vs. 44.9 ng·h/mL), indicating that suppressive effect of inflammation on CYP3A4 was reversed and CYP3A4 activity was increased approaching "normalization." After stopping TCZ treatment (5 weeks postinfusion), SIM $AUC_{0-\infty}$ increased compared to the exposure at week 1 (65.2 vs 44.9 ng·h/mL), as the effect of the single-dose TCZ administration decreased with decreasing TCZ concentration. The clinical relevance of these changes can be revealed when compared to the effect of rifampin (a

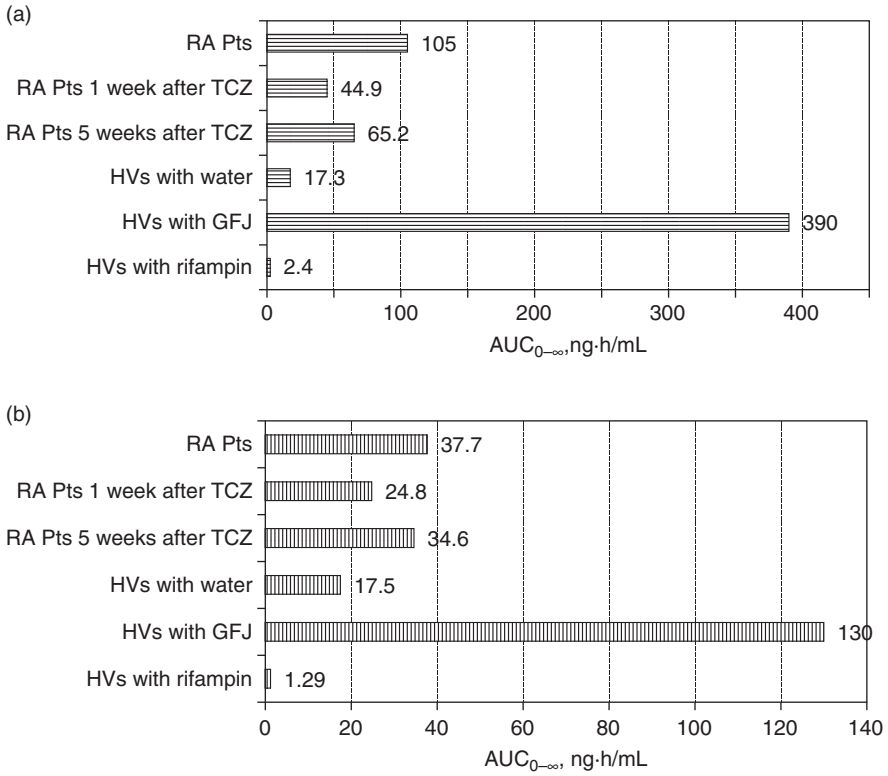


**FIGURE 11-3** Comparison of (a) simvastatin and (b) simvastatin acid metabolite. AUC change in rheumatoid arthritis patients (RA Pts) and in RA patients treated with tocilizumab (TCZ) rifampin, and grapefruit juice (GFJ) compared to healthy volunteers (HVs). Simvastatin was administered with water alone, 24 h after the last dose of 200 mL GFJ (3 times a day for 3 days) and 5 days after 600 mg rifampin daily treatment.

typical CYP3A4 inducer[34] and grapefruit juice (GFJ; a CYP3A4 inhibitor).[37] SIM $AUC_{0-\infty}$ was decreased by 87% after 5 daily doses of 600 mg rifampin,[34] and SIM $AUC_{0-\infty}$ was increased by ninefold after 3 days dosing of 200 mL GFJ three times daily.[37] The decrease in SIM $AUC_{0-\infty}$ upon initiation of TCZ (1 week postdosing) was much less than the change by rifampin (Figure 11-3a) and the lincrease in SIM $AUC_{0-\infty}$ increase after stopping of TCZ treatment (5 weeks postinfusion) was also much less than the change by GFJ (Figure 11-3b). A similar conclusion can also be drawn for the $AUC_{0-\infty}$ of SIM acid metabolite (Figure 11-3b). Interpretation of the findings of the TCZ DDDI assessments in the context of the effects of other well-known DDI perpetrators supports the recommendation that "the effect of tocilizumab on CYP enzymes may be clinically relevant for CYP450 substrates with narrow therapeutic index, where the dose is individually adjusted."[2]

## REFERENCES


1. Nishimoto N, Kishimoto T. Humanized antihuman IL-6 receptor antibody, tocilizumab. In: Chernajovsky Y, Nissim A, ed. Therapeutic antibodies. Handbook of experimental pharmacology 181. Berlin, Springer, 2008: 151–60.
2. Genentech. Package insert for Actemra. South San Francisco, CA, USA, 2012.
3. Hoffmann-La Roche. Package insert for Actemra. Mississauga, Ontario, Canada, 2010.
4. Roche Products. Package insert for Actemra. Dee Why, Australia, 2010.
5. RoActemra. Package insert for Actemra. Welwyn, Garden City, UK, 2010.
6. Chugai Pharmaceutical. Package insert for Actemra. Tokyo, Japan, 2008.
7. Hushaw LL, Sawaqed R, Sweis G, et al. Critical appraisal of tocilizumab in the treatment of moderate to severe rheumatoid arthritis. *Ther Clin Risk Manag* **6**, 143–52 (2010).
8. Okuda Y. Review of tocilizumab in the treatment of rheumatoid arthritis. *Biologics* **2**(1), 75–82 (2008).
9. Matsumoto T, Tsurumoto T, Shindo H. Interleukin-6 levels in synovial fluids of patients with rheumatoid arthritis correlated with the infiltration of inflammatory cells in synovial membrane. *Rheumatol. Int* **26**(12), 1096–100 (2006).
10. Milman N, Karsh J, Booth RA. Correlation of a multi-cytokine panel with clinical disease activity in patients with rheumatoid arthritis. *Clin Biochem* **43**(16–17), 1309–14 (2010).
11. Navarro-Millán I, Singh JA. Systematic review of tocilizumab for rheumatoid arthritis: a new biologic agent targeting the interleukin-6 receptor. *Clin Thera* **34**(4), 788–802 (2012).
12. Frey N, Grange S, Woodworth T. Population pharmacokinetic analysis of tocilizumab in patients with rheumatoid arthritis. *J Clin Pharmacol* **50**(7), 754–66 (2010).
13. Zhang X, Peck R Clinical pharmacology of tocilizumab for the treatment of patients with rheumatoid arthritis. *Expert Rev Clin Pharmacol* **4**(5), 539–58 (2011).
14. Nishimoto N, Terao K, Mima T, et al. Mechanisms and pathologic significances in increase in serum interleukin-6 (IL-6) and soluble IL-6 receptor after administration of an anti-IL-6 receptor antibody, tocilizumab, in patients with rheumatoid arthritis and Castleman disease. *Blood* **112**(10), 3959–64 (2008).
15. Uchiyama Y, Yoshida H, Koike N, et al. Anti-IL-6 receptor antibody increases blood IL-6 level via the blockade of IL-6 clearance, but not via the induction of IL-6 production. *Int Immunopharmacol* **8**(11), 1595–601 (2008).
16. Morgan ET. Impact of infectious and inflammatory disease on cytochrome P450-mediated drug metabolism and pharmacokinetics. *Clin Pharmacol Ther* **85**(4), 434–38 (2009).
17. Renton KW. Regulation of drug metabolism and disposition during inflammation and infection. *Expert Opin Drug Metab Toxicol* **1**(4), 629–40 (2005).
18. Morgan ET, Goralski KB, Piquette-Miller M, et al. Regulation of drug-metabolizing enzymes and transporters in infection, inflammation, and cancer. *Drug Metab Dispos* **36**(2), 205–16 (2008).
19. Williams SJ, Farrell GC. Inhibition of antipyrine metabolism by interferon. *Br J Clin Pharmacol* **22**, 610–12 (1986).

20. Williams SJ, Baird-Lambert JA, Farrell GC. Inhibition of theophylline metabolism by interferon. *Lancet* **8565**, 939–41 (1987).
21. Islam M, Frye RF, Richards TJ, et al. Differential effect of IFNa- 2b on the cytochrome P450 enzyme system: a potential basis of IFN toxicity and its modulation by other drugs. *Clin Cancer Res* **8**, 2480–87 (2002).
22. Jover R, Bort R, Gómez-Lechón MJ, Castell JV. Down-regulation of human CYP3A4 by the inflammatory signal interleukin-6: molecular mechanism and transcription factors involved. *FASEB J* **16**, 1799–01 (2002).
23. Warren GW, van Ess PJ, Watson AM, et al. Cytochrome P450 and antioxidant activity in interleukin-6 knockout mice after induction of the acute phase response. *J Interferon Cytokine Res* **21**, 821–26 (2001).
24. Ward LD, Howlett GJ, Hammacher A, et al. Use of a biosensor with surface plasma response detection for the determination of binding constants: measurement of interleukin-6 binding to the soluble interleukin-receptor. *Biochemistry* **34**, 2901–07 (1995).
25. Robak T, Gladalska A, Stepien H, Robak E. Serum levels of interleukin-6 type cytokines and soluble interleukin-6 receptor in patients with rheumatoid arthritis. *Mediators Inflamm* **7**, 347–53 (1998).
26. Zhang X, Schmitt C, Grange S, et al. Disease-drug interaction studies of tocilizumab with cytochrome P450 substrates in vitro and in vivo. *Clin Pharmacol Ther* **85**, S59 (2009).
27. Schmitt C, Kuhn B, Zhang X, et al. Disease-drug-drug interaction involving tocilizumab and simvastatin in patients with rheumatoid arthritis. *Clin Pharmacol Ther* **89**(5), 735–40 (2011).
28. Hartmann G, Cheung AK, Piquette-Miller M. Inflammatory cytokines, but not bile acids, regulate expression of murine hepatic anion transporters in endotoxemia. *J Pharmacol Exp Ther* **303**(1), 273–81 (2002).
29. Schmitt C, Kuhn B, Banken L, et al. Tocilizumab has no relevant effect on methotrexate pharmacokinetics in patients with rheumatoid arthritis. *International Journal of Clinical Pharmacology and Therapeutics.* **50**(3), 218–23 (2012).
30. Arnett F, Edworthy SM, Bloch DA, et al. The American Rheumatism Association 1987 revised criteria for the classification of rheumatoid arthritis. *Arthritis Rheum* **31**(3), 315–24 (1988).
31. US Food and Drug Administration, US Department of Health and Human Services, Center for Drug Evaluation and Research, Center for Biologics Evaluation and Research (CBER). Draft guidance for industry: drug interaction studies—study design, data analysis, implications for dosing, and labeling. Silver Spring, MD, 2012. Available at www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM292362.pdf.
32. Najib NM, Idkaidek N, Adel A, et al. Pharmacokinetics and bioequivalence evaluation of two simvastatin 40 mg tablets (Simvast and Zocor) in healthy human volunteers. *Biopharm Drug Dispos* **24**, 183–89 (2003).
33. Lilja JJ, Neuvonen M, Neuvonen PJ. Effects of regular consumption of grapefruit juice on the pharmacokinetics of simvastatin. *Br J Clin Pharmacol* **58**, 56–60 (2004).
34. Kyrklund C, Backman JT, Kivistö KT, et al. Rifampin greatly reduces plasma simvastatin and simvastatin acid concentrations. *Clin Pharmacol Ther* **68**, 592–97 (2000).

35. Kantola T, Kivistö KT, Neuvonen PJ. Erythromycin and verapamil considerably increase serum simvastatin and simvastatin acid concentrations. *Clin Pharmacol Ther* **64**, 177–82 (1998).

36. Elsby R, Fox L, Stresser D, et al. In vitro risk assessment of AZD9056 perpetrating a transporter-mediated drug-drug interaction with methotrexate. *Eur J Pharm Sci* **43**(1–2), 41–9 (2011).

37. Lilja JJ, Kivistö KT, Neuvonen PJ. Duration of effect of grapefruit juice on the pharmacokinetics of the CYP3A4 substrate simvastatin. *Clin Pharmacol Ther* **68**(4), 384–90 (2000).

# DRUG–DRUG INTERACTIONS FOR ETANERCEPT—A FUSION PROTEIN

JOAN KORTH-BRADLEY

The treatment of immune-mediated inflammatory diseases, such as rheumatoid arthritis, psoriasis, and psoriatic arthritis has been changed with the introduction of biological agents in the 1990s. Previously available biological agents, such as insulin, were replacements for endogenous proteins that were missing in patients with diabetes, while newer agents such as etanercept were constructs not found in nature. In this case study of drug–drug interaction assessments for etanercept, a fusion protein targeting tumor necrosis factor α (TNF-α) used in the treatment of several immune-mediated inflammatory diseases, background information about etanercept will be discussed, followed by a general assessment on the likelihood of drug interactions with fusion proteins such as etanercept and finally the results of specific etanercept–drug interaction investigations will be discussed.

## 12.1 ETANERCEPT BACKGROUND

Etanercept is a human tumor necrosis factor receptor (TNFR) p75 Fc fusion protein produced by recombinant DNA technology in a Chinese hamster ovary (CHO) mammalian cell expression system. It is a dimer of a chimeric protein genetically engineered by fusing the extracellular ligand binding domain of human tumor necrosis factor receptor 2 (TNFR2/p75) to the Fc domain of human immunoglobulin G1 (IgG1). The Fc component of etanercept contains the hinge, CH2 and CH3 regions, but not the CH1 region of IgG1. Etanercept has a molecular weight of approximately 150 kDa. It causes immune suppression by binding tumor necrosis factor, which is increased in patients with inflammatory diseases.[1]

Initially approved for the treatment of rheumatoid arthritis (RA), etanercept has subsequently received approval for the treatment of juvenile rheumatoid arthritis

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*,
Edited by Shayne C. Gad
Copyright © 2013 John Wiley & Sons, Inc.

(JRA), psoriatic arthritis (PsA), psoriasis, and ankylosing spondylitis (AS). It was also investigated for the treatment of congestive heart failure (CHF), without success. Initially, the approved dose was 25 mg administered by subcutaneous injection twice weekly. The dose for children was 0.4 mg/kg twice weekly. Subsequently, 50 mg administered weekly was also approved, as was 0.8 mg/kg administered once weekly in children.

After subcutaneous injection, etanercept is slowly absorbed into systemic circulation, with maximal concentration reached in 2 to 3 days. Dose proportionality has not been formally evaluated, but saturation of clearance is not apparent over the dosing range studied, and systemic exposure appears to be linear. Bioavailability after subcutaneous administration is 58%.[2] Bioequivalence of absorption was shown after administration of either liquid or reconstituted lyophilized etanercept to healthy volunteers.[3] The volume of distribution of etanercept is small. After administration of single subcutaneous etanercept doses in healthy subjects, the apparent volume of distribution was $12 \pm 6$ L.[4] In 3 subjects with RA, repeated samples of synovial fluid showed etanercept concentrations were similar to those observed in serum. Etanercept is slowly cleared from the body with a terminal half-life of 3 to 4 days. It is assumed that after binding of etanercept to TNF, the etanercept–TNF complex is then metabolized through peptide and amino acid pathways with either recycling of amino acids or elimination in bile and urine.[4] There was elimination of radioactivity observed in urine after administration of radiolabeled etanercept to patients and healthy subjects,[5] but almost no intact etanercept was observed in the collected urine.[6]

The potential impact of various demographic and disease characteristics on the pharmacokinetics of etanercept has been evaluated. A study of 30 healthy Japanese volunteers who received single subcutaneous 10-, 25-, or 50-mg doses of etanercept showed comparable pharmacokinetics with a similar group of healthy American volunteers.[6] A small pharmacokinetic study ($n = 6$) was performed in subjects with end stage renal disease requiring dialysis. The pharmacokinetics in these subjects was similar to subjects with psoriasis not requiring dialysis.[7] The pharmacokinetic profile of etanercept in subjects with RA showed parameters similar to those observed in healthy subjects.[8] Pharmacokinetics in subjects with AS were similar to those observed in subjects with RA.[9] The pharmacokinetics of etanercept were studied in 11 subjects with chronic CHF after repeated etanercept doses of 12 mg/m$^2$.[10] The pharmacokinetic results in subjects with CHF were similar to those observed in healthy subjects and in subjects with RA.

Several population pharmacokinetic analyses have been performed to further characterize etanercept pharmacokinetics. Lee and co-workers fit a one-compartment model to concentration data collected from 77 subjects who had received either 25- or 50-mg doses twice weekly for 24 weeks.[11] Significant covariates for apparent clearance were sex and race (white vs. nonwhite). White women had mean (95% confidence intervals) clearance estimates of 0.117 (0.108–0.130) L/h, and white men had clearance estimates of 0.138 (0.118–0.163) L/h; nonwhite patients had mean clearances that were 138% (104–160%) compared with white patients. Zhou and co-workers fit a two-compartment model to

concentration data collected from 10 studies, including 265 healthy subjects and patients with JRA and RA.[9] One of the studies had data after IV administration. The final model showed that age (<17 years) and body weight (<60 kg) were significant covariates for apparent clearance, which was lower compared with other subjects. The clearance in children (4 to 17 years of age) was lower than clearance in adults; however, when clearance values were adjusted for body weight, clearance was comparable with that of adults. A modest effect of the presence of RA on absorption was observed, and nonwhite individuals had larger central volumes of distribution, which disappeared if weight normalized. Finally, Nestorov's group fit a one-compartment model to data collected from three studies in 1077 patients with psoriasis.[12] Gender, weight, and time were significant covariates for apparent clearance and body weight for apparent volume of distribution. In contrast to what was observed in the previous analysis in patients with RA, in this analysis, the population mean (95% confidence interval) for clearance in women (0.148 L/h, 0.137–0.161 L/h) was higher than in men (0.129 L/h, 0.123–0.136 L/h). The finding of lower clearance with lower body weight may counteract the discrepancy. Apparent clearance increased over the first 4 weeks. With the exception of dosing for children <60 kg, none of the covariates identified is considered clinically significant or needing specific dosing.

## 12.2 MECHANISMS OF DRUG INTERACTIONS

Formal drug interaction studies were not performed during the initial clinical development of etanercept. It was believed that etanercept has little potential for causing drug interactions, or being influenced by administration of either inducers or inhibitors of important metabolic enzymes for other medications and thus drug interaction studies would not provide information about the metabolic fate of etanercept.

Parenteral administration excludes potential for absorption interactions with either food or other concomitantly administered medications. The volume of the injection is 1 mL when etanercept is administered to adults. It is unlikely that patients would want to combine other medications for simultaneous injection. Such combinations would not be recommended owing to potential physicochemical problems as well as simply too much volume.

Etanercept is a biological product. Methods to assess the extent of protein binding for small, synthetic molecules are not capable of adequately assessing protein binding for large molecules that are proteins themselves. It is believed there is relatively low potential for displacement reactions between etanercept and other medications that are highly protein bound.

As noted previously, it is assumed that after binding of etanercept to TNF, the complex is then metabolized through peptide and amino acid pathways with either recycling of amino acids or elimination in bile and urine. The cytochrome P450 system (CYP) and other commonly used metabolic pathways are not involved in the metabolism of etanercept, so there is little potential for metabolic drug interactions. The role of transporters in etanercept pharmacokinetics has not been studied.

Finally, etanercept is not believed to change biliary or urinary production and flow, so interactions associated with elimination are also not likely to occur. Information about metabolism and excretion, and thus potential for interaction, is obtained by studying the pharmacokinetics of drug in patients with isolated end organ dysfunction. In the case of etanercept, no evidence was obtained to suggest that either renal or hepatic failure alters elimination of the drug. A study of patients with severe sepsis received a single intravenous dose of etanercept to test whether etanercept had activity in this condition.[13] Serial serum concentrations were measured and, although not reported in the paper, no difference in concentrations in different subjects with different degrees of acute hepatic impairment or acute renal impairment.[5] As noted earlier, a study of subjects with chronic renal failure showed no difference in pharmacokinetics.

In the approximately 15 years since the development of etanercept, many more therapeutic proteins have been studied and guidelines for evaluation of the potential for drug interactions have been developed.[14] In addition to the various mechanisms by which drug interactions may occur already discussed, the importance of cytokine induction of important CYP enzymes has become better understood. Etanercept is not a cytokine and so again would not be expected to directly alter CYP activity. A potential indirect effect could be caused through inhibition of TNF-α by etanercept, but there have been no reports in the literature of interactions of this type. In conclusion, etanercept has a low likelihood of causing a drug interaction or being affected by other medications.

## 12.3 PHARMACODYNAMIC DRUG INTERACTIONS

Although etanercept may have a low potential for interactions from a pharmacokinetic perspective, pharmacodynamic interactions may occur as a result of its activity. It is always reasonable to evaluate the potential for interactions among medications that are commonly administered for the same indications. Rather than performing pharmacokinetic studies, researchers can collect safety and efficacy data as well as pharmacokinetic data in Phase II and III trials; the results are reviewed carefully for evidence of interactions.[15] This type of investigation was the basis for the recommendations that etanercept not be used concomitantly with anakinra,[16] abataept,[17] cyclophosphamide,[18] or sulfasalazine.[19] In each case, the patients receiving combination therapy had more serious adverse reactions, usually infections, than those subjects in the control group without benefit of improved efficacy.

In addition to too much immunosupression, etanercept may interfere with the desired response to vaccines. Patients exposed to varicella virus should temporarily discontinue etanercept and consider prophylactic treatment with varicella zoster immune globulin.[20] A study of patients with PsA showed that patients treated with etanercept or placebo had similar responses to pneumococcal vaccine.[21] Approximately 20% of patients failed to show a twofold response to any of the antigens. It was concluded that there was no increased risk of poor response with etanercept treatment given alone or with methotrexate.

## 12.4 RESULTS OF DRUG INTERACTION STUDIES WITH ETANERCEPT

Formal drug interaction studies have been performed to evaluate the interactions of etanercept with methotrexate, digoxin, and warfarin. Etanercept concentrations were also measured in clinical trials for patients with RA who were also treated with anakinra.

There was no mutual pharmacokinetic interaction between etanercept and warfarin.[22] In a nonrandomized, three-period study, 12 healthy male subjects received a single oral 25-mg dose of warfarin after an overnight fast, followed by twice-weekly 25-mg subcutaneous doses of etanercept for seven doses. The last dose of etanercept was administered concurrently with a second dose of warfarin. Serial blood samples for plasma warfarin concentration measurement and international normalized ratio (INR) assessment were collected before and up to 144 h after dose administration. Serial blood samples for serum etanercept concentration measurement were collected before and up to 60 h after the sixth dose and 264 h after the seventh dose. Etanercept did not affect the pharmacokinetics and pharmacodynamics of warfarin. All ratios of maximum serum concentration ($C_{max}$) and area under the serum concentration versus time curve (AUC) for pharmacokinetics (R- and S-enantiomers of warfarin) and INR fell within the confidence interval of 0.8 to 1.25. Warfarin also did not cause a clinically significant alteration in the pharmacokinetics of etanercept. In conclusion, coadministration of etanercept and warfarin would not be expected to change the pharmacokinetics of either medication; therefore, no dosage adjustment is needed in cases in which warfarin and etanercept are coadministered.

There was also no effect of coadministration of etanercept on the pharmacokinetics of digoxin.[23] In a crossover, open-label, nonrandomized, three-period study, 12 healthy male subjects received loading oral doses of digoxin 0.5 mg every 12 h on day 1 and 0.25 mg every 12 h on day 2, followed by a daily maintenance dose of 0.25 mg for a total of 27 days. Etanercept was administered as a twice weekly 25-mg subcutaneous dose beginning on day 9 and continuing up to day 37 for a total of 9 doses. All ratios of $C_{max}$ and AUC for pharmacokinetics of digoxin fell within the confidence interval of 0.8 to 1.25. Although not considered clinically relevant, the mean $C_{max}$ and AUC of etanercept were 4.2% and 12.5% lower, respectively, when etanercept was given with digoxin than when administered alone. There were no clinically relevant changes in the electrocardiogram parameters, and adverse events did not increase when both drugs were combined. In conclusion, there is no clinically relevant interaction between etanercept and digoxin, and both drugs can be safely coadministered without the need for a dosage adjustment.

The results from the population pharmacokinetic analysis demonstrated that the coadministration of methotrexate had no significant effect on the pharmacokinetics of etanercept in subjects with RA.[24] Using the population pharmacokinetic model previously described, which included data from healthy subjects and patients with RA and JRA, nine Bayesian-derived individual pharmacokinetic parameters were

estimated for 100 individuals, 50 receiving methotrexate and etanercept and 50 receiving etanercept alone. The patients had been randomly selected from the TEMPO trial[25] and each had had a single serum measurement collected after 24 weeks of treatment. The predicted weight-normalized CL values were $0.071 \pm 0.013$ and $0.072 \pm 0.011$ for the etanercept-treated patients and for the combination treated patients, respectively. Thus no dose adjustment of etanercept dosage is needed when coadministered with methotrexate.

The study of combined treatment of etanercept and anakinra mentioned previously[16] also collected blood samples for the measurement of anakinra and etanercept concentrations in plasma before dosing and then at 4, 12 and 24 weeks after the start of treatment. Although there was large intersubject variability, there appeared to be no difference in either anakinra or etanercept concentrations between those on combination therapy and those receiving etanercept alone.

There are references in the literature to an interaction between paclitaxel and etanercept, where coadministration of paclitaxel causes an increase in etanercept concentrations of 1.5-fold,[14,26] but the specifics of the interaction are not well described in the literature or in the prescribing information for any of the commercially available forms of paclitaxel.[27,28]

## 12.5 CONCLUSIONS

In conclusion, much more is known now about the potential for drug interaction involving therapeutic proteins than when etanercept was being developed. However, a systematic review of potential causes for drug interactions and careful review of safety and efficacy of studies where combinations of medications are used has been successful in identifying pharmacodynamic interactions so that potentially unsafe combinations of treatments can be avoided, even when pharmacokinetic interactions are unlikely.

## ACKNOWLEDGMENTS AND CONFLICTS OF INTEREST

The author is an employee of Pfizer Inc. who comarkets Enbrel (etanercept) along with Amgen. The statements or opinions expressed in this paper are her own and do not necessarily represent those of Pfizer.

## REFERENCES

1. Alldred A. Etanercept in rheumatoid arthritis. *Expert Opin Pharmacother* **2**, 1137–48 (2001).
2. Lebsack ME, Hanna RK, Lange MA, et al. Absolute bioavailability of TNF receptor fusion protein following subcutaneous injection in healthy volunteers. *Pharmacotherapy* **17**, 1118–19 (1997).

3. Sullivan JT, Ni L, Sheelo C, Salfi M, Peloso PM. Bioequivalence of liquid and reconstituted etanercept subcutaneous injections. *J Clin Pharmacol* **46**, 654–61 (2006).
4. Korth-Bradley JM, Rubin ASR, Hanna RK, et al. The pharmacokinetics of etanercept in healthy volunteers. *Ann Pharmacother* **34**, 161–64 (2000).
5. European Medicinal Agency. Enbrel: EPAR - Scientific Discussion. London, 18 Oct 2006. Available at www.ema.europa.eu/docs/en_GB/document_library/EPAR_-_Scientific_Discussion/human/00262/WC500027358.pdf. Accessed 15 Mar 2013.
6. Kawai S, Sekino H, Yamashita N, et al. The comparability of etanercept pharmacokinetics in healthy Japanese and American subjects. *J Clin Pharmacol* **46**, 418–23 (2006).
7. Don BR, Spin G, Nestorov I, et al. The pharmacokinetics of etanercept in patients with end-stage renal disease on haemodialysis. *J Pharm Pharmacol* **57**, 1407–13 (2010).
8. Kremer JM, Spencer-Green GT, Hanna RK, et al. ENBREL (etanercept) pharmacokinetics in patients with rheumatoid arthritis. *Arthritis Rheum* **43**, S229 (2000).
9. Zhou H, Buckwalter M, Mayer P, et al. Population-based pharmacokinetics of the soluble TNFr etanercept: a clinical study in 43 patient with ankylosing spondylitis compared with post hoc data from patients with rheumatoid arthritis. *Int J Clin Pharmacol Ther* **42**, 267–76 (2004).
10. Soran O, Feldman AM, Schneider R, et al. The pharmacokinetics of etanercept in patients with heart failure. *Br J Clin Pharmacol* **51**, 197–99 (2001).
11. Lee H, Kimko HC, Rogge M, et al. Population pharmacokinetic and pharmacodynamic modeling of etanercept using logistic regression analysis. *Clin Pharm Ther* **73**, 348–65 (2003).
12. Nestorov I, Zitnik R, Ludden T. Population pharmacokinetic modeling of subcutaneously administered etanercept in patients with psoriasis. *J Pharmacokin Pharmacodyn* **31**, 463–90 (2004).
13. Fisher CJ, Agosti JM, Opal SM et al. Treatment of septic shock with the tumor necrosis factor receptor: Fc fusion protein. *N Engl J Med* **334**, 1697–02 (1996).
14. Huang S-M, Zhao H, Reynold SK et al. Therapeutic protein-drug interactions and implications for drug development. *Clin Pharm Ther* **87**, 497–503 (2010).
15. Van Roon EN, van den Bernt PMLA, Jansen TLThA, et al. An evidence-based assessment of the clinical significance of drug-drug interactions between disease-modifying antirheumatic drugs and non-antirheumatic drugs according to rheumatologists and pharmacists. *Clin Ther* **31**, 1737–46 (2009).
16. Genovese MC, Cohen S, Moreland L, et al. Combination therapy with etanercept and anakinra in the treatment of patients with rheumatoid arthritis who have been treated unsuccessfully with methotrexate. *Arthritis Rheum* **50**, 1412–19 (2004).
17. Weinblatt M, Schiff M, Goldman A, et al. Selective costimulation modulation using abatacept in patients with active rheumatoid arthritis while receiving etanercept: a randomized clinical trial. *Ann Rheum Dis* **66**, 228–34 (2007).
18. The Webener's Granulomatosis Etanercerpt Trial (WGET) Research Group. Etanercept plus standard therapy for Wegener's granulomatosis. *N Eng J Med* **352**, 351–61 (2005).
19. Combe B, Codreanu C, Fiocco U, et al. Efficacy, safety and patient-reported outcomes of combination etanercept and sulfasalazine versus etanercept alone in patients

with rheumatoid arthritis: a double-blind randomized 2-year study. *Ann Rheum Dis* **68**, 1146–52 (2009).

20. Amgen, Pfizer. Package insert for Enbrel for subcutaneous injection. Thousand Oaks, CA 2012.
21. Mease PJ, Ritchlin CT, Martin RW, et al. Pneumococcal vaccine response in psoriatic arthritis patients during treatment with etanercept. *J Rheumatol* **31**, 1356–61 (2004).
22. Zhou H, Patat A, Parks V, et al. Absence of a pharmacokinetic interactions between etanercept and warfarin. *J Clin Pharmacol* **44**, 543–50 (2004).
23. Zhou H, Parks V, Patat A, et al. Absence of a pharmacokinetic interactions between etanercept and digoxin. *J Clin Pharmacol* **44**, 1244–51 (2004).
24. Zhou H, Mayer PR, Wajdula J, Fatenejad S. Unaltered etanercept pharmacokinetics with concurrent methotrexate in patients with rheumatoid arthritis. *J Clin Pharmacol* **44**, 1235–43 (2004).
25. Klareskog L, van der Heijde D, de Jager JP, et al. Therapeutic effect of the combination of etanercept and methotrexate compared with each treatment alone in patients with rheumatoid arthritis: double blind, randomized, controlled trial. *Lancet* **363**, 675–81 (2004).
26. Mahood I, Green MD. Drug interaction studies of therapeutic proteins and monoclonal antibodies. *J Clin Pharmacol* **47**, 1540–54 (2007).
27. Celegene Corp. Package insert for Abraxane. Summit, NJ, 2012.
28. Bristol-Myers Squibb.Package insert for Taxol. Princeton, NJ, 2011.

# DRUG INTERACTIONS OF CYTOKINES AND ANTICYTOKINE THERAPEUTIC PROTEINS

J. GREG SLATTER, LARRY C. WIENKERS, and LESLIE J. DICKMANN

*If it were not for the great variability among individuals, medicine might as well be a science and not an art.*

Sir William Osler, 1892[1]

## 13.1 INTRODUCTION

The suppression of drug metabolism by proinflammatory cytokines and the relief of suppression by disease treatment are a current issue in the development of drug interaction labels for new biologic products. This topic has been the subject of a number of recent reviews that capture the history of related work as well as current regulatory and pharmaceutical industry thinking.[1–12] This chapter provides an overview of some recent data developed in industrially relevant preclinical models and outlines the pros and cons of these models with respect to the generation of actionable data–that is data that are robust enough to preclude the need for a clinical drug–disease interaction study, and/or a class label for a new cytokine or anticytokine therapeutic agent.

Because small molecule clearance mechanisms are suppressed by inflammation, proinflammatory cytokine treatment and/or active inflammatory disease may afford patient populations with impaired drug clearance.[6] As such, the effective treatment of inflammatory disease would be expected to normalize or "desuppress" drug clearance. When clearance is mediated by cytochromes P450 (CYP), suppression works in the same direction as CYP inhibition, leading to increased systemic exposure. In contrast, CYP desuppression increases CYP activity (similar to CYP induction), leading to lower drug concentrations. The caveat with desuppression is that the upper limit of the effect is the clearance

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*,
Edited by Shayne C. Gad
Copyright © 2013 John Wiley & Sons, Inc.

observed in healthy patients. The clinical literature that have accurately measured the magnitude of these effects are limited. Where clinical pharmacokinetic effects have been characterized, several drug labels appropriately note that small molecule clearance changes are possible when anti-inflammatory biologic agents or proinflammatory cytokines are coadministered with narrow therapeutic index small molecules that are cleared mainly by CYP enzymes (labels are reviewed in Ref. 5). Many pharma/biotech pharmacokinetics and drug metabolism labs are now discussing best practices for *in vitro* experiment design and interpretation, meaningful *in vitro* to *in vivo* correlations (IVIVC), and the design of suitable clinical investigations involving therapeutic proteins.[13,14]

## 13.2 CLINICAL RELEVANCE OF CYTOKINE-MEDIATED SUPPRESSION AND DESUPPRESSION OF ADME ENZYMES

When proinflammatory cytokine-mediated effects on CYPs are compared to induction and inhibition of CYPs by agents such as ketoconazole and rifampin, it is so far apparent that the clinical magnitude of these effects are relatively subtle. In many cases, small study populations have made it difficult to discern a clinically measurable effect on transcriptionally variable CYP isoenzymes. The magnitude of known clinical changes of therapeutic protein DDI, along with useful data on dosimetry and necessary mechanistic details are compiled in three useful tables published in Ref. 4.

### 13.2.1 Treatment with Interferons and Related Proinflammatory Agents

Clinical interest in the suppression of drug metabolism by inflammatory disease rose after the 1978 account of side effects in asthmatic children on theophylline during an influenza outbreak. Theophylline half-life increases averaged about twofold,[15] and this highlights that subtle disease-mediated changes in oral clearance can be clinically significant in narrow therapeutic index drugs. Shedlofsky et al.[16] showed that changes in theophylline and antipyrine clearance in healthy subjects could be elicited with lipopolysaccharide (LPS) endotoxin. Interferon-mediated suppression of CYP in influenza was proposed as a potential mechanism in a clinical study in healthy subjects that compared antipyrine clearance during influenza (78% of healthy value) and after administration of interferon A (85% of healthy value) or interferon B (85% of healthy value).[17]

Over the time period when disease-related CYP effects were first linked to interferons, these cytokines were also of interest as drug candidates. Type I interferons ($\alpha$2a, $\alpha$2b, $\beta$1a, $\beta$2b) are a key component of polytherapy of Hepatitis B and Hepatitis C viral infections (HBV and HCV), autoimmune diseases, and cancers, by virtue of their ability to elicit a host immune response. As a consequence, CYP suppression by interferons has been investigated largely clinically because combination therapy entails risk of higher exposure to concomitant, narrow therapeutic index CYP substrates.[18] *In vitro* studies and a variety of preclinical *in vivo* studies have also been undertaken to understand the mechanism of CYP suppression, relative to other

cytokines such as the interleukins. Flaman et al.[18] recently examined the mechanism of CYP3A4 suppression by interferon α in human hepatoma (HepG2) cells and demonstrated a time-dependent and physiologically relevant concentration-dependent decrease in CYP3A4 transcription at nontoxic concentrations of interferon α. They proposed that suppression was transcriptional and involved an interferon-mediated JAK1-STAT1 signaling cascade that is independent of ERK1/2 and p38 MAP kinases. Aitken and Morgan[19] investigated a type II interferon, interferon γ, and showed that it mediated isoform-specific effects on CYPs in human hepatocytes, with CYP1A2, 2A6, 2B6, and 3A4 activities being selectively suppressed in a nitric oxide–independent manner. Isoenzyme-specific upregulation and downregulation of CYPs by interferon α2b has also been observed at therapeutic concentrations in human hepatocyte coculture with Kupffer cells.[20] Kupffer cells are a contributor to cytokine-related signaling in the liver, and there is current interest in whether data from human hepatocyte/Kupffer cell cocultures are any more informative than hepatocytes alone, in terms of recapitulating *in vivo* acute phase response effects on CYPs.[20,21]

A variety of clinical studies on therapeutic interferons have revealed that probe substrate clearances were generally in the range of 85% to 40% relative to control values. Oral clearance decreases were generally less than twofold to threefold (50% to 33% of control values) and covered a variety of major CYP isoenzymes.[4] Accordingly, there are relatively few drug interactions listed in the product labels of interferons. Label warnings include use of interferon α2a and α2b with theophylline, the latter doubling theophylline levels. In 18 peginterferon α2b–naive chronic hepatitis C subjects on stable methadone maintenance therapy, mean methadone AUC increased by about 16% after 4 weeks of peginterferon α2b treatment and was about twofold higher in two subjects. In 22 HCV patients, peginterferon α2b caused a modest increase in clearance of a CYP2C8/9 probe substrate (mean increase 28%), a variable increase in clearance of a CYP2D6 probe substrate (mean increase 66% with 13 subjects increased, 5 subjects decreased, and 4 subject unchanged) and with no significant effect on CYP1A2 and CYP3A4 probe substrates.[5] With these counterintuitive increases in CYP substrate clearance, it is apparent that suppression of CYPs by the interferons used for HCV therapy is not a straightforward problem. Data from Nakai et al.[22] may explain the subtle effects and counterintuitive data because the authors have shown that suppression of drug-metabolizing enzymes increases with the progression of liver fibrosis in chronic HCV patients. Accordingly, when interferons are used successfully in the treatment of HCV infection, sustained virologic responses over time may reduce hepatic inflammation and therefore mitigate the suppressive effects of the proinflammatory cytokine treatments. This highlights the temporal efficacy-related aspects of this drug–disease interaction that make the quantitative *in vivo* extrapolation of *in vitro* data tenuous.

### 13.2.2 Treatment of Inflammatory Conditions through the Mitigation of Cytokine-Mediated Signaling

It is now well established that inflammation contributes to pathology in a wide range of diseases and that inflammation can also contribute to altered drug pharmacokinetics and pharmacodynamics.[11,23] For example, drug–disease interactions of

this sort have been described for HIV patients whose oral midazolam clearance was 50% (a twofold decrease) or 82% relative to that of healthy subjects.[24,25] Losartan oral clearance was unaffected by rheumatoid arthritis, but its active metabolite AUC was 40% of control in subjects with active arthritis.[26] In other examples, antipyrine clearance was impaired in sepsis,[27] and clozapine clearance was impaired in infection.[28,29]

Haas et al.[30] showed an inverse correlation of serum interleukin 6 (IL-6) with CYP3A activity in postsurgical patients, as measured by the Erythromycin Breath Test. Frye et al.[31] showed inverse correlations of CYP1A2- and CYP2C19-mediated probe substrate metabolism with serum IL-6 in patients with congestive heart failure, and Rivory et al.[32] showed a correlation of the acute phase marker, serum C-reactive protein (CRP), with the Erythromycin Breath Test in cancer patients. Advanced cancer patients with reduced clearance of docetaxel and elevated markers of inflammation experienced more severe adverse effects of chemotherapy.[33] Chen et al.[34] revealed a strong correlation between IL-6 concentrations and increased cyclosporine levels in six bone marrow transplant patients receiving a continuous cyclosporine infusion over the posttransplant interval. On average IL-6 increased to a maximum of 19.8-fold and CRP was elevated to a maximum of 8.8-fold. Cyclosporine levels increased by 3.60-fold at 4.83 days after the maximum IL-6 concentration. The Spearman rank order correlation $R^2$ for the correlation between $C_{\text{max IL-6}}$ and dose normalized $C_{\text{max CsA}}$ was 0.794. For the cyclosporine metabolite AM1, the correlation with $C_{\text{max IL-6}}$ was 0.975.[34] Taken together, these selected studies show how systemic manifestations of inflammation are correlated with CYP status and how CYP suppression could be a contributor to adverse events, especially during acute care polytherapy.

As the studies cited indicate, the magnitude of CYP effects can be mild. Accordingly, there are also studies that failed to reveal significant cytokine-mediated effects on CYPs, such as when arthritis treatment with sodium aurothiomalate, sulfasalazine or penicillamine had no significant effect on the CYP3A marker, urinary 6β-hydroxycortisol in rheumatoid arthritis patients.[35] Similarly, there was no difference in cyclosporine pharmacokinetics in elderly patients with and without arthritis.[36]

To date, the upper magnitude of inflammatory disease treatment effects on CYP-related drug metabolism appears to be in the range of onefold to fourfold, with most examples occurring at the low end of the range. Upper limits of CYP suppression could be better defined with adequately powered studies comparing probe substrate pharmacokinetics and inflammation biomarkers in patients with severe inflammatory disease, measured either against predisease status or age and sex matched healthy subjects. The Chen et al.[34] study on cyclosporine pharmacokinetics in transplant patients stands out as a good example of the value of this approach.

### 13.2.3 The Tocilizumab–Simvastatin Interaction

So far there is only one published clinical study investigating a drug–disease interaction involving cytokine desuppression by anticytokine therapy, the tocilizumab–

simvastatin interaction.[37] Rheumatoid arthritis patients were treated with the anti-IL6R monoclonal antibody, tocilizumab (Actemra™),[36] and simvastatin pharmacokinetics were measured before and after tocilizumab administration. Rheumatoid arthritis patients experienced an 2.3 fold and (43% of control) decrease in the plasma AUC of the CYP3A4 substrate simvastatin after an 10 mg/kg tocilizumab dose.[36,37] In comparison, in separate pharmacokinetic studies, the decreases in simvastatin AUC after treatment with the prototypical inducer rifampin were 7.2- and 11.1-fold.[38,39] The view that desuppression normalized but did not induce simvastatin oral clearance in tocilizumab-treated rheumatoid arthritis patients is supported by a simvastatin $AUC_{0-\infty}$ comparison across separate studies ($44.5 \pm 19$ ng·h/mL in tocilizumab-treated rheumatoid arthritis patients versus 29 ng·h/mL (range 8–56 ng∗h/mL) in healthy subjects).[37,39]

### 13.2.4 Can C-reactive Protein (CRP) and/or IL-6 Be Used to Predict the Clinical Magnitude of CYP Suppression and Desuppression?

Schmitt et al.[37] noted that the plasma concentration versus time profile of the acute phase reaction protein, serum CRP, "paralleled the time course of tocilizumab's effect on simvastatin pharmacokinetics." Production of CRP by hepatocytes and release into the systemic circulation is primarily elicited by IL-6 and to a lesser extent by IL-1β and tumor necrosis factor and is mediated by the transcription factors C/EBPβ and NF-κB.[40] This observation and the other independent correlations of CYP3A activity with CRP and IL-6 discussed earlier,[31,32] may indicate that initial and posttreatment CRP concentrations are a potential quantitative biomarker for overall hepatic inflammation of sufficient magnitude to suppress CYPs. Therefore, although it is widely acknowledged to be a variable inflammation biomarker, large changes in circulating CRP (and/or IL-6) may be inexpensively predictive of the clinical magnitude of suppression and/or desuppression of CYPs in adequately powered studies. Clinical data on other cytokine and anticytokine drugs will be needed to discern whether clinically measurable suppression and desuppression of CYP enzymes is possible in the absence of significant changes in CRP and IL-6. These data will be useful in understanding the relative magnitude of hypothetical "direct" effects of various individual cytokines on CYPs against a possible central role of transcriptional CYP suppression that occurs secondary to inflammation-related hepatocyte exposure to IL-6 and/or IL-1β.

## 13.3 MECHANISM

### 13.3.1 Cytokine-Mediated Suppression of CYPs

No single global mechanism has yet been identified that accounts for all inflammation-mediated changes for all drug-metabolizing enzymes. Ashino et al.[41] and others have conducted work related to the premise that blanket downregulation of pro-oxidant CYP oxidative pathways is hepatoprotective against immune-activation-related oxidative stress. Morgan[42] reviewed the complexities of possible adaptive

mechanisms for CYP suppression, including the relief of oxidative stress. In this insightful review, he posed some excellent "rational speculation" on why animals that experience inflammation might find it beneficial to suppress hepatic CYP enzymes. Transcriptional effects are the most investigated and may be coordinately regulated by a variety of different nuclear receptors and transcription factors[43] as they are in liver disease.[44] Because a number of cytokines mimic the effects of infection when injected *in vivo* or when incubated with hepatocytes in primary culture, downregulation of many CYP isoenzymes has been associated with elevated levels of a variety of cytokines.[19,42,45] Inflammation-related CYP suppression is mainly transcriptionally mediated, although posttranscriptional mechanisms have also been proposed.[46,47] Consequently, NF-κB-mediated regulation and related crosstalk with the pregnane X receptor (PXR, NR1I2) are often inferred, though an overall mechanism for suppression has not been fully elucidated.[48–52] Ghose et al.[53] proposed that suppression of many CYP isoenzymes may be mediated by inflammation-induced alteration of the subcellular location of retinoid X receptor α (RXRα), which is an obligate partner in heterodimeric nuclear receptors liver X receptor (LXR), farnesoid X receptor (FXR), retinoic acid receptor (RAR), constitutive androstane receptor (CAR), pregnane X receptor (PXR), and peroxisome proliferator activated receptor alpha (PPARα). Gu et al.[54] demonstrated direct interactions of NF-κB P65 with the DNA binding domain of retinoid X receptor alpha (RXRα). Taken together, these studies indicated that effects on RXRα are a potential explanation for cytokine-mediated suppressive effects that are observed across the spectrum of enzymes involved in steroid, lipid, bile acid, and drug metabolism.

A recent genomewide association study on CYP regulation profiled liver tissue from over 427 donors. The study revealed negative correlations between the expression of genes in a gene module related to the acute phase response (APR) and genes in an hepatic steroid and lipid metabolism gene module (see Figure 3 in Ref.[55]). Subtle downregulation of steroid and lipid metabolism, in consort with up-regulation of the APR related genes is apparently a core component of the hepatic response to systemic inflammation. This is illustrated in Figure13-1, for which supplemental data from Yang et al.[55] were used to plot pairwise correlations among mRNA of CYP3A4, PXR, IL-6, IL-1β, and the APR marker CRP. CYP3A4 expression in this population was correlated with many genes under the regulatory control of LXR, FXR, RAR, CAR, PXR, and PPARα. CYP3A4 was also weakly correlated with a variety of other hepatic acute phase markers associated with the APR.[55,56] It is attractive to speculate, based on these gene expression correlations, that hepatic inflammation is a covariate of basal CYP3A4 expression in human populations. If this were true, the dynamic range needed to discern these correlations in this population of liver donors may have been related to individual differences in disease state or suitability for transplant, hepatic inflammation related to the process of death and dying, and/or liver collection and preservation.

### 13.3.2 A Related Mechanism: Drug Binding to α1-Acid Glycoprotein

There is a second mechanism that alters small molecule pharmacokinetics in inflammatory diseases. Jamali and others have shown that drugs that bind

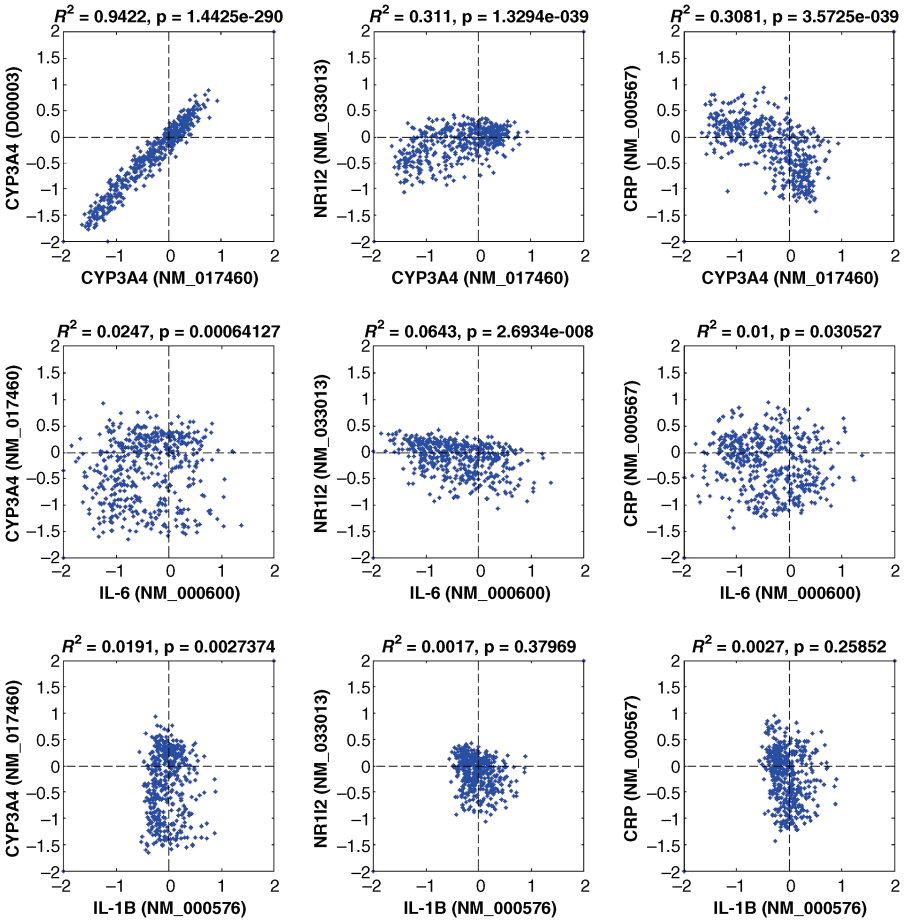


**FIGURE 13-1** Gene to gene mRNA expression correlations. CYP3A4, IL-6, and IL1β are correlated with a second CYP3A4 probe, the nuclear pregnane X receptor (PXR; gene name NR1I2) and the acute phase response marker, C-reactive protein (CRP). *Top row, CYP3A4 correlations*: (left) Two unique CYP3A4 probes were present on the array and show the expected high correlation between different probes for the same enzyme across the donor population; (middle) expected positive correlation between CYP3A4 and PXR (NR1I2) expression; (right) negative correlation between CYP3A4 and CRP. *Middle row, IL-6 correlations*: (left) Lack of correlation with CYP3A4; (middle) negative correlation between IL-6 and PXR (NR1i2) expression; (right) lack of correlation between IL-6 and CRP. *Bottom row IL-1β correlations*: (left) Dynamic range of IL-1β expression is narrower than IL-6 ($< \pm$ 10-fold), no correlation with CYP3A4; (middle) no correlation between IL-1β and PXR (NR1i2) expression; (right) no correlation between IL-1β and CRP. Plots prepared by Dr. Yudong He (Amgen Corp), from open access data at sagebase.org and data from Ref. 55.

extensively and exclusively to the acute phase reaction protein α1-acid glycoprotein (AAG) can exhibit inflammatory disease-dependent clearance.[11,23] A recent study on the kinase inhibitor vismodegib illustrates how modern biophysical techniques can be used to assess the role of AAG-binding-dependent pharmacokinetics in patient populations.[57] The AAG drug–disease interaction mainly occurs with basic oral drugs that have intermediate to high extraction ratios, such as verapamil and β-blockers.[58] Most AAG correlations with drug clearance were developed using *in vivo* experiments in rats or mice, for which treatment-mediated relief of experimental inflammation was accomplished with ketoprofen and infliximab.[59,60] In these cases, binding to elevated AAG may work in consort with inflammation-impaired CYP-mediated clearance to increase drug exposure. Accordingly, changes in oral clearance of AAG-binding drugs with rate-limiting CYP-mediated metabolism may be greater than in drugs that do not bind AAG. For example, rheumatoid arthritis resulted in a fourfold increase in total verapamil AUC, relative to control subjects.[58]

## 13.4 CAN PRECLINICAL MODELS BE USED TO PREDICT CLINICAL SUPPRESSION OR DESUPPRESSION?

### 13.4.1 Cytokine Effects in Hepatocyte Culture: Studies on IL-6

Human hepatocytes are an *in vitro* model of choice for investigating cytokine effects on CYPs.[21,19,61] IL-6 is the most studied cytokine to date and is known to decrease CYP mRNA and the activities of major human CYP isoenzymes.[19] For the interleukins, virtually all *in vitro* hepatocyte studies conducted before 2010 were conducted at cytokine concentrations that were much higher than human serum cytokine concentrations. For example, in healthy individuals, serum IL-6 concentrations range from 1.3 to 10.3 pg/mL and are increased to 2.6–123 pg/mL in inflammatory diseases.[61]

Our laboratory carried out an analysis of CYP regulation by IL-6 and desuppression of CYP by an anti-IL-6 monoclonal antibody.[61] IL-6 was chosen as a model cytokine to define industrially relevant pros and cons of *in vitro* experimentation on cytokine-mediated suppression of CYPs in primary human hepatocytes. IL-6 was also selected because well-designed clinical and related *in vitro* work became the foundation of the tocilizumab drug–disease–drug interaction label[37] and because of the multiple roles of IL-6 in liver physiology.[62] mRNA expression of diverse drug metabolizing enzymes and CYP1A2 and 3A4 activities were measured.

Cytokines and antibodies from commercial vendors required special characterization that is outside of the realm of standard small molecule *in vitro* drug metabolism protocols. For example: Experiments were benchmarked to physiologically relevant serum cytokine concentrations. Binding affinities of monoclonal antibodies for their targets were characterized and compared to the affinity of the cytokine for its receptor. Lastly, the expression of relevant cytokine receptors on hepatocytes, and the cytokine-mediated generation of acute phase proteins (serum amyloid A and/or CRP) were followed through the course of the experiment.[61,63]

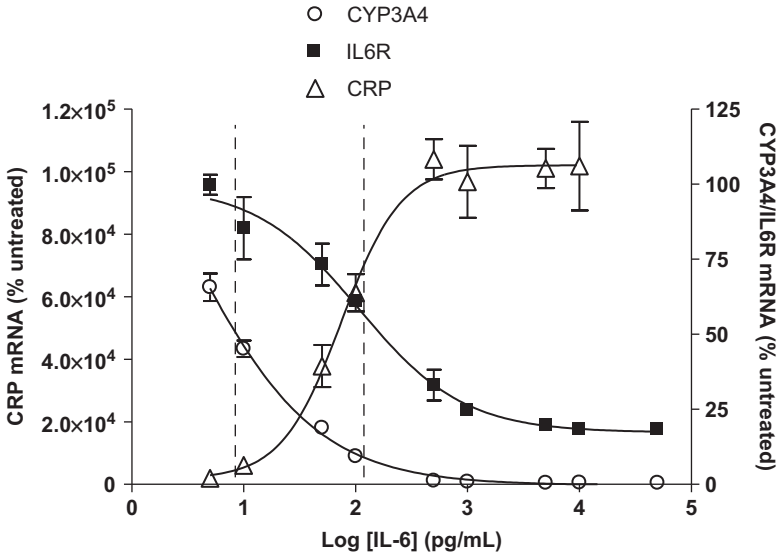


**FIGURE 13-2** Single donor data showing the dose-response of human hepatocytes to interleukin 6 (IL-6). mRNA expression of CYP3A4 is compared to mRNA of the acute phase response marker C-reactive protein (CRP) and mRNA expression of the IL-6 receptor. Dashed vertical bars show the normal physiological range of IL-6 and IL-1β serum concentrations in health and disease. Selected data from Ref. 61 were replotted with permission.

Across the spectrum of major drug metabolizing enzymes, CYP3A4 was the most sensitive to suppression by IL-6 and CYP1As were the least sensitive. The $EC_{50}$ values of many CYP isoenzymes were above the range of physiological IL-6 concentrations. The dynamic range of IL-6 responsiveness was determined in human hepatocytes and compared to the development of the hepatocyte APR, as measured by CRP production. In Figure13-2, a single hepatocyte donor dose-response to IL-6 shows that CYP suppression, CRP production, and IL-6 receptor (IL-6R) downregulation all occurred within or near the physiological range of serum IL-6 concentrations. This suggests two intriguing possibilities. First, that the hepatocytes in this model corroborate clinical indications that CYP suppression and the APR are correlated, indicating that CRP could be a bona fide biomarker of CYP suppression and desuppression in a large clinical population. Second, IL-6R downregulation may indicate an adaptive and/or self-limiting response to sustained cytokine exposure, and this could translate into the relatively modest downregulation of the CYP isoenzymes that has been observed clinically to date.

We also investigated the dynamic range of inflammation effects on CYP expression *in vitro* by studying cytokine-mediated CYP suppression during sustained induction pressure. Rifampin and omeprazole were used to induce CYP3A4 and CYP1A2, respectively. Both inducers shifted the IL-6 dose-response curve to the right, affording new $EC_{50}$ values that were well above the physiologically relevant range of IL-6 concentrations. Therefore, at least *in vitro*, CYP suppression by

cytokines may not necessarily overcome CYP induction by concomitant administration of a CYP inducer. The effect of CYP induction and suppression in rodents *in vivo* has been studied,[64] but a definitive study on the magnitude of CYP induction in inflammatory disease patients, relative to healthy individuals, could not be found.

No *in vitro* experiments have yet addressed the antibody-mediated abrogation of IL-6-mediated suppression of CYP that is the basis of the tocilizumab–simvastatin interaction. In human hepatocytes, an anti-IL-6 monoclonal antibody shifted the IL-6 dose-response curve for suppression of CYP3A4 activity to the right. Although this expected result indicated that the human hepatocyte model can indeed measure desuppression, the magnitude of the effect was a consequence of a long list of experimental variables, including antibody concentration and the competing affinities of antibodies, cytokine ligands, and their receptors.

Overall, these data illustrated that an individual cytokine can be studied in hepatocytes *in vitro* and can afford quantitative CYP suppression and desuppression data. However, these data represent only a small component of the complicated biology of the APR and/or chronic inflammation in the intact liver. This limits the value of the model for quantitative prediction of clinical drug–disease interactions related to CYP suppression and desuppression.

### 13.4.2 Is Interleukin-6 Always Involved When Hepatic CYP Suppression Occurs?

IL-6 is secreted by several different cell types, including T-cells, osteoblasts, and macrophages such as Kupffer cells in the liver. Acting via the IL-6 receptor on hepatocytes, IL-6 uniquely mediates diverse hepatic effects that include the APR, hepatic regeneration, and decreases in apoptosis and necrosis.[62] Tocilizumab efficacy data[65] have shown that the IL-6/IL-6R axis is a valid target for the treatment of rheumatoid arthritis. Similarly, the tocilizumab–simvastatin interaction has illustrated that IL-6 is also a key mediator of inflammatory disease manifestations in the liver.[37] Many other cytokines have been shown to cause CYP suppression *in vitro* (interferons, IL-1, IL-2, IL-10, TNF, etc.), yet the diverse roles of IL-6 in the liver may (or may not) render it preeminent among all cytokines as a biomarker or rate-limiting mediator of a clinically measurable effect. As such, cytokine combination experiments *in vitro* might help determine which cytokines are key drivers of hepatic CYP responses to systemic inflammation.

### 13.4.3 Recapitulating Hepatic Inflammation *in vitro* with Cytokine Combinations

The evaluation of a single cytokine in isolation in hepatocyte culture *in vitro* has little chance of recapitulating the complexity of acute or chronic inflammation in the liver *in vivo*. Many studies have examined the effect of LPS endotoxin,[66,67] and others have systematically dissected components of the APR through the study of individual cytokines.[19,21,68–70] To probe whether two clinically significant cytokines could act additively or synergistically *in vitro*, we examined the effects of

IL-1β alone and in combination with IL-6, on drug-metabolizing enzymes in human hepatocytes.[63,71,72]

It has been shown that IL-1β can suppress hepatic drug-metabolizing enzyme and transporter levels, implying a role of IL-1β in CYP suppression.[19,41] As with IL-6, isoform-specific differences have been observed in CYP suppression in IL-1β-treated hepatocytes *in vitro*.[19,46] However, all studies conducted to date have examined supraphysiologic concentrations of IL-1β relative to physiologically relevant serum concentrations.

Like IL-6, IL-1β alone suppressed mRNA of several CYP isoenzymes, with $EC_{50}$ values that differed by isoenzyme.[63] $EC_{50}$ values of IL-1β were markedly higher than those of IL-6. IL-1β also caused upregulation of CRP and downregulation of IL-6R mRNA. These two events occurred near the physiological range of serum IL-1β concentrations.[73,74] As shown in Figure13-3,[63] IL-1β and IL-6 combinations at concentrations from 1 to 100 pg/mL in human hepatocytes demonstrate that IL-6 was the dominant cytokine in terms of increases in CRP mRNA as well as decreases in CYP3A4 and 2B6 mRNA.

The effects of pairwise or ternary combinations of other cytokines on CYPs, when benchmarked to physiological serum cytokine concentrations, may inform

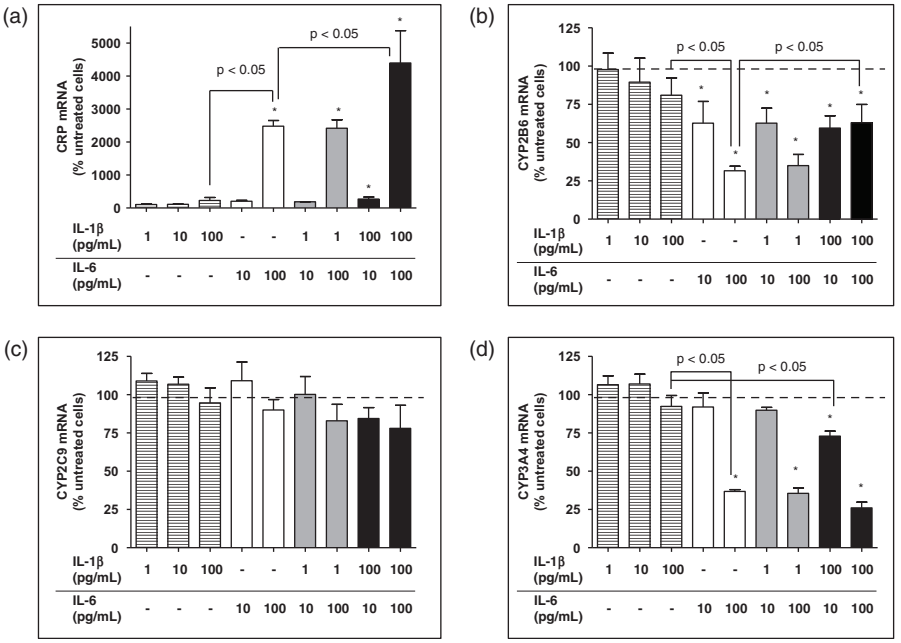


**FIGURE 13-3** A comparison of the acute phase reactant (a) C-reactive protein (CRP), (b) CYP2B6, (c) CYP2C9, and (d) CYP3A4 mRNA responses in human hepatocytes treated with interleukin 6 (IL-6) and IL-1β. Asterisk: a p-value <0.05 when compared to untreated cells. Adapted from Ref. 63, Figures 2 and 3 with permission.

which cytokines are dominant in effects on CYP in isolated hepatocytes. These experiments may help in the development of a mechanistic understanding of which cytokines are the most important, even if this approach is certain to fall short of the *in vitro* reconstitution of human disease in liver. Important data in these investigations are the presence of relevant cytokine receptors on hepatocytes, the measurement of CRP and the downregulation of relevant cytokine receptors. Cosgrove et al.[75] developed a cytokine combination matrix and coincubated major cytokines with drugs known to elicit idiosyncrasy with human hepatocytes. The study evaluated synergy among major cytokines in the manifestation of drug-induced hepatocellular toxicity and may represent a useful approach for the systematic assessment of effects of ternary cytokine combinations on CYP suppression.

### 13.4.4 CYP Suppression and Desuppression in Murine Models of Inflammatory Disease *In Vivo*

There are many published studies on suppression of drug-metabolizing enzymes by cytokines in rodent models. A general inflammatory stimulus from bacterial LPS in preclinical *in vivo* models suppresses CYP probe substrate clearances generally by onefold to twofold.[66] However, there are relatively few studies on disease treatment effects on CYP suppression. Piquette-Miller and Jamali[59] studied nonsteroidal anti-inflammatory drug (NSAID) treatment effects on propranolol pharmacokinetics in severe adjuvant arthritis in rats. They investigated CYP markers, protein binding, and cytokine levels for this basic, AAG-binding drug. Ling and Jamali[60] studied the effect of the anti-TNF monoclonal antibody infliximab on the resolution of early adjuvant arthritis in rats by examining effects on CYPs, inflammation markers, and pharmacokinetics of the AAG-binding molecule verapamil. Ashino et al.[76] studied the effect of a murine anti-IL-6 monoclonal antibody treatment on arthritis in a transgenic mouse arthritis model. The effects of disease and treatment on CYP activity and expression were measured. In the Ashino study, IL-6 was markedly upregulated and CYP3A was the most sensitive CYP isoenzyme to suppression in arthritic rats. CYP3A protein content fell to 60% of that in nonarthritic transgenic rats, and an anti-IL-6 antibody normalized this response.

The collagen antibody-induced arthritis (CAIA) mouse model of rheumatoid arthritis is used to screen therapeutic candidates. In this model, arthritis is induced by initial administration of a monoclonal antibody cocktail directed toward epitopes in collagen type II that are highly autoantigenic, followed by administration of an endotoxin such as LPS.[77] Arthritis is induced within a week of initiation, with the severity of arthritis corresponding to the dose of monoclonal antibody cocktail.[78]

We adopted murine CAIA as a potential fast-turnaround model that might help in the development of *in vitro* to *in vivo* correlations.[79,80] On days 5 and 7 after initiation of the CAIA protocol, serum IL-1β, IL-6, and the APR marker serum amyloid A (SAA) were significantly elevated (Figure13-4), whereas both IL-2 and TNF-α were below limits of quantitation (not shown in the figure). Hepatic CYP mRNA levels were generally downregulated in the CAIA model, although CYP1b1, 2d9, and 3a13 were upregulated, relative to arthritis-naive animals at day 5 of treatment.

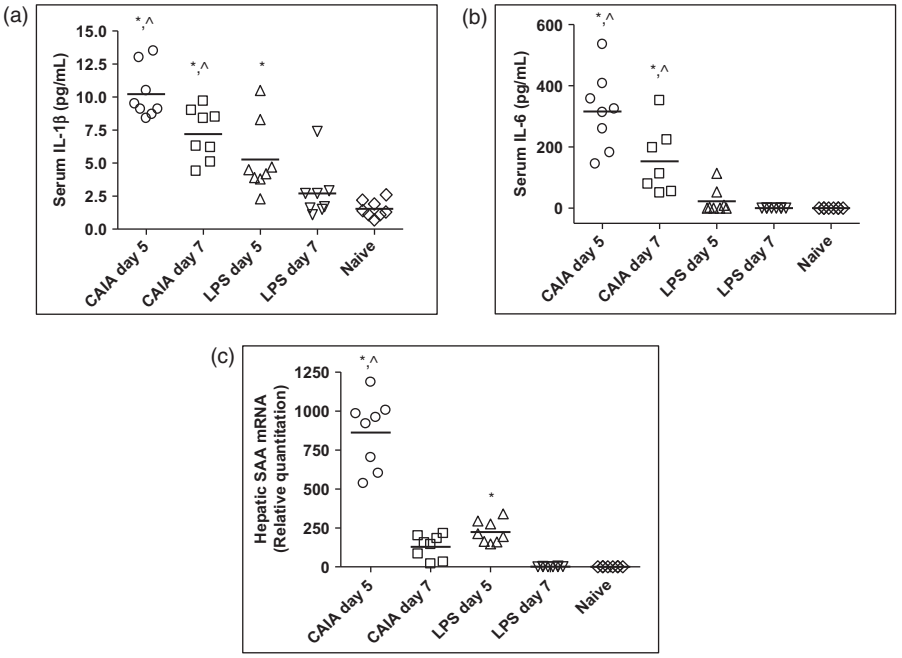


**FIGURE 13-4** Serum cytokine and hepatic serum amyloid A (SAA) levels measured in naive, lipopolysaccharide (LPS) treated and collagen antibody-induced arthritis (CAIA). mice (a) Interleuken 1β (IL-1β) protein, (b) IL-6 protein, and (c) SAA mRNA. Serum IL-2 and tumor necrosis factor α were below the limits of quantitation. Adapted from Ref. 79 with permission.

CYP1a2, 2c29, 2b9, and 3a11 mRNA were significantly downregulated to 36%, 29%, 44%, and 45% of naive values at day 5, whereas CYP2e1 levels were unchanged (Table13-1).[79,80] These data were consistent with results obtained using a transgenic mouse model of arthritis where CYP2c29 and CYP3a11 hepatic mRNA were downregulated to 75% and 45%, respectively, compared to wild type animals.[76] Other models are known to afford strain- and isoform-dependent CYP suppression data, such as in a bacterial LPS model of acute inflammation.[67] Taken together, these studies indicate that the direction and magnitude of changes in hepatic CYP enzymes in preclinical inflammation models depend on both the model of inflammation and the mouse strain used.

Intrinsic clearance data obtained using hepatic S9 with phenacetin, bupropion, diclofenac, dextromethorphan, and testosterone CYP probe substrate activities suggests that the maximum decrease in CYP activity in the CAIA model is approximately twofold (Figure13-5).[79,80] With the exception of dextromethorphan, the CAIA day 5 group and the LPS day 5 group had nearly identical profiles. This suggested that the observed decreases in the CAIA day 5 intrinsic clearance may be due to the administration of LPS alone and not the development of CAIA. Indeed, paw swelling in the CAIA model did not correlate with CYP suppression because CAIA

**TABLE 13-1 Relative mRNA Levels Compared to Either Naive Animals (CAIA or LPS-Treated Mice) or Untreated Cells (Mouse Hepatocytes)[a]**

| | CAIA Day 5 | CAIA Day 7 | Hepatocytes, 10/500[b] IL-1β/IL-6 | Hepatocytes 10/100[b] IL-1β/IL-6 | LPS Day 5 | LPS Day 7 |
|---|---|---|---|---|---|---|
| CYP1a2 | 0.36*,** | 0.93* | 0.13* | 0.30* | 0.97 | 1.28 |
| CYP2b9 | 0.44*,** | 0.77* | 0.21* | 0.39* | 0.77 | 0.84 |
| CYP2c29 | 0.29*,** | 0.79 | 0.38* | 0.41* | 0.56* | 0.78* |
| CYP2d9 | 2.28* | 0.95 | 0.51* | 0.64* | 2.47* | 1.14 |
| CYP2e1 | 1.10 | 0.77 | 0.02* | 0.03* | 1.06 | 0.72 |
| CYP3a11 | 0.45*,** | 0.54* | 0.19* | 0.38* | 0.93 | 0.65* |
| CYP3a13 | 4.37*,** | 1.97*,** | 0.90 | 0.88 | 1.59 | 0.66 |
| POR | 1.84* | 1.38 | 0.46* | 0.60 | 1.87* | 0.92 |
| SAA | 918* | 144* | 464* | 18.0* | 223* | 0.98 |

[a]Adapted from Ref. 79 with permission.
[b]Units of pg/mL.
*$p < 0.05$ when compared to the naive group (CAIA) or untreated cells (hepatocytes);
**$p < 0.05$ when compared to the respective LPS counterpart.
CAIA: collagen antibody-induced arthritis; IL: interleukin; LPS: bacterial lipopolysaccharide; POR: P450 oxidoreductase; SAA: serum amyloid A.

day 5 mice had, on average, a clinical score of 2.6 (out of 16) while the CAIA day 7 mice had an average score of 6.4. The differences observed between day 5 and day 7 for CYP suppression, cytokine markers levels, and disease severity show that this model is time dependent. The CAIA model measured maximum CYP suppression

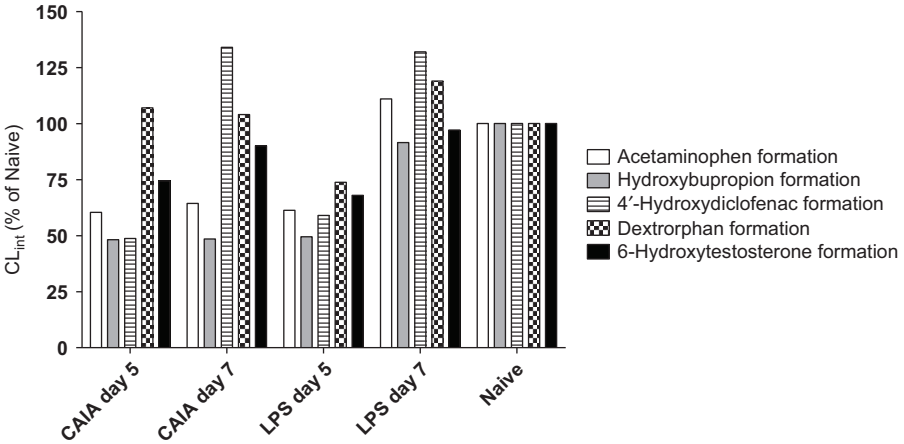


**FIGURE 13-5** *In vitro* intrinsic clearance ($CL_{int}$) of several cytochrome CYP substrates in naive, lipopolysaccharide (LPS) treated and collagen antibody-induced arthritis (CAIA) mice. Values were obtained from pooled S9 fractions ($n = 8$ animals/group). Eadie-Hofstee plots indicated a multiple enzyme system for acetaminophen and 6β-hydroxytestosterone clearance. Changes in $CL_{int}$ reflected changes in both $V_{max}$ and $K_m$. Adapted from Ref. 79 with permission.

that was in the same general range as other murine models. However, aside for a short experiment cycle time, there were no compelling advantages over other arthritis models.

### 13.4.5 *In Vivo*/*In Vitro* Similarities and Differences between the CAIA Model and Cultured Mouse Hepatocytes Treated with a Cytokine Cocktail

Serum cytokine levels from the CAIA study were used as a guide for experiments on CYP suppression in mouse hepatocyte culture. While some CYP isoforms (CYP1a2, 2b9, 2c29, and 3a11) were in agreement in terms of magnitude and direction of suppression in both the *in vitro* and *in vivo* models, several isoforms (CYP2d9, 2e1, 3a13) were not (Table13-1).[79,80] Although this *in vitro*/*in vivo* comparison was rudimentary, it highlights the difficulty in properly reconstructing a complex inflammatory disease state in an *in vitro* hepatocyte model.

### 13.4.6 Are *In Vitro* Data Adequate to Preclude the Need for a Clinical DDI Study?

The pros of investigating cytokine-mediated effects on CYP expression in human hepatocytes are that *in vitro* experimentation affords quantitative data showing that ablating a particular inflammatory cytokine might de-suppress/normalize CYP activity. As such, mechanisms can be studied and hypotheses developed using this mainstream industrial technology.[61] However, it is at least technically challenging if not impossible, to accurately re-create an inflammatory disease state in hepatocytes *in vitro* and then treat it. This would be a crucial foundation for a clinically relevant and actionable *in vitro-in vivo* extrapolation. The single cytokine approach in hepatocytes is not likely to quantitatively predict subtle disease treatment effects on CYPs in humans *in vivo*. The diverse variables that are not controlled in hepatocytes include relevant hepatic cytokine concentrations, time-dependent adaptive response mechanisms like cytokine receptor downregulation, the absence of other interacting cytokines, the absence (or unquantified presence) of crucial cell types like Kupffer cells in primary hepatocyte culture, and of course the nuances of patient and disease-specific aspects of inflammatory disease.[61] Other models, such as Kupffer cell coculture may, theoretically, better recapitulate cytokine effects in whole liver. However, CYP3A suppression by IL-1β and IL-6 was not significantly altered by the addition of Kupffer cells in the first report of the application of this model to CYP suppression.[21] As such, human hepatocyte coculture is still in development, and the benefits and limitations of the technique for the study of cytokine-mediated suppression of CYPs have not been established.

Attempts are under way to simulate human pharmacokinetic outcomes based on currently available *in vitro* and *in vivo* data on IL-6. For example, cyclosporine pharmacokinetics and IL-6 concentrations in transplant patients[20] and *in vitro* human hepatocyte data from Aitken and Morgan[19] were used with a simCYP minimal physiologically based pharmacokiuetic (PBPK) model to retrospectively

predict a 2-5-fold increase in cyclosporine levels, against an actual 2-7-fold change.[81] Kenny[82] tested *in vitro* drug–drug interaction calculations using human hepatocyte data on IL-6-mediated suppression of CYP3A4 to calculate $K_i$ and $k_{\text{inact}}$ values. Population simulations were developed via simCYP using these *in vitro* data, along with clinical IL-6 and simvastatin concentrations from Schmitt et al.[37] to retrospectively simulate the magnitude of the tocilizumab–simvastatin interaction at low and high systemic IL-6 concentrations. They confirmed that IL-6 was a key variable in the interaction because closer predictions were obtained at higher IL-6 serum concentrations. More examples of clinically measurable effects of diverse cytokines and related therapeutic proteins on small molecule pharmacokinetics are needed before the potential merits of predictions based on preclinical *in vitro* experiments can be established for cytokines other than IL-6.

## 13.5 CURRENT REGULATORY PERSPECTIVE

The tocilizumab–simvastatin interaction was an important event from a regulatory perspective. The clinical observation of increased simvastatin oral clearance after tocilizumab treatment in patients with rheumatoid arthritis, along with some related *in vitro* characterization of IL-6 effects on CYPs in human hepatocytes, afforded a drug interaction label recommending caution for tocilizumab use with narrow therapeutic index CYP substrates.[4] Subsequently, regulatory authorities have requested *in vitro* and/or *in vivo* study data for other cytokine-modulating biologics such as golimumab[5] and for some therapeutic proteins that do not have direct connections to inflammatory disease. A science-driven article by regulatory scientists at the FDA on therapeutic protein–drug interactions (TP-DIs) was published recently,[5] and elements of that paper have reappeared in the revised draft drug interaction guidance that was issued for comment in 2012.[83] For cytokines or cytokine modulators, the new draft guidance advocates clinical investigation in targeted patient populations using probe substrates or probe substrate cocktails. Consistent with the complexities outlined in this chapter, the draft guidance also states that *in vitro* or animal studies have limited value in the qualitative and quantitative projection of clinical interactions because translation of *in vitro* to *in vivo* and animal to human results to date have been inconsistent.[83]

## 13.6 CLINICAL OPTIONS

Drug–disease interactions associated with an inflammatory disease could be investigated in a compound independent way by comparing CYP probe clearances in adequate numbers of patients with a given disease, against age- and sex-matched healthy controls. This would help establish the absolute upper magnitude of drug-metabolizing enzyme suppression for a given disease state. For disease treatment effects on CYP-mediated metabolism, midazolam AUC could be compared before and after treatment with the biologic agent. Alternatively, if it is thought that

CYP3A4 may not be the most sensitive CYP isoenzyme, a CYP probe cocktail approach may be pursued in a similar fashion.[84] To date, the study conducted by Roche on the tocilizumab–simvastatin interaction has a hallmark design, as it included data on the biomarkers IL-6 and CRP and included two sampling points after the administration of the long half-life antibody.[37] Whether the 80–125% confidence intervals employed in the tocilizumab–simvastatin interaction study to discern a clinically measurable interaction are appropriate for other compounds or diseases is open to discussion. Alternatively, confidence intervals could be set in a case-by-case fashion, possibly based on statistical assessment of treatment-induced changes in inflammatory disease severity biomarkers such as CRP, known population pharmacokinetic norms for narrow therapeutic index victim drugs in health and disease states, or variability in the expression of CYPs of interest.

In most cases, therapeutic protein drug–disease interaction studies will need to be conducted in the target population and are complicated by the long duration of the effects of most therapeutic antibodies. These and other factors may place practical limitations on the number of treatment arms in the study and may exclude some simple DDI study designs. Some study design options for DDI are overviewed in Huang et al.[5] When a cotherapy is likely to be employed extensively in the intended clinical population and plasma concentrations of the concomitant medication are monitored in larger clinical trials, a population pharmacokinetics–based approach may discern a measurable change in exposure to a select victim drug.

Industry scientists and regulators alike are interested in knowing whether *in vitro* data will ever be understood well enough to be used to develop a therapeutic protein drug interaction label that will effectively protect patients and inform healthcare practitioners. New therapeutic cytokines, anticytokine antibodies, and other biologic agents may have direct effects on CYP expression in hepatocytes, or indirect effects via hepatic manifestations of systemic inflammation, or both. It will continue to be difficult to rule out the need for a clinical pharmacokinetic study based on *in vitro* data alone because current *in vitro* models entail risk of false positive or false negative data and expected clinical effects are mild relative to classical CYP inducers and inhibitors.

The alternative is to proceed directly to clinical investigation. Some factors in determining the need for a clinical study might include a lack of direct effect on CYP3A4 in hepatocytes *in vitro* but might also consider whether relevant cytokine receptors are present on hepatocytes and whether large changes in systemic concentrations of inflammatory markers or APR proteins are expected in the target patient population. A risk-based, case-by-case approach is advised.

## 13.7 CONCLUSIONS

The chapter focused on how inflammatory disease can affect the expression of CYPs and other drug-metabolizing enzymes. In particular, clinical experience with inflammation-related effects on drug metabolism was reconciled with the *in vitro*

and *in vivo* models that could be used to model or predict the effects of disease, or cytokine/anticytokine treatments on drug metabolism. Recent work was used to illustrate the pros and cons of preclinical experimentation and to make recommendations with respect to actionable approaches to the development of a robust and informative drug interaction label.

## ACKNOWLEDGMENTS

Sonal Patel, Dan Rock, Josh Pearson, Graham Jang, Chris Endres (Amgen PKDM), Yudong He (Amgen Comparative Biology and Safety Sciences), Helen McBride, Kent Miner (Amgen Inflammation research). Raymond Evers (Merck Drug Metabolism) and the *in vitro* LM DDI working group of the IQ Consortium.

## DECLARATION OF INTEREST

The authors were employed by Amgen, Inc.

## REFERENCES

1. Piquette-Miller M, Grant DM. The art and science of personalized medicine. *Clin Pharmacol Ther* **81**, 311–15 (2007).
2. Kraynov E, Martin SW, Hurst S, et al. How current understanding of clearance mechanisms and pharmacodynamics of therapeutic proteins can be applied for evaluation of their drug-drug interaction potential. *Drug Metab Dispos* **39**, 1779–83 (2011).
3. Zhou H, Mascelli MA. Mechanisms of monoclonal antibody-drug interactions. *Annu Rev Pharmacol Toxicol* **51**, 359–72 (2011).
4. Lee JI, Zhang L, Men AY, et al. CYP-mediated therapeutic protein-drug interactions: clinical findings, proposed mechanisms and regulatory implications. *Clin Pharmacokinet* **49**, 295–310 (2010).
5. Huang SM, Zhao H, Lee JI, et al. Therapeutic protein-drug interactions and implications for drug development. *Clin Pharmacol Ther* **87**, 497–503 (2010).
6. Morgan ET. Impact of infectious and inflammatory disease on cytochrome P450-mediated drug metabolism and pharmacokinetics. *Clin Pharmacol Ther* **85**, 434–38 (2009).
7. Zhou H, Davis HM. Risk-based strategy for the assessment of pharmacokinetic drug-drug interactions for therapeutic monoclonal antibodies. *Drug Discov Today* **14**, 891–98 (2009).
8. Seitz K, Zhou H. Pharmacokinetic drug-rug interaction potentials for therapeutic monoclonal antibodies: reality check. *J Clin Pharmacol* **47**, 1104–18 (2007).
9. Petrovic V, Teng S, Piquette-Miller M. Regulation of drug transporters during infection and inflammation. *Mol Interv* **7**, 99–111 (2007).
10. Mahmood I, Green MD. Drug interaction studies of therapeutic proteins or monoclonal antibodies. *J Clin Pharmacol* **47**, 1540–54 (2007).

11. Kulmatycki KM, Jamali F. Drug disease interactions: role of inflammatory mediators in disease and variability in drug response. *J Pharm Pharm Sci* **16**, 8: 602–25 (2005).
12. Haas CE. Drug-cytokine interactions. In Piscitelli SC, ed *Drug interactions in infectious diseases*. Totawa NJ, Humana Press, 2005: 431–61.
13. Pharmacokinetics and Disposition Expert Working Group Drug-Drug Interaction Taskforce, Lyoyd P, Zhou H, Theil FP, Kakkar T, Nestorov I, roberts SA. Highlights from a recent BIO survey on therapeutic protein-drug interactions. *J. clin Pharmacol*. **52**: 1755–63 (2012).
14. Girish S, Martin SW, Peterson MC, et al. AAPS workshop report: strategies to address therapeutic protein-drug interactions during clinical development. *AAPS J* **13**, 405–16 (2011).
15. Chang KC, Bell TD, Lauer BA, Chai H. Altered theophylline pharmacokinetics during acute respiratory viral illness. *Lancet* **1**(8074), 1132–33 (1978).
16. Shedlofsky SI, Israel BC, McClain CJ, et al. Endotoxin administration to humans inhibits hepatic cytochrome P450-mediated drug metabolism. *J Clin Invest* **94**, 2209–14 (1994).
17. Brockmeyer NH, Barthel B, Mertins L, Goos M. Changes of antipyrine pharmacokinetics during influenza and after administration of interferon-alpha and -beta. *Int J Clin Pharmacol Ther* **36**, 309–11 (1998).
18. Flaman AS, Gravel C, Hashem AM, et al. The effect of interferon-α on the expression of cytochrome P450 3A4 in human hepatoma cells. *Toxicol Appl Pharmacol* **253**, 130–36 (2011).
19. Aitken AE, Morgan ET. Gene-specific effects of inflammatory cytokines on cytochrome P450 2C, 2B6 and 3A4 mRNA levels in human hepatocytes. *Drug Metab Disp* **35**, 1687–93 (2007).
20. Chen C, Han YH, Yang Z, Rodrigues AD. Effect of interferon-α2b on the expression of various drug-metabolizing enzymes and transporters in co-cultures of freshly prepared human primary hepatocytes. *Xenobiotica*. **41**:476–85 (2011).
21. Sunman JA, Hawke RL, LeCluyse EL, Kashuba AD. Kupffer cell-mediated IL-2 suppression of CYP3A activity in human hepatocytes. *Drug Metab Dispos* **32**, 359–363 (2004).
22. Nakai K, Tanaka H, Hanada K, et al. Decreased expression of cytochromes P450 1A2, 2E1, and 3A4 and drug transporters Na+-taurocholate-cotransporting polypeptide, organic cation transporter 1, and organic anion-transporting peptide-C correlates with the progression of liver fibrosis in chronic hepatitis C patients. *Drug Metab Dispos* **36**, 1786–93 (2008).
23. Kulmatycki KM, Jamali F. Therapeutic relevance of altered cytokine expression. *Cytokine* **14**, 1–10 (2001).
24. Jetter A, Fätkenheuer G, Frank D, et al. Do activities of cytochrome P450 (CYP)3A, CYP2D6 and P-glycoprotein differ between healthy volunteers and HIV-infected patients? *Antivir Ther* **15**, 975–83 (2010).
25. Jones AE, Brown KC, Werner RE, et al. Variability in drug metabolizing enzyme activity in HIV-infected patients. *Eur J Clin Pharmacol;* **66**, 475–85 (2010).
26. Daneshtalab N, Lewanczuk RZ, Russell AS, Jamali F. Drug-disease interactions: losartan effect is not downregulated by rheumatoid arthritis. *J Clin Pharmacol* **46**, 1344–55 (2006).
27. Carcillo JA, Doughty L, Kofos D, et al. Cytochrome P450 mediated-drug metabolism is reduced in children with sepsis-induced multiple organ failure. *Intensive Care Med* **29**, 980–84 (2003).

28. Haack MJ, Bak ML, Beurskens R, et al. Toxic rise of clozapine plasma concentrations in relation to inflammation. *Eur Neuropsychopharmacol* **13**, 381–85 (2003).
29. Raaska K, Raitasuo V, Neuvonen PJ. Effect of influenza vaccination on serum clozapine and its main metabolite concentrations in patients with schizophrenia. *Eur J Clin Pharmacol* **57**, 705–08 (2001).
30. Haas CE, Kaufman DC, Jones CE, et al. Cytochrome P450 3A4 activity after surgical stress. *Crit Care Med* **31**, 1338–46 (2003).
31. Frye RF, Schneider VM, Frye CS, Feldman AM. Plasma levels of TNF-alpha and IL-6 are inversely related to cytochrome P450-dependent drug metabolism in patients with congestive heart failure. *J Card Fail* **8**, 315–19 (2002).
32. Rivory LP, Slaviero KA, Clarke SJ. Hepatic cytochrome P450 3A drug metabolism is reduced in cancer patients who have an acute-phase response. *Br J Cancer* **87**, 277–80 (2002).
33. Charles KA, Rivory LP, Stockler MR, et al. Predicting the toxicity of weekly docetaxel in advanced cancer. *Clin Pharmacokinet* **45**, 611–22 (2006).
34. Chen YL, Le Vraux V, Leneveu A, et al. Acute-phase response, interleukin-6, and alteration of cyclosporine pharmacokinetics. *Clin Pharmacol Ther* **55**, 649–60 (1994).
35. Beyeler C, Frey BM, Bird HA. Urinary 6 beta-hydroxycortisol excretion in rheumatoid arthritis. *Br J Rheumatol* **36**, 54–58 (1997).
36. Kovarik JM, Koelle EU. Cyclosporin pharmacokinetics in the elderly. *Drugs Aging* **15**, 197–205 (1999).
37. Schmitt C, Kuhn B, Zhang X, et al. Disease-drug-drug interaction involving tocilizumab and simvastatin in patients with rheumatoid arthritis. *Clin Pharmacol Ther* **89**, 735–40 (2011).
38. Kyrklund C, Backman JT, Kivistö KT, et al. Rifampin greatly reduces plasma simvastatin and simvastatin acid concentrations. *Clin Pharmacol Ther* **68**, 592–97 (2000).
39. Chung E, Nafziger AN, Kazierad DJ, Bertino JS Jr. Comparison of midazolam and simvastatin as cytochrome P450 3A probes. *Clin Pharmacol Ther* **79**, 350–61 (2006).
40. Rhodes B, Fürnrohr BG, Vyse TJ. C-reactive protein in rheumatology: biology and genetics. *Nat Rev Rheumatol* **7**, 282–89 (2011).
41. Ashino T, Oguro T, Shioda S, et al. Involvement of interleukin-6 and tumor necrosis factor alpha in CYP3A11 and 2C29 down-regulation by Bacillus Calmette-Guerin and lipopolysaccharide in mouse liver. *Drug Metab Dispos* **32**, 707–14 (2004).
42. Morgan ET. Regulation of cytochrome p450 by inflammatory mediators: why and how? *Drug Metab Dispos* **29**, 207–12 (2001).
43. Zhou C, Verma S, Blumberg B. The steroid and xenobiotic receptor (SXR), beyond xenobiotic metabolism. *Nucl Recept Signal* **7**, e001 (2009).
44. Congiu M, Mashford ML, Slavin JL, Desmond PV. Coordinate regulation of metabolic enzymes and transporters by nuclear transcription factors in human liver disease. *J Gastroenterol Hepatol* **24**, 1038–44 (2009)
45. Renton KW and Nicholson TE. Hepatic and central nervous system cytochrome P450 are down-regulated during lipopolysaccharide-evoked localized inflammation in brain. *J Pharmacol Exp Ther* **294**, 524–30 (2000).
46. Lee, CM, Pohl J, Morgan ET. Dual mechanisms of CYP3A protein regulation by pro-inflammatory cytokine stimulation in primary hepatocyte cultures. *Drug Metab Dispos* **37**, 865–72 (2009).

47. Ferrari L, Peng N, Halpert JR, Morgan ET. Role of nitric oxide in down-regulation of CYP2B1 protein, but not RNA, in primary cultures of rat hepatocytes. *Mol Pharmacol* **60**, 209–16 (2001).
48. Jover R, Bort R, Gómez–Lechón MJ, Castell JV. Down-regulation of human CYP3A4 by the inflammatory signal interleukin-6: molecular mechanism and transcription factors involved. *FASEB J* **16**, 1799–01 (2002).
49. Zhou C, Tabb MM, Nelson EL, et al. Mutual repression between steroid and xenobiotic receptor and NF-kappaB signaling pathways links xenobiotic metabolism and inflammation. *J Clin Invest* **116**, 2280–89 (2006).
50. Xie W, Tian Y. Xenobiotic receptor meets NF-kappaB, a collision in the small bowel. *Cell Metab* **4**, 177–78 (2006).
51. Zordoky BN, El-Kadi AO. Role of NF-kappaB in the regulation of cytochrome P450 enzymes. *Curr Drug Metab* **10**, 164–78 (2009).
52. Pascussi JM, Gerbal–Chaloin S, Pichard-Gracia L, et al. Interleukin-6 negatively regulates the expression of pregnane X receptor and constitutively activated receptor in primary human hepatocytes. *Biochem Biophys Res Commun* **274**, 707–13 (2000).
53. Ghose R, Zimmerman TL, Thevananther S, Karpen SJ. Endotoxin leads to rapid subcellular re-localization of hepatic RXRalpha: a novel mechanism for reduced hepatic gene expression in inflammation. *Nucl Recept* **2**, 4 (2004).
54. Gu X, Ke S, Liu D, et al. Role of NF-kappaB in regulation of PXR-mediated gene expression: a mechanism for the suppression of cytochrome P-450 3A4 by pro-inflammatory agents. *J Biol Chem* **281**, 17882–889 (2006).
55. Yang X, Zhang B, Molony C, et al. Systematic genetic and genomic analysis of cytochrome P450 enzyme activities in human liver. *Genome Res* **20**, 1020–36 (2010).
56. Slatter JG. Preclinical models for drug interactions with biologics: P450 based DDIs following inflammatory disease treatment with biologics. Paper presented at the Thirteenth Annual Land O' Lakes Conference on Drug Metabolism and Applied Pharmacokinetics, Meerimac, WI, September 13-17, 2010.
57. Giannetti AM, Wong H, Dijkgraaf GJ, et al. Identification, characterization, and implications of species-dependent plasma protein binding for the oral Hedgehog pathway inhibitor vismodegib (GDC-0449). *J Med Chem* **54**, 2592–01 (2011).
58. Mayo PR, Skeith K, Russell AS, Jamali F. Decreased dromotropic response to verapamil despite pronounced increased drug concentration in rheumatoid arthritis. *Br J Clin Pharmacol* **50**, 605–13 (2000).
59. Piquette-Miller M, Jamali F. Influence of severity of inflammation on the disposition kinetics of propranolol enantiomers in ketoprofen-treated and untreated adjuvant arthritis. *Drug Metab Dispos* **23**, 240–45 (1995).
60. Ling S, Jamali F. The effect of infliximab on hepatic cytochrome P450 and pharmacokinetics of verapamil in rats with pre-adjuvant arthritis: a drug-disease and drug-drug interaction. *Basic Clin Pharmacol Toxicol* **105**, 24–29 (2009).
61. Dickmann LD, Patel SK, Rock DA, et al. Effects of interleukin-6 (IL-6) and an anti-IL-6 monoclonal antibody on drug metabolizing enzymes in human hepatocyte culture. *Drug Metab Dispos* **39**, 1415–22 (2011).
62. Taub R. Liver regeneration: from myth to mechanism. *Nat Rev Mol Cell Biol* **5**, 836–47 (2004).

63. Dickmann LJ, Patel SK, Wienkers LC, Slatter JG. Effects of interleukin 1β (IL-1β) and IL-1β /interleukin 6 (IL-6) combinations on drug metabolizing enzymes in human hepatocyte culture. *Curr Drug Metab* **13**, 930–7 (2012).
64. Li-Masters T, Morgan ET. Effects of bacterial lipopolysaccharide on phenobarbital-induced CYP2B expression in mice. *Drug Metab Dispos* **29**, 252–57 (2001).
65. Rose-John S, Waetzig GH, Scheller J, et al. The IL-6/sIL-6R complex as a novel target for therapeutic approaches. **11**, 613–24 (2007).
66. Yang KH, Lee MG. Effects of endotoxin derived from Escherichia coli lipopolysaccharide on the pharmacokinetics of drugs. *Arch Pharm Res* **31**, 1073–86 (2008).
67. Richardson TA, Morgan ET. Hepatic cytochrome P450 gene regulation during endotoxin-induced inflammation in nuclear receptor knockout mice. *J Pharmacol Exp Ther* **314**, 703–09 (2005).
68. Morgan ET, Thomas KB, Swanson R, et al. Selective suppression of cytochrome P-450 gene expression by interleukins 1 and 6 in rat liver. *Biochim Biophys Acta* **1219**, 475–83 (1994).
69. Le Vee M, Lecureur V, Moreau A, et al. Differential regulation of drug transporter expression by hepatocyte growth factor in primary human hepatocytes. *Drug Metab Dispos* **37**, 2228–35 (2009).
70. Assenat, E. Gerbal-Chaloin S, Larrey D, et al., Interleukin 1β inhibits CAR-induced expression of hepatic genes involved in drug and bilirubin clearance. *Hepatology* **40**, 951–60 (2004).
71. Dickmann LJ, Patel SK, Wienkers LC, Slatter JG. *In vitro* effects of cytokines on cytochrome P450 expression in human hepatocyte culture. Paper presented at the 17th North American Regional Meeting of the International Society for the Study of Xenobiotics, Atlanta, GA., October 16–20, 2011.
72. Slatter JG. P450 Based drug-disease interactions following inflammatory disease treatment with biologics: studies in the CAIA murine arthritis model and related primary hepatocyte *in vitro*/*in vivo* correlations. Paper presented at the AAPS National Biotechnology Conference, San Francisco, May 15–18, 2011.
73. Hidaka T, Suzuki K, Kawakami M, et al. Dynamic changes in cytokine levels in serum and synovial fluid following filtration leukocytapheresis therapy in patients with rheumatoid arthritis. *J Clin Apher* **16**, 74–81 (2001).
74. Yilmaz M, Kendirli SG, Altintas D, et al. Cytokine levels in serum of patients with juvenile rheumatoid arthritis. *Clin Rheumatol* **20**, 30–35 (2001).
75. Cosgrove BD, King BM, Hasan MA, et al. Synergistic drug-cytokine induction of hepatocellular death as an *in vitro* approach for the study of inflammation-associated idiosyncratic drug hepatotoxicity. *Toxicol Appl Pharmacol* **237**, 317–30 (2009).
76. Ashino T, Arima Y, Shioda S, et al. Effect of interleukin-6 neutralization on CYP3A11 and metallothionein-1/2 expressions in arthritic mouse liver. *Eur J Pharmacol* **558**, 199–207 (2007).
77. Khachigian LM. Collagen antibody–induced arthritis. *Nat Protoc* **1**, 2512–16 (2006).
78. Chondrex. Adjuvant induced arthritis, 2011. Available at www.chondrex.com/animal-models/adjuvant-induced-arthritis-adjuvants.
79. Dickmann LJ, McBride HJ, Patel SK, et al. Murine collagen antibody induced arthritis (CAIA) and primary mouse hepatocyte culture as models to study cytochrome P450 suppression. *Biochem Pharmacol* **83**, 1682–9 (2012).

80. Dickmann LJ, Patel SK, McBride H, et al. Collagen antibody induced arthritis (CAIA) in mice and primary mouse hepatocyte culture as models to study P450 suppression: *in vivo/in vitro* similarities and differences. Paper presented at the 17th North American Regional Meeting of the International Society for the Study of Xenobiotics, Atlanta, GA, October 16–20, 2011.
81. Machavaram KK, Almond LM, Jamei M, Rostami-Hodjegan A. Predicting the effect of interleukin-6, an inflammatory cytokine, on time-dependent reduction of cyclosporine clearance: an application of the Simcyp Population-Based Simulator to suppression of CYP450 by biologics. Paper presented at the AAPS National Biotechnology Conference. San Francisco, May 16–18, 2011.
82. Kenny J. Cytokine mediated regulation of human cytochrome P450s—can we translate in vitro evaluation to clinical relevance? Paper presented at the AAPS National Biotechnology Conference. San Francisco, May 16–18, 2011.
83. US Food and Drug Administration, US Department of Health and Human Services, Center for Drug Evaluation and Research, Center for Biologics Evaluation and Research (CBER). Draft guidance for industry: drug interaction studies–study design, data analysis, implications for dosing, and labeling. Silver Spring, MD, 2012. Available at www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM292362.pdf.
84. Goh BC, Reddy NJ, Dandamudi UB, et al. An evaluation of the drug interaction potential of pazopanib, an oral vascular endothelial growth factor receptor tyrosine kinase inhibitor, using a modified Cooperstown 5+1 cocktail in patients with advanced solid tumors. *Clin Pharmacol Ther* **88**, 652–59 (2010).

# DRUG INTERACTIONS FOR GROWTH FACTORS AND HORMONES

YOW-MING C. WANG and TARUNDEEP KAKKAR

## 14.1 INTRODUCTION

According to the dictionary, the definition of *growth factor* is any of various proteins that promote the growth, organization, and maintenance of cells and tissues. And, the definition of *hormones* is any of various internally secreted compounds that affect the functions of specifically receptive organs or tissues when transported to them by the body fluids. Many of the proteins discussed in other chapters of this book, such as interferons and cytokines, may also be regarded as growth factors or hormones by these definitions. This chapter will discuss the traditional growth factors, primarily hematopoietic growth factors, and hormones, primarily growth hormones and glucagon.

The clinical applications of biotechnology-derived products began with growth factors and hormones in the 1980s. Two decades later, growth factors and hormones ranked the second and the third, respectively, of the top biologic drugs in terms of the U.S. sales[1] only behind therapeutic monoclonal antibodies. The total sales of growth factors and hormones combined accounted for approximately $20 billion (about $10 billion each).

In the growth factor category, epoetin α (recombinant human erythropoietin; rhEPO) approved by the U.S. Food and Drug Administration (FDA) in 1989 represents the first commercially available recombinant growth factor.[2] Filgrastim (recombinant human granulocyte colony stimulating factor, G-CSF) followed closely with an FDA approval in 1991. Subsequently, improvements to these recombinant human growth factors, by hyperglycosylation in the case of darbepoetin α[2,3] and by pegylation in the case of pegfilgrastim[4–6] and methoxy polyethylene glycol-epoetin β,[7,8] resulted in new generations of recombinant proteins with a longer

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*,
Edited by Shayne C. Gad
Copyright © 2013 John Wiley & Sons, Inc.

elimination half-life, enabling less frequent dose administration than the first-generation products. Palifermin (recombinant human keratinocyte growth factor; rhKGF) and romiplostim (a thrombopoietin stimulating agent) are growth factors that were approved in the past 10 years. These growth factors made substantial contributions to the therapeutic treatment of several chronic conditions.

Hormones, as a class, consist of a vast range of compounds to cover various ailments, and many have been on the market for decades. Most of the marketed drugs are of synthetic origin, including sex hormones such as estrogens, progesterone, thyroid hormone, and corticosteroids such as betamethasone. In 1960, the FDA approved Lilly's recombinant glucagon, which came more than 30 years after Murlin hypothesized that a hormone with a blood sugar–raising effect was present in the insulin product of low purity in 1923.[9] The first synthetic growth hormone, called somatrem, was developed in 1980s by Genentech in collaboration with a Swedish pharmaceutical company (Kabi), which commercialized the human growth hormone harvested and purified from cadavers.[10] The recombinant growth hormone somatropin, on the market now, is manufactured by an improved method.[10] Today, multiple recombinant human growth hormone and recombinant glucagon products are commercially available globally. Pegvisomant, a growth hormone analog and growth hormone receptor antagonist, was approved by European Medicines Agency (EMA) and the FDA in 2002 and 2003, respectively.

Pharmacokinetics (PK) drug interaction studies were not conducted as part of the clinical drug development program and did not play a role in the regulatory decision for many of the recombinant growth factors and hormone products. The PK drug interaction issue has been deemed unique to the small molecule chemical drugs and not of concern to the biotherapeutic agents because such agents are catabolized by proteolytic enzymes and do not share the metabolic pathways responsible for the *in vivo* disposition of small molecule drugs. Consequently, the understanding of PK drug interactions, or lack thereof, for these products primarily came from clinical experience in patient care over time and research conducted after the approval of marketing authorizations that was done to address these patient care findings.

Administration of two drugs indicated for treating the same disease condition is expected to show pharmacodynamic drug interactions because of additive or synergistic effects; often this type of drug interaction is factored into the study design and monitored during clinical trials and clinical use. For instance, romiplostim administration enables a reduction in the use of corticosteroids in treatment of thrombocytopenia, and the pharmacodynamic synergism of sex hormones is the basis for the use of follitropin α in combination with other agents to stimulate ovulation (e.g., human chorionic gonadotropin, clomiphene citrate) to potentiate the follicular response. This type of "polypharmacy" for additive therapeutic effect is not discussed in this chapter.

For many of the recombinant technology derived hormone products such as gonadotropins (luteinizing hormone, follicle-stimulating hormone, and human chorionic gonadotropin), minimal information is available in the literature and on product labels regarding drug interactions. We have opted to not discuss insulin because many drugs interact with insulin pharmacodynamically; some drugs alter the blood

glucose control and some drugs affect insulin sensitivity, which necessitate insulin dose adjustment. On the other hand, few reports on drug interaction involve insulin as a perpetrator even though in *in vitro* systems insulin has been implicated in modulating the expression of hepatic enzymes via its engagement in a multitude of cellular signaling pathways.

The focus of this chapter is on four recombinant growth factors products (erythropoietins, G-CSF, KGF, and thrombopoietins) and three hormone products (glucagon, somatropin, and pegvisomant). We compiled the information related to drug interactions found in scientific publications and regulatory agency approved labeling documents. The referenced U.S. prescribing information (PI) can be found at the Drugs@FDA website, and the European summary of product characteristics (SPC) can be found on the EMA's website for European public assessment reports (EPARs) by keyword search, using the international nonproprietary name (INN). In addition to the postings on regulatory agencies' websites, manufacturers and other interest groups also post up-to-date labeling documents on their product websites. Readers are referred to these online resources for the product labeling information cited in this chapter.

## 14.2 GROWTH FACTORS

### 14.2.1 Erythropoiesis-Stimulating Agents

Multiple commercial recombinant erythropoiesis-stimulating agents (ESAs)[2] are produced in Chinese hamster ovary (CHO) cells using recombinant DNA technology. They stimulate erythropoiesis by the same mechanism as the endogenous erythropoietin[11,12] and are indicated for treatment of anemia due to chronic kidney disease or chemotherapy. With the identical amino acid sequence as the natural erythropoietin produced by the kidney, epoetin α has 165 amino acids and a molecular weight (MW) of 30 kDa. Darbepoetin α, a second-generation ESA (MW 37 kDa), has two additional *N*-linked oligosaccharide chains compared to epoetin α, resulting in a prolonged terminal elimination half-life. The third-generation ESA product manufactured by Roche is methoxy polyethylene glycol (PEG)-epoetin β (MW 60 kDa), which differs from epoetin β in that PEG is covalently linked to some of the lysine residues in rhEPO, which also leads to a reduced clearance and a prolonged half-life. The prolonged half-life for these new ESAs afforded greater *in vivo* activity, despite the decrease of *in vitro* potency when compared to their respective first-generation counterpart. Other epoetins (e.g., epoetin β, epoetin δ, epoetin ζ) may have some structural differences from epoetin α and may or may not have differences in the biological activities.

The SPCs of epoetin α and darbepoetin α mentioned a potential for an interaction with substances that are highly bound to red blood cells, such as cyclosporine and tacrolimus, and recommended monitoring of the blood levels of the concomitant medications for assessment of potential dosage adjustment as the hemoglobin rises. Other implications of ESAs on clinical management of patients have been

reported where the observed interactions are primarily results of the pharmacodynamic effects of the ESAs. One example is the effect of ESA on the hemoglobin A1C level.[13] Hemoglobin A1C is a measure of the amount of glycosylated hemoglobin in the blood and is a commonly accepted indicator of the long-term glycemic control in diabetic patients.[14,15] But, its level depends on the number of circulating erythrocytes and the duration of red blood cells in circulation because glucose binding to hemoglobin occurs continuously during the life span of a red blood cell.[16] The hemoglobin A1C level decreases on initiation of ESAs and increases on discontinuation of ESA treatment; however, the glucose levels were reported to be stable. These data indicate that caution should be applied in the use of hemoglobin A1C level in patients who are initiating ESA treatment or in those whose ESA doses are being adjusted. Another example occurred in a German multicenter rhEPO trial,[17] where stroke patients receiving rhEPO after tissue plasminogen activator (tPA)-induced thrombolysis appeared to have increased mortality. A subsequent investigation in the mouse cerebral artery occlusion model suggested interaction of t-PA with EPO at the blood–brain interface–that is, promotion of vascular permeability and extracellular matrix breakdown may account for the unfavorable observation. As such, the authors concluded that after t-PA-induced thrombolysis, the use of EPO may not be suitable.[17] This hypothesis will need to be verified with additional studies.

### 14.2.2 Granulocyte Colony Stimulating Factor

Filgrastim, pegfilgrastim, and lenograstim are three commercially available G-CSF products. Filgrastim is a nonglycosylated recombinant human G-CSF (rhG-CSF) produced by *E. coli* containing 175 amino acids (MW 19 kDa). Pegfilgrastim is the pegylated filgrastim (MW 39 kDa), with a 20 kDa mono-methoxy-polyethylene glycol molecule covalently bound to the *N*-terminal methionyl residue of filgrastim. Lenograstim, a glycosylated rhG-CSF product, produced by CHO cells with 174 amino acids (MW 19 kDa), is approved globally outside the United States. Barring minor differences, these products are indicated for the treatment of neutropenia caused by chemotherapy. Except for pegfilgrastim, they are also approved for the mobilization of the peripheral blood progenitor cells.

Drug interactions were called out in the U.S. PI and EU SPC for filgrastim (with lithium and 5-FU) and pegfilgrastim (with 5-FU). Lithium products are oral medications for the treatment of psychiatric patients with bipolar disorder. Lithium has broad effects on the hematopoietic system and is known to increase neutrophil counts from the research of the 1980s. Mechanistically, lithium dose-dependently reduces the levels of surface CXCR4 protein and mRNA in neutrophils.[18] Because CXCR4 protein plays a critical role in the retention of hematopoietic cells within the bone marrow and in their mobilization into the circulation, a reduction in the CXCR4 is associated with an increased release of neutrophils into the circulation. In addition, lithium induces the production of cytokines, including G-CSF, which contributes to the increase in neutrophils.[19] Preliminary evidence from a small number of patients treated concomitantly with filgrastim and 5-fluorouracil indicates

that the severity of neutropenia may be exacerbated, a phenomenon believed to be associated with an increase in the topoisomerase sensitivity in the presence of G-CSF.[20] Furthermore, the product labels indicate that the use of filgrastim or lenograstim is not recommended from 24 h before until 24 h after chemotherapy ends, in view of the sensitivity of rapidly dividing myeloid cells to cytotoxic chemotherapy.

### 14.2.3 Keratinocyte Growth Factor

Marketed globally, palifermin is a truncated, nonglycosylated form of endogenous human keratinocyte growth factor (KGF) with the first 23 *N*-terminal amino acids deleted to improve protein stability. This recombinant KGF contains 140 amino acids (MW 16.3 kDa). It is a mucocutaneous epithelial growth factor and indicated for treatment of oral mucositis in chemotherapy setting. Palifermin has been shown to bind to heparin *in vitro* because KGF is a heparin-binding member of the fibroblast growth factor family,[21–23] and heparin has been shown to be involved in modulating the KGF binding to its receptor.[21] The potential of palifermin binding to low molecular weight heparin was a concern because heparin is used in the indicated clinical setting. As such, its U.S. PI advises against coadministration with heparin and advises to rinse heparin-containing IV lines with saline before administering palifermin into the line. The FDA's approval letter (December 15, 2004) listed three postmarketing-required studies related to drug interactions between palifermin and heparin: an *in vivo* study in healthy subjects to evaluate the drug interaction of palifermin with heparin, an *in vitro* study to evaluate the drug interaction of palifermin with low molecular weight heparins, and an *in vivo* study in healthy subjects, contingent on the results of the *in vitro* study, to evaluate the drug interaction of palifermin with low molecular weight heparins.

Two clinical studies evaluating the interaction between palifermin and heparin were posted on the ClinicalTrials.gov website, presumably to fulfill the postmarketing requirements. The first one was a single-dose study[24] in 27 healthy subjects receiving palifermin (60 μg/kg) coadministered with and without therapeutic levels of un-fractionated heparin. The primary objective was to evaluate the effect of heparin on the PK of palifermin and the secondary objective was to evaluate the effect (activated partial thromboplastin time, aPTT) of a single dose of palifermin on unfractionated heparin pharmacodynamics (PD) based on the PD measures of $AUCaPTT_{0–6}$ and $AUCaPTT_{0–24}$. As reflected in the drug interactions section of the PI, coadministration of heparin resulted in a fivefold increase in AUC and an 80% decrease in the mean clearance rate of palifermin. This case of drug interaction is one of the early examples to challenge the general perception that drug interactions were not of concern when dealing with biologics products. Although the PD results from this same study showed no significant effect of palifermin on heparin activity with respect to aPTT, the clinical impact of the increased palifermin exposure remains to be investigated.

The second study,[25] presumably conducted to further explore the clinical relevance of the observed PK interaction for the palifermin PD effect, was designed to

evaluate the impact of a continuous intravenous (IV) infusion of unfractionated heparin on the multiple-dose pharmacodynamics of palifermin. Forty-four healthy adult subjects were enrolled and randomized into three treatment groups in a ratio of 20 : 15 : 8 receiving either (1) daily palifermin dose of 40 μg/kg for three consecutive days as IV bolus injections and continuous heparin IV infusion, (2) daily palifermin dose of 40 μg/kg for three consecutive days as IV bolus injections, or (3) no treatment (a control group). The PK of palifermin and its PD response were measured, with PD defined as the ratio to baseline of epithelial cell proliferation as assessed by Ki67 staining of buccal mucosal tissue on study day 4. Consistent with the first study, the palifermin exposure was fivefold higher when coadministered with heparin. The PD response was similar with and without heparin, as the geometric mean ratio was 0.863 and the 90% confidence interval was within the predefined interval of 0.7 to 1.43, an equivalence criteria.

Similar to the G-CSF labeling, it is not recommended to administer palifermin within 24 h before or after chemotherapy. Because of the likelihood of coadministration with G-CSF, preclinical studies in murine and nonhuman primate chemotherapy/radiotherapy models explored the possibility of interaction between G-CSF and palifermin and found no evidence of any drug interaction.

### 14.2.4 Thrombopoiesis-Stimulating Agents

Two recombinant thrombopoiesis-stimulating agents (TSAs) are marketed: recombinant human interleukin 11 (rhIL-11) and romiplostim, a thrombopoietin (TPO) mimetic Fc-fusion protein. These recombinant products are indicated for the treatment of thrombocytopenia as their endogenous counterparts play a key role in platelet production.

Romiplostim, produced by recombinant DNA technology in *E. coli*, is a nonglycosylated Fc-peptide fusion protein (peptibody) that has a specific function to increase platelet production via binding to the TPO receptor,[26,27] despite the lack of amino acid sequence homology to the endogenous TPO. By covalent binding of the TPO mimetic peptides to the human immunoglobulin IgG1 Fc domain, this second-generation TSA has prolonged elimination half-life compared to the first-generation product, which failed to reach the market because of the immunogenicity issue.[26,28] No formal drug interaction studies were conducted in the development program, apart from the expected pharmacodynamic interaction with other medications used to treat thrombocytopenia, such as corticosteroids, danazol, azathioprine, intravenous immunoglobulin (IVIG), and anti-D immunoglobulin.

Oprelvekin, a rhIL-11 produced in *E. coli*, is a nonglycosylated protein containing 177 amino acids (MW 19 kDa). It differs from native IL-11 (178 amino acids) in lacking the amino-terminal proline residue. This alteration has not resulted in measurable differences in bioactivity either *in vitro* or *in vivo*. The endogenous IL-11 binds to IL-11 receptors and exerts pleiotropic activities, including stimulation of hemopoiesis, thrombopoiesis, and many nonhematopoietic activities, such as anti-inflammation. In terms of the potential association with the liver metabolic enzymes, IL-11 has been shown to be involved in modulating the transcriptional expression of aromatase in several tissues[29,30] and has minor effect on other hepatic

metabolic enzymes, as shown in the *in vitro* hepatocyte studies.[31] Aromatase is a member of the cytochrome P450 superfamily, which aromatize androgens to form estrogens.[32] Inhibitors of aromatase are clinically used as adjuvants in breast cancer treatment.[32,33] No formal drug interaction studies have been conducted for oprelvekin. Further evaluation of drug interactions related to IL-11 modulation could become important because IL-11 produced by inflammatory cells in the intestine has been implicated in the pathophysiology of cancer, and IL-11 has been proposed as a new target for anticancer therapeutics.[34]

## 14.3 HORMONES

### 14.3.1 Human Growth Hormone

Somatropin is the active ingredient of more than 10 recombinant human growth hormone (rhGH) products approved by the FDA and at least three approved by the EMA. The amino acid sequence of somatropin is identical to that of human growth hormone of pituitary origin, which is a single-chain protein of 191 amino acids (MW 20 kDa), including four cysteine residues present as two intrachain disulfides. They are produced in *E. coli* cells or *Saccharomyces cerevisiae* or mammalian cell line by recombinant DNA technology. Somatropin is indicated for the treatment of growth hormone deficiency, growth failures or short stature due to various causes, and wasting and cachexia secondary to HIV.

Growth hormone has been shown to modulate the cytochrome P450 gene expression and activity in primary culture of human hepatocytes, animal models, and humans.[35–38] It affects the PK of small molecule drugs by which the clearance of some drugs is increased, such as antipyrine and theophylline, and for others decreased, such as amobarbital. Much research is still ongoing to elucidate the precise mechanism of the regulatory effect of growth hormone on various CYP450 enzymes.[39] Consequently the labeling of growth hormone products indicates that monitoring is advisable when it is administered in combination with medicinal products known to be metabolized by CYP450 liver enzymes, such as corticosteroids, sex steroids, anticonvulsants, and cyclosporine. The labeling also advises that patients taking insulin for diabetes mellitus should be carefully monitored during treatment with somatropin because an adjustment of the insulin dose may be required.

On the other hand, the treatment effects of growth hormone can be modulated by the coadministered drugs. Patients with coexisting adrenocorticotropic hormone (ACTH) deficiency should have their glucocorticoid replacement dose carefully adjusted to avoid an inhibitory effect on growth. For women taking oral estrogens, a higher dose of growth hormone may be required to achieve the treatment goal.

***14.3.1.1 Interaction with Glucocorticoid Therapy*** Glucocorticoids may cause growth retardation and reduce the effect of exogenously administered rhGH in GH-deficient patients that appears to be mediated by multiple mechanisms.[40] In a study conducted with seven children who were on prednisone therapy, rhGH was

given to counteract the impaired linear growth. Based on the change in growth velocity, six of seven children responded to rhGH. One patient who received high doses of prednisone (>0.5 mg/kg per day) throughout the rhGH therapy did not experience an increased growth velocity. The response to rhGH was found to be related to the prednisone dose; no increase in growth velocity at prednisone doses >0.35 mg/kg per day was noted, and the response increased with decreased prednisone dose.[41] Downregulation of the expression of GH receptors and the receptor activities, alteration of the GH binding protein, and regulation of the local IGF-1 secretion and activities are a few of the possible mechanisms reported.[40]

Conversely, treatment effects of glucocorticoids are affected by growth hormone because it inhibits 11β-hydroxysteroid dehydrogenase type 1 (11β-HSD1), an enzyme that converts inactive cortisone to active cortisol.[42] In patients with GH deficiency, 11β-HSD1 activity is enhanced, which is subsequently normalized after they receive rhGH treatment.[42,43] Hence the U.S. PI of GH products indicate that GH-deficient patients being treated with rhGH may require initiation of glucocorticoid therapy and patients treated with glucocorticoid replacement for previously diagnosed hypoadrenalism may require an increase in their maintenance doses when rhGH therapy is started.

***14.3.1.2 Interaction with Amobarbital*** In a clinical study in children amobarbital was used as a probe to detect hepatic cytochrome P450 enzyme-mediated activity because it is eliminated almost completely through oxidation via hepatic cytochrome P450 mixed-function oxidases. Amobarbital PK after single doses at 3 to 5 mg/kg was compared before and after 6 weeks of rhGH treatment (~2 mg dosed intramuscularly 3 times a week) in four GH-deficient children. The results showed that plasma clearance of amobarbital decreased by 50% after rhGH treatment.[44]

***14.3.1.3 Interaction with Antipyrine*** In a clinical study with GH-deficient adults, the antipyrine clearance test was used as a marker for hepatic cytochrome P450 activity. Antipyrine (10 mg/kg) was given orally before and after 3 months of rhGH therapy in which daily SC doses of rhGH were given at 0.125 IU/kg/week for 1 month followed by daily SC doses at 0.25 IU/kg/week. Median antipyrine clearance increased significantly (by 52%) and antipyrine clearance increased for all subjects except for one, who was a smoker with high baseline clearance.[45]

In another clinical study with 12 GH-deficient children, antipyrine clearance was used to detect short-term and long-term rhGH effect on enzyme activity. Antipyrine (18 mg/kg oral) was given before and after either daily rhGH dose at 1U/day for 5–7 days (short-term effect) or 3 times per week treatment of rhGH at 2U/dose for 6 weeks to 12 months (long-term effect) via intramuscular injection. In both cases, the mean clearance did not change statistically relative to baseline.[46]

***14.3.1.4 Interaction with Theophylline*** In two clinical studies in GH-deficient children, the effect of rhGH on PK of theophylline was assessed.[47] In both studies, 4 mg/kg theophylline was administered (either orally or IV as

aminophylline) before and after 2 IU rhGH given intramuscularly 3 times a week for 3–5 weeks (first study) or 3 weeks (second study). In the first study in four children, the elimination half-life of theophylline decreased by 55% relative to pretreatment (clearance data not reported). In the second study in eight children, even though the half-life did not change, the apparent clearance of theophylline increased by 28% relative to pretreatment.

***14.3.1.5 Interaction with Oral Estrogens*** The effect of oral estrogen therapy on rhGH treatment was demonstrated in eight hypopituitary GH-deficient women with hypogonadism.[48] Daily doses of 2 mg oral estradiol valerate were administered to all subjects for 8 weeks (over 2 estrogen cycles of 4 weeks each). Starting from the second estrogen cycle, subjects received stepwise increase of rhGH daily dose from 0.5 IU in week 1, to 1 IU in week 2, to 2 IU in week 3. The level of insulin-like growth factor 1 (IGF-1) was measured at baseline before estrogen treatment, during the estrogen-alone treatment phase, and during rhGH treatment at the end of weeks 1, 2, and 3 to evaluate the dose-response relationship of rhGH treatment. Orally administered estrogen reduced the IGF-1 level to below the baseline, which was followed by a weekly stepwise increase in IGF-1 levels owing to increasing rhGH dose, indicating a dose response for rhGH treatment. A daily rhGH dose of 1 IU restored the mean IGF-1 level to approximately the baseline level. Therefore, patients who are on oral estrogen therapy will require a higher dose of rhGH for the desired treatment effect. It is important to note that the same study suggested that the antagonistic effect was dose route dependent, because estrogen administered via transdermal patch did not cause a decrease in IGF-1 levels. The differential effect of estrogen from two routes of administration was also demonstrated using other metabolic endpoints for lipid metabolism and protein metabolism. Although the mechanism of this interaction between estrogen and growth hormone is unclear, estrogen has been implicated in modulating the metabolic action of GH *in vitro* as well as in patients.[49,50]

***14.3.1.6 Interaction with Thyroxin*** In a clinical study the effect of rhGH on biomarkers related to thyroxin ($T_4$) therapy (owing to central hypothyroidism) was studied in eight GH-deficient adults.[51] These patients were on stable $T_4$ substitution (0.1–0.2 mg daily dose) for at least 1 year and were studied over four consecutive 4-week periods receiving no rhGH therapy, followed by daily SC doses of 1 IU/m$^2$, then 2 IU/m$^2$, and finally 4 IU/m$^2$ rhGH. At the end of each period, the thyroid hormone levels were measured.

Serum total and free $T_4$ levels decreased with increasing doses of rhGH; however, the decrease was not statistically significant in both cases. Serum total and free triiodothyronine ($T_3$) levels, the active thyroid hormone derived from $T_4$ by deiodinase, increased with increasing doses of rhGH and were significantly greater than no rhGH therapy. The $T_3/T_4$ ratio exhibited a significant dose-dependent increase, and the reverse $T_3$, an inactive isomer of $T_3$ that binds to the thyroid hormone receptor, showed a significant dose-dependent decrease relative to no rhGH dose. In the product label for the GH product Omnitrope it is speculated that in

patients receiving replacement therapy with thyroxin mild hyperthyroidism may occur and the label advises to test thyroid function after starting treatment with rhGH and after dose adjustments.

***14.3.1.7 Interaction with Insulin*** The effect of rhGH on insulin dose-response was studied in six normal male subjects in a crossover design[52] with a 72-h washout in between two treatment occasions in which subjects were started on either 0.45% saline or GH (2 μg/kg/h) 12 h before insulin dose-response studies. During each occasion, subjects were administered insulin doses of 0.4, 1.0, 2.0, and 10.0 mU/kg/min sequentially for 2 h, and each insulin dose was given simultaneously with the infusion of glucose as 50% solution to maintain arterial glucose concentrations at 95 mg/dL. The glucose production and utilization were measured before the first insulin infusion and during the last 40 min of each subsequent insulin infusion. The effect of rhGH on insulin receptor function was also determined before the first insulin infusion by measuring the proportion of monocytes in the mononuclear cell preparation and the insulin binding to the monocytes.

In the presence of rhGH, both dose-response curves of insulin vs. glucose production and insulin vs. glucose utilization shifted to the right—that is, the half-maximal response was significantly increased in the presence of rhGH without affecting the maximum response—suggesting that rhGH increases insulin resistance. Treatment with rhGH did not result in differences in the binding of insulin to the receptors on monocytes. Therefore, the insulin resistance effect was postulated to be the result of a postreceptor defect.

### 14.3.2 Pegvisomant

Pegvisomant (Somavert) is an analog of GH that has been structurally altered to act as a GH receptor antagonist and is indicated for treatment of acromegaly. Pegvisomant is produced in *E. coli* and contains 191 amino acid residues to which several PEG polymers are covalently bound (mostly 4–6 PEG/protein molecule). The predominant molecular weights of pegvisomant are approximately 42, 47, and 52 kDa, with the protein portion being 22 kDa and the PEG being ~5 kDa per unit.

Per the U.S. PI, the treatment regimen for blood sugar control for acromegalic patients with diabetes mellitus who are being treated with insulin and/or oral hypoglycemic agents may require dose reductions after the initiation of therapy with pegvisomant. This is consistent with the report that growth hormone can increase the insulin resistance, as stated in the previous sections.

The effect of growth hormone blockade on insulin sensitivity was studied by Williams *et al.* in seven young adults with type 1 diabetes who were on a combination of short- and long-acting insulins two to four times per day.[53] In this crossover study subjects were given two treatments, either 5 or 10 mg daily SC pegvisomant, over a 3-week period after a run-in period and a 3-week washout in between treatments. At baseline and after each treatment, variable rate insulin infusion was administered overnight to maintain blood glucose at 5 mmol/L. In the morning, this was followed by a two-step hyperinsulinemic euglycemic clamp, using [6,6 $^2H_2$]

glucose and $^2H_5$ glycerol, to measure glucose and glycerol turnover. The investigators showed that overnight insulin requirements significantly decreased for both treatments when compared to baseline. Another study suggested that the decrease in overnight free fatty acid led to an improvement in insulin sensitivity.[54]

The second drug interaction in the U.S. PI stated that in clinical studies patients on opioids often needed higher serum pegvisomant concentrations to achieve appropriate IGF-I suppression compared to patients not receiving opioids. In our literature review, we did not locate any clinical reports of formal drug interaction studies; however, we came across reports of complex interactions between the opioid system and somatostatin, which appears to be an active subject of current research. Morphine and opioid antagonists have been shown to regulate cell growth *in vitro* by interacting with the somatostatin receptors;[55–57] opioids with proliferative effect and opioid antagonists with antiproliferative effect. Analgesic effect associated with octreotide, a peptide analog of somatostatin, administration was reported for a patients treated for acromegaly.[58] The involvement of opioids in regulating the functions and/or secretion of growth hormone has been suggested in clinical setting.[59,60] These research reports were consistent with the advice in the product labeling.

### 14.3.3 Glucagon

Multiple recombinant glucagon products are available commercially in the United States and one is available in the EU. Recombinant glucagon, a single-chain polypeptide hormone with 29 amino acids (MW 3.5 kDa), is synthesized in *E. coli* and is identical to human glucagon produced by the pancreas. Indicated for the treatment of hypoglycemia and also used in diagnostic testing of the stomach and other digestive organs, it increases blood glucose and relaxes smooth muscle of the gastrointestinal tract.

The labeling of glucagon products refer to interactions with various drugs, and most of them are pharmacology-based: Potential interaction with β-blockers might lead to a greater increase in both pulse and blood pressure, which may require therapy in patients with pheochromocytoma or coronary artery disease. When used with indomethacin, glucagon may lose its ability to raise blood glucose or may even produce hypoglycemia. Coadministration with an anticholinergic drug is not recommended owing to increased gastrointestinal side effects. Glucagon may increase the anticoagulant effect of warfarin. Insulin reacts antagonistically toward glucagon. Summarized in the following sections by product are the results of our literature search, which are consistent with the product labeling.

***14.3.3.1 Interaction with Indomethacin and Prostaglandins*** Some prostaglandins have been demonstrated to evoke the release of glucagon from the α-cells in the pancreas.[61] On the other hand, inhibition of the synthesis of prostaglandins by cyclooxygenase inhibitors, such as indomethacin and acetyl salicylic acids, can inhibit the release of glucagon. The effects of indomethacin on glucose levels and on insulin levels were demonstrated in anesthetized male Wistar rats.[62]

The glucose and insulin levels were monitored over 2 h in rats that were administered (1) glucagon only as IV infusion for 120 min, (2) a pretreatment of indomethacin at 10 mg/kg administered intraperitoneally at 15 and 3 h before the infusions of glucagon, or (3) a pretreatment of indomethacin followed by infusions of saline. The basal glucose levels were similar across the three groups. Indomethacin alone does not affect glucose levels or insulin levels over 2 h. In response to glucagon treatment alone, the glucose level and the insulin level increased and remained elevated over 90 min. On the other hand, glucagon treatment in rats pretreated with indomethacin raised the glucose level and the insulin level above baseline for a short time and then declined to below the basal level after 30 min; peak glucose and insulin levels were lower than in rats not administered indomethacin.

In another set of experiments assessing the effect of indomethacin on hepatic glucose output, 2-$^{3}$H-glucose was mixed with normal glucose and infused intravenously for 210 min. At 90 min, which is after establishing steady-state, glucagon was intravenously infused for 120 min either alone or with indomethacin. Results showed that indomethacin reduced both glucose production and glucose utilization.

The same research group conducted a study in conscious female dogs with similar design to evaluate the effect of indomethacin or acetyl salicylic acid on glucagon's ability to produce glucose. Pretreatment with indomethacin and salicylate did not alter basal glucose levels but reduced the effect of glucagon in terms of the increase in glucose levels and the glucose production and utilization.[63]

***14.3.3.2 Interaction with Warfarin*** During 9 months of a study in Massachusetts General Hospital, 469 patients on warfarin were monitored to identify drugs that modify the hypoprothrombinemic effect of warfarin.[64] Twenty-four subjects were on glucagon concomitant therapy. Of these, 13 subjects were on glucagon for at least 2 days with a total dose >50 mg of glucagon. In 8 of these, the hypoprothrombinemic effect of warfarin was enhanced, indicating that they became markedly more sensitive to warfarin during glucagon treatment. They had much prolonged prothrombin time and experienced bleeding while on both warfarin and glucagon. The authors postulated that glucagon may potentiate the anticoagulant action of warfarin by acting synergistically to depress hepatic synthesis of vitamin-K-sensitive clotting proteins or by increasing the affinity of warfarin for its receptor sites.

***14.3.3.3 Interaction with Aminophylline*** Glucagon has myocardial inotropic action in experimental animals and in men. In a study conducted in anesthetized dogs the effect of aminophylline and glucagon when coadministered on coronary and systemic circulation was compared to either agent given alone.[65] Comparisons were made on baseline and posttreatment values of multiple coronary and systemic hemodynamic endpoints, including heart rate, coronary sinus blood flow, cardiac output, and pressures. Treatment with aminophylline alone or glucagon alone was delivered as a constant rate infusion at two different infusion rates, one being half of the other. The combination treatment with aminophylline and glucagon were infused at the lower of two infusion rates for each agent. In total, 21 repeated runs

of the treatment procedures were made in 10 dogs, and summary statistics were determined for various measurements. The average doses used were 2.7 and 5.5 mg/min for aminophylline and 21 and 42 μg/min for glucagon.

When compared to the treatment effect of glucagon alone at the same dose rate, aminophylline enhanced the cardiovascular effects of glucagon. Furthermore, the effects of the glucagon and aminophylline combined treatment were greater than the glucagon treatment effect as a single agent at twice the dose. This interaction was thought to be the result of an accumulation of cyclic 3′,5′-AMP because glucagon activates adenyl cyclase and aminophylline inhibits 3′,5′ nucleotide phosphodiesterase.

## 14.4 CONCLUSIONS

Our literature research had limited success in finding drug interactions that involve the discussed growth factor products, whereas information is more readily available for the covered hormone products. We speculate that the target receptor specificity of these growth factors in combination with the specificity of their postreceptor functions may play a role in the limited findings of drug interactions with these products. The observed interactions are mostly consequences of pharmacological effect of either the perpetrator drug or the victim drug. On the other hand, the hormone products we discussed generally have much broader involvement in biological/physiological functions. They can affect not only the cytochrome P450 enzyme system but also other metabolic functions and enzyme systems in addition to pharmacological interactions mediated at the receptor levels. Furthermore, the complexity of the network of signaling pathways for hormones and the intricate interconnectivities with other complex biochemical signaling processes present opportunities for pharmacodynamic interactions. This can explain the higher volume of available drug interaction information in the literature for hormones as compared to growth factors.

Palifermin is the only case of a PK drug interaction we found in which heparin causes an increase in the palifermin systemic exposure; however, the clinical significance of this interaction is yet to be elucidated. Several cases of PK interactions were reported for recombinant hormone products. Many more cases of pharmacodynamic interactions were observed for both recombinant growth factors and hormones. These reported interactions have clear impact on the clinical use of the products in that they are included in the product labeling whether they are caused by PK alteration or PD modulation.

With the advent of the advanced tools for research at the molecular and cellular level we also expect that the empirically identified drug interactions will gradually find the mechanistic explanations. For instance, lithium's effect on neutrophil count was an empirical finding that was called out in the filgrastim label as a potential interaction. This was later mechanistically attributed to its interaction with the chemokine receptor CXCR4 a few years after the initial approval of filgrastim. As growth factor product pipelines in pharmaceutical development rapidly expand, it is

very likely that drug interactions may become more common for new growth factors that reach the market. We expect that new reports of drug interaction on currently marketed products will be forthcoming as more and more patients take multiple medications given the increasing average age of the population. In summary, our research indicates that investigations on drug interactions for growth factors and hormone products are important for clinical use of these medications and that this is an area for continued investment by scientists in academia, industry, and regulatory agencies.

## REFERENCES


1. Aggarwal S. What's fueling the biotech engine—2009–2010. *Nat Biotechnol* **28**, 1165–71 (2010).
2. American Society of Clinical Oncology. Erythropoiesis-stimulating agents: continued challenges. *J Oncol Pract* **3**, 248–253 (2007).
3. Macdougall IC, Padhi D, Jang G. Pharmacology of darbepoetin alfa. *Nephrol Dial Transplant* **22** (Suppl 4), iv2–iv9 (2007).
4. Molineux G. Pegfilgrastim: using pegylation technology to improve neutropenia support in cancer patients. *Anticancer Drugs* **14**, 259–64 (2003).
5. Renwick W, Pettengell R, Green M. Use of filgrastim and pegfilgrastim to support delivery of chemotherapy: twenty years of clinical experience. *BioDrugs* **23**, 175–86 (2009).
6. Molineux G. The design and development of pegfilgrastim (PEG-rmetHuG-CSF, Neulasta). *Cur Pharmaceut Design* **10**, 1235–44 (2004).
7. Macdougall IC. CERA (Continuous Erythropoietin Receptor Activator): a new erythropoiesis-stimulating agent for the treatment of anemia. *Curr Hematol Rep* **4**, 436–40 (2005).
8. Topf JM. CERA: third-generation erythropoiesis-stimulating agent. *Expert Opin Pharmacother* **9**, 839–49 (2008).
9. Seige K. [Glucagon, its discovery and description and the work of Max Burger.] *Z Gesamte Inn Med* **41**, 568–71 (1986).
10. Somatropin. Human growth hormone history. Available at www.somatropin.net/hgh-history.htm.
11. Lombardero M, Kovacs K, Scheithauer BW. Erythropoietin: a hormone with multiple functions. *Pathobiology* **78**, 41–53 (2011).
12. Chateauvieux S, Grigorakaki C, Morceau F, et al. Erythropoietin, erythropoiesis and beyond. *Bioch Pharmacol* **81**, 1291–303 (2011).
13. Brown JN, Kemp DW, Brice KR. Class effect of erythropoietin therapy on hemoglobin A (1c) in a patient with diabetes mellitus and chronic kidney disease not undergoing hemodialysis. *Pharmacotherapy* **29**, 468–72 (2009).
14. Okada T, Nakao T, Matsumoto H, et al. Influence of erythropoietin-stimulating agent treatment on glycated hemoglobin values in diabetic patients with chronic kidney disease. *Clin Exper Nephrol* **15**, 316–17 (2011).
15. Joy MS, Cefalu WT, Hogan SL, Nachman PH. Long-term glycemic control measurements in diabetic patients receiving hemodialysis. *Am J Kidney Dis* **39**, 297–307 (2002).

16. Inaba M, Okuno S, Kumeda Y, et al. Glycated albumin is a better glycemic indicator than glycated hemoglobin values in hemodialysis patients with diabetes: effect of anemia and erythropoietin injection. *J Am Soc Nephrol* **18**, 896–903 (2007).
17. Zechariah A, ElAli A, Hermann DM. Combination of tissue-plasminogen activator with erythropoietin induces blood-brain barrier permeability, extracellular matrix disaggregation, and DNA fragmentation after focal cerebral ischemia in mice. *Stroke* **41**, 1008–12 (2010).
18. Kim HK, Kim JE, Chung J, P et al. Lithium down-regulates the expression of CXCR4 in human neutrophils. *J Trace Elements Med Biol* **21**, 204–09 (2007).
19. Hager ED, Dziambor H, Hohmann D, et al. Effects of lithium on thrombopoiesis in patients with low platelet cell counts following chemotherapy or radiotherapy. *Biolo Trace Element Res* **83**, 139–48 (2001).
20. Kreuser ED, Wadler S, Thiel E. Biochemical modulation of cytotoxic drugs by cytokines: molecular mechanisms in experimental oncology. *Recent Results Cancer Res* **139**, 371–82 (1995).
21. Berman B, Ostrovsky O, Shlissel M, et al. Similarities and differences between the effects of heparin and glypican-1 on the bioactivity of acidic fibroblast growth factor and the keratinocyte growth factor. *J Biol Chem* **274**, 36132–38 (1999).
22. MacDonald KP, Hill GR. Keratinocyte growth factor (KGF) in hematology and oncology. *Curr Pharm Des* **8**, 395–403 (2002).
23. Rubin JS, Osada H, Finch PW, et al. Purification and characterization of a newly identified growth factor specific for epithelial cells. *Proc Nat Acad Sci U S A* **86**, 802–06 (1989).
24. Clinical Development, Biovitrum AB., An open-label, randomized, 2-part, parallel design study to characterize the effect of heparin on palifermin pharmacokinetics and the effect of palifermin on heparin pharmacodynamics in healthy subjects. NCT00361348, February 27, 2009. Available at http://clinicaltrials.gov/ct2/show/NCT00361348?term=palifermin+heparin&rank=2.
25. Swedish Orphan Biovitrum. An open-label, randomized, parallel-design study to characterize the effect of heparin on palifermin activity in healthy adult subjects. NCT01163097, November 23, 2012. Available at http://clinicaltrials.gov/ct2/show/study/NCT01163097?term=palifermin+heparin&rank=1.
26. Wang YM, Krzyzanski W, Doshi S, et al. Pharmacodynamics-mediated drug disposition (PDMDD) and precursor pool lifespan model for single dose of romiplostim in healthy subjects. *AAPS J* **12**, 729–40 (2010).
27. Wei P. Thrombopoietin factors. *Cancer Treatment Res* **157**, 75–93 (2011).
28. Neumann TA, Foote M. Megakaryocyte growth and development factor (MGDF): an Mpl ligand and cytokine that regulates thrombopoiesis. *Cytokines Cell Molec Ther* **6**, 47–56 (2000).
29. Simpson ER, Michael MD, Agarwal VR, H. et al. Cytochromes P450 11: expression of the CYP19 (aromatase) gene: an unusual case of alternative promoter usage. *FASEB J* **11**, 29–36 (1997).
30. Zhao Y, Agarwal VR, Mendelson CR, Simpson ER. Transcriptional regulation of CYP19 gene (aromatase) expression in adipose stromal cells in primary culture. *J Steroid Biochem Molec Biol* **61**, 203–10 (1997).
31. Guillen MI, Donato MT, Jover R, et al. Oncostatin M down-regulates basal and induced cytochromes P450 in human hepatocytes. *J Pharmacol Exper Ther* **285**, 127–34 (1998).

32. Brodie A. Aromatase and its inhibitors—an overview. *J Steroid Biochem Mol Biol* **40**, 255–261 (1991).
33. Bickenbach KA, Jaskowiak N. Aromatase inhibitors: an overview for surgeons. *J Am Coll Surgeons* **203**, 376–89 (2006).
34. Putoczki T, Ernst M. More than a sidekick: the IL-6 family cytokine IL-11 links inflammation to cancer. *J Leukocyte Biol* **88**, 1109–17 (2010).
35. Liddle C, Goodwin BJ, George J, *et al*. Separate and interactive regulation of cytochrome P450 3A4 by triiodothyronine, dexamethasone, and growth hormone in cultured hepatocytes. *J Clin Endocrinol Metab* **83**, 2411–16 (1998).
36. Jarukamjorn K, Sakuma T, Jaruchotikamol A, et al. Modified expression of cytochrome P450 mRNAs by growth hormone in mouse liver. *Toxicology* **219**, 97–105 (2006).
37. Thangavel C, Dworakowski W, Shapiro BH. Inducibility of male-specific isoforms of cytochrome p450 by sex-dependent growth hormone profiles in hepatocyte cultures from male but not female rats. *Drug Metab Dispos* **34**, 410–19 (2006).
38. Jurgens G, Lange KH, Reuther LO, et al. Effect of growth hormone on hepatic cytochrome P450 activity in healthy elderly men. *Clin Pharmacol Ther* **71**, 162–68 (2002).
39. Kim SK, Novak RF. The role of intracellular signaling in insulin-mediated regulation of drug metabolizing enzyme gene and protein expression. *Pharmacol Ther* **113**, 88–120 (2007).
40. Mushtaq T, Ahmed SF. The impact of corticosteroids on growth and bone health. *Arch Dis Childhood* **87**, 93–96 (2002).
41. Rivkees SA, Danon M, Herrin J. Prednisone dose limitation of growth hormone treatment of steroid-induced growth failure. *J Pediatr* **125**, 322–25 (1994).
42. Moore JS, Monson JP, Kaltsas G, et al. Modulation of 11beta-hydroxysteroid dehydrogenase isozymes by growth hormone and insulin-like growth factor: in vivo and in vitro studies. *J Clin Endocrinol Metab* **84**, 4172–77 (1999).
43. Swords FM, Carroll PV, Kisalu J, et al. The effects of growth hormone deficiency and replacement on glucocorticoid exposure in hypopituitary patients on cortisone acetate and hydrocortisone replacement. *Clin Endocrinol* **59**, 613–20 (2003).
44. Redmond GP, Bell JJ, Perel JM. Effect of human growth hormone on amobarbital metabolism in children. *Clin Pharmacol Ther* **24**, 213–18 (1978).
45. Cheung NW, Liddle C, Coverdale S, et al. Growth hormone treatment increases cytochrome P450-mediated antipyrine clearance in man. *J Clin Endocrinol Metab* **81**, 1999–2001 (1996).
46. Rifkind AB, Saenger P, Levine LS., et al. Effects of growth hormone on antipyrine kinetics in children. *Clin Pharmacol Ther* **30**, 127–32 (1981).
47. Redmond GP. Regulation of theophylline metabolism by growth hormone. *Semin Perinatol* **5**, 374–77 (1981).
48. Wolthers T, Hoffman D M, Nugent AG, et al. Oral estrogen antagonizes the metabolic actions of growth hormone in growth hormone-deficient women. *Am J Physiol Endocrinol Metabol* **281**, E1191–96 (2001).
49. Lopez-Guerra A, Chirino R, Navarro D, et al. Estrogen antagonism on T3 and growth hormone control of the liver microsomal low-affinity glucocorticoid binding site (LAGS). *J Steroid Biochem Molec Biol* **63**, 219–28 (1997).

50. Wolthers T, Hoffman DM, Nugent AG, et al. Oral estrogen antagonizes the metabolic actions of growth hormone in growth hormone-deficient women. *Am J Physiol Endocrinol Metabol* **281**, E1191–96 (2001).
51. Jorgensen JO, Moller J, Laursen T, et al. Growth hormone administration stimulates energy expenditure and extrathyroidal conversion of thyroxine to triiodothyronine in a dose-dependent manner and suppresses circadian thyrotrophin levels: studies in GH-deficient adults. *Clin Endocrinol* **41**, 609–14 (1994).
52. Rizza R A, Mandarino LJ, Gerich J E. Effects of growth hormone on insulin action in man. Mechanisms of insulin resistance, impaired suppression of glucose production, and impaired stimulation of glucose utilization. *Diabetes* **31**, 663–69 (1982).
53. Williams RM, Amin R, Shojaee-Moradie F, et al. The effects of a specific growth hormone antagonist on overnight insulin requirements and insulin sensitivity in young adults with Type 1 diabetes mellitus. *Diabetologia* **46**, 1203–10 (2003).
54. Gormsen LC, Jessen N, Gjedsted J, Gjedde, et al. Dose-response effects of free fatty acids on glucose and lipid metabolism during somatostatin blockade of growth hormone and insulin in humans. *J Clin Endocrinol Metab* **92**, 1834–42 (2007).
55. Hatzoglou A, Kampa M, Castanas E. Opioid-somatostatin interactions in regulating cancer cell growth. *Frontiers Biosci* **10**, 244–56 (2005).
56. Notas G, Kampa M, Nifli AP, et al. The inhibitory effect of opioids on HepG2 cells is mediated via interaction with somatostatin receptors. *Eur JPharmacol* **555**, 1–7 (2007).
57. Hatzoglou A, Ouafik L, Bakogeorgou E, et al. Morphine cross-reacts with somatostatin receptor SSTR2 in the T47D human breast cancer cell line and decreases cell growth. *Cancer Res* **55**, 5632–36 (1995).
58. Pascual J, Freijanes J, Berciano J, Pesquera C. Analgesic effect of octreotide in headache associated with acromegaly is not mediated by opioid mechanisms. Case report. *Pain* **47**, 341–44 (1991).
59. De Marinis L, Mancini A, Valle D, et al. Physiological role of the opioid-cholinergic interaction in growth hormone neuroregulation: effect of sex and food intake. *Metabolism* **46**, 740–44 (1997).
60. Pende A, Musso NR, Montaldi ML, et al. Interaction between morphine, an opioid agonist, and clonidine, an alpha-adrenergic agonist, on the regulation of anterior pituitary hormone secretion in normal male subjects. *Biomed Pharmacother* **41**, 243–47 (1987).
61. Giugliano D, Torella R, D'Onofrio F. Prostaglandins and the alpha-cell. *Prostagland Med* **6**, 283–97 (1981).
62. Ganguli S, Sperling MA, Frame C, Christensen R. Inhibition of glucagon-induced hepatic glucose production by indomethacin. *Am J Physiol* **236**, E358–65 (1979).
63. Miller JD, Ganguli S, Sperling MA. Prostaglandin synthesis inhibitors impair hepatic glucose production in response to glucagon and epinephrine stimulation. *Diabetes* **32**, 439–44 (1983).
64. Koch-Weser J. Potentiation by glucagon of the hypoprothrombinemic action of warfarin. *Ann Intern Med* **72**, 331–35 (1970).
65. Afonso S, Hansing CE, Ansfield TJ, et al. Enhancement of cardiovascular effects of glucagon by aminophylline. *Cardiovasc Res* **6**, 235–39 (1972).

# DRUG–DRUG INTERACTIONS FOR NUCLEIC ACID-BASED DERIVATIVES

JIAN WANG and HONG ZHAO

## 15.1 INTRODUCTION

In the last decade, oligonucleotide-based therapeutics have been under active development for treatment of diseases such as cancers, genetic disorders, and or infections. Fomivirsen was the first antisense oligonucleotide drug approved by the FDA in 1998 for the treatment of cytomegalovirus retinitis in AIDS patients who have failed therapy with other anticytomegalovirus drugs or who are intolerant or have a contraindication to other therapies. Six years later, pegaptanib (pegylated aptamer), a selective vascular endothelial growth factor (VEGF) antagonist was approved for the treatment of neovascular (wet) age-related macular degeneration (AMD). Both drugs are administered via intravitreal injection.

Antisense oligonucleotides (ASOs) are perhaps the most direct therapeutic strategy to target RNA. Currently ASOs make up the major investigational new drug (IND) submissions for oligonucleotide drug development. These ASOs are designed to bind to the target RNA by well-characterized Watson-Crick base pairing, and once bound to the target RNA, modulate its function through a variety of postbinding events. In this chapter, the primary information used for summarizing drug–drug interactions (DDIs) of oligonucleotides comes from the first- and second-generation ASOs because the currently published DDI studies are mostly focused on these classes of oligonucleotides. When possible, information on other classes of oligonucleotides is presented in the context of these base classes. Caution should be taken in making generalized conclusions based on the limited number of reported studies due to the fact that most of the oligonucleotides are still in the development stage. More information needs to be accumulated to better define the DDI potential between oligonucleotides and other drugs.

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*,
Edited by Shayne C. Gad
Copyright © 2013 John Wiley & Sons, Inc.

This chapter will review the results from both preclinical and clinical DDI studies for the first- and second-generation ASOs. Phosphorothioate ASOs involve substituting sulfur for one of the two nonbridging oxygens in each internucleotide linkage, which are the most extensively studied oligonucleotide drugs and are referred to as first-generation ASOs. MOE-modified ASOs are the products with 2″-*O*-(2-methoxyethyl) (2″-*O*-MOE) ribose sugar modifications for the 4–5 nucleotides at each terminus of ASOs. This modification increases the resistance of ASOs to exonucleases and increases the binding affinity of ASOs to mRNA. The chimeric structure (oligonucleotide core surrounded by MOE-modified nucleotides) as second-generation ASOs preserves their capacity to serve as ribonuclease (RNase) H substrates. The clinical pharmacokinetics (PK) of the first-generation ASOs (single-stranded phosphorothioate modified DNA) and the second-generation ASOs (single-stranded phosphorothioate modified chimeric with partial 2″-*O*-MOE ribose sugar modifications) have been well characterized.

There are very limited reports available for alternative backbone modified oligonucleotides including the chemical classes of peptide nucleic acids (PNAs), locked nucleic acids (LNAs) and morpholino (neutral backbone) nucleic acids (MNAs). The small interfering RNA (siRNA), a double-stranded RNA motif containing both sense and antisense RNA strands, precludes its practical systemic delivery without chemical modification, conjugation or special formulation.[1]

## 15.2 CLINICAL PHARMACOKINETICS

The absorption, distribution, metabolism, and excretion of ASOs have been characterized in clinical trials. Their PK properties largely depend on their physical and chemical properties.

After parenteral administration, the first-generation phosphorothioate oligonucleotides are highly bound to plasma proteins through their hydrophilic sites. The protein binding is >90% for ASOs and >60% for siRNA. The binding of oligonucleotides to plasma proteins limits these compounds from glomerular filtration and urinary excretion. As the degree of protein binding decreases, the amount of free or unbound oligonucleotides increases, leading to an increase in urinary excretion. Oligonucleotides distribute rapidly from blood to tissues with a distribution half-life less than 2 h. They highly accumulate in tissues such as kidney, liver, spleen, lymph nodes, adipocytes, and bone marrow. Oligonucleotides do not cross the blood–brain barrier and poorly distribute to muscle, heart, and lung. Oligonucleotides are metabolized by nucleases ubiquitously expressed in tissue cells. The elimination half-life of oligonucleotides reflects the metabolism in tissues, the equilibration of drugs and metabolites between tissues and blood, and excretion by the kidneys.[2]

Chemical modifications of oligonucleotides result in marked difference in their PK profiles.[3] The first-generation ASOs modify the phosphodiester backbone of DNA to a phosphorothioate backbone, which enhances their hydrophobicity and protein binding and increases their nuclease resistance. The 2″-*O*-MOE modification at the 5″ and 3″ termini for the second-generation ASOs blocks the exonuclease

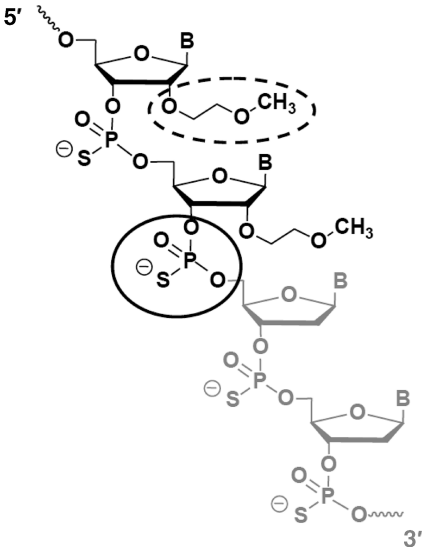


**FIGURE 15-1** Chemical modifications. Solid circle: phosphorothioate modification; dotted circle: 2′-methoxyethyl modification

cleavage and therefore they are primarily cleared by slow endonuclease digestion and ultimate excretion of parent drug and metabolites in urine. Figure 15-1 illustrates modification sites of oligonucleotides to produce the first- and second-generation ASOs.

While the terminal elimination half-lives for the first-generation ASOs are <40 h, the second-generation ASOs have longer half-lives up to a month, which allow for a less frequent (once weekly or longer interval) dosing regimen as compared to the first-generation ASOs. Many of the advantages associated with moving from the first-generation to second-generation ASOs are directly related to their improved PK profiles.[1]

The oligonucleotides that lack charge and/or have low binding to plasma proteins are prone to rapid clearance from plasma, with substantially high excretion in urine in the first day after administration, resulting in less overall tissue accumulation and ultimate low target bioavailability. This phenomenon is seen in the classes of PNAs, morpholinos, and unmodified siRNA. The PK properties of unmodified RNAi oligonucleotides preclude systemic delivery. Therefore chemical conjugation or formation approaches, such as pegylation or lipoidal delivery systems are often needed to improve the RNAi stability and distribution into target tissues.

## 15.3 DRUG–DRUG INTERACTIONS

### 15.3.1 *In Vitro* Studies

***15.3.1.1 DDI via Protein Binding Replacement*** Oligonucleotides are highly bound to plasma proteins, which protects them from urinary excretion and

subsequently increases their plasma exposure. Albumin has been identified as one of the major plasma proteins with high binding capacity to oligonucleotides. The binding of oligonucleotides to plasma proteins is relatively weak with low dissociation constant of 5–20 μmol/L, and the binding sites for this hydrophilic drug class differ from the binding sites for hydrophobic small molecular drugs. Therefore, it is speculated that the potential for PK interactions mediated by protein binding is expected to be low at clinically relevant concentrations. Several *in vitro* and clinical studies have been reported in the literature regarding the protein binding displacement for oligonucleotide-based drugs. Three examples are provided.

ISIS 113715 is a 20-mer first-generation ASO that is complementary to the protein tyrosine phosphatase 1B (PTP-1B) messenger RNA and subsequently reduces translation of the PTP-1B protein, a negative regulator of insulin receptor. ISIS 113715 is currently being studied in early Phase II clinical trials to determine its ability to improve or restore insulin receptor sensitivity in patients with type 2 diabetes mellitus. *In vitro* protein binding displacement studies in human plasma samples showed minor effects on rosiglitazone and no effect on glipizide when co-incubated with ISIS 113715.[4]

ISIS 301012 (mipomersen) is a 20-nucleotide second-generation ASO targeting human ApoB mRNA. Its protein binding is >95% at clinically relevant concentrations. An *in vitro* protein binding interaction study with atorvastatin and simvastatin (both with >95% protein binding) showed no displacement between mipomersen sodium and atorvastatin or simvastatin at clinically relevant plasma concentrations.[5,6]

ISIS 2302, a first-generation ASO that targets human intercellular adhesion molecule-1 (ICAM-1) mRNA, is currently in clinical development for treatment of ulcerative colitis.[7] *In vitro* drug displacement studies with ISIS 2302 and other highly protein-bound drugs, including warfarin, salicylic acid, valproic acid, tolbutamide, acetylsalicylic acid, ibuprofen, and prednisone, were conducted. The results suggest that coadministration of ISIS 2302 with these highly protein-bound drugs did not displace one another at supraclinical plasma concentrations. Based on these findings, it is not anticipated to see PK interactions in the clinical settings as a result of the displacement of plasma protein binding.

In summary, *in vitro* data from protein binding displacement studies suggest that the potential for PK interactions caused by displacement of plasma protein binding is low when ASOs are administrated concomitantly with other highly protein bound drugs. The analyses by Benet et al. have shown that changes in plasma protein binding usually do not influence the clinical exposure of a patient to a drug.[8] As a consequence, no adjustments in dosing regimens will be necessary except in rare cases of a drug with a high extraction ratio and narrow therapeutic index that is given parenterally or, even rarer, a drug with a narrow therapeutic index that is given orally and has a very rapid pharmacokinetic-pharmacodynamic equilibration time.[8]

***15.3.1.2 DDI via Metabolic Enzymes*** Oligonucleotides are mainly metabolized by nucleases and are unlikely substrates for the cytochrome P450 (CYP) mediated oxidative metabolism. Therefore, the PK of oligonucleotides are not expected to be affected by concomitant use of drugs that are CYP inducers or

inhibitors. However, the potential for oligonucleotides to act as inhibitors or inducers of CYP enzymes cannot be precluded. *In vitro* studies using human microsomes have shown that several oligonucleotide drugs inhibited multiple CYP enzymes, including CYP3A4, CYP1A2, CYP2D6, CYP2E1, CYP2B6, CYP2A6, and CYP2C8 (unpublished data).

Several studies with negative results are reported in literature. For example, the ability of mipomersen to inhibit the major CYP enzymes—namely CYP1A2, CYP2C9, CYP2C19, CYP2D6, and CYP3A4—in cryo-preserved human hepatocytes was evaluated.[6] The results of this *in vitro* study revealed <10% inhibition in the activity of all evaluated enzymes at concentrations up to 800 μg/mL (~100 μmol/L), and dose-dependent inhibition was not observed. This study suggested mipomersen is unlikely to be an inhibitor of CYP1A2, CYP2C9, CYP2C19, CYP2D6, and CYP3A4.

### 15.3.2 DDI Studies in Animals

G3139 is a first-generation ASO targeting the open reading frame of oncogene *Bcl-2* mRNA. The PK interaction between G3139 and anthracycline anticancer drug doxorubixin was studied in severely compromised immunodeficient (SCID) mice model bearing human breast cancer xenografts.[9,10] When doxorubixin was coadministrated intravenously with intraperitoneal (IP) G3139, elevated G3139 peak concentration ($C_{max}$) was observed, with a fourfold increase in the absorption rate of G3139 from the IP cavity into plasma in the presence of doxorubixin. The systemic exposure (AUC) was only slightly increased.[10]

It is interesting that examination of the exposure of doxorubixin after coadministration with G3139 revealed marked elevation of tumor doxorubixin levels as compared to those after doxorubixin administered alone. Although the plasma exposure of doxorubixin in mice treated with G3139 was minimally reduced compared with no G3139 treatment, tumor exposure to doxorubixin increased twofold to threefold, indicating possible oligonucleotide-drug interactions at the tumor site. The mechanism behind the observed elevation of tumor tissue distribution is not fully understood. Findings from histology slides suggest that it could be through enhancing drug permeability of tumor tissue treated with G3139, thereby increased tumor uptake of doxorubixin-G3139 complexes that could form in the circulation, and increased binding of doxorubixin to tumor cells arising from G3139-induced DNA lesions caused by apoptosis activation.[10]

Similar findings have been reported by Wang et al. for an antisense anti-MDM2 mixed-backbone oligonucleotide, which is a second-generation ASO.[11] There were increased tissue levels of irinotecan and its active metabolite SN-38 in nude xenograft mice models when the two drugs were given in combination.[11]

### 15.3.3 DDI Studies in Humans

Two ASOs, fomivirsen and pegaptanib have been approved in the United States in 1998 and 2004, respectively. Two early clinical studies conducted for pegaptanib

alone and in combination with photodynamic therapy (PDT) or verteporfin showed no apparent difference in the plasma PK of pegaptanib.[12] However, no dedicated DDI studies have been conducted in humans between either of these two ASOs and other drugs.

A number of oligonucleotide-based drugs are currently on clinical trials. For the drugs that are in the development stage, a number of clinical DDI studies have been reported in the literature. An overview of the effects of oligonucleotides on the PK of coadministrated drugs is presented in Table 15-1. The effects of coadministrated small molecular drugs on the PK of oligonucleotides are summarized in Table 15-2. Some of these DDI studies and their study designs tailored based on the PK characteristics of individual drugs are further discussed in the following section. Among them, there are a few formal or dedicated DDI studies. The approach of a cross-study comparison was also used in several studies. A literature search found that the published DDI studies conducted in humans were to assess the potential interactions between first-generation or second-generation ASOs and other chemical drugs that are either in combination therapies with these ASOs, or are likely to be coadministrated to patients.

***15.3.3.1 Mipomersen (ISIS 301012)*** Mipomersen is being developed as an adjunct to maximally tolerated lipid-lowering therapies in patients with familial hypercholesteromia who have severely elevated low-density lipoprotein cholesterol (LDL-C) and are at high risk for coronary heart disease. A dedicated DDI study was conducted to characterize the PK interactions between mipomersen and lipid-lowering agents, including simvastatin and ezetimibe.[6] Both drugs and their metabolites are highly protein bound (> 90%). In addition, simvastatin is a substrate of CYP3A4 and CYP2D6, and ezetimibe is primarily metabolized via glucuronide conjugation with subsequent biliary and renal excretion. In this study, 10 healthy subjects per cohort received an oral dose of 40 mg simvastatin or 10 mg ezetimibe before and after four 2-h intravenous doses of 200 mg mipomersen sodium over an 8-day period. For cohort A, mipomersen sodium was given on days 4, 6, 8, and 11. Mipomersen PK profiles were assessed after the dose on day 8 (mipomersen alone) and on day 11 (in combination with simvastatin). For cohort B, mipomersen sodium was given on days 8, 10, 12, and 15. Mipomersen PK profiles were assessed after the dose on day 12 (mipomersen alone) and on day 15 (in combination with ezetimibe). The dosing regimen for mipomersen sodium was designed to achieve near steady-state tissue accumulation by the end of dosing.

The values for area under the plasma concentration-time curve ($AUC_{24h}$), maximum plasma concentration ($C_{max}$), and apparent elimination half-life ($t_{½}$) of each drug were calculated, and their 90% confidence intervals (CIs) of the geometric least-squares means (GLSM) ratios are shown in Table 15-3, which indicates that the systemic exposure of mipomersen was not altered when coadministered with simvastatin or ezetimibe. However, mipomersen coadministration resulted in approximately 50% lower $C_{max}$ and 28% higher $AUC_{24\,h}$ for simvastatin; and a 42% lower $C_{max}$ with no appreciable change in $AUC_{72\,h}$ for ezetimibe. The underlying mechanisms for the observed results are to be elucidated.

**TABLE 15-1 Effects of Oligonucleotides on the Pharmacokinetics of Other Drugs**

| Interacting ASOs | Class | Substrates | Study Design | PK Parameters | Results | References |
|---|---|---|---|---|---|---|
| Mipomersen | Gen 2 | Ezetimibe, simvastatin | Healthy subjects; parallel design | $AUC_{72\,h}$, $AUC_{24\,h}$, $C_{max}$, $T_{max}$, $t_{½}$ | Ezetimibe $C_{max}$ ↓ 42%, $AUC_{72h}$ no change; simvastatin $C_{max}$ ↓ 49%, $AUC_{24h}$, ↑ 28% | 5,6 |
| ISIS 113715 | Gen 1 | Glipizide, metformin, rosiglitazone | Cancer patients; sequential crossover design | $AUC_{24\,h}$, $C_{max}$, $t_{½}$ | Rosiglitazone $C_{max}$ ↑ 46%, $AUC_{24}$ ↑ 15%; glipizide and metforminno change | 4 |
| ISIS 2503 | Gen 1 | Gemcitabine | Cancer patients; sequential crossover design | $C_{ss}$ | Gemcitabine $C_{max}$ ↓, AUC ↓; not statistically significant | 13 |
| ISIS 3521 | Gen 1 | 5-fluorouracil | Cancer patients; sequential crossover design | $C_{ss}$ | No changes | 14 |
| | | Gemcitabine, cisplatin | Cancer patients; sequential crossover design | AUC, $C_{max}$ | No changes | 15 |
| | | Paclitaxel, carboplatin | Cancer patients; sequential crossover design | AUC, $C_{max}$ | No changes | 16,17 |

*(continued)*

**TABLE 15-1** (*Continued*)

| Interacting ASOs | Class | Substrates | Study Design | PK Parameters | Results | References |
|---|---|---|---|---|---|---|
| G3139 | Gen 1 | Carboplatin, paclitaxel | Cancer patients; cross-study PK comparison | AUC, CL, $V_{ss}$, $t_{½}$ | No changes | 18 |
| GEM231 | Gen 2 | Paclitaxel, docetaxel | Cancer patients; parallel design | $C_{ss}$ | Ongoing | 19 |
| G3139 | Gen 1 | Doxorubucin | Mice; parallel study | AUC, $V_d$, $t_{½}$ | Tumor exposure to DOX ↑ twofold to threefold | 9,10 |
| Oligo AS | Gen 2 | Irinotecan | Mice; parallel study | Tissue concentration | Tissue irinotecan and SN-38 ↑ | 20 |

ASOs: antisense oligonucleotides; AUC: area under the concentration-time curve; CL: clearance; $C_{max}$: maximum concentration; $C_{ss}$: steady-state concentrations; DOX: Doxorubucin; Gen 1: first-generation phosphorothioate oligonucleotide; Gen 2: second-generation antisense chimeric 2′-*O*-(2-methoxyethyl) phosphorothioate oligonucleotides; PK: pharmacokinetics; $T_{max}$: time to maximum concentration; $V_d$: volume of distribution; $V_{ss}$: steady-state volume of distribution; $t_{½}$: elimination half-life.

**TABLE 15-2 Effects of Other Drugs on the Pharmacokinetics of Oligonucleotides**

| Interacting Drugs | Substrates | Class | Study Design | PK | Results | References |
|---|---|---|---|---|---|---|
| Ezetimibe; simvastatin | Mipomersen | Gen 2 | Healthy subjects; parallel design | $AUC_{24\,h}$, $C_{max}$, $T_{max}$, $t_{½}$ | No changes | 5,6 |
| Glipizide; metformin; rosiglitazone | ISIS 113715 | Gen 1 | Cancer patients; sequential crossover design | $AUC_{24\,h}$, $C_{max}$, $t_{½}$ | No changes | 4 |
| 5-Fluorouracil; leucovorin | ISIS 3521 | Gen 1 | Cancer patients; sequential crossover design | AUC, CL, $V_{ss}$, $t_{½}$ | No changes | 14 |
| Gemcitabine; cisplatin | ISIS 3521 | Gen 1 | Cancer patients; sequential crossover design | $C_{ss}$ | $C_{ss}$ ↑ 60% with cisplatin | 15 |
| Gemcitabine | ISIS 2503 | Gen 1 | Cancer patients; sequential crossover design | AUC, $C_{max}$, CL, $V_{ss}$, $t_{½}$ | No changes | 13 |
| Doxorubucin | G3139 | Gen 1 | Mice, parallel study | AUC, $C_{max}$, CL, $t_{½}$ | G3139 $C_{max}$ ↑; absorption rate↑ | 10 |

AUC: area under the concentration-time curve; CL: clearance; $C_{max}$: maximum concentration; $C_{ss}$: steady-state concentrations; Gen 1: first generation phosphorothioate oligonucleotide; Gen 2: second-generation antisense chimeric 2′-*O*-(2-methoxyethyl) phosphorothioate oligonucleotides; PK: pharmacokinetics; $V_{ss}$: steady-state volume of distribution; $t_{½}$: elimination half-life.

**TABLE 15-3 Geometric Least-Squares Means (GLSM) Ratios and 90% Confidence Intervals (CIs) of Pharmacokinetic Parameters After Coadministration of Substrates and Interacting Drugs[a]**

| Substrates | Interacting Drugs | PK Parameters | GLSM Ratio | 90% CI |
|---|---|---|---|---|
| Simvastatin | Mipomersen | $C_{max}$ (ng/mL) | 0.51 | (0.34, 0.77) |
| | | $AUC_{24\,h}$ (h·ng/mL) | 1.28 | (0.90, 1.82) |
| Ezetimibe | Mipomersen | $C_{max}$ (ng/mL) | 0.58 | (0.40, 0.82) |
| | | $AUC_{72\,h}$ (h·ng/mL) | 1.00 | (0.78, 1.27) |
| Mipomersen | Simvastatin | $C_{max}$ (ng/mL) | 0.98 | (0.93, 1.03) |
| | | $AUC_{24\,h}$ (h·ng/mL) | 1.00 | (0.94, 1.07) |
| Mipomersen | Ezetimibe | $C_{max}$ (ng/mL) | 1.05 | (0.86, 1.28) |
| | | $AUC_{24\,h}$ (h·ng/mL) | 1.01 | (0.92, 1.11) |

[a] $n = 10$.
CI: confidence interval; PK: pharmacokinetics.

The clinical trial design used in this study took into consideration the long elimination half-life of mipomersen and the relatively large interindividual variability in the PK of the oral lipid-lowering drugs. Thus the design of the study included a loading dose approach for mipomersen sodium that ensured achievement of near steady-state concentrations in the target tissues (the liver and kidneys) before testing of potential interactions with concomitantly administered oral lipid-lowering agents. The sequential design allowed each individual in the study to be his or her own control to reduce intraindividual variability. The limitation of the study is that a 24-h PK sampling duration was too short relative to the 30-day elimination half-life of mipomersen; therefore the $AUC_{0-24\,h}$ may not be an appropriate parameter for assessing the effect of coadministered drugs on the PK of mipomersen.

ISIS 2503 is an ASO targeting H-ras mRNA. The potential PK interaction between ISIS 2503 and gemcitabine was evaluated in a Phase I trial.[13,21] A total of 19 patients were treated with a fixed dose of gemcitabine 1000 mg/m$^2$ on days 1 and 8 and two escalating doses of ISIS 2503 (4 and 6 mg/kg/day) as a 14-day continuous infusion starting on day 1. An additional 8 patients received a flat dose of ISIS 2503 based on their ideal body weight. Blood samples for gemcitabine PK were collected on day 1 (gemcitabine alone) and day 8 (with ISIS 2503). Only steady-state concentration ($C_{ss}$) data were collected for ISIS 2503, at pregemcitabine and postgemcitabine infusion on day 8.

At three dose levels studied, the $C_{ss}$ of ISIS 2503 achieved during infusion were not altered, and the fraction of total oligonucleotide composed of parent oligonucleotide (intact 20-mer ISIS 2503) was also unchanged when ISIS 2503 was coadministered with gemcitabine. There was a consistent decrease in the gemcitabine plasma concentration observed at the end of the 30-min gemcitabine infusion in all ISIS 2503 three dose cohorts. Alteration of gemcitabine concentration at this single time point resulted in a 31–48% decrease in mean AUC and a 49–57% decrease in mean $C_{max}$ in patients receiving a flat dose or 6 mg/kg/day of ISIS 2503.

ISIS 113715 is a 20-mer first-generation phosphorothioate ASO that is complementary to the protein tyrosine phosphatase 1B (PTP-1B) mRNA. In a clinical DDI study, 23 healthy subjects received a single oral dose of an antidiabetic drug (either metformin 500 mg, glipizide 5 mg or rosiglitazone 2 mg) alone and together with 200 mg of ISIS 113715 subcutaneously in a sequential-crossover design.[4] The results of this study indicated that there were no large deviations outside of the boundaries of 80–125% for PK parameters ($C_{max}$ and AUC) of glipizide and metformin, but some differences were observed for rosiglitazone with the mean ratios and 90% CIs for $C_{max}$ and $AUC_{24\,h}$ of 1.46 (1.04–2.05) and 1.15 (0.80–1.65), respectively when coadministered with ISIS 113715. Conversely, there were no observed significant alterations in the PK of ISIS 113715 when administered in combination with these oral antidiabetic drugs.

#### *15.3.3.2 ISIS 3521 (LY900003)*

*In Combination with 5-Fluorouracil and Leucovorin* ISIS 3521 (LY900003) is a sodium salt form of a 20-mer phosphorothioate oligonucleotide that hybridizes to the 3′-untranslated region of human protein kinase C (PKC) mRNA and inhibits its expression through RNase-mediated cleavage of hybridized PKC mRNA. In a clinical study, patients with refractory solid tumors received ISIS 3521 as a 21-day continuous infusion simultaneously with 5-fluorouracil (5-FU) and leucovorin given daily for 5 days and repeated every 4–5 weeks (one cycle).[14] A PK analysis for 5-FU and ISIS 3521 was performed on samples taken during the first cycle in all patients. Fifteen patients received ISIS 3521 at one of three dose levels: 1.0 mg/kg/day ($n = 3$), 1.5 mg/kg/day ($n = 3$), and 2.0 mg/kg/day ($n = 9$). All patients concomitantly received 5-FU (425 mg/m$^2$/day) and leucovorin (20 mg/m$^2$/day) for 5 consecutive days. The washout time period was not specified in the paper. The steady-state plasma concentration ($C_{ss}$) of ISIS 3521 without 5-FU was determined by averaging the measured plasma concentrations at time 24 h to time 21 days during the continuous infusion. After 5-FU administration on day 4, ISIS 3521 concentrations measured at multiple time points within 90 min of 5-FU administration were averaged to obtain $C_{ss}$.

The plasma concentration of ISIS 3521 increased in a dose-dependent manner over the range of doses investigated. There was no significant difference in the ISIS 3521 steady-state plasma concentrations when ISIS 3521 was given with or without 5-FU. Furthermore, there was no significant change in the percentage of intact ISIS 3521 in the presence or absence of 5-FU. Because there were only two to three patient observations recorded in each dose treatment cohort, a statistical test cannot be performed with confidence; however, on visual inspection of the data, there was considerable overlap of the ISIS 3521 $C_{ss}$ values in the presence and absence of 5-FU. The PK parameters [AUC, $t_{½}$, CL, and volume of distribution at steady-state ($V_{ss}$)] of 5-FU in the presence or absence of ISIS 3521 also showed no evidence for PK interactions.[14]

*In Combination with Cisplatin and Gemcitabine* A clinical trial of ISIS 3521 in combination with cisplatin/gemcitabin was conducted in patients with advanced

non–small cell lung cancer.[15] The PK interaction study design was similar to the previously mentioned study reported by Mani et al. For gemcitabine, the respective GLSM ratios on study days 1 and 8 for each parameter of interest ($AUC_{0-\infty}$ and $C_{max}$) were ranged from 0.69 to 1.27. Similar findings were observed for cisplatin, and the GLSM ratios for each parameter of interest in cycles 1 and 2 were close to 1.

This study compared the $C_{ss}$ of ISIS 3521 and total oligonucleotide in the absence and presence of gemcitabine, cisplatin, or both drugs. In the presence of gemcitabine, the GLSM ratio for $C_{ss}$ of ISIS 3521 and total oligonucleotide are 0.81 and 0.70, respectively. In the presence of cisplatin, the GLSM ratio for $C_{ss}$ of ISIS 3521 and total oligonucleotide are 1.60 and 1.61, respectively. The clinical significance of this magnitude of change in ISIS 3521 $C_{ss}$ when coadministered with gemcitabine and cisplatin is unknown.

*In Combination with Carboplatin and Paclitaxel* A clinical trial was conducted with ISIS 3521 in combination with carboplatin and paclitaxel for the treatment of patients with non-small cell lung cancer. Patients initially received carboplatin (target AUC: 5 or 6 mg/mL/min) and paclitaxel (175 mg/m$^2$) alone in cycle 1. In cycle 2 and beyond, patients received escalating doses of ISIS 3521 (1.0–2.0 mg/kg/day) by continuous intravenous infusion on days 1–14, with carboplatin and paclitaxel administered on day 4. In the expanded phase, all patients received carboplatin (target AUC: 6 mg/mL/min) and paclitaxel (175 mg/m$^2$) with ISIS 3521 (2.0 mg/kg/day) starting in cycle 1. Treatment was repeated every 21 days until maximum benefit. There was no evidence of a PK interaction between ISIS 3521 and either chemotherapy agent.[16]

Overall, these literature reports conclude that ISIS 3521 has no clinically relevant effect on the PK of those drugs administered concomitantly with ISIS 3521. However, a 60% increase in ISIS 3521 $C_{ss}$ in the presence of cisplatin was observed and the underlying mechanisms are not fully explained. In addition, there are several limitations for these reported DDI studies. Firstly, some of the analyses did not include other PK parameters such as $C_{max}$ which was limited by the design of DDI studies. Secondly, the studies suffered from a small numbers of patients (e.g., only two to three patient observations recorded in each dose cohort in the study with 5-FU and leucovorin). Lastly, the dose and regimen in these studies may not be the maximal clinical doses to detect a significant PK interaction.

*Drug Transporters* Despite over a decade of research, the mechanism of cellular membrane transport of oligonucleotide-based drugs are still unclear, possibly involving multiple pathways.[22] Oligonucleotides usually have molecular weight greater than 4 kDa and contain more than 12 negative charges per molecule. The general belief is that only lipophilic or small hydrophilic molecules can traverse biological membranes. The mechanism of cellular uptake of naked DNA/RNA is still poorly defined, but one of these pathways may require a cell membrane transporter in mammalian cells.[23]

There are no published reports on the role of P-glycoprotein (P-gp) in oligonucleotide absorption or intracellular transport. A study suggested that at a high

dose of 10 μM, an ASO can be transported by P-gp *in vitro* (unpublished data). Until the mechanism of oligonucleotide transport is clearly delineated, the potential for oligonucleotides to be as substrates of transporters such as P-gp should be considered during the drug development.

## 15.4 OTHER CONSIDERATIONS

### 15.4.1 Indirect Effects on CYP Enzymes

Some oligonucleotides contain unmethylated CpG motifs within the oligonucleotide sequence that have been shown to be potent immune stimulators.[24] The immunostimulatory oligonucleotides can activate signaling cascades to regulate the expression of a number of host-defense proteins such as interferons and interferon-inducible genes. Cytokines, such as interleukin 6 (IL-6), may affect the oxidative metabolic process by modulating the hepatic microsomal CYP isozymes, and thereby change the metabolism of concomitantly administrated drugs that are substrates of these enzymes.

The delivery systems and formulations of oligonucleotides, which include conjugation with other molecules, such as pegylation, liposome, or lipid nanoparticles (LNPs) are often used to extend elimination half-lives of the siRNA or ASOs or to have a better delivery of them to the target cells. Some delivery systems or formulations may have the ability to modulate cytokine release and increase the complexities of drug interaction potential.

### 15.4.2 Immunogenicity and Its Effects on PK

Experimental induction of antibodies in mice suggests the potential for DNA to be immunogenic. In normal subjects, anti-DNA B cell precursors make up as much as 1–5% of the total B cell population. The introduced antisense DNA or RNA may induce autoantibody or anti-ASO antibody production (immunogenicity) that would not normally occur for the endogenous sense sequence. There are several mechanisms that may be responsible for the incidence of ASO immunogenicity: (1) the use of modified analogs designed to increase nuclease resistance may cause immunogenicity, (2) protection from digestion and enhanced cellular uptake achieved by linking oligonucleotides to protein carriers causes additional problems, and (3) some oligonucleotide drugs produce potent immune stimulation *in vivo*, thereby enhancing the anti-PEG antibody response when these immunostimulatory agents are encapsulated in PEGylated liposomes.[25,26]

Altered PK profiles associated with antiproduct antibody formation have been seen in therapeutic biological products. The potential for immunogenicity of oligonucleotide-based drugs may also alter their PK. Currently most of the oligonucleotide-based drugs are still in the preclinical or early clinical development stages; therefore, the immunogenicity data are not available for most compounds under development.

### 15.4.3 Considerations of DDI Studies

The first-generation ASOs often need to be administrated chronically to attain persistent modulation of target gene expression, especially in oncology. When both the substrate and the interacting drug are likely to be given over an extended period of time, the often used DDI study design is to administer the substrate to steady state with collection of blood samples over one or more dosing intervals followed by multiple dose administration of an interacting drug.[27] The second-generation ASOs have long elimination half-life of 10–30 days. Design of the DDI studies for such long half-life drugs may consider inclusion of a loading dose to ensure faster achievement of the steady-state concentrations of the study drug. In addition, a cross-over design may not be practical because of the long elimination half-life.

### 15.4.4 Regulatory Perspective

Currently, there is no specific regulatory guidance on clinical pharmacology studies for oligonucleotide-based drugs. The current regulatory recommendations for evaluation of oligonucleotide drug–drug interactions are similar to the approaches for small molecule drugs.[27] In general, the metabolism and transport of an investigational new drug and its interactions with other drugs as a substrate, inhibitor, or inducer needs to be assessed during drug development as part of an adequate assessment of the drug's safety and effectiveness. At the early stage of clinical development, *in vitro* screens need to be conducted to determine if the parent or active metabolite(s) is a substrate, inhibitor or inducer of CYP isozymes or P-gp transporter. These studies will help determine the need for *in vivo* drug–drug interactions.

In oncology, many oligonucleotides are developed in combination with the current standard therapeutics. If oligonucleotides are to be used in combination with other drugs such as chemotherapy, the potential for a PK drug–drug interaction between oligonucleotides and the drugs in the combination will need to be evaluated during the development of such combination therapy. The evaluation can be conducted in a dedicated drug–drug interaction trial or incorporated in a clinical efficacy and safety trial. This evaluation is important because many oncology drugs have narrow therapeutic indices and altered drug systemic exposure could increase drug-related toxicities or compromise its therapeutic effects.

## 15.5 SUMMARY

The reported clinical data suggest that there are limited PK drug–drug interactions between oligonucleotide-based drugs and small molecular drugs. A few animal studies demonstrated that tumor exposures of the anticancer drugs are affected by the oligonucleotides coadministration without affecting their plasma pharmacokinetics. *In vitro* data from protein binding displacement studies suggest that the

potential for pharmacokinetic interactions caused by displacement of plasma protein binding is low when oligonucleotides are administrated concomitantly with other highly protein-bound drugs.

Oligonucleotides are mainly metabolized by nucleases in tissues and are unlikely to be substrates of CYP metabolic enzymes. However, the current data regarding the CYP enzyme involvement are very limited and cannot exclude the possibility that the exposure of oligonucleotides can be affected by CYP enzyme inducers or inhibitors. Given that *in vitro* studies using human microsomes have shown that several oligonucleotide drugs inhibited multiple CYP enzymes, the inhibition and induction potential of oligonucleotides for CYP enzymes should be addressed during the oligonucleotide-based drug development.

## REFERENCES


1. Crooke ST, ed. *Antisense drug technology: principles, strategies, and applications*. Boca Raton, CRC Press, 2007.
2. Geary RS, Antisense oligonucleotide pharmacokinetics and metabolism. *Expert Opin Drug Metab Toxicol* **5**, 381–91 (2009).
3. Geary RS, Watanabe TA, Truong L, et al. Pharmacokinetic properties of 2′-*O*-(2-methoxyethyl)-modified oligonucleotide analogs in rats. *J Pharmacol Exp Ther* **296**, 890–97 (2001).
4. Geary RS, Bradley JD, Watanabe T, et al. Lack of pharmacokinetic interaction for ISIS 113715, a 2′-*O*-methoxyethyl modified antisense oligonucleotide targeting protein tyrosine phosphatase 1B messenger RNA, with oral antidiabetic compounds metformin, glipizide or rosiglitazone. *Clin Pharmacokinet* **45**, 789–801 (2006).
5. Ito MK, ISIS 301012 gene therapy for hypercholesterolemia: sense, antisense, or nonsense? *Ann Pharmacother* **41**, 1669–78 (2007).
6. Yu RZ, Geary RS, Flaim JD, et al. Lack of pharmacokinetic interaction of mipomersen sodium (ISIS 301012), a 2′-*O*-methoxyethyl modified antisense oligonucleotide targeting apolipoprotein B-100 messenger RNA, with simvastatin and ezetimibe. *Clin Pharmacokinet* **48**, 39–50 (2009).
7. Watanabe TA, Geary RS, Levin AA, Plasma protein binding of an antisense oligonucleotide targeting human ICAM-1 (ISIS 2302). *Oligonucleotides* **16**, 169–80 (2006).
8. Benet LZ, Hoener BA, Changes in plasma protein binding have little clinical relevance. *Clin Pharmacol Ther* **71**, 115–21 (2002).
9. Lopes de Menezes DE, Hudon N, McIntosh N, Mayer LD, Molecular and pharmacokinetic properties associated with the therapeutics of bcl-2 antisense oligonucleotide G3139 combined with free and liposomal doxorubicin. *Clin Cancer Res* **6**, 2891–902 (2000).
10. Lopes de Menezes DE, Mayer LD, Pharmacokinetics of Bcl-2 antisense oligonucleotide (G3139) combined with doxorubicin in SCID mice bearing human breast cancer solid tumor xenografts. *Cancer Chemother Pharmacol* **49**, 57–68 (2002).

11. Wang H, Nan L, Yu D, et al. Antisense anti-MDM2 oligonucleotides as a novel therapeutic approach to human breast cancer: in vitro and in vivo activities and mechanisms. *Clin Cancer Res* **7**, 3613–24 (2001).
12. Gilead Sciences. Package insert for Macugen (pegaptanib). San Dimas, CA, 2011. Available at http://www.accessdata.fda.gov/drugsatfda_docs/label/2011/021756s018lbl.pdf.
13. Adjei AA, Dy GK, Erlichman C, et al. A phase I trial of ISIS 2503, an antisense inhibitor of H-ras, in combination with gemcitabine in patients with advanced cancer. *Clin Cancer Res* **9**, 115–23 (2003).
14. Adjei AA, Rowinsky EK, Novel anticancer agents in clinical development. *Cancer Biol Ther* **2**, S5–15 (2003).
15. Mani S, Rudin CM, Kunkel K, et al. Phase I clinical and pharmacokinetic study of protein kinase C-alpha antisense oligonucleotide ISIS 3521 administered in combination with 5-fluorouracil and leucovorin in patients with advanced cancer. *Clin Cancer Res* **8**, 1042–48 (2002).
16. Villalona-Calero MA, Ritch P, Figueroa JA, et al. A phase I/II study of LY900003, an antisense inhibitor of protein kinase C-alpha, in combination with cisplatin and gemcitabine in patients with advanced non-small cell lung cancer. *Clin Cancer Res* **10**, 6086–93 (2004).
17. Yuen AR, Halsey J, Fisher GA, et al. Phase I study of an antisense oligonucleotide to protein kinase C-alpha (ISIS 3521/CGP 64128A) in patients with cancer. *Clin Cancer Res* **5**, 3357–63 (1999).
18. Motl SE, LY900003: a novel compound for the treatment of non-small cell lung cancer. *Cancer Ther* **1**, 237–44 (2003).
19. Liu G, Kolesar J, McNeel DG, et al. A phase I pharmacokinetic and pharmacodynamic correlative study of the antisense Bcl-2 oligonucleotide g3139, in combination with carboplatin and paclitaxel, in patients with advanced solid tumors. *Clin Cancer Res* **14**, 2732–39 (2008).
20. Mani S, Goel S, Nesterova M, et al. Clinical studies in patients with solid tumors using a second-generation antisense oligonucleotide (GEM 231) targeted against protein kinase A type I. *Ann N Y Acad Sci* **1002**, 252–62 (2003).
21. Wang H, Wang S, Nan L, et al. Antisense anti-MDM2 mixed-backbone oligonucleotides enhance therapeutic efficacy of topoisomerase I inhibitor irinotecan in nude mice bearing human cancer xenografts: in vivo activity and mechanisms. *Int J Oncol* **20**, 745–52 (2002).
22. Wu-Pong S, Alternative interpretations of the oligonucleotide transport literature: insights from nature. *Adv Drug Deliv Rev* **44**, 59–70 (2000).
23. Bennett RM, As nature intended? The uptake of DNA and oligonucleotides by eukaryotic cells. *Antisense Res Dev* **3**, 235–41 (1993).
24. Klinman DM, Klaschik S, Tomaru K, et al. Immunostimulatory CpG oligonucleotides: effect on gene expression and utility as vaccine adjuvants. *Vaccine* **28**, 1919–23 (2010).
25. Dawson KH, Bell DA, Production and pathogenic effects of anti-DNA antibodies: relevance to antisense research. *Antisense Res Dev* **1**, 351–60 (1991).
26. Judge AD, Bola G, Lee AC, MacLachlan I, Design of noninflammatory synthetic siRNA mediating potent gene silencing in vivo. *Mol Ther* **13**, 494–505 (2006).

27. US Food and Drug Administration, US Department of Health and Human Services, Center for Drug Evaluation and Research, Center for Biologics Evaluation and Research (CBER). Draft guidance for industry: drug interaction studies—study design, data analysis, implications for dosing, and labeling. Silver Spring, MD, 2012. Available at http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM292362.pdf.

# MONOGRAPHS FOR DRUG-DRUG INTERACTIONS OF THERAPEUTICS BIOLOGICS

JOCELYN LEU, SIMONE KASEK, CHRISTINE LI, WARARAT LIMOTHAI, DORA BABU MADHURA, CHETAN RATHI, SUMIT RAWAL, JOSIAH RYMAN, MARGARET THOMSON, ASHIT TRIVEDI, HONGHUI ZHOU, and BERND MEIBOHM

Over the last decade, the unprecedented increase in the clinical use of therapeutic biologics for a variety of indications including immune-mediated inflammatory disease indications such as rheumatoid arthritis and inflammatory bowel diseases, and oncology indications has resulted in a new class of drug-drug interactions, therapeutic biologic-drug interactions. However, this kind of drug-drug interactions is generally understudied or underreported due to a perception that drug-drug interactions involving therapeutic biologics are scarce and of limited relevance. Moreover, the underlying mechanisms of many of those drug-drug interactions are still poorly understood.

In order to better assess the landscape of drug-drug interactions for therapeutic biologics, we have compiled drug-drug interaction monographs involving therapeutic biologics either as victims or perpetrators based on current US labeling information and scientific literature. To broaden our knowledgebase in this area, our monographs also include some important 'theoretical' drug-drug interactions with well established plausible mechanisms. In order to make those monographs more useful to the readers, we have also included other relevant references if available from which the drug-drug interaction contents was obtained. We feel these monographs might be useful for readers as they might guide expectations in what might be encountered as potential therapeutic biologic-drug interaction in clinical development programs and applied pharmacotherapy.

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*,
Edited by Shayne C. Gad
Copyright © 2013 John Wiley & Sons, Inc.

**Drug-Drug Interaction Monograph for ACCRETROPIN (Somatropin) #1**

**Perpetrator drug:**

| | |
|---|---|
| INN name/US brand name: | Somatropin/Accretropin, Bio-Tropin, |
| Class/Molecular weight: | Hormone/~22 kDa |

**Victim drug:**

| | |
|---|---|
| INN name/US brand name: | Cortisone, prednisone |
| Class/Molecular weight: | Corticosteroids |

**Description of the effect of the DDI:**
Inhibits metabolism of corticosteroids to the biologically active metabolites (e.g., cortisone to cortisol).

**Description of the (proven, putative or hypothesized) mechanism of the DDI:**
Inhibits 11 β-hydroxysteroid dehydrogenase type 1 (11βHSD-1) in adipose/hepatic tissue since conversion of these agents to the biologically active metabolites is dependent on the activity of the 11βHSD-1.

**Clinical consequences:**
Patients treated with glucocorticoid replacement therapy for previously diagnosed hypoadrenalism may require an increase in their maintenance or stress doses.

**Reference:** Accretropin prescribing information

**Drug-Drug Interaction Monograph for ACCRETROPIN (Somatropin) #2**

**Perpetrator drug:**

| | |
|---|---|
| INN name/US brand name: | Somatropin/Accretropin, Bio-Tropin |
| Class/Molecular weight: | Hormone/~22 kDa |

**Victim drug:**

| | |
|---|---|
| INN name/US brand name: | CYP450 3A4 substrates |
| Class/Molecular weight: | Corticosteroids, sex steroids, anticonvulsants, cyclosporine |

**Description of the effect of the DDI:**
Potentially increases clearance of CYP450 3A4 substrates.

**Description of the (proven, putative or hypothesized) mechanism of the DDI:**
Increases CYP450-mediated antipyrine clearance. These data suggest that somatropin administration may alter the clearance of compounds known to be metabolized by CYP450 liver enzymes (e.g., corticosteroids, sex steroids, anticonvulsants, cyclosporine). No formal DDI studies have been conducted.

**Clinical consequences:**
Careful monitoring is advisable.

**Reference:** Accretropin US prescribing information

Lee JI, Zhang L, Men AY, Kenna LA, Huang SM. CYP-mediated therapeutic protein-drug interactions: clinical findings, proposed mechanisms and regulatory implications. Clin Pharmacokinet. 2010 May;49(5):295–310.

**Drug-Drug Interaction Monograph for ACCRETROPIN (Somatropin) #3**

**Perpetrator drug:**
INN name/US brand name: Somatropin/Accretropin
Class/Molecular weight: Hormone/~22 kDa

**Victim drug:**
INN name/US brand name: Insulin and oral antidiabetic agents
Class/Molecular weight: Hormone/5.8 kDa

**Description of the effect of the DDI:**
Pharmacodynamic effect inducing insulin resistance.

**Description of the (proven, putative or hypothesized) mechanism of the DDI:**
Decreases insulin sensitivity at high doses. Growth hormone induces insulin resistance by the stimulation of lipolysis and a concomitant switch from oxidation of glucose to oxidation of lipids, during both acute and chronic treatment.

**Clinical consequences:**
This could lead to glucose intolerance and possibly diminish the effects of insulin. The dose of insulin and/or oral agents may require adjustment when somatropin therapy is initiated.

**Reference:**

Accretropin prescribing information

Attallah H, Friedlander AL, Hoffman AR: Visceral obesity, impaired glucose tolerance, metabolic syndrome, and growth hormone therapy. Growth Horm IGF Res (2006) 16 (Suppl A/Jul): S62–7.

Groop L, Segerlantz M, Bramnert M: Insulin sensitivity in adults with growth hormone deficiency and effect of growth hormone treatment. Horm Res (2005) 64 (Suppl 3): 45–50.

**Drug-Drug Interaction Monograph for ACCRETROPIN (Somatropin) #4**

**Perpetrator drug:**
INN name/US brand name: Glucocorticoids
Class/Molecular weight: Small Molecule

**Victim drug:**
INN name/US brand name: Somatropin/Accretropin
Class/Molecular weight: Hormone/~22 kDa

**Description of the effect of the DDI:**
Pharmacodynamic effect on growth promoting effects of somatropin.

**Description of the (proven, putative or hypothesized) mechanism of the DDI:**
Excessive glucocorticoid therapy may attenuate the growth promoting effects of somatropin in children.

**Clinical consequences:**
Carefully adjust glucocorticoid replacement therapy with concomitant growth hormone.

*(Continued)*

**Reference:** Accretropin prescribing information

Mehls I, Tonshoff B, Kovacs G, Mayer C, Scerek J, Oh J: Interaction between glucocorticoids and growth hormone. Acta Paediatrica (2008) 82 (Suppl s389): 77–82.

---

**Drug-Drug Interaction Monograph for ACTEMRA (Tocilizumab) #1**

---

**Perpetrator drug:**
INN name/US brand name: Methotrexate
Class/Molecular weight: Small Molecule/476 Da

**Victim drug:**
INN name/US brand name: Tocilizumab/Actemra
Class/Molecular weight: Monoclonal Antibody/~148 kDa

**Description of the effect of the DDI:**
No effect on tocilizumab clearance by coadministered methotrexate evaluated by population PK analysis.

**Description of the (proven, putative or hypothesized) mechanism of the DDI:**
None.

**Clinical consequences:**
None.

**Reference:** Actemra US prescribing information

Frey N, Grange S, Woodworth T. Population pharmacokinetic analysis of tocilizumab in patients with rheumatoid arthritis. J Clin Pharmacol. 2010;50(7):754–66.

---

**Drug-Drug Interaction Monograph for ACTEMRA (Tocilizumab) #2**

---

**Perpetrator drug:**
INN name/US brand name: NSAIDS
Class/Molecular weight: Small Molecule

**Victim drug:**
INN name/US brand name: Tocilizumab/Actemra
Class/Molecular weight: Monoclonal Antibody/~148 kDa

**Description of the effect of the DDI:**
No effect on tocilizumab clearance by coadministered NSAIDS evaluated by population PK analysis.

**Description of the (proven, putative or hypothesized) mechanism of the DDI:**
None.

**Clinical consequences:**
None.

**Reference:** Actemra US prescribing information

Frey N, Grange S, Woodworth T. Population pharmacokinetic analysis of tocilizumab in patients with rheumatoid arthritis. J Clin Pharmacol. 2010 Jul;50(7):754–66.

---

**Drug-Drug Interaction Monograph for ACTEMRA (Tocilizumab) #3**

**Perpetrator drug:**

| | |
|---|---|
| INN name/US brand name: | Corticosteroids |
| Class/Molecular weight: | Small Molecule |

**Victim drug:**

| | |
|---|---|
| INN name/US brand name: | Tocilizumab/Actemra |
| Class/Molecular weight: | Monoclonal Antibody/~148 kDa |

**Description of the effect of the DDI:**
No effect on tocilizumab clearance by coadministered corticosteroids evaluated by population PK analysis.

**Description of the (proven, putative or hypothesized) mechanism of the DDI:**
None.

**Clinical consequences:**
None.

**Reference:** Actemra US prescribing information

Frey N, Grange S, Woodworth T. Population pharmacokinetic analysis of tocilizumab in patients with rheumatoid arthritis. J Clin Pharmacol. 2010; 50(7):754–66.

**Drug-Drug Interaction Monograph for ACTEMRA (Tocilizumab) #4**

**Perpetrator drug:**

| | |
|---|---|
| INN name/US brand name: | Tocilizumab/Actemra |
| Class/Molecular weight: | Monoclonal Antibody/~148 kDa |

**Victim drug:**

| | |
|---|---|
| INN name/US brand name: | Simvastatin/Zocor |
| Class/Molecular weight: | Small Molecule/419 Da |

**Description of the effect of the DDI:**
In 12 RA patients not treated with Actemra, receiving 40 mg simvastatin, exposures of simvastatin and its metabolite, simvastatin acid, was 4- to 10-fold and 2-fold higher, respectively, than the exposures observed in healthy subjects. One week following administration of a single infusion of Actemra (10 mg/kg), exposure of simvastatin and simvastatin acid decreased by 57% and 39%, respectively, to exposures that were similar or slightly higher than those observed in healthy subjects. Exposures of simvastatin and simvastatin acid increased upon withdrawal of Acemtra in RA patients.

**Description of the (proven, putative or hypothesized) mechanism of the DDI:**
Simvastatin is a CYP3A4 and OATP1B1 substrate. *In vitro* data suggested that IL-6 reduced mRNA expression for CYP3A4, and this reduced expression was reversed by co-incubation with tocilizumab at clinically relevant concentrations. Accordingly, inhibition of IL-6 signaling in RA patients treated with tocilizumab may restore CYP3A4 activity to higher levels than those in the absence of tocilizumab leading to increased metabolism of drugs that are CYP3A4 substrate.

*(Continued)*

**Clinical consequences:**
Selection of a particular dose of simvastatin in RA patients should take into account the potentially lower exposures that may result after initiation of Actemra (due to normalization of CYP3A4) or higher exposures after discontinuation of Actemra.

**Reference:**

Actemra US prescribing information

Kim S. Rheumatol Int (2012) 32:2601 – 2604.

Schmitt C, et al. Disease-drug-drug interaction involving tocilizumab and simvastatin in patients with rheumatoid arthritis. Clin Pharmacol Ther 2011; 89(5): 735–740.

Zhang X, Schmitt C, Grange S, Terao K, Miya K. 2009. Disease-drug interaction studies of tocilizumab with cytochrome p450 substrate in vitro and in vivo. Clin. Pharmacol. Ther. 85:S59 (Abstr.)

---

**Drug-Drug Interaction Monograph for ACTEMRA (Tocilizumab) #5**

---

**Perpetrator drug:**

| | |
|---|---|
| INN name/US brand name: | Tocilizumab/Actemra |
| Class/Molecular weight: | Monoclonal Antibody/~148 kDa |

**Victim drug:**

| | |
|---|---|
| INN name/US brand name: | Omeprazole/Prilosec |
| Class/Molecular weight: | Small Molecule/345 Da |

**Description of the effect of the DDI:**
In RA patients receiving 10 mg omeprazole, exposure to omeprazole was approximately 2 fold higher than that observed in healthy subjects. In RA patients receiving 10 mg omeprazole, before and one week after Actemra infusion (8 mg/kg), the omeprazole $AUC_{inf}$ decreased by 12% for poor (N = 5) and intermediate metabolizers (N = 5) and by 28% for extensive metabolizers (N = 8) and were slightly higher than those observed in healthy subjects.

**Description of the (proven, putative or hypothesized) mechanism of the DDI:**
Omeprazole is a CYP2C19 and CYP3A4 substrate. *In vitro* data suggested that IL-6 reduced mRNA expression for several CYP450 isoenzymes including CYP1A2, CYP2B6, CYP2C9, CYP2C19, CYP2D6 and CYP3A4, and this reduced expression was reversed by co-incubation with tocilizumab at clinically relevant concentrations. Accordingly, inhibition of IL-6 signaling in RA patients treated with tocilizumab may restore CYP450 activities to higher levels than those in the absence of tocilizumab leading to increased metabolism of drugs that are CYP450 substrates.

**Clinical consequences:**
None specifically mentioned for omeprazole.

**Reference:**

Actemra US prescribing information

Zhou H and Mascelli M, Annu. Mechanisms of Monoclonal Antibody-Drug Interactions. Rev. Pharmacol. Toxicol. 2011. 51:359–72.

---

**Drug-Drug Interaction Monograph for ACTEMRA (Tocilizumab) #6**

**Perpetrator drug:**

| | |
|---|---|
| INN name/US brand name: | Tocilizumab/Actemra |
| Class/Molecular weight: | Monoclonal Antibody/~148 kDa |

**Victim drug:**

| | |
|---|---|
| INN name/US brand name: | Dextromethorphan/Robitussin, Delsym |
| Class/Molecular weight: | Small Molecule/271 Da |

**Description of the effect of the DDI:**
In 13 RA patients receiving 30 mg dextromethorphan, exposure to dextromethorphan was comparable to that in healthy subjects. However, exposure to its metabolite, dextrorphan (a CYP3A4 substrate), was a fraction of that observed in healthy subjects. One week following administration of a single infusion of Actemra (8 mg/kg), dextromethorphan exposure was decreased by approximately 5%. However, a larger decrease (29%) in dextrorphan levels was noted after Actemra infusion.

**Description of the (proven, putative or hypothesized) mechanism of the DDI:**
Dextromethorphan is a CYP2D6 and CYP3A4 substrate. *In vitro* data suggested that IL-6 reduced mRNA expression for several CYP450 isoenzymes including CYP1A2, CYP2B6, CYP2C9, CYP2C19, CYP2D6 and CYP3A4, and this reduced expression was reversed by co-incubation with tocilizumab at clinically relevant concentrations. Accordingly, inhibition of IL-6 signaling in RA patients treated with tocilizumab may restore CYP450 activities to higher levels than those in the absence of tocilizumab leading to increased metabolism of drugs that are CYP450 substrates.

**Clinical consequences:**
None specifically mentioned for dextromethorphan.

**Reference:**

Actemra US prescribing information

Zhang X, Schmitt C, Grange S, Terao K, Miya K. 2009. Disease-drug interaction studies of tocilizumab with cytochrome p450 substrate in vitro and in vivo. Clin. Pharmacol. Ther. 85:S59 (Abstr.)

**Drug-Drug Interaction Monograph for ACTIMMUNE (Interferon Gamma-1B)**

**Perpetrator drug:**

| | |
|---|---|
| INN name/US brand name: | Interferon Gamma-1B/Actimmune |
| Class/Molecular weight: | Polypeptide/~17 kDa |

**Victim drug:**

| | |
|---|---|
| INN name/US brand name: | CYP450 substrates |
| Class/Molecular weight: | Small Molecule |

**Description of the effect of the DDI:**
Depression of hepatic metabolism.

**Description of the (proven, putative or hypothesized) mechanism of the DDI:**
Preclinical studies in rodents demonstrated a decrease in hepatic microsomal CYP450 concentrations.

*(Continued)*

**Clinical consequences:**
None mentioned.

**Reference:**

Actimmune prescribing information

Mahmood I, Green MD. Drug interaction studies of therapeutic proteins or monoclonal antibodies. J Clin Pharmacol. 2007 Dec;47(12):1540–54.

---

**Drug-Drug Interaction Monograph for ADCETRIS (Brentuximab vedotin) #1**

---

**Perpetrator drug:**

| | |
|---|---|
| INN name/US brand name: | Brentuximab vedotin/Adcetris |
| Class/Molecular weight: | Antibody-drug conjugate/~153 kDa |

**Victim drug:**

| | |
|---|---|
| INN name/US brand name: | Midazolam or other CYP3A4 substrates |
| Class/Molecular weight: | Small molecule/326 Da |

**Description of the effect of the DDI:**
Co-administration of Adcetris did not affect exposure to midazolam, a CYP3A4 substrate. Monomethyl auristatin E (MMAE) does not inhibit other CYP enzymes at relevant clinical concentrations. Adcetris is not expected to alter the exposure to drugs that are metabolized by CYP3A4 enzymes.

**Description of the (proven, putative or hypothesized) mechanism of the DDI:**
None.

**Clinical consequences:**
None mentioned.

**Reference:**
Adcetris prescribing information

---

**Drug-Drug Interaction Monograph for ADCETRIS (Brentuximab vedotin) #2**

---

**Perpetrator drug:**

| | |
|---|---|
| INN name/US brand name: | Ketoconazole/Nizoral |
| Class/Molecular weight: | Small molecule/531 Da |

**Victim drug:**

| | |
|---|---|
| INN name/US brand name: | Brentuximab vedotin/Adcetris |
| Class/Molecular weight: | Antibody-drug conjugate/~153 kDa |

**Description of the effect of the DDI:**
Co-administration of Adcetris with ketoconazole, a potent CYP3A4 inhibitor, increased exposure to MMAE by approximately 34%.

**Description of the (proven, putative or hypothesized) mechanism of the DDI:**
Adcetris is a CD30-directed antibody-drug conjugate (ADC) consisting of three components: 1) the chimeric IgG1 antibody cAC10, specific for human CD30, 2) the microtubule

disrupting agent monomethyl auristatin E (MMAE), and 3) a protease-cleavable linker that covalently attaches MMAE to cAC10. MMAE is primarily metabolized by CYP3A.

**Clinical consequences:**
Patients who are receiving strong CYP3A4 inhibitors concomitantly with Adcetris should be closely monitored for adverse reactions.

**Reference:**
Adcetris prescribing information

---

**Drug-Drug Interaction Monograph for ADCETRIS (Brentuximab vedotin) #3**

---

**Perpetrator drug:**

| | |
|---|---|
| INN name/US brand name: | Rifampin/Rifadin, Rimactane |
| Class/Molecular weight: | Small molecule/823 Da |

**Victim drug:**

| | |
|---|---|
| INN name/US brand name: | Brentuximab vedotin/Adcetris |
| Class/Molecular weight: | antibody-drug conjugate/~153 kDa |

**Description of the effect of the DDI:**
Co-administration of Adcetris with rifampin, a potent CYP3A4 inducer, reduced exposure to MMAE by approximately 46%.

**Description of the (proven, putative or hypothesized) mechanism of the DDI:**
Adcetris is a CD30-directed antibody-drug conjugate (ADC) consisting of three components: 1) the chimeric IgG1 antibody cAC10, specific for human CD30, 2) the microtubule disrupting agent monomethyl auristatin E (MMAE), and 3) a protease-cleavable linker that covalently attaches MMAE to cAC10. MMAE is primarily metabolized by CYP3A.

**Clinical consequences:**
None mentioned.

**Reference:**
Adcetris prescribing information

---

**Drug-Drug Interaction Monograph for AMEVIVE (Alefacept)**

---

**Perpetrator drug:**

| | |
|---|---|
| INN name/US brand name: | Alefacept/Amevive |
| Class/Molecular weight: | Fc fusion protein/91.4 kDa |

**Victim drug:**

| | |
|---|---|
| INN name/US brand name: | CYP450 substrates |
| Class/Molecular weight: | Small molecule |

**Description of the effect of the DDI:**
No drug interaction studies have been conducted.

*(Continued)*

**Description of the (proven, putative or hypothesized) mechanism of the DDI:**
The formation of CYP450 enzymes may be suppressed by increased levels of cytokines (e.g., TNFα, IL-1, IL-6, IL-10, IFN) during chronic inflammation. Therefore, a molecule that exerts its pharmacological effect by inhibiting the release of cytokines, such as Amevive, could normalize the formation of CYP450 enzymes.

**Clinical consequences:**
Upon initiation or discontinuation of Amevive in patients being treated with CYP450 substrates with a narrow therapeutic index, monitoring of the effect (e.g., warfarin) or drug concentration (e.g., cyclosporine or theophylline) is recommended and the individual dose of the drug product may be adjusted as needed.

**Reference:**
Amevive prescribing information

Lee JI, Zhang L, Men AY, Kenna LA, Huang SM. CYP-mediated therapeutic protein-drug interactions: clinical findings, proposed mechanisms and regulatory implications. Clin Pharmacokinet. 2010 May;49(5):295–310.

---

**Drug-Drug Interaction Monograph for ANGIOMAX (Bivalirudin)**

---

**Perpetrator drug:**

| | |
|---|---|
| INN name/US brand name: | Bivalirudin/Angiomax |
| Class/Molecular weight: | Peptide/2.2 kDa |

**Victim drug:**

| | |
|---|---|
| INN name/US brand name: | Anti-coagulants (e.g., heparin, warfarin, thrombolytics, glycoprotein IIb/IIIa-receptor inhibitors (GPI)) |
| Class/Molecular weight: | Small Molecule |

**Description of the effect of the DDI:**
In clinical trials in patients undergoing Percutaneous Coronary Intervention (PCI) or Percutaneous Transluminal Coronary Angioplasty (PTCA), co-administration of Angiomax with heparin, warfarin, thrombolytics, or GPIs was associated with increased risks of major bleeding events compared to patients not receiving these concomitant medications.

**Description of the (proven, putative or hypothesized) mechanism of the DDI:**
Bivalirudin is a direct thrombin inhibitor used as an anticoagulant. Thus use with other anticoagulants can increase the risk of bleeding events.

**Clinical consequences:**
None mentioned in label.

**Reference:**
Angiomax prescribing information

Kastrati A, Neumann FJ, Mehilli J et al: Bivalirudin versus unfractionated heparin during percutaneous coronary intervention. N Engl J Med (2008) 359 (7/Aug 14): 688–96.

---

**Drug-Drug Interaction Monograph for APIDRA, HUMULIN, HUMALOG, NOVOLOG, LANTUS (Insulin) #1**

**Perpetrator drug:**

INN name/US brand name: Oral antidiabetic products
Class/Molecular weight: Small molecule

**Victim drug:**

INN name/US brand name: Insulin/Apidra, Humulin, Humalog, Novolog, Lantus, etc
Class/Molecular weight: Hormone/~5.8 kDa

**Description of the effect of the DDI:**
Pharmacodynamic interaction: Drugs that may increase the blood glucose-lowering effect of insulins can increase the risk of hypoglycemia. Insulin may be used as adjunctive therapy with oral antihyperglycemic agents when glycemic control deteriorates on oral antidiabetic agents in patients with type 2 diabetes mellitus. However lower daily doses of insulin are required to maintain desirable glycemic control.

**Description of the (proven, putative or hypothesized) mechanism of the DDI:**
Both drugs reduce blood glucose.

**Clinical consequences:**
May require insulin dose adjustment and particularly close monitoring

**Reference:**
Apidra prescribing information, Humulin prescribing information, Humalog prescribing information, Lantus prescribing information, Novolog prescribing information

Kabadi MU, Kabadi UM. Efficacy of sulfonylureas with insulin in type 2 diabetes mellitus. Ann Pharmacother. 2003 Nov;37(11):1572–6.

**Drug-Drug Interaction Monograph for APIDRA, HUMULIN, HUMALOG, NOVOLOG, LANTUS (Insulin) #2**

**Perpetrator drug:**

INN name/US brand name: Pramlintide/Symlin
Class/Molecular weight: Peptide/4 kDa

**Victim drug:**

INN name/US brand name: Insulin/Apidra, Humulin, Humalog, Novolog, Lantus, etc
Class/Molecular weight: Hormone/~5.8 kDa

**Description of the effect of the DDI:**
Pharmacodynamic interaction: Drugs that may increase the blood glucose-lowering effect of insulins can increase the risk of hypoglycemia. Pramlintide is used as an adjunct to insulin and has been associated with an increased risk of insulin-induced severe hypoglycemia, particularly in patients with type 1 diabetes.

**Description of the (proven, putative or hypothesized) mechanism of the DDI:**
Both drugs reduce blood glucose.

*(Continued)*

**Clinical consequences:**
May require insulin dose adjustment and particularly close monitoring. To minimize the risk of insulin-induced hypoglycemia, all patients should reduce premeal insulin by 50% upon initiation of pramlintide therapy, monitor blood glucose frequently, and contact their healthcare provider if symptoms of nausea and/or hypoglycemia are persistent. Careful patient selection, patient instruction, and insulin dose adjustments are critical for reducing the risk of hypoglycemia.

**Reference:**
Pullman J, Darsow T, Frias J. Pramlintide in the Management of Insulin-Using Patients with Type 2 and Type 1 Diabetes. Vasc Health Risk Manag. 2006 September; 2(3): 203–212.

Apidra prescribing information, Humulin prescribing information, Humalog prescribing information, Lantus prescribing information, Novolog prescribing information
Symlin prescribing information.

---

**Drug-Drug Interaction Monograph for APIDRA, HUMULIN, HUMALOG, NOVOLOG, LANTUS (Insulin) #3**

---

**Perpetrator drug:**

| | |
|---|---|
| INN name/US brand name: | ACE Inhibitors (e.g. lisinopril, enalapril, captopril) |
| Class/Molecular weight: | Small molecule/peptides |

**Victim drug:**

| | |
|---|---|
| INN name/US brand name: | Insulin/Apidra, Humulin, Humalog, Novolog, Lantus, etc |
| Class/Molecular weight: | Hormone/~5.8 kDa |

**Description of the effect of the DDI:**
Pharmacodynamic interaction: Drugs that may increase the blood glucose-lowering effect of insulins can increase the risk of hypoglycemia. A nested case-control study was performed with diabetic patients treated with insulin or with oral antidiabetic drugs, who were admitted to hospital with hypoglycemia. Hypoglycemia was significantly associated with current use of ACE inhibitors (odds ratio 2.8 [95% CI 1.4–5.7]). In another nested case-control study, hypoglycemia was associated with the use of ACE inhibitors (odds ratio [OR] 3.2, 95% CI 1.2–8.3, $P = 0.023$),

**Description of the (proven, putative or hypothesized) mechanism of the DDI:**
Insulin resistance appears to be a syndrome that is associated with a clustering of metabolic disorders, including non-insulin-dependent diabetes mellitus, obesity, hypertension, lipid abnormalities, and atherosclerotic cardiovascular disease. Thus if a facet of metabolic disorders is normalized, this may perhaps increase insulin sensitivity

**Clinical consequences:**
May require insulin dose adjustment and particularly close monitoring

**Reference:**
Herings RM, de Boer A, Stricker BH, Leufkens HG, Porsius A. c. Hypoglycemia associated with use of inhibitors of angiotensin converting enzyme. Lancet. 1995;345(8959):1195–8.

DeFronzo R and Ferrannini E. Insulin resistance: a multifaceted syndrome responsible for NIDDM, obesity, hypertension, dyslipidemia, and atherosclerotic cardiovascular disease. Diabetes Care 1991;14:173–194.

Morris AD, Boyle DI, McMahon AD, Pearce H, Evans JM, Newton RW, Jung RT, MacDonald TM. ACE inhibitor use is associated with hospitalization for severe hypoglycemia in patients with diabetes. DARTS/MEMO Collaboration. Diabetes Audit and Research in Tayside, Scotland. Medicines Monitoring Unit. Diabetes Care. 1997 Sep; 20(9):1363–7.

Apidra prescribing information, Humulin prescribing information, Humalog prescribing information, Lantus prescribing information, Novolog prescribing information.

---

**Drug-Drug Interaction Monograph for APIDRA, HUMULIN, HUMALOG, NOVOLOG, LANTUS (Insulin) #4**

---

**Perpetrator drug:**

| | |
|---|---|
| INN name/US brand name: | Disopyramide/Norpace, Rythmodan |
| Class/Molecular weight: | Small molecule/339 Da |

**Victim drug:**

| | |
|---|---|
| INN name/US brand name: | Insulin/Apidra, Humulin, Humalog, Novolog, Lantus, etc |
| Class/Molecular weight: | Hormone/~5.8 kDa |

**Description of the effect of the DDI:**
Pharmacodynamic interaction: Drugs that may increase the blood glucose-lowering effect of insulins can increase the risk of hypoglycemia. Case reports describe disopyramide-induced hypoglycemia as well as hypoglycemia on disopyramide treatment with concomitant insulin.

**Description of the (proven, putative or hypothesized) mechanism of the DDI:**
Both drugs have the potential to cause hypoglycemia.

**Clinical consequences:**
May require insulin dose adjustment and particularly close monitoring

**Reference:**
Cacoub P, Deray G, Baumelou A, Grimaldi A, Soubrie C, Jacobs C. Disopyramide-induced hypoglycemia: case report and review of the literature. Fundam Clin Pharmacol. 1989; 3(5):527–35.

Taketa K, Yamamoto Y. Hypoglycemic effect of disopyramide in a case of diabetes mellitus under insulin treatment. Acta Med Okayama. 1980 Sep;34(4):289–92.Croxson MS, Shaw DW, Henley PG, Gabriel HD. Disopyramide-induced hypoglycaemia and increased serum insulin.

Apidra prescribing information, Humulin prescribing information, Humalog prescribing information, Lantus prescribing information, Novolog prescribing information.

---

## Drug-Drug Interaction Monograph for APIDRA, HUMULIN, HUMALOG, NOVOLOG, LANTUS (Insulin) #5

**Perpetrator drug:**

INN name/US brand name: Fibrates (e.g. clofibrate, gemfibrozil, fenofibrate)
Class/Molecular weight: Small molecule

**Victim drug:**

INN name/US brand name: Insulin/Apidra, Humulin, Humalog, Novolog, Lantus, etc
Class/Molecular weight: Hormone/~5.8 kDa

**Description of the effect of the DDI:**
Pharmacodynamic interaction: Drugs that may increase the blood glucose-lowering effect of insulins can increase the risk of hypoglycemia. In a case report where the patient added gemfibrozil to her medications, the patient experienced recurrent hypoglycemic episodes concurrant with a massive reduction in insulin requirements.

**Description of the (proven, putative or hypothesized) mechanism of the DDI:**
Insulin resistance appears to be a syndrome that is associated with a clustering of metabolic disorders, including non-insulin-dependent diabetes mellitus, obesity, hypertension, lipid abnormalities, and atherosclerotic cardiovascular disease. Thus if a facet of metabolic disorders is normalized, this may perhaps increase insulin sensitivity.

**Clinical consequences:**
May require insulin dose adjustment and particularly close monitoring.

**Reference:**
Klein J, Ott V, Schütt M, Klein HH. Recurrent hypoglycaemic episodes in a patient with Type 2 diabetes under fibrate therapy. J Diabetes Complications. 2002;16(3):246–8.

DeFronzo R and Ferrannini E. Insulin resistance: a multifaceted syndrome responsible for NIDDM, obesity, hypertension, dyslipidemia, and atherosclerotic cardiovascular disease. Diabetes Care 1991;14:173–194.

Apidra prescribing information, Humulin prescribing information, Humalog prescribing information, Lantus prescribing information, Novolog prescribing information.

## Drug-Drug Interaction Monograph for APIDRA, HUMULIN, HUMALOG, NOVOLOG, LANTUS (Insulin) #6

**Perpetrator drug:**

INN name/US brand name: Fluoxetine/Prozac
Class/Molecular weight: Small molecule/309 Da

**Victim drug:**

INN name/US brand name: Insulin/Apidra, Humulin, Humalog, Novolog, Lantus, etc
Class/Molecular weight: Hormone/~5.8 kDa

**Description of the effect of the DDI:**
Pharmacodynamic interaction: Drugs that may increase the blood glucose-lowering effect of insulins can increase the risk of hypoglycemia. Fluoxetine treatment was associated with a large decline in HbA1c and insulin levels in obese diabetics compared to placebo. Fasting glucose levels also declined significantly.

**Description of the (proven, putative or hypothesized) mechanism of the DDI:**
Serotonergic antidepressants appear to enhance insulin sensitivity, reduce hyperglycemia and facilitate the storage of glucose as glycogen in muscle tissue.

**Clinical consequences:**
May require insulin dose adjustment and particularly close monitoring.

**Reference:**
MA Deeg and EW Lipkin. Hypoglycemia associated with the use of fluoxetine. West J Med. 1996 March; 164(3): 262–263.

McIntyre RS, Soczynska JK, Konarski JZ, Kennedy SH. The effect of antidepressants on glucose homeostasis and insulin sensitivity: synthesis and mechanisms. Expert Opin Drug Saf. 2006 Jan;5(1):157–68.

Gray DS, Fujioka K, Devine W et al.: A randomized double-blind clinical trial of fluoxetine in obese diabetics.Int. J. Obes. Relat. Metab. Disord. (1992)16(Suppl. 4): S67–S72.

Maheux P, Ducros F, Bourque J et al.: Fluoxetine improves insulin sensitivity in obese patients with non-insulindependent diabetes mellitus independently of weight loss. Int. J. Obes. Relat. Metab. Disord. (1997) 21(2):97–102.

Apidra prescribing information, Humulin prescribing information, Humalog prescribing information, Lantus prescribing information, Novolog prescribing information.

---

**Drug-Drug Interaction Monograph for APIDRA, HUMULIN, HUMALOG, NOVOLOG, LANTUS (Insulin) #7**

---

**Perpetrator drug:**

| | |
|---|---|
| INN name/US brand name: | Monoamine oxidase inhibitors (e.g. hydralazine, phenelzine) |
| Class/Molecular weight: | Small molecule |

**Victim drug:**

| | |
|---|---|
| INN name/US brand name: | Insulin/Apidra, Humulin, Humalog, Novolog, Lantus, etc |
| Class/Molecular weight: | Hormone/~5.8 kDa |

**Description of the effect of the DDI:**
Pharmacodynamic interaction: Drugs that may increase the blood glucose-lowering effect of insulins can increase the risk of hypoglycemia. Depressed subjects responding to monoamine oxidase inhibitor (MAOI) therapy displayed a decrease in peak glucose concentration, a lower fasting blood glucose level and shorter duration of the hyperglycemic period after the administration of glucose ($p < 0.05$).

**Description of the (proven, putative or hypothesized) mechanism of the DDI:**
MAOIs are associated with hypoglycemia.

**Clinical consequences:**
May require insulin dose adjustment and particularly close monitoring.

**Reference:**
McIntyre RS, Soczynska JK, Konarski JZ, Kennedy SH. The effect of antidepressants on glucose homeostasis and insulin sensitivity: synthesis and mechanisms. Expert Opin Drug Saf. 2006;5(1):157–68.

*(Continued)*

Adnitt PI. Hypoglycemic action of monoamineoxidase inhibitors (MAOI'S). Diabetes. 1968;17(10):628–33.

Apidra prescribing information, Humulin prescribing information, Humalog prescribing information, Lantus prescribing information, Novolog prescribing information.

---

**Drug-Drug Interaction Monograph for APIDRA, HUMULIN, HUMALOG, NOVOLOG, LANTUS (Insulin) #8**

---

**Perpetrator drug:**

| | |
|---|---|
| INN name/US brand name: | Propoxyphene/Darvon |
| Class/Molecular weight: | Small molecule/339 Da |

**Victim drug:**

| | |
|---|---|
| INN name/US brand name: | Insulin/Apidra, Humulin, Humalog, Novolog, Lantus, etc |
| Class/Molecular weight: | Hormone/~5.8 kDa |

**Description of the effect of the DDI:**
Pharmacodynamic interaction: Drugs that may increase the blood glucose-lowering effect of insulins can increase the risk of hypoglycemia. Case reports of this interaction were reported in patients with chronic renal failure.

**Description of the (proven, putative or hypothesized) mechanism of the DDI:**
Both drugs reduce blood glucose.

**Clinical consequences:**
May require insulin dose adjustment and particularly close monitoring.

**Reference:**
Almirall J, Montoliu J, Torras A, Revert L Propoxyphene-Induced Hypoglycemia in a Patient with Chronic Renal Failure. Nephron 1989;53:273–275.

Shah P, Aniszweski J, Service FJ. Propoxyphene-induced hypoglycemia in renal failure. Endocrine Practice 2006:12: 170–173.

Apidra prescribing information, Humulin prescribing information, Humalog prescribing information, Lantus prescribing information, Novolog prescribing information.

---

**Drug-Drug Interaction Monograph for APIDRA, HUMULIN, HUMALOG, NOVOLOG, LANTUS (Insulin) #9**

---

**Perpetrator drug:**

| | |
|---|---|
| INN name/US brand name: | Pentoxifylline/Trental |
| Class/Molecular weight: | Small molecule/278 Da |

**Victim drug:**

| | |
|---|---|
| INN name/US brand name: | Insulin/Apidra, Humulin, Humalog, Novolog, Lantus, etc |
| Class/Molecular weight: | Hormone/~5.8 kDa |

**Description of the effect of the DDI:**
Pharmacodynamic interaction: Drugs that may increase the blood glucose-lowering effect of insulins can increase the risk of hypoglycemia.

**Description of the (proven, putative or hypothesized) mechanism of the DDI:**
Both drugs reduce blood glucose.

**Clinical consequences:**
May require insulin dose adjustment and particularly close monitoring.

**Reference:**
Apidra prescribing information, Humulin prescribing information, Humalog prescribing information, Lantus prescribing information, Novolog prescribing information.

---

**Drug-Drug Interaction Monograph for APIDRA, HUMULIN, HUMALOG, NOVOLOG, LANTUS (Insulin) #10**

---

**Perpetrator drug:**

| | |
|---|---|
| INN name/US brand name: | Salicylates (e.g. aspirin) |
| Class/Molecular weight: | Small molecule |

**Victim drug:**

| | |
|---|---|
| INN name/US brand name: | Insulin/Apidra, Humulin, Humalog, Novolog, Lantus, etc |
| Class/Molecular weight: | Hormone/~5.8 kDa |

**Description of the effect of the DDI:**
Pharmacodynamic interaction: Drugs that may increase the blood glucose-lowering effect of insulins can increase the risk of hypoglycemia.

**Description of the (proven, putative or hypothesized) mechanism of the DDI:**
Salicylates lower blood sugar and enhance glucose-stimulated insulin secretion in normal and diabetic subjects. Although the hypoglycemic effect appears to be mediated by enhanced insulin secretion, extrapancreatic mechanisms cannot be excluded. The mechanism of the enhanced insulin secretion appears to be mediated by prostaglandin synthesis inhibition.

**Clinical consequences:**
May require insulin dose adjustment and particularly close monitoring.

**Reference:**
Limbeck G, Ruvalcaba R, Samols E, Kelley V. Salicylates and Hypoglycemia. Archives of Pediatrics & Adolescent Medicine. 1965:109:165–167.

Baron S. Salicylates as Hypoglycemic Agents. Diabetes Care. 1982:5:64–71.

Apidra prescribing information, Humulin prescribing information, Humalog prescribing information, Lantus prescribing information, Novolog prescribing information.

---

## Drug-Drug Interaction Monograph for APIDRA, HUMULIN, HUMALOG, NOVOLOG, LANTUS (Insulin) #11

**Perpetrator drug:**

| | |
|---|---|
| INN name/US brand name: | Somatostatin analogs/(see Accretropin) |
| Class/Molecular weight: | Hormone |

**Victim drug:**

| | |
|---|---|
| INN name/US brand name: | Insulin/Apidra, Humulin, Humalog, Novolog, Lantus, etc |
| Class/Molecular weight: | Hormone/~5.8 kDa |

**Description of the effect of the DDI:**
Pharmacodynamic interaction: Drugs that may increase the blood glucose-lowering effect of insulins can increase the risk of hypoglycemia. Infusion of somatostatin (combined deficiency of glucagon and growth hormone) accentuated insulin-induced hypoglycemia (plasma glucose nadir: 36+/−2 ng/dl during infusion of somatostatin vs. 47+/−2 mg/dl in control studies, $P < 0.01$) and impaired restoration of normoglycemia (plasma glucose at min 90: 73+/−3 mg/dl at end of somatostatin infusion vs. 92+/−3 mg/dl in control studies, $P < 0.01$).

**Description of the (proven, putative or hypothesized) mechanism of the DDI:**
Impaired recovery of plasma glucose was due to blunting of the compensatory increase in glucose appearance since glucose disappearance was not augmented, and was attributable to suppression of glucagon secretion.

**Clinical consequences:**
May require insulin dose adjustment and particularly close monitoring.

**Reference:**
Rizza R, Cryer P, and Gerich J. Role of glucagon, catecholamines, and growth hormone in human glucose counterregulation. J Clin Invest. 1979;64: 62–71.

Apidra prescribing information, Humulin prescribing information, Humalog prescribing information, Lantus prescribing information, Novolog prescribing information.

## Drug-Drug Interaction Monograph for APIDRA, HUMULIN, HUMALOG, NOVOLOG, LANTUS (Insulin) #12

**Perpetrator drug:**

| | |
|---|---|
| INN name/US brand name: | Sulfonamide antibiotics |
| Class/Molecular weight: | Small molecule |

**Victim drug:**

| | |
|---|---|
| INN name/US brand name: | Insulin/Apidra, Humulin, Humalog, Novolog, Lantus, etc |
| Class/Molecular weight: | Hormone/~5.8 kDa |

**Description of the effect of the DDI:**
Pharmacodynamic interaction: Drugs that may increase the blood glucose-lowering effect of insulins can increase the risk of hypoglycemia.

**Description of the (proven, putative or hypothesized) mechanism of the DDI:**
Because of their structural similarities to sulfonylureas, sulfonamides are liable to facilitate hypoglycemia by increasing insulin release in susceptible individuals.

**Clinical consequences:**
May require insulin dose adjustment and particularly close monitoring.

**Reference:**
Arem R, Garber A, Field J. Sulfonamide-induced hypoglycemia in chronic renal failure. Archives of Int Med. 1983;143:827–829.

Hekimsoy Z, Biberoğlu S, Cömlekçi A, Tarhan O, Mermut C, Biberoğlu K. Trimethoprim/sulfamethoxazole-induced hypoglycemia in a malnourished patient with severe infection. Eur J Endocrinol. 1997;136(3):304–6.

Apidra prescribing information, Humulin prescribing information, Humalog prescribing information, Lantus prescribing information, Novolog prescribing information.

---

**Drug-Drug Interaction Monograph for APIDRA, HUMULIN, HUMALOG, NOVOLOG, LANTUS (Insulin) #13**

---

**Perpetrator drug:**

| | |
|---|---|
| INN name/US brand name: | Corticosteroids |
| Class/Molecular weight: | Small molecule |

**Victim drug:**

| | |
|---|---|
| INN name/US brand name: | Insulin/Apidra, Humulin, Humalog, Novolog, Lantus, etc |
| Class/Molecular weight: | Hormone/~5.8 kDa |

**Description of the effect of the DDI:**
Pharmacodynamic interaction: Drugs that may reduce the blood glucose-lowering effect of insulins can increase the risk of hypoglycemia. Case report of hypoglycemic seizure occurring after high-dose inhaled corticosteroids.

**Description of the (proven, putative or hypothesized) mechanism of the DDI:**
None.

**Clinical consequences:**
May require insulin dose adjustment and particularly close monitoring.

**Reference:**
Dunlop KA, Carson DJ, Shields MD. Hypoglycemia due to adrenal suppression secondary to high-dose nebulized corticosteroid. Pediatr Pulmonol. 2002;34(1):85–6.

Apidra prescribing information, Humulin prescribing information, Humalog prescribing information, Lantus prescribing information, Novolog prescribing information.

---

**Drug-Drug Interaction Monograph for APIDRA, HUMULIN, HUMALOG, NOVOLOG, LANTUS (Insulin) #14**

---

**Perpetrator drug:**

| | |
|---|---|
| INN name/US brand name: | Niacin |
| Class/Molecular weight: | Small molecule/123 Da |

**Victim drug:**

| | |
|---|---|
| INN name/US brand name: | Insulin/Apidra, Humulin, Humalog, Novolog, Lantus, etc |
| Class/Molecular weight: | Hormone/~5.8 kDa |

*(Continued)*

**Description of the effect of the DDI:**
Pharmacodynamic interaction: Drugs that may reduce the blood glucose-lowering effect of insulins can increase the risk of hypoglycemia. Case reports of high dose niacin causing hypoglycemia.

**Description of the (proven, putative or hypothesized) mechanism of the DDI:**
None.

**Clinical consequences:**
May require insulin dose adjustment and particularly close monitoring.

**Reference:**
Arcinegas-Rodriguez S et al. Metabolic acidosis, hypoglycemia, and severe myalgias: an attempt to mask urine drug screen results. Pediatr Emerg Care. 2011;27(4): 315–7.

Apidra prescribing information, Humulin prescribing information, Humalog prescribing information, Lantus prescribing information, Novolog prescribing information.

---

**Drug-Drug Interaction Monograph for APIDRA, HUMULIN, HUMALOG, NOVOLOG, LANTUS (Insulin) #15**

---

**Perpetrator drug:**

| | |
|---|---|
| INN name/US brand name: | Danazol/Danocrine |
| Class/Molecular weight: | Small molecule/338 Da |

**Victim drug:**

| | |
|---|---|
| INN name/US brand name: | Insulin/Apidra, Humulin, Humalog, Novolog, Lantus, etc |
| Class/Molecular weight: | Hormone/~5.8 kDa |

**Description of the effect of the DDI:**
Pharmacodynamic interaction: Drugs that may reduce the blood glucose-lowering effect of insulins can increase the risk of hypoglycemia. In subjects treated with danazol for lupus thrombocytopenia, side effects including hypoglycemia were reported.

**Description of the (proven, putative or hypothesized) mechanism of the DDI:**
None.

**Clinical consequences:**
May require insulin dose adjustment and particularly close monitoring.

**Reference:**
Wong KL. Danazol in treatment of lupus thrombocytopenia. Asian Pac J Allergy Immunol. 1991;9(2):125–9.

Apidra prescribing information, Humulin prescribing information, Humalog prescribing information, Lantus prescribing information, Novolog prescribing information.

---

**Drug-Drug Interaction Monograph for APIDRA, HUMULIN, HUMALOG, NOVOLOG, LANTUS (Insulin) #16**

**Perpetrator drug:**

| | |
|---|---|
| INN name/US brand name: | Diuretics (e.g. furosemide, hydrochlorothiazide) |
| Class/Molecular weight: | Small molecule |

**Victim drug:**

| | |
|---|---|
| INN name/US brand name: | Insulin/Apidra, Humulin, Humalog, Novolog, Lantus, etc |
| Class/Molecular weight: | Hormone/~5.8 kDa |

**Description of the effect of the DDI:**
Pharmacodynamic interaction: Drugs that may reduce the blood glucose-lowering effect of insulins can increase the risk of hypoglycemia.

**Description of the (proven, putative or hypothesized) mechanism of the DDI:**
Insulin resistance appears to be a syndrome that is associated with a clustering of metabolic disorders, including non-insulin-dependent diabetes mellitus, obesity, hypertension, lipid abnormalities, and atherosclerotic cardiovascular disease. Thus if a facet of metabolic disorders is normalized, this may perhaps increase insulin sensitivity

**Clinical consequences:**
May require insulin dose adjustment and particularly close monitoring.

**Reference:**
DeFronzo R and Ferrannini E. Insulin resistance: a multifaceted syndrome responsible for NIDDM, obesity, hypertension, dyslipidemia, and atherosclerotic cardiovascular disease. Diabetes Care 1991;14:173–194.

Apidra prescribing information, Humulin prescribing information, Humalog prescribing information, Lantus prescribing information, Novolog prescribing information.

**Drug-Drug Interaction Monograph for APIDRA, HUMULIN, HUMALOG, NOVOLOG, LANTUS (Insulin) #17**

**Perpetrator drug:**

| | |
|---|---|
| INN name/US brand name: | Sympathomimetic agents (eg. epinephrine, albuterol, terbutaline) |
| Class/Molecular weight: | Small molecule |

**Victim drug:**

| | |
|---|---|
| INN name/US brand name: | Insulin/Apidra, Humulin, Humalog, Novolog, Lantus, etc |
| Class/Molecular weight: | Hormone/~5.8 kDa |

**Description of the effect of the DDI:**
Pharmacodynamic interaction: Drugs that may reduce the blood glucose-lowering effect of insulins can increase the risk of hypoglycemia.

*(Continued)*

**Description of the (proven, putative or hypothesized) mechanism of the DDI:**
None.

**Clinical consequences:**
May require insulin dose adjustment and particularly close monitoring.

**Reference:**
Apidra prescribing information, Humulin prescribing information, Humalog prescribing information, Lantus prescribing information, Novolog prescribing information.

---

**Drug-Drug Interaction Monograph for APIDRA, HUMULIN, HUMALOG, NOVOLOG, LANTUS (Insulin) #18**

---

**Perpetrator drug:**

| | |
|---|---|
| INN name/US brand name: | Glucagon |
| Class/Molecular weight: | Peptide hormone |

**Victim drug:**

| | |
|---|---|
| INN name/US brand name: | Insulin/Apidra, Humulin, Humalog, Novolog, Lantus, etc |
| Class/Molecular weight: | Hormone/~5.8 kDa |

**Description of the effect of the DDI:**
Pharmacodynamic interaction: Drugs that may reduce the blood glucose-lowering effect of insulins can increase the risk of hypoglycemia.

**Description of the (proven, putative or hypothesized) mechanism of the DDI:**
None.

**Clinical consequences:**
May require insulin dose adjustment and particularly close monitoring.

**Reference:**
Apidra prescribing information, Humulin prescribing information, Humalog prescribing information, Lantus prescribing information, Novolog prescribing information.

---

**Drug-Drug Interaction Monograph for APIDRA, HUMULIN, HUMALOG, NOVOLOG, LANTUS (Insulin) #19**

---

**Perpetrator drug:**

| | |
|---|---|
| INN name/US brand name: | Isoniazid/Laniazid, Nydrazid |
| Class/Molecular weight: | Small molecule/137 Da |

**Victim drug:**

| | |
|---|---|
| INN name/US brand name: | Insulin/Apidra, Humulin, Humalog, Novolog, Lantus, etc |
| Class/Molecular weight: | Hormone/~5.8 kDa |

**Description of the effect of the DDI:**
Pharmacodynamic interaction: Drugs that may reduce the blood glucose-lowering effect of insulins can increase the risk of hypoglycemia.

**Description of the (proven, putative or hypothesized) mechanism of the DDI:**
None.

**Clinical consequences:**
May require insulin dose adjustment and particularly close monitoring.

**Reference:**
Apidra prescribing information, Humulin prescribing information, Humalog prescribing information, Lantus prescribing information, Novolog prescribing information.

---

**Drug-Drug Interaction Monograph for APIDRA, HUMULIN, HUMALOG, NOVOLOG, LANTUS (Insulin) #20**

---

**Perpetrator drug:**

| | |
|---|---|
| INN name/US brand name: | Phenothiazine derivatives (e.g. chlorpromazine, thioridazine) |
| Class/Molecular weight: | Small molecule |

**Victim drug:**

| | |
|---|---|
| INN name/US brand name: | Insulin/Apidra, Humulin, Humalog, Novolog, Lantus, etc |
| Class/Molecular weight: | Hormone/~5.8 kDa |

**Description of the effect of the DDI:**
Pharmacodynamic interaction: Drugs that may reduce the blood glucose-lowering effect of insulins can increase the risk of hypoglycemia.

**Description of the (proven, putative or hypothesized) mechanism of the DDI:**
None.

**Clinical consequences:**
May require insulin dose adjustment and particularly close monitoring.

**Reference:**
Apidra prescribing information, Humulin prescribing information, Humalog prescribing information, Lantus prescribing information, Novolog prescribing information.

---

**Drug-Drug Interaction Monograph for APIDRA, HUMULIN, HUMALOG, NOVOLOG, LANTUS (Insulin) #21**

---

**Perpetrator drug:**

| | |
|---|---|
| INN name/US brand name: | Thyroid hormones |
| Class/Molecular weight: | Small molecule |

**Victim drug:**

| | |
|---|---|
| INN name/US brand name: | Insulin/Apidra, Humulin, Humalog, Novolog, Lantus, etc |
| Class/Molecular weight: | Hormone/~5.8 kDa |

**Description of the effect of the DDI:**
Pharmacodynamic interaction: Drugs that may reduce the blood glucose-lowering effect of insulins can increase the risk of hypoglycemia.

*(Continued)*

**Description of the (proven, putative or hypothesized) mechanism of the DDI:**
None.

**Clinical consequences:**
May require insulin dose adjustment and particularly close monitoring.

**Reference:**
Apidra prescribing information, Humulin prescribing information, Humalog prescribing information, Lantus prescribing information, Novolog prescribing information.

---

**Drug-Drug Interaction Monograph for APIDRA, HUMULIN, HUMALOG, NOVOLOG, LANTUS (Insulin) #22**

---

**Perpetrator drug:**

| | |
|---|---|
| INN name/US brand name: | Estrogens |
| Class/Molecular weight: | Small molecule |

**Victim drug:**

| | |
|---|---|
| INN name/US brand name: | Insulin/Apidra, Humulin, Humalog, Novolog, Lantus, etc |
| Class/Molecular weight: | Hormone/~5.8 kDa |

**Description of the effect of the DDI:**
Pharmacodynamic interaction: Drugs that may reduce the blood glucose-lowering effect of insulins can increase the risk of hypoglycemia.

**Description of the (proven, putative or hypothesized) mechanism of the DDI:**
none.

**Clinical consequences:**
May require insulin dose adjustment and particularly close monitoring.

**Reference:**
Apidra prescribing information, Humulin prescribing information, Humalog prescribing information, Lantus prescribing information, Novolog prescribing information.

---

**Drug-Drug Interaction Monograph for APIDRA, HUMULIN, HUMALOG, NOVOLOG, LANTUS (Insulin) #23**

---

**Perpetrator drug:**

| | |
|---|---|
| INN name/US brand name: | Progestins (e.g., in oral contraceptics) |
| Class/Molecular weight: | Small molecule |

**Victim drug:**

| | |
|---|---|
| INN name/US brand name: | Insulin/Apidra, Humulin, Humalog, Novolog, Lantus, etc |
| Class/Molecular weight: | Hormone/~5.8 kDa |

**Description of the effect of the DDI:**
Pharmacodynamic interaction: Drugs that may reduce the blood glucose-lowering effect of insulins can increase the risk of hypoglycemia.

**Description of the (proven, putative or hypothesized) mechanism of the DDI:**
None.

**Clinical consequences:**
May require insulin dose adjustment and particularly close monitoring.

**Reference:**
Apidra prescribing information, Humulin prescribing information, Humalog prescribing information, Lantus prescribing information, Novolog prescribing information.

---

**Drug-Drug Interaction Monograph for APIDRA, HUMULIN, HUMALOG, NOVOLOG, LANTUS (Insulin) #24**

---

**Perpetrator drug:**

| | |
|---|---|
| INN name/US brand name: | Protease inhibitors |
| Class/Molecular weight: | Small molecule |

**Victim drug:**

| | |
|---|---|
| INN name/US brand name: | Insulin/Apidra, Humulin, Humalog, Novolog, Lantus, etc |
| Class/Molecular weight: | Hormone/~5.8 kDa |

**Description of the effect of the DDI:**
Pharmacodynamic interaction: Drugs that may reduce the blood glucose-lowering effect of insulins can increase the risk of hypoglycemia.

**Description of the (proven, putative or hypothesized) mechanism of the DDI:**
None.

**Clinical consequences:**
May require insulin dose adjustment and particularly close monitoring.

**Reference:**
Apidra prescribing information, Humulin prescribing information, Humalog prescribing information, Lantus prescribing information, Novolog prescribing information.

---

**Drug-Drug Interaction Monograph for APIDRA, HUMULIN, HUMALOG, NOVOLOG, LANTUS (Insulin) #25**

---

**Perpetrator drug:**

| | |
|---|---|
| INN name/US brand name: | Atypical antipsychotics |
| Class/Molecular weight: | Small molecule |

**Victim drug:**

| | |
|---|---|
| INN name/US brand name: | Insulin/Apidra, Humulin, Humalog, Novolog, Lantus, etc |
| Class/Molecular weight: | Hormone/~5.8 kDa |

**Description of the effect of the DDI:**
Pharmacodynamic interaction: Drugs that may reduce the blood glucose-lowering effect of insulins can increase the risk of hypoglycemia.

*(Continued)*

**Description of the (proven, putative or hypothesized) mechanism of the DDI:**
none.

**Clinical consequences:**
May require insulin dose adjustment and particularly close monitoring.

**Reference:**
Apidra prescribing information, Humulin prescribing information, Humalog prescribing information, Lantus prescribing information, Novolog prescribing information.

---

**Drug-Drug Interaction Monograph for APIDRA, HUMULIN, HUMALOG, NOVOLOG, LANTUS (Insulin) #26**

---

**Perpetrator drug:**

| | |
|---|---|
| INN name/US brand name: | Beta-blockers |
| Class/Molecular weight: | Small molecule |

**Victim drug:**

| | |
|---|---|
| INN name/US brand name: | Insulin/Apidra, Humulin, Humalog, Novolog, Lantus, etc |
| Class/Molecular weight: | Hormone/~5.8 kDa |

**Description of the effect of the DDI:**
Pharmacodynamic interaction: May either increase or decrease the blood-glucose-lovwering effect of insulin.

**Description of the (proven, putative or hypothesized) mechanism of the DDI:**
None.

**Clinical consequences:**
May require insulin dose adjustment and particularly close monitoring.

**Reference:**
Apidra prescribing information, Humulin prescribing information, Humalog prescribing information, Lantus prescribing information, Novolog prescribing information.

---

**Drug-Drug Interaction Monograph for APIDRA, HUMULIN, HUMALOG, NOVOLOG, LANTUS (Insulin) #27**

---

**Perpetrator drug:**

| | |
|---|---|
| INN name/US brand name: | Clonidine/Catapres |
| Class/Molecular weight: | Small molecule/230 Da |

**Victim drug:**

| | |
|---|---|
| INN name/US brand name: | Insulin/Apidra, Humulin, Humalog, Novolog, Lantus, etc |
| Class/Molecular weight: | Hormone/~5.8 kDa |

**Description of the effect of the DDI:**
Pharmacodynamic interaction: May either increase or decrease the blood-glucose-lovwering effect of insulin.

**Description of the (proven, putative or hypothesized) mechanism of the DDI:**
None.

**Clinical consequences:**
May require insulin dose adjustment and particularly close monitoring.

**Reference:**
Apidra USPI, Feb 2009, Humulin USPI, May 2011, Humalog USPI, Oct 2012, Lantus USPI, Mar Apidra prescribing information, Humulin prescribing information, Humalog prescribing information, Lantus prescribing information, Novolog prescribing information.

---

**Drug-Drug Interaction Monograph for APIDRA, HUMULIN, HUMALOG, NOVOLOG, LANTUS (Insulin) #28**

---

**Perpetrator drug:**
INN name/US brand name: Lithium salts
Class/Molecular weight: Small molecule

**Victim drug:**
INN name/US brand name: Insulin/Apidra, Humulin, Humalog, Novolog, Lantus, etc
Class/Molecular weight: Hormone/~5.8 kDa

**Description of the effect of the DDI:**
Pharmacodynamic interaction: May either increase or decrease the blood-glucose-lovwering effect of insulin.

**Description of the (proven, putative or hypothesized) mechanism of the DDI:**
None.

**Clinical consequences:**
May require insulin dose adjustment and particularly close monitoring.

**Reference:**
Apidra prescribing information, Humulin prescribing information, Humalog prescribing information, Lantus prescribing information, Novolog prescribing information.

---

**Drug-Drug Interaction Monograph for APIDRA, HUMULIN, HUMALOG, NOVOLOG, LANTUS (Insulin) #29**

---

**Perpetrator drug:**
INN name/US brand name: Alcohol
Class/Molecular weight: Small molecule

**Victim drug:**
INN name/US brand name: Insulin/Apidra, Humulin, Humalog, Novolog, Lantus, etc
Class/Molecular weight: Hormone/~5.8 kDa

**Description of the effect of the DDI:**
Pharmacodynamic interaction: May either increase or decrease the blood-glucose-lowering effect of insulin.

*(Continued)*

**Description of the (proven, putative or hypothesized) mechanism of the DDI:**
None.

**Clinical consequences:**
May require insulin dose adjustment and particularly close monitoring.

**Reference:**
Apidra prescribing information, Humulin prescribing information, Humalog prescribing information, Lantus prescribing information, Novolog prescribing information.

---

**Drug-Drug Interaction Monograph for ARCALYST (Rilonacept) #1**

---

**Perpetrator drug:**
INN name/US brand name: TNF-blocking agents
Class/Molecular weight:
**Victim drug:**
INN name/US brand name: Rilonacept/Arcalyst
Class/Molecular weight: Fusion protein/251 kDa

**Description of the effect of the DDI:**
Pharmacodynamic interaction: Concomitant administration of another drug that blocks IL-1 with a TNF-blocking agent in another patient population has been associated with an increased risk of serious infections and an increased risk of neutropenia. Concomitant administration of rilonacept with TNF-blocking agents may also result in similar toxicities and is not recommended.

**Description of the (proven, putative or hypothesized) mechanism of the DDI:**
Both are immunomodulators which carry the risk of increased infections.

**Clinical consequences:**
Increased risk of serious infections and neutropenia.

**Reference:**
Arcalyst prescribing information.

---

**Drug-Drug Interaction Monograph for ARCALYST (Rilonacept) #2**

---

**Perpetrator drug:**
INN name/US brand name: Rilonacept/Arcalyst
Class/Molecular weight: Fusion protein/251 kDa

**Victim drug:**
INN name/US brand name: CYP450 substrates
Class/Molecular weight:

**Description of the effect of the DDI:**
Normalization of CYP450 enzymes.

**Description of the (proven, putative or hypothesized) mechanism of the DDI:**
The formation of CYP450 enzymes is suppressed by increased levels of cytokines (e.g., IL-1) during chronic inflammation. Thus it is expected that for a molecule that binds to IL-1, such as rilonacept, the formation of CYP450 enzymes could be normalized.

**Clinical consequences:**
Clinically relevant for CYP450 substrates with narrow therapeutic index. Upon initiation of rilonacept, in patients being treated with these types of medicinal products, therapeutic monitoring of the effect or drug concentration should be performed and the individual dose of the medicinal product may need to be adjusted as needed.

**Reference:**
Arcalyst prescribing information

Lee JI, Zhang L, Men AY, Kenna LA, Huang SM. CYP-mediated therapeutic protein-drug interactions: clinical findings, proposed mechanisms and regulatory implications. Clin Pharmacokinet. 2010;49(5):295–310.

---

**Drug-Drug Interaction Monograph for AVASTIN (Bevacizumab) #1**

---

**Perpetrator drug:**

| | |
|---|---|
| INN name/US brand name: | Bevacizumab/Avastin |
| Class/Molecular weight: | Monoclonal antibody/149 kDa |

**Victim drug:**

| | |
|---|---|
| INN name/US brand name: | Irinotecan/Camptosar, Campto |
| Class/Molecular weight: | Small molecule/587–677 Da |

**Description of the effect of the DDI:**
No significant effect of bevacizumab on the pharmacokinetics of irinotecan or its active metabolite SN38.

**Description of the (proven, putative or hypothesized) mechanism of the DDI:**
None.

**Clinical consequences:**
None.

**Reference:**
Horita Y, Yamada Y, Hirashima Y, Kato K, Nakajima T, Hamaguchi T, Shimada Y. Effects of bevacizumab on plasma concentration of irinotecan and its metabolites in advanced colorectal cancer patients receiving FOLFIRI with bevacizumab as second-line chemotherapy. Cancer Chemother Pharmacol. 2010;65(3):467–71. Epub 2009 Jun 25.

Avastin prescribing information.

---

**Drug-Drug Interaction Monograph for AVASTIN (Bevacizumab) #2**

---

**Perpetrator drug:**

| | |
|---|---|
| INN name/US brand name: | Bevacizumab/Avastin |
| Class/Molecular weight: | Monoclonal antibody/149 kDa |

**Victim drug:**

| | |
|---|---|
| INN name/US brand name: | Carboplatin/Paraplatin |
| Class/Molecular weight: | Small molecule/371 Da |

*(Continued)*

**Description of the effect of the DDI:**
Based on limited data, there did not appear to be a difference in the mean exposure of carboplatin when each was administered alone or in combination with bevacizumab.

**Description of the (proven, putative or hypothesized) mechanism of the DDI:**
None.

**Clinical consequences:**
None.

**Reference:**
Avastin prescribing information.

---

**Drug-Drug Interaction Monograph for AVASTIN (Bevacizumab) #3**

---

**Perpetrator drug:**

| | |
|---|---|
| INN name/US brand name: | Bevacizumab/Avastin |
| Class/Molecular weight: | Monoclonal antibody/149 kDa |

**Victim drug:**

| | |
|---|---|
| INN name/US brand name: | Paclitaxel/Taxol |
| Class/Molecular weight: | Small Molecule/854 Da |

**Description of the effect of the DDI:**
Based on limited data, there did not appear to be a difference in the mean exposure of paclitaxel when each was administered alone or in combination with bevacizumab. However, 3 of the 8 patients receiving bevacizumab plus paclitaxel/carboplatin had substantially lower paclitaxel exposure after 4 cycles of treatment (Day 63) than those at Day 0. Patients receiving paclitaxel/carbopatin without Avastin had a greater paclitaxel exposure at Day 63 than at Day 0.

**Description of the (proven, putative or hypothesized) mechanism of the DDI:**
None.

**Clinical consequences:**
None.

**Reference:**
Avastin prescribing information.

---

**Drug-Drug Interaction Monograph for AVASTIN (Bevacizumab) #4**

---

**Perpetrator drug:**

| | |
|---|---|
| INN name/US brand name: | Bevacizumab/Avastin |
| Class/Molecular weight: | Monoclonal antibody/149 kDa |

**Victim drug:**

| | |
|---|---|
| INN name/US brand name: | Interferon alfa/Multiferon |
| Class/Molecular weight: | Interferon |

**Description of the effect of the DDI:**
No difference in the mean exposure of interferon alfa when administered alone or in combination with bevacizumab.

**Description of the (proven, putative or hypothesized) mechanism of the DDI:**
None.

**Clinical consequences:**
None.

**Reference:**
Avastin prescribing information.

---

**Drug-Drug Interaction Monograph for BENLYSTA (Belimumab) #1**

---

**Perpetrator drug:**
INN name/US brand name: Corticosteroids
Class/Molecular weight: Small molecule

**Victim drug:**
INN name/US brand name: Belimumab/Benlysta
Class/Molecular weight: Monoclonal antibody/147 kDa

**Description of the effect of the DDI:**
Formal drug interaction studies have not been performed. In clinical trials of patients with systemic lupus erythematosus (SLE), there was no evidence of a clinically meaningful effect of concomitant medications on belimumab pharmacokinetics.

**Description of the (proven, putative or hypothesized) mechanism of the DDI:**
None.

**Clinical consequences:**
None.

**Reference:**
Benlysta prescribing information.

---

**Drug-Drug Interaction Monograph for BENLYSTA (Belimumab) #2**

---

**Perpetrator drug:**
INN name/US brand name: Antimalarials
Class/Molecular weight: Small molecule

**Victim drug:**
INN name/US brand name: Belimumab/Benlysta
Class/Molecular weight: Monoclonal antibody/147 kDa

**Description of the effect of the DDI:**
Formal drug interaction studies have not been performed. In clinical trials of patients with systemic lupus erythematosus (SLE), there was no evidence of a clinically meaningful effect of concomitant antimalarials on belimumab pharmacokinetics.

**Description of the (proven, putative or hypothesized) mechanism of the DDI:**
None.

**Clinical consequences:**
None.

**Reference:**
Benlysta prescribing information.

---

**Drug-Drug Interaction Monograph for BENLYSTA (Belimumab) #3**

**Perpetrator drug:**

| | |
|---|---|
| INN name/US brand name: | Immunomodulatory and immunosuppressive agents (incl azathioprine, methotrexate, and mycophenolate) |
| Class/Molecular weight: | |

**Victim drug:**

| | |
|---|---|
| INN name/US brand name: | Belimumab/Benlysta |
| Class/Molecular weight: | Monoclonal antibody/147 kDa |

**Description of the effect of the DDI:**
Formal drug interaction studies have not been performed. In clinical trials of patients with systemic lupus erythematosus (SLE), there was no evidence of a clinically meaningful effect of concomitant immunomodulatory and immunosuppressive agents on belimumab pharmacokinetics.

**Description of the (proven, putative or hypothesized) mechanism of the DDI:**
None.

**Clinical consequences:**
None.

**Reference:**
Benlysta prescribing information.

**Drug-Drug Interaction Monograph for BENLYSTA (Belimumab) #4**

**Perpetrator drug:**

| | |
|---|---|
| INN name/US brand name: | Angiotensin pathway antihypertensives |
| Class/Molecular weight: | Small molecules |

**Victim drug:**

| | |
|---|---|
| INN name/US brand name: | Belimumab/Benlysta |
| Class/Molecular weight: | Monoclonal antibody/147 kDa |

**Description of the effect of the DDI:**
Formal drug interaction studies have not been performed. In clinical trials of patients with systemic lupus erythematosus (SLE), there was no evidence of a clinically meaningful effect of concomitant angiotensin pathway antihypertensives on belimumab pharmacokinetics.

**Description of the (proven, putative or hypothesized) mechanism of the DDI:**
None.

**Clinical consequences:**
None.

**Reference:**
Benlysta prescribing information.

**Drug-Drug Interaction Monograph for BENLYSTA (Belimumab) #5**

**Perpetrator drug:**
INN name/US brand name: HMG-CoA reductase inhibitors (statins)
Class/Molecular weight: Small molecules

**Victim drug:**
INN name/US brand name: Belimumab/Benlysta
Class/Molecular weight: Monoclonal antibody/147 kDa

**Description of the effect of the DDI:**
Formal drug interaction studies have not been performed. In clinical trials of patients with systemic lupus erythematosus (SLE), there was no evidence of a clinically meaningful effect of concomitant HMG-CoA reductase inhibitors on belimumab pharmacokinetics.

**Description of the (proven, putative or hypothesized) mechanism of the DDI:**
None.

**Clinical consequences:**
None.

**Reference:**
Benlysta prescribing information.

**Drug-Drug Interaction Monograph for BENLYSTA (Belimumab) #6**

**Perpetrator drug:**
INN name/US brand name: NSAIDS (e.g. ibuprofen)
Class/Molecular weight: Small molecules

**Victim drug:**
INN name/US brand name: Belimumab/Benlysta
Class/Molecular weight: Monoclonal antibody/147 kDa

**Description of the effect of the DDI:**
Formal drug interaction studies have not been performed. In clinical trials of patients with systemic lupus erythematosus (SLE), there was no evidence of a clinically meaningful effect of concomitant NSAIDS on belimumab pharmacokinetics.

**Description of the (proven, putative or hypothesized) mechanism of the DDI:**
None.

**Clinical consequences:**
None.

**Reference:**
Benlysta prescribing information.

**Drug-Drug Interaction Monograph for BIO-TROPIN/(Somatropin)**

See ACCRETROPIN (Somatropin) drug-drug interaction monographs #1– #4
Reference: Bio-Tropin US prescribing information

**Drug-Drug Interaction Monograph for BOTOX, DYSPORT (Onabotulinumtoxin A) #1**

**Perpetrator drug:**

| | |
|---|---|
| INN name/US brand name: | Aminoglycosides or other agents interfering with neuromuscular transmission (curare-like compounds) |
| Class/Molecular weight: | Small molecules |

**Victim drug:**

| | |
|---|---|
| INN name/US brand name: | Onabotulinumtoxin A/Botox, Dysport |
| Class/Molecular weight: | Protein Toxin/149.3 kDa |

**Description of the effect of the DDI:**
<Note: Besides cosmetic uses, Botox is also used for the treatment of destusor overactivity associated with a neurologic condition, chronic migraine, upper limb spasticity, cervical dystonia, primary axillary hyperhidrosis, blepharospasm and strabismus>

Effect of the onabotulinumtoxin A may be potentiated with coadministration of aminoglycosides or other agents interfering with neuromuscular transmission (curare-like compounds).

**Description of the (proven, putative or hypothesized) mechanism of the DDI:**
None.

**Clinical consequences:**
None.

**Reference:**
Botox prescribing information.

Dysport prescribing information.

**Drug-Drug Interaction Monograph for BOTOX, DYSPORT (Onabotulinumtoxin) #2**

**Perpetrator drug:**

| | |
|---|---|
| INN name/US brand name: | Onabotulinumtoxin A/Botox, Dysport |
| Class/Molecular weight: | Protein Toxin/149.3 kDa |

**Victim drug:**

| | |
|---|---|
| INN name/US brand name: | Anticholinergic drugs |
| Class/Molecular weight: | |

**Description of the effect of the DDI:**
<Note: Besides cosmetic uses, Onabotulinumtoxin A is also used for the treatment of destusor overactivity associated with a neurologic condition, chronic migraine, upper limb spasticity, cervical dystonia, primary axillary hyperhidrosis, blepharospasm and strabismus>

Use of anticholinergic drugs after administration of Botox may potentiate systemic anticholinergic effects.

**Description of the (proven, putative or hypothesized) mechanism of the DDI:**
None.

**Clinical consequences:**
None.

**Reference:**
Botox prescribing information.

Dysport prescribing information.

**Drug-Drug Interaction Monograph for BOTOX, DYSPORT (Onabotulinumtoxin) #3**

**Perpetrator drug:**
INN name/US brand name: Muscle relaxants
Class/Molecular weight:

**Victim drug:**
INN name/US brand name: Onabotulinumtoxin A/Botox, Dysport
Class/Molecular weight: Protein Toxin/149.3 kDa

**Description of the effect of the DDI:**
<Note: Besides cosmetic uses, Botox is also used for the treatment of destusor overactivity associated with a neurologic condition, chronic migraine, upper limb spasticity, cervical dystonia, primary axillary hyperhidrosis, blepharospasm and strabismus>

Excessive weakness may be exaggerated by administration of a muscle relaxant before or after administration of onabotulinumtoxin A.

**Description of the (proven, putative or hypothesized) mechanism of the DDI:**
None.

**Clinical consequences:**
None.

**Reference:**
Botox prescribing information.

Dysport prescribing information.

**Drug-Drug Interaction Monograph for BYDUREON (Exenatide) #1**

**Perpetrator drug:**
INN name/US brand name: Orally administered drugs
Class/Molecular weight:

**Victim drug:**
INN name/US brand name: Exenatide/Bydureon
Class/Molecular weight: Peptide/4.2 kDa

**Description of the effect of the DDI:**
Exenatide has the potential to reduce the extent and rate of absorption of orally administered drugs.

**Description of the (proven, putative or hypothesized) mechanism of the DDI:**
Exenatide slows gastric emptying.

**Clinical consequences:**
Use caution when administering oral medications with exenatide. Exenatide should be used with caution in patients receiving oral medications that have narrow therapeutic index or require rapid gastrointestinal absorption. For oral medications that are dependent on threshold concentrations for efficacy, such as contraceptives and antibiotics, patients should be advised to take those drugs at least 1 hour before exenatide injection. If such drugs are to

*(Continued)*

be administered with food, patients should be advised to take them with a meal or snack when exenatide is not administered

**Reference:**
Bydureon prescribing information.

Byetta prescribing information.

---

**Drug-Drug Interaction Monograph for BYDUREON (Exenatide) #2**

---

**Perpetrator drug:**

| | |
|---|---|
| INN name/US brand name: | Acetaminophen/Tylenol |
| Class/Molecular weight: | Small molecule/151 Da |

**Victim drug:**

| | |
|---|---|
| INN name/US brand name: | Exenatide/Bydureon |
| Class/Molecular weight: | Peptide/4.2 kDa |

**Description of the effect of the DDI:**
In patients with type 2 diabetes, exenatide did not affect the absorption of orally administered acetaminophen to any clinically relevant degree.

**Description of the (proven, putative or hypothesized) mechanism of the DDI:**
No DDI.

**Clinical consequences:**
None.

**Reference:**
Bydureon prescribing information.

(not in Byetta labeling)

---

**Drug-Drug Interaction Monograph for BYDUREON (Exenatide) #3**

---

**Perpetrator drug:**

| | |
|---|---|
| INN name/US brand name: | Warfarin/Coumadin |
| Class/Molecular weight: | Small molecule/308 Da |

**Victim drug:**

| | |
|---|---|
| INN name/US brand name: | Exenatide/Bydureon, Byetta |
| Class/Molecular weight: | Peptide/4.2 kDa |

**Description of the effect of the DDI:**
There are postmarketing reports of increased INR with concomitant use of warfarin, sometimes associated with bleeding. In a drug interaction study, exenatide did not have a significant effect on INR.

**Description of the (proven, putative or hypothesized) mechanism of the DDI:**
None.

**Clinical consequences:**
In patients taking warfarin, the INR should be monitored more frequently after initiating BYDUREON. Once a stable INR has been documented, the INR can be monitored at the intervals usually recommended for patients on warfarin.

**Reference:**
Bydureon prescribing information.

Byetta prescribing information.

---

**Drug-Drug Interaction Monograph for BYETTA/(Exenatide)**

---

See BYDUREON (Exenatide) drug-drug interaction monographs #1 and #3
Reference: Bio-Tropin US prescribing information

---

**Drug-Drug Interaction Monograph for CIMZIA (Certolizumab)**

---

**Perpetrator drug:**
INN name/US brand name: TNF-blocking agents and other immune modulators (e.g. anakinra, abatacept, rituximab, and natalizumab)
Class/Molecular weight:

**Victim drug:**
INN name/US brand name: Certolizumab pegol /Cimzia
Class/Molecular weight: Antibody Fab' fragment/91 kDa

**Description of the effect of the DDI:**
An increased risk of serious infections has been seen in clinical studies of other TNF-blocking agents used in combination with anakinra or abatacept, with no added benefit. Formal drug interaction studies have not been performed with rituximab or natalizumab. Because of the nature of the adverse events seen with these combinations with TNF blocker therapy, similar toxicities may also result from the use of Cimzia in these combinations.

**Description of the (proven, putative or hypothesized) mechanism of the DDI:**
Immunomodulatory effects of both agents.

**Clinical consequences:**
There is not enough information to assess the safety and efficacy of such combination therapy. Therefore, the use of certolizumab pegol in combination with anakinra, abatacept, rituximab, or natalizumab is not recommended

**Reference:**
Cimzia prescribing information.

Johnston BL, Conly JM. Tumour necrosis factor inhibitors and infection: What is there to know for infectious diseases physicians? Can J Infect Dis Med Microbiol. 2006; 17(4): 209–212.

Katikireddi VS, Whittle SL, Hill CL. Tumour necrosis factor inhibitors and risk of serious infection in rheumatoid arthritis. Int J Rheum Dis. 2010;13(1):12–26.

---

## Drug-Drug Interaction Monograph for COPAXONE (Glatiramer acetate)

**Perpetrator drug:**

INN name/US brand name: Glatiramer acetate/Copaxone
Class/Molecular weight: Peptide/5 kDa–9 kDa

**Victim drug:**

INN name/US brand name: Corticosteroids
Class/Molecular weight:

**Description of the effect of the DDI:**

Results from existing clinical trials do not suggest any significant interactions of glatiramer acetate with therapies commonly used in multiple sclerosis (MS) patients, including the concurrent use of corticosteroids for up to 28 days.

**Description of the (proven, putative or hypothesized) mechanism of the DDI:**

None.

**Clinical consequences:**

None.

**Reference:**

Magraner MJ, Coret F, Navarré A, Boscá I, Simó M, Escutia M, Bernad A, Navarro L, Casanova B. Pulsed steroids followed by glatiramer acetate to prevent inflammatory activity after cessation of natalizumab therapy: a prospective, 6-month observational study. J Neurol. 2011;258(10):1805–11.

De Stefano N, Filippi M, Hawkins C. Short-term combination of glatiramer acetate with i.v. steroid treatment preceding treatment with GA alone assessed by MRI-disease activity in patients with relapsing-remitting multiple sclerosis. J Neurol Sci. 2008;266(1–2):44–50.

Copaxone prescribing information.

## Drug-Drug Interaction Monograph for CUBICIN (Daptomycin)

**Perpetrator drug:**

INN name/US brand name: Daptomycin/Cubicin
Class/Molecular weight: Cyclic lipopeptide/1.6 kDa

**Victim drug:**

INN name/US brand name: Simvastatin/Zocor
Class/Molecular weight: Small molecule/419 Da

**Description of the effect of the DDI:**

In healthy subjects, concomitant administration of daptomycin and simvastatin had no effect on plasma trough concentrations of simvastatin, and there were no reports of skeletal myopathy.

**Description of the (proven, putative or hypothesized) mechanism of the DDI:**

None.

**Clinical consequences:**

Experience with the coadministration of HMG-CoA reductase inhibitors and daptomycin in patients is limited; therefore, consideration should be given to suspending use of HMG-CoA reductase inhibitors temporarily in patients receiving daptomycin.

**Reference:**

Cubicin prescribing information.

**Drug-Drug Interaction Monograph for DDAVP (Desmopressin) #1**

**Perpetrator drug:**
INN name/US brand name: Pressor Agents
Class/Molecular weight:

**Victim drug:**
INN name/US brand name: Desmopressin/DDAVP
Class/Molecular weight: Peptide/1.2 kDa

**Description of the effect of the DDI:**
Desmopressin generally has very low pressor activity compared to the antidiuretic activity. However large doses of desmopressin with other pressor agents could potentiate the pressor effect.

**Description of the (proven, putative or hypothesized) mechanism of the DDI:**
Pharmacodynamic synergy.

**Clinical consequences:**
Although the pressor activity of desmopressin is very low compared to the antidiuretic activity, use of large doses of intranasal mesmopressin with other pressor agents should only be done with careful patient monitoring.

**Reference:**
DDAVP prescribing information.

**Drug-Drug Interaction Monograph for DDAVP (Desmopressin) #2**

**Perpetrator drug:**
INN name/US brand name: Drugs that may cause hyponatremia (e.g. tricyclic antidepressants, selective serotonin re-uptake inhibitors, chlorpromazine, opiate analgesics, NSAIDs, lamotrigine and carbamazepine)
Class/Molecular weight:

**Victim drug:**
INN name/US brand name: Desmopressin/DDAVP
Class/Molecular weight: Peptide/1.2 kDa

**Description of the effect of the DDI:**
Increased risk of hyponatremia. Several cases of mild to moderate hyponatremia have been reported as well as a severe case of hyponatremia after administration of desmopressin.

**Description of the (proven, putative or hypothesized) mechanism of the DDI:**
Pharmacodynamic synergy. Desmopressin is an synthetic analog of vasopressin and stimulates collecting duct aquaporin expression and subsequent resorption of water to the renal parenchyma.

*(Continued)*

**Clinical consequences:**
The concomitant administration of drugs that may increase the risk of water intoxication with hyponatremia, (e.g. tricyclic antidepressants, selective serotonin re-uptake inhibitors, chlorpromazine, opiate analgesics, NSAIDs, lamotrigine and carbamazepine) should be performed with caution.

**Reference:**
Kelleher HB, Henderson SO. Severe hyponatremia due to desmopressin. J Emerg Med. 2006;30(1):45–7.

DDAVP prescribing information.

---

**Drug-Drug Interaction Monograph for ERBITUX (Cetuximab)**

---

**Perpetrator drug:**

| | |
|---|---|
| INN name/US brand name: | Cetuximab/Erbitux |
| Class/Molecular weight: | Monoclonal Antibody/152 kDa |

**Victim drug:**

| | |
|---|---|
| INN name/US brand name: | Irinotecan/Camptosar |
| Class/Molecular weight: | Small Molecule/677.2 Da |

**Description of the effect of the DDI:**
There was no evidence of any pharmacokinetic interactions between cetuximab and irinotecan.

**Description of the (proven, putative or hypothesized) mechanism of the DDI:**
Not applicable.

**Clinical consequences:**
No dose adjustment is needed.

**Reference:**
Delbaldo C et al (2005). Pharmacokinetic profile of cetuximab (Erbitux) alone and in combination with irinotecan in patients with advanced EGFR-positive adenocarcinoma. Eur J Cancer 41(12):1739–1745.

---

**Drug-Drug Interaction Monograph for ENBREL (Etanercept) #1**

---

**Perpetrator drug:**

| | |
|---|---|
| INN name/US brand name: | Abatacept/Orencia |
| Class/Molecular weight: | Fusion Protein/92 kDa |

**Victim drug:**

| | |
|---|---|
| INN name/US brand name: | Etanercept/Enbrel |
| Class/Molecular weight: | Fusion Protein/150 kDa |

**Description of the effect of the DDI:**
The incidence of total serious adverse events in the combination therapy of abatacept and background biologic treatment (64.1% etanercept) was much higher (23%) than that (8%) of background biologic treatment (65.6% etanercept).

**Description of the (proven, putative or hypothesized) mechanism of the DDI:**
Mechanism is not known.

**Clinical consequences:**
Use of etanercept with abatacept is not recommended.

**Reference:**
Enbrel prescribing information

Weinblatt M et al (2006). Safety of the selective costimulation modulator abatacept in rheumatoid arthritis patients receiving background biologic and nonbiologic disease-modifying antirheumatic drugs: A one-year randomized, placebo-controlled study. Arthritis Rheumatism 54(9): 2807–2816.

---

**Drug-Drug Interaction Monograph for ENBREL (Etanercept) #2**

---

**Perpetrator drug:**

| | |
|---|---|
| INN name/US brand name: | Anakinra/Kineret |
| Class/Molecular weight: | IL-1 Receptor Antagonist/17.3 kDa |

**Victim drug:**

| | |
|---|---|
| INN name/US brand name: | Etanercept/Enbrel |
| Class/Molecular weight: | Fusion Protein/150 kDa |

**Description of the effect of the DDI:**
In a study in which patients with active rheumatoid arthritis were treated for up to 24 weeks with concurrent etanercept and anakinra therapy, a 7% rate of serious infections was observed, which was higher than that observed with etanercept alone (0%). Two percent of patients treated concurrently with etanercept and anakinra developed neutropenia (ANC $< 1 \times 109$/L).

**Description of the (proven, putative or hypothesized) mechanism of the DDI:**
Mechanism is not known.

**Clinical consequences:**
Use of etanercept with anakinra is not recommended.

**Reference:**
Enbrel prescribing information.

Genovese MC et al (2004). Combination therapy with etanercept and anakinra in the treatment of patients with rheumatoid arthritis who have been treated unsuccessfully with methotrexate. Arthritis Rheumatism 50(5): 1412–1419.

---

**Drug-Drug Interaction Monograph for ENBREL (Etanercept) #3**

---

**Perpetrator drug:**

| | |
|---|---|
| INN name/US brand name: | Etanercept/Enbrel |
| Class/Molecular weight: | Fusion Protein/150 kDa |

**Victim drug:**

| | |
|---|---|
| INN name/US brand name: | Digoxin/Lanoxin |
| Class/Molecular weight: | Small Molecule/780.95 Da |

*(Continued)*

**Description of the effect of the DDI:**
Digoxin maximum plasma concentration ($C_{max}$) and area under the plasma concentration versus time curve(AUC) were not altered by the coadministration of etanercept. The mean Cmax and AUC of etanercept were 4.2% and 12.5% lower, respectively, when etanercept was given with digoxin than when administered alone. There were no clinically relevant changes in the electrocardiogram (ECG) parameters when both drugs were combined.

**Description of the (proven, putative or hypothesized) mechanism of the DDI:**
Not applicable.

**Clinical consequences:**
There is no clinically relevant interaction between etanercept and digoxin, and both drugs can be safely coadministered without the need for a dosage adjustment.

**Reference:**
Zhou H et al (2004). Absence of a clinically relevant interaction between etanercept and digoxin. J Clin Pharmacol 44: 1244–1251.

---

**Drug-Drug Interaction Monograph for ENBREL (Etanercept) #4**

---

**Perpetrator drug:**

| | |
|---|---|
| INN name/US brand name: | Etanercept/Enbrel |
| Class/Molecular weight: | Fusion Protein/150 kDa |

**Victim drug:**

| | |
|---|---|
| INN name/US brand name: | Pneumococcal Polysaccharide Vaccine/Pneumovax |
| Class/Molecular weight: | Vaccine/NA |

**Description of the effect of the DDI:**
Most patients with psoriatic arthritis receiving etanercept were able to mount effective B-cell immune responses to pneumococcal polysaccharide vaccine, but titers in aggregate were moderately lower and fewer patients had 2-fold rises in titers compared to patients not receiving Enbrel.

**Description of the (proven, putative or hypothesized) mechanism of the DDI:**
Mechanism is not known.

**Clinical consequences:**
The clinical relevance is unknown.

**Reference:**
Enbrel prescribing information.

---

**Drug-Drug Interaction Monograph for ENBREL (Etanercept) #5**

---

**Perpetrator drug:**

| | |
|---|---|
| INN name/US brand name: | Etanercept/Enbrel |
| Class/Molecular weight: | Fusion Protein/150 kDa |

**Victim drug:**

| | |
|---|---|
| INN name/US brand name: | Warfarin/Coumadin, Jantoven, Marevan, Lawarin, etc. |
| Class/Molecular weight: | Small Molecule/330.3 Da |

**Description of the effect of the DDI:**
Etanercept did not affect the pharmacokinetics and pharmacodynamics of warfarin. All ratios of maximum serum concentration ($C_{max}$) and area under the serum concentration versus time curve (AUC) for pharmacokinetics (R- and S-enantiomers of warfarin) and International Normalized Ratio (INR) fell within the confidence interval of 0.8 to 1.25.

Warfarin also did not cause a clinically significant alteration in the pharmacokinetics of etanercept.

**Description of the (proven, putative or hypothesized) mechanism of the DDI:**
Not applicable.

**Clinical consequences:**
Coadministration of etanercept and warfarin would not be expected to change the pharmacokinetics of either medication; therefore, no dosage adjustment is needed in cases in which warfarin and etanercept are coadministered.

**Reference:** Zhou H et al (2004). Absence of a pharmacokinetic interaction between etanercept and warfarin. J Clin Pharmacol 44: 543–550.

---

**Drug-Drug Interaction Monograph for ENBREL (Etanercept) #6**

**Perpetrator drug:**
INN name/US brand name: Methotrexate/Mexate, Folex, Rheumatrex
Class/Molecular weight: Small Molecule/454.5 Da

**Victim drug:**
INN name/US brand name: Etanercept/Enbrel
Class/Molecular weight: Fusion Protein/150 kDa

**Description of the effect of the DDI:**
Based on the bioequivalence criteria, the Bayesian-estimated clearance for patients receiving etanercept alone (mean: 0.070 L/h) was comparable to that for patients receiving a combination of etanercept and methotrexate (mean = 0.066 L/h). The pharmacokinetics of etanercept were not altered by the concurrent administration of methotrexate in patients with rheumatoid arthritis.

**Description of the (proven, putative or hypothesized) mechanism of the DDI:**
Not applicable.

**Clinical consequences:**
No etanercept dose adjustment is needed for patients taking concurrent methotrexate.

**Reference:**
Zhou H et al (2004). Unaltered etanercept pharmacokinetics with concurrent methotrexate in patients with rheumatoid arthritis. J Clin Pharmacol 44: 1235–1243.

---

**Drug-Drug Interaction Monograph for EXUBERA (Insulin human)**

**Perpetrator drug:**

| | |
|---|---|
| INN name/US brand name: | Albuterol/Ventolin, Proventil, Proventil-HFA, AccuNeb, Vospire, ProAir |
| Class/Molecular weight: | Small Molecule/576.7 Da |

**Victim drug:**

| | |
|---|---|
| INN name/US brand name: | Insulin Human/Exubera |
| Class/Molecular weight: | Peptide Hormone/5.8 kDa |

**Description of the effect of the DDI:**
Administration of albuterol 30 minutes prior to administration of insulin in non-diabetic subjects with both mild asthma (n = 36) and moderate asthma (n = 31) resulted in a mean increase in insulin AUC and Cmax of between 25 and 50% compared to when insulin was administered alone.

**Description of the (proven, putative or hypothesized) mechanism of the DDI:**
The mechanism is not known. Albuterol, a short-acting beta-agonist that works as a bronchodilator, may enhance insulin's inhalation/absorption through its bronchodilatory effect.

**Clinical consequences:**
Consistent timing of dosing of bronchodilators relative to insulin administration, close monitoring of blood glucose concentrations and dose titration as appropriate are recommended.

**Reference:**
http://www.accessdata.fda.gov/drugsatfda_docs/nda/2006/021868s000_ClinPharmR.pdf.
http://www.accessdata.fda.gov/drugsatfda_docs/label/2008/021868s016s017lbl.pdf.

**Drug-Drug Interaction Monograph for HUMIRA (Adalimumab) #1**

**Perpetrator drug:**

| | |
|---|---|
| INN name/US brand name: | Adalimumab/Humira |
| Class/Molecular weight: | Monoclonal Antibody/148 kDa |

**Victim drug:**

| | |
|---|---|
| INN name/US brand name: | Methotrexate/Rheumatrex, Trexall |
| Class/Molecular weight: | Small Molecule/454.5 Da |

**Description of the effect of the DDI:**
Repeated administration of adalimumab had no statistically significant effect on the pharmacokinetics of methotrexate.

**Description of the (proven, putative or hypothesized) mechanism of the DDI:**
None.

**Clinical consequences:**
No dose adjustment is needed.

**Reference:** Weisman M et al. Efficacy, pharmacokinetic, and safety assessment of adalimumab, a fully human anti-tumor necrosis factor-alpha monoclonal antibody, in adults with rheumatoid arthritis receiving concomitant methotrexate: a pilot study. Clin Therapeutics 2003;25(6):1700–1721.

---

**Drug-Drug Interaction Monograph for HUMIRA (Adalimumab) #2**

---

**Perpetrator drug:**

INN name/US brand name: Methotrexate/Rheumatrex, Trexall
Class/Molecular weight: Small Molecule/454.5 Da

**Victim drug:**

INN name/US brand name: Adalimumab/Humira
Class/Molecular weight: Monoclonal Antibody/148 kDa

**Description of the effect of the DDI:**
Methotrexate reduced adalimumab apparent clearance after single and multiple dosing by 29% and 44%, respectively.

**Description of the (proven, putative or hypothesized) mechanism of the DDI:**
Methotrexate lowers adalimumab's clearance by decreasing the incidence of immunogenicity and/or operating via interacting Fcγ receptor.

**Clinical consequences:**
No dose adjustment is needed.

**Reference:** Humira prescribing information.

---

**Drug-Drug Interaction Monograph for INTRON A (Interferon alfa-2B)**

---

**Perpetrator drug:**

INN name/US brand name: Interferon Alfa-2B/Intron A
Class/Molecular weight: Cytokine/19.3 kDa

**Victim drug:**

INN name/US brand name: Theophylline/Elixophyllin, Theo-24, Theolair, etc.
Class/Molecular weight: Small Molecule/180.2 Da

**Description of the effect of the DDI:**
Concomitant use of alpha interferon and theophylline decreases theophylline clearance, resulting in a significant increase in serum theophylline levels.

**Description of the (proven, putative or hypothesized) mechanism of the DDI:**
Intron A inhibits metabolism of theophylline by CYP1A2.

**Clinical consequences:**
No dose adjustment is needed.

**Reference:**
Intron A prescribing information.

Willimas SJ et al. Inhibition of theophylline metabolism by interferon. Lancet 1987;24; 2(8565):939–41.

---

**Drug-Drug Interaction Monograph for IPRIVASK (Desirudin) #1**

**Perpetrator drug:**
INN name/US brand name: Desirudin Recombinant/Iprivask
Class/Molecular weight: Polypeptide/19.2 kDa

**Victim drug:**
INN name/US brand name: Warfarin/Coumadin, Jantoven, Marevan, Lawarin, etc.
Class/Molecular weight: Small Molecule/330.3 Da

**Description of the effect of the DDI:**
Warfarin (2 × 5 mg daily oral administration) did not significantly affect the pharmacokinetics of desirudin.

There was an additive effect of desirudin on activated partial thromboplastin time (aPTT), prothrombin time (PT), and international normalized ratio (INR) when co-administered with warfarin.

**Description of the (proven, putative or hypothesized) mechanism of the DDI:**
The exact mechanism is not yet understood.

**Clinical consequences:**
If a patient is switched from oral anticoagulants to desirudin therapy or from desirudin to oral anticoagulants, the anticoagulant activity should continue to be closely monitored with appropriate methods. That activity should be taken into account in the evaluation of the overall coagulation status of the patient during the switch.

**Reference:**
Iprivask prescribing information.

http://www.accessdata.fda.gov/drugsatfda_docs/nda/2003/21-271_Iprivask_BioPharmr.pdf.

**Drug-Drug Interaction Monograph for IPRIVASK (Desirudin) #2**

**Perpetrator drug:**
INN name/US brand name: Acetylsalicylic acid/Aspirin
Class/Molecular weight: Small Molecule/180.1 Da

**Victim drug:**
INN name/US brand name: Desirudin Recombinant/Iprivask
Class/Molecular weight: Polypeptide/19.2 kDa

**Description of the effect of the DDI:**
Acetylsalicylic acid (ASA) had no effect on the pharmacokinetics of desirudin. Administration of ASA had no effect on aPTT and PT produced by intravenous infusion of desirudin. Prolonged bleeding times observed in some patients.

**Description of the (proven, putative or hypothesized) mechanism of the DDI:**
The exact mechanism is not yet understood.

**Clinical consequences:**
Desirudin should be used with caution in conjunction with drugs which affect platelet function. These medications include systemic salicylates, NSAIDS including ketorolac,

acetylsalicylic acid, ticlopidine, dipyridamole, sulfinpyrazone, clopidogrel, abciximab and other glycoprotein IIb/IIIa antagonists.

**Reference:**
Iprivask prescribing information.

http://www.accessdata.fda.gov/drugsatfda_docs/nda/2003/21-271_Iprivask_BioPharmr.pdf.

---

**Drug-Drug Interaction Monograph for IPRIVASK (Desirudin) #3**

---

**Perpetrator drug:**
INN name/US brand name: Piroxicam/Feldene
Class/Molecular weight: Small Molecule/331.35 Da

**Victim drug:**
INN name/US brand name: Desirudin Recombinant/Iprivask
Class/Molecular weight: Polypeptide/19.2 kDa

**Description of the effect of the DDI:**
When desirudin was coadministered with piroxicam (half of the usual daily dose of 20 mg), no significant pharmacokinetic interaction were observed. Neither was there any obvious alteration of the anticoagulant effect of desirudin.

**Description of the (proven, putative or hypothesized) mechanism of the DDI:**
The exact mechanism is not yet understood.

**Clinical consequences:**
As with other anticoagulants, desirudin should be used with caution in conjunction with drugs which affect platelet function. These medications include systemic salicylates, NSAIDS including ketorolac, acetylsalicylic acid, ticlopidine, dipyridamole, sulfinpyrazone, clopidogrel, abciximab and other glycoprotein IIb/IIIa antagonists

**Reference:**
Iprivask prescribing information.

http://www.accessdata.fda.gov/drugsatfda_docs/nda/2003/21-271_Iprivask_BioPharmr.pdf.

---

**Drug-Drug Interaction Monograph for IPRIVASK (Desirudin) #4**

---

**Perpetrator drug:**
INN name/US brand name: Heparin/Calciparine and Hep-Lock
Class/Molecular weight: Sulfated Glycosaminoglycan/3–30 kDa

**Victim drug:**
INN name/US brand name: Desirudin Recombinant/Iprivask
Class/Molecular weight: Polypeptide/19.2 kDa

**Description of the effect of the DDI:**
The pharmacokinetics of desirudin was not altered by concomitant infusion of heparin (bolus of 5000 units followed by infusion of 1000 units/hour).

*(Continued)*

**Description of the (proven, putative or hypothesized) mechanism of the DDI:**
The exact mechanism is not yet understood.

**Clinical consequences:**
During prophylaxis of venous thromboembolism, concomitant treatment with heparins (unfractionated and low-molecular weight heparins) or dextrans is not recommended. The effects of desirudin and unfractionated heparins on prolongation of aPTT are additive.

**Reference:**
Iprivask prescribing information.

http://www.accessdata.fda.gov/drugsatfda_docs/nda/2003/21-271_Iprivask_BioPharmr.pdf.

---

**Drug-Drug Interaction Monograph for IPRIVASK (Desirudin) #5**

---

**Perpetrator drug:**

| | |
|---|---|
| INN name/US brand name: | Desmopressin/DDAVP, DesmoMelt, Stimate, Minirin |
| Class/Molecular weight: | Peptide/1069.2 Da |

**Victim drug:**

| | |
|---|---|
| INN name/US brand name: | Desirudin Recombinant/Iprivask |
| Class/Molecular weight: | Polypeptide/19.2 kDa |

**Description of the effect of the DDI:**
There was no effect of desmopressin on the pharmacokinetics of desirudin. However, it was shown that an infusion of desmopressin (0.3 μg/kg) can partially reverse the desirudin induced increase in aPTT and can dramatically increase factor VIII:C levels depending on the rate of ongoing desirudin infusion.

**Description of the (proven, putative or hypothesized) mechanism of the DDI:**
The exact mechanism is not yet understood.

**Clinical consequences:**
The clinical effectiveness of desmopressin in treating bleeding due to desirudin overdose has not been studied.

**Reference:**
Iprivask prescribing information.

http://www.accessdata.fda.gov/drugsatfda_docs/nda/2003/21-271_Iprivask_BioPharmr.pdf.

---

**Drug-Drug Interaction Monograph for KINERET (Anakinra)**

---

**Perpetrator drug:**

| | |
|---|---|
| INN name/US brand name: | Methotrexate/Mexate, Folex, Rheumatrex |
| Class/Molecular weight: | Small Molecule/454.5 Da |

**Victim drug:**

| | |
|---|---|
| INN name/US brand name: | Anakinra/Kineret |
| Class/Molecular weight: | IL-1 Receptor Antagonist/17.3 kDa |

**Description of the effect of the DDI:**
Toxicologic and toxicokinetic studies in rats did not demonstrate any alterations in the clearance or toxicologic profile of either methotrexate or anakinra when the two agents were administered together.

**Description of the (proven, putative or hypothesized) mechanism of the DDI:**
The exact mechanism is not yet understood.

**Clinical consequences:**
No dose adjustment is needed.

**Reference:**
Kineret prescribing information.

---

**Drug-Drug Interaction Monograph for LUVERIS (Lutropin alfa)**

---

**Perpetrator drug:**

| | |
|---|---|
| INN name/US brand name: | Follitropin Alfa/Gonal-f |
| Class/Molecular weight: | Human hormone/22.7 kDa |

**Victim drug:**

| | |
|---|---|
| INN name/US brand name: | Lutropin Alfa/Luveris |
| Class/Molecular weight: | Recombinant human hormone/23.4 kDa |

**Description of the effect of the DDI:**
There are no pharmacokinetic interactions between lutropin alfa (150 IU) with follitropin alfa (150 IU) when administered simultaneously.

**Description of the (proven, putative or hypothesized) mechanism of the DDI:**
The exact mechanism is not yet understood.

**Clinical consequences:**
No dose adjustment is needed.

**Reference:**
Luveris prescribing information.

http://www.accessdata.fda.gov/drugsatfda_docs/nda/2004/021322s000_Luveris_biopharmr.pdf.

---

**Drug-Drug Interaction Monograph for MIACALCIN (Calcitonin)**

---

**Perpetrator drug:**

| | |
|---|---|
| INN name/US brand name: | Calcitonin-salmon/Miacalcin |
| Class/Molecular weight: | Polypeptide/3455 Da |

**Victim drug:**

| | |
|---|---|
| INN name/US brand name: | Lithium |
| Class/Molecular weight: | Small Molecule |

**Description of the effect of the DDI:**
Concurrent use of calcitonin and lithium may result in decreased lithium concentrations and loss of lithium efficacy. The co-administration of calcitonin and lithium causes a significant decrease in lithium plasma levels, with the average decrease reaching a maximum of 30%.

**Description of the (proven, putative or hypothesized) mechanism of the DDI:**
Increased renal lithium excretion and/or decreased lithium intestinal absorption.

*(Continued)*

**Clinical consequences:**
Lithium levels should be closely monitored in patients receiving calcitonin therapy. Lithium dose adjustments may be necessary.

**Reference:** Miacalcin prescribing information.

Passiu G, et al. Calcitonin decreases lithium plasma levels in man. Preliminary report. Int J Clin Pharm Res 1998; 18:179–181.

Micromedex

---

**Drug-Drug Interaction Monograph for MIRCERA (Methoxy polyethylene glycol-epoetin beta)**

---

**Perpetrator drug:**
INN name/US brand name: Methoxy polyethylene glycol-epoetin beta/Mircera
Class/Molecular weight: Erythropoietin analog/60 kDa

**Victim drug:**
INN name/US brand name: None
Class/Molecular weight:

**Description of the effect of the DDI:**
No drug interactions have been identified.

**Description of the (proven, putative or hypothesized) mechanism of the DDI:**
None available.

**Clinical consequences:**
None.

**Reference:** Mircera prescribing information

---

**Drug-Drug Interaction Monograph for MYOBLOC (Rimabotulinumtoxin B)**

---

**Perpetrator drug:**
INN name/US brand name: Aminoglycosides/other agents interfering neuromuscular transmission (e.g., curare-like compounds)
Class/Molecular weight: Small molecule

**Victim drug:**
INN name/US brand name: Rimabotulinumtoxin B/Myobloc
Class/Molecular weight: Two chain polypeptide/150 kDa

**Description of the effect of the DDI:**
Caution should be exercised when co-administering rimabotulinumtoxin B and aminoglycosides or other agents interfering with neuromuscular transmission (e.g., curare-like compounds) as the effect of the toxin may be potentiated.

**Description of the (proven, putative or hypothesized) mechanism of the DDI:**
None available

**Clinical consequences:**
Toxicity of rimabotulinumtoxin B may increase.

**Reference:** Myobloc prescribing information.

---

**Drug-Drug Interaction Monograph for MYOZYME (Alglucosidase alfa)**

**Perpetrator drug:**
INN name/US brand name: Alglucosidase
Class/Molecular weight: Enzyme/110 kDa

**Victim drug:**
INN name/US brand name: None
Class/Molecular weight:

**Description of the effect of the DDI:**
No drug interactions have been identified.

**Description of the (proven, putative or hypothesized) mechanism of the DDI:**
None available.

**Clinical consequences:**
None.

**Reference:** Myozyme prescribing information.

---

**Drug-Drug Interaction Monograph for NEUMEGA (Oprelvekin)**

**Perpetrator drug:**
INN name/US brand name: Oprelvekin/Neumega
Class/Molecular weight: Recombinant interleukin/19 kDa

**Victim drug:**
Class/Molecular weight: Opioids

**Description of the effect of the DDI:**
Patient on treatment with opioids and oprelvekin developed respiratory depression and sedation suggested from a sedative medication.

**Description of the (proven, putative or hypothesized) mechanism of the DDI:**
Oprelvekin alters renal handling of water and sodium. Morphine and its metabolites are known to undergo renal elimination, and oprelvekin may have interfered with this process.

**Clinical consequences:**
Precaution should be taken while prescribing oprelvekin to patients already on opioid therapy due to possibility of drug interaction.

**Reference:** Moorehead et al. Suspected interaction of oprelvekin (rhIL-11) with a long-acting opioid. Pediatr Blood Cancer. 2008 Dec;51(6):837

---

**Drug-Drug Interaction Monograph for NEUPOGEN (Filgrastim)**

**Perpetrator drug:**
INN name/US brand name: Filgrastim/Neupogen
Class/Molecular weight: Hematopoietic growth factor/18.8 kDa

**Victim drug:**
INN name/US brand name: Oprelvekin/Neumega
Class/Molecular weight: Recombinant interleukin/19 kDa

**Description of the effect of the DDI:**
No change in pharmacokinetic parameters or toxicity has been observed under co-administration of filgrastim and oprelvekin.

**Description of the (proven, putative or hypothesized) mechanism of the DDI:**
NA

**Clinical consequences:**
No clinical consequences have been observed.

**Reference:** Neupogen prescribing information.

**Drug-Drug Interaction Monograph for NULOJIX (Belatacept) #1**

**Perpetrator drug:**
INN name/US brand name: Belatacept/Nulojix
Class/Molecular weight: FC-CTLA-4 fusion protein/90.6 kDa

**Victim drug:**
INN name/US brand name: Mycophenolic acid (MPA)
Class/Molecular weight: Small molecule/320.3 Da

**Description of the effect of the DDI:**
In a pharmacokinetic study, the plasma concentrations of mycophenolic acid (MPA) was measured in 41 patients who received fixed mycophenolate mofetil (MMF) doses of 500 mg to 1500 mg twice daily with either 5 mg per kg of belatacept or cyclosporine. The mean dose-normalized MPA $C_{max}$ and $AUC_{0\text{-}12}$ were approximately 20% and 40% higher, respectively, with belatacept coadministration than with cyclosporine coadministration.

**Description of the (proven, putative or hypothesized) mechanism of the DDI:**
Belatacept may cause a decrease in CYP450 activity leading to a decrease in systemic clearance of MPA.

**Clinical consequences:**
Clinicians should be aware that there is a potential change of MPA exposure after crossover from cyclosporine to belatacept or from belatacept to cyclosporine in patients concomitantly receiving MMF.

**Reference:** Nulojix prescribing information.

**Drug-Drug Interaction Monograph for NULOJIX (Belatacept) #2**

**Perpetrator drug:**
INN name/US brand name: Belatacept/Nulojix
Class/Molecular weight: FC-CTLA-4 fusion protein/90.6 kDa

**Victim drug:**
INN name/US brand name: CYP450 substrates
Class/Molecular weight:

**Description of the effect of the DDI:**
No formal drug interaction studies have been conducted with belatacept, but biologic therapies that are cytokines or cytokine modulators have been shown to affect the expression and/or functional activities of CYP450 enzymes.

**Description of the (proven, putative or hypothesized) mechanism of the DDI:**
*In vitro* studies have shown that belatacept inhibits the production of certain cytokines during an alloimmune response.

**Clinical consequences:**
The potential for belatacept to alter the systemic concentrations of drugs that are CYP450 substrates has not been studied; however, in the event that kidney transplant patients receiving belatacept exhibit signs and symptoms of altered efficacy or adverse events associated with co-administered drugs which are known to be metabolized by CYP450, the clinician should be aware of potentially altered CYP450 metabolism of these drugs.

**Reference:** Nulojix prescribing information.

**Drug-Drug Interaction Monograph for OMNITROPE (Somatropin recombinant)**

**See ACCRETROPIN (Somatropin) drug-drug interaction monographs #1–#4**

**Reference:** Omnitrope prescribing information

**Drug-Drug Interaction Monograph for ONCASPAR (Pegaspargase)**

**Perpetrator drug:**
INN name/US brand name: N/A
Class/Molecular weight:

**Victim drug:**
INN name/US brand name: N/A
Class/Molecular weight:

**Description of the effect of the DDI:**
No formal drug interaction studies between pegaspartase and other drugs have been performed, as per drug label.

**Description of the (proven, putative or hypothesized) mechanism of the DDI:**
No dose adjustments have been recommended.

**Clinical consequences:**
No clinical consequences have been observed.

**Reference:** Oncaspar prescribing information.

**Drug-Drug Interaction Monograph for ORENCIA (Abatacept)**

**Perpetrator drug:**

| | |
|---|---|
| INN name/US brand name: | Abatacept/Orencia |
| Class/Molecular weight: | CTLA-4/human IgG1 Fc fusion protein<br>Immune Modulator (Antirheumatic)/92 kDa |

**Victim drug:**

| | |
|---|---|
| Class: | TNF antagonists/inhibitors (etanercept, adalimumab, anakinra, infliximab, golimumab) |

**Description of the effect of the DDI:**
Increased risk of serious infection.

**Description of the (proven, putative or hypothesized) mechanism of the DDI:**
Potentiating effect on inhibition of immune reactivity.

**Clinical consequences:**
Concomitant therapy is not recommended.

**Reference:** Micromedex.

**Drug-Drug Interaction Monograph for ORTHOCLONE OKT3 (Muromonab-CD3)**

**Perpetrator drug:**

| | |
|---|---|
| INN name/US brand name: | Echinacea preparations |
| Class/Molecular weight: | Herbal preparation |

**Victim drug:**

| | |
|---|---|
| INN name/US brand name: | Muromonab-CD3/Orthoclone OKT3 |
| Class/Molecular weight: | Murine IgG2α monoclonal antibody/146 kDa |

**Description of the effect of the DDI:**
Concurrent use of Echinacea preparations and muromonab-CD3 may result in decreased effectiveness of muromonab-CD3.

**Description of the (proven, putative or hypothesized) mechanism of the DDI:**
The probable mechanism of interaction is stimulation of the immune system by Echinacea, resulting in increased formation of neutralizing anti-drug antibodies.

**Clinical consequences:**
Since reduced muromonab effectiveness may be life threatening in organ transplantation, the concomitant use of Echinacea and muromonab-CD3 should be avoided.

**Reference:** Micromedex.

**Drug-Drug Interaction Monograph for PERJETA (Pertuzumab) #1**

**Perpetrator drug:**

| | |
|---|---|
| INN name/US brand name: | Pertuzumab/Perjeta |
| Class/Molecular weight: | Monoclonal Antibody/148 kDa |

**Victim drug:**
INN name/US brand name: Trastuzumab/Herceptin
Class/Molecular weight: Monoclonal Antibody/148 kDa

**Description of the effect of the DDI:**
No change in pharmacokinetic parameters or toxicity has been observed under co-administration of pertuzumab and trastuzumab.

**Description of the (proven, putative or hypothesized) mechanism of the DDI:**
No clinical consequences have been observed.

**Clinical consequences:**
No dose adjustments have been recommended.

**Reference:** Perjeta prescribing information.

---

**Drug-Drug Interaction Monograph for PERJETA (Pertuzumab) #2**

---

**Perpetrator drug:**
INN name/US brand name: Pertuzumab/Perjeta
Class/Molecular weight: Monoclonal Antibody/148 kDa

**Victim drug:**
INN name/US brand name: Docetaxel/Taxotere
Class/Molecular weight: Chemotherapeutic/807.9 Da

**Description of the effect of the DDI:**
No change in pharmacokinetic parameters or toxicity has been observed under co-administration of pertuzumab and docetaxel.

**Description of the (proven, putative or hypothesized) mechanism of the DDI:**
No clinical consequences have been observed.

**Clinical consequences:**
No dose adjustments have been recommended.

**Reference:** Perjeta prescribing information.

---

**Drug-Drug Interaction Monograph for PRIALT (Ziconotide Acetate)**

---

**Perpetrator drug:**
INN name/US brand name: Antidepressants, Anxiolytics, Antiepileptics,
Class/Molecular weight: Neuroleptics, Sedatives

**Victim drug:**
INN name/US brand name: Ziconotide Acetate/Prialt
Class/Molecular weight: Peptide/2.6 kD

**Description of the effect of the DDI:**
The use of drugs with CNS-depressant activities in conjunction with ziconotide acetate may be associated with an increased incidence of CNS-depressant effects.

*(Continued)*

**Description of the (proven, putative or hypothesized) mechanism of the DDI:**
The patient may experience cognitive impairment when ziconotide acetate is used in conjunction with CNS-depressant drugs due to the additive effects of the two therapies.

**Clinical consequences:**
It may be necessary to decrease the dose of the perpetrator medication.

**Reference:** Prialt prescribing information.

---

**Drug-Drug Interaction Monograph for PROLEUKIN (Aldesleukin) #1**

**Perpetrator drug:**
Class: Glucocorticoids

**Victim drug:**
INN name/US brand name: Aldesleukin/Proleukin
Class/Molecular weight: Cytokine/15.3 kDa

**Description of the effect of the DDI:**
Reduced antitumor effectiveness of aldesleukin with a concomitantly administered glucocorticoid.

**Description of the (proven, putative or hypothesized) mechanism of the DDI:**
This is a class related pharmacodynamics interaction of unknown mechanism.

**Clinical consequences:**
NA

**Reference:** Proleukin prescribing information.

---

**Drug-Drug Interaction Monograph for PROLEUKIN (Aldesleukin) #2**

**Perpetrator drug:**
Class Antihypertensives

**Victim drug:**
INN name/US brand name: Aldesleukin/Proleukin
Class/Molecular weight: Cytokine/15.3 kDa

**Description of the effect of the DDI:**
Antihypertensives potentiate the hypotensive effects of aldesleukin.

**Description of the (proven, putative or hypothesized) mechanism of the DDI:**
This is a class related pharmacodynamics interaction of unknown mechanism.

**Clinical consequences:**
Monitor patient's blood pressure.

**Reference:** Proleukin prescribing information.

---

**Drug-Drug Interaction Monograph for RAPTIVA (Efazulimab)**

**Perpetrator drug:**
INN name/US brand name: Efalizumab/Raptiva
Class/Molecular weight: Monoclonal Antibody/150 kDa

**Victim drug:**
Class/Molecular weight: Immunosuppressants

**Description of the effect of the DDI:**
Interactions have not been studied but an increased risk of infection or malignancy is expected.

**Description of the (proven, putative or hypothesized) mechanism of the DDI:**
Efalizumab and other immunosuppressants have concurrent mechanisms of action and increase the risk of toxicity.

**Clinical consequences:**
Do not coadminister efalizumab and immunosuppressants.

**Reference:** Raptiva prescribing information.

**Drug-Drug Interaction Monograph for REGRANEX (Becaplermin)**

**Perpetrator drug:**
INN name/US brand name: Becaplermin/Regranex
Class/Molecular weight: Peptide/25 kDa

**Victim drug:**
INN name/US brand name: NA
Class/Molecular weight:

**Description of the effect of the DDI:**
No drug-drug interaction studies have been performed.

**Description of the (proven, putative or hypothesized) mechanism of the DDI:**
NA

**Clinical consequences:**
None

**Reference:** Regranex prescribing information.

**Drug-Drug Interaction Monograph for REMICADE (Infliximab) #1**

**Perpetrator drug:**
INN name/US brand name: Infliximab/Remicade
Class/Molecular weight: Chimeric Monoclonal Antibody/148 kDa

**Victim drug:**
INN name/US brand name: Abatacept/Orencia
Class/Molecular weight: Fusion Protein/92 kDa

*(Continued)*

**Description of the effect of the DDI:**
Concomitant administration of infliximab and abatacept appeared to increase the risk for severe infection.

**Description of the (proven, putative or hypothesized) mechanism of the DDI:**
The increased risk of infection is hypothesized to be due to increased immunosuppression by TNF-alpha blockade if both drugs are given together.

**Clinical consequences:**
Infliximab should not be given concomitantly with abatacept.

**Reference:** Remicade prescribing information.

---

**Drug-Drug Interaction Monograph for REMICADE (Infliximab) #2**

---

**Perpetrator drug:**
INN name/US brand name: Infliximab/Remicade
Class/Molecular weight: Chimeric Monoclonal Antibody/148 kDa

**Victim drug:**
INN name/US brand name: Anakinra/Kineret
Class/Molecular weight: Recombinant Peptide/17 kDa

**Description of the effect of the DDI:**
Concomitant administration of infliximab and anakinra appeared to increase the risk for severe infection.

**Description of the (proven, putative or hypothesized) mechanism of the DDI:**
The increased risk of infection is hypothesized to be due to increased immunosuppression by TNF-alpha blockade if both drugs are given together.

**Clinical consequences:**
Infliximab should not be given concomitantly with anakinra.

**Reference:** Remicade prescribing information.

---

**Drug-Drug Interaction Monograph for REOPRO (Abciximab)**

---

**Perpetrator drug:**
INN name/US brand name: Abciximab/Reopro
Class/Molecular weight: chimeric monoclonal antibody fragment/47.6 kDa

**Victim drug:**
INN name/US brand name: Heparin
Class/Molecular weight: Anticoagulant

**Description of the effect of the DDI:**
Increased risk of bleeding has been described.

**Description of the (proven, putative or hypothesized) mechanism of the DDI:**
No change in pharmacokinetic parameters has been described. This appears to be a pharmacodynamic interaction.

**Clinical consequences:**
Use of heparin and abciximab during PTCA procedures should be monitored. Starting doses of heparin should not be higher than 70 IU/kg.

**Reference:** Reopro prescribing information.

The CAPTURE Investigators. Randomised placebo-controlled trial of abciximab before and during coronary intervention in refractory unstable angina: the CAPTURE Study. Lancet 1997, 349, 1429–35.

---

**Drug-Drug Interaction Monograph for SANDOSTATIN (Octreotide acetate)**

**Perpetrator drug:**
INN name/US brand name: Octreotide acetate/Sandostatin
Class/Molecular weight: Peptide/1.02 kDa

**Victim drugs:**
INN name/US brand name: Various small molecule drugs
Class/Molecular weight: Small molecules

**Description of the effect of the DDI:**
Increased risk of cardiotoxicity (QT prolongation, torsades de pointes, cardiac arrest)

**Description of the (proven, putative or hypothesized) mechanism of the DDI:**
Additive effects on QT prolongation was the mechanism involved in drug interactions of Sandostatin with ziprasidone, bepridil, thioridazine, pimozide, terfenadine, mesoridazine, sparfloxacin, dronedarone, cisapride, levomethadyl, gemifloxacin, Phenothiazines, Erythromycin, Iloperidone, moxifloxacin, enflurane, apomorphine, perflutren lipid microsphere, pazopanib, ondansetron, trazodone, crizotinib, astemizole, spiramycin, solifenacin, fingolimod, mifepristone, vandetanib, sodium phosphate, clarithromycin, chloral hydrate, droperidol, telavancin, saquinavir, isoflurane, toremifene, pentamidine, foscarnet, salmeterol, domperidone, levofloxacin, class ia antiarrhythmics, antipsychotics, gatifloxacin, paliperidone, lidoflazine, dasatinib, telithromycin, chloroquine, artemether/lumefantrine, quinine, dolasetron, tricyclic antidepressants, voriconazole, arsenic trioxide, tetrabenazine,. methadone, clozapine, vardenafil, isradipine, norfloxacin, lopinavir/ritonavir, sorafenib, ofloxacin, cotrimoxazole, probucol, vasopressin, nilotinib, clomipramine, sunitinib, azithromycin, ciprofloxacin, class i antiarrhythmic agents, fluoxetine, lapatinib, asenapine, ranolazine, zolmitriptan, mefloquine, class iii antiarrhythmics, halofantrine, citalopram, fluconazole, posaconazole, halothane.

**Clinical consequences:**
Concomitant use of drugs with other drugs that prolong the QT interval, such as Sandostatin/octreotide is contraindicated due to the potential for additive effects on QT interval prolongation and an increased risk of serious ventricular arrhythmias, serious cardiac events including torsade de pointes.

**Reference:** Sandostatin prescribing information, Micromedex.

---

**Drug-Drug Interaction Monograph for SEROSTIM/SAIZEN (Somatropin)**

**See ACCRETROPIN (Somatropin) drug-drug interaction monographs #1– #4**

**Reference:** Serostim prescribing information
Saizen prescribing information

**Drug-Drug Interaction Monograph for SIMPONI (Golimumab) #1**

**Perpetrator drug:**
INN name/US brand name: Golimumab/Simponi
Class/Molecular weight: Human IgG1 monoclonal antibody/150 kDa

**Victim drugs:**
INN name/US brand name: Abatacept/Orencia
Class/Molecular weight: CTLA-4/human IgG1 Fc fusion protein/92 kDa

**Description of the effect of the DDI:**
Increased risk of serious infection.

**Description of the (proven, putative or hypothesized) mechanism of the DDI:**
Interference with the immune system at multiple levels.

**Clinical consequences:**
In controlled trials, the concurrent administration of another TNF-blocker and abatacept was associated with a greater proportion of serious infections than the use of a TNF-locker alone; and the combination therapy, compared to the use of a TNF-blocker alone, has not demonstrated improved clinical benefit in the treatment of psoriatic arthritis. Therefore, the combination of TNF-blockers including golimumab and abatacept is not recommended.

**Reference:** Simponi prescribing information, Micromedex.

**Drug-Drug Interaction Monograph for SIMPONI (Golimumab) #2**

**Perpetrator drug:**
INN name/US brand name: Golimumab/Simponi
Class/Molecular weight: Human IgG1 monoclonal antibody/150 kDa

**Victim drugs:**
INN name/US brand name: Anakinra/Kineret
Class/Molecular weight: recombinant, nonglycosylated form of the human interleukin-1 receptor antagonist/17.3 kDa
INN name/US brand name: Various
Class/Molecular weight: Live Vaccines

**Description of the effect of the DDI:**
Increased risk of serious infections and neutropenia.

**Description of the (proven, putative or hypothesized) mechanism of the DDI:**
Increased risk of serious infections has been seen in clinical rheumatoid arthritis studies of other TNF-blockers used in combination with anakinra, with no added benefit; therefore, use of golimumab with anakinra is not recommended. A higher rate of serious infections has also been observed in rheumatoid arthritis patients treated with rituximab who received subsequent treatment with a TNF-blocker. There is insufficient information to provide recommendations regarding the concomitant use of golimumab and other biologic products approved to treat rheumatoid arthritis or psoriatic arthritis.

**Clinical consequences:**
Concurrent administration of anakinra (an interleukin-1 antagonist) and another TNF-blocker was associated with a greater portion of serious infections and neutropenia and no additional benefits compared to TNF-blocker alone. Therefore, the combination of anakinra with TNF blockers, including golimumab, is not recommended.

**Reference:** Simponi prescribing information, Micromedex.

---

**Drug-Drug Interaction Monograph for SIMPONI (Golimumab) #3**

---

**Perpetrator drug:**
INN name/US brand name: Golimumab/Simponi
Class/Molecular weight: Human IgG1 monoclonal antibody/150 kDa

**Victim drugs:**
INN name/US brand name: Various
Class/Molecular weight: Live Vaccines

**Description of the effect of the DDI:**
Live vaccines: Increased risk of secondary transmission of infection by live vaccines.

**Description of the (proven, putative or hypothesized) mechanism of the DDI:**
IgG antibodies are known to cross the placenta during pregnancy and have been detected in the serum of infants born to patients treated with these antibodies. Since golimumab is an IgG antibody, infants born to women treated with golimumab during their pregnancy may be at increased risk of infection for up to 6 months.

**Clinical consequences:**
Administration of live vaccines to infants exposed to golimumab *in utero* is not recommended for 6 months following the mother's last golimumab injection during pregnancy.

**Reference:** Simponi prescribing information, Micromedex

---

**Drug-Drug Interaction Monograph for SIMPONI (Golimumab) #4**

---

**Perpetrator drug:**
INN name/US brand name: Golimumab/Simponi
Class/Molecular weight: Human IgG1 monoclonal antibody/150 kDa

*(Continued)*

**Victim drugs:**

| | |
|---|---|
| INN name/US brand name: | Various |
| Class/Molecular weight: | CYP 450 substrates |

**Description of the effect of the DDI:**
Coadministration of Simponi may affect the clearance of concurrently administered CYP 450 substrates.

**Description of the (proven, putative or hypothesized) mechanism of the DDI:**
The formation of CYP450 enzymes may be suppressed by increased levels of cytokines (e.g., TNF-$\alpha$) during chronic inflammation. Therefore, it is expected that for a molecule that antagonizes cytokine activity, such as golimumab, the formation of CYP450 enzymes could be normalized.

**Clinical consequences:**
Upon initiation or discontinuation of golimumab in patients being treated with CYP450 substrates with a narrow therapeutics index, monitoring of the effect (e.g., warfarin) or drug concentration, cyclosporine or theophylline) is recommended and the individual dose of drug product may be adjusted as needed.

**Reference:** Simponi prescribing information, Micromedex.

---

**Drug-Drug Interaction Monograph for SIMULECT (Basiliximab) #1**

**Perpetrator drug:**

| | |
|---|---|
| INN name/US brand name: | Basiliximab/Simulect |
| Class/Molecular weight: | Monoclonal antibody/143.8 kDa |

**Victim drug:**

| | |
|---|---|
| INN name/US brand name: | Tacrolimus/Prograf |
| Class/Molecular weight: | Immunosuppressant/804 Da |

**Description of the effect of the DDI:**
Concurrent use of basiliximab and tacrolimus may result in increased plasma tacrolimus trough concentrations and increased risk of tacrolimus toxicity.

**Description of the (proven, putative or hypothesized) mechanism of the DDI:**
Cytokine-induced alterations in CYP3A4-mediated metabolism of tacrolimus.

**Clinical consequences:**
Monitor tacrolimus plasma concentrations closely in patients receiving concomitant basiliximab treatment. Tacrolimus blood trough levels should be closely monitored 1 to 2 months post-transplant and dose-adjustments made as necessary.

**Reference:** Simulect labeling information.
Micromedex.

---

**Drug-Drug Interaction Monograph for SIMULECT (Basiliximab) #2**

**Perpetrator drug:**
INN name/US brand name: Echinacea
Class/Molecular weight: Herbal preparation

**Victim drug:**
INN name/US brand name: Basiliximab/Simulect
Class/Molecular weight: Monoclonal antibody/143.8 kDa

**Description of the effect of the DDI:**
Concurrent use of Echinacea and basiliximab may result in decreased effectiveness of basiliximab. Theoretically, Echinacea may antagonize the immunosuppressive effect of basiliximab. Human studies have shown an increase in granulocytes, monocytes, and lymphocytes in response to treatment with Echinacea, increased phagocytic activity, and increased natural killer cell function. Animals treated with Echinacea have shown increased immune response to antigens, and increased macrophage activation.

**Description of the (proven, putative or hypothesized) mechanism of the DDI:**
Immune system stimulation by Echinacea.

**Clinical consequences:**
Concomitant use of Echinacea and basiliximab should be avoided due to the antagonistic effect an immunostimulant such as echinacea may have on basiliximab, an immunosuppressant.

**Reference:** Simulect labeling information. Micromedex.

**Drug-Drug Interaction Monograph for SOLIRIS (Eculizumab)**

**Perpetrator drug:**
INN name/US brand name: Eculizumab/Soliris
Class/Molecular weight: Human monoclonal antibody/148 kDa

**Victim drugs:**
INN name/US brand name: Bioflavonoids; zinc glycinate; vitamin E; cholecalciferol, zinc sulfate
Class/Molecular weight: Vitamins and immunostimulants

**Description of the effect of the DDI:**
Drug interaction studies have not been performed with eculizumab.

**Description of the (proven, putative or hypothesized) mechanism of the DDI:**
Theoretically, agents that are thought to have immunostimulant properties such as echinacea, vitamin E and zinc may antagonize the pharmacologic effects of immunosuppressants.

**Clinical consequences:**
Clinical cases of drug interactions have not been reported.

**Reference:** Soliris prescribing information.

**Drug-Drug Interaction Monograph for SOMAVERT (Pegvisomat)**

**Perpetrator drug:**

| | |
|---|---|
| INN name/US brand name: | Pegvisomant/Somavert |
| Class/Molecular weight: | Recombinant protein/52 kDa |

**Victim drugs:**

| | |
|---|---|
| INN name/US brand name: | Octreotide/Sandostatin<br>Somatuline/Lanreotide |
| Class/Molecular weight: | Sandostatin: Cyclic octapeptide/1,02 kDa<br>Lanreotide: Cyclical octapeptide analog of the natural hormone, somatostatin/1,1 kDa |

**Description of the effect of the DDI:**
Concurrent use of octreotide or lanreotide and pegvisomant may result in elevated liver enzymes.

**Description of the (proven, putative or hypothesized) mechanism of the DDI:**
Concomitant use of somatostatin analogs, like octreotide and lanreotide, and pegvisomant may result in increased hepatic enzyme levels. Monitoring of liver function tests and appropriate dosage adjustment of pegvisomant is recommended during coadministration.

**Clinical consequences:**
Concomitant use of octreotide or lanreotide with pegvisomant may result in increased hepatic enzyme levels.

**Reference:** Somavert prescribing information, Micromedex.

**Drug-Drug Interaction Monograph for STELARA (Ustekinumab) #1**

**Perpetrator drug:**

| | |
|---|---|
| INN name/US brand name: | Ustekinumab/Stelara |
| Class/Molecular weight: | Human IgG1 monoclonal antibody/146 kDa |

**Victim drugs:**

| | |
|---|---|
| INN name/US brand name: | Various |
| Class/Molecular weight: | Live Vaccines |

**Description of the effect of the DDI:**
Concurrent use of ustekinumab and live vaccines may result in an increased risk of secondary transmission of infection by the live vaccine.

**Description of the (proven, putative or hypothesized) mechanism of the DDI:**
Altered cellular immune response.

**Clinical consequences:**
Although data are not available on the effects of concurrent vaccination, the manufacturer does not recommend the administration of live vaccines in patients receiving ustekinumab. Bacillus of Calmette and Guerin live vaccines should not be administered during treatment with ustekinumab or for 1 year prior to initiation or 1 year after discontinuation of ustekinumab.

**Reference:** Stelara prescribing information, Micromedex.

**Drug-Drug Interaction Monograph for STELARA (Ustekinumab) #2**

**Perpetrator drug:**
INN name/US brand name: Ustekinumab/Stelara
Class/Molecular weight: Human IgG1 monoclonal antibody/146 kDa

**Victim drugs:**
INN name/US brand name: Various
Class/Molecular weight: CYP 450 substrates

**Description of the effect of the DDI:**
Coadministration of Stelara may affect the clearance of concurrently administered CYP 450 substrates.

**Description of the (proven, putative or hypothesized) mechanism of the DDI:**
The formation of CYP450 enzymes can be altered by increased levels of certain cytokines (e.g., IL-1, 289 IL-6, IL-10, TNFα, IFN) during chronic inflammation. Thus, ustekinumab, an antagonist of IL-12 and IL-23, could normalize the formation of CYP450 enzymes.

**Clinical consequences:**
Upon initiation of ustekinumab in patients who are receiving concomitant CYP450 substrates, particularly those with a narrow therapeutic index, monitoring for therapeutic effect (e.g., for warfarin) or drug concentration (e.g., for cyclosporine) should be considered and the individual dose of the drug adjusted as needed.

**Reference:** Stelara prescribing information, Micromedex.

**Drug-Drug Interaction Monograph for STIMATE (Desmopressin acetate)**

**See DDAVP (Desmopressin) drug-drug interaction monographs #1-#2**

**Reference:** Stimate prescribing information

**Drug-Drug Interaction Monograph for SYLATRON (Peginterferon)**

**Perpetrator drug:**
INN name/US brand name: Peginterferon alfa-2B/Sylatron
Class/Molecular weight: Recombinant interferon alfa-2b)/40 kDa

**Victim drug:**
Class: Drugs metabolized by CYP2C9 or CYP2D6

**Description of the effect of the DDI:**
When administering peginterferon alfa with medications metabolized by CYP2C9 or CYP2D6, the therapeutic effect of these drugs may be altered.

The coadministration of interferons and zidovudine may result in severe bone marrow toxicity, most often manifested as granulocytopenia and/or thrombocytopenia.

The risk of seizures may be increased during coadministration of tramadol or bupropion with any substance that can reduce the seizure threshold, including inferons. Interferons also exhibit CNS- and/or respiratory-depressant effects, which may be enhanced during their concomitant use with tramadol.

*(Continued)*

**Description of the (proven, putative or hypothesized) mechanism of the DDI:**
In a study in healthy subjects, a measure of CYP2C9 activity increased to 125% (90% CI: 116% to 135%) of baseline, whereas a measure of CYP2D6 activity decreased to 51% (90% CI: 38% to 67%) of baseline when coadministered with peginterferon alfa-2b at 1 mcg/kg/week. A measure of CYP1A2 and CYP3A4 activity did not show clinically meaningful changes.

Interferons and zidovudine agents are individually myelotoxic, with potentially additive or synergistic effects.

**Clinical consequences:**
Close monitoring for drugs metabolized by CYP2C9 or CYP2D6.

Caution is advised if concomitant therapy is necessary, particularly with the alfa interferons. Complete blood counts, differential and platelet counts should be performed regularly and drug dosages adjusted accordingly. Patients should be advised to promptly report symptoms such as chills, fever, sore throat, pale skin, or unusual fatigue to their physician. Cessation of one or all drugs may be necessary if toxicity develops

Caution is advised if tramadol or bupropion is administered with any substance that can reduce the seizure threshold such as interferons, particularly in the elderly and in patients with epilepsy, a history of seizures, or other risk factors for seizures (e.g., head trauma, brain tumor, metabolic disorders, alcohol and drug withdrawal, CNS infections).

**Reference:** Sylatron prescribing information.

---

**Drug-Drug Interaction Monograph for SYNAGIS (Palivizumab)**

---

**Perpetrator drug:**

| | |
|---|---|
| INN name/US brand name: | Palivizumab/Synagis |
| Class/Molecular weight: | Humanized IgG1 monoclonal antibody/148 kDa |

**Victim drug:**

| | |
|---|---|
| Class: | Abciximab/Reopro; Belimumab/Benlysta |
| Class/Molecular weight: | Monoclonal antibodies |

**Description of the effect of the DDI:**
When abciximab or belimumab is used in combination with palivizumab, drug interactions could occur. As a result, the risk for allergic reactions, infections, or low blood cell counts can increase.

**Description of the (proven, putative or hypothesized) mechanism of the DDI:**
If patients received abciximab in the past, they may have developed antibodies to the medication. These antibodies can increase the chance of having an allergic reaction to Synagis or other monoclonal antibody medications.

Using palivizumab with belimumab may increase the risk for side effects, such as infection or low blood cell counts.

**Clinical consequences:**
No specific recommendations.

**Reference:** Synagis prescribing information.

---

**Drug-Drug Interaction Monograph for TEV-TROPIN (Somatropin)**

**See ACCRETROPIN (Somatropin) drug-drug interaction monographs #1– #4**

**Reference:** Tev-Tropin prescribing information

**Drug-Drug Interaction Monograph for TNKASE (Tenecteplase)**

**Perpetrator drug:**

INN name/US brand name: Tenecteplase/TNKase
Class/Molecular weight: Enzyme/58.9 kDa

**Victim drug:**

Class: Low molecular-weight heparins (enoxaparin, dalteparin, parnaparin, nadroparin, reviparin, tinzaparin, ardeparin, and certoparin)

Class: Anticoagulants (heparin, drotrecogin alfa, dabigatran, warfarin, phenprocoumon, phenindione, acenocoumarol, danaparoid, bivalirudin, argatroban, lepirudin, desirudin, fondaparinux, rivaroxaban, protein C (Human)).

Class: Platelet aggregation inhibitors (acetylsalicylic acid, dipyridamole, GP IIb/IIIa inhibitors, ticlopidine)

Class: Thrombolytic agents (streptokinase, urokinase, alteplase, anistreplase, reteplase, lepirudin)

**Description of the effect of the DDI:**
Increasing risk of bleeding might be observed if anticoagulants and drugs that alter platelet function are administered prior to, during, or after TNKase therapy.

**Description of the (proven, putative or hypothesized) mechanism of the DDI:**
NA

**Clinical consequences:**
Monitor for signs of bleeding.

**Reference:** Tnkase prescribing information, Micromedex.

**Drug-Drug Interaction Monograph for TRASYLOL (Aprotinin) #1**

**Perpetrator drug:**

INN name/US brand name: Aprotinin/Trasylol
Class/Molecular weight: Bovine proteinase inhibitor/6.5 kDa

**Victim drug:**

Class: Fibrinolytic agents

*(Continued)*

**Description of the effect of the DDI:**
Aprotinin is known to have antifibrinolytic activity and, therefore, may inhibit the effects of fibrinolytic agents.

**Description of the (proven, putative or hypothesized) mechanism of the DDI:**
NA

**Clinical consequences:**
No specific recommendations are available.

**Reference:** Trasylol prescribing information.

---

**Drug-Drug Interaction Monograph for TRASYLOL (Aprotinin) #2**

---

**Perpetrator drug:**
INN name/US brand name: Aprotinin/Trasylol
Class/Molecular weight: Bovine proteinase inhibitor/6.5 kDa

**Victim drug:**
INN name/US brand name: Captopril/Capoten
Class/Molecular weight: Small molecule

**Description of the effect of the DDI:**
In study of nine patients with untreated hypertension, aprotinin infused intravenously in a dose of 2 million KIU over two hours blocked the acute hypotensive effect of 100 mg of captopril.

**Description of the (proven, putative or hypothesized) mechanism of the DDI:**
NA

**Clinical consequences:**
No specific recommendations are available.

**Reference:** Trasylol prescribing information.

---

**Drug-Drug Interaction Monograph for TRASYLOL (Aprotinin) #3**

---

**Perpetrator drug:**
INN name/US brand name: Aprotinin/Trasylol
Class/Molecular weight: Bovine proteinase inhibitor/6.5 kDa

**Victim drug:**
INN name/US brand name: Heparin
Class/Molecular weight: Carbohydrate/12–15 kDa

**Description of the effect of the DDI:**
Aprotinin, in the presence of heparin, has been found to prolong the activated clotting time

**Description of the (proven, putative or hypothesized) mechanism of the DDI:**
Aprotinin prolongs whole blood clotting times by a different mechanism than heparin. In the presence of aprotinin, prolongation is dependent on the type of whole blood clotting test employed. If an activated clotting time (ACT) is used to determine the effectiveness of heparin anticoagulation, the prolongation of the ACT by aprotinin may lead to an overestimation of the degree of anticoagulation, thereby leading to inadequate anticoagulation.

**Clinical consequences:**
No specific recommendations are available.

**Reference:** Trasylol prescribing information.

---

**Drug-Drug Interaction Monograph for TYSABRI (Natalizumab) #1**

---

**Perpetrator drug:**
INN name/US brand name: Interferon beta-1a/Avonex
Class:/Molecular weight: Protein/20.0 kDa

**Victim drug:**
INN name/US brand name: Natalizumab/Tysabri
Class/Molecular weight: Monoclonal Antibody/149 kDa

**Description of the effect of the DDI:**
Natalizumab exposure parameters (i.e., $C_{max}$, AUC) and terminal elimination half-life appeared to be somewhat higher when administered in combination with interferon beta 1a. $C_{max}$ increased from 151.9 to 174.2 μg/L while the half-life increased from 142.6 to 175.3 h.

**Description of the (proven, putative or hypothesized) mechanism of the DDI:**
No dose adjustments have been recommended.

**Clinical consequences:**
No clinical consequences have been observed

**Reference:** Tysabri prescribing information.

Vollmer et al. An open-label safety and drug interaction study of natalizumab (Antegren™) in combination with interferon-beta (Avonex®) in patients with multiple sclerosis. Mult Scler. 2004;10(5):511–20.

---

**Drug-Drug Interaction Monograph for TYSABRI (Natalizumab) #2**

---

**Perpetrator drug:**
INN name/US brand name: Immunosuppressants

**Victim drug:**
INN name/US brand name: Natalizumab/Tysabri Change formatting
Class/Molecular weight: Monoclonal Antibody/149 kDa

**Description of the effect of the DDI:**
Because of the potential for increased risk of PML(Progressive multifocal leukoencephalopathy) and other infections, Crohn's disease patients receiving natalizumab should not be treated with concomitant immunosuppressants (e.g., 6- mercaptopurine, azathioprine, cyclosporine, or methotrexate) or inhibitors of TNF-α. Corticosteroids should be reduced in patients with Crohn's disease who will begin with natalizumab therapy.

**Description of the mechanism of the DDI:**
NA

*(Continued)*

**Clinical consequences:**
In Crohn's disease clinical studies, opportunistic infections (*Pneumocystis carinii* pneumonia, pulmonary *Mycobacterium avium intracellulare*, bronchopulmonary aspergillosis, and *Burkholderia cepacia*) have been observed in <1% of TYSABRI-treated patients; some of these patients were receiving concurrent immunosuppressants.

Multiple Sclerosis patients receiving chronic immunosuppressant or immunomodulatory therapy should not be treated with natalizumab.

**Reference:** Tysabri prescribing information.

---

**Drug-Drug Interaction Monograph for ULTRESA/[Pancrelipase(Amylase, lipase, protease)]**

---

**Perpetrator drug:**
INN name/US brand name: Pancrelipase/Ultresa
Class/Molecular weight: Enzyme mixture/~131 kDa

**Victim drug:**
INN name/US brand name: L-Methylfolate
Class/Molecular weight: Small Molecule/459 Da

**Description of the effect of the DDI:**
Concurrent use of L-methylfolate and pancrelipase may result in decreased L-methylfolate serum concentrations.

**Description of the (proven, putative or hypothesized) mechanism of the DDI:**
Decreased folate absorption.

**Clinical consequences:**
No clinical consequences have been observed.

**Reference:** Micromedex.

---

**Drug-Drug Interaction Monograph for VORAXAZE (Glucarpidase)**

---

**Perpetrator drug:**
INN name/US brand name: Leucovorin/Wellcovorin
Class/Molecular weight: Folate Analog

**Victim drug:**
INN name/US brand name: Glucarpidase/Voraxaze
Class/Molecular weight: Homodimer Protein/83 kDa

**Description of the effect of the DDI:**
Coadministration of leucovorin and glucarpidase results in increased plasma concentrations of methotrexate.

**Description of the (proven, putative or hypothesized) mechanism of the DDI:**
Leucovorin competes with methotrexate for the same binding sites on glucarpidase. This action diminishes the efficacy of glucarpidase in decreasing elevated methotrexate concentrations.

**Clinical consequences:**
Do not give leucovorin within 2 hours of glucarpidase administration.

**Reference:** Voraxaze prescribing information.

---

**Drug-Drug Interaction Monograph for WYDASE (Hyaluronidase)**

**Perpetrator drug:**

| | |
|---|---|
| INN name/US brand name: | Hyaluronidase/Wydase |
| Class/Molecular weight: | Enzyme/53.9 kDa |

**Victim drug:**

| | |
|---|---|
| Class: | Anesthetics (Mepivacaine, bupivacaine, benzocaine, etidocaine, cocaine, lidocaine, propxycaine, tetracaine, ropivacaine, bucaine, proparacaine, dibucaine, procaine, chloroprocaine, prilocaine) |

**Description of the effect of the DDI:**
Concurrent use of anesthetics and hyaluronidase may result in an increased incidence of a systemic reaction to the anesthetic.

**Description of the (proven, putative or hypothesized) mechanism of the DDI:**
Hyaluronidase increases the duration of anesthetics in inferior alveolar nerve blocks.

**Clinical consequences:**
Monitor patients for local anesthetic toxicity. Hyaluronidase effectively increases topical diffusion of local anesthetics. However, due to its propensity to increase the incidence of systemic reactions to local anesthetics and because there are newer anesthetic agents which penetrate very effectively, its use can no longer be justified.

**Reference:** Horliana TAC, et al. Hyaluronidase increases the duration of mepivacaine in inferior alveolar nerve blocks. J Oral Maxillofac Surg. 2008;66(2):286–90.
Micromedex.

**Drug-Drug Interaction Monograph for XGEVA (Denosumab)**

**Perpetrator drug:**

| | |
|---|---|
| INN name/US brand name: | Denosumab/Xgeva |
| Class/Molecular weight: | Human IgG2 Monoclonal Antibody/147 kDa |

**Victim drug:**

| | |
|---|---|
| Class: | Anticancer chemotherapeutics |

**Description of the effect of the DDI:**
No change in pharmacokinetic parameters or toxicity has been observed under co-administration of denosumab and anticancer agents.

Serum denosumab concentrations at 1 and 3 months were not altered by concomitant chemotherapy.

**Description of the (proven, putative or hypothesized) mechanism of the DDI:**
No dose adjustments have been recommended.

**Clinical consequences:**
No clinical consequences have been observed

**Reference:** Xgeva prescribing information.

**Drug-Drug Interaction Monograph for XIAFLEX (Collagenase *Clostridium histolyticum)***

**Perpetrator drug:**

| | |
|---|---|
| INN name/US brand name: | Collagenase *Clostridium histolyticum*/Xiaflex |
| Class/Molecular weight: | Two polypeptides/114 kDa and 113 kDa |

**Victim drug:**

| | |
|---|---|
| Class: | Anticoagulants |

**Description of the effect of the DDI:**
Possibility of developing an ecchymosis/contusion or an injection site hemorrhage.

**Description of the (proven, putative or hypothesized) mechanism of the DDI:**
Additive anticoagulation.

**Clinical consequences:**
Collagenase should be used with caution in patients receiving concomitant anticoagulants (except for low-dose aspirin).

**Reference:** Xiaflex prescribing information.

**Drug-Drug Interaction Monograph for XIGRIS (Drotrecogin alfa)**

**Perpetrator drug:**

| | |
|---|---|
| INN name/US brand name: | Drotrecogin alfa/Xigris |
| Class/Molecular weight: | Glycoprotein/55 kDa |

**Victim drug:**

| | |
|---|---|
| Class: | Anticoagulants |

**Description of the effect of the DDI**
Concurrent use of drotrecogin alfa and anticoagulants may result in increased risk of bleeding.

**Description of the (proven, putative or hypothesized) mechanism of the DDI:**
Additive anticoagulation.

**Clinical consequences:**
Use caution with coadministration of drotrecogin alfa and anticoagulants in therapeutic dosages for thromboembolism. No drotrecogin alfa dosage adjustment is necessary when coadministered with heparin using the prophylactic doses for venous thromboembolism.

**Reference:** Xigris prescribing information.

**Drug-Drug Interaction Monograph for ZALTRAP (Ziv-Aflibercept) #1**

**Perpetrator drug:**

| | |
|---|---|
| INN name/US brand name: | Ziv-Aflibercept/Zaltrap |
| Class/Molecular weight: | Recombinant fusion protein/155 kDa |

**Victim drug:**

| | |
|---|---|
| INN name/US brand name: | Irinotecan/Camptosar |
| Class/Molecular weight: | Antineoplastic agent/677.2 Da |

**Description of the effect of the DDI:**
No change in pharmacokinetic parameters or toxicity has been observed under co-administration of ziv-aflibercept and irinotecan.

**Description of the (proven, putative or hypothesized) mechanism of the DDI:**
No dose adjustments have been recommended.

**Clinical consequences:**
No clinical consequences have been observed.

**Reference:** Zaltrap prescribing information.

---

**Drug-Drug Interaction Monograph for ZALTRAP (Ziv-Aflibercept) #2**

---

**Perpetrator drug:**

| | |
|---|---|
| INN name/US brand name: | Ziv-Aflibercept/Zaltrap |
| Class/Molecular weight: | Recombinant fusion protein/155 kDa |

**Victim drug:**

| | |
|---|---|
| INN name/US brand name: | Fluorouracil/Efudex |
| Class/Molecular weight: | Antineoplastic antimetabolite/130.1 Da |

**Description of the effect of the DDI:**
No change in pharmacokinetic parameters or toxicity has been observed under co-administration of ziv-aflibercept and fluorouracil.

**Description of the (proven, putative or hypothesized) mechanism of the DDI:**
No dose adjustments have been recommended.

**Clinical consequences:**
No clinical consequences have been observed.

**Reference:** Zaltrap prescribing information.

---

**Drug-Drug Interaction Monograph for ZENAPAX (Daclizumab) #1**

---

**Perpetrator drug:**

| | |
|---|---|
| INN name/US brand name: | Daclizumab/Zenapax |
| Class/Molecular weight: | Humanized IgG1 Monoclonal Antibody/144 kDa |

**Victim drug:**

| | |
|---|---|
| Class: | Immunosuppressive agents (cyclosporine, mycophenolate mofetil, ganciclovir, acyclovir, azathioprine, corticosteroids) |

*(Continued)*

**Description of the effect of the DDI:**
In cardiac transplant recipients (n = 434) receiving concomitant cyclosporine, mycophenolate mofetil, and corticosteroids, mortality was increased in patients randomized to receive daclizumab compared with those randomized to receive placebo.

The concomitant administration of daclizumab and mycophenolate mofetil had no effect on the pharmacokinetics of mycophenolic acid.

**Description of the (proven, putative or hypothesized) mechanism of the DDI:**
No dose adjustments have been recommended.

**Clinical consequences:**
No clinical interventions have been recommended.

**Reference:** Zenapax prescribing information.

---

**Drug-Drug Interaction Monograph for ZENAPAX (Daclizumab) #2**

---

**Perpetrator drug:**

| | |
|---|---|
| Class: | Anti-Lymphocyte Antibody<br>Antibody/144 kDa |

**Victim drug:**

| | |
|---|---|
| INN name/US brand name: | Daclizumab/Zenapax |
| Class/Molecular weight: | Immunosuppressive, Humanized IgG1 Monoclonal |

**Description of the effect of the DDI:**
The use of daclizumab was associated with a higher incidence of mortality in patients receiving concomitant anti-lymphocyte antibody therapy, compared to placebo, in a large (n = 434) randomized controlled study of patients receiving cardiac transplants.

**Description of the (proven, putative or hypothesized) mechanism of the DDI:**
No dose adjustments have been recommended.

**Clinical consequences:**
No clinical interventions have been recommended.

**Reference:** Zenapax prescribing information.

---

**Drug-Drug Interaction Monograph for ZEVALIN (Ibritumomab tiuxetan)**

---

**Perpetrator drug:**

| | |
|---|---|
| INN name/US brand name: | Ibritumomab tiuxetan/Zevalin |
| Class/Molecular weight: | Immunoconjugate between monoclonal antibody and linker-chelator/148 kDa |

**Victim drug:**

| | |
|---|---|
| Class: | Anticoagulants/Antiplatelet agents |

**Description of the effect of the DDI:**
Concurrent use of ibritumomab tiuxetan and anticoagulants/antiplatelet agents may result in an increased risk of bleeding and hemorrhage.

**Description of the (proven, putative or hypothesized) mechanism of the DDI:**
NA

**Clinical consequences:**
Monitor for thrombocytopenia more frequently when using ibritumomab and anticoagulants/antiplatelet agents concomitantly. The transfusion practices may need to be modified.

**Reference:** Zevalin prescribing information.

---

**Drug-Drug Interaction Monograph for ZOLADEX (Goserelin Acetate)**

**Perpetrator drug:**
INN name/US brand name: Goserelin Acetate/Zoladex
Class/Molecular weight: Decapeptide/1,269 Da

**Victim drug:**
INN name/US brand name: Tamoxifen Citrate/Nolvadex
Class/Molecular weight: Nonsteroidal Antiestrogen/563.6 Da

**Description of the effect of the DDI:**
The effect of concurrent goserelin and tamoxifen in treating premenopausal women with early breast cancer is not clear, but may have less benefit than a mono therapy.

**Description of the (proven, putative or hypothesized) mechanism of the DDI:**
No change in pharmacokinetic parameters or toxicity has been observed under co-administration of goserelin and tamoxifen.

**Clinical consequences:**
The combination of goserelin and tamoxifen reduces the risk of first event by 24%, compared to 32% with goserelin and 27% with tamoxifen alone.

**Reference:** Zoladex prescribing information.
Sverrisdottir A, et al. Interaction between goserelin and tamoxifen in a prospective randomised clinical trial of adjuvant endocrine therapy in premenopausal breast cancer. Breast Cancer Res Treat. 2011;128(3):755–63.

---

**Drug-Drug Interaction Monograph for ZORBTIVE (Somatropin)**

**See ACCRETROPIN (Somatropin) drug-drug interaction monographs #1– #4**

**Reference:** Zorbtive prescribing information

# TUMOR-TARGETED NANOPARTICLES: STATE-OF-THE-ART AND REMAINING CHALLENGES

GAURAV BAJAJ

*Division of Clinical Pharmacology and Therapeutics, The Children's Hospital of Philadelphia, Philadelphia, PA, USA*

YOON YEO

*Department of Industrial and Physical Pharmacy, Purdue University, West Lafayette, IN, USA*

## 1.1 INTRODUCTION

Bringing a drug only to target tissues without spilling any molecule in unwanted places would be an ideal goal for any pharmacological therapies [1]. Owing to the dose-limiting side effects of chemotherapy, a number of drug delivery strategies have been developed in the context of cancer, where "targeted" therapy is most anticipated. Most tumor-targeted drug delivery systems are based on the fact that cancer cells express various molecular markers that are distinguished from those of normal cells. In particular, nanomedicines have received enormous attention in the past decades as a potential tool to increase the selectivity of chemotherapy and diagnosis, due to the small size conducive to circulation and the large surface area to volume ratio that facilitates surface functionalization. Several nanomedicine products have been launched in the market or in the clinical development stage as summarized in Table 1.1. Newer approaches are actively developed to increase target selectivity, although the number of targeted nanomedicines that reached the later phase of

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*,
Edited by Shayne C. Gad
Copyright © 2013 John Wiley & Sons, Inc.

**TABLE 1.1 Nanomedicines on the Market or in Phase II/III Clinical Development**

| Type of Nanoparticles | Formulation | Drug | Brand Name (Company) | Indication(s) | Current Status and Reference |
|---|---|---|---|---|---|
| Lipid-based or liposomes | Lipid–drug complex | Amphotericin B | Abelcet (Sigma-Tau) | Antimicrobial | Market [3] |
| | PEGylated liposome | Doxorubicin | Doxil and Caelyx (Janssen) | Breast cancer, ovarian cancer, multiple myeloma, Kaposi's sarcoma | Market [4] |
| | Non-PEGylated liposome | Doxorubicin | Myocet (Enzon) | Breast cancer | Market [5] |
| | Non-PEGylated liposome | Daunorubicin | DaunoXome (Galen) | Kaposi's sarcoma | Market [6] |
| | Non-PEGylated liposome | Cytarabine | DepoCyt (Sigma-Tau) | Lymphomatus meningitis, leukemia, glioblastoma | Market [7] |
| | Non-PEGylated liposome | Amphotericin B | Ambisome (Gilead) | Fungal infection, cryptococcal meningitis | Market [8] |
| | PEGylated liposome | Cisplatin | Lipoplatin (Regulon) | Various malignancies | Phase III, investigational drug [9–13] |
| | Heat-activated PEGylated liposome | Doxorubucin | Thermodox (Celsion Corporation) | Hepatocellular carcinoma, recurrent chest wall breast cancer | Phase III [14] |
| | Sphingomyelin-based liposomes | Vincristine sulfate (targeted, Optisome™, details not provided) | Marqibo® (Talon Therapeutics) | Non-Hodgkin's lymphoma | NDA submitted [15,16] |
| Albumin-bound particles | Albumin bound nanoparticle | Paclitaxel (Albumin) | Abraxane® (Astellas) | Breast cancer | Market [17,18] |
| Polymer-drug/protein/nucleic acid conjugates | PEG conjugated | L-Asparaginase | Oncaspar® (Sigma-Tau) | Acute lymphoblastic leukemia | Market [19] |
| | Poliglumex (poly-L-glutamic acid) conjugated | Paclitaxel | Opaxio™ or previously Xyotax (Cell Therapeutics Inc.) | Nonsmall cell lung cancer, ovarian cancer | Phase III [20–23] |

| | | | | | |
|---|---|---|---|---|---|
| | Conjugated with uniquely engineered macromolecular polymer core using specialized linkers | Irinotecan | Etirinotecan Pegol (Nektar Therapeutics) | Breast cancer, ovarian cancer, colorectal cancer | Phase III [24] |
| | PEGylated anti-VEGF aptamer | Pegaptanib sodium (VEGF) | Macugen (Eyetech, Inc.) | Age-related macular degeneration | Market [25] |
| | PEG conjugated | Recombinant human granulocyte colony-stimulating factor | Neulasta (Amgen) | Prevention of chemotherapy-associated neutropenia | Market [26] |
| Polymeric micelles | Methoxy PEG-PLA copolymer micelle | Paclitaxel | Genexol-PM® (Samyang Co.) | Breast cancer, lung cancer, ovarian cancer | Phase II [27,28] |
| Nanocrystals | Nanocrystal drug particles | Aprepitant | Emend (Merck) | Antiemetic | Market [29] |
| | Nanocrystal drug particles | Sirolimus | Rapamune (Pfizer) | Immunosuppressive | Market [29] |
| | Nanocrystal drug particles | Fenofibrate | Tricor (Abbott) | Hypercholesterolemia or mixed dyslipidemia | Market [29] |
| | Nanocrystal drug particles | Megestrol acetate | Megace ES (BMS) | Synthetic progestin | Market [29] |
| Dendrimers | Carbopol® gel (cross-linked acrylic acid) | SPL7013 (antimicrobial agent) | VivaGel (Starpharma, Dendritic Nanotechnologies Inc., United States) | Bacterial vaginosis | Market, Phase III [3,30] |
| Cyclodextrin-based nanoparticles | Transferrin receptor-targeted cyclodextrin NPs, RONDEL™ | siRNA targeting the M2 subunit of ribonucleotide reductase | CALAA-01 (Calando Pharmaceuticals) | Inhibition of tumor growth | Phase I trial, [31,32] |
| Inorganic nanoparticles | Streptavidin covalently attached to quantum nanocrystals (QDot) | — | Qdot® 800 Streptavidin Conjugate (Invitrogen) | Imaging, diagnostics, detection of proteins, nucleic acids | Market [33,34] |
| | Superparamagnetic iron oxide nanoparticles coated with carboxydextran | — | Resovist (Schering) | Hepatocellular carcinoma | Market [35] |
| | Iron nanoparticles | — | Feridex (AMAG Pharmaceuticals) | Detection of liver lesions | Market [36] |

*Notes:* The table was created through internet search, product websites, and previously published articles [37,38]. Antibody–drug conjugates are omitted, but some publications list them under nanomedicine [38].

clinical trials is relatively low considering the prevalent excitement and the investment made till date [2].

This chapter discusses the rationale of targeted nanomedicines, different approaches to achieve this goal, current challenges, and considerations for their future development. Here, nanomedicines refer to particulate materials with a diameter in nanometer scale, prepared with natural or synthetic polymers, lipids, or inorganic solids, in the form of polymer–drug conjugates, liposomes, polymeric nanoparticles (NPs) or micelles, dendrimers, nanotubes, nanocrystals, nanorods, nanoshells, or nanocages [2]. Nanomedicines used as a carrier of a drug are often called nanocarriers. Nanoparticles and nanocarriers are also interchangeably used. We will use the term "nanoparticles" to represent various types of nanocarriers in this chapter.

## 1.2 FUNCTIONS OF NANOPARTICLES

Fates of a drug entering circulation largely depend on the properties of the drug. Small hydrophilic drug molecules do not bind to proteins in the blood and are quickly eliminated from the body. Hydrophobic drugs are eliminated relatively slowly through kidneys due to plasma protein binding, metabolized in the liver to a hydrophilic metabolite, and then eliminated from the body. On the other hand, the disposition and clearance of drugs encapsulated in NPs depend relatively less on the drug itself than on the properties of NPs during circulation [39].

Systemically administered NPs should be stable enough to retain a drug during circulation and able to extravasate at the target tissues and release the drug in the target tissues and/or inside the target cells to exert an action. Advancement in the nanotechnology has made it possible to incorporate therapeutic drugs in various types of NPs. NPs have contributed to the delivery of poorly water-soluble drugs and improving the bioavailability of drugs with poor stability [2,40,41]. One of the major concerns in cancer chemotherapy is multidrug resistance as most anticancer drugs are substrates of drug efflux transporters such as P-glycoprotein. NPs can be designed to overcome the resistance by bypassing efflux transporters [2,42]. Most significantly, NPs have gained enormous interest as a way of increasing drug delivery to target tissues, in particular, tumors. A detailed discussion of the use of NPs in drug delivery is available in other review articles [2,43].

NPs can be combined with contrast or imaging agents to visualize pathological tissues and monitor the progression of diseases [44]. Liposomes, micelles, and dendrimers are used to carry paramagnetic ions of indium-111 ($^{111}$In), technetium ($^{99m}$Tc), manganese (Mn), and gadolinium (Gd). Semiconductor quantum dots or gold NPs have been widely explored for target-specific imaging as combined with various functionalization strategies [45–47].

## 1.3 TUMOR-TARGETED NANOPARTICLES

Several strategies have been employed to achieve tumor-targeted drug delivery based on NPs. We briefly introduce these targeting strategies before we discuss challenges in translating NPs to commercial products.

### 1.3.1 Passive Targeting

Delivering anticancer agents to the tumors is a multistep process. Drugs delivered into the bloodstream should reach intratumoral region from systemic circulation by crossing tumor vasculature so that drug molecules have access to the tumor cells and stroma. Many solid tumors are characterized by leaky vasculature and poor lymphatic drainage [43]. New vasculature in the tumor is formed to sustain the tumor growth and provide nutrients; however, the structure is usually abnormal and defective with holes ranging from 400 to 600 nm in diameter [48–50]. The defective structure of blood vessels surrounding tumors allows NPs to extravasate and accumulate in the tumor tissues. The impaired lymphatic drainage further helps retain NPs in the tumors. Therefore, NPs with a size less than the cutoff of holes in the vasculature have a selective advantage in reaching tumors. This phenomenon is referred to as the enhanced permeability and retention (EPR) effect [51–53].

For NPs to take advantage of the EPR effect, it is important that the NPs circulate for a long period of time. It is widely known that NPs with hydrophobic or charged surfaces are readily opsonized and taken up by the reticuloendothelial system (RES), which is responsible for clearing macromolecules or particles present in the blood before they reach the target tissues [54,55]. Therefore, most NPs are surface-modified with a hydrophilic and electrically neutral polymer such as polyethylene glycol (PEG) (PEGylated) [54,56,57]. The PEGylated surface is resistant to protein binding and thus avoids the recognition by the RES [55,58,59]. For long-term circulation and effective extravasation via leaky vasculature of tumors, ~50–300 nm is considered an optimal size [60,61].

### 1.3.2 Active Targeting

Active targeting strategy was developed to further enhance tumor accumulation of NPs beyond the level possible with long-circulating NPs. Here, NP surface is decorated with targeting ligands that can actively bind to cell receptors overexpressed on the tumor cells [2,49,62–65]. Frequently used targeting ligands include antibodies, peptides, nucleic acids, and small molecular weight ligands binding to specific receptors [2,65,66]. For example, NPs are conjugated with the monoclonal antibody teratuzumab or herceptin to target HER2 receptor on breast cancer cells [59,67]. A docetaxel-NP formulation containing a small molecule ligand which targets prostate-specific membrane antigen has recently entered a Phase I clinical trial [68].

The rationale of active targeting is that a specific interaction between targeting molecules and cell receptors will increase the amount of NPs retained in tumors. An example is shown in a biodistribution study of $^{111}$In-labeled polymeric micelle NPs [69]. Nontargeted NPs (25 nm) were rapidly cleared from plasma as compared to nontargeted NPs with a size of 60 nm, resulting in twofold less total tumor accumulation. On the other hand, another 25 nm NPs functionalized with epidermal growth factor achieved a greater tumor accumulation than nontargeted ones in mice with tumors overexpressing EGF-receptors, reaching the same level as that of 60 nm nontargeted NPs [69]. However, it should be borne in mind that the

contribution of a targeting ligand is predicated on the extravasation of NPs to the tumors; thus, the ability to survive circulation for a long period of time still remains one of the most important requirements in the active targeting strategy. For effective active targeting, the density of target receptors should be in range of $10^4$ or $10^5$ copies per cell [2,70]. In addition to the number of binding sites, binding affinity also plays an important role. Multivalency of ligands can further enhance the affinity for target cells [71]. Receptors in a cell are internalized via different pathways and go through a continuous process of recycling [72]. NPs with targeting ligands can be internalized via the receptor-mediated endocytosis pathways and release drugs inside the cells.

### 1.3.3 Target-Activated Systems

Target-activated systems refer to NPs that remain stable until they reach target tissues and are activated by the extracellular or intracellular cues [64,73]. The NPs are designed to circulate similar to passively targeted NPs and transform to a more cell-interactive form in response to conditions unique to tumor tissues or intracellular environments. These conditions include acidic pH or enzymes abundant in the tumor extracellular matrix (ECM).

***1.3.3.1 pH-Activated Systems.*** NPs responsive to acidic pH can be classified into two categories: one activated in the tumor ECM, which may have a pH as low as ~6, and the other activated in the intracellular lysosomes, which acidifies to pH 4–5. Tumors exhibit a slightly acidic pH in the range 6.8–7.2 compared to the physiological pH of 7.4 [74]. This acidity is due to the accumulation of lactic acid, induced by the increased glycolysis of tumor cells [75]. The acidosis of tumor is further enhanced as tumors develop hypoxia, which induces or selects for hyperglycolic cells [76]. The pH-sensitive NPs are composed of polymers, which are protonated at acidic pHs. Upon protonation of the polymer, the NPs are disintegrated to release drugs or change the surface properties to enter cells [64,77].

Most NPs entering cells via receptor-mediated endocytosis pathways end up in endosomes, which undergo a maturation process into late endosomes and lysosomes [72]. The luminal pH in these vesicles is acidic, varying from pH 6 (endosomes), pH 5–6 (late endosomes), to pH 4.5 (lysosomes) [72]. NPs are designed to escape the vesicles and/or release the payload in response to the low pH. Polycationic polymers such as polyethyleneimine or fusogenic peptides can be used in the formulations [78]. On entering the cell, a drug should reach its final destination to produce an action, which can be a nucleus or other organelles such as mitochondria, lysosome, and endoplasmic reticulum. NP trafficking to the specific organelles may be achieved by the use of specific peptide sequences [47,79].

***1.3.3.2 Enzymatically Activated Systems.*** Enzymes overexpressed in the tumor microenvironment are used for activating the NPs at tumor sites. An example of such enzymes is matrix metalloproteinases (MMPs) [80,81], which play a critical role in the invasion of tumor cells and angiogenesis [82]. Several studies have employed

MMPs for removing the PEG protection of NP surface at the tumor sites [83–85]. In these approaches, PEG is linked to the NP surface via a peptide linker sensitive to MMPs. Recently, cathepsin B has been utilized to activate a nanoprobe in a tumor-specific manner [86]. Cathepsin B is a lysosomal cysteine protease that plays a role in tumor progression [87]. The nanoprobe consists of glycol chitosan conjugated with a cathepsin B substrate peptide, a near infrared fluorescent dye, and a dark quencher, which self-assemble into 280 nm NPs. Upon the entrance into tumor cells, the cathepsin B substrate peptide is cleaved by intracellular cathepsin B, allowing for dequenching of the dye and emission of the fluorescence signals [86].

## 1.4 REMAINING CHALLENGES IN THE DEVELOPMENT OF TUMOR-TARGETED NANOPARTICLES

Several NP products, such as liposomes and polymer–drug conjugates, are currently in the market or in the late stage of development process, improving therapeutic efficacy and safety profiles of chemotherapeutics [36,40]. A large number of targeted NPs have been developed in the last three decades in an attempt to further improve the efficacy of NP-based therapy. However, several challenges remain to be overcome for developing NPs that can attract commercial interest and achieve significant clinical outcomes. Current limitations and challenges in the advancement of nanomedicine have been thoroughly discussed in recent review articles [36,60,88–92].

### 1.4.1 NP Stability

Stability of circulating NPs matters in at least two contexts. NPs carrying a drug must survive the body's natural ability to clear the NPs and should be able to retain a drug during circulation. The former requirement is largely addressed by surface modification with a "stealth" coating with polymers such as PEG. However, recent studies have identified potential disadvantages of PEG [73]. For example, PEG can interfere with NP–cell interactions [93] and endosomal escape of NPs [93] after the disposition of NPs. It is also suggested that PEGylated liposomes are implicated with immune responses, which lead to an accelerated blood clearance of the second dose liposomes [94,95]. Alternative polymers or surface protection strategies are actively explored [73].

Stable retention of a drug during circulation is another important challenge. Despite the thermodynamic stability in buffers, some polymeric micelles or matrix-type NPs have been shown to quickly disintegrate and/or leak the payload upon contact with various blood components [64,96–98]. Chen et al. demonstrated that polymeric micelles composed of PEG-poly(D,L-lactic acid) block-copolymer dissociated and released the encapsulated fluorescent dyes within 15 min after intravenous injection [96]. Another study showed that a drug and the carrier (liposomes) exhibited different pharmacokinetics in blood, indicating that the drug was escaping the carrier prematurely [99]. The instability of NPs in circulation is a significant (yet often neglected) problem in the development stage. Success of a new NP system

depends on the availability of a method that can reliably predict the stability of NPs in a biological environment.

### 1.4.2 Heterogeneous Tumor Vasculature

Tumor accumulation of NPs mainly depends on their passive extravasation via the leaky vasculature feeding the tumors. While the defective architecture of tumor vasculature is widely exploited by the NP-based drug delivery, it is often ignored that the tumor blood vessels are heterogeneous [100,101] and do not exhibit the same degree of leakiness [102]. The heterogeneity of the EPR effect is even greater when different tumor models are compared. A biodistribution study of $^{111}$In-labeled PEGylated liposomes in 17 patients with locally advanced cancers showed that the levels of tumor liposome uptake varied from 0.5% to 3.6% of injected dose (ID) at 72 h, which translated into 2.7–53% of ID/kg of tumor (ID/kg) [103]. This study also showed that the tumor uptake of liposomes varied significantly with the site of the primary tumors, from $5.3 \pm 2.6\%$ ID/kg in breast cancers, $18.3 \pm 5.7\%$ ID/kg in lung tumors, to $33.0 \pm 15.8\%$ ID/kg in head and neck cancers [103]. Therefore, NPs relying on the EPR effect alone are unlikely to achieve reliable and predictable clinical outcomes.

### 1.4.3 NP Distribution in Tumors

The lack of functional lymphatics and the high vascular permeability result in elevation of the pressure in the interstitial tissue of solid tumors [104]. Furthermore, tumors develop various conditions to increase the stiffness of tissue [105]. The high interstitial fluid pressure and the stiffness of the tissue interfere with migration of a drug in tumor, limiting its effect on the central region of the tumor, a potential source of tumor relapse and metastasis [106–108].

Drug delivery with NPs is more significantly influenced by this challenge because of the size and surface properties [105]. Goodman et al. predicted that movement of 100–200 nm NPs in tumor spheroids would be much restricted even with the treatment of an ECM-degrading enzyme [109]. It was also experimentally demonstrated that the intratumoral distribution of 60 nm polymeric micelles was limited and localized in proximity to the blood vessel [69]. The size restriction is particularly problematic in the treatment of tumors with hypovascular and hypopermeable properties such as pancreatic tumors [110]. In the case of NPs employing active targeting strategy, high affinity binding to the receptors can lead to the "binding site barrier effect" in solid tumors, where the NP–receptor interactions may prevent migration of NPs to the interior of tissues [2,69,111,112]. While an active targeting strategy has the potential to improve the delivery of drugs through the intracellular uptake of NPs, the effect may be limited to the periphery of tumors when the size and binding site barrier play a dominant role in intratumoral disposition of the NPs.

### 1.4.4 Regulatory Considerations

The unique properties of NPs enabled by the size and surface pose different safety issues than free drugs and larger dosage forms [113]. The U.S. Food and Drug

Administration's Nanotechnology Task Force, formed in August 2006, solicited the agency's action to develop a regulatory guidance for manufacturers and researchers [114]. However, much remains uncertain as to how their safety should be tested and how much time and cost will be required for the approval of NP products, especially those designed with multiple functionalities. Such uncertainties negatively influence the investors and development partners, when the new product brings about only mild improvement over existing products.

New drug products must pass the regulatory scrutiny that examines their safety, effectiveness, and potential hazards. Reproducibility and predictability of the product's performance are the most important quality criteria under any test methods and models. Accordingly, it is important to ensure that NPs have consistent properties, such as size, charge, shape, and density of surface functional groups. Increasing complexity of a formulation increases the difficulty of quality control as well as the development cost. Therefore, the current effort to develop new types of NPs must undergo a careful cost–benefit analysis prior to a significant investment.

## 1.5 FUTURE PERSPECTIVES

Several recent studies attempt to address the challenges discussed so far [105]. For example, Maeda et al. proposed the use of drugs that artificially augment the EPR effect [102]. Angiotensin II was used to elevate systemic blood pressure [115], and nitroglycerin or other agents inducing nitric oxide production were used to facilitate vascular blood flow and permeability [116]. These approaches have shown the potential to overcome the heterogeneity of the EPR effect. To overcome the limitations in NP distribution in tumors, ECM-degrading enzymes such as hyaluronidase and collagenase were proposed as a potential aid to reduce the stiffness of ECM and the interstitial fluid pressure [109,117]. While ongoing studies are likely to make these challenges under control, additional issues await future efforts. Here we discuss the remaining homework for the advancement of nanomedicine.

### 1.5.1 *In Vitro* Models

Preclinical studies of anticancer drugs and NPs are first carried out in cell culture models. The cell culture model is a convenient and inexpensive way of screening the effectiveness of a drug or an NP formulation. However, it is not unusual that these drugs and NPs showing great efficacy in cell culture models are often proven much less effective *in vivo*. The lack of predictability is due in part to the fact that a single cell population grown in a two-dimensional (2D) layer barely reflects the nature of human tumors, which involves an intricate cross talk between different cell populations and three-dimensional (3D) organization of cells and ECM [118]. There is an increasing understanding of the significance of 3D architecture of tumors, cell populations, matrix composition and mechanics, and the gradients of signaling materials in modeling tumors *in vitro*. Accordingly, active efforts are made to build

more realistic *in vitro* tumor models [119–121]. Lessons learned from tissue engineering and regenerative medicine are urgently solicited in these efforts.

### 1.5.2 *In Vivo* Models

Owing to the relatively low cost and well-established protocols, mouse models with allograft or human xenograft tumors are widely used in the evaluation of *in vivo* performance of NPs. In these models, cancer cells are inoculated (typically subcutaneously) in immunodeficient mice, allowed to grow into visually identifiable tumors, and treated with a new formulation to examine the pharmacokinetics, biodistribution, and pharmacological effects. However, several cases show that the results of preclinical studies are poorly correlated with Phase III outcomes. For example, in a preclinical study with mice bearing C-26 colon carcinoma, PEGylated liposomes carrying doxorubicin achieved significant survival benefits than free doxorubicin, attributable to their accumulation in tumors [122]. On the other hand, the PEGylated liposomal doxorubin was at best equivalent to free doxorubin in clinical efficacy (progression-free survival and overall survival) in a randomized Phase III trial with metastatic breast cancer patients [123]. In another example, a macromolecular conjugate of PTX and poly(L-glutamic acid) (PTX poliglumex) demonstrated a prolonged circulation half-life and greater tumor uptake and antitumor activity as compared to Taxol (PTX solubilized with Cremophor EL) at the same dose in a mouse model with syngeneic ovarian carcinoma [124]. However, the developer officially withdrew its application for a marketing authorization of PTX poliglumex for the first-line treatment of lung cancer patients because of the lack of therapeutic advantages over the comparators and unexplained toxicity [125].

In explaining the gap between the results of xenograft models and clinical outcomes, several limitations of current animal models may be considered [118,126]. First, the size and growth rate of tumors in mice are not comparable to those of human patients. Human patients with tumors large enough to detect with visual inspection would be candidates for surgical debulking rather than chemotherapy. Metastatic or microscopic residual tumors would be more appropriate subjects of targeted chemotherapy, but not much is known about the vasculature structure of those tumors and the effectiveness of the EPR effect. Second, since many animal models employ allografts or human tumor xenografts, the use of immunodeficient mice is inevitable. The potential consequences of immune responses to NPs are underrated in these models. Third, when human xenografts are inoculated in mouse models, the tumors recruit substances to build ECM from mice as they grow. The potential impact of this artificial arrangement on the architecture of ECM, cell–ECM interactions, and tumor propagation is barely considered in the interpretation of preclinical studies. Fourth, even though a xenograft can represent important attributes of the original tumors, it is uncertain whether it captures the genetic and epigenetic variability of tumors in its entirety [118].

When preclinical studies in animal models are essential steps providing the rationale for clinical trials, successful development of NP products depends critically

on the availability of reliable animal models. At this point, it is yet unclear what would be the best animal model for predicting clinical outcomes of a drug product. One potential alternative to the subcutaneous xenograft is an orthotopic model, where a xenograft grown in proximity to the tissues or organs from which the tumor cell line was derived [118]. This model has advantages over the subcutaneous model as it provides an environment closer to a normal milieu of the tumor. One may also envision that genetically modified animals will have a promise in recapitulating the progression of human tumors and increasing the predictability of clinical outcomes.

### 1.5.3 New Targets

Although the current state-of-the-art targeting strategies have largely relied on specific interactions between cell surface markers and ligands, a number of studies have also found that the efficacy demonstrated *in vitro* or *in vivo* model does not translate at the clinical level. This has partly to do with the fact that a tumor is not a collective mass of a single-cell population expressing a consistent level of markers at all times. To the contrary, a tumor is highly heterogeneous and dynamic both in composition and genetic makeup. This makes it difficult to achieve a consistent outcome with an NP targeting a single type of molecular target [60,91]. Targeting multiple markers simultaneously is a conceivable alternative to the current targeting strategy. In addition, quiescent cancer cells or tumor-initiating cells are suggested as a potential target of drug delivery [127]. The tumor-initiating cells or cancer stem cells (CSCs), a subpopulation of cancer cells similar to stem cells, are believed to differentiate into new tumors when specific conditions are met. CSCs are known to be resistant to current radio- and chemotherapy, and some believe that the surviving CSCs are partly responsible for metastasis and recurrence of the disease [127]. Potential targets for NP-based drug delivery to CSCs include cellular markers such as CD44 and CD133 and hypoxic conditions that lead to the formation of CSC niches [127].

### 1.5.4 New Therapeutic Agents

While NPs have contributed to reducing toxic side effects of traditional chemotherapeutic drugs [123,128], their accumulation (and side effects) in the RES and bystander organs is currently inevitable. NPs may be a more useful tool for the delivery of alternative therapeutic agents, which are inherently specific to tumor cells (thus benign to normal tissues) but unstable in circulation or do not have an effective means to access target cells. Nucleic acid therapeutics such as small-interfering RNA, microRNA, or antisense oligonucleotides have proven effective in suppressing intrinsic drug resistance and survival of cancer cells, but their delivery is challenging. Various NPs are currently developed for the systemic delivery of these gene therapeutics [129], and one of them has entered a Phase I clinical trial in 2008 [130].

**Acknowledgments**

This work was supported by NSF DMR-1056997 and NIH R21 CA135130.

## REFERENCES

1. Strebhardt, K., Ullrich, A. (2008). Paul Ehrlich's magic bullet concept: 100 years of progress. *Nature Reviews Cancer 8*, 473–480.
2. Peer, D., Karp, J. M., Hong, S., Farokhzad, O. C., Margalit, R., Langer, R. (2007). Nanocarriers as an emerging platform for cancer therapy. *Nature Nanotechnology 2*, 751–760.
3. Diamond, M. and The Sepracoat Adhesion Study Group. (1998). Reduction of de novo postsurgical adhesions by intraoperative precoating with Sepracoat (HAL-C) solution: a prospective, randomized, blinded, placebo-controlled multicenter study. *Fertility and Sterility 69*, 1067–1074.
4. Doxil Available at http://www.doxil.com/assets/DOXIL_PI_Booklet.pdf Accessed June 30, 2012.
5. Myocet Available at http://www.ema.europa.eu/ema/index.jsp?curl=pages/medicines/human/medicines/000297/human_med_000916.jsp&mid=WC0b01ac058001d124 Accessed June 30, 2012.
6. Daunoxome Available at http://www.daunoxome.com/ Accessed June 30, 2012.
7. Depocyt Available at http://depocyt.com/patients/about.asp Accessed June 30, 2012.
8. Ambisome Available at http://www.ambisome.com/MOA.aspx Accessed June 30, 2012.
9. Lipoplatin Available at http://lipoplatin.com/info_drug.php Accessed June 30, 2012.
10. Boulikas, T. (2009). Clinical overview on Lipoplatin: a successful liposomal formulation of cisplatin. *Expert Opinion on Investigational Drugs 18*, 1197–1218.
11. Fantini, M., Gianni, L., Santelmo, C., Drudi, F., Castellani, C., Affatato, A., Nicolini, M., Ravaioli, A. (2011). Lipoplatin treatment in lung and breast cancer. *Chemotherapy Research and Practice 2011*, 125–192.
12. Mylonakis, N., Athanasiou, A., Ziras, N., Angel, J., Rapti, A., Lampaki, S., Politis, N., Karanikas, C., Kosmas, C. (2010). Phase II study of liposomal cisplatin (Lipoplatin) plus gemcitabine versus cisplatin plus gemcitabine as first line treatment in inoperable (stage IIIB/IV) non-small cell lung cancer. *Lung Cancer 68*, 240–247.
13. Stathopoulos, G. P., Boulikas, T. (2012). Lipoplatin formulation review article. *Journal of Drug Delivery 2012*, 581363.
14. Thermodox http://celsion.com/docs/technology_thermodox Accessed December 12, 2012.
15. Marqibo Available at http://www.talontx.com/pipeline.php Accessed June 30, 2012.
16. Rodriguez, M. A., Pytlik, R., Kozak, T., Chhanabhai, M., Gascoyne, R., Lu, B., Deitcher, S. R., Winter, J. N. (2009). Vincristine sulfate liposomes injection (Marqibo) in heavily pretreated patients with refractory aggressive non-Hodgkin lymphoma: report of the pivotal phase 2 study. *Cancer 115*, 3475–3482.
17. Green, M. R., Manikhas, G. M., Orlov, S., Afanasyev, B., Makhson, A. M., Bhar, P., Hawkins, M. J. (2006). Abraxane, a novel Cremophor-free, albumin-bound particle form

of paclitaxel for the treatment of advanced non-small-cell lung cancer. *Annals of Oncology 17*, 1263–1268.

18. Abraxane http://www.abraxane.com/mbc/hcp/download/Abraxane_Prescribing_Information .pdf Accessed December 12, 2012.
19. Oncaspar Available at http://www.oncaspar.com/providers-oncaspar.asp Accessed June 30, 2012.
20. Dipetrillo, T., Milas, L., Evans, D., Akerman, P., Ng, T., Miner, T., Cruff, D., Chauhan, B., Iannitti, D., Harrington, D., Safran, H. (2006). Paclitaxel poliglumex (PPX-Xyotax) and concurrent radiation for esophageal and gastric cancer: a phase I study. *American Journal of Clinical Oncology 29*, 376–379.
21. Singer, J. W. (2005). Paclitaxel poliglumex (XYOTAX, CT-2103): a macromolecular taxane. *Journal of Controlled Release 109*, 120–126.
22. Singer, J. W., Shaffer, S., Baker, B., Bernareggi, A., Stromatt, S., Nienstedt, D., Besman, M. (2005). Paclitaxel poliglumex (XYOTAX; CT-2103): an intracellularly targeted taxane. *Anticancer Drugs 16*, 243–254.
23. Opaxio http://www.celltherapeutics.com/opaxio Accessed December 12, 2012.
24. Erinotecan Available at http://www.nektar.com/product_pipeline/oncology_nktr-102. html Accessed June 30, 2012.
25. Ng, E. W., Shima, D. T., Calias, P., Cunningham, E. T., Jr., Guyer, D. R., Adamis, A. P. (2006). Pegaptanib, a targeted anti-VEGF aptamer for ocular vascular disease. *Nature Reviews Drug Discovery 5*, 123–132.
26. Molineux, G. (2004). The design and development of pegfilgrastim (PEG-rmetHuG-CSF, Neulasta). *Current Pharmaceutical Design 10*, 1235–1244.
27. Kim, D. W., Kim, S. Y., Kim, H. K., Kim, S. W., Shin, S. W., Kim, J. S., Park, K., Lee, M. Y., Heo, D. S. (2007). Multicenter phase II trial of Genexol-PM, a novel Cremophor-free, polymeric micelle formulation of paclitaxel, with cisplatin in patients with advanced non-small-cell lung cancer. *Annals of Oncology 18*, 2009–2014.
28. Lee, K. S., Chung, H. C., Im, S. A., Park, Y. H., Kim, C. S., Kim, S. B., Rha, S. Y., Lee, M. Y., Ro, J. (2008). Multicenter phase II trial of Genexol-PM, a Cremophor-free, polymeric micelle formulation of paclitaxel, in patients with metastatic breast cancer. *Breast Cancer Research and Treatment 108*, 241–250.
29. Junghanns, J. U., Muller, R. H. (2008). Nanocrystal technology, drug delivery and clinical applications. *International Journal of Nanomedicine 3*, 295–309.
30. Rupp, R., Rosenthal, S. L., Stanberry, L. R. (2007). VivaGel (SPL7013 Gel): a candidate dendrimer—microbicide for the prevention of HIV and HSV infection. *International Journal of Nanomedicine 2*, 561–566.
31. CALAA-01 (clinical-trial) Available at http://clinicaltrials.gov/ct2/show/ NCT00689065?term=calando&rank=1 Accessed June 30, 2012.
32. CALAA-01 http://www.arrowheadresearch.com/programs/calaa-01 Accessed December 12, 2012.
33. Qdot Nanocrystal Available at http://www.invitrogen.com/site/us/en/home/brands/ Molecular-Probes/Key-Molecular-Probes-Products/Qdot.html Accessed June 30, 2012.
34. Qdot 800 Available at http://products.invitrogen.com/ivgn/product/Q10171MP Accessed June 30, 2012.
35. Resovist Available at http://radiologie-uni-frankfurt.de/sites/radiologie-uni-frankfurt. de/content/e43/e2321/e2331/resofinal_eng.pdf Accessed June 30, 2012.

36. Feridex Available at http://www.amagpharma.com/products/feridex_iv.php Accessed June 30, 2012.
37. Wagner, V., Dullaart, A., Bock, A. K., Zweck, A. (2006). The emerging nanomedicine landscape. *Nature Biotechnology 24*, 1211–1217.
38. Davis, M. E., Chen, Z. G., Shin, D. M. (2008). Nanoparticle therapeutics: an emerging treatment modality for cancer. *Nature Reviews Drug Discovery 7*, 771–782.
39. Zamboni, W. C. (2008). Concept and clinical evaluation of carrier-mediated anticancer agents. *Oncologist 13*, 248–260.
40. Farokhzad, O. C., Langer, R. (2009). Impact of nanotechnology on drug delivery. *ACS Nano 3*, 16–20.
41. Li, S. D., Huang, L. (2008). Pharmacokinetics and biodistribution of nanoparticles. *Molecular Pharmaceutics 5*, 496–504.
42. Wong, H. L., Rauth, A. M., Bendayan, R., Manias, J. L., Ramaswamy, M., Liu, Z., Erhan, S. Z., Wu, X. Y. (2006). A new polymer-lipid hybrid nanoparticle system increases cytotoxicity of doxorubicin against multidrug-resistant human breast cancer cells. *Pharmaceutical Research 23*, 1574–1585.
43. Jain, R. K., Stylianopoulos, T. (2010). Delivering nanomedicine to solid tumors. *Nature Reviews Clinical Oncology 7*, 653–664.
44. Vladimir, T. (2009). Multifunctional and stimuli-sensitive pharmaceutical nanocarriers. *European Journal of Pharmaceutics and Biopharmaceutics 71*, 431–444.
45. Pericleous, P., Gazouli, M., Lyberopoulou, A., Rizos, S., Nikiteas, N., Efstathopoulos, E. P. (2012). Quantum dots hold promise for early cancer imaging and detection. *International Journal of Cancer 131*, 519–528.
46. Shi, D., Bedford, N. M., Cho, H.-S. (2011). Engineered multifunctional nanocarriers for cancer diagnosis and therapeutics. *Small 7*, 2549–2567.
47. Tkachenko, A. G., Xie, H., Coleman, D., Glomm, W., Ryan, J., Anderson, M. F., Franzen, S., Feldheim, D. L. (2003). Multifunctional gold nanoparticle−peptide complexes for nuclear targeting. *Journal of the American Chemical Society 125*, 4700–4701.
48. Allen, T. M., Cullis, P. R. (2004). Drug delivery systems: entering the mainstream. *Science 303*, 1818–1822.
49. Gabizon, A. A., Shmeeda, H., Zalipsky, S. (2006). Pros and cons of the liposome platform in cancer drug targeting. *Journal of Liposome Research 16*, 175–183.
50. Yuan, F., Dellian, M., Fukumura, D., Leunig, M., Berk, D. A., Torchilin, V. P., Jain, R. K. (1995). Vascular permeability in a human tumor xenograft: molecular size dependence and cutoff size. *Cancer Research 55*, 3752–3756.
51. Greish, K. (2007). Enhanced permeability and retention of macromolecular drugs in solid tumors: a royal gate for targeted anticancer nanomedicines. *Journal of Drug Targeting 15*, 457–464.
52. Maeda, H. (2001). The enhanced permeability and retention (EPR) effect in tumor vasculature: the key role of tumor-selective macromolecular drug targeting. *Advances in Enzyme Regulation 41*, 189–207.
53. Matsumura, Y., Maeda, H. (1986). A new concept for macromolecular therapeutics in cancer chemotherapy: mechanism of tumoritropic accumulation of proteins and the antitumor agent smancs. *Cancer Research 46*, 6387–6392.
54. Oku, N., Tokudome, Y., Asai, T., Tsukada, H. (2000). Evaluation of drug targeting strategies and liposomal trafficking. *Current Pharmaceutical Design 6*, 1669–1691.

55. Owens, D. E., 3rd, Peppas, N. A. (2006). Opsonization, biodistribution, and pharmacokinetics of polymeric nanoparticles. *International Journal of Pharmaceutics 307*, 93–102.
56. van Vlerken, L. E., Vyas, T. K., Amiji, M. M. (2007). Poly(ethylene glycol)-modified nanocarriers for tumor-targeted and intracellular delivery. *Pharmaceutical Research 24*, 1405–1414.
57. Allen, T. M., Chonn, A. (1987). Large unilamellar liposomes with low uptake into the reticuloendothelial system. *FEBS Letters 223*, 42–46.
58. Torchilin, V. P. (2007). Targeted pharmaceutical nanocarriers for cancer therapy and imaging. *AAPS Journal 9*, E128–E147.
59. Allen, T. M. (2002). Ligand-targeted therapeutics in anticancer therapy. *Nature Reviews Cancer 2*, 750–763.
60. Bae, Y. H., Park, K. (2011). Targeted drug delivery to tumors: myths, reality and possibility. *Journal of Controlled Release 153*, 198–205.
61. Moghimi, S. M., Hunter, A. C., Murray, J. C. (2001). Long-circulating and target-specific nanoparticles: theory to practice. *Pharmacology Reviews 53*, 283–318.
62. Yokoyama, M. (2005). Drug targeting with nano-sized carrier systems. *Journal of Artificial Organs 8*, 77–84.
63. Wang, M., Thanou, M. (2010). Targeting nanoparticles to cancer. *Pharmacological Research 62*, 90–99.
64. Gullotti, E., Yeo, Y. (2009). Extracellularly activated nanocarriers: a new paradigm of tumor targeted drug delivery. *Molecular Pharmaceutics 6*, 1041–1051.
65. Yu, B., Tai, H. C., Xue, W., Lee, L. J., Lee, R. J. (2010). Receptor-targeted nanocarriers for therapeutic delivery to cancer. *Molecular Membrane Biology 27*, 286–298.
66. Mann, A. P., Somasunderam, A., Nieves-Alicea, R., Li, X., Hu, A., Sood, A. K., Ferrari, M., Gorenstein, D. G., Tanaka, T. (2010). Identification of thioaptamer ligand against E-selectin: potential application for inflamed vasculature targeting. *PLoS One*, *5*. e13050.
67. Albanell, J., Baselga, J. (1999). Trastuzumab, a humanized anti-HER2 monoclonal antibody, for the treatment of breast cancer. *Drugs Today (Barc) 35*, 931–946.
68. Hrkach, J., Von Hoff, D., Ali, M. M., Andrianova, E., Auer, J., Campbell, T., De Witt, D., Figa, M., Figueiredo, M., Horhota, A., Low, S., McDonnell, K., Peeke, E., Retnarajan, B., Sabnis, A., Schnipper, E., Song, J. J., Song, Y. H., Summa, J., Tompsett, D., Troiano, G., Van Geen Hoven, T., Wright, J., LoRusso, P., Kantoff, P. W., Bander, N. H., Sweeney, C., Farokhzad, O. C., Langer, R., Zale, S. (2012). Preclinical development and clinical translation of a PSMA-targeted docetaxel nanoparticle with a differentiated pharmacological profile. *Science Translational Medicine 4*, 128ra39.
69. Lee, H., Fonge, H., Hoang, B., Reilly, R. M., Allen, C. (2010). The effects of particle size and molecular targeting on the intratumoral and subcellular distribution of polymeric nanoparticles. *Molecular Pharmaceutics 7*, 1195–1208.
70. Park, J. W., Hong, K., Kirpotin, D. B., Colbern, G., Shalaby, R., Baselga, J., Shao, Y., Nielsen, U. B., Marks, J. D., Moore, D., Papahadjopoulos, D., Benz, C. C. (2002). Anti-HER2 immunoliposomes: enhanced efficacy attributable to targeted delivery. *Clinical Cancer Research 8*, 1172–1181.
71. Hong, S., Leroueil, P. R., Majoros, I. J., Orr, B. G., Baker, J. R., Jr., Banaszak Holl, M. M. (2007). The binding avidity of a nanoparticle-based multivalent targeted drug delivery platform. *Chemical Biology 14*, 107–115.

72. Alberts, B., Wilson, J. H., Hunt, T., *Molecular Biology of the Cell*, Garland Science, New York, 2008.
73. Amoozgar, Z., Yeo, Y. (2012). Recent advances in stealth coating of nanoparticle drug delivery systems. *Wiley Interdisciplinary Reviews: Nanomedicine and Nanobiotechnology 4*, 219–233.
74. Gerweck, L. E., Seetharaman, K. (1996). Cellular pH gradient in tumor versus normal tissue: potential exploitation for the treatment of cancer. *Cancer Research 56*, 1194–1198.
75. Vander Heiden, M. G., Cantley, L. C., Thompson, C. B. (2009). Understanding the Warburg effect: the metabolic requirements of cell proliferation. *Science 324*, 1029–1033.
76. Raghunand, N., Gatenby, R. A., Gillies, R. J. (2003). Microenvironmental and cellular consequences of altered blood flow in tumours. *British Journal of Radiology 76*, S11–S22.
77. Lee, E. S., Gao, Z., Bae, Y. H. (2008). Recent progress in tumor pH targeting nanotechnology. *Journal of Controlled Release 132*, 164–170.
78. Breunig, M., Bauer, S., Goepferich, A. (2008). Polymers and nanoparticles: intelligent tools for intracellular targeting? *European Journal of Pharmaceutics and Biopharmaceutics 68*, 112–128.
79. Szeto, H. (2006). Cell-permeable, mitochondrial-targeted, peptide antioxidants. *The AAPS Journal 8*, E277–E283.
80. Chau, Y., Dang, N. M., Tan, F. E., Langer, R. (2006). Investigation of targeting mechanism of new dextran-peptide-methotrexate conjugates using biodistribution study in matrix-metalloproteinase-overexpressing tumor xenograft model. *Journal of Pharmaceutical Sciences 95*, 542–551.
81. Hatakeyama, H., Akita, H., Kogure, K., Harashima, H. (2007). Development of a novel systemic gene delivery system for cancer therapy with a tumor-specific cleavable PEG-lipid. *Yakugaku Zasshi-Journal of the Pharmaceutical Society of Japan 127*, 1549–1556.
82. Egeblad, M., Werb, Z. (2002). New functions for the matrix metalloproteinases in cancer progression. *Nature Reviews Cancer 2*, 161–174.
83. Terada, T., Iwai, M., Kawakami, S., Yamashita, F., Hashida, M. (2006). Novel PEG-matrix metalloproteinase-2 cleavable peptide-lipid containing galactosylated liposomes for hepatocellular carcinoma-selective targeting. *Journal of Controlled Release 111*, 333–342.
84. Hatakeyama, H., Akita, H., Ito, E., Hayashi, Y., Oishi, M., Nagasaki, Y., Danev, R., Nagayama, K., Kaji, N., Kikuchi, H., Baba, Y., Harashima, H. (2011). Systemic delivery of siRNA to tumors using a lipid nanoparticle containing a tumor-specific cleavable PEG-lipid. *Biomaterials 32*, 4306–4316.
85. Niidome, T., Ohga, A., Akiyama, Y., Watanabe, K., Niidome, Y., Mori, T., Katayama, Y. (2010). Controlled release of PEG chain from gold nanorods: targeted delivery to tumor. *Bioorganic & Medicinal Chemistry 18*, 4453–4458.
86. Ryu, J. H., Kim, S. A., Koo, H., Yhee, J. Y., Lee, A., Na, J. H., Youn, I., Choi, K., Kwon, I. C., Kim, B.-S., Kim, K. (2011). Cathepsin B-sensitive nanoprobe for *in vivo* tumor diagnosis. *Journal of Materials Chemistry 21*, 17631–17634.

87. Lopez-Otin, C., Matrisian, L. M. (2007). Emerging roles of proteases in tumour suppression. *Nature Reviews Cancer 7*, 800–808.
88. Ruenraroengsak, P., Cook, J. M., Florence, A. T. (2010). Nanosystem drug targeting: facing up to complex realities. *Journal of Controlled Release 141*, 265–276.
89. Gaumet, M., Vargas, A., Gurny, R., Delie, F. (2008). Nanoparticles for drug delivery: the need for precision in reporting particle size parameters. *European Journal of Pharmaceutics and Biopharmaceutics 69*, 1–9.
90. Florence, A. T. (2007). Pharmaceutical nanotechnology: more than size. Ten topics for research. *International Journal of Pharmaceutics 339*, 1–2.
91. Bae, Y. H. (2009). Drug targeting and tumor heterogeneity. *Journal of Controlled Release 133*, 2–3.
92. Bawa, R. (2007). Patents and nanomedicine. *Nanomedicine 2*, 351–374.
93. Du, H., Chandaroy, P., Hui, S. W. (1997). Grafted poly-(ethylene glycol) on lipid surfaces inhibits protein adsorption and cell adhesion. *Biochimica et Biophysica Acta (BBA)—Biomembranes 1326*, 236–248.
94. Ishihara, T., Maeda, T., Sakamoto, H., Takasaki, N., Shigyo, M., Ishida, T., Kiwada, H., Mizushima, Y., Mizushima, T. (2010). Evasion of the accelerated blood clearance phenomenon by coating of nanoparticles with various hydrophilic polymers. *Biomacromolecules 11*, 2700–2706.
95. Ishihara, T., Takeda, M., Sakamoto, H., Kimoto, A., Kobayashi, C., Takasaki, N., Yuki, K., Tanaka, K.-i., Takenaga, M., Igarashi, R., Maeda, T., Yamakawa, N., Okamoto, Y., Otsuka, M., Ishida, T., Kiwada, H., Mizushima, Y., Mizushima, T. (2009). Accelerated blood clearance phenomenon upon repeated injection of PEG-modified PLA-nanoparticles. *Pharmaceutical Research 26*, 2270–2279.
96. Chen, H., Kim, S., Li, L., Wang, S., Park, K., Cheng, J. X. (2008). Release of hydrophobic molecules from polymer micelles into cell membranes revealed by Forster resonance energy transfer imaging. *Proceedings of National Academy of Sciences of the United States of America 105*, 6596–6601.
97. Chen, H., Kim, S., He, W., Wang, H., Low, P. S., Park, K., Cheng, J. X. (2008). Fast release of lipophilic agents from circulating PEG-PDLLA micelles revealed by *in vivo* forster resonance energy transfer imaging. *Langmuir 24*, 5213–5217.
98. Gullotti, E., Yeo, Y. (2012). Beyond the imaging: limitations of cellular uptake study in the evaluation of nanoparticles. *Journal of Controlled Release 164*, 172–176.
99. de Smet, M., Heijman, E., Langereis, S., Hijnen, N. M., Grüll, H. (2011). Magnetic resonance imaging of high intensity focused ultrasound mediated drug delivery from temperature-sensitive liposomes: an *in vivo* proof-of-concept study. *Journal of Controlled Release 150*, 102–110.
100. Dvorak, H. F., Weaver, V. M., Tlsty, T. D., Bergers, G. (2011). Tumor microenvironment and progression. *Journal of Surgical Oncology 103*, 468–474.
101. Sitohy, B., Nagy, J. A., Dvorak, H. F. (2012). Anti-VEGF/VEGFR therapy for cancer: reassessing the target. *Cancer Research 72*, 1909–1914.
102. Maeda, H. (2010). Tumor-selective delivery of macromolecular drugs via the EPR effect: background and future prospects. *Bioconjugate Chemistry 21*, 797–802.
103. Harrington, K. J., Mohammadtaghi, S., Uster, P. S., Glass, D., Peters, A. M., Vile, R. G., Stewart, J. S. W. (2001). Effective targeting of solid tumors in patients with locally

advanced cancers by radiolabeled pegylated liposomes. *Clinical Cancer Research 7*, 243–254.

104. Jain, R. K. (1994). Barriers to drug delivery in solid tumors. *Scientific American 271*, 58–65.
105. Holback, H., Yeo, Y. (2011). Intratumoral drug delivery with nanoparticulate carriers. *Pharmaceutical Research 28*, 1819–1830.
106. Heldin, C. H., Rubin, K., Pietras, K., Ostman, A. (2004). High interstitial fluid pressure—an obstacle in cancer therapy. *Nature Reviews Cancer 4*, 806–813.
107. Minchinton, A. I., Tannock, I. F. (2006). Drug penetration in solid tumours. *Nature Reviews Cancer 6*, 583–592.
108. Jang, S. H., Wientjes, M. G., Lu, D., Au, J. L. S. (2003). Drug delivery and transport to solid tumors. *Pharmaceutical Research 20*, 1337–1350.
109. Goodman, T. T., Chen, J. Y., Matveev, K., Pun, S. H. (2008). Spatio-temporal modeling of nanoparticle delivery to multicellular tumor spheroids. *Biotechnology and Bioengineering 101*, 388–399.
110. Cabral, H., Matsumoto, Y., Mizuno, K., Chen, Q., Murakami, M., Kimura, M., Terada, Y., Kano, M. R., Miyazono, K., Uesaka, M., Nishiyama, N., Kataoka, K. (2011). Accumulation of sub-100 nm polymeric micelles in poorly permeable tumours depends on size. *Nature Nanotechnology 6*, 815–823.
111. Juweid, M., Neumann, R., Paik, C., Perez-Bacete, M. J., Sato, J., van Osdol, W., Weinstein, J. N. (1992). Micropharmacology of monoclonal antibodies in solid tumors: direct experimental evidence for a binding site barrier. *Cancer Research 52*, 5144–5153.
112. Saga, T., Neumann, R. D., Heya, T., Sato, J., Kinuya, S., Le, N., Paik, C. H., Weinstein, J. N. (1995). Targeting cancer micrometastases with monoclonal antibodies: a binding-site barrier. *Proceedings of National Academy of Sciences of the United States of America 92*, 8999–9003.
113. Prescott, C. (2010). Regenerative nanomedicines: an emerging investment prospective? *Journal of The Royal Society Interface 7*, S783–S787.
114. US Food and Drug Administration. 2007. Nanotechnology Task Force Report 2007.
115. Nagamitsu, A., Greish, K., Maeda, H. (2009). Elevating blood pressure as a strategy to increase tumor-targeted delivery of macromolecular drug SMANCS: cases of advanced solid tumors. *Japanese Journal of Clinical Oncology 39*, 756–766.
116. Seki, T., Fang, J., Maeda, H. (2009). Enhanced delivery of macromolecular antitumor drugs to tumors by nitroglycerin application. *Cancer Science 100*, 2426–2430.
117. Eikenes, L., Tari, M., Tufto, I., Bruland, O. S., Davies, C. D. (2005). Hyaluronidase induces a transcapillary pressure gradient and improves the distribution and uptake of liposomal doxorubicin (Caelyx™) in human osteosarcoma xenografts. *British Journal of Cancer 93*, 81–88.
118. Kamb, A. (2005). What's wrong with our cancer models? *Nature Reviews Drug Discovery 4*, 161–165.
119. Hearnden, V., MacNeil, S., Battaglia, G. (2011). Tracking nanoparticles in three-dimensional tissue-engineered models using confocal laser scanning microscopy. *Methods in Molecular Biology 695*, 41–51.
120. Hosoya, H., Kadowaki, K., Matsusaki, M., Cabral, H., Nishihara, H., Ijichi, H., Koike, K., Kataoka, K., Miyazono, K., Akashi, M., Kano, M. R. (2012). Engineering fibrotic

tissue in pancreatic cancer: a novel three-dimensional model to investigate nanoparticle delivery. *Biochemical and Biophysical Research Communications 419*, 32–37.

121. Mitra, M., Mohanty, C., Harilal, A., Maheswari, U. K., Sahoo, S. K., Krishnakumar, S. (2012). A novel *in vitro* three-dimensional retinoblastoma model for evaluating chemotherapeutic drugs. *Molecular Vision 18*, 1361–1378.
122. Huang, S. K., Mayhew, E., Gilani, S., Lasic, D. D., Martin, F. J., Papahadjopoulos, D. (1992). Pharmacokinetics and therapeutics of sterically stabilized liposomes in mice bearing C-26 colon carcinoma. *Cancer Research 52*, 6774–6781.
123. O'Brien, M. E. R., Wigler, N., Inbar, M., Rosso, R., Grischke, E., Santoro, A., Catane, R., Kieback, D. G., Tomczak, P., Ackland, S. P., Orlandi, F., Mellars, L., Alland, L., Tendler, C., Grp, C. B. C. S. (2004). Reduced cardiotoxicity and comparable efficacy in a phase III trial of pegylated liposomal doxorubicin HCl (CAELYX™/Doxil®) versus conventional doxorubicin for first-line treatment of metastatic breast cancer. *Annals of Oncology 15*, 440–449.
124. Li, C., Yu, D.-F., Newman, R. A., Cabral, F., Stephens, L. C., Hunter, N., Milas, L., Wallace, S. (1998). Complete regression of well-established tumors using a novel water-soluble poly(L-glutamic acid)-paclitaxel conjugate. *Cancer Research 58*, 2404–2409.
125. European Medicines Agency. 2009. Questions and answers on the withdrawal of the marketing authorisation application for Opaxio (Paclitaxel poliglumex).
126. Damia, G., D'Incalci, M. (2009). Contemporary pre-clinical development of anticancer agents—what are the optimal preclinical models? *European Journal of Cancer 45*, 2768–2781.
127. Vinogradov, S., Wei, X. (2012). Cancer stem cells and drug resistance: the potential of nanomedicine. *Nanomedicine 7*, 597–615.
128. Hawkins, M. J., Soon-Shiong, P., Desai, N. (2008). Protein nanoparticles as drug carriers in clinical medicine. *Advanced Drug Delivery Reviews 60*, 876–885.
129. David, S., Pitard, B., Benoît, J.-P., Passirani, C. (2010). Non-viral nanosystems for systemic siRNA delivery. *Pharmacological Research 62*, 100–114.
130. Davis, M. E., Zuckerman, J. E., Choi, C. H. J., Seligson, D., Tolcher, A., Alabi, C. A., Yen, Y., Heidel, J. D., Ribas, A. (2010). Evidence of RNAi in humans from systemically administered siRNA via targeted nanoparticles. *Nature 464*, 1067–1070.

# APPLICATIONS OF LIGAND-ENGINEERED NANOMEDICINES

Gayong Shim, Joo Yeon Park, Lee Dong Roh, and Yu-Kyoung Oh
*College of Pharmacy, Seoul National University, Daehak-dong, Seoul, South Korea*
Sangbin Lee
*School of Life Sciences and Biotechnology, Korea University, Seoul, South Korea*

## 2.1 INTRODUCTION

Over the past two decades, a variety of therapeutic modalities have been investigated for the treatment of cancer. Among the most promising new modalities in the field of medical applications is nanomedicine, which has proven to be an outstanding approach for improving therapeutic index [1,2]. Nanomedicines can achieve remarkable improvements in anticancer efficacy not only by increasing bioavailability and half-life, but also by promoting tumor-accumulation of therapeutic entities through the enhanced permeability and retention (EPR) effect [3]. Moreover, through modifications of the surface of nanoparticles, nanomedicines have showed the ability to improve target-specificity and further enhance the efficacy [4].

### 2.1.1 Nanoparticulate Drug Delivery

Nanoparticles are a proven modality for altering the *in vivo* fates of various therapeutic entities, such as small molecules, nucleic acids, and peptides [5]. Nanoparticulate drugs can achieve prolonged retention *in vivo*, increased accumulation in tumor tissues, and enhanced pharmacodynamics [1]. Depending on the materials used to construct them, nanoparticles are mainly classified as lipid- or

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*,
Edited by Shayne C. Gad
Copyright © 2013 John Wiley & Sons, Inc.

polymer-based. The series of nanomedicines on the market and in clinical trials substantiates the potential of nanoparticles as drug-delivery platforms [5,6]. Among nanomedicines, lipid-based nanomedicines, particularly liposomes, have a long history of commercialization. Following *in vivo* administration, the surface of liposomes easily interacts with serum proteins, promoting their elimination by the reticuloendothelial system. More prolonged circulation in the blood has been achieved by introducing poly(ethylene glycol) (PEG) moieties on liposomal surfaces, thereby reducing the interaction of liposomes with blood components. Such surface coating with PEG has been applied not only to liposomes but also to other nanoparticles. The current status of nanomedicines on the market is summarized in Table 2.1.

**TABLE 2.1 Current Status of Nanomedicine on the Market**

| Product Name | Active Drug | Company | Indications |
|---|---|---|---|
| *Lipid-based* | | | |
| Daunoxome | Daunorubicin | Nexstar | Kaposi's sarcoma |
| Myocet | Doxorubicin | Zeneus | Combinational therapy of recurrent breast cancer |
| Doxil/Caelyx | Doxorubicin | Tibotec | Refractory Kaposi's sarcoma |
| | Doxorubicin | ALZA | Ovarian cancer/recurrent breast cancer |
| DepoCyt | Cytarabine | Enzon | Lymphomatous meningitis |
| Onco TCS | Vincristine | Enzon/Inex | Non-Hodgkin's lymphoma |
| Ambisome | Amphotericin | Nexstar | Fungal infection in immunocompromised patients |
| Visudyne | Verteforfin | QLT/Norvatis | Wet age-related macular degeneration |
| Depo Dur | Morphine sulfate | Pacira Pharm | Postsurgical pain |
| *Polymer-based* | | | |
| Neulasta | PEG–GCSF | Amgen | Neutropenia associated with cancer chemotherapy |
| Pegasis | PEG–IFN2a | Genentech | Melanoma/Chronic myeloid leukemia/renal-cell carcinoma |
| Oncaspar | PEG–L-asparaginase | Enzon | Acute lymphoblastic leukemia |
| Zinostatin stimalamer | SMANCS | Yamanouchi | Hepatocellular carcinoma |
| Genexol-PM | Paclitaxel | Samyang | Metastatic breast cancer |
| *Others* | | | |
| Abaxane | Albumin-bound paclitaxel | Abraxis bioScience/ AstraZeneca | Metastatic breast cancer |

GCSF, granulocyte colony-stimulating factor; IFNα, interferon-α; SMANCS, styrene maleic anhydride neocarzinostatin.

### 2.1.2 Engineered Nanoparticulate Drug Delivery

Although nanoparticulate drugs have been equipped with attractive features such as improvements in pharmacokinetic and pharmacodynamic profiles, they still suffer from the lack of an active driving force for tissue or cell-specific delivery, relying instead on the EPR effect. Under *in vivo* conditions, nanoparticles confront formidable biological barriers that impede their delivery to target tissues, including the reticuloendothelial system, vascular endothelium, and blood enzymes [7]. Coating nanoparticle surfaces with PEG increases their stability in the bloodstream, but it does not enhance their interactions with specific types of tissues or cells. To the contrary, there have been reports that higher PEG content in nanomedicines can even reduce interactions of nanoparticles with cells [8]. Numerous approaches have been developed to overcome such limitations through modification of nanoparticles with various ligands, including antibodies, peptides and proteins, sugars and polysaccharides, and chemicals.

## 2.2 LIGAND-MEDIATED DELIVERY

### 2.2.1 Antibodies

Because of their specific antigen-recognition properties, antibodies have long been studied as a way of increasing the specificity of nanomedicine delivery [9,10]. The types of antibody variants considered for this purpose include single-chain variable fragments (scFv), $F(ab')_2$ fragments, and whole IgG antibodies. In cancer-related applications, targeting is typically achieved by tagging nanoparticles with antibodies against proteins known to be overexpressed on the surface of tumor cells relative to normal cells. Commonly studied cancer target proteins include HER2, selectins, and the epidermal growth factor receptor (EGFR). The primary methods for tagging antibodies to nanoparticles are covalent chemical conjugation and specific biomimetic interaction methods. *In vivo* applications of antibody-tagged nanomedicines are summarized in Table 2.2.

#### *2.2.1.1* In Vivo *Applications of Antibody-Tagged Nanomedicines*

*Nanoparticles Targeted to Tumors.* Taking advantage of the fact that HER2 is overexpressed on several cancer cells, researchers have tagged liposomes entrapping vincristine [11] or doxorubicin [12] with HER2-specific antibodies. Liposomes entrapping the anticancer drug vincristine were tagged with anti-HER2 antibodies by coupling to maleimide-modified PEG lipids [11]. Modification of vincristine liposomes with anti-HER2 antibody was shown to provide greater efficacy than free vincristine or unmodified liposomal vincristine against the HER2-overexpressing breast cancer cell lines, BT474-M2 and SKBR3. For *in vivo* applications, anti-HER2 antibody-modified immunoliposomes, unmodified liposomes, and free vincristine were intravenously administered to BT474-M2-xenografted mice at a dose of 1 mg/kg once a week for 3 weeks. Compared to other groups, immunoliposomes

**TABLE 2.2** ***In Vivo*** **Applications of Antibody-Tagged Nanomedicines**

| Target | Disease or Target Tissue | Nanoparticles | Active Substances | *In Vivo* Model | References |
|---|---|---|---|---|---|
| HER2 | Breast cancer | Liposome | Vincristine | BT474-M2-bearing mice | 11 |
| HER2 | Breast cancer | Liposome | Doxorubicin | MCF-7/HER2-bearing mice | 12 |
| C-met | Lung cancer | Liposome | Doxorubicin | H1993-bearing mice | 13 |
| BRCAA1 | Gastric cancer | $Fe_3O_4$ nanoparticles | Fluorescent magnetic nanoprobes | MGC803-bearing mice | 14 |
| EGFR | Glioblastoma multiforme | Liposome | Sodium borocaptate | EGFR-overexpressing U87-bearing mice | 16 |
| EGFR | Glioblastoma multiforme | Liposome | Fluorescent dye (8-hydroxypyrene–1,3,6-trisulfonic acid trisodium salt) | EGFR-overexpressing U87-bearing mice | 25 |
| EGFR | Hepatocellular carcinoma | Poly(lactic acid-*co*-L-lysine) nanoparticle | Fluorescent dye (rhodamine B) | SMMC7721-bearing mice | 17 |
| E-selectin | MCa-4 mammary tumor | Liposome | Combretastatin | MCa-4-bearning mice | 18 |
| THY 1.1 | Mesangial cells in kidney | Liposome | Doxorubicin | Wistar rats | 19 |
| P-selectin | Myocardial infarction | Liposome | VEGF | Myocardial infarction-induced rats | 20 |
| LFA-1 | Human immunodeficiency virus-infected cells | Liposome | Anti-CCR5 siRNA | Humanized mice | 21 |

produced a greater reduction in tumor volumes and were well tolerated without causing substantial weight loss.

In addition to anti-HER2 antibody, quantum-dot (Q-dot) nanoparticles have been further conjugated to doxorubicin-loaded liposomes to create multifunctional imaging and therapeutic nanomedicines [12]. In this study, anti-HER2/ErbB2 scFv fragments were conjugated to one end of liposomal surface PEG, and carboxyl Q-dots were chemically tethered to the other end of liposomal surface PEG. *In vivo* imaging revealed a greater accumulation of anti-HER2 scFv- and Q-dot-conjugated doxorubicin immunoliposomes than Q-dot-conjugated liposomes in tumor tissues following intravenous administration in MCF-7/HER2-xenografted mice.

Anti-cMet scFv has been used to modify doxorubicin-loaded liposomes and Q-dot nanoparticles [13]. Doxorubicin-loaded liposomes were modified with anti-cMet scFv by coupling to maleimide-modified PEG chains on the external surface of liposomes. Anti-cMet scFv was similarly linked to maleimide-activated Q-dots. In H460-derived lung cancer xenografted mice, intravenous administration of 2 mg/kg of doxorubicin in an anti-cMet scFv-modified liposome formulation was shown to be almost twice as effective in reducing tumor volume as unmodified liposomes. Molecular imaging in mice revealed a greater distribution of anti-cMet scFv-modified Q-dots in tumors than unmodified Q-dots. However, a substantial distribution of Q-dots to tissues other than tumors was noted in the upper tail region, regardless of scFv modification. Such a distribution of Q-dots to the normal tail region may limit the targeted imaging of tumor tissues.

In addition to Q-dots, fluorescent magnetic nanoparticles have been conjugated with anti-BRCAA1 antibody to allow *in vivo* fluorescence and magnetic resonance imaging of gastric cancers [14]. In these applications, fluorescent magnetic nanoparticles were first functionalized with a carboxyl group and then conjugated with anti-BRCAA1 antibody using 1-ethyl-3-(3-dimethylaminopropyl)-carbodiimide and *N*-hydroxylsuccinimide as coupling agents. Treatment of the MGC803 gastric cancer cell line with anti-BRCAA1 antibody-tethered fluorescent magnetic nanoparticles provided a twofold increase in cellular-uptake efficiency compared to plain fluorescent magnetic nanoparticles. *In vivo* fluorescence imaging showed that the tumor tissue distribution of anti-BRCAA1 antibody-coupled nanoparticles in MGC803-xenografted mice peaked 6 h after intravenous administration. Magnetic resonance imaging obtained within 12 h after administration revealed a higher tumor distribution of anti-BRCAA1 antibody-coated magnetic nanoparticles compared to plain nanoparticles.

Antibodies against EGFR, which is overexpressed in malignant glioblastoma but weakly expressed or undetectable in normal brain [15], have been used for ligand modification of liposomes and polymeric nanoparticles. In particular, sodium borocaptate-loaded liposomes tagged with anti-EGFR antibody have been used for boron neutron capture therapy [16]. After intravenous administration of immunoliposomes to U87-EGFR cell-xenografted mice, the distribution of sodium borocaptate-loaded liposomes to the tumor tissues was tested by quantitation of boron compounds. Compared to unmodified liposomes, anti-EGFR antibody-tagged

liposomes delivered higher amounts of boron compounds to tumor tissues, but not to the normal brain, liver, or blood.

Poly(lactic acid-*co*-L-lysine) nanoparticles have been modified with anti-EGFR-antibodies [17]. For chemical tethering, these antibodies were linked to the surface of poly(lactic acid-*co*-L-lysine) nanoparticles using the coupling agents, 1-ethyl-3-(3-dimethylaminopropyl)-carbodiimide and *N*-hydroxylsuccinimide, and rhodamine B was loaded into the nanoparticles as a tracer of cellular uptake. In the SMMC-7721 hepatocellular carcinoma cell line, the internalization of anti-EGFR-antibody-conjugated poly(lactic acid-*co*-L-lysine) nanoparticles was decreased in the presence of excess amounts of free anti-EGFR antibodies, indicating that nanoparticles were endocytosed via EGFR receptors. In SMMC-7721-bearing mice, polymeric nanoparticles with or without anti-EGFR antibody were distributed to tumor tissues 4 h after intravenous administration, with anti-EGFR antibody-conjugated nanoparticles showing more prolonged retention within tumor tissues during the 48 h after dosing.

Modification of PEGylated liposomes with an anti-E-selectin antibody has been studied as a means to improve distribution to mammary tumor tissues [18]. In this application, an anti-E-selectin antibody was coupled to the distal end of maleimide (2 mol%)-functionalized PEG formulated in liposomes entrapping the antivascular drug, combretastatin A4 disodium phosphate (CA4DP). At equivalent doses (15 mg/kg), anti-E-selectin antibody-tagged liposomes incorporating CA4DP showed greater anticancer efficacy in mice bearing MCa-4 mammary tumor cell xenografts than plain liposomes incorporating CA4DP and was more efficacious than a higher dose (81 mg/kg) of free CA4DP. Moreover, when used in combination therapy, CA4DP chemotherapy using E-selectin antibody-tagged liposomes synergized with radiation therapy to produce greater anticancer effects than chemotherapy or radiation therapy alone.

*Nanoparticles Targeted to Other Diseases.* Although many antibody-modified nanomedicines have been studied for targeting tumor tissues, recent studies have reported applications of immunoliposomes for kidney diseases [19], vascular diseases [20], and antiviral therapy [21]. The therapeutic entities carried by these immunoliposomes include doxorubicin, cytokine proteins, and small interfering RNA (siRNA).

The $F(ab')_2$ fragment of an anti-OX7 antibody, which has an affinity for Thy 1.1 antigen on mesangial cells in the kidney, has been used to enhance the delivery of doxorubicin-loaded liposomes to rat kidney [19]. In this application, anti-OX7 $F(ab')_2$ was conjugated to the surfaces of liposomes via maleimide functionalized PEG lipids. Intravenously administered anti-OX7 $F(ab')_2$-conjugated liposomes were shown to specifically enhance delivery to mesangial cells in both kidneys. In contrast, free OX7 $F(ab)_2$ fragments coadministered with liposomes did not enhance renal delivery. Treatment of rats with low-dose doxorubicin (0.24 mg/kg) resulted in extensive glomerular damage in the case of immunoliposomes, whereas such treatment with unmodified liposomes or free drug did not. This study indicates the utility of immunoliposomes for treatment of renal diseases.

In myocardial infarction, ischemic myocardium sites are reported to exhibit upregulation of cell-adhesion molecules, such as P-selectin. On the basis of this pathology, an anti-P-selectin antibody was designed to target altered vascular endothelial surfaces [20]. In this study, anti-P-selectin antibody was conjugated to the maleimide group on the 1,2-distearoyl-*sn*-glycero-3-phosphoethanolamine-*N*-[(polyethylene glycol)2000] (DSPE–PEG2000) component of the liposome. Next, vascular endothelial growth factor was loaded into anti-P-selectin-conjugated liposomes by rehydrating freeze-dried immunoliposomes with vascular endothelial growth factor (VEGF) solutions. In the rat myocardial infarction model, treatment with VEGF-loaded immunoliposomes was reported to significantly improve cardiac function and the number of perfused vessels compared to controls [20].

For siRNA-based antiviral therapy, liposomes have been modified with an antibody against lymphocyte function-associated antigen-1 (LFA-1) [21]. In this approach, rather than using PEG as a stabilizing moiety, hyaluronan 751 kDa was used as a liposome-surface modifier to increase stability in the bloodstream. Hyaluronan-modified liposomes were then conjugated with anti-LFA-1 antibody using an amine-coupling method. CCR5-specific siRNA was entrapped by rehydrating lyophilized liposomes with CCR5-specific siRNA and a protamine complex solution. The resulting immunoliposomal siRNA was injected into humanized mice via the tail vein at a dose of 50 μg/mouse. Subsequent challenge of humanized mice with human immunodeficiency virus was shown to provide enhanced resistance to infection by reducing viral load.

#### *2.2.1.2 Methods for Tagging Nanomedicines with Antibodies*

*Covalent Conjugation.* Because antibodies must retain their three-dimensional conformation to specifically interact with antigen epitopes, methods used to tag nanomedicines with antibodies should be carefully controlled. Although new methods of antibody tagging have been developed, the most widely used method has been the chemical conjugation method, where the antibody is linked to activated groups on the surfaces of nanoparticles via covalent bonding [22].

Maleimide groups have been used to provide activated ends to nanoparticles, allowing conjugation to sulfhydryl groups in cysteine residues of the antibody. This method has been used to modify drug-entrapped liposomes with various antibodies, including the anti-cMet scFv [13], and anti-OX7 $F(ab')_2$ fragments [19], and anti-HER2 [11], anti-E-selectin [18], and anti-P-selectin [20] antibodies, described earlier.

Alternatively, the surface of nanoparticles can be chemically modified using 1-ethyl-3-(3-dimethylaminopropyl)-carbodiimide and *N*-hydroxylsuccinimide as coupling agents. Examples of this include the modification of poly(lactic acid-*co*-L-lysine) nanoparticles with anti-EGFR-antibody [17] and modification of fluorescent magnetic nanoparticles with anti-BRCAA1 antibody [14].

Chemical coupling methods, however, suffer from several potential limitations. First, chemical coupling may affect the three-dimensional conformation of antibodies. Second, antibodies bound by nonspecific coupling are randomly oriented.

Both altered conformation and random orientation may result in the loss of antibody activity. Moreover, tagging antibodies to nanoparticles through complex reaction procedures requires time and effort.

*Noncovalent Binding.* To overcome the limitations inherent in chemical coupling methods, researchers have developed noncovalent biomimetic linking methods based on specific interactions. The prototypical example of this is based on protein A, a protein produced by *Staphylococcus aureus* that binds to the Fc domain of antibodies through its ZZ motif. Protein A has been exploited to prevent opsonization in the bloodstream [23] and has been applied to biotechnology products for the separation of antibodies from protein mixtures and the design of protein chips [24]. In the current context, the antibody-affinity ZZ motif of protein A has been used to tag antibodies to the surface of liposomes without chemical coupling. In a previous study by Feng and colleagues, a ZZ-His peptide was first bound to the nickel component of Ni lipid-containing liposomes through Ni–His interactions [16]. The resulting liposomal surface ZZ was used to tag an anti-EGFR antibody by simply mixing at room temperature. This approach, however, is limited by the potential cytotoxicity of the liposomal Ni used for His binding. Moreover, directly tagging an antibody to the surface of liposomes with PEG lipid chains may reduce the probability of antibody interaction because of steric hindrance of neighboring PEG chains.

In a more recent study, Feng and colleagues sought to improve the ability of antibodies to recognize and bind to target cells by designing liposomes that positioned the antibody further out from the liposome surface than did the Ni liposomes used previously. To accomplish this, they first modified maleimide-functionalized PEG chains of liposomes with a thiolated form of Gaussia luciferase-ZZ-His [25]. Gaussia luciferase is a 19.9 kDa bioluminescent reporter protein that provides an imaging functionality. Anti-EGFR antibodies were then bound to the ZZ domain at the distal end of PEG chains on the liposomal surface. The presence of Gaussia luciferase enabled monitoring of the *in vivo* fate of immunoliposomes by bioluminescence imaging.

In addition to the ZZ motif, a cyclic 13-mer peptide has been reported to be a crucial sequence for the Fc-binding function of protein A [26]. Taking advantage of the Fc-binding 13-mer peptide (FcBP), Oh and colleagues linked the peptide to functionalized PEG lipid derivatives [27], and then used the FcBP-linked PEG lipids to formulate liposomes to which antibodies could be bound via the specific affinity of FcBP. Subsequent attachment of anti-CXCR4 antibody to FcBP-modified liposomes resulted in enhanced cellular uptake by a CXCR4-positive Ramos cell line compared to plain liposomes.

Compared with the conventional chemical conjugation methods (Figure 2.1a), protein A-based affinity methods for noncovalently tagging antibodies (Figure 2.1b) may have advantages in retaining the three-dimensional conformation of liposomes and controlling the orientation of antibodies. However, this Fc moiety affinity-based approach is also limited in that it cannot be used to coat liposomes with scFv or $F(ab')_2$ fragment type antibodies, which lack in the Fc portion.

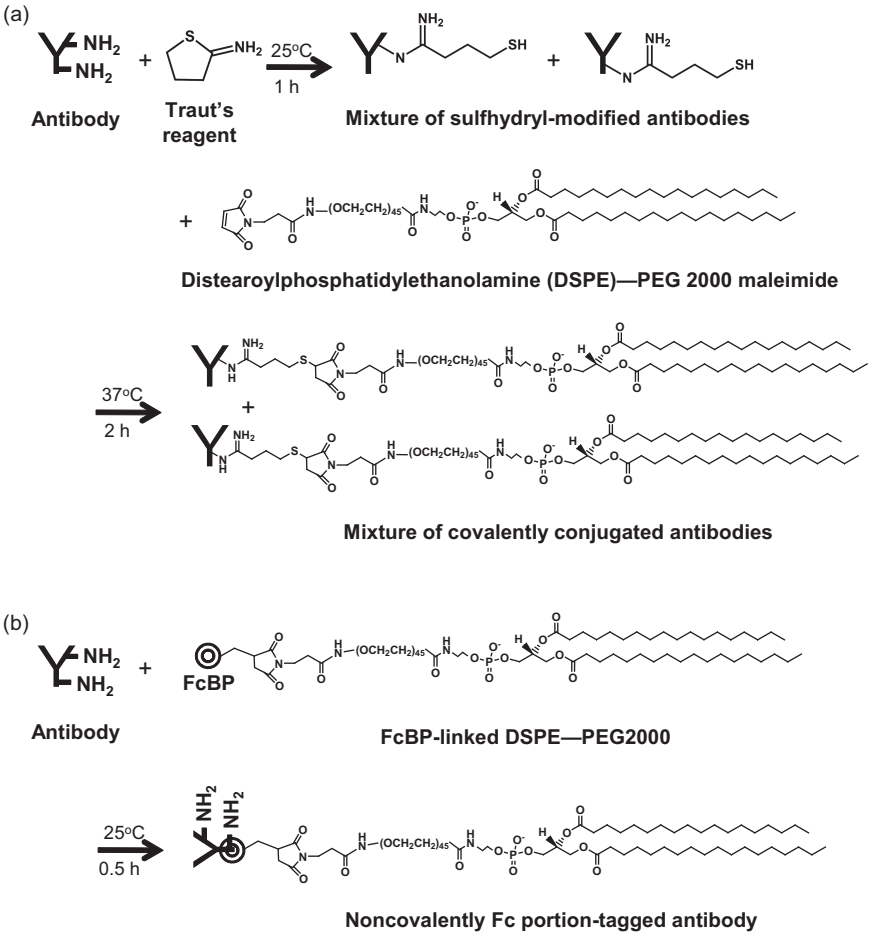


**FIGURE 2.1** Covalent and noncovalent antibody-tagging methods. (a) Covalent tagging of antibody to the distal end of maleimide-functionalized PEG derivative of lipid. (b) FcBP-based noncovalent and specific tagging of antibody to a PEG derivative of lipid.

### 2.2.2 Peptides and Proteins

Compared to the variety of antibodies that can confer specific binding to overexpressed antigens on pathogenic cells, the menu of proteins and peptide ligands available for modification of nanoparticles is limited. With the exception of the widely studied transferrin, few proteins have been used for modifying nanoparticles. Peptide ligands have been more actively studied owing to the recent development of phage-display technologies, which allow for discovery of new peptides with high affinity for specific targets. Compared to protein ligands, peptide ligands are advantageous in that they can be easily synthesized and are thus commercially available. Moreover, their relatively small size makes it easy to handle in chemical

coupling methods. For example, functional amino acids can be added synthetically near each terminal of peptides, providing control over coupling sites and reaction methods.

***2.2.2.1 Transferrin.*** The iron-binding protein transferrin has been widely studied for modifying anticancer drug-loaded nanomedicines owing to the high expression of transferrin receptors on various cancer cells [28,29]. Iron-bound transferrin specifically interacts with transferrin receptors, resulting in receptor-mediated endocytosis and delivery of iron into cells [30]. The expression levels of transferrin receptors are reported to be relatively high in rapidly proliferating cells, such as epithelial cells and cancer cells, compared to normal cells. Cancers known to express high levels of transferrin receptors include lung adenocarcinoma, breast cancer, chronic lymphocytic leukemia, and brain tumors [31]. The use of transferrin modification for enhancing the delivery of anticancer drugs to tumor tissues has been reported for various nanoparticles, including liposomes [32,33], polymeric nanoparticles [34–36], albumin nanoparticles [37], and gold nanoparticles [38].

Transferrin modification of doxorubicin-loaded stealth liposomes was shown to enhance cellular delivery of doxorubicin to HepG2 cells compared to unmodified, plain doxorubicin liposomes [32]. In this application, transferrin was linked to the surfaces of liposomes using 1-ethyl-3-(3-dimethylaminopropyl)-carbodiimide and *N*-hydroxylsuccinimide as coupling agents. Following intravenous administration in rats at a dose of 5 mg/kg, the pharmacokinetics of transferrin-modified liposomal doxorubicin was not substantially different from those of unmodified liposomal doxorubicin. Notably, despite the similar pharmacokinetics, intravenously administered transferrin-modified doxorubicin showed a greater distribution to tumor tissues in HepG2 cell-bearing mice, with peak accumulation occurring 24 h after administration. The similar pharmacokinetics in rats but enhanced anticancer effects observed for transferrin-modified doxorubicin liposomes reflect enhanced tumor tissue distribution and cellular uptake.

Transferrin-modified liposomes have been studied as a delivery system for an antisense oligonucleotide [33]. In this application, G3139, an antisense oligonucleotide targeting Bcl-2, was complexed with a mixture of liposomes and protamines to yield lipoplexes, after which transferrin-conjugated PEG lipids were inserted into the lipoplexes. Thirty-six hours after intravenous administration of G3139 (5 mg/kg) in K562-bearing xenografted mice, tumor uptake of transferrin-modified lipoplexes was higher than that of the free form of G3139. Because there was no comparison between transferrin-modified liposomes and plain liposomes in this study, it is difficult to conclude that the observed increase in tumor uptake was due to transferrin, given the EPR effect noted for nanoparticles.

In addition to liposomes, various polymeric nanoparticles have been modified with transferrin. For example, biodegradable poly(DL-lactide-*co*-glycolide) (PLGA) nanoparticles have been loaded with paclitaxel and modified with transferrin [35]. The surface of paclitaxel-loaded PLGA nanoparticles was epoxy activated and reacted with transferrin. A pharmacokinetic study of paclitaxel (20 mg/kg) in Sprague–Dawley rats

showed that the mean residence time of paclitaxel delivered by plain PLGA nanoparticles and transferrin-modified PLGA increased by 2.68 and 3.76 folds, respectively, as compared to a free paclitaxel solution. The distribution of paclitaxel to tumor tissues was highest with transferrin-modified PLGA, reaching a peak accumulation 2 h after intravenous administration.

Polycyanoacrylate nanoparticles carrying PEG-hydroxycamptothecin have also been modified with transferrin [39]. In normal Wistar rats, the pharmacokinetic profile of hydroxycamptothecin delivered in transferrin-modified PEG nanoparticles did not substantially differ from that in unmodified PEG nanoparticles. A fit to a two-compartment model showed that the elimination half-life of free PEG-hydroxycamptothecin was 0.68 h, but unmodified PEG nanoparticles and transferrin-modified PEG nanoparticles increased the half-life to 23.21 and 20.92 h, respectively. The tumor distribution of transferrin-modified PEG nanoparticles in S180 tumor-bearing mice, measured as area under the curve, was 2.12-fold higher than that of unmodified PEG nanoparticles, and reached a peak accumulation 4 h after intravenous administration.

Doxorubicin-loaded human serum albumin nanoparticles have recently been dual-modified with transferrin and tumor necrosis factor-related apoptosis-inducing ligand (TRAIL) [37]. Dual labeling with transferrin and TRAIL was reported to enable the delivery of nanoparticles to multiple tumor types, including HCT 116 and CAPAN-1 cells, and drug-resistant MCF-7/ADR cells. Moreover, dual-ligand modification with transferrin and TRAIL was shown to provide synergistic tumor cell-killing activity, reflecting the cytotoxicity of doxorubicin and the induction of apoptosis via the TRAIL ligand. Transferrin/TRAIL-modified doxorubicin nanoparticles intravenously administered in HCT116-xenografted mice showed accumulation in tumor tissues, reaching a peak 16 h after injection.

One concern regarding transferrin-mediated delivery is rapid intracellular trafficking kinetics [30]. After receptor-mediated endocytosis, transferrin receptor/transferrin complexes release iron in the endosome and then rapidly recycle back to the cell surface. The short residence time of transferrin receptor-containing endosomes within cells may reduce the ability of transferrin to deliver the drugs entrapped in nanomedicines. One approach to improving the efficacy of transferrin-modified nanomedicines is genetically engineering variants of transferrin with decreased iron release rates [40]. This study tested genetic variants of transferrin in the context of a transferrin-conjugated cytotoxin, but this approach may be applicable to other transferrin-modified nanomedicines.

***2.2.2.2 Arginine–Glycine–Aspartic Acid (RGD).*** Integrins are among the most widely studied receptors for the tumor-targeted delivery of nanomedicines. Integrins, composed of a heterodimer of α- and β-subunits, play a role in such cellular functions as cellular adhesion, migration, and proliferation [41]. Of several types of integrins, integrin isoforms αυβ3, α5β1, and αυβ6 are overexpressed in various tumor tissues [42] and are thus attractive targets for cancer therapy. Integrins αυβ3 and αυβ5 have been well studied for cancer targeting since they are overexpressed not only in tumor cells but also in angiogenic endothelial cells [43].

The tripeptide, arginine–glycine–aspartic acid (RGD), is a part of certain extracellular proteins such as fibronectin and vitronectin that interact with integrins. This integrin-binding ligand has been used to modify the surfaces of nanoparticles carrying anticancer drugs. Since Pierschbacher and Ruoslahti first reported that RGD was involved in cell attachment [44], many groups have studied RGD for integrin-directed delivery of nanomedicines. Cyclic RGD is preferred to linear RGD because linear RGD has poor binding affinity for integrins [45]. Cyclic RGD has been linked to polymeric nanoparticles [46], dendrimers [47,48], and liposomes [49,50].

The cyclic pentapeptide RGDyK has been co-conjugated to PEGylated polyamidoamine dendrimers with the anticancer drug doxorubicin for targeting to tumor neovascular endothelial cells and tumor tissues [47]. In this application, doxorubicin was conjugated via an acid-sensitive *cis*-aconityl linkage to allow the release of free doxorubicin in the acidic endosomal compartment after endocytosis. The cyclic RGDyK pentapeptide was then linked using bifunctional *N*-hydroxysulfosuccinimide-polyoxyethylene-maleimide. Although no quantitative biodistribution data were provided, tumor sections were immunofluorescence-stained using an antivascular endothelial cadherin antibody to monitor the extent of tumor growth. Immunofluorescence staining revealed that vascular endothelial cadherin levels and, thus, tumor size were lowest in tumor tissues treated with cyclic RGDyK pentapeptide-linked conjugates.

In a recent study by the same group, cyclic RGD-linked PEGylated polyamidoamine dendrimers were conjugated with doxorubicin using an acid-insensitive succinic linkage or acid-sensitive aconityl linkage [48]. Quantitative biodistribution data showed that tumor accumulation of cyclic RGD–polymer conjugates with doxorubicin attached through an acid-sensitive aconityl linkage was twofold lower than that with doxorubicin attached through an acid-insensitive succinic linkage. Despite the lower tumor accumulation, cyclic RGD–polymer conjugates with doxorubicin attached via an acid-sensitive aconityl linkage promoted the longest survival times in mice. The prolonged survival effect is attributed to the acid-labile release of doxorubicin in endolysosomes following uptake of the conjugates by the tumor cells. Unlike acid-insensitive linkage, acid-sensitive linkage may facilitate the liberation of free doxorubicin from the conjugates and increase the diffusion of the liberated doxorubicin from acidic lysosomes to the cytoplasm and subsequently to the nucleus in an actively intercalating form. This study highlights the importance of free drug release at tumor sites, especially within intracellular environments.

Polycationic liposomes have also been modified with a cyclic RGD–PEG lipid derivative for systemic delivery of siRNA [49]. In this application the cationic charges of liposomes for electrostatic complexation with negatively charged siRNA were provided by dicetylphosphate-tetraethylenepentamine. The cyclic RGD modification was introduced into liposomes using a PEG lipid derivative of cyclic RGD in the liposome preparation. After intravenous administration into C57BL/6 mice bearing luciferase-expressing B16F10-luc2 cells, cyclic RGD–PEG polycationic liposomes containing luciferase siRNA produced the greatest degree of luciferase knockdown in B16F10 metastatic lung tissues [49].

Although most RGD-modified nanomedicines have been developed for the purpose of tumor-targeted delivery, other *in vivo* applications have been reported based on the high expression of αυβ3 integrins on angiogenic vascular endothelial cells. For example, instead of anticancer drugs, the anti-inflammatory drug dexamethasone phosphate has been loaded into RGD–PEGylated liposomes and tested in a rat arthritis model [50]. In this study, the authors observed an antiarthritic effect of the dexamethasone phosphate loaded RGD–PEGylated liposomes.

***2.2.2.3 Bombesin.*** Bombesin is a 14-amino-acid peptide with a high affinity for the gastrin-releasing peptide receptor, also called the bombesin receptor, which is highly expressed on the surface of prostate, breast, and small-cell lung carcinoma cells. Exploiting this overexpression of bombesin receptors, researchers have modified nanoparticles with bombesin peptide for tumor targeting and demonstrated enhanced delivery via bombesin receptors *in vitro* [51,52]. Bombesin-functionalized gold nanoparticles have also been designed as a delivery system for X-ray contrast agents for molecular computed tomography imaging of prostate tumors in mice [53].

***2.2.2.4 Other Peptides.*** Peptides that can distinguish between tumor and normal tissues are attractive in the development of cancer diagnostics and targeted-delivery systems for anticancer agents. The *in vivo* applications of other peptide-modified nanomedicines are summarized in Table 2.3. Peptides can be used as antibody replacements for modification of nanomedicines if the corresponding peptide-binding receptors are clearly identified [54]. The phage display technique has contributed greatly to the discovery of numerous potent peptides that bind to specific tumor cells with high affinity. However, phage display-derived peptides are limited by the lack of understanding of their cognate receptors. The development of techniques that allow the identification of specific receptors for peptides is needed for the advancement of this field.

### 2.2.3 Sugars and Polysaccharides

***2.2.3.1 Galactose.*** Galactose, a simple type of sugar, is a primary metabolic energy source for humans. Apart from the oral absorption of galactose in nutrients such as sugar beets, gum, and milk products, this monosaccharide is biochemically synthesized in the body as glucoproteins and glycolipids. In the nanomedicine field, galactose has been utilized for liver-targeted delivery because galactose-bearing asialoglycoproteins can bind the asialoglycoprotein receptors localized to parenchymal liver cells (i.e., hepatocytes) [64]. Thus, it is possible to target nanoparticles to liver cells and liver-derived cells by tethering galactose molecules to the nanoparticle surface.

Galactosylated multimodular lipoplexes have been studied as a means to deliver plasmid DNA into primary hepatocytes [65]. Galactosylated cationic vectors, including lipids and polyethyleneimines, have been shown to increase the efficiency of nucleic acid transfection into hepatocytes compared with the vectors lacking galactose. On the other hand, it is possible that the galactosylated cationic

**TABLE 2.3** ***In Vivo*** **Applications of Peptide-Modified Nanomedicines**

| Peptide | Target | Nanoparticles | Active Substances | *In Vivo* Model | References |
|---|---|---|---|---|---|
| SFSIIHTPILPL | Hepatocellular carcinoma cells | Liposome | Doxorubicin | Mahlavu-derived hepatocellular carcinoma-bearing mice | 55 |
| GGGGYSAYPDSVPMMSK | Ephrin type-A receptor 2 | Magnetic nanoparticle | Fluorescent dye (rhodamine) | BG-1-bearing mice | 56 |
| YHWYGYTPQNVI/LARLLT | Epidermal growth factor receptor | Liposome | Fluorescent dye (Cy5.5) | H1299-bearing mice | 57,58 |
| HVGGSSV | Irradiated tumors | Liposome | Doxorubicin | H460-bearing mice | 59,60 |
| CDPGYIGSR | Laminin receptor | Liposome | Fluorouracil | B16F10-bearing mice | 61 |
| HAIYPRH | Transferrin receptor | Polyamidoamine dedrimer nanoparticle | Doxorubicin | Bel-7402-bearing mice | 62,63 |

nanoparticles might also have promoted nonspecific cellular uptake and cytotoxicity due to their positive charges. To reduce nonspecific cellular delivery, Letrou-Bonneval and colleagues developed multimodular assemblies with a guanidinium-cholesterol-based cationic liposomes core, which contained internally condensed plasmid DNA and external PEG with galactose at distal ends as a steric stabilizer. These colloidally stabilized, galactosylated lipoplexes were shown to increase the transfection efficiency of plasmid DNA in primary hepatocytes of Sprague–Dawley rats, whereas ungalactosylated lipoplexes did not.

*N*-Acetylgalactosamine, an amino sugar derivative of galactose, has also been used to target asialoglycoprotein receptors, specifically for receptor-targeted delivery of siRNA-carrying lipid nanoparticles [66]. To ascertain the role of *N*-acetylgalactosamine and asialoglycoprotein receptor interaction in the liver delivery of siRNA, these researchers intravenously administered factor VII-specific siRNA-loaded ionizable lipid nanoparticles containing various molar amounts of *N*-acetylgalactosamine-modified lipids to apolipoprotein $E^{-/-}$ mice at an siRNA dose of 0.2 mg/kg. A maximal gene-silencing effect in serum was observed at 0.15 mol% *N*-acetylgalactosamine moiety, indicating the importance of exogenous galactose modification of nanoparticles [66].

***2.2.3.2 Mannose and Its Derivatives.*** Mannose, a sugar monomer of a hexose, has several important biological roles, including the glycosylation of proteins. Unlike galactose interactions with asialoglycoprotein receptors on parenchymal liver cells, D-mannose is recognized by a mannose receptor generally present on nonparenchymal liver cells, such as Kupffer cells, macrophages resident in the liver.

Mannosylated liposomes have been studied for targeted delivery to alveolar macrophages [67]. Liposomes containing 9.1 mol% of mannosylation exhibited the most efficient uptake by alveolar macrophages *in vitro*. Tests showed that [$^3$H] cholesteryl hexadecyl ether-labeled mannosylated liposomes administered by pulmonary aerosolization in Sprague–Dawley rats accumulated in alveolar macrophages, with a peak level 2.5-fold greater than that observed with plain liposomes occurring 2 h after administration.

Mannose as a targeting moiety has also been used to modify ultrasound-responsive bubble liposomes for gene delivery [68]. Complexes of such mannose-modified liposomes with plasmid DNA have demonstrated mannose receptor-expressing cell-specific gene transfer *in vivo*. In the conventional lipoplex form of plasmid DNA delivery, endosomal trafficking of lipoplexes induced the production of tumor necrosis factor-α through recognition of plasmid DNA by endosomal toll-like receptor 9. It was expected that ultrasound stimulation would allow the delivery of lipoplexes directly into the cytoplasm, reducing the likelihood of DNA recognition by tumor necrosis factor-α. As predicted, exposure to ultrasound following intravenous administration of mannose-modified bubble liposomes to mice was shown to significantly reduce toll-like receptor 9 production in the exposed tissue [69].

Mannan, a mannose-containing polysaccharide recognized by the mannose receptor of antigen-presenting cells, has been used to coat superparamagnetic

iron oxide nanoparticles and applied to imaging of lymph nodes [70]. Mannan-coated superparamagnetic iron oxide nanoparticles injected into mice were shown to distribute to the lymph node to a greater degree than dextran-coated or poly(vinyl alcohol)-coated iron oxide nanoparticles. The accumulation of mannan-coated superparamagnetic iron oxide nanoparticles in the lymph node peaked after 1 h and gradually decreased until 24 h. Unlike mannan-coated nanoparticles, poly(vinyl alcohol)-coated iron oxide nanoparticles were mainly distributed in the spleen.

Mannose-6-phosphate has been used to modify human serum albumin for targeted delivery to cells expressing the mannose-6-phosphate/insulin-like growth factor-II receptor [71]. Such conjugates demonstrated an affinity for insulin-like growth factor-II receptor expressing B16 melanoma cells and were internalized by them. Two hour after intravenous injection into B16 bearing C57BL/6 mice, $^{125}$I-doxorubin-linked conjugates were mainly distributed in the liver (50% of injected dose), followed by the lung, kidney, stomach, small intestine, and tumor. A total of 3% of the injected dose of $^{125}$I-doxorubin-linked conjugates was distributed in tumor tissues, an amount less than 1/10 that in the liver. A particularly valuable aspect of this study is its quantitative presentation of biodistribution data as percentage of injected dose in 20 organs and tissues. Although the terminology "tumor targeting" is commonly used, the absolute percentage of distribution to the tumors as reported in this study suggests that the nanoparticles distribute in other tissues of the body to substantial extents.

***2.2.3.3 Hyaluronic Acid.*** Hyaluronic acid, an anionic glycosaminoglycan, is a carbohydrate abundant in the body. Notably, it is a major component of the extracellular matrix, which participates in cell proliferation, migration, and interaction with cell surface receptors, including intracellular adhesion molecule-1 and CD44, the receptor for hyaluronic acid-mediated motility [72,73].

Reducible polyethyleneimine-grafted hyaluronic acid has been synthesized and used for systemic delivery of siRNA against liver cirrhosis [74]. Systemic administration of apolipoprotein B-specific siRNA using reducible polyethyleneimine-grafted hyaluronic acid complexes was shown to reduce target gene expression and increase therapeutic efficacy in a mouse model of liver cirrhosis.

In liver fibrosis, the expression of matrix metalloproteinases is imbalanced; thus, one therapeutic strategy is to restore metalloproteinase activity by delivering plasmid DNA encoding matrix metalloproteinases. To this end, hyaluronic acid, which is negatively charged, can be used to form electrostatic triple complexes with preformed dual complexes of plasmid DNA encoding matrix metalloproteinase 13 and polyethyleneimine [75]. The resulting DNA–polyethyleneimine is shielded with hyaluronic acid by charge interactions. Intravenous administration of the hyaluronic acid-shielded plasmid DNA complexes into mice was shown to ameliorate the progression of liver fibrosis as evidenced by a reduction in collagen deposition. In this study, the shielding with hyaluronic acid was shown to reduce the toxicity of the polyethyleneimine vector. Without hyaluronic acid shielding, the maximum dose of plasmid DNA that could be systemically administered was 1 mg/kg, owing to the toxicity of polyethyleneimine. However, the shielding of

hyaluronic acid reduced the toxicity of polyethyleneimine, allowing 100% survival of mice at a plasmid DNA dose of at least 3 mg/kg.

In a recent study, hyaluronic acid itself was used as nanoparticle material for anticancer drug delivery [76]. Hyaluronic acid is biodegraded by the endogenous enzyme hyaluronidase, which is known to be particularly abundant in malignant tissues and metastatic cancers. To exploit this abundance of hyaluronidase and facilitate rapid release of drug at tumor sites, researchers have formulated PEG-conjugated hyaluronic acid nanoparticles loaded with an anticancer drug, camptothecin. Studies of the biodistribution of camptothecin-loaded PEGylated hyaluronic acid nanoparticles showed tumor localization in mice bearing CD44-expressing tumors. Camptothecin-loaded PEGylated hyaluronic acid nanoparticles prevented significant increases in tumor size for at least 35 days.

### 2.2.4 Small Molecule Chemicals

***2.2.4.1 Folate.*** Folate—water soluble Vitamin $B_9$—plays several roles in biosynthesis, including nucleic acid metabolism (e.g., DNA and RNA synthesis), prevention of DNA changes, and amino acid production. Notably, folate is required for the replication of DNA during cell division and proliferation. Cancer cells, which grow very rapidly, express high surface levels of folate receptors, which mediate cellular uptake of folate. Given the higher relative expression of folate receptors on tumor cells, it is not surprising that folate has been widely studied as a ligand molecule for nanomedicines carrying anticancer drugs or imaging agents [77,78]. *In vivo* applications of folate-modified nanomedicines are summarized in Table 2.4.

In the study of Yamada and colleagues [79], different *in vitro* and *in vivo* results were observed between doxorubicin in free form and folate-modified liposomes. *In vitro*, free doxorubicin showed higher antitumor activity than doxorubicin in folate-modified liposomes. In folate receptor-expressing KB cells, the concentration leading to 50% of cell death was 0.65 and 1.27 μg/ml for doxorubicin in free form and in liposomes with folate linked to the distal end of PEG lipid chain (mw 5000), respectively. In contrast, *in vivo* antitumor activity of doxorubicin was found to be higher in folate-liposomes as compared to free form. This difference was explained by the different *in vivo* distribution pattern of doxorubicin. The distribution of doxorubicin to tumor tissues was found only in the mice intravenously injected with folate-modified liposomes but not in the mice treated with free doxorubicin.

The importance of drug release at the target site has motivated development of an approach that controls the release of drug in a pH-responsive manner. Here, folate has been linked to pH-sensitive polymeric micelles composed of PEG–poly (aspartate–hydrazone–doxorubicin) for tumor tissue delivery of doxorubicin [80]. In this application, the anticancer drug doxorubicin was conjugated to pH-sensitive hydrazone to facilitate the drug release in acidic intracellular compartments, such as lysosomes and endosomes, where pH ranges from 5 to 6. Conjugation of folate to pH-sensitive micelles dramatically increased cellular uptake and cytotoxicity *in vitro* compared to unmodified micelles. However, an *in vivo* distribution study revealed doxorubicin accumulation in tumors after delivery using micelles was similar with or

**TABLE 2.4 *In Vivo* Applications of Folate-Modified Nanomedicines**

| Nanomedicines | Active Substances | *In Vivo* Application | References |
|---|---|---|---|
| Liposome | Doxorubicin | M109-bearing mice | 79 |
| Liposome | Doxorubicin | M109-bearing mice, J6456-bearing mice, KB-bearing mice | 82 |
| Liposome | Docetaxel | Sarcoma-180-bearing mice | 83 |
| Liposome | Plasmid DNA | M109-bearing mice | 84 |
| Liposome | Plasmid DNA | SCC7-bearing mice | 85 |
| Liposome | Antisense Oligodeoxynucleotide | KB-bearing mice | 86 |
| Polymeric micelle | Doxorubicin | A2780/doxorubicin-resistant tumor-bearing mice | 87 |
| Polymeric micelle | Adriamycin | KB-bearing mice | 80 |
| Heparin-based nanoparticle | Paclitaxel | KB-3-1-bearing mice | 81 |
| Polyamidoamine dendrimer | EGFR-antisense oligonucleotides | C6-bearing rats | 88 |
| Nanohydrogel | 5-Fluorouracil | Wistar rats | 89 |
| Albumin nanoparticles | $^{188}$Re, cisplatin | SKOV3-bearing mice | 90 |
| Chelator conjugate | $^{99m}$Tc | KB-bearing mice | 91 |

without folate modification. Despite the similar tumor accumulation, folate-conjugated micelles showed lower *in vivo* toxicity and a reduced effective dose, increasing the therapeutic window by fivefold compared to the free drug. The absence of a difference in tumor accumulation between micelles with and without folate conjugation was explained by the dominance of the EPR effect in determining the tumor accumulation of drug carriers. The higher therapeutic window of folate-modified micelles indicates that folate conjugation may affect the tumor distribution of micelles after extravasation, rather than the migration of micelles from the bloodstream into the tumor tissues.

Consistent with the results observed in this micelle study, another group reported no difference in the tumor tissue accumulation of a folate-modified nanomedicine [81]. Specifically, molecular imaging revealed that the tumor tissue distribution of heparin-paclitaxel nanoparticles modified with folate ligands was similar to that of unmodified particles. To determine whether nanoparticles simply accumulated outside tumor cells or actually entered them, these researchers stained tumor cells in the tumor tissues with an antihuman CD326 antibody. This tumor cell-identification strategy revealed that folate-modified nanoparticles were predominantly taken up by tumor cells, whereas unmodified nanoparticles were present in the extracellular space. Moreover, tumor volume was found to be significantly smaller in the

group treated with folate-modified nanoparticles compared to groups treated with unmodified nanoparticles or free drug. This increase in the *in vivo* anticancer efficacy of folate-modified nanoparticles suggests tumor-cell targeting after extravasation.

***2.2.4.2 Alendronate.*** Although small molecule chemical ligands have many advantages such as lower immunogenicity, ease of handling in modification processes, and possibly lower costs, only a few small molecule chemical ligands other than folate have been studied.

Alendronate, a drug used to prevent and treat osteoporosis, has been studied as a bone-targeting moiety in a mouse model of bone metastasis [92]. In this application, alendronate was linked to an *N*-(2-hydroxypropyl) methacrylamide-paclitaxel conjugate. An *in vivo* study showed that the alendronate-linked conjugate-treated group achieved greater tumor regression than did groups treated with free paclitaxel or a mixture of free paclitaxel and alendronate. Moreover, the alendronate-linked conjugate showed less hemolytic activity than Cremophor EL, a commercial vehicle for paclitaxel. Although the authors speculated that the higher efficacy and safety of alendronate-linked conjugate was due to a more favorable pharmacokinetic profile of the conjugate, quantitative data on tumor tissue distribution were not provided in this study.

***2.2.4.3 Anisamide.*** Sigma receptors are membrane-bound proteins with a high affinity for neuroleptics. Although sigma receptors are expressed on normal tissues, including liver, lungs, kidneys, and central nervous system, they are known to be overexpressed in diverse tumors. Huang and colleagues tested liposomes modified with the sigma receptor-binding ligand, anisamide, first studying anisamide-modified liposomes as delivery vehicles for doxorubicin [93]. A PEG-conjugated phospholipid was derivatized with an anisamide ligand, which was then incorporated into the doxorubicin-loaded liposome. *In vivo* imaging revealed that anisamide-conjugated liposomal doxorubicin showed significantly improved accumulation in the tumor.

In a recent study by the same group, anisamide-modified cationic liposomes were studied for systemic delivery of c-Myc-specific siRNA [94]. Four hours after intravenous administration to B16F10 melanoma tumor-bearing mice, anisamide-modified liposomes showed increased accumulation in tumors compared to unmodified liposomes. Western blots of tumor tissues showed that the c-Myc protein was silenced in the group treated with anisamide-modified liposomes, but not in other groups.

## 2.3 CURRENT CHALLENGES IN CLINICAL DEVELOPMENT OF LIGAND-MODIFIED NANOMEDICINE

### 2.3.1 Nonclinical and Clinical Studies

In addition to the *in vivo* case studies highlighted earlier, numerous animal studies are ongoing. Despite these efforts, very few powerful ligand-modified nanomedicine candidates have entered clinical trials. One such case is the transferrin-conjugated

immunotoxin, TransMid (Xenova Biomedix Co.), which has entered clinical trials as a locally administered treatment for brain tumors [95]. In this conjugate, transferrin is bound to diphtheria toxin through a lysine linker and a thioether. After binding to glioma cells, the conjugate enters the cells via receptor-mediated endocytosis and releases toxin, which kills the cells. This immunotoxin completed Phase I and II clinical trials, but was stopped during Phase III trials in late 2006 by the sponsor, Xenova Biomedix, according to a conditional power analysis showing that the conjugate would not significantly improve upon standard treatments for high-grade brain tumors.

A second case, EC-145, is under development by Endocyte Inc. [96], although putting this chemical drug conjugate in the category of nanomedicine may be arguable. EC-145 is a conjugate of folate and vinka alkaloid, in which folate is linked to the anticancer agent desacetylvinblastine monohydrazide through a peptide spacer. After promising preclinical trial results [97], the conjugate is currently in Phase II clinical trials in patients with advanced ovarian cancer.

One other nanomedicine undergoing clinical testing is Calando Pharmaceutical's nanomaterial, denoted CALAA-01, which is now in Phase I clinical trials [5,98] after completion of multi-dosing tests in nonhuman primates [99,100]. CALAA-01 consists of a human transferrin-tagged cyclodextrin-based polymer for systemic administration of siRNA against the M2 subunit of ribonucleotide reductase. Researchers at Calando have reported that systemic delivery of siRNA-loaded nanoparticles significantly reduced target gene expression in patients. Because CALAA-01 is not only a targeted nanoparticle but also an siRNA therapeutic, this first clinical trial for the treatment of solid cancer represents a significant step for the field of nanomedicine. It is worthwhile to note that the focus of the study in question was on the systemic administration of RNAi therapeutics rather than on the targeting ligand, and comparative studies of siRNA-loaded nanoparticles with and without transferrin are lacking. It remains to be seen whether the transferrin ligand played a vital role in the results of the clinical study.

### 2.3.2 Challenges and Future Perspectives

Numerous ligand-modified nanomedicine candidates have shown promise *in vitro*, but there are multiple *in vivo* barriers that can nullify the targeting capabilities of nanoparticles. One of the biggest barriers against the delivery to target tissues is the massive uptake of nanoparticles in the reticuloendothelial systems such as liver. As quantitatively reported in the study using mannose-6-phosphate as a targeting moiety [71], described earlier, 50% of the injected dose was found in the liver, with less than 5% of the dose reaching the tumor tissue. To increase the delivery to tumors or other target sites, it is essential that the substantial distribution of nanoparticles to the liver be decreased.

One approach for reducing such off-target delivery is to increase the selectivity of the nanomedicine. To date, most nanomedicines have been modified with a single ligand whose receptors are overexpressed on target cells. However, it is practically not easy to identify suitable receptors that exist exclusively on target cells. As a

result, off-target binding of nanomedicines to other normal cells that also express the same receptor is unavoidable. To circumvent this limitation and increase the selectivity for tumor cells, which actually overexpress multiple types of surface receptors, Saul and coworkers have recently suggested a dual-ligand approach [101]. In their *in vitro* study, liposomes were labeled with folate and an anti-EGFR antibody to further improve the specificity of tumor targeting. Although additional proof-of-concept experiments are needed for *in vivo* applications, a dual-ligand approach might be a future direction for enhancing the specificity of delivery.

One general expectation in ligand-modified nanomedicine is that the ligands actively improve the migration of nanoparticles from the bloodstream to target sites, such as tumor cells. This idea of "smart" ligands that guide nanoparticles from the rapidly flowing bloodstream to specific tumor sites is not fully realized in practice. Although several recent studies have shown dramatic increase in tumor accumulation of ligand-modified nanoparticles, others have consistently reported the lack of enhanced tumor tissue accumulation, even after high-density ligand modification. As discussed by the authors of these recent studies [80], most ligand-grafted nanoparticles arrive at tumor sites, not through the active guidance of ligands but through the predominant EPR effect.

Several approaches have been suggested for improving the penetration of nanoparticles through vascular endothelial barriers. One such approach is to increase the action of the EPR effect for both passive and targeted delivery of nanomedicines. Several groups have attempted to enhance EPR effects by administration of nitric oxide donor chemicals [102]. Additional approaches include the use of components for vascular targeting [98] and the design of multifunctional nanoparticles with bio-inspired penetration enhancers [103].

Although conjugated ligands do not appear to have a profound effect on the tumor accumulation of nanomedicines, they may contribute to enhancing tumor cell uptake of nanoparticles after extravasation, as suggested by the case of folate-modified micelles [80], described earlier. These and similar studies suggest that the greater anticancer efficacy of ligand-modified nanoparticles, despite similar accumulation at tumor sites, may be due to enhanced intracellular delivery.

After the arrival of ligand-modified nanomedicines at tumor tissues, the remaining challenge is penetration through the tumor mass. Owing to their size, nanoparticles at tumor sites have been shown to accumulate mainly in the peripheral area of the vasculature, rather than penetrating deeply into the tumor mass [78]. Reducing the sizes of nanoparticles in the tumor microenvironment has been suggested as a way of overcoming the size-dependent hindrance of nanoparticle penetration into the tumor matrix [104]. In this study, 100 nm nanoparticles were shrunk to a size of 10 nm by the action of proteases that are highly expressed in the tumor microenvironment. Such an approach might provide a framework for subsequent efforts to overcome the penetration limitation.

The composition of engineered nanoparticles is an important topic in the commercialization of nanomedicine. Currently, the most widely used method—chemical tethering of ligands onto preformed nanoparticles—has several limitations with respect to manufacturing and quality control [105]. The underlying chemistry

requires optimization to prevent batch-to-batch variability and variations according to the types of ligands and vehicles [106]. Complicated synthesis procedures also require complex purification procedures, which typically lead to expensive and time-consuming scaling-up efforts. With this manufacturing issue in mind, recent studies have demonstrated the advantages of self-assembly, prefunctionalization, and genetic engineering strategies in eliminating or minimizing complex manufacturing processes.

Ligand-engineered nanomedicines need to be used in conjunction with patient-tailored diagnostic tools to achieve the best therapeutic outcome. Depending on the cancer type, the expression levels of the receptors for targeting ligands may vary. Moreover, only patients that overexpress the specific receptors would be responsive to the ligand-modified nanomedicine. Thus, for optimal therapeutic efficacy, patients may need to be diagnosed in advance to determine whether they are likely to be responsive to a specific ligand-modified nanomedicine.

## 2.4 CONCLUSIONS

A variety of *in vivo* applications of ligand-engineered nanoparticles have been described. Despite the merits of engineered nanoparticles, remaining *in vivo* barriers can severely limit the activity of these particles after reaching tumor sites. At present, an ideal nanoparticle that satisfies all the requirements for successful clinical trials does not exist. However, the first and difficult steps toward developing a concrete understanding of the drawbacks have been taken, and related research has already been started to conquer them. As these obstacles are surmounted, the list of clinically successful nanomedicines is expected to grow in the near future.

### Acknowledgments

This work was supported by research grants from the Ministry of Education, Science, and Technology (2012007005), from the Bio & Medical Technology Development Program of the National Research Foundation funded by the Korean government (MEST) (No. 20120006123), and from the Korean Health Technology R&D Project (No. A092010), Ministry for Health, Welfare & Family Affairs, Republic of Korea.

## REFERENCES

1. Davis, M. E., Chen, Z., Shin, D. M. (2008). Nanoparticle therapeutics: an emerging treatment modality for cancer. *Nature Reviews Drug Discovery 7*, 771–782.
2. Petros, R. A., DeSimone, J. M. (2010). Strategies in the design of nanoparticles for therapeutic applications. *Nature Reviews Drug Discovery 9*, 615–627.

3. Matsumura, Y., Maeda, H. (1986). A new concept for macromolecular therapeutics in cancer chemotherapy: mechanism of tumoritropic accumulation of proteins and the antitumor agent smancs. *Cancer Research 46*, 6387–6392.
4. Mohanty, C., Das, M., Kanwar, J. R., Sahoo, S. K. (2011). Receptor mediated tumor targeting: an emerging approach for cancer therapy. *Current Drug Delivery 8*, 45–56.
5. Zhang, L., Gu, F. X., Chan, J. M., Wang, A. Z., Langer, R. S., Farokhzad, O. C. (2008). Nanoparticles in medicine: therapeutic applications and developments. *Clinical Pharmacology and Therapeutics 83*, 761–769.
6. Irache, J. M., Esparza, I., Gamazo, C., Agüeros, M., Espuelas, S. (2011). Nanomedicine: novel approaches in human and veterinary therapeutics. *Veterinary Parasitology 180*, 47–71.
7. Jain, R. K., Stylianopoulos, T. (2010). Delivering nanomedicine to solid tumors. *Nature Reviews Clinical Oncology 7*, 653–664.
8. Hatakeyama, H., Akita, H., Harashima, H. (2011). A multifunctional envelope type nano device (MEND) for gene delivery to tumours based on the EPR effect: a strategy for overcoming the PEG dilemma. *Advances in Drug Delivery Review 63*, 152–160.
9. Yu, B., Tai, H. C., Xue, W., Lee, L. J., Lee, R. J. (2010). Receptor-targeted nanocarriers for therapeutic delivery to cancer. *Molecular Membrane Biology 27*, 286–298.
10. Fay, F., Scott, C. J. (2011). Antibody-targeted nanoparticles for cancer therapy. *Immunotherapy 3*, 381–394.
11. Noble, C. O., Guo, Z., Hayes, M. E., Marks, J. D., Park, J. W., Benz, C. C., Kirpotin, D. B., Drummond, D. C. (2009). Characterization of highly stable liposomal and immunoliposomal formulations of vincristine and vinblastin. *Cancer Chemotherapy and Pharmacology 64*, 741–751.
12. Weng, K. C., Noble, C. O., Papahadjopoulos-Sternberg, B., Chen, F. F., Drummond, D. C., Kirpotin, D. B., Wang, D., Hom, Y. K., Hann, B., Park, J. W. (2008). Targeted tumor cell internalization and imaging of multifunctional quantum dot-conjugated immunoliposomes *in vitro* and *in vivo*. *Nano Letters 8*, 2851–2857.
13. Lu, R. M., Chang, Y. L., Chen, M. S., Wu, H. C. (2011). Single chain anti-c-Met antibody conjugated nanoparticles for *in vivo* tumor-targeted imaging and drug delivery. *Biomaterials 32*, 3265–3274.
14. Wang, K., Ruan, J., Qian, Q., Song, H., Bao, C., Zhang, X., Kong, Y., Zhang, C., Hu, G., Ni, J., Cui, D. (2011). BRCAA1 monoclonal antibody conjugated fluorescent magnetic nanoparticles for *in vivo* targeted magnetofluorescent imaging of gastric cancer. *Nanobiotechnology 9*, 23.
15. Sauter, G., Maeda, T., Waldman, F. M., Davis, R. L., Feuerstein, B. G. (1996). Patterns of epidermal growth factor receptor amplification in malignant gliomas. *The American Journal of Pathology 148*, 1047–1053.
16. Feng, B., Tomizawa, K., Michiue, H., Miyatake, S., Han, X. J., Fujimura, A., Seno, M., Kirihata, M., Matsui, H. (2009). Delivery of sodium borocaptate to glioma cells using immunoliposome conjugated with anti-EGFR antibodies by ZZ-His. *Biomaterials 30*, 1746–1755.
17. Liu, P., Li, Z., Zhu, M., Sun, Y., Li, Y., Wang, H., Duan, Y. (2010). Preparation of EGFR monoclonal antibody conjugated nanoparticles and targeting to hepatocellular carcinoma. *Journal of Materials Science: Materials in Medicine 21*, 551–556.

18. Pattillo, C. B., Venegas, B., Donelson, F. J., Del Valle, L., Knight, L. C., Chong, P. L., Kiani, M. F. (2009). Radiation-guided targeting of combretastatin encapsulated immunoliposomes to mammary tumors. *Pharmaceutical Research 26*, 1093–1100.
19. Tuffin, G., Waelti, E., Huwyler, J., Hammer, C., Marti, H. P. (2005). Immunoliposome targeting to mesangial cells: a promising strategy for specific drug delivery to the kidney. *Journal of the American Society of Nephrology 16*, 3295–3305.
20. Scott, R. C., Rosano, J. M., Ivanov, Z., Wang, B., Chong, P. L., Issekutz, A. C., Crabbe, D. L., Kiani, M. F. (2009). Targeting VEGF-encapsulated immunoliposomes to MI heart improves vascularity and cardiac function. *The FASEB Journal 23*, 3361–3367.
21. Kim, S. S., Peer, D., Kumar, P., Subramanya, S., Wu, H., Asthana, D., Habiro, K., Yang, Y. G., Manjunath, N., Shimaoka, M., Shankar, P. (2010). RNAi-mediated CCR5 silencing by LFA-1-targeted nanoparticles prevents HIV infection in BLT mice. *Molecular Therapy 18*, 370–376.
22. Manjappa, A. S., Chaudhari, K. R., Venkataraju, M. P., Dantuluri, P., Nanda, B., Sidda, C., Sawant, K. K., Murthy, R. S. (2011). Antibody derivatization and conjugation strategies: application in preparation of stealth immunoliposome to target chemotherapeutics to tumor. *Journal of Controlled Release 150*, 2–22.
23. Schalén, C., Truedsson, L., Christensen, K. K., Christensen, P. (1985). Blocking of antibody complement-dependent effector functions by streptococcal IgG Fc-receptor and staphylococcal protein A. *Acta Pathologica, Microbiologica, et Immunologica Scandinavica. Section B, Microbiology 93*, 395–400.
24. Jung, Y., Kang, H. J., Lee, J. M., Jung, S. O., Yun, W. S., Chung, S. J., Chung, B. H. (2008). Controlled antibody immobilization onto immunoanalytical platforms by synthetic peptide. *Analytical Biochemistry 374*, 99–105.
25. Feng, B., Tomizawa, K., Michiue, H., Han, X. J., Miyatake, S., Matsui, H. (2010). Development of a bifunctional immunoliposome system for combined drug delivery and imaging *in vivo*. *Biomaterials 31*, 4139–4145.
26. De Lano, W. L., Ultsch, M. H., de Vos, A. M., Wells, J. A. (2000). Convergent solutions to binding at a protein–protein interface. *Science 287*, 1279–1283.
27. Oh, Y. K., Chang, R. S., Yu, Y. H., Kim, W. K. Lipopeptides with specific affinity to Fc region of antibodies and antigen-recognizing lipid nanoparticles comprising the same. 10-2010-0119486 (Korea, Patent).
28. Daniels, T. R., Delgado, T., Helguera, G., Penichet, M. L. (2006). The transferrin receptor part II: targeted delivery of therapeutic agents into cancer cells. *Clinical Immunology 121*, 159–176.
29. Daniels, T. R., Bernabeu, E., Rodríguez, J. A., Patel, S., Kozman, M., Chiappetta, D. A., Holler, E., Ljubimova, J. Y., Helguera, G., Penichet, M. L. (2012). Transferrin receptors and the targeted delivery of therapeutic agents against cancer. *Biochimica et Biophysica Acta 1820*, 291–317.
30. Yoon, D. J., Liu, C. T., Quinlan, D. S., Nafisi, P. M., Kamei, D. T. (2011). Intracellular trafficking considerations in the development of natural ligand-drug molecular conjugates for cancer. *Annals of Biomedical Engineering 39*, 1235–1251.
31. Daniels, T. R., Delgado, T., Rodriguez, J. A., Helguera, G., Penichet, M. L. (2006). The transferrin receptor part I: biology and targeting with cytotoxic antibodies for the treatment of cancer. *Clinical Immunology 121*, 144–158.

32. Li, X., Ding, L., Xu, Y., Wang, Y., Ping, Q. (2009). Targeted delivery of doxorubicin using stealth liposomes modified with transferrin. *International Journal of Pharmaceutics 373*, 116–123.
33. Zhang, X., Koh, C. G., Yu, B., Liu, S., Piao, L., Marcucci, G., Lee, R. J., Lee, L. J. (2009). Transferrin receptor targeted lipopolyplexes for delivery of antisense oligonucleotide G3139 in a murine K562 xenograft model. *Pharmaceutical Research 26*, 1516–1524.
34. Xu, Z., Gu, W., Huang, J., Sui, H., Zhou, Z., Yang, Y., Yan, Z., Li, Y. (2005). *In vitro* and *in vivo* evaluation of actively targetable nanoparticles for paclitaxel delivery. *International Journal of Pharmaceutics 288*, 361–368.
35. Shah, N., Chaudhari, K., Dantuluri, P., Murthy, R. S. R., Das, S. (2009). Paclitaxel-loaded PLGA nanoparticles surface modified with transferrin and Pluronic®P85, an *in vitro* cell line and *in vivo* biodistribution studies on rat model. *Journal of Drug Targeting 17*, 533–542.
36. Ren, W., Chang, J., Yan, C., Qian, X., Long, L., He, B., Yuan, X., Kang, C., Betbeder, D., Sheng, J., Pu, P. (2010). Development of transferrin functionalized poly(ethylene glycol)/poly(lactic acid) amphiphilic block copolymeric micelles as a potential delivery system targeting brain glioma. *Journal of Materials Science: Materials in Medicine 21*, 2673–2681.
37. Bae, S., Ma, K., Kim, T. H., Lee, E. S., Oh, K. T., Park, E-S., Lee, K. C., Youn, Y. S. (2012). Doxorubicin-loaded human serum albumin nanoparticles surface-modified with TNF-related apoptosis-inducing ligand and transferrin for targeting multiple. *Biomaterials 33*, 1536–1546.
38. Choi, C. H. J., Alabi, C. A., Webster, P., Davis, M. E. (2010). Mechanism of active targeting in solid tumors with transferrin-containing gold nanoparticles. *Proceedings of the National Academy of Sciences of the United States of America 107*, 1235–1240.
39. Hong, M., Zhu, S., Jiang, Y., Tang, G., Sun, C., Fang, C., Shi, B., Pei, Y. (2010). Novel anti-tumor strategy: PEG–hydroxycamptothecin conjugate loaded transferrin-PEG-nanoparticles. *Journal of Controlled Release 141*, 22–29.
40. Yoon, D. J., Chu, D. S., Ng, C. W., Pharm, E. A., Mason, A. B., Hudson, D. M., Smith, V. C., MacGillivray, R. T., Kamei, D. T. (2009). Genetically engineering transferrin to improve its *in vitro* ability to deliver cytotoxins. *Journal of Controlled Release 133*, 178–184.
41. Hsu, A. R., Veeravagu, A., Cai, W., Hou, L. C., Tse, V., Chen, X. (2007). Integrin $\alpha\upsilon\beta3$ antagonists for anti-angiogenic cancer treatment. *Recent Patents on Anti-Cancer Drug Discovery 2*, 143–160.
42. McCabe, N. P., De, S., Vasanji, A. (2007). Prostate cancer specific integrin $\alpha\upsilon\beta3$ modulates bone metastatic growth and tissue remodeling. *Oncogene 26*, 6238–6243.
43. Shim, G., Lee, S., Choi, H., Lee, J., Kim, C-W., Byun, Y., Oh, Y-K. (2011). Nanomedicines for receptor-mediated tumor targeting. *Recent Patents on Nanomedicine 1*, 138–148.
44. Pierschbacher, M. D., Ruoslahti, E. (1984). Cell attachment activity of fibronectin can be duplicated by small fragments of the molecule. *Nature 309*, 30–33.
45. Juliano, R. L., Alam, R., Dixit, V., Kang, H. M. (2009). Cell-targeting and cell-penetrating peptides for delivery of therapeutic and imaging agents. *Wires Nanomedicine and Nanobiotechnology 1*, 324–335.

46. Danhier, F., Vroman, B., Lecouturier, N., Crokart, N., Pourcelle, V., Freichels, H., Jérôme, C., Marchand-Brynaert, J., Feron, O., Préat, V. (2009). Targeting of tumor endothelium by RGD-grafted PLGA-nanoparticles loaded with paclitaxel. *Journal of Controlled Release 140*, 166–173.
47. Zhu, S., Qian, L., Hong, M., Zhang, L., Pei, Y., Jiang, Y. (2011). RGD-modified PEG–PAMAM–DOX conjugate: *in vitro* and *in vivo* targeting to both tumor neovascular endothelial cells and tumor cells. *Advanced Materials 23*, 84–89.
48. Zhang, L., Zhu, S., Qian, L., Pei, Y., Qiu, Y., Jiang, Y. (2011). RGD-modified PEG–PAMAM–DOX conjugates: *in vitro* and *in vivo* studies for glioma. *European Journal of Pharmaceutics and Biopharmaceutics 79*, 232–240.
49. Yonenaga, N., Kenjo, E., Asai, T., Tsuruta, A., Shimizu, K., Dewa, T., Nango, M., Oku, N. (2012). RGD-based active targeting of novel polycation liposomes bearing siRNA for cancer treatment. *Journal of Controlled Release 160*, 177–181.
50. Koning, G. A., Schiffelers, R. M., Wauben, M. H. M., Kok, R. J., Mastrobattista, E., Molema, G., ten Hagen, T. L. M., Storm, G. (2006). Targeting of angiogenic endothelial cells at sites of inflammation by dexamethasone phosphate–containing RGD peptide liposomes inhibits experimental arthritis. *Arthritis and Rheumatism 54*, 1198–1208.
51. Martin, A. L., Hickey, J. L., Ablack, A. L., Lewis, J. D., Luyt, L. G., Gillies, E. R. (2010). Synthesis of bombesin-functionalized iron oxide nanoparticles and their specific uptake in prostate cancer cells. *Journal of Nanoparticle Research 12*, 1599–1608.
52. Honer, M., Mu, L., Stellfeld, T., Graham, K., Martic, M., Fischer, C. R., Lehmann, L., Schubiger, P. A., Ametamey, S. M., Dinkelborg, L., Srinivasan, A., Borkowski, S. (2011). $^{18}$F-labeled bombesin analog for specific and effective targeting of prostate tumors expressing gastrin-releasing peptide receptors. *The Journal of Nuclear Medicine 52*, 270–278.
53. Chanda, N., Kattumuri, V., Shukla, R., Zambre, A., Katti, K., Upendran, A., Kulkarni, R. R., Kan, P., Fent, G. M., Casteel, S. W., Smith, C. J., Boote, E., Robertson, J. D., Cutler, C., Lever, J. R., Katti, K. V., Kannan, R. (2010). Bombesin functionalized gold nanoparticles show *in vitro* and *in vivo* cancer receptor specificity. *Proceedings of the National Academy of Sciences of the United States of America 107*, 8760–8765.
54. Brown, K. C. (2010). Peptidic tumor targeting agents: the road from phage display peptide selections to clinical applications. *Current Pharmaceutical Design 16*, 1040–1054.
55. Lo, A., Lin, C-T., Wu, H-C. (2008). Hepatocellular carcinoma cell-specific peptide ligand for targeted drug delivery. *Molecular Cancer Therapeutics 7*, 579–589.
56. Scarberry, K. E., Dickerson, E. B., McDonald, J. F., Zhang, Z. J. (2008). Magnetic nanoparticle-peptide conjugates for *in vitro* and *in vivo* targeting and extraction of cancer cells. *Journal of the American Chemical Society 130*, 10258–10262.
57. Song, S., Liu, D., Peng, J., Sun, Y., Li, Z., Gu, J-R., Xu, Y. (2008). Peptide ligand-mediated liposome distribution and targeting to EGFR expressing tumor *in vivo*. *International Journal of Pharmaceutics 363*, 155–161.
58. Song, S., Liu, D., Peng, J., Deng, H., Guo, Y., Xu, L. X., Miller, A. D., Xu, Y. (2009). Novel peptide ligand directs liposomes toward EGF-R high-expressing cancer cells *in vitro* and *in vivo*. *The FASEB Journal 23*, 1396–1404.
59. Han, Z., Fu, A., Wang, H., Diaz, R., Geng, L., Onishko, H., Hallahan, D. E. (2008). Noninvasive assessment of cancer response to therapy. *Nature Medicine 14*, 343–349.

60. Lowery, A., Onishko, H., Hallahan, D. E., Han, Z. (2011). Tumor-targeted delivery of liposome-encapsulated doxorubicin by use of a peptide that selectively binds to irradiated tumors. *Journal of Controlled Release 150*, 117–124.
61. Dubey, P. K., Singodia, D., Vyas, S. P. (2010). Liposomes modified with YIGSR peptide for tumor targeting. *Journal of Drug Targeting 18*, 373–380.
62. Han, L., Huang, R., Liu, S., Huang, S., Jiang, C. (2010). Peptide-conjugated PAMAM for targeted doxorubicin delivery to transferrin receptor overexpressed tumors. *Molecular Pharmaceutics 7*, 2156–2165.
63. Han, L., Huang, R., Li, J., Liu, S., Huang, S., Jiang, C. (2011). Plasmid pORF-hTRAIL and doxorubicin co-delivery targeting to tumor using peptide-conjugated polyamidoamine dendrimer. *Biomaterials 32*, 1242–1252.
64. Zhu, L., Ye, Z., Cheng, K., Miller, D. D., Mahato, R. I. (2008). Site-specific delivery of oligonucleotides to hepatocytes after systemic administration. *Bioconjugate Chemistry 19*, 290–298.
65. Letrou-Bonneval, E., Chèvre, R., Lambert, O., Costet, P., André, C., Tellier, C., Pitard, B. (2008). Galactosylated multimodular lipoplexes for specific gene transfer into primary hepatocytes. *The Journal of Gene Medicine 10*, 1198–1209.
66. Akinc, A., Querbes, W., De, S., Qin, J., Frank-Kamenetsky, M., Jayaprakash, K. N., Jayaraman, M., Rajeev, K. G., Cantley, W. L., Dorkin, J. R., Butler, J. S., Qin, L. L., Racie, T., Sprague, A., Fava, E., Zeigerer, A., Hope, M. J., Zerial, M., Sah, D. W. Y., Fitzgerald, K., Tracy, M. A., Manoharan, M., Koteliansky, V., Fougerolles, A., Maier, M. A. (2010). Targeted delivery of RNAi therapeutics with endogenous and exogenous ligand-based mechanisms. *Molecular Therapy 18*, 1357–1364.
67. Chono, S., Kaneko, K., Yamamoto, E., Togami, K., Morimoto, K. (2010). Effect of surface-mannose modification on aerosolized liposomal delivery to alveolar macrophages. *Drug Development and Industrial Pharmacy 36*, 102–107.
68. Un, K., Kawakami, S., Suzuki, R., Maruyama, K., Yamashita, F., Hashida, M. (2010). Development of an ultrasound-responsive and mannose-modified gene carrier for DNA vaccine therapy. *Biomaterials 31*, 7813–7826.
69. Un, K., Kawakami, S., Yoshida, M., Higuchi, Y., Suzuki, R., Maruyama, K., Yamashita, F., Hashida, M. (2011). The elucidation of gene transferring mechanism by ultrasound-responsive unmodified and mannose-modified lipoplexes. *Biomaterials 32*, 4659–4669.
70. Vu-Quang, H., Yoo, M-K., Jeong, H-J., Lee, H-J., Muthiah, M., Rhee, J. H., Lee, J-H., Cho, C-S., Jeong, Y. Y., Park, I-K. (2011). Targeted delivery of mannan-coated superparamagnetic iron oxide nanoparticles to antigen-presenting cells for magnetic resonance-based diagnosis of metastatic lymph nodes *in vivo*. *Acta Biomaterialia 7*, 3935–3945.
71. Prakash, J., Beljaars, L., Harapanahalli, A. K., Zeinstra-Smith, M., Jager-Krikken, A., Hessing, M., Steen, H., Poelstra, K. (2010). Tumor-targeted intracellular delivery of anticancer drugs through the mannose-6-phosphate/insulin-like growth factor II receptor. *International Journal of Cancer 126*, 1966–1981.
72. Koo, H., Huh, M. S., Sun, I-C., Yuk, S. H., Choi, K., Kim, K., Kwon, I. C. (2011). *In vivo* targeted delivery of nanoparticles for theranosis. *Accounts of Chemical Research 44*, 1018–1028.
73. Kouvidi, K., Berdiaki, A., Nikitovic, D., Katonis, P., Afratis, N., Hascall, V. C., Karamanos, N. K., Tzanakakis. G. N. (2011). Role of receptor for hyaluronic acid-

mediated motility (RHAMM) in low molecular weight hyaluronan (LMWHA)-mediated fibrosarcoma cell adhesion. *Journal of Biological Chemistry 286*, 38509–38520.

74. Park, K., Hong, S. W., Hur, W., Lee, M-Y., Yang, J-A., Kim, S. W., Yoon, S. K., Hahn, S. K. (2011). Target specific systemic delivery of TGF-β siRNA/(PEI-SS)-*g*-HA complex for the treatment of liver cirrhosis. *Biomaterials 32*, 4951–4958.
75. Kim, E. J., Cho, H. J., Park, D., Kim, J. Y., Kim, Y. B., Park, T. G., Shim, C. K., Oh, Y. K. (2011). Antifibrotic effect of MMP13-encoding plasmid DNA delivered using polyethylenimine shielded with hyaluronic acid. *Molecular Therapy 19*, 355–361.
76. Choi, K. Y., Yoon, H. Y., Kim, J-H., Bae, S. M., Park, R-W., Kang, Y. M., Kim, I-S., Kwon, I. C., Choi, K., Jeong, S. Y., Kim, K., Park, J. H. (2011). Smart nanocarrier based on PEGylated hyaluronic acid for cancer therapy. *ACS Nano 5*, 8591–8599.
77. Low, P. S., Kularatne, S. A. (2009). Folate-targeted therapeutic and imaging agents for cancer. *Current Opinion in Chemical Biology 13*, 256–262.
78. Kolhatkar, R., Lote, A., Khambati, H. (2011). Active tumor targeting of nanomaterials using folic acid, transferrin and integrin receptors. *Current Drug Discovery Technology 8*, 197–206.
79. Yamada, A., Taniguchi, Y., Kawano, K., Honda, T., Hattori, Y., Maitani, Y. (2008). Design of folate-linked liposomal doxorubicin to its antitumor effect in mice. *Clinical Cancer Research 14*, 8161–8168.
80. Bae, Y., Nishiyama, N., Kataoka, K. (2007). *In vivo* antitumor activity of the folate-conjugated pH-sensitive polymeric micelle selectively releasing adriamycin in the intracellular acidic compartments. *Bioconjugate Chemistry 18*, 1131–1139.
81. Wang, X., Li, J., Wang, Y., Cho, K. J., Kim, G., Gjyrezi, A., Koenig, L., Giannakakou, P., Shin, H. J., Tighiouart, M., Nie, S., Chen, Z. G., Shin, D. M. (2009). HFT-T, a targeting nanoparticle, enhances specific delivery of paclitaxel to folate receptor-positive tumors. *ACS Nano 3*, 3165–3174.
82. Gabizon, A., Horowitz, A. T., Goren, D., Tzemach, D., Shmeeda, H., Zalipsky, S. (2003). *In vivo* fate of folate-targeted polyethylene-glycol liposomes in tumor-bearing mice. *Clinical Cancer Research 9*, 6551–6559.
83. Li, X., Tian, X., Zhang, J., Zhao, X., Chen, X., Jiang, Y., Wang, D., Pan, W. (2011). *In vitro* and *in vivo* evaluation of folate receptor-targeting amphiphilic copolymermodified liposomes loaded with docetaxel. *International Journal of Nanomedicine 6*, 1167–1184.
84. Hofland, H. E. J., Masson, C., Iginla, S., Osetinsky, I., Reddy, J. A., Leamon, C. P., Scherman, D., Bessodes, M., Wils, P. (2002). Folate-targeted gene transfer *in vivo*. *Molecular Therapy 5*, 739–744.
85. Duarte, S., Faneca, H., De Lima, M. C. P. (2012). Folate-associated lipoplexes mediate efficient gene delivery and potent antitumoral activity *in vitro* and *in vivo*. *International Journal of Pharmaceutics 423*, 365–377.
86. Leamon, C. P., Cooper, S. R., Hardee, D. E. (2003). Folate-liposome-mediated antisense oligodeoxynucleotide targeting to cancer cells: evaluation *in vitro* and *in vivo*. *Bioconjugate Chemistry 14*, 738–747.
87. Kim, D., Gao, Z., Lee, E. S., Bae, Y. H. (2009). *In vivo* evaluation of doxorubicin-loaded polymeric micelles targeting folate receptors and early endosomal pH in drug-resistant ovarian cancer. *Molecular Pharmaceutics 6*, 1353–1362.
88. Kang, C., Yuan, X., Li, F., Pu, P., Yu, S., Shen, C., Zhang, Z., Zhang, Y. (2009). Evaluation of folate-PAMAM for the delivery of antisense oligonucleotides to rat C6

glioma cells *in vitro* and *in vivo*. *Journal of Biomedical Materials Research A 93*, 585–594.

89. Blanco, M. D., Guerrero, S., Benito, M., Fernández, A., Teijón C., Olmo, R., Katime, I., Teijón, J. M. (2011). *In vitro* and *in vivo* evaluation of a folate-targeted copolymeric submicrohydrogel based on *N*-isopropylacrylamide as 5-fluorouracil delivery system. *Polymers 3*, 1107–1125.
90. Tang, Q-S., Chen, D-Z., Xue, W-Q., Xiang, J-Y., Gong, Y-C., Zhang, L., Guo, C-Q. (2011). Preparation and biodistribution of $^{188}$Re-labeled folate conjugated human serum albumin magnetic cisplatin nanoparticles ($^{188}$Re-folate-CDDP/HSA MNPs) *in vivo*. *International Journal of Nanomedicine 6*, 3077–3085.
91. Mindt, T. L., Müller, C., Melis, M., De Jong, M., Schibli, R. (2008). "Click-to-chelate": *in vitro* and *in vivo* comparison of a $^{99m}Tc(CO)_3$-labeled N(τ)-histidine folate derivative with its isostructural, clicked 1,2,3-triazole analogue. *Bioconjugate Chemistry 19*, 1689–1695.
92. Miller, K., Eldar-Boock, A., Polyak, D., Segal, E., Benayoun, L., Shaked, Y., Satchi-Fainaro, R. (2011). Antiangiogenic antitumor activity of HPMA copolymer-paclitaxel-alendronate conjugate on breast cancer bone metastasis mouse model. *Molecular Pharmaceutics 8*, 1052–1062.
93. Banerjee, R., Tyagi, P., Li, S., Huang, L. (2004). Anisamide-targeted stealth liposomes: a potent carrier for targeting doxorubicin to human prostate cancer cells. *International Journal of Cancer 112*, 693–700.
94. Chen, Y., Bathula, S. R., Yang, Q., Huang, L. (2010). Targeted nanoparticles deliver siRNA to melanoma. *Journal of Investigational Dermatology 130*, 2790–2798.
95. Kratz, F., Elsadek, B. (2012). Clinical impact of serum proteins on drug delivery. *Journal of Controlled Release 161*, 429–445.
96. Dosio, F., Milla, P., Cattel, L. (2010). EC-145, a folate-targeted vinca alkaloid conjugate for the potential treatment of folate receptor-expressing cancers. *Current Opinion on Investigational Drugs 11*, 1424–1433.
97. Reddy, J. A., Dorton, R., Westrick, E., Dawson, A., Smith, T., Xu, L. C., Vetzel, M., Kleindl, P., Vlahov, I. R., Leamon, C. P. (2007). Preclinical evaluation of EC145, a folate-vinca alkaloid conjugate. *Cancer Research 67*, 4434–4442.
98. Timko, B. P., Whitehead, K., Gao, W., Kohane, D. S., Farokhza, O., Anderson, D., Langer, R. (2011). Advances in drug delivery. *Annual Review of Materials Research 41*, 1–20.
99. Davis, M. E. (2009). The first targeted delivery of siRNA in humans via a self-assembling, cyclodextrin polymer-based nanoparticle: from concept to clinic. *Molecular Pharmaceuticals 6*, 659–668.
100. Davis, M. E., Zuckerman, J. E., Choi, C. H. J., Seligson, D., Tolcher, A., Alabi, C. A., Yen, Y., Heidel, J. D., Ribas, A. (2010). Evidence of RNAi in humans from systemically administered siRNA via targeted nanoparticles. *Nature 464*, 1067–1070.
101. Saul, J. M., Annapragada, A. V., Bellamkonda, R. V. (2006). A dual-ligand approach for enhancing targeting selectivity of therapeutic nanocarriers. *Journal of Controlled Release 114*, 277–287.
102. Chrastina, A., Massey, K. A., Schnitzer, J. E. (2011). Overcoming *in vivo* barriers to targeted nanodelivery. *Wiley Interdisciplinary Reviews: Nanomedicine and Nanobiotechnology 3*, 421–437.

103. Yoo, J-W., Irvine, D. J., Discher, D. E., Mitragotri, S. (2011). Bio-inspired, bioengineered and biomimetic drug delivery carriers. *Nature Reviews Drug Discovery 10*, 521–535.

104. Wong, C., Stylianopoulos, T., Cui, J., Martin, J., Chauhan, V. P., Jiang, W., Popovic, Z., Jain, R. K., Bawendi, M. G., Fukumura, D. (2011). Multistage nanoparticle delivery system for deep penetration into tumor tissue. *Proceedings of the National Academy of Sciences of the United States of America 108*, 2426–2431.

105. Hirata, K., Maruyama, T., Watanabe, H., Maeda, H., Nakajou, K., Iwao, Y., Ishima, Y., Katsumi, H., Hashida, M., Otagiri, M. (2010). Genetically engineered mannosylated-human serum albumin as a versatile carrier for liver-selective therapeutics. *Journal of Controlled Release 145*, 9–16.

106. Patil, Y., Toti, U., Khdair, A., Ma, L., Panyam, J. (2009). Single-step surface functionalization of polymeric nanoparticles for targeted drug delivery. *Biomaterials 30*, 859–866.

# LIPID NANOPARTICLES FOR THE DELIVERY OF NUCLEIC ACIDS

YUHUA WANG AND LEAF HUANG

*Division of Molecular Pharmaceutics and Center for Nanotechnology in Drug Delivery, Eshelman School of Pharmacy, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA*

## 3.1 INTRODUCTION

As with the development of recombinant DNA technology, gene therapy emerged as a promising technology that would drastically improve the practice of medicine for treating inherited and acquired diseases. The idea underlying gene therapy is to treat human disease by the transfer of genetic materials into specific cells of the patient [1]. Nucleic acids, the fundamental elements that precisely controlled the expression of proteins, are very appealing therapeutic candidates due to the simplicity of the drug development strategy and minimal side effects compared with conventional drugs. Moreover, the discovery of RNA interference (RNAi) by Fire and Mello in the late 1990s has opened up an entirely new field of "gene therapy," using small RNA fragments that potently knockdown the target gene expression with high specificity and efficiency [2]. Gene therapy has hardly made its mark in medicine. Successful implementation of gene transfer in the clinic requires drug delivery techniques that can efficiently and specifically deliver nucleic acids to the site of action after systemic administration.

The physicochemical properties of nucleic acid, such as vulnerability to nucleases, high anionic charge content, and high molecular weight (MW), preclude naked unformulated molecules from performing their functions after systemic delivery. Therefore, a delivery vector should facilitate nucleic acid therapeutics to access the

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*,
Edited by Shayne C. Gad
Copyright © 2013 John Wiley & Sons, Inc.

target cell, cross the cell membrane, escape the endosome compartments, and reach the sites of action, which is the cytoplasm for small interfering RNA (siRNA) and microRNA, and the nucleus for plasmid DNA (pDNA). Apart from physical methods such as electroporation [3] and hydrodynamic injection [4], there are two major classes of gene delivery vehicles that have been extensively investigated and developed to deliver nucleic acids to the target cells *in vivo*: viral and nonviral vectors. The former class harnesses the extremely high transduction efficiency of virus particles to infect target cells. For this purpose, the viral genome is genetically modified to deplete virulent and proliferation elements, leaving only the expression machinery in most cases. However, this class of vectors could stimulate serious immune responses [5] or result in unwanted host genome integration, which in rare cases would induce cancer development [6,7]. The latter vectors are composed of liposomes, polymers, inorganic materials, and nucleic acid conjugates. The synthetic vectors, compared with their counterparts, are easier for scale-up production, lower in immunogenicity, and cheaper in manufacture cost. Although favorable in many aspects, nonviral vectors suffer from relatively low delivery efficiency due to multiple barriers between the site of administration and the site of action. Ever since the first landmark publications in 1987 [8], cationic liposome based lipid nanoparticles (LNPs) turn out to be the most widely used and efficient nanometric device for *in vitro* delivery of nucleic acid. Cationic liposomes condense DNA into an ill-defined particulate structure (lipoplex) with excess positive charges, which allow efficient, but nonspecific, binding with the negatively charged cell surface and subsequent internalization [9]. The condensation of DNA also results in protection from enzymatic degradation. Effective as they appear *in vitro*, most of the LNPs failed or behaved far from satisfactory *in vivo*, especially for systemic delivery to solid tumor.

In this chapter, we first elucidate the extracellular and the intracellular barriers the LNPs encounter after systemic administration. Understanding the challenges in the delivery will lead to a rational design of novel lipid molecules or nanometric devices that could address these challenges. Next, we use a bottom-up strategy to identify the latest progression in lipid component development in the LNPs, and then compare various formulations that were exploited to encapsulate nucleic acid for successful *in vivo* delivery. The nucleic acids in this review include pDNA, siRNA, and antisense oligonucleotide (ODN), which either serve for gain of function or loss of function to treat various diseases.

## 3.2 BARRIERS TO SYSTEMIC GENE DELIVERY

Except for some directly accessible tumor sites after surgery ablation, tumor sites are typically only reachable through systemic administration of therapeutics in the bloodstream. Unfortunately, for the LNP-based nucleic acid delivery systems, multiple barriers exist extracellularly and intracellularly. These barriers substantially impede the successful delivery and therefore undermine the efficacy of delivered therapeutics.

### 3.2.1 Extracellular Barriers

Extracellular barriers refer to the hurdles imposed on the LNPs in the blood circulation before tumor cells take them up. These barriers include the liposome instability in serum, reticuloendothelial system (RES) uptake, vascular endothelium, and extracellular matrix (Figure 3.1).

Positively charged surface, found in most LNPs, results in strong protein adsorption when they are exposed to the blood serum where a high concentration of proteins with countercharges are present. The protein adsorption leads to agglomeration of the LNPs, followed by entrapment in the lung capillary and in the spleen macrophage enriched mesh tissue [10]. Some of the adsorbed blood proteins, for example, IgM and C4, are the known opsonins that facilitate the uptake of the LNPs by the RES cells, mainly the liver Kupffer cells and the splenic macrophages [10]. Conventional LNPs that encapsulate nucleic acid usually have a short blood circulation half-life, which could be substantially prolonged by coating the particles with polysaccharides or poly(ethylene glycol) (PEG) residues [10,11]. The hydrophilic coating endows the LNP with resistance to RES and hence the coated formulation is named *stealth liposomes*. Prolonged circulation provides more opportunity for LNPs to access the tumor site.

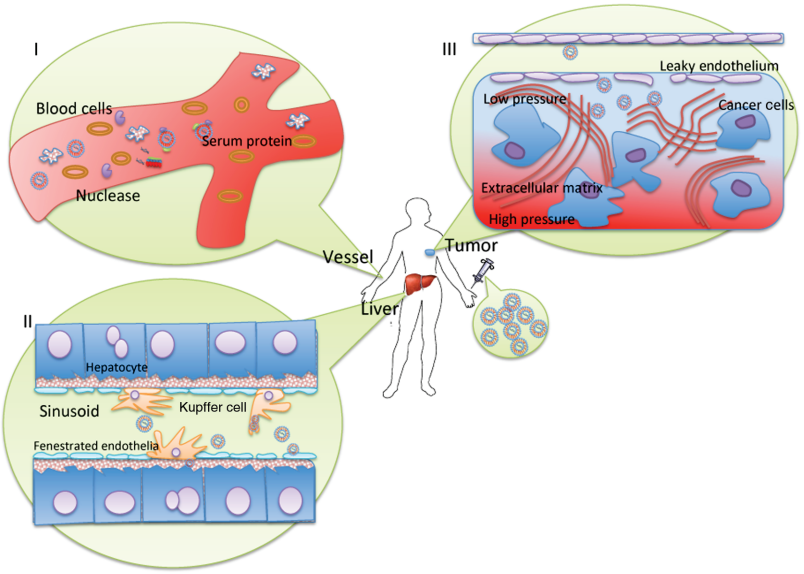


**FIGURE 3.1** Illustration of barriers to intravenous delivery of gene therapeutics. After LNP is intravenously injected into human body, serum protein tends to be adsorbed to the surface of LNP, destabilizing colloidal stability of LNP and causing the release of cargo (Inset I). The RES in the liver may take up LNPs, which reduces half-life of LNPs in the blood circulation (Inset II). Circulating LNPs extravasate into the interstitial space tumor tissue through the leaky endothelial lining and result in high accumulation in the tumor tissue due to the lack of lymphatic drainage (Inset III). However, LNPs have to traverse through the crowded extracellular matrix against the increasing fluid pressure to access the proximity of cancer cell.

Egress from the bloodstream and across the endothelial barriers also poses an insurmountable obstacle for liposome-mediated nucleic acid delivery. Molecules larger than 5–10 nm cannot readily traverse the capillary endothelium of healthy tissue [12] but keep circulating until they are cleared from the blood by renal clearance or other mechanisms. Most solid tumors, liver, and spleen allow the extravasation of large molecules or particles due to different anatomical characteristics of the endothelium in these organs [13]. It was reported by Maeda that tumor vascular has elevated permeability, which accounted for higher extravasation of macromolecules into interstitial tissue spaces [14] and a prolonged retention due to deficient lymphatic drainage [15]. This tumor-specific feature was named *enhanced permeability and retention* (EPR) effect, which has been extensively exploited as the major passive tumor-targeting strategy to achieve a higher accumulation in the tumor than plasma or other tissues. After extravasation from the blood vessel to the tumor site by the EPR effect, the LNPs are required to penetrate into the extracellular matrix filled with polysaccharide and protein network to gain access to the tumor cell membrane. This matrix permits only diffusion of small molecules and proteins yet not the conventional LNPs which are usually 100 nm or larger. Furthermore, the defective lymphatic elevates the interstitial pressure in tumor mass, which leads to radially outward convection opposing the inward diffusion [16].

### 3.2.2 Intracellular Barriers

Provided that the LNPs have circumvented all the potential extracellular barriers and traveled to the cellular surface after the "Odyssey Journey," with physical integrity and biological activity, there are still several major intracellular barriers that hinder foreign genetic materials from being delivered to the nucleus of the cells for expression. They are (i) the cytoplasmic membrane, (ii) the endosome compartment, (iii) the intracellular trafficking, and (iv) the nuclear membrane (Figure 3.2).

The LNPs may nonspecifically interact with cell surface followed by endocytosis. Studies have also demonstrated that internalization pathway and intracellular routing of particles is size dependent. The clathrin-mediated endocytic pathway has a size limit for internalization of ~200 nm in diameter. As the size increases, caveolae-mediated internalization becomes the dominant internalization pathway [17], and the LNPs are less likely to go through the endosome route. In order to realize targeted gene delivery, the LNPs are modified with various biologically active moieties such as arginine-glycine-aspartate (RGD) peptide, folate, anisamide, and transferrin, which can be specifically recognized by certain receptors that are overexpressed on the cancer cell surface and facilitate the internalization of the LNPs. The receptor-mediated internalization of LNPs is primarily via endocytotic pathway.

After cellular uptake via endocytosis, the LNPs are encapsulated in the membrane vesicles derived by invagination of clathrin-coated pits, namely, endosomes. The pH in the lumen of early endosomes gradually experiences a drop from neutral to pH 6.0 as the vacuolar ATPase proton pump is activated. As endosomes mature, the pH value continues to drop to 5.0 when late endosomes fuse with lysosomes [18]. Nucleic

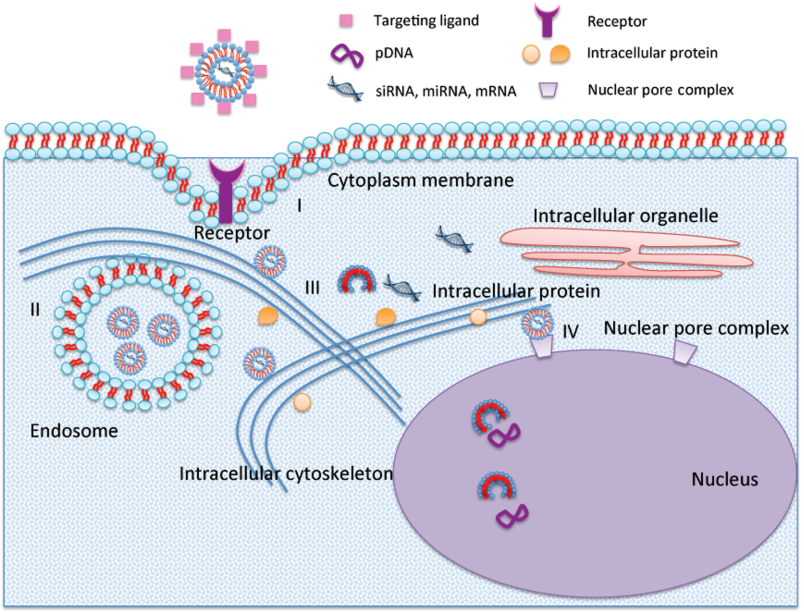


**FIGURE 3.2** Schematic illustration of proposed pathway of gene delivery mediated by the LNPs. The targeting motif on the surface of the LNPs binds to the receptor overexpressed on the surface of cancer cells, allowing the internalization of the complexes via receptor-mediated endocytosis (Step I). The LNPs will be encapsulated in the endosomal compartment. The cargo needs to be released into the cytoplasm before the endosomes fuse with lysosomes (Step II). The cargos such as siRNA, mRNA, or miRNA will take effect in the cytoplasm as released. For pDNA, the carrier has to escort the cargo through the intracellular matrix until it reaches the periphery of the nucleus (Step III). If the size of the LNP complexes is more than 40 nm, the complex can enter the nucleus only at the mitosis phase of the cell cycle when the nuclear membrane dissolves. If the LNP complex is small enough (<40 nm), the complex may pass through the nuclear pore complex (NPC) (Step IV). Inside the nucleus, the pDNA will be decomplexed from the carrier for transcription.

acids that are entrapped in the endosome are subject to degradation if they have not escaped from the endosome before endosome–lysosome fusion occurs. This usually leads to poor efficacy because nucleic acids would not ultimately reach the cytoplasm or nucleus. Therefore, endosome entrapment is always considered one of the major intracellular barriers to nonviral gene delivery. In general, when the LNPs are endocytosed, the cationic lipids can interact with anionic membrane lipids and form ion pairs. Hence, the endosome membrane is destabilized by the formation of hexagonal phase ($H_{II}$), and nucleic acids are released into the cytoplasm subsequently [19].

Once successfully released into the cytoplasm, cargos such as siRNA, antisense RNA, miRNA, and mRNA could directly take effect without further transportation. However, pDNA must traverse through cytoplasm to reach the nucleus, which is a poorly characterized event so far and also considered a rate-limiting step in the

pDNA-mediated gene transfer. According to Goodsell's depiction [20], cytoplasm is a milieu that is crowded with high concentration of soluble proteins, subcellular organisms and compartments, cytoskeleton structure, and all other small molecules. All of these cellular components, especially three types of cytoskeletal filaments, namely, microtubules, microfilaments, and intermediate filaments, impose steric constraints on the diffusion of the macromolecules, slowing down long-range transport [21]. As a result, the passive diffusion coefficients of macromolecules are inversely related to their size. It seems very unlikely that Brownian motion is the major driving force for this colloidal nonviral gene delivery system whose particle size is between 80 and 150 nm [21]. In nature, adenovirus overcomes this obstacle by utilizing microtubule structure to travel to the perinuclear area after endosome escape [22]. This phenomenon has also been observed in nonviral gene delivery. Suh et al. employed multiple particle tracking (MPT) technique to determine that the velocity of polyethyleneimine (PEI)–DNA complexes transport was 0.2 μm/s, which is the same magnitude as movement of motor proteins dyneins and kinesins along the microtubule [23]. Although a few mechanistic studies have been done to get an in-depth understanding of the microtubule-mediated trafficking of these nanodevices to date, it is speculated that the motion of the LNPs could be greatly improved if they bear a signal molecule that can bind specifically to the microtubule [24,25].

The nuclear envelope is always considered a crucial barrier to efficient nonviral gene delivery. Over the past 10 years, mechanisms for the nuclear transport of pDNA have been investigated and described. Many approaches have been developed based on exploitation of these mechanisms. In general, when carrier–DNA complexes reach perinuclear membrane, there are three scenarios for the LNPs to finally reach nucleoplasm, which are categorized according to the particle size. When the particle size is below 10 nm, it can passively diffuse through the nuclear pore complex (NPC). However, this does not apply to most of the LNPs. When the particle size is above 40 nm, it tends to accumulate around the nuclear membrane and wait for the mitosis phase to enter the nuclear compartment when the nuclear envelope is dissolved. This is partly why growth-arrested cells and other non-dividing cells are considered difficult to transfect. Also, this accounts for the observed cell-cycle dependent expression pattern of the cells transfected by nonviral vectors. When the particle size is between 10 and 40 nm and the particle carries a nuclear localization signal (NLS), it could be actively transported into the nucleus through the nuclear pore complex (NPC) via the NLS-mediated nuclear importing machinery. NLS is a short basic amino acid sequence that can be recognized by importin-α, an NLS receptor, which then dimerizes with importin-β. The importin-αβ heterodimer complex then carries the NLS bearing cargo through NPC for nuclear transport. NLS has been incorporated into cationic lipids [26,27] and proved to facilitate pDNA nuclear entry. The location of NLS in the vector is of significant importance regarding the active functionality. The signal should be readily accessible to the importin protein. If NLS overlaps with the DNA-binding domain, there is a good chance that the NLS-directed nuclear transport would be compromised.

## 3.3 COMPONENTS OF LNPs

The components in the LNPs usually include various lipids, polymer (polypeptide or synthetic polymer) and lipid-PEG conjugates, and so on. There are extensive investigations on a rational design basis, to ameliorate the components of the LNPs for improved delivery efficiency. Here we review the most representative ones.

### 3.3.1 Cationic Lipid

The encapsulation of nucleic acids of various molecular weights, ranging from 20 base pairs to several thousand base pairs, essentially depends on the cationic charges of carrier molecules. A multivalent cation is the key component because the electrostatic interaction between the cationic lipid and the anionic cargos results in condensation of nucleic acid into nanometric particles through an entropy-driven process [28].

Commonly used cationic lipids are usually classified into three categories [29] based on the charges on their headgroups: (1) monovalent aliphatic lipids such as *N*-(1-(2,3-dioleyloxy) propyl)-*N*,*N*,*N*-trimethylammonium chloride (DOTMA) [8], 1,2-dioleyl-3-trimethylammonium propane (DOTAP) [30]; (2) multivalent aliphatic lipids with headgroups containing several amine functions such as dioctadecylamidoglycylspermine (DOGS) [31], 2,3-dioleyloxy-*N*-(2-[sperminecarboxyamido] ethyl-*N*,*N*-dimethyl-1-propaniminium bromide (DOSPA) [32]; (3) cationic lipid derivatives such as 3β-(*N*-(*N*′,*N*′-dimethylaminoethane)-carbamoyl)cholesterol (DC-Chol) [33] and bis(guanidinium)-tren-cholesterol (BGTC) [34].

The cationic lipids not only are responsible for DNA condensation but also play a critical role in interacting with the negatively charged cellular membrane and triggering an efficient endocytosis for uptake. This advantage is well exploited in nonspecific tissue culture transfection *in vitro*. Nevertheless, the advantage *in vitro* turns out to be the disadvantage *in vivo*. The positive charge is detrimental to *in vivo* transfection because the charges on the particle surface tend to attract negatively charged serum proteins, subsequently leading to fast blood clearance and poor tissue distribution for the aforementioned reasons.

This problem is addressed by either coating liposomes with neutrally charged PEG polymer, which not only screen the positive charge but also provide a hydrophilic steric barrier that significantly reduces protein adsorption. Alternatively, cationic lipids with a pH-sensitive ionizable headgroup emerge. With $pK_a$ slightly lower than 7.0, ionizable cationic lipid can efficiently condense nucleic acids at acidic pH and maintain a neutral or low cationic surface charge at physiological pH (7.4) after *in vivo* administration. This strategy provides sophisticated control over surface charge of the LNPs in the blood circulation and prevents nonspecific cellular uptake, representing a recent trend in the development of lipid materials for nucleic acid delivery, especially for siRNA.

The idea of incorporating ionizable lipids in nucleic acid delivery was first reported by Semple et al. [35]. In order to overcome the unfavorable pharmacokinetics (PK) and toxicity profiles for cationic lipid based LNPs, a novel ionizable

aminolipid (1,2-dioleoyl-3-dimethylammonium propane, DODAP) was used to efficiently encapsulate antisense ODN in the LNPs. Later, another ionizable lipid DLinDMA (1,2-dilinoleyloxy-*N*,*N*-dimethyl-3-aminopropane) has proven to be highly effective regarding *in vivo* transfection efficiency [35]. The liposome made of DLinDMA showed close to neutral surface charge in the blood circulation at physiological pH. After endocytosis, the pH drops in the endosomal compartment, and the lipid is protonated. The cationic lipid is supposed to interact with naturally occurring anionic phospholipids in the endosomal membrane by electrostatic interaction, forming ion pairs that induce conformational transition from bilayer to hexagonal $H_{II}$ phase, which destabilizes endosomal membrane for nucleic acid release. The LNPs formed with ionizable lipids have been tested in rodents, nonhuman primates [36], and is now being evaluated in human clinical trials.

As a benchmark for the design of novel ionizable cationic lipid, a pool of ~1200 lipid molecules with various modifications on the hydrocarbon chain, linker, and headgroup have been implicated [37]. A sensitive mouse Factor VII model was used as the *in vivo* screening system to evaluate liposome-based hepatocyte siRNA delivery efficiency [38]. A series of modifications have been implemented in the headgroups, linkers, and hydrocarbon chains [37]. The strategy for the headgroup modifications was to change the size, acid-dissociation constant, and number of ionizable groups and to evaluate the corresponding delivery efficiency. Piperrazino, morpholino, trimethylamino, or bis-dimethylamino cations were tested, with none showing better performance than dimethylamino headgroup of 2,2-dilinoleyl-4-dimethylaminomethyl-[1,3]-dioxolane (DLin-K-DMA). In addition, the distance between the dimethylamino group and the dioxlane linker was varied by introducing various numbers of the methylene group. The results showed that insertion of a single additional methylene group (2,2-dilinoleyl-4-(2-dimethylaminoethyl)-[1,3]-dioxolane) produced a dramatic increase in potency with an *in vivo* ED50 as low as 0.1 mg/kg, which turned out to be the lowest dose for silencing Factor VII *in vivo* [37].

Linker region was optimized by testing different rates of chemical or enzymatic stability and hydrophobicity according to modifications. The results indicated that rapidly degradable linker did not necessarily correlate with high *in vivo* activity. Highly reactive ester-containing lipid showed reduced *in vivo* activity compared to less reactive alkoxy-containing lipid. The lipids were speculated to undergo an unwanted degradation before the vector successfully delivered a cargo to the cells.

For hydrocarbon chains, a correlation between number of *cis* double bonds in the chains and delivery activity was studied. It was determined that linoleyl lipid containing two double bonds per hydrocarbon chain is optimal in a transfection model, mediated by fusion with endosomal membrane for cytoplasmic release of the nucleic acid [39].

### 3.3.2 Neutral Lipid

In order to avoid unfavorable PK caused by positive surface charge after systemic delivery, Sood and colleagues have developed a method to formulate siRNA in a

neutral lipid dioleoylphosphatidylcholine (DOPC)-based-liposome for systemic delivery [40]. Since DOPC has an amphiphilic choline headgroup, which displays neutral charge at physiological pH, a conventional approach to encapsulate siRNA with a charge–charge interaction is not applicable here. Therefore, a *tert*-butanol/Tween 20 system was employed to encapsulate siRNA through a series of process including evaporation, lyophilization, and rehydration [40]. The loading efficiency of siRNA was estimated to be about 65%. The DOPC liposome was able to deliver anticancer siRNA in metastatic liver carcinoma [41,42] and orthotopic ovarian cancer model [40,43–45] with high efficacy. Although the results are impressive, the mechanism of siRNA encapsulation in neutral LNPs was not elucidated, and physicochemical properties of the LNPs were poorly characterized.

With the same aim in mind (to reduce cationic charge on the surface of liposome), Herscovici and colleagues designed a series of lipid molecules with thiourea headgroup on the basis of the observation that bis-thiourea moieties selectively bind to dihydrogen phosphate via multitopic hydrogen bonding [46]. Nevertheless, the transfection was unimpressive due to the hydrophobic nature of the lipopolythiourea liposome, which hinders the interaction with the cell membrane [46]. This feature, however, could be exploited to avoid nonspecific cell uptake and be overcome by postgrafting with targeting ligands to facilitate internalization [46]. The absence of cationic charges endowed the liposome with an improved PK profile *in vivo*, yet the benefit did not increase the tumor accumulation according to the proposed EPR effect [47]. It is also worth noticing that thiourea LNPs achieved a similar transfection level as cationic LNPs, although internalization of the thiourea LNPs was determined to be sixfold less than that of the cationic LNPs. This high expression level despite low uptake efficiency indeed revealed a favorable release profile for lipothiourea after internalization, which was measured to be three times faster than the cationic liposome. This fast unpackaging characteristic is attributable to the relatively low hydrogen bond-based interaction between phosphate group of nucleic acid and thiourea lipid [48] when compared with electrostatic interaction.

### 3.3.3 Anionic Lipid

Although anionic lipids alone cannot condense nucleic acid through electrostatic interaction, they are usually formulated into nanometric particle to reduce the nonspecific cellular uptake and adsorption by positively charged serum protein for *in vivo* nucleic acid delivery [49–52]. In this approach, nucleic acids are usually condensed with excessive polycations such as PEI or protamine to form a core complex with positive surface charge. Preformed anionic liposomes are subsequently added to the system, coating the core complexes with bilayer lipid membrane. Chen et al. formulated liposome/polycation/DNA (LPD)-II by coating liposome, composed of 1,2-dioleoyl-*sn*-glycero-3-phosphate (DOPA), 1,2-dioleoyl-*sn*-glycero-3-phosphoethanolamine (DOPE), and cholesterol, on protamine-condensed siRNA. The LPD-II formulation showed much lower immunogenicity compared with the cationic liposome coated LPD nanoparticles, which widened the therapeutic window for the therapeutic applications [52].

Another approach, reported by Srinivasan and Burgess, utilized anionic liposome as a gene carrier [53]. In their study, divalent cations (such as $Ca^{2+}$, $Mg^{2+}$, $Mn^{2+}$, and $Ba^{2+}$) were incorporated into the system to achieve the complexation between the nucleic acid and the anionic lipids. High percent of transfection was achieved *in vitro* in the presence and absence of serum and demonstrated minimal cytotoxicity. Although the hydrodynamic size of the complexed particles was at the submicron level, which was not suitable for *in vivo* applications, this formulation provided a novel way of using anionic liposome for transfection.

### 3.3.4 PEG–Lipid Conjugates

As proposed by Allen, "If you want to be invisible, look like water" [54]. For longer blood circulation, which is a prerequisite for efficient gene delivery after systemic administration, it is necessary to coat the surface of the LNPs with neutral and hydrophilic polymers such as PEG. Klibanov et al. have reported, for the first time, a prolonged liposome circulation after PEGylation back in 1990 [11], a landmark publication for long circulation liposomes. Later, Campbell observed an increased tumor accumulation of PEGylated cationic LNP due to EPR effect [55]. To date, PEG is the most appreciated hydrophilic polymer with a tunable size (normally from 300 to 25,000 Da), particularly high biocompatibility and versatility. Surface modification with PEG stabilizes liposome in the presence of serum and extends the blood circulation after administration, as a result of reduced level of protein adsorption to the hydrophobic area of the LNPs. This effect predominantly depends on the molecular weight, density, conformation, and flexibility of the PEG chain. Vonarbourg and Benoit did an excellent review on this subject [10]. It was demonstrated that the protein adsorption efficiency decreased with an increase of MW of PEG chain [56]. Most authors advocate an efficient MW in the range 1500–3500 Da [10]. In addition, the PEG density and conformation, which not only depend on the MW of PEG chain, but also the size of LNPs, are of great importance to resist opsonin penetration. Briefly, smaller LNPs with larger curvature, or smaller radius, need a higher density of PEG compared with larger ones. The conformation of PEG chain on the surface is dictated by the relationship between the distance ($D$) of the anchorage sites and the gyration radius ($R_F$). When $D > 2R_F$, the grafted PEG chain is in a mushroom-like conformation. However, when there is more PEG grafted on the surface of the lipid bilayer, which makes polymer chains closer (in such case, $D \ll R_F$), polymer chains tend to appear as a brush-like conformation. Brush conformation was believed to be most effective in repelling against the complement components, more resistant to phagocytosis (Figure 3.3).

Various lipid-PEG conjugates are inserted into the lipid bilayer membrane during or after the formation of the LNPs. The hydrophobic interaction between the anchor lipid of the conjugate and hydrophobic chain of the membrane lipid is the major force that keeps the PEG on the surface of the LNPs. Since the lipid-PEG conjugate is soluble in the aqueous bulk, the molecule also tends to shed from the liposome surface especially under a sink condition. Cullis et al. have taken the initiative to

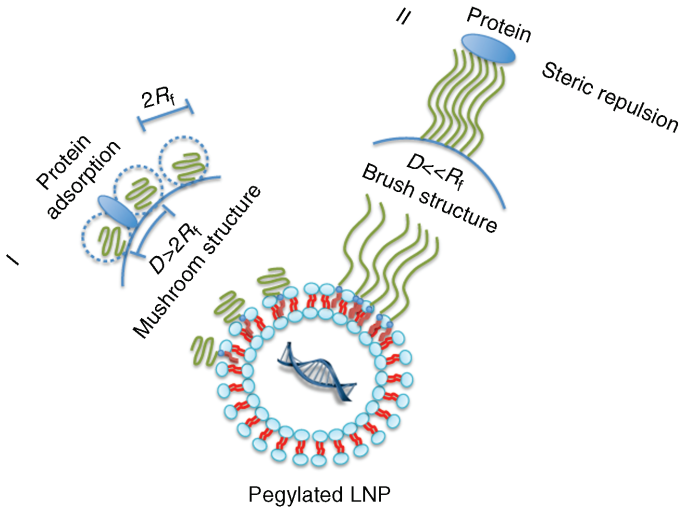


**FIGURE 3.3** Putative PEG conformation with respect to PEG concentration in the surface of LNPs. Conformations of the polymer chains are dictated by the relationship between the distance ($D$) between the anchorage sites of two chains and the gyration radius ($R_f$). When $D > 2R_f$, PEG chains form a mushroom-like structure. Protein tends to be adsorbed to the surface of the LNPs. When $D \ll R_f$, PEG chains assume brush-like conformation. The high-density, brush-like PEG configurations would sterically suppress the deposition of large proteins.

investigate and have tried to figure out the correlation between different lipid-PEG conjugates and transfection efficiency of the LNPs prepared by the dialysis method [57–60]. It was demonstrated that longer acyl chain in the lipid-PEG conjugates stabilized the particle, whereas the transfection efficiency was compromised due to less interaction between the cationic lipid and cell membrane. On the other hand, shorter acyl chain length resulted in a readily dissociable surface coating that could transform the LNPs from a stable form into a transfection-competent form. It was also demonstrated that the composition of acyl chain determined the rate and extent of PEG extraction, which was reflected on the elimination rate in the bloodstream. Adlakha-Hutcheon et al. reported the relationship between the anchor chain length and liposome circulation time [59]. In their study, a PEGylated programmable fusogenic vesicles (PFV) stabilized with PEG was employed to exhibit time-dependent destabilization in body. Three types of PEG-lipid with different acyl chain length were used to coat the surface of liposomes: 1,2-dimyristoyl-*sn*-glycero-3-phosphoethanolamine (DMPE, C14), 1,2-dipalmitoyl-*sn*-glycero-3-phosphoethanolamine (DPPE, C16), and 1,2-distearoyl-*sn*-glycero-3-phosphoethanolamine (DSPE, C18). The results revealed that formulation with the shortest acyl chain (C14) were rapidly eliminated from the circulation. On the other hand, formulations with DPPE–PEG (C16) and DSPE–PEG (C18) showed a prolonged half-life for 4 and 24 h, respectively [59]. A complete protection is favorable for long-circulating liposome in the blood stream.

Shedding of the PEG coating is preferred once the LNPs are in the proximity of targeted cells, because the hydrophilic PEG coating tends to prevent the interaction between cell membrane and liposome, impeding the internalization of the LNPs. Grafting of a ligand on the distal end of PEG chain can circumvent this dilemma, facilitating internalization through receptor-mediated endocytosis. However, caution needs to be taken, as the distal end ligand can lead to protein adsorption, which speeds up the blood clearance especially when the ligand density is high [61]. After the LNPs are internalized into the cells through an endocytotic pathway, either the cationic lipid needs to interact with negatively charged endosome membrane or some fusogenic lipid, such as DOPE, needs to interact with and destabilize the endosome membrane structure. Under such circumstances, the presence of PEG chain might sterically hinder the interaction between the carrier lipids and endosomal membrane lipid, substantially reducing cytosolic release efficiency. Therefore, it is desirable that PEG coating is shed away to allow the interaction between fusogenic lipid and endosome membrane. Although the continuous shedding effect of PEG takes place after systemic delivery due to the sink condition in the blood, it is difficult to have a complete control over the shedding rate to make sure that shedding is complete when LNPs are entrapped in the endosome of the targeted cells. This kind of "passive" shedding demands extensive characterization of lipid-PEG conjugates and their interaction with the *in vivo* environment. Therefore, a series of stimulus-responsive-lipid-PEG have been developed to realize a controlled shedding of PEG. The idea is to introduce a linker between the lipid and PEG chain, which is cleavable under certain stimulus exclusively in endosome such as low pH [62–65], reducing agents [66], or enzymes [67,68]. Some linkers are designed to respond to extracellular environment of the tumor cells, such as low pH and the presence of matrix metalloproteinases [69]. Such rational designs of shedding process significantly promote the drug release and target cell interaction process.

## 3.4 FORMULATIONS OF THE LNP FOR NUCLEIC ACIDS ENCAPSULATION

The most conventional way of encapsulating nucleic acids with cationic liposome is to mix the two components to form a lipoplex, which however does not have much control over the structure and homogeneity of the complex. Various approaches have been investigated and developed to form the LNPs with desired structure so as to have more control over the performance of nanoparticles *in vivo*.

### 3.4.1 Self-Assembly

One of the most exploited strategies of self-assembly is to use a polycation to condense nucleic acid into anionic compact core complex, upon which cationic liposomes are coated on the surface to form a membrane-core structure. In the mid-1990s, the Huang laboratory used poly(L-lysine) and protamine [70] to precondense pDNA before incorporating the pDNA into the preformed cationic liposome

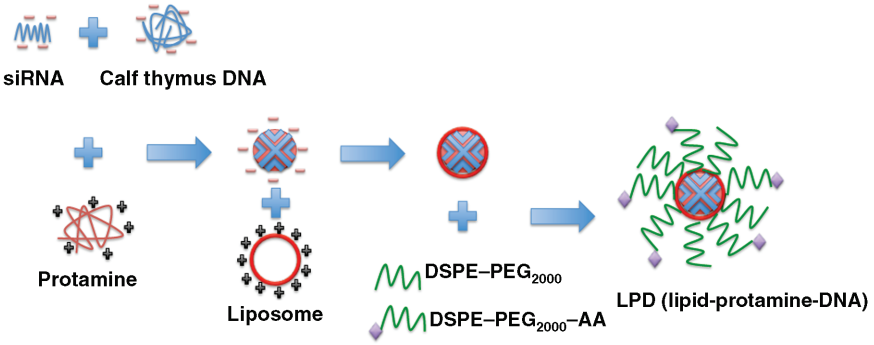


**FIGURE 3.4** Schematic illustration of LPD preparation. Briefly, anionic mix of siRNA and calf thymus DNA is condensed by protamine. The negatively charged condensates are then coated by cationic DOTAP–cholesterol liposome to form the core-membrane structure complex. The nanoparticle is PEGylated by incubation with DSPE–$PEG_{2000}$–anisamide (AA) and DSPE–$PEG_{2000}$ micelles.

containing 3β-[*N*-(*N*′,*N*′-dimethylaminoethane)-carbamoyl] cholesterol (DC-Chol) and DOPE. The membrane-core-structured LNPs were named as lipid-polycation-DNA (LPD) nanoparticles. The precondensation dramatically decreased the particle size of the pDNA–liposome complex and effectively protected DNA from nuclease degradation [71]. This formulation not only enhanced the transfection efficiency of the LNPs by up to 28-fold *in vitro* [71] but also demonstrated *in vivo* transfection capacity with minor modifications [70]. Later, Li et al. modified the formulation by coating the surface with DSPE–PEG and a sigma receptor targeting ligand DSPE–PEG–anisamide (Figure 3.4), which substantially increased *in vivo* siRNA delivery efficiency by fourfold and the gene silencing effect by two- to threefold [72]. Successful delivery of LPD to subcutaneous and lung metastatic tumors was believed to be due to reduced RES uptake by PEG coating, which resulted in higher tumor accumulation because of the EPR effect [73,74]. It is worth noticing that the negatively charged protamine/calf thymus DNA/siRNA core played a critical role in supporting the lipid bilayer, allowing more surfactant-like DSPE–PEG conjugates to be inserted in the membrane without ripping off the lipid membrane [75]. This high density of sheddable PEG significantly improved the PK of LPD nanoparticles in the blood stream [76].

With a similar approach, Harashima et al. have devised a series of lipid based core-shell structured nanocarriers, namely, multifunctional envelope-type nano-device (MEND) [68,77–93]. In these formulations, stearyl-octaarginine and cholesteryl-GALA(WEAALAEALAEALAEHLAEALAEALEALAA) were incorporated into the preformed lipid membrane mixed with DOPE and cholesteryl hemisuccinate (CHEMS) [81,83,84,90,91]. Slightly different from the LPD formulations, the pDNA/protamine or PEI condensates in water are added to the blown lipid film. The hydration of lipid membrane initiates the formation of liposome, which coats the condensates subsequently. The final nanoparticle is sonicated to ensure homogeneity. MEND has been employed to deliver genes to

the liver [83] or the lung [82] with the surface modification of octaarginine or IRQ peptide (IRQRRRR). The octaarginine or IRQ peptide enabled the internalization of MEND nanoparticles through macropinocytosis and caveolar endocytosis, respectively. The GALA peptide on the surface was also used as a pH-responsive fusogenic peptide that destabilizes membrane structure at acidic pH in the endosome for cytoplasmic release of the cargo.

### 3.4.2 Detergent Dialysis

As an alternative to self-assembly, which usually involves preformed liposomes, a detergent dialysis method is used for lipid-based nanocarriers to encapsulate negatively charged nucleic acid [94]. The idea is to use nonionic detergent with a high critical micelle concentration to solubilize amphiphilic lipid molecule in aqueous solution in the presence of the nucleic acid so that nucleic acids were bound to free lipid instead of liposome. As the detergent is removed gradually by dialysis, cationic lipid and the associated nucleic acid are rearranged and form a vesicle structure that entraps the nucleic acid inside the liposomes. The name, stabilized plasmid-lipid particle (SPLP), is well recognized and represents a class of formulated LNPs prepared by the detergent dialysis method. This method forms 100 nm liposomes with higher trapping efficiency (60–70%) than those achieved by conventional methods such as reverse evaporation phase, extrusion, or dehydration-rehydration, which also usually result in large particle size due to aggregation [59,95].

With a similar idea, Jeffs et al. have developed a new procedure-spontaneous vesicle formation (SVF) to encapsulate pDNA [96]. A lipid solution containing cholesterol, distearoylphosphatidylcholine (DSPC), 1,2-dioleyloxy-*N*,*N*-dimethylaminopropane (DODMA), and 3-*O*-(2′(-methoxypolyethyleneglycol)$_{2000}$)-1,2-distearoyl-*sn*-glycerol (PEG–S-DSG) in 90% ethanol was used to mix with pDNA in the buffer. The mixing procedure resulted in the drop of ethanol concentration, which was insufficient to support lipid solubility. A lipid vesicle was subsequently formed to encapsulate pDNA. The pDNA-containing liposome was further diluted with buffer to further stabilize the metastable vesicles, resulting in a significant increase in DNA encapsulation from 60% to 90% [96]. The employment of a T-shaped mixing chamber ensured the reproducibility of SPLP, which showed the potential for scale-up manufacture for clinical studies.

Morrissey et al. adapted the detergent dialysis method to encapsulate chemically stabilized siRNA with a superb efficiency of 93% [97]. The specialized liposome, namely, stable nucleic-acid-lipid particle (SNALP), was composed of DSPC, cholesterol, PEG–C-DMA, and ionizable lipid DLinDMA. The LNP is formed with similar nucleic acid to lipid ratio and mean particle size (140 nm) regardless of the nucleic acid payload. The *in vivo* knockdown efficiency was evaluated in a mouse model of hepatitis B virus replication. The formulation drastically reduced the effective dose of siRNA delivery to 3 mg/kg/day and demonstrated a persistent knockdown effect up to 6 weeks with a weekly injection. Later, Zimmermann et al. has examined the siRNA delivery efficiency of SNALP by knocking down disease target apolipoprotein B (ApoB) in the liver of nonhuman primates. It was

demonstrated that a single injection of 2.5 mg/kg resulted in a sustained 80% silencing of ApoB mRNA expression in the liver till 11 days after injection [36]. The degree and duration of the gene-silencing effect was way beyond the data acquired from rodent animal studies. The outstanding effects may be attributed to species differences with respect to turnover time of RNAi. The report supported the clinical potential to realize RNAi therapies with the aid of SNALP delivery system.

### 3.4.3 Nanometric Calcium Phosphate/siRNA Precipitation

Co-precipitates formed between calcium phosphate and DNA have been used as an efficient method to transfect cultured mammalian cells to express exogenous genes since it was first reported by Graham et al. [98]. However, the application of this technique has been limited to *in vitro* transfection as a result of uncontrollable size and batch-to-batch variability of transfection efficiency [99]. Li and Huang have harnessed the low toxicity and pH-induced degradability of the co-precipitates and formulated a nano-sized calcium phosphate particles coated with liposomes for siRNA systemic delivery, namely, lipid–calcium–phosphate (LCP) [100,101]. The calcium phosphate particle is precipitated in an aqueous phase of a water/oil (cyclohexane) reverse microemulsion system stabilized with surfactants (IGPAL-520 and DOPA) (Figure 3.5). Size of the formed calcium phosphate amorphous precipitate is well controlled, as the surfactants on the surface prevent aggregation of the precipitates. siRNA is entrapped within the nanoprecipitate by the interaction between phosphate groups of siRNA and calcium ions. The lipid monolayer of DOPA is stabilized on the surface of the nanoprecipitate as a result of interactions

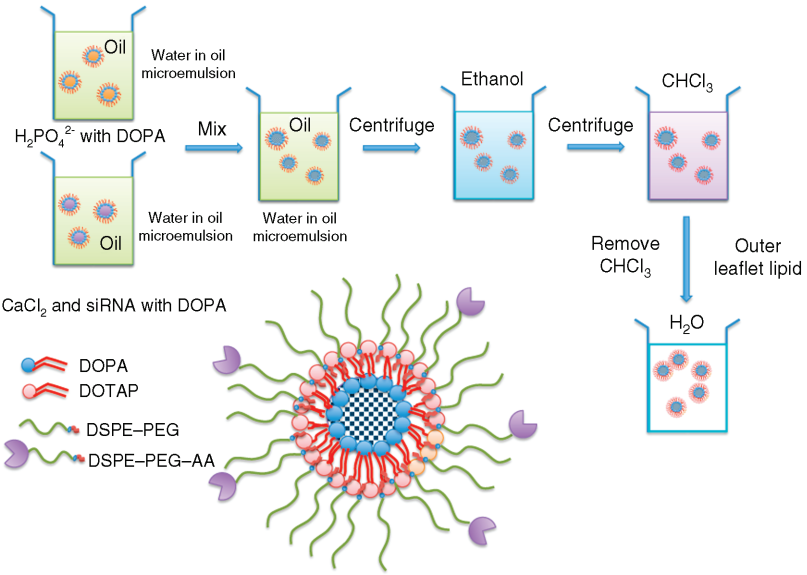


**FIGURE 3.5** Preparation of liposome–calcium–phosphate (LCP) nanoparticles.

between the phosphate groups of DOPA and calcium ions. IGEPA-520 is subsequently removed by repeated centrifugation and washing with ethanol. The final LCP is formed by incubating DOPA-stabilized calcium phosphate nanoprecipitates with outer leaflet lipids in an aqueous solution [100,101].

The LCP demonstrates a pH-responsive siRNA release profile. As the pH in the endosome starts to drop, calcium phosphate precipitates dissolve and release cargo. Meanwhile, the released calcium ions increase the osmotic pressure of endosomal compartments, destabilize the endosomal membrane, and facilitate the release of siRNA into the cytosol. This property substantially enhances delivery efficiency of the LCP. The *in vitro* half maximal inhibitory concentration $IC_{50}$ was determined to be 60 nM, which was three to four times lower than that of LPD (200 nM) [100,101]. The LCP nanoparticle has been demonstrated to knockdown the luciferase activity in a B16F10 lung metastasis model with an $ED_{50}$ of 0.06 mg/ml. The co-delivery of VEGF, MDM2, and c-myc siRNA formulated in LCP nanoparticle significantly suppressed the lung metastasis (~70%) with a relatively low dose (0.36 mg/kg).

## 3.5 ORGAN-SPECIFIC DELIVERY OF NUCLEIC ACIDS WITH LNPs

The advancement of nanoparticles has triggered the site-specific delivery of nucleic acid based therapeutics. Moreover, owing to the different anatomical structures and physiological functions of major organs, particularly the endothelium status and the RES, various strategies harnessing the organ-specific features have been developed so as to specifically deliver nucleic acid loaded LNPs into targeted organs.

### 3.5.1 Brain-Targeted Delivery

The brain is highly protected from the periphery by two major barriers, the blood–brain barrier (BBB) and blood–cerebrospinal fluid barrier (BCSFB). The former barrier is primarily responsible for the maintenance of homeostasis of the central nervous system (CNS). The BBB is composed of nonfenestrated microvessel endothelial cells with continuous tight junctions that limit the transport of macrophages, xenobiotics, and some other endogenous metabolites [102]. In addition to the physical barrier reflected by the tight junction, there is a metabolism-driven barrier that is manifested by the expression of several receptors, ion channels, and influx/efflux transport proteins [103]. Therefore, the BBB has been the major barrier that compromises the efficacy of most therapeutics including macromolecular drugs.

Of many brain diseases, primary malignant brain tumors are the most lethal ones. As the primary tumors or metastases begin to grow beyond 1–2 mm in diameter within the brain parenchyma, the BBB is compromised both structurally and functionally with features of highly tortuous, disorganized, and permeable vessels [104]. These features are not different from tumors arising elsewhere and could be exploited for the delivery of therapeutics based on the EPR effect. Nevertheless, for other brain diseases such as neurodegenerative disorders or

ischemic stroke, the BBB still remains intact and is able to block the pharmacological reagents from crossing the endothelium to access the target site. The advancement of nanotechnology has tackled the problem with great success for CNS drug delivery since the nanocarriers could selectively interact with endothelial microvessels after modification with targeting moieties and result in a high local concentration of therapeutics in the brain parenchyma [103]. The LNPs are probably the most investigated and clinically tested due to the long circulation time and low toxicity [105]. The major strategies for the development of LNPs to cross the BBB is through transcytosis, initially observed as a mechanism by which iron-loaded transferrin is transported to the parenchyma of the brain [106]. Transferrin or antibodies against transferrin receptor have been extensively used for the surface modification of these LNPs for brain delivery [107–113].

The most outstanding LNP that has been well studied and applied to various brain disease models is the Trojan horse liposome (THL), developed by Pardridge et al. [108]. The surface modifications with targeting ligands for insulin receptor or transferrin receptor expressed on the surface of BBB enable selective transport of the LNP from the luminal side of the endothelium to the brain parenchyma. Additional grafting of neuron-specific ligands facilitates the nanoparticle uptake via receptor-mediated endocytosis. The delivery efficiency of the LNPs has been demonstrated in gene therapy for the treatment of Parkinson's disease model in rats [109,114,115], VII mucopolysaccharidosis model in mice [111], and RNAi therapy targeting epidermal growth factor receptor with shRNA in mouse glioblastoma [116].

An alternative strategy to enhance brain delivery is to coat the nanoparticle with cell-penetrating peptides such as TAT and octaarginine (R8) since these peptides increase the selectivity of BBB and promote drug trafficking [117]. Although the mechanism of action was not well elucidated, it was observed that the surface charge of the particle played an important role in the brain delivery [117,118]. This implies that electrostatic interaction between the positively charged particle and the lectins on the microvessel endothelial cells might trigger the endocytosis.

### 3.5.2 Liver-Targeted Delivery

Among various organs, the liver has drawn most attention due to its critical role in the metabolism. Moreover, the liver has a unique sinusoidal endothelium, which processes open pores or fenestrae lacking a diaphragm and a basal lamina underneath the endothelial lining with the pore size around 100 nm in diameter [119]. This property of the well-perfused organ makes it easy for the LNPs to extravasate to reach the interstitial space of the liver tissue. However, Kupffer cells, the macrophages of the liver, lay in ambush in the endothelium to capture and destroy foreign particles. Any successful nanoparticle drug delivery vectors will have to evade the Kupffer cells before accessing other cells of the liver. In this regard, a dense PEG coating usually disguises the LNPs from being recognized by the RES. The presence of a hepatocyte-specific targeting ligand diverts the LNPs from highly phagocytic cells and enhances the uptake by hepatocytes. Since hepatocytes uniquely express

asialoglycoprotein receptor on the cell surface, galactose is widely exploited as a ligand for liver-specific gene transfer [120–122].

Another strategy to direct LNPs to hepatocytes is to take advantage of apolipoprotein (Apo), a component of high-density lipoprotein (HDL) that guides the transport of lipids to the liver for processing as a natural ligand for the scavenger receptor on hepatocytes. The hydrophobic domain in the Apo prompts the interaction between the Apo and the LNPs, essentially forming an artificial HDL. The incorporation of the Apo as a targeting ligand makes the LNPs predominantly taken up by the liver cells with well-characterized mechanism [123,124]. LNPs have proven to be highly efficient carriers for the liver delivery. This is probably because the circulating LNPs could interact with serum proteins and acquire protein components that guide the transport of the LNPs to specific cells [125,126]. Apolipoprotein E (ApoE), shown to be capable of enhancing the uptake into hepatocytes through low-density lipoprotein receptor (LDLR), was exploited as an "endogenous ligand" and an alternative to an exogenous targeting ligand [127]. The functionality of the proposed endogenous ligand was examined by Akinc et al. in LDLR−/− mice and ApoE−/− mice, respectively. It was demonstrated that the hepatocyte-specific delivery of the neutral liposome is ApoE mediated and LDLR dependent [127].

In addition to the aforementioned rational-based design of LNPs for the liver delivery, some molecules, which have demonstrated liver-specific properties, are also explored for the modification of LNPs for liver specificity. It was observed that R8-peptide-modified LNPs largely accumulated in the liver after intravenous administration and transfected hepatocytes with high efficiency. The uptake was proven to be R8 density dependent [128,129]. The liver-specific biodistribution is likely attributable to a unique property of R8 given that this type of peptide represents a promising device for specific liver delivery [130].

### 3.5.3 Lung-Targeted Delivery

Owing to the anatomical structure of the lung, rich in microcapillaries, gene transfer to the lung could be performed through systemic administration similar to other organs. Genetic therapeutics could also be administered noninvasively via airways. Because of the limited scope of this review, we focus the discussion mainly on the former route. The LNPs with cationic surface charge will preferably accumulate in the lung tissue without any modification since there is an excellent chance for them to interact with microcapillary endothelium and trigger the endocytosis [131]. Although the non-PEGylated LNPs are more likely to be opsonized and taken up by the macrophages in the liver and the spleen, it still results in a high level of transgene expression in the lung [131]. Recent studies have revealed that various tissue disorders are related to endothelium, particularly tumor angiogenesis. Therefore, the endothelium could serve as a therapeutic target for primary or metastatic lung tumors. Surface modifications with PEG and a tumor-specific targeting ligand can enhance the accumulation of liposome in the interstitial space of solid tumor and significantly increase the uptake of liposome by cancer cells [74,132,133].

### 3.5.4 Kidney-Targeted Delivery

Although the prevention and treatment of renal diseases is highly demanding, nonviral gene therapists have not paid as much attention to the kidney as to the liver and the brain. For one of the reasons, the kidney is not an organ that has the large accumulation of nanoparticles after systemic administration. Moreover, the signal from the kidney in the biodistribution study was often proven to be the fragments of dissociated nanoparticle, which was disposed through the glomerular filtration. Lai et al. tried to use Lipofectin formulation (DOTMA–DOPE) to deliver β-galactosidase gene to the kidney via intra-renal-arterial injection. The expression of the reporter gene was detected in the cortex and the outer medulla area and exclusively observed in the tubular epithelial cells [134]. Ito et al. used the same formulation to introduce inducible nitric oxide synthase (iNOS) gene into the kidney via intraureteral injection. The iNOS expression was detected in the collecting ducts, distal tubules, and glomerulus in the injected kidney and improved the renal function in unilateral ureteral obstruction [135]. Nevertheless, most of the kidney gene therapy has been conducted by viral vectors for high specificity and transgene expression.

## 3.6 CURRENT CHALLENGES TO THE DEVELOPMENT OF LNPs

The term *targeted gene delivery* indicated a predominant accumulation of genetic therapeutics in the intended tissues or even in the specific cell populations. This idea was originally conceived by Paul Ehrlich as the "magic bullet" that goes straight to the target cells to deliver cargos after systemic administration. The LNPs at the current stage, regardless of materials and formulations, are far from the ideal "magic bullet" because the biodistribution of most nanoparticles relies on both the properties of the nanoparticles and the vascular endothelium status of the organs or tumors. Although PEGylation could significantly improve the PK profiles of nanoparticles, drastically reducing the liver and spleen accumulations by the RES uptake and increasing the blood circulation time, the tumor accumulation of LNPs is reported to be around 5–10% of the injected dose. It has been noticed that the accumulation would not increase even if there were excessive LNPs circulating in the bloodstream. This suggested that a saturation state existed for the EPR-mediated tumor accumulation. Maeda et al. have investigated factors that affect tumor vasculature permeability and described a bradykinin-mediated pathway to account for the EPR effect [136]. Some intermediate molecules in the pathway such as bradykinin or NOS were targeted for the modulation of vascular permeability to enhance the EPR effect [136]. Nevertheless, there was only marginal increase in the tumor accumulation. Strictly speaking, there is no passive "targeting" if more than 90% of the administered dose goes elsewhere other than the targeted site. Furthermore, the degree of the leakiness of tumor vasculature is highly variable and heterogeneous even for the same type of xenograft implanted at the same site [137]. So far, the EPR effect has been the major driving force for nanoparticle delivery to

solid tumors. It is imperative that some alternative mechanisms be exploited in order to enhance tumor accumulation of nanoparticles to meet the therapeutic challenge.

## 3.7 FUTURE PERSPECTIVES

The lack of satisfactory gene delivery systems impedes the application of the therapy in the medical practice. Liposome-mediated gene delivery appears to be most feasible and reliable compared with other nonviral vectors. This is by and large due to their relatively high rate of gene transfer efficiency, biodegradability, and cost-effectiveness. The emergence of ionizable catonic lipids, such as conditional catonic lipids, has revealed its translational potential. Apparently, the development of this class of lipids has been the focus of the research due to the promising therapeutic activity in animal disease models. Meanwhile, we begin to understand the factors determining the delivery efficiency such as the status of nanoparticles in the blood circulation and the impact of the major organs on the nanoparticles due to their anatomical structure and physiological function, tumor spatial and temporal heterogeneity, and conditions of tumor microenvironment. With this knowledge in hand, hundreds of rationally designed molecules including lipids and other macromolecules have been developed materialwise to improve the PK profile of the LNPs and transfection efficiency. All components are expected to contribute to the delivery efficiency according to their designs. Despite that these sophisticated components work wonderfully on their own, there was hardly any synergistic effect observed when combined together. Thus, in assembling all the components, it is important to ascertain that an optimization of formulation is conducted to achieve better delivery efficiency. It is envisaged that with the advent of novel lipids and formulation procedures, the seemingly insurmountable hurdle of gene therapy will be overcome in due course.

### Acknowledgments

The original work in authors' laboratory has been supported by NIH grants CA129835, CA149363, CA151652, and CA151455.

## REFERENCES

1. Mulligan, R. C. (1993). The basic science of gene therapy. *Science 260*, 926–932.
2. Fire, A., Xu, S., Montgomery, M. K., Kostas, S. A., Driver, S. E., Mello, C. C. (1998). Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature 391*, 806–811.
3. Heller, L. C., Ugen, K., Heller, R. (2005). Electroporation for targeted gene transfer. *Expert Opinion on Drug Delivery 2*, 255–268.
4. Knapp, J. E., Liu, D. (2004). Hydrodynamic delivery of DNA. *Methods in Molecular Biology 245*, 245–250.
5. Thomas, C. E., Ehrhardt, A., Kay, M. A. (2003). Progress and problems with the use of viral vectors for gene therapy. *Nature Reviews Genetics 4*, 346–358.

6. Cavazzana-Calvo, M., Hacein-Bey, S., de Saint Basile, G., Gross, F., Yvon, E., Nusbaum, P., Selz, F., Hue, C., Certain, S., Casanova, J. L., Bousso, P., Deist, F. L., Fischer, A. (2000). Gene therapy of human severe combined immunodeficiency (SCID)-X1 disease. *Science 288*, 669–672.
7. Kaiser, J. (2003). Gene therapy. Seeking the cause of induced leukemias in X-SCID trial. *Science 299*, 495.
8. Felgner, P. L., Gadek, T. R., Holm, M., Roman, R., Chan, H. W., Wenz, M., Northrop, J. P., Ringold, G. M., Danielsen, M. (1987). Lipofection: a highly efficient, lipid-mediated DNA-transfection procedure. *Proceedings of the National Academy of Sciences of the United States of America 84*, 7413–7417.
9. Radler, J. O., Koltover, I., Salditt, T., Safinya, C. R. (1997). Structure of DNA-cationic liposome complexes: DNA intercalation in multilamellar membranes in distinct interhelical packing regimes. *Science 275*, 810–814.
10. Vonarbourg, A., Passirani, C., Saulnier, P., Benoit, J. P. (2006). Parameters influencing the stealthiness of colloidal drug delivery systems. *Biomaterials 27*, 4356–4373.
11. Klibanov, A. L., Maruyama, K., Torchilin, V. P., Huang, L. (1990). Amphipathic polyethyleneglycols effectively prolong the circulation time of liposomes. *FEBS Letters 268*, 235–237.
12. Hobbs, S. K., Monsky, W. L., Yuan, F., Roberts, W. G., Griffith, L., Torchilin, V. P., Jain, R. K. (1998). Regulation of transport pathways in tumor vessels: role of tumor type and microenvironment. *Proceedings of the National Academy of Sciences of the United States of America 95*, 4607–4612.
13. Whitehead, K. A., Langer, R., Anderson, D. G. (2009). Knocking down barriers: advances in siRNA delivery. *Nature reviews Drug Discovery 8*, 129–138.
14. Maeda, H. (2010). Tumor-selective delivery of macromolecular drugs via the EPR effect: background and future prospects. *Bioconjugate Chemistry 21*, 797–802.
15. Matsumura, Y., Maeda, H. (1986). A new concept for macromolecular therapeutics in cancer chemotherapy: mechanism of tumoritropic accumulation of proteins and the antitumor agent smancs. *Cancer Research 46*, 6387–6392.
16. Baxter, L. T., Jain, R. K. (1989). Transport of fluid and macromolecules in tumors. I. Role of interstitial pressure and convection. *Microvascular Research 37*, 77–104.
17. Rejman, J., Oberle, V., Zuhorn, I. S., Hoekstra, D. (2004). Size-dependent internalization of particles via the pathways of clathrin- and caveolae-mediated endocytosis. *The Biochemical Journal 377*, 159–169.
18. Maxfield, F. R., McGraw, T. E. (2004). Endocytic recycling. *Nature Reviews Molecular Cell Biology 5*, 121–132.
19. Tseng, Y. C., Mozumdar, S., Huang, L. (2009). Lipid-based systemic delivery of siRNA. *Advanced Drug Delivery Reviews 61*, 721–731.
20. Goodsell, D. S. (1991). Inside a living cell. *Trends in Biochemical Sciences 16*, 203–206.
21. Luby-Phelps, K. (2000). Cytoarchitecture and physical properties of cytoplasm: volume, viscosity, diffusion, intracellular surface area. *International Review of Cytology 192*, 189–221.
22. Leopold, P. L., Kreitzer, G., Miyazawa, N., Rempel, S., Pfister, K. K., Rodriguez-Boulan, E., Crystal, R. G. (2000). Dynein- and microtubule-mediated translocation of adenovirus serotype 5 occurs after endosomal lysis. *Human Gene Therapy 11*, 151–165.

23. Suh, J., Wirtz, D., Hanes, J. (2004). Real-time intracellular transport of gene nanocarriers studied by multiple particle tracking. *Biotechnology Progress 20*, 598–602.
24. Pichon, C., Billiet, L., Midoux, P. (2010). Chemical vectors for gene delivery: uptake and intracellular trafficking. *Current Opinion in Biotechnology 21*, 640–645.
25. Midoux, P., Pichon, C., Yaouanc, J. J., Jaffres, P. A. (2009). Chemical vectors for gene delivery: a current review on polymers, peptides and lipids containing histidine or imidazole as nucleic acids carriers. *British Journal of Pharmacology 157*, 166–178.
26. Aronsohn, A. I., Hughes, J. A. (1998). Nuclear localization signal peptides enhance cationic liposome-mediated gene therapy. *Journal of Drug Targeting 5*, 163–169.
27. Subramanian, A., Ranganathan, P., Diamond, S. L. (1999). Nuclear targeting peptide scaffolds for lipofection of nondividing mammalian cells. *Nature Biotechnology 17*, 873–877.
28. Matulis, D., Rouzina, I., Bloomfield, V. A. (2002). Thermodynamics of cationic lipid binding to DNA and DNA condensation: roles of electrostatics and hydrophobicity. *Journal of the American Chemical Society 124*, 7331–7342.
29. Morille, M., Passirani, C., Vonarbourg, A., Clavreul, A., Benoit, J. P. (2008). Progress in developing cationic vectors for non-viral systemic gene therapy against cancer. *Biomaterials 29*, 3477–3496.
30. Alexander, M. Y., Akhurst, R. J. (1995). Liposome-medicated gene transfer and expression via the skin. *Human Molecular Genetics 4*, 2279–2285.
31. Remy, J. S., Sirlin, C., Vierling, P., Behr, J. P. (1994). Gene transfer with a series of lipophilic DNA-binding molecules. *Bioconjugate Chemistry 5*, 647–654.
32. Zhang, S., Xu, Y., Wang, B., Qiao, W., Liu, D., Li, Z. (2004). Cationic compounds used in lipoplexes and polyplexes for gene delivery. *Journal of Controlled Release 100*, 165–180.
33. Gao, X., Huang, L. (1991). A novel cationic liposome reagent for efficient transfection of mammalian cells. *Biochemical and Biophysical Research Communications 179*, 280–285.
34. Vigneron, J. P., Oudrhiri, N., Fauquet, M., Vergely, L., Bradley, J. C., Basseville, M., Lehn, P., Lehn, J. M. (1996). Guanidinium-cholesterol cationic lipids: efficient vectors for the transfection of eukaryotic cells. *Proceedings of the National Academy of Sciences of the United States of America 93*, 9682–9686.
35. Semple, S. C., Klimuk, S. K., Harasym, T. O., Dos Santos, N., Ansell, S. M., Wong, K. F., Maurer, N., Stark, H., Cullis, P. R., Hope, M. J., Scherrer, P. (2001). Efficient encapsulation of antisense oligonucleotides in lipid vesicles using ionizable aminolipids: formation of novel small multilamellar vesicle structures. *Biochimica et Biophysica Acta 1510*, 152–166.
36. Zimmermann, T. S., Lee, A. C., Akinc, A., Bramlage, B., Bumcrot, D., Fedoruk, M. N., Harborth, J., Heyes, J. A., Jeffs, L. B., John, M., Judge, A. D., Lam, K., McClintock, K., Nechev, L. V., Palmer, L. R., Racie, T., Rohl, I., Seiffert, S., Shanmugam, S., Sood, V., Soutschek, J., Toudjarska, I., Wheat, A. J., Yaworski, E., Zedalis, W., Koteliansky, V., Manoharan, M., Vornlocher, H. P., MacLachlan, I. (2006). RNAi-mediated gene silencing in non-human primates. *Nature 441*, 111–114.
37. Semple, S. C., Akinc, A., Chen, J., Sandhu, A. P., Mui, B. L., Cho, C. K., Sah, D. W., Stebbing, D., Crosley, E. J., Yaworski, E., Hafez, I. M., Dorkin, J. R., Qin, J., Lam, K., Rajeev, K. G., Wong, K. F., Jeffs, L. B., Nechev, L., Eisenhardt, M. L., Jayaraman, M.,

Kazem, M., Maier, M. A., Srinivasulu, M., Weinstein, M. J., Chen, Q., Alvarez, R., Barros, S. A., De, S., Klimuk, S. K., Borland, T., Kosovrasti, V., Cantley, W. L., Tam, Y. K., Manoharan, M., Ciufolini, M. A., Tracy, M. A., de Fougerolles, A., MacLachlan, I., Cullis, P. R., Madden, T. D., Hope, M. J. (2010). Rational design of cationic lipids for siRNA delivery. *Nature Biotechnology 28*, 172–176.

38. Akinc, A., Zumbuehl, A., Goldberg, M., Leshchiner, E. S., Busini, V., Hossain, N., Bacallado, S. A., Nguyen, D. N., Fuller, J., Alvarez, R., Borodovsky, A., Borland, T., Constien, R., de Fougerolles, A., Dorkin, J. R., Narayanannair Jayaprakash, K., Jayaraman, M., John, M., Koteliansky, V., Manoharan, M., Nechev, L., Qin, J., Racie, T., Raitcheva, D., Rajeev, K. G., Sah, D. W., Soutschek, J., Toudjarska, I., Vornlocher, H. P., Zimmermann, T. S., Langer, R., Anderson, D. G. (2008). A combinatorial library of lipid-like materials for delivery of RNAi therapeutics. *Nature Biotechnology 26*, 561–569.
39. Heyes, J., Palmer, L., Bremner, K., MacLachlan, I. (2005). Cationic lipid saturation influences intracellular delivery of encapsulated nucleic acids. *Journal of Controlled Release 107*, 276–287.
40. Landen, C. N., Jr., Chavez-Reyes, A., Bucana, C., Schmandt, R., Deavers, M. T., Lopez-Berestein, G., Sood, A. K. (2005). Therapeutic EphA2 gene targeting *in vivo* using neutral liposomal small interfering RNA delivery. *Cancer Research 65*, 6910–6918.
41. Gray, M. J., Dallas, N. A., Van Buren, G., Xia, L., Yang, A. D., Somcio, R. J., Gaur, P., Mangala, L. S., Vivas-Mejia, P. E., Fan, F., Sanguino, A. M., Gallick, G. E., Lopez-Berestein, G., Sood, A. K., Ellis, L. M. (2008). Therapeutic targeting of Id2 reduces growth of human colorectal carcinoma in the murine liver. *Oncogene 27*, 7192–7200.
42. Villares, G. J., Zigler, M., Wang, H., Melnikova, V. O., Wu, H., Friedman, R., Leslie, M. C., Vivas-Mejia, P. E., Lopez-Berestein, G., Sood, A. K., Bar-Eli, M. (2008). Targeting melanoma growth and metastasis with systemic delivery of liposome-incorporated protease-activated receptor-1 small interfering RNA. *Cancer Research 68*, 9078–9086.
43. Landen, C. N., Merritt, W. M., Mangala, L. S., Sanguino, A. M., Bucana, C., Lu, C., Lin, Y. G., Han, L. Y., Kamat, A. A., Schmandt, R., Coleman, R. L., Gershenson, D. M., Lopez-Berestein, G., Sood, A. K. (2006). Intraperitoneal delivery of liposomal siRNA for therapy of advanced ovarian cancer. *Cancer Biology & Therapy 5*, 1708–1713.
44. Mangala, L. S., Han, H. D., Lopez-Berestein, G., Sood, A. K. (2009). Liposomal siRNA for ovarian cancer. *Methods in Molecular Biology 555*, 29–42.
45. Merritt, W. M., Lin, Y. G., Spannuth, W. A., Fletcher, M. S., Kamat, A. A., Han, L. Y., Landen, C. N., Jennings, N., De Geest, K., Langley, R. R., Villares, G., Sanguino, A., Lutgendorf, S. K., Lopez-Berestein, G., Bar-Eli, M. M., Sood, A. K. (2008). Effect of interleukin-8 gene silencing with liposome-encapsulated small interfering RNA on ovarian cancer cell growth. *Journal of the National Cancer Institute 100*, 359–372.
46. Leblond, J., Mignet, N., Leseurre, L., Largeau, C., Bessodes, M., Scherman, D., Herscovici, J. (2006). Design, synthesis, and evaluation of enhanced DNA binding new lipopolythioureas. *Bioconjugate Chemistry 17*, 1200–1208.
47. Leblond, J., Mignet, N., Largeau, C., Seguin, J., Scherman, D., Herscovici, J. (2008). Lipopolythiourea transfecting agents: lysine thiourea derivatives. *Bioconjugate Chemistry 19*, 306–314.

48. Breton, M., Leblond, J., Seguin, J., Midoux, P., Scherman, D., Herscovici, J., Pichon, C., Mignet, N. (2010). Comparative gene transfer between cationic and thiourea lipoplexes. *The Journal of Gene Medicine 12*, 45–54.
49. Lee, R. J., Huang, L. (1996). Folate-targeted, anionic liposome-entrapped polylysine-condensed DNA for tumor cell-specific gene transfer. *The Journal of Biological Chemistry 271*, 8481–8487.
50. Mastrobattista, E., Kapel, R. H., Eggenhuisen, M. H., Roholl, P. J., Crommelin, D. J., Hennink, W. E., Storm, G. (2001). Lipid-coated polyplexes for targeted gene delivery to ovarian carcinoma cells. *Cancer Gene Therapy 8*, 405–413.
51. Sun, X., Provoda, C., Lee, K. D. (2010). Enhanced *in vivo* gene expression mediated by listeriolysin O incorporated anionic LPDII: Its utility in cytotoxic T lymphocyte-inducing DNA vaccine. *Journal of Controlled Release 148*, 219–225.
52. Chen, Y., Bathula, S. R., Li, J., Huang, L. (2010). Multifunctional nanoparticles delivering small interfering RNA and doxorubicin overcome drug resistance in cancer. *The Journal of Biological Chemistry 285*, 22639–22650.
53. Srinivasan, C., Burgess, D. J. (2009). Optimization and characterization of anionic lipoplexes for gene delivery. *Journal of Controlled Release 136*, 62–70.
54. Allen, T. M. (1994). The use of glycolipids and hydrophilic polymers in avoiding rapid uptake of liposomes by the mononuclear phagocyte system. *Advanced Drug Delivery Reviews 13*, 285–309.
55. Campbell, R. B., Fukumura, D., Brown, E. B., Mazzola, L. M., Izumi, Y., Jain, R. K., Torchilin, V. P., Munn, L. L. (2002). Cationic charge determines the distribution of liposomes between the vascular and extravascular compartments of tumors. *Cancer Research 62*, 6831–6836.
56. Moghimi, S. M., Hunter, A. C. (2001). Capture of stealth nanoparticles by the body's defences. *Critical Reviews in Therapeutic Drug Carrier Systems 18*, 527–550.
57. Mok, K. W., Lam, A. M., Cullis, P. R. (1999). Stabilized plasmid-lipid particles: factors influencing plasmid entrapment and transfection properties. *Biochimica et Biophysica Acta 1419*, 137–150.
58. Tam, P., Monck, M., Lee, D., Ludkovski, O., Leng, E. C., Clow, K., Stark, H., Scherrer, P., Graham, R. W., Cullis, P. R. (2000). Stabilized plasmid-lipid particles for systemic gene therapy. *Gene Therapy 7*, 1867–1874.
59. Adlakha-Hutcheon, G., Bally, M. B., Shew, C. R., and Madden, T. D. (1999). Controlled destabilization of a liposomal drug delivery system enhances mitoxantrone antitumor activity. *Nat Biotechnol 17*, 775–779.
60. Zhang, Y. P., Sekirov, L., Saravolac, E. G., Wheeler, J. J., Tardi, P., Clow, K., Leng, E., Sun, R., Cullis, P. R., Scherrer, P. (1999). Stabilized plasmid-lipid particles for regional gene therapy: formulation and transfection properties. *Gene Therapy 6*, 1438–1447.
61. Allen, T. M., Brandeis, E., Hansen, C. B., Kao, G. Y., Zalipsky, S. (1995). A new strategy for attachment of antibodies to sterically stabilized liposomes resulting in efficient targeting to cancer cells. *Biochimica et Biophysica Acta 1237*, 99–108.
62. Guo, X., Szoka, F. C. Jr. (2001). Steric stabilization of fusogenic liposomes by a low-pH sensitive PEG–diortho ester–lipid conjugate. *Bioconjugate Chemistry 12*, 291–300.
63. Li, W., Huang, Z., MacKay, J. A., Grube, S., Szoka, F. C. Jr. (2005). Low-pH-sensitive poly(ethylene glycol) (PEG)-stabilized plasmid nanolipoparticles: effects of PEG chain

length, lipid composition and assembly conditions on gene delivery. *The Journal of Gene Medicine 7*, 67–79.

64. Choi, J. S., MacKay, J. A., Szoka, F. C. Jr. (2003). Low-pH-sensitive PEG-stabilized plasmid-lipid nanoparticles: preparation and characterization. *Bioconjugate Chemistry 14*, 420–429.
65. Guo, X., MacKay, J. A., Szoka, F. C. Jr. (2003). Mechanism of pH-triggered collapse of phosphatidylethanolamine liposomes stabilized by an ortho ester polyethyleneglycol lipid. *Biophysical Journal 84*, 1784–1795.
66. Maeda, T., Fujimoto, K. (2006). A reduction-triggered delivery by a liposomal carrier possessing membrane-permeable ligands and a detachable coating. *Colloids and Surfaces B: Biointerfaces 49*, 15–21.
67. Zhang, J. X., Zalipsky, S., Mullah, N., Pechar, M., Allen, T. M. (2004). Pharmaco attributes of dioleoylphosphatidylethanolamine/cholesterylhemisuccinate liposomes containing different types of cleavable lipopolymers. *Pharmacological Research 49*, 185–198.
68. Hatakeyama, H., Akita, H., Kogure, K., Oishi, M., Nagasaki, Y., Kihira, Y., Ueno, M., Kobayashi, H., Kikuchi, H., Harashima, H. (2007). Development of a novel systemic gene delivery system for cancer therapy with a tumor-specific cleavable PEG-lipid. *Gene Therapy 14*, 68–77.
69. Romberg, B., Hennink, W. E., Storm, G. (2008). Sheddable coatings for long-circulating nanoparticles. *Pharmaceutical Research 25*, 55–71.
70. Li, S., Huang, L. (1997). *In vivo* gene transfer via intravenous administration of cationic lipid-protamine-DNA (LPD) complexes. *Gene Therapy 4*, 891–900.
71. Gao, X., Huang, L. (1996). Potentiation of cationic liposome-mediated gene delivery by polycations. *Biochemistry 35*, 1027–1036.
72. Li, S. D., Huang, L. (2006). Surface-modified LPD nanoparticles for tumor targeting. *Annals of the New York Academy of Sciences 1082*, 1–8.
73. Li, S. D., Chen, Y. C., Hackett, M. J., Huang, L. (2008). Tumor-targeted delivery of siRNA by self-assembled nanoparticles. *Molecular Therapy 16*, 163–169.
74. Li, S. D., Chono, S., Huang, L. (2008). Efficient gene silencing in metastatic tumor by siRNA formulated in surface-modified nanoparticles. *Journal of Controlled Release 126*, 77–84.
75. Li, S. D., Huang, L. (2009). Nanoparticles evading the reticuloendothelial system: role of the supported bilayer. *Biochimica et Biophysica Acta 1788*, 2259–2266.
76. Li, S. D., Huang, L. (2010). Stealth nanoparticles: high density but sheddable PEG is a key for tumor targeting. *Journal of Controlled Release 145*, 178–181.
77. Akita, H., Kudo, A., Minoura, A., Yamaguti, M., Khalil, I. A., Moriguchi, R., Masuda, T., Danev, R., Nagayama, K., Kogure, K., Harashima, H. (2009). Multi-layered nanoparticles for penetrating the endosome and nuclear membrane via a step-wise membrane fusion process. *Biomaterials 30*, 2940–2949.
78. El-Sayed, A., Masuda, T., Khalil, I., Akita, H., Harashima, H. (2009). Enhanced gene expression by a novel stearylated INF7 peptide derivative through fusion independent endosomal escape. *Journal of Controlled Release 138*, 160–167.
79. Hatakeyama, H., Akita, H., Harashima, H. (2011). A multifunctional envelope type nano device (MEND) for gene delivery to tumours based on the EPR effect: a strategy for overcoming the PEG dilemma. *Advanced Drug Delivery Reviews 63*, 152–160.

80. Hatakeyama, H., Akita, H., Ito, E., Hayashi, Y., Oishi, M., Nagasaki, Y., Danev, R., Nagayama, K., Kaji, N., Kikuchi, H., Baba, Y., Harashima, H. (2011). Systemic delivery of siRNA to tumors using a lipid nanoparticle containing a tumor-specific cleavable PEG-lipid. *Biomaterials 32*, 4306–4316.
81. Hatakeyama, H., Ito, E., Akita, H., Oishi, M., Nagasaki, Y., Futaki, S., Harashima, H. (2009). A pH-sensitive fusogenic peptide facilitates endosomal escape and greatly enhances the gene silencing of siRNA-containing nanoparticles *in vitro* and *in vivo*. *Journal of Controlled Release 139*, 127–132.
82. Ishitsuka, T., Akita, H., Harashima, H. (2011). Functional improvement of an IRQ-PEG-MEND for delivering genes to the lung. *Journal of Controlled Release 154*, 77–83.
83. Khalil, I. A., Hayashi, Y., Mizuno, R., Harashima, H. (2011). Octaarginine- and pH sensitive fusogenic peptide-modified nanoparticles for liver gene delivery. *Journal of Controlled Release 156*, 374–380.
84. Khalil, I. A., Kogure, K., Futaki, S., Hama, S., Akita, H., Ueno, M., Kishida, H., Kudoh, M., Mishina, Y., Kataoka, K., Yamada, M., Harashima, H. (2007). Octaarginine-modified multifunctional envelope-type nanoparticles for gene delivery. *Gene Therapy 14*, 682–689.
85. Kogure, K., Moriguchi, R., Sasaki, K., Ueno, M., Futaki, S., Harashima, H. (2004). Development of a non-viral multifunctional envelope-type nano device by a novel lipid film hydration method. *Journal of Controlled Release 98*, 317–323.
86. Kuramoto, H., Park, Y. S., Kaji, N., Tokeshi, M., Kogure, K., Shinohara, Y., Harashima, H., Baba, Y. (2008). On-chip fabrication of mutifunctional envelope-type nanodevices for gene delivery. *Analytical and Bioanalytical Chemistry 391*, 2729–2733.
87. Masuda, T., Akita, H., Nishio, T., Niikura, K., Kogure, K., Ijiro, K., Harashima, H. (2008). Development of lipid particles targeted via sugar-lipid conjugates as novel nuclear gene delivery system. *Biomaterials 29*, 709–723.
88. Moriguchi, R., Kogure, K., Akita, H., Futaki, S., Miyagishi, M., Taira, K., Harashima, H. (2005). A multifunctional envelope-type nano device for novel gene delivery of siRNA plasmids. *International Journal of Pharmaceutics 301*, 277–285.
89. Mudhakir, D., Akita, H., Tan, E., Harashima, H. (2008). A novel IRQ ligand-modified nano-carrier targeted to a unique pathway of caveolar endocytic pathway. *Journal of Controlled Release 125*, 164–173.
90. Nakamura, Y., Kogure, K., Futaki, S., Harashima, H. (2007). Octaarginine-modified multifunctional envelope-type nano device for siRNA. *Journal of Controlled Release 119*, 360–367.
91. Nakamura, Y., Kogure, K., Yamada, Y., Futaki, S., Harashima, H. (2006). Significant and prolonged antisense effect of a multifunctional envelope-type nano device encapsulating antisense oligodeoxynucleotide. *The Journal of Pharmacy and Pharmacology 58*, 431–437.
92. Sasaki, K., Kogure, K., Chaki, S., Kihira, Y., Ueno, M., Harashima, H. (2005). Construction of a multifunctional envelope-type nano device by a SUV*-fusion method. *International Journal of Pharmaceutics 296*, 142–150.
93. Shaheen, S. M., Akita, H., Nakamura, T., Takayama, S., Futaki, S., Yamashita, A., Katoono, R., Yui, N., Harashima, H. (2011). KALA-modified multi-layered nano-particles as gene carriers for MHC class-I mediated antigen presentation for a DNA vaccine. *Biomaterials 32*, 6342–6350.

94. Wang, C. Y., Huang, L. (1987). pH-sensitive immunoliposomes mediate target-cell-specific delivery and controlled expression of a foreign gene in mouse. *Proceedings of the National Academy of Sciences of the United States of America 84*, 7851–7855.
95. Hofland, H. E., Shephard, L., Sullivan, S. M. (1996). Formation of stable cationic lipid/DNA complexes for gene transfer. *Proceedings of the National Academy of Sciences of the United States of America 93*, 7305–7309.
96. Jeffs, L. B., Palmer, L. R., Ambegia, E. G., Giesbrecht, C., Ewanick, S., MacLachlan, I. (2005). A scalable, extrusion-free method for efficient liposomal encapsulation of plasmid DNA. *Pharmaceutical Research 22*, 362–372.
97. Morrissey, D. V., Lockridge, J. A., Shaw, L., Blanchard, K., Jensen, K., Breen, W., Hartsough, K., Machemer, L., Radka, S., Jadhav, V., Vaish, N., Zinnen, S., Vargeese, C., Bowman, K., Shaffer, C. S., Jeffs, L. B., Judge, A., MacLachlan, I., Polisky, B. (2005). Potent and persistent *in vivo* anti-HBV activity of chemically modified siRNAs. *Nature Biotechnology 23*, 1002–1007.
98. Graham, F. L., van der Eb, A. J. (1973). A new technique for the assay of infectivity of human adenovirus 5 DNA. *Virology 52*, 456–467.
99. Jordan, M., Schallhorn, A., Wurm, F. M. (1996). Transfecting mammalian cells: optimization of critical parameters affecting calcium-phosphate precipitate formation. *Nucleic Acids Research 24*, 596–601.
100. Li, J., Chen, Y. C., Tseng, Y. C., Mozumdar, S., Huang, L. (2010). Biodegradable calcium phosphate nanoparticle with lipid coating for systemic siRNA delivery. *Journal of Controlled Release 142*, 416–421.
101. Li, J., Yang, Y., Huang, L. (2012). Calcium phosphate nanoparticles with an asymmetric lipid bilayer coating for siRNA delivery to the tumor. *Journal of Controlled Release 158*, 108–114.
102. Weksler, B. B., Subileau, E. A., Perriere, N., Charneau, P., Holloway, K., Leveque, M., Tricoire-Leignel, H., Nicotra, A., Bourdoulous, S., Turowski, P., Male, D. K., Roux, F., Greenwood, J., Romero, I. A., Couraud, P. O. (2005). Blood-brain barrier-specific properties of a human adult brain endothelial cell line. *The FASEB Journal 19*, 1872–1874.
103. Wong, H. L., Wu, X. Y., Bendayan, R. (2011). Nanotechnological advances for the delivery of CNS therapeutics. *Advanced Drug Delivery Reviews 64*, 686–700.
104. Jain, R. K., di Tomaso, E., Duda, D. G., Loeffler, J. S., Sorensen, A. G., Batchelor, T. T. (2007). Angiogenesis in brain tumours. *Nature Reviews Neuroscience 8*, 610–622.
105. Costantino, L., Tosi, G., Ruozi, B., Bondioli, L., Vandelli, M. A., Forni, F. (2009). Chapter 3—Colloidal systems for CNS drug delivery. *Progress in Brain Research 180*, 35–69.
106. Fishman, J. B., Rubin, J. B., Handrahan, J. V., Connor, J. R., Fine, R. E. (1987). Receptor-mediated transcytosis of transferrin across the blood-brain barrier. *Journal of Neuroscience Research 18*, 299–2304.
107. Ko, Y. T., Bhattacharya, R., Bickel, U. (2009). Liposome encapsulated polyethylenimine/ODN polyplexes for brain targeting. *Journal of Controlled Release 133*, 230–237.
108. Zhang, Y., Schlachetzki, F., Zhang, Y. F., Boado, R. J., Pardridge, W. M. (2004). Normalization of striatal tyrosine hydroxylase and reversal of motor impairment in experimental parkinsonism with intravenous nonviral gene therapy and a brain-specific promoter. *Human Gene Therapy 15*, 339–350.

109. Zhang, Y., Calon, F., Zhu, C., Boado, R. J., Pardridge, W. M. (2003). Intravenous nonviral gene therapy causes normalization of striatal tyrosine hydroxylase and reversal of motor impairment in experimental parkinsonism. *Human Gene Therapy 14*, 1–12.

110. Boado, R. J. (2007). Blood-brain barrier transport of non-viral gene and RNAi therapeutics. *Pharmaceutical Research 24*, 1772–1787.

111. Zhang, Y., Wang, Y., Boado, R. J., Pardridge, W. M. (2008). Lysosomal enzyme replacement of the brain with intravenous non-viral gene transfer. *Pharmaceutical Research 25*, 400–406.

112. Boado, R. J., Pardridge, W. M. (2011). The Trojan horse liposome technology for nonviral gene transfer across the blood-brain barrier. *Journal of Drug Delivery 2011*, 296151.

113. Pardridge, W. M. (2010). Preparation of Trojan horse liposomes (THLs) for gene transfer across the blood-brain barrier. *Cold Spring Harbor Protocols 2010*, pdb prot5407.

114. Zhang, Y., Pardridge, W. M. (2009). Near complete rescue of experimental Parkinson's disease with intravenous, non-viral GDNF gene therapy. *Pharmaceutical Research 26*, 1059–1063.

115. Xia, C. F., Boado, R. J., Zhang, Y., Chu, C., Pardridge, W. M. (2008). Intravenous glial-derived neurotrophic factor gene therapy of experimental Parkinson's disease with Trojan horse liposomes and a tyrosine hydroxylase promoter. *The Journal of Gene Medicine 10*, 306–315.

116. Zhang, Y., Zhang, Y. F., Bryant, J., Charles, A., Boado, R. J., Pardridge, W. M. (2004). Intravenous RNA interference gene therapy targeting the human epidermal growth factor receptor prolongs survival in intracranial brain cancer. *Clinical Cancer Research 10*, 3667–3677.

117. Qin, Y., Chen, H., Yuan, W., Kuai, R., Zhang, Q., Xie, F., Zhang, L., Zhang, Z., Liu, J., He, Q. (2011). Liposome formulated with TAT-modified cholesterol for enhancing the brain delivery. *International Journal of Pharmaceutics 419*, 85–95.

118. Qin, Y., Zhang, Q., Chen, H., Yuan, W., Kuai, R., Xie, F., Zhang, L., Wang, X., Zhang, Z., Liu, J., He, Q. (2012). Comparison of four different peptides to enhance accumulation of liposomes into the brain. *Journal of Drug Targeting 20*, 235–245.

119. Braet, F., Wisse, E. (2002). Structural and functional aspects of liver sinusoidal endothelial cell fenestrae: a review. *Comparative Hepatology 1*, 1.

120. Hashida, M., Nishikawa, M., Yamashita, F., Takakura, Y. (2001). Cell-specific delivery of genes with glycosylated carriers. *Advanced Drug Delivery Reviews 52*, 187–196.

121. Kawakami, S., Fumoto, S., Nishikawa, M., Yamashita, F., Hashida, M. (2000). *In vivo* gene delivery to the liver using novel galactosylated cationic liposomes. *Pharmaceutical Research 17*, 306–313.

122. Nishikawa, M., Kawakami, S., Yamashita, F., Hashida, M. (2003). Glycosylated cationic liposomes for carbohydrate receptor-mediated gene transfer. *Methods in Enzymology 373*, 384–399.

123. Kim, S. I., Shin, D., Choi, T. H., Lee, J. C., Cheon, G. J., Kim, K. Y., Park, M., Kim, M. (2007). Systemic and specific delivery of small interfering RNAs to the liver mediated by apolipoprotein A-I. *Molecular Therapy 15*, 1145–1152.

124. Kim, S. I., Shin, D., Lee, H., Ahn, B. Y., Yoon, Y., Kim, M. (2009). Targeted delivery of siRNA against hepatitis C virus by apolipoprotein A-I-bound cationic liposomes. *Journal of Hepatology 50*, 479–488.

125. Chonn, A., Semple, S. C., Cullis, P. R. (1992). Association of blood proteins with large unilamellar liposomes *in vivo*. Relation to circulation lifetimes. *The Journal of Biological Chemistry 267*, 18759–18765.

126. Cullis, P. R., Chonn, A., Semple, S. C. (1998). Interactions of liposomes and lipid-based carrier systems with blood proteins: relation to clearance behaviour *in vivo*. *Advanced Drug Delivery Reviews 32*, 3–17.

127. Akinc, A., Querbes, W., De, S., Qin, J., Frank-Kamenetsky, M., Jayaprakash, K. N., Jayaraman, M., Rajeev, K. G., Cantley, W. L., Dorkin, J. R., Butler, J. S., Qin, L., Racie, T., Sprague, A., Fava, E., Zeigerer, A., Hope, M. J., Zerial, M., Sah, D. W., Fitzgerald, K., Tracy, M. A., Manoharan, M., Koteliansky, V., Fougerolles, A., Maier, M. A. (2010). Targeted delivery of RNAi therapeutics with endogenous and exogenous ligand-based mechanisms. *Molecular Therapy 18*, 1357–1364.

128. Khalil, I. A., Hayashi, Y., Mizuno, R., Harashima, H. (2011). Octaarginine- and pH sensitive fusogenic peptide-modified nanoparticles for liver gene delivery. *Journal of Controlled Release 156*, 374–380.

129. Mudhakir, D., Akita, H., Khalil, I. A., Futaki, S., Harashima, H. (2005). Pharmacokinetic analysis of the tissue distribution of octaarginine modified liposomes in mice. *Drug Metabolism and Pharmacokinetics 20*, 275–281.

130. Sarko, D., Beijer, B., Garcia Boy, R., Nothelfer, E. M., Leotta, K., Eisenhut, M., Altmann, A., Haberkorn, U., Mier, W. (2010). The pharmacokinetics of cell-penetrating peptides. *Molecular Pharmacology 7*, 2224–2231.

131. Li, S., Rizzo, M. A., Bhattacharya, S., Huang, L. (1998). Characterization of cationic lipid-protamine-DNA (LPD) complexes for intravenous gene delivery. *Gene Therapy 5*, 930–937.

132. Chen, Y., Bathula, S. R., Yang, Q., Huang, L. (2010). Targeted nanoparticles deliver siRNA to melanoma. *The Journal of Investigative Dermatology 130*, 2790–2798.

133. Chen, Y., Zhu, X., Zhang, X., Liu, B., Huang, L. (2010). Nanoparticles modified with tumor-targeting scFv deliver siRNA and miRNA for cancer therapy. *Molecular Therapy 18*, 1650–1656.

134. Lai, L. W., Moeckel, G. W., Lien, Y. H. (1997). Kidney-targeted liposome-mediated gene transfer in mice. *Gene Therapy 4*, 426–431.

135. Ito, K., Chen, J., Khodadadian, J. J., Seshan, S. V., Eaton, C., Zhao, X., Vaughan, E. D., Jr., Lipkowitz, M., Poppas, D. P., Felsen, D. (2004). Liposome-mediated transfer of nitric oxide synthase gene improves renal function in ureteral obstruction in rats. *Kidney International 66*, 1365–1375.

136. Maeda, H., Wu, J., Sawa, T., Matsumura, Y., Hori, K. (2000). Tumor vascular permeability and the EPR effect in macromolecular therapeutics: a review. *Journal of Controlled Release 65*, 271–284.

137. Yuan, F., Salehi, H. A., Boucher, Y., Vasthare, U. S., Tuma, R. F., Jain, R. K. (1994). Vascular permeability and microcirculation of gliomas and mammary carcinomas transplanted in rat and mouse cranial windows. *Cancer Research 54*, 4564–4568.

# PHOTOSENSITIVE LIPOSOMES AS POTENTIAL TARGETED THERAPEUTIC AGENTS

David H. Thompson, Pochi Shum, Oleg V. Gerasimov, and Marquita Qualls

*Department of Chemistry, Purdue University, West Lafayette, IN, USA*

## 4.1 INTRODUCTION

Many biological processes, including those involved in endosomal uptake, tumor growth and inflammation, occur in locally low pH or hypoxic environments [1,2]. In the case of rapidly growing tumors or inflammatory sites, the tissues also are often characterized by porous or poorly formed vasculature, providing limited nanoparticle access to these target tissues via the enhanced permeability and retention effect [3,4]. Taken together, these features have led to the development of a vast array of materials designs and delivery schemes to enhance the efficacy of therapeutic agents in acidic tumors and inflamed tissue. Herein we report recent efforts to utilize photochemical processes to initiate drug release from liposomal carrier systems in a controlled and potentially localized manner.

Liposomes are clinically proven, commercially successful nanoscale carrier systems for delivering high concentrations of small molecule agents such as doxorubicin [5], amphotericin B [6], verteporfin [7], cytarabin [8], and morphine [9]. Widespread clinical application of liposomes to other classes of drugs, however, has been hindered by both their limited tissue penetration [10,11] and their premature release of contents before they reach the target site. Acid-sensitive liposomes bearing targeting ligands can promote site-specific internalization via receptor-mediated

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*,
Edited by Shayne C. Gad
Copyright © 2013 John Wiley & Sons, Inc.

endocytosis [12] and exploit the time-dependent decrease in endosomal pH to promote escape of the liposomal contents into the cytosol [13–15]. They are particularly effective if they are designed with fusogenic lipids; however, if the formulation lacks an efficient triggering and endosomal escape mechanism, the drug cargo is destined for lysosomal degradation [16,17]. Bioactivation that promotes liposome–endosome membrane fusion is an effective strategy to promote cargo escape from both the liposomal and endosomal compartments. This is commonly achieved using phospholipid 1,2-dioleoyl-*sn*-glycero-3-phosphoethanolamine (DOPE), which adopts a bilayer state only at pH above the $pK_a$ of the phosphoethanolamine headgroup (i.e., >pH 8.0). At physiological pH and temperature, pure DOPE liposomes form an $H_{II}$ phase (i.e., an inverted micelle structure) that is "nonbilayer" and unable to retain the contents [18]. It is possible to increase the stability of DOPE-rich liposomes at physiological pH and temperature by formulating the liposome with stabilizing components, such as phosphoserine (PS) lipid [19,20], cholesterol hemisuccinate (CHEMS) [21,22], cleavable polyethylene glycol (PEG) lipids [23–26], or other cleavable polymers [27]. *In vitro* studies at neutral pH have shown that DOPE dispersions with CHEMS, PS, or PEG–vinyl ether–lipids [25,28,29] retain encapsulated fluorophores, suggesting that these liposomes should be structurally stable during circulation but remain responsive to low pH media. A family of PEG–vinyl ether–lipids recently reported by our group provide tunable pH-sensitivity over 10 orders of magnitude, making it now possible to engineer the liposomal release properties to display the desired release properties for a given target pH and timescale [30,31].

A greater control of acid-induced drug release may be achieved by photochemical acidification of target tissues. Light-activated dendrimeric photosensitizers [32], sensitizer-grafted polymer nanoparticles [33], gold nanorods [34], gold nanoshells [35], and liposomal-borne photosensitizers [16,36–38] have been described as agents to exert photodynamic effects on tumor cells. Liposomes offer many advantages over these systems as delivery vehicles because of the modularity of their design that offers a large aqueous compartment for cargo entrapment and a membrane shell that provides an adjustable surface-to-volume ratio as well as a site for hosting targeting ligands, fusogenic species, or other functional components. Thompson and coworkers have studied plasmenylcholines [39] and PEG–vinyl ether–lipids [25,28–31] for their potential as gene and drug delivery vehicles that are activated under low pH conditions and hydrolyzed via vinyl ether protonation reaction leading to cargo release from the carrier system [40]. It is well known that various onium salts are capable of producing Brønsted acids upon photo-induced intramolecular electron transfer bond cleavage [41–43]. One class of onium salts that exhibits this property is an anthracene-based photoacid generator (An-PAG), which liberates protons on exposure to UV illumination. Incorporation of lipid-modified An-PAG into liposomal bilayers, therefore, was anticipated to provide a means to stimulate contents release from these vehicles via acid-catalyzed vinyl ether hydrolysis in a light intensity-dependent manner. Incorporation of photosensitizers that can efficiently generate singlet oxygen within the vicinity of vinyl ether lipid membranes is another method of stimulating liposome release on aerobic irradiation.

This work describes the synthesis of a hexadecyl-modified An-PAG for use as a light-activated trigger, which will induce contents release from DOPE:lysoplasmenylcholine liposomes. We also describe the liposomes containing photosensitizers such as bacteriochlorophyll a (Bchl a) or aluminum phthalocyanine tetrasulfonate ($AlPcS_4^{4-}$). Finally, biodistribution of $AlPcS_4^{4-}$ -loaded liposomes in a murine tumor xenograft model is described.

## 4.2 EXPERIMENTAL PROCEDURES

### 4.2.1 Materials

DOPE and lysoplasmenylcholine (LysoPlsC) were obtained from Avanti Polar Lipids; all chemicals were obtained from Aldrich Chemical Company. Synthetic intermediates and final products were characterized by 300 MHz $^1$H NMR and MALDI MS using the Purdue University Center for Cancer Research's Interdepartmental NMR Facility and Campus-Wide Mass Spectrometry Center, respectively.

### 4.2.2 Synthesis of 1-Bromo-3-hexadecylthiobenzene (1)

To a solution of 1-bromo-3-thiophenol (2 ml, 19.4 mmol) in 20 ml of dry THF at 20°C was added NaH (20.0 mmol) with stirring under an Ar atmosphere. After gas evolution ceased (~10 min), 1-bromohexadecane (6.1 ml, 20 mmol) was added, along with a trace of $Bu_4NI$ as phase transfer catalyst, to the reaction mixture and heated at reflux for 20 h. The reaction mixture was then concentrated by rotary evaporation, the unreacted NaH destroyed by addition of $H_2O$ (5 ml), and the product extracted with $CH_2Cl_2$ (3 × 20 ml). The combined $CH_2Cl_2$ extracts were dried over $MgSO_4$ before filtration, evaporation, and purification by silica gel column chromatography with 95:5 hexane:EtOAc as eluent.

### 4.2.3 Synthesis of 9-[3-Hexadecylthiophenyl] anthracene (2)

Compound (**1**) (1 g, 2.42 mmol) was dissolved in 20 ml dry THF and Mg turnings (57.6 mg, 2.5 mmol) added along with a small $I_2$ crystal. The mixture was stirred under an Ar atmosphere for 10 min and then heated at reflux until all of the Mg had dissolved (about 1 h). Anthrone (580 mg, 3 mmol) dissolved in 10 ml dry benzene was then added all at once and the mixture stirred at 20°C for 18 h. The reaction was quenched by addition of 30% $H_2SO_4$ in water (10 ml) and the mixture extracted with $CHCl_3$ (3 × 20 ml). The extracts were combined, dried over $MgSO_4$, filtered, and the filtrate concentrated by rotary evaporation.

### 4.2.4 Synthesis of 9-[3-Hexadecyl-3'-cyanobenzylsulfonium] phenylanthracene Trifluoromethanesulfonate (C16-An-PAG) (3)

Compound (**2**) was dissolved in $CH_2Cl_2$ with 3-cyanobenzyl bromide. $Ag(OTf)_2$-dioxane complex was added to the reaction mixture as a solid and the reaction stirred

for 18 h at 20°C before filtration and evaporation of the solvent to give a crude residue that was purified by silica gel column chromatography using hexane:$Et_2O$ as eluent to give $C_{16}$-An-PAG.

### 4.2.5 Liposome Preparation

Stock solutions of DOPE (20 mg/ml), LysoPlsC (20 mg/ml), and $C_{16}$-An-PAG (4 mg/ml) were prepared in 95:5 benzene:MeOH and combined to give 10 mmol total of a 35:35:30 molar ratio of components. The lipid stock solutions were combined at the desired ratios and vortexed for 2 min before flash freezing in $LN_2$ for 3 min and lyophilization (15 h) to produce a powder. The lyophilized lipid mixtures were then hydrated in 1 ml of 100 mM $CaCl_2$ solution via 10 freeze–thaw–vortex cycles. The lipid suspension then was extruded 10 times through 100 nm polycarbonate membranes filters at 45°C. The liposome solution was passed through a Sephadex G-50 column equilibrated with a 150 mM NaCl solution to remove unentrapped $Ca^{2+}$ and the liposomes used immediately. Remote loading of doxorubicin was performed using the ammonium sulfate gradient method as described previously [37]. $AlPcS_4^{4-}$ was loaded using the passive entrapment method using DSPE–PEG2000–folate as targeting ligand as previously described [16].

### 4.2.6 Dynamic Light Scattering

Liposome sizes were measured by dynamic light scattering using a Coulter N4-Plus instrument and the distributions calculated using the manufacturer's supplied software. The experiments were run in triplicate and the average diameter reported.

### 4.2.7 Photoexcitation Method

Samples of the liposome solution (3 ml) were placed in a 1 cm path length quartz cuvette and placed in a temperature-controlled cuvette holder with stirring. A frequency-doubled continuous wave, mode-locked Nd:YAG laser provided 10 mW/s at 368 nm with a maximum energy of 1 mJ/pulse was used to activate the $C_{16}$-An-PAG-containing liposomes. Lenses were used to focus the beam to a typical spot size of 6 mm diameter. Bchl a- and $AlPcS_4^{4-}$ -containing liposomes were irradiated with an Al/Ga/As diode array laser emitting at 800 nm and coupled into a 1 mm optical fiber that was focused onto a temperature-controlled cuvette containing the stirred solution of liposomes.

### 4.2.8 Calcium Ion Release Assay

Absorption spectra were measured with a Hewlett-Packard 8453 UV–Vis spectrophotometer equipped with Chem Station software. Calcium-loaded liposomes were mixed with 2.85 ml Arsenazo III in 20 mM HEPES, pH 7.4. The absorption at 656 nm was measured and then Triton X-100 added to disrupt the liposomes and the absorption at 656 nm re-measured. The percent release was quantified using the

ratio method: % release $= 100 \times [(T - T_0)/(1 - T_0)]$, where $T =$ (fluorescence before Triton addition/fluorescence after Triton addition) at time $T$; $T_0 =$ fluorescence ratio at $T = 0$ min [44].

### 4.2.9 Negative Stain Transmission Electron Microscopy

A small drop of liposome solution was placed on a carbon-coated G-300 mesh copper grid and the excess liquid blotted away using filter paper after 1 min. A 1% solution of $UO_2OAc_2$ was then added for 1 min and the excess again removed using the filter paper. The sample was dried in a desiccator overnight before imaging on a Philips EM-400 transmission electron microscope operating at 80 kV accelerating voltage.

### 4.2.10 Animal Studies

Balb/c mice on folate-deficient diets were implanted with $5 \times 10^5$ M109 lung tumor cells via subcutaneous injection in a 100 μl volume with folate-deficient DMEM. The implants were allowed to grow for 3 week before imaging using a custom-built long-focal length microscope and camera system.

## 4.3 RESULTS AND DISCUSSION

### 4.3.1 DOPE:LysoPlsC Liposomes Using $C_{16}$-An-PAG as a Photosensitizer

***4.3.1.1 Synthesis of $C_{16}$-An-PAG.*** Indirect photoinitiation of liposomal content release was sought by using a membrane-anchored photoacid generator. We elected to prepare a hexadecyl-modified version of the anthracene-based photoacid generator [41–43] that transfers an excited $\pi^*$ electron from the anthracene moiety to the $\sigma^*$ lowest unoccupied molecular orbital localized on the phenylsulfonium moiety on irradiation at 368 nm, leading to bond rearrangement and production of a Brønsted acid. Previous work also suggests that the cyanobenzyl-substituted sulfonium salts are cleaved by a concerted electron transfer-bond fragmentation process (Scheme 4.1). The rate of C—S bond cleavage is exceedingly fast and dependent on several factors, including the thermodynamic stability of the cation radical–radical pair. Compound (**3**) was prepared as shown in Scheme 4.2 using a modified procedure reported for the methyl analog [41].

***4.3.1.2 Solution Acidification on Photolysis of $C_{16}$-An-PAG.*** $C_{16}$-An-PAG was dissolved in dry acetonitrile at a concentration of 1.5 mM and the absorption spectrum recorded, revealing peaks at 334, 350, 369, and 388 nm (Figure 4.1). Irradiation of this solution at 368 nm (35 mW/cm$^2$ illumination), with periodic sampling and analysis on 1:10 dilution into water, produced a rapid acidification over 2.5 pH units within 2 min at 37°C (Figure 4.2). The rapid acidification produced by this photoacid generator suggested that it could be useful for initiating acid-triggered release

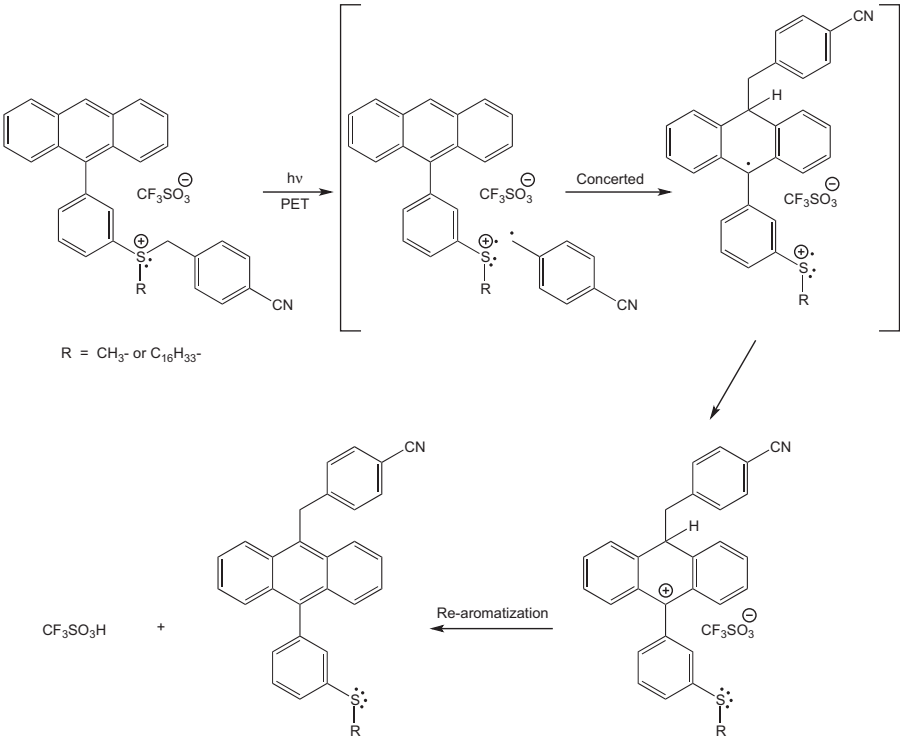


**SCHEME 4.1** Photoinduced electron transfer process leading to the formation of triflic acid.

from appropriately designed pH-sensitive liposomes. Prior work by Thompson and coworkers [36,37] suggested that plasmenyl-type lipids could be used to effect rapid contents release from liposomes on acidification. This potential was tested using blends of LysoPlsC and DOPE.

***4.3.1.3 Light-Induced Liposome Leakage and Phase Change Using C16-An-PAG.*** The performance of 100 nm 35:35:30 DOPE:LysoPlsC:$C_{16}$-An-PAG liposomes was evaluated using a $Ca^{2+}$-Arsenazo III assay [45]. This formulation was selected because the $H_{II}$ phase of DOPE can be transformed into a stable lamellar phase by blending it with LysoPlsC, a pH-sensitive vinyl ether lipid of opposite curvature propensity that ultimately produces fatty acid degradation products upon acid-catalyzed hydrolysis in aerobic solution (Scheme 4.3). Cleavage of the LysoPlsC fraction within the liposome membrane causes a change in the intrinsic curvature of the fatty acid fragments, leading to a loss of $L_\alpha$ phase stability and a collapse of the liposome into the preferred $H_{II}$ phase of DOPE with concomitant release of the liposomal contents. Irradiation of the extruded, unbuffered liposome solution at 368 nm lead to 75% $Ca^{2+}$ release within 4 h, with accompanying change in the lipid morphology from spherical $L_\alpha$ phase liposomes to collapsed $H_{II}$ phase aggregates (Figure 4.3). Nonirradiated and irradiated buffered controls produced

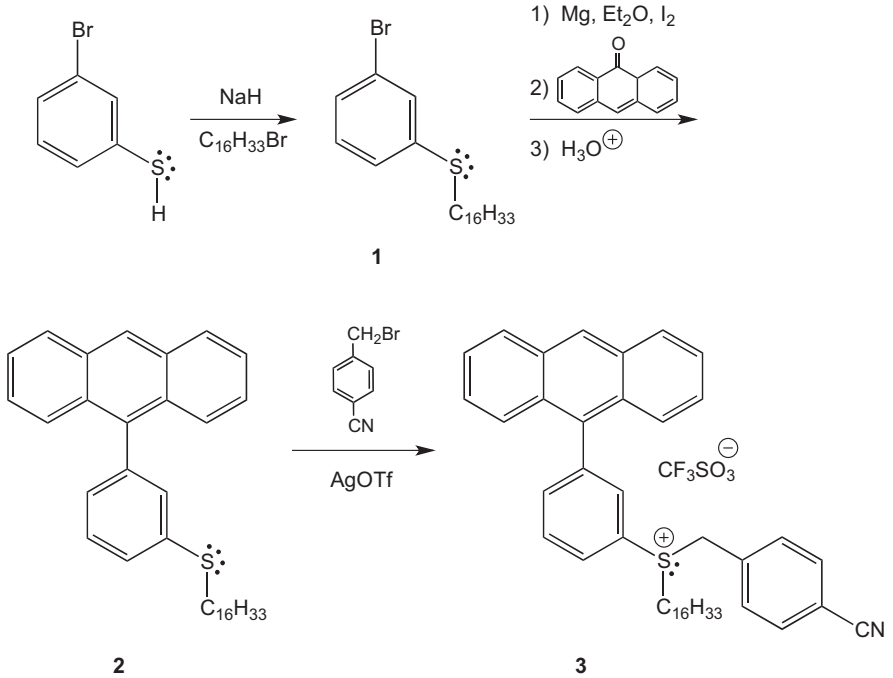


**SCHEME 4.2** Synthesis pathway for $C_{16}$-An-PAG.

13% and 20% release, respectively, under the same illumination conditions. While the system containing $C_{16}$-An-PAG shows light-activated content release in an unbuffered condition, a faster release kinetics was desired. Therefore, we investigated another light-sensitive liposome systems.

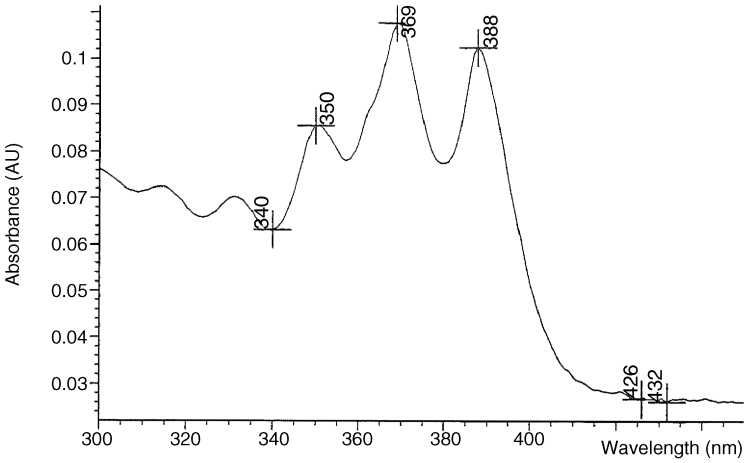


**FIGURE 4.1** Absorption spectrum of $C_{16}$-An-PAG (**3**) in acetonitrile at 20°C.

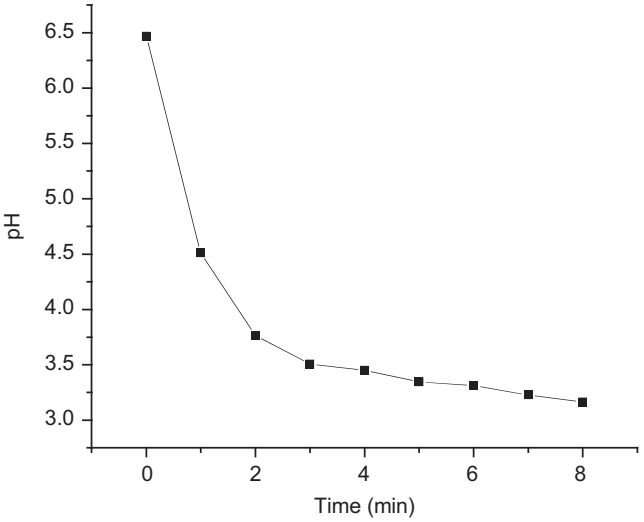


**FIGURE 4.2** Rate of photochemical acidification produced by 1.5 mM (**3**) in acetonitrile at 37°C on illumination at 368 nm, 35 mW/cm$^2$.

### 4.3.2 PPlsC Liposomes Using Bchl a as a Photosensitizer

***4.3.2.1 Light-Induced Liposome Leakage and Phase Change Using Photosensitized Production of Singlet Oxygen ($^1O_2$).*** Prior work has shown that rapid calcein release occurs on photosensitized oxidation of palmitoylplasmenylcholine (PPlsC) [36,37] and dipalmitoylplasmenylcholine (DPPlsC) [46,47] liposomes at neutral pH using a variety of different chromophores. We evaluated the potential of PPlsC liposomes for drug delivery using Bchl a (Figure 4.4a) as a membrane-soluble

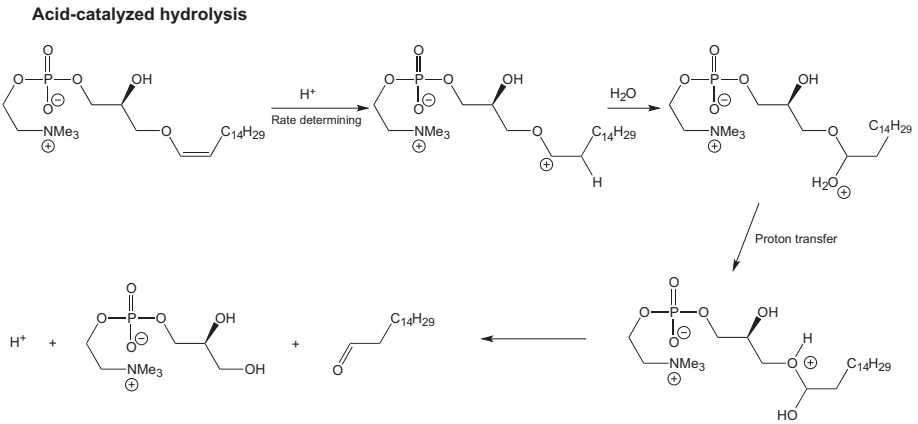


**SCHEME 4.3** Pathway for acid-catalyzed lysoplasmenylcholine (LysoPlsC) lipid hydrolysis. Aerobic oxidation of the initially formed fatty aldehyde hydrolysis product generates the corresponding fatty acid.

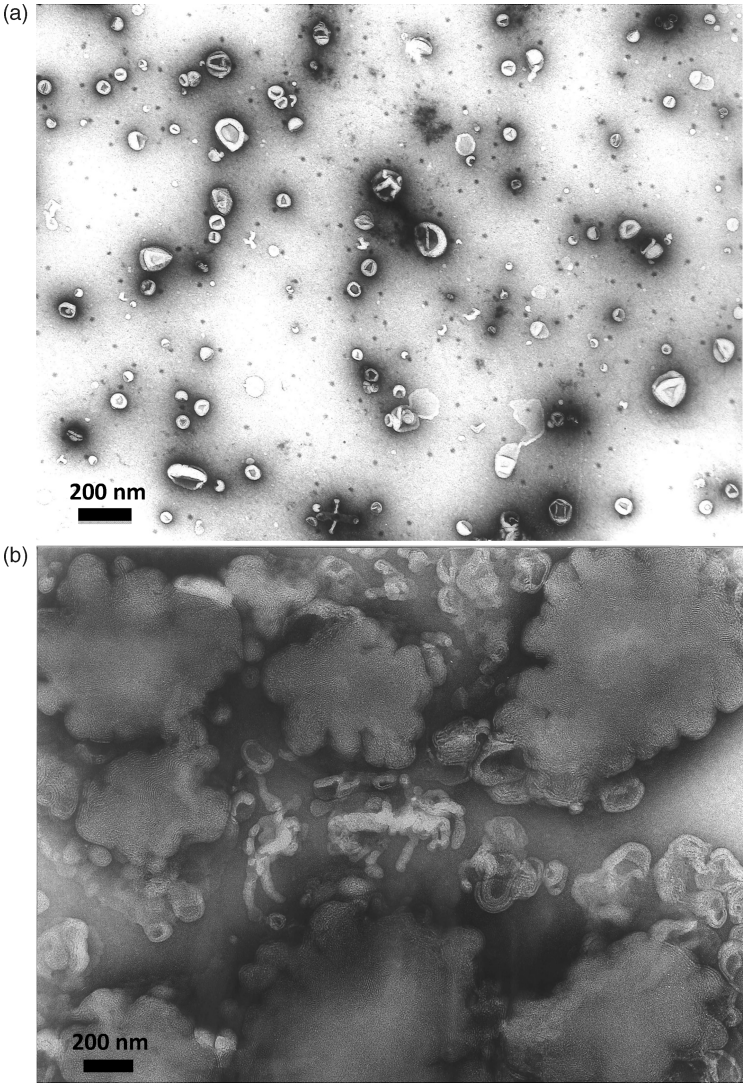


**FIGURE 4.3** Negative-stain TEM of 35:35:30 DOPE:LysoPlsC:$C_{16}$-An-PAG liposomes before (a) and 4 h after (b) irradiation at 368 nm.

photosensitizer. Since Bchl a is a near-infrared sensitizer that efficiently generates $^1O_2$ on aerobic irradiation, it was envisioned that rapid drug release should be possible on irradiation of drug-loaded PPlsC liposomes doped with Bchl a.

Liposome samples were prepared by evaporating a $CHCl_3$ solution of PPlsC and Bchl a into a thin film before hydrating the sample in ammonium citrate buffer and extruding it through 100 nm track-etch membranes. After buffer exchange to HEPES buffered saline via gel chromatography, doxorubicin was added and the solution incubated at 50°C for 10 min. Analysis of both the lipid and doxorubicin content indicated that a drug:lipid ratio of 0.085 was achieved, or approximately

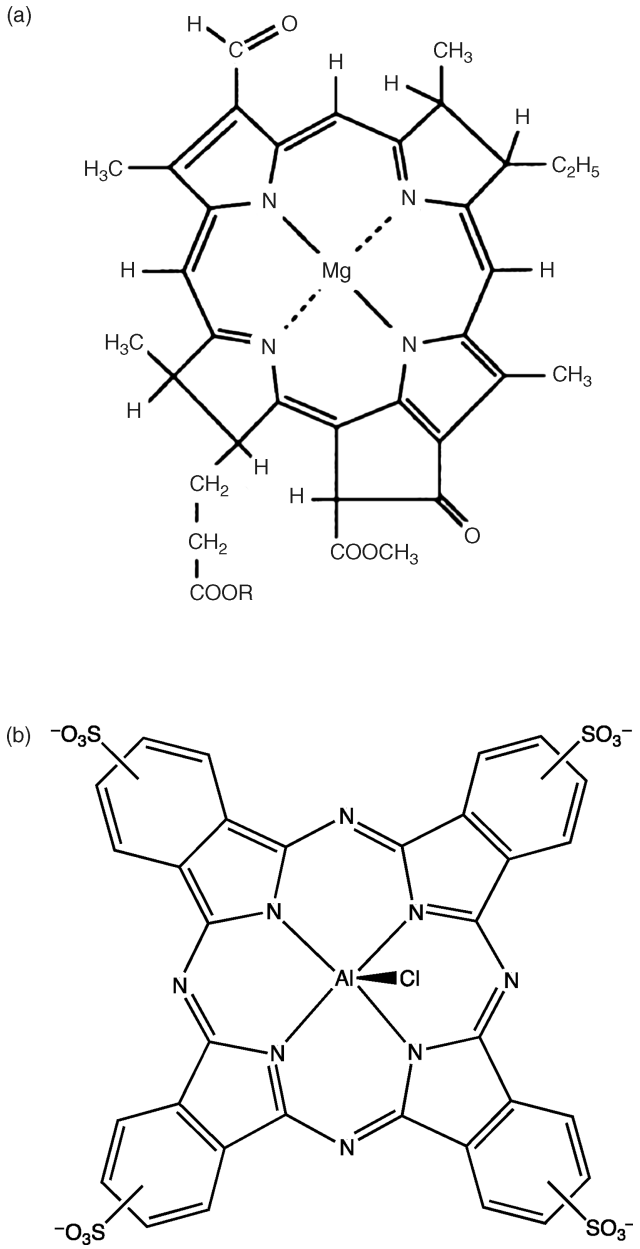


**FIGURE 4.4** Photosensitizers used in this study. Bacteriochlorophyll a; (Bchl a, a) and chloroaluminum phthalocyaninetetrasulfonate ($AlPcS_4^{4-}$, b).

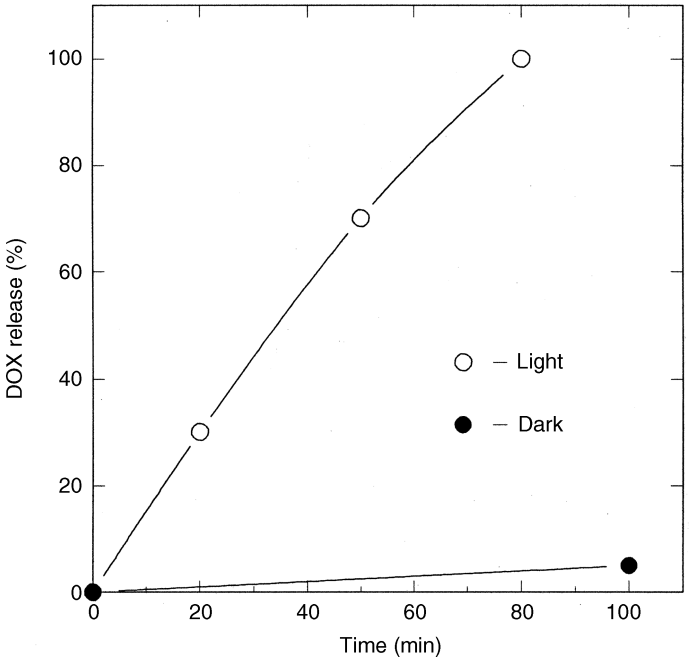


**FIGURE 4.5** Rate of doxorubicin release from remote-loaded 2-palmitoylplasmenylcholine (PPlsC) liposomes sensitized with Bchl a. [PPlsC] = 1.12 mM; [dihydrocholesterol] = 0.28 mM; [DOX] = 120 μM; [Bchl a] = 34 μM; 500 mW illumination at 800 nm, 38°C, 20 mM HEPES, 150 mM NaCl, pH 7.4.

2000 doxorubicin molecules per liposome. Illumination of this sample at 38°C with an Al/Ga/As diode laser emitting 800 nm excitation at 500 mW intensity led to complete doxorubicin release within 80 min (Figure 4.5). Since Bchl a is known to photobleach rapidly within the first 10 min of irradiation [37], the observed rate of drug release is actually a minimum drug efflux rate under these conditions since the extent of PPlsC photooxidation (estimated to be less than ~25%) is lower than desired to effect rapid release. The photoreleased doxorubicin remained cytotoxic within experimental error of the measurement [37].

### 4.3.3 Folate Receptor (FR)-Targeted DPPlsC Liposomes Using $AlPcS_4^{4-}$ as a Photosensitizer

***4.3.3.1 Rate of Liposome Internalization and Cargo Leakage Using the Folic Acid-Based Receptor-Mediated Endocytosis Pathway.*** Low and coworkers have demonstrated the utility of folate conjugates for targeting folate receptor (FR) positive tumor cells *in vitro* and *in vivo* [48]. Folate-modified liposomes have also been used to deliver cytosine arabinose to the cytoplasm of target cells [13]. In order to assess the kinetic timescale of this process, we loaded FR-targeted DPPlsC liposomes with propidium iodide (PI) as a probe that produces little fluorescence

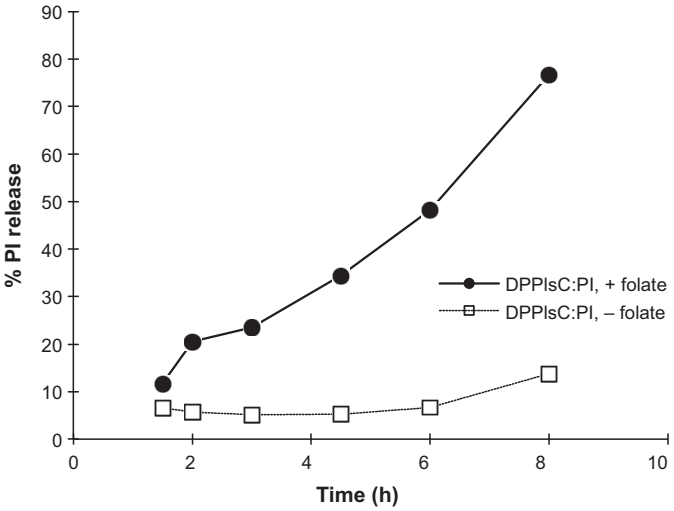


**FIGURE 4.6** Rate of propidium iodide (PI) release from DPPlsC liposomes in KB cell culture with and without FR targeting. KB cells were treated with DPPlsC:PI liposomes containing 5% mPEG2000–DSPE (DPPlsC:PI, − folate) or 4.9% mPEG2000 and 0.1% folate–PEG3400–DSPE (DPPlsC:PI, + folate).

until it intercalates into DNA. Once inside the acidic endosomal pathway, we anticipated that the vinyl ether lipid would be hydrolyzed in a manner similar to Scheme 4.3 and the corresponding fatty aldehyde/fatty acid fragments would facilitate PI escape from the endosome and intercalate into cellular DNA. As the data in Figure 4.6 show, nearly 80% of the encapsulated PI is gradually released from FR-targeted liposomes over a 2–8 h timescale, whereas $\leq$14% PI staining of KB cells over the same time period when nontargeted DPPlsC:PI liposomes were used. These results show that the FR-targeted liposomes can deliver a cargo to FR-positive cells using receptor-mediated internalization.

***4.3.3.2 Pharmacokineticss and Biodistribution of FR-Targeted DPPlsC Liposomes Bearing $AlPcS_4^{4-}$ as a Photosensitizer Cargo.*** We sought to combine other water-soluble cargo, such as chloroaluminum phthalocyaninetetrasulfonate ($AlPcS_4^{4-}$) (Figure 4.4b), in the FR-targeted liposomes for labeling the cytoplasm of target cells for photodynamic therapy applications. We investigated the potential of this approach by loading FR-targeted DPPlsC liposomes with $AlPcS_4^{4-}$ at self-quenching concentrations so that the internalized sensitizer would be retained in a nonphotoactive state until reaching the target cell and escaping from the endosomes into the surrounding cytoplasm [16]. This photosensitizer was chosen because it is (i) much more photostable than Bchl a; (ii) rapidly cleared from blood in its free form (i.e., will be rapidly excreted if lost from the liposomal carrier during systemic circulation, thereby limiting the dermatologic sensitivity that is commonly observed for hematoporphyrin derivative formulations); (iii) capable of triggering efflux from DPPlsC liposomes as shown in Scheme 4.4; and (iv) highly water soluble, thus

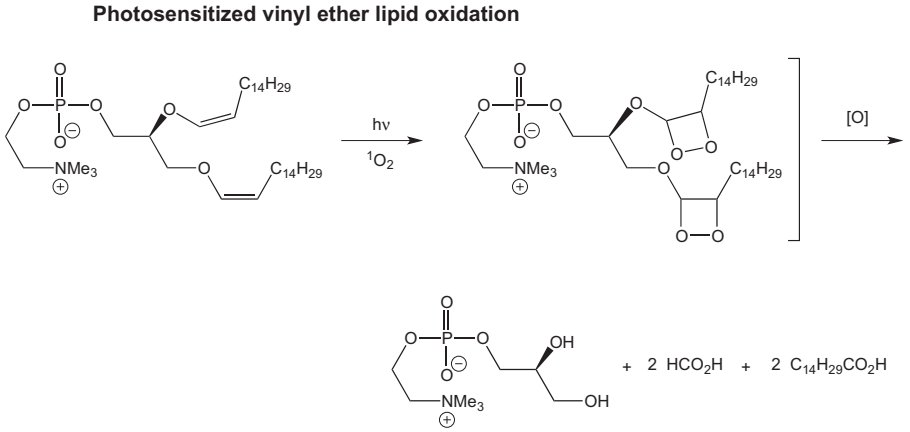


**SCHEME 4.4** Pathway for photooxidative cleavage of the divinyl ether lipid, diplasmenylcholine (DPPlsC). Stepwise addition of $^1O_2$ to the electron rich vinyl ether bond leads to the thermally unstable, putative bis-dioxetane intermediate shown in brackets. Initial fragmentation of this species produces 2 equiv. of glycerophosphocholine bisformate ester that are rapidly hydrolyzed to formate and glycerophosphocholine. Aerobic oxidation of the initially formed fatty aldehyde photolysis product generates the corresponding fatty acid.

potentially limiting its volume of distribution *in vivo* to only those cells that have internalized the $AlPcS_4^{4-}$-loaded liposomes, such as FR-positive tumor cell targets and nontarget cells of the reticuloendothelial system that commonly scavenge nanoparticulate carriers (i.e., macrophages within the liver and spleen).

As the data in Figure 4.7 show, substantial $AlPcS_4^{4-}$ fluorescence appears rapidly in the 1–2 mm subcutaneous M109 FR-positive xenograft tumors within 2 h compared to nontarget surrounding skin tissue. Approximately 35% of this tumor-localized $AlPcS_4^{4-}$ fluorescence is lost within 5 h, with slow depletion of an additional 12% occurring over the next 67 h, to give a total retained dose of ~53% 72 h after injection (<0.3% of skin-localized dose was detectable in healthy skin at this time point). Since the $AlPcS_4^{4-}$ cargo was encapsulated at self-quenching concentrations, we cannot entirely exclude the possibility that normal skin was capable of harboring liposomes but not stimulating $AlPcS_4^{4-}$ release as in FR-targeted KB cells [16]. However, in the absence of specific adhesion mechanisms that would promote liposome accumulation in vital skin, we infer from these findings that the low observed $AlPcS_4^{4-}$ fluorescence is due to low levels of skin-accumulated $AlPcS_4^{4-}$. This is an important advantage since one of the most common side effects of the hydrophobic photosensitizers typically used in photodynamic therapy is cutaneous photosensitivity. The low levels of skin-accumulated $AlPcS_4^{4-}$ fluorescence suggest that this issue may not be as problematic as other hydrophobic photosensitizers that display a long persistence in skin.

The biodistribution of $AlPcS_4^{4-}$ in the mouse model was evaluated 2 h after infusion of FR-targeted DPPlsC/$AlPcS_4^{4-}$ liposomes via tail vein injection (Figure 4.8). Liver, tumor, spleen, and kidney were found to have the highest $AlPcS_4^{4-}$

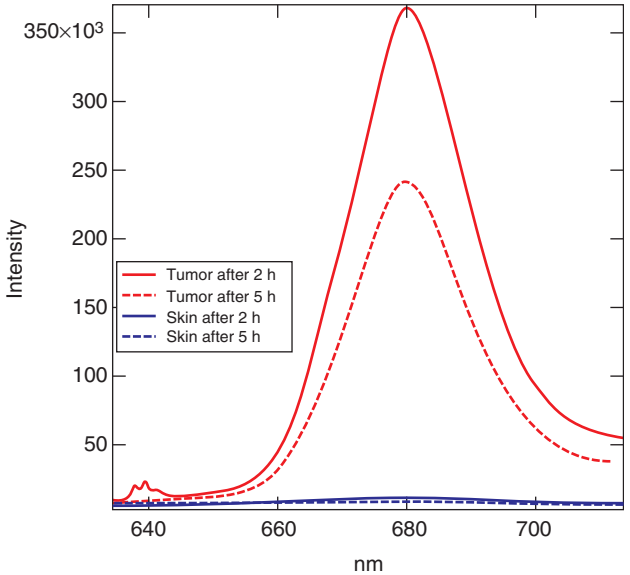


**FIGURE 4.7** Fluorescence kinetics of FR-targeted DPPlsC liposomes loaded with $AlPcS_4^{4-}$ in a 1–2 mm subcutaneous folate-receptor positive M109 FR-positive tumor murine xenograft relative to nontarget skin tissue. DPPlsC:Chol:mPEG2000–DSPE:folate–PEG3400–DSPE liposomes (59.9:35:5:0.1) containing 9.8 μM $AlPcS_4^{4-}$ were injected via tail vein and imaged using a home-built fluorescence imaging microscope at the times indicated.

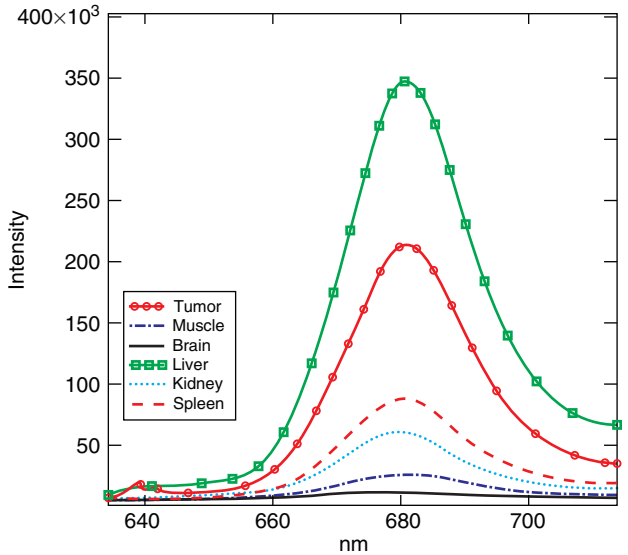


**FIGURE 4.8** Biodistribution of $AlPcS_4^{4-}$ fluorescence 48 h after infusion of FR-targeted DPPlsC/$AlPcS_4^{4-}$ liposomes as described in Figure 4.7.

fluorescence levels, with liver and tumor displaying 2–3 times higher accumulation than spleen and kidney over this time period. Extensive uptake by liver and spleen is commonly observed for nanoparticles [8], as is kidney uptake due to scavenger receptors for folic acid in these tissues [49]. It is noteworthy that the levels present in tumor were comparatively high, however, additional experiments are needed to determine their specific localization (i.e., deeply penetrated into tumor tissue or near the extravasation site) and whether they have been internalized by FR-positive tumor cells or are simply adherent to the surface of the cells.

## 4.4 CONCLUSIONS

Several different photochemical schemes were developed and employed using a variety of vinyl ether-based liposome systems. Photoacid generation using (**3**), a membrane-immobilized, anthracene-based sulfonium salt, induced cargo release on the hours timescale even though the rate of $H^+$ release occurs on the minutes timescale. This suggests that, in spite of membrane localization of the $C_{16}$-An-PAG derivative, the protons are released into bulk solution rather than remain associated with the membrane where vinyl ether hydrolysis must occur. Photooxidative drug release was also demonstrated on the minutes timescale using the near infrared photosensitizer, Bchl a. FR targeting of liposomal PI and $AlPcS_4^{4-}$ to FR-positive KB cells was also demonstrated in cell culture and *in vivo*, with the latter displaying promising characteristics with regard to tumor:nontarget tissue distributions. Taken together, these results suggest that it is possible to design light-responsive liposomes to provide a level of functional control and reaction kinetics that are difficult to achieve with other triggering modalities.

## 4.5 PROSPECTS FOR LIGHT-ACTIVATED LIPOSOMAL DRUG DELIVERY

Our findings suggest that opportunities remain for the development of effective phototherapeutic strategies using specific targeting ligands that can selectively deliver cytotoxic agents to tumor cells *in vivo*. Major challenges remain, however, in the development of (i) chromophores that can be activated in the near infrared to elicit their action in the absence of oxygen, since most solid tumors have regions that are substantially hypoxic; (ii) strategies that display large bystander effects due to tumor heterogeneity; and (iii) techniques that enable the real-time assessment of therapeutic progress so that the treatment can be modulated as required for individual patient outcome.

### Acknowledgments

We gratefully acknowledge financial support by National Institutes of Health (GM087016) and the 21st Century Research and Technology Fund of Indiana. We would like to thank the Purdue University Center for Cancer Research

Interdepartmental NMR Facilities and Campus-Wide Mass Spectrometry Center (supported by NCI CCSG CA23168 to Purdue University Center for Cancer Research) for their materials characterization support. We also thank Mike Kennedy, Karim Jallad, Dor Ben-Amotz, and Phil Low for fruitful discussions.

## REFERENCES

1. Mellman, I., Fuchs, R., Helenius, A. (1986). Acidification of the endocytic and exocytic pathways. *Annual Review of Biochemistry 55*, 663–700.
2. Gerweck, L. E. (1998). Tumor pH: implications for treatment and novel drug design. *Seminars in Radiation Oncology 8*, 176–182.
3. Maeda, H. (2010). Tumor-selective delivery of macromolecular drugs via the EPR effect: background and future prospects. *Bioconjugate Chemistry 21*, 797–802.
4. Jain, R. K. (2001). Delivery of molecular and cellular medicine to solid tumors. *Advanced Drug Delivery Reviews 46*, 149–168.
5. Drummond, D. C., Noble, C. O., Hayes, M. E., Park, J. W., Kirpotin, D. B. (2008). Pharmacokinetics and in vivo drug release rates in liposomal nanocarrier development. *Journal of Pharmaceutical Sciences 97*, 4696–4740.
6. Guo, L. S. S. (2001). Amphotericin B colloidal dispersion: an improved antifungal therapy. *Advanced Drug Delivery Reviews 47*, 149–163.
7. Cruess, A. F., Zlateva, G., Pleil, A. M., Wirostko, B. (2009). Photodynamic therapy with verteporfin in age-related macular degeneration: a systematic review of efficacy, safety, treatment modifications and pharmacoeconomic properties. *Acta Opthalmologica 87*, 118–132.
8. Duncan, R., Gaspar, R. (2011). Nanomedicine(s) under the microscope. *Molecular Pharmaceutics 8*, 2101–2141.
9. Allen, T. M., Cullis, P. R. (2004). Drug delivery systems: entering the mainstream. *Science 303*, 1818–1822.
10. Kirpotin, D. B., Drummond, D. C., Shao, Y., Shalaby, M. R., Hong, K. L., Nielsen, U. B., Marks, J. D., Benz, C. C., Park, J. W. (2006). Antibody targeting of long-circulating lipidic nanoparticles does not increase tumor localization but does increase internalization in animal models. *Cancer Research 66*, 6732–6740.
11. Wong, C., Stylianopoulos, T., Cui, J., Martin, J., Chauhan, V. P., Jiang, W., Popovic, Z., Jain, R. K., Bawendi, M. G., Fukumura, D. (2011). Multistage nanoparticle delivery system for deep penetration into tumor tissue. *Proceedings of the National Academy of Sciences of the United States of America 108*, 2426–2431.
12. Mayor, S., Pagano, R. E. (2007). Pathways of clathrin-independent endocytosis. *Nature Reviews Molecular Cell Biology 8*, 603–612.
13. Rui, Y., Wang, S., Low, P. S., Thompson, D. H. (1998). Diplasmenylcholine-folate liposomes: an efficient vehicle for intracellular drug delivery. *Journal of the American Chemical Society 120*, 11213–11218.
14. Drummond, D. C., Noble, C. O., Guo, Z., Hayes, M. E., Connolly-Ingram, C., Gabriel, B. S., Hann, B., Liu, B., Park, J. W., Hong, K. L., Benz, C. C., Marks, J. D., Kirpotin, D. B. (2010). Development of a highly stable and targetable nanoliposomal formulation of topotecan. *Journal of Controlled Release 141*, 13–21.

15. Biswas, S., Dodwadkar, N. S., Sawant, R. R., Torchilin, V. P. (2011). Development of the novel PEG–PE-based polymer for reversible attachment of specific ligands to liposomes: synthesis and in vitro characterization. *Bioconjugate Chemistry 22*, 2005–2013.
16. Qualls, M. M., Thompson, D. H. (2001). Chloroaluminum phthalocyanine tetrasulfonate delivered via acid-labile diplasmenylcholine-folate liposomes: intracellular localization and synergistic phototoxicity. *International Journal of Cancer 93*, 384–392.
17. Coon, B. G., Crist, S., Gonzalez-Bonet, A., Kim, H.-K., Sowa, J., Thompson, D. H., Ratliff, T. L., Aguilar, R. C. (2011). Fibronectin attachment protein (FAP) from Bacillus Calmette-Guerin as a targeting agent for bladder tumor cells. *International Journal of Cancer 131*, 591–600.
18. Ellens, H., Bentz, J., Szoka, F. C. (1986). Fusion of phosphatidylethanolamine-containing liposomes and mechanism of the La-HII phase transition. *Biochemistry 25*, 4141–4147.
19. Bally, M. B., Tilcock, C. P. S., Hope, M. J., Cullis, P. R. (1983). Polymorphism of phosphatidylethanolamine phosphatidylserine model systems: influence of cholesterol and $Mg^{2+}$ on $Ca^{2+}$-triggered bilayer to hexagonal (HII) transitions. *Canadian Journal of Biochemistry & Cell Biology 61*, 346–352.
20. Ellens, H., Bentz, J., Szoka, F. C. (1984). pH-Induced destabilization of phosphatidylethanolamine-containing liposomes: role of bilayer contact. *Biochemistry 23*, 1532–1538.
21. Hafez, I. M., Ansell, S., Cullis, P. R. (2000). Tunable pH-sensitive liposomes composed of mixtures of cationic and anionic lipids. *Biophysical Journal 79*, 1438–1446.
22. Slepushkin, V. A., Simoes, S., deLima, M. C. P., Duzgunes, N. (2004). Sterically stabilized pH-sensitive liposomes. *Methods in Enzymology 387*, 134–147.
23. Kirpotin, D., Hong, K. L., Mullah, N., Papahadjopoulos, D., Zalipsky, S. (1996). Liposomes with detachable polymer coating: destabilization and fusion of dioleoylphosphatidylethanolamine vesicles triggered by cleavage of surface-grafted poly(ethylene glycol). *FEBS Letters 388*, 115–118.
24. Zalipsky, S., Qazen, M., Walker, J. A., Mullah, N., Quinn, Y. P., Huang, S. K. (1999). New detachable poly(ethylene glycol) conjugates: cysteine-cleavable lipopolymers regenerating natural phospholipid, diacylphosphatidylethanolamine. *Bioconjugate Chemistry 10*, 703–707.
25. Boomer, J. A., Inerowicz, H. D., Zhang, Z.-Y., Bergstrand, N., Edwards, K., Kim, J.-M., Thompson, D. H. (2003). Acid-triggered release from sterically-stabilized fusogenic vesicles via a hydrolytic dePEGylation strategy. *Langmuir 19*, 6408–6415.
26. Shin, J., Shum, P., Thompson, D. H. (2003). Acid-triggered release via dePEGylation of DOPE liposomes containing acid-labile vinyl ether PEG-lipids. *Journal of Controlled Release 91*, 187–200.
27. Sakaguchi, N., Kojima, C., Harada, A., Kono, K. (2008). Preparation of pH-sensitive poly (glycidol) derivatives with varying hydrophobicities: their ability to sensitize stable liposomes to pH. *Bioconjugate Chemistry 19*, 1040–1048.
28. Bergstrand, N., Arfvidsson, M. C., Kim, J.-M., Thompson, D. H., Edwards, K. (2003). Interactions between pH-sensitive liposomes and model membranes. *Biophysical Chemistry 104*, 361–379.
29. Boomer, J. A., Qualls, M. M., Inerowicz, H. D., Haynes, R. H., Patri, G. V. S., Kim, J.-M., Thompson, D. H. (2009). Cytoplasmic delivery of liposomal contents mediated by an acid-labile cholesterol–vinyl ether–PEG conjugate. *Bioconjugate Chemistry 20*, 47–59.

30. Shin, J., Shum, P., Grey, J., Fujiwara, S., Malhotra, G. S., González-Bonet, A. M., Moase, E., Allen, T. M., Thompson, D. H. (2012). Acid-labile mPEG-vinyl ether-1,2-dioleylglycerol lipids with tunable pH sensitivity: synthesis and structural effects on hydrolysis rates, DOPE liposome release performance and pharmacokinetics. *Molecular Pharmaceutics 9*, 3266–3276.
31. Kim, H.-K., Van den Bossche, J., Hyun, S.-H., Thompson, D. H. (2012). Acid-triggered release via dePEGylation of fusogenic liposomes mediated by heterobifunctional phenyl substituted vinyl ethers with tunable pH-sensitivity. *Bioconjugate Chemistry 23*, 2071–2077.
32. Nishiyama, N., Morimoto, Y., Wang, W. D., Kataoka, K. (2009). Design and development of dendrimer photosensitizer-incorporated polymeric micelles for enhanced photodynamic therapy. *Advanced Drug Delivery Reviews 61*, 327–338.
33. Kuruppuarachchi, M., Savoie, H., Lowry, A., Alonso, C., Boyle, R. W. (2011). Polyacrylamide nanoparticles as a delivery system in photodynamic therapy. *Molecular Pharmaceutics 8*, 920–931.
34. Huff, T. B., Tong, L., Zhao, Y., Hansen, M. N., Cheng, J.-X., Wei, A. (2007). Hyperthermic effects of gold nanorods on tumor cells. *Nanomedicine 2*, 125–132.
35. Loo, C., Lowery, A., Halas, N. J. J. W., Drezek, R. (2005). Immunotargeted nanoshells for integrated cancer imaging and therapy. *Nano Letters 5*, 709–711.
36. Thompson, D. H., Gerasimov, O. V., Wheeler, J. J., Anderson, V. C. (1996). Triggerable plasmalogen liposomes: improvement of system efficiency. *Biochimica et Biophysica Acta 1279*, 25–34.
37. Thompson, D. H., Rui, Y., Gerasimov, O. V, Triggered release from liposomes mediated by physically- and chemically-induced phase transitions, In *Vesicles: Surfactant Science Series*, ed. Rosoff, M., Marcel-Dekker, New York, NY, Vol. *62*, 1996, pp. 679–746.
38. Shum, P., Kim, J.-M., Thompson, D. H. (2001). Phototriggering of liposomal drug delivery systems. *Advanced Drug Delivery Reviews 53*, 273–284.
39. Gerasimov, O., Schwan, A., Thompson, D. H. (1997). Acid-catalyzed plasmenylcholine hydrolysis and its effect on bilayer permeability: a quantitative study. *Biochimica et Biophysica Acta 1324*, 200–214.
40. Keeffe, J. R., Kresge, A. J., In *The Chemistry of Enols*, ed., Rappoport, Z., Wiley, New York, NY, 1990, p. 437.
41. Saeva, F. D., Breslin, D. T., Luss, H. R. (1991). Intramolecular photoinduced rearrangements via electron transfer-induced, concerted bond cleavage and cation radical coupling. *Journal of the American Chemical Society 113*, 5333–5337.
42. Saeva, F. D., Martic, P. A., Garcia, E. (1993). Intramolecular photoinduced electron transfer (PET) bond cleavage in some sulfonium salt derivatives: effect of distance and thermodynamics on PET rate. *Journal of Physical Organic Chemistry 6*, 333–340.
43. Saeva, F. D., Garcia, E., Martic, P. A. (1995). Comparative photochemical behavior of some anthracenyl and naphthacenyl sulfonium salt derivatives. *Journal of Photochemistry and Photobiology A: Chemistry 86*, 149–154.
44. Allen, T. M., Hong, K., Papahadjopoulos, D. (1990). Membrane contact, fusion and hexagonal (HII) transitions in phosphatidylethanolamine liposomes. *Biochemistry 29*, 2976–2985.
45. Wymer, N., Gerasimov, O. V., Thompson, D. H. (1998) Cascade liposomal triggering: light-induced $Ca^{2+}$ release from plasmenylcholine liposomes triggers PLA2-catalyzed

hydrolysis and contents leakage from DPPC liposomes. *Bioconjugate Chemistry 9*, 305–308.

46. Collier, J. H., Hu, B.-H., Ruberti, J. W., Zhang, Z.-Y., Shum, P., Thompson, D. H., Messersmith, P. B. (2001). Thermally and photochemically triggered self-assembly of peptide hydrogels. *Journal of the American Chemical Society 123*, 9463–9464.
47. Zhang, Z.-Y., Shum, P., Yates, M., Messersmith, P. B., Thompson, D. H. (2002). Formation of fibrinogen-based hydrogels using phototriggerable diplasmalogen liposomes. *Bioconjugate Chemistry 13*, 640–646.
48. Sega, E. I., Low, P. S. (2008). Tumor detection using folate receptor-targeted imaging agents. *Cancer and Metastasis Reviews 27*, 655–664.
49. Kennedy, M. D., Jallad, K., Lu, J., Low, P. S., Ben-Amotz, D. (2003). Evaluation of folate conjugate uptake and transport by the choroid plexus of mice. *Pharmaceutical Research 20*, 714–719.

# MULTIFUNCTIONAL DENDRITIC NANOCARRIERS: THE ARCHITECTURE AND APPLICATIONS IN TARGETED DRUG DELIVERY

Ryan M. Pearson, Jin Woo Bae, and Seungpyo Hong

*Department of Biopharmaceutical Sciences, College of Pharmacy, University of Illinois, Chicago, IL, USA*

## 5.1 RECENT ADVANCES IN DENDRITIC NANOMATERIALS

Dendrimers are hyperbranched, monodispersed macromolecules with chemically well-defined structures. Synthesized by either the convergent or divergent method, a dendrimer consists of three functional domains (core, interior, and periphery) as shown in Figure 5.1. Each domain of a dendrimer can be tailored to precisely and easily control its molecular weight, size (hydrodynamic radius), and surface charge and functionality [1]. The highly controllable physical properties of dendrimers enable the fine-tuning of their biological properties, such as cytotoxicity, biodistribution, and intracellular uptake/trafficking/fate [2–4]. Furthermore, the large number of the surface groups also allows facile multifunctionalization of dendrimers through incorporation of a variety of functional molecules such as therapeutic and diagnostic moieties. The high deformability and the capability to precisely control their surface functional groups make dendrimers ideally suited for mediating strong multivalent interactions between ligands and receptors [5–7]. These unique properties of dendrimers have led these nanomaterials to become one of the most promising platforms for targeted drug delivery. Over the past two

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*,
Edited by Shayne C. Gad
Copyright © 2013 John Wiley & Sons, Inc.

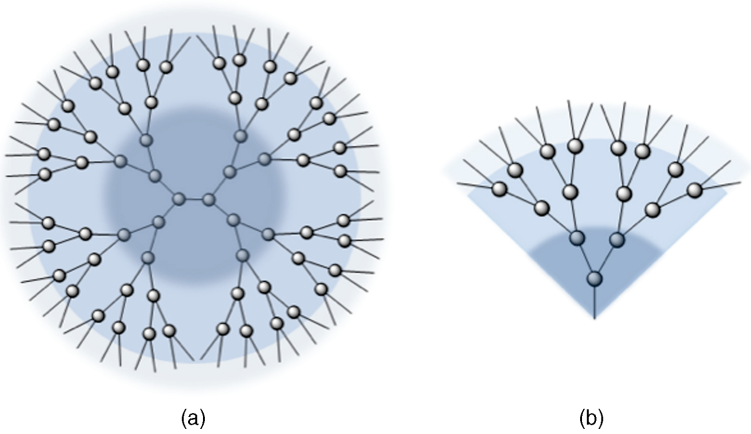


**FIGURE 5.1** General structures of a dendrimer (a) and dendron (b). Dark to light shaded regions represent the characteristic regions of the dendrimer (core, interior, and periphery) and characteristic regions of the dendron (focal point, interior, and periphery).

decades, significant progress has been made toward the development of dendrimer-based nanotherapeutics for cancer treatments, as highlighted in several recent reviews [8–12]. Although dendrimers have shown great promise in improving cancer treatments and diagnoses, these materials have some limitations that may negatively impact the outcomes of clinical trials and hinder their advancement to the clinic. The drawbacks include complex, multistep synthetic procedures that present scalability issues and limited options for carrying drug molecules, that is, covalent conjugation. Conjugation of drug molecules to the dendrimer often results in low drug payload, in addition to introducing unnecessary chemical modifications to drug molecules [1,13]. Furthermore, extensive surface conjugation with many functional molecules often induces the deterioration of the material properties and structural regularity of the dendrimer [5].

In parallel with the progress of dendrimers, dendron-based nanomaterials have also attracted a great deal of scientific interests due to their capability to hybridize the multifunctional and multivalent properties of dendrimers with various organic and inorganic materials [14]. Dendrons are monodisperse wedge-shaped, hyperbranched sections that comprise dendrimers. Their unique molecular architecture enables chemical modifications of the focal point and/or the periphery, allowing a multitude of complex structures to be realized [14,15]. The use of dendrons for cancer treatments has not yet reached the same level as dendrimers; however, because of their versatility and modular characteristics, they have rapidly emerged as a potential targeted drug delivery platform. In particular, amphiphilic dendrons capable of self-assembly can be engineered to form dendron-based copolymer micelles in aqueous solutions at polymer concentrations higher than the critical micelle concentration (CMC). These micelles can carry significantly higher numbers of drug molecules within their core than dendrimers, while maintaining multiple functional groups on their surface available for chemical modifications [16–18].

However, another level of structural complexity arises from their nonunimolecular structure. Optimization of each structural component is thus necessary because the morphology of self-assembled structures is difficult to predict and often has a marked effect on its biological properties [19].

In this chapter, we highlight promising dendritic systems and discuss their limitations and recent approaches to overcome those drawbacks. The discussion progresses through sections describing important architectural parameters used to rationally design various dendritic materials. Dendritic structure-related toxicity, biodistribution, cellular uptake, and multivalency are also described in detail. Additionally, we discuss through examples how dendrimers and dendrons have been designed and used as targeted drug delivery vehicles for cancer diagnosis and treatments in particular. We anticipate that dendritic materials possessing multifunctional architectures will have a significant impact on the way by which cancer is prevented, diagnosed, treated, and ultimately cured in the future.

## 5.2 EFFECTS OF DENDRITIC ARCHITECTURES ON BIOLOGICAL SYSTEMS

Many types of dendritic macromolecules have been developed spanning various fields since 1978 [20,21]. However, the clinical translation of these exciting materials has been disappointingly slow due to the issues related to nonspecific toxicity and biodistribution. To date, VivaGel$^{TM}$, a topically applied vaginal microbicide [22], is the only dendrimer-based product that has been approved by the FDA for the clinical trials. Other promising dendrimer-based therapeutics that are in preparation for clinical trials include Gadomer-17, a magnetic resonance imaging (MRI) contrast agent [23], and a colony-stimulating factor 1 receptor (CSF1R)-targeted dendrimer for the treatment of rheumatoid arthritis [24]. To design highly effective dendrimers, a number of parameters must be considered and optimized. In this section, we discuss how architectural parameters, such as generation (size), surface charge, and chemical composition, can be balanced to control dendrimer-related toxicity and biodistribution, and to achieve highly specific biological interactions.

### 5.2.1 Structure-Related Toxicity and Biodistribution

A review by Duncan and Izzo described the biocompatibility and structure-related toxicity of dendrimers with various backbone structures [25]. The most important factors found to affect the biocompatibility of dendrimers were surface charges and generation. Malik et al. studied the effects that dendrimer surface functionality and generation had on hemolysis and cell morphology [26]. Amine-terminated poly(amidoamine) (PAMAM) and poly(propylenimine) (PPI) dendrimers exhibited generation-dependent toxicity. The increased toxicity of higher generation dendrimers has been attributed to (i) the large number of surface groups at the same molar concentration, that is, greater charge density; and (ii) the increased radius of curvature, which increases the contact area at the dendrimer–cell

membrane interface [2]. The toxicity by the amine-terminated (positively charged) dendrimers can be substantially reduced by acetylation, carboxylation, or PEGylation (covering the surface with a hydrophilic nonfouling layer of poly (ethylene glycol) (PEG)) of the dendrimer surfaces [27]. Chen et al. modified melamine-based dendrimers to amine, guanidine, carboxylate, sulfonate, phosphate, and PEG-terminated ones [28]. The dendrimers with amine groups exhibited significantly higher cytotoxicity than all other types of dendrimers, whereas PEGylated dendrimers had significantly reduced cytotoxicity after 24 h incubation at a concentration of 10 mg/ml.

Amine-terminated dendrimers are generally known to cause cytotoxicity through destabilization of cell membranes because of the electrostatic interactions, but the underlying mechanism has not been fully understood. To study these nanoscale interactions between PAMAM dendrimers and lipid bilayers, Hong et al. used atomic force microscopy (AFM), a series of enzymatic assays, a fluorescence-activated cell sorter (FACS), and fluorescence microscopy [2,3,29]. For AFM observations *in situ*, 1,2-dimyristoyl-*sn*-glycero-3-phosphocholine (DMPC) was prepared as supported lipid bilayers and subsequently treated with various PAMAM dendrimers. It was observed that cationic dendrimers induced the formation of nanoscale holes in the supported DMPC lipid bilayers in a size-dependent manner. Generation 7 (G7) PAMAM dendrimers were able to initiate hole formation but G5 PAMAM dendrimers could only expand the size of pre-existing defects. Leakage of cytosolic proteins luciferase (Luc) and lactate-dehydrogenase (LDH) from KB and Rat2 cells also demonstrated that PAMAM dendrimers permeabilized cell membranes with size dependency, that is, G7 induced more enzyme leakage than G5 [2]. The membrane integrity was restored 2 h after the removal of G5 PAMAM dendrimers from the incubation medium, indicating the dendrimer-induced membrane permeabilization was reversible. On the other hand, acetamide-terminated (charge neutral) dendrimers neither caused defects nor induced enzyme leakage at the concentrations tested with either generation. This study provided direct evidence that amine-terminated PAMAM dendrimers could disrupt lipid bilayers in a concentration- and generation-dependent manner.

Given that the high toxicity and poor degradability of PAMAM and PPI dendrimers hinder their translation into clinical applications, several new dendrimers have been synthesized and tested as alternatives. For example, hydroxyl-terminated, biodegradable polyester dendrimers based on a 2,2-bis(hydroxymethyl) propionic acid (polyester) backbone were synthesized [30,31]. While $IC_{50}$ (half maximal inhibitory concentration) of amine-terminated dendrimers have been reported as low as 50 μg/ml after 72 h incubation [26], polyester dendrimers have shown much less toxicity, where 60% of the growth rate of B16F10 cells was maintained at a concentration of 20 mg/ml after 48 h incubation [32]. The relatively low toxicity of polyester dendrimers is attributed to hydroxyl groups on the surface, which likely decreased nonspecific cellular interactions [33]. To address the nondegradability of dendrimers, stimuli-responsive functional groups have been incorporated into the dendrimer structure that respond to enzymes, light, reduction, or pH to ensure degradation [34].

Biodistribution reflects the *in vivo* fate of dendrimers. The biodistribution of dendrimers has been traditionally studied in the context of diagnostic applications, since their multiple surface functional groups offer conjugation sites to create multimodal imaging probes [35]. As observed in numerous *in vitro* studies, the generation and surface properties of dendrimers largely affects their biodistribution profiles. Margerum et al. first reported a generation-dependent biodistribution of PAMAM dendrimers, radiolabeled with gadolinium (Gd) using MRI [36]. G5 dendrimers accumulated more preferentially in the liver after 7 days as compared to lower generations G2–4. However, the biodistribution of these Gd-dendrimer chelates cannot be universally applied to other dendrimers because the Gd ions may influence the molecular weight and surface composition. $^{125}$I-labeled cationic and anionic PAMAM dendrimers have shown structure and size dependent biodistribution profiles [26]. Anionic PAMAM dendrimers administered intravenously to Wistar rats at a concentration of 10 μg/ml showed a 10- to 20-fold longer circulation time than cationic PAMAM dendrimers. It was also noted that lower generation dendrimers exhibited longer circulation times. Both anionic and cationic dendrimers were found to primarily accumulate in the liver (30–90%) 1 h after administration, indicating that the dendrimer surface groups need to be modified to avoid hepatic uptake if tumor targeting is intended. In this regard, PEGylation has been exploited as a means to enhance the blood circulation half-life and decrease opsonization [37]. Kojima et al. have shown using lysine-modified PEGylated PAMAM dendrimers that the increased circulation time was most probably attributable to a large increase in molecular weight and full surface coverage of PEG on the dendrimer [38].

As for dendrons and dendron-based hybrid materials, it is difficult to make a general conclusion with respect to structure-related toxicity and biodistribution due to a large variety of the possible structures. For this reason, each type of dendron should be evaluated independently to ensure that it elicits appropriate biological responses. Because of the versatility and uniqueness of the structures that can be created, a vast number of different architectures have been synthesized [14]. To date, there have been only a few biodistribution studies of dendron-based materials. $^{111}$In-labeled linear dendritic block-copolymer micelles comprised of a hydrophobic polypeptide (β-benzyl-L-aspartate) conjugated to a carboxyl-terminated PEGylated G4 polyester dendron were evaluated in tumor-bearing nude mice [39]. The micelles were shown to accumulate to all vessel-rich organs initially, and most of the micelles were cleared from the body by day 5. Gillies et al. synthesized polyester "bow-tie" hybrid dendrimers comprised of two asymmetric polyester dendrons and evaluated their biodistribution profiles in both tumored and nontumored mice [40]. Polymers with molecular weights of 40,000 Da or greater were found to have significantly enhanced circulation times compared to lower molecular weight polymers. The degree of branching was found to play a role in determining circulation times as well. Higher generation bow-tie dendrimers circulated for a longer time and showed relatively less renal clearance, likely due to decreased molecular flexibility and difficulty to pass glomerular filtration [40].

### 5.2.2 Mechanisms for Cell Entry of Dendrimers

The ability of dendrimers to interact with cells and cross cell membranes is critical to the effective delivery of therapeutic and diagnostic payloads. Multiple pathways have been proposed as potential mechanisms of cell entry of dendrimers, which are largely affected by molecular size, shape, charge, and surface chemistry [41]. It is particularly important to understand the cellular uptake, intracellular trafficking, and fate of dendrimers when designing novel nanocarriers with subcellular targeting capabilities [42]. Many groups have revealed various mechanisms of PAMAM dendrimer internalization as summarized in Table 5.1. Cationic PAMAM dendrimers have high affinities for plasma membranes that are negatively charged because of proteoglycans and phospholipids, resulting in a nonspecific adsorptive uptake mechanism. Therefore, modification of the dendrimer surface is necessary to achieve fine control over their cellular interactions.

The mechanism by which dendrimers are internalized into cells is a subject of considerable debate. For example, in Caco-2 cells, cationic PAMAM and carboxylated PAMAM dendrimers were shown to internalize by a clathrin-dependent endocytosis mechanism [43]. These results were further confirmed in studies performed using HeLa cells, where cationic PAMAM and partially acetylated PAMAM dendrimers were internalized via both clathrin-mediated endocytosis and macropinocytosis [44]. In other studies, carboxylated PAMAM dendrimers were internalized into A549 lung epithelial cells by a caveolae-mediated pathway, and neutral or cationic PAMAM dendrimers were taken up by nonclathrin/noncaveolae-mediated mechanisms [45]. Recently, Hong et al. reported a systematic study regarding the cellular internalization mechanisms of G7 PAMAM dendrimers with various surface charges such as amine, acetyl, and carboxyl groups [3]. Fluorescently labeled cholera-toxin subunit B (CTB), transferrin (Tf), and ganglioside $GM_1$-pyrene were employed as endocytic markers to be co-incubated with PAMAM dendrimers. G7 PAMAM dendrimers were found to co-localize with CTB in KB and Rat2 cells whereas acetylated and carboxylated dendrimers did not interact with the cells. Co-localization with CTB is generally taken as indirect evidence for $GM_1$ interaction and a caveolae-mediated uptake mechanism since CTB does not internalize without $GM_1$. To further assess the role that $GM_1$ may have in facilitating the uptake of G7 PAMAM dendrimers, similar experiments were performed using C6 cells, which are $GM_1$ deficient. Interestingly, C6 cells that lack $GM_1$ still allowed G7 PAMAM dendrimers to be internalized, indicating that caveolae-mediated endocytosis is not the only mechanism for dendrimer internalization. Additionally, G7 PAMAM dendrimers were internalized into cells even at 4°C, suggesting that the internalization mechanism may not be energy-dependent. This study, together with the series of publications [2,3,29], presented strong evidence that nanoscale hole formation is a mechanism of cellular internalization for cationic dendrimers, in addition to other endocytotic mechanisms. Results of these studies strongly suggest that the mechanism of dendrimer internalization is multifaceted and depends on numerous factors including material properties (i.e., charge, generation), concentration, and cell type. It is most likely

**TABLE 5.1 Proposed Mechanisms for Cellular Internalization of Dendrimers**

| Materials | Endocytic Pathway | Cells | References |
|---|---|---|---|
| PAMAM G4-$(NH_2)_{64}$ | Cholesterol-dependent pathway | Caco-2 cells | [43] |
| PAMAM G2-$(NH_2)_{16}$ | Clathrin-dependent pathway | Caco-2 cells | [116] |
| PAMAM G2-$(NH_2)_{16}$<br>PAMAM G4-$(NH_2)_{64}$<br>PAMAM G6-$(NH_2)_{256}$ | Clathrin-dependent and macropinocytosis pathway (G6 > G4 > G2) | HeLa cells | [44] |
| PAMAM G4-$(NH_2)_{64}$<br>PAMAM G4-$(OH)_{64}$ | Nonclathrin, noncaveolae-mediated mechanism involving electrostatic interactions or other nonspecific fluid-phase endocytosis (G4-$NH_2$ > G3.5-COOH > G4-OH) | A549 lung epithelial cells | [45] |
| PAMAM G3.5-COOH | Caveolae-mediated endocytosis | A549 lung epithelial cells | [45] |
| PAMAM G3.5-$NH_2$ | Clathrin-dependent and dynamin-dependent pathway | Caco-2 cells | [117] |
| PAMAM G4-$(NH_2)_{64}$<br>PAMAM G3-$(NH_2)_{32}$<br>PAMAM G2-$(NH_2)_{16}$ | Cholesterol-dependent pathway (G4 >> G3 > G2) | B16F10 melanoma cells | [118] |
| PAMAM G3-$(NH_2)_{30}$-$(Propanolol)_2$ | Caveolae-dependent and macropinocytosis pathways (G3 >> G3P2) | HT-29 cells | [119] |
| PAMAM G3-$(NH_2)_{30}$-$(Lauroyl)_2$ | Caveolae-dependent, clathrin-dependent, and macropinocytosis pathways (G3L2 >> G3) | HT-29 cells | [119] |
| PAMAM G3-$(NH_2)_{28}$-$(Lauroyl)_2$-$(Propanolol)_2$ | Caveolae-dependent, and possibly clathrin-dependent, endocytosis pathways (G3 = G3L2P2) | HT-29 cells | [119] |
| PAMAM G5-$(NH_2)_{125.6}$-$(PEG\text{-}lactoferrin)_{2.4}$/pDNA complex | Clathrin-dependent endocytosis, caveolae-mediated endocytosis, and macropinocytosis | Brain capillary endothelial cells | [120] |
| PEPE G2-$(acrylate)_{16}$ | Caveolae-dependent, clathrin-dependent pathways | Brain vascular endothelial cells | [121] |
| PAMAM G7-$(NH_2)_{512}$ | Nanoscale hole formation | KB, Rat2, and C6 cells | [3] |

that multiple cellular uptake mechanisms are involved and govern the cellular uptake of dendrimers.

### 5.2.3 Multivalent Interactions

Multivalent interactions, which naturally occur in many biological systems, are characterized by simultaneous binding of multiple ligands to multiple receptors [46,47]. These interactions are critical to many pathological and physiological processes and can confer several advantages, such as improved and tight binding through amplification of an existing interaction, increased specificity of binding, efficient signaling communication with cells, and prevention of undesired interactions [46,48,49]. Therefore, understanding of multivalent ligand–receptor interactions is important for designing dendritic nanocarriers that are highly specific to target cells. Several possible mechanisms for multivalent interactions have been proposed. These include the chelate effect, subsite binding, steric stabilization, statistical rebinding, and receptor clustering [46,50]. In particular, the statistical effect (rebinding by a high local concentration of ligands) and the multivalent effect (multivalent binding between multiple ligands and receptors, which may lead to receptor clustering) have been mainly investigated [51,52]. Figure 5.2 depicts the two major mechanisms.

Early studies of multivalent targeting were primarily performed to understand carbohydrate–protein interactions and to gain insight into how proteins present on

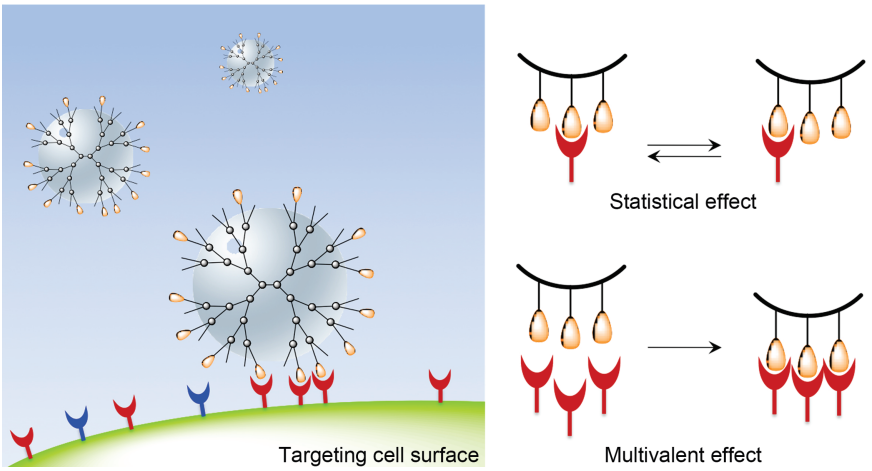


**FIGURE 5.2** Schematic representation of multivalent binding of ligand-conjugated dendrimers to target cell surfaces. Dendrimers functionalized with multiple copies of targeting ligands can bind to specific receptors overexpressed on cell surfaces via multivalent interactions. Multiple ligand presentation on the dendrimer surface can lead to the statistical binding effect if their complementary receptors are not clustered on the cell surface. Simultaneous binding of multiple ligands to multiple receptors is responsible for the multivalent effect, which can affect signaling functions of receptors.

cell surfaces can recognize carbohydrates. These studies focused on interactions of mannoside-concanavalin A (Con A) with pea lectin [53], sialic acid with CTB with heat labile enterotoxin of *E. coli* [54], sialic acid with virus hemagglutinin [55], lactose with CTB [56], and hyaluronan with CD44 [57]. Carbohydrate–protein interactions are dependent upon multiple factors including hydrogen bonding and hydrophobic interactions. Additionally, these interactions are known to be involved in a number of important biological phenomena such as cell–cell interactions, antibody–antigen interactions, and viral infection by strengthening their interactions through receptor clustering [58–60]. Since then, many researchers have attempted to establish a rational dendritic design that captures the essence of the effective multivalent binding observed between carbohydrates and proteins.

Some of the studies to understand how dendritic structures may affect multivalent interactions were performed using glycodendrimers exploiting the well-known carbohydrate–protein interactions. Glycodendrimers with 16 mannoside residues possess much higher binding avidities toward Con A and pea lectins compared to single carbohydrate residues [53]. However, fully mannose-substituted larger glycodendrimers (36-mer with three mannose residues per arm) did not achieve a significant increase in binding. It appeared to be a result of steric hindrance of neighboring carbohydrates and/or saturation of binding sites available on the cell surface [61]. It implies that there should be an optimal number of ligands that can participate in multivalent interactions. Andre et al. further supported the notion of an optimal ligand number using lactose-functionalized G5 PAMAM dendrimers [62]. These dendrimers displayed an optimal level of carbohydrate density for effective multivalent interactions with different proteins. Similar results have been reported elsewhere [63–67], providing additional evidence that multivalent binding is highly dependent on the number of surface ligands.

Further modifications to the dendrimer structure were shown to enhance the multivalent binding events of carbohydrates. Simple modifications of the dendrimer surface with short and long chemical linkers increased the conformational flexibility of carbohydrates bound to dendrimers, which might increase the mobility of ligands and probability of binding to its appropriate receptor. Flexible linkers can facilitate optimal geometric orientation of ligands toward binding sites in a thermodynamically favorable manner, compared to rigid linkers [68,69]. Although the effect of chemical nature and the length of linkers on multivalent interactions have not been fully understood, the use of a PEG linker has proven effective in inducing efficient multivalent binding [70,71].

In addition to the number and the flexibility of bound carbohydrates, a variety of architectural parameters such as size, shape, and density of binding elements can affect multivalent interactions. Gestwicki et al. explored the effect of ligand architecture on receptor binding and clustering by synthesizing a series of 28 mannose-containing ligands that varied in size, shape, and density [72]. These ligands were derived from five structurally different classes of materials (low-molecular weight compound, PAMAM dendrimer, bovine serum albumin, linear, and polydisperse polymers). The multivalent binding modes of the various mannosylated ligands with the lectin Con A were explored using high-throughput assays. As Con A can be clustered by

multivalent ligands (e.g., mannose), Con A is an excellent model protein for examining different aspects of multivalent binding, such as the number of clustered receptors, receptor-clustering rate, and receptor proximity. It was found that ligand architectures resulted in different degrees of Con A inhibition and clustering. Although dendrimers did not generate the most effective ligand architecture for either Con A inhibition or clustering, it was shown that the dendrimer with higher mannose content was a more effective inhibitor of Con A than the dendrimer with a lower mannose content. In addition, the rapid receptor-clustering rate from the highly functionalized, large dendrimers was observed. However, further investigation is necessary to determine how architectural parameters influence multivalent binding modes in other systems.

On the basis of these initial efforts on multivalent targeting, many researchers have sought ways to treat and detect cancer effectively by ligand-conjugated dendritic materials. There are several studies supporting that appropriate presentation of tumor targeting ligands on dendritic surfaces enables the nanocarriers to achieve high binding avidity and specificity toward cancer cells [5,7,73–76]. Hong et al. first reported a quantitative and systematic study of the multivalent effect using G5 PAMAM dendrimers with different numbers of folate as a targeting ligand [5]. The measurements using surface plasmon resonance (SPR) revealed that increasing the number of folate per dendrimer can dramatically enhance the binding avidity toward folate receptors (FRs) up to 170,000-fold, compared to free folic acid (FA). However, high-folate conjugation (over $\sim$7 folate molecules per dendrimer) compromised the material properties of the dendrimer (increased polydispersity index (PDI) and decreased solubility). Saturation behaviors were observed in cell-level experiments using FACS. When the number of folate per dendrimer was increased, dendrimer binding to FR-overexpressing KB cells was increased; however, the improvement became marginal when more than seven folates were conjugated to a dendrimer molecule. This study clearly demonstrated that an optimal number of targeting agents (five folate molecules per dendrimer in this particular case) needs to be identified, in order to achieve the maximal binding avidity without compromising materials properties.

Despite a number of studies on multivalent interactions, it is still unclear which architectural factors are most crucial to designing novel dendritic structures. Obviously, ligand number and flexibility of ligand presentation should be considered in optimizing multivalent interactions; however, evaluation of dendritic constructs with multivalent targeting should vary on a case-by-case basis depending on the types of ligands and target receptors.

## 5.3 FUNCTIONALIZATION OF DENDRITIC STRUCTURES VIA SURFACE MODIFICATION

Controlling biological properties of dendritic materials via surface modification has been well established as an effective method to create multifunctional architectures of dendritic materials. Depending on the purpose of each study, various molecules have been covalently conjugated to dendrimers at varying numbers and densities. To date, a variety of conjugation methods have been reported to introduce drugs,

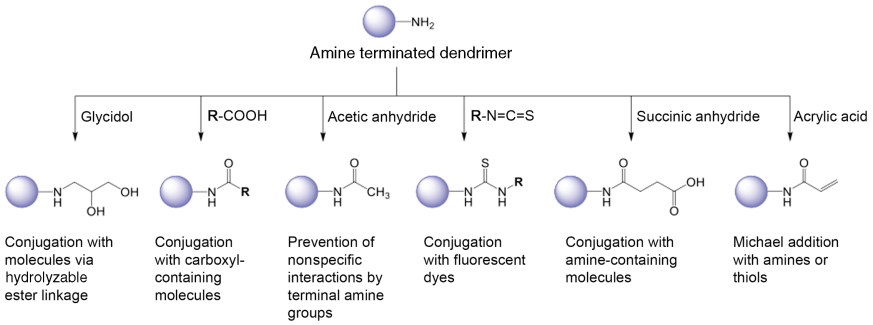


**FIGURE 5.3** Various surface modification methods of amine-terminated dendrimers. Starting from amine-terminated dendrimers, many types of surface functional groups can be easily introduced to dendrimers for further applications.

hydrophilic spacers, targeting molecules, and imaging modalities onto dendritic surfaces. Although the conjugation chemistry depends on the molecular functionality to be linked, ideally the method for surface modification should be simple and well defined without causing any loss of inherent reactant properties. Various reagents are commercially available for chemical coupling of molecules and dendritic materials. Figure 5.3 depicts frequently used methods for surface modification of amine-terminated dendrimers such as PAMAM, PLL, and PPI dendrimers. These strategies allow conjugation of a variety of functional groups onto the dendrimer surface, enabling further modification of dendrimers. As previously mentioned, PEGylation is one of the effective surface modification methods that can increase the blood circulation time of dendritic materials through prevention of nonspecific interactions with biological substances. In general, two types of PEGylation strategies have been widely adopted. *N*-Hydroxysuccinimide (NHS)- and *p*-nitrophenyl chloroformate (*p*-NPC)-based chemistries have been commonly used for water- and organic solvent-based reactions, respectively. Zhu et al. compared the PEGylation efficiency of G3 PAMAM dendrimers using these two strategies [77]. Dendrimers PEGylated using *p*-NPC-PEG achieved significantly higher levels of PEGylation compared to those using PEG–NHS [78].

## 5.4 DENDRIMER-BASED THERAPEUTIC AND DIAGNOSTIC APPROACHES

Various dendrimer-based nanocarriers have been extensively exploited for effective cancer treatment. Table 5.2 lists the most recent achievements in cancer therapy and diagnosis. The surface properties of the dendrimers were rationally modified with bioactive entities (drugs, genes, targeting ligands, and imaging agents) and their *in vivo* efficacy was studied. In this section, we discuss how dendritic properties affect *in vivo* efficacy using examples in three major delivery applications: anticancer drugs, genes, and imaging agents.

**TABLE 5.2 Recent Progress in Cancer Research Using Various Dendrimers**

| Purpose of Study | Dendritic Materials | Therapeutics or Diagnostics | References |
|---|---|---|---|
| Folate-mediated targeting of conjugated drugs | PAMAM dendrimer | Methotrexate, paclitaxel | [86,112] |
| Tumor accumulation by different PEGylation degrees and drug conjugation style | PAMAM dendrimer | Doxorubicin | [77] |
| Convection enhanced delivery of boronated dendrimer-epidermal growth factor bioconjugate for boron neutron capture therapy | PAMAM dendrimer | Boron | [122] |
| Folate-mediated targeting of drug encapsulated PEG-dendrimer | PAMAM dendrimer | 5-Fluorouracil | [114] |
| Targeted siRNA delivery of cRGD conjugated dendrimers through spheroid model of malignant glioma | PAMAM dendrimer | siRNA | [123] |
| *In vivo* tumor imaging by activatable cell penetrating peptide-coated dendrimer labeled with Cy5 and gadolinium | PAMAM dendrimer | Cy5, gadolinium | [124] |
| Folate-mediated *in vivo* tumor imaging using gadolinium-conjugated dendrimer | PAMAM dendrimer | Gadolinium | [125] |
| *In vivo* imaging using multimodal and multicolor nanoprobes | PAMAM dendrimer | $^{111}$In, Cy5, Alexa 660, 680, 700, and 750 | [126] |
| Systemic antiangiogenic activity of dendrimer itself | PLL dendrimer | — | [127] |
| Delivery of anticancer drug using PEG-dendrimer | PLL dendrimer | Camptothecin | [113] |
| *In vivo* tumor imaging using gadolinium-conjugated dendrimer | PLL dendrimer | Gadolinium | [128] |
| Targeted siRNA delivery using cross linked PEG-dendrimers with LHRH peptide | PPI dendrimer | siRNA | [129] |
| *In vivo* visualization of gene transfer mediated by systemic injection of dendrimer-pDNA complex | PPI dendrimer | pDNA | [130] |
| *In vivo* imaging and biodistribution of radiolabeled dendrimer | PE dendrimer | $^{99m}$Tc | [131] |
| Biodistribution of drug-conjugated dendrimer and comparison with drug-loaded liposome | PE dendrimer | Doxorubicin | [132] |
| Noninvasive imaging of angiogenesis using PEG-dendrimer coated with cRGD | PE dendrimer | $^{76}$Br | [133] |
| Targeting of glial tumor and enhanced permeability across blood brain barrier using glucosylated dendrimer | PEPE dendrimers | Methotrexate | [134] |
| Photodynamic therapy using dendrimer phthalocyanine-encapsulated polymeric micelles | PAE dendrimer | Phthalocyanine | [135] |
| *In vivo* efficacy of conjugated drugs using PEG-dendrimer | PEA dendrimer | Doxorubicin | [136] |

PAMAM, polyamidoamine; PLL, poly(L-lysine); PPI, poly(propylenimine); PE, polyester; PEPE, polyether-copolyester; PAE, poly(acryl ether); PEA, poly(ester-amide); LHRH, luteinizing hormone-releasing hormone; cRGD, cyclic (Arg–Gly–Asp) peptide.

### 5.4.1 Delivery of Anticancer Drugs

Dendrimers have been investigated on numerous occasions for the delivery of small molecule anticancer drugs. Through implementation of rational design parameters, their efficacy has been reported to be superior to traditional chemotherapeutics. There are two major methods utilized to incorporate anticancer drugs with the dendrimer [18]. The first method is through encapsulation [79–81], where drugs noncovalently interact with the nonpolar interior core of the dendrimer forming a host-guest complex; and the second, drugs can be covalently attached to one of the many surface groups at the dendrimer periphery [82–84].

As drug encapsulation into dendrimers has several limitations such as low drug loading and uncontrollable drug release, conjugation of therapeutics to dendrimers has been preferred and well established as a means to increase drug solubility, circulating time, and accumulation in tumor tissue *in vivo* [9]. However, it should be noted that there is a limit to the number of molecules that can be conjugated to the surface of dendrimers to maintain structural regularity and solubility [5]. Ideal dendrimer-based carriers should mitigate any surface charges that cause cytotoxicity, possess an imaging capability for *in vivo* visualization, and carry multiple synergistic bioactive compounds to treat the target disease. Figure 5.4 illustrates a set of dendrimer-based

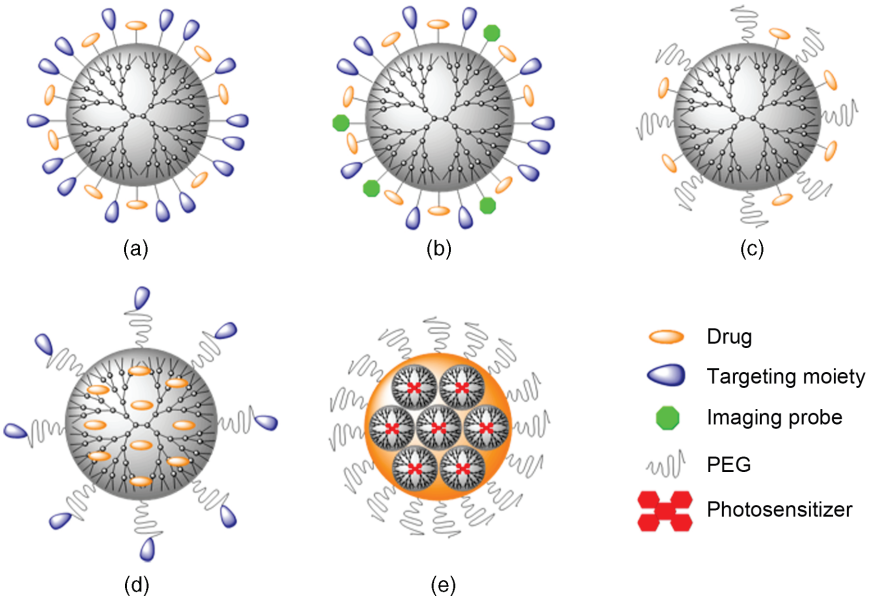


**FIGURE 5.4** Recently studied dendrimer-based anticancer drug delivery systems. (a) PAMAM dendrimer–folic acid–methotrexate [111], (b) PAMAM dendrimer–folic acid–paclitaxel–FITC [112], (c) camptothecin conjugated PLL dendrimer–PEG [113], (d) 5-fluorouracil encapsulated PAMAM dendrimer–PEG–folic acid [114], (e) phthalocyanine dendrimer encapsulated PEG–PLL micelle [115].

**BOX 5.1 IMPORTANT DESIGN CONSIDERATIONS FOR THE DEVELOPMENT OF DENDRIMER-BASED THERAPEUTICS**

**Structure**

1. *Generations* determine sizes and numbers of surface functional groups, which affects toxicity.
2. *Surface charges* determine cellular interactions and toxicity.
3. *Backbone compositions* determine degradation rates and drug encapsulation potential.
4. *Cores* determine branching degrees and impart potential to add additional functionalities.

**Function**

1. *Targeting ligands* enable selectivity and promote cellular uptake.
2. *Multivalent binding* substantially improves binding avidity by orders of magnitude.
3. *Stimuli-responsiveness* enhances drug delivery through site-specific degradation.
4. *Therapeutic agents* allow killing of target cells.
5. *Imaging agents* enable visual tracking of dendritic nanocarriers *in vitro* and *in vivo*.
6. *PEGylation* reduces nonspecific interactions and increases *in vivo* circulation time.

nanocarriers used for anticancer therapy that have had successful pre-clinical results. Here we will describe some of the most recent progress in dendrimer-based anticancer therapeutics with the aim to provide a set of criteria (summarized in Box 5.1) that should be considered in designing novel dendrimer-based therapeutics.

As noted earlier, coating the surface of dendrimers with a PEG layer is a well-established method used to decrease recognition by the reticuloendothelial system (RES) and to increase the circulation time, thereby enhancing drug delivery efficiencies. Recently, Lim et al. studied a series of PEGylated G2 triazine dendrimers with 12 paclitaxel molecules conjugated through either cleavable ester or ester/disulfide linkages using PC-3 prostate cancer xenografted SCID mice [85]. These dendrimers were well tolerated *in vivo* and could be administered at doses up to 200% of the maximum tolerated dose of free paclitaxel. However, low-level tumor localization and reduced circulation times were observed, suggesting that conjugation of multiple drug molecules may negatively influence the *in vivo* properties of the parent dendrimers. Thus, a key factor in designing dendrimers for drug delivery is to optimize the number of conjugated drug molecules so that important features such as tumor localization and circulating half-life times are minimally deteriorated. To

further enhance the specific delivery of drug molecules from the dendrimer to the tumor site, environmentally sensitive dendrimer-drug linkages have been incorporated. Zhu et al. compared various degrees of PEGylation of PAMAM dendrimers conjugated to doxorubicin (DOX) through acid-sensitive *cis*-aconityl and acid-insensitive amide linkages [77]. Although increasing degrees of PEGylation resulted in a decrease in cytotoxicity, higher PEGylation led to lower cellular uptake of DOX-conjugated dendrimers. At the same PEGylation degree, the acid-labile DOX–dendrimer conjugates were more effective in killing cells than the amide-linked DOX-dendrimers because of the facilitated release of DOX in the lysosomes. *In vivo* evaluation showed that the acid-labile DOX–dendrimers had an improved antitumor activity with an increase in the PEGylation degree because of longer circulation time and more effective tumor accumulation. These results show that the optimally designed conjugation of the drug to a PEGylated dendrimer via a cleavable bond may be beneficial for the treatment of a solid tumor.

One major drawback of dendrimer-based drug delivery systems is their multistep synthesis, which is difficult to scale up. Multifunctional dendrimer-based therapeutics can still be produced with batch-to-batch consistency and specificity toward cancer cells [86]. To improve the processing speed of multifunctional dendrimers, a "one-pot" synthesis method was devised where G5 PAMAM dendrimers were conjugated to predetermined ratios of methotrexate (MTX) and FA. Utilizing this simple approach, similar therapeutic results were obtained to those based on the multistep approach [87]. Although the creation of multifunctional dendrimers is complex, these innovative designs, along with the novel synthetic methodologies, make the dendrimer-based nanocarriers highly promising to reach clinical implementation.

With the rapid advances in genetic analysis using DNA-microarray technologies, the conventional approach to cancer therapy has shifted toward a personalized approach [88]. Gene-expression profiling can now be used to identify prognostic and predictive profiles for each patient, which can give clinicians vital information about which course of treatment would be most advantageous. Furthermore, investigation of cancer-specific surface markers and development of genome-wide association studies provide the tools necessary to bring cancer therapy to a next level by identifying genetic predispositions and their influences on the progression of a disease state [89]. To ultimately succeed in clinical translation, dendrimer-based cancer therapeutics must adapt to utilize the information provided from these biochemical studies.

### 5.4.2 Gene Delivery

The transfer of generic material into a cell to induce, inhibit, or replace gene expression is an additional therapeutic strategy that can be utilized as an effective option for the treatment of multiple diseases. Traditionally, viruses have been used due to their high transfection efficiencies; however, owing to the many safety concerns (potential for mutagenicity, oncogenicity, immunogenicity, and cytotoxicity), a lack of target-specificity, and high-production costs, alternatives that are safe and yet effective are required [90,91]. Cationic nonviral gene delivery vectors that form polyplexes with negatively charged genes are viable options due to their low

cost of preparation, low immunogenicity, and pathogenicity [92]. Unfortunately, the transfection efficiency of nonviral vectors is much lower than that of viruses, and the presence of a large positive charge density often induces nonspecific cytotoxicity. Extensive optimization in nonviral vectors is thus necessary.

PAMAM dendrimers were first reported to mediate the transfection of mammalian cells in culture by Haensler and Szoka [93]. Since then, researchers have designed numerous dendrimers to deliver a wide range of nucleic acids such as antisense oligonucleotides [94], microRNA [95], small interfering RNA [96], and plasmid DNA [97]. In a follow-up study by Szoka, "activated" dendrimers, or dendrimers partially degraded by solvolysis, had a greater than 50-fold enhancement in transfection efficiencies compared to intact dendrimers, which was attributed to an increase in flexibility [98]. Kuo et al. measured the expression of 44,000 genes found in HeLa cells treated with "activated" or intact dendrimers alone or in complex with pDNA [99]. Interestingly, it was found that the differences in molecular architecture induced almost completely different gene-expression profiles. Not surprisingly, surface modification and/or spatial modulation of the positive charges can enhance transfection efficiencies. Kim et al. synthesized G2 PPI dendrimers terminated with arginine. At concentrations as high as 150 μg/ml, a significantly improved transfection efficiency (by fourfold) was obtained over control polyethylenimine (PEI25K), while maintaining 80–90% cell viability [100]. Another group attempted to remove the positive surface charge by synthesizing a PAMAM-like dendrimer with internally quaternized amines and a fully acetylated surface [101]. The modification of the surface and addition of positively charged internal groups drastically reduced the cytotoxicity and enhanced cellular uptake. Surface modification of dendrimers with targeting ligands allows highly effective delivery of genes to their target site. Huang et al. developed a Tf-conjugated PAMAM-PEG dendrimer for brain targeting that showed 2.25-fold higher brain uptake than PAMAM and PAMAM-PEG *in vivo* [102]. Additionally, some groups have designed stimuli responsive dendrimers to increase cellular uptake and transfection efficiency. When short, highly charged oligocations were attached to a PEI-based dendrimeric vector via acid-cleavable linkers, a nearly fivefold increase in transfection efficiency was achieved without increasing cytotoxicity [103]. Rapid progression toward the development of better, more biocompatible, highly effective dendrimer-based gene delivery vectors illuminates the promising impact that dendrimers will have in the future for gene therapy applications.

### 5.4.3 Imaging

The necessity to obtain accurate imaging for *in vitro* and *in vivo* applications has led to the development of many novel nanoparticle-based imaging agents that can be utilized with current micro- and macroscale imaging technologies [104]. The inclusion of imaging agents for nanoparticle tracking and visualization of tumor sites is of great interest to both preclinical researchers and clinicians. In particular, dendrimer-based imaging agents have potential to deliver bioactive compounds and imaging agents simultaneously. Fluorescent imaging agents such as rhodamines,

fluoresceins, and AlexaFluors have been conjugated to dendrimers to allow visualization in cell culture and within tumor sections using fluorescence/confocal microscopy. For *in vivo* imaging, there is a need to develop highly effective and safe contrast agents for imaging that can be used for early detection using MRI. Dendrimers are well suited for the uses as novel MRI contrast agents because of their multivalency, monodisperse characteristics, and relatively large size, compared to low molecular weight gadolinium (Gd) chelates [105]. Several groups have modified the surface of dendrimers with Gd(III) chelates (dendrimer–Gd(III)) and found large proton relaxation enhancements and high relaxivities. For example, PEGylated dendrimers were modified with a tumor-targeting peptide HAIYPRH and modified to chelate Gd(III) [106]. It was found that 98 Gd(III) ions could be loaded per PAMAM dendrimer molecule and that this system could increase the diagnostic efficiency to identify liver cancer. Luo et al. synthesized a set of multifunctional peptide dendritic probes of differing generations that utilized a Gd(III) chelate for imaging and clustered galactose moieties for targeting to liver hepatocytes [107]. The inclusion of targeting ligands with dendrimer-based imaging agents represents a method that can be used to increase image contrast and specificity as opposed to nontargeted molecules. The multifunctionality is a clear advantage of the dendrimers, which allows them to be developed as multifunctional imaging agents that both enhance and expedite the detection and diagnosis of cancers.

## 5.5 DENDRON-BASED HYBRID NANOMATERIALS

It is clear that dendrimers possess many advantageous properties for developing novel therapeutics. Along those lines, each individual dendritic wedge-shaped section, or dendron, that comprises the dendrimer can be used individually or hybridized with other materials to allow creation of a multitude of architectures. Dendrons possess the unique ability to merge the highly desirable properties of dendrimers with the advantages of linear-block copolymers. Unlike the unimolecular structured dendrimers, when engineered with an appropriate hydrophile-lipophile balance (HLB) [108], dendron-based copolymers form self-assembled supramolecular structures. For drug delivery, a variety of dendron-based architectures have been developed and are depicted in Figure 5.5.

Some of the most promising types of dendron-based materials are “bow-tie” hybrid dendrimers [40], linear dendritic block copolymers [16], and dendrimersomes [109]. These materials possess many highly favorable characteristics necessary for drug delivery over comparable dendrimers and linear-block copolymers. First, dendron-based copolymer micelles can overcome the low drug payload disadvantage of dendrimer systems because physical encapsulation into the core is an additional option for the micellar systems. Note that the stability and controllability are the issues in dendrimer encapsulation and the *in vivo* feasibility of the so-called host–guest system is questionable [110]. Encapsulation itself has a significant advantage that there is no need to chemically modify the structure of the drug molecule. Second, the presence of the dendron and its multiple surface functional groups allow

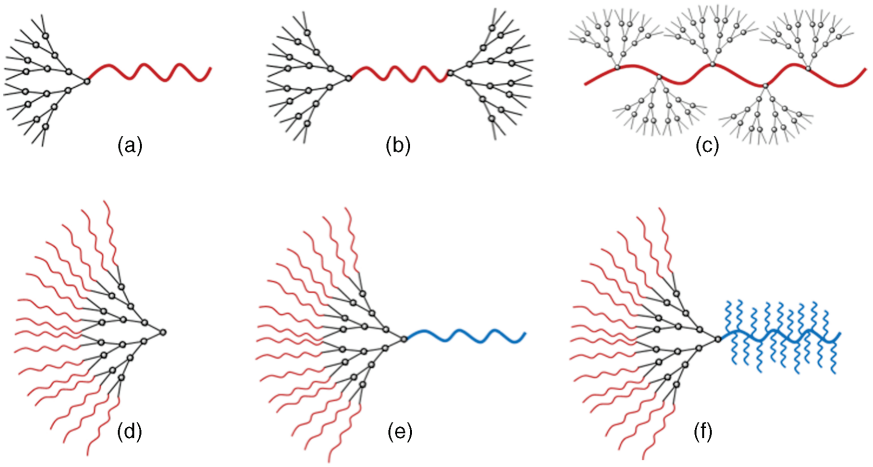


**FIGURE 5.5** A variety of dendron-based architectures for drug delivery. (a) A linear polymer linked to the focal point of a dendron, (b, c) a linear polymer bearing two or multiple dendrons at the end or backbone, (d–f) multiple polymer chains linked to the focal point or the periphery of a dendron. The HLB of dendrons can be engineered by controlling the ratio of hydrophilic to hydrophobic polymers.

multivalent interactions to be integrated. Poon et al. synthesized a micellar drug delivery system consisting of a PEGylated polyester dendron block and a rigid hydrophobic linear polypeptide core-forming block. Multivalent targeting was achieved through incorporation of differing degrees of FA on the dendron surface. These micelles selectively accumulated in xenografted tumors and delivered paclitaxel with a fourfold enhancement and reduced systemic toxicity [17,39]. Third, the dendritic architecture can substantially enhance the thermodynamic stability of self-assembled structures as measured by the CMC. Linear-dendron based copolymers have been reported to self-assemble at extremely low CMCs and at high HLBs. To investigate this phenomenon, we previously synthesized highly PEGylated dendron coils (PDCs) and compared their self-assembly behaviors to linear-block copolymers at similar HLBs [16]. The results showed that PDCs at similar HLBs had CMCs that were 1–2 orders of magnitude lower than those of the linear-block copolymer counterparts. The decrease in CMC is attributable to the preorganized, conical architecture of PDC, which decreases entropic cost during self-assembly [16]. Recently, Percec et al. synthesized 11 polymer libraries consisting of over 100 amphiphilic Janus dendrimers and found that they could self-assemble into highly monodispersed supramolecular structures denoted as “dendrimersomes” [109]. These dendrimersomes combine the mechanical strength of polymersomes and the biological mimicry of liposomes, and their highly branched structures offers the ease of surface functionalization. Even though a vast number of distinct morphologies have been reported in the current literature, the exact parameters that govern the phase-separation behavior and self-assembly of these dendron-based nanomaterials have not been fully understood.

## 5.6 FUTURE PERSPECTIVES

Dendrimers and dendron-based materials are one of the most promising platforms used for drug delivery to date. Extensive *in vitro* and *in vivo* research including those presented in this chapter has illuminated the bright future of dendrimers for drug delivery. Drawbacks of dendrimers, such as limited drug payload and toxicity, can be overcome by engineering the dendrimer structure (e.g., PEGylation), modifying the surface groups (e.g., charge neutralization), and incorporation of functional moieties (e.g., targeting, imaging, therapeutic agents). The choice of dendrimer generation (size) is an additional factor to consider when designing new dendrimer-based drug delivery platforms, because it determines the number of surface functional groups available for conjugation of molecules as well as cellular interactions. However, key challenges, such as limited payload and complex synthetic steps, still remain for dendrimers to be clinically implemented.

Dendron-based nanomaterials have emerged for their potential application in drug delivery. Hybridization of dendrons and linear polymers confers many additional benefits that are not seen in either system alone. This combination of dendritic and linear polymer melds high drug loading potential with multifunctional possibilities and multivalent interactions. However, the dendron systems are relatively new and need to undergo extensive validation processes to prove their potential as effective nanocarriers. Nonetheless, the dendritic structures clearly have significant advantages over the linear polymer counterparts and therefore hold a great promise for various delivery applications. With an optimized design considering the backbone structure, size, surface groups, and multifunctionalities, dendritic nanomaterials will undoubtedly play a key role as a versatile delivery vector in the new era of personalized medicine.

## REFERENCES

1. Lee, C. C., MacKay, J. A., Frechet, J. M. J., Szoka, F. C. (2005). Designing dendrimers for biological applications. *Nature Biotechnology 23*, 1517–1526.
2. Hong, S. P., Bielinska, A. U., Mecke, A., Keszler, B., Beals, J. L., Shi, X. Y., Balogh, L., Orr, B. G., Baker, J. R., Holl, M. M. B. (2004). Interaction of poly(amidoamine) dendrimers with supported lipid bilayers and cells: hole formation and the relation to transport. *Bioconjugate Chemistry 15*, 774–782.
3. Hong, S., Rattan, R., Majoros, I. J., Mullen, D. G., Peters, J. L., Shi, X., Bielinska, A. U., Blanco, L., Orr, B. G., Baker, Jr., J. R., Holl, M. M. (2009). The role of ganglioside GM1 in cellular internalization mechanisms of poly(amidoamine) dendrimers. *Bioconjugate Chemistry 20*, 1503–1513.
4. Leroueil, P. R., Hong, S., Mecke, A., Baker, Jr., J. R., Orr, B. G., Banaszak Holl, M. M. (2007). Nanoparticle interaction with biological membranes: does nanotechnology present a Janus face? *Accounts of Chemical Research 40*, 335–342.
5. Hong, S., Leroueil, P. R., Majoros, I. J., Orr, B. G., Baker, J. R., Banaszak Holl, M. M. (2007). The binding avidity of a nanoparticle-based multivalent targeted drug delivery platform. *Chemistry & Biology 14*, 107–115.

6. Kukowska-Latallo, J. F., Candido, K. A., Cao, Z., Nigavekar, S. S., Majoros, I. J., Thomas, T. P., Balogh, L. P., Khan, M. K., Baker, J. R. (2005). Nanoparticle targeting of anticancer drug improves therapeutic response in animal model of human epithelial cancer. *Cancer Research 65*, 5317–5324.
7. Myung, J. H., Gajjar, K. A., Saric, J., Eddington, D. T., Hong, S. (2011). Dendrimer-mediated multivalent binding for the enhanced capture of tumor cells. *Angewandte Chemie International Edition 50*, 11769–11772.
8. Wolinsky, J. B., Grinstaff, M. W. (2008). Therapeutic and diagnostic applications of dendrimers for cancer treatment. *Advanced Drug Delivery Reviews 60*, 1037–1055.
9. Tekade, R. K., Kumar, P. V., Jain, N. K. (2009). Dendrimers in oncology: an expanding horizon. *Chemical Reviews 109*, 49–87.
10. Fox, M. E., Szoka, F. C., Frechet, J. M. (2009). Soluble polymer carriers for the treatment of cancer: the importance of molecular architecture. *Accounts of Chemical Research 42*, 1141–1151.
11. Menjoge, A. R., Kannan, R. M., Tomalia, D. A. (2010). Dendrimer-based drug and imaging conjugates: design considerations for nanomedical applications. *Drug Discovery Today 15*, 171–185.
12. Cheng, Y., Zhao, L., Li, Y., Xu, T. (2011). Design of biocompatible dendrimers for cancer diagnosis and therapy: current status and future perspectives. *Chemical Society Reviews 40*, 2673–2703.
13. Patri, A. K., Kukowska-Latallo, J. F., Baker, J. R. (2005). Targeted drug delivery with dendrimers: comparison of the release kinetics of covalently conjugated drug and non-covalent drug inclusion complex. *Advanced Drug Delivery Reviews 57*, 2203–2214.
14. Rosen, B. M., Wilson, C. J., Wilson, D. A., Peterca, M., Imam, M. R., Percec, V. (2009). Dendron-mediated self-assembly, disassembly, and self-organization of complex systems. *Chemical Reviews 109*, 6275–6540.
15. Wurm, F., Frey, H. (2011). Linear-dendritic block copolymers: the state of the art and exciting perspectives. *Progress in Polymer Science 36*, 1–52.
16. Bae, J. W., Pearson, R. M., Patra, N., Sunoqrot, S., Vukovic, L., Kral, P., Hong, S. (2011). Dendron-mediated self-assembly of highly PEGylated block copolymers: a modular nanocarrier platform. *Chemical Communications 47*, 10302–10304.
17. Poon, Z., Chen, S., Engler, A. C., Lee, H. -i., Atas, E., von Maltzahn, G., Bhatia, S. N., Hammond, P. T. (2010). Ligand-clustered "Patchy" nanoparticles for modulated cellular uptake and *in vivo* tumor targeting. *Angewandte Chemie International Edition 49*, 7266–7270.
18. D'Emanuele, A., Attwood, D. (2005). Dendrimer–drug interactions. *Advanced Drug Delivery Reviews 57*, 2147–2162.
19. Whitesides, G. M., Grzybowski, B. (2002). Self-assembly at all scales. *Science 295*, 2418–2421.
20. Buhleier, E., Wehner, W., Vogtle, F. (1978). Cascade-chain-like and nonskid-chain-like syntheses of molecular cavity topologies. *Synthesis-Stuttgart*, 155–158.
21. Tomalia, D. A., Baker, H., Dewald, J., Hall, M., Kallos, G., Martin, S., Roeck, J., Ryder, J., Smith, P. (1985). A new class of polymers: starburst-dendritic macromolecules. *Polymer Journal 17*, 117–132.

22. Bernstein, D. I., Bourne, N., Ayisi, N. K., Ireland, J., Matthews, B., McCarthy, T., Sacks, S. (2003). Evaluation of formulated dendrimer SPL7013 as a microbicide. *Antiviral Research 57*, A66–A66.
23. Herborn, C. U., Barkhausen, J., Paetsch, I., Hunold, P., Mahler, M., Shamsi, K., Nagel, E. (2003). Coronary arteries: contrast-enhanced MR imaging with SH L 643A—experience in 12 volunteers. *Radiology 229*, 217–223.
24. Hayder, M., Poupot, M., Baron, M., Nigon, D., Turrin, C. -O., Caminade, A. -M., Majoral, J. -P., Eisenberg, R. A., Fournie, J. -J., Cantagrel, A., Poupot, R., Davignon, J. -L. (2011). A phosphorus-based dendrimer targets inflammation and osteoclastogenesis in experimental arthritis. *Science Translational Medicine 3*, 81ra35.
25. Duncan, R., Izzo, L. (2005). Dendrimer biocompatibility and toxicity. *Advanced Drug Delivery Reviews 57*, 2215–2237.
26. Malik, N., Wiwattanapatapee, R., Klopsch, R., Lorenz, K., Frey, H., Weener, J. W., Meijer, E. W., Paulus, W., Duncan, R. (2000). Dendrimers: relationship between structure and biocompatibility *in vitro*, and preliminary studies on the biodistribution of $^{125}$I-labelled polyamidoamine dendrimers *in vivo*. *Journal of Controlled Release 65*, 133–148.
27. Jain, K., Kesharwani, P., Gupta, U., Jain, N. K. (2010). Dendrimer toxicity: let's meet the challenge. *International Journal of Pharmaceutics 394*, 122–142.
28. Chen, H. T., Neerman, M. F., Parrish, A. R., Simanek, E. E. (2004). Cytotoxicity, hemolysis, and acute *in vivo* toxicity of dendrimers based on melamine, candidate vehicles for drug delivery. *Journal of the American Chemical Society 126*, 10044–10048.
29. Hong, S., Leroueil, P. R., Janus, E. K., Peters, J. L., Kober, M. M., Islam, M. T., Orr, B. G., Baker, Jr., J. R., Banaszak Holl, M. M. (2006). Interaction of polycationic polymers with supported lipid bilayers and cells: nanoscale hole formation and enhanced membrane permeability. *Bioconjugate Chemistry 17*, 728–734.
30. Ihre, H., Padilla De Jesus, O. L., Frechet, J. M. (2001). Fast and convenient divergent synthesis of aliphatic ester dendrimers by anhydride coupling. *Journal of the American Chemical Society 123*, 5908–5917.
31. Ihre, H. R., Padilla De Jesus, O. L., Szoka, Jr., F. C., Frechet, J. M. (2002). Polyester dendritic systems for drug delivery applications: design, synthesis, and characterization. *Bioconjugate Chemistry 13*, 443–452.
32. Padilla De Jesus, O. L., Ihre, H. R., Gagne, L., Frechet, J. M., Szoka, Jr., F. C., (2002). Polyester dendritic systems for drug delivery applications: *in vitro* and *in vivo* evaluation. *Bioconjugate Chemistry 13*, 453–461.
33. Winnicka, K., Bielawski, K., Rusak, M., Bielawska, A. (2009). The effect of generation 2 and 3 poly(amidoamine) dendrimers on viability of human breast cancer cells. *Journal of Health Science 55*, 169–177.
34. Kojima, C. (2010). Design of stimuli-responsive dendrimers. *Expert Opinion on Drug Delivery 7*, 307–319.
35. Kobayashi, H., Brechbiel, M. W. (2003). Dendrimer-based macromolecular MRI contrast agents: characteristics and application. *Molecular Imaging 2*, 1–10.
36. Margerum, L. D., Campion, B. K., Koo, M., Shargill, N., Lai, J. J., Marumoto, A., Sontum, P. C. (1997). Gadolinium(III) DO3A macrocycles and polyethylene glycol coupled to dendrimers—effect of molecular weight on physical and biological properties

of macromolecular magnetic resonance imaging contrast agents. *Journal of Alloys and Compounds 249*, 185–190.

37. Harris, J. M., Chess, R. B. (2003). Effect of pegylation on pharmaceuticals. *Nature Reviews Drug Discovery 2*, 214–221.
38. Kojima, C., Regino, C., Umeda, Y., Kobayashi, H., Kono, K. (2010). Influence of dendrimer generation and polyethylene glycol length on the biodistribution of PEGylated dendrimers. *International Journal of Pharmaceutics 383*, 293–296.
39. Poon, Z., Lee, J. A., Huang, S., Prevost, R. J., Hammond, P. T. (2011). Highly stable, ligand-clustered patchy "micelle" nanocarriers for systemic tumor targeting. *Nanomedicine: Nanotechnology, Biology, and Medicine 7*, 201–209.
40. Gillies, E. R., Dy, E., Frechet, J. M. J., Szoka, F. C. (2005). Biological evaluation of polyester dendrimer: poly(ethylene oxide) "bow-Tie" hybrids with tunable molecular weight and architecture. *Molecular Pharmaceutics 2*, 129–138.
41. Zhao, F., Zhao, Y., Liu, Y., Chang, X., Chen, C., Zhao, Y. (2011). Cellular uptake, intracellular trafficking, and cytotoxicity of nanomaterials. *Small 7*, 1322–1337.
42. Rajendran, L., Knolker, H. J., Simons, K. (2010). Subcellular targeting strategies for drug design and delivery. *Nature Reviews Drug Discovery 9*, 29–42.
43. Kitchens, K. M., Kolhatkar, I. B., Swaan, P. W., Ghandehari, H. (2008). Endocytosis inhibitors prevent poly(amidoamine) dendrimer internalization and permeability across Caco-2 cells. *Molecular Pharmaceutics 5*, 364–369.
44. Albertazzi, L., Serresi, M., Albanese, A., Beltram, F. (2010). Dendrimer internalization and intracellular trafficking in living cells. *Molecular Pharmaceutics 7*, 680–688.
45. Perumal, O. P., Inapagolla, R., Kannan, S., Kannan, R. M. (2008). The effect of surface functionality on cellular trafficking of dendrimers. *Biomaterials 29*, 3469–3476.
46. Mammen, M., Choi, S. K., Whitesides, G. M. (1998). Polyvalent interactions in biological systems: implications for design and use of multivalent ligands and inhibitors. *Angewandte Chemie International Edition 37*, 2755–2794.
47. Badjic, J. D., Nelson, A., Cantrill, S. J., Turnbull, W. B., Stoddart, J. F. (2005). Multivalency and cooperativity in supramolecular chemistry. *Accounts of Chemical Research 38*, 723–732.
48. Kiessling, L. L., Gestwicki, J. E., Strong, L. E. (2006). Synthetic multivalent ligands as probes of signal transduction. *Angewandte Chemie International Edition 45*, 2348–2368.
49. Vance, D., Martin, J., Patke, S., Kane, R. S. (2009). The design of polyvalent scaffolds for targeted delivery. *Advanced Drug Delivery Reviews 61*, 931–939.
50. Kiessling, L. L., Gestwicki, J. E., Strong, L. E. (2000). Synthetic multivalent ligands in the exploration of cell–surface interactions. *Current Opinion in Chemical Biology 4*, 696–703.
51. Wolfenden, M. L., Cloninger, M. J. (2006). Carbohydrate-functionalized dendrimers to investigate the predictable tunability of multivalent interactions. *Bioconjugate Chemistry 17*, 958–966.
52. Liu, S., Maheshwari, R., Kiick, K. L. (2009). Polymer-based therapeutics. *Macromolecules 42*, 3–13.
53. Page, D., Aravind, S., Roy, R. (1996). Synthesis and lectin binding properties of dendritic mannopyranoside. *Chemical Communications* 1913–1914.

54. Thompson, J. P., Schengrund, C. L. (1997). Oligosaccharide-derivatized dendrimers: defined multivalent inhibitors of the adherence of the cholera toxin B subunit and the heat labile enterotoxin of *E. coli* GM1. *Glycoconjugate Journal 14*, 837–845.
55. Reuter, J. D., Myc, A., Hayes, M. M., Gan, Z., Roy, R., Qin, D., Yin, R., Piehler, L. T., Esfand, R., Tomalia, D. A., Baker, Jr., J. R., (1999). Inhibition of viral adhesion and infection by sialic-acid-conjugated dendritic polymers. *Bioconjugate Chemistry 10*, 271–278.
56. Vrasidas, I., de Mol, N. J., Liskamp, R. M. J., Pieters, R. J. (2001). Synthesis of lactose dendrimers and multivalency effects in binding to the cholera toxin B subunit. *European Journal of Organic Chemistry* 4685–4692.
57. Toole, B. P. (2009). Hyaluronan-CD44 interactions in cancer: paradoxes and possibilities. *Clinical Cancer Research 15*, 7462–7468.
58. Lee, Y. C., Lee, R. T. (1995). Carbohydrate–protein interactions—basis of glycobiology. *Accounts of Chemical Research 28*, 321–327.
59. Lis, H., Sharon, N. (1998). Lectins: carbohydrate-specific proteins that mediate cellular recognition. *Chemical Reviews 98*, 637–674.
60. Sears, P., Wong, C. H. (1999). Carbohydrate mimetics: a new strategy for tackling the problem of carbohydrate-mediated biological recognition. *Angewandte Chemie International Edition 38*, 2301–2324.
61. Ashton, P. R., Hounsell, E. F., Jayaraman, N., Nilsen, T. M., Spencer, N., Stoddart, J. F., Young, M. (1998). Synthesis and biological evaluation of alpha-D-mannopyranoside-containing dendrimers. *Journal of Organic Chemistry 63*, 3429–3437.
62. Andre, S., Ortega, P. J. C., Perez, M. A., Roy, R., Gabius, H. J. (1999). Lactose-containing starburst dendrimers: influence of dendrimer generation and binding-site orientation of receptors (plant/animal lectins and immunoglobulins) on binding properties. *Glycobiology 9*, 1253–1261.
63. Baird, E. J., Holowka, D., Coates, G. W., Baird, B. (2003). Highly effective poly (ethylene glycol) architectures for specific inhibition of immune receptor activation. *Biochemistry 42*, 12739–12748.
64. Woller, E. K., Cloninger, M. J. (2002). The lectin-binding properties of six generations of mannose-functionalized dendrimers. *Organic Letters 4*, 7–10.
65. Choudhury, A. K., Kitaoka, M., Hayashi, K. (2003). Synthesis of a cellobiosylated dimer and trimer and of cellobiose-coated polyamidoamine (PAMAM) dendrimers to study accessibility of an enzyme, cellodextrin phosphorylase. *European Journal of Organic Chemistry* 2462–2470.
66. Samuelson, L. E., Sebby, K. B., Walter, E. D., Singel, D. J., Cloninger, M. J. (2004). EPR and affinity studies of mannose-TEMPO functionalized PAMAM dendrimers. *Organic & Biomolecular Chemistry 2*, 3075–3079.
67. Schlick, K. H., Udelhoven, R. A., Strohniever, G. C., Cloninger, M. J. (2005). Binding of mannose-functionalized dendrimers with pea (*Pisum sativum*) lectin. *Molecular Pharmaceutics 2*, 295–301.
68. Kane, R. S. (2010). Thermodynamics of multivalent interactions: influence of the linker. *Langmuir 26*, 8636–8640.
69. Shewmake, T. A., Solis, F. J., Gillies, R. J., Caplan, M. R. (2008). Effects of linker length and flexibility on multivalent targeting. *Biomacromolecules 9*, 3057–3064.

70. Kim, Y., Hechler, B., Gao, Z. G., Gachet, C., Jacobson, K. A. (2009). PEGylated dendritic unimolecular micelles as versatile carriers for ligands of G protein-coupled receptors. *Bioconjugate Chemistry 20*, 1888–1898.

71. Rele, S. M., Cui, W. X., Wang, L. C., Hou, S. J., Barr-Zarse, G., Tatton, D., Gnanou, Y., Esko, J. D., Chaikof, E. L. (2005). Dendrimer-like PEO glycopolymers exhibit anti-inflammatory properties. *Journal of the American Chemical Society 127*, 10132–10133.

72. Gestwicki, J. E., Cairo, C. W., Strong, L. E., Oetjen, K. A., Kiessling, L. L. (2002). Influencing receptor-ligand binding mechanisms with multivalent ligand architecture. *Journal of the American Chemical Society 124*, 14922–14933.

73. Dijkgraaf, I., Rijnders, A. Y., Soede, A., Dechesne, A. C., van Esse, G. W., Brouwer, A. J., Corstens, F. H., Boerman, O. C., Rijkers, D. T., Liskamp, R. M. (2007). Synthesis of DOTA-conjugated multivalent cyclic-RGD peptide dendrimers via 1,3-dipolar cycloaddition and their biological evaluation: implications for tumor targeting and tumor imaging purposes. *Organic & Biomolecular Chemistry 5*, 935–944.

74. McNerny, D. Q., Kukowska-Latallo, J. F., Mullen, D. G., Wallace, J. M., Desai, A. M., Shukla, R., Huang, B., Banaszak Holl, M. M., Baker, J. R. (2009). RGD Dendron bodies; synthetic avidity agents with defined and potentially interchangeable effector sites that can substitute for antibodies. *Bioconjugate Chemistry 20*, 1853–1859.

75. Martin, A. L., Bernas, L. M., Rutt, B. K., Foster, P. J., Gillies, E. R. (2008). Enhanced cell uptake of superparamagnetic iron oxide nanoparticles functionalized with dendritic guanidines. *Bioconjugate Chemistry 19*, 2375–2384.

76. Vannucci, L., Fiserova, A., Sadalapure, K., Lindhorst, T. K., Kuldova, M., Rossmann, P., Horvath, O., Kren, V., Krist, P., Bezouska, K., Luptovcova, M., Mosca, F., Pospisil, M. (2003). Effects of *N*-acetyl-glucosamine-coated glycodendrimers as biological modulators in the B16F10 melanoma model *in vivo*. *International Journal of Oncology 23*, 285–296.

77. Zhu, S. J., Hong, M. H., Zhang, L. H., Tang, G. T., Jiang, Y. Y., Pei, Y. Y. (2010). PEGylated PAMAM dendrimer-doxorubicin conjugates: *in vitro* evaluation and *in vivo* tumor accumulation. *Pharmaceutical Research 27*, 161–174.

78. Pan, G. F., Lemmouchi, Y., Akala, E. O., Bakare, O. (2005). Studies on PEGylated and drug-loaded PAMAM dendrimers. *Journal of Bioactive and Compatible Polymers 20*, 113–128.

79. Kojima, C., Kono, K., Maruyama, K., Takagishi, T. (2000). Synthesis of polyamidoamine dendrimers having poly(ethylene glycol) grafts and their ability to encapsulate anticancer drugs. *Bioconjugate Chemistry 11*, 910–917.

80. Beezer, A. E., King, A. S. H., Martin, I. K., Mitchel, J. C., Twyman, L. J., Wain, C. F. (2003). Dendrimers as potential drug carriers; encapsulation of acidic hydrophobes within water soluble PAMAM derivatives. *Tetrahedron 59*, 3873–3880.

81. Morgan, M. T., Carnahan, M. A., Immoos, C. E., Ribeiro, A. A., Finkelstein, S., Lee, S. J., Grinstaff, M. W. (2003). Dendritic molecular capsules for hydrophobic compounds. *Journal of the American Chemical Society 125*, 15485–15489.

82. Majoros, I. J., Myc, A., Thomas, T., Mehta, C. B., Baker, J. R. (2006). PAMAM dendrimer-based multifunctional conjugate for cancer therapy: synthesis, characterization, and functionality. *Biomacromolecules 7*, 572–579.

83. Khandare, J., Kolhe, P., Pillai, O., Kannan, S., Lieh-Lai, M., Kannan, R. M. (2005). Synthesis, cellular transport, and activity of polyamidoamine dendrimer–methylprednisolone conjugates. *Bioconjugate Chemistry 16*, 330–337.

84. Lai, P. -S., Lou, P. -J., Peng, C. -L., Pai, C. -L., Yen, W. -N., Huang, M. -Y., Young, T. -H., Shieh, M. -J. (2007). Doxorubicin delivery by polyamidoamine dendrimer conjugation and photochemical internalization for cancer therapy. *Journal of Controlled Release 122*, 39–46.
85. Lim, J., Chouai, A., Lo, S. -T., Liu, W., Sun, X., Simanek, E. E. (2009). Design, synthesis, characterization, and biological evaluation of triazine dendrimers bearing paclitaxel using ester and ester/disulfide linkages. *Bioconjugate Chemistry 20*, 2154–2161.
86. Myc, A., Douce, T. B., Ahuja, N., Kotlyar, A., Kukowska-Latallo, J., Thomas, T. P., Baker, J. R. (2008). Preclinical antitumor efficacy evaluation of dendrimer-based methotrexate conjugates. *Anti-Cancer Drugs 19*, 143–149.
87. Zhang, Y., Thomas, T. P., Desai, A., Zong, H., Leroueil, P. R., Majoros, I. J., Baker, J. R. (2010). Targeted dendrimeric anticancer prodrug: a ethotrexate-folic acid-poly(amido-amine) conjugate and a novel, rapid, "one pot" synthetic approach. *Bioconjugate Chemistry 21*, 489–495.
88. La Thangue, N. B., Kerr, D. J. (2011). Predictive biomarkers: a paradigm shift towards personalized cancer medicine. *Nature Reviews Clinical Oncology 8*, 587–596.
89. van 't Veer, L. J., Dai, H., van de Vijver, M. J., He, Y. D., Hart, A. A., Mao, M., Peterse, H. L., van der Kooy, K., Marton, M. J., Witteveen, A. T., Schreiber, G. J., Kerkhoven, R. M., Roberts, C., Linsley, P. S., Bernards, R., Friend, S. H. (2002). Gene expression profiling predicts clinical outcome of breast cancer. *Nature 415*, 530–536.
90. Glover, D. J., Lipps, H. J., Jans, D. A. (2005). Towards safe, non-viral therapeutic gene expression in humans. *Nature Reviews Genetics 6*, 299–310.
91. Pack, D. W., Hoffman, A. S., Pun, S., Stayton, P. S. (2005). Design and development of polymers for gene delivery. *Nature Reviews Drug Discovery 4*, 581–593.
92. Khalil, I. A., Kogure, K., Akita, H., Harashima, H. (2006). Uptake pathways and subsequent intracellular trafficking in nonviral gene delivery. *Pharmacological Reviews 58*, 32–45.
93. Haensler, J., Szoka, Jr., F. C. (1993). Polyamidoamine cascade polymers mediate efficient transfection of cells in culture. *Bioconjugate Chemistry 4*, 372–379.
94. Yoo, H., Juliano, R. L. (2000). Enhanced delivery of antisense oligonucleotides with fluorophore-conjugated PAMAM dendrimers. *Nucleic Acids Research 28*, 4225–4231.
95. Ren, Y., Kang, C. -S., Yuan, X. -B., Zhou, X., Xu, P., Han, L., Wang, G. X., Jia, Z., Zhong, Y., Yu, S., Sheng, J., Pu, P. -Y. (2010). Co-delivery of as-miR-21 and 5-FU by poly (amidoamine) dendrimer attenuates human glioma cell growth *in vitro*. *Journal of Biomaterials Science, Polymer Edition 21*, 303–314.
96. Zhou, J., Wu, J., Hafdi, N., Behr, J. -P., Erbacher, P., Peng, L. (2006). PAMAM dendrimers for efficient siRNA delivery and potent gene silencing. *Chemical Communications* 2362–2364.
97. Arima, H., Kihara, F., Hirayama, F., Uekama, K. (2001). Enhancement of gene expression by polyamidoamine dendrimer conjugates with alpha-, beta-, and gamma-cyclodextrins. *Bioconjugate Chemistry 12*, 476–484.
98. Tang, M. X., Redemann, C. T., Szoka, Jr., F. C. (1996). *In vitro* gene delivery by degraded polyamidoamine dendrimers. *Bioconjugate Chemistry 7*, 703–714.
99. Kuo, J. -h. S., Liou, M. -j., Chiu, H. -c. (2010). Evaluating the gene-expression profiles of HeLa cancer cells treated with activated and nonactivated poly(amidoamine) dendrimers, and their DNA complexes. *Molecular Pharmaceutics 7*, 805–814.

100. Kim, T. -i., Baek, J. -u., Zhe Bai, C., Park, J. -s. (2007). Arginine-conjugated polypropylenimine dendrimer as a non-toxic and efficient gene delivery carrier. *Biomaterials 28*, 2061–2067.

101. Patil, M. L., Zhang, M., Betigeri, S., Taratula, O., He, H., Minko, T. (2008). Surface-modified and internally cationic polyamidoamine dendrimers for efficient siRNA delivery. *Bioconjugate Chemistry 19*, 1396–1403.

102. Huang, R. -Q., Qu, Y. -H., Ke, W. -L., Zhu, J. -H., Pei, Y. -Y., Jiang, C. (2007). Efficient gene delivery targeted to the brain using a transferrin-conjugated polyethyleneglycol-modified polyamidoamine dendrimer. *The FASEB Journal 21*, 1117–1125.

103. Steele, T., Shier, W. (2010). Dendrimeric alkylated polyethylenimine nano-carriers with acid-cleavable outer cationic shells mediate improved transfection efficiency without increasing toxicity. *Pharmaceutical Research 27*, 683–698.

104. Jin, S. -E., Bae, J. W., Hong, S. (2010). Multiscale observation of biological interactions of nanocarriers: from nano to macro. *Microscopy Research and Technique 73*, 813–823.

105. Langereis, S., Dirksen, A., Hackeng, T. M., van Genderen, M. H. P., Meijer, E. W. (2007). Dendrimers and magnetic resonance imaging. *New Journal of Chemistry 31*, 1152–1160.

106. Han, L., Li, J., Huang, S., Huang, R., Liu, S., Hu, X., Yi, P., Shan, D., Wang, X., Lei, H., Jiang, C. (2011). Peptide-conjugated polyamidoamine dendrimer as a nanoscale tumor-targeted T1 magnetic resonance imaging contrast agent. *Biomaterials 32*, 2989–2998.

107. Luo, K., Liu, G., He, B., Wu, Y., Gong, Q., Song, B., Ai, H., Gu, Z. (2011). Multifunctional gadolinium-based dendritic macromolecules as liver targeting imaging probes. *Biomaterials 32*, 2575–2585.

108. Becher, P., Schick, M. J., *Nonionic Surfactants Physical Chemistry*. Surfactants Science Series, New York, Marcel Dekker, 1987, pp. 435–491.

109. Percec, V., Wilson, D. A., Leowanawat, P., Wilson, C. J., Hughes, A. D., Kaucher, M. S., Hammer, D. A., Levine, D. H., Kim, A. J., Bates, F. S., Davis, K. P., Lodge, T. P., Klein, M. L., DeVane, R. H., Aqad, E., Rosen, B. M., Argintaru, A. O., Sienkowska, M. J., Rissanen, K., Nummelin, S., Ropponen, J. (2010). Self-assembly of janus dendrimers into uniform dendrimersomes and other complex architectures. *Science 328*, 1009–1014.

110. Jang, W. -D., Kamruzzaman Selim, K. M., Lee, C. -H., Kang, I. -K. (2009). Bioinspired application of dendrimers: from bio-mimicry to biomedical applications. *Progress in Polymer Science 34*, 1–23.

111. Myc, A., Kukowska-Latallo, J., Cao, P., Swanson, B., Battista, J., Dunham, T., Baker, J. R. (2010). Targeting the efficacy of a dendrimer-based nanotherapeutic in heterogeneous xenograft tumors *in vivo*. *Anti-Cancer Drugs 21*, 186–192.

112. Majoros, I. J., Myc, A., Thomas, T., Mehta, C. B., Baker, Jr., J. R. (2006). PAMAM dendrimer-based multifunctional conjugate for cancer therapy: synthesis, characterization, and functionality. *Biomacromolecules 7*, 572–579.

113. Fox, M. E., Guillaudeu, S., Frechet, J. M., Jerger, K., Macaraeg, N., Szoka, F. C. (2009). Synthesis and *in vivo* antitumor efficacy of PEGylated poly(L-lysine) dendrimer-camptothecin conjugates. *Molecular Pharmaceutics 6*, 1562–1572.

114. Singh, P., Gupta, U., Asthana, A., Jain, N. K. (2008). Folate and folate-PEG-PAMAM dendrimers: synthesis, characterization, and targeted anticancer drug delivery potential in tumor bearing mice. *Bioconjugate Chemistry 19*, 2239–2252.

115. Nishiyama, N., Morimoto, Y., Jang, W. D., Kataoka, K. (2009). Design and development of dendrimer photosensitizer-incorporated polymeric micelles for enhanced photodynamic therapy. *Advanced Drug Delivery Reviews 61*, 327–338.

116. Kitchens, K. M., Foraker, A. B., Kolhatkar, R. B., Swaan, P. W., Ghandehari, H. (2007). Endocytosis and interaction of poly (amidoamine) dendrimers with Caco-2 cells. *Pharmaceutical Research 24*, 2138–2145.

117. Goldberg, D. S., Ghandehari, H., Swaan, P. W. (2010). Cellular entry of G3.5 poly (amido amine) dendrimers by clathrin- and dynamin-dependent endocytosis promotes tight junctional opening in intestinal epithelia. *Pharmaceutical Research 27*, 1547–1557.

118. Seib, F. P., Jones, A. T., Duncan, R. (2007). Comparison of the endocytic properties of linear and branched PEIs, and cationic PAMAM dendrimers in B16f10 melanoma cells. *Journal of Controlled Release 117*, 291–300.

119. Saovapakhiran, A., D'Emanuele, A., Attwood, D., Penny, J. (2009). Surface modification of PAMAM dendrimers modulates the mechanism of cellular internalization. *Bioconjugate Chemistry 20*, 693–701.

120. Huang, R. Q., Ke, W. L., Han, L., Liu, Y., Shao, K., Ye, L. Y., Lou, J. N., Jiang, C., Pei, Y. Y. (2009). Brain-targeting mechanisms of lactoferrin-modified DNA-loaded nanoparticles. *Journal of cerebral blood flow and metabolism 29*, 1914–1923.

121. Dhanikula, R. S., Hammady, T., Hildgen, P. (2009). On the mechanism and dynamics of uptake and permeation of polyether-copolyester dendrimers across an *in vitro* blood-brain barrier model. *Journal of Pharmaceutical Sciences 98*, 3748–3760.

122. Yang, W. L., Barth, R. F., Wu, G., Huo, T. Y., Tjarks, W., Ciesielski, M., Fenstermaker, R. A., Ross, B. D., Wikstrand, C. J., Riley, K. J., Binns, P. J. (2009). Convection enhanced delivery of boronated EGF as a molecular targeting agent for neutron capture therapy of brain tumors. *Journal of Neuro-Oncology 95*, 355–365.

123. Waite, C. L., Roth, C. M. (2009). PAMAM-RGD conjugates enhance siRNA delivery through a multicellular spheroid model of malignant glioma. *Bioconjugate Chemistry 20*, 1908–1916.

124. Olson, E. S., Jiang, T., Aguilera, T. A., Nguyen, Q. T., Ellies, L. G., Scadeng, M., Tsien, R. Y. (2010). Activatable cell penetrating peptides linked to nanoparticles as dual probes for *in vivo* fluorescence and MR imaging of proteases. *Proceedings of the National Academy of Sciences of the United States of America 107*, 4311–4316.

125. Cheng, Z., Thorek, D. L., Tsourkas, A. (2010). Gadolinium-conjugated dendrimer nanoclusters as a tumor-targeted T1 magnetic resonance imaging contrast agent. *Angewandte Chemie International Edition 49*, 346–350.

126. Kobayashi, H., Koyama, Y., Barrett, T., Hama, Y., Regino, C. A., Shin, I. S., Jang, B. S., Le, N., Paik, C. H., Choyke, P. L., Urano, Y. (2007). Multimodal nanoprobes for radionuclide and five-color near-infrared optical lymphatic imaging. *ACS Nano 1*, 258–264.

127. Al-Jamal, K. T., Al-Jamal, W. T., Akerman, S., Podesta, J. E., Yilmazer, A., Turton, J. A., Bianco, A., Vargesson, N., Kanthou, C., Florence, A. T., Tozer, G. M., Kostarelos, K. (2010). Systemic antiangiogenic activity of cationic poly-L-lysine dendrimer delays tumor growth. *Proceedings of the National Academy of Sciences of the United States of America 107*, 3966–3971.

128. Kaneshiro, T. L., Jeong, E. K., Morrell, G., Parker, D. L., Lu, Z. R. (2008). Synthesis and evaluation of globular Gd-DOTA-monoamide conjugates with precisely controlled nanosizes for magnetic resonance angiography. *Biomacromolecules 9*, 2742–2748.

129. Taratula, O., Garbuzenko, O. B., Kirkpatrick, P., Pandya, I., Savla, R., Pozharov, V. P., He, H., Minko, T. (2009). Surface-engineered targeted PPI dendrimer for efficient intracellular and intratumoral siRNA delivery. *Journal of Controlled Release 140*, 284–293.

130. Chisholm, E. J., Vassaux, G., Martin-Duque, P., Chevre, R., Lambert, O., Pitard, B., Merron, A., Weeks, M., Burnet, J., Peerlinck, I., Dai, M. S., Alusi, G., Mather, S. J., Bolton, K., Uchegbu, I. F., Schatzlein, A. G., Baril, P. (2009). Cancer-specific transgene expression mediated by systemic injection of nanoparticles. *Cancer Research 69*, 2655–2662.

131. Parrott, M. C., Benhabbour, S. R., Saab, C., Lemon, J. A., Parker, S., Valliant, J. F., Adronov, A. (2009). Synthesis, radiolabeling, and bio-imaging of high-generation polyester dendrimers. *Journal of the American Chemical Society 131*, 2906–2916.

132. Guillaudeu, S. J., Fox, M. E., Haidar, Y. M., Dy, E. E., Szoka, F. C., Frechet, J. M. J. (2008). PEGylated dendrimers with core functionality for biological applications. *Bioconjugate Chemistry 19*, 461–469.

133. Almutairi, A., Rossin, R., Shokeen, M., Hagooly, A., Ananth, A., Capoccia, B., Guillaudeu, S., Abendschein, D., Anderson, C. J., Welch, M. J., Frechet, J. M. J. (2009). Biodegradable dendritic positron-emitting nanoprobes for the noninvasive imaging of angiogenesis. *Proceedings of the National Academy of Sciences of the United States of America 106*, 685–690.

134. Dhanikula, R. S., Argaw, A., Bouchard, J. F., Hildgen, P. (2008). Methotrexate loaded polyether–copolyester dendrimers for the treatment of gliomas: enhanced efficacy and intratumoral transport capability. *Molecular Pharmaceutics 5*, 105–116.

135. Nishiyama, N., Nakagishi, Y., Morimoto, Y., Lai, P. S., Miyazaki, K., Urano, K., Horie, S., Kumagai, M., Fukushima, S., Cheng, Y., Jang, W. D., Kikuchi, M., Kataoka, K. (2009). Enhanced photodynamic cancer treatment by supramolecular nanocarriers charged with dendrimer phthalocyanine. *Journal of Controlled Release 133*, 245–251.

136. van der Poll, D. G., Kieler-Ferguson, H. M., Floyd, W. C., Guillaudeu, S. J., Jerger, K., Szoka, F. C., Frechet, J. M. (2010). Design, synthesis, and biological evaluation of a robust, biodegradable dendrimer. *Bioconjugate Chemistry 21*, 764–773.

# CHITOSAN-BASED NANOPARTICLES FOR BIOMEDICAL APPLICATIONS

Heebeom Koo, Kuiwon Choi, Ick Chan Kwon, and Kwangmeyung Kim

*Center for Theragnosis, Biomedical Research Institute, Korea Institute of Science and Technology, Seongbuk-gu, Seoul, Republic of Korea*

## 6.1 CHITOSAN AS A BIOPOLYMER

In recent years, progress in the field of biotechnology and nanotechnology has made a huge impact on biomedical fields. In addition to the improvement in mechanical devices, nanoparticles have emerged as a multifunctional platform for both diagnosis and therapy of various diseases [1,2]. Researchers have developed various kinds of nanoparticles using polymers, lipids, iron, gold, silica, and so on [3,4]. Among them, polymers are the most widely used components in fabricating nanoparticles [5]. Their major advantage is that their physical or chemical properties can be easily controlled by changing the molecular weights or chemical structures [6]. Therefore, a number of polymer-based nanoparticles have been developed and applied to biomedical fields [7].

Chitosan is one of the representative natural polymers, which is derived from chitin and frequently used for the development of nanoparticles [8,9]. Chitin is the main component of crustacean shell and is composed of β-(1,4)-2-acetamido-D-glucose linked via (1–4) glycosidic bonds. Chitosan is produced by deacetylation of chitin, and the number of amine groups is dependent on the degree of deacetylation. These amine groups endow many useful characteristics with chitosan.

Water solubility of chitosan is mainly dependent on the protonation of amine groups. Because the $pK_a$ value of this amine group is about 6.5, chitosan is generally

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*,
Edited by Shayne C. Gad
Copyright © 2013 John Wiley & Sons, Inc.

soluble in solvents with pH below 6.5 but insoluble in solvents with neutral or basic pH [10]. Even though low-molecular-weight chitosans below ~50 kDa are soluble at neutral pH, most chitosans originally have relatively low solubility in water, which is a major hurdle for further development. To enhance the water solubility of chitosan, various chitosan derivatives have been developed by chemical modification. One way to enhance water solubility is to introduce additional cationic or anionic charged groups. Quaternized chitosan (trimethyl chitosan), sulfated chitosan, and carboxymethyl chitosan are representative chitosan derivatives which have charges at a neutral condition [11–13]. Alternatively, conjugation with highly water soluble poly(ethylene glycol) (PEG) or glycol groups (one unit of ethylene oxide) can enhance the solubility of chitosan without additional charges [14–16].

Chitosan is distinguished from many biopolymers by its amine groups providing useful information on its properties. First, this amine group can bind to anionic surface of cellular membrane, enabling efficient cellular adhesion and uptake [17]. This property is very useful in biomedical applications such as drug or gene delivery. Second, chitosan can also bind to anionic DNA or RNA by electronic interactions with these amine groups [18,19]. Consequently, chitosan is a very attractive material for developing gene carriers [20]. Finally, many other molecules can be conjugated to chitosan via these amine groups [21]. This point is particularly important because amide coupling of amine and carboxylic acid groups is the most widely used conjugation method. In addition, a number of functional molecules such as fluorescence dyes are commercially available as activated with *N*-hydroxysuccinimide ester (NHS) for conjugation to these amine groups [22].

Biodegradability and biocompatibility of chitosan facilitate its use for biomedical applications. The oral $LD_{50}$ value of chitosan is more than 16 g/kg in mice, which is very high and demonstrates the biocompatibility of chitosan [23]. It can be degraded to low-molecular-weight molecules by various enzymes secreted from the kidney. Owing to this biocompatibility, its use as a wound dressing agent was approved for human use by the U.S. Food and Drug Administration (FDA) [24]. On the basis of these points, chitosan is generally believed as a biocompatible biopolymer by many researchers. However, modified chitosans may show more delayed renal excretion due to limited interactions with enzymes [25].

In this review, we focus on chitosan-based nanoparticles for imaging and drug/gene delivery. We introduce many studies on chitosan nanoparticles, which evaluated biological effects in living systems such as cultured cells or animal models. Furthermore, we describe important considerations in developing chitosan-based nanoparticles for diagnosis and therapy of diseases.

## 6.2 CHITOSAN NANOPARTICLES FOR IMAGING

### 6.2.1 Optical Imaging

Optical imaging is generally based on fluorescence or luminescence. The advantages are easy handling, high sensitivity, and relatively low cost of imaging equipment [26]. Optical imaging nanoprobes can be prepared by simple conjugation or loading

of commercial fluorescence dyes to nanoparticles [27]. Nanoparticles with targeting ability can be used for optical imaging of a specific disease through long-term blood circulation and high accumulation in the target site. For *in vivo* optical imaging, the near-infrared (NIR) region of fluorescence (about 650–900 nm) is frequently used due to the relatively low interference by hemoglobin and water [28].

Ranjan et al. showed that chitosan-based polymeric nanoparticles can be used in *in vivo* optical imaging by simple loading of indocyanine green (ICG), an FDA-approved commercial fluorescence dye [29]. They developed nanoparticles with poly(L-lactide-*co*-epsilon-caprolactone) and poloxamer, and ICG was encapsulated in the nanoparticles. The surface of nanoparticles was further modified by chitosan. The positive charge of chitosan enhanced cellular uptake and tissue accumulation during both *in vitro* and *in vivo* experiments. These results were obtained by NIR optical imaging of the loaded ICG.

Recently, Kim's group demonstrated an excellent tumor-targeting ability of their glycol chitosan nanoparticles by optical imaging [30]. Glycol chitosan is a chitosan derivative with good water solubility. Glycol groups also provide cationic amine groups with PEG-like shielding effects inhibiting interactions of nanoparticles with anionic serum protein. Glycol chitosan nanoparticles were synthesized by conjugating hydrophobic 5β-cholanic acids to hydrophilic glycol chitosan [31] (Figure 6.1a, b ). The average size of glycol chitosan nanoparticles was about 250 nm, and NIR fluorescence dye Cy 5.5 was conjugated to it for optical imaging. Owing to the small size, deformability, fast uptake into tumor cells, and the shielding effect of glycol groups, glycol chitosan nanoparticles could be efficiently accumulated within tumors via the enhanced permeability and retention (EPR) effect in newly generated tumor vessels [32–34]. The tumor-targeting ability of glycol chitosan nanoparticles was analyzed in various tumor models. In a flank tumor model, it showed long-term circulation and high accumulation in tumor tissues for a prolonged period (Figure 6.1c). In particular, the EPR effect in tumor tissues and fast penetration of nanoparticles were monitored with live optical imaging [35] (Figure 6.1d). Subsequently, a significant correlation between angiogenesis in tumors and *in vivo* behavior of glycol chitosan nanoparticles was shown in a brain tumor model in a time-dependent manner (Figure 6.1e). In a liver tumor model, the high fluorescence signals in tumor tissues than in surrounding normal liver tissues demonstrated the ability of glycol chitosan nanoparticles to escape from the reticuloendothelial system (RES) of the liver (Figure 6.1f). Finally, specific detection of metastatic tumors with them was shown in a lung metastasis tumor model (Figure 6.1g). Furthermore, successful optical imaging of unintended secondary metastatic tumors on forefoot showed the potential of the early detection of metastasis in patients. These results demonstrate the utility of fluorescently labeled chitosan nanoparticles in optical imaging of various cancers.

### 6.2.2 Magnetic Resonance (MR) Imaging

Nowadays, magnetic resonance (MR) imaging is indispensable for the diagnosis of many diseases in clinical settings. MR imaging mainly provides anatomical imaging

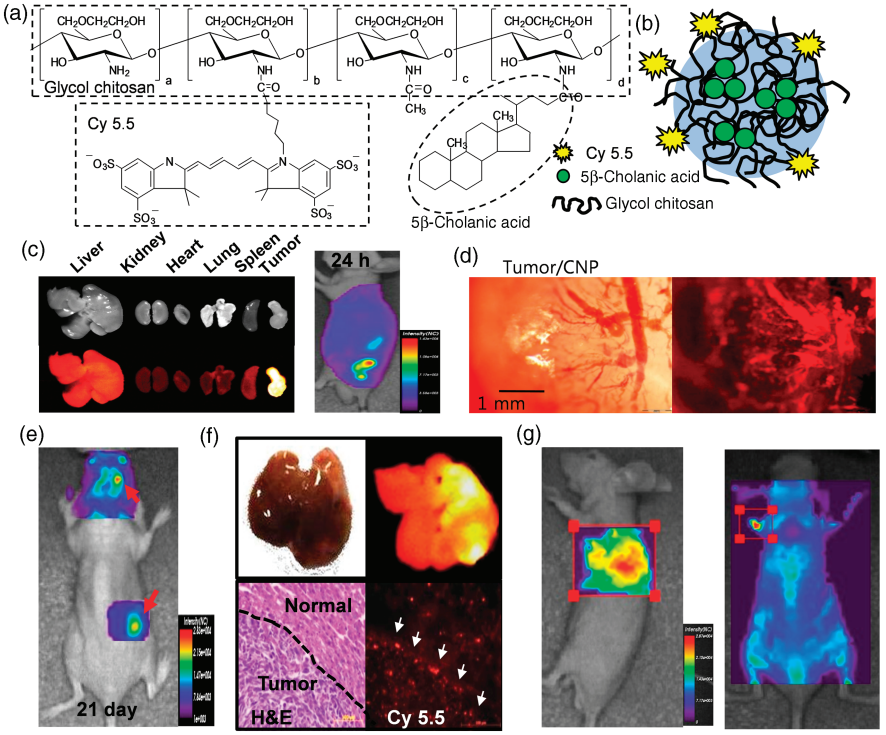


**FIGURE 6.1** Chitosan nanoparticles for optical imaging. (a) Chemical structure of Cy5.5 and 5β-cholanic acid conjugated glycol chitosan. (b) Schematic illustration of a self-assembled glycol chitosan nanoparticle. (c) *In vivo* and *ex vivo* tumor imaging with chitosan nanoparticles in left flank tumor model. (d) Vascular penetration imaging in left flank tumor tissue with chitosan nanoparticles. (e) Brain tumor imaging with chitosan nanoparticles. (f) Liver tumor imaging with chitosan nanoparticles (white arrows marks the margin between normal and tumor tissue.). (g) Lung and secondary metastatic tumor imaging with chitosan nanoparticles. (*Source*: Figures reproduced with permission from Reference [30].)

of organs and tissues, and can visualize internal disorders of the body resulting from diseases. MR imaging is generally categorized into two modes: longitudinal (T1) mode and transverse (T2) mode [36].

In T1 MR imaging, gadolinium (Gd)-based contrast agents can reduce T1 relaxation time by interacting with water and provide bright signals. Gd ions are generally entrapped by chelators such as tetraazacyclododecane tetraacetic acid (DOTA) or diethylenetriaminepentaacetic acid (DTPA) for MR imaging [37]. These chelators can be chemically conjugated to functional nanoparticles for improvement of T1 contrast imaging of target tissues [38]. Since Gd-based T1 MR imaging is based on interactions with water molecules, Gd chelators should be present on the outer environment rather than inside of the nanoparticles. Recently, Kim's group

developed a dual imaging probe by conjugation of Gd-DOTA and Cy 5.5 to glycol chitosan backbone of the nanoparticles [39]. In addition to optical imaging by Cy 5.5, intense T1 MR signal of the resulting nanoparticles revealed bright tumor tissues in MR images of tumor-bearing mice.

On the other hand, T2 MR imaging agents, such as iron oxide nanoparticles, shorten the transverse relaxation time and provide dark images of target tissues [40]. Iron oxide nanoparticles are generally synthesized at a high temperature using fatty acids as a surfactant. The iron oxide nanoparticles with such a hydrophobic surface should be hydrophilically modified for applications in aqueous environment of the body [41]. Consequently, many researchers conjugated hydrophilic polymers to them or encapsulated them in hydrophilic nanoparticles to increase water solubility. Chitosan can be employed for varying purposes. For example, Lee et al. developed iron oxide-loaded chitosan nanoparticles with hydrophobic linoleic acid-conjugated water-soluble chitosan and hydrophobic fatty acid-coated iron oxide nanoparticles [42]. These chitosan nanoparticles were stable in an aqueous condition for a long time and generated dark T2 signals in hepatocytes in mice after intravenous injection. The same group also synthesized chitosan nanoparticles conjugated with oleic acid for self-assembly and Cy 5.5 dyes for optical imaging [43] (Figure 6.2a). Subsequently, oleic acid-coated iron oxide nanoparticles were encapsulated in these chitosan nanoparticles for T2 MR imaging (Figure 6.2b). The resulting nanoparticles could provide both optical imaging and T2 MR signals (Figure 6.2c, d). In a U87MG brain-tumor-bearing mouse model, the tumor site was marked by optical fluorescence imaging and T2 MR imaging simultaneously (Figure 6.2e–g).

### 6.2.3 Computed Tomography (CT) Imaging

Computed tomography (CT) is a powerful imaging technique, which can provide two- or three-dimensional images of living subjects. CT can provide anatomical images of bones and tissues with very high spatial resolution. For enhanced CT imaging, iodine-based contrast agents are used in clinical settings [44]. However, owing to the renal toxicity and high viscosity of iodine solutions, there is an urgent need for alternative CT imaging agents. Recently, many researchers have focused on gold nanoparticles as a new CT imaging agent because gold nanoparticles present a high X-ray absorption coefficient for intense CT signals [45]. Consequently, several groups have tried targeted delivery of gold nanoparticles for *in vivo* CT imaging [46].

Recently, Sun et al. developed a gold nanoparticle-based optical/CT dual imaging probe with glycol chitosan for *in vivo* tumor diagnosis [47]. They used glycol chitosan as a reducing and stabilizing agent during fabrication of gold nanoparticles. Glycol chitosan polymers tightly adsorbed onto the surface of a gold nanoparticle via primary amine groups as reducing agents. Then, matrix metalloproteinase (MMP) activatable peptide probes were conjugated to the glycol chitosan-coated gold nanoparticles for optical imaging [48] (Figure 6.3a). This glycol chitosan-coated gold nanoprobe showed a long circulation time and high accumulation in tumor sites for diagnosis. In a tumor-bearing mouse model, optical and CT imaging of tumor

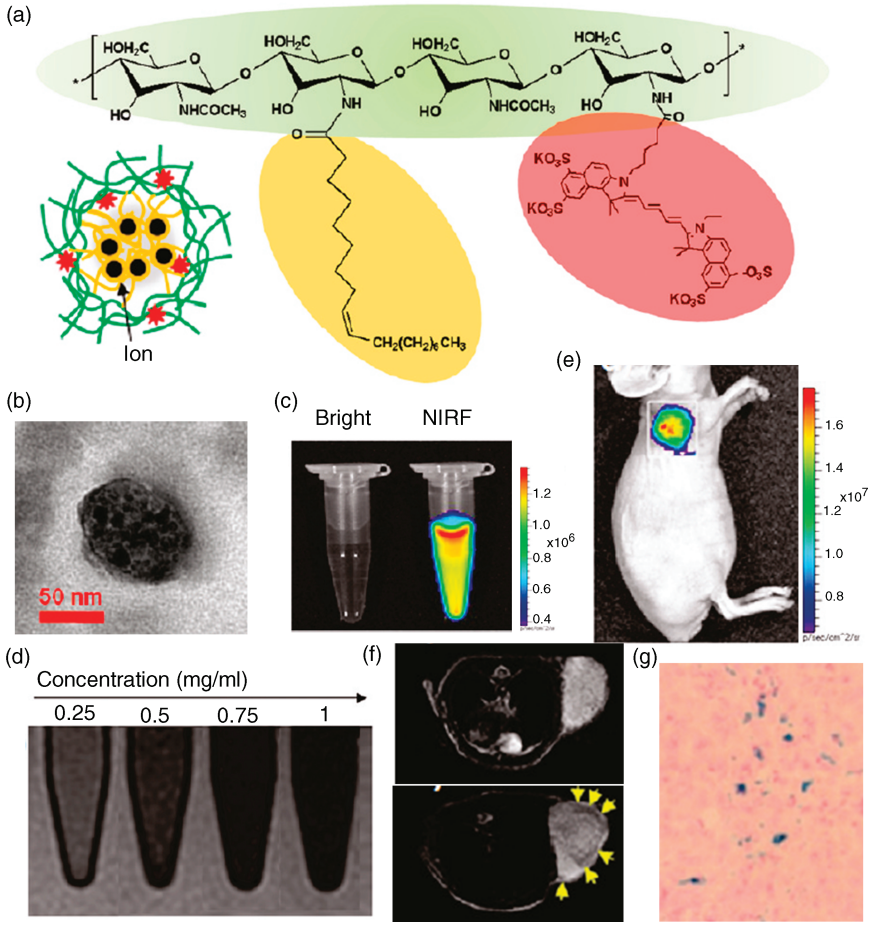


**FIGURE 6.2** Chitosan nanoparticles for optical/MR imaging. (a) Chemical structure and illustration of a Cy5.5 and oleic acid conjugated chitosan nanoparticle containing iron oxide nanoparticles (green, yellow, and red color denotes chitosan, oleic acid, and Cy5.5, respectively). (b) Transmission emission microscopy (TEM) image of a chitosan nanoparticle. (c) Bright field and NIR fluorescence images of chitosan nanoparticle solution. (d) MR phantom images of chitosan nanoparticles. (e) NIR fluorescence image of a tumor-bearing mouse after injection of chitosan nanoparticles. (f) MR image of tumor-bearing mouse after injection of chitosan nanoparticles (yellow arrows denote the dark MR signals from chitosan nanoparticles). (g) Prussian blue stained tumor tissue of chitosan nanoparticle-injected mouse. (*Source*: Figures reproduced with permission from Reference [43].)

tissues was successfully obtained with this nanoprobe (Figure 6.3b, c). Importantly, this research is a good example of a multimodal imaging nanoprobe with synergetic effects [49]. CT imaging can provide three-dimensional anatomical imaging of tumor tissues in the body, while highly sensitive optical imaging can provide

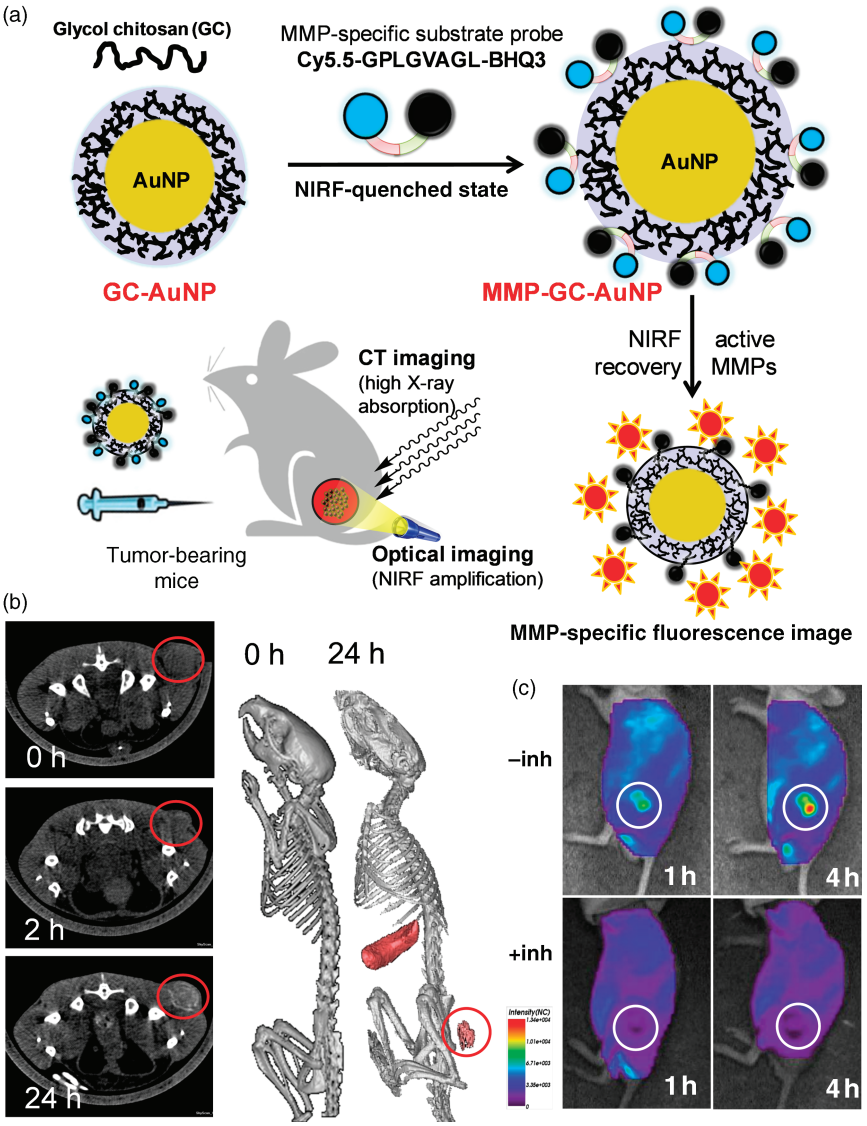


**FIGURE 6.3** Chitosan nanoparticles for optical/CT imaging. (a) Schematic illustration of an MMP-specific glycol chitosan-coated gold nanoparticle and their MMP triggered fluorescence activation. (b) CT image of tumor-bearing mouse after injection of chitosan gold nanoparticles (red circles mark the tumor site). (c) NIR fluorescence image of tumor-bearing mouse after injection of chitosan gold nanoparticles (white circles mark the tumor site). (*Source*: Figures reproduced with permission from Reference [47].)

precise biological information of MMP enzymes in tumor tissues. Benefits of combination should be further explored in the development of multimodal imaging nanoprobes [40].

### 6.2.4 Positron Emission Tomography (PET) Imaging

Positron emission tomography (PET) images are obtained by tracking positron emitting radioisotopes such as $^{11}C$, $^{18}F$, $^{64}Cu$, and $^{68}Ga$ in living systems. A principal advantage of PET is that it is the most widely used functional imaging technique in current clinical settings. Molecularly targeted radioisotopes and high sensitivity of PET can provide biological information of the disease [50]. Moreover, PET imaging enables quantitative analysis of the pharmacokinetics and pharmacodynamics of injected molecules with radiolabeling [51]. $^{18}F$-fluorodeoxyglucose ($^{18}F$-FDG) is the most commonly used and commercialized imaging agent for PET, and many other agents are currently developed.

Yang et al. synthesized $^{18}F$-labeled galactosylated chitosan for hepatocyte imaging with PET [52]. The galactosyl group can bind to hepatocytes in the liver via asialoglycoprotein receptors [53]. The galactosylated chitosan was synthesized by coupling the carboxylic acid group of lactobionic acid and the amine group of chitosan. The resulting chitosan conjugate was labeled with $^{18}F$ by *N*-succinimidyl-4-$^{18}F$-fluorobenzoate. They showed high accumulation in liver tissue after intravenous injection. Injection of excess neogalactosylalbumin as competitive molecules for receptors prevented accumulation of the galactosylated chitosan nanoparticles.

Recently, Tsao et al. reported that $^{68}Ga$-glycopeptide composed of glutamate peptide and chitosan could be applied for PET imaging of tumor angiogenesis [54]. $^{68}Ga$ is one of the most attractive radioisotopes that can be produced from generators for PET imaging [55]. This glycopeptide could efficiently enter tumor cells via glutamate transporters. Tumor tissues in New Zealand rabbits were observed in PET images with $^{68}Ga$-glycopeptide, and their accumulation was higher than that of commercial $^{18}F$-FDG. In addition to these papers, a large number of chitosan nanoparticles showing superior target specificity have been developed for PET imaging [9,56].

### 6.2.5 Ultrasound (US) Imaging

The basic mechanism of ultrasound (US) imaging is the analysis of acoustic signals resulting from reflection or scattering of applied ultrasonic waves inside the body [57]. US imaging has many advantages such as relatively low capital costs, simplicity of the procedure, direct imaging without delay, and low cost of imaging. [58] For the enhancement of US contrast, gas-encapsulated nano- or microbubbles are generally used during imaging [59]. As the core materials of these bubbles, air or inert perfluorocarbon gas has been used. These gas cores are encapsulated by shell materials such as lipids or polymers to enhance the stability and water solubility of bubbles for intravenous injection to patients. Albumin-coated Albunex® and lipid-coated Sonovue® are the most widely used US contrast agents at present.

Liu et al. used chitosan as coating materials of perfluorocarbon gas in developing as an US contrast agent [60]. They synthesized a core-shell structure with silica and chitosan and made hollow nanoparticles by cross-linking of chitosan shell with glutaraldehyde and removing silica cores in hydrofluoride solution. Subsequently, liquid-form perfluoropentane was loaded into the core of nanoparticles. The resulting chitosan/perfluoropentane nanobubbles generated intense US signals in PBS solution similar to Sonovue. Although only a few US contrast agents have been developed with chitosan nanoparticles so far, many chitosan derivatives may be useful for the fabrication of microbubbles for *in vivo* US imaging due to their biocompatibility and wide availability.

## 6.3 CHITOSAN NANOPARTICLES FOR THERAPY

### 6.3.1 Drug Delivery

Owing to the positive charge of chitosan, some chitosan nanoparticles have been developed based on electronic interactions with anionic polymers. Sung's group developed pH-responsive nanoparticles with cationic chitosan and anionic poly-γ-glutamic acid [61]. The ionic cross-linking in this particle was further enhanced with tripolyphosphate (TPP) and magnesium sulfate [62] (Figure 6.4a). From pH 2.5 to 6.6, the nanoparticle is relatively stable because both chitosan and poly-γ-glutamic acid are ionized. However, chitosan is deprotonated at pH above 7.0, and nanoparticles become unstable and can release drugs. This study showed that pH-responsive protonation of chitosan can be applied for controlling drug release.

Many previous papers proved that the amine and carboxylic acid groups of chitosan could bind to glycoproteins in mucus by hydrogen bonding [63]. Moreover, recent studies showed that the tight junctions in monolayers of enterocytes could be opened after binding with chitosan molecules [64,65]. Consequently, these properties make chitosan nanoparticles promising drug carriers for oral drug delivery [66,67]. Sung group applied their pH-responsive chitosan nanoparticles to oral delivery of insulin [68,69]. They showed superior transport of insulin across Caco-2 cell monolayers by chitosan nanoparticles using confocal microscopy (Figure 6.4b). The efficient insulin delivery to the duodenum and jejunum by these particles has also been successfully proved in rat models. Furthermore, the time-dependent blood glucose level data in a diabetic rat model demonstrated the great potential of the chitosan nanoparticles as a protein drug carrier for oral delivery.

Because of its cationic charge, chitosan can bind to anionic drugs directly. Heparin is a highly sulfated glycosaminoglycan with intense anionic charge. It is clinically used as a potent anticoagulant, which can form an inhibitory complex with anitithrombin [70]. Sung's group showed that cationic chitosan can bind to anionic heparin and form self-assembled nanoparticles [70]. The chitosan–heparin nanoparticles were successfully applied to the oral delivery of heparin. Chitosan enhanced the absorption of heparin in the gastrointestinal tract, resulting in effective anticoagulation after oral administration in a rat model.

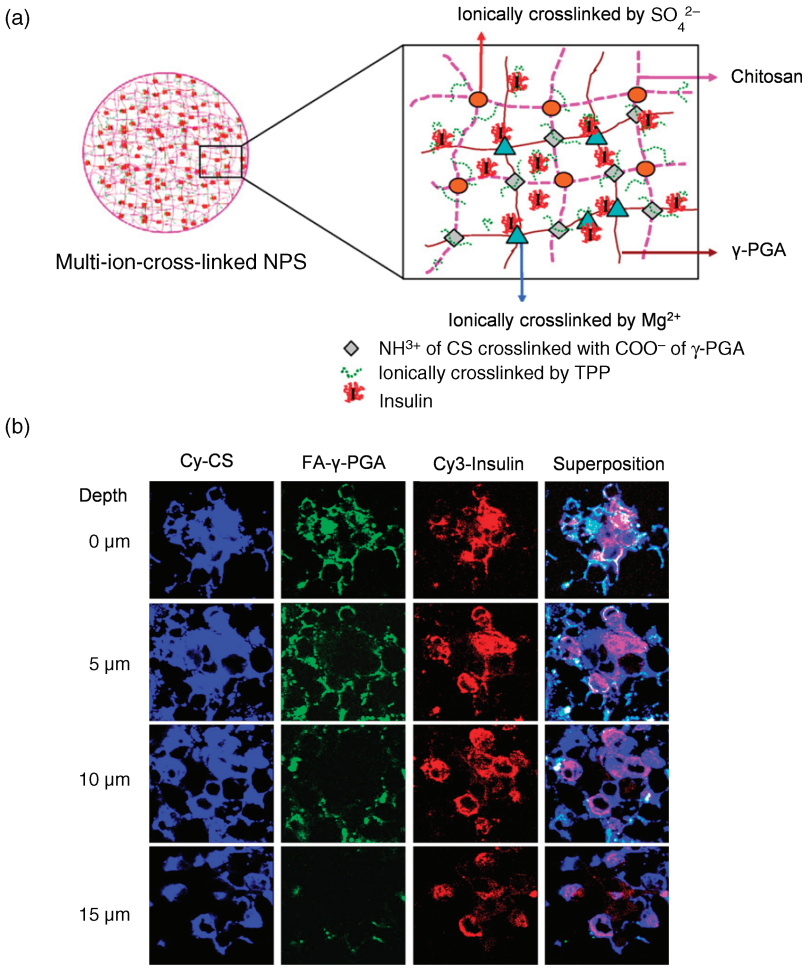


**FIGURE 6.4** Chitosan nanoparticles for drug delivery. (a) Schematic illustration of multi-ion-cross-linked chitosan nanoparticles with poly-γ-glutamic acid (γ-PGA). (b) Fluorescence images of Caco-2 cell monolayers with chitosan nanoparticles. (*Source*: Figures reproduced with permission from Reference [61].)

In case of hydrophobic drugs, amphiphilic chitosan nanoparticles may be a good carrier candidate [71]. Zhao et al. synthesized chitosan nanoparticles by conjugation of hydrophobic linoleic acid and hydrophilic poly(β-maleic acid) to chitosan backbone [72]. Poly(β-maleic acid) could improve the water solubility of original chitosan and facilitate self-assembly of the chitosan conjugate in nanoparticles. A hydrophobic anticancer drug, paclitaxel, was stably loaded into this chitosan nanoparticle, and its therapeutic efficacy has been proved in a tumor-bearing mouse model. Recently, Lee et al. synthesized *N*-acetyl histidine conjugated glycol chitosan

nanoparticles and showed successful therapeutic results following an intravenous injection of doxorubicin-loaded nanoparticles in tumor-bearing mice [73]. The biodistribution of nanoparticles and doxorubicin was monitored by optical and radioisotope imaging. In particular, they showed that the ratio of hydrophobic moiety in nanoparticles affected the amount of doxorubicin accumulated in tumor tissues. This study shows that the optimization of nanoparticles for efficient drug delivery is important, and that real-time *in vivo* imaging technique can be a useful tool for optimization [74].

### 6.3.2 Gene Delivery

Chitosan has long been considered a highly attractive material for developing gene carriers due to its primary amine groups and cationic charges. Chitosan can form a stable complex with nucleotides and increase their cellular uptake. A number of gene delivery carriers were developed with original or modified forms of chitosan, and these efforts are summarized in recent reviews [18,75]. We introduce some representative examples in this section.

Chitosan can be used as a gene carrier in its original form. Leong and colleagues reported that the Factor VIII DNA polyplexes with chitosan could be used in oral gene delivery [76]. Molecular weight and the degree of deacetylation of chitosan have a significant influence on transfection efficiency. In this study, they used chitosan with a molecular weight of 390 kDa and a degree of deacetylation of 84%. Owing to the mucoadhesive character of chitosan, the transfection efficiency with chitosan by oral delivery was more enhanced compared with naked DNA. Moreover, the increased thrombin generation and phenotypic correction in hemophilia A mouse models proved that chitosan was an attractive gene carrier for oral delivery.

Many researchers who are interested in gene therapy of lung diseases have tried several administration routes, including pulmonary administration. Mohri et al. introduced spray–freeze–dried powder of chitosan and DNA for inhalation [77]. They achieved high recovery of dried genes through this method and showed that an optimized amount of chitosan could enhance gene transfection. With dual imaging based on fluorescence and luminescence, the pulmonary administration and gene expression were evaluated in a mouse model. The intense luminescence signal in lung tissues showed that chitosan powder produced with spray–freeze–dried method could be used as pulmonary gene carriers.

To enhance the target cell specificity for gene delivery, some receptor-binding moieties have been conjugated to chitosan nanoparticles [78]. Zheng et al. used folate as a targeting moiety, which can bind to folate receptors abundant in various tumor cells [79]. They synthesized folate conjugated *N*-trimethyl chitosan with high water solubility and DNA condensing ability [80]. The resulting polymer showed high transfection efficiency *in vitro*, and its cytotoxicity was lower than that of synthetic polymers such as polyethylenimine. For hepatocyte targeting, galactosylated chitosan-polyethylenimine (PEI) was synthesized by Jiang and colleagues [81]. It enhanced the transfection efficiency in liver cells by binding to asialoglycoprotein receptors.

In recent years, small interfering RNA (siRNA) is emerging as a promising therapeutic modality, which can specifically suppress the expression of target protein with no adverse effects [82,83]. Oh group developed poly-L-arginine derivatives of chitosan (CS–PLR), where PEG was also conjugated to inhibit their aggregation under aqueous conditions [84] (Figure 6.5a). This chitosan nanoparticle can form a stable complex with siRNA, and its gene silencing effect was confirmed under cell culture conditions with green fluorescence protein. This PEG–CS–PLR showed enhanced gene silencing efficiency and cellular uptake *in vitro* compared to Lipofectamine™ 2000, commercialized transfection agent and non-PEGylated CS–PLR (Figure 6.5b, c). Finally, they made a tumor-bearing mouse model with B16F10 tumor cells expressing red fluorescence protein (RFP). After intratumoral injection of chitosan nanoparticles containing RFP sequence siRNA, the red fluorescence in tumor site was reduced, proving the *in vivo* gene silencing effect (Figure 6.5d).

Howard and colleagues showed therapeutic results of chitosan/siRNA nanoparticles by intraperitoneal injection [85]. They used an unmodified chitosan with a molecular weight of 114 kDa and a deacetylation degree of 84%. At pH 5.5, chitosan nanoparticles were formed with siRNA of tumor necrosis factor-α (TNF-α)

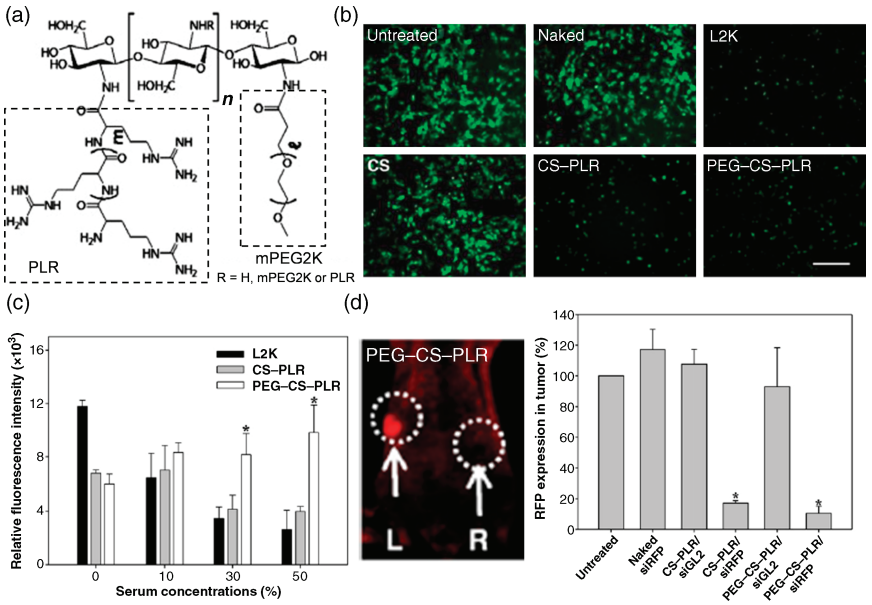


**FIGURE 6.5** Chitosan nanoparticles for gene delivery. (a) Chemical structure of PEG–polyarginine–chitosan (PEG–CS–PLR). (b) *In vitro* silencing of green fluorescence protein (GFP) expression with chitosan nanoparticles. (c) Cellular uptake analysis using fluorescently labeled siRNA. *denotes significant difference from other groups ($p < 0.05$) (d) *In vivo* silencing of red fluorescence protein (RFP) expression with chitosan nanoparticles (white circles denote tumor site). (*Source*: Figures reproduced with permission from Reference [84].)

sequence. During *in vitro* transfection in murine peritoneal macrophages, the chitosan nanoparticles showed efficient knockdown of TNF-α gene. Then, a collagen type II-induced arthritis (CIA) mouse model was established by immunization with an arthrogen-CIA collagen emulsion. After intraperitoneal injection of chitosan nanoparticles, significant anti-inflammatory effect was observed in the CIA model.

For intravenous injection, chitosan nanoparticles containing genes should provide long-term circulation and resist aggregation in blood. Huh et al. reported that glycol chitosan/PEI nanoparticles could be applied to tumor-targeted siRNA delivery via intravenous injection [86]. They made glycol chitosan/PEI nanoparticles by conjugation with 5β-cholanic acid to each polymer. The resulting mixture with two polymer conjugates were self-assembled into a nanoparticle form in an aqueous condition prior to siRNA addition. PEI enhanced the electronic interactions with siRNA, and the glycol group of glycol chitosan enhanced water solubility and inhibited aggregation with serum proteins [87]. In tumor-bearing mice models, this nanoparticle formulation showed approximately fivefold higher accumulation of siRNA compared to the free form of siRNA. *In vivo* gene silencing was also demonstrated in an RFP/B16F10 tumor-bearing mouse model. These results demonstrate the huge potential of chitosan nanoparticles for clinical application as gene carriers.

### 6.3.3 Photodynamic Therapy

Photodynamic therapy is one of the therapeutic modalities in clinical fields, mainly used in treatments of cancers or skin diseases [88]. It employs photosensitizers, which can generate cytotoxic singlet oxygen when irradiated at an appropriate wavelength. Consequently, targeted delivery of these photosensitizers can result in cell death or tissue destruction in disease site for therapeutic purpose [89]. If they are not delivered to target disease site properly, the therapeutic efficacy reduces and they can result in unintended phototoxicity in normal tissues. Therefore, efficient delivery systems of photosensitizers are needed for successful photodynamic therapy [90].

In 2010, Kim's group applied the amphiphilic glycol chitosan nanoparticles for tumor-targeted delivery of a hydrophobic photosensitizer, protophorphyrin IX (PpIX) [91]. The nanoparticles achieved faster cellular uptake of PpIX than free PpIX, and this uptake rate reflected the efficient death of tumor cells. The *in vivo* data showed high accumulation of PpIX in tumor site during optical imaging and successful tumor suppression in a SCC7 tumor-bearing mouse model upon laser irradiation.

One of the attractive features of photosensitizers is the simultaneous generation of fluorescence with singlet oxygen upon irradiation [92]. This intrinsic fluorescence can provide easy tracking of the photosensitizers in both cell and animal models [93,94]. Kim's group also reported that this fluorescence can be applied to the time-dependent tracking of photosensitizers in a tumor-bearing mouse model by optical imaging [95]. They synthesized two types of glycol chitosan nanoparticles based on physical loading and chemical conjugation of photosensitizers, and imaged biodistribution of the photosensitizers noninvasively by NIR fluorescence (Figure 6.6a).

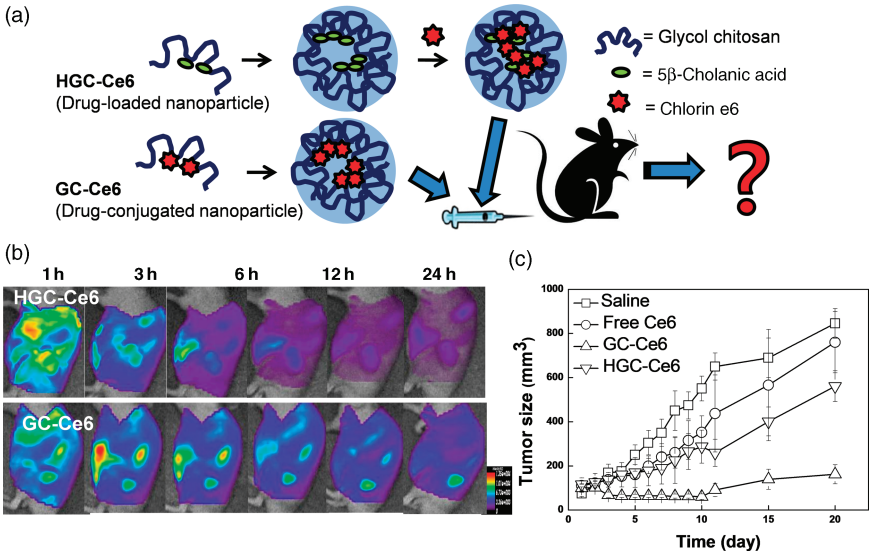


**FIGURE 6.6** Chitosan nanoparticles for photodynamic therapy. (a) Schematic illustration of two kinds of glycol chitosan nanoparticles containing photosensitizer, chlorin e6 (Ce6). (b) *In vivo* biodistribution of photosensitizers delivered by chitosan nanoparticles. (c) *In vivo* photodynamic therapy with chitosan nanoparticles. (*Source*: Figures reproduced with permission from Reference [95].)

The results showed that the loss of photosensitizers by unintended burst release was significantly large in case of physical loading (Figure 6.6b). This loss of photosensitizers could result in lower therapeutic efficacy when compared to photosensitizer-conjugated nanoparticles (Figure 6.6c). This problem of burst release during blood circulation should be overcome by innovative methods to obtain improved therapeutic results [96].

## 6.4 PROMISES, LIMITATIONS, AND FUTURE PERSPECTIVES OF CHITOSAN NANOPARTICLES

Until now, various forms of nanoparticles have been developed based on chitosan for biomedical applications including imaging, drug delivery, and gene delivery [97]. Chitosan is a biodegradable and biocompatible natural polymer with primary amine groups and cationic charges, and many researchers have explored chitosan due to its unique properties. Consequently, valuable results have been obtained with chitosan nanoparticles in cell and animal models. However, some hurdles remain to be overcome for their clinical applications.

The first challenge is the approval from the Food and Drug Administration (FDA) [98]. For commercialization of chitosan nanoparticles for clinical applications, only an FDA-approved form of chitosan should be used. Various types of modifications or

conjugations are performed to improve the functionality of chitosan [99]. However, additional synthetic modification steps are likely to delay the commercialization and clinical translation of nanoparticles.

The second challenge is the quality control of chitosan nanoparticles during manufacture. Physicochemical properties such as size, potential, and molecular weight of chitosan nanoparticles are not as uniform as those of single molecular drugs or imaging agents. This heterogeneity may not be a significant issue at an experimental level, but it may become a potential obstacle from a commercial standpoint. In fact, quality control is a common challenge for most nanoparticle systems for biomedical applications.

While chitosan nanoparticles have a long way to go, there are a number of exciting ongoing studies that employ chitosan in developing nanoparticles for biomedical applications [100]. Investigations are warranted with regard to the potential of chitosan nanoparticles as a system to improve the quality of clinical treatments in the foreseeable future.

## REFERENCES

1. Petros, R. A., DeSimone, J. M. (2010). Strategies in the design of nanoparticles for therapeutic applications. *Nature Reviews Drug Discovery 9*, 615–627.
2. Koo, H., Huh, M. S., Ryu, J. H., Lee, D.-E., Sun, I.-C., Choi, K., Kim, K., Kwon, I. C. (2011). Nanoprobes for biomedical imaging in living systems. *Nano Today 6*, 204–220.
3. Park, K., Lee, S., Kang, E., Kim, K., Choi, K., Kwon, I. C. (2009). New generation of multifunctional nanoparticles for cancer imaging and therapy. *Advanced Functional Materials 19*, 1553–1566.
4. Xie, J., Lee, S., Chen, X. (2010). Nanoparticle-based theranostic agents. *Advanced Drug Delivery Reviews 62*, 1064–1079.
5. Kim, J.-H., Park, K., Nam, H. Y., Lee, S., Kim, K., Kwon, I. C. (2007). Polymers for bioimaging. *Progress in Polymer Science 32*, 1031–1053.
6. Fox, M. E., Szoka, F. C., Fréchet, J. M. J. (2009). Soluble polymer carriers for the treatment of cancer: the importance of molecular architecture. *Accounts of Chemical Research 42*, 1141–1151.
7. Bae, Y., Kataoka, K. (2009). Intelligent polymeric micelles from functional poly(ethylene glycol)-poly(amino acid) block copolymers. *Advanced Drug Delivery Reviews 61*, 768–784.
8. Yhee, J., Koo, H., Lee, D., Choi, K., Kwon, I., Kim, K. (2011). Multifunctional chitosan nanoparticles for tumor imaging and therapy. *Advances in Polymer Science 243*, 139–161.
9. Agrawal, P., Strijkers, G. J., Nicolay, K. (2010). Chitosan-based systems for molecular imaging. *Advanced Drug Delivery Reviews 62*, 42–58.
10. Dash, M., Chiellini, F., Ottenbrite, R. M., Chiellini, E. (2011). Chitosan—A versatile semi-synthetic polymer in biomedical applications. *Progress in Polymer Science 36*, 981–1014.

11. Liu, X.-P., Zhou, S.-T., Li, X.-Y., Chen, X.-C., Zhao, X., Qian, Z.-Y., Zhou, L.-N., Li, Z.-Y., Wang, Y.-M., Zhong, Q., Yi, T., Li, Z.-Y., He, X., Wei, Y.-Q. (2010). Anti-tumor activity of *N*-trimethyl chitosan-encapsulated camptothecin in a mouse melanoma model. *Journal of Experimental and Clinical Cancer Research 29*, 76.
12. Bhattacharya, D., Das, M., Mishra, D., Banerjee, I., Sahu, S. K., Maiti, T. K., Pramanik, P. (2011). Folate receptor targeted, carboxymethyl chitosan functionalized iron oxide nanoparticles: a novel ultradispersed nanoconjugates for bimodal imaging. *Nanoscale 3*, 1653–1662.
13. Artan, M., Karadeniz, F., Karagozlu, M. Z., Kim, M.-M., Kim, S.-K. (2010). Anti-HIV-1 activity of low molecular weight sulfated chitooligosaccharides. *Carbohydrate Research 345*, 656–662.
14. Bae, K., Moon, C., Lee, Y., Park, T. (2009). Intracellular delivery of heparin complexed with chitosan-g-poly(ethylene glycol) for inducing apoptosis. *Pharmaceutical Research 26*, 93–100.
15. Lee, S. J., Koo, H., Lee, D.-E., Min, S., Lee, S., Chen, X., Choi, Y., Leary, J. F., Park, K., Jeong, S. Y., Kwon, I. C., Kim, K., Choi, K. (2011). Tumor-homing photosensitizer-conjugated glycol chitosan nanoparticles for synchronous photodynamic imaging and therapy based on cellular on/off system. *Biomaterials 32*, 4021–4029.
16. Hwang, H.-Y., Kim, I.-S., Kwon, I. C., Kim, Y.-H. (2008). Tumor targetability and antitumor effect of docetaxel-loaded hydrophobically modified glycol chitosan nanoparticles. *Journal of Controlled Release 128*, 23–31.
17. Huang, M., Khor, E., Lim, L.-Y. (2004). Uptake and cytotoxicity of chitosan molecules and nanoparticles: effects of molecular weight and degree of deacetylation. *Pharmaceutical Research 21*, 344–353.
18. Mao, S., Sun, W., Kissel, T. (2010). Chitosan-based formulations for delivery of DNA and siRNA. *Advanced Drug Delivery Reviews 62*, 12–27.
19. Noh, S. M., Han, S. E., Shim, G., Lee, K. E., Kim, C.-W., Han, S. S., Choi, Y., Kim, Y. K., Kim, W.-K., Oh, Y.-K. (2011). Tocopheryl oligochitosan-based self assembling oligomersomes for siRNA delivery. *Biomaterials 32*, 849–857.
20. Kim, Y., Jiang, H., Choi, Y., Park, I., Cho, M., Cho, C. (2011). Polymeric nanoparticles of chitosan derivatives as DNA and siRNA carriers. *Advances in Polymer Science 243*, 1–21.
21. Kim, K., Kim, J. H., Park, H., Kim, Y.-S., Park, K., Nam, H., Lee, S., Park, J. H., Park, R.-W., Kim, I.-S., Choi, K., Kim, S. Y., Park, K., Kwon, I. C. (2010). Tumor-homing multifunctional nanoparticles for cancer theragnosis: simultaneous diagnosis, drug delivery, and therapeutic monitoring. *Journal of Controlled Release 146*, 219–227.
22. Park, K., Kim, J.-H., Nam, Y. S., Lee, S., Nam, H. Y., Kim, K., Park, J. H., Kim, I.-S., Choi, K., Kim, S. Y., Kwon, I. C. (2007). Effect of polymer molecular weight on the tumor targeting characteristics of self-assembled glycol chitosan nanoparticles. *Journal of Controlled Release 122*, 305–314.
23. Baldrick, P. (2010) The safety of chitosan as a pharmaceutical excipient. *Regulatory Toxicology and Pharmacology 56*, 290–299.
24. Wedmore, I., McManus, J. G., Pusateri, A. E., Holcomb, J. B. (2006). A special report on the chitosan-based hemostatic dressing: experience in current combat operations. *The Journal of Trauma 60*, 655–658.

25. Kean, T., Thanou, M. (2010). Biodegradation, biodistribution and toxicity of chitosan. *Advanced Drug Delivery Reviews 62*, 3–11.
26. Hilderbrand, S. A., Weissleder, R. (2010). Near-infrared fluorescence: application to *in vivo* molecular imaging. *Current Opinion in Chemical Biology 14*, 71–79.
27. Lee, S., Park, K., Kim, K., Choi, K., Kwon, I. C. (2008). Activatable imaging probes with amplified fluorescent signals. *Chemical Communications (Cambridge, England)* 4250–4260.
28. Weissleder, R. (2001). A clearer vision for *in vivo* imaging. *Nature Biotechnology 19*, 316–317.
29. Ranjan, A. P., Zeglam, K., Mukerjee, A., Thamake, S., Vishwanatha, J. K. (2011). A sustained release formulation of chitosan modified PLCL: poloxamer blend nanoparticles loaded with optical agent for animal imaging. *Nanotechnology 22*, 295104.
30. Na, J. H., Koo, H., Lee, S., Min, K. H., Park, K., Yoo, H., Lee, S. H., Park, J. H., Kwon, I. C., Jeong, S. Y., Kim, K. (2011). Real-time and non-invasive optical imaging of tumor-targeting glycol chitosan nanoparticles in various tumor models. *Biomaterials 32*, 5252–5261.
31. Park, J. H., Kwon, S., Nam, J.-O., Park, R.-W., Chung, H., Seo, S. B., Kim, I.-S., Kwon, I. C., Jeong, S. Y. (2004). Self-assembled nanoparticles based on glycol chitosan bearing 5[beta]-cholanic acid for RGD peptide delivery. *Journal of Controlled Release 95*, 579–588.
32. Torchilin, V. (2011). Tumor delivery of macromolecular drugs based on the EPR effect. *Advanced Drug Delivery Reviews 63*, 131–135.
33. Kim, J.-H., Kim, Y.-S., Park, K., Lee, S., Nam, H. Y., Min, K. H., Jo, H. G., Park, J. H., Choi, K., Jeong, S. Y., Park, R.-W., Kim, I.-S., Kim, K., Kwon, I. C. (2008). Antitumor efficacy of cisplatin-loaded glycol chitosan nanoparticles in tumor-bearing mice. *Journal of Controlled Release 127*, 41–49.
34. Nam, H. Y., Kwon, S. M., Chung, H., Lee, S.-Y., Kwon, S.-H., Jeon, H., Kim, Y., Park, J. H., Kim, J., Her, S., Oh, Y.-K., Kwon, I. C., Kim, K., Jeong, S. Y. (2009). Cellular uptake mechanism and intracellular fate of hydrophobically modified glycol chitosan nanoparticles. *Journal of Controlled Release 135*, 259–267.
35. Maeda, H. (2010). Tumor-selective delivery of macromolecular drugs via the EPR effect: background and future prospects. *Bioconjugate Chemistry 21*, 797–802.
36. Na, H. B., Song, I. C., Hyeon, T. (2009). Inorganic nanoparticles for MRI contrast agents. *Advanced Materials 21*, 2133–2148.
37. Lu, Z.-R., Wang, X., Parker, D. L., Goodrich, K. C., Buswell, H. R. (2003). Poly(l-glutamic acid) Gd(III)-DOTA conjugate with a degradable spacer for magnetic resonance imaging. *Bioconjugate Chemistry 14*, 715–719.
38. Yim, H., Yang, S.-G., Jeon, Y. S., Park, I. S., Kim, M., Lee, D. H., Bae, Y. H., Na, K. (2011). The performance of gadolinium diethylene triamine pentaacetate-pullulan hepatocyte-specific T1 contrast agent for MRI. *Biomaterials 32*, 5187–5194.
39. Nam, T., Park, S., Lee, S.-Y., Park, K., Choi, K., Song, I. C., Han, M. H., Leary, J. J., Yuk, S. A., Kwon, I. C., Kim, K., Jeong, S. Y. (2010). Tumor targeting chitosan nanoparticles for dual-modality optical/MR cancer imaging. *Bioconjugate Chemistry 21*, 578–582.
40. Cheon, J., Lee, J.-H. (2008). Synergistically integrated nanoparticles as multimodal probes for nanobiotechnology. *Accounts of Chemical Research 41*, 1630–1640.

41. Gao, J., Gu, H., Xu, B. (2009). Multifunctional magnetic nanoparticles: design, synthesis, and biomedical applications. *Accounts of Chemical Research 42*, 1097–1107.

42. Lee, C.-M., Jeong, H.-J., Kim, S.-L., Kim, E.-M., Kim, D. W., Lim, S. T., Jang, K. Y., Jeong, Y. Y., Nah, J.-W., Sohn, M.-H. (2009). SPION-loaded chitosan-linoleic acid nanoparticles to target hepatocytes. *International Journal of Pharmaceutics 371*, 163–169.

43. Lee, C.-M., Jang, D., Kim, J., Cheong, S.-J., Kim, E.-M., Jeong, M.-H., Kim, S.-H., Kim, D. W., Lim, S. T., Sohn, M.-H., Jeong, Y. Y., Jeong, H.-J. (2011). Oleyl-chitosan nanoparticles based on a dual probe for optical/MR imaging *in vivo*. *Bioconjugate Chemistry 22*, 186–192.

44. Kweon, S., Lee, H.-J., Hyung, W., Suh, J., Lim, J., Lim, S.-J. (2010). Liposomes coloaded with iopamidol/lipiodol as a RES-targeted contrast agent for computed tomography imaging. *Pharmaceutical Research 27*, 1408–1415.

45. Sun, I.-C., Eun, D.-K., Na, J. H., Lee, S., Kim, I.-J., Youn, I.-C., Ko, C.-Y., Kim, H.-S., Lim, D., Choi, K., Messersmith, P. B., Park, T. G., Kim, S. Y., Kwon, I. C., Kim, K., Ahn, C.-H. (2009). Heparin-coated gold nanoparticles for liver-specific CT imaging. *Chemistry - A European Journal 15*, 13341–13347.

46. Kim, D., Jeong, Y. Y., Jon, S. (2010). A drug-loaded aptamer−gold nanoparticle bioconjugate for combined CT imaging and therapy of prostate cancer. *ACS Nano 4*, 3689–3696.

47. Sun, I.-C., Eun, D.-K., Koo, H., Ko, C.-Y., Kim, H.-S., Yi, D. K., Choi, K., Kwon, I. C., Kim, K., Ahn, C.-H. (2011). Tumor targeting gold particles for CT/optical dual cancer imaging. *Angewandte Chemie International Edition 50*, 9348–9351.

48. Lee, S., Ryu, J. H., Park, K., Lee, A., Lee, S.-Y., Youn, I.-C., Ahn, C.-H., Yoon, S. M., Myung, S.-J., Moon, D. H., Chen, X., Choi, K., Kwon, I. C., Kim, K. (2009). Polymeric nanoparticle-based activatable near-infrared nanosensor for protease determination *in vivo*. *Nano Letters 9*, 4412–4416.

49. Louie, A. (2010). Multimodality imaging probes: design and challenges. *Chemical Reviews 110*, 3146–3195.

50. Chen, K., Conti, P. S. (2010). Target-specific delivery of peptide-based probes for PET imaging. *Advanced Drug Delivery Reviews 62*, 1005–1022.

51. Cho, Y. W., Park, S. A., Han, T. H., Son, D. H., Park, J. S., Oh, S. J., Moon, D. H., Cho, K.-J., Ahn, C.-H., Byun, Y., Kim, I.-S., Kwon, I. C., Kim, S. Y. (2007). *In vivo* tumor targeting and radionuclide imaging with self-assembled nanoparticles: mechanisms, key factors, and their implications. *Biomaterials 28*, 1236–1247.

52. Yang, W., Mou, T., Guo, W., Jing, H., Peng, C., Zhang, X., Ma, Y., Liu, B. (2010). Fluorine-18 labeled galactosylated chitosan for asialoglycoprotein-receptor-mediated hepatocyte imaging. *Bioorganic & Medicinal Chemistry Letters 20*, 4840–4844.

53. Frisch, B., Carriere, M., Largeau, C., Mathey, F., Masson, C., Schuber, F., Scherman, D., Escriou, V. (2004). A new triantennary galactose-targeted PEGylated gene carrier, characterization of its complex with DNA, and transfection of hepatoma cells. *Bioconjugate Chemistry 15*, 754–764.

54. Tsao, N., Wang, C.-H., Her, L.-J., Tzen, K.-Y., Chen, J.-Y., Yu, D.-F., Yang, D. J. (2011). Development of 68Ga-glycopeptide as an imaging probe for tumor angiogenesis. *Journal of Biomedicine and Biotechnology 2011*, 267206.

55. Zhernosekov, K. P., Filosofov, D. V., Baum, R. P., Aschoff, P., Bihl, H., Razbash, A. A., Jahn, M., Jennewein, M., Rösch, F. (2007). Processing of generator-produced 68Ga for medical application. *Journal of Nuclear Medicine 48*, 1741–1748.
56. Min, K. H., Park, K., Kim, Y.-S., Bae, S. M., Lee, S., Jo, H. G., Park, R.-W., Kim, I.-S., Jeong, S. Y., Kim, K., Kwon, I. C. (2008). Hydrophobically modified glycol chitosan nanoparticles-encapsulated camptothecin enhance the drug stability and tumor targeting in cancer therapy. *Journal of Controlled Release 127*, 208–218.
57. Deshpande, N., Needles, A., Willmann, J. K. (2010). Molecular ultrasound imaging: current status and future directions. *Clinical Radiology 65*, 567–581.
58. Voigt, J.-U. (2009). Ultrasound molecular imaging. *Methods 48*, 92–97.
59. Lindner, J. R. (2004). Microbubbles in medical imaging: current applications and future directions. *Nature Reviews Drug Discovery 3*, 527–533.
60. Liu, Y.-L., Wu, Y.-H., Tsai, W.-B., Tsai, C.-C., Chen, W.-S., Wu, C.-S. (2011). Core-shell silica@chitosan nanoparticles and hollow chitosan nanospheres using silica nanoparticles as templates: preparation and ultrasound bubble application. *Carbohydrate Polymers 84*, 770–774.
61. Lin, Y.-H., Sonaje, K., Lin, K. M., Juang, J.-H., Mi, F.-L., Yang, H.-W., Sung, H.-W. (2008). Multi-ion-crosslinked nanoparticles with pH-responsive characteristics for oral delivery of protein drugs. *Journal of Controlled Release 132*, 141–149.
62. Nasti, A., Zaki, N., de Leonardis, P., Ungphaiboon, S., Sansongsak, P., Rimoli, M., Tirelli, N. (2009). Chitosan/TPP and chitosan/TPP-hyaluronic acid nanoparticles: systematic optimisation of the preparative process and preliminary biological evaluation. *Pharmaceutical Research 26*, 1918–1930.
63. Suknuntha, K., Tantishaiyakul, V., Worakul, N., Taweepreda, W. (2011). Characterization of muco- and bioadhesive properties of chitosan, PVP, and chitosan/PVP blends and release of amoxicillin from alginate beads coated with chitosan/PVP. *Drug Development and Industrial Pharmacy 37*, 408–418.
64. Ranaldi, G., Marigliano, I., Vespignani, I., Perozzi, G., Sambuy, Y. (2002). The effect of chitosan and other polycations on tight junction permeability in the human intestinal Caco-2 cell line. *The Journal of Nutritional Biochemistry 13*, 157–167.
65. Lee, E., Kim, H., Lee, I.-H., Jon, S. (2009). *In vivo* antitumor effects of chitosan-conjugated docetaxel after oral administration. *Journal of Controlled Release 140*, 79–85.
66. Lee, E., Lee, J., Lee, I.-H., Yu, M., Kim, H., Chae, S. Y., Jon, S. (2008). Conjugated chitosan as a novel platform for oral delivery of paclitaxel. *Journal of Medicinal Chemistry 51*, 6442–6449.
67. George, M., Abraham, T. E. (2006). Polyionic hydrocolloids for the intestinal delivery of protein drugs: alginate and chitosan—a review. *Journal of Controlled Release 114*, 1–14.
68. Sonaje, K., Lin, Y.-H., Juang, J.-H., Wey, S.-P., Chen, C.-T., Sung, H.-W. (2009). *In vivo* evaluation of safety and efficacy of self-assembled nanoparticles for oral insulin delivery. *Biomaterials 30*, 2329–2339.
69. Sonaje, K., Lin, K.-J., Wey, S.-P., Lin, C.-K., Yeh, T.-H., Nguyen, H.-N., Hsu, C.-W., Yen, T.-C., Juang, J.-H., Sung, H.-W. (2010). Biodistribution, pharmacodynamics and pharmacokinetics of insulin analogues in a rat model: oral delivery using pH-Responsive nanoparticles vs. subcutaneous injection. *Biomaterials 31*, 6849–6858.

70. Chen, M.-C., Wong, H.-S., Lin, K.-J., Chen, H.-L., Wey, S.-P., Sonaje, K., Lin, Y.-H., Chu, C.-Y., Sung, H.-W. (2009). The characteristics, biodistribution and bioavailability of a chitosan-based nanoparticulate system for the oral delivery of heparin. *Biomaterials 30*, 6629–6637.
71. Park, J. H., Saravanakumar, G., Kim, K., Kwon, I. C. (2010). Targeted delivery of low molecular drugs using chitosan and its derivatives. *Advanced Drug Delivery Reviews 62*, 28–41.
72. Zhao, Z., He, M., Yin, L., Bao, J., Shi, L., Wang, B., Tang, C., Yin, C. (2009). Biodegradable nanoparticles based on linoleic acid and poly(β-malic acid) double grafted chitosan derivatives as carriers of anticancer drugs. *Biomacromolecules 10*, 565–572.
73. Lee, B. S., Park, K., Park, S., Kim, G. C., Kim, H. J., Lee, S., Kil, H., Oh, S. J., Chi, D., Kim, K., Choi, K., Kwon, I. C., Kim, S. Y. (2010). Tumor targeting efficiency of bare nanoparticles does not mean the efficacy of loaded anticancer drugs: importance of radionuclide imaging for optimization of highly selective tumor targeting polymeric nanoparticles with or without drug. *Journal of Controlled Release 147*, 253–260.
74. Willmann, J. K., van Bruggen, N., Dinkelborg, L. M., Gambhir, S. S. (2008). Molecular imaging in drug development. *Nature Reviews Drug Discovery 7*, 591–607.
75. Rudzinski, W. E., Aminabhavi, T. M. (2010). Chitosan as a carrier for targeted delivery of small interfering RNA. *International Journal of Pharmaceutics 399*, 1–11.
76. Bowman, K., Sarkar, R., Raut, S., Leong, K. W. (2008). Gene transfer to hemophilia A mice via oral delivery of FVIII-chitosan nanoparticles. *Journal of Controlled Release 132*, 252–259.
77. Mohri, K., Okuda, T., Mori, A., Danjo, K., Okamoto, H. (2010). Optimized pulmonary gene transfection in mice by spray-freeze dried powder inhalation. *Journal of Controlled Release 144*, 221–226.
78. Byrne, J. D., Betancourt, T., Brannon-Peppas, L. (2008). Active targeting schemes for nanoparticle systems in cancer therapeutics. *Advanced Drug Delivery Reviews 60*, 1615–1626.
79. Zheng, Y., Cai, Z., Song, X., Yu, B., Bi, Y., Chen, Q., Zhao, D., Xu, J., Hou, S. (2009). Receptor mediated gene delivery by folate conjugated N-trimethyl chitosan *in vitro*. *International Journal of Pharmaceutics 382*, 262–269.
80. Sahni, J. K., Chopra, S., Ahmad, F. J., Khar, R. K. (2008). Potential prospects of chitosan derivative trimethyl chitosan chloride (TMC) as a polymeric absorption enhancer: synthesis, characterization and applications. *Journal of Pharmacy and Pharmacology 60*, 1111–1119.
81. Jiang, H. L., Kwon, J. T., Kim, Y. K., Kim, E. M., Arote, R., Jeong, H. J., Nah, J. W., Choi, Y. J., Akaike, T., Cho, M. H., Cho, C. S. (2007). Galactosylated chitosan-graft-polyethylenimine as a gene carrier for hepatocyte targeting. *Gene Therapy 14*, 1389–1398.
82. Castanotto, D., Rossi, J. J. (2009). The promises and pitfalls of RNA-interference-based therapeutics. *Nature 457*, 426–433.
83. Oh, Y.-K., Park, T. G. (2009). siRNA delivery systems for cancer treatment. *Advanced Drug Delivery Reviews 61*, 850–862.

84. Noh, S. M., Park, M. O., Shim, G., Han, S. E., Lee, H. Y., Huh, J. H., Kim, M. S., Choi, J. J., Kim, K., Kwon, I. C., Kim, J.-S., Baek, K.-H., Oh, Y.-K. (2010). Pegylated poly-l-arginine derivatives of chitosan for effective delivery of siRNA. *Journal of Controlled Release 145*, 159–164.
85. Howard, K. A., Paludan, S. R., Behlke, M. A., Besenbacher, F., Deleuran, B., Kjems, J. (2008). Chitosan/siRNA nanoparticle-mediated TNF-[alpha] knockdown in peritoneal macrophages for anti-inflammatory treatment in a murine arthritis model. *Molecular Therapy 17*, 162–168.
86. Huh, M. S., Lee, S.-Y., Park, S., Lee, S., Chung, H., Lee, S., Choi, Y., Oh, Y.-K., Park, J. H., Jeong, S. Y., Choi, K., Kim, K., Kwon, I. C. (2010). Tumor-homing glycol chitosan/polyethylenimine nanoparticles for the systemic delivery of siRNA in tumor-bearing mice. *Journal of Controlled Release 144*, 134–143.
87. Saravanakumar, G., Min, K. H., Min, D. S., Kim, A. Y., Lee, C.-M., Cho, Y. W., Lee, S. C., Kim, K., Jeong, S. Y., Park, K., Park, J. H., Kwon, I. C. (2009). Hydrotropic oligomer-conjugated glycol chitosan as a carrier of paclitaxel: synthesis, characterization, and *in vivo* biodistribution. *Journal of Controlled Release 140*, 210–217.
88. Moore, C. M., Pendse, D., Emberton, M. (2009). Photodynamic therapy for prostate cancer—a review of current status and future promise. *Nature Clinical Practice Urology 6*, 18–30.
89. Chatterjee, D. K., Fong, L. S., Zhang, Y. (2008). Nanoparticles in photodynamic therapy: an emerging paradigm. *Advanced Drug Delivery Reviews 60*, 1627–1637.
90. van Nostrum, C. F. (2004). Polymeric micelles to deliver photosensitizers for photodynamic therapy. *Advanced Drug Delivery Reviews 56*, 9–16.
91. Lee, S. J., Park, K., Oh, Y.-K., Kwon, S.-H., Her, S., Kim, I.-S., Choi, K., Lee, S. J., Kim, H., Lee, S. G., Kim, K., Kwon, I. C. (2009). Tumor specificity and therapeutic efficacy of photosensitizer-encapsulated glycol chitosan-based nanoparticles in tumor-bearing mice. *Biomaterials 30*, 2929–2939.
92. Celli, J. P., Spring, B. Q., Rizvi, I., Evans, C. L., Samkoe, K. S., Verma, S., Pogue, B. W., Hasan, T. (2010). Imaging and photodynamic therapy: mechanisms, monitoring, and optimization. *Chemical Reviews 110*, 2795–2838.
93. Koo, H., Lee, H., Lee, S., Min, K. H., Kim, M. S., Lee, D. S., Choi, Y., Kwon, I. C., Kim, K., Jeong, S. Y. (2010). *In vivo* tumor diagnosis and photodynamic therapy via tumoral pH-responsive polymeric micelles. *Chemical Communications 46*, 5668–5670.
94. Park, S., Lee, S. J., Chung, H., Her, S., Choi, Y., Kim, K., Choi, K., Kwon, I. C. (2010). Cellular uptake pathway and drug release characteristics of drug-encapsulated glycol chitosan nanoparticles in live cells. *Microscopy Research and Technique 73*, 857–865.
95. Lee, S. J., Koo, H., Jeong, H., Huh, M. S., Choi, Y., Jeong, S. Y., Byun, Y., Choi, K., Kim, K., Kwon, I. C. (2011). Comparative study of photosensitizer loaded and conjugated glycol chitosan nanoparticles for cancer therapy. *Journal of Controlled Release 152*, 21–29.
96. Lee, H. J., Kim, S. E., Kwon, I. K., Park, C., Kim, C., Yang, J., Lee, S. C. (2010). Spatially mineralized self-assembled polymeric nanocarriers with enhanced robustness and controlled drug-releasing property. *Chemical Communications 46*, 377–379.
97. Peniche, H., Peniche, C. (2011). Chitosan nanoparticles: a contribution to nanomedicine. *Polymer International 60*, 883–889.

98. Eifler, A. C., Thaxton, C. S., Nanoparticle therapeutics: FDA approval, clinical trials, regulatory pathways, and case study, In *Biomedical Nanotechnology*, ed. Hurst, S. J., Humana Press, 2011, pp 325–338.
99. Alves, N. M., Mano, J. F. (2008). Chitosan derivatives obtained by chemical modifications for biomedical and environmental applications. *International Journal of Biological Macromolecules 43*, 401–414.
100. Liu, Z., Jiao, Y., Wang, Y., Zhou, C., Zhang, Z. (2008). Polysaccharides-based nanoparticles as drug delivery systems. *Advanced Drug Delivery Reviews 60*, 1650–1662.

# POLYMER–DRUG NANOCONJUGATES

Rong Tong

*Department of Chemical Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA Laboratory for Biomaterials and Drug Delivery, Department of Anesthesiology, Division of Critical Care Medicine, Children's Hospital Boston, Harvard Medical School, Boston, MA, USA*

Li Tang, Nathan P. Gabrielson, Qian Yin, and Jianjun Cheng

*Department of Materials Science and Engineering, University of Illinois at Urbana–Champaign, Urbana, IL, USA*

## 7.1 INTRODUCTION

Polymeric nanomedicine, an emerging field that involves the use of drug-containing polymeric nanoparticles (NPs) for cancer treatment, is expected to alter the landscape of oncology [1]. The medical application of nanotechnology has been extensive: roughly 40 nanomedicines have already been approved by the Food and Drug Administration (FDA) for clinical use [2–4] and a handful of NPs are currently in preclinical investigations [2,5]. Incorporation of chemotherapeutic agents in NP delivery vehicles can improve water solubility, reduce clearance, reduce drug resistance, and enhance therapeutic effectiveness [6]. Broadly speaking, two approaches have been used to load drugs in NPs for delivery. One is to encapsulate drugs within NPs via noncovalent bonds, that is, distribution of the drug throughout a polymeric matrix during formulation (Figure 7.1a) [7]. The second loading strategy is the formulation of NPs using polymer–drug conjugates, a technique first proposed in 1975 (Figure 7.1b) [8]. In this chapter, we discuss the pros and cons of the two strategies and introduce newly developed polymer-drug conjugates—so-called nanoconjugates—as a formulation strategy which addresses challenges faced by both encapsulates and conjugates

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*,
Edited by Shayne C. Gad
Copyright © 2013 John Wiley & Sons, Inc.

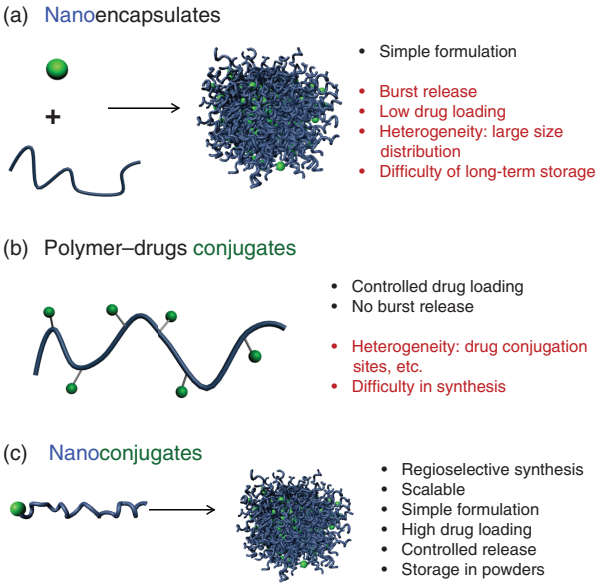


**FIGURE 7.1** Pros (black) and cons (gray) of (a) nanoencapsulates (NEs) and (b) polymer–drug conjugates. We propose (c) nanoconjugates (NCs) with reduced heterogeneity to address the challenges in both encapsulated and conjugated systems.

(Figure 7.1c). Along the way, we also discuss topics related to regioselective polymer–drug conjugation chemistry, *in vitro* formulation and characterization, and present preliminary *in vivo* studies that highlight the potential clinical translation of nanoconjugates.

## 7.2 CURRENT STATUS OF NANOENCAPSULATES AND POLYMER–DRUG CONJUGATES

### 7.2.1 Nanoencapsulates

Nanoparticles loaded with a drug via encapsulation within a polymer matrix are termed nanoencapsulates. Although nanoencapsulate formulation is straightforward in both concept and fabrication, their clinical use is challenged by a variety of issues (Figure 7.1a). First, nanoencapsulates often show "burst" drug release profiles in aqueous solution, with as much as 80–90% of the encapsulated drug released during the first few to tens of hours [9]. The rapid drug release, also called dose dumping, can prevent the drug from operating within its therapeutic window and cause severe systemic toxicity [10]. Second, drug loading in nanoencapsulates is very low—typically in the range of 1–5% for most NPs studied—and is widely variable depending on the amount of drug being used, the hydrophobicity/hydrophilicity of the drug as well as the compatibility between the drug and the polymer [9,11]. Drug loading, in particular, is a critical measure of the feasibility of nanoencapsulate delivery systems in clinical

settings [11]. At low drug loading, large amounts of delivery vehicles are needed to achieve therapeutic concentrations. However, because of the limited body weight and blood volume of animals, administration volumes are fixed. For mice with 20–30 g body weight, the intravenously injected volume must be kept around 100–200 μL [12]. Thus, the intravenous administration of NPs with 1% drug loading in a 100 μL solution at a dose of 50 mg/kg (e.g., docetaxel) would require a 1 g/mL NP solution. In practice, it is impossible to formulate such concentrated solutions and inject them intravenously. Another problem related to drug loading is the lack of a general strategy to achieve quantitative drug encapsulation in many NPs (e.g., polylactide (PLA) and poly(lactic-*co*-glycolic acid) (PLGA) NPs). In this instance, nonencapsulated drugs may self-aggregate, thus complicating their removal from the NPs [13]. Ultimately, these formulation challenges significantly impact the processability and the clinical translation of nanoencapsulate delivery vehicles for cancer therapy.

### 7.2.2 Polymer–Drug Conjugates

As difficulties facing nanoencapsulate drug loading began emerging, considerable interest was shifted to drug delivery strategies employing polymer–drug conjugates (Figure 7.1b). In fact, a polymer–protein drug conjugate has already been approved for use in Japan, and a handful of other polymer–drug conjugates are presently transitioning from later clinical trials to the wider market [10,14]. One of the first systems explored, *N*-(2-hydroxylproply)methacrylamine (HPMA) polymer–drug conjugates developed by Duncan and Kopecek in the late 1970s, has led to various clinical trial candidates [14,15]. Other interesting types of polymer–drug conjugates include cyclodextrin-containing polymers with conjugated camptothecin developed by Davis laboratory (IT-101) [16–20] and dendritic polyester–drug conjugates developed by Frechet and coworkers [21,22].

One of the primary challenges facing polymer–drug conjugate delivery systems is the actual conjugation step. Not only does the conjugation often require extensive post-conjugation purification steps, the heterogeneity of polymer–drug conjugates arising from nonsite-specific coupling reactions may also present bottlenecks to the clinical translation [10]. As shown in Figure 7.2, heterogeneities of polymer–drug conjugates result from (1) molecular weight distributions (MWDs, Mw/Mn) of the polymers (Figure 7.2a), (2) lack of control of the drug conjugation site on the polymer backbone (Figure 7.2b), and (3) lack of regioselectivity with regard to the conjugation site on drugs with multiple conjugation-amenable functional groups (Figure 7.2c). This last point is particularly relevant, as many of the best-selling anticancer small molecule drugs (e.g., paclitaxel, docetaxel, and doxorubicin) contain multiple functional groups (e.g., multihydroxyl groups in all three agents; ketone and amine groups in doxorubicin) for conjugation [23]. In the past several decades, there have been numerous efforts to minimize heterogeneities within polymer–drug conjugates. These include the development of polymers with low MWDs (e.g., dendrimers) [22,24–31], the conjugation of therapeutic agents to specific sites along the polymer backbone (e.g., the termini) [16,32–38], and the activation of specific functional groups on the therapeutic agents by means of

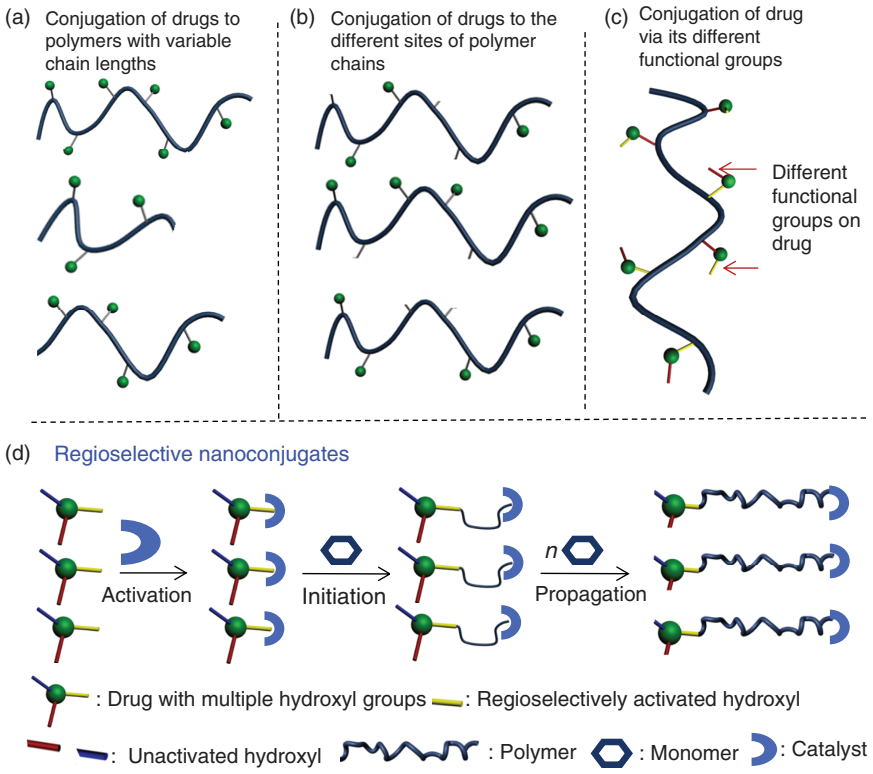


**FIGURE 7.2** Heterogeneities in polymer-drug conjugates: (a) polymer–drug conjugates with a broad distribution of polymer chain lengths; (b) polymer–drug conjugates in which the conjugation site on the polymer backbone is uncontrolled; (c) polymer–drug conjugates in which a multifunctional therapeutic agent is conjugated without regioselectivity. Our solution to the heterogeneity problem: (d) drug-initiated, controlled polymerization for the synthesis of polymer–drug conjugates with low MWDs; the drug molecules are conjugated regioselectively to the polymer termini.

protection/deprotection chemistry [39,40]. However, concise synthetic strategies that yield polymer–drug conjugates with minimal heterogeneity for clinical application are still lacking (Figure 7.2a–c).

## 7.3 NANOCONJUGATES: DESIGN AND SYNTHESIS

### 7.3.1 Design and General Consideration

We present here a new synthetic strategy to address challenges faced by both nanoencapsulates and polymer–drug conjugates. In particular, we focus on chemistry that allows for site-specific conjugation of a drug to a polymer with a narrow molecular weight distribution (MWD) with quantitative incorporation efficiency.

In designing such a synthetic strategy, we first considered chemical systems which already benefit from controlled and precise chemistry, like the well-established and controlled polymerization methodologies allowing for the preparation of polyesters [41,42], polypeptides [43,44], and hydrocarbon-based synthetic biopolymers [45] with precisely controlled molecular weights (MWs) and narrow MWD. By incorporating these existing strategies in our design, polymer–drug conjugates can be produced with precise and homogeneous chemical structure and, ultimately, improved performance.

Ring-opening polymerization (ROP) for the preparation of PLA polyesters has been investigated extensively [46,47]. The polymerization reaction typically involves lactide (LA) ring opening by a metal–alkoxide ($L_nM$–OR) to form a RO-terminated LA–metal alkoxide ($ROOCCH(CH_3)O\text{-}ML_n$) followed by chain propagation to form RO-terminated PLA [41,46,47]. According to this mechanism, the RO group ultimately is connected to the PLA terminus through an ester bond (Figure 7.3a). This process is well understood and has been used extensively for the

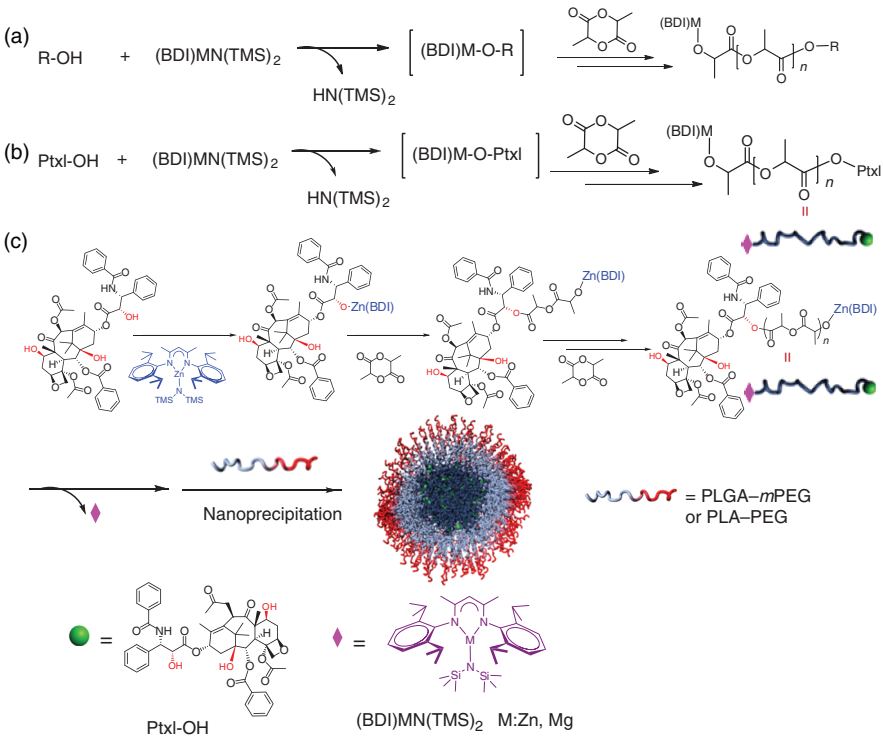


**FIGURE 7.3** (a) Mechanism of ROP of PLA initiated by R-OH/BDI-metal catalyst. We propose the mechanism can be applied to (b) complex molecules with dense hydroxyl groups, for example, Ptxl, to initiate ROP. (c) Regioselective coordination of (BDI-II)$ZnN(TMS)_2$ onto 2′-OH of Ptxl to initiate ROP of LA. The polymers can be nanoprecipitated with PLGA–mPEG or PLA–mPEG to formulate sub-120 nm NPs.

incorporation of hydroxyl-containing small molecules [38], macromolecules [48], and NPs [49] to the termini of PLA.

In the context of delivery technology, hydroxyl groups of drugs can be utilized to achieve controlled living polymerization of LA. The hydroxyl group is the most abundant functional group in natural products, being found in about 65% of the ~78,000 known pharmacophores as well as in several of the best-selling anticancer small molecule drugs. Thus, the active initiator for PLA polymerization can be prepared *in situ* by mixing a hydroxyl(s)-containing drug with an active metal complex, such as a metal-amido [41,50]. If well designed, the *in situ* formed M–ORs can initiate controlled polymerization of LA, resulting in quantitative incorporation of OR with 100% monomer conversion (Figure 7.3b) [41,46]. By judicious control of the metal catalyst ligand, it is also possible to control the regioselective coordination of a structurally complex drug molecule bearing multiple hydroxyl groups (Figure 7.2d). After ROP, the linear PLA will be covalently linked onto the specific hydroxyl group of the drug via an ester bond. PLA–drug conjugates can then be driven to form NPs—called nanoconjugates (NCs)—by nanoprecipitation methods (Figure 7.3c). Unlike nanoencapsulates, the use of a degradable ester linkage between a drug and a polymer prevents burst release from the NPs and facilitates the controlled release of drugs from the conjugates.

ROP of LA to generate polymer–drug conjugates with precise and homogenous chemistry has been examined in several studies performed by the Cheng laboratory [10,51,52]. For these studies, the ROP of LA was catalyzed by (BDI-X)MN(TMS)$_2$ (BDI = 2-((2,6-dialkylphenyl)amino)-4-((2,6-dialkylphenyl)imino)-2-pentene, M = Mg or Zn, Table 7.1), a class of catalysts developed by Coates and coworkers for the controlled ROP of LA [41,50]. The strategy allows for quantitative incorporation of multihydroxyl drugs [51] (e.g., paclitaxel (Ptxl), docetaxel (Dtxl), and doxorubicin (Doxo)), other hydroxyl-containing therapeutic molecules [52] (e.g., camptothecin (Cpt), cyanine, cyclopamine) and even small peptides (goserelin) via ester bonds. By tuning the substituents on the *N*-aryl groups ($R_1$ and $R_2$) and the β-position ($R_3$) of

**TABLE 7.1 (BDI-X)ZnN(TMS)$_2$ for NCs Synthesis**

| | Ligand | $R_1$ | $R_2$ | $R_3$ |
|---|---|---|---|---|
| 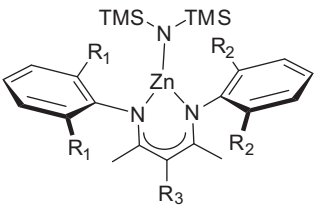  (BDI-**X**)ZnN(TMS)$_2$ | BDI-II | *i*Pr | *i*Pr | H |
| | BDI-EE | Et | Et | H |
| | BDI-EI | Et | *i*Pr | H |
| | BDI-IICN | *i*Pr | *i*Pr | CN |

the BDI ligand (Table 7.1) [53], regioselective polymerization on specific hydroxyl groups can be achieved (Table 7.2). The controlled polymerization also can be expanded to other biopolymers, such as poly(δ-valerolactone), poly(trimethylene carbonate), and poly(ε-caprolactone) (Table 7.3 for various polymers loaded with Ptxl and Dtxl). Apart from their controlled chemistry, drug–PLA conjugates can be nanoprecipitated to generate NPs with sub-120 nm diameter, high drug loading, nearly quantitative loading efficiencies, controlled release profiles without burst release, and narrow particle size distributions [51,52,54,55].

The controlled chemistry mentioned earlier was validated in the synthesis of Ptxl–PLA, and Doxo–PLA NCs. In the following discussion of these NCs, we examine the catalyst metal selection to achieve selective activation of hydroxyls in Cpt without disrupting the lactone ring structure [54]. Regioselective polymerization chemistry is also explored in the synthesis of Ptxl–PLA and Doxo–PLA NCs, whose initiator drugs possess multiple hydroxyl groups as well as other functional moieties. Combined, these experiments demonstrate that the proposed NC synthetic strategy is widely applicable for the synthesis of drug–polymer conjugates with various hydroxyl-containing therapeutics and with various polymer backbones.

### 7.3.2 Synthesis of Cpt–PLA Nanoconjugates

20(*S*)-Camptothecin (Cpt), a topoisomerase I inhibitor isolated from the Chinese tree *Camptotheca acuminate*, exhibits a broad range of anticancer activity in various animal models [56,57]. In terms of usability, Cpt has low aqueous solubility in its therapeutically active lactone form and is transformed rapidly to its carboxylate analog at physiological pH, producing a highly toxic and therapeutically inactive molecule (Figure 7.4a) [58,59]. While Cpt–polymer conjugates have been prepared to overcome the solubility limitation of the drug, conjugates prepared with conventional coupling chemistry are plagued by various heterogeneities. For example, experiments exploring the conjugation of Cpt via condensation reactions yield MWDs over a range 1.5–2.5 [16]. Meanwhile, direct conjugation of Cpt through its C20-tertiary hydroxyl is a difficult multistep reaction: Cpt must first be converted to a Cpt–amino ester and then conjugated to a polymer containing carboxylate groups via the amine end group of the Cpt–amino ester [16,60]. If the polymer has pendant functional groups, Cpt conjugation is further complicated.

To synthesize Cpt–PLA NCs via ROP, we first explored its ability to form Cpt–metal complexes with various BDI catalysts. When mixed with (BDI-II)MgN$(TMS)_2$, the C20-OH of Cpt formed a (BDI-II)Mg-Cpt alkoxide *in situ*. ROP of LA with this complex resulted in 100% LA conversion. HPLC analysis showed that Cpt was entirely conjugated to PLA with no detectable free Cpt in the polymerization solution. Although the $M_n$ of the resulting polymer was in good agreement with the expected $M_n$, the MWD of Cpt–$LA_{100}$ was relatively broad (1.31) due in part to chain transfer during polymerization [41]. In order to achieve a better controlled polymerization, we next tested (BDI-II)ZnN$(TMS)_2$. However, the (BDI-II)ZnN$(TMS)_2$/Cpt-mediated ROP resulted in only 61% Cpt incorporation, indicating inefficient formation of Cpt–Zn complexes during the initiation step (Table 7.4)

**TABLE 7.2 Formulation of NCs with Hydroxyl-Containing Therapeutic and Dye Molecules[a]**

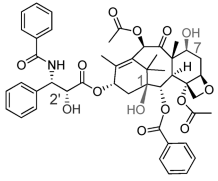

**Paclitaxel (Ptxl)**
Mitotic inhibitor

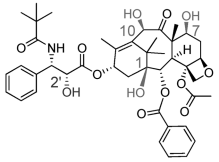

**Docetaxel (Dtxl)**
Mitotic inhibitor

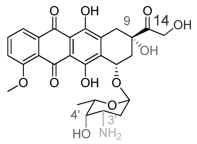

**Doxorubicin (Doxo)**
Intercalating DNA, topoisomerase II inhibitor

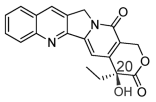

**Camptothecin (Cpt)**
Topoisomerase I inhibitor

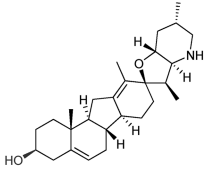

**Cyclopamine (Cpa)**
Hedgehog pathway inhibitor

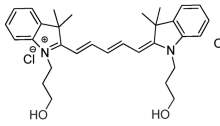

**Cy5-OH (Cy5)**
Fluorescent molecule

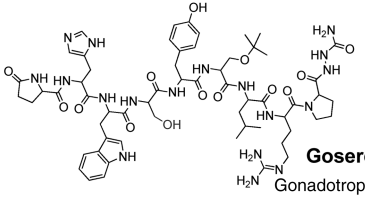

**Goserelin (Gos)**
Gonadotropin releasing hormone agonist

| Drug | [LA]/[Drug] | NCs | Loading (wt%) | LA Conversion (%) | Loading Efficiency (%) | Size (nm) | PDI | Reference |
|---|---|---|---|---|---|---|---|---|
| Cpt | 10 | Cpt–$LA_{10}$ | 19.5 | >99 | >99 | $72.5 \pm 0.7$ | $0.06 \pm 0.02$ | [54] |
| Ptxl | 25 | Ptxl–$LA_{25}$ | 19.2 | >99 | 97 | $55.6 \pm 0.5$ | $0.04 \pm 0.01$ | [51] |
| Ptxl | 15 | Ptxl–$LA_{15}$ | 28.3 | >99 | 95 | $85.5 \pm 1.4$ | $0.09 \pm 0.03$ | [51] |
| Dtxl | 10 | Dtxl–$LA_{10}$ | 35.9 | >99 | 97 | $77.9 \pm 1.5$ | $0.06 \pm 0.02$ | [51] |
| Doxo | 25 | Doxo–$LA_{25}$ | 13.1 | >99 | 98 | $90.8 \pm 0.9$ | $0.09 \pm 0.01$ | [52] |
| Doxo | 10 | Doxo–$LA_{10}$ | 27.4 | >99 | 94 | $125.2 \pm 2.3$ | $0.11 \pm 0.01$ | [52] |
| Cpa | 50 | Cpa–$LA_{50}$ | 5.4 | >99 | >99 | $78.0 \pm 2.4$ | $0.12 \pm 0.01$ | [52] |
| Cy5 | 25 | Cy5–$LA_{25}$ | 12.3 | >99 | >99 | $76.3 \pm 3.8$ | $0.06 \pm 0.01$ | [52] |
| Gos | 10 | Cy5–$LA_{10}$ | 46.8 | >99 | >99 | $120.6 \pm 2.7$ | $0.01 \pm 0.01$ | [52] |

LA, lactide; NCs, nanoconjugates; PDI, polydispersity.

[a]In molecule structures, the hydroxyl groups are differentiated in dark gray (regioselective conjugated to NCs) and in light gray. The functional group competing with hydroxyl groups (i.e., amine) is labeled in lightest gray.

**TABLE 7.3 Ptxl (or Dtxl)/(BDI-II)ZnN(TMS)$_2$ -Mediated ROP of LA, VL, CL, and TMC**

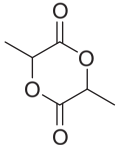


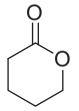


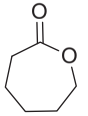


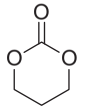


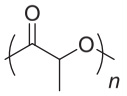


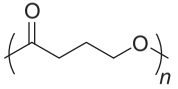


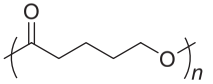


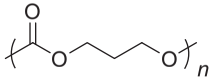


| Initiator (R) | Monomer | [M]/[R] | Time | Temperature (°C) | Conversion (%) | Incorporation efficiency (%) | $M_n/M_{exp}$ ($\times 10^3$ g/mol) | MWD ($M_w/M_n$) |
|---|---|---|---|---|---|---|---|---|
| Ptxl | LA | 200 | 12 | r.t. | >99 | >99 | 29.7/28.1 | 1.02 |
| Ptxl | VL | 100 | 12 | r.t. | >99 | >99 | 15.1/10.8 | 1.17 |
| Ptxl | CL | 200 | 10 | r.t. | >99 | >99 | 20.3/23.7 | 1.07 |
| Ptxl | TMC | 100 | 6 | 50 | >99 | >99 | 14.7/11.1 | 1.10 |
| Dtxl | LA | 200 | 12 | r.t. | >99 | >99 | 29.6/25.2 | 1.03 |
| Dtxl | CL | 100 | 10 | r.t. | >99 | >99 | 11.2/12.2 | 1.05 |
| Dtxl | TMC | 200 | 6 | 50 | >99 | >99 | 25.9/21.3 | 1.19 |
| Dtxl | VL | 100 | 12 | r.t. | >99 | >99 | 14.2/10.7 | 1.15 |

M, monomer; $M_{exp}$, expected molecular weight; MWD, molecular weight distribution; r.t., room temperature.

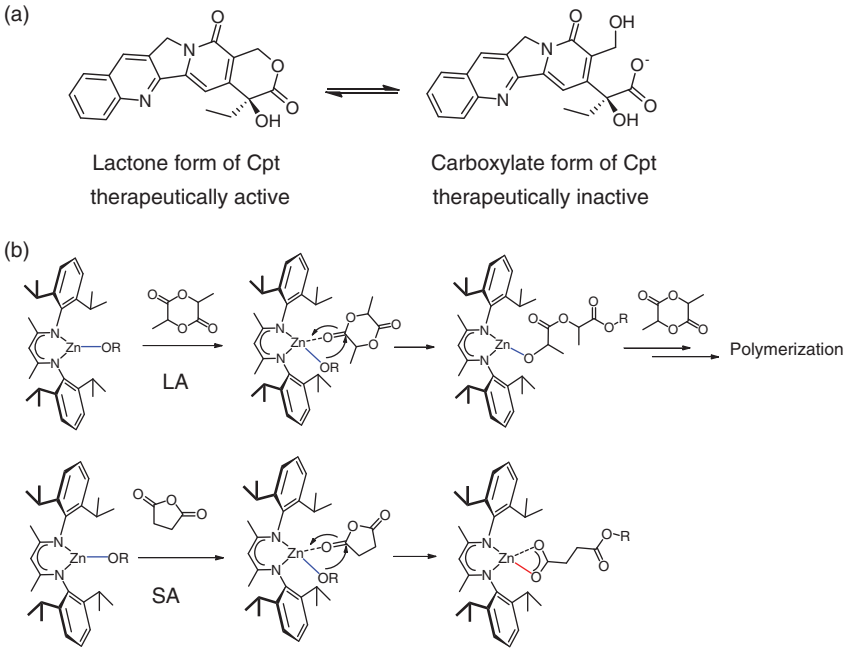


**FIGURE 7.4** (a) Equilibrium of Cpt Lactone and carboxylate forms. (b) Suggested insertion-coordination mechanism of (BDI)Zn-OR mediated ring-opening of lactide (LA) and succinic anhydride (SA). R group represents the PLA polymer chain or agents containing hydroxyl group(s) (e.g., Cpt).

[54]. The insufficient activation of the C20-OH of Cpt by (BDI-X)Zn was addressed by subtly tuning the 2- and 6-substituents of the *N*-aryl groups on the catalyst. Doing so, 100% incorporation efficiencies were observed in both (BDI-EE)ZnN(TMS)$_2$/Cpt- and (BDI-EI)ZnN(TMS)$_2$/Cpt-mediated polymerizations (Table 7.4) [54].

**TABLE 7.4 Ring Opening Reaction of SA and Polymerization of LA, Mediated by Cpt– (BDI-X)MN(TMS)$_2$ (M=Mg or Zn)[a]**

| Catalyst | Cpt loading efficiency (%) | Polymer MWD | Cpt–SA (%) | Cpt–Carboxylate[a] |
|---|---|---|---|---|
| (BDI-II)MgN(TMS)$_2$ | >99 | >1.3 | 78 | Yes |
| (BDI-II)ZnN(TMS)$_2$ | 61 | <1.1 | 19 | Yes |
| (BDI-EE)ZnN(TMS)$_2$ | >99 | >1.2 | N.D. | N.D. |
| (BDI-EI)ZnN(TMS)$_2$ | >99 | ~1.1 | 89 | No |

MWD, molecular weight distribution.
[a]Determined by HPLC analysis of Cpt–SA reaction mediated by different metal catalysts. The Cpt–carboxylate form indicated that the catalyst might have deleterious effect toward Cpt.

In addition to concerns regarding the ROP of LA, we also examined the retention of the lactone ring of Cpt throughout polymerization. As mentioned earlier, the lactone ring of Cpt is unstable and subject to ring opening in the presence of a nucleophile. To ensure that the therapeutically active form of Cpt is released in physiological conditions, the lactone ring of the drug must be maintained throughout the conjugation reaction. As LA is subject to rapid polymerization and the resulting Cpt–PLA conjugate is difficult to be characterized precisely, we used succinic anhydride (SA) as a model monomer to study (BDI-X)ZnN$(TMS)_2$/Cpt-mediated initiation. The ring opening of SA follows the same coordination–insertion mechanism as the initiation step of LA ROP but does not involve the subsequent chain propagation. Thus, the resulting product, Cpt–succinic acid (Cpt–SA), is a small molecule instead of a polymer. As a small molecule, the structure of Cpt–SA can be determined by routine characterization methods (Figure 7.4b). Experiments with SA revealed that Cpt activation by (BDI-II)MgN$(TMS)_2$ resulted in Cpt lactone ring opening (Table 7.4) [54]. The catalyst (BDI-EI)ZnN$(TMS)_2$, however, was able to prevent Cpt carboxylate formation. Using this catalyst, controlled polymerizations were observed over a broad range of LA/Cpt ratios from 75 to 400 in excellent agreement with the expected MWs and with narrow MWDs (1.02–1.18) [54].

With enhanced polymerization chemistry, Cpt–PLA NCs have improved formulation properties. Because both monomer conversion and drug incorporation are quantitative in Cpt–PLA synthesis, drug loadings can be predetermined by adjusting LA/Cpt feeding ratios. At a low monomer/initiator (M/I) ratio of 10, the drug loading can be as high as 19.5% (Cpt–$LA_{10}$, Table 7.2). To our knowledge, this Cpt–PLA NC has one of the highest loadings of Cpt ever reported [54]. Even at this high drug loading, sustained release of Cpt from Cpt–$LA_{10}$ NC was observed through the hydrolysis of the ester linker that connects the Cpt and the PLA without any observed burst release. Furthermore, the released Cpt (in PBS) had an HPLC elution time identical to authentic Cpt and was confirmed to have an identical molecular structure after isolation and characterization by $^1$H NMR [54]. In contrast, PLA/Cpt NPs prepared by co-precipitation have been previously reported to give low drug loading (0.1–1.5%), low loading efficiency (2.8–38.3%), and poorly controlled release kinetics (20–90% of the encapsulated Cpt released within 1 h in PBS) [61]. Our unique conjugation technique allows for formation of Cpt-containing PLA NCs with superbly controlled formulation parameters, thus making Cpt–PLA NCs potentially useful agents for sustained treatment of cancer *in vivo*.

### 7.3.3 Synthesis of Ptxl-PLA (Dtxl-PLA) Nanoconjugates

Initially isolated from the bark of the yew tree, Ptxl is a potent anticancer drug that was originally used as a treatment for ovarian and breast cancer. Its use has since expanded and the drug is now also used to treat lung, liver, and other types of cancer. Clinical application of Ptxl is often accompanied with severe, undesirable side effects partially due to the solvent Cremophor EL in the commercial formulation of Taxol$^{TM}$. To reduce the side effects, various nanoparticulate delivery vehicles have been developed and investigated in the past few decades [62–65], including

the FDA-approved albumin-bound Ptxl NPs (Abraxane™). Ptxl's analog, Dtxl (Taxotere™), is currently one of the best-selling chemotherapeutic agents and is used in the treatment of breast, ovarian, prostate, and nonsmall cell lung cancer. Many current NP formulations of both taxanes typically have low drug loadings, uncontrolled encapsulation efficiencies, and significant drug burst release effects when used *in vivo* [10,13,66,67]. Adding to formulation troubles, the densely functionalized structures of both molecules make it difficult to produce homogeneous Ptxl– and Dtxl–polymer conjugates.

Ptxl has three hydroxyl groups at its C2′, C1, and C7 positions which can initiate LA polymerization, resulting in Ptxl–PLA conjugates with 1–3 PLA chains attached (Figure 7.5a). The three hydroxyl groups differ in steric hindrance in the order of 2′-OH < 7-OH < 1-OH. The tertiary 1-OH is least accessible and is typically inactive [68]. The 7-OH, however, can potentially compete with 2′-OH [69] for coordination with metal catalysts. In order to differentiate the initiation site between the two hydroxyl groups, we postulated that a metal catalyst with a bulky chelating ligand would preferentially form Ptxl–metal complexes through the 2′-OH for site-specific LA polymerization.

A selective reduction reaction was used to divide Ptxl into two fragments—one with the 2′-OH and the other with the 1- and 7-OHs—in order to study the impact of catalyst on regioselectivity [70]. Tetrabutylammonium borohydride ($Bu_4NBH_4$) can selectively and quantitatively reduce the C13-ester bond of Ptxl to produce baccatin III (BAC) and (1*S*,2*R*)-*N*-1-(1-phenyl-2,3-dihydroxypropyl)benzamide (PDB; Figure 7.5a) [70]. Thus, Ptxl–$LA_5$ was generated using a variety of catalysts, reduced with $Bu_4NBH_4$ and examined using HPLC–MS analysis. Doing so, we found that Mg $(N(TMS)_2)_2$—a catalyst without a chelating ligand—initiates polymerization nonpreferentially at both the 2′-OH and the 7-OH. Its counterpart with a bulky BDI-II ligand (e.g., (BDI-II)$MgN(TMS)_2$) preferentially initiates polymerization at the 2′-OH position. However, the resulting Ptxl–PLAs displayed fairly broad MWD (>1.2). We reasoned that the observation was attributable to fast propagation relative to initiation for the Mg catalysts [41]. Thus, to reduce the MWD of polymers, we utilized a zinc analog, (BDI-II)$ZnN(TMS)_2$, to give a more controlled LA polymerization with a narrow MWD (1.02, Figure 7.5b) [41].

We further examined the effect of the ligand on initiation regioselectivity and LA polymerization by (BDI-X)$ZnN(TMS)_2$/Ptxl by varying the steric bulk of the *N*-aryl substituents ($R_1$ and $R_2$) and the electronic properties of $R_3$ (Table 7.5) [41,53,71,72]. While both PDB–PLA and BAC–PLA are observed for Ptxl–$LA_5$ initiated by (BDI-EE)$ZnN(TMS)_2$/Ptxl, analysis of the fragments of the reductive cleavage of (BDI-EI) $ZnN(TMS)_2$/Ptxl showed relatively reduced amounts of BAC–PLA. This indicates that a Zn catalyst with bulky *N*-aryl substituents will preferentially coordinate with the 2′-OH of Ptxl to initiate PLA polymerization. Of note, various studies suggest that the polymerization process does not lead to deleterious effect on Ptxl—the Ptxl can be loaded and released with intact structure.

SA was again used as a model monomer to study the initiation step of ROP. Such reactions yield a small molecule, Ptxl–succinic acid (Ptxl–SA), whose structure can be determined easily by routine characterization methods (Table 7.5). The results

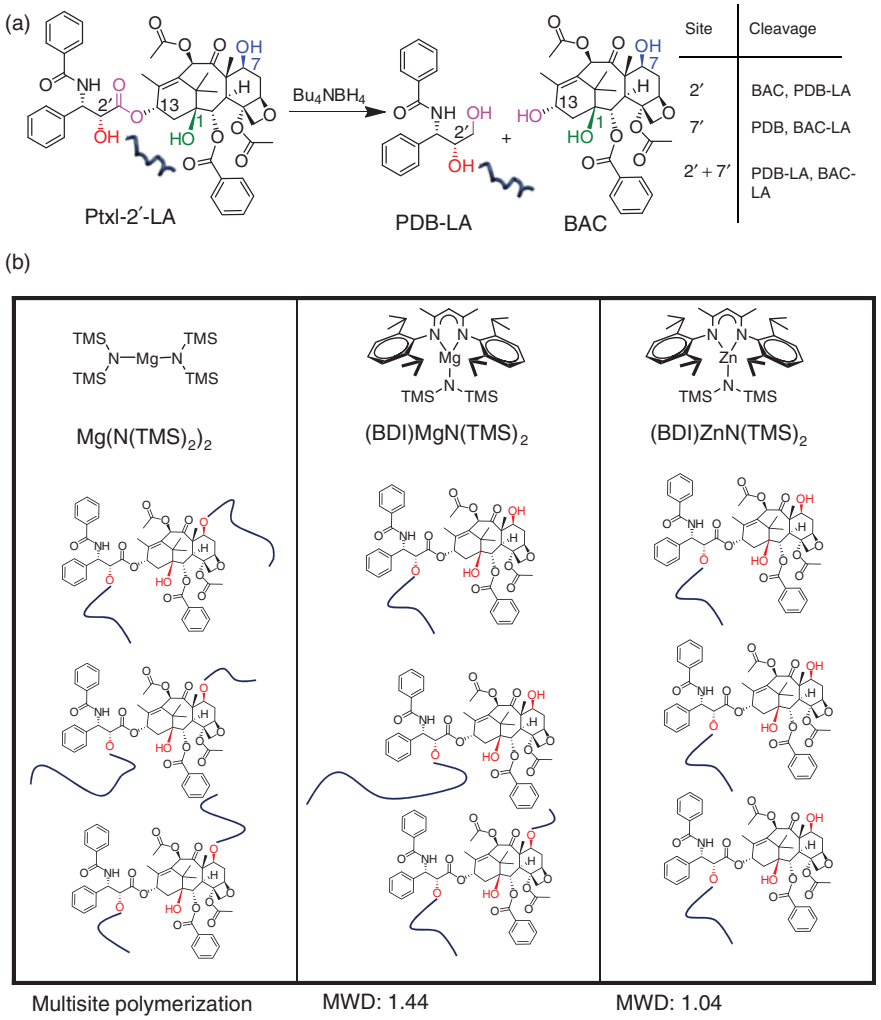


**FIGURE 7.5** (a) $Bu_4NBH_4$-induced site-specific degradation of Ptxl for the formation of PDB and baccatin (BAC). (b) Scheme of Ptxl–PLA conjugates mediated by different catalysts, with the indication of regioselectivity and molecular weight distributions (MWDs).

show that the regioselectivity of the Ptxl/SA reaction increases as the sizes of $R_1$ and $R_2$ increase, in good agreement with the reductive reaction study using $Bu_4NBH_4$. The catalyst (BDI-II)ZnN(TMS)$_2$ showed the best regioselectivity while (BDI-EE) ZnN(TMS)$_2$ showed the worst. Changing $R_3$ from –H (BDI-II) to the electron-withdrawing –CN group (BDI-IICN) did not change the regioselectivity of the Ptxl/SA reaction. However, the addition of the CN group enhanced the reactivity of the resulting catalyst ((BDI-IICN)ZnN(TMS)$_2$), leading to a higher yield of Ptxl–2′-SA (Table 7.5). Controlled polymerization of PLA was observed when (BDI-II)ZnN

**TABLE 7.5 Regioselectivity of the Ptxl/Metal Catalyst-Mediated Ring Opening Reactions**

| Catalyst | Reduction of Ptxl–$LA_5$[a] | Ptxl–2′-SA (%) | Regioselectivity[a] | Regioselectivity for 2′-OH[b] | Polymer MWD |
|---|---|---|---|---|---|
| $Mg[N(TMS)_2]_2$ | PDB–LA, BAC–LA | N.D. | 2′-OH and 7-OH | N.D. | >1.4 |
| (BDI-II)$MgN(TMS)_2$ | Majorly PDB–LA, BAC | N.D. | Majorly 2′-OH | N.D. | >1.4 |
| (BDI-II)$ZnN(TMS)_2$ | PDB–LA, BAC | 40 | 2′-OH | 100 | <1.1 |
| (BDI-EE)$ZnN(TMS)_2$ | PDB–LA, BAC–LA | 30 | 2′-OH and 7-OH | 35 | 1.3 |
| (BDI-EI)$ZnN(TMS)_2$ | Majorly PDB–LA, BAC | 45 | Majorly 2′-OH | 77 | ~1.2 |
| (BDI-IICN)$ZnN(TMS)_2$ | PDB–LA, BAC | 54 | 2′-OH | 100 | <1.1 |

[a]Determined by HPLC analysis of Ptxl–$LA_5$ site-specific reduced by $Bu_4NBH_4$.
[b]Determined by HPLC analysis of Ptxl–SA ring opening reactions mediated by (BDI-X)Zn catalyst.

$(TMS)_2$/Ptxl was used at various [LA]/[Ptxl] ratios (50/1–300/1), with the obtained MWs in excellent agreement with the expected MWs and the monomodal MWDs in the range of 1.02–1.09.

Dtxl, the Ptxl analog, has four hydroxyl groups at C2′, C1, C7, and C10 (Table 7.2). To examine the regioselectivity of Dtxl-initiated ROP of LA, we examined the Dtxl/(BDI-X)ZnN$(TMS)_2$-mediated ring opening reaction of SA. When complexed with (BDI-II)ZnN$(TMS)_2$, Dtxl reacted with SA to yield Dtxl–2′-SA with 100% regioselectivity and 71.5% yield. (BDI-II)ZnN$(TMS)_2$/Dtxl also showed excellent control for the ROP of LA, affording Dtxl–PLA with predictable MWs and very narrow MWDs. NCs formed with either Ptxl–PLA or Dtxl–PLA showed unprecedented high loading—up to 28.3 wt% in Ptxl–LA NCs and up to 35.9 wt% in Dtxl–LA NCs. The detailed formulation methods and efficacy of both conjugates are further discussed in Section 7.4.

Poly(δ-valerolactone) (PVL), poly(trimethylene carbonate) (PTMC), and poly (ε-caprolactone) (PCL) have been used extensively as alternatives to PLA in suturing, drug delivery, and tissue engineering applications. We were therefore interested as to whether the Ptxl (or Dtxl)/(BDI-II)ZnN$(TMS)_2$-mediated ROP of LA could be extended to the synthesis of Ptxl–PCL, Ptxl–PVL, and Ptxl–PTMC conjugates. Ptxl/(BDI-II)ZnN$(TMS)_2$ showed excellent control over the polymerization of δ-valerolactone (VL), trimethylene carbonate (TMC), and ε-caprolactone (CL). All the polymerization reactions gave drug–polymer conjugates with the expected MWs and narrow MWDs ($M_w/M_n < 1.2$) (Table 7.3). Furthermore, the polymerizations of VL and CL proceeded at room temperature, and the monomer conversions were quantitative. The polymerization of TMC, however, required a slightly elevated reaction temperature (6 h at 50°C).

### 7.3.4 Synthesis of Doxo–PLA Nanoconjugates

Doxo is commonly used in the treatment of a wide range of cancers and leukemia. Clinically, Doxo is administered as Doxil™, a PEGylated liposome-encapsulated form of the drug [74,75]. Owing to its high hydrophilicity, the encapsulation of Doxo within a hydrophobic polymer matrix can be challenging. For micelles formed by co-precipitating Doxo and PLGA–PEG, reported values indicate Doxo loading and loading efficiency as low as 0.51% and 23%, respectively [76]. Doxo loadings of 0.6–8.7% with loading efficiencies of 11.4–43.6% have also been reported by Hubbell and coworkers in their encapsulation studies using inverse emulsion polymerization of nonprotonated Doxo [77]. However, even with techniques allowing increased loading, burst release of Doxo is reported in many NP delivery systems [40,76,77].

Conjugation of Doxo to polymers can be difficult due to the diverse functional groups presented by the drug. Doxo has three hydroxyl groups, two phenolic hydroxyls, one ketone group, and one amine group. Doxo is also sensitive to pH, heat, metal ions, and light, further complicating its conjugation chemistry [78]. One conjugation strategy is to couple the terminal carboxylate of PLA with Doxo by creating an amide linkage through the 3′-$NH_2$ of Doxo [39,79]. Such Doxo–PLA conjugates, however, cannot release Doxo in its original form by hydrolysis.

Instead, Doxo-3′-lactamide, a prodrug of Doxo, is formed, which does not easily degrade *in vivo* due to the stability of amide bond [80]. Other efforts have been devoted to conjugation between the C13-ketone group of Doxo and hydrazine groups of polymeric carriers by forming an acid-labile hydrazone bond [22,81–83]. However, clinical studies of immunoconjugates with Doxo connected to monoclonal antibodies via hydrazone linkers gave unsatisfactory antitumor effects, ultimately leading to the termination of the clinical development of such immunoconjugates [84].

In previous studies involving the use of metal catalysts for LA polymerization, ROP of LA proceeded predominately by metal–alkoxides (M–ORs) rather than by metal-amides (M-NHRs) [47]. M–OR complexes typically have higher activities for LA ring opening than their amine analogs. For instance, Coates and coworkers reported that (BDI)ZnOCH$(CH_3)_2$ initiated and completed an LA polymerization within 20 min at a M/I ratio of 200 while a similar polymerization mediated by (BDI) ZnN$(TMS)_2$ required 10 h to complete [41]. The chemoselectivity of -OH instead of -$NH_2$ for Zn catalysts has also been confirmed by another molecular pair, 1-pyrenemethanol (Pyr-OH) and 1-pyrenemethylamine (Pyr-$NH_2$) [52].

Doxo has three hydroxyl groups, one at each of its C4′, C9, and C14 positions. Theoretically, LA polymerization can be initiated by any or all of these hydroxyl groups. C9-OH is the most sterically hindered and thus unlikely to initiate polymerization. To evaluate the initiation regioselectivity, we mixed Doxo with the catalyst (BDI-II)ZnN$(TMS)_2$ and succinic anhydride (SA) to mimic the initiation step of LA polymerization. ESI–MS results coupled with NMR analysis revealed that the SA ring was opened by the C14-OH of Doxo rather than by the C4′-OH or C3′-$NH_2$ of Doxo. When (BDI-II)ZnN$(TMS)_2$ was replaced by Zn$(N(TMS)_2)_2$, a Zn catalyst without ligands, the initiation regioselectivity completely disappeared, with Doxo–4′, 14-bissuccinic ester (Doxo–2SA), and Doxo–4′, 9, 14-trisuccinic ester (Doxo–-3SA) being the predominant products (Figure 7.6). Interestingly, the metal activity also had a profound effect on regioselectivity. When the more highly active (BDI-II) MgN$(TMS)_2$ catalyst was used in a similar reaction, the most prevalent product was Doxo–2SA. Thus, by rationally designing ROP metal catalysts, Doxo–PLA conjugates with highly controlled regio- and chemoselectivity are achievable within one step without having to protect the C3′-$NH_2$ and other competing hydroxyl groups of Doxo [52].

Having completed preliminary investigations with SA, we next studied Doxo-initiated LA polymerization. Doxo/(BDI-II)ZnN$(TMS)_2$-mediated LA polymerization resulted in Doxo–PLA with narrow MWD ($<1.2$) and the expected MWs. At a low M/I ratio of 10, the drug loading was as high as 27.4% (Doxo–$LA_{10}$). To our knowledge, this is by far the highest loading ever reported in Doxo-containing polymeric NPs [52]. HPLC analysis revealed that Doxo–$LA_{10}$ conjugates incubated in PBS at 30°C released Doxo in its original, therapeutically active form. In a separate experiment, Doxo–$LA_{100}$ treated with 0.1 M NaOH was shown to release 88–92% of the Doxo in its original form. In this case, the incomplete recovery is likely due to instability of Doxo in NaOH. Combined, these studies suggest that Doxo molecules were linked to PLA through its hydroxyl group(s) by forming ester linker(s) with PLA, which could be hydrolyzed in an alkaline condition. Even at high

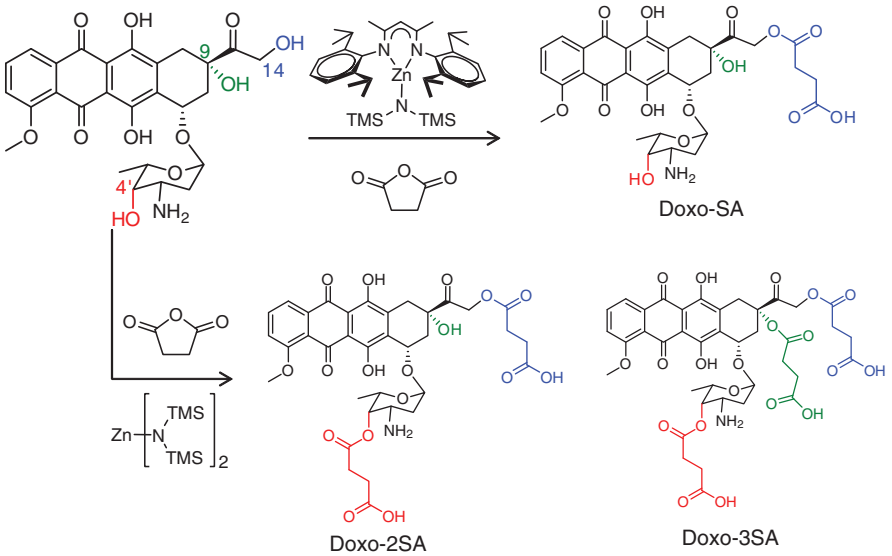


**FIGURE 7.6** Scheme of Doxo–SA reaction (1:3 molar ratio) mediated by different Zn catalysts.

drug loading (27.4%), sustained release of Doxo from Doxo–$LA_{10}$ NC was observed without burst release. This is in sharp contrast to the burst release of PLA/Doxo NPs prepared by co-precipitation, in which 90% of the Doxo is released within 3 h [52]. The sustained release of Doxo from NCs with high loading may help alleviate the systemic side effects of Doxil$^{TM}$ without reducing the overall dosage.

## 7.4 NANOCONJUGATES: FORMULATION AND POTENTIAL APPLICATION

### 7.4.1 Formulation

Particle size is an important aspect of drug delivery. For cancer therapeutics, formulated particles should have a uniform, monomodal distribution and a diameter <200 nm to take advantage of the enhanced permeability and retention (EPR) effect. As such, particles formed by nanoprecipitation of Ptxl–PLA were evaluated for their size. In general, Ptxl–PLA yielded particles with monomodal distributions while their nanoencapsulate (NEs, see Section 7.2.1) counterparts were polydisperse (Figure 7.7a) [13,85]. As the multimodal distribution of NEs is due in part to the aggregation of nonencapsulated free drug [13], the monomodal distribution observed with NCs is likely related to the unimolecular structure of Ptxl–PLA conjugates. The actual size of Ptxl–PLA NPs prepared by nanoprecipitation can be manipulated by changing the solvent as well as polymer concentration. At a fixed Ptxl–PLA concentration, the size of NCs prepared by precipitating a DMF solution of conjugate

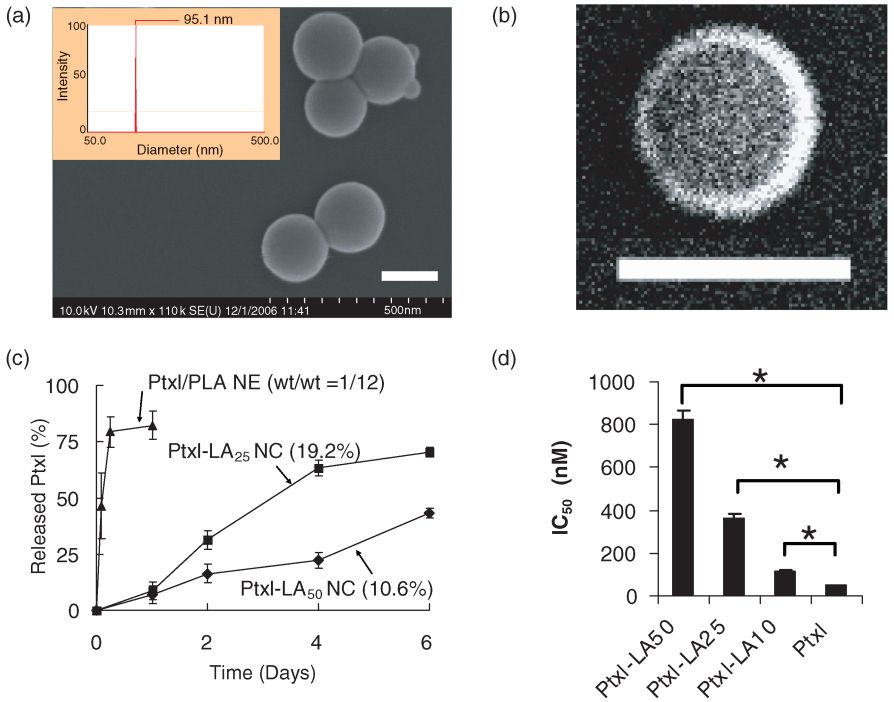


**FIGURE 7.7** (a) Ptxl–$LA_{25}$ NCs were analyzed by dynamic light scattering (DLS) and scanning electron microscopy (SEM). (b) SEM image of Ptxl–$LA_{25}$/PLGA–mPEG NCs (Scale bar = 150 nm). The different density of materials were indicated with different gray scale: PEG layer (white color) and hydrophobic PLA part (gray color) (c) Release kinetics of Ptxl from Ptxl–PLA NCs and Ptxl–PLA NE (Prepared by nanoprecipitation of a mixture of Ptxl and PLA (Ptxl–PLA (wt/wt) = 1/12) at 37°C in PBS. (d) $IC_{50}$ values determined by MTT assay of Ptxl–PLA NCs and free Ptxl, which were incubated with PC-3 cells for 24 h. (*Source*: Figures reproduced with permission from Reference [51].)

is typically in a range 60–100 nm, 20–30 nm smaller than those prepared with acetone or THF as a solvent [13]. When nanoprecipitation is carried out using a 1:20 (v:v) mixture of DMF and water, the size of Ptxl–LA NCs shows a linear correlation with the conjugate concentration and can be used to precisely form particles with diameters from 60 to 100 nm.

Surface modification of NPs with PEG is widely used for prolonged systemic circulation and reduced aggregation of NPs in blood [86]. To avoid the removal of unreacted reagents and by-products, poly(lactide)-*b*-methoxylated PEG (PLA–mPEG)—an amphiphilic copolymer that has a 14 kDa PLA and a 5 kDa PEG segment [87]—was used to PEGylate NCs instead of covalently conjugating PEG to NCs [65,85]. The PEGylated structure of the NCs can be seen by scanning transmission electron microscopy (Figure 7.7b). Ptxl–LA NCs with a PEGylated surface have a negative surface zeta potential and thus remain nonaggregated in water and PBS solution by means of surface charge repulsion. In addition, we also

developed a one-step co-precipitation method to formulate NCs that could stay nonaggregated in a salt solution with an amphiphilic triblock copolymer PLA–PEG–PLA, a feature that diblock PLA–PEG did not provide [55].

Drug burst release is a long-standing formulation challenge to NEs and often results in undesirable side effects and reduced therapeutic efficacy [67]. Conventional NEs typically show burst release of 60–90% of their payload within a few to tens of hours [9]. Since the Ptxl release kinetics of Ptxl–PLA NCs is determined not simply by diffusion—as is the case with NEs—but by the hydrolysis of the Ptxl–PLA ester linker followed by diffusion out of NCs, the release kinetics of Ptxl from NCs are more controlled and show significantly reduced burst release (Figure 7.7c). For example, Ptxl released from Ptxl–$LA_{25}$ NCs (19.2 wt%) was 8.7% at Day 1 and 70.4% at Day 6. In comparison, 89% of Ptxl was released within 24 h from Ptxl/PLA NE (Figure 7.7c). Release of Ptxl from Ptxl–$LA_{50}$ NCs was slower than from Ptxl–$LA_{25}$ NCs, presumably because of the higher MW of Ptxl–$LA_{50}$ and more compact particle structure. With its controlled drug release, the *in vitro* toxicity of Ptxl–LA NCs correlates with the amount of Ptxl released (Figure 7.7d). For example, Ptxl–$LA_{15}$ NCs have a nearly identical $IC_{50}$ to free Ptxl (87 nM) while the $IC_{50}$ of Ptxl–$LA_{50}$ NCs is an order of magnitude higher. As a result, the toxicity of NCs can be tuned in a wide range simply by controlling NC drug loading.

Aptamers are either single-stranded DNA or RNA that specifically binds to target ligands [88,89]. When used for cancer targeting, aptamers are capable of binding to target antigens with extremely high affinity and specificity in a manner resembling antibody-mediated cancer targeting [90,91]. Unlike antibodies, synthesis of aptamers is an entirely chemical process and thus shows negligible batch-to-batch inconsistency [92,93]. Moreover, aptamers are typically nonimmunogenic and exhibit remarkable stability against pH, temperature, and solvent. The A10 aptamer with 2′-fluoro-modified ribose on all pyrimidines and a 3′-inverted deoxythymidine cap has been identified and utilized to target extracellular prostate-specific membrane antigen (PSMA) [94]. A10 binds to PSMA-positive LNCaP prostate cancer cells but not PSMA-negative PC-3 prostate cancer cells. To demonstrate NC targeting, amine-terminated A10 aptamer was conjugated to PLA–PEG–COOH/Cy5–PLA NCs through carboxylic acid–amine coupling in the presence of EDC and NHS to give aptamer/PLA–PEG–COOH/Cy5–PLA NCs (aptamer–Cy5 NC) [90]. Incubation of aptamer–Cy5 NCs with LNCaP cells for 6 h resulted in substantially increased NC internalization (Figure 7.8) compared to PC-3 cells. These *in vitro* studies demonstrated that NCs conjugated with aptamer-targeting ligand can potentially be used for prostate cancer targeting.

Small-scale NPs that stay nonaggregated in PBS for *in vitro* or *in vivo* laboratory studies are straightforward. However, in order to facilitate clinical translation, NPs need to be prepared in large quantity with well-controlled properties that remain unchanged during the processes of manufacturing, storage, and transport. Because Ptxl is covalently conjugated to PLA through an ester bond that is subject to hydrolysis upon exposure to water, handling of Ptxl–PLA NCs in aqueous solution is undesirable. Thus, NCs have to be formulated in solid form for clinical use. By screening various molecules, we discovered that albumin functions as an excellent

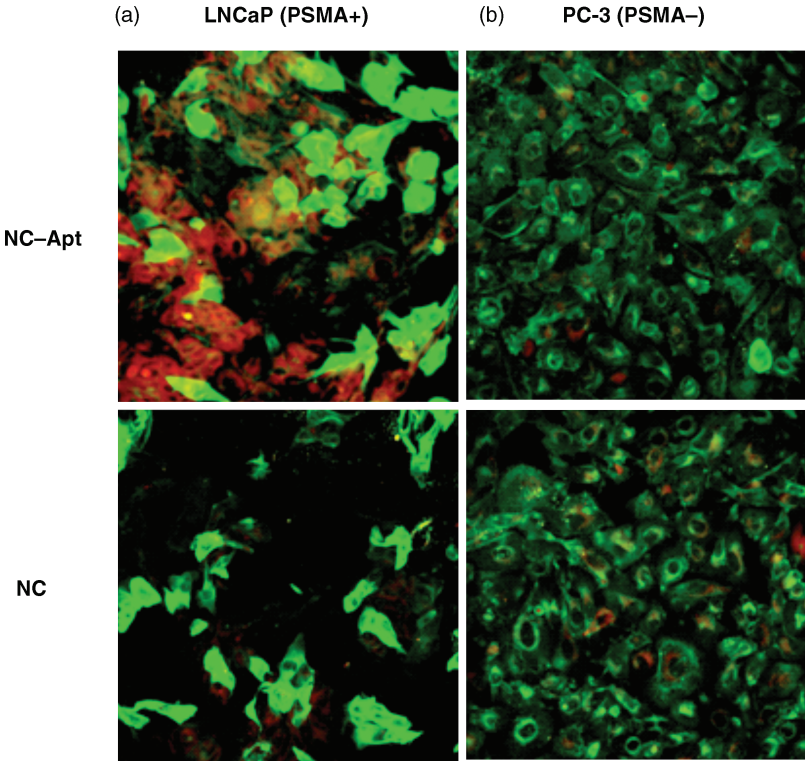


**FIGURE 7.8** Confocal images of LNCaP (a) and PC-3 (b) cells treated with aptamer-functionalized nanoconjugates (NC–Apt, top) and nanoparticles without aptamer (NC, bottom).The Cy-5 incorporated NC or NC–Apt are shown in red. The cells counter stained with Alexa-Flour 488 Phalloidin (binding to cellular actin) are shown in green.

lyoprotectant for Ptxl–PLA and yields solid NCs that do not aggregate when reconstituted in PBS. By incorporating the albumin-based lyoprotection technique, we demonstrated for the first time that polymer nanoparticles containing a conjugated nucleic acid targeting ligand can be prepared in solid form and still be reconstituted to well-dispersed, nonaggregated particles with functional targeting capability [55].

### 7.4.2 Theranostic Nanoconjugates

There is growing interest in developing noninvasive, whole-body fluorescent imaging techniques to assess the biodistribution of drug delivery systems or diagnostic agents within patients. To ensure effective measurement of fluorescent signal *in vivo*, it is crucial to use red or near-IR dyes. Quantum dots, a class of inorganic nanocrystals with excellent fluorescent intensity and photostability, can readily be prepared to have a far-red emission band. However, there is a general consensus

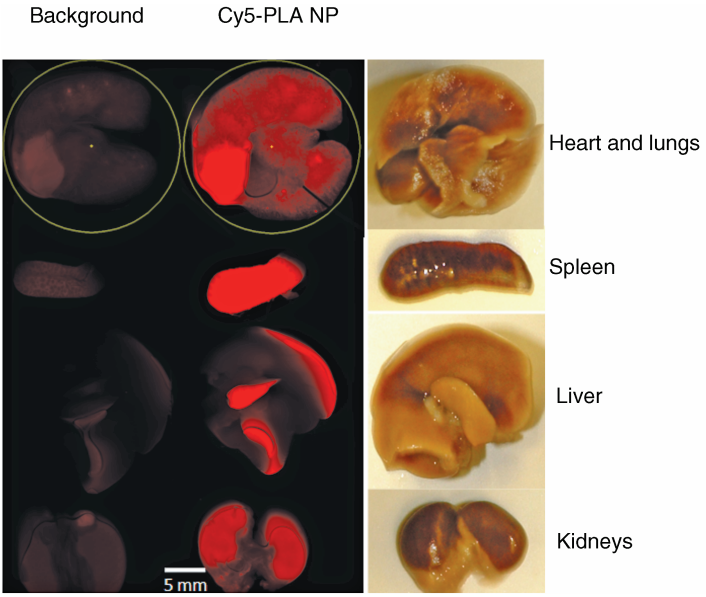


**FIGURE 7.9** Distribution of Cy5 dye-labeled PLA NCs in various visceral mice organs. Mice were sacrificed 24 h after i.v. injection of Cy5–PLA NCs.

that quantum dots cannot be used in humans because of their toxicity. Currently, small molecule organic dyes are more promising probes to be coupled with imaging systems for clinical applications as compared to quantum dots. Polymeric nanoparticles can function as good carriers for the delivery of small molecule imaging materials because they can provide prolonged systemic circulation and improved tumor accumulation compared to unformulated drugs. To meet this challenge, it is particularly important that NPs be formulated with stably incorporated fluorescent ligands and controlled formulation parameters (size, surface properties, etc.). Accordingly, our monomodal, narrowly distributed PEGylated Cy5–PLA NPs have the potential to be candidates for whole body *in vivo* imaging. Preliminary *in vivo* studies demonstrated that Cy5–PLA NCs can be easily visualized in various visceral organs with low autofluorescent background ratios (Figure 7.9). Owing to their sub-100 nm size, the Cy5–PLA NCs also can be used to study the lymphatic biodistribution of NPs when coupled with whole-body optical imaging [95]. These promising results support the further development of Cy5–PLA NCs as a model system to assess the *in vivo* pharmacological and pharmacokinetics profiles of NCs.

### 7.4.3 Nanoconjugates Against Other Diseases

Considering that molecular targeting of cell-based targets may be confounded by inter- or intra-patient heterogeneity in cell surface antigen expression, targeted NPs that can recognize the extracellular matrix have attracted considerable attention for

therapeutic/diagnostic delivery. We have recently engineered a peptide-conjugated NP to target the vascular basement membrane for the treatment of injured vasculature. The high affinity C11 peptide was screened from a combinatorial phage library of hepta-peptide ligands against human collagen IV, which represents 50% of the vascular basement membrane (Figure 7.10) [96]. Angioplasty-injured carotid artery

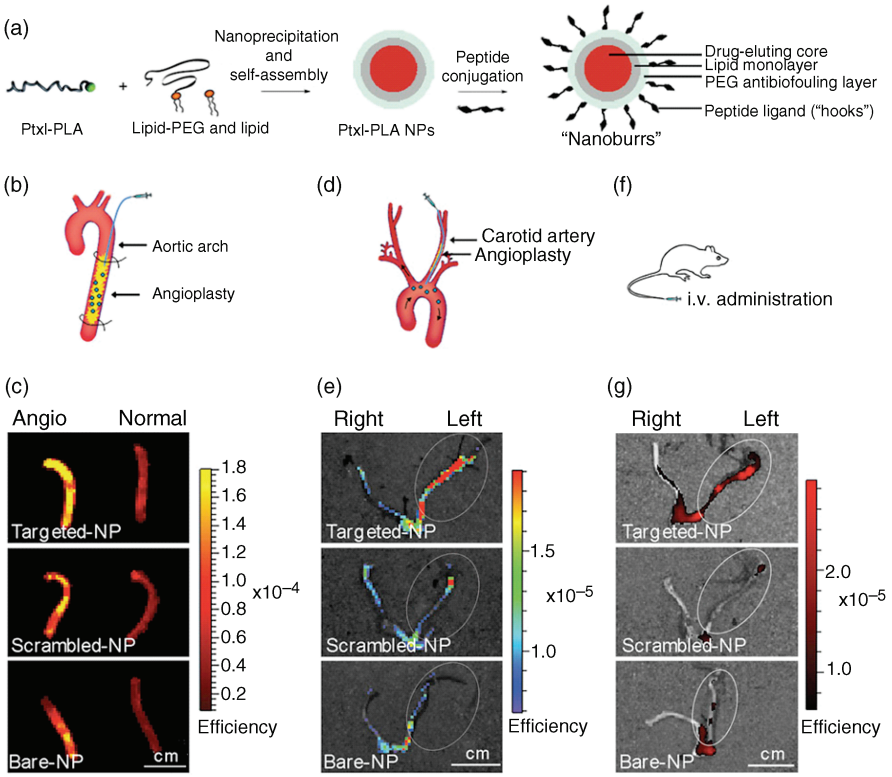


**FIGURE 7.10** (a) Schematic of NPs synthesis by nanoprecipitation and self-assembly of Ptxl–LA NCs with lipids and peptide ligands to adhere to the exposed basement membrane during vascular injury. (b) Scheme of *ex vivo* abdominal aorta injury model; samples were delivered into the aorta segment for 5 min. (c) Fluorescence images overlaid on photographs of balloon-injured aortas incubated with NCs with a targeting peptide, compared with scrambled-peptide and nontargeted NPs. (d) *In vivo* intra-aortic administration in a carotid injury model: a catheter was inserted via the external carotid into the common carotid and advanced into the aortic arch. (e) Fluorescence images overlayed on photographs of carotid arteries incubated with NCs with a targeting peptide, compared with scrambled-peptide and nontargeted NPs. (f) *In vivo* systemic administration in a carotid angioplasty model. (g) Fluorescence images overlayed on photographs of carotid arteries incubated with NCs with a targeting peptide, compared with scrambled-peptide and nontargeted NPs. For imaging, Alexa Fluor 647–PLGA dye conjugates were encapsulated in place of Ptxl–PLA drug conjugates. (Scale bar, 1 cm.)

was used as a model of compromised vasculature to examine the targeting capacity of the C11 peptide-conjugated polymeric NPs. The targeted NPs were delivered via both intra-arterial and i.v. administration and, when compared to nontargeted NPs, showed greater *in vivo* vascular retention at sites of injured vasculature in rats (Figure 7.10) [96]. Although the initial application was for vessel wall targeting in cardiovascular disease, the utility of this peptide-targeted NP system is broad and could be used to diagnose and treat different human diseases where the endothelial lining is compromised.

Immunosuppressive agents have played a pivotal role in ensuring the success of organ transplantation and greatly improved the outcomes of patients with life-threatening, immune-mediated diseases [97]. However, the use of immunosuppressive agents is hindered by the lack of selectivity as well as major adverse drug reactions. The immunosuppressive agent cyclosporine (CsA), for example, results in a dramatic improvement in short-term allograft survival but also poses a risk of chronic nephrotoxicity because of its narrow therapeutic window [98–100]. We recently developed CsA–PLA NCs to achieve targeted immunosuppression for a wide variety of immune-mediated disorders [101]. CsA–PLA NCs showed superior physicochemical properties with excellent size control, narrow size distribution, and very well controlled release kinetics without a noticeable burst release. CsA–PLA NCs also showed excellent stability in biological media with negligible aggregation. Because of their ability to mediate the sustained release of CsA *in vitro* while suppressing T-cell mediated immune responses, CsA–PLA NCs are an excellent system for immunosuppression of organ rejection. We also developed a novel strategy combining CsA–PLA NCs with dendritic cells (DCs) to efficiently deliver CsA to draining lymph nodes to inhibit T-cell priming in a locally controlled and sustained manner without systemic release. This innovative delivery strategy constitutes a strong basis for future targeted delivery of immunosuppressive drugs with improved efficiency and reduced toxicity.

## 7.5 CONCLUSIONS AND OUTLOOK

LA ROP-mediated drug conjugation allows for facile regio- and chemoselective incorporation of drugs onto PLA. This in turn allows for the formation of drug delivery vehicles with low polydispersity, predetermined drug loadings (up to ~30%), and quantitative loading efficiencies. The BDI–metal chelating complexes described earlier for Cpt, Ptxl, and Doxo conjugation do not have deleterious effects on drugs and can be easily removed by solvent extraction. Because both Zn and Mg ions are biocompatible and, in fact, are key elements in our dietary mineral supplements, there should be no significant safety concerns regarding the use of these two metal catalysts. Multigram scales of drug-PLA conjugates can be readily prepared within hours using the one-pot polymerization approaches described here. Because drug molecules are covalently conjugated to PLA, the post-reaction formulation process (precipitation, removal of catalysts, nanoprecipitation, sterilization, lyophilization, shipping and handling, etc.) has minimal impact on sample

property as compared to drug–polymer NPs prepared via encapsulation methods. This polymerization-mediated conjugation method may be utilized for the formulation of polymer–drug conjugates not only for drug delivery, but also for other controlled release applications (scaffolds, coatings of stents, etc.). Other cyclic esters (e.g., ε-caprolactone and δ-valerolactone) are likely to replace LA and find use as monomers in such drug-initiated polymerizations [102]. The drug/(BDI-X) $ZnN(TMS)_2$-initiated polymerization of these cyclic esters at room temperature has recently been achieved in our laboratory, which will provide further tunability of the release profiles and other physicochemical properties. Given that the lack of a controlled formulation for nanoparticulate drug delivery vehicles is one of the bottlenecks to their clinical development, this unique, ROP-mediated conjugation methodology may contribute to the development of clinically applicable nanomedicines.

Although there has been impressive progress in nanomedicine for cancer treatment in the past few years, enormous tasks remain. The convergence of seemingly disparate scientific fields (e.g., cancer biology, electronics, bioimaging, biomicroelectromechanical systems, computer science, polymer and materials chemistry, biophysics) will only accelerate the development of nanomedicine [103]. Our work will continue in the daunting task of pushing targeted nanoconjugates for clinical evaluation, with the expansion of our technology to other disease-related formulations (e.g., imaging agents or implantable devices). It also will be important to improve intratumoral penetration for enhanced efficacy [104,105]. Given the limitations of spatial and temporal changes in the expression of the target [106], the seemingly pedestrian issues on targeting ligand functionality are also likely to be challenging *in vivo* [107,108]. The prevention and treatment of metastasis will be of particular interest because metastasis is responsible for 90% of cancer deaths [109]. The coordination of different approaches will help optimize the delivery efficiency. A deep understanding of all aspects of the biology of cancers, including tumor microenvironment, will continue to be crucial for design and understanding [110]. We are far from being able to create delivery vehicles to meet the expectation of complete tumor eradication. However, as the field of nanomedicine matures, the objective may be on the horizon.

## Acknowledgments

Jianjun Cheng acknowledges support from the NSF (Career Award Program DMR-0748834) and the NIH (NIH Director's New Innovator Award 1DP2OD007246-01, 1R21EB009486A, and 1R21CA139329Z). Rong Tong acknowledges a student fellowship from the Siteman Center for Cancer Nanotechnology Excellence (University of Washington & University of Illinois at Urbana-Champaign) from 2007 to 2010. Li Tang was funded at University of Illinois at Urbana-Champaign from NIH National Cancer Institute Alliance for Nanotechnology in Cancer "Midwest Cancer Nanotechnology Training Center" Grant R25 CA154015A.

## REFERENCES


1. Langer, R. (1998). Drug delivery and targeting. *Nature 392*, 5.
2. Davis, M. E., Chen, Z., Shin, D. M. (2008). Nanoparticle therapeutics: an emerging treatment modality for cancer. *Nature Reviews Drug Discovery 7*, 771.
3. Petros, R. A., DeSimone, J. M. (2010). Strategies in the design of nanoparticles for therapeutic applications. *Nature Reviews Drug Discovery 9*, 615.
4. Wagner, V., Dullaart, A., Bock, A. K., Zweck, A. (2006). The emerging nanomedicine landscape. *Nature Biotechnology 24*, 1211.
5. Zhang, L., Gu, F. X., Chan, J. M., Wang, A. Z., Langer, R. S., Farokhzad, O. C. (2008). Nanoparticles in medicine: therapeutic applications and developments. *Clinical Pharmacology and Therapeutics 83*, 761.
6. Farokhzad, O. C., Langer, R. (2009). Impact of nanotechnology on drug delivery. *ACS Nano 3*, 16.
7. Cohen, S., Yoshioka, T., Lucarelli, M., Hwang, L. H., Langer, R. (1991). Controlled delivery systems for proteins based on poly(lactic glycolic acid) microspheres. *Pharmaceutical Research 8*, 713.
8. Ringsdorf, H. (1975). Structure and properties of pharmacologically active polymers. *Journal of Polymer Science Part C-Polymer Symposium 51*, 135.
9. Musumeci, T., Ventura, C. A., Giannone, I., Ruozi, B., Montenegro, L., Pignatello, R., Puglisi, G. (2006). PLA/PLGA nanoparticles for sustained release of docetaxel. *International Journal of Pharmaceutics 325*, 172.
10. Tong, R., Cheng, J. J. (2007). Anticancer polymeric nanomedicines. *Polymer Reviews (Philadelphia, PA, US) 47*, 345.
11. Hamblett, K. J., Senter, P. D., Chace, D. F., Sun, M. M. C., Lenox, J., Cerveny, C. G., Kissler, K. M., Bernhardt, S. X., Kopcha, A. K., Zabinski, R. F., Meyer, D. L., Francisco, J. A. (2004). Effects of drug loading on the antitumor activity of a monoclonal antibody drug conjugate. *Clinical Cancer Research 10*, 7063.
12. Schluep, T., Hwang, J., Cheng, J. J., Heidel, J. D., Bartlett, D. W., Hollister, B., Davis, M. E. (2006). Preclinical efficacy of the camptothecin-polymer conjugate IT-101 in multiple cancer models. *Clinical Cancer Research 12*, 1606.
13. Cheng, J., Teply, B. A., Sherifi, I., Sung, J., Luther, G., Gu, F. X., Levy-Nissenbaum, E., Radovic-Moreno, A. F., Langer, R., Farokhzad, O. C. (2007). Formulation of functionalized PLGA–PEG nanoparticles for *in vivo* targeted drug delivery. *Biomaterials 28*, 869.
14. Duncan, R. (2006). Polymer conjugates as anticancer nanomedicines. *Nature Reviews Cancer 6*, 688.
15. Duncan, R. (2003). The dawning era of polymer therapeutics. *Nature Reviews Drug Discovery 2*, 347.
16. Cheng, J., Khin, K. T., Jensen, G. S., Liu, A. J., Davis, M. E. (2003). Synthesis of linear, beta-cyclodextrin-based polymers and their camptothecin conjugates. *Bioconjugate Chemistry 14*, 1007.
17. Schluep, T., Hwang, J., Hildebrandt, I. J., Czernin, J., Choi, C. H. J., Alabi, C. A., Mack, B. C., Davis, M. E. (2009). Pharmacokinetics and tumor dynamics of the nanoparticle IT-101 from PET imaging and tumor histological measurements. *Proceedings of the National Academy of Sciences of the United States of America 106*, 11394.

18. Davis, M. E. (2009). Design and development of IT-101, a cyclodextrin-containing polymer conjugate of camptothecin. *Advanced Drug Delivery Reviews 61*, 1189.
19. Davis, M. E., Zuckerman, J. E., Choi, C. H. J., Seligson, D., Tolcher, A., Alabi, C. A., Yen, Y., Heidel, J. D., Ribas, A. (2010). Evidence of RNAi in humans from systemically administered siRNA via targeted nanoparticles. *Nature 464*, 1067.
20. Davis, M. E. (2009). The first targeted delivery of siRNA in humans via a self-assembling, cyclodextrin polymer-based nanoparticle: from concept to clinic. *Molecular Pharmaceutics 6*, 659.
21. Fox, M. E., Szoka, F. C., Frechet, J. M. J. (2009). Soluble polymer carriers for the treatment of cancer: the importance of molecular architecture. *Accounts of Chemical Research 42*, 1141.
22. Lee, C. C., Gillies, E. R., Fox, M. E., Guillaudeu, S. J., Frechet, J. M. J., Dy, E. E., Szoka, F. C. (2006). A single dose of doxorubicin-functionalized bow-tie dendrimer cures mice bearing C-26 colon carcinomas. *Proceedings of the National Academy of Sciences of the United States of America 103*, 16649.
23. http://en.wikipedia.org/wiki/List_of_bestselling_drugs.
24. Lee, C. C., MacKay, J. A., Frechet, J. M. J., Szoka, F. C. (2005). Designing dendrimers for biological applications. *Nature Biotechnology 23*, 1517.
25. Srinivasachari, S., Fichter, K. M., Reineke, T. M. (2008). Polycationic beta-cyclodextrin "Click Clusters": monodisperse and versatile scaffolds for nucleic acid delivery. *Journal of the American Chemical Society 130*, 4618.
26. Lu, H., Cheng, J. (2008). Controlled ring-opening polymerization of alpha-amino acid *N*-carboxyanhydrides and facile end group functionalization of polypeptides. *Journal of the American Chemical Society 130*, 12562.
27. Lu, H., Cheng, J. (2007). Hexamethyldisilazane-mediated controlled polymerization of alpha-amino acid *N*-carboxyanhydrides. *Journal of the American Chemical Society 129*, 14114.
28. Cheng, J., Deming, T. J. (2001). Controlled polymerization of beta-lactams using metal-amido complexes: synthesis of block copoly(beta-peptides). *Journal of the American Chemical Society 123*, 9457.
29. Medina, S. H., El-Sayed, M. E. H. (2009). Dendrimers as carriers for delivery of chemotherapeutic agents. *Chemical Reviews 109*, 3141.
30. Gillies, E. R., Frechet, J. M. J. (2002). Designing macromolecules for therapeutic applications: polyester dendrimer-poly(ethylene oxide) "bow-tie" hybrids with tunable molecular weight and architecture. *Journal of the American Chemical Society 124*, 14137.
31. Liu, G., Jia, L. (2004). Design of catalytic carbonylative polymerizations of heterocycles. Synthesis of polyesters and amphiphilic poly(amide-block-ester)s. *Journal of the American Chemical Society 126*, 14716.
32. Zhang, X. F., Li, Y. X., Chen, X. S., Wang, X. H., Xu, X. Y., Liang, Q. Z., Hu, J. L., Jing, X. B. (2005). Synthesis and characterization of the paclitaxel/MPEG–PLA block copolymer conjugate. *Biomaterials 26*, 2121.
33. Cheng, J., Khin, K. T., Davis, M. E. (2004). Antitumor activity of beta-cyclodextrin polymer–camptothecin conjugates. *Molecular Pharmaceutics 1*, 183.
34. Shen, Y. Q., Jin, E. L., Zhang, B., Murphy, C. J., Sui, M. H., Zhao, J., Wang, J. Q., Tang, J. B., Fan, M. H., Van Kirk, E., Murdoch, W. J. (2010). Prodrugs forming high drug loading

multifunctional nanocapsules for intracellular cancer drug delivery. *Proceedings of the National Academy of Sciences of the United States of America 132*, 4259.

35. Kim, S. C., Kim, D. W., Shim, Y. H., Bang, J. S., Oh, H. S., Kim, S. W., Seo, M. H. (2001). *In vivo* evaluation of polymeric micellar paclitaxel formulation: toxicity and efficacy. *Journal of Controlled Release 72*, 191.
36. Benny, O., Fainaru, O., Adini, A., Cassiola, F., Bazinet, L., Adini, I., Pravda, E., Nahmias, Y., Koirala, S., Corfas, G., D'Amato, R. J., Folkman, J. (2008). An orally delivered small-molecule formulation with antiangiogenic and anticancer activity. *Nature Biotechnology 26*, 799.
37. Lee, S. H., Zhang, Z. P., Feng, S. S. (2007). Nanoparticles of poly(lactide): tocopheryl polyethylene glycol succinate (PLA–TPGS) copolymers for protein drug delivery. *Biomaterials 28*, 2041.
38. Zhang, Z. P., Feng, S. S. (2006). Nanoparticles of poly(lactide)/vitamin E TPGS copolymer for cancer chemotherapy: synthesis, formulation, characterization and *in vitro* drug release. *Biomaterials 27*, 262.
39. Sengupta, S., Eavarone, D., Capila, I., Zhao, G. L., Watson, N., Kiziltepe, T., Sasisekharan, R. (2005). Temporal targeting of tumour cells and neovasculature with a nanoscale delivery system. *Nature 436*, 568.
40. Yoo, H. S., Lee, K. H., Oh, J. E., Park, T. G. (2000). *In vitro* and *in vivo* anti-tumor activities of nanoparticles based on doxorubicin-PLGA conjugates. *Journal of Controlled Release 68*, 419.
41. Chamberlain, B. M., Cheng, M., Moore, D. R., Ovitt, T. M., Lobkovsky, E. B., Coates, G. W. (2001). Polymerization of lactide with zinc and magnesium beta-diiminate complexes: stereocontrol and mechanism. *Journal of the American Chemical Society 123*, 3229.
42. O'Keefe, B. J., Hillmyer, M. A., Tolman, W. B. (2001). Polymerization of lactide and related cyclic esters by discrete metal complexes. *Journal of the Chemical Society-Dalton Transactions* 2215.
43. Cheng, J., Deming, T. J. (2001). Synthesis and conformational analysis of optically active poly(beta-peptides). *Macromolecules 34*, 5169.
44. Lu, H., Cheng, J. (2008). *N*-trimethylsilyl amines for controlled ring-opening polymerization of amino acid *N*-carboxyanhydrides and facile end group functionalization of polypeptides. *Journal of the American Chemical Society 130*, 12562.
45. Ouchi, M., Terashima, T., Sawamoto, M. (2009). Transition metal-catalyzed living radical polymerization: toward perfection in catalysis and precision polymer synthesis. *Chemical Reviews 109*, 4963.
46. Dechy-Cabaret, O., Martin-Vaca, B., Bourissou, D. (2004). Controlled ring-opening polymerization of lactide and glycolide. *Chemical Reviews 104*, 6147.
47. du Boullay, O. T., Marchal, E., Martin-Vaca, B., Cossio, F. P., Bourissou, D. (2006). An activated equivalent of lactide toward organocatalytic ring-opening polymerization. *Journal of the American Chemical Society 128*, 16442.
48. Hu, Y., Jiang, X., Ding, Y., Zhang, L., Yang, C., Zhang, J., Chen, J., Yang, Y. (2003). Preparation and drug release behaviors of nimodipine-loaded poly(caprolactone)–poly(ethylene oxide)–polylactide amphiphilic copolymer nanoparticles. *Biomaterials 24*, 2395.

49. Chen, F., Gao, Q., Hong, G., Ni, J. (2008). Synthesis of magnetite core–shell nanoparticles by surface-initiated ring-opening polymerization of L-lactide. *Journal of Magnetism and Magnetic Materials 320*, 1921.

50. Cheng, M., Attygalle, A. B., Lobkovsky, E. B., Coates, G. W. (1999). Single-site catalysts for ring-opening polymerization: synthesis of heterotactic poly(lactic acid) from rac-lactide. *Journal of the American Chemical Society 121*, 11583.

51. Tong, R., Cheng, J. (2008). Paclitaxel-initiated, controlled polymerization of lactide for the formulation of polymeric nanoparticulate delivery vehicles. *Angewandte Chemie (International edition) 47*, 4830.

52. Tong, R., Cheng, J. (2009). Ring-opening polymerization-mediated controlled formulation of polylactide-drug nanoparticles. *Journal of the American Chemical Society 131*, 4744.

53. Moore, D. R., Cheng, M., Lobkovsky, E. B., Coates, G. W. (2002). Electronic and steric effects on catalysts for $CO_2$/epoxide polymerization: subtle modifications resulting in superior activities. *Angewandte Chemie (International edition) 41*, 2599.

54. Tong, R., Cheng, J. (2010). Controlled synthesis of camptothecin–polylactide conjugates and nanoconjugates. *Bioconjugate Chemistry 21*, 111.

55. Tong, R., Yala, L., Fan, T. M., Cheng, J. (2010). The formulation of aptamer-coated paclitaxel–polylactide nanoconjugates and their targeting to cancer cells. *Biomaterials 31*, 3043.

56. Hertzberg, R. P., Caranfa, M. J., Hecht, S. M. (1989). On the mechanism of topoisomerase-I inhibition by camptothecin: evidence for binding to an enzyme DNA complex. *Biochemistry 28*, 4629.

57. Muggia, F. M., Dimery, I., Arbuck, S. G., *Camptothecin and its analogs: an overview of their potential in cancer therapeutics*, Academy of Sciences, New York, 1996.

58. Mi, Z. H., Burke, T. G. (1994). Differential interactions of camptothecin lactone and carboxylate forms with human blood components. *Biochemistry 33*, 10325.

59. Mi, Z. H., Burke, T. G. (1994). Marked interspecies variations concerning the interactions of camptothecin with serum albumins: a frequency-domain fluorescence spectroscopic study. *Biochemistry 33*, 12540.

60. Cheng, J., Khin, K. T., Davis, M. E. (2004). Antitumor activity of beta-cyclodextrin polymer: camptothecin conjugates. *Molecular Pharmaceutics 1*, 183.

61. Kunii, R., Onishi, H., Machida, Y. (2007). Preparation and antitumor characteristics of PLA/(PEG–PPG–PEG) nanoparticles loaded with camptothecin. *European Journal of Pharmaceutics and Biopharmaceutics 67*, 9.

62. Li, C. (1998). Complete regression of well-established tumors using a novel water-soluble poly (L-glutamic acid)–paclitaxel conjugate. *Cancer Research 58*, 2404.

63. Singer, J. W. (2005). Paclitaxel poliglumex (XYOTAX, CT-2103): a macromolecular taxane. *Journal of Controlled Release 109*, 120.

64. Fonseca, C., Simoes, S., Gaspar, R. (2002). Paclitaxel-loaded PLGA nanoparticles: preparation, physicochemical characterization and *in vitro* anti-tumoral activity. *Journal of Controlled Release 83*, 273.

65. Gref, R., Minamitake, Y., Peracchia, M., Trubetskoy, V. S., Torchilin, V. P., Langer, R. (1994). Biodegradable long-circulating polymeric nanospheres. *Science 263*, 1600.

66. Panyam, J., Labhasetwar, V. (2003). Biodegradable nanoparticles for drug and gene delivery to cells and tissue. *Advanced Drug Delivery Reviews 55*, 329.

67. Soppimath, K. S., Aminabhavi, T. M., Kulkarni, A. R., Rudzinski, W. E. (2001). Biodegradable polymeric nanoparticles as drug delivery devices. *Journal of Controlled Release 70*, 1.
68. Mastropaolo, D., Camerman, A., Luo, Y. G., Brayer, G. D., Camerman, N. (1995). Crystal and molecular-structure of paclitaxel (taxol). *Proceedings of the National Academy of Sciences of the United States of America 92*, 6920.
69. Mathew, A. E., Mejillano, M. R., Nath, J. P., Himes, R. H., Stella, V. J. (1992). Synthesis and evaluation of some water-soluble prodrugs and derivatives of taxol with antitumor-activity. *Journal of Medicinal Chemistry 35*, 145.
70. Magri, N. F., Kingston, D. G. I., Jitrangsri, C., Piccariello, T. (1986). Modified taxols. 3. Preparation and acylation of baccatin-III. *The Journal of Organic Chemistry 51*, 3239.
71. Moore, D. R., Cheng, M., Lobkovsky, E. B., Coates, G. W. (2003). Mechanism of the alternating copolymerization of epoxides and $CO_2$ using beta-diiminate zinc catalysts: evidence for a bimetallic epoxide enchainment. *Journal of the American Chemical Society 125*, 11911.
72. Cheng, M., Moore, D. R., Reczek, J. J., Chamberlain, B. M., Lobkovsky, E. B., Coates, G. W. (2001). Single-site beta-diiminate zinc catalysts for the alternating copolymerization of $CO_2$ and epoxides: catalyst synthesis and unprecedented polymerization activity. *Journal of the American Chemical Society 123*, 8738.
73. http://www.doxil.com/doxil-supply-shortage.
74. http://www.doxil.com/ovarian-cancer/hand-foot-syndrome.
75. Lorusso, D., Di Stefano, A., Carone, V., Fagotti, A., Pisconti, S., Scambia, G. (2007). Pegylated liposomal doxorubicin-related palmar-plantar erythrodysesthesia ('hand-foot' syndrome). *Annals of Oncology 18*, 1159.
76. Yoo, H. S., Park, T. G. (2001). Biodegradable polymeric micelles composed of doxorubicin conjugated PLGA–PEG block copolymer. *Journal of Controlled Release 70*, 63.
77. Missirlis, D., Kawamura, R., Tirelli, N., Hubbell, J. A. (2006). Doxorubicin encapsulation and diffusional release from stable, polymeric, hydrogel nanoparticles. *European Journal of Pharmaceutical Sciences 29*, 120.
78. Altreuter, D. H., Dordick, J. S., Clark, D. S. (2002). Nonaqueous biocatalytic synthesis of new cytotoxic doxorubicin derivatives: exploiting unexpected differences in the regioselectivity of salt-activated and solubilized subtilisin. *Journal of the American Chemical Society 124*, 1871.
79. Yoo, H. S., Oh, J. E., Lee, K. H., Park, T. G. (1999). Biodegradable nanoparticles containing doxorubicin–PLGA conjugate for sustained release. *Pharmaceutical Research 16*, 1114.
80. Nishiyama, N., Kataoka, K. (2006). Nanostructured devices based on block copolymer assemblies for drug delivery: designing structures for enhanced drug function. *Advances in Polymer Science 193*, 67.
81. Greenfield, R. S., Kaneko, T., Daues, A., Edson, M. A., Fitzgerald, K. A., Olech, L. J., Grattan, J. A., Spitalny, G. L., Braslawsky, G. R. (1990). Evaluation in vitro of adriamycin immunoconjugates synthesized using an acid-sensitive hydrazone linker. *Cancer Research 50*, 6600.
82. Kaneko, T., Willner, D., Monkovic, I., Knipe, J. O., Braslawsky, G. R., Greenfield, R. S., Vyas, D. M. (1991). New hydrazone derivatives of adriamycin and their

immunoconjugates: a correlation between acid stability and cytotoxicity. *Bioconjugate Chemistry 2*, 133.

83. Ulbrich, K., Subr, V. (2004). Polymeric anticancer drugs with pH-controlled activation. *Advanced Drug Delivery Reviews 56*, 1023.
84. Florent, J. C., Monneret, C. (2008). Doxorubicin conjugates for selective delivery to tumors. *Anthracycline Chemistry and Biology II: Mode of Action, Clinical Aspects and New Drugs 283*, 99.
85. Farokhzad, O. C., Cheng, J., Teply, B. A., Sherifi, I., Jon, S., Kantoff, P. W., Richie, J. P., Langer, R. (2006). Targeted nanoparticle–aptamer bioconjugates for cancer chemotherapy *in vivo*. *Proceedings of the National Academy of Sciences of the United States of America 103*, 6315.
86. Caliceti, P., Veronese, F. M. (2003). Pharmacokinetic and biodistribution properties of poly(ethylene glycol)–protein conjugates. *Advanced Drug Delivery Reviews 55*, 1261.
87. Pierri, E., Avgoustakis, K. (2005). Poly(lactide)–poly(ethylene glycol) micelles as a carrier for griseofulvin. *Journal of Biomedical Materials Research. Part A 75A*, 639.
88. Tuerk, C., Gold, L. (1990). Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage-T4 DNA-polymerase. *Science 249*, 505.
89. Ellington, A. D., Szostak, J. W. (1992). Selection *in vitro* of single-stranded-DNA molecules that fold into specific ligand-binding structures. *Nature 355*, 850.
90. Farokhzad, O. C., Jon, S. Y., Khadelmhosseini, A., Tran, T. N. T., LaVan, D. A., Langer, R. (2004). Nanoparticle–aptamer bioconjugates: a new approach for targeting prostate cancer cells. *Cancer Research 64*, 7668.
91. Keefe, A. D., Pai, S., Ellington, A. (2010). Aptamers as therapeutics. *Nature Reviews Drug Discovery 9*, 537.
92. Nimjee, S. M., Rusconi, C. P., Sullenger, B. A. (2005). Aptamers: an emerging class of therapeutics. *Annual Review of Medicine 56*, 555.
93. Cao, Z. H., Tong, R., Mishra, A., Xu, W. C., Wong, G. C. L., Cheng, J. J., Lu, Y. (2009). Reversible cell-specific drug delivery with aptamer-functionalized liposomes. *Angewandte Chemie (International edition) 48*, 6494.
94. Lupold, S. E., Hicke, B. J., Lin, Y., Coffey, D. S. (2002). Identification and characterization of nuclease-stabilized RNA molecules that bind human prostate cancer cells via the prostate-specific membrane antigen. *Cancer Research 62*, 4029.
95. Chaney, E. J., Tang, L., Tong, R., Cheng, J. J., Boppart, S. A. (2010). Lymphatic biodistribution of polylactide nanoparticles. *Molecular Imaging 9*, 153.
96. Chan, J. M., Zhang, L. F., Tong, R., Ghosh, D., Gao, W. W., Liao, G., Yuet, K. P., Gray, D., Rhee, J. W., Cheng, J. J., Golomb, G., Libby, P., Langer, R., Farokhzad, O. C. (2010). Spatiotemporal controlled delivery of nanoparticles to injured vasculature. *Proceedings of the National Academy of Sciences of the United States of America 107*, 2213.
97. Kaplan, B., Meier-Kriesche, H. U. (2004). Renal transplantation: a half century of success and the long road ahead. *Journal of the American Society of Nephrology 15*, 3270.
98. Wong, W., Venetz, J. P., Tolkoff-Rubin, N., Pascual, M. (2005). Immunosuppressive strategies in kidney transplantation: which role for the calcineurin inhibitors? *Transplantation 80*, 289.
99. Chapman, J. R., Nankivell, B. J. (2006). Nephrotoxicity of ciclosporin A: short-term gain, long-term pain? *Nephrology, Dialysis, Transplantation 21*, 2060.

100. Neto, A. B., Haapalainen, E., Ferreira, R., Feo, C. F., Misiako, E. P., Vennarecci, G., Porcu, A., Dib, S. A., Goldenberg, S., Gomes, P. O., Nigro, A. T. (1999). Metabolic and ultrastructural effects of cyclosporin A on pancreatic islets. *Transplant International 12*, 208.

101. Azzi, J., Tang, L., Moore, R., Tong, R., El Haddad, N., Akiyoshi, T., Mfarrej, B., Yang, S. M., Jurewicz, M., Ichimura, T., Lindeman, N., Cheng, J. J., Abdi, R. (2010). Polylactide-cyclosporin A nanoparticles for targeted immunosuppression. *The FASEB Journal 24*, 3927.

102. Rieth, L. R., Moore, D. R., Lobkovsky, E. B., Coates, G. W. (2002). Single-site beta-diiminate zinc catalysts for the ring-opening polymerization of beta-butyrolactone and beta-valerolactone to poly(3-hydroxyalkanoates). *Journal of the American Chemical Society 124*, 15239.

103. Sharp, P. A., Langer, R. (2011). Promoting convergence in biomedical science. *Science 333*, 527.

104. Wong, C., Stylianopoulos, T., Cui, J. A., Martin, J., Chauhan, V. P., Jiang, W., Popovic, Z., Jain, R. K., Bawendi, M. G., Fukumura, D. (2011). Multistage nanoparticle delivery system for deep penetration into tumor tissue. *Proceedings of the National Academy of Sciences of the United States of America 108*, 2426.

105. Sugahara, K. N., Teesalu, T., Karmali, P. P., Kotamraju, V. R., Agemy, L., Greenwald, D. R., Ruoslahti, E. (2010). Coadministration of a tumor-penetrating peptide enhances the efficacy of cancer drugs. *Science 328*, 1031.

106. Jain, R. K., Stylianopoulos, T. (2010). Delivering nanomedicine to solid tumors. *Nature Reviews Clinical Oncology 7*, 653.

107. Torchilin, V. P. (2006). Multifunctional nanocarriers. *Advanced Drug Delivery Reviews 58*, 1532.

108. Sutton, D., Nasongkla, N., Blanco, E., Gao, J. M. (2007). Functionalized micellar systems for cancer targeted drug delivery. *Pharmaceutical Research 24*, 1029.

109. Chaffer, C. L., Weinberg, R. A. (2011). A perspective on cancer cell metastasis. *Science 331*, 1559.

110. Hanahan, D., Weinberg, Robert A. (2011). Hallmarks of cancer: the next generation. *Cell 144*, 646.

# NANOCRYSTALS PRODUCTION, CHARACTERIZATION, AND APPLICATION FOR CANCER THERAPY

Christin P. Hollis

*College of Pharmacy, University of Kentucky, Lexington, KY, USA*

Tonglei Li

*Department of Industrial and Physical Pharmacy, Purdue University, West Lafayette, IN, USA*

## 8.1 INTRODUCTION

It is estimated that two-thirds of the newly synthesized drugs are poorly soluble in water [1]. Delivery of these compounds requires significant efforts and, in some cases, becomes extremely difficult. Poor solubility limits the bioavailability when delivered through the oral route; for many compounds, parental delivery, which requires special formulation treatment to overcome the solubility limit, remains to be the only choice. A common practice is to solubilize a poorly soluble drug with specialized organic molecules, which include organic solvents, surfactants, and macromolecules. Unfortunately, adverse effects resulted from the use of such solubilizing molecules, for example, hypersensitivity reactions, nephrotoxicity, neurotoxicity, and neutropenia, have been reported [2,3]. In addition, the individually solubilized drug molecules can also reach healthy organs and tissues indiscriminately and inflict undesired effects. Hence, there is a need for developing formulation that can deliver poorly water-soluble drugs, including anticancer agents.

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*,
Edited by Shayne C. Gad
Copyright © 2013 John Wiley & Sons, Inc.

Attempting to target tumors, anticancer drugs are generally formulated as nanoparticles in order to take advantage of the so-called enhanced permeability and retention (EPR) effect [4], which characterizes tumors' underdeveloped vasculatures and ineffective lymphatic systems. Various nanosized carrier designs have been developed, including liposomes [5,6], micelles [7], solid lipid nanoparticles [8,9], polymeric nanoparticles [10,11], and dendrimers [12]. In many of these nanoparticle designs, drug molecules are solubilized, dispersed, and encapsulated by specialized chemicals.

Despite a significant amount of efforts on the development of new delivery systems over the last few decades, there are only a few Food and Drug Administration (FDA)-approved cancer therapy products in which poorly soluble drugs are formulated as nanoparticles. Two are liposomal formulations, Doxil® (doxorubicin) and Daunosome® (daunorubicin), and one other is Abraxane®, albumin-bound paclitaxel (PTX) nanoparticles. Impeding the progress in the anticancer drug product development are several factors that lie in the current formulation design strategy. When a drug is solubilized into a nanoconstruct, the multicomponent system is prone to physical instability and structural destruction during storage and/or *in vivo* passage, resulting in drug leakage [13,14]. The drug loading is typically limited [7,15]; high manufacturing cost, multiple steps in preparation and synthesis, and difficulties in scale-up have also limited the potential of the aforementioned vehicles for clinical usage.

Clearly, it has been and will remain a significant challenge to deliver a poorly soluble anticancer compound. Current solubilization approaches face inherent limitations that need to be overcome for the clinical application, particularly when tumor targeting is desired. It is recognized that in order to fully take advantage of the EPR effect, a clinically viable tumor-targeted delivery system has to be stable, should have a high drug loading, and be capable of sustaining the release of drug molecules in the tumor [16]. It is rather futile to develop an intricate vehicle that has little clinical potential. New approaches need to be developed to shift the paradigm of delivering poorly soluble drugs, including antineoplastic agents.

Over the last few years, nanocrystal formulations have been of considerable interest for delivering poorly soluble drugs via oral, parenteral, and other routes [17]. Administered directly as nanosized solid particles, drug nanocrystals require no solubilizing and/or encapsulating chemicals, thus circumventing side effects and instability drawbacks impeding many existing delivery systems. More importantly, production of nanocrystals at the large scale is feasible due to the simplicity of formulation and absence of solubilizing, encapsulating, or conjugating steps. Nanocrystals are expected to target the tumor via the EPR effect and offer less toxicity, improving the anticancer efficacy and pharmacokinetic properties. Nanocrystal formulations may likely rejuvenate the delivery of poorly soluble anticancer compounds, existing and to be discovered.

In this chapter, we first review the production of nanocrystals of pharmaceutical organic compounds. We then focus on the application of nanocrystals for cancer therapy.

## 8.2 NANOCRYSTAL PRODUCTION

Crystals are regular, periodic supramolecular structures assembled by organic or other types of molecules or atoms. Characteristically, a crystal shows distinct X-ray diffraction patterns, and has a specific melting point, density, and solubility, in addition to other well-defined physicochemical properties. Compared with the amorphous state, in which molecules are randomly packed, the crystalline state is more stable, exemplified by the fact that amorphous materials can readily undergo spontaneous phase transition and recrystallization. As such, using crystalline materials in a formulation assures better physical stability than amorphous formulations.

Nanocrystals of poorly soluble drugs can offer several advantages over solubilized and encapsulated formulations. Besides the physical stability, the dissolution rate of nanocrystals can be significantly enhanced—under sink condition—due to the increase in the surface area. More importantly, manufacturing at the industrial scale is possible. In the following sections, we discuss two major approaches for producing nanocrystals of organic molecules. The top-down approach processes large crystals and physically comminute into smaller pieces; the bottom-up approach yields nanocrystals from solution by crystallization. Several review articles have focused on the techniques, mainly the top-down approach [17–21].

### 8.2.1 Top-Down Approach

There are two basic diminution techniques: wet milling and high-pressure homogenization. Wet milling involves mechanical attrition, where drug particles are wetted by an aqueous solution of surfactants and subjected to shearing and grinding by mechanical toolings, such as milling balls in a milling container. The particle size is reduced and may reach a few hundred nanometers. The Nanocrystal® technique [22], a wet milling process patented by Élan Co., has yielded several oral products including Rapamune®, Emend®, Tricor®, and Megace ES® [23]. One drawback of the technique is potential contamination from the erosion of metal milling balls or pearls. The amount of contamination is determined by the hardness of drug particles as well as the milling time, which can take up to several days [24]. The use of polymeric beads (e.g., polystyrene derivatives) may be helpful in minimizing the erosion.

High-pressure homogenization is another fragmentation technique. One development is based on jet-stream microfluidization [25], in which two fluid streams of particle suspensions collide under high pressure in a Y-shaped chamber, leading to particle collision and subsequent particle rupture. The marketed product Triglide® developed by SkyePharma Co. is produced by Insoluble Drug Delivery-Particles (IDD-P™) technology, a microfluidization method [26]. Piston-gap homogenizers are also used to produce nanosized solid particles by forcing a suspension of drug particles with a piston through a thin gap under high pressure [27]. The combination of cavitations, high shear forces, and turbulent flow fractures the particles. The outcome of particle fragmentation is decided by several factors, including the power of homogenization, particle hardness, and number of the piston-moving cycles. Compared with wet milling, the homogenization method may yield less

contamination during the production process [28]. There are currently two patented technologies for nanocrystal production that employ the piston-gap homogenization method, DissoCubes® and Nanopure®. In the former, developed by SkyePharma Co., drug powders are dispersed in an aqueous solution of surfactants and subjected to homogenization [27]; in the latter, previously owned by PharmaSol GmbH and now acquired by Abbott Co., drug particles are suspended in either nonaqueous dispersion media (e.g., oils and liquefied polyethylene glycol) or aqueous-organic cosolvents (e.g., glycerol–water and ethanol–water mixtures) prior to homogenization [27]. These operations require that one starts with drug particles no larger than 25 μm so that the blockage of the homogenizer can be minimized [29].

Currently, the top-down approach dominates the production of nanocrystals and is used to manufacture several products on the market [21]. Despite its popularity, there are inherent drawbacks in this approach, including the high-energy input, prolonged operation time, possibility of contamination, and decreased crystallinity or stability [30]. The stability concern is particularly critical for a nanocrystal system to be considered for parenteral drug delivery. The amorphous content produced during the high-energy process can lead to phase instability, which in turn causes uncertainty and potential variability in pharmacokinetic behavior, which should be tightly controlled.

### 8.2.2 Bottom-Up Approach

Nanocrystals can be grown directly from solution. Nucleation and crystal growth thus play the essential role in the bottom-up approach. To induce nucleation, a solution needs to become supersaturated. Cooling, evaporating the solvent, and mixing with an antisolvent are among several ways to create supersaturation [31]. Still, a solution can become supersaturated to some extent without producing any nuclei. This is because nucleation is energetically unfavorable. According to the classical nucleation theory [32–34], the change in free energy when a nucleus starts to grow includes two energy components, one associated with the creation of the new surface, and another with the packing of molecules in the bulk crystal. As depicted in Figure 8.1, the surface-related term, $\Delta G_S$, is always positive (energetically unfavorable) because creation of a new surface of a solid particle requires energy input. On the other hand, the bulk-related term, $\Delta G_V$, is negative (energetically favorable) due to the formation of intermolecular interactions between molecules in the nucleus. When a nucleus is small, $\Delta G_S$ dominates, and the overall free energy change is positive so that the development of the nucleus is not spontaneous. As the nucleus grows and surpasses, the so-called critical nucleus size, $\Delta G_V$ becomes dominant and the overall term becomes negative. Subsequently, the nucleus can grow freely as long as supersaturation is maintained.

In reality, homogeneous nucleation—nucleation without the aid of substances other than the solute molecules—seldom occurs. Foreign materials, impurities, or even the surface of the crystallization vessel often help to induce nucleation, which is denoted as the heterogeneous nucleation [35]. According to the classical nucleation theory, the presence of a foreign surface lowers the $\Delta G_S$ term for a nucleus to form on

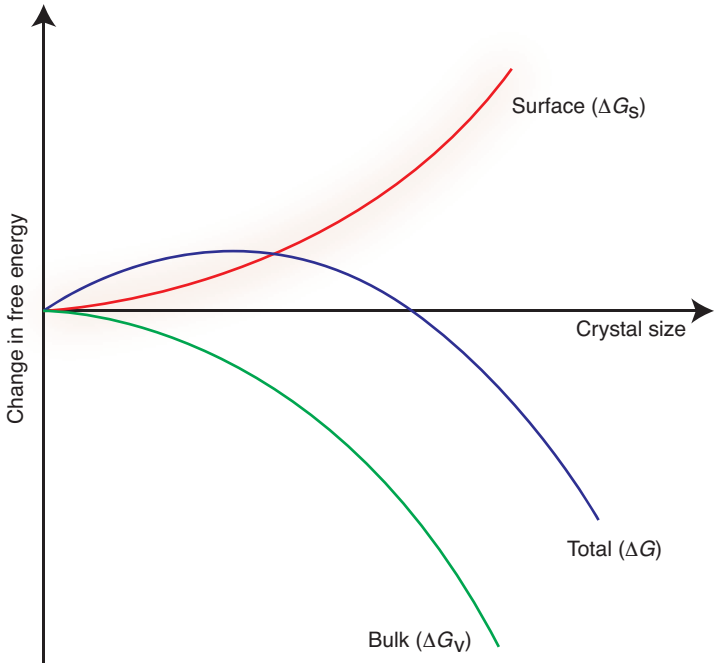


**FIGURE 8.1** The change in free energy as a function of crystal size during the nucleation and crystal growth process.

the surface. In other words, the nucleus requires less energy for surface creation when it forms on a pre-existing surface.

In order to produce nanosized crystals from solution, it is thus extremely critical to control the nucleation process. It is important not only for controlling the particle size but also for the size distribution. Ideally, nucleation of a large number of crystalline embryos must occur concurrently from the supersaturated solution. Rapid depletion of the solute molecules from the solution limits further growth of each nucleus so as to achieve a nanosize. Having a large number of nuclei produced at the same time also keeps the particle size distribution narrow. Similar particle sizes are not only preferred for ensuring product quality but also important for maintaining the stability of nanocrystals. A narrow distribution of particle size can greatly minimize the so-called Ostwald ripening phenomenon, in which larger particles grow at the expense of redissolving smaller ones because of their differences in surface energy [36]. For this purpose, the induction of nucleation needs to be abrupt and homogeneous throughout the growth medium. One common strategy is to mix the solution of a drug compound with an antisolvent so that the supersaturation is reached immediately; at the same time, the mixture is subjected to intense sonication and/or mechanical stirring. It is known that air cavities created by sonication may trigger nucleation because of the presence of air–liquid interfaces and locally concentrated liquid pockets. Nevertheless,

because nucleation is such a difficult process to control [37], subtle changes in growth conditions can result in drastic variations in product quality. For producing nanocrystals, extreme care is needed for identifying a suitable set of peripheral conditions [38], which mostly likely vary for distinct compounds. Figure 8.2 shows a few samples of nanocrystals that were produced by the bottom-up approach in our laboratory.

There are several approaches that have been developed to create supersaturation and induce nucleation [20]. They include hydrosols [36], high-gravity controlled

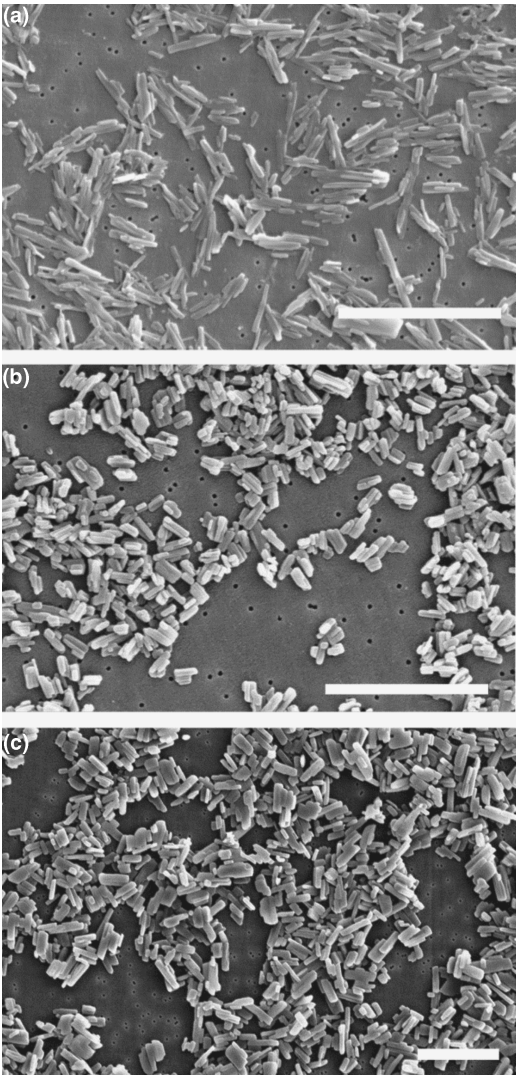


**FIGURE 8.2** SEM images of nanocrystals grown in our laboratory: camptothecin (a), paclitaxel (b), and ZSTK474 (c). Scale bar: 2 μm.

precipitation technology (HGCP) [39,40], confined impinging jets [41], supercritical fluid (SCF) technology [42], and sonoprecipitation [43], as discussed further.

***8.2.2.1 Hydrosols.*** Considered to be a more mature method, hydrosols or aqueous nanosuspensions of poorly water-soluble drugs have been developed by Sandoz (now Novartis) [44]. The active ingredient is dissolved in aqueous-organic miscible cosolvents (e.g., ethanol, isopropanol, and acetone), while various peptisers and/or stabilizing agents (e.g., citric acid, gelatin, ethyl cellulose N7, lecithin) may be added to the aqueous phase to prevent aggregation. A matrix former, generally sugar or sugar alcohol (e.g., mannitol), is also added to the solution to prevent agglomeration and aid in the resuspension process. In the laboratory setting, solid nanoparticles can be produced by mechanically mixing the various components in a beaker while removing the organic solvent through evaporation. For the industrial production, the organic and aqueous phases are pumped into a static mixer and forced through an atomizing nozzle, leading to the formation and drying of nanoparticles [44–47]. One study reported that spray-dried cyclosporin A contained 38 μm particle agglomerates that were then reduced to 120 nm upon redispersion in water [45]; the nanoparticles produced by this method were nonetheless amorphous [45,46].

***8.2.2.2 High-Gravity Controlled Precipitation (HGCP).*** HGCP is considered to be one of the most promising techniques to produce nanoprecipitations at the commercial scale [20]. The key component in HGCP is a rotating packed bed (RPB), where two liquid streams are pumped in and mixed together vigorously by centrifugation. Subsequently, the mixture is spread into thin films and ultimately broken into tiny droplets. In addition to using antisolvents, mixing two streams of reactants can also produce precipitates chemically in HGCP. Depending on the compound and conditions, a product may be amorphous or crystalline. Nanoparticles of a few pharmaceutical compounds, including cephradine [48], danazol [49], and salbutamol sulfate [50,51], have been successfully produced by this method.

***8.2.2.3 Flash Nanoprecipitation by Confined Liquid Impinging Jets (CLIJ).*** In this technique, two respective jet streams of a drug solution and an antisolvent are forced together by opposing nozzles that are mounted in a small chamber to create localized, intensified mixing. The mixing causes precipitation of nanoparticles within the residence time of mixing. The velocity, drug concentration, and volume ratio of the two streams are considered to be the factors determining the particle size distribution of a product. CLIJ has been utilized in the production of salbutamol sulfate, ibuprofen, cyclosporin A, and amphotericin B [52–54]. GRAS (generally regarded as safe) additives may be added to assist the production.

***8.2.2.4 Supercritical Fluid (SCF) Technology.*** This method takes advantage of the unique physical properties of supercritical fluids, including low density and viscosity as well as high diffusivity of solute molecules to attain rapid mixing. In addition, quick and easy removal of a SCF without excessive drying can greatly facilitate the precipitation of nanoparticles. Supercritical carbon dioxide ($SCO_2$) is

mostly used and considered green or environmentally friendly for processing most pharmaceuticals. Depending on the solubility of a compound in $SCO_2$, production can be accomplished by allowing the drug–$SCO_2$ mixture to expand under ambient conditions (method called rapid expansion of supercritical solution or RESS [55]) or by using $SCO_2$ as the antisolvent when the drug is poorly soluble in $SCO_2$ (a process is called supercritical antisolvent or SAS [55]). In both the methods, it is critical that other solvents used are miscible with $SCO_2$; water is actually immiscible. Amoxicillin, ampicillin, and rifampicin nanoparticles were successfully produced by the SCF approach [56]. 100–500 nm protein nanoparticles of insulin, rhDNase, lysosome, and albumin were precipitated from aqueous ethanol solution by using ethanol as the co-solvent [57].

***8.2.2.5 Sonoprecipitation.*** Ultrasonic waves can create cavitations that subsequently collapse, releasing shock waves that may assist in rapid and more uniform nucleation [58]. Ultrasound can also reduce particle agglomeration by minimizing particle contacts [43]. Drug microparticles were reportedly produced by the sonocrystallization method [59–61]; it is also shown that the method could produce nanoparticles of amorphous cefuroxime axetil [62]. Factors that influence the quality of a final product include (i) ultrasound frequency, intensity, and power, (ii) sonication probe size and immersion depth, (iii) solution volume, and (iv) duration. Although the experimental setup appears to be straightforward, this technology has not been adopted for commercial production.

The quality of nanocrystals produced by the bottom-up approach is expected to be superior to that by the top-down method. Particle size and morphology can be better controlled, and it is possible to produce pure nanocrystals with little or no amorphous content. Shorter production time, less energy consumption, and reduced contamination also make the precipitation techniques desirable. Clearly, the bottom-up approach holds great promise for developing nanocrystals for parenteral delivery, which demands uncompromised quality and stability. Although no such product is currently in the market, there is certainly a growing interest and need in this field. Challenges remain in controlling nucleation and subsequent crystal growth so that batch-to-batch reproducibility can be achieved. It is certain that among various methods that are being developed, ones that are simple, reproducible, and yet cost-effective for production scale-up should have great potential in commercial production [36].

### 8.2.3 Combined Approach

There have been efforts to combine the two aforementioned approaches. The NanoEdge™ patent by Baxter claims that nanocrystals, ranging from 400 to 2 μm, can be prepared by crystallization followed by high-pressure homogenization [63]. Our research group used an antisolvent approach to grow nanocrystals by rapidly precipitating anticancer drugs under intense stirring and sonication [64,65]. To further reduce the size, the nanocrystal suspension can subsequently be homogenized. Nanocrystals with an average particle size ranging from 200 to 800 nm have

been successfully produced in our laboratory. The second step of homogenization breaks down larger crystals and, particularly, aggregates, achieving a much narrower particle size distribution.

## 8.3 PARTICLE STABILIZATION

Nanocrystals and nanoparticles of drugs, whether prepared by the top-down or bottom-up approach, often require surfactants to be physically attached to the surface of each particle in order to prevent the particles from aggregation during their preparation, storage, and administration. Particle aggregation mutually facilitates Ostwald's ripening and eventually becomes irreversible, leading to larger particles and precipitates. Surfactant usage is particularly necessary when the surface of nanocrystals is uncharged. In general, polymeric surfactants are commonly used to introduce steric repulsion between particles; electrostatic repulsion also promotes particle segregation when ionic surfactants are used. Povidones, pluronics, and cellulosics are the most commonly used surfactants. Natural biological molecules, such as lecithins and cholic acid derivatives, are utilized as well [66].

Finding a suitable surfactant for a particular drug is experimentally demanding [67]. A stabilizer obviously needs to have sufficient affinity for the particle surface and the amount used should provide sufficient steric and/or electrostatic repulsion between the particles [68]. Excess amounts of surfactants do not always lead to better coverage due to the tendency of surfactants to form micelles when the concentration surpasses the critical micelle concentration (CMC) [69,70]. A recent study of the effects of 13 stabilizers during wet milling of 9 drug compounds [71] concluded that the semisynthetic polymers, including hydroxypropylmethylcellulose (HPMC), methylcellulose (MC), hydroxyethylcellulose (HEC), hydroxypropylcellulose (HPC), carboxymethylcellulose sodium salt (NaCMC), and alginic acid sodium salt (NaAlg), yielded rather poor stabilizing performance, likely because of their high viscosity. Less viscous stabilizers, including linear synthetic polymers (PVP K30 and K90) and synthetic copolymers (poloxamer 188 and Kollicoat IR®), showed better stabilizing capabilities, especially at higher concentrations. Tween® 80 and TPGS (D-α-tocopherol polyethylene glycol 1000 succinate) gave the best stabilizing performance. Moreover, the timing of surfactant addition may affect particle properties. Stabilizers are commonly added along with the starting microcrystalline materials for wet milling or homogenization. However, the addition of a stabilizer prior to crystallization may alter the nucleation and subsequent crystal growth outcome. For example, the morphology of the nanocrystals of ZSTK474, a novel PI3K inhibitor [72], that were prepared in our laboratory by the bottom-up approach varied from an elongated to a plate-like shape due to the addition of 1% (w/v) Pluronic® F127 (Figure 8.3).

Because nanocrystals undergo consistent dynamic changes, especially in solution or in contact with moisture—dissolution and surface recrystallization—albeit at a slow rate, surface adsorption by surfactants may not be able to sustain stability of nanocrystals for the long-term storage. Alternatively, nanocrystal products may be

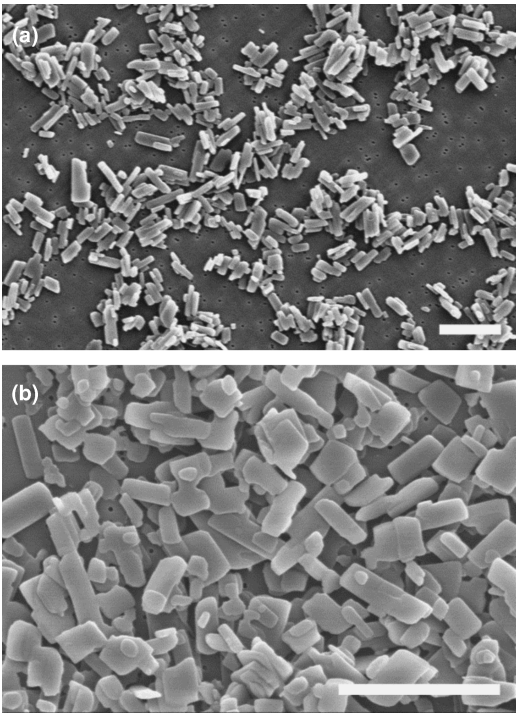


**FIGURE 8.3** SEM images of ZSTK474 nanocrystals grown in the absence (a) and presence (b) of Pluronic F127. Scale bar: 2 μm.

further processed by freeze drying, spray drying, or pelletization [19]; water-soluble sugars, such as mannitol, trehalose, and dextran, are commonly added as matrix formers prior to drying [73]. In the absence of a solvent, the surface dynamics and mobility are greatly minimized. For further sterilization, nanosuspensions for intravenous use are often subjected to terminal heat and gamma irradiation [74] as well as filtration [75], and stored as liquid suspension or lyophilized solid form.

## 8.4 PARTICLE CHARACTERIZATION

### 8.4.1 Particle Size

Nanocrystals need to be fully evaluated to determine their (i) particle size and size distribution, (ii) surface charge, (iii) crystallinity, and (iv) dissolution rate [18,29,73,76]. For rapid and noninvasive determination of the mean particle size and size distribution, photon correlation spectroscopy (PCS) is mostly used for particles ranging from 3 nm to 3 μm. The size distribution is indicated by polydispersity index (PDI), which is indicative of the width of particle size distribution. A PDI value of 0.1–0.25 suggests that the size distribution is fairly narrow, while a PDI value greater than 0.5 indicates a very broad

distribution. To achieve long-term stability, it is important to maintain the PDI parameter as low as possible. When the particle size is >3 μm, laser diffractometry (LD) can be used. Note that the particle size data obtained by LD are volume based, and those by PCS are light intensity weighted. Also note that these techniques work well for spherical particles; for nanocrystals, which tend to have large aspect ratios (i.e., length/width), the size results often deviate significantly from those determined by electron microscope imaging. For nonspherical particles, LD may overestimate the particle size distribution [77,78]. Over-prediction of the median diameters is also observed [79]. Thus, caution is needed when measuring and reporting particle size of irregular shape particles [80]. A coulter counter, which yields the absolute number of particles per volume unit of different size classes, can be used as an additional method, for example, to quantify any contamination from microparticles. When a formulation is intended for intravenous delivery, it is essential to ensure that the particle size is smaller than the smallest blood capillary (a few micrometers) so that potential capillary blockade or emboli formation can be avoided [29].

### 8.4.2 Surface Charge

Surface charging is important for the stability of nanocrystals in solution. According to the Derjaguin, Verwey, Landau, and Overbeek (DLVO) theory, the stability of a colloidal system depends on the balance between two counteracting forces, attractive (van der Waals) and repulsive (electrical double layer) [81]. The electrical double layer surrounding a particle consists of an inner region (Stern layer), where counterions are strongly bound to the particle, and an outer region (diffuse layer), where counterions are loosely associated. Within the diffuse layer, there exists a notional boundary, named the slipping plane, which separates counterions that move along with the particle and those that stay with the bulk fluid. The electric potential difference at the slipping plane versus a point in the bulk medium is defined as the zeta potential. It can be measured by using light scattering to monitor the electrophoretic mobility when the colloidal system is subjected to an electric field. In addition to the sample's surface chemistry, pH of the medium, conductivity of the fluid, and concentration of additives can influence the distributions of counterions surrounding the interface and thereby the zeta potential. To resist flocculation and aggregation, particles need to have sufficiently high repulsion forces. It is generally recognized that a specific value of zeta potential can be useful in predicting the physical stability of a nanosuspension formulation. ±30 mV or larger (absolute value) is desired for an electrostatically stabilized suspension. For particles with smaller zeta potentials, surfactants may be needed to provide further steric and/or electrostatic repulsions [82]. Numerous studies have shown that surface potential can have a drastic influence on cellular uptake or even tumor targeting of nanoparticles [83–87]. One study of cationic and anionic polymeric nanoparticles with similar particle size (90–100 nm) reported that anionic nanoparticles were transited through the degradative lysosomal pathway within the cell, while the cationic particles were transcytosed and accumulated at the basolateral membrane [84]. On the other hand, positively charged micelles caused higher hemolytic activity—due to strong

electrostatic interactions between the particles and the anionic erythrocyte membrane—and accumulated in the liver proportionally to the positive charge density; in contrast, negatively charged micelles were not hemolytic [87]. Highly charged nanoparticles had a much higher opsonization rate than neutral or slightly charged nanoparticles of the same size [87,88].

### 8.4.3 Crystallinity

Crystallinity of a product is often assessed by X-ray diffraction (XRD), complemented by differential scanning calorimetry (DSC). When a compound is known to form hydrates or solvates, thermogravimetric analysis (TGA) is used as well to determine any weight loss at elevated temperature. These analyses also play a key role in characterizing polymorphic forms of a crystalline formulation. In one study to formulate SN 30191, an anticancer drug (PI3K inhibitor), by high-pressure homogenization, the premilled form II particles were converted to form I when the sample was subjected to a pressure between 500 and 1500 bar [89]. Further increase in pressure to 1750 bar resulted in the formation of a hydrate form. Polymorphic transition can considerably change the solubility and dissolution rate, which may alter a compound's *in vivo* performance, such as its bioavailability.

### 8.4.4 Dissolution

Dissolution behavior of nanocrystals can be determined by the USP dissolution apparatus 2 (paddle) or similar methods. A large volume of dissolution medium (e.g., more than 500 mL) may be needed to minimize the influence of withdrawing samples during the measurement. Careful attention is needed when carrying out the solubility measurement of nanocrystals. It is reported that solubility results determined by centrifugation, ultracentrifugation, or filtration may be questionable because of poor separation of solid particles from the solution [90]. In the presence of high-concentration stabilizers or surfactants, solubility of a drug may be apparently increased, likely due to micellar solubilization of the drug [90]. It must be noted that although the dissolution rate of a nanocrystal formulation can be enhanced considerably because of the increase in surface area, solubility increase is minimal. For instance, the solubility of loviride in a nanosuspension was reportedly increased by 15% when the particle size was about 150 nm [90]. For a poorly soluble drug, the absolute increase in the dissolution rate of its nanocrystal formulation is still marginal.

## 8.5 NANOCRYSTALS FOR CANCER THERAPY

Combinatorial chemistry and high throughput screening have promoted a steady increase in the number of drug candidates over the years but the majority have been poorly soluble [91–93]. Among various formulation strategies, reducing the size of solid particles of a drug compound has gained considerable attention as a delivery

method [73]. Although the marketed nanocrystal-based products are mainly for oral delivery, formulated mostly by the top-down approach, nanocrystals have been exploited for other routes as well, including parenteral, dermal, mucosal, ocular, and pulmonary [17]. Relative to the conventional delivery approaches, nanocrystal formulations are free of organic solvents or other solubilizing chemicals, which may cause adverse effects. The administration of escalated doses of nanocrystals, which can be up to 3–10-fold higher than the maximum tolerated dose of a conventional formulation, may result in improved patient tolerability and enhanced therapeutic efficacy [94,95]. More importantly, nanocrystals offer a much stable and well-integrated platform for drug delivery, circumventing many drawbacks associated with the existing delivery designs, such as drug leakage and systemic breakdown during the storage or even administration stage. Such merits become even more desirable for the intravenous delivery of poorly soluble chemotherapeutic agents. The high drug loading and sustained release of drug molecules allow the nanocrystals to fully take advantage of the EPR effect for targeting tumors. To date, the number of studies reported to use nanocrystals for cancer therapy is limited (Table 8.1). Because of the promising values, more studies are expected to emerge. Below, we briefly review some recent studies that characterize the behaviors and performance of nanocrystal formulations of antineoplastic agents *in vitro* and *in vivo*.

Free drug molecules dissolved from nanocrystals enter cells through passive diffusion. Additionally, nanocrystals can be taken up by cells via active transport process, including the clathrin- and caveolae-mediated endocytosis. Particles of a size of 200 nm or less are preferentially internalized by clathrin-coated pits, while particles between 200 and 500 nm are taken up by the caveolae-mediated pathway [96,97]. The maximum size of particles that can internalized through cell membrane is 500 nm [98]. In one of our studies, the addition of chlorpromazine (CPZ), which disrupts clathrin and hence clathrin-mediated endocytosis, reduced the cellular uptake of camptothecin (CPT) nanocrystals by about 20% compared with that of CPT nanocrystals without the inhibitor [64]. Moreover, nanocrystals are often observed to exert greater cytotoxicity than a solution formulation, particularly when the incubation time is relatively short. One study shows that, against PC-3 prostate cancer cells, the $IC_{50}$ value of deacetymycoepoxydiene (DM) nanocrystals was much lower than that of the drug solution after 12 h incubation; the $IC_{50}$ value of the nanosuspension and solution converged when observed at 48 h [99]. Similar observations were made in our laboratory when $IC_{50}$ values of paclitaxel nanocrystals and its solution against the breast cancer cell MDA-MB-231 were compared (not published).

One of the early *in vivo* studies of nanocrystals was conducted in a murine adenocarcinoma model by testing nanocrystals of piposulfan, camptothecin, etoposide, and paclitaxel that were prepared by ball milling [94]. Solubility values of these drugs range from 200 to $<4$ μg/ml. It was found that, when given intravenously using a multiple-dose regimen at the maximum tolerated dose (MTD), the nanocrystal formulations of the four chemotherapeutic drugs were able to suppress the tumor burden with no death reported, while control formulations using organic solvents and surfactants to solubilize the drugs caused 20% or more mortality. The significant

**TABLE 8.1 Nanocrystal Formulations of Anticancer Drugs**

| Compound | Method | Stabilizer(s) | Average Size (nm) | Zeta Potential (mV) | Reference |
|---|---|---|---|---|---|
| Piposulfan | MM | Tween 80, Span80 | 210.2 ± 38.9 | N/A | [94] |
| Camptothecin | MM | Pluronic F108 | 202.3 ± 30.5 | N/A | [94] |
| Etoposide | MM | Pluronic F127 | 256.2 ± 53.0 | N/A | [94] |
| Paclitaxel | MM | Pluronic F127 | 279.2 ± 29.60 | N/A | [94] |
| Asulacrine | HPH | Poloxamer 188 (Pluronic F68) | 702 ± 0.02(d90) | N/A | [105] |
| SN30191 | HPH | Poloxamer 407, Solutol HS15, Mannitol | 1200–1300 (d90) | N/A | [89] |
| Paclitaxel | 3PNET | Pluronic F127 | 122 ± 35 | 0.82 | [95,118] |
| Oridonin | HPH | Pluronic F68, lecithin, HPMC, PVP | 322.7; 103.3 ± 1.5 and 897.2 ± 14.2; 912.5 ± 17.6 | −26.74 ± 2.68; −20.3 ± 0.4, −21.8 ± 0.8; N/A | [70,106,119] |
| DeacetyMycoepoxydiene | HPH | Lecithin, Pluronic F68, HPMC, PVP | 423 ± 11 | −23.1 ± 3.5 | [99] |
| Quercetin | HPH, EPAS | Pluronic F68, Lecithin | 213.6 ± 29.3 (HPH); 282.6 ± 0.02 (EPAS) | −22.48 ± 4.6 (HPH); −21.12 ± 2.7 (EPAS) | [120] |
| Paclitaxel | SLbL, US | Polyelectrolytes, BSA | 100 ± 20 nm (SLbL); 220 ± 20 (US) | −45 ± 3 | [121] |
| Camptothecin | Anti-solvent | Without stabilizer | 234.4 ± 35.6 | −41.8 ± 1.3 | [64] |
| Paclitaxel | Anti-solvent | Without stabilizer | 438.5 ± 8.3 | −15–22 | [65] |

HPH, high pressure homogenization; MM, media milling; EPAS, evaporative precipitation into aqueous solution; 3PNET, three-phase nanoparticle engineering technology; SLbL, sonication assisted layer-by-layer polyelectrolyte coating; US, ultrasonication.

increase in the MTD of nanocrystals is also reported by a recent study in which paclitaxel nanocrystals were formulated by a three-phase nanoparticle engineering technique and tested in two murine tumor models [95]. The nanocrystals were given at a dose equivalent to the MTD of a Cremophor-EL solubilized dosage form (20 mg/kg), and both formulations showed similar treatment efficacy. The nanocrystals showed much improved therapeutic outcome when administered at 60 mg/kg. The PTX nanosuspensions were also tested by oral administration but found to be less effective than the parenteral dosage form.

Intravenously administered nanocrystals are expected to reach tumors via the EPR effect. We found that camptothecin nanocrystals, administered in a MCF-7 xenografted murine tumor model, yielded significantly better antitumor activity ($p < 0.01$) than the drug salt solution [64]. Biodistribution results at 24 h after the injection further revealed that CPT nanocrystals accumulated in the tumor at a much higher level than the salt solution. Note that a majority of nanocrystals (or any nanoparticle drug delivery systems) are still being taken up by the reticuloendothelial system (RES), mainly liver and spleen [100]. A common practice to reduce the RES uptake and enhance biocompatibility is to modify the surface of a nanosystem is commonly modified, physically or chemically, with biocompatible polymers, such as polyethylene glycol (PEG), that is known to prolong the systemic circulation [101]. Because nanocrystals continue to dissolve after their administration, surface coating by polymeric surfactants onto the nanocrystals may not work. Thus, there is a critical need to investigate new ways to incorporate biocompatible molecules (and other functional molecules) to nanocrystals. Nonetheless, relative to solution-based delivery systems, nanocrystals are expected to accumulate in the tumor by at least an order of magnitude more (from less than one to a few percent) at the same dose level. Given the stable integrity, nanocrystals hold a great promise in targeted cancer therapy.

The particle size of a nanoparticle delivery system has a fundamental influence on the pharmacokinetics (PK) and biodistribution of the nanoparticles. It has been suggested that particles between 50 and 300 nm can achieve prolonged blood circulation [100,102], which helps their accumulation in solid tumors via the EPR effect [100,103,104]. With the same surface charge of −25 mV, polymeric nanoparticles with the size of 150 nm were reported to have significantly longer circulation time due to the low uptake by the RES than particles of 500 nm [85]. Compared with a solution formulation, optimally designed nanovehicles are expected to bear improved PK behaviors, including increased half-life ($t_{½}$), reduced clearance (Cl), increased area under the curve (AUC), and increased mean residence time (MRT) [104]. The PK of intravenously injected nanocrystals of chemotherapeutics (e.g., paclitaxel, asulacrine, and oridonin) was reported to be quite distinctive, best described by a two-compartmental model, where the elimination was nonzero order [105–107]. In a rabbit model, the pharmacokinetic profile of 103 nm oridonin nanocrystals, which were shown to dissolve rapidly, was similar to that of the drug solution [70]. On the contrary, the 300 nm paclitaxel [107] and 700 nm asulacrine nanocrystals [105] were rapidly taken up by the RES organs, such as liver, spleen, and lungs, resulting in shortened distribution half-lives ($t_{½}\alpha$) and decreased AUC. The decreased and yet prolonged AUC in plasma was found to be beneficial to

minimize the toxic drug exposure. Interestingly, the 900 nm oridonin nanocrystals were found to achieve decreased Cl and $t_{1/2}\alpha$ and increased AUC as compared with the dissolved drug [70]. Overall, prolonged elimination half-life ($t_{1/2}\beta$) and MRT were observed for nanocrystals between 300 and 900 nm [105–107]. The sustained blood concentrations likely resulted from slow release of drug molecules from nanocrystals trapped by the RES organs [105,106]. Size has also been found to affect particle accumulation and retention in the tumor. Small particles may accumulate faster in tumors, but larger ones can be retained longer [108]. Moreover, particle accumulation in the tumor relies on the leakiness of the tumor vasculature, which is further determined by the tumor type [109,110], histological grade, and malignant severity [111]. The cutoff sizes of pores in the tumor vasculature range from 200 nm to as large as 1.2 μm, with the majority between 380 and 780 nm [110]. Collectively, it has been suggested that in order to take advantage of the EPR effect and target tumors, nanoparticles should be smaller than 400 nm [108]. The particle size is nonetheless one of many factors that determine the pharmacokinetics and biodistribution of nanoparticles. Surface chemistry plays an important role as well. Because their surface is decorated with hydroxyl groups, amorphous silica particles with sizes of 50 and 250 nm, respectively, exhibited significantly higher blood to liver ratios compared with 180 nm PEG-coated solid lipid nanoparticles in a Wistar rat model [102]. Despite their intrinsic hydrophilicity, which resulted in decreased RES clearance, detailed toxicological studies of amorphous silica particles are still limited [112].

A recent clinical study of Panzem®, a nanocrystal dispersion of 2-methoxyestradiol (2ME2), an endogenous estradiol-17β metabolite that bears antimitotic and antioangiogenic activities, demonstrated an improved pharmacokinetic performance and antitumor activity relative to its capsule formulation when given orally to 16 patients with advanced solid malignancies and 14 patients with taxane-refractory, metastatic castrate-resistant prostate cancer [113,114]. Despite the fact that the 2ME2 nanocrystal formulation was well-tolerated and showed evidence of biologic activity, it did not appear to have sufficient clinical activity and the study was terminated at phase II. No other clinical studies to test parenteral nanocrystal formulations for cancer therapy have been reported.

## 8.6 FUTURE PERSPECTIVES

Nanocrystals have shown promising potentials for drug delivery, particularly for delivering poorly water-soluble anticancer drugs. Major challenges still remain and must be overcome before nanocrystal-based formulation can be fully embraced for treating patients. One such difficulty is production of high-quality nanocrystals with a uniform particle size and morphology. These properties are especially desired for parenteral dosage forms. The top-down approach lacks a well-defined mechanism to control the particle size; the high-energy process may damage the crystallinity of fragmented particles, which presents a concern for the physical stability of a final product. On the other hand, the bottom-up approach stems from solution

crystallization, which has been extensively investigated over decades with regard to mechanism and kinetics. The broad understanding will eventually lead to technical breakthroughs in controlling the particle size, size distribution, and shape of nanocrystals produced from solution. More studies nevertheless remain to be done, particularly for fine-tuning the nucleation process at the industrial scale to ensure uniformity and reproducibility.

Another challenge of using nanocrystals for cancer therapy is seemingly a lack of feasible means to enhance biocompatibility and improve cancer targeting. Tumor-targeting moieties to augment cellular uptake cannot be chemically attached to the crystal surface. Physical adsorption directly or conjugated with surface-coated surfactants may not be suitable either because of continuous dissolution of nanocrystals. Also because of the dynamic nature of dissolution, any surfactants used to prevent nanocrystals from agglomeration and/or to enhance systemic circulation can be easily detached from the nanocrystal surface during blood circulation and extravasation. These limitations seem to drastically dampen the application of nanocrystals for cancer therapy. Similarly, it seems to be difficult, if not impossible, to develop theranostic nanocrystal systems that integrate imaging or diagnostic agents with a drug substance.

So comes the hybrid nanocrystal concept. Over the last few years, our laboratory has developed nanosized hybrid crystals, through which a functional material is physically integrated. The concept was inspired by a well-known phenomenon in solid-state chemistry wherein impurities are entrapped in the crystal lattice of a host particle. The existence of foreign substances in a host is typically insignificant, having little effect on the structure and integrity of the host crystal structure but capable of casting drastic influence on the optical appearance and other physical properties of the host. Examples are abundant. Colored diamonds are one famous example; synthetic hybrid crystals include alloys and semiconductors. The most relevant examples are dyeing crystals, where guest dye molecules are integrated in organic crystals [115]. Hybrid nanocrystals of paclitaxel were developed in our laboratory to integrate fluorescent probes and tested in a murine tumor model [65]. The results indicated that the nanocrystals were capable of not only exerting antitumor effect but permitting real-time imaging of the delivery system in animals. Figure 8.4 shows the hybrid paclitaxel nanocrystals and

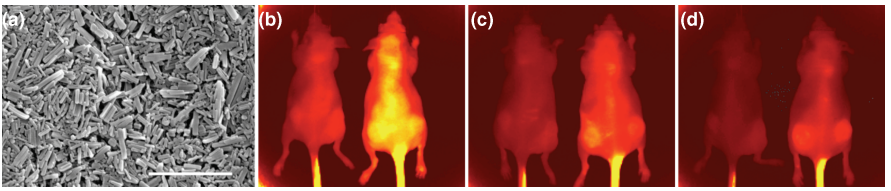


**FIGURE 8.4** SEM image of paclitaxel/fluorophore (FPR-749) nanocrystals (a, scale bar: 2 μm), and fluorescence images obtained by IVIS of two mice bearing MCF-7 xenograft tumors in rear flanks. FPR-749 solution (left) and the hybrid nanocrystals (right) were intravenously injected via tail vein and images were captured at 25 min (b), 4 h (c) and 48 h (d), respectively. Exposure time was 0.25 s, F-stop 2, and binning medium.

whole-body fluorescence images of mice intravenously injected with either dissolved or integrated fluorophore. We are currently expanding the concept to integrate other types of functional compounds, including radioisotopes, ligands, and polymers. Because the guest substance is integrated within a nanocrystal host, hybrid nanocrystals are expected to advance the usage of nanocrystals for cancer therapy.

Finally, although the focus of this chapter is organic nanocrystals, we cannot discount the development and application of inorganic nanocrystals in drug delivery and cancer treatment. Quantum dots (QDs) have been widely used in cellular imaging [116,117]. Other inorganic nanosystems, including iron oxide, silica, gold, and even carbon nanotubes have been explored as drug delivery systems.

## Acknowledgments

The project described was supported by Grant Number R25CA153954 from the National Cancer Institute. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Cancer Institute or the National Institutes of Health.

## REFERENCES

1. Hauss, D. J. (2007). Oral lipid-based formulations. *Advanced Drug Delivery Reviews 59*, 667–676.
2. Singla, A. K., Garg, A., Aggarwal, D. (2002). Paclitaxel and its formulations. *International Journal of Pharmaceutics 235*, 179–192.
3. Rowinsky, E. K., Eisenhauer, E. A., Chaudhry, V., Arbuck, S. G., Donehower, R. C. (1993). Clinical toxicities encountered with paclitaxel (Taxol(R)). *Seminars in Oncology 20*, 1–15.
4. Maeda, H., The enhanced permeability and retention (EPR) effect in tumor vasculature: the key role of tumor-selective macromolecular drug targeting, In *Advances in Enzyme Regulation*, Vol. 41, ed. Weber, G., Pergamon, Great Britain, 2001, 189–207.
5. Maruyama, K. (2011). Intracellular targeting delivery of liposomal drugs to solid tumors based on EPR effects. *Advanced Drug Delivery Reviews 63*, 161–169.
6. Andresen, T. L., Jensen, S. S., Jorgensen, K. (2005). Advanced strategies in liposomal cancer therapy: problems and prospects of active and tumor specific drug release. *Progress in Lipid Research 44*, 68–97.
7. Torchilin, V. P. (2007). Micellar nanocarriers: pharmaceutical perspectives. *Pharmaceutical Research 24*, 1–16.
8. Wong, H. L., Bendayan, R., Rauth, A. M., Li, Y. Q., Wu, X. Y. (2007). Chemotherapy with anticancer drugs encapsulated in solid lipid nanoparticles. *Advanced Drug Delivery Reviews 59*, 491–504.
9. Bummer, P. M. (2004). Physical chemical considerations of lipid-based oral drug delivery—solid lipid nanoparticles. *Critical Reviews in Therapeutic Drug Carrier Systems 21*, 1–19.

10. Tong, R., Cheng, J. J. (2007). Anticancer polymeric nanomedicines. *Polymer Reviews 47*, 345–381.
11. Pridgen, E. M., Langer, R., Farokhzad, O. C. (2007). Biodegradable, polymeric nanoparticle delivery systems for cancer therapy. *Nanomedicine 2*, 669–680.
12. Wolinsky, J. B., Grinstaff, M. W. (2008). Therapeutic and diagnostic applications of dendrimers for cancer treatment. *Advanced Drug Delivery Reviews 60*, 1037–1055.
13. Gulati, M., Grover, M., Singh, S., Singh, M. (1998). Lipophilic drug derivatives in liposomes. *International Journal of Pharmaceutics 165*, 129–168.
14. Mayer, L. D., Krishna, R., Webb, M., Bally, M. (2000). Designing liposomal anticancer drug formulations for specific therapeutic applications. *Journal of Liposome Research 10*, 99–115.
15. Garcia-Fuentes, M., Alonso, M. J., Torres, D. (2005). Design and characterization of a new drug nanocarrier made from solid–liquid lipid mixtures. *Journal of Colloid and Interface Science 285*, 590–598.
16. Ruenraroengsak, P., Cook, J. M., Florence, A. T. (2010). Nanosystem drug targeting: facing up to complex realities. *Journal of Controlled Release 141*, 265–276.
17. Shegokar, R., Müller, R. H. (2010). Nanocrystals: industrially feasible multifunctional formulation technology for poorly soluble actives. *International Journal of Pharmaceutics 399*, 129–139.
18. Müller, R. H., Jacobs, C., Kayser, O. (2001). Nanosuspensions as particulate drug formulations in therapy rationale for development and what we can expect for the future. *Advanced Drug Delivery Reviews 47*, 3–19.
19. Müller, R. H., Moschwitzer, J., Bushrab, F. N., Manufacturing of nanoparticles by milling and homogenization techniques, In *Nanoparticle Technology for Drug Delivery, Drugs, and Pharmaceutical Sciences*, eds. Gupta, R. B., Kompella, U. B., Taylor & Francis Group, LLC, New York, 2006, pp. 21–51.
20. Chan, H. K., Kwok, P. C. L. (2011). Production methods for nanodrug particles using the bottom-up approach. *Advanced Drug Delivery Reviews 63*, 406–416.
21. Rabinow, B. E. (2004). Nanosuspensions in drug delivery. *Nature Reviews Drug Discovery 3*, 785–796.
22. Liversidge, G. G., Cundy, K. C., Bishop, J. F., Czekai, D. A. Surface modified drug nanoparticles. U.S. Patent 5,145,684. 1991.
23. Junghanns, J., Müller, R. H. (2008). Nanocrystal technology, drug delivery and clinical applications. *International Journal of Nanomedicine 3*, 295–309.
24. Merisko-Liversidge, E., Liversidge, G. G., Cooper, E. R. (2003). Nanosizing: a formulation approach for poorly-water-soluble compounds. *European Journal of Pharmaceutical Sciences 18*, 113–120.
25. Haynes, D. H. Phospholipid-coated microcrystals: injectable formulations of water insoluble drugs. U.S. Patent 5,091,187. 1992.
26. Keck, C. M., Müller, R. H. (2006). Drug nanocrystals of poorly soluble drugs produced by high pressure homogenisation. *European Journal of Pharmaceutics and Biopharmaceutics 62*, 3–16.
27. Müller, R. H., Becker, R., Kruss, K. P., Peters, K. Pharmaceutical nanosuspensions for medicament administration as systems with increased saturation solubility and rate of solution. U.S. Patent 5,858,410. 1999.

28. Illig, K. J., Mueller, R. L., Ostrander, K. D., Swanson, J. R. (1996). Use of microfluidizer processing for preparation of pharmaceutical suspensions. *Pharmaceutical Technology 20*, 78–88.
29. Patravale, V. B., Date, A. A., Kulkarni, R. M. (2004). Nanosuspensions: a promising drug delivery strategy. *Journal of Pharmacy and Pharmacology 56*, 827–840.
30. Begat, P., Young, P. M., Edge, S., Kaerger, J. S., Price, R. (2003). The effect of mechanical processing on surface stability of pharmaceutical powders: visualization by atomic force microscopy. *Journal of Pharmaceutical Sciences 92*, 611–620.
31. Panagiotou, T., Fisher, R. J. (2008). Form nanoparticles via controlled crystallization. *Chemical Engineering Progress 104*, 33–39.
32. Becker, R., Döring, W. (1935). Kinetische behandlung der keimbildung in übersättingten dämpfen. *Annals of Physics 24*, 719–752.
33. Kashchiev, D., van Rosmalen, G. M. (2003). Review: nucleation in solutions revisited. *Crystal Research and Technology 38*, 555–574.
34. Volmer, M., *Kinetik der Phasenbildung*, Steinkopf., Dresden, 1939.
35. Mullin, J. W., *Crystallization*, Butterworth-Heinemann, Oxford, 2001.
36. Horn, D., Rieger, J. (2001). Organic nanoparticles in the aqueous phase—theory, experiment, and use. *Angewandte Chemie-International Edition 40*, 4331–4361.
37. Oxtoby, D. W. (1998). Nucleation of first-order phase transitions. *Accounts of Chemical Research 31*, 91–97.
38. Dirksen, J. A., Ring, T. A. (1991). Fundamentals of crystallization—kinetic effects on particle-size distributions and morphology. *Chemical Engineering Science 46*, 2389–2427.
39. Chen, J. F., Wang, Y. H., Guo, F., Wang, X. M., Zheng, C. (2000). Synthesis of nanoparticles with novel technology: high-gravity reactive precipitation. *Industrial & Engineering Chemistry Research 39*, 948–954.
40. Hu, T. T., Wang, J. X., Shen, Z. G., Chen, J. F. (2008). Engineering of drug nanoparticles by HGCP for pharmaceutical applications. *Particuology 6*, 239–251.
41. Panagiotou, T., Mesite, S. V., Fisher, R. J. (2009). Production of norfloxacin nanosuspensions using microfluidics reaction technology through solvent/antisolvent crystallization. *Industrial & Engineering Chemistry Research 48*, 1761–1771.
42. Byrappa, K., Ohara, S., Adschiri, T. (2008). Nanoparticles synthesis using supercritical fluid technology—towards biomedical applications. *Advanced Drug Delivery Reviews 60*, 299–327.
43. Luque de Castro, M. D., Priego-Capote, F. (2007). Ultrasound-assisted crystallization (sonocrystallization). *Ultrasonics Sonochemistry 14*, 717–724.
44. List, M., Sucker, H. Pharmaceutical colloidal hydrosols for injection. G.B. Patent 2,200,048. 1988.
45. Gassmann, P., Sucker, H. Improvements in pharmaceutical compositions. WO 018105. 1992.
46. List, M., Sucker, H. Hydrosols of pharmacologically active agents and their pharmaceutical compositions comprising them. U.S. Patent 5,389,382. 1995.
47. Gassmann, P. S., H. Pharmaceutical compositions comprised of stabilized peptide particles. U.S. Patent 6,447,806. 2002.

48. Zhong, J., Shen, Z. G., Yang, Y., Chen, J. F. (2005). Preparation and characterization of uniform nanosized cephradine by combination of reactive precipitation and liquid antisolvent precipitation under high gravity environment. *International Journal of Pharmaceutics 301*, 286–293.
49. Zhao, H., Wang, J. X., Wang, Q. A., Chen, J. F., Yun, J. (2007). Controlled liquid antisolvent precipitation of hydrophobic pharmaceutical nanoparticles in a microchannel reactor. *Industrial & Engineering Chemistry Research 46*, 8229–8235.
50. Chiou, H., Li, L., Hu, T. T., Chan, H. K., Chen, J. F., Yun, J. (2007). Production of salbutamol sulfate for inhalation by high-gravity controlled antisolvent precipitation. *International Journal of Pharmaceutics 331*, 93–98.
51. Hu, T., Chiou, H., Chan, H. K., Chen, J. F., Yun, J. (2008). Preparation of inhalable salbutamol sulphate using reactive high gravity controlled precipitation. *Journal of Pharmaceutical Sciences 97*, 944–949.
52. Chiou, H., Chan, H. K., Heng, D., Prud'homme, R. K., Raper, J. A. (2008). A novel production method for inhalable cyclosporine A powders by confined liquid impinging jet precipitation. *Journal of Aerosol Science 39*, 500–509.
53. Chiou, H., Chan, H. K., Prud'homme, R. K., Raper, J. A. (2008). Evaluation on the use of confined liquid impinging jets for the synthesis of nanodrug particles. *Drug Development and Industrial Pharmacy 34*, 59–64.
54. Kumar, V., Adamson, D. H., Prud'homme, R. K. (2010). Fluorescent polymeric nanoparticles: aggregation and phase behavior of pyrene and amphotericin B molecules in nanoparticle cores. *Small 6*, 2907–2914.
55. Rogers, T. L., Johnston, K. P., Williams, R. O. (2001). Solution-based particle formation of pharmaceutical powders by supercritical or compressed fluid $CO_2$ and cryogenic spray-freezing technologies. *Drug Development and Industrial Pharmacy 27*, 1003–1015.
56. Reverchon, E., De Marco, I., Torino, E. (2007). Nanoparticles production by supercritical antisolvent precipitation: a general interpretation. *Journal of Supercritical Fluids 43*, 126–138.
57. Bustami, R. T., Chan, H. K., Dehghani, F., Foster, N. R. (2000). Generation of microparticles of proteins for aerosol delivery using high pressure modified carbon dioxide. *Pharmaceutical Research 17*, 1360–1366.
58. Lancaster, R. W., Singh, H., Theophilus, A. L. Apparatus and process for preparing crystalline particles. WO 0038811. 2000.
59. Abbas, A., Srour, M., Tang, P., Chiouc, H., Chan, H. K., Romagnoli, J. A. (2007). Sonocrystallisation of sodium chloride particles for inhalation. *Chemical Engineering Science 62*, 2445–2453.
60. Dhumal, R. S., Biradar, S. V., Paradkar, A. R., York, P. (2009). Particle engineering using sonocrystallization: salbutamol sulphate for pulmonary delivery. *International Journal of Pharmaceutics 368*, 129–137.
61. Kaerger, J. S., Price, R. (2004). Processing of spherical crystalline particles via a novel solution atomization and crystallization by sonication (SAXS) technique. *Pharmaceutical Research 21*, 372–381.
62. Dhumal, R. S., Biradar, S. V., Yamamura, S., Paradkar, A. R., York, P. (2008). Preparation of amorphous cefuroxime axetil nanoparticles by sonoprecipitation for enhancement of bioavailability. *European Journal of Pharmaceutics and Biopharmaceutics 70*, 109–115.

63. Kipp, J. E., Wong, J. C. T., Doty, M. J., Rebbeck, C. L. Microprecipitation method for preparing submicron suspensions. U.S. Patent 6,869,617. 2001.
64. Zhang, H., Hollis, C. P., Zhang, Q., Li, T. L. (2011). Preparation and antitumor study of camptothecin nanocrystals. *International Journal of Pharmaceutics 415*, 293–300.
65. Zhao, R. S., Hollis, C. P., Zhang, H., Sun, L. L., Gemeinhart, R. A., Li, T. L. (2011). Hybrid nanocrystals: achieving concurrent therapeutic and bioimaging functionalities toward solid tumors. *Molecular Pharmaceutics 8*, 1985–1991.
66. Van Eerdenbrugh, B., Van den Mooter, G., Augustijns, P. (2008). Top-down production of drug nanocrystals: nanosuspension stabilization, miniaturization and transformation into solid products. *International Journal of Pharmaceutics 364*, 64–75.
67. Kocbek, P., Baumgartner, S., Kristl, J. (2006). Preparation and evaluation of nanosuspensions for enhancing the dissolution of poorly soluble drugs. *International Journal of Pharmaceutics 312*, 179–186.
68. Lee, J., Lee, S. J., Choi, J. Y., Yoo, J. Y., Ahn, C. H. (2005). Amphiphilic amino acid copolymers as stabilizers for the preparation of nanocrystal dispersion. *European Journal of Pharmaceutical Sciences 24*, 441–449.
69. Deng, Z., Xu, S., Li, S. (2008). Understanding a relaxation behavior in a nanoparticle suspension for drug delivery applications. *International Journal of Pharmaceutics 351*, 236–243.
70. Gao, L., Zhang, D. R., Chen, M. H., Zheng, T. T., Wang, S. M. (2007). Preparation and characterization of an oridonin nanosuspension for solubility and dissolution velocity enhancement. *Drug Development and Industrial Pharmacy 33*, 1332–1339.
71. Van Eerdenbrugh, B., Vermant, J., Martens, J. A., Froyen, L., Van Humbeeck, J., Augustijns, P., Van Den Mooter, G. (2009). A screening study of surface stabilization during the production of drug nanocrystals. *Journal of Pharmaceutical Sciences 98*, 2091–2103.
72. Yaguchi, S. I., Fukui, Y., Koshimizu, K., Yoshimi, H., Matsuno, T., Gouda, H., Hirono, S., Yamazaki, K., Yamori, T. (2006). Antitumor activity of ZSTK474, a new phosphatidylitiositol 3-kinase inhibitor. *Journal of the National Cancer Institute 98*, 545–556.
73. Kesisoglou, F., Panmai, S., Wu, Y. H. (2007). Nanosizing—oral formulation development and biopharmaceutical evaluation. *Advanced Drug Delivery Reviews 59*, 631–644.
74. Na, G. C., Stevens, H. J., Yuan, B. O., Rajagopalan, N. (1999). Physical stability of ethyl diatrizoate nanocrystalline suspension in steam sterilization. *Pharmaceutical Research 16*, 569–574.
75. Zheng, J. Y., Bosch, H. W. (1997). Sterile filtration of NanoCrystal(TM) drug formulations. *Drug Development and Industrial Pharmacy 23*, 1087–1093.
76. Wu, L. B., Zhang, J., Watanabe, W. (2011). Physical and chemical stability of drug nanoparticles. *Advanced Drug Delivery Reviews 63*, 456–469.
77. Matsuyama, T., Yamamoto, H. (2004). Particle shape and laser diffraction: a discussion of the particle shape problem. *Journal of Dispersion Science and Technology 25*, 409–416.
78. Muhlenweg, H., Hirleman, E. D. (1998). Laser diffraction spectroscopy: influence of particle shape and a shape adaptation technique. *Particle & Particle Systems Characterization 15*, 163–169.

79. Endoh, S., Kuga, Y., Ohya, H., Ikeda, C., Iwata, H. (1998). Shape estimation of anisometric particles using size measurement techniques. *Particle & Particle Systems Characterization 15*, 145–149.
80. Stevens, N., Shrimpton, J., Palmer, M., Prime, D., Johal, B. (2007). Accuracy assessments for laser diffraction measurements of pharmaceutical lactose. *Measurement Science & Technology 18*, 3697–3706.
81. Buckton, G., *Interfacial Phenomena in Drug Delivery and Targeting*, Harwood Academic Publishers, Switzerland, 1995, 133–134.
82. Jacobs, C., Müller, R. H. (2002). Production and characterization of a budesonide nanosuspension for pulmonary administration. *Pharmaceutical Research 19*, 189–194.
83. Chung, T. H., Wu, S. H., Yao, M., Lu, C. W., Lin, Y. S., Hung, Y., Mou, C. Y., Chen, Y. C., Huang, D. M. (2007). The effect of surface charge on the uptake and biological function of mesoporous silica nanoparticles 3T3-L1 cells and human mesenchymal stem cells. *Biomaterials 28*, 2959–2966.
84. Harush-Frenkel, O., Rozentur, E., Benita, S., Altschuler, Y. (2008). Surface charge of nanoparticles determines their endocytic and transcytotic pathway in polarized MDCK cells. *Biomacromolecules 9*, 435–443.
85. He, C. B., Hu, Y. P., Yin, L. C., Tang, C., Yin, C. H. (2010). Effects of particle size and surface charge on cellular uptake and biodistribution of polymeric nanoparticles. *Biomaterials 31*, 3657–3666.
86. Juliano, R. L., Stamp, D. (1975). Effect of particle-size and charge on clearance rates of liposomes and liposome encapsulated drugs. *Biochemical and Biophysical Research Communications 63*, 651–658.
87. Xiao, K., Li, Y. P., Luo, J. T., Lee, J. S., Xiao, W. W., Gonik, A. M., Agarwal, R. G., Lam, K. S. (2011). The effect of surface charge on in vivo biodistribution of PEG-oligocholic acid based micellar nanoparticles. *Biomaterials 32*, 3435–3446.
88. Dobrovolskaia, M. A., Aggarwal, P., Hall, J. B., McNeil, S. E. (2008). Preclinical studies to understand nanoparticle interaction with the immune system and its potential effects on nanoparticle biodistribution. *Molecular Pharmaceutics 5*, 487–495.
89. Sharma, P., Zujovic, Z. D., Bowmaker, G. A., Denny, W. A., Garg, S. (2011). Evaluation of a crystalline nanosuspension: polymorphism, process induced transformation and in vivo studies. *International Journal of Pharmaceutics 408*, 138–151.
90. Van Eerdenbrugh, B., Vermant, J., Martens, J. A., Froyen, L., Van Humbeeck, J., Van den Monter, G., Augustijns, P. (2010). Solubility increases associated with crystalline drug nanoparticles: methodologies and significance. *Molecular Pharmaceutics 7*, 1858–1870.
91. Gardner, C. R., Walsh, C. T., Almarsson, O. (2004). Drugs as materials: valuing physical form in drug discovery. *Nature Reviews Drug Discovery 3*, 926–934.
92. Lipinski, C. A. (2000). Drug-like properties and the causes of poor solubility and poor permeability. *Journal of Pharmacological and Toxicological Methods 44*, 235–249.
93. Lipinski, C. A., Lombardo, F., Dominy, B. W., Feeney, P. J. (1997). Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Advanced Drug Delivery Reviews 23*, 3–25.
94. MeriskoLiversidge, E., Sarpotdar, P., Bruno, J., Hajj, S., Wei, L., Peltier, N., Rake, J., Shaw, J. M., Pugh, S., Polin, L., Jones, J., Corbett, T., Cooper, E., Liversidge, G. G.

(1996). Formulation and antitumor activity evaluation of nanocrystalline suspensions of poorly soluble anticancer drugs. *Pharmaceutical Research 13*, 272–278.

95. Liu, F., Park, J. Y., Zhang, Y., Conwell, C., Liu, Y., Bathula, S. R., Huang, L. (2010). Targeted cancer therapy with novel high drug-loading nanocrystals. *Journal of Pharmaceutical Sciences 99*, 3542–3551.
96. Rejman, J., Oberle, V., Zuhorn, I. S., Hoekstra, D. (2004). Size-dependent internalization of particles via the pathways of clathrin-and caveolae-mediated endocytosis. *Biochemical Journal 377*, 159–169.
97. Zuhorn, I. S., Kalicharan, R., Hoekstra, D. (2002). Lipoplex-mediated transfection of mammalian cells occurs through the cholesterol-dependent clathrin-mediated pathway of endocytosis. *Journal of Biological Chemistry 277*, 18021–18028.
98. Yu, X., Valmikinathan, C. M., Rogers, A., Wang, J., Nanotechnology and drug delivery, In *Biomedical Nanostructures*, eds. Gonsalves, K. E., Halberstadt, C. R., Laurencin, C. T., Wiley, Hoboken, NJ, 2007, 93–113.
99. Wang, Y. C., Liu, Z. P., Zhang, D. R., Gao, X. H., Zhang, X. Y., Duan, C. X., Jia, L. J., Feng, F. F., Huang, Y. J., Shen, Y. M., Zhang, Q. A. (2011). Development and in vitro evaluation of deacety mycoepoxydiene nanosuspension. *Colloids and Surfaces B-Biointerfaces 83*, 189–197.
100. Moghimi, S. M., Hunter, A. C., Murray, J. C. (2001). Long-circulating and target-specific nanoparticles: theory to practice. *Pharmacological Reviews 53*, 283–318.
101. Shenoy, D., Little, S., Langer, R., Amiji, M. (2005). Poly(ethylene oxide)-modified poly (beta-amino ester) nanoparticles as a pH-sensitive system for tumor-targeted delivery of hydrophobic drugs: Part 2. In vivo distribution and tumor localization studies. *Pharmaceutical Research 22*, 2107–2114.
102. Barbe, C., Bartlett, J., Kong, L. G., Finnie, K., Lin, H. Q., Larkin, M., Calleja, S., Bush, A., Calleja, G. (2004). Silica particles: a novel drug-delivery system. *Advanced Materials 16*, 1959–1966.
103. Alexis, F., Pridgen, E., Molnar, L. K., Farokhzad, O. C. (2008). Factors affecting the clearance and biodistribution of polymeric nanoparticles. *Molecular Pharmaceutics 5*, 505–515.
104. Li, S. D., Huang, L. (2008). Pharmacokinetics and biodistribution of nanoparticles. *Molecular Pharmaceutics 5*, 496–504.
105. Ganta, S., Paxton, J. W., Baguley, B. C., Garg, S. (2009). Formulation and pharmacokinetic evaluation of an asulacrine nanocrystalline suspension for intravenous delivery. *International Journal of Pharmaceutics 367*, 179–186.
106. Gao, L., Zhang, D. R., Chen, M. H., Duan, C. X., Dai, W. T., Jia, L. J., Zhao, W. F. (2008). Studies on pharmacokinetics and tissue distribution of oridonin nanosuspensions. *International Journal of Pharmaceutics 355*, 321–327.
107. Wang, Y. L., Li, X. M., Wang, L. Y., Xu, Y. L., Cheng, X. D., Wei, P. (2011). Formulation and pharmacokinetic evaluation of a paclitaxel nanosuspension for intravenous delivery. *International Journal of Nanomedicine 6*, 1497–1507.
108. Bae, Y. H., Park, K. (2011). Targeted drug delivery to tumors: myths, reality and possibility. *Journal of Controlled Release 153*, 198–205.
109. Hashizume, H., Baluk, P., Morikawa, S., McLean, J. W., Thurston, G., Roberge, S., Jain, R. K., McDonald, D. M. (2000). Openings between defective endothelial cells explain tumor vessel leakiness. *American Journal of Pathology 156*, 1363–1380.

110. Hobbs, S. K., Monsky, W. L., Yuan, F., Roberts, W. G., Griffith, L., Torchilin, V. P., Jain, R. K. (1998). Regulation of transport pathways in tumor vessels: role of tumor type and microenvironment. *Proceedings of the National Academy of Sciences of the United States of America 95*, 4607–4612.

111. Daldrup, H., Shames, D. M., Wendland, M., Okuhata, Y., Link, T. M., Rosenau, W., Lu, Y., Brasch, R. C. (1998). Correlation of dynamic contrast-enhanced MR imaging with histologic tumor grade: comparison of macromolecular and small-molecular contrast media. *American Journal of Roentgenology 171*, 941–949.

112. Fadeel, B., Garcia-Bennett, A. E. (2010). Better safe than sorry: understanding the toxicological properties of inorganic nanoparticles manufactured for biomedical applications. *Advanced Drug Delivery Reviews 62*, 362–374.

113. Tevaarwerk, A. J., Holen, K. D., Alberti, D. B., Sidor, C., Arnott, J., Quon, C., Wilding, G., Liu, G. (2009). Phase I trial of 2-methoxyestradiol NanoCrystal dispersion in advanced solid malignancies. *Clinical Cancer Research 15*, 1460–1465.

114. Harrison, M. R., Hahn, N. M., Pili, R., Oh, W. K., Hammers, H., Sweeney, C., Kim, K., Perlman, S., Arnott, J., Sidor, C., Wilding, G., Liu, G. (2011). A phase II study of 2-methoxyestradiol (2ME2) NanoCrystalA (R) dispersion (NCD) in patients with taxane-refractory, metastatic castrate-resistant prostate cancer (CRPC). *Investigational New Drugs 29*, 1465–1474.

115. Kahr, B., Gurney, R. W. (2001). Dyeing crystals. *Chemical Reviews 101*, 893–951.

116. Smith, A. M., Gao, X. H., Nie, S. M. (2004). Quantum dot nanocrystals for in vivo molecular and cellular imaging. *Photochemistry and Photobiology 80*, 377–385.

117. Smith, A. M., Nie, S. M. (2010). Semiconductor nanocrystals: structure, properties, and band gap engineering. *Accounts of Chemical Research 43*, 190–200.

118. Deng, J. X., Huang, L., Liu, F. (2010). Understanding the structure and stability of paclitaxel nanocrystals. *International Journal of Pharmaceutics 390*, 242–249.

119. Zhang, Z., Zhang, X. M., Xue, W., YangYang, Y. N., Xu, D. R., Zhao, Y. X., Lou, H. Y. (2010). Effects of oridonin nanosuspension on cell proliferation and apoptosis of human prostatic carcinoma PC-3 cell line. *International Journal of Nanomedicine 5*, 735–742.

120. Gao, L., Liu, G. Y., Wang, X. Q., Liu, F., Xu, Y. F., Ma, J. (2011). Preparation of a chemically stable quercetin formulation using nanosuspension technology. *International Journal of Pharmaceutics 404*, 231–237.

121. Pattekari, P., Zheng, Z., Zhang, X., Levchenko, T., Torchilin, V., Lvov, Y. (2011). Top-down and bottom-up approaches in production of aqueous nanocolloids of low solubility drug paclitaxel. *Physical Chemistry Chemical Physics 13*, 9014–9019.

# CLEARANCE OF NANOPARTICLES DURING CIRCULATION

Seung-Young Lee and Ji-Xin Cheng
*Weldon School of Biomedical Engineering, Purdue University, West Lafayette, IN, USA*

## 9.1 INTRODUCTION

Administration of conventional therapeutic agents for cancer therapy is often limited due to their chemical properties, such as low solubility and poor stability in blood. These properties cause a short circulation half-life and low therapeutic efficacy, thus necessitating frequent administrations. In addition, toxic side effects of these agents due to nonspecific biodistribution have been additional issues in cancer chemotherapy. To overcome the limitations of conventional therapeutic agents, researchers have explored drug delivery systems that can safely and efficiently carry drugs and deliver them to the tumor site without affecting normal tissues [1–3]. In this regard, nanoparticle systems have gained a lot of interest because of the drug loading capacity and the ability to easily pass through anatomical barriers. Delivery of nanoparticles to solid tumors is facilitated by the leakiness of tumor blood vessels [4,5], and the discovery of various tumor-specific ligands has enabled more efficient delivery of nanoparticles to cancer cells [6,7].

One of the key factors to determine the targeting efficacy of nanoparticles is their blood-circulation time because long-circulating nanoparticles have more chances to transport into tumor tissues [8,9]. The circulation half-life of the nanoparticles primarily depends on the rate of the biological clearance [10,11]. The clearance mechanisms can be classified into three categories: (i) disintegration of nanoparticles by protein adsorption, (ii) opsonization-mediated nanoparticle removal by immune

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*,
Edited by Shayne C. Gad
Copyright © 2013 John Wiley & Sons, Inc.

cells, and (iii) filtration by organs with fenestrated vasculature. Biological clearance of nanoparticles is influenced by physiochemical properties of the nanoparticles, such as surface charge, composition, size, shape, and deformability [12–14]. To extend the circulation time of nanoparticles in the blood stream, nanoparticles are modified in various ways.

## 9.2 NANOPARTICLE-BASED DRUG CARRIERS IN CANCER THERAPY

### 9.2.1 Nanoparticle-Based Drug Carriers

Nanoparticles with a size ranging from 1 to 1000 nm have been considered promising drug carriers. Their distinct advantages include flexible control of particle size and surface properties, the compatibility with various routes of administration (e.g., topical, enteral, and parenteral), and the ability to protect anticancer drugs against clearance in the blood and to provide controlled drug release such as sustained or pulsatile drug release [15–18].

Nanoparticle-based drug carriers were first introduced in 1950–1960s as liposomes and polymer–drug conjugates [19,20]. Since then, various nanoparticle-based drug carriers have been developed for the enhancement of chemotherapy [21,22]. In particular, advancement in nanotechnology and biomaterials has contributed to the improvement of existing nanoparticle systems and the development of new drug delivery systems (Table 9.1) [11,23–43]. These nanoparticle-based drug carriers help enhance drug localization in tumor tissues and reduce cytotoxicity for normal tissues compared to free drug.

### 9.2.2 Tumor Targeting Strategies

An ideal nanoparticle-based drug carrier for chemotherapy should first guide a drug to the targeted tumor tissues while preserving therapeutic activity of the drug during circulation. Secondly, after reaching tumor tissues, nanoparticles should release the drug to kill tumor tissues selectively and site specifically, reducing collateral damages to normal tissues [44–46]. Currently, passive and active tumor targeting strategies are used to deliver nanoparticles to the tumor sites.

***9.2.2.1 Passive Tumor Targeting.*** Most solid tumors possess leaky vasculature induced by vascular permeability factors such as bradykinin, nitric oxide (NO), peroxynitrite ($ONOO^-$), prostaglandins, vascular endothelial growth factor (VEGF), matrix metalloproteinases (MMPs), kallikrein, and tumor necrosis factor (TNF) [4,47]. High concentrations of these secreted factors around tumor tissues accelerate the formation of new blood vessels (neovascularization or angiogenesis) to supply oxygen and nutrients for tumor growth. This results in hyperpermeable vasculature, which allows macromolecules or nanoparticles to access tumor tissue. In addition, many solid tumors are known to develop a defective lymphatic drainage system, causing the molecules and nanoparticles to be retained there [48]. The cut-off size of

**TABLE 9.1 Nanoparticle-Based Drug Delivery Systems**

| System | Structure | Structure Illustration | Example of Compounds | Reference |
|---|---|---|---|---|
| Polymeric nanoparticles | Hydrophobic moieties or polymers are conjugated to hydrophilic backbone polymer and form self-aggregation | | PEG–PLA/PLGA<br>LMWC–PLGA<br>GC–CA<br>HA–CA | [23]<br>[24]<br>[11]<br>[25] |
| Polymersome | Amphiphilic block copolymers assemble and form a polymer bilayer shell | | PEO–PEE<br>PEG–PLA<br>PEO–PBO | [26]<br>[27]<br>[28] |
| Polymeric micelles | Amphiphilic block copolymers assemble into small spherical particles of a hydrophobic core and hydrophilic shell | | PEG–PLA<br>PEG–PCL<br>PEG–PLGA<br>PEO–PPO–PEO | [29]<br>[30]<br>[31]<br>[32] |
| Dendrimer | Well-defined branched synthetic polymer with a spherical three-dimensional shape | | PEG–PAMAM<br>PGLSA<br>Polypeptide | [33]<br>[34]<br>[35] |
| Liposome | Amphiphilic lipids assemble into spherical particles with lipid bilayer shell | | PC–cholesterol<br>PEG–DSPE | [36]<br>[37] |
| Inorganic nanoparticles | Metal-based nanoparticles with polymer surface modification | | Gold<br>Iron oxide | [38]<br>[39] |
| Protein nanoparticles | Protein-based formulation using the interaction with drugs | | Albumin | [40] |
| Hybrid nanoparticles | At least two different materials are used to form spherical particles composed of a core and shell | | Albumin–PMMA<br>PLGA–QD–siRNA<br>PLGA-dendrimer | [41]<br>[42]<br>[43] |

Hydrophobic polymer; hydrophilic polymer. Lipid; drug.
PEG, poly(ethylene glycol); PLA, poly(lactic acid); PLGA, poly(lactic-*co*-glycolic acid); LMWC, low molecular-weight chitosan; GC, glycol chitosan; CA, cholanic acid; HA, hyaluronic acid; PEE, poly(ethyl ethylene); PEO, poly(ethylene oxide); PBO, poly(butylene oxide); PCL, poly(caprolactone); PPO, poly (propylene oxide); PAMAM, poly(amido amine); PGLSA, poly(glycerol succinic acid); PC, phosphatidylcholine; DSPE, distearoylphosphatidylethanolamine; PMMA, poly(methyl methacrylate); Dox, doxorubicin; QD, quantum dot.

gap junctions between endothelial cells in permeable tumor blood vessels is around 200–800 nm [5], which enables accumulation of nanoparticles having the similar size. This phenomenon is known as the enhanced permeability and retention (EPR) effect, a critical principle for the passive tumor targeting of nanoparticles [49]. After accumulation in the tumor tissues, the drug is released into extracellular matrices, where it diffuses into intracellular sites.

Most nanoparticle-based drug carriers rely entirely on this EPR mechanism to facilitate delivery to the tumor. Despite a wealth of documentations on the EPR effect, tumor targeting efficacy through the EPR effect remains controversial. This controversy stems from the fact that there is no major driving force facilitating the passive extravasation of nanoparticles through the leaky vasculature [50]. It is currently believed that passive nanoparticle extravasation via the EPR effect occurs by diffusion and convection in blood flow. However, these forces actually play a minor role in the transport of nanoparticles to the interstitium of tumors compared to bulk fluid flow [51]. Interstitial fluid pressure in tumor is higher than intravascular pressure due to the defective lymphatic drainage [52]. Furthermore, blood pressure in solid tumors tends to increase toward the tumor center, which drives overall mass transport of fluid away from the center [53,54]. As a result, nanoparticle transport to central tumor regions is remarkably limited. A recent report describes that this problem was experimentally addressed by increasing systemic blood pressure with angiotensin II (AT-II) [55] or facilitating tumor blood flow using nitroglycerin (NG) [56]. This approach resulted in temporary augmentation of the EPR effect and drug delivery to tumor tissues [57].

***9.2.2.2 Active Tumor Targeting.*** Another strategy for tumor targeted drug delivery is active targeting using tumor-binding ligands [58,59]. Here nanoparticles are decorated with a ligand as a targeting moiety on the surface, which helps their entry into the cells. Cellular internalization of these nanoparticles occurs via receptor-mediated endocytosis [60,61] and is especially important in biotherapeutic delivery such as DNA and small interfering RNA (siRNA) [62]. Various targeting ligands for cancer therapy, such as RGD peptides for targeting $\alpha_v\beta_3$ integrins in tumor vasculature and folic acid for targeting folate receptors on cancer cells, have been identified and used for targeted drug delivery (Table 9.2) [63–71]. However, the mechanism of active targeting nanoparticles is not entirely clear [72], and the effectiveness of ligands remains controversial. Several recent reports have shown only a modest improvement in tumor accumulation of the ligand-decorated nanoparticles compared to nanoparticles with no ligands [73,74]. The results of these studies suggest that the tumor localization of nanoparticles is independent of the ligand and primarily relies on the EPR effect.

## 9.3 BIOLOGICAL CLEARANCE OF NANOPARTICLES

When the nanoparticle delivery to tumor tissues largely depends on the EPR effect, circulation time of nanoparticles in the blood is one of the most important properties [8]. The long circulation time provides nanoparticles with a great opportunity to transport

**TABLE 9.2 Tumor Molecules and Ligands for Active Targeting**

| Target Site | Target Molecule (or Receptor) | Ligand | Reference |
|---|---|---|---|
| Tumor vascular targeting | VEGF receptor | CBO-P11 | [63] |
| | $\alpha_v\beta_3$ | cRGD peptide | [64] |
| | VCAM-1 | Anti-VCAM-1 antibody | [65] |
| | MT1-MMP | Anti-MT1-MMP antibody | [66] |
| Tumor cell targeting | EGF receptor | EGF | [67] |
| | Transferrin receptor | Transferrin | [68] |
| | Folate receptor | Folic acid | [69] |
| | PSMA | Anti-PSMA aptamer | [70] |
| | ASGP receptor | Carbohydrate | [71] |

VEGF, vascular endothelial growth factor; cRGD, cyclic arginine–glycine–aspartic acid; VCAM-1, vascular cell adhesion molecule-1; EGF, epidermal growth factor; PSMA, prostate specific membrane antigen; ASGP, asialoglycoprotein.

into tumor tissues through the leaky vasculature and interact with target receptors on tumor cells [75]. The circulation time of the nanoparticles in the blood is determined by the rate of biological clearance from the blood, spleen, liver, and kidney (Figure 9.1) [10,76]. While biological clearance is a natural protection mechanism against invasion by foreign organisms, it is a significant obstacle for the application of nanoparticles in cancer therapy. In addition, the localization of drugs or nanoparticles in filtering organs such as the liver and the spleen can cause hypersensitivity reactions such as exaggerated immune responses, making the concept of targeted drug delivery to eliminate secondary toxicity pointless [77–80]. To evade rapid biological clearance, various approaches have been developed focusing on surface modification or controlling the physiochemical properties (size, shape, surface charge, and deformability) of nanoparticles, which play critical roles in their biological clearance [81–84].

### 9.3.1 Blood Clearance

***9.3.1.1 Protein Adsorption and Opsonization.*** Plasma proteins immediately cover the surface of administered nanoparticles, which they recognize as foreign materials [85,86]. Mechanisms of plasma protein binding to nanoparticles are not yet fully understood due to the complexity of the binding events between the proteins and nanoparticles. Proteins associated with various nanoparticles have been determined by protein separation and identification techniques such as centrifugation, gel filtration, two-dimensional polyacrylamide gel electrophoresis (2D-PAGE), and atomic force microscopy [87–89]. The composition of proteins adsorbed to the nanoparticles primarily depends on the protein concentration, kinetics of protein binding, and surface characteristics of the nanoparticles (Table 9.3) [90–104]. Abundant plasma proteins such as albumin, immunoglobulin G (IgG), fibrinogen, and apolipoproteins are first adsorbed on most types of nanoparticles. Over time, these proteins are sequentially exchanged with other

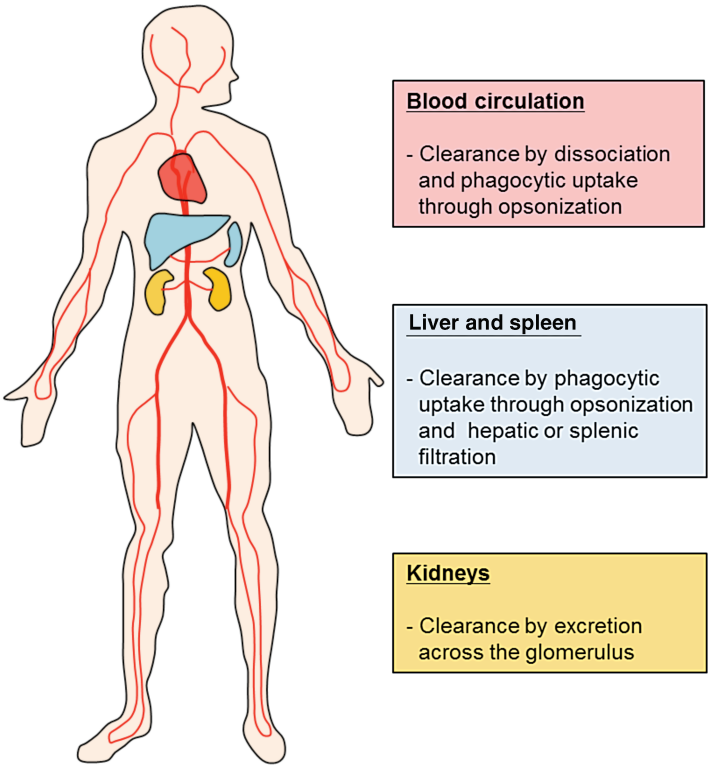


**FIGURE 9.1** Biological clearance of nanoparticles in blood circulation, liver, spleen, and kidneys. (*Source:* Modified from Reference [10].)

high affinity proteins. The exchange rates are relative to the chemical properties and stability of the protein [105]. For instance, fibrinogen binds immediately onto the surface of solid lipid nanoparticles and is rapidly displaced by apolipoproteins after 0.5 min incubation in blood [106]. Protein binding to iron oxide particles also changes with the plasma concentration and incubation time [107].

Physiochemical properties of nanoparticles that influence protein adsorption to their surface include surface charge, hydrophobicity, particle size, surface curvature, and surface area. Electrically neutral nanoparticles have distinctly less protein adsorption than charged nanoparticles [108–110], whereas protein adsorption increases in general for negatively charged nanoparticles [111]. The composition of adsorbed proteins depends on the functional groups and its charge. For example, proteins with an isoelectric point below 5.5 (e.g., albumin) have affinity for positively charged nanoparticles with amine groups ($–NH_2$), while those with isoelectric point above 5.5 (e.g., IgG) was shown to be predominantly adsorbed to negatively charged nanoparticles with carboxyl acid (–COOH) [112]. Macrophage uptake of positively charged nanoparticles was greatly increased by albumin adsorption [113]. Hydrophobicity of a nanoparticle surface is another key factor

**TABLE 9.3 Nanoparticle Systems and Opsonins**

| System | Identified Oposnins | Reference |
|---|---|---|
| Polystyrene with poloxamer 184, 188, 407 | Factor B, transferrin, albumin, fibrinogen, IgG, apolipoproteins | [92] |
| Liposomes | Albumin, fibrinogen, apolipoproteins, IgG, α1-antitrypsin, α2-macroglobulin, IgM | [93] |
| Solid lipid nanoparticles with Tween 80 | Fibrinogen, IgG, IgM, apolipoproteins (including ApoE), transthyretin | [94] |
| Solid lipid nanoparticles with poloxamer 188 | Fibrinogen, IgG, IgM, apolipoproteins (excluding ApoE), transthyretin, albumin | [94] |
| Poly(lactic acid) nanoparticles with PEG | Albumin, fibrinogen, apolipoproteins, IgG | [95] |
| Poly(hexadecyl cyanoacrylate) nanoparticles | Albumin, apolipoproteins, IgG, transferrin | [96] |
| Poly(ε-caprolactone) nanoparticles | IgG, apolipoproteins | [97] |
| Polycyanoacrylate nanoparticles | Albumin, IgG, IgM, fibrinogen, apolipoproteins | [98] |
| Iron oxide nanoparticles | Albumin, IgG, IgM, fibrinogen | [99] |
| Poly(D,L-lactic acid) nanoparticles | Albumin, IgG, fibrinogen, IgM, apolipoproteins, antithrombin III | [100] |
| Polystyrene with Rhodamine B | Albumin, IgG, fibrinogen, apolipoproteins, PLS:6 | [101] |
| Poly(isobutylcyanoacrylate) with Dextran | Albumin, IgG, fibrinogen, apolipoproteins, serotransferrin, transthyretin | [102] |
| Poly(butyl cyanoacrylate) nanoparticles with polysorbate 80 | Albumin, IgG, fibrinogen, IgM, apolipoproteins | [103] |
| Solid lipid nanoparticles with poloxamer or poloxamine coating | Albumin, fibrinogen, apolipoproteins | [104] |

PEG, poly(ethylene glycol); IgG, immunoglobulin G; IgM, immunoglobulin M; PLS, unidentified protein.
*Source:* Modified from Reference [87].

in plasma protein adsorption [108]. Higher hydrophobicity of the nanoparticle surface resulted in greater and faster protein binding [114–116]. The effects of size, surface curvature, and surface area of the nanoparticles on protein adsorption have also been evaluated. A study using *N*-isopropylacrylamide (NIPAM) and *N*-*tert*-butylacrylamide (BAM) copolymer nanoparticles with sizes ranging from 70 to 700 nm found the same amount of protein adsorption [117]. In contrast, another study on poly methoxypolyethyleneglycol cyanoacrylate-*co*-*n*-hexadecyl cyanoacrylate (mPEG–PHDCA) copolymer nanoparticles with three different sizes (80, 170, and 240 nm) demonstrated that larger nanoparticles had greater amounts of protein binding. The amount of proteins adsorbed to nanoparticles with a particle

size of 80, 170, and 240 nm was 6, 23, and 34% of serum protein, respectively [118]. These conflicting results may be ascribed to the difference in material systems. Alternatively, change in one property of nanoparticles may have been accompanied by inadvertent change in other properties, which can influence the protein adsorption profile.

***9.3.1.2 Disintegration of Nanoparticles by Protein Adsorption.*** Protein adsorption can trigger the disintegration of nanoparticles, compromising their function as drug carriers. The instability of polymeric micelles by plasma protein binding has been investigated [119]. The disintegration of poly(ethylene glycol)–poly(caprolactone) (PEG–PCL) micelles in serum solution was shown using fluorescence self-quenching method [120]. *In vitro* and *in vivo* disintegration of poly(ethylene glycol)–poly(DL-lactic acid) (PEG–PDLLA) micelles by α- and β-globulins was also shown using the fluorescence resonance energy transfer (FRET) technique [121,122]. There are three possible mechanisms associated with the disintegration of micelles by plasma protein: protein adsorption, protein penetration, and drug extraction by proteins (Figure 9.2) [123]. Unstable nanoparticles may be quickly eliminated in systemic circulation by protein-mediated disintegration [121].

***9.3.1.3 Opsonization-Mediated Nanoparticle Uptake by Immune Cells.*** Another main clearance pathway of nanoparticles is the uptake of nanoparticles by immune

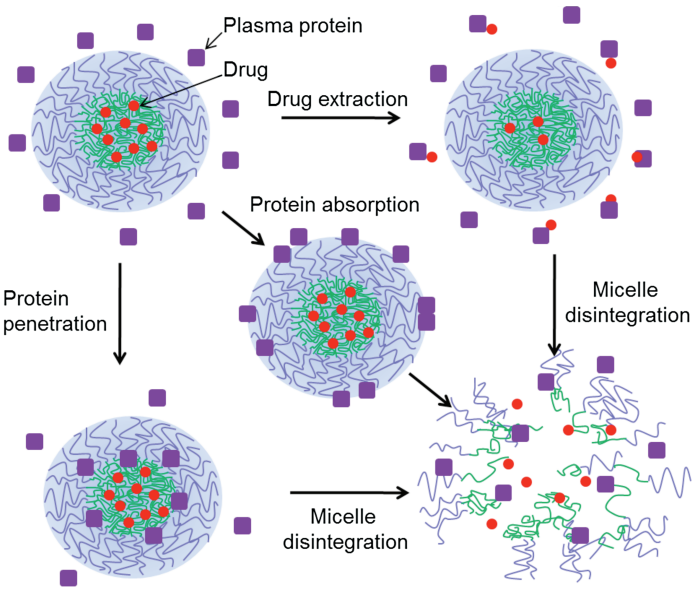


**FIGURE 9.2** Disintegration of nanoparticles by plasma protein; protein adsorption, protein penetration, and drug extraction by proteins (*Source:* Modified from Reference [123].)

cells. Immune cells both in the blood stream (leukocytes, monocytes, platelets, and dendritic cells) and in organs (Kupffer cells in liver, macrophage and B cells in spleen, and dendritic cells in lymph nodes) engulf and remove the nanoparticles from the blood circulation. This process is generally mediated by binding of specific receptor proteins called opsonins to the surface of nanoparticles [124]. The bound opsonins, such as plasma proteins, immunoglobulins, and component factors in complement system, facilitate the nanoparticle uptake into immune cells, negatively affecting the biodistribution of nanoparticles and hence drug delivery to target tumor tissues [125,126]. The opsonins on the nanoparticle undergo conformational changes and are recognized by the immune cells via specialized receptors [108].

Macrophages are mainly responsible for the clearance of large molecules like nanoparticles. The receptor-mediated phagocytic pathways are classified into four categories (mannose receptor-, complement receptor-, immunoglobulin Fcγ receptor-, and scavenger receptor- mediated phagocytosis) according to the opsonins adsorbed onto the nanoparticles [124]. Through these mechanisms, opsonins such as IgG, apolipoproteins, and fibrinogen promote phagocytosis of nanoparticles by macrophages in the reticuloendothelial system (RES) [106,107,127]. These nanoparticles are subject to localization in RES organs such as the liver and spleen, and eventually eliminated by resident immune cells [113,128,129]. Adsorption of fetuin and fibronectin triggers the phagocytosis of polystyrene nanoparticles by Kupffer cells via the scavenger receptor [130,131]. Increasing the amount of C3 and IgG adsorbed on lecithin-coated polystyrene nanoparticles caused an increased uptake of Kupffer cells [132]. Preopsonization of poloxamine-coated polystyrene nanoparticles by incubation in serum solution accelerated the phagocytosis by macrophages [133,134]. Furthermore, during the clearance process, immune cells produce proinflammatory responses, which involve the secretion of cytokines such as tumor necrosis factor-α, interleuckin-12, and interferon-γ. These cytokines recruit more immune cells to take up nanoparticles. This can cause immunotoxicity and severe damages to the RES organs [77]. However, the effect of nanoparticle opsonization is not always destructive. Some studies have shown that opsonins can enhance tumor targeting and blood circulation time of nanoparticles. For example, dysopsonins, such as albumin, extended the retention time of nanoparticles in the blood [135,136]. In addition, various apolipoproteins enabled nanoparticles to transport across the blood–brain barrier (BBB) for brain tumor therapy [137].

### 9.3.2 Organ Filtrations

Some organs and tissues have naturally permeable vasculature in order to conduct their physiological functions. The inherent gaps between endothelial cells in vessels of those organs, called fenestration (Figure 9.3) [138], play a major role in determining the biological fate of nanoparticles.

While the fenestration of nanoparticles can be used for tumor targeting similar to the EPR effect, it can also cause rapid clearance and nontargeting effect of

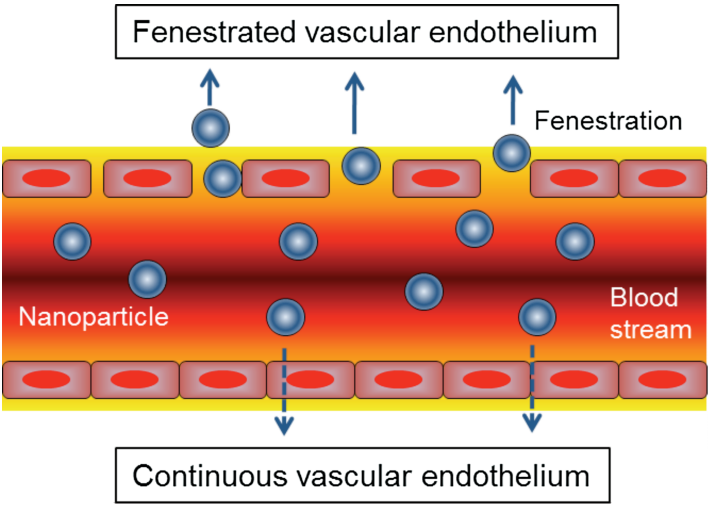


**FIGURE 9.3** Schematic illustration of vascular extravasations of nanoparticles in either continuous or fenestrated tissues.

nanoparticles. The size and structure of a fenestration depend on the physiological functions of the organs and tissues (Table 9.4) [5,47,139–150]. Furthermore, the size of a fenestration can change under various pathological conditions [151]. In the spleen, nanoparticles should pass through splenic fenestrations (splenic sinuses) to remain in blood circulation, while, in the liver, the nanoparticles should avoid going through hepatic fenestrations (hepatic sinuses). Otherwise, nanoparticles may be captured by the hepatocyte tissues (hepatocytes cluster), where the nanoparticles are eliminated by Kupffer cells and hepatocytes. Therefore, to improve the blood circulation and tumor targeting of nanoparticles, an understanding of anatomical fenestrations is essential [138]. Of many organs and tissues that contain

**TABLE 9.4 Sizes of Fenestration in Different Organs and Tissues**

| Organ or Tissue | Fenestration Size | Animal Model | Reference |
|---|---|---|---|
| Spleen | 150 nm | Mice | [139,140] |
| Liver | 100–150 nm | Rat | [141,142] |
| Kidney | 4.5–5 nm | Rat | [143] |
| Lung | 1–400 nm | Dog | [144] |
| Bone marrow | 85–150 nm | Guinea-pig, rabbit, rat | [145,146] |
| Skeletal, cardiac, and smooth muscle | <6 nm | Mice | [147] |
| Skin, subcutaneous, and mucous membrane | <6 nm | Mice | [147] |
| Blood–brain barrier (BBB) | No fenestrations | — | [148,149] |
| Tumor | 200–800 nm | Mice | [5,47] |
| Inflamed organs | 80 nm–1.4 μm | Hamster | [150] |

*Source:* Modified from Reference [138].

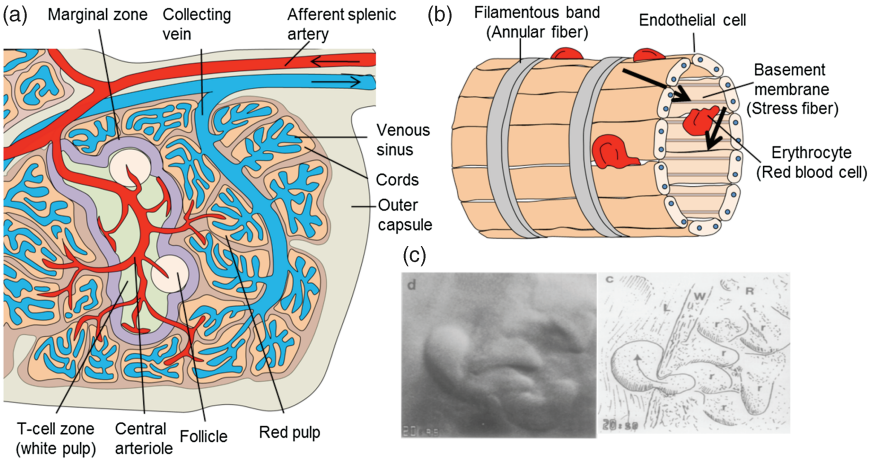


**FIGURE 9.4** Structure of spleen. (a) Schema of the spleen, (b) venous sinuses in the red pulp of the spleen (*Source:* Modified from References [154 and 155]), (c) photomicrograph of an erythrocyte squeezing through sinus wall (*Source:* Reproduced with permission from Reference [155]).

fenestrations, we focus on the organs mainly responsible for the clearance of nanoparticles, such as spleen, liver, and kidneys.

***9.3.2.1 Spleen.*** The main function of the spleen is to filter the blood [152,153]. The spleen possesses tree-shaped branching arterial vessels that are linked to a venous sinusoidal system (Figure 9.4a) [154,155]. The frame of the spleen is mainly composed of fibrous connective tissue called trabecula. The smaller arterial vessels are surrounded by the lymphoid tissues, which form the white pulp of the spleen. Some of the smallest arterial branches end in the marginal sinus located between the white pulp and the marginal zone, while other branches cross the marginal zone to form the venous system of the red pulp, where mechanical blood filtration occurs [154]. To efficiently filter antigens, microorganisms, and old red blood cells (erythrocytes) from the blood, the venous system of the red pulp has a unique microanatomical structure. Arterial vessels eventually reach into cords in the red pulp, where they are structured as an open blood system without an endothelial barrier [156]. A large number of splenic macrophages are resident in the cords to remove unfiltered organisms. After the blood is purified in the venous sinuses of the red pulp, it is collected in the efferent veins, and finally leaves the spleen through the collecting veins.

To transport through the venous sinuses, red blood cells must pass through the sinus wall. The sinus wall consists of a lining of endothelial cells with basement membrane (stress fiber) and filamentous bands (annular fiber). The rod-shaped endothelial cells are positioned in parallel to the longitudinal axis of the sinus and connect to each other by the stress fiber running along the cellular axis. The annular fiber composed of extracellular matrix components connects to the

basement membrane along the interendothelial slits and forms the endothelial cells into tubular pattern (Figure 9.4b) [157,158]. The unique arrangement of the endothelial cells lining the sinus wall forces the blood to pass through the slits to travel from the cords into the sinus [155]. The slits seldom exceed 0.2–0.5 μm in width [157]. However, these narrow slits do not hinder the passage of healthy red blood cells, whose diameters range from 6 to 8 μm, because of their high degree of deformability, whereas unhealthy blood cells having rigid membranes get stuck in these slits and are ultimately eliminated by the splenic macrophages in the cords (Figure 9.4c) [155]. Large and stiff nanoparticles when injected into the blood stream can also be removed by the same splenic filtration process. Although the recognition of pathogens by macrophages occurs via specific opsonin receptors as discussed earlier, the physical filtration system of the spleen potentially helps splenic macrophages to recognize pathogens stuck in sinus slits [152]. Therefore, physiochemical properties of nanoparticles such as size, flexibility, shape, and surface properties influence passage across the sinus slits in the spleen. It has been reported that splenic clearance of polystyrene nanoparticles increased with their size [159,160]. Also, rigid chitosan nanoparticles with the highest degree of substitution (DS) of 5β-cholanic acid showed rapid accumulation in the mouse spleen via the filtration mechanism, compared to softer chitosan nanoparticles with low DS of 5β-cholanic acid [11]. These studies indicate that the physical properties of nanoparticles such as size and deformability play a pivotal role in splenic passage.

***9.3.2.2 Liver.*** The liver has multiple functions including metabolic control, nutrient (e.g., glycogen) storage, plasma protein synthesis, hormone production, and the elimination of foreign or toxic substances (detoxification) [161]. Here, we focus on the detoxification process in the liver that directly relates to the clearance of nanoparticles. The liver receives blood from both the hepatic portal vein and the hepatic arteries. Around 75% of the liver's blood supply is obtained through the hepatic portal vein that drains the blood from the small intestine, stomach, spleen, and pancreas. The remaining 25% of the blood supply comes from the hepatic arteries. The hepatic portal vein and the hepatic arteries branch out in the liver and are combined at the entrance to the hepatic sinusoids [161]. The blood then passes through the sinusoids and collects in the central vein that drains into the hepatic vein (Figure 9.5a) [162].

The basic functional unit of the liver is the hepatic lobule (Figure 9.5b), which consists of a concentric arrangement of hepatocytes (hepatic cords) radiating outward from the central vein and those separated by the sinusoids. The portal triads include a bile duct, the terminal branches of the hepatic artery, and portal vein, which are regularly located at the vertices of the lobule. The sinusoids are lined by fenestrated endothelial cells and have no recognizable basement membrane. These unique fenestrations allow hepatocytes to have direct exposure to the blood in the space of Disse. The space of Disse generally contains extracellular matrix proteins (e.g., reticular fibers), fibroblasts (e.g., hepatic stellate cells), and dendritic cells. Hepatic macrophages (Kupffer cells) reside in the sinusoidal endothelium [162].

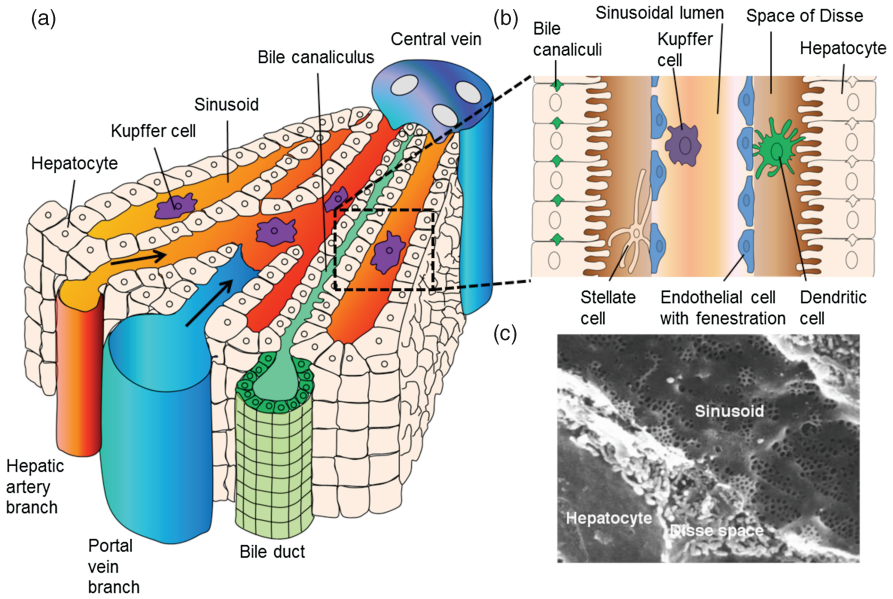


**FIGURE 9.5** Structure of a liver lobule. (a) Schema of the liver lobule, (b) sinusoid in the liver lobule (*Source:* Modified from Reference [162]), (c) scanning electron micrograph (SEM) of the sinusoid (*Source:* Professor Robin Fraser, University of Otago, New Zealand; reproduced per the copyright release to the public domain).

During the detoxification process, Kupffer cells exclusively phagocytose large particles (e.g., 800 nm in diameter), whereas hepatocytes mainly endocytose small particles (e.g., 50 nm in diameter) transported through the endothelial fenestrations [163,164]. Particles taken up by hepatocytes are excreted through the bile by biliary excretion [165]. The size of the fenestration for rats is reported to be around 100 nm in mean diameter under physiological conditions, and the fenestrations are clustered in groups of 10–100 fenestration occupying 2–20% of the sinusoidal endothelial cell surfaces (Figure 9.5c) [142,166,167]. These fenestrations provide a sieving function of sinusoidal blood. Since the liver is a key organ that stores nutrients and removes toxic substances, the internal blood flow rate in the liver is quite slow. This enhances the absorption of plasma nutrients and foreign organisms by the hepatocytes inside the fenestration [161].

This mechanism also contributes to the clearance of small nanoparticles that can pass through the hepatic fenestrations. Most small sized nanoparticles (<100 nm in diameter) have shown substantial accumulation in the liver. Examples are shell cross-linked PEG–poly(*t*-butyl acrylate) nanoparticles (18–37 nm in diameter) [168], DNA–PEI complex with crystallized ligand shell (<50 nm in diameter) [169], PEGylated iron oxide (SPIO) (6.2 nm in average diameter) [170], cyclodextrin–camptothecin nanoparticles (30–40 nm in diameter) [171], core cross-linked *N*-isopropylacrylamide (NIPAm) nanoparticles (around 50 nm in diameter) [172], and quantum dots (3.4–5.5 nm in diameter) [173]. Hepatotoxicity testing

of silica nanoparticles with different sizes (70, 300, and 1000 nm in diameter) found that only the small nanoparticles (70 nm) caused hepatic toxicity [174]. Another study on the biodistribution of PEGylated polystyrene nanoparticles (50 and 140 nm in diameter) in mice demonstrated that most small nanoparticles were found in the cytoplasm of hepatocytes after 24 h post injection, whereas relatively large nanoparticles mainly internalized into Kupffer cells [175]. Recently, a study exhibited silica nanoparticles (~20 nm in diameter) were excreted through the fecal matter (via the liver and the bile), indicating that these nanoparticles had passed through the hepatic fenestrations and endocytosized by hepatocytes [176]. It is therefore evident that relatively small nanoparticles tend to be mechanically trapped in the liver through the sinusoidal fenestrations and preferentially removed by hepatocytes rather than Kupffer cells located on the sinusoidal endothelium [177].

***9.3.2.3 Kidneys.*** The kidneys play major role in homeostasis by maintaining the purity and constancy of internal fluids in the body. The unique filtration system of the kidneys allows wastes and excess ions to be excreted by urine, while essential substances are reabsorbed to the blood [161]. The kidneys cleanse approximately one-quarter of total blood supply in the body each minute. The blood enters the kidneys through the renal artery, which divides into several branches called interlobar arteries. Then, at the medulla-cortex border, the interlobar arteries branch into the arcuate arteries, which curve over the medullary pyramids. Smaller interlobar arteries branch off the arcuate arteries, which supply the cortex at the outer region of the kidneys. Once cleansed, the venous blood collects and leaves the kidneys through the renal vein [161]. Nephrons are the basic functional units of the kidneys responsible for eliminating toxins, wastes, and foreign materials from the body via urine. Each nephron consists of two main structures: a glomerulus and a renal tubule (Figure 9.6) [161]. The closed end of the renal tubule is cup-shaped and completely covers the glomerulus, which is called the Bowman's capsule. Entangled podocytes in the glomerulus form porous membranes called filtration slits. This slits allow small wastes to transport from the plasma to the renal tubule.

The renal filtration is a nonselective, passive, and mechanical process. Thus, glomerular filtration of an object is highly dependent on its size. The filtrate follows the renal tubule, reaches the bladder, and is eventually excreted as urine. Useful substances in the filtrates are reabsorbed by the tubule. The size of the renal filtration slits in rats is around 39 nm [179]. However, the actual slit size is about 4.5–5 nm due to the stratified nature of the glomerular capillary walls [143]. The filtration size threshold of PAMMA dendrimer for mice has been reported to be around 5.4 nm [180,181]. While quantum dots (QDs) with a diameter of 4.36–5.52 nm were rapidly cleared by renal clearance in mice [173,182], QDs >8 nm in diameter did not show the renal clearance but exhibited substantial accumulation in the liver. Glomerular filtration is also influenced by other properties such as charge, shape, and deformability of filtrate. For example, anionic glycolated Fab fragments showed significantly less renal filtration compared to cationic glycolated Fab fragments owing to the electrostatic repulsion between the anionic Fab fragments and the anionic

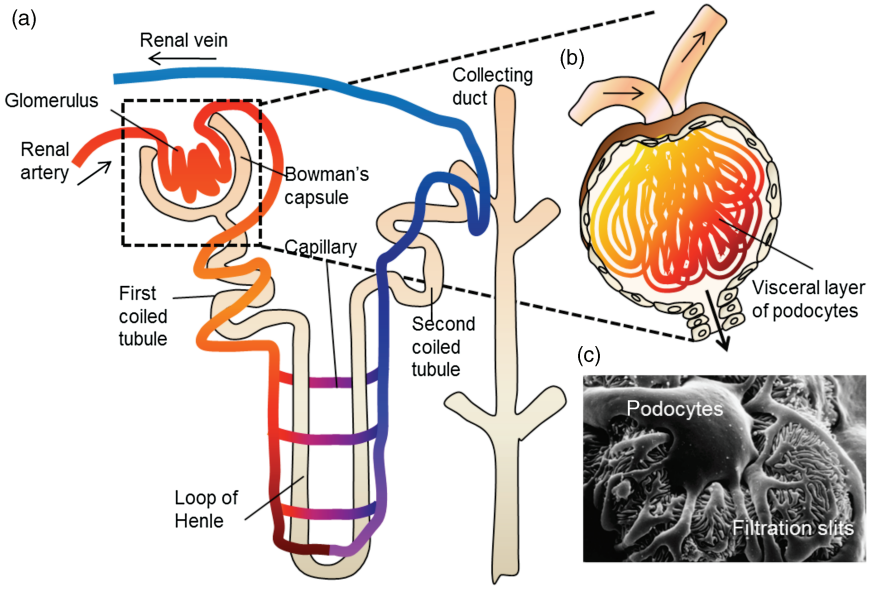


**FIGURE 9.6** Structure of a nephron. (a) Schema of the nephron, (b) glomerulus in the nephron (*Source:* Modified from Reference [161]), (c) SEM of the podocytes in the glomerulus (*Source:* Reproduced with permission from Reference [178]).

membrane of the glomerular capillary walls [183]. Opsonization stimulated by the surface charge of nanoparticles increases the size, which precludes renal filtration [173]. On the other hand, flexible Ficoll and asymmetrically shaped dextran demonstrated higher renal filtration efficacy than globular proteins, even though their sizes (4.6 and 5.5 nm, respectively) are relatively larger than that of the proteins (3.75 nm) [184]. Although these studies used biological molecules, the results are applicable for nanoparticles in renal clearance.

## 9.4 STRATEGIES TO MINIMIZE THE BIOLOGICAL CLEARANCE OF NANOPARTICLES

Since the emergence of nanoparticles in the 1950s, various strategies to minimize their biological clearance have been developed. Some of them are highlighted in the following discussion.

### 9.4.1 Chemical Cross-Linking

As we have discussed earlier in Section 9.3.1.2, polymeric micelles are unstable under physiological conditions. To solve the problem, various covalent cross-linking methods have been introduced such as photo cross-linking [185] and disulfide cross-linking [186].

There are generally three cross-linkable locations for a triblock copolymer micelle: shell, interface, and core [187–190]. However, to maintain the surface properties and increase drug loading capacity of the micelle, cross-linking is typically done at the interfacial positions [191,192]. Another type of cross-linked nanoparticle is a nano-sized hydrogel (e.g., Pluronic polymer), of which stability can be readily increased by manipulating the cross-linking degree of the nanogel matrix [193,194].

### 9.4.2 Surface Modification

The surface properties of nanoparticles greatly influence their degree of opsonization and biodistribution. For example, neutral or weakly positive nanoparticles are shown to evade opsonization [81,108]. Also, the hydrophilic surface of nanoparticles helps them to escape opsonin-mediated macrophage capture [195,196]. PEG, a nonionic hydrophilic polymer, is widely used to increase hydrophilicity of nanoparticles. Brush configurations of PEG on nanoparticle surfaces reduce protein adsorption. The resistance to protein adsorption greatly increases the blood-circulation time of PEGlyated nanoparticles [9,12,197,198]. However, proteins like albumin, α-chymotrypsin, and hen-egg-white lysozyme can adsorb to the PEG-coated surfaces by forming hydrogen bonding interactions between their carboxyl groups and oxygen atoms in PEG [199–201]. Despite this protein adsorption, PEG surface coating is still the most widely accepted method to reduce opsonization onto nanoparticles. On the other hand, albumin-coated liposomes have demonstrated more prolonged circulation time in blood compared to PEG-coated liposomes, indicating that albumin coating can prevent the adsorption of other proteins more efficiently [202]. Therefore, various albumin-coated nanoparticles have been developed to extend their circulation time in blood [203,204].

### 9.4.3 Control of the Size, Shape, and Deformability of Nanoparticles

The size of nanoparticles is a key attribute that can be manipulated in order to reduce the biological clearance associated with anatomical fenestrations discussed earlier. Although the vascular heterogeneity across different animal models needs to be considered, the diameter range 100–200 nm has been considered an appropriate size in rodent models for both long-term circulation and passive accumulation in tumors [22]. Other properties of nanoparticles such as shape and deformability (flexibility) are also believed to play a role [83,84]. Recent advances in nanotechnology such as the template technology have enabled the development of particles with various sizes and shapes [205,206]. The template technology, which utilizes a template for making homogenous nanoparticles in size and shape, has been used to investigate the effect of particle shape on the phagocytosis by macrophages [207,208]. Macrophages contacting elliptical shaped particles along their major axis rapidly phagocytosed the particles within 6 min, while macrophages that initially contacted the same particles along their minor axis did not internalize the particles. This result suggests that macrophage uptake depends on the particle shape upon initial contact. Another

example pertaining to the effect of nanoparticle shape on biological clearance is filamentous micelles (filomicelles) that have two-dimensional shape (~18 μm in length and ~20–60 nm in diameter). These filamentous micelles exhibited a significantly prolonged blood circulation time (circulation half-lives: ~5 days) and improved drug delivery [30,209]. This shape not only facilitates micelles passage through RES organs such as the spleen and liver but also minimizes their internalization into macrophages. In addition, red blood cells (RBCs) mimicking polymeric particles have recently been engineered using a poly(lactic-*co*-glycolic acid) (PLGA) RBC-shaped template [210]. Although the deformability of these particles was less than that of mouse red blood cells, it was still more flexible than spherical PLGA particles due to their special shape. These particles also demonstrated oxygen and drug carrying capabilities [210]. Another recent study on the deformability of particles showed that micro-sized *N*-isopropylacrylamide (NIPAm) hydrogel particles (~1 μm in diameter) were able to pass through membrane pores at least 10 times smaller in size [211].

It was also reported that the difference in deformability of nanoparticles affected how the nanoparticles were internalized into macrophages [212]. In this study, soft nanoparticles were preferentially internalized via macropinocytosis while rigid nanoparticle uptake was found to be mediated by clathrin-mediated routes. As for nanoparticles with intermediate flexibility, they were internalized via multiple mechanisms. Furthermore, soft chitosan nanoparticles have shown prolonged blood circulation and higher tumor targeting efficacy compared to rigid chitosan nanoparticles [11]. Chitosan nanoparticles with higher degree of substitution of 5β-cholanic acid have demonstrated lower deformability and, thus, higher accumulation in the liver and spleen even though they had a similar particle size (300–400 nm).

## 9.5 FUTURE PERSPECTIVES

We have discussed mechanisms by which nanoparticles are biologically cleared from the body and current strategies to enhance the blood-circulation and tumor targeting effect of nanoparticles. The surface properties of nanoparticles are indirectly related to their blood clearance through plasma protein adsorption and opsonization, while physical properties such as size, shape, and deformability directly impact the removal of nanoparticles via the filtering organs. To extend the retention time of nanoparticles in blood circulation, a number of strategies manipulating the physiochemical properties of nanoparticles have been developed. Nevertheless, the fraction of nanoparticles that actually reach the target tissues still remains very small (~5%), and the majority of administered nanoparticles are found in the RES organs [213].

To overcome this critical challenge, it is necessary to revisit nanoparticle-based delivery strategies and key factors in the design of nanoparticles. One of the most important prerequisites in the design of a nanoparticle system is to stably retain the loaded drug in blood circulation before accessing cancer cells. Disintegration of nanoparticles by plasma protein adsorption is the first clearance mechanism encountered by the systemically administered nanoparticles. A vast majority of nanoparticle

studies focus on postextravasation events such as cellular uptake and intracellular trafficking, but what is the use if most drug has leached out of the system and blank nanoparticles finally reach the targets? Without solving the first problem, targeted drug delivery strategies will have difficulty in taking the next step.

In this regard, the absence of appropriate techniques to investigate whether the nanoparticles remained intact after systemic administration is the major obstacle in the development of new nanoparticle systems. Optical imaging of fluorescently labeled nanoparticles can provide information on pharmacokinetics and biodistribution of the nanoparticles; however, from this technique it is not obvious whether the fluorescent signals come from the intact nanoparticles loaded with drug or disintegrated pieces of nanoparticles. Fluorescence imaging of labeled nanoparticles does not provide sufficient evidence that the nanoparticle has been successfully delivered to the tumor. In an effort to overcome these limitations, our group has designed and synthesized disulfide cross-linked nanoparticles [214]. Using FRET imaging, we monitor the intactness of these nanoparticles in blood vessels and tissues spatiotemporally. The FRET imaging has allowed us to prove that our nanoparticles could stably retain anticancer drugs in bloodstream and increase its blood half-life and delivery to tumors.

## REFERENCES

1. Ashley, C. E., Carnes, E. C., Phillips, G. K., Padilla, D., Durfee, P. N., Brown, P. A., et al. (2011). The targeted delivery of multicomponent cargos to cancer cells by nanoporous particle-supported lipid bilayers. *Nature Materials 10*, 389–397.
2. Cabral, H., Matsumoto, Y., Mizuno, K., Chen, Q., Murakami, M., Kimura, M., et al. (2011). Accumulation of sub-100 nm polymeric micelles in poorly permeable tumours depends on size. *Nature Nanotechnology 6*, 815–823.
3. von Maltzahn, G., Park, J. -H., Lin, K. Y., Singh, N., Schwöppe, C., Mesters, R., et al. (2011). Nanoparticles that communicate in vivo to amplify tumour targeting. *Nature Materials 10*, 545–552.
4. Matsumura, Y., Maeda, H. (1986). A new concept for macromolecular therapeutics in cancer chemotherapy: mechanism of tumoritropic accumulation of proteins and the antitumor agent smancs. *Cancer Research 46*, 6387–6392.
5. Torchilin, V. (2011). Tumor delivery of macromolecular drugs based on the EPR effect. *Advanced Drug Delivery Reviews 63*, 131–135.
6. Byrne, J. D., Betancourt, T., Brannon-Peppas, L. (2008). Active targeting schemes for nanoparticle systems in cancer therapeutics. *Advanced Drug Delivery Reviews 60*, 1615–1626.
7. Wang, J., Tian, S., Petros, R. A., Napier, M. E., DeSimone, J. M. (2010). The complex role of multivalency in nanoparticles targeting the transferrin receptor for cancer therapies. *Journal of the American Chemical Society 132*, 11306–11313.
8. Moghimi, S. M., Hunter, A. C., Murray, J. C. (2001). Long-circulating and target-specific nanoparticles: theory to practice. *Pharmacological Reviews 53*, 283–318.

9. Lee, S., Lee, S. -Y., Park, S., Ryu, J. H., Na, J. H., Koo, H., et al. (2012). In vivo NIRF imaging of tumor targetability of nanosized liposomes in tumor-bearing mice. *Macromolecular Bioscience 12*, 849–856.
10. Alexis, F., Pridgen, E., Molnar, L. K., Farokhzad, O. C. (2008). Factors affecting the clearance and biodistribution of polymeric nanoparticles. *Molecular Pharmaceutics 5*, 505–515.
11. Na, J. H., Lee, S. -Y., Lee, S., Koo, H., Min, K. H., Jeong, S. Y., et al. (2012). Effect of the stability and deformability of self-assembled glycol chitosan nanoparticles on tumor-targeting efficiency. *Journal of Controlled Release 163*, 2–9.
12. Perry, J., Reuter, K. G., Kai, M. P., Herlihy, K. P., Jones, S. W., Luft, J. C., et al. (2012). PEGylated PRINT nanoparticles: the impact of PEG density on protein binding, macrophage association, biodistribution, and pharmacokinetics. *Nano Letters 12*, 5304–5310.
13. Meng, H., Xue, M., Xia, T., Ji, Z., Tarn, D. Y., Zink, J. I., et al. (2011). Use of size and a copolymer design feature to improve the biodistribution and the enhanced permeability and retention effect of doxorubicin-loaded mesoporous silica nanoparticles in a murine xenograft tumor model. *ACS Nano 5*, 4131–4144.
14. Shroff, K., Kokkoli, E. (2012). PEGylated liposomal doxorubicin targeted to α5β1-expressing MDA-MB-231 breast cancer cells. *Langmuir 28*, 4729–4736.
15. Hu, C. -M. J., Zhang, L. (2012). Nanoparticle-based combination therapy toward overcoming drug resistance in cancer. *Biochemical Pharmacology 83*, 1104–1111.
16. Blanco, E., Hsiao, A., Ruiz-Esparza, G. U., Landry, M. G., Meric-Bernstam, F., Ferrari, M. (2011). Molecular-targeted nanotherapies in cancer: enabling treatment specificity. *Molecular Oncology 5*, 492–503.
17. Anderson, L. J. E., Hansen, E., Lukianova-Hleb, E. Y., Hafner, J. H., Lapotko, D. O. (2010). Optically guided controlled release from liposomes with tunable plasmonic nanobubbles. *Journal of Controlled Release 144*, 151–158.
18. Huynh, L., Neale, C., Pomès, R., Allen, C. (2012). Computational approaches to the rational design of nanoemulsions, polymeric micelles, and dendrimers for drug delivery. *Nanomedicine: nanotechnology, biology and medicine 8*, 20–36.
19. Zatzkewitz, H. (1954). Incorporation of physiologically-active substances into a colloidal blood plasma substitute. I. Incorporation of mescaline peptide into polyvinylpyrrolidone. *Hoppe-Seylers Journal of Physiological Chemistry 297*, 149–156.
20. Bangham, A. D., Horne, R. W. (1964). Negative staining of phospholipids and their structural modification by surface-active agents as observed in the electron microscope. *Journal of Molecular Biology 8*, 660–668.
21. Yallapu, M. M., Jaggi, M., Chauhan, S. C. (2011). Design and engineering of nanogels for cancer treatment. *Drug Discovery Today 16*, 457–463.
22. Petros, R. A., DeSimone, J. M. (2010). Strategies in the design of nanoparticles for therapeutic applications. *Nature Reviews Drug Discovery 9*, 615–627.
23. Gref, R., Minamitake, Y., Peracchia, M., Trubetskoy, V., Torchilin, V., Langer, R. (1994). Biodegradable long-circulating polymeric nanospheres. *Science 263*, 1600–1603.
24. Amoozgar, Z., Park, J., Lin, Q., Yeo, Y. (2012). Low molecular-weight chitosan as a pH-sensitive stealth coating for tumor-specific drug delivery. *Molecular Pharmaceutics 9*, 1262–1270.

25. Choi, K. Y., Yoon, H. Y., Kim, J. -H., Bae, S. M., Park, R. -W., Kang, Y. M., et al. (2011). Smart nanocarrier based on PEGylated hyaluronic acid for cancer therapy. *ACS Nano 5*, 8591–8599.

26. Discher, B. M., Won, Y. -Y., Ege, D. S., Lee, J. C. -M., Bates, F. S., Discher, D. E., et al. (1999). Polymersomes: tough vesicles made from diblock copolymers. *Science 284*, 1143–1146.

27. Ahmed, F., Pakunlu, R. I., Srinivas, G., Brannan, A., Bates, F., Klein, M. L., et al. (2006). Shrinkage of a rapidly growing tumor by drug-loaded polymersomes: pH-triggered release through copolymer degradation. *Molecular Pharmaceutics 3*, 340–350.

28. Smart, T. P., Ryan, A. J., Howse, J. R., Battaglia, G. (2009). Homopolymer induced aggregation of poly(ethylene oxide)n-b-poly(butylene oxide)m polymersomes. *Langmuir 26*, 7425–7430.

29. Shen, J., Zhan, C., Xie, C., Meng, Q., Gu, B., Li, C., et al. (2011). Poly(ethylene glycol)-block-poly(D,L-lactide acid) micelles anchored with angiopep-2 for brain-targeting delivery. *Journal of Drug Targeting 19*, 197–203.

30. Loverde, S. M., Klein, M. L., Discher, D. E. (2012). Nanoparticle shape improves delivery: rational coarse grain molecular dynamics (rCG-MD) of taxol in worm-like PEG-PCL micelles. *Advanced Materials 24*, 3823–3830.

31. Yoo, H. S., Park, T. G. (2001). Biodegradable polymeric micelles composed of doxorubicin conjugated PLGA–PEG block copolymer. *Journal of Controlled Release 70*, 63–70.

32. Batrakova, E. V., Li, S., Brynskikh, A. M., Sharma, A. K., Li, Y., Boska, M., et al. (2010). Effects of pluronic and doxorubicin on drug uptake, cellular metabolism, apoptosis and tumor inhibition in animal models of MDR cancers. *Journal of Controlled Release 143*, 290–301.

33. Kojima, C., Kono, K., Maruyama, K., Takagishi, T. (2000). Synthesis of polyamidoamine dendrimers having poly(ethylene glycol) grafts and their ability to encapsulate anticancer drugs. *Bioconjugate Chemistry 11*, 910–917.

34. Morgan, M. T., Carnahan, M. A., Immoos, C. E., Ribeiro, A. A., Finkelstein, S., Lee, S. J., et al. (2003). Dendritic molecular capsules for hydrophobic compounds. *Journal of the American Chemical Society 125*, 15485–15489.

35. King, H. D., Dubowchik, G. M., Mastalerz, H., Willner, D., Hofstead, S. J., Firestone, R. A., et al. (2002). Monoclonal antibody conjugates of doxorubicin prepared with branched peptide linkers: inhibition of aggregation by methoxytriethyleneglycol chains. *Journal of Medicinal Chemistry 45*, 4336–4343.

36. Serrano-Luna, J., Gutiérrez-Meza, M., Mejía-Zepeda, R., Galindo-Gómez, S., Tsutsumi, V., Shibayama, M. (2010). Effect of phosphatidylcholine–cholesterol liposomes on *Entamoeba histolytica* virulence. *Canadian Journal of Microbiology 56*, 987–995.

37. Yoshizawa, Y., Kono, Y., Ogawara, K. -i., Kimura, T., Higaki, K. (2011). PEG liposomalization of paclitaxel improved its in vivo disposition and anti-tumor efficacy. *International Journal of Pharmaceutics 412*, 132–141.

38. Giljohann, D. A., Seferos, D. S., Daniel, W. L., Massich, M. D., Patel, P. C., Mirkin, C. A. (2010). Gold nanoparticles for biology and medicine. *Angewandte Chemie International Edition 49*, 3280–3294.

39. Mahmoudi, M., Sant, S., Wang, B., Laurent, S., Sen, T. (2011). Superparamagnetic iron oxide nanoparticles (SPIONs): development, surface modification and applications in chemotherapy. *Advanced Drug Delivery Reviews 63*, 24–46.

40. Gradishar, W. J. (2006). Albumin-bound paclitaxel: a next-generation taxane. *Expert Opinion on Pharmacotherapy 7*, 1041–1053.
41. Ge, J., Neofytou, E., Lei, J., Beygui, R. E., Zare, R. N. (2012). Protein–polymer hybrid nanoparticles for drug delivery. *Small 8*, 3573–3578.
42. Kim, J. H., Noh, Y. -W., Heo, M. B., Cho, M. Y., Lim, Y. T. (2012). Multifunctional hybrid nanoconjugates for efficient in vivo delivery of immunomodulating oligonucleotides and enhanced antitumor immunity. *Angewandte Chemie International Edition 51*, 9670–9673.
43. Yang, H., Tyagi, P., Kadam, R. S., Holden, C. A., Kompella, U. B. (2012). Hybrid dendrimer hydrogel/PLGA nanoparticle platform sustains drug delivery for one week and antiglaucoma effects for four days following one-time topical administration. *ACS Nano 6*, 7595–7606.
44. Lee, S. J., Huh, M. S., Lee, S. Y., Min, S., Lee, S., Koo, H., et al. (2012). Tumor-homing poly-siRNA/glycol chitosan self-cross-linked nanoparticles for systemic siRNA delivery in cancer treatment. *Angewandte Chemie International Edition 51*, 7203–7207.
45. Murakami, M., Cabral, H., Matsumoto, Y., Wu, S., Kano, M. R., Yamori, T., et al. (2011). Improving drug potency and efficacy by nanocarrier-mediated subcellular targeting. *Science Translational Medicine 3*, 64ra2.
46. Ko, J. Y., Park, S., Lee, H., Koo, H., Kim, M. S., Choi, K., et al. (2010). pH-sensitive nanoflash for tumoral acidic pH imaging in live animals. *Small 6*, 2539–2544.
47. Fang, J., Nakamura, H., Maeda, H. (2011). The EPR effect: unique features of tumor blood vessels for drug delivery, factors involved, and limitations and augmentation of the effect. *Advanced Drug Delivery Reviews 63*, 136–151.
48. Maeda, H. (2010). Tumor-selective delivery of macromolecular drugs via the EPR effect: background and future prospects. *Bioconjugate Chemistry 21*, 797–802.
49. Maeda, H. (2001). SMANCS and polymer-conjugated macromolecular drugs: advantages in cancer chemotherapy. *Advanced Drug Delivery Reviews 46*, 169–185.
50. Simonsen, T. G., Gaustad, J. -V., Leinaas, M. N., Rofstad, E. K. (2012). High interstitial fluid pressure is associated with tumor-line specific vascular abnormalities in human melanoma xenografts. *PLoS ONE 7*, e40006.
51. Kratz, F., Warnecke, A. (2012). Finding the optimal balance: challenges of improving conventional cancer chemotherapy using suitable combinations with nano-sized drug delivery systems. *Journal of Controlled Release*. 23178950.
52. Heldin, C. -H., Rubin, K., Pietras, K., Ostman, A. (2004). High interstitial fluid pressure—an obstacle in cancer therapy. *Nature Reviews Cancer 4*, 806–813.
53. Yu, M., Tannock, Ian F. (2012). Targeting tumor architecture to favor drug penetration: a new weapon to combat chemoresistance in pancreatic cancer? *Cancer Cell 21*, 327–329.
54. Lee, H., Hoang, B., Fonge, H., Reilly, R., Allen, C. (2010). In vivo distribution of polymeric nanoparticles at the whole-body, tumor, and cellular levels. *Pharmaceutical Research 27*, 2343–2355.
55. Nagamitsu, A., Greish, K., Maeda, H. (2009). Elevating blood pressure as a strategy to increase tumor-targeted delivery of macromolecular drug SMANCS: cases of advanced solid tumors. *Japanese Journal of Clinical Oncology 39*, 756–766.
56. Seki, T., Fang, J., Maeda, H. (2009). Enhanced delivery of macromolecular antitumor drugs to tumors by nitroglycerin application. *Cancer Science 100*, 2426–2430.

57. Maeda, H. (2012). Vascular permeability in cancer and infection as related to macromolecular drug delivery, with emphasis on the EPR effect for tumor-selective drug targeting. *Proceedings of the Japan Academy, Series B: Physical and Biological Sciences 88*, 53–71.
58. Gullotti, E., Yeo, Y. (2009). Extracellularly activated nanocarriers: a new paradigm of tumor targeted drug delivery. *Molecular Pharmaceutics 6*, 1041–1051.
59. Yu, M. K., Park, J., Jon, S. (2012). Targeting strategies for multifunctional nanoparticles in cancer imaging and therapy. *Theranostics 2*, 3–44.
60. Farokhzad, O. C., Langer, R. (2009). Impact of nanotechnology on drug delivery. *ACS Nano 3*, 16–20.
61. Ruoslahti, E., Bhatia, S. N., Sailor, M. J. (2010). Targeting of drugs and nanoparticles to tumors. *The Journal of Cell Biology 188*, 759–768.
62. Miele, E., Spinelli, G. P., Miele, E., Fabrizio, E. D., Ferretti, E., Tomao, S., et al. (2012). Nanoparticle-based delivery of small interfering RNA: challenges for cancer therapy. *International Journal of Nanomedicine 7*. 3637–3657.
63. Deshayes, S., Maurizot, V., Clochard, M. -C., Baudin, C., Berthelot, T., Esnouf, S., et al. (2011). "Click" conjugation of peptide on the surface of polymeric nanoparticles for targeting tumor angiogenesis. *Pharmaceutical Research 28*, 1631–1642.
64. Danhier, F., Pourcelle, V., Marchand-Brynaert, J., Jérôme, C., Feron, O., Préat, V., (2012). Targeting of tumor endothelium by RGD-grafted PLGA-nanoparticles, *Methods in Enzymology 508*, 157–175.
65. Tsourkas, A., Shinde-Patil, V. R., Kelly, K. A., Patel, P., Wolley, A., Allport, J. R., et al. (2005). In vivo imaging of activated endothelium using an anti-VCAM-1 magnetooptical probe. *Bioconjugate Chemistry 16*, 576–581.
66. Sano, K., Temma, T., Azuma, T., Nakai, R., Narazaki, M., Kuge, Y., et al. (2011). A pre-targeting strategy for MR imaging of functional molecules using dendritic Gd-based contrast agents. *Molecular Imaging and Biology 13*, 1196–1203.
67. Li, X., Qiu, L., Zhu, P., Tao, X., Imanaka, T., Zhao, J., et al. (2012). Epidermal growth factor–ferritin H-chain protein nanoparticles for tumor active targeting. *Small 8*, 2505–2514.
68. Li, Y., He, H., Jia, X., Lu, W. -L., Lou, J., Wei, Y. (2012). A dual-targeting nanocarrier based on poly(amidoamine) dendrimers conjugated with transferrin and tamoxifen for treating brain gliomas. *Biomaterials 33*, 3899–3908.
69. Zhang, H., Cai, Z., Sun, Y., Yu, F., Chen, Y., Sun, B. (2012). Folate-conjugated β-cyclodextrin from click chemistry strategy and for tumor-targeted drug delivery. *Journal of Biomedical Materials Research Part A 100*, 2441–2419.
70. Javier, D. J., Nitin, N., Levy, M., Ellington, A., Richards-Kortum, R. (2008). Aptamer-targeted gold nanoparticles as molecular-specific contrast agents for reflectance imaging. *Bioconjugate Chemistry 19*, 1309–1312.
71. Wang, Y. -C., Liu, X. -Q., Sun, T. -M., Xiong, M. -H., Wang, J. (2008). Functionalized micelles from block copolymer of polyphosphoester and poly(ε-caprolactone) for receptor-mediated drug delivery. *Journal of Controlled Release 128*, 32–40.
72. Canton, I., Battaglia, G. (2012). Endocytosis at the nanoscale. *Chemical Society Reviews 41*, 2718–2739.
73. Bartlett, D. W., Su, H., Hildebrandt, I. J., Weber, W. A., Davis, M. E. (2007). Impact of tumor-specific targeting on the biodistribution and efficacy of siRNA nanoparticles

measured by multimodality in vivo imaging. *Proceedings of the National Academy of Sciences of the United States of America 104*, 15549–15554.

74. Kirpotin, D. B., Drummond, D. C., Shao, Y., Shalaby, M. R., Hong, K., Nielsen, U. B., et al. (2006). Antibody targeting of long-circulating lipidic nanoparticles does not increase tumor localization but does increase internalization in animal models. *Cancer Research 66*, 6732–6740.
75. Torchilin, V. P., Passive and active drug targeting: drug delivery to tumors as an example drug delivery, In *Handbook of Experimental Pharmacology*, ed. Schäfer-Korting, M., Springer, Berlin, 2010, pp. 3–53.
76. Huang, R. B., Mocherla, S., Heslinga, M. J., Charoenphol, P., Eniola-Adefeso, O. (2010). Dynamic and cellular interactions of nanoparticles in vascular-targeted drug delivery (review). *Molecular Membrane Biology 27*, 312–327.
77. Desai, N. (2012). Challenges in development of nanoparticle-based therapeutics. *The AAPS Journal 14*, 282–295.
78. Vega-Villa, K. R., Takemoto, J. K., Yáñez, J. A., Remsberg, C. M., Forrest, M. L., Davies, N. M. (2008). Clinical toxicities of nanocarrier systems. *Advanced Drug Delivery Reviews 60*, 929–938.
79. Jong, W. H. D., Borm, P. J. (2008). Drug delivery and nanoparticles: applications and hazards. *International Journal of Nanomedicine 3*, 133–149.
80. Maynard, A. D., Warheit, D. B., Philbert, M. A. (2011). The new toxicology of sophisticated materials: nanotoxicology and beyond. *Toxicological Sciences 120*, S109–S129.
81. Hirn, S., Semmler-Behnke, M., Schleh, C., Wenk, A., Lipka, J., Schäffler, M., et al. (2011). Particle size-dependent and surface charge-dependent biodistribution of gold nanoparticles after intravenous administration. *European Journal of Pharmaceutics and Biopharmaceutics 77*, 407–416.
82. Huang, K., Ma, H., Liu, J., Huo, S., Kumar, A., Wei, T., et al. (2012). Size-dependent localization and penetration of ultrasmall gold nanoparticles in cancer cells, multicellular spheroids, and tumors in vivo. *ACS Nano 6*, 4483–4493.
83. Longmire, M. R., Ogawa, M., Choyke, P. L., Kobayashi, H. (2011). Biologically optimized nanosized molecules and particles: more than just size. *Bioconjugate Chemistry 22*, 993–1000.
84. Venkataraman, S., Hedrick, J. L., Ong, Z. Y., Yang, C., Ee, P. L. R., Hammond, P. T., et al. (2011). The effects of polymeric nanostructure shape on drug delivery. *Advanced Drug Delivery Reviews 63*, 1228–1246.
85. Cedervall, T., Lynch, I., Lindman, S., Berggård, T., Thulin, E., Nilsson, H., et al. (2007). Understanding the nanoparticle–protein corona using methods to quantify exchange rates and affinities of proteins for nanoparticles. *Proceedings of the National Academy of Sciences of the United States of America 104*, 2050–2055.
86. Walkey, C. D., Chan, W. C. W. (2012). Understanding and controlling the interaction of nanomaterials with proteins in a physiological environment. *Chemical Society Reviews 41*, 2780–2799.
87. Aggarwal, P., Hall, J. B., McLeland, C. B., Dobrovolskaia, M. A., McNeil, S. E. (2009). Nanoparticle interaction with plasma proteins as it relates to particle biodistribution, biocompatibility and therapeutic efficacy. *Advanced Drug Delivery Reviews 61*, 428–437.

88. Lartigue, L., Wilhelm, C., Servais, J., Factor, C., Dencausse, A., Bacri, J. -C., et al. (2012). Nanomagnetic sensing of blood plasma protein interactions with iron oxide nanoparticles: impact on macrophage uptake. *ACS Nano 6*, 2665–2678.
89. Schaefer, J., Schulze, C., Marxer, E. E. J., Schaefer, U. F., Wohlleben, W., Bakowsky, U., et al. (2012). Atomic force microscopy and analytical ultracentrifugation for probing nanomaterial protein interactions. *ACS Nano 6*, 4603–4614.
90. Lundqvist, M., Stigler, J., Elia, G., Lynch, I., Cedervall, T., Dawson, K. A. (2008). Nanoparticle size and surface properties determine the protein corona with possible implications for biological impacts. *Proceedings of the National Academy of Sciences of the United States of America 105*, 14265–14270.
91. Hulander, M., Lundgren, A., Berglin, M., Ohrlander, M., Lausmaa, J., Elwing, H. (2011). Immune complement activation is attenuated by surface nanotopography. *International Journal of Nanomedicine 6*, 2653–2666.
92. Blunk, T., Hochstrasser, D. F., Sanchez, J. C., Muller, B. W., Muller, R. H. (1993). Colloidal carriers for intravenous drug targeting: plasma protein adsorption patterns on surface-modified latex particles evaluated by two-dimensional polyacrylamide gel electrophoresis. *Electrophoresis 14*, 1382–1387.
93. Diederichs, J. E. (1996). Plasma protein adsorption patterns on liposomes: establishment of analytical procedure. *Electrophoresis 17*, 607–611.
94. Goppert, T. M., Muller, R. H. (2003). Plasma protein adsorption of Tween 80- and poloxamer 188-stabilized solid lipid nanoparticles. *Journal of Drug Targeting 11*, 225–231.
95. Gref, R., Lück, M., Quellec, P., Marchand, M., Dellacherie, E., Harnisch, S., et al. (2000). 'Stealth' corona-core nanoparticles surface modified by polyethylene glycol (PEG): influences of the corona (PEG chain length and surface density) and of the core composition on phagocytic uptake and plasma protein adsorption. *Colloids and Surfaces B: Biointerfaces 18*, 301–313.
96. Kim, H. R., Andrieux, K., Gil, S., Taverna, M., Chacun, H., Desmaële, D., et al. (2007). Translocation of poly(ethylene glycol-co-hexadecyl)cyanoacrylate nanoparticles into rat brain endothelial cells: role of apolipoproteins in receptor-mediated endocytosis. *Biomacromolecules 8*, 793–799.
97. Lemarchand, C., Gref, R., Passirani, C., Garcion, E., Petri, B., Müller, R., et al. (2006). Influence of polysaccharide coating on the interactions of nanoparticles with biological systems. *Biomaterials 27*, 108–118.
98. Peracchia, M. T., Harnisch, S., Pinto-Alphandary, H., Gulik, A., Dedieu, J. C., Desmaële, D., et al. (1999). Visualization of in vitro protein-rejecting properties of PEGylated stealth® polycyanoacrylate nanoparticles. *Biomaterials 20*, 1269–1275.
99. Thode, K., Lück, M., Semmler, W., Müller, R. H., Kresse, M. (1997). Determination of plasma protein adsorption on magnetic iron oxides: sample preparation. *Pharmaceutical Research 14*, 905–910.
100. AlléAaemann, E., Gravel, P., Leroux, J. -C., Balant, L., Gurny, R. (1997). Kinetics of blood component adsorption on poly(D,L-lactic acid) nanoparticles: Evidence of complement C3 component involvement. *Journal of Biomedical Materials Research 37*, 229–234.
101. Müller, R. H., Rühl, D., Lück, M., Paulke, B. R. (1997). Influence of fluorescent labelling of polystyrene particles on phagocytic uptake, surface hydrophobicity, and plasma protein adsorption. *Pharmaceutical Research 14*, 18–24.

102. Labarre, D., Vauthier, C., Chauvierre, C., Petri, B., Müller, R., Chehimi, M. M. (2005). Interactions of blood proteins with poly(isobutylcyanoacrylate) nanoparticles decorated with a polysaccharidic brush. *Biomaterials 26*, 5075–5084.

103. Göppert, T. M., Müller, R. H. (2005). Polysorbate-stabilized solid lipid nanoparticles as colloidal carriers for intravenous targeting of drugs to the brain: Comparison of plasma protein adsorption patterns. *Journal of Drug Targeting 13*, 179–187.

104. Göppert, T. M., Müller, R. H. (2005). Protein adsorption patterns on poloxamer- and poloxamine-stabilized solid lipid nanoparticles (SLN). *European Journal of Pharmaceutics and Biopharmaceutics 60*, 361–372.

105. Noh, H., Vogler, E. A. (2007). Volumetric interpretation of protein adsorption: Competition from mixtures and the Vroman effect. *Biomaterials 28*, 405–422.

106. Göppert, T. M., Müller, R. H. (2005). Adsorption kinetics of plasma proteins on solid lipid nanoparticles for drug targeting. *International Journal of Pharmaceutics 302*, 172–186.

107. Jansch, M., Stumpf, P., Graf, C., Rühl, E., Müller, R. H. (2012). Adsorption kinetics of plasma proteins on ultrasmall superparamagnetic iron oxide (USPIO) nanoparticles. *International Journal of Pharmaceutics 428*, 125–133.

108. OwensIII, D. E., Peppas, N. A. (2006). Opsonization, biodistribution, and pharmacokinetics of polymeric nanoparticles. *International Journal of Pharmaceutics 307*, 93–102.

109. Roser, M., Fischer, D., Kissel, T. (1998). Surface-modified biodegradable albumin nano- and microspheres. II: effect of surface charges on in vitro phagocytosis and biodistribution in rats. *European Journal of Pharmaceutics and Biopharmaceutics 46*, 255–263.

110. Salvador-Morales, C., Zhang, L., Langer, R., Farokhzad, O. C. (2009). Immunocompatibility properties of lipid–polymer hybrid nanoparticles with heterogeneous surface functional groups. *Biomaterials 30*, 2231–2240.

111. Gessner, A., Lieske, A., Paulke, B. R., Müller, R. H. (2002). Influence of surface charge density on protein adsorption on polymeric nanoparticles: analysis by two-dimensional electrophoresis. *European Journal of Pharmaceutics and Biopharmaceutics 54*, 165–170.

112. Gessner, A., Lieske, A., Paulke, B. -R., Müller, R. H. (2003). Functional groups on polystyrene model nanoparticles: Influence on protein adsorption. *Journal of Biomedical Materials Research Part A 65*, 319–326.

113. Xiao, K., Li, Y., Luo, J., Lee, J. S., Xiao, W., Gonik, A. M., et al. (2011). The effect of surface charge on in vivo biodistribution of PEG-oligocholic acid based micellar nanoparticles. *Biomaterials 32*, 3435–3446.

114. Yadav, K. S., Jacob, S., Sachdeva, G., Chuttani, K., Mishra, A. K., Sawant, K. K. (2011). Long circulating PEGylated PLGA nanoparticles of cytarabine for targeting leukemia. *Journal of Microencapsulation 28*, 729–742.

115. Vij, N., Min, T., Marasigan, R., Belcher, C., Mazur, S., Ding, H., et al. (2010). Development of PEGylated PLGA nanoparticle for controlled and sustained drug delivery in cystic fibrosis. *Journal of Nanobiotechnology 8*, 22.

116. Ratzinger, G., Länger, U., Neutsch, L., Pittner, F., Wirth, M., Gabor, F. (2009). Surface modification of PLGA particles: the interplay between stabilizer, ligand size, and hydrophobic interactions. *Langmuir 26*, 1855–1859.

117. Cedervall, T., Lynch, I., Foy, M., Berggård, T., Donnelly, S. C., Cagney, G., et al. (2007). Detailed identification of plasma proteins adsorbed on copolymer nanoparticles. *Angewandte Chemie International Edition 46*, 5754–5756.

118. Fang, C., Shi, B., Pei, Y. -Y., Hong, M. -H., Wu, J., Chen, H. -Z. (2006). In vivo tumor targeting of tumor necrosis factor-α-loaded stealth nanoparticles: effect of MePEG molecular weight and particle size. *European Journal of Pharmaceutical Sciences 27*, 27–36.

119. Yokoyama, M. (2010). Polymeric micelles as a new drug carrier system and their required considerations for clinical trials. *Expert Opinion on Drug Delivery 7*, 145–158.

120. Savić, R., Azzam, T., Eisenberg, A., Maysinger, D. (2006). Assessment of the integrity of poly(caprolactone)-b-poly(ethylene oxide) micelles under biological conditions: a fluorogenic-based approach. *Langmuir 22*, 3570–3578.

121. Chen, H., Kim, S., He, W., Wang, H., Low, P. S., Park, K., et al. (2008). Fast release of lipophilic agents from circulating PEG-PDLLA micelles revealed by in vivo Förster resonance energy transfer imaging. *Langmuir 24*, 5213–5217.

122. Chen, H., Kim, S., Li, L., Wang, S., Park, K., Cheng, J. -X. (2008). Release of hydrophobic molecules from polymer micelles into cell membranes revealed by Förster resonance energy transfer imaging. *Proceedings of the National Academy of Sciences of the United States of America 105*, 6596–6601.

123. Kim, S., Shi, Y., Kim, J. Y., Park, K., Cheng, J. -X. (2010). Overcoming the barriers in micellar drug delivery: loading efficiency, in vivo stability, and micelle–cell interaction. *Expert Opinion on Drug Delivery 7*, 49–62.

124. Dobrovolskaia, M. A., McNeil, S. E. (2007). Immunological properties of engineered nanomaterials. *Nature Nanotechnology 2*, 469–478.

125. Moghimi, S. M., Patel, H. M. (1998). Serum-mediated recognition of liposomes by phagocytic cells of the reticuloendothelial system—the concept of tissue specificity. *Advanced Drug Delivery Reviews 32*, 45–60.

126. Frank, M. M., Fries, L. F. (1991). The role of complement in inflammation and phagocytosis. *Immunology Today 12*, 322–326.

127. Camner, P., Lundborg, M., Låstbom, L., Gerde, P., Gross, N., Jarstrand, C. (2002). Experimental and calculated parameters on particle phagocytosis by alveolar macrophages. *Journal of Applied Physiology 92*, 2608–2616.

128. Kim, K., Kim, J. H., Park, H., Kim, Y. -S., Park, K., Nam, H., et al. (2010). Tumor-homing multifunctional nanoparticles for cancer theragnosis: Simultaneous diagnosis, drug delivery, and therapeutic monitoring. *Journal of Controlled Release 146*, 219–227.

129. Tong, L., He, W., Zhang, Y., Zheng, W., Cheng, J. -X. (2009). Visualizing systemic clearance and cellular level biodistribution of gold nanorods by intrinsic two-photon luminescence. *Langmuir 25*, 12454–12459.

130. Nagayama, S., Ogawara, K. -i., Minato, K., Fukuoka, Y., Takakura, Y., Hashida, M., et al. (2007). Fetuin mediates hepatic uptake of negatively charged nanoparticles via scavenger receptor. *International Journal of Pharmaceutics 329*, 192–198.

131. Dutta, D., Sundaram, S. K., Teeguarden, J. G., Riley, B. J., Fifield, L. S., Jacobs, J. M., et al. (2007). Adsorbed proteins influence the biological activity and molecular targeting of nanomaterials. *Toxicological Sciences 100*, 303–315.

132. Nagayama, S., Ogawara, K. -i., Fukuoka, Y., Higaki, K., Kimura, T. (2007). Time-dependent changes in opsonin amount associated on nanoparticles alter their hepatic uptake characteristics. *International Journal of Pharmaceutics 342*, 215–221.

133. Moghimi, S. M., Hedeman, H., Muir, I., Illum, L., Davis, S. (1933). An investigation of the filtration capacity and the fate of large filtered sterically-stabilized microspheres in rat spleen. *Biochimica et Biophysica Acta 1157*, 233–240.

134. Moghimi, S. M. (1995). Mechanisms of splenic clearance of blood cells and particles: towards development of new splenotropic agents. *Advanced Drug Delivery Reviews 17*, 103–115.

135. Estevanato, L., Cintra, D., Baldini, N., Portilho, F., Barosa, L., Martins, O., et al. (2011). Preliminary biocompatibility investigation of magnetic albumin nanosphere designed as a potential versatile drug delivery system. *International Journal of Nanomedicine 6*, 1709–1717.

136. Piao, L., Li, H., Teng, L., Yung, B. C., Sugimoto, Y., Brueggemeier, R. W., et al. (2013). Human serum albumin-coated lipid nanoparticles for delivery of siRNA to breast cancer. *Nanomedicine: Nanotechnology, Biology and Medicine 9*, 122–129.

137. van Rooy, I., Cakir-Tascioglu, S., Hennink, W., Storm, G., Schiffelers, R., Mastrobattista, E. (2011). In vivo methods to study uptake of nanoparticles into the brain. *Pharmaceutical Research 28*, 456–471.

138. Gaumet, M., Vargas, A., Gurny, R., Delie, F. (2008). Nanoparticles for drug delivery: the need for precision in reporting particle size parameters. *European Journal of Pharmaceutics and Biopharmaceutics 69*, 1–9.

139. Takakura, Y., Mahato, R. I., Hashida, M. (1998). Extravasation of macromolecules. *Advanced Drug Delivery Reviews 34*, 93–108.

140. Wadenvik, H., Kutti, J. (1988). The spleen and pooling of blood cells. *European Journal of Haematology 41*, 1–5.

141. Cogger, V. C., McNerney, G. P., Nyunt, T., DeLeve, L. D., McCourt, P., Smedsrød, B., et al. (2010). Three-dimensional structured illumination microscopy of liver sinusoidal endothelial cell fenestrations. *Journal of Structural Biology 171*, 382–388.

142. Fahimi, H. D., *The Liver: Biology and Pathobiology*, Raven Press, New York, 1982.

143. Ohlson, M., Sörensson, J., Haraldsson, B. (2001). A gel-membrane model of glomerular charge and size selectivity in series. *American Journal of Physiology—Renal Physiology 280*, F396–F405.

144. Conhaim, R. L., Eaton, A., Staub, N. C., Heath, T. D. (1988). Equivalent pore estimate for the alveolar-airway barrier in isolated dog lung. *Journal of Applied Physiology 64*, 1134–1142.

145. Nakaoka, R., Tabata, Y., Yamaoka, T., Ikada, Y. (1997). Prolongation of the serum half-life period of superoxide dismutase by poly(ethylene glycol) modification. *Journal of Controlled Release 46*, 253–261.

146. Moghimi, S. M. (1995). Exploiting bone marrow microvascular structure for drug delivery and future therapies. *Advanced Drug Delivery Reviews 17*, 61–73.

147. Seymour, L. W. (1992). Passive tumor targeting of soluble macromolecules and drug conjugates. *Critical Reviews in Therapeutic Drug Carrier Systems 9*, 135–187.

148. Cucullo, L., McAllister, M. S., Kight, K., Krizanac-Bengez, L., Marroni, M., Mayberg, M. R., et al. (2002). A new dynamic in vitro model for the multidimensional study of

astrocyte–endothelial cell interactions at the blood–brain barrier. *Brain Research 951*, 243–254.

149. Kas, H. S. (2004). Drug delivery to brain by microparticulate systems. *Advances in Experimental Medicine and Biology 553*, 221–230.
150. Arfors, K. E., Rutili, G., Svensiö, E. (1979). Microvascular transport of macromolecules in normal and inflammatory conditions. *Acta Physiologica Scandinavica, Supplementum 463*, 93–103.
151. Hirano, A., Kawanami, T., Llena, J. F. (1994). Electron microscopy of the blood-brain barrier in disease. *Microscopy Research and Technique 27*, 543–556.
152. Kraal, G. (1992). Cells in the marginal zone of the spleen. *International Review of Cytology 132*, 31–74.
153. Steiniger, B., Barth, P. (2000). Microanatomy and function of the spleen. *Advances in Anatomy, Embryology and Cell Biology 151*, 1–101.
154. Mebius, R. E., Kraal, G. (2005). Structure and function of the spleen. *Nature Reviews Immunology 5*, 606–616.
155. MacDonald, I. C., Ragan, D. M., Schmidt, E. E., Groom, A. C. (1987). Kinetics of red blood cell passage through interendothelial slits into venous sinuses in rat spleen, analyzed by in vivo microscopy. *Microvascular Research 33*, 118–134.
156. Groom, A. C., Schmidt, E. E., MacDonald, I. C. (1991). Microcirculatory pathways and blood flow in spleen: new insights from washout kinetics, corrosion casts, and quantitative intravital videomicroscopy. *Scanning Microscopy 5*, 159–173.
157. Chen, L. -T., Weiss, L. (1973). The role of the sinus wall in the passage of erythrocytes through the spleen. *Blood 41*, 529–537.
158. Drenckhahn, D., Wagner, J. (1986). Stress fibers in the splenic sinus endothelium in situ: molecular structure, relationship to the extracellular matrix, and contractility. *The Journal of Cell Biology 102*, 1738–1747.
159. Moghimi, S. M., Porter, C. J. H., Muir, I. S., Illum, L., Davis, S. S. (1991). Non-phagocytic uptake of intravenously injected microspheres in rat spleen: influence of particle size and hydrophilic coating. *Biochemical and Biophysical Research Communications 177*, 861–866.
160. Demoy, M., Andreux, J. -P., Weingarten, C., Gouritin, B., Guilloux, V., Couvreur, P. (1999). Spleen capture of nanoparticles: influence of animal species and surface characteristics. *Pharmaceutical Research 16*, 37–41.
161. Marieb, E. N., *Essentials of Human Anatomy and Physiology*, Benjamin Cummings, San Francisco, 2002.
162. Adams, D. H., Eksteen, B. (2006). Aberrant homing of mucosal T cells and extra-intestinal manifestations of inflammatory bowel disease. *Nature Reviews Immunology 6*, 244–251.
163. Dan, C., Wake, K. (1985). Modes of endocytosis of latex particles in sinusoidal endothelial and Kupffer cells of normal and perfused rat liver. *Experimental Cell Research 158*, 75–85.
164. Wisse, E., *Ultrastructure and function of Kupffer cells and other sinusoidal cells in the liver. In: Kupffer cells and other liver sinusoidal cells*, Elsevier, Amsterdam, 1977.
165. Kuntz, E., Kuntz, H. -D. *Hepatology: Principles and Practice: History, Morphology, Biochemistry, Diagnostics, Clinic, Therapy*, Springer, Berlin, 2006.

166. Vidal-Vanaclocha, F., Barberá-Guillem, E. (1985). Fenestration patterns in endothelial cells of rat liver sinusoids. *Journal of Ultrastructure Research 90*, 115–123.

167. Fraser, R., Dobbs, B. R., Rogers, G. W. (1995). Lipoproteins and the liver sieve—the role of the fenestrated sinusoidal endothelium in lipoprotein metabolism, atherosclerosis, and cirrhosis. *Hepatology 21*, 863–874.

168. Sun, X., Rossin, R., Turner, J. L., Becker, M. L., Joralemon, M. J., Welch, M. J., et al. (2005). An assessment of the effects of shell cross-linked nanoparticle size, core composition, and surface PEGylation on in vivo biodistribution. *Biomacromolecules 6*, 2541–2554.

169. Kren, B. T., Unger, G. M., Sjeklocha, L., Trossen, A. A., Korman, V., Diethelm-Okita, B. M., et al. (2009). Nanocapsule-delivered Sleeping Beauty mediates therapeutic Factor VIII expression in liver sinusoidal endothelial cells of hemophilia A mice. *The Journal of Clinical Investigation 119*, 2086–2099.

170. Glaus, C., Rossin, R., Welch, M. J., Bao, G. (2010). In vivo evaluation of 64Cu-labeled magnetic nanoparticles as a dual-modality PET/MR imaging agent. *Bioconjugate Chemistry 21*, 715–722.

171. Schluep, T., Hwang, J., Hildebrandt, I. J., Czernin, J., Choi, C. H. J., Alabi, C. A., et al. (2009). Pharmacokinetics and tumor dynamics of the nanoparticle IT-101 from PET imaging and tumor histological measurements. *Proceedings of the National Academy of Sciences of the United States of America 106*, 11394–11399.

172. Hoshino, Y., Koide, H., Furuya, K., Haberaecker, W. W., Lee, S. -H., Kodama, T., et al. (2012). The rational design of a synthetic polymer nanoparticle that neutralizes a toxic peptide in vivo. *Proceedings of the National Academy of Sciences of the United States of America 109*, 33–38.

173. Choi, H. S., Liu, W., Liu, F., Nasr, K., Misra, P., Bawendi, M. G., et al. (2010). Design considerations for tumour-targeted nanoparticles. *Nature Nanotechnology 5*, 42–47.

174. Nishimori, H., Kondoh, M., Isoda, K., Tsunoda, S. -i., Tsutsumi, Y., Yagi, K. (2009). Silica nanoparticles as hepatotoxicants. *European Journal of Pharmaceutics and Biopharmaceutics 72*, 496–501.

175. Popielarski, S. R., Hu-Lieskovan, S., French, S. W., Triche, T. J., Davis, M. E. (2005). A nanoparticle-based model delivery system to guide the rational design of gene delivery to the liver. 2. In vitro and in vivo uptake results. *Bioconjugate Chemistry 16*, 1071–1080.

176. Kumar, R., Roy, I., Ohulchanskky, T. Y., Vathy, L. A., Bergey, E. J., Sajjad, M., et al. (2010). In vivo biodistribution and clearance studies using multimodal organically modified silica nanoparticles. *ACS Nano 4*, 699–708.

177. Longmire, M., Choyke, P. L., Kobayashi, H. (2008). Clearance properties of nano-sized particles and molecules as imaging agents: considerations and caveats. *Nanomedicine 3*, 703–717.

178. Smoyer, W. E., Mundel, P. (1998). Regulation of podocyte structure during the development of nephrotic syndrome. *Journal of Molecular Medicine 76*, 172–183.

179. Deen, W. M., Lazzara, M. J., Myers, B. D. (2001). Structural determinants of glomerular permeability. *American Journal of Physiology—Renal Physiology 281*, F579–F596.

180. Kobayashi, H., Brechbiel, M. W. (2004). Dendrimer-based nanosized MRI contrast agents. *Current Pharmaceutical Biotechnology 5*, 539–549.

181. Kobayashi, H., Brechbiel, M. W. (2005). Nano-sized MRI contrast agents with dendrimer cores. *Advanced Drug Delivery Reviews 57*, 2271–2286.

182. Choi, H. S., Ashitate, Y., Lee, J. H., Kim, S. H., Matsui, A., Insin, N., et al. (2010). Rapid translocation of nanoparticles from the lung airspaces to the body. *Nature Biotechnology 28*, 1300–1303.

183. Kobayashi, H., Le, N., Kim, I. -s., Kim, M. -K., Pie, J. -E., Drumm, D., et al. (1999). The pharmacokinetic characteristics of glycolated humanized anti-Tac Fabs are determined by their isoelectric points. *Cancer Research 59*, 422–430.

184. Venturoli, D., Rippe, B. (2005). Ficoll and dextran vs. globular proteins as probes for testing glomerular permselectivity: effects of molecular size, shape, charge, and deformability. *American Journal of Physiology—Renal Physiology 288*, F605–F613.

185. van Nostrum, C. F. (2011). Covalently cross-linked amphiphilic block copolymer micelles. *Soft Matter 7*, 3246–3259.

186. Cheng, R., Feng, F., Meng, F., Deng, C., Feijen, J., Zhong, Z. (2011). Glutathione-responsive nano-vehicles as a promising platform for targeted intracellular drug and gene delivery. *Journal of Controlled Release 152*, 2–12.

187. Sun, Y., Yan, X., Yuan, T., Liang, J., Fan, Y., Gu, Z., et al. (2010). Disassemblable micelles based on reduction-degradable amphiphilic graft copolymers for intracellular delivery of doxorubicin. *Biomaterials 31*, 7124–7131.

188. Sun, P., Zhou, D., Gan, Z. (2011). Novel reduction-sensitive micelles for triggered intracellular drug release. *Journal of Controlled Release 152*, Supplement 1, e85–e87.

189. Kim, J. O., Sahay, G., Kabanov, A. V., Bronich, T. K. (2010). Polymeric micelles with ionic cores containing biodegradable cross-links for delivery of chemotherapeutic agents. *Biomacromolecules 11*, 919–926.

190. O'Reilly, R. K., Hawker, C. J., Wooley, K. L. (2006). Cross-linked block copolymer micelles: functional nanostructures of great potential and versatility. *Chemical Society Reviews 35*, 1068–1083.

191. Li, Y., Xiao, K., Luo, J., Xiao, W., Lee, J. S., Gonik, A. M., et al. (2011). Well-defined, reversible disulfide cross-linked micelles for on-demand paclitaxel delivery. *Biomaterials 32*, 6633–6645.

192. Koo, A. N., Min, K. H., Lee, H. J., Lee, S. -U., Kim, K., Chan Kwon, I., et al. (2012). Tumor accumulation and antitumor efficacy of docetaxel-loaded core-shell-corona micelles with shell-specific redox-responsive cross-links. *Biomaterials 33*, 1489–1499.

193. Choi, W. I., Yoon, K. C., Im, S. K., Kim, Y. H., Yuk, S. H., Tae, G. (2010). Remarkably enhanced stability and function of core/shell nanoparticles composed of a lecithin core and a pluronic shell layer by photo-crosslinking the shell layer: In vitro and in vivo study. *Acta Biomaterialia 6*, 2666–2673.

194. Choi, W. I., Lee, J. H., Kim, J. -Y., Kim, J. -C., Kim, Y. H., Tae, G. (2012). Efficient skin permeation of soluble proteins via flexible and functional nano-carrier. *Journal of Controlled Release 157*, 272–278.

195. Moghimi, S. M., Szebeni, J. (2003). Stealth liposomes and long circulating nanoparticles: critical issues in pharmacokinetics, opsonization and protein-binding properties. *Progress in Lipid Research 42*, 463–478.

196. Chen, J., Yan, G. -j., Hu, R. -r., Gu, Q. -w., Chen, M. -l., Gu, W., et al. (2012). Improved pharmacokinetics and reduced toxicity of brucine after encapsulation into stealth liposomes: role of phosphatidylcholine. *International Journal of Nanomedicine 7*, 3567–3577.

197. Rojnik, M., Kocbek, P., Moret, F., Compagnin, C., Celotti, L., Bovis, M. J., et al. (2012). In vitro and in vivo characterization of temoporfin-loaded PEGylated PLGA nanoparticles for use in photodynamic therapy. *Nanomedicine 7*, 663–677.

198. Jiang, W., Lionberger, R., Yu, L. X. (2011). In vitro and in vivo characterizations of PEGylated liposomal doxorubicin. *Bioanalysis 3*, 333–344.

199. Ragi, C., Sedaghat-Herati, M. R., Ouameur, A. A., Tajmir-Riahi, H. A. (2005). The effects of poly(ethylene glycol) on the solution structure of human serum albumin. *Biopolymers 78*, 231–236.

200. Topchieva, I. N., Sorokina, E. M., Efremova, N. V., Ksenofontov, A. L., Kurganov, B. I. (1999). Noncovalent adducts of poly(ethylene glycols) with proteins. *Bioconjugate Chemistry 11*, 22–29.

201. Furness, E. L., Ross, A., Davis, T. P., King, G. C. (1998). A hydrophobic interaction site for lysozyme binding to polyethylene glycol and model contact lens polymers. *Biomaterials 19*, 1361–1369.

202. Yokoe, J. -i., Sakuragi, S., Yamamoto, K., Teragaki, T., Ogawara, K. -i., Higaki, K., et al. (2008). Albumin-conjugated PEG liposome enhances tumor distribution of liposomal doxorubicin in rats. *International Journal of Pharmaceutics 353*, 28–34.

203. Xie, J., Wang, J., Niu, G., Huang, J., Chen, K., Li, X., et al. (2010). Human serum albumin coated iron oxide nanoparticles for efficient cell labeling. *Chemical Communications 46*, 433–435.

204. Wu, Y., Chakrabortty, S., Gropeanu, R. A., Wilhelmi, J., Xu, Y., Er, K. S., et al. (2010). pH-responsive quantum dots via an albumin polymer surface coating. *Journal of the American Chemical Society 132*, 5012–5014.

205. Acharya, G., Shin, C. S., McDermott, M., Mishra, H., Park, H., Kwon, I. C., et al. (2010). The hydrogel template method for fabrication of homogeneous nano/microparticles. *Journal of Controlled Release 141*, 314–319.

206. Merkel, T. J., Chen, K., Jones, S. W., Pandya, A. A., Tian, S., Napier, M. E., et al. (2012). The effect of particle size on the biodistribution of low-modulus hydrogel PRINT particles. *Journal of Controlled Release 162*, 37–44.

207. Champion, J. A., Mitragotri, S. (2006). Role of target geometry in phagocytosis. *Proceedings of the National Academy of Sciences of the United States of America 103*, 4930–4934.

208. Sharma, G., Valenta, D. T., Altman, Y., Harvey, S., Xie, H., Mitragotri, S., et al. (2010). Polymer particle shape independently influences binding and internalization by macrophages. *Journal of Controlled Release 147*, 408–412.

209. Geng, Y., Dalhaimer, P., Cai, S., Tsai, R., Tewari, M., Minko, T., et al. (2007). Shape effects of filaments versus spherical particles in flow and drug delivery. *Nature Nanotechnology 2*, 249–255.

210. Doshi, N., Zahr, A. S., Bhaskar, S., Lahann, J., Mitragotri, S. (2009). Red blood cell-mimicking synthetic biomaterial particles. *Proceedings of the National Academy of Sciences of the United States of America 106*, 21495–21499.

211. Hendrickson, G. R., Lyon, L. A. (2010). Microgel translocation through pores under confinement. *Angewandte Chemie International Edition 49*, 2193–2197.

212. Banquy, X., Suarez, F., Argaw, A., Rabanel, J. -M., Grutter, P., Bouchard, J.-F., et al. (2009). Effect of mechanical properties of hydrogel nanoparticles on macrophage cell uptake. *Soft Matter 5*, 3984–3991.

213. Bae, Y. H., Park, K. (2011). Targeted drug delivery to tumors: myths, reality and possibility. *Journal of Controlled Release 153*, 198–205.

214. Lee, S.-Y., Kim, S., Tyler, J. Y., Park, K., Cheng, J.-X. (2013). Blood-stable, tumor-adaptable disulfide bonded mPEG-(Cys)4-PDLLA micelles for chemotherapy. *Biomaterials 34*, 552–561.

# DRUG DELIVERY STRATEGIES FOR COMBATING MULTIPLE DRUG RESISTANCE

JOSEPH W. NICHOLS
*Department of Bioengineering, University of Utah, Salt Lake City, UT, USA*
YOU HAN BAE
*Department of Pharmaceutics, University of Utah, Salt Lake City, UT, USA*

## 10.1 INTRODUCTION

It has been 40 years since the United States Congress passed the National Cancer Act, which directed a great deal of effort into research combating cancer. Since that time, only modest gains have been made in reducing cancer mortality rates in the United States and elsewhere, and the most significant gains have come as a result of prevention campaigns and declining smoking rates [1,2]. Table 10.1 shows that death rates for many cancers have actually increased since 1975 due to rising incidence rates. A major challenge facing researchers has been multiple drug resistance (MDR) in solid tumors [3–6]. Multiple drug resistance is defined as the simultaneous resistance to a wide range of drugs that are unrelated in structure or mechanism of action, and is encountered in up to 80% of nonresectable and 40% of all cancer cases [7,8]. The presence of MDR markers in a patient is strongly associated with poor clinical outcomes [9]. The mechanisms of MDR are extremely varied and can include both chemical and physical barriers to drug action. An MDR tumor generally displays some combination of several MDR phenotypes. The expression of these phenotypes may also vary from cell to cell within a tumor [10].

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*,
Edited by Shayne C. Gad
Copyright © 2013 John Wiley & Sons, Inc.

**TABLE 10.1 Cancer Incidence, Survival, and Death Rates in the United States for the Years from the Mid-1970s to Early 2000s**

| | Incidence Rates[a] | | Mortality Rates[b] | | Survival Rates[a] | |
|---|---|---|---|---|---|---|
| | 1975 | 2003 | 1975[c] | 2000–2004 | 1975 (%) | 2003 (%) |
| All sites | 400.4 | 472.9 | 162.2 | 192.7 | 48.9 | 66.7 |
| Brain, nervous system | 5.9 | 6.6 | 3.8 | 4.4 | 23.0 | 33.7 |
| Breast | 105.1 | 126.5 | 14.6 | 14.5 | 75.5 | 89.9 |
| Colon and rectum | 59.5 | 50.6 | 21.7 | 19.4 | 48.7 | 65.6 |
| Leukemia | 12.8 | 13.3 | 6.6 | 7.5 | 33.4 | 55.6 |
| Lung and bronchus | 52.3 | 64.5 | 36.8 | 54.7 | 11.5 | 16.0 |
| Non-Hodgkin lymphoma | 11.1 | 20.6 | 4.7 | 7.6 | 46.0 | 70.2 |
| Ovary | 16.3 | 13.6 | 4.7 | 5.0 | 33.7 | 44.1 |
| Pancreas | 11.8 | 11.6 | 8.6 | 10.6 | 3.0 | 4.8 |
| Prostate | 94.0 | 169.0 | 8.4 | 10.5 | 66.4 | 99.4 |

Incidence and mortality rates reported as per 100,000 population per year. Survival rates are the percent of patients alive 5 years from diagnosis.
[a]Reference [1].
[b]Reference [2].
[c]As reported in Reference [11].

Many innovative strategies are in development to overcome individual resistance mechanisms. Most modern drug delivery technologies use nanoparticles as the main delivery platform because of their versatility in both form and function. New treatments are frequently successful in overcoming individual drug delivery barriers (such as resistance mechanisms) and show promise in preclinical studies, but successful application of these technologies in a clinical setting has been rare. Transferring new technologies from laboratory to clinic may require a broader perspective in formulating a drug delivery strategy. Drug carrier characteristics should be optimized to deliver the highest concentration possible from drug administration to drug action, rather than narrowly focus on singular characteristics.

This chapter outlines major mechanisms of drug resistance in tumors, emphasizing the contribution of each mechanism to the whole classification called MDR. The chapter then samples innovative strategies to circumvent or overcome these resistance mechanisms.

## 10.2 PATHWAYS TO DRUG RESISTANCE

An effective cancer treatment must deliver sufficient quantities of drug to its site of action for a sufficiently long period of time to allow drug action to be effective. Otherwise, the treatment will fail to annihilate the cancer regardless of the drug's predicted effectiveness. Achieving a lethal drug dosage at the intracellular site of action for all cancerous cells in a tumor is the perennial challenge in applied drug delivery research. The difficulty of drug delivery arises from the relatively large

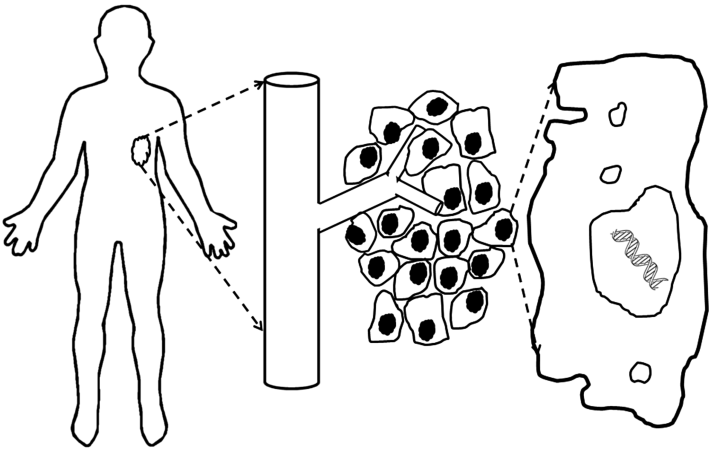

**FIGURE 10.1** Before reaching an intracellular site of action, a drug must first pass through several stages of drug distribution. The first is systemic circulation, which distributes the drug carriers throughout the entire circulatory system. In this stage, carriers are subject to clearance through renal filtration or the MPS. A small number of particles will avoid clearance and extravasate out of the disorganized neovasculature into the interstitial space of the tumor. Drug carrier concentration in the tumor may be high relative to other tissues due to the EPR effect but most of the administered drug will not encounter the tumor. Once in the tumor the drug carriers must pass through the tumor interstitial space and extracellular matrix to encounter a cell. The final phase of distribution involves crossing the lipid bilayer membrane to enter the cell then diffusing to the site of action. Drugs targeted to the nucleus must also enter the nuclear envelope. (*Source:* Adapted from Reference [10] and modified.)

distances and diverse environments through which an individual drug particle must traverse before reaching its target. The total path of a drug can be summarized in Figure 10.1, showing the various environments a particle must travel through before an intracellular target may be reached.

### 10.2.1 Resistance of the Tumor Microenvironment

A great deal of drug resistance is conferred to tumors by characteristics of the tumor microenvironment which prevent a drug from ever reaching the subcellular site of action. Most cancer therapeutics are administered via intravenous injection and within minutes are distributed throughout the entire circulatory system. The first major barrier to drug delivery is to move the drug out of the circulatory system and into the tumor before the drug can be cleared by renal filtration [12] and the monocyte phagocyte system (MPS) [13]. Long circulating particles have the opportunity to collect in the tumor by means of the enhanced permeability and retention (EPR) effect, which is the tendency of drugs to preferentially accumulate in a solid tumor due to the tumor's leaky vasculature and poor lymphatic drainage [14,15]. The EPR effect is widely used in modern drug therapeutics, but its usefulness can be mitigated by several countervailing trends [16].

To take advantage of the EPR effect, a drug delivery system must be long circulating, which generally means using large particles which are able to avoid renal filtration modified with a polymer such as poly(ethylene glycol) (PEG) to slow phagocytic uptake [17]. Unfortunately, large particles are less able to extravasate into the tumor interstitium from the capillaries [18]. Thus, there is a trade off between extended circulation half-life and tumor extravasation [19].

Extravasation is further limited by the elevated hydrostatic pressure of the tumor interior, which can sometimes nearly equal the microvascular pressure, severely diminishing the driving force for extravasation into the tumor [17]. This elevated interstitial pressure is a result of the leaky blood vessels that freely allow proteins and other solutes normally excluded by the capillaries of normal tissues. This causes a higher osmotic pressure in interstitial space of the tumor and results in edema [20]. The additional water cannot be cleared because the lymph vessels are often destroyed by a high static pressure resulting from rapidly dividing cells in the limited tumor space [21]. Volume expansion cannot equalize the pressure because of the dense highly cross-linked extracellular matrix (ECM) of the tumor [22].

The pressure at the tumor periphery is lower than that in the center of the tumor creating a driving force for convective flow from the tumor center to the outside [23]. The vasculature is not uniformly distributed in the tumor but is mostly concentrated at the tumor periphery. This is because the solid stress of the tumor destroys the original blood vessels [24], which must be replaced by new vasculature through angiogenesis. These new vessels form from existing vessels outside the tumor and are disorganized, rarely penetrate to the tumor center, and leave large spaces between capillaries [25–27]. This distance becomes a large diffusive barrier and results in chronically hypoxic and undernourished regions in the tumor, which can be seen in Figure 10.2 [28], which is aggravated by the convective flow away from the center of the tumor. This presents a critical challenge for researchers to consider because the conditions of the hypoxic regions select for the most dangerous cell phenotypes such as aggressive growth and metastatic behavior, but these are the very cells least likely to be exposed to lethal drug concentrations [29].

Diffusion is made more difficult by the dense collagen network of the tumor, which may slow or even stop the movement of large particles entirely [30,31]. In this capacity, the collagen matrix can be considered a gate or filter, physically excluding some particles based on size. Those particles that are not excluded must travel a highly tortuous path navigating both the collagen fibers and the extremely dense cell mass. This tends to exacerbate the distance barrier to particle diffusion.

### 10.2.2 Cellular Resistance

Once a drug reaches its site of action, it must enter the cell and reach the site of action before it can be effective. Drug-resistant cells have several mechanisms that can prevent the drug from entering the cell and distributing to the target site. One of the principal and best understood mechanisms of MDR is the active efflux of small molecule drugs by membrane-bound, nonspecific, ATP-driven molecular pumps, of which P-glycoprotein (Pgp) is the most common [32–35]. Pgp pumps are normally

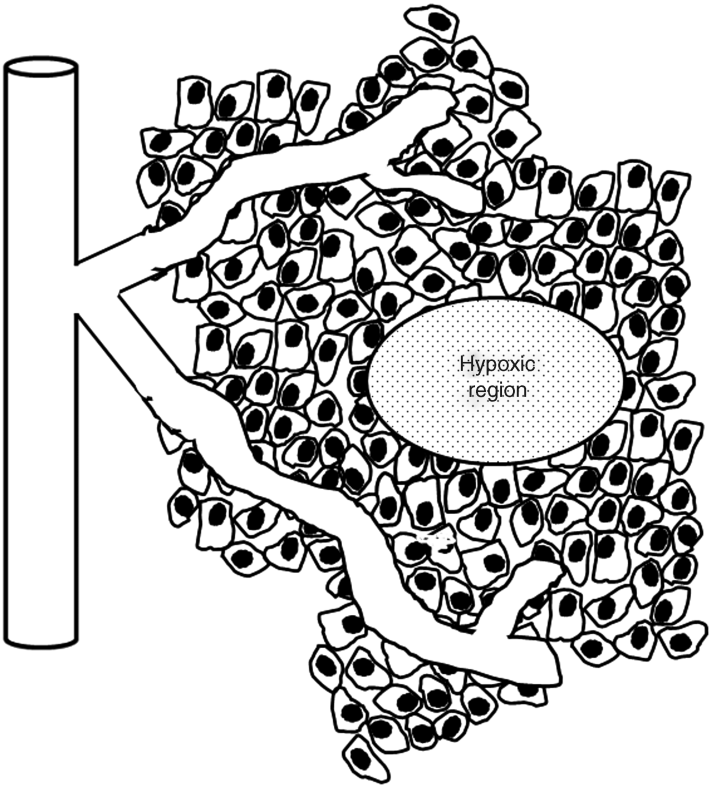


**FIGURE 10.2** Angiogenesis originates from existing blood vessels outside the tumor. New blood vessels tend to be highly irregular and disorganized; they also permeate primarily at the tumor periphery rather than penetrating directly into the tumor interior. This leaves some cells distant from the nearest capillaries creating hypoxic regions in the interior regions of the tumor. Also note that the pressure gradient runs toward the outside of the tumor, so a slight convective flow is expected to run away from the tumor core.

found in the liver, pancreas, kidney, colon, testes, and jejunum of healthy tissues and play an important role in protecting and detoxifying cells commonly exposed to toxic environments [36,37]. It is part of a large family of proteins with similar functions encoded by the MDR1 gene known as the ATP-binding cassette (ABC) family [38]. Pgp can transport a large variety of unrelated compounds from the membrane to the cell exterior and is capable of maintaining large concentration gradients across the cell membrane [39,40].

Another mechanism for drug efflux not related to the ABC pumps is the lung resistance-related protein (LRP) most commonly found in lung carcinomas [41]. LRP is not a membrane pump, but rather a major vault protein. Vaults are thought to be associated with the nuclear pore complex and regulate entry into the nucleus. Drugs can be sequestered into a vault at the nuclear pore complex and then transported into the cytoplasm where it is subsequently exocytosed [42].

Many drugs may also be sequestered into a number of intracellular compartments especially the mitochondria, Golgi, and endosomes. Drugs collect into these compartments by pH partitioning and active pumping in the case of endosomal vesicles and by increased solubility in membrane-rich organelles such as the Golgi and mitochondria [43]. The drug sequestration can be quite marked due to large solubility differences [44]. These drugs are held away from their sites of action until they can be either metabolized or removed from the cell entirely via exocytosis.

Cells can gain resistance to drugs targeting specific signaling pathways by modifying part of the targeted chemical pathway to compensate for the drug's action. The most commonly encountered forms of chemical resistance modulate key pathways targeted by cancer therapeutics such as the apoptosis pathway. Almost all cancers must shut down the protein p53, which is a key choke point in apoptosis [45]. The cell is rendered more resistant by upregulation of antiapoptosis factors such as bcl2 [46,47]. DNA repair mechanisms may also be upregulated to weaken apoptotic signals triggered by damaged DNA [48].

Resistance may also be conferred by increasing the cellular metabolism of the drug [49]. Glutathione is a key component of the glutathione detoxification system, which can be used to metabolize drugs in the cytoplasm [50]. It operates by binding to the toxic substance by means of a sulfur bond [51]. This mechanism can confer resistance to a number of anticancer compounds [52–55]. It can also act as an antioxidant, which would confer resistance toward drugs designed to generate free radicals [51].

### 10.2.3 Perspective on Drug Resistance

This review of MDR mechanisms is not meant to be comprehensive and does not provide depth on any of the mechanisms described. The intent of this brief description of MDR mechanisms is to provide perspective on the breadth of the challenge in designing a drug delivery system to overcome the diverse mechanisms of resistance. Any approach to combat MDR that only takes a single mechanism into account is unlikely to succeed in a clinical setting. Oftentimes, a delivery strategy targeted to overcome one MDR mechanism will leave it especially vulnerable to another. For example, using large particles as drug carriers will help increase circulation time and accumulation in the tumor; however, once in the tumor the diffusion of large particles is severely limited and is unlikely to reach the majority of tumor cells. Large particles become helpful again as a way to circumvent Pgp pumps because they can cross the cell membrane via endocytosis rather than diffusion. All drug delivery systems must compromise between these countervailing effects.

Different cells within the same tumor will have different phenotypes, express Pgp at varying levels, have different apoptotic thresholds, or metabolize toxins at different rates [56]. In the face of treatment, natural selection favors resistant cell types upon tumor recurrence [57]. Any plausible technology to combat MDR should account for the diversity of mechanisms present in a single tumor and seek to combat the cancer as holistically as possible. Otherwise, each successive treatment will increase the risk of a more resistant recurrent tumor.

## 10.3 COMBATING MDR

The diversity of MDR mechanisms requires that any drug delivery system designed to combat MDR be extremely versatile. Nanoparticles are among the most versatile tools available to researcher for combating cancer [58]. Platforms for nanotherapeutics can take many forms as shown in Figure 10.3. The different classes of these nanoparticles include dendrimers [59], linear polymeric carriers [60], inorganic nanoparticles [61], micelles [62], polymersomes [63], and liposomes [64]. Attaching drugs to or encapsulating them in nanocarriers helps solubilize and stabilize most drug formulations [65]. The large particle size of drug delivery systems excludes clearance by renal filtration, which increases circulation time, allowing for drug

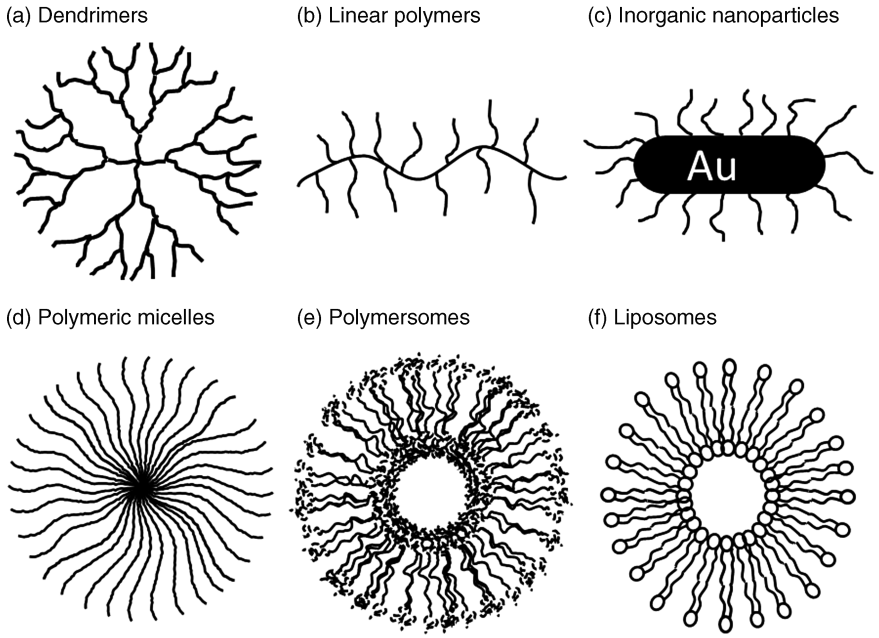


**FIGURE 10.3** Various physical forms of nanoparticles used in cancer therapy. (a) Dendrimers are highly branched polymers that form into a tree or bush-like structure. (b) Linear macromolecular carriers are composed of a degradable polymer backbone to which drugs or targeting moieties can be attached. (c) Inorganic nanoparticles can be made from various metals or other compounds such as SiO; the surface of these particles is often modified with PEG, targeting moieties or drugs. (d) Micelles are often composed of block copolymers bearing a hydrophilic and hydrophobic region. The hydrophobic region of these polymers interacts in aqueous environments forming a stable core into which drugs can be loaded. The hydrophilic region of the polymer can be modified in other ways. (e) Polymersomes are generally made from triblock copolymers that assemble to form a vesicular structure with a hydrophilic core. Compounds may be loaded into both the hydrophilic and hydrophobic regions. (f) Liposomes are similar in structure to polymersomes but are made using phospholipids similar to those in natural lipid bilayer membranes.

accumulation through the EPR effect. The circulation time of nano-sized drug carriers can be further enhanced by modifying the surface to inhibit MPS uptake. The most common and effective surface modification for increasing circulation time is grafting PEG on the particle surface. PEG is a highly hydrophilic polymer that extends out and sterically prevents proteins from adsorbing the particle surface and MPS uptake [66]. Other modifications can also be added to a nanoparticle to confer the desired properties. Targeting moieties are a common modification designed to increase tumor localization and cellular uptake [67]. Drug can also be attached to the surface of the particle or be stored in the internal hydrophobic or hydrophilic compartments. This versatility makes co-delivery of multiple compounds with different characteristics possible [68]. In addition to drug compounds, dyes or other imaging agents can also be included with the drug carrier. This provides the capability to image the tumor during the course of treatment, a capability known as theranostics [69]. The ability of nanotherapeutics to carry on such a broad range of function is what brings it to the forefront of modern cancer research.

Long circulating nanoparticles are, in most instances, able to accumulate preferentially in the tumor due to the EPR effect [70,16]. From the point of entry into the tumor, however, the strategies of drug delivery may vary tremendously. The majority of methods are targeted to combating ABC pumps such as Pgp, which are generally considered the principal mechanism of MDR [11]. Efflux from ABC pumps is sometimes even described as synonymous with MDR [71,72].

The main strategies for combating the ABC pump mechanism of multiple drug resistance are to either circumvent it by bypassing membrane diffusion or by delivering Pgp disabling compounds with the drug. The capacity of a nanoparticle to be a platform for multiple compounds provides the capability to combat other MDR mechanisms. Co-delivery of anticancer drugs with apoptosis modulators can be used to increase drug sensitivity, and co-delivery of the nanoparticles with collagenases can be used to enhance drug penetration into the tumor.

### 10.3.1 Bypassing MDR by Endocytosis

ABC pumps are membrane-embedded proteins which capture and then efflux small molecule compounds while dissolved in the membrane. The process of diffusing across the membrane exposes small molecular weight drugs to efflux [73]. Encapsulating the drug in a large particle allows bulk uptake through endocytic pathways, thus bypassing MDR pumps such as Pgp. Bringing the intact drug carrier into the cytoplasm through endocytosis and phagocytosis enables the drug to be released directly into the cytoplasm resulting in a higher sustained cytoplasmic concentration [74].

A major challenge to achieving sufficiently cytotoxic-free drug concentrations is for the cell to uptake a sufficient quantity of drug carriers. The rate of endocytosis is dependent on a number of factors including size, shape, and particle–membrane interactions. Optimal particle uptake appears to occur for particles in the 30–50 nm diameter range [75,76]. The aspect ratio of the particles also plays a role in active uptake by the cell. Studies done with silica nanoparticles have shown that slightly

elongated particles (with an aspect ratio of 2.1–2.5) show the highest cellular uptake rates [77,78].

Active targeting improves cellular uptake by conjugating ligands to the nanoparticle, which will interact with receptors on the cell membrane. These targeting moieties can increase the frequency and strength of membrane interactions, giving greater likelihood of internalization. Targeting moieties are often antibodies against proteins that are generally overexpressed on cancer cells. A broad range of antibodies can be used depending on the cancer type [79,80]. Among the most common targets for active targeting are folate receptors [81] and Her2 receptor [82], which are overexpressed in many cancers. Targeting to specific protein receptors suffers from the differential expression of these receptors within a single tumor. An example of this phenomenon is the drug Trastuzumab (Herceptin®) which targets the Her2 antigen which is overexpressed in some 20–30% breast cancer patients. The use of this treatment governed by the immunohistochemistry (IHC) scoring system which rates the overall Her2 expression in a tumor. The highest score under this system is assigned to tumors in which only 10% of tumor cells stain strongly for Her2. Among patients with the highest levels of Her2 expression, only half appear to benefit from Herceptin targeted treatment [83]. Other targeted cancer treatments also must labor under this burden of differential protein expression in cancer cells.

Nonspecific targeting can also be used to enhance cellular uptake. The cell penetrating peptide TAT is associated with HIV infectivity and allows it to infect a wide variety of cells [84]. The effectiveness of TAT at entering cells has led to its use as a means of encouraging endocytosis in drug delivery [85]. TAT works in an enormous variety of cells, but this lack of specificity risks increasing the nonspecific toxicity of the drug [86]. Specificity can be increased (and negative side effects reduced) by shielding the TAT until the drug carrier reaches the tumor site. Deshielding can be tied to environmental stimuli unique to the tumor, such as acidity or hypoxia [87].

The effectiveness of a drug carrier is in part dependent on the mechanism by which it is internalized. Nanoparticles can enter the cell in a variety of ways including diffusion, clathrin-mediated endocytosis, caveolae-mediated endocytosis, and pinocytosis [88]. The means by which the particle enters the cell will affect the drug release and intracellular distribution of the drug. Particles that diffuse across the membrane will emerge near the membrane on the intracellular side. Drug release then is likely to occur close to the membrane where the drug molecules can then reenter the membrane by diffusion and be effluxed [89]. Most drug carriers are actively brought into the cell through either endocytosis or phagocytosis and then enter the endocytic pathway [90]. The lysosome is the end stage of the endocytic pathway and presents a severe problem for drug release. The harsh, acidic, and degradative environment in the lysosome is designed to degrade any phagocytosed particles before they are allowed entry into the cytosol to prevent harm to the cell [91].

Avoiding the lysosome is critical for effective delivery from endocytosed particles. Caveolae-mediated endocytosis operates with a unique pathway that may not involve the lysosome [92,93], suggesting a strategy for circumventing drug destruction by targeting the caveolae-mediated pathway [94]. Targeting a single endocytic

pathway can be challenging; pinocytosis is a common uptake mechanism and is nonspecific. Unfortunately, the data was unclear as to whether the lysosomal phase is actually avoided [95,96]. If the lysosome is the end result of all endocytic pathways, it is necessary to design the drug delivery system to exit the endosome before they can merge with the lysosome. pH-sensitive micelles provide one strategy for endosomal escape [97,98]. Micelles can be tuned to break up at the endosomal pH releasing the contents into the vesicle. The sudden release of solute drug and micelle polymer in the endosome causes a spike in osmotic pressure driving water to rush into the vesicle. The vesicle may then swell to the breaking point, emptying the contents of the vesicle directly into the cytoplasm.

### 10.3.2 Co-Delivery of Drugs with Other Compounds

One of the major strengths of using nanoparticles as a platform for drug delivery is the ability to deliver finely tuned doses of multiple drugs into the local environment. Many nanoparticle drug formulations have been designed and studied to target MDR by co-delivering chemotherapeutic agents with chemosensitizers, oligonucleotides, proteins, and so on. Co-delivery of drugs with other compounds can allow synergistic cooperation, maximizing the effect of the delivered drug.

Drugs can be incorporated in nanoparticles in a variety of ways, which can enable various strategies for combinational drug therapy [99]. Drugs can be covalently linked to the nanoparticle or linear polymer chains, often by a degradable linker which is tuned to release the drug at the desired time [100]. Often, the drug is stored in hydrophilic and/or hydrophobic compartments of the nanoparticle. Micelles often store drugs in their hydrophobic core, and liposomes and polymersomes can store dissolved hydrophobic drug in the membrane and can additionally store hydrophilic drugs in the aqueous center [68]. Drug can also be physically entrapped in a polymer mesh such as in a dendrimer or hydrogel [101,102].

Co-delivery of drug with a Pgp modulator is a common strategy to combat ABC pump-mediated MDR [103]. This strategy has been employed for more than 30 years since Pgp was identified as a major MDR mechanism. Several generations of ABC pump modulators have been developed to deal with the weaknesses of the first modulators [104]. Use of these modulators has effectively increased the intracellular concentration and drug sensitivity of the cells [105]. On the other hand, ABC pumps are also important features of cells in healthy tissues, which are routinely exposed to toxins such as the liver, kidneys, and intestines [36]. Therefore, using Pgp modulators leads to nonspecific toxicity issues that limit the clinical relevance of the method [106]. Third generation modulators, currently being developed, have been shown to more effectively sensitize cells while maintaining tolerable toxicity [107,108]. Drugs delivered in nanoparticles collect in the liver, kidney, and spleen where ABC pumps play an important role in detoxifying healthy tissues. This makes ABC pump modulation a less than ideal tactic unless true tumor targeting can be achieved.

One important strategy to mitigate emergence of MDR is to deliver drugs of very different structures and functions simultaneously [68]. The ability to incorporate

drugs in a nanoparticle platform by several mechanisms makes it possible to deliver disparate drugs. This strategy is used in agriculture, where applying multiple pesticides to a field simultaneously slows evolution by creating divergent pressures and mitigating the advantage of individuals with resistant traits [109]. Furthermore, well-chosen drug combinations can take advantage of synergistic effects among drugs to increase sensitivity of the target cells allowing the total amount of drug to be minimized [110].

The sensitivity of cells can be greatly increased by co-delivering apoptosis modulators with the drug. Apoptosis modulators can take the form of antisense oligonucleotides or other chemical compounds such as ceramide [111]. The antisense oligonucleotides generally work by knocking down expression of proteins, such as bcl2, which inhibit the apoptosis pathway. The resulting drug combination can be as much as 100-fold more potent than using drug alone [112]. Of course the apoptosis pathway is common to all cells, so carefully targeted drug release is again a challenge as lowering the apoptotic threshold in normal cells could lead to tissue damage.

Nanoparticles can also be used to modify the tumor microenvironment and improve drug distribution through the tissue. The most commonly explored means of improving drug distribution at the tissue and microenvironmental levels is to deliver collagenases with the nanoparticles, which can degrade the collagen matrix of the tumor tissue [113,31]. Degrading the collagen matrix has a twofold benefit. The most obvious is that the dense network of protein fibers can be a significant barrier to particle diffusion, especially the relatively large particles used as drug delivery systems. Drug bound nanoparticles fail to move more than a few microns from the point of extravasation into the tumor. The collagen matrix also contributes to the elevated interstitial pressure observed in most tumors by preventing volume expansion and providing resistance to bulk fluid flow [30]. The interstitial pressure in the tumor can approach that of the microvasculature, thus moving the convective driving force for extravasation of the capillary into the tumor and reducing overall tumor drug accumulation [24,114,115]. Degradation of the tumor protein matrix temporarily lowers the oncotic pressure of the tumor and restores the driving force for extravasation [115]. Collagen matrix degradation has also been shown to improve diffusion within the tumor tissue [31], thus improving overall distribution. One concern with using this approach is that a host of cells may be released and metastasize. Further study is needed to evaluate this risk.

### 10.3.3 Stimuli-Based Methods

Nanoparticles may be designed to respond to a variety of stimuli. Micelles sensitive to surrounding pH were discussed as a means to deliver drug directly to the cytoplasm of the cell in a previous section [98]. Nanoparticles may also be designed to respond to external stimuli such as radiation, ultrasound, and temperature, which can then be used in the clinic to localize drug release, sensitize cells, or directly kill cells [116–118]. Cancer treatments relying on external stimuli have some tremendous advantages over more passive treatment methods. The main

advantage of using external stimuli as a trigger for treatment is the level of control it offers the physician. The therapeutic effect occurs precisely where and when the physician determines is appropriate.

Photothermal ablation is a form of external-stimuli-based nanotherapy and has an additional advantage of not requiring drug sensitive tumor cells. Treatment by photothermal ablation is accomplished by administering nanoparticles to the patient, then focusing particular wavelengths of light on a tumor. The light is then absorbed by the nanoparticles, causing heating of the surrounding tissue [119]. Ablative therapy can result in local temperatures within the tissue approaching the boiling point of water [120]. This gives ablative therapy the advantage of not requiring cellular entry, and the specificity of the treatment can be consciously controlled by the practitioner. Particle distribution can be a significant challenge just as for drug-based therapies; however, heat diffusion may be less hindered than particle diffusion depending on the tumor type and composition [121]. The effectiveness of the therapy depends in part on the ability to predict and control the heat distribution, which generally requires high-level mathematical modeling of the tumor heat profile during treatment [122,123]. Achieving sufficient tissue penetration to treat deep tumors is also a distinct challenge for ablation therapy, but a broad range of stimuli have been shown to accomplish therapeutic heat ranges with different particles providing a large range of treatment options [116,123,124]. In situations where necrosis is undesirable, more gentle heating (in the 45°C range) can be used to sensitize cells to traditional drug therapies [125]. Cell sensitization in hyperthermic conditions results from the inhibition of intracellular repair mechanisms due to heat shock [126].

There exist other forms of external-stimuli-based therapies designed to improve therapeutic efficacy. This includes using ferromagnetic particles, which can actively move deeper into a tumor in response to a magnetic field [127], and ultrasound-sensitive drug packets, which can be released specifically at a specific site with focused ultrasound [117]. Many of these therapies have theranostic potential, meaning that the particles used can be simultaneously used as imaging contrast agents to visualize the tumor even as the treatment is occurring [128].

Unfortunately, these methods are only able to treat "visible" tumors, or tumors that are large enough and can collect enough particles to stand out against the background noise for imaging. Smaller tumors and metastatic cells must be treated by more traditional drug-based therapies. Also, the inorganic nanoparticles, which are frequently used for stimuli-based therapies, are not readily cleared by the body and may accumulate permanently in the liver and spleen posing serious long-term toxicity issues [129].

## 10.4 FUTURE PERSPECTIVES

MDR is a grave development in the lifetime of a tumor. MDR tumors are generally composed of more resilient and potentially more metastatic cells, which lead to a poor prognosis for a patient. The diversity of mechanisms contributing to MDR

creates a difficult problem in designing effective strategies to treat resistant cancers. Research has revealed an impressive arsenal of technologies capable of combating individual resistance mechanisms. Unfortunately, the multimodal nature of resistance can severely limit the effectiveness of narrowly designed drug delivery systems. This in part explains the limited success of modern therapies in clinical settings. The next step in drug delivery to resistant tumors will require a more holistic view of distribution and resistance, accounting for each stage of delivery from circulatory distribution to efficacy at the site of action. This approach may require trade-offs to cope with the exacting demands of tumor delivery and distribution, cellular uptake, and drug efficacy. Designing drug delivery systems from a comprehensive perspective requires a more profound understanding of the interplay between the MDR mechanisms and drug delivery.

**Acknowledgments**

This work is supported by NIH CA 101850.

## REFERENCES

1. Howlader, N. et al. *SEER Cancer Statistics Review, 1975–2009*. National Cancer Institute, Bethesda, MD, 2010.
2. American Cancer Society. *Cancer Facts & Figures 2012*. American Cancer Society, Atlanta. 2012, p. 18.
3. Ling, V. (1997). Multidrug resistance: molecular mechanisms and clinical relevance. *Cancer Chemotherapy and Pharmacology 40*, 3–8.
4. Raderer, M., Scheithaur, W. (1993). Clinical trials of agents that reverse multidrug resistance, a literature review. *Cancer 72*, 3553–3563.
5. Wind, N. S., Holen, I. (2011). Multidrug resistance in breast cancer: from *in vitro* models to clinical studies. *International Journal of Breast Cancer 2011*, 1–12.
6. Szakács, G., Paterson, J. K., Ludwig, J. A., Booth-Genthe, C., Gottesman, M. M. (2006). Targeting multidrug resistance in cancer. *Nature Reviews Drug Discovery 5*, 219–234.
7. Gottesman, M. (1993). How cancer cells evade chemotherapy: sixteenth Richard and Hinda Rosenthal Foundation Award Lecture. *Cancer Research 53*, 747–754.
8. Gottesman, M. M., Fojo, T., Bates, S. E. (2002). Multidrug resistance in cancer: role of ATP–dependent transporters. *Nature Reviews Cancer 2*, 48–58.
9. Sotiriou, C. (2003). Breast cancer classification and prognosis based on gene expression profiles from a population-based study. *Proceedings of the National Academy of Sciences 100*, 10393–10398.
10. Mikhail, A. S., Allen, C. (2009). Block copolymer micelles for delivery of cancer therapy: transport at the whole body, tissue and cellular levels. *Journal of Controlled Release 138*, 214–223.
11. Bailar, J. C. (1987). Rethinking the war on cancer. *Issues in Science and Technology 4*, 16–21.

12. Choi, H. S. et al. (2007). Renal clearance of quantum dots. *Nature Biotechnology 25*, 1165–1170.
13. Hume, D. (2006). The mononuclear phagocyte system. *Current Opinion in Immunology 18*, 49–53.
14. Matsumura, Y., Maeda, H. (1986). A new concept for macromolecular therapeutics in cancer chemotherapy: mechanism of tumoritropic accumulation of proteins and the antitumor agent smancs. *Cancer Research 46*, 6387–6392.
15. Maeda, H. (2001). SMANCS and polymer-conjugated macromolecular drugs: advantages in cancer chemotherapy. *Advanced Drug Delivery Reviews 46*, 169–185.
16. Boucher, Y., Salehi, H., Witwer, B., Harsh, G. R., Jain, R. K. (1997). Interstitial fluid pressure in intracranial tumours in patients and in rodents. *British Journal of Cancer 75*, 829–836.
17. Gabizon, A., Shmeeda, H., Horowitz, A. T., Zalipsky, S. (2004). Tumor cell targeting of liposome-entrapped drugs with phospholipid-anchored folic acid–PEG conjugates. *Advanced Drug Delivery Reviews 56*, 1177–1192.
18. Woodle, M. C. (1995). Sterically stabilized liposome therapeutics. *Advanced Drug Delivery Reviews 16*, 249–265.
19. Dreher, M. R. et al. (2006). Tumor vascular permeability, accumulation, and penetration of macromolecular drug carriers. *Journal of the National Cancer Institute 98*, 335–344.
20. Sylven, B., Bois, I. (1960). Protein content and enzymatic assays of interstitial fluid from some normal tissues and transplanted mouse tumors. *Cancer Research 20*, 831–836.
21. Roose, T., Netti, P. A., Munn, L. L., Boucher, Y., Jain, R. K. (2003). Solid stress generated by spheroid growth estimated using a linear poroelasticity model. *Microvascular Research 66*, 204–212.
22. Venn, M., Maroudas, A. (1977). Chemical composition and swelling of normal and osteoarthrotic femoral head cartilage. I. Chemical composition. *Annals of the Rheumatic Diseases 36*, 121–129.
23. Boucher, Y., Baxter, L. T., Jain, R. K. (1990). Interstitial pressure gradients in tissue-isolated and subcutaneous tumors: implications for therapy. *Cancer Research 50*, 4478–4484.
24. Boucher, Y., Jain, R. K. (1992). Microvascular pressure is the principal driving force for interstitial hypertension in solid tumors: implications for vascular collapse. *Cancer Research 52*, 5110–5114.
25. Fukumura, D., Duda, D. G., Munn, L. L., Jain, R. K. (2010). Tumor microvasculature and microenvironment: novel insights through intravital imaging in pre-clinical models. *Microcirculation (New York, NY, 1994) 17*, 206–225.
26. Holash, J., Wiegand, S. J., Yancopoulos, G. D. (1999). New model of tumor angiogenesis: dynamic balance between vessel regression and growth mediated by angiopoietins and VEGF. *Oncogene 18*, 5356–5362.
27. Schofield, J. W., Gaffney, E. A., Gatenby, R. A., Maini, P. K. (2011). Tumour angiogenesis: the gap between theory and experiments. *Journal of Theoretical Biology 274*, 97–102.
28. Primeau, A. J., Rendon, A., Hedley, D., Lilge, L., Tannock, I. F. (2005). The distribution of the anticancer drug Doxorubicin in relation to blood vessels in solid tumors. *Clinical Cancer Research 11*, 8782–8788.

29. Ameri, K. et al. (2010). Circulating tumour cells demonstrate an altered response to hypoxia and an aggressive phenotype. *British Journal of Cancer 102*, 561–569.
30. Netti, P. A., Berk, D. A., Swartz, M. A., Grodzinsky, A. J., Jain, R. K. (2000). Role of extracellular matrix assembly in interstitial transport in solid tumors. *Cancer Research 60*, 2497–2503.
31. Ramanujan, S. et al. (2002). Diffusion and convection in collagen gels: implications for transport in the tumor interstitium. *Biophysical Journal 83*, 1650–1660.
32. Sharom, F. J. (1997). The P-glycoprotein efflux pump: how does it transport drugs? *The Journal of Membrane Biology 160*, 161–175.
33. Eckford, P. D. W., Sharom, F. J. (2009). ABC efflux pump-based resistance to chemotherapy drugs. *Chemical Reviews 109*, 2989–3011.
34. Endicott, J. A., Ling, V. (1989). The biochemistry of P-glycoprotein-mediated multidrug resistance. *Annual Review of Biochemistry 58*, 137–171.
35. Nielsen, D., Skovsgaard, T. (1992). P-glycoprotein as multidrug transporter: a critical review of current multidrug resistant cell lines. *Biochimica et Biophysica Acta (BBA)—Molecular Basis of Disease 1139*, 169–183.
36. Thiebaut, F. et al. (1987). Cellular localization of the multidrug-resistance gene product P-glycoprotein in normal human tissues. *Proceedings of the National Academy of Sciences of the United States of America 84*, 7735–7738.
37. Tanigawara, Y. (2000). Role of P-glycoprotein in drug disposition. *Therapeutic Drug Monitoring 22*, 137–140.
38. Dean, M., Hamon, Y., Chimini, G. (2002). The human ATP-binding cassette (ABC) transporter superfamily. *The Human ATP-Binding Cassette (ABC) Transporter Superfamily 42*, 1007–1017.
39. Seelig, A. (1998). A general pattern for substrate recognition by P-glycoprotein. *European Journal of Biochemistry/FEBS 251*, 252–261.
40. Mayer, L. D., Bally, M. B., Hope, M. J., Cullis, P. R. (1985). Uptake of antineoplastic agents into large unilamellar vesicles in response to a membrane potential. *Biochimica et Biophysica Acta 816*, 294–302.
41. Izquierdo, M. A. et al. (1996). Broad distribution of the multidrug resistance-related vault lung resistance protein in normal human tissues and tumors. *The American Journal of Pathology 148*, 877–887.
42. Chugani, D. C., Rome, L. H., Kedersha, N. L. (1993). Evidence that vault ribonucleoprotein particles localize to the nuclear pore complex. *Journal of Cell Science 106*(Pt 1), 23–29.
43. Duvvuri, M., Krise, J. P. (2005). Intracellular drug sequestration events associated with the emergence of multidrug resistance: a mechanistic review. *Frontiers in Bioscience 10*, 1499–1509.
44. Hurwitz, S. J., Terashima, M., Mizunuma, N., Slapak, C. A. (1997). Vesicular anthracycline accumulation in doxorubicin-selected U-937 cells: participation of lysosomes. *Blood 89*, 3745–3754.
45. Kerr, J. F., Winterford, C. M., Harmon, B. V. (1994). Apoptosis. Its significance in cancer and cancer therapy. *Cancer 73*, 2013–2026.
46. Dole, M. G. et al. (1995). Bcl-xL is expressed in neuroblastoma cells and modulates Bcl-xL is expressed in neuroblastoma cells and modulates chemotherapy-induced apoptosis. *Cancer Research 55*, 2576–2582.

47. Teixeira, C., Reed, J., Pratt, M. (1995). Estrogen promotes chemotherapeutic drug resistance by a mechanism involving Bcl-2 proto-oncogene expression in human breast cancer cells. *Cancer Research 55*, 3902–3907.
48. Masuda, H. et al. (1988). Increased DNA repair as a mechanism of acquired resistance to cis-diamminedichloroplatinum (II) in human ovarian cancer cell lines. *Cancer Research 48*, 5713–5716.
49. Tew, K. D. (1994). Glutathione-associated enzymes in anticancer drug resistance. *Cancer Research 54*, 4313–4320.
50. Raijmakers, M. T. M., Bruggeman, S. W. M., Steegers, E. A. P., Peters, W. H. M. (2002). Distribution of components of the glutathione detoxification system across the human placenta after uncomplicated vaginal deliveries. *Placenta 23*, 490–496.
51. Townsend, D. M., Tew, K. D. (2003). The role of glutathione-S-transferase in anti-cancer drug resistance. *Oncogene 22*, 7369–7375.
52. Lewis, A. D. et al. (1988). Amplification and increased expression of alpha class glutathione S-transferase-encoding genes associated with resistance to nitrogen mustards. *Proceedings of the National Academy of Sciences of the United States of America 85*, 8511–8515.
53. Black, S. M. et al. (1990). Expression of human glutathione S-transferases in *Saccharomyces cerevisiae* confers resistance to the anticancer drugs adriamycin and chlorambucil. *The Biochemical Journal 268*, 309–315.
54. Clapper, M., Hoffman, S., Tew, K. (1990). Sensitization of colon tumor xenografts to L-phenylalanine mustard using ethacrynic acid. *Journal of Cell Pharmacology 1*, 71–78.
55. Efferth, T., Volm, M. (2005). Glutathione-related enzymes contribute to resistance of tumor cells and low toxicity in normal organs to artesunate. *In Vivo (Athens, Greece) 19*, 225–232.
56. Heppner, G., Dexter, D., DeNucci, T., Calabresi, P. (1978). Heterogeneity in drug sensitivity among tumor cell subpopulations of a single mammary tumor. *Cancer Research 38*, 3758–3763.
57. Richardson, M. E., Siemann, D. W. (1997). Tumor cell heterogeneity: impact on mechanisms of therapeutic drug resistance. *International Journal of Radiation Oncology, Biology, Physics 39*, 789–795.
58. Gindy, M. E., Prud'homme, R. K. (2009). Multifunctional nanoparticles for imaging, delivery and targeting in cancer therapy. *Expert Opinion on Drug Delivery 6*, 865–878.
59. Gillies, E. R., Fréchet, J. M. J. (2005). Dendrimers and dendritic polymers in drug delivery. *Drug Discovery Today 10*, 35–43.
60. Minko, T., Kopecková, P., Kopecek, J. (2000). Efficacy of the chemotherapeutic action of HPMA copolymer-bound doxorubicin in a solid tumor model of ovarian carcinoma. *International Journal of Cancer. Journal International du Cancer 86*, 108–117.
61. Ghosh, P., Han, G., De, M., Kim, C. K., Rotello, V. M. (2008). Gold nanoparticles in delivery applications. *Advanced Drug Delivery Reviews 60*, 1307–1315.
62. Nishiyama, N., Kataoka, K. (2006). Current state, achievements, and future prospects of polymeric micelles as nanocarriers for drug and gene delivery. *Pharmacology & Therapeutics 112*, 630–648.
63. Christian, D. A. et al. (2009). Polymersome carriers: from self-assembly to siRNA and protein therapeutics. *European Journal of Pharmaceutics and Biopharmaceutics:*

*Official Journal of Arbeitsgemeinschaft für Pharmazeutische Verfahrenstechnik e.V 71*, 463–474.

64. Gabizon, A. A. (1995). Liposome circulation time and tumor targeting: implications for cancer chemotherapy. *Advanced Drug Delivery Reviews 16*, 285–294.
65. Sutton, D., Nasongkla, N., Blanco, E., Gao, J. (2007). Functionalized micellar systems for cancer targeted drug delivery. *Pharmaceutical Research 24*, 1029–1046.
66. Li, S.-D., Huang, L. (2008). Pharmacokinetics and biodistribution of nanoparticles. *Molecular Pharmaceutics 5*, 496–504.
67. Farokhzad, O. C. et al. (2006). Targeted nanoparticle-aptamer bioconjugates for cancer chemotherapy *in vivo*. *Proceedings of the National Academy of Sciences of the United States of America 103*, 6315–6320.
68. Zhang, L. et al. (2007). Co-delivery of hydrophobic and hydrophilic drugs from nanoparticle-aptamer bioconjugates. *ChemMedChem 2*, 1268–1271.
69. Kim, K. et al. (2010). Tumor-homing multifunctional nanoparticles for cancer theragnosis: simultaneous diagnosis, drug delivery, and therapeutic monitoring. *Journal of Controlled Release 146*, 219–227.
70. Noguchi, Y. et al. (1998). Early phase tumor accumulation of macromolecules: a great difference in clearance rate between tumor and normal tissues. *Cancer Science 89*, 307–314.
71. Ozben, T. (2006). Mechanisms and strategies to overcome multiple drug resistance in cancer. *FEBS Letters 580*, 2903–2909.
72. Choi, C.-H. (2005). ABC transporters as multidrug resistance mechanisms and the development of chemosensitizers for their reversal. *Cancer Cell International 5*, 30.
73. Romsicki, Y., Sharom, F. J. (1999). The membrane lipid environment modulates drug interactions with the P-glycoprotein multidrug transporter. *Biochemistry 38*, 6887–6896.
74. Wong, H. L. et al. (2006). A mechanistic study of enhanced doxorubicin uptake and retention in multidrug resistant breast cancer cells using a polymer-lipid hybrid nanoparticle system. *The Journal of Pharmacology and Experimental Therapeutics 317*, 1372–1381.
75. Zhang, S., Li, J., Lykotrafitis, G., Bao, G., Suresh, S. (2009). Size-dependent endocytosis of nanoparticles. *Advanced Materials (Deerfield Beach, Florida) 21*, 419–424.
76. Jiang, W., Kim, B. Y. S., Rutka, J. T., Chan, W. C. W. (2008). Nanoparticle-mediated cellular response is size-dependent. *Nature Nanotechnology 3*, 145–150.
77. Chithrani, B. D., Ghazani, A. A., Chan, W. C. W. (2006). Determining the size and shape dependence of gold nanoparticle uptake into mammalian cells. *Nano Letters 6*, 662–668.
78. Meng, H. et al. (2011). Aspect ratio determines the quantity of mesoporous silica nanoparticle uptake by a small GTPase-dependent macropinocytosis mechanism. *ACS Nano 5*, 4434–4447.
79. Cheever, M. A. et al. (2009). The prioritization of cancer antigens: a national cancer institute pilot project for the acceleration of translational research. *Clinical Cancer Research 15*, 5323–5337.
80. Schrama, D., Reisfeld, R. A., Becker, J. C. (2006). Antibody targeted drugs as cancer therapeutics. *Nature Reviews Drug Discovery 5*, 147–159.

81. Quintana, A. et al. (2002). Design and function of a dendrimer-based therapeutic nanodevice targeted to tumor cells through the folate receptor. *Pharmaceutical Research 19*, 1310–1316.
82. Liu, X., Wang, Y., Nakamura, K., Kubo, A., Hnatowich, D. J. (2008). Cell studies of a three-component antisense MORF/tat/Herceptin nanoparticle designed for improved tumor delivery. *Cancer Gene Therapy 15*, 126–132.
83. Vogel, C. L. et al. (2002). Efficacy and safety of trastuzumab as a single agent in first-line treatment of HER2-overexpressing metastatic breast cancer. *Journal of Clinical Oncology 20*, 719–726.
84. Vogel, J., Hinrichs, S. H., Reynolds, R. K., Luciw, P. A., Jay, G. (1988). The HIV tat gene induces dermal lesions resembling Kaposi's sarcoma in transgenic mice. *Nature 335*, 606–611.
85. Sethuraman, V. A., Bae, Y. H. (2007). TAT peptide-based micelle system for potential active targeting of anti-cancer agents to acidic solid tumors. *Journal of Controlled Release 118*, 216–224.
86. Mann, D. A., Frankel, A. D. (1991). Endocytosis and targeting of exogenous HIV-1 Tat protein. *The EMBO Journal 10*, 1733–1739.
87. van Vlerken, L. E., Vyas, T. K., Amiji, M. M. (2007). Poly(ethylene glycol)-modified nanocarriers for tumor-targeted and intracellular delivery. *Pharmaceutical Research 24*, 1405–1414.
88. Thurn, K. T. et al. (2011). Endocytosis of titanium dioxide nanoparticles in prostate cancer PC-3M cells. *Nanomedicine: Nanotechnology, Biology, and Medicine 7*, 123–130.
89. Colin de Verdière, A. et al. (1994). Uptake of doxorubicin from loaded nanoparticles in multidrug-resistant leukemic murine cells. *Cancer Chemotherapy and Pharmacology 33*, 504–508.
90. Khalil, I. A., Kogure, K., Akita, H., Harashima, H. (2006). Uptake pathways and subsequent intracellular trafficking in nonviral gene delivery. *Pharmacological Reviews 58*, 32–45.
91. Trouet, A., Deprez-de Campeneere, D., De Duve, C. (1972). Chemotherapy through lysosomes with a DNA-daunorubicin complex. *Nature: New Biology 239*, 110–112.
92. Ferrari, A. et al. (2003). Caveolae-mediated internalization of extracellular HIV-1 tat fusion proteins visualized in real time. *Molecular Therapy 8*, 284–294.
93. Le, P. U., Nabi, I. R. (2003). Distinct caveolae-mediated endocytic pathways target the Golgi apparatus and the endoplasmic reticulum. *Journal of Cell Science 116*, 1059–1071.
94. Gabrielson, N. P., Pack, D. W. (2009). Efficient polyethylenimine-mediated gene delivery proceeds via a caveolar pathway in HeLa cells. *Journal of Controlled Release 136*, 54–61.
95. Kiss, A. L., Botos, E. (2009). Endocytosis via caveolae: alternative pathway with distinct cellular compartments to avoid lysosomal degradation? *Journal of Cellular and Molecular Medicine 13*, 1228–1237.
96. Tran, D., Carpentier, J. L., Sawano, F., Gorden, P., Orci, L. (1987). Ligands internalized through coated or noncoated invaginations follow a common intracellular pathway. *Proceedings of the National Academy of Sciences of the United States of America 84*, 7957–7961.

97. Oishi, M., Kataoka, K., Nagasaki, Y. (2006). pH-responsive three-layered PEGylated polyplex micelle based on a lactosylated ABC triblock copolymer as a targetable and endosome-disruptive nonviral gene vector. *Bioconjugate Chemistry 17*, 677–688.
98. Mohajer, G., Lee, E. S., Bae, Y. H. (2007). Enhanced intercellular retention activity of novel pH-sensitive polymeric micelles in wild and multidrug resistant MCF-7 cells. *Pharmaceutical Research 24*, 1618–1627.
99. Cho, K., Wang, X., Nie, S., Chen, Z. G., Shin, D. M. (2008). Therapeutic nanoparticles for drug delivery in cancer. *Clinical Cancer Research 14*, 1310–1316.
100. Minko, T., Kopecková, P., Kopecek, J. (1999). Comparison of the anticancer effect of free and HPMA copolymer-bound adriamycin in human ovarian carcinoma cells. *Pharmaceutical Research 16*, 986–996.
101. Na, K., Park, K. H., Kim, S. W., Bae, Y. H. (2000). Self-assembled hydrogel nanoparticles from curdlan derivatives: characterization, anti-cancer drug release and interaction with a hepatoma cell line (HepG2). *Journal of Controlled Release 69*, 225–236.
102. D'Emanuele, A., Attwood, D. (2005). Dendrimer-drug interactions. *Advanced Drug Delivery Reviews 57*, 2147–2162.
103. Patil, Y., Sadhukha, T., Ma, L., Panyam, J. (2009). Nanoparticle-mediated simultaneous and targeted delivery of paclitaxel and tariquidar overcomes tumor drug resistance. *Journal of Controlled Release 136*, 21–29.
104. Modok, S., Mellor, H. R., Callaghan, R. (2006). Modulation of multidrug resistance efflux pump activity to overcome chemoresistance in cancer. *Current Opinion in Pharmacology 6*, 350–354.
105. Wu, J. et al. (2007). Reversal of multidrug resistance by transferrin-conjugated liposomes co-encapsulating doxorubicin and verapamil. *Journal of Pharmacy & Pharmaceutical Sciences: A Publication of the Canadian Society for Pharmaceutical Sciences, Société Canadienne des Sciences Pharmaceutiques 10*, 350–357.
106. Krishna, R., Mayer, L. D. (2000). Multidrug resistance (MDR) in cancer. Mechanisms, reversal using modulators of MDR and the role of MDR modulators in influencing the pharmacokinetics of anticancer drugs. *European Journal of Pharmaceutical Sciences: Official Journal of the European Federation for Pharmaceutical Sciences 11*, 265–283.
107. Pusztai, L. et al. (2005). Phase II study of tariquidar, a selective P-glycoprotein inhibitor, in patients with chemotherapy-resistant, advanced breast carcinoma. *Cancer 104*, 682–691.
108. Newman, M. J. et al. (2000). Discovery and characterization of OC144-093, a novel inhibitor of P-glycoprotein-mediated multidrug resistance. *Cancer Research 60*, 2964–72.
109. Tabashnik, B. E. (1989). Managing resistance with multiple pesticide tactics: theory, evidence, and recommendations. *Journal of Economic Entomology 82*, 1263–1269.
110. Martello, L. A. et al. (2000). Taxol and discodermolide represent a synergistic drug combination in human carcinoma cell lines. *Clinical Cancer Research 6*, 1978–1987.
111. van Vlerken, L. E., Duan, Z., Seiden, M. V., Amiji, M. M. (2007). Modulation of intracellular ceramide using polymeric nanoparticles to overcome multidrug resistance in cancer. *Cancer Research 67*, 4843–4850.
112. Pakunlu, R. I. et al. (2006). *In vitro* and *in vivo* intracellular liposomal delivery of antisense oligonucleotides and anticancer drug. *Journal of Controlled Release 114*, 153–162.
113. Goodman, T. T., Olive, P. L., Pun, S. H. (2007). Increased nanoparticle penetration in collagenase-treated multicellular spheroids. *International Journal of Nanomedicine 2*, 265–274.

114. Netti, P. A., Baxter, L. T., Boucher, Y., Skalak, R., Jain, R. K. (1995). Time-dependent behavior of interstitial fluid pressure in solid tumors: implications for drug delivery. *Cancer Research 55*, 5451–5458.

115. Eikenes, L., Bruland, Ø. S., Brekken, C., Davies, C. D. L. (2004). Collagenase increases the transcapillary pressure gradient and improves the uptake and distribution of monoclonal antibodies in human osteosarcoma xenografts. *Cancer Research 64*, 4768–4773.

116. O'Neal, D. P., Hirsch, L. R., Halas, N. J., Payne, J. D., West, J. L. (2004). Photo-thermal tumor ablation in mice using near infrared-absorbing nanoparticles. *Cancer Letters 209*, 171–176.

117. Dromi, S. et al. (2007). Pulsed-high intensity focused ultrasound and low temperature-sensitive liposomes for enhanced targeted drug delivery and antitumor effect. *Clinical Cancer Research 13*, 2722–2727.

118. Kong, G., Braun, R. D., Dewhirst, M. W. (2000). Hyperthermia enables tumor-specific nanoparticle delivery: effect of particle size. *Cancer Research 60*, 4440–4445.

119. Zhang, J. Z. (2010). Biomedical applications of shape-controlled plasmonic nanostructures: a case study of hollow gold nanospheres for photothermal ablation therapy of cancer. *The Journal of Physical Chemistry Letters 1*, 686–695.

120. Hirsch, L. R. et al. (2003). Nanoshell-mediated near-infrared thermal therapy of tumors under magnetic resonance guidance. *Proceedings of the National Academy of Sciences of the United States of America 100*, 13549–13554.

121. Liu, Z. et al. (2005). Radiofrequency tumor ablation: insight into improved efficacy using computer modeling. *American Journal of Roentgenology 184*, 1347–1352.

122. Ahmed, M., Liu, Z., Humphries, S., Goldberg, S. N. (2008). Computer modeling of the combined effects of perfusion, electrical conductivity, and thermal conductivity on tissue heating patterns in radiofrequency tumor ablation. *International Journal of Hyperthermia 24*, 577–588.

123. Curley, S. A., Izzo, F., Ellis, L. M., Nicolas Vauthey, J., Vallone, P. (2000). Radiofrequency ablation of hepatocellular cancer in 110 patients with cirrhosis. *Annals of Surgery 232*, 381–391.

124. Hill, C. R., ter Haar, G. R. (1995). High intensity focused ultrasound--potential for cancer treatment. *The British Journal of Radiology 68*, 1296–1303.

125. Visaria, R. K. et al. (2006). Enhancement of tumor thermal therapy using gold nanoparticle-assisted tumor necrosis factor-alpha delivery. *Molecular Cancer Therapeutics 5*, 1014–1020.

126. Kampinga, H. H. (2006). Cell biological effects of hyperthermia alone or combined with radiation or drugs: a short introduction to newcomers in the field. *International Journal of Hyperthermia 22*, 191–196.

127. Sun, C., Lee, J. S. H., Zhang, M. (2008). Magnetic nanoparticles in MR imaging and drug delivery. *Advanced Drug Delivery Reviews 60*, 1252–1265.

128. McCarthy, J. R., Weissleder, R. (2008). Multifunctional magnetic nanoparticles for targeted imaging and therapy. *Advanced Drug Delivery Reviews 60*, 1241–1251.

129. Sadauskas, E. et al. (2009). Protracted elimination of gold nanoparticles from mouse liver. *Nanomedicine: Nanotechnology, Biology, and Medicine 5*, 162–169.

# INTRACELLULAR TRAFFICKING OF NANOPARTICLES: IMPLICATIONS FOR THERAPEUTIC EFFICACY OF THE ENCAPSULATED DRUG

Lin Niu

*Department of Pharmaceutics, College of Pharmacy, University of Minnesota, Minneapolis, MN,USA*

Jayanth Panyam

*Department of Pharmaceutics, College of Pharmacy, University of Minnesota, Minneapolis, MN, USA; Masonic Cancer Center, University of Minnesota, Minneapolis, MN, USA*

## 11.1 INTRODUCTION

Many drug molecules have intracellular sites of action and, therefore, require entry into the cell to exert their therapeutic effects. These include both macromolecules such as nucleic acids and small molecules such as taxanes. However, the cell membrane presents a formidable barrier to the entry of many drugs, either because the molecules are too large and/or polar to diffuse across the membrane (nucleic acids, proteins) or because they are actively pumped out of the cell by membrane-bound efflux transporters (paclitaxel, doxorubicin). Delivery systems that facilitate the intracellular accumulation of these drug molecules could enhance their therapeutic efficacy. Additionally, it has become obvious that for certain drugs, the localization into specific intracellular organelles is critical for their biological performance. For instance, to achieve effective gene expression, plasmid DNA

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*,
Edited by Shayne C. Gad
Copyright © 2013 John Wiley & Sons, Inc.

needs to be delivered not only into the cell but also into the nucleus. In the past decade, studies from several laboratories have shown that encapsulation of a drug in polymeric nanoparticles can improve its intracellular delivery. This chapter examines the intracellular trafficking of polymeric nanoparticles and the impact of modulating this trafficking on the therapeutic performance of the encapsulated drug.

## 11.2 PATHWAYS OF CELLULAR ENTRY

In order to understand the interaction of nanoparticles with cells, it is imperative to first introduce the lipid bilayer. The cell membrane defines the boundary between cellular contents and the external environment, separating intracellular biochemical activities from the extracellular surroundings. This seemingly excluding barrier is also the major interface where cells exchange information and materials with the surroundings. Phagocytosis and endocytosis are processes that either originate at or involve the plasma membrane to fulfill this mission. Similarly, at the subcellular level, it is again the compartmentalization by lipid bilayer membranes that enable various specialized biochemical processes proceeding simultaneously within the cells without interfering with one another. However, the necessity of exchanging materials and information still holds true at the organelle level. Intracellular membrane systems are involved in sorting and transporting materials within the cell in an orderly manner. Thus, plasma membrane and endomembrane system as a whole constitute dynamic continuums that relay material and information within the cell and with its surroundings. The interplay between nanoparticles and cells with regard to cellular uptake and intracellular trafficking is also intimately related with this dynamic cellular membrane system.

### 11.2.1 Endocytosis: Many Different Versions

Cells acquire large biomolecules and foreign substances that otherwise would be impermeable to the cell membrane through a process termed endocytosis. This term encompasses many different active (energy-dependent) membrane-based uptake processes. The major pathways involved in nanoparticles uptake are illustrated in Figure 11.1 [1]. Receptor-mediated endocytosis is a process by which cells internalize receptor-bound ligands. Clathrin-mediated endocytosis is so far the best-characterized pathway by which cells internalize specific molecules and modulate the expression level of membrane-bound receptors [2]. Morphologically, a salient feature of this endocytic mode is the formation of clathrin-coated pits and vesicles [3]. Recently, it was shown that this form of endocytic activity is initiated through the nucleation of FCHo proteins and various scaffold proteins such as eps15 and intersectin that promote the formation of cell membrane curvature, followed by recruitment of adaptor protein 2 (AP2) and clathrin triskelia [4]. Once the assembly of the clathrin lattice on the membrane is achieved, the scission of the invaginated clathrin-coated vesicle from the cell membrane is triggered. This process is facilitated by the mechanochemical enzyme dynamin, which prefers to localize

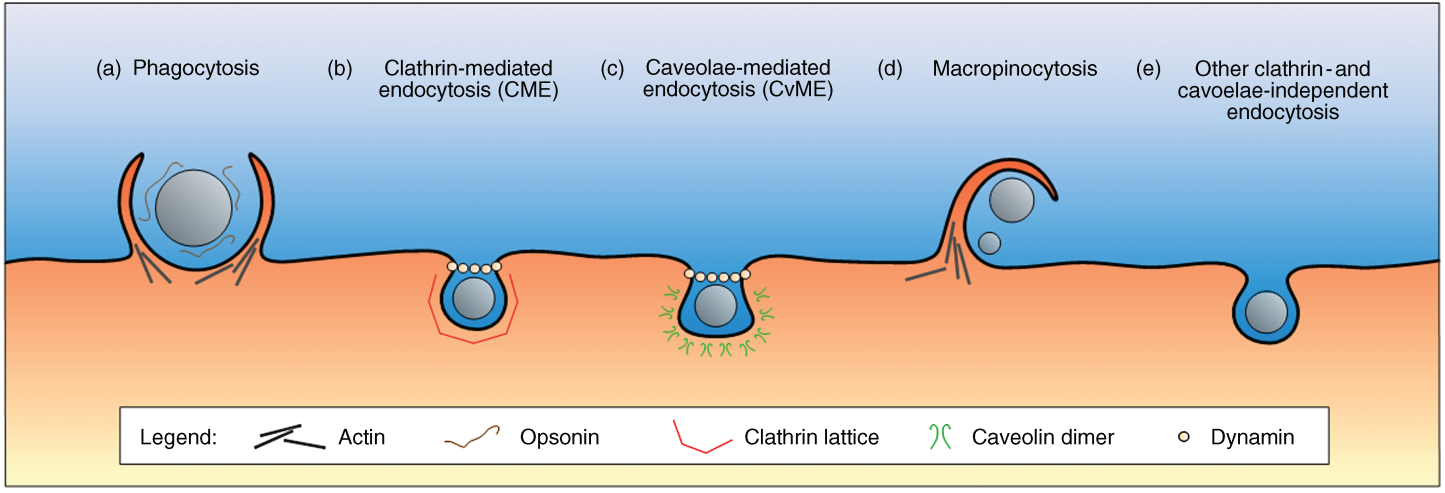


**FIGURE 11.1** Principal nanocarrier internalization pathways in mammalian cells. (*Source:* Reproduced from Reference [1] with permission from Springer Science and Business Media.)

to the constricted neck of the clathrin-coated vesicle because of the curvature, to complete the vesicle fission procedure in a GTP-dependent manner, pinching off the vesicle from plasma membrane [5]. The diameter of clathrin-coated pits in mouse epithelial cells ranges from 120 to 150 nm; while in mouse neuron, the diameter of clathrin-coated pits is from 70 to 90 nm [6]. The size of clathrin-coated pits can increase to accommodate the size of cargo [7]. The upper limit observed so far is ~200 nm in the case of clathrin-dependent cellular entry of vesicular stomatitis virus [8], although the definite upper limit in terms of transporting nanoparticles through this pathway is not yet elucidated. Clathrin-mediated endocytosis is involved in the internalization and recycling of transferrin receptor (TfR) and epidermal growth factor receptor (EGFR) [9].

Distinct from the classical clathrin-dependent pathway, several alternative clathrin-independent routes have been identified [10]. Caveolae-mediated endocytosis has recently gained a lot of attention. Despite this pathway's dependence on actin and dynamin for assembly and scission from plasma membrane (similar to the clathrin pathway), caveolae exhibits an invaginated flask-shaped morphology with caveolin-1 as the defining component [11]. In addition, caveolae are also rich in cholesterol and sphingolipid and are similar to detergent-resistant membrane microdomain lipid rafts in lipid composition. It has been suggested that caveolae share a common endocytic pathway with lipid rafts in that both processes are sensitive to cholesterol depletion [11]. The size of caveolae is relatively smaller than clathrin-coated vesicle. The reported diameter range is 50–80 nm under normal physiological conditions [12,13]. Caveolae and raft pathways facilitate the uptake of sphingolipid binding toxins (e.g., cholera toxin and Shiga toxin), albumin, glycosylphosphatidylinositol (GPI)-anchored proteins, growth hormone, and SV40 virus [14].

Macropinocytosis is a tightly regulated, actin-dependent fluid phase process that allows cells to capture materials in close proximity to the cell membrane [15]. This so-called cell drinking process relies on the ruffling of cell membrane, engulfing solute molecules and nutrients from extracellular fluid, followed by the formation of a large vacuole termed macropinosome within the cells [16]. Unlike receptor-mediated endocytosis, macropinocytosis is generally considered a nonspecific sampling process [17]. However, a variety of viruses including vaccinia, adeno, and picorna have been reported to take advantage of macropinocytosis to gain access to cytosol [18]. In addition, macropinocytosis has been proven to be involved in the cellular entry of arginine containing cationic peptides, such as the TAT peptide derived from human immunodeficiency virus (HIV) [19,20].

Phagocytosis is typically considered a procedure for internalization of large particles that are 0.5 μm in size [21]. Phagocytosis is often a feature of distinct cell types. Professional phagocytes, such as macrophages and dendritic cells, typically possess much higher phagocytic ability compared to other normal cell types [22]. As a result, phagocytosis is not considered a major pathway for nanoparticle internalization in non-phagocytic cells. Thus, this chapter does not focus on this topic for the discussion of cellular entry and subsequent intracellular fate of nanoparticles. However, in most cases, it is a significant uptake and clearance pathway when nanoparticles are administered *in vivo* due to the widely documented role of the

**TABLE 11.1 Agents Used to Inhibit Specific Endocytic Pathways**

| Dynamin Dependent | Clathrin | Caveolin | Macropinocytosis | Microtubule Dependent |
|---|---|---|---|---|
| Dynasore | Chlorpromazine<br>Amantadine<br>$K^+$ depletion | Filipin<br>Genistein | Cytochalasin<br>5-(*N*-ethyl-*N*-isopropyl) amiloride (EIPA) | Nocodazole |

reticuloendothelial system (RES) in nonselective elimination of nanoparticles [23]. Nanoparticle aggregates that are too large for endocytosis are likely cleared by professional phagocytes via phagocytosis. Moreover, phagocytosis could serve as a significant pathway for indirect nanoparticle acquisition. When successful intracellular nanoparticle delivery results in cell death, with an anticancer payload for example, the resulting apoptotic bodies and cellular debris will be cleared by professional phagocytes [24]. It is conceivable that a significant amount of nanoparticle components are still present within the apoptotic bodies. Thus, phagocytosis should not be overlooked for its contribution to the disposition of nanoparticles.

Often, several pathways contribute to the cellular uptake of nanoparticles. For instance, by using selective inhibitors to dissect endocytic pathways involved in the uptake of hydrophobically modified glycol chitosan nanoparticles by HeLa cells, caveolae-mediated pathway was found to account for 40% of uptake, clathrin-mediated pathway contributed to 20% of uptake, and another 30% was attributable to macropinocytosis [25]. A list of commonly used agents that selectively inhibit specific endocytic pathways is provided in Table 11.1.

### 11.2.2 Factors Affecting Nanoparticle Uptake into Cells

***11.2.2.1 Size and Shape.*** It is generally accepted that size is the primary determinant of efficiency for the cellular entry of nanoparticles. It was reported that 50 nm gold nanoparticles show maximum uptake efficiency in HeLa cells compared to smaller or larger particles in the tested size range from 14 to 100 nm [26]. This trend was maintained in the presence or absence of transferrin, even though transferrin adsorption did facilitate uptake across the entire size range. Dos Santos et al. studied the effect of wider range of particle sizes (40, 100, 200, 500, 1000, 2000 nm) on the uptake efficiency of carboxyl-modified polystyrene nanoparticles in several cell lines (HeLa, A549 epithelial cells, 1321N1 astrocytes, HCMEC D3 endothelial cells, and RAW 264.7 macrophages) [27]. Regardless of the particle size, macrophages were the most efficient at internalizing the particles. Internalization of nanoparticles was highly size-dependent in all the cell types studied, with uptake efficiency decreasing with increasing particle size.

The size dependence of nanoparticle uptake and the presence of an optimal size for the highest uptake efficiency have been attributed to the smallest wrapping time to generate a suitable membrane curvature around nanoparticle/nanoparticle aggregates [28]. This size is predicted to be about 30 nm, which is in agreement with several experimental observations with both viral and nonviral nanoparticles [29,30].

On the basis of the wrapping time theory, it is also proposed that uptake of nanoparticles with optimal size could occur independent of clathrin- and caveolin-mediated endocytosis. A similar experimental study and mechanistic modeling have been conducted for single-walled carbon nanotubes (SWNTs). The model asserts that a size of 25 nm confers the optimal radius for membrane capture. For nanomaterials with a dimension smaller than 25 nm, clusters could form on the plasma membrane surface to generate sufficient enthalpic contribution via nanoparticle–cell membrane interaction to overcome the membrane elastic energy and entropic barriers related to vesicle formation. It was found that SWNTs demonstrate 1000 times higher endocytosis rate constant ($10^{-3}$ $\text{min}^{-1}$) than spherical gold nanoparticles with the same radius ($10^{-6}\,\text{min}^{-1}$) [31]. Moreover, a length dependence of SWNT uptake has been reported with a threshold of ~189 nm, indicating nanotubes longer than this threshold could not be taken up by cells [32].

Conflicting data have been reported about the impact of shape on cellular uptake. The uptake of rod-shaped gold nanoparticles was reported as being less efficient compared with their spherical counterparts. Cells internalized 5 and 3.75 times more 74 nm and 14 nm spherical gold nanoparticles than $74 \times 14$ nm rod-shaped gold nanoparticles, respectively [26]. In addition, higher aspect ratio (1:5) nanoparticles were taken up poorly in comparison with lower aspect ratio (1:3) nanoparticles. In contrast, a study using cationic, cross-linked poly(ethylene glycol) hydrogel nanoparticles manufactured through PRINT technique has generated opposite results. HeLa cells internalized the high-aspect-ratio cylindrical particles (diameter, $d = 150$ nm, height, $h = 450$ nm) ~4 times greater than the low-aspect-ratio symmetric cylindrical particles ($d = 200$ nm, $h = 200$ nm) [33]. Similarly, greater internalization of larger aspect ratio was observed for mesoporous silica nanoparticles in A375 melanoma cell lines [34]. On the other hand, Meng et al. showed that the internalization of rod-shaped silica nanoparticles in HeLa and A549 cell lines was much greater than that of spherical particles, but rod-shaped silica nanoparticles with aspect ratio of 2.1–2.5 demonstrated the highest uptake efficiency than either longer or shorter length rods. This was attributed to stronger activation by nanoparticles with this particular aspect ratio of GTP-binding protein Rac1, which is activated during macropinocytosis, and the Rho subfamily member protein CDC42 involved in regulating actin polymerization-dependent macropinocytosis processes [35].

***11.2.2.2 Surface Properties.*** The surface characteristic of nanoparticles not only affects cellular uptake but also the intracellular fate following the cellular entry. Characteristics affecting nanoparticles uptake include surface charge, chemical composition, presence of targeting ligand, and the number of ligand molecules.

It is generally thought that neutral charge on the surface minimizes the interaction of nanoparticles with cells, partly because of weaker protein adsorption on the surface of nanoparticles in biological milieu [36,37]. PEGylation (incorporation of poly(ethylene glycol), PEG, chains on the surface) is the most well-established surface modification technique to provide "stealth" capability and improve the pharmacokinetic profile. The neutral and hydrophilic PEG coating

lowers nonspecific protein adsorption, opsonization, and nanoparticle aggregation so that receptor-mediated endocytosis by cells of RES is dampened [37]. On the other hand, charged nanoparticles typically exhibit stronger ability to interact with cells. This is especially true for cationic nanoparticles that are highly charged, although higher cytotoxicity could also be expected [38]. Using an *in vitro* polarized epithelial Madin–Darby canine kidney (MDCK) cell model, it was found that both cationic and anionic nanoparticles are taken up by the epithelial cells through clathrin-dependent pathway and macropinocytosis, followed by significant transcytosis and accumulation at the basolateral membrane. Only a fraction of anionic but not cationic nanoparticles transited through the degradative lysosomal pathway [39]. In the context of *in vivo* distribution, since nanoparticles with high surface charge and larger size are cleared by macrophages more efficiently, 150 nm nanoparticles with slight negative surface charge has been shown to be the most effective in achieving prolonged circulation, a prerequisite for enhanced tumor accumulation [40].

A novel mechanism for how positive surface charge affects nanoparticle–cell interaction was recently proposed. Binding of positively charged nanoparticles to the negatively charged plasma membrane could modulate the membrane potential, induce membrane depolarization, and influx of $Ca^{2+}$ [41]. To take advantage of the fact that positively charged surfaces lead to improved cellular entry, Poon et al. constructed a layer-by-layer nanoparticle system, consisting of quantum dot core, coated with iminobiotin-modified poly-L-lysine inner layer and biotin-functionalized PEG outer layer, bridged by neutravidin. Because of the pH-sensitive nature of iminobiotin–neutravidin interaction, it was expected that the outer PEG layer that conferred long circulating property would be shed in the acidic tumor microenvironment. The subsequent exposure of positively charged poly-L-lysine coating allowed for improved cellular uptake of nanoparticles. The benefit of this strategy was shown *in vivo* in MDA–MB-435 breast cancer model [42].

Another study suggested that it might be possible to modulate the mechanism of cellular entry by tuning the surface properties of nanoparticles. It was reported that 6 nm nanoparticles, decorated with a shell of alternating hydrophobic and anionic hydrophilic ligands in ribbon-like domains, display membrane penetration without bilayer disruption even at 4°C and altered intracellular localization compared to the "isomer" nanoparticle counterpart with identical surface functionalization but completely disordered surface pattern arrangement [43].

### 11.2.3 Targeting Ligand-Mediated Cellular Entry of Nanoparticles

In order to maximize efficacy and minimize off-target side effects, targeting proteins and receptors that are abundantly expressed on malignant cells have been routinely investigated. The mechanism of entry and intracellular trafficking of nanoparticles with the surface-bound ligand could, to some degree, mirror that of the specific target receptors [44]. Although overexpression of several different receptors has been reported, this chapter discusses transferrin and folic acid receptors as prototypical clathrin- and caveolae-based receptors, respectively.

***11.2.3.1 Transferrin Receptor (TfR).*** Uptake of transferrin by TfR via clathrin-coated pit is one of the most well-characterized endocytic processes [45,46]. TfR is overexpressed in a wide variety of cancers [47]. Naturally, numerous groups have chosen TfR-mediated endocytosis as both a targeting and uptake enhancing strategy. Even though the presence of TfR targeting moiety on drug delivery carrier may not necessarily improve the overall accumulation in solid tumor, TfR targeting strategy has shown to significantly improve the amount of nanoparticles localized within cancer cells [48]. The study that provided the first evidence of successful RNA interference in human utilized TfR-targeted nanoparticle as a delivery platform for systemic administration of siRNA [49].

By using a 3-glysidoxypropyltrimethoxysilane (3-GPTMS) linker with epoxy terminal, the surface of mesoporous silica nanoparticles was decorated with Tf in another study. This allowed for enhanced accumulation of silica nanoparticles in TfR transfected HFF cell line compared to the wild type HFF cell line, indicating TfR-mediated endocytic uptake process in TfR overexpressed cells [50]. Cylindrical PEG-based PRINT nanoparticles conjugated with either human Tf or anti-TfR monoclonal antibody was internalized more readily in a panel of six common cancer cell lines relative to nanoparticles labeled with either bovine Tf or isotype control antibody, suggesting the potential for exploiting TfR-mediated endocytosis in improving intracellular delivery in TfR positive tumors [51]. TfR-targeted lipid nanoparticles encapsulated with antisense oligonucleotides against the antiapoptotic factor Bcl-2 were found to be internalized more effectively and possessed much better gene knockdown efficiency in leukemia cell lines than nontargeted lipid nanoparticle and free oligonucleotides [52]. Highest Bcl-2 downregulation and apoptosis-inducing efficacy was found in K562 cell line, which expressed the highest TfR level. The Bcl-2 downregulation and apoptosis induction were blocked by excessive free Tf. Combination of TfR targeting with pH responsiveness is particularly well suited to capitalize on the acidification of endosome subsequent to TfR-mediated endocytosis. Methacrylic acid copolymer complexed with poly(amido amine) (PAMAM) dendrimers and siRNA formed a core-shell polyelectrolyte micelle, which was decorated with anti-TfR Fab (antibody fragment of antigen binding) [53]. Targeted micelles showed stronger uptake by PC-3 prostate cancer cells than nontargeted control, and excess free anti-TfR antibody abolished this difference. Once the anionic portion of methacrylic acid copolymer was neutralized during acidification, polycationic dendrimer core containing siRNA was exposed, followed by endosomal escape and mRNA transcript silencing.

***11.2.3.2 Folate Receptor.*** Unlike other transmembrane receptor proteins, folate receptor is a glycosylphosphatidylinositol (GPI)-anchored membrane protein, potentially associated with lipid raft [54]. Internalization of folate receptor follows a clathrin-independent mechanism. Folate receptor is sorted into recycling endosomal compartment, where GPI-anchored proteins are enriched and then trafficked to perinuclear sites before eventually returning folate receptor to the cell membrane. The synthesis of a disulfide-linked folate conjugate as a reduction-responsive fluorescence resonance energy transfer (FRET) probe enabled the elucidation of

folate receptor life cycle, potential location of reduction reaction, and distinct sorting destination of cleaved yield products [55]. The disulfide disruption of folate-(BODIPY)-SS-rhodamine conjugate (folate-FRET) proceeded with a half-life of 6 h at recycling endosomal vesicle with reductive environment, which was independent of redox machinery in lysosome and Golgi apparatus. Furthermore, folate receptors (indicated by folate BODIPY fluorescent dye) were recycled back to plasma membrane, while released rhodamine was sorted into a distinct vesicle free of membrane-bound folate receptors. Since folate receptor is recycled to the plasma membrane, the timely release of the carrier-bound drug during the intracellular phase is crucial to attaining optimal intracellular drug delivery for a folate receptor-targeted system. Accordingly, folate receptor targeting drug delivery platforms with built-in reduction responsive drug unloading trigger have been formulated. In one of such systems, disulfide bond between hexaethylene glycol outer layer and cucurbit[6]uril (CB[6]) inner core was disrupted in the reducing environment of endosomes following folate receptor-mediated endocytic uptake, releasing doxorubicin entrapped within this self-assembled amphiphilic platform [56].

***11.2.3.3 Other Receptors.*** It was shown that a delivery system based on beta-cyclodextrin-containing polycations possessed strong transfection efficiency against hepatoma cell line because of asialoglycoprotein receptor-mediated uptake [57]. The protein adsorption on the surface of nanoparticles also plays a major role in nanoparticle uptake. At the initial stages of nanoparticle development, unmodified nanoparticles were prone to opsonization by RES due to the protein adsorption and subsequent uptake by Kupffer cells in the liver. On the other hand, it was reported that polysorbate 80-coated poly(butyl cyanoacrylate) nanoparticles show promising ability to cross blood–brain barrier. Follow up studies elucidated the critical role of apolipoprotein adsorption that triggers receptor-mediated endocytosis by endothelial cells lining the blood–brain barrier [58].

## 11.3 INTRACELLULAR FATE OF INTERNALIZED NANOPARTICLES

Even though it is impossible to generalize a unifying theme of intracellular fate for distinct nanoparticle systems in distinct cell types, a set of potential itinerary repetitively observed for nanoparticles could be enumerated here. Upon cellular entry, a nanoparticle starts the intracellular journey either at cytosol directly via direct membrane fusogenic entry or from within membrane-confined vesicles such as early endosomes following endocytosis and macropinocytosis. As the early endosome undergoes maturation, a fraction of cargo within the compartment could be transported and sorted either to recycling pathway or toward multivesicular body and lysosome, the most widely documented intracellular sequestration destination for nanoparticles. Meanwhile, a typical trajectory of cargo transport mediated by motor molecules along the microtubule highway towards perinuclear region could also be observed. Throughout the aforementioned endosome maturation processes, nanoparticles with specific surface properties could escape endolysosomes and enter

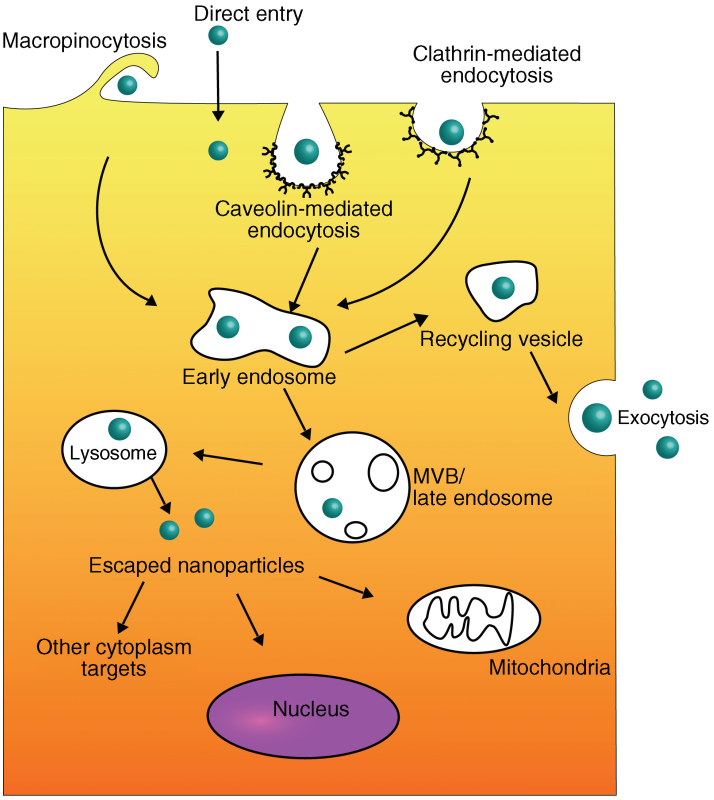


**FIGURE 11.2** Potential intracellular trafficking pathways for nanoparticles. MVB: multivesicular body.

cytosol at this stage, leading to either intracellular localization with or without further distribution to other subcellular organelles. Some common pathways for nanoparticle intracellular trafficking are depicted in Figure 11.2.

The intracellular fate is intimately linked with the route of entry, which is dependent on the surface properties of nanoparticles. Targeting moiety often has major influence on both entry and intracellular trafficking since endocytosis route could dictate the subsequent intracellular sorting processes. Oba et al. have demonstrated that cyclic arginine-glycine-aspartic acid (RGD) targeting peptide not only selectively boosts PEG–poly-lysine–DNA polyplex cellular entry into cells expressing $\alpha_v\beta_3$ and $\alpha_v\beta_5$ integrin receptors but also ensures preferential perinuclear localization so that transfection efficiency against HeLa cells is significantly augmented. It was shown that this was not due to enhanced uptake but due to favorable intracellular trafficking following caveolae-mediated endocytosis, which directs more plasmid DNA cargo through caveosome pathway rather than degradative lysosome pathway as suggested by Figure 11.3 [59,60].

An elegant study on the uptake and intracellular fate of gold nanoparticles in HeLa cells demonstrated the correlation between surface chemistry and cellular fate

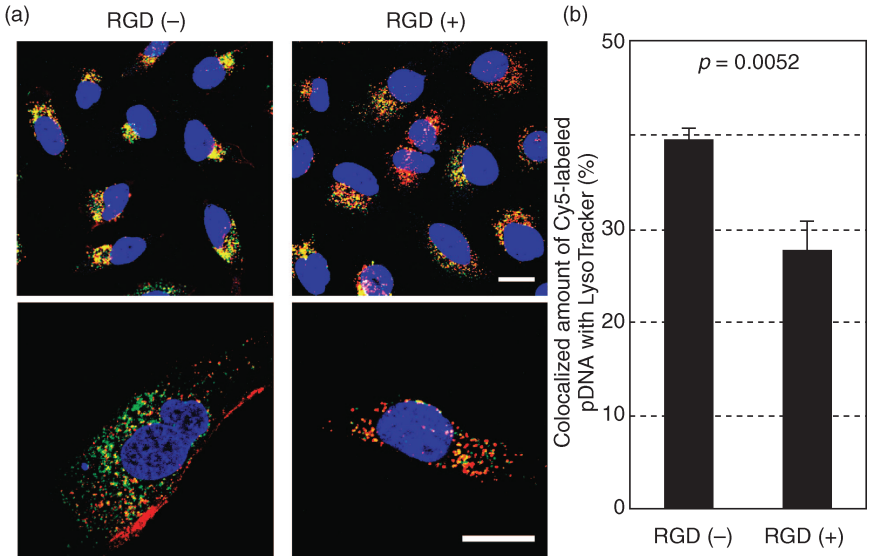


**FIGURE 11.3** Distribution of RGD (+) and RGD (−) polyplex micelles in late endosomes and lysosomes. Polyplex micelles loading Cy5-labeled pDNA (red) were incubated with HeLa cells for 1 h. After replacement with fresh medium, the cells were reincubated for 11 h. The cell nuclei were stained with Hoechst 33342 (blue), and the acidic late endosomes and lysosomes were stained with LysoTracker Green (green). (a) CLSM images of the cells transfected with RGD (−) micelles (left) and RGD (+) micelles (right). The scale bars represent 20 μm. (b) Quantification of Cy5-labeled pDNA colocalized with LysoTracker Green in the inner cytoplasm. Error bars in the graph represent SEM ($n = 10$). (*Source:* Reprinted with permission from Reference [60]. Copyright (2008) American Chemical Society.)

of gold nanoparticles. Bare gold nanoparticles were taken up by endocytosis and were confined exclusively in endosomes, while PEGylation prevented cellular uptake completely. However, the inclusion of these PEGylated gold nanoparticles within liposomes or functionalization with cell penetrating peptide resulted in cytosolic distribution, circumventing the endosomal fate of unmodified counterpart as demonstrated in Figure 11.4 [61]. The cytosolic disposition was attributed to direct entry of nanoparticles across the cell membrane and/or disruptive surface property that damaged endosomal membranes [61].

### 11.3.1 Organelle Targeting

Specific organelle targeting is an important determinant of nucleic acid delivery efficiency of nonviral systems. RNA interference occurs in the cytosol while DNA transfection can only be achieved following transport of plasmid DNA into the nucleus. For successful siRNA delivery into cytosol, nanocarriers that enter the cell through endocytosis must exhibit endosomal escape [62,63]. For successful gene

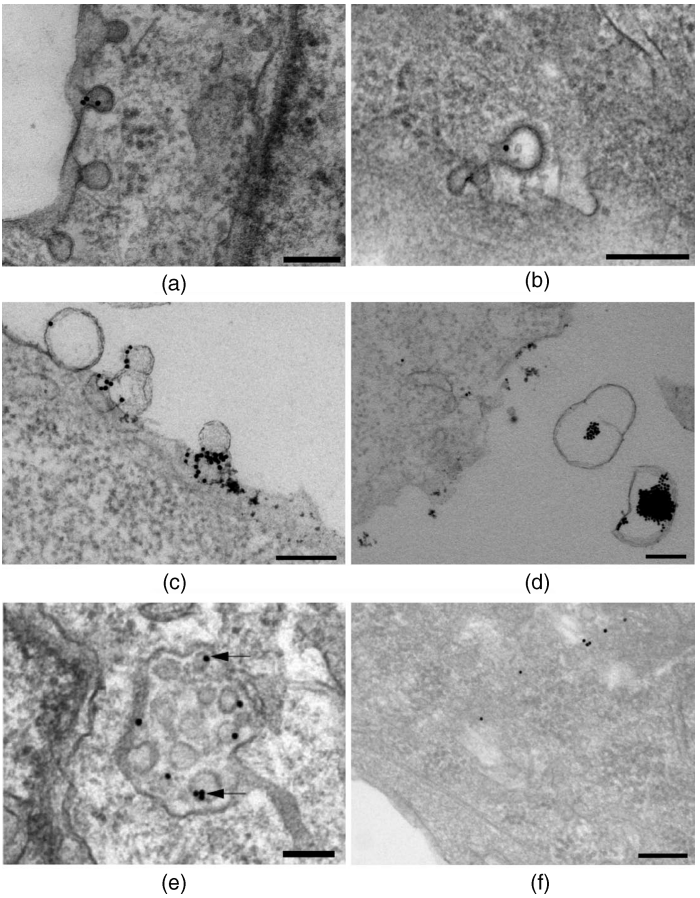


**FIGURE 11.4** Uptake of PEG-modified gold nanoparticles in the presence of liposomes. Observed uptake mechanisms without apparent direct involvement of the liposomes include (a) caveolae and (b) clathrin-mediated endocytosis. Particle delivery by liposomes and subsequent uptake are shown in (c) and (d). Perhaps reflecting the wealth of different uptake mechanisms available under these conditions, nanoparticles (free) are found in the endosome (e) or free in the cytosol (f). Scale bars are 200 nm. (*Source:* Reprinted with permission from Reference [61]. Copyright (2008) American Chemical Society.)

delivery, nanocarriers must achieve nuclear entry after endosomal escape so that the genetic materials are delivered directly into the nucleus. To this end, nuclear localization signal (NLS) peptide has been utilized to enable nuclear targeting. Two different targeting ligands may be conjugated to the same nanoparticle surface to achieve sequential cell and organelle targeting. For example, conjugation with RGD and NLS peptides imparted both tumor targeting and nucleus targeting property to 30 nm gold nanoparticles [64].

Mitochondria serve as the powerhouse of the cell, producing ATP through oxidative phosphorylation. In addition, mitochondria contain a distinct set of DNA inherited from maternal source. Abnormal functionality in mitochondria has been linked with a diverse range of neurodegenerative disorders including Alzheimer's. Thus, a mitochondria-targeted delivery system is desirable for mitochondrial gene therapy, metabolic diseases, and proapoptotic strategy for cancer treatment. Lipid nanoparticles loaded with proapoptotic drug SV30, in spite of lacking a specific mitochondria targeting functionality, exhibited a preferential accumulation at sites adjacent to mitochondria, as suggested by colocalization of MitoTracker (a mitochondrion labeling fluorescent dye) and Nile red-labeled lipid nanoparticles [65]. On the other hand, liposomes functionalized with mitochondriotropic triphenylphosphonium cations promoted preferential localization of nanocarrier at mitochondria and improved the therapeutic efficacy of ceramide both *in vitro* and *in vivo* [66]. A multifunctional iron oxide "nanoworm" functionalized with both glioblastoma targeting peptide (CGKRK) and mitochondria targeting peptide ($_{D}$[KLAKLAK]$_2$) has shown promising efficacy in treating glioblastoma bearing mice by inducing proapoptotic processes. Neither tumor targeting peptide nor mitochondria targeting peptide alone could confer this beneficial activity, highlighting that the sequential, cooperative action of both peptides was indispensable to achieving intracellular delivery followed by modulation of intracellular trafficking and localization [67].

### 11.3.2 Exocytosis of Nanoparticles

The story of nanoparticle–cell interaction would be incomplete without a discussion of the elimination mechanisms of nanoparticles, which has important bearing on nanoparticles' retention, drug release profile, and toxicity. In addition to the possible disassembly and degradation of nanoparticles in the lysosomes, exocytosis is another clearance mechanism for internalized nanoparticles. Cargo that has been sorted to secretory/recycling vesicles is destined to the fate of exocytosis. The relatively scarce data on exocytosis of nanoparticles prevent generalization of rules governing nanoparticle excretion. Nevertheless, according to the studies on nanoparticle internalization and sorting, it is conceivable that for a specific nanoparticle of interest (with unique size, shape, and surface property) whether or not exocytosis occurs at all, to what extent, and how fast likely depends on the cell line under study and could be highly context dependent.

Even though the data with regard to exocytosis are not as abundant as internalization, exocytosis of poly(D,L-lactide-*co*-glycolide) nanoparticles has been studied *in vitro* using human arterial vascular smooth muscle cells (VSMCs). These studies showed that intracellular concentration of nanoparticles was maintained as long as the cells were in contact with excess nanoparticles. Once the extracellular nanoparticle concentration gradient was removed, exocytosis of nanoparticles occurred with about 65% of the internalized fraction undergoing exocytosis in 30 min [68]. Energy depletion reduced exocytosis, indicating the ATP-dependent nature of this active sorting process. In addition, protein adsorption on nanoparticle surface has a major influence on sorting of nanoparticles into

exocytic pathway as exocytosis was abrogated in serum-free culture medium. Slowing et al. have elegantly shown that when two separately cultured flasks of human umbilical vein endothelial cell (HUVECs) incubated with either FITC-labeled mesoporous silica nanoparticles or TRITC-labeled nanoparticles were pooled, individual cells contained both FITC- and TRITC-labeled nanoparticles, providing indirect evidence that internalized mesoporous silicon nanoparticles could undergo exocytosis by HUVEC, leading to reuptake and exchanging of expelled nanoparticles [69]. However, no such effect was observed in the HeLa cell line. In another experiment, about 50% of internalized D-penicillamine-coated zwitterionic quantum dots (DPA–QDs) of 4 nm radius was expelled by HeLa cells. The exocytosed fraction increased with a half-life of 21 min and reached saturation after 2 h [70]. Similar to quantum dots, nanodiamonds made of nanocarbon allotrope with negatively charged nitrogen defects as fluorescent centers have good photostability and therefore are candidates for long-term labeling of stem cells. Retention rate is an important parameter to consider before using fluorescent nanodiamonds to track a specific cell of interest. There was little (~15% or less) excretion of the endocytosed nanodiamond after 6 days of labeling from both HeLa and 489-2.1 cells, but exocytosis occurred more readily (up to 30%) for 3T3-L1 preadipocytes [71]. On the contrary, addition of silver nanoparticles at a concentration of 50 μg/ml to RBL-2H3 mast cell line triggered degranulation instantaneously [72]. Decisive *in vivo* and mechanistic experiments investigating exocytosis of nanoparticles are lacking. It remains to be seen whether exocytosis patterns seen *in vitro* would be observed *in vivo* and whether exocytosis has any major implications for nanoparticle pharmacokinetics and pharmacodynamics.

## 11.4 IN-DEPTH ELUCIDATION OF INTRACELLULAR TRAFFICKING EVENTS ENTAILS HIGH RESOLUTION IMAGING TECHNIQUES

The advent and application of more advanced imaging techniques will enable in-depth understanding of the intracellular trafficking events. Spinning disk confocal microscopy has increasingly become the method of choice for live cell imaging due to its quick data acquisition compared to more widely available conventional confocal instrumentation. In conjunction with short-term overexpression of fluorescently tagged Rab family proteins by transfection, the intracellular trafficking of 40 nm fluorescent polystyrene nanoparticles in HeLa cells has been observed with high spatial and temporal resolution [73]. The fraction of nanoparticles co-localized with Rab5, a marker for early endosome, decreased over time, while the fraction of nanoparticles co-localized with Rab7/9, markers for late endosome, increased correspondingly until reaching a plateau of 80%, which was consistent with the data achieved by using LAMP1 as the lysosome marker. Only miniscule co-localization of nanoparticles with Rab11 was observed, suggesting that nanoparticle trafficking into recycling vesicles was not a major sorting route in this experimental setup.

The resolution of conventional optical microscopy, including fluorescent confocal microscopy, is diffraction limited. For most studies, the resolving power of the

conventional techniques cannot surpass the limit of 200–300 nm, depending on the instrumentation and fluorescent emission wavelength under study. However, the diameter of most nanoparticles employed in current research is below this diffraction-limited resolution. Conceivably, as researchers gain access to more advanced imaging techniques, such as super-resolution microscopy capable of generating subdiffraction limit resolution, our understanding of nanoparticle intracellular trafficking and fate can progress dramatically. It was reported that two 40 nm nanoparticles separated by an 85 nm center-to-center distance could be resolved by stimulated emission depletion (STED) microscopy, a scenario unresolvable by diffraction-limited conventional confocal microscopy [74]. The uptake and aggregation of silica nanoparticle with 128 nm diameter in lung carcinoma cell line A549 have also been visualized by STED microscopy [75]. Another report shows that 32 nm silica nanoparticle could gain access to nucleus and form aggregates within nucleus progressively. In contrast, 80 nm silica nanoparticles not only migrate at a slower rate intracellularly but also fail to penetrate nuclear membrane [76]. This report also demonstrated the unique strength of super-resolution technique in imaging intracellular fate of nanoparticles.

## 11.5 FUTURE PERSPECTIVES

Nanoparticulate systems hold great promise in enhancing the therapeutic efficacy of drugs with intracellular sites of action [77] and improving the efficacy of nucleic acid-based therapies encompassing siRNA and gene delivery [78]. The last decade has witnessed tremendous advancement of our understanding of cellular entry and intracellular trafficking of nanoparticles. This progress could guide the rational design of next generation of nanoparticles with improved therapeutic efficacy. However, one limitation in the field has been the understandable lack of consistency across research groups in the cell and animal models used. Nanoparticle–cell interaction is highly dependent on the cell line and treatment conditions used. For example, it has been shown that even closely related cell lines such as PC3, PC3-flu, and PC3-PSMA (the latter two were derived from the former) demonstrate dramatically differential interaction profiles with quantum dots. In PC3 cells, punctate cytoplasmic localization was observed, while in PC-PSMA cell line, co-localization of nanoparticles with transferrin was observed at the perinuclear recycling compartment [79]. The preferential sequestration of DNA polyplexes in perinuclear microtubule organizing compartment (MTOC) of PC3-PSMA cells (similar to that seen with quantum dots) was associated with the poor transfection efficiency. In contrast, the cytoplasmic localization of DNA polyplexes in PC3 cells resulted in relatively higher transfection efficiency [80]. Thus, caution should be exercised when reaching broad conclusions from a specific experimental setup and extrapolating observations from an experiment done with one cancer cell type to another cancer cell type or worse, to noncancerous cells and professional phagocytes.

From a therapeutic perspective, heterogeneity in tumor cell–nanoparticle interaction within a given population of tumor cells could pose significant problems. Differences in uptake of cytotoxin-armed nanoparticles within a given population

could result in differential cytotoxicity in the population and impose a selection pressure that could result in the enrichment of drug-resistant clones. Thus, in the context of clinical applications, future nanoparticle design may have to consider nanoparticles' interaction with the entire tumor cell population as well as a whole gamut of other tumor-associated cell types (tumor-associated macrophages, endothelial cells, mesenchymal stem cells, to name a few) *in vivo*. This would require a "systems biology" approach to better understand the complex interactions that nanocarriers likely have with biological systems.

## REFERENCES


1. Hillaireau, H., Couvreur, P. (2009). Nanocarriers' entry into the cell: relevance to drug delivery. *Cellular and Molecular Life Sciences 66*, 2873–2896.
2. Le Roy, C., Wrana, J. L. (2005). Clathrin- and non-clathrin-mediated endocytic regulation of cell signalling. *Nature Reviews Molecular Cell Biology 6*, 112–126.
3. Kirchhausen, T. (2000). CLATHRIN. *Annual Review of Biochemistry 69*, 699–727.
4. Henne, W. M., Boucrot, E., Meinecke, M., Evergren, E., Vallis, Y., Mittal, R., Mcmahon, H. T. (2010). FCHo proteins are nucleators of clathrin-mediated endocytosis. *Science 328*, 1281–1284.
5. Sweitzer, S. M., Hinshaw, J. E. (1998). Dynamin undergoes a GTP-dependent conformational change causing vesiculation. *Cell 93*, 1021–1029.
6. Mcmahon, H. T., Boucrot, E. (2011). Molecular mechanism and physiological functions of clathrin-mediated endocytosis. *Nature Reviews Molecular Cell Biology 12*, 517–533.
7. Ehrlich, M., Boll, W., Van Oijen, A., Hariharan, R., Chandran, K., Nibert, M. L., Kirchhausen, T. (2004). Endocytosis by random initiation and stabilization of clathrin-coated pits. *Cell 118*, 591–605.
8. Cureton, D. K., Massol, R. H., Saffarian, S., Kirchhausen, T. L., Whelan, S. P. J. (2009). Vesicular stomatitis virus enters cells through vesicles incompletely coated with clathrin that depend upon actin for internalization. *PLoS Pathogenesis 5*, e1000394.
9. Traub, L. M. (2009). Tickets to ride: selecting cargo for clathrin-regulated internalization. *Nature Reviews Molecular Cell Biology 10*, 583–596.
10. Mayor, S. Pagano, R. E. (2007). Pathways of clathrin-independent endocytosis. *Nature Reviews Molecular Cell Biology 8*, 603–612.
11. Nabi, I. R., Le, P. U. (2003). Caveolae/raft-dependent endocytosis. *The Journal of Cell Biology 161*, 673–677.
12. Conner, S. D., Schmid, S. L. (2003). Regulated portals of entry into the cell. *Nature 422*, 37–44.
13. Bareford, L. M., Swaan, P. W. (2007). Endocytic mechanisms for targeted drug delivery. *Advanced Drug Delivery Reviews 59*, 748–758.
14. Pelkmans, L., Helenius, A. (2002). Endocytosis via caveolae. *Traffic 3*, 311–320.
15. Swanson, J. A., Watts, C. (1995). Macropinocytosis. *Trends in Cell Biology 5*, 424–428.
16. Falcone, S., Cocucci, E., Podini, P., Kirchhausen, T., Clementi, E. Meldolesi, J. (2006). Macropinocytosis: regulated coordination of endocytic and exocytic membrane traffic events. *Journal of Cell Science 119*, 4758–4769.

17. Lim, J. P., Gleeson, P. A. (2011). Macropinocytosis: an endocytic pathway for internalising large gulps. *Immunology and Cell Biology 89*, 836–843.
18. Mercer, J., Helenius, A. (2009). Virus entry by macropinocytosis. *Nature Cell Biology 11*, 510–520.
19. Kaplan, I. M., Wadia, J. S., Dowdy, S. F. (2005). Cationic TAT peptide transduction domain enters cells by macropinocytosis. *Journal of Controlled Release 102*, 247–253.
20. Nakase, I., Niwa, M., Takeuchi, T., Sonomura, K., Kawabata, N., Koike, Y., Takehashi, M., Tanaka, S., Ueda, K., Simpson, J. C., Jones, A. T., Sugiura, Y., Futaki, S. (2004). Cellular uptake of arginine-rich peptides: roles for macropinocytosis and actin rearrangement. *Molecular Therapy 10*, 1011–1022.
21. Groves, E., Dart, A. E., Covarelli, V., Caron, E. (2008). Molecular mechanisms of phagocytic uptake in mammalian cells. *Cellular and Molecular Life Sciences 65*, 1957–1976.
22. Jutras, I., Desjardins, M. (2005). Phagocytosis: at the crossroads of innate and adaptive immunity. *Annual Review of Cell and Developmental Biology 21*, 511–527.
23. Brannon-Peppas, L., Blanchette, J. O. (2004). Nanoparticle and targeted systems for cancer therapy. *Advanced Drug Delivery Reviews 56*, 1649–1659.
24. Erwig, L. P., Henson, P. M. (2007). Clearance of apoptotic cells by phagocytes. *Cell Death Differentiation 15*, 243–250.
25. Nam, H. Y., Kwon, S. M., Chung, H., Lee, S.-Y., Kwon, S.-H., Jeon, H., Kim, Y., Park, J. H., Kim, J., Her, S., Oh, Y.-K., Kwon, I. C., Kim, K., Jeong, S. Y. (2009). Cellular uptake mechanism and intracellular fate of hydrophobically modified glycol chitosan nanoparticles. *Journal of Controlled Release 135*, 259–267.
26. Chithrani, B. D., Ghazani, A. A., Chan, W. C. W. (2006). Determining the size and shape dependence of gold nanoparticle uptake into mammalian cells. *Nano Letters 6*, 662–668.
27. dos Santos, T., Varela, J., Lynch, I., Salvati, A., Dawson, K. A. (2011). Quantitative assessment of the comparative nanoparticle-uptake efficiency of a range of cell lines. *Small 7*, 3341–3349.
28. Gao, H., Shi, W., Freund, L. B. (2005). Mechanics of receptor-mediated endocytosis. *Proceedings of the National Academy of Sciences of the United States of America 102*, 9469–9474.
29. Nakai, T., Kanamori, T., Sando, S., Aoyama, Y. (2003). Remarkably size-regulated cell invasion by artificial viruses. Saccharide-dependent self-aggregation of glycoviruses and its consequences in glycoviral gene delivery. *Journal of the American Chemical Society 125*, 8465–8475.
30. Jiang, W., Kimbetty, Y. S., Rutka, J. T., Chanwarren, C. W. (2008). Nanoparticle-mediated cellular response is size-dependent. *Nature Nanotechnology 3*, 145–150.
31. Jin, H., Heller, D. A., Sharma, R., Strano, M. S. (2009). Size-dependent cellular uptake and expulsion of single-walled carbon nanotubes: single particle tracking and a generic uptake model for nanoparticles. *ACS Nano 3*, 149–158.
32. Becker, M. L., Fagan, J. A., Gallant, N. D., Bauer, B. J., Bajpai, V., Hobbie, E. K., Lacerda, S. H., Migler, K. B., Jakupciak, J. P. (2007). Length-dependent uptake of DNA-wrapped single-walled carbon nanotubes. *Advanced Materials 19*, 939–945.
33. Gratton, S. E. A., Ropp, P. A., Pohlhaus, P. D., Luft, J. C., Madden, V. J., Napier, M. E., Desimone, J. M. (2008). The effect of particle design on cellular internalization pathways. *Proceedings of the National Academy of Sciences of the United States of America 105*, 11613–11618.

34. Huang, X., Teng, X., Chen, D., Tang, F., He, J. (2010). The effect of the shape of mesoporous silica nanoparticles on cellular uptake and cell function. *Biomaterials 31*, 438–448.

35. Meng, H., Yang, S., Li, Z., Xia, T., Chen, J., Ji, Z., Zhang, H., Wang, X., Lin, S., Huang, C., Zhou, Z. H., Zink, J. I., Nel, A. E. (2011). Aspect ratio determines the quantity of mesoporous silica nanoparticle uptake by a small GTPase-dependent macropinocytosis mechanism. *ACS Nano 5*, 4434–4447.

36. Khan, J. A., Pillai, B., Das, T. K., Singh, Y., Maiti, S. (2007). Molecular effects of uptake of gold nanoparticles in HeLa cells. *ChemBioChem 8*, 1237–1240.

37. Verma, A., Stellacci, F. (2010). Effect of surface properties on nanoparticle–cell interactions. *Small 6*, 12–21.

38. Schaeublin, N. M., Braydich-Stolle, L. K., Schrand, A. M., Miller, J. M., Hutchison, J., Schlager, J. J., Hussain, S. M. (2011). Surface charge of gold nanoparticles mediates mechanism of toxicity. *Nanoscale 3*, 410–420.

39. Harush-Frenkel, O., Rozentur, E., Benita, S., Altschuler, Y. (2008). Surface charge of nanoparticles determines their endocytic and transcytotic pathway in polarized MDCK cells. *Biomacromolecules 9*, 435–443.

40. He, C., Hu, Y., Yin, L., Tang, C., Yin, C. (2010). Effects of particle size and surface charge on cellular uptake and biodistribution of polymeric nanoparticles. *Biomaterials 31*, 3657–3666.

41. Arvizo, R. R., Miranda, O. R., Thompson, M. A., Pabelick, C. M., Bhattacharya, R., Robertson, J. D., Rotello, V. M., Prakash, Y. S., Mukherjee, P. (2010). Effect of nanoparticle surface charge at the plasma membrane and beyond. *Nano Letters 10*, 2543–2548.

42. Poon, Z., Chang, D., Zhao, X., Hammond, P. T. (2011). Layer-by-layer nanoparticles with a pH-sheddable layer for in vivo targeting of tumor hypoxia. *ACS Nano 5*, 4284–4292.

43. Verma, A., Uzun, O., Hu, Y., Hu, Y., Han, H.-S., Watson, N., Chen, S., Irvine, D. J., Stellacci, F. (2008). Surface-structure-regulated cell-membrane penetration by monolayer-protected nanoparticles. *Nature Materials 7*, 588–595.

44. Yoon, D. J., Liu, C. T., Quinlan, D. S., Nafisi, P. M., Kamei, D. T. (2011). Intracellular trafficking considerations in the development of natural ligand-drug molecular conjugates for cancer. *Annals of Biomedical Engineering 39*, 1235–1251.

45. Doherty, G. J., Mcmahon, H. T. (2009). Mechanisms of endocytosis. *Annual Review of Biochemistry 78*, 857–902.

46. Grant, B. D., Donaldson, J. G. (2009). Pathways and mechanisms of endocytic recycling. *Nature Reviews Molecular Cell Biology 10*, 597–608.

47. Qian, Z. M., Li, H., Sun, H., Ho, K. (2002). Targeted drug delivery via the transferrin receptor-mediated endocytosis pathway. *Pharmacological Reviews 54*, 561–587.

48. Choi, C. H. J., Alabi, C. A., Webster, P., Davis, M. E. (2010). Mechanism of active targeting in solid tumors with transferrin-containing gold nanoparticles. *Proceedings of the National Academy of Sciences of the United States of America 107*, 1235–1240.

49. Davis, M. E., Zuckerman, J. E., Choi, C. H. J., Seligson, D., Tolcher, A., Alabi, C. A., Yen, Y., Heidel, J. D., Ribas, A. (2010). Evidence of RNAi in humans from systemically administered siRNA via targeted nanoparticles. *Nature 464*, 1067–1070.

50. Ferris, D. P., Lu, J., Gothard, C., Yanes, R., Thomas, C. R., Olsen, J.-C., Stoddart, J. F., Tamanoi, F., Zink, J. I. (2011). Synthesis of biomolecule-modified mesoporous silica nanoparticles for targeted hydrophobic drug delivery to cancer cells. *Small 7*, 1816–1826.

51. Wang, J., Tian, S., Petros, R. A., Napier, M. E., Desimone, J. M. (2010). The complex role of multivalency in nanoparticles targeting the transferrin receptor for cancer therapies. *Journal of the American Chemical Society 132*, 11306–11313.

52. Yang, X., Koh, C. G., Liu, S., Pan, X., Santhanam, R., Yu, B., Peng, Y., Pang, J., Golan, S., Talmon, Y., Jin, Y., Muthusamy, N., Byrd, J. C., Chan, K. K., Lee, L. J., Marcucci, G., Lee, R. J. (2008). Transferrin receptor-targeted lipid nanoparticles for delivery of an antisense oligodeoxyribonucleotide against Bcl-2. *Molecular Pharmaceutics 6*, 221–230.

53. Felber, A. E., Castagner, B., Elsabahy, M., Deleavey, G. F., Damha, M. J., Leroux, J.-C. (2011). siRNA nanocarriers based on methacrylic acid copolymers. *Journal of Controlled Release 152*, 159–167.

54. Sabharanjak, S., Mayor, S. (2004). Folate receptor endocytosis and trafficking. *Advanced Drug Delivery Reviews 56*, 1099–1109.

55. Yang, J., Chen, H., Vlahov, I. R., Cheng, J.-X., Low, P. S. (2006). Evaluation of disulfide reduction during receptor-mediated endocytosis by using FRET imaging. *Proceedings of the National Academy of Sciences of the United States of America 103*, 13872–13877.

56. Park, K. M., Lee, D. W., Sarkar, B., Jung, H., Kim, J., Ko, Y. H., Lee, K. E., Jeon, H., Kim, K. (2010). Reduction-sensitive, robust vesicles with a non-covalently modifiable surface as a multifunctional drug-delivery platform. *Small 6*, 1430–1441.

57. Pun, S. H., Davis, M. E. (2002). Development of a nonviral gene delivery vehicle for systemic application. *Bioconjugate Chemistry 13*, 630–639.

58. Kreuter, J. R., Shamenkov, D., Petrov, V., Ramge, P., Cychutek, K., Koch-Brandt, C., Alyautdin, R. (2002). Apolipoprotein-mediated transport of nanoparticle-bound drugs across the blood-brain barrier. *Journal of Drug Targeting 10*, 317–325.

59. Oba, M., Fukushima, S., Kanayama, N., Aoyagi, K., Nishiyama, N., Koyama, H., Kataoka, K. (2007). Cyclic RGD peptide-conjugated polyplex micelles as a targetable gene delivery system directed to cells possessing alphavbeta3 and alphavbeta5 integrins. *Bioconjugate Chemistry 18*, 1415–1423.

60. Oba, M., Aoyagi, K., Miyata, K., Matsumoto, Y., Itaka, K., Nishiyama, N., Yamasaki, Y., Koyama, H., Kataoka, K. (2008). Polyplex micelles with cyclic RGD peptide ligands and disulfide cross-links directing to the enhanced transfection via controlled intracellular trafficking. *Molecular Pharmaceutics 5*, 1080–1092.

61. Nativo, P., Prior, I. A., Brust, M. (2008). Uptake and intracellular fate of surface-modified gold nanoparticles. *ACS Nano 2*, 1639–1644.

62. Panyam, J., Zhou, W.-Z., Prabha, S., Sahoo, S. K., Labhasetwar, V. (2002). Rapid endo-lysosomal escape of poly(DL-lactide-*co*-glycolide) nanoparticles: implications for drug and gene delivery. *The FASEB Journal 16*, 1217–1226.

63. Patil, Y., Panyam, J. (2009). Polymeric nanoparticles for siRNA delivery and gene silencing. *International Journal of Pharmaceutics 367*, 195–203.

64. Kang, B., Mackey, M. A., El-Sayed, M. A. (2010). Nuclear targeting of gold nanoparticles in cancer cells induces DNA damage, causing cytokinesis arrest and apoptosis. *Journal of the American Chemical Society 132*, 1517–1519.

65. Weyland, M., Manero, F., Paillard, A., Gree, D., Viault, G., Jarnet, D., Menei, P., Juin, P., Chourpa, I., Benoit, J. P., Gree, R., Garcion, E. (2011). Mitochondrial targeting by use of lipid nanocapsules loaded with SV30, an analogue of the small-molecule Bcl-2 inhibitor HA14-1. *Journal of Controlled Release 151*, 74–82.

66. Boddapati, S. V., D'souza, G. G. M., Erdogan, S., Torchilin, V. P., Weissig, V. (2008). Organelle-targeted nanocarriers: specific delivery of liposomal ceramide to mitochondria enhances its cytotoxicity in vitro and in vivo. *Nano Letters 8*, 2559–2563.

67. Agemy, L., Friedmann-Morvinski, D., Kotamraju, V. R., Roth, L., Sugahara, K. N., Girard, O. M., Mattrey, R. F., Verma, I. M., Ruoslahti, E. (2011). Targeted nanoparticle enhanced proapoptotic peptide as potential therapy for glioblastoma. *Proceedings of the National Academy of Sciences of the United States of America 108*, 17450–17455.

68. Panyam, J., Labhasetwar, V. (2003). Dynamics of endocytosis and exocytosis of poly (D,L-lactide-*co*-glycolide) nanoparticles in vascular smooth muscle cells. *Pharmaceutical Research 20*, 212–220.

69. Slowing, I. I., Vivero-Escoto, J. L., Zhao, Y., Kandel, K., Peeraphatdit, C., Trewyn, B. G., Lin, V. S. Y. (2011). Exocytosis of mesoporous silica nanoparticles from mammalian cells: from asymmetric cell-to-cell transfer to protein harvesting. *Small 7*, 1526–1532.

70. Jiang, X., RoĈker, C., Hafner, M., Brandholt, S., DoŘlich, R. M., Nienhaus, G. U. (2010). Endo- and exocytosis of zwitterionic quantum dot nanoparticles by live HeLa cells. *ACS Nano 4*, 6787–6797.

71. Fang, C. Y., Vaijayanthimala, V., Cheng, C. A., Yeh, S. H., Chang, C. F., Li, C. L., Chang, H. C. (2011). The exocytosis of fluorescent nanodiamond and its use as a long-term cell tracker. *Small 7*, 3363–3370.

72. Yang, W., Lee, S., Lee, J., Bae, Y., Kim, D. (2010). Silver nanoparticle-induced degranulation observed with quantitative phase microscopy. *Journal of Biomedical Optics 15*, 045005.

73. Sandin, P., Fitzpatrick, L. W., Simpson, J. C., Dawson, K. A. (2012). High-speed imaging of Rab family small GTPases reveals rare events in nanoparticle trafficking in living cells. *ACS Nano 6*, 1513–1521.

74. Willig, K. I., Keller, J., Bossi, M., Hell, S. W. (2006). STED microscopy resolves nanoparticle assemblies. *New Journal of Physics 8*, 106.

75. Schübbe, S., Cavelius, C., Schumann, C., Koch, M., Kraegeloh, A. (2010). STED microscopy to monitor agglomeration of silica particles inside A549 cells. *Advanced Engineering Materials 12*, 417–422.

76. Schubbe, S., Schumann, C., Cavelius, C., Koch, M., Müller, T., Kraegeloh, A. (2011). Size-dependent localization and quantitative evaluation of the intracellular migration of silica nanoparticles in Caco-2 cells. *Chemistry of Materials 24*, 914–923.

77. Patil, Y., Sadhukha, T., Ma, L., Panyam, J. (2009). Nanoparticle-mediated simultaneous and targeted delivery of paclitaxel and tariquidar overcomes tumor drug resistance. *Journal of Controlled Release 136*, 21–29.

78. Patil, Y. B., Swaminathan, S. K., Sadhukha, T., Ma, L., Panyam, J. (2010). The use of nanoparticle-mediated targeted gene silencing and drug delivery to overcome tumor drug resistance. *Biomaterials 31*, 358–365.

79. Barua, S., Rege, K. (2009). Cancer-cell-phenotype-dependent differential intracellular trafficking of unconjugated quantum dots. *Small 5*, 370–376.

80. Barua, S., Rege, K. (2010). The influence of mediators of intracellular trafficking on transgene expression efficacy of polymer-plasmid DNA complexes. *Biomaterials 31*, 5894–5902.

# TOXICOLOGICAL ASSESSMENT OF NANOMEDICINE

Hayley Nehoff and Sebastien Taurin

*Department of Pharmacology and Toxicology, Otago School of Medical Sciences, University of Otago, Dunedin, New Zealand*

Khaled Greish

*Department of Pharmacology and Toxicology, Otago School of Medical Sciences, University of Otago, Dunedin, New Zealand; Department of Oncology, Faculty of Medicine, Suez Canal University, Egypt*

## 12.1 INTRODUCTION

The rapid evolution of nanotechnology has led to the realization of nanomedicine, defined as the use of engineered nanostructures for molecular level monitoring, repair, construction, and control of human biological systems [1]. Projections indicate that the nano pharmaceutical market will reach US$180 billion by 2015 [2]. Nanomedicine has originally focused on drug delivery and the treatment of cancer; however, new applications are emerging such as vaccine adjuvants, medical imaging, and gene therapy, the latter being applicable to a wide range of therapies ranging from Parkinson's disease treatment to osteogenesis [3]. The potential of nanotechnology to improve medicine is not limited to the use as a vehicle for medicines but also applicable for streamline discovery, design, development, and production of medications. These advances may reduce the time it takes to move a medicine from the bench to the clinic (currently 8–14 years) and the cost (current average of US$1.8 billion) [4]. On the other hand, the abundant use of and exposure to nanomaterials naturally raises

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*,
Edited by Shayne C. Gad
Copyright © 2013 John Wiley & Sons, Inc.

the concern of potential toxicity. However, limited data are available regarding toxicological evaluation of these new nanoconstructs, despite the accelerating rate of research papers describing new nanomaterials. The situation is similar regarding the investments in the emerging nanotechnology. For example, only 6.6% of the investment of the U.S. government in nanotechnologies is being spent on investigating the health and environmental consequences of their use [5].

In the United States, nanomaterials may, at the discretion of regulatory bodies, be deemed to fall under regulations pertaining to laws such as the toxic substances control act, while in the European Union they may fall under the regulation of the Registration, Evaluation, and Authorization of Chemicals Act [6]. Nanomedicine approval by regulatory bodies is subject to the standard preclinical, clinical, and postrelease regulations of traditional drugs. The majority of current nanomaterial regulation outlines are based on regulations developed for standard materials [7]. Early experimental studies to assess the toxicity of nanoparticles (NPs) rely on knowledge obtained from the studies on the nuisance of metal fumes, exposure to asbestos, synthetic fibers, silica, and dust. However, the increasing diversity of engineered nanomaterials such as metals, dendrimers, chitosan, silica, carbon nanotubes, and others used for medical applications have resulted in the need for improved methodologies to evaluate the safety of these new nanoconstructs [8]. Standard tests to evaluate toxicity are not always applicable to nanomedicine. For example, carbon nanotubes (CNTs) have been shown to cause a false cytotoxic effect in the 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide (MTT) assay, and titanium NPs (TiNPs) bind lactic dehydrogenase (LDH) to interfere with LDH assays [9–11]. A glaring example of the shortcomings of classical toxicological evaluation is dose measurements. Traditional mg/kg dose measurements still need to be investigated for comparison with other medicines and poisons. However, for the purposes of having meaningful measurements of the toxicity of nanomedicine, surface area should be considered as the most important determinant of reactivity and hence possible toxic interaction with biological systems. A battery of other characteristics such as shape, size, charge, and porosity are as important as surface area. Currently there is no satisfactory standard or consensus on how to define nanomedicine exposure and assess their toxicity. As shown in Table 12.1, most of the toxic attributes of a nanomedicine are governed by the physicochemical parameters of the nanoconstructs. Complete understanding of these parameters is essential for the prediction of a specific nanomedicines safety. Section 12.2 is a summary of the current knowledge on safety of commonly used nanoconstructs in medicine, followed by our recommendations for conducting safety studies for novel nanomedicines.

## 12.2 SAFETY OF COMMONLY USED NANOCONSTRUCTS IN NANOMEDICINE

### 12.2.1 Liposomes

Liposomes are widely used for drug and gene delivery [46,47]. Several types have been approved by the Food and Drug Administration (FDA) for clinical use [48] such

**TABLE 12.1 Description of the Administration Routes, Characteristics Affecting the Biodistribution and Molecular Effects of Various Types of Nanoparticles**

| | Liposome | Polymeric Micelle | Metal Nanospheres and Nanorods | Carbon | Dendrimers | Silica Nanoparticles |
|---|---|---|---|---|---|---|
| Administration route | Parenteral [12]<br>Subcutaneous [13]<br>Dermal [14]<br>Wound infiltration [15] | Parenteral [16]<br>Oral [17] | Parenteral [18,19]<br>Oral [18]<br>Inhalation [18] | Parenteral [20] | Parenteral [21,22]<br>Topical [21]<br>Oral [21]<br>Transdermal [21]<br>Ocular [21] | Parenteral [22] |
| Characteristics affecting biodistribution | Size [23]<br>Charge [24] | Size [25]<br>Core [25]<br>Surface [25]<br>Charge [25] | Size [26]<br>Charge [27]<br>Surface [28] | Agglomeration [29]<br>Size [30] | Size [22]<br>Charge [22]<br>Surface [22] | Size [31]<br>Shape [31]<br>Charge [32] |
| Molecular effects | Opsonization [33]<br>Astrocytosis [34]<br>Edema [34]<br>Inflammation [34] | Chronic liver toxicity [35] | Oxidative stress [36]<br>DNA damage [36]<br>DNA binding [37]<br>Inflammation [38]<br>Apoptosis [39] | Inflammation [40] | Hemotoxicity [22]<br>Hepatotoxicity [41] | Oxidative stress [42,43]<br>Inflammation [43–45] |

as Doxil, Danuxome, Myocet, Depocyt, and MEPACT [49–53]. Several studies have demonstrated that liposomes accumulate in organs such as the liver and spleen and are consumed by macrophages making these tissues susceptible to toxicity (Figure 12.1) [54]. Excessive liposome deposition in macrophages may affect their phagocytic capacity or alter their gene expression. Several *in vitro* experiments have shown that liposomes may downregulate the production of nitric oxide and tumor necrosis factor (TNF)-α by macrophages and may cause macrophage cell death [55]. However, there is no scientific evidence that therapeutic doses of various liposomes can induce clinically relevant immunosuppressive effects [56].

Depending on their composition, liposomes have also been observed to stimulate the innate immune responses [56]. Doxil, a polyethylene glycol (PEG)-grafted liposomal carrier for doxorubicin, has been shown to promote allergic reactions, despite the view of PEG as immunologically inert [57]. In the central nervous system (CNS) of rats, liposome administration has been observed to cause focal astrogliosis, extracellular edema, and inflammation [34]. The dependence of liposome toxicity on

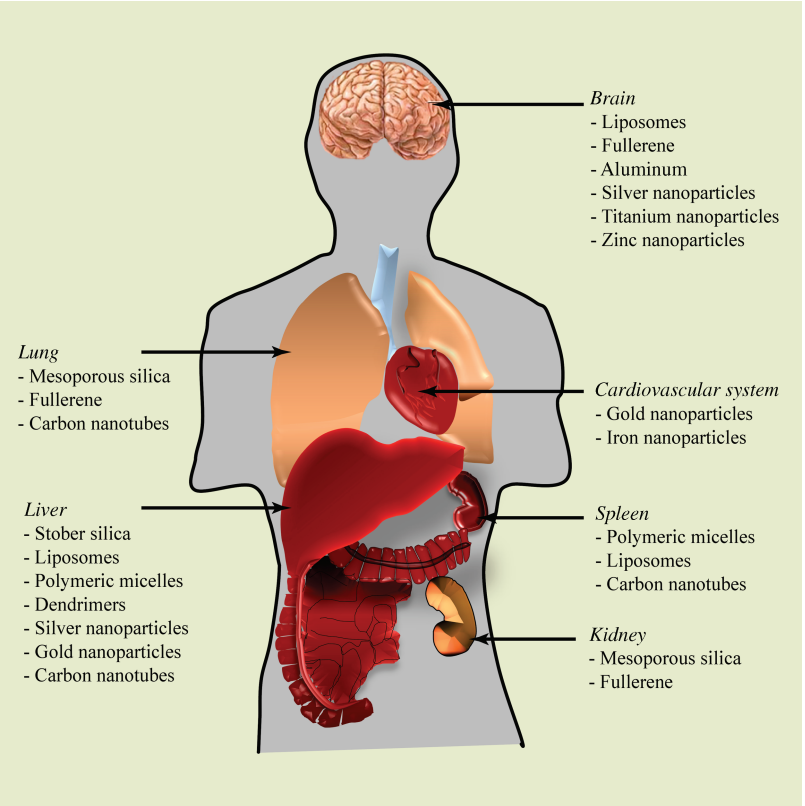


**FIGURE 12.1** Schematic representation of the toxicity associated with specific nanomedicines on human organs.

their composition has been documented both *in vitro* [58] and *in vivo* [59]. Lecithin–cholesterol–dicetyl phosphate or lecithin–cholesterol–stearylamine administered intracerebrally to mice produced neuronal edema and/or necrosis, generalized epileptic seizures and death due to respiratory failure immediately after injection, whereas lecithin–cholesterol–phosphatidic acid liposomes produced relatively little effect [59].

### 12.2.2 Polymer Conjugates and Polymeric Micelles

Biodegradable polymers including poly(D,L-lactic acid) and poly(D,L-lactic-*co*-glycolic acid) have been commonly used to structure NPs and encapsulate a high load of poorly soluble material [60]. Since the approval of the first polymeric conjugates, PEG conjugated adenosine deaminase (ADA) [61] and styrene maleic acid conjugated neocarzinostatin polymer (SMANCS), for cancer treatment in the early 1990s, the field of polymer-based nanomedicine has grown exponentially [35,36]. Synthetic polymers such as *N*-(2-hydroxypropyl) methacrylamide (HPMA) copolymers, poly (ethyleneimine) (PEI), divinyl ether and maleic anhydride (DIVEMA), poly(vinylpyrrolidone) (PVP), PEG, and linear polyamidoamines as well as pseudosynthetic polymers such as poly(amino acids), and natural polymers including dextran, dextrin, hyaluronic acid, and chitosans have been used at a clinical level as a polymeric platform for drug delivery [62].

A major drawback of the use of polymeric carriers is related to their rapid uptake and accumulation by the reticuloendothelial system (RES) located in the liver, spleen, and lung [63] (Figure 12.1), which reduce the dose that can reach the tumors and raise concerns of chronic toxicity in these organs [64]. An example is HPMA copolymers that accumulate in RES organs in a size-dependent manner [65]. Furthermore, polymers can influence the biological responses. For example, SMANCS and DIVEMA caused increased secretion of interferon-γ and -β and increased the activity of natural killer and cytotoxic T-lymphocyte cells [66–69]. Another aspect limiting the use of certain polymers is their degradation profile, essentially HPMA and PEG copolymers which are nonbiodegradable, a parameter essential to consider for nanomedicine development [70].

### 12.2.3 Dendrimers

Dendrimers have gained a lot of interest over the last few years as drug carriers because of their well-defined polymeric structure, low polydispersity index (PDI), and functionalized surface diversity. In 1985, Tomalia et al. described the synthesis of polyamidoamine (PAMAM) dendrimers [71], and subsequently an array of dendrimers has been synthesized utilizing Tomalia's principles. More recently, on the basis of their ability to disrupt protein aggregates, dendrimers have been studied for the treatment of prion diseases and neurodegenerative diseases and as an antiviral agent for the treatment of AIDS [72].

To date, only two dendrimers have reached clinical development. These are the contrast agent Gadomer®, which is in Phase II clinical trial for the delineation of

microvascular obstruction regions [73], and the intravaginal topical virucidal agent VivaGel® in Phase I clinical trial [74]. The complexity of obtaining chemically pure dendrimers is a major challenge for clinical development. Moreover, PAMAM dendrimers showed a distinct *in vivo* toxicity profile depending on the functional groups on their surface. Amine-terminated PAMAM dendrimers showed higher toxicity than carboxyl- or hydroxyl-terminated PAMAM dendrimers. Amine-terminated dendrimers triggered fatal hemotoxicity manifested as disseminated intravascular coagulation-like condition (DIC) and hemolysis.

### 12.2.4 Carbon Nanotubes

Carbon-based nanomaterials are used either as hollow spheres, ellipsoids or cylinders. Spheres and ellipsoids are referred to as fullerenes while cylinders are categorized as nanotubes. Fullerenes, especially C60, have exhibited some promise for drug delivery and for antiviral treatment as photosensitizers [75]. Pristine CNT and C60 are highly hydrophobic, chemically inert, and insoluble. Various biological and bioactive molecules such as proteins, enzymes, vaccines, DNA, and siRNA have been attached to CNTs and used for different biomedical applications [76]. More recently, CNT's photothermal properties have been exploited for their effect on cancer ablation [77]. CNTs exist in different forms such as single wall nanotubes (SWNT) and multiwalled nanotubes (MWNT), which can both be chemically modified and/or functionalized with bioactive molecules to enhance their solubility and drug targeting [78].

However, there is a growing concern on the use of fullerene and CNTs on the basis of *in vitro* and *in vivo* animal studies (Figure 12.1). For example, incubation of keratinocytes and bronchial epithelial cells with SWNTs has been shown to cause oxidative stress and generation of reactive oxygen species (ROS) [79]. Other reports described the oxidative stress and production of ROS following treatment by CNTs on macrophages and other cell lines, either be due to metal impurities [80] or as a result of the experimental conditions [81].

Acute and subacute toxicity studies of fullerenes by intraperitoneal injection in rats determined an approximate $LD_{50}$ of 600 mg/kg with kidney impairment often observed in deceased animals [82]. There was no mortality in an acute oral and skin toxicity test [82]. Lipid peroxidation in the brain of largemouth bass was observed after 48 h exposure to 0.5 ppm of soluble fullerene C60 [83]; although these effects were later found to be associated with impurities from the solvent used for the preparation of soluble fullerene [84], pure soluble fullerene showed antioxidant activity [84]. Soluble fullerene modified with two cationic chains (fulleropyrrolidine derivatives) triggered hemolysis and appeared cytotoxic to a variety of cell lines at concentration between 20 and 60 μM [85]. Several studies have looked at the potential effects of fullerene on embryonic development. For example, intraperitoneal injection of 50 mg/kg CNTs to pregnant rats (11 days gestation) with fullerene C60 solubilized with polyvinyl pyrrolidone was shown to cross the placental barrier and distribute into embryonic tissues, causing the death of the embryo as a consequence of several anomalies and malformations [86].

CNTs bear remarkable structural similarities to the carcinogenic asbestos fibers and have been shown to induce asbestos-like inflammation and foreign body granulomas in mice [87]. Pulmonary exposure of rodents to CNTs has been shown to induce inflammation and granuloma formation but evidence that these granulomas progress to mesotheliomas is yet to be found [88,89]. Pulmonary exposure to fullerene in rodents has also shown to induce lung inflammation, fibrosis, and DNA damage [90,91]. Exposure of adult male rats to 200 μg of CNTs has been observed to cause decreased sperm production [92]. One major concern for the application of fullerenes and CNTs in biomedicine is the lack of consistent data regarding the effect of functionalized particles and impurities associated with their synthesis on the biological environment. All of these parameters fundamentally contribute to the difference in biological responses observed across the different studies.

### 12.2.5 Silica NPs

Mesoporous silica NPs (MSNs) have a well-defined structure that can be functionalized with a wide range of organic functional groups [93]. In recent years, MSNs have been investigated for biomedical applications such as biosensors, enzyme immobilization, detection of cancer cells *in vivo*, targeted cancer therapy, DNA delivery, and drug delivery [94–99]. Recently, the FDA approved the first human clinical trial of Cornell dots, multimodal silica NPs that enclose several dye molecules to be used for the detection of cancer cells [100].

There are a few studies investigating toxic effects of MSNs. Studies determine that the cytotoxicity of MSNs depends on the shape and size and functionalized surface type [101]. Size-dependent cytotoxicity is well documented with smaller particles generally being more cytotoxic. For example, MSNs with an identical geometry can trigger a size-dependent necrosis in human endothelial cells *in vitro* [102]. Although subcutaneously injected MSNs did not show toxicity, intraperitoneal and intravenous injection of nonfunctionalized MSNs with sizes varying from 150 to 4000 nm, have been shown to be lethal in mice [103]. The effect is partially based on the hydrodynamic sizes of $SiO_2$ postprotein exposure in relation with the physicochemical characteristics of $SiO_2$ NPs and the vasculature which further promoted multiple organ failures [104]. This toxicity appears to result from the particles themselves and not from contaminants or degradation products [103]. MSN was also shown to increase hemolysis at concentrations above 100 μg/ml, an effect influenced by the surface area and the geometry of the MSNs [105]. Furthermore, silica NPs were shown to enhance thrombus formation following intratracheal instillation or intravenous injection in a hamster model of thrombosis [106]. Intravenous injection of MSNs has been found to induce liver injury such as inflammation, silicotic nodular-like lesions, fibrosis, Kupffer cell activation, and the development of oxidative stress-associated liver damage at high doses (50 mg/kg) in mice [45]. Altered biliary function and glomerular function have also been observed in mice following 20 mg/kg MSNs intravenous injection [107]. MSNs have been shown to cross the blood–brain barrier (BBB) and accumulate in the striatum, where they deplete dopamine and induce oxidative stress and inflammation, and to damage the murine placenta and enter the fetal brain [43,108].

### 12.2.6 Metal-Based NPs

Metal-based NPs such as aluminum, gold, silver, iron, zinc, and titanium are currently being developed as nanomedicines; however, their unique properties make it difficult to predict their toxicity. Several studies were conducted *in vitro* and *in vivo* to demonstrate the influence of intrinsic parameters such as size and shape on their adverse toxicological effects. Figure 12.2 summarizes the physicochemical properties of the different nanomedicines that may influence biocompatibility and the possible pathways involved in their toxicity.

Aluminum (Al) is currently utilized for medical applications such as antacids, buffered aspirin, and vaccine adjuvants [109]. Al has no known biological function but has been demonstrated to cause significant toxicity such as dialysis encephalopathy, osteomalacia, and microcytic anemia [110]. The role of Al in dialysis encephalopathy is of particular concern and has been recognized in humans as early as 1979 when its ability to pass through the BBB and accumulate in the CNS to cause dementia was first reported in patients on chronic hemodialysis [111]. Of the various forms of dementia thought to be related to Al exposure, Alzheimer's disease has since been the most intensely studied [112]. Despite this research as well as observations of increased Al in the brains of Alzheimer patients and its ability to enhance amyloid beta penetration [113,114], there is no concrete evidence for the causality and this topic is still controversial [115].

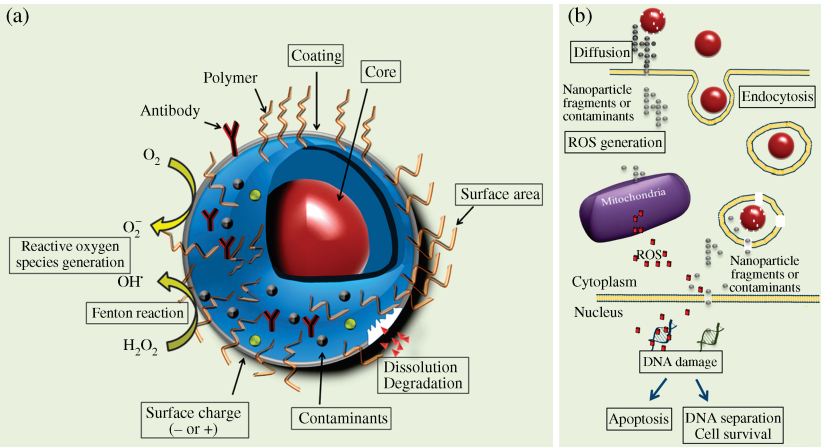


**FIGURE 12.2** Physicochemical properties of nanomedicines that may influence biocompatibility. (a) Illustration of the importance of the nanomaterial composition, surface coating, solubility, degradation, dissolution, and contaminants with biological tissue. (b) Possible mechanistic pathway for toxicity induced by exposure to nanomedicines. Nanomedicines may degrade rapidly outside the cells or following endocytosis. The resulting fragments or contaminants may target mitochondria and promote oxidative stress due to the increased production of ROS. The ROS and the nanoparticles fragments or contaminants trigger downstream signaling responses that promote genotoxicity. ROS, reactive oxygen species.

Gold NPs (Au NPs) have been extensively studied for use in nanomedicine. Colloidal gold has been used since the 1920s for the treatment of tuberculosis before the emergence of antibiotics and has recently found some applications for the treatment of rheumatic disease [116]. More recently, Au NPs have been investigated for photothermal therapy, imaging, gene therapy, and drug delivery [117–120]. Biodistribution studies of Au NPs indicate that they preferentially accumulate in the liver, spleen, kidney, and lungs in a size-dependent manner with smaller particles (<50 nm) showing dispersion throughout all tissues [121]. Several studies have examined the effect of Au NP size on toxicity and have demonstrated that nanoclusters of 1.4 nm exhibit a higher cytotoxicity compared to nanoclusters of 15 nm [122]. Animal studies have also demonstrated that Au NPs can cause damage to the liver, cardiovascular system, and immune system [123,124] in a size-dependent manner. Additional studies have determined the influence of the charge, shape, and surfactant used for synthesis of the Au NPs on the potential cytotoxicity [122].

Metal-based NPs have also been demonstrated to induce oxidative stress following administration. Iron (Fe) NPs are being developed for use in magnetic resonance imaging, photothermal therapy, tissue repair, and drug delivery. The Fe NP Feraheme was approved in 2009 for the treatment of iron deficiency [125,126]. Oxidative stress is consistently observed following Fe NP treatment [125,127,128] and is highly concerning due to its potential implication in a variety of diseases including neurodegenerative diseases, cardiovascular diseases, and cancer [129–131]. This oxidative stress is also likely to be the cause of the Fe NP-induced genotoxicity [132,133]. The enhanced permeability observed *in vitro* at 50 μg/ml and the enhanced microvasculature permeability observed *in vivo* at 0.8 mg/kg [134,135] indicate that Fe NPs may play a role in the development of cardiovascular diseases, which is supported by the observation of Fe NPs in human atherosclerotic plaques [130,136]. Silver (Ag) NPs have been shown to cross the BBB and trigger oxidative stress in the CNS and impair short-term memory [137,138]. Titanium (Ti) NPs and zinc (Zn) NPs have also been shown to be associated with CNS toxicity and impairment of the spatial recognition and memory [139–141].

## 12.3 EVALUATION OF NANOMATERIAL TOXICITY

A classical scheme for toxicity assessment is represented in Figure 12.3. In Section 12.3.1, we briefly summarize the principles of toxicological assessment of nanomaterials. Toxicity assessments are classified into acute, subacute, subchronic, and chronic studies. Absorption, distribution, metabolism, excretion, and pharmacokinetics (ADME/PK), carcinogenic, and teratogenic studies are also required for complete evaluation of nanomaterial toxicity.

### 12.3.1 Characterization

Thorough and systematic physicochemical characterization of materials is essential prior to studying their toxic effects *in vivo*. One of the main reasons is to ensure the

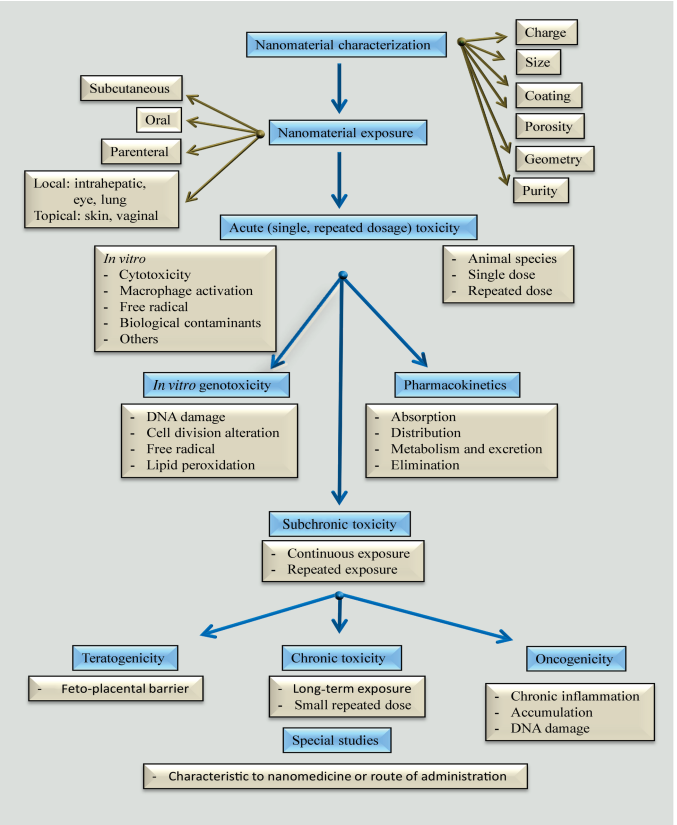


**FIGURE 12.3** Proposed paradigm for toxicity testing of nanomedicines.

accuracy and reproducibility of the observed toxic effects of the studied NPs. Toxicity of NPs is significantly influenced by their physicochemical properties such as size, shape, surface charge, charge density, composition, density of structure, presence of pores, and surface activating sites. Hence, documentation of the characteristics of the NP under evaluation becomes crucial in order to correlate with the observed biological effects *in vivo*. Characterization data are important to facilitate comparison of toxicity results of a given NP with other nanoconstructs. Instrumentation techniques have advanced considerably in the past decades to sufficiently characterize the physicochemical properties of NPs, including dynamic light scattering techniques for size determination, zeta potential measurements for relative particle charge, electron microscopy (scanning and transmission) to qualitatively analyze the size and shape, nuclear magnetic resonance to accurately determine chemical structures, X-ray photoelectron spectroscopy to measure chemical compositions and chemical and electronic states, mass spectrometry for mass and elemental analysis and spectroscopy techniques that measure absorption, emission, or scattering of either wavelength or frequency for the determination

of the size and distribution. An important note of caution while characterizing nanomaterials is to evaluate these properties under physiologically relevant conditions.

### 12.3.2 Absorption, Distribution, Metabolism, and Excretion (ADME)

Understanding the pharmacokinetics of NPs is an integral part of assessing their toxicity as it can facilitate an in-depth understanding of nanomaterial behavior in a biological system and provide valuable information which serves as a prelude to Phase I clinical study in humans. ADME studies are performed in at least one rodent species and one nonrodent species. In humans, ADME studies (usually a Phase I clinical trial) are performed in 8–10 healthy male subjects, unless the drug in question is a cytotoxic cancer treatment which is studied in patients instead of healthy volunteers [142]. The route of administration should be the proposed route for clinical application and usually a single dose is sufficient [142].

Absorption studies investigate the systemic exposure relative to the route of administration. *In vitro* absorption studies can be used to estimate *in vivo* absorption such as the Caco-2 cell monolayer to estimate intestinal absorption or a human skin equivalent model for dermal absorption [143,144]. *In vivo* oral absorption is usually investigated in a rodent model, where the plasma concentration is measured at certain time points following oral administration [145]. Microscopy methods such as scanning electron microscopy and multiphoton microscopy can be used to determine dermal NP absorption *in vivo* [146]. The *in vivo* whole body distribution of a radiolabeled nanomedicine may be quantitatively evaluated using whole body autoradiography or inductively coupled plasma mass spectroscopy [147,148]. At specified intervals, plasma samples may be taken to determine multiple pharmacokinetic parameters such as $T_{1/2}$, $C_{max}$, AUC, $T_{max}$ and $V_d$ [142]. It is critical that the radiolabel/fluorophore does not dissociate from the nanomedicine during metabolism and does not interfere with the activity of the nanomedicine, its metabolism, and/or stability [142]. The hepatic microsome test investigates potential metabolism of nanomedicines by the enzymes of the smooth endoplasmic reticulum (SER), for example, CYP450 enzymes and glucuronosyltransferases. Studies utilizing isolated hepatocytes are also conducted as these retain SER, cytosolic, and mitochondrial enzymes which may also metabolize the nanomedicine [149,150]. Excreted metabolites may be identified to determine the likely metabolic pathways *in vivo* such as hydroxylation, glucuronidation, sulfation, and methylation [142]. Following the dosing, urine, fecal, and bile samples are obtained at specified intervals to study the excretion mechanism of the nanomedicine (for more information consult [142]).

### 12.3.3 Acute and Chronic Toxicity

Acute and subacute toxicity studies (Table 12.2) may be measured based on a single or repeated dose administration via various methods of administration such as intravenous, oral (one dose per day), dermal, or inhalation. Typically, acute toxicity studies require five animals/sex/dose to be monitored over a period of ~14 days with

**TABLE 12.2 *In Vitro* and *In Vivo* Assays to Evaluate the Potential Toxic Effect of Nanomedicines**

| Toxicity | *In Vitro* | *In Vivo* |
|---|---|---|
| Genotoxicity | Ames test<br>Comet test | Micronuclei test<br>Unscheduled DNA test |
| Carcinogenicity | Ames test<br>Comet test<br>Syrian hamster embryo assay<br>Immortalized aneuploid mouse cell assay | Two-year rodent cancer bioassay |
| Reproductive toxicity | Mouse morula assay<br>Mouse blastocyst assay<br>Sperm viability assay | One generation rodent reproductive toxicity study<br>Two generation rodent reproductive toxicity study<br>Prenatal development toxicity study |
| Acute and subacute toxicity | — | Rodent acute and subacute toxicity studies<br>Dog acute and subacute toxicity studies<br>Primate acute and subacute toxicity studies |
| Chronic and subchronic toxicity | — | Rodent chronic and subchronic toxicity studies<br>Primate chronic and subchronic toxicity studies |

outcomes of interest being weight change, response to the dose, function of organ systems, mortality, clinical pathology, and gross necropsy [151]. For subacute studies, 10 rodents/sex/dose are monitored over 4–5 weeks. Subchronic and chronic studies measure the effects of chronic dosing and exposure over the majority of the lifespan of the animal. Subchronic studies span ~13 weeks while chronic studies span 18–30 months. Subchronic and chronic studies measure the same outcomes as acute studies with the addition of ocular toxicity, cardiovascular function, neurotoxicity, and immunotoxicity [151].

### 12.3.4 Genotoxicity

Genotoxicity is a vital test (see Table 12.2) for any new medication as it is an important indicator of the potential for the medication to induce teratogenic or carcinogenic effects. The most common *in vitro* tests are the Ames test and the comet test. The Ames involves histidine-dependent cell lines of *Salmonella typhimurium,* which are unable to grow on histidine-deficient media. If the bacteria are able to grow on the media following incubation with the nanomedicine in significantly higher numbers than the control, it is likely that the material is mutagenic [152]. The comet

test involves mammalian cells, usually lymphocytes, being exposed to the material and then lysed, their genetic material extracted and then run on an electrophoresis gel. Broken DNA unwinds and so is unable to move through the gels with the same speed as intact DNA, resulting in a dense ball of intact DNA with a spread of DNA following in a pattern much similar to the tail of a comet [153].

The *in vivo* tests commonly performed are the micronuclei test and the unscheduled DNA synthesis test. The micronuclei test is reliable and well validated. The premise is that, as an erythroblast develops into an erythrocyte, the nucleus of the cell is excised but the micronuclei (a signal of incorrect chromosome distribution during anaphase) may be left behind; therefore, the presence of a significantly larger number of micronuclei than the negative control signals genotoxicity of the material [154]. The unscheduled DNA synthesis test measures DNA synthesis in mammalian liver cells that are not undergoing S phase scheduled DNA synthesis. DNA synthesis that is not scheduled is assumed to be a result of DNA repair following damage by the compound being investigated [155].

### 12.3.5 Carcinogenicity

Carcinogenicity is a serious concern for any medication. It may occur via two main mechanisms known as genotoxic and nongenotoxic carcinogenicity with genotoxic carcinogens being up to 50 times more potent than nongenotoxic carcinogens [156]. Nongenotoxic carcinogens have their effect via disruption of cellular metabolic processes and signaling pathways. Nongenotoxic carcinogens may, for example, alter the production of hormones that regulate the cellular mitotic rate (17β-estradiol), act as immunosuppressant (cyclosporin) or induce tissue-specific toxicity and inflammatory responses (arsenic) [157].

*In vitro* carcinogenicity tests (see Table 12.2) generally involve much of the same genotoxicity tests as above but also involve cell transformation assays using Syrian hamster embryo cell lines and immortalized aneuploid mouse cell lines [158]. The Syrian hamster embryo cell test is often used and signals carcinogenicity of a substance via morphological transformation of cell cultures [158]. The immortalized aneuploid mouse cell lines signals carcinogenicity via a progression from immortality to tumorigenicity [158]. The two-year cancer bioassay in rodents is considered the gold standard of carcinogenicity tests as it monitors the rodent for the greater portion of their lifespan. This test usually requires at least three different dose groups with at least 50 animals/sex/dose. The dose is administered daily for 24 months during which time tumor development, time of tumor onset, tumor location, dimensions, appearance, and progression are closely monitored [159].

### 12.3.6 Reproductive and Developmental Toxicity

Reproductive toxicity studies are necessary to gauge the possible effects of nanomedicine on reproductive ability and teratogenicity. Although some teratogenic effects are a consequence of genotoxicity, some are nongenotoxic such as retinoic

acid. *In vitro* studies to evaluate possible reproductive and developmental toxicity include studies of sperm viability and murine embryos (Table 12.2). The evaluation of various parameters of sperm health including motility [160], apoptosis and necrosis, mitochondrial function, and LDH leakage [161] following incubation with the nanomedicines may also be informative. The effects of the nanomedicine in question on the development and proliferation of mouse morulas and blastocysts may also be studied *in vitro* to identify the effects of the nanomedicines on the viability and health of the embryo [162]. A one-generation *in vivo* study involves both males and females [163]. Sufficient numbers of animals are used in order to give 20 litters (assuming sterility is not an issue). The compound of interest is administered to the animals, males 2 weeks prior to, during, and after mating and females for 54 days to allow for 14 days premating, 14 days mating, 22 days gestation, and 4 days lactation. The parameters are pregnancy duration, stillbirths, live births, and litter number, size, sex, and abnormalities. Gross necropsy and histopathological examinations are also performed.

## 12.4 CONCLUSIONS

Nanomedicine has great potential to improve medical competences. Inconsistent classification, inadequate regulation, limited knowledge of its safety, and flaws in toxicity screening need to be addressed for this technology to provide the maximum benefit to mankind in general and in particular to cancer patients.

### Acknowledgments

This work has been supported by Departmental fund no. PL. 108403.01.S. LM to KG and HRC Emerging Researcher First Grant (PL.108360.01.P.LM) to Sebastien Taurin from department of pharmacology and toxicology, Otago. Khaled Greish thanks Ms Rebecca Cookson for proofreading the manuscript.

## REFERENCES

1. Webster, P., (2005). World Nanobiotechnology Market, Frost & Sullivan. *Nanomedicine*, *1*, 140–142.
2. Hobson, D. W. (2009). Commercialization of nanotechnology. *Wiley Interdisciplinary Reviews. Nanomedicine and Nanobiotechnology 1* (2), 189–202.
3. Krebs, M. D., Salter, E., Chen, E., Sutter, K. A., Alsberg, E. (2010). Calcium phosphate-DNA nanoparticle gene delivery from alginate hydrogels induces *in vivo* osteogenesis. *Journal of Biomedical Materials Research Part A 92* (3), 1131–1138.
4. Paul, S. M., Mytelka, D. S., Dunwiddie, C. T., Persinger, C. C., Munos, B. H., Lindborg, S. R., Schacht, A. L. (2010). How to improve R&D productivity: the pharmaceutical industry's grand challenge. *Nature Reviews Drug Discovery 9* (3), 203–214.

5. Roco, M. C. The long view of nanotechnology development: The National Nanotechnology Initiative at 10 years. 1 ed. Nanotechnology research directions for societal needs in 2020, ed. M.C.M. Roco, C. A. Hersam, M. C. 2011.
6. Gwinn, M. R., Tran. L. (2010). Risk management of nanomaterials. *Wiley Interdisciplinary Reviews. Nanomedicine and Nanobiotechnology 2* (2), 130–137.
7. Chan, V. S. (2006). Nanomedicine: an unresolved regulatory issue. *Regulatory Toxicology and Pharmacology 46* (3), 218–224.
8. Fadeel, B., Garcia-Bennett, A. E. (2010). Better safe than sorry: understanding the toxicological properties of inorganic nanoparticles manufactured for biomedical applications. *Advanced Drug Delivery Reviews 62* (3), 362–374.
9. Worle-Knirsch, J. M., Pulskamp, K., Krug, H. F. (2006). Oops they did it again! Carbon nanotubes hoax scientists in viability assays. *Nano Letters 6* (6), 1261–1268.
10. Casey, A., Herzog, E., Davoren, M., Lyng, F. M., Byrne, H. J., Chambers, G. (2007). Spectroscopic analysis confirms the interactions between single walled carbon nanotubes and various dyes commonly used to assess cytotoxicity. *Carbon 45* (7), 1425–1432.
11. Zaqout, M. S., Sumizawa, T., Igisu, H., Wilson, D., Myojo, T., Ueno, S. (2011). Binding of titanium dioxide nanoparticles to lactate dehydrogenase. *Environmental Health and Preventive Medicine 17* (4), 341–345.
12. Cortes, J., Di Cosimo, S., Climent, M. A., Cortes-Funes, H., Lluch, A., Gascon, P., Mayordomo, J. I., Gil, M., Benavides, M., Cirera, L., Ojeda, B., Rodriguez, C. A., Trigo, J. M., Vazquez, J., Regueiro, P., Dorado, J. F., Baselga, J. (2009). Nonpegylated liposomal doxorubicin (TLC-D99), paclitaxel, and trastuzumab in HER-2-overexpressing breast cancer: a multicenter phase I/II study. *Clinical Cancer Research 15* (1), 307–314.
13. Buse, J., El-Aneed, A. (2010). Properties, engineering and applications of lipid-based nanoparticle drug-delivery systems: current research and advances. *Nanomedicine (Lond) 5* (8), 1237–1260.
14. Verma, D. D., Verma, S., Blume, G., Fahr, A. (2003). Particle size of liposomes influences dermal delivery of substances into skin. *International Journal of Pharmaceutics 258* (1–2), 141–151.
15. Golf, M., Daniels, S. E., Onel, E. (2011). A phase 3, randomized, placebo-controlled trial of DepoFoam(R) bupivacaine (extended-release bupivacaine local analgesic) in bunionectomy. *Advances in Therapy 28* (9), 776–788.
16. Gaucher, G., Marchessault, R. H., Leroux, J. C. (2010). Polyester-based micelles and nanoparticles for the parenteral delivery of taxanes. *Journal of Controlled Release 143* (1), 2–12.
17. Gaucher, G., Satturwar, P., Jones, M. C., Furtos, A., Leroux, J. C. (2010). Polymeric micelles for oral drug delivery. *European Journal of Pharmaceutics and Biopharmaceutics 76* (2), 147–158.
18. Arvizo, R., Bhattacharya, R., Mukherjee, P. (2010). Gold nanoparticles: opportunities and challenges in nanomedicine. *Expert Opinion on Drug Delivery 7* (6), 753–763.
19. De Jong, W. H., Hagens, W. I., Krystek, P., Burger, M. C., Sips, A. J., Geertsma, R. E. (2008). Particle size-dependent organ distribution of gold nanoparticles after intravenous administration. *Biomaterials 29* (12), 1912–1919.
20. Liu, Z., Tabakman, S., Welsher, K., Dai, H. (2009). Carbon nanotubes in biology and medicine: *in vitro* and *in vivo* detection. *Imaging and Drug Delivery. Nano Research 2* (2), 85–120.

21. Jain, N. K., Asthana, A. (2007). Dendritic systems in drug delivery applications. *Expert Opinion on Drug Delivery 4* (5), 495–512.
22. Greish, K., Thiagarajan, G., Herd, H., Price, R., Bauer, H., Hubbard, D., Burckle, A., Sadekar, S., Yu, T., Anwar, A., Ray, A., Ghandehari, H. (2011). Size and surface charge significantly influence the toxicity of silica and dendritic nanoparticles. *Nanotoxicology. 6*, 713–723.
23. Liu, D., Mori, A., Huang, L. (1992). Role of liposome size and RES blockade in controlling biodistribution and tumor uptake of GM1-containing liposomes. *Biochimica et Biophysica Acta 1104* (1), 95–101.
24. Henriksen-Lacey, M., Christensen, D., Bramwell, V. W., Lindenstrom, T., Agger, E. M., Andersen, P., Perrie, Y. (2010). Liposomal cationic charge and antigen adsorption are important properties for the efficient deposition of antigen at the injection site and ability of the vaccine to induce a CMI response. *Journal of Controlled Release 145* (2), 102–108.
25. Alexis, F., Pridgen, E., Molnar, L. K., Farokhzad, O. C. (2008). Factors affecting the clearance and biodistribution of polymeric nanoparticles. *Molecular Pharmacology 5* (4), 505–515.
26. Sonavane, G., Tomoda, K., Makino, K. (2008). Biodistribution of colloidal gold nanoparticles after intravenous administration: effect of particle size. *Colloids and Surfaces B: Biointerfaces 66* (2), 274–280.
27. Hirn, S., Semmler-Behnke, M., Schleh, C., Wenk, A., Lipka, J., Schäffler, M., Takenaka, S., Möller, W., Schmid, G., Simon, U., Kreyling, W. G. (2011). Particle size-dependent and surface charge-dependent biodistribution of gold nanoparticles after intravenous administration. *European Journal of Pharmaceutics and Biopharmaceutics 77* (3), 407–416.
28. Duffin, R., Tran, L., Brown, D., Stone, V., Donaldson, K. (2007). Proinflammogenic effects of low-toxicity and metal nanoparticles *in vivo* and *in vitro*: highlighting the role of particle surface area and surface reactivity. *Inhalation Toxicology 19* (10), 849–856.
29. Wick, P., Manser, P., Limbach, L. K., Dettlaff-Weglikowska, U., Krumeich, F., Roth, S., Stark, W. J., Bruinink, A. (2007). The degree and kind of agglomeration affect carbon nanotube cytotoxicity. *Toxicology Letters 168* (2), 121–131.
30. Sato, Y., Yokoyama, A., Shibata, K., Akimoto, Y., Ogino, S., Nodasaka, Y., Kohgo, T., Tamura, K., Akasaka, T., Uo, M., Motomiya, K., Jeyadevan, B., Ishiguro, M., Hatakeyama, R., Watari, F., Tohji, K. (2005). Influence of length on cytotoxicity of multi-walled carbon nanotubes against human acute monocytic leukemia cell line THP-1 *in vitro* and subcutaneous tissue of rats *in vivo*. *Molecular Biosystems 1* (2), 176–182.
31. Decuzzi, P., Godin, B., Tanaka, T., Lee, S. Y., Chiappini, C., Liu, X., Ferrari, M. (2010). Size and shape effects in the biodistribution of intravascularly injected particles. *Journal of Controlled Release 141* (3), 320–327.
32. Tao, Z., Toms, B. B., Goodisman, J., Asefa, T. (2009). Mesoporosity and functional group dependent endocytosis and cytotoxicity of silica nanomaterials. *Chemical Research in Toxicology 22* (11), 1869–1880.
33. Patel, H. M., Moghimi, S. M. (1998). Serum-mediated recognition of liposomes by phagocytic cells of the reticuloendothelial system—the concept of tissue specificity. *Advanced Drug Delivery Reviews 32* (1–2), 45–60.
34. Bell, H., Kimber, W. L., Li, M., Whittle, I. R. (1998). Liposomal transfection efficiency and toxicity on glioma cell lines: *in vitro* and *in vivo* studies. *Neuroreport 9* (5), 793–798.

35. Kasai, T., Matsumura, S., Iizuka, T., Shiba, K., Kanamori, T., Yudasaka, M., Iijima, S., Yokoyama, A. (2011). Carbon nanohorns accelerate bone regeneration in rat calvarial bone defect. *Nanotechnology 22* (6), 065102.
36. Trouiller, B., Reliene, R., Westbrook, A., Solaimani, P., Schiestl, R. H. (2009). Titanium dioxide nanoparticles induce DNA damage and genetic instability *in vivo* in mice. *Cancer Research 69* (22), 8784–8789.
37. Goodman, C. M., McCusker, C. D., Yilmaz, T., Rotello, V. M. (2004). Toxicity of gold nanoparticles functionalized with cationic and anionic side chains. *Bioconjugate Chemistry 15* (4), 897–900.
38. Cho, W. S., Kim, S., Han, B. S., Son, W. C., Jeong, J. (2009). Comparison of gene expression profiles in mice liver following intravenous injection of 4 and 100 nm-sized PEG-coated gold nanoparticles. *Toxicology Letters 191* (1), 96–102.
39. Wang, X., Wang, Y., Chen, Z. G., Shin, D. M. (2009). Advances of cancer therapy by nanotechnology. *Cancer Research and Treatment 41* (1), 1–11.
40. Crouzier, D., Follot, S., Gentilhomme, E., Flahaut, E., Arnaud, R., Dabouis, V., Castellarin, C., Debouzy, J. C. (2010). Carbon nanotubes induce inflammation but decrease the production of reactive oxygen species in lung. *Toxicology 272* (1–3), 39–45.
41. Jain, R. K., Stylianopoulos, T. (2010). Delivering nanomedicine to solid tumors. *Nature Reviews Clinical Oncology 7* (11), 653–664.
42. Park, E. J., Park, K. (2009). Oxidative stress and pro-inflammatory responses induced by silica nanoparticles *in vivo* and *in vitro*. *Toxicology Letters 184* (1), 18–25.
43. Wu, J., Wang, C., Sun, J., Xue, Y. (2011). Neurotoxicity of silica nanoparticles: brain localization and dopaminergic neurons damage pathways. *ACS Nano 5* (6), 4476–4489.
44. Nishimori, H., Kondoh, M., Isoda, K., Tsunoda, S., Tsutsumi, Y., Yagi, K. (2009). Silica nanoparticles as hepatotoxicants. *European Journal of Pharmaceutics and Biopharmaceutics 72* (3), 496–501.
45. Liu, T., Li, L., Fu, C., Liu, H., Chen, D., Tang, F. (2012). Pathological mechanisms of liver injury caused by continuous intraperitoneal injection of silica nanoparticles. *Biomaterials 33* (7), 2399–2407.
46. Kshirsagar, N. A., Pandya, S. K., Kirodian, G. B., Sanath, S. (2005). Liposomal drug delivery system from laboratory to clinic. *Journal of Postgraduate Medicine 51*S5–S15.
47. Sonoke, S., Ueda, T., Fujiwara, K., Kuwabara, K., Yano, J. (2011). Galactose-modified cationic liposomes as a liver-targeting delivery system for small interfering RNA. *Biological & Pharmaceutical Bulletin 34* (8), 1338–1342.
48. Boisseau, P., Loubaton, B. (2011). Nanomedicine, nanotechnology in medicine. *Comptes Rendus Physique 12* (7), 620–636.
49. FDA. (1996). FDA approves DaunoXome as first-line therapy for Kaposi's sarcoma. Food and Drug Administration. *Journal of International Association of Physicians in AIDS Care 2* (5), 50–51.
50. Gordon, A. N., Fleagle, J. T., Guthrie, D., Parkin, D. E., Gore, M. E., Lacave, A. J. (2001). Recurrent epithelial ovarian carcinoma: a randomized phase III study of pegylated liposomal doxorubicin versus topotecan. *Journal of Clinical Oncology 19* (14), 3312–3322.
51. Chhikara, B. S., Parang, K. (2010). Development of cytarabine prodrugs and delivery systems for leukemia treatment. *Expert Opinion on Drug Delivery 7* (12), 1399–1414.

52. Batist, G., Barton, J., Chaikin, P., Swenson, C., Welles, L. (2002). Myocet (liposome-encapsulated doxorubicin citrate): a new approach in breast cancer therapy. *Expert Opinion on Pharmacotherapy 3* (12), 1739–1751.
53. Ando, K., Mori, K., Corradini, N., Redini, F., Heymann, D. (2011). Mifamurtide for the treatment of nonmetastatic osteosarcoma. *Expert Opinion on Pharmacotherapy 12* (2), 285–292.
54. Yatvin, M. B., Lelkes, P. I. (1982). Clinical prospects for liposomes. *Medical Physics 9* (2), 149–175.
55. Filion, M. C., Phillips, N. C. (1997). Toxicity and immunomodulatory activity of liposomal vectors formulated with cationic lipids toward immune effector cells. *Biochimica et Biophysica Acta 1329* (2), 345–356.
56. Szebeni, J., Moghimi, S. M. (2009). Liposome triggering of innate immune responses: a perspective on benefits and adverse reactions. *Journal of Liposome Research 19* (2), 85–90.
57. Hunter, A. C., Moghimi, S. M. (2002). Therapeutic synthetic polymers: a game of Russian roulette? *Drug Discovery Today 7* (19), 998–1001.
58. Mayhew, E., Ito, M., Lazo, R. (1987). Toxicity of non-drug-containing liposomes for cultured human cells. *Experimental Cell Research 171* (1), 195–202.
59. Adams, D. H., Joyce, G., Richardson, V. J., Ryman, B. E., Wisniewski, H. M. (1977). Liposome toxicity in the mouse central nervous system. *Journal of the Neurological Sciences 31* (2), 173–179.
60. Kuijpers, S. A., Coimbra, M. J., Storm, G., Schiffelers, R. M. (2010). Liposomes targeting tumour stromal cells. *Molecular Membrane Biology 27* (7), 328–340.
61. Chun, J. D., Lee, N., Kobayashi, R. H., Chaffee, S., Hershfield, M. S., Stiehm, E. R. (1993). Suppression of an antibody to adenosine-deaminase (ADA) in an ADA-deficient patient receiving polyethylene glycol modified adenosine deaminase. *Annals of Allergy 70* (6), 462–466.
62. Duncan, R. (2003). The dawning era of polymer therapeutics. *Nature Reviews Drug Discovery 2* (5), 347–360.
63. Nishiyama, N., Kataoka, K. (2006). Current state, achievements, and future prospects of polymeric micelles as nanocarriers for drug and gene delivery. *Pharmacology & Therapeutics 112* (3), 630–648.
64. Garnett, M. C., Kallinteri, P. (2006). Nanomedicines and nanotoxicology: some physiological principles. *Occupational Medicine 56* (5), 307–311.
65. Allmeroth, M., Moderegger, D., Biesalski, B., Koynov, K., Rosch, F., Thews, O., Zentel, R. (2011). Modifying the body distribution of HPMA-based copolymers by molecular weight and aggregate formation. *Biomacromolecules 12* (7), 2841–2849.
66. Suzuki, F., Pollard, R. B., Maeda, H. (1989). Stimulation of non-specific resistance to tumors in the mouse using a poly(maleic-acid-styrene)-conjugated neocarzinostatin. *Cancer Immunology, Immunotherapy 30* (2), 97–104.
67. Breslow, D. S., Edwards, E. I., Newburg, N. R. (1973). Divinyl ether-maleic anhydride (pyran) copolymer used to demonstrate the effect of molecular weight on biological activity. *Nature 246* (5429), 160–162.
68. Suzuki, F., Matsumoto, K., Schmitt, D. A., Pollard, R. B., Maeda, H. (1993). Immunomodulating activities of orally administered SMANCS, a polymer-conjugated derivative of the proteinaceous antibiotic neocarzinostatin, in an oily formulation. *International Journal of Immunopharmacology 15* (2), 175–183.

69. Pratten, M. K., Duncan, R., Cable, H. C., Schnee, R., Ringsdorf, H., Lloyd, J. B. (1981). Pinocytic uptake of divinyl ether-maleic anhydride (pyran copolymer) and its failure to stimulate pinocytosis. *Chemico-Biological Interactions 35* (3), 319–330.
70. Duncan, R., Vicent, M. J. (2010). Do HPMA copolymer conjugates have a future as clinically useful nanomedicines? A critical overview of current status and future opportunities. *Advanced Drug Delivery Reviews 62* (2), 272–282.
71. Tomalia, D. A., Baker, H., Dewald, J., Hall, M., Kallos, G., Martin, S., Roeck, J., Ryder, J., Smith, P. (1985). A new class of polymers: starburst-dendritic macromolecules. *Polymer Journal 17* (1), 117–132.
72. Gajbhiye, V., Palanirajan, V. K., Tekade, R. K., Jain, N. K. (2009). Dendrimers as therapeutic agents: a systematic review. *The Journal of Pharmacy and Pharmacology 61* (8), 989–1003.
73. Barkhausen, J., Ebert, W., Heyer, C., Debatin, J. F., Weinmann, H. J. (2003). Detection of atherosclerotic plaque with Gadofluorine-enhanced magnetic resonance imaging. *Circulation 108* (5), 605–609.
74. O'Loughlin, J., Millwood, I. Y., McDonald, H. M., Price, C. F., Kaldor, J. M., Paull, J. R. (2010). Safety, tolerability, and pharmacokinetics of SPL7013 gel (VivaGel): a dose ranging, phase I study. *Sexually Transmitted Diseases 37* (2), 100–104.
75. Bakry, R., Vallant, R. M., Najam-ul-Haq, M., Rainer, M., Szabo, Z., Huck, C. W., Bonn, G. K. (2007). Medicinal applications of fullerenes. *International Journal of Nanomedicine 2* (4), 639–649.
76. Varkouhi, A. K., Foillard, S., Lammers, T., Schiffelers, R. M., Doris, E., Hennink, W. E., Storm, G. (2011). SiRNA delivery with functionalized carbon nanotubes. *International Journal of Pharmaceutics 416* (2), 419–425.
77. Burke, A., Ding, X., Singh, R., Kraft, R. A., Levi-Polyachenko, N., Rylander, M. N., Szot, C., Buchanan, C., Whitney, J., Fisher, J., Hatcher, H. C., D'Agostino, R., Kock, N. D., Ajayan, P. M., Carroll, D. L., Akman, S., Torti, F. M., Torti, S. V. (2009). Long-term survival following a single treatment of kidney tumors with multiwalled carbon nanotubes and near-infrared radiation. *Proceedings of the National Academy of Sciences of the United States of America. 106*, 12897–12902.
78. Gomez-Gualdron, D. A., Burgos, J. C., Yu, J., Balbuena, P. B. (2011). Carbon nanotubes: engineering biomedical applications. *Progress in Molecular Biology and Translational Science 104*, 175–245.
79. Shvedova, A. A., Castranova, V., Kisin, E. R., Schwegler-Berry, D., Murray, A. R., Gandelsman, V. Z., Maynard, A., Baron, P. (2003). Exposure to carbon nanotube material: assessment of nanotube cytotoxicity using human keratinocyte cells. *Journal of Toxicology and Environmental Health. Part A 66* (20), 1909–1926.
80. Liu, D., Yi, C., Zhang, D., Zhang, J., Yang, M. (2010). Inhibition of proliferation and differentiation of mesenchymal stem cells by carboxylated carbon nanotubes. *ACS Nano 4* (4), 2185–2195.
81. Nel, A., Xia, T., Mädler, L., Li, N. (2006). Toxic potential of materials at the nanolevel. *Science 311* (5761), 622–627.
82. Chen, H. H., Yu, C., Ueng, T. H., Chen, S., Chen, B. J., Huang, K. J., Chiang, L. Y. (1998). Acute and subacute toxicity study of water-soluble polyalkylsulfonated C60 in rats. *Toxicologic Pathology 26* (1), 143–151.

83. Oberdorster, E. (2004). Manufactured nanomaterials (fullerenes, C60) induce oxidative stress in the brain of juvenile largemouth bass. *Environmental Health Perspectives 112* (10), 1058–1062.

84. Andrievsky, G., Klochkov, V., Derevyanchenko, L. (2005). Is the C60 fullerene molecule toxic?! *Fullerenes, Nanotubes and Carbon Nanostructures 13* (4), 363–376.

85. Bosi, S., Feruglio, L., Da Ros, T., Spalluto, G., Gregoretti, B., Terdoslavich, M., Decorti, G., Passamonti, S., Moro, S., Prato, M. (2004). Hemolytic effects of water-soluble fullerene derivatives. *Journal of Medicinal Chemistry 47* (27), 6711–6715.

86. Tsuchiya, T., Oguri, I., Yamakoshi, Y. N., Miyata, N. (1996). Novel harmful effects of [60]fullerene on mouse embryos *in vitro* and *in vivo*. *FEBS Letters 393* (1), 139–145.

87. Poland, C. A., Duffin, R., Kinloch, I., Maynard, A., Wallace, W. A., Seaton, A., Stone, V., Brown, S., Macnee, W., Donaldson, K. (2008). Carbon nanotubes introduced into the abdominal cavity of mice show asbestos-like pathogenicity in a pilot study. *Nature Nanotechnology 3* (7), 423–428.

88. Morimoto, Y., Hirohashi, M., Ogami, A., Oyabu, T., Myojo, T., Todoroki, M., Yamamoto, M., Hashiba, M., Mizuguchi, Y., Lee, B. W., Kuroda, E., Shimada, M., Wang, W. N., Yamamoto, K., Fujita, K., Endoh, S., Uchida, K., Kobayashi, N., Mizuno, K., Inada, M., Tao, H., Nakazato, T., Nakanishi, J., Tanaka, I. (2011). Pulmonary toxicity of well-dispersed multi-wall carbon nanotubes following inhalation and intratracheal instillation. *Nanotoxicology. 6* (6), 587–599.

89. Varga, C., Szendi, K. (2010). Carbon nanotubes induce granulomas but not mesotheliomas. *In Vivo 24* (2), 153–156.

90. Kamata, H., Tasaka, S., Inoue, K., Miyamoto, K., Nakano, Y., Shinoda, H., Kimizuka, Y., Fujiwara, H., Ishii, M., Hasegawa, N., Takamiya, R., Fujishima, S., Takano, H., Ishizaka, A. (2011). Carbon black nanoparticles enhance bleomycin-induced lung inflammatory and fibrotic changes in mice. *Experimental Biology and Medicine 236* (3), 315–324.

91. Wessels, A., Van Berlo, D., Boots, A. W., Gerloff, K., Scherbart, A. M., Cassee, F. R., Gerlofs-Nijland, M. E., Van Schooten, F. J., Albrecht, C., Schins, R. P. (2011). Oxidative stress and DNA damage responses in rat and mouse lung to inhaled carbon nanoparticles. *Nanotoxicology 5* (1), 66–78.

92. Yoshida, S., Hiyoshi, K., Oshio, S., Takano, H., Takeda, K., Ichinose, T. (2010). Effects of fetal exposure to carbon nanoparticles on reproductive function in male offspring. *Fertility and Sterility 93* (5), 1695–1699.

93. Qhobosheane, M., Santra, S., Zhang, P., Tan, W. (2001). Biochemically functionalized silica nanoparticles. *The Analyst 126* (8), 1274–1278.

94. Zhang, Q., Zhang, L., Liu, B., Lu, X., Li, J. (2007). Assembly of quantum dots-mesoporous silicate hybrid material for protein immobilization and direct electrochemistry. *Biosensors & Bioelectronics 23* (5), 695–700.

95. Wang, Y., Caruso, F. (2004). Enzyme encapsulation in nanoporous silica spheres. *Chemical Communications (Cambridge, England)* (13) 1528–1529.

96. Benezra, M., Penate-Medina, O., Zanzonico, P. B., Schaer, D., Ow, H., Burns, A., DeStanchina, E., Longo, V., Herz, E., Iyer, S., Wolchok, J., Larson, S. M., Wiesner, U., Bradbury, M. S. (2011). Multimodal silica nanoparticles are effective cancer-targeted probes in a model of human melanoma. *The Journal of Clinical Investigation 121* (7), 2768–2780.

97. Rosenholm, J. M., Peuhu, E., Bate-Eya, L. T., Eriksson, J. E., Sahlgren, C., Linden, M. (2010). Cancer-cell-specific induction of apoptosis using mesoporous silica nanoparticles as drug-delivery vectors. *Small 6* (11), 1234–1241.
98. Popat, A., Hartono, S. B., Stahr, F., Liu, J., Qiao, S. Z., Qing Max, G. Lu. (2011). Mesoporous silica nanoparticles for bioadsorption, enzyme immobilisation, and delivery carriers. *Nanoscale 3* (7), 2801–2818.
99. Liu, Q., Zhang, J., Sun, W., Xie, Q. R., Xia, W., Gu, H. (2012). Delivering hydrophilic and hydrophobic chemotherapeutics simultaneously by magnetic mesoporous silica nanoparticles to inhibit cancer cells. *International Journal of Nanomedicine 7*, 999–1013.
100. Choi, J., Burns, A. A., Williams, R. M., Zhou, Z., Flesken-Nikitin, A. Zipfel, W. R., Wiesner, U., Nikitin, A. Y. (2007). Core-shell silica nanoparticles as fluorescent labels for nanomedicine. *Journal of Biomedical Optics 12* (6), 064007.
101. Vallhov, H., Gabrielsson, S., Stromme, M., Scheynius, A., Garcia-Bennett, A. E. (2007). Mesoporous silica particles induce size dependent effects on human dendritic cells. *Nano Letters 7* (12), 3576–3582.
102. Napierska, D., Thomassen, L. C., Rabolli, V., Lison, D., Gonzalez, L., Kirsch-Volders, M. Martens, J. A., Hoet, P. H. (2009). Size-dependent cytotoxicity of monodisperse silica nanoparticles in human endothelial cells. *Small 5* (7), 846–853.
103. Hudson, S. P., Padera, R. F., Langer, R., Kohane, D. S. (2008). The biocompatibility of mesoporous silicates. *Biomaterials 29* (30), 4045–4055.
104. Yu, T., Greish, K., McGill, L. D., Ray, A., Ghandehari, H. (2012). Influence of geometry, porosity, and surface characteristics of silica nanoparticles on acute toxicity: their vasculature effect and tolerance threshold. *ACS Nano. 6* (3), 2289–2301.
105. Yu, T., Malugin, A., Ghandehari, H. (2011). Impact of silica nanoparticle design on cellular toxicity and hemolytic activity. *ACS Nano 5* (7), 5717–5728.
106. Nemmar, A., Nemery, B., Hoet, P. H., Van Rooijen, N., Hoylaerts, M. F. (2005). Silica particles enhance peripheral thrombosis: key role of lung macrophage-neutrophil cross-talk. *American Journal of Respiratory and Critical Care Medicine 171* (8), 872–879.
107. Huang, X., Li, L., Liu, T., Hao, N., Liu, H., Chen, D., Tang, F. (2011). The shape effect of mesoporous silica nanoparticles on biodistribution, clearance, and biocompatibility *in vivo*. *ACS Nano 5* (7), 5390–5399.
108. Yamashita, K., Yoshioka, Y., Higashisaka, K., Mimura, K., Morishita, Y., Nozaki, M., Yoshida, T., Ogura, T., Nabeshi, H., Nagano, K., Abe, Y., Kamada, H., Monobe, Y., Imazawa, T., Aoshima, H., Shishido, K., Kawai, Y., Mayumi, T., Tsunoda, S., Itoh, N., Yoshikawa, T., Yanagihara, I., Saito, S., Tsutsumi, Y. (2011). Silica and titanium dioxide nanoparticles cause pregnancy complications in mice. *Nature Nanotechnology 6* (5), 321–328.
109. Exley, C., Siesjo, P., Eriksson, H. (2010). The immunobiology of aluminium adjuvants: how do they really work? *Trends in Immunology 31* (3), 103–109.
110. Becaria, A., Campbell, A., Bondy, S. C. (2002). Aluminum as a toxicant. *Toxicology and Industrial Health 18* (7), 309–320.
111. Arieff, A. I., Cooper, J. D., Armstrong, D., Lazarowitz, V. C. (1979). Dementia, renal failure, and brain aluminum. *Annals of Internal Medicine 90* (5), 741–747.
112. Bondy, S. C. (2010). The neurotoxicity of environmental aluminum is still an issue. *Neurotoxicology 31* (5), 575–581.

113. Yumoto, S., Kakimi, S., Ohsaki, A., Ishikawa, A. (2009). Demonstration of aluminum in amyloid fibers in the cores of senile plaques in the brains of patients with Alzheimer's disease. *Journal of Inorganic Biochemistry 103* (11), 1579–1584.

114. Banks, W. A., Niehoff, M. L., Drago, D., Zatta, P. (2006). Aluminum complexing enhances amyloid beta protein penetration of blood-brain barrier. *Brain Research 1116* (1), 215–221.

115. Frisardi, V., Solfrizzi, V., Capurso, C., Kehoe, P. G., Imbimbo, B. P., Santamato, A., Dellegrazie, F., Seripa, D., Pilotto, A., Capurso, A., Panza, F. (2010). Aluminum in the diet and Alzheimer's disease: from current epidemiology to possible disease-modifying treatment. *Journal of Alzheimer's Disease 20* (1), 17–30.

116. Hafejee, A., Burke, M. J. (2004). Acute pneumonitis starting 2 hours after intramuscular gold administration in a patient with rheumatoid arthritis. *Annals of the Rheumatic Diseases 63* (11), 1525–1526.

117. Elsherbini, A. A., Saber, M., Aggag, M., El-Shahawy, A., Shokier, H. A. (2011). Laser and radiofrequency-induced hyperthermia treatment via gold-coated magnetic nanocomposites. *International Journal of Nanomedicine 6*, 2155–2165.

118. Reuveni, T., Motiei, M., Romman, Z., Popovtzer, A., Popovtzer, R. (2011). Targeted gold nanoparticles enable molecular CT imaging of cancer: an *in vivo* study. *International Journal of Nanomedicine 6*, 2859–2864.

119. Rink, J. S., McMahon, K. M., Chen, X., Mirkin, C. A., Thaxton, C. S., Kaufman, D. B. (2010). Transfection of pancreatic islets using polyvalent DNA-functionalized gold nanoparticles. *Surgery 148* (2), 335–345.

120. Kumari, S., Singh, R. P. (2011). Glycolic acid-*g*-chitosan-gold nanoflower nanocomposite scaffolds for drug delivery and tissue engineering. *International Journal of Biological Macromolecules*. *50* (3), 878–883.

121. Sun, Y. N., Wang, C. D., Zhang, X. M., Ren, L., Tian, X. H. (2011). Shape dependence of gold nanoparticles on *in vivo* acute toxicological effects and biodistribution. *Journal of Nanoscience and Nanotechnology 11* (2), 1210–1216.

122. Aillon, K. L., Xie, Y., El-Gendy, N., Berkland, C. J., Forrest, M. L. (2009). Effects of nanomaterial physicochemical properties on *in vivo* toxicity. *Advanced Drug Delivery Reviews 61* (6), 457–466.

123. Abdelhalim, M. A., Jarrar, B. M. (2011). Gold nanoparticles induced cloudy swelling to hydropic degeneration, cytoplasmic hyaline vacuolation, polymorphism, binucleation, karyopyknosis, karyolysis, karyorrhexis and necrosis in the liver. *Lipids in Health and Disease 10*, 166.

124. Zhang, X. D., Wu, D., Shen, X., Liu, P. X., Yang, N., Zhao, B., Zhang, H., Sun, Y. M., Zhang, L. A., Fan, F. Y. (2011). Size-dependent *in vivo* toxicity of PEG-coated gold nanoparticles. *International Journal of Nanomedicine 6*, 2071–2081.

125. Naqvi, S., Samim, M., Abdin, M., Ahmed, F. J., Maitra, A., Prashant, C., Dinda, A. K. (2010). Concentration-dependent toxicity of iron oxide nanoparticles mediated by increased oxidative stress. *International Journal of Nanomedicine 5*, 983–989.

126. Rosner, M. H., Auerbach, M. (2011). Ferumoxytol for the treatment of iron deficiency. *Expert Review of Hematology 4* (4), 399–406.

127. Voinov, M. A., Sosa Pagan, J. O., Morrison, E., Smirnova, T. I., Smirnov, A. I. (2011). Surface-mediated production of hydroxyl radicals as a mechanism of iron oxide nanoparticle biotoxicity. *Journal of the American Chemical Society 133* (1), 35–41.

128. Mahmoudi, M., Laurent, S., Shokrgozar, M. A., Hosseinkhani, M. (2011). Toxicity evaluations of superparamagnetic iron oxide nanoparticles: cell "vision" versus physicochemical properties of nanoparticles. *ACS Nano 5* (9), 7263–7276.

129. Patel, V. P., Chu, C. T. (2011). Nuclear transport, oxidative stress, and neurodegeneration. *International Journal of Clinical Experimental Pathology 4* (3), 215–229.

130. Lee, S., Park, Y., Zuidema, M. Y., Hannink, M., Zhang, C. (2011). Effects of interventions on oxidative stress and inflammation of cardiovascular diseases. *World Journal of Cardiology 3* (1), 18–24.

131. Klaunig, J. E., Kamendulis, L. M., Hocevar, B. A. (2010). Oxidative stress and oxidative damage in carcinogenesis. *Toxicologic Pathology 38* (1), 96–109.

132. Maiti, S. (2011). Nanotoxicity of gold and iron nanoparticles. *Journal of Biomedical Nanotechnology 7* (1), 65.

133. Singh, N., Jenkins, G. J., Nelson, B. C., Marquis, B. J., Maffeis, T. G., Brown, A. P., Williams, P. M., Wright, C. J., Doak, S. H. (2012). The role of iron redox state in the genotoxicity of ultrafine superparamagnetic iron oxide nanoparticles. *Biomaterials 33* (1), 163–170.

134. Apopa, P. L., Qian, Y., Shao, R., Guo, N. L., Schwegler-Berry, D., Pacurari, M., Porter, D., Shi, X., Vallyathan, V., Castranova, V., Flynn, D. C. (2009). Iron oxide nanoparticles induce human microvascular endothelial cell permeability through reactive oxygen species production and microtubule remodeling. *Particle and Fibre Toxicology 6*, 1.

135. Zhu, M. T., Feng, W. Y., Wang, B., Wang, T. C., Gu, Y. Q., Wang, M., Wang, Y., Ouyang, H., Zhao, Y. L., Chai, Z. F. (2008). Comparative study of pulmonary responses to nano- and submicron-sized ferric oxide in rats. *Toxicology 247* (2–3), 102–111.

136. Kooi, M. E., Cappendijk, V. C., Cleutjens, K. B., Kessels, A. G., Kitslaar, P. J., Borgers, M., Frederik, P. M., Daemen, M. J., van Engelshoven, J. M. (2003). Accumulation of ultrasmall superparamagnetic particles of iron oxide in human atherosclerotic plaques can be detected by *in vivo* magnetic resonance imaging. *Circulation 107* (19), 2453–2458.

137. Rahman, M. F., Wang, J., Patterson, T. A., Saini, U. T., Robinson, B. L., Newport, G. D., Murdock, R. C., Schlager, J. J., Hussain, S. M., Ali, S. F. (2009). Expression of genes related to oxidative stress in the mouse brain after exposure to silver-25 nanoparticles. *Toxicology Letters 187* (1), 15–21.

138. Hritcu, L., Stefan, M., Ursu, L., Neagu, A., Mihasan, M., Tartau, L., Melnig, V. (2011). Exposure to silver nanoparticles induces oxidative stress and memory deficits in laboratory rats. *Central European Journal of Biology 6* (4), 497–509.

139. Gao, X., Yin, S., Tang, M., Chen, J., Yang, Z., Zhang, W., Chen, L., Yang, B., Li, Z., Zha, Y., Ruan, D., Wang, M. (2011). Effects of developmental exposure to TiO2 nanoparticles on synaptic plasticity in hippocampal dentate gyrus area: an *in vivo* study in anesthetized rats. *Biological Trace Element Research 143* (3), 1616–1628.

140. Han, D., Tian, Y., Zhang, T., Ren, G., Yang, Z. (2011). Nano-zinc oxide damages spatial cognition capability via over-enhanced long-term potentiation in hippocampus of Wistar rats. *International Journal of Nanomedicine 6*, 1453–1461.

141. Hu, R., Zheng, L., Zhang, T., Gao, G., Cui, Y., Cheng, Z., Cheng, J., Hong, M., Tang, M., Hong, F. (2011). Molecular mechanism of hippocampal apoptosis of mice following exposure to titanium dioxide nanoparticles. *Journal of Hazardous Materials 191* (1–3), 32–40.

142. Zhang, D. C. S., ADME studies in animals and humans: experimental design, metabolite profiling and identification, and data presentation, In *Drug Metabolism. Drug Design and Development: Basic Concepts and Practice*, ed. Zhang, D., Zhu, M.; Humphreys, W. G., Wiley, Hoboken, NJ, 2007.

143. Kobori, T., Watanabe, J., Nakao, H. (2012). Gold nanoparticles as localization markers for direct and live imaging of particle absorption through a Caco-2 cell monolayer using dark-field microscopy. *Analytical Sciences 28* (1), 61.

144. Jeong, S. H., Kim, J. H., Yi, S. M., Lee, J. P., Kim, J. H., Sohn, K. H., Park, K. L., Kim, M. K., Son, S. W. (2010). Assessment of penetration of quantum dots through *in vitro* and *in vivo* human skin using the human skin equivalent model and the tape stripping method. *Biochemical and Biophysical Research Communications 394* (3), 612–615.

145. Iqbal, J., Vigl, C., Moser, G., Gasteiger, M., Perera, G., Bernkop-Schnurch, A. (2011). Development and *in vivo* evaluation of a new oral nanoparticulate dosage form for leuprolide based on polyacrylic acid. *Drug Delivery 18* (6), 432–440.

146. Zvyagin, A. V., Zhao, X., Gierden, A., Sanchez, W., Ross, J. A., Roberts, M. S. (2008). Imaging of zinc oxide nanoparticle penetration in human skin *in vitro* and *in vivo*. *Journal of Biomedical Optics 13* (6), 064031.

147. Castellino, S., Groseclose, M. R., Wagner, D. (2011). MALDI imaging mass spectrometry: bridging biology and chemistry in drug development. *Bioanalysis 3* (21), 2427–2441.

148. Arnida, M. M., Janat-Amsbury, A., Ray, C. M., Peterson, Ghandehari, H. (2011). Geometry and surface characteristics of gold nanoparticles influence their biodistribution and uptake by macrophages. *European Journal of Pharmaceutics and Biopharmaceutics 77* (3), 417–423.

149. Winters, D. K., Cederbaum, A. I. (1992). Expression of a catalytically active human cytochrome P-4502E1 in Escherichia coli. *Biochimica et Biophysica Acta 1156* (1), 43–49.

150. Li, A. P. (2001). Screening for human ADME/Tox drug properties in drug discovery. *Drug Discovery Today 6* (7), 357–366.

151. Tyner, K., Sadrieh, N. (2011). Considerations when submitting nanotherapeutics to FDA/CDER for regulatory review. *Methods in Molecular Biology 697*, 17–31.

152. Mortelmans, K., Zeiger, E. (2000). The Ames Salmonella/microsome mutagenicity assay. *Mutation Research 455* (1–2), 29–60.

153. Cotelle, S., Ferard, J. F. (1999). Comet assay in genetic ecotoxicology: a review. *Environmental and Molecular Mutagenesis 34* (4), 246–255.

154. Weber, E., Bidwell, K., Legator, M. S. (1975). An evaluation of the micronuclei test using triethylenemelamine, trimethylphosphate, hycanthone and niridazole. *Mutation Research 28* (1), 101–106.

155. OECD, *Test No. 486: Unscheduled DNA Synthesis (UDS) Test with Mammalian Liver Cells* in vivo: OECD Publishing.

156. Parodi, S., Malacarne, D., Romano, P., Taningher, M. (1991). Are genotoxic carcinogens more potent than nongenotoxic carcinogens? *Environmental Health Perspectives 95*, 199–204.

157. Hernandez, L. G., van Steeg, H., Luijten, M., van Benthem, J. (2009). Mechanisms of non-genotoxic carcinogens and importance of a weight of evidence approach. *Mutation Research 682* (2–3), 94–109.

158. Creton, S., Aardema, M. J., Carmichael, P. L., Harvey, J. S., Martin, F. L., Newbold, R. F., O'Donovan, M. R., Pant, K., Poth, A., Sakai, A., Sasaki, K., Scott, A. D., Schechtman, L. M., Shen, R. R., Tanaka, N., Yasaei, H. (2012). Cell transformation assays for prediction of carcinogenic potential: state of the science and future research needs. *Mutagenesis 27* (1), 93–101.

159. OECD, *Test No. 451: Carcinogenicity Studies*: OECD Publishing.

160. Wiwanitkit, V., Sereemaspun, A., Rojanathanes, R. (2009). Effect of gold nanoparticles on spermatozoa: the first world report. *Fertility and Sterility 91* (1), e7–e8.

161. Brendler-Schwaab, S., Hartmann, A., Pfuhler, S., Speit, G. (2005). The *in vivo* comet assay: use and status in genotoxicity testing. *Mutagenesis 20* (4), 245–254.

162. Chan, W. H., Shiao, N. H. (2008). Cytotoxic effect of CdSe quantum dots on mouse embryonic development. *Acta Pharmacologica Sinica 29* (2), 259–266.

163. OECD, *Test No. 415: One-Generation Reproduction Toxicity Study*: OECD Publishing.

# An Overview of the Discovery and Development Process for Biologics

HEATHER H. SHIH, PAULA MILLER, and DOUGLAS C. HARNISH

## 1.1 INTRODUCTION

Biologics, also called biotherapeutics or biopharmaceuticals, are drug substances derived from living organisms or produced using biotechnology that are composed of biological entities such as proteins, peptides, nucleic acids, or cells [1]. They differ from small molecule (SM) drugs that are chemically synthesized and have low molecular weights. Some biologics, such as antibody–drug conjugates, consist of both a protein moiety and an SM component, both of which are required for the therapeutic action of the drug. Traditional biologics that have reached the market include vaccines and blood-derived factors. The advancement in modern biotechnology has brought forth new classes of biologics as exemplified by monoclonal antibodies (mAbs), Fc fusion proteins, recombinant proteins, and peptide drugs. Some early clinical success is now seen in several novel classes of biologics, which include antibody variants, novel protein scaffolds, RNA therapeutics, and cell-based therapies [2–5]. This chapter focuses on protein-based biologics, particularly mAbs because they represent the largest class of biologic drugs. By the end of 2011, the US Food and Drug Administration (FDA) had approved close to 40 mAbs and antibody variants as summarized in Table 1.1. Details on other forms of biologics such as vaccines and RNA drugs can be found in later chapters.

The first protein-based biologic drug, recombinant insulin Humulin, was approved in the United States in 1982 [6]. Since then the field of biologics grew steadily, with the biotechnology sector laying the foundation for both the drug discovery process and technology innovation. Around late 1990s, the pharmaceutical industry started to invest more in the development of biologics. This shift from a primary focus on SM drugs was largely due to patent expiration on these drugs and the concurrent fierce competition from generic SM drugs. In addition, the increasing difficulty to bring new drugs to the market because of tightened regulations and a lack of breakthroughs in the drug discovery process has also contributed to this shift.

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*,
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

**TABLE 1.1 List of Food and Drug Administration–Approved Antibody-Based Therapeutics Up to 2011 as Categorized by Types**

| Type of Ab-Based Therapeutics | Nonproprietary/ Trade Name | Ab Target or Fc Fusion Partner | Company |
|---|---|---|---|
| Human antibodies | Ipilimumab/Yervoy | CTLA4 | Bristol-Myers Squibb |
| | Belimumab/Benlysta | B-lymphocyte stimulator | Human Genome Sciences |
| | Ustekinumab/Stelara | p40 subunit of IL-12 and IL-23 | Johnson & Johnson |
| | Canakinumab/Ilaris | IL-1β | Novartis |
| | Denosumab/Prolia/Xgeva | RANKL | Amgen |
| | Ofatumumab/Arzerra | CD20 | Genmab |
| | Golimumab/Simponi | TNF | Centocor |
| | Panitumumab/Vectibix | EGFR | Amgen |
| | Adalimumab/Humira | TNF | Abbott |
| Humanized antibodies | Tocilizumab/Actemra | IL-6R | Roche |
| | Eculizumab/Soliris | C5 | Alexion |
| | Natalizumab/Tysabri | Alpha4 integrin | Biogen/Elan |
| | Bevacizumab/Avastin | VEGFa | Genentech |
| | Efalizumab/Raptiva | CD11a | Genentech |
| | Omalizumab/Xolair | Human IgE Fc | Genentech |
| | Alemtuzumab/ Campath-IH | CD52 | Genzyme |
| | Trastuzumab/Herceptin | Her2 | Genentech |
| | Palivizumab/Synagis | RSV protein F | MedImmune |
| | Daclizumab | CD25 | Roche |
| Chimeric antibodies | Cetuximab/Erbitux | EGFR | Imclone |
| | Infliximab/Remicade | TNFα | Centocor |
| | Basiliximab/Simulect | CD25 | Novartis |
| | Rituximab/Rituxan | CD20 | IDEC |
| Murine antibody | Muromonab-CD3/Orthoclone OKT3 | CD3 | Janssen-Cilag |
| Fab fragment | Abciximab/Reopro | CD43 | Centocor |
| | Ranibizumab/Lucentis | VEGFa | Genentech |
| | Certolizumab pegol/Cimzia | TNFα | UCB |
| Antibody conjugates | Brentuximab vedotin/Adcetris | CD30 | Seattle Genetics |
| | Tositumomab-I131/Bexxar | CD20 | GlaxoSmithKline |
| | Ibritumomab tiuxetan/Zevalin | CD20 | IDEC |
| | Gemtuzumab ozogamicin/Mylotarg | CD33 | Wyeth |

**TABLE 1.1** (*Continued*)

| Type of Ab-Based Therapeutics | Nonproprietary/ Trade Name | Ab Target or Fc Fusion Partner | Company |
|---|---|---|---|
| Fc fusions | Afilibercept/Eylea | VEGFR1 and 2 ECD | Regeneron, Bayer |
| | Belatacept/Nulojix | CTLA4 ECD | Bristol-Myers Squibb |
| | Romiplostim/Nplate | Peptide thrombopoietin mimetic | Amgen |
| | Rilonacept/Arcalyst | IL-1R ECD | Regeneron |
| | Abatacept/Orencia | CTLA4 ECD | Bristol-Myers Squibb |
| | Alefacept/Amevive | LFA-3 ECD | Biogen IDEC |
| | Etanercept/Enbrel | TNFRII ECD | Wyeth/Amgen |

Presently, the number of biologics on the market has reached more than 200, and the sales of biologics in 2009 reached $93 billion, with approximately one third of current pharmaceutical pipelines consisting of biologics [7]. Given that almost all of the large pharmaceutical companies have acquired infrastructures and committed resources to develop biologics, we will continue to see a robust growth in this sector in the coming years.

Compared with SM drugs, protein-based biologics have unique therapeutic features. A therapeutic protein usually exhibits exquisite specificity when binding to and modulating its molecular target, which often translates into low off-target toxicity and clinical safety. For example, therapeutic mAbs bind to their target molecules with affinities in the picomolar to low nanomolar range (e.g., [8]). Furthermore, the interaction occurs over a broad interface with multiple physical and chemical bonds formed between an antibody and its cognate antigen, resulting in an extraordinary binding specificity that allows the differentiation of binding partners that differ by as few as one amino acid or subtle conformational difference. On the contrary, the small size of an SM drug makes it prone to off-target binding to proteins other than its intended target, which may result in unacceptable levels of toxicities. A potentially short development cycle is another advantage for the development of biologics, particularly mAbs and recombinant proteins. A clinical candidate for mAb or recombinant protein can be generated and selected in as short as 3 to 5 years compared with typically 7 to 8 years for SMs.

Protein-based biologics have their own limitations. Presently, almost all protein-based drugs must be administered as intravenous or subcutaneous injections because oral delivery is not yet a viable route of administration. Furthermore, protein drugs do not readily penetrate cell membrane and blood–brain barrier (BBB) and therefore are limited to the modulation of peripherally located extracellular targets. The cost of goods to manufacture protein drugs is significantly higher than for SM drugs, which translates into a high drug price that exacerbates health management cost issues [9].

Based on these pros and cons associated with the development of biologics, presently the pharmaceutical industry strives to achieve a balanced portfolio consisting of both SM and biologic drugs. This chapter provides an overview of the discovery and development process for protein therapeutics with a primary focus on mAbs (Fig. 1.1). Additionally, the chapter summarizes the current status of the protein-based biologics field and discusses several future trends.

## 1.2 THE DISCOVERY PROCESS FOR MONOCLONAL ANTIBODIES

Monoclonal antibodies and mAb variants such as antibody–drug conjugates and Fc fusion proteins are a major class of biologics. This section describes in detail the discovery process for mAbs. Later chapters illustrate the process for several other forms of biologics such as vaccines and RNA therapeutics.

In mammals, antibodies are proteins found in the blood that are produced by B cells from the humoral immune system in defense of foreign organisms recognized by the host system. Also known as immunoglobulins, antibodies can be categorized into five classes or isotypes, namely IgM, IgD, IgG, IgE, and IgA (reviewed by Schroeder and Cavacini [10]). IgG is the predominant class in therapeutic antibodies. An IgG molecule consists of two heavy chains and two light chain interlinked by disulfide bonds (Fig. 1.2). Each chain has an N-terminal variable region (Fv) and a C-terminal constant region. The variable region of each pair of heavy and light chains has six hypervariable loops (three per chain) called

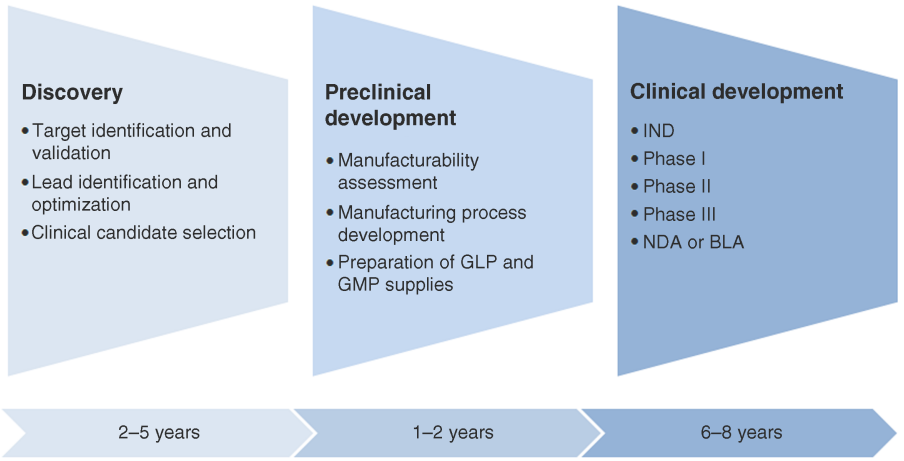


**FIGURE 1.1** Major phases and general timelines of the drug development process for biologics. BLA: biologics license application; GLP: good laboratory practice; GMP: good manufacturing practices; IND: investigational new drug; NDA: new drug application.

*complementarity determining regions* (CDRs) that form the antigen binding region at the tip of the IgG molecule. The heavy and light chains in an Fv fragment can be joined via a linker using recombinant DNA technology and produced as a single-chain Fv (scFv) molecule, which is used in phage display technology as a method to generate therapeutic antibody candidates (see later discussion). Papain digestion of an IgG yields three components: two Fab fragments and an Fc fragment. A Fab molecule can be generated via molecular engineering into a therapeutic agent. Currently, three Fab-based antibody drugs are on the market: abciximab, ranibizumab, and certolizumab pegol (Table 1.1). The Fc fragment exhibits effector functions, namely the ability to engage immune system to kill antibody opsonized molecules. In addition, the binding of Fc to bacterial protein A and protein G has been applied to the purification of therapeutic antibodies at industrial manufacturing scales. Furthermore, Fc binds to the neonatal Fc receptor (FcRn) expressed on endothelial cells. Upon uptake by endothelial cells, FcRn recycles IgG molecules back into circulation, thus conferring a long *in vivo* half-time to IgG [11]. Via molecular engineering, the Fc fragment can be fused to another protein fragment such as a soluble factor or the extracellular domain of a cell surface receptor. The resulting Fc fusion proteins represent one type of antibody-based therapeutics (Table 1.1 and Fig. 1.2).

The advancement in molecular and cellular biology has transformed the isolation, molecular engineering, and production of recombinant mAbs into an industrial drug-making process. Compared with other protein drugs, mAbs demonstrate superior properties as therapeutic molecules. They typically exhibit exquisite specificity to their molecular targets and minimal off-targeting binding. The bivalency of each antibody molecule contributes to its extraordinarily high binding strength (avidity) as the summation of the affinity from each half of the molecule. It also confers a cross-linking function that can be applied to its therapeutic function. The effector functions associated with the Fc domain allow the molecule to effectively mobilize the immune system to attack and kill tumor cells when used to treat cancers. mAbs are highly stable proteins with natural resistance to biological and chemical degradation. They tend to be amenable to expression and purification at manufacturing scales. They typically exhibit long *in vivo* half-lives, which allow infrequent administration in patients. As a result of these multifaceted advantages associated with mAbs, the pharmaceutical industry has focused on the development of mAbs as a major class of biologic drugs.

In many pharmaceutical and biotechnology companies, the development process for therapeutic mAbs is well established and analogous to that for SM drugs. It can be generalized into four stages: target selection, screening preparation, lead selection and optimization, and clinical candidate selection. The discovery phase is followed by a preclinical development process and ultimately the clinical testing of the selected mAb candidate in human subjects. This section describes relevant research activities that take place in each of the four discovery stages leading to the selection of a clinical candidate protein. The major differences in the respective processes for developing biologics and SMs are discussed at the end of this section.

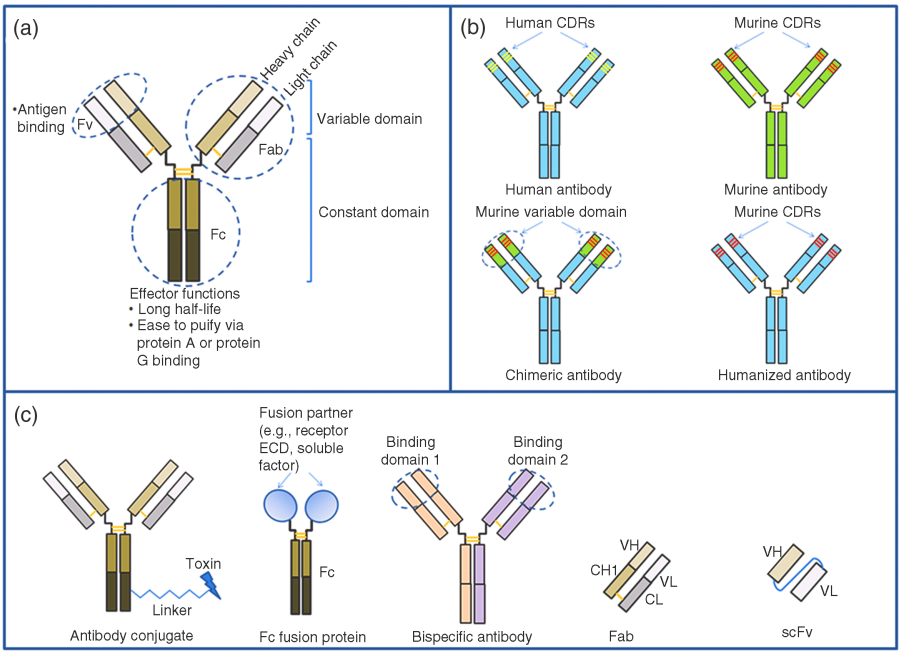


**FIGURE 1.2** Major types and properties of therapeutic antibodies and antibody variants. (a) Depiction of a whole immunoglobulin molecule. The major domains and their associated functions are indicated. *Light lines* represent disulfide bonds. Fab: antigen binding fragment; Fc: crystallizable fragment; Fv: variable fragment. (b) Major forms of therapeutic antibodies. CDR: complementarity determining region. (c) Major forms of antibody variants. All but one (single-chain variable fragment [scFv]) drug is currently approved. The scFv is used in drug discovery to identify drug candidates.

### 1.2.1 Target Selection

A target is a biological entity in patients that can be specifically and effectively modulated by a drug to ameliorate or cure a pathological condition. Selecting a drug target is usually the first step in a drug development program, although there are exceptions to the rule when drug candidate screening is carried out using a functional readout without predefined targets. In a majority of such cases, the target is identified retrospectively after the functional candidates have been selected.

A novel target is often identified either during the studies of biological pathways underlying a disease or as a result of disease target identification efforts frequently using genomic technologies such as transcriptional profiling, proteomics, and genome-wide gene association studies. Usually a target must be "validated" before the initiation of a drug discovery program. *Target validation* refers to a process of collecting clinical and experimental data to predict a beneficial therapeutic outcome from a hypothesized modulation of a selected target. A typical target validation data package often includes some or all of the following elements: (1) the target is

aberrantly expressed in the disease tissue(s) (the polymorphism of the target gene in humans has a strong association with the disease); (2) in cellular and animal models, overexpression or deficiency of the target leads to a biological outcome consistent with the expected therapeutic outcome; and (3) pharmacologic modulation of the target gives rise to expected results in cellular and animal models. A drug discovery project may begin with some evidence that the target is tied to the pathogenesis of a disease, in which case target validation becomes a continuous endeavor carried out in parallel to the generation of therapeutic candidates.

### 1.2.2 Screening Preparation

With enough confidence that therapeutic modulation of a selected target can treat a pathological condition, screening preparation can be initiated. The goal of this drug discovery stage is to generate all the reagents, functional assays, and cellular and animal models to carry out a variety of functional screens in order to select the best therapeutic candidate for human testing. This process starts with a well-thought-out screening plan that defines the primary, secondary, and tertiary screening methods encompassing *in vitro* and cellular assays and animal studies. The plan should end with the selection of a clinical candidate. Included in the plan are estimated timelines for each step, "go" and "no-go" decision points, foreseeable challenges, and mitigation plans.

Screening reagents and assays are the "bricks and mortar" to a successful selection of a therapeutic candidate. The reagents include all the essential materials needed to develop and carry out the screening assays such as cDNA and expression vector for the target, purified target protein, and cells that express the target for testing candidate antibody functions. The screening assays allow the selection of mAb candidates based on target binding and biological activities.

For drug discovery projects, animal models are used to determine the efficacy, pharmacokinetics (PK), pharmacodynamics (PD), and safety of a drug candidate. The development of relevant animal models often takes months to years and should be initiated early in the drug discovery process.

### 1.2.3 Lead Selection and Optimization

Lead selection is a process in which a rigorous screening scheme is applied in order to identify a candidate that meets the defined drug-like criteria of functional and molecular properties. The leads from this process are early drug candidates and may not have all the drug-like properties such as minimal immunogenicity profile, favorable biochemical and biophysical properties, and optimal PK properties. These properties can be subsequently optimized using a variety of technologies and ultimately developed into a viable clinical candidate.

Three major technologies have been used to develop the mAb drugs currently on the market: hybridoma, phage display, and human transgenic mouse technologies.

The hybridoma technology is a classic mAb generation technology first developed by Kohler and Milstein in 1975 [12]. Using this method, rodents are immunized

with target proteins to induce a high-titer response indicative of the generation of high-affinity rodent antibodies. The splenocytes are subsequently harvested from responding animals and fused with myeloma cells to create clonal hybridoma cells capable of secreting antibody molecules. Each hybridoma clone generates a single IgG, which is the basis for the name of mAbs. The hybridoma clones can be perpetually expanded in cell culture to provide nondepletable sources of therapeutic candidate molecules. For clinical manufacturing, however, hybridoma culture is not well suited for an industry process because of the inefficiency of these cells producing high concentrations of antibodies. Instead, DNA sequences of the IgGs from hybridoma clones are determined, and recombinant mAbs are generated in industrial quantities for clinical testing. The ultimate therapeutic recombinant mAbs have often been modified from the original rodent hybridoma mAbs using antibody engineering methods such as humanization (see later discussion).

As demonstrated by its long history, hybridoma technology is a reliable method to generate high-affinity antibodies to many protein targets and is still widely used in industry to derive early mAb therapeutic candidates. The shortcomings of this method include a relatively long cycle time from immunization to the identification of desirable mAbs (3–4 months) and difficulties to generate antibodies to proteins that are toxic to or immunotolerated by the host animal.

In contrast to the utilization of an *in vivo* system central to the hybridoma technology, phage display is a purely *in vitro* system. As first demonstrated by McCafferty et al. [13], a scFv or the Fab fragment of an IgG can be displayed on the surface of a phage particle as a fusion protein to the gene 3 phage coat protein. Built upon modern recombinant DNA technology, phage display is an elegant technology to derive antibody fragments that bind to a protein of interest. A library constituting a vast number of diverse scFv or Fab sequences up to the order of $10^{11}$ is generated. The diversity of the scFv or Fab sequences is derived from either a natural human antibody repertoire isolated from immune tissues or a synthetic repertoire where recombinant DNA technology is used to generate sequence variations [14]. During the screening process, the phage-displayed scFv or Fab is allowed to bind to the target protein and is separated from the nonbinders. The sequence of the binding scFvs or Fabs can be determined from the bacterial clones producing these "functional" phage particles. The selected scFv candidates are genetically converted back to full IgG molecules that are functionally indistinguishable from natural IgGs.

Compared with the hybridoma technology, phage display technology offers a quick way to find functional antibody fragments in a matter of weeks. In addition, the *in vitro* experimental conditions are highly amenable to manipulations and can facilitate the isolation of antibodies to difficult targets that may not be easily generated via a hybridoma approach. However, phage display has been a proprietary technology and thus is less accessible to general use than hybridoma technology. This may change given that the patent on gene 3 fusion technology has expired in 2012. To date, the majority of approved mAbs have been derived from hybridoma technology with adalimumab and recently belimumab being the only phage-derived mAb on the market. It is believed that this technology is reaching its maturity and will give rise to many more approved mAb drugs in the coming years [15].

The human transgenic mouse technology is built upon the hybridoma technology to directly generate fully human mAbs in immunized transgenic mice that express human IgGs in place of murine IgGs. The best known human transgenic mouse technologies are Medarex's HuMab-Mouse and Abgenix's XenoMouse. New strains of human transgenic mice will continue to be generated. Five therapeutic mAbs derived from this technology have gained FDA approval (Table 1.1). Despite the success of using this technology to bring forth marketed therapeutic mAbs, its proprietary nature limits general access and wide usage.

After a mAb lead is identified, it often needs to be optimized before being selected as a therapeutic candidate. Humanization has become widely used to maximize the content of the human sequence in a mAb of rodent origin to greater than 95% (reviewed by Almagro and Fransson [16]). Compared with the parental rodent mAb, the "humanized" mAb has significantly reduced "foreignness" and thus an improved safety profile in relation to the immunogenicity of the molecule (see later discussion). The process in general takes 3 to 6 months with a near 100% success rate to yield a humanized molecule with activity comparable to the parental rodent mAb.

Affinity maturation is another widely used antibody optimization method used to significantly increase the antigen binding affinity of the parental antibody [17]. Either selected CDR sequences or the entire variable domains of the parent antibody are randomized to create a library of mutant antibody molecules, from which those with improved antigen binding affinity are selected. Two widely used enabling technologies for affinity maturation are phage display and ribosomal display. An increase in binding affinity in the range of 10- to 100-fold can be achieved from this optimization process.

Fc engineering is a technology aimed at endowing optimal effector functions and half-life extension to the parental mAb. The effector function of a mAb often plays a role in its therapeutic function such as the killing of cancer cells via antibody-dependent cellular cytotoxicity (ADCC) or antibody-mediated complement-dependent cytotoxicity (CDC). In other cases such as targeting receptors on normal cells, it is beneficial to reduce or abrogate the effector function of a mAb to avoid deleterious effects resulting from potential ADCC or CDC. Mutations in the Fc that up-or down-regulate the effector functions of an IgG have been vigorously studied and presently applied to candidate molecules in clinical testing [18]. Additionally, Fc mutations that increase IgG interaction with FcRn, a receptor on endothelial cells that helps recycle a bound IgG back into circulation, have been shown to significantly improve the serum half-life of an IgG [19].

### 1.2.4 Selection of a Clinical Candidate

The selection of a clinical candidate is a process in which a lead mAb has been thoroughly evaluated for its biological activities and molecular characteristics. The biological activities refer to desirable biological functions as characterized in a combination of *in vitro* systems, cellular systems, and animal models. Based on these characterizations, a clinical candidate molecule is deemed to have met predefined efficacy and safety criteria for clinical testing in human subjects. The molecular

characteristics refer to the intrinsic behavior of a candidate molecule that must be amenable for manufacturing as well as delivering into human patients. The molecule must demonstrate acceptable levels of expression in a manufacturing production system, display favorable physiochemical characteristics, and exhibit desirable solubility in a formulation used for clinical administration. Other considerations for clinical candidate selection that need to be taken into account include commercial assessment, competitive landscape, and intellectual properties. After a molecule is selected as a clinical candidate, it triggers a number of preclinical development activities, leading to investigational new drug (IND) filing and the commencement of clinical studies. It is not unusual that a candidate selected from early stages of discovery process fails candidate selection because of various reasons associated with the criteria described.

### 1.2.5 Key Differences in the Discovery Process for Monoclonal Antibodies Versus Small Molecule Drugs

Although the general concept and overall drug discovery process from selecting a target to nominating a clinical candidate are similar for mAbs and SMs, there are some key differences. In general, the discovery cycle time for mAb drugs can be significantly shorter than that for SMs. Therapeutic mAb discovery is becoming a platform process that allows rapid identification, characterization, and production of lead candidates. Whereas it is very challenging to optimize an initial SM "hit" to turn it into a drug candidate amenable for human testing, many drug-like properties of mAbs are intrinsic to this class of molecule. This section discusses several specific areas that illustrate key differences in this process for the two different classes of drugs.

***1.2.5.1 Therapeutic Targets*** Protein therapeutics including mAbs are best suited for modulating molecular targets expressed on the surface and outside of the cells. Proteins lack intrinsic abilities to penetrate cell membranes to reach intracellular targets. SMs can access targets localized either inside or outside of a cell. In addition to the cellular location of a target, physiological barriers such as the BBB in the central nervous system (CNS) can greatly limit the passage of large molecules such as mAbs from the circulation into the brain.

The mode of target modulation can be different for mAb and SM drugs. Being structurally small, an SM drug typically modulates a target by docking onto a small pocket in a protein such as an adenosine triphosphate (ATP) binding site of a kinase and functionally altering the target molecule. It is difficult for SM drugs to disrupt protein–protein interactions that take place over large binding interfaces, a task that can be easily achieved by mAbs.

***1.2.5.2 Screening Scale*** The screening scales are significantly different for the two classes of drugs during the lead discovery phase. Whereas screening tens to hundreds of thousands of chemical compounds is a routine practice for a SM project, the screening scale is much smaller for mAbs, which is typically in the range

of thousands of test molecules. For mAbs, a build-in enrichment process before the screening step facilitates a reduction in the screening scale. For hybridoma antibody generation, the immunized rodents are an *in vivo* selection system that filters out nonspecific antibodies from the naïve antibody repertoire and enriches for high-affinity and antigen-specific antibodies. For phage antibody generation, the initial *in vitro* selection process enriches populations of phage particles that display target-binding antibody fragments. In both processes, the screening is carried out primarily to deconvolute single clone binding activities from an enriched binder pool. SM lead generation lacks this selection and enrichment process and generally relies on large-scale screening to derive candidate molecules.

***1.2.5.3 Need for Surrogate Molecule*** Surrogate antibodies are pertinent to many mAb drug discovery projects. Because of its exquisite specificity, a mAb developed to target a human protein in many cases fails to bind to its orthologous proteins in other animal species. The evaluation of a drug candidate in animals for efficacy, safety, and PK and PD, however, is important, and these are required preclinical studies. Therefore, in many cases, in parallel to the generation of the therapeutic candidate, a surrogate antibody is also created that is functionally equivalent to the therapeutic candidate and binds the target ortholog expressed in the appropriate animal species used for drug discovery models, most often mouse. It is noteworthy that SMs are not entirely excluded from the issue of lacking species cross-reactivity. However, the surrogate approach is not a routine practice during the discovery process of an SM drug.

***1.2.5.4 Drug Properties*** Lipinski's rule of five is a general set of rules to evaluate drug-like molecular properties of a SM compound [20]. Similarly, there are emerging rules to describe the drug-like physicochemical properties for protein drug candidates. Given their distinctive molecular compositions, SMs and mAbs each have uniquely defined drug-like properties. For SMs, these properties are associated with the presence of H-bond donors and acceptors, molecular weight, and partition coefficient; for mAbs, they are based on expression levels, aggregation propensity, stability, solubility, and posttranslational modifications.

The two classes of drugs also bear different intrinsic toxicity risks. A SM is prone to bind to unintended molecular targets in humans and elicit off-target toxicity, which may not be manifested in an *in vitro* screen or animal toxicity models. On the contrary, mAbs in general exhibit low off-target toxicity because of highly specific binding to their molecular targets.

Immunogenicity is a unique safety concern for protein therapeutics, including mAbs. The immune system of a patient who receives a protein drug can recognize it as a foreign entity and react with an antidrug response. The consequence of such immunogenicity response ranges from negligible or mild effects to severe anaphylactic response [21]. Presently, there are few methods at the preclinical stages to accurately predict the immunogenicity of a clinical protein drug candidate, albeit that a number of practices are used to reduce the likelihood of immunogenicity, which include the prediction and elimination of T-cell epitopes in the mAb molecule and the

utilization of *ex vivo* assays to monitor the potential of a mAb to activate T cells [22]. A vigorous determination of antidrug antibody response is mandated to monitor the immunogenicity of a protein drug during clinical development.

## 1.3 MANUFACTURING PROCESS DEVELOPMENT FOR BIOLOGICS

### 1.3.1 Introduction

After a candidate with appropriate efficacy and safety has been identified, the research focus shifts to process development and manufacture of clinical trial materials. A robust manufacturing process is critical to ensure that appropriate quality is reproducibly achieved in the packaged clinical supplies. Compared with SMs, biotherapeutics are large, complex molecules that fold to incorporate specific tertiary structure and are subject to a variety of posttranslational modifications. As such, a biotherapeutic drug is a complex mixture of species. Furthermore, the specific composition of the mixture is profoundly affected by the manufacturing process as well as handling and storage conditions. Therefore, to reproducibly control the specific composition of the mixture and the quality of the resulting clinical supplies, it is necessary to understand the critical processing parameters during manufacture and handling and storage conditions and how they ultimately impact product quality.

### 1.3.2 Early Assessment of Discovery Candidate(s)

During the transition from the discovery research phase to the clinical development phase, a candidate or several candidates are typically assessed for suitability for manufacturing process development. mAbs are a special class of biologics that share common structural features that make them well suited to platform production and purification. Platform production involves defining a set of conditions for production that are suitable for most antibodies, albeit the platform may not be optimal for any given candidate. Antibodies are typically assessed for fit in a defined platform for process development and production. This usually involves an assessment of expression in mammalian cell lines or titer based on data available from the discovery process, which may include data from transient expression or pools derived from stable transfection into a Chinese hamster ovary (CHO) host cell line, for example. In transient HEK-293 systems, titers (i.e., expression levels) of less than 50 mg/l may present a potential concern and would likely require close monitoring during development to ensure acceptable expression titers are achieved in mammalian cell lines resulting from stable transfection.

Evaluation of a candidate's propensity to aggregate and to undergo degradation in a preferred formulation or set of formulations is an important part of the early assessment process. Aggregation can occur during all phases of production, and controlling the levels of aggregate in the final product can be challenging. In addition to aggregation, significant degradation pathways such as oxidation,

deamidation, isomerization, and peptide bond cleavage are also evaluated early, typically at multiple temperatures. Often accelerated stability studies are carried out under more extreme conditions to understand the major degradation pathways for a specific candidate or set of candidates. It is important to recognize that because different degradation pathways may be accelerated at different rates, these studies need to be analyzed carefully and may not represent the distribution or even the specific composition of the various impurities under standard conditions [23,24].

The early assessment of candidates is largely intended to identify those that may present significant challenges during development. If multiple candidates are being considered for development, the selection can be based on a panel of data including, but not limited to, efficacy, tolerability, and stability. Early formulation studies can help to inform the selection decision. If a candidate shows particularly poor stability during the early assessment, it can be a significant or determining factor in candidate selection. It is also important to understand that although an early assessment can highlight potential challenges for a specific candidate, it does not replace subsequent development work; rather, it is the foundation for the preclinical development that follows.

### 1.3.3 Bioprocess Development and Manufacture

After a development candidate has been identified, cell line development begins with transfection of a suitable expression vector into a host cell line. Mammalian expression systems predominate for the large-scale production of antibodies because they can perform complex posttranslational modifications that are important for correct protein folding, stability, multimer formation, and secretion into the medium [25]. CHO cell lines are some of the more commonly used host cell lines for antibody production and use selectable markers based on dihydroxy folate reductase (dhfr) and glutamine synthetase (GS) that are available in suitable stable expression vectors for these cell lines [26,27]. Cell line stability is another key consideration during development. Typically, a good production cell line will be stable for 70 to 100 generations. The primary goal is to create a stable cell line with the appropriate growth properties and a high specific productivity ($Q_p$). The selection process is carried out over progressively increasing scales from microtiter plates to small bioreactors. During the selection process, the material that is produced is used to develop or confirm a suitable formulation and downstream processing methods and to evaluate analytical platform methods. The availability of this material and material produced during subsequent cell culture process development for parallel analytical and pharmaceutics development is critical for an efficient and integrated strategy for production of clinical trial material. After a clone is selected, a master cell bank (MCB) is prepared. The MCB, or a working cell bank (WCB) prepared from the MCB, is used in the scale-up and production of material for enabling regulatory toxicology studies and clinical trials. It is worth noting that the MCB is prepared under good manufacturing practices (GMP) conditions and is used to generate all the antibody supplies for the lifetime of the product if the candidate is successful. Thus, the preparation of an MCB represents the earliest component of the commercial process.

For mAbs, platform cell culture process development presents a number of challenges, including the need to adequately control for protein misfolding, aggregation, oxidation, deamidation, proteolysis, and glycosylation variants. Each of these product-related impurities should be monitored and controlled under the platform conditions. In some cases, the platform process may need to be modified to address candidate specific issues.

Scaling the process involves thawing vials from an MCB or WCB. This initial inoculum is expanded in shake flasks and small bioreactors and is then transferred to progressively larger seed reactors before transfer of the cell mass to the production bioreactor. There are two growth phases during fermentation. During the initial growth phase, the primary objective is to increase the viable cell mass; during the protein production phase, cell growth slows, and antibody expression and secretion ensue. A typical mAb cycle time in the production bioreactor is 10 to 12 days.

The common structural features of antibodies make them well suited to platform purification. Most, if not all, large-scale purification schemes incorporate protein A or MabSelect chromatography as the first downstream purification step. The protein A–immobilized ligand binds to the Fc region of antibodies while host cell proteins (HCPs), DNA, and other process-related impurities flow through in the mobile phase. Typically, protein A chromatography affords product in greater than 90% purity after elution from the column at low pH [28]. Ion exchange chromatography is commonly used to further purify the antibody product from HCPs, DNA, and other process- and product-related impurities. Additional filters or chromatography steps can be used to address candidate-specific issues.

Because antibodies are produced in mammalian cell culture, which can harbor viruses harmful to humans, two orthogonal methods to remove viral particles are typically incorporated in the downstream process in addition to the chromatography steps. These generally include a low pH viral inactivation step followed by viral filtration.

For first in human (FIH) clinical trials, it is generally acceptable to demonstrate clearance of two representative viruses, assayed once using new resin; however, because subsequent trials in Europe would fall under the 2008 European Medicines Agency (EMA) guidelines and because viral clearance studies are expensive and time consuming, more and more companies are choosing to comply with the more recent EMA guidelines in US trials even at early stages of development. Viral clearance validation for phase 3 and biologic license application (BLA) filing is much more comprehensive, including, but not limited to, demonstrating clearance for a minimum panel of four representative viruses, assayed in duplicate using new and recycled resins [29].

The last downstream step is commonly ultrafiltration/diafiltration (UF/DF) to reduce the volume, increase the protein concentration to an appropriate level for formulation, and incorporate selected stabilizing excipients to facilitate storage and preparation of the fully formulated drug product. Production of clinical-grade material involves execution of multistep processes, and often multiple facilities are required to support manufacture of the drug substance, the drug product, and the required analytical testing. To simplify logistics, it is often desirable to develop and

define hold conditions for downstream process intermediates and storage conditions for the drug substance. The downstream process for production of drug substance is designed to eliminate or reduce impurities and degradation products within acceptable limits and to minimize formation during processing to afford a drug substance with the appropriate quality.

### 1.3.4 Formulation Process Development and Manufacture

All proteins are subject to various degradation pathways. Degradation can occur during the process to prepare the drug substance or drug product and during their storage in the liquid, frozen, or solid state. An understanding of the various degradation pathways is critical not only to minimize and remove impurities during production but also to maintain the appropriate quality attributes of the final drug product during and after production and during shipping and storage.

Degradation of proteins can be generally categorized as either physical or chemical. Physical degradation includes surface adsorption, denaturation, and aggregation. Examples of chemical degradation include deamidation, oxidation, isomerization, fragmentation, and cross-linking.

As the demand for subcutaneous and intramuscular injection routes of administration to support patient focused delivery approaches increases, the ability to formulate proteins at increasingly higher concentrations has become a major focus in the pharmaceutical industry. These high concentrations present unique challenges, including the possibility for aggregation, precipitation, and high-viscosity solutions that can be difficult to administer to patients.

Although antibodies share a common framework, many of their degradation pathways are associated with specific primary sequences or the unique antigen-binding regions (i.e., CDRs). Additionally, the rates of various degradation pathways for a single antibody differ under a specific set of conditions, including pH, temperature, and concentration, as well as processing and handling conditions. Therefore, determining the optimal formulation and storage conditions can be very candidate specific. Most platform strategies strive to define a formulation or set of formulations that are suitable for the majority of candidates but not optimized for any specific candidate. This approach has been quite successful for antibodies but does not eliminate the need to understand the degradation pathways for a specific candidate and document that the selected formulation(s) are suitable for that antibody. This is particularly important for antibody drug candidates that ultimately become successful commercial products because the dose and dosage form generally evolve as development progresses from FIH studies to regulatory approval and launch.

### 1.3.5 Analytical Method Development and Manufacture

In the previous sections, a common theme is the importance of understanding how the production process, formulation, handling, and storage conditions affect the critical quality attributes of biological drug candidates. Key to this understanding is the ability to monitor and measure the process- and product-related impurities

observed during development and production and relate them to the efficacy and stability of a given candidate. A variety of methods can be used to identify, monitor, and quantify the various process- and product-related impurities.

Of the various process-related impurities, HCPs, DNA, and endotoxins are of primary concern because they can have negative side effects on patients. Endotoxins, or pyrogenic lipopolysaccharides (LPS), are derived from Gram-negative bacteria such as *Escherichia coli*. In mammalian production hosts, endotoxins can be introduced into a process via raw materials [30,31]. DNA can be extracted and quantified by standard techniques, including quantitative polymerase chain reaction (qPCR). Additionally, a number of host cell–specific enzyme-linked immunosorbent assay– (ELISA-) based assays to quantify HCP and DNA are also available [32].

Product-related impurities are equally important to monitor, and a variety of analytical methods have been used to quantify the physical and chemical degradation pathways of proteins. Aggregation is one of the more significant types of physical degradation. A variety of methods for detection and quantification can be used, depending on the size and type of aggregates. Size exclusion chromatography–high-performance liquid chromatography (SEC-HPLC) supplemented with appropriate orthogonal methods is a commonly used analytical strategy. Analytical methods to sufficiently characterize glycosylation patterns are extremely important for antibodies. Although heterogeneity in glycosylation patterns is not unusual, proper glycosylation and distribution of variants may be critical for proper antibody function [24]. Oxidation, deamidation, and fragmentation are some of the more common chemical degradation pathways observed for antibodies. Fragmentation can be detected using methods that separate intact and fragmented molecules based on size. Deamidation is the primary source of charge heterogeneity, and initial identification may involve detection of differences in charge distribution by isoelectric focusing (IEF) or high-performance cation exchange chromatography [33]. Similar methods can be used to assess oxidative degradation.

The methods described can be used for most protein therapeutics, and many are amenable to platform development for antibodies. Platform analytical methods are designed to detect and quantify the expected degradation pathways for antibodies, although the impurity profile for any specific antibody tends to be unique. Ultimately, however, for each candidate, it is necessary to develop candidate-specific identity and bioassays. Often simple binding ELISA-based (or BIACore) bioassays are initially developed for early clinical development. It is necessary, however, to develop a cell-based or functional bioassay for later stage clinical trials.

### 1.3.6 Project Management and Critical Path to the Clinic

Production of clinical supplies requires close collaboration among multiple scientific lines. Often multiple teams, each with a different focus, are working in an integrated and iterative way during research, process development, and manufacture with multiple hand-offs between these teams. More often than not, research, process development, and manufacture occur over multiple locations and long periods of time, during which personnel move in and out of various roles on the teams.

Furthermore, outsourcing some or all of the process development or manufacturing to contract manufacturing organizations (CMOs) is becoming increasingly common, which brings with it additional complexity. The outsourced work typically requires a technology transfer and some level of oversight personnel to facilitate and act as technical consultants. The lead times to identify a CMO, conduct a quality audit, negotiate a contract, and schedule a slot for the work can cause delays if appropriate planning is not in place. Moreover, managing the contract and the payment schedule can require additional nontechnical oversight. Although the pharmaceutical industry's focus on speed to the clinic in recent years has fueled the implementation of platform process development for antibodies, the technical and business complexities have necessitated careful planning and project management to facilitate and streamline platform antibody production across scientific lines and to ensure there is an appropriate balance between speed, cost, and quality.

In an accelerated development strategy, the availability of representative protein generally defines the critical path to the clinic: first the material from cell culture development to enable the parallel development of the downstream purification process, analytical methods, and a suitable formulation; then the good laboratory practice (GLP) material for a regulatory toxicology study; and finally, packaged GMP supplies with supporting stability studies to enable clinical dosing. Because GMP supplies can be prepared while the regulatory toxicology is underway, from a production perspective, it is the process development and manufacture of GLP material for regulatory toxicology that is on the critical path to the clinic.

## 1.4 REGULATORY REVIEW AND APPROVAL FOR BIOLOGICS

### 1.4.1 US Drug Law

The basis of US drug law is the Federal Food, Drug, and Cosmetic Act (FD&C Act) of 1938. Its passage by the US Congress required new drugs (prescription and over-the-counter drugs, medical devices, foods, and cosmetics) to be shown safe and effective before marketing and ushered in the modern area of drug development. In 1944, the Public Health Service Act (PHS Act) was subsequently passed to provide the pathway to regulate the licensure of biologic products such as vaccines, blood products, allergenic products, and other products such as mAbs. The significance of this is that for marketing of a new product, a New Drug Application (NDA) is used for SM drugs subject to the drug approval provisions of the FD&C Act, and a biologics license application (BLA) is required for biological products subject to licensure under the PHS Act. The Prescription Drug User Fee Act (PDUFA) required drug and biologics manufacturers to pay fees for product applications and supplements, and other services.

***1.4.1.1 Food and Drug Administration Modernization Act of 1997*** The Food and Drug Administration Modernization Act (FDAMA) reauthorized PDUFA and, at the time, mandated the most wide-ranging reforms in FDA practices since 1938.

Provisions include measures to regulate advertising of unapproved uses of approved drugs, to increase patient access to experimental drugs, and to accelerate review of important new medications. In addition, the law provided for an expanded database on clinical trials accessible by patients.

***1.4.1.2 Best Pharmaceuticals for Children Act of 2002*** The Best Pharmaceuticals for Children Act (BPCA) continued the exclusivity provisions for pediatric drugs as mandated under the FDAMA, in which market exclusivity of a drug is extended by 6 months, and in exchange, the manufacturer carries out studies of the effects of drugs when taken by children.

***1.4.1.3 Pediatric Research Equity Act of 2003*** The FDA is given clear authority under the Pediatric Research Equity Act (PREA) to require sponsors to conduct clinical research into pediatric applications for new drugs and biological products.

***1.4.1.4 Food and Drug Administration Amendments Act (FDAAA) of 2007*** Among other things, the FDAAA reauthorized PDUFA, PREA, and the BPCA. Other major topics covered within this wide-ranging legislation are new FDA authorities for requiring label changes and postmarketing studies, new and expanded requirements for registration of clinical trials and disclosure of results, formation of the Reagan-Udall Institute in support of the FDA's Critical Path initiative, new conflict of interest rules for the FDA Advisory Committee members, and new rules on citizen petitions.

Another landmark act was the Federal Register Act of 1935, which established the basic legal structure of the US regulatory system. The Administrative Procedure Act of 1946 added procedural requirements to ensure due process and public participation. As a result of the Federal Register Act, the daily *Federal Register* provides a single, uniform publication for executive agency rules, notices, and presidential documents, thus providing an official notice of a document's existence, its contents, and its legal effect. From this came the Code of Federal Regulations (CFR). It is a codification of rules published in the *Federal Register* by the executive departments and agencies of the federal government.

### 1.4.2 Food and Drug Administration Overview

The FDA is a consumer protection agency and falls within the executive branch of the US government under the Department of Health and Human Services (DHHS). The FDA has a broad range of responsibilities, including regulation of drugs, biologics, medical devices, animal drugs, food, and cosmetics. The FDA's major activities are: (1) reviewing new products; (2) ensuring safe drug manufacturing and handling; (3) monitoring for new risks, standards, and regulations (4) research; and (5) enforcing actions to protect public health. There are five FDA centers with the Center for Drug Evaluation and Research (CDER) and Center for Biologics Evaluation and Research (CBER) responsible for providing regulatory guidance for new drug products while ensuring that marketed products are safe and effective.

The CBER reviews products according to type (e.g., vaccines), and the CDER reviews products by area of medicine (e.g., oncology). The CDER does review certain biologic agents, proteins intended for therapeutic use that are extracted from animals or microorganisms. These include most of the recombinant proteins such as mAbs, cytokines, and growth factors as well as the more traditional pharmaceuticals. The other biologic products, including blood and blood components, cellular and gene therapy products, vaccines, antitoxins, allergenic extracts, and venoms, are reviewed by the CBER.

### 1.4.3 Drug Development in the United States

The major phases of the drug development process in the United States are discovery or preclinical testing, phase 1 FIH studies (typically a single ascending dose and multiple ascending dose study), a phase 2 proof of mechanism study, and phase 3 pivotal registration studies in the intended patient population followed by market approval with potential phase IV postmarketing commitments. In the United States, it takes anywhere from 10 to 15 years for an experimental drug to move from the laboratory to the patient. Approximately only one in 5000 synthetic SM pharmaceuticals proceed from discovery to approval.

To get approval in the United States to begin clinical testing, the sponsor must submit an IND application to the FDA. The contents of the IND provide the data and rationale to support proceeding with human clinical testing of the drug. The IND application must contain information in three broad areas. The first is animal pharmacology and toxicology studies that permit an assessment as to whether the product is reasonably safe for initial testing in humans. The second is manufacturing information pertaining to the composition, manufacture, stability, and controls used for manufacturing the drug substance and the drug product. This information is assessed to ensure that the company can adequately produce and supply consistent batches of the drug. Finally, the clinical development plan is required with detailed protocols for the proposed clinical studies to allow an assessment of whether the initial-phase trials will expose subjects to unnecessary risks. Also, information should be provided on the qualifications of clinical investigators as well as commitments to obtain informed consent from the research subjects, to obtain review of the study by an institutional review board (IRB) and to adhere to the IND regulations.

In general, an NDA or BLA is filed at the completion of all three clinical trial phases (with the exception of accelerated approval). It contains all of the scientific information (studies in animals and humans) and addresses the safety and efficacy of the drug. The goals of a NDA or BLA are to provide sufficient information to permit the FDA reviewer to reach the following key decisions: (1) that the drug is safe and effective in its proposed use(s); (2) the benefits of the drug outweigh the risks; (3) the proposed labeling (package insert) is appropriate for what it should contain; and (4) methods used in manufacturing the drug and the controls used to maintain the drug's quality are adequate to preserve the drug's identity, strength, quality, and purity.

After the FDA receives a NDA or BLA, the review period is defined by the PDUFA, and the FDA has 60 days to accept or refuse to file. The FDA begins the review by assembling a review team. Based on requirements or to gain expert input, the FDA may also seek the advice of an advisory committee. Its role is to provide independent advice that contributes to the quality of the agency's regulatory decision making and lends credibility to the product review process. The ultimate decision to approve is based on whether the drug's benefits outweigh the potential risks. After a recommendation (e.g., approved or complete response) on the application is reached by the reviewers and their supervisors, the director of the applicable Drug Review Division or office evaluates the decision. After the product is approved, the product can be legally marketed in the United States starting on the date of approval.

### 1.4.4 US Generic and Biosimilar Legislation

After a product is marketed, it has a finite time of exclusivity before a generic competitor can enter the marketplace. This time frame is dictated by the type and quality of approved patents associated with the innovator product. This patent protection provides the innovator drug maker time to recoup its development costs associated with bringing the therapy to market and thus also provides incentive for continual new drug development. The current system of generic drug approval was established by the Drug Price Competition and Patent Term Restoration Act of 1984 (also known as the Hatch-Waxman Act) [34]. This act amended the FD&C Act and provided two abbreviated pathways for the approval of generic drugs, including natural source products and recombinant proteins.

Section 505(b)(2) of the Hatch-Waxman Act eliminated the necessity of the generic competitor to duplicate all of the preclinical and clinical studies performed by the innovator for drug approval. This allowed the direct comparison of the generic drug with the innovator product already approved for the same indication while using abbreviated clinical trials of 3 to 6 months in duration to demonstrate comparability. The second abbreviated pathway as outlined in Section 505( j) sets forth the process by which the producer of a generic drug that is identical to a previously approved innovator product can file an Abbreviated New Drug Application (ANDA) to seek FDA approval. An ANDA allows the applicant to rely on the FDA's previous finding of safety and efficacy for the already approved drug; the FDA can request only bioavailability studies as supportive data. These allowances for an ANDA application and ultimate approval were tied to a couple of assumptions. The first assumption was that the generic drugs would be exactly the same as the innovator drug. The FDA instituted a policy that the bioavailability of the generic drug needs to be within $\pm 20\%$ of active ingredient over a period of time compared with the innovator product. For drugs with a very narrow therapeutic range, this may be problematic. The second assumption was that bioequivalence data were accurate surrogates to an innovator drug's safety and efficacy results.

The Hatch-Waxman Act does provide the innovator a period of exclusivity, so-called "data exclusivity," of up to 5 years upon approval before a generic version can enter the market. However, the act is also intended to provide a mechanism for

generic competitors to challenge the innovator's patents and provide the first in time challenger with 180 days of generic market exclusivity. To file an ANDA, the competitor must make one of four certifications: (1) the drug has not been patented, (2) the patent has expired, (3) the generic will not go on the market until the patent expires, or (4) the generic will not infringe on the patent or the patent is invalid.

In March 2010, as a part of the Patient Protection and Affordable Care Act (PPACA), or health care reform, Congress passed the Biologics Price Competition and Innovation Act (BPCI Act) of 2009, which amends Section 351 of the Public Health Service Act, creating a statutory pathway for FDA approval of follow-on biologics as "biosimilar" to or "interchangeable" with the innovator marketed biologic products [35]. The biosimilar product must have the same presumed mechanism of action, route of administration, dosage form, and potency as the innovator product. It may only be reviewed and approved for indications for which the FDA already has approved the innovator product.

The necessity for the BPCI Act is because although Section 505(b)(2) of the FD&C Act allows for the approval of generic drugs involving natural source products and recombinant proteins, the complexity of the large biopharmaceuticals makes it difficult to demonstrate that a biosimilar product is structurally identical to an already approved biopharmaceutical. Unlike SMs, for which analytical methods are adequate to ensure the sameness of a generic product to the innovator product, the size and complexity of the molecular structure of biologics preclude any exact structural comparison between the innovator and the biosimilar product. Moreover, the manufacture of recombinant proteins in living cells can result in subtle differences with respect to glycosylation and other posttranslational modifications. Even nonrecombinant proteins that are purified from their natural sources can exhibit different posttranslational modifications and have alterations in their amino acid sequences. Because the correlation between the efficacy and safety of a biopharmaceutical and its structural characteristics is rarely established, it is uncertain how these slight modifications may impact its effectiveness. In light of these factors, a follow-on biologic cannot be expected to replicate the innovator product precisely; therefore, these molecules are referred to as *biosimilars*. A biosimilar product is defined as one that is "highly similar" to the reference product notwithstanding minor differences in clinically inactive components and for which there are no clinically meaningful differences between the biological product and the reference product in terms of safety, purity, and potency of the product [36].

Under the BPCI Act, biosimilars also have the opportunity to meet a higher standard of similarity to a reference product, "interchangeability," reflecting an FDA assessment that pharmacists can make substitutions between biologics without the prescriber's intervention. A biologic will be considered interchangeable with a reference product if the developer demonstrates that it can be expected to produce the same clinical result in any given patient and that the risk associated with alternating or switching between the two products is not greater than that involved in continuing to use the reference product.

In response to the BPCI Act, the FDA has recently issued draft guidelines to define the approval process and requirements for biosimilars [37–39]. In these draft

guidelines, the FDA is advocating a stepwise approach to demonstrate biosimilarity. The first step relies on extensive comparative analytical data to the reference product to set the stage for the clinical development requirements. The more robust the analytical data demonstrating comparability to the reference product, the potential for a more targeted, streamlined clinical development program. The next step would involve consideration as to the toxicity studies that should be conducted to support biosimilarity. To avoid large head-to-head phase 3 clinical trials, the FDA is putting an emphasis on immunogenicity and PK and PD studies to mitigate this "residual uncertainty" of the biosimilarity of the product. At a minimum, the FDA anticipates that at least one clinical study to assess immunogenicity will likely be required. In addition, a clinical trial will also be required for a biosimilar to be considered interchangeable with the reference product. In addition, postmarketing surveillance may also be a key point of emphasis for any biosimilars application.

## 1.5 BIOLOGICS: THE PAST, THE PRESENT, AND THE FUTURE

With the ongoing shift in the pharmaceutical industry from an SM-centric to a balanced portfolio with equal shares of SMs and biologics, the development of biological drugs has gained a great deal of momentum as reflected by a large number of biologics in preclinical and clinical pipelines. Most major pharmaceutical companies have now implemented sophisticated processes and infrastructures to develop biologics, particularly mAbs and other protein therapeutics.

Cutting-edge technologies have significantly impacted the field of biologics discovery and development. At the molecular level, considerable protein engineering efforts have led to the development of novel molecular scaffolds, optimized molecular functions, and enhanced drug properties. At the process level, novel technologies have resulted in an increase in screening throughput, innovative methods to generate antibody candidates, reduced cycle time of early drug discovery process, and new capabilities to tackle challenging issues.

This section discusses several major trends in the discovery and development of biologics with a focus on mAb therapeutics.

### 1.5.1 Biosimilars

By McCamish and Woollett's definition, a biosimilar is a "follow-on biologic that meets extremely high standards for comparability or similarity to the originator biologic drug that is approved for use in the same indication" [40]. Having suffered from severe loss of profits immediately after a brand-name drug comes off patent, the industry has taken a drastically different approach with generic biologics, or biosimilars. Large pharmaceutical manufacturers in addition to generic drug makers have both aggressively entered the play field. Led by Merck, pharmaceutical giants Eli Lilly, Pfizer, and several others have publicly announced their endeavor in the making of biosimilars and are actively developing pipeline portfolios. Since the approval of the first biosimilar omnitrope (recombinant human growth hormone;

Sandoz) in 2006 in Europe, a number of biosimilars have now been approved by EMEA, including two biosimilar somatropins, five biosimilar rHuEPOs (recombinant human erythropoietins), and four biosimilar filgrastims [41].

The regulatory path for approving biosimilars is emerging. The EMA has led the way and issued a series of guidelines to regulate the approval of biosimilars in Europe. In the United States, FDA has started to define the approval processes and requirements (see earlier discussion). In contrast to the approval of generic SM drugs, the regulations for biosimilars are considerably more sophisticated because of the complex nature of biologics. In addition to the demonstration of composition equivalency, for a biosimilar, the proof for biological equivalency is required. As a result, the cost for developing biosimilars is considerably higher than that for generic SMs, and their market share relative to the brand name counterpart could fall in the 30% range as opposed to 70 to 90% for generic SMs (http://www.gabionline.net/Biosimilars/Research/How-profitable-will-biosimilars-be). In the coming years, we will learn much more about the regulations for biosimilar approval, the safety profiles of marketed biosimilars, the market shares relative to their brand-name counterparts, and the competitive landscape composed of both large pharmaceutical and generic drug makers.

### 1.5.2 Novel Scaffolds

One actively explored area in protein therapeutic discovery is the development of novel scaffolds that demonstrate unique and potentially superior features as compared with the established scaffolds best represented by mAbs. Novel molecular scaffolds from fibronectin-based scaffold to camelid nanobodies have been incepted first in academia more than a decade ago [42,43]. A myriad of biotechnology companies have devoted their efforts to transforming novel scaffolds into drug discovery platforms. A key challenge is to endow these novel scaffolds with drug-like attributes comparable to the well-established mAb-based scaffold. Although molecules derived from many novel scaffolds exhibit extraordinarily high target binding affinity and specificity, it is not an easy task for these molecules to achieve long *in vivo* half-lives, low immunogenicity, and good manufacturability, which are the drug-like features associated with mAbs that make them such a successful class of therapeutics. Despite these challenges, the biologics field has not been deterred to invest in the development of novel scaffolds [44]. Over the next 5 to 10 years the industry may see the maturation of some of these novel scaffold-based drugs coming onto the market. It is likely that mAbs will remain a formidable mainstream scaffold with novel scaffold-based proteins fulfilling niche applications.

Antibody–drug conjugates (ADCs) are becoming a promising new class of antibody variants for cancer treatment. An ADC is a therapeutic molecule consisting of an antibody conjugated to a toxin. The antibody functions as a targeting vehicle via its specific binding to a tumor cell surface antigen and brings the toxin to the targeted cancer cell. After the ADC is internalized by the targeted cancer cell, the toxin becomes dissociated from the antibody and subsequently leads to cell killing [45]. With the recent FDA approval of the first ADC brentuximab vedotin in 2011 and

promising data on Genentech's T-DM1 in late-stage development, there is a lot of excitement in the field to pursue this class of antibody-based therapeutics to effectively kill both solid tumor and hematologic cancer cells [46,47].

Although presently still in development, combinatorial biologics therapy may become an important treatment strategy in the future. Human diseases are generally complex in nature, and intervention of more than one molecular target and pathway is often required for successful treatment. For example, most cancers result from mutations in a large number of genes that cause dysregulation in both growth-inhibitory as well as growth-promoting pathways. In the clinic, it is already a common practice to treat patients with combinatorial therapies, and many cancer trials testing a novel biologic drug are done in combination with anticancer SM drugs. For many inflammatory diseases such as rheumatoid arthritis and inflammatory bowel disease, inhibiting one cytokine may not put a potent brake on a raging disease triggered by a multicytokine storm and simultaneous inhibition of different cytokines and their receptors may be required to manifest clinical efficacy.

Currently, there are two main approaches for combinatorial protein therapies. A multifunctional antibody (often referred to as bispecific antibody) combines two antigen-binding domains in a single molecule capable of simultaneously binding to and modulating the functions of two different molecular targets. Currently, a "trifunctional" bispecific antibody, catumaxomab, is a drug marketed in Europe for the treatment of malignant ascites. The "trifunction" refers to the simultaneous binding to EpCAM expressed on tumor cells, CD3 expressed on T cells, and Fc binding to FcγR on accessory immune cells [48]. The challenge for developing a multifunctional antibody arises from the difficulty to manufacture such an extensively engineered molecule, which often exhibits issues such as low expression levels, insolubility at high concentration, and high propensity to aggregation.

The second approach for combinatorial protein therapy is to apply a cocktail of two or more mAbs. Symphogen is one of the pioneers in this approach and has several therapeutic antibody combinations in the pipeline. The most complex one, rozrolimupab for the treatment of idiopathic thrombocytopenic purpura, consists of 25 IgG1 antibodies in one mixture [49]. A key challenge of developing therapeutic antibody mixtures is an unclear regulatory path. In theory, both the individual components and the combination in the antibody cocktail need to be proven safe and efficacious in clinical trials. This scenario would require large, complex, and costly clinical trials and could be cost prohibitive.

Even with many formidable obstacles, progress continues to be made developing combinatorial biologics. We expect this field to overcome some of these hurdles in the coming years to deliver novel antibody-based therapeutics and treatment methods into the clinic.

### 1.5.3 Drug Delivery

The delivery of protein-based biologics is an actively investigated area in the pharmaceutical industry. Unlike SM drugs, which are primarily delivered orally as tablets or capsules, the delivery of protein-based biologics is largely limited to

parenteral administration. Nasal and local delivery methods have also been applied in some cases. However, oral delivery remains largely infeasible and represents the holy grail of the field. Many efforts are under way, including novel nanoparticle formulation and the development of small and highly stable scaffolds that may survive trafficking through the digestive tracks to reach circulation. However, it will likely take years before oral delivery of protein drugs becomes a reality.

Another delivery challenge is for biologics to cross BBB to reach therapeutic targets that reside in the CNS. A number of biotechnology companies are applying novel technologies to this area using both targeted delivery and nanoparticle-based method to deliver biologics across BBB. So far the success has been limited.

### 1.5.4 Immunogenicity

Because of the complex multifactorial causes of immunogenic response in patients to protein drugs, it is extremely difficult to accurately predict the immunogenicity of a therapeutic candidate. It is possible, however, to assess the potential immunogenicity risks associated with a candidate molecule during preclinical development and devise clinical plans that monitor and manage the antidrug response in human patients. Because this is a critical safety aspect of biologic drugs, the field will strive to unravel the immunologic basis of immunogenicity at the molecular, cellular, and physiological levels, as well as design and implement new testing methods to accurately measure the clinical antidrug response.

### 1.5.5 Streamline the Drug Discovery and Development Process

Cutting-edge technologies are constantly applied to streamline both early drug discovery and manufacturing processes for the development of biologics. In discovery, automation is routinely implemented at various stages of the workflow to reduce manual labor and increase the throughput of various screening steps. In manufacturing, there is a strong trend toward switching from the traditional permanent bioreactors to disposable bioreactors. Meanwhile, production of extraordinarily high levels of therapeutic proteins using optimized protein expression systems has become a reality. These technology innovations will ultimately lead to a significant cost reduction in making biologics, ultimately improving the commercial viability for this class of drugs.

## 1.6 CONCLUSION

An industrial focus on the development of biologic drugs, including both new therapeutic entities and biosimilars, will continue for years. Currently, a sophisticated process built on the foundation of cutting-edge technologies has been put in place to support biologic drug development. Technological advancement will continue to make a major impact to this field, shaping both the process and the therapeutic entity. There are similarities as well as differences in the development of SM and biologic drugs. A

balanced portfolio with an equal share of both SM drugs and biologics is becoming a typical business strategy for large pharmaceutical companies.

## REFERENCES


1. CEBR. Science and the Regulation of Biological Products: From a Rich History to a Changing Future. Silver Spring, MD: U.S. Food and Drug Administration; 2001.
2. Highleyman L. FDA approves fomivirsen, famciclovir, and thalidomide. Food and Drug Administration. BETA 1998;5. Available from: http://www.ncbi.nlm.nih.gov/pubmed/11365993.
3. Schmidt C. FDA approves first cell therapy for wrinkle-free visage. Nat Biotechnol 2011;29(8):674–675.
4. Seimetz D, Lindhofer H, Bokemeyer C. Development and approval of the trifunctional antibody catumaxomab (anti-EpCAM × anti-CD3) as a targeted cancer immunotherapy. Cancer Treat Rev 2010;36(6):458–467.
5. Thompson CA. FDA approves kallikrein inhibitor to treat hereditary angioedema. Am J Health Syst Pharm 2010;67(2):93.
6. Altman LK. A new insulin given approval for use in U.S. The New York Times; October 30, 1982. Available from: http://www.nytimes.com/1982/10/30/us/a-new-insulin-given-approval-for-use-in-us.html.
7. Walsh G. Biopharmaceutical benchmarks. Nat Biotechnol 2010;28(9):917–924.
8. Kaymakcalan Z, Sakorafas P, Bose S, Scesney S, Xiong L, Hanzatian DK, et al. Comparisons of affinities, avidities, and complement activation of adalimumab, infliximab, and etanercept in binding to soluble and membrane tumor necrosis factor. Clin Immunol 2009;131(2):308–316.
9. Scolnik PA. mAbs: a business perspective. mAbs 2009;1(2):179–184.
10. Schroeder HWJr, Cavacini L. Structure and function of immunoglobulins. J Allergy Clin Immunol 2010;125(2 Suppl 2):S41–S52.
11. Roopenian DC, Akilesh A. FcRn: the neonatal Fc receptor comes of age. Nat Rev Immunol 2007;7(9):715–725.
12. Kohler G, Milstein C. Continuous cultures of fused cells secreting antibody of predefined specificity. Nature 1975;256(5517):495–497.
13. McCafferty J, Griffiths AD, Winter G, Chiswell DJ. Phage antibodies: filamentous phage displaying antibody variable domains. Nature 1990;348(6301):552–554.
14. Ponsel D, Neugebauer J, Ladetzki-Baehs K, Tissot K. High affinity, developability and functional size: the holy grail of combinatorial antibody library generation. Molecules 2011;16(5):3675–3700.
15. Nelson AL, Dhimolea E, Reichert JM. Development trends for human monoclonal antibody therapeutics. Nat Rev Drug Discov 2010;9(10):767–774.
16. Almagro JC, Fransson J. Humanization of antibodies. Front Biosci 2008;13:1619–1633.
17. Wark KL, Hudson PJ. Latest technologies for the enhancement of antibody affinity. Adv Drug Deliv Rev 2006;58(5-6):657–670.
18. Kaneko E, Niwa R. Optimizing therapeutic antibody function: progress with Fc domain engineering. BioDrugs 2011;25(1):1–11.

19. Lazar GA, Dang W, Karki S, Vafa O, Peng JS, Hyun L, et al. Engineered antibody Fc variants with enhanced effector function. Proc Natl Acad Sci USA 2006;103(11):4005–4010.
20. Lipinski CA, Lombardo F, Dominy BW, Feeney PJ. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. Adv Drug Deliv Rev 2001;46(1-3):3–26.
21. Maggi E, Vultaggio A, Matucci A. Acute infusion reactions induced by monoclonal antibody therapy. Expert Rev Clin Immunol 2011;7(1):55–63.
22. Baker MP, Jones TD. Identification and removal of immunogenicity in therapeutic proteins. Curr Opin Drug Discov Dev 2007;10(2):219–227.
23. Daugherty AL, Randall JM. Formulation and delivery issues for monoclonal antibody therapeutics. Adv Drug Deliv Rev 2006;58:686–706.
24. Wang W, Singh S, Zeng DL, King K, Nema S. Antibody structure, instability, and formulation. J Pharm Sci 2007;96(1):1–26.
25. Jenkins N, Murphy L, Tyther R. Post-translational Modifications of recombinant proteins: significance for biopharmaceuticals. Mol Biotechnol 2008;39:113–118.
26. Bebbington CR, Renner G, Thomson S, King D, Abrams D, Yarranton GT. High-level expression of a recombinant antibody from myeloma cells using a glutamine synthetase gene as an amplifiable selectable marker. Biotechnology (NY) 1992;10:169–175.
27. Owyang AM, Issafras H, Corbin J, Ahluwalia K, Larsen P, Pongo E, et al. XOMA 052, a potent, high-affinity monoclonal antibody for the treatment of IL-1beta-mediated diseases. mAbs 2011;3(1):49–60.
28. Darcy E, Leonard P, Fitzgerald J, Danaher M, O'Kennedy R. Purification of antibodies using affinity chromatography. Methods in Mol Biol 2011;681:369–382.
29. Zhou JX, Tressel T. Current practice and future strategies for viral clearance in mAb downstream production. Am Pharm Rev 2008;11(1):109–110, 112, 114–121.
30. Brass JM, Krummen K, Moll-Kaufmann C. Quality assurance after process changes of the production of a therapeutic antibody. Pharm Acta Helv 1996;71:395–403.
31. GE Healthcare. Pilot-scale capture of monoclonal antibody IgG4 on MabSelect Xtra with high purity and yield. Application Note 11-0011-58 AB; 2007.
32. Glynn J, Hagerty T, Pabst T, Annathur G, Thomas K, Johnson P, et al. The development and application of a monoclonal antibody purification platform. BioPharm Int 2009; (March):16–20.
33. Harris RJ, Kabakoff B, Macchi FD, Shen FJ, Kwong M, Andya JD, et al. Identification of multiple sources of charge heterogeneity in a recombinant antibody. J Chromatogr B Biomed Sci Appl 2001;752:233–245.
34. Food and Drug Administration. Drug Price Competition and Patent Term Restoration Act of 1984; 1984. Available from: http://www.kenyon.com/Resources/Hatchman/HTMLHelp/!SSL!/WebHelp/Public_Laws/P_L__98_417__1984_htm.
35. Food and Drug Administration. Biologics Price Competition and Innovation Act; 2009. Available from: http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/UCM216146.pdf.
36. Food and Drug Administration. FDA guidance for industry: quality considerations in demonstrating biosimilarity to a reference protein product. Silver Spring, MD: U.S. Food and Drug Administration; 2012.

37. Food and Drug Administration. Regulation of biological products; 1999. Available from: http://www.fda.gov/RegulatoryInformation/Legislation/ucm149278.htm.
38. European Medicines Agency. Guideline on similar biological medicinal products. London. 2005. Available from: http://www.emea.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC500003517.pdf.
39. European Medicines Agency. Draft: guideline on similar biological medicinal products containing monoclonal antibodies. London, UK: European Medicines Agency; 2010.
40. McCamish M, Woollett G. Worldwide experience with biosimilar development. mAbs 2011;3(2):209–217.
41. Schellekens H. Biosimilar therapeutics: what do we need to consider? NDT Plus 2009; 2(Suppl 1):i27–i36.
42. Hamers-Casterman C, Atarhouch T, Muyldermans S, Robinson G, Hamers C, Songa EB, et al. Naturally occurring antibodies devoid of light chains. Nature 1993;363(6428): 446–448.
43. Koide A, Bailey CW, Huang X, Koide S. The fibronectin type III domain as a scaffold for novel binding proteins. J Mol Biol 1998;284(4):1141–1151.
44. Gebauer M, Skerra A. Engineered protein scaffolds as next-generation antibody therapeutics. Curr Opin Chem Biol 2009;13(3):245–255.
45. Alley SC, Okeley NM, Senter PD. Antibody–drug conjugates: targeted drug delivery for cancer. Curr Opin Chem Biol 2010;14(4):529–537.
46. Gualberto A. Brentuximab Vedotin (SGN-35), an antibody–drug conjugate for the treatment of CD30-positive malignancies. Expert Opin Investig Drugs 2012;21(2): 205–216.
47. Krop IE, LoRusso P, Miller KD, Modi S, Yardley D, Rodriguez G, et al. A phase II study of trastuzumab emtansine in patients with human epidermal growth factor receptor 2-positive metastatic breast cancer who were previously treated with trastuzumab, lapatinib, an anthracycline, a taxane, and capecitabine. J Clin Oncol 2012;30(26):3234–3241.
48. Ruf P, Gires O, Jäger M, Fellinger K, Atz J, Lindhofer H. Characterisation of the new EpCAM-specific antibody HO-3: implications for trifunctional antibody immunotherapy of cancer. Br J Cancer 2007;97(3):315–321.
49. Stasi R. Rozrolimupab, symphobodies against rhesus D, for the potential prevention of hemolytic disease of the newborn and the treatment of idiopathic thrombocytopenic purpura. Curr Opin Mol Ther 2010;12(6):734–740.

# Nonclinical Safety Assessment of Biologics, Including Vaccines

LIANGBIAO GEORGE HU and DAVID W. CLARKE

## 2.1 INTRODUCTION

There are several definitions as to what constitutes a biopharmaceutical. The official Food and Drug Administration (FDA) definition of "biological products" or "biologics" can be summarized as "any virus, therapeutic serum, toxin, antitoxin or analogous product applicable to the prevention, treatment or cure of diseases or injuries of man" as codified in 21 CFR 600.3 [1]. The European Union regulations define "biological medicinal products" as "a protein or nucleic acid–based pharmaceutical substance used for therapeutic or *in vivo* diagnostic purposes, which is produced by means other than direct extraction from a native (non-engineered) biological source" [1].

The definition of a biotechnology-derived product (biologics) has changed over time. Biologics can generally be divided into various classes, which include recombinant proteins, monoclonal antibodies (mAbs), vaccines, gene transfer products, and somatic cell therapy products. However, some other non–biotechnology-derived biologics, including synthetic peptides and synthetic oligonucleotide-based technologies produced by chemical synthesis, conventional biologics, and tissue products, may also be classified as biologics [2,3].

For any pharmaceutical, either chemically synthesized small molecules (SMs) or biologics, the regulatory expectations for nonclinical testing is outlined in various regulatory documents [4]. The initial standardized international guidance was put forth by the International Conference on Harmonization (ICH), which dealt directly with the nonclinical testing of biopharmaceuticals (ICH S6, 1996) [5]. Revised regulatory guidances ICH M3(R2) [6], ICH S6(R1) [7], and ICH S9 [8] expanded on the guidelines for biotechnology-derived products.

Biologics have many properties that are different from traditional chemically synthesized SMs. Biologics are generally large molecules or heterogenic mixture of large molecules; have complex physicochemical characteristics; are heat sensitive, which makes them not easily characterized; often display species specificity and pleiotropism in their pharmacologic effects; and are often immunogenic. SMs are

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*,
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

usually synthetic, organic compounds with well-defined structures and relative stability. Biologics generally have high selectivity and specificity for their targets, significantly minimizing the risk of off-target adverse effects, except antibody drug conjugate (to be discussed later) and are often catabolized to amino acids with limited metabolite-related toxicities. In contrast, SM drugs have high potential for off-target activity, and such toxicity is often mediated through metabolism by liver enzymes, such as cytochrome P450. Whereas biologics are usually administered by the parenteral route (i.e., intravenous [IV], subcutaneous [SC], or intramuscular [IM]), SM drugs are often able to be administered orally. Unlike most SM drugs with a short half-life (usually hours), many biologics, particularly those with a large molecular weight, can have a long half-life (up to a few days or weeks), which allows weekly to monthly dosing clinically. Biologics also have some unique advantages in that targets presumed nontractable by chemical means might be suitable for biologics, which creates new opportunities to manipulate the well-defined functional domain(s) of a biologic to improve its safety and efficacy, such as bispecific mAbs and vaccines [4,9,10].

The increase in the number and vast array of biologics in development has been paralleled by innovations in bioengineering and production technologies that make industrial-scale manufacture of biotechnology-derived therapeutic agents possible. Nonclinical safety testing programs must be rationally designed with a strong scientific understanding of the product, including its method of manufacture, purity, sequence, structure, species specificity, pharmacologic and immunologic effects, and intended clinical use. In addition, a firm understanding of the regulatory requirements for a particular product type, such as mAb, antibody–drug conjugates (ADCs), or vaccine will ensure that the most appropriate and regulatory-compliant test systems are used to achieve the best chance of gaining regulatory approval for clinical testing and marketing authorization in the shortest possible research and development cycle [11–13].

Vaccines represent a unique class of biopharmaceutical products. Two major classes of vaccines have been approved or are under development: prophylactic and therapeutic (refer to other chapters in this book). Requirements for nonclinical toxicity studies depend on the vaccine's potential antigenic target, risk-to-benefit consideration, target population, available clinical data from relevant products, product features, and the availability of animal models [2,14–16]. In general, a dialog with the regulatory agencies would help manufacturers and vaccine developers clarify what preclinical toxicology studies are needed, approaches to study design, and the extent of nonclinical toxicology study documentation required before initiation of clinical trials [17].

An ADC is a hybrid that carries both properties of a biologic (mAb portion) and an SM (linker and payload). The ADC concept is based on combining the strengths of the individual large and SMs while eliminating the weaknesses of each individual approach. A well-designed ADC should minimize the systemic toxicity of the free drug (payload) by selectively delivering it to the target cells. The mAb portion of the ADC would bind to the target protein on the cell surface, which in turn triggers the internalization of the antibody–antigen complex. After release within the endosome or lysosome, the payload or linker payload would exert cytotoxic effects on the

cells expressing the target. The mAb may in itself have biological activity; however, this is not a strict requirement for an ADC. To date, the majority of ADCs in development are for cancer indications, and the most common drugs or payloads include tubulin inhibitors, such as auristatins and maytansinoids; compounds that induce DNA double strand break, such as calicheamicin; and DNA minor groove binder alkylators, such as duocarmycin. The payloads are generally attached to the mAb via a cleavable linker (susceptible to acid or catalytic cleavage) or a non-cleavable linker that requires proteolytic degradation of the antibody to release the payload. A strong scientific understanding of the biology and distribution of the mAb component, the nature of the linker (cleavable vs. noncleavable) and pharmacology of the payload is critical to rationally design the nonclinical toxicity studies. A characteristic of the payloads used in ADCs is that they are extremely potent (nM $EC_{50}$s) and therefore are extremely toxic, which can lead to toxicity in cells that express the target. Therefore, it is critical to identify a target that is expressed primarily on tumor cells to limit the toxicity in nontumor cells. So far, the clinical data generated with the leading ADC platforms indicate that their toxicities overlap and include thrombocytopenia, neutropenia, peripheral neuropathies, and elevation in liver enzymes, which appear to be serious safety concerns. Nonclinically, in addition to the clinical observations, certain dose-dependent renal, lung, and liver changes represent serious nonclinical complications [18]. Given the differences in the mechanism of action of different payloads and the biological roles of the respective cell surface antigens, many of the dose-limiting toxicities observed clinically and nonclinically may be related to off-target toxic effects, potentially caused by binding to FcRγ or nonspecific endocytosis or pinocytosis-mediated uptake of ADCs in certain cell types. The nonspecific release of the toxic payload in bone marrow or liver might be another off-target contributor to the toxicities in these organs [19–24]. To date, there are no specific regulatory guidelines for the nonclinical safety assessment of ADCs.

The remainder of this chapter discusses nonclinical factors that need to be considered when developing a biologic with respect to both program strategy and the nonclinical evaluation of toxicity.

## 2.2 CONSIDERATIONS IN THE SELECTION OF THE ANIMAL SPECIES

### 2.2.1 General Introduction

Based on the specificity of the target for most biologics, the selection of relevant species for nonclinical studies, including toxicology studies, is a critical issue to allow accurate translation of animal data into humans. The key principle in ICH S6 (R1) for selection of a toxicology animal species is the pharmacologic relevance of the species [5,7]. ICH S6 indicates "comparisons of target sequence homology between species can be an appropriate starting point, followed by cell-based assays to make qualitative and quantitative cross-species comparisons of relative target binding affinities and receptor/ligand occupancy and kinetics. Assessments of

functional activities are also recommended" [5]. Therefore, a pharmacologically relevant species should have the following minimum properties: expression of the target protein or amino acids, the target to be recognized by the therapeutic biologics with sufficient affinity to allow the evaluation of the pharmacological activity, and a target with pharmacologic functions similar to those in humans. Also, the biologic should possess similar functional potency in humans and animals. Other factors, such as tissue distribution of the target in the selected animal species, should be comparable to that in humans.

Even though the tissue expression profile may not accurately predict the toxicity of most biologics, such as mAbs, it is critical for the selection of a relevant animal species.

Vaccines represent a unique type of biologic. In particular, prophylactic vaccines generally target infectious diseases; therefore, they may not demonstrate sequence homology with mammalian species. Unlike toxicology programs for SMs and other biologics, vaccine programs typically use a single species, which must be demonstrated to be a relevant animal model based on the immunogenicity or efficacy of the vaccine in the selected animal species. Routinely, either rats or mice (rodent species) or rabbits (nonrodent species) are used for toxicology studies. Nonhuman primates (NHPs) may be used if it is the only species that has homology with the human antigen, which often occurs with therapeutic vaccines [2,17].

### 2.2.2 Sequence Homology

The requirement of two toxicology species, one rodent and one nonrodent species, is still a standard in nonclinical toxicology programs. Based on the literature or public databases, the homology of the DNA, RNA, or amino acid sequence among the various animal species and human should be analyzed. Even though sequence similarity is not sufficient to prove a relevant species, the data may be an important factor in the selection of toxicology animal species because the sequence homology may directly correlate to the biological function of the target [9]. For example, erythropoietin (Epo) is a hematopoietic cytokine that regulates red blood cell production and has a high sequence homology across species (human Epo is 91% identical to monkey; 85% to cat and dog; and 80 to 82% to pig, sheep, mouse, and rat), which may explain the biological and immunologic cross-reactivity that has been observed in a number of species [25]. Thus, sequence analysis would be the first step in selection of relevant toxicology species.

### 2.2.3 Target Affinity, Distribution, Biology, and Biochemical Pathways

Nonclinical *in vivo* safety studies for new chemical entities (NCEs) are designed to identify a broad range of potential adverse effects that may or may not be related to the pharmacologic action of the drug. However, biological products, such as mAbs, have high target specificity, and their toxicities are usually the result of exaggerated pharmacology, which may allow for better prediction of the adverse effects, provided that the pharmacologic actions across species are similar. To define a species as

pharmacologically relevant, ideally, the target of interest should be modulated in a manner similar to that in humans, such as activation of downstream signaling pathways or effector function [26]. Therefore, a thorough review of the biological activity of the target and comparison of the expression profile of the target between human and animal species would gain substantial information in the selection of toxicology relevant species. Concerns arising from target biology–related toxicities have been highlighted by past experiences with two therapeutic mAbs, natalizumab (α4-intergrin) and TGN1412 (anti-CD28) [27,28]. Administration of TGN1412, a super-agonist mAb, in six healthy volunteers led to devastating adverse effects related to activation of T cells [29]. Natalizumab, indicated for the treatment of multiple sclerosis, was recalled from the market because it induced a rare fatal viral demyelinating disease, progressive multifocal leukoencephalopathy, in two patients [30]. It can be deduced from these examples that without a thorough understanding of the distribution and function, mAbs with immunomodulatory properties have the potential to induce unexpected toxicities.

***2.2.3.1 Target Binding and Specificity Assay*** Surface plasmon resonance–(SPR-) derived technology, Biacore assay, is well suited to carry out qualitative studies to confirm the specificity of target interactions as well as quantitative measurements for affinity, kinetics, and concentration determination. The optical biosensors characterize binding events with samples ranging from proteins, nucleic acids, and SMs to complex mixtures, lipid vesicles, viruses, bacteria, and eukaryotic cells. A small volume of analyte can be tested easily for selective binding to 2 to 400 targets simultaneously, depending on the instrument platform chosen. Therefore, Biacore assays have become one of the most favorable options for characterizing the target binding affinity and specificity of the biologics [31]. Alternatively, evaluation of antibody binding to cells by flow-activated cell sorting (FACS) is sensitive and widely used in the development of biologics. Enzyme-linked immunosorbent assay (ELISA) and other *in vitro* assays are also available and validated for binding and specificity assays. Both Biacore and FACS analysis can provide information as to whether the biologic binds to the same site as the endogenous ligand or at a different site. With both techniques, the association or disassociation rate constants can be determined to provide an understanding of the binding affinity.

***2.2.3.2 Biodistribution*** Biodistribution of the target antigen *in vivo* should be comparable between the relevant animal species and humans to allow evaluation of toxicity arising from binding to the target antigen, which is referred to as on-target toxicity. This can be accomplished through an understanding of the RNA or protein expression in various tissues. Tissue cross-reactivity studies may also provide information with respect to the distribution of the target (see Section 2.3.4.5).

***2.2.3.3 Target Functional Activity Assay*** Per ICH guideline S6, the functional aspects of a biologic binding to the target are essential to determine if the selected species is pharmacologically relevant. Cell lines and primary cell cultures have

been used to determine the direct effects on cellular phenotype, the live cell activities, such as cytokine and chemokine release and proliferation induced by the biologics. The data may be further used to predict specific aspects of *in vivo* activities and assess quantitatively the relative sensitivity of various species including human. For example, CD28, an agonist mAb, is involved in T-cell proliferation [32], an *in vitro* cell-based assay can be used to determine if the therapeutic mAb induces the desired pharmacologic effects. In contrast, an antagonist mAb (e.g., Remicade) would functionally block the effects of administered human tumor necrosis factor α (TNF-α) in an *in vitro* cell-based assay [33]. However, an *in vitro* assay may not be readily available for all targets and may require extensive characterization of the target to understand the functional implications of binding to the target.

### 2.2.4 Use of Genetically Altered Rodents and Surrogate Antibodies

If it is not possible to identify a relevant species for toxicologic evaluations, alternative toxicology models might be required for a valid toxicology study [34].

***2.2.4.1 Genetically Altered Rodent Model*** Transgenic models that express the appropriate human transgene, such as knock-out or human knock-in mouse models, enable biologics to be tested on the human target antigen. An example is keliximab, a primatized anti-CD4 mAb. The anti-CD4 mAb, originally developed for the treatment of asthma and rheumatoid arthritis, only binds to human and chimpanzee CD4. A humanized transgenic mouse expressing human CD4 was used for pharmacokinetic (PK) and pharmacodynamic (PD) studies, single- and repeat-dose toxicity studies, and host defense and to address the safety concerns of anti-CD4 immunosuppression as an adverse side effect. However, there are disadvantages of using transgenic mouse model for biologics. If the target does not exist inherently in a rodent, it is logical to extrapolate that the target pathway might not exist; therefore, binding to the target may not have the same functional effect. Furthermore, the breeding of the transgenic mouse line is potentially challenging, such as the expression and retention of the transgene in the colony [9,35,36].

***2.2.4.2 Use of Surrogate Antibodies for Testing*** An alternative may be to raise a surrogate mAb against the homologous antigen present in a rodent. The surrogate protein is not the therapeutic biologic; however, it may be used to evaluate the pharmacologic or toxicologic effects of the pharmacologic mechanism in a rodent that is considered comparable to that of the biologic in human. The transgenic mouse and surrogate antibodies are more acceptable when the only other species that is relevant to the testing drug candidate is chimpanzees. To date, surrogate antibodies have been used frequently in reproductive toxicity studies. For example, infliximab, a humanized anti-TNF mAb, only cross-reacts with human and chimpanzee. The surrogate anti-TNF mAb was successfully used to evaluate single-dose escalation studies, a 6-month toxicity study, and reproductive toxicity studies. The disadvantage of surrogate antibodies represents a parallel development program, which is very

expensive and might not necessarily satisfy all of the regulatory concerns because the clinical candidate is not being evaluated [9,37,38].

In general, if one pharmacologically relevant species is available and used as the primary toxicology species, studies in a second species with a surrogate or homologous product are not considered to add further value for risk assessment and are not recommended [7].

### 2.2.5 Use of Animal Models of Disease

For biologics that are used to correct an excess or deficiency of a protein in a disease condition, it may not be possible to evaluate the safety in normal animals because it may lead to an unacceptable exaggeration of the pharmacology. In this case, understanding the safety in an animal model of disease, if available and appropriate, should be considered. For example, a biotherapeutic drug used to reduce calcium levels under conditions of extreme hypercalcemia can result in hypocalcemia or toxicity in a normal healthy animal.

Because of the nature of the biologics, many drug products (DPs) only cross-react and have pharmacologically activity in primates. A disease model may not be available in primates. In this case, a homologous protein could be used if available. Transgenic or KO mice may also be challenged with the disease to understand safety issues in a condition that more closely mimics the clinical situation [9,36]. In general, lack of historic control data, validation of the animal model, and uncertainty around the regulatory agencies are the major obstacles in using this approach.

## 2.3 CONSIDERATIONS FOR THE NONCLINICAL SAFETY EVALUATION

The nonclinical safety evaluation of a biologic is not substantively different from any other programs. A number of guidance documents are available to assist. The most critical one is ICH S6 guideline with respect to the types of studies, and the ICH M3 guideline provides guidance on the general timing of the studies.

The first steps in developing a nonclinical safety evaluation plan for a new biopharmaceutical are to understand the product attributes, including the biochemical and biophysical characteristics and pharmacologic activity of the product and the target. Studies should be conducted only after the relevant species have been properly identified (see Section 2.2) and can be conducted in rodents (typically rats) and NHPs (typically cynomolgus monkeys) or in the NHP only, depending on which species are considered relevant. Studies should not be conducted in a nonpharmacologically relevant species; exception to this may be with the ADCs, in which studies in a nonpharmacologically relevant species may provide information on the toxicity of the payload portion of the molecule.

In some cases, toxicity endpoints can be incorporated into a discovery pharmacology study, as appropriate. If the highest dose used in a pharmacology study adequately characterizes the toxicity profile of the molecule, it may be acceptable to

use these data to support the first in human– (FIH-) enabling studies rather than conducting a stand-alone exploratory toxicity study.

Numerous regulatory guidelines specifically detail the requirements of nonclinical safety evaluations. Among all ICH guidelines, ICH document S6: Preclinical Safety Evaluation of Biotechnology-Derived Pharmaceuticals (CPMP/ICH/302/95) [5,7] is the key guideline for all biological product development. For mAbs, the Center for Biologics Evaluation and Research (CBER) published "Point to Consider in the Manufacture and Testing of Monoclonal Antibody Products for Human Use, 1997." For vaccines, there are a few guidelines specifying different aspects of vaccine development, such as Evaluation of New Adjuvants in Vaccines (CPMP/BWP/6622/02) [39], reproductive toxicology studies for preventive vaccines for infectious disease indications (CBER, FDA, 2006) [40], plasmid DNA vaccines for preventive infectious disease indication (CBER, FDA, 2007) [41], pharmaceutical and biological aspects of combined vaccines (CPMP/BWP/477/98) [42], preclinical pharmacologic and toxicologic testing of vaccines (CPMP/SWP/465/95) [16], and WHO guidelines on nonclinical evaluation of vaccines [15].

In general, the safety assessment programs should normally include two pharmacologically relevant species. However, for many biological products, one species can be used if no other relevant species can be identified. For example, many mAbs do not cross-react with rodents, so a single species is usually sufficient. For testing mAbs, the primate is usually the species of choice for humanized or human mAbs; for vaccines, often rodents or rabbits are often used; and for gene–cell therapy vectors, rodents are the most common species. The ADCs are hybrid of SM and biologics; therefore, even though most humanized mAb portion only cross-reacts in NHPs, the toxicology studies are typically conducted in two species, rodents (typically rats) and nonrodents (typically NHPs).

### 2.3.1 Route of Administration

The route of administration for biologics is dependent on the intended clinical regimen. The oral route of administration is generally not applicable for a biologic. Most anticancer biologics, including mAbs and ADCs, especially for late stage cancer, are clinically delivered by the IV route. Most biologics used for chronic indications are typically administered by the SC route. For vaccines, historically, the IM, SC, and intradermal routes have been used, although the oral and intranasal routes are now actively exploited. Nonclinical studies should use the intended clinical route of administration when practicable.

### 2.3.2 Selection of Doses for Toxicity Studies

As outlined in ICH S6, a rationale for dose selection should take into account the exposure–response relationship. Pharmacology and PK and PD information can be used to assist in dose selection. Dose selection in all studies should be based on exposure– (Area under the concentration time curve (AUC), efficacious concentration ($C_{eff}$), etc.) response relationships rather than dose only (mg/kg) because

the exposure or response at a specific dose in one species may not directly translate to a similar exposure and response at the same dose in another species or human.

There is no scientific information to support that high doses of protein cannot be administered to animals for long periods of time. The concept of "protein burden toxicity" has not been proven, and there are examples in which large doses of protein have been administered chronically with no associated toxicity. Therefore, there is currently no rationale to limit the high dose based on these arguments. The most recent revision of the ICH S6 provides some guidance on the selection of the high dose in that one needs to consider: (1) a dose (exposure) that provides the maximum intended pharmacologic effect and (2) a dose that provides an approximately 10-fold higher exposure than the maximum dose exposure in the clinic (note: this is *not* the minimum efficacious dose). The higher level of these two doses should be selected as the high dose of the repeat dose studies unless scientific data support a lower dose [7]. A relatively large difference in binding affinity or *in vitro* potency might suggest a need for higher doses in nonclinical studies. In the event that toxicities cannot be demonstrated at the doses selected using this approach, then traditional toxicity studies with higher multiples of human dose levels are unlikely to provide additional useful information. However, a limit dose approach in alignment with ICH M3 and similar to that used for SMs can be used as appropriate.

When *in vivo/ex vivo* PD endpoints are not available, the high dose selection can be based on PK data and available *in vitro* binding or pharmacology data.

Many companies have established default dose levels for a new biologics that achieve these objectives. The highest dose is often defaulted to 100 mg/kg in NHPs and 200 mg/kg in rodents, with the exception of ADCs, oligos, and vaccines. For ADCs, the calculated high dose levels using the assumptions described above may be far over its maximum tolerated dose. Therefore, a lower dose should be used to ensure the success of the toxicology studies [4]. For vaccines, it is generally acceptable to set the high dose to be equal to the maximum human clinical dose [15].

### 2.3.3 Safety Pharmacology

In addition to the ICH S6 [5], S6(R1) [7], and S9 [8] guidances, the ICH S7a [43] and ICH S7b [44] represent specific guidances on Safety Pharmacology studies for human pharmaceuticals. In accordance with regulatory requirements, assessments of cardiovascular (CV), central nervous system (CNS), and pulmonary function must be performed before FIH. Some biologics carry a higher level of risk for safety pharmacology issues based on target tissue expression, mechanism of action, therapeutic indication, or patient population (e.g., those for diabetes, CV, or metabolic indications or targets that are expressed on cardiac tissues). However, for most standard biopharmaceuticals it is appropriate to incorporate the evaluations into the single-dose or repeat-dose toxicity studies. When there is a concern for an effect on the CV system, a stand-alone study is likely warranted. However, the standard Latin-square design used with SMs may not be applicable because of both the long half-life and the possibility for formation of antidrug antibodies (ADAs) [45]. Rodent telemetry may be used for any biologics with a predefined CV risk and

that have pharmacologic activity in rodents (rats or mice). These studies could incorporate echocardiography if there is an elevated concern for modulation of cardiac structure (e.g., left ventricular hypertrophy) or function (percent ejection fraction or contractility); however, the biologic must be biologically active in rodents (rats or mice). If there is no biological activity in the rodent, then NHP telemetry studies (heart rate, blood pressure, electrocardiography [ECG]) may be used for early CV assessment. For biologics with no known CV risk, the collection and evaluation of basic CV endpoints (surface lead ECG in chair-restrained animals) as part of the FIH-enabling Good Laboratory Practice (GLP) NHP toxicity study is sufficient to meet regulatory guidelines [46]. It is generally not advisable to collect CV endpoints in rodent FIH-enabling GLP toxicity studies.

The *in vitro* human ether-a-go-go (hERG) assay is generally not required for mAbs, vaccines, therapeutic proteins, and therapeutic peptides because of the low propensity for binding of biologics to the potassium ion channel.

Certain biologics carry a greater risk for CNS safety issues, such as those being developed for neurologic or psychiatric indications or pain and indications in which the blood–brain barrier (BBB) may be compromised (e.g., diabetes, oncology, CNS, and peripheral inflammatory disorders) or whose targets are expressed in the central or peripheral nervous system. For biologics that fit these categories, it may be necessary to include detailed observations in the form of a neurologic examination (for NHPs) or Functional Observational Battery or Irwin Test (for rodents) in single-dose or repeat-dose studies. For biologics with little known CNS risk, the evaluation of CNS endpoints, including body temperature measurements, obtained as appropriately timed clinical observations in FIH-enabling studies in rodents or nonrodent GLP toxicity studies, should be sufficient to satisfy regulatory requirements.

Likewise, evaluation of the effect of a biologic on the respiratory system can be achieved via addition of pulmonary measurements (e.g., respiratory rate) to the FIH-enabling nonrodent GLP toxicity study. Stand-alone respiratory safety pharmacology studies are generally not performed unless there is a strong scientific justification. If both the rodent and the NHP are appropriate species for the toxicology evaluation, adding respiratory measurements to the rodent study allows for the collection of data in a species with a more defined historical database.

In general, gastrointestinal and renal assessments are not considered necessary (either as stand-alone or incorporated into toxicity studies) unless there is a perceived risk of serious concern. In such cases, specialized assessments may be designed either as additions to toxicity studies or as separate stand-alone studies.

### 2.3.4 General Toxicology Studies

In general, toxicology studies should be conducted by the same route and schedule of administration as proposed in human subjects. To gain a successful approval from regulatory agencies, a complete general toxicology package should include, but is not limited to, the following *in vivo* toxicology studies: (1) single-dose toxicity study, (2) repeat-dose toxicity studies, (3) developmental and reproductive toxicity studies,

and (4) genotoxicity and carcinogenicity studies. In addition, ICH M3R2, ICH S6, and ICH S9 provide comprehensive reviews on the duration of nonclinical studies relative to clinical development [5–8]. Safety evaluation programs should normally include two relevant species. However, although two species may be necessary to characterize toxicity in short-term studies, it may be possible to justify the use of only one species for subsequent long-term toxicity studies (e.g., if the toxicity profile in the two species is comparable in the short term). If the NHP is the only relevant species, it is generally recommended to use mature NHPs (typically cynomolgus monkeys). Mature NHPs may provide useful information for fertility and reproductive toxicity (refer to relevant sections).

***2.3.4.1 Single-Dose Toxicity Studies*** Single-dose toxicity studies with biologics may, in theory, be useful in describing the relationship between dose and systemic and local effects and in defining dose levels for use in repeat-dose studies but are not intended to define the no-observed-adverse-event level (NOAEL). These studies should be conducted in a pharmacologically relevant species, rodent (typically rat) and NHPs (typically cynomolgus monkeys). For most biologics, for which NHP is the only relevant species, one species should be sufficient. ADCs represent a somewhat unique situation in that they are a hybrid of both SM and mAb. As such, a single- or two-dose non-GLP toxicity study in rodents (typically rats), in addition to NHPs, is recommended to screen the off-target related toxicities. For prophylactic vaccines, single-dose toxicity studies are not usually conducted.

Single-dose toxicity information may be obtained from pharmacology and PK studies provided they are conducted using the relevant species and appropriate dose level to characterize the toxicity profile.

Typically, single-dose toxicity studies are conducted at the early phase of the drug development using research-grade drug substance (DS). The resource constraints, such as the amount of available DS, may effectively limit the high dose selection. These studies are generally not conducted according to GLPs [6].

For single-dose toxicity studies in NHP, animals should be observed for at least 2 weeks after dosing, and blood samples for measurements of plasma drug concentrations should be included. If there is a cause for concern, animals can be necropsied, and a limited major organ panel and suspected target organs should be examined microscopically. The panel should include the liver, lung, kidneys, spleen, bone or bone marrow, eyes, heart, testis or ovaries, and injection site. Based on the type of biological modality, the dose level and study duration should be adjusted to obtain reliable data to allow for dose selection for longer repeat dose studies. If the biologic cross-reacts with rodents, then a 2-week observation period with serum collection for drug level determinations followed by a necropsy may be appropriate. Microscopic evaluations should consist of the major organs, as outlined for nonrodents. Collection of blood samples for immunogenicity analysis may be helpful but not essential.

If a biologic is to be administered only once in the clinic, there may be value in conducting a more detailed GLP single-dose toxicity study, as outlined in the ICH M3(R2) [6] to support dosing in early clinical trials.

***2.3.4.2 Repeat-Dose Toxicity Studies*** Many biologics are intended for repeat dosing in the clinic, and hence repeat-dose toxicity studies are usually required to support drug safety for the first human clinical trial. The DP used for repeat-dose toxicity studies should be highly comparable to, if not the same as, the DP to be used for studies in human in terms of manufacturing processes and analytical quality. There are likely to be more than one repeat-dose study of different durations conducted during the development of biologics. The duration of dosing in these studies should be sufficient to support clinical development as outlined in ICH S6 and ICH M3(R2) [5–7]. The longest duration of dosing is generally 6 months for chronic toxicology studies to support biological product registration except for late stage cancer indications in which a 3-month toxicology study in duration is considered adequate [6,8,34,47].

An exception to this is for vaccines, in which the number of doses administered generally exceeds maximum clinical doses ($N$) by at least one administration ($N+1$). The vaccine dosing frequency should be sufficient to induce the desired immune response, typically biweekly, and the route of administration is typically IM. A subset of animals is terminated 3 days after the last dose, and the remaining animals approximately 28 days after the last dose. With vaccines it is strongly recommended to conduct a pre–investigational new drug meeting with CBER reviewers to discuss nonclinical studies, clinical study designs, data requirements, and other scientific issues that need resolution before the initiation of clinical trials [17].

The repeat-dose toxicity studies are usually conducted according to GLP regulations. If appropriate, the studies should be conducted using two relevant species, rodents (typically rats) and nonrodents (typically NHPs). Typically, animals are dosed once every half-life for at least four doses with an additional dose 2 to 3 days before termination. Analysis may include the appearance of the injection site, clinical signs, hematology, clinical chemistry, urinalysis, and macroscopic and microscopic examination of all organs. For the ADC programs, animals are typically given two to four doses either every other week or every 3 weeks depending on the proposed clinical dosing frequency. Animals in the dosing phase are terminated for pathology analysis 3 days after the last dose.

The repeat-dose toxicity studies typically include three dosing groups and a vehicle control group, and doses should be selected to cover the range from a no effect level to one that induces toxicity. However, dose selection for a biologic can be complicated in that all proposed dose levels may exceed a maximum pharmacologically active dose (see Section 2.3.2). The low dose should represent a level that provides an equivalent exposure to at least an efficacious exposure and the mid-dose is included for further characterization the dose response of any observed effect [6].

In many development programs, the definitive clinical route of administration has not been established when the repeat-dose toxicity studies are initiated. Therefore, an additional high-dose group using alternative route of administration may be added to the study as the fourth dosing group. In this way, the repeat dose study would cover two potential routes of administration in clinic. In addition, some safety pharmacology endpoints (e.g., ECG if the NHP is selected), samples for ADA and TK and immunotoxicology should be included in the study [48].

A recovery period should be included to examine the reversibility of any toxic effects observed during the dosing phase studies. However, the evaluation of recovery is not warranted if there are no adverse effects at the end of the dosing period or sufficient scientific justification can be provided. The nondosing recovery period is not designed for the assessment of delayed toxicity or solely for the determination of immunogenicity. Recovery animals are usually added to the high-dose and control groups unless a high death rate is expected at the high dose. The recovery period should be equal to 5 half-lives [6,11,32]. As a default model, for mAbs and ADCs, an 8-week recovery period is usually sufficient; for vaccines, it is generally 2-4 weeks. However, for certain biologics, the recovery period may need to be 3 months or longer in order to evaluate the reversal of target organ toxicity (e.g. testicular changes) or prolonged PD effects. Unless there are specific toxicity concerns, histopathology assessment of the target organs is sufficient to determine the reversibility in recovery animals. A complete reversal of toxic effects is not necessary as per ICH guidelines S6. An extended recovery period would not add additional value.

#### *2.3.4.3 Reproductive Toxicity*

*Considerations of Species Selection for Reproductive Toxicity* A number of recent reviews have discussed considerations for assessing the developmental and reproductive toxicity potential of biologics [49–51]. The principles of developmental and reproductive toxicity studies for biologics are not different from those for other SMs; however, biologics pose some unique challenges. The need to conduct developmental and reproductive toxicity studies is dependent on the product, clinical indication, and patient population [5–7]. Because biologics tend to be highly specific for their target in the human body, they often show limited cross-reactivity with the traditional species used for reproductive studies (i.e., rats and rabbits).

Although traditional SMs (molecular weight $<1000$ daltons) readily diffuse across membranes and the placenta, biologics do not and therefore have limited access to the conceptus and are not likely to interact with DNA or other intracellular organelles. However, there are specific mechanisms for transporting particular classes of large molecules across the membranes, including the placenta. An example is the transport of antibodies across the placenta by endocytosis via the FC receptor [52]. Immunoglobulin G (IgG) is the predominant means of fetal immunity, and it is transported across the placenta. This transport happens in a linear fashion as the pregnancy progresses, with the largest amount transferred in the third trimester. Preferential transport occurs for IgG1 followed by IgG4 and IgG3 with IgG2 being the least detected [53,54].

The choice of the species to be used must be carefully determined. Although many biologics are less cross-reactive with any of the standard species (rats, rabbits) used for reproductive toxicity studies compared with NHPs, ICH S6 provides some latitude in that it only indicates the studies should be conducted in a relevant species, not necessarily the most relevant species. Although NHPs may be physiologically similar to humans, it is extremely difficult to perform reproductive toxicity studies in

NHPs because of the high rate of spontaneous abortion (up to 28%), low fertility, few offspring, and long gestation [9,34]. As a general principle, when the clinical candidate is pharmacologically active in rodents and rabbits, these species should be used for reproductive toxicity studies unless there is a scientific reason to use an NHP.

The use of traditional animal models has many advantages in that the study designs, methods, and historical data are readily available. However, if the NHP is the only relevant species, one needs to consider their use. A large historical database exists for both rhesus and cynomolgus monkeys, and many contract research organizations have experience with both strains. Cynomolgus monkeys are more appropriate for reproductive toxicity testing than rhesus monkeys because the latter are seasonal breeders, which makes the toxicity testing especially difficult [25]. However, compared with rodents, in which antibodies can cross the placenta throughout gestation, in human and NHPs, transfer of antibodies occurs primarily during the latter part of pregnancy (i.e., after organogenesis). This could lead to teratogenic effects being noted in rodents but not in NHPs or humans. Several recent papers have dealt with many of the issues involved with conducting reproductive toxicity studies in NHPs [49,50].

An alternative reproductive toxicology strategy may be used in place of NHPs if a sound scientific justification can be provided. As discussed in the previous sections, use of transgenic mice expressing the human target or use of a homologous protein in a species expressing an analogous copy of the human target can be considered when the biologic only cross-reacts with humans and chimpanzees [9]. The use of genetically modified mice as a tool for evaluating developmental and reproductive toxicity is gaining some acceptance for those products where bioactivity is limited to humans and NHPs. The use of these models may allow for the identification of the reproductive hazard associated with the biologic. For example, a knock-out animal that was viable and without discernible development issues may support that a biologic with antagonist activity may be of limited hazard with respect to reproductive and developmental effects.

*Fertility* When a clinical candidate is only pharmacologically active in the NHP, the potential effects on male and female fertility can be assessed by standard histopathology evaluations, organ weights, and assessment of menstrual cyclicity in repeat-dose toxicity studies for at least 3 months using sexually mature NHPs. This study can be included in the repeat-dose toxicity study protocol. However, one must ensure that sexually mature animals are available and used in these studies. If there are specific concerns over reproductive toxicity, sperm assessment, and reproductive hormone analysis may be included in the study protocol.

If rodents are a relevant species, a standard fertility toxicity study using the protocols consistent with those for SMs is appropriate.

*Embryo–Fetal Development and Pre- and Postnatal Development* The objective of the developmental toxicity assessment is to provide safety information on the risk of exposure during pregnancy, pre- and postnatal development. Biological candidates

with Fc-containing modalities, such as mAbs, do have the capability of crossing the placenta with a species-specific profile over the course of gestation [52]. Because of the differences in placental transfer across species, the information on embryo–fetal exposure in nonclinical studies would be important for human risk assessment.

For both SM and biological candidates, the most preferred toxicology species for embryo–fetal development (EFD) studies are rabbits and rodents. If the biological candidate molecule is pharmacologically active in both rodents and rabbits, both species should be used to evaluate for effects on EFD. However, one species should be acceptable for EFD evaluation if the species selection is scientifically justified or if lethality or teratogenicity is identified in one of the two species. For pre- and postnatal development (PPND) studies, one species (rodents, rabbits, or NHPs) should be sufficient. However, the NHP is the least favorable species [6].

If the clinical candidate is only active in NHPs, the regulatory preference is to test the clinical candidate in the NHP, accepting the limitations of NHP studies and alternative approaches, including the use of surrogate molecules in rodent studies. If the sponsor chooses to use NHPs as the only relevant species, an enhanced PPND (ePPND) study [50] is typically initiated after the proof of conception with dosing continuing throughout gestation and into the postnatal period.

In general, the EFD toxicity studies should be conducted before enrolling a large number of women of childbearing potential in clinical trials (phase III) [6], and the PPND and ePPND studies should be available to support the submission of a marketing application. If NHPs represent the only pharmacologically relevant species, the ePPND study may be conducted earlier in development to support enrollment of women of childbearing potential rather than conducting a separate EFD study in NHPs.

*Juvenile Toxicity* The nonclinical safety information on the risk of direct exposure during postnatal development supports the use of biological products in pediatric populations. There is no specific regulatory guidance for nonclinical toxicity studies. However, the same approaches used with SMs should be used for the biological candidates. Overall, the need for a juvenile animal study is usually determined on a case-by-case basis, and the study should be conducted before initiation of clinical trials in pediatric populations.

***2.3.4.4 Genotoxicity*** Based on the ICH guideline S6 [5,7], standard genotoxicity assays, such as Ames assay, *in vitro* micronucleus, and *in vivo* micronucleus assays, are not required to be conducted for biological products because they would not be expected to interact directly with DNA or other chromosomal material [5]. However, for ADCs, although the complete ADC should not enter the nucleus to interact with DNA, the linker or linker and payload may; therefore, a genotoxicity assessment on the linker alone or linker and payload combination may be warranted.

***2.3.4.5 Carcinogenicity Studies*** As indicated in the ICH S6 guideline [5,7], standard carcinogenicity bioassays are generally inappropriate for biologics. However, assessment of carcinogenic potential may still be needed based on the duration of clinical dosing, patient population, or biological activity of the product (e.g.,

growth factors, immunosuppressive agents). In cases when the product is biologically active and nonimmunogenic in rodents, an assessment of carcinogenic potential may need to be evaluated in a single rodent species. When the product is not biologically active in rodents or results in a significant immunologic response, alternative approaches must be considered. These scientifically justified alternative approaches should be discussed with the regulatory agencies before conducting the work; this could include use of KO animals, *in vitro* cell proliferation assays, or standard 2-year carcinogenicity studies, and the use of a surrogate (although generally not favored) may be considered [55].

***2.3.4.6 Immunogenicity*** Almost all therapeutic proteins induce some level of an antibody response [36]. An immune response to a biologic in animals does not mean that it cannot be developed. Although such an antibody response does not necessarily cause severe side effects, clinical implications can occur. Antibody formation may lead to a loss of efficacy, neutralization of the endogenous counterpart, or general immune system effects (i.e., serum sickness–like disease, infusion reactions, anaphylaxis, and anaphylactoid reactions). In nonclinical studies, immunogenicity assessments are conducted to assist in the interpretation of the study results, although immunogenicity assessments from the nonclinical studies are not likely relevant for predicting the immunogenicity of human or humanized proteins.

Therefore, measurement of antibodies associated with administration of biological products should be performed when conducting repeated dose toxicity studies to aid in the interpretation of these studies. Antibody responses should be characterized (e.g., titer, number of responding animals, neutralizing or non-neutralizing), and their appearance should be correlated with any pharmacologic or toxicologic changes. Specifically, the effects of antibody formation on PK and PD parameters, incidence and severity of adverse effects, complement activation, and the emergence of new toxic effects should be considered when interpreting the data. Attention should also be paid to the evaluation of possible pathological changes related to immune complex formation and deposition. Humans may develop serum antibodies against humanized proteins, and frequently the therapeutic response persists in their presence. The nonclinical safety studies with biologics need to include an assessment of the immunogenicity of the protein. This is used to help to be able to interpret the results from the study [9,36].

***2.3.4.7 Tissue Cross-Reactivity*** There has been considerable discussion on use of the tissue cross-reactivity (TCR) study [56]. General consensus is that it should not be used as the basis to justify species selection for the toxicity program. TCR studies are used to provide an understanding of the biological distribution of the target in humans compared with relevant species used in the toxicology studies. For an ADC in which one is delivering a potent toxin to the cells, understanding the TCR may be critical to interpreting the results of the toxicology studies.

TCR studies should be conducted for therapeutic mAbs and related antibody-like biologics (defined as containing a complementarity-determining region [CDR]). The primary objective of the TCR study is to identify off-target binding and secondarily

to identify sites of on-target binding that have not been previously identified. A comparison of the patterns of staining between human and animal tissues may also be used to support the relevance of species evaluated in nonclinical toxicology studies but should not be used as the primary rationale for species selection; the selection of species for nonclinical toxicity testing should primarily be based on *in vitro* or *in vivo* pharmacology data (see Section 2.2). The results from the TCR studies are used to guide nonclinical and clinical investigators about potential target organs of toxicity based on the presence of staining in tissues.

A more detailed description of TCR studies, including a discussion of methodologies and interpretation, has been recently published [56].

***2.3.4.8 Cytokine Storm Assay*** There has been an awareness that administration of particular mAbs has been associated with a generalized increase in multiple cytokines (cytokine release syndrome) [57]. There is a heightened awareness that biologics that activate the innate immune system or that stimulate specific immune cells or malignant cells for immune-mediated destruction may be potential target for immune activation. Stimulation of these cells can result in a massive release of cytokines into the systemic circulation, such as what was seen as part of the cytokine storm that followed the administration of TGN1412. If the target of a biologic is likely to result in immune stimulation that could result in the release of cytokines, one needs to consider the potential for the cytokine release syndrome. The potential for this should be evaluated using human cells and cells from the appropriate animal species in which toxicity studies will be conducted.

## 2.4 PHARMACOKINETICS

The evaluation of the PK properties is important for biologics to allow for evaluation of the PK–PD relationship and to allow for data extrapolation from animals to humans. Although there are many similarities with what is required for an SM, there are some important differences, which are discussed in the following sections. Studies evaluating PK of vaccine candidates are generally not required [15,16].

### 2.4.1 Absorption

Most biologics generally have poor oral bioavailability, so administration is generally by the IV, IM, or SC route. Absorption of biologics after injection is influenced by the physicochemical properties and molecular weight of the protein, product excipients, blood flow at the injection site, temperature, and so on. Oral delivery is limited by barriers such as enzymatic and pH-dependent degradation in the gastrointestinal tract, and low membrane permeability. In general, oral bioavailability of biologics is erratic and less than 1% [58]. There have been several recent advances in increasing the absorption by nonparenteral routes of administration, but to date none have advanced into the market except a few through pulmonary delivery [59].

### 2.4.2 Distribution

Because of their large size, distribution of biologics is generally limited to the central compartment. Large proteins (molecular weight >30 kDa) cross blood capillaries slowly; however, they may bind to specific binding proteins involved in their transport and regulation, which may alter the distribution. In addition, there can be specialized mechanism for distribution into specific organs (e.g., CNS and fetus). The BBB is effective in preventing unregulated entry of substances from the blood, and it is equally effective at selectively permitting entry of necessary blood components. Extensive characterization of the transport of mAbs across the BBB has been started for potential utility in cancer therapy. Active transporters do exist for nutrients and regulatory molecules [60]. Of special relevance for antibody therapeutics, FcRn, the Fc receptor that protects antibodies from degradation in serum, is highly expressed in brain vessels [61]. In other organs, another function of FcRn is to transport antibody across tissue barriers. For the brain, a role in removal of harmful antibody from the brain has been suggested [62], but the actual roles are not yet known. In other tissues, FcRn-mediated transport is bidirectional, and the predominant direction can be modified experimentally [63]. The FcRn may also play a role in transporting mAbs across the BBB—or might be exploited for that purpose.

### 2.4.3 Metabolism

Biologics are generally degraded to smaller peptides or individual amino acids via pathways generally understood for the endogenous compounds. The individual metabolites (amino acids) would then be reused in the endogenous pool for the synthesis of structural and functional proteins. Because of the known metabolism to endogenous amino acids, metabolism studies generally required for new chemical entities are not needed for biologics.

The major site for the metabolism of biologics would be the same as for endogenous proteins, being the liver, kidney, the blood, and extravascular site of administration. The extravascular metabolism likely accounts for the low bioavailability observed after SC or IM administration compared with IV.

However, there are special instances when the metabolism of the biologic needs to be considered. ADCs, made of a mAb with an SM linker and payload, need to be evaluated for metabolism, release of the payload, and the PKs of the released product, including its metabolism and distribution. A more detailed understanding of the metabolism of the released substance may be accomplished through separate studies when the released substance is administered alone. However, these results cannot replace studies with the complete ADC because the distribution of the released product when administered alone may be different from that when released from the ADC.

### 2.4.4 Excretion

As outlined earlier, because biologics are generally degraded to individual amino acids and enter the endogenous pool, biologics are generally not excreted as unchanged proteins; therefore, evaluation of excretion is generally not required for a biologic, except ADCs where excretion of linker, payload, and metabolites, may need to be evaluated.

## 2.5 SELECTION OF THE INITIAL CLINICAL DOSE

Several good guidance documents outline the procedures for the selection of the initial starting dose for a human clinical trial [64,65]. Although these guidelines are generally focused on SMs, they may still apply to biologics. With respect to biologics, the criteria for setting the initial starting dose changed on March 13, 2006, when TGN1412, a mAb, was administered to humans for the first time, using initial starting doses consistent with the FDA guidance (i.e., using a fraction of the NOAEL dose level in nonclinical species). Severe, life-threatening toxicities ensued. The science behind these adverse events and the immediate responses to this event has been described in another article [29]. The incident with TGN1412 was a wake-up call to the pharmaceutical industry, the clinical trials community, and the regulatory agencies. The incident was investigated thoroughly by several different groups, including the Expert Group on Phase One Clinical Trials (chaired by Professor Gordon Duff) [66], the Royal Statistical Society, and the Early Stage Clinical Trial Task-force. Each of these groups issued reports that summarized the causes of the adverse events and proposed ways that such adverse events could be avoided in future FIH studies. This has led the European Union to issue "Guideline on Strategies to Identify and Mitigate Risks for First-in-Human Clinical Trials with Investigational Medicinal Products" [67]. The guideline is wide ranging in scope and is intended to assist sponsors in the transition from nonclinical to early clinical development. It identifies factors influencing risk for new investigational medicinal products and considers quality aspects, nonclinical and clinical testing strategies, and designs for FIH clinical trials.

Consistent with what we have discussed previously, the European Medicines Agency (EMA) guideline [67] describes the assessment of the relevance of the animal models and the need to ensure that the species used in the toxicology studies are pharmacologically relevant, with lesser emphasis on similarity of the metabolism of the clinical candidate in animals and humans (as has typically been the case for NCEs). The use of nonrelevant species in toxicology studies is strongly discouraged.

The guideline states that all available information has to be taken into consideration for the dose selection, and this has to be made on a case-by-case basis. In general, the NOAEL determined in nonclinical safety studies performed in the most sensitive and relevant animal species, adjusted with allometric factors or on the basis of PKs, gives the most important information. The relevant dose is then reduced or adjusted by appropriate safety factors according to the particular aspects of the molecule and the design of the clinical trials. However, for investigational medicinal products for which risk factors have been identified, the use of the NOAEL with appropriate adjustments may not be appropriate, and the use of the minimal anticipated biological effect level (MABEL) approach is recommended. The MABEL is the anticipated dose level leading to a minimal biological effect in humans. The guideline states that the calculation of MABEL should use all *in vitro* and *in vivo* information available from PK and PD data when such data may include (1) target binding and receptor occupancy studies *in vitro* in target cells from human and the relevant animal species, (2) concentration–response curves *in vitro* in target

cells from human and the relevant animal species and dose/exposure–response *in vivo* in the relevant animal species, and (3) exposures at pharmacologic doses in the relevant animal species [67].

When calculating the FIH dose using NOAEL data, safety factors should be applied to the calculation, and when using the MABEL approach, similar safety factors also need to be incorporated. However, the EMA guideline does not give insight as to the magnitude of the safety factor that should be used.

Although the EMA guideline describes two different (NOAEL and MABEL) approaches for the calculation of the FIH dose, the sponsor should look at the totality of the data available to them and evaluate several different approaches to the calculation of a safe starting dose and then take the most conservative approach. The calculation of the safe starting dose, as described in the EMA guideline, applies only to studies in normal healthy volunteers. The guideline acknowledges that "other approaches may also be considered in specific situations, e.g. for studies with conventional cytotoxic in oncology patients" [67].

## 2.6 FUTURE DIRECTIONS

To date the most common forms of biologics in development have been mAb, recombinant proteins, and vaccines. However, numerous new modalities are being developed. One of these new modalities is bispecific antibodies, which are able to recognize two distinct targets at the same time and might bridge two targets together, which present an opportunity for therapeutic gains (e.g., bringing cancer cells within closer proximity to cytotoxic T lymphocytes). When one considers species selection, if both CDRs do not have cross-reactivity to the same species, one needs to understand how to appropriately assess the safety if the full pharmacologic effect cannot be generated in one species.

As discussed throughout this chapter, ADCs are another new area for biologics that requires a unique approach for the nonclinical toxicity evaluation. There are currently two approved ADCs (Mylotarg® and ADCetris®), so the development path is a little more defined. However, the development of an ADC represents a hybrid approach between a traditional mAb and an SM.

A number of nucleic acid–based products are in development. These can generally be divided into gene therapy products, DNA vaccines, and antisense and RNA-based modalities. To date these nucleic acid–based products have not been approved for commercial use in humans, and therefore there is a less defined development path. However, some unique characteristics for each of these approaches need to be considered.

Gene therapy seems elegantly simple in its concept–supply the human body with a gene that can correct a biological malfunction that causes a disease. Researchers are testing several approaches to gene therapy, including replacing a mutated gene that causes disease with a healthy copy of the gene; inactivating, or "knocking out," a mutated gene that is functioning improperly; or introducing a new gene into the body to help fight a disease [68]. Some of the issues for the

nonclinical evaluation of gene therapy products deal specifically with the delivery system used and its acceptability in animal models. Additional issues are similar to those covered in previous sections: comparison of gene function in animals with humans, replacement of gene in normal animals, and relevance of a human gene in an animal. These all need to be considered when designing a toxicology program for a new gene therapy agent.

DNA vaccines represent another novel technology whereby genetic material (usually a plasmid) encoding an antigen is transferred to cells of a vaccine recipient. These cells express the antigen and present it to the recipient's immune system. The concern with DNA vaccines is slightly different from the concern for gene therapy because a DNA vaccine is not intended to be incorporated into the host DNA; therefore the concern revolves around incorporation of the plasmid DNA into the host chromosome DNA, resulting in long-term persistence and increasing the risk of mutagenesis and oncogenesis and the potential that the integration leads to vertical transmission [69,70]. This must be evaluated in a biodistribution study looking at the persistence and incorporation of the plasmid DNA into the host genome. An additional concern involves the potential for development of immune tolerance to the protein or the indication of autoimmunity through the induction of an anti-DNA immune response.

Antisense RNAs represent another new class of nucleic acid–based products that are currently in development. Antisense RNA acts by introducing a synthetic complementary (antisense) RNA (RNAi) into the cell. However, the uptake of RNAi has been somewhat slowed because of issues associated with delivery (how to target the RNAi to the cells of interest) and safety [71]. One also need to understand the sequence homology between the human RNAi and the animal species being used; if there is no match, then the consequence of the RNAi cannot be fully evaluated in the animal.

Stem cells represent a unique type of biologics; although not expressed by cells, they represent novel technologies. Several recent review papers describe some of the unique characteristics of stem cells that need to be considered in their development [72]. Of particular concern to stem cell products include the biodistribution (where do the cell go and how long do they last), immunogenicity (whether the cells are likely to be immunogenic and result in immune-based rejection), and tumorigenicity (whether the cells will differentiate into tumor cells and developing tumors, such as teratomas). Therefore, when developing a stem cell product, one needs to understand the distribution of the product after administration and the level of any residual undifferentiated cells and their ability to form a teratoma.

These new biologic modalities offer great promise for the treatment of diseases but also challenges in developing the nonclinical toxicology testing paradigm to ensure the safety of the patients in the clinical trials.

## REFERENCES

1. Rader RA. (Re)defining biopharmaceutical. Nat Biotechnol 2008;26(7):743–751.
2. Brennan FR, Shaw L, Wing MG, Robison C. Preclinical safety testing of biotechnology-derived pharmaceuticals. Mol Biotechnol 2004;27:59–74.

3. King DJ, Adair JR. Recombinant antibodies for diagnosis and therapy of human disease. Curr Opin Drug Discov 1999;2:110–117.
4. Baldrick P. Safety evaluation of biological drugs: what are toxicology studies in primates tells us. Regul Toxicol Pharmacol 2011;59(2):227–236.
5. International Conference on Harmonization. ICH topic S6: preclinical evaluation of biotechnology-derived pharmaceuticals; 1996. Available from: http://www.ich.org?LOB/media/MEDIA503.pdf.
6. International Conference on Harmonization. ICH Topic M3 (R2): note for guidance on non-clinical safety studies for the conduct of human clinical trials and marketing authorization for pharmaceuticals; 2009. Available from: http://www.ich.org/LOB/media/MEDIA5544.pdf.
7. International Conference on Harmonization. ICH S6(R1). Addendum to ICH S6: Preclinical safety evaluation of biotechnology-derived pharmaceuticals S6(R1); 2009. Available from: http://www.fda.gov/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/default.htm.
8. International Conference on Harmonization. ICH S9. International Conference on Harmonization (ICH) of technical requirements for registration of pharmaceuticals for human use: nonclinical evaluation for anticancer pharmaceuticals; 2010. Available from: http://www.fda.gov/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/default.htm.
9. Chapman K, Pullen N, Graham M, Ragan I. Preclinical safety testing of monoclonal antibodies: the significance of species relevance. Nature 2007;6:120–126.
10. Douthwaite J, Jermutus L. Exploiting directed evolution for the discovery of biologics. Curr Opin Drug Discov 2006;8:268–275.
11. Lynch CM, Hart BW, Grewal IS. Practical considerations for nonclinical safety evaluation of therapeutical monocloncal antibodies. mAbs 2009;1(1):2–11.
12. Cavagnaro JA. The principals of ICH S6 and the case-by-case approach. In: Cavagnaro JA, editor. Preclinical Safety Evaluation of Biopharmaceuticals. Hoboken, NJ: Wiley; 2008, p. 45–65.
13. Chapman K, Pullen N, Coney L, Dempster M, Andrews L, Bajramovic J, et al. Preclinical development of monoclonal antibodies, considerations for the use of non-human primates. mAbs 2009;1(5):505–516.
14. Wolf JJ, Kaplanski CV, Lebron JA. Nonclinical safety assessment of vaccines and adjuvants. In: Davies G, editor. Vaccine Adjuvants, Methods in Molecular Biology. New York, NY: Springer; 2010, p. 29–40.
15. World Health Organization: WHO guidelines on nonclinical evaluation of vaccine. WHO/BS/03. Geneva, Switzerland: World Health Organization; 2005.
16. European Medicines Agency. Note for guidance on preclinical pharmacological and toxicological testing of vaccines. EMEA. CPMP/SWP/465/95; 1997.
17. Marshall V, Baylor NW. Food and Drug Administration regulation and evaluation of vaccines. Pediatrics 2011;127(Suppl):S22–S30.
18. Junutula JR, Flagella KM, Graham RA, Parsons KL, Ha E, Raab H, et al. Engineered thio-trastuzumab-DM1 conjugate with an improved therapeutic index to target HER2-positive breast cancer. Clin Cancer Res 2010;16(19);4769–4778.
19. Viola G, Bortolozzi R, Hamel E, Moro S, Brun P, Castagliuolo I, et al. MG-2477, a new tubulin inhibitor, induces autophagy through inhibition of the Akt/mTOR pathway and delayed apoptosis in A549 cells. Biochem Pharmacol 2011;83:16–26.

20. Sapra P, Hooper AT, O'Donnell CJ, Gerber HP. Investigational antibody drug conjugates for solid tumors. Expert Opin Invest Drug 2011;20(8):1131–1149.
21. Carter PJ, Senter PD. Antibody-drug conjugates for cancer therapy. Cancer J 2008;14:154–169.
22. Kim KM, McDonagh CF, Westendorf L, Brown LL, Sussman D, Feist T, et al. Anti-CD30 diabody–drug conjugates with potent antitumor activity. Mol Cancer Ther 2008;7 (8):2486–2497.
23. Erickson HK, Widdison WC, Mayo MF, Whiteman K, Audette C, Wilhelm SD, et al. Tumor delivery and *in vivo* processing of disulfide-linked and thioether-linked antibody–maytansinoid conjugates. Bioconjug Chem 2010;21(1):84–92.
24. Erickson HK, Park PU, Widdison WC, Kovtun YV, Garrett LM, Hoffman K, et al. Antibody–maytansinoid conjugates are activated in targeted cancer cells by lysosomal degradation and linker-dependent intracellular processing. Cancer Res 2006;66(8):4426–4433.
25. Wen DY, Boissel JR, Tracy TE, Gruninger RH, Mulcahy LS, Czelusniak J, et al. Erythropoietin structure–function relationships: high degree of sequence homology among mammals. Blood 1993;5:1507–1516.
26. Zavitz KH, Bartel PL, Nobden AN. Drug targets, target identification, validation and screening. In: Wermuth's The Practice of Medicinal Chemistry. St. Louis: Elsevier; 2008.
27. Muller PY, Brennan FR. Safety assessment and dose selection for first-in-human clinical trials with immunomodulatory monoclonal antibodies. Clin Pharmacol Ther 2009;85 (3):247–258.
28. Descotes J. Immunotoxicity of monoclonal antibodies. mAbs 2009;1(2):104–111.
29. Horvath CJ, Milton MN. The TeGenero incident and the Duff Report conclusions: a series of unfortunate events or an avoidable event? Toxicol Pathol 2009;37:372–383.
30. Molloy ES. PML and rheumatology: the contribution of disease and drugs. Cleve Clin J Med 2011;78(Suppl 2):S28–S32.
31. Jason-Moller L, Murphy M, Bruno JA. Overview of Biacore systems and their applications. Curr Protoc Protein Sci 2006;19:13.1–19.13.
32. Brennan FR, Morton LD, Spindeldreher S, Kiessling A, Allenspach R, Hey A, et al. Safety and immunotoxicity assessment of immunomodulatory monoclonal antibodies. mAbs 2010;2(3):233–255.
33. Clarke J, Hurst C, Martin P, Vahle J, Ponce R, Mounho B, et al. Duration of chronic toxicity studies for biotechnology-derived pharmaceuticals: is 6 months still appropriate? Regul Toxicol Pharmacol 2008;50:2–22.
34. Bussiere JL, Martin P, Horner M, Couch J, Flaherty M, Andrews L, et al. Alternative strategies for toxicity testing of species-specific biopharmaceuticals. Int J Toxicol 2009;28 (3):230–253.
35. Sharma A, Davis CB, Tobia LA, Kwok DC, Tucci MG, Gore ER, et al. Comparative pharmacodynamics of keliximab and clenoliximab in transgenic mice bearing human CD4. J Pharmacol Exp Ther 2000;293(1):33–41.
36. Bussiere JL. Species selection considerations for preclinical toxicology studies for biotherapeutics. Expert Opin Drug Metab Toxicol 2008;4(7):871–877.
37. Wu B, Joshi A, Ren S, Ng C. The application of mechanism-based PK/PD modelling in pharmacodynamic-based dose selection of muM17, a surrogate monoclonal antibody for efalizumab. J Pharm Sci 2006;95:1258–1268.

38. Clarke J, Leach W, Pippig S, Joshi W, Wu B, House R, et al. Evaluation of a surrogate antibody for preclinical safety testing of an anti-CD11a monoclonal antibody. Regul Toxicol Pharmacol 2004;40:219–226.
39. European Medicines Agency. Guideline on adjuvants in vaccines for human use (EMEA/CHMP/VEG/134716/2004); 2005. Available from: http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC500003809.pdf.
40. Food and Drug Administration. Guidance for industry: considerations for developmental toxicity studies for preventive and therapeutic vaccines for infectious disease indications; 2006. Available from: http://www.fda.gov/BiologicsBloodVaccines/GuidanceComplianceRegulatoryInformation/Guidances/Vaccines/ucm074827.htm.
41. Food and Drug Administration. Guidance for industry: considerations for plasmid DNA vaccines for infectious disease indications; 2007. Available from: http://www.fda.gov/BiologicsBloodVaccines/GuidanceComplianceRegulatoryInformation/Guidances/Vaccines/ucm074770.htm.
42. European Medicines Agency. Note for guidance on pharmaceutical and biological aspects of combined vaccines; 1998. Available from: http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC500003939.pdf.
43. ICH S7a. Safety pharmacology studies for human pharmaceuticals; 2000. Available from: http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Safety/S7A/Step4/S7A_Guideline.pdf.
44. ICH S7B. The non-clinical evaluation of the potential for delayed ventricular repolarization (qt interval prolongation) by human pharmaceuticals; 2005. Available from: http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Safety/S7B/Step4/S7B_Guideline.pdf.
45. Cavero I. Cardiovascular system assessment best practices: a Safety Pharmacology Society meeting. Expert Opin Drug Saf 2010;9(5):855–866.
46. Vargas HM, Bass AS, Breidenbach A, Feldman HS, Gintant GA, Harmer AR, et al. Scientific review and recommendations on preclinical cardiovascular safety evaluation of biologics. J Pharmacol Toxicol Methods 2008;58:72–76.
47. Chapman KL, Andrews L, Bajramovic JJ, Baldrick P, Black LE, Bowman CJ, et al. The design of chronic toxicology studies of monoclonal antibodies: implications for the reduction in use of non-human primates. Regul Toxicol Pharmacol 2012;62(2):347–354.
48. Ponce R, Abad L, Amaravadi L, Gelzleichter T, Gore E, Green J, et al. Immunogenicity of biologically-derived therapeutics: assessment and interpretation of nonclinical safety studies. Regul Toxicol Pharmacol 2009;54:164–182.
49. Morford LL, Bowman CJ, Blanset DL, Bøgh EB, Chellman CJ, Halpern WG, et al. Preclinical safety evaluations supporting pediatric drug development with biopharmaceuticals: strategy, challenges, current practices. Birth Defects Res B Dev Reprod Toxicol 2011;92:359–380.
50. Bowman CJ, Martin PL. Developmental and reproductive safety assessment of biopharmaceuticals. In: RD Hood, editor. Developmental and Reproductive Toxicology: A Practical Approach, 3rd ed. London, UK: Informa Healthcare.
51. Martin PL, Breslin W, Rocca M. et al. Considerations in assessing the developmental and reproductive toxicity potential of biopharmaceuticals. Birth Defects Res B Dev Reprod Toxicol 2009;86:176–203.

52. Pentšuk N, van der Laan JW. An interspecies comparison of placental antibody transfer: new insights into developmental toxicity testing of monoclonal antibodies. Birth Defects Res B Dev Reprod Toxicol 2009;86:328–344.
53. Garty BZ, Ludomirsky A, Danon YL Peter JB, Douglas SD. Placental transfer of immunoglobulin G subclasses. Clin Diagn Lab Immunol 1994;1(6):667–669.
54. Kane SV, Acquah LA. Placental transport of immunoglobulins: a clinical review for gastroenterologists who prescribe therapeutic monoclonal antibodies to women during conception and pregnancy. Am J Gastroenterol 2009;104:228–233.
55. Nakazaw T, Kurokawa M, Kimura K, Wakata A, Hisada S, Inoue T, et al. Safety assessment of biopharmaceuticals: Japanese perspective on ICH S6 guideline maintenance. J Toxicol Sci 2008;33(3):277–282.
56. Leach MW, Halpern WG, Johnson CW, Rojko JL, MacLachlan TK, Chan CM, et al. Use of tissue cross-reactivity studies in the development of antibody-based biopharmaceuticals: history, experience, methodology, and future directions. Toxicol Pathol 2010;38 (7):1138–1166.
57. Danilenko DM, Wang H. The yin and yang of immunomodulatory biologics: assessing the delicate balance between benefit and risk. Toxicol Pathol 2012;40(2):272–287.
58. Wang W, EQ Wang EQ, Balthasar JP. Monoclonal antibody pharmacokinetics and pharmacodynamics. Clin Pharmacol Ther 2008;84(5):548–558.
59. Gonda I. Systemic delivery of drugs to humans via inhalation. J Aerosol Med 2006;19 (1):47–53.
60. Lampson LA. Monoclonal antibodies in neuro-oncology. Getting past the blood–brain barrier. mAbs 2011;3(2):153–160.
61. Roopenian DC, Akilesh S. FcRn: the neonatal Fc receptor comes of age. Nat Rev Immunol 2007;7:715–725.
62. Ward ES, Ober RJ. Chapter 4: Multitasking by exploitation of intracellular transport functions the many faces of FcRn. Adv Immunol 2009;103:77–115.
63. Kuo TT, de Muinck EJ, Claypool SM, Yoshida M, Nagaishi T, Aveson VG, et al. N-glycan moieties in neonatal Fc receptor determine steady-state membrane distribution and directional transport of IgG. J Biol Chem 2009;284:8292–8300.
64. Food and Drug Administration. Guidance for industry estimating the maximum safe starting dose in initial clinical trials for therapeutics in adult healthy volunteers; 2005. Available from: http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/ucm078932.pdf.
65. DeGeorge JJ, Ahn C, Andrews PA, Brower ME, Giorgio DW, Goheer MA, et al. Regulatory considerations for preclinical development of anticancer drugs. Cancer Chemother Pharmacol 1998;41:173–185.
66. Expert Group on Phase One Clinical Trials (chairman: Professor Gordon W. Duff). Expert Group on Phase One Clinical Trials: Final report. The Stationary Office; 2006.
67. European Medicines Agency. Guideline on strategies to identify and mitigate risks for first-in human clinical trials with investigational medicinal products, EMEA/CHMP/SWP/28367/07; 2007. Available from: http://www.emea.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC500002988.pdf.
68. Genetics Home Reference. Handbook: Gene Therapy. Experimental techniques, safety, ethics, and availability. U.S. National Library of Medicine. Available from: http://ghr.nlm.nih.gov/handbook/therapy?show=all.

69. Schalk JA, Mooi FR, Berbers GA, van Aerts LA, Ovelgönne H, Kimman TG. Preclinical and clinical safety studies on DNA vaccines. Hum Vaccines 2006;2(2):45–53.
70. Glenting J, Wessels S. Ensuring safety of DNA vaccines. Microb Cell Fact 2005;4:26.
71. Grimm D, Kay MA. Therapeutic application of RNAi: is mRNA targeting finally ready for prime time? J Clin Invest 2007;117:3633–3641.
72. Goldring CEP, Duffy PA, Benvenisty N, Andrews PW, Ben-David U, Eakins R, et al. Assessing the safety of stem cell therapeutics. Cell Stem Cell 2011;8(6):618–628.

# Clinical Assessment of Biologic Agents

LESLEY ANN SAKETKOO, SHIKHA MITTOO, and LUIS R. ESPINOZA

## 3.1 INTRODUCTION

Biologic agents have become part of a powerful armament in the treatment of immune-mediated diseases, although long-term data are still being collected. It appears that biologics may have favorably altered outcomes in mortality and disability in recalcitrant and life-threatening diseases. At the same time, these "miraculous" agents come with important and potentially life-threatening concerns related to immune suppression that can give rise to serious infection, heart failure, demyelinating neural disease, and malignancy. Therefore, continued assessment of efficacy and adverse effects is central to the management of patients treated with biologics. This chapter outlines important considerations in the clinical assessment during and after a biologic clinical trial. This chapter covers structural aspects of clinical trial design and surveillance for adverse effects, provides an overview of endpoints in clinical trials design, and discusses the basis for generic (not disease-specific) and disease-specific instruments (i.e., outcome measures, endpoints) used in trials. This chapter also provides information on novel forms of developing clinical endpoints from the patient's perspective.

## 3.2 CLINICAL TRIAL STRUCTURE

### 3.2.1 Study Objective and Hypothesis Testing

Every investigation begins with a question in the mind of an investigator. This question is then transformed into a statement, according to scientific method, called a *hypothesis*. The hypothesis of the study is the foundation of the study design and drives the overall objective of the study. For example, an investigator may wonder if

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*,
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

drug A has better efficacy in controlling disease X than placebo; the hypothesis would be "drug A has better efficacy in controlling disease X than placebo." Here are some other examples: drug X has the same efficacy as drug Y but has less toxicity than drug Y, the test dose of drug X at 10 mg is more effective than 5 mg of drug X, and drug X at 10 mg is as safe as drug X at 5 mg. There is an unlimited number of hypotheses that exist or can exist. A study is designed with utility and efficiency in mind to gain data that will be useful on the performance of an intervention (in the case of biologics, administration of a drug). These data can then be tested against the hypothesis.

When there is a hypothesis, there is always a *null hypothesis*, which states the contrary to the hypothesis. The investigators set out to reject the null hypothesis. In our very first example, this would be "drug A does not have better efficacy in controlling disease X than placebo." Often, two types of hypotheses are tested when comparing one treatment arm with another–either that the two treatments have similar efficacy (*non-inferiority*) or that one treatment has a superior effect compared with the other treatment.

In addition to the hypothesis, other discoveries may be made during the course of the investigation. These discoveries may be unanticipated but important qualities of the drug, or there may be discoveries regarding the behavior of the disease in question; this is especially true if the disease is a rare disease. These findings may spark further investigations either within the same trial or in the context of a new investigation.

### 3.2.2 Standard Phases of Clinical Trials

Clinical trials involving medications that are seeking agency approval (e.g., US Food and Drug Administration [FDA], European Medicines Agency [EMA]) traditionally move through "phases." There are four phases; however, it typically requires passage of only the first three phases for the drug for application to the approval process. The phases of clinical trials investigate the study drug in a stepwise fashion. As the study drug's safety and efficacy are proven, each subsequent phase augments the number of research subjects to sequentially include more subjects with each phase.

*Phase I* is usually performed in a much smaller number of patients (~50 patients) than the other phases. It is designed to understand the pharmacokinetic effects in humans; to identify issues related to safety, toxicity (ability of patients to tolerate the drug), and side effects; and to understand a safe dosing range. A *phase II* clinical trial has a patient cohort that is larger, perhaps five to 10 times the number (~200) of a phase I trial. A phase II trial usually tests the efficacy of the study drug in different dosages to determine safe and efficacious doses for a particular disease or diseases. A *phase III* clinical trial is a confirmatory trial that attempts to verify the drug's safety and efficacy, which were demonstrated in the smaller trials, in a much larger group of patients (~1000 patients). A *phase IV* clinical trial is usually conducted to gain further understanding of the properties and performance qualities of the medicine in regard to dosing, risks, benefits, and so on. A phase IV trial is the only phase that is not necessary before agency approval of the study drug is sought.

### 3.2.3 Types of Studies

The gold standard of study design is the *prospective, double-blind, randomized, placebo-controlled clinical trial*. This has four components that are each considered the gold standard of each category:

1. The study collects data over the forward movement of time (rather than investigating data that were collected in the past). This is referred to as a *prospective* trial. Collecting data prospectively ensures that all expected points of data are able to be collected and in a uniform standardized method. This is the highest level of data collection. *Retrospective* data collection, whereby investigators attempt to answer a question or hypothesis by looking back in time, is fraught with many pitfalls because the data collection is rarely standardized at the points of collection and reporting. Data collected retrospectively are often culled from incomplete documentation. The sources of the data may be multiple or unknown reporters, rendering a proportion of the data difficult to interpret.
2. At all levels of the investigation, especially regarding the investigator and the patient, there is *blinding* (or *masking*) to whether the patient is receiving the study drug (or which dose of the study drug if multiple doses are used). A double-blinded study reduces bias from both parties that could potentially affect their responses in regard to endpoints and how they interact with each other. For example, if a physician is aware of whether or not a patient is receiving the study drug, the physician may unwittingly behave differently or speak in a different tone or phraseology that may have a positive or negative effect on the patient's response. It may also affect the physician's perspective of the patient's health and affect the results of data reporting, especially instruments such as the physician global score of disease activity. For many trials, complex efforts are undertaken to extend blinding to include the monitoring safety agency, recruitment personnel, and the statistician.
3. The patients are *randomized* to whether or not they will receive the study drug or what dose of the study drug will be received. Patients are randomly assigned without the cognizance of the investigating team assigned to one of the groups. This is often done via a computerized mechanism. Randomization reduces several levels of bias.
4. The study drug is compared with "no drug" in the form of a *placebo* that is delivered and looks exactly like the study drug. Humans have a tendency to respond to an intervention even when it is an "empty" intervention. Using a placebo group helps to control for the effect of this. A "no drug" placebo is often not ethical, as discussed later in this chapter.

Sometimes in rare diseases in which it is difficult to assemble a large cohort, for instance, it is not always feasible to perform a randomized, double-blind, placebo-controlled clinical trial. The following types of trials should be considered in contrast to the above "gold standard" components.

***3.2.3.1 Prospective Open-Label Observational Clinical Trial*** In this type of investigation, both the investigator and the patient are aware of the drug being administered. The data are collected over the forward movement of time, which allows for quality control of the data being collected, but both patient and investigator biases are introduced into the study, potentially affecting the results. It is commonly accepted that prospective open-label studies overestimate the efficacy of a study drug. This type of study, if the results are positive, may be the gateway to then conducting the much more expensive and labor-intensive randomized, double-blind, placebo-controlled clinical trial.

***3.2.3.2 Retrospective Analysis*** This type of study investigates how patients in the past responded to a drug. In this type of study, charts are reviewed. The data that were collected and documented by the treating physician at that point in time are often the only data available. There is often little uniformity of the data collected across patients. Often, there is much important data missing from the notes of the physicians in regard to confirmation of diagnosis, criteria that establish disease severity, and accurate documentation by the physician of quantifiable change of disease over time. In addition to challenges related to the physician's documentation, tests that would normally be obtained in a prospective study are often incomplete for any number of reasons (e.g., cost to patient, noncompliance of patient with testing, physician forgetting to order a test, lack of standardized reporting or interpretation of the test). It is very difficult to be certain of the quality of the data represented in these types of studies. Last, bias is inherent on many levels.

## 3.3 INCORPORATING PATIENTS INTO THE STRUCTURE OF A TRIAL

### 3.3.1 Treatment Groups and Cohort Enrichment

For a clinical trial, one must define the appropriate patients for inclusion. A group of patients to be studied is called a *cohort*. After diagnosis is confirmed by using strict and standard diagnostic criteria, it is important to identify patients with active disease who are likely to respond to an efficacious intervention; this is often referred to as *cohort enrichment*.

It is of great importance to understand that although a patient may have a diagnosis of a disease, the disease may not actually be active in that patient or not severe enough to warrant treatment. As an example, some autoimmune diseases may "burn out," and although the patient may have residual damage from a disease process that was once active (e.g., loss of the kidney's normal function in glomerular nephritis or stark abnormalities of the hands in rheumatoid arthritis [RA]), the disease is no longer progressive or active. Diseases also carry various rates of progression and severity of aggression. Sometimes a slow-moving disease may only warrant minimal intervention and observation.

For this reason, a minimal criterion of inclusion is established that defines the level of disease activity sufficient for inclusion in a particular clinical trial. A

minimal criterion for inclusion often targets patients with moderate to severe disease; however, some clinical trials are designed for patients with mild active disease. Patients may also be stratified by the severity of their disease into several groups within one clinical trial. Other stipulating descriptions of the cohort, depending on the hypothesis of the study, might be that patients are either naïve to treatment, have failed nonbiologic therapy, or perhaps have failed treatment on another biologic. (Treatment failure may be defined as either a lack of efficacy or patient intolerability to the treatment.)

An important component of a cohort that needs to be carefully defined before initiation of the clinical trial is a description of the exclusion criteria. Exclusion criteria are designed to exclude potential mimics of the target disease that would confound the results. Patients may also have other conditions or diseases in addition to the disease being studied. These are called *comorbid conditions*. It may be necessary to exclude patients with certain comorbid diseases such as renal failure, heart failure, lung disease, and so on. The natural progression of the patient's comorbid disease could potentially confound the interpretation of the safety results of the chosen clinical endpoints, or in certain instances, the study drug could potentially harm patients (e.g., in a patient with renal failure with a study drug that is known to rely on renal excretion). As a result of such stringent inclusion and exclusion criteria, inferences and conclusions from clinical trials are not likely representative of "real-world" clinic patients, patients who are more heterogeneous in their clinical disease activity and course. Often, phase IV trials provide further data in this arena.

### 3.3.2 Comparative Groups Within Clinical Trials

The most basic construct of comparison is the *treatment group versus control group*. To be able to quantify the efficacy and safety of an intervention, the performance of an intervention is compared against a *control group* that receives either a placebo or another medicine. Different treatment groups within a clinical trial are often referred to as *arms*, and frequently there are more than just two arms in a clinical trial (Fig. 3.1).

It is true that double-blind, randomized, placebo-controlled trials are the gold standard among the types of studies. However, it may not be ethical to use a placebo-only design in a comparative trial that could potentially result in harm, disability, or death to a patient who does not receive some form of efficacious treatment. In these cases, a number of modifications may be instituted with nonbiologic or other biologic therapy that is known to be efficacious in treating the disease in question. Additionally, if there is no promise of some potentially efficacious treatment (whether the new promising drug or the standard medicine), physicians are less likely to enroll very sick patients into a clinical trial. This concept would introduce a *selection bias* of patients who have less severe disease because physicians will not take the risk of losing a treatment opportunity in a patient with aggressive and progressive disease.

As mentioned earlier, there may be multiple "arms" within a clinical trial. The intervention group may have a number of comparative arms built into the trial design,

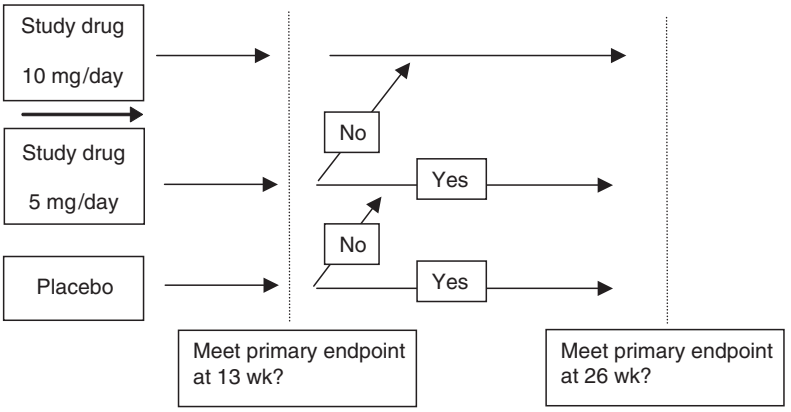


**FIGURE 3.1** An example of a cross-over study. In this example, the patient not meeting an endpoint over a specified period of time "crosses over" to the group with the medicine or with a higher dose of the study drug.

the most common being patients comparatively receiving different doses of the same medication. These groups may also be compared with an additional non-intervention group. Furthermore, clinical trials with multiple arms may provide the ability of the control or one of the comparative treatment arms to *cross over* if a clinical trial endpoint (see Fig. 3.1) has not been met within a specified period of time. A *cross-over design* builds into the design an opportunity for the research subject to receive the study drug or the study drug at a higher dose. Cross-over designs provide three important mechanisms to an investigation. First, cross-over studies allow patients with uncontrolled disease the opportunity to advance to a potentially more efficacious treatment if the assigned substance does not appear to be working. Second, cross-over designs provide an added dimension to the research, which allows for yet another opportunity to demonstrate potential superior efficacy with use of the study drug or with use of the study drug at a higher dose. Third, to continually allow patients in the less efficacious arms to cross over to the potentially more efficacious arm sequentially augments the final sample size of the potentially effective arm.

## 3.4 EFFICACY

### 3.4.1 Endpoints

Clinical trials are designed to meet clinical endpoints. Endpoints are a single measurement or combination of measurements that are thought to accurately reflect the current state of the activity of the disease compared with the same measurements obtained in the past or at future times. Endpoints must be able to reflect *sensitivity* and *accuracy to change over time*. For example, measuring joint deformity caused by

irreversible damage in RA is unlikely to show improvement of disease even if the disease has been effectively treated. However, swollen joints are a temporary state of inflammation; when the disease is controlled, joints are no longer swollen, or they may even worsen quantifiably if not treated. The number of swollen joints can be measured and compared over time. In this case, the endpoint is a physical examination finding, but it can be any parameter that is measured and reported at discrete points over time, such as laboratory values; scores on questionnaires; quantifiable changes on imaging; the decreasing dosage of another medication such as steroids; and even decreased number of events such as visits to the emergency department or hospital admissions, which are sometimes measured in asthma.

Qualities of a good endpoint are that it is *easy to interpret*, it *performs reliably* with repeated measurements, and it *directly measures what it is intended to measure*. Additionally, feasibility in regard to patient safety, cost, and availability of the test are important considerations [1–4].

Before detailing primary and secondary endpoints, it may be important to clarify what is a *surrogate marker*. A surrogate marker is a marker or endpoint that can be useful in measuring the activity of a particular disease because the marker is recognized to reflect disease activity; however, there is no guaranteed relationship that changes in the surrogate marker directly reflect disease activity. For example, serologic acute phase reactants, such as erythrocyte sedimentation rate (ESR) and C-reactive protein (CRP), are markers of inflammation and therefore rise in relation to inflammatory diseases such as RA and Crohn's disease. However, these markers may also be elevated in other situations such as infection. Ultimately, the change in levels of acute phase reactants is a peripheral phenomenon of the disease, not necessarily the action of the disease itself.

***3.4.1.1 Primary Endpoints*** The primary endpoint of a clinical trial is the defining measurement that the investigators agree on before the trial begins that will determine efficacy of the clinical trial. The primary endpoint is sometimes called the *primary outcome*. It is chosen because it is believed to be the single best reflection of disease activity for that particular disease in that particular clinical trial. As in RA, there may be a choice of several primary endpoints from which the investigators may choose. It may be a single measure or a combined response index that contains several measurements as part of a formula that provides a single result. An example of this is the ACR (American College of Rheumatology) 20 (see later discussion), which combines features of a physical examination, laboratory values, the patient's and the physician's overall perception of disease activity, and a score that reflects a patient's reported physical ability (may also be called *function*) (see later discussion for full details).

Additionally, a primary endpoint may be specified to be reached at a specified period of time within the duration of the clinical trial. For example, a trial may last for 1 year, but the primary endpoint is specified to be reached at week 26 with continued performance evaluation for the remainder of the year. Potential disease-specific primary endpoints are discussed for each disease below.

***3.4.1.2 Secondary Endpoints*** Secondary endpoints are important measures that are incorporated into a clinical trial to help investigators understand the intervention's impact on other areas of disease or the patient's life. These secondary endpoints may demonstrate the intervention's effect on parameters specifically related to the disease itself or to effects of the study medication on the patient. Secondary endpoints might be related to quality of life; assessment of fatigue, depression, or functioning; or the patient's or the physician's overall assessment of disease activity. Although we think of these endpoints as a secondary measure of disease activity, they can also be an indication of side effects from the study drug. For example, if patients in the treatment arm are showing worsening fatigue or mental health scores that are discordant with the trend of the primary endpoint, it could be a signal that these symptoms are related to the study drug itself.

An overview of generic (i.e., not disease-specific or validated across several different diseases) assessment tools most often used as secondary endpoints is provided in the following sections.

*Mental Health* Several instruments measure disease and medication effects on mental health. A few examples of the most commonly used instruments to measure depression and anxiety are described here.

The *Center for Epidemiology Studies—Depression Scale (CES-D)* is a 20-item questionnaire composed of short phrases that are scored on a scale of 0 to 3. The questionnaire asks about the patient's feelings and involvement in activities that might indicate depression.

The *Hospital Anxiety and Depression Scale (HADS)* is a 15-item questionnaire that has a depression component and an anxiety component that can be separated out if the investigator desires.

*Sleep* The *Medical Outcomes Study—Sleep Scale (MOS-SS)* is the most commonly used sleep scale. It is a 12-item questionnaire that asks about sleep disturbance, difficulty falling and staying asleep, feeling as though one has gotten enough sleep, and ability to stay awake during the day. It is important to remember that disease manifestations such as pain as well as medicines can affect sleep; prednisone is a good example.

*Pain* The Visual Analogue Scale (VAS) is the most commonly used method to assess pain (Fig. 3.2). It is often incorporated into other instruments such as the Health Assessment Questionnaire (HAQ).

*Physical Function* The HAQ and the HAQ-DI (Disability Index) are paper or sometimes electronic questionnaires given to the patient asking questions about physical ability and functioning in daily life as well as pain and the patient's perception of his or her health, both of which are done on a 15-mm VAS that the patient puts a mark through. A score is totaled at the end. To view the HAQ or the HAQ-DI, go to http://aramis.stanford.edu/HAQ.html. The HAQ and the HAQ-DI have also been adopted into other instruments such as the RAPID3 (see later discussion) [5,6].

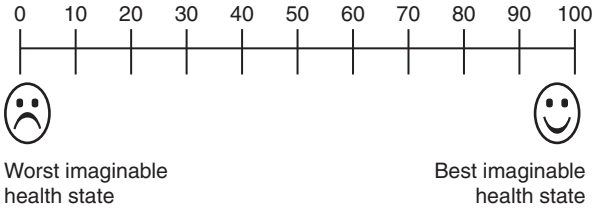


**FIGURE 3.2** Visual analogue scale. The respondent, whether it is the physician or the patient, draws a line through the scale that corresponds to his or her perception.

*Health-Related Quality of Life* The SF-36 (Medical Outcomes Study, Short Form 36) measures health-related quality of life and has several components that measure disease impact on general health, pain, physical ability and function, mental health, energy, effect of work, and social activities. To view the SF-36, go to http://www.sf-36.org/tools/sf36.shtml.

*Worker Productivity* A field of secondary endpoints that has received much attention with great challenges in its measurement is worker productivity. Many diseases treated with biologics cause disability to the extent that work performance, productivity, presence, and imminent job loss (work instability) are affected [7–9]. One of the important goals of disease treatment is to have the patient return to as full a working life as possible. Developing and incorporating measurements that gauge disease effects on remunerative work are complex but are beginning to be introduced into the clinical trial setting [7–9].

***3.4.1.3 Obtaining Patient Perspective in the Development of Clinical Trial Endpoints*** Many of the instruments used in clinical trials for secondary endpoints are "patient-reported" endpoints; this means that the patient is asked to answer questions either as posed by the investigator, on a questionnaire (e.g., the SF-36, HAQ, or mental health scales), or on a VAS (e.g., for pain or perceived global health). These are very important; at one time, clinical trial data were limited to traditional "objective" measures of disease.

*Patient-Reported Outcomes* As discrete descriptors of domains, patient-reported measures are increasingly recognized as equal to or more important than traditional measures of disease activity [10,11]. For example, in patients with idiopathic pulmonary fibrosis, patient-reported dyspnea is recognized as an accurate and robust prognosticator of death [12]. To develop and select valid patient-reported outcome measures, disease experts need to consider incorporating the perspectives of patients because these perspectives supply crucial information. The reasoning behind obtaining the patient perspective supports that patients are a type of expert in the disease they are experiencing, and they see the disease from a different lens than that of medical experts. An example of this is fatigue in the clinical assessment of RA. Focus groups with

patients who have RA were conducted. It was through this forum whereby patients discussed what was important to them in relation to their disease that a new clinical endpoint was identified—fatigue. Fatigue was validated as an accurate and reliable reflection of disease activity that is sensitive to change over time [13]. Fatigue is now a mandatory endpoint in clinical trial design for RA.

Patients as consultants and advisors included early in the development and the process of clinical trials design and implementation is recommended by patient advocacy groups, investigators, and organizations such as OMERACT (Outcome Measures in Rheumatology), the World Health Organization (WHO), the US FDA, and the EMA [2–4,10,11,14]. We now realize there is much important information to be learned from the inclusion of patients' points of view.

## 3.5 PHARMACOVIGILANCE

*Pharmacovigilance* is the term that relates to scientific strategies to detect, assess, gain pathophysiological knowledge of, and ultimately prevent drug-related problems that arise in the postmarketing phase of a drug. Pharmacovigilance is driven by patient care and patient safety. It essentially includes monitoring for adverse outcomes, with the aim of ultimately providing clear information to health agencies and consumers developing risk–benefit profiles to aid in the most appropriate management decisions. Pharmacovigilance begins with surveillance that occurs at many levels, including the international (e.g., the WHO), national, and even institutional levels. Many countries have their own biologic registries and have been able to determine trends in certain types of malignancies, serious infections, and lung and heart involvement over short- and long-term periods after availability of a biologic.

### 3.5.1 Adverse Side Effects

There are many adverse side effects, which may be considered as follows: (1) common, but nuisance, side effects; (2) uncommon, but clinically relevant, side effects that include paradoxical adverse events; and (3) serious adverse effects. The three types of adverse effects from biologic clinical trial data are discussed in this section.

Before initiation of most treatment with biologics, whether for clinical trials or clinical practice, patients must be tested for tuberculosis (TB), hepatitis B, and hepatitis C. It is wise to get a history of recent travel as an indication of whether to screen further for the presence of latent infectious diseases [15].

Clinical trials for biologics usually monitor for changes in kidney function; liver function with alanine aminotransferase, aspartate aminotransferase, and bilirubin; lipid profile; and bone marrow suppression, which could be suggested by a decreased white blood cell count, anemia, and platelets. As a matter of course, electrocardiograms are often obtained periodically (or at least at baseline) to monitor for the development of cardiac conduction abnormalities (arrhythmias). Chest radiography is often obtained at baseline. The monitoring physical examination for adverse effects includes changes in weight, blood pressure, and cardiac

and pulmonary examination as well as examination for the development of edema and skin changes.

Common side effects are defined as effects that occur with a frequency of greater than 5% in the treated population. Symptoms experienced by patients are often nonspecific (e.g., gastrointestinal [GI] distress, headache), reversible, and self-limiting and are often experienced by both the treatment and placebo groups (if experienced by the placebo group, it is referred to as the *placebo effect*). Attributing an adverse effect to the experimental treatment requires statistically comparing the symptoms experienced by the treatment group with the those experienced by the placebo group. Examples of common side effects significantly associated with biologic therapies are injection site reactions and upper respiratory symptoms.

***3.5.1.1 Paradoxical Adverse Events*** Potentially unique effects of biologics are paradoxical adverse events. Biologic therapies for RA are designed to block upregulation or augmentation of select proteins that are involved in a complicated cascade of events leading to a particular outcome, such as joint inflammation. The proteins or cellular events that are antagonized often have multiple potentially opposing roles. In the case of tumor necrosis factor α (TNF-α), the role of protein may differ by tissue subtype or by stage of disease (in the case of cardiac failure) or may be expressed differently in different organ systems (e.g., there is less expression in skin). As a result of these variations and the fact that a variety of other relevant cytokine and protein networks are at play at any given moment, paradoxical adverse events may arise. As an example, TNF-α may promote or inhibit pulmonary fibrosis. By definition, TNF antagonism is suppression of inflammation and autoimmune consequences, yet it has also been the trigger of autoimmunity, being implicated in development of lupus-like reactions [16]. TNF inhibition in clinical trials has demonstrated efficacy in improving psoriasis, yet rare, paradoxical events of palmopustular psoriasis have been reported among treated patients [17].

***3.5.1.2 Serious Adverse Events*** The approval of the TNF inhibitor infliximab in August 1998 was preceded by only one case of TB being identified. The immediate months to follow saw a wave of TB cases reported to the FDA in striking temporal association with anti–TNF-α use with greater than 50% of patients presenting with disseminated and extrapulmonary disease involving lymph tissue, bone, bladder, bowel, and meninges [15]. After this devastating revelation, other serious infections also became apparent during the postmarketing period involving life-threatening bacterial, fungal, and parasitic infections. As an example, reactivation of JC (John Cunningham) virus is uniquely reported among patients who have received rituximab. Such surveillance and adverse outcome reporting are parts of an ongoing evaluation after a clinical trial demonstrates efficacy and safety and after agency approval. In the United States, all adverse events should be reported to the Adverse Events Reporting System of the FDA. Other countries maintain "biologic registries" that track many aspects of biologic use, including adverse events.

The importance of biologic registries across different ethnic populations cannot be underscored because certain ethnic populations may have predilections to

serious adverse events compared with other ethnicities. For example, the prevalence of methotrexate- and leflunomide-related pulmonary toxicity is reportedly higher among the Japanese registries compared with North American cohorts. Although it is recommended that clinical trials incorporate patients of diverse ethnic and social backgrounds, it is not always feasible. Moreover, because of the expense and logistics of conducting clinical trials, many trials are now being conducted in developing or select European communities where the socioeconomic fabric may differ significantly from the populations at risk. As a result, the inferences drawn from clinical trials may not necessarily be applicable to countries outside of those where the trials are conducted. Additionally, the location where a clinical trial is conducted may adversely impact ethnic health disparities and raise important ethical concerns. Although recruiting for marginalized ethnic groups has been a challenge in North American trials, this process provides the possibility to assess the outcome differences in various subgroups. The importance of clinical trials as it relates to ethnicity was highlighted in a landmark clinical trial of antihypertensive medication in African American patients. Angiotensin-converting enzyme inhibitors or angiotensin receptor blockade therapy was less effective in this ethnic group compared with whites. In the same vein, African American women with systemic lupus erythematosus are at increased risk for renal disease compared with whites, and North American Native American populations with RA are more likely to have higher disease activity and greater severity or burden of disease compared with whites with the same condition [18]. It is ethically important to be able to offer promising therapeutic options to patients who may have the greatest need but also to test the efficacy and impact of these therapies in these subgroups.

A continual surveillance for malignancy and serious infections is essential in the clinical assessment of biologic therapy. It is important to remember that although postmarketing surveillance is an important way to uncover harmful signals that may not be apparent in a shorter term clinical trial involving fewer patients or to uncover rare but clinically relevant adverse effects (e.g., TB), there is a potential reporting bias. This means that events should be interpreted cautiously, and one should also consider the lack of compliance with such monitoring systems, with events also potentially being underreported.

## 3.6 DISEASES TREATED WITH BIOLOGICS

Wide ranges of autoimmune inflammatory disorders are treated with biologics. In general, these autoimmune inflammatory disorders are characterized by perturbations in both innate and adaptive immune functions and probably are triggered by unknown exogenous or endogenous antigens. This section provides an overview of some of the diseases treated with biologic agents. This is not a comprehensive list, but it provides a background to help better understand the utility of endpoints in clinical trials for biologics.

### 3.6.1 Rheumatoid Arthritis

Rheumatoid arthritis is a disease of systemic inflammation that most obviously affects the joints with inflammation and destruction. However, RA can affect many organ systems, including the eyes, lungs, heart, and skin, with inflammation and vasculitis. Anti-TNF agents were the first widely used biologics in the treatment of RA. However, RA now has a number of FDA-approved biologics with different working mechanisms that include B-cell depletion with rituximab; interference of T-cell co-stimulation of cell interactions necessary for immune response with abatacept; and inhibition of interleukins IL-1 or IL-6 with anakinra and tocilizumab, respectively. Other biologics are in the pipeline and include small molecules such as syk kinase inhibitors JAK (Janus kinase) and STAT (signal transducer and activator of transcription).

Rheumatoid arthritis has been carefully studied for endpoints. However, no single component has been identified as being a pure reflection of disease activity in RA, although the following components are agreed to be important: the number of swollen joints, the number of tender joints, the duration of morning stiffness, the patient's report of global disease activity, pain, fatigue, progression of joint damage and erosions of bones as seen on radiographs, blood "surrogate" biomarkers reflective of inflammation such as ESR and CRP, physical function, and the broader evaluation of health-related quality of life.

The following are some of the disease-specific indices for RA commonly used as endpoints in clinical trials for RA.

In the ACR 20, ACR 50, and ACR 70, the number in the title corresponds to the percent of reduction in both tender and swollen joints and in at least three of five component parameters: pain by VAS, physician or patient global assessment of disease activity by VAS, function (by some accepted method such as HAQ), or the acute phase response (ESR or CRP). For example, an ACR 20 response would be a patient who at the beginning of the trial had 20 swollen joints, 25 tender joints, a pain scale score of 8, an HAQ score of 0.8, a physician global score of 6, and a patient global score of 8 with an ESR of 80 and after treatment was found to have a 20% improvement in the number of tender and swollen joints and improvement in at least three other parameters but no changes in physician global and HAQ. In this example, the patient has had large change in several areas, and although he or she has met the endpoint of ACR20, the HAQ and physician global scores have not changed much; however, this might be because the patient might have continued joint involvement in the hands, which would interfere greatly with functioning. This example illustrates the inherent challenges in these composite indices: although this patient has met this endpoint, many physicians would not consider that he or she has been treated satisfactorily.

The Disease Activity Score 28 (DAS28) is a composite endpoint; the 28 refers to the number of joints that are examined. This composite score is "weighted," which means that certain variables weigh more heavily than others. The formula is complex, involving mathematical applications such as square roots and logarithms to produce a score. In this case, the DAS28 gives more weight to the blood markers of inflammation (acute phase reactants) ESR and CRP. There is a concern that such

outcome measures are too reliant on this "surrogate" marker of disease activity, resulting in an overinterpretation of improvement [19].

Some disease activity composites do not include acute phase reactants. The Clinical Disease Activity Index (CDAI) is an unweighted calculation of the number of tender joints and swollen joints based on a count of 28 joints performed by the physician as well as patient global assessment and physician global assessment based on a 10-point scale distributed over a 150-mm line. The Simplified Disease Activity Index (SDAI) incorporates the same clinical measures as the CDAI but also includes the unweighted addition of the CRP level.

In contrast to the above activity scores, the Routine Assessment Patient Index Data (RAPID3) provides a fast (<10 seconds), feasible, informative quantitative index in patients with RA, and RAPID scores correlate significantly with the DAS28 and CDAI.

### 3.6.2 Osteoporosis and Malignant Bone Metastasis

Osteoporosis and malignant bone metastasis are diseases that are not directly immune mediated but are an indication for treatment with a biologic called denosumab, which is a monoclonal antibody against RANK ligand (receptor activator to nuclear factor κB). This ligand reduces the proliferation and activity of osteoclasts, which are central effectors of bone breakdown leading to osteopenia. RANK ligand may also have a role to play in RA, but at this time, there is no indication for treatment in RA. Endpoints to assess for efficacy in osteoporosis are changes in bone mineral density comparatively in the lumbar spine and the hip and a reduction in fractures of the treatment group compared with the control group. For metastasis, endpoints are a reduction in pain intensity, reduction in sites of pain, and reduction of new sites of bony metastasis as evaluated on imaging.

### 3.6.3 Inflammatory Bowel Disease

Inflammatory bowel disease (IBD), such as Crohn's disease and ulcerative colitis (UC), is an inflammatory and immune cell infiltration of the mucosa of the GI tract. The introduction of biologics seems to have decreased the need for surgical bowel removal and sequelae of IBD such as persistent bleeding, fistulas (nonhealing holes along the tract that allow fecal matter to communicate with areas of the body that should be protected from such). In UC, an immune-mediated inflammatory disease of the small bowel and colon, clinical trial endpoints are often a combination of the degree of bleeding, the frequency of stools, changes in the appearance of the mucosa on endoscopic examination, and the physician's global assessment of disease activity. These are most commonly assessed by the Mayo Score, which evaluates these properties on a 4-point scale that includes none or normal, mild, moderate, and severe.

Crohn's disease differs from UC in that it can occur anywhere along the GI tract from the mouth to the anus. Additionally, the immune-mediated effects may occur outside of the immune system involving the eyes, joints, and skin. The clinical trial

endpoint in Crohn's disease has consistently been the Crohn's Disease Activity Index (CDAI), which is a composite of items that are each weighted from a factor of 2 to 30 in a formula to ultimately produce a score. The components of the CDAI include the frequency of diarrhea, abdominal pain, the amount of antidiarrhea medications used, anemia caused by GI blood loss, and the existence of other manifestations thought to be related to the disease (e.g., arthritis, uveitis [inflammation of the eye], skin disease). These are then summed with a "weight factor" that takes into account the change in the patient's body weight, which should be improving if the disease is controlled and may be declining if the disease is active and the patient is unable to absorb the nutrients required to maintain a normal body weight.

### 3.6.4 Psoriasis and Psoriatic Arthritis

Psoriasis and psoriatic arthritis (PsA) are autoimmune-mediated diseases that can lead to joint destruction, disability, social isolation, and depression. In addition to skin lesions, PsA causes arthritis, dactylitis (sausage-like swelling of the digits caused by inflammation of the tendons running along the digit), inflammation, and fusing together of the spinal column and sacroiliac joints of the pelvis. Biologics have proved to have a high impact on recalcitrant disease.

Assessment of biologic agents in PsA include skin scoring and swollen and tender joint counts. The use of RA disease activity scales is acceptable and validated, but these scales do not perform optimally in individuals with PsA. Central to the assessment of PsA are skin assessment, peripheral joint assessment, patient global assessment, pain, physical function, and quality of life [20,21]. Other aspects that could be considered in clinical trial design are enthesitis (inflammation of tendon insertion to bone), dactylitis, spinal involvement, radiologic changes, physician global assessment of disease activity, and acute phase reactants.

Regarding skin involvement, a number of methods may be used to assess the degree of body surface area affected by skin disease.

### 3.6.5 Reactive Airway Disease

Asthma is a disease that is triggered by an outside or foreign antigen; the host's immune response overreacts to the antigenic presence. Omalizumab has been approved for moderate to severe persistent allergic asthma as demonstrated by serum immunoglobulin with a total IgE level greater than 30 IU/ml. Omalizumab is a monoclonal antibody that binds to free IgE, thereby preventing ensuing the binding of IgE (with bound allergen) to mast cells and basophils (important immune cells related to allergic reactions) and averting the allergic cascade that causes the bronchoconstriction of lungs resulting in an asthma attack.

Endpoints that have been specific to asthma in measuring efficacy of biologic agents are number of exacerbations, number of hospitalizations, ability to decrease use of inhaled steroids by 50%, and ability to discontinue use of an inhaled steroid. Other endpoints could include breathing tests both done by the patient and in the clinical setting.

## 3.7 FUTURE PERSPECTIVES

With the advent of newer and more powerful immunosuppressive and antiinflammatory drugs including the use of combination therapy, pharmacodynamics and pharmacokinetics are difficult to predict. Therefore, newer strategies including pharmacogenetics and biomarkers are being used to predict metabolic responses to a variety of immunosuppressive drugs to enable a more accurate prediction of their metabolic responses [22–24]. This, however, is an area of intense investigation, which will require large prospective trials with careful validation of data obtained.

## REFERENCES

1. Tugwell P, Bombardier C. A methodologic framework for developing and selecting endpoints in clinical trials. J Rheumatol 1982;9:758.
2. Boers M, Brooks P, Strand CV, Tugwell P. The OMERACT filter for outcome measures in rheumatology. J Rheumatol 1998;25:198–199.
3. Boers M, Tugwell P, Felson DT, et al. World Health Organization and International League of Associations for Rheumatology core endpoints for symptom modifying antirheumatic drugs in rheumatoid arthritis clinical trials. J Rheumatol Suppl 1994;41:86.
4. Stucki G, Boonen A, Tugwell P, Cieza A, Boers M. The World Health Organisation International Classification of Functioning, Disability and Health: a conceptual model and interface for the OMERACT process. J Rheumatol 2007;34(3):600–606.
5. Felson DT, Anderson JJ, Boers M, Bombardier C, et al. American College of Rheumatology preliminary definition of improvement in rheumatoid arthritis. Arthritis Rheum 1995;38:727.
6. Felson DT, Anderson JJ, Boers M, et al. The American College of Rheumatology preliminary core set of disease activity measures for rheumatoid arthritis clinical trials. The Committee on Outcome Measures in Rheumatoid Arthritis Clinical Trials. Arthritis Rheum 1993;36:729.
7. Gilworth G, Chamberlain MA, Harvey A, Woodhouse A, Smith J, Smyth MG, et al. Development of a work instability scale for rheumatoid arthritis. Arthritis Rheum 2003; 49(3):349–354.
8. Wolfe F, Michaud K, Choi HK, Williams R. Household income and earnings losses among 6,396 persons with rheumatoid arthritis. J Rheumatol 2005;32:1875–1883.
9. Osterhaus JT, Purcaru O, Richard L. Discriminant validity, responsiveness and reliability of the rheumatoid arthritis-specific work productivity survey (WPS-RA). Arthritis Res Ther 2009;11:R73.
10. Bottomley A, Jones D, Claassens L. Patient-reported outcomes: assessment and current perspectives of the guidelines of the Food and Drug Administration and the reflection paper of the European Medicines Agency. Eur J Cancer 2009;45(3):347–353.
11. Speight J, Barendse SM. FDA guidance on patient reported outcomes. BMJ 2010; 340:c2921.

12. Martinez FJ, Safrin S, Weycker D, Starko KM, Bradford WZ, King TE, Jr, et al.; IPF Study Group. The clinical course of patients with idiopathic pulmonary fibrosis. Ann Intern Med 2005;142(12 Pt 1):963–967.
13. Minnock P, Kirwan J, Bresnihan B. Fatigue is a reliable, sensitive and unique outcome measure in rheumatoid arthritis. Rheumatology 2009;48(12):1533–1536.
14. Saketkoo LA, Matteson EL, Brown KK, Seibold JR, Strand V. Developing disease activity and response criteria in connective tissue disease related interstitial lung disease. J Rheumatol 2011;38(7):1514–1518.
15. Saketkoo LA, Espinoza LR. Impact of biologic agents on infectious diseases. Infect Dis Clin North Am 2006;20(4):931–961, viii.
16. Clowse ME. The use of anti-TNFα medications for rheumatologic disease in pregnancy. Int J Womens Health 2010;2:199–209.
17. Cuchacovich R, Espinoza CG, Virk Z, Espinoza LR. Biologic therapy (TNF-alpha antagonists)–induced psoriasis: a cytokine imbalance between TNF-alpha and IFN-alpha? J Clin Rheumatol 2008;14:353–356.
18. Peschken CA, Hitchon CA, Robinson DB, Smolik I, Barnabe CR, Prematilake S, et al. Rheumatoid arthritis in a North American native population: longitudinal followup and comparison with a white population. J Rheumatol 2010;37(8):1589–1595.
19. Smolen JS, Aletaha D. Interleukin-6 receptor inhibition with tocilizumab and attainment of disease remission in rheumatoid arthritis: the role of acute-phase reactants. Arthritis Rheum 2011;63:43.
20. Ritchlin CT, Kavanaugh A, Gladman DD, Mease PJ, Helliwell P, Boehncke WH, et al.; Group for Research and Assessment of Psoriasis and Psoriatic Arthritis (GRAPPA). Treatment recommendations for psoriatic arthritis. Ann Rheum Dis 2009;68(9):1387.
21. Kavanaugh A, Ritchlin C, Boehncke WH. Quality indicators in psoriatic arthritis. Clin Exp Rheumatol 2007;25(6 Suppl 47):98.
22. Ware N. The role of genetics in drug dosing. Pediatr Nephrol 2012;27(9):1489–1498.
23. Schlickeiser S, Sawitzki B. Peripheral biomarkers for individualizing immunosuppression in transplantation—regulatory T cells. Clin Chim Acta 2012;413(17-18):1406–1413.
24. Cravedi P, Heeger PS. Immunologic monitoring in transplantation revisited. Curr Opin Organ Transplant 2012;17:26–32.

# Key Regulatory Guidelines for the Development of Biologics in the United States and Europe[1]

RICHARD KINGHAM, GABRIELA KLASA, and KRISTA HESSLER CARVER[2]

## 4.1 INTRODUCTION

The United States and the European Union have distinct but overlapping schemes for the regulation of biologics, ranging from the definition of a biologic itself to the technical requirements for approval. In the United States, the definition of "biological product" was developed over time, and historical context continues to inform its interpretation. In the European Union, biologics are largely defined in terms of their active substances and methods of manufacture. Despite these differences, both jurisdictions recognize that biologics warrant special treatment because of their distinct characteristics, such as their complex structures and susceptibility to variation during manufacturing. Whereas in the United States, Congress enacted a separate statute for biologics, in the EU, the general approval scheme and certain specific requirements apply to biologics. Nevertheless, US and EU authorities have undertaken harmonization efforts with respect to some technical requirements for biologics applications; thus, there is significant overlap in requirements imposed by both regions. This chapter provides an overview of the US and EU regulatory schemes, from nonclinical trials through clinical trials to approval. It then discusses considerations for global development of biologics, and it ends by discussing special issues for developing vaccines.

[1] The chapter uses both the terms "drug" and "medicine" as they are used in the US and EU regulatory schemes, respectively.

[2] Mr. Kingham is a partner at Covington & Burling LLP in Washington, DC, and London; Ms. Carver is an associate in the firm's Washington, DC, office; and Ms. Klasa is an associate in the firm's Brussels office. The views expressed here are solely those of the authors and do not necessarily reflect the views of the firm or its clients.

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*,
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

## 4.2 GENERAL UNITED STATES REGULATORY SCHEME

### 4.2.1 The Definition of "Biological Product" and Its Significance

In the United States, "biological products" are subject to a different premarket pathway and differing intellectual property protections than products regulated only as "drugs."[3] Whereas a biological product must be licensed pursuant to a biologics license application (BLA) showing it is "safe, pure, and potent," the sponsor of a nonbiologic drug must submit a New Drug Application (NDA) showing the drug is safe and effective.[4] Certain new biological products receive 12 years of data protection, but new drugs receive up to 5 years of this protection.[5] Biologic and drug legislation also provide different schemes for resolving patent issues regarding entry of follow-on products.[6] Thus, determining whether a product meets the definition of "biological product" is enormously important.

Unfortunately, this inquiry is not straightforward. Drug and biologic legislation developed separately, and Congress did not provide detailed guidance for distinguishing biologics from other drugs. As a result, the Food and Drug Administration (FDA) and other agencies administering the biologics law over time have made these distinctions mostly on an *ad hoc* basis based on history and relevant precedents.[7] The FDA has recently proposed a bright-line rule for distinguishing "proteins" that qualify for treatment as biologics from certain other products that do not. This approach remains in the proposal stages, however, and history and precedent continue to play important roles in product jurisdiction determinations. This section outlines that relevant history.

In 1902, Congress passed the Biologics Control Act, which applied to "any virus, therapeutic serum, toxin, antitoxin, or analogous product applicable to the prevention and cure of diseases of man" and required licensure of facilities making these products.[8] Over the next hundred-plus years, Congress expanded this list of covered products to include, among other things, the following products and those "analogous" to them: vaccines, blood, blood products, allergenic products, and proteins (except chemically synthesized polypeptides).[9] Despite all of these

[3] Products meeting the "biological product" definition also meet either the definition of "drug" or "device" under the Federal Food, Drug, and Cosmetic Act (FDCA). Intercenter Agreement Between CDER and CBER (effective October 31, 1991) [hereinafter ICA], § II.

[4] Public Health Service Act (PHSA) § 351(a)(2)(C)(i)(I); FDCA § 505(b) & (d).

[5] PHSA § 351(k)(7); FDCA §§ 505(c)(3)(E), 505(j)(5)(F).

[6] PHSA § 351(l); see, e.g., FDCA §§ 505(b)(1), 505(c)(2), 505(c)(3)(A)-(D), 505(j)(2)(A)(vii) and (B), 505(j)(5).

[7] Michael S. Labson, Krista Hessler Carver, and Marie C. Boyd, *FDA Regulation of Biological Products*, in THE PHARM. REGULATORY PROCESS 103, 110 (Ira R. Berry & Robert P. Martin, eds., 2008).

[8] Pub. L. No. 57-244, 32 Stat 728, 728 (1902). Congress added a product licensure requirement in 1944 and then replaced the dual license requirement with a single requirement for a BLA in 1997. Pub. L. No. 78-410, § 351, 58 Stat. 682, 702 (1944); Pub. L. No. 105-115, § 123(a), 111 Stat. 2296, 2323 (1997).

[9] Pub. L. No. 111-148, Title VII, Subtitle A, § 7002 (b)(2), 124 Stat. 119, 814 (2010); Pub. L. No. 105-115, § 123(d), 111 Stat. 2296, 2324 (1997); Pub. L. No. 91-515, § 291, 84 Stat. 1297, 1308 (1970); Pub. L. No. 78-410, 58 Stat. 682, 702 (1944).

amendments, Congress never defined the listed terms and, in particular, never defined "analogous," so the scope of the biological product definition remained unclear.

The overlapping definition of "drug" added to this complexity. The Food and Drugs Act of 1906 and the Federal Food, Drug, and Cosmetic Act of 1938 (FDCA) defined "drug" broadly to include, among other things, substances intended for use in the cure, mitigation, or prevention of disease, and the latter statute mandated submission of an NDA before marketing of a drug.[10] Although these "drug" definitions encompassed many biologics, the statutes did not provide concrete parameters for distinguishing nonbiological drugs from biological products. In 1944, when Congress revised and recodified the 1902 Act in the Public Health Service Act (PHSA), it clarified that the NDA requirement did not apply to biologics, but it did not elucidate the scope of the biological product definition.[11]

Regulators attempted to fill this gap by promulgating regulatory definitions of virus, therapeutic serum, toxin, antitoxin, and analogous product.[12] For example, the 1947 regulations, which are essentially similar to the current regulations,[13] defined products "analogous" to a toxin or antitoxin as those intended for preventing, treating, or curing diseases or injuries "through specific immunization."[14] The 1947 definition of products analogous to therapeutic serums excluded hormones.[15] Hormones such as insulin and human growth hormone were approved under the FDCA, not the PHSA.[16] Despite the 1947 regulations, differentiating biologics from drugs remained challenging at the margins.

[10] Pub. L. No. 59-384 § 6, 34 Stat 768 (1906); Pub. L. No. 75-717 §§ 201(g)(2), 52 Stat. 1040, 1041, 1052-53 (1938).

[11] See Pub. L. No. 78-410 § 351(g), 58 Stat. 682, 703 (1944) ("Nothing contained in this Act shall be construed as in any way affecting, modifying, repealing, or superseding the provisions of the [FDCA]"); H. R. Rep. No. 1364, 78th Cong., at 23 (1944) ("Subsection (g) is an explicit statement, confirming the present legal situation, that products subject to this section are not exempted from the [FDCA], except for the provision of that act relating to the licensing of new drugs").

[12] 42 C.F.R. § 73.1(g) (1949); see, e.g., Treasury Dep't, U.S. Pub. Health & Marine-Hospital Serv., Regulations for the Sale of Viruses, Serums, Toxins, and Analogous Products, Miscellaneous Publication No. 20, ¶ 16 (1909).

[13] 21 C.F.R. § 600.3(h) (2012). The FDA has not updated 21 C.F.R. § 600.3(h) to reflect that the statutory definition of biological product now includes products applicable to the prevention, treatment, or cure of "condition[s]" of human beings. For this reason, one court has concluded that the regulation is invalid "to the extent that [it] purports to eliminate the application... to 'conditions.'" *United States v. Livdahl*, 459 F. Supp. 2d. 1255, 1261 (S.D. Fla. 2005).

[14] 42 C.F.R. § 73.1(g)(5)(iii) (1949).

[15] *Id.* § 73.1(g)(5)(ii) (defining a product as analogous to a serum if it was "composed of whole blood or plasma or containing some organic constituent or product *other than a hormone* or an amino acid, derived from whole blood, plasma, or serum and intended for administration by a route other than ingestion") (emphasis added).

[16] Letter of Steven K. Galson, M.D., M.P.H., Director, CDER, to Kathleen M. Sanzo, Esq., Morgan, Lewis & Bockius LLP; Stephan E. Lawton, Esq., BIO; and Stephen J. Juelsgaard, Esq., Genentech, Docket No. 2004P-0231, PDN1 (May 30, 2006) (2006 Consolidated Response), at 44 & n.82; FDA, Ever Approved Drug Products Listed by Active Ingredient, at 2291 (printout dated August 2, 1989) (obtained via FOIA request) (listing NDA 1-754 for Sterling Drug Inc.'s insulin suspension product, approved November 27, 1939).

The advent of biotechnology, along with agency organizational disputes, brought this issue to the forefront of FDA's focus. In 1986, the FDA issued a policy statement stating that it would determine whether biotechnology products constituted biologics "based on the intended use of each product on a case-by-case basis."[17] Thus, the FDA continued to make product-specific determinations informed by history and precedent, and different units of the FDA had to agree on the approval pathway for a given product. This proved to be difficult, with press reports of "turf battles" between the Center for Drug Evaluation and Research (CDER) and the Center for Biologics Evaluation and Research (CBER) for jurisdiction over blockbuster biotechnology products[18] and claims that the decisions were "inconsisten[t]."[19] For example, epidermal growth factors were regulated as drugs because their first approved indications were traditionally drug indications.[20] Most monoclonal antibodies (mAbs) were licensed as biologics because of their biological source material and immunologic function.[21] Recombinant insulin and human growth hormone, similar to their naturally derived counterparts, were approved pursuant to NDAs.[22]

CDER and CBER subsequently executed an Intercenter Agreement (ICA) that attempted to clarify the governing authorities for products derived from living material. The agreement provided that the following products, among others, were subject to licensure under the PHSA: vaccines; proteins, peptides, and carbohydrates produced by cell culture (other than hormones and products previously derived from human or animal tissue and approved as drugs); proteins made in transgenic animals; blood and blood products; and allergenic products.[23] NDAs were required for, among other things, hormones (regardless of method of manufacture), synthetic mononucleotide and polynucleotide products, and naturally derived products other than vaccines or allergenics.[24] Twelve years later, the FDA consolidated review of most therapeutic proteins in CDER, but this transfer did not modify the governing statutory scheme for any ICA product, and the FDA continued to decide whether new products were biological products or nonbiologic drugs on a case-by-case basis using the principles of the ICA and historical precedent.[25]

In February 2012, the FDA issued draft guidance aimed at implementing recent legislation that added "protein (except any chemically synthesized polypeptide)" to the biological product definition.[26] In this draft guidance, the FDA proposed a

[17] 51 Fed. Reg. 23309, 23310 (June 26, 1986).
[18] *FDA Triage System for Drugs/Biologics Questioned*, U.S. Regulatory Reporter (November 1990), at 3.
[19] *FDA's Handling of Biotech Approvals*, Scrip (August 26, 1988), at 14.
[20] *The FDA and Drug-Biologic Classifications*, U.S. Regulatory Reporter (August 1988), at 6.
[21] *Id.*
[22] Suzanne White Junod, Ph.D., FDA Historian, *Celebrating a Milestone: FDA's Approval of First Genetically-Engineered Product*, Update, September/October 2007, at 43, 44; 2006 Consolidated Response, at 44 and n.82.
[23] ICA, § III(B)(1).
[24] *Id.* § III(A).
[25] 68 Fed. Reg. 38067, 38068 (June 26, 2003).
[26] Patient Protection and Affordable Care Act (PPACA), Pub. L. No. 111-148 § 7002(b)(2), 124 Stat 119, 814 (2010) (amending section 351(i) of the PHSA).

bright-line rule distinguishing proteins from "peptides" and "chemically synthesized polypeptide[s]" that the FDA proposes to approve under the FDCA.[27] The agency proposed to define "protein" as "any alpha amino acid polymer with a specific defined sequence that is greater than 40 amino acids in size."[28] According to the draft guidance, "peptides" have 40 or fewer amino acids and are not "proteins."[29] The agency also proposed to define "chemically synthesized polypeptide" as an alpha amino acid polymer that is made entirely by chemical synthesis and that has fewer than 100 amino acids.[30] Until the draft guidance is finalized, these definitions must be considered proposals. Nevertheless, they signal that the FDA might be shifting from its traditional, ad hoc approach to jurisdictional decisions to a new approach guided by bright-line rules.[31]

### 4.2.2 Nonclinical Studies for Biologics

Similar to other drugs, biologics must undergo laboratory and animal testing to define their pharmacologic and toxicologic effects before they can be studied in humans.[32] The legal framework for preclinical testing of biologics is essentially similar to that for drugs; for example, the FDA's good laboratory practice (GLP) regulations typically apply.[33] Nevertheless, biologics present special issues, necessitating a "flexible, case-by-case, science-based approach" to preclinical testing.[34]

For biotechnology-derived pharmaceuticals, the FDA has adopted the International Conference on Harmonization of Technical Requirements for the Registration of Pharmaceuticals for Human Use (ICH) S6 guidance, which describes the unique

[27] FDA, *Draft Guidance for Industry: Biosimilars: Questions and Answers Regarding Implementation of the Biologics Price Competition and Innovation Act of 2009* 13 (February 2012) (Q&A Draft Guidance). Sponsors of certain transitional proteins will remain eligible to submit their applications as NDAs during a transition period that ends on March 23, 2020. PPACA § 7002(e), 124 Stat. at 817. This option is available if a protein is the same "product class" as a biological product that was approved under the FDCA on or before March 23, 2010. Id. § 7002(e)(2). In the draft guidance, the FDA proposes to consider two products to be in the same product class if they "are homologous to the same gene-coded sequence," with allowances "for additional novel flanking sequences." Q&A Draft Guidance, at 14. Under the draft guidance, two products may be in the same product class even if their differences result in changes in pharmacokinetics, but not if the changes "alter[] a biological target or effect." *Id.* at 14–15.

[28] Q&A Draft Guidance, at 13.

[29] *Id.*

[30] *Id.*

[31] The FDA also has special rules for determining the regulatory status of tissue engineering and cell processing products. See 21 C.F.R. Parts 1270 & 1271. Detailed discussion of these rules is beyond the scope of this chapter.

[32] FDA, *Guidance for Industry: S6 Preclinical Safety Evaluation of Biotechnology-Derived Pharmaceuticals* 3 (July 1997) (S6 Guidance).

[33] 21 C.F.R. Part 58. For some studies using specialized test systems, it may be impossible to fully comply with GLP, but this might not preclude use of the studies to support clinical trial initiation and marketing authorization. S6 Guidance, at 3.

[34] S6 Guidance, at 1.

approach needed to selection of animal species and immunogenicity testing as overarching considerations and outlines typical preclinical testing.[35] Also, in May 2012, the FDA adopted the addendum to that ICH guidance.[36] Because Europe's Committee for Medicinal Products for Human Use (CHMP) approved that guidance nearly a year earlier, in July 2011, the addendum is discussed in Section 3.1, *infra*, but it is now equally applicable in the United States. Therefore, for a full understanding of nonclinical testing standards in the United States, readers should also review Section 3.1 of this chapter.

***4.2.2.1 Relevant Species*** Many biologics cannot be tested in commonly used animal species, such as rats and dogs, because of their biological activity and species- or tissue-specific activity.[37] Instead, sponsors must use a variety of tests, such as *in vitro* binding assays and functional tests, to identify a "relevant species," that is, "one in which the test material is pharmacologically active due to the expression of the receptor or an epitope (in the case of monoclonal antibodies)."[38] Generally, a sponsor should identify two relevant species, although one species may suffice when the product's biological activity is well understood or only one relevant species exists.[39]

In some cases, the chimpanzee—which cannot be sacrificed at the end of the study—is the only relevant species, and in other cases, identifying a relevant species might be impossible.[40] In these situations, the sponsor might need to consider alternative approaches to gathering animal data, such as the use of homologous proteins (which recognize the target protein or epitope in the animal),[41] transgenic animals that express the human receptor, or other animal models.[42]

***4.2.2.2 Immunogenicity*** Many biologics elicit immune responses, which can affect preclinical study results.[43] In some cases, these effects are desired (e.g., with a vaccine), but unwanted immunogenicity might be harmful. Potential undesired effects include neutralizing or prolonging the biologic's activity, forming immune complexes, or cross-reacting with endogenous substances.[44] As a result, sponsors

[35] *Id.* at 1 n.1. This guidance applies to biotechnology-derived proteins and peptides, their derivatives, and products of which they are components. It may also apply to recombinant DNA protein vaccines, chemically synthesized peptides, and plasma-derived products, among other things. It does not apply to allergenic extracts, cellular blood components, conventional bacterial or viral vaccines, DNA vaccines, or cellular and gene therapies. *Id.* at 2.

[36] FDA, *Guidance for Industry S6 Addendum to Preclinical Safety Evaluation of Biotechnology-Derived Pharmaceuticals* (May 2012) as adopted by FDA at 77 Fed. Reg. 29665, 29666 (May 18, 2012).

[37] S6 Guidance, at 4.

[38] *Id.*

[39] *Id.*

[40] Andrea B. Weir, Ph.D., DABT, Pharmacologist/Toxicologist, Office of Drug Evaluation VI, CDER, FDA, *Preclinical Safety Assessment of Therapeutic Proteins and Monoclonal Antibodies*, in BIOLOGICS DEVELOPMENT: A REGULATORY OVERVIEW17, 19 (Mark Mathieu, ed., 3rd ed., 2004).

[41] *Id.*

[42] S6 Guidance, at 4–5.

[43] Weir, *supra* note 40, at 19.

[44] *Id.* at 20.

should obtain necessary samples for antibody testing during repeat-dose toxicity studies and, when interpreting the data, should consider the effects of antibody formation on pharmacokinetics (PK), pharmacodynamics (PD), and adverse events.[45] Detection of antibodies should not prompt termination of a preclinical study unless the immune response neutralizes the biologic's effects in "a large proportion" of the test animals.[46] Finally, sponsors should be aware that animals' immune responses are not indicative of those in humans.[47]

***4.2.2.3 Typical Preclinical Testing*** Sponsors usually must conduct PD studies, such as *in vitro* binding assays and *in vivo* studies that assess the product's pharmacologic activity and define its mechanism of action.[48] Biologics typically undergo single- and repeat-dose toxicity studies using relevant species, as noted earlier.[49] Safety pharmacology studies, which evaluate the product's functional effects on major body systems and specific organs, and local tolerance testing can be done separately or subsumed in toxicity testing.[50]

Sponsors also usually conduct single- and multiple-dose PK and/or toxicokinetic studies to assess absorption, disposition, exposure, and clearance (in particular, antibody-mediated clearance) and explore dose–response relationships.[51] This information is used to predict margins of safety for human studies. Immunogenicity testing might include screening and mechanistic studies, but animal models are not highly predictive of human immunogenicity.[52]

Typical carcinogenicity bioassays are "generally inappropriate" for biologics, although the S6 guidance calls for assessment of carcinogenicity when warranted based on the "duration of clinical dosing, patient population, and/or biological activity."[53] If concern exists regarding carcinogenic potential, the sponsor can consider several approaches to assess risk, including testing in a variety of malignant and normal human cells and further testing in relevant species.[54] According to ICH S6, reproductive and developmental toxicity studies may or may not be recommended, depending on "the product, clinical indication, and intended patient population."[55] Such studies using primate species pose challenges because of these animals' heterogeneous drug responses, high background abortion rate, and low number of offspring.[56]

[45] S6 Guidance, at 6.
[46] *Id.*
[47] *Id.*
[48] Weir, *supra* note 40, at 20; S6 guidance, at 4.
[49] Weir, *supra* note 40, at 21.
[50] S6 guidance, at 7, 10.
[51] *Id.* at 7–8; Weir, *supra* note 40, at 21.
[52] Weir, *supra* note 40, at 22; S6 guidance, at 9.
[53] S6 guidance, at 10.
[54] *Id.*
[55] *Id.* at 9.
[56] Weir, *supra* note 40, at 22.

Because biologics generally degrade into peptides and amino acids, classic biotransformation studies are unnecessary.[57] Genotoxicity studies also usually are not applicable to biotechnology-derived drugs because they are not expected to interact with DNA or chromosomes.[58]

### 4.2.3 Clinical Studies for Biologics

***4.2.3.1 The Investigational New Drug Application*** If a sponsor plans to perform clinical testing of a biologic in the United States, it must first have an investigational new drug application (IND) in effect.[59] An IND generally goes into effect 30 days after the FDA receives it.[60] During this 30-day time period, the FDA reviews the IND for any safety issues and may place a clinical hold on the study if, among other things, it presents an "unreasonable" risk to patients.[61] The IND must contain "[a]dequate" information from preclinical studies, on which the sponsor bases its conclusion that clinical trials are reasonably safe.[62] For well-characterized therapeutic biotechnology products, the IND should describe the product's pharmacologic effects and mechanism of action and provide information on its absorption, distribution, metabolism, and excretion.[63] Sponsors must include a description of the overall investigational plan and a protocol for each planned study; protocols not submitted in the initial IND should be submitted as protocol amendments.[64] The IND also must contain chemistry, manufacturing, and controls information sufficient to allow evaluation of safety.[65] This information is particularly important for many biologics, which may raise concerns because of their impurity profiles or the use of materials with unknown components in their manufacture.[66] The FDA recognizes that sponsors likely will change their manufacturing processes as development progresses. Section 2.3.4, infra, discusses the effects of these changes on product development.

***4.2.3.2 Good Clinical Practices*** Traditionally, the FDA used the phrase "good clinical practices" (GCP) to collectively describe a number of regulations and guidance documents with two overarching goals: (1) to ensure the integrity of data collected in clinical trials and (2) to protect clinical trial subjects.[67] In the mid

[57] S6 guidance, at 8.
[58] *Id.* at 9–10.
[59] 21 C.F.R. § 312.20(a) & (b); 21 C.F.R. § 312.40(b)(1); FDA, *Guidance for Industry: Content and Format of Investigational New Drug Applications (INDs) for Phase 1 Studies of Drugs, Including Well-Characterized, Therapeutic, Biotechnology-Derived Products 2* (November 1995) (Phase 1 Guidance).
[60] 21 C.F.R. § 312.40(b)(1).
[61] 21 C.F.R. § 312.42(b)(i).
[62] 21 C.F.R. § 312.23(a)(8).
[63] Phase 1 Guidance, at 10.
[64] 21 C.F.R. § 312.23(a)(3)(iv) & (6).
[65] 21 C.F.R. § 312.23(a)(7); Phase 1 Guidance, at 4.
[66] Phase 1 Guidance, at 4–5. This Guidance applies only to well-characterized therapeutic biotech products. *Id.* at 2.
[67] *Good Clinical Practices (GCP)*, in BIOLOGICS DEVELOPMENT: A REGULATORY OVERVIEW 115, 115 (Mark Mathieu, ed., 3rd ed., 2004); see also 62 Fed. Reg. 25691, 25692 (May 9, 1997).

1990s, however, the ICH developed a consolidated GCP guideline, known as the E6 guidance, to harmonize standards for clinical study design, conduct, reporting, and recordkeeping.[68] The FDA has adopted this guidance. The agency recommends that sponsors use it when generating data for submission to the agency[69] and has stated that it will deem studies complying with ICH GCP as meeting the FDA's GCP standards.[70] This guidance supplements and clarifies FDA regulations on institutional review boards (IRBs) (21 C.F.R. Part 56), informed consent (Part 50), and clinical studies for drugs and biologics (Part 312). It describes the overarching principles for conducting clinical trials, the responsibilities of various parties involved with the clinical trial (IRB, sponsor, investigator), and the necessary documents for conducting a clinical study (e.g., the study protocol and investigator's brochure).[71] Sponsors should consider it in combination with the above-cited regulations, more recent FDA regulations (such as Part 54 on financial disclosures for clinical investigators), and more recently released FDA guidance on specific GCP topics.

***4.2.3.3 Study Design Considerations*** As with new drugs, clinical development of biologics typically involves three phases, which may or may not overlap. Biologics present several unique clinical considerations, however. Often, their clinical development programs must include an assessment of immunogenicity, which is typically not an issue for small molecule drugs. Also, because many biologics treat serious or life-threatening illnesses, their development may be compressed.[72]

Phase 1 studies involve the "initial introduction" of the biologic into a small number of humans to assess the product's metabolism, pharmacology, and safety at escalating doses.[73] Unlike phase 1 trials for nonbiologic drugs, phase 1 studies of biologics frequently involve administration to patients rather than healthy volunteers who will not derive benefit from them to ensure the risk–benefit profile of the product is acceptable for ethical purposes.[74] For example, studies may enroll patients when the biologic is known or suspected to be "unavoidably toxic," when there is a risk of antibody development to a native protein or mAb, or when the product's bioactivity is disease specific.[75] Phase 1 studies should determine the maximum tolerated dose and assess the product's bioactivity and PK to determine the optimum biological dose.[76] With respect to immunogenicity, these studies should assess subjects' antibody

[68] 62 Fed. Reg. at 25692.

[69] *Id.*; FDA, *Guidance for Industry: E6 Good Clinical Practice: Consolidated Guidance* (E6 Guidance) 1 (April 1996).

[70] *Good Clinical Practices (GCP)*, *supra* note 67, at 118.

[71] See generally E6 Guidance.

[72] See Barbara G. Matthews, M.D., M.P.H., *Clinical Testing of Biologically Derived Therapeutics*, in BIOLOGICS DEVELOPMENT: A REGULATORY OVERVIEW 79, 84 (Mark Mathieu, ed., 3rd ed., 2004).

[73] 21 C.F.R. § 312.21(a)(1) & (2).

[74] FDA, *Guidance for Industry, Estimating the Maximum Safe Starting Dose in Initial Clinical Trials for Therapeutics in Adult Healthy Volunteers* 3 (July 2005) (MRSD Guidance); Matthews, *supra* note 72, at 83.

[75] MRSD Guidance, at 3; Matthews, *supra* note 72, at 83.

[76] Matthews, *supra* note 72, at 81–82, 83.

development, both directly after administration and at least 28 days thereafter, and determine whether these results are linked to adverse events, PK, or PD.[77]

Phase 2 trials are controlled studies that evaluate short-term adverse events and effectiveness for a specific use in several hundred patients.[78] These studies further explore exposure–response relationships and the biologic's PK, PD, and immunogenicity, and they provide information to help refine the phase 3 protocol, including with respect to size, population, and endpoints.[79] Biologics sponsors often combine phase 2 studies with phase 1 or phase 3 studies.[80]

Phase 3 studies enroll patients and provide primary evidence for labeling claims and risk–benefit assessment.[81] They are larger than phase 2 studies, but their size ranges considerably depending on the patient population and the availability of alternative therapies.[82] They typically are randomized, double blinded, controlled, and performed at multiple study centers.[83] Placebo controls are used when ethical considerations permit, but in some cases (e.g., when effective treatment is already available and withholding treatment would expose subjects to unreasonable risks), active controls are used.[84] The studied patient population, as defined in the protocol's exclusion and inclusion criteria, should be representative of the population for which the sponsor seeks approval.[85]

Endpoint selection is critical to a successful phase 3 trial. The endpoint must demonstrate clinical benefit in the intended patient population.[86] Ideally, the endpoint is an established clinical outcome measure, although validated surrogate endpoints may be used in some cases.[87] If the endpoint is not well defined, the sponsor might have to combine use of several effectiveness outcomes.[88]

Several alternatives to traditional endpoints are available. Under the accelerated approval scheme, eligible sponsors may obtain approval based on either (1) a surrogate endpoint that is "reasonably likely to predict clinical benefit" or (2) a clinical endpoint that is measurable earlier than irreversible morbidity or mortality and that is "reasonably likely to predict an effect on irreversible morbidity or mortality or other clinical benefit."[89] The evidence supporting an endpoint's

[77] *Id.*
[78] 21 C.F.R. § 312.21(b).
[79] Matthews, *supra* note 72, at 84–85, 90–92.
[80] *Id.* at 84.
[81] 21 C.F.R. § 312.21(c).
[82] See *id.*; Matthews, *supra* note 72, at 80.
[83] Matthews, *supra* note 72, at 85–86.
[84] *Id.*; FDA, Inside Clinical Trials: Testing Medical Products in People, http://www.fda.gov/Drugs/ResourcesForYou/Consumers/ucm143531.htm (last accessed November 29, 2012).
[85] Matthews, *supra* note 72, at 85.
[86] *Id.*
[87] *Id.* at 85–86.
[88] *Id.* at 85.
[89] FDCA § 506(c)(1)(A). In July 2012, Congress amended the criteria for accelerated approval as part of the Food and Drug Administration Safety and Innovation Act (FDASIA). Pub. L. No. 112-144 §§ 901-902, 126 Stat. 993, 1082-1088 (2012). The amended statutory provisions are discussed here. The FDA has not yet updated its accelerated approval regulations to reflect these statutory changes. See 21 C.F.R. Part 601, Subpart E.

likelihood of predicting clinical benefit may include epidemiologic, pathophysiologic, therapeutic, pharmacologic, or other evidence developed using biomarkers or other scientific tools.[90] To be eligible for accelerated approval, a medicine must be intended for a serious or life-threatening disease or condition.[91] The FDA may require that the sponsor of an accelerated approval product conduct post-approval studies to "verify and describe the predicated effect on irreversible morbidity or mortality or other clinical benefit."[92]

Another alternative, the "animal rule," applies only to biologics that address serious or life-threatening conditions caused by "exposure to lethal or permanently disabling toxic biological, chemical, radiological, or nuclear substances" and for which human efficacy studies are unethical and infeasible.[93] Under the animal rule, a sponsor may obtain approval based on human safety data and adequate and well-controlled animal studies show the product is "reasonably likely to produce clinical benefit in humans" if, among other things, available data allow selection of an effective human dose.[94] Sponsors obtaining approval under the animal rule must commit to conduct postmarket clinical studies to confirm clinical benefit.[95]

***4.2.3.4 Manufacturing Process Changes During Development*** Sponsors often change the manufacturing process of biologics before approval (e.g., to scale up from pilot production to full-scale manufacturing, to improve manufacturing efficiency, or to change the production facility).[96] Biologics are much more sensitive to process changes than chemically synthesized drugs, and process changes have the potential to adversely affect a biological product. As a result, the FDA will determine whether the sponsor must conduct additional studies to support licensure of the postchange biological product. The FDA has issued guidance that describes this inquiry.[97]

According to this guidance, the FDA deems pre- and postchange products to be "comparable" when testing shows that "the manufacturing change does not affect safety, identity, purity, or potency."[98] The FDA may require that the sponsor perform some combination of analytical testing, *in vitro* and *in vivo* biological assays, animal

[90] FDCA § 506(c)(1)(B).
[91] FDCA § 506(c)(1)(A).
[92] FDCA § 506(c)(2)(A).
[93] 21 C.F.R. § 601.90.
[94] 21 C.F.R. § 601.91; see also FDA, *Draft Guidance for Industry: Animal Models—Essential Elements to Address Efficacy Under the Animal Rule* 16 (January 2009).
[95] 21 C.F.R. § 601.91(b)(1).
[96] FDA, *Guidance Concerning Demonstration of Comparability of Human Biological Products, Including Therapeutic Biotechnology-derived Products* § II (April 1996) (Comparability Guidance). The FDA has also adopted ICH guidance regarding quality issues in comparability determinations. FDA, Q5E Comparability of Biotechnological/Biological Products Subject to Changes in their Manufacturing Process (June 2005).
[97] Comparability Guidance § II.
[98] *Id.*

tests (PK, PD, and/or toxicity), and clinical testing (pharmacology, immunogenicity, safety, and/or efficacy studies) to make this showing.[99] In determining the type and amount of data necessary to support a specific change, the FDA will consider the extent of manufacturing changes, the stages of manufacturing at which they occur, the profile of the prechange product, the ease of characterizing the product, and the results of the necessary tests.[100] Data from in-process assays at the affected phase of manufacture can be particularly important to this determination.[101]

For example, the FDA might require human pharmacology tests to assess changes—such as those in product formulation—that could affect PK or PD. If the change results in differences in structure, bioactivity, or PK that "are meaningful with respect to potential impact on the product's safety, purity, or potency," the FDA generally requires additional clinical study of the product's safety, effectiveness, or both.[102] The same is true when "analytical and other preclinical testing is not sufficiently sensitive or broad enough to detect such meaningful differences."[103] In contrast, changes in filling site might only require comparative data as to final release specifications and stability.[104] Because the FDA has broad discretion in formulating the testing requirements for a comparability showing, the agency encourages sponsors to consult with it before implementing the proposed process change.[105] In any event, the sponsor must fully describe the change in the IND or BLA.[106]

***4.2.3.5 Meetings with the Food and Drug Administration Before and During the Clinical Trial Period*** Sponsor meetings with the FDA are often critical to successful, streamlined clinical development. They can help sponsors avoid conducting expensive clinical studies that the agency would ultimately reject. The FDA has promulgated regulations and issued guidance on FDA–sponsor meetings. Except when otherwise noted, agreements reached at these meetings do not bind the agency; the FDA's regulations provide that action on meetings does not constitute final agency action.[107]

Sponsors can obtain several types of pre-approval meetings with the FDA. 21 C.F.R. § 312.82 describes two types of such meetings. First, under this regulation, the sponsor can seek a pre-IND meeting to reach agreement with the FDA on the design of preclinical studies needed to commence human testing.[108] Although the regulation notes that these meetings are available to sponsors of therapies for life-threatening or

[99] *Id.* § III.
[100] *Id.*
[101] *Id.*
[102] *Id.*
[103] *Id.*
[104] *Id.*
[105] *Id.* § IV.
[106] *Id.* § III.
[107] 21 C.F.R. § 10.65(a).
[108] 21 C.F.R. § 312.82(a).

severely debilitating illnesses,[109] other sponsors can also obtain them. For example, the FDA has indicated that pre-IND meetings can be useful in various challenging development scenarios (e.g., when the sponsor has identified a concerning safety signal).[110] Second, after phase 1 data are available for a therapy intended for life-threatening or severely debilitating illnesses, the sponsor may meet with the FDA to reach agreement on phase 2 study design, "with the goal that such testing will . . . provide sufficient data . . . to support a decision on its approvability for marketing."[111]

Sponsors also may request End of Phase 2 (EOP2) meetings and pre-BLA meetings.[112] At EOP2 meetings, the FDA aims to evaluate whether proceeding to phase 3 is safe, assess the protocols for phase 3, and determine whether any other information will be needed to approve the product.[113] The FDA will record any agreements reached in minutes and transmit them to the sponsor, and studies conducted in accordance with an agreement will be "presumed sufficient" for approval unless a "significant scientific development . . . requires otherwise."[114] Pre-BLA meetings are intended to identify "any major unresolved problems," discuss analysis of the data, and agree on the approach to formatting the BLA.[115]

The FDA classifies its meetings with sponsors into three categories. Type A meetings relate to "an otherwise stalled product development plan."[116] These include certain meetings to discuss clinical holds, meetings to resolve disputes with the FDA, and meetings to discuss the FDA's evaluation of Special Protocol Assessments (SPAs), described later.[117] Pre-IND, EOP1, EOP2, and pre-BLA meetings typically are Type B meetings.[118] All other meetings are Type C meetings.[119] Type A and Type C meetings have the shortest and longest time lead times to meeting dates, respectively.[120] FDA guidance and regulations describe the recommended content of the meeting request and pre-meeting briefing package, procedures for pre-meeting communication to narrow the topics for discussion (if possible), and the FDA's

[109] 21 C.F.R. § 312.82.

[110] FDA, *Small Business Assistance: Frequently Asked Questions on the Pre-Investigational New Drug* (IND) Meeting, http://www.fda.gov/Drugs/DevelopmentApprovalProcess/SmallBusinessAssistance/ucm069906.htm (last accessed November 29, 2012).

[111] 21 C.F.R. § 312.82(b).

[112] 21 C.F.R. § 312.47(b).

[113] 21 C.F.R. § 312.47(b)(1)(i).

[114] 21 C.F.R. § 312.47(b)(1)(v). The FDA also holds End of Phase 2A (EOP2A) meetings, which occur after completion of phase 1 trials and the first set of clinical exposure-response studies, to discuss dose selection for later trials and methods (e.g., trial design, modeling, or simulation) for improving quantification of exposure–response data from early trials. FDA, *Guidance for Industry: End-of-Phase 2A Meetings* 1, 4 (September 2009).

[115] 21 C.F.R. § 312.47(b)(2).

[116] FDA, *Guidance for Industry: Formal Meetings Between the FDA and Sponsors or Applicants* 2 (May 2009).

[117] *Id.*

[118] *Id.* at 3.

[119] *Id.*; FDA, *Guidance for Industry: End-of-Phase 2A Meetings*, at 5.

[120] FDA, *Guidance for Industry: Formal Meetings Between the FDA and Sponsors or Applicants*, at 3.

practices for documenting the substance of the meeting in official meeting minutes and for resolving disputes concerning these minutes.[121]

Sponsors also may request an SPA and accompanying meeting. Pursuant to Section 505(b)(5) of the FDCA and FDA guidance, the FDA must grant "reasonable written request[s]" to meet with BLA sponsors to achieve agreement on the design and size of the following kinds of trials: (1) clinical trials intended to form the primary basis of an efficacy claim, (2) animal carcinogenicity protocols, and (3) final product stability protocols.[122] The agency must reduce to writing any agreements made through this SPA process and make them part of the administrative record.[123] The FDA cannot deviate from any such agreement unless the sponsor consents or the FDA identifies a significant scientific issue essential to determine the drug's safety or efficacy.[124]

### 4.2.4 The Biologics License Application

***4.2.4.1 Contents of the Biologics License Application*** Unlike the drug regulations, which specify the required contents of an NDA in great detail, the regulation on BLA content is quite brief. Under 21 C.F.R. § 601.2, the BLA must contain, among other things, nonclinical and clinical data showing the biologic's safety, purity, and potency; a "full description of manufacturing methods" for the product; stability data substantiating the expiration date; product samples and a summary of test results for the lot from which they derived; proposed labeling, enclosures, and containers; and the addresses of manufacturing facilities.[125] Although this regulation is far less prescriptive than its counterpart in the NDA regulations,[126] the FDA expects BLAs to contain essentially the same information and data as NDAs, and the electronic Common Technical Document (eCTD) format is the FDA's standard format for both.[127] The FDA's approach thus accords with Congress' 1997 directive that the agency "shall take measures to minimize differences in the review and approval of products required to have approved [BLAs and NDAs]."[128]

***4.2.4.2 Food and Drug Administration Review*** The Prescription Drug User Fee Act (PDUFA), which applies to most innovative biologics,[129] and the FDA's good

[121] *Id.* at 4–5, 7–10; see also 21 C.F.R. § 10.65(e).
[122] FDCA § 505(b)(5)(B); FDA, *Guidance for Industry: Special Protocol Assessment* 2 (May 2002).
[123] FDCA § 505(b)(5)(C).
[124] *Id.*
[125] 21 C.F.R. § 601.2(a).
[126] 21 C.F.R. § 314.50.
[127] See, e.g., FDA, Electronic Common Technical Document (eCTD), http://www.fda.gov/Drugs/DevelopmentApprovalProcess/FormsSubmissionRequirements/ElectronicSubmissions/ucm153574.htm (last accessed December 9, 2012).
[128] Pub. L. No. 105–115, § 123(f), 111 Stat. at 2324.
[129] FDCA § 735(1) (excluding whole blood, blood components for transfusion, allergenic extracts, and certain other biologics from the definition of "human drug application" for purposes of PDUFA).

review management principles and practices (GRMPs) govern agency review of BLAs.[130] Pursuant to PDUFA, the FDA levies "user fees" to defray part of its costs from reviewing applications and commits to performance goals for its review of those applications through a letter to Congress.[131] PDUFA sunsets every 5 years and was reauthorized for the fifth time in July 2012 (PDUFA V).[132]

After a sponsor submits a BLA, the agency assembles a review team comprising reviewers focusing on varying disciplines, such as clinical and toxicology issues.[133] The FDA then decides, within the first 60 days after submission, whether it can "file" the application (i.e., whether the BLA contains all information needed to permit a substantive review).[134] The FDA may issue a refuse-to-file decision if the BLA does not meet this threshold. During the filing period, the FDA will also decide whether to designate the BLA as a "priority" or "standard" application. For CBER-regulated applications, "priority" means that the biologic would constitute a significant improvement in the safety or effectiveness of treatment, diagnosis, or prevention of a serious or life-threatening disease.[135] In CDER, the same basic standard applies except that treatments for nonserious diseases also are eligible.[136]

In connection with PDUFA V, the FDA committed to "review and act on"—that is, issue an action letter for—90% of priority, original BLAs within 6 months of the 60-day filing date; the goal date for standard BLAs is 10 months from the 60-day filing date.[137] In the past, the FDA's actual performance has varied, with the agency meeting its PDUFA performance goals in some years but not others.[138] As part of its PDUFA V performance goals, the FDA committed to establish a new review model, known as "the Program," for original BLAs and certain NDAs. The Program offers

[130] FDA, *Guidance for Review Staff and Industry: Good Review Management Principles and Practices for PDUFA Products* 2 (April 2005) (GRMP Guidance) ("The principles of this guidance also pertain, in general, to all preapproval reviews of NDAs, BLAs, and efficacy supplements").

[131] See generally FDCA §§ 735, 736, 736B; FDA, *PDUFA Reauthorization Performance Goals and Procedures Fiscal Years 2013 Through 2017*, available at: http://www.fda.gov/downloads/ForIndustry/UserFees/PrescriptionDrugUserFee/ucm270412.pdf (last accessed November 28, 2012) (PDUFA V Performance Goals).

[132] FDASIA, Pub. L. No. 112–144, Title I, 126 Stat. at 996–1002.

[133] GRMP Guidance, at 11.

[134] 21 C.F.R. § 601.2(a); CBER, SOPP 8404: *Refusal to File Procedures for Biologic License Applications* (Version #3) (August 27, 2007); see also CDER, *New Drug Evaluation Guidance Document: Refusal to File* (1993); CDER, *21st Century Review Process Desk Reference Guide* 16–20 (Version September 2012) (Desk Reference Guide).

[135] CBER, SOPP 8405: *Complete Review and Issuance of Action Letters* (Version #4) (September 20, 2004).

[136] *Id.*; CDER MaPP 6020.3: *Review Classification Policy: Priority (P) and Standard (S)* 2 (February 2012).

[137] PDUFA V Performance Goals, § I(E).

[138] Compare, e.g., FDA, *FY 2005 Performance Report to the President and Congress for the Prescription Drug User Fee Act* 11(indicating that, during fiscal year 2004, the FDA reviewed and acted on 97% of priority and standard applications within the specified time frames) with FDA, *FY 2009 Performance Report to the President and Congress for the Prescription Drug User Fee Act* 19 (indicating that, during fiscal year 2008, the FDA reviewed and acted on 85% of standard applications and 68% of priority applications within the specified time frames).

the opportunity for additional communication with the agency, including in a new "late-cycle" meeting.[139]

After FDA has begun substantive review of the BLA, reviewers may issue information request (IR) and discipline review (DR) letters to the applicant.[140] IR letters ask for specific information while review is in progress.[141] Reviewers issue DR letters at the end of a particular discipline review "to convey early thoughts on possible deficiencies." These letters do not necessarily reflect the input of their supervisors.[142] IR and DR letters do not stop the review clock.[143] Applicants may respond to IR and DR letters with additional information. This type of submission might constitute a "major amendment" to the application; if so, the FDA might extend the PDUFA goal date.[144]

Next, the FDA might seek an advisory committee's (AC's) advice on the application. By statute, the FDA must refer an original BLA to an AC or explain, in the action letter for the application, why that step was not taken.[145] The FDA typically requests that the AC address particular questions and vote on the answers, but the AC's advice does not bind the FDA.[146]

After the agency completes its review of the BLA, it will issue an approval letter, or a complete response letter (CRL), which states that the agency cannot approve the BLA in its current form.[147] A CRL lists identified deficiencies and, when possible, recommends sponsor actions to place the BLA in a position for approval.[148] An applicant may file a "resubmission" to address the deficiencies.[149] The review timeline for a resubmission depends on its content but is either 2 or 6 months from receipt.[150] An applicant also can request resolution of any dispute concerning the CRL.[151] If the FDA denies approval of the application, the applicant may request, and the Commissioner must issue, a notice of opportunity for hearing.[152]

***4.2.4.3 Approval Standard*** The FDA must approve a BLA if it shows that the proposed product is "safe, pure, and potent" and the facilities where the product

[139] PDUFA V Performance Goals, § II.
[140] See generally FDA, *Guidance for Industry: Information Request and Discipline Review Letters Under the Prescription Drug User Fee Act* (November 2001).
[141] *Id.* at 2.
[142] *Id.* at 2–3.
[143] *Id.* at 3.
[144] CDER, MaPP 6010.8: *NDAs and BLAs: Communication to Applicants of Planned Review Timeline*s 3 (June 2008); CBER, SOPP 8402: *Designation of Amendments as Major* (Version #4) (October 2012).
[145] FDCA § 505(s); Desk Reference Guide 22.
[146] 21 C.F.R. § 14.5(b).
[147] 21 C.F.R. §§ 601.3(a), 601.4(a).
[148] 21 C.F.R. § 601.3(a)(1) & (3).
[149] 21 C.F.R. § 601.3(b)(1).
[150] See generally CDER and CBER, *Guidance for Industry: Classifying Resubmissions in Response to Action Letters* (April 1998); see also PDUFA V Performance Goals, §§ I(E), XVI(D)&(E).
[151] FDA, *Guidance for Industry: Formal Dispute Resolution: Appeals Above the Division Level* (February 2000). Dispute resolution can occur at any time in the BLA review process, GRMP Guidance, at 26, but typically occurs after receipt of a CRL.
[152] 21 C.F.R. § 601.4(b).

made, processed, packed, or held comply with good manufacturing practice (GMP).[153] Two FDA regulations define "safety" to mean "relative freedom from harmful effect" in light of the patient's underlying condition, assuming that the biologic is "prudently administered."[154] In determining whether this standard is met, the FDA must consider the risks of the product against its benefits.[155] Proof of safety comprises "adequate tests by methods reasonably applicable," including reports of "significant human experience" with the product.[156] "Purity" means that the finished product is "relative[ly] free[]" from "extraneous matter," including moisture and pyrogens.[157]

"Potency" means the product's "specific ability or capacity... to effect a given result" based on laboratory testing or controlled clinical data.[158] Thus, the FDA has interpreted "potency" to include effectiveness.[159] Nevertheless, the FDCA's requirement for "adequate and well-controlled trials," which typically means at least two pivotal clinical studies, does not apply to biologics in all circumstances.[160] Instead, this is a default requirement for biologics. Proof of efficacy must comprise adequate and well-controlled trials unless the sponsor shows that this requirement (1) "is not reasonably applicable" to the biologic or "essential to the validity" of the trial and (2) an alternative method is "adequate to substantiate effectiveness."[161] For example, serologic response evaluations may be sufficient when the correlation between the marker and clinical effectiveness has been established.[162]

## 4.3 EUROPEAN UNION GUIDELINES

In the European Union, biological medicinal product is an umbrella term covering a broad spectrum of medicinal products, all of which are larger and more complex than chemically synthesized products. Biological medicines are defined as "product[s], the active substance of which is a biological substance."[163] A "biological substance,"

[153] PHSA § 351(a)(2)(C); 21 C.F.R. § 601.3(a).
[154] 21 C.F.R. §§ 600.3(p), 601.25(d)(1).
[155] 21 C.F.R. § 601.25(d)(3).
[156] 21 C.F.R. § 601.25(d)(1).
[157] 21 C.F.R. § 600.3(r).
[158] 21 C.F.R. § 600.3(s).
[159] FDA, *Guidance for Industry: Providing Clinical Evidence of Effectiveness for Human Drug and Biological Products* 3–4 (May 1998) (Effectiveness Guidance).
[160] 46 Fed. Reg. 4634, 4635 (January 16, 1981) ("While it is clear that... the applicable statutory requirement for potency in the [PHSA] has been interpreted as requiring that a [biological] product be effective, the specific statutory criteria governing new drugs, 'adequate and well-controlled clinical studies,' have not been applied to biological drugs"); see also 21 C.F.R. § 201.57(c)(2)(F)(iv) and (v) (noting that "substantial evidence of effectiveness based on adequate and well-controlled studies" must support indications labeling for non-biologic drugs, whereas "[f]or biological products, all indications listed in this section must be supported by substantial evidence of effectiveness").
[161] 21 C.F.R. § 601.25(d)(2); see also Effectiveness Guidance, at 4.
[162] 21 C.F.R. § 601.25(d)(2).
[163] Section 3.2.1.1. Module 3, Part I, Annex I to Directive 2001/83/EC of the European Parliament and of the Council of 6 November 2001 on the Community code relating to medicinal products for human use (as amended by Directive 2009/120/EC of 14 September 2009).

in turn, is defined as "a substance that is produced by or extracted from a biological source and that needs for its characterization and the determination of its quality a combination of physico-chemical-biological testing, together with the production process and its control."[164] Annex II to the EU GMP guidelines notes that biologics "can be defined... largely by reference to their method of manufacture."[165] Examples of biological medicines include immunologic medicines; medicines derived from human blood and plasma; medicines developed by means of recombinant DNA technology, "controlled expression of genes coding for biologically active proteins in prokaryotes and eukaryotes including transformed mammalian cells,"[166] or hybridoma and mAb methods; and advanced therapy medicinal products.[167]

### 4.3.1 Nonclinical Studies

Similar to the FDA,[168] the CHMP has adopted ICH S6 as a guideline governing preclinical testing of biologics.[169] In July 2011, the CHMP adopted the addendum to this guideline, and the addendum came into effect in Europe in December 2011. The addendum complements, clarifies, and updates ICH S6 and is intended to further harmonize the standards for nonclinical studies. As explained in Section 2.2, we discuss the addendum here in light of the CHMP's earlier approval of it. The addendum and ICH S6 are applicable in both the United States and EU, however, and readers interested in understanding the nonclinical standards in either jurisdiction should review both sections.

***4.3.1.1 General Principles*** Although the addendum does not alter the scope of the ICH S6, it prevails whenever there are differences between the two.[170] The

[164] *Id.*

[165] Annex 2 on the Manufacture of Biological Medicinal Products for Human Use, Eudralex Volume 4—Guidelines for Good Manufacturing Practices for Medicinal Products for Human and Veterinary Use. The same is confirmed in the revised EU guidelines on the manufacture of biological active substances and medicinal products for human use (deadline for coming into operation is January 31, 2013) [Annex 2 on the Manufacture of Biological active substances and Medicinal Products for Human Use, Eudralex Volume 4—Guidelines for good manufacturing practices for medicinal products for human and veterinary use].

[166] Section 1 of the Annex to Regulation 726/2004/EC of the European Parliament and of the Council of 31 March 2004 laying down Community procedures for the authorisation and supervision of medicinal products for human and veterinary use and establishing a European Medicines Agency (as amended by Regulation (EC) No. 1394/2007 of 13 November 2007).

[167] Section 3.2.1.1. Module 3, Part I, Annex I to Directive 2001/83/EC of the European Parliament and of the Council of 6 November 2001 on the Community code relating to medicinal products for human use (as amended by Directive 2009/120/EC of 14 September 2009).

[168] See *supra* Section 4.2.2.

[169] Guideline on pre-clinical safety evaluation of biotechnology-derived pharmaceuticals (ICH S6) (CPMP/ICH/302/95) (in operation from March 1998).

[170] See generally Addendum to Guideline on pre-clinical safety evaluation of biotechnology-derived pharmaceuticals (ICH S6), EMA/CHMP/ICH/731268/1998 (July 2011) (ICH S6 Addendum). Unlike ICH S6, the addendum advises that sponsors of nonclinical trials for biologics used in cancer treatment should consult the ICH S9 guideline. ICH S6 Addendum, at Section 1.3; Note for guidance on nonclinical evaluation for anticancer pharmaceuticals, EMEA/CHMP/ICH/646107/2008 (November 2009).

addendum covers the following five topics: species selection, study design, immunogenicity, reproductive and developmental toxicity, and carcinogenicity.[171]

***4.3.1.2 Species Selection*** The addendum discusses the factors that sponsors should consider in selecting relevant species for nonclinical testing.[172] According to the addendum, initial testing should compare target sequence homology between species, with subsequent *in vitro* assays making qualitative and quantitative cross-species comparisons of relative target binding affinities, receptor–ligand occupancy, and kinetics. Sponsors also should assess functional activity. This testing should permit identification of a species model that can demonstrate potentially adverse consequences of target modulation.

When the preceding approaches cannot be used to determine relevant species, the sponsor may conduct tissue cross-reactivity studies. If no relevant species exists, the sponsor may consider homologous molecules or transgenic models, as noted in ICH S6.[173] Specific instructions are provided for mAbs, for which a short-term safety study in one species—and no additional toxicity studies—are recommended.[174]

If there are two relevant species (one rodent and one nonrodent), the sponsor should conduct short-term studies in both. If the toxicologic findings from these studies are similar for both species, long-term studies may involve one of those species, usually the rodent species.[175]

***4.3.1.3 Study Design*** Sponsors should consider PK–PD approaches—such as exposure–response relationships, modeling, or simulation approaches—when selecting the high dose for toxicity testing. The high dose should be the higher of (1) the dose providing the maximum intended pharmacologic effect in the preclinical species and (2) the dose providing "an approximately 10-fold exposure multiple over the maximum exposure to be achieved in the clinic."[176] When no PD endpoint is available, the sponsor should select the high dose based on PK data, as well as available *in vitro* binding and/or pharmacology data. Generally, repeat-dose toxicity tests should have a duration of 6 months; studies of longer duration are not considered valuable.[177]

Finally, the sponsor may need to assess subject recovery from the medicine's pharmacologic and toxicologic effects when these effects occur at clinically relevant exposures. One approach is to include, in at least one study, a nondosing period assessing the reversibility of the toxic effects.[178]

[171] ICH S6 Addendum, at Section 1.1.
[172] *Id.* at Section 2.1.
[173] See *supra* notes 40–42 and accompanying text.
[174] ICH S6 Addendum, at Section 2.1.
[175] *Id.* at Section 2.2.
[176] *Id.* at Section 3.1.
[177] *Id.* at Section 3.2.
[178] *Id.* at Section 3.3.

***4.3.1.4 Immunogenicity*** As noted in ICH S6, nonclinical studies are not useful in predicting potential immunogenicity of human or humanized proteins in humans.[179] The addendum provides more detail than ICH S6 regarding situations when the sponsor should measure antidrug antibodies (ADAs), namely when (1) there is evidence of altered PD activity, (2) there are unexpected changes in exposure in the absence of a PD marker, or (3) there is evidence of immune-mediated reactions. Collection of appropriate samples during the study is recommended, as noted in Section 2.2, because it is hard to predict the need for ADA measurement before the completion of the in-life phase of the study.[180]

***4.3.1.5 Reproductive and Developmental Toxicity*** The addendum first provides general advice on reproductive and developmental testing and then discusses more specific recommendations for fertility studies, embryo–fetal development (EFD) studies and pre- and postnatal development (PPND) studies, and the timing of studies in nonhuman primates (NHPs).[181]

The addendum first discusses appropriate species for testing.[182] Reproductive studies should occur in a relevant species, but no such studies are required for products directed at foreign targets, such as bacteria and viruses. If the product is pharmacologically active in both rodents and rabbits, EFD studies should occur in both species unless teratogenicity or embryo–fetal lethality is identified in one of them. Sponsors should not use NHPs in developmental testing unless they are the only relevant species, and even then, the sponsor can provide scientific justification to use an alternative model. If no relevant species exists, the sponsor may consider using transgenic mice or homologous proteins. If the weight of the evidence (e.g., information regarding the mechanism of action, phenotypic data from genetically modified animals, or class effects) suggests an adverse effect on fertility or pregnancy, this information may permit communication of the risks, and additional studies might not be warranted.

Fertility studies should occur in mice or rats when either is a relevant species. Mating studies are impractical for NHPs, but sponsors can evaluate the reproductive tract in repeat-dose toxicity studies of sexually mature NHPs that last at least 3 months.[183] Concerns about effects on conception or implantation should be addressed through studies of NHPs, a transgenic model, or a homologous protein or through risk management, informed consent, and labeling. With respect to EFD and PPDN, the sponsor may conduct separate studies or may consider one study that covers day 20 of gestation to birth (an enhanced PPDN or ePPND).[184]

[179] See *supra* note 47 and accompanying text.
[180] ICH S6 Addendum, at Section 4.
[181] *Id.* at Section 5.
[182] *Id.* at Section 5.1.
[183] *Id.* at Section 5.2.
[184] *Id.* at Section 5.3.

If the candidate enters clinical trials before completion of EFD studies, appropriate risk management techniques (e.g., contraception) should be used in any clinical trial involving women of childbearing potential.[185] If these precautions are in place and NHPs are the only relevant species, the sponsor can conduct EFD and ePPND studies during phase III. When these precautions are not possible, the sponsor should submit the complete EFD report or interim ePPND study before beginning phase III.[186] When the product's mechanism of action raises serious developmental toxicity concerns and NHPs are the only relevant species, no study is necessary; instead, the labeling should disclose the concern, and the sponsor should avoid administering the candidate to women of childbearing potential.

***4.3.1.6 Carcinogenicity*** As noted, carcinogenicity assessments of biologics are not always warranted, but the addendum provides advice for use in situations when they are appropriate.[187] According to the addendum, the sponsor may design a strategy addressing potential carcinogenicity based on a weight of evidence approach, including a review of relevant information, such as literature; information on class effects, target biology, and mechanisms of action; *in vitro* data; clinical data; and data from chronic toxicity studies.[188] In some cases, this review will be sufficient to address the carcinogenic potential.[189]

In situations when the mechanism of action raises concerns and the weight of the evidence supports them, the hazard should be addressed through product labeling and risk management practices.[190] If the weight of the evidence regarding a mechanism-based concern is instead unclear, the sponsor can propose additional studies to address it.[191]

When insufficient information exists on product characteristics and mechanism of action, a more extensive assessment might be appropriate, including, for example, additional endpoints in toxicity studies.[192] If this assessment suggests a carcinogenicity concern, sponsors may propose additional studies or labeling to address the concern.[193] If this assessment instead suggests no carcinogenicity concern, additional nonclinical testing is not recommended.[194]

### 4.3.2 Clinical Studies in Compliance with the Clinical Trials Directive

After complying with the preclinical testing requirements, biologics also need to undergo clinical trials before a marketing authorization application (MAA) can be

[185] *Id.* at Section 5.4.
[186] *Id.* at Section 5.4.
[187] See *supra* Section 4.2.2.3.
[188] ICH S6 Addendum, at Section 6.
[189] *Id.*
[190] *Id.*
[191] *Id.*
[192] *Id.*
[193] *Id.*
[194] *Id.*

submitted. The Clinical Trials Directive sets forth the general requirements for clinical trials of medicinal products, including biologics.[195] Because some general standards may not be relevant or appropriate for biologics, however, regulators must take a flexible approach to trials of these products. This section summarizes the requirements of the Clinical Trials Directive, noting special considerations for biologics when necessary.

***4.3.2.1 Clinical Trial Authorization*** The Clinical Trials Directive and European Commission guidance[196] describe the steps that a sponsor must take before commencing a clinical trial. According to these documents, a clinical trial may commence only if (1) the anticipated therapeutic and public health benefits outweigh any foreseeable risks and inconveniences to the subjects; (2) the trial subjects understand the objectives and risks of the trial and give informed, written consent to participate; (3) the trial safeguards the physical and mental integrity of the subjects; and (4) insurance covers the liability of the sponsor and investigator.

To comply with these requirements, the trial sponsor must take certain steps. In general, the sponsor must take responsibility for the following: trial conduct, appointment of an appropriate investigator, selection of the institution that will conduct the trial, quality control, data collection standards, protocol drafting, and creation of the investigator's brochure.[197] The sponsor then must apply for approval from both the ethics committee in the country where the trial will be conducted and competent authorities of the Member States. Written authorization may be required for all biologics trials and is required for trials involving medicines containing genetically modified organisms, medicines for gene therapy, and medicines for somatic cell therapy (including xenogenic cell therapy).[198] The opinion of the ethics committee should be issued within 60 days. A review period of 30 days can be added for medicines requiring written authorization noted earlier, and for xenogenic cell therapy, there are no time limits for authorization. This time frame can be extended by an additional 90 days (in addition to the original 90 days) if the ethics committee consults a national group or committee. The trial may begin only if (1) the ethics committee has issued a favorable opinion

[195] Directive 2001/20/EC of the European Parliament and of the Council of 4 April 2001 on the approximation of the laws, regulations and administrative provisions of the Member States relating to the implementation of good clinical practice in the conduct of clinical trials on medicinal products for human use (as amended by Regulation (EC) No. 596/2009 of 18 June 2009).

[196] Communication from the Commission—Detailed guidance on the request to the competent authorities for authorisation of a clinical trial on a medicinal product for human use, the notification of substantial amendments and the declaration of the end of the trial (CT-1), 2010 O.J. (C 82), 1.

[197] S. Shorthose, "*Guide to EU pharmaceutical regulatory law*," Alphen aan den Rijn: Kluwer Law International, 2011, p. 139.

[198] Article 9(5) and (6), Directive 2001/20/EC of the European Parliament and of the Council of 4 April 2001 on the approximation of the laws, regulations and administrative provisions of the Member States relating to the implementation of good clinical practice in the conduct of clinical trials on medicinal products for human use (as amended by Regulation (EC) No. 596/2009 of 18 June 2009).

and (2) no competent authority has informed the applicant of grounds for non-acceptance.[199]

#### *4.3.2.2 Good Clinical Practices and Other Considerations for Clinical Trials*

Clinical trials of biologics must comply with GCP, as described in Directive 2005/28/EC on Good Clinical Practice[200] and the ICH E6 guideline, which the CHMP has adopted.[201] The directive and guideline describe general governing principles for clinical trials. The rights, safety, and well-being of trial subjects must prevail over the interests of science and society. Investigators must obtain freely given informed consent from every trial subject before each subject is enrolled. Clinical trial information must be handled, recorded, and stored with respect for relevant confidentiality and privacy rules. Trials must comply with the ethical principles of the World Medical Association's Declaration of Helsinki. Specific GCP guidelines apply to trials of advanced therapy medicinal products.[202] These guidelines regulate issues such as the donation, procurement, and testing of human tissues and cells; the implementation of a traceability system; and specific rules on safety reporting and long-term follow-up.

Under the Clinical Trials Directive, special requirements apply to clinical trials conducted on minors and other persons not able to give informed legal consent.[203] These requirements are intended to preserve the dignity of the trial subjects, confirm that the benefits of the trial outweigh the risks, and ensure that subjects' representatives give consent with as much involvement of the subject as possible. Competent authorities must record information regarding trials in the European database of clinical trials (EudraCT), which is accessible only to other competent authorities, the European Medicines Agency (EMA), and the European Commission.[204]

CHMP has issued a guideline on quality requirements during the clinical trial period for investigational medicinal products (IMPs) containing biological or biotechnology-derived substances.[205] The guideline describes quality documentation that should be submitted to the competent authority as part of the sponsor's investigational medicinal product dossier (IMPD). Given the importance of the production process for a biologic's properties, as described in Section 3, the

[199] *Id.* Article 9(1).

[200] Commission Directive 2005/28/EC of 8 April 2005 laying down principles and detailed guidelines for good clinical practice as regards investigational medicinal products for human use, as well as the requirements for authorisation of the manufacturing or importation of such products, 2005 O.J. (L 91) 13.

[201] Note for guidance on good clinical practice, CPMP/ICH/135/95 (1997; revised July 2002) (ICH E6 Guideline).

[202] Detailed guidelines on good clinical practice specific to advanced therapy medicinal products, ENTR/F/2/SF/dn D(2009) 35810 (December 3, 2009).

[203] Article 4 and Article 5, Directive 2001/20/EC of the European Parliament and of the Council of 4 April 2001 on the approximation of the laws, regulations and administrative provisions of the Member States relating to the implementation of good clinical practice in the conduct of clinical trials on medicinal products for human use (as amended by Regulation (EC) No. 596/2009 of 18 June 2009).

[204] *Id.* Article 11.

[205] Guideline on the requirements for quality documentation concerning biological investigational medicinal products in clinical trials, EMA/CHMP/BWP/534898/2008 (March 15, 2012).

guideline states that the IMPD should include, among other things (1) an adequate description of the process and process controls, including a flow chart of all successive steps and details of in-process testing, and (2) a description and justification of "any reprocessing during manufacture of the drug substance."[206]

The guideline also recognizes that sponsors will improve and optimize their manufacturing processes during clinical development and describes the steps sponsors should take following these changes.[207] Specifically, the sponsor must compare the quality attributes of the pre- and postchange biological active substances and relevant intermediates, and "[d]epending on the consequences of the change introduced and the stage of development, a comparability exercise may be necessary."[208] For first-in-human (FIH) clinical trials, sponsors should use product representative of the material used during the nonclinical testing phase.[209] Finally, with regard to characterization, the guideline requires details on the biological activity to be provided, recognizing that the extent of characterization data will further increase in later phases.[210]

***4.3.2.3 Study Design Considerations*** General guidance on study design applies to biologics as well as small molecule medicines.[211] According to the guidance, there is a "close, but variable correlation" between phase of development and type of study, but one type of trial can occur in several different phases.[212] The guidance therefore identifies the "[m]ost typical kind of study for each phase."[213]

Phase I usually involves the initial introduction of the investigational product into human subjects, and studies in this phase usually have nontherapeutic objectives. Specifically, phase I studies typically investigate one or more of the following: (1) initial safety and tolerability; (2) PK, which are "particularly important to assess the clearance of the drug and to anticipate possible accumulation of parent drug or metabolites and potential drug-drug interactions"; (3) PD; and (4) drug activity, to preliminarily determine the potential therapeutic benefit of a medicine.[214] The most typical phase I study is the human pharmacology study. According to the guidance, phase I studies may be conducted in healthy volunteers or certain types of patients (e.g., patients with mild hypertension). If the medicine has significant potential toxicity (e.g., cytotoxic products), the trial will usually be conducted in patients.

The CHMP introduced new guidelines on FIH studies after the 2006 TeGenero incident. That case involved a FIH clinical trial of a novel mAb, during which

[206] *Id.* at S.2.2.
[207] *Id.* at S.2.6.
[208] *Id.*
[209] *Id.*
[210] *Id.* at S.3.1.
[211] Note for guidance on general considerations for clinical trials, CPMP/ICH/291/95 (March 1998) (ICH E8 Guideline).
[212] *Id.* at 3.1.3.
[213] *Id.* at 3.1.3.1, 3.1.3.2, 3.1.3.3.
[214] *Id.* at 3.1.3.1.

subjects experienced severe adverse events, including multiorgan failure.[215] The new guideline describes risk factors helpful in identifying potential severe adverse reactions, including concerns derived from knowledge or lack thereof regarding (1) the product's mode of action, (2) the nature of the target, or (3) the relevance of animal models.[216] The guideline also discusses quality, nonclinical, and clinical study design considerations for minimizing risk to human subjects.[217]

The most typical phase II study is a therapeutic exploratory study that explores efficacy in narrowly defined, relatively homogenous groups of patients. Initially, studies may use a variety of designs (e.g., concurrent controls and comparisons with baseline status). Subsequent phase II trials usually are randomized and concurrently controlled, allowing for evaluation of the medicine's safety and efficacy for a particular indication. A major goal of this phase is to determine the dose(s) for phase III trials.[218]

Phase III typically involves therapeutic confirmatory studies that are designed to verify the preliminary evidence obtained in phase II and to provide a sufficient basis for marketing authorization. Phase III studies "may also further explore the dose-response relationship, or explore the drug's use in wider populations, in different stages of disease, or in combination with another [medicine]."[219] With regard to medicines administered for long periods, extended exposure trials ordinarily occur during phase III, although the sponsor may start them in phase II.[220]

To ensure that clinical trials in all three phases of development will be adequate to support an MAA, sponsors should design these trials with the MAA requirements in mind. Biologics in general need to comply with the requirements set out in Part III of the CTD, and advanced therapy medicinal products need to comply with the requirements described in Part IV of the CTD. Section 3.3, *infra*, provides more information regarding the MAA.

***4.3.2.4 Consultation with the European Medicines Agency*** A sponsor may obtain, from the EMA, scientific advice regarding clinical trial protocols.[221] Although this advice does not bind the ethics committee and national competent authority and is not binding for purposes of a future MAA, it can be useful to guide revisions to the protocol.[222] The agency's remarks will only address scientific issues

[215] Ganesh Suntharalingam et al., *Cytokine Storm in a Phase 1 Trial of the Anti-CD28 Monoclonal Antibody TGN1412*, 355(10) N Engl. J. Med. 1018, 1018 (2006).

[216] CHMP Guideline on strategies to identify and mitigate risks for first-in-human clinical trials with investigational medicinal products, EMEA/CHMP/SWP/28367/07 (July 19, 2007), at Section 4.1.

[217] *Id.* at Sections 4.2–4.4.

[218] ICH E8 Guideline, at 3.1.3.2.

[219] *Id.* at 3.1.3.3.

[220] *Id.*

[221] Article 57(1)(n), Regulation 726/2004/EC of the European Parliament and of the Council of 31 March 2004 laying down Community procedures for the authorisation and supervision of medicinal products for human and veterinary use and establishing a European Medicines Agency (as amended by Regulation (EU) No. 1235/2010 of 15 December 2010).

[222] EMA, Scientific Advice, http://www.ema.europa.eu/ema/index.jsp?curl=pages/regulation/general/general_content_000050.jsp&mid=WC0b01ac05800229bb&jsenabled=true (last accessed December 3, 2012).

and will generally focus on matters such as the selection of endpoints and comparator, the duration of treatment or follow-up, and the design of pivotal studies.[223] Advice also might address a sponsor's proposal to deviate from a CHMP guideline.[224] If the applicant decides not to follow the EMA's advice, it should justify this decision in its MAA.[225] EMA guidance details the procedures for requesting scientific advice.[226]

The fact that an applicant requests advice from EMA does not preclude it from also seeking advice from national competent authorities or from foreign regulators, such as the FDA. The process of obtaining advice from the national competent authorities is often less formal than requesting advice from the EMA, and such advice can prove helpful. Consequently, seeking such advice is a common choice among applicants.[227]

Applicants also may seek parallel scientific advice from the EMA and FDA.[228] Generally, the parallel scientific procedure is available for "important breakthrough drugs," products that the EMA and FDA have identified as falling within therapeutic areas of overlapping interest (e.g., oncology products, vaccines, and blood products), and products with "significant clinical safety, animal toxicology, or unique manufacturing concerns that could impede... development."[229] The goal of these meetings is to provide clarity regarding the regulatory requirements of each region and the reasons for any differences between them.[230] A sponsor requesting parallel advice should authorize the agencies to exchange all information about the product, including trade secrets.[231] After the parallel scientific advice procedure, each agency will provide its own independent advice on the questions at issue.[232] There is no guarantee of harmonized advice or identical regulatory decisions on the approvability of the product.[233] Nevertheless, sponsors are increasingly requesting parallel scientific advice. For example, in the period ranging from September 2009 to September 2010, the agencies received seven requests for such advice.[234]

[223] See Shorthose, *supra* note 197, at 141.
[224] European Medicines Agency guidance for companies requesting scientific advice and protocol assistance, EMEA-H-4260-01-Rev. 6 (May 21, 2010).
[225] *Id.* Item 20.
[226] See generally *id.*
[227] See Shorthose, *supra* note 197, at 143.
[228] EMA/FDA, General Principles: EMEA-FDA Parallel Scientific Advice, EMEA/24517/2009 (July 2009), at 1.
[229] *Id.* at 2.
[230] *Id.*
[231] *Id.*
[232] *Id.* at 3.
[233] *Id.*
[234] EMA/FDA, Interactions between the European Medicines Agency and U.S. Food and Drug Administration September 2009–September 2010, EMA/705027/2010 (June 2011), Section 6.

### 4.3.3 The Marketing Authorization Application: Contents and Approval Standard

Many biologics fall under the scope of the centralized marketing authorization procedure, which is mandatory for medicines developed through biotechnological methods (recombinant DNA technology; controlled expression of genes coding for biologically active proteins in prokaryotes and eukaryotes, including transformed mammalian cells; and hybridoma and mAb methods).[235] For example, the following are subject to the centralized procedure: cell therapy, gene therapy, vaccines from strains developed through recombinant DNA technology (including gene deletion), and "any medicinal product for which a monoclonal antibody is used at any stage in the manufacturing process."[236]

Nonetheless, some biologics are still approved at the Member State level. For example, many vaccines do not fall within the scope of the centralized procedure. The EMA has published a guideline intended to harmonize the summaries of product characteristics and patient information leaflets for human vaccines.[237] This chapter discusses the requirements of the EU centralized procedure.

The approval standards for biotechnology products are the same as for chemically synthesized medicines. Both types of products must be safe and effective and have appropriate quality. Because of their special characteristics, however, biotechnology products must comply with several additional dossier requirements.

The MAA for a biotechnology product must meet the standard dossier submission requirements, as described in Article 8 of the Medicines Directive.[238] Consequently, the MAA must generally comply with the CTD format, including with respect to Module I (administrative information, including labeling and mock-ups), Module 2 (various summaries), Module 3 (chemical, pharmaceutical, and biological information), Module 4 (nonclinical reports), and Module 5 (clinical study reports).[239]

MAAs for biologics also must meet special requirements. The applicant must thoroughly describe the manufacturing process and must (1) provide information on the origin and history of the starting materials; (2) demonstrate that the active substance complies with specific measures for preventing the transmission of animal spongiform encephalopathies; (3) if cell banks are used, demonstrate that cell

[235] Article 3(1) and Annex to Regulation 726/2004/EC of the European Parliament and of the Council of 31 March 2004 laying down Community procedures for the authorisation and supervision of medicinal products for human and veterinary use and establishing a European Medicines Agency (as amended by Regulation (EC) No. 1394/2007 of 13 November 2007).

[236] European Commission, Notice to Applicants, vol. 2A, ch. 4, of April 2006 (ENTR/F2/BL D(2006)), Section 1.1.

[237] CPMP Guideline on pharmaceutical aspects of the product information for human vaccines, EMEA/CPMP/BWP/2758/02 (November 23, 2003). Revisions to this guideline are under consideration. Concept paper on a revision of the guideline on pharmaceutical aspects of the product information for human vaccines EMEA/CHMP/BWP/290688/2009 (June 17, 2009).

[238] 8, Directive 2001/83/EC of the European Parliament and of the Council of 6 November 2001 on the Community code relating to medicinal products for human use (as amended by Directive 2011/62/EU of 8 June 2011).

[239] *Id.* at Annex I, Table of Contents.

characteristics remain unchanged at the passage level for production (and beyond); (4) provide information as to whether there are adventitious agents in seed materials, cell banks, pools of serum or plasma, and all other materials of biological origin, and, if it is not possible to avoid the presence of potentially pathogenic adventitious agents, show that further processing ensures elimination or inactivation of the agents; (5) if possible, base vaccine production on a seed lot system and established cell banks; (6) in case of medicines derived from human blood or plasma, describe the origin, criteria, and procedures for the collection, transportation, and storage of the starting material; and (7) describe the manufacturing facilities and equipment.[240]

Other special rules apply certain types of biological medicines. For example, for plasma-derived medicinal products, the applicant must provide an information dossier, the Plasma Master File. MAAs for vaccines other than for influenza need to contain a Vaccine Antigen Master File. Special rules also apply to advanced therapy medicinal products, including gene therapies, somatic cell therapies, and tissue-engineered products.[241]

## 4.4 REGULATORY STRATEGIES FOR WORLDWIDE MARKETING OF BIOLOGICAL PRODUCTS

### 4.4.1 Acceptance of Foreign Clinical Studies in the United States and Europe

***4.4.1.1 United States*** The FDA has adopted two regulations governing its acceptance of foreign clinical data, one applicable to supportive data and one applicable to data that form the sole basis for approval. Both regulations require the sponsor to meet certain conditions before the FDA will agree to use of the data.

First, the FDA accepts "well-designed and well-conducted" foreign, non-IND studies as "support" for an IND or BLA if two conditions are met.[242] The FDA generally must be able to conduct an onsite inspection of the data, if necessary.[243] The sponsor also must have conducted the study using GCP, as defined in 21 C.F.R. § 312.120.[244] For purposes of that regulation, GCP means standards that ensure the credibility of the results and the protection of subjects, including independent ethics board approval and documentation of subjects' informed consent.[245] Complying with ICH E6, the GCP guidance, is one way—but not the only way—to meet this requirement.[246] The FDA recently issued guidance on

[240] *Id.* at Annex I, Section 3.2.1.2.
[241] Advanced therapy medicinal products are regulated under a separate regime, Regulation 1394/2007/EC of the European Parliament and of the Council of 13 November 2007 on advanced therapy medicinal products and amending Directive 2001/83/EC and Regulation (EC) No. 726/2004 (as amended by Regulation (EU) No. 1235/2010 of 15 December 2010).
[242] 21 C.F.R. § 312.120(a)(1).
[243] 21 C.F.R. § 312.120(a)(1)(ii).
[244] 21 C.F.R. § 312.120(a)(1)(i).
[245] *Id.*; see also 73 Fed. Reg. 22800, 22807 (April 28, 2008).
[246] 73 Fed. Reg. at 22807.

submitting information to demonstrate compliance with 21 C.F.R. § 312.120.[247] After these threshold criteria are met, the FDA will use the factors described in ICH E5, *Ethnic Factors in the Acceptability of Foreign Clinical Data*, to determine the scientific relevance of the data with respect to the US population and the need for additional bridging data to confirm their value. This exercise is described in Section 4.2, *infra*.

Second, when foreign data are intended to form the sole basis for United States marketing approval, the sponsor must comply with GCP and meet three other criteria.[248] The FDA must deem the foreign data to be "[a]pplicable to the U.S. population and U.S. medical practice" using the criteria described in ICH E5.[249] The clinical investigators must have "recognized competence." Finally, the data must be considered valid without the need for an onsite FDA inspection, or the FDA must be able to validate the data through such an inspection.

***4.4.1.2 Europe*** Directive 2001/83/EC allows for clinical trials conducted outside the European Union to be taken into consideration during the review of an MAA in the European Union if such trials have been designed, implemented, and reported based on principles equivalent to those of the Clinical Trials Directive with regard to good clinical practice and ethical principles. Moreover, they should comply with the ethical principles outlined in the Declaration of Helsinki.[250] The applicant must submit a statement declaring such compliance as part of the MAA.[251]

In December 2008, the EMA published a strategy paper on the acceptance of data from foreign clinical trials conducted in "third countries," particularly those outside the "'traditional' Western European and North American research areas."[252] According to this strategy paper, there is a "growing concern both among regulators and in public debate about how well these trials are conducted from an ethical and scientific/organizational standpoint."[253] The EMA has called for increased cooperation between international regulatory authorities involved in the supervision of clinical trials and has put forth other proposals to address these issues.[254]

[247] FDA, *Guidance for Industry and FDA Staff: FDA Acceptance of Foreign Clinical Studies Not Conducted Under an IND Frequently Asked Questions* (March 2012).

[248] 21 C.F.R. § 314.106(b); Murray M. Lumpkin, M.D., M.Sc., Deputy Commissioner, Int'l Programs, FDA, *FDA Perspective on Int'l Clinical Trials*, Presentation at FDA Clinical Trials Workshop (November 8, 2011), Slide 7.

[249] Lumpkin, *supra* note 248, Slide 14.

[250] Recital 8, Annex I to Directive 2001/83/EC of 6 November 2001 on the Community code relating to medicinal products for human use (as amended by Directive 2009/120/EC of 14 September 2009).

[251] Article 8 of Directive 2001/83/EC of 6 November 2001 on the Community code relating to medicinal products for human use (as amended by Directive 2011/62/EU of 8 June 2011).

[252] EMEA strategy paper: Acceptance of clinical trials conducted in third countries, for evaluation in Marketing Authorisation Applications, December 5, 2008, EMEA/228067/2008.

[253] *Id.*

[254] See, e.g., *id.*; "Reflection paper on ethical and GCP aspects of clinical trials of medicinal products for human use conducted outside of the EU/EEA and submitted in marketing authorisation applications to the EU Regulatory Authorities," April 16, 2012, EMA/121340/2011; "International workshop on the ethical and good-clinical-practice aspects of clinical trials conducted in third countries", held from June 9, 2010, to July 9, 2010, in London, UK.

### 4.4.2 Foreign and Multinational Clinical Studies: Addressing Ethnic Factors

ICH E5 describes strategies to extrapolate clinical data generated in one region to support approval in another based on an assessment of ethnic factors' impact on the medicine's safety and efficacy.[255] "Ethnic factors" include intrinsic factors, such as genetics and age, and extrinsic factors, such as regional clinical trial conduct and medical practice.[256]

For the ICH E5 framework to apply, the clinical trial must meet the regulatory requirements (e.g., choice of control, trial endpoints, and key design features) of the region where approval is sought.[257] Regulators then assess the medicine's sensitivity to ethnic factors using information about its PK and PD and their relationship to safety and efficacy.[258] Based on the level of ethnic sensitivity, regulators will then determine whether existing data show the trials' relevance to the new region or whether a bridging study is necessary to confirm their relevance.[259] For example, when the medicine is ethnically insensitive and extrinsic factors in the two regions are similar, regulators might not require a bridging study.[260] Generally, one bridging study will suffice for extrapolation under ICH E5 unless the bridging study is too small to assess safety or does not confirm the relevance of the foreign data to the new region's population.[261] In these cases, regulators in the new region likely will require additional data. Depending on the circumstances, a bridging study might use pharmacologic endpoints or might constitute a controlled clinical trial.[262] Separate safety data might be needed when the sponsor does not need to perform a bridging efficacy study or when the efficacy study is not powered for safety.[263]

ICH E5 and its companion Q&A guidance also discuss strategies for a multinational trial to support simultaneous registration applications in multiple countries.[264] The trial's goals would be to show efficacy in each region and to compare regional results to show insensitivity to ethnic factors.[265] In designing these types of studies, sponsors should choose a primary endpoint that is acceptable to all regional regulatory authorities (or when this is impossible, collect data on all primary endpoints in all regions for comparison).[266] The sponsor should use a common

[255] FDA, *ICH E5: Ethnic Factors in the Acceptability of Foreign Clinical Data*, as adopted by FDA at 63 Fed. Reg. 31790, 31791 (June 10, 1998).
[256] *Id.* at 31791, 31793–31794.
[257] *Id.* at 31791–31792; FDA/ICH, *Guidance for Industry: E5 – Ethnic Factors in the Acceptability of Foreign Clinical Data: Questions & Answers* 4 (September 2006) (ICH E5 Q&A).
[258] 63 Fed. Reg. at 31792.
[259] *Id.*
[260] *Id.*
[261] *Id.*
[262] *Id.* at 31792–31793.
[263] *Id.* at 31793.
[264] See *id.* at 31793; ICH E5 Q&A, at 6.
[265] ICH E5 Q&A, at 6.
[266] *Id.*

protocol to collect endpoint data and should power the study to permit an efficacy showing in each region.[267] Collection of safety data should be as uniform as possible across regions, and the study should meet all regional requirements for design and analysis.[268] Other factors relevant to study design, such as definition of disease and choice of control group, should be discussed with regional regulatory authorities.[269] The sponsor should provide the results by region and assess the consistency of the product's effect across regions.[270] When the regional results are persuasive, this showing can confirm the relevancy of the foreign results as further support for the regional marketing application.[271]

## 4.5 PREVENTIVE VACCINE DEVELOPMENT: SPECIAL CONSIDERATIONS

### 4.5.1 General Considerations for Vaccine Development in the United States

Development of a vaccine for FDA approval presents special issues. For example, vaccines are often intended for use in healthy populations; thus, they present distinct risk–benefit issues from therapeutic products.[272] As another example, data regarding concomitant use of other vaccinations are important to licensure.[273] Section 5.1 provides an overview of vaccine development issues in the United States, including necessary studies by phase of development, study endpoints, and the possible need for bridging studies.

***4.5.1.1 Types of Studies by Phase of Development*** For vaccines, phase 1 studies involve an initial assessment of safety and immunogenicity in a small number of healthy adult volunteers, ordinarily individuals at low risk of contracting the disease of interest.[274] The phase 1 study primarily assesses safety. Investigators should monitor patients for local and systemic adverse events at specified times in the week after administration and in the months that follow (including an assessment at 6 months after the last dose).[275] The protocol should include a toxicity grading scale for these events.[276] The stopping criteria generally must be more conservative than in therapeutic settings because vaccine trials enroll healthy individuals.[277] When the

[267] *Id.* at 6–7.
[268] *Id.* at 7.
[269] *Id.* at 2.
[270] *Id.* at 7.
[271] *Id.* at 6.
[272] Karen L. Goldenthal, M.D., Loris D. McVittie, Ph.D., Cynthia Kleppinger, M.D., and Antonia Gerber, M.D., Div. of Vaccines and Related Products Applications, CBER, FDA, in BIOLOGICS DEVELOPMENT: A REGULATORY OVERVIEW 99, 100 (Mark Mathieu, ed., 3rd ed., 2004).
[273] *Id.* at 101.
[274] *Id.*
[275] *Id.* at 102.
[276] See generally FDA, *Guidance for Industry: Toxicity Grading Scale for Healthy Adult and Adolescent Volunteers Enrolled in Preventive Vaccine Clinical Trials* (September 2007).
[277] See *id.* at 1; Goldenthal et al., *supra* note 272, at 102–103.

sponsor studies a live vaccine, the phase 1 study should assess the "shedding" of live vaccine organisms in bodily substances, and investigators might need to isolate vaccinated individuals to evaluate shedding and any reversion of the vaccine strain to wild type.[278] The sponsor also might need to conduct additional studies assessing secondary transmission of the disease to third parties coming into contact with the vaccinated individual.[279]

Phase 2 studies should enroll individuals "at clear risk" for the disease.[280] These studies should produce "more definitive" immunogenicity data that allow the sponsor to determine whether an adjuvant is needed and to select the vaccine formulation, dose, dosing schedule, and route of administration for phase 3 trials.[281] Phase 2 studies also should evaluate the immune response to the vaccine upon administration with likely concomitant vaccines.[282] In addition to assessing the vaccine's effects, these studies also should assess the disease that it is intended to prevent, to allow refinement of one or more "case definitions" of the disease or infection to be prevented.[283] For example, the sponsor should gather epidemiologic data on the disease, including seroincidence data when applicable, in at-risk individuals and should determine geographic strain specificity.[284] By the end of phase 2, the sponsor should have developed and validated laboratory assays that will be used for the case definition for the efficacy trials (e.g., those used to distinguish wild-type immune responses from those that the vaccine elicits).[285]

Phase 3 studies should be controlled, randomized, and double blinded.[286] In formulating sample size calculations, sponsors should consider that multiple immunizations might be needed to achieve maximum efficacy.[287] When appropriate, however, sponsors may conduct a detailed safety assessment in only a subset of subjects as long as active monitoring for serious adverse events is in place for all subjects.[288] The FDA typically requires long-term follow-up, which might take the form of a postmarket commitment, to assess the duration of immunogenicity and efficacy, long-term safety, and the need for different doses.[289] "Ideally," the sponsor will evaluate the correlation of protection with immune response at specific time points after immunization as part of the phase 3 program.[290]

Sponsors should consider use of a Data Safety Monitoring Board (DSMB), which may conduct an interim review of the data, for the phase 3 vaccine trials.[291] The

[278] Goldenthal et al., *supra* note 272, at 103.
[279] *Id.*
[280] *Id.* at 104.
[281] *Id.*
[282] *Id.*
[283] *Id.*
[284] *Id.*
[285] *Id.*
[286] *Id.* at 105.
[287] *Id.*
[288] *Id.*
[289] *Id.*
[290] *Id.* at 107.
[291] *Id.*

protocol should specify the conditions that trigger any planned interim review, the statistical analysis plan for the interim analysis, and specific early termination criteria (e.g., criteria based on a toxicity grading scale).[292]

***4.5.1.2 Endpoints in Vaccine Studies*** The FDA accepts three types of endpoints for showing vaccine efficacy: (1) clinical endpoints (i.e., prevention of the disease in question); (2) immune response endpoints; and (3) pursuant to the animal rule described earlier in Section 2.3.3, animal study endpoints that are "clearly related" to the desired benefit in humans, such as survival or prevention of major morbidity.[293] First, the FDA generally mandates use of a clinical endpoint for vaccines that are novel or the first of their kind for the population, among other things.[294] Second, as noted earlier, the FDA will accept a serologic endpoint "where a previously accepted correlation between [this endpoint] and clinical effectiveness already exists"[295] (e.g., based on prior successful clinical studies using clinical endpoints or population-based studies of immunized individuals).[296] Because serologic endpoints may allow for smaller efficacy trials, however, their use in pivotal efficacy studies might result in a need for additional safety studies.[297] Third, sponsors can use the animal rule only when studies using clinical or serologic endpoints are unethical or infeasible; this might be the case for vaccines to address smallpox or anthrax, for example.[298]

***4.5.1.3 Necessary Bridging Studies*** Bridging data might be needed in several circumstances. For example, these data might be needed when the sponsor conducts efficacy studies abroad. In this case, the frameworks described in Sections 4.1.1 and 4.2 will govern the acceptability of the foreign data and the type of bridging studies needed to assess ethnic factors, respectively.[299] For vaccines studied in foreign efficacy trials, the FDA typically requires bridging data regarding safety and immunogenicity.[300] Similarly, when the manufacturer implements process changes after the pivotal study for a vaccine, the FDA may require clinical bridging studies between the pre- and postchange vaccines to assess whether the immune responses to both are equivalent and to evaluate safety.[301] If no immune correlate of protection has been identified and the new product elicits a decreased immune response, a comparability showing might not be possible.[302]

[292] *Id.*; Rosemary Tiernan, M.D., M.P.H., FDA, Div. of Vaccines and Related Products Applications, CBER, FDA, Regulatory Perspective on Development of Preventive Vaccines for Global Infectious Diseases: Presentation at the Nat'l Institute for Allergy & Infectious Diseases (October 30, 2007), Slide 24.
[293] Tiernan, *supra* note 292, Slide 11; 21 C.F.R. § 601.91(a)(3).
[294] Tiernan, *supra* note 292, Slide 12.
[295] 21 C.F.R. § 601.25(d)(2).
[296] Goldenthal et al., *supra* note 272, at 107–108.
[297] *Id.* at 109.
[298] 21 C.F.R. § 601.90; Tiernan, *supra* note 292, Slide 21.
[299] See Goldenthal et al., *supra* note 272, at 107.
[300] *Id.*
[301] *Id.* at 108.
[302] *Id.* at 109.

### 4.5.2 Special Procedures for Centralized Review of Vaccine Dossiers in the European Union

Under EU law, a fast track authorization procedure is available for pandemic influenza vaccines. The sponsor may submit, and obtain approval of, the core pandemic dossier during the interpandemic period[303] based on a "mock-up" vaccine (i.e., a vaccine containing an active ingredient that mimics the novelty of a pandemic virus).[304] After a pandemic is declared, the special procedure allows for fast track approval of a pandemic variation.[305] The manufacturers of influenza vaccines may submit the core pandemic dossier and the subsequent pandemic variation either to the EMA for a centralized approval or to the national authorities for a national or mutual recognition procedure approval.[306] The centralized procedure is obligatory for certain techniques of vaccine preparation, such as reverse genetics.[307]

### 4.5.3 Development of Vaccines to Protect Against Global Infectious Diseases

***4.5.3.1 United States*** The FDA reviews and approves vaccines for infectious diseases that are not endemic or have not been reported to occur in the United States, such as cholera, malaria, and tuberculosis.[308] The regulatory pathway, approval standard, and required level of evidence are no different for these vaccines and those intended to prevent diseases endemic in the United States.[309] For example, review under the animal rule is available under the same conditions for both types of vaccines.[310] The FDA has licensed several vaccines, including ones for Japanese encephalitis and typhoid, based on efficacy trials performed solely in disease-endemic regions.[311]

Because foreign efficacy trials likely will be necessary when the disease has a low incidence in the United States, the FDA's regulations governing acceptability of foreign data and ICH guidance on ethnic factors again are likely to apply.[312] In these situations, the "bridge" might comprise data from a single-arm safety and

[303] For a detailed explanation of the procedure, see CHMP Guideline on the submission of marketing authorization applications for pandemic influenza vaccines through the centralized procedure of 5 April 2004, EMEA/CHMP/VEG/4986/03.
[304] WHO, Safety of pandemic vaccines, http://www.who.int/csr/disease/swineflu/notes/h1n1_safety_vaccines_20090805/en/index.html (last accessed December 3, 2012).
[305] CPMP Guideline on Dossier Structure and Content for Pandemic Influenza Vaccine Marketing Authorization Application of 18 December 2008, EMEA/CPMP/VEG/4717/2003.
[306] CHMP Guideline on the submission of marketing authorization applications for pandemic influenza vaccines through the centralized procedure of 5 April 2004, EMEA/CHMP/VEG/4986/03.
[307] Part 1 of Annex to Regulation 726/2004/EC of 31 March 2004 laying down Community procedures for the authorization and supervision of medicinal products for human and veterinary use and establishing a European Medicines Agency" (as amended by Regulation (EC) No. 1394/2007 of 13 November 2007).
[308] FDA, *Guidance for Industry: General Principles for the Development of Vaccines to Protect Against Global Infectious Diseases* 3 (December 2011) (Global Infectious Disease Guidance).
[309] *Id*. at 3, 7.
[310] See *id*. at 4, 6–7.
[311] *Id*. at 5, 8.
[312] See *supra* Sections 4.4.1.1 and 4.4.2.

immunogenicity study in the United States, from which comparisons can be drawn to pivotal foreign data.[313] Sponsors should discuss with FDA their plans for development of global infectious disease vaccines, including bridging study design and special issues for concomitant vaccines in this context.[314]

***4.5.3.2 Europe*** The CHMP may "give a scientific opinion, in the context of cooperation with the World Health Organization" regarding medicines that are intended exclusively for markets outside the Community and "used to prevent or treat diseases of major public health interest."[315] Eligible medicines include vaccines used in the World Health Organization Expanded Program on Immunization or vaccines for protection against a public health priority disease.[316]

## 4.6 CONCLUSION

Despite clear differences in the US and EU regulatory regimes for biologics—for example, with respect to the definitional frameworks for biologics—the systems share a number of similarities. Both regimes use certain harmonized scientific testing standards, and one can safely conclude that both regimes clearly recognize the specific nature of biologics and take appropriate measures to address the possible issues resulting from it. The regions have distinct regulatory approval procedures because of the general structural differences in their medicine authorization regimes, as well as different historical developments in both systems. Overall, the US and EU regulatory regimes for biologics are more similar than they are different.

[313] Tiernan, *supra* note 292, Slides 48–49.
[314] *Id*., Slide 52; Global Infectious Disease Guidance, at 5.
[315] Article 58 of Regulation 726/2004/EC of 31 March 2004 laying down Community procedures for the authorization and supervision of medicinal products for human and veterinary use and establishing a European Medicines Agency (as amended by Regulation (EU) No. 1235/2010 of 15 December 2010); EMA, Opinions on medicines for use outside the European Union, http://www.ema.europa.eu/ema/index.jsp?curl=pages/regulation/general/general_content_000312.jsp&mid=WC0b01ac058001d12c&jsenabled=true.
[316] EMA, Opinions on medicines for use outside the European Union, http://www.ema.europa.eu/ema/index.jsp?curl=pages/regulation/general/general_content_000312.jsp&mid=WC0b01ac058001d12c&jsenabled=true.

# Landscape and Consideration of Intellectual Property for Development of Biosimilars

SRIKUMARAN MELETHIL

Prediction is very difficult, especially if it's about the future.

—*Niels Bohr*

## 5.1 INTRODUCTION

Biological products,[1] compared with drugs, are complex molecules such as proteins, hormones, vaccines, enzymes, and monoclonal antibodies (mAbs). Because of their general instability in the gastrointestinal tract, such products are generally administered by non-oral routes (e.g., intravenous, intramuscular). A visual representation of the magnitude of differences between drugs and biologics is depicted in a recent publication showing structural representations of a complex mAb and the small molecule drug aspirin.[2] Biologicals are often quite expensive. For example, the annual cost of Cerezyme used in the treatment of Gaucher disease is $200,000.[3]

[1] The term "biological product" means a virus, therapeutic serum, toxin, antitoxin, vaccine, blood, blood component or derivative, allergenic product, protein (except any chemically synthesized polypeptide) or analogous product, or arsphenamine or derivative of arsphenamine (or any other trivalent organic arsenic compound), applicable to the prevention, treatment, or cure of a disease or condition of human beings (42 U.S.C. § 262(i)(1)).

[2] Kozlowski S, Woodcock J, Midthun K, Sherman RB, Noonan KE. Developing the nation's biosimilars program. N Engl J Med 2011;365:385–388.

[3] Johnson JA. FDA regulation of follow-on biologics, April 26, 2010. Available from: http://www.primaryimmune.org/advocacy_center/pdfs/health_care_reform/Biosimilars_Congressional_Research_Service_Report.pdf (henceforth, Johnson, 2010).

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*,
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

Patents on biological products with an estimated $10 billion in sales expired in 2010; this number is expected to double by 2015.[4] Therefore, cheaper versions of biological products are possible, at least in theory, and meet an important societal need.[5]

Biological products, with a few exceptions, are approved under § 351 of the Public Health Service Act,[6] which did not previously provide for the marketing of generic versions of such products until amended in 2010. Clearly, the approval of biosimilars in Europe starting in 2006 put considerable political pressure on the US Congress to act.[7] These amendments[8] can be found in the Biologics Price Competition and Innovation Act of 2009 (BPCI Act), which allows for the approval of generic biological products, also called follow-on biologics and biosimilars (this latter term will be used in this chapter). It is superficially modeled after the Drug Price Competition and Patent Term Restoration Act of 1984[9] (commonly referred to as the Hatch-Waxman Act), which provided a similar pathway for the entry of generic drugs, which are mostly composed of relatively smaller and simpler molecules.

Given the complex nature of biological products, many legitimate scientific issues regarding the safety and efficacy of biosimilars have been raised.[10] For example, minor lot-to-lot variations in manufacturing processes can lead to structural variability (e.g., extent of protein glycosylation), and the current limitations in analytical methodology may fail to detect these minor differences; such differences can affect the efficacy and safety (e.g., immunogenicity) of a biological product. Therefore, the challenges for a company to market a generic version of a biological product that is biosimilar to or interchangeable[11] with an existing biological product is considerably greater than with small molecule drugs. These scientific challenges are discussed in elaborate detail in other chapters in this book. A glimpse of these challenges of providing cost-effective biologicals to American consumers can be found in the following hopeful statement regarding the BPCI Act: "It is the sense of the Senate that a biosimilars pathway balancing innovation and consumer interests should be established."[12] Interestingly, independent analyses of cost savings to the federal

[4] Chu R, Pugtach M. Biogenerics or Biosimilars? Discussing the Present, Considering the Future. Available from: http://www.stockholm-network.org/downloads/publications/Biosimilars_FINAL.pdf.

[5] Carver KH, Elikhan J, Lietzan E. An unofficial history of the Biologic Price Competition and Innovation Act of 2009. Food Drug Law J 2010;65:671–818.

[6] 42 U.S.C. § 262.

[7] Anonymous. Europe approves two follow-on human growth hormones. Nature Biotech 2009;24:601–602.

[8] The Biologics Price Competition and Innovation Act of 2009 (henceforth the BPCI Act) is embedded in the omnibus law titled the Patient Protection and Affordable Care Act of 2010, as Title VII—Improving Access to Innovative Medical Therapies, Subtitle A.

[9] Federal Food, Drug, and Cosmetic Act. Federal Food, Drug, and Cosmetic Act. Section 505(j) (21 U.S.C. § 355(j)).

[10] Woodcock J. Assessing the impact of a safe and equitable biosimilar policy in the United States. Statement before the Subcommittee on Health, Committee on Energy and Commerce, United States House of Representatives. Available from: http://www.hhs.gov/asl/testify/2007/05/t20070502h.html.

[11] Legal definitions for "biosimilar" and "interchangeable" are provided later in this chapter.

[12] Sense of the Senate (Patient Protection and Affordable Care Act, Pub. Law No.111-148, 124 Stat. 252, Title VII 7001(b).

government over a 10-year period have come up with estimates ranging from no (zero) savings up to $14 billion.[13] A 2009 Federal Trade Commission Report was not optimistic about the cost savings with the market introduction of biosimilars as compared with that of generic drugs.[14] Given the high cost of generating the required data to meet regulatory standards for biosimilars, it is likely that the market competition will "resemble brand-to-brand competition than the dynamics of brand-name competition under Hatch-Waxman."[15] In other words, cost differences between generic and innovator biologicals will not be as high as the comparable differences with nonbiologic drugs. The final answer to this question is likely to emerge when the savings are reported, as required by law.[16]

The purpose of this chapter is to introduce readers to the recently passed BPCI Act with special emphasis on the intellectual property (IP) provisions of the law, especially as they relate to patents and trade secrets. It begins with a brief legislative history to provide readers with an overview of the scientific, legal, economic, and political debate before the passage of BPCI Act. It is followed by a brief presentation of approval guidelines for biosimilarity to enable readers to obtain a better understanding of the scientific challenges associated with manufacturing biosimilars. Next, key IP provisions under the BPCI Act for innovator and generic manufacturers are presented. This is followed by certain legal issues that have been raised, such as challenges to the Food and Drug Administration's (FDA's) use of innovator data and postapproval applicability of safe harbor provisions. The chapter concludes with a future perspective that draws the attention of readers to certain widely debated IP and regulatory issues that will soon be or can be expected to be decided by the FDA and the Supreme Court. These decisions will likely have a major impact on the IP landscape and dialogue for biosimilars. Because scientists and other nonlawyers are expected to be a significant part of the reading audience, it is written with an explanatory tone to facilitate a better understanding of legal issues by these groups.

## 5.2 LEGISLATIVE HISTORY

A key point to be made is the diametrically opposite views of the two competing stakeholders—namely, the innovator and generic industries—over biosimilar approval. Although this can be expected, total disagreement over scientific issues, where there can be room for more agreement, may come as a surprise. For example, the effect of the manufacturing process on protein structure and function is a central

[13] Johnson, 2010.

[14] Federal Trade Commission. Emerging Health Issue: Follow-on Biologic Drug Competition, Washington, DC, June 10, 2009. Available from: http://www.ftc.gov/os/2009/06/P083901biologicsreport.pdf (henceforth, FTC 2009 Report).

[15] FTC 2009 Report, at p. iii.

[16] "The Secretary of the Treasury, in consultation with the Secretary of Health and Human Services, shall for each fiscal year determine the amount of savings to the Federal Government as a result of the enactment of [the BPCI Act]." Patient Protection and Affordable Care Act, Pub. Law No.111–148, Title VII, § 7003, Mar. 23, 2010, 124 Stat. 821.

question when it comes to biosimilars, which, as noted, are mostly complex molecules. During the preapproval debate over the BPCI Act, generic companies held the view that "[s]pecific aspects and parts of the manufacturing process do not determine the characteristics of a protein product"[17] because the "FDA had allowed innovator companies to make changes in manufacture, without subsequent clinical testing of the altered product."[18] In contrast, Pharmaceutical Research and Manufacturers of America (PhRMA) and Biotechnology Industry Organization (BIO) took the position that even minor changes in the manufacturing process can have a "potentially profound clinical impact".[19] The unintended consequences of this innovator industry position with respect to enablement of biological patents are discussed later in this chapter.[20] Therefore, these diametrically opposing views taken by the brand and generic industries with respect to biosimilars can be expected to result in considerable and extended patent litigation between these two segments of the pharmaceutical industry, as has been witnessed with litigation between these two entities under the 1984 Hatch-Waxman Act. A more exhaustive discussion of the history of the passage of the BPCI can be found elsewhere.[21] It also provides insights into the legal battles expected to come between and among the FDA and innovator and generic companies.

## 5.3 APPROVAL GUIDELINES FOR A GENERIC BIOLOGICAL PRODUCT

The FDA, which has been delegated the authority to develop appropriate rules and guidelines for the approval of biosimilars,[22] has yet to provide these details; they are expected in fiscal year 2013.[23] Although optional,[24] in February 2012, the agency provided three preliminary science-oriented guidance documents under the familiar major heading "Guidance for Industry." These "Draft Guidance" documents have the following self-explanatory subtitles: (1) "Biosimilars: Questions and Answers Regarding Implementation of the Biologics Price Competition and Innovation Act of 2009,"[25] (2) "Quality Considerations in Demonstrating Biosimilarity to a

[17] Letter from Suzanne M. Senbaugh, SICOR, Inc. to FDA, Docket No. FDA-2004-N-0059, November 12, 2004.
[18] *Id.*
[19] Letter from Sara Redcliffe, BIO to FDA, Docket no. FDA-2004-N-0059, December 13, 2004.
[20] See Section 5.5.1.2 (Lack of Patent Enablement).
[21] Carver KH, Elikan J, Lietzan E. An unofficial legislative history of Biologics Price Competition and Innovation Act. Food Drug Law J 2010;65:671–818.
[22] Regulation of biological products. 42 U.S.C. § 262.
[23] Johnson, 2010.
[24] 42 U.S.C. § 262 (k)(8).
[25] Food and Drug Administration. Guidance for industry on biosimilars: Q & As regarding implementation of the BPCI Act of 2009: questions and answers part I; 2012. Available from: http://www.fda.gov/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/ucm259809.htm.

Reference Protein Product,"[26] and (3) "Scientific Considerations in Demonstrating Biosimilarity to a Reference Product."[27] The practical implications of these recommendations for biosimilars manufacturers will crystallize only after the FDA issues its final rules.

The BPCI Act defines a generic biological product as a biosimilar[28] or an interchangeable.[29] The fact that an interchangeable biosimilar can be substituted for a reference product without the knowledge or consent of the prescriber could raise a conflict between state and federal laws on drug substitution. Currently, in most (if not all) states, generics for small molecule drugs cannot be substituted for the reference product if the prescriber checks the "Dispense as written" box on the prescription. It would appear that state pharmacy laws may have to be modified for interchangeable biosimilars because the dispenser, most likely a pharmacist, can under the BPCI Act make a substitution without the consent or knowledge of the prescribing physician. Some states have moved to remedy this uncertainty (see Section 5.6).

An applicant for a biosimilar product is required to provide the following supportive evidence:

1. Analytical data in support of similarity[30]
2. Preclinical (animal) studies[31]
3. Extensive clinical data[32]

[26] Food and Drug Administration. Guidance for industry: quality considerations in demonstrating biosimilarity to a reference protein product; 2012. Available from: http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM291134.pdf.

[27] Food and Drug Administration. Guidance for industry: scientific considerations in demonstrating biosimilarity to a reference product; 2012. Available from: http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM291128.pdf.

[28] "The term 'biosimilar' or 'biosimilarity'... means-(A) the biological product is highly similar to the reference product notwithstanding minor differences in clinically inactive components; and (B) there are no clinically meaningful differences between the biological product and the reference product in terms of safety, purity, and potency of the product" (42 U.S.C. § 262(i)(2)).

[29] "The term 'interchangeable' or 'interchangeability' in reference to a biological product that is shown to meet the standards [for a reference product] means the biological product may be substituted for the reference product *without the intervention of the health care provider who prescribed the reference product*" (emphasis added) (42 U.S.C. § 262(i)(3)).

[30] "[A]nalytical studies demonstrat[ing] that the biological product is *highly similar* to the reference product, not withstanding minor differences in clinically inactive component" (emphasis added) (42 U.S.C. § 262(k)(2)(A)(i)(I)(aa)).

[31] "[A]nimal studies (including the assessment of toxicity)" (42 U.S.C. § 262(k)(2)(A)(i)(I)(bb)).

[32] "[A] clinical study or studies (including the assessment of immunogenicity and pharmacokinetics or pharmacodynamics) that are sufficient to demonstrate safety, purity, and potency in 1 or more appropriate conditions of use for which the reference product is licensed and intended be used and for which licensure is sought for the biological product" (42 U.S.C. § 262(k)(2)(A)(i)(I)(cc)).

4. Mechanism of action data, when such information is known for the reference product[33]
5. Labeling information consistent with that approved for the reference product[34]
6. Route, dose, and dosage form information[35]
7. Assurance the manufacturing facility meets quality assurance standards[36]

An applicant is also required to provide information regarding the safety and efficacy of the reference product,[37] which the FDA makes public. The BPCI Act provides the applicant an option to provide additional publically available supportive information regarding the reference product or other, presumably related, products.[38]

An interchangeable biosimilar must, in addition to meeting the above requirements for a biosimilar product, also provide evidence that:

1. It produces the same clinical outcome.[39]
2. Under chronic use, the interchangeable product will be similar to the reference product in terms of safety and tolerance.[40]

It is not clear why there are two types of biosimilars, namely interchangeable and non-interchangeable biosimilars. Because the requirements for a biosimilar ensure similarity in outcome and almost all innovator biologicals and their corresponding biosimilars are more often than not administered "more than once," the additional (and confusing) requirements for interchangeability appears clinically unnecessary.[41]

[33] "[T]he biological product and reference product utilize the same mechanism or mechanisms of action for the condition or conditions of use prescribed, recommended, or suggested in the proposed labeling, but *only* to the extent the mechanism or mechanisms are known with the reference product" (emphasis added) (42 U.S.C. § 262(k)(2)(A)(i)(II)).

[34] "[T]he condition or conditions of use prescribed, recommended or suggested in the labeling proposed for the biological product have been previously approved for the reference product" (42 U.S.C. § 262(k)(2)(A)(i)(III)).

[35] "[T]he route of administration, the dosage form, and the strength of the biological product are the same as those of the reference product" (42 U.S.C. § 262(k)(2)(A)(i)(IV)).

[36] "[T]he facility in which the biological product is manufactured, processed, packed or held meets standards designed to assure that the biological product continues to be safe, pure, and potent " (42 U.S.C. § 262(k)(2)(A)(i)(V)).

[37] "An application . . . shall include publically-available information regarding the Secretary's previous determination that the reference product is safe, pure, and potent" (42 U.S.C. § 262(k)(2)(A)(iii)(I)).

[38] "[The applicant] may include any additional information in support of the application, including publicly-available information with respect to the reference product or another biological product." (42 U.S.C. § 262 (k)(2)(A)(iii)(II)).

[39] "[It] can be expected to produce the *same* clinical result as the reference product in any given patient" (emphasis added) (42 U.S.C. § 262(k)(4)(A)(ii)).

[40] "[F]or a biological product that is administered more than once to an individual, the risk in terms of safety or diminished efficacy of alternating or switching between use of the biological product and the reference product is *not greater* than the risk of using the reference product without such alternation or switch" (emphasis added) (42 U.S.C. § 262(k)(4)(B)).

[41] *Id.*

## 5.4 INTELLECTUAL PROPERTY PROTECTION UNDER THE BPCI ACT

The innovator enjoys dual monopolies for a limited time, which run in parallel. The first one is the protection afforded by market exclusivities. The second is via the traditional patent protection. Such dual IP protection is provided to incentivize innovation and marketing of new biological products.

### 5.4.1 Market Exclusivities

The reference product innovator is allowed an exclusivity period of 12 years from the date of approval.[42] This exclusivity does not apply to a supplement of the original reference product[43] or "to a subsequent application . . . for a change (not including a modification to the structure of the biological product) that results in a new indication, route of administration, dosing schedule, dosage form, delivery system, delivery device or strength"[44] or to a subsequent application . . . for "a modification to the structure of the biological product that does not result in a change in safety, purity or potency."[45] It appears that for these changes, the sponsor will have to seek patent protection and include them under the provisions provided to add newly issued or licensed patents[46]. In addition, as with the Hatch-Waxman Act, an additional 6 months of exclusivity is allowed if the innovator applicant provides clinical data in pediatric subjects.[47]

Furthermore, a generic product applicant is prohibited from filing an application for 4 years after the approval of the reference product.[48] From a practical standpoint, unless the FDA approval takes more than 8 years, this provision does not appear to be of additional benefit to the innovator given the 12-year market exclusivity.

There are incentives for the first manufacturer to market an interchangeable product with respect to the reference product. The first marketed generic

[42] "Approval of an application [for a generic biological product] may not be made effective by the Secretary until the date that is 12 years after date on which the reference product was first licensed" (42 U.S.C. § 262(k)(7)(A)).
[43] 42 U.S.C. § 262(k)(7)(C)(i).
[44] 42 U.S.C. § 262(k)(7)(C)(ii)(I).
[45] 42 U.S.C. § 262(k)(7)(C)(ii)(II).
[46] 42 U.S.C. § 262(l)(7).
[47] "If, prior to the approval of an application that is submitted [by the generic applicant] . . . the Secretary determines that information relating to the use of a new biological product in the pediatric population may produce health benefits in that population, the Secretary makes a written request for pediatric studies (which shall include a timeframe for completing such studies), the applicant *agrees* to the request, such studies are completed using appropriate formulations for each age group for which the study is requested within any such timeframe, . . . the periods for such biological product . . . are deemed to be . . . 12 years and 6 months. . . . " (emphasis added) (42 U.S.C. § 262 (m) (2)(A)).
[48] An application [for a generic product] may not be submitted for 4 years after the date on which the reference product was first licensed. . . . " (42 U.S.C. § 262(k)(7)(B)).

interchangeable biological product is granted one of several exclusivity protections, with the provision that the earliest of these protections applies:

1. A second or subsequent interchangeable product cannot be approved for sale for 1 year from the marketing of the first.[49]
2. When patent infringement action is brought against the first applicant for an interchangeable product, the exclusivity period is 18 months[50] after a final court decision[51] or "dismissal with or without prejudice of [patent infringement] action."[52]
3. The exclusivity period is 42 months after the approval of the first interchangeable product if patent litigation against the applicant is still ongoing.[53]
4. If the applicant is not sued, then the exclusivity period is 18 months from the approval of the first interchangeable product.[54]

### 5.4.2 Patent Protection and Infringement Issues

The BPCI Act appears to discourage patent litigation, as indicated by its provisions, which are quite different from those of the 1984 Hatch-Waxman Act. One big difference is that under this earlier act, the innovator asks the FDA to list all patents relevant to the innovator drug in the "Orange Book"[55] at the time of filing a New Drug Application (NDA). These relevant patents are called "sponsor patents." The generic drug applicant has the option to challenge these patent(s) on the basis of invalidity or non-infringement. Under the BPCI Act, there is no such listing of sponsor patents. Instead, the reference sponsor and the generic applicant are required to engage in lengthy prelitigation negotiations; included in these negotiations are technologies the sponsor is willing to license to the generic applicant. By the same token, the applicant is not required to accept these offers. Obviously, this provision is an attempt to minimize possible patent infringement litigation. This Act also limits the number of patents that can be litigated. These matters will be discussed in detail later in this section.

The patent negotiation process begins after the generic applicant is notified by the Secretary of Health and Human Services that the application has been accepted for review. Within 20 days after such notice,[56] the applicant must provide the reference product sponsor with information relating to the application and details of the

[49] 42 U.S.C. § 262(k)(6)A.
[50] 42 U.S.C. § 262(k)(6)B(i).
[51] "... the term 'final court decision' means a final decision of a court from which no appeal (other than a petition to the United States Supreme Court) for writ of certiorari has been or can be taken)." (42 U.S.C. § 262 (k)(6)).
[52] 42 U.S.C. § 262(k)(6)B(ii).
[53] 42 U.S.C. § 262(k)(6)(C)(i).
[54] 42 U.S.C. § 262(k)(6)(C)(ii).
[55] Orange Book: Approved Drug Products with Therapeutic Equivalence Evaluations (current through June 2013). (availabe at http://www.accessdata.fda.gov/scripts/cder/ob/default.cfm).
[56] "Not later than 20 days after Secretary notifies ... the applicant that the application has been accepted for review, the applicant shall provide the sponsor..." (42 U.S.C. § 262 (l)(2)).

manufacturing process(es)[57] and may optionally provide other information sought by the sponsor.[58] It does not appear there are any incentives for the applicant to comply with this last provision. After receipt of this information from the applicant for a biosimilar or an interchangeable product, the reference product sponsor shall provide the applicant with a list of relevant patents[59] and offers and licensing options, if any, within 60 days.[60] Reminiscent of Hatch-Waxman provisions,[61] the applicant has 60 days to respond to the sponsor addressing issues relating to invalidity and unenforceability of the listed patents or non-infringement of these patents by the applicant. In addition, the applicant should provide the sponsor with a response to its offer of licensing.[62] Interestingly, there is an optional self-incriminatory provision in which the applicant can admit to innovators' patents that it is likely to infringe.[63] Again, it does not appear the applicant has any incentive to respond to this option.

### 5.4.3 Patent Negotiations and Litigation Phase

Negotiations begin after the exchange of correspondence between the applicant and the sponsor regarding the validity, enforcement, or infringement of patents submitted by the sponsor. The two parties "shall engage in good faith negotiations to agree on which, if any, of the listed patents by the . . . applicant or the . . . sponsor shall be the subject of an action for patent infringement. . . ."[64]. If there is non-agreement between the parties within 15 days of negotiations, the next step is litigation.[65]

Before the start of litigation, the applicant for the generic biological product shall notify the reference product sponsor with the number of patents the applicant will submit to the sponsor.[66] This is the second chance for the applicant to provide such information to the sponsor, except that this time, it is not optional as in the first time.

[57] "[T]he applicant shall provide to the reference product sponsor a copy of the application submitted, . . . and other such other information that describes the process or processes used to manufacture the biological product. . . ." (42 U.S.C. § 262(l)(2)(A)).

[58] The applicant "*may* provide to the . . . sponsor additional information requested by or on behalf of the sponsor" (emphasis added) (42 U.S.C. § 262(l)(2)(B)).

[59] ". . . the . . . sponsor shall provide the . . . applicant a list of patents for which the . . . sponsor believes a claim of patent infringement could reasonably be . . . asserted by the . . . sponsor . . ." (42 U.S.C. § 262 (l) (3) (A) (i).

[60] ". . . an identification of the patents on such list that the . . . sponsor would be prepared to license to the . . . applicant." (42 U.S.C. 262(l)(3)(A)(ii)).

[61] Food and Drug Law and Regulation, Editors: Adams DG, Cooper RM, Hahn, MJ and Kahan JS, FDLI, Washington DC (2008), pp. 365–368.

[62] ". . . the applicant shall provide the . . . sponsor [with] a response [to patents the sponsor would be prepared to license], . . . under subparagraph (A)(ii)." (42 U.S.C. § 262(l)(3)(B)(iii)).

[63] ". . . the applicant may provide to the sponsor a list of patents to which the . . . applicant believes a claim of patent infringement could reasonably be asserted by the . . . sponsor if a person not licensed by the . . . sponsor engaged in making, using, offering to sell, selling or importing into the United States of the biological product that is subject of . . . the application" (42 U.S.C. § 262(l)(3)(B)(i)).

[64] 42 U.S.C. § 262(l)(4)(A).

[65] "If within 15 days of beginning negotiations, . . . the . . . applicant and the . . . sponsor fail to agree on a final and complete list of which, if any, patents listed . . . by the . . . applicant or the . . . sponsor shall be the subject of an action of patent infringement . . ." (42 U.S.C. § 262(l)(4)(B)).

[66] 42 U.S.C. § 262(l)(5)(A).

Note that the applicant is required only to submit the number of patents and not a list.[67] After this notification, further actions shall proceed as follows: On a mutually agreed date, but within 5 days after such notification by the applicant, the two parties shall "simultaneously exchange"[68] a list of patents. Such an exchange is a deterrent to modifying patents of interest, based on knowledge of the other party's patents. The list includes patents that each party believes are in contention.[69] The number of patents a sponsor can list cannot exceed the number of patents listed by the applicant.[70] If the applicant does not include any patents in its list, the sponsor is limited to a single (i.e., one) patent that can be litigated. The sponsor must initiate infringement action within 30 days, irrespective of the fact that the sponsor and the applicant have agreed[71] or not agreed[72] on the patent(s) to be litigated. The applicant must report and provide a copy of the complaint to the Secretary of Health and Human Services,[73] who makes it public knowledge.[74]

The sponsor is allowed to add newly obtained or licensed patents to the list (mutually agreed or not) within 30 days of obtaining such patents[75] with notification to the applicant. Upon receipt of this notification, the applicant shall, *inter alia*, inform the sponsor about the invalidity or non-enforceability of the added patent(s), or non-infringement of the new patent(s) by the applicant's product, or state that its product will not be marketed before the expiration of this newly added patent(s). This

[67] *Id.*

[68] 42 U.S.C. § 262(l)(5)(B)(i).

[69] ".... the list of patents that the applicant believes should be the subject of an action for patent infringement..." (42 U.S.C. § 262(l)(5)(B)(i)(I)) and " the list of patents that the sponsor believes should be the subject of an action for patent infringement..." (42 U.S.C. Section § 262(l)(5)(B)(i)(II)).

[70] 42 U.S.C. § 262(l)(5)(B)(ii).

[71] "If...the...applicant and the...sponsor agree on patents...not later than 30 days after such agreement, the...sponsor shall bring an action for patent infringement with respect to each such patent." (42 U.S.C. § 262(l)(6)(A)).

[72] "If [the is failure to reach agreement between the parties]...not later than 30 days after the exchange of lists....the...sponsor shall bring an action for patent infringement with respect to each patent that is included in such lists" (42 U.S.C. § 262(l)(6)(B)).

[73] "Not later than 30 days after a complaint is served...to a[n] applicant in an action of patent infringement...the applicant shall provide the Secretary with notice and copy of such compliant." (42 U.S.C. Ï262 (l)(6)(C)(i)).

[74] "The Secretary [of Health and Human Services] shall publish in the Federal Register notice of a complaint..." (42 U.S.C. § 262(l)(6)(C)(ii)).

[75] "In the case of a patent that (A) is issued to, or exclusively licensed by, the...sponsor after the date that the...sponsor provided the list to the...applicant...and (B) the...sponsor reasonably believes that, due to the issuance of such patent, a claim of infringement could reasonably be asserted by the...sponsor if a person not licensed by the...sponsor engaged in making, using, offering to sell, selling, or importing into the United States the biological product that is the subject of the...application, and not later than 30 days after such issuance or licensing the...sponsor shall provide the...applicant a supplement to the list provided [earlier] by the...sponsor that includes such patent...." (42 U.S.C. § 262(l)(7)(A) and (B)).

part of the response by the applicant is identical to the response during earlier patent negotiations.[76]

The applicant is required to notify the sponsor within 180 days before proposed marketing of its generic product.[77] Upon such notice, the sponsor is allowed to seek a preliminary injunction[78] to stop the intended marketing until the matter is resolved in the courts.[79] The courts examine the relevant patents for validity, enforceability, and infringement by the generic applicant.[80] Both parties "shall reasonably cooperate to expedite such further discovery as is needed in connection with the preliminary injunction motion."[81]

Incentives are provided to the generic applicant to follow procedures provided in the BPCI Act. If the applicant provides the sponsor with a copy of the biosimilar application containing all of the details as described,[82] both parties are prohibited from bringing a declaratory judgment action[83] for patent infringement, validity, or enforcement until the generic applicant notifies the sponsor about its intent to market the generic.[84] This action applies only to patents that were presented by both parties during negotiations.[85] If the applicant fails to provide the sponsor with the applications, the penalties may be severe in that only the sponsor "may bring an action . . . for a declaration of infringement, validity or enforceability of *any* patent that claims the biological product or use of the biological product (emphasis added)."[86] Additional incentives are also provided for the applicant to comply with provisions of the BPCI Act relating to patents.[87] If the applicant fails to state that (1) each of the patents originally submitted by the sponsor, including any new patent(s), are either invalid

[76] The applicant "shall provide to the . . . sponsor . . . a detailed statement that describes, on a claim by claim basis, the factual and legal basis of the opinion of the . . . applicant that such a patent is invalid, unenforceable, or will not be infringed by the commercial marketing of the biological product. . . ." (42 U.S.C. § 262(l)(3)(B)(ii)(I)).

[77] "The . . . applicant shall provide notice to the . . . sponsor not later than 180 days before the date of the first commercial marketing of the biological product. . . ." (42 U.S.C. § 262(l)(8)(A)).

[78] An order from the court whereby the generic applicant is restrained from marketing the product. "An injunction commands an act that the court regards as essential to justice, or it prohibits an act that is deemed to be contrary to good conscience. It is an extraordinary remedy. . . ." TheFreeDictionary. Injunction. Available from: http://legal-dictionary.thefreedictionary.com/injunction.

[79] "After receiving [such] notice and before such date of the first commercial marketing of such biological product the . . . sponsor may seek a preliminary injunction prohibiting the . . . applicant from engaging in the commercial manufacture or sale of such biological product until the court decides the issue of patent validity, enforcement, and infringement with respect to any patent that is (i) included in the list provided by the . . . sponsor, [including newly issued or licensed patents] or in the list provided by the . . . . applicant . . . ) (42 U.S.C. § 262(8)(B)(i) and (ii)).

[80] 42 U.S.C. § 262(l)(8)(B).

[81] 42 U.S.C. § 262(l)(8)(C).

[82] 42 U.S.C. § 262(l)(2)(A).

[83] "A statement by the court declaring the rights and duties of the parties in a case or stating an opinion on a question of law without awarding relief. This judgment is binding." Gilbert Law Summaries Pocket Size Law Dictionary. Harcourt and Brace Company; Orlando, Florida 1994.

[84] 42 U.S.C. § 262(l)(9)(A).

[85] 42 U.S.C. § 262(l)(8)(B)(i) and (ii).

[86] 42 U.S.C. § 262(l)(9)(C).

[87] 42 U.S.C. § 262(l)(9).

or unenforceable; (2) the marketed product will not infringe any of these patents; or (3) marketing of the applicant's product will not commence before the expiration of these patents or fails to notify the secretary of a notice of litigation, then the sponsor can immediately commence a declaratory judgment action with respect to validity, infringement, or enforceability of the pertinent patents.[88]

## 5.5 LEGAL ISSUES

One major legal issue focuses on a central question: How legal is the use of the sponsor's data by the FDA in approving a competitor's generic version of the sponsor's biological product? Clearly, the discretionary power of the FDA under the BPCI Act allows the agency to use innovator data for approval of biosimilars.[89] The anticipated litigation is most likely to focus on the extent of such use allowed by law. The following discussion provides some insight into these matters.

### 5.5.1 Food and Drug Administration Use of Sponsor's Data in Generic Biosimilar Approval

Trade secrets submitted by the reference sponsor under the Biologics License Application (BLA) are considered proprietary and protected by federal trade secret laws from disclosure. It is interesting to note that the FDA's definition of "trade secret" has evolved over the years. The current definition reads as follows: "A trade secret may consist of any commercially available plan, formula, process, or device that is used for the making, preparing, compounding, or processing of trade commodities, and that can be said to be the end of product innovation or substantial effort. There must be a direct relationship between the trade secret and the productive process."[90] The older definition included wording to protect unfair competitive edge.[91] This change was implemented after a court[92] refused to accept the agency's broader definition of a trade secret; it used a narrower definition, which was adopted by the FDA. Trade secrets are also protected from disclosure under the FDA's rules governing the Freedom of Information Act[93] and prohibited acts.[94] The FDA's past decisions also support the belief that information is proprietary.[95] In one case, a

[88] 42 U.S.C. § 262(l)(9)(B) and (C).
[89] 42 U.S.C. § 262(k)(2)(ii).
[90] 21 C.F.R. § 20.61(a).
[91] "A trade secret may consist of any formula, pattern, device, compilation of information which is used in business, and gives him an opportunity to obtain an advantage over competitors who do not know it or use it," 21.C.F.R. 20.61(a) (1982).
[92] Pub. Citizen Health Research Group v. FDA, 704 F.2d 1280 (D.C. Cir., 1983).
[93] 21 C.F.R.§ 20.61(c) .
[94] "The using by any person to his own advantage or revealing, other than to the Secretary or officers or employees of the Department, or to the courts when relevant in any judicial proceeding under this Act, any information acquired . . . concerning any method or process which is a trade secret is entitled to protection; . . . " (21 U.S.C. § 331(j)).
[95] *Tri-Bio Lab., Inc. v. United States,* 836, F.2d 135, 141.

manufacturer asked FDA to approve its generic version of a veterinary drug by incorporating information submitted earlier by the innovator for the same drug. The FDA refused to consider the application based on "its policy [of] treating such information as proprietary and confidential."[96] The court agreed with the agency's decision. But this court made the distinction that internal use of data provided by the reference sponsor to approve a generic product does not constitute disclosure.[97]

It would be informative, and perhaps also predictive of FDA's approach to biosimilar approvals, to examine certain provisions of Section 505(b)(2) of the Federal Food, Drug, and Cosmetic Act, under which biologicals have been approved. These provisions are quite similar to those of the BPCI Act in that it allows a later applicant to rely on a previous (innovator) applicant's safety and efficacy data. It should be noted that a 505(b)(2) application is considered a new drug application. Section 505(b)(2) states:

> An application [may be] submitted under . . . section 505 (b)(2)for a drug for which the [safety and effectiveness] investigations . . . relied upon by the applicant for approval of the application were not conducted by or for the applicant and for which the applicant has not obtained a right of reference or use from the person by or for whom the investigations were conducted . . . and [shall] also include . . . a certification to each patent which claims the drug for which such investigations were conducted or which claims a use for such drug for which the applicant is seeking approval.

A couple of examples are presented to illustrate the nature of innovator challenges to the FDA regarding such reliance. One case involved the potential marketing of amlodipine maleate (Amvaz) by Dr. Reddy's Laboratories under the Section 505(b)(2) paradigm. This applicant chose to rely on Pfizer's NDA, which was for the besylate salt of amlodipine (Norvasc). Pfizer filed a Citizen Petition asking the FDA to "immediately revoke its acceptance for filing and receipt of [the application] and/or deny [it's] approval" on the argument that FDA could not approve the Amvaz application using the Norvasc NDA application because that would constitute unauthorized use of proprietary information. In response, Dr. Janet Woodcock, director of the Center of Drug Evaluation Research, denied the petition and stated that no proprietary information is used in Section 505(b)(2) decision making.[98]

Other innovator companies have also petitioned the FDA challenging information sharing under a Section 505(b)(2) application. In a second instance, Genentech petitioned the FDA to "refrain from taking steps that prejudice [Genentech's]

[96] "Any reference to information furnished by a person other than the applicant may not be considered unless its use is authorized in a written statement signed by the person who submitted it," 21 C.F.R. § 514.

[97] "Because the Trade Secret Act . . . prohibits only public disclosure of application data, it does not bar internal agency use of submitted data," *Tri-Bio Lab., Inc. v. United States,* 836, F.2d 135, 141.

[98] "FDA may rely on its earlier conclusions regarding safety and effectiveness to whatever extent the conclusions are appropriate for the drug under review in the 505(b)(2) application. Although reliance of an FDA finding of safety and effectiveness for an NDA is certainly indirect reliance on the data submitted in the original NDA, reliance on the conclusions supported by that data is not the same as manipulating those data to reach new conclusions not evident form the existing approval," Docket No. FDA-2003-P-0014.

property rights in trade secret and confidential commercial data and information provided by the agency for the limited purpose of reviewing and approving [Genentech's'] products."[99] In the matter of the Omnitrope application filed by Sandoz, the petitioner (Pfizer) raised both scientific and legal concerns. Of interest to the discussion here are the legal aspects of the challenge: In its petition, Pfizer stated: "It is . . . legally improper for FDA to rely on, reference, or otherwise use the clinical and manufacturing information that establishes the safety and effectiveness of Genotropin . . . to approve Omnitrop[e]." In a collective response to several petitioners, the FDA did not address the issues relating to ". . . . the legality of relying on the approval or licensing of another product to approve or license follow-on protein products, . . . [t]rade secret and Constitutional law (Fifth Amendment takings clause) arguments based on the use of trade secret data and information (e.g., chemistry, manufacturing, and controls (CMC)) data in one approved application to review and approve another application for a follow-on protein product . . ."[100] and stated that "the FDA has not and does not need to review or use any trade secret information from the approved Genotropin application, or any other application, to approve the Omnitrope NDA."[101] In its lengthy explanation in support of denial, the agency also provided a broader assurance to the petitioner that it had "not compromised the protections offered under the law to [innovative] data and information."[102]

Citing two provisions of the BPCI Act, one author has concluded that "[t]he question whether FDA is authorized to examine and consider the information in the reference product application, when reviewing biosimilar applications should therefore be regarded as fully settled by the statute itself: FDA may not do so."[103] One of the provisions is the requirement that the applicant can only use nonproprietary information in its application. The second one is that the FDA can only use "the information submitted in the application (or a supplement)" in its decision making. To add to the debate, another commentator, citing a 2005 article, stated: "A follow-on biologic manufacturer must use the pioneer biologic's manufacturing process to demonstrate interchangeability. . . ."[104] There is no such assertion in the BPCI Act; whether it is a biosimilar or an interchangeable, the requirements are that ". . . . the facility in which the biological product is manufactured, processed, packed, or held, meets standards designated to assure that the biological product continues to be safe, pure and potent."[105]

[99] Letter from Stephen G. Juelsgarrd, Genentech, Inc. to FDA, Docket No. FDA-2004-P-0214.
[100] Letter from Stephen K. Galson, M.D., M.P.H., Director CDER to Kathleen M. Sanzo, Esq., Docket No. FDA 2004P-0231.
[101] *Id.*
[102] *Id.*
[103] Epstein RA. The constitutional protection of trade secrets and patents under the Biologics Price Competition and Innovation Act of 2009. Food Drug L J 2011;66:285–328, 291 (hereinafter Epstein).
[104] Montas A. Cheaper clinical trials: the real solution to the biologic industry's Gordian Knot. Am J Med 2011;37:172–193, 183.
[105] 42 U.S.C. § 262(k)(2)(A)(V).

With regard to the assertion,[106] the history of generic approval legislation and implementation shows FDA reliance on innovator information. This is reflected in the agency's response to the Omnitrope and other petitions. It appears that the agency will not respond to legal matters, leaving such challengers no option but the courts. This might be because the FDA believes it can win in court because of the agency favorable Supreme Court decision in the *Chevron*[107] case.

The Court held the following: "First, always, is the question whether the Congress has spoken directly to the precise question at issue. If the intent of Congress is clear, that is the end of the matter; for the court, as well as the agency, must give effect to the unambiguously expressed intent of Congress. . . . if the statue is silent or ambiguous with respect to a specific question, the issue for the court is whether the agency's answer is based on a permissible construction of the statute."

The phrase any "permissible construction" of the law is indeed a low threshold for the FDA to meet, if challenged. This might explain why Pfizer did not further contest the FDA's denial of its petition in the Omnitrope case (*supra*).

In one author's opinion, if anything is "settled,"[108] it is that the FDA can use and has used data submitted by the innovator in approving a competitor's generic product. The FDA has repeatedly stated that no proprietary information has or will be used. This remains a contentious issue between the industry and the agency.

***5.5.1.1 Takings Issue*** Concerns have been expressed regarding the possible use by the FDA of proprietary information submitted by innovator companies in the approval of biosimilars. Some specifics are discussed later (see Section 5.6). This section presents a brief summary of pertinent legal issues. The uncompensated use of an innovator's trade secret information may be unconstitutional under the Fifth Amendment of the United States Constitution.[109] Two Supreme Court cases have to be considered in analyzing the takings issue. In one case,[110] the Supreme Court ruled that if the law warns the sponsor that confidential information submitted for regulatory purposes will be disclosed, then there is no takings issue because the data was submitted voluntarily by the sponsor.[111] It appears that BPCI Act provides such notice, even if it is indirect.[112] The second issue deals with the matter of compensation. In the second Supreme Court case,[113] a similar regulatory situation involving the EPA had to deal with the agency using the sponsor's pesticide safety and efficacy data to approve a generic version. Because the Federal Insecticide, Fungicide, and Rodenticide Act required generic manufacturers to compensate the sponsor, the Supreme Court ruled that this Act did not violate the takings clause.

[106] Epstein (*supra*).
[107] *Chevron U.S.A. Inc. v. Natural Resources Defense Council, Inc.*, 467 U.S. 837 (1984).
[108] Epstein (*supra*) at 294.
[109] "... nor shall private property be taken for public use, without just compensation."
[110] *Penn Central Transportation Co. V. City of New York,* 438 U.S. 104, 124 (1978).
[111] This is a disingenuous argument because the sponsor has no choice but to submit the required data for market approval.
[112] 42 U.S.C. § 262(k)(2)(A)(ii).
[113] *Ruckelshaus v. Monsanto Co.,* 467 U.S. 986, 1000-01 (1984).

Given the long history of the approval of drugs and biologicals, it appears that court challenges on the takings would favor the FDA. As noted, that might be the reason why Pfizer did not challenge the agency's denial of its Omnitrope petition (*supra*). In addition, it could be argued that market exclusivities under the BPCI Act provide just compensation for use of an innovator's trade secret by the FDA to approve a generic.

***5.5.1.2 Lack of Patent Enablement*** This is a second legal issue for biological products in general. Brand name companies, by their statements about the general inabilities of generic companies to produce biosimilars, may have inadvertently raised the issue of enablement, with respect to validity of existing biological patents and future patentability of biologicals.[114] Patent law requires that any patent for a product (in this case, a biological) disclose sufficient information for a skilled individual to be able to manufacture that product (i.e., enablement).[115] In addition, "undue" experimentation on the part of this skilled individual to make the biological product would not meet the enablement requirement.[116] The *Genentech, Inc. v. Novodisk* case involved the validity of US Patent 5,424,199, which deals with a recombinant human growth hormone. This patent was declared "invalid for failure of the specification to meet the enablement requirement . . . "[117]

***5.5.1.3 Safe Harbor Provisions*** This issue is related to patent infringement. It is not clear if the "safe harbor" provisions under the Hatch-Waxman Act, which apply to small molecule drugs, would also apply to biologicals. Under these provisions, a generic drug manufacturer is allowed to use patented information without infringement for preparing its Abbreviated New Drug Application (ANDA) filing to the FDA.[118] A strict reading of the statutory language suggests that these provisions may not apply to biosimilars. However, before the passage of the BPCI Act, proteins, now included in the definition of a biological product, had been approved under the

[114] Mandel GN. The generic biologics debate: industry's unintended admission that biotech patent fail enablement. Virginia J. Law & Tech. Vol. 11, No. 8 , 1-25, Fall 2006.

[115] "The specification shall contain a written description of the invention, and of the manner and process of making and using it, in such full, clear, concise, and exact terms as to enable any person skilled in the art to which it pertains, or with which it is most nearly connected, to make and use the same, and shall set forth the best mode contemplated by the inventor of carrying out his invention," 35 U.S.C. § 112 Specification (1994).

[116] In re Vaeck, 947, F.2d 488 (Fed. Cir., 1991).

[117] 108 F.3d 1361, 1368.

[118] "It shall not be an act of infringement to make, use, offer to sell, or sell within the United States or import into the United States a patented invention (other than a new animal drug or veterinary biological product (as those terms are used in the Federal Food, Drug, and Cosmetic Act and the Act of March 4, 1913) which is primarily manufactured using recombinant DNA, recombinant RNA, hybridoma technology, or other processes involving site specific genetic manipulation techniques) *solely* for uses reasonably related to the development and submission of information under a Federal law which regulates the manufacture, use, or sale of drugs or veterinary biological products." (emphasis added) (35 U.S.C 271(e)(1)).

FDCA (Food, Drug, and Cosmetic Act), blurring the difference between these two classes of therapeutic agents. Furthermore, similarities in statutory definitions of a drug[119] and biological product[120] are not helpful in resolving this issue.

A recent decision by the Federal Court of Appeals involving the generic version of the anticoagulant drug Lovenox (generic name, enoxaparin) would suggest that biosimilar manufacturers may also be protected under these provisions.[121] Enoxaparin, similar to many biologicals, is not a single entity; it consists of a mixture of oligosaccharides of varying molecular weights and compositions derived from the chemical breakdown of the complex drug heparin (itself a mixture of polysaccharides of different molecular weights and compositions) obtained from pig intestines. Momenta is the assignee of a patent[122] that is directed toward measurement of sulfated polysaccharides of varying molecular weights. Amphastar had used technology covered by this patent in preparing its FDA filing (i.e., ANDA) for approval for its generic version of Lovenox. Two days after the FDA approved Amphastar's generic version, Momenta, which markets its own generic version of Lovenox, filed a patent infringement suit against Amphastar.[123] The district court ruled that Amphastar had infringed Momenta's patent and granted Momenta a preliminary injunction, thus preventing market entry of Amphastar's generic version of Lovenox. Amphastar appealed, arguing that the district court had taken "an unduly restrictive view of the safe harbor... [provisions] of 35 U.S.C. § 271(e)(1)."[124] The appeals court ruled that these provisions protect both pre- and postapproval activities. Preapproval activities relate to information use of patented information to prepare an ANDA, and postapproval activities refer to the use of the patented information after the ANDA is granted. For continued marketing of its generic product, "Amphastar is required [by the FDA] to conduct a laboratory determination of the identity and strength of the active ingredient for each batch of enoxaparin."[125] Therefore, the appeals court concluded that Amphastar's postapproval activities were covered by the safe harbor provisions because "[h]ere, the submissions are not 'routine submissions' to the FDA..."[126] In a previous case, this court had ruled that these safe harbor provisions "[do] not apply to information that may be routinely reported to the

[119] Drugs are defined as "...articles intended for use in the diagnosis, cure, mitigation, treatment, or prevention of disease in man or other animals" or " articles (other than food) intended to affect the structure or any function of the body of man or other animals...." (21 U.S.C. § 321(g)(1)).

[120] A "biological product" is defined as a "virus, therapeutic serum, toxin, antitoxin, vaccine, blood, blood component or derivative, allergenic product, protein (except any chemically synthesized polypeptide) or analogous product, or arsphenamine or derivative of arsphenamine (or any other trivalent organic arsenic compound) applicable to the prevention, treatment, or cure of a disease or condition of human beings," 42 U.S.C. § 262(i)(1).

[121] *Momenta Pharmaceuticals Inc. v. Amphastar Pharmaceuticals, Inc.* Nos. 12-1062, -1103, -1104 (Fed. Cir., August 3, 2012). (hereinafter, *Momenta*)(*available at http://www.cafc.uscourts.gov/opinions-orders/search/report.html, then search keyword Momenta*).

[122] United States Patent No. 7,575,886.

[123] *Momenta.*

[124] *Id.* at 9.

[125] *Id.* at 18.

[126] *Id.* at 18.

FDA, long after marketing approval has been granted."[127] Upon the advice of the Solicitor General of the United States, the Supreme Court denied a petition for certiorari (i.e., to hear the case) in the *Classen* case. Stay tuned. This matter is not over. Momenta plans to appeal the federal circuit decision to the Supreme Court.[128]

## 5.6 FUTURE PERSPECTIVES

This concluding section presents several important and unresolved issues littering the biosimilar IP landscape. In this author's opinion at this time (February 2013), the major uncertainties facing innovator and prospective biosimilar manufacturers can be grouped under three major headings: regulatory, legal, and political. From a regulatory perspective, it was only recently (*supra*) that the FDA started providing preliminary regulatory and scientific guidances to the industry for biosimilar approval under the BPCI Act. As might be expected given the complexity of biological drug products (compared with small molecule drugs), the FDA will initially take a "case-by-case" approach to approval of biosimilar products. Understandably, it will take several years for the various stakeholders connected with the biotechnology and pharmaceutical industries, especially marketers of innovator and generic products, to fully understand the FDA's regulatory enforcement scheme and to develop strategies for individual success in a profit-driven environment. Given the high cost of biological product development, the amount of nonclinical and clinical information required by the FDA of a prospective biosimilar applicant to prove similarity or interchangeability to an innovator product will be one big deciding factor for market introduction of biosimilars. If such requirements are deemed excessive vis-à-vis profitability by the industry, obviously biosimilar development will be adversely affected.

A second matter pending before the FDA relates to the constitutionality of the use innovator data by the FDA to approve generics. The Citizen Petition filed by Abbott Laboratories on April 2, 2012, asserts that the BPCI Act does not apply to innovator biologicals filed before the passage of this act on March 10, 2010. The petition asks the FDA to:

> [C]onfirm that it will not accept for filing, file, approve, or discuss with any company, or otherwise take any action indicating that the agency will consider, any application or any investigational new drug application (IND) for a biosimilar that cites, as its reference product, BLA 125057 for Humira® (adalimumab) or any other product for which the biologics license application (BLA) was submitted to FDA prior to March 23, 2010, the date on which the BPCIA was signed into law.[129]

[127] *Classen Immunotherapeutics, Inc. v. Biogen IDEC,* 659 F.3d 1057, 1070 (Fed. Cir., 2011).

[128] Momenta. U.S. Court of Appeals denied Momenta Pharmaceuticals petition for rehearing en banc in "safe-harbor" case; 2012. Available from: http://ir.momentapharma.com/releasedetail.cfm?releaseid=722630.

[129] Abbott Laboratories Covington and Burling LLP Citizen petition, FDA Docket No. 2012-P_0317. Available from: http://www.regulations.gov/#!documentDetail;D=FDA-2012-P-0317.

Because Abbott "had no notice, or reasonable expectation, that the agency would use its trade secrets to approve another company's product," when the BLA for Humira was submitted to the FDA in 2002, the company believes that such uncompensated use would be unconstitutional under the Fifth Amendment[130] (issue discussed earlier). As of October 1, 2012, the FDA had not decided on this petition.[131] Given the importance of this issue, it is very likely that any decision by the FDA will be legally challenged. In this connection, it is interesting to note the Patient Protection and Affordable Care Act, which includes the BPCI Act, was found constitutional by the Supreme Court.[132] This could weaken the takings argument.

Another regulatory challenge for the FDA is for an applicant to take a "short cut" to biosimilar approval by filing it as 351(a) application, where it would enjoy the benefits of an abbreviated filing of a 351(k) application; such filings have been referred to as "skinny BLAs."[133] There is some evidence in support of such a strategy. Teva admits that its product, Neutroval (a granulocyte colony-stimulating factor) is a biosimilar to Neupogen, marketed by Amgen. However, Teva filed a 351(a) application. The FDA claims it has no authority to prevent such filings,[134] but others have insisted that the agency does have such authority.[135] It is anticipated that the debate on this subject will continue, especially if others also chose such filings to bypass the BPCI Act.

The legality of "reverse payments," in which a generic company agrees not to compete with an innovator until its patents have essentially expired, is also an unresolved matter of general interest to the drug and biotechnology industries. The Federal Trade Commission (FTC) has estimated that such "pay for delay agreements" will cost "American consumers $3.5 billion per year—$35 billion over the next 10 years."[136] After disagreements between the Third[137] and Eleventh[138] Circuit Courts, the Supreme Court heard, on March 25, 2013, oral arguments on the question posed by petitioner FTC in its brief[139]: "Whether reverse-payments are per se lawful unless the underlying patent litigation was a sham or the patent was obtained by fraud, or instead are presumptively anticompetitive and unlawful?" The Supreme Court ruled on June 17, 2013 that reverse payments are in violation of

[130] *Id.*

[131] "FDA has been unable to reach a decision on your petition because it raises complex issues requiring extensive review and analysis by Agency officials." Letter from Jane A. Axelrod, Associate Director for Policy, Center for Drug Evaluation and Research, FDA, to Erika Lietzan, Covington and Burling LLP.

[132] National Federation of Independent Business et al. v. Sebelius, Secretary of Health and Human Services, et al. Available from: http://www.supremecourt.gov/opinions/11pdf/11-393c3a2.pdf.

[133] Epstein (*supra*) at 325.

[134] FDA Week (December 17, 2010).

[135] Epstein (*supra*) at 326.

[136] Federal Trade Commission. Pay-for-delay: how drug company pay-offs cost consumers billions; 2010. Available from: http://www.ftc.gov/os/2010/01/100112payfordelayrpt.pdf.

[137] *Federal Trade Commission v. Watson Pharmaceuticals, Inc.* 677 F.3d 1298 (Eleventh Circuit) (the FTC lost this antitrust case making such payments legal).

[138] In re K-Dur Antitrust Litigation, 686 F.3d 197 (Third Circuit) (this court disagreed with the opinion of the Eleventh Circuit). Available from: http://sblog.s3.amazonaws.com/wp-content/uploads/2013/01/12-416tsUnitedStates.pdf.

[139] Federal Trade Commission v. Watson Pharmaceuticals, Inc., et al. (http://www.justice.gov/atr/cases/watson.html).

antitrust laws (FTC v. Actavis, Inc., 570 U.S. ____2013). Therefore, innovator companies would find it difficult to delay the entry of biosimilars into the market through pay for delay arrangements with generic manufacturers.

Because biological products are often derived from natural substances, it has raised a fundamental question: Do they meet the legal requirements of patentability?[140] The "hot" debate over the patentability of naturally occurring substances (e.g., genes) has finally reached the Supreme Court. On April 15, 2013, it heard oral arguments on the question "Are human genes patentable?"[141] in the famous *Myriad* case. The trial court had ruled that the *BRAC1* and *BRAC2* genes (and perhaps, genes in general) are not patent-eligible subject matter;[142] this decision was reversed by the Federal Court of Appeals, which maintained that isolated genes are patentable.[143] A second question for arguments heard on April 15 was "Did the court of appeals err in upholding a method claim by Myriad that is irreconcilable with [the Supreme] Court's ruling in Mayo Collaborative Servs. v. Prometheus Lab., 132 S. Ct. 1289 (2012)?" In the *Prometheus* case, the Supreme Court ruled that a diagnostic test involving measurement of 6-thioguanine in blood for possible drug dosage adjustment is not patentable subject matter. The Supreme Court decided on June 13, 2013 that isolated genes (genomic DNA) are not patentable because they are products of nature. However, complementary DNA, where introns have been removed from the DNA, is patentable. The impact of this much awaited decision on biologicals remains to be seen. One recent commentary (Rai et al., Science, 341, 137–138, 2013 (July 12)) stated that the impact of this decision can be expected to be "modest", since "[many] human gDNA patents have expired or will do so soon."

The general applicability of the safe harbor provisions to biological drug development, especially postapproval activities, remains unclear. The *Classen* and *Momenta* cases (*supra*) have let open the possibility of Supreme Court intervention, given statements made by the Solicitor General in the *Classen* case.[144]

Finally, a relatively new (political) hurdle for the biosimilar industry is state-level lobbying by innovator companies. Some of these companies (e.g., Genentech,

[140] 35 U.S.C. 101 "Whoever invents or discovers any new and useful process, machine, manufacture, or composition of matter, or any new and useful improvement thereof, may obtain a patent therefore, subject to the conditions and requirements of this title." 35 U.S.C. 101; "Laws of nature, natural phenomena, and abstract ideas are not patentable" Diamond v. Diehr, 450, U.S. 303, 309 (1981).

[141] http://www.americanbar.org/publications/preview_home/12-398.html (under Merit Briefs, click on Brief for Petitioners).

[142] Association for Molecular Pathology, et al. v. United States Patent and Trade Mark Office, et al., 702 F. Supp. 2d 181.

[143] Association for Molecular Pathology, et al., v. United States Patent and Trade Mark Office, 689 F.3d 1003 (August 16, 2012).

[144] "It is an open question whether Section 271(e)(1) applies at all to all patented research methods of the kind at issue in this case." Amicus Brief of the Solicitor General of the United States in the *GlaxoSmithKline v. Classen Immunotherapies, Inc.* at p. 20. Available from: http://www.justice.gov/osg/briefs/2012/2pet/6invit/2011-1078.pet.ami.inv.pdf.

Eli Lilly) have initiated efforts in Virginia, North Dakota, and Indiana to reduce biosimilar market share by limiting the ability of pharmacists to make generic substitutions for innovator biologicals.[145] Such efforts are being opposed by generic and insurance companies. Although early, both sides appear to have been successful. The Virginia House of Delegates passed such a law by a vote of 91 to 6, and the Indiana House Public Health Committee "approved a bill, but law makers responding to objections from the generic association, removed the requirement that the patient consent to any substitution." [146] Because, in general, patient consent will be dictated by price and biosimilars can be expected to less expensive (otherwise the very goal of the BPCI Act would be in vain), the need for such a requirement seems unwarranted. As a recent newspaper report put it, "The trench fighting at the state levels is the latest phase in a battle over the rules for adding competition to the biotechnology drug market. . . . "[147] As noted (*supra*), the BPCI Act allows for such substitution without the knowledge of the subscriber,[148] and state laws govern dispensing of drugs and biologicals.

In conclusion, it would not be surprising that there will be many regulatory challenges and much litigation as has been observed with approval of generics under the 1984 Hatch-Waxman Act; patent challenges under provisions of this Act increased fivefold between 2003 and 2009. To date, no application for a biosimilar has been submitted to the FDA.[149] However, ample incentives for biosimilar manufacturers will soon become available because biological patents worth an additional $47 billion in sales will expire between 2015 and 2020.[150] The impact of the BPCI Act on the biologicals industry will begin to crystallize only as the FDA starts to review and decide on applications for biosimilars and interchangeables and as patent and regulatory challenges are resolved by the courts.

[145] Pollak A. Biotech firms, billions at risk, lobby states to limit generics. The New York Times. Available from: http://www.nytimes.com/2013/01/29/business/battle-in-states-on-generic-copies-of-biotech-drugs.html?_r=0.
[146] *Id.*
[147] *Id.*
[148] 42 U.S.C. § 262(i)(3).
[149] Malkin BJ. Challenges to the development of biosimilars industry in the United States. Thomson Reuters/Aspatore; 2012.
[150] Available from: http://www.gabionline.net/Biosimilars/General/US-67-billion-worth-of-biosimilar-patents-expiring-before-2020.

# Scientific Aspects of Sterility Assurance, Sterility, Asepsis, and Sterilization

JAMES E. AKERS and JAMES P. AGALLOCO

## 6.1 INTRODUCTION

This chapter provides introductory-level information relating to sterile product manufacturing and the control of microbiological contamination in that endeavor. However, many of the principles discussed here are relevant to any process in which the control of microbiological contamination is paramount. Microbiology is a broad and intensely complex science, and unfortunately, novices often fail to grasp the complexity of the discipline and further fail to appreciate the diversity of the microbial world. So diverse are the organisms in the science of microbiology that they fall into two primary phylogenetic kingdoms, the prokaryotes and the eukaryotes. In fact, the incredible diversity of the microbial world is such that some organisms, viruses, are neither prokaryotes nor eukaryotes and instead are obligate intracellular parasites possessing genetic material and a limited complement of functioning enzymes but lacking any innate metabolic apparatus. Although in sterile product manufacturing, the focus is primarily on the bacterial and mycologic world (prokaryotes and eukaryotes), this is most assuredly not true in the case of some products of biological origin in which contamination with adventitious viruses could be problematic.

Fortunately, only a tiny percentage of the microorganisms with which we humans share the earth have shown any capability to produce disease. The overwhelming majority of organisms found on our bodies and in our environment are completely harmless commensals or even organisms with which humans have a symbiotic relationship.

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*,
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

## 6.2 CONTAMINATION CONTROL

In the control of microbial contamination, there are only two primary methods available. The manufacturer can use systems intended to prevent microbial contamination from entering the production pathway or it can use chemical or physical means in the form of heat, chemicals, or irradiation to kill microorganisms. Some manufacturing processes, of course, use a combination of microbial exclusion and microbial destruction. It is necessary for industrial microbiologists to have sufficient understanding both of processes for microbiological exclusion and destruction. Of course, it is also useful to understand environmental circumstances that favor microbial colonization or survival because, given the ubiquity of microbes in nature, complete exclusion of microbes is an impossible rather than merely a difficult task to accomplish.

## 6.3 FUNDAMENTAL CONCEPTS IN "STERILE" PRODUCT MANUFACTURING

Two separate, distinct, and very different methods are used to manufacture health care products that carry the word "sterile" on their labels. In the method called *terminal sterilization*, the product is sterilized in its final container before commercial release. Terminal sterilization performed with any reasonable degree of technical competence will result in a product that is very unlikely to be nonsterile, which is to say that each and every unit within a lot will be free of viable microorganisms. Terminally, sterilization will render any microbial contamination that might be present fully inert with a high degree of certainty. Importantly, this certainty is readily and very reliably quantifiable. The manufacturer can, with sufficient accuracy ensure reliability and reproducibility, measure the delivered lethality using thermographic data or some form of dosimetry depending on the method of sterilization used. So, it can be stated that precise analytical and monitoring methodologies can be applied to terminal sterilization, ensuring lot and unit sterility, and that these methods can be applied in process and during validation as required.

The second and more common method used in the manufacture of products labeled "sterile" is known as *aseptic processing*. As an industry, we would like to think that both terminal sterilization and aseptic processing attain the same endpoint, which, of course, is "sterility," and that both meet the regulatory expectation for the process control requirement that has come to be known as "sterility assurance." However, these two approaches, although aimed at the same apparent goal, have enormous differences obvious at both the practical and theoretical levels.

## 6.4 THE EVOLUTION OF ASEPTIC PROCESSING TECHNOLOGY

During a period that extended from the 1970s until the late 1980s, there was only one general method by which aseptic manufacturing was accomplished, and that was

using gowned human operators in large-scale classified suites of cleanrooms known as *aseptic processing areas*. It should be noted that during this period, there was considerable performance variability among aseptic processing operations relating to cleanroom design, the types of products manufactured, and the level of automation used. By the late 1980s, separative technologies such as isolators began to appear, and the past two decades have also seen dramatic increases in the capability of automation systems, further decreasing industry's reliance on personnel. Although there is solid evidence that aseptic processes have improved in terms of "sterility assurance," the health care industry is confronted with decisions and regulatory challenges that did not exist a quarter of a century ago.

Aseptic processing, although essential because of the large (and increasing) number of products that will not withstand terminal sterilization treatments, is a vastly more complex discipline than it was a comparatively short time ago. Among the reasons for this complexity are:

1. Multiple technologies in operation simultaneously: Industry at present is using conventional cleanrooms, isolators, restricted access barriers (RABs), and various combinations thereof with varying levels of automation.
2. There has never been a clear understanding within the regulations regarding what was possible in aseptic processing and perhaps more importantly what is impossible analytically. The idea that microbiological control when humans are involved is measurable in some quantifiable manner and therefore can be validated, although understandable in regulatory terms, is a scientifically formidable (and probably impossible) task.
3. There remains a poor understanding of the probabilistic nature of "sterility assurance" as it relates to process capability. There also exists an unfortunate tendency to believe that definitive means are available to prove that the state called sterility exists in each and every unit in each lot of product.
4. As a corollary to item 3, there is a persistent belief that it is possible to measure levels of sterility and know when sterility does or does not exist in a lot based on microbiological analytical methods. Although it may seem logical that methods used to detect microorganisms in a product or the environment should make the assessment of sterility possible, in reality, this capability continues to escape our grasp. The sheer diversity of microorganisms makes simple and easily applied assays for sterility a technical impossibility. Tests to prove sterility would require the ability to test each and every unit within a lot. This is of, course, both impractical and unrealistic. Testing of the environment is both statistically limited and unreliable and more significantly is not a direct test on the condition of the product but rather an assessment of transient contamination within the environment in which the product is made or assembled.
5. When confronted with concerns regarding the adequacy of growth-based analytical microbiological technology, there is the unfounded expectation that rapid or instrumental microbiological methods will obviate these uncertainties and actually allow for the measurement of "sterility." Sadly, there are

no instrumental methods on the horizon that could enable us to measure the sterility of individual units directly in a noninvasive or nondestructive manner. Hence, the advantage of rapid methods is largely speed of analysis as opposed to a lower limit of detection. The measurement of sterility would require an assay with perfect sensitivity combined with the ability to take a sample of infinite size. The technical challenge of proving the negative absolute that is sterility is an enormous one indeed; scientists and engineers working in this discipline must keep this in mind.

6. There is an expectation created by a combination of regulatory suggestion and vendor-generated "buzz" that genotypic microbial identification methods can provide a near infallible and therefore extremely reliable means of tracking down routes and sources of contamination. Reality in this case falls well short of perfection, and in fact, some rapid assays, because they are not growth based, make identification difficult or even impossible regardless of the identification method used. The contamination of an article with nucleic acid may be indicative of the presence of contaminants at some point in the manufacturing process; however, this does not mean that the material contains infectious organisms or microbial toxins. Certainly, in some investigations, genotypic identification has proven valuable; in others, it has provided no value. This is not surprising given the diversity of organisms among the microflora and the differences in distribution of these organisms that arise from multiple sources of contamination. Thus, there is no certainty that an airborne contaminant isolated from one environment may match an isolator recovered from filtration of the product. A single human operator may be colonized by multiple varieties of the same organisms. Vegetative cells released into an aseptic processing environment may not remain viable for more than a few hours, making the results of resampling activities uncertain.
7. New process control concepts are being added to the industry workload at an unprecedented rate. Concepts such as quality by design, process analytical technology, quality risk management, and various statistical process control methodologies may all have merit, but it is unclear how they apply to sterility assurance and how they correlate to preexisting and still required validation and monitoring programs. Readers are urged to consider each of these process control and management methodologies independently. Each of these concepts is sufficiently complex to require a complete chapter in and of itself. The minimization of risk in sterile product manufacturing requires that the product be safe, and this requires that the risk of microbial contamination in any product labeled sterile must be extremely low—in fact, lower than that of any other product type. That being said, the absolute elimination of microbiological risk in the administration of any product labeled "sterile" cannot be proven.

Although the concept of sterilization (or terminal sterilization of a finished product) in which a defined input for validation can be chosen and attained is

generally easier to understand, there are difficulties, both microbiological and statistical, that must be confronted. The mere fact that the term "sterility assurance," typically given as "sterility assurance level" (SAL), is applied to both exclusionary and lethal processes causes confusion. This has led to attempts to add the SAL results from aseptic processing and sterilization processes together. This seems quite reasonable mathematically given the apparent common unit of measure, but it is far from reasonable microbiologically.

## 6.5 STERILITY AND ASEPSIS: TOWARD A COMMON SCIENTIFIC UNDERSTANDING

It is clear in any discussion of sterility assurance that words really do matter and that when we talk of such things as viability, monitoring, sterile, sterilization, and aseptic, we have to be working from a common and well-accepted lexicon if we are to have any hope of reaching a broad and rather uniform level of understanding. These words appear in every regulatory standard, regulatory guidance document, industry standard, and international performance standard relating to sterile product manufacturing and those for microbiological controlled or restricted bioburden processes as well.

The authors have frequently asked training audiences to define two words that are used frequently the aforementioned documents, namely "aseptic" and "sterile." There have been a surprisingly wide range of definitions offered. Many individuals who attempt to define these words end up concluding that they are more or less synonymous. This could be because in many regulatory documents, the manner in which these words appear leads readers to conclude that in the regulatory context and therefore in practical usage, they are synonymous. In fact, the Food and Drug Administration's (FDA's) latest guidance document on aseptic processing contains the construction "sterile drug products produced by aseptic processing" in its title [1].

However, these words are not synonymous; there are very clear scientific distinctions between the words "aseptic" and "sterile." It is only when the real technical differences between the conditions these words describe are acknowledged that a path can be found to a more rational standard setting, regulation, and day-to-day process control.

Frequently, when the terms "sterile" and "aseptic" are under discussion, it is said that the differences in meaning between these words are not technically significant and therefore of only a semantic nature. One definition found in *Webster's New Collegiate Dictionary* for general semantics is "the language used in political propaganda or advertising to achieve a desired effect on an audience" [2]. However, "sterile" and "aseptic," when used in a regulatory sense, *should* be used with a clear and unambiguous *scientific* understanding. We must at all times avoid mistaking semantic conveniences for scientific fact.

If our industry is to have harmonized and universally applicable standards for products bearing the label "sterile," it is essential that we agree on a common lexicon.

Nearly a quarter of a century ago, the 1987 *FDA Guideline on Sterile Drug Products Produced by Aseptic Processing* implied in its title that aseptic processing can achieve a truly sterile result [3]. This construction was repeated in the title of the most recent version of this guideline published in 2004 [1]. This time interval should have been sufficient for thought to congeal around this fundamental concept of relatedness between "sterile" and "aseptic."

This implication of the confusion can be further examined by considering the definitions medical microbiologists apply to the word "sterile." Not surprisingly, although there is quite a lot of agreement on the meaning of the word "sterile," even the definition of this absolute term is not without conflict. In Jawetz et al.'s *Medical Microbiology*, "sterile" is defined as "Free of life of every kind. Sterilization may be accomplished by filtration (in the case of liquids or air) or by treatment with microbiocidal agents. Since the criterion of death for microorganisms is the ability to reproduce, sterile material may contain intact metabolizing microbial cells" [4]. It can be argued that the authors of this definition are having it both ways because they clearly state that "sterile" means free of life and then go on to define a condition in which metabolizing organisms are present. It is probably more reasonable to conclude that the authors have simply realized that the continuation of life requires the ability to reproduce. In fact, the term "sterile" is often applied to advanced multicellular organisms that lack the ability to reproduce.

However, one can clearly see the problem this definition poses for the detection of a "sterile" condition. If metabolic activity alone is not an indicator of non-sterility, does this not imply that test methods that rely on only biochemical activity cannot prove or disprove that living (reproducing) organisms are, in fact, present? Certainly, the condition defined as sterility cannot practically rule out the possibility that all evidence of the possible existence of microorganisms be absent from a preparation. Although there are few data on the subject, it is not unreasonable to think that genetic remnants of microorganisms lacking the ability to reproduce might indeed be present in a container labeled "sterile," perhaps even a material subjected to a sterilization process. Thus, it seems clear that a genetic screening method that yielded a signal but that does not allow the ability to discern whether the DNA came from a living or dead organism would not serve well as a "sterility" test. Certainly, such a signal might give useful information about potential contamination sources in an input material but could neither prove nor disprove "sterility."

Zinsser's *Medical Microbiology* text defines sterilization as "an absolute (term) that implies the total inactivation of all forms of microbial life in terms of the organism's ability to reproduce" [5]. These authors, perhaps wisely, steered clear of considerations regarding metabolic and simply based their definition upon the ability to reproduce.

*Microbiology* by Davis et al., a common text used in graduate microbiology and medical schools, defines "sterile" as "the use of either physical or chemical agents to eliminate all viable microorganisms" [6]. This definition avoids the debate about metabolism and growth but introduces the term "viability" instead. Fortunately, the authors do define viability, which they state is the "ability of the organism to

propagate indefinitely." Hence, their definition is again based on the ability of an organism to reproduce.

Although other microbiology texts and dictionaries offer slightly differing perspective regarding the definition of "sterility," all are in general agreement that the destruction of an organism's ability to replicate is central to the concept of sterilization and hence the attainment of the condition of "sterility." This is logical because it is quite clear that organisms that lack the ability to reproduce are incapable of causing an infectious disease, which, of course, is the risk modality by which microbial contamination produces adverse patient effect. Interestingly, it follows from this line of reasoning that viable but nonculturable organisms are in fact "sterile" unless they are actually put into conditions in which they can grow. It can be argued that the description "viable but nonculturable" is an oxymoron unless one expands the definition of "viable" to include metabolic activity, which is not justified, in the authors' opinion. Metabolic activity is an unreliable indicator of either the attribute of viability or the condition of "sterility." However, microbiologists are aware that an organism may fail to grow in one set of conditions (e.g., nutrition, oxygen level, incubation temperature) but grow quite well in other environmental circumstances. Does this mean that "sterility" can in fact be a conditional attribute? Yes; although this is a scientific reality, it is clear this is difficult from both a regulatory and legal perspective because it falls short of the expected (or hoped for) absoluteness.

## 6.6 STERILITY, VIABILITY, AND VIRULENCE

The ruling out of metabolic activity as an indicator of viability is essential. Among the microorganisms that are present in the environment and cause human and animal disease are a large number of obligate intracellular parasites, most notably viruses. Viruses are marvelously adapted parasites that lack any inherent metabolic activity and simply take over the metabolic and macromolecular synthesis machinery of their host in order to effect replication. Therefore, a material could be both free of detectable growth on all bacterial or mycological media and have no metabolic activity and still be "nonsterile."

In pure scientific terms, whereas a sterile product is one that contains no organisms capable of reproducing, an aseptically produced product is one that will not produce infection. An aseptically produced product may be sterile in the strictest sense, or it may not; however, at all times, it is expected to be noninfectious.

If we consider the term "sterility" a bit more critically, while properly an absolute, in reality it is situational! Although viable organisms may be present in a material, they would only be able to replicate if appropriate growth conditions were present. In terms of infectious disease, the term "virulence" is used to describe the ability of an organism to replicate in a susceptible host and cause infection. According to Zinsser's *Microbiology*, virulence is "a quantitative measure of the degree of pathogenicity of a particular microorganism. Virulence is usually measured by

the numbers of microorganisms necessary to kill or alter a particular organism under standardized test conditions" [7].

Not surprisingly, virulence has been and continues to be an area of very active research within the general study of host–parasite relationships. Virulence depends not only on the route of exposure and the number of organisms to which a host is exposed but also on the general health of the host organism studied. Virulence is far too complicated a subject to be covered in depth in a relatively brief chapter, but some general information can be introduced. Of the many thousands of species of microorganisms known, only about 300 have been reported to have the ability to produce disease [7]. It is staggering to think that there may be statistically more microbial species not yet known than those that are known and that there are by some estimates roughly $10^{30}$ viable microbiological cells present on earth at any given time. On the average human, there are roughly 10 bacterial cells per human cell, which means the risk of disease from microbial infection is remarkably low considering their ubiquity and astonishing diversity [8]. This, of course, means that the virulence of most environmental organisms is zero, and fortunately, among those that produce disease, virulence is, in the case of most pathogenic species, generally quite low as well. Readers may find this information surprising because in the pharmaceutical industry, the fear of microorganisms is such that there is a tendency to imagine that infectious disease is a distinct likelihood in the wake of any microbial contamination event.

Even among pathogenic organisms, which are those known to have some level of virulence in humans or animals, it can require a substantial contamination insult to result in a measurable pathological effect. Finlay and Falkow reported in *Microbiological Reviews* that inoculation of healthy volunteers intradermally with virulent bacteria could require as many as 5 to 10 million organisms to initiate an infection [9]. Even extremely virulent organisms such as *Francisella tularensis* may typically require more than 10 organisms capable of replicating to initiate infection. Studies on viruses have shown that typically several hundred to several thousand infectious virions (individual viruses) are required to initiate an infection; however, more recent studies indicate that around 20 Norovirus (genus *Calicivirus*) particles can initiate an infection [10].

A potentially pathogenic organism, even in a weakened host, faces a typically hostile, well-defended environment, and thus the barriers to initiation of infection are high. Thus, it is not surprising that, for example, the terrorist attacks with *Bacillus anthracis* (even though they involved millions of specially purified and weaponized spores designed to disseminate in the environment) produced very limited morbidity and mortality.

In no way do is this meant to imply that the data concerning microbial virulence mean that the health care industry in general or the pharmaceutical industry specifically can or should take a relaxed attitude toward the dangers of microbial contamination. Laxity with respect to contamination control is something that should be diligently avoided. Rather, it is possible and reasonable for industry and regulators to take a more scientific and hopefully more realistic approach to risk analysis and management regarding microbial contamination. Or to put it another

way, overreaction to microbiological risk and the resulting emphasis on absolute proof of contamination risk mitigation can introduce very significant cost increases with no real benefit to the end user of the product. In fact, in some circumstances, it can result in a quite safe product being rejected, which is wasteful and perhaps even dangerous to overall population health should the product be in short supply. The question should be what are realistic expectations for the production and dissemination of "sterile" health care products? Scientifically enlightened regulation would then not consume resources by focusing on comparatively low-risk elements in the product supply chain but rather on elements known to produce a greater risk of patient risk as measured by actual injury.

## 6.7 REGULATIONS AND ASEPTIC CONTAMINATION RATES

There is substantial ambiguity in the regulations concerning the ability of manufacturers to produce "sterile" products. For example, although one can find numerous suggestions in government regulations and guidelines that a target of zero contamination in media fills is appropriate, one cannot find any guideline or standard that completely rules out the likelihood of any single process simulation test containing positive units. Parenteral Drug Association (PDA) Technical Report No. 22, *Process Simulation Testing for Aseptically Filled Products*, states in the section on acceptance criteria and interpretation of results that:

> despite the number of units filled during a process simulation test or the number of positives allowed, the ultimate goal for the number of positives for any process simulation should be zero. A sterile product is, after all one that contains no viable organisms. There are, however, numerous technical problems in achieving this goal . . . thus, a limit of some low number other than zero is often chosen [11].

The PDA position on this point is realistic and scientifically defensible.

This statement and similar statements in other guidance documents, however, need further dissection if the full context of their meaning is to be grasped. There is an unfortunate assumption often made regarding media fill testing, which is that a zero result means that the ability of the process to make "sterile" product has been validated. There are two reasons why this assumption must never be made.

First, the sensitivity and limit of detection of media fill tests are impossible to establish. "Sensitivity" refers to the range of environmental organisms that could be grown on the media fill test media system. Put another way, this is the number of species that could be grown and recovered. This sensitivity data are really impossible to derive and are no doubt highly variable because they depend significantly on the condition of organisms in the environment. Many, if not most, organisms in the environment are nutritionally and chemically stressed to one degree or another, and by many estimates, the number of nonculturable organisms might exceed those that are actually culturable under some condition. Also, the limit of detection of the media fill test is quite a lot higher than zero. Fertility or growth promotion testing on

media fill test media uses inocula in the range of 10–100 CFU. Lower inoculum levels run the risk of negative growth results. Therefore, the limit of detection is greater than 0 even among organisms well adapted to grow in the soya bean casein digest broth typically used, and not all organisms likely to be in the environment will be grown and recovered by this medium.

Second, it is microbiologically incorrect to think that media fill tests can in any way demonstrate the complete absence of viable organisms; they can only show that there are not enough viable, nutritionally adapted organisms present to initiate growth on the chosen medium. As previously discussed, no bacteriologic media can recover viruses, and some viruses, as we all know, are important pathogens. Thus, the phrase "sterile media fill," which is a common linguistic construction in pharmaceutical industry, is really only true with clear and very significant qualifications. These qualifications are in fact so significant in nature that for scientific veracity, the term "sterile media fill" should never be used. It certainly is true that there were not enough cells in the environment that entered into each negative container to produce visible growth; however, this does not come at all close to proving the absolute that is "sterility" by any definition of that word.

It is reasonable at this point for readers to ask two questions:

1. Why don't we try to create media fill conditions more conducive to the recovery of any organism that might be present in the environment?
2. Why don't we require a growth promotion or fertility test outcome that shows growth at an inoculum much lower than 10 CFU and preferably at 1 CFU?

The answers in both cases are the same–because doing so is simply not possible. With a great deal of study, it might be possible to find a media that would recover a broader spectrum of organisms, but doing so would no doubt cause us to lose the ability to detect some organisms that are currently detected. Indeed, sometimes firms do conduct media fills using what are often called anaerobic media. (It should be noted true or "strict" anaerobic conditions are not a realistic possibility in an aseptic manufacturing facility.) Technically, anaerobic organisms do not require oxygen for growth; in fact, strict anaerobes do not tolerate the presence of any oxygen, meaning air is poisonous to them. Therefore, these anaerobes would be recovered only in highly reduced chemical conditions in which $O_2$ is quantitatively absent. So, strict anaerobic conditions are unlikely to exist in any production facility; what is far more likely to exist are conditions in which the presence of $O_2$ is significantly reduced. So, really what so-called anaerobic media fills do is bias the recovery toward organisms that prefer to grow at low $O_2$ levels, which are known to microbiologists as microaerophiles; they may also be known as facultative anaerobes or aerotolerant. Of course, conditions of growth media and incubation favoring these organisms will disfavor the growth of organisms requiring the presence of greater amounts of $O_2$. For this reason, the need for media fill tests done under reduced $O_2$ tension has always been a debatable point. Some industry authorities favor the idea, but others do not. What can be safely said is that neither standard media fills nor so-called

anaerobic tests will grow *everything*, but they will probably grow *different things*. Also, there will be a subset of environmental organisms adaptable enough to grow in both conditions. We must then conclude that the value of specific media fills done under so-called anaerobic conditions is limited because it seems reasonable to consider the media fill not as an absolute measure of anything but rather as a generalized evaluation of microbiological control conditions at a specific point in time.

The authors realize that what we have suggested in this section of the chapter is a significant reduction in the value many firms and some regulatory investigators place upon the media fill test. We are suggesting simply that there is a correct level of valuation or emphasis to be placed on the media fill. We have observed an increasing tendency to think of media fill tests as true validation events, which we believe is scientifically and statistically wrong. There has been a tendency to put the media fill test on a plane with biological challenges performed in the support of sterilization process validation, which is scientifically inappropriate. It is not possible to calculate a probability of nonsterility or SAL from a media fill test. Only a maximum contamination rate can be determined, and even then it must be considered that no test could recover all potential microbial contaminants. A legitimate claim for a media fill test with zero visible contamination is that an insufficient number of organisms capable of growth under the situational nutritional conditions were present and hence no growth was observed. We hope readers will agree that this falls well short of the claim of "sterile media fill" and certainly does not support a claim that a SAL can be derived from the number of units filled in any given test. It must then follow that it is also wrong to add the total number of media fill units on a given aseptic production line and then claim that the SAL is equal to that contamination rate value. Perhaps most importantly, it is a serious mistake to think that conducting any given intervention or series of interventions in multiple media fill tests and observing no growth means that these interventions are "validated."

The authors have proposed a risk analysis method regarding aseptic processing that we believe can be more accurate than a media fill in predicting microbial contamination potential [10,11]. This method, referred to as the Akers-Agalloco method, is beyond the scope of this chapter. However, some firms have used it successfully, and some evidence indicates that it can be predictive of weaknesses in contamination control [12]. Other aseptic processing risk evaluative methods may be effective and should, in our view, be considered as an adjunct to media fill testing. The limitations of microbiological analysis as a measure of the attribute of sterility have been poorly recognized in industry, although we believe this is slowly changing. The limitations of microbiological analysis are not limited to media fill testing but rather apply to such highly valued activities as environmental monitoring and sterility testing as well. A concept of acceptable commercial sterility or safety would be better applied to aseptic production than absolute sterility. The idea that we should expect every unit in an aseptically product lot to be sterile, when it is not scientifically possible to prove or even expect such an outcome, is wrong.

## 6.8 CONSIDERATIONS REGARDING ENVIRONMENTAL MONITORING

In the previous section, it was pointed out that current international regulatory guidelines allow that a media fill test might occasionally have a positive unit. In fact, under some circumstances in validation, a media fill with a single positive might not be disqualifying. This is a clear admission that at some level the authorities understand that "sterility" in aseptic processing is neither absolute nor a universally attainable outcome. Similarly, we know of no standard or guideline that suggests that an environment must be completely free of all detectable organisms at all times. Most guidelines, including those issued by the FDA and the European Medicines Agency, suggest that to pick one sample type, samples taken in ISO (International Organization for Standardization) 5 environments should show that airborne viable contamination averages less than one viable colony per cubic meter. Therefore, both media fill acceptance criteria and environmental monitoring acceptance criteria allow that some low level of recoverable microbial growth is likely to occur at least occasionally.

To summarize the key points covered so far, the following conclusions can be drawn:

- "Sterility" is an absolute concept, and the act of sterilization is one that renders all organisms present in or on a material incapable of reproducing.
- A material need not be "sterile" to be noninfectious, and the risk of infection from a small number of organisms is in most cases quite low.
- The number of species known to be capable of producing infectious disease is a very small subset of the constellation of known microorganisms.
- The existing standards and guidelines may suggest that aseptically produced products are "sterile," but there is realization, as found in the words of PDA's TR No. 22, that there are "numerous technical problems in achieving this goal." In fact, careful reading of the standards and guidelines confirms that they allow for nonsterile environments and imperfect media fill results.

We are then brought face to face with a central problem in aseptic processing. If we cannot truly prove that "aseptic" products are "sterile," and, in fact, our standards recognize that contamination in process simulation testing and environmental sampling is likely, why do we persist in calling these products "sterile?" A further exploration of the concept of asepsis is required to adequately consider this question.

As previously mentioned, it appears that much of the pharmaceutical industry and governing competent regulatory authorities considers "aseptic" and "sterile" to be more or less synonymous. However, in academic science, as reflected in common technical usage, it is clear that in the science of microbiology, "sterility" and "asepsis" have never had the same technical meaning.

The word "sepsis" is derived from the Greek verb meaning "to decay or to rot." The prefix "a" means "without." Therefore, in the purest sense, the word "asepsis"

means "the elimination of decay." In the medical sense, the word "sepsis" has come to mean something very close to infectious. The Jawetz *Medical Microbiology* text defines condition of sepsis or septic as "the presence of pathogenic microbes in living tissue" [4]. It follows then that "aseptic" is defined as "the absence of *pathogenic* microbes." Thus, in the clearest technical sense, an "aseptic" product is certainly safe because it lacks infectivity, but it may not be "sterile" in the absolute sense of that term [6]. So, if products produced in industrial aseptic processing were indeed infectious and therefore routinely contained organisms capable of producing infection, we should expect significant morbidity and perhaps even some mortality associated with this subset of "sterile" products. However, there is no evidence that there are any significant risk associated with industrial aseptic processing conducted in compliance with standards that have been promulgated in the past two decades.

The authors suggest that PDA dealt with this contentious issue in a scientifically honest manner in the current version of that association's Technical Report No. 22 [10]. Perhaps PDA's statement has gotten little notice because it is found in "Appendix 2—Definitions" rather than the main body of the technical report. PDA defines "sterility" as "free of viable organisms;" however, below the definition is a note that says "in practice, no such absolute statement regarding the absence of microorganisms can be proven." However, we believe that many thoughtful readers would hasten to point out that scientific evidence and the opinion of PDA regarding the provability of "sterility" in aseptic processing matters very little if the regulatory authorities hold industry to a standard of absolute sterility [1]. In a practical sense, we would have to agree with that observation, although we are required to point out again that there is no current standard in force that appears to consider "sterility" in cleanrooms possible, even using the technically limited measure of environmental recovery and subsequent growth of contamination [13,14]. We point out again that it is not appropriate to consider a zero outcome to be the equivalent of "sterile." We cannot prove a negative absolute, so we must reiterate that zero does not mean "sterile"; it only means that on the occasion of this test or sample, no organisms that would grow to a point where they could be detected by inspection were present.

It is not appropriate to simply state that contamination in aseptic processing areas is a fact of life; in the interest of scientific balance, we should look at the technical literature with regard to the effects of human occupancy on cleanrooms. First, we should examine the level of skin contamination present on humans. Data published in Ljungqvist and Reinmüller's *Ventilation and Airborne Contamination in Cleanrooms* indicates that the total population of aerobic bacteria on human skin is typically greater than 1.2 million/$m^2$ in the head and neck regions of both male and female subjects [15]. The numbers of organisms present on the hands and arms is typically in the range of 0.9 to 3 million/$m^2$ on healthy subjects. A number of studies have been done on the release of total particulates by humans in cleanroom clothing [16,17]. In the same publication referenced above, a study conducted by Takasago Thermal Engineering found that a fully gowned person sitting in a cleanroom would release about 15,000 particles per minute. A walking individual, according to this study, releases roughly 157,000 total particulates per minute. In a study published by Reinmüller in 2001, it was reported that the typical ratio between total particles

larger than 0.5 μm and viable aerobic organisms recovered is in the range of 600 to 7000 to one [15]. Studies done on personnel clothed in new full coverage cleanroom gowns found that these people release 600 to 1300 total particulates per hour in the larger than 0.5-μm size range and that among these were as many as 40 CFU of viable aerobic organisms. These data are generally consistent with the findings published by Dr. William Whyte [13]. Reinmüller's data also showed that as the gowns aged and were subjected to washing and resterilization, both the number of total particulate and the level of microbial contamination increased. Not surprisingly, these studies indicate that there is also a correlation between the amount of physical activity undertaken by personnel as a source of contamination, or put another way, the amount of contamination they contribute to their surrounding environment. Reinmüller also reported that microbial contamination is strongly associated with particulates in the 0.5-μm size class. It has been widely assumed that microorganisms in aseptic processing areas are associated with particulates in the 5- to 10-μm size range, and possibly the emphasis placed on particulates in this size class in some regulatory recommendations is an upshot of this assumption. However, there appears to be little factual basis for an emphasis on 5-μm particulates, and given the low frequency at which particles in this size range are recovered, there seems little justification for using them in a classification exercise because the counting statistics are poor.

From the work of Whyte, Reinmüller, and others, it can be estimated conservatively that a typical, properly gowned cleanroom worker will contribute 10 to 100 CFU of viable aerobic organisms to the environment per hour [13,14,15,16]. It is obvious then, given the high air exchange (which results in correspondingly high contaminant dilution rates) used in today's ISO 5 cleanrooms, that the total recovered microbial concentration should be very low. However, these data also clearly indicate that even the best cleanrooms do not under any circumstance approach "sterility." In fact, it could be argued that they can only attain asepsis if we are fortunate enough to have facilities staffed only by personnel who do not release pathogens, a circumstance that is quite beyond our ability to prove.

Let us relate these experimental data to a frequently discussed and quite current regulatory issue. It has recently been suggested in regulatory findings that firms should test their most critical process equipment for the presence of microorganisms and take action against product if viable organisms are recovered on samples taken from a production operation. Because operators must in many aseptic operations physically assemble production equipment such as filling pumps and needles and because people release microorganisms during work in cleanrooms, it clearly should not be considered unreasonable to recover low-level contamination in and around areas where people work. In fact, given the variability associated with the sample and recovery methods and what we know of their sensitivity, the expectation is that some relatively low frequency of contamination would be observed.

The stopper bowl is frequently cited as a critical surface that must be tested, and it was been suggested that contamination found on this surface is indicative of a loss of sterility assurance. However, because the recharging of stopper feed hoppers is typically a manual operation, it seems unreasonable to believe that the occasional

recovery of organisms should be considered indicative of a loss of control. We are unaware of any research done on gowned cleanroom employees that indicates they do not consistently shed viable organisms. In fact, from the data published in refereed journals, it appears that release of organisms by personnel is not only inevitable but also completely normal. Therefore, to suggest that remedial action must always follow the recovery of organisms at any location in a cleanroom is wrong. It would also follow that requiring investigations to be done whenever an isolate is recovered in an ISO 5 room will only create busy work rather than enhance process control. Investigations into circumstances that are expected and normal are of no real operational value.

It is clear given the research and empirical data that a low but random level of recovery is completely normal in manned aseptic processing and therefore noteworthy only if there is a significant change in level or incidence rate. Scientific thoroughness requires us to consider than in almost every case, environmental monitoring is conducted by a gowned member of the staff and that this sampling activity is also prone to some level of adventitious contamination in the same manner as it might be derived from production personnel.

So, it is unreasonable to think that the discovery of organisms that one should expect to be present in low numbers represents a loss of sterility assurance. It is scientifically and statistically more appropriate for industry standards and regulations to recognize that "sterility" cannot be assured in any cleanroom under normal operational circumstances. This is true even if we define "sterility" as merely the absence of viable organisms, which is a very liberal definition of the word. Operations that carefully and scrupulously comply with the letter and spirit of current Good Manufacturing Processes (cGMPs) cannot be expected to yield uniformly "no growth" results, which clearly means that some recovery of viable organisms in any human-scale cleanroom is in fact *normal*. As long as people are present, they are the most likely source of microbial contamination, and because there is nothing they can realistically do to completely eliminate microbial and particulate shedding, it follows that no exceptional assignable cause can be identified in an investigation.

Because the presence of personnel in the environment always ensures that contamination will be present, it can be argued that it is the total absence of contamination that would reflect an exceptional condition! Certainly, it is expected that the recovery of viable organisms from critical ISO 5 or EU grade A environments should be observed at a very low rate. However, that very low rate will always be greater than zero, which implies that a background level of contamination is present, which—not surprisingly—our intensive monitoring efforts sometimes recover.

## 6.9 PRACTICAL CONSIDERATIONS IN CURRENT GOOD MANUFACTURING PROCESSES–COMPLIANT ASEPTIC PROCESSING: ACHIEVING STERILITY BY DESIGN [17]

The scientific considerations regarding aseptic processing covered to this point in this chapter can help readers better understand the difficulties associated with the

manufacturing of sterile products using aseptic technologies. In practice, however, regardless of the regulatory expectations or the technical challenges resulting from a given product, the fact remains that at its basis, aseptic processing is quite simply nothing more than the control of contamination at a very high level. Provided an aseptic processing achieves that objective, product safety, which is the most critical attribute associated with an aseptically produced product, is assured. As can be seen from the research done in the area of aseptic processing already discussed in this chapter, in principle, high-level contamination control is primarily a function of how well a process manages risks associated with human contamination.

The authors have coined the term "sterility by design" (SbD) to underscore an aseptic production focus on facility design, process design, process control, and operational effectiveness that results in the reduction of personnel interaction with sterile materials. There is no more effective way to achieve high performance regarding contamination control (and hence risk management) in aseptic processing than optimal management of risk arising from human workers. By far the best approach to aseptic processing is the complete elimination of contact between human operators and the area in which aseptic processing is accomplished. This eliminates all contact among personnel, sterile components, and product, and it obviates all direct participation by operations in the aseptic setup, fill, and seal phases of manufacturing. This aseptic processing ideal has been released and will likely become the norm during the careers of most of the readers of this text.

At the present, however, manned cleanrooms, or aseptic processes with some degree of direct human intervention requirement, are typical. The concepts of SbD have three simple fundamental performance objectives:

1. To ensure separation between human operators and the most risk-intensive locations in aseptic processing to the greatest extent possible
2. To limit personnel interaction with sterile materials at all times and ensure that when such contact is unavoidable, it occurs only under conditions that are most favorable to the reduction of contamination risk
3. To the greatest extent possible, reduce the presence of personnel in the critical aseptic environments

The following are key SbD considerations, which can be used to form the basis for the identification of critical process design considerations, validation features, and critical control points:

1. Product considerations
   - The product must be compounded and sterilized using validated methods that both ensure sterility and the conservation of the product's chemical or biological therapeutic attributes.
   - Physical sterilization using heat or ionizing radiation is generally straightforward and low in risk.
   - Sterilization of products by filtration is highly reliable but more risk intensive than physical sterilization processes.

- Control of bioburden before sterilizing filtration is critical because it can ensure that the filter is used within its defined use conditions and that microbial toxins are avoided. *Under no circumstances is the presence of conditions that favor microbial proliferation in the product acceptable because the proliferation of contaminants is universally objectionable*.
- Active ingredients, excipients, and all product components should be packaged in ways that allow surface decontamination of their containers. When product is supplied to the critical zone in sterile containers (e.g., sterile powders, sterile components, or subunits), it should be introduced in a manner that retains at least one hermetic layer until entry into the ISO 5 critical zone.
- Product contact materials that enter aseptic environments should always be sterilized using a validated process.
- Product supply and connection should be made within the ISO 5 critical zone.

2. Containers and closures
   - Containers and closures must be cleaned, prepared (e.g., lubricated), and sterilized and depyrogenated using validated processes.
   - Containers and closures should be introduced in a manner that retains at least one hermetically sealed layer until entry into the ISO 5 critical zone.
   - Containers and closures should be selected and specified for reliability of handling in the processing equipment because operator interventions are frequently required as a result of poor component feeding and handling.
   - Containers and closures should be of suitable quality for their intended use. Higher Acceptance Quality Levels (AQL) for defects can result in a reduction in the need for corrective interventions.
   - Containers and closures bulk containers should be designed for ease of opening and minimization and isolation of operator activity. Inventories of these materials within the aseptic environment should be minimized.
3. Facilities
   - Facilities should be designed for ease of cleaning and sanitization and decontamination.
   - The facility should be designed to minimize the potential for transfer of contamination into the aseptic processing area from surrounding environments.
   - The aseptic processing area should be maintained at a positive pressure to prevent the airborne ingress of contamination from less clean environments.
   - An air pressure differential cascade should be established in which lower classified environments are balanced so as to remain negative to higher classified environments (e.g., ISO 5 > ISO 6).
   - HEPA-filtered air should be supplied through all classified areas, and HEPA filters should be subjected to annual or semiannual efficiency testing.

- Differential pressures between areas should be continuously monitored and alarmed.
- Temperature and humidity established to maximize personnel comfort during operations, humidity should be approximately 50% or less, and temperatures should obviate any possibility of operator discomfort or perspiration.
- Interlocks should be used to prevent inadvertent pressure reversal.
- Closed RABs and isolators should be the preferred processing environments.
- Facility surfaces should be resistant to the action of cleaning, sanitizing, and decontamination chemicals.
- The aseptic facility should be maintained in "clean" operational state at all times and periodically sanitized or decontaminated. Isolators and closed RABs should be decontaminated with sporicidal agents at regularly scheduled and validated intervals.
- A "just-in-time" approach to material management should be observed in all aseptic areas to the greatest extent possible.
- Smoke studies, expectations for unidirectional air flow, absence of eddies, and other subjective expectations imposed on HVAC systems should be recognized as nondefinitive means for assessing environmental performance.
- In facility design, air exchange rates (air changes per hour) are the preferred indicator of the contamination control capacity of the environment.

4. Equipment
   - Product contact surfaces should be sterilized by validated methods. (Vibratory feed systems may be exempted from this requirement provided they are high-level decontaminated with a sporicidal agent in situ as is typical in isolator or closed RABs systems.)
   - Sterilization-in-place and clean-in-place procedures should be used whenever possible; these methods are always superior to aseptic assembly from a contamination control perspective.
   - Equipment that cannot be sterilized in place should be fully assembled before sterilization to reduce the need for aseptic assembly.
   - Equipment and utensils should be sterilized in sealed containers whenever possible. When they must be sterilized as wrapped articles, the wrapping must be tear and puncture resistant and impervious to moisture.
   - Equipment and utensils should be introduced in a manner that retains at least one hermetically sealed layer until entry into the ISO 5 critical zone.
   - Equipment and utensils should be sterilized and depyrogenated using a just-in-time approach.
   - Equipment should be selected for high reliability, ease of change over, and adjustment; equipment that can be set up with minimum manual intervention results in superior contamination risk management.

- Remote adjustment of equipment operational parameters should be used whenever possible to further reduce contamination risk from intervention.
- Tool-free change over from one format to another should be used when possible.
- Equipment should be tolerant of minor deviations in component specification and should be easy to keep in proper alignment.
- Containers should be handled and positioned positively rather than passively to avoid corrective interventions as a result of jams, tip-overs, or misfeeds.
- Equipment (and to some extent critical facility systems) should use feedback systems for ease of control, operation, and documentation.
- Nonproduct contact portions of the equipment should be easily decontaminated and should be designed to preserve the integrity of the ISO 5 critical zone.
- All equipment surfaces should be resistant to the potential corrosive action of sanitizing and decontamination chemicals as well as cleaning agents.

5. Procedures
   - Procedures should be critically reviewed to emphasize the need to reduce the numbers of interventions of all types and to simplify all aseptic processes to the greatest extent possible.
   - Interventions should be designed to ensure minimal risk of contaminating sterile materials.
   - All interventions performed during aseptic processing must be recognized as increasing the risk of contamination dissemination.
   - All interventions should be performed using sterilized tools.
   - Intervention procedures should be established in detail for all inherent interventions and more broadly for corrective interventions.
6. Personnel
   - Personnel must receive initial and periodic formal training regarding CGMP, aseptic processing, general contamination control science, aseptic gowning, and job-specific tasks.
   - When appropriate, personnel should be initially qualified and periodically assessed for their proficiency in aseptic gowning.
   - Personnel should be trained and then initially qualified and periodically assessed for proficiency in aseptic technique.
   - Personnel should be disciplined and detail oriented so they may conform to the highest standards of aseptic technique at all times.
   - Personnel should be monitored upon each exit from the aseptic core. (Gloves inside aseptic enclosures should be monitored at the end of the batch or campaign.)
   - Gown materials should be cleaned and sterilized using validated methods. A fresh sterile gown should be used by the operator on each entry into the aseptic environment.

- Gloves on aseptic enclosures should be replaced periodically, sterilized, and integrity tested.
- Manual filling by aseptically gowned personnel should be recognized as an anachronistic throwback to an earlier time and no longer used for aseptic processing.

7. Monitoring
   - Overreliance on monitoring of the environment is a common mistake; it is important to consider that it is impossible to monitor quality into a process or product.
   - Environmental monitoring activities are human interventions and must be treated as such.
   - Monitoring of operational facility and equipment parameters in all stages of aseptic processing is very important, and when advanced aseptic technologies such as isolators or closed RABs are used, is generally more important than environmental monitoring.
   - It is critical to avoid a false sense of security taken from the review of environmental microbiological data and media fill test results; one must be aware of the fact that in microbiology zero does not mean "sterile." Rather, it means nothing detected.
   - Avoid overconfidence from environmental monitoring data and always rely on physical data and careful expert supervisory oversight to ensure operational integrity in aseptic processing.

Attention to these rather simple concepts in sterility by design will ensure a design space optimized for the manufacture of safe and well-controlled sterile product. This will in parallel ensure an aseptic processing culture that is both reliable and compliant at all times.

## 6.10 THE NEED FOR SCIENTIFICALLY VALID STANDARDS AND REGULATIONS

We hope that after reviewing the concepts of "sterility" and "asepsis" that it can be agreed that achieving either of these goals in cleanroom product manufacturing is impossible to ensure, much less to prove, at the current state of technology. We do not object to the label "aseptic processing" for the production of product of high microbiological purity in cleanrooms. However, we should all understand that the actual demonstration of "asepsis" in each and every unit of product is not possible and that it is equally impossible to demonstrate that some general condition called "sterile" exists. It is inappropriate scientifically to consider that a cleanroom with aseptically gowned personnel could be "sterile." A "sterile" cleanroom would require "sterile" human operators using perfect contamination-preventing clothing while working in a "sterile" environment, and these conditions cannot presently be attained.

Because these environments are not "sterile" and because humans release organisms at various rates depending on movement and gowning effectiveness, we should not be surprised at the periodic random discovery of some viable organisms in even the most critical locations within a cleanroom. We cannot agree with the notion that organisms found on two closely spaced days on a single site in a cleanroom are indicative of an adverse trend. We believe that this contention is statistically unreasonable because workers will release organisms continuously, albeit at varying rates per unit time. With that in mind and considering the fact that the samples themselves are taken by other humans who also continuously release organisms, how can low-level contamination legitimately be considered a trend even if seen 2 or 3 days in a row? Also, how can one consider the occasional positive result in a media fill test as anything but the natural consequence of the environmental conditions under which the study is conducted?

Upon careful analysis, we suggest that the aseptic processing industry has fallen prey to the inappropriate application of absolutist thinking totally unsupported by scientific fact. We believe that in the context of aseptic processing, we are applying the absolute meaning of the word "sterile" in a wholly inappropriate and ultimately self-defeating manner. There are well-meaning industry practices and regulatory compliance expectations that may lead to an unfortunate use of material and human resources. Among them are:

1. The imposition of expanded environmental monitoring schemes because it is erroneously thought that as contamination is reduced more testing is required to detect lower microbial levels as cleanroom facilities improve and advanced aseptic processing systems such as isolators and close restricted assess barriers are used.
2. The imposition of zero microbial contamination as a fixed requirement at any location in a cleanroom, including on so-called "product contact surfaces."
3. The suggestion that the recovery of any mold at all in a cleanroom is indicative of a loss of control. Mold should be infrequently observed, but it must be realized that it is ubiquitous in nature and can be found as a passenger on human skin, hair, and clothing. Certainly, cleanrooms should not be colonized by mold, but this is a different matter from the rare random recovery of mold.
4. The linkage of product release to environmental monitoring data. If we are to reject product for random low-level background contamination in manned cleanrooms, we are merely adding to the cost of production with no commensurate benefit in patient safety.
5. The dramatic increase in personnel monitoring and testing in terms of both frequency and location on personnel.
6. The suggestion that any contamination found in a media fill test is indicative of a loss of process control.
7. The suggestion that the cooling zone of depyrogenation tunnels should be sterilized in the complete absence of real data indicating that a risk exists.

It should also be recognized that the term "sterility assurance level" in the context of aseptic processing has no scientific merit and has been unfortunately but understandably misapplied in efforts to correlate what are really media fill contamination rates to the label claim of "sterility." We should also recognize that many of the numerical standards that are in use for testing and monitoring of aseptically filled products were not scientifically derived and have never taken the analytical methods sensitivity into a account. Additionally, many standards and expectations were implemented without the benefit of careful and critical peer review. Scientific honesty also requires that we admit that the methods used to evaluate the microbial content of the environment vary tremendously with respect to the results they yield as studies by a number of scientists have already shown.

## 6.11 AN OVERVIEW OF STERILIZATION AND DECONTAMINATION

Sterilization is, as already mentioned, a very different thing from aseptic processing, although obviously, the product pathway and product contact parts are sterilized before use in aseptic processing. Thus, sterilization is required in all sterile product manufacturing activities. In this section of the chapter, we are referring only to sterilization processes that actively destroy microorganisms. To be clear, this means only methods that physically sterilize using, for example, heat, energy from radiation, or chemicals.

Again, we must begin with definitions. As a noted authority on the subject of sterilization and decontamination, Dr. S.S. Block, wrote in the definitions chapter of his widely referenced book titled *Disinfection, Sterilization and Preservation*: "Definitions are man-made; they do not come to us from on high" [18]. He goes on to write:

> Death is not exact. When is a person dead? When is he no longer breathing? When he is brain dead? When all the cells in his body are dead? When is a product sterile? When all life in it has been killed or when it has been heated at X temperature for Y minutes or is it when its microbial population (bioburden) has been reduced by z D values? It is what we say it is according to the way we construct our definition.

In the case of sterilization processes in the health care industry, we say sterilization begins is at a probability of nonsterility (PNSU) of 1 unit in a million. The way in which this definition, often given as a "SAL of $10^{-6}$," is used in our industry, it is almost as though it is an immutable law of science. We often forget that the selection of this value as a target or specification was more or less arbitrary. As Dr. Block put it, that is the way we have constructed our particular industry specific definition. We could have chosen another value, but we didn't; we chose this one.

There are various historical perspectives that have been advanced regarding the choice of a PNSU of not more than 1 unit in a million. Some have written that perhaps we borrowed this value from the NASA program for the sterilization of spacecraft.

Perhaps it evolved out of a general understanding in the medical community that it is a suitable risk value for parenteral products that could be sterilized in their containers. Whatever its actual genesis might have been, what is clear is that those who chose this value understood clearly that they were not defining a technical limit for the condition of "sterility." Rather, they were defining an acceptable level of risk, which is clearly quite a different thing altogether.

We often hear people speak as though the SAL or more precisely PNSU of $10^{-6}$ is an attribute that all terminally sterilized products have. This if, of course, untrue because $10^{-6}$ is not a process capability; rather, it is a minimum acceptable level of performance. Also, the sterilization of components or other production-required equipment and materials always results in SALs well in excess of $10^{-6}$. In fact, the failure to understand how low the true risk associated with well-designed and validated sterilization processes really is quite often results in extremely excessive margins of safety being arbitrarily built in to cycles.

So if someone says that an aseptically produced product has a SAL of at least $10^{-3}$, the reader should now know that this is a meaningless statement because we have no means of tabulating a SAL or PNSU for an aseptically produced project. In the case of terminal sterilization, we can demonstrate quantitatively that we have met or exceeded the established minimal risk value for sterility, which is a SAL or PNSU of $10^{-6}$. Do not fall into the trap of thinking that this means physical sterilization is, as often heard, "three logs better" than aseptic processing; this is not at all the case. As already noted, any attempted direct quantitative risk comparison between physical sterilization and exclusionary methods such as aseptic processing is doomed to fail because they operate in entirely different ways such that there is no scientifically valid means of initiating a comparison.

A noted regulatory investigator, Mr. John Sharp, once said (we suspect partly in jest) that "sterilization probability of one in a million is fine if you are not the millionth bloke" [19]. However, in the case of physical sterilization, the way in which we sterilize our materials; components; and in many cases, final products means that the probability of any user of those products being affected by a correctly run sterilization process is so close to zero as to be no cause for technical debate at all. In our combined seven decades of experience in microbiology and sterilization processes, we have actually never seen a process that yielded a minimal PNSU of exactly $10^{-6}$. This is because all of them provide substantially less risk of a nonsterile event than that, and most of them deliver actual lethality values far in excess of the minimum. The purpose of any sterilization process is the elimination of bioburden; however, the mechanisms used for development and validations of these processes rely on the use of very highly resistant biological indicators. These biological indicators are not only as much as 1 million times more resistant than typical bioburden; they are also challenge the system at enormously higher population numbers than could be present in any normal environment. A common misunderstanding is that the destruction of $10^6$ spores on a biological indicator equates to a PNSU of $10^{-6}$. Actually, the complete kill of, for example, a moist heat biological indicator inoculated with *Geobacillus stearothermophilus* spores at a population of $10^6$ ensures a PNSU substantially lower (better) than $10^{-6}$.

So, an important fact to remember is that when physical sterilization is done with any degree of competence, it is so reliable and so well quantified that it is a truly insignificant risk modality. This does not mean that developing and validating sterilization processes is easy, but it does mean the technologies, whether they are heat, chemical, or radiation, are well understood and comparatively easy to quantify. The result is that we do not see sterilization cycles in terminal sterilization or materials sterilization in which there is the slightest risk of injury even if the entire population of the planet was administered the product. In fact, even this criterion is often exceeded by tens, if not hundreds, of orders of magnitude. Readers are urged to not only delve further into published technical literature but to also spend some time with a calculator or computer investigating the true risk of a nonsterile unit at the lethality levels afforded by the typical industry autoclave cycle.

## 6.12 MICROBIAL DEATH CURVES

The death of microorganisms, when subjected to lethal treatments regardless of sterilization method, follows a fairly close approximation of first-order kinetics. What this means is that when the death curve, which indicates the destruction of microbial population per unit time, is plotted on semilogarithmic graph paper, something close to a straight line results. When subjected to linear regression analysis, the result should be a fit that confirms reasonable linearity. What this essentially means is that the slope is consistent over time, which is consistent with microbial death or kill, being the result of a uniform effect that has only minor variability in reaction time.

In practice, one will normally see a very close fit to the first-order kinetics with heat sterilization, whether it is dry heat or moist heat, and other forms of sterilization such as chemical and radiation. Under ideal chemical reaction conditions, broadly lethal sporicidal agents should also produce a close approximation of first-order kinetics. When nonlinear death curves are observed, quite often a latent or lag period is observed followed by a more characteristic linear curve. This latency is believed to be the result of manner in which the suspension culture of microorganisms interacts with the material onto which they were inoculated. In some cases, they form clumps of organisms in a matrix of inorganic salts and organic materials that result in the appearance of a type of protective biofilm that is difficult to penetrate.

This phenomenon probably plays a significant role in the observation of the so-called *substrate effect* in which the kill rate of an organisms appears to depend to one degree or another on the substrate upon which it was inoculated. Substrate effects may also related to hydrophobicity because moisture is often required to produce optimal spore kill rates. Bacterial spores, because of their extreme resistance to environmental stress in the form of heat, and chemicals are often used in the evaluation of sterilization processes, and the absence of moisture may reduce kill efficacy.

The effect of moisture can be seen quite clearly from the simple fact that it is necessary to run dry heat processes at far higher temperatures than moist heat

processes to achieve the same result. In 1933, Barker reported that moisture was essential to the denaturation and coagulation of proteins in moist heat sterilization [20]. Lewith reported similar findings; thus, it appears that moist heat kills by a different mechanism than dry heat, which is probably largely an oxidative process [21]. Wang et al. showed experimentally that activation energy and entropy changes for killing of spores and denaturing of protein are quite similar, which further supports the concept suggested by Barker [22]. It was further demonstrated in these studies and others that spores are quite permeable and freely exchange water with the environment.

Dry heat sterilization, because water vapor is absent, results in low water activity of the spores. Stated another way, dry heat sterilization occurs at low relative humidity. Dry heat sterilization is performed at a water activity ($A_w$) of zero (0), which is a relative humidity (RH) of 0%. In steam sterilization, the conditions are an $A_w$ of 1 and RH of 100%. Angelotti et al. showed that spore resistance actually increased from an $A_w$ of 0 to 0.3 [23]. They reported that spore kill rate increases of up to 100-fold were possible at reduced moisture levels. These data help explain the criticality of air removal, a factor often emphasized in moist heat sterilization. Because cold air and hot steam do not mix easily, insufficient air removal can result in low moisture, low temperature, and consequently poor kill at the sterilization target. Young points out that "all the air does not need to be removed from a sterilizer, but all surfaces requiring sterilization must be exposed to adequate moisture" [24]. This is further reinforced by the fact that in some moist heat processes, materials air is purposely left in the chamber to minimize pressure gradients across products being sterilized. However, provided that the air–steam mixture is adequately uniform, something that can be assisted by a circulating fan or water circulation, consistent sterilization is observed.

Moisture is not a factor only in heat sterilization; many sporicides are formulated in an aqueous base, and their effectiveness owes no doubt in part to the ability of water to permeate spores freely. It should be noted that moisture has been observed to be critical to the efficacy of gas sterilization by chemicals such as ethylene oxide, ozone, and chlorine dioxide.

## 6.13 TOWARD A MEASURE OF REAL PRODUCT SAFETY

It is increasing clear that what is needed in industry are real measures of regulatory compliance based on scientific and engineering fact rather than compliance based on supposition or tradition rather than science. In no area of compliance is the need for science- and risk-based regulation more critical than in the area of aseptic processing. The absolute nature of "sterility" as a concept has led to the evolution of compliance expectations that are, in some cases, both unscientific and unreasonable.

It makes no sense to set standards that cannot reasonably be attained. Doing so will do nothing but result in claims that sterility assurance is somehow lost when actually a cleanroom is being operated using current best practices and the contamination rate is at the established background level.

If we fail to base aseptic processing requirements on what is attainable rather than some scientifically excessive theoretical consideration of "sterility," we will do nothing but waste resources and product while at the same time failing to increase product safety. It may seem like we are doing something to improve product by appearing to tighten the standards, but in reality, unattainable standards will do nothing to improve real compliance or real product quality. Rather, they will simply require what can best be described as work without real tangible benefit.

It is understandable that many people involved in standard setting and regulatory policy would attempt to impose a standard of absolute safety because this appears to be the ultimate in assurance of safety. However, we simply lack the ability to operate cleanrooms in a "sterile" fashion, and none of our test methods can ensure that the attribute of "sterility" has been attained. The so-called "sterility test" is least capable of all.

The test for sterility has been applied to preparations for many years. In fact, tests for absence of viable organisms have been used in microbiology since the 19th century when the ability to culture and grow microorganisms in a laboratory setting was developed. The sterility test has been included in the USP for many decades, and the introduction of membrane filtration brought about broader applications of this growth and recovery methodology in the 1950s and 1960s.

However, in retrospect, the naming of the "absence of recovered viable organisms" test as a "sterility test" was an unfortunate occurrence.

From a purely scientific perspective, the "sterility test" cannot confirm the condition of "sterility" for three very good scientific reasons. First, we do not know what the actual sensitivity and limit of detection of the "sterility test" really is. It is reasonable to assume, however, that it is not zero, even among organisms that could possible be cultured with the media stipulated in the test instructions. Second, because we use a test that uses only two media intended to grow a broad spectrum of bacteria and mold, we know that it will not recover more fastidious organisms or important pathogens such as viruses that are obligate intracellular parasites and will not grow in culture media such as trypticase soya broth or fluid thioglycollate. Finally, because the "sterility test" is a sample depending test and the sample size is quite small given typical lot sizes, the statistical reliability of the test is severely limited.

Thus, the term "sterility test" is and always has been a misnomer in many respects. The test, if the result is positive, certainly shows that a lot of material is surely nonsterile. However, as is the case with any "zero" growth result in microbiology zero cannot correctly be interpreted as equivalent to "sterile." Actually, "no growth" only means that no growth was observed in this sample under the conditions of the defined "sterility test." What this means is that the "sterility test" is a test for the presence of detectable contamination in a limited sample population rather than a test for sterility. It also means that it is really a "sentinel" test from which it is presumed that the absence of organisms that could grow in the media and incubation conditions defined in the test, meaning that a sufficient level of microbiological safety is likely. We emphasize again, however, that we do not at this point have a true "sterility test."

Requiring industry to attain and confirm "sterility" is not the same thing as achieving that condition. We can only achieve what is possible; we cannot and should not be expected to conform to some theoretical ideal.

## 6.14 A FEW SCIENTIFICALLY MODEST PROPOSALS

1. Ensure that microbiologists or engineers who fully understand the underlying microbiological principles write all standards on microbiological quality. Microbiologists should not write chemical or pharmacologic standards, and it seems both pragmatic and reasonable that chemists and pharmacologists ought to leave microbiology to microbiologists.
2. Improve dialogue between industry regulatory inspectors and reviewers on the science of aseptic processing and the tools we use for aseptic process control.
3. Base decisions on process control and performance requirements for aseptic processing on measurable safety rather than theoretical or abstract requirements such as "sterility." The need to change validation requirements or process control requirements should be based on a real need to improve safety rather than on the perception that a closer approach to an absolute ideal is required. Thus, standards would be based on quantitative safety data rather than subjective opinion.
4. Ensure that numerical standards as they apply to microbiological control take full consideration of the limits of detection, accuracy, and precision of the methods. All too often, microbiological methods are assumed to be far more accurate and reproducible than they actually are. The difference between 5 CFU and 15 CFU may seem very significant to a chemist, but in microbiology, such differences can easily be the result of the method, the conditions of the organisms, or both. The arithmetic difference between $5 \times 10^5$ and $1 \times 10^6$ would be huge in a chemical assay; however, in microbiological serial dilutions, it is barely significant.

It seems reasonable to assert that those in the sterile health care product manufacturing field have to work to higher standards than food scientists, who only have to achieve "industrial sterility" in aseptic food processing. Actually, there is no real difference between "commercial sterility" in the food industry and the "sterility assurance" of our pharmaceutical processes; both terms correctly reflect the fact that these processes are not capable of "sterility" in the strictest scientific meaning of that word. However, as an industry, we must take care that our evolving attitudes and expectations regarding perfection do not blind us to reality or that we allow perfection as reflected in expectations to become the enemy of good (or better).

It is critical to note that none of the foregoing discussion is meant to suggest that we should relax our expectations that products manufactured aseptically and labeled "sterile" be made under stringent process control and uniformly diligent attention to detail. There is ample evidence that our industry has a rigorous commitment to ensuring that our "aseptic" processes are as good as they can be. So much are our "hearts in the right spot" that because of a misunderstanding of scientific principles,

expectations are established that are higher than is technically achievable. We believe it is reasonable to state that based on in-use safety data that products manufactured in compliance with current industry practices are safe.

The authors call the reader's attention to the PDA's *White Paper on Aseptic Processing* published in 1988 [25]. In our experience, products manufactured today using the approach described in that document would still be microbiologically safe. In fact, the case can be made that in the present day, the real danger to patients from sterile products is at the level of admixture in the pharmacy and administration.

It was reported by the World Health Organization (WHO) [26] that of each 100 hospitalized patients in the developed world seven will develop at least one healthcare associated infection. The rate of hospital acquired infections on the developing world is considerably higher with 10 patients out of each 100 admitted having at least one healthcare associated infection. These infections are not the result of contaminated product but rather result from treatment including parenteral injection. The same WHO report states that in high income countries, up to 30% of intensive care unit (ICU) patients are affected by at least one ICU acquired infection. The WHO further states that the costs associated with these infections are about $10 billion in Europe and $6.5 billion in the United States, and also asserts that most of these infections could be prevented by simple and relatively inexpensive measures such as improvement hand hygiene. These are real consumer safety issues to which industry, health care providers, and the regulatory community should focus their collective energy. Humankind will be better served if we were to devote our efforts to solving these problems rather than devoting unreasonable levels of resources to areas in the delivery chain where risk is already low.

Certainly, when organizations are found that are making product under conditions that do not comply with the cGMPs that exist in this industry for aseptic processing, aggressive enforcement action should always be taken. However, this does not mean that the vast majority of current practices are inadequate. Nothing is accomplished by punishing everyone for the sins of a few. In recent years, too many regulatory initiatives appear to be solutions in search of a problem.

Will we ever achieve demonstrable "sterility" in human-scale cleanrooms? No, we will not. Will we achieve absolute "sterility" in the current generation of isolators? No, we will not, but we will experience a closer approach to the ideal by manufacturing in a germ-free environment. Technological improvements should be implemented when they are beneficial. We believe that when aseptic products are manufactured in compliance with industry practices that have existed for at least the past 20 years, they are reliably safe. Let us stop worrying about theoretical conditions that can neither be proven nor measured and focus instead on safety, which can be both proven and measured.

## REFERENCES

1. Food and Drug Administration. Guideline on sterile drug products produced by aseptic processing. Washington, DC: Food and Drug Administration; 2004.
2. Webster's New Collegiate Dictionary.

3. Food and Drug Administration. Guideline on sterile drug products produced by aseptic processing. Washington, DC: Food and Drug Administration; 1987.
4. Brooks GF, Butel JS, Ornston LN, editors. Jawetz Melnick & Adelberg's Medical Microbiology. Norwalk, CT: Appleton & Lange; 1993.
5. Willett HP. Sterilization and disinfection. In: Joklik WF, Willett HP, Amos DB, Wilfer WM, editors. Zinsser Microbiology. Norwalk, CT: Appleton & Lange; 1992, p. 188–199.
6. Davis BD, Dulbecco R, Eisen HN, Ginsberg HS, Wood WBMcCarty M, editors. Microbiology, Including Immunology and Molecular Genetics, Hagerstown MD: Harper and Row; 1973.
7. Osterhout S. Host-parasite relationships. In: Joklik WF, Willett HP, Amos DB, Wilfer WM, editors. Zinsser Microbiology. Norwalk, CT: Appleton & Lange; 1992, p. 87–392.
8. Iglewski BH, Clark VL. Molecular Basis of Bacterial Pathogenesis. New York: Academic Press; 1990.
9. Finlay BB, Falkow S. Common themes in bacterial pathogenicity. Microbiol Rev 1990;53: 210.
10. Agalloco J, Akers J. Risk analysis for aseptic processing: the Akers-Agalloco Method. Pharm Technol 2005;29(11):74–88.
11. Agalloco J, Akers J. Simplified risk analysis for aseptic processing: the Akers-Agalloco method. Pharm Technol 2006;30(7):60–76.
12. Katayama H, Toda A, Tokunaga Y, Katoh S. Proposal for a new categorization of aseptic processing facilities based on risk assessment scores. J Pharm Sci Technol 2008;62 (4):235–243. Agalloco JP, Akers JE. Validation of aseptic processing. In: Carleton FJ, Agalloco JP, editors. Validation of Pharmaceutical Products: Sterile Products, 2nd ed. New York: Marcel Dekker; 1999.
13. Whyte W. Sterility assurance and models for assessing bacterial contamination. J Parenteral Sci Technol 1986;40:188–197.
14. Reinmueller B. Dispersion and risk assessment of airborne contaminants in pharmaceutical clean rooms. Bulletin No. 56, Royal Institute of Technology, Building Services Engineering; Stockholm, Sweden, 2001.
15. Ljungqvist BJ, Reinmueller B. Ventilation and Airborne Contamination in Clean Rooms. Tryckeri PriCor, KTH Sweden; 1995.
16. Reinmueller B. Some observations on comparative studies on active air sampling. Proceedings of the 15th ICCCS International Symposium on Contamination Control, Copenhagen, p. 350–359.
17. Khawar M. Pharmaceutical briefings; April 2001.

# Cell Culture Processes in Monoclonal Antibody Production

FENG LI, AMY SHEN, and ASHRAF AMANULLAH

## 7.1 INTRODUCTION

With the increasing number of therapeutic candidates, especially monoclonal antibodies (mAbs), entering early stage process development, biopharmaceutical companies are increasing looking at innovative solutions to deliver the pipeline. Time to market, cost effectiveness, and manufacturing flexibility while maintaining the desired product quality attributes are key issues in today's competitive market, where several companies are working on therapies for similar clinical indications. Because most antibody therapies require high doses over a long period of time, antibody manufacturing capacity remains a big challenge. There is a continuing need for developing high-yield processes to produce large quantities with cost and time efficiency to meet clinical requirements and pave the way toward commercialization. In response to this strong demand, the current trend in the antibody industry is to make continuous improvements to develop a high-titer upstream process in a timely manner. Mammalian cell culture is the dominant expression system for therapeutic antibody production. However, mammalian cells have been considered difficult to work with in term of manufacturability and scalability because of low yield, medium complexity, serum requirement, shear sensitivity, and so on. After a decade of development work in cell line, media, and bioreactor condition optimization, high titers up to approximately 5 g/l and cell densities of more than 20 million cells/ml have been recently reported in fed-batch processes [1,2]. Cell-specific productivity of more than 20 pg/cell/day can be routinely achieved for production cell lines. Enhancement of specific productivity per cell is achieved not only from selection of the highly productive clones but also from optimization of medium composition and bioreactor operation conditions.

Cell culture process development starts with cell line generation and selection followed by media and culture condition optimization in small-scale systems, including 96-well plate, shaker flasks, and bench-scale bioreactors, for high-

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*,
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

throughput screening purposes. After the process conditions are defined, the process is often transferred to the pilot scale to get scale-up information and to produce toxicology materials and later to current good manufacturing process (cGMP) manufacturing for production of clinical material. As development of a commercial cell culture process for production of a biological product is completed at the laboratory and pilot scales, the commercialization process begins with process characterization, scale up, technology transfer, and validation for the manufacturing process. The strong demand for therapeutic antibody production, relatively modest titers, and reduction of cost of goods has resulted in a trend toward large-scale manufacturing. Today the largest biotech companies are using several hundred to 2000-l scale bioreactor for early clinical production and 10,000- to 25,000-l stirred tank bioreactor to produce commercial products.

## 7.2 MAMMALIAN EXPRESSION SYSTEM AND CELL LINE ENGINEERING

### 7.2.1 Expression Systems

Therapeutic antibodies are mainly produced in mammalian host cell lines, including murine myeloma cells, PER.C6 human cells, and Chinese hamster ovary (CHO) cells [3]. The choice of which expression system to choose is determined by an individual company's experience and by the host cell's ability to deliver high productivity with acceptable product quality attributes.

***7.2.1.1 NS0 Expression System*** Murine NS0 cells are non–immunoglobulin-secreting myeloma cells and are cholesterol auxotrophs requiring the presence of cholesterol in culture medium for growth [4]. However, cholesterol-independent NS0 cells also have been established [5–7]. NS0 cells lack endogenous glutamine synthetase (GS) enzyme activity, making them suitable for use with GS as a selectable marker for recombinant antibody expression [8]. However, antibody production up to 3 g/l has been reported from non-GS NS0 cell lines as well [9,10]. Most mouse-derived cell lines, including NS0, produce *N*-glycosylneuraminic acid (NGNA), a sialic acid that does not exist in human antibodies and that might be immunogenic in humans [11,12]. Although NS0 cells have been used in industry to produce therapeutic antibodies, this immunogenicity concern might at least partially make NS0 cells not as widely used as CHO cells for therapeutic antibody production.

***7.2.1.2 PER.C6 Expression System*** Compared with NS0 and CHO cells, PER.C6 cells are a relatively new technology and are derived from human embryonic retina cells, which have been immortalized by transfecting the E1 genes from adenovirus 5 DNA [13]. PER.C6 cells can proliferate indefinitely in suspension under serum-free conditions similar to NS0 and CHO cells. Their human origin might favor the argument that it is advantageous to produce therapeutic antibodies in PER.C6 cells because the posttranslational modifications are human-like and might be less immunogenetic. However, the human origin may make cells more vulnerable

to adventitious agents that can infect humans and cause diseases. This may raise some safety concerns and require more stringent requirements for cell line characterization and downstream viral removal capability.

***7.2.1.3 Chinese Hamster Ovary Expression System*** CHO cells are the predominant host used to produce therapeutic proteins. About 70% of all recombinant proteins [14] produced today are made in CHO cells, including DUXB11, DG44, and CHOK1 cells. Whereas DUXB11 and DG44 cells do not have dihydrofolate reductase (DHFR) activity, CHOK1 cells have endogenous DHFR activity.

A stable cell line development process starts with expression vector construction and transfection. After being transfected with plasmids bearing the antibody light chain gene, the heavy chain gene, and a selectable marker or markers, the cells are screened for highly productive cell lines after growth recovery, serum-free suspension adaptation and amplification (if necessary), and clone selection. To screen and select a highly productive and stable clone from the transfectant population in a limited time frame is a major challenge. These steps are further described below.

### 7.2.2 Cell Line Development

***7.2.2.1 Expression Vector*** The first step in developing a stable antibody-producing cell line is to construct an expression vector. Typically, the mammalian expression vector contains a cassette for the antibody genes, selectable marker gene (s) in mammalian cells, and a second cassette for the genes enabling plasmid replication in bacteria. To achieve a high level of antibody expression, strong promoter–enhancers, such as the cytomegalovirus (CMV) promoter [15] and elongation factor alpha (EF1$\alpha$) promoter [16], are typically used to drive antibody heavy chain and light chain expression. Often an intron sequence in the 5′ untranslated region is included after the promoter–enhancer to increase the export of transcribed messages to the cytoplasm from the nucleus, and one or more 3′ polyadenylation signal sequences are also included to maximize mRNA levels. Commonly used polyadenylation signal sequences are the SV40 late or early polyadenylation signal sequences and the bovine growth hormone polyadenylation sequence. In addition to transcription, translation and secretion are also required for antibody production. Typically, a consensus Kozak sequence [17] is created by placing GCCGCC(A/G)CC immediately in front of the first translation initiation codon to enhance translation initiation, and a signal peptide sequence is placed immediately in front of the mature antibody peptide to direct antibody secretion [8].

***7.2.2.2 Transfection and Selection*** A variety of transfection methods have been developed to stably introduce vector DNA into mammalian cells, including calcium phosphate, electroporation, cationic lipid-based lipofection, and the polymer- or dendrimer-based methods [18]. All four methods have been used for stable transfection, but electroporation and lipofection are the most common choices. Transfected cells are then selected, relying on different selectable markers that can be

**TABLE 7.1 Selectable Makers Often Used in Mammalian Expression Vectors**

| Selectable Marker | Selective Reagent |
|---|---|
| ***Metabolic Selectable Markers*** | |
| Dihydrofolate reductase (DHFR) | Methionine sulfoximine (MSX) |
| Glutamine synthase (GS) | Methotrexate (NTX) |
| ***Antibiotic Selectable Markers*** | |
| Puromycin acetyltransferase | Puromycin |
| Blasticidin deaminase | Blastcidin |
| Histidinol dehydrogenase | Histidinol |
| Hygromycin phosphotransferase | Hygromycin |
| Zeocin resistance gene | Zeocin |
| Bleomycin resistance gene | Bleomycin |
| Aminoglycoside phosphotransferase | Neomycin (G418) |

categorized into two groups, metabolic selectable markers and antibiotic selectable markers. Some commonly used selectable markers are listed in Table 7.1.

CHO stable cell lines have often been selected using metabolic selective markers, including methotrexate (MTX) (dihydrofolate reductase gene mediated) and methionine sulphoximine (MSX) (GS gene mediated) [8]. Table 7.2 shows a selected list of approved antibody products produced in CHO cells. It has also been reported that an antibody-producing cell line can be generated using two different selectable markers together [9]. In such cases when two selective markers have been used to select stable antibody-producing cell lines, the selective markers are either located in one expression construct also containing genes for heavy chain and light chain or are

**TABLE 7.2 Selected List of Approved Antibodies Produced in Chinese Hamster Ovary Cells**

| Product | Type | Therapeutic Use | Manufacturer | Year of FDA Approval | Selection System |
|---|---|---|---|---|---|
| Vectibix | Anti-EGFR | Metastatic colorectal cancer | Amgen | 2006 | DHFR |
| Avastin | Anti-VEGF | Metastatic colorectal cancer and lung cancer | Genentech | 2004 | DHFR |
| Xolair | Anti-IgE | Moderate and severe asthma | Genentech, Novartis | 2004 | DHFR |
| Raptiva | Anti-CD11a | Chronic psoriasis | Genentech | 2004 | DHFR |
| Humira | Anti-TNF-α | Rheumatoid arthritis | Abbott | 2002 | DHFR |
| Campath | Anti-CD52 | Chronic lymphocytic leukemia | Genzyme, Bayer | 2001 | DHFR |
| Herceptin | Anti-HER2 | Metastatic breast cancer | Genentech | 1998 | DHFR |
| Rituxan | Anti-CD20 | Non-Hodgkin's lymphoma | Genentech, Biogen-Idec | 1997 | DHFR |

DHFR: dihydrofolate reductase; EGFR: endothelial growth factor; FDA: Food and Drug Administration; IgE, immunoglobulin E; TNF-α: tumor necrosis factor α; VEGF: vascular endothelial growth factor.

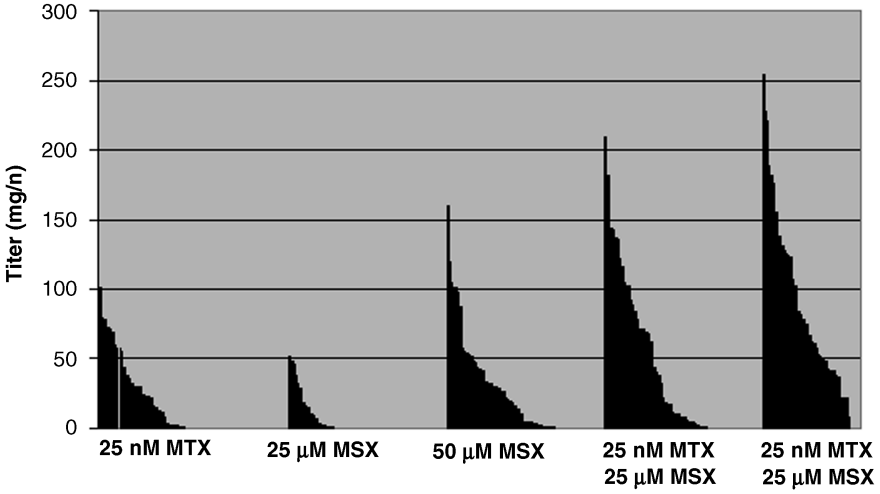


**FIGURE 7.1** 96-well enzyme-linked immunosorbent assay of clones from single and double selection. For each case, 72 clones were screened. MSX: methionine sulphoximine; MTX: methotrexate.

located on two separate plasmids, each containing either the heavy chain or light chain gene.

Recently, we have explored the "double selection" approach, which simultaneously transfects two plasmids into CHO cells for stable antibody cell line generation. The unique feature of this approach is that each plasmid contains both the heavy chain and light chain genes and one selective marker. Cells transfected with two plasmids are selected with two different selective reagents such as MTX and MSX. The productivity of clones generated using this approach resulted in higher productivity than clones transfected with either plasmid alone, as shown in our 96-well productivity assay (Fig. 7.1). To our knowledge, this is the first time that simultaneous transfection with two expression constructs and selection with two different selective markers has been successfully demonstrated as a means to generate highly productive stable CHO cell lines expressing antibody.

***7.2.2.3 Stable Cell Line Screening*** To generate stable cell lines with adequate productivity for clinical or commercial material production, typically hundreds to thousands of clones are screened. The primary screen is usually an enzyme-linked immunosorbent assay with or without cell number normalization to eliminate non- or low producers. While high producers are scaled up, additional assays are performed to measure cell growth, cell-specific productivity, and volumetric productivity (titer) to choose the top 12 to 24 clone candidates, which are typically further analyzed in a fed-batch cell culture scale-down models such as shake flasks or laboratory scale bioreactors. Cell growth, specific productivity, volumetric productivity, product quality (often including glycosylation profile, charge variants, aggregate levels,

and protein sequence heterogeneity), and clone stability are assessed to pick the top four to six clones for further evaluation in bioreactors to choose the final production clone and backup clone (refer to Section 7.4.1).

### 7.2.3 Cell Line Engineering

A significant amount of work has been done to genetically engineer the production host cells to improve or modify the product quality or to improve the host cell robustness. One area that has received a lot of attention is glycosylation control because glycan structures on antibodies can have substantial effects on clearance rate and bioactivities [19]. Antibodies produced in CHO cells typically have complex biatennary structures with very low or no bisecting *N*-acetylglucosamine (bisecting GlcNAc) and high levels of core fucosylation [20]. Overexpression of *N*-acetylglucosaminyltransferase III has been used to increase the fraction of bisecting GlcNAc that resides on antibodies to improve antibody-dependent cellular cytotoxicity (ADCC). RNA interference (RNAi) and gene deletion technologies have also been used to decrease or eliminate the fucose on antibodies to dramatically increase ADCC activity [21,22]. Other areas of host cell engineering include approaches to decrease programmed cell death, to reduce lactate accumulation, and to manipulate cell growth. Overexpression of antiapoptotic gene or genes and RNAi technology–mediated knock-down expression of apoptotic gene or genes have been used to extend the culture viability, leading to the improved productivity [23]. Because CHO cell and other continuously cultured cells have low efficiency to completely oxidize glucose to $CO_2$ and $H_2O$, one of the byproducts of the cell culture process is lactate accumulation, which could cause the acidification of culture medium and lead to high osmolarity and low viability caused by the alkali added to control the medium pH. A significant amount of work [24–26] has been performed to reduce lactate accumulation. However, the usefulness of this approach may be very much clone dependent [26].

## 7.3 CELL LINE AND CELL BANK TESTS

Cell line testing is a critical component of the quality control system for the production of mAbs. Production cell lines need to be qualified at different stages of product development to ensure consistent product quality, cell line purity, and cell line stability. After cell line development through transfection, amplification, and final clone selection, a master cell bank (MCB) and a working cell bank (WCB) are created. The history of the host cell lines in terms of their original source, cultivation history, and the record of animal-derived raw material usage should be documented. Cell line transfection, selection, and subcloning methods are also described in regulatory documents. Antibody production cell banking typically uses a two-tiered cell banking system in which the MCB is created first either from an initial clone or from a preliminary cell bank derived from an initial clone, and the WCB is then generated from the MCB [27,28]. Both the WCB and MCB can provide the cell

supply for a product manufacturing campaign. Whereas the MCB is usually used in early stage clinical phase I and II manufacturing, the WCB is used in phase III and commercial manufacturing. To preclude contamination with cell lines of the same or other species of origin, the MCB and WCB should be created under cGMP, and no other cell lines should be handled or be present in the same cell banking suite.

### 7.3.1 Testing for Master Cell Bank, Working Cell Bank, End-of-Production Cells, and Cells at the *In Vitro* Cell Age Limit

As the production cell lines move forward from the cell line development to different phases of cGMP manufacturing, various tests are required to characterize and qualify the cells at different stages from the MCB to the cells at the *in vitro* cell age limit. The production cell lifespan from cell line generation to the *in vitro* cell age limit is illustrated in Figure 7.2.

As the original cell source to ensure an adequate supply of equivalent cells for production during the entire lifespan of the product, the MCB is subject to intensive testing to demonstrate cell identity and freedom from adventitious agents. Because the WCB is derived from one or more ampoules of the MCB and expanded by a few passages, full retesting for virus contamination is typically not necessary. Besides MCB and WCB testing, end-of-production (EOP) cells should be tested to further evaluate whether new contaminants have been introduced during manufacturing operations or induced by the process conditions. The testing of cells at the *in vitro* cell age limit should be performed at least as the product moves into commercial production. Based on the International Conference on Harmonization of Technical Requirements for the Registration of Pharmaceuticals for Human Use (ICH) guideline, the definition of *in vitro* cell age is the measure of time between thaw of the MCB vial(s) to harvest of the production vessel measured by culture time, by population doubling level of the cells, or by passage level of the cells [29]. The limit of *in vitro* cell age for production should be set based on data derived from production cells expanded under pilot plant scale or full-scale conditions to the proposed *in vitro* cell age or beyond. The tests to support establishing the limit of *in vitro* cell age are often done by extending the seed train culture to beyond potential manufacturing age and then using it to start a culture at either full scale or in a qualified scale-down model. Another way to generate aged cells is to collect cells from a full-scale production run and then passage them further in small scale until the cell age equals or exceeds the proposed limit. Cell line identity, genetic stability, and sequence of the product should all be addressed using these cells to define the maximum cell age for manufacturing product with acceptable performance and product quality attributes.

### 7.3.2 Cell Line Qualification Testing

For the purposes of clinical and commercial manufacturing, the cell banks and EOP cells need to be fully characterized and qualified before the release of bulk drug substance for market. The testing results are also required for submissions of the Biologic License Application (BLA), and some tests must be completed for the

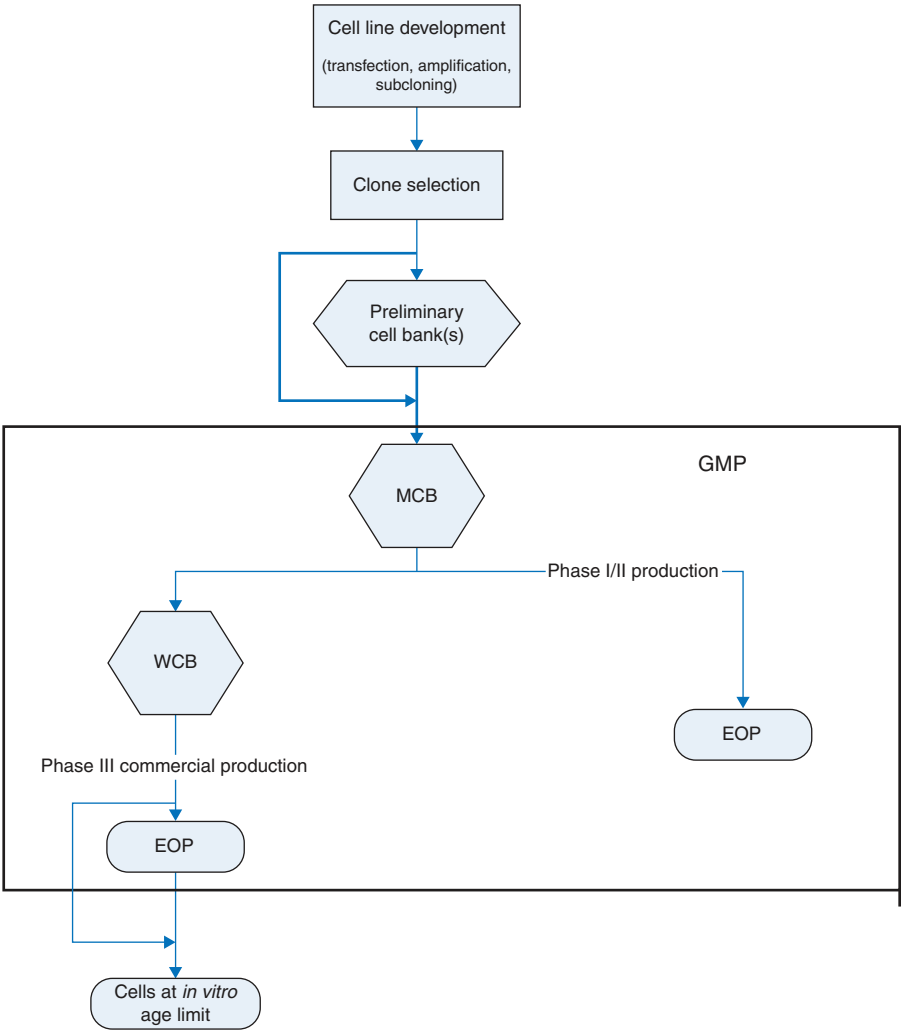


**FIGURE 7.2** Cell line lifespan during product development. EOP: end-of-production; GMP: good manufacturing process; MCB: master cell bank; WCB: working cell bank.

investigational new drug (IND). In general, these tests can be categorized into cell line identity, purity, and stability as described in the next sections.

***7.3.2.1 Identity*** A cell line identity assay is used to demonstrate that the species of origin of banked cells or EOP cells is consistent with the species of the host cells. Isoenzyme analysis is commonly used to confirm cell line species of origin by checking electrophoretic banding patterns for a number of intracellular enzymes [30]. This analysis is often used for EOP cells and cells at the limit of *in vitro* cell age

to provide assurance of no cross-contamination with cell lines of other species during the cell culture manufacturing process.

***7.3.2.2 Purity*** This assay is used to demonstrate that the production banks are biologically pure and free from adventitious microbial and cellular contaminants. Intensive purity assays should be performed to test for sterility and to demonstrate the absence of *Mycoplasma* and viruses. Adventitious virus contamination can be present in the cell bank or can be introduced during production through contaminated raw material (e.g., animal-derived cell culture media component) and manufacturing operations. Moreover, most rodent cell lines used in antibody production are known to carry endogenous viruses. For example, CHO cell lines harbor levels of non-infections retrovirus, and NS0 is considered capable of producing an infectious retrovirus (murine leukemia virus). Retrovirus infectivity can be detected by incubating samples with detector cell lines, such as mink $S^{+}L^{-}$ cells and rat XC cells. These retrovirus infectivity assays are often performed in direct or extended format because low levels of infection that may not be detected by direct assays can be detected in extended assays by serial passage of inoculated detector cells to amplify low levels of retrovirus [31]. The level of retroviral-like particles (RVLPs) can be quantified by thin-section electron microscopy (TEM) or reverse transcriptase (RT) assays. These assays should be conducted for the EOP cells to establish the extent of virus clearance that the purification process should achieve. Because RVLP count could change with cell density and process conditions, EOP cells should be reexamined when significant changes in the production process are made [32] to ensure that RVLP levels present in cell culture fluid do not exceed the downstream viral clearance capability. Process conditions affecting the RVLP expression in CHO cells have been extensively tested. Such studies concluded that RVLP specific productivity (particles produced per cell per unit time) is an intrinsic property of individual cell lines. Only cell culture process changes (e.g., temperature and pH shifts, butyrate addition) that alter the cell physiological state can impact RVLP specific productivity. In these cases, the EOP RVLP recount is recommended. Other operational changes (e.g., bioreactor scale, seed density, feed volume) do not generally impact RVLP-specific productivity. Therefore, the potential to increase RVLP levels can be estimated based on the increase in accumulated cell mass [33].

***7.3.2.3 Stability*** Cell line stability can be assessed from the phenotypic, genotypic, and product quality changes as a function of *in vitro* cell age. Phenotypic stability is established by evaluating cell growth and productivity after serial cultivations. A common approach is to passage cells in seed train medium to mimic the production seed culture process. Periodically, production process cultures in a scale-down model are initiated with aged cells to assess the process performance with regard to viable cell density, viability, titer, and specific productivity. Genotypic stability can be tested by assessing the cell expression construct, which should be analyzed in the MCB and at the limit of *in vitro* cell age to establish that the correct coding sequence is maintained in the production cell line during the manufacturing campaign. The gene copy number and integration site can be assessed by quantitaive

polymerase chain reaction and Southern blot, respectively. The fidelity of the product gene transcripts can be determined by cDNA sequencing. Because cDNA sequencing cannot detect low levels of protein sequence variants, protein amino acid analysis by peptide mapping with liquid chromatography/tandem mass spectrometry (LC/MS/MS) can be used to demonstrate that the protein amino acid sequence does not change during the cell culture process.

There is no defined guidance on how to apply these assays to different cell lines because the cell line testing requirements often vary case by case because of the complexity of product development stages, specifics of different cell types, and country-dependent regulatory requirements. Table 7.3 summarizes the cell line testing requirements for MCB, WCB, EOP cells, and cells at *in vitro* age limit during late stage product development as an example. In this case, the production cell line is derived from a NS0 host obtained from the European Collection of Cell Cultures (ECACC). The isoenzyme assay is used in every stage from the MCB to the cell age limit to confirm the species origin. Tests for sterility (bacteria and fungi) and *Mycoplasma* are also applied to every stage because these adventitious contaminants can be introduced any time during the cell culture operation. Adventitious virus tests include routine *in vitro* and *in vitro* virus tests. Mouse antibody production and bovine virus tests are also added to test adventitious virus because of the host cell species origin and the contact of the host cells with bovine serum during the initial cell line development. All of the tests for the presence of adventitious virus are performed on the MCB but not all of them on the WCB. Routine *in vivo* and *in vitro* tests are also performed for EOP cells and cells at the *in vitro* age limit. Endogenous virus tests include TEM, extended $S^+L^-$, and extended XC plaque, which are performed on the MCB but not on the WCB. TEM assay is also used for EOP cells and cells at the *in vitro* age limit to estimate RVLP levels. Cell line stability is assessed through phenotypic (cell growth and productivity), genotypic, and product identity tests. Cell growth and productivity are tested at all stages as they are in-process control assays. Product identity testing by peptide mapping with LC/MS/MS is performed for the antibody product purified from cell culture fluid derived from EOP cells and cells at the *in vitro* age limit. Genetic analysis of the expression construct in terms of cDNA sequence is performed for MCB, EOP cells, and cells at the *in vitro* age limit to demonstrate that the expression construct structure does not change during the cell culture process.

## 7.4 PROCESS DEVELOPMENT, SCALE UP, AND TECH TRANSFER

Because of the high level of uncertainty of biopharmaceutical clinical studies, process development is often divided into early stage and late stage with different focuses. The responsibility of early stage development is to develop bioprocesses to produce materials for phase I and II clinical trials and animal toxicology studies. To accelerate IND filing for "proof of concept" clinic studies, it is becoming increasingly necessary for companies to deliver the pipeline efficiently with streamlined cell culture platform processes using standardized process conditions and procedures.

**TABLE 7.3 Cell Line Qualification Testing Requirements During Late Stage Product Development**

| Assay | Test Purpose | Specification | Stage | | | |
|---|---|---|---|---|---|---|
| | | | MCB | WCB | EOP | Cell Age Limit |
| Isoenzyme | Cell line identity | Consistent with species origin | √ | √ | √ | √ |
| Sterility | Microbial contamination | No growth | √ | √ | √ | √ |
| *Mycoplasma* | *Mycoplasma* contamination | None/detected | √ | √ | √ | √ |
| *In vitro* adventitious viruses | Routine adventitious virus | None/detected | √ | | √ | √ |
| *In vivo* adventitious viruses | Routine adventitious virus | None/detected | √ | | √ | √ |
| Bovine viruses | Selected adventitious virus | None/detected | √ | | | |
| Mouse antibody production | Selected adventitious virus | None/detected | √ | | | √ |
| TEM | Retroviral test | RVLP count and type | √ | | √ | √ |
| Extended $S^+L^-$ | Retroviral test | None/detected | √ | | | |
| Extended XC plaque | Retroviral test | None/detected | √ | | | |
| Cell growth and productivity | Phenotypic stability | Consistent performance | √ | √ | √ | √ |
| cDNA sequence | Genotypic stability | No mutation | √ | | √ | √ |
| Peptide mapping (drug substance) | Product identity | Consistent amino acid sequence | √ | | √ | √ |

EOP: end-of-production; MCB: master cell bank; RVLP: retroviral-like particle; TEM: thin-section electron microscopy; WCB: working cell bank.

The use of platform process allows the acceleration of early stage cell culture process development activities such as clone selection, process lock, and tech transfer to clinical manufacturing with increased efficiency. The similarity of molecular characteristics and properties among different mAbs makes process platform approach feasible, although the processes may not be fully optimized for every molecule. As the product moves into the late stage for phase III and commercial production, the processes can be further optimized during the time frame between early phase and late phase production when clinical studies are being conducted. Late stage process development is then focused on improvements of process yield, robustness, scalability, and regulatory compliance. With process changes during both early and late stage development, product quality and comparability need to be closely monitored for patient safety. To gain biologics license application (BLA) common technical document (CTD) regulatory approval for commercialization, the processes also need to be characterized to evaluate the effects of process parameters on process performance and product quality and then validated to demonstrate process consistency before commercial cGMP production. Figure 7.3 illustrates the process development activities at different clinical development stages.

A cell culture process is composed of three parts: cell line, cell culture media, and bioreactor environmental conditions. Therefore, the cell culture process optimization is an integrated activity involving clone selection, medium development, and bioreactor condition optimization. After the cell culture process is locked, it needs to be transferred from the development laboratory to a production facility. In early stage development, speed to clinical studies is critical, and toxicity and phase I and II

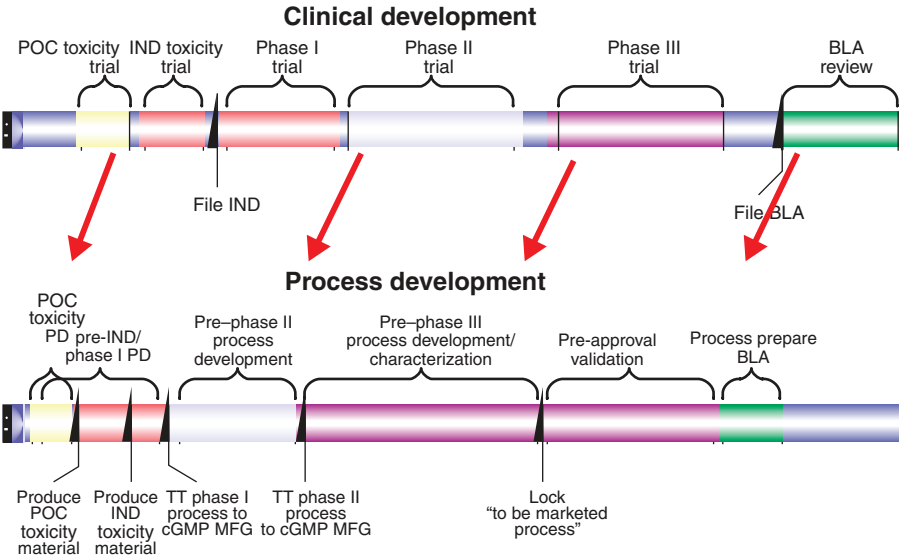


**FIGURE 7.3** Clinical and process development flowchart. BLA: Biologic License Application; cGMP: current good manufacturing process; IND: investigational new drug; MFG: manufacturing; POC: proof of concept.

material deliveries are often on the critical path. In the late stage, costly large-scale production runs are deferred until results from early clinical trials are available. After the product succeeds in its pivotal clinical trial, speed to market becomes critical to achieve commercial success. This approach does not allow much time for engineering and shakedown runs. Therefore, successful scale up in a limited time frame is highly desirable.

### 7.4.1 Clone Selection

Selection of the final production clone is often considered one of the most critical decisions in cell culture process development because changing production cell line during clinical development is considered a major process change, which requires considerable efforts to demonstrate product comparability before and after the change. For example, changing the cell line after phase III production typically requires additional human clinical studies. Thus, it is important to select the right clone before phase III clinical production and preferably at the phase I stage.

As described in Section 7.2.2, after being transfected, the cells are diluted and cultivated in 96-well plates with a basal growth medium and screened for clones with robust cell growth and high productivity. The selected clone candidates based on productivity and growth characteristics are then adapted to production media and suspension culture. At this stage, to predict clone performance in large-scale production bioreactor, an enriched medium similar to the final production medium formulation with a similar feeding regimen can be applied in shake flasks or small-scale bioreactors. Several clone attributes should be considered and evaluated, including product quality, manufacturability, and productivity.

***7.4.1.1 Product Quality*** Maintaining consistent and comparable product quality is often challenging because of changes of cell line, process, and scale conditions as product candidates are moved forward from development laboratory to pilot plant and cGMP manufacturing. Among these variable conditions, the product cell line has the most significant impact on many quality attributes because some product quality attributes are often clone dependent. Table 7.4 summarizes some common analytical assays and quality assessment criteria used during clone selection to test mAb molecular properties such as sequence homogeneity, charge variants, aggregation, and glycosylation.

***7.4.1.2 Productivity, Stability, and Scalability*** Process yield obviously is another critical criterion to select the production clone. Because antibody expression rate in mammalian cells is non–growth associated, the final titer is equal to specific productivity $Q_p$ multiplied by the integral of viable cell density over culture duration ($\text{Titer} = Q_p \cdot \int X \, dt$). In today's industry standard, a production cell line could have a specific productivity from 20 to 80 pg/cell/day and peak viability cell densities from 5 to $20 \times 10^6$ cells/ml in a 10- to 14-day fed-batch process. Cell line stability is another factor that needs to be considered. It is sometimes found that productivity declines as cell age increases for some cell lines. Unstable clones are not suitable for

**TABLE 7.4 Product Quality Assessment During Clone Selection**

| Quality Attribute | Analytical Assay | Quality Assessment Criteria |
|---|---|---|
| Molecule integrity | cDNA sequence (qPCR), peptide map (LC/MS/MS), CE-SDS, or SDS-PAGE | Avoid amino acid sequence mutation and truncated antibodies |
| Aggregation | SEC | Avoid high level of aggregation, which could be immunogenic |
| Glycosylation | HPLC- or CE-based glycan assay | Avoid high levels of unusually glycosylation forms |
| Charge heterogeneity | IEF or iCIEF, IEC, HPLC | Avoid usually high levels of acidic or basic variant or extra acidic or basic peaks |

CE: capillary electrophoresis; HPLC: high-performance liquid chromatography; iCIEF: imaged capillary isoelectric focusing; IEC: ion-exchange chromatography; IEF: isoelectric focusing; LC: liquid chromatography; MS: mass spectrometry; qPCR: quantitative polymerase chain reaction; SDS-PAGE: sodium dodecyl sulfate polyacrylamide gel electrophoresis; SEC: size exclusion chromatography.

large-scale production because the cell culture process is scaled up through serial culture passages of the seed train and inoculum train, and therefore the cell age increases with scale. In addition, cell growth and metabolic characteristics, which can impact cell culture process robustness and scalability, also need to be assessed. For example, robust cell growth with high viability is desirable, but high lactate-producing clones are not preferable to avoid dramatic culture pH changes.

Screening and selection of highly productive and scalable clones among the transfectant population in a limited time frame are still major challenges because the product quality, productivity, and even cell metabolic profiles often depend on cell culture conditions. As described in Section 7.5.2, using miniaturized high-throughput bioreactors with full-process parameter controllability to mimic the large-scale bioreactor environment can help to identify the best production clone at the very early stage.

### 7.4.2 Medium and Feeding Strategy Development

In general, medium development for a fed-batch process involves batch medium and feed concentrate development as well as feeding strategy optimization. Optimization of cell culture media is often considered as cell line dependent and needs to be based on the metabolism and nutrient consumption by a specific cell line. Several approaches can be used systematically, such as single-component titration, spent medium analysis, and medium blending [34]. Because of safety concerns for transmissible spongiform encephalopathy (TSE) and other transmissible contaminants, bovine serum and animal-derived raw materials should be avoided if possible. After two decades of media development and host cell adaptation, fully chemically defined media consisting of amino acids, vitamins, trace elements, inorganic salts, lipids, and insulin or insulinlike growth factors have been developed and implemented in large-scale mAb production. But not all antibody production cell lines can achieve

high yield in fully chemically defined media. Addition of animal component–free hydrolysates to the chemically defined media is a common approach to increase cell density, culture viability, and productivity in a timely manner. Hydrolysates are protein digests composed of amino acids, small peptides, carbohydrates, vitamins, and minerals, which provide nutrient supplements to the media. In addition, hydrolysate peptides can also act as growth factors and production stimulators. Non–animal-derived hydrolysates from soy, wheat, and yeast are commonly used in cell culture media and feeds. However, because of its composition complexity and lot-to-lot variations, hydrolysate can be a significant source of medium variability.

The most common approach to develop a feed medium is to concentrate basal media without salts to avoid high osmolality. Some key feeding components such as phosphate have also been identified [35]. During medium preparation, pH, temperature, or both may need to be adjusted to completely dissolve some low-solubility components [36]. Optimization of a feeding strategy needs to consider nutrient consumption, byproduct accumulation, and growth–production balance. Previous studies indicated that byproducts such as lactate and ammonia could be minimized by maintaining low-glucose and low-glutamine concentrations through frequent feeding [37]. However, the operation complexity for frequent or continuous feeding makes it less desirable for large-scale manufacturing, although this should not be a major challenge because they are routinely applied to microbial fermentations. Stepwise bolus addition of the feed solution to the production bioreactor is most widely used in industry because of its simplicity and scalability. In general, medium development is labor intensive and time consuming. A combination of high-throughput cell culture scale-down systems with statistical design of experiment approaches is commonly applied to shorten the development time.

### 7.4.3 Bioreactor Optimization and Scale Up

A typical stirred tank bioreactor is equipped with temperature, pressure, agitation, pH, and dissolved oxygen controls. Culture operating parameter optimization is required to achieve high expression of product with acceptable product quality profiles. These parameters are physical, chemical, and biological in nature. Physical parameters include temperature, gas flow rate, and agitation speed, and chemical parameters include dissolved oxygen and carbon dioxide, pH, osmolality, redox potential, and metabolite levels, including substrate, amino acid, and waste byproducts. Biological parameters are used for determining the physiological state of the culture and include viable cell concentration, viability, and a variety of intracellular and extracellular measurements such as NADH (nicotinamide adenine dinucleotide + hydrogen), lactate dehydrogenase levels, mitochondrial activity, and cell cycle analysis. Variations in the microenvironment parameters from optimal levels can have a dramatic impact on culture performance, productivity, and product quality. Figure 7.4 illustrates the cell culture operating parameters' impacts on the culture environments such as dissolved oxygen (DO), pH, osmolality, dissolved $CO_2$, mixing, hydrodynamic shear, and process performances (e.g., cell growth, metabolite concentrations, product titer, product quality).

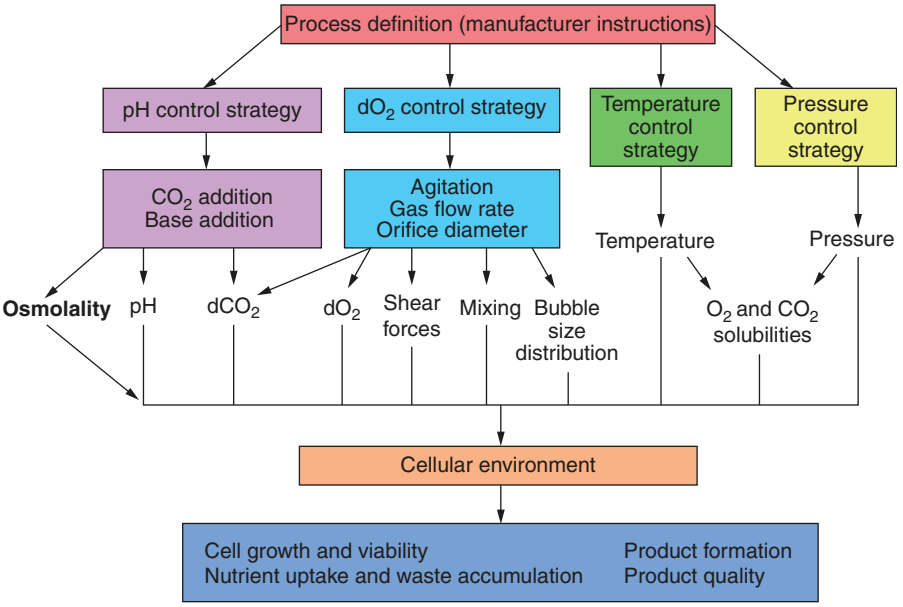


**FIGURE 7.4** Cell culture process operating parameters affect process performance and product quality.

Because of the complexity of protein products with isoforms and microheterogeneities, the performance of the cell culture process can have significant impacts on product quality and potency, especially with respect to glycosylation, post-transcriptional modifications, and impurity profiles. Therefore, the bioreactor operating parameters have to be optimized and characterized thoroughly to improve process performance and to understand the process impact on product quality. Significant progress has been made in the past two decades in understanding the impact of the cellular environment on culture physiology and its subsequent impact on productivity, product quality, and downstream processes. Another reason for the resurgence of research activity and innovation in this area can be attributed to the Food and Drug Administration's (FDA) process analytical technology (PAT). The PAT directive (http://www.fda.gov/Cder/OPS/) is a framework for controlling manufacturing consistency and product quality through measurements and control of (both online and offline) of critical quality attributes of materials as well as processes. In addition, it also encourages a better understanding of the manufacturing process, identifying relationships among process parameters and reproducibility, culture performance, and product quality attributes. Although originally conceptualized for small molecule manufacturing processes, it is increasingly being extended to biological processes, which are inherently more complex because of the nonlinear impact of a large number of variables.

Process parameters can be measured either online (or at-line by direct connection to calibrated analyzers) or offline via operator intervention. Typical examples of

offline measurements include pH (usually for verification of online pH readings and compensation for sensor drift), cell counting and viability measurements using a hemocytometer or automated cell counters, packed cell volume, osmolality, and metabolites. Accurate gas and liquid flow measurements are typically conducted using mass flow meters and magnetic meters.

Temperature can be measured accurately using commercially available resistance temperature devices, although thermocouples can also be used. A temperature shift to manipulate the cell cycle in CHO cell culture has been reported to extend culture longevity. A temperature shift from 37 to 30°C after 48 hours postinoculation can retain cells in G1 phase longer and therefore delay the onset of apoptosis [38].

Dissolved oxygen is typically between 20 and 50% air saturation to prevent dissolved oxygen limitation and excessively high values that can lead to cytotoxicity. The Clark-type electrode is still the most commonly used for dissolved oxygen. In this type of electrode, oxygen diffuses through a permeable membrane and is reduced at a negatively polarized cathode (platinum) with respect to a reference anode. The cathode and anode are separated by an electrolyte solution, and the reduction of oxygen generates a voltage-dependent current that is directly proportional to the dissolved oxygen concentration. Although cell growth is relatively insensitive to dissolved oxygen in the range 20 to 100% air saturation [39,40], the dissolved oxygen level could have significant impact on product quality. It has been reported that reduction in dissolved oxygen caused a decreased glycosylation of antibody *N*-glycan chains [41].

Dissolved $CO_2$ is also an important process variable that can accumulate to inhibitory levels at values greater than 120 to 150 mm Hg [42–44] or impact product quality at high cell density cultivations or scale up because of inadequate $CO_2$ stripping. Commercially available sensors are available from Yellow Springs Instruments (YSI, Yellow Springs, OH) and Metler Toledo (Columbus, OH). However, in many cell culture processes, dissolved $CO_2$ is still measured using offline blood gas analyzers such as those from Nova Biomedical (Waltham, MA). The use of the *in situ* sensors could allow for the online control of dissolved $CO_2$ via air sparging, although this is rarely implemented in cell culture applications, perhaps because of the interrelationship of air sparging and dissolved oxygen control or that it is generally possible to match $CO_2$ profiles at small and large scales by the correct choice of bioreactor sparger design, agitation speed, and gas sparge rates. Measurement of oxygen uptake and $CO_2$ production rates can also be achieved using online MS (generally multiplexed to multiple bioreactors). Unlike microbial fermentations, this type of analysis is rarely conducted either because of the cost of such instrumentation or the sensitivity of measurements at very low rates of metabolism.

For mammalian cell culture, pH is an important variable to measure and is normally controlled tightly. Even a small deviation of 0.1 unit from the optimal pH value can significantly impact culture growth and metabolism, particularly glucose consumption and lactate production [45,46]. High pH ($\geq$7.0) is usually preferred for initial cell growth; however, a high pH condition is also associated with increased glucose metabolism by converting more glucose into lactate. When lactate accumulation exceeds the medium buffer capacity, the culture medium pH decreases.

Because most processes control pH at a certain level, lower pH of the culture medium triggers base addition to the culture, which increases the osmolality of the culture medium. This high pH, high lactate, and high osmolality cascade often causes delayed cell growth and accelerated cell death. Thus, the pH control condition needs to be optimized at the beginning of the cell culture process to maximize the initial cell growth. When fast cell growth is achieved, the pH set point can be shifted to a lower level to avoid lactate accumulation and to facilitate antibody production. It is also important to note that the pH set point and control strategy (e.g., dead band) are intimately linked to dissolved $CO_2$ levels and base consumption for pH control and therefore osmolality. These considerations make the accurate measurement of pH even more imperative. pH probes typically consist of a 0.2- to 0.5-mm glass membrane fabricated from silicate groups containing sodium, which form a hydrated gel layer on sides of the membrane when placed in liquid. An $Ag/AgCl_2$ electrode in KCl electrolyte saturated with $AgCl_2$ is used to measure the change in potential of the outer surface of the membrane caused by changes in the medium pH. Some issues related to the use of pH probes for extended culture duration include drift and reduced sensitivity.

Most cell culture media are designed to have an osmolality in a range of 270 to 330 mOsm/kg. The impact of osmolality is cell line specific, with reports of growth inhibition with increasing osmolality but no effect on cell specific productivity [47]. In many cases, it is difficult to decouple the impact of dissolved $CO_2$ and culture osmolality. Culture osmolality is routinely measured using freezing-point depression osmometry. In this method, the total osmolality of an aqueous solution is determined by comparing the freezing point of pure water and the freezing point of the sample. For example, water has a freezing point of 0°C, and a solution with saline concentration of 1 Osm/kg has a freezing point of −1.86°C. Various commercial systems are available, although the analyzers from Advanced Instruments (Norwood, MA) are the most commonly used in cell culture and are capable of measuring osmolality in the range of 0 to 1500 mOsm/kg of water.

Cell culture metabolites such as glucose, lactate, glutamine, and glutamate are commonly measured offline using enzymatic biosensors specific to the measured analyte. These measurements are important not only for maintaining substrate levels above critical levels via feeding strategies but also in developing processes with reduced byproduct formation. Commercially available instruments include the YSI 2700 from YSI Nova Bioprofile 400 from Nova Biomedical. These analyzers use biosensors in which the reaction of a substrate with the enzyme produces a product that can be readily measured. For example, glucose can be detected using glucose oxidase in the presence of excess oxygen to produce gluconic acid and hydrogen peroxide. A second reaction between fluorescein and the $H_2O_2$ produced from the first reaction produces a fluorescent signal that correlates directly to the glucose concentration. In addition to the quantification of metabolites, the Nova Bioprofile 400 can also measure pH, dissolved gases ($O_2$ and $CO_2$), $NH_4^+$ (0.20–25.00 mM), $Na^+$ (40.0–220.0 mM), $K^+$ (1.00–25.0 mM), and $Ca^{2+}$ (0.10–10.0 mM). It is important to point out that accurate measurements with these systems require proper maintenance and calibration procedures.

Cell concentration and viability measurements are critical in developing cell culture processes and are used for determining culture physiology in response to operating conditions and calculation of growth rates, specific consumption and production rates of metabolites, and cell-specific productivity. The most common method for measuring cell concentration and viability is via trypan blue exclusion microscopy. This method depends on the ability of cells with intact membranes to exclude trypan blue (live cells), and those that are stained blue are classified as dead cells. A number of automated analyzers exploiting this method and coupled to image analysis are commercially available (Cedex, Innovatis, Bielefeld, Germany; Vi-Cell, Beckman Coulter, Fullerton, CA). These analyzers also allow the determination of cell size. In many cell culture processes, the mean cell size can change by as much as 50% (or more than a threefold change cell volume). In these cases, it is insufficient to measure viable or total cell concentration. Instead, packed cell volume (or viable packed cell volume) is determined using centrifugation-based hematocrit methods. This is a more accurate representation of the cell mass, which is ultimately linked to metabolite profiles and product formation. The trypan blue exclusion method does not allow measurements other than live or dead. Multiparameter flow cytometry using fluorescent dyes such as propidium iodide and annexin V or caspase 5 can be used to quantify live or dead cells as well as apoptotic cell populations. This is now possible in cell culture laboratory settings using relatively simple 96-well plate–based systems and specifically designed measurement kits for such analyses as well as a multitude of other cell-based assays (Guava Technologies, Hayward, CA). Measurements of cell metabolic activity can be quantified by online oxygen uptake rate measurements or offline cell mitochondrial activity using fluorescence-activated cell sorting (FACS).

Process scale up implements process at larger scale to achieve similar process performances as in small scale. A typical large-scale cell culture process consists of a number of unit operations, including seed train, inoculum train, and production run

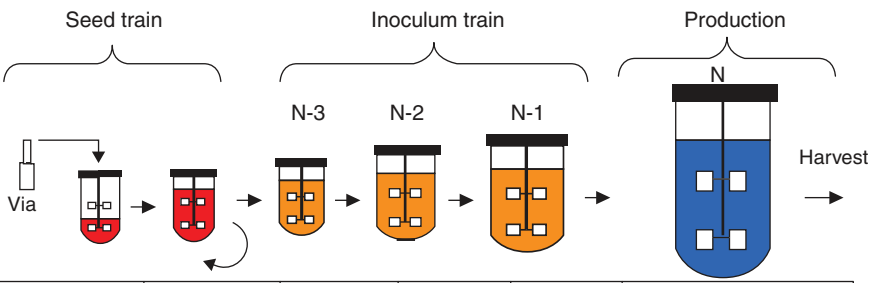


| Manufacturing scale | 20 l | 80 l | 400 l | 2000 l | 12,000 l |
|---|---|---|---|---|---|
| Scale-down model | 2 l | 2 l | 2 l | 2 l | 2 l |
| Key performance parameter | Growth rate, final viability | Growth rate, final viability | | | Titer, product quality |

**FIGURE 7.5** A large-scale cell culture process steps and their key process performance parameters.

as described in Figure 7.5. Each step has its key process performance parameters such as cell growth, viability, titer, and product quality. The success of cell culture process scale up is measured by these key process parameters meeting their predefined criteria.

Mammalian cells are generally considered difficult to be scaled up because of their sensitivity to physicochemical conditions. Scale up of high cell density process to large-scale bioreactors by providing enough oxygen supply and $CO_2$ removal can be challenging in some cases and requires a proper selection of large-scale bioreactor operating parameters. Bioreactor operating parameters can be categorized into volume-dependent parameters (e.g., working volume, feed volume, agitation, aeration) and volume-independent parameters (e.g., pH, dissolved oxygen, temperature). A general strategy for scale up is to proportionally scale up the volume-dependent parameters while maintaining the volume-independent parameters at the same set points used in the small-scale process. However, some volume-dependent parameters are difficult to be scaled up or down linearly because of differences in bioreactor geometry, ratio of liquid surface to volume, gassing regimen, and control capability. During bioreactor operating condition scale up, agitation and aeration are two important conditions to be scaled up properly to achieve comparable process performance across different scales. Whereas agitation is required to provide sufficient mixing and oxygen mass transfer and is generally scaled using equal specific energy dissipation rate, aeration is used to supply of oxygen and remove of excess $CO_2$. In small-scale bioreactors, the operable ranges of operating parameters are wide because mixing is generally not a problem even at relatively low agitation speeds, and dissolved $CO_2$ removal is efficient because of a large surface-to-volume ratio. In large-scale bioreactors, these parameters need to be chosen carefully because agitation-related shear stress and aeration-related $CO_2$ accumulation increase with scale.

Besides large-scale bioreactor operation issues, other common pitfalls of cell culture process scale up include lot-to-lot variations of raw material, consistency of media preparation, media hold stability, and cell line stability during the long duration of manufacturing process as described previously. These issues can often be identified in scale-down models. A qualified cell culture scale-down model, which is able to reproduce large-scale process performance at small scale, provides crucial information needed to assess and resolve any unexpected performance problem of the large-scale system. Beyond manufacturing support, the scale-down model can also be used for process characterization and validation experiments to study acceptable ranges of process operating parameters and determine the critical operating parameters.

### 7.4.4 Process Characterization and Validation

The FDA defines process validation as "establishing documented evidence that a specific process will consistently produce a product meeting its predetermined specifications and quality attributes." For a biological product, all process validation data should be submitted and the specifications are established and approved as part of the BLA. This regulatory requirement reflects the recognition that the product

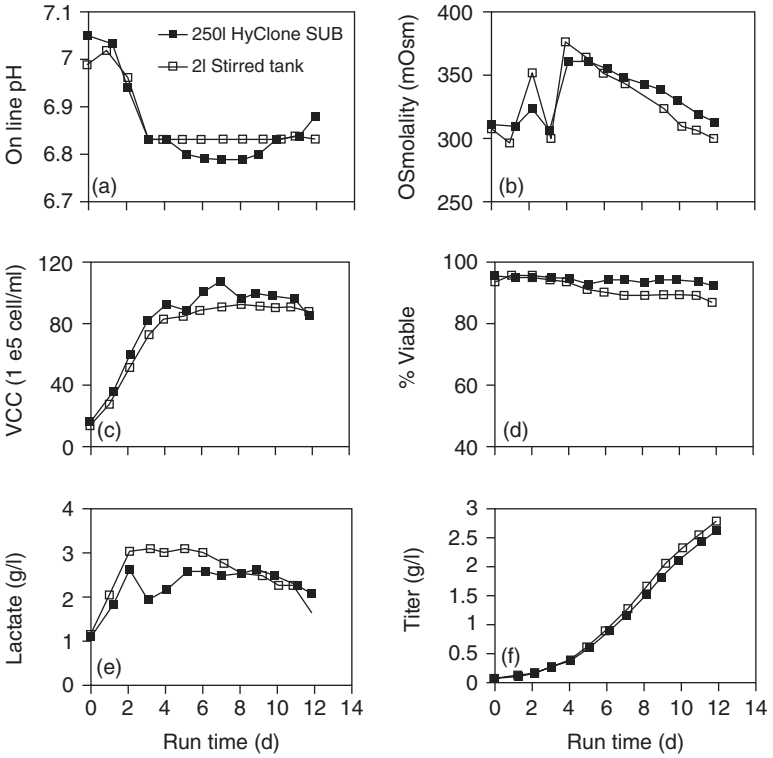


**FIGURE 7.6** Cell culture performance in a 2-l stirred tank bioreactor (□) and 250-l HyClone single-use bioreactor (SUB, ■): online pH (a), osmolality (b), viable cell concentration (c), percent viability (d), lactate concentration (e), and titer (f).

quality can be substantially affected by the manufacturing process. To meet this requirement, the process needs to be fully characterized and validated to understand the impact of process inputs (operating parameters) on process outputs (performance parameters). Figure 7.6 illustrates a typical cell culture manufacturing process and the key performance parameters of several unit operations, including seed train, inoculum train, and production. Because key performance parameters are often different for different cell culture steps, each operation usually needs to be characterized and validated separately. The full process characterization and validation usually occurs after the completion of phase III process development when the manufacturing process is locked and no further changes are expected. Although some process validation work can be performed using qualified scale-down models, full-scale process validation is required to demonstrate consistent process performance and product quality when controlling critical process parameters within predefined acceptable ranges at commercial scale. For validation purposes, normally three consecutive full-scale runs are required for BLA licensure in the United States.

The objectives of process characterization include identification of key operational and performance parameters, establishment of acceptable ranges for key parameters,

and demonstration of process robustness [1,48,49]. Technical information from the characterization study has also become a regulatory expectation in recent years as a prerequisite for manufacturing process validation as well as for long-term commercial manufacturing support [50,51]. Because performing a characterization study at the manufacturing scale is not practically feasible because of the cost of operation and limited availability of large-scale bioreactors, scale-down models that represent the performance of manufacturing scale process are usually used. It is also important to use qualified analytical methods and raw materials throughout the characterization study for consistency and accuracy. Cell culture process characterization requires support from the purification and analytical groups to test process impact and to monitor product quality. Some cell culture materials from characterization studies can be used as feed stream for downstream process characterization, especially for the robustness study.

### 7.4.5 Process Technology Transfer and Current Good Manufacturing Process Manufacturing

After a scalable process is locked at the laboratory and pilot scales, it needs to be transferred to a manufacturing facility for cGMP production of clinical or commercial quantity of product. The purpose of process technology transfer is to transfer of all of the necessary process technology, information, documentation, equipment, materials, and tools in order to implement the manufacturing process at a specific facility for clinical and commercial material production and obtain the regulatory approval.

The tech transfer typically requires collaborations among different groups, including process development, manufacturing, and quality. The activities required to successfully transfer a cell culture process can be organized into several stages, including:

1. Process–facility gap analysis
2. Process, facility, equipment modification, and qualification
3. Engineering runs at full scale if needed
4. cGMP commissioning
5. cGMP production (comparability study for process and product)

As a new process being introduced into an existing facility, a gap analysis should be conducted to identify the gaps and potential risks to fit the manufacturing process into the facility. The facility and equipment modifications and installation qualification/operational qualification/performance qualification (IQ/OQ/PQ) activities should be completed before the full-scale production runs. Moreover, as the product moving from early stage (phase I and II clinical) to late stage (phase III commercial) production at different scale facilities, the process may need to be modified to reflect scale and facility changes. In this case, the modifications should be minimized to changes needed to enable the facility fit and to focus on critical quality attributes and the changes need to be qualified and incorporated into the comparability evaluation to ensure the process and product quality consistency before and after manufacturing

site change. In general, regulatory authorities do not require that the pre- and postchange product is identical but rather that its physicochemical properties and biological activity are highly comparable and that these changes have no impact on the safety or efficacy of the product.

Non-GMP engineering runs are conducted to test process performance at full scale to ensure process and product quality consistency. Before the initiation of the cGMP production, cGMP commissioning should also be completed to make sure facility, raw material, documentation, personnel training, quality control, and production systems are ready.

## 7.5 PROCESS ADVANCES AND FUTURE PERSPECTIVES

### 7.5.1 Advances in Process Online Monitoring and Control

Monitoring of cell cultures is essential for the development and optimization of highly productive processes. In addition, the ability to develop control strategies depends on the availability of robust and accurate *in situ* sensors or at-line instruments coupled to a sampling system to measure relevant parameters. This section gives a brief discussion of the use of commercially available sensors, advances in optical and noninvasive sensor technologies, and autosampling systems to enable at-line monitoring.

***7.5.1.1 Dielectric Permittivity and Electrical Impedance Spectroscopy*** Offline measurements of cell mass do not easily permit the use of growth-based feedback control strategies for feeding. Online optical cell density probes based on light backscatter have been used successfully to monitor cell cultures [52]. Such measurements are generally linear with cell concentration only at high viabilities and deviate significantly from linearity with decreasing culture viability toward the latter stages of fed-batch cultures. Dielectric permittivity and electrical impedance spectroscopy can be used successfully to monitor viable cell volume [53]. These measurements make use of the fact that live cells consist of poorly conductive lipid membranes that separate a conductive cytoplasm from a conductive extracellular matrix. When an electric field is applied, the cells become polarized and behave as tiny electric capacitors. The capacitance of the cell suspension gives a measure of the extent of the field-induced polarization, which is related to the viable cell volume or concentration, provided the cell diameter does not change appreciably.

***7.5.1.2 Near Infrared Spectroscopy*** Spectroscopy, particularly using near or mid infrared (NIR or MIR), is an attractive alternative for the measurement of cell culture components, including substrates, waste products, amino acids, cell concentration, and viability. Sample preparation is not required, and this technique can be used to measure a large number of parameters simultaneously. A light source is introduced to the sample via a steam-sterilizable probe mounted using standard bioreactor ports. Up to nine channels are available on a commercially available unit

from Foss NIR Systems (Laurel, MD). The resulting transmittance measurement can be correlated to individual components and concentrations in a sample matrix. Because of the complex nature of the spectra, partial least-square (PLS) regression methods are used for deconvolution. Initially, the NIR spectral scans are trained using offline measurements of cell counts, osmolality, metabolites, and amino acids or other cell culture media components. This calibration model is then used to predict culture components in spectral scans in independent bioreactor cultures. Several studies have been published demonstrating the usefulness of this technique for monitoring cell culture media components [54–57]. Despite this, NIR is rarely applied in cell culture processes (even in a laboratory setting). This may be because the spectroscopic measurements (position and intensity of spectrum) are sensitive to changes in process parameters such as pH, ionic strength, and temperature. This is especially true for low-concentration components in the presence of a complex matrix background. Significant challenges remain for the successful application of this promising technology for cell culture monitoring in terms of establishment of robust calibration methods and PLS-based data analysis that can be implemented routinely by nonexpert operators.

***7.5.1.3 Noninvasive Fluorescence Sensors*** Noninvasive fluorescence sensor technologies have rapidly emerged as an important development for online monitoring of cell culture parameters such as optical density, pH, and dissolved oxygen, especially for microtiter plates for high-throughput applications [58]. Such sensors are commercially available from PreSens Sensing GmbH (Regensburg, Germany) and Fluorometrix (Stowe, MA). This technology has also been incorporated in optical probes for pH, dissolved oxygen, and dissolved $CO_2$ (Polestar Technologies Inc., Needham Heights, MA), Sartorius (Edgewood, NY), and YSI, as well in shake flasks (Fluorometrix and PreSens Sensing GmbH). The dissolved oxygen and pH detection uses dynamic luminescence quenching of the specific fluorophore by oxygen or protons per the Stern-Volmer equation [59]. These measurements use dual referenced optical sensing whereby the relative emission ratio is measured from the shared spectrum from an indicator fluorophore sensitive to pH or DO. The recessed and external light protected LEDs under the emit light at 470 nm for pH sensing and 505 nm for DO. Further research is ongoing to develop robust dissolved $CO_2$ sensors and biosensors for culture substrate and metabolites.

***7.5.1.4 Automated Sampling Systems*** In addition to online measurements, a number of measurements have to be conducted using offline analyzers with manual intervention, which increases contamination risks as well as places a burden on resources. However, such measurements can also be conducted at line using automated sampling systems and with data transfer to control systems can be used in control strategies. Until recently, there has been a paucity of commercial autosamplers, and when available, they have been limited to fraction collectors. In most cases, this requires operator intervention for sample analysis. Online sampling could fully eliminate the labor-intensive, low-value-added processes of manual reactor sampling and data entry and eliminate operator-introduced sampling

variability. However, autosamplers also pose concerns for increasing contamination risks, robustness, ease of use, and hold-up sample volume requirements (especially at small scales). Recently, Groton Biosystems (Groton, CT) has introduced four- and eight-bioreactor sampling systems (ARS400 and ARS800), which are capable of user-defined sterile sampling and can be connected up to four separate analytical devices. The system is also configured with cleaning and sanitizing reagents to allow a low bioburden environment in the sampling tubes. The sterility barrier between the analyzer and bioreactor contents is provided via a reactor valve interface.

### 7.5.2 Miniaturized High-Throughput Scale-Down Models for Cell Culture

To deliver a robust and productive process while maintaining aggressive timelines to introduce the molecular candidate in to the clinic, increasing throughput of bioprocess development has become progressively important to rapidly screen and optimize process parameters [60,61]. Traditionally, small-scale systems such as shake flasks and spinner flasks have been used to screen a large number of clones and experimental conditions. However, because of the inability to monitor and control environmental parameters and the inability to conduct fed-batch cultivations on a routine basis and the laborious nature of shake flask experimentation, these models are less desirable for bioprocess optimization.

De Jesus et al. [62] developed a scaled-down system for suspension cell culture, coined the "TubeSpin," using 50-ml centrifugation tubes as culture vessels mounted on rotational shakers. The TubeSpin can be configured with vented caps, thus allowing the exchange of gases via the headspace. Although this system does not permit the *in situ* measurement and control of pH and dissolved oxygen, it is nevertheless suitable for large screening experiments in which such information may not be crucial. It is also conceivable that the TubeSpin could be linked to a liquid handling system to allow an extension of its capabilities. A number of micro-bioreactors are also commercially available at scales of 50 to 300 ml (Fluorometrix, AC Biotech, Julich, Germany; Infors AG, Bottmingen, Switzerland; DASGIP, Julich, Germany). Bioprocessors Corp. (Woburn, MA) has developed a fully integrated robotics platform (Simcell) with its proprietary microbioreactor arrays (each array consisting of six gas-permeable microreactors with submilliliter working volumes) for true high-throughput cell culture applications. The Simcell system appears to be very promising and has been acquired by several companies. Although results have been presented at conferences [63], there are no published reports in the literature.

### 7.5.3 Disposable Cell Culture Technologies

Traditionally, stainless steel tanks have been used at laboratory and pilot scales for process development and production of research grade, toxicology, and phase I clinical materials. Stainless steel tanks also dominate large-scale manufacture (>1–25,000 l) of biotherapeutics. However, the use of fixed plant equipment is costly, requiring long lead times for installation of the tanks and supporting infrastructure

and qualification. There is also a high burden from validation efforts related to sterility and cleaning, as well as maintenance.

Disposable technologies offer significant advantages over traditional fixed plant equipment, particularly at pilot scales of operation. They can be introduced rapidly into a laboratory and manufacturing facility because installation, qualification, and personnel training efforts are minimal, require few utilities, are provided clean and presterilized, and offer increased reliability because of their one-time use. In addition, they can increase plant capacity and flexibility by reducing turnaround time (especially in the event of contaminations); decreasing set-up time; and demanding a smaller footprint because of significantly reduced piping, valve, and instrumentation requirements. Implementing design changes is also more rapid with disposables, allowing for continuous improvement and integration of new technologies such as online monitoring systems. Operational flexibility is essential because of changing priorities in development. In addition, facility change-over time and effort can be reduced, especially in cGMP applications in which preventing cross-product contamination is of crucial importance. Overall, these advantages lead to significantly lower capital costs and lower resource (personnel) requirements, which are key factors for both large and small companies alike.

Cell culture processes are becoming more productive because of advances in cell line engineering and process development, which in turn allows the use of smaller scales. As a result, disposables are gaining popularity. Use of disposables facilitates the production of gram quantities of research grade material, scale up, and handling of multiple products in a current good laboratory practice (cGLP) or cGMP facility for the manufacture of toxicology or even phase I material. Although disposable technologies in bioprocessing covers a wide range of components such as bioreactors, tubing and pipes, pumps, filters and membranes, valves, sampling devices, bottles, bags for media and buffer preparation, and sensors (to name a few) and the applications range from culture cultivation, purification (cell separation, buffer bags, in-process pool containment, prepacked chromatography columns), and bulk drug storage, the focus of this section is on single-use bioreactors (SUBs) for cell culture processes.

***7.5.3.1 Single-Use Bioreactors Integrated with Disposable Bags*** Single-use bioreactors at the laboratory and pilot scales (1–1000 l) incorporate disposable bag technology at the core of their design and are capable of providing mixing and mass transfer and control of process parameters (air flow rate, gas mixing, temperature, dissolved oxygen, pH) similar to conventional stirred tanks. Both traditional and noninvasive sensors have been used. These are described later. The bags are generally manufactured with class VI United States Pharmacopeia (USP) materials. For many years, liners and bags composed of these materials have been used for media make-up, storage, and transfer, so cultivating cells inside such materials should not be viewed as a substantial risk. Before the implementation and use of SUBs, they must be qualified in terms of extractables or leachables (e.g., total organic content) and vapor transmission from the material of construction of the bags used as well as integrity and sterility testing using standard methodologies such as the use of model solvents and media hold tests under processing conditions. This is

generally done with the complete bag design, which includes the bag, fittings, and tubing. The extractables or leachables information is also generally available from the vendors. In addition, SUBs have to be characterized in terms of their mixing, oxygen and $CO_2$ mass transfer, and process control capabilities (response time and accuracy) for dissolved oxygen, pH, and temperature.

There are several commercially available SUBs, and although not a comprehensive list, the companies marketing these systems include Wave Biotech (http://www.wavebiotech.com, now part of GE Healthcare), Sartorius (Edgewood, NY, http://www.sartorius.com), HyClone (http://www.hyclone.com), Hynetics (http://www.hynetics.com), ATMI (http://www.atmi.com), Xcellerex (http://www.xcellerex.com), Applikon (http://www.applikon-bio.com), and CELL-ution (http://www.cellutionbiotech.com). The systems differ in their methods of mixing, impeller integration (if one is used), aeration strategy, and probe incorporation. Two of these SUBs, the Wave Bioreactor and the HyClone, are described next.

***7.5.3.2 Wave Bioreactor*** One of the first SUBs introduced for cell culture applications was by Wave Biotech and was coined the Wave Bioreactor [64]. The Wave Bioreactor consists of a disposable bag, which contains cells and media, placed on a heated rocker. Headspace aeration is used to inflate the "cellbag," and the rocking motion ensures mixing and mass transfer. Oxygen mass transfer coefficients up to $60\,hr^{-1}$ at 30 rpm have been reported [65]. Dissolved oxygen and pH are measured by optical and polarographic probes, respectively. Gas blending to maintain dissolved oxygen and pH are supplied using mass flow controllers and integrated controllers. The manufacturers of this system, GE Healthcare Bioscience Bioprocess Corp. (Somerset, NJ), claim suitability for a wide variety of cell types and have manufactured units with up to 500 l of working volume. However, challenges exist with certain cell types, such as cholesterol-dependent cell lines [66], and mechanical issues have been noted at larger scales when repeatedly stopping and starting a 500-l wave of liquid places a major strain on the motor and rocker assembly.

GE also produces a system capable of operation under perfusion conditions. This system involves the use of a cell bag specifically designed for perfusion applications that contains an internal polyethylene filter. The floating 7-μm pore size polyethylene filter is approximately $100\,cm^2$ and allows the continuous withdrawal of cell-free spent medium from the system, which is replaced with fresh medium to maintain culture volume. Filter fouling, a common problem in membrane-based cell retention devices, is reduced because of the wave motion of culture fluid across the filter. Perfusion control based on constant culture weight relies on a load cell integrated into the Wave Bioreactor platform.

There are relatively few publications related to the Wave Bioreactor. Although publications exist on batch cultivation of mammalian, insect, yeast, and plant cells within the Wave Bioreactor [67–69], little literature exists regarding the use of this system for more intensive processes, such as fed-batch or perfusion operation. Pierce and Shabram [70] used Wave Bioreactors coupled to an external hollow fiber 0.2-μm pore size microfiltration cartridge for perfusion culture of a CHO-based cell line for

production of fusion protein at scales up to 500l. Very recently, Tang et al. [71] reported the use of a 1-l perfusion Wave Bioreactor for the production of IgG2 mAb using hybridoma cells. A number of similar wavelike systems, operating on similar principles, have been recently introduced, including those from Applikon, Sartorius, and CELL-ution. Although both systems from Applikon and Sartorious are equipped with invasive fluorescence-based optical dissolved oxygen and pH sensors (an improvement over the current sensors used in the Wave Bioreactor), the system from CELL-ution relies on more desirable noninvasive sensor technology. In addition, according to the manufacturer of this system, it is possible to obtain very high oxygen mass transfer rates because of the combined axial and vertical movement of the rocker.

***7.5.3.3 HyClone Single-Use Bioreactor*** The HyClone SUB, sold through Thermo Fisher Scientific, Inc., is intended as a retrofit product to replace the stainless steel bioreactor vessel in existing bioreactor skids rather than as a complete turn-key bioreactor system with integrated control systems. Alternatively, the HyClone SUB can be integrated using a number of customizable control systems such as Delta V. The current line of products consists of customizable 50-, 250-, and 100-l maximum working volume units. The disposable bags are mounted inside a permanent stainless steel jacket and secured by clips. Heat transfer is provided by an electrically heated jacket or customizable heat exchangers. The bags, made of class VI USP materials, are fitted with presterilized air inlet and vent filters (with optional heaters to prevent condensation) and ports for media/base addition and sampling. A single 45-degree pitched blade impeller made of molded polyethylene is linked to the seal–bearing assembly by C-flex tubing, which forms the contact material to the shaft. Aeration can be provided either via overlay or sparge. In the latter case, gas is provided by an integrated gas-permeable disc at the bottom of the bag. Overpressurization of the bag can be prevented using disposable pressure transducers linked to a control system. The sealing–bearing assembly links with the motor and allows impeller rotation while maintaining bag integrity. The bags can be used with presterilized conventional 12-mm sensors for pH and dissolved oxygen, which can be inserted in a sterile manner into the bag using novel quick connects. These sensors must also be connected to a control system for data gathering and process control.

We have used the HyClone system to successful scale from 2 to 200 l using CHO cells producing an IgG1 recombinant protein. The HyClone SUB was controlled by a DeltaV-based BioNet system (Broadley James, Irvine, CA). Additionally, a number of customized modifications were made to the standard bag to address shortcomings in the original design, such as extension of the vent filter tube and pressure transducers to reduce foam-outs and avoid bag rupture. Cell culture maintenance, expansion, and process conditions were similar to those described in Section 7.5.2. The initial cell culture process conditions for the production stage were 37°C with a pH of 7 and shifted lower to 33°C and a pH of 6.8 on day 2. The 2-l bioreactor (Applikon, Foster City, CA) was operated at process conditions similar to those of the HyClone system. Samples were taken for cell growth, viability, osmolality, metabolites, titer, and product quality attributes (data not shown) analysis. Excellent scale equivalency was obtained.

### 7.5.4 Future Prospects

All approved mAbs are produced in mammalian cells such as CHO, mouse myeloma (NSO), baby hamster kidney (BHK), human embryonic kidney (HEK-293), and PER.C6 (Section 7.2). Secreted antibodies from these cell lines are glycoslylated in a similar, but not identical, way to their human counterparts. The glycan structures in mAbs can influence their interaction with immune natural killer cells that bind to the constant region (Fc) and destroy antibody targeted cells. This process known as antibody-dependent cell toxicity (ADCC) depends on a specific *N*-glycosylation site at Asn 297 in the Fc domain of the heavy chain of IgG1s [72]. In addition to glycoform structure, the lack of fucose or presence of *N*-acetylglucosamine can impact ADCC potency. Although many mAbs do not exhibit ADCC activity, glycosylation is difficult to control precisely in mammalian cells and may be dependent on a variety of factors such as clonal variations, media, and culture conditions.

Recently, several alternative expression systems such as *Pichiapastoris* and *Escherichia coli* have emerged as promising hosts for mAb secretion when it may be possible engineer specific antibodies [73]. Researchers at Glycofi (a subsidiary of Merck & Co. Inc., Whitehouse Station, NJ) have successfully demonstrated the feasibility of glycol-engineered *P. pastoris* cell lines to produce mAbs with highly specific glycoforms [74]. Several different glycoforms of commercially available rituximab (manufactured as Rituxan by Genentech Inc., CA) were generated, and binding to Fcγ receptors and ADCC activity were measured. This study demonstrated a 10-fold increase in binding affinity as well as enhanced ADCC activity with the glycan-engineered proteins compared with Rituxan. Thus, controlling the glycan composition and structure of IgGs appears to be a promising method for improving the efficacy of therapeutic mAbs when ADCC activity is deemed important. Coupled with the use of well-established *P. pastoris* as a platform, including high cell density cultures, scalability, cost effectiveness, and existing large-scale fermentation capacity, can allow for the high-fidelity production of human glycosylated therapeutic proteins.

*E. coli* has been most commonly used for the production of antibody fragments such as Fabs when Fc-mediated effector functions are not required or deleterious [73]. Simmons et al. [75] demonstrated that efficient secretion of heavy and light chains in a favorable ratio resulted in the high-level expression and assembly of full-length IgGs in the *E. coli* periplasm. The technology described offers a rapid and potentially inexpensive method for the production of full-length aglycosylated therapeutic antibodies when ADCC function is not required. Mazor et al. [76] also showed that it was possible to obtain full-length antibodies from combinatorial libraries expressed in *E. coli*. The full-length secreted heavy and light chains assembled into aglycosylated IgGs which were captured by a Fc-binding protein located on the inner membrane. Flow cytometry was used after permeabilization of the membrane and attachment of the antibody to a fluorescent antigen. The use of cell-free protein synthesis for recombinant protein production is also emerging as an important technology. There are also reports of using *Aspergillus niger* for the production of mAbs or antibody fragments [77]. These authors used *N*-terminal fusion to glucoamylase for both heavy and light chains to express a

full length IgG. Goerke and Swartz [78] recently demonstrated the utility of this technology using *E. coli* cell extracts to produce a number of proteins, antibody fragments and vaccine fusion proteins, with correct folding and presence of disulfide bonds.

The past two decades have seen significant advances in cell culture technology that have increased the expression of recombinant proteins from 100 mg/l to several g/l concentrations. These advances have resulted from intensive research in cell line engineering (vector improvements, high-throughput screening methods for high-producer clones, and introduction of antiapoptotic genes in cell lines to prolong culture viability), media development, feeding strategies, cell metabolism, and better process understanding and their impact on product quality and scale-up studies. Further increases in titers will result from a combination of the above. In particular, improvements in media, including the development of high-yielding chemically defined media with replacement of hydrosylates with specific peptides, is an important area for further work [79,80].

Current fed-batch technology typically does not yield bioreactor cultures in excess of 10% packed cell volume. In contrast in microbial and fungal fed-batch cultures, the packed cell volume can be as high as 30 to 40% of the total bioreactor volume and therefore presents an opportunity to target further increases in cell concentration. Application of robust online sensors (e.g., capacitance sensors and NIR) for cell stoichiometric–based online feeding strategies with lower byproduct formation can be used achieving higher cell density cultures. However, prolongation of culture viability is also required. Better understanding of cell biology, including the application of gene expression analysis using CHO chips and genomic-scale models, will also be key [81,82]. The factors contributing to optimal producer cell lines are often complex and not simply due to one gene, which necessitates an understanding of system-wide properties for better engineering of cell lines [83]. Advances in genomics, proteomics, and metabolomics will fuel these advances; current application of these technologies is still in its infancy. A complete genomic scale-model requires the full annotated gene sequences of CHO cells, and this is currently unavailable. As a result, gene expression study by cross-species hybridization using human or mouse DNA microarrays is an option [84], although it is not very reliable. Researchers at the University of Minnesota BTI, National University Singapore, in collaboration with a consortium have made significant progress in CHO transcript sequencing by developing a CHO cDNA array and mouse oligonucleotide array for examining gene expression during apoptosis in fed-batch cultures. Information from such studies can be used for targeting apoptosis signaling genes to prolong culture viability. The most common approach for engineering antiapoptotic genes reported in the literature involves the overexpression of Bcl-2 or Bcl-xL to prevent mitochondrial release of cytochrome C into the cytoplasm [85]. Integration of small interfering RNAs (siRNA) into CHO cells to modulate gene expression is also a promising technology that can be used to tailor cell line properties. Kim and Lee [86] used siRNA to downregulate lactate dehydrogenase for reduced lactic acid formation in CHO cells.

Another area of research has focused on the use of inducible expression systems in mammalian cells. This strategy has the advantage of decoupling cell growth from product formation. Growth arrest in NS0 cultures via the inducible expression of p21 was reported to result in a significant increase in cell-specific productivity [87]. Current practice for the selection of high-producing clones involves a time-consuming and laborious screening strategy because the random integration of the transgene of interest gives rise to very different expression levels because of the variation in the chromosomal surroundings or because of unpredictable cross-interactions of multiple gene copies [88]. Several strategies have been proposed to overcome the negative position effect of random integration. One of the most promising methods is the use of heterologous recombination systems for site-specific gene integration such as the Flp/FRT system from *Saccharomyces cerevisiae* and the Cre-lox system from the bacteriophage P1 in eukaryotic cells. Coroadinha et al. [89] used the Flp/FRT system to target gene insertion into a high-expression chromosomal locus and succeeded in establishing a viral packaging 293 cell line for consistent high-titer virus production. Fukushige and Sauer [90] demonstrated the use of a lox recombination vector to obtain stable transformants with predictable gene expression profiles. The positive selection vector system was designed to directly select Cre-mediated DNA integration at a lox target (using an inactive lox-neo fusion gene) previously placed into the genome of cultured cells. This technique should allow the rapid and efficient exchange of a single copy of the transgene of interest with no change in expression levels.

A prerequisite to the application of this promising technology is the identification of integration sites that lead to high expression. However, it should be possible to investigate the position effects on gene expression using the site-specific DNA integration Flp/FRT and Cre-lox methods. For example, transfection of DHFR-deficient CHO cells with a loxP-GFP fusion and DHFR genes and high-throughput screening via FACS was used to identify transcription active sites [91]. Simultaneous transfection with two expression constructs and selection with two different selective markers as a means to generate highly productive stable CHO cell lines also appears promising (Section 7.2.2). Overall, these methods should result in greater efficiencies in cell line screening for high producers and a further reduction in development timelines.

Integration of high-throughput technologies, online monitoring and control capability, and automation will allow researchers to broaden the experimental design space as well as lower the cost of process development. From an operational perspective, disposable technologies should continue to see widespread adoption because of the advantages discussed earlier (operational simplicity, reduced turnaround time, facility change-over time, lower capital costs, and lower resource requirement).

## ACKNOWLEDGMENTS

The authors are grateful for the valuable contributions and technical discussions by Domingos Ng, Darren Brown, Cary Opel, Aaron Chen, Gayle Derfus, Dan Abramzon, Terrence Allotta, Terry Hudson, Brad Snedecor, David Chang, John Joly, and Dana Andersen.

## REFERENCES


1. Li F, Hashimura Y, Pendleton R, Harms J, Collins E, Lee B. A systematic approach for scale-down model development and characterization of commercial cell culture processes. Biotechnol Prog 2006;22:696–703.
2. Wurm FM. Production of recombinant protein therapeutics in cultivated mammalian cells. Nat Biotechnol 2004;22(11):1393–1398.
3. Andersen DC, Krummen L. Recombinant protein expression for therapeutic applications. Curr Opin Biotech 2002;13:117–123.
4. Keen MJ, Steward TW. Adaptation of cholesterol-requiring NS0 mouse myeloma cells to high density growth in a fully defined protein-free and cholesterol-free culture medium. Cytotechnology 1995;17:203–211.
5. Gorfien S, Paul B, Walowitz J, Keem R, Biddle W, Jayme D. Growth of NS0 cells in protein-free, chemically defined medium. Biotechnol Prog 2000;16(5):682–687.
6. Seth G, Ozturk M, Hu WS. Reverting cholesterol auxotroph of NS0 cells by altering epigenetic gene silencing. Biotechnol Bioeng 2006;93(4):820–827.
7. Taymar E, Hartman NS, Genereux K, Barritt DS, He Y, Burky JE, et al. Derivation and characterization of cholesterol-independent non-GS NS0 cell lines for production of recombinant antibodies. Biotechnol Bioeng 2007;96(2):294–306.
8. Birch JR, Racher JR. Antibody production. Adv Drug Deliv Rev 2006;58:671–685.
9. Li F, Kao E, Ryll T. A generic high cell mass fed-batch process for NS0 cells. Cell Culture Engineering IX, poster presentation; Cancun, Mexico; 2004.
10. Burky JE, Wesson MC, Young A, Farnsworth S, Dionne B, Zhu Y, et al. Protein-free fed-batch culture of non-GS NS0 cell lines for production of recombinant antibodies. Biotechnol Bioeng 2006;96(2):281–293.
11. Bebbington CR, Renner G, Thomson S, King D, Abrams D, et al. High level expression of recombinant antibody from myeloma cells using glutamine synthetase as an amplifiable selectable marker. Biotechnology 1992;10:169–175.
12. Chung CH, Mirakhur B, Chan E, Le QT, Berlin J. Cetuximab-induced anaphylaxis and IgE specific for galactose-α-1,3-galactose. N Eng J Med 2008;358(11):1109–1117.
13. Jones DH, Kroos N, Anema R, Van Montfort B, Vooys A, Van der Kraats S, et al. High-level expression of recombonant IgG in the human cell line PER.C6. Biotechnol Prog 2003;19(1):163–168.
14. Jayapal KP, Wlaschin KF, Hu W-S, Yap MGS. Recombinant protein therapeutics from CHO cells-20years and counting. Chem Eng Prog 2007;103:40–47.
15. Boshart M, Weber F, Jahn G, Dorsch-Hasler K, Fleckenstein B, Schaffher W. A very strong enhancer is located upstream of an immediate early gene of human cytomegalovirus. Cell 1985;41:521.
16. Deer JR, Allison DS. High-level expression of proteins in mammalian cells using transcription regulatory sequencings from the Chinese hamster EF-1a gene. Biotechnol Prog 2004;20:880–889.
17. Kozak K. Point mutations define a sequence flanking the AUG initiator codon that modules translation by eucarbotic ribosomes. Cell 1986;44:283.

18. Shen AY, Van de Goor J, Zheng L, Reyes AE, Krummen LA. Recombinant DNA technology and cell line development. Cell Culture Technology for Pharmaceutical and Cell-Based Therapies; Taylor and Francis Group, 6000 Broken Sound Parkway NW, Boca Raton, FL, 2006, p. 15–40.
19. Davis J, Jiang LY, Pan L-Z, LaBarre MU, Andersen D, Reff M. Expression of GnTIII in recombinant anti-CD20 CHO production cell line: expression of antibodies with altered blycomes leads to an increase in ADCC through higher affinity for FcyRIII. Biotechnol Bioeng 2001;74:288.
20. Saba JA, Kunkel JP, Jan DCH, Ens WE, Standing KG, Butler M, et al. A study of immunoglobulin G glycosylation in monoclonal and polyclonal species by electrospray and matrix-assisted laser desorption/ionization mass spectrometry. Anal Biochem 2002;305:16–31.
21. Kanda Y, Imai-Nishiya H, Kuni-Kamochi R, Katsuhiro M, Inoue M, Kitajima-Miyama K, et al. Establishment of a GDP-mannose 4,6-dehydrase (GMD) knockout host cell line: a new strategy for generating completely non-fucosylated recombinant therapeutics. J Biotechnol 2007;130:300–310.
22. Mori K, Kamochi RK, Ohnuki NY. Engineering Chinese hamster ovary cells to maximize effector function of produced antibodies using FUT8 siRNA. Biotechnol Bioeng 2004;88:901–908.
23. Sauerwald TM, Figueroa BJ, Hardwick J., Oyler GA, Betenbaugh MJ. Combining caspase and mitochondrial dysfunction inhibitors of apoptosis to limit cell death in mammalian cell cultures. Biotechnol Bioeng 2006;94(2):362–372.
24. Chen K, Liu Q, Xie L-Z, Sharp PA, Wang DI. Engineering of a mammalian cell line for reduction of lactate romation and high monoclonal antibody production. Biotechnol Bioeng, 2001;72:55–61.
25. Jeong D-W, Cho IT, Kim, TS, Bae GW, Kim I-H, Kim IY, Effects of lactate dehydrogenase suppression and glycerol-3-phosphate dehydrogenase overexpression on cellular metabolism. Mol Cell Biochem 2006;284:1–8.
26. Kim SH, Lee GM. Down-regulation of lactate dehydrogenase-A by siRNAs for reduced lactic acid formation of Chinese hamster ovary cells producing thrombopoietin. Appl Microbiol Biotechnol 2007;74:152–159.
27. International Conference on Harmonization of Technical Requirements for the Registration of Pharmaceuticals for Human Use. ICH Q5D, Derivation and Characterization of Cell Substrates Used for Production of Biotechnological/Biological Products; 1998.
28. Food and Drug Administration. Points to Consider in the Characterization of Cell Lines Used to Produce Biologicals. Washington, DC: Food and Drug Administration; 1993.
29. International Conference on Harmonization of Technical Requirements for the Registration of Pharmaceuticals for Human Use. ICH Q5B, Quality of Biotechnotogical Products: Analysis of the Expression Construct in Cells Used for Production of r-DNA Derived Protein Products. 1996.
30. Nims RW, Herbstritt CJ. Cell line authentication using isoenzyme analysis. BioPharm Int 2005;18:76–82.
31. Shepherd AJ, Wilson NJ, Smith KT. Characterisation of endogenous retrovirus in rodent cell lines used for production of biologicals. Biologicals 2003;31:251–260.
32. International Conference on Harmonization of Technical Requirements for the Registration of Pharmaceuticals for Human Use. ICH Q5A, Guidance for Industry: Viral Safety

Evaluation of Biotechnology Products Derived from Cell Lines of Human and Animal Origin; 1998.

33. Brorson K, De Wit C, Hamilton E, Mustafa M, Swann PG, Kiss R, et al. Impact of cell culture process changes on endogenous retrovirus expression. Biotechnol Bioeng, 2002;80(3):257–267.
34. Fletcher T. Designing culture media for recombinant protein production: a rational approach. BioProcess Int 2005;3(1):30–36.
35. deZengotita VM, Miller WM, Aunins J, Zhou WC. Phosphate feeding improves high-cell-concentration for monoclonal antibody production. Biotechnol Bioeng 2000;69 (5):566–576.
36. Robinson DK, Seamans TC, Gould SL, DiStefano DJ, Chan CP, Lee DK, et al. Optimization of a fed-batch process for production of a recombinant antibody. Ann NY Acad Sci 1994;745:285–296.
37. Xie LZ, Wang DIC. High cell density and high monoclonal antibody production through medium design and rational control in a bioreactor. Biotechnol Bioeng 1996;51(6):725–729.
38. Moore A, Mercer J, Dutina G, Donahue CJ, Bauer KD, Mather JP, et al. Effects of temperature shift on cell cycle, apoptosis and nucleotide pools in CHO cell batch cultures. Cytotechnology 1997;23:47–54.
39. Miller WM, Wilke CR, Blanch HW. Effect of dissolved oxygen on hybridoma growth and metabolism in continuous culture. J Cell Physiol 1987;132:524–530.
40. Ozturk SS, Palsson BO. Effects of dissolved oxygen on hybridoma cell growth, metabolism and antibody production kinetics in continuous cultures. Biotechnol Prog 1990;6:437–446.
41. Kunkel JP, Jan DC, Jamieson JC, Butler M. Dissolved oxygen concentration in serum-free continuous culture affects N-linked glycosylation of a monoclonal antibody. J Biotechnol 1998;62:55–71.
42. Garnier A, Voyer R, Tom R, Perret S, Jardin B, Kamen A. Dissolved carbon dioxide accumulation in a large scale high cell density production of TGFβ receptor with baculovirus infected Sf-9 cells. Cytotechnology 1996;22:53–63.
43. Gray DR, Cgen S, Howarth W, Inlow D, Maiorella L. $CO_2$ in large-scale and high density CHO cell perfusion culture. Cytotechnology 1996;22:65–78.
44. Kimura R, Miller WM. Glycosylation of CHO-derived recombinant tPA produced under elevated pCO2. Biotechnol Prog 1997;73:311–317.
45. Wayte J, Boraston R, Bland H, Varley J, Brown M. pH: effects on growth productivity of cell lines producing monoclonal antibodies: control in large scale fermenters. Cytotechnology 1997;22:87–94.
46. Schmid G, Blanch HW, Wilke CR. Hybridoma growth, metabolism and product formation in HEPES-buffered media: II: effect of pH. Biotechnol Lett 1990;12: 633–638.
47. deZengotita VM, Schmelzer AE, Miller WM. Chracterization of hybridoma cell responses to elevated pCO2 and osmolality: intracellular pH, cell size, apoptosis and metabolism. Biotechnol Bioeng 2002;77:369–380.
48. Seely JE, Seely R. A rational, step-wise approach to process characterization. BioPharm Int 2003;(August):24–34.

49. Rathore AS, Wang A, Menon M, Riske F, Campbell J, Goodrich E, et al. Optimization, scale-up and validation issues in filtration of biopharmaceuticals: Parts I and II. BioPharm Int 2004;(August/September):42–50.
50. Rathore, A, et al, Process validation how much to do and when to do it. BioPharm Int 2002;(October):18–28.
51. Food and Drug Administration. FDA Guidance for Industry: Q5E Comparability of Biotechnological/Biological Products Subject to Changes in Their Manufacturing Process; 2005. Available from: http://www.fda.gov/cder/guidance/index.htm.
52. Konstantinov KB, Pambuyan R, Matanguihan R, Yoshida T, Purusich CM, Hu WS. On-line monitoring of hybridoma cell growth using a laser turbidity sensor. Biotechnol Bioeng 1992;40:1337–1342.
53. Zeiser A, Bedard C, Voyer R, Jardin B, Tom R, Kamen A. On-line monitoring of the progress of infection in Sf-9 insect cell cultures using relative permittivity measurements. Biotechnol Bioeng 1999;63:122–126.
54. Chung H, Arnold MA, Rhiel M, Murhammer DW. Simultaneous measurements of glucose, glutamine, ammonia, lactate and glutamate in aqueous solutions using near-infrared spectroscopy. Appl Spectrosc 1996;50:270–276.
55. Riley MR, Okeson CD, Frazier BL. Rapid calibration development for near-infrared spectroscopic monitoring of mammalian cell cultivations. Biotechnol Prog 1999;15(6): 1133–1141.
56. Riley MR, Crider HM, Nite ME, Garcia RA, Woo J, Wegge RM. Simultaneous measurements of 19 components in animal cell culture media by near infrared spectroscopy. Biotechnol Prog 2001;17(2):376–378.
57. Yano T, Harata M. Prediction of the concentration of several constituents in a mouse-mouse hybridoma culture by near infrared spectroscopy. J Ferment Bioeng 1994;77: 659–662.
58. Harms P, Kostov Y, French JA, Soliman M, Anjanappa M, Ram A, et al. Design and performance of a 24-station high throughput microbioreactor. Biotechnol Bioeng 2006; 93:6–13.
59. Kostov Y, Harms P, Randers-Eichhorn L, Rao G. Low-cost microbioreactor for high-throughput bioprocessing. Biotechnol Bioeng 2001;72(3):346–352.
60. Kensy F, John GT, Hofmann B, Büchs J. Characterization of operation conditions and online monitoring of physiological culture parameters in shaken 24-well microtiter plates. Bioprocess Biosyst Eng 2005;75:75–81.
61. Micheletti M, Barrett T, Doig SD, Baganz F, Levy MS, Woodley JM, et al. Fluid mixing in shaken bioreactors: implications for scale-up predictions from microliter-scale microbial and mammalian cell cultures. Chem Eng Sci 2006;61:2939–2949.
62. De Jesus MJ, Girard M, Bourgeois M, Baumgartner G, Jacko B, Amstutz H, et al. TubeSpin satellites: a fast track approach for process development with animal cells using shaking technology. Biochem Eng J 2004;17:217–223.
63. Amanullah A, Otero JM, Mikola MR, Hsu A, Zhang J, Aunins J, et al. Evaluation of Simcell technology for cell culture process development. Biochemical Engineering XV, Quebec City, Canada; July 15–19, 2007.
64. Singh V. Disposable bioreactor for cell culture using wave-induced agitation. Cytotechnology 1999;30:149–158.

65. Mikola M, Seto J, Amanullah A. Evaluation of a novel Wave Bioreactor cellbag for aerobic yeast cultivation. Bioprocess Biosys Eng 2007;30(4):231–241.
66. Kadarusman J, Bhatia R, McLaughlin J, Lin WR. Growing cholesterol-dependent NS0 myeloma cell line in the wave bioreactor system: overcoming cholesterol-polymer interaction by using pretreated polymer or inert fluorinated ethylene propylene. Biotechnol Prog 2005;21(4):1341–1346.
67. Namdev PK, Lio P. Assessing a disposable bioreactor for attachment-dependent cell cultures. BioPharm Int 2000;13:44–50.
68. Palazón J, Mallol A, Eibl R, Lettenbauer C, Cusidó RM, Piñol MT. Growth and ginsenoside production in hairy root cultures of Panax ginseng using a novel bioreactor. Planta Med 2003;69:344–349.
69. Weber W, Weber E, Geisse S, Memmert K. Optimisation of protein expression and establishment of the Wave Bioreactor for Baculovirus/insect cell culture. Cytotechnology 2002;38:77–85.
70. Pierce L, Shabram PW. Scalability of a disposable bioreactor from 25L-500L in perfusion mode with a CHO-based cell line: a tech review. BioProc J 2004;4:51–56.
71. Tang YJ, Ohashi R, Hamel JF. Perfusion culture of hybridoma cells for hyperproduction of IgG(2a) monoclonal antibody in a wave bioreactor-perfusion culture system. Biotechnol Prog 2007;23(1):255–264.
72. Rothman RJ, Perussia B, Herlyn D, Warren L. Antibody-dependent cytotoxicity mediated by natural killer cells is enhanced by castanospermine-induced alterations of IgG glycosylation. Mol Immunol 1989;26:1113–1123.
73. Andersen DC, Reilly DE. Production technologies for monoclonal antibodies and their fragments. Curr Opin Biotechnol 2004;15:456–462.
74. Li H, Sethuraman N, Stadheim TA, Zha D, Prinz B, Ballew N, et al. Optimization of humanized igGs in glycoengineered Pichia pastoris. Nat Biotechnol 2006;24(2):210–215.
75. Simmons LC, Reilly D, Klimowski L, Raju TS, Meng G, Sims P, et al. Expression of full-length immunoglobulins in Escherichia coli: rapid and efficient production of a glycosylated antibodies. J Immunol Methods 2002;263:133–147.
76. Mazor Y, Van Blarcom V, Mabry R, Iverson BL, Georgiou G. Isolation of engineered, full length antibodies from libraries expressed in *Escherichia coli*. Nat Biotechnol 2007; 25(5):563–565.
77. Ward M, Lin C, Victoria DC, Fox BP, Fox JA, Wong DL, et al. Characterization of humanized antibodies secreted by *Aspergillus niger*. Appl Environ Microbiol 2004;70(5): 2567–2576.
78. Goerke AR, Swartz JR. Development of cell-free protein synthesis platforms for disulfide bonded proteins. Biotechnol Bioeng 2008;99(2):351–367.
79. Franek F, Eckschlager T, Katinger H. Enhancement of monoclonal antibody production by lysine-containing peptides. Biotechnol Prog 2003;19:169–174.
80. Spens E, Häggström L. Defined protein and animal component-free NS0 fed-batch culture. Biotechnol Bioeng 2007;98(6):1183–1194.
81. Seth G, Philp RJ, Lau A, Jiun KY, Yap M, Hu WS. Molecular portrait of high productivity in recombinant NS0 cells. Biotechnol Bioeng 2006;97(4):933–951.
82. Wlaschin KF, Seth G, Hu WS. Toward genomic cell culture engineering. Cytotechnology 2006;50:121–140.

83. Griffin TJ, Seth G, Xie H, Banhakavi S, Hu WS. Advancing mammalian cell culture engineering using genome-scale technologies. Trends Biotechnol 2007;25(9):401–408.
84. Wong DCF, Wong KTM, Lee YY, Morin PN, Heng CK, Yap MGS. Transcriptional profiling of apoptotic pathways in batch and fed-batch CHO cell cultures. Biotechnol Bioeng 2006;94(2):373–382.
85. Arden N, Betenebaugh MJ. Life and death in mammalian cell culture: strategies for apoptosis inhibition. Trends Biotechnol 2004;22:74–180.
86. Kim SH, Lee GM. Down-regulation of lacatate dehydrogenase-A by siRNAs for reduced lactic acid formation of Chinese hamster ovary cells producing thrombopoietin. Appl Microbiol Biotechnol 2007;74(1):152–159.
87. Bi J-X, Shuttleworth J, Al-Rubeai M. Uncoupling of cell growth and proliferation results in enhancement of productivity in p21cip1-arrested CHO cells. Biotechnol Bioeng 2004;85:741–749.
88. Wolf AP. When more is less. Nat Genet 1998;18:5–6.
89. Coroadinho AS, Schucht R, Gama-Norton L, Wirth D, Hauser H, Carrondo MJ. The use of recombinase mediated cassette exchange in retroviral vector producer cell lines: predictability and efficiency by transgene exchange. J Biotechnol 2006;124(2):457–468.
90. Fukushige S, Sauer B, Genomic targeting with a positive-selection lox integration vector allows highly reproducible gene expression in mammalian cells. Proc Natl Acad. Sci USA 1992;89:7905–7909.
91. Kito M, Itami S, Fukano Y, Yamana K, Shibui T, Construction of engineered CHO strains for high-level production of recombinant proteins. Appl Microbiol Biotechnol 2002; 60:442–448.

# Protein and Peptide Purification

BECKLEY K. NFOR, ESTEBAN J. FREYDELL, and MARCEL OTTENS

## 8.1 INTRODUCTION

The increasing demand for therapeutic proteins and peptides has led to a higher market growth potential for these products [1]. Although advances in upstream technology have enabled improved production titers of recombinant proteins and peptides from a few milligrams per liter to several grams per liter [2], their total manufacturing cost remains extremely high, with a significant proportion (50–80%) coming from downstream processing [3–9].

Downstream processing (DSP) deals with the isolation and purification of biological products from complex cell cultures [3,4,8]. The need to purify these products timely, efficiently, economically and to sufficient purity and quantity applies to every large-scale purification process. The extent of purification depends on the end use of the protein or peptide product. Hence, from the DSP point of view, it is important to know the application of the product of interest. Commercially available proteins and peptides can be classified into food and food additives or nutraceuticals, pharmaceuticals or therapeutics, industrial enzymes, and diagnostics. Some examples of proteins in these different categories have been summarized by Ghosh [10]. Therapeutic proteins and peptides destined for administration by injection generally have the most stringent quality and safety requirements compared with the other categories of products [11,12]. Table 8.1 shows the typical specifications for therapeutic protein products such as monoclonal antibodies (mAbs) [11].

Irrespective of the product application, the general structure of DSP is the same for cell-derived proteins and peptides. The objective of this chapter is to present a comprehensive overview of the key aspects of DSP development of proteins and peptides. The chapter opens with a general description of the DSP steps and highlights their objectives and the basic bioseparation techniques used (Section 8.2). After introducing the general aspects of DSP, recent developments in inclusion body (IB) refolding are presented (Sections 8.3). There are several challenges facing

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*,
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

**TABLE 8.1 Typical Specifications for Therapeutic Proteins [11]**

| Product Specification | Typical Range |
|---|---|
| Purity | >99% |
| Host cell protein | <10 ng/dose |
| Nucleic acid | <10 pg/dose |
| Endotoxins | <1 EU/kg body wt |
| Virus | >12 log reduction |
| Cells | Not detectable |
| Leachables | <10 ppm |

EU: endotoxin units.

DSP development, as outlined in Section 8.4. Section 8.5 reviews strategies for protein and peptide purification process development and how each tries to address or overcome the DSP challenges. Finally, expert views on future directions of purification process development efforts are presented (Section 8.6). An extensive list of carefully selected literature on various aspects of DSP development is also included for further reading.

## 8.2 OVERVIEW OF DOWNSTREAM PROCESSING

Recombinant DNA technology has permitted the large-scale production of almost any biological product via the fermentation route using any suitable expression system (e.g., bacterial cells, yeasts, mammalian cells, insect cells). Depending on the expression system used or the localization of the product (intracellular or extracellular), the material leaving the bioreactor or the fermentation broth may contain the target protein in a complex mixture of cells, cell debris, host cell proteins, nucleic acids, endotoxins, viruses, and other contaminants (e.g., unconsumed media components), which must be cleared or reduced to acceptable levels during DSP (see Table 8.1). Otherwise the product is unlikely to make it through clinical trials and into the market. DSP of cell-derived proteins and peptides is broadly divided into different stages: primary recovery and clarification, IB processing, protease inactivation or nucleic acid removal, concentration or preconditioning, purification, virus reduction, and product formulation or storage, each of which may involve one or more process steps with distinct objectives as depicted in Figure 8.1 and outlined Table 8.2. Separation units and reagents commonly used for different DSP steps are also presented in Table 8.2. The unit operations exploit physicochemical or thermodynamic property differences between the target molecule and impurities to achieve separation [13–15]. This section provides a comprehensive overview of DSP, highlighting the separation techniques commonly used at the industrial scale. More in-depth description of the

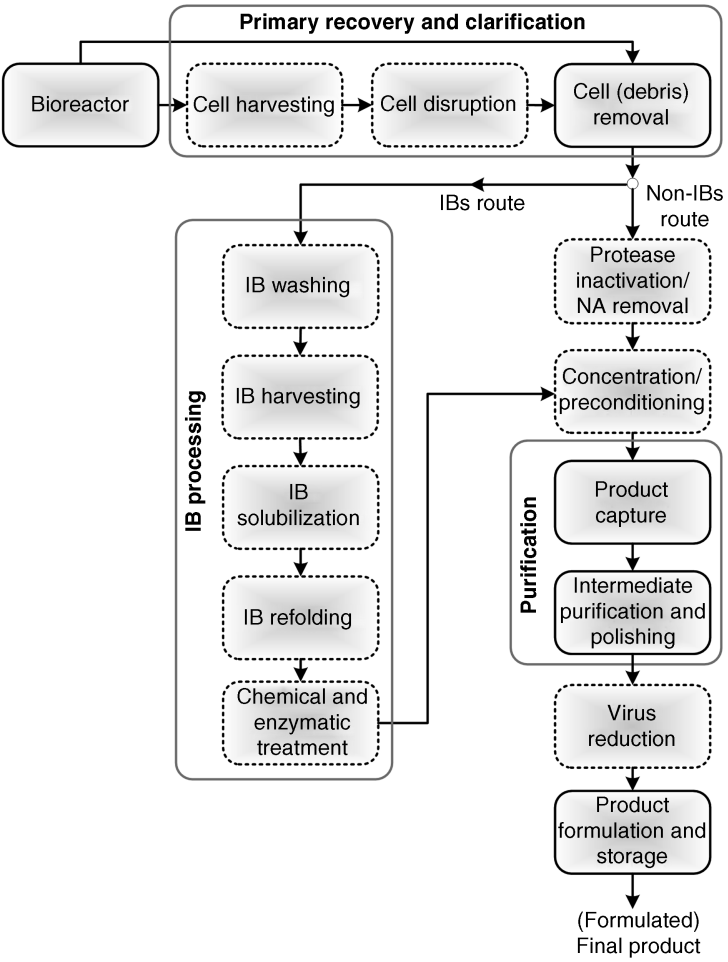


**FIGURE 8.1** Generalized block scheme of downstream processing. *Dotted boxes* show the steps that may or may not be present depending on expression system used and the target product localization and chemistry. IB: inclusion body; NA: nucleic acid.

downstream unit operations and their implication for DSP can be found in several dedicated monographs [16–20] and reviews [13,21–23].

### 8.2.1 Primary Recovery or Clarification

The main objectives of the primary recovery or clarification steps are to release the target product from the cells (for intracellular products only) and to remove cells and cell debris from the process stream containing the target product. Intracellular product release is accomplished by harvesting and disrupting the cells before cell and

**TABLE 8.2 Principal Steps and Unit Operations Used in Downstream Processing of Proteins and Peptides**

| Steps | | Objectives | Suitable Techniques or Reagents |
|---|---|---|---|
| Primary recovery or clarification | Cell harvesting | Harvest cells from fermentation broth (intracellular products only) | • Centrifugation<br>• Microfiltration (rotary vacuum filtration, cross-flow filtration, depth filtration) |
| | Cell disruption | Release of intracellular products and byproducts | • Homogenization<br>• Bead milling<br>• Osmotic shock<br>• Chemical or enzymatic permeabilization or lysis |
| | Cell (debris) removal | Separate cells or cell debris from intra- or extracellular fluid containing the product and byproducts | • Centrifugation<br>• Microfiltration (rotary vacuum filtration, cross-flow filtration, depth filtration) |
| Protease inactivation and nucleic acid removal | Protease inactivation | Inactivate or inhibit protease action through binding to the protease | • Serine protease inhibitors (phenylmethylsulfonyl fluoride PMSF, 4-aminophenylmethylsulfonyl fluoride–APMSF, benzamidine HCl, leupeptin)<br>• Cysteine protease inhibitors (leupeptin, chymostatin, antipain HCl)<br>• Aspartic protease inhibitor (Pepstatin)<br>• Metalloprotease inhibitors (EDTA, EGTA) |
| | Nucleic acid removal | Reduce viscosity by removing or reducing nucleic acids | • Precipitation using polyethyleneimine, $MgCl_2$, or streptomycin sulfate<br>• Nuclease enzyme addition (e.g., DNase, RNase) |
| Concentration and preconditioning | Concentration | Concentrate the target protein or peptide | • UF<br>• (Salt-induced) precipitation<br>• Crystallization<br>• (Aqueous two-phase) extraction |
| | Preconditioning | Buffer exchange or salt removal | • DF |

| | | | |
|---|---|---|---|
| IB processing | IB washing | Remove contaminants that adsorb onto IBs | • Detergents: Triton X-100 (0.1–4% v/v) + high salt + EDTA, sodium deoxycholate (2% w/v), sarkosyl<br>• Chaotrophs: 0.5–1 M guanidine HCl or urea |
| | IB harvesting | Separate IBs from wash solution | • Centrifugation |
| | IB solubilization | Denature and solubilize IBs | • Chaotrophs: 8-M urea, 4–6 M guanidine HCl<br>• Detergents: SDS (0.5–2%)<br>• Reducing agents: Dithiothreitol or β-mercaptoethanol (5–100 mM) |
| | IB refolding | Remove denaturants and thiol reagents and allow protein to refold | • Dilution (batch, fed batch or continuous)<br>• Membranes (e.g., dialysis, diafiltration)<br>• Fixed-bed chromatography (ion-exchange, hydrophobic interaction, size-exclusion, or gel filtration)<br>• SMB chromatography<br>• Reverse micelles<br>• Refolding aids: Chaperones, foldases |
| | Chemical and enzymatic treatment | Cut off tag or fusion (poly)peptide to restore protein or enzyme activity or for posttranslational modification (e.g., glycosylation) | • Chemical: citraconic anhydrate (for reversible blocking of lysine side chains)<br>• Enzymes: trypsin, carboxypeptidase B, protease V8 |
| Purification | Product capture | Capture and purify the target product and reduce the process volume as much as possible | • Fixed-bed chromatography (affinity, mixed mode, ion-exchange, hydrophobic interaction)<br>• Membrane chromatography<br>• EBA chromatography<br>• SMB chromatography<br>• ATPS |

(*continued*)

**TABLE 8.2** (*Continued*)

| Steps | | Objectives | Suitable Techniques or Reagents |
|---|---|---|---|
| | | | • UF or DF (mainly for interstage conditioning) |
| | Intermediate purification | Further removal of host cell proteins, nucleic acids, and endotoxins while achieving some virus clearance | • Fixed-bed chromatography (affinity, mixed mode, ion-exchange, hydrophobic interaction, size-exclusion, or gel filtration) |
| | Polishing | Remove product-related impurities such as product aggregates, fragments, and variants | • Membrane chromatography<br>• SMB chromatography<br>• UF or DF (mainly for interstage conditioning) |
| Virus reduction | | Inactivate and clear or reduce viruses to acceptable levels | • Virus inactivation (low pH or 60% EtOH)<br>• Virus retention (nano)filters |
| Product formulation and storage | Product formulation | Formulate the target protein or peptide into the correct dosage form based on the route of administration | • Crystallization<br>• Drying (spray, freeze, tray, fluid bed, drum) |
| | Product storage | Use storage conditions or product forms that ensure long-term stability and storage of the active protein or peptide product | • Granulation, tableting, packaging<br>• Pegylation |

APMSF: 4-aminophenylmethylsulfonyl fluoride; ATPS: aqueous two-phase extraction; DF: diafiltration; EDTA: ethylenediaminetetraacetic acid; EGTA: ethylene glycol tetraacetic acid; EtOH: ethanol; IB: inclusion body; SDS: sodium dodecyl sulfate; SMB: simulated moving bed; UF: ultrafiltration.

cell debris removal as depicted in Figure 8.1. Techniques commonly used for cell disruption can be classified as mechanical (bead milling, high-pressure homogenization) and nonmechanical (e.g., enzymatic or chemical lysis or permeabilization). The nonmechanical methods are more suitable for periplasmic products [24]. Enzymatic or chemical lysis may be used in combination with a mechanical disruption method to improve protein release. When mechanical disruption is used, there is an inevitable release of other cellular components such as proteases, nucleic acids, endotoxins, and other protein impurities, thereby complicating further DSP. Unit operations commonly used for harvesting or removal of cells (cell debris) include microfiltration (cross-flow microfiltration, depth filtration, rotary vacuum filtration) and centrifugation. Early efforts on primary recovery were typically centered on microfiltration because of its low capital cost advantage over centrifugation [25]. However, the operational cost of the microfiltration unit may be higher because of membrane fouling that can occur, especially at high cell densities, which increases the cartridge replacement frequency. With the advent of low shear centrifuges, centrifugation is now commonly used industrially for high-density cell cultures with disk-stack centrifuge being the most widely used [26]. On the other hand, for removal of small debris ($<0.5\,\mu m$), centrifugation is not adequate because of the large g-force or energy input required [18]. For primary recovery of recombinant therapeutic proteins such as mAbs from mammalian cell cultures, a hybrid disk-stack centrifugation–(depth) filtration system has become the industry standard [3,4,27].

### 8.2.2 Protease Inactivation and Nucleic Acid Removal

The presence of proteases in the product stream may lead to proteolytic degradation of the target protein. Hence, proteases must be inactivated or removed from the product stream as quickly as possible. Protease removal is achieved in the chromatographic purification steps further downstream. However, the addition of protease inhibitors to cell lysis buffers may be required to stop the protease activity before their removal. Protease inhibitors are biological or chemical compounds that function by reversibly or irreversibly binding to the protease [28]. Because proteases generally belong to one of four evolutionarily distinct enzyme families (i.e., serine proteases, cysteine proteases, carboxyl and aspartyl proteases, and metalloproteases) with different functional groups and cleavage sites on a protein [19], a "cocktail" of different inhibitors may be required to protect proteins from proteolysis (see Table 8.2) [28,29]. Therefore, when possible, the addition of protease inhibitors should be avoided because these chemicals would have to be removed, further complicating the DSP.

High nucleic acid content increases viscosity. As a general rule, if the viscosity of the clarified liquid stream is higher than 4 cP, then viscosity reduction is required [11]. Nucleic acids and endotoxins are negatively charged macromolecules; hence, their removal can be achieved by using cationic precipitants such as polyethyleneimine (a long chain cationic polymer with a molecular mass of $\sim$24 kDa) [11,20]. Alternatively, a nuclease enzyme (e.g., DNase or RNase) may be added to hydrolyze

the nucleic acids [29]. Other techniques such as anion exchange chromatography (AEX) may also be used. However, because of the relatively high liquid viscosity and the sensitivity of chromatographic techniques to fouling, they are seldom used at this stage. Further removal of nucleic acids or endotoxins (or both) is achieved in the purification steps.

### 8.2.3 Concentration and Preconditioning

The clarified liquid stream leaving the primary recovery section is virtually free of particulate matter and may be obtained as a dilute or concentrated solution of the target protein in a complex mixture of impurities. If present as a dilute protein solution, this stream may need to be concentrated before the purification steps. Otherwise, it is directly sent to the purification steps after preconditioning. The goal of the concentration step is to reduce the process volume and hence the sizes of purification units further downstream. The only real limits to the desirable level of protein concentration are the liquid viscosity and the tendency for proteins to aggregate or precipitate. Ultrafiltration and precipitation are two widely used concentration techniques [19,20]. Aqueous two-phase extraction (ATPS) may also be used, although this is not a routine industrial practice. Direct product crystallization from the clarified liquid stream also sounds attractive but is not suitable at this stage because of the complexity of the process stream. When product concentration is achieved by (salt-induced) precipitation, the precipitate needs to be resolubilized in a suitable buffer.

The preconditioning step is required to adjust the process stream conditions (e.g., buffer or salt type, pH, conductivity) through buffer exchange or salt removal or addition to levels required by the first purification step further downstream. Techniques used include dialysis and diafiltration, with diafiltration being more suitable at the industrial scale because of the large process volumes [19]. Although buffer exchange can also be accomplished by size-exclusion chromatography (SEC), this technique is seldom used at this stage because of its low loading capacity and significant dilution of the effluent stream.

### 8.2.4 Purification

After primary recovery, concentration, or preconditioning, the clarified liquid stream to the purification section often contains the target protein in a complex mixture of host cell proteins, nucleic acids, endotoxins, viruses, and other host cell impurities. When confronted with such a complex mixture, many different purification routes can be followed to achieve the desired product quality, often involving one or more chromatographic steps because of their noninvasive nature and unparalleled selectivity [29]. One generally seeks to achieve the most efficient purification with the minimum number of steps [29–31]. Protein purification generally consists of at least two main steps, (1) product capture and (2) intermediate purification or polishing, both with defined objectives (see Table 8.2). Fixed-bed chromatography is mostly applied for the different purification steps as outlined in Table 8.2, but alternative unit operations that

may offer better throughput are also gaining importance [32,33]. These include monoliths, ATPS, three-phase partitioning, membrane filtration, crystallization, membrane chromatography, and simulated moving bed (SMB) chromatography. Direct product capture from fermentation broth containing cells or cell debris is also possible. This can be achieved by expanded-bed adsorption (EBA) chromatography based on ion exchange or another interaction mechanism [20,34].

After the chromatographic methods to be applied for a given separation problem have been defined, the selection of a suitable chromatographic resin and operating conditions for each method becomes crucial because it may dictate the success or failure of the purification step for which it is intended and hence of the overall separation. The general requirements of chromatographic resins used for protein purification include high capacity, selectivity, recovery yield, and low backpressure at relatively high linear velocities. For product capture, high-capacity, linear velocity (speed), and selectivity are the most important requirements. Resins used for product polishing must exhibit high selectivity for separation of product-related impurities such as product aggregates, fragments, and variants.

### 8.2.5 Virus Reduction

For biological products derived from animal or plant cells, viral clearance or reduction to acceptable levels may be required (depending on the product application). For therapeutic proteins, it is required by the Food and Drug Administration to have at least two orthogonal steps for virus inactivation and removal. Hence, although some level of viral clearance is achieved in the chromatographic purification steps, one or more additional viral clearance steps may be required for these products. In mAb purification platforms, virus inactivation at low pH and virus removal by filtration are typically used [4,27]. Viral filters can be classified on the basis of pore sizes into retroviral (<50 nm) and parvoviral (<20 nm) grade filters [35]. The filter size and position in the purification train can vary from one product or process to another. Their small pore sizes (especially parvoviral filters) make the filters susceptible to fouling and clogging. Thus, pressurized tanks are commonly used to drive fluid through the filters in preference to the use of pumps [27].

### 8.2.6 Product Formulation and Storage

The main objective of product formulation and storage is to get the protein or peptide product in a form or condition desired for final consumption or for long-term product stability and storage. Common methods used for product formulation and storage are summarized in Table 8.2. Protein formulation and related aspects are comprehensively reviewed elsewhere [36,37]. In general, crystallization would be an ideal storage method because the crystallized product is likely to be very pure and stable. However, finding suitable crystallization conditions for most (therapeutic) proteins is still a challenge [38]. Another alternative is complete dehydration of protein product, preferably using a gentle technique such as freeze drying (or lyophilization) as opposed to direct product drying or freezing, which may result in product

inactivation [19]. For product storage in solution or liquid form, a buffer exchange step (e.g., SEC, dialysis or diafiltration, ultrafiltration) may be necessary to get the desired storage condition. However, because storage conditions may be highly protein product dependent, prior screening for suitable storage conditions is generally required [39]. For some proteins, additives (e.g., salt) may be required for the final application or to improve the stability and shelf life of the product [19]. Finally, protein engineering techniques such as PEGylation may be used to improve protein stability and half-life, enabling extended storage or activity of the PEGylated protein products [40,41].

### 8.2.7 Inclusion Body Processing

Prokaryotes such as the bacteria *Escherichia coli* offer several advantages as expression systems, comprising well-understood molecular genetics, that is, the genome can be readily modified, relatively inexpensive culturing procedures, and high fermentation yields [42–44]. One important limitation of these expression systems is that they lack an excretory mechanism; hence, overexpression of certain gene sequences often results in the intracellular accumulation of the product in an inactive, insoluble aggregate known as IBs. If the target protein is present as IBs, these need to be solubilized and refolded after cell debris removal (Fig. 8.1) [45–48]. Chemical and enzymatic treatment may be required after protein refolding to remove tags or fusion polypeptides, which may otherwise reduce the protein activity, or for posttranslational modifications (e.g., glycosylation) of the recombinant protein product. The IB washing step is used to remove cellular contaminants that adsorb onto the hydrophobic IB surfaces. Typical IB washing reagents are shown in Table 8.2. The nonionic detergent Triton X-100 is the most commonly used [49]. Harvesting of IBs (size $\sim$1 mm) from the wash solution can be accomplished by centrifugation. Thereafter, solubilization is typically attained using a solution comprising chaotropes (e.g., urea, guanidine hydrochloride), reducing agents (e.g., β-mercaptoethanol, dithiothreitol), and alkaline pH. This solution breaks the intermolecular interactions in the aggregated protein, freeing the monomer in a denatured and reduced (D&R) form. Ideal solubilization conditions should result in maximum protein solubility, decreasing the fraction of soluble aggregates formed and increasing the fraction of soluble monomer [45,50–52]. Decreasing the concentration of chaotropes and reducing agents in the protein solution initiates refolding, meaning the soluble protein refolds and forms its disulfide bridges. The change in chemical composition is a buffer exchange that can be realized by performing direct dilution (e.g., continuous stirred tank dilution, batch dilution, fed-batch dilution), membranes (e.g., dialysis, diafiltration), or by liquid chromatography (chromatographic refolding). These methods belong to the class of chemical-induced refolding, discussed in more detail later. A separate class is the so-called pressure-induced refolding, not discussed in this chapter. This method employs hydrostatic pressures in the range of 150 to 200 MPa to disrupt aggregated protein, favoring the formation of the native product [53–59].

## 8.3 PROTEIN REFOLDING METHODS

### 8.3.1 Dilution Refolding

Dilution refolding is the most widely used method to refold proteins at an industrial scale because it is considered the simplest way to initiate protein refolding. This approach consists of diluting the denatured, often D&R, protein IB solution with refolding buffer at dilution factors of 50% or higher. This buffer generally does not contain denaturant agents or reducing agents, but it would most likely contain refolding additives (e.g., aggregation suppressors, redox couples) [60]. Protein refolding is initiated as a result of the instant decrease in the concentration of the denaturant and reducing agent. The refolding yield, defined as the amount of native protein per amount of D&R protein loaded, is negatively correlated to the concentration of D&R protein in the refolding reactor [61]. This implies that refolding at high protein concentrations generally leads to a low refolding yield. This correlation is explained by the kinetics of protein refolding and protein aggregation and their dependence on the concentration of D&R protein [61]. Protein refolding is a monomolecular reaction; thus, its reaction rate is linearly dependent on the D&R protein concentration. Conversely, aggregation is a multimolecular reaction; hence, its reaction rate is a power function of the D&R protein concentration [62]. To minimize the product mass loss (improving refolding yield) toward the competing aggregation reaction, batch dilution refolding is usually operated at a total protein concentration between 0.01 and 0.2 mg/ml [61]. This concentration range can be increased by adding additives that mitigate protein aggregation [60,63,64]. In terms of reactor configuration, dilution refolding can be done in a batch [61], fed-batch [65,66], or continuous stirred tank reactor (CSTR) [67]. Based on reactor design theory, a batch reactor is not suitable to maximize yield (conversion) when the product is formed via a first-order reaction competing with a higher order reaction (folding vs. aggregation). Fed-batch refolding (also known as pulse renaturation) addresses this limitation by controlling the rate of D&R protein addition, minimizing the aggregation reaction rate [64,66,68]. Finally, to improve throughput, dilution refolding can be perform on a CSTR [68], or a CSTR in conjunction with fed-batch addition of the D&R protein solution [66]. Fed-batch refolding and CSTR refolding may be operated such that the native product can be concentrated during the refolding operation, which is not feasible using batch dilution refolding alone.

### 8.3.2 Chromatographic Protein Refolding

Refolding using liquid chromatography delivers a relatively high degree of process intensification compared to refolding by dilution. Process intensification encompasses the options of performing (a) protein refolding, (b) change in chemical composition using gradients, (c) purification on the product, and (d) its concentration in one unit operation. Refolding by chromatography has effectively been shown for hydrophobic interaction chromatography [69–71], size-exclusion chromatography

[72,73], immobilized metal affinity chromatography (IMAC) [74,75], and ion-exchange chromatography [76–78]. This section addresses IEX and SEC because most of the recently published literature promotes these two approaches as the ones that could find application at a preparative scale.

***8.3.2.1 Refolding by Ion-Exchange Chromatography*** Ion-exchange refolding (IExR) was first described by Creighton [78]. He showed that a protein denatured using urea could be bound reversibly to an ion-exchange media. Moreover, he showed that urea could be eliminated using a gradient without eluting the bound protein, resulting in refolding and formation of disulfide bridges. Finally, using a salt gradient, the refolded protein could be eluted, obtaining it in a relatively pure fraction, free of urea. Throughout the years, the scientific community gained interest in IExR plausibly because (a) IEX is broadly employed as a purification step in the biopharmaceutical industry, and thus it could be readily introduced in existing processes; (b) compared to SEC, ion exchange has a high loading capacity and the product can be concentrated during elution using a salt gradient; (c) compared to HIC, the protein denatured using urea can be bound to the IEX media using a much lower salt concentration, not the case for HIC refolding; and (d) protein–protein interactions could plausibly be reduced by the spatial separation obtained upon the binding of the protein to the IEX media, presumably leading to less protein aggregation, increasing the refolding yield. Creighton's method has been throughout the years modified to include a simultaneous pH, urea, and salt gradient. This approach has been tested with lysozyme on a cation exchanger [79,80] and iron superoxide dismutase (Fe-SOD) on an anion exchanger [81]. Recently, IExR employing an incubation period at no flow has effectively been used to refold bovine serum albumin [76] and α-fetoprotein [82]. The latter application demonstrated that IExR can be used to refold complex proteins, as in the latter two examples proteins having 16 and 17 disulfide bridges were successfully refolded. Examples of proteins that have been refolded using IEX are presented in Table 8.3. Table 8.3 also presents the operational factors reported and their corresponding values. Refolding yield in this unit operation is influenced by the protein load, the speed of denaturant removal, the adsorbent's backbone type, and the resident time [76,80,83].

***8.3.2.2 Refolding by Size-Exclusion Chromatography*** Refolding by size exclusion was first reported in the work of Werner et al. and afterwards by other groups [73,88,89]. Essentially, the principle of SEC refolding can be described as follows. A pulse containing a denaturant, a reducing agent, and the protein is loaded to the SEC column, already equilibrated using the refolding buffer. Due to the differences in average distribution coefficients, the denaturant, reducing agent, and the protein concentration fronts separate as they progress down the column. This separation results in a decrease in the concentration around the protein of the denaturant and reducing agent, resulting in refolding and the formation of disulfide bridges. Various examples of SEC refolded proteins are presented in Table 8.4.

**TABLE 8.3 Examples of Proteins Refolded by Ion-Exchange Chromatography**

| Protein | Resin | Bonds ($n$) | RSH (mM) | RSSR (mM) | pH | $C_{Urea0}$ (M) | $C_{UreaE}$ (M) | $\xi^*$ | Load† | Rec (%)‡ | Reference |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Cytochrome C | CM cellulose | 1 | 0 | 0 | 8.7 | 8 | 0 | −0.10 | 3.7 | 100 | Creighton [78] |
| Hen ovalbumin | DEAE–cellulose | 1 | 0 | 0 | 8.7 | 8 | 0 | −0.11 | 3.7 | 50 | Creighton [78] |
| BPTI | CM cellulose | 3 | 0.1 | 1 | 8.7 | 8 | 0 | −0.05 | 3.7 | 90 | Creighton [78] |
| α-Latalbumin | DEAE–cellulose | 4 | 0 | 0 | 8.7 | 8 | 0 | −0.11 | 3.7 | 10 | Creighton [78] |
| SOD–*Escherichia coli* | Q-Sepharose | NR | 0 | 0 | 7.4 | 6 | 1 | −0.20 | 0.8 | NR | Li and Su [81] |
| Lysozyme | SP-Sepharose | 4 | 3 | 0.3 | 8.7 | 6 | 1 | −0.4 | 1.6 | NR | Li et al. [79] |
| Lysozyme | SP-Sepharose | 4 | 3 | 0.3 | 8 | 6 | 1 | −0.21 | 1.14 | NR | Li and Su [80] |
| BSA | Q-Sepharose | 17 | 2.2 | 1.1 | 8.5 | 8 | 0.079 | ∞ | 2-20 | 67-25 | Langenhof et al. [76] |
| TGase | SP-Sepharose | NR | 0 | 0 | 6 | 6 | 0.5 | −0.15 | 0.4 | 49.7 | Liu et al. [84] |
| rhG-CSF | Q-Sepharose-FF | 2 | 0 | 0 | 8.0 | 3 | 1.2 | −0.08 | NR | 43 | Wang et al. [85] |
| LK-68 | Q-Sepharose | 9 | 0 | 0 | 9.0 | 8 | 0 | −0.22 | 2.5 | NR | Choi et al. [86] |
| NS3 | SP-Sepharose | NR | 10 | 0 | 4 | 8 | 1 | −0.08 | 0.059 | NR | Li et al. [87] |
| | DEAE | NR | 10 | 0 | 9 | 8 | 1 | −0.09 | 0.059 | NR | Li et al. [87] |

*$\xi$: Urea slope (mol urea/l/min).
†Load: milligrams of denatured protein per milliliter of packed bed.
‡Recovery: mass recovered (protein eluted/protein injected).
∞: step change; BSA: bovine serum albumin; NR: not reported.
BPTI: bovine pancreatic trypsin inhibitor; CM: carboxymethyl; DEAE: diethylaminoethyl; FF: fast flow; SP: sulfopropyl; LK-68: lipoprotein kringle; NS3: non-structural protein 3; rhG-CSF: recombinant human granulocyte colony-stimulating factor; Q: quaternary ammonium; SOD: super oxide dismutase; TGase: transglutaminase; Rec: recovery; RSH: reducing agent; RSSR: oxidized agent.

**TABLE 8.4 Examples of Proteins Refolded Using Size-Exclusion Chromatography**

| Protein | Gel Type | $V_{Inj}$ (%BV) | D&R (mg/ml) | $\dot{Q}$ (ml/min) | Reference |
|---|---|---|---|---|---|
| rETS-1 | Superdex 75 | 4.2–8.5 | 1–10 | 0.2 | Werner et al. [88] |
| RNase A | Sephacryl S100 | 0.18–0.37 | 1–10 | 0.5 | Werner et al. [88] |
| Lysozyme | Sephacryl S100 | 0.32 | 9.6 | 2.5 | Batas and Chaudhuri [73] |
| Lysozyme | Mixed gel | 0.67 | NR | 2.5 | Batas and Chaudhuri [73] |
| Carbonic anhydrase | Sephacryl S100 | 0.67 | 11 | 1.1 | Batas and Chaudhuri [73] |
| Lysozyme | Sephacryl S100 | 4.24 | 10 | 0.3 | Batas et al. [101] |
| u-PA | Sephacryl S300 | 0.20–3.31 | 2–16 | 0.5–4.5 | Fahey et al. [102] |
| Lysozyme | Sephacryl S100 | 0.32 | 40 | 2.5 | and Chaudhuri [103] |
| | Superdex 75 | 0.70 | 40 | 2 | and Chaudhuri [103] |
| u-PA | Sephacryl S300 | 0.42 | 8 | 0.5 | Fahey et al. [92] |
| Bovine carbonic anhydrase B | Superdex 75 | 4.24 | 5.2 | 0.4 | Gu et al. [93] |
| Lysozyme | Superdex 75 | 0.08 | 5 | 1 | Wang et al. [104] |
| Lysozyme | Sephacryl S100 | 0.25 | 11.8 | 0.8 | Park et al. [99] |
| β-Lactamase | Sephacryl S300 | 0.47–0.94 | 8 | 0.83–4 | Harrowing and Chaudhuri [105] |
| Lysozyme | Sephacryl S200 | 0.62 | 8.3 | 2 | Lanckriet and Middelberg [98] |
| α-Lactalbumin | Superdex 75 PrepGrade | 1.50 | 10 | 1.01 | Schlegl et al. [97] |

D&R: denatured and reduced protein; rETS: recombinant human ETS-1; u-PA: urokinase plasminogen activator; RNase A: bovine ribonuclease A.

Refolding using SEC has been done using protein feed concentrations in the range of 0.5 mg/ml [90] all the way to 80 mg/ml [73]. Furthermore, refolding using SEC has been studied using various model proteins and several types of size-exclusion media [91]. In the studies performed, the objective shared was the optimization of the refolding yield. Several approaches have been tested comprising the effect of the type of gel filtration media [92,93], the impact of using a urea gradient and/or a pH gradient [94,95], and the effect of using a chaperone system [96]. Though these strategies could potentially increase the refolding yield, they should not have a substantial impact on specific eluent consumption, volumetric productivity, and final product concentration, as these process performance indicators depend primarily on the mode of operation (i.e., batch vs. continuous).

Batch SEC refolding has the same known limitations of the standard batch SEC purification step. These limitations comprise the use of long columns and a significant product dilution, due to the large column volume to loading volume ratio. The aforementioned limitations result in the need for a large inventory of SEC media, long process times, and a large consumption of eluent(s). These limitations can be offset by changing the mode of operation from batch to continuous. Continuous size-exclusion refolding has been shown on preparative annular chromatography (P-CAC) for α-lactalbumin [97] and lysozyme [98] and simulated moving bed chromatography for lysozyme and an industrial fusion protein [72,99,100].

### 8.3.3 Refolding by Membranes

Membrane refolding can be accomplished using dialysis or diafiltration [101,106–108]. Dialysis is a relatively slow process because the mass transfer–driving force is diffusion. Conversely, diafiltration is a faster process that offers better controllability because the permeate, retentate, and refolding buffer flow rates can be accurately controlled. Furthermore, diafiltration can be operated in a continuous manner, which translates into relatively high volumetric productivities, and could also be used to concentrate the final (native) product. Refolding yields in these systems can be adversely affected by (1) the unspecific binding of the protein to the filtration membrane; (2) losses of protein through the membrane; and (3) concentration polarization, leading to membrane fouling and protein aggregation [49].

## 8.4 CHALLENGES IN PURIFICATION

The purification of proteins and peptides from crude biological cell cultures poses several challenges. To begin with, the product may be present in soluble form or as insoluble IBs, which have to be resolubilized and renatured or refolded. Although IB harvesting from the clarified cell extract may be easy to accomplish, IB refolding poses a significant challenge, with low overall refolding yields and significant product losses [45]. In the soluble form, the products are often present

at very low concentrations in a complex mixture of other impurities, which are ill defined with respect to the composition and the physicochemical (e.g., molecular mass or size, charge, hydrophobicity) and thermodynamic (e.g., solubility) properties of the mixture components, as well as the flow characteristics (e.g., viscosity). This fact renders purification process development and especially the removal of product-related impurities such as other protein isoforms, fragments, and aggregates with similar physicochemical properties to the target molecule particularly challenging [30]. Furthermore, it is important to distinguish between impurities that may be tolerated and the so-called "critical impurities" or "critical contaminants" that must be removed as soon as possible (e.g., proteases that may degrade the target protein). Moreover, most protein products are thermolabile and are sensitive to extreme pH and to chemical substances such as surfactants and organic solvents. On the process perspective, there may be several bioseparation unit operations to choose from at each processing step, each of which may have several process parameters and operating conditions to be optimized, giving rise to an optimization problem. This is further complicated by the large number of possible interconnections of the unit operations into candidate process flowsheets, leading to a combinatorial problem in the process synthesis, and the fact that many bioprocess unit operations operate in batch or semi-continuous mode and require regular cleaning. Finally, therapeutic proteins and peptides must not only meet stringent product quality and safety requirements by regulatory agencies (e.g., product purity and biological activity and the permitted levels of host cell proteins, nucleic acids, endotoxins, and viruses), but their time to market also needs to be as short as possible [4].

## 8.5 PURIFICATION STRATEGIES

Despite the technical challenges, the short time to market for cell culture–derived therapeutic products means that their purification process development must be performed timely, and the products must satisfy the strict product requirements by regulatory authorities. Strategies for protein and peptide purification process development can be broadly classified into the heuristics or expert knowledge–based method, high-throughput experimentation (HTE), platform strategies, algorithmic- or optimization-based methods, and hybrid methods involving a combination of strategies [1,30,109].

### 8.5.1 Heuristics or Expert Knowledge–Based Method

Before the advent of HTE, purification process development was mostly performed traditionally (i.e., using a set of heuristics or rules of thumb based on past experience, supported by some laboratory-scale experiments). Because of the time and material cost of performing experiments at the laboratory scale, heuristics are first used to synthesize one or a few candidate purification processes for a target product [18,110,111] followed by laboratory-scale testing to find a working process and

suitable operating conditions, mainly on the basis of trial and error, and finally by the scale up of any working process as long as it satisfies the customer and regulatory needs. This purely traditional approach often results in overall suboptimal processes with inefficient feedstock and auxiliary materials utilization and high DSP costs. There have been attempts to capture expert knowledge on the computer in the form of expert systems [14,31,112–115] supported by physicochemical properties characterization of the target protein and major protein contaminants [116].

### 8.5.2 High-Throughput Experimentation

Nowadays, the use of heuristics for initial process screening has been largely replaced by high-throughput process development (HTPD, i.e., involving the use of HTE) [117]. Although the general protein purification guidelines are still very useful, chromatographic resin screening and the optimization of buffers and operating conditions are now largely performed by high-throughput screening (HTS) using 96-well plate format and robotic liquid handling technology [118,119]. HTS of chromatographic resins enables the exploration of a much wider space of process variables than possible with conventional column experiments because of miniaturization and parallelization of experiments, with significant savings in time and material costs [118–127]. Data obtained by HTS of chromatographic resins are very useful for fast resin and process evaluation during the early stages of process development [118,123–125,127–131].

The major biopharmaceutical companies have adopted a more structured HTPD approach involving the combined use of high-throughput resin screening followed by column scouting and optimization and scale up of the optimized process showing the best performance [132]. The main idea is to first reduce the large experimental space as cheaply and as fast as possible by means of HTS and then focus the column-scouting experiments on the most promising resin options only. With the availability of miniaturized column formats, it is now possible to perform scouting experiments also at high throughput [127,131].

A good overview of commercially available tools for HTPD can be found in Bhambure et al. [117]. The major challenge of using HTS is in the handling and analysis of the large amount of experimental data generated, for which better techniques and tools are needed. Combined with statistical design of experiments or evolutionary search algorithms, however, an optimal system can be reached in which the number of experiments performed in HTS mode is minimized while maximizing information gained from the data [125]. This approach even allows the optimization of unit operations when no physical model exists and process knowledge is minimal.

### 8.5.3 Platform Strategies

After an economically viable purification process has been developed for a given protein product by HTPD or otherwise, this might serve as a platform or template for other products within the same category, such that purification schemes for the

different proteins in that category show a high degree of commonality [4,27]. For example, the purification platforms for most mAbs contain at least two chromatographic steps, including a protein A affinity chromatography capture step, two or more chromatographic polishing steps, and at least two orthogonal virus removal steps to ensure product safety [3,4,27,133]. Platform approaches for mAb purification are mainly driven by the projected large increase in the number of antibody therapies and time-to-market considerations [4,8,27].

The advantages of using a platform approach include savings in time and effort and greater harmonization of practices and information across different functions and sites within a biotech organization [4,27]. Optimization mainly involves improvements of individual steps because the purification cascade is more or less fixed. In the case of the mAb purification platforms, for example, optimization efforts have been focused on improving the dynamic binding capacity of the protein A affinity step from the current range of 30 to 50 g/l to 80 to 100 g/l [3,4], as well as on designing new-affinity ligands [6]. There have also been attempts to replace or eliminate at least one of the conventional IEX steps in the mAb process. In this regard, the potential of membrane IEX as a replacement of the flow-through mode anion exchange chromatographic step for removal of low-level impurities such as host cell proteins and DNA has been demonstrated [134,135].

### 8.5.4 Algorithmic or Model-Based Methods

The algorithmic methods of process development involve the use of mathematical models and computer simulations to predict, optimize, and scale up a purification process [1,30,136], as well as for robustness studies to ensure consistent process performance [137–139]. In this way, the number of experiments is kept to a minimum. This has been made possible through the development of reliable state-of-the-art unit operation models and sophisticated computer programs to assist in DSP development [16,140,141]. Key aspects in the model-based approach include the development of suitable analytic procedures for model parameters acquisition and model validation [136].

Despite the potential benefits, model-aided process development based on sound thermodynamic property data is still lagging behind in the biopharmaceutical industry because of a variety of reasons: (1) the complex and ill-defined nature of typical bioprocess feed materials, (2) the lack of suitable analytical tools and methods to deal with these and the time and material costs that may be involved, and (3) other practical considerations such as the large number of variables and auxiliary materials (e.g., resins, buffers, salts) to be screened. This is changing with the advent of HTE, as discussed under the hybrid approach in Section 8.5.5.

### 8.5.5 Hybrid Methods

By definition, a hybrid approach combines two or more different strategies for process development by taking advantage of the strengths of each. Here, we focus on the model-based HTPD approach involving the integrated use of HTE and process

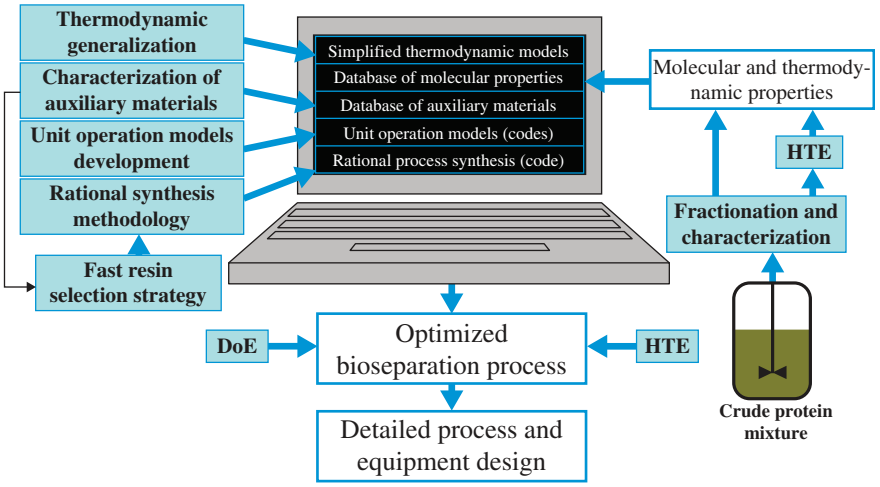


**FIGURE 8.2** Overview of model-based high-throughput protein purification process development approach by Nfor et al. [142]. DoE: design of experiment; HTE: high-throughput experimentation.

modeling for purification process development. In the model-based HTPD strategy, in addition to using HTE for screening auxiliary materials and operating conditions, the latter is also used for fast acquisition of thermodynamic properties data of proteins and proteins assemblies of common host organisms, bringing the possibility of model-based prediction and optimization of fairly complex systems to practical reality.

A typical example of how this concept of model-based HTPD could be applied in practice is provided by Ottens and coworkers [30,142–145], as depicted in Figure 8.2.

In their approach, HTE and modern proteomics tools such as mass spectrometry are used for the fractionation and characterization of crude protein mixtures to obtain the necessary physicochemical and thermodynamic properties of the target protein and the major protein contaminants (i.e., those with close selectivity to the target protein in each defined separation method) needed for process modeling [144,146]. Using the obtained parameters as inputs, a computer algorithm based on properly validated unit operations models is used to find an optimum purification cascade [145]. The selected purification cascade then undergoes further experimental verification, optimization, and robustness analysis using a combination of process modeling, HT and laboratory-scale experiments, and design of experiments tools to identify the critical process parameters and define their operating windows. This is followed by model-assisted, detailed design and economic evaluation of the best process. An important aspect of the modeling framework is that the thermodynamic models or adsorption isotherms used to capture protein–ligand interactions in process chromatography must be reliable. Attempts have been made to establish

generalized mechanistic thermodynamic frameworks that cut across multiple separation mechanisms and methods [13,147,148]. Rational model-based methodologies for fast and rational resin or process evaluation and selection have also been developed [143–145], as well as a multidimensional experimental scheme for fast acquisition of the relevant model parameters [146].

The main characteristics of the process development strategies discussed were recently summarized by Rathore and coworkers [117], and the strengths and weaknesses of each are outlined by Nfor et al. [30].

## 8.6 FUTURE DIRECTIONS

This chapter would be incomplete without probing into future bioseparation process development directions. The discussion here was consciously limited to process development for therapeutic proteins and peptides that show the greatest potential for growth and have the greatest regulatory concerns.

Although purification platforms are currently safely positioned within the biopharmaceutical industry because of the accrued benefits [27], an interesting debate has emerged over the capability of existing purification platforms and facilities to adequately respond to the DSP needs of the future, such as higher capacity and product quality; shorter time to market; and better process monitoring and control [3,5]. Moreover, it has been demonstrated that an inflexible generic platform with fully templated operating conditions is hardly possible because of dramatic differences in elution behavior because of property differences, as observed for mAbs [27]. Hence, some biopharmaceutical companies (e.g., Amgen) have adopted a flexible generic platform instead in which optimization of the individual process steps is still a must for each mAb [27].

Furthermore, as patent rights on protein therapeutics expire and other manufacturers step into the same or similar product market (the so-called biosimilars market), competition among biosimilars manufacturers will become rife, and so will the need to cut significantly on purification costs, although others have argued that stricter comparability testing for biosimilars may diminish their potential as market drivers [5].

In either case, process development tools and strategies that emphasize better process understanding, monitoring, and control through the quality by design and process analytical technology initiatives are likely to be more responsive to these future needs [1,117,149,150]. In our view, the hybrid method based on a combination of HTPD and process modeling is best suited to meet these challenges.

As demand and throughput requirements on therapeutic proteins and peptides continue to rise, we will probably also see a rising need for better scalable, non–packed-bed chromatographic techniques such as aqueous two-phase systems, selective precipitation with polyelectrolytes or polymers, high-resolution ultrafiltration, crystallization, monoliths, and membrane chromatography [32–34], although packed-bed chromatography will continue to be the workhorse for protein purification for decades to come because of its unparalleled selectivity. Innovations in

packed-bed chromatography will mainly involve the development of better quality adsorbents with better mass transfer characteristics, higher capacity, and less susceptibility to fouling and reuse. EBA chromatography for direct product capture from crude cell cultures will continue to make its (re)entry (e.g., DSM's RHOBUST technology [34]). SMB technology will also gain importance for large biopharmaceuticals as capacity and productivity needs increase. When product expression titers are much higher than impurity levels, simple but innovative solutions such as designing chromatographic ligands for capture operations in flow-through mode could be used instead (i.e., designed to bind impurities) in which case the binding capacity will no longer be considered a bottleneck for increasing the throughput of the capture step because the impurity levels are much lower. Protein engineering strategies for selective protein purification through the use of protein tags (e.g., PEGylation) or through the development of ligands for selective affinity purification of native target proteins will also gain ground [151].

There is increasing pressure within biopharmaceutical companies to minimize extensive investments in new facilities and optimize the usage of existing facilities because of a combination of factors such as the increasing number of drug candidates, the slowing rate of new drug approvals, and increased competition. Hence, effective utilization of manufacturing facilities will become a major driver for process innovation, and innovative technologies that can operate within existing facilities are likely to shape the market landscape [5]. For therapeutic products with small niche markets, for example, high product titers could bring a radical shift toward lower production volumes, which in turn will increase the need for facility flexibility and the use of disposable technologies [5].

With the ever-growing importance of HTE tools such as miniaturized column and well formats and microchips, there is already a growing need for analytical methods that are compatible with these miniaturized systems [131]. We shall also see important developments in microfluidics as the application areas of microchips in biopharmaceutical process development increase (e.g., second viral coefficient or solubility determination on chip for product stability or shelf-life studies [152], ATPS on chip [153], diffusivity measurement on chip [154]).

Only within the past decade or so, there has been significant progress in the way process development in the biopharmaceutical industry is performed, although there are still many challenges to overcome, as highlighted in this chapter. Effective communication among the regulatory bodies, industrial practitioners, and academic researchers is key in addressing the remaining DSP challenges and in bringing to realization, the industrial implementation of new (innovative) DSP technologies.

## REFERENCES

1. Pavlou AK, Reichert JM. Recombinant protein therapeutics—success rates, market trends and values to 2010. Nat Biotechnol 2004;22(12):1513–1519.
2. Wurm FM. Production of recombinant protein therapeutics in cultivated mammalian cells. Nat Biotech 2004;22(11):1393–1398.

3. Kelley B, Blank G, Lee A. Downstream processing of monoclonal antibodies: current practices and future opportunities. In: Gottschalk U, editor. Process Scale Purification of Antibodies. New Jersey: John Wiley & Sons; 2009, p. 1–23.
4. Low D, O'Leary R, Pujar NS. Future of antibody purification. J Chromatogr B 2007;848 (1):48–63.
5. Shukla AA, Thömmes J. Recent advances in large-scale production of monoclonal antibodies and related proteins. Trends Biotechnol 2010;28(5):253–261.
6. Lowe CR, Lowe AR, Gupta G. New developments in affinity chromatography with potential application in the production of biopharmaceuticals. J Biochem Biophys Methods 2001;49(1–3):561–574.
7. Farid SS. Process economics of industrial monoclonal antibody manufacture. J Chromatogr B 2007;848(1):8–18.
8. Sommerfeld S, Strube J. Challenges in biotechnology production—generic processes and process optimization for monoclonal antibodies. Chem Eng Process 2005;44 (10):1123–1137.
9. Steinmeyer DE, McCormick EL. The art of antibody process development. Drug Disov Today 2008;13(13–14):613–618.
10. Ghosh R. Protein bioseparation: an overview. In: Protein Bioseparation Using Ultrafiltration: Theory, Applications and New Developments. London: Imperial College Press; 2003.
11. Sofer G, Hagel L. Purification design, optimization, and scale-up. In: Sofer G, Hagel L, editors. Handbook of Process Chromatography: A Guide to Optimization, Scale-up, and Validation. London: Academic Press; 1997, p. 27–113.
12. Headon DR, Walsh G. The industrial production of enzymes. Biotechnol Adv 1994; 12(4):635–646.
13. Ahamed T, Ottens M, Nfor BK, van Dedem GWK, van der Wielen LAM. A generalized approach to thermodynamic properties of biomolecules for use in bioseparation process design. Fluid Phase Equilib 2006;241(1–2):268–282.
14. Leser EW, Asenjo JA. Rational design of purification processes for recombinant proteins. J Chromatogr A 1992;584(1):43–57.
15. Wai PPC, Bogle IDL, Bagherpour K, Gani R. Process synthesis and simulation strategies for integrated biochemical process design. Computers Chem Eng 199620(Suppl 1): S357–S362.
16. Harrison RG. Bioseparations Science and Engineering. New York: Oxford University Press; 2003.
17. Asenjo JA. Separation Processes in Biotechnology. New York: Marcel Dekker; 1990.
18. Asenjo JA, Patrick I. Large-scale protein purification. In: Harris ELV, Angal S, editors. Protein Purification Applications: A Practical Approach. Oxford: Oxford University Press; 1990, p. 1–28.
19. Scopes RK. Protein Purification: Principles and Practice. 3rd ed. New York: Springer-Verlag; 1994, p. 132.
20. Janson J-C, Ryden L. Protein Purification. Principles, High Resolution Methods, and Applications. 2nd ed. New York: John Wiley & Sons; 1998.
21. Hubbuch J, Kula M-R. Isolation and purification of biotechnological products. J Non-Equilib Thermodyn 2007;32:99–127.

22. Lightfoot EN, Moscariello JS. Bioseparations. Biotechnol Bioeng 2004;87(3):259–273.
23. Strube J, Grote F, Josch JP, Ditz R. Process development and design of downstream processes. Chemie Ingenieur Technik 2011;83(7):1044–1065.
24. Naglak TJ, Hettwer DJ, Wang HY. Chemical permeabilization of cells for intracellular product release. In: Separation Processes in Biotechnology. New York: Marcel Dekker; 1990.
25. van Reis R, Leonard LC, Hsu CC, Builder SE. Industrial scale harvest of proteins from mammalian cell culture by tangential flow filtration. Biotechnol Bioeng 1991;38 (4):413–422.
26. Kempken R, Preissmann A, Berthold W. Assessment of a disc stack centrifuge for use in mammalian cell separation. Biotechnol Bioeng 1995;46(2):132–138.
27. Shukla AA, Hubbard B, Tressel T, Guhan S, Low D. Downstream processing of monoclonal antibodies—application of platform approaches. J Chromatogr B 2007;848(1):28–39.
28. Thermo Scientific. Pierce Protein Research Products. Protease and phosphatase inhibitors. Available from: http://www.piercenet.com.
29. GE HealthCare. Protein Purification Handbook 2007, p. 18–1132–1129.
30. Nfor BK, Ahamed T, van Dedem GWKV, van der Wielen LAM, van de Sandt EJ, Eppink M, et al. Design strategies for integrated protein purification processes: challenges, progress and outlook. J Chem Technol Biotechnol 2008;83:124–132.
31. Asenjo JA, Andrews BA. Is there a rational method to purify proteins? From expert systems to proteomics. J Mol Recognit 2004;17(3):236–247.
32. Przybycien TM, Pujar NS, Steele LM. Alternative bioseparation operations: life beyond packed-bed chromatography. Curr Opin Biotechnol 2004;15(5):469–478.
33. Thommes J, Etzel M. Alternatives to chromatographic separations. Biotechnol Prog 2007;23(1):42–45.
34. Douwenga R. Rhobust® Technology. BioProcess Int 2011;9(7).
35. Ireland T, Siwak M, Bolton G, Lutz. Viral filtration of plasma-derived human IgG: a case study using Viresolve NFP. BioPharm Int 2004;11:38.
36. Kelner DN, Wang W, Wang DQ. Protein formulation. Biotechnol Bioproc 2003;27:549–604.
37. Sellers SP, Maa Y-F. Principles of biopharmaceutical protein formulation: an overview. Methods Mol Biol 2005;308:243–263.
38. Ahamed T, Esteban BNA, Ottens M, van Dedem GWK, van der Wielen LAM, Bisschops MAT, et al. Phase behavior of an intact monoclonal antibody. Biophys J 2007;93(2):610–619.
39. Capelle MAH, Gurny R, Arvinte T. High throughput screening of protein formulation stability: practical considerations. Eur J Pharm Biopharm 2007;65(2): 131–148.
40. Marshall SA, Lazar GA, Chirino AJ, Desjarlais JR. Rational design and engineering of therapeutic proteins. Drug Disov Today 2003;8(5).
41. Jevševar S, Kunstelj M, Porekar VG. PEGylation of therapeutic proteins. Biotechnol J 2010;5(1):113–128.
42. Schmidt FR. Recombinant expression systems in the pharmaceutical industry. Appl Microbiol Biotechnol 2004;65(4):363–372.

43. Terpe K. Overview of bacterial expression systems for heterologous protein production: from molecular and biochemical fundamentals to commercial systems. Appl Microbiol Biotechnol 2006;72(2):211–222.
44. Demain AL, Vaishnav P. Production of recombinant proteins by microbes and higher organisms. Biotechnol Adv 2009;27(3):297–306.
45. Freydell EJ, Ottens M, Eppink M, van Dedem G, van der Wielen L. Efficient solubilization of inclusion bodies. Biotechnol J 2007;2(6):678–684.
46. Choe W-S, Nian R, Lai W-B. Recent advances in biomolecular process intensification. Chem Eng Sci 2006;61(3):886–906.
47. Cabrita LD, Bottomley SP. Protein expression and refolding—a practical guide to getting the most out of inclusion bodies. Biotechnol Annu Rev 2004;10:31–50.
48. De Bernardez Clark E. Refolding of recombinant proteins. Curr Opin Biotechnol 1998; 9(2):157–163.
49. Clark EDB. Protein refolding for industrial processes. Curr Opin Biotechnol 2001;12:202–207.
50. Tsumoto K, Ejima D, Kumagai I, Arakawa T. Practical considerations in refolding proteins from inclusion bodies. Protein Expres Purif 2003;28:1–8.
51. Umetsu M, Tsumoto K, Nitta S, Adschiri T, Ejima D, Arakawa T, et al. Nondenaturing solubilization of [beta]2 microglobulin from inclusion bodies by L-arginine. Biochem Biophys Res Commun 2005;328(1):189–197.
52. Singh S, Panda A. Solubilization and refolding of bacterial inclusion body proteins. J Biosci Bioeng 2005;99(4):303–310.
53. Gorovits BM, Horowitz PM. High hydrostatic pressure can reverse aggregation of protein folding intermediates and facilitate acquisition of native structure. Biochemistry 1998;37(17):6132–6135.
54. John RJS, Carpenter JF, Randolph TW. High pressure fosters protein refolding from aggregates at high concentrations. Proc Natl Acad Sci USA 1999;96(23):13029–13033.
55. Kim Y, Randolph TW, Seefeldt MB, Carpenter JF. High-pressure studies on protein aggregates and amyloid fibrils. Methods Enzymol 2006;413(Part C):237–253.
56. Phelps DJ, Hesterberg LK. Protein disaggregation and refolding using high hydrostatic pressure. J Chem Technol Biotechnol 2007;82(7):610–613.
57. Randolph TW, Seefeldt M, Carpenter JF. High hydrostatic pressure as a tool to study protein aggregation and amyloidosis. Biochim Biophys Acta 2002;1595(1–2):224–234.
58. Ribo M, Font J, Benito A, Torrent J, Lange R, Vilanova M. Pressure as a tool to study protein-unfolding/refolding processes: the case of ribonuclease A. Biochim Biophys Acta 2006;1764(3):461–469.
59. St. John RJ, Carpenter JF, Randolph TW. High pressure fosters protein refolding from aggregates at high concentrations. Proc Natl Acad Sci USA 1999;96(23):13029–13033.
60. Yasuda M, Murakami Y, Sowa A, Ogino H, Ishikawa H. Effect of additives on refolding of a denatured protein. Biotechnol Prog 1998;14:601–606.
61. Hevehan DL, Clark EDB. Oxidative renaturation of lysozyme at high concentrations. Biotechnol Bioeng 1997;54(3):221–230.
62. Kiefhaber T, Rudolph R, Kohler H-H, Buchner J. Protein aggregation in vitro and in vivo: a quantitative model of the kinetic competition between folding and aggregation. Nat Biotechnol 1991;9(9):825–829.

63. Clark EDB, Schwarz E, Rudolph R. Inhibition of aggregation side reactions during in vitro protein folding. Methods Enzymol 1999;309:217–236.
64. Middelberg APJ. Preparative protein refolding. Trends Biotechnol 2002;20:437–443.
65. Winter J, Lilie H, Rudolph R. Renaturation of human proinsulin—a study on refolding and conversion to insulin. Anal Biochem 2002;310(2):148–155.
66. Katoh S, Katoh Y. Continuous refolding of lysozyme with fed-batch addition of denatured protein solution. Process Biochem 2000;35:1119–1124.
67. Schlegl R, Tscheliessnig A, Necina R, Wandl R, Jungbauer A. Refolding of proteins in a CSTR. Chem Eng Sci 2005;60(21):5770–5780.
68. Lilie H, Schwarz E, Rudolph R. Advances in refolding of proteins produced in *E. coli*. Curr Opin Biotechnol 1998;9:497–501.
69. Geng X, Chang X. High- performance hydrophobic interaction chromatography as a tool for protein refolding. J Chromatogr 1992;599:185–194.
70. Bai Q, Kong Y, Geng X-d. Studies on renaturation with simultaneous purification of recombinant human proinsulin from *E. coli* with high performance hydrophobic interaction chromatography. J Liq Chromatogr Relat Technol 2003;26(5):683–695.
71. Wang C, Geng X, Wang D, Tian B. Purification of recombinant bovine normal prion protein PrP(104-242) by HPHIC. J Chromatogr B 2004;806(2):185–190.
72. Freydell EJ, Bulsink Y, Van Hateren SH, van der Wielen LAM, Eppink M, Ottens M. Size-exclusion simulated moving bed chromatographic protein refolding. Chem Eng Sci 2010;65:4701–4713.
73. Batas B, Chaudhuri JB. Protein refolding at high concentration using size-exclusion chromatography. Biotechnol Bioeng 1996;50(1):16–23.
74. Rogl H, Kosemund K, Kühlbrandt W, Collinson I. Refolding of *Escherichia coli* produced membrane protein inclusion bodies immobilised by nickel chelating chromatography. FEBS Lett 1998;432(1–2):21–26.
75. Zahn R, von Schroetter C, Wüthrich K. Human prion proteins expressed in Escherichia coli and purified by high-affinity column refolding. FEBS Lett 1997;417 (3):400–404.
76. Langenhof M, Leong SSJ, Pattenden LK, Middelberg APJ. Controlled oxidative protein refolding using an ion-exchange column. J Chromatogr A 2005;1069:195–201.
77. Machold C, Schlegl R, Buchinger W, Jungbauer A. Matrix assisted refolding of proteins by ion exchange chromatography. J Biotechnol 2005;117(1):83–97.
78. Creighton TE. Folding of proteins adsorbed reversibly to ion-exchange resins. UCLA Symp Mol Cell Biol New Ser 1986;39:249–257.
79. Li M, Zhang G, Su Z. Dual gradient ion-exchange chromatography improved refolding yield of lysozyme. J Chromatogr A 2002;959:113–120.
80. Li M, Su Z. Refolding human lysozyme produced as an inclusion body by urea concentration and pH gradient ion exchange chromatography. Chromatographia 2002;56:33–38.
81. Li M, Su Z-G. Refolding of superoxide dismutase by ion-exchange chromatography. Biotechnol Lett 2002;24(11):919–923.
82. Chen Y, Leong SSJ. Adsorptive refolding of a highly disulfide-bonded inclusion body protein using anion-exchange chromatography. J Chromatogr A 2009;1216(24):4877–4886.

83. Freydell EJ, van der Wielen LAM, Eppink M, Ottens M. Ion-exchange chromatographic protein refolding J Chromatogr A 2010;1217:7265–7274.
84. Liu X-Q, Yang X-Q, Xie F-H, Song L-Y, Zhang G-Q, Qian S-J. On-column refolding and purification of transglutaminase from *Streptomyces fradiae* expressed as inclusion bodies in Escherichia coli. Protein Expres Purif 2007;51 (2):179–186.
85. Wang CZ, Liu JF, Geng XD. Refolding with simultaneous purification of recombinant human granulocyte colony-stimulating factor from *Escherichia coli* using strong anion exchange chromatography. Chin Chem Lett 2005;16(3):389–392.
86. Choi WC, Kim MY, Suh CW, Lee EK. Solid-phase refolding of inclusion body protein in a packed and expanded bed adsorption chromatography. Process Biochem 2005; 40(5):1967–1972.
87. Li M, Poliakov A, Danielson UH, Su Z, Janson J-C. Refolding of a recombinant full-length non-structural (NS3) protein from hepatitis C virus by chromatographic procedures. Biotechnol Lett 2003;25(20):1729–1734.
88. Werner MH, Clore GM, Gronenborn AM, Kondoh A, Fisher RJ. Refolding proteins by gel filtration chromatography. FEBS Lett 1994;345(2–3):125–130.
89. Batas B, Jones HR, Chaudhuri JB. Studies of the hydrodynamic volume changes that occur during refolding of lysozyme using size-exclusion chromatography. J Chromatogr A 1997;766:109–119.
90. Shalongo W, Ledger R, Jagannadham MV, Stellwagen E. Refolding of denatured thioredoxin observed by size-exclusion chromatography. Biochemistry 1987;26: 3135–3141.
91. Jungbauer A, Kaar W, Schlegl R. Folding and refolding of proteins in chromatographic beds. Curr Opin Biotechnol 2004;15:487–494.
92. Fahey EM, Chaudhuri JB, Binding P. Refolding and purification of a urokinase plasminogen activator fragment by chromatography. J Chromatogr B 2000;737:225–235.
93. Gu Z, Zhu X, Ni S, Zhou H, Su Z. Inhibition of aggregation by media selection, sample loading and elution in size exclusion chromatographic refolding of denatured bovine carbonic anhydrase B. J Biochem Biophys Methods 2003;56:165–175.
94. Gu Z, Su Z, Janson J-C. Urea gradient size-exclusion chromatography enhanced the yield of lysozyme refolding. J Chromatogr A 2001;918:311–318.
95. Gu Z, Weidenhaupt M, Ivanova N, Pavlov M, Xu B, Su Z-G, et al. Chromatographic methods for the isolation and refolding of proteins from *E. coli* inclusion bodies. Protein Expres Purif 2002;25:174–179.
96. Dong X-Y, Wang Y, Shi J-H, Sun Y. Size exclusion chromatography with an artificial chaperone system enhanced lysozyme renaturation. Enzyme Microb Technol 2002;30:792–797.
97. Schlegl R, Iberer G, Machold C, Necina R, Jungbauer A. Continuous matrix-assisted refolding of proteins. J Chromatogr A 2003;1009:119–132.
98. Lanckriet H, Middelberg APJ. Continuous chromatographic protein refolding. J Chromatogr A 2004;1022:103–113.
99. Park B-J, Lee C-H, Mun S, Koo Y-M. Novel application of simulated moving bed chromatography to protein refolding. Process Biochem 2006;41(5):1072–1082.

100. Park B-J, Lee C-H, Koo Y-M. Development of novel protein refolding using simulated moving bed chromatography. Korean J Chem Eng 2005;22(3):425–432.
101. Batas B, Schiraldi C, Chaudhuri JB. Inclusion body purification and protein refolding using microfiltration and size exclusion chromatography. J Biotechnol 1999;68:149–158.
102. Fahey EM, Chaudhuri JB, Binding P. Refolding of low molecular weight urokinase plasminogen activator by dilution and size exclusion chromatography—a comparative study. Sep Sci Technol 2000;35(11):1743–1760.
103. Batas B, Chaudhuri J. The influence of operational parameters on lysozyme refolding using size-exclusion chromatography. Bioprocess Biosyst Eng 2001;24(4):255–259.
104. Wang SSS, Chang C-K, Liu H-S. Step change of mobile phase flow rates to enhance protein folding in size exclusion chromatography. Biochem Eng J 2006;29(1–2):2–11.
105. Harrowing SR, Chaudhuri JB. Effect of column dimensions and flow rates on size-exclusion refolding of b-lactamase. J Biochem Biophys Methods 2003;56:177–188.
106. Leong SSJ, Middelberg APJ. Dilution versus dialysis: a quantitative study of the oxidative refolding of recombinant human alpha-fetoprotein. Food Bioprod Proc 2006;84(C1):9–17.
107. Sorensen HP, Sperling-Petersen HU, Mortensen KK. Dialysis strategies for protein refolding: preparative streptavidin production. Protein Expres Purif 2003;31(1):149–154.
108. West SM, Chaudhuri JB, Howell JA. Improved protein refolding using hollow-fibre membrane dialysis. Biotechnol Bioeng 1998;57(5):590–599.
109. Guiochon G, Beaver LA. Separation science is the key to successful biopharmaceuticals. J Chromatogr A 2011;1218(49):8836–8858.
110. Petrides D. Bioprocess design and economics. In: Harrison RG, editor Bioseparations Science and Engineering. New York: Oxford University Press; 2003.
111. Asenjo JA, Herrera L, Byrne B. Development of an expert system for selection and synthesis of protein purification processes. J Biotechnol 1989;11:275–298.
112. Leser EW, Asenjo JA. The rational selection of purification processes for proteins: an expert system for downstream processing design. Ann NY Acad Sci 1994;721:337–347.
113. Lienqueo ME, Asenjo JA. Use of expert systems for the synthesis of downstream protein processes. Computers Chem Eng 2000;24(9–10):2339–2350.
114. Lienqueo ME, Leser EW, Asenjo JA. An expert system for the selection and synthesis of multistep protein separation processes. Computers Chem Eng 1996;20(Suppl 1):S189–S194.
115. Leser EW, Lienqueo ME, Asenjo JA. Implementation in an expert system of a selection rationale for purification processes for recombinant proteins. Ann NY Acad Sci 1996;782:441–455.
116. Andrews AT, Noble I, Keeratipibul S, Asenjo JA. Physicochemical properties of the matrix proteins of three main culture vehicles. Biotechnol Bioeng 1994;44:29–37.
117. Bhambure R, Kumar K, Rathore AS. High-throughput process development for biopharmaceutical drug substances. Trends Biotechnol 2011;29(3):127–135.
118. Bensch M, Wierling PS, von Lieres E, Hubbuch J. High throughput screening of chromatographic phases for rapid process development. Chem Eng Technol 2005;28(11):1274–1284.

119. Coffman JL, Kramarczyk JF, Kelley BD. High-throughput screening of chromatographic separations: I. Method development and column modeling. Biotechnol Bioeng 2008;100(4):605–618.
120. Bergander T, Nilsson-Välimaa K, Öberg K, Lacki KM. High-throughput process development: determination of dynamic binding capacity using microtiter filter plates filled with chromatography resin. Biotechnol Prog 2008;24(3):632–639.
121. Kelley BD, Switzer M, Bastek P, Kramarczyk JF, Molnar K, Yu T, et al. High-throughput screening of chromatographic separations: IV. Ion-exchange. Biotechnol Bioeng 2008;100(5):950–963.
122. Kramarczyk JF, Kelley BD, Coffman JL. High-throughput screening of chromatographic separations: II. Hydrophobic interaction. Biotechnol Bioeng 2008;100(4):707–720.
123. Rege K, Pepsin M, Falcon B, Steele L, Heng M. High-throughput process development for recombinant protein purification. Biotechnol Bioeng 2006;93(4):618–630.
124. Susanto A, Knieps-Grünhagen E, von Lieres E, Hubbuch J. High throughput screening for the design and optimization of chromatographic processes: assessment of model parameter determination from high throughput compatible data. Chem Eng Technol 2008;31(12):1846–1855.
125. Susanto A, Treier K, Knieps-Grünhagen E, von Lieres E, Hubbuch J. High throughput screening for the design and optimization of chromatographic processes: automated optimization of chromatographic phase systems. Chem Eng Technol 2009;32(1):140–154.
126. Wensel DL, Kelley BD, Coffman JL. High-throughput screening of chromatographic separations: III. Monoclonal antibodies on ceramic hydroxyapatite. Biotechnol Bioeng 2008;100(5):839–854.
127. Wiendahl M, Wierling PS, Nielsen J, Christensen DF, Krarup J, Staby A, et al. High throughput screening for the design and optimization of chromatographic processes–miniaturization, automation and parallelization of breakthrough and elution studies. Chem Eng Technol 2008;31(6):893–903.
128. Coffman JL, Kelley B. High-throughput screening in downstream process development of protein therapeutics. Paper presented at the 229th ACS National Meeting Division of Biochemical Technology, San Diego, CA; 2005
129. Londo T, Lynch P, Kehoe T, Meys M, Gordon N. Accelerated recombinant protein purification process development: automated, robotics-based integration of chromatographic purification and analysis. J Chromatogr A 1998;798(1–2):73–82.
130. Wiendahl M, Völker C, Husemann I, Krarup J, Staby A, Scholl S, et al. A novel method to evaluate protein solubility using a high throughput screening approach. Chem Eng Sci 2009;64(17):3778–3788.
131. Wierling PS, Bogumil R, Knieps-Grünhagen E, Hubbuch J. High-throughput screening of packed-bed chromatography coupled with SELDI-TOF MS analysis: monoclonal antibodies versus host cell protein. Biotechnol Bioeng 2007;98(2):440–450.
132. Eppink MHM, Schreurs R, Gijsen A, Verhoeven K. Platform technology for developing purification processes. BioPharm Int 2007;20(3):44–50.
133. Ishihara T, Kadoya T. Accelerated purification process development of monoclonal antibodies for shortening time to clinic: design and case study of chromatography processes. J Chromatogr A 2007;1176(1–2):149–156.

134. Knudsen HL, Fahrner RL, Xu Y, Norling LA, Blank GS. Membrane ion-exchange chromatography for process-scale antibody purification. J Chromatogr A 2001;907 (1–2):145–154.

135. Yang H, Viera C, Fischer J, Etzel MR. Purification of a large protein using ion-exchange membranes. Ind Eng Chem Res 2002;41(6):1597–1602.

136. Mollerup JM, Hansen TB, Kidal S, Staby A. Quality by design–thermodynamic modelling of chromatographic separation of proteins. J Chromatogr A 2008;1177 (2):200–206.

137. Jakobsson N, Degerman M, Nilsson B. Optimisation and robustness analysis of a hydrophobic interaction chromatography step. J Chromatogr A 2005;1099:157–166.

138. Jakobsson N, Degerman M, Stenborg E, Nilsson B. Model based robustness analysis of an ion-exchange chromatography step. J Chromatogr A 2007;1138:109–119.

139. Jakobsson N, Karlsson D, Axelsson JP, Zacchi G, Nilsson B. Using computer simulation to assist in the robustness analysis of an ion-exchange chromatography step. J Chromatogr A 2005;1063(1–2):99–109.

140. Petrides DP, Koulouris A, Lagonikos PT. The role of process simulation in pharmaceutical process development and product commercialization. Pharm Eng 2002;22(1):1–8.

141. Guiochon G, Shirazi SG, Katti AM. Fundamentals of Preparative and Nonlinear Chromatography. London: Academic Press; 1994.

142. Nfor BK, Verhaert PD, van der Wielen LAM, Hubbuch J, Ottens M. Rational and systematic protein purification process development: the next generation. Trends Biotechnol 2009;27(12):673–679.

143. Nfor BK, Zuluaga DS, Verheijen PJT, Verhaert PD, van der Wielen LAM, Ottens M. Model-based rational strategy for chromatographic resin selection. Biotechnol Prog 2011;27(6):1629–1643.

144. Nfor BK. Model-based rational and systematic protein purification process development: a knowledge-based approach. PhD thesis, Delft University of Technology, Delft, Netherlands; 2011.

145. Nfor BK, Ahamed T, van Dedem GWK, Verhaert PD, van der Wielen LAM, Eppink MHM, et al. Model-based rational methodology for protein purification process synthesis. Chem Eng Sci 2013;89:185–195.

146. Nfor BK, Ahamed T, van Dedem GWK, Verhaert PD, van der Wielen LAM, Eppink MHM, et al. Multi-dimensional fractionation and characterization of crude protein mixtures: towards establishment of a database of protein purification process development parameters. Biotechnol Bioeng 2012;109(12):3070–3083.

147. Nfor BK, Noverraz M, Chilamkurthi S, Verhaert PD, van der Wielen LAM, Ottens M. High-throughput isotherm determination and thermodynamic modeling of protein adsorption on mixed mode adsorbents. J Chromatogr A 2010;1217(44):6829–6850.

148. Nfor BK, Hylkema NN, Wiedhaup KR, Verhaert PD, van der Wielen LAM, Ottens M. High-throughput protein precipitation and hydrophobic interaction chromatography: salt effects and thermodynamic interrelation. J Chromatogr A 2011;1218(49):8958–8973.

149. Rathore AS. Roadmap for implementation of quality by design (QbD) for biotechnology products. Trends Biotechnol 2009;27(9):546–553.

150. Rathore AS, Winkle H. Quality by design for biopharmaceuticals. Nat Biotech 2009;27(1):26–34.

151. Hedhammar M, Gräslund T, Hober S. Protein engineering strategies for selective protein purification. Chem Eng Technol 2005;28(11):1315–1325.
152. Deshpande K, Ahamed T, van der Wielen LAM, Horst JHt, Jansens PJ, Ottens M. Protein self-interaction chromatography on a microchip. Lab Chip 2009;9(4):600–605.
153. Meagher RJ, Light YK, Singh AK. Rapid, continuous purification of proteins in a microfluidic device using genetically-engineered partition tags. Lab Chip 2008;8(4).
154. Häusler E, Domagalski P, Ottens M, Bardow A. Microfluidic diffusion measurements: the optimal H-cell. Chem Eng Sci 2012;72:45–50.

# Chemical and Genetic Modification

MONIKA FARYS, CLAIRE L. GINN, GEORGE O. BADESCU, KAROLINA PECIAK, ESTERA M. PAWLISZ, HANIEH KHALILI, and STEVE J. BROCCHINI

## 9.1 INTRODUCTION

There are many natural posttranslational modifications (PTMs) of proteins (e.g., glycosylation, ubiquitination, acylation, amino acid derivatization) [1–4]. Glycosylation is common for therapeutic proteins. Classes of proteins modified by glycosylation include monoclonal antibodies (mAbs), blood factors, hematopoietic proteins, and cytokines. Excluding mAbs, many therapeutic proteins were developed as a replacement for a corresponding endogenous protein. Maintaining control of protein structure to be as close as possible to the structure of the endogenous protein has been a key goal.

Second-generation versions of many protein therapeutics involve some type of modification engineered to improve the protein [5,6]. For example, erythropoietin (EPO) [7] has been modified to have two more glycosylation sites (darbepoetin alfa; Aranesp), and EPO has also been covalently conjugated to poly(ethylene glycol) (PEG; methoxy polyethylene glycol-epoetin β; Micera). Broadly, these modifications of EPO were made to increase its circulation half-life.

As medicines, proteins are generally highly selective and potent, but they suffer from substantial limitations that can include the potential to be immunogenic (e.g., formation of antidrug antibodies), physicochemical instabilities (e.g., propensity to aggregate), and suboptimal pharmacokinetics (e.g., rapid clearance). Proteins have not naturally evolved to be produced and purified on a large scale as they are when used as medicines. Production and downstream processing can considerably influence protein homogeneity and physicochemical characteristics.

Proteins have certainly not evolved to be administered and used as medicines. Many endogenous proteins of therapeutic relevance are often regulated at low concentration within the body. Their clinical administration can result in highly fluctuating *in vivo* concentrations that also do not reflect their natural abundance and

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*,
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

biodistribution within the body. In this context, the potential for immunogenicity can increase.

The widespread clinical use of proteins is further complicated by side effects such as an increased susceptibility to infections or malignancies that may develop even after treatment has stopped [8]. Regulatory vigilance has resulted in an increase of postapproval safety-related regulatory actions [9] and some withdrawals (e.g., efalizumab, gemtuzumab, ozogamicin).

Despite these limitations, protein-based medicines are established with more than 174 protein therapeutics in the clinic [10,11]. Many types of modifications are being made to improve the therapeutic properties of proteins. These are becoming increasingly important as lifecycle management leads to biobetter versions of existing medicines and as nonendogenous proteins [12] are being introduced to the clinic. There are many reasons to modify proteins for therapeutic use. This chapter focuses on (1) optimization of protein pharmacokinetics, (2) improvement of protein properties to be used as medicines (this primarily includes reduction of immunogenic sequences and improvement of protein stability), (3) addition of a therapeutic effect, and (4) to use of proteins as diagnostics. Protein fusion methods are only discussed in the context of increasing blood circulation times. There is no effort in this review to discuss the important and highly active field of bispecific antibodies.

Modifications to proteins can be made by both recombinant [3,4] (e.g., protein fusion, structure based engineering, humanization) and chemical conjugation [13–18] (e.g., polymer conjugation, antibody–drug conjugates [ADCs]) approaches. Recombinant approaches include modifying protein structure to remove epitopes that could lead to increased propensity for immunogenic reactions [19,20] and to substitute amino acids that are prone to cause protein aggregation. Protein aggregation is often thought to contribute to immunogenicity [21,22]. Modifications that decrease aggregation and remove obvious epitopes or conformational sources of immunogenicity are therefore important [23]. Therapeutic proteins are large molecules that often exert their mode of action in the blood compartment, but increasingly, they are being used for intracellular targets (i.e., ADCs). Many potential applications for proteins do exist [24,25] and much early preclinical research has been described in the literature; however, this review focuses primarily on research that has at least gone to late preclinical development, and preferably to at least clinical evaluation.

## 9.2 PROTEIN MODIFICATION TO INCREASE CIRCULATION HALF-LIFE

The continued development of protein-based medicines relies on achieving optimal pharmacokinetics in the blood compartment. Although many therapeutic monoclonal antibodies (mAbs) (e.g., $IgG_1$) can display extended half-lives by recycling via interaction with the Fc receptor when a large enough dose is given, other proteins tend to clear rapidly (minutes to a few hours). The frequency for dosing is often increased when a protein rapidly clears from circulation. Although this is inconvenient to the patient and increases the cost of treatment, efficacy may be

compromised because rapid clearance can result in a subtherapeutic plasma concentration between doses. Increased dosing frequency also means the patient will experience side effects more frequently. Optimizing pharmacokinetics is often a key challenge that should be addressed early in preclinical development.

Many strategies have been evaluated to increase the circulation half-life of therapeutic proteins [26,27]. Clinically successful strategies include those based essentially on increasing the size of the protein by conjugation of a flexible water-soluble polymer or by addition of more glycosyl groups onto the protein. This is known as glycoengineering [5,28–32], which is generally accomplished recombinantly, but can also be accomplished by chemical modification strategies [32]. Polymer conjugation has focused primarily around PEGylation [33–39]. Another strategy involving the modification of a protein to prolong its circulation time is to exploit endocytic recycling mechanisms [40–48]. This is most often accomplished clinically by recombinantly fusing the protein of interest to the Fc portion of an immunoglobulin G (IgG). Much work has also been accomplished to fuse albumin or albumin binding proteins to the therapeutic protein of interest. Pumps to prolong dosing into the bloodstream have been successful with smaller polypeptides (e.g., insulin) and less often with proteins, probably because of formulation limitations to maintain protein structure in such conditions. The same can be said of colloidal systems where formulation limitations may also dominate. Although in principle the protein can be introduced to the circulatory system in unmodified form when released from a pump or colloid, the protein will still clear quickly. High flux may be necessary to ensure there is a therapeutic level of protein over the necessary time period required for efficacy.

### 9.2.1 Chemical Conjugation of Water-Soluble Polymers

Rapid clearance *in vivo* as a result of glomerular filtration and proteolytic degradation is a common limitation to the efficacy of many therapeutic proteins. Conjugation of a flexible, water-soluble polymer can increase the circulating half-life of proteins by tackling both of these issues [49,50]. The most common polymer used to modify proteins is PEG. The solution size of a PEG–protein conjugate is generally dominated by the PEG rather than the protein [51–54], and the resulting conjugate can have a hydrodynamic volume 10 times that of the unconjugated protein [52]. This size increase of the protein conjugate can be above the kidney filtration threshold, resulting in a significant extension of the blood circulation time. In addition, the PEG molecules can mask or block the protein surface by steric shielding effects to reduce interactions of the protein from proteolytic enzymes [55]. Increased blood residence time results in reduced dosing frequency because the therapeutic effect can be maintained for longer time periods. Steric shielding effects of the polymer can, in principle, reduce interactions of the protein with the immune system and antigen-presenting cells.

In addition to the limitations posed by their rapid clearance, protein-based medicines are complex molecules prone to aggregation. Protein aggregation can make it difficult to achieve dose reproducibility and is generally thought to increase the risk of immunogenicity. Regulatory agencies enforce stringent criteria for characterizing and minimizing aggregation phenomena in final dosage forms of

protein-based medicines. Because protein-based medicines are administered parenterally, formulation options are limited. Conjugation of a flexible, water-soluble polymer such as PEG can often decrease the susceptibility for a protein to aggregate in solution. Again, steric shielding effects of the conjugated polymer cause a decreased propensity for protein aggregation. Together, the requirements for optimized pharmacokinetics [56] and the need for stable formulations pose challenges that must be addressed when developing protein-based medicines.

Many different polymers have been conjugated to proteins by a wide range of methods [16,57–69] in efforts to extend their circulation half-life [17,39,59–63]. Notable polymers that have been examined in recent years include the degradable polysaccharides; polysialic acid [64–66], which has been conjugated to EPO and is in clinical evaluation; HEPylation (heparosan polymer); and hesylation (hydroxyethyl starch) [59]. In adition to PEG and some of its variants [67], polyoxazoline is a synthetic, nondegradable, flexible water-soluble polymer [68] that has also been examined for protein conjugation.

***9.2.1.1 Protein PEGylation*** Clinically, it is PEG which has been used as a conjugating polymer and it has been conjugated to proteins in at least 11 clinically used products (Table 9.1) [33,38,69] with approximately 20 PEGylated proteins in clinical development [39]. Hence, protein modifications derived from PEG will form the basis of this section.

As with almost any protein therapeutic, there are issues with the manufacture, characterization, and use of PEGylated proteins [70–75], including the potential induction of anti-PEG antibodies [76], especially in the case of hyperPEGylated products that contain large numbers of conjugated PEG molecules [77–78]. Many proteins used as therapeutics display some propensity for immunogenicity [79–84]. While PEG may have the potential to decrease the immunogenicity of the therapeutic protein, there is a risk of secondary antibodies against PEG. One aspect of anti-PEG antibodies may be attributable to the ubiquity of PEG in wide range of health care and consumer products as a contributing cause for anti-PEG antibodies [77,79]. Much work remains to be done to understand what might be the clinical features of any potential immunogenic reaction toward PEG in PEG-modified proteins [85], but it is generally thought that any observed toxicity of a PEG-protein conjugate is due to the protein [86].

The fact is that PEGylated proteins have an enviable track record of safe clinical use and benefit [86]. PEGylation can provide significant unexpected, and still not completely understood, therapeutic benefits [50,87–90]. There is still also much to be learned about how *in vitro* biological properties for PEGylated proteins obtained during preclinical studies correlate to what is observed *in vivo* [87,91] and in the clinic. The steric shielding effects of the polymer that prevent proteolytic degradation can also decrease biological activity, especially when observed using *in vitro* assays. While decreased *in vitro* activity is the norm for PEG-protein conjugates, it is essential that PK/PD data are obtained to determine the potential of preclinical candidates.

Steric shielding tends to decrease the on-rates of the PEG–protein conjugate for receptor binding [59,92–94]. The off rate, though, is often little changed from the

**TABLE 9.1 Clinically Registered Proteins Modified by the Covalent Conjugation of Poly(ethylene glycol)**

| PEGylated Product | Year Approved | Company | Type of Modification | PEG Used | Indication |
|---|---|---|---|---|---|
| PEG–adenosine deaminase (Adagen) | 1990 | Enzon | HyperPEGylation, random | Linear, 5 kDa | SCID |
| PEG–asparaginase (Oncaspar) | 1994 | Enzon | HyperPEGylation, random | Linear, 5 kDa | Acute lymphoblastic leukemia |
| PEG–INF-α2b (PEG-Intron) | 2000 | Schering-Plough/ Enzon | MonoPEGylation, random | Linear, 12 kDa | Chronic hepatitis C |
| PEG–INF-α2a (Pegasys) | 2002 | Hoffmann-La Roche | MonoPEGylation, random | Branched chain, 40 kDa | Chronic hepatitis C |
| PEG–G-CSF (Neulasta) | 2002 | Amgen/Nektar | MonoPEGylation, site specific | Linear, 20 kDa | Neutropenia |
| PEG–growth hormone receptor antagonist (Somavert) | 2002 | Pfizer/Nektar | HyperPEGylation, random | Linear, 5 kDa | Acromegaly |
| PEG–anti-VEGF aptamer (Macugen) | 2004 | Pfizer | MonoPEGylation, site specific | Branched chain, 40 kDa | Age-related macular degeneration |
| PEG–epoetin-β (Mircera) | 2007 | Hoffmann-La Roche | MonoPEGylation, random | Linear, 30 kDa | Anemia associated with chronic renal failure |
| PEG–anti-TNF-α Fab' (Cimzia) | 2008 | UCB | MonoPEGylation, site specific | Branched chain, 40 kDa | Crohn's disease and rheumatoid arthritis |
| PEG–uricase (Krystexxa) | 2010 | Savient | HyperPEGylation, random | Linear, 10 kDa | Gout |
| PEG–EPO analogue (Omontys) | 2012* | Affymax/Takeda | MonoPEGylation, site specific | Branched chain, 40 kDa | Anemia associated with chronic renal failure |

EPO, erythropoietin; INF: interferon; G-CSF: granulocyte colony-stimulating factor; PEG: poly(ethylene glycol); SCID: severe combined immunodeficiency disease; TNF: tumor necrosis factor; VEGF: vascular endothelial growth factor. *Recalled in 2013.

native protein. Although the site of PEG conjugation on the protein can influence the relative biological activity [95–100], PEGylation does not cause a protein to display a biological function that differs from the native protein [101–103]. Translating preclinical observations of PEG protein efficacy to the clinic must be done using relevant *in vivo* models in which increased circulation can compensate for decreased *in vitro* activities. The key appears to be that the extension of circulation time can result in an extended *in vivo* pharmacologic effect that has been clinically exploited for the different PEG-modified protein products now being used [49–50].

In several cases, the PEGylated protein has become first line treatment. Often the two clinically used PEGylated interferon-α2 (INF-α2) products are cited as definitive examples of how a PEGylated protein has shown real benefit [104]. PEGylated INF-α2 is used to treat hepatitis C, so efficacy can be measured empirically by reduction of the sustained viral load [105–108]. Only with the use of PEG–INF-α2 and ribavirin has it become possible to cure patients of some genotypes of hepatitis C. Although now first-line treatments, there was significant loss of *in vitro* activity observed during preclinical development for both PEGylated INF-α2 products (Pegasys, ∼7%; PEG-Intron, ∼28%) [109]. It is difficult to predict preclinical efficacy based only on observations of *in vitro* activity. Even though the *in vitro* activity is diminished for PEGylated INF-α2, the native protein is very active (∼9 pg/ml $ED_{50}$), which means even a decrease to 1% of the *in vitro* activity for the PEG protein may still be clinically efficacious if half-life can be extended.

Other PEGylated proteins (e.g., EPO, granulocyte colony-stimulating factor [G-CSF]) are also now standard treatments in noninfectious and often chronic conditions. In these examples, PEG modification of a clinically used protein has provided a means to make a biobetter version of the protein. A recently approved PEGylated product is peginesatide (Omontys, approved in 2012), which is a peptide that is PEGylated during solid-phase synthesis [110] and has been developed to treat patients with anemia who have developed pure blood red cell aplasia (PRCA), which can occur from treatment with EPO. While this product was not inferior with either epotin or darpepotin in anemia and chronic kidney disease, respectively, it was recalled in 2013 for safety concerns related to hypersensitivity [111] presumably caused by the therapeutic polypeptide.

Reduced activity *in vitro* accompanied with improved biological activity *in vivo* has also been demonstrated for other PEGylated protein therapeutics such as IL-2 [112], IL-6 [113], and tumor necrosis factor α (TNF-α) [114], for example. It is thought that preclinical development could be accelerated if site-specific PEGylation is easier to accomplish because in some cases, it appears that when PEG is conjugated to the same site on a protein, *in vitro* activity can be broadly maintained as PEG size is changes [115]. However, other studies indicate that PEG shape (morphology), when conjugated to the same site, can influence *in vitro* activity and that ion exchange chromatographic retention can correlate to activity [116].

Also, the possibility exists that if higher potency can be achieved while the protein displays extended circulation times, then lower doses or less frequent dosing schedules may be possible, either of which may be important as a means to minimize side effects over a given course of treatment. The defined incorporation of non-native

amino acids for site-specific conjugation in an optimal location on the protein may provide the option to achieve good potency for a polymer conjugated protein [117]. It has also been hypothesized that selective PEGylation at either termini of a protein might offer the potential to achieve good potency [118].

Although PEG is currently the only polymer used clinically for conjugation to therapeutic proteins for parenteral administration, much research continues to be focused on the development of protein–polymer conjugates using other polymers [59,119,120] and molecules (e.g., lipids). Many strategies used with PEG will also broadly apply to other polymers.

***9.2.1.2 The evolution of protein modification with PEG*** PEGylation was first described by Davis, Abuchowski and Palczuk in the late 1970's and it is generally recognized that there has been little toxicity observed for PEG in the wide range of consumer and health care products in which it is used [86]. PEG is approved by the Food and Drug Administration (FDA) as an excipient that is widely used in pharmaceuticals, including injectable, topical, rectal, and nasal formulations. This lack of toxicity is also generally observed when a PEGylated product is administered parenterally [86,122]. Although some concerns have been noted because of high-dose accumulation in animal models [123], which would be expected for any water-soluble polymer, recent regulatory approvals suggest that parenteral doses of up to 200 mg or more weekly of PEG as part of a PEGylated product are possible for extended periods [86,122,124–127].

In the early 1990s, the first PEGylated protein (PEG–adenosine or Adagen) was approved for the treatment of severe combined immunodeficiency disease (SCID) [128]. PEG–asparaginase (Pegaspargase, Oncaspar, approved in 1994) for acute lymphoblastic leukemia soon followed [129]. These are PEGylated nonhuman bacterial enzymes, and PEGylation allowed these medicines to be used more effectively in the clinic [130]. PEG-asparaginase is produced by hyperconjugation of 5 kDa PEG. The surface of PEG–asparaginase is saturated with PEG molecules (5 kDa). This is known as hyperPEGylation. Recently, PEG–uricase (pegloticase) (see Table 9.1) was approved for clinical use. This porcine-derived PEGylated protein is a tetramer and is used for the treatment of chronic gout. It is hyperPEGylated with 36 ($9 \pm 1$ per subunit) hydroxy PEG (10 kDa) molecules in an effort to reduce clearance by anti-PEG antibodies [131]. HyperPEGylation of uricase protects it from proteolytic enzymes to prolong its half-life and to minimize immunogenicity to uricase so it can be used to treat a chronic condition.

PEGylation has evolved from the initial random hyperPEGylation to monoPEGylation. Many of the current products in the clinic are monoPEGylated with PEG at different sites on the protein. Many of the conjugation strategies used with PEGylation have been used with other polymers. It should be noted, however, that a conjugation moiety on different polymers may have very different reactivity, and purification of the polymer conjugate may vary [132]. For example, purification of polysialylated conjugates will be considerably more anionic than the unconjugated protein. In contrast, PEG conjugates will not have any charge imparted to the conjugate by the PEG because it is a neutral polymer. Hence, purification by ion

exchange chromatography requires very different conditions for conjugates derived from these two polymers.

Typically, the strategy for polymer conjugation is directed toward the nucleophilic amino acid side chains (e.g., lysines, terminal amino group, serine, histidine). Although strategies have been described that result in some degree of selectivity for lysine conjugation [30], conjugation usually occurs at a number of different sites to yield positional conjugation isomers [133,134]. Another approach to vary protein reactivity [135,136] is recombinantly to deplete the number of lysines on the protein surface and replace them with less reactive residues to reduce the number of conjugation sites.

For monoPEGylated proteins where PEG is not selectively conjugated to a specific amino acid residue in the protein, each positional PEG–protein isomer will display different physiochemical, biological, and pharmaceutical properties [137–139]. The same would be expected for any polymer conjugated nonspecifically to a protein. Many of the PEG conjugation reagents (e.g., PEG–*N*-hydroxysuccinimide esters) also tend to undergo competitive hydrolysis reactions with water, which requires large excess of reagent for good conjugation conversion while complicating downstream processing.

Thiol-specific conjugations have been extensively examined to achieve more homogeneous monoPEG protein conjugates. Thiol-based conjugations are often more efficient and site specific than conjugations to other protein nucleophiles (e.g. lysine $\varepsilon$-amino groups). Most therapeutic proteins do not have a free cysteines. If cysteines are present in a therapeutic protein, there is usually an even number of them and they pair up to form protein disulfide bonds. Much work has been described to add a cysteine recombinantly to a protein for PEGylation [140,141]. However, if there is a disulfide present in the protein, these engineered proteins suffer from limitations associated with disulfide scrambling and aggregation. Despite these limitations, the advantages of thiol-specific conjugation were exploited to develop certolizumab pegol, which was approved in 2008. This PEG modified protein is an anti–TNF-$\alpha$ Fab with a free cysteine that has been engineered into its structure for conjugation to a PEG reagent that comprises two PEG molecules of 20 kDa each (total, 40 kDa; often simply called a 40-kDa branched PEG reagent).

Incorporation of a single cysteine is probably most appropriate for site-specific conjugation of proteins that do not have any disulfides. For example, a protein scaffold derived from a fibronectin type III domain that does not have a disulfide in its structure has been engineered to have a free cysteine for PEGylation and is currently in clinical trials [142,143].

Interestingly, many thiol-selective conjugations are conducted with maleimide-based reagents, and recent evidence indicates that these reagents are labile to hydrolysis [144] and can undergo exchange reactions in *vivo* [145]. To address maleimide instability, new thiol-selective reagents have been developed [146].

Another approach is to use conjugation reagents capable of bis-alkylation that can conjugate to both cysteine thiols from an existing disulfide in the protein [147,148]. Conjugation results in a 3-carbon bridge spanning the two cysteine thiols of the

original disulfide with PEG conjugated site-specifically on the bridge. An alternative approach for site-specific conjugation using bis-alkylating reagents is that they can conjugate a protein His-tag site-specifically [118]. Because it is rare to have multiple histidines next to each other in a protein [149,150], it makes the His tag a good target for selective conjugation of the PEG. This strategy offers the potential for site-specific conjugation to be applied to virtually any protein.

In addition to cysteine-specific modification, other amino acids such as arginine [151], glutamic acid [152,153], and even methionine [154] have been targets for site-specific conjugation. Glycosyl moieties have also been described as targets for protein modification [155–157]. PEGylation of either the C- [158,159] or the N-terminus [160–167] of proteins has also been described.

A particularly interesting approach for C-terminal polymer conjugation is to produce intein–fusion proteins that can be cleaved with hydrazine to give a C-terminal hydrazide that can then undergo ligation reactions with a suitable keto substituted polymer [168]. Reductive amination has been extensively investigated for protein N-terminal conjugation [165,169–171], and this process is used for pegfilgrastim (PEG–G-CSF). Imine formation is the first step in reductive amination with water as the byproduct, and so homogeneous imine formation is not generally efficient in terms of reagent stoichiometry or site specificity [170,172]. Efforts have been published to optimize reductive amination to G-CSF [173] and other proteins (e.g. staphylokinase, growth hormone) [170,174] as a means to maximize N-terminal selectivity and conjugation efficiency. The His tag–selective PEGylation mentioned earlier [118] is a potential alternative to C- and N-terminal PEGylation.

Other approaches for site-specific of modification of proteins have been described using enzymes, such as transglutaminase [175], glycosyl-transferases [176], and sortase [177] as well as genetic insertion of unnatural amino acids to enable generation of selectively PEGylated proteins [178–183]. Incorporation of non-native amino acids requires that cellular biochemical pathways probably need to be robustly modified [181,184] to ensure that the efficiency of protein production is not compromised by reversion to the wild-type phenotype.

Nonetheless, non-native amino acids with alkenyl, iodo, and keto moieties for reaction can in principle be incorporated and used for site-specific conjugation. For this strategy to work, tRNAs with the desired codon, usually the amber stop codon, are aminoacylated with the amino acid of interest. The reengineered tRNA is then added into the expression system along with the DNA template. This technique was used to insert an azide group at a specific site in the sequence of superoxide dismutase [178]. In this instance, conjugation with an alkyne terminated PEG yielded conjugates with similar activity to the native enzyme. The use of copper-mediated click reactions can be difficult to achieve with proteins, generally because of excessive purification processes and the sensitivity of proteins to the reaction conditions. Yields of amino acid incorporation can also be poor because of their competition with termination codons. This approach is currently limited to the addition of a single type of non-natural amino acid because only the action of the single amber stop codon can be suppressed.

### 9.2.2 Glycosylation Modifications

Glycosylation is the most common posttranslational modification of proteins. Modifying protein glycosylation to extend half-life is a clinically proven strategy, but it has not been as extensively used as PEGylation or protein fusion strategies. Many therapeutic proteins that are endogenous replacement proteins (e.g., cytokines, blood factors) or antibodies are naturally glycosylated and produced in glycosylated forms.

Glycosylation can play a significant role in protein bioactivity and pharmacokinetics and can increase solubility [28,29]. Additionally, glycosylation has been found to prevent protein instabilities, including: oxidation, cross-linking, precipitation, pH, chemical or thermal denaturation, aggregation, and proteolytic degradation [185]. Improved stability has been observed for glycoproteins such as G-CSF (Granocyte) [186] and INF-γ (Actimmune) [187]. Many excellent reviews describe the scope of protein glycosylation and the analytical challenges involved with characterization [32,188–191].

Clinically, the dominant platform for glycoprotein production is mammalian expression systems, especially Chinese hamster ovary (CHO) cell lines [32,192,193]. Other methods, including enzymatic, chemical, and chemoenzymatic *in vitro* glycosylation methods, have also been described [32]. CHO expression systems have the ability to produce proteins with N-glycans similar to human proteins; however, effort is required to minimize development timelines from gene to production cell line. Often, control of glycosylation structure is a key challenge with multiple glycoforms being produced [194]. This technology has been applied, for example, to the production of darbepoetin-α (Aranesp), a hyperglycosylated analogue of EPO. Aranesp has two additional N-linked carbohydrates, resulting in a threefold increase in serum half-life and increased *in vivo* activity compared with recombinant human EPO [195,196]. Darbepoetin-α is primarily used to treat anemia associated with chronic kidney disease and HIV infection [196,197].

Much preclinical research has been conducted to find other methods to prepare glycosylated proteins. Yeast-based expression systems have also been examine to produce therapeutic proteins, including mAbs [198]. The major advantages are robust expression, scalable fermentation, and the ability to perform N-glycosylation [199]. Pichia pastoris can be engineered to produce glycoproteins with a high degree of homogeneity, and this yeast system has the ability to optimize the glycosylation for specific therapeutic functions [198,200,201]. Glycoproteins produced in yeast can have a high mannose content, which can induce immunogenicity and can result in a shorter serum half-life [202]. Transgenic plants are also being examined as expression systems to make glycoproteins [32,203–205]. Nonhuman glycan structures that lack galactose and terminal sialic acid that contain xylose and α1,3-fucose can be present [206,207]. Synthetic strategies have also been explored [208,209]. This approach has been applied to make glycosylated EPO. A cysteine tag was used in a solid-phase ligation strategy to introduce an N-glycosylation site followed by chemoselective reaction with glycosyl iodoacetamide [210] or bromoactetamide [211].

### 9.2.3 Fusion Proteins

Another approach to increase the plasma half-life of therapeutic proteins is by recombinantly linking or fusing the protein of interest to another protein that displays extended circulation half-life. Fusion of the protein of interest to the Fc fragment of an IgG is designed to exploit the recycling mechanism with the pH-dependent binding of the neonatal Fc receptor (FcRn) [212]. Human serum albumin also binds to the Fc receptor and has also been used to fuse with a therapeutic protein. Fusion proteins have displayed decreased toxicity [213] and active targeting [214]. Although the vast majority of fusion proteins are produced by recombinant strategies, chemically based conjugation strategies to link two proteins together have also been described [215].

The use of recombinant strategies to produce Fc fusion proteins has been clinically successful. Examples include (Enbrel), Alefacept (Amevive) [216, 217], Abatacept (Orencia) [218], Rilonacept (Arcalyst), Romiplostim (Nplate) [219–221], and Belatacept (Nulojix) [222, 223] (Table 9.2). In addition to half-life

**TABLE 9.2 Examples of fusion proteins [27,212]**

| Product | Company | Fusion Partner 1 | Fusion Partner 2 |
|---|---|---|---|
| *Examples of Clinically Approved Fusion Proteins* | | | |
| Amevive (alefacept) | Astellas Pharma US | IgG Fc | LFA-3 |
| Arcalyst (rilonacept) | Regeneron | IgG Fc | IL1Rs |
| Enbrel (etanercept) | Amgen | IgG Fc | TNFR |
| Eylea (aflibercept) | Regeneron | IgG Fc | VEGFR1 /VEGFR2 |
| Nplate (romiplostim) | Amgen | IgG Fc | TPO-receptor (c-mpl) binding peptide |
| Ulojix (belatacept) | Bristol-Myers Squibb | IgG Fc | CTLA4 |
| Orencia (abatacept) | Bristol-Myers Squibb | IgG Fc | Mutated CTLA4 |
| Ontak (denileukin diftitox) | Ligand Pharmaceuticals | IL-2 | Diphtheria toxin |
| *Examples of Fusion Proteins in Clinical Trials (Phase III)* | | | |
| LEA29Y | BMS | IgG Fc | CTLA4 |
| VEGF Trap | Regeneron | IgG Fc | VEGF |
| IL13-PE38QQR | NeoPharm | IL-13 | *Pseudomonas* exotoxinPE38 |
| Proximium/Vicinium | Viventia | Anti-EpCAM FAb | *Pseudomonas* exotoxin A |

EpCAM: Epithelial cell adhesion molecule; CTLA4: Cytotoxic T-Lymphocyte Antigen 4; Ig: immunoglobulin; IL: interleukin; IL1Rs: Interleukin-1 receptor; LFA-3: lymphocyte function-associated antigen; TNFR: Tumor necrosis factor receptor; TPO: thrombopoietin; VEGF: vascular endothelial growth factor.

extension, in many cases, fusion of a therapeutic protein with the Fc domain can improve its solubility or stability (e.g., protection from proteolysis) [224]. Interestingly, the serum half-life differs for many of the Fc fusion proteins in the clinic (ranging from 4 to 13 days). For example, the half-life of Enbrel is 4.3 days, but for Alefacept and Abatacept, the half-lives are 11.3 and 13.1 days, respectively [225]. It remains unclear whether the receptor regions of Fc-fusion proteins alter the conformation of the CH2–CH3 domain interface or the regions cause steric hindrance on the binding site of FcRn.

Albumin fusion has also been examined to extend the circulatory half-life of therapeutic proteins. Albumin also bind to FcRn, which is necessary for observed circulation half-life of 19 days for albumin [226]. Although albumin is a complex protein, one appeal for using it as a fusion partner that was often cited is the albumin avoids effector functions that can be associated with some Fc regions. The first reported albumin fusion was with CD4. There was a 140-fold increase in the circulation half-life of $0.25 \pm 0.1$ h for CD4 to $34 \pm 4$ h for the albumin–CD4 fusion in the rabbit. The binding and *in vitro* anti-viral activity of CD4 for the albumin–CD4 fusion was similar to those of soluble CD4 [227]. Further development of albumin fusion technologies has focused on proteins that have been used clinically, for example, cytokines including INF-β [48], INF-α2b (Albuferon) [47,228], growth hormone (Albutropin) [229], G-CSF (Albugranin) [46], EPO [230] and interleukin-2 (IL-2) (Albuleukin) [231]. Additional examples include hirudin [232], factor VIIa [233], and thioredoxin-1 (Trx) [234] and bispecific antibodies [235]. Peptides include glucagon-like peptide-1 (GLP-1) (Albugon) [236], insulin (Albulin) [237] and recombinant B-type natriuretic peptide (AlbuBNP) [238] to treat congestive heart failure. Several of these albumin fusion proteins have been evaluated in preclinical studies. A few albumin fusions have been evaluated in clinical trials, including Albutropin, AlbuBNP, and albumin IL-2. Albuferon (INF–albumin fusion) appears to have progressed the furthest in clinical development; unfortunately, phase III trials were stopped.

Recombinant fusions of therapeutic proteins to transferrin (Tf) have also been evaluated preclinically. Tf is a blood plasma glycoprotein that reversibly binds ferric ions at physiological pH and is important for the control of the amount of iron in the blood. When Tf is bound to iron, it can bind to a Tf receptor (Tfr) on the cell surface and is transported into the cell by receptor-mediated endocytosis. The acidic pH in the endosomal compartment of the cell causes Tf to release its iron ions, and then Tf together with Tfr is recycled to the cell surface [239]. Fusion with Tf, which is resistant to enzymatic digestion, allows it to be considered a stable protein partner for use in fusions with the potential for extended half-life being possible. Tfr is expressed in high numbers on epithelial cells. Therefore, Tf fusions have been considered for oral drug delivery. Recently, human growth hormone–Tf fusion was evaluated in preclinical studies [240]. Tf fusion technology has also been applied to GLP-1 and exendin-4 (EX-4) in efforts to improve their pharmacology [241].

Fundamentally, there are several disadvantages with attempting to recombinantly fuse proteins together. It is often difficult to isolate a fusion protein pure in sufficient

amounts during early development and at scale for production [242,243] because there is often an increased propensity for aggregation and enhanced degradation [244,245]. Therapeutic fusion proteins have not evolved to be together [246–248] so it can be difficult to ensure that each of the fused proteins is correctly folded and that there is an optimal sequence of amino acids that link the two proteins together. The flexibility, composition, and appropriate length of the polypeptide linker between the proteins to be fused are important factors [249] that must be optimized [250–255]. It also is unfavorable to have a linker sequence with a high propensity for forming α-helical or β-strand structures because this limits the flexibility of the fused proteins and affects their functional activities. Moreover, the linker potentially can be a source of immunogenicity [256]. Although computational methods can be used to design the linker sequence [257], linker selection still remains a challenge in the construction of functional fusion proteins.

## 9.3 STRUCTURE-BASED ENGINEERING TO MODIFY PROTEIN AND PEPTIDE THERAPEUTICS

The continuous improvement in DNA technologies has resulted in strategies to modify protein structure. It is not only possible to produce recombinant protein-based therapeutics but also to engineer favorable properties that can improve biological and physiochemical characteristics of a protein of interest [258].

### 9.3.1 Sequence-Based Modifications

Endogenous human proteins are expressed *in vivo*. Many are often expressed in low concentration, sometimes as pro-variants, and often have an immediate effect on their target. These proteins have also evolved to function within the physicochemical environment of the body. In contrast, therapeutic proteins are exposed to a variety of chemical and physical stresses during production and downstream purification. They are also stressed during formulation processes, and while stored and used as final dosage forms.

Recombinant variants of endogenous and novel proteins that are being developed as therapeutics can often be engineered to display improved solubility and stability. Maintaining protein stability throughout its lifetime from production to clinical use poses many challenges. Site-directed mutagenesis to modify protein structure can be achieved with high fidelity [259] to make the protein more robust in regards to proteolysis, oxidative stress, or temperature and pH changes [5]. Such efforts can help to decrease the propensity for unfolding, aggregation, and degradation processes that can compromise protein function. Other strategies include diversity methods, such as phage display, or semirational protein design methods [5] (Table 9.3).

For example, proteins can potentially be stabilized by introduction of disulfide bond [260], removal of free cysteine [261], or modulation of the internal hydrophobicity of the protein [262]. Replacement of free cysteines can prevent the formation

**TABLE 9.3 Recombinant Strategies for Protein Modification [258]**

| Issue Addressed | | Engineering Strategy | Example |
|---|---|---|---|
| Lack of posttranslational modifications in *Escherichia coli* | Glycosylation | Expression without the glycosylation (only if the biological activity is not impaired) | IL-2 [355] |
| | | Mutation of residues on the surface to create more soluble protein | EPO [356] |
| | | Mutation of glycosylation sites to cysteines to allow subsequent glycosylation *in vitro* | EPO [210] |
| | Proteolytic maturation | Expression in two separate strains or cleavage of the precursor *in vitro* | Insulin [357] |
| | Disulfide bridge formation | Expression in the periplasmic space | hGH, proinsulin [358] |
| Protein stability | | Decreasing number of free cysteine residues by mutation to alanine | IL-2 [359,360], IFN-β1 [261] |
| | | Deletion of the hydrophobic region | KGF [361] |
| Modulation of protein activity | | Design of the rapid-acting or long-acting protein version | Insulin [362] |
| | | Enhanced activity by improved affinity to the target molecule | DNaseI [363] |
| | | Design of protein consisting of a consensus sequence | IFN-α-con [364] |

hGH: human growth hormone; IFN: interferon; IL: interleukin; KGF: keratinocyte growth factor.

of undesired intermolecular or intramolecular disulfide bonds. This approach has been successfully demonstrated on several therapeutically relevant proteins with important clinical applications, such as INF-β1b [263], G-CSF [264], and human fibroblast growth factor (FGF) [5].

Another methodology is to increase the stabilizing factors that influence protein folding. As a good first approximation, protein folding is driven by hydrophobic properties of nonpolar amino residues. Strengthening of internal hydrophobicity by amino acid substitution may increase protein stability [265]. This approach is often aided by using computational design methods, which are used to optimize packing and hydrophobic interactions within the folded protein [266,267].

Protein structure can be also improved by optimizing the propensity for localized secondary structure, hydrogen bonds, and electrostatic interactions [268]. This approach has been successfully applied to clinically relevant proteins, including G-CSF and human growth hormone [269]. The optimized variant of site-specifically mutated G-CSF displays enhanced thermal stability and prolonged shelf life while maintaining the desired biological activity [5].

An additional aspect of protein stabilization is to reduce the susceptibility of a protein to proteolytic degradation. If the sequence within a protein that is prone to proteolysis can be determined, it can be genetically modified so it is no longer recognized by the putative protease. Frequently, protease cleavage sites are located in a flexible loop of the protein structure; therefore, one approach of introducing site-specific modification, which reduces the flexibility of the loop, may lead to the reduction of protein's propensity for proteolytic degradation.

As expected, care must be taken to maintain protein structure and biological activity [5], but this approach has been used for factor VIII-a. For blood clotting factors, specific proteolytic susceptibility is important for activation and inactivation. The coagulation factor VIII-a with increased resistance to proteolytic inactivation was developed by genetic mutation of two arginines, which are required for cleavage by thrombin, factor Xa, and activated protein C [270].

Maintaining protein solubility during production, formulation, and administration processes is another important issue associated with the development of effective therapeutics. Proteins are usually expressed and then formulated at relatively high concentrations, which can result in aggregate formation. Aggregation can cause a decrease in protein activity, reduced bioavailability, and potentially increased immunogenicity. Several strategies to minimize aggregation by protein engineering have been developed and demonstrated on clinically important proteins.

For example, replacement of an unpaired cysteine residue prevents formation of undesired intermolecular disulfide bonds during processing, which can lead to aggregation. Substitution of solvent exposed nonpolar residues with polar residues may lead to improvement in protein's solubility. An example is the A1 domain of cholera toxin, in which a variant with significantly improved solubility and retained biological activity was obtained [271]. Another approach is to alter the net charge of protein to change the isoelectric point (pI), which also can lead to improvement of protein solubility. An example of this approach is with a single-chain antibody targeting renal cell carcinoma. The protein sequence was altered by adding five glutamic acid residues to the C-terminus, leading to lowering the pI from 7.5 to 6.1 and therefore increasing its solubility [272]. Many of these modification strategies have been used with the clinically used insulin analogues that have been developed.

Identification and reduction of potential immunogenicity are key safety concerns [19,80]. The clinical use of protein-based medicines is complex. Many patients treated with proteins may develop immune response [23]. There are consequences from the processes used in the manufacture of proteins as well as the inherent complexity of protein structure and how they are administered. In addition to immunogenic effects, unwanted side effects such as an increased susceptibility to infections or malignancies may develop during treatment or after treatment has

stopped [9]. Postapproval safety-related regulatory actions have been issued for approximately 23% of protein-based medicines in the United States and European Union [9]. Immune responses to a protein therapeutic may also alter pharmacokinetics, reduce efficacy, cross-react with an endogenous protein, and cause an adverse reaction [19]. Immunogenicity may also lead to life-threatening situations if the produced antibodies not only neutralize the administrated therapeutic but also neutralize the endogenous protein [79].

Protein immunogenicity [6,81] may be caused by numerous factors, which may be categorized into three main groups: structure-related characteristics of the protein, product related-factors [22,83], and patient-related factors [273]. The presence of heterogeneous mixtures may contribute to long-term safety related issues of protein-based medicines [83]. A significant structural factor is the degree of nonhuman sequence within the protein. Additionally, immunogenicity may be caused by noncovalent-derived structural changes of the protein, such as protein mis- or unfolding and protein aggregation, which will result in presentation of new amino acid arrangements on the surface of the protein.

Although this still remains a complex task, recent advances in mass spectrometry and computational epitope mapping can be used in efforts to reduce immunogenicity [274,275]. This approach is very useful for nonhuman proteins. A clinical example is staphylokinase, which is a bacterially derived protein that activates human plasminogen for thrombolysis. Staphylokinase (and streptokinase) are generally used in acute situations unlike many protein therapeutics that are dosed over a period of time. However, avoidance of immunogenic reactions is still important for patient safety. A reduction in its immunogenicity was obtained by site-specific mutagenesis in which 13 amino acids were substituted to alanine [276].

### 9.3.2 Humanization of Nonhuman Proteins

Investigation of the potential for protein immunogenicity and how it can be reduced is a necessary aspect in the development process [19,23,81,277]. Protein engineering provides valuable tools in efforts to manage immunogenicity, such as humanization of nonhuman proteins (e.g., insulin analogues and mAbs) or structure-based modifications to remove immunogenic epitopes [20,22,278,279]. In this second approach, the primary sequence of a protein, which is recognized by the immune system, can in principle be reengineered so it is no longer recognized as foreign. This process can be facilitated using epitope mapping, which allows T- and B-cell epitopes to be identified and subsequently modified [275,280–282]. For many therapeutics, especially mAbs, a decrease in therapeutic immunogenicity is obtained by changing the protein sequence from murine to fully human [283].

The widespread use of therapeutically relevant mAbs has led to development of technologies to humanize their sequence. The human anti-mouse antibody (HAMA) response is common for murine antibodies. The HAMA response neutralizes the administered antibody to reduce efficacy and can also cause anaphylactic shock [284]. One study found that out of the 40 antibodies that have been in the clinic, 84%

had marked HAMA reactions, 7% had tolerable HAMA, and only 9% were found to exhibit negligible HAMA reactions [285].

The strategy of "humanization" [286] was accomplished first with the constant regions of antibodies to make chimeric and increasingly more humanized versions. Humanization of antibody variable regions has been more recent. The variable regions comprise the complementarity-determining regions (CDRs), which constitute the antigen-binding site and framework regions (FRs), which form the scaffold around the binding site. CDR grafting has been successful to reduce immunogenicity, although there can be a reduction in the binding affinity of the antibody [287]. As the amino acids of the scaffold interact with those of the CDR regions, it is also necessary to back mutate some of the key residues found here that are important for maintain binding site integrity with the aim of retaining as much binding affinity as is possible. This approach has proved so effective that humanized antibodies now represent half of the current FDA-approved mAbs on the market [286].

One such antibody, Zenapax (daclizumab), was the first humanized antibody approved by the FDA in 1997 for human therapeutic for use in the United States and is used to prevent renal allograft rejection. It is a humanized IgG1 form of the mouse mAb anti-Tac, an anti-human IL-2 receptor chain (CD25) antibody that blocks the interaction of IL-2 with the IL-2 receptor and thus prevents the activation of T cells [287]. Humanization of this antibody led to 14% low titer of anti-idiotype A antibodies and 34% lower in pediatric patients compared with the nonhumanized antibody [285].

Another more recently approved humanized antibody that benefited from humanization was the murine anti-CD3 antibody (muromonab), which was found to elicit neutralizing antibodies after a single dose in all patients tested; a humanized version was immunogenic in only 25% of patients after multiple injections [285]. Novel antibodies are being developed as their humanized version. Although immunogenicity may be reduced compared with murine and chimeric versions, many mAbs still display considerable toxicities that have to be monitored, especially in oncology and inflammation.

## 9.4 PROTEIN MODIFICATION TO ADD A THERAPEUTIC EFFECT

The modification approaches previously described in this chapter are aimed at improving the properties of a therapeutic protein or peptide. In contrast to this, conjugation of drugs to proteins or peptides mainly aims to improve the therapeutic properties of the small molecule drug. The main clinical area is oncology as many of the small molecule drugs are cytotoxic with the need to increase their therapeutic window to improve efficacy.

The protein or peptide component of such a conjugate as well as the linker can be regarded as a carrier that helps to improve the solubility, stability, pharmacokinetics, bioavailability, biodistribution or localization, selectivity, toxicity profile, targeting potential, and, ultimately, therapeutic index of the drug component. However, the efficacy of a therapeutically active protein or peptide can also be enhanced by conjugation to a pharmacologically active drug (payload) that would give it new therapeutic functions.

Currently, at least five protein-drug conjugates (PDCs) are approved for treatment, and many more are at different stages of development. Due to the clinical success of protein–drug conjugates, the interest in this field has increased at an accelerating rate over the past decade. This is reflected by the growing number of research publications, candidate compounds in development, and conferences and commercial deals, as well as a large range of excellent reviews for the main types of conjugates, such as albumin–drug conjugates, antibody-drug conjugates (ADCs), and radioimmunoconjugates.

The most important motivation for conjugating drugs to proteins and peptides is the potential to widen the therapeutic window of the drug by targeting its delivery to specific cell types, tissues, or organs. The concept of targeted chemotherapy is more than a century old and was first described by Paul Ehrlich, who proposed the idea of creating "magic bullets" to treat all human diseases.

Despite the numerous and significant advances for therapies in oncology, most current therapeutic agents have relatively low selectivity, a low therapeutic index, and a narrow therapeutic window. Most anticancer drugs have to be dosed at near their maximum tolerated dose (MTD) to achieve a therapeutic effect. In addition to the development of drug resistance, many severe side effects are frequently associated with current chemotherapies. These include nephrotoxicity, hepatotoxicity, bone marrow toxicity, neurotoxicity, cardiovascular toxicity, mucositis, and gastrointestinal toxicity. These unwanted side effects can lead to serious complications that are difficult and expensive to treat, are dose-limiting, and can prevent effective treatment and can force the discontinuation of lifesaving treatment.

As a result, cures are achieved only in a small set of cancers and patients. However, there have been some clinical successes for several targeted therapy approaches in the past 15 years that have brought a new wave of optimism. A few notable examples are the development of highly efficacious mAbs, tumor-specific kinase inhibitors, and, maybe the most promising of all, ADCs that combine the specificity of an antibody with the cytotoxicity of a highly potent chemotherapy drug.

### 9.4.1 Protein– and Peptide–Drug Conjugates as Strategies for Targeted Drug Delivery

These conjugates comprise a protein, a linker, and an active agent that is covalently conjugated to the protein or peptide via the linker. The principle underlying protein– and peptide–drug conjugate (PDC)-based targeted drug delivery strategies is that the protein component has affinity for the target site, where the drug can be released (often intracellularly). Selective drug release is a key requirement for most active agents because their pharmacological activity relies on binding to cellular targets. However, radioisotopes have a different mode of action and do not have to be released at the target site [288–290].

PDCs have a multistep mechanism of action, with the different steps having different and even contradictory requirements. Ideally, the PDC should be stable during storage, nontoxic, stable, and inactive in circulation during its journey to the target site, where drug should then be released. The whole PDC can be regarded as a drug precursor, which should be systemically nontoxic. The requirement for this

conditional stability is key to the specific delivery of a drug to the intended target site and makes the development of PDC therapeutics challenging.

The accumulation of the protein at the target exploits the morphological and physiological differences between the disease target site and healthy cells, tissues, and organs. Passive targeting strategies have been used to exploit the passive accumulation of large macromolecules in some malignant and inflamed tissue caused by a leaky vascular system combined with a defective drainage system, which is known as enhanced permeability and retention (EPR) [291,292–295].

Strategies that aim at increasing the molecular mass and hydrodynamic volume of a drug (e.g., conjugates to large proteins or polymers, nanoparticles, micelles) mainly rely on EPR. However, EPR is not specific for the drug conjugate because it occurs for all other macromolecules in circulation. Although enhanced accumulation in some tumors has been demonstrated for large drug conjugates compared with the free drug, in some cases (e.g., nanoparticles that are devoid of any active targeting mechanisms) 95% of the administered dose accumulated in other nontarget organs such as the liver, spleen, and lungs [295]. Because the majority of the administered dose accumulates in nontarget sites, Bae and Park [295] have suggested that the term "passive targeting" is a misnomer, and it should be replaced with "blood circulation and extravasation" to reflect its nonselective nature.

Active targeting exploits the affinity of the carrier protein for specific antigens or receptors that are either uniquely expressed or overexpressed on the target site [296,297]. The targeting or guiding molecule can be a peptide, mAb, antibody fragment, growth factor, hormone, or any other protein that binds specifically to the target antigen or receptor. Of all of the protein–drug conjugates, ADCs are attracting by far the most attention.

### 9.4.2 Antibody–Drug Conjugates

Monoclonal antibodies are established as very useful therapeutic agents in oncology. However, only a few antibodies that target hematological malignancies, such as rituximab (anti-CD20) and alemtuzumab (anti-CD52), are effective as single-agent therapies to date. Antibodies targeting solid tumors, such as trastuzumab and pertuzumab (anti-HER2), cetuximab and panitumumab (anti-EGFR), or bevacizumab (anti-VEGF), are ineffective as single-agent therapies and have to be used in combination with cytotoxic drugs. There are many different mechanisms through which antibodies exert their anticancer activity, such as antibody-dependent cytotoxicity (ADCC), complement-dependent cytolysis (CDC), antibody-dependent phagocytosis (ADP), interference with cellular signaling, depletion of circulating tumor cells, or induction of apoptosis. Despite this, they are unable to sustain a lasting effect and thus have a very limited curative potential.

At the same time, although many small molecule drugs are very potent against cancer cells, their effective clinical use is limited by low selectivity and high toxicity to nontarget tissues. ADCs constitute a very promising approach to combine the best features of antibodies, such as high target selectivity, low systemic toxicity, and long circulating half-life with the high potency of cytotoxic drugs or radionuclides.

ADC technology is being investigated by a wide range of pharma and biotech companies, reflecting their confidence in the potential for clinical and commercial success of this approach. As of April 2013, Thomson Reuters' Integrity database listed 170 ADCs, with 9 products launched and 43 products at different stages of clinical development (including drug and radionuclide conjugates for both therapeutic and diagnostic use).

***9.4.2.1 Brief Overview of ADC History*** Only three therapeutic ADCs have ever been approved for treatment. The first one, gemtuzumab ozogamicin (Mylotarg), was conditionally approved in 2000. In postapproval phase III trials, gemtuzumab ozogamicin failed to demonstrate a survival benefit over conventional therapy and was associated with a higher rate of fatal toxicity than in the control arm [298]. Despite controversy in the medical literature about this study, Pfizer voluntarily withdrew it from the market in 2010. Currently, there are several ongoing phase II and III trials examining gemtuzumab ozogamicin in a range of indications and patient subpopulations. Recent results from phase III trials in patients with acute myeloid leukemia (AML) showed a significant clinical benefit for combining standard induction therapy with lower dose or fractioned lower doses of gemtuzumab ozogamicin [298].

The second ADC, brentuximab vedotin (SGN-35, Adcetris), was granted accelerated approval by the FDA in August 2011 based on results from single-arm phase II trials in Hodgkin's lymphoma and anaplastic large cell lymphoma, which demonstrated an acceptable toxicity profile and remarkable objective response rates of 73% and 86%, respectively [299]. The third ADC, trastuzumab emtansine (T-DM1, Kadcyla), demonstrated significantly prolonged progression-free and overall survival compared with standard treatment in patients with HER2-positive advanced breast cancer who were previously treated with trastuzumab and a taxane (the EMILIA trial) [300]. It was granted approval by the FDA in February 2013.

The recent successes in the ADC field follow on from over three decades of sustained research efforts and are a tribute to the lessons learnt from the many previous failures. Historically, the different phases of ADC development follow very closely the development of cytotoxic drugs.

***9.4.2.2 First-Generation ADCs*** During the 1980s, chemotherapy drugs such as doxorubicin, methotrexate, cisplatin, or vinca alkaloids were used. These were well-characterized and already approved drugs to treat cancer [301–304]. When examined *in vivo* in xenograft tumor models in mice, several ADCs demonstrated significant antitumor activity because of enhanced tumor accumulation (10–30% ID/g of tumor) compared with the free drug.

However, clinical evaluation of three such conjugates failed to demonstrate a therapeutic benefit or clinical response. Moreover, the ADC that progressed to phase II trials in gastric adenocarcinoma and metastatic breast cancer, BR96–doxorubicin (SGN-15), proved less effective than unconjugated doxorubicin [301,303]. The explanation is that the ADC failed to deliver sufficient (therapeutic) concentrations

of drug to the tumor cells. Several contributing factors have been identified, such as insufficient drug potency, low density of antigen expression, poor internalization and drug release, low tumor localization, linker instability leading to premature release of drug in circulation, and the occurrence of anti-ADC neutralizing antibodies.

The discord between the observed *in vivo* animal studies and clinical antitumor activity can also be explained by the very different tumor uptake of specific antibodies in patients (0.003–0.08% ID/g of tumor) and in mouse xenograft models (10–30% ID/g) [299]. An average human is 3000-fold greater in body weight than a mouse; therefore, each gram of tumor is diluted into 3000-fold more body weight and that 10 to 30% ID/g in mouse is approximately equivalent to 0.01 to 0.003% ID/g of tumor in human patients [303]. Also, in some cases, the human tumor-associated antigens targeted had no or insufficient homology with the rodent antigens, leading to significant underestimation of on-target toxicity and an overestimation of tolerability and therapeutic index. Consequently, since the 1990s, as well as developing more potent cytotoxic drugs, which are active in the concentration range of antibody accumulation in tumors ($IC_{50}$ values of 0.01 to 0.1 nM) [301], and novel linkers that are stable in circulation and allow rapid release of the drug in its most active form inside the tumor cells, the focus has also been on developing more predictive *in vivo* models.

***9.4.2.3 Second-Generation ADCs*** Second-generation ADCs use cytotoxic drugs that are 100- to 1000-fold more potent *in vitro* than the drugs used by the first-generation ADCs. Clinical-stage ADCs use one of the following cytotoxic drugs: Wyeth/Pfizer's calicheamicin ($IC_{50}$ $10^{-10}$–$10^{-9}$ M), ImmunoGen's maytansinoids DM1 ($IC_{50}$ $10^{-11}$–$10^{-9}$ M) and DM4 ($IC_{50}$ $10^{-12}$–$10^{-11}$ M), Seattle Genetics' auristatin E ($IC_{50}$ $10^{-11}$–$10^{-9}$ M) and auristatin F ($IC_{50}$ $10^{-8}$–$10^{-7}$ M), or the Medarex/BMS's MGBA (minor groove binding agent) duocarmycin analog (CC-1065) ($IC_{50} < 10^{-9}$ M).

ADCs using a double prodrug approach are sometimes referred to as *third-generation ADCs*. One such example is the Medarex UltiMab approach involving conjugation of MGBA to antibodies through a cleavable peptidic linker. The cleaved MGBA–linker adduct has reduced potency and requires a further activation step by intracellular carboxyesterases before releasing the active drug [305]. One such ADC targeting CD70 (MDX-1203) is currently being evaluated in a proof-of-concept phase I clinical trial.

***9.4.2.4 Key Features of ADCs*** ADCs are a clinically validated strategy for generating therapeutics with improved efficacy and tolerability. However, the ADC strategy has yet not met the promise of the "magic bullet." Only a small proportion of the administered drug ever reaches the tumor and a similar profile of dose-limiting nontarget toxicities was observed for all ADCs for which clinical results have been reported [306–309]. The past couple of decades have seen major advances in understanding the *in vivo* behavior of ADCs. However, there are very few general rules and there is little about ADCs that can be rationally predicted and designed. The ADC field remains largely experimental with the design and optimization of novel ADCs still dominated by an empirical trial and error approach.

Ongoing efforts are focused around key features of ADCs, such as the target antigen and antibody, payload, conjugation site, stability and stoichiometry, type of linker, and mechanism of payload release. Currently, there are over 30 ADCs under clinical evaluation; clinical data from the many ongoing trials will become available over the next few years and will guide further optimization of ADCs.

***9.4.2.5 Selection of Suitable Target Antigen and Antibody*** Antigen selection is one of the most important factors determining the selectivity of payload delivery to the target cancer cells. The ideal target antigen would be uniquely expressed on the cancer cell; however, such antigens have not yet been identified. Most antigens targeted by ADCs are tumor-associated rather than tumor-specific and are expressed on the cancer cells only at higher densities than on normal cells [297]. In most cases, binding to the antigen triggers the internalization of the ADC into the target cell and subsequent trafficking to specific subcellular compartments where the drug component is released or converted to its fully active form. In general, the antigens targeted by ADCs are internalized by receptor-mediated endocytosis and delivered into the endocytic pathway, often ending up in the lysosomal compartment [301,310,311]. However, the rate and extent of internalization and subcellular trafficking depend on the biology of the cell surface antigen or receptor. Some receptors internalize continuously regardless of whether a ligand is bound or not, but other receptors, such as the epidermal growth factor receptor, internalize at an accelerated rate when the natural ligand or ligand-mimetic protein is bound [301].

A small number of antigens and receptors, such as CD20 or CD21, are very poorly internalized or reside permanently on the cell surface [312]. Even when targeting a well-characterized antigen, the behavior of an ADC cannot be predicted. For instance, conjugation of the cytotoxic drugs doxorubicin or auristatin to rituximab, a noninternalizing anti-CD20 antibody, yielded ADCs that were internalized and delivered to the lysosome with great efficiency [313]. However, the same was not true when rituximab was conjugated to calicheamicin, a different cytotoxic drug [314].

For other antigens, such as CD30, both the naked antibody and the auristatin conjugates have been shown to internalize with similar rates [315]. Another interesting example is given by L49, a normally noninternalizing antibody targeting the p97 (melanotransferrin) antigen. In contrast with the naked antibody, the auristatin conjugates exhibited very efficient internalization and trafficking to the lysosome in antigen-expressing sensitive cell lines. However, in a subgroup of antigen-positive but nonresponsive cell lines, the internalized ADC was trafficked away from the lysosome through caveolae-mediated transport to locations lacking proteolytic activity, where drug release and function were prevented [311,316].

These examples suggest that internalization and subcellular trafficking depend not only on the nature of the antigen but also on the nature of the drug molecule, conjugation technology used, and cell phenotype, among many other potential factors. The antigen density on the target cells, as well as rate of internalization and subcellular trafficking, will determine the concentrations of payload that can be achieved in the target cell, and hence will drive efficacy. However, the relative level

of expression on normal cells compared with target cells is also extremely important as this will drive the on-target toxicity and tolerability profiles.

It is not surprising that the two recently approved ADCs that have shown favorable safety profiles and remarkable clinical efficacy, brentuximab vedotin and trastuzumab emtansine, target CD30 and HER2, respectively, which are antigens that are highly overexpressed on cancer cells compared with normal cells and that are internalized very efficiently [298]. For a long period, it was thought that internalization or trafficking to the lysosome is essential for optimal ADC activity [311]. However, there have been reports of good *in vivo* and clinical efficacy with ADCs targeting antigens expressed at relatively low levels on the cancer cells [298] or ADCs that are poorly internalized and thus release the drug extracellularly [314,317]. Alternative strategies that have recently given encouraging results target noninternalizing antigens overexpressed on the endothelial cells or subendothelial extracellular matrix of tumor vasculature [302,318,319].

***9.4.2.6 Selection of Payload Release Mechanism: Linker Design*** Selective release of the drug at the target site is usually achieved by incorporating special features into the linker between the drug and the antibody [305,320–323]. Three main types of linkers have been described: (1) chemically cleavable linkers, (2) enzymatically cleavable linkers, and (3) noncleavable linkers. Linker design strategies are based on the physiological differences between the extra- and intracellular compartments for internalizing ADCs or the differences between the extracellular space of target and nontarget tissues for noninternalizing ADCs.

Early developments focused on exploiting the pH difference between these compartments and used acid-labile hydrazone linkers, which have been described in many preclinical studies. A hydrazone linking moiety is more stable in the neutral environment of the blood (pH 7.3–7.5) than in the acidic intracellular environment of the endosome (pH 5.0–6.5), lysosome (pH 4.5–5.0), or tumor pericellular space, where such linkers can be readily hydrolyzed to release the active drug [305,322].

The first ADC to be granted FDA approval, gemtuzumab ozogamicin (Mylotarg), involved random conjugation of calicheamicin through an acid-labile hydrazone linker to the solvent-exposed lysine residues of an anti-CD33 mAb. The disappointing performance of Mylotarg was associated with insufficient plasma stability of the acid-labile linker leading to premature release of toxic amounts of calicheamicin in circulation [321,324]. Acid-labile linkers are generally considered relics of the early days of ADC development, which are unable to confer sufficient plasma stability to an ADC. However, inotuzumab ozogamicin (CMC-544), a calicheamicin conjugate to an anti-CD22 antibody using identical linkage chemistry to Mylotarg, has demonstrated very good serum and plasma stability and promising clinical efficacy with an objective response rate of 68% for follicular non-Hodgkin's lymphoma and 15% for diffuse large B-cell lymphoma. It is currently being evaluated in phase III trials [305,325,326]. Other examples of linkers that undergo proton-catalyzed hydrolysis at low pH incorporate biodegradable polymers, such as hydrophilic polyols [327,328].

Another type of chemically cleavable linkers are those containing a disulfide, which is cleaved upon exposure to a free thiol [305]. Cleavage selectivity is based on

the presence of glutathione (a thiol-containing tripeptide) inside the cell at concentrations that are three orders of magnitude higher than in blood, that is, millimolar versus micromolar. Moreover, the hypoxic state of many cancer cells can lead to further accumulation of glutathione in the intracellular or pericellular space. However, it is well known that thiol-selective maleimide moieties can undergo exchange reactions *in vivo* with free thiols from other proteins such as albumin [145], so a disulfide-linked ADC would be expected to be labile during circulation.

Recently, Bernardes et al. [319] have reported a novel linkerless approach that involves direct conjugation through disulfide bonds of a thiol-containing cytotoxic drug to the engineered cysteine residues of an antibody. Selective cleavage of the disulfide bond in the reducing tumor microenvironment generated only two species, the antibody and the cytotoxic drug, making this ADC traceless. Although the strong antitumor activity seen for this noninternalizing vascular-targeting ADC in a murine tumor model is very encouraging, it remains to be established whether such a conjugate would have sufficient plasma stability for human treatment.

Another approach is to use enzymatically cleavable linkers containing peptide sequences, for example, valine–citrulline, phenylalanine–lysine, or the more hydrophilic valine–lysine sequence, that are cleaved by lysosomal proteases such as cathepsin or plasmin [305,329]. Such peptide linkers can exhibit very good serum stability often 100-fold greater than their chemically cleavable counterparts. Alternative enzyme-cleavable linkers contain β-glucuronide, a carbohydrate moiety, as the cleavage site [330–333]. Such linkers are more hydrophilic than peptidic linkers and are readily cleaved by β-glucuronidase, an enzyme present in the lysosomal compartment and in the interstitial space of many tumors. Enzyme-cleavable linkers often contain self-immolating spacers between the drug and the cleavage site. Such spacers not only distance the drug from the cleavage site to avoid steric hindrance allowing efficient enzymatic cleavage, but also, upon enzymatic cleavage, initiate a cascade of reactions that result in the release of the drug in its most active form, free from any linker adducts [305].

Another option is to use stable, noncleavable linkers to link the drug to the protein- or peptide-guiding molecule. The use of such linkers requires ADC internalization and delivery to the highly catabolic lysosomal compartment, where the protein or peptide backbone is completely degraded, releasing the drug and its linker. The condition for the success of this approach is that the drug remains active when conjugated to the linker moiety [334,335].

Noncleavable linkers are by far the most stable in circulation but are often hydrophobic. This can lead to issues with aggregation, can limit the drug loading, and can have a negative impact on the functionality of the targeting molecule, the pharmacokinetic profile of the ADC, and the pharmacology of the released drug–linker compound [305]. To address these limitations, novel hydrophilic linkers containing either a negatively charged α-sulfonic acid group or a polar short PEG chain have been synthesized [336,337]. Additionally, phosphonic acid and tetraalkyl ammonium derivatives have also been used to improve the solubility of the linker and the resulting ADC [338].

***9.4.2.7 Payload Selection: Matching the Payload and Linker*** Depending on their mechanism of action, the payloads adopted in ADCs at clinical stages of development can be grouped into tubulin-targeting payloads (Seattle Genetics' auristatins or ImmunoGen's maytansinoid drugs) and DNA-intercalating payloads (Pfizer's calicheamicin and the CC-1065 duocarmycin analog from Medarex). There is also one ADC in phase I that uses SN-38, a topoisomerase 1 inhibitor (Labetuzumab-SN-38, Immunomedics), and one that uses doxorubicin, a first-generation tubulin inhibitor (Milatuzumab-doxorubicin, Immunomedics) [298,339,340].

ImmunoGen and Seattle Genetics have licensed their drug-linker technologies to multiple biopharmaceutical companies, so most ADCs under clinical development use one of the two technologies.

Seattle Genetics' auristatin cytotoxic drugs are conjugated to the reduced cysteine residues of the interchain disulfide bonds of IgGs through either enzyme cleavable maleimidocaproyl-valine-citruline-*p*-aminobenzyl linkers (MC-vc-PAB-MMAE/F) or noncleavable maleimidocaproyl linkers (MC-MMAF). Brentuximab vedotin (Adcetris) consists of an anti-CD30 mAb with MMAE conjugated to an average of four cysteine residues via an enzyme cleavable valine–citrulline linker. Upon internalization, the linker is cleaved by lysosomal proteases, and the self-eliminating PAB spacer releases unmodified free MMAE intracellularly [304,321].

All clinical-stage ADCs with MMAE use a -vc- cleavable linker because noncleavable linker MMAE constructs display reduced potency. On the other hand, MMAF is more tolerant to linker modifications and has been used with both cleavable -vc- linkers (e.g., SGN-75, AGS-16M8F) and noncleavable -mc- linkers (e.g., SGN-19A). The higher potency of free MMAE compared with free MMAF seen *in vitro* is mainly attributable to the higher cell penetrability of MMAE. However, when delivered inside the cell by an ADC, the two compounds display similar potency [334].

ImmunoGen's Tumor-Activated Prodrug (TAP) technology and antibody–maytansinoid conjugates (AMCs) consist of antibodies linked at lysine residues via either a cleavable disulfide linker (SPP-DM1 or the disulfide hindered SPDB-DM4) or a noncleavable thioether linker (MCC-DM1) to three to four cytotoxic drug molecules [341]. For some AMCs, such as trastuzumab emtansine (T-DM1), the disulfide cleavable and thioether noncleavable variants showed very similar antitumor activities *in vivo* in xenograft models, but the noncleavable-linked conjugate displayed better serum stability, so it was chosen for clinical development [342]. Cantuzumab emtansine (huC242-SPP-DM1) was evaluated in phase I trials but was subsequently discontinued because of linker instability. The half-life of the conjugate was significantly shorter than the half-life of the antibody component ($t_{1/2}$ conjugate = 41 h vs. $t_{1/2}$ antibody = 230 h).

Kellogg et al. [341] generated a panel of conjugates incorporating disulfide cleavable linkers with different degrees of steric hindrance around the disulfide bond, as well as a control conjugate with a noncleavable linker. Increasing the degree of steric hindrance around the disulfide bond resulted in increased serum stability of the conjugates, but all conjugates displayed similar *in vitro* activities. However, when evaluated *in vivo* in xenograft models, the conjugates with highly hindered disulfide linkers, as well as the conjugates with noncleavable linkers, displayed only marginal

activity, but the huC242-SPDB-DM4 conjugate with a linker with an intermediate level of disulfide hindrance displayed the best efficacy. This can be partially explained by the observation that the DM4 maytansinoid metabolite released from the SPDB-DM4 linker was more potent at inhibiting tumor growth than DM1, suggesting that the *in vivo* antitumor activity of an AMC depends on other factors in addition to conjugate potency and half-life [341]. huC242-SPDB-DM4 (IMGN242) has been evaluated clinically (I/II) but has been discontinued because of the requirement for personalized dosing and slow patient recruitment. However, several other ADCs conjugated with SPP-DM1, such as lorvotuzumab mertansine (IMNG901, phase I/II), or with SPDB-DM4, such as BT-062 (phase I/II), IMNG-388 (phase I), and SAR3419 (phase I), are currently under clinical evaluation.

Currently, there are approximately 16 compounds in clinical development that use Seattle Genetics' technology, including one with FDA approval, and seven compounds use ImmunoGen's TAP technology, including one with recent FDA approval. Only one ADC using Seattle Genetics' technology has been discontinued from clinical development for undisclosed reasons, and six of ImmunoGen's TAP compounds have been discontinued for reasons including narrow therapeutic window, linker instability, lack of activity, requirement for individualized dosing, or change in business strategy [306–309].

After the drug is released, it diffuses across the lysosomal or endosomal membrane to bind to its intracellular target or leaves the cell through active or passive processes [301,305,310,343]. Depending on their cell permeability, drugs can diffuse across to neighboring cells or, if the drug is released extracellularly, enter nearby cells. For instance, the two forms of auristatin most commonly used, monomethyl auristatin E (MMAE) and monomethyl auristatin F (MMAF), differ in their cell permeability properties. MMAF cannot diffuse across the cell membrane, but MMAE can readily diffuse into the nearby cells and exert its cytotoxic activity, regardless of whether the nearby cells are antigen positive or not—the bystander effect [316,334].

In contrast, drugs attached to noncleavable linkers are usually unable to diffuse across the cellular membrane and so are unable to exert a bystander effect [317,335]. Bystander effects can be beneficial, particularly when targeting solid tumor cells, which can be very heterogeneous in terms of antigen expression [311]. However, the potential benefit of killing all tumor cells has to be balanced with the risk of normal cell toxicity.

An absolute requirement for a successful ADC is that therapeutic concentrations of active drug are achieved and maintained at the target site. The concentration of drug in the target cell is a balance between the amount of drug reaching the cell and the amount of drug leaving the cell. The amount of drug that reaches target cells depends on multiple factors, such as the pharmacokinetic, biodistribution, and plasma stability profiles of the ADC, the drug loading, rate of internalization, efficiency of drug release, efficiency of prodrug activation, and, for non-internalizing ADCs, the rate of diffusion from neighboring cells or the pericellular space. Efflux of drug from the target cells depends mainly on the cell permeability of the drug and on the occurrence of active efflux mediated by drug transporters. The emergence of multidrug resistance (MDR) in tumor cells is an important factor that may limit the

activity of ADCs and is mainly attributed to upregulation of active drug efflux mediated by drug transporters such as the P-glycoprotein (Pgp) [336]. Most cytotoxic drugs (e.g., calicheamicin, doxorubicin, taxanes, maytansinoids, dolostatin 10 derivatives, CC-1065) are good substrates for Pgp, mainly because of their hydrophobicity. Successful approaches to bypass MDR have involved modifying the drug molecule or the linker. For instance, the charged MMAF molecule is a poorer substrate for Pgp than the neutral MMAE [334]. Also, substitution of a hydrazide group with an amide in the linker of a calicheamicin conjugate resulted in greater accumulation in MDR cells [311]. Another approach has been to modify the drugs with hydrophilic noncleavable linkers [336,337]. The more hydrophilic linker–drug compounds demonstrated enhanced retention by Pgp-expressing cells because it is thought that more hydrophilic compounds are poorer substrates for the efflux pumps.

The cytotoxic payload has to be matched with the biology of the target cancer cells. For instance, tubulin-targeting cytotoxic drugs, such as auristatins and maytansinoids, which at the moment dominate the clinical development pipelines, are highly potent against rapidly dividing cells but are less active against slowly dividing cells. This selectivity may contribute to the therapeutic index of ADCs with such payloads by limiting unwanted side effects to toxicities in rapidly dividing healthy cells, which are generally manageable. However, these payloads are inactive against slowly dividing target cells such as tumor-inducing cells (TICs), also known as cancer stem cells (CSCs). DNA-targeting payloads, such as calicheamicin, are more active against slowly dividing cells. On one hand, this makes them better suited to targeting TICs, but, on the other hand, they are generally less well tolerated. A high proportion of the current research efforts in the ADC field are focused on developing new highly potent DNA-targeting payloads such as duocarmycins (Synthon) and pyrrolobenzodiazepine dimers (PBDs) (Spirogen), as well as payloads with a new mode of action for ADCs such as the RNA polymerase inhibitor α-amanitin (Heidelberg Pharma) [298,321,339,344].

***9.4.2.8 Conjugation: Payload Stoichiometry and Stability of Conjugation*** The conjugation site(s), the stoichiometry of drug loading, and the stability of conjugation impact not only the function of the targeting protein molecule but also the aggregation, pharmacokinetic, and biodistribution profiles of the ADC. The effect of drug conjugation on the targeting properties of a drug conjugate has been extensively studied for both albumin [291] and ADCs [145,345–347]. Although a higher drug loading may seem preferable for its potential to achieve higher drug concentrations at the target site, in practice it often has the opposite effect [291,345,347]. The ideal number of drug molecules per antibody appears to be between 2 and 4. It was shown that increasing the drug loading to over 4 per antibody results in higher plasma clearance rates, lower tolerability, decreased efficacy, and actually results in less drug reaching the tumor [347,348].

Most clinical-stage ADCs have an average drug–antibody ratio (DAR) of around 4. However, the two main conjugation technologies currently adopted, namely conjugation to the side chains of solvent-exposed lysine residues (ImmunoGen, Pfizer) or to cysteine residues liberated by reducing the interchain disulfides (Seattle Genetics),

generate heterogeneous mixtures with a broad DAR distribution, generally between 0 and 8. Each DAR species can have its own distinct biological profile with potential differences in efficacy, pharmacokinetics, distribution, immunogenicity, and safety. Moreover, the naked antibody (DAR 0) can act as a competitive inhibitor, limiting the ADC uptake into the target cells and thus potentially reducing efficacy. Mylotarg was a very heterogeneous compound, with an average DAR of between 2 and 3, a DAR distribution of between 0 and 8, and 50% unconjugated antibody [305,321,325]. It is postulated that, in addition to the instability of the acid-labile linker, the high proportion of unconjugated antibody could be a contributing factor to the disappointing results seen for Mylotarg. Interestingly, inotuzumab ozogamicin (CMC-544), another ADC utilizing the same linker-payload and conjugation chemistry, but with a higher average DAR (5–7) and with less than 10% unconjugated antibody, shows very promising clinical efficacy [305,325,326].

Another source of heterogeneity with implications for both efficacy and tolerability is the instability of conjugation. It is well known that maleimide conjugates can undergo retro-addition reactions *in vivo* in the presence of competing free thiols, such as the cysteine 34 of circulating serum albumin [144]. This can lead not only to deconjugation and premature loss of payload from the ADC, but also to cross-conjugation to albumin and the formation of albumin–drug adducts. The clinical implications of maleimide instability are not yet fully understood.

One of the most prominent preclinical stage approaches to overcome the broad DAR distribution of current ADCs is to engineer unpaired cysteine residues at defined locations on the antibody. Conjugation of the drug to the unpaired cysteine residues using maleimide chemistry yields ADCs with homogeneous DAR [147,346,348]. Genentech's THIOMABs is the most advanced technology using engineered unpaired cysteine residues, and this approach is currently being investigated by several other companies, including MedImmune, Seattle Genetics, and Pfizer. Shen et al. [145] showed that the stability of the maleimide linkage and the extent of cross-conjugation to circulating albumin depend on the local environment of the engineered cysteine residue.

The work with THIOMAB ADC variants highlighted the importance of conjugation stability and DAR homogeneity. When examined *in vivo*, the less stable variants were associated with increased off-target toxicity and lower efficacy [145]. Moreover, a homogeneous DAR 2 THIOMAB ADC showed equivalent *in vivo* efficacy and a higher maximum tolerated dose (MTD) compared with a standard ADC with an average DAR of 4 and a DAR distribution of 0 to 8, despite having only about half the amount of drug [145,346,348]. This is an elegant solution to the broad DAR distribution issue but it is not without its limitations. It requires antibody reengineering, which can be time consuming and costly, the engineered cysteine residues can become capped during expression and uncapping can lead to reduction of interchain disulfides and heterogeneous conjugation, and it does not fully solve the issue of maleimide instability.

An alternative approach that avoids the use of maleimide linkers involves conjugation at engineered non-natural amino acids (Ambrx, Allozyne). However, antibody reengineering and expression of variants containing non-natural amino

acids can be complex and expensive. Moreover, the ability to manufacture such ADCs remains challenging as currently reported methods of conjugation require high reagent stoichiometry (up to 30 molar equivalents) and long reaction times (up to 4 days) [349].

***9.4.2.9 Efforts to Control Drug Release*** The potential of ADCs to widen the therapeutic window of small molecule drugs stems from their ability to reduce the systemic exposure of the free drug while increasing the dose of the drug at the target site (i.e., tissue, cells, or both). To achieve this, most ADC strategies incorporate multiple levels of control of drug delivery, often with a cumulative effect on the therapeutic index. However, each such level comes with its own limitations and challenges and has to be tailored to the biology of the target site. The first level of control is given by the targeting mechanism. Passive targeting is nonspecific and has a low degree of selectivity because the majority of the administered dose often accumulates in nontarget tissues [295]. Although active targeting is more selective, it is rarely specific because most antigens are overexpressed rather than uniquely expressed on the target cells [296,297]. Moreover, even for highly specific targeting molecules such as antibodies, the peak concentrations accumulated at the tumor site are less than 0.01% of the administered dose per gram of tumor [303], while the rest of the administered dose remains in circulation or is deposited at nontarget sites.

The second level of control is incorporated in the linker design to achieve selective release of the drug molecule at the target site. The main limitation of cleavable linkers is their plasma instability [305]. Even if the drug is released in circulation at a very low rate, this can cumulate over the many days that most ADCs can remain in circulation. Although noncleavable linkers are very stable in circulation, they are unsuitable for noninternalizing ADCs because they rely on internalization and lysosomal degradation and are also incompatible with bystander effects. Moreover, even when a stable linker is used, drug release in nontarget cells can occur because of antigen-independent internalization through pinocytosis [311].

Strategies to overcome these limitations and reduce toxicity have focused on improving linker stability and on integrating additional levels of drug release control. Double prodrug strategies [305,343] require additional chemical or enzymatic activation steps of the released prodrug at the target site, so even if nondesired release occurs in circulation or at other nontarget sites, the prodrug should have no or reduced toxicity. Other approaches use prodrugs that require photoactivation at the target site (photodynamic therapy) [350] or use thermoresponsive peptide drug carriers that can be activated by locally induced hyperthermia [351].

Alternative approaches to improve ADC therapy may use techniques already established for radioimmunotherapy, such as antibody pretargeting [288,299] and the shielding, sequestration, or removal of antigen-positive nontarget cells before ADC therapy [338]. Another recent approach involved removing the unbound fraction of an ADC from circulation 24 h after administration by extracorporeal affinity adsorption treatment [351]. Such approaches have the potential to optimize the tumor-to-nontumor ratio and thus improve efficacy and reduce toxicity.

## 9.5 PROTEIN MODIFICATION USED FOR DIAGNOSTICS

Diagnostics refers to the process of trying to identify a disease or disorder and usually requires the use of clinical tests. The aim of these tests is to detect or quantify certain biomolecules that are of diagnostic significance or alternatively to identify therapeutic targets for treatment purposes. Proteins, especially antibodies, are particularly suited to this role because of their ability to bind to a wide range of biological targets with unparalleled affinity and specificity. To perform this role, however, proteins often require some form of modification. These modifications can generally be divided into two categories, those allow the antibody detection on binding and those that facilitate their attachment to surfaces. Consequently, antibodies often require the conjugation of different molecular entities such as fluorophores, radiolabels, or even enzymes. These modifications endow the new molecules with new properties that allow them to be used for a wide range of applications from immunohistochemistry (IHC) and immunoassays to imaging and the production of protein microarrays.

Immunoassays are some of the most commonly used diagnostic tools for the analysis of biomolecules, with enzyme linked immunosorbent assays (ELISAs) one of the most currently used immunoassays. The technique first requires the presence of an immobilized ligating reagent, either an antibody or antigen to a solid support, usually a plate surface. An antibody–antigen binding complex is then formed on the addition of its binding partner. After a series of washing steps to remove unbound components, quantification of the successful binding can be evaluated by the addition of a detection reagent or an antibody that is able to bind specifically to the formed complex. Detection requires the conjugation of a measurable label either an enzyme such as horseradish peroxidase or alkaline phosphatase and the subsequent addition of a substrate or the attachment of a fluorophore, such as fluorescein or rhodamine, to the detection antibody. The generated signal can then be quantified and is proportional to the amount of analyte in the sample. ELISA assays have been used diagnostically for the screening and antibody detection in patients with suspected HIV. Hepatitis C infection is also diagnosed by ELISA using antigens to detect antibodies directed against the hepatitis virus [352]. A similar technique, IHC, refers to the process of detecting antigens such as proteins in cells of a tissue section by exploiting the principle of antibodies to bind specifically to antigens present on biological tissues. Again, in this method, the antibodies carry an indicator molecule for their detection and subsequent examination by a microscope. IHC is widely used clinically for the detection of antigens present on abnormal cells such as those found in cancerous tumors. Fluorescently labelled antibodies can also be used to determine cell populations in biological samples. One example of this is the use of antibodies against CD4 and CD8 antigens, which are used routinely to determine the population and of T lymphocytes to follow the progress of patients with AIDs.

Radiolabeled protein-based imaging agents have been particularly useful in the diagnosis and early detection and targeting of certain pathological conditions such as cancer, myocardial injury, and the presence of sites of infection (Table 9.4) [11].

**TABLE 9.4 Protein-Based Imaging Agents [11,347,365]**

| Imaging Agent | Company | Protein Features | Isotope | Clinical Uses |
|---|---|---|---|---|
| Acutect | Diatide, Inc. | Synthetic peptide that binds to GP IIb/IIa on activated platelets | $^{99}$Tc | Deep vein thrombosis |
| Bexxar | SmithKline Beecham | Mouse IgG2a antibody against CD20 | $^{131}$I | Non-Hodgkin lymphoma |
| CEA-Scan | Immunomedics | Mouse IgG1 anti-CEA Fab fragment | $^{99}$Tc | Breast and colorectal cancer |
| Protascint | Cytogen | Mouse IgG1 anti-PSA antibody | $^{111}$In | Metastatic prostate cancer |
| Myoscint | Centocor | Mouse IgG2a Fab against human cardiac myosin | $^{111}$In | Myocardial injury |
| Octreoscan | Mallinckrodt Medical | Labelled octreotide | $^{111}$In | Neuroendocrine tumors |
| Verluma | Boehringer Ingelheim | Mouse IgG2b anti-CEA Fab fragment | $^{99}$Tc | Small cell lung cancer |
| Oncoscint | Cytogen | Mouse IgG1 antibody against the glycoprotein TAG-72 | $^{111}$In | Colorectal and ovarian cancer |
| NeutroSpec | Palatin Technologies | Mouse IgM anti-CD15 antibody | $^{99}$Tc | Appendicitis |
| Zevalin | Spectrum Pharmaceuticals | Mouse IgG1 against CD20 | $^{111}$In | Non-Hodgkin's lymphoma |

CE: carcinogenic antigen; GP: glycoprotein; Ig: immunoglobulin; In: Indium; PSA: prostate-specific antigen; TAG: Tumor Associated Glycoprotein; Tc, technetium.

Several coupling reactions have been devised to attach the radiolabel without blocking the antibody binding site. One such method is to label the antibody via its oligosaccharide chain; an alternative approach is to attach a chelator to the $NH_2$ group of lysine [353].

Heavy metals such as indium and technetium are particularly attractive for this purpose because free isotopes are nontoxic. For imaging purposes, these imaging agents must also possess a high specificity for their target while showing little nonspecific binding to nontarget molecules on healthy tissue (hence why proteins are ideal carriers). Moreover, these molecules must be able to withstand the harsh conjugation conditions required for labeling, and they must be relatively small to localize at the target site quickly and any unbound molecules able to clear from the rapidly body before clinical assessment. Whole antibodies have been used for this purpose, but smaller antibody fragments and peptides are preferred because of their ability to penetrate tumors more quickly while showing a lower circulation half-life that sometimes results in images with improved contrast [354]. Among those that have received FDA approval are capromab pendetide (Prostascint) an indium-111–labeled anti–prostate-specific antigen (PSA) antibody able to recognize intracellular PSA that has been used for the detection of prostate cancer. Apcitude (Acutect)

technetium-labeled synthetic peptide that binds to glycoprotein IIb/IIIa receptors on activated platelets for the imaging of deep vein thrombosis, and Imciromabpentetate (Myoscint), an indium-111–labeled antibody specific for human cardiac myosin used to detect the presence and location of myocardial injury in patients with suspected myocardial infarction [11].

## 9.6 FUTURE DIRECTIONS

There will continue to be many opportunities to improve protein therapeutics. New formulations, such as highly concentrated dosage forms for subcutaneous administration, may require modification to decrease propensities for aggregation. Multifunctional proteins, such as bispecific antibodies are being developed and modifications to optimize each function may become important. Certainly, as non-endogenous proteins are developed, modifications of primary structure will be required to minimize the potential to elicit anti-drug antibodies. Prolonging and optimizing protein half-life is increasingly being recognized as a key property in development. Conjugation of polymers, drugs, and detection moieties will continue because some of the current chemical methods can be made more efficient. Cost of manufacture has not usually been considered a priority, but innovations being made in conjugation modifications so that they can be made earlier in downstream processing may become important. Protein modifications will continue to be developed to optimize formulation properties including maintaining protein stability, utilizing high concentrations and minimizing viscosity. As with all medicines, safety is paramount, and modifications will continue to be developed to make proteins more safe for use in the clinic.

## ACKNOWLEDGMENTS

SB is grateful for funding from NIHR Biomedical Research Centre at Moorfields Hospital and the UCL Institute of Ophthalmology, Moorfields Special Trustees, the Helen Hamlyn Trust (in memory of Paul Hamlyn), Fight for Sight, and Freemasons Grand Charity. SB is also grateful for funding from the UK Engineering & Physical Sciences Research Council (EPSRC) for the EPSRC Centre for Innovative Manufacturing in Emergent Macromolecular Therapies. Financial support from the consortium of industrial and governmental users for the EPSRC Centre is also acknowledged.

## REFERENCES

1. Uy R. Wold F. Posttranslational covalent modification of proteins. Science 1977;198 (4320):890–896.
2. Wold F. *In vivo* chemical modification of proteins (post-translational modification). Annu Rev Biochem 1981;50:783–814.
3. Walsh G, Jefferis R. Post-translational modifications in the context of therapeutic proteins. Nat Biotechnol 2006;24(10):1241–1252.

4. Van Kasteren S. Synthesis of post-translationally modified proteins. Biochem Soc Trans 2012;40(5):929–944.
5. Marshall SA, Lazar GA, Chirino AJ, Desjarlais JR. Rational design and engineering of therapeutic proteins. Drug Discov Today 2003;8(5):212–221.
6. Descotes J, Gouraud A. Clinical immunotoxicity of therapeutic proteins. Expert Opin Drug Metab Toxicol 2008;4(12):1537–1549.
7. Jelkmann W. Biosimilar recombinant human erythropoietins ("epoetins") and future erythropoiesis-stimulating treatments. Expert Opin Biol Ther 2012;12(5):581–592.
8. Mancini S, Amorotti E, Vecchio S, Ponz De Leon M, Roncucci L. Infliximab-related hepatitis: discussion of a case and review of the literature. Intern Emerg Med 2010;5(3): 193–200.
9. Giezen T, Mantel-Teeuwisse A, Straus S, Schellekens H, Leufkens H, Egberts A. Safety-related regulatory actions for biologicals approved in the United States and the European Union. JAMA 2008;300(16):1887–1896.
10. Kontermann RE. Half-life modulating strategies—an introduction. In: Kontermann R, editor. Therapeutic Proteins: Strategies to Modulate Their Plasma Half-Lives. Weinheim, Germany: Wiley-VCH Verlag GmbH; 2012.
11. Leader B, Baca QJ, Golan DE. Protein therapeutics: a summary and pharmacological classification. Nat Rev Drug Discov 2008;7(1):21–39.
12. Wurch T, Pierré A, Depil S. Novel protein scaffolds as emerging therapeutic proteins: from discovery to clinical proof-of-concept. Trends Biotechnol 2012;30(11):575–582.
13. Shaw E. Selective chemical modification of proteins. Physiol Rev 1970;50(2):244–296.
14. Glazer AN. Specific chemical modification of proteins. Annu Rev Biochem 1970;39: 101–130.
15. Means GE, Feeney RE. Chemical modifications of proteins: history and applications. Bioconjug Chem 1990;1(1):2–12.
16. Basie E, Joubert N, Pucheault M. Protein chemical modification on endogenous amino acids. Chem Biol 2010;17(3):213–227.
17. Stephanopoulos N, Francis MB. Choosing an effective protein bioconjugation strategy. Nat Chem Biol 2011;7(12):876–884.
18. Veronese FM, Morpurgo M. Bioconjugation in pharmaceutical chemistry. Farmaco 1999;54(8):497–516.
19. Scott DW, De Groot AS. Can we prevent immunogenicity of human protein drugs? Ann Rheum Dis 2009;69(Suppl 1):i72–i76.
20. Sethu S, Govindappa K, Alhaidari M, Pirmohamed M, Park K, Sathish J. Immunogenicity to biologics: mechanisms, prediction and reduction. Arch Immunol Ther Exp (Warsz) 2012;60(5):331–344.
21. Joubert MK, Hokom M, Eakin C, Zhou L, Deshpande M, Baker MP, et al. Highly aggregated antibody therapeutics can enhance the in vitro innate and late-stage T-cell immune responses. J Biol Chem 2012;287(30):25266–25279.
22. Wang W, Singh SK, Li N, Toler MR, King KR, Nema S. Immunogenicity of protein aggregates—concerns and realities. Int J Pharm 2012;431(1-2):1–11.
23. De Groot AS, Scott DW. Immunogenicity of protein therapeutics. Trends Immunol 2007;28(11):482–490.

24. Duncan R, Gaspar R. Nanomedicine(s) under the microscope. Mol Pharm 2011;8(6): 2101–2141.
25. Duncan R, Richardson SCW. Endocytosis and intracellular trafficking as gateways for nanomedicine delivery: opportunities and challenges. Mol Pharm 2012;9(9):2380–2402.
26. Pollaro L, Heinis C. Strategies to prolong the plasma residence time of peptide drugs. MedChemComm 2010;1(5):319.
27. Kontos S, Hubbell JA. Drug development: longer-lived proteins. Chem Soc Rev 2012; 41(7):2686.
28. Sola RJ, Grievenow K. Glycosylation of therapeutic proteins an effective strategy to optimize efficacy. Biodrugs 2010;24(1):9–21.
29. Elliott S, Lorenzini T, Asher S, Aoki K, Brankow D, Buck L, et al. Enhancement of therapeutic protein *in vivo* activities through glycoengineering. Nat Biotechnol 2003; 21(4):414–421.
30. Chen X, Muthoosamy K, Pfisterer A, Neumann, Weil T. Site-selective lysine modification of native proteins and peptides via kinetically controlled labeling. Bioconjug Chem 2012;23:500–508.
31. Styslinger TJ, Zhang N, Bhatt VS, Pettit N, Palmer AF, Wang PG. Site-selective glycosylation of hemoglobin with variable molecular weight oligosaccharides: Potential alternative to PEGylation. J Am Chem Soc 2012;134(17):7507–7515.
32. Wang LX, Lomino JV. Emerging technologies for making glycan-defined glycoproteins. ACS Chem Biol 2012;7(1):110–22.
33. Jevsevar S, Kunstelj M, Porekar VG. PEGylation of therapeutic proteins. Biotechnol J 2010;5(1):113–128.
34. Hamidi M, Rafiei P, Azadi A. Designing PEGylated therapeutic molecules: advantages in ADMET properties. Expert Opin Drug Discov 2008;3(11):1293–1307.
35. Payne RW, Murphy BM, Manning MC. Product development issues for PEGylated proteins. Pharm Dev Technol 2011;16(5):423–440.
36. Pisal DS, Kosloski MP, Balu-Iyer SV. Delivery of therapeutic proteins. J Pharm Sci 2010;99(6):2557–2575.
37. Veronese F. Peptide and protein PEGylation: a review of problems and solutions. Biomaterials 2001;22(5):405–417.
38. Bailon P, Won C. PEG-modified biopharmaceuticals. Expert Opin Drug Deliv 2009;6(1): 1–16.
39. Kang J, Deluca P, Lee K. Emerging PEGylated drugs. Expert Opin Emerg Drugs 2009; 14(2):363–380.
40. Kim H, Robinson SB, Csaky KG. FcRn receptor-mediated pharmacokinetics of therapeutic IgG in the eye. Mol Vis 2009;15:2803–2812.
41. Andersen JT, Terje J, Sandlie I. The versatile MHC class I-related FcRn protects IgG and albumin from degradation: implications for development of new diagnostics and therapeutics. Drug Metabol Pharmacokinet 2009;24(4):318–332.
42. Dennis MS, Zhang M, Meng Y, Kadkhodayan M, Kirchhofer D, Combs D, et al. Albumin binding as a general strategy for improving the pharmacokinetics of proteins. J Biol Chem 2002;277(38):35035–35043.
43. Otagiri M, Chuang VTG. Pharmaceutically important pre-and posttranslational modifications on human serum albumin. Biol Pharm Bull 2009;32(4):527–534.

44. Stork R, Campigna E, Robert B, Muller D, Kontermann RE. Biodistribution of a bispecific single-chain diabody and its half-life extended derivatives. J Biol Chem 2009;284(38):25612–25619.
45. Trussel S, Dumelin C, Frey K, Villa A, Buller F, Neri D. 2009. New strategy for the extension of the serum half-life of antibody fragments. Bioconjug Chem 20: 2286–2292.
46. Halpern W, Riccobene TA, Agostini H, Baker K, Stolow DD, Gu ML, et al. Albugranin tm, a recombinant human granulocyte colony stimulating factor (G-CSF) genetically fused to recombinant human albumin induces prolonged myelopoietic effects in mice and monkeys. Pharm Res 2002;19(11):1720–1729.
47. Subramanian GM, Fiscella M, Lamousé-Smith A, Zeuzem S, Mchutchison JG. Albinterferon α-2b: a genetic fusion protein for the treatment of chronic hepatitis C. Nat Biotechnol 2007;25(12):1411–1419.
48. Sung C, Nardelli B, Lafleur DW, Blatter E, Corcoran M, Olsen HS, et al. An IFN-beta-albumin fusion protein that displays improved pharmacokinetic and pharmacodynamic properties in nonhuman primates. J Interferon Cytokine Res 2003;23(1):25–36.
49. Mehvar R. Modulation of the pharmacokinetics and pharmacodynamics of proteins by polyethylene glycol conjugation. J Pharm Sci 2000;3(1):125–36.
50. Fishburn C. The pharmacology of PEGylation: balancing PD with PK to generate novel therapeutics. J Pharm Sci 2008;97(10):4167–4183.
51. Fee C. Size comparison between protein PEGylation with branched and linear poly (ethylene glycol) molecules. Biotechnol Bioeng 2007;98(4):725–731.
52. Fee C, Van Alstine J. Prediction of the viscosity radius and the size exclusion chromatography behavior of PEGylated proteins. Bioconjug Chem 2004;15(6):1304–1313.
53. Gokarn YR, Mclean M, Laue TM. Effect of PEGylation on protein hydrodynamics. Mol Pharm 2012;9(4):762–773.
54. Vugmeyster Y, Harrold J, Xu X. Absorption, distribution, metabolism, and excretion (ADME) studies of biotherapeutics for autoimmune and inflammatory conditions. AAPS J 2012;14:714–727.
55. Veronese FM, editor. PEGylated protein drugs: basic science and clinical applications. In: Parnham MJ, Bruinvels J, editors. Milestones in Drug Therapy. Basel: Birkhauser; 2009.
56. Mahmood I, Green M. Pharmacokinetic and pharmacodynamic considerations in the development of therapeutic proteins. Clin Pharmacokinet 2005;44(4):331–347.
57. Kalia J, Raines R. Advances in bioconjugation. Curr Organ Chem 2010;14(2):138–147.
58. Gauthier M, Klok H. Peptide/protein-polymer conjugates: synthetic strategies and design concepts. Annu Rev Phys Chem 2008;2591–2611.
59. Chen C, Constantinou A, Deonarain M. Modulating antibody pharmacokinetics using hydrophilic polymers. Expert Opin Drug Deliv 2011;8(9):1221–1236.
60. Kochendoerfer GG. Site-specific polymer modification of therapeutic proteins. Curr Opin Chem Biol 2005;9(6):555–560.
61. Gil E, Hudson S. Stimuli-responsive polymers and their bioconjugates. Prog Polym Sci 2004;29(12):1173–1222.
62. Klok H-A. Peptide/protein–synthetic polymer conjugates: quo vadis. Macromolecules 2009;42(21):7990–8000.

63. Li K, Liu B. Water-soluble conjugated polymers as the platform for protein sensors. Polym Chem 2010;1(3):252.
64. Constantinou A, Epenetos AA, Hreczuk-Hirst D, Jain S, Wright M, Chester KA, et al. Site-specific polysialylation of an antitumor single-chain Fv fragment. Bioconjug Chem 2009;20(5):924–931.
65. Gregoriadis G, Jain S, Papaioannou I, Laing P. Improving the therapeutic efficacy of peptides and proteins: a role for polysialic acids. Int J Pharm 2005;300(1–2):125–130.
66. Gregoriadis G, Fernandes A, Mital M, McCormack B. Polysialic acids: potential in improving the stability and pharmacokinetics of proteins and other therapeutics. Cell Mol Life Sci 2000;57(13):1964–1969.
67. Ryan SM, Wang X, Mantovani G, Sayers CT, Haddleton DM, Brayden DJ. Conjugation of salmon calcitonin to a combed-shaped end functionalized poly(poly(ethylene glycol) methyl ether methacrylate) yields a bioactive stable conjugate. J Control Release 2009;135:51–59.
68. Viegas TX, Bentley MD, Harris JM, Fang Z, Yoon K, et al. Polyoxazoline: chemistry, properties, and applications in drug delivery. Bioconjug Chem 2011;22(5):976–986.
69. Pasut G, Veronese FM. State of the art in PEGylation: the great versatility achieved after forty years of research. J Control Release 2012;161(2):461–472.
70. Gaber-Porekar V, Zore I, Podobnik B, Menart V. Obstacles and pitfalls in the PEGylation of therapeutic proteins. Curr Opin Drug Discov Dev 2008;11(2):242–250.
71. Jevševar S, Kunstelj M. Half-Life Extension through PEGylation. In: Kontermann R, editor. Therapeutic Proteins: Strategies to Modulate Their Plasma Half-Lives. Weinheim, Germany: Wiley-VCH Verlag GmbH & Co. KGaA 2012;41–61.
72. Fee C, Van Alstine J. PEG proteins: reaction engineering and separation issues. Chem Eng Sci 2006;61(3):924–939.
73. González-Valdez J, Rito-Palomares M, Benavides J. Advances and trends in the design, analysis, and characterization of polymer–protein conjugates for "PEGylaided" bioprocesses. Anal Bioanal Chem 2012;403(8):2225–2235.
74. Mayolo-Deloisa K, González-Valdez J, Guajardo-Flores D, Aguilar O, Benavides J, Rito-Palomares M. Current advances in the non-chromatographic fractionation and characterization of PEGylated proteins. J Chem Technol Biotechnol 2010;86(1):18–25.
75. Rebolj K, Pahovnik D, Žagar E. Characterization of a protein conjugate using an asymmetrical-flow field-flow fractionation and a size-exclusion chromatography with multi-detection system. Anal Chem 2012;84(17):7374–7383.
76. Armstrong JK. The occurrence, induction, specificity and potential effect of antibodies against poly(ethylene glycol). In: PEGylated Protein Drugs: Basic Science and Clinical Applications. Basel: Birkhäuser; 2009. p. 147–168.
77. Garay RP, El-Gewely R, Armstrong JK, Garratty G, Richette P. Antibodies against polyethylene glycol in healthy subjects and in patients treated with PEG-conjugated agents. Expert Opin Drug Deliv 2012;9(11):1319–1323.
78. Judge A, McClintock K, Phelps J, Maclachlan I. Hypersensitivity and loss of disease site targeting caused by antibody responses to PEGylated liposomes. Mol Ther 2006;13(2):328–337.
79. Kromminga A, Schellekens H. Antibodies against erythropoietin and other protein-based therapeutics: an overview. Ann NY Acad Sci 2005;1050:257–65.

80. Nechansky A, Kircheis R. Immunogenicity of therapeutics: a matter of efficacy and safety. Expert Opin Drug Discov 2010;5(11):1–13.
81. Schellekens H. Bioequivalence and the immunogenicity of biopharmaceuticals. Nat Rev Drug Discov 2002;1(6):457–462.
82. Schellekens H. Immunogenicity of therapeutic proteins: clinical implications and future prospects. Clin Ther 2002;24(11):1720–1740.
83. Singh S. Impact of product-related factors on immunogenicity of biotherapeutics. J Pharm Sci 2011;100(2):354–387.
84. Sominanda A, Lundkvist M, Fogdell-Hahn A, Hemmer B, Hartung H-P, Hillert J, et al. Inhibition of endogenous interferon beta by neutralizing antibodies against recombinant interferon beta. Arch Neurol 2010;67(9):1095–1101.
85. Schellekens H, Hennink WE, Brinks V. The immunogenicity of polyethylene glycol: facts and fiction. Pharm Res 2013;30:1729–1734.
86. Webster R, Didier E, Harris P, Siegel N, Stadler J, Tilbury L, et al. PEGylated proteins: evaluation of their safety in the absence of definitive metabolism studies. Drug Metabol Dispos 2007;35(1):9–16.
87. Caliceti P, Veronese F. Pharmacokinetic and biodistribution properties of poly(ethylene glycol)–protein conjugates. Adv Drug Deliv Rev 2003;55(10):1261–1277.
88. Fried M, Shiffman M, Reddy K, Smith C, Marinos G, Goncales F, et al. PEG-interferon alfa-2a plus ribavirin for chronic hepatitis C virus infection. N Engl J Med 2002;34: 975–982.
89. Hadziyannis S. PEG-interferon-alpha2a and ribavirin combination therapy in chronic hepatitis C: a randomized study of treatment duration and ribavirin dose. Ann Intern Med 2004;140:346–355.
90. Talal A, AL. Pharmacodynamics of PEG-IFN alpha differentiate HIV/HCV coinfected sustained virological responders from nonresponders. Hepatology 2006;43: 943–953.
91. Hamidi M, Azadi A, Rafiei P. Pharmacokinetic consequences of PEGylation. Drug Deliv 2006;13(6):399–409.
92. Kubetzko S, Sarkar CA, Plückthun A. Protein PEGylation decreases observed target association rates via a dual blocking mechanism. Mol Pharmacol 2005;68(5): 1439–1454.
93. Mabry R, Rani M, Geiger R, Hubbard GB, Carrion R, Brasky K, et al. Passive protection against anthrax by using a high-affinity antitoxin antibody fragment lacking an Fc region. Infect Immun 2005;73(12):8362–8368.
94. Khalili H, Godwin A, Choi J, Lever R, Brocchini S. Comparative binding of disulfide-bridged PEG-Fabs. 2012;23(11):2262–2277.
95. Youngster S, Wang YS, Grace M, Bausch J, Bordens R, Wyss DF. Structure, biology, and therapeutic implications of PEGylated interferon alpha-2b. Curr Pharm Des 2002;8(24):2139–2157.
96. Pettit D, Bonnert T, Eisenman J, Srinivasan S, Paxton R, Beers C, et al. Structure-function studies of interleukin-15 using site-specific mutagenesis, poly(ethylene glycol) conjugation, and homology modeling. J Biol Chem 1997;272(4):2312.
97. Wang Y, Youngster S, Bausch J, Zhang R, Mcnemar C, Wyss D. Identification of the major positional isomer of PEGylated interferon alpha-2b. Biochemistry 2000;39: 10634–10640.

98. Wang Y, Youngster S, Grace M, Bausch J, Wyss D. Structural and biological characterization of PEGylated recombinant interferon alpha-2b and its therapeutic implications. Adv Drug Del Rev 2002;54:547–570.
99. Foser S, Schacher A, Weyer K, Brugger D, Dietel E, Marti S, et al. Isolation, structural characterization, and antiviral activity of positional isomers of monoPEGylated interferon α-2a (PEGASYS). Protein Expres Purif 2003;30:78–87.
100. Dhalluin C, Ross A, Leuthold L, Foser S, Gsell B, Muller F, et al. Structural and biophysical characterization of the 40 kDa PEG-interferon-a2a and its individual positional isomers. Bioconjug Chem 2005;16:504–517.
101. Gonnelli M, Strambini GB. No effect of covalently linked poly(ethylene glycol) chains on protein internal dynamics. Biochim Biophys Acta 2009;1794(3):569–576.
102. Rodríguez-Martínez J, Solá R, Castillo B, Cintrón-Colón H, Rivera-Rivera I, Barletta G, et al. Stabilization of α-chymotrypsin upon PEGylation correlates with reduced structural dynamics. Biotechnol Bioeng 2008;101(6):1142–1149.
103. Lu Y, Harding SE, Turner A, Smith B, Athwal DS. Effect of PEGylation on the solution conformation of antibody fragments. J Pharm Sci 2008;97(6):2062–2079.
104. Foster GR. PEGylated interferons for the treatment of chronic hepatitis C: pharmacological and clinical differences between peginterferon-2a and peginterferon-2b. Drugs 2010;70(2):147–165.
105. Birerdinc A, Younossi ZM. Emerging therapies for hepatitis C virus. Expert Opin Emerg Drugs 2010;15(4):535–544.
106. Munir S, Saleem S, Idrees M, Tariq A, Butt S, Rauff B, et al. Hepatitis C treatment: current and future perspectives. Virol J 2010;7(1):296–301.
107. Vezali E, Aghemo A, Colombo M. Interferon in the treatment of chronic hepatitis C: a drug caught between past and future. Expert Opin Biol Ther 2011(3):301–313.
108. Petta S, Craxì A. Optimal therapy in hepatitis C virus genotypes 2 and 3 patients. Liver Int 2011;31:36–44.
109. Bailon P, Palleroni A, Schaffer C, Spence C, Fung W, Porter JE, et al. Rational design of a potent, long-lasting form of interferon: a 40 kDa branched polyethylene glycol-conjugated interferon α-2a for the treatment of hepatitis C. Bioconjug Chem 2001; 12:195–202.
110. Mooler I, Thomas A, Geyer H, Schanzer W, Thevis M. Synthesis, characterisation, and mass spectrometric detection of a PEGylated EPO-mimetic peptide for sports drug testing purpose. Rapid Commun Mass Spectrom 2011;25:2115–2123.
111. European Medicines Agency. Withdrawal of the marketing authorisation application for Omontys (peginesatide). EMA/384566/2013;2013.
112. Katre NV, Knauf MJ, Laird WJ. Chemical modification of recombinant interleukin 2 by polyethylene glycol increases its potency in the murine Meth A sarcoma model. Proc Natl Acad Sci USA 1987;86(6):1487–1491.
113. Inoue H, Kadoya T, Kabaya K, Tachibana K, Nishi N, Sato M, et al. A highly enhanced thrombopoietic activity by monomethoxy polyethylene glycol-modified recombinant human interleukin-6. Lab Clin Med 1994;124(4):529–536.
114. Tsutsumi Y, Kihira T, Yamamoto S, Kubo K, Nakagawa S, Miyake M, et al. Chemical modification of natural human tumor necrosis factor-alpha with polyethylene glycol increases its anti-tumor potency. Jpn J Cancer Res 1994;85(1):9–12.

115. Shaunak S, Godwin A, Choi J, Balan S, Pedone E, Vijayarangam D, et al. Site-specific PEGylation of native disulfide bonds in therapeutic proteins. Nat Chem Biol 2006;2 (6):312–313.
116. Caserman S, Kusterle M, Milunović T, Schiefermeier M, Jevsevar S, Porekar V. Correlations between *in vitro* potency of PEG-protein conjugates and their chromatographic behavior. Anal Biochem 2009;389:27–31.
117. Cho H, Daniel T, Buechler YJ, Litzinger DC, Maio Z, Putnam AMH, et al. Optimized clinical performance of growth hormone with an expanded genetic code. Proc Natl Acad Sci USA 2011;108(22):9060–9065.
118. Cong Y, Pawlisz E, Bryant P, Balan S, Laurine E, Tommasi R, et al. Site-specific PEGylation at histidine tags. Bioconjug Chem 2012;23(2):248–263.
119. Cao Z, Jiang S. Super-hydrophilic zwitterionic poly(carboxybetaine) and amphiphilic non-ionic poly(ethylene glycol) for stealth nanoparticles. Nano Today 2012;7(5): 404–413.
120. Duncan R, Spreafico F. Polymer conjugates: pharmacokinetic considerations for design and development. Clin Pharmacokinet 1994;27:290–306.
121. Abuchowski A, ES T, Palczuk NC, Davis FF. Alteration of immunological properties of bovine serum albumin by covalent attachment of polyethylene glycol. J Biol Chem 1977;252(11):3578–3581.
122. Young M, Malavalli A, Winslow N, Candegriff K, Winslow R. Toxicity and hemodynamic effects after single dose administration of MalPEG-hemoglobin (MP4) in rhesus monkeys. Translational Res 2007;149(6):333–342.
123. Bendele A, Seely J, Richey C, Sennello G, Shopp G. Short communication: renal tubular vacuolation in animals treated with polyethylene-glycol-conjugated proteins. Toxicol Sci 1998;42(2):152–157.
124. Sandborn W, Feagan B, Stoinov S, Honiball P, Rutgeerts P, Mason D, et al. Certolizumab pegol for the treatment of Crohn's disease. N Engl J Med 2007;357(3):228–238.
125. Cimzia (certolizumab pegol) package insert medication guide. UCB, Inc. Symrna, GA, USA; 2012.
126. Parkinson C, Scarlett J, Trainer P. Pegvisomant in the treatment of acromegaly. Adv Drug Del Rev 2003;55(10):1303–1314.
127. Yin D, Vreeland F, Schaaf L, Millham R, Duncan B, Sharma A. Clinical pharmacodynamic effects of the growth hormone receptor antagonist pegvisomant: implications for cancer therapy. Clin Cancer Res 2007;13(3):1000–1009.
128. Levy Y, Hershfield MS, Fernandez-Mejia C, Polmar SH, Scudiery D, Berger M, et al. Adenosine deaminase deficiency with late onset of recurrent infections: response to treatment with polyethylene glycol-modified adenosine deaminase. J Pediatr 1988;113(2):312–317.
129. Abuchowski A, Kazo G, Verhoest CJ, Van Es T, Kafkewitz D, Nucci M, et al. Cancer therapy with chemically modified enzymes. I. Antitumor properties of polyethylene glycol-asparaginase conjugates. Cancer Biochem Biophys 1984;7(2):175–186.
130. Kodera Y, Sekine T, Yasukohchi T, Kiriu Y, Hiroto M, Matsushima A, Inada Y. Stabilization of L-Asparaginase Modified with Comb-Shaped Poly(ethylene glycol) Derivatives, in vivo and in vitro. Bioconjugate Chem 1994;5(4):283–286.

131. Sherman MR, Saifer MGP, Ruiz FP. PEG-uricase in the management of treatment-resistant gout and hyperuricemia. Adv Drug Del Rev 2008;60:59–68.
132. Lewis A, Tang Y, Brocchini S, Choi J, Godwin A. Poly(2-methacryloyloxyethyl phosphorylcholine) for protein conjugation. Bioconjug Chem 2008;19:2144–2155.
133. Luxon BA, Grace M, Brassard D, Bordens R. PEGylated interferons for the treatment of chronic hepatitis C infection. Clin Ther 2002;24(9):1363–1383.
134. Veronese F, Pasut G. PEGylation, successful approach to drug delivery. Drug Discovery Today 2005;10(21):1451–1458.
135. Hershfield M, Chaffee S. Use of site-directed mutagenesis to enhance the epitope-shielding effect of covalent modification of proteins with polyethylene glycol. Proc Natl Acad Sci USA 1991;88(16):7185–7189.
136. Yamamoto Y, Tsutsumi Y, Yoshioka Y, Nishibata T, Kobayashi K, Okamoto T, et al. Site-specific PEGylation of a lysine-deficient TNF-alpha with full bioactivity. Nat Biotechnol 2003;21:546–552.
137. Roberts MJ, Bentley MD, Harris JM. Chemistry for peptide and protein PEGylation. Adv Drug Deliv Rev 2002;54:459–476.
138. Harris J, Chess R. Effect of PEGylation on pharmaceuticals. Nat Rev Drug Dis 2003;2 (3):214–221.
139. Zalipsky S, Seltzer R, Menon-Rudolph S. Evaluation of a new reagent for covalent attachment of polyethylene glycol to proteins. Biotechnol Appl Biochem 1992;15(1):100–114.
140. Chapman AP. PEGylated antibodies and antibody fragments for improved therapy: a review. Adv Drug Deliv Rev 2002;54:531–545.
141. Rosendahl MS, Doherty DH, Smith SJ, Carlson SJ, Chlipala EA, Cox GN. A long-acting, highly potent interferon α-2 conjugate created using site-specific PEGylation. Bioconjug Chem 2005;16:200–207.
142. Bloom L, Calabro V. Fn3: a new protein scaffold reaches the clinic. Drug Discov Today 2009;14:949–955.
143. Lipovsek D. Adnectins: engineered target-binding protein therapeutics. Protein Eng 2011;24(1-2):3–9.
144. Alley SC, Benjamin DR, Jeffrey SC, Okeley NM, Meyer DL, Sanderson RJ, et al. Contribution of linker stability to the activities of anticancer immunoconjugates. Bioconjug Chem 2008;19:759–765.
145. Shen BQ, Xu K, Liu L, Raab H, Bhakta S, Kenrick M, et al. Conjugation site modulates the *in vivo* stability and therapeutic activity of antibody-drug conjugates. Nat Biotechnol 2012;30(2):184–189.
146. Godwin A. Novel reagents and method for conjugating biological molecules. PCT WO/2010/010324;2010.
147. Brocchini S, Balan S, Godwin A, Choi JW, Zloh M, Shaunaik S. PEGylation of native disulfide bonds in proteins. Nat Protoc 2006;1(5):2241–2252.
148. Brocchini S, Godwin A, Balan S, Choi Jw, Zloh M, Shaunak S. Disulfide bridge based PEGylation of proteins. Adv Drug Deliv Rev 2008;60(1):3–12.
149. Bolanos-Garcia V, Davies O. Structural analysis and classification of native proteins from *E. coli* commonly co-purified by immobilised metal affinity chromatography. Biochim Biophys Acta 2006;1760(9):1304–1313.

150. Knecht S, Ricklin D, Eberle AN, Ernst B. Oligohis-tags: mechanisms of binding to $Ni^{2+}$-NTA surfaces. J Mol Recognit 2009;22(4):270–279.
151. Gauthier MA, Klok HA. Arginine-specific modification of proteins with polyethylene glycol. Biomacromolecules 2010;12:482–493.
152. Sato H. Enzymatic procedure for site-specific PEGylation of proteins. Adv Drug Deliv Rev 2002;54:487–504.
153. Rabuka D. Chemoenzymatic methods for site-specific protein modification. Curr Opin Chem Biol 2010;14:790–796.
154. Kramer JR and Deming TJ. Preparation of multifunctional and multireactive polypeptides via methionine alkylation. Biomacromolecules 2012;13(6):1719–1723.
155. Karpf DM, Sørensen BB, Hermit MB, Holmberg HL, Tranholm M, Bysted BV, Stennicke HR. Prolonged half-life of glycoPEGylated rFVIIa variants compared to native rFVIIa. Thrombosis Research 2011;128(2):191–195.
156. Plesner B, Westh P, Nielsen A. Biophysical characterisation of glycoPEGylated recombinant human factor VIIa. Int J Pharm 2011;406:62–68.
157. Stennicke HR, Ostergaard H, Bayer RJ, Kalo MS, Kinealy K, Holm PK, et al. Generation and biochemical characterization of glycoPEGylated factor VIIIa derivatives. Thromb Haemost 2008;100(5):920–928.
158. Yi L, Sun H, Wu Y-W, Triola G, Waldmann H, Goody RS. A highly efficient strategy for modification of proteins at the C terminus. Angew Chem Int Ed Engl 2010;49(49):9417–9421.
159. Gault VA, Kerr BD, Irwin N, Flatt PR. C-terminal mini-PEGylation of glucose-dependent insulinotropic polypeptide exhibits metabolic stability and improved glucose homeostasis in dietary-induced diabetes. Biochem Pharmacol 2008;75 (12):2325–2333.
160. Rajan RS, Li T, Aras M, Sloey C, Sutherland W, Arai H, et al. Modulation of protein aggregation by polyethylene glycol conjugation: GCSF as a case study. Protein Sci 2006;15(5):1063–1075.
161. Basu A, Yang K, Wang M, Liu S, Chintala R, Palm T, et al. Structure-function engineering of interferon-β-1b for improving stability, solubility, potency, immunogenicity, and pharmacokinetic properties by site-selective mono-PEGylation. Bioconjug Chem 2006;17(3):618–630.
162. Gaertner HF, Offord RE. Site-specific attachment of functionalized poly(ethylene glycol) to the amino terminus of proteins. Bioconjug Chem 1996;7(1):38–44.
163. Na DH, Lee KC, Deluca P.P. PEGylation of octreotide: II. Effect of N-terminal mono-PEGylation on biological activity and pharmacokinetics. Pharm Res 2005;22(5): 743–749.
164. Nie Y, Zhang X, Wang X, Chen J. Preparation and stability of N-terminal mono-PEGylated recombinant human endostatin. Bioconjug Chem 2006;17(4):995–999.
165. Lee H, Jang IH, Ryu SH, Park TG. N-terminal site-specific mono-PEGylation of epidermal growth factor. Pharm Res 2003;20(5):818–825.
166. Gilmore JM, Scheck RA, Esser-Kahn AP, Joshi NS, Francis MB. N-terminal protein modification through a biomimetic transamination reaction. Angew Chem Int Ed 2006;45(32):5307–5311.
167. Lee B, Kwon J, Kim H, Yamamoto S. Solid-phase PEGylation of recombinant interferon α-2a for site-specific modification: process performance, characterization, and *in vitro* bioactivity. Bioconjug Chem 2007;18(6):1728–1734.

168. Thom J, Anderson D, McGregor J, Cotton G. Recombinant protein hydrazides: application to site-specific protein PEGylation. Bioconjug Chem 2011;22(6):1017–1020.
169. Kinstler O, Molineux G, Treuheit M, Ladd D, Gegg C. Mono-N-terminal poly(ethylene glycol)-protein conjugates. Adv. Drug Deliv Rev 2002;54(4):477–485.
170. Wang J, Hu T, Liu Y, Zhang G, Ma G, Su Z. Kinetic and stoichiometric analysis of the modification process for N-terminal PEGylation of staphylokinase. Anal Biochem 2011;412(1):114–116.
171. Tong Y, Zhong K, Tian H, Gao X, Xu X, et al. Characterization of a monoPEG20000-Endostar. Int J Biol Macromol 2010;46(3):331–336.
172. Cindric M, Cepo T, Galic N, Bukvic-Krajacic M, Tomczyk N, Vissers JPC, et al. Structural characterization of PEGylated rHuG-CSF and location of PEG attachment sites. J Pharmaceut Biomed 2007;44(2):388–395.
173. Puchkov IA, Kononova NV, Bobruskin AI, Bairamashvili DI, Mart'yanov VA, Shuster AM. Recombinant granulocyte colony-stimulating factor (filgrastim): optimization of conjugation conditions with polyethylene glycol. Bioorg Khim 2012;38(5):479–487.
174. Wu L, Ho SV, Wang W, Gao J, Zhang G, Su Z, & Hu T. N-terminal mono-PEGylation of growth hormone antagonist: Correlation of PEG size and pharmacodynamic behavior. Int J Pharm 2013;453(2):533–540.
175. Fontana A, Spolaore B, Mero A, Veronese F. Site-specific modification and PEGylation of pharmaceutical proteins mediated by transglutaminase. Adv Drug Deliv Rev 2008;60:13–28.
176. Defrees S, Wang ZG, Xing R, Scott AE, Wang J, Zopf D, et al. GlycoPEGylation of recombinant therapeutic proteins produced in *Escherichia coli*. Glycobiology 2006;16(9):833–843.
177. Popp MW, Dougan SK, Chuang T-Y, Spooner E, Ploegh HL. Sortase-catalyzed transformations that improve the properties of cytokines. Proc Natl Acad Sci USA 2011;108(8):3169–3174.
178. Deiters A, Cropp T, Summerer D, Mukherji M, Schultz P. Site-specific PEGylation of proteins containing unnatural amino acids. Bioorg Med Chem Lett 2004;14(23):5743–5745.
179. Johnson JA, Lu YY, Van Deventer JA, Tirrell DA. Residue-specific incorporation of non-canonical amino acids into proteins: recent developments and applications. Curr Opin Chem Biol 2010;14:774–780.
180. Liu CC, Schultz PG. Adding new chemistries to the genetic code. Annu Rev Biochem 2010;79:413–444.
181. Voloshchuk N, Montclare JK. Incorporation of unnatural amino acids for synthetic biology. Mol Biosyst 2009;6(1):65–80.
182. Connor RE, Tirrell DA. Non-canonical amino acids in protein polymer design. Polym Rev 2007;47(1):9–28.
183. Carrico I, Carlson B, Bertozzi C. Introducing genetically encoded aldehydes into proteins. Nat Chem Biol 2007;3(6):321–322.
184. Filipovska A, Rackham O. Building a parallel metabolism within the cell. ACS Chem Biol 2008;3(1):51–63.
185. Sola RJ, Griebenow K. Effects of glycosylation on the stability of protein pharmaceuticals. J Pharm Sci 2009;98(4):1223–1245.

186. Carter CR, Whitmore KM, Thorpe R. The significance of carbohydrates on G-CSF: differential sensitivity of G-CSFs to human neutrophil elastase degradation. J Leukoc Biol 2004;75(3):515–22.
187. Sareneva T, Pirhonen J, Cantell K, Julkunen I. N-glycosylation of human interferon-gamma: glycans at Asn-25 are critical for protease resistance. Biochem J 1995;308(Pt 1):9–14.
188. Sinclair AM, Elliott S. Glycoengineering: the effect of glycosylation on the properties of therapeutic proteins. J Pharm Sci 2005;94(8):1626–1635.
189. Li H, d'Anjou M. Pharmacological significance of glycosylation in therapeutic proteins. Curr Opin Biotechnol 2009;20:678–684.
190. Jain K, Kesharwani P, Gupta U, Jain NK. A review of glycosylated carriers for drug delivery. Biomaterials 2012;33(16):4166–4186.
191. Werner RG, Kopp K, Schlueter M. Glycosylation of therapeutic proteins in different production systems. Acta Paediatr Suppl 2007;96(455):17–22.
192. Hossler P, Mulukutla BC, Hu WS. Systems analysis of N-glycan processing in mammalian cells. PLoS One 2007;2(8):e713.
193. Hossler P, Khattak SF, Li ZJ. Optimal and consistent protein glycosylation in mammalian cell culture. Glycobiology 2009;19(9):936–949.
194. Schiestl M, Stangler T, Torella C, Cepeljnik T, Toll H, Grau R. Acceptable changes in quality attributes of glycosylated biopharmaceuticals. Nat Biotech 2011;29 (4):310–312.
195. Koury MJ. Sugar coating extends half-lives and improves effectiveness of cytokine hormones. Trends Biotechnol 2003;21(11):462–464.
196. Egrie JC, Browne JK. Development and characterization of novel erythropoiesis stimulating protein (NESP). Br J Cancer 2001;84(Suppl 1):3–10.
197. Kiss Z, Elliott S, Jedynasty K, Tesar V, Szegedi J. Discovery and basic pharmacology of erythropoiesis-stimulating agents (ESAS), including the hyperglycosylated ESA, darbepoetin alfa: an update of the rationale and clinical impact. Eur J Clin Pharmacol 2010;66(4):331–340.
198. Gerngross TU. Advances in the production of human therapeutic proteins in yeasts and filamentous fungi. Nat Biotech 2004;22(11):1409–1414.
199. Daly R, Hearn MTW. Expression of heterologous proteins in *Pichia pastoris*: a useful experimental tool in protein engineering and production. J Mol Recognit 2005;18(2):119–138.
200. Gerngross T. Production of complex human glycoproteins in yeast In: Axford J, editor. Glycobiology and Medicine. Netherlands: Springer; 2005, p. 139.
201. Wildt S, Gerngross TU. The humanization of N-glycosylation pathways in yeast. Nat Rev Microbiol 2005;3(2):119–128.
202. Ghaderi D, et al. Production platforms for biotherapeutic glycoproteins. Occurrence, impact, and challenges of non-human sialylation. Biotechnol Genet Eng Rev 2012;28: 147–175.
203. Chen M, Liu X, Wang Z, Song J, Qi Q, Wang PG. Modification of plant N-glycans processing: the future of producing therapeutic protein by transgenic plants. Med Res Rev 2005;25(3):343–360.
204. Whitelam GC. The production of recombinant proteins in plants. J Sci Food Agric 1995;68(1):1–9.

205. Gomord V, Faye LC. Posttranslational modification of therapeutic proteins in plants. Curr Opin Plant Biol 2004;7(2):171–181.

206. Cabanes-Macheteau M, Fitchette-Laine AC, Loutelier-Bourhis C, Lange C, Vine ND, Ma JK, et al. N-glycosylation of a mouse IgG expressed in transgenic tobacco plants. Glycobiology 1999;9(4):365–72.

207. Chargelegue D, Vine ND, Van Dolleweerd CJ, Drake PMW, Ma JKC. A murine monoclonal antibody produced in transgenic plants with plant-specific glycans is not immunogenic in mice. Transgen Res 2000;9(3):187–194.

208. Chalker JM, Bernardes GJ, Davis BG. A "tag-and-modify" approach to site-selective protein modification. Acc Chem Res 201144(9):730–741.

209. Gamblin DP, Scanlan EM, Davis BG. Glycoprotein synthesis: an update. Chem Rev 2009;109(1):131–163.

210. Macmillan D, Bill RM, Sage KA, Fern D, Flitsch SL. Selective in vitro glycosylation of recombinant proteins: semi-synthesis of novel homogeneous glycoforms of human erythropoietin. Chem Biol 2001;8(2):133–145.

211. Hirano K, Macmillan D, Tezuka K, Tsuji T, Kajihara Y. Design and synthesis of a homogeneous erythropoietin analogue with two human complex-type sialyloligosaccharides: combined use of chemical and bacterial protein expression methods. Angew Chem Int Ed 2009;48(50):9557–9560.

212. Jazayeri JA, Carroll GJ. Half-Life Extension by Fusion to the Fc Region. In: Kontermann R, editor. Therapeutic Proteins: Strategies to Modulate Their Plasma Half-Lives. Weinheim, Germany: Wiley-VCH Verlag GmbH & Co. KGaA 2012;157–188.

213. Chang C, Gupta P, Goldenberg D. Advances and challenges in developing cytokine fusion proteins as improved therapeutics. Expert Opin Drug Discov 2009;4(2):181–194.

214. Schrama D, Reisfeld RA, Becker JC. Antibody targeted drugs as cancer therapeutics. Nat Rev Drug Discov 2006;5(2):147–59.

215. Stoddart CA, Nault G, Galkina SA, Thibaudeau K, Bakis P, Bousquet-Gagnon N, et al. Albumin-conjugated C34 peptide HIV-1 fusion inhibitor: equipotent to C34 and T-20 *in vitro* with sustained activity in SCID-hu Thy/Liv mice. J Biol Chem 2008;283(49):34045–34052.

216. Goldenberg MM. Etanercept, a novel drug for the treatment of patients with severe, active rheumatoid arthritis. Clin Ther 1999;21(1):75–87.

217. Strober BE, Menon K. Alefacept for the treatment of psoriasis and other dermatologic diseases. Dermatol Ther 2007;20(4):270–276.

218. Korhonen R, Moilanen E. Abatacept, a novel CD80/86-CD28 T cell co-stimulation modulator, in the treatment of rheumatoid arthritis. Basic Clin Pharmacol Toxicol 2009;104(4):276–284.

219. Molineux G. The development of romiplostim for patients with immune thrombocytopenia. Ann NY Acad Sci 2011;1222:55–63.

220. Kapur S, Bonk ME. Rilonacept (Arcalyst), an interleukin-1 trap for the treatment of cryopyrin-associated periodic syndromes. P T 2009;34(3):138–141.

221. Terkeltaub R, Sundy JS, Schumacher HR, Murphy F, Bookbinder S, Biedermann S, et al. The interleukin 1 inhibitor rilonacept in treatment of chronic gouty arthritis: results of a placebo-controlled, monosequence crossover, non-randomised, single-blind pilot study. Ann Rheum Dis 2009;68(10):1613–1617.

222. Vincenti F, Dritselis A, Kirkpatrick P. Belatacept. Nat Rev Drug Discov 2011;10(9):655–656.

223. Walsh G. Biopharmaceutical benchmarks 2010. Nat Biotechnol 2010;28(9):917–924.

224. Huang C. Receptor–Fc fusion therapeutics, traps, and MIMETIBODY technology. Curr Opin Biotechnol 2009;20(6):692–699.
225. Suzuki T, Ishii-Watabe A, Tada M, Kobayashi T, Kanayasu-Toyoda T, Kawanishi T, et al. Importance of neonatal FcR in regulating the serum half-life of therapeutic proteins containing the Fc domain of human IgG1: a comparative study of the affinity of monoclonal antibodies and Fc-fusion proteins to human neonatal FcR. J Immunol 2010;184(4):1968–1976.
226. Chaudhury C, Brooks Cl, Carter DC, Robinson JM, Anderson CL. Albumin binding to FcRn: distinct from the FcRn-IgG interaction. Biochemistry 2006;45(15):4983–4990.
227. Yeh P, Landais D, Lemaitre M, Maury I, Crenne JY, Becquart J, et al. Design of yeast-secreted albumin derivatives for human therapy: biological and antiviral properties of a serum albumin-CD4 genetic conjugate. Proc Natl Acad Sci USA 1992;89(5):1904–1908.
228. Zeuzem S, Yoshida EM, Benhamou Y, Pianko S, Bain VG, et al. Albinterferon alfa-2b dosed every two or four weeks in interferon-naive patients with genotype 1 chronic hepatitis C. Hepatology 2008;48(2):407–417.
229. Osborn BL, Sekut L, Corcoran M, Poortman C, Sturm B, Chen G, et al. Albutropin: a growth hormone-albumin fusion with improved pharmacokinetics and pharmacodynamics in rats and monkeys. Eur J Pharmacol 2002;456(1–3):149–158.
230. Joung CH, Shin JY, Koo JK, Lim JJ, Wang JS, Lee SJ, et al. Production and characterization of long-acting recombinant human albumin-EPO fusion protein expressed in CHO cell. Protein Expr Purif 2009;68(2):137–145.
231. Yao Z, Dai W, Perry J, Brechbiel MW, Sung C. Effect of albumin fusion on the biodistribution of interleukin-2. Cancer Immunol Immunother 2004;53(5):404–410.
232. Syed S, Schuyler PD, Kulczycky M, Sheffield WP. Potent antithrombin activity and delayed clearance from the circulation characterize recombinant hirudin genetically fused to albumin. Blood 1997;89(9):3243–3252.
233. Schulte S. Use of albumin fusion technology to prolong the half-life of recombinant factor VIIa. Thromb Res 2008;122(Suppl 4):S14–S19.
234. Ikuta S, Chuang VT, Ishima Y, Nakajou K, Furukawa M, Watanabe H, et al. Albumin fusion of thioredoxin—the production and evaluation of its biological activity for potential therapeutic applications. J Control Release 2010;147(1):17–23.
235. Muller D, Karle A, Meissburger B, Hofig I, Stork R, Kontermann RE. Improved pharmacokinetics of recombinant bispecific antibody molecules by fusion to human serum albumin. J Biol Chem 2007;282(17):12650–12660.
236. Baggio LL, Huang Q, Brown TJ, Drucker DJ. A recombinant human glucagon-like peptide (GLP)-1-albumin protein (Albugon) mimics peptidergic activation of GLP-1 receptor-dependent pathways coupled with satiety, gastrointestinal motility, and glucose homeostasis. Diabetes 2004;53(9):2492–2500.
237. Duttaroy A, Kanakaraj P, Osborn BL, Schneider H, Pickeral OK, Chen C, et al. Development of a long-acting insulin analog using albumin fusion technology. Diabetes 2005;54(1):251–258.
238. Wang W, Ou Y, Shi Y. AlbuBNP, a recombinant B-type natriuretic peptide and human serum albumin fusion hormone, as a long-term therapy of congestive heart failure. Pharm Res 2004;21(11):2105–2111.
239. Ponka P, Lok CN. The transferrin receptor: role in health and disease. Int J Biochem Cell Biol 1999;31(10):1111–1137.

240. Amet N, Wang W, Shen WC. Human growth hormone-transferrin fusion protein for oral delivery in hypophysectomized rats. J Control Release 2010;141(2):177–182.
241. Kim BJ, Zhou J, Martin B, Carlson OD, Maudsley S, Greig NH, et al. Transferrin fusion technology: a novel approach to prolonging biological half-life of insulinotropic peptides. J Pharmacol Exp Ther 2010;334(3):682–692.
242. Kufer P, Lutterbuse R, Baeuerle PA. A revival of bispecific antibodies. Trends Biotechnol 2004;22(5):238–244.
243. Asano R, Nakayama M, Kawaguchi H, Kubota T, Nakanishi T, Umetsu M, et al. Construction and humanization of a functional bispecific EGFR×CD16 diabody using a refolding system. FEBS J 2011;279(2):223–233.
244. Schmidt SR. Fusion Proteins: Applications and Challenges. In: Schmidt SR, editor. Fusion Protein Technologies for Biopharmaceuticals: Applications and Challenges. Hoboken, NJ, USA: John Wiley & Sons, Inc. 2013;3–24.
245. Cordes AA, Platt CW, Carpenter JF, Randolph TW. Selective domain stabilization as a strategy to reduce fusion protein aggregation. J Pharm Sci 2012;101(4):1400–1409.
246. Lengyel JS, Stott KM, Wu X, Brooks BR, Balbo A, Schuck P, et al. Extended polypeptide linkers establish the spatial architecture of a pyruvate dehydrogenase multienzyme complex. Structure 2008;16(1):93–103.
247. Wriggers W, Chakravarty S, Jennings PA. Control of protein functional dynamics by peptide linkers. Biopolymers 2005;80(6):736–746.
248. Ma B, Tsai C-J, Haliloğlu T, Nussinov R. Dynamic allostery: linkers are not merely flexible. Structure 2011;19(7):907–917.
249. Robinson CR, Sauer RT. Optimizing the stability of single-chain proteins by linker length and composition mutagenesis. Proc Natl Acad Sci USA 1998;95(11): 5929–5934.
250. Gustavsson M, Lehtiö J, Denman S, Teeri TT, Hult K, Martinelle M. Stable linker peptides for a cellulose-binding domain–lipase fusion protein expressed in pichia pastoris. Protein Eng 2001;14(9):711–715.
251. Arai R, Ueda H, Kitayama A, Kamiya N, Nagamune T. Design of the linkers which effectively separate domains of a bifunctional fusion protein. Protein Eng 2001;14(8):529–532.
252. Le Gall F, Reusch U, Little M, Kipriyanov SM. Effect of linker sequences between the antibody variable domains on the formation, stability and biological activity of a bispecific tandem diabody. Protein Eng Des Sel 2004;17(4):357–366.
253. Fang M. Effects of interlinker sequences on the biological properties of bispecific single-chain antibodies. Chin Sci Bull 2003;48(21):2277.
254. Zhang J, Yun J, Shang Z, Zhang X, Pan B. Design and optimization of a linker for fusion protein construction. Prog Nat Sci 2009;19(10):1197–1200.
255. Weatherill EE, Cain KL, Heywood SP, Compson JE, Heads JT, Adams R, et al. Towards a universal disulphide stabilised single chain Fv format: importance of interchain disulphide bond location and vL-vH orientation. Protein Eng Des Sel 2012;25(7):321–329.
256. Camacho CJ, Katsumata Y, Ascherman DP. Structural and thermodynamic approach to peptide immunogenicity. PLoS Comput Biol 2008;4(11):e1000231.
257. Crasto CJ, Feng JA. Linker: a program to generate linker sequences for fusion proteins. Protein Eng 2000;13(5):309–312.

258. Kamionka M. Engineering of therapeutic proteins production in *Escherichia coli*. Curr Pharm Biotechnol 2011;12(2):268–274.
259. Ulmer KM. Protein engineering. Science 1983;219(4585):666–671.
260. Perry LJ, Wetzel R. Disulfide bond engineered into T4 lysozyme: stabilization of the protein toward thermal inactivation. Science 1984;226(4674):555–557.
261. O'Rourke EC, Drummond RJ, Creasey AA. Binding of $^{125}$I-labeled recombinant beta interferon (IFN-beta Ser17) to human cells. Mol Cell Biol 1984;4(12):2745–2749.
262. Matsumura M, Becktel WJ, Matthews BW. Hydrophobic stabilization in T4 lysozyme determined directly by multiple substitutions of Ile 3. Nature 1988;334 (6181):406–410.
263. Lin L. Betaseron. Dev Biol Stand 1998;96:97–104.
264. Willemsen T, Hagemann U, Jouaux E, Stebel S, Mason J, Müller K, et al. Protein engineering. In: Walker J, Rapley R, editors, Germany Molecular Biomethods Handbook. Germany: Humana Press; 2008, p. 587–629.
265. Rose GD, Geselowitz AR, Lesser GJ, Lee RH, Zehfus MH. Hydrophobicity of amino acid residues in globular proteins. Science 1985;229(4716):834–838.
266. Desjarlais JR, Handel TM. De novo design of the hydrophobic cores of proteins. Protein Sci 1995;4(10):2006–2018.
267. Hurley JH, Baase WA, Matthews BW. Design and structural analysis of alternative hydrophobic core packing arrangements in bacteriophage T4 lysozyme. J Mol Biol 1992;224(4):1143–1159.
268. Marshall SA, Morgan CS, Mayo SL. Electrostatics significantly affect the stability of designed homeodomain variants. J Mol Biol 2002;316(1):189–199.
269. Filikov AV, Hayes RJ, Luo P, Stark DM, Chan C, Kundu A, et al. Computational stabilization of human growth hormone. Protein Sci 2002;11(6):1452–1461.
270. Pipe SW, Kaufman RJ. Characterization of a genetically engineered inactivation-resistant coagulation factor VIIIa. Proc Natl Acad Sci USA 1997;94(22):11851–11856.
271. Agren L, Norin M, Lycke N, Lowenadler B. Hydrophobicity engineering of cholera toxin A1 subunit in the strong adjuvant fusion protein CTA1-DD. Protein Eng 1999;12(2): 173–178.
272. Tan PH, Chu V, Stray JE, Hamlin DK, Pettit D, Wilbur DS, et al. Engineering the isoelectric point of a renal cell carcinoma targeting antibody greatly enhances SCFV solubility. Immunotechnology 1998;4(2):107–114.
273. Van De Weert M. Immunogenicity of Biopharmaceuticals. Series Vol. VIII. New York: Springer; 2008.
274. Bryson C, Jones TD and Baker MP. Prediction of immunogenicity of therapeutic proteins: validity of computational tools. Biodrugs 2010;24(1):1–8.
275. Kramer A, Schneider-Mergener J. Synthesis and screening of peptide libraries on continuous cellulose membrane supports. Methods Mol Biol 1998;87:25–39.
276. Wan M, Wang Y, Rabideau S, Moreadith R, Schrimsher J, Conn G. An enzyme-linked immunosorbent assay for host cell protein contaminants in recombinant PEGylated staphylokinase mutant SY161. J Pharm Biomed Anal 2002;28(5):953–963.
277. Baker MP, Reynolds HM, Lumicisi B, Bryson CJ. Immunogenicity of protein therapeutics: The key causes, consequences and challenges. Self Nonself. 2010;1(4):314–322.

278. Qu Z, Griffiths GL, Wegener WA, Chang CH, Govindan SV, Horak ID, et al. Development of humanized antibodies as cancer therapeutics. Methods 2005;36(1):84–95.
279. Jones TD, Crompton LJ, Carr FJ, Baker MP. Deimmunization of monoclonal antibodies. Methods Mol Biol. 2009;525:405–23.
280. Degroot A, Knoph P, Martin W. De-immunization of therapeutic protein by T-cell epitope modification. Dev Biol (Basel) 2005;122:171–194.
281. Onda M, Beers R, Xiang L, Nagata S, Wang Q, Pastan I. An immunotoxin with greatly reduced immunogenicity by identification and removal of B cell epitopes. Proc Natl Acad Sci USA 2008;105(32):11311–11316.
282. Weber CA, Mehta PJ, Ardito M, Moise L, Martin B, De Groot AS. T cell epitope: friend or foe? Immunogenicity of biologics in context. Adv Drug Deliv Rev 2009;61(11): 965–976.
283. Weiner LM. Fully human therapeutic monoclonal antibodies. J Immunother 2006;29(1): 1–9.
284. Dall'acqua WF, Damschroder MM, Zhang J, Woods RM, Widjaja L, Yu J, et al. Antibody humanization by framework shuffling. Methods 2005;36(1):43–60.
285. Hwang WYK, Foote J. Immunogenicity of engineered antibodies. Methods 2005;36(1): 3–10.
286. Presta LG. Engineering of therapeutic antibodies to minimize immunogenicity and optimize function. Adv Drug Deliv Rev 2006;58(5–6):640–656.
287. Tsurushita N, Hinton PR, Kumar S. Design of humanized antibodies: from anti-Tac to Zenapax. Methods 2005;36(1):69–83.
288. Sharkey RM, Goldenberg DM. Cancer radioimmunotherapy. Immunotherapy 2011; 3(3):349–370.
289. Wu AM, Senter PD. Arming antibodies: prospects and challenges for immunoconjugates. Nat Biotechnol 2005;23(9):1137–1146.
290. Sykes TR, Sykes CJ. Selected regulatory elements in the development of protein and peptide targeted radiotherapeutic agents. In: Monoclonal Antibody and Peptide-Targeted Radiotherapy of Cancer. Hoboken, NJ: John Wiley & Sons, Inc.; 2010. p. 571–611.
291. Kratz F. Albumin as a drug carrier: design of prodrugs, drug conjugates and nanoparticles. J Control Release 2008;132(3):171–183.
292. Noguchi Y, Wu J, Duncan R, Strohalm J, Ulbrich K, Akaike T, et al. Early phase tumor accumulation of macromolecules: a great difference in clearance rate between tumor and normal tissues. Jpn J Cancer Res 1998;89:307–314.
293. Matsumura Y, Maeda H. A new concept for macromolecular therapeutics in cancer chemotherapy: mechanism of tumoritropic accumulation of proteins and the antitumour agent smancs. Cancer Res 1986;6:6387–6392.
294. Elsadek B, Kratz F. Impact of albumin on drug delivery—new applications on the horizon. J Control Release 2012;157(1):4–28.
295. Bae YH, Park K. Targeted drug delivery to tumors: myths, reality and possibility. J Control Release 2011;153(3):198–205.
296. Teicher BA, Chari RV. Antibody conjugate therapeutics: challenges and potential. Clin Cancer Res 2011;17(20):6389–6397.
297. Teicher BA. Antibody–drug conjugate targets. Curr Cancer Drug Targets 2009;9(8):982–1004.

298. Sievers EL, Senter PD. Antibody–drug conjugates in cancer therapy. Annu Rev Med 2013;64: 15–29.
299. Minich SS. Brentuximab vedotin: a new age in the treatment of Hodgkin lymphoma and anaplastic large cell lymphoma. Ann Pharmacother 2012;46(3):377–383.
300. Verma S, Miles D, Gianni L, et al. Trastuzumab emtansine for HER2-positive advanced breast cancer. N Engl J Med 2012;367(19):1783–1791.
301. Chari RVJ. Targeted cancer therapy: conferring specificity to cytotoxic drugs. Acc Chem Res 2007;41(1):98–107.
302. Casi G, Neri D. Antibody–drug conjugates: basic concepts, examples and future perspectives. J Control Release 2012;161(2):422–428.
303. Lambert JM. Drug-conjugated monoclonal antibodies for the treatment of cancer. Curr Opin Pharmacol 2005;5(5):543–549.
304. Senter PD. Potent antibody drug conjugates for cancer therapy. Curr Opin Chem Biol 2009;13(3):235–244.
305. Ducry L, Stump B. Antibody–drug conjugates: linking cytotoxic payloads to monoclonal antibodies. Bioconjug Chem 2010;21(1):5–13.
306. Erickson HK, Lewis Phillips GD, Leipold DD, Provenzano CA, Mai E, Johnson HA, et al. The effect of different linkers on target cell catabolism and pharmacokinetics/pharmacodynamics of trastuzumab maytansinoid conjugates. Mol Cancer Ther 2012;11(5):1133–1142.
307. Hawkins MJ, Soon-Shiong P, Desai N. Protein nanoparticles as drug carriers in clinical medicine. Adv Drug Deliv Rev 2008;60(8):876–885.
308. Lopus M. Antibody–DM1 conjugates as cancer therapeutics. Cancer Lett 2011;307 (2):113–118.
309. Wakankar A, Chen Y, Gokarn Y, Jacobson FS. Analytical methods for physicochemical characterization of antibody drug conjugates. mAbs 2011;3(2):161–172.
310. Alley SC, Okeley NM, Senter PD. Antibody–drug conjugates: targeted drug delivery for cancer. Curr Opin Chem Biol 2010;14(4):529–537.
311. Kovtun YV, Goldmacher VS. Cell killing by antibody–drug conjugates. Cancer Lett 2007;255(2):232–240.
312. Vangeepuram N, Ong GL, Mattes MJ. Processing of antibodies bound to B-cell lymphomas and lymphoblastoid cell lines. Cancer 1997; 80(12 Suppl): 2425–2430.
313. Law CL, Cerveny CG, Gordon KA, Klussman K, Mixan BJ, Chace DF, et al. Efficient elimination of B-lineage lymphomas by anti-CD20–auristatin conjugates. Clin Cancer Res 2004;10(23):7842–7851.
314. Dijoseph JF, Dougher MM, Armellino DC, Kalyandrug L, Kunz A, Boghaert ER, et al. CD20-specific antibody-targeted chemotherapy of non-Hodgkin's B-cell lymphoma using calicheamicin-conjugated rituximab. Cancer Immunol Immunother 2007;56 (7):1107–1117.
315. Sutherland MS, Sanderson RJ, Gordon KA, Andreyka J, Cerveny CG, Yu C, et al. Lysosomal trafficking and cysteine protease metabolism confer target-specific cytotoxicity by peptide-linked anti-CD30–auristatin conjugates. J Biol Chem 2006;281 (15):10540–10547.
316. Smith LM, Nesterova A, Alley SC, Torgov MY, Carter PJ. Potent cytotoxicity of an auristatin-containing antibody–drug conjugate targeting melanoma cells expressing melanotransferrin/p97. Mol Cancer Ther 2006;5(6):1474–1482.

317. Polson AG, Calemine-Fenaux J, Chan P, Chang W, Christensen E, Clark S, et al. Antibody–drug conjugates for the treatment of non-Hodgkin's lymphoma: target and linker-drug selection. Cancer Res 2009;69(6):2358–2364.

318. Gerber HP, Senter PD, Grewal IS. Antibody drug-conjugates targeting the tumor vasculature: current and future developments. mAbs 2009;1(3):247–253.

319. Bernardes GJL, Casi G, Trüssel S, Hartmann I, Schwager K, Scheuermann J, et al. A traceless vascular-targeting antibody–drug conjugate for cancer therapy. Angew Chem Int Ed 2012;124(4):965–968.

320. Sapra P, Hooper AT, O'Donnell CJ, Gerber HP. Investigational antibody drug conjugates for solid tumors. Expert Opin Investig Drugs 2011;20(8):1131–1149.

321. Beck A, Haeuw JF, Wurch T, Goetsch L, Bailly C, Corvaia N. The next generation of antibody–drug conjugates comes of age. Discov Med 2010;10(53):329–339.

322. Dosio F, Brusa P, Cattel L. Immunotoxins and anticancer drug conjugate assemblies: the role of the linkage between components. Toxins (Basel) 2011;3(7):848–883.

323. Goldmacher VS, Kovtun YV. Antibody–drug conjugates: using monoclonal antibodies for delivery of cytotoxic payloads to cancer cells. Ther Deliv 2011;2(3):397–416.

324. Lowenberg B, Beck J, Graux C, Van Putten W, Schouten HC, Verdonck LF, et al. Gemtuzumab ozogamicin as postremission treatment in AML at 60 years of age or more: results of a multicenter phase 3 study. Blood 2010;115(13):2586–2591.

325. Ricart AD. Antibody–drug conjugates of calicheamicin derivative: gemtuzumab ozogamicin and inotuzumab ozogamicin. Clin Cancer Res 2011;17(20):6417–6427.

326. Advani A, Coiffier B, Czuczman MS, Dreyling M, Foran J, Gine E, et al. Safety, pharmacokinetics, and preliminary clinical activity of inotuzumab ozogamicin, a novel immunoconjugate for the treatment of B-cell non-Hodgkin's lymphoma: results of a phase I study. J Clin Oncol 2010;28(12):2085–2093.

327. Papisov MI, Hiller A, Yurkovetskiy A, Yin M, Barzana M, Hillier S, et al. Semisynthetic hydrophilic polyals. Biomacromolecules 2005;6(5):2659–2670.

328. Yurkovetskiy A, Choi S, Hiller A, Yin M, Mccusker C, Syed S, et al. Fully degradable hydrophilic polyals for protein modification. Biomacromolecules 2005;6(5):2648–2658.

329. Doronina SO, Bovee TD, Meyer DW, Miyamoto JB, Anderson ME, Morris-Tilden CA, et al. Novel peptide linkers for highly potent antibody–auristatin conjugate. Bioconjug Chem 2008;19(10):1960–1963.

330. Burke PJ, Senter PD, Meyer DW, Miyamoto JB, Anderson M, Toki BE, et al. Design, synthesis, and biological evaluation of antibody–drug conjugates comprised of potent camptothecin analogues. Bioconjug Chem 2009;20(6):1242–1250.

331. Jeffrey SC, Andreyka JB, Bernhardt SX, Kissler KM, Kline T, Lenox JS, et al. Development and properties of beta-glucuronide linkers for monoclonal antibody–drug conjugates. Bioconjug Chem 2006;17(3):831–840.

332. Jeffrey SC, Nguyen MT, Moser RF, Meyer DL, Miyamoto JB, Senter PD. Minor groove binder antibody conjugates employing a water soluble beta-glucuronide linker. Bioorg Med Chem Lett 2007;17(8):2278–2280.

333. Jeffrey SC, De Brabander J, Miyamoto J, Senter PD. Expanded utility of the β-glucuronide linker: ADCs that deliver phenolic cytotoxic agents. ACS Med Chem Lett 2010;1 (6):277–280.

334. Doronina SO, Mendelsohn BA, Bovee TD, Cerveny CG, Alley SC, Meyer DL, et al. Enhanced activity of monomethylauristatin F through monoclonal antibody delivery:

effects of linker technology on efficacy and toxicity. Bioconjug Chem 2006;17(1):114–124.

335. Erickson HK, Park PU, Widdison WC, Kovtun YV, Garrett LM, Hoffman K, et al. Antibody–maytansinoid conjugates are activated in targeted cancer cells by lysosomal degradation and linker-dependent intracellular processing. Cancer Res 2006;66 (8):4426–4433.
336. Zhao RY, Wilhelm SD, Audette C, Jones G, Leece BA, Lazar AC, et al. Synthesis and evaluation of hydrophilic linkers for antibody–maytansinoid conjugates. J Med Chem 2011;54(10):3606–3623.
337. Kovtun YV, Audette CA, Mayo MF, Jones GE, Doherty H, Maloney EK, et al. Antibody–maytansinoid conjugates designed to bypass multidrug resistance. Cancer Res 2010;70 (6):2528–2537.
338. Sammet B, Steinkühler C, Sewald N. Antibody–drug conjugates in tumor therapy. Pharm Patent Anal 2012;1(1):65–73.
339. Adair JR, Howard PW, Hartley JA, Williams DG, Chester KA. Antibody–drug conjugates—a perfect synergy. Expert Opin Biol Ther 2012;12: 1191–1206.
340. Trail PA. Antibody drug conjugates as cancer therapeutics. Antibodies 2013;2: 113–129.
341. Kellogg BA, Garrett L, Kovtun Y, Lai KC, Leece B, Miller M, et al. Disulfide-linked antibody–maytansinoid conjugates: optimization of *in vivo* activity by varying the steric hindrance at carbon atoms adjacent to the disulfide linkage. Bioconjug Chem 2011;22 (4):717–727.
342. Lewis Phillips GD, Li G, Dugger DL, Crocker LM, Parsons KL, Mai E, et al. Targeting HER2-positive breast cancer with trastuzumab-DM1, an antibody–cytotoxic drug conjugate. Cancer Res 2008;68(22):9280–9290.
343. Zhao RY, Erickson HK, Leece BA, Reid EE, Goldmacher VS, Lambert JM, et al. Synthesis and biological evaluation of antibody conjugates of phosphate prodrugs of cytotoxic DNA alkylators for the targeted treatment of cancer. J Med Chem 2011;55(2):766–782.
344. Moldenhauer G, Salnikov AV, Luttgau S, et al. Therapeutic potential of amanitin-conjugated anti-epithelial cell adhesion molecule monoclonal antibody against pancreatic carcinoma. J Natl Cancer Inst 2012;104: 622–634.
345. Boswell CA, Mundo EE, Zhang C, Bumbaca D, Valle NR, Kozak KR, et al. Impact of drug conjugation on pharmacokinetics and tissue distribution of anti-STEAP1 antibody–drug conjugates in rats. Bioconjug Chem 2011;22(10):1994–2004.
346. Junutula JR, Raab H, Clark S, Bhakta S, Leipold DD, Weir S, et al. Site-specific conjugation of a cytotoxic drug to an antibody improves the therapeutic index. Nat Biotechnol 2008;26(8):925–932.
347. Hamblett KJ, Senter PD, Chace DF, Sun MM, Lenox J, Cerveny CG, et al. Effects of drug loading on the antitumor activity of a monoclonal antibody drug conjugate. Clin Cancer Res 2004;10(20):7063–7070.
348. Junutula JR, Flagella KM, Graham R, Parsons KL, Ha E, et al. Engineered thio-trastuzumab-DM1 conjugate with an improved therapeutic index to target human epidermal growth factor receptor 2-positive breast cancer. Clin Cancer Res 2010;16 (19):4769–4778.
349. Axup JY, Bajjuri KM, Ritland M, Hutchins BM, Kim CH, et al. Synthesis of site-specific antibody–drug conjugates using unnatural amino acids. Proc Natl Acad Sci USA 2012;109(40):16101–16106.

350. Hussain AF, Kampmeier F, Von Felbert V, Merk HF, Tur MK, Barth S. SNAP-tag technology mediates site specific conjugation of antibody fragments with a photosensitizer and improves target specific phototoxicity in tumor cells. Bioconjug Chem 2011;22(12):2487–2895.

351. Nilsson R, Mårtensson L, Eriksson SE, Sjögren H-O, Tennvall J. Toxicity-reducing potential of extracorporeal affinity adsorption treatment in combination with the auristatin-conjugated monoclonal antibody BR96 in a syngeneic rat tumor model. Cancer 2010;116(S4):1033–1042.

352. Cao C, Shi C, Li P, Tong Y, Ma Q. Diagnosis of hepatitis C virus (HCV) infection by antigen-capturing ELISA. Clin Diagn Virol 1996;6(2–3):137–145.

353. Berkower I. The promise and pitfalls of monoclonal antibody therapeutics. Curr Opin Biotechnol 1996;7(6):622–628.

354. Weisser NE, Hall JC. Applications of single-chain variable fragment antibodies in therapeutics and diagnostics. Biotechnol Adv 2009;27(4):502–520.

355. Roifman CM, Mills GB, Chu M, Gelfand EW. Functional comparison of recombinant interleukin 2 (IL-2) with IL-2-containing preparations derived from cultured cells. Cell Immunol 1985;95(1):146–156.

356. Narhi Lo, Arakawa T, Aoki K, Wen J, Elliott S, Boone T, et al. Asn to Lys mutations at three sites which are N-glycosylated in the mammalian protein decrease the aggregation of *Escherichia coli*-derived erythropoietin. Protein Eng 2001;14(2):135–140.

357. Zündorf I, Dingermann T. Vom rinder-, schweine-, pferde-insulin zum humaninsulin: die biotechnische und gentechnische insulin-herstellung: bereitstellung ausreichender men-gen von humaninsulin. Pharmazie in unserer Zeit 2001;30(1):27–32.

358. Winter J, Neubauer P, Glockshuber R, Rudolph R. Increased production of human proinsulin in the periplasmic space of *Escherichia coli* by fusion to DSBA. J Biotechnol 2000;84(2):175–185.

359. Doyle MV, Lee MT, Fong S. Comparison of the biological activities of human recombinant interleukin-2(125) and native interleukin-2. J Biol Response Mod 1985;4(1):96–109.

360. De Oliveira JE, Soares CRJ, Peroni CN, Gimbo E, Camargo IMC, Morganti L, et al. High-yield purification of biosynthetic human growth hormone secreted in Escherichia coli periplasmic space. J Chromatogr A 1999;852(2):441–450.

361. Hsu E, Osslund T, Nybo R, Chen BL, Kenney WC, Morris CF, et al. Enhanced stability of recombinant keratinocyte growth factor by mutagenesis. Protein Eng Des Sel 2006;19 (4):147–153.

362. Brange J. The new era of biotech insulin analogues. Diabetologia 1997;40(2):S48–S53.

363. Pan CQ, Lazarus RA. Engineering hyperactive variants of human deoxyribonuclease I by altering its functional mechanism. Biochemistry 1997;36(22):6624–6632.

364. Blatt LM, Davis JM, Klein SB, Taylor MW. The biologic activity and molecular characterization of a novel synthetic interferon-alpha species, consensus interferon. J Interferon Cytokine Res 1996;16(7):489–499.

365. Kaur S, Venktaraman G, Jain M, Senapati S, Garg PK, Batra SK. Recent trends in antibody-based oncologic imaging. Cancer Lett 2012;315(2):97–111.

# Analytical Characterization of Proteins and Peptides

YAJUN JENNIFER WANG and BRIAN HOSKEN

## 10.1 INTRODUCTION

The biopharmaceutical manufacturing of therapeutics produces a heterogeneous product composed of populations of structurally related molecules, referred to as product-related variants. The quality and heterogeneity of every protein product is highly dependent on its manufacturing process [1]. This fact has two important implications for the biopharmaceutical industry. First, the manufacturing process must be tightly controlled. After a process has been optimized and locked, changes cannot be made without first assessing the impact on product quality and comparability. The second point, which is the basis for this chapter, is that protein therapeutics must be well-characterized by physicochemical, immunochemical, and biological methods. The definition for a well-characterized biologic was published in the *Federal Register* in 1996 [2]. The concept of a well-characterized biologic was later summarized as "a recombinant DNA-derived protein with known amino acid sequence, disulfide linkages, glycosylation sites, and post-translational modifications. The purity, impurities, and biological activity must be quantifiable and the impurities should be identified" [3,4].

It is often stated that biologics, such as protein therapeutics, cannot be fully characterized because of their complex nature. Although this statement is generally accepted, it should also be acknowledged that advances in bioanalytical techniques and more than 30 years of industry experience with therapeutic proteins are closing the gap between the industry standard of a well-characterized biologic and the goal of a fully characterized biologic [5]. In practice, the extent of characterization is linked to the level of risk associated with each phase of drug development (International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use [ICH] Q9) [6]. For example, although there may not be sufficient time or resources for extensive characterization of a protein therapeutic during early stage development, it is expected that the molecule must

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*,
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

be well characterized before the Biologic License Application is filed with regulatory agencies.

Protein therapeutics and their variants are characterized and classified by many physicochemical properties, including hydrated size (Stokes radius), molecular weight (MW), charge, hydrophobicity, electrophoretic mobility, isoelectric point (pI), sedimentation velocity, glycosylation, and spectral properties. The nature of each variant can be related to differences in their primary, secondary, tertiary, or quaternary levels of protein structure. Protein variants can be created throughout the biological manufacturing process, which includes the following steps: cell culture, harvest, purification, formulation, and filling into the final container. Some variants may also be introduced during analysis as artifacts. Therefore, product characterization and quality control testing are required at critical points throughout clinical development and manufacturing [7].

This chapter describes the protein therapeutics, including monoclonal antibodies (mAbs), in terms of their expected protein structure and their commonly encountered variants. Current methods and strategies used for the physicochemical characterization of protein therapeutics and their variants will be discussed.

## 10.2 STRUCTURAL CHARACTERIZATION AND CONFIRMATION

The primary structure of a protein therapeutic is its most fundamental property. The amino acid sequence of each protein produced via recombinant DNA technology is known from the vector inserted into each clone. Therefore, the first step in protein characterization should be interrogation of the expected primary structure using widely available software programs. This *in silico* analysis can be used to calculate theoretical MWs, reveal potential glycosylation sites, identify enzymatic cleavage sites, predict hydrophobicity, determine isoelectrical point, compare sequence alignment with other protein products, and identify potential hotspots for unwanted posttranslational modifications.

Variants associated with primary structure (sequence variants) can be attributed to unusual translational events such as the misincorporation of amino acids [8], expression of introns [9], cross-over events [10], or the presence of an uncleaved leader sequence [11]. More commonly, sequence variants are caused by genetic impurities in the cell line. These mutations, which occur during transfection and amplification of the clone [12], can lead to cell lines that produce molecules with varying levels of amino acid substitutions [13]. For this reason, the primary structure of a recombinant protein must be verified to ensure the genetic fidelity of the cell line used to manufacture the pharmaceutical. The following techniques are usually used to verify the amino acid composition and primary structure of therapeutic proteins.

### 10.2.1 Amino Acid Composition by Amino Acid Analysis

Protein amino acid composition and quantitation can be analyzed upon hydrolysis [14]. Several derivatization techniques, such as ninhydrin [15], phenylisothiocyanate

(PITC), 6-aminoquinolyl-*N*-hydroxysuccinimidyl carbamate (AQC) [16], and *o*-phthalaldehyde/9-fluoromethyl-chloroformate (OPA/FMOC) [17], with either precolumn or postcolumn derivatization, have been used. Direct analysis of amino acids using the integrated pulsed amperometric detection (IPAD) method is also possible. The liberated amino acids are separated using reversed-phase high-performance liquid chromatography (RP-HPLC), ion-exchange chromatography (IEC), or capillary electrophoresis (CE). Recently, methods that combine AccQ·Tag chemical derivatization with the exceptional chromatographic resolution of ultra-high-performance LC (UHPLC) [18], and in conjunction with the high sensitivity and selectivity of tandem mass spectrometry (MS/MS) were reported [19].

Free amino acid composition and their levels in the cell culture media are analyzed directly. Another example of amino acid analysis is the detection and quantification of nor-Leucine misincorporation for methionine during *Escherichia coli* fermentation processes [15].

### 10.2.2 Amino Acid Sequence via N-Terminal Sequencing

N-terminal sequencing via Edman degradation [20] was routinely used to confirm the protein amino acid sequence until tandem LC–MS was introduced more than a decade ago. N-terminal sequencing still has its value as an identity method to distinguish isobaric amino acid residues, such as Leu/Ile and deamidation product isoaspartate (isoAsp) or aspartic acid (Asp), and to provide complementary sequence information along with MS/MS [21].

### 10.2.3 Peptide Mapping via Tandem Liquid Chromatography–Mass Spectrometry

Peptide mapping is the main tool for confirming the primary structure of proteins and identifying sequence variants. The general procedure for generating a peptide map is to denature, reduce, alkylate, and digest the protein with a highly specific protease. The resulting peptides are separated and profiled by RP-HPLC with ultraviolet (UV) detection [8,22,23]. A recent article described the optimization of peptide maps for mAbs with emphasis on reducing sample preparation time and eliminating artifacts [24].

An early example of the identification of a sequence variant by peptide mapping was reported by Harris et al. in 1993 [12]. A combination of techniques (amino acid analysis, N-terminal sequencing, and MS) was used to identify every peak in the chromatogram, resulting in 98% coverage of the antibody. The mutation, which occurred in 21% of the heavy chain, was also confirmed at the DNA level with a polymerase chain reaction (PCR) assay. The sequence variant detection sensitivity improves with the use of modern MS. It has been reported that a comparison of UV peptide map chromatograms only allows the detection of sequence variants occurring at a frequency greater than 3% [25]. With MS/MS detection, peptide maps are capable of resolving and quantifying low-level variant peptides that co-elute with more abundant peptides. Under some circumstances, the LC/MS/MS analysis of

peptide maps can enable the identification of sequence variants occurring with a frequency as low as 0.01% [26]. A workflow for identifying posttranslational modifications and amino acid substitutions from high-resolution and mass accuracy LC/MS/MS instruments has recently been described [27]. Demonstrating the absence of sequence variants in a therapeutic antibody requires a significant amount of time and highly skilled analysts. A systematic analytical approach combining tryptic and chymotryptic peptide mapping with a mascot error tolerant search (ETS) has been developed to detect and identify low-level protein sequence variants in recombinant mAbs [25].

### 10.2.4 Protein Identity by Matrix-Assisted Laser Desorption/Ionization–Mass Spectrometry Peptide Mass Fingerprinting

Peptide mass fingerprinting (PMF) was first described and demonstrated with fast atom bombardment MS by Henzel et al. 1989 [28]. Despite its promise, the method was seldom used until 1992, with the coming of significantly more sensitive commercial instrumentation based on matrix-assisted laser desorption/ionization–time-of-flight–mass spectrometry (MALDI-TOF-MS). Since then, the method grew from a need for a faster, more efficient method to identify frequently observed proteins in electrophoresis gels. PMF laid the foundation for high-throughput, high-sensitivity methods of protein analysis, now known as proteomics [29]. PMF is now being used routinely as a protein identification method in therapeutic protein product manufacturing because of its high specificity and fast analysis time. The method can be validated per ICH guideline Q6B [4] for identity testing.

### 10.2.5 Sulfhydryl Group(s) and Disulfide Bridges

Characterization at the secondary and tertiary levels of protein structure often focuses on the many intra- and intermolecular disulfide bonds. Disulfide bonds in the antibody hinge region are difficult to assign because of closely spaced cysteines. The analysis of disulfide-linked peptides containing multiple cysteine residues by N-terminal sequencing and LC/MS allows for the unambiguous assignment of disulfide linkages [30]. The complete set of disulfide bond linkages was assigned for an IgG4 molecule using a strategy that involved LC/MS/MS analysis of reduced and nonreduced Lys-C digests [31].

Disulfide variants related to incorrect disulfide linkages can be unique to specific IgG isotypes. For example, IgG4 antibodies have a tendency to form intramolecular disulfides in the hinge region [32]. The result of this modification is an intact antibody held together by noncovalent interactions. The relative percentage of noncovalently bound antibodies (i.e., noncovalently bound light chain and heavy chain) can be measured by denaturing methods such as sodium dodecyl sulfate (SDS)–capillary gel electrophoresis [33]. IgG2 antibodies have a tendency to form intermolecular disulfide bonds between the hinge regions and the light or heavy chains [34]. These mixed disulfides produce structural isoforms that have been separated by electrophoresis, reversed-phase and IEC [34–35].

There are also unusual variants related to the loss or addition of a sulfur atom in disulfide bonds. A nonreducible thioether bond has been identified in IgG1 molecules [36]. This cross-link occurs between the heavy (Cys-223) and light chain (Cys-213) after the loss of a sulfur atom. A decomposition pathway involving β-elimination has been proposed [37]. Recently, the existence of trisulfide bonds in the hinge of an IgG2 mAb has been reported [38]. This modification is unusual in proteins and may be a result of the fermentation process.

Every cysteine residue in the antibody should be engaged in an intra- or intermolecular disulfide bond. Although intramolecular disulfides are not required for the correct protein conformation, they do have a stabilizing effect on the tertiary and quaternary structure of the antibody [39,40]. Therefore, it is expected that the presence of unpaired cysteines could lower the thermodynamic stability of the molecule. A free thiol variant could affect product quality during storage and limit the shelf life of the therapeutic antibody. The thiol side chain on an unpaired cysteine residue is very reactive and could be a potential source of product heterogeneity. For example, cysteinylation of an unpaired thiol in a mAb has been identified by LC/MS analysis of the Fab region [41]. Free sulfhydryls have also been demonstrated to reduce the potency of a therapeutic antibody [42]. Free thiols are typically assayed with Ellman's reagent [43], although a fluorescence method for measuring unpaired cysteines has also been developed [44].

### 10.2.6 Carbohydrate Structure

The glycosylation of proteins occurs during and after translation and involves multiple glycosyltransferases and glycosidases. The resulting oligosaccharide structures are complex and highly dependent on the expression system and cell culture conditions. Although glycoproteins may have many different forms of glycan structures at either O-linked or N-linked sites (or both), all IgG antibodies contain a single, N-linked, glycosylation site in each of the CH2 domains of the Fc region. In addition, 15 to 20% of naturally occurring polyclonal antibodies and one licensed therapeutic antibody also contain an additional glycosylation site in the Fab regions [45]. When expressed in hamster or mouse cell lines, the Fc sites of recombinant antibodies are occupied by oligosaccharide structures similar to those found in the human population. The dominant glycan species are related to a common biantenary, core-fucosylated structure with the addition of zero (G0), one (G1), or two (G2) terminal galactoses. In addition to the major glycan species, there are many minor species related to incomplete processing or the presence of additional sugars on the terminal galactose. Common glycan species observed in recombinant mAbs expressed in mammalian cell lines and their corresponding mass are shown in Table 10.1.

Glycosylation is essential for the proper structure and function of antibodies [46]. The effects of certain glycan structures or the absence of glycans on the biological activity of therapeutic antibodies is well established and has been thoroughly reviewed [47,48]. For example, the absence of core fucose or the presence of terminal galactose increases certain effector functions of antibodies. Glycosylation is a critical attribute of therapeutic antibodies whose mode of action involves cell

**TABLE 10.1 Typical N-Linked Glycans on Recombinant Monoclonal Antibodies**

| Structure | Abbreviation | Average Mass (Da) of N-Linked Glycans [M-18]* | Average Mass (Da) of Released Glycans $[M+Na]^+$ |
|---|---|---|---|
| 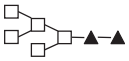 | Man5 | 1217.1 | 1258.1 |
| 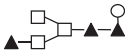 | G0-GlcNAc | 1242.1 | 1283.1 |
| | G0-Fuc | 1299.2 | 1340.2 |
| 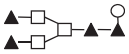 | G0 | 1445.3 | 1486.3 |
| 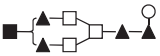 | G1/G1′ | 1607.5 | 1648.5 |
| 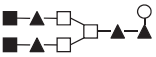 | G2 | 1769.6 | 1810.6 $[M-H]^-$ |
| 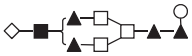 | G1+SA | 1898.7 | 1915.7 |
| 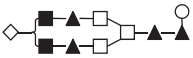 | G2+SA | 2060.8 | 2077.8 |
| 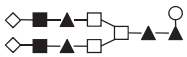 | G2+2SA | 2352.1 | 2369.1 |

*Dehydrated masses (M−18) are useful for identifying glycans on peptides and proteins by liquid chromatography–mass spectrometry. The masses of released glycans ($[M+Na]^+$ or $[M-H]^-$) are used to identify glycans by matrix-assisted laser desorption/ionization analysis. ◯ Fucose (Fuc), □ mannose (Man), ▲ GlcNAc, ■ galactose (Gal), ◇ sialic acid (SA).

killing through effector functions. Glycosylation may also pose a safety concern. Antibodies produced in mammalian cell lines can incorporate non-human and potentially immunogenic sugars [49]. For example, whereas murine cell lines add *N*-glycolyl neuraminic acid (NGNA), native human antibodies contain only *N*-acetyl neuraminic acid (NANA). Oligosaccharide linkages can also vary among species. Mouse cell lines are also capable of producing galactose-α-1,3-galactose

(α-Gal), but humans lack the necessary glycosyltransferase to make this linkage. The Gal-α-1,3-Gal disaccharide present in the Fab portion of a therapeutic antibody has been linked to immune responses in patients [45]. Because glycosylation can affect the efficacy, pharmacokinetics (PK), and safety of mAbs and protein therapeutics in general, it is necessary to perform a thorough characterization of their carbohydrates. Because of the complex nature and heterogeneity of oligosaccharides, detailed characterization may require multiple analytical techniques.

Glycans can be characterized at several levels of detail. The most basic analysis involves hydrolyzing the oligosaccharide into its individual sugars. Compositional analysis is used to determine the relative percentage of carbohydrates present on the antibody and to determine the concentrations of individual monosaccharides. The next level of analysis involves profiling the intact glycans released from the protein. The highest level is glycosylation site profiling. This type of analysis requires protein digestion and separation of the glycopeptides before glycan profiling. Because compositional analysis provides little information and glycosylation site profiling is not required for typical antibodies, the focus of this review is on glycan profiling of the Fc region of mAbs. Glycan profiling techniques can be further divided into those that are used for the characterization of enzymatically released glycans and those that are applied to the characterization of the intact, reduced, or digested antibody containing bound glycans.

***10.2.6.1 Analysis of Released Glycans*** The analysis of released, N-linked glycans usually requires several labor-intensive and time-consuming steps. Glycans must first be released from the antibody, purified to remove the proteins, and derivatized with a fluorescent dye for detection purposes [50] and then frequently purified again to remove the free dye before separation by chromatography or electrophoresis. Glycans can be released from the antibody by chemical or enzymatic methods. Hydrazinolysis completely removes N-linked glycans from proteins [51]; however, this method uses hazardous chemicals and degrades the protein [52]. Enzymatic approaches are more commonly used for antibodies because they typically contain only N-linked glycosylation sites, and the only protein modification from digestion is deamidation of the asparagine (Asn) glycosylation site [53]. Most protocols use the endoglycosidase peptide-N-glycosidase F (PNGase F) because this enzyme cleaves the amide bond between Asn and the first GlcNAc [53]. Methods for the enzymatic release of glycans from protein in solution or bound to polyvinylidene fluoride (PVDF) membranes have been described [54,55]. Both chemical and enzymatic methods for oligosaccharide release create an intact glycan with a reducing terminus (aldehyde functional group) that can be derivatized through the process of reductive amination with fluorescent probes containing aromatic amines [50]. Fluorescent detection provides the relative molar concentrations of each glycan because each structure has a single label at its reducing end.

***10.2.6.2 High-Performance Liquid Chromatography Profiling of Glycans*** HPLC profiling of fluorescently labeled glycans provides both high resolution and sensitivity. Critical steps of this analysis (oligosaccharide release, selection

of fluorescent probes and derivatization chemistry, purification methods, modes of chromatography) and strategies for sequence determination have been reviewed [56,57]. With regard to mAbs, most separations are performed using normal phase chromatography with amide or amine functional groups on the stationary phase [58–61]. Amide columns separate glycans by hydrophilic interactions (HILIC), which increase with the size of the glycan structure. Amine columns provide a mixed-mode separation based on the charge and hydrophilicity of the oligosaccharides. UHPLC separation of released glycans in HILIC mode was compared with reversed-phase (RP) and porous graphitic carbon chromatography (PGC) modes. The high selectivities, as well as the commercial availability of ultra-high-performance stationary phases render HILIC the chromatography method of choice for the analysis of glycans [62]. Nonlabeled oligosaccharides can be separated by high pH anion exchange chromatography with pulsed amperometric detection (HPAEC-PAD) [63,64]. Although not as sensitive as methods using fluorescent detection, HPAEC-PAD has the benefit of much simpler sample preparation.

The identification of chromatographic peaks is typically determined by comparing the retention times of unknown peaks to those of known standards or by performing sequential digestions with highly specific exoglycosidases [65]. Unknown peaks can also be identified by fraction collection and off-line MS analysis [66]. The use of an LC/MS system allows for the direct identification of glycan species. For example, glycans from a recombinant IgG have been labeled with 2-aminobenzamide (2-AB), separated by reverse phase ($C_{18}$) chromatography and analyzed by electrospray ionization (ESI)/MS [67,68].

***10.2.6.3 Capillary Electrophoresis Profiling of Glycans*** Electrophoretic techniques are capable of providing separations similar to HPLC with higher resolution in less time. The evolution of electrophoretic methods for glycan profiling, from paper to capillary electrophoresis, has been reviewed [69]. Capillary electrophoresis with laser-induced fluorescence (CE-LIF) detection is a viable alternative to chromatography methods because of its high resolution, sensitivity, and speed. Separations can be completed in 5 to 25 minutes depending on the protocol. The separation and detection of neutral oligosaccharides requires derivatization with a charged fluorophore such as APTS (8-aminopyrene-1,3,6-trisulfonic acid) [70,71] or 2-AA (2-aminobenzoic acid) [72]. With each glycan carrying the same charge at its reducing terminus, differences in mobility are related entirely to differences in the hydrated volume of each structure. This mechanism permits the complete resolution of both structural isomers related to the presence of a single terminal galactose (G1 and G1′).

Identification of glycan species in the electropherogram is usually determined by comparing the migration time of unknown peaks to those of oligosaccharide standards or by performing sequential digestions of oligosaccharide samples with specific exoglycosidases. Another strategy for characterizing CE glycan profiles involves first purifying the labeled oligosaccharides by HPLC, identifying them by offline MS, and then analyzing the known glycan species by CE-LIF [72]. Overcoming the challenges of coupling CE to MS has made it possible to directly

characterize the CE profiles. Recently, a CE-LIF method was modified and coupled to an ESI-TOF MS to characterize unknown glycan species observed during routine CE-LIF profiling [73,74].

***10.2.6.4 Mass Spectrometry Profiling of Released Glycans*** Released glycans are also characterized by MS techniques. Labeled glycans contain the necessary charge for detection, but unlabeled, neutral glycans are detected in the positive mode as adducts of alkali metals. Sialylated glycans can be detected in the negative mode [54]. For mAbs produced in well-characterized expression systems, the MW information from MALDI is usually sufficient to identify glycan species. Confirmation of the linear sequence and linkages requires a more complicated ESI-MS/MS analysis. A multimethod strategy for the complete structural characterization of 2-AA labeled glycans released from the Fc and Fab sites of a therapeutic antibody has been described [66]. Normal phase HPLC, MALDI/MS, sequential enzymatic digestion, and MS/MS fragmentation were used to identify 21 oligosaccharide structures. The sequence and linkages were defined by the activity of highly specific exoglycosidases and MS/MS fragmentation patterns.

***10.2.6.5 Glycan Profiling of Intact, Reduced, and Digested Monoclonal Antibodies*** Glycan profiling can also be accomplished by the LC/MS analysis of intact, reduced, or digested antibodies. The performance and limitations for each approach have been reviewed [75]. Analysis of the intact antibody has the benefit of requiring very little sample preparation. However, the resolution is limited by several factors: (1) the high MW of the intact antibody, (2) molecule heterogeneity, and (3) broad isotopic distribution. Analysis of the reduced or papain-digested antibody provides higher mass resolution as a result of the lower MW and less heterogeneity. Another benefit of the LC/MS analysis of papain digested antibodies is that the physical separation of the Fab and Fc portions allows for the site specific profiling of mAbs with an additional glycosylation site in the complimentary determination regions (CDR) [76]. The performance of LC/MS for quantifying glycoproteins and glycopeptides from five different sample preparations (intact IgG, Fc fragment, HC, HC Fc fragment, and the tryptic peptide) was recently compared with N-glycan HPLC profiling. LC/MS analysis of the HC Fc fragment gave comparable results in terms of the number of glycans identified and their relative concentrations to the standard HPLC profiling method [77].

## 10.3 PHYSICOCHEMICAL PROPERTIES

### 10.3.1 Molecular Weight (or Size)

Molecular weight (or size) is one of the fundamental physicochemical properties of proteins. Size profiling analytical techniques includes size exclusion chromatography (SEC) with light scattering detection, ultracentrifugation, electrophoresis, and MS. These techniques are summarized in this section.

***10.3.1.1 Size Exclusion Chromatography*** Size exclusion chromatography is the simplest and most routinely used technique in the biopharmaceutical industry for measuring size heterogeneity. Size variants, including high-molecular-weight aggregates and low-molecular-weight fragments, are separated from the intact proteins by differences in their hydrated sizes (Stokes radius). Proteins diffuse in and out of pores in the stationary phase at rates that depend on their hydrated size. The differences in diffusion rates are measured by the volume (or retention time) required for each variant to elute from the column. The elution volume of each variant can be used to calculate its relative MW compared with those of protein standards with a known MW and Stokes radii.

Size exclusion chromatography is widely used throughout industry for many reasons. HPLCs are standard equipment, and this technique has few parameters (column, flow rate, mobile phase, and protein load) to optimize. SEC can also be performed under denaturing conditions (dSECs) to measure size variants that are noncovalently associated and therefore undetectable under native conditions [78].

Size exclusion chromatography has several limitations, such as relatively low resolution. Although SEC can be used to quantify irreversible and slowly dissociating aggregates, fast-dissociating aggregates may not be observed because of dilution that occurs during sample preparation and chromatography. This implies that the measured aggregate profile may not accurately reflect the aggregate profile of the formulated proteins. The separation can also be affected by nonspecific interactions between the protein and column. These types of interactions lead to broad peak shapes, delayed elution volumes, and poor column recovery. The nature of protein–column interactions and the importance of mobile phase composition has been reviewed [79].

***10.3.1.2 Light Scattering*** The addition of refractive index and multiangle laser light scattering (MALLS) detectors to the SEC HPLC system allows for the simultaneous determination of aggregate concentrations and absolute MWs [80–82].

The MW value obtained from static light scattering is considered absolute because it is calculated directly from the measurement using equations derived from first principles and does not rely on size standards. Accurate MW calculations depend on well-resolved species. The refractive index and light scattering detectors must also be calibrated before the analysis. Light scattering detectors are very sensitive to higher order aggregates, making this analysis suitable for measuring the removal of aggregates during purification steps [83] and for measuring product stability [82].

Dynamic light scattering (DLS) measures the diffusion coefficient of proteins in solution. The measured diffusion coefficient can then be related to the hydrated diameter of a protein by the Stokes–Einstein equation. DLS measurements are simple and fast because they do not require calibration, and it is only necessary to know the temperature, refractive index, and viscosity of the solution. Another benefit is that measurements can be made under conditions that are more representative of the formulated antibody. Similar to static light scattering, accurate size determinations are only possible for homogenous protein solutions. Applications for DLS in

the pharmaceutical industry include rapid determination of size homogeneity, measurement of thermal stability of antibodies, and online characterization of size variants from SEC [84]. This technique has also been used to study the effect of aggregation and protein concentration on the opalescence of antibody formulations [85] and to demonstrate the removal of size variants by successive purification steps [83].

***10.3.1.3 Analytical Ultracentrifugation*** Analytical ultracentrifugation (AUC) provides absolute MWs under sedimentation equilibrium conditions. A protein solution reaches equilibrium in a centrifuge when the sedimentation force from radial acceleration is countered by the opposing force of diffusion. The equilibrium concentration profile can be fit to an exponential equation that provides the MW of the protein. As with light scattering measurements, a sedimentation equilibrium experiment only provides a true MW value for a homogenous protein solution and therefore has limited applications for the characterization of protein variants.

Sedimentation velocity (or moving boundary) is a second type of AUC experiment that is more relevant to the characterization of therapeutic antibodies [86,87]. Sedimentation velocity experiments are performed at higher rotor speeds than sedimentation equilibrium experiments to overcome the opposing diffusion force. Under these conditions, each size variant sediments at a velocity that is dependent on its MW and hydrodynamic properties. The concentration profile is measured at intervals throughout the separation and then deconvoluted to yield sedimentation coefficients and relative amounts for each size variant.

Sedimentation velocity experiments provide information similar to SEC with good agreement [88]. SEC separations provide better sensitivity for low levels of aggregates. However, the limit of quantitation for AUC has been separately reported to be 0.6% of aggregates [89,90]. For the characterization of complex aggregate profiles, sedimentation velocity experiments have several key advantages over SEC separations. AUC provides higher resolution across a larger range of protein sizes. More important, AUC experiments are performed on a homogenous solution without prior separation or preparation, which eliminates the artifacts sometimes observed during chromatography. AUC experiments can be performed under many different solution conditions, including pH, protein concentration, ionic strength, and temperature. This benefit allows the characterization of aggregate profiles under conditions that are more representative of the formulated product. Despite these advantages, AUC is not widely used for the routine characterization of antibodies. Factors limiting its use include the requirement of specialized instrumentation, complicated data analysis, and the requirement of highly trained personnel.

***10.3.1.4 Electrophoresis*** Gel electrophoresis is an established method for measuring size variants under denaturing conditions. This analysis resolves fragments, which are only held together by noncovalent interactions and are therefore undetected by SEC. Although there are now several platforms for performing electrophoresis, the separation principles remain the same. In the presence of SDS, proteins obtain a constant mass-to-charge ratio that enables their separation in a molecular sieving matrix under the influence of an applied electric field.

The smaller proteins migrate through the gel faster than the larger ones. Within the pharmaceutical industry, the standard polyacrylamide slab gel format (SDS-PAGE) has been largely replaced by capillary electrophoresis in a gel-filled capillary (CE-SDS) for several reasons. First, the CE instrument is fully automated in contrast to the labor-intensive SDS-PAGE analysis. Second, UV or LIF signal near the capillary outlet is recorded digitally. Online detection eliminates the staining, destaining, imaging, and densitometry steps required for slab gel analysis. Third, the analytical performance of CE-SDS is better than that of SDS-PAGE with regards to accuracy, precision, and dynamic range. Both methods can have comparable sensitivity depending on the staining and detection methods.

CE-SDS methods have been developed for both UV [33,91,92] and fluorescence detection [93–97] strategies. Fluorescence detection requires that the antibody sample first be derivatized with a fluorescent probe. This labeling step complicates sample preparation and adds the possibility of introducing an artifact. However, the sensitivity of the method is greatly increased. The 3-(2-furoyl)-quinoline-2-carboxaldehyde (FQ) dye fluorescence method takes the advantage of nonfluorescing free dye so that it does not need to be removed after labeling [94,97]. Fluorescent probes and labeling strategies for CE can be found in a review by Lacroix et al. [98].

Gel electrophoresis has also been adapted to microfluidic devices [99]. Applications of this technology for protein analysis have been described in several reviews [100–102]. The suitability of microchip electrophoresis (MCE) with commercially available instrumentation and chips has been demonstrated for the sizing and quantitation of antibody size variants [103–105]. This format has all the advantages of CE-SDS with one major distinction. The separation of antibody size variants can be performed in a fraction of the time as a result of the much shorter separation distance. The analytical performance of MCE has been compared with that of conventional CE-SDS with UV detection. The chip format was 70 times faster and had comparable resolution, precision, and sensitivity [106].

***10.3.1.5 Liquid Chromatography–Mass Spectrometry*** As the first step of protein structural characterization, the mass of intact or reduced molecules is usually determined using MS. Modern instruments can determine the mass of a mAb to a mass accuracy of close to 10 ppm ($\sim$1.5 Da) [107].

Theoretical considerations include instrument resolution, mass accuracy, and mass range. An ideal instrument for mass determination of intact mAbs should have a resolution of 8000 to 10,000, a mass accuracy in the mid-ppm range, and an m/z range of up to 5000 for ESI, and 175,000 for MALDI [107]. The first successful attempts for mass determination of intact antibodies with reasonable accuracy were performed using MALDI-TOF instrumentation or ESI with a quadrupole mass analyzer. However, the method of choice today for accurate mAb mass determination is by LC–MS on an ESI-TOF or ESI-Q-TOF instrument [108].

To increase the throughput of mass determination, coupling LC directly to the mass spectrometer becomes necessary. However, analytical characterization of therapeutic proteins, especially mAbs, had been hindered for a long time by the lack of appropriate chromatographic methods to be used in conjunction with

high-resolution MS. Those methodologies for standard RP–HPLC, especially the RP-chromatographic resins, are incompatible with antibodies because of irreproducibility, low recovery, short column lifetimes, and poor resolution of degradation products. Improved analytical RP-HPLCMS methods were developed for monitoring and characterizing recombinant proteins [109], intact antibodies [108], reduced antibodies [110], human antibody IgG2 domains [111], and limited digested Fc-fusion proteins [112].

Another challenge for MS analysis of chromatography using a salt mobile phase, such as IEC and SEC, is the incompatibility of subsequent detection. However, by modifying the mobile phase with the MS detection to exclude salt, the method becomes compatible with MS for LC–MS analysis of large proteins, with some sacrifice in chromatography resolution. To enable the MS detection from typical LC separations, Alvarez et al. demonstrated a generic online LC–MS method operated in a two-dimensional format for the rapid characterization of mAb charge and size variants [113]. Using a single chromatographic system capable of running two independent gradients, up to six fractions of interest from a size exclusion (SEC) or an ion exchange (IEC) separation can be identified by trapping and desalting the fractions onto a series of reversed phase trap cartridges with subsequent online analysis by MS. An online disulfide reduction step was successfully incorporated into the workflow, allowing more detailed characterization of modified mAbs by providing chain-specific information.

### 10.3.2 Extinction Coefficient

The extinction coefficient of a biological product is one of the physicochemical properties that ICH Q6B requires to use if UV absorption is used to measure protein content [4]. The mass extinction coefficient ($E^{0.1\%}$, referred to as extinction coefficient) represents the absorbance of a 1 mg/ml or 0.1% mass/volume protein solution in a 1-cm cuvette. The extinction coefficient can be predicted based on the amino acid composition of the protein under the experimental conditions [114,115], experimentally determined via amino acid analysis [15,116], or both.

Formulation changes can have an effect on protein absorbance and should be considered. Protein UV absorbance is partially dependent on the ionization state of tyrosine residues. Because the $pK_a$ of tyrosine is about 10, the extinction coefficients of all formulation buffers below pH 8 are considered equivalent. At a pH above 8, differences in pH can affect absorbance significantly. Therefore, if the formulation pH changes across pH 8, the experimental extinction coefficient must be redetermined with the new formulation.

### 10.3.3 Charge Profiling

Therapeutic proteins are also characterized by differences in their charge states (i.e., net charges at a certain pH). Charge profiling of intact proteins, under native conditions, can resolve variants related to protein conformation, size, sequence variants, glycosylation, and posttranslational modifications. Charge profiling

methods are an integral component of characterization studies and quality control strategies because they are sensitive to many types of modifications.

Ion-exchange chromatography and isoelectric focusing (IEF) are widely used to characterize proteins within the pharmaceutical industry. These two charge profiling methods and the commonly observed modifications have been thoroughly reviewed [117]. Both techniques use very different separation mechanisms and may or may not provide complementary information. Whereas IEC is based on electrostatic interactions between the stationary phase and charged groups on the protein surface, IEF separations are based on differences in the overall net charge of the protein isoforms.

***10.3.3.1 Ion-Exchange Chromatography*** Ion-exchange chromatography separations are developed by choosing a cation or anion exchange column and varying the buffer system, pH, and ionic strength of the elution buffer. IEC of mAbs is typically performed with a weak cation exchanger (WCX) and salt gradients [118]. However, a recent publication described the optimization of a pH gradient IEC method for the separation of mAb charge variants [119]. IEC separations are popular within the industry because they use commonly available and reliable HPLC instruments and perform consistently. IEC profiles are also amenable to characterization because fractions can be collected from analytical or preparative scale separations and subjected to downstream analysis (e.g., LC/MS, peptide mapping, potency, N-terminal sequencing). A strategy for the assignment of peaks from a WCX separation has been published [118]. One negative aspect of IEC is that method development can be labor intensive and is usually specific for each protein product. On the other hand, IEC separation based on pH-gradient elution correlates well with proteins' pI, thus allowing for the development of generic separation conditions that can be applied to many therapeutic proteins [119].

***10.3.3.2 Isoelectric Focusing*** Isoelectric focusing is a high-resolution technique that can separate protein isoforms with very minor differences in their pI. Under the influence of an electric field, charge variants migrate through a pH gradient until they reach a position in which they have no net charge. IEF separations can be performed in gels with immobilized pH gradients or in capillaries with self-forming gradients. Both formats work well, but IEF gels have been replaced with capillary IEF (CIEF) for the same reasons that SDS-PAGE has been replaced by CE-SDS. Automation is not possible with gels, and CIEF is more quantitative because of the on-capillary detection. Examples of the development and validation of CIEF methods for mAbs have been published [120–122].

Recently developed imaged capillary isoelectric focusing (ICIEF) technology uses a different capillary format. The separation takes place in a short, transparent capillary that enables detection by a digital camera. The short focusing distance and lack of a mobilization step reduce the analysis time of ICIEF. Method development and validation of ICIEF methods for therapeutic mAbs have also been described [123,124].

Capillary IEF methods offer several benefits compared with IEC. They are typically faster, provide higher resolution, and can be generic for many different

antibodies. A significant drawback of CIEF methods is that they require specialized instrumentation, and the charge profiles are difficult to characterize because of the nanoliter separation volumes (i.e., fractions are difficult to collect). IEC fractions have been collected, characterized, and analyzed by ICIEF in an effort to characterize the charge profile and correlate the two different separations [124]. A better approach to characterizing CIEF charge profiles is to use a preparative IEF technique. Several different platforms for preparative IEF exist [125]; however, there are few examples of the application of preparative IEF methods for the characterization of mAb charge isoforms [126,127].

### 10.3.4 Hydrophobicity Profiling

HPLC is one of the most frequently used techniques for characterization of biotherapeutics. Besides SEC and IEC methods mentioned earlier, other commonly used chromatographic profiling methods include RP [128] and hydrophobic interaction chromatography (HIC) [129,130].

Reduced RP chromatography has been a popular method for the characterization and identification of clips and variants [128]. Stackhouse et al. [128] have developed a high-throughput UHPLC method with a 1.7-μm phenyl column. This UHPLC method allows quantitation of hinge clipping in an IgG1 molecule and acid-induced Asp/proline (D/P) clip in an IgG2 molecule. The results from the UHPLC method were comparable to those obtained with reduced CE-SDS. Additionally, the phenyl column offered partial resolution of oxidation and other chemical modifications, making this technique an attractive method for high-throughput process characterization and formulation screens.

Valliere-Douglass et al. [129] described the use of HIC to separate and characterize variants of mAbs resulting from variable N- and C-terminal processing, stress-induced covalent modifications, and conformationally altered populations. Boyd et al. presented a case study using HIC to separate an oxidized Trp-containing variant, oxidized methionine (Met) and isomerization/deamidation variants in an IgG1 [130].

## 10.4 BIOPHYSICAL PROPERTIES (HIGHER ORDER STRUCTURE AND STABILITY)

Biophysical techniques such as circular dichroism (CD), second-derivative UV spectroscopy, differential scanning calorimetry (DSC), Fourier-transform infrared spectroscopy (FTIR), AUC, and fluorescence spectroscopy are used to measure the higher order structure of proteins. Because these methods are well suited to measuring changes in protein structure and conformation relative to an initial condition, they are useful for measuring antibody stability in different formulations or conditions [131,132]. These biophysical techniques are not commonly applied to the routine characterization of mAbs because they are not sensitive to small structural changes that may only occur in a small fraction of the population. There are, however, many published examples of the application of biophysical techniques

to the structural analysis of enriched populations of antibody variants. Some examples follow:

- Analytical and biophysical (CD, tryptophan fluorescence, DSC) techniques were used to demonstrate that modification of a cysteine residue in the CDR region resulted in changes to the tertiary but not secondary structure of an antibody [133].
- CD and second-derivative UV spectroscopy were used to demonstrate that an IgG1 dimer had identical secondary structure compared with the monomer [134].
- Differences in the tertiary structure of IgG2 disulfide isoforms were observed by AUC, DSC, and CD experiments [135].
- CD data linked decreased stability with increased aspartate isomerization in two closely related antibodies. X-ray crystallography was also used to demonstrate that solvent accessibility increased Asp reactivity [136].
- Aggregation of a therapeutic antibody was characterized by light scattering, CD, and fluorescence spectroscopy [137].
- FTIR analysis of mAb particles that were generated by agitation indicated that the secondary structure was relatively unchanged from the native mAb [138].

## 10.5 OTHER POSTTRANSLATIONAL MODIFICATIONS AND PRODUCT-RELATED IMPURITIES

Other than glycosylation, common modifications observed in therapeutic IgGs include N-terminal glutamine (Gln) conversion to pyroglutamic acid (pGlu) [139–142], C-terminal Lys enzymatic processing [143] and C-terminal amidation [144], deamidation and isomerization [118,145–148], oxidation [110,135,149–154], and glycation [155–161]. This section reviews these common posttranslational and chemical modifications and the analytical methods used for their separation, detection, and structural elucidation.

### 10.5.1 N-Terminal Pyroglutamic Acid Formation

It is well known that the N-termini of mAbs can be modified through acetylation, formylation, and pyroglutamylation [162]. Among these, pGlu formation is of special interest because most mAbs have Gln or glutamic acid (Glu) at their N-terminus. There are three key suggestions in the literature as to the origin of this N-terminal variant, namely (1) enzyme glutaminyl cyclase in the cell culture process [140,163–166], (2) nonenzymatic spontaneous conversion [142], and (3) analytical artifacts [110,142]. The enzymatic reaction in the cell culture process plays a major role, leading to the most conversion. In the nonenzymatic case, the pseudo first-order reaction is driven by the formation of a product with a relatively stable five-member ring structure and the release of ammonia (Fig. 10.1). The

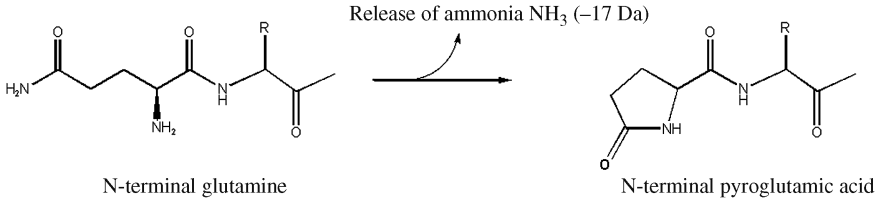


**FIGURE 10.1** Mechanism of formation of pyroglutamic acid from N-terminal glutamine.

reaction rate depends on temperature; pH; and, to a lesser extent, conductivity. Certain buffers, such as phosphate, may act as catalysts to increase the pGlu formation rate. The conditions presented in most biopharmaceutical processes can accelerate the formation of this variant. Therefore, nearly complete conversion (>95%) of N-terminal Gln to pGlu is commonly observed for some antibodies. N-terminal Gln may also convert to pGlu as an assay artifact. Rehder et al. [110] measured the N-terminal pGlu conversion of a degraded IgG1 using tryptic and Glu–C peptide map analyses by LC–MS, and his data suggested that N-terminal Gln may convert to pGlu during peptide map analysis, which is typically conducted at a pH of 7.5 and 37°C. This condition facilitates the formation of pGlu on the N-termini of unfolded peptides. Therefore, to accurately quantitate the content of pGlu/Gln species, an analytical strategy has to be carefully designed to minimize the unintended conversion during analysis.

Nonenzymatic formation of pGlu from N-terminal Glu has also been reported for recombinant mAbs [139,140,148], but this chemical change usually occurs under much harsher conditions, happens less often, and occurs at a slower rate than the formation of pGlu from Gln. This difference is likely attributable to the unfavorable hydrogen bonding propensity of Glu compared with Gln [167].

Formation of N-terminal pGlu of a protein or peptide from Gln or Glu decreases the MW by 17 Da or 18 Da, respectively. More significantly, cyclization of Gln to form pGlu causes a loss of the positively charged N-terminal primary amine, which results in an acidic shift. When separated by charge-based analytical methods, such as IEC and CIEF, mAbs with pGlu are usually the main peak, but the antibodies with Gln elute in the basic region. The identification and quantitation of pGlu-containing variants can also be done via LC/MS/MS after enzymatic digestion. Using LC coupled with electrospray ionization on a quadrupole TOF mass spectrometer (Q-TOF), Yu et al. [168] quantified the levels of pGlu at the N-terminus of the light chain of a recombinant mAb. These quantitative results were also comparable to those obtained with a more conventional peptide mapping method. Rehder et al. [110] developed an RP LC/MS method for reduced antibodies that provided efficient separation of the light chain and two variants of the heavy chain containing N-terminal Gln and pGlu. Yan et al. [141] used an RPLC–MS method to demonstrate that about 99% of the N-terminus of the heavy chain of a recombinant IgG1 contains pGlu, and about 2% of the light chain N-terminus contains pGlu [141]. Further proof of the N-terminal pGlu presence may be achieved by digestion of the mAb with

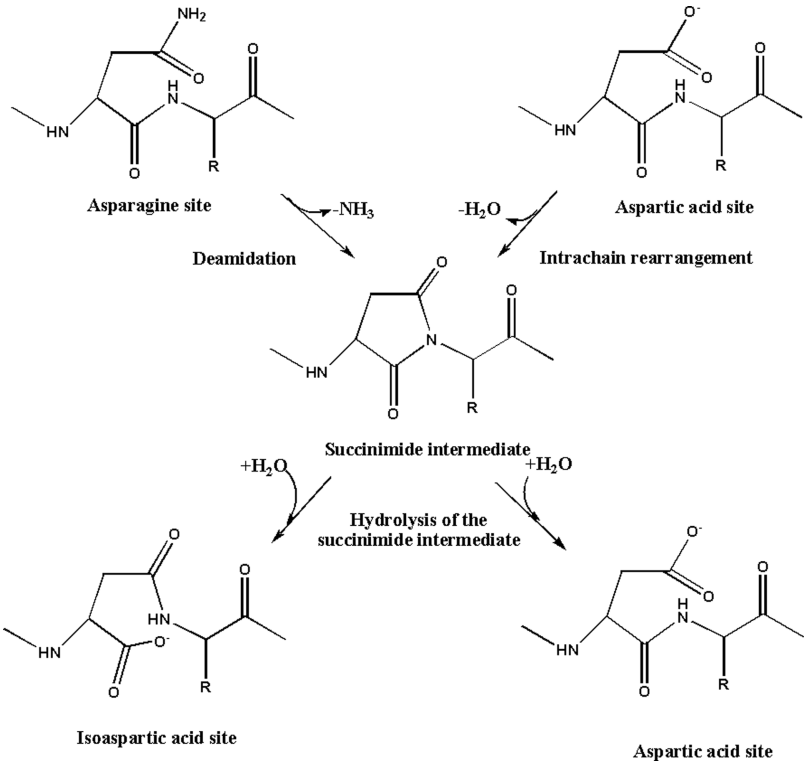


**FIGURE 10.2** Illustration of the formation of succinimide intermediate from asparagine or aspartic acid (Asp) residues and subsequent hydrolysis to isoaspartate or Asp.

pyroglutamate aminopeptidase. Samples that have been treated with this enzyme can then be analyzed using MS [148] or Edman sequencing [169].

### 10.5.2 C-Terminal Heterogeneity

***10.5.2.1 Lysine Variants*** C-terminal lysine (Lys) and arginine (Arg) residues, whose presence was expected based on gene sequence information, are often absent in proteins isolated from mammalian cell culture. This modification was first reported by Harris et al. [143,170]. This loss of Lys is presumed to be attributable to the activity of a basic carboxypeptidase in the host cells [143,171]. The degree of C-terminal Lys heterogeneity reflects the manufacturing consistency of recombinant IgGs and may be monitored for product consistency [172]. The presence or absence of C-terminal Lys residues on antibody heavy chains does not influence antigen binding [173], which is mediated by the distant CDRs. Binding to the Fcγ receptor or complement Clq is also unaffected because these interactions involve residues in the hinge-CH2 region and CH2 domains, respectively [143]. To date, the latter assumption has not been confirmed by testing the bioactivity and bioavailability *in vivo*.

Because the removal of C-terminal Lys reduces one positive charge, this charge difference can be resolved and monitored by charge based methods, such as IEF and IEC. Antibodies with different numbers of C-terminal Lys residues can also be separated by HIC [136] and RP [108]. The relative percentage of C-terminal Lys containing peptide can be determined using RP-HPLC after cyanogen bromide (CNBr) cleavage of the antibody [143]. Treatment with carboxypeptidase B (CpB) readily removes C-terminal Lys and reduces the charge heterogeneity caused by the presence of 0, 1, or 2 C-terminal Lys residues.

The removal of one C-terminal Lys residue also decreases the MW by 128 Da. Therefore, mass-based methods, such as MALDI-TOF-MS and LC–ESI-MS, can be used directly to resolve the variants with different numbers of C-terminal Lys residues on an intact mAb after N-linked glycans are enzymatically removed using PNGase F [143,171,174]. Dick et al. [171] determined that tryptic peptide mapping with LC–MS was reproducible, provides the relative percentage of Lys variants, and has significant advantages over charge-based methods.

***10.5.2.2 Amidation*** C-terminal amidation after Lys and subsequent glycine (Gly) removal is generally found in bioactive peptide hormones [175] and was observed in a monoclonal IgG1 antibody [144]. Many bioactive peptides must be amidated at their carboxy terminus to exhibit full activity. The hormones are synthesized from Gly-extended intermediates that are transformed into active amidated hormones by oxidative cleavage of the Gly N–C α bond. In higher organisms, this reaction is catalyzed by a single bifunctional enzyme, peptidylglycine α-amidating monooxygenase (PAM). Two redox-active copper atoms play a crucial role in this Gly amidation process. After being reduced by ascorbate, they catalyze the reduction of molecular oxygen for the hydroxylation of Gly-extended substrates [176]. The enzymatic mechanism requires a consensus sequence site, −XG(B), in which the amino acid to be amidated, X, is followed by Gly, G, which donates the amide group. Gly may be followed by a basic (B) amino acid [144]. In IgG1 mAbs, this consensus sequence site is present at the C-terminus of the HC amino acid sequence, SLSLS**PGK**. Studies also demonstrated that HC C-terminal amidation has no effect on Fab binding activity as measured by enzyme-linked immunosorbent assay (ELISA) and no observed differences in either effector function or antigen binding using a specific cell-based assay when main peak was compared with basic peak species [144].

Because amidation adds one positive charge to the molecule, this variant appears in the basic region of charge-based separation methods. It usually co-elutes or migrates with the C-terminal Lys-containing species. Complete removal of C-terminal Lys via CpB digestion will reveal the variants containing amidated species on the charge-based assays. The theoretical monoisotopic mass difference for the amidated peptide and the unmodified peptide ending in proline is −0.98 Da. Together with the loss of Gly residue, the mass difference between the amidated species and the nonamidated one ending in Gly is −58 Da. This mass difference can be readily resolved by reduced LC–MS. LC/MS/MS after enzymatic digestion can be used to further confirm the C-terminal peptide sequence and amidation [144].

### 10.5.3 Deamidation and Isomerization

Deamidation of Asn and isomerization of Asp are common modifications of proteins. Deamidation or isomerization proceeds via a succinimide intermediate, with the favored product being isoAsp [177,178]. Formation of atypical isoaspartyl sites via the deamidation-linked isomerization of asparaginyl peptide bonds or direct isomerization of aspartyl peptide bonds is a major contributor to spontaneous protein degradation under mild conditions. This nonenzymatic reaction reroutes the Asx–Xaa peptide bond through the β-carbonyl of asparaginyl or aspartyl residues, thereby adding an extra carbon to the polypeptide backbone. Formation of isoAsp has been implicated in protein inactivation, aggregation, degradation, and autoimmunity [118,179–182]. The succinimide intermediate is typically not stable at neutral and basic pH, but its presence has been reported for multiple mAbs [118,183].

Deamidation and isomerization occur both *in vitro* and *in vivo*. The reaction rate depends on many factors, including primary structure, three-dimensional structure, pH, temperature, ionic strength, buffer ions, and other solution properties [147,184,185]. Deamidation of Asn residues is observed most frequently at Asn–Gly and Asn–Ser sites within proteins [186], but it has recently been observed at the Asn–Asn site in the CDR region of a mAb. Deamidation can also be introduced during sample preparation for peptide mapping [170].

Although Asn is the most frequently modified residue, Gln has also been reported to undergo deamidation through the formation of a six-member ring. Gln deamidation occurs much less frequently than Asn deamidation and Asp rearrangement [187].

At neutral pH, Asn deamidation to Asp or isoAsp introduces a negative charge to antibodies and generates acidic species, which results in an earlier elution on cation exchange chromatography and an acidic shift by IEF methods [8,123,188]. Succinimide is more basic and less polar than aspartate and isoAsp and therefore elutes later on cation exchange and HIC columns [118]. A remarkable resolution between Asp- and isoAsp-containing species is possible with cation exchange chromatography given the trivial difference in charge. This separation may be caused by isomerization changing the structure in such a way that the isoAsp-containing variants present a difference in surface charge features [118].

Although charge-based methods are widely used to detect and quantitate deamidation, several methods have been used to identify deamidation sites [141,147]. Among them, tryptic peptide mapping via LC/MS/MS is used ubiquitously [8,11,147,148,188–190]. The advent of high-resolution mass spectrometers in recent years makes it possible to distinguish the parent peptide from the deamidated peptide with a mass difference of only 1 Da. Tandem MS via collision-induced dissociation (CID), which cleaves N–CO peptide bonds to produce b/y fragment ions, provides peptide sequence information and facilitates the identification of deamidation sites. Complementary to CID, electron transfer dissociation (ETD) uses low-energy electrons to irradiate the ions of interest to produce charge reduced radical ions. ETD cleaves the N–C α bonds to produce c/z fragments to allow for protein sequence analysis and amino acid side chain modification identification [191–196].

The relatively nonselective ETD fragmentation pattern makes the distinction between isoAsp and Asp residues possible using MS. Asp-N digestion has also been successfully used to distinguish aspartyl from isoaspartyl residues based on the relatively high resistance of the isoaspartyl form to Asp-N digestion [118,190,197].

Traditional N-terminal sequencing via Edman degradation is still useful not only to identify deamidation sites but also to provide information about the nature of the deamidation product, whether it is Asp, isoAsp, or succinimide. In the case of isoAsp or succinimide, no further sequence information will be observed because of the β-aspartyl residue or the five-member ring structure in the backbone, which prohibits Edman chemistry [198,199]. For the succinimide intermediate, Kwong et al. [200] determined that incubation in 2-M hydroxylamine and 0.2-M Tris buffer at a pH of 9 for 2 h at 45°C without denaturants (e.g., guanidine) will specifically cleave on the C-terminal side of succinimides without cleavage at Asn–Gly bonds. N-terminal sequence analysis can then be used to identify the internal sequence generated by cleavage of the succinimide, hence identifying the succinimide site. Care must be taken when using hydroxylamine cleavage. In earlier work done by Canova-Davis et al. [201] on the fusion protein rhIGF-I (recombinant human insulin like growth factor I), it was shown that hydroxylamine cleavage also produced hydroxamates of the Asn and Gln residues as a side product.

Enzymatic methods have been used to detect and quantitate isoAsp in peptides and proteins [146,202,203]. The enzyme protein L-isoaspartyl-*O*-methyltransferase (PIMT) catalyzes the transfer of a methyl group from *S*-adenosyl methionine (SAM) to the α-carboxyl group of an isoAsp residue and generates S-adenosyl homocysteine (SAH) as a byproduct. The methylated peptide species can spontaneously demethylate to form a succinimide intermediate and release methanol. A commercially available kit ISOQUANT from Promega (Madison, WI, USA) provides a protocol to detect the byproduct SAH using RP-HPLC. A combination of the PIMT enzymatic method and alkaline hydroxylamine cleavage can be used to selectively cleave isoAsp linkages in high yield. Location of the isoAsp linkage is then inferred by identifying the cleavage products by MS or N-terminal Edman sequencing [180].

### 10.5.4 Oxidation

Protein oxidation is a covalent modification of an amino acid that is induced by reactive oxygen species and is often a result of stress conditions or contamination [141,204]. In human antibodies, methionine residues are most susceptible to oxidation. However oxidation can also occur on cysteine, tryptophan, tyrosine, and lysine residues [150,205,206]. In IgG1, whereas Met residues at positions 255 and 431 in the Fc region are the most likely to be oxidized to methionine sulfoxide, Met residues in the Fab region were relatively stable [141,154,206]. Tryptophan oxidation was first reported by Finley et al. [152] in α-crystallin exposed to the products of a nonenzymatic, metal-catalyzed oxidation system (the Fenton reaction). The structures of oxidation products in peptides and proteins were identified. A recent study showed that oxidation of Trp 94 and deamidation of Asn 93, located in the light chain CDR of a therapeutic IgG1 mAb, correlates with loss of bioactivity upon photodegradation [207].

For therapeutic antibodies, oxidation is typically studied with thermal, chemically stressed and UV-irradiated samples [153,206,208]. Samples are either stored at an elevated temperature for extended periods of time or in the presence of oxidizing agents such as *tert*-butylhydroperoxide (tBHP), hydrogen peroxide ($H_2O_2$), or a free-radical generator, 2,2′-azobis(2-amidinopropane) dihydrochloride (AAPH) [209]. Photo irradiation with UV light also causes oxidation.

Oxidized species can be separated by a variety of chromatographic techniques, such as RP-HPLC, IEC, and HIC. Because oxidation likely occurs on the surface of the protein, it may change the overall charge distribution, resulting in subtle changes in peak shape or even resolved peaks. When separated by weak IEC, antibodies with oxidized Met residues eluted later [206]. Lam et al. [154] described the analysis of methionine oxidation in an antibody digested with papain and separated by HIC. The RP separation of oxidized antibody fragments after papain digestion was also reported [141]. Column chemistries with aromatic selectivity, such as phenyl stationary phases, were successfully used for separating tryptophan-oxidized species of antibodies. Yang et al. [151] developed an RP-HPLC method for reduced mAbs that separates the heavy chain containing oxidized tryptophan from the nonoxidized heavy chain.

Although chromatographic techniques are useful for quantifying oxidation, they do not provide site-specific information on their own. Peptide map analysis by LC/MS/MS is overwhelmingly used to identify the oxidation sites. Oxidized peptides usually elute earlier than the corresponding native peptides in a RP-HPLC peptide map. Methionine oxidation to methionine sulfoxide causes a mass shift of +16 Da. Yan et al. also observed a −64 Da fragment as the major signal in a MS/MS spectrum of the oxidized DSM(O)LISR peptides in a recombinant human monoclonal IgG1 antibody caused by the elimination of –SOH–CH3 fragment from Met side chains during MS/MS sequencing [141]. Because the formation of a sulfoxide is the main product of Met oxidation, this modification is usually observed as a discrete peak in UV peptide maps. In contrast, oxidized tryptophan has many products, including hydroxy-tryptophan (+16), kynurenine (+4), and *N*-formylkynurenine (+32) [152]. The multiple tryptophan products dilute the oxidized peptide intensity, thus making peptide mapping via UV detection alone a less sensitive choice to detect tryptophan oxidation.

### 10.5.5 Glycation

Glycation is a posttranslational chemical reaction that results in the nonenzymatic glycosylation of the N-terminal primary amine or the ε-amine group of Lys side chains by reducing sugars such as glucose and lactose. A primary amine in the protein reacts with the aldehyde group of glucose to form a labile aldimine (Schiff base) derivative of the protein. This intermediate slowly isomerizes to the more stable ketoamine adduct via the Amadori rearrangement [55,158,160,210]. Glycation of therapeutic antibodies can occur during the cell culture process [155,157] and during formulation and storage [161]. Zhang et al. [157] observed that varying levels of glycation (up to 90%) at a light chain hotspot in a mAb, and its 12 recombinant

mutants manufactured in small-scale transient CHO (Chinese hamster ovary) cell lines were related to glucose concentration in the cell culture media. Sequence alignment suggested that the coexistence of LC-Asp31 and LC-Lys49 in a mAb lead to the high rate of glycation. The substitution of either residue leads to decreased glycation. The LC-Asp31 residue acts as an internal catalyst, facilitating the Amadori rearrangement of the aldimine adduct on LC-Lys49 [157]. Harris [149] reported that glycation of Lys residues occurred in the light chain and the Fab portion of the heavy chain. A minor peak detected in a deglycosylated, humanized mAb by MS was also hypothesized to be partially caused by glycation [11]. In a separate study [76], an additional peak was observed in the mass spectrum of the light chain after papain digestion and DTT reduction. The mass of this peak is 162 Da higher than that of the light chain. Studies with different cell culture conditions, including media glucose concentrations, suggests that the +162 Da modification was most likely attributable to glycation rather than O-linked glycosylation because the levels of the +162 Da modification correlated with the concentration of glucose in the cell culture medium [76]. Andya et al. [161] reported that a spray-dried powder formulation of a recombinant humanized mAb (rhumAb) with a 300:1 lactose to antibody molar ratio stored at 30°C for 9 months was completely glycated at no less than one site, analyzed by boronate affinity chromatography. Tryptic map and MS analysis of the retained peak from boronate chromatography indicated that 16 of a total of 44 Lys residues in the primary structure were lactosylated. Although glycation is a commonly observed modification, low-level glycation on a Lys located within a heavy chain CDR of an IgG1 mAb did not significantly affect its binding potency [155].

The glycated variants can be separated and quantitated by boronate affinity chromatography using optimized shielding reagents [155]. The isolated glycated material can then be characterized by peptide mapping analysis with LC/MS/MS. Glycation masks the positive charge on the N-terminal primary amino group or on the side chain of Lys residues by converting the primary amine to a secondary amine. This conversion makes the glycated antibody slightly more acidic than the unmodified antibody. Therefore, charge-based separation techniques such as CIEF and IEC are sometimes used for measuring glycation in the absence of other acidic variants [155].

## 10.6 AGGREGATES AND PARTICULATES

Proteins commonly exhibit a concentration-dependent propensity for self-association, which often leads to the formation of aggregates and particulates (all referred as aggregates herein) that range in size from nanometers (oligomers) to microns (subvisible and visible particles). A standardized nomenclature and classification scheme that can be applied to describe all protein aggregates was proposed [211]. Five categories are described under which a given aggregate may be independently classified (Table 10.2).

Although SEC is the method of choice for quantitating soluble aggregates, the most common technique for quantitating subvisible particles between 2 and 100

**TABLE 10.2 Classifications of Particles and Aggregates [211]**

| Category | Classification |
|---|---|
| Size | Submicron |
| | 1–100 μm |
| | >100 μm |
| Reversibility | Reversible |
| | Irreversible |
| | Dissociable |
| | Nondissociable |
| | Dissociable under physiological conditions |
| Secondary and tertiary structure | Folded |
| | Partially unfolded |
| | Unfolded |
| | Amyloid |
| Covalent nodification | Cross-linked |
| | Intramolecular chemical modification |
| Morphology | Aspect ratio |
| | Surface roughness |
| | Optical properties |
| | Internal morphology |

microns is the light obscuration method [212]. This technique can determine the size and the amount of particles but cannot differentiate among the types of particles, such as protein particles, foreign material, micro bubbles, or silicone oil droplets, that can be present in protein solutions. The microflow imaging techniques can determine morphology and optical characteristics of the particles but still not identify the chemical composition. Multiple characterization and identification methods are available, such as scanning electron microscope–energy dispersive spectroscopy (SEM–EDS), micro-X-ray fluorescence, Raman microscopy (with different detection systems), microspectroscopy, microchemistry, micro-FTIR, and "rapid Identity" [212]. Such technologies can detect extremely small quantities of elements and components. The primary method for quantitating visible particles is visual inspection, a method that can be subjective and relies on adequate training of the human inspectors. Automated methods for visible particle determination are being developed.

Aggregates are a major safety concern because of their putative ability to trigger an immune response [213–215]. One ultimate goal of aggregate and particulate analysis is to ensure product consistency while minimizing the safety risk to patients. Thus, it is important to understand what is in the products and whether those aggregates present a safety issue. It is necessary to develop a particle characterization program to help address concerns during process and formulation development and ultimately, to minimize aggregate formation.

## 10.7 SUMMARY

A major task for characterization scientists in the biopharmaceutical industry is to fully characterize heterogeneous protein products in terms of their physicochemical structure; their extent of modification; and their impact on safety, efficacy, and PK. Nearly all of the major variants described in this chapter may be introduced during the manufacturing process, storage, and sample analysis [216]. With recent advances in analytical techniques and instrumentation, one can examine these variants with much greater sensitivity, resolution, accuracy, and confidence. This chapter has covered the major aspects that should be considered when characterizing a typical protein therapeutic, but there are still new variants being discovered such as the existence of trisulfide bonds [38], C-mannosylation [217], amino acid misincorporation [26,218], and N-glycosylation at nonconsensus sites [219]. Therefore, characterization scientists ought to be vigilant for new modifications.

The impact of these variants on safety, efficacy, immunogenicity, and PK cannot be generalized and must be assessed on a case-by-case basis. This usually involves additional toxicology, potency, or animal model studies. These studies can be used to link specific physicochemical attributes with biological activity and provide justification for the assignment of critical quality attributes (CQAs). The CQA knowledge is essential in understanding the critical process parameters, and designing overall control system strategies [220].

## ACKNOWLEDGMENTS

The authors would like to thank Patricia Rancatore, Boyan Zhang, Jennifer Rea, and Victor Ling for their critical review of this chapter.

## REFERENCES

1. Kozlowski S, Swann P. Current and future issues in the manufacturing and development of monoclonal antibodies. Adv Drug Deliv Rev 2006;58(5–6):707–722.
2. FDA. Well-characterized biotechnology products; elimination of establishment license application. Federal Register 1996;61(19):2733–2739.
3. Henry C. FDA, reform, and the well-characterized biologic. Anal Chem 1996; 68(21):674A–677A.
4. International Conference on Harmonisation Guideline. Specifications: test procedures and acceptance criteria for biotechnological/biological products (Q6B). Federal Register 1999;64:44928.
5. Bussel JB, Giulino L, Lee S, Patel VL, Sandborg C, Stiehm ER 2007. Update on therapeutic monoclonal antibodies. Curr Probl Pediatr Adolesc Health Care 37(4): 118–135.
6. International Conference on Harmonisation Guideline: Quality risk management (Q9). Federal Register 2006;71(106):32105–32106.

7. Harris RJ, Shire SJ, Winter C. Commercial manufacturing scale formulation and analytical characterization of therapeutic recombinant antibodies. Drug Devel Res 2004;61(3):137–154.
8. Kroon DJ, Baldwin-Ferro A, Lalan P. Identification of sites of degradation in a therapeutic monoclonal antibody by peptide mapping. Pharm Res 1992;9(11):1386–1393.
9. Beck A, Bussat MC, Zorn N, Robillard V, Klinguer-Hamour C, Chenu S, et al. Characterization by liquid chromatography combined with mass spectrometry of monoclonal anti-IGF-1 receptor antibodies produced in CHO and NS0 cells. J Chromatogr B Analyt Technol Biomed Life Sci 2005;819(2):203–218.
10. Wan M, Shiau FY, Gordon W, Wang GY. Variant antibody identification by peptide mapping. Biotechnol Bioeng 1999;62(4):485–488.
11. Lyubarskaya Y, Houde D, Woodard J, Murphy D, Mhatre R. Analysis of recombinant monoclonal antibody isoforms by electrospray ionization mass spectrometry as a strategy for streamlining characterization of recombinant monoclonal antibody charge heterogeneity. Anal Biochem 2006;348(1):24–39.
12. Harris RJ, Murnane AA, Utter SL, Wagner KL, Cox ET, Polastri GD, et al. Assessing genetic heterogeneity in production cell lines: detection by peptide mapping of a low level Tyr to Gln sequence variant in a recombinant antibody. Biotechnology (NY) 1993;11(11):1293–1297.
13. Guo D, Gao A, Michels DA, Feeney L, Eng M, Chan B, et al. Mechanisms of unintended amino acid sequence changes in recombinant monoclonal antibodies expressed in Chinese hamster ovary (CHO) cells. Biotechnol Bioeng 2010;107(1):163–171.
14. Schegg KM, Denslow ND, Andersen TT, Bao YA, Cohen SA, Mahrenholz AM, et al. Quantitation and identification of proteins by amino acid analysis. In: Marshak DR, editor. Techniques in Protein Chemistry. San Diego: Academic Press; 1997, p. 207–216.
15. Macchi FD, Shen FJ, Keck RG, Harris RJ. Amino acid analysis, using postcolumn ninhydrin detection, in a biotechnology laboratory. Methods Mol Biol 2000;159:9–30.
16. Strydom DJ, Cohen SA. Comparison of amino acid analyses by phenylisothiocyanate and 6-aminoquinolyl-*N*-hydroxysuccinimidyl carbamate precolumn derivatization. Anal Biochem 1994;222(1):19–28.
17. Schuster R. Determination of amino acids in biological, pharmaceutical, plant and food samples by automated precolumn derivatization and high-performance liquid chromatography. J Chromatogr 1988;431(2):271–284.
18. Boogers I, Plugge W, Stokkermans YQ, Duchateau AL. Ultra-performance liquid chromatographic analysis of amino acids in protein hydrolysates using an automated pre-column derivatisation method. J Chromatogr 2008;1189(1–2):406–409.
19. Salazar C, Armenta JM, Cortes DF, Shulaev V. Combination of an AccQ. Tag-ultra performance liquid chromatographic method with tandem mass spectrometry for the analysis of amino acids. Methods Mol Biol 2012;828:13–28.
20. Shively JE. The chemistry of protein sequence analysis. EXS 2000;88:99–117.
21. Pham V, Henzel WJ, Arnott D, Hymowitz S, Sandoval WN, Truong BT, et al. De novo proteomic sequencing of a monoclonal antibody raised against OX40 ligand. Anal Biochem 2006;352(1):77–86.
22. Hoff ER, Chloupek RC. Analytical peptide mapping of recombinant DNA-derived proteins by reversed-phase high-performance liquid chromatography. Methods Enzymol 1996;271:51–68.

23. Kannan K, Mulkerrin MG, Zhang M, Gray R, Steinharter T, Sewerin K, et al. Rapid analytical tryptic mapping of a recombinant chimeric monoclonal antibody and method validation challenges. J Pharm Biomed Anal 1997;16(4):631–640.
24. Dick LW Jr, Mahon D, Qiu D, Cheng KC. Peptide mapping of therapeutic monoclonal antibodies: improvements for increased speed and fewer artifacts. J Chromatogr B Analyt Technol Biomed Life Sci 2009;877(3):230–236.
25. Yang Y, Strahan A, Li C, Shen A, Liu H, Ouyang J, et al. Detecting low level sequence variants in recombinant monoclonal antibodies. mAbs 2(3):285–928.
26. Yu XC, Borisov OV, Alvarez M, Michels DA, Wang YJ, Ling V. Identification of codon-specific serine to asparagine mistranslation in recombinant monoclonal antibodies by high-resolution mass spectrometry. Anal Chem 2009;81(22):9282–9290.
27. Zybailov B, Sun Q, van Wijk KJ. Workflow for large scale detection and validation of peptide modifications by RPLC-LTQ-Orbitrap: application to the *Arabidopsis thaliana* leaf proteome and an online modified peptide library. Anal Chem 2009;81(19):8015–8024.
28. Henzel WJ, Stults JT, Hsu CA, Aswad DW. The primary structure of a protein carboxyl methyltransferase from bovine brain that selectively methylates L-isoaspartyl sites. J Biol Chem 1989;264(27):15905–15911.
29. Henzel WJ, Watanabe C, Stults JT. Protein identification: the origins of peptide mass fingerprinting. J Am Soc Mass Spectrom 2003;14(9):931–942.
30. Zhang B, Cockrill SL. Methodology for determining disulfide linkage patterns of closely spaced cysteine residues. Anal Chem 2009;81(17):7314–7320.
31. Zhang W, Marzilli LA, Rouse JC, Czupryn MJ. Complete disulfide bond assignment of a recombinant immunoglobulin G4 monoclonal antibody. Anal Biochem 2002;311(1):1–9.
32. Bloom JW, Madanat MS, Marriott D, Wong T, Chan SY. Intrachain disulfide bond in the core hinge region of human IgG4. Protein Sci 1997;6(2):407–415.
33. Lee HG. High-performance sodium dodecyl sulfate-capillary gel electrophoresis of antibodies and antibody fragments. J Immunol Methods 2000;234(1-2):71–81.
34. Wypych J, Li M, Guo A, Zhang Z, Martinez T, Allen MJ, et al. Human IGG2 antibodies display disulfide mediated structural isoforms. J Biol Chem 2008;283(23):16194–16205.
35. Guo A, Han M, Martinez T, Ketchem RR, Novick S, Jochheim C, et al. Electrophoretic evidence for the presence of structural isoforms specific for the IgG2 isotype. Electrophoresis 2008;29(12):2550–2556.
36. Tous GI, Wei Z, Feng J, Bilbulian S, Bowen S, Smith J, et al. Characterization of a novel modification to monoclonal antibodies: thioether cross-link of heavy and light chains. Anal Chem 2005;77(9):2675–2682.
37. Cohen SL, Price C, Vlasak J. Beta-elimination and peptide bond hydrolysis: two distinct mechanisms of human IgG1 hinge fragmentation upon storage. J Am Chem Soc 2007;129(22):6976–6977.
38. Pristatsky P, Cohen SL, Krantz D, Acevedo J, Ionescu R, Vlasak J. Evidence for trisulfide bonds in a recombinant variant of a human IgG2 monoclonal antibody. Anal Chem 2009;81(15):6148–6155.
39. McAuley A, Jacob J, Kolvenbach CG, Westland K, Lee HJ, Brych SR, et al. Contributions of a disulfide bond to the structure, stability, and dimerization of human IgG1 antibody CH3 domain. Protein Sci 2008;17(1):95–106.
40. Thies MJ, Talamo F, Mayer M, Bell S, Ruoppolo M, Marino G, et al. Folding and oxidation of the antibody domain C(H)3. J Mol Biol 2002;319(5):1267–1277.

41. Gadgil HS, Bondarenko PV, Pipes GD, Dillon TM, Banks D, Abel J, et al. Identification of cysteinylation of a free cysteine in the Fab region of a recombinant monoclonal IgG1 antibody using Lys-C limited proteolysis coupled with LC/MS analysis. Anal Biochem 2006;355(2):165–174.
42. Chaderjian WB, Chin ET, Harris RJ, Etcheverry TM. Effect of copper sulfate on performance of a serum-free CHO cell culture process and the level of free thiol in the recombinant antibody expressed. Biotechnol Prog 2005;21(2):550–553.
43. Hansen RE, Winther Jr. An introduction to methods for analyzing thiols and disulfides: reactions, reagents, and practical considerations. Anal Biochem 2009;394(2):147–158.
44. Zhang W, Czupryn MJ. Free sulfhydryl in recombinant monoclonal antibodies. Biotechnol Prog 2002;18(3):509–513.
45. Chung CH, Mirakhur B, Chan E, Le QT, Berlin J, Morse M, et al. Cetuximab-induced anaphylaxis and IgE specific for galactose-alpha-1,3-galactose. N Engl J Med 2008;358 (11):1109–1117.
46. Krapp S, Mimura Y, Jefferis R, Huber R, Sondermann P. Structural analysis of human IgG-Fc glycoforms reveals a correlation between glycosylation and structural integrity. J Mol Biol 2003;325(5):979–989.
47. Jefferis R. Glycosylation of recombinant antibody therapeutics. Biotechnol Prog 2005;21(1):11–16.
48. Jefferis R. Antibody therapeutics: isotype and glycoform selection. Expert Opin Biol Ther 2007;7(9):1401–1413.
49. Raju TS, Briggs JB, Borge SM, Jones AJ. Species-specific variation in glycosylation of IgG: evidence for the species-specific sialylation and branch-specific galactosylation and importance for engineering recombinant glycoprotein therapeutics. Glycobiology 2000;10(5):477–486.
50. Lamari FN, Kuhn R, Karamanos NK. Derivatization of carbohydrates for chromatographic, electrophoretic and mass spectrometric structure analysis. J Chromatogr B Analyt Technol Biomed Life Sci 2003;793(1):15–36.
51. Patel T, Bruce J, Merry A, Bigge C, Wormald M, Jaques A, et al. Use of hydrazine to release in intact and unreduced form both N- and O-linked oligosaccharides from glycoproteins. Biochemistry 1993;32(2):679–693.
52. Mizuochi T, Hounsell EF. Methods in Molecular Biology: Release of N-Linked Oligosaccharide Chains by Hydrazinolysis. Vol. 108. Totowa, NJ: Humana Press Inc.; 1996. p. 653–656.
53. O'Neill RA. Enzymatic release of oligosaccharides from glycoproteins for chromatographic and electrophoretic analysis. J Chromatogr 1996;720(1–2):201–215.
54. Keck RG, Briggs JB, Jones AJ. Oligosaccharide release and MALDI-TOF MS analysis of N-linked carbohydrate structures from glycoproteins. Methods Mol Biol 2005;308:381–396.
55. Papac DI, Briggs JB, Chin ET, Jones AJ. A high-throughput microscale method to release N-linked oligosaccharides from glycoproteins for matrix-assisted laser desorption/ionization time-of-flight mass spectrometric analysis. Glycobiology 1998;8(5):445–454.
56. Anumula KR. High-sensitivity and high-resolution methods for glycoprotein analysis. Anal Biochem 2000;283(1):17–26.
57. Anumula KR. Advances in fluorescence derivatization methods for high-performance liquid chromatographic analysis of glycoprotein carbohydrates. Anal Biochem 2006; 350(1):1–23.

58. Anumula KR, Dhume ST. High resolution and high sensitivity methods for oligosaccharide mapping and characterization by normal phase high performance liquid chromatography following derivatization with highly fluorescent anthranilic acid. Glycobiology 1998;8(7):685–694.
59. Guile GR, Rudd PM, Wing DR, Prime SB, Dwek RA. A rapid high-resolution high-performance liquid chromatographic method for separating glycan mixtures and analyzing oligosaccharide profiles. Anal Biochem 1996;240(2):210–226.
60. Kamoda S, Nakano M, Ishikawa R, Suzuki S, Kakehi K. Rapid and sensitive screening of N-glycans as 9-fluorenylmethyl derivatives by high-performance liquid chromatography: a method which can recover free oligosaccharides after analysis. J Proteome Res 2005;4(1):146–152.
61. Yuen CT, Gee CK, Jones C. High-performance liquid chromatographic profiling of fluorescent labelled N-glycans on glycoproteins. Biomed Chromatogr 2002;16(4): 247–254.
62. Melmer M, Stangler T, Premstaller A, Lindner W. Comparison of hydrophilic-interaction, reversed-phase and porous graphitic carbon chromatography for glycan analysis. J Chromatogr 2011;1218(1):118–123.
63. Basa LJ, Spellman MW. Analysis of glycoprotein-derived oligosaccharides by high-pH anion-exchange chromatography. J Chromatogr 1990;499:205–220.
64. Hardy MR, Townsend RR. Separation of positional isomers of oligosaccharides and glycopeptides by high-performance anion-exchange chromatography with pulsed amperometric detection. Proc Natl Acad Sci USA 1988;85(10):3289–3293.
65. Prime S, Dearnley J, Ventom AM, Parekh RB, Edge CJ. Oligosaccharide sequencing based on exo- and endoglycosidase digestion and liquid chromatographic analysis of the products. J Chromatogr 1996;720(1–2):263–274.
66. Qian J, Liu T, Yang L, Daus A, Crowley R, Zhou Q. Structural characterization of N-linked oligosaccharides on monoclonal antibody cetuximab by the combination of orthogonal matrix-assisted laser desorption/ionization hybrid quadrupole-quadrupole time-of-flight tandem mass spectrometry and sequential enzymatic digestion. Anal Biochem 2007;364(1):8–18.
67. Chen X, Flynn GC. Analysis of N-glycans from recombinant immunoglobulin G by on-line reversed-phase high-performance liquid chromatography/mass spectrometry. Anal Biochem 2007;370(2):147–161.
68. Prater BD, Connelly HM, Qin Q, Cockrill SL. High-throughput immunoglobulin G N-glycan characterization using rapid resolution reverse-phase chromatography tandem mass spectrometry. Anal Biochem 2009;385(1):69–79.
69. Raju TS. Electrophoretic methods for the analysis of N-linked oligosaccharides. Anal Biochem 2000;283(2):125–132.
70. Ma S, Nashabeh W. Carbohydrate analysis of a chimeric recombinant monoclonal antibody by capillary electrophoresis with laser-induced fluorescence detection. Anal Chem 1999;71(22):5185–5192.
71. Mechref Y, Muzikar J, Novotny MV. Comprehensive assessment of N-glycans derived from a murine monoclonal antibody: a case for multimethodological approach. Electrophoresis 2005;26(10):2034–2046.
72. Kamoda S, Ishikawa R, Kakehi K. Capillary electrophoresis with laser-induced fluorescence detection for detailed studies on N-linked oligosaccharide profile of therapeutic recombinant monoclonal antibodies. J Chromatogr 2006;1133(1–2):332–339.

73. Gennaro LA, Salas-Solano O. On-line CE-LIF-MS technology for the direct characterization of N-linked glycans from therapeutic antibodies. Anal Chem 2008;80(10): 3838–3845.
74. Gennaro LA, Salas-Solano O, MA S. Capillary electrophoresis-mass spectrometry as a characterization tool for therapeutic proteins. Anal Biochem 2006;355(2):249–258.
75. Wagner-Rousset E, Bednarczyk A, Bussat MC, Colas O, Corvaia N, Schaeffer C, et al. The way forward, enhanced characterization of therapeutic antibody glycosylation: comparison of three level mass spectrometry-based strategies. J Chromatogr B Analyt Technol Biomed Life Sci 2008;872(1–2):23–37.
76. Lim A, Reed-Bogan A, Harmon BJ. Glycosylation profiling of a therapeutic recombinant monoclonal antibody with two N-linked glycosylation sites using liquid chromatography coupled to a hybrid quadrupole time-of-flight mass spectrometer. Anal Biochem 2008;375(2):163–172.
77. Sinha S, Pipes G, Topp EM, Bondarenko PV, Treuheit MJ, Gadgil HS. Comparison of LC and LC/MS methods for quantifying N-glycosylation in recombinant IgGs. J Am Soc Mass Spectrom 2008;19(11):1643–1654.
78. Xiao G, Bondarenko PV. Identification and quantification of degradations in the Asp-Asp motifs of a recombinant monoclonal antibody. J Pharm Biomed Anal 2008;47 (1):23–30.
79. Arakawa T, Ejima D, Li T, Philo JS. The critical role of mobile phase composition in size exclusion chromatography of protein pharmaceuticals. J Pharm Sci 2010;99(4): 1674–1692.
80. Hartmann WK, Saptharishi N, Yang XY, Mitra G, Soman G. Characterization and analysis of thermal denaturation of antibodies by size exclusion high-performance liquid chromatography with quadruple detection. Anal Biochem 2004;325(2):227–239.
81. Oliva A, Llabres M, Farina JB. Applications of multi-angle laser light-scattering detection in the analysis of peptides and proteins. Curr Drug Discov Technol 2004; 1(3):229–242.
82. Ye H. Simultaneous determination of protein aggregation, degradation, and absolute molecular weight by size exclusion chromatography-multiangle laser light scattering. Anal Biochem 2006;356(1):76–85.
83. Ahrer K, Buchacher A, Iberer G, Jungbauer A. Thermodynamic stability and formation of aggregates of human immunoglobulin G characterised by differential scanning calorimetry and dynamic light scattering. J Biochem Biophys Methods 2006;66(1-3): 73–86.
84. Nobbmann U, Connah M, Fish B, Varley P, Gee C, Mulot S, et al. Dynamic light scattering as a relative tool for assessing the molecular integrity and stability of monoclonal antibodies. Biotechnol Genet Eng Rev 2007;24:117–128.
85. Sukumar M, Doyle BL, Combs JL, Pekar AH. Opalescent appearance of an IgG1 antibody at high concentrations and its relationship to noncovalent association. Pharm Res 2004;21(7):1087–1093.
86. Laue T. Analytical ultracentrifugation: a powerful “new” technology in drug discovery. Drug Disov Today: Technol 2004;1(3):309–315.
87. Liu J, Andya JD, Shire SJ. A critical review of analytical ultracentrifugation and field flow fractionation methods for measuring protein aggregation. AAPS J 2006;8(3): E580–E589.

88. Berkowitz SA. Role of analytical ultracentrifugation in assessing the aggregation of protein biopharmaceuticals. AAPS J 2006;8(3):E590–E605.
89. Gabrielson JP, Randolph TW, Kendrick BS, Stoner MR. Sedimentation velocity analytical ultracentrifugation and SEDFIT/c(s): limits of quantitation for a monoclonal antibody system. Anal Biochem 2007;361(1):24–30.
90. Pekar A, Sukumar M. Quantitation of aggregates in therapeutic proteins using sedimentation velocity analytical ultracentrifugation: practical considerations that affect precision and accuracy. Anal Biochem 2007;367(2):225–237.
91. Lee HG, Chang S, Fritsche E. Rational approach to quantitative sodium dodecyl sulfate capillary gel electrophoresis of monoclonal antibodies. J Chromatogr 2002;947(1):143–149.
92. Thorne JM, Goetzinger WK, Chen AB, Moorhouse KG, Karger BL. Examination of capillary zone electrophoresis, capillary isoelectric focusing and sodium dodecyl sulfate capillary electrophoresis for the analysis of recombinant tissue plasminogen activator. J Chromatogr 1996;744(1–2):155–165.
93. Hunt G, Nashabeh W. Capillary electrophoresis sodium dodecyl sulfate nongel sieving analysis of a therapeutic recombinant monoclonal antibody: a biotechnology perspective. Anal Chem 1999;71(13):2390–2397.
94. Michels DA, Brady LJ, Guo A, Balland A. Fluorescent derivatization method of proteins for characterization by capillary electrophoresis-sodium dodecyl sulfate with laser-induced fluorescence detection. Anal Chem 2007;79(15):5963–5971.
95. Salas-Solano O, Tomlinson B, Du S, Parker M, Strahan A, Ma S. Optimization and validation of a quantitative capillary electrophoresis sodium dodecyl sulfate method for quality control and stability monitoring of monoclonal antibodies. Anal Chem 2006; 78(18):6583–6594.
96. Ma S. Analysis of protein therapeutics by capillary electrophoresis: applications and challenges. Dev Biol (Basel) 2005;122:49–68.
97. Michels DA, Parker M, Salas-Solano O. Quantitative impurity analysis of monoclonal antibody size heterogeneity by CE-LIF: example of development and validation through a quality-by-design framework. Electrophoresis 2012;33(5):815–826.
98. Lacroix M, Poinsot V, Fournier C, Couderc F. Laser-induced fluorescence detection schemes for the analysis of proteins and peptides using capillary electrophoresis. Electrophoresis 2005;26(13):2608–2621.
99. Bousse L, Mouradian S, Minalla A, Yee H, Williams K, Dubrow R. Protein sizing on a microchip. Anal Chem 2001;73(6):1207–1212.
100. Dolnik V, Liu S. Applications of capillary electrophoresis on microchip. J Sep Sci 2005;28(15):1994–2009.
101. Hawtin P, Hardern I, Wittig R, Mollenhauer J, Poustka A, Salowsky R, et al. Utility of lab-on-a-chip technology for high-throughput nucleic acid and protein analysis. Electrophoresis 2005;26(19):3674–3681.
102. Little MJ, Paquette DM, Roos PK. Electrophoresis of pharmaceutical proteins: status quo. Electrophoresis 2006;27(12):2477–2485.
103. Forrer K, Hammer S, Helk B. Chip-based gel electrophoresis method for the quantification of half-antibody species in IgG4 and their by- and degradation products. Anal Biochem 2004;334(1):81–88.
104. Ohashi R, Otero JM, Chwistek A, Hamel JF. Determination of monoclonal antibody production in cell culture using novel microfluidic and traditional assays. Electrophoresis 2002;23(20):3623–3629.

105. Vasilyeva E, Woodard J, Taylor FR, Kretschmer M, Fajardo H, Lyubarskaya Y, et al. Development of a chip-based capillary gel electrophoresis method for quantification of a half-antibody in immunoglobulin G4 samples. Electrophoresis 2004;25(21–22): 3890–3896.
106. Chen X, Tang K, Lee M, Flynn GC. Microchip assays for screening monoclonal antibody product quality. Electrophoresis 2008;29(24):4993–5002.
107. Zhang Z, Pan H, Chen X. Mass spectrometry for structural characterization of therapeutic antibodies. Mass Spectrom Rev 2009;28(1):147–176.
108. Dillon TM, Bondarenko PV, Speed Ricci M. Development of an analytical reversed-phase high-performance liquid chromatography-electrospray ionization mass spectrometry method for characterization of recombinant antibodies. J Chromatogr 2004;1053(1–2):299–305.
109. Chen G, Pramanik BN. LC–MS for protein characterization: current capabilities and future trends. Expert Rev Proteomics 2008;5(3):435–444.
110. Rehder DS, Dillon TM, Pipes GD, Bondarenko PV. Reversed-phase liquid chromatography/mass spectrometry analysis of reduced monoclonal antibodies in pharmaceutics. J Chromatogr 2006;1102(1–2):164–175.
111. Yan B, Eris T, Yates Z, Hong RW, Steen S, Kleemann G, et al. Analysis of human antibody IgG2 domains by reversed-phase liquid chromatography and mass spectrometry. J Chromatogr B Analyt Technol Biomed Life Sci 2009;877(16–17):1613–1620.
112. Kleemann GR, Beierle J, Nichols AC, Dillon TM, Pipes GD, Bondarenko PV. Characterization of IgG1 immunoglobulins and peptide-Fc fusion proteins by limited proteolysis in conjunction with LC–MS. Anal Chem 2008;80(6):2001–2009.
113. Alvarez M, Tremintin G, Wang J, Eng M, Kao YH, Jeong J, et al. On-line characterization of monoclonal antibody variants by liquid chromatography-mass spectrometry operating in a two-dimensional format. Anal Biochem 2011;419(1):17–25.
114. Pace CN, Vajdos F, Fee L, Grimsley G, Gray T. How to measure and predict the molar absorption coefficient of a protein. Protein Sci 1995;4(11):2411–2423.
115. Kuipers BJH, Gruppen H. Prediction of molar extinction coefficients of proteins and peptides using UV absorption of the constituent amino acids at 214 nm to enable quantitative reverse phase high-performance liquid chromatography–mass spectrometry analysis. J Agric Food Chem 2007;55(14):5445–5451.
116. Gill SC, von Hippel PH. Calculation of protein extinction coefficients from amino acid sequence data. Anal Biochem 1989;182(2):319–326.
117. Vlasak J, Ionescu R. Heterogeneity of monoclonal antibodies revealed by charge-sensitive methods. Curr Pharm Biotechnol 2008;9(6):468–481.
118. Harris RJ, Kabakoff B, Macchi FD, Shen FJ, Kwong M, Andya JD, et al. Identification of multiple sources of charge heterogeneity in a recombinant antibody. J Chromatogr B Biomed Sci Appl 2001;752(2):233–245.
119. Farnan D, Moreno GT. Multiproduct high-resolution monoclonal antibody charge variant separations by pH gradient ion-exchange chromatography. Anal Chem 2009;81(21):8846–8857.
120. Hunt G, Hotaling T, Chen AB. Validation of a capillary isoelectric focusing method for the recombinant monoclonal antibody C2B8. J Chromatogr 1998;800(2):355–367.
121. Hunt G, Moorhouse KG, Chen AB. Capillary isoelectric focusing and sodium dodecyl sulfate-capillary gel electrophoresis of recombinant humanized monoclonal antibody HER2. J Chromatogr 1996;744(1–2):295–301.

122. Santora LC, Krull IS, Grant K. Characterization of recombinant human monoclonal tissue necrosis factor-alpha antibody using cation-exchange HPLC and capillary isoelectric focusing. Anal Biochem 1999;275(1):98–108.
123. Janini G, Saptharishi N, Waselus M, Soman G. Element of a validation method for MU-B3 monoclonal antibody using an imaging capillary isoelectric focusing system. Electrophoresis 2002;23(11):1605–1611.
124. Sosic Z, Houde D, Blum A, Carlage T, Lyubarskaya Y. Application of imaging capillary IEF for characterization and quantitative analysis of recombinant protein charge heterogeneity. Electrophoresis 2008;29(21):4368–4376.
125. Righetti PG, Castagna A, Herbert B, Reymond F, Rossier JS. Prefractionation techniques in proteome analysis. Proteomics 2003;3(8):1397–1407.
126. Meert CD, Brady LJ, Guo A, Balland A. Characterization of antibody charge heterogeneity resolved by preparative immobilized pH gradients. Anal Chem 2010;82(9):3510–3518.
127. Timm V, Gruber P, Wasiliu M, Lindhofer H, Chelius D. Identification and characterization of oxidation and deamidation sites in monoclonal rat/mouse hybrid antibodies. J Chromatogr B Analyt Technol Biomed Life Sci 2010;878(9-10):777–784.
128. Stackhouse N, Miller AK, Gadgil HS. A high-throughput UPLC method for the characterization of chemical modifications in monoclonal antibody molecules. J Pharm Sci 2011;100(12):5115–5125.
129. Valliere-Douglass J, Wallace A, Balland A. Separation of populations of antibody variants by fine tuning of hydrophobic-interaction chromatography operating conditions. J Chromatogr 2008;1214(1–2):81–89.
130. Boyd D, Kaschak T, Yan B. HIC resolution of an IgG1 with an oxidized Trp in a complementarity determining region. J Chromatogr B Analyt Technol Biomed Life Sci 2011;879(13-14):955–960.
131. Harn N, Allan C, Oliver C, Middaugh CR. Highly concentrated monoclonal antibody solutions: direct analysis of physical structure and thermal stability. J Pharm Sci 2007;96 (3):532–546.
132. Ejima D, Tsumoto K, Fukada H, Yumioka R, Nagase K, Arakawa T, et al. Effects of acid exposure on the conformation, stability, and aggregation of monoclonal antibodies. Proteins 2007;66(4):954–962.
133. Banks DD, Gadgil HS, Pipes GD, Bondarenko PV, Hobbs V, Scavezze JL, et al. Removal of cysteinylation from an unpaired sulfhydryl in the variable region of a recombinant monoclonal IgG1 antibody improves homogeneity, stability, and biological activity. J Pharm Sci 2008;97(2):775–790.
134. Remmele RL Jr, Callahan WJ, Krishnan S, Zhou L, Bondarenko PV, Nichols AC, et al. Active dimer of Epratuzumab provides insight into the complex nature of an antibody aggregate. J Pharm Sci 2006;95(1):126–145.
135. Dillon TM, Speed Ricci M, Vezina C, Flynn GC, Liu YD, Rehder DS, et al. Structural and functional characterization of disulfide isoforms of the human IGG2 subclass. J Biol Chem 2008;283(23):16206–16215.
136. Wakankar AA, Borchardt RT, Eigenbrot C, Shia S, Wang YJ, Shire SJ, eta l. Aspartate isomerization in the complementarity-determining regions of two closely related monoclonal antibodies. Biochemistry 2007;46(6):1534–1544.
137. Demeule B, Lawrence MJ, Drake AF, Gurny R, Arvinte T. Characterization of protein aggregation: the case of a therapeutic immunoglobulin. Biochimica Biophysica Acta 2007;1774(1):146–153.

138. Brych SR, Gokarn YR, Hultgen H, Stevenson RJ, Rajan R, Matsumura M. Characterization of antibody aggregation: role of buried, unpaired cysteines in particle formation. J Pharm Sci 2010;99(2):764–781.
139. Chelius D, Jing K, Lueras A, Rehder DS, Dillon TM, Vizel A, et al. Formation of pyroglutamic acid from N-terminal glutamic acid in immunoglobulin gamma antibodies. Anal Chem 2006;78(7):2370–2376.
140. Yu L, Vizel A, Huff MB, Young M, Remmele RL Jr, He B. Investigation of N-terminal glutamate cyclization of recombinant monoclonal antibody in formulation development. J Pharm Biomed Anal 2006;42(4):455–463.
141. Yan B, Valliere-Douglass J, Brady L, Steen S, Han M, Pace D, et al. Analysis of post-translational modifications in recombinant monoclonal antibody IgG1 by reversed-phase liquid chromatography/mass spectrometry. J Chromatogr 2007;1164(1–2):153–161.
142. Dick LWJr, Kim C, Qiu D, Cheng KC. Determination of the origin of the N-terminal pyro-glutamate variation in monoclonal antibodies using model peptides. Biotechnol Bioeng 2007;97(3):544–553.
143. Harris RJ. Processing of C-terminal lysine and arginine residues of proteins isolated from mammalian cell culture. J Chromatogr 1995;705(1):129–134.
144. Johnson KA, Paisley-Flango K, Tangarone BS, Porter TJ, Rouse JC. Cation exchange-HPLC and mass spectrometry reveal C-terminal amidation of an IgG1 heavy chain. Anal Biochem 2007;360(1):75–83.
145. Zhang W, Czupryn JM, Boyle PTJr, Amari J. Characterization of asparagine deamidation and aspartate isomerization in recombinant human interleukin-11. Pharm Res 2002;19(8):1223–1231.
146. Zhang W, Czupryn MJ. Analysis of isoaspartate in a recombinant monoclonal antibody and its charge isoforms. J Pharm Biomed Anal 2003;30(5):1479–1490.
147. Chelius D, Rehder DS, Bondarenko PV. Identification and characterization of deamidation sites in the conserved regions of human immunoglobulin gamma antibodies. Anal Chem 2005;77(18):6004–6011.
148. Liu H, Gaza-Bulseco G, Sun J 2006. Characterization of the stability of a fully human monoclonal IgG after prolonged incubation at elevated temperature. J Chromatogr B Analyt Technol Biomed Life Sci 837(1–2):35–43.
149. Harris RJ 2005. Heterogeneity of recombinant antibodies: linking structure to function. Dev Biol (Basel) 122:117–127.
150. Houde D, Kauppinen P, Mhatre R, Lyubarskaya Y 2006. Determination of protein oxidation by mass spectrometry and method transfer to quality control. J Chromatogr 1123(2):189–198.
151. Yang J, Wang S, Liu J, Raghani A. Determination of tryptophan oxidation of monoclonal antibody by reversed phase high performance liquid chromatography. J Chromatogr 2007;1156(1–2):174–182.
152. Finley EL, Dillon J, Crouch RK, Schey KL. Identification of tryptophan oxidation products in bovine alpha-crystallin. Protein Sci 1998;7(11):2391–2397.
153. Usami A, Ohtsu A, Takahama S, Fujii T. The effect of pH, hydrogen peroxide and temperature on the stability of human monoclonal antibody. J Pharm Biomed Anal 1996;14(8–10):1133–1140.
154. Lam XM, Yang JY, Cleland JL. Antioxidants for prevention of methionine oxidation in recombinant monoclonal antibody HER2. J Pharm Sci 1997;86(11):1250–1255.

155. Quan C, Alcala E, Petkovska I, Matthews D, Canova-Davis E, Taticek R, et al. A study in glycation of a therapeutic recombinant humanized monoclonal antibody: where it is, how it got there, and how it affects charge-based behavior. Anal Biochem 2008;373(2):179–191.
156. Gadgil HS, Bondarenko PV, Pipes G, Rehder D, McAuley A, Perico N, et al. The LC/MS analysis of glycation of IgG molecules in sucrose containing formulations. J Pharm Sci 2007;96(10):2607–2621.
157. Zhang B, Yang Y, Yuk I, Pai R, McKay P, Eigenbrot C, et al. Unveiling a glycation hot spot in a recombinant humanized monoclonal antibody. Anal Chem 2008;80(7): 2379–2390.
158. Quan CP, Wu S, Dasovich N, Hsu C, Patapoff T, Canova-Davis E. Susceptibility of rhDNase I to glycation in the dry-powder state. Anal Chem 1999;71(20):4445–4454.
159. Watkins NG, Neglia-Fisher CI, Dyer DG, Thorpe SR, Baynes JW. Effect of phosphate on the kinetics and specificity of glycation of protein. J Biol Chem 1987;262(15): 7207–7212.
160. Watkins NG, Thorpe SR, Baynes JW. Glycation of amino groups in protein. Studies on the specificity of modification of RNase by glucose. J Biol Chem 1985;260(19): 10629–10636.
161. Andya JD, Maa YF, Costantino HR, Nguyen PA, Dasovich N, Sweeney TD, et al. The effect of formulation excipients on protein stability and aerosol performance of spray-dried powders of a recombinant humanized anti-IgE monoclonal antibody. Pharm Res 1999;16(3):350–358.
162. Krishna RG WF, Imahori K, Skiyama F, editors. Methods in Protein Sequence Analysis. New York: Plenum; 1993, p 167.
163. Sykes PA, Watson SJ, Temple JS, Bateman RC Jr. Evidence for tissue-specific forms of glutaminyl cyclase. FEBS Lett 1999;455(1–2):159–161.
164. Bateman A, Solomon S, Bennett HP. Post-translational modification of bovine pro-opiomelanocortin. Tyrosine sulfation and pyroglutamate formation, a mass spectrometric study. J Biol Chem 1990;265(36):22130–22136.
165. Busby WH Jr, Quackenbush GE, Humm J, Youngblood WW, Kizer JS. An enzyme(s) that converts glutaminyl-peptides into pyroglutamyl-peptides. Presence in pituitary, brain, adrenal medulla, and lymphocytes. J Biol Chem 1987;262(18):8532–8536.
166. Huang KF, Liu YL, Cheng WJ, Ko TP, Wang AH. Crystal structures of human glutaminyl cyclase, an enzyme responsible for protein N-terminal pyroglutamate formation. Proc Natl Acad Sci USA 2005;102(37):13117–13122.
167. Eswar N, Ramakrishnan C. Deterministic features of side-chain main-chain hydrogen bonds in globular protein structures. Protein Eng 2000;13(4):227–238.
168. Yu L, Remmele RL Jr, He B. Identification of N-terminal modification for recombinant monoclonal antibody light chain using partial reduction and quadrupole time-of-flight mass spectrometry. Rapid Commun Mass Spectrom 2006;20(24):3674–3680.
169. Mozdzanowski J, Bongers J, Anumula K. High-yield deblocking of amino termini of recombinant immunoglobulins with pyroglutamate aminopeptidase. Anal Biochem 1998;260(2):183–187.
170. Harris RJ, Chamow SM, Gregory TJ, Spellman MW. Characterization of a soluble form of human CD4. Peptide analyses confirm the expected amino acid sequence, identify glycosylation sites and demonstrate the presence of three disulfide bonds. Eur J Biochem 1990;188(2):291–300.

171. Dick LW Jr, Qiu D, Mahon D, Adamo M, Cheng KC. C-terminal lysine variants in fully human monoclonal antibodies: investigation of test methods and possible causes. Biotechnol Bioeng 2008;100(6):1132–1143.
172. Schenerman MA, Brooks RS, Kozlowski S, Webber K, Gazzano-Santoro H, Mire-Sluis A. CMC Strategy Forum Report: Analysis and structure characterization of monoclonal antibodies. Bioprocess Int 2004;2(2):8.
173. Antes B, Amon S, Rizzi A, Wiederkum S, Kainer M, Szolar O, et al. Analysis of lysine clipping of a humanized Lewis-Y specific IgG antibody and its relation to Fc-mediated effector function. J Chromatogr B Analyt Technol Biomed Life Sci 2007;852(1–2): 250–256.
174. Lazar AC, Kloczewiak MA, Mazsaroff I. Matrix-assisted laser desorption/ionization mass spectrometry for the evaluation of the C-terminal lysine distribution of a recombinant monoclonal antibody. Rapid Commun Mass Spectrom 2004;18(3):239–244.
175. Prigge ST, Kolhekar AS, Eipper BA, Mains RE, Amzel LM. Amidation of bioactive peptides: the structure of peptidylglycine α-hydroxylating monooxygenase. Science 1997;278(5341):1300–1305.
176. Prigge ST, Mains RE, Eipper BA, Amzel LM. New insights into copper monooxygenases and peptide amidation: structure, mechanism and function. Cell Mol Life Sci 2000;57(8–9):1236–1259.
177. Clarke S. Propensity for spontaneous succinimide formation from aspartyl and asparaginyl residues in cellular proteins. Int J Pept Protein Res 1987;30(6):808–821.
178. Bornstein P, Balian G. Cleavage at Asn-Gly bonds with hydroxylamine. Methods Enzymol 1977;47:132–145.
179. Robinson AB, McKerrow JH, Cary P. Controlled deamidation of peptides and proteins: an experimental hazard and a possible biological timer. Proc Natl Acad Sci USA 1970;66 (3):753–757.
180. Zhu JX, Aswad DW. Selective cleavage of isoaspartyl peptide bonds by hydroxylamine after methyltransferase priming. Anal Biochem 2007;364(1):1–7.
181. Rehder DS, Chelius D, McAuley A, Dillon TM, Xiao G, Crouse-Zeineddini J, et al. Isomerization of a single aspartyl residue of anti-epidermal growth factor receptor immunoglobulin gamma2 antibody highlights the role avidity plays in antibody activity. Biochemistry 2008;47(8):2518–2530.
182. Doyle HA, Mamula MJ. Post-translational protein modifications in antigen recognition and autoimmunity. Trends Immunol 2001;22(8):443–449.
183. Cacia J, Keck R, Presta LG, Frenz J. Isomerization of an aspartic acid residue in the complementarity-determining regions of a recombinant antibody to human IgE: identification and effect on binding affinity. Biochemistry 1996;35(6):1897–1903.
184. Robinson NE, Robinson AB. Deamidation of human proteins. Proc Natl Acad Sci USA 2001;98(22):12409–12413.
185. Tyler-Cross R, Schirch V. Effects of amino acid sequence, buffers, and ionic strength on the rate and mechanism of deamidation of asparagine residues in small peptides. J Biol Chem 1991;266(33):22549–22556.
186. Wright HT. Nonenzymatic deamidation of asparaginyl and glutaminyl residues in proteins. Crit Rev Biochem Mol Biol 1991;26(1):1–52.
187. Liu DT. Deamidation: a source of microheterogeneity in pharmaceutical proteins. Trends Biotechnol 1992;10(10):364–369.

188. Perkins M, Theiler R, Lunte S, Jeschke M. Determination of the origin of charge heterogeneity in a murine monoclonal antibody. Pharm Res 2000;17(9):1110–1117.
189. Huang L, Lu J, Wroblewski VJ, Beals JM, Riggin RM. *In vivo* deamidation characterization of monoclonal antibody by LC/MS/MS. Anal Chem 2005;77(5):1432–1439.
190. Wang L, Amphlett G, Lambert JM, Blattler W, Zhang W. Structural characterization of a recombinant monoclonal antibody by electrospray time-of-flight mass spectrometry. Pharm Res 2005;22(8):1338–1349.
191. Creese AJ, Cooper HJ. The effect of phosphorylation on the electron capture dissociation of peptide ions. J Am Soc Mass Spectrom 2008;19(9):1263–1274.
192. Syka JE, Coon JJ, Schroeder MJ, Shabanowitz J, Hunt DF. Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. Proc Natl Acad Sci USA 2004;101(26):9528–9533.
193. Udeshi ND, Compton PD, Shabanowitz J, Hunt DF, Rose KL. Methods for analyzing peptides and proteins on a chromatographic timescale by electron-transfer dissociation mass spectrometry. Nat Protoc 2008;3(11):1709–1717.
194. Udeshi ND, Shabanowitz J, Hunt DF, Rose KL. Analysis of proteins and peptides on a chromatographic timescale by electron-transfer dissociation MS. FEBS J 2007;274 (24):6269–6276.
195. Chi A, Bai DL, Geer LY, Shabanowitz J, Hunt DF. Analysis of intact proteins on a chromatographic time scale by electron transfer dissociation tandem mass spectrometry. Int J Mass Spectrom 2007;259(1–3):197–203.
196. Chi A, Huttenhower C, Geer LY, Coon JJ, Syka JE, Bai DL, et al. Analysis of phosphorylation sites on proteins from *Saccharomyces cerevisiae* by electron transfer dissociation (ETD) mass spectrometry. Proc Natl Acad Sci USA 2007;104(7): 2193–2198.
197. Mimura Y, Kabat EA, Tanaka T, Fujimoto M, Takeo K, Nakamura K. Microheterogeneity of mouse antidextran monoclonal antibodies. Electrophoresis 1995;16(1):116–123.
198. DiAugustine RP, Gibson BW, Aberth W, Kelly M, Ferrua CM, Tomooka Y, et al. Evidence for isoaspartyl (deamidated) forms of mouse epidermal growth factor. Anal Biochem 1987;165(2):420–429.
199. Bischoff R, Lepage P, Jaquinod M, Cauet G, Acker-Klein M, Clesse D, et al. Sequence-specific deamidation: isolation and biochemical characterization of succinimide intermediates of recombinant hirudin. Biochemistry 1993;32(2):725–734.
200. Kwong MY, Harris RJ. Identification of succinimide sites in proteins by N-terminal sequence analysis after alkaline hydroxylamine cleavage. Protein Sci 1994;3(1): 147–149.
201. Canova-Davis E, Eng M, Mukku V, Reifsnyder DH, Olson CV, Ling VT. Chemical heterogeneity as a result of hydroxylamine cleavage of a fusion protein of human insulin-like growth factor I. Biochem J 1992;285(Pt 1):207–213.
202. Johnson BA, Aswad DW. Optimal conditions for the use of protein L-isoaspartyl methyltransferase in assessing the isoaspartate content of peptides and proteins. Anal Biochem 1991;192(2):384–391.
203. Aswad DW, Paranandi MV, Schurter BT. Isoaspartate in peptides and proteins: formation, significance, and analysis. J Pharm Biomed Anal 2000;21(6):1129–1136.
204. Shacter E. Quantification and significance of protein oxidation in biological samples. Drug Metab Rev 2000;32(3–4):307–326.

205. Liu H, Gaza-Bulseco G, Xiang T, Chumsae C. Structural effect of deglycosylation and methionine oxidation on a recombinant monoclonal antibody. Mol Immunol 2008;45(3):701–708.

206. Chumsae C, Gaza-Bulseco G, Sun J, Liu H. Comparison of methionine oxidation in thermal stability and chemically stressed samples of a fully human monoclonal antibody. J Chromatogr B Analyt Technol Biomed Life Sci 2007;850(1–2):285–294.

207. Qi P, Volkin DB, Zhao H, Nedved ML, Hughes R, Bass R, et al. Characterization of the photodegradation of a human IgG1 monoclonal antibody formulated as a high-concentration liquid dosage form. J Pharm Sci 2009;98(9):3117–3130.

208. Yu J. Intentionally degrading protein pharmaceuticals to validate stability-indicating analytical methods. BioPharm Int 2000;(13):46–50.

209. Ji JA, Zhang B, Cheng W, Wang YJ. Methionine, tryptophan, and histidine oxidation in a model protein, PTH: mechanisms and stabilization. J Pharm Sci 2009;98(12): 4485–4500.

210. Baynes JW, Watkins NG, Fisher CI, Hull CJ, Patrick JS, Ahmed MU, D, et al. The Amadori product on protein: structure and reactions. Prog Clin Biol Res 1989;304:43–67.

211. Narhi LO, Schmit J, Bechtold-Peters K, Sharma D. Classification of protein aggregates. J Pharm Sci 2012;101(2):493–498.

212. Narhi LO, Jiang Y, Cao S, Benedek K, Shnek D. A critical review of analytical methods for subvisible and visible particles. Curr Pharm Biotechnol 2009;10(4):373–381.

213. Rosenberg AS. Effects of protein aggregates: an immunologic perspective. AAPS J 2006;8(3):E501–E507.

214. Carpenter JF, Randolph TW, Jiskoot W, Crommelin DJ, Middaugh CR, Winter G, et al. Overlooking subvisible particles in therapeutic protein products: gaps that may compromise product quality. J Pharm Sci 2009;98(4):1201–1205.

215. Singh SK, Afonina N, Awwad M, Bechtold-Peters K, Blue JT, Chou D, et al. An industry perspective on the monitoring of subvisible particles as a quality attribute for protein therapeutics. J Pharm Sci 2010;99(8):3302–3321.

216. Liu H, Gaza-Bulseco G, Faldu D, Chumsae C, Sun J. Heterogeneity of monoclonal antibodies. J Pharm Sci 2008;97(7):2426–2447.

217. Wang LW, Leonhard-Melief C, Haltiwanger RS, Apte SS. Post-translational modification of thrombospondin type-1 repeats in ADAMTS-like 1/punctin-1 by C-mannosylation of tryptophan. J Biol Chem 2009;284(44):30004–30015.

218. Wen D, Vecchi MM, Gu S, Su L, Dolnikova J, Huang YM, et al. Discovery and investigation of misincorporation of serine at asparagine positions in recombinant proteins expressed in Chinese hamster ovary cells. J Biol Chem 2009;284(47): 32686–32694.

219. Valliere-Douglass JF, Eakin CM, Wallace A, Ketchem RR, Wang W, Treuheit MJ, et al. Glutamine-linked and non-consensus asparagine-linked oligosaccharides present in human recombinant antibodies define novel protein glycosylation motifs. J Biol Chem 2010;285(21):16012–16022.

220. International Conference on Harmonisation. Guidance on Q8(R1) pharmaceutical development; addition of annex; availability. Notice. Federal Register 2009;74(109): 27325–27326.

# Protein and Peptide Formulation Development

SATOSHI OHTAKE and WEI WANG

## 11.1 INTRODUCTION

Commercial production of proteins was enabled following significant advancements in the development of recombinant DNA technology. However, proteins are inherently unstable and possess unique chemical and physical properties that present challenges in their purification, separation, formulation, storage, and delivery (Fig. 11.1).

The aggregation tendency of proteins is arguably the most commonly faced challenge in the development of stable protein biotherapeutics [1]. Protein aggregates have been reported to exhibit either reduced or no biological activity [2,3]; for example, direct correlation was established between the loss of recombinant factor VIII activity and the degree of aggregation [4]. Many factors affect the mechanism and rate of protein aggregation, which can be divided into two general categories: internal and external. Internal effects arise from the structure of the protein (hence the amino acid sequence) and are specific to the protein being evaluated. External factors are driven by the environment of the protein, such as temperature, solution conditions, and processing steps involved (e.g., shear stress, freeze–thaw). In situations that require drying of the protein to improve storage stability (e.g., lyophilization), solid-state condition and composition play important roles.

Because the folded state of protein is only marginally more stable than its unfolded state [5], any change to the environment can trigger degradation or inactivation. Besides aggregation, several other degradation pathways involve covalent modifications. One example is fragmentation, which involves peptide bond cleavage. Others include oxidation, deamidation, isomerization, and disulfide bond shuffling. Furthermore, proteins may undergo posttranslational modifications, including glycosylation. In addition, formulation excipients have the potential to interact with the side chain groups of a protein (e.g., Maillard

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*,
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

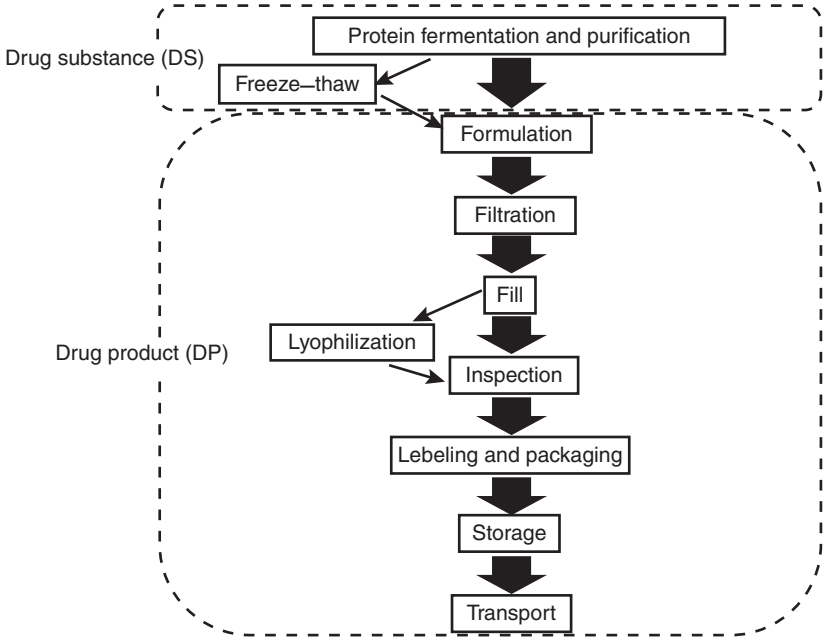


**FIGURE 11.1** Formulation fill-finish unit operations.

reaction) [6]. For many fill-finish operations, the physical stability of a protein can be severely impacted. Stability can be affected during freeze–thaw cycles, mixing, and filtration, which are all common process operations involved in the commercial production of protein therapeutics (see Fig. 11.1). Protein stability can be further impacted through its interaction with packaging and delivery device components.

The ideal formulation would consist of the pure, native protein by itself. For reasons mentioned earlier, it is not practical to only have the protein as the drug product (DP). This is because the protein must be purified from a complex mixture of cell debris and culture media followed by removal of any buffers or salts that are typically used for purification. Most importantly, the protein may not be stable on its own at the target concentration. For this purpose, excipients are typically included to achieve the desired stability profile. Selection of appropriate excipients and their concentrations requires a good understanding of the broad spectrum of degradation pathways affecting the protein of interest and appropriate expertise in a wide variety of analytical methods.

Regulatory guidelines are critical elements to reference during formulation development because they recommend information needed to demonstrate the safety and efficacy of the biotherapeutic. Furthermore, the results obtained allow the formulation scientist to write the appropriate pharmaceutical development section for regulatory filings. Development of commercial formulations requires a clear

understanding of the intended market. For example, information on indication, patients, method of delivery, typical dose requirement, and market distribution provides directions for the design of a successful formulation. Also, it is important to consider the competitiveness of the formulation in comparison to other products already on the market.

## 11.2 GENERAL FORMULATION DEVELOPMENT STRATEGIES

### 11.2.1 Preformulation Development

Preformulation studies are designed to probe the susceptibility of proteins to a variety of stresses. The main objectives of the preformulation study include: (1) general characterization of the product, (2) investigation of potential stability issues, (3) development of relevant analytical methods, (4) establishment of a stability profile, and (5) identification of major formulation challenges.

To predict potential stability problems within a short period of time, proteins are exposed to excessive stresses, and the resulting degradation products are analyzed. The findings from these accelerated stability studies may also be useful to predict the kinetics of degradation processes under realistic handling conditions, especially if there is a shortage of real-time results available due to time and resource constraints. The accelerated stability study is not acceptable to determine the expiry of the product. Rather, it is best used to rank order the severity of various degradation pathways observed. Not all of the degradation products are equivalent in terms of their contribution to the quality of the protein product. Thus, a qualitative assessment of each degradation product is important when weighing their significance. For example, if the degradation product compromises the safety or efficacy of the product, it should be considered more critical; certain chemical changes (e.g., deamidation) may not alter the activity of a given product, whereas other reactions (e.g., oxidation) may render the same product inactive.

The critical parameters affecting the pharmaceutical quality of a protein therapeutic are defined during the preformulation study. In many cases, formulation design is based on compromises that deal with several different problems (Fig. 11.2), as will be described in more detail later. Furthermore, the formulation will have to be optimized based on assumptions and extrapolations of results obtained during a limited time.

### 11.2.2 Formulation Development

A reliable formulation is required to support various preclinical and clinical studies. To obtain regulatory approval for the initiation of clinical trials, demonstration of safety, manufacturability, and stability of the drug is required. A sufficiently stable formulation is necessary at this stage because it is important to maintain the quality of the protein for the entire duration of preclinical studies,

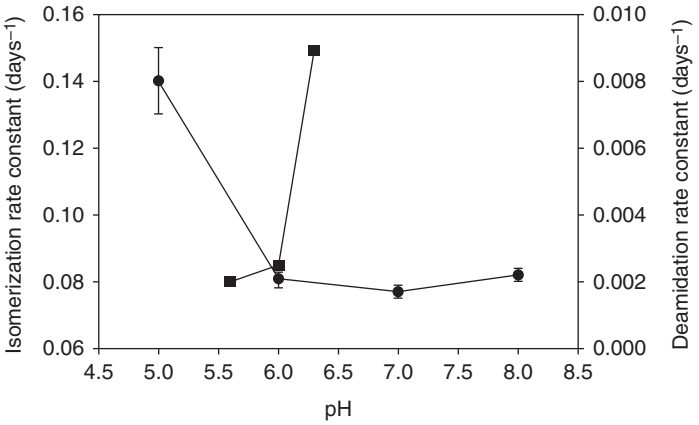


**FIGURE 11.2** Pseudo first-order rate constants for isomerization (●) and deamidation (■) versus pH for a monoclonal antibody. (Adapted from Ref. [299].)

from which results will be used to support the registration of the DP with the regulatory agencies.

Different types of formulations need to be developed based on clinical needs, patient compliance, delivery method, stability of the drug, storage and distribution, and market competitiveness. Having a clear plan about what type of formulation is desired will allow the scientist to design better study plans. Typically, liquid formulations have been preferred due to their convenience of manufacture and use. However, protein drugs in liquid dosage form may not be sufficiently stable for storage and handling. Dried formulations (e.g., lyophilized) and suspension formulations (e.g., insulin zinc suspension) have been successfully prepared to overcome stability problems. In addition, specific applications and intended mode of delivery may demand the use of an alternative dosage form (e.g., spray-dried powder for pulmonary delivery).

In addition to the recommended storage condition, the effects of temperature excursion should also be tested because the quality of the DP can be severely affected. In addition, exposure to multiple freeze–thaw cycles is typically conducted as a way to determine whether the formulation can withstand unexpected freezing during distribution and storage. Likewise, sensitivity to agitation, or surface denaturation, requires evaluation to support the selection of formulation, container closure systems, and mode of shipping and handling. For formulations containing low protein concentration, special attention is required to avoid impurity-related instability (e.g., oxidation fostered by metal contaminants in excipients).

The typical starting point for formulation development is pH and buffer screen. Most chemical reactions are affected by pH (e.g., deamidation and fragmentation), as well as by the choice of buffer salt. Generally, it is preferred to select excipients that have been included in marketed products with the same route of delivery as the therapeutic under evaluation. In addition, preference is given to excipients that have

been used at similar level and frequency of dosing, history of chronic use, and on similar patient populations. Otherwise, the approval and safety of the excipients need to be examined carefully. Depending on the intended use several different pharmaceutical grades are available (e.g., USP, Ph EUR, and JP). These pharmaceutical grade materials should be considered as primary resource, however the quality provided may be insufficient. Furthermore, regulatory guidelines discourage the use of animal- or human-derived excipients.

As mentioned previously, the preference is to develop DPs in the liquid format due to ease of manufacturing; however, if sufficient stability cannot be attained, lyophilization may be considered (see Fig. 11.1). Recent improvements in device design, including dual-chamber syringes and cartridges, as well as convenient reconstitution devices, have helped the pharmaceutical industry develop lyophilized products without significant concerns surrounding patient compliance. However, it is important to note that lyophilization can present its own challenges, particularly in the design of appropriate formulations and process cycles that are economical. More detail on lyophilization is provided in Section 11.3.2.5.

Commercially viable and market competitive formulations share several common features. Most notably, the formulation should maintain the safety and efficacy profile of the protein DP throughout the handling and use period specified on the label. Shelf life needs to be adequate, under conditions to which the product will be exposed, throughout the commercial distribution system. Finally, a number of systematic studies need to be carried out to comply with regulatory requirements for registration.

### 11.2.3 Typically Observed Degradation Mechanisms

Degradation pathways for proteins can be separated into two general categories, those involving physical instability and those involving chemical instability (Fig. 11.3). Physical instability does not require covalent modification of the protein;

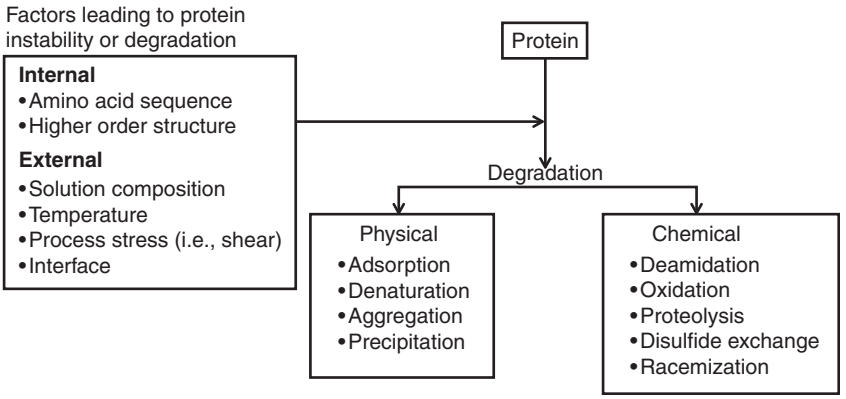


**FIGURE 11.3** Summary of commonly observed physical and chemical degradation mechanisms for protein pharmaceuticals.

rather, it refers to changes to its higher order structure (secondary and above). Chemical instability can be defined as any process that involves modification to the protein via bond formation or cleavage, yielding a new chemical entity. Both mechanisms are described in more detail in the following sections.

***11.2.3.1 Physical Instability*** Proteins, due to their polymeric nature and ability to adopt a higher order structure, can undergo a variety of structural changes independent of chemical modification. Typically, globular proteins have a folded (or compact) structure to minimize the exposure of hydrophobic groups to the solvent [7–9].

After the first successful expression of recombinant proteins, the formation of insoluble aggregates was observed in the host bacteria [10]. These insoluble aggregates, referred to as inclusion bodies (IBs), reduced the yield of soluble active protein and necessitated their dissociation in order to recover the bioactivity of the expressed protein [10,11]. Formation of IBs appears to be a general phenomenon and independent of the degree of overexpression. Unfortunately, a thorough understanding of the mechanism of IB formation remains elusive, which is compounded by the lack of their characterization [12,13].

In addition to IB formation during expression, several other forms of physical instability are often encountered for proteins. These include surface adsorption, denaturation, aggregation, and precipitation, some of which are interrelated. Adsorption is defined as the loss of protein caused by adherence to a surface. It is operative under all concentrations, with the loss being most significant at low concentrations. Adhesion of proteins to surfaces is a well-known phenomenon in the field of biomaterials because biocompatibility is essential for artificial limbs, organs, and contact lenses [14]. Precipitation is the macroscopic equivalent of aggregation. Protein precipitation has been known to occur for a long time, usually in conjunction with denaturation [15,16], but detailed studies have mostly been limited to the case of insulin. Recent work on insulin frosting has demonstrated that the process is accelerated by the presence of large headspace within the vial, which offers a large air–water interface at which insulin denaturation occurs [17]. Denaturation and aggregation are discussed in more detail in the next section, with emphasis on the mechanism, trigger, and formulation approaches used for stabilization.

*Denaturation* In solution, the native state of the protein is not a rigid structure. Rather, it can fluctuate to form several stable (and active) states that are difficult to detect [18]. Denaturation often refers to the alteration of the global fold of the protein, that is, a disruption of secondary and tertiary structures [16,19,20]. This process can be either reversible or irreversible and is caused by a variety of conditions, including changes in temperature, shifts in pH, and addition of organic solvents or other denaturants. The progression of protein denaturation is thought to occur in a sequential series of intermediate formation [21–23]. In fact, the denatured state is an ensemble average of several non-native protein conformations. These misfolded intermediates are often thermodynamically stable, and their rate of refolding to the native state can be quite

slow. One reason may be that the misfolded intermediate contains non-native disulfide bond(s) that require reduction for proper folding [24].

A commonly reported cause for protein denaturation is surface interaction. Adsorption of proteins to interfaces has been extensively reviewed [25,26] and thus will only be briefly discussed here. Proteins have been reported to adhere strongly to hydrophobic surfaces, but their adsorption to hydrophilic surfaces is typically weak [27]. One detrimental consequence for interfacial protein adsorption is decrease in its thermostability [28] This can be reduced, or even eliminated, by using a more hydrophilic container or through the addition of excipients. These excipients inhibit surface adsorption by competitive binding at the interface or through binding with the protein itself. A commonly used excipient for this purpose is the surfactant. When adding surfactants or other stabilizers to the formulation, the effects of the additive on the pharmacologic properties of the drug must be considered; the additives may cause leaching of the container closure components, and thus stability studies must address this possibility [29].

In addition to the container closure surface, proteins can also denature at other interfaces. Gas interface and shear forces, derived from pumping or filtration, can both lead to significant protein denaturation. Gas interface is nonpolar and can lead to denaturation via similar mechanism as described earlier for hydrophobic surfaces. The rate of denaturation at a gas–liquid interface is highly dependent on the residence time of the protein at such interface [30]. For example, the rate of protein denaturation is increased upon vigorous shaking for a prolonged period of time, resulting in the formation of soluble as well as insoluble protein aggregates [31]. Typically, the stability of a therapeutic protein in the final configuration is assessed by subjecting it to shaking for several hours, simulating the worst case level of stress caused by transportation and handling.

*Aggregation* Aggregation is the most prevalent means by which the physical stability of protein structure is compromised. Many factors lead to protein aggregation. The external environments (e.g., pH, osmolality, ionic strength) of the protein affect the structure, as will be described in more detail later. The native structure of the protein results from a delicate balance between multiple competing forces and interactions. Thus, a change in the external environment of the protein can easily disrupt such balance, leading to change in structure, which may subquently lead to aggregation. In addition to the factors listed above, temperature, exposure to light, and process-related factors (e.g., shear stress, freezing, thawing, and drying) can all contribute to destabilization. Typically, aggregation is preceded by a conformational change, which results in non-native configuration, and can proceed through multiple pathways [32,33]. It should be noted that aggregates can be soluble and their formation reversible.

In contrast to protein association, which is a reversible process that involves two or more proteins in the native state, protein aggregation typically requires the interaction of two or more denatured protein molecules. Thus, protein denaturation must occur before aggregate formation. Under typical conditions, native proteins in solution are in equilibrium with their unfolded intermediates [34], which in turn are in equilibrium with the denatured state. Several reports suggest these intermediates

to be the precursors of the aggregation process [35–37] because they expose a greater fraction of hydrophobic patches to the solvent and possess higher structural flexibility compared to the folded state [38]. In fact, simulation studies support the assertion that aggregation is accelerated by the formation of partially unfolded conformations with accessible hydrophobic areas on the protein surface [39]. Many other proteins have been reported to aggregate through similar mechanisms, including bovine serum albumin (BSA) [40], recombinant human growth hormone [41], and granulocyte-colony stimulating factor [42]. Theoretically, the completely unfolded state of a protein can form aggregates directly. The initial protein aggregates are soluble oligomers but gradually become insoluble as they exceed certain size and solubility limits. This can be present in various forms (e.g., amorphous precipitates or fibrils) and depends on the protein structure and solvent conditions. It should be noted that oligomerization is not a requirement in the process of fibril formation; for Aβ42, additives and conditions have been identified that can inhibit oligomerization without affecting fibrillization [43]. Nucleation-dependent polymerization is another mechanism that has been proposed to describe the aggregation tendency of several proteins [44,45].

Some proteins associate directly to aggregates from the native state without proceeding through an intermediate state. These associations can be driven by electrostatic, hydrophobic, or van der Waals interactions, depending on the solution condition [46]. Such association often leads to the formation of reversible aggregates, which can be considered the precursors of irreversible aggregates, or precipitates [47,48]. This process has been described as association-limited aggregation [44]. A key parameter to quantify the tendency of a protein to self-associate is the second virial coefficient, B22. Whereas a positive value indicates net repulsion, a negative value indicates net attraction. In the latter case, protein–protein interactions are favored over protein–solvent interactions, potentially leading to protein aggregation [49]. It should be noted, however, that the B22 values typically do not indicate the relative aggregation tendency of the protein [50,51]. In general, minimization of protein surface charge will likely lead to increased aggregation or particulate formation [52].

There is a narrow temperature range in which the proteins are relatively stable. Excursion from this range will destabilize the protein and potentially promote aggregation [53,54]. Typically, thermally induced protein aggregation is irreversible [55,56], although the degree of reversibility depends on the stage of aggregation process and the size of aggregates being formed [57]. In addition to the factors already mentioned, higher temperature promotes aggregation by increasing diffusion, frequency of molecular collisions, and hydrophobic interactions [38,58]. All proteins possess a unique melting or denaturation temperature ($T_m$), at which they unfold. $T_m$ is a good indication of the relative thermal stability of the protein; the higher the $T_m$, the less likely the protein will unfold. Thermally induced unfolding is often followed by aggregation due to the exposure of hydrophobic residues. There are exceptions, however. Hen egg white lysozyme with a $T_m$ of 55°C at pH 2 has been reported to resist aggregation at its $T_m$ [59]. The solution condition (e.g., salt concentration) may have an overriding effect in controlling protein aggregation.

Instability at a low temperature may facilitate the generation of unfolded intermediates, although aggregation may be retarded because of reduced hydrophobic interactions. As such, low temperature-induced protein aggregation is often reversible in contrast to aggregation induced by elevated temperature.

The solution pH dictates the type and distribution of charges on the protein surface [60,61]. These charges affect both intra- and intermolecular interactions. Due to the delicate nature of the protein structure, resulting from the balance of various interaction forces, proteins are typically stable only within a narrow pH range [62,63]. Aggregation caused by pH change can simply be an effect of altered charge–charge interactions or pH-induced protein destabilization [51,64]. At extreme pHs, proteins are heavily charged, which can lead to increased repulsive interactions. As a consequence, pH-induced protein unfolding typically does not lead to immediate aggregation [65,66]. Furthermore, changing the solution pH could potentially modify the aggregation mechanism [67]; formation of amyloid fibrils has been reported at high net charge, but particulate formation was noted under conditions of low net charge [52]. As described further below, prevention of protein aggregation needs to take into account all possible pathways. At the isoelectric point (pI) of the protein, solubility is low because of minimal protein charge–charge repulsions. Currently, no clear relationship between aggregation and pI has been established because there are examples demonstrating both positive [68,69] and negative [64,70] correlations. pH may also impact the interaction of the protein with formulation components, potentially leading to protein aggregation. In addition, pH has been reported to play a significant role in affecting the morphology of aggregates [52,59]. To maintain the pH of the solution to provide structural stability, a buffering agent is often included in the formulation. The choice of buffering agent and its concentration can have a significant impact on the aggregation behavior of proteins [62,71,72]. For example, phosphate buffers have been reported to increase aggregation in several studies [73,74].

Ionic strength (and the type of salt used to vary ionic strength) is another formulation parameter that plays a key role in affecting the aggregation tendency of a protein [75]. Increasing ionic strength often neutralizes the surface charge of proteins, thus reducing both intra- and intermolecular charge–charge interactions through enhanced electrostatic screening [76]. Its impact on aggregation depends on whether such interactions favor protein folding or association. If so, then aggregation tends to increase [77,78]. If not, aggregation will be inhibited [70,79]. Generalization is difficult because the effect of ionic strength on aggregation tendency is protein specific and can vary depending on solution conditions. Although NaCl is commonly used for adjusting ionic strength, other inorganic salts can be used and may exert a significant effect on protein aggregation, possibly through binding. Formulation approaches to inhibit or control the extent of aggregation are described in more detail later.

Protein aggregation can be induced during various processing steps (see Fig. 11.1), including fermentation and expression [80,81], unfolding and refolding [82,84], purification [85,86], freeze-thaw [87,88], shaking and shearing [89,90], pressurization [91,92], drying [93,94], preparation of protein drug delivery systems,

administration, and so on. Many impurities and contaminants that are carried with proteins during various production processes (from protein expression to manufacture of final DP) can potentially affect protein stability and aggregation. Such potential impurities include host cell proteins, enzymes, cell membrane components, residual metal ions, etc. Organic solvents can be introduced during the production of proteins and have been reported to affect their stability both positively and negatively, depending on the protein and the relative quantity or quality of the solvent [95]. The most commonly used organic solvent in protein production is glycerol, which can stabilize proteins and inhibit aggregation under certain experimental conditions [96]. Rubber stoppers are often siliconized for ease of processing during production. Trace amount of silicone has been reported to cause protein aggregation due to its hydrophobic nature; the aggregation propensity of four model proteins (concanavalin A, BSA, ribonuclease A, and lysozyme) possessing varying molecular weights and isoelectric points was found to increase significantly in the presence of 0.5% silicone oil in solution [97].

Aggregates can potentially impact the activity, half-life, and immunogenicity of the protein therapeutic [2,98,99]. As such, studies must be performed to determine the extent of aggregate formation that is acceptable for patient administration. Several reports on the relationship between aggregation and immune response have been published [100–102]. However, aggregation can at times be used advantageously, for example, to increase process yield during protein expression or purification. In these cases, aggregation is induced intentionally and is then reversed during subsequent processing. Other notable examples for desirable aggregation include crystallization and precipitation (e.g., isoelectric, salt induced). For example, the zinc hexamer of insulin is a complex of insulin and zinc, which results in the formation of insulin crystals [103]. This form of insulin has been developed for its long-acting properties because the hexamer dissociates slowly into dimers and eventually to monomers [104].

Although aggregation has been categorized under physical instability, it can also be classified under chemical degradation if it is induced through direct cross-linking. The most common cross-linking reaction is the intermolecular disulfide bond formation and exchange [62,105,106]; for example, the heat-induced aggregation of β-lactoglobulin B (β-Lg) and κ-casein [107]. Proteins not possessing any free cysteines can also undergo aggregation through a variety of mechanisms, including disulfide exchanges via β-elimination [108], oxidation [109], and the Maillard reaction [110]. It is also worth noting that multiple aggregation pathways often occur. For example, insulin can undergo aggregation via both physical (e.g., formation of soluble hexamers or insoluble fibril) and chemical (e.g., formation of soluble dimers via cyclic anhydride intermediate) processes [111–113].

***11.2.3.2 Chemical Instability*** In addition to physical instability, a variety of chemical reactions is known to occur in proteins. These reactions can involve hydrolysis, deamidation, oxidation, β-elimination, and racemization, among others. Some of these degradation pathways are described in more detail in the following sections.

*Deamidation* Deamidation is typically observed in proteins following the loss of an amine group from asparagine (Asn) and glutamine (Gln) residues, resulting in the formation of aspartic acid and glutamic acid, respectively, through a succinimide intermediate. This degradation pathway is of particular interest because this is believed to serve as a biological regulator of protein clearance mechanism *in vivo* [114]. Deamidation has been reported to have no detectable impact on the activity of some proteins (e.g., human growth hormone and interleukin-1α) [115,116] while reducing the activity of others (e.g., lysozyme and cytochrome C) [117,118]. Protein conformational changes are expected to occur upon deamidation due to the modification of surface charge distribution [7]. For example, altered forms of proteins have been reported for lysozyme [119], human growth hormone (hGH) [120], and epidermal growth factor (EGF) [121].

Deamidation of Asn and Gln residues is an acid- and base-catalyzed reaction based on elimination and hydrolysis. Whereas the acid-catalyzed reaction involves protonation of the amide leaving group, the base-catalyzed reaction involves nucleophilic attack of the amide carbonyl carbon by hydroxide or conjugate base. Deamidation of Asn residues, which occurs most often at the Asn–glycine (Gly) sequence, is reported to be accelerated under neutral to alkaline conditions [122,123]. The proposed mechanism involves the formation of a cyclic imide intermediate through nucleophilc attack of the carbonyl carbon of the Asn side-chain amide by the main-chain peptide nitrogen from the adjacent C-terminal residue [124,125]. This is reported to be the rate-determining step [126]. The cyclic imide is then base catalyzed at either one of the nitrogen–carbonyl bonds, producing a mixture of peptides in which the polypeptide backbone is connected via α-carboxyl linkage, resulting in aspartate (Asp) formation, or via β-carboxyl linkage, resulting in iso-Asp [123,127]. Thermodynamically, isopeptide formation is favored over the α-linked Asp peptide at equilibrium, typically at a 3:1 ratio [127]. Because deamidation is an acid- and base-catalyzed reaction, there is a pH at which the rate of deamidation is minimal, as has been reported for cytochrome C [128].

As the formation of the cyclic intermediate involves the participation of the neighboring amino acid, the size and physicochemical properties of the neighboring amino acid side chain are expected to play an important role in its rate of formation. Robinson and Rudd [129] reported the presence of large bulky hydrophobic groups on the N-terminal side of the amide residue to result in lower rates of deamidation; a hexapeptide with leucine (Leu) as the neighboring amino acid to Asn has been reported to be approximately 50 times slower in the rate of cyclic imide formation than that with Gly as the neighboring amino acid [126]. On the other hand, polar residues preceding Asn and Gln have been reported to result in enhanced deamidation rates. On a larger scale, the secondary and tertiary structures of the protein play a significant role in affecting the rate of Asn deamidation (e.g., an Asn that ordinarily does not deamidate in the native structure of ribonuclease was reported to deamidate in the denatured protein) [130].

Consequences of deamidation are many, including its effect on metabolic half-life, charge distribution, and conformation [131,132]. Deamidation can be a cooperative process, whereby the initial deamidation increases the rate of deamidation at

additional sites. This cooperative process may involve a structural change that moves Asn closer to residues that catalyze deamidation, promoting further flexibility and eventually leading to unfolding. If irreversible unfolding occurs, proteins are more susceptible to irreversible aggregation. On the other hand, an increase in net negative charge on the protein may lead to decrease in the rate of aggregation, caused by enhanced repulsive interactions. Very little is known about the effects of deamidation on immunogenicity because contrasting conclusions have been reported [116,133,134]. Formulation strategies to reduce or inhibit deamidation are described later.

*Oxidation* Protein oxidation (or catalyzed auto-oxidation) is one of the major chemical degradation pathways, both in solution and lyophilized formulations. This degradation pathway is complicated because any of the steps involved (initiation, propagation) can be rate limiting, and the rate-limiting step can change during the course of the reaction. This is further compounded by the variability in the location of the reactive amino acids within the protein structure; some are buried within the hydrophobic regions of the molecule and are fairly inert, while others are exposed to the solvent and are thus more reactive [135]. For example, the methionine (Met) residues in trypsin [136] and myoglobin [137] are unreactive under mild oxidation conditions, but Met in hGH [135] and transferrin [138] are partially exposed to the solvent and react at varying oxidation rates.

Several reactive amino acids (e.g., Met, cysteine [Cys], cystine, histidine, tryptophan, and tyrosine) can undergo oxidation through various mechanisms. For example, Met has been reported to be oxidized to methionine sulfoxide by atmospheric oxygen, hydrogen peroxide, and periodate, to name a few [139]. Under extremely oxidative conditions, methionine sulfone can be formed. This conversion has been associated with loss in biological activity for many peptide hormones, including parathyroid hormone [140] and corticotropin [141]. In many cases, reduction of Met sulfoxide by thiols has been reported to result in nearly full recovery of biological activity [142,143], and there are several examples for which Met oxidation does not lead to loss in protein function [144,145].

Cysteine is also easily oxidized. For a thiol group of Cys, oxidation can occur in a stepwise manner to sulfenic acid, disulfide, sulfinic acid, and sulfonic acid, depending on the reaction conditions [146]. The rate of Cys oxidation is typically accelerated at higher pH, in which the thiol is deprotonated [147], although exceptions to this rule have been reported [148,149]. In fact, pH dependence on protein oxidation may be used to distinguish between Met and Cys oxidation; whereas hydrogen peroxide–catalyzed oxidation of Cys decreases as the pH is lowered from pH 5 to 1, the rate of Met oxidation has been reported to increase slightly [150]. Cys oxidation is also strongly catalyzed by transition metal ions, especially copper.

Disulfide formation is not necessarily restricted to intramolecular reactions. If the thiol group within the protein is hindered (or only a single thiol group is present), intermolecular disulfide bonds can be formed, resulting in aggregation [151]. For example, oxidation of the well-protected Met [4] residue in the variable domain of an

IgG light chain has been reported to induce noticeable structural changes and sensitize the protein to stirring-induced aggregation [152].

Oxidizing agents that trigger the reaction include peroxides, light, free radicals, dissolved oxygen, and metal ions [153,154]. There are several reactive oxygen species, including singlet oxygen, superoxide radical, alkyl or hydrogen peroxide, hydroxyl radicals, and halide oxygen complexes. The oxidative potential is dependent on the oxidizing agent and the location of the labile residue in the protein structure.

Pharmaceutical consequences of oxidation include loss in bioactivity [33], reduced half-lives [155], and changes in antigenicity caused by conformational shifts [156]. For example, both unoxidized and oxidized Insulin B chains were reported to stimulate insulin-specific T-cell hybridomas, however xenogenic insulins did not, suggesting the importance of conformation on antigenicity [157].

*Proteolysis* Amino acids that make up the protein backbone are linked via peptide bonds. Under certain conditions, these bonds can be easily hydrolyzed to yield constituent amino acids or peptides, resulting in the disruption of protein structure. The primary impact of hydrolysis is loss in potency or activity of the protein.

It has been reported that peptide bonds of Asp residues are cleaved in dilute acid at a rate at least 100 times faster than the other peptide bonds [158]. Hydrolysis can take place at either the N- or C-terminal peptide bonds, which are adjacent to the Asp residue. The mechanism proposed by Inglis [159] involves cleavage of the N-terminal peptide bond through a six-membered ring intermediate. In a separate study, Marcus [160] has compared the lability of the Asp–Proline (Pro) bonds with that of other peptide bonds, particularly those of Asp residues. The results indicate the Asp-Pro bonds to be eight- to 20-fold more labile than other Asp-X or X-Asp peptide bonds. Peptide bonds that do not involve Asp were found to be stable to hydrolysis under similar conditions.

*Other Chemical Degradation Pathways* In addition to deamidation, oxidation, and proteolysis, several other chemical degradation pathways are possible for proteins, depending on the solvent condition and amino acid sequence. Incorrect disulfide formation results from interchange of disulfide bonds, which can lead to an altered three-dimensional structure and thus loss of catalytic activity. The reaction mechanism is different in alkaline and acidic media [161,162]; in neutral and alkaline media, the reaction is catalyzed by thiols, which in the form of thiolate ions carry out nucleophilic attack on the sulfur atom of the disulfide. Disulfide exchange can be prevented if thiol scavengers (e.g., *N*-ethylmaleimide, copper ion) are present [163].

Proteins can be sensitive to light exposure, leading to many types of light-induced degradation [164]. Recent investigations indicate ultraviolet irradiation to induce cross-linking reactions between Cys and Trp residues and photolysis of disulfide bonds, leading to new disulfide bond formation and exchange [165].

Racemization is typically considered to proceed through the removal of the α-methine hydrogen by base to form a carbanion. Racemization of Asp is

particularly interesting because the free amino acid racemizes $10^5$-fold slower than that in a protein structure, but all other amino acid residues under similar condition undergoes racemization at a two- to fourfold slower rate [166].

High temperature inactivation of proteins is often a consequence of the destruction of disulfide bonds [117,167]. Chemically, this is the result of β-elimination from the cystine residue. Besides cystine, several other amino acids (serine, threonine, phenylalanine, and lysine) have been reported to degrade under alkaline conditions [168,169]. In many cases, β-elimination is influenced by pH, temperature, and the presence of metal ions.

## 11.3 KEY CHALLENGES IN DRUG PRODUCT FORMULATION AND FILL-FINISH

### 11.3.1 Variables Affecting Protein and Peptide Stability

Formulation and fill-finish processes introduce various stresses on proteins that can cause significant degradation, as described earlier. Major stresses include temperature, pH, and shear, among others. All proteins undergo thermal unfolding when exposed to heat. The degree of thermal stability (typically analyzed and expressed as $T_m$) varies among proteins, but many proteins can denature at temperatures below their $T_m$. In fact, this is the reason why protein pharmaceuticals are sterilized by membrane filtration rather than autoclaving. Furthermore, degradative reactions described in the previous section are accelerated as the temperature is increased. Temperature can also indirectly affect the stability of the protein by modifying the pH of buffering agents as well as the solubility of oxygen in solution. pH affects both the solubility and stability of proteins, as described previously, and thus the buffering capacity of the protein-containing solution must be controlled appropriately to support all manufacturing processes involved in DP production.

Proteins may undergo aggregation or precipitation when exposed to mechanical stresses, including agitation or a high-shear environment such as pumping, filtration, mixing, and fill-finish processes [31,111,170]. Proteins also have a tendency to bind to surfaces [171]. A significant decrease in recovery has been reported in cases for which a diluted protein solution is exposed to a large surface area. Freezing, another commonly used processing step, introduces complex physical and chemical changes during which proteins can denature. As water crystallizes out of solution, the remaining solutes are concentrated along with the protein. Some solutes may crystallize, or phase separate, due to their low solubility limit, resulting in pH shift [172,173]. The surface denaturation of proteins at the ice–water interface has also been reported [174]. During thawing, the concentration gradient may not dissipate immediately, leading to protein instability. In addition, removal of water may result in alteration of protein conformation because the native structure of a protein is strongly influenced by the aqueous environment.

Extractables, leachables, contaminants, and impurities in the excipients may significantly affect protein stability. Exposure of proteins to metal ions can lead to

undesirable structural changes or precipitation [175]. Trace amount of pro-oxidants, (e.g., peroxide and hydroxyl radical) can oxidize proteins directly [33,176]. They can be introduced to the product as impurities or generated in the product by other free radical–mediated reactions of impurities. Light exposure can also result in protein degradation, which is dependent on light intensity and wave length [177].

Degradation routes at higher temperatures are typically not the same as those at lower temperatures. It has become common practice to conduct accelerated stability testing for proteins and to use the accelerated rate profiles to compare formulations and minimize degradation. This approach may provide misleading results because degradation occurring at higher temperatures may require vastly different formulation conditions for maximum stability compared to that occurring at lower storage temperatures. For example, when the stability of interleukin-1β was studied by different analytical techniques at or below 30°C, the degradation pathway was deamidation, but at or above 39°C, it was aggregation and precipitation [178]. The fact that this protein degraded by two different mechanisms precluded the prediction of shelf life using simple Arrhenius behavior. Disruption of the protein conformation can alter the spatial arrangement of charges, which in turn can affect the way in which the protein interacts with solvent molecules [179].

### 11.3.2 Compatibility with Manufacturing Process

The operations that occur after protein purification, such as freezing and thawing of the purified protein bulk, formulation (excipient addition), sterile filtration, filling, freeze drying, and inspection, are commonly referred to as formulation and fill-finish operations (see Fig. 11.1). Each step is described in detail in the following sections.

***11.3.2.1 Bulk Freeze–Thawing*** Purified proteins, also called bulk drug or drug substance (DS), are often frozen before formulation operations to facilitate storage, transportation, and handling while maintaining product quality. As such, freezing and thawing operations often constitute the first step in the formulation operation. The freezing process can be detrimental to proteins in several ways. For example, all water-soluble species in the formulation concentrate as the water phase separates (ice formation), which can lead to an enhanced rate of degradation [180,181]. Furthermore, protein-stabilizing excipients can crystallize during freezing, leaving the protein unprotected; for example, buffer components can crystallize, leading to pH shift of the unfrozen solution [172,173,182]. Proteins can denature through adsorption onto the ice surface [183] or by an decrease in temperature (cold denaturation) [184]. During the thawing process, proteins can be degraded through excessive agitation (shear and foaming) or precipitation caused by the concentration gradient.

Parameters that are critical for large-scale freeze–thaw process include: (1) temperature and rate of freeze-thaw, (2) use of agitation during thawing, (3) container selection, and (4) scale of operation. Freezing temperature dictates the thermodynamic state of the protein and its interaction with formulation excipients; if the product is frozen to a temperature below the eutectic temperature ($T_e$) of the crystalline additive, it is no longer available for proper interaction. Typically, the

target freezing temperature range is determined based on thermal analysis data obtained via differential scanning calorimetry (DSC), which provides information on various phase transitions for the components present in the formulation. Temperature–time profiles can be obtained experimentally by freezing the bulk in the selected container or freezer and positioning temperature recorders at various locations within the solution. Proteins have an optimum freezing rate; very high rates can be detrimental to the stability of proteins due to the creation of an excessively large ice surface area. In contrast, slow rates expose proteins to freezing-induced conditions for a prolonged period of time. For example, increased protein aggregation and decreased activity have been reported for liquid nitrogen–based freezing systems [87,183]. Fast freezing can also trap air that could be released during thawing and may cause protein denaturation at the air–liquid interface [185,186]. Controlled rate technologies, such as Celsius-Paks or CryoVessels, can achieve faster freeze–thaw rates by using a combination of small path length and increased flux for heat transfer [187]. Freezing time is another critical parameter controlled to ensure completion of the freezing process. Placebo experiments at the maximum intended bulk load can provide information about the longest freezing times and slowest freezing rates. After obtaining freeze–thaw profiles at full scale, the impact of the rate on protein integrity is tested by subjecting the protein to the worst-case freeze–thaw profile. It should be mentioned that concentrated protein solutions are often more resistant to freezing-induced protein aggregation [187,188]. The increased stability at high protein concentrations during freezing is likely attributable to one or more of the following factors: (1) limited area of ice–water interface for denaturation, (2) self-stabilization of proteins [47,183], and (3) conversion of monomers to active and more stable dimers or multimers [189,190].

Thawing temperature dictates the time required for the process and the rate of liquefaction of the freeze-concentrated protein. To expedite the thawing process, it is typically advisable to select the thawing temperature reasonably high, based on available stability data. Thawing time is determined by performing a full-scale experiment with the protein or placebo (if protein supply is limited). Completion is confirmed by visually ensuring that there are no residual ice crystals in solution. Similar to freezing rates, the faster the thawing rates, the better the product stability (although up to an optimum level). Slow thawing rates can result in ice recrystallization, with small ice crystals growing into larger ones; proteins may denature at the ice–liquid interface and lose their activity [191,192]. Cryoconcentration created during freezing can further harm the protein during thawing. The thawing method may or may not involve some type of agitation. Agitation helps maintain a uniform temperature, if the thawing temperature is different from the storage temperature, and minimize recrystallization and cryoconcentration-related effects. Very high thawing rates may result in foaming and thus to shear-induced and surface-mediated protein denaturation.

All of the factors described are linked to the scale of operation, material of construction, and configuration of the container. The container size is often dictated by the volume of the production batch and the frequency of production. Product stability and the batch volume typically play a role in the selection of an appropriate

container configuration. If material availability is a concern, preliminary experiments can be performed with a commercially available scale-down tool (e.g., cryo-wedges). Cryo-wedges are designed to subject the protein at small scales (~0.5 l) to the same freeze–thaw profile experienced at a larger scale. Some other considerations for container selection include its ease of use and transportation, sterilization method, chemical resistance, tamper evidence, and compendia requirements. The selected container is validated prior to experimentation, and container closure integrity should be ensured for the proposed freeze–thaw process. For a review of large-scale protein freeze–thaw operation, readers are referred to publications by Wisniewski [193,194] and Wisniewski and Wu [195].

***11.3.2.2 Formulation*** The purified form of the protein is usually stored in buffer, which typically does not contain all of the ingredients required in the final formulation. Rather, if the DS is frozen, cryoprotectants are present, which may or may not be required in the final DP formulation. Furthermore, the protein concentration at the end of purification may be different from that required for DP. This is because it is operationally more favorable to store the DS at a higher concentration than that intended for the final formulation. Thus, after thawing of the frozen DS, its concentration is adjusted, and the desired excipients are added (and extraneous excipients are removed). In other cases, buffer exchange may be required between the DS and DP, typically using an ultrafiltration (UF) or diafiltration (DF) step. Although this sounds simple, the labile nature of proteins and the scale of operation involved offer many possibilities for protein degradation. Exposure of proteins to air–liquid interfaces during processing (e.g., pumping, mixing) and their interactions with large surfaces of tanks and tubing material over extended periods of time can lead to unfolding or adsorption to surfaces. On the other hand, insufficient mixing during the addition of excipients could alter product quality (solution inhomogeneity), resulting in the final DP being out of specifications. In certain applications, a concentration step may be required if the DP is to be formulated at a higher concentration than the DS. High product concentration and viscosity could pose further challenges for membrane filtration during the concentration step. Thus, careful consideration must be given to each unit operation to ensure that protein integrity and activity are maintained.

The formulation process should be carefully designed by taking into consideration all possible concentrations of the purified and formulated bulk, as well as process consistency and ease of operation at multiple manufacturing sites. The constituent buffer should be carefully selected so that the operation is simple yet robust, with appropriate tolerances in place. Buffer tolerances are established such that the process is practically feasible while yielding a product that is within established specifications for purity, potency, safety, and efficacy.

Various excipients are included in the formulation to protect the protein against multiple degradation pathways. Common attributes of desirable excipients are aqueous solubility, non-toxicity, non-reactivity, rapid clearance from the body, and lack of immunogenicity. In addition, the excipients should be capable of stabilizing the native conformation of the protein during processing, storage, and

administration to the patient. Some of the simplest and most effective are salts [196–198] and polyalcohol compounds, such as glycerol and sugars [199–201]. The purity of excipients is another key factor affecting product quality [202]. Certain impurities in the raw materials can trigger degradation reactions; for example, peroxides in polysorbate can lead to protein oxidation, and trace amounts of metals can lead to aggregation [203,204]. Use of animal-derived excipients may carry the risk of causing transmissible spongiform encephalopathies, thus requiring careful evaluation. The choice of raw material suppliers is critical. The order of addition of excipients can also impact product quality; in some cases, interim interactions between excipients may occur at concentrations higher than their equilibrium concentration in the post-mixed, formulated bulk. For example, addition of polysorbate is often performed following the UF/DF step to minimize its loss caused by interaction with the membrane component. In addition, surfactants could be adsorbed onto surfaces (i.e., tubing and tanks), thus resulting in lowered surfactant concentration in the formulated bulk. Leachables and extractables from various contact surfaces also have the potential to affect product integrity.

Small-scale characterization studies should be conducted to evaluate protein stability under appropriate formulation conditions. Buffers should be characterized to establish tolerances around excipient concentrations, pH, acid–base volumes required, conductivity, and osmolality. Hold time studies using the worst-case scenario for surface area exposed per unit volume should be designed to examine the impact of different contact materials on protein stability. Temperature excursions should also be evaluated through hold time studies. Characterization of the mixing process should include both product homogeneity testing and impact of mixing shear on product quality. Also, compatibility of the protein with product-contacting surfaces (e.g., pumps and tubing) should be examined. These findings can then be verified with fewer runs at commercial scale to determine their impact on DP quality.

Formulation approaches to counteract several degradation mechanisms described in the previous section will now be reviewed. Focus will be on denaturation, aggregation, deamidation, and oxidation.

*Formulation Approaches for Countering Denaturation and Aggregation* A protein DP may contain a specified acceptable level of aggregates not only after initial product release but also during the shelf life of the product. The aggregation rate of the protein needs to be estimated and controlled during product development so that the limit will not be exceeded during manufacture and storage. Because the solvent environment dictates the ability of the protein to maintain its native state and resist denaturation from various stresses, several excipients and their combinations are empirically tested. One popular mechanism that describes the efficacy of additives, or co-solvents, to inhibit protein unfolding is preferential hydration. Excipients are referred to as co-solvents because they affect the solvation of the protein. Co-solvents include salts, sugars, buffers, and amino acids. A co-solvent is excluded from the protein surface as a result of either inert steric exclusion effects (e.g., increase in the surface tension of water) or chemical incompatibility with the protein surface (e.g., charge repulsion), thus leaving

the protein to be hydrated with water [196,205–208]. In contrast, co-solvent binding describes mechanism of protein destabilization through binding of co-solvents, which can be either non-specific or specific [205]. Several studies have demonstrated that many denaturants cause global changes to protein conformation rather than bind to specific sites [209,210]. The classification of co-solvents into these two categories is often difficult because several additional factors influence their behavior [211]. For example, although surfactants are generally considered co-solvents, they display unique features that are not observed for other co-solvents [212]. Other mechanisms are also operable, such as an increased rate of protein folding [213,214], a reduction of solvent accessibility and conformational mobility [215], and an increase in solvent viscosity [216].

Based on the information gathered during preformulation studies, as described earlier, the aggregation tendency of the protein candidate can be quickly evaluated at various pH, ionic strengths, and storage temperatures. The assessment of formulation efficacy can be conducted by the use of appropriate analytical techniques, including size exclusion high-performance liquid chromatography (SE-HPLC), isoelectric capillary eletrophoresis (iCE), light scattering, and so on. If a high dose is projected for clinical use by subcutaneous (SC) injection, information on the solubility and viscosity of the protein candidate is required. Four major parameters are typically examined during formulation development; pH, buffering agent, salt, and stabilizers (if required). A wide pH range needs to be examined because the relationship between protein aggregation and pH is often complex and difficult to predict. Several types of relationships have been reported, including well-shaped [178,217], bell-shaped [76,218], sigmoidal [219], and linear [220] relationships.

Next step in formulation development is the identification of an optimum ionic strength to minimize protein aggregation, which can be highly protein dependent [5]; it should be noted that the salt type can have a significant impact on protein aggregation [75]. Although many buffering agents are available for pH adjustment, the aggregation behavior of proteins can vary drastically in different buffer systems and concentrations [62,71,221]. Many stabilizers, including sugars, polyols, amino acids, amines, salts, polymers, and surfactants, can reduce protein aggregation [1,222,223].

Sugars and polyols can help maintain the correct conformation of proteins, serving as co-solvents [224]; however, they do not appear to work equally well on all stages of the aggregation process [225]. Surfactants can be generally divided into two classes, non-ionic and ionic. Both classes have been demonstrated to affect protein aggregation induced by different stresses. The major rationale for using surfactants in protein formulations is their ability to bind to hydrophobic interfaces that are potential denaturation sites [226,227]. Non-ionic surfactants are a well-known class of excipients, often used to prevent or inhibit protein aggregation [212] and adsorption to surfaces [228,229], as well as to hinder denaturation [230,231]. A commonly used type of non-ionic surfactant for this purpose is polysorbate (Tween) [212]. The reported effects of non-ionic surfactants on protein aggregation have not been consistent among different proteins because their effects are multifaceted [49,178,232,233]. Ionic surfactants, or surfactant-like compounds, have been

reported to affect protein aggregation more intensely than non-ionic surfactants. This is due to their strength of protein binding, facilitated by their interaction with both polar and nonpolar groups [82,234,235]. In several cases, however, ionic surfactants have been reported to induce protein aggregation [236]. The presence of other co-solvents may alter the physical characteristics of surfactants. For example, addition of co-solvents that cause preferential hydration of the protein (e.g., sugars) may also cause hydration of the surfactant, resulting in phase separation. Some protein DPs need to be manufactured with preservatives to accommodate multiple administrations. Yet these preservatives may not be compatible with the protein of interest. Preservatives often contain a hydrophobic component, which can bind to proteins, as discussed earlier for surfactants.

Several types of small charged additives have been reported to affect the aggregation tendency of proteins [237]. This may be influenced by the modification of solution ionic strength or through specific interactions with proteins [74,238]. Some positively charged amino acids (histidine [239], lysine [240], and arginine [241]) are particularly effective in inhibiting protein aggregation. The mechanism for the protective effect of arginine has been studied in several laboratories. This was thought to be attributable to the ability of arginine to solubilize proteins through the weakening of protein–protein interactions rather than a protein stabilization effect [242]. Additional evidence suggests that arginine deters aggregation by slowing protein–protein association reactions through preferential exclusion [243]. Metal ions play a critical role in maintaining the correct conformations for several proteins [244,245]. In many cases, metal ions can facilitate aggregation of proteins. The pro-aggregation effects are generally mediated through binding to proteins [90,246]. Binding can result in small changes in the tertiary structure of proteins, leading to the formation of aggregation-competent conformers. The effect of metal ions is often concentration dependent [247]. Metal ions also influence the kinetics of protein aggregation and the properties of aggregates. Polymers have been extensively used with proteins. They may inhibit protein aggregation through: (1) surface activity, (2) preferential exclusion, (3) weak binding, (4) steric hindrance of protein–protein interactions, and (5) increased viscosity, limiting protein structural movement [66,248]. It seems that the overall effect is a balance between protein binding and macromolecular crowding effect. Among these interaction forces, hydrophobic interactions appear to be more effective in controlling protein aggregation than charge effects [249].

Many studies have been conducted to evaluate the effect of protein concentration on aggregation. Because of the existence of several aggregation pathways (e.g., partial unfolding induced, simple native association-initiated aggregation), increasing protein concentration could potentially result in a number of effects, including increased aggregation caused by a greater chance of association [46,48], decreased aggregation caused by a crowding effect [98,250,251], and precipitation [81,252].

If proteins demonstrate a high tendency to aggregate in the liquid state and no effective formulation approaches can be found, preparation of a dried dosage form may be a viable alternative. The commonly used method for preparing a solid product is lyophilization, as will be discussed later.

*Formulation Approaches for Countering Deamidation* The main strategy to reduce deamidation in proteins involves optimization of solution conditions (excipients, pH, ionic strength, and temperature). Knowledge of the deamidation mechanism can help in the search for an optimum pH. Furthermore, the anion present in the buffer solution have been reported to play a significant role in impacting deamidation rates [253]. These effects are described in more detail below. Moreover, because the degradation pathways can be temperature dependent, data from accelerated stability studies may not always be extrapolated to real-time studies.

The mechanism of deamidation often dictates the formulation pH. If deamidation proceeds by a general acid–base-catalyzed mechanism, then it may be possible to formulate at an optimum pH, at which both rates are minimized. For example, deamidation that proceeds through the cyclic imide intermediate is base catalyzed, and thus lowering the pH should greatly reduce the rate of deamidation.

It is widely known that deamidation is markedly dependent on the type of buffer anion used [254]. Gilbert et al. [253] investigated the deamidation of Gln to Glu in the presence of different sodium buffer salts at pH 8 and 37°C. The rates of deamidation were high in phosphate, arsenate, and bicarbonate. In contrast, there was no observed rate enhancement by sulfate, nitrate, acetate, chloride, or pyruvate anions. The deamidation rate decreased upon the addition of organic solvents, with the reduction in rate constant being directly proportional to the concentration (and choice) of organic solvent. For example, solvents such as ethanol reduced the rate of deamidation while polyethylene glycols (PEG 4000 and 8000) demonstrated only a slight retardation in deamidation rate for several peptides as a function of PEG concentration [253]. For the hexapeptide Val-Thr-Pro-Asn-Gly-Ala, the main reported effect of co-solvent addition has been the decrease in the dielectric constant of solution [255]. More specifically, the decreased deamidation rate is attributable to the destabilization of the deprotonated peptide bond nitrogen anion in solvents with low dielectric constants. The effect of ionic strength on the rate of deamidation was studied using the glutaminyl pentapeptides (Gly-Leu-Gln-Ala-Gly and Gly-Arg-Gln-Ala-Gly) [256] and the asparaginyl pentapeptide (Gly-Arg-Asn-Arg-Gly) [257]. These studies demonstrated deamidation to proceed at a faster rate with higher ionic strength, which supports the previous work with Gln [253].

*Formulation Approaches for Countering Oxidation* Protein oxidation in solution is highly dependent on the temperature, dissolved oxygen, light, and excipients used. The addition of polysorbate 80 to lyophilized interleukin-2 formulations has been reported to stabilize the product nominally but to increase the amount of protein oxidation [258]. Oxidation was caused by the high peroxide content in polysorbate 80 remaining after purification and bleaching. Peroxide is known to increase upon storage, particularly at higher temperatures [259,260]. There are three classes of antioxidants. The typical antioxidants are phenolic compounds, including butylated hydroxytoluene, butylated hydroxyanisole, propyl gallate, and vitamin E. Reducing agents form the second class of antioxidants, which includes ascorbic acid, sodium sulfite, thioglycerol, and thioglycolic acid. These act as oxygen scavengers in a closed system, wherein they oxidize preferentially over the protein. The third class of antioxidants includes

chelating agents such as EDTA (ethylenediaminetetraacetic acid), citric acid, and thioglycolic acid. These antioxidants act by removing trace metal ions that play a major role in initiating auto-oxidation reactions.

Typically, the first set of strategies for minimizing protein oxidation is to refrigerate, protect from light, remove oxygen, and adjust the pH. Because trace contaminants present in various excipients can also impact oxidation, excipient screening and an effective selection process based on drug stability data are critical. Replacement of headspace oxygen in both solution and dry-powder formulations may also be required. When oxidation is light induced, protection of protein solutions from light may be required.

***11.3.2.3 Filtration*** Following formulation, the active product is sterile filtered (using a 0.22-μm pore size membrane) to ensure that the product is free of bioburden. Polyvinylidene fluoride (PVDF) membranes are commonly used for this purpose. Product compatibility with the filter membrane material is tested during validation studies prior to the use of filters at a commercial manufacturing site. Filter size recommendation is critical and can be product specific. Too small a filter could prolong the filtration time, but too large a filter could result in large hold-up volume, which in turn leads to greater product loss. Other critical parameters that require careful consideration include: (1) allowable pressure drop across the filter, (2) duration over which the filtered product can be held to maintain product sterility, (3) temperature of the operation, and (4) the mode of liquid transfer during filtration. All filters are integrity tested before and after use, per established procedures.

An additional in-line filtration step might be incorporated before filling. Dual filtration may also be used for risk mitigation in the scenario of filter failure. Thus, the protein solution may be exposed to multiple filtration steps prior to filling as a final dosage form in the selected container. It is therefore important to evaluate the impact of filtration process on product quality. Sterile filtration at a high trans-membrane pressure could stress the protein. The protein can also bind selectively to the membrane, resulting in either misfolding or protein loss. Similarly, other formulation components (e.g., surfactants that are added as stabilizers) may get adsorbed onto the membrane surface. This causes the surfactant concentration in the solution to decrease below the target value. Thus, the order of addition is a key consideration.

***11.3.2.4 Fill*** After the DS has been formulated and sterile filtered, it is filled into primary DP containers, which are typically vials or syringes for prefilled injectables. All of the container closure surfaces are often treated with different coatings to minimize protein binding and facilitate their processing (e.g., insertion of stoppers into vials). During this step, the DP not only comes into contact with the primary container but also with the various components such as stoppers, plungers, and so on. All of these components are subjected to sterilization processes separately and brought together under aseptic processing conditions [261]. The container closure system and the environment of the fill chamber are qualified to be of the highest standards needed to ensure product quality and sterility.

Prior to fill, the vials are washed, depyrogenated, and then cooled to room temperature. Stoppers are generally washed, siliconized, and sterilized, although pre-washed and pre-siliconized stoppers are commercially available. Some of the critical attributes include: (1) the type of silicone used, (2) the extent of siliconization, (3) the stopper composition, (4) the residual moisture of the stoppers, and (5) the moisture permissibility of stoppers. Rubber stoppers are lubricated with silicone to ensure appropriate closure fitting into the vial neck, and adequate siliconization ensures proper machinability of the stoppers on the fill equipment. Silicone is considered relatively inert, and its concentration in DP containers is significantly low to ensure no risk with respect to biological incompatibility. However, interaction between proteins and silicone oil has been reported to induce protein aggregation or product inactivation [97]. Knowledge of the residual moisture content of stoppers and the moisture-vapor transmissibility are critical for lyophilized formulations, for which moisture transfer from and through the stopper to the product is a concern [262]. Also, the possibility of any deleterious interactions between the product and any leachables from the vials or stoppers requires investigation.

Besides silicone oil, other constituents of the primary container have the potential to contribute to leachables and extractables at a concentration that can affect protein stability and activity [263,264]. Leachables are contaminants that have migrated from packaging components (e.g., plastic bottles, glass vials, prefilled syringes, rubber stoppers) or manufacturing equipment (e.g., tanks, tubing) into the product during DP processing through prolonged exposure. Extractables refer to species that are released from the material of construction of the contacting surface caused by prolonged interaction with the product under specific solvent and temperature exposure conditions. In addition to toxicologic concerns, these chemicals can also impact protein stability.

Filling of liquid protein pharmaceuticals is performed in a clean room of appropriate classification (class 100). Most manufacturing facilities use automatic fill machines with multiple needles. Studies are conducted with the product to determine the fill-accuracy capabilities of the equipment and the amount of unrecoverable product (i.e., holdup volume) from vials and disposable syringes. Based on these values, a target fill weight and tolerances around that value are established. Automatic fill machines typically have the capability to periodically test fill weights online. The filling device's materials of construction (for product contact) must be compatible with the protein. In addition, the shape and height of the filling nozzle, as well as the fill speed, could contribute to foaming in vials that could possibly result in exposing the hydrophobic residues of the protein to air and initiate protein aggregation at the air–water interface. Although it is necessary to keep the filling speed as high as practically achievable to maintain high throughput, very high filling speeds can lead to foaming, and subsequently to protein denaturation [265]. The maximum time the product can be held at room temperature during fill is based on product stability data, which are recorded for ensuring product stability. For some liquid-dose products, the vial headspace is overlayed with nitrogen gas to reduce potential for product oxidation by headspace oxygen

[266]. Light exposure during fill and post-fill processes may also result in protein degradation for certain products.

***11.3.2.5 Lyophilization*** There are two major product forms for commercial protein pharmaceuticals: liquid and freeze-dried solid. Due to their ease of processing, the low cost of production, and the convenience of administration, liquid protein products are typically preferred unless aggregation or other forms of instability is difficult to control adequately. If the product shelf life is insufficient in liquid dosage form, freeze-dried formulation is a viable alternative. Freeze-drying, or lyophilization, is the process by which the water content of the product is reduced by freezing followed by removal through sublimation under vacuum. Although a freeze-dried protein formulation generally offers enhanced protection against protein modification, the freeze-drying process itself could cause significant protein degradation.

Freeze-drying improves product stability by: (1) maintaining the protein in an amorphous phase with its stabilizers, (2) immobilizing the protein in a glassy phase below the glass transition temperature ($T'_g$) of the formulation, and (3) reducing the residual water content to a low, desirable value. The lyophilization process primarily consists of three stages. The first stage involves freezing the product and creating a frozen matrix suitable for drying. This step impacts the drying characteristics in the subsequent two stages. The second stage is primary drying, which involves the removal of ice by sublimation through reduction in chamber pressure while maintaining the product temperature at a low, desirable level. The third stage in the process is secondary drying, during which the water content is further reduced until the target level is reached. Some of the crystalline additives, such as mannitol and glycine, may not crystallize completely during freezing. These excipients have low $T'_g$ values and, if not crystallized fully, they contribute to the overall $T'_g$ of the formulation (i.e., $T'_g$ depression). Low $T'_g$ necessitates drying at lower temperatures, which prolongs the cycle time. This can be avoided through the use of pre-sublimation annealing, which requires holding the system for some time at a temperature between the ice melt and $T'_g$ of the freeze concentrate [267]. There are two main purposes behind this step: (1) to crystallize bulking agents such as glycine and mannitol and (2) to allow larger ice crystals to grow (Ostwald ripening). There are two physical phenomena that should be avoided during a lyophilization process: collapse and meltback.

Collapse is described as the loss of microscopic and macroscopic structure of the freeze-dried product. Collapse is typically observed if the product temperature exceeds a characteristic temperature called the collapse temperature ($T_c$). When this occurs, the matrix collapses under its own weight. In addition to losing pharmaceutical elegance, collapsed product may lead to increased residual water content, uneven distribution of moisture, longer reconstitution times, and possible conformational changes of the protein. $T_c$ is unique to a given formulation and correlates well with the $T'_g$ of the formulation; the $T_c$ is usually slightly higher than the $T'_g$. Primary-to-secondary drying transition is critical because a premature transition to a higher shelf temperature can lead to product collapse; as the water

is removed, $T'_g$ of the amorphous phase increases, but caution must be taken to implement a slow rate of increase in temperature, so as to always stay below the glass transition temperature of the amorphous phase [188].

Most formulations contain crystalline components such as mannitol or glycine as bulking agents. Eutectic melting of these crystalline components leads to meltback. To avoid meltback, the product needs to be held below the eutectic melting temperature ($T_e$) of the crystalline components contained in the formulation. $T_e$ is the crystalline analogue of $T'_g$. Because the $T_e$'s of commonly used bulking agents are typically higher than the $T'_g$ of the amorphous phase, meltback can be prevented by keeping the product below the $T'_g$ of the formulation.

For freeze-dried products, the formulation and the lyophilization process are intricately interrelated. As mentioned earlier, to maintain product stability, the product temperature must be maintained below $T'_g$ both during drying and storage. Therefore, a formulation with a higher $T'_g$ allows drying at a higher temperature compared to that with a lower $T'_g$, which subsequently expedites the freeze-drying time. Typically, the $T'_g$ of a given formulation is determined experimentally before a lyophilization cycle is developed. This can be done using several techniques, including DSC. $T_c$ can be determined using a freeze-drying microscope by freezing the material and raising the temperature to observe the onset of collapse [268].

Sugars, particularly the disaccharides, are the most often cited stabilizers used in lyophilized protein formulations. The two hypotheses put forth for describing the efficacy of disaccharides in stabilizing proteins are the water replacement and glassy state models. The water replacement hypothesis postulates that disaccharides stabilize dried proteins through "hydration" at sites previously occupied by water [269]. The glassy state hypothesis suggests that protein stabilization is the result of increased glass transition temperature [270]. It should be noted that the most commonly used disaccharides, sucrose and trehalose, have been reported to demonstrate varying degrees of protection, depending on the protein and the protein-to-disaccharide ratio [271,272]. The excipients that can act as water substitutes often allow for a reduction in residual water content. In addition, many of these excipients (e.g., mannitol) are required to produce a dry cake of high quality and are often included to produce a uniform powder that can easily be reconstituted. In many formulations, salts are added to control the osmolality of the solution. However, lyophilization in the presence of salts such as sodium chloride has caused several proteins to precipitate [273,274]. The effect of sodium chloride on the freezing process (freezing point depression) may result in aggregation [72] or increased binding to the protein [275]. To circumvent these issues, mannitol or other sugars are often used as isotonicifiers. In addition, non-ionic surfactants, such as polysorbates, are commonly included in formulations to prevent stresses related to freeze–thaw cycles, presumably by inhibiting denaturation at the ice–water interface [170].

The residual water content of the freeze-dried cake has been reported to have a significant impact on the structure and stability of the dried protein. Several studies have suggested the existence of an optimum residual water content required to maintain the stability of various protein formulations [276]. After determining the

optimum level, the maintenance of this water content becomes critical to the stability of the dried formulation.

Dried protein products generally require a reconstitution step before administration. Such a step could lead to different levels of protein degradation, depending on the reconstitution procedure, reconstitution medium, and presence of other excipients. It is generally recommended that reconstitution be conducted carefully, not to create any air bubbles. In this regard, high vacuum in the product container should be avoided to allow adequate control of the reconstitution medium stream in contact with the inside wall of the product container. Inclusion of a protein stabilizer in the reconstitution medium has been reported to minimize protein degradation [277].

In addition to the factors already described, problems can be encountered during process scale up. Factors that introduce heterogeneity between the vials need to be optimized. Furthermore, differences in heat and mass transfer between laboratory-scale, pilot-scale, and production-scale lyophilizers need to be taken into account.

Besides freeze-drying, several other processes are available to remove water from the protein formulation, including spray drying [278,279], spray freeze-drying [280,281], vacuum drying [282,283], and foam drying [284,286], to name a few. These are not commonly used due to various operational and scale-up challenges.

***11.3.2.6 Storage, Transport, and Delivery*** Filled vials are crimped and inspected. Appropriate inspection criteria need to be in place for both liquid and solid dosage forms. Liquid dosage forms should be inspected for particulate matter. Inspection methods can vary from manual (i.e., visual) to the use of automated machines. An allowable reject ion rate for inspection should be established. Following inspection, the vials are labeled and packaged. These steps, combined with inspection (both manual and automated), typically result in the exposure of DP to light for a prolonged period of time. Photoexposure could affect product quality, so it should be evaluated a priori for the worst-case conditions (maximum intensity and duration).

Unexpected environmental changes can be encountered during distribution and handling of products, including accidental freezing, exposure to temperatures outside of the recommended range, and vigorous agitation. It is critical to develop the formulation with these stresses in mind because product quality can be compromised.

Container closure systems could be a major factor in influencing the activity and immunogenicity of the protein DP [266]. After DP fill-finish operations are complete, they need to be transferred either to various storage sites or to the clinic. Transportation-related stresses include those that are related to vibration, shock, and changes in pressure and temperature, which might be experienced during air or ground transportation. When protein solutions are moved or vibrated during shipping or mixing, proteins are subjected to certain degrees of shaking stress. Exposure to light could induce protein oxidation, as discussed earlier. The International Conference on Harmonization (ICH) Q1B guideline recommends performing photostability testing in a controlled light chamber. It is likely that biologics, given their nature, may fail the recommended levels specified in ICH Q1B. However, it is possible that

formulation components, such as antioxidants, may slow down the rate and extent of photodegradation. A time-dependence study would need to be designed to test the resilience of the biologic to light-induced degradation. From this study, recommendation could be made in regards to processing times and light exposure conditions during manufacture, transport, and storage.

## 11.4 SUMMARY

Protein instability encompasses many complicated, as well as interrelated, chemical and physical processes. Destabilization can occur during any of the fill-finish operations described in this chapter (see Fig. 11.1). Even with the advancements in the development of protein pharmaceuticals, several unmet needs still remain. Of them, two stand out: high-concentration formulations and high-throughput analytics.

For products that require frequent and chronic administration, the SC route of administration is preferred. In particular, when coupled with the use of prefilled syringe and autoinjector device technology, SC delivery allows for self-administration and improved patient compliance, which may result in expansion of product market. Treatment with high doses may require development of formulations at concentrations exceeding 100 mg/ml because of the small volume limited by the SC route of administration. Development of formulations at high concentrations also poses stability, manufacturing, and delivery challenges; for example, increasing protein concentration may lead to higher viscosity, which may be difficult to administer by injection [48]. The other alternative is the delivery of higher volumetric dose.

Aggregation is expected to be the primary degradation pathway in high-concentration protein formulations. However, the relationship among concentration, rate, and extent of aggregate formation depends highly on the mechanism of association [287]. In contrast, many of the chemical reactions, such as deamidation, Asp isomerization, and oxidation, generally demonstrate kinetics that follow lower order concentration dependency. Macromolecular crowding (i.e., excluded volume effects) can also affect the physical properties of proteins, such as viscosity, which can impact the capability to manufacture high-concentration formulations.

Although it may be possible to create a formulation that demonstrates the desired stability profile during development, the design of processes to allow large-scale manufacture is critical and cannot be postponed. Two commonly used processes at large scale are tangential flow filtration (TFF) and lyophilization. TFF is the main technology for buffer exchange and concentration of proteins that have been used at commercial scale [288,289]. The main problem that has been identified during processing is the shear sensitivity of the protein, which may lead to the formation of aggregates and particulates during filling of the final product. Another parameter that could limit the use of a TFF system to concentrate a protein solution is its high viscosity. Although TFF is the industry standard for concentrating macromolecules at large scale, there are alternatives. High-concentration protein formulation can be achieved either by decreasing the solvent content while retaining the protein (e.g., using TFF) or by removal of the solvent (i.e., drying) followed by reconstitution with

reduced solvent volume. Bulk-drying processes include spray drying [278], fluid bed drying [290], vacuum drying [283], and lyophilization in tray units [291]. For parenteral protein products, bulk-drying processes require operation under aseptic conditions. As the protein-loading concentration is increased, reconstitution occurs much more slowly. Investigation of the morphology of the lyophilized solid by scanning electron microscopy has revealed a very dense compact structure at 110 mg/ml compared with a loosely packed layer of flakes at 40 mg/ml [292].

Although high-concentration formulations have the cost saving advantage of decreasing bulk storage space or number of product fills, they suffer from undesirable overall cost of goods caused by unrecoverable volumes from various product-contacting surfaces and containers. The loss could be minimized by decreasing product contact surface areas, such as with the use of smaller or narrower based vials or a prefilled syringe, in which the solution can be pushed off the surface with the plunger head.

Generally, analytical techniques used to characterize proteins involve either dilution to lower concentrations or exposure of protein to solvent conditions that are drastically different from the initial formulation composition. This may have considerable impact on the results of the assay because change in solvent composition or concentration may alter the physical state of the protein, resulting in a non-representative sample. Technologies such as static and dynamic light scattering [293,294] and analytical ultracentrifugation [295,296] provide more representative information about the self-association states of proteins at high concentrations due to the maintenance of the sample concentration during analysis.

With the number of monoclonal antibody candidates increasing (with compressed timelines for their development), there is a strong need for high-throughput analytical methods. A common high-throughput formulation development (HTFD) format is the use of commercially available multiwell microplates [297]. These can be placed in automated liquid-processing systems for sample preparation, dispensing, and handling. The low sample volume required for analysis can potentially save a large amount of protein DS. The HTFD microplate format is especially useful for the rapid screening of protein aggregation through turbidity measurement because this can be conveniently and non-invasively analyzed using a spectrophotometer [298]. It should be noted, however, that the small sample volume in microplates may not be sufficient to allow for a detailed analysis of other degradation pathways by traditional analytical methods, and protein surface adsorption and evaporation may contribute significantly to the variations in the results. There is certainly more development to come in this field.

## REFERENCES

1. Wang W. Protein aggregation and its inhibition in biopharmaceutics. Int J Pharm 2005;289(1–2):1–30.
2. Runkel L, Meier W, Pepinsky RB, Karpusas M, Whitty A, Kimball K, et al. Structural and functional differences between glycosylated and non-glycosylated forms of human interferon-beta (IFN-beta). Pharm Res 1998;15(4):641–649.

3. Nesta DP, Elliott JS, Warr JP. Supercritical fluid precipitation of recombinant human immunoglobulin from aqueous solutions. Biotechnol Bioeng 2000;67(4):457–464.
4. Wang W, Kelner DN 2003. Correlation of rFVIII inactivation with aggregation in solution. Pharm Res 20(4):693–700.
5. Wang W. Instability, stabilization, and formulation of liquid protein pharmaceuticals. Int J Pharm 1999;185(2):129–188.
6. Gadgil HS, Bondarenko PV, Pipes G, Rehder D, McAuley A, Perico N, et al. The LC/MS analysis of glycation of IgG molecules in sucrose containing formulations. J Pharm Sci 2007;96(10):2607–2621.
7. Kauzmann W. Some factors in the interpretation of protein denaturation. Adv Protein Chem 1959;14:1–63.
8. Privalov PL. Stability of proteins: small globular proteins. Adv Protein Chem 1979;33:167–241.
9. Dill KA. Theory for the folding and stability of globular proteins. Biochemistry 1985;24 (6):1501–1509.
10. Williams DC, Van Frank RM, Muth WL, Burnett JP. Cytoplasmic inclusion bodies in *Escherichia coli* producing biosynthetic human insulin proteins. Science 1982;215 (4533):687–689.
11. Marston FA. The purification of eukaryotic polypeptides synthesized in *Escherichia coli*. Biochem J 1986;240(1):1–12.
12. Gribskov M, Burgess RR. Overexpression and purification of the sigma subunit of *Escherichia coli* RNA polymerase. Gene 1983;26(2–3):109–118.
13. Marston FAO, Lowe PA, Doel M, Schoemaker JM, White S, Angal S. Purification of calf prochymosin (prorennin) synthesized in *Escherichia coli*. Biotechnology 1984;2:800–804.
14. Horbett TA. Protein adsorption on biomaterials. In: Cooper SL, Peppas NA, Hoffman AS, Ratner BD, editors. Biomaterials: Interfacial Phenomena and Application. American Chemical Society; 1982. p. 233–244.
15. Mirsky AE, Pauling L. On the structure of native, denatured, and coagulated proteins. Proc Natl Acad Sci USA 1936;22(7):439–447.
16. Neurath H, Greenstein JP, Putnam FW, Erickson JA. The chemistry of protein denaturation. Chem Rev 1944;34(2):157–265.
17. Massey EH, Sheliga TA. Human insulin (HI) isophane suspension (NPH) with improved physical stability. Pharm Res 1988;5:S34.
18. Lakowicz JR, Weber G. Quenching of fluorescence by oxygen. A probe for structural fluctuations in macromolecules. Biochemistry 1973;12(21):4161–4170.
19. Havel HA, Kauffman EW, Plaisted SM, Brems DN. Reversible self-association of bovine growth hormone during equilibrium unfolding. Biochemistry 1986;25(21):6533–6538.
20. Tanford C. Protein denaturation. Adv Protein Chem 1968;23:121–282.
21. Kiefhaber T, Rudolph R, Kohler HH, Buchner J. Protein aggregation in vitro and in vivo: a quantitative model of the kinetic competition between folding and aggregation. Biotechnology 1991;9(9):825–829.
22. Brems DN, Plaisted SM, Kauffman EW, Havel HA. Characterization of an associated equilibrium folding intermediate of bovine growth hormone. Biochemistry 1986;25(21): 6539–6543.

23. DeFelippis MR, Alter LA, Pekar AH, Havel HA, Brems DN. Evidence for a self-associating equilibrium intermediate during folding of human growth hormone. Biochemistry 1993;32(6):1555–1562.
24. Creighton TE. Disulphide bonds and protein stability. BioEssays 1988;8(2):57–63.
25. Sadana A. Interfacial protein adsorption and inactivation. Bioseparation 1992;3(5): 297–320.
26. Norde W, Lyklema J. Why proteins prefer interfaces. J Biomater Sci Polym Ed 1991; 2(3):183–202.
27. Andrade JD, Hlady V, Wei A-P, Ho C-H, Lea AS, Jeon S.I. et al. Proteins at interfaces: principles, multivariate aspects, protein resistant surfaces, and direct imaging and manipulation of adsorbed proteins. Clin Mater 1992;11(1–4):67–84.
28. Steadman BL, Thompson KC, Middaugh CR, Matsuno K, Vrona S, Lawson EQ, et al. The effects of surface adsorption on the thermal stability of proteins. Biotechnol Bioeng 1992;40(1):8–15.
29. Chawla AS, Hinberg I, Blais P, Johnson D. Aggregation of insulin, containing surfactants, in contact with different materials. Diabetes 1985;34(5):420–424.
30. Donaldson TT, Boonstra EF, Hammond JM. Kinetics of protein denaturation at gas–liquid interfaces. J Colloid Interface Sci 1980;74(2):441–450.
31. Henson AF, Mitchell JR, Musselwhite PR. The surface coagulation of proteins during shaking. J Colloid Interface Sci 1970;32(1):162–165.
32. Wetzel R. Mutations and off-pathway aggregation of proteins. Trends Biotechnol 1994;12(5):193–198.
33. Cleland JL, Powell MF, Shire SJ. The development of stable protein formulations: a close look at protein aggregation, deamidation, and oxidation. Crit Rev Ther Drug Carrier Syst 1993;10(4):307–377.
34. Clark AC. Protein folding: are we there yet? Arch Biochem Biophys 2008;469(1):1–3.
35. Kendrick BS, Carpenter JF, Cleland JL, Randolph TW. A transient expansion of the native state precedes aggregation of recombinant human interferon-gamma. Proc Natl Acad Sci USA 1998;95(24):14142–14146.
36. Carpenter JF, Randolph TW, Jiskoot W, Crommelin DJ, Middaugh CR, Winter G, et al. Overlooking subvisible particles in therapeutic protein products: gaps that may compromise product quality. J Pharm Sci 2009;98(4):1201–1205.
37. Murphy KP, Bhakuni V, Xie D, Freire E. Molecular basis of co-operativity in protein folding. III. Structural identification of cooperative folding units and folding intermediates. J Mol Biol 1992;227(1):293–306.
38. Damodaran S, Song KB. Kinetics of adsorption of proteins at interfaces: role of protein conformation in diffusional adsorption. Biochim Biophysic Acta 1988;954(3):253–264.
39. Zhang L, Lu D, Liu Z. How native proteins aggregate in solution: a dynamic Monte Carlo simulation. Biophys Chem 2008;133(1–3):71–80.
40. Militello V, Casarino C, Emanuele A, Giostra A, Pullara F, Leone M. Aggregation kinetics of bovine serum albumin studied by FTIR spectroscopy and light scattering. Biophys Chem 2004;107(2):175–187.
41. Bam NB, Cleland JL, Randolph TW. Molten globule intermediate of recombinant human growth hormone: stabilization with surfactants. Biotechnol Prog 1996;12 (6):801–809.

42. Raso SW, Abel J, Barnes JM, Maloney KM, Pipes G, Treuheit MJ, et al. Aggregation of granulocyte-colony stimulating factor in vitro involves a conformationally altered monomeric state. Protein Sci 2005;14(9):2246–2257.
43. Necula M, Kayed R, Milton S, Glabe CG. Small molecule inhibitors of aggregation indicate that amyloid beta oligomerization and fibrillization pathways are independent and distinct. J Biol Chem 2007;282(14):10311–10324.
44. Weiss WF 4th, Young TM, Roberts CJ. Principles, approaches, and challenges for predicting protein aggregation rates and shelf life. J Pharm Sci 2009;98(4): 1246–1277.
45. Gsponer J, Vendruscolo M. Theoretical approaches to protein aggregation. Protein Peptide Lett 2006;13(3):287–293.
46. Liu J, Nguyen MD, Andya JD, Shire SJ. Reversible self-association increases the viscosity of a concentrated monoclonal antibody in aqueous solution. J Pharm Sci 2005;94(9):1928–1940.
47. Saluja A, Kalonia DS. Nature and consequences of protein–protein interactions in high protein concentration solutions. Int J Pharm 2008;358(1–2):1–15.
48. Kanai S, Liu J, Patapoff TW, Shire SJ. Reversible self-association of a concentrated monoclonal antibody solution mediated by Fab–Fab interaction that impacts solution viscosity. J Pharm Sci 2008;97(10):4219–4227.
49. Chi EY, Krishnan S, Randolph TW, Carpenter JF. Physical stability of proteins in aqueous solution: mechanism and driving forces in nonnative protein aggregation. Pharm Res 2003;20(9):1325–1336.
50. Broersen K, Weijers M, de Groot J, Hamer RJ, de Jongh HH. Effect of protein charge on the generation of aggregation-prone conformers. Biomacromolecules 2007;8(5):1648–1656.
51. Bajaj H, Sharma VK, Badkar A, Zeng D, Nema S, Kalonia DS. Protein structural conformation and not second viral coefficient relates to long-term irreversible aggregation of a monoclonal antibody and ovalbumin in solution. Pharm Res 2006;23(6): 1382–1394.
52. Krebs MR, Devlin GL, Donald AM. Protein particulates: another generic form of protein aggregation? Biophys J 2007;92(4):1336–1342.
53. Jaenicke R. Protein stability and molecular adaptation to extreme conditions. Eur J Biochem 1991;202(3):715–728.
54. Rees DC, Robertson AD. Some thermodynamic implications for the thermostability of proteins. Protein Sci 2001;10(6):1187–1194.
55. Boctor AM, Mehta SC. Enhancement of the stability of thrombin by polyols: microcalorimetric studies. J Pharm Pharmacol 1992;44(7):600–603.
56. Charman SA, Mason KL, Charman WN. Techniques for assessing the effects of pharmaceutical excipients on the aggregation of porcine growth hormone. Pharm Res 1993;10(7):954–962.
57. Calamai M, Canale C, Relini A, Stefani M, Chiti F, Dobson CM. Reversal of protein aggregation provides evidence for multiple aggregated States. J Mol Biol 2005; 346(2):603–616.
58. Speed MA, King J, Wang DI. Polymerization mechanism of polypeptide chain aggregation. Biotechnol Bioeng 1997;54(4):333–343.

59. Arnaudov LN, de Vries R. Thermally induced fibrillar aggregation of hen egg white lysozyme. Biophys J 2005;88(1):515–526.
60. Schein CH. Solubility as a function of protein structure and solvent components. Biotechnology 1990;8(4):308–317.
61. Kohn WD, Kay CM, Hodges RS. Salt effects on protein stability: two-stranded alpha-helical coiled-coils containing inter- or intrahelical ion pairs. J Mol Biol 1997;267(4):1039–1052.
62. Wang YJ, Shahrokh Z, Vemuri S, Eberlein G, Beylin I, Busch M. Characterization, stability, and formulations of basic fibroblast growth factor. Pharm Biotechnol 1996;9:141–180.
63. Advant SJ, Braswell EH, Kumar CV, Kalonia DS. The effect of pH and temperature on the self-association of recombinant human interleukin-2 as studied by equilibrium sedimentation. Pharm Res 1995;12(5):637–641.
64. Chi EY, Krishnan S, Kendrick BS, Chang BS, Carpenter JF, Randolph TW. Roles of conformational stability and colloidal stability in the aggregation of recombinant human granulocyte colony-stimulating factor. Protein Sci 2003;12(5):903–913.
65. Gaikwad SM, Islam Khan M. pH-dependent aggregation of oligomeric *Artocarpus hirsuta* lectin on thermal denaturation. Biochem Biophys Res Commun 2003;311 (2):254–257.
66. Munishkina LA, Ahmad A, Fink AL, Uversky VN. 2008. Guiding protein aggregation with macromolecular crowding. Biochemistry 47(34):8993–9006.
67. Van Buren N, Rehder D, Gadgil H, Matsumura M, Jacob J. Elucidation of two major aggregation pathways in an IgG2 antibody. J Pharm Sci 2009;98(9):3013–3030.
68. Chiti F, Stefani M, Taddei N, Ramponi G, Dobson CM. Rationalization of the effects of mutations on peptide and protein aggregation rates. Nature 2003;424(6950):805–808.
69. DuBay KF, Pawar AP, Chiti F, Zurdo J, Dobson CM, Vendruscolo M. Prediction of the absolute aggregation rates of amyloidogenic polypeptide chains. J Mol Biol 2004;341 (5):1317–1326.
70. Majhi PR, Ganta RR, Vanam RP, Seyrek E, Giger K, Dubin PL. Electrostatically driven protein aggregation: beta-lactoglobulin at low ionic strength. Langmuir 2006; 22(22):9150–9159.
71. Paborji M, Pochopin NL, Coppola WP, Bogardus JB. Chemical and physical stability of chimeric L6, a mouse-human monoclonal antibody. Pharm Res 1994;11(5):764–771.
72. Pikal MJ, Dellerman KM, Roy ML, Riggin RM. The effects of formulation variables on the stability of freeze-dried human growth hormone. Pharm Res 1991;8(4):427–436.
73. Katayama DS, Nayar R, Chou DK, Valente JJ, Cooper J, Henry CS, et al. Effect of buffer species on the thermally induced aggregation of interferon-tau. J Pharm Sci 2006; 95(6):1212–1226.
74. Raibekas AA, Bures EJ, Siska CC, Kohno T, Latypov RF, Kerwin BA. Anion binding and controlled aggregation of human interleukin-1 receptor antagonist. Biochemistry 2005;44(29):9871–9879.
75. Kita Y, Arakawa T. Salts and glycine increase reversibility and decrease aggregation during thermal unfolding of ribonuclease-A. Biosci Biotechnol Biochem 2002; 66(4):880–882.
76. Giger K, Vanam RP, Seyrek E, Dubin PL. Suppression of insulin aggregation by heparin. Biomacromolecules 2008;9(9):2338–2344.

77. McGuffey MK, Epting KL, Kelly RM, Foegeding EA. Denaturation and aggregation of three alpha-lactalbumin preparations at neutral pH. J Agric Food Chem 2005; 53(8):3182–3190.
78. Baussay K, Bon CL, Nicolai T, Durand D, Busnel JP. Influence of the ionic strength on the heat-induced aggregation of the globular protein beta-lactoglobulin at pH 7. Int J Biol Macromol 2004;34(1–2):21–28.
79. Alford JR, Kendrick BS, Carpenter JF, Randolph TW. High concentration formulations of recombinant human interleukin-1 receptor antagonist: II. Aggregation kinetics. J Pharm Sci 2008;97(8):3005–3021.
80. Georgiou G, Valax P. Isolating inclusion bodies from bacteria. Methods Enzymol 1999;309:48–58.
81. Finke JM, Roy M, Zimm BH, Jennings PA. Aggregation events occur prior to stable intermediate formation during refolding of interleukin 1beta. Biochemistry 2000; 39(3):575–583.
82. Yazdanparast R, Esmaeili MA, Khodagholi F. Control of aggregation in protein refolding: cooperative effects of artificial chaperone and cold temperature. Int J Biol Macromol 2007;40(2):126–133.
83. De Bernardez Clark E, Schwarz E, Rudolph R. Inhibition of aggregation side reactions during in vitro protein folding. Methods Enzymol 1999;309:217–236.
84. Clark EDB. Refolding of recombinant proteins. Curr Opin Biotechnol 1998;9(2):157–163.
85. Heavner GA, Arakawa T, Philo JS, Calmann MA, Labrenz S. Protein isolated from biopharmaceutical formulations cannot be used for comparative studies: follow-up to "a case study using Epoetin Alfa from Epogen and EPREX". J Pharm Sci 2007; 96(12):3214–3225.
86. Shukla AA, Gupta P, Han X. Protein aggregation kinetics during Protein A chromatography. Case study for an Fc fusion protein. J Chromatogr A 2007;1171(1–2):22–28.
87. Eckhardt BM, Oeswein JQ, Bewley TA. Effect of freezing on aggregation of human growth hormone. Pharm Res 1991;8(11):1360–1364.
88. Wang W. Lyophilization and development of solid protein pharmaceuticals. Int J Pharm 2000;203(1–2):1–60.
89. Kiese S, Papppenberger A, Friess W, Mahler HC. Shaken, not stirred: mechanical stress testing of an IgG1 antibody. J Pharm Sci 2008;97(10):4347–4366.
90. Vessely CR, Carpenter JF, Schwartz DK. Calcium–induced changes to the molecular conformation and aggregate structure of beta–casein at the air–water interface. Biomacromolecules 2005;6(6):3334–3344.
91. Huppertz T, Fox PF, de Kruif KG, Kelly AL. High pressure-induced changes in bovine milk proteins: a review. Biochim Biophys Acta 2006;1764(3):593–598.
92. Gracia-Julia A, Rene M, Cortes MM, Picart L, Lopez PT, Chevalier D, et al. Effect of dynamic high pressure on whey protein aggregation: a comparison with the effect of continuous short-time thermal treatments. Food Hydrocolloids 2008;22(6):1014–1032.
93. Andya JD, Hsu CC, Shire SJ. Mechanisms of aggregate formation and carbohydrate excipient stabilization of lyophilized humanized monoclonal antibody formulations. AAPS PharmSci 2003;5(2):E10.
94. Morales AA, Nunez-Gandolff G, Perez NP, Veliz BC, Caballero-Torres I, Duconge J, et al. Freeze-dried formulation for direct $^{99m}$Tc-labeling ior-egf/r3 MAb: additives, biodistribution, and stability. Nucl Med Biol 1999;26(6):717–723.

95. Pace CN, Trevino S, Prabhakaran E, Scholtz JM. Protein structure, stability and solubility in water and other solvents. Philos Trans R Soc Lond B Biol Sci 2004;359 (1448):1225–1234; discussion 1234–1225.
96. Feng S, Yan YB. Effects of glycerol on the compaction and stability of the wild type and mutated rabbit muscle creatine kinase. Proteins 2008;71(2):844–854.
97. Jones LS, Kaufmann A, Middaugh CR. Silicone oil induced aggregation of proteins. J Pharm Sci 2005;94(4):918–927.
98. Ellis RJ, Minton AP. Protein aggregation in crowded environments. Biol Chem 2006; 387(5):485–497.
99. Hull E, Spoja C, Cordova M, Cohlberg JA. Neurofilament protein aggregation in a cell line model system. Biochem Biophys Res Commun 2008;366(1):73–79.
100. Hermeling S, Schellekens H, Maas C, Gebbink MF, Crommelin DJ, Jiskoot W. Antibody response to aggregated human interferon alpha2b in wild-type and transgenic immune tolerant mice depends on type and level of aggregation. J Pharm Sci 2006;95(5):1084–1096.
101. Rosenberg AS. Effects of protein aggregates: an immunologic perspective. AAPS J 2006;8(3):E501–507.
102. Braun A, Kwee L, Labow MA, Alsenz J. Protein aggregates seem to play a key role among the parameters influencing the antigenicity of interferon alpha (IFN-alpha) in normal and transgenic mice. Pharm Res 1997;14(10):1472–1478.
103. Lougheed WD, Albisser AM, Martindale HM, Chow JC, Clement JR. Physical stability of insulin formulations. Diabetes 1983;32(5):424–432.
104. Brange J, Owens DR, Kang S, Volund A. Monomeric insulins and their experimental and clinical implications. Diabetes Care 1990;13(9):923–954.
105. Jordan GM, Yoshioka S, Terao T. The aggregation of bovine serum albumin in solution and in the solid state. J Pharm Pharmacol 1994;46(3):182–185.
106. Yoshioka S, Aso Y, Izutsu K, Terao T. Aggregates formed during storage of beta-galactosidase in solution and in the freeze-dried state. Pharm Res 1993;10(5):687–691.
107. Livney YD, Dalgleish DG. Specificity of disulfide bond formation during thermal aggregation in solutions of beta-lactoglobulin B and kappa-casein A. J Agric Food Chem 2004;52(17):5527–5532.
108. Costantino HR, Langer R, Klibanov AM. Moisture-induced aggregation of lyophilized insulin. Pharm Res 1994;11(1):21–29.
109. Khossravi M, Shire SJ, Borchardt RT. Evidence for the involvement of histidine A(12) in the aggregation and precipitation of human relaxin induced by metal-catalyzed oxidation. Biochemistry 2000;39(19):5876–5885.
110. Liang JN, Rossi MT. In vitro non-enzymatic glycation and formation of browning products in the bovine lens alpha-crystallin. Exp Eye Res 1990;50(4):367–371.
111. Sluzky V, Tamada JA, Klibanov AM, Langer R. Kinetics of insulin aggregation in aqueous solutions upon agitation in the presence of hydrophobic surfaces. Proc Natl Acad Sci USA 1991;88(21):9377–9381.
112. Brange J, Havelund S, Hougaard P. Chemical stability of insulin. 2. Formation of higher molecular weight transformation products during storage of pharmaceutical preparations. Pharm Res 1992;9(6):727–734.
113. Sluzky V, Klibanov AM, Langer R. Mechanism of insulin aggregation and stabilization in agitated aqueous solutions. Biotechnol Bioeng 1992;40(8):895–903.

114. Robinson AB, McKerrow JH, Cary P. Controlled deamidation of peptides and proteins: an experimental hazard and a possible biological timer. Proc Natl Acad Sci USA 1970; 66(3):753–757.

115. Wingfield PT, Mattaliano RJ, MacDonald HR, Craig S, Clore GM, Gronenborn AM, et al. Recombinant-derived interleukin-1 alpha stabilized against specific deamidation. Protein Eng 1987;1(5):413–417.

116. Skottner A, Forsman A, Skoog B, Kostyo JL, Cameron CM, Adamafio NA, et al. Biological characterization of charge isomers of human growth hormone. Acta Endocrinol 1988;118(1):14–21.

117. Ahern TJ, Klibanov AM. The mechanisms of irreversible enzyme inactivation at 100C. Science 1985;228(4705):1280–1284.

118. Flatmark T. Multiple molecular forms of bovine heart cytochrome c. V. A comparative study of their physicochemical properties and their reactions in biological systems. J Biol Chem 1967;242(10):2454–2459.

119. Tallan HH, Stein WH. Chromatographic studies on lysozyme. J Biol Chem 1953;200 (2):507–514.

120. Lewis UJ, Cheever EV, Hopkins WC. Kinetic study of the deamidation of growth hormone and prolactin. Biochim Biophys Acta 1970;214(3):498–508.

121. DiAugustine RP, Gibson BW, Aberth W, Kelly M, Ferrua CM, Tomooka Y, et al. Evidence for isoaspartyl (deamidated) forms of mouse epidermal growth factor. Anal Biochem 1987;165(2):420–429.

122. Battersby AR, Robinson JC. Studies on specific chemical fission of peptide links. Part I. The rearrangement of aspartyl and glutamyl peptides. J Chem Soc 1955; 259–269.

123. Haley EE, Corcoran BJ, Dorer FE, Buchanan DL. Beta-aspartyl peptides in enzymatic hydrolysates of protein. Biochemistry 1966;5(10):3229–3235.

124. Meinwald YC, Stimson ER, Scheraga HA. Deamidation of the asparaginyl-glycyl sequence. Int J Peptide Protein Res 1986;28(1):79–84.

125. Bornstein P, Balian G. Cleavage at Asn-Gly bonds with hydroxylamine. Methods Enzymol 1977;47:132–145.

126. Geiger T, Clarke S. Deamidation, isomerization, and racemization at asparaginyl and aspartyl residues in peptides. Succinimide-linked reactions that contribute to protein degradation. J Biol Chem 1987;262(2):785–794.

127. Tyler-Cross R, Schirch V. Effects of amino acid sequence, buffers, and ionic strength on the rate and mechanism of deamidation of asparagine residues in small peptides. J Biol Chem 1991;266(33):22549–22556.

128. Robinson AB, Tedro S. Sequence dependent deamidation rates for model peptides of hen egg-white lysozyme. Int J Peptide Protein Res 1973;5(4):275–278.

129. Robinson AB, Rudd CJ. Deamidation of glutaminyl and asparaginyl residues in peptides and proteins. Curr Top Cell Regul 1974;8:247–295.

130. Wearne SJ, Creighton TE. Effect of protein conformation on rate of deamidation: ribonuclease A. Proteins 1989;5(1):8–12.

131. Chazin WJ, Kordel J, Thulin E, Hofmann T, Drakenberg T, Forsen S. Identification of an isoaspartyl linkage formed upon deamidation of bovine calbindin D9k and structural characterization by 2D $^{1}$H NMR. Biochemistry 1989;28(21):8646–8653.

132. Rogers SW, Rechsteiner M. Degradation of structurally characterized proteins injected into HeLa cells. Tests of hypotheses. J Biol Chem 1988;263(36):19850–19862.

133. Verma A, McNichol B, Dominguez-Castillo RI, Amador-Molina JC, Arciniega JL, Reiter K, et al. Use of site-directed mutagenesis to model the effects of spontaneous deamidation on the immunogenicity of *Bacillus anthracis* protective antigen. Infect Immun 2012;81(1):278–284.

134. Hermeling S, Crommelin DJ, Schellekens H, Jiskoot W. Structure-immunogenicity relationships of therapeutic proteins. Pharm Res 2004;21(6):897–903.

135. Teh LC, Murphy LJ, Huq NL, Surus AS, Friesen HG, Lazarus L, et al. Methionine oxidation in human growth hormone and human chorionic somatomammotropin. Effects on receptor binding and biological activities. J Biol Chem 1987;262(14):6472–6477.

136. Holeysovsky V, Lazdunski M. The structural properties of trypsinogen and trypsin. Alkylation and oxidation of methionines. Biochim Biophys Acta 1968;154(3):457–467.

137. Harris CM, Hillrl. The carboxymethylation of human metmyoglobin. J Biol Chem 1969;244(8):2195–2203.

138. Penner MH, Yamasaki RB, Osuga DT, Babin DR, Mearea CF, Feeney RE. Comparative oxidations of tyrosines and methionines in transferrins: human serum transferrin, human lactotransferrin, and chicken ovotransferrin. Arch Biochem Biophys 1983;225(2):740–747.

139. Brot N, Weissbach H. Biochemistry of methionine sulfoxide residues in proteins. BioFactors 1991;3(2):91–96.

140. Tashjian AH Jr, Ontjes DA, Munson PL. Alkylation and oxidation of methionine in bovine parathyroid hormone: effects on hormonal activity and antigenicity. Biochemistry 1964;3:1175–1182.

141. Dedman ML, Farmer TH, Morris CJ. Studies on pituitary adrenocorticotrophin. 3. Identification of the oxidation-reduction centre. Biochem J 1961;78:348–352.

142. Kenkare UW, Richards FM. The histidyl residues in ribonuclease-S. Photooxidation in solution and in single crystals; the iodination of histidine-12. J Biol Chem 1966;241 (13):3197–3206.

143. Jori G, Galiazzo G, Marzotto A, Scoffone E. Selective and reversible photo-oxidation of the methionyl residues in lysozyme. J Biol Chem 1968;243(16):4272–4278.

144. George-Nascimento C, Gyenes A, Halloran SM, Merryweather J, Valenzuela P, Steimer KS, et al. Characterization of recombinant human epidermal growth factor produced in yeast. Biochemistry 1988;27(2):797–802.

145. Heath WF, Merrifield RB. A synthetic approach to structure-function relationships in the murine epidermal growth factor molecule. Proc Natl Acad Sci USA 1986;83(17):6367–6371.

146. Jocelyn PC. Biochemistry of the SH Groups: The Occurrence, Chemical Properties, Metabolism, and Biological Function of Thiols and Disulfides. New York: Academic Press; 1972.

147. Philipson L. On the oxidation of 2,3-dimercaptopropanol. Biochim Biophys Acta 1962;56:375–377.

148. Neumann NP, Moore S, Stein WH. Modification of the methionine residues in ribonuclease. Biochemistry 1962;1:68–75.

149. Jaques LB, Bell HJ. The reduction of hydrogen peroxide by fibrin. Can J Res 1946; 24(Sect E):79–83.

150. Means GE, Feeney RE. Chemical Modifications of Proteins. San Francisco: Holden-Day; 1971.
151. Little C, O'Brien PJ. 1967. Products of oxidation of a protein thiol group after reaction with various oxidizing agents. Arch Biochem Biophys 122(2):406–410.
152. Hu D, Qin Z, Xue B, Fink AL, Uversky VN. Effect of methionine oxidation on the structural properties, conformational stability, and aggregation of immunoglobulin light chain LEN. Biochemistry 2008;47(33):8665–8677.
153. Stadtman ER. Metal ion-catalyzed oxidation of proteins: biochemical mechanism and biological consequences. Free Radic Biol Med 1990;9(4):315–325.
154. Hiller KO, Masloch B, Goebl M, Asmus KD. Mechanism of the hydroxyl radical induced oxidation of methionine in aqueous solution. J Am Chem Soc 1981;103 (10):2734–2743.
155. Davies KJ, Lin SW, Pacifici RE. Protein damage and degradation by oxygen radicals. IV. Degradation of denatured protein. J Biol Chem 1987;262(20):9914–9920.
156. Shima K, Sawazaki N, Tanaka R, Tarui S, Nishikawa M. Effect of an exposure to chloramine-T on the immunoreactivity of glucagon. Endocrinology 1975;96(5):1254–1260.
157. Glimcher LH, Schroer JA, Chan C, Shevach EM. Fine specificity of cloned insulin-specific T cell hybridomas: evidence supporting a role for tertiary conformation. J Immunol 1983;131(6):2868–2874.
158. Schultz J. Cleavage at aspartic acid. Methods Enzymol 1967;11:255–263.
159. Inglis AS. Cleavage at aspartic acid. Methods Enzymol 1983;91:324–332.
160. Marcus F. Preferential cleavage at aspartyl-prolyl peptide bonds in dilute acid. Int J Peptide Protein Res 1985;25(5):542–546.
161. Cecil R, McPhee JR. The sulfur chemistry of proteins. Adv Protein Chem 1959;14:255–389.
162. Lumper L, Zahn H. [Chemistry and biochemistry of disulfide exchanges]. Adv Enzymol Relat Areas Mol Biol 1965;27:199–237.
163. Ryle AP, Sanger F. Disulphide interchange reactions. Biochem J 1955;60(4):535–540.
164. Kerwin BA, Remmele RL Jr. Protect from light: photodegradation and protein biologics. J Pharm Sci 2007;96(6):1468–1479.
165. Vanhooren A, Devreese B, Vanhee K, Van Beeumen J, Hanssens I. Photoexcitation of tryptophan groups induces reduction of two disulfide bonds in goat alpha-lactalbumin. Biochemistry 2002;41(36):11035–11043.
166. Friedman M, Masters PM. Kinetics of racemization of amino acid residues in casein. J Food Sci 1982;47:760–764.
167. Volkin DB, Klibanov AM. Thermal destruction processes in proteins involving cystine residues. J Biol Chem 1987;262(7):2945–2950.
168. Florence TM. Degradation of protein disulphide bonds in dilute alkali. Biochem J 1980;189(3):507–520.
169. Whitaker JR, Feeney RE. Chemical and physical modification of proteins by the hydroxide ion. Crit Rev Food Sci Nutr 1983;19(3):173–212.
170. Kreilgaard L, Jones LS, Randolph TW, Frokjaer S, Flink JM, Manning MC, et al. Effect of Tween 20 on freeze-thawing- and agitation-induced aggregation of recombinant human factor XIII. J Pharm Sci 1998;87(12):1597–1603.

171. Andrade JD, Hilady V, Feng L, Tingey K. Proteins at interfaces: principles, problems, and potential. In: Brash JL, Wojciechowskik PW, editors. Interfacial Phenomena and Bioproducts. New York: Marcel Dekker; 1996. p. 19–55.

172. Van Den Berg L, Rose D. Effect of freezing on the pH and composition of sodium and potassium phosphate solutions; the reciprocal system $KH_2PO_4$-$Na_2$-$HPO_4$-$H_2O$. Arch Biochem Biophys 1959;81(2):319–329.

173. Van Den Berg L. The effect of addition of sodium and potassium chloride to the reciprocal system: $KH_2PO_4$-$Na_2HPO_4$-$H_2O$ on pH and composition during freezing. Arch Biochem Biophys 1959;84:305–315.

174. Randolph TW. Phase separation of excipients during lyophilization: effects on protein stability. J Pharm Sci 1997;86(11):1198–1203.

175. Bauer R, Muller A, Richter M, Schneider K, Frey J, Engelhardt W. Influence of heavy metal ions on antibodies and immune complexes investigated by dynamic light scattering and enzyme-linked immunosorbent assay. Biochim Biophys Acta 1997;1334(1):98–108.

176. Stadtman ER, Levine RL. Protein oxidation. Ann NY Acad Sci 2000;899:191–208.

177. Potier M, Thauvette L, Michaud L, Giroux S, Beauregard G. Inactivation mechanism of tetrameric beta-galactosidase by gamma-rays involves both fragmentation and temperature-dependent denaturation of protomers. Biochemistry 1991;30(33):8151–8157.

178. Gu LC, Erdos EA, Chiang HS, Calderwood T, Tsai K, Visor GC, et al. Stability of interleukin 1 beta (IL-1 beta) in aqueous solution: analytical methods, kinetics, products, and solution formulation implications. Pharm Res 1991;8(4):485–490.

179. Matthew JB, Gurd FR, Garcia-Moreno B, Flanagan MA, March KL, Shire SJ. pH-dependent processes in proteins. CRC Crit Rev Biochem 1985;18(2):91–197.

180. Franks F. Freeze-drying: from empiricism to predictability. The significance of glass transitions. Dev Biol Stand 1992;74:9–18; discussion 19.

181. Hatley RHM, Franks F, Day H. The effect on the oxidation of ascorbic acid of freeze-concentration and cooling. Biophys Chem 1986;24:187–192.

182. Pikal-Cleland KA, Cleland JL, Anchordoquy TJ, Carpenter JF. Effect of glycine on pH changes and protein stability during freeze-thawing in phosphate buffer systems. J Pharm Sci 2002;91(9):1969–1979.

183. Chang BS, Kendrick BS, Carpenter JF. Surface-induced denaturation of proteins during freezing and its inhibition by surfactants. J Pharm Sci 1996;85(12):1325–1330.

184. Privalov PL. Cold denaturation of proteins. Crit Rev Biochem Mol Biol 1990;25(4):281–305.

185. Millqvist-Fureby A, Malmsten M, Bergenstahl B. Surface characterisation of freeze-dried protein/carbohydrate mixtures. Int J Pharm 1999;191(2):103–114.

186. Miller R, Fainerman VB, Makievski AV, Kragel J, Grigoriev DO, Kazakov VN, et al. Dynamics of protein and mixed protein/surfactant adsorption layers at the water/fluid interface. Advances in colloid and interface science 2000;86(1–2):39–82.

187. Lashmar UT, Vnanderburgh M, Little SJ. Bulk freeze-thaw of macromolecules: effect of cryoconcentration on their formulation and stability. BioProcess Int 2007;5(6):44–54.

188. Carpenter JF, Pikal MJ, Chang BS, Randolph TW. Rational design of stable lyophilized protein formulations: some practical advice. Pharm Res 1997;14(8):969–975.

189. Mozhaev VV, Martinek K. Structure-stability relationships in proteins: new approaches to stabilizing enzymes. Enzyme Microb Technol 1984;6:50–59.

190. Allison SD, Dong A, Carpenter JF. Counteracting effects of thiocyanate and sucrose on chymotrypsinogen secondary structure and aggregation during freezing, drying, and rehydration. Biophys J 1996;71(4):2022–2032.
191. Hsu CC, Nguyen HM, Yeung DA, Brooks DA, Koe GS, Bewley TA, et al. Surface denaturation at solid-void interface--a possible pathway by which opalescent particulates form during the storage of lyophilized tissue-type plasminogen activator at high temperatures. Pharm Res 1995;12(1):69–77.
192. Strambini GB, Gabellieri E. Proteins in frozen solutions: evidence of ice-induced partial unfolding. Biophys J 1996;70(2):971–976.
193. Wisniewski R. Principles of large-scale cryopreservation of cells, microorganisms, protein solutions, and biological products. In: Avis KE, Wagner CM, editors. Cryopreservation: Applications in Pharmaceuticals and Biotechnology. Denver: Interpharm Press; 1999. p 11–179.
194. Wisniewski R. Large-scale cryopreservation: process development for freezing and thawing of large volume of cell suspensions, protein solutions, and biological products. In: Avis KE, Wagner CM, editors. Cryopreservation: Applications in Pharmaceuticals and Biotechnology. Denver: Interpharm Press; 1999. p. 181–313.
195. Wisniewski R, Wu V. Large-Scale Freezing and Thawing of Biopharmaceutical Products. Buffalo Grove, IL: Interpharm Press; 1996.
196. Arakawa T, Timasheff SN. Mechanism of protein salting in and salting out by divalent cation salts: balance between hydration and salt binding. Biochemistry 1984;23 (25):5912–5923.
197. Arakawa T, Timasheff SN. Preferential interactions of proteins with salts in concentrated solutions. Biochemistry 1982;21(25):6545–6552.
198. Stellwagen E, Babul J. Stabilization of the globular structure of ferricytochrome c by chloride in acidic solvents. Biochemistry 1975;14(23):5135–5140.
199. Lee JC, Timasheff SN. The stabilization of proteins by sucrose. J Biol Chem 1981;256 (14):7193–7201.
200. Gekko K, Timasheff SN. Mechanism of protein stabilization by glycerol: preferential hydration in glycerol–water mixtures. Biochemistry 1981;20(16):4667–4676.
201. Kuntz ID Jr, Kauzmann W. Hydration of proteins and polypeptides. Adv Protein Chem 1974;28:239–345.
202. Chang D, Chang R. Review of current issues in pharmaceutical excipients. Pharm Technol 2007.
203. Navarra G, Leone M, Militello V. Thermal aggregation of beta-lactoglobulin in presence of metal ions. Biophys Chem 2007;131(1–3):52–61.
204. Mounsey JS, O'Kennedy BT. Conditions limiting the influence of thiol-disulphide interchange reactions on the heat-induced aggregation kinetics of beta-lactoglobulin. Int Dairy J 2007;17(9):1034–1042.
205. Timasheff SN. Solvent effects on protein stability. Curr Opin Struct Biol 1992;2(1): 35–39.
206. Timasheff SN, Arakawa T. Mechanism of protein precipitation and stabilization by co-solvents. J Crystal Growth 1988;90(1–3):39–46.
207. Arakawa T, Timasheff SN. Mechanism of poly(ethylene glycol) interaction with proteins. Biochemistry 1985;24(24):6756–6762.

208. Arakawa T, Timasheff SN. The stabilization of proteins by osmolytes. Biophys J 1985; 47(3):411–414.

209. Dill KA, Shortle D. Denatured states of proteins. Annu Rev Biochem 1991;60:795–825.

210. Schellman JA. Selective binding and solvent denaturation. Biopolymers 1987; 26(4):549–559.

211. Cleland JL, Wang DIC. Cosolvent Effects on Refolding and Aggregation. Washington, DC: American Chemical Society; 1993.

212. Kerwin BA. Polysorbates 20 and 80 used in the formulation of protein biotherapeutics: structure and degradation pathways. J Pharm Sci 2008;97(8):2924–2935.

213. Wang A, Robertson AD, Bolen DW. Effects of a naturally occurring compatible osmolyte on the internal dynamics of ribonuclease A. Biochemistry 1995;34(46): 15096–15104.

214. Frye KJ, Royer CA. The kinetic basis for the stabilization of staphylococcal nuclease by xylose. Protein Sci 1997;6(4):789–793.

215. Kendrick BS, Chang BS, Arakawa T, Peterson B, Randolph TW, Manning MC, et al. Preferential exclusion of sucrose from recombinant interleukin-1 receptor antagonist: role in restricted conformational mobility and compaction of native state. Proc Natl Acad Sci USA 1997;94(22):11917–11922.

216. Jacob M, Schmid FX. Protein folding as a diffusional process. Biochemistry 1999;38 (42):13773–13779.

217. Soenderkaer S, Carpenter JF, van de Weert M, Hansen LL, Flink J, Frokjaer S. Effects of sucrose on rFVIIa aggregation and methionine oxidation. European J Pharm Sci 2004; 21(5):597–606.

218. Rezaei-Ghaleh N, Ramshini H, Ebrahim-Habibi A, Moosavi-Movahedi AA, Nemat-Gorgani M. Thermal aggregation of alpha-chymotrypsin: role of hydrophobic and electrostatic interactions. Biophys Chem 2008;132(1):23–32.

219. Hamada H, Takahashi R, Noguchi T, Shiraki K. Differences in the effects of solution additives on heat- and refolding-induced aggregation. Biotechnol Prog 2008;24(2):436–443.

220. Saluja A, Badkar AV, Zeng DL, Kalonia DS. Ultrasonic rheology of a monoclonal antibody (IgG2) solution: implications for physical stability of proteins in high concentration formulations. J Pharm Sci 2007;96(12):3181–3195.

221. Narhi LO, Philo JS, Sun B, Chang BS, Arakawa T. Reversibility of heat-induced denaturation of the recombinant human megakaryocyte growth and development factor. Pharm Res 1999;16(6):799–807.

222. Tsai PK, Volkin DB, Dabora JM, Thompson KC, Bruner MW, Gress JO, et al. Formulation design of acidic fibroblast growth factor. Pharm Res 1993;10(5):649–659.

223. Gonzalez M, Murature DA, Fidelio GD. Thermal stability of human immunoglobulins with sorbitol. A critical evaluation. Vox Sang 1995;68(1):1–4.

224. Xia Y, Park YD, Mu H, Zhou HM, Wang XY, Meng FG. The protective effects of osmolytes on arginine kinase unfolding and aggregation. Int J Biol Macromol 2007; 40(5):437–443.

225. Rondeau P, Armenta S, Caillens H, Chesne S, Bourdon E. Assessment of temperature effects on beta-aggregation of native and glycated albumin by FTIR spectroscopy and PAGE: relations between structural changes and antioxidant properties. Arch Biochem Biophys 2007;460(1):141–150.

226. Randolph TW, Jones LS. Surfactant-protein interactions. Pharm Biotechnol 2002;13:159–175.
227. Chou DK, Krishnamurthy R, Randolph TW, Carpenter JF, Manning MC. Effects of Tween 20 and Tween 80 on the stability of Albutropin during agitation. J Pharm Sci 2005;94(6):1368–1381.
228. Twardowski ZJ, Nolph KD, McGary TJ, Moore HL. Nature of insulin binding to plastic bags. Am J Hosp Pharm 1983;40(4):579–582.
229. Bohnert J, Horbett T. Changes in adsorbed fibrinogen and albumin interactions with polymers indicated by decreases in detergent elutability. J Coll Interface Sci 1986;111:363–377.
230. Tandon S, Horowitz PM. Detergent-assisted refolding of guanidinium chloride-denatured rhodanese. The effects of the concentration and type of detergent. J Biol Chem 1987;262(10):4486–4491.
231. Piatigorsky J, Horwitz J, Simpson RT. Partial dissociation and renaturation of embryonic chick delta-crystallin. Characterization by ultracentrifugation and circular dichroism. Biochim Biophys Acta 1977;490(2):279–289.
232. Kerstens S, Murray BS, Dickinson E. Confocal microscopy of heat-induced aggregation and gelation of beta-lactoblobulin in presence of non-ionic surfactant. Food Hydrocolloids 2005;19(3):625–633.
233. Won CM, Molnar TE, McKean RE, Spenlehauer GA. Stabilizers against heat-induced aggregation of RPR 114849, an acidic fibroblast growth factor (AFGF). Int J Pharm 1998;167:25–36.
234. Collins T, D'Amico S, Georlette D, Marx JC, Huston AL, Feller G. A nondetergent sulfobetaine prevents protein aggregation in microcalorimetric studies. Anal Biochem 2006;352(2):299–301.
235. Rozema D, Gellman SH. Artificial chaperone-assisted refolding of carbonic anhydrase B. J Biol Chem 1996;271(7):3478–3487.
236. Panyukov YV, Nemykh MA, Rafikova ER, Kurganov BI, Yaguzhinsky LS, Arutyunyan AM, et al. Low cetyltrimethylammonium bromide concentrations induce reversible amorphous aggregation of tobacco mosaic virus and its coat protein at room temperature. Int J Biochem Cell Biol 2006;38(4):533–543.
237. Maclean DS, Qian Q, Middaugh CR. Stabilization of proteins by low molecular weight multi-ions. J Pharm Sci 2002;91(10):2220–2229.
238. Unterhaslberger G, Schmitt C, Sanchez C, Appolonia NC, Raemy A. Heat denaturation and aggregation of beta-lactoglobulin enriched WPI in the presence of arginine HCl, NaCl and guanidinium HCl at pH 4.0 and 7.0. Food Hydrocolloids 2006;20(7):1006–1019.
239. Chen B, Bautista R, Yu K, Zapata GA, Mulkerrin MG, Chamow SM. Influence of histidine on the stability and physical properties of a fully human antibody in aqueous and solid forms. Pharm Res 2003;20(12):1952–1960.
240. Ecroyd H, Carver JA. The effect of small molecules in modulating the chaperone activity of alphaB-crystallin against ordered and disordered protein aggregation. FEBS J 2008;275(5):935–947.
241. Lyutova EM, Kasakov AS, Gurvits BY. Effects of arginine on kinetics of protein aggregation studied by dynamic laser light scattering and tubidimetry techniques. Biotechnol Prog 2007;23(6):1411–1416.

242. Tsumoto K, Ejima D, Kita Y, Arakawa T. Review: why is arginine effective in suppressing aggregation? Protein Peptide Lett 2005;12(7):613–619.
243. Baynes BM, Wang DI, Trout BL. Role of arginine in the stabilization of proteins against aggregation. Biochemistry 2005;44(12):4919–4925.
244. Huus K, Havelund S, Olsen HB, van de Weert M, Frokjaer S. Thermal dissociation and unfolding of insulin. Biochemistry 2005;44(33):11171–11177.
245. Alcaraz LA, Donaire A. Unfolding process of rusticyanin: evidence of protein aggregation. Eur J Biochem 2004;271(21):4284–4292.
246. Derrick TS, Kashi RS, Durrani M, Jhingan A, Middaugh CR. Effect of metal cations on the conformation and inactivation of recombinant human factor VIII. J Pharm Sci 2004;93(10):2549–2557.
247. Giese A, Levin J, Bertsch U, Kretzschmar H. Effect of metal ions on de novo aggregation of full-length prion protein. Biochem Biophys Res Commun 2004;320(4):1240–1246.
248. Vrkljan M, Foster TM, Powers ME, Henkin J, Porter WR, Staack H, et al. Thermal stability of low molecular weight urokinase during heat treatment. II. Effect of polymeric additives. Pharm Res 1994;11(7):1004–1008.
249. Calamai M, Taddei N, Stefani M, Ramponi G, Chiti F. Relative influence of hydrophobicity and net charge in the aggregation of two homologous proteins. Biochemistry 2003;42(51):15078–15083.
250. Treuheit MJ, Kosky AA, Brems DN. Inverse relationship of protein concentration and aggregation. Pharm Res 2002;19(4):511–516.
251. Minton AP. Macromolecular crowding. Curr Biol 2006;16(8):R269–271.
252. Schlieben NH, Niefind K, Schomburg D. Expression, purification, and aggregation studies of His-tagged thermoalkalophilic lipase from *Bacillus thermocatenulatus*. Protein Expr Purif 2004;34(1):103–110.
253. Gilbert JB, Price VE, Greenstein JP. Effect of anions on the non-enzymatic desamidation of glutamine. J Biol Chem 1949;180(1):209–218.
254. Steinhardt J, Fugitt C. Catalyzed hydrolysis of amide and peptide bonds in proteins. J Res Natl Bur Stand (US) 1942;29:315–327.
255. Brennan TV, Clarke S. Spontaneous degradation of polypeptides at aspartyl and asparaginyl residues: effects of the solvent dielectric. Protein Sci 1993;2(3):331–338.
256. Scotchler JW, Robinson AB. Deamidation of glutaminyl residues: dependence on pH, temperature, and ionic strength. Anal Biochem 1974;59(1):319–322.
257. McKerrow JH, Robinson AB. Deamidation of asparaginyl residues as a hazard in experimental protein and peptide procedures. Anal Biochem 1971;42(2):565–568.
258. Hora MS, Rana RK, Wilcox CL, Katre NV, Hirtzer P, Wolfe SN, et al. Development of a lyophilized formulation of interleukin-2. Dev Biol Stand 1992;74:295–303.
259. Donbrow M, Azaz E, Pillersdorf A. Autoxidation of polysorbates. J Pharm Sci 1978;67 (12):1676–1681.
260. Hamburger R, Azaz E, Donbrow M. Autoxidation of polyoxyethylenic non-ionic surfactants and of polyethylene glycols. Pharm Acta Helv 1975;50(1–2):10–17.
261. Stockdale D. Overview of aseptic fill/finish manufacturing. Am Pharm Rev 2004.
262. Pikal MJ, Shah S. Moisture transfer from stopper to product and resulting stability implications. Dev Biol Stand 1991;74:165–179.

263. Allain LK, Wang Q. Impact of package leachables on the stability of pharmaceutical products. Am Pharm Rev 2007;10(4):38–44.
264. DeGrazio FL. Focus on fill and finish: parenteral packaging concerns for biotech drugs: compatibility is key. BioProcess Int 2006;4(2):12–16.
265. Maa YF, Hsu CC. Protein denaturation by combined effect of shear and air-liquid interface. Biotechnol Bioeng 1997;54(6):503–512.
266. Sharma B. Immunogenicity of therapeutic proteins. Part 2: impact of container closures. Biotechnol Adv 2007;25(3):318–324.
267. Lu X, Pikal MJ. Freeze-drying of mannitol-trehalose-sodium chloride-based formulations: the impact of annealing on dry layer resistance to mass transfer and cake structure. Pharm Devel Technol 2004;9(1):85–95.
268. Nail SL, Her LM, Proffitt CP, Nail LL. An improved microscope stage for direct observation of freezing and freeze drying. Pharm Res 1994;11(8):1098–1100.
269. Levine H, Slade L. Another view of trehalose for drying and stabilizing biological materials. BioPharm 1992;5:36–40.
270. Koster KL, Webb MS, Bryant G, Lynch DV. Interactions between soluble sugars and POPC (1-palmitoyl-2-oleoylphosphatidylcholine) during dehydration: vitrification of sugars alters the phase behavior of the phospholipid. Biochim Biophys Acta 1994;1193 (1):143–150.
271. Ohtake S, Wang YJ. Trehalose: current use and future applications. J Pharm Sci 2011;100(6):2020–2053.
272. Salnikova MS, Middaugh CR, Rytting JH. Stability of lyophilized human growth hormone. Int J Pharm 2008;358(1–2):108–113.
273. Carpenter JF, Arakawa T, Crowe JH. Interactions of stabilizing additives with proteins during freeze-thawing and freeze-drying. Dev Biol Stand 1992;74:225–238.
274. Pikal MJ. Freeze-drying of proteins. Part II: formulation selection. BioPharm 1990; 3(9):26–30.
275. Carpenter JF, Crowe JH, Arakawa T. Comparison of solute-induced protein stabilization in aqueous solution and in the frozen and dried states. J Dairy Sci 1990;73(12):3627–3636.
276. Greiff D. Protein structure and freeze-drying: the effects of residual moisture and gases. Cryobiology 1971;8(2):145–152.
277. Webb SD, Cleland JL, Carpenter JF, Randolph TW. A new mechanism for decreasing aggregation of recombinant human interferon-gamma by a surfactant: slowed dissolution of lyophilized formulations in a solution containing 0.03% polysorbate 20. J Pharm Sci 2002;91(2):543–558.
278. Masters K. Spray Drying Handbook. 4th ed. New York: John Wiley & Sons; 1985.
279. Ohtake S, Martin RA, Yee L, Chen D, Kristensen DD, Lechuga-Ballesteros D, et al. Heat-stable measles vaccine produced by spray drying. Vaccine 2010;28(5):1275–1284.
280. Maa YF, Ameri M, Shu C, Payne LG, Chen D. Influenza vaccine powder formulation development: spray-freeze-drying and stability evaluation. J Pharm Sci 2004; 93(7):1912–1923.
281. Costantino HR, Firouzabadian L, Hogeland K, Wu C, Beganski C, Carrasquillo KG, et al. Protein spray-freeze drying. Effect of atomization conditions on particle size and stability. Pharm Res 2000;17(11):1374–1383.

282. Sharma VK, Kalonia DS. Effect of vacuum drying on protein-mannitol interactions: the physical state of mannitol and protein structure in the dried state. AAPS PharmSciTech 2004;5(1):E10.

283. Miller DP, Anderson RE, de Pablo JJ. Stabilization of lactate dehydrogenase following freeze thawing and vacuum-drying in the presence of trehalose and borate. Pharm Res 1998;15(8):1215–1221.

284. Ohtake S, Martin RA, Saxena A, Lechuga-Ballesteros D, Santiago AE, Barry EM, et al. Formulation and stabilization of *Francisella tularensis* live vaccine strain. J Pharm Sci 2011;100(8):3076–3087.

285. Ohtake S, Martin R, Saxena A, Pham B, Chiueh G, Osorio M, et al. Room temperature stabilization of oral, live attenuated *Salmonella enterica* serovar Typhi-vectored vaccines. Vaccine 2011;29(15):2761–2771.

286. Pisal S, Wawde G, Salvankar S, Lade S, Kadam S. Vacuum foam drying for preservation of LaSota virus: effect of additives. AAPS PharmSciTech 2006;7(3):60.

287. Glatz CE. Modeling of aggregation-precipitation phenomena. In: Ahern TJ, Manning MC, editors. Stability of Protein Pharmaceuticals. New York: Plenum Press; 1992. p. 135–166.

288. Shiloach J, Martin N, Moes H. Tangential flow filtration. Adv Biotechnol Processes 1988;8:97–125.

289. Genovesi CS. Several uses for tangential-flow filtration in the pharmaceutical industry. J Parenter Sci Technol 1983;37(3):81–86.

290. Poupitch G. Fluid bed drying in the laboratory. Am Biotechnol Lab 1994;12(4):30–34.

291. Avis KE. Parenteral preparations. In: Gennaro AR, editor. Remington's Pharmaceutical Sciences. 18th ed. Easton, PA: Mack Publishing Company; 1990. p. 1545–1569.

292. Shire SJ, Shahrokh Z, Liu J. Challenges in the development of high protein concentration formulations. J Pharm Sci 2004;93(6):1390–1402.

293. Pecora R. Quasi-elastic light scattering from macromolecules. Annu Rev Biophys Bioeng 1972;1:257–276.

294. Georgalis Y, Saenger W. Light scattering studies on supersaturated protein solutions. Sci Prog 1999;82(Pt 4):271–294.

295. Schuck P, Perugini MA, Gonzales NR, Howlett GJ, Schubert D. Size-distribution analysis of proteins by analytical ultracentrifugation: strategies and application to model systems. Biophys J 2002;82(2):1096–1111.

296. Minton AP. Analytical centrifugation with preparative ultracentrifuges. Anal Biochem 1989;176(2):209–216.

297. Capelle MA, Gurny R, Arvinte T. High throughput screening of protein formulation stability: practical considerations. Eur J Pharm Biopharm 2007;65(2):131–148.

298. Capelle MA, Gurny R, Arvinte T. A high throughput protein formulation platform: case study of salmon calcitonin. Pharm Res 2009;26(1):118–128.

299. Shire S. Formulation and manufacturability of biologics. Curr Opin Biotech 2009;20 (6):708–714.

# Regulatory Strategies and Lessons in the Development of Biosimilars

UMANG S. SHAH

## 12.1 BACKGROUND AND OBJECTIVE

In 1982, human insulin was introduced as the first recombinant-derived therapeutic protein (cloned by Genentech scientists and licensed to Eli Lilly and Company). Since then, more than 200 different lifesaving or greatly improving "quality of life" protein drugs have been introduced, and this class of therapeutics has become firmly established [1]. However, protein therapeutics are typically very expensive compared with small molecule drugs (one study has stated that the average biologic is about 22 times more expensive than the average small molecule innovative drug) [2]. A year's worth of breast cancer–fighting Herceptin, a biopharmaceutical, for example, averages $54,000; Cerezyme for treating a rare disorder known as Gaucher disease costs around $200,000 a year; a Enbrel for rheumatoid arthritis averages more than $19,000 a year [3]. A more affordable version of such protein therapeutic could result in the possibility of utilization and accessibility to a larger population of patients as well as cost savings to patients, insurance companies, and federal budgets. The first phase of the first generation of modern biotechnology-derived protein drugs is coming to an end with the expiration of their patents (Table 12.1), and an ongoing debate about the scientific and regulatory issues associated with the development of generic (i.e., follow-on) protein therapeutics has emerged and is continuously evolving. The term "generic" protein therapeutics is not a common term used by the scientific community, the US Food and Drug Administration (FDA), or the European Medicines Agency (EMA) for such molecules, and these molecules are typically referred to as biosimilars, reflecting the difficulties in their manufacture,

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*,
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

**TABLE 12.1 Biopharmaceuticals (Patent Expiration, 2009 Sales)**

| Product | Active Substance | Molecular Weight (Da) | Indication | Company | Approval Date | US Patent or Market Exclusivity Expires | EU Patent or Market Exclusivity Expires | Worldwide Sales (US$ Millions, 2011) |
|---|---|---|---|---|---|---|---|---|
| Cerezyme | Imiglucerase | 60,430 | Gaucher disease | Genzyme | May 23, 1994 | Expired | Expired | 716.5 |
| Humulin | Human insulin | ~5807 | Diabetes mellitus | Eli Lilly | October 28, 1982 | Expired | Expired | 1248.8 |
| Novolin | Human insulin | 5808 | Diabetes mellitus | Novo Nordisk | August 30, 1983 | Expired | XX | 1916 |
| Intron-A | Interferon alpha2b | ~19,269 | Chronic hepatitis C | Schering-Plough | October 4, 1983 | Expired | Expired | 194 |
| Avonex | Interferon-β | 22,500 | Multiple sclerosis | Biogen Idec | May 17, 1996 | Expired | XX | 2700 |
| Humatrope | Somatropin | 22,125 | Growth hormone | Eli Lilly | June 23, 1987 | Expired | NA | 400 |
| Nutropin | Somatropin | 22,125 | Growth hormone | Genentech | November 17, 1996 | Expired | Expired | 348 |
| Epogen | Erythropoietin | 30,400 | Anemia | Amgen | June 1, 1989 | 2013 | 2004 | 3728 |
| Neupogen | Filgrastim | 18,800 | Neutropenia | Amgen | February 20, 1991 | 2015 | 2006 | 1442 |

Published on Friday, 06 March 2009 http://www.pipelinereview.com/index.php/Table/FREE-Reports/.

characterization, and comparison with a reference innovator product. In Europe, the EMA has passed guidelines for the abbreviated approval of biosimilars and has even approved several small, simple, and less complex biopharmaceuticals. In the United States, efforts to create regulatory pathways for similar versions of biopharmaceuticals have been fraught with scientific, political, and financial issues. After extensive debate about creating a formal regulatory pathway for biosimilars, the most recent proposal appears to be similar to the 505(b)(2) drug approval process. On March 23, 2010, President Obama signed into law a reconciliation bill that finalizes Congress's health care reform legislation, the Patient Protection and Affordable Care Act. Now after years of public debate, biosimilars and interchangeable biological products are closer to becoming reality for patients in the United States as President Obama's health care reform bill amends the Public Health Service Act (PHSA) to strike a balance between "innovation and consumer interest" because in addition to the much-discussed provisions regarding health insurance, this act creates a regulatory pathway for biosimilars [4]. Passage of the PHSA has resulted in the FDA's considering standards to govern authorization for marketing of biosimilars and interchangeables. The FDA released three draft guidances on implementation of the Biologics Price Competition and Innovation Act in February 2012. These are:

1. Biosimilars: Questions and Answers Regarding Implementation of the Biologics Price Competition and Innovation Act of 2009
2. Quality Considerations in Demonstrating Biosimilarity to a Reference Protein Product
3. Scientific Considerations in Demonstrating Biosimilarity to a Reference Product

Although these guidances create new and different standards, they also raise new barriers to entry in their definitions of data exclusivity and patent litigation procedures.

The objective of this chapter is to understand the scientific challenges and the main elements of the EU pathway and the proposed US FDA regulatory pathway for biosimilars while considering potential strategies in the development of a successful biosimilar.

## 12.2 HISTORY

Over the past several decades, biomedical research has enhanced our understanding of molecular principles of health and diseases and resulted in revealing many polypeptides that possess significant medicinal potential. In the 1970s, the development of two new techniques—recombinant DNA technology (genetic engineering) and monoclonal antibody (mAb) technology (hybridoma technology; Drs César Milstein and Georges Köhler of the MRC Laboratory)—marked the birth of the modern biotech era [5].

Recombinant DNA technology has fourfold impact on production of pharmaceutically important proteins:

1. Overcomes the problem of source availability
2. Overcomes problems of product safety (direct extraction may unwittingly transmit diseases such as hepatitis B or C and HIV)
3. Provides an alternative to direct extraction from inappropriate or dangerous source material (e.g., fertility hormone follicle-stimulating hormone obtained from the urine of postmenopausal women and a related hormone, human chorionic gonadotropin, from urine of pregnant woman, and Ancrod, an anticoacoa extracted from Malaysian pit vipers)
4. Facilitates generation of engineered therapeutic proteins displaying clinical advantage over native proteins (site-directed mutagenesis facilitates predefines changes, e.g., slow- and fast-acting insulins, chimeric or humanized antibodies greatly reduce immunogenicity)

## 12.3 MARKET

The advent of the aforementioned two breakthrough technologies underpinned the establishment of start-up biopharmaceutical companies in the 1980s founded by academic and technical experts and financed by venture capitalists. By mid 2002, about 120 biopharmaceutical products gained marketing approval in the United States or European Union, representing a market share of $15 billion. The sales of biopharmaceuticals reached $20 billion in 2003, and by 2007, worldwide sales reached US $94 billion, accounting for about 15.56% of the pharmaceutical industry's gross revenue. From 2000 to 2006, biopharmaceuticals accounted for 33% of all new active substances that were launched and are reflected to a similar degree in the current development pipelines of the pharmaceutical industry. The sales of the top 10 biologic brands increased significantly in 2011. The market leader was Humira with sales of $7.9 billion followed by Enbrel and Remicade. Drug products such as epoetin-α (Epogen, Eprex, Procrit), abciximab (ReoPro), interferon-α (Intron A, Roferon A), interferon-β (Avonex, Rebif, Betaseron), anti–tumor necrosis factor α agents (Enbrel, Remicade, Humira), and trastuzumab (Herceptin) are all examples of highly successful biotech drugs that revolutionized the pharmacotherapy of previously unmet medical needs (see Table 12.1 for first-in class biopharmaceuticals) [6].

EuropaBio (www.europabio.org) reports that by 2007, more than 325 million patients had benefited from approved medicines manufactured through biotechnology and gene technology.

Prevenar was the first vaccine to have sales of $2.7 billion in 2008 and revived the current interest in vaccines. The sales of the biopharmaceutical brands continued to increase significantly through 2009 (Table 12.2). The top 30 biologics sales exceeded $24 billion in 2011 [7]. Among the various segments, vaccines and mAb-based products represent the two biggest product categories followed by regulatory factors

**TABLE 12.2 Top Five Biopharmaceuticals by US Sales (2011)**

| Product | Active Substance | Molecular Weight (Da) | Indications | Company | Approval Date | Worldwide Sales, mm (USD, 2009) |
|---|---|---|---|---|---|---|
| Humira | Adalimumab | ~148,000 | Rheumatoid arthritis, psoriasis, Crohn's disease, ulcerative colitis, psoriatic arthritis | Abbott and Eisai | December 31, 2002 | 7932 |
| Enbrel | Etanercept dimeric fusion protein | ~150,000 | Rheumatoid arthritis, psoriatic arthritis plaque psoriasis | Immunex* | November 2, 1998 | 7867 |
| Remicade | TNF-α antagonist | ~149,100 | Rheumatoid arthritis | Centocor (J&J) | August 24, 1998 | 7157 |
| Rituxan | Rituximab | | Non–Hodgkin's lymphoma | Roche | November 26, 1997 | 6859 |
| Avastin | VEGF | | Metastatic colorectal cancer | Roche | February 26, 2004 | 5813 |

* Purchased by Amgen.
TNF: tumor necrosis factor; VEGF: vascular endothelial growth factor.
Adapted from http://www.pipelinereview.com.

(e.g., hormones and cytokines) and gene therapy and antisense-based products. mAbs began to reveal their full potential in 1986, when Medical Research Council researcher Dr. Gregory Winter pioneered a technique to humanize mouse mAbs [6] because they were much less likely to elicit an inappropriate immune response in patients. Dr. Gregory founded the Cambridge Antibody Technology, which licensed the technique to more than 50 companies worldwide. Subsequently, Humira became the first fully human mAb drug launched in 2003 as a treatment for rheumatoid arthritis. Cambridge Antibody Technology was subsequently acquired by Astra Zeneca in 2006 for £702 million [8].

Some interesting timelines in the mAb landscape include the following:

1975: Method devised to isolate and reproduce mAb
1986: Techniques pioneered to humanize mouse mAb
1990: Test tube production of highly specific human mAb
2003: First fully human mAb, Humira, launched in the United Kingdom
2005: Humira sales reach more than $200 million

Consensus sales forecasts for world's top 10 drugs in 2014 include eight that are of biopharmaceutical origin (Table 12.3).

Opportunities for the biosimilars market are immense because biologics worth $25 billion are going to be off patent by 2016, and the recent approval pathway in the United States will perhaps open up to what is expected to become the largest market [9]. The global biosimilars market is expected to be worth $19.4 billion by 2014,

**TABLE 12.3 Consensus World Sales Forecasts for the Top 10 Drugs in 2014**

| Name (Indication) | Company | Worldwide Sales (US$, billions) |
|---|---|---|
| Avastin* (cancer) | Genentech† | 8.9 |
| Humira* (arthritis) | BASF Pharma‡ | 8.5 |
| Enbrel (arthritis) | Immunex§ | 8.0 |
| Crestor (cholesterol) | AstraZeneca | 7.7 |
| Remicade* (arthritis) | Merck | 7.6 |
| Rituxan* (cancer) | Genentech* | 7.4 |
| Lantus (diabetes) | Sanofi-Aventis | 7.1 |
| Advair (asthma, COPD) | GlaxoSmithKline | 6.8 |
| Herceptin* (cancer) | Genentech† | 6.4 |
| NovoLog (diabetes) | Novo Nordisk | 5.7 |

*Monoclonal antibody.
†A member of the Roche Group.
‡Acquired by Abbott in 2001.
§Acquired by Amgen in 2002.
COPD: chronic obstructive pulmonary disease.
Adapted from Source: Thomson Reuters (Reporting by Ben Hirschler, Editing by Sitaraman Shankar). Available from: http://www.reuters.com/article/idUSLDE63C0BC20100413.

growing at an expected average annual growth rate of about 89% from 2009 to 2014 [9]. The biosimilar market is also similarly segmented into peptides, recombinant glycosylated proteins, recombinant nonglycosylated proteins, and others. Recombinant nonglycosylated proteins form the largest segment in the biosimilars market, with an estimated $11.5 billion in sales by 2014 [9].

Several biosimilars have already been approved in the European Union under EMA biosimilars regulations and guidelines, including human growth hormone (hGH) Omnitrope and Valtropin and erythropoietins Binocrit, epoetin alfa HEXAL, Abseamed, Silapo, and Retacrit. In the United States, several follow-on versions of protein therapeutics originally approved under the FDC Act have gained approval under Section 505(b)(2), including Omnitrope. However, until March 2010, the approvals in the United States were limited to the original products approved under the FDC Act, not the majority of protein therapeutics approved under the PHSA.

## 12.4 WHY ARE BIOPHARMACEUTICALS DIFFERENT FROM SMALL MOLECULES?

Traditional pharmaceuticals are low-molecular-weight chemicals and are typically manufactured by direct synthesis. Biopharmaceuticals, in contrast, are typically large, complex, heterogeneous, protein-based molecules produced by a living cell–based system that cannot be fully characterized by analytical methods in the laboratory [10]. The differences in production, molecular size, and complexity between chemical medicines and biopharmaceuticals pose the question of whether or not the established legal and regulatory standards for approval of generic medicines, whose very basis is one of sameness or identicality, could be applied to attempted copies of innovator biopharmaceuticals (Table 12.4).

Following are some important considerations for biopharmaceutical products [11]:

- Patient safety: Because of differences between biopharmaceuticals, challenging issues exist relating to the development, approval, and marketing of biosimilar products. Patient safety must be of paramount consideration when evaluating these issues.
- Need for clinical trials and indication-specific approval: Because of the complexity of biopharmaceuticals as well as the distinctions introduced by differing manufacturing processes, each biosimilar must be shown to be safe and effective on the basis of its own adequate and well-controlled clinical studies.
- No automatic substitution or interchangeability: Differing manufacturing processes mean that each biosimilar is likely to be inherently different from the corresponding innovator biologic. As such, pharmacists should not be allowed to freely substitute the follow-on biologic for the innovator's product.
- Postmarketing risk management plans required: Postmarketing risk management plans must be an essential element of any approval of a biosimilar product.

**TABLE 12.4 Differences Between Small Molecular Weight Drugs and Biopharmaceuticals**

| Small-Molecular-Weight Drugs | Biopharmaceuticals |
|---|---|
| Low molecular weight (<5000 Da) | High molecular weight (>5000 Da)<br>Typically they are 100- to 1000-fold larger molecules |
| Simple chemical structure | Complex three-dimensional structure |
| Bioactivity: defined | Difficult to quantify bioactivity |
| Chemically synthesized: predictable and well defined | Produced by living organism; affected by minor changes<br>*Inherent product variability*<br>Carboxylation<br>O-linked glycosylation<br>N-linked glycosylation<br>Oxidation<br>Amino acid substitution<br>Truncation<br>Aggregation<br>Denaturation<br>Amidation, deamidation, and so on |
| Easy characterization: simple definitive details or standards | Difficult or impossible to fully characterize |
| Synthetically pure | Often heterogeneous |
| Relatively stable | Stability: proteins are inherently unstable molecules and may structurally be damaged by heat, prolonged storage, denaturants, organic solvents, oxygen, pH changes, and other factors, leading to reduction or complete loss of biological activity |
| Impurities: defined standards or specifications | Difficult to define and quantify (viral, bacterial, fungal, TSE) |
| Rarely produce immune response | Prone to eliciting immune response |

TSE: transmissible spongiform encephalopathies.

- Similarity and comparability are distinct concepts: The concepts of similarity and comparability are distinct and must be properly applied to any approval process of a biosimilar product.
- No reliance on trade secret or confidential commercial information: Regulatory agencies should not refer to or rely on innovator trade secrets and confidential commercial information to approve a biosimilar product.

## 12.5 TYPICAL BIOPHARMACEUTICAL MANUFACTURING STEPS AND ANALYTICAL TECHNIQUES FOR CHARACTERIZATION

Almost all biopharmaceuticals on the market are produced by genetic engineering using various recombinant expression systems. Proteins are made by living cells such as bacteria, yeast, plant, or mammalian cells acting as the "manufacturing facility."

The manufacturing of recombinant protein products includes cloning, transfecting, screening for the cells, subcloning and developing this cell further concerning expression yield, growth properties, and so on into a master and working cell bank; growing the recombinant cell in large bioreactor vessels (up to and even exceeding 10,000-l scale) depending on the supply needs; purifying the target protein using a multistep downstreaming process; and finally bringing it into a formulation and device suitable for transport, storage, and application to the patients.

The process runs under strictly controlled, validated conditions in closed systems to assure consistency to avoid any contamination in accordance with good manufacturing process (GMP) requirements. The manufacturing process of protein products has a significant impact on the quality of the final product. Figure 12.1 illustrates a typical platform process used for manufacturing of mAb products [12]. Any variability in the quality of the final product depends on the variability of each process step and variability of the starting materials. Media components used in cell culture processes are known to have an impact on the performance of the production bioreactor. Various examples demonstrate that changes in cell lines, expression systems, bioreactor operating conditions, purification process, or even the use of different syringes and vials can lead to changes in quality attributes (Table 12.5).

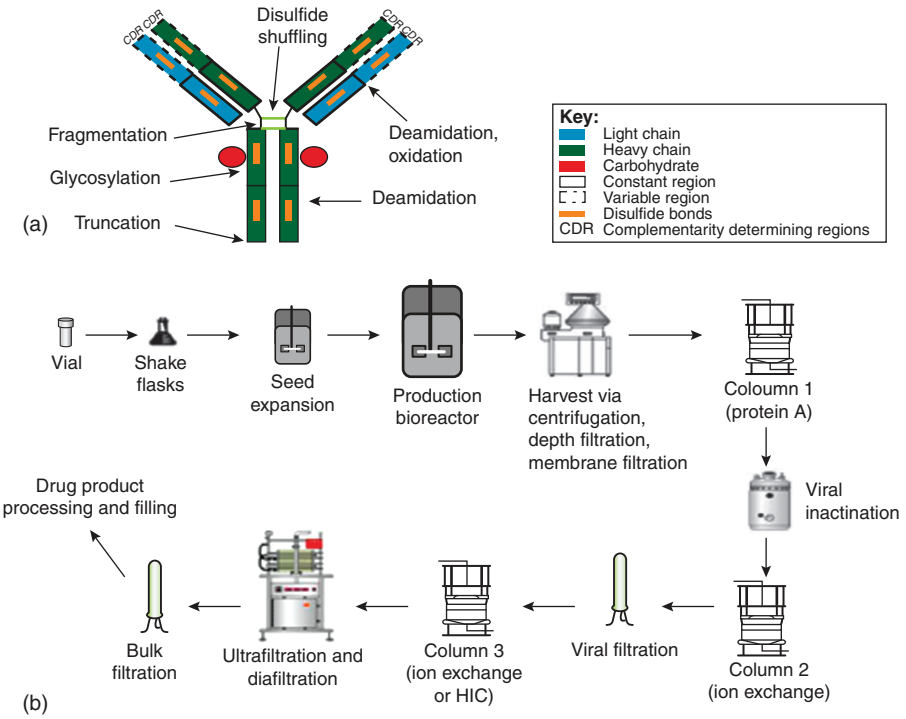


**FIGURE 12.1** Illustration of complexity of protein product and processes involved. (a) Structure of a typical monoclonal antibody (mAb) along with sites of heterogeneity. (b) Typical platform process for an mAb. HIC: hydrophobic interaction chromatography.

**TABLE 12.5 Typical Product Quality Attributes for a Biological Product***

| Class | Sample Quality Attributes (Commonly Used Tests) |
|---|---|
| Identity | Primary sequence (peptide map and amino acid sequence analysis), immunogenicity (immunoassay), other indications (IE HPLC, gel electrophoresis) |
| Potency | Potency (cell-based bioassay, gene expression bioassay, ADCC, CDC) |
| General characteristics | Appearance or color (visual), clarity (turbidimetric), subvisible particles, pH (potentiometric), osmolality (freezing point depression), conformation (near or far UV, circular dichroism, spectroscopy, Fourier transform infrared spectroscopy, X-ray crystallography and differential scanning calorimetry), effector function (receptor binding), concentration (UV absorbance) |
| Desired host cell modifications | Glycosylation (monosaccharide composition analysis, oligosaccharide profile, CE, LC—MS, MS/MS, ESI, MALDI-TOF), phosphorylation (peptide mapping with MS) |
| Undesirable host cell modifications | Truncation (SE HPLC, gel electrophoresis, analytical ultracentrifugation (AUC), peptide mapping with MS, RP HPLC), glycation (peptide mapping with MS HPLC), methylation, isomerization (RP HPLC) |
| Desired process modifications | PEGylation (HPLC, CE) |
| Undesirable process modifications (product variants and impurities) | Aggregation (SE-HPLC, gel electrophoresis, light scattering, and AUC), oxidation (peptide mapping with MS), deamidation (cIEF), peptide mapping with MS and cation exchange (CEX-HPLC), C terminal lysine (cIEF, peptide mapping with MS, and CEX-HPLC), C-terminal lysine (cIEF, peptide mapping with MS, and CEX-HPLC) misfolds (RP-HPLC) |
| Host cell–related impurities | Host cell proteins (ELISA), DNA, endotoxin (Limulus Amebocyte Lysate assay) |
| Process-related impurities | Antifoam (compendial), extractables (miscellaneous), leachables (miscellaneous), bioburden (membrane filtration test), pluronic (compendial), protein A (protein A ELISA) |

*This list is nonexhaustive.
ADCC: antibody-dependent cell-mediated cytotoxicity; AUC: analytical ultracentrifugation; CDC: complement dependent cytotoxicity; CE: capillary electrophoresis, CEX: cation exchange; cIEF: capillary isoelectric focusing; ELISA: enzyme-linked immunosorbent assay; ESI: electrospray ionization; HPLC: high-performance liquid chromatography; IE: ion exchange; IEF: isoelectric focusing; LC: liquid chromatography; MALDI: matrix-assisted laser desorption ionization; MS: mass spectrometry; RP: reversed phase; SE: size exclusion; TOF: time of flight; UV: ultraviolet.

Potency can be affected by physicochemical or structural changes. A 20% reduction in sialic acid of a glycoprotein can result in 50% loss in amount of circulating protein [13]. For some mAbs, monitoring carbohydrate heterogeneity throughout product manufacturing and comparability assessment is critical because differentiation sialylation, galactosylation, and fucosylation can affect clinical function, including complement-dependent cytotoxicity, antibody-dependent cellular cytotoxicity (ADCC), and antigen binding. Generally, a decrease in core fucose content increases ADCC. However, variations in galactosylation do not affect this activity. The role of glycosylation in mAb efficacy is product dependent; therefore, glycosylation analysis is dictated by the product. Schiestl et al. [13] reported change of the isoform distribution between two sets of batches of Aranesp expiring between November 2008 and April 2011 (Fig. 12.2). They reported that a higher sialylated isoform decreased by an average of 10%, and the less sialylated isoform increased by 3%. However, within each set there was high batch-to-batch consistency. They attributed this change as a consequence of manufacturing process change. Aranesp is an engineered analog of human erythropoietin and differs in amino acid sequence that introduces two additional N-glycosylation sites. The *in vivo* biological activity is known to be dependent on number of sialic acid units per molecule (a result of available sites, completeness of sialylation, and antennarity of the N-glycans).

In the same article, Martin Schiestl et al. [13] also characterized commercial batches of Rituxan with expiration dates from September 2007 to October 2011 using glycan mapping, cation exchange chromatography (CEX), and ADCC *in vitro* bioactivity (Fig. 12.3). For Rituxan, the glycosylation-dependent Fc effector functions are essential contributors to its therapeutic mode of action and clinical efficacy. CEX revealed differences in C-terminal lysine and N-terminal glutamine variants between the 2008 and 2010 batches. The basic variants reduced from 30 to 50% to 10%. The glycan map revealed a threefold increase in unfucosylated G0 glycan, and the corresponding ADCC activity also increased.

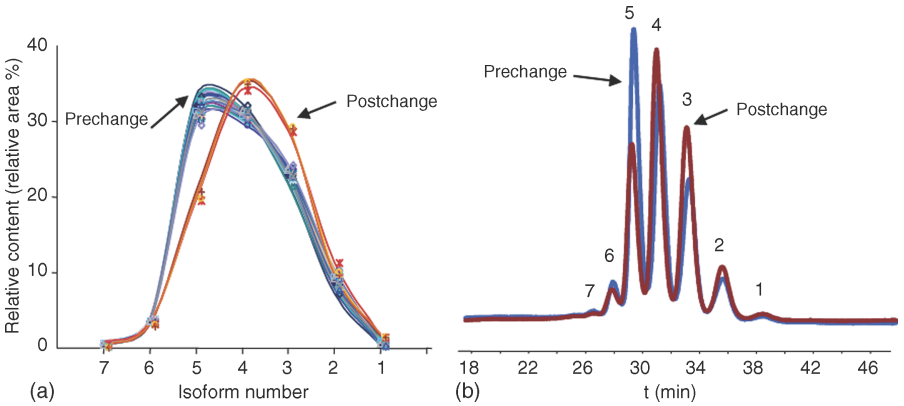


**FIGURE 12.2** Comparison of pre- and postchange Aranesp batches measured by capillary zone electrophoresis. (a) Relative content of the individual isoforms of the prechange ($n = 18$) and postchange batches ($n = 4$) batches. (b) Representative electropherograms; peaks are labeled with isoform number.

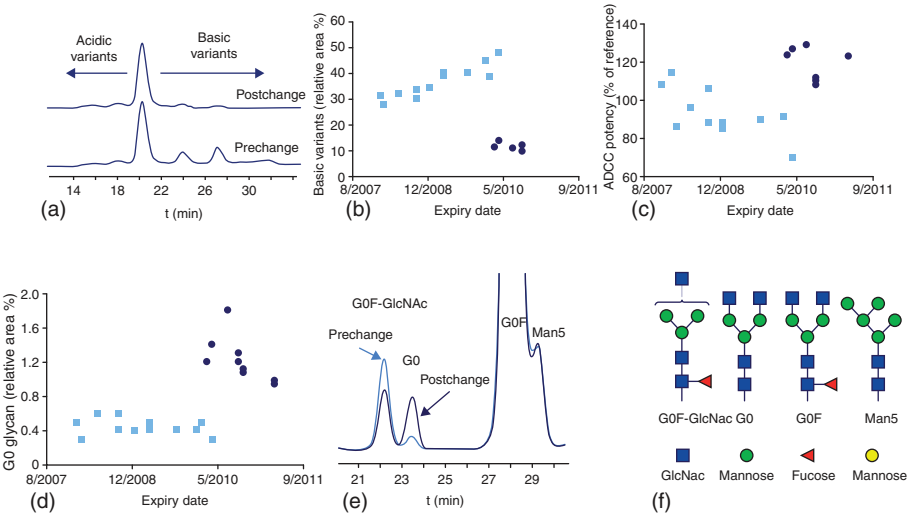


**FIGURE 12.3** The glycan map revealing a threefold increase in unfucosylated G0 glycan and the corresponding antibody-dependent cellular cytotoxicity (ADCC) activity increase. (a) Acidic vs. basic variants (pre- and postprocess change). (b) Relative % area of basic variants (multiple lots). (c) ADCC potency vs. various lots manufactured pre- and post-manufacturing changes. (d) Relative % area of G0 glycan vs. multiple lots. (e) Increase in unfucosylated G0 glycans attributed to manufacturing process changes. (f) Various glycan isoforms. G = glycan isoforms; GlcNAc = *N*-acetylglucosamine; Man = mannose; Fuc = F = fucose; G0 = $GlcNAc_2Man_3GlcNAc_2$; G0F = $GlcNAc_2Man_3GlcNAc_2Fuc$; Man5 = $Man_5GlcNAc_2$; G2F = $Gal_2GlcNAc_2Man_3GlcNAc_2Fuc$; (1,3)G1F or (1,6)G1F = variants of $GalGlcNAc_2Man_3GlcNAc_2Fuc$.

The same authors [13] also reported that for two batches of Enbrel (expiration dates of 2009 and 2011, respectively) (Fig. 12.4), glycan mapping revealed a decrease from 50 to 30% in the N-glycans G2F, but CEX analysis reported an increase to 40 to 60% from 15 to 30% in the C-terminal lysine variants attributable to manufacturing process changes. Figure 12.4 suggests that batch-to-batch variability is small during routine manufacturing, major shifts in product attributes may occur as a result of manufacturing changes, the variance induced by manufacturing changes establishes "goal posts" that a biosimilar company can used for establishing comparability, and it is apparent that originators are essentially manufacturing biosimilars of their original products [13].

From a regulatory perspective, it is expected that a company will provide some comparability data after any change in manufacturing likely to affect product quality during the development process. Before pivotal trials, these data are required to show that the postchange material is similar enough to allow safety data based on earlier material to be applied to the new material.

The consequence of inadequate comparability data (or illustrating significant changes) is the possibility of having to repeat toxicology and safety studies before continuing the clinical development plan. Changes made during or after the pivotal trial require comparability to ensure both safety and efficacy of the postchange

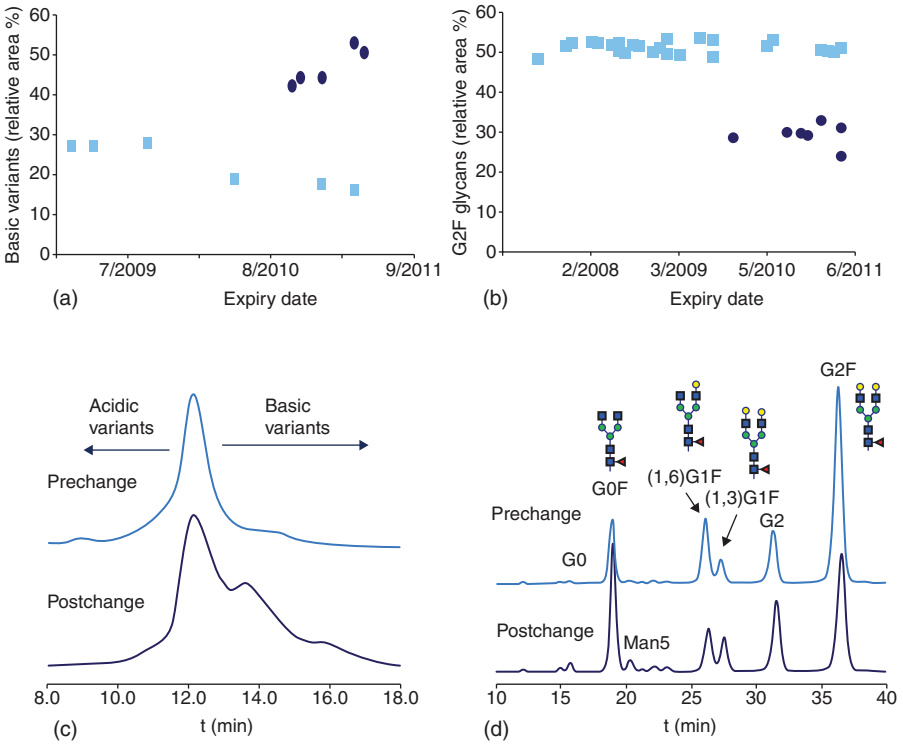


**FIGURE 12.4** Glycan mapping revealing a decrease from 50 to 30% in the N-glycans G2F; cation exchange chromatography analysis reported an increase to 40 to 60% from 15 to 30% in the C-terminal lysine variants.

material. This leads to either significantly extended pivotal trials or new trials altogether. The extent of the comparability exercise is dictated by not only the timeline in product development when the manufacturing change occurs but also by the nature of the change and the nature of the product. The regulatory importance of appropriate assessment of comparability has resulted in the development of the International Committee for Harmonization (ICH) guideline Q5E, "Comparability of Biotechnological/Biological Products Subject to Changes in Their Manufacturing Process," and the EMA has published two guidelines, titled "Guideline on Comparability of Medicinal Products Containing Biotechnology-Derived Proteins as Drug Substance—Non clinical and Clinical" and "Guideline Comparability of Medicinal Products Containing Biotechnology-Derived Proteins as Active Substance—Quality Issues."

Typically, 20 to 30 unit operations are performed during the manufacture of a protein product; therefore, significant variability in product quality is likely by the time the final stages are reached. Minor changes in the process could lead to major changes in the product. It was often stated that for biopharmaceuticals, "the process is the product" (Table 12.6). Even highly purified proteins never consist of one single molecular entity but are mixtures of many closely related molecular species. It has

**TABLE 12.6 Examples Highlighting the Impact of Processes on Product Quality and Clinical Performance**

*Changes in Manufacturing Process*

- The choice of cell line or expression system affects host-cell impurities, which may act as adjuvants or be directly immunogenic.
- Protein glycosylation may be affected by concentrations of glucose, ammonium ions and hormones, cell culture conditions, and hydrodynamic shear.
- mAb product produced by B-cell hybridoma vs. CHO cell transfectoma has been shown to possess varying levels of C-terminal lysine.
- Medium components manufactured by two different suppliers have been shown to lead to differences in product quality.
- Trace amount of contaminants or impurities have been shown to lead to antibody development against insulin and growth hormone products.
- Purification products have been known to have an impact on the incidence of antisomatotropin molecules (Omnitrope vs. Genotrope).

*Leachables and Extractables on Product Safety and Efficacy*

- Uncoated plunger stoppers from syringes leached vulcanizing agent are speculated to cause increased PRCA with Eprex.
- Residual amounts of tungsten used to create syringe have been known to cause aggregation of two protein therapeutic products.
- Glass surfaces, air–liquid interfaces, and lubricants can mediate protein denaturization; phthalates in plastics and latex rubber may contaminate product and cause allergic reactions and increase immunogenicity.

CHO: Chinese hamster ovary; mAB: monoclonal antibody; PRCA: pure blood red cell aplasia.

been estimated that in an immunoglobulin G (Ig G) molecule, there might theoretically be up to 108 different species considering only a limited number of all possible glycosylation variants [12].

Hence, a second manufacturer aiming to replicate a protein product has to run through an analogous procedure without access to the originator's materials (e.g., cell banks) and information (standard operating procedures). For glycosylated proteins, differences in the glycopattern may significantly influence receptor binding, protein–protein interaction, and pharmacokinetics (PKs) of protein substances. For immunoglobulins, small differences in core fucosylation can lead to big changes in Fcc receptor binding and consequently impact immune effector functions such as ADCC, which is believed to be a major mechanism of action contributing to the potency of many mAbs, particularly in oncologic indications. Differences in protein product characteristics may not be detectable by commonly used analytical characterization (Table 12.7).

## 12.6 GENERICS VERSUS BIOSIMILARS

The debate essentially has been about whether technology exists to create and validate a copy of a product made through complex biological processes; hence, neither the scientific community nor the FDA is comfortable designating a

**TABLE 12.7 Analytical Methods and Their Utility in Characterizing Drug Substance and Drug Product (Partial List)**

| Parameter | Methodology | Utility |
|---|---|---|
| Appearance | Visual, turbidity | Physical stability; container-closure integrity |
| Purity-degradation products and related substances | GPC/SEC-HPLC, RP-HPLC, gel electrophoresis, immunoassays | Relative purity of API and drug product; impurities related to expression system |
| Molecular weight and size | GPC/SEC-HPLC, gel electrophoresis, LC—ESI-MS, MALDI-TOF-MS | Product single molecular weight or polydisperse? |
| pH | Potentiometric assays | Is the product pH labile, or do pH changes affect potency? |
| Primary structure | Amino acid sequencing, N-term degradation, Edman sequencing, peptide mapping | Verifies primary amino acid sequence |
| Secondary structure | CD, 2D-NMR | Help determine overall 3D folding |
| Tertiary structure | Disulfide content or position, CD | Correct folding and integrity of 3D product |
| Agglomeration and aggregation | Subvisual and visual particle size analysis, immunogenicity | Physical stability and immunogenic potential |
| Carbohydrate analysis | RP-HPLC, HPLC—MS, gel electrophoresis, MALDI-MS, ESI-MS | Ensures posttranslational modification and carbohydrate content |
| Water content (lyophilized product) | Karl Fischer, NIR | Indicator of hydrolytic potential |
| Immunogenic potential | Surface plasmon response, ELISA, immunoprecipitation | Positive or negative indicators of immunogenic potential |
| Sterility | Membrane filtration | Microbial contamination from manufacturing operations |
| Bacterial endotoxins | LAL | Processing contaminants and host organism contaminants |

AE: anion exchange; CD: circular dichroism; CE: capillary electrophoresis; ELISA: enzyme-linked immunosorbent assay; ESI: electrospray ionization; GPC: gel permeation chromatography; HPLC: high-performance liquid chromatography; LAL: Limulus amebocyte lysate; MALDI: matrix-assisted laser desorption ionization; MS: mass spectrometry; NIR: near infrared; NMR: nuclear magnetic resonance; RP: reversed phase; SEC: size exclusion chromatography; TGA: thermogravimetric analysis; 2D: two-dimensional; 3D: three-dimensional.

biogeneric as a substitutable replacement for the innovator product. Hence, these molecules are typically referred to as biosimilars, reflecting the difficulties in their manufacture, characterization, and comparison with a reference innovator product. Compared with the manufacture of small molecule entities, the manufacture of biopharmaceuticals requires a greater number of batch records (>250 vs. <10), a large number of product quality tests (>2000 vs. <100), an increased amount of

critical process steps (>5000 vs. <100), and significantly larger process data entries (>60,000 vs. <4000) [11]. The molecular weight of a traditional small molecule drug can be measured accurately within 1/100 Da. In contrast, the molecular weight of biopharmaceuticals can vary by as much as 1 kDa because of the heterogeneity of the production processes and of products originating from the same process. This complexity of protein products is central to the challenges associated with defining their safety and efficacy.

Similarity is a natural extension of comparability, and scientific exercises need to be conducted to demonstrate that a patient is receiving a biological drug that is comparable to the same drug produced before a change in the manufacturing process. Changes to manufacturing process can range from the relatively simple procedure of changing the cellular harvesting method (in which recovery is not affected) to more complex procedures, such as the replacement of a stabilizer or introduction of a new cell line. A comparability exercise is a stepwise approach to development using side-by-side comparisons of the biological product before and after manufacturing changes. This exercise begins with extensive analytical and biological characterization. Animal or human studies are then conducted, if needed, to answer scientific questions still remaining after analytical and biological characterization and comparison. If "no differences" exist between these products, the new product can be substituted for the old. A similar approach is deemed necessary for biosimilars. Key factors that influence the approval process include evidence of integrity and consistency of the manufacturing process; conformance of manufacturing standards to applicable regulations; demonstration of product consistency with appropriate reference standards or comparators (using relevant assays), including comparative PK and pharmacodynamic (PD) data; and the extent of clinical data and experience with the approved product. Immunogenicity assessment has always been a key concern related to approval of follow-on protein products. Although significant advances have been made in predicting immunogenicity using *in silico, in vitro*, and *in vivo* tools, human clinical trials continue to remain the centerpiece of immunogenicity assessment.

The International Nonproprietary Names (INN) program directed through the World Health Organization has the purpose of assigning nonproprietary names to medicinal substances. Assigning INNs to protein therapeutics is often a difficult process because of their complexity. For example, when products have significant differences in glycosylation, such as with the epoetin products, different Greek letters are attached to their INNs. The idea of a biosimilar is based on regulations, but the system for assigning INNs needs to be based on scientific characterization of the biosimilar and comparison with a reference product. The focus of this discussion also is the issue of separate INNs for the reference product and the biosimilar. Justification, for example, for separate INNs has been to ensure adequate postapproval pharmacovigilance and to prevent inappropriate substitution of these products [14]. Regulatory agencies in both the United States and Europe have realized that a pathway for marketing approval of biosimilars requires a different structure and guidelines compared with conventional generics to ensure quality, safety, and efficacy. One of the foundational principles of the European framework is that

biosimilars are expected to be similar, not identical, to the innovator biologics they seek to copy [15].

## 12.7 CURRENT REGULATORY STATUS AND PATHWAYS TO BIOSIMILAR PRODUCTS

### 12.7.1 Europe

***12.7.1.1 Approval Process for Generic Synthetic Small Molecules*** In the European Union, rising health care costs had a profound influence on the growth of the generics industry. Currently, there are several marketing authorization procedures in the EU: a centralized procedure, a mutual recognition procedure (MRP), and a decentralized procedure (along with the individual national registration process). Generic manufacturers of centrally approved reference products have the option of applying through centralized or decentralized procedure. Generic applications normally are submitted through the MRP, which enables national assessment done on part of the market authorization in one reference member state (RMS) to be mutually recognized by other member states, that is, concerned member states (CMS). The legal text of the MRP is provided in Directive 2001/83/EC as amended in by Directive 2004/27/EC, and further guidance is given in the Notice to Applicants, which forms Chapter 2 of the Rules Governing Medicinal Products in the European Union. The Directive 2001/83/EC as amended gives the legislative basis for bioequivalence. Although some similarities to the generic pathway for biosimilars have indeed been built in with respect to small molecules, the one main factor discussed earlier that distinguishes biosimilars from generic drugs is their capacity to elicit an immune response [16].

***12.7.1.2 Approval Process for Biosimilars in Europe*** The regulation of biosimilar drugs was a contentious issue in the EMA. The EMA had been involved in preparing a regulatory framework for biosimilars since 1998 (Concept Paper on the Development of a CHMP Guideline on Comparability of Biotechnology-derived Products [CHMP/BWP/1113/98]). Subsequently, the Committee for Medicinal Products for Human Use (CHMP) issued a Guideline on Similar Biological Medicinal Products (EMEA/CHMP/437/04) [17]. The biological medicinal products in Europe were established with Directive 2001/83/EC, which provided a clear foundation for EMA for future approval process for biosimilars. Specific guidelines were introduced as amendment to Article 10 (Article 10[4] and section 4, Part II, Annex I to the Directive 2001/83/EC known as "Annex 1"). Annex I (Part 2, Paragraph 4) became incorporated as part of the Commission Directive 2003/63/EC, which approved the concepts of biosimilar and gave the EMA mandate to develop its own guidelines on preclinical and clinical data. The EU Directive 2003/63/EC, Annex Part II, Point 4 for "Similar Biological Medicinal Products" states that "the information to be supplied shall not be limited to pharmaceutical, chemical and biological data, supplemented with bioequivalence and bio-availability data." In fact,

the type and amount of additional data (i.e., toxicologic and other nonclinical and appropriate clinical data) shall be determined on a case-by-case basis in accordance with relevant scientific guidelines published by EMA. Directive 2004/27/EC details negotiations between the Council of the European Union and European Parliament on amendments to Directive 2001/83/EC, including the "Bolar Clause," which allowed full development of biosimilars before the originator's patent expired. The European Directive 2004/27/EC addresses the biosimilar medicines subject as follows: when a biological medicinal product that is similar to a reference biological product does not meet the conditions in the definition of generic medicinal products, owing to, in particular, differences relating to raw materials or differences in manufacturing processes of the biological medicinal product and the reference biological medicinal product, the results of appropriate preclinical tests or clinical trials relating to these conditions must be provided. The type and quantity of supplementary data to be provided must comply with the relevant criteria stated in Annex I and the related detailed guidelines. In 2005, the EMA CHMP issued an overarching biosimilars guideline, "Guideline on Similar Biological Medicinal Products," (EMEA/CHMP/437/04) and two guidelines for consultation on the quality issues (EMEA/CPMP/BWP/3207/00 rev 1) and on the nonclinical and clinical issues (CHMP/3079/02 and EMEA/CHMP/BMWP/101695/2006) to be considered for the development of biosimilars.

As stated in Section 4, Part II, Annex I to Directive 2001/83/EC, the EMA may develop additional guidance documents addressing quality and nonclinical and clinical aspects. These include:

- Guideline on Comparability of Medicinal Products Containing Biotechnology-Derived Proteins as Active Substance—Quality Issues (EMEA/CPMP/BWP/49348/05)
- Guideline on Comparability of Medicinal Products Containing Biotechnology-Derived Proteins as Active Substance—Clinical and Non-clinical Issues (EMEA/CHMP/BMWP/42832/05)

The key EMA guidelines are listed in Table 12.8.

Essentially, Directive 2001/83/EC and updates 2003/63/EC and 2004/27/EC lay out a general pathway for approval of biosimilars. According to these directives, comparability studies will be required to establish the similarity between the biosimilar and the reference protein therapeutic (which must already have marketing authorization in the European Union) in terms of quality, safety, and efficacy. According to Article 57 (1)(n) of the Regulation (EC) No. 726/2004, the EMA shall provide acting particularly through its committees and advice on the conduct of the various tests and trials necessary to demonstrate the quality, safety, and efficacy of medicinal products. For most biosimilars, the comparability exercises proceed in a stepwise fashion (Fig. 12.5) [18]. In the first step, a comparison of the quality of the test and the reference protein therapeutic is made. Any differences or concerns in the quality between the test and the reference product will affect the details and amount of comparative nonclinical and

**TABLE 12.8 Current EU Biosimilar Guidances**

| Reference Number | Guideline | Scope | Release Date | Effective Date |
|---|---|---|---|---|
| CPMP/BWP/3207/00 Rev.1 | Guideline on Comparability of Medicinal Products Containing Biotechnology-Derived Proteins as Active Substance—Quality Issues | All biosimilars | December 2003 | December 2003 |
| CHMP/437/04 | Guideline on Similar Biological Medicinal Products | All biosimilars | September 2005 | October 2005 |
| EMEA/CHMP/BWP/49348/05 | Guideline on Similar Biological Medicinal Products Containing Biotechnology-Derived Proteins as Active Substance—Quality Issues | All biosimilars | February 2006 | June 2006 |
| EMEA/CHMP/42832/05 | Guideline on Similar Biological Medicinal Products Containing Biotechnology-Derived Proteins as Active Substance—Nonclinical and Clinical Issues | All biosimilars | February 2006 | June 2006 |
| EMEA/CHMP/BMWP/101695/06 | Guideline on Comparability of Biotechnology-Derived medicinal Products After a Change in the Manufacturing Process—Nonclinical and Clinical Issues | All biosimilars | July 2007 | November 2007 |
| EMEA/CHMP/BMWP/14327/06 | Guideline on Immunogenicity Assessment of Biotechnology-Derived Therapeutic Proteins | All biosimilars | January 2008 | April 2008 |

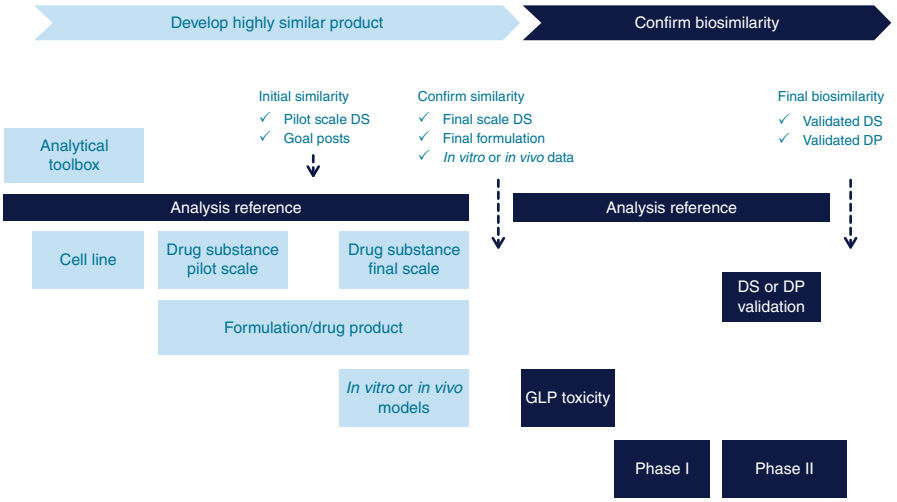


**FIGURE 12.5** An overview of Sandoz biosimilar development approach. DP: drug product; DS: drug substance; GLP: good laboratory practices.

**TABLE 12.9 EU Biosimilar Approvals (as of February 2009)**

| Biosimilar Product | Company | International Nonproprietary Name | Reference Product | Date of Approval |
|---|---|---|---|---|
| Omnitrope | Sandoz | Somatropin | Genotropin | April 12, 2006 |
| Valtropin | Biopartners | Somatropin | Humatrope | April 24, 2006 |
| Binocrit | Sandoz | Epoetin alfa | Erypo/Eprex | August 28, 2007 |
| Epoetin alfa Hexal | Hexal | Epoetin alfa | Erypo/Eprex | August 28, 2007 |
| Abseamed | Medice | Epoetin alfa | Erypo/Eprex | December 18, 2007 |
| Silapo | Stada | Epoetin alfa | Erypo/Eprex | December 18, 2007 |
| Retacrit | Hospira | Epoetin alfa | Erypo/Eprex | September 15, 2008 |
| Ratiograstim | Ratiopharm | Filgrastim | Neupogen | September 15, 2008 |
| Tevagrastim | Teva | Filgrastim | Neupogen | September 15, 2008 |
| Biograstim | CT Arzneimittel | Filgrastim | Neupogen | September 15, 2008 |
| Filgrastim ratiopharm | Ratiopharm | Filgrastim | Neupogen | September 15, 2008 |
| Zarzio | Sandoz | Filgrastim | Neupogen | February 6, 2009 |
| Filgrastim Hexal | Hexal | Filgrastim | Neupogen | February 6, 2009 |

clinical bridging studies that will be required in further steps. Clinical comparability studies usually consist of PK and PD studies in healthy volunteers followed by comparative safety and efficacy trials in patients with the most sensitive and relevant indication. (Specific target study populations are recommended in the EMA guidelines for each of the current categories of biosimilars.)

The EMA CHMP has released for consultation several product-class specific guidelines for the development of several biosimilars, including recombinant human granulocyte; colony-stimulating factor; recombinant human insulin; somatropin; erythropoietin; and, most recently, low-molecular-weight heparins, which have been helpful in several approvals to date (Table 12.9). The product class-specific guidelines stop short of providing precise equivalence margins and do not prescribe a fixed standard for approval. There is still considerable room to maneuver when it comes to negotiating precise standards for approval of any one biosimilar.

Biosimilars that cannot currently be fully characterized because of their complexity (structural, manufacturing) or lack acceptable analytical procedures and technology or clinical and regulatory experience may not be acceptable using the "similar biological medicinal product" approach. In 2010, the EMA released draft versions of two guidelines relating to the regulatory approval of mAbs: the "Guideline on Biological Medicinal Products Containing Monoclonal Antibodies" and the "Guideline on Immunogenicity Assessment of Monoclonal Antibodies Intended for *In Vivo* Clinical Use." Both guidelines were subject to public consultation until May 31, 2011. The former regulation for the approval of

biosimilar mAbs was long awaited by the pharmaceutical industry because many basic patents for therapeutic mAbs will expire within the next couple of years. The present guideline sets forth the nonclinical and clinical requirements for mAb-containing medicinal products claiming to be similar to another one already marketed (i.e., biosimilars) and complements several previous guidelines for biosimilar medicinal products based on which 13 such products have been approved so far in Europe. The guideline is intended to provide product-specific guidance that presents the current view of the EMA on the demonstration of biosimilarity of two mAb-containing medicinal products. Although the new guideline is specifically directed to mAbs, it states that "the principles discussed may also, on a case-by-case basis, be relevant to related substances like, for example, fusion proteins based on IgG Fc. However, second- or next-generation biologicals that are structurally or functionally altered, in comparison with already licensed reference products, to gain an improved or different clinical performance" (i.e., biobetters) are beyond the scope of the guideline.

The second guideline on the immunogenicity assessment of mAb that was released by the EMA is applicable to all biological medicinal products containing mAbs, not just biosimilars. The guideline puts emphasis on the evaluation of clinical consequences of unwanted immune response for each mAb. The mode of mAb action is critical in this assessment because many mAb lyse or induce apoptosis in cells. Also important is whether adverse events such as infusion reactions can be properly handled by neutralizing the mAb with other antibodies or with medication. One standard requirement is that all mAbs will require a validated screening assay followed by a validated neutralizing assay. All antibodies must be classified as neutralizing or non-neutralizing, regardless of their risk level.

The "centralized" procedure has been made mandatory for all products derived from biotechnology. Under the centralized procedure, application for market approval is submitted to the EMA. The EMA issues an opinion (favorable or unfavorable) to the European Commission (EC), which then makes the final approval. Approval by the EC is valid in all EU member states. Any quality data for the biosimilar must fulfill all requirements for Common Technical Document (CTD) Module 3 as defined in Annex I to Directive 2001/83/EC and satisfy the technical requirements of the monographs of the European Pharmacopoeia and any additional requirements, such as defined in relevant CHMP and ICH guidelines.

Note: CTD is an assembly of all the quality, safety, and efficacy information in a common format for regulatory review processes and strives to eliminate the need to reformat the information for submission to the different ICH regulatory authorities.

*Module 3* A full quality dossier (chemical, pharmaceutical, and biologic information) is required for biosimilars. This should be supplemented by the demonstration of comparability (EMEA/CPMP/BWP/3207/00 and EMEA/CHMP/BWP/49348/05).

*Module 4* Before initiating clinical development, nonclinical studies should be conducted. The approach taken to reveal the similarities and differences between the biosimilar and reference product will need to be fully justified. Normally, other routine toxicologic studies, such as safety pharmacology, reproduction toxicology, mutagenicity, and carcinogenicity studies, are not required for biosimilars unless indicated by the results of repeat-dose studies (EMEA/CHMP/BMWP/42832/05).

*Module 5* Even though efficacy can be comparable, the biosimilar can show differences in the safety profile (in terms of severity or adverse events incidence). Clinical studies (rather than the bioequivalence studies used for traditional generics) are needed to support the biosimilar product's safety and effectiveness. In particular, the studies must address immunogenicity.

*Extrapolation of Indication* Under certain circumstances, the EMEA allows extrapolation of indication, that is, a biosimilar that has demonstrated comparable safety and efficacy in the "most sensitive" indication can be assumed to be able to extrapolate that safety and efficacy to other indications of the reference product (EMEA/CHMP/BMWP/42832/05).

The EU pharmaceutical legislation review [19] has resulted in a period of 8 years of data exclusivity for innovator product with an additional 2 years of market exclusivity, which can be extended further by a year (8 + 2 + 1) if a new indication is added within first 8 years for all registration procedures. In terms of patents the European Union now has the highest patent protection level in the world because of various regulatory situations, including patents for biotech products; market protection periods for 20 years; patents extendable to additional 5 years by Supplementary Certificate Regulation (1768/92/EEC); and patents that are also granted for new uses, dosages, and changes in formulations. The Supplementary Protection Certificate (Council Regulation [EEC] 1768/92) is a regulatory initiative introduced to give extended market protection to patented pharmaceutical products to compensate for time lost between patent filing and market authorization.

### 12.7.2 The United States

***12.7.2.1 Approval Process for Generic Synthetic Small Molecules*** Conventional small molecule drugs are regulated under the FDC Act passed in 1938 in the United States. Under this act, new innovator drugs must apply for marketing approval through a New Drug Application (NDA) under Section 505(b)(1). The Drug Price Competition and Patent Term Restoration Act (Hatch-Waxman Act) [20], enacted in 1984, amended the FDC Act and laid out a pathway for approval of generic versions of innovator conventional drugs. The Hatch-Waxman Act provides 5 years of market exclusivity to the sponsor of a new chemical entity (NCE), during which time applications cannot be submitted for alternate versions of the NCE). Furthermore, sponsors can receive 3 more years of market exclusivity for modifications to existing

products that require new clinical investigations. A method of permission to a generic drug sponsor to market its product is provided through FDA approval of an Abbreviated New Drug Application (ANDA) as described in Section 505(j) of the FDCA under the principle that "it is wasteful and unnecessary to carry out studies to demonstrate what is already known about a drug." The generic version of an innovator drug must essentially be identical to the innovator drug in active ingredient, dosage form, strength, route of administration, labeling, quality, performance characteristics, and intended use. Another ANDA route for generic drug approval also added to the FDC Act by the Hatch-Waxman Act, Section 505(b)(2) can be used by drugs that have been modified compared with the previously approved innovator drug yet are comparable. Such an application contains full reports of safety and investigation, but some of the investigations "were not conducted by or for the applicant and for which the applicant has not obtained a right of reference or use from the person by or for whom the investigations were conducted" (21 U.S.C. 355(b)(2). Drugs using this approach can include new chemical or molecular entities and drugs that have been changed compared with the previously approved version. According to the Office of Generic Drugs (OGD), which is part of the US FDA's Center for Drug Evaluation and Research (CDER), "a generic drug is identical, or bioequivalent to a brand name drug in dosage form, safety, strength, route of administration, quality, performance characteristics and intended use." Bioequivalence is defined in the United States in the Code of Federal Regulations, Title 21, Part 320.1 [21].

***12.7.2.2 Approval Process for Biosimilars*** An approval scheme, analogous to an ANDA under the Hatch-Waxman Act, cannot be applied to biosimilars because of the inability to declare the follow-on version identical to the innovator product and described differences to conventional small molecule drugs. As explained by Janet Woodcock of the FDA in testimony to the US House of Representatives on March 26, 2007, because of "the variability and complexity of protein molecules, current limitations of analytical methods, and the difficulties in manufacturing a consistent product, it is unlikely that, for most proteins, a manufacturer of a follow-on protein product could demonstrate that its product is identical to an already approved product. Therefore, the Section 505(j) generic drug approval pathway will not ordinarily be available for protein products" [22].

Starting with the Biologics Control Act, Congress has regulated most biologics separately from small molecules. In 1944, the Biologics Control Act was incorporated into the PHSA, under which biologics are still regulated today (for historical reasons, certain biologics are regulated under the FDCA, including hGH, calcitonin, and hyaluronidase) (Table 12.10). Historically, some simpler protein therapeutics have been approved under the FDC Act, including insulin, hyaluronidase, menotropins, and hGHs (Table 12.11). When recombinant proteins and monoclonal antibodies began development in the 1970s and 1980s, these were regulated as follows:

1. By the CDER under the Food, Drug and Cosmetic Act (FDCA) as drugs when they were hormones such as insulin, hGH, and PTH derivatives

**TABLE 12.10 Food and Drug Administration Center Regulatory Responsibility for Therapeutic Biological Products**

| CDER | CBER |
|---|---|
| mAbs (*in vivo* use) | Cellular product, including products composed of human, bacterial, or animal cells |
| Proteins intended for therapeutic use: cytokines (e.g., interferons), enzymes (e. g., thrombolytics) | Vaccines |
| Immunomodulators (nonvaccine, nonallergenic) | Allergenic extracts |
| Growth factors, cytokines, hormones, mAbs intended to mobilize, stimulate, decrease, or alter production of hematopoietic cells *in vivo* | Antitoxins, antivenoms, venoms |
| — | Blood, blood components, plasma-derived products, recombinant and transgenic versions of plasma derivatives |

CBER: Center for Biologics Evaluation and Research; CDER: Center for Drug Evaluation and Research; mAb: monoclonal antibody.

**TABLE 12.11 Examples of a Few Therapeutic Peptide and Protein Molecules Currently Marketed in the United States**

| | |
|---|---|
| Peptides | Antibiotics: bacitracin, bleomycin<br>Hormones: corticotrophin, glucagon, oxytocin, vasopressin<br>Others: polymixin B, cyclosporine etc |
| Nonglycosylated proteins | Interleukins: andresleukin (IL-1), denileukin diftitox (fusion, protein-IL2 + DT), anakinra (IL2)<br>Interferons: interferon-α-n1, interferon-α2a, interferon-α2b<br>Enzymes and inhibitors: lactase, trypsin, chymotrypsin, urokinase, papain<br>Growth factors and hormones: filigrastim, somatropin, menotropins<br>Others: insulin, hemoglobin |
| Glycosylated proteins | Interferon-β1a, alteplase, erythropoietin, normal immune globulin, factor VII, factor IX, etanercept, sargromostin |
| Monoclonal antibodies | Avciximab, palivizumab, basiliximab, gentuzumab, satumomab |

2. By the Center for Biologics Evaluation and Research (CBER) under the PHSA as biologics when they were cytokines or blood factors such as factor VIII for hemophilia
3. As other recombinant proteins and mAbs came under development, the CBER held primary responsibility for this review, with the CDER retaining responsibility for hormones such as insulin and hGH.

In 2003, all therapeutic proteins were transferred from the CBER to the CDER. This reassignment of review responsibility did not impact the legal classification of these protein products, such that the CDER assumed responsibility for the review and approval of biologics approved under Section 351 of the PHSA.

Table 12.11 shows categories of a few therapeutic peptide and protein molecules currently marketed in United States. To date, at least six biosimilars have been approved in the United States by the FDA, perhaps the most prominent of which is the follow-on of somatropin. Sandoz's Omnitrope intended to replicate the recombinant hGH somatropin (Genotropin, Pfizer), which is regulated under the FDCA. Because hGH products are approved under the FDCA, the abbreviated approval of Omnitrope did not establish a pathway for follow-on versions of biopharmaceuticals regulated under the PHSA. Several other biosimilars approved in the United States under Section 505(b)(2) include Hylenex (hyaluronidase), Hydase (hyaluronidase), Amphadase (hyaluronidase), Fortical (calcitonin), and GlucaGen (glucagon). Despite approval of these biosimilars by the FDA, these products have not been declared as directly substitutable for the reference products. The appropriateness of a Section 505(b)(2) application and the information required for each of the biosimilar approvals have been decided on a case-by-case basis by the FDA. Some of the factors that have influenced the amount and type of data required by the FDA for marketing approval of biosimilar include:

- The robustness of the manufacturing process
- The degree to which structural similarity can be assessed; minor differences in clinically inactive components are permitted.
- The extent to which mechanism of action was understood
- The existence of valid, mechanistically related PD assays
- The comparative PK
- The comparative immunogenicity
- The amount of clinical data available and extent of experience with the original product, or products

Because no defined regulatory pathway for the FDA to review or approve biosimilars in the United States was available, several bills to set up such a pathway were introduced in Congress. The challenge was to approve necessary regulations that protect innovator companies while maintaining the safety and efficacy of biosimilar products that would increase competition and lower drug costs. Five bills were introduced in the 110th Congress (January 3, 2007–January 3, 2009); however, none received a vote. Subsequently, two unique bills were introduced in the 111th Congress (January 3, 2009–January 3, 2011) related to biosimilars, including a reintroduction of The Access to Life Saving Medicine Act, now titled The Promoting Innovation and Access to Life Saving Medicine Act by Representative Waxman and Senator Schumer, and a reintroduction of The Pathway for Biosimilars Act by

Representative Eshoo (Table 12.12). The two bills survived committee votes in US Congress and the Senate, respectively, before being reconciled and passed by the House of Representatives in March 2010.

On March 23, 2010, President Obama signed into law the reconciliation bill that finalized Congress's health care reform legislation, the Patient Protection and Affordable Care Act (Affordable Care Act) (Pub. L. 111–148) [23]. The Affordable Care Act contains a subtitle called the Biologics Price Competition and Innovation Act of 2009 (BPCI Act) that by inserting subsection (k) amends the PHSA and other statutes to create an abbreviated approval pathway for biological products (chemically synthesized polypeptides are excluded) shown to be biosimilar to or interchangeable with an FDA-licensed reference biological product (Sections 7001–7003 of the BPCI Act). Included in this law is a provision to balance consumer interests and innovation for biological products through the Biologics Price Competition and Innovation Act.

The definition of "biological product" under the BPCI Act has been redefined to include:

(b)(2) The term "biosimilar" or "biosimilarity" in reference to a biological product that is the subject of an application under subsection (k), means

(A) that the biological product is highly similar to the reference product notwithstanding minor differences in clinically inactive components; and

(B) there are no clinically meaningful differences between the biological product and the reference product in terms of the safety, purity, and potency of the product.

(3) The term 'interchangeable' or 'interchangeability', in reference to a biological product that is shown to meet the standards described in subsection (k)(4), means that the biological product may be substituted for the reference product without the intervention of the health care provider who prescribed the reference product.

(4) The term "reference product" means the single biological product licensed under subsection (a) against which a biological product is evaluated in an application submitted under subsection (k).

The objectives of the BPCI Act are conceptually similar to those of the Drug Price Competition and Patent Term Restoration Act of 1984 (Pub. L. 98–417). Biosimilars are required to have the same route of administration, dosage form, and strength as the reference product and use the same mechanisms of action (MOAs) for the conditions of use but only to the extent that the mechanism of action is known. A biological product may be demonstrated to be "biosimilar" to a biological reference product based on data derived from analytical studies, animal studies, and a clinical study or studies if the product is shown to be highly similar to the reference product, notwithstanding minor differences in clinically inactive components, and if there are no clinically meaningful differences between the biological product and the reference product in terms of safety, purity, and potency.

**TABLE 12.12 Comparison of H.R. 1427, The Access to Life Saving Medicine Act (Now Titled The Promoting Innovation and Access to Life Saving Medicine Act), and H.R. 1548, a Reintroduction of The Pathway for Biosimilars Act**

| Provision | Waxman Bill (H.R. 1427) | Eshoo Bill (H.R. 1548) | Hatch-Waxman Act (1984) |
|---|---|---|---|
| Exclusivity | New product: 5 years<br>New indication: 3 years<br>Pediatric applications: additional 6 months | New product: 12 years*<br>New indication: 14 years†<br>Pediatric applications: additional 6 months‡ | New product: 5 years*<br>New indication: 3 years†<br>Pediatric applications: additional 6 months‡ |
| New clinical trials required? | No | Yes | No |
| First to file exclusivity | 1 year | 2 years | 180 days |
| Patent notification | • Biosimilar applicant or prospective applicant may send written request to approval for patent information at any time<br>• Approval holder has 60 days to provide patent information to biosimilar applicant<br>• Approval holder must update patent information for 2 years from date of first request<br>• Biosimilar applicant may provide notice of application to approval holder at any time; must include certification that product will not infringe or that patent claims are invalid | The FDA would provide a copy of a biosimilar application to the reference product approval holder, who would then be required to:<br>• Provide the biosimilar applicant with a list of patents that purport to cover the product or its use<br>• Explain in writing to the biosimilar applicant why the reference product approval holder believes the biosimilar product would infringe on one or more patents on that list | • Patents covering product or use must be listed by approval holder in the *Orange Book*<br>• Generic applicants must certify at time of filing of generic application that product will not infringe or that patent claims are invalid<br>• Generic applicant must notify patentee or approval holder of ANDA submission and explain response for clarification<br>• Approval holder or patentee has 45 days from notice to sue generic applicant for infringement |

(*continued*)

**TABLE 12.12** (*Continued*)

| Provision | Waxman Bill (H.R. 1427) | Eshoo Bill (H.R. 1548) | Hatch-Waxman Act (1984) |
|---|---|---|---|
| | • Approval holder or patentee has 45 days for notice to sue biosimilar applicant for infringement | | |
| Naming of generic | Biosimilar has same official name as reference product | Biosimilar must have a distinct official name from reference product | Generic product has same official name as reference product |

*Measured from date of first approval of new product using a given biologic.
†Measured from date of first approval of new indication.
‡Added to the end of exclusivity period provided for new product or indication.
ANDA: Abbreviated New Drug Application; FDA: Food and Drug Administration.

The BPCI Act also includes, among other provisions, a 12-year period of marketing exclusivity from the date of first licensure of the reference product, during which approval of a 351(k) application referencing that product cannot be made effective; an exclusivity period for the first biological product submitted in a 351(k) application that has been determined to be interchangeable with the reference product for any condition of use, during which a second or subsequent biological product may not be determined interchangeable to that reference product; and a transition provision for protein products that have been or will be approved under section 505 of the FD&C Act (21 U.S.C. 355) prior to March 23, 2020.

***12.7.2.3 Impact on Patent Prosecution, the Food and Drug Administration, and Litigation Practices (Fig. 12.6) [24]*** Several provisions of the BPCI Act will impact patent prosecution and litigation, as well as FDA practice. Areas of particular interest and opportunity are listed below:

*Pioneer Exclusivity* The BPCI Act grants pioneers 12 years of marketing exclusivity from application approval (with a potential further 6 months for pediatric studies). In addition, during the first 4 years of such exclusivity (or 4 years plus 6 months for pediatric studies), the follow on biologics (FOB) applicant cannot submit an FOB application.

*Biosimilar Applicant Exclusivity* Biosimilar applicants are awarded exclusivity for being the first product determined to be "interchangeable" with the pioneer biologic. The law contains provisions for marketing exclusivity for both the interchangeable biological product and the reference product. For biological products labeled biosimilar, no marketing exclusivity is available. The first product to be granted interchangeability against a specific reference standard receives 12 months of marketing exclusivity. If changes are made to a reference product, 12 years of exclusivity may be granted if a new Biologic License Application (BLA) must be filed. Two conditions are described in the legislation to be granted this exclusivity:

1. A structural change made to the molecule that results in a change in safety, purity, or potency
2. No change to the molecular structure; the indication, route of administration, dosing schedule, dosage form, delivery system, delivery device, or strength must also remain the same

Marketing exclusivity is extended for the conduct of pediatric studies and orphan drugs.

The marketing periods described above are extended by 6 months for new biosimilar or interchangeable biological products or already marketed biological products if studies are completed in the pediatric population and the reports are acceptable to the agency. This period may not be extended if the determination is made later than 9 months before the expiration period. For drugs for rare diseases, exclusivity is extended an additional 6 months to 7 years, 6 months.

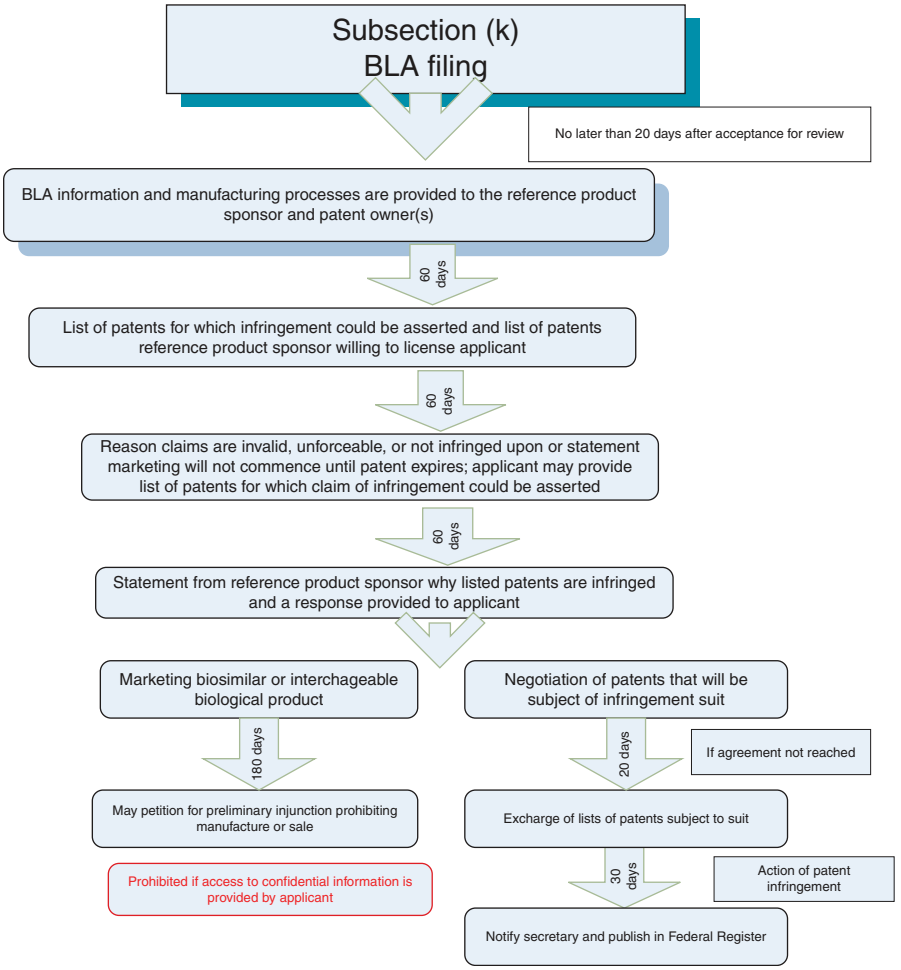


**FIGURE 12.6** Patent infringement resolution and negotiations: process and timelines. BLA: Biologic License Application.

*Patent Notices and Pre-litigation Disclosures (see Fig. 12.6)* Unlike in the Hatch-Waxman context, in which pioneers provide notice of the patents that protect their product in a public forum (the *Orange Book*), the BPCI Act provides for private notices of such patents.

*Notice to Pioneer of Product and Process* The biosimilar applicant must provide a copy of its application and a description of its manufacturing process to the pioneer within 20 days of acceptance of the application for review. This information and data must be used for the sole and exclusive purpose of determining whether a claim of patent infringement could be reasonably asserted. Written consent from the applicant is required to disclose this confidential information to "any other person or entity," and

this consent cannot be reasonably withheld. Because of the irreparable harm experienced by the applicant if this confidential information is disclosed to others, the court may consider injunctive relief for violations.

- Pioneers' list of relevant patents: Within 60 days of the production of the application and manufacturing process, the pioneer must identify the relevant patents it could assert, the patents to which it has an exclusive license, and the patents it is willing to license. After receipt of this confidential information, a formal resolution process takes place that involves the mutual exchange of information relating to patents. Timing constraints have been put in place to ensure that the process moves forward as expeditiously as possible. Within 60 days of the reference product sponsor's receipt of the subsection (k) BLA and manufacturing processes data and information, the sponsor is to provide the applicant with a list of patents for which a claim of infringement could be reasonably asserted. In addition, the reference product sponsor needs to identify patents on the list that it would be willing to license to the applicant. Within 60 days of receiving this list, the subsection (k) applicant must provide a detailed statement of the reason why the claims listed by the reference product sponsor are invalid, unenforceable, or will not be infringed upon and a response to the licensing offer or a statement that marketing of the biosimilar or interchangeable biological product will not begin before the patent expires. The applicant may provide to the reference product sponsor a list of patents for which a claim of infringement could reasonably be asserted. Within 60 days of receipt of this response, the reference product sponsor must provide to the applicant a detailed statement of why the listed patents are infringed and a response to the validity and enforceability statements.
- Biosimilar applicant's contentions: Within 60 days of the production of the list of patents by the pioneer, the biosimilar applicant shall provide a detailed (i.e., claim-by-claim) statement that describes the factual and legal basis for why each patent listed by the pioneer is not valid and enforceable or would not be infringed by commercial manufacture of the biosimilar product or state that the FOB applicant does not intend to commercially market any biosimilar product until the date one or more patents expire.

*Determining Which Patents Will Be Litigated* The BPCI Act also provides that the parties shall engage in good faith negotiations with a goal of agreeing on which, if any, patents should be litigated. The parties are given 15 days to resolve any differences.

*Limits on Staying Biosimilar Approval* The BPCI Act does not grant pioneers an automatic stay to prevent generic approval during the pendency of patent litigation. Because the disclosure periods in the BPCI Act are very short and far more analysis is demanded up front than in the Hatch-Waxman setting, this is the time for pioneers to audit their portfolios of patents relating to biologic patents to identify and fix any potential issues. They should stress test the patents for the validity and

non-infringement positions biosimilar applicants are likely to take hence pioneers need to plan and refine their portfolios accordingly.

On November 2, 2010 there was a public hearing because the FDA was seeking comments on several issues (partial list below) [25]:

1. What scientific and technical factors should the agency consider in determining whether the biological product is highly similar to the reference product notwithstanding minor differences in clinically inactive components?
2. What range of structural differences between a proposed biosimilar product and the reference product is consistent with the standard "highly similar" and may be acceptable in a 351(k) application if the applicant can demonstrate the absence of any clinically meaningful differences between the proposed biosimilar product and the reference product?
3. Under what circumstances should the agency consider finding that animal studies or a clinical study or studies are "unnecessary" for submission of a 351(k) application?
4. What factors should the agency consider in evaluating the potential risk related to alternating or switching between use of the proposed interchangeable biological product and the reference product or among interchangeable biological products?
5. What types of guidance documents for industry should be a priority for the agency during the early period of implementation?

Subsequently, on February 9, 2012 the FDA issued the following three draft guidance documents on biosimilar product development to assist industry in developing such products in the United States. Comments to the draft guidances were due by April 16, 2012 followed by a public meeting on May 11, 2012.

- "Scientific Considerations in Demonstrating Biosimilarity to a Reference Product": This draft guidance is intended to assist companies in demonstrating that a proposed therapeutic protein product is biosimilar to a reference product for the purpose of submitting an application, called a "351(k)" application, to the FDA. This draft guidance describes a risk-based "totality-of-the-evidence" approach that the FDA intends to use to evaluate the data and information submitted in support of a determination of biosimilarity of the proposed product to the reference product. As outlined in the draft guidance, the FDA recommends a stepwise approach in the development of biosimilar products.
- "Quality Considerations in Demonstrating Biosimilarity to a Reference Protein Product": The draft guidance provides an overview of analytical factors to consider when assessing biosimilarity between a proposed therapeutic protein product and a reference product for the purpose of submitting a 351(k) application. This includes the importance of extensive analytical, physicochemical, and biological characterization in demonstrating that the proposed

biosimilar product is highly similar to the reference product notwithstanding minor differences in clinically inactive components.

- "Biosimilars: Questions and Answers Regarding Implementation of the Biologics Price Competition and Innovation Act of 2009": The draft guidance provides answers to common questions from people interested in developing biosimilar products. The question-and-answer format addresses questions that may arise in the early stages of product development, such as how to request meetings with the FDA, addressing differences in formulation from the reference product, how to request exclusivity, and other topics.

Following summarizes a few aspects of the draft guidances:

1. *Stepwise approach and totality of evidence to demonstrate and assess biosimilarity:* The concept of a totality-of-the-evidence approach can serve as an appropriate and efficient approach to demonstrating biosimilarity. The more comprehensive and robust the comparative structural and functional characterization, the more useful such characterization can be in determining what additional studies may be needed. So if a more rigorous structural and functional comparison shows minimal or no difference between the proposed product and reference product, the stronger the scientific justification for a selective and targeted approach to animal or clinical testing to support a demonstration of biosimilarity. Such assessments should used state-of-the-art methods, consistent test material and reference product, consistent reagents, and sensitive assays. Safety and immunogenicity need to be evaluated in appropriate preapproval clinical studies and, when appropriate, postmarketing studies of sufficient duration should be done to detect and assess differences in a sensitive population. Also, the draft guidances imply that in some circumstances, the amino acid sequence of a biosimilar and its reference product may be different. The FDA may need to clarify the extent to which sequence variants can exist within the concept of biosimilarity. Also, the guidance suggests FDA expects injectable biologicals to have both the same total content of drug substance (in mass or units of activity in a container closure) and the same concentration of drug substance (in mass or units of activity per unit volume) as the reference product to have same strength. The FDA also expects that applicants characterize, identify, and quantify impurities (product and process related as defined in ICH Q6B) in the proposed biosimilar product and the reference product. Some innovator companies, however, have argued that biosimilar development program should include, at a minimum, side-by-side analytical evaluations as well as comparative nonclinical studies (*in vitro, in vivo*, or both), comparative human PK and PD studies (when a relevant response marker exists), and a comparative clinical study addressing safety and efficacy separately or jointly.
2. *Interchangeability:* The initial guidances do not reflect in detail on interchangeability. Hence, the sponsor may need to seek biosimilar designation first followed by additional data to request designation of interchangeability later. The

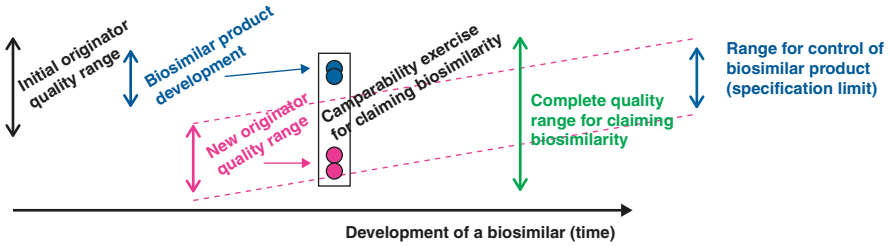


**FIGURE 12.7** A model of developing "goal posts."

opportunity to seek interchangeability at the time of initial application appears to be vague. This raises concerns whether interchangeability is a higher standard of approval than biosimilarity and thus presents a complex set of regulatory challenges to be addressed fully in a separate guidance. McCamish and Woollett [18] have explained a model of developing "goal posts" whereby a scientific basis for establishing a highly similar biosimilar is initiated (Figure 12.7) [18]. They suggest that even originator biologic products undergo changes over time (e.g., when manufacturing process is modified by scaling up or transferred to alternate sites). Changes in originator product attributes provide a distribution of attributes that should be considered acceptable because the products were subsequently administered to patients successfully. The limits to distribution of such attributes are what they call "goal posts," which could be used to demonstrate acceptability of a biosimilar. After "goal posts" are established for an originator reference product, an iterative process can be used in developing a biosimilar in which the product is fully evaluated by physicochemical and biological characterization compared with the originator product. If product attributes fall outside established "goal posts," various process steps can be modified to produce product attributes that are within the variability of the originator product. This process can be repeated until highly similar product attributes are achieved. The degree of similarity that is established can be based on analytical characterization that is run in parallel with iterative process development activities. Depending on the extent that product attributes of the biosimilar overlap with those of the originator product, preclinical and clinical development can be abbreviated compared with developing a novel biologic.

The goal posts of biosimilarity are established by the biosimilar sponsor by their analysis of the distribution of product attributes present in the reference product pre- and postmanufacturing change. They then use these to select the design space for their biosimilar candidate. Although the complete quality range may be quite broad for the life of the reference product, the biosimilar sponsor will select a tighter range of control for its biosimilar product.

In May 2006, Sandoz achieved an approval for Omnitrope (somatropin) based on a demonstration of "high similarity" or comparability to Pfizer's Genotropin (somatropin), and in July 2010, the approval of enoxaparin as a fully

substitutable generic to Sanofi-Aventis' Lovenox (enoxaparin). In reviewing the enoxaparin file, the FDA established five criteria to determine "sameness" of Sandoz's enoxaparin to Lovenox. These were physiochemical properties, source material, the sugar building blocks, biologic assays, and *in vitro* PD profile. Each of these criteria had to be comparable for the FDA to rule that the two products were the "same." In this case, the analytical characterization, along with other information, provided sufficient data for the FDA to approve the product. Again, enoxaparin was not approved as a biosimilar under the new BPCI Act pathway, but the scientific approach used by the FDA is equally applicable.

3. *Foreign reference product issues:* The draft "Scientific Considerations Guidance" states that "[t]o obtain licensure of a proposed product under section 351(k) of the Public Health Service (PHS) Act, a sponsor must demonstrate that the proposed product is biosimilar to a single reference product that previously has been licensed by FDA. However, under certain circumstances, a sponsor may seek to use data derived from animal or clinical studies comparing a proposed product with a non-U.S.-licensed product to address, in part, the requirements under section 351(k)(2)(A) of the PHS Act."

   The challenge here is foreign versions of US-licensed reference products can significantly vary in the extent to which they are similar to the US product. In some cases, a single manufacturer may release the same drug product (with differences only in packaging and labeling) from the same manufacturing plant(s) into both the US and non-US markets. In other cases, products intended for different markets may be manufactured independently or comply with locally specific quality or pharmacopeial standards. When products are manufactured by more than one entity, two products that were once related may have diverged from each other over time in subtle but important ways. Moreover, some innovative products manufactured abroad may have even more fundamental differences from the products intended for marketing in the United States (e.g., because of the use of different cell lines, formulations, container closure, or shipping methods. The foreign comparator product may not have the same formulation as the US-licensed reference product. The manufacturing facilities for the foreign comparator product may not be in compliance with current good manufacturing practices (cGMPs) and may not be subject to regularly scheduled inspections by the FDA or equivalent regulatory agency. When a foreign comparator product can be determined to be fully representative of the US-licensed reference product, a sponsor is allowed to use data derived from a study conducted with this foreign comparator to address components of the statutory requirements. This helps avoid unethical conduct of duplicative trials. Because a scientific stepwise approach for bridging US-licensed product to non–US-licensed biologic is the basis of allowing use of foreign sourced biologic as a comparator agent in any characterization, animal, or clinical efficacy and safety trials that are needed. As suggested by McCamish and Woollett [18], structural and functional data can be the foundation for bridging of the US reference product to the foreign comparator biologic (Fig. 12.8). If residual

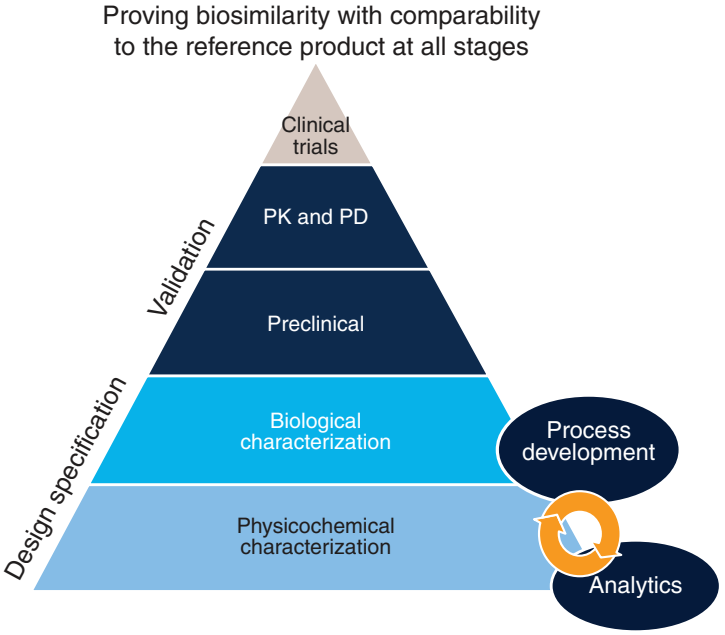


**FIGURE 12.8** Iterations toward biosimilar development. PD: pharmacodynamics; PK: pharmacokinetics.

uncertainty remains with regards to comparability of the biologic, PK and PD data could be used.

Because of changes an innovative product sponsor may make to its manufacturing process, a reference product studied in a biosimilar development program may change (either during or after product development) in subtle ways that are unknown to the biosimilar applicant. To address this concern, the FDA could recommend that a biosimilar applicant conduct its analytical, nonclinical, and clinical comparisons during a limited period of time and then establish an internal reference standard. Neither the biosimilar nor the reference product would thus be constrained from improving their processes and complying with cGMP after approval of a biosimilar product. The biosimilar and reference products could be treated as independent products, each standing on their own merits for future manufacturing changes and quality evaluations. To do otherwise could constrain manufacturing improvements.

4. *Animal data:* Nonclinical *in vivo* safety evaluation of a biosimilar is generally required to be conducted before clinical studies. The PHSA states that an application must include information demonstrating that the biologic is biosimilar to a reference product based on data derived from animal studies (including the assessment of toxicity), suggesting that the toxicology study

should be designed to compare the two products (not reestablish the toxicity profile of the reference product). The FDA could provide greater clarity with regard to when animal toxicity studies will be appropriate to help inform biosimilar development program planning, and this issue could be further addressed in class-specific guidances.

5. *Clinical studies and immunogenicity assessment:* The draft guidance suggests that an applicant should use a population PK modeling approach to explore potential variability in PK. The draft also suggests that establishing a similar human PK and PD profile may provide a basis for a selective and targeted approach to subsequent clinical testing. However, the FDA has been given the authority to waive any unnecessary analytical and preclinical studies. This is because if only clinical data were to be meaningful, then all requirements to demonstrate analytical similarity and conduct animal study would produce data of little value in determination of biosimilarity or interchangeable biosimilarity. Hence, the scope and magnitude of clinical studies depend on the extent of residual uncertainty about the biosimilarity of the two products after conducting structural and functional characterization and possible animal studies. The guidance also suggests that scope of the clinical program for the purpose of determining clinically meaningful differences may be narrowed to the conduct of the PK and PD studies when appropriate PD markers are available.

   The draft guidance states that if immune response to the reference product is rare, two studies may be needed to evaluate immunogenicity: (1) a premarket study to detect major differences in responses and (2) a postmarket study to detect subtle differences.

6. *Clinical safety and effectiveness data:* The guidance states that comparative clinical safety and effectiveness data are necessary if there are "residual uncertainties" about biosimilarity based on structural and functional characterization, animal testing, human PK and PD data, and a clinical immunogenicity assessment. When the similarity exercise leaves considerable residual uncertainty about the structural and functional similarity of the biosimilar and the reference product, a more rigorous (equivalence) assessment with more traditional endpoints may be necessary.

7. *Clinical trial design:* For biosimilars, when clinical trials are determined to be a useful confirmation of biosimilarity, sponsors may have the option of a variety of study designs. The study design could vary considerably depending on the product, the patient population, and the indication being considered for the initial study (recognizing that not all indications for all biosimilars will need to be studied in human trials). For example, in the approval of Sandoz's granulocyte colony-stimulating factor (GCSF) as a biosimilar in Europe, the clinical trial designs relied primarily on phase 1 studies to show that comparability can be well established using PK and PD studies. Specifically, Sandoz [18] refers to a recent peer-reviewed publication in which they describe the approach to biosimilarity used to develop commercial GCSF, or filgrastims, in some detail. Filgrastim, being a nonglycosylated peptide, is

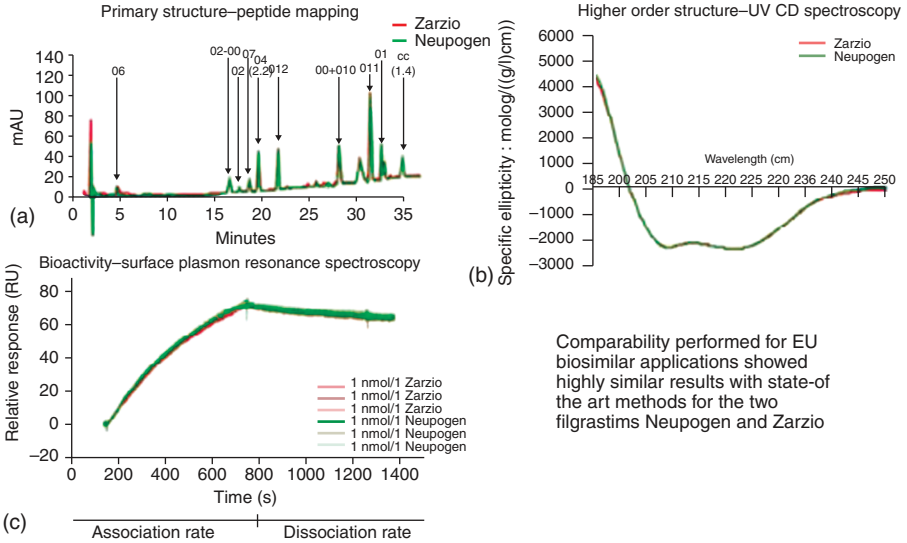


**FIGURE 12.9** Biosimilar filgrastim compared with the reference product (Neupogen) using three state-of-the-art analytical techniques. (a) Peptide mapping (Zarzio vs. Neupogen). (b) UV CD spectroscopy (higher order structure). (c) Bioactivity—surface plasmon resonance spectroscopy. CD: circular dichroism; UV: ultraviolet.

one of the easiest biologics to re-create, but that clinical studies can be minimal is critically important to consider as the United States implements the abbreviated pathway here. The first step in the process was to demonstrate chemical and biological comparability to the originator. In Figure 12.9, Sandoz scientists [26] documented overlapping peptide mapping, higher order structures comparing multiple batches of the biosimilar with the originator product.

8. *Extrapolation between conditions of use (indications):* According to the draft guidance, a sponsor of a proposed product may seek licensure only for a condition of use that has been previously licensed for the referenced product. In addition, a biosimilar applicant should scientifically justify the extrapolation of clinical data to support a determination of biosimilarity for each condition of use of the reference product for which licensure is sought. Such scientific justification would need to address the MOA(s) in each condition of use for which licensure is sought. The FDA recommends that the sponsor consider whether tested condition of use is the most sensitive one in detecting clinically meaningful differences in safety and effectiveness.

   For a biosimilar sponsor, however, extrapolation is one of the leading reasons why a sponsor may choose to apply to the FDA for a product licensure, hence the guidance's preclude any real opportunity for an abbreviated approach toward a biosimilar approval. Also, it is unconceivable that a biosimilar shown to be

"highly similar" to its reference product not share the same MOA; hence, receptor binding between biosimilar and originator in head-to-head studies should suffice without further elaboration of downstream intracellular signaling. An FDA guidance during early development regarding which populations should be studied to allow extrapolating to all remaining indications would be invaluable.

9. *Pediatric studies and waivers:* The FDA states that a noninterchangeable biosimilar biologic product should be considered to have a new active ingredient. This leads to consequence of pediatric studies being expected for the approval of a biosimilar but not for an interchangeable biosimilar unless the FDA issues a pediatric waiver. The FDA has made it clear in the public statement and guidances encouraging sponsors not to apply for interchangeable biosimilars at this time but to consider a two-step process whereby they achieve a biosimilar approval and then consider a supplement to gain the interchangeable designation. Given that studies with biosimilars will be necessary as a scientific matter if biosimilarity has been established in adults, performing scientifically meaningless studies is unethical. It is therefore important that the FDA make clear the possibility of a waiver for these studies rather than suggest that clinical studies in pediatric populations will be the norm.
10. *Postmarketing safety and assessment:* Postmarketing safety monitoring applies equally to small and large molecules and to originator, biosimilar, and generic drugs irrespective of the sponsors. Postmarketing surveillance is fundamental to defining the safety profile and effectiveness of any product in the marketed setting because clinical development programs are necessarily finite in size and duration and will not always identify rare and unexpected adverse effects. However, it is also inherent in the approval of a biosimilar that the connection has been built as a scientific, regulatory, and legal matter to the prior finding of safety, purity, and potency of the originator product. Hence, the unknown at the time of biosimilar approval is vastly smaller than at the time of the originator product. It is possible for antidrug antibody response to be induced after long-term exposure to a foreign protein. However, it should not be assumed a priori that such events will occur because the biologic is a biosimilar. Nonetheless, if large-scale immunogenicity studies are considered necessary to assess this possibility, they could be conducted after licensure but would also be needed for the requirements for the originator, which may use comparability protocols during transfer or process improvement endeavors.
11. *User fees:* The Federal Food, Drug, and Cosmetic Act, as amended by the Biosimilar User Fee Act of 2012, authorized the FDA to assess and collect fees for biosimilar biological products from October 2012 through September 2017 (Table 12.13). The FDA dedicates these fees to expediting the review process for biosimilar biological products.

**TABLE 12.13 Fee Schedule (2013)**

| Fee Category | Fee Rate for Fiscal Year 2013 |
|---|---|
| Initial BPD | $195,880 |
| Annual BPD | $195,880 |
| Reactivation | $391,760 |
| Biosimilar Biological Product Application | |
| Requiring clinical data | $1,958,800 |
| Not requiring clinical data | $979,400 |
| Biosimilar biological product supplement requiring clinical data | $979,400 |
| Biosimilar biological product establishment | $526,500 |
| Biosimilar biological product | $98,380 |

BPD: biological process development.

## 12.8 CONCLUSION

In the past two decades, the biotechnology industry has made great advances in improving our understanding of manufacturing processes and their starting materials, as well as of the impact they have on product quality. Whereas early biopharmaceutical regulation was grounded by technical limitations, modern biopharmaceutical regulation is driven by tremendous advances in scientific knowledge. Development of analytical tools and techniques has dramatically increased the ability to characterize proteins and substantiate the structure, composition, and function of the therapeutic molecule. These advances enable the detection of small differences in molecular weight; elucidation of primary, secondary, and tertiary protein structures; detection and quantification of posttranslational modifications (i.e., patterns of glycosylation); and improved understanding of structure–function relationships and potential immunogenic responses. Simultaneously, developments in the fields of pharmaceutical and biotechnological manufacturing have greatly improved process efficiency and control.

Sandoz, the generics arm of Novartis, which has three approved biosimilar products, has begun a phase II clinical trial for a biosimilar version of rituximab, the active ingredient in Rituxan/MabThera, an mAb used to treat non-Hodgkin's lymphoma and rheumatoid arthritis, which is marketed by Roche/Genentech and Biogen Idec. The phase II study is targeted to patients with rheumatoid arthritis and aims to demonstrate bioequivalence to the reference product by assessing data on PK and PD as well as efficacy and safety. Sandoz also reported that it has developed a high-yield and large-scale process for rituximab at its facilities in Schaftenau, Austria.

In Japan, Sandoz has gained marketing approval of Japan's first therapeutic biosimilar, Somatropin BD. Also, in Canada, Sandoz has received market authorization for Omnitrope, which is the first biosimilar to achieve approval from Health Canada under Information and Submission Requirements for Subsequent Entry Biologics. An Indian company, Dr. Reddy's Laboratories, has already developed and

received marketing approval for a biosimilar version of the monoclonal antibody rituximab (Genentech, Inc. and Biogen Idec, Inc.), called Reditux.

In summary, the European Union has a well-established and well-documented legal and regulatory pathway for the review and approval of biosimilars. With a biosimilar registration pathway in place in the United States, it is expected to pave the way for the market entry of biosimilars in 2013. The FDA may commit to publishing class-specific guidances that discuss the approval standards and other key implementation issues for particular product classes. Class-specific guidance will provide the most meaningful guidance for manufacturers and will foster the confidence in biosimilars needed for physician and patient acceptance of this new class of product.

Other organizations and countries around the world, including Australia, Canada, Japan, and Switzerland, are also following the same scientific principles for an abbreviated approval pathway for biosimilars. Many Southeast Asian countries, including China, India, and South Korea, also have regulations in place to deal with biosimilars.

## REFERENCES


1. McLeod LD, de Groot A, Martin B. Biotherapeutics. BioProcess Int 2009;7(S5):28–33.
2. Everett E, Wright EL. Citizens Against Government Waste. Biogenerics: What They Are, Why They Are Important, and Their Economic Value to Taxpayers and Consumers. Final Report; 2007. Available from: http://councilfor.cagw.org/site/DocServer/Biogenerics_FINAL.pdf?docID=2221.
3. Radelate A. AARP VIVA, Making "Miracle" Drugs Affordable. 2009. http://www.aarp.org/about-aarp/info-05-2010/aarp_web_site_copyright_information.html.
4. Title VII—Improving Access to Innovative Medical Therapies, Subtitle A—Biologics Price Competition and Innovation. Available from: http://www.healthcare.gov/center/authorities/title_vii_improving_access_to_innovative.pdf.
5. Walsh G. Biopharmaceuticals and biotechnology medicine: an issue of nomenclature. Eur J Pharm Sci 2002;15(2):135–138.
6. Pisani J, Bonduelle Y. Consultants at Pricewaterhouse Coopers LLP. Pharmaceutical Strategy. Opportunities and Barriers in the Biosimilar Market: Evolution or Revolution for Generic Companies? Available from: http://www.ableindia.org/biosimilars.pdf.
7. Medical Research Council. Therapeutic Antibodies. Available from: http://www.mrc.ac.uk/Achievementsimpact/Storiesofimpact/Therapeuticantibodies/index.htm.
8. Visiongain Ltd. Therapeutic Monoclonal Antibodies: World Market 2010–2025. Available from: http://prlog.org/10985333.
9. Mansell P. Scrip Report: New Dynamics and Challenges in the Generics Market. 2007 Owner Informa plc Publisher Philip Jarvis Editor Mike Ward Founded March 1972.
10. Woodcock J. The FDA's assessment of follow-on protein products: a historical perspective. Nat Rev Drug Discov 2007;6:437–442.
11. Genentech. Biosimilars or Follow-on Biologics. Available from: http://www.gene.com/gene/about/views/followon-biologics.html.
12. Rathore AS. Follow-on protein products: scientific issues, developments and challenges. Trends Biotechnol 2009;27(12):698–705.

13. Schiestl M, Stangler T, Torella C, Cepeljnik T, Toll H, Grau R. Acceptable changes in quality attributes of glycosylated biopharmaceuticals. Nat Biotechnol 2011;29(4): 310–312.
14. World Health Organization. Various government publications. Available from: http://www.who.int/medicines/services/inn/inn_bio/en/index.html.
15. European Generic Medicines Association. Various government publications. Available from: http://www.egagenerics.com/doc/bio_rossignol-testimony2007.pdf.
16. Zuniga L, Calvo B. Regulatory aspects of biosimilars in Europe. Trends Biotechnol 2009;27:385–387.
17. European Medicines Agency, Committee for Medicinal Products for Human Use. Guideline on Similar Biological Medicinal Products. Available from: http://www.ema.europa.eu/pdfs/human/biosimilar/043704en.pdf.
18. McCamish M, Woollett G. The state of the art in the development of biosimilars. Clin Pharmacol Ther 2012;91(3):405–417.
19. European Generics Medicine Association. Various government publications. Available from: http://www.egagenerics.com/pr-2010-09-02.htm.
20. Library of Congress. Various government publications. Available from: http://thomas.loc.gov/cgi-bin/bdquery/z?d098:SN01538:@@@D&summ2=m&|TOM:/bss/d098query.html.
21. Code of Federal Regulations. Various government publications. Available from: http://edocket.access.gpo.gov/cfr_2003/aprqtr/pdf/21cfr320.1.pdf.
22. Statement of Janet Woodcock, M.D. Deputy Commissioner, Chief Medical Officer Food and Drug Administration Before the Committee on Oversight and Government Reform United States House of Representatives, "Follow-On Protein Products", March 26, 2007. Available from: http://www.reform.democrats.house.gov/documents/20070326104056-22106.pdf.
23. The Patient Protection and Affordable Care Act. Various government publications. Available from: http://www.gpo.gov.
24. Sensabaugh SM. Requirements for biosimilars and interchangeable biological drugs in the United States. Drug Info J 2011;45 pp 155–162.
25. Federal Register Notice. Government publication. Docket No. FDA-2101-N-0477. 2010;75(192).
26. McCamish M, Woollett G. Worldwide experience with biosimilar development. mAbs 2011;3(2):209–217.

# Vaccine Development: History, Current Status, and Future Trends

LEO VAN DER POL and JEAN-PIERRE AMORIJ

## 13.1 INTRODUCTION

Vaccines consist of a heterogeneous group of biopharmaceuticals that show little synergy to allow a standardized approach for product and process development. For a more homogeneous group of recombinant protein biopharmaceuticals, such as monoclonal antibodies (mAbs), a modular approach has been accomplished that reduces the risk and required time for a new development project. These protein products are well-characterized, facilitating a quality by design (QbD) approach. For vaccines, a prospective design approach is complicated by the limited knowledge on host–pathogen interaction, in particular the responses of the human immune system, and the great variability of many pathogens. This frustrates the formulation of product quality features of a vaccine to guarantee the exertion of the desired (protecting) effect. The classical trial-and-error approach for vaccine development is often insufficient to combat variable and immune-deceiving pathogens and generates limited scientific understanding for the present-day desired level of control on process and product. Therefore, it seems promising that progress in scientific understanding of, in particular, the innate immune system enables tailored vaccine design. Knowledge-based adjuvant design and the upcoming recombinant subunit vaccines, conjugate vaccines and virus-like particle (VLP) vaccines, are expected to facilitate a vaccine design based on a structure–function relationship in the future. This chapter presents an overview of vaccine development with special attention on process development and the route toward QbD for vaccines.

Vaccines are a group of biopharmaceuticals [1] that show heterogeneity as broad as the whole group of biopharmaceuticals. This is because pathogens are diverse, and the human immune response to these invaders is complex. For most viral and

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*,
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

bacterial intruders, such as tetanus and diphtheria, the immune response elicited by the classical vaccines was sufficient to reach a high level of protection in the developed countries in a very cost-effective manner. However, for the (highly variable) pathogens with immune-deceiving mechanisms, assistance by antiviral agents or antibiotics is required. The enormous impact of the classical vaccines in reducing the disease burden of diseases such as diphtheria, tetanus, whooping cough (pertussis), and polio (DTP-P) is fading from the common active knowledge, but incidents receive considerable attention in the Internet era, creating a great focus on product safety.

The introduction of many recombinant protein products, such as insulin, erythropoietin (EPO), and mAbs, has changed both production techniques and regulatory requirements. For vaccine development and the market, this had great impact, including on trying to meet increasing regulatory demands, including good manufacturing process (GMP) guidelines and substantial clinical trials, because a traditional low-cost product created a serious threshold for the introduction of new vaccines and a great reduction in the number of introduction of new vaccines on the market as well as the number of vaccine manufacturers.

Vaccine development projects that were started after the introduction of the classical vaccines had to face increasing demands in regulatory guidelines such as GMP [2], comparability [3] (product + impurity profile), International Conference on Harmonization (ICH) Q8 [4] (pharmaceutical development, QbD), ICH Q9 [5] (risk assessment), the process analytical technology (PAT) initiative [6], and toward continued process and product improvement using QbD approaches [7,8]. A common theme is gaining sufficient scientific understanding of the product and the process to be able to explain the specific product quality features and explain the rationale behind performing the different process steps under given conditions.

For small molecules and recombinant protein pharmaceutical products, the requirements have been implemented in the development approach and have had a beneficial effect on development time and production performance by providing better control and troubleshooting capabilities. This has resulted in several modular and standard approaches for these biopharmaceutical products. Although the classical vaccines were among the first biotech pharmaceutical products to be introduced in the past century, there exist little modular or standardized development approaches for vaccines. For vaccines, the implementation of the continuously evolving regulatory guidelines is an immense challenge. This is mainly because of the lack of knowledge on host–pathogen interaction (the interaction between the pathogen and the human immune system resulting in an immune response with the required phenotype (markers or correlates of protection [CoP])) that is needed to combat an attack of a pathogen and therefore the associated lack of specific criteria for product and process for state-of-the-art vaccine development. In addition, several features of vaccines (Table 13.1), such as the low risk-to-benefit ratio as a result of the (predominantly) healthy target population for vaccination, large study populations for efficacy studies as a result of low infection rates, the need for combination vaccines with coming challenges of immunologic and pharmaceutical interference, and the low therapeutic dose and subsequent analytical challenges make

**TABLE 13.1 Specific Features of Therapeutic Biopharmaceuticals and Prophylactic Vaccines**

| | Therapeutic Biopharmaceuticals | Preventive Vaccines | Trend Vaccine |
|---|---|---|---|
| Target population | Patients with specific disease: majority adults and elderly adults | Healthy population: majority infants and children | Toward elderly (including therapeutic vaccines) |
| Acceptable risk-to-benefit ratio | High (depending on disease) | Low | |
| Phase III, number of subjects | 100–1000 | >3000 (depending on incidence of infection) | |
| Therapeutic dose | Milligram range | Microgram range | |
| Number of doses or subject | Multi = regular injections daily or weekly for years | 1–4 | |
| Need for combinations or needle-free delivery | No | Yes | |
| Basis for product target profile | Structure activity relation | Complex immunology: multiple correlates of protection (e.g., T-cell response, systemic B-cell responses, mucosal immune responses) | |
| Average development time | 8–12 years, decreasing | 15–20 years | Shorter for bioterrorism vaccine application without clinical trials |
| "Success rate" | One in six when reaching toxicology studies and clinical trials | Lower than one in 10 when entering clinical studies | |
| Main features bill of testing | PK and PD<br>Repeated toxicology | Efficacy<br>Safety<br>Repeated toxicology | |
| Cost | 1000 Euros per gram | | Increasing in developing world |

PD: pharmacodynamics; PK: pharmacokinetics.

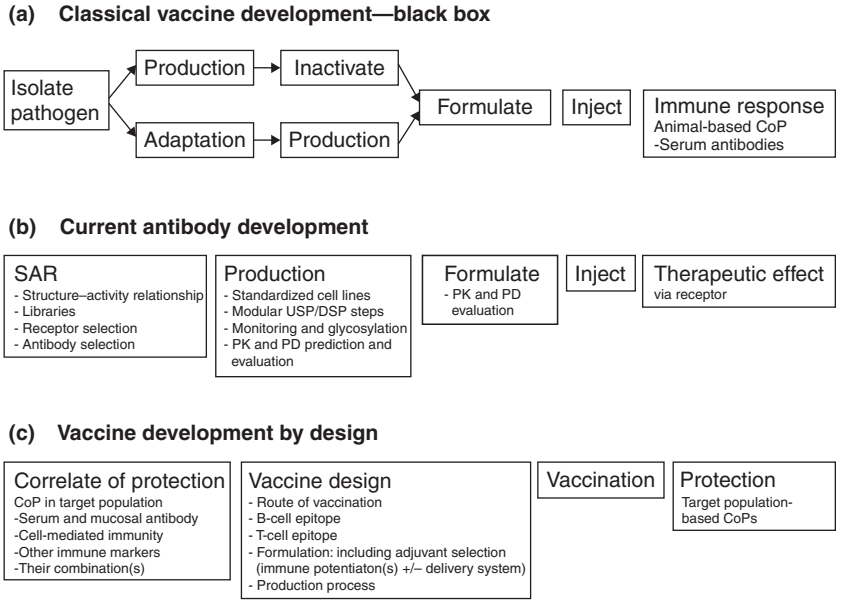


**FIGURE 13.1** Biopharmaceutical product and process development. (a) Process equals product and vaccine selection based on pathogen (+/− attenuation), upstream processing (USP), downstream processing (DSP), inactivation (not for life attenuated (LA)), fill and finish, and correlates of protection (CoP) by animal models. (b) Monoclonal antibody selection, receptor selection, structure–activity relationship (SAR), design of mAb or libraries, Exclude: . . . , USP, DSP, and biochemical equivalence. (c) Quality by design–based vaccine design. Host–pathogen interaction (e.g., reverse genetics), product target profile based on aimed CoPs (human), vaccine design, USP, DSP, fill and finish, relevant *in vitro* and *in vivo* models for CoP qualification (difficulty quantification), and fingerprint approaches (consistent production).

implementation of current guidelines highly challenging compared with biopharmaceuticals such as mAbs [9].

The classical process development for vaccines has been based on a trial-and-error approach based on the Pasteur principle of isolate, inactivate, and inject [10]. A modification of this early development includes the adaptation of strains to attenuated versions of the pathogen that do not cause the disease but do generate a protective immune response. Examples of attenuated strains for bacterial and viral vaccines against tuberculosis and polio are bacillus Calmette-Guérin (BCG) and Sabin polio virus, respectively. The development of the classical vaccines (DTP-P and mumps, measles, rubella [MMR]) resulted in a black box processes with limited scientific understanding (Fig. 13.1a). This was in the pre-GMP era with considerably lower requirements for process rationale and validation. The production processes for classical vaccines were implemented for genetically stable pathogens with the exception of influenza, which shows yearly seasonal changes [11].

The classical approach starts with the isolation of the pathogen and selection of methods to culture and store it in safe conditions. Then an animal model is used to

demonstrate that the chemical (formalin or β-propiolactone) inactivated pathogen as vaccine induces an antibody immune response and provides protection against the wild-type pathogen in a challenge study. A feasible small-scale production process is developed to produce the inactivated pathogen, and a practical approach is used for scale up for market volume, confirmed by animal studies. Because a vaccine dose contains about 1000-fold less product than required for treatment with therapeutic recombinant proteins, the required increase in production volume and scale is usually smaller.

For bacterial vaccines, a pathogen strain has been cultured and inactivated to produce a "whole-cell" vaccine (e.g., *Bordetella pertussis*), an attenuated bacterium has been used (BCG), or pathogen bacterial strains have been cultured to produce purified and inactivated toxins or virulence factors (*Corynebacterium diphtheriae*, *Clostridium tetani*, *B. pertussis*).

For viral vaccines, usually adherent primary cells were cultured (rat, chicken cells for MMR, eggs for influenza, monkey cells for polio) in glass (Roux) bottles or plastic roller bottles, on membrane plates, and in bioreactors when microcarriers made it possible [12].

Probably because of their proven great beneficial effect for public health and the existing vaccine surveillance programs, monitoring the effect of released vaccines the regulatory agencies have not demanded upgrade of classical vaccine manufacturing to current guidelines. In fact, probably because of the rise of regenerative medicine products (which, besides most vaccines, form another category of "less well-defined" products), the most recent ICH Q11 guideline [7] provides or offers a choice in approach between "traditional" and "enhanced" (with more scientific understanding of product and process) vaccine development approaches.

As opposed to the development of product quality features (critical quality attributes [CQAs]) confirming the process feasibility criteria (design space) for well-characterized biopharmaceuticals, the vaccines, as less defined products, adopted the "process equals product" paradigm [13,14].

The black box status presents the risk that when a pathogen that is considered to be stable will change over decades the existing vaccines will become less effective. This should be observed by the surveillance programs, as seem to be the case for *B. pertussis* [15–17], mumps [18,19], and measles [20–22].

Another risk is that the black box production process cannot be kept constant during half a century of manufacturing, for instance, because of obligate changes in suppliers of raw materials, disposables, and manufacturing equipment. Over the years, minor changes that have no effect on the product efficacy that are accepted could add up to larger process changes yielding a vaccine product with lower efficacy than the original product. This cannot be demonstrated with actual CQA (Table 13.2) of product quality but has to be deduced from the often nonaccurate animal models used as read out for a correlate of protection (antibody titers, pathogen-neutralizing serum titers) and the postmarketing surveillance monitoring.

Many vaccine manufacturers of classical vaccines have already upgraded their production processes by introducing improvements such as the application of serum-free media and the change from primary cells to immortal adherent cell lines.

**TABLE 13.2 The Quality by Design Roadmap for General Pharmaceutical Product Development and Vaccine Development**

| | General | Vaccine |
|---|---|---|
| Quality target product profile (QTPP) | • Desired product performance, intended use, route of administration, dosage form, and delivery system<br>• Dosage strength (general: milligrams per dose)<br>• Container closure system<br>• Therapeutic moiety release or delivery and attributes affecting pharmacokinetic characteristics<br>• Drug product quality criteria (e.g., sterility, purity, stability, drug release) | • Induction of protection against pathogen: correlate of protection, no side effects in the healthy population, safety, intended use, target population, route of administration, dosage form and delivery system<br>• Dosage strength (general: $<15$ μg/dose; $10^{5-7}$ particles/dose)<br>• Container closure system<br>• Presentation form or adjuvant (immunopotentiator or delivery system) affecting immune responses<br>• Drug product quality criteria (e.g., sterility, purity (e.g., HCP, DNA), stability) |
| Critical quality attributes (CQAs)<br>Risk assessment: linking material attributes and process parameters to drug product CQAs | • Series of unit operation<br>• Physical, chemical, or microbiological properties or characteristics of an input or output material is defined as an attribute<br>• Process parameters include type of equipment, setting, batch size, operating conditions (e.g., temperature, pH, time, speed) and environmental conditions<br>• The quality and properties of drug substances and excipients are considered attributes of raw materials<br>• Identification of the critical raw material attributes, process parameters, and quality attributes<br>• CQA-desired product performance, intended use, route | • USP, DSP, fill and finish<br>• Media, cell lines, virus seeds, bacterial seeds, excipients, output characteristics<br>• Process parameters include type of equipment, setting, batch size, operating conditions (e.g., temperature, pH, time, speed), and environmental conditions<br>• The quality and properties of cell lines, virus seeds, bacterial seeds, and excipients are considered attributes of raw materials<br>• Identification of the critical raw material attributes, process parameters, and quality attributes<br>• CQA-desired product performance, intended use, route |

| | | |
|---|---|---|
| Design space | • CQA controlled directly or indirectly (input from materials)<br>• Investigation of critical process parameters (CPPs): process inputs that have direct impact on CQA<br>• Definition of design space: acceptable input for robust product quality, acceptable range of the CQA | • CQA controlled directly or indirectly (input from materials)<br>• Investigation of CPPs: process inputs that have direct impact on CQA<br>• Definition of design space: acceptable input for robust product quality, acceptable range of the CQA |
| Control strategy | • Definition of input for robust product quality, acceptable range of the CQA, and therewith compliant with QTPP | • Definition of input for robust product quality, acceptable range of the CQA, and therewith compliant with QTPP |
| Product life cycle management and continual improvement | • Product life cycle management and continual improvement | • Product life cycle management and continual improvement |

HCP: Host cell proteins.

Vaccine development and the vaccine market have seen a revival since the introduction of the pneumococcal conjugate vaccine Prevnar, illustrating that higher cost processes for vaccines are feasible for the developed world. Other factors that have contributed include the threats of pandemics and bioterrorism, development of therapeutic vaccines, and the support of institutions such as the Bill & Melinda Gates Foundation for vaccination in developing countries.

With the introduction of new technologies (Table 13.3), such as conjugate vaccines for bacterial pathogens and VLP vaccine concepts for viral pathogens, more targeted approaches for specific pathogen groups are emerging.

## 13.2 CURRENT VACCINE DEVELOPMENT (PRODUCT AND PROCESS)

For vaccine development, there seem to be two basic strategies: the classical strategy close to the pathogen (isolate, inactivate, and inject) and an alternative strategy to build a vaccine from individual components.

Staying close to the (whole) pathogen, producing an inactivated pathogen product or a less pathogenic attenuated strain (either obtained by natural selection or recombinant method), has a potential benefit of incorporating several unknown epitopes that will support either specific or nonspecific (adjuvants) immune response. The benefit of such potentially broad immunogenic responses might be masked by evasive mechanisms of the pathogen, such as high variability or the presence of components that interfere with the normal immune response.

The other option is to develop a vaccine based on building blocks. Such a vaccine may consist of (recombinant) antigens (e.g., proteins, saccharides, or both) that are formulated as a particulate vaccine. The build vaccine product consisting of (mainly) known components will have a (mainly) known structure–function relationship: each component in combination with the formulation will induce known immune activities to elicit the target immune response defined by the correlates of protection in the target population. This option would be open for fully characterized vaccine products as mentioned earlier and could facilitate a QbD approach for vaccine development.

The current vaccine development seems to favor a "best of both worlds" or hybrid strategy. These options hold that additional information is gathered during the vaccine development trajectory (from literature, "omics," previous projects, or guidelines). Such complex (vaccine) development trajectories ask for an integrated project approach (or multidisciplinary project approach).

The development of a vaccine provides a costly, long-lasting, and complex challenge. To deal with such a complex challenge that requires input from several experts to give target specifications for process and product, the integrated project approach is warranted. In this integrated approach, all of the relevant experts are included in the project team to collect the target criteria and to translate it in a plan with specific criteria for milestone achievements for the different phases of the development.

A vaccine development can fail because of various reasons, such as economic (unacceptable high cost of goods), regulatory (insufficient quality level),

**TABLE 13.3 Classical and New, Improved Vaccine Technologies for Vaccine Development**

| Technologies | Disease | Pathogen | Vaccine |
|---|---|---|---|
| | | *Classical* | |
| Isolate and inactive | Diphtheria | Diphtheria bacteria | Diphtheria toxoid* |
| | Tetanus | Bacillus tetani | Tetanus toxoid* |
| | Polio | Polio virus (3 subtypes) | IPV (Salk) |
| | Whooping cough | Pertussis bacteria | Cellular pertussis vaccine |
| Isolate and adapt | Measles | Measle virus | MMR |
| | Mumps | Mumps virus | |
| | Rubella | Rubella virus | |
| | Polio | Polio virus (three subtypes) | OPV (Sabin) |
| | | *New and Improved* | |
| Cell culture | Hepatitis A | | Hepatitis A |
| | Rotavirus | | Oral live rotavirus |
| | Epidemic or pandemic flu | Influenza | LAIV |
| | | | Subunit or split |
| Conjugation | *Haemophilus* | *Haemophilus influenzae* type B | HiB |
| | Meningitis | Meningococcus C | Men C |
| | Meningitis | Meningococcus ACWY | Men ACWY |
| | Pneumococcal-induced pneumonia | Pneumococcus type 7 | PnC-7 |
| | | Pneumococcus type 10 | PnC-10 |
| | | Pneumococcus type 13 | PnC-13 |
| Recombinant or VLP | Hepatitis | HBV | HBsAg* |
| | HPV-induced cancer | HPV cancer | HPV* |
| | Whooping cough | Pertussis bacteria | Acellular pertussis vaccine |

*Vaccine is formulated by adsorption on aluminum particles, alum particles act as adjuvant ($AlOH_3$ or $AlPO_4$)
HBsAg: hepatitis B surface antigen; HBV: hepatitis B virus; HiB: *Haemophilus* influenza type B; HPV: human papilloma virus; IPV: inactivated polio virus vaccine; LAIV: live-attenuated influenza vaccine; Men: Meningococcus; MMR: measles, mumps, rubella; OPV: oral polio virus vaccine; PnC: Pneumococcus; VLP: virus-like particle.

technological (process unreliable, insufficient scientific understanding for appropriate control; instability of product, raw materials, producing organism), or failure in clinical trials (insufficient level of predictability of assays for efficacy). Therefore, it is important that the experts of the project team have final responsibility for their expertise in the development project. It is important that the initial development plan contains all of the criteria from the different kinds of expertise that reflect the best feasible compromise (from current knowledge at that time) and that the project team can reach a feasible compromise on how to proceed when certain criteria linked to milestone achievements are not met. This by accepting a lower performance of product or process than given in the target specifications or invest extra effort in meeting the criterion (or presenting this as advice to higher management).

### 13.2.1 Product Development

The development of a new vaccine product starts with a disease caused by a pathogen that generates an unacceptable disease burden (identifying a therapeutic application for which the immune system can help to fight it). Based on the type of pathogen, knowledge on pathogen–host interaction, knowledge from previous comparable development projects, and the current literature, criteria for the product are defined. The level of detail depends on the existing knowledge on the required immune responses to induce protection; most vaccines primarily aim at eliciting an antibody response, although T-cell–mediated immune responses receive increasing attention.

With existing knowledge, several topics have to be addressed to define the product criteria [9,23,24]. These include the target population that should be vaccinated (e.g., infants, adults, elderly adults, subjects with an underlying disease), the required immune response (systemic or mucosal humoral immune responses, the extent and type of cell-mediated immune responses), the route of administration, the optimal (adjuvanted) vaccine formulation, the optimal vaccination schema (one, two, or more doses at specific time points or after specific periods), the regulatory boundaries (new product or second generation), the known appearance of immune evasion by the pathogen, known or expected safety problems (e.g., vaccine-induced enhanced disease as shown in the 1970s for formalin-inactivated respiratory syncytial virus [RSV]).

Unfortunately, to date, existing knowledge on host–pathogen interactions and the requirements for the optimal immune responses (defined by the correlates of protection in the target population) are too limited to allow a full QbD approach (no defined structure–function relationship).

In general, such concept criteria for the vaccine product [9,23,24] can be given by:

1. Repetitive molecular structures, epitopes that evoke a specific immune response against a pathogen:
   a. B-cell epitopes responsible for humoral antibody responses (adaptive)
   b. T-cell epitopes responsible for T-cell–mediated response (innate; that promote adaptive responses)

2. Pathogen-associated molecular patterns (PAMPs), repetitive molecular structures that evoke a supportive general (innate) immune stimulation effect
3. Do not contain molecular structures that block, reduce, or overstimulate the desired immune response (e.g., unwanted T-help (Th) skewing toward Th2 as is the case of many aluminum-adjuvanted vaccines)
4. Structural features: size, correct presentation of (repetitive) epitopes, stable configuration (chemical and biological stability), correct transport for target tissue
5. Additional safety besides point 3: standard, sterile, no viruses, DNA, endotoxin, HCP, leachables
6. Features that support proper delivery, such as powders with a defined particle size for nasal or pulmonary delivery; formulations with mucoadhesive properties for longer nasal residence time; formulations (coated tablets or antiacid co-delivery) that guarantee integrity of the vaccine during passage of the stomach during oral delivery

### 13.2.2 Product Assays, Correlates of Protection

Process development following QbD principles requires fast, accurate assays to determine product quality and quantity to explore and confirm the design space for a target process. There is considerable discussion on what parameters give appropriate correlates of protection [25] and which fast biochemical assays are good indicators of product potency and can predict performance in animals and in humans [26]. The latter is particularly difficult because there is often discrepancy between results obtained with specific animal models and clinical results in humans. Moreover, similar to humans, animal models give rise to a high level of variation and as such are not good indicators for consistency of the final process and product.

### 13.2.3 Process Development

After definition of the product (and intermediate products) and subsequent specifications, a suitable expression systems can be selected that meets the product criteria (for the whole product or for the different subunits). Emphasis in the QbD approach is on product quality attributes, but for economic reasons, the potential yield is relevant because it will determine part of the cost price. For risk reduction, many pharmaceutical companies favor a proven expression system for a comparable new product.

The different phases include:

- The initial process to support product development, part I (will this process or mainly the used expression system provide a feasible process to meet product criteria: and the reverse, can an alternative expression system with proven performance for other vaccines meet the criteria for product quality?)
- A feasible small-scale target process
- Scale up to pilot, prevalidation studies and reproducibility

- Production for toxicology, clinical phases I and II
- Scale up to market scale, validation (mix down scale + consistency runs)
- Clinical phase III and market launch

***13.2.3.1 Selection of a Suitable Expression System*** For a "close to the pathogen" product, the expression system for the vaccine is usually the pathogen or modified pathogen itself. This means culturing the pathogenic bacteria for bacterial vaccines and culturing viruses in animal cells.

For whole inactivated vaccines such as for whooping cough and for live attenuated vaccine as for tuberculosis (TBC), the culturing of the bacterial pathogens (*B. pertussis* and BCG, respectively) is obligatory, but also for the production of purified inactivated toxins such as for tetanus and diphtheria, the upstream process makes use of the original pathogens from *C. tetani* and *C. diphtheriae*, and no shift has been made to produce these toxins as recombinant in the standard bacterial host *Escherichia coli*. For the conjugate vaccines (*Haemophilus influenzae* type b; pneumococcus; meningococcus A, C, W), the capsular sugar components are derived from the target pathogen, and the carrier protein is the standard tetanus or diphtheria toxoid, the CRM protein from a diphtheria mutant, or OmpB membrane protein from *Neissera meningitidis*. For the viral vaccines as close to pathogen products, replication in mammalian (human) cells is usually required. For original vaccines, animals were used as expression system, such as pox virus in cattle and rabies in rabbits. Even today for some applications of viral vaccines (Japanese encephalitis, Crimean Congo hemorrhagic fever), for fast-track development and manufacturing, mouse brain is used for virus replication. Influenza is traditionally cultured in eggs because of its broad replication ability in mammalian (horse, human) and avian (duck, chicken) cells.

From animals, the culture of viruses went to primary cells and later derived cells with extended but still finite life spans (tertiary monkey cells, human diploid lung cells as MRC-5 and WI-38). From these mortal cells, the developments shifted to immortal cells such as Vero (African monkey) and MDCK (Madin-Darby canine kidney) that are cultured as adherent cells that still have contact inhibition to stop growth at confluency. For many viruses, it was also thought that the orientation of an attached (epithelial) cell was necessary to obtain an appropriate quality and quantity of virus. Vero cells have been approved for polio, rabies, Japanese encephalitis, rotavirus, vaccinia, and influenza. MDCK cells are used for approved influenza with both adherent and suspension cells. Concerns with immortal cells are the risk of tumorigenicity or carcinogenicity and the presence of nonproduct viruses because for a viral vaccine product, it is difficult to reduce contaminating DNA and RNA to a very low level during inactivation of purification without damaging or losing the viral product [27].

The first recombinant vaccine, which is for hepatitis B virus (HBV), was approved for the market in 1986 by Merck Sharp & Dohme (MSD) only 5 years after the introduction of the first recombinant biopharmaceutical, insulin. Although insulin was followed by a stream of other recombinant protein biopharmaceuticals (e.g., tissue plasminogen activator, EPO, mAbs, human growth factor hGF), it took until

1998 until the next recombinant vaccine (against Lyme disease) was launched (by Glaxo Smith Kline [GSK]). Because this product was withdrawn from the market, the only approved recombinant vaccines besides that for HBV are the human papillomavirus (HPV) vaccines (MSD, GSK).

The used expression systems for these VLPs are yeast cells (MSD, *Saccharomyces cerevisiae*, HBV, HPV) and insect cells (GSK, HPV). The current VLP-based vaccines make use of adjuvants based on aluminum and/or monophosporyl lipid A or squalene-based oil-in-water emulsion [9]. Because vaccines often require membrane proteins (or other membrane components), it will be interesting to follow whether the same standard expression systems will be used as for the soluble recombinant protein products (*E. coli*, *Saccharomyces*, *Pichia*, Chinese hamster ovary [CHO]).

***13.2.3.2 Process Development Target Upstream Process*** The selection of an expression system largely determines the upstream process. Despite the broad variation in expression systems for vaccines, there are also a number of common features for the upstream process for vaccine manufacturing.

It starts with the seed lot system: a master and working seed lot(s) to start an expansion of bacteria, cells, or virus. For the preculture expansion, many different—usually disposable—systems are used: agar plates, culture flasks, shake flasks, roller bottles, spinner flasks, plastic bags, and membrane plate systems. The expansion culture is continued until a sufficient volume has been generated for the start of the common upstream production unit, the stirred tank reactor (but different operation regimens exist with respect to medium feed, including batch, fed batch, continuous stirred tank reactor (CSTR) retention of cells (perfusion equipment), stirring [type and number of impellers with and without vortex], pH control, and aeration [headspace, sparging]).

During this bioreactor process, a crude product is generated that can be used for further concentration, purification, and inactivation during the DSP.

***13.2.3.3 Process Development: Downstream Processing*** For the classical (whole) modified pathogen vaccines the DSP is limited to concentration and inactivation of product. With the rise of subunit vaccine products the purification of the subunits and the formation of a complete particle product generates considerable more work for the DSP and formulation groups. Standard purification techniques, used for vaccines and biopharmaceuticals in general, to purify product include centrifugation, filtration (dead-end and tangential flow), precipitation, and column chromatography (SEC, IEC, HIC, IAC). As for the upstream processing, the DSP of vaccines from pathogen disruption until final formulation is more characterized by the different methods than by accepted standardization.

### 13.2.4 Formulation and Vaccine Delivery

The development of a vaccine is highly dependent on tailored vaccine delivery approaches to guarantee broad and strong B-cell and T-cell responses. The design of tailored vaccine delivery approaches has several needs and challenges, such as

restrictions posed by specific target populations, potentials of dedicated stable formulations and devices, and the use of adjuvants. These needs and challenges of tailored vaccine delivery approaches, including the current status and perspectives of vaccine delivery via several routes of administration, are discussed in detail by Amorij et al. [9].

Ideally, a vaccine is a simple and stable formulation that is suitable for convenient administration and provides lifelong immunity against the pathogen. Unfortunately, in the traditional vaccine approach, there was little focus on DSP and formulation development other than concentration of the inactivated or attenuated pathogen.

For most classical whole-cell and whole-virus vaccines, proper products have been delivered because the pathogens are stable and lack immune evasive properties. There was little attention on purification and formulation because the "black box" products worked.

Live-attenuated or recombinant whole-pathogen approaches can be attractive as a fast alternative approach (for the wild type in high containment), but the bonus of having several unknown antigens causing a wanted specific and nonspecific supportive immune response might be accompanied by the presence of molecular structures reducing or exaggerating this response (RSV, severe acute respiratory syndrome [SARS]).

The testing of different adjuvants was limited to testing the effect after mixing of different concentrations with the vaccine product (modified pathogen of purified subunit). Probably the same for combination vaccines: mix and test (in animals).

Despite the successes of the classical vaccines, there is still a number of infectious diseases for which no effective vaccine is available (e.g., HIV, malaria) or for which existing vaccines provide insufficient immunity (e.g., tuberculosis, whooping cough) or are unaffordable for those most in need (e.g., pneumococcal disease). In addition, not only the protective immune response induced by the vaccine determines it success but also the willingness of the public to be vaccinated as well as the availability of vaccine and trained personal to administer of the vaccine. Attention to safety has resulted in the development of safer vaccines (subunit vaccines, which in general are less immunogenic), adjuvants to enhance efficacy and facilitate dose sparing, and the exploration of noninvasive administration approaches.

Key aspects for formulation and delivery that have to be considered in vaccine design are [9]:

- The increase in knowledge (as indicated earlier): Knowledge on identification of immune mechanisms and the definition of clear CoPs and development of robust assays for their determination
- Identification of the optimal delivery of the antigen
  - Specific sites in the body
  - Delivery of the antigen to and activation of relevant immune cells
- The success of (recombinant) subunit vaccines (conjugate and VLP) recombinant proteins: Subunit vaccines are less immunogenic (need adjuvants) and are as more pure product less stable (degradation, aggregation)

- Induction of the desired immune response: Depending on the intrinsic immunogenicity of the antigen and the intended route of administration, an adjuvant (delivery system, immune potentiator, or both) may be selected to steer the magnitude and quality of the innate immune response and the uptake and processing of the antigens by antigen-presenting cells
  - Delivery systems, including nanoparticles, microspheres, liposomes, and VLPs
  - Immune potentiators or modulators, which are compounds that directly activate immune cells through specific receptors or pathways
- Proper formulation of the key component, antigen, and adjuvant in a device or container depending on the intended route and storage conditions
- Design of combination vaccines to minimize the risk of decreased willingness to be vaccinated as a result of increased numbers of injections and bear the risk of pharmaceutical and immunological interference
- Cold chain problems in the developing countries (live vaccines often need to be lyophilized, but with the introduction of subunit vaccines, it became feasible to have stable liquid formulations)
- The increasing number of vaccines in pediatric programs and the wish to combine them (the challenge of formulation of combination vaccines is pharmaceutical and immunologic interference)
- Route of administration: Circumvent the use of needles and specific delivery to specific tissues to increase effectiveness

Better vaccines will be designed in the near future when knowledge on CoPs, optimization of vaccine delivery by administration route, delivery system and immune modulator, and stable formulations are integrated.

## 13.3 SPECIFIC RELEVANT TRENDS

### 13.3.1 Regulatory Quality by Design and Process Analytical Technology

***13.3.1.1 Quality by Design*** The ultimate design of vaccines includes QbD. QbD requires a thorough scientific understanding of all of the CQAs of the vaccine product in order to follow the QbD roadmap for pharmaceutical product development (see Table 13.3). To date, many challenges for vaccine development have to be faced before the first vaccine is fully designed by QbD. These challenges include unknown host–pathogen interactions; definition of *all* CoPs; and understanding of potential side effects of the final product, including vaccine-related PK and PD studies.

For a number of subunit vaccines that go in the direction of "well-characterized product" that have been approved for the market, a QbD approach should become feasible. In this category belong VLPs (HBV, HPV) and conjugate vaccines (Hib, pneumococcus, meningitis). Although a full QbD approach for vaccines is not

possible today by an overall design of the product and processes, an approach that is closest to full QbD is the application of QbD on the several unit steps of the production process. This approach, QbD for nonplatform products and processes typical for vaccines, has been described in the A-VAX case study by the CMC-Vaccines working group (comparable to the A-mAb case) [8]. They described a "toward QbD" approach for development of a fictive vaccine, including USP for (inactivated) polysaccharide, VLP and live virus production, DSP steps, polysaccharide–VLP conjugation, formulation based on lyophilization, and aluminum adjuvant adsorption.

***13.3.1.2 Process Analytical Technology*** The PAT initiative [28] was introduced by the US Regulatory Authority of the FDA to promote scientific understanding of the manufacturing process to improve process control and facilitate fast release based on online monitoring. It was written for the pharmaceutical industry in general, so it is also for chemicals and small molecules as products or for the release of water, raw materials, buffers and culture media, and sterilized items in an autoclave.

Process analytical technology uses QbD as a basal approach. To successfully apply PAT and to be able to indicate the critical process parameters per unit process, CQAs must describe the correct product quality (Table 13.4). The last step for PAT is to identify an online signal that adequately reflects the other critical process parameters within a certain used range for this online measured parameter so that within this range, a product with correct predefined quality is always obtained (to be demonstrated in validation studies).

Main benefits that are usually given for applying PAT include better process understanding, shorter release time between process steps, and to the ability to shift a process in used set points and criteria for specific parameters to other values after approval of the process within the given design space for these parameters.

Because demonstrating successful PAT application requires specific CQA for the (intermediate) product and accurate assays to demonstrate this, the standard use of PAT is limited to well-characterized products. For well-characterized products with fast assays for product quality, it is questionable whether PAT application provides an extra benefit over just applying QbD (see Table 13.4).

For less well-characterized products, it is important to generate a good database for process development with all the process specifications that support a specific outcome such as animal (toxicology) studies and clinical studies. In this case, a design space is actually confirmed by the variation within specs of the product, which is monitored in postmarketing surveillance. If an omics assay program, including genomics or microarray, proteomics (two-dimensional gel electrophoreses (GE), MS), and metabolomics testing, has supported the process development, the outcome of the animal and clinical testing and postmarketing surveillance can be coupled to specific information to generate new CQAs in a parallel or retrospective quality approach. Results can then be used for future process changes (see Table 13.4).

*Process Analytical Technology Case Study* An example of implementation of PAT for a human vaccine process is illustrated by the use of an online near infra red (NIR)

**TABLE 13.4 Main Features of Process Analytical Technology and Quality by Design for Process Development and Their Relation**

| Process Analytical Technology | Quality by Design |
|---|---|
| Online measurement to get fast release of product (intermediate, bulk, final) or raw material, medium, or buffer | Product design approach providing criteria based on scientific understanding<br>Design = prospective approach |
| Usually applied per unit process (step) | Usually applied for the whole product development from knowledge on a disease to a therapeutic hypothesis for cure or prevention to product design (clear structure–function relationship) |
| Online signal should reflect all other critical process parameters | Risk reduction based on knowledge; gives defined criteria per milestone for product and process development |
| Critical quality attributes are required to demonstrate that the chosen online parameters indeed reflect the other relevant parameters within the defined design space → scientific understanding of the process | Definition of control strategy using critical process parameters within design space |
| Could be done as a parallel or concomitant approach | Product quality is guaranteed by design space "product release by design" |
| Primarily process database | Uses (new) knowledge of the disease (host pathogen interaction) and relevant information of product and process from comparable products (PK and PD in humans; product stability) |

PD: pharmacodynamics; PK: pharmacokinetics.

signal for the fermentation step of *Bordetella pertussis* [29]. Here the metabolic status of the bacteria is coupled to the production of the gene cluster containing the relevant virulence factors as demonstrated with array technology. The production of virulence factors serves as a marker for product quality that could be linked to the NiR signal.

### 13.3.2 Comparison with Monoclonal Antibodies

After the discovery of Kohler and Milstein to produce mAbs with murine hybridoma cells, a boost in therapeutic drugs was expected [31]. However, the first generation of murine antibody products failed for therapeutic applications because of an unwanted human immune response against the mouse product and because the role of glycosylation of the antibody was not recognized, causing reduced clinical performance such as rapid clearance from the blood of these intravenously added therapeutic antibodies.

The expression systems were modified to be able to generate fully human antibodies, and new assays were developed for screening the glycosylation pattern

and added to the concept product specifications, bill of testing. After this, the development of humanized antibodies led to the great success of the mAbs as group of therapeutic products [32]. mAbs form a homogeneous group of recombinant proteins. As a result, standardized target manufacturing processes as well as fast-track modular development routes for new mAb products have been established (Fig. 13.1b). The standard manufacturing process for an mAb includes an upstream expansion of suspension cells from the working cell bank using animal component–free medium and fed-batch fermentation in a large-scale stirred tank (10,000 l). The standard DSP contains tangential flow filtration (TFF) or centrifugation as primary recovery followed by protein A and ion exchange chromatography and dead-end filtration for removal of virus and sterile final filtration step. The current standard host cell is the CHO cell line; a few other mammalian-derived suspension cells (hybridoma, myeloma) fit in comparable processes.

The better scientific understanding of essential product features and the experiences from repeated process development for comparable mAb projects have led to better risk management, better process control, and more freedom to operate with respect to process changes. Extensive scientific understanding is built in as a result of extensive experience (and data) with the standard expression system and the well-defined target production process. This in combination with detailed characterization of antibody products that allow demonstration of biochemical equivalence of mAbs (including their impurity profile) supports that a process change will not give rise to increased safety risks. Together with the use of a standard expression system—CHO cells on a standard animal component–free medium—with a target production process that was approved for another mAb product supports that the impurity profile of the mAb product will not cause increased safety risk after the process change.

This freedom to operate for process changes has been facilitated by the regulatory authorities such as the FDA. The FDA's point to consider on mAb manufacture and testing [33] includes a section on demonstrating product comparability related to manufacturing changes. If convincing *in vitro* evidence of product comparability (biochemical equivalence) can be provided in combination with data that animal testing reveals no differences, extensive clinical data demonstrating equivalence may not be necessary. The Guidance for Comparability states: "The manufacturer's ability to establish sensitive and validated assays for characterizing the product and biological activity and to evaluate the significance of differences noted in such cases can provide the basis for FDA to assess product comparability without the necessity of repeating clinical efficacy studies" [3].

The extensive scientific understanding of mAb products and their production processes has led to companies specialized in development and manufacturing of new therapeutic mAbs. Most of those companies offer fast-track development of mAbs from (provided) mAb-encoding cDNA to production of the mAb product for clinical phase I trials within 1 year. Others make use of the freedom to operate by first using a transient expression system with insect cells or Human embryonic kidney (HEK) 293 cells for fast product screening and later transferring the product DNA to a CHO cell line.

The scientific understanding, as is the case for mAbs, helps to fulfill a regulatory requirement by giving the rationale why a product contains specific features or why a specific process step is performed in a specific manner. In addition, scientific understanding provides better product performance, improved process control, and databases for troubleshooting ability. This presents a general driving force to acquire the same level of understanding for vaccines as for most other biopharmaceuticals. Unfortunately, this is extremely challenging for vaccines because current knowledge on immunology of host–pathogen interaction and working mechanism of existing vaccines, especially the markers that predict protection by a vaccine (CoPs), is still evolving and limited. As a result, current vaccine development is not possible via full QbD (Fig. 13.1c) and builds on knowledge from general science and previous comparable projects. To add understanding of product and process, vaccine development applies on omics tools (explained later) and other PAT tools.

The fact that many pathogens are highly variable or have other ways to deceive the immune system prevents successful application of the classical Pasteur approach of injecting a whole inactivated pathogen as vaccine. For these pathogens, fundamental research is needed to elucidate the exact host–pathogen interaction in order to be able to "design" a suitable vaccine strategy or to screen vaccine candidates from a reverse vaccinology approach [11] (see Section 13.3.2.1). In recent years, considerable progress has been made in the understanding of the pathogen–host interaction, in particular with respect to the innate immune system (e.g., Toll-like receptor activation). It has become clear that all adaptive immune responses are initiated by the innate immune system and that the recognition of common microbial components by specific receptors is a first crucial step for the response. This has opened the way to a rational design of adjuvants [9,34–36].

***13.3.2.1 Omics*** The availability of high-throughput technologies to sequence genomes of pathogens made a holistic approach for vaccine development possible. This approach, indicated as reverse vaccinology [37], can be used to identify suitable epitopes or antigens for a vaccine. This has been applied for the development of a meningococcus B vaccine by Novartis. The *Neisseria* genome was sequenced and the structural genes suitable for production in *E. coli*, identified as genes encoding for outer membrane proteins, were selected for a after round of testing an immune response. For meningococcus B, the serum bactericidal assay is an accepted correlate of protection to evaluate a protective antibody response against *Neisseria meningitis* type B. Five proteins were selected to be incorporated in the final vaccine product. In recent years, this omics approach has been extended to the more scientific knowledge–based systems biology approach, also called systems vaccinology [38,39], which (preferentially) makes use of markers of the innate immune response in humans to predict the appropriate vaccine (protective) effect [39].

### 13.3.3 Enabling Technologies

A number of technologies have become available for the (general) development of biopharmaceuticals that enable the call for more scientific understanding and

reduction of risk while not causing extra overall development time. In this way, the development of vaccines adopts more generic methods and avoids the "tailor-made" development of specific technologies that is associated with the very long (>15 years) development time of several vaccines a decade ago.

***13.3.3.1 Disposable Technology*** The use of disposable technology can significantly reduce the investment in time and cost to select, implement, and validate stainless steel hardware.

The use of disposable technology is not new. For the freezing of cells in cryotubes; expansion of cells and bacteria in plates, shake flasks, roller bottles, and spinner flasks; and the use of membranes, hollow fibers, and other filters, many disposables items were already available. More recent trends include the use of bags on shaking platforms and disposable bioreactors for the expansion of larger culture volumes and the fermentation process. Also, for DSP, a trend is clear toward the use of disposable filter material.

***13.3.3.2 High-Throughput Technology*** There has been considerable development to create miniature systems to generate scientific understanding faster. This can include the screening for cell lines or particular culture media in multiwell systems or the exploration and confirmation of critical process parameters in downscaled multireactor systems (in parallel testing from one inoculum) or automated DSP systems (automated serial column settings). Usually, design of experiments is used in these multi-experiment systems to reduce the number of required experiments to generate sufficient scientific understanding to define the design space for a specific process step.

In addition, many high-throughput developments have been realized for assay applications facilitating process development (not further discussed here).

***13.3.3.3 Platform Technology*** Platform technology is technology that is already available to use because it has been developed for a group of products that are sufficiently comparable to allow a general approach. For fast development of a process for a new product, it is highly beneficial if there is already a target process using platform technology for several, if not, all unit process steps because it greatly reduces the required time for the process development, and the risk of failure is greatly reduced (provided the products are sufficiently similar).

If for a new product it is feasible to use an existing large-scale reactor for the fermentation process, it saves the time in selecting, purchasing, implementing, and validating the new bioreactor. In addition, there will exist a database on used conditions for several cell lines that can be used for the new expression system, and there is usually a proven downscaled bioreactor system that can be operated under comparable conditions as the large-scale bioreactor, which greatly supports the process development for this process step and reduces the risk for process transfer and scale up. Blueprints for process validation are available, and there is experience with the (nonspecific) in-process controls associated with operating the system. If the same expression system in standardized culture medium is used for a new

comparable product, it is great advantage in knowledge of the impurity profile associated with this production method.

For a product group such as mAbs, there are targeted or modular approaches that only use platform technology. This enables the development of a suitable cell line from cDNA containing the antigen in 3 to 6 months and the generation of material with a standardized process according to cGMP guidelines after another 6 months [7,28].

Risks of the use of platform technology include that different process steps are usually optimized to give an acceptable (feasible with other criteria) value not necessarily the optimal performance, and if a certain technique is used for too long, there might be new, better platform technology (applied by competitors). In this way, use (addiction) of platform technology can slow down improvements in process technology for vaccine development and manufacturing.

### 13.3.4 Emergency Vaccine

The national health agencies of all countries have preparedness for outbreaks of infectious diseases in their standard programs. The preparedness includes all measures to take during an outbreak to try to control the epidemic, such as stockpiling of syringes, antivirals, antibiotics, and potential cross-protective vaccines; reducing traveling; and closing schools. The fast development and production of a vaccine to control are often also included in the scenario [40].

This provides quite a challenge for vaccine development because a fast-track route with minimum risk and maximum predictability has to be provided, which cannot be given for vaccines in general. In the past, little outbreaks have been dealt with successfully by means of producing a vaccine, as tried for HIV and SARS. A successful example of outbreak control by a vaccine the application of a monovalent outer membrane vesicle (OMV) vaccine against meningitis type B in Cuba, Norway, France, and New Zealand.

Key elements for such an emergency vaccine route are:

1. Fast-track development, including proven platform technology
2. Proper correlates of protection
3. A fast-track release schedule, including risk assessment

Because there are few proven routes for vaccine development, a targeted approach for an emergency vaccine would include the traditional Pasteur principle of isolate, inactivate, and inject, ideally combined with QbD-oriented routes. At present, QbD-oriented routes for emergency vaccines against (pandemic) influenza have been applied successfully for related pathogens. In this targeted approach, only unit processes are improved that would block the development and manufacturing. For similar risk reduction for animal studies, several suitable models should be available for testing. Crucial is also a fast-track decision tree for approvals of vaccines by the government, possibly outside the normal regulatory authorities.

The optimum situation is to have a modular approach (fast-track scenario) for specific groups of pathogens that would give a high chance of success to deliver a vaccine fast for an outbreak of a related pathogen. Such a "tunnel" for vaccine development, regulatory approval, and manufacturing exist only for the classical variations of influenza virus.

It is important to update the strategies when new proven routes, such as the targeted approach for conjugate vaccines, have become available. Because the outbreak control is a public responsibility but vaccine manufacturing is a private enterprise, a proper interface between governmental agencies and (local) vaccine companies is essential.

## 13.4 FUTURE TRENDS FOR VACCINE DEVELOPMENT

An increase in vaccine development activities is expected for "the decade of vaccines" (Gates), driven by both public and private investments and promoted by the increased understanding of the immunology of vaccines. As is described in ICH Q11, it is expected that the classical route will be chosen for attenuated of inactivated modified (whole) pathogen vaccines in addition to the structural (or QbD-like) approach for well-characterized subunit vaccines. Both development routes will improve to reduce the risk of failure of the development project.

The classical approach is expected to improve by:

1. Partial characterization of the complex product (by omics)
2. Better selection of relevant animal models such as specific disease animal models that mimic human disease or show a more human-like immune responses to a pathogen
3. Improvement of human correlates of protection
4. A properly defined production process according to PAT to improve the definition of products that are not fully defined

The more current structural—or QbD-like—approach based on scientific understanding of product and process will further gain importance with the increase in knowledge of immunology and the findings from many vaccine developments (either successful or not).

It would also be beneficial to gain knowledge for future approaches by analyzing the way the classical vaccines work.

For successful application of a structural approach, endproduct criteria have to support the different aspects of the whole process (or design space, see figure 13.2) because the weakest (unknown) link will reduce the predictability of vaccine development. Therefore, an integrated approach using input from different relevant experts seems to be the best approach for vaccine development in the near future. In this way, vaccine development might be faster in the near future than pathogens are able to develop their modes of deceiving our immune systems.

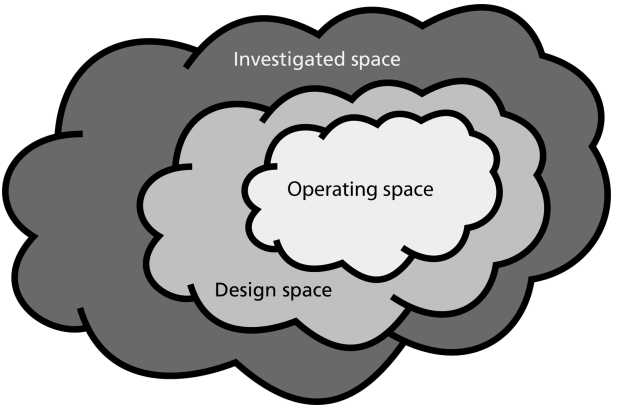


**FIGURE 13.2** The design space as a subspace of the investigated space (also known as the knowledge space). The operating space (where the process is intended to run) can be a large or small part of the design space. (Adapted from Ref. [30].)

## ACKNOWLEDGMENTS

This publication arises from the project FastVac, which has received funding from the European Union, in the framework of the Health programme. The authors thank Ing. Paul van Herpen for help with graphics.

## REFERENCES

1. Butler D. Translational research: crossing the valley of death. Nature 2008;453 (7197):840–842.
2. International Conference on Harmonization. Good Manufacturing Practice Guide for Active Pharmaceutical Ingredients Q7; 2000. Available from: http://wwwichorg/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Quality/Q7/Step4/Q7_Guidelinepdf.
3. Food and Drug Administration. FDA Guidance Concerning Demonstration of Comparability of Human Biological Products, Including Therapeutic Biotechnology-Derived Products; 1996. Available from: http://wwwfdagov/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/ucm122879htm.
4. International Conference on Harmonization. Pharmaceutical Development Q8; 2009. Available from: http://wwwichorg/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Quality/Q8_R1/Step4/Q8_R2_Guidelinepdf.
5. International Conference on Harmonization. Quality Risk Management Q9; 2005. Available from: http://wwwichorg/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Quality/Q9/Step4/Q9_Guidelinepdf.
6. Food and Drug Administration. Guidance for industry. PAT—A Framework for Innovative Pharmaceutical Development, Manufacturing, and Quality Assurance; 2004. Available from: http://www.fda.gov/downloads/drugs/guidancecomplianceregulatoryinformation/guidances/ucm070305.pdf.

7. International Conference on Harmonization. Development and Manufacturing of Drug Substances (Chemical Entities and Biotechnological/Biological Entities) Q11; 2012. Available from: http://wwwichorg/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Quality/Q11/Q11_Step_4pdf.
8. CMC-Vaccines, Working, Group. A-VAX: Applying Quality by Design to Vaccines; 2012. Available from: http://wwwpdaorg/Home-Page-Content/CMC-VWG-A-VAXaspx.
9. Amorij JP, Kersten GF, Saluja V, Tonnis WF, Hinrichs WL, Slutter B, et al. Towards tailored vaccine delivery: needs, challenges and perspectives. J Control Release 2012;161 (2):363–376.
10. Rappuoli R. Bridging the knowledge gaps in vaccine design. Nat Biotechnol 2007;25 (12):1361–1366.
11. Rappuoli R, Miller HI, Falkow S. Medicine. The intangible value of vaccination. Science 2002;297(5583):937–939.
12. Rader RA. (Re)defining biopharmaceutical. Nat Biotechnol 2008;26(7):743–751.
13. Buckland BC. The process development challenge for a new vaccine. Nat Med 2005;11 (4 Suppl):S16–S19.
14. Marshall V, Baylor NW. Food and Drug Administration regulation and evaluation of vaccines. Pediatrics 2011;127(Suppl 1):S23–S30.
15. de Greeff SC, de Melker HE, van Gageldonk PG, Schellekens JF, van der Klis FR, et al. Seroprevalence of pertussis in The Netherlands: evidence for increased circulation of *Bordetella pertussis*. PLoS One 2010;5(12):e14183.
16. Mooi FR. *Bordetella pertussis* and vaccination: the persistence of a genetically monomorphic pathogen. Infect Genet Evol 2010;10(1):36–49.
17. Sutherland JN, Chang C, Yoder SM, Rock MT, Maynard JA. Antibodies recognizing protective pertussis toxin epitopes are preferentially elicited by natural infection versus acellular immunization. Clin Vaccine Immunol 2011;18(6):954–962.
18. Hviid A, Rubin S, Muhlemann K. Mumps. Lancet 2008;371(9616):932–944.
19. Xu P, Li Z, Sun D, Lin Y, Wu J, Rota PA, et al. Rescue of wild-type mumps virus from a strain associated with recent outbreaks helps to define the role of the SH ORF in the pathogenesis of mumps virus. Virology 2011;417(1):126–136.
20. Leuridan E, Sabbe M, Van Damme P. Measles outbreak in Europe: susceptibility of infants too young to be immunized. Vaccine 2012;30(41):5905–5913.
21. Putz MM, Bouche FB, de Swart RL, Muller CP. Experimental vaccines against measles in a world of changing epidemiology. Int J Parasitol 2003;33(5-6):525–545.
22. Plemper RK, Snyder JP. Measles control—can measles virus inhibitors make a difference? Curr Opin Investig Drugs 2009;10(8):811–820.
23. Bachmann MF, Jennings GT. Vaccine delivery: a matter of size, geometry, kinetics and molecular patterns. Nat Rev Immunol 2010;10(11):787–796.
24. Zepp F. Principles of vaccine design—lessons from nature. Vaccine 2010;28(Suppl 3): C14–C24.
25. Plotkin SA. Correlates of protection induced by vaccination. Clin Vaccine Immunol 2010;17(7):1055–1065.
26. Metz B, van den Dobbelsteen G, van Els C, van der Gun J, Levels L, van der Pol L, et al. Quality-control issues and approaches in vaccine development. Expert Rev Vaccines 2009;8(2):227–238.

27. Hess RD, Weber F, Watson K, Schmitt S. Regulatory, biosafety and safety challenges for novel cells as substrates for human vaccines. Vaccine 2012;30(17):2715–2727.
28. CMC Biotech Working Group. A-Mab Case Study; 2009. Available from: http://wwwcasssorg/associations/9165/files/A-Mab_case_study_Version_2–1pdf.
29. Streefland M, Van Herpen PF, Van de Waterbeemd B, Van der Pol LA, Beuvery EC, Tramper J, et al. A practical approach for exploration and modeling of the design space of a bacterial vaccine cultivation process. Biotechnol Bioeng 2009;104(3):492–504.
30. Harms J, Wang X, Kim T, Yang X, Rathore AS. Defining process design space for biotech products: case study of *Pichia pastoris* fermentation. Biotechnol Prog 2008;24(3):655–662.
31. Kohler G, Milstein C. Continuous cultures of fused cells secreting antibody of predefined specificity. Nature 1975;256(5517):495–497.
32. Lindl T. Development of human monoclonal antibodies: a review. Cytotechnology 1996;21(3):183–193.
33. Food and Drug Administration. Points to Consider in the Manufacture and Testing of Therapeutic products for Human Use Derived from Transgenic Animals; 1995. Available from: http://wwwfdagov/downloads/BiologicsBloodVaccines/GuidanceComplianceRegulatoryInformation/OtherRecommendationsforManufacturers/UCM153306pdf.
34. Steeghs L, Keestra AM, van Mourik A, Uronen-Hansson H, van der Ley P, Callard R, et al. Differential activation of human and mouse Toll-like receptor 4 by the adjuvant candidate LpxL1 of *Neisseria meningitidis*. Infect Immun 2008;76(8):3801–3807.
35. van der Ley P, van den Dobbelsteen G. Next-generation outer membrane vesicle vaccines against *Neisseria meningitidis* based on nontoxic LPS mutants. Hum Vaccin 2011;7(8):886–890.
36. Fransen F, Boog CJ, van Putten JP, van der Ley P. Agonists of Toll-like receptors 3, 4, 7, and 9 are candidates for use as adjuvants in an outer membrane vaccine against *Neisseria meningitidis* serogroup B. Infect Immun 2007;75(12):5939–5946.
37. Rappuoli R, Black S, Lambert PH. Vaccine discovery and translation of new vaccine technology. Lancet 2011;378(9788):360–368.
38. Bernstein A, Pulendran B, Rappuoli R. Systems vaccinomics: the road ahead for vaccinology. OMICS 2011;15(9):529–531.
39. Rappuoli R, Aderem A. A 2020 vision for vaccines against HIV, tuberculosis and malaria. Nature 2011;473(7348):463–469.
40. Soborg C, Molbak K, Doherty TM, Ulleryd P, Brooks T, Coenen C, et al. Vaccines in a hurry. Vaccine 2009;27(25–26):3295–3298.

# Role and Application of Adjuvants and Delivery Systems in Vaccines

RAJESH K. GUPTA

## 14.1 INTRODUCTION[1]

Adjuvants have been described for use with vaccines for almost a century [1–4]. The word "adjuvant" is derived from the Latin word *adjuvare*, which means to help. Therefore, the term "adjuvant" has been used for any substance or material that helps vaccine antigens induce an early, high, broad (providing cross-protection against different variants or strains of the pathogen), or longer lasting immune response (Table 14.1) [5–18]. Adjuvanted vaccines often require less antigen to achieve a protective immune response, thus extending availability of vaccines to a larger population, which is important during pandemics and vaccine shortages. Although the mechanisms of most commonly evaluated adjuvants have not been fully understood, the adjuvant effect or adjuvanticity of these substances was considered to be (1) by modulation of immune system (e.g., activation of cells of the immune system); (2) by targeting of antigens to antigen-presenting or other immune cells; or (3) by controlled delivery of antigens, including slow-release or depot formation, at the site of injection. Because of the broad definition of adjuvants and the lack of a clear understanding of mechanism of action (MOA) for most commonly used adjuvants, many diverse substances that, when added to vaccine antigens, result in an increased immune response to vaccine antigens have been considered as adjuvants. The increase in immune response by certain substances could be attributable to stabilization of the antigen from degradation or by facilitating complete delivery of antigen to subjects by preventing antigens from being adsorbed onto the final container or syringe used for delivery of vaccines. Such increases in

[1] This chapter was submitted by the author, when he worked at the Food and Drug Administration (FDA). However, the findings and conclusions in this chapter are an informal communication and represent author's own views and judgment. These comments do not bind or obligate the Food and Drug Administration.

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*,
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

**TABLE 14.1 Role of Vaccine Adjuvants and Delivery Systems**

| |
|---|
| Enhance the immune response (higher antibody levels) |
| Develop an early and longer lasting protective immune response |
| Develop a broad immune response providing cross-protection against related strains |
| Develop protective immune response with lesser antigen, reducing the cost and increasing availability of vaccines to broader population |
| Facilitate development of combination vaccines by avoiding antigen competition, carrier specific epitope suppression, original antigenic sin, and so on |
| Modulate immune response to desired type, cellular (MHC class I) or humoral (MHC class II) |
| Modulate immune response to appropriate T-helper cells Th1, Th2, or other types to elicit appropriate response against a defined pathogen |
| Modulate antibody avidity, type (isotype, IgG subclass) |
| Elicit immune response in immunologically hyporesponsive populations such as elderly, immunocompromised, and pediatric populations |

MHC: major histocompatibility complex.

immune response should be considered as a formulation aspect in the development of vaccines and distinct from true adjuvant effect. Until the past two decades, not much attention was paid to the formulation during development and manufacture of vaccines, with the exception of the development of optimal formulations to maintain virus viability in live viral vaccines suitable for freezing or lyophilization.

As with most discoveries in medicine, use and development of adjuvants have been empirical without knowledge about the mechanism of adjuvanticity. Use of aluminum-based adjuvants in human vaccines started in the 1920s and 1930s as alum was originally used to partially purify protein antigens, mainly tetanus and diphtheria toxoids, by precipitating them in the presence of anions, including phosphate, sulfate, and bicarbonate ions, resulting in a mixture of compounds, mainly aluminum sulfate phosphate and aluminum hydroxyl phosphate [4,19–25]. Under specific conditions of pH and ionic strengths, these toxoids bound or adsorbed onto the aluminum compounds and were separated from other impurities. Over the years, several MOAs for aluminum-based adjuvants have been proposed, which is discussed later in this chapter. However, the role of aluminum-based adjuvants is still not clear.

Because of the importance of adjuvants in human vaccines established by the widespread use of vaccines with aluminum-based adjuvants and because of limitations of these adjuvants [6,19,20], there has been a search for alternate safe and potent adjuvants for more than 50 years. This became important and necessary with the development of purified, subunit, and synthetic vaccines, which are poor immunogens and require potent adjuvants to provide protective immune response. Furthermore, interest in development of new adjuvants grew because of their potential to selectively modulate the immune response to Th1 or Th2 types by directing antigen presentation through major histocompatibility complex (MHC) class I or MHC class II [7,26]. Selective modulation of immune response is

considered important for certain vaccines eluding successful development, for example, vaccines for protection against certain infectious diseases (e.g., respiratory syncytial virus, human immunodeficiency virus [HIV], malaria, tuberculosis), cancer, fertility, and allergenic and autoimmune diseases.

## 14.2 CHALLENGES IN DEVELOPING VACCINE ADJUVANTS

Despite advances in our knowledge about the immune response to infections, diseases and vaccination, and a better understanding in the MOA of various adjuvants, mineral compounds (mainly aluminum adjuvants) continued to be the only adjuvants used with human vaccines for almost a century. The development of adjuvants for human vaccines has been one of the slowest progresses in the history of medicine [27]. However, during past two decades, vaccines against influenza, hepatitis B virus (HBV), and human papilloma viruses (HPVs) have been approved in the Europe, the United States, and other countries with new adjuvants, such as oil-based emulsions and combination of oil-based emulsions, aluminum adjuvants, and monophosphoryl lipid A (MPL) [14,16,18].

There are a number of challenges in the development and use of adjuvants for human vaccines. The biggest problem with the use of adjuvants for routine vaccines are toxicity and adverse side effects after vaccination [6]. Because routine vaccines are given to healthy individuals, any side effect or even the perception of a side effect is not acceptable. An adjuvant for routine immunization of healthy infants, children, and adults must have a very low or no adverse side effects. Severe side effects, even those occurring rarely, would not be acceptable for routine vaccines unless there was a clear benefit against burden from disease and no other treatment alternatives are available for the indication for which the vaccine is given. When a new vaccine is introduced, there are often clear and visible benefits of a vaccine against morbidity and mortality of the disease (high risk). But as the vaccine continues to be used and controls the disease successfully (low risk of disease burden with the use of vaccine), the public forgets about the burden from the disease and remembers only side effects, even if they are rare and may not be caused by the vaccine. Thus, the success of a vaccine becomes an impediment for its continued use. There have been many examples of increase in disease burden with decrease in vaccination rates caused by adverse publicity against vaccines [28,29]. As our knowledge in immunology and vaccinology continues to grow, there is a need for continuous improvement in existing vaccines and adjuvants to stay ahead of the declining risk-to-benefit ratio for diseases controlled by vaccination.

At present, the choice of adjuvants for routine human vaccines reflects a compromise between a requirement for adjuvanticity and an acceptable low level of side effects. For high-risk groups such as patients with cancer or AIDS or for therapeutic vaccines, a certain level of toxicity may be acceptable based on benefits. However, the development of adjuvants for such applications has been challenging to get an appropriate immune response that could provide required protection or treatment for these diseases, which do not yet have successful vaccines. Other

problems in the development of adjuvants have already been discussed in several publications [7,10,13], which include restricted adjuvanticity of certain formulations to a few antigens, use of aluminum adjuvants as reference adjuvant preparations under suboptimal conditions, nonavailability of reliable animal models for most diseases, use of nonstandard assays, and biological differences between animal models and humans. These problems may lead to the failure of promising formulations to show adjuvanticity in clinical trials and delay advances in development of adjuvants.

## 14.3 VACCINE ADJUVANTS, ADJUVANT EFFECT, AND TOXICITY

Adjuvants belong to a heterogenous group of chemical compounds with diverse MOAs [5,7,8]. This is clearly evident from substances described as adjuvants for the first time in 1926 by Ramon showing increased antitoxin response to tetanus and diphtheria toxoids [2,3]. These substances included agar, tapioca, lecithin, starch oil, saponin, and even breadcrumbs. Major mechanisms of adjuvanticity are summarized in Table 14.2. However, it is not easy to categorize adjuvants on this basis because of multiple MOAs proposed for many adjuvants. Furthermore, because of the empirical development of vaccines and adjuvants with an emphasis on developing a formulation that works, several adjuvant preparations contain multiple substances to form complex formulations with an objective to elicit optimal immune response. For example, Freund's complete adjuvant (FCA) is an oil-based vehicle and contains *Mycobacterium* as a potent immunostimulator [30]; Syntex adjuvant formulation (SAF) contained threonyl muramyl dipeptide (MDP) as an immunostimulator and oil and pluronic block polymers as vehicles; certain liposome formulations contained encapsulated MDP, MPL, or cytokines and aluminum compounds; and GlaxoSmithKline (GSK)'s proprietary adjuvant systems (AS) contain combinations of aluminum compounds, oil-based components, and immunostimulators [14,18].

Because of a broad definition of adjuvants in helping to increase the immune response to an antigen, another aspect of adjuvant effect has been modification of

**TABLE 14.2 Major Mechanisms of Actions by Vaccine Adjuvants**

| Mechanism | Examples |
|---|---|
| Depot formation | Mineral compounds, oil-based adjuvants, liposomes. biodegradable polymer microspheres $>10\ \mu m$ |
| Delivery vehicles and targeting immune competent cells | Liposomes, oil adjuvants, biodegradable polymer microspheres $<10\ \mu m$, nanoparticles, non-ionic block polymer surfactants |
| Modulation of response (activation of immune-competent cells) | Lipid A, monophosphoryl lipid A, cPG, QS21, pertussis toxin, cytokines |

antigen in a controlled manner, resulting in an enhanced immunogenicity for that antigen. Modifications of antigens to increase their immunogenicity include (1) polymerization of protein antigens (tetanus toxoid and diphtheria toxoid) by glutaraldehyde treatment [31–33], (2) cationization of proteins (diphtheria toxoid) [34], and (3) conjugation of proteins to haptens and polysaccharides [7]. Primarily, conjugation is done to provide T-cell-dependent antigen characteristics to haptens and polysaccharides by proteins [35–37]. Carrier proteins, when coupled to T-cell-independent antigens, haptens, or weak immunogens, enhance the immunogenicity of the antigens coupled to these proteins. Carrier proteins usually act by providing T-cell epitopes to the coupled antigens or by increasing the molecular weight of the antigen. Carrier proteins are immunogenic to themselves, and it was observed that if the conjugation between polysaccharides and proteins is performed carefully without destroying important epitopes on a carrier protein, polysaccharides also enhance immunogenicity of the proteins [7]. This aspect is important in the development of combination vaccines to avoid problems with antigen overload and carrier-specific epitope suppression. As discussed earlier, any substances contributing to stabilization and protection of the antigen from degradation or preventing antigen from adsorption to the final container or syringe used for delivery of vaccines, thus ensuring complete delivery of antigen, may also have been considered an adjuvant. It is suggested that such enhanced immune response by modification or stabilization of antigens should not be considered an adjuvant effect.

During the past 50 years and even earlier, several adjuvant formulations were developed, and a few of them were evaluated in clinical trials. Most of them were never used for routine vaccines either because of lack of adjuvanticity or their toxicity and side effects. Because adjuvanticity is often associated with adverse side effects, a balance between these two attributes has to be maintained to obtain a defined or clear benefit for the use of an adjuvant. Some of the adverse effects are ascribed to mechanisms involved in the adjuvanticity of these formulations. For example, local reactions may be attributable to depot formation at the site of injection, which is one of the mechanisms of adjuvanticity of several adjuvant formulations (Table 14.2). Another mechanism of adjuvant action is stimulation of cells of the immune system to secrete various cytokines that may lead to systemic side effects depending on the type and magnitude of the cytokine response. Other side effects may be attributable to general adverse pharmacologic reactions of various formulations.

## 14.4 CHARACTERISTICS OF AN "OPTIMAL" ADJUVANT

With our current knowledge about the immune system, it seems that adjuvanticity and toxicity or side effects are integral parts of the adjuvant formulation. Depending on the required risk-to-benefit ratio, a balance between adjuvanticity and side effects can be optimized, particularly when using a rational design in development of vaccines. Theoretically, an ideal adjuvant without any side effects is desirable. Because of safety concerns of the adjuvant formulations for prophylactic vaccines,

particularly for routine immunization of infants and children, Edelman [5,10] listed a number of criteria to ensure the safety of adjuvanted vaccines. In addition to safety of adjuvanted vaccines for the absence of local reactions, systemic reactions (general toxicity, pyrogenicity), autoimmune diseases, hypersensitivity reactions, carcinogenicity, teratogenicity, and so on, an ideal adjuvant would be chemically defined so that it can be manufactured consistently. The preparation would elicit a protective immune response with a variety of weak antigens and would allow use of lower doses of antigens, achieving protection with fewer injections. The adjuvant would be effective in infants and young children, ideally at birth, and would also be effective in hyporesponsive population such as elderly and immunocompromised patients. The adjuvant would elicit a more persistent response of high quality (high affinity antibodies or desired IgG isotype) and breadth covering cross reacting strains of pathogens and of desired type (humoral, cellular, local, or systemic). The adjuvant would be stable with regard to adjuvanticity and toxicity with no or minimal interaction with the antigen. It would be biodegradable and nonimmunogenic to itself. However, there will never be a single adjuvant having all of these characteristics; however, there are possibilities to develop an adjuvant with the required characteristics based on risk-to-benefit ratio and striking a compromise or balance between desired and nondesired characteristics.

Even with the development of an optimal or ideal adjuvant with no toxicity and desired adjuvanticity, the adjuvant should only be used with vaccine antigen if it is required for effectiveness and safety of the product or provides a clear or distinct benefit. The need and benefits of using an adjuvant with a vaccine should be demonstrated in humans because animal models have not been good predictors of effectiveness of most human vaccines. This requires an extra group in clinical trials for the vaccine without an adjuvant. For traditional vaccines, aluminum adjuvants have been used by convention based on adjuvanticity in animals, sometimes without a clearly demonstrated enhanced immune response or protection in humans. It may be possible that aluminum adjuvants have been stabilizing the formulation, thus indirectly contributing to enhanced immune response.

## 14.5 REGULATORY ASPECTS FOR ADJUVANTS USED IN HUMAN VACCINES

Adjuvants are not licensed per se for use in humans; rather, they are approved as part of a vaccine formulation. However, it is generally mentioned or discussed at various meetings and in publications that aluminum-containing adjuvants have been the only adjuvants licensed by regulatory agencies. Adjuvants are not and should not be licensed alone for use with any vaccine because their physicochemical characteristics, such as size of particles and surface charge on the adjuvant, change after mixing antigen with or adsorbing antigen onto an adjuvant, particularly for aluminum-containing adjuvants. Such changes in physicochemical properties have the potential to affect the adjuvanticity and safety profile of the adjuvanted or adsorbed formulation. It is essential to know the chemistry,

manufacturing, and control aspects, including the toxicity, of an adjuvant intended for use with a vaccine. It may be of value to perform extensive toxicology and pharmacology studies on the adjuvant alone to learn about this material. However, such studies may not truly represent the toxicologic and pharmacologic profile of the vaccine formulated with adjuvant. Historically, toxicology studies performed typically for drugs have not been performed for vaccines because vaccines are given in minute amounts (microgram quantities) in only a few doses spread over several months. However, in recent years, typical toxicology studies performed for drugs have been recommended and required for vaccines. These studies may be useful for certain vaccines, adjuvants, or vaccine formulations, but there must be a scientific basis for performing extensive standard toxicology studies for vaccines. In a recent toxicology meeting, a concern about the immunotoxicity of adjuvanted vaccine formulation observed during toxicology studies was raised. This clearly shows a lack of scientific or regulatory understanding because vaccines and adjuvants are designed to modulate or elicit an immune response, when such response for a therapeutic drug is immunotoxicity, but is required for efficacy of the vaccine.

Historically, specific safety tests were designed and routinely performed at various stages of vaccine manufacture based on science; perceived risks; and the knowledge of the microorganism, attenuation history, inactivation, and so on. These tests were specifically designed to include susceptible animal species (e.g., guinea pigs for tetanus and diphtheria toxicity and detection of risk of tuberculosis for vaccines made from monkey cells, monkeys for neurovirulence test of oral polio vaccine, rabbits for detection of cercopithecoid herpesvirus 1, adult mice for detection of lymphocytic choriomeningitis virus, suckling mice for detection of *Coxsackie* viruses), or other systems (cell cultures from monkeys and humans to detect a number of simian and human viruses, rabbit kidney cells to detect herpes virus), appropriate doses (five human doses for residual toxins in tetanus and diphtheria vaccines), duration (6 weeks for specific toxicity test for diphtheria), appropriate stage of manufacture (adventitious agents and *Mycoplasma* testing on products made from cell cultures during early stage of manufacture, such as control cells and viral harvests), and route of inoculation (intraperitoneal, subcutaneous, intraspinal, intracerebral) rather than one or two standard animal species (mice, rats, rabbits) in an acute or chronic toxicity tests. Similar safety testing should be designed for the evaluation of novel adjuvants, including testing of raw materials, the adjuvant itself, and the formulated product. Testing should be based on perceived risks and science rather than only performing standard toxicology studies. If there is a scientific rationale to perform standard acute, chronic, reproductive, carcinogenic toxicology studies, these must be performed.

Because adjuvants are desired to be active components in a vaccine formulation, these should be manufactured and characterized like an active component. Dose of vaccine antigen and immunization schedule are based on the efficacy and effectiveness of the formulation, including the adjuvant. The formulated product should be fully characterized, and the consistency in manufacture of the formulated product should be demonstrated with regard to various characteristics, such as size of

particles, degree of adsorption or encapsulation, amount of active components of adjuvant, and so on.

## 14.6 ADJUVANTS USED IN LICENSED VACCINES

Until almost two decades ago, aluminum compounds were the only adjuvants used with human vaccines. Calcium phosphate has been used as an adjuvant with a number of vaccines, mainly in France, but has not been in use for many years. During the past two decades, several vaccines with other adjuvants have been approved or licensed for human use (Table 14.3). These include influenza vaccines with an alternate adjuvant system, oil-in-water emulsions (MF-59 and AS03), and HBV and HPV vaccines with a combination adjuvant formulation composed of MPL and aluminum hydroxide (AS04), all of which have been approved in Europe, and HPV vaccine with AS04, which is licensed in the United States [14,16,18]. Other adjuvants that have been used with approved human vaccines include virus-like particles (self-assembling particles composed of one or more viral protein); for example, immunopotentiating reconstituted influenza virosomes are registered as a component of the hepatitis A vaccine in Europe, Asia, and South America.

### 14.6.1 Aluminum Compounds

Aluminum compounds, including aluminum hydroxide ($Al[OH]_3$), aluminum phosphate ($AlPO_4$), a mixture of these compounds, and alum-precipitated vaccines, have been the most commonly used adjuvants with vaccines [19–25]. For inactivated and subunit vaccines, aluminum adjuvants are usually the first choice because of their good safety track record, low cost, and adjuvanticity with a variety of antigens. These adjuvants are often referred to as "alum" in the literature and during discussions at scientific and regulatory meetings, which is misleading because alum, chemically

**TABLE 14.3 Adjuvants Used in Licensed Vaccines**

| Adjuvant | Vaccine | Usage |
|---|---|---|
| Aluminum compounds | Numerous | In use for more than 70 years worldwide |
| Calcium phosphate | DTP combinations | In France for many years<br>Not used anymore |
| MF-59 (oil in water) | Influenza | Europe |
| AS03 (oil in water with α-tocopherol) | Influenza | Europe |
| AS04 (MPL and aluminum) | HBV, HPV | Europe, United States |
| Virosome (liposome) | Influenza<br>Hepatitis A virus | Europe |

DTP: diphtheria, tetanus, pertussis vaccine; HBV: hepatitis B virus; HPV: human papillomavirus; MPL: monophosphoryl lipid A.

potassium aluminum sulfate ($KAl[SO_4]_2 \cdot 12H_2O$), is not an adjuvant as such; and two commonly used aluminum adjuvants, referred to as "alum", aluminum hydroxide and aluminum phosphate, have very different physical characteristics [24,25] and differ in their adjuvant properties [20–22]. Alum was originally used to partially purify protein antigens, mainly tetanus and diphtheria toxoids, by precipitating them in the presence of anions, including phosphate, sulfate, and bicarbonate ions, resulting in a mixture of compounds, mainly aluminum phosphate sulfate or aluminum sulfate. Such vaccines were called "alum-precipitated vaccines," which led to the use of term "alum" for describing vaccines with aluminum-containing adjuvants [19–25].

Alum-precipitated tetanus and diphtheria toxoids were the first routine human vaccines using aluminum adjuvants and were used for human immunization for many years [7,19]. Although "alum" was not the adjuvant in these vaccines, the use of these preparations left a legacy for the term "alum," which has been used to describe all aluminum-containing adjuvants. The legacy of this inaccurate description seems to be in part contributing to our lack of understanding about the use and MOA of these adjuvants, leading to describing all adjuvants as a "dirty little secret" [38]. Researchers and industry may be responsible for maintaining the "dirty little secret" status for aluminum adjuvants by upholding the tradition and convenience of referring to all aluminum-containing adjuvants as "alum" despite differences in physicochemical and adjuvant properties and the MOA of aluminum hydroxide and aluminum phosphate. Perhaps because of proprietary reasons for keeping vaccine formulations secret or because of a lack of clear understanding of the science, aluminum adjuvants have been incorrectly or incompletely described in several vaccines formulations. For example, certain vaccines described to be adsorbed on aluminum hydroxide actually have aluminum hydroxyphosphate in the final formulation. This happened because phosphate buffer used in the formulation converted aluminum hydroxide to aluminum hydroxyphosphate by ligand exchange mechanism [39–41], either before or when antigens were added to the adjuvant.

The stability, potency, and immunogenicity of certain vaccines, particularly those with ligands, such as phosphate groups in HBV and *Haemophilus influenzae* type b (Hib) conjugate vaccines, that can exchange with the hydroxyl group on aluminum hydroxide, depend on mechanism and strength of binding of antigens to the aluminum adjuvant [25,40,41]. Antigens with phosphate groups bind through a covalent bond to aluminum hydroxide by the ligand exchange mechanism. Such a binding is so strong that bound antigens are not immunogenic, which may be attributable to distortion of secondary and tertiary structure of antigens bound chemically to aluminum through ligand exchange mechanism. It may also be possible that protective epitopes on antigens chemically bound to aluminum adjuvants are not presented to B cells or such antigens are not able to get processed by dendritic or other antigen processing cells. When a formulation starts with an aluminum hydroxide adjuvant and a phosphate buffer is used during formulation, the original aluminum hydroxide adjuvant immediately becomes aluminum hydroxyphosphate because of ligand exchange mechanism. In such a formulation, the

antigen with ligands that can exchange with hydroxyl group cannot replace the phosphate group. Such formulations are stable and immunogenic.

Use of alum-precipitated vaccines declined considerably because of variability in production of such formulations and the development of defined and controlled aluminum hydroxide and aluminum phosphate preparations [7,19,20,42–44]. Because aluminum compounds have been widely used for so many years with human vaccines, they have become the benchmark or reference for evaluating new adjuvant formulations. It is very important to prepare optimal formulations of vaccines adsorbed onto aluminum adjuvants in order to correctly evaluate new adjuvants. Two methods have commonly been used to prepare vaccines and toxoids with aluminum compounds—an *in situ* precipitation of aluminum compounds in the presence of antigen (developed originally to purify toxoids by precipitation with alum) and adsorption of antigens onto a preformed aluminum gel [5–7,19–23,42–44]. Adsorption of antigens on aluminum adjuvants depends on the physical and chemical characteristics of the antigen, type of aluminum adjuvant, and conditions of adsorption. These conditions are often not adequately described in publications or not optimized, resulting in poorly formulated aluminum adjuvant preparations that do not exhibit optimal adjuvanticity.

Another complexity in understanding the MOA of aluminum adjuvants is considering aluminum phosphate, aluminum hydroxide, and all of the vaccine formulations containing these adjuvants similar. Characteristics of aluminum adjuvants, for example, their physical structure being crystalline or amorphous, size of the gel particles, adsorption capacity, point of zero charge, ratio of aluminum to phosphate, and so on depend upon the conditions of making these gels, including the order of adding reagents, speed at which the reagents are added and mixed, mixing speed, time taken to adjust pH, and scaling up the production [20–23,25,44]. Such variables, if not controlled, affect batch-to-batch consistency of the adjuvant; therefore, aluminum adjuvants have been described as difficult to manufacture in a physicochemically reproducible way. However, over the years with the implementation of current good manufacturing practices (cGMP) to the vaccine production industry, manufacture of aluminum adjuvants has been standardized, and consistency in manufacture has been demonstrated within a manufacturer. But there may still be a high degree of heterogeneity in physicochemical properties of aluminum adjuvants from various manufacturers. Physicochemical properties, including the size of gel particles and charge of the adjuvanted vaccine, quality (strength), and quantity (degree of adsorption/binding) of antigens adsorbed onto the adjuvant, are distinct for each product, which may have an impact on the adjuvant properties and MOA [25,40]. The physicochemical properties for the same vaccine (e.g., a simple adsorbed tetanus toxoid) will be different for different manufacturers because of heterogeneity in quality of aluminum adjuvant and different formulation conditions. It will not be surprising then to find a different MOA for a vaccine antigen adsorbed onto different aluminum adjuvants or the same aluminum adjuvant under different conditions. For example, activation of immune competent cells may depend on an overall charge and other characteristics of the vaccine formulation, and the size of gel particles may lead to either depot formation for a formulation with large particles

**TABLE 14.4 Important Characteristics and Parameters for Adsorption of Antigen onto Aluminum Adjuvants**

| Characteristic or Parameter | Optimal Conditions |
|---|---|
| Ligand exchange (ions or groups on antigens, e.g., phosphate) | Avoid ligand exchange between antigen and adjuvant<br>Use phosphate buffer with aluminum hydroxide for antigens with ligands for potential exchange |
| Electrostatic forces (charge on antigen and adjuvant) | Opposite charge on antigen and adjuvant at pH for formulation |
| Hydrophobic interactions | |
| Van der Waals forces | |
| Hydrogen bonding | |
| pH | 6.0–7.5, at optimal pH adjuvant and antigen with opposite charges |
| Temperature 4–25°C | Depending on the stability of antigen |
| Size of gel particles | $<10\,\mu m$ |
| Ionic strength of reaction mixture | Low, absence of phosphate ions for aluminum phosphate |

($>10\,\mu m$) or direct delivery to macrophages or dendritic cells for a formulation with small particles ($<10\,\mu m$). In many publications, the MOA of aluminum adjuvants has been evaluated by using the aluminum adjuvant alone or the type of aluminum adjuvant is not specified, often referred to as "alum." Because of the different physicochemical properties of an aluminum adjuvant alone and vaccine formulated with the aluminum adjuvant, the MOA may be different for various preparations. Therefore, it is expected that there are a number of hypotheses, sometimes conflicting, about the MOAs of aluminum adjuvants (discussed later).

The characteristics and parameters affecting the adsorption of antigens onto aluminum adjuvants are summarized in Table 14.4. The adsorption of antigens onto aluminum adjuvants depends heavily on the electrostatic forces between adjuvant and antigen and ligand exchange [24,25,39–41]. Adsorption based on electrostatic forces is a preferred method, and ligand exchange, in general, between adjuvant and antigen should be avoided. Other interactions, including hydrophobic, van der Waals, and hydrogen bonding, contribute to the adsorption of antigens on aluminum adjuvants. However, these forces may not suffice to cause adsorption of antigen if the same charge or electrostatic repulsive force is present on antigen and adjuvant. The two most commonly used aluminum adjuvants (aluminum hydroxide and aluminum phosphate) have different points of zero charge [25]. At neutral pH, these gels have opposite charges, wherein aluminum phosphate is negatively charged and aluminum hydroxide is positively charged. It is important to select the aluminum adjuvant carefully on the basis of the charge of the antigen at neutral pH or any other desirable pH. Other physical conditions affecting adsorption of antigens on mineral adjuvants include pH, temperature, size of the gel particles, and ionic strength of the reaction mixture. Whereas the pH and ionic strength affect adsorption by altering

charge on the gel and the antigens, the temperature may affect the rate of interaction between the gel and the antigen. The size of gel particles affects the surface area of gel available for adsorption: small particles have more surface area than large particles. For example, the amount of diphtheria toxoid adsorbed onto aluminum hydroxide gels was inversely proportional to the gel particle size [45].

An acidic pH of less than 6 has been commonly used for adsorption of tetanus and diphtheria toxoids onto the aluminum phosphate adjuvant [42] because diphtheria toxoid and aluminum phosphate are both negatively charged at neutral pH, resulting in poor or no adsorption. At a pH of 6.0, aluminum phosphate is positively charged, thus improving adsorption of negatively charged diphtheria toxoid. *In situ* adsorption of diphtheria toxoid resulted in higher adsorption than the commercial aluminum phosphate preparation, probably because of trapping of some antigen in the gel [20–22]. Adsorption of tetanus and diphtheria toxoids onto aluminum hydroxide gel (Alhydrogel) was not sensitive to the pH commonly used in formulation of these vaccines (pH 6.0–7.0) and to excess phosphate ions because aluminum hydroxide is positively charged and toxoids are negatively charged at this pH range [20–22]. Adsorption of bovine serum albumin onto aluminum hydroxide or lysozyme onto aluminum phosphate was inversely proportional to their ionic strength [46]. Excess anions, particularly phosphate ions, and impurities, such as amino acids, peptides, and polysaccharides, reduce protein adsorption, probably by competing with antigen for adsorption sites [47]. Multiple-charged negative ions, especially phosphate ions, interfere with the adsorption capacity of aluminum hydroxide because of conversion of aluminum hydroxide to aluminum hydroxyphosphate and therefore affecting its charge. These negative ions may be used for elution of adsorbed antigens from the gel. Lindblad and Sparck [47] recommended avoiding phosphate buffers in the reaction mixture for adsorption of antigen onto aluminum adjuvants, but use of phosphate buffer has been helpful in avoiding binding of certain antigens to aluminum hydroxide by a ligand exchange mechanism. In general, a low ionic strength and absence of excess phosphate ions and impurities are recommended for optimal adsorption of antigens on aluminum phosphate gel [48].

It is recommended that each formulation be evaluated and optimized for adsorption. Although it may be beneficial to have maximum adsorption of antigen to the adjuvant, binding by ligand exchange mechanism may be detrimental to the immunogenicity and effectiveness of many antigens. The temperature of adsorption has been considered important for complete adsorption of antigen onto aluminum phosphate, although most of the adsorption, up to 80 to 90% of diphtheria toxoid, occurred within a few minutes at temperatures ranging from 4 to 45°C [42]. There has been some discussion on desorption of antigen from adjuvant after injection into the body when a physiologically neutral pH and presence of body fluids containing proteins and anions might desorb the antigen from the gel. Earlier studies showed that freshly made preparations of aluminum phosphate adsorbed diphtheria toxoid had more antigen desorption than aged preparations when exposed to neutral pH or serum [42]. Thus, aging of aluminum-adsorbed vaccines appears to improve their immunogenicity [43].

The MOA of aluminum-containing adjuvants is still not clearly understood despite their extensive use in human vaccines for almost a century [49,50]. A number of mechanisms have been proposed over the years for the adjuvant action of aluminum-containing adjuvants, starting with a concept of depot formation at the site of injection. Glenny et al. [51], who first described and used aluminum adjuvants, proposed in 1931 that aluminum adjuvants act by depot formation at the site of injection, allowing for a slow release of antigen and thus prolonging the time for the interaction between antigen and antigen-presenting cells (APC) and lymphocytes. The local depot mechanism was challenged when Holt [43] observed that antibody formation continued after removal of the adjuvant–antigen depot from the site of injection. In another study, the depot formation could not be confirmed, when approximately 90% of radiolabeled tetanus toxoid adsorbed onto aluminum phosphate disappeared from the site of injection within 24 hours of subcutaneous injection, but tetanus toxoid encapsulated within biodegradable polymer microspheres stayed at the site of injection for prolonged periods [52]. However, the amount of aluminum phosphate adsorbed tetanus toxoid at the site of injection was higher than the soluble tetanus toxoid for 4 weeks, thus demonstrating some localized depot formation. The depot concept, which is still considered one of the mechanisms for aluminum adjuvants, has generated numerous discussions. Is it a short-term depot, to which macrophages are attracted, or a long-term depot from which antigen is released over an extended time? There are also questions regarding site of the depot. Is it formed at the site of injection or in the draining lymph nodes? These questions remain unanswered with evidence to support or contradict these theories. White et al. [53] showed that antibody-producing cells in the regional, popliteal lymph nodes of rabbits injected with 150 Lf of soluble diphtheria toxoid completely disappeared in 3 weeks, but rabbits injected with 10 Lf of aluminum phosphate precipitated toxoid had antibody producing cells in the nodes at 3 to 4 weeks, suggesting a depot in the draining lymph nodes. The most direct evidence for a local depot effect comes from experiments in which local granulomas, formed after injection of aluminum-adsorbed vaccines, were able to induce immune responses when these granulomas were excised from the site of injection 7 weeks later, macerated, and injected into other animals [54]. The antigen in the granuloma was apparently not available to the animals for a secondary response because minute doses of antigen injected adjacent to the granuloma produced a secondary antibody response [55]. White [55] postulated that antigen within the granuloma was unable to penetrate the fibrous tissue surrounding the granuloma, and antibody may be reacting with antigen to form an antigen–antibody precipitate within the fibers of the peripheral zone of the granuloma, thus preventing the diffusion of antigen from granuloma and sequestering it from APC.

Another hypothesis on the MOA of aluminum-containing adjuvants has been considered as their potential to target APC. The adjuvanticity of mineral adjuvants may be due to their ability to convert soluble antigens to particulate forms, which are more readily phagocytosed. The particle size of commercially available aluminum adjuvant gels is less than 10 μm, the average size for aluminum hydroxide, Alhydrogel was 3.07 μm and for aluminum phosphate, Adju-phos was 4.26 μm.

It has been shown that polylactide glycolide (PLGA) microspheres smaller than 10 μm are taken up by APC and provide strong adjuvant effects [56]. The size of aluminum adjuvant particles changes after adsorption of antigens onto the adjuvant and over time when adsorbed vaccines are stored. So this hypothesis of targeting to APC will be different for various formulations depending on the size of particles in the formulation. Antigen adsorbed onto aluminum hydroxide is more readily taken up by human monocytes than free antigen, and human monocytes exposed to aluminum hydroxide secrete interleukin-1 (IL-1) [57]. The adjuvanticity of aluminum gel particles via targeting to APC emphasizes the importance of degree of adsorption of antigens onto gels. However, it has been reported that antigens partially adsorbed onto aluminum-containing adjuvants are as effective as those with complete adsorption.

Activation of the immune system by aluminum-containing adjuvants has also been discussed as a possible MOA. An increased antibody response to a soluble antigen has been observed when aluminum adjuvant was injected at a different site, suggesting a systemic stimulatory effect on immunocompetent cells, possibly mediated by cytokines [58]. However, these results were not confirmed in a subsequent study, when antibody responses of mice to diphtheria toxoid and aluminum hydroxide injected at different sites versus soluble diphtheria toxoid alone were very low and similar, even after 2 doses [20]. Aluminum compounds have been reported to induce eosinophilia [59] and activate complement [60], which may lead to a local inflammatory response, thereby enhancing the antibody response. It was also shown that mice injected with alum (aluminum hydroxide) with or without antigen had an increase in IL-4–producing cells bearing the monocyte and neutrophil marker $Gr1^{+}$(5) [61]. Mineral adjuvants augment mainly humoral immune response, particularly IgG1 and IgE antibody responses, through IL-4 [62] by activation of Th2-type cells [8,63]. This type of immune response is similar to that elicited by most soluble protein antigens. Cooper [63] described aluminum adjuvants as good stimulants for Th2-type cell-mediated immune response, especially eosinophils, which is similar to that elicited by some helminth parasites [44]; thus, aluminum adjuvants may be a good candidate for antiparasitic vaccines. In a special mouse hybrid model of schistosomiasis, Horowitz et al. [64] demonstrated protection in animals injected with sonicated parasite antigen and aluminum hydroxide. In contrast, antigen injected with FCA elicited lower IgE antibodies and lower levels of protection. Aluminum adjuvants are not efficient in raising cell-mediated immune responses [8,65]. The induction of delayed type hypersensitivity by aluminum adjuvants in mice and guinea pigs has not been clearly demonstrated [48,66].

Lately, there has been much attention in understanding the MOA of aluminum adjuvants, the most used adjuvant in the history of vaccination [25,27,30,49,50]. During the past 10 years, there have been a number of reports on the mechanism of adjuvanticity for aluminum-containing adjuvants at the molecular level. Tritto et al. [49] suggested two independent mechanisms of adjuvanticity for aluminum-containing adjuvants: (1) presenting antigen in a particulate multivalent form to APC, thus enhancing immunogenicity and (2) promoting a local proinflammatory

environment by Toll-like receptor– (TLR-) independent mechanism, which results in blood cell recruitment and dendritic cell differentiation. However, the molecular target for alum proinflammatory activity is still unknown [49].

Recent data from different laboratories demonstrated that the molecular target for aluminum-containing adjuvants is nucleotide binding and an oligomerization domain (NOD)-like receptor protein 3 (NLRP3) [49,67–69]. NLRP3 interacts with the apoptosis-associated speck-like protein containing a caspase recruitment domain, ASC, and the protease caspase 1 to form a protein complex called inflammasome. When the inflammasome is activated, it processes proinflammatory cytokines, including IL-1β, IL-18, and IL-33, into their mature form. *In vitro*, alum activates caspase 1 through NLRP3 and synergizes with TLR agonists such as lipopolysaccharide for IL-1β production [67,68]. This process is blocked by actin polymerization inhibitors, indicating that NALP3 activation requires phagocytosis [67]. It was further demonstrated that NLRP3 activation contributed to alum adjuvanticity when antibody responses to a number of antigens adjuvanted with alum were significantly reduced in NLRP3-deficient mice compared with wild-type animals [68]. In an another study, Th2 cell priming and antibody responses to ovalbumin or human serum albumin co-administered with alum were impaired in NLRP3-, ASC-, or caspase 1–deficient mice [70]. These data could not be reproduced because human serum albumin injected with alum to NLRP3-deficient mice showed comparable antibody responses to those elicited in wild-type animals, suggesting that NLRP3 is not required for alum adjuvanticity [71]. These conflicting data might be the result of different immunization routes and different protocols for measuring the immune response [49].

Another hypothesis on the adjuvanticity of aluminum-containing adjuvants is based on activation of NLRP3 through the induction of necrosis, which causes the release of danger signals such as uric acid and adenosine triphosphate [72,73]. A higher concentration of uric acid was observed locally after alum administration in the peritoneum, and uricase treatment decreased T-cell priming induced by alum, suggesting that uric acid release at injection site is required for alum adjuvanticity [72]. However, necrosis, as a mechanism of alum adjuvanticity, was not supported by *in vitro* data showing that macrophage treatment with uricase does not affect the response to alum [68,70]. Furthermore, Flach et al. [74] also suggested that this may not be possible because alum injection triggered a mild tissue irritation, suggesting little cell death or uric acid release. However, more insights on activation of NLRP3 by alum came from Hornung et al. [75], who showed that phagocytosis was essential for NLRP3 activation and IL-1β release by alum, silica crystals, or fibrillar amyloid β. A recent report showed that alum interacted with membrane lipids on the surface of dendritic cells, triggering signaling cascades that promote $CD4^+$ T-cell activation and humoral immune responses [50].

It seems that the mechanism of adjuvanticity for aluminum-containing adjuvants is far from clear [49,50] despite a clear understanding of the physicochemical properties of aluminum-containing adjuvants [24,25,39,40], description of various hypotheses on the adjuvanticity [27,49,50], and an optimism that adjuvants are no longer a “dirty little secret” because of an increase in our understanding of innate

immunity [16,26,38]. This may be due to a unique MOA for every aluminum adjuvant-containing formulation, as discussed earlier, or there may be other possibilities. Aluminum-containing compounds have been shown to be potent adjuvants in various animals models, particularly in rodents, mice, and guinea pigs, with a number vaccine antigens, especially tetanus and diphtheria toxoids [76,77]. Their role in providing enhanced antibody response in humans has been questioned for tetanus and diphtheria vaccines from time to time, indicating that these adjuvants may not be potent adjuvants in humans. Despite a lack of an understanding about the MOA, it is clear that aluminum adjuvants have an important role in the effectiveness of existing vaccines. One of the roles that aluminum adjuvants may have in the enhanced immune response of many vaccines could be their ability to stabilize vaccine formulations and ensuring the complete delivery of vaccine antigen dose. Because vaccines are used in minute amounts (microgram quantities), there is a possibility of a small but significant amount of these active ingredients lost because of protein aggregation or binding on the surface of delivery containers, such as vials and syringes. Aluminum adjuvants provide stability to certain antigens when adsorbed on to them and additionally ensure complete delivery of small amounts of antigens because antigens, when not formulated with aluminum adjuvants, can adsorb on the surfaces of containers. This role in helping vaccine antigens may be considered as adjuvant effect within the broader definition of adjuvants, but this is a formulation issue, which can also be achieved by developing alternate formulations.

The enormous heterogeneity in physicochemical properties of vaccines formulated with aluminum-containing adjuvants, leading to a chaos in our understanding of their MOA, has an advantage in the flexibility in designing these adjuvants to characteristics required for a specific formulation. By carefully adding varying amounts of phosphate to aluminum hydroxide, a formulation of aluminum hydroxyphosphate having zero charge at the desired pH between 5 and 10 can be achieved. This type of formulation work will lead to rationale design of vaccines with aluminum adjuvants. Such formulations, with varying amounts of phosphate buffer added to aluminum hydroxide to optimize the adjuvanticity, have been developed empirically without understanding the science involved in the conversion of aluminum hydroxide to aluminum hydroxyphosphate, thus changing mechanism of antigen binding to the adjuvant. Such well-characterized formulations may help in our understanding of the MOA of aluminum adjuvants, taking advantage of all the knowledge and tools gained over the years in understanding aluminum adjuvants.

Despite so much confusion about the adjuvanticity of aluminum-containing adjuvants, these adjuvants have been in use for almost a century and will continue to be used with currently licensed vaccines and new vaccines because of their long track record of use, low cost, and extensive experience in manufacture of these products. With enhanced understanding of the physicochemical characteristics of aluminum compounds and flexibility in modulating these characteristics to manage the quality and quantity of binding or adsorbing vaccine antigens onto these adjuvants, a rationale design in developing vaccines with these adjuvants is possible and desirable. This will enable a better understanding of the MOA of these

adjuvants and ensure control over manufacturing process to achieve consistency in manufacture.

### 14.6.2 Oil Emulsions

Another class of adjuvants used with human vaccines intermittently for almost 50 years is oil emulsions, which also have a long history of use with animal vaccines [78–81]. Almost a century ago and before aluminum adjuvants were described, Pinoy and Le Moignic [1] in 1916, reported that a suspension of killed *Salmonella typhimurium* in mineral oil increased the immune response. Later, in 1925, Ramon [2] described starch oil as one of the substances augmenting the antitoxic response to diphtheria toxoid. The oil emulsions as adjuvants did not receive further attention until Freund [30] in 1937, demonstrated the adjuvant effect of mineral (paraffin) oil mixed with killed *Mycobacteria*, referred to as FCA. FCA has been one of the most potent adjuvants described so far and has been used extensively to augment the immune response of laboratory animals mainly in generating hyperimmune serum. However, its use in animals has been banned or restricted for several years because of immense suffering and pain to animals. FCA has been too toxic to be used in humans.

The water-in-oil emulsion without *Mycobacteria*, called Freund's incomplete adjuvant (FIA), has been used in a number of veterinary vaccines [78–81]. FIA was used in humans with influenza and killed poliomyelitis vaccines, enhancing the immunogenicity of these vaccines [82–87]. The MOA of FIA was attributed to depot formation at the site of injection and slow release of antigen with stimulation of antibody-producing cells [88–90]. Use of FIA was discontinued in humans because of the side effects such as local reactions at the site of injection (granuloma and cyst formation), oil-induced neoplasms in mice, and Arlacel A induced carcinogenicity in mice [91,92]. Nevertheless, FIA-based influenza vaccine produced fewer side effects than aqueous vaccines [83] probably because of use of less antigen with this adjuvant and also because of the slow release of antigen from the adjuvanted vaccine. Hundreds of thousands of doses of FIA-adjuvanted vaccines had been inoculated into humans before the possibility of carcinogenicity was raised based on a study in mice. A long-term follow-up of the subjects injected with FIA-adjuvanted vaccines failed to show an increase in mortality, tumors, or autoimmune diseases caused by oil adjuvants [93]. The FIA-based bacterial vaccines such as cholera [94] and tetanus [95] produced better immunity than aqueous preparations, but side effects, including sterile abscesses, ulcers, and delayed-type local reactions, were recorded. These severe local reactions were considered to be attributable to the presence of lipopolysaccharide (LPS) in cholera vaccine and to certain enzymes in the bacterial vaccine components that degraded the Arlacel A to release toxic fatty acids [96–98]. Based on severe local reactions and the possibility of carcinogenicity as observed in mice, it was concluded that the use of mineral oil adjuvants in the human population may be hazardous and should not be recommended for general use in humans [92,99].

In search of an alternate adjuvant to FIA, various kinds of emulsions with different oils were evaluated to develop a stable, nontoxic, and effective adjuvant for human

vaccines. Hilleman and colleagues described an oil-based adjuvant, "Adjuvant 65," consisting of 86% peanut oil, 10% Arlacel A (the ester mannide monooleate) as the emulsifier, and 4% aluminum monostearate as the stabilizer [82,100]. Unlike FIA, this adjuvant contained an oil component, which was metabolized or excreted by the body. In humans, Adjuvant 65 was safe and potent but exhibited less adjuvanticity than FIA. Nevertheless, the use of Adjuvant 65 was discontinued because of reactogenicity in humans with certain lots of vaccine and a reduction of adjuvanticity when a purified or synthetic emulsifier was used in place of Arlacel A [101]. Different emulsions such as oil-in-water and water-in-oil-in-water emulsions using other oils, including sesame oil and peanut oil, were also tried as adjuvants, but none of these emulsions has been used for routine human vaccines [102–108].

During the 1980s and early 1990s, several reports described oil-in-water emulsions using squalene or such emulsions with emulsifiers, surfactants, or immunomodulators to stabilize the emulsion or enhance the adjuvanticity of the formulation [109–112]. The emulsifiers commonly used to stabilize oil-in-water emulsions included fatty acid esters of polyhydric alcohols such as Tween and Span series. Synthetic organic surfactant adjuvants used with squalane-in-water emulsion were the block polymer surfactants [111]. Immunostimulators used in oil-in-water emulsions included MDP and its derivatives, trehalose dimycolate (TDM) and MPL, with or without emulsifiers or non-ionic block copolymer surfactants [109–114]. It was suggested that antigen should be incorporated into the internal phase of oil emulsions for an effective adjuvanticity (e.g., the water phase of water-in-oil emulsion and the oil phase of oil-in-water emulsion) [110].

During the early 1990s, an oil-in-water emulsion named MF-59 containing 4.3% squalene (w/v), 0.5% Tween 80 (w/v), and 0.5% Span 85 (w/v) was developed by scientists at Chiron Vaccines. This emulsion gave consistently higher antibody levels to a variety of antigens in animals and humans than aluminum adjuvants [16,115–118]. Early toxicology studies of MF-59 in animals and in humans demonstrated its safety with vaccines against herpes simplex virus (HSV), HIV, and influenza virus. MF-59 has been shown to be a stimulator of both humoral and cellular immune responses to subunit antigens in animals as well as humans. Based on early radiolabeled studies, it was suggested that MF-59 does not appear to act by depot formation [16]. In a recent article [118], the MOA of MF-59 has been proposed from all the work performed on this adjuvant. It has been described that MF-59 creates a transient "immunocompetent" local environment at the injection site, resulting in the recruitment of key immune cells, which take up the antigen and the adjuvant and transport them to the local lymph nodes, where the immune response is induced. It seems that MF-59 induces local cells to create chemokine-driven gradient, which recruits immune cells to the injection site.

Inactivated influenza vaccine with MF-59 has been licensed for more than 10 years in Europe and has been administered to more than 50 million people worldwide with a good safety record [16,116,117]. The development of MF-59 over many years shows challenges in developing vaccines in general and adjuvants in particular. Successful development of MF-59 adjuvant was possible because of the persistence and patience of the Chiron scientists in developing this formulation and because of

lessons learned from failures and setbacks in evaluating various formulations for more than 20 years [16]. Originally, Chiron Vaccines developed an oil-in-water emulsion as a delivery system for an immunostimulator called MTP-PE, a synthetic version of MDP [115]. In clinical testing, this formulation showed high reactogenicity for use with routine vaccines [119,120]. So this formulation was another failed attempt to develop a vaccine adjuvant because of safety problems. Further evaluation of this formulation showed that the emulsion vehicle alone, without the immunostimulator, was an effective adjuvant [121]. The emulsion alone (MF-59) showed an adjuvant effect with inactivated influenza vaccine with no additional benefit of MTP-PE. Exclusion of MTP-PE prevented significant and unacceptable local and systemic reactions. These observations led to development of MF-59 as a stand-alone adjuvant for influenza vaccines [122], which has benefits of extending the supply of influenza vaccine during a pandemic because of use of a lower vaccine dose. Furthermore, MF-59 has been shown to induce cross-protective immune response against divergent strains, which is important both for pandemic and seasonal vaccines [116,117]. MF-59 has been effective in elderly adults as well as infants who are most vulnerable to influenza and respond poorly to influenza vaccines [16].

Other lessons learned by Chiron scientists during development of MF-59 included importance of selection of appropriate antigen, good trial design, and relevant immunologic endpoints for the success of a vaccine [16]. A good adjuvant cannot compensate for poor selections in these areas.

Initially, the MF-59 adjuvant might be perceived to have "failed," but what really failed were the assumptions on which the vaccine candidate was based. For example, MF-59 was used with HSV vaccine in a phase III efficacy trial for protection against genital herpes simplex infection, and the vaccine failed to offer sustained protective efficacy [123]. The same vaccine formulation had also failed to show efficacy as a therapeutic vaccine candidate in a previous trial [124]. The adjuvant, MF-59, was selected because of its ability to induce high levels of neutralizing antibodies and T-cell responses, and the adjuvant achieved that with the HSV vaccine [125]. It was found that serum-neutralizing antibodies did not offer protection against HSV infection, which is a sexually transmitted disease, for which there is often repeated exposure of high viral load [126]. There are additional examples in which MF-59 performed as expected, but a vaccine antigen or the concept was not appropriate. MF-59 has induced potent neutralizing antibody responses against HIV gp120 envelope [127] and high levels of protective titers against hepatitis B surface antigen [128]. However, the antibody responses induced against HIV gp120 were not able to neutralize relevant "field" strains, which has been a problem for all HIV vaccine candidates evaluated so far. Therefore, a good adjuvant cannot fix a bad antigen and cannot compensate for an incorrect assumption in a clinical trial design [16].

Another oil emulsion–based adjuvant licensed in Europe with inactivated influenza vaccines is GlaxoSmithKline's AS03, which is a combination of α-tocopherol (vitamin E) and squalene in an oil-in-water emulsion with a droplet diameter of approximately 150 nm [14,18,129]. Vitamin E is known to have immune-enhancing properties and maintains the integrity of cellular membranes by protecting their physical stability [18]. Both monocytes and macrophages respond to AS03 with a

local production of a range of cytokines and chemokines. An AS03-adjuvanted pandemic influenza vaccine has allowed for antigen sparing with a high level of cross-reactive immune response to heterologous strains of H5N1 [130–132]. Both the MF-59 and AS03 adjuvant formulations are designed to mix at the time of use, which provides flexibility in the use of adjuvant. This has raised a question about regulatory agencies approving an adjuvant alone, which can be used with an approved vaccine. An approved vaccine must be evaluated with an already approved adjuvant for another vaccine for its safety and efficacy. As discussed earlier, mixing of an antigen and adjuvant may create a unique formulation with distinct physicochemical and adjuvant characteristics depending on the nature and characteristics of the antigen. Therefore, licensing of an adjuvant will not represent the safety profile of the adjuvant when formulated with a particular vaccine antigen. When a vaccine is licensed as two separate entities or components, vaccine antigen and adjuvant, which are mixed immediately before use, consistency in physicochemical characteristics and the safety and stability of the emulsion for a few hours (e.g., up to 6 hours) must be demonstrated for the mixed product.

### 14.6.3 Other Adjuvant Systems Used in Licensed Vaccines

Liposomes have been around for almost 40 years [133] and have been evaluated extensively as adjuvants for the last 30 years, alone as delivery vehicles or for targeting antigens to APC, as well as in combination with other immunostimulators, such as MDP, LPS, lipid A, MPL, QS-21, and so on [16–18,133–139] Liposomes are typically made up of concentric lipid membranes containing phospholipids and other lipids in a bilayer configuration separated by aqueous compartments. These were originally developed to be used as carriers of biologically active substances and have been safely administered parenterally to patients as carriers of therapeutic drugs without formation of granulomas at the site of injection. They acted as adjuvants in a number of animal studies with a variety of antigens. Both humoral and cell-mediated immune responses have been augmented by liposomes [140]. Not only is the clearance of antigens incorporated into liposomes markedly prolonged, but the liposomes may also target antigen to APC, including macrophages. The cell-mediated immunity is believed to be stimulated because of the hydrophobic nature of liposomes. Adjuvants such as LPS, lipid A, MPL, and MDP when incorporated into liposomes had better adjuvanticity and less toxicity [138]. Nonphospholipid liposomes (Novasome vesicle) have been used commercially for two poultry vaccines approved by the US Department of Agriculture. These nonphospholipid liposome vesicles were evaluated with human vaccine antigens (tetanus and diphtheria toxoids) in rabbits and mice [139]. Two common approaches to formulate antigens with liposomes include adsorption on its surface or entrapment inside the liposomes [17]. Tetanus and diphtheria toxoids encapsulated in or mixed with these liposomes elicited antitoxin levels similar to those elicited by antigens given with FCA or adsorbed onto aluminum adjuvants [139].

Another type of liposome, referred to as virosomes or fusogenic liposomes, is created by inserting virus fusion proteins into a liposome bilayer [141]. Virosome-based

influenza and hepatitis A vaccines have been licensed for human use in Europe for more than 10 years [16,17,142]. These products have been shown to be well tolerated and immunogenic in all age groups, including elderly adults, adults, and children. The suggested MOA of influenza-based virosomes involves a direct interaction of virosome particles with APC or, in some cases, with B cells, which in turn, activate T cells. Despite the limited use of virosome-based vaccines in Europe, liposomes have not been used for routine human vaccines. However, in recent years, a number of liposome-based therapeutic vaccines, particularly against cancer, have been evaluated [143–145].

Combination of two or more adjuvants has been evaluated in the past in animals and in human clinical trials. During the last 10 years, two vaccines, hepatitis B vaccine and HPV vaccine, with a combination of two adjuvants, MPL and aluminum hydroxide, named AS04, were approved for human use [14,16,18]. MPL, a derivative of the bacterial cell wall LPS, is a potent immunostimulant. The adjuvant effect of LPS was described in 1956 [146], although it was known earlier that certain Gram-negative bacteria or vaccines from these organisms had an adjuvant effect [6]. LPS and MPL activate TLR-4, a receptor expressed on APC [26,49]. The engagement of TLR-4 promotes cytokine expression, antigen presentation and migration of APC to the T-cell area of draining lymph nodes, allowing for an efficient priming of naïve T cells.

Most of the biological activities, including adjuvanticity and toxicity, of LPS are due to lipid A region of the molecule [147,148], thereby making LPS and lipid A too toxic for use in human vaccines. Incorporation of LPS or lipid A into liposomes reduced the toxicity while retaining or even increasing the adjuvant effect [138]. Attempts have been made to detoxify LPS without affecting its adjuvanticity. The most promising derivative of lipid A, which has been shown to have most of the adjuvant properties of LPS but with markedly diminished side effects, is MPL [149–152]. The lower toxicity of MPL compared with LPS has been ascribed to the removal of one of the phosphates on lipid A. However, when lipid A was completely dephosphorylated with hydrofluoric acid, the dephosphorylated preparation still had same amount of pyrogenic activity in rabbits as the lipid A (Gupta, R., unpublished data). During preparation of MPL by mild acid hydrolysis of lipid A from *Salmonella minnesota*, one or two fatty acid chains on lipid A backbone bound by ester bond are also removed along with one phosphate group. Normally, lipid A with complete toxicity and adjuvant properties has six or seven fatty acid chains. Lipid A with less than six fatty acid chains has demonstrated significantly less toxicity and pyrogenic activity while retaining some of the adjuvant activity of unmodified lipid A (Gupta, R., unpublished data). The balance between toxicity and adjuvanticity of MPL without one phosphate group and fewer than six fatty acid chains makes it suitable for human use as compared with the toxicity–adjuvanticity balance of lipid A and LPS.

In AS04, MPL is adsorbed onto aluminum hydroxide or aluminum phosphate, depending on the vaccine and its desired adsorption [153,154]. MPL activates the innate immune by directly stimulating TLR-4, leading to maturation of dendritic cells and the expression of cytokines that enhance the adaptive immune response by stimulating the maturation of Th cells to the Th1 type. Thus, the recognition of MPL

by TLR-4 leads to enhanced humoral and cellular immune responses. The aluminum adjuvant appears to modulate and prolong the cytokine responses to MPL at the injection site. These results support a model in which addition of MPL to an aluminum adjuvant enhances the vaccine response by prompting increased activation of APC and stimulation of Th1-type responses [155].

AS04-adjuvanted HBV vaccine significantly enhanced the immune response to hepatitis B antigen and induced rapid, higher, and longer lasting seroprotection and enhanced cell-mediated immune response compared with aluminum-adjuvanted vaccine for prehemodialysis and hemodialysis patients, who are relatively poor responders to aluminum-adjuvanted HBV vaccine [156], Similarly, the AS04-adjuvanted HPV vaccine induced higher antibody levels compared with the same antigen formulated with an aluminum adjuvant. Furthermore, the AS04-adjuvanted HPV vaccine provides cross-protection against certain other high-risk HPV types not contained in the vaccine [157].

## 14.7 FUTURE DIRECTIONS AND PROSPECTS

For more than 50 years, a number of novel adjuvants have been evaluated with human vaccines in preclinical and clinical studies. Before the past two decades, aluminum adjuvants were the only adjuvants used with routinely administered human vaccines. During the past 20 years, the licensure of a number of vaccines with alternate adjuvants has renewed optimism and enthusiasm in the development of novel adjuvants.

Expanded knowledge about pathogenesis and protection from diseases, either after a natural infection or against early vaccines, containing whole organisms (inactivated or live attenuated) and purified or partially purified virulence factors has led to an understanding about recognition of pathogen-associated molecular patterns (PAMPs) by the immune system through pathogen-recognition receptors, including the TLRs, NLRs, and RIG-I-like receptors (RLRs) [26]. Most currently used vaccines made from whole organisms and virulence factors of pathogens, including detoxified toxins, have inherent adjuvanticity due to PAMPs that can bind various receptors. For example, bacterial cell walls, peptide sequences in toxins and other virulence factors, DNA, and viral RNA commonly present in current vaccines can bind specific TLRs and activate the TLR-expressing APC. Such an understanding of innate immune responses and their impact on modulation of B- and T-cell responses facilitates the development of adjuvants with the potential to modulate the quality and quantity of immune response with an acceptable risk-to-benefit ratio for the particular indication for which the adjuvant is being developed.

With the growing knowledge of the innate immune system and a better understanding of mechanisms of adjuvanticity and toxicity of various adjuvants, it seems possible for novel vaccines to be rationally designed or tailored when the benefits outweigh the risk. For example, a low- or weakly modulated immune response for a prophylactic vaccine and a strongly modulated immune response for a therapeutic vaccine against cancer or HIV with a high risk and higher benefits will both be acceptable and desirable.

## ADDENDUM

When this chapter was being drafted, there were initial reports of rise in cases of narcolepsy among children and adolescents, mainly from Sweden and Finland, receiving pandemic H1N1 influenza vaccine containing the adjuvant AS03 (Pandemrix) [158,159]. Since then, more cases of narcolepsy have been reported from children and adults receiving this vaccine containing AS03 adjuvant [160–162]. No increase in cases of narcolepsy has been reported from individuals receiving H1N1 vaccine without an adjuvant and with another oil-based adjuvant MF-59 [163]. While the relationship between the H1N1 vaccine containing AS03 and narcolepsy is being further evaluated, this shows the challenges in developing adjuvants and maintaining a balance between adjuvanticity and toxicity for vaccines used in masses. Narcolepsy has been associated with human leukocyte antigen (HLA) DQB1*06024 [164], with likely autoimmune link to this HLA, T-cell receptor polymorphisms [165], Tribbles 2 autoantibodies [166], and other environmental factors [158]. Adjuvants are used to modulate or enhance the immune response in an individual to the vaccine antigens. As discussed in this chapter, maintaining a balance between adjuvanticity and toxicity of the adjuvant is critical for prophylactic vaccines given to millions of healthy individuals.

## ACKNOWLEDGMENTS

I am thankful to Dr. James L. Kenney and Dr. Ramjay Vatsan for a careful review of the manuscript and valuable comments.

## REFERENCES

1. Moignic L, Pinoy. Les vaccins en emulsion dans les corps gras ou 'lipo-vaccins'. C. R. Soc. Biol. 1916;79:201–203.
2. Ramon G. Sur l'augmentation anormale de l'antitoxine chez les chevaux producteurs de serum antidiphterique. Bull Soc Cent Med Vet 1925;101:227–234.
3. Ramon G. Procedes pour accroitre la production des antitoxines. Ann Inst Pasteur 1926;40:l–10.
4. Glenny AT, Pope CG, Waddington H, Wallace U. The antigenic value of toxoid precipitated by potassium alum. J Pathol Bacterial 1926;29:38–45.
5. Edelman R. Vaccine adjuvants. Rev Infect Dis 1980;2:370–383.
6. Gupta RK, Relyveld EH, Lindblad EB, Bizzini B, Ben-Efraim S, Gupta CK. Adjuvants—a balance between toxicity and adjuvanticity. Vaccine 1993;11:293–306.
7. Gupta RK, Siber GR. Adjuvants for human vaccines—current status, problems and future prospects. Vaccine 1995;13:1263–1276.
8. Cox JC, Coulter AR. Adjuvants—a classification and review of their modes of action. Vaccine 1997;15:248–256.

9. Singh M, O'Hagan D. Advances in vaccine adjuvants. Nat Biotechnol 1999;17:1075–1081.
10. Edelman R. An overview of adjuvant use. Methods Mol Med 2000;42:1–27.
11. O'Hagan DT, MacKichan ML, Singh M. Recent developments in adjuvants for vaccines against infectious diseases. Biomol Eng 2001;18:69–85.
12. Singh M, O'Hagan DT. Recent advances in vaccine adjuvants. Pharm Res 2002;19: 715–728.
13. Kenney RT, Edelman R. Survey of human-use adjuvants. Expert Rev Vaccines 2003;2:167–188.
14. Garçon N, Chomez P, Van Mechelen M. GlaxoSmithKline adjuvant systems in vaccines: concepts, achievements and perspectives. Expert Rev Vaccines 2007;6:723–739.
15. Wilson-Welder JH, Torres MP, Kipper MJ, Mallapragada SK, Wannemuehler MJ, Narasimhan B. Vaccine adjuvants: current challenges and future approaches. J Pharma Sci 2009;98:1278–1316.
16. O'Hagan DT, De Gregorio E. The path to a successful vaccine adjuvant—"the long and winding road." Drug Discov Today 2009;14:541–551.
17. Dey AK, Srivastava IK. Novel adjuvants and delivery systems for enhancing immune responses induced by immunogens. Expert Rev Vaccines 2011;10(2):227–251.
18. Garçon N, Leroux-Roels G, Cheng W-F. Vaccine adjuvants. Perspect Vaccinol 2011;1:89–113.
19. Aprile MA, Wardlaw AC. Aluminum compounds as adjuvants for vaccine and toxoids in man: a review. Can J Pubilc Health 1966;57:343–354.
20. Gupta RK, Rost BE, Relyveld E, Siber GR. Adjuvant properties of aluminum and calcium compounds. In: Powell MF, Newman MJ, editors. Vaccine Design: The Subunit and Adjuvant Approach. New York: Plenum; 1995, p. 229–248.
21. Gupta RK. Aluminum compounds as vaccine adjuvants. Adv Drug Delivery Rev 1998;32:155–172.
22. Gupta RK, Rost BE. Aluminum compounds as vaccine adjuvants. Methods Molecular Med 2000;42:65–89.
23. Lindblad EB. Aluminium compounds for use in vaccines. Immunol Cell Biol 2004;82:497–505.
24. HogenEsch H. Mechanisms of stimulation of the immune response by aluminum adjuvants. Vaccine 2002;20(Suppl 3):S34–S39.
25. Hem SL, HogenEsch H. Relationship between physical and chemical properties of aluminum-containing adjuvants and immunopotentiation. Expert Rev Vaccines 2007;6:685–698.
26. Duthie MS, Windish HP, Fox CB, Reed SG. Use of defined TLR ligands as adjuvants within human vaccines. Immunol Rev 2011;239:178–196.
27. De Gregorio E, Tritto E, Rappuoli R. Alum adjuvanticity: unraveling a century old mystery. Eur J Immunol 2008;38:2068–2071.
28. Gangarosa EJ, Galazka AM, Wolfe CR, Phillips LM, Gangarosa RE, Miller E, et al. Impact of anti-vaccine movements on pertussis control: the untold story. Lancet 1998;351:356–361.
29. Omer SB, Salmon DA, Orenstein WA, deHart MP, Halsey N. Vaccine refusal, mandatory immunization, and the risks of vaccine-preventable diseases. N Engl J Med 2009;360:1981–1988.

30. Freunds J, Casals J, Hosmer EP. Sensitization and antibody formation after injection of tubercle bacilli and paraffin oil. Proc Soc Exp Biol Med 1937;37:509–513.
31. Relyveld EH. Detoxification of microbial toxins with glutaraldehyde and their use in the preparation of vaccines. In: Rosenberg P, editor. Toxins: Animal, Plant and Microbial. Oxford: Pergamon Press; 1978, p. 1049–1065.
32. Relyveld EH. Current developments in production and testing of tetanus and diphtheria vaccines. In: Mizrahi A, Hertman I, Klingberg M, Kohn A, editors. New Developments with Human and Veterinary Vaccines. New York: Alan R. Liss; 1980, p. 51–76.
33. Relyveld EH, Ben-Efraim S. Preparation of vaccines by the action of glutaraldehyde on toxins, bacteria, viruses, allergens, and cells. In: Langone JJ, van Vunakis H, editors. Methods in Enzymology. New York: Academic Press; 1983, p. 24–60.
34. Michael JG, Pesce AJ, Flanagan M. Enhanced immunogenicity of cationized proteins. In: Spriggs DR, Koff WC, editors. Topics in Vaccine Adjuvant Research. Boca Raton, FL: CRC Press; 1991, p. 99–107.
35. Herrington DA, Clvde DF, Losonskv G, et al. Safety and immunogenicity in man of a synthetic peptide in malaria vaccine against *Plasmodium falciparum* sporozoites. Nature (London) 1987;328:257–259.
36. Nash H, Talwar GP, Segal SJ, et al. Observations on the antigenicity and clinical effects of a candidate anti-pregnancy vaccine: beta subunit of human chorionic gonadotropin linked to tetanus toxoid. Fertil Steril 1980;34:328–335.
37. Robbins JB, Schneerson R. Polysaccharide-protein conjugates: a new generation of vaccines. J Infect Dis 1990;161:821–832.
38. Arakawa T. Adjuvants: no longer a 'dirty little secret', but essential key players in vaccines of the future. Expert Rev Vaccines 2011;10:1–5.
39. Iyer S, Robinett RSR, HogenEsch H, Hem SL. Mechanism of adsorption of hepatitis B surface antigen by aluminum hydroxide adjuvant. Vaccine 2004;22:1475–1479.
40. Hansen B, Belfast M, Soung G, Song L, Egan PM, Capen R, et al. Effect of the strength of adsorption of hepatitis B surface antigen to aluminum hydroxide adjuvant on the immune response. Vaccine 2009;27:888–892.
41. Egan PM, Belfast MT, Giménez JA, Sitrin RD, Mancinelli RJ. Relationship between tightness of binding and immunogenicity in an aluminum-containing adjuvant-adsorbed hepatitis B vaccine. Vaccine 2009;27:3175–3180.
42. van Ramshorst JD. The adsorption of diphtheria toxoid on aluminum phosphate. Recueil Travaux Chimiques des Pays-Bas 1949;68:169–180.
43. Holt LB. Developments in Diphtheria Prophylaxis. London: William Heinemann Medical Books; 1950, p. 1–181.
44. Lindblad EB. Aluminium adjuvants. In: Stewart-Tull DES, editor. The Theory and Practical Application of Adjuvants. Chichester, UK: Wiley; 1995, p. 21–35.
45. Souza Santos P, Vallejo-Freire A, Furlanetto RS, Andrade MC. Correlation between the adsorption of diphtheria toxoid and alizarin by aluminum hydroxide hydrate gels. Mem Inst Butantan 1957;28:221–231.
46. Al-Shakhshir RH, Regnier FE, White JL, Hem SL. Contribution of electrostatic and hydrophobic interactions to the adsorption of proteins by aluminium-containing adjuvants. Vaccine 1995;13:41–44.
47. Lindblad EB, Sparck JV. Basic concepts in the application of immunological adjuvants. Scand J Lab Anim Sci 1987;14:1–13.

48. Bomford R. Aluminium salts: perspectives in their use as adjuvants. In: Gregoriadis G, Allison AC, Poste G, editors. Immunological Adjuvants and Vaccines. London: Plenum; 1989, p. 35–41.
49. Tritto E, Mosca F, De Gregorio E. 2009. Mechanism of action of licensed vaccine adjuvants. Vaccine 27:3331–3334.
50. Mbow ML, De Gregorio E, Ulmer JB. Alum's adjuvant action: grease is the word. Nat Med 2011;17:415–416.
51. Glenny AT, Buttle GAH, Stevens MF. Rate of disappearance of diphtheria toxoid injected into rabbits and guinea pigs: toxoid precipitated with alum. J Pathol Bacteriol 1931;34:267–275.
52. Gupta RK, Chang A-C, Griffin P, Rivera R, Siber GR. In vivo distribution of radioactivity in mice after injection of biodegradable polymer microspheres containing $^{14}$C-labeled tetanus toxoid. Vaccine 1996;14:1412–1416.
53. White RG, Coons AH, Connolly JM. Studies on antibody production III. The alum granuloma. J Exp Med 1955;102:73–82.
54. Harrison WT. Some observations on the use of alum-precipitated diphtheria toxoid. Am J Publ Health 1935;25:298–300.
55. White RG. Concepts relating to the mode of action of adjuvants. Symp Series Immunobiol Stand 1967;6:3–12.
56. Eldridge JH, Staas JK, Meulbroek JA, Tice TR, Gilley RM. Biodegradable and biocompatible poly (DL-lactide-co-glycolide) microspheres as an adjuvant for staphylococcal enterotoxin B toxoid which enhances the level of toxin-neutralizing antibodies. Infect Immun 1991;59:2978–2986.
57. Mannhalter JW, Neychev HO, Zlabinger GJ, Ahmad R, Eibl MM. Modulation of the human immune response by the non-toxic and nonpyrogenic adjuvant aluminium hydroxide: effect of antigen uptake and antigen presentation. Clin Exp Immunol 1985;61:143–151.
58. Flebbe LM, Braley-Mullen H. Immunopotentiating effects of the adjuvants SGP and Quil A.I. Antibody responses to T-dependent and T-independent antigens. Cell Immunol 1986;99:119–127.
59. Walls RS. Eosinophil response to alum adjuvants: involvement of T cells in non-antigen-dependent mechanisms. Proc Soc Exp Biol Med 1977;156:431–435.
60. Ramanathan VD, Badenoch-Jones P, Turk JL. Complement activation by aluminium and zirconium compounds. Immunology 1979;37:881–888.
61. Jordan MB, Mills DM, Kappler J, Marrack P, Cambier JC. Promotion of B cell immune responses via an alum-induced myeloid cell population. Science 2004;304:1808–1810.
62. Grun JL, Maurer PH. Different T helper cell subsets elicited in mice utilizing two different adjuvant vehicles: the role of endogenous interleukin-1 in proliferative responses. Cell Immunol 1989;121:134–145.
63. Cooper PD. The selective induction of different immune responses by vaccine adjuvants. In: Ada GR, editor. Strategies in Vaccine Design. Austin, TX: R.G. Lands; 1994, p. 125–158.
64. Horowitz S, Smolarsky M, Arnon R. Protection against *Schistosoma mansoni* achieved by immunization with sonicated parasite. Eur J Immunol 1982;12:327–332.
65. Alving CR, Glass M, Detrick B. Summary: adjuvants/clinical trials working group. AIDS Res Human Retroviruses 1992;8:1427–1430.

66. Bomford R. The comparative selectivity of adjuvants for humoral and cell-mediated immunity. II. Effect on delayed-type hypersensitivity in the mouse and guinea pig, and cell-mediated immunity to tumour antigens in the mouse of Freund's incomplete and complete adjuvants, alhydrogel, *Corynebacterium parvum, Bordetella pertussis*, muramyl dipeptide and saponin. Clin Exp Immunol 1980;39:435–443.
67. Li H, Nookala S, Re F. Aluminum hydroxide adjuvants activate caspase-1 and induce IL-1beta and IL-18 release. J Immunol 2007;178:5271–5276.
68. Wack A, Baudner BC, Hilbert AK, Manini I, Nuti S, Tavarini S, et al. Combination adjuvants for the induction of potent, long-lasting antibody and T-cell responses to influenza vaccine in mice. Vaccine 2008;26:552–561.
69. Sharp FA, Ruane D, Claass B, Creagh E, Harris J, Malyala P, et al. Uptake of particulate vaccine adjuvants by dendritic cells activates the NALP3 inflammasome. Proc Natl Acad Sci USA 2009;106:870–875.
70. Eisenbarth SC, Colegio OR, O'Connor W, Sutterwala FS, Flavell RA. Crucial role for the Nalp3 inflammasome in the immunostimulatory properties of aluminium adjuvants. Nature 2008;453:1122–1126.
71. Franchi L, Nunez G. The Nlrp3 inflammasome is critical for aluminium hydroxide-mediated IL-1β secretion but dispensable for adjuvant activity. Eur J Immunol 2008;38:2085–2089.
72. Kool M, Soullié T, van Nimwegen M, Willart MAM, Muskens F, Jung S, et al. Alum adjuvant boosts adaptive immunity by inducing uric acid and activating inflammatory dendritic cells. J Exp Med 2008;205:869–882.
73. Kool M, Pétrilli V, De Smedt T, Rolaz A, Hammad H, van Nimwegen M, et al. Cutting edge: alum adjuvant stimulates inflammatory dendritic cells through activation of the NALP3 inflammasome. J Immunol 2008;181:3755–3759.
74. Flach TL, Ng G, Hari A, Desrosiers MD, Zhang P, Ward SM, et al. Alum interaction with dendritic cell membrane lipids is essential for its adjuvanticity. Nat Med 2011;17:479–487.
75. Hornung V, Bauernfeind F, Halle A, Samstad EO, Kono H, Rock KL, et al. Silica crystals and aluminum salts activate the NALP3 inflammasome through phagosomal destabilization. Nat Immunol 2008;9:847–856.
76. Gupta RK, Siber GR. Comparison of adjuvant activities of aluminium phosphate, calcium phosphate and stearyl tyrosine for tetanus toxoid. Biologicals 1994;22:53–63.
77. Gupta RK, Griffin P, Rivera R, Siber GR. Development of an animal model to assess the immunogenicity of single-dose tetanus and diphtheria vaccines based on controlled release from biodegradable polymer microspheres. Dev Biol Stand 1998;92:277–287.
78. McKercher PO, Graves JH. A review of the current status of oil adjuvants in foot-and-mouth disease vaccines. Dev Biol Stand 1977;35:107–112.
79. Freund J, Lipton MM, Pisani TM. Immune response to rabies vaccine in water-in-oil emulsion. Proc Soc Exp Biol 1948;68:609–610.
80. Ott GL, Yaillen AJ, Korschak J. Adjuvant killed hog cholera vaccine. Vet Med 1962;57:1060–1063.
81. McClelland L, Hampil B, Hamparian VV, Potahs L, Ketler A, Hilleman MR. Laboratory and field investigations of bovine myxovirus parainfluenza 3 virus and vaccine. II. Development and appraisal of potency of SF-4 (shipping fever) virus vaccine. J lmmunol 1961;87:134–138.

82. Hilleman MR. Critical appraisal of emulsified oil adjuvants applied to viral vaccines. Prog Med Virol 1966;8:131–182.
83. Davenport FM. Seventeen years' experience with mineral oil adjuvant influenza virus vaccines. Ann Allergy 1968;26:288–292.
84. Stuart-Harris CH. Adjuvant influenza vaccines. Bull WHO 1969;41:617–621.
85. Meiklejohn G. Adjuvants: present status and future needs. In: Proceedings of the Pan American Health Organization Conference on the Application of Vaccines against Viral, Rickettsial and Bacterial Diseases of Man. Scientific Publication No. 226, Pan American Health Organization, Washington, DC; 1971, p. 413–417.
86. Hobson D. The potential role of immunological adjuvants in influenza vaccines. Postgrad Med J 1973;49:180–184.
87. Salk J, Salk D. Control of influenza and poliomyelitis with killed virus vaccines. Science 1977;195:834–846.
88. Freund J. The mode of action of immunological adjuvants. Adv Tuberc Res 1956;7:130–148.
89. Herbert WJ. Antigenicity of soluble protein in the presence of high levels of antibody: a possible mode of action of the antigen adjuvants. Nature (London) 1966;210:747–748.
90. White RG. Concepts of the mechanism of action of adjuvants. In: Borek F, editor. lmmunogenicity. Amsterdam: North Holland Publishing Company; 1972, p. 122–130.
91. Potter M, Boyce CR. Induction of plasma cell neoplasmas in strain BALBc mice with mineral oil and mineral oil adjuvants. Nature (London) 1962;193:1086–1087.
92. Murray R, Cohen P, Hardegree MC. Mineral oil adjuvants: biological and chemical studies. Ann Allergy 1972;30:146–151.
93. Beebe GW, Simon AH, Vivona LS. Long-term mortality follow-up of army recruits who received adjuvant influenza virus vaccine in 1951–1953. Am J Epidemiol 1972;95:337–346.
94. Ogonuki H, Hashizume S, Takashashi B. Preparation and laboratory tests of oil-adjuvant cholera vaccine. Bull WHO 1967;37:729–736.
95. MacLennan R, Schofield FD, Pittman M, Hardegree MC, Barlie MF. Immunization against neonatal tetanus in New Guinea. Antitoxin response of pregnant women to adjuvant and plain toxoids. Bull WHO 1965;32:683–697.
96. World Health, Organization. Immunological Adjuvants Tech Report Series No 595. Geneva: World Health Organization; 1976, p. 3–40.
97. Hardegree MC, Pittman M. Influence of antigens on release of free fatty acids from Arlacel A (mannide monooleate). Proc Soc Exp Biol Med 1966;123:179–184.
98. Stewart-Tull DES. Freund-type mineral oil adjuvant emulsions. In: Stewart-Tull DES, editor. The Theory and Practical Application of Adjuvants. Chichester, UK: Wiley; 1995, p. 1–19.
99. Pittman M. A life with biological products. Annu Rev Microbiol 1990;44:1–25.
100. Woodhour AF, Metzgar DP, Stim TB, Tytell AA, Hilleman MR. New metabolizable immunological adjuvant for human use. I. Development and animal immune response. Proc Soc Exp Biol Med 1964;116:516–523.
101. Hilleman MR, Beale AJ. Adjuvants for human use. In: Bell R, Torrigiani G, editors. New Approaches to Vaccine Development. Basel: Schwabe; 1983, p. 157–166.
102. Herbert WJ. Multiple emulsion adjuvants. Symp Series lmmunobiol Stand 1967;6:89–94.

103. Fukumi H. Effectiveness and untoward reactions of oil adjuvant influenza vaccines. Symp Series lmmunobiol Stand 1967;6:237–240.
104. Taylor PJ, Miller CL, Pollock TM, Perkins FT, Westwood MA. Antibody response and reactions to aqueous influenza vaccine, simple emulsion vaccine and multiple emulsion vaccine. J Hyg (Camb) 1969;67:485–490.
105. Freestone DS, Hamilton-Smith S, Schild GC, Buckland R, Chinn S, Tyrrell DAJ. Antibody responses and resistance to challenge in volunteers vaccinated with live attenuated, detergent split and oil adjuvant A2/Hong Kong/68 (H3N2) influenza vaccines. J Hyg (Camb) 1972;70:531–535.
106. Kimura J, Nariuchi H, Watanabe T, Matuhasi T, Okayasu I, Hatakeyama S. Studies on the adjuvant effect of water-in-oil-in-water emulsion in sesame oil. I. Enhanced and persistent antibody formation by antigen incorporated into the water-in-oil-in-water emulsion. Jpn J Exp Med 1978;48:149–152.
107. Brugh M, Stone HD, Lipton HW. Comparison of inactivated Newcastle disease viral vaccines containing different emulsion adjuvants. Am J Vet Res 1983;44:722–778.
108. Woodard LF, Jasman RL. Stable oil-in-water emulsions: preparation and use as vaccine vehicles for lipophilic adjuvants. Vaccine 1985;3:137–153.
109. Allison AC, Byars NE. Adjuvants for a new generation of vaccines. In: Woodrow GC, Levine MM, editors. New Generation Vaccines. New York: Marcel Dekker; 1990, p. 129–140.
110. Woodard LF. Surface chemistry and classification of vaccine adjuvants and vehicles. In: Mizrahi A, editor. Bacterial Vaccines, Advances in Biotechnological Processes, vol. 73. New York: Wiley-Liss; 1990, p. 281–306.
111. Hunter R, Olsen M, Buynitzky S. Adjuvant activity of non-ionic block copolymers. IV. Effect of molecular weight and formulation on titre and isotype of antibody. Vaccine 1991;9:250–256.
112. Powell MF, Newman MJ. Vaccine Design: The Subunit and Adjuvant Approach. New York: Plenum Publishing Corporation; 1995.
113. Woodard LF. Adjuvant activity of water-insoluble surfactants. Lab Anim Sci 1989;39:222–225.
114. Masihi KN, Lange W, Brehmer W, Ribi E. Immunobiological activities of non-toxic lipid A: enhancement of nonspecific resistance in combination with trehalose dimycolate against viral infections and adjuvant effects. Int J Immunopharmacol 1986;8:339–345.
115. Ott G, Barchfeld GL, Chernoff D, Radhakrishnan R, van Hoogevest P, van Nest G. MF59: design and evaluation of a safe and potent adjuvant for human vaccines. In: Powell MF, Newman MJ, editors. Vaccine Design: The Subunit and Adjuvant Approach. New York: Plenum Publishing Corporation; 1995, p. 277–296.
116. O'Hagan DT. MF59 is a safe and potent vaccine adjuvant that enhances protection against influenza virus infection. Expert Rev Vaccines 2007;6:699–710.
117. O'Hagan DT, Rappuoli R, De Gregorio E, Tsai T. Del Giudice G. MF59 adjuvant: the best insurance against influenza strain diversity. Expert Rev Vaccines 2011;10:447–462.
118. O'Hagan DT, Ott GS, De Gregorio E, Seubert A. The mechanism of action of MF59 - An innately attractive adjuvant formulation. Vaccine 2012;30:4341–4348.
119. Keitel W, Couch R, Bond N, Adair S, van Nest G, Dekker C. Pilot evaluation of influenza virus vaccine (IVV) combined with adjuvant. Vaccine 1993;11:909–913.

120. Keefer MC, Graham BS, McElrath MJ, Matthews TJ, Stablein DM, Corey L, et al. Safety and immunogenicity of Env 2–3, a human immunodeficiency virus type 1 candidate vaccine, in combination with a novel adjuvant MTP-PE/MF59. NIAID AIDS Vaccine Evaluation Group. AIDS Res Hum Retroviruses 1996;12:683–693.
121. Ott G, Barchfeld G, van Nest G. Enhancement of humoral response against human influenza vaccine with the simple submicron oil/water emulsion adjuvant MF59. Vaccine 1995;13:1557–1562.
122. Cataldo DM, van Nest G. The adjuvant MF59 increases the immunogenicity and protective efficacy of subunit influenza vaccine in mice. Vaccine 1997;15:1710–1715.
123. Corey L, Langenberg AGM, Ashley R, Sekulovich RE, Izu AE, Douglas AM Jr. et al. Recombinant glycoprotein vaccine for the prevention of genital HSV-2 infection: two randomized controlled trials. JAMA 1999;282:331–340.
124. Straus SE, Wald A, Kost RG, McKenzie R, Langenberg AGM, Hohman P, et al. Immunotherapy of recurrent genital herpes with recombinant herpes simplex virus type 2 glycoproteins D and B: results of a placebo-controlled vaccine trial. J Infect Dis 1997;176:1129–1134.
125. Langenberg AG, Burke RL, Adair SF, Sekulovich R, Tigges M, Dekker CL, et al. A recombinant glycoprotein vaccine for herpes simplex virus type 2: safety and immunogenicity. Ann Intern Med 1995;122:889–898.
126. Mascola JR. Herpes simplex virus vaccines—why don't antibodies protect? JAMA 1999;282:379–380.
127. Nitayaphan S, Khamboonruang C, Sirisophana N, et al. A phase I/II trial of HIV SF2 gp120/MF59 vaccine in seronegative Thais. Vaccine 2000;18:1448–1455.
128. Heineman TC, Clements-Mann ML, Poland GA, et al. A randomized, controlled study in adults of the immunogenicity of a novel hepatitis B vaccine containing MF59 adjuvant. Vaccine 1999;17:2769–2778.
129. Morel S, Didierlaurent A, Bourguignon P, Delhaye S, Baras B, Jacob V, et al. Adjuvant system AS03 containing α-tocopherol modulates innate immune response and leads to improved adaptive immunity. Vaccine 2011;29:2461–2473.
130. Leroux-Roels I, Borkowski A, Vanwolleghem T, Dramé M, Clement F, Hons E, et al. Antigen sparing and cross-reactive immunity with an adjuvanted rH5N1 prototype pandemic influenza vaccine: a randomised controlled trial. Lancet 2007;370:580–589.
131. Leroux-Roels I, Bernhard R, Gérard P, Drame M, Hanon E, Leroux-Roels G. Broad clade 2 cross-reactive immunity induced by an adjuvanted clade 1 rH5N1 pandemic influenza vaccine. PLoS One 2008l3:e1665.
132. Roman F, Vaman T, Gerlach B, Markendorf A, Gillard P, Devaster, J-M. Immunogenicity and safety in adults of one dose of influenza A H1N1v 2009 vaccine formulated with and without AS03A-adjuvant: preliminary report of an observer-blind, randomised trial. Vaccine 2010;28:1740–1745.
133. Allison AG. Gregoriadis G. Liposomes as immunological adjuvants. Nature 1974; 252(5480):252.
134. van Rooijen N, van Nieuwmegen R. Liposomes in immunology: evidence that their adjuvant effect results from surface exposition of the antigens. Cell Immunol 1980; 49(2):402–407.
135. Gregoriadis G, Davis D, Davies A. Liposomes as immunological adjuvants: antigen incorporation studies. Vaccine 1987;5(2):145–151.

136. Davis D, Davies A, Gregoriadis G. Liposomes as adjuvants with immunopurified tetanus toxoid: the immune response. Immunol Lett 1987;14(4):341–348.
137. Alving CR. Liposomes as carriers of antigens and adjuvants. J Immunol Methods 1991;140(1):1–13.
138. Alving CR. Lipopolysaccharide, lipid A, and liposomes containing lipid A as immunologic adjuvants. Immunobiology 1993;187(3-5):430–446.
139. Gupta RK, Varanelli C, Griffin P, Wallach DFH, Siber GR. Adjuvant properties of non-phospholipid liposomes (Novasomes®) in experimental animals for human vaccine antigens. Vaccine 1996;14:219–225.
140. Pietrobon PJF. Liposome design and vaccine development. In: Powell MF, Newman MJ, editors. Vaccine Design: The Subunit and Adjuvant Approach. New York: Plenum Publishing; 1995, p. 347–361.
141. Gluck R, Mischler R, Brantschen S, Just M, Althaus B, Cryz SJ Jr. Immunopotentiating reconstituted influenza virus virosome vaccine delivery system for immunization against hepatitis. Am J Clin Invest 1992;90:2491–2495.
142. Gluck R, Metcalfe IC. New technology platforms in the development of vaccines for the future. Vaccine 2002;20(Suppl 5):B10–B16.
143. Butts C, Murray N, Maksymiuk A, et al. Randomized phase IIB trial of BLP25 liposome vaccine in stage IIIB and IV non-small-cell lung cancer. J Clin Oncol 2005;23(27):6674–6681.
144. North S, Butts C. Vaccination with BLP25 liposome vaccine to treat non-small cell lung and prostate cancers. Expert Rev Vaccines 2005;4(3):249–257.
145. Sangha R, North S. L-BLP25: a MUC1-targeted peptide vaccine therapy in prostate cancer. Expert Opin Biol Ther 2007;7(11):1723–1730.
146. Johnson AG, Gaines S, Landy M. Studies on the 0 antigen of *Salmonella typhosa* V. Enhancement of antibody response to protein antigens by the purified lipopolysaccharide. J Exp Med 1956;103:225–246.
147. Luderitz O, Galanos C, Reitschel ET. Endotoxins of Gram-negative bacteria. Pharm Ther 1982;15:383–402.
148. Quershi N, Takayama K. Structure and function of lipid A. In: The Iglewski BH, Clark VL, editors. Bacteria, vol. XI. Molecular Basis of Bacteria/Pathogenesis. San Diego: Academic Press; 1990, p. 319–338.
149. Ribi E, Amano K, Cantrell JL, Schwartzman S, Parker R, Takayama K. Preparation and antitumor activity of nontoxic lipid A. Cancer Immunol Immunother 1982;12:91–96.
150. Qureshi N, Takayama K, Ribi E. Purification and structural determination of nontoxic lipid A obtained form the lipopolysaccharide of *Salmonella typhimurium*. J Biol Chem 1982;257(19):11808–11815.
151. Baker PJ, Haslov KR, Fauntleroy MB, Stashak PW, Myers K, Ulrich JT. Enrichment of suppressor T cells by means of binding to monophosphoryl lipid A. Infect Immun 1990;58(3):726–731.
152. Rudbach JA, Cantrell JL, Ulrich JT, Mitchell MS. Immunotherapy with bacterial endotoxin. In: Friedman H, Klein TW, Nakano M, Nowotny A, editors. Endotoxin. New York: Plenum; 1990, p. 865–676.
153. Garcon N, Segal L, Tavares F, Van Mechelen M. The safety evaluation of adjuvants during vaccine development: the AS04 experience. Vaccine 2011;29:4453–4459.

154. Garcon N, Wettendorff M, Van Mechelen M. Role of AS04 in human papillomavirus vaccine: mode of action and clinical profile. Expert Opin Biol Ther 2011;11:667–677.
155. Didierlaurent AM, Morel S, Lockman L, Giannini SL, Bisteau M, Carlsen H, et al. AS04, an aluminum salt- and TLR-4 agonist-based adjuvant system, induces a transient localized innate immune response leading to enhanced adaptive immunity. J Immunol 2009;83:6186–6197.
156. Kong NCT, Beran J, Kee SA, Miguel JL, Sánchez C, Bayas JM, et al. Immunogenicity and safety of an adjuvanted hepatitis B vaccine in pre-hemodialysis and hemodialysis patients. Kidney Int 2005;68:2298–2303.
157. Paavonen J, Naud P, Salmeron J, et al. Efficacy of human papillomavirus (HPV)-16/18 AS04-adjuvanted vaccine against cervical infection and precancer caused by oncogenic HPV types (PATRICIA): final analysis of a double-blind, randomised study in young women. Lancet 2009;374:301–314.
158. Dauvilliers Y, Montplaisir J, Cochen V, Desautels A, Einen M, Lin L, Kawashima M, Bayard S, Monaca C, Tiberge M, Filipini D, Tripathy A, Nguyen BH, Kotagal S, Mignot E. Post-H1N1 narcolepsy-cataplexy. Sleep 2010; 33: 1428–1430.
159. National Institute for Health and, Welfare. National Institute for Health and Welfare recommends discontinuation of Pandemrix vaccinations. National Institute for Health and Welfare; 2010. Available from: www.thl.fi/en_US/web/en/pressrelease?id=22930.
160. Nohynek H, Jokinen J, Partinen M, VaaralaO, Kirjavainen T, et al. AS03 adjuvanted AH1N1 vaccine associated with an abrupt increase in the incidence of childhood narcolepsy in Finland. PLoS One 2012;7:e33536. doi:10.1371/journal.pone.0033536.
161. Partinen M, Saarenpää-Heikkilä O, Ilveskoski I, Hublin C, Linna M, Olsén P, et al. Increased incidence and clinical picture of childhood narcolepsy following the 2009 H1N1 pandemic vaccination campaign in Finland. PLoS One 2012;7:e33723. doi:10.1371/journal.pone.0033723.
162. Miller E, Andrews N, Stellitano L, Stowe J, Winstone AM, Shneerson J, Verity C. Risk of narcolepsy in children and young people receiving AS03 adjuvanted pandemic A/H1N1 2009 influenza vaccine: retrospective analysis. Br Med J 2013;346:f794.
163. Tsai TF, Crucitti A, Nacci P, Nicolay U, Cioppa GD, Ferguson J, Clemens R. Explorations of clinical trials and pharmacovigilance databases of MF59®-adjuvanted influenza vaccines for associated cases of narcolepsy. Scand J Infect Dis 2011;43:702–706.
164. Mignot E, Lin L, Rogers W, et al. Complex HLA-DR and -DQ interactions confer risk of narcolepsy-cataplexy in three ethnic groups. Am J Hum Genet 2001;68:686–699.
165. Hallmayer J, Faraco J, Lin L, et al. Narcolepsy is strongly associated with the T-cell receptor alpha locus. Nat Genet 2009;41:708–711.
166. Cvetkovic-Lopes V, Bayer L, Dorsaz S, et al. Elevated Tribbles homolog 2-specific antibody levels in narcolepsy patients. J Clin Invest 2010;120:713–719.

# Methods for Characterizing Proteins in Aluminum Adjuvant Formulations

MARTINUS A.H. CAPELLE, EMILIE POIRIER, and TUDOR ARVINTE

## 15.1 INTRODUCTION

Bovine serum albumin (BSA) mixed with aluminum hydroxide adjuvant (Alhydrogel) was used as a vaccine model for physicochemical studies. This chapter presents methods that allow studies of protein in vaccine formulations without separation of the protein from the aluminum salt adjuvant, including (1) UV-Vis transmission, (2) 90-degree light scatter, (3) fluorescence spectroscopy, (4) time-correlated single photon counting fluorescence lifetime, (5) light microscopy, (6) attenuated total reflection Fourier-transform infrared spectroscopy (ATR-FTIR), (7) UV resonance Raman spectroscopy, (8) sedimentation by gravitation, (9) analytical centrifugation with the LUMiSizer, (10) circular dichroism (CD), and (11) ultrasound resonator technology. Advantages of using fluorescence emission compared with UV-Vis absorbance for the determination of amount of protein bound to the adjuvant are discussed. The presented methods can be applied to characterize the vaccines' stability and improve the formulation. The physicochemical properties (e.g., conformation, binding strengths, type of aggregates or formed complexes) together with *in vivo* activity data will contribute to a better understanding of the mechanism(s) of action that can be applied to design more potent vaccines.

Few methods are described in literature to study the physical and chemical properties of vaccines consisting of a protein adsorbed onto an aluminum salt adjuvant. In this chapter, different methods are discussed for the characterization of BSA as a model protein after adsorption onto aluminum hydroxide adjuvant, Alhydrogel. The properties of BSA adsorbed onto the adjuvant are compared with BSA in water and to the adjuvant alone. Conformation and stability of proteins might change after adsorption onto aluminum adjuvants. The adjuvant properties,

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*,
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

such as turbidity, morphology, compressibility, mass, and size, might also change after adsorption of a protein. The final vaccine obtained by mixing the antigen with the adjuvant will likely have different physical and chemical properties than the individual components before mixing.

Adjuvants are usually incorporated into vaccines to effectively stimulate and enhance the immune response. Aluminum adjuvants have been the most frequently used adjuvants in both human and veterinary vaccines for more than 60 years [1]. These adjuvants are generally recognized as safe, but despite their widespread use, their mechanism of action is still not completely understood. Aluminum-based adjuvants are low cost and have shown adjuvanticity with a variety of antigens [2]. The two mainly used aluminum adjuvants are Alhydrogel (aluminum hydroxide) and Adju-Phos (aluminum phosphate). Alhydrogel typically exists in a crystalline state and is mineralogically known as boehmite, which consists of corrugated sheets of octahedral aluminum [3,4]. Aluminum hydroxide particles are known to have a high surface area, about 500 $m^2/g$ [5], and needle-like or fibril morphology with a diameter of approximately 2 $nm^3$. In general, proteins adsorb better to oppositely charged adjuvants through electrostatic interactions [6]. Aluminum hydroxide has an excellent adsorptive capacity for many proteins because of the point of zero charge (PZC, analogous to the isoelectric point [IEP] of a protein) of approximately 11.

Bovine serum albumin adsorbed on Alhydrogel was used as a model protein vaccine for physicochemical studies. BSA adsorption on Alhydrogel is favored because of the electrostatic interaction between the negatively charged BSA (IEP of 4.7) and the positively charged nature of the aluminum hydroxide. Adsorption of negatively charged ovalbumin to negatively charged adjuvants has been observed, although the adsorptive capacity was greatest when the surface charge was positive [7]. Hem and coworkers proposed that structurally flexible proteins, such as ovalbumin, display more complex adsorption behavior than structurally rigid proteins, such as lysozyme, for which adsorptive behavior can be explained by electrostatic forces [7]. Other forces that contribute to the adsorption of proteins on aluminum adjuvants are hydrophobic, van der Waals, and hydrogen bonding [2].

Protein adsorbed to the adjuvant leads to more immunostimulation than unadsorbed protein [1]. Some of the described mechanisms involved in this process are the rendering of the protein multimeric, the delivery of multiple copies to the same cell. and the targeting of the immune system [8]. The main analytical studies performed on the final vaccine often consist of the determination of adsorption profiles and biological activity [1]; however, these biological assays are generally time consuming and of low accuracy and precision [9]. They also provide no direct information on the mechanisms of destabilization [9].

Chemical and physical stability studies can provide insight into the origin of the change responsible for vaccine inactivation. Within the academic and industrial communities, there is a need of high-resolution analytical methods that can detect small changes in complex macromolecular systems of multiple components, as was also stated by Middaugh and coworkers [9]. Previous studies from our group on

MMA383 antibody vaccine [10,11] and from another group on ricin A–chain vaccine [12] have highlighted the importance of formulation and characterization in regard to the immune response. Recently, several articles have appeared that report studies of the structure and stability of proteins after adsorption to aluminum adjuvants using calorimetry, fluorescence, electron microscopy, and infrared spectroscopy [10,11,13,14].

In this chapter, vaccine analyses were performed using the following spectroscopy and microscopy methods: (1) UV-Vis transmission, (2) 90-degree light scatter, (3) fluorescence spectroscopy, (4) time-correlated single photon counting fluorescence lifetime, (5) attenuated total reflection Fourier-transform infrared spectroscopy, and (6) light microscopy. In addition, new methods for characterizing protein adsorbed onto adjuvant vaccines are discussed: (1) UV resonance Raman spectroscopy, (2) CD, (3) ultrasound resonator technology, (4) sedimentation by gravitation, and (5) analytical centrifugation with the LUMiSizer. The majority of the developed methods are noninvasive and might also be applied for studying complex formulations such as virosomes, virus-like particles, nanoparticles, crystals, hydrogels, and other drug delivery systems. The methods were adapted with the aim of detecting small differences against a complex background and to be used for the characterization of vaccine formulations without dilution or protein desorption in conditions as near as possible to those used in medical applications [15].

## 15.2 ANALYTICAL METHODS

### 15.2.1 Sample Preparation

The samples were prepared by mixing 2% Alhydrogel with a BSA solution. The protein was added on top of the adjuvant. All measurements, except for the CD, were done on samples containing 3.4 mg/ml aluminum and 1.7 mg/ml BSA (0.5 mg/ml BSA per mg/ml aluminum). For the CD measurements, samples with a lower protein (0.1 mg/ml BSA) and adjuvant concentration (0.2% Alhydrogel) were used to avoid saturation of the detector. In both cases, no BSA was detected in the supernatant after centrifugation, indicating that all protein was either bound to the adjuvant or otherwise present within the pellet. This is expected as the protein concentration is below the adsorptive capacity of 2.2 mg BSA per milligram of aluminum of Alhydrogel in MOPS pH 7.4 [13]. The adsorptive capacity is defined as the maximum mass of protein adsorbed per milligram of aluminum.

### 15.2.2 UV-Vis Transmission

UV-Vis transmission spectra of BSA in water, BSA–Alhydrogel, and Alhydrogel are shown in Figure 15.1a. Measurements were performed with a Cintra 40 UV-Vis spectrometer (GBC Scientific, Dandenong, Australia) and in quartz cuvettes with 0.1-cm path length at 25°C. Adsorption of BSA onto the aluminum adjuvant resulted in a more turbid mixture (increased absorbance) and with a slope that is more parallel

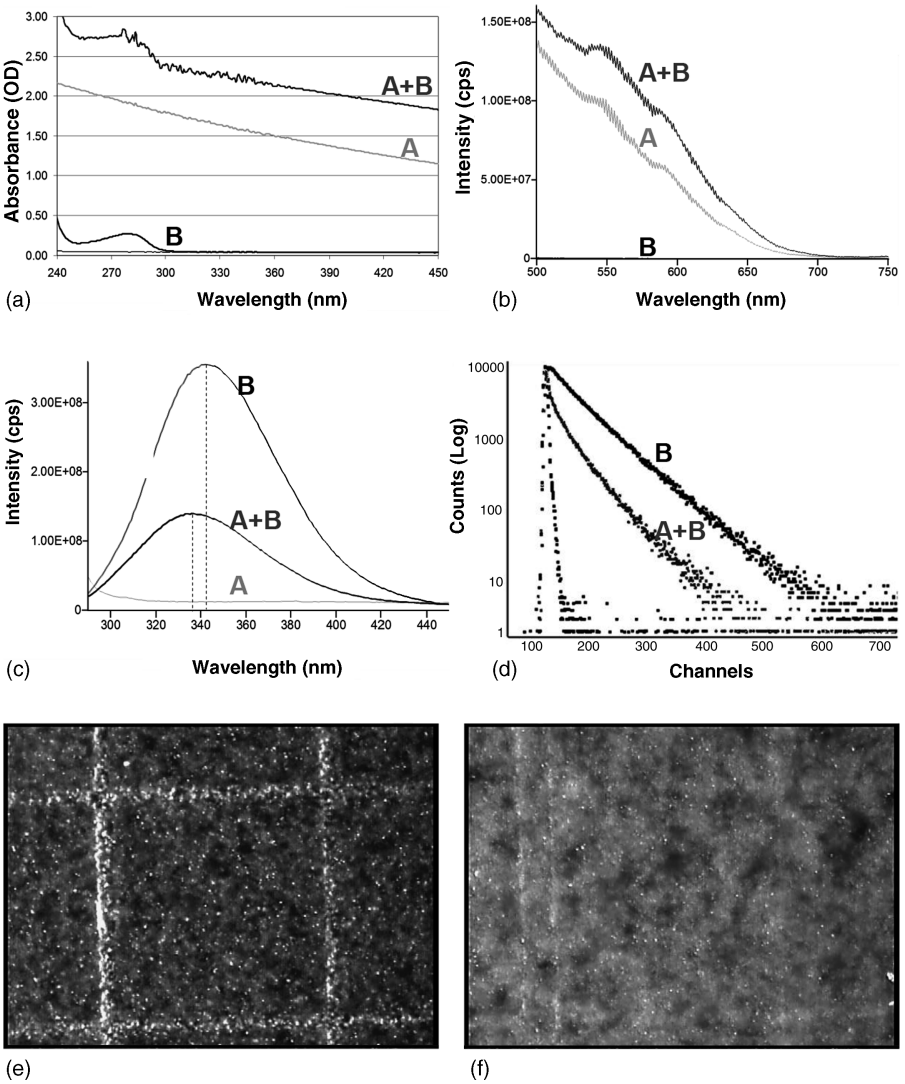


**FIGURE 15.1** 1.7 mg/ml bovine serum albumin (BSA) in water (B), 1.7 mg/ml BSA with 0.6% Alhydrogel (A + B), 0.6% Alhydrogel (A) and water were analyzed by (a) UV-Vis absorbance spectroscopy, (b) 90-degree light scattering, (c) steady-state fluorescence emission (the wavelength with maximum emission intensities for BSA and BSA with Alhydrogel were 336 and 343 nm, respectively), and (d) steady-state fluorescence lifetime. The black decay is the light-scatter signal. (e and f) Light microscopy pictures of 0.6% Alhydrogel and BSA–Alhydrogel, respectively; the background square is 250 μm × 250 μm.

to the $x$-axis compared with the adjuvant without protein. The addition of protein to the adjuvant resulted in an increase in particle size, which resulted in a stronger Mie and Thompson scatter. A similar difference in scatter was also observed using the Tecan Safire plate reader (data not shown). In case of no interaction between the protein and adjuvant, the absorbance spectrum would be the sum of the two individual spectra. UV-Vis absorbance measurements in 96-well plates can be performed to determine the turbidity of vaccines similar to the high-throughput turbidimetry [16] and nephelometry [17] methods.

### 15.2.3 Ninety-Degree Light Scatter

Ninety-degree light-scatter measurements were performed with a Fluoromax spectrofluorometer (Spex Instruments S.A., Stanmore, UK). Hellma quartz cuvettes were used with a volume of 0.5 ml and a path length of $1.0 \times 0.2$ cm. The 90-degree light-scatter scan is a synchronous scan (i.e., the intensity was monitored at the same wavelength as the excitation light). The measurement range was nm to 750 nm with slits of 0.5 mm for the excitation and emission. Spectra were recorded with 0.01-s integration time per 1-nm increment. To avoid saturation of the detector, a neutral 90% absorbing density filter was placed in front of the detector.

Ninety-degree light-scatter intensity depends on the refractive index, particle size, and concentration. The BSA–Alhydrogel sample scattered significantly more light than the Alhydrogel sample without protein (Fig. 15.1b). The major contributor to the scatter signal is the adjuvant. Because the adjuvant concentration is the same, we can assume that the difference in scatter signal is caused by an increase in particle size after addition of protein to the adjuvant. In our experiments, we used 0.6% Alhydrogel, a relatively high adjuvant concentration. In such cases, to avoid saturation caused by strong scattering, the sample measurements can be performed at higher wavelengths and using short path length cuvettes. At lower adjuvant concentrations, the filter may not be needed. Also, at lower adjuvant concentrations of 0.3% Alhydrogel, we observed that the formulation can induce large differences in light-scatter properties of an antibody-Alhydrogel vaccine [10].

### 15.2.4 Fluorescence Spectroscopy

Steady-state fluorescence measurements were performed with a Fluoromax spectrofluorometer. The fluorescence intensities were measured at 25°C in a thermostated cuvette holder. Hellma quartz cuvettes were used with a volume of 0.5 ml and a path length of $1.0 \times 0.2$ cm: excitation was at 280 nm, and the emission was monitored between 300 and 450 nm. Spectra were recorded with a 0.1-s integration time per 1-nm increment and slits of 0.3 and 1.0 mm for excitation and emission, respectively.

The intrinsic tryptophan fluorescence intensity of BSA decreased and shifted to lower wavelengths after adsorption onto Alhydrogel. The maximum intensity for BSA in solution was at 343 nm and after adsorption onto Alhydrogel was 336 nm (Fig. 15.1c). This blue shift for BSA–Alhydrogel in relation to BSA in

solution is in concordance with the results of Jones et al. [13] and likely due to adsorption-induced conformational changes. The intensity of BSA in water is higher than in the presence of adjuvant, which is largely due to the turbidity of adjuvant containing samples; less fluorescence is recorded by the detector as a large portion is scattered and does not reach the detector. Because of the scatter issue, a comparison of protein adjuvanted vaccines should be mainly based on the position of the emission maximum.

Fluorescence lifetime of the tryptophan of BSA adsorbed on the aluminum adjuvant was measured with a time-correlated single-photon counting (TCSPC) IBH 5000U fluorescence lifetime spectrophotometer (IBH, Glasgow, UK). The samples were excited with a 279-nm NanoLED source. The spectrophotometer is equipped with a monochromator at the emission side. Optimum emission wavelength was determined to be at 345 nm. Data analysis was performed using the DAS6 software (IBH, Glasgow, UK). After reconvolution of the intensity decay with the instrument response function, the calculated data were analyzed by linear and nonlinear least-squares modeling. The mean weighted average lifetimes were calculated using the equation

$$\tau_F = \sum \alpha_i \tau_i^2 \Big/ \sum \alpha_i \tau_i \tag{1}$$

where $\tau_F$ is the fluorescence decay time and $\alpha$ is the normalized pre-exponential factor.

The mean weighted fluorescence lifetime of BSA in water (about 6.2 ns) is significant longer than the lifetimes of BSA–Alhydrogel (5.1 ns) (Table 15.1 and Fig. 15.1d). A difference of 0.9 ns is a large change in fluorescence lifetime; the experimental variation is around 0.05 ns. The blue shift of the intrinsic fluorescence and longer fluorescence lifetime of the BSA after adsorption on the adjuvant indicate that the protein is adsorbed in different conformations. The measurement of fluorescence in these turbid samples provides additional information on the local

**TABLE 15.1 Fluorescence and absorbance of BSA in water and BSA with Alhydrogel***

| | $A_{350}$ | $A_{350}\Delta$ | $I_{max}$ | $\lambda_{max}$ (nm) | $\tau_1$ (ns) | $\alpha_1$ (%) | $\tau_2$ (ns) | $\alpha_2$ (%) | $\tau_3$ (ns) | $\alpha_3$ (%) | $\tau_F$ (ns) |
|---|---|---|---|---|---|---|---|---|---|---|---|
| BSA in water | 0.039 | 0.00 | 1.6E + 08 | 343 | 0.7 | 0.8 | 3.5 | 20.7 | 6.6 | 78.5 | 6.2 |
| BSA with Alhydrogel | 2.57 | 0.38 | 7.1E + 07 | 336 | 1.1 | 6.6 | 2.8 | 36.1 | 5.9 | 57.3 | 5.1 |

*Fluorescence and absorbance data of the bovine serum albumin (BSA)–Alhydrogel and BSA solutions were compared. The UV-Vis absorbance at 350 nm ($A_{350}$) and the UV-Vis absorbance at 350 nm corrected with the absorbance of the adjuvant controls ($A_{350}\Delta$) are listed. The fluorescence emission maximum ($\lambda_{max}$) and the fluorescence intensity at emission maximum ($I_{max}$) were measured using quartz cuvettes and were derived from Figure 15.3. The best fit (X was between 0.9 and 1.1) to fluorescence lifetimes of the BSA formulations was obtained using three exponents; $\tau_1$, $\tau_2$, and $\tau_3$ are the decay times, and $\alpha_1$, $\alpha_2$, and $\alpha_3$ are the exponential factors.

environment of the tyrosine and tryptophan amino acids (e.g., aggregation, conformation, folding, or chemical degradation).

UV-Vis absorbance measurements are commonly performed to determine the antigen concentration in the supernatant. However, for vaccines with low antigen concentration, the UV-Vis absorbance is not sensitive enough for protein concentration determinations, and fluorescence measurements should be performed. The amount of protein bound to the adjuvant is calculated from the concentration in the supernatant after pelleting the aluminum adjuvant by centrifugation. A Tecan (Tecan Group Ltd., Männedorf, Switzerland) Safire 2 microplate reader was used to determine intrinsic fluorescence emission of BSA remaining in the supernatant after centrifugation. Fluorescence emission was measured from the bottom after excitation at 280 nm using bandwidths of 5 nm for both excitation and emission. The fact that fluorescence is more sensitive in protein detection is documented by the following experiment. Three samples consisting of 0.3 mg/ml BSA with 0.025%, 0.03%, and 0.3% Alhydrogel were centrifuged, and the supernatant was measured by UV-Vis absorbance and fluorescence emission in a 96-well plate (Fig. 15.2). No BSA was found in the supernatant of the sample with 0.3% Alhydrogel by both methods. However, the fluorescence emission of the supernatant of BSA with 0.03% Alhydrogel showed a clear fluorescence spectrum, but nothing was detected by UV-Vis absorbance spectroscopy; the fluorescence intensity corresponded to 1% of the initial 0.3 mg/ml BSA (99% bound). For the BSA 0.025% Alhydrogel samples the UV-Vis and fluorescence methods gave values for bound BSA of 84% and 87%, respectively. The fluorescence emission method was more sensitive, easier to reproduce, and less influenced by the scatter of "left-over" adjuvant in the supernatant. Fluorescence spectroscopy is, in our experience, the method of choice for determination of antigen concentration in supernatants.

### 15.2.5 Light Microscopy

Light microscopy observations were performed with a 10× objective on a Leica DM RXE microscope (Leica Microsystems GmbH, Wetzlar, Germany) equipped with a mercury lamp. The images were acquired with a Camera Ecoline Digital and processed using the Cyberlink Power Director software (Cyberlink, Vaals, Netherlands) operating under Windows 7. Microscopy slides with a fixed distance of 0.1 mm between the slide and the cover glass were used. The size of the 4 × 4 grids is 1 mm × 1 mm and corresponds to a volume of 0.1 μl. BSA induced changes in the morphology of the complexes between protein and adjuvant; in the presence of protein, the aluminum structures appear more cloudy and turbid (Fig. 15.1e and f). These induced changes in morphology are immediate and are in agreement with the spectroscopic and light-scatter results.

### 15.2.6 Attenuated Total Reflection–Fourier Transform Infrared Spectroscopy FTIR

The secondary structure of BSA adsorbed on Alhydrogel was characterized using an IdentifyIR (Smiths Detection, Danbury CT, USA) ATR-FTIR spectrometer. Samples

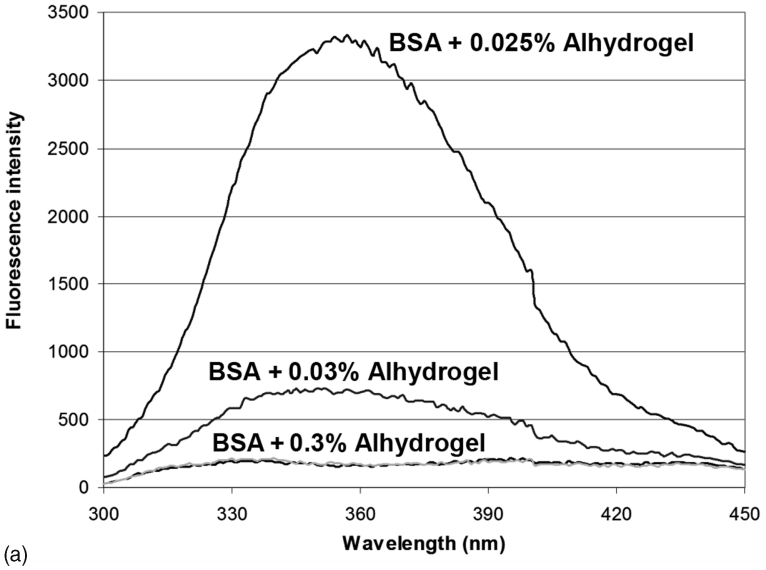


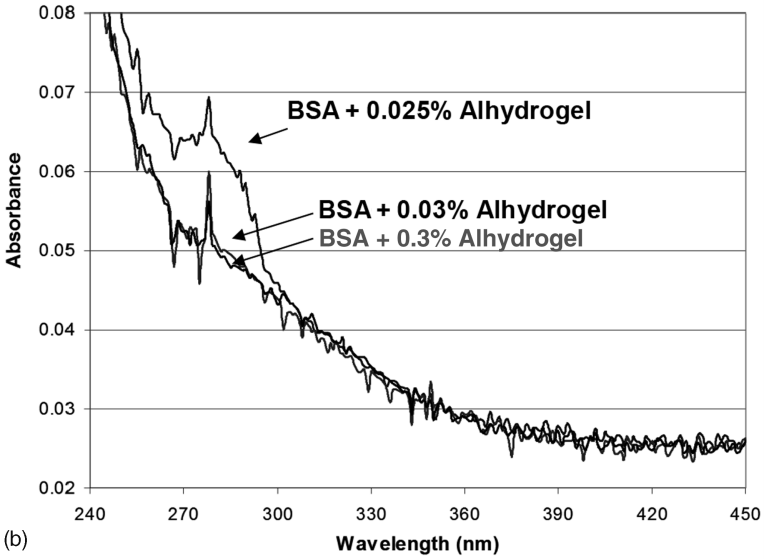


**FIGURE 15.2** Fluorescence emission (a) and UV-Vis absorbance (b) of the supernatant of 0.3 mg/ml bovine serum albumin (BSA) with 0.025% Alhydrogel, 0.03% Alhydrogel, and 0.3% Alhydrogel, as indicated in the figures.

were measured directly in the intact vaccine formulation without any form of sample treatment. The IdentifyIR contained a three reflections diamond-attenuated total reflection (ATR) crystal. In ATR-FTIR, the absorption is measured of the evanescent wave that penetrates the solution when the incident light is reflected at the crystal–liquid interface [18]. The ATR crystal was covered completely with BSA–adjuvant suspension and imaged with the light microscope accessory with a 100x magnification. Data collection were between 4000 and 650 $cm^{-1}$ with a 4–wave number resolution. The average of 256 scans was used. The acquired ATR-FTIR spectra were corrected for absorption of water.

ATR-FTIR spectra showed that the secondary structure of BSA changed after adsorption to Alhydrogel (Fig. 15.3a). The absorption bands that were attributed to the adjuvants structure are between 3000 and 3600 $cm^{-1}$ and around 1100 $cm^{-1}$. The strong shoulder observed in the Alhydrogel samples at 3100 $cm^{-1}$ and the peaks at 3300 and 1100 $cm^{-1}$ are characteristic of aluminum hydroxide and indicates the existence of structural hydroxyl environments [3]. The BSA–Alhydrogel sample had distinct peaks at wave numbers of 2750 to 3000 $cm^{-1}$, which was ascribed to protein-inflicted changes in the structural hydroxyl environments of Alhydrogel and not protein characteristic peaks. No peaks were determined for the BSA in water sample because of the low protein concentration (data not shown); a concentration of about 10 mg/ml BSA is required to obtain a proper ATR-FTIR spectrum. Because the BSA in presence of the adjuvant is no longer in solution as it is adsorbed onto the suspended aluminum particles, the ATR-FTIR intensities are stronger. Therefore, ATR-FTIR spectroscopy measures mainly the adsorbed antigen and less the solubilized antigen.

Amide I (1650 $cm^{-1}$) and amide II (1520 $cm^{-1}$) absorption bands are characteristic for the peptide bonds in proteins. The amide I band for BSA–Alhydrogel is typical for α-helix, and amide II band at 1520 $cm^{-1}$ was assigned to β-sheet [19]. BSA–Alhydrogel had similar intensities for both bands, as was previously reported [13], which can be attributed to presence of both α-helices and β-sheet.

ATR-FTIR spectroscopy for secondary structure analysis of different proteins adsorbed onto adjuvant has been described [13,19–21]. The sample handling, such as loading and cleaning of the crystal, is rapid and simple. Measurements are highly reproducible because of the constant "optical path" that depends on the crystal and solution refractive indices [18,22]. A few crystals or volumes of around 10 μl are sufficient to cover the surface of the crystal. With the IdentifyIR, the crystal surface was continuously monitored by an integrated light microscope and shown on a LCD screen.

### 15.2.7 Ultraviolet Raman Spectroscopy

Ultraviolet resonance Raman spectroscopy with a high-throughput sample setup at the Central Laser Facility of the Rutherford Appleton Laboratory, UK, was used [23,24]. Raman spectra were recorded using a 244-nm laser excitation generated by frequency doubling an Innova FRED 300c Ar+ laser; the experimental conditions were similar to those described by Patois et al. [23]. The Raman spectrum of BSA

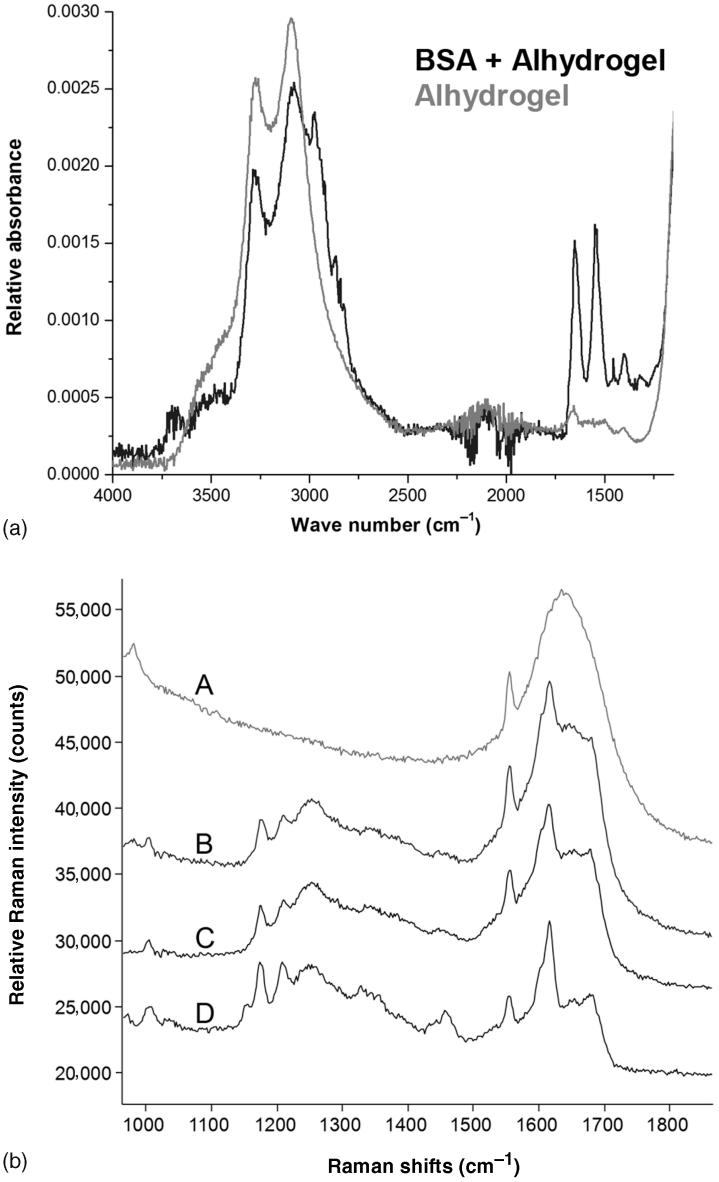


**FIGURE 15.3** 0.6% Alhydrogel (A, gray) and 1.7 mg/ml bovine serum albumin (BSA) with 0.6% Alhydrogel (B, black) were analyzed by (a) ATR-FTIR (attenuated total reflection–Fourier transform infrared) absorbance spectra, BSA–Alhydrogel had distinct amide absorption bands and no absorbance was detected for Alhydrogel around the amide I and II bands; (b) ultraviolet Raman spectroscopy, BSA–Alhydrogel corrected for Alhydrogel (C); and BSA–Alhydrogel corrected for BSA (D).

was obtained after subtracting the adjuvant signal from the BSA–Alhydrogel sample (Fig. 15.3b). In Figure 15.3b, the spectra are presented with an offset and not overlaid as is custom for UV resonance Raman spectroscopy. Interestingly, this BSA spectrum from the BSA–Alhydrogel sample was similar to the Raman spectrum of BSA in water, showing that the overall structure of BSA remained unchanged; the average of the structure of BSA in solution as well as BSA adsorbed onto the adjuvant. These data show that UV resonance Raman spectroscopy is a suitable technique for the investigation of proteins bound to vaccine adjuvants. For identification of the various Raman peaks, refer to our recent publication on Raman spectroscopy [23].

### 15.2.8 Sedimentation by Gravitation

Aluminum salt adjuvants consist of particles in suspensions that can sediment in time. The sedimentation by gravitation can be documented by (1) measurement of 90-degree light scatter [10] or turbidity in time and (2) photographic measurement of the pellet height in time. The first method was applied for an antibody vaccine with Alhydrogel adjuvant in which we reported the measurements of gravitational sedimentation of large particles using 90-degree light scatter [10]. At the same concentration of antibody and adjuvants, different sedimentation properties were observed [10]. The results showed that sedimentation behavior of protein aluminum adjuvant vaccines is influenced by different factors, such as the particle mass, viscosity, density, and buoyancy.

An alternative method to characterize the sedimentation kinetics is the photographic measurement of the pellet height in time. In the present study, the BSA–Alhydrogel vaccines in 1-ml syringes and 3-ml cuvettes were photographed at different time intervals (data not shown) against a black background to record the sedimentation process. Sedimentation equilibrium was observed after 2 weeks at room temperature. Advantages of using syringes include the measurement in small volumes using standard and low-cost laboratory consumables, good control of pellet height, and the possibility to measure noninvasively in prefilled syringes. In the case that the sedimentation is fast (within minutes), the cuvette measurements using 90-degree light scatter provide a more accurate method to study sedimentation kinetics [10].

### 15.2.9 Sedimentation Analysis Using Analytical Centrifugation

Accelerated sedimentation analysis on the BSA adjuvant samples was performed using an analytical centrifuge LUMiSizer 610 (L.U.M. GmbH, Berlin, Germany). The LUMiSizer 610 measured the intensity of the transmitted near infrared (NIR) light as function of time and position over the full sample length. The centrifugal force applied was $146 \times g$. NIR transmission profiles were recorded with an interval of 10 s during 40 min of separation at a thermostated temperature of 20°C. Data were analyzed using the LUMiSizer software SEPView (L.U.M. GmbH) as a function of the position inside the sample as a distance from the center of rotation. The

progression of the transmission profiles contains information on stability and on the kinetics of the sedimentation processes. Transmission profiles are representative of the variation of particle concentration inside the sample (low transmission means high particle concentration; high transmission means low particle concentration).

Sedimentation analysis with the LUMiSizer 610, using centrifugal force and detection in the NIR, can distinguish between BSA–Alhydrogel and Alhydrogel without protein. This analytical centrifugation method is fast (20–40 min) and permits the determination of sedimentation profiles of up to 12 different samples simultaneously [25,26]. The evolution of transmission profiles in time for BSA–Alhydrogel is shown in Figure 15.4a. The separation process was characterized by a sharp sedimentation-front moving toward the bottom of the cell. The decreasing spacing between consecutive profiles of BSA–Alhydrogel (Fig. 15.4a) can be explained by the consolidation of a closed three-dimensional network formed in case of flocculated particles. The resistance against compaction increases as sedimentation proceeds because of increased interactions between the higher formed local concentration of particles. Near the cell bottom, a compact sediment was formed (corresponding to the lowest transmission). The length of the sediment or pellet height and the volume or the filling height are measured directly with the LUMiSizer (Fig. 15.4a). As the sample is centrifuged, it becomes increasingly more clear, and the turbid sediment front moves to the right. The change in time of the distance from the top of the cuvette to the turbid sediment front for a given transmission value, (e.g., at 60% transmission in our case) is called the front-track curve (Fig. 15.4b). The slope in each point (first derivative) of the front-tracking curve defines the sedimentation velocity at that time point. The initial sedimentation velocities (velocity $= \Delta$ Position/$\Delta$ Time) during the first minute for BSA–Alhydrogel and Alhydrogel are 10 and 3 mm/min, respectively (see Fig. 15.4b). In our case, after 30 min, the position of the front no longer changed significantly, and the pellet percentage of total volume was calculated. Figure 15.4b shows that the adjuvant with BSA sediments faster and forms a larger pellet than the adjuvant alone. The mass of the formed BSA–adjuvant complexes is larger than those without the protein, resulting in an increased sedimentation rate. The size of the pellet is not directly correlated to the mass of the particles and depends more on the particle morphology, charges, and colloidal properties. In this case, the larger pellet might be formed because of interactions (e.g., repulsions) between the BSA–adjuvant particles.

These measurements permit a good characterization of vaccines by all of these parameters—the tracking curves pattern, the sedimentation velocities, and the pellet height as a percentage of total volume. Thus, the LUMiSizer provides information on colloidal properties and particle size of turbid vaccines, and it may be also be a useful method for the characterization of other vaccine adjuvants such as microparticles [8,27].

### 15.2.10 Circular Dichroism

Circular dichroism spectra were obtained with a Jasco J-815 spectropolarimeter (Tokyo, Japan) equipped with a thermostated cell holder. Measured samples were

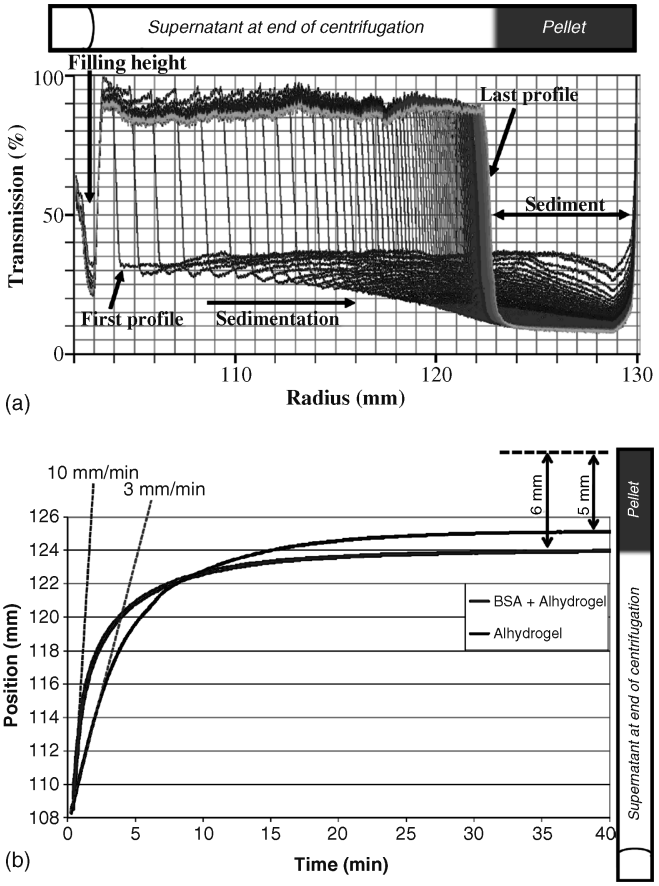


**FIGURE 15.4** LUMiSizer experiments. (a) The changes in time of transmission profiles for 1.7 mg/ml bovine serum albumin (BSA) adsorbed on 0.6% Alhydrogel analyzed during centrifugation at 146 × *g* forces. The separation process is characterized by a sharp sedimentation front moving toward the bottom of the cell. The sharp dip in the transmission profiles at about 103 mm is due to the liquid–air interface and represents the filling height of the sample. The position of the cell bottom is at about 130 mm. The LUMiSizer cuvette is drawn on the top (a) and on the side (b) of the figures. (b) Front-track sedimentation curves by tracking the sedimentation front at 60% transmission value. BSA–Alhydrogel (dark gray) had a faster sedimentation compared with Alhydrogel (black). BSA–Alhydrogel and Alhydrogel had initial sedimentation velocities (slope) of approximately 10 and 3 mm/min and pellet heights of 6 and 5 mm, respectively.

(1) 0.1 mg/ml BSA in water, (2) 0.1 mg/ml BSA with 0.2% Alhydrogel (corresponds to 1 mg/ml aluminum), and (3) 0.2% Alhydrogel. Far-UV spectra were measured between 190 and 260 nm in quartz cuvettes with 0.1-mm path length. Spectra were recorded at a speed of 5 nm/min, with a step size of 0.2 nm, a response time of 16 s, and a spectral bandwidth of 1 nm. An average of three scans was recorded.

To study the thermal stability of the vaccines, the ellipticity at 220 nm of the protein and protein adsorbed onto adjuvant were measured between 20 and 90°C in a cuvette with 0.1-cm path length. The thermal stability method is described in detail in a recent publication from our group [28]. The kinetic melting of the protein was recorded with a scan speed of 60°C/h, response time of 4 s, bandwidth of 1 nm, and data pitch of 0.1°C.

Circular dichroism measurements showed that whereas the thermally induced unfolding of BSA in water was 69.7°C, for the BSA–Alhydrogel, the unfolding temperature was 72.3°C (Fig. 15.5). This shows that the BSA conformation is protected against thermal unfolding by interactions with aluminum adjuvant. To our knowledge, this is the first publication describing the use of CD to determine the thermal stability of protein–adjuvant vaccines without prior sample treatment. In the present CD study, a concentration of 0.2% Alhydrogel was used, and measurements were performed in a cuvette with a path length of 0.1 mm. Smaller path length CD cuvettes could be used if the sample is more turbid. Turbidity limits the use of CD for protein-adjuvant vaccines. However, the aim is often to use a low adjuvant concentration (e.g., 0.1%), and CD could add complementary information regarding the stability and conformation of the adsorbed antigen.

### 15.2.11 Ultrasound Resonator Technology

Ultrasonic resonator technology (URT) measurements were carried out by the ResoScan system from TF instruments (TF instruments GmbH, Heidelberg, Germany). The ResoScan instrument has two measurement cells with each a path length of 7 mm, which act as ultrasonic resonators. Cells were filled with a volume of 200 μl and closed to prevent evaporation. The samples were measured in one cell with in the reference cell water or 0.6% Alhydrogel for BSA in water and BSA–Alhydrogel, respectively. An optimal resonance peak (master peak) at 7.8 MHz was used for the ultrasound velocity measurements. Before the start of the thermal scans, the samples were kept for 5 min at 20°C. The difference in ultrasound velocity ($\Delta U$ in m/s) between the sample and its reference was recorded between 20 and 85°C, temperature increase rate of 300 mK/min. Data were reprocessed by the UltraView 3.2 software (TF Instruments GmbH).

Ultrasound resonator technology (URT) measures the sound velocity in solutions, which is mainly influenced by the hydration shell of the solute molecules. The sound velocity is influenced by protein solutions properties such as protein size, conformation, and aggregation as well as by the other constituents (e.g., sugars, salts, buffer). Measurement of the ultrasound can be used for very turbid protein samples, protein–adjuvant vaccines, emulsion, and creams [29,30].

Figure 15.6 presents the temperature dependence of ultrasound velocities of BSA in water and BSA–Alhydrogel. The ultrasound velocity in BSA–water sample decreased steadily between 20 and 85°C. The sound velocity in the BSA–Alhydrogel sample was different; it decreased between 20 and 30°C and increased between 35 and 70°C. These changes are likely due to the strong effect of BSA on the structure and morphology of aluminum (Fig. 15.1e and f). Thus, URT measurements

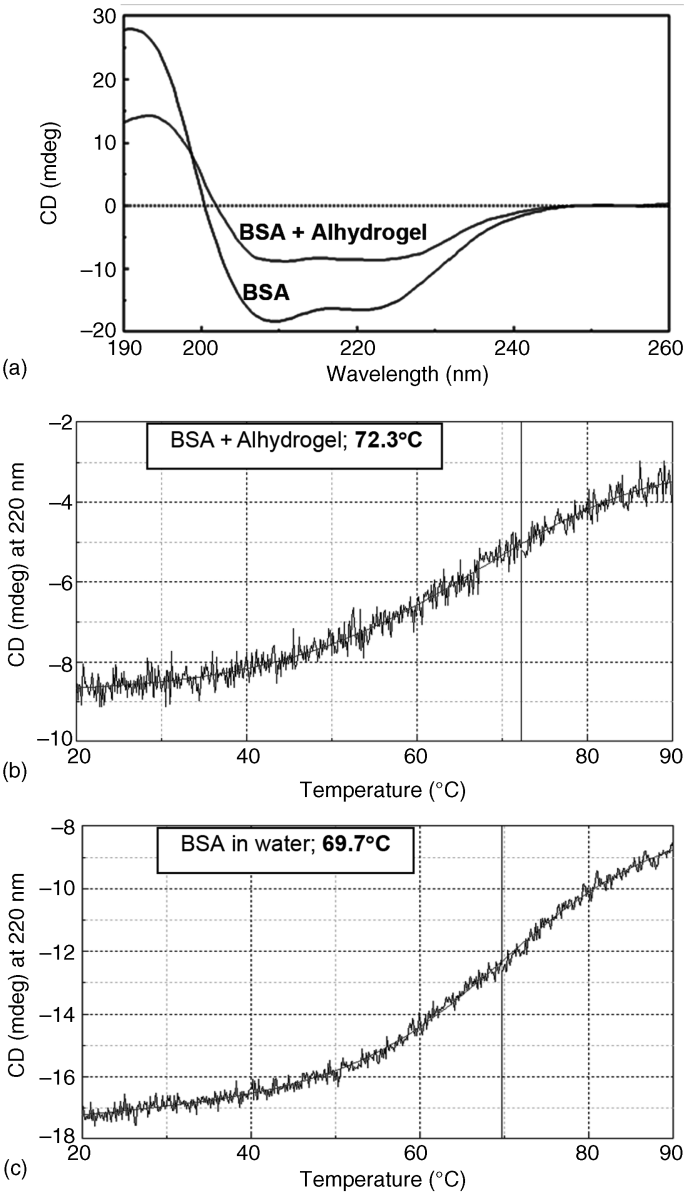


**FIGURE 15.5** (a) Circular dichroism spectra of 0.1 mg/ml bovine serum albumin (BSA) and 0.1 mg/ml BSA with 0.2% Alhydrogel. Thermal stability of (b) BSA–Alhydrogel and (c) BSA in water. The melting temperature was determined after fitting with a two-term van't Hoff equation as described by our group in an earlier publication [28].

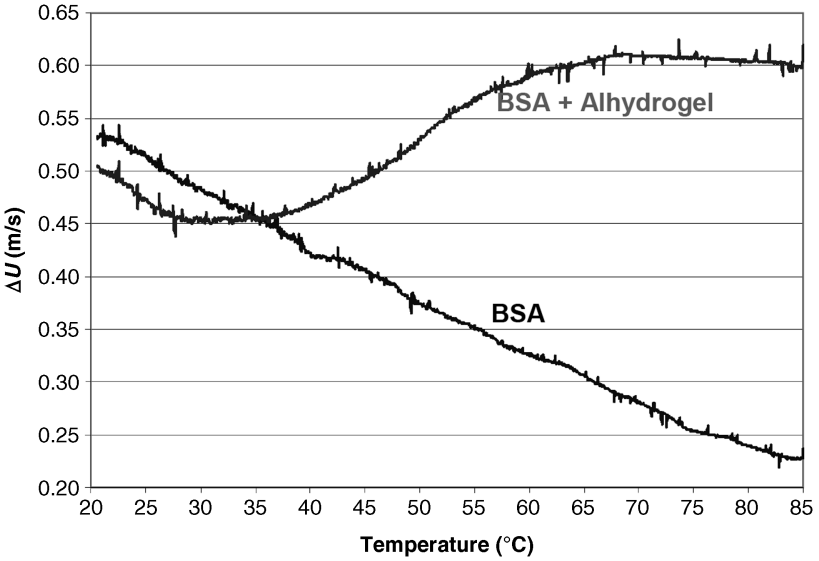


**FIGURE 15.6** Ultrasound resonance measurements: 1.7 mg/ml BSA in water, and 1.7 mg/ml BSA with 0.6% Alhydrogel after subtraction of the corresponding background.

bring new information for the comparison of different vaccine formulations and stability studies complementary to CD analysis.

## 15.3 CONCLUSION

In the early development phase of a new vaccine, many analytical methods are required to study its chemical and physical stability. This chapter has discussed different methods for the characterization of the vaccine model BSA–Alhydrogel. It was shown that BSA changed conformations after adsorption on Alhydrogel and was more stable toward thermal conformation changes. BSA strongly interacted with the aluminum hydroxide particles, resulting in more turbid suspensions that sedimented much faster than Alhydrogel suspensions. Many of the analytical methods can be adapted to 96-well plates or to small volumes. The noninvasive methods permit the characterization of protein vaccines in conditions as near as possible to the one used in therapeutic applications and open possibilities for new approaches for a faster development of stable and active protein vaccines.

## ACKNOWLEDGMENTS

The authors wish to thank Filippo Mulinacci for his help with the CD experiments and Dr. Steven Bell and Dr. Iain Larmour from Queen's University Belfast for their

help with performing the UV Raman analyses. The authors acknowledge the experimental time granted by the UV Raman facilities at the Central Laser Facility of the Rutherford Appleton Laboratory, UK, and the local technical assistance from Sue Tavender.

## REFERENCES

1. Lindblad EB. Aluminium compounds for use in vaccines. Immunol Cell Biol 2004;82: 497–505.
2. Gupta RK. Aluminum compounds as vaccine adjuvants. Adv Drug Deliv Rev 1998;32: 155–172.
3. Shirodkar S, Hutchinson RL, Perry DL, White JL, Hem SL. Aluminum compounds used as adjuvants in vaccines. Pharm Res 1990;7:1282–1288.
4. Vogel FR, Powell MF. A compendium of vaccine adjuvants and excipients. Pharm Biotechnol 1995;6:141–228.
5. Johnston CT, Wang SL, Hem SL. Measuring the surface area of aluminum hydroxide adjuvant. J Pharm Sci 2002;91:1702–1706.
6. Callahan PM, Shorter AL, Hem SL. The importance of surface charge in the optimization of antigen-adjuvant interactions. Pharm Res 1991;851–858.
7. Chang MF, White JL, Nail SL, Hem SL. Role of the electrostatic attractive force in the adsorption of proteins by aluminum hydroxide adjuvant. PDA. J Pharm Sci Technol 1997;51:25–29.
8. Friede M, Aguado MT. Need for new vaccine formulations and potential of particulate antigen and DNA delivery systems. Adv Drug Deliv Rev 2005;57:325–331.
9. Brandau DT, Jones LS, Wiethoff CM, Rexroad J, Middaugh CR. Thermal stability of vaccines. J Pharm Sci 2003;92:218–231.
10. Capelle MAH, Brugger P, Arvinte T. Spectroscopic characterization of antibodies adsorbed to aluminium adjuvants: correlation with antibody vaccine immunogenicity. Vaccine 2005;23:1686–1694.
11. Taschner N, Muller SA, Alumalla VR, Goldie KN, Drake AF, Aebi U, et al. Modulation of antigenicity related to changes in antibody flexibility upon lyophilization. J Mol Biol 2001;310:169–179.
12. Carra JH, Wannemacher RW, Tammariello RF, Lindsey CY, Dinterman RE, Schokman RD, et al. Improved formulation of a recombinant ricin A-chain vaccine increases its stability and effective antigenicity. Vaccine 2007;25:4149–4158.
13. Jones LS, Peek LJ, Power J, Markham A, Yazzie B, Middaugh CR. Effects of adsorption to aluminum salt adjuvants on the structure and stability of model protein antigens. J Biol Chem 2005;280:13406–13414.
14. Agopian A, Ronzon F, Sauzéat E, Sodoyer R, El Habib R, Buchet R, et al. Secondary structure analysis of HIV-1-gp41 in solution and adsorbed to aluminum hydroxide by Fourier transform infrared spectroscopy. Biochim Biophys Acta 2007;1774:351–358.
15. Arvinte T. Concluding remarks: analytical methods for protein formulations. In: Jiskoot W, Crommelin DJ, editors. Methods for Structural Analysis of Protein Pharmaceuticals. 1st ed. Arlington, VA: AAPS Press; 2005, p. 661–666.

16. Smith AM, Chan J, Oksenberg D, Urfer R, Wexler DS, Ow A, et al. A high-throughput turbidometric assay for screening inhibitors of protein disulfide isomerase activity. J Biomol Screen 2004;9:614–620.
17. Bevan CD, Lloyd RS. A high-throughput screening method for the determination of aqueous drug solubility using laser nephelometry in microtiter plates. Anal Chem 2000;72:1781–1787.
18. Goldberg ME, Chaffotte AF. Undistorted structural analysis of soluble proteins by attenuated total reflectance infrared spectroscopy. Protein Sci 2005;14:2781–2792.
19. Cantor CR, Schimmel PR. Biophysical Chemistry. Part II. Techniques for the Study of Biological Structure and Function. New York: W.H. Freeman and Co.; 1980.
20. Raussens V, de Jongh H, Pezolet M, Ruysschaert JM, Goormaghtigh E. Secondary structure of the intact $H^+$,$K^+$-ATPase and of its membrane-embedded region. An attenuated total reflection infrared spectroscopy, circular dichroism and Raman spectroscopy study. Eur J Biochem 1998;252:261–267.
21. Smith BM, Franzen S. Single-pass attenuated total reflection Fourier transform infrared spectroscopy for the analysis of proteins in $H_2O$ solution. Anal Chem 2002;74:4076–4080.
22. Hind AR, Bhargava SK, McKinnon A. At the solid/liquid interface: FTIR/ATR—the tool of choice. Adv Colloid Interface Sci 2001;93:91–114.
23. Patois E, Larmour IA, Bell SE, Palais C, Capelle MA, Gurny R, et al. Ultraviolet resonance Raman spectroscopy used to study formulations of salmon calcitonin, a starch–peptide conjugate and TGF-β3. Eur J Pharm Biopharm 2012;81:392–398.
24. Mulinacci F, Bell SE, Capelle MA, Gurny R, Arvinte T. Oxidized recombinant human growth hormone that maintains conformational integrity. J Pharm Sci 2011;100:110–122.
25. Kuentz M, Rothlisberger D. Rapid assessment of sedimentation stability in dispersions using near infrared transmission measurements during centrifugation and oscillatory rheology. Eur J Pharm Biopharm 2003;56:355–361.
26. Detloff T, Sobisch T, Lerche D. Particle size distribution by space or time dependent extinction profiles obtained by analytical centrifugation. Part Part Syst Charact 2006;23:184–187.
27. Gander B. Trends in particulate antigen and DNA delivery systems for vaccines. Adv Drug Deliv Rev 2005;57:321–323.
28. Mulinacci F, Capelle MA, Gurny R, Drake AF, Arvinte T. Stability of human growth hormone: influence of methionine oxidation on thermal folding. J Pharm Sci 2011;100:451–463.
29. Shah RB, Zidan AS, Funck T, Tawakkul MA, Nguyenpho A, Khan MA. Quality by design: characterization of self-nano-emulsified drug delivery systems (SNEDDs) using ultrasonic resonator technology. Int J Pharm 2007;341:189–194.
30. Stillhart C, Kuentz M. Comparison of high-resolution ultrasonic resonator technology and Raman spectroscopy as novel process analytical tools for drug quantification in self-emulsifying drug delivery systems. J Pharm Biomed Anal 2012;59:29–37.

# The State of the Art and Future of Gene Medicines

FRANK JACOBS, STEPHANIE C. GORDTS, and BART DE GEEST

## 16.1 INTRODUCTION AND HISTORIC PERSPECTIVE

Genetic diseases are a major source of hospitalizations and mortality worldwide [1]. Although there has been a stark improvement in diagnostic methods for detecting the underlying genetic defects over the past decades, the development of novel therapeutics has been lagging. Gene therapy holds great promise as a cure for these disorders. However, the application of gene therapy is not restricted to the treatment of inherited diseases such as severe combined immune deficiency, familial hypercholesterolemia, cystic fibrosis, and hemophilia but also involves the treatment of a broad spectrum of disorders such as cancer, Parkinson's and Alzheimer's disease, arthritis, diabetes, and cardiovascular diseases.

This chapter discusses the significant technological progress that has been made in the development of gene therapy strategies and the application of these strategies in specific therapeutic niches. More specifically, the chapter aims to provide an overview of selected highlights from the past (preclinical development), current (first-in-human [FIH] clinical trials), and future (remaining hurdles and regulatory aspects) of different gene therapy approaches. In these discussions, the chapter focuses on gene therapy approaches using viral vectors because they are generally considered to be more efficacious and have yielded the most prominent clinical successes. As such, an in-depth discussion of gene therapy strategies that are based on nonviral vector systems is not included in this chapter, and readers are referred to excellent reviews covering this aspect of the gene therapy field [2–4].

The term "gene therapy" was first coined in the 1960s and was based on novel contemporary insights into the genetic code. This is clearly illustrated in the 1963 assay by Joshua Lederberg in which he stated: "The ultimate application of

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*,
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

molecular biology would be the direct control of nucleotide sequences in human chromosomes, coupled with recognition, selection, and integration of the desired genes" [5]. The idea of using modified viruses to carry this genetic load and transport into the affected cells came not much later in 1966, when Edward Tatum postulated that "Finally we can anticipate that viruses will be used effectively for man's benefit, in theoretical studies concerning somatic cell genetics and possibly in genetic therapy" [6,7]. In 1967, Nobel Laureate Marshall Nirenberg wrote in *Science* about the possibility of reprogramming cells to produce "synthetic" messages and warned that society may not be ready to cope with this new human power over genetics [8]. However, the idea of gene therapy remained largely theoretical, and significant progress was hampered by various technological and ethical restrictions. The long and winding road that has since been traveled from this first notion of gene therapy has recently been elegantly reviewed [9–11].

In the 1970s, the onset of new technologies, such as restriction endonucleases, polymerase chain reaction (PCR), and sequencing techniques, finally allowed researchers to translate the elusive idea of gene therapy into practice. Initial *in vitro* studies showed that it was indeed possible to transfer DNA into cells and correct an underlying genetic defect [12]. However, the transfer of naked DNA was tedious, and the efficiency was low, underscoring the need for more efficient gene delivery systems. The following decades saw the development of different viral vector systems, including retroviral vectors [13], adenoviral vectors [14], and adeno-associated viral (AAV) vectors [15]. Using these novel technologies, various groups showed efficient correction of genetic disorders in animal models. An in-depth discussion of the preclinical development of these vector systems falls outside of the scope of this chapter and has been extensively reviewed elsewhere [16–18].

Although it was beyond any doubt that gene therapy had great potential, the field had been hyped, and progress was hampered by the rush for quick application studies and clinical trials. Unfortunately, this hype resulted in a lack of studies aimed at evaluating basic vectorology, innate and adaptive immune responses, vector biology, vector targeting, and basic mechanistic aspects of gene therapy approaches. The tragic death in October 1999 of Jesse Gelsinger, who was treated with an adenoviral vector for the treatment of a urea cycle disorder [19], indicated again that the field required more time to mature scientifically before significant therapeutic advances could be expected.

Since then, continuous refinement of viral vector systems, in combination with an increased understanding of the underlying vector biology and host immunology, has led to successful development of clinically viable gene therapy strategies. As will be discussed in the following paragraphs, the gene therapy field has yet again reached an important crossroad. Whereas the efficacy and safety of last-generation viral vectors has been scrutinized in various animal models, the question now remains how these results can be extrapolated to humans. Therefore, data from a new generation of FIH clinical trials are urgently needed to assess the current status of the gene therapy field and to identify the remaining challenges for ensuring clinical efficacy.

## 16.2 GENE THERAPY IN THE CLINIC

As briefly alluded to in the introduction, significant efforts have been made over the past decades to assess the potency of different gene transfer systems to correct numerous diseases in preclinical animal models. These consorted efforts have led to the successful translation of several gene therapy approaches into the clinic (Fig. 16.1). Indeed, some of the most promising results have only been obtained in the past few years [1,20,21]. In the following paragraphs, we will discuss the most salient examples of successful gene therapy trials. Interestingly, the development of gene transfer strategies for these highly diverse diseases has been intensely interconnected. With regard to technological aspects, the generation, development, and characterization of different vector platforms have been a collaborative effort by many investigators across the world. On the other hand, relevant clinical data in the field have always been scarce, and therefore the willingness to openly share and discuss preliminary data has been an important factor in ensuring efficient optimization of clinical protocols (see below). As such, it is clear that although the following paragraphs will highlight some of the most encouraging and notable studies, many other groups have also contributed heavily, albeit indirectly, to the further development of these gene transfer strategies.

### 16.2.1 Gene Therapy for Severe Combined Immunodeficiencies

One of the most fitting examples of the mercurial history of the gene therapy field is the development of a therapy for patients with severe combined immunodeficiency (SCID). The only currently available treatment for SCID is allogeneic hematopoietic stem cell transplantation. However, this technique is associated with inherent

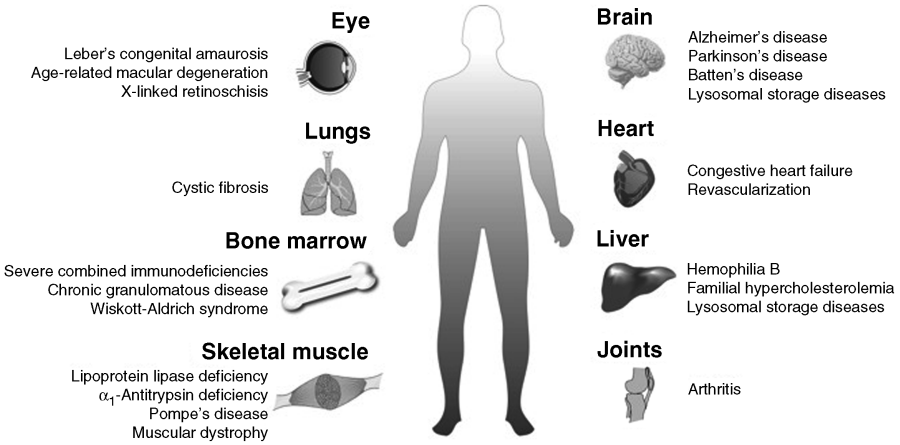


**FIGURE 16.1** Clinical gene therapy strategies: an overview of selected target organs and diseases that are currently pursued in clinical gene therapy trials.

complications and is known to possibly result in graft-versus-host disease. These limitations are even more pronounced in patients without suitable human leukocyte antigen (HLA)–matched donors. The general lack of therapeutic options for patients with SCID, together with the life-threatening complications of the disease, have identified SCID as a prime candidate for gene therapy. To date, gene therapy approaches have focused on the two most common forms of SCID: *adenosine deaminase* SCID (ADA-SCID) and *X-linked* SCID (SCID-X1).

ADA-SCID is the second most common cause of SCID (10–20% of all SCID patients) and is caused by a defect in the enzyme adenosine deaminase (ADA), which is required for the breakdown of purines. As a result of this defect, the conversion of ribonucleotides to deoxyribonucleotides is disturbed, leading to abnormalities in B-cell, T-cell, and natural killer (NK) lymphocyte development and function. Without treatment, ADA-SCID is generally fatal within the first year of life, underscoring the necessity for early interventions [22]. A unique feature for ADA-SCID is that, in addition to allogeneic hematopoietic stem cell transplantation, patients also respond to enzyme replacement therapy (ERT) with modified ADA-enzyme. However, because response to ERT is highly variable, the long-term prognosis on ERT is poor [22]. Initial gene therapy trials for ADA-SCID performed in the early 1990s focused on *ex vivo* transduction of lymphocytes [23,24] and hematopoietic progenitor cells [25] in patients who were receiving ERT. Although the procedure was shown to be safe, patients did not show any signs of efficacy, which was later attributed to the low levels of engraftment of genetically corrected cells and a poor selective advantage of these cells because of the maintenance of the ERT [26]. Additional studies have since then identified a more efficient gene transfer protocol using $CD34^+$ bone marrow cells combined with the administration of a nonmyeloablative dose of chemotherapy before cell infusion [27]. Based on these novel insights, three additional trials have recently been performed in Italy, the United Kingdom, and the United states. These studies included a combined total of 30 patients, and patients have now been followed for up to 10 years without any evidence of leukemic transformations [21,28]. Results have been encouraging, and most patients displayed progressive immune reconstitution, efficient systemic detoxification, and long-term multilineage engraftment. Moreover, 21 of 30 patients (i.e., 67%) who have been treated were able to stop ERT completely, indicating that the metabolic and immunologic correction in these patients has been achieved solely by the gene transfer vector infusion. Based on these positive results, Gaspar et al. [21] postulated that, given the high level of safety and efficacy, gene therapy could be considered as the first treatment option in the absence of suitable HLA-matched bone marrow donors, especially because failure of gene therapy does not preclude the patients from further treatment options such as ERT or bone marrow transplantation.

SCID-X1 is caused by mutations in the gene encoding the interleukin-2 receptor $\gamma$ chain (IL2RG), and this is the most common cause of SCID (accounting for 40–50% of all SCID patients) [20]. Patients display a complete absence of functional T cells and NK cells, and the function of B cells is severely impaired. SCID-X1 is a particularly attractive target for gene therapy because of the profound growth and survival advantage of genetically corrected cells. To date, two trials totaling

20 patients have been performed by Hacein-Bey-Abina et al. [29,30] and Gaspar et al. [20]. The technology that was used in these studies is essentially the same as was used for the trials for ADA-SCID [21,28]. Although there were some notable differences in the study design of the two SCID-X1 trials (including vector design, vector envelope protein composition, and *in vitro* transduction conditions), the combined results of these studies indicate some clear converging results [31]. First, the survival of gene transfer recipients compared favorably with recipients who had received mismatched allogeneic bone marrow transplants. Seventeen patients (85%) showed sustained presence of transduced T lymphocytes. Interestingly, the kinetics of T-cell recovery after gene transfer in SCID-X1 was markedly faster than what was observed in the setting of ADA-SCID. This may be caused by remaining purine toxicity in the thymic epithelium or peripheral tissues in the context of ADA-SCID. Alternatively, repopulation of T cells in SCID-X1 may be artificially accelerated by the accumulation of T-cell precursors because of their blocked differentiation from the stem cell pool. Second, whereas T-cell counts have been shown to be stable over time, long-term persistence of NK cells and B cells has been less impressive. Nevertheless, patients have been shown able to fight off infections without additional pharmacotherapy and approximately half of the patients do not require any immunoglobulin substitution [32].

However, these promising results have been severely marred by the development of a hitherto unforeseen complication: the development of T-cell acute lymphoblastic leukemia (T-ALL) because of insertional oncogenesis caused by the gammaretroviral vector. In total, five patients across both studies developed T-cell leukemia within 2 to 5.5 years after gene transfer, and this has since led to the discontinuation of these trials [32]. After chemotherapy, one patient succumbed from refractory leukemia, but the remaining four patients recovered and showed sustained T-cell immunity. Extensive studies have since shown that in all cases, the underlying cause was integration of the provirus into a proto-oncogene locus (predominantly into the *LIM domain only 2 [LMO2]* locus). These studies further highlighted that retroviruses preferably integrated into genes, especially genes that are actively transcribed. It quickly became clear that the viral *late terminal repeats* (LTRs) displayed an enhancer activity that could turn on transcription of the target gene and could thus trigger the leukemic progress [33]. Interestingly, none of the patients from the ADA-SCID trials developed leukemia, although the vectors used in these studies were highly similar. This strongly suggests that one or more disease-related factors influence retroviral integration. Whereas the underlying mechanisms for these differences remain highly speculative [26], these observations underscore the need for developing alternative vector systems.

Taken together, the hard-learned lesson of two decades of gene therapy for SCID is one of cautionary optimism. In the absence of HLA-matched bone marrow donors, gene therapy has been shown to be a viable, safe, and efficacious treatment option. However, the observations of an increased risk for leukemia have highlighted the inherent complications of the first generation of retroviral vectors. Lessons from these studies have led to the design of improved vectors that are less likely to transactivate proto-oncogenes by including internal promoters to drive transgene

expression or, alternatively, the inclusion of insulating DNA sequences that shield the surrounding genomic DNA from the enhancer effects of the LTRs [34]. Given the genotoxic risk associated with gammaretroviral vectors, future trials for SCID, as well as other immunodeficiencies, will be performed using self-inactivating lentiviral or gammaretroviral vectors devoid of viral regulatory elements [35,36]. Trials evaluating the safety and efficacy of these vectors have been initiated, and their results are eagerly awaited. On the basis of the seminal observations with SCID, a broad array of other blood cell diseases are being approached with gene therapy, including Wiskott-Aldrich syndrome, chronic granulomatous disease, X-adrenoleukodystrophy, metachromatic leukodystrophy, Hurler's syndrome, and β-thalassemia.

### 16.2.2 Hemophilia B

Hemophilia B is an X-linked recessive bleeding disorder. The disease is caused by mutations in the gene encoding the blood coagulation factor IX (FIX), a serine protease that is essential for formation of the fibrin clot. Spontaneous bleeds into joints and soft tissues of inadequately treated patients with severe hemophilia B (i.e., patients with $<1\%$ remaining FIX activity) lead to chronic debilitating arthropathies. In addition, spontaneous bleeds in the brain can result in severe neurologic complications and death. Extensive studies on the natural history of hemophilia B have indicated that raising the circulating levels of FIX into the range of a few percent of normal suffices to greatly improve the symptoms of the disease. The current mainstay therapy for hemophilia B is infusion of recombinant or plasma-derived FIX protein. However, protein infusion therapy is not always efficient, and the relatively short half-life of FIX requires frequent readministrations (two to three times a week). The average cost of protein therapy has been estimated to be more than $300,000 per year for an adult patient who receives prophylactic treatment, resulting in a lifetime treatment cost of more than $20,000,000 [1,37]. Additional features that have attracted interest in hemophilia as a target for gene transfer include (1) a wide therapeutic window (levels up to 150% are still within the normal range), (2) standardized endpoint assays, (3) FIX can be expressed from liver as well as extrahepatic tissues such as muscle, and (4) the availability of both small and large animal models.

Initial studies focused on the muscle as a target organ because myocytes had previously been shown to be able to maintain long-term transgene expression of reporter genes after direct intramuscular (IM) delivery of AAV vectors. In addition, at that time, IM injections were regarded as a more conservative and safe initial approach with reduced risks for the patients [38]. Animal studies in mice and dogs later confirmed the possibility of generating long-term FIX expression in the absence of serious adverse effects after IM injection [39,40]. Based on these results, the first phase I clinical trial for IM delivery of AAV2 in patients with severe hemophilia caused by missense mutations was initiated in 1999. Consequently, this was also the first clinical trial evaluating AAV gene transfer to the muscle. Initial results indicated that although parenteral vector administration was not associated with toxicity at any dose tested, circulating FIX levels were subtherapeutic (i.e., $\leq 1\%$ of normal)

[41–43]. In addition, the large number of injections required to transduce adequate amounts of skeletal muscle precluded injection of the predicted therapeutic dose in humans. *Post-hoc* analysis indicated no inhibitor formation against the expressed FIX protein in the muscle, and FIX could be detected for several years via immunohistochemistry performed on muscle biopsies. However, attempts to further improve gene transfer efficiency to the muscle in dog models indicated that depending on the context, inhibitory anti-FIX antibodies could indeed be generated after IM injections, precluding further development of this strategy [44]. Based on these findings, it was postulated that changing the target organ to liver could substantially increase the therapeutic index of AAV gene transfer vectors for hemophilia B. Indeed, previous studies in mice, dogs, and nonhuman primates (NHPs) had indicated that there was a strong dose-advantage in favor of liver, in the order of 10- to 100-fold, which was likely attributable to the higher secretion efficiency of hepatocytes into the circulation [45,46]. In addition, gene delivery to the liver also appeared to promote tolerance to the transgene product [47].

Based on the results of preclinical studies in animal models, the first clinical trial for liver-directed gene transfer was initiated in 2001 [48]. In this open-label, dose-escalation study, a recombinant AAV2 vector encoding human FIX was delivered through the hepatic artery in humans with severe hemophilia B. This was also the first clinical study in which AAV vectors were administered into the circulation. Taking into account that a previous trial evaluating systemic gene transfer with an adenoviral vector resulted in a fatal complication in one of the enrolled patients [19], the regulatory agencies therefore (understandably) required this trial to start at doses that were several-fold lower than the therapeutic dose that had been predicted based on data from animal studies. As could be expected, none of the subjects enrolled in the lower two dose cohorts presented any evidence of vector-related toxicity, and no measurable FIX activity could be determined. However, one subject in the highest dose cohort showed a therapeutic increase in FIX activity of up to 10–12% of normal levels (up from $<1\%$ at baseline). In contrast to long-lasting expression in hemophilic dogs and NHPs using similar vectors, expression at therapeutic levels in this subject was short lived. Starting at 5 weeks after vector administration, FIX expression began to fall and gradually declined to baseline at 10 weeks after vector infusion. The gradual decline in FIX levels coincided with an asymptomatic rise in serum transaminase levels between 4 and 8 weeks after vector administration. Antibodies directed against the FIX transgene could not be detected, and possible toxic or infectious causes for the increase in liver enzymes were excluded. A second subject receiving an identical dose of the AAV2 vector did not show any signs of FIX expression or liver toxicity, which was attributed to impaired transduction efficiency caused by higher levels of pretreatment AAV2 neutralizing antibodies in this patient (see below). Finally, one additional subject who was dosed at a fivefold lower dose also displayed a self-limited rise in liver enzymes with similar kinetics. Because none of the human subjects in these studies displayed a *null* mutation of FIX and because both antibody and T-cell responses directed against FIX were absent, studies focused on cellular immune responses directed toward the AAV capsid. A series of follow-up experiments suggested that T cell–mediated immunity to AAV capsid antigens

caused destruction of AAV2-transduced hepatocytes [48]. A strong correlation was shown between the time course of $CD8^+$ T cells responses to capsid, the observed transient increase in liver enzymes, and the loss of transgene expression [49]. Nevertheless, this association does not prove causality. Further studies evaluating the underlying mechanism are currently ongoing.

With regard to further clinical trials, the question remains whether it is possible to avoid the occurrence of an immune response against the vector even with an incomplete understanding of the mechanisms that may drive cellular immunity. Strategies that are currently under evaluation include the combination of immunosuppressive drugs in order to block activation and proliferation of the capsid-specific T cells until the capsid has been degraded and cleared from the transduced hepatocytes (Ragni et al., NTC00515710). Alternatively, Nathwani et al. [50] developed a novel AAV vector that was optimized in order to reduce the vector dose required to achieve therapeutic FIX expression, thus possibly reducing activation of the immune system. First, the authors chose to use an alternative AAV serotype, AAV8, which had been shown to be extremely hepatotropic after systemic administration [50]. Second, the vector genome was designed as a self-complementary vector in order to bypass the synthesis of a second DNA strain after hepatocyte transduction [51]. Finally, the authors also optimized the codon usage of the transgene cDNA sequence in order to improve the efficiency of translation [52]. This vector was evaluated in a dose-escalation study in six patients with severe hemophilia B; patients have been followed for 6 to 16 months [50]. Interestingly, all patients showed a therapeutic response, which was roughly dose dependent, and with FIX levels ranging from 2 to 11% of normal. Four of six patients discontinued prophylactic FIX treatment and remained free of spontaneous hemorrhages. The remaining two patients who remained on FIX protein therapy were able to extend the time interval between prophylactic injections. Intriguingly, one of the two patients who had received the highest vector dose displayed an asymptomatic increase in serum transaminases, which coincided with the presence of AAV8-capsid specific T cells in the peripheral blood. The other patient in the same cohort also displayed a slight increase in liver enzyme levels, although the cause was less clear. Both patients received a short course of glucocorticoid therapy. As a result, serum transaminase levels quickly returned to normal, and FIX expression levels were maintained in the range of 3–11% of normal. The gene transfer vector production costs for this study have been estimated at $30,000 for a patient treated with the middle dose of vector. In light of the annual costs of protein infusion therapy, it is evident that dramatic cost reductions have already been achieved.

Taken together, these studies indicate that within the right context, AAV-mediated gene transfer to the liver has the potential to convert a severe hemophilic phenotype into milder form of the disease or even reverse it entirely. In this regard, the study by Nathwani et al. [50] is truly a landmark study in that it is the first to achieve long-term expression after systemic gene transfer to the liver at levels that are therapeutically relevant. However, in the midst of these almost euphoric results, it remains of paramount importance to remain vigilant and identify all possible factors that may influence the outcome of gene therapy in these patients. Currently, it is still unclear if

there are any risks associated with this procedure. Of note, the critical determinants of cellular immune responses directed against the AAV vector capsid are still largely unknown. It is highly likely that interindividual differences in antigen processing and presentation may influence the type and magnitude of the immunologic response against the AAV particles [53]. In addition, pre-exposure to natural AAV infections may vary highly among patients, resulting in a completely different repertoire of memory T cells. In this regard, it also remains to be determined to which extent increases in serum transaminases are clinically relevant. Finally, another question is how to optimize treatment for the large proportion of patients (up to 30% [54]) who harbor AAV8-specific antibodies. Many of these questions can only be addressed after inclusion of additional participants. If further studies indicate that this approach is safe, it may replace the cumbersome and expensive prophylactic FIX therapy. Moreover, the use of similar vectors may spark significant progress in gene therapy for other metabolic disorders, including familial hypercholesterolemia, lysosomal storage diseases, and urea cycle defects.

### 16.2.3 Leber's Congenital Amaurosis

Leber's congenital amaurosis (LCA) is a group of retinal degenerative disorders that results in early onset blindness and displays an autosomal recessive inheritance [1]. Symptoms appear during infancy or early childhood, and the diagnosis is typically made when the parents notice that the baby fails to track visually. There are no treatments available for LCA. Although there may be some low level of vision during childhood, patients eventually become completely blind by the third or fourth decade of life. To date, 14 different genes have been identified that can result in LCA [55]. In the development for gene transfer strategies to cure LCA, attention quickly moved to a variant caused by the deficiency of the gene called RPE65 (so-called type 2 LCA), which accounts for 6% of all patients with LCA [55]. This selection was mainly based on the availability of a naturally occurring dog model of LCA with RPE65 deficiency. RPE65 encodes the retinal pigment epithelium (RPE)–specific 65KDa protein, which is essential in the conversion of 11-*cis* retinal during the generation of rhodopsin in rod photoreceptors. In contrast, cone photoreceptors have access to 11-*cis* retinaldehyde chromophore through an alternative pathway that does not require RPE65. Thus, in patients with RPE65 deficiency, vision via cone photoreceptors is preserved, although progressive degeneration of rod photoreceptors eventually results in the complete loss of cone photoreceptors and, ultimately, complete blindness. However, vision could theoretically be restored by simply providing the missing enzyme, thus preventing retinal degeneration. In addition, the eye is generally regarded as an immunoprivileged site, which may prevent immunologic complications (*cfr.* the AAV2 hemophilia B trial [48]). The development of a successful therapy for LCA is regarded by many as one of the most compelling illustrations of the significant progress that has been made by the gene therapy field over the past decades.

Initial studies in genetically engineered knock-out mice and a naturally occurring dog model of RPE65 deficiency had shown the feasibility of reversing the degeneration of the retinal epithelium via subretinal injections with an AAV2 vector

encoding RPE65 [56,57]. Based on these promising findings, three groups in the United Kingdom and the United States began clinical trials for LCA caused by RPE65 deficiency in 2007. Under general anesthesia, patients received a subretinal low-dose injection of an AAV2 vector encoding the RPE65 protein in one of both eyes (generally the eye with worst function at baseline). To date, there have been no serious adverse effects in any of the more than 25 subjects who have been treated so far. Not more than 1 year after initiation of the trial, all three groups reported a certain degree of positive results in most patients who had been included in the trial. The relatively large variation in patient responses has been partly attributed to differences in vector genome structure between trials (i.e., trials in the United Kingdom used elements of the endogenous RPE65 promoter, but researchers from the Philadelphia trial used the ubiquitous cytomegalovirus promoter). Results were reported as improved psychophysical measures of visual function (including visual acuity and visual fields), as well as objective measures (e.g., pupillary light response, electroretinograms, and full-field light sensitivity) and functional measures (e.g., the ability to navigate through an obstacle course in a dimly light room) [58–60]. These beneficial effects of RPE65 gene therapy on visual function became evident as early as 2 weeks after vector injection. As was expected based on the immunoprivileged position of the eye, immune responses to the vector and the transgene product were minimal. Importantly, animal studies had previously indicated that restoration of vision after RPE65 gene therapy was most effective when performed in younger animals [61]. Although most clinical trials did not specifically assess this possible age-dependent efficacy, *post-hoc* analysis indicated that the response to subretinal gene therapy strongly corresponded to the degree of retinal degeneration at the time of injection, which had previously been shown to be strongly age dependent [60]. In the Philadelphia trial, these positive results have prompted researchers to start treating the second eye in patients who had previously been treated. Results from these studies are eagerly awaited because they will provide invaluable information on the possibility of redosing patients in the absence of immunologic complications.

In conclusion, these studies have indicated that subretinal injections of AAV vectors are safe and well tolerated. Moreover, a therapeutic effect was observed in all studies, albeit with varying success. These trials also showed that the therapeutic effect of gene transfer is highly dependent on the stage of the disease at the time of intervention. Further studies are required to assess additional variables that may influence the outcome of gene transfer studies to the retina. With regard to extrapolation of these results to other diseases of the retina and the eye, studies have indicated that whereas AAV2 can efficiently transduce cells within RPE, transduction of other cell types (including photoreceptors) is lower. Intriguingly, subretinal injection of any of the AAV serotypes tested to date can result in very efficient transduction of RPE (usually 100% transduction in the area of detachment) and variable transduction of photoreceptors (between 20% and 80% depending on the serotype and vector dose) [62]. As such, whereas interventions requiring subretinal injections are currently quite rare, the onset of gene therapy for genetic retinal disorders may lead to an increased demand among vitroretinal surgeons.

Nevertheless, an intravitreal approach allows a more homogenous transduction and poses a lower risk of surgical complications. This is also relevant for nongenetic diseases, such as age-related macular degeneration. However, it is currently unclear which AAV serotypes are most suited for this approach (see Ref. [62]). Continued progress is ongoing, and planned future trials will likely change certain practical aspects in ophthalmology [1]. Whereas identification of the underlying genetic defects in patients with hereditary retinal disorders was previously of less importance because the absence of available therapies, correct genotypic information is likely to become more important in the near future. Early detection may further be of paramount importance given the better phenotypical correction in younger patients.

### 16.2.4 Lipoprotein Lipase Deficiency

Lipoprotein lipase (LPL) is an enzyme found on the luminal surface of endothelial cells in capillaries and is mostly concentrated in adipose tissue, skeletal muscle, and the heart [63]. LPL plays a central role in the catabolism of triglyceride-rich lipoproteins, chylomicrons, and very-low density lipoproteins via hydrolysis of triglycerides into fatty acids and monoacylglycerol. In addition, LPL promotes cellular uptake of chylomicron remnants, cholesterol-rich lipoproteins, and fatty acids. Under normal conditions, LPL serves to blunt any excessive postprandial increase in triglycerides. LPL deficiency (LPLD) is an autosomal recessive disorder caused by loss of function mutations within the LPL gene, located on chromosome 8p22. To date, more than 70 different mutations in the LPL gene have been documented, and most of them are associated with loss of the catalytic activity [64]. Patients are either homozygous for these mutations or compound heterozygotes. LPLD is the most common cause of hyperchylomicronemia, with a worldwide incidence of one to two in 1,000,000, although founder effects occur in certain populations [65]. Patients present with extremely elevated levels of large chylomicrons, severe hypertriglyceridemia, and low levels of high-density lipoprotein (HDL). The most severe clinical complication is acute pancreatitis, although patients also have eruptive xanthomas, lipemia retinalis, and hepatosplenomegaly. Acute pancreatitis is a heterogeneous inflammatory condition that results in significant morbidity or mortality in 20–30% of patients [66]. Currently, there is no pharmacologic treatment for LPLD, and patients are maintained on a fat-restricted diet in combination with standard lipid-lowering therapy. However, these treatments are insufficient in preventing the onset of pancreatitis and preventing complications associated with LPLD.

Gene transfer of LPL to the muscle using an AAV1 vector is being developed as a novel therapeutic to treat LPLD, and this program is sponsored by UniQure (formerly Amsterdam Molecular Therapeutics) under the name alipogene tiparvec (Glybera, AMT-011, AAV1-LPL$^{S447X}$). The final product encodes a natural gain-of-function variant of LPL that is present in 20% of whites. This variant is associated with an enhanced removal of proatherogenic apo$B_{100}$-containing lipoproteins, resulting in lower levels of low-density lipoproteins, lower triglyceride

levels, and increased levels of high-density lipoproteins or HDL cholesterol, and is associated with a decreased incidence of cardiovascular disease [67]. Animal studies in LPL-deficient mice indicated that AAV1-LPL$^{S447X}$ resulted in a persistent and dose-dependent reduction of plasma triglycerides of up to 95% [63]. Safety studies did not indicate any vector-associated toxicities.

Initiation of a clinical development program for alipogene tipoparvec in 2005 proved cumbersome; because of the low number of patients, there was a reasonable paucity of studies evaluating the natural history, presentation, and evolution of burden in LPLD patients. Therefore, the program also included prospective observational studies aimed at establishing reference values for plasma triglyceride levels, assessing the efficacy of a fat-restricted diet on LPLD-associated pathologies, and establishing a framework of safety and efficacy parameters for future intervention studies. Since then, three phase I/II and phase II/III trials have been performed in the Netherlands and Canada, including a total of 27 patients. In all studies, alipogene tipoparvec was administered as a single series of IM injections (30–70 injections depending on the body weight) in the upper and lower limbs. In selected studies, patients also received a combination of immunosuppressive drugs to inhibit immune responses aimed against the AAV1 vector. In addition to standard clinical results and data related to efficacy, these studies included analysis of vector shedding in body fluids (urine, semen, saliva) as well as immune responses directed against both the LPL transgene product and the AAV1 vector. An in-depth discussion of these results was recently published by Gaudet et al. [63]. Across the board, administration of alipogene tipoparvec was well tolerated, and none of the studies reported any dose-limiting toxicities. Long-term follow-up of patients has underscored the excellent safety profile of AAV vectors in human clinical trials. Nearly all patients showed a decrease in plasma triglyceride levels, and muscle biopsies taken 6 months after vector administration indicated persistent LPL expression and biological activity. Cox proportional hazard regression modeling indicated a 70% reduction of the risk for acute pancreatitis, although this did not reach statistical significance [63]. It is currently unclear to which extent the reduction in plasma triglyceride levels will be predictive of hard clinical endpoints. Therefore, generation of additional long-term follow up data is indicated, and patients will be followed until 15 years after vector administration for further assessments of safety and efficacy. In June of 2011, the European Medicines Agency (EMA) adopted a negative opinion, recommending refusal of the marketing approval for Glybera. The basis for this negative opinion was that studies had not shown a consistent long-lasting effect on plasma lipids and that there were insufficient patients for whom long-term follow-up data were available. After a request for reevaluation by Amsterdam Molecular Therapeutics, the Committee for Advanced Therapeutics indicated that most of the original comments could be addressed by additional postmarket studies and that approval should be granted. Nevertheless, the EMA upheld their initial objections for not granting marketing approval. As a result, the project sponsor faced severe financial turmoil, and a new private company was founded that pertained the rights to Amsterdam Molecular Therapeutics' portfolio [68]. After months of intensive discussions, UniQure was able to convince the European Commission to request a reassessment

of the marketing approval of Glybera for a specific subset of LPLD patients with multiple or severe attacks of pancreatitis. This third review by the Committee for Medicinal Products for Human Use (CHMP) again did not grant a positive opinion, although a narrow majority of the CHMP was in favor of approval. Nevertheless, UniQure was able to convince the CHMP to invite a fourth review of Glybera, which resulted in a majority vote in favor of marketing authorization on July 19, 2012. As such, Glybera became the first gene therapy product to receive marketing approval in Europe.

### 16.2.5 Heart Failure

Interestingly, the growing experience with AAV in clinical trials has also contributed to the application of gene transfer therapeutics into the field of nongenetic diseases such as heart failure. Chronic heart failure is a progressive condition in which the heart is unable to pump sufficient blood to meet the body's demands, and this remains an important cause of morbidity and mortality. Hospitalization costs in the United States because of chronic heart failure were estimated to reach more than $32 billion for 2009 [69]. The prognosis for patients with chronic heart failure remains poor, and nonpharmacologic treatments (e.g., heart transplantations and cardiac assist devices) suffer from limited availability and high costs. These restrictions have underscored the large unmet need for novel therapeutic options for patients with chronic heart failure. On the one hand, many groups have focused on cell-based therapies. However, clinical trials have struggled to show clinical benefits on hard clinical endpoints, and the isolation and selection of the desired cell populations remains a matter of debate [70]. In contrast, several recent studies have shown that gene therapy might provide an effective option for treating the failing myocardium [71,72].

One of the most common abnormalities in heart failure is a dysfunction of the sarcoplasmic reticulum, resulting in a disturbance of intracellular calcium handling. The enzyme responsible for reloading the sarcoplasmic reticulum with calcium during relaxation is the sarcoplasmic reticulum $Ca^{2+}$-ATPase (SERCA2a). Several studies have indicated that the observed dysfunction is largely attributable to a decreased activity of SERCA2a [73]. Indeed, low levels of SERCA2a have been shown to correlate with abnormally high diastolic levels of cytosolic calcium, impaired calcium release from the sarcoplasmic reticulum during systole, and poor clinical outcomes in patients with heart failure [74]. In preclinical animal models of heart failure, including rodents [75], pigs [72], and sheep [71], it has been shown that restored expression of SERCA2a protein levels after AAV-mediated gene transfer to the myocardium resulted in a significant improvement of cardiac function and energetics.

Based on these findings, a FIH phase I/II trial was initiated in 2007. The double-blind, placebo-controlled Calcium Upregulation by Percutaneous Administration of Gene therapy in Cardiac Disease (CUPID) trial aims to restore SERCA2a expression in patients with chronic heart failure. In total, 39 patients with advanced heart failure were administered an AAV1 vector encoding SERCA2a via anterograde epicardial

coronary artery infusion using standard catheterization techniques. Patients were screened for the presence of neutralizing antibodies against AAV1 before being included in one of three dose groups (low, middle, or high) and were followed for 12 months after vector administration. All patients tolerated the procedure well, and there were no untoward safety findings. Contrastingly, the incidence of adverse effects was inversely correlated to the dose of AAV1-SERCA2a. Because the possible impact of SERCA2a expression on clinical outcomes could not be predicted, this study used an endpoint approach requiring concordant changes across multiple endpoints to demonstrate efficiency. Using this methodology, the authors reported a false-positive rate of 2.7% in the placebo group [76]. Patients on placebo continuously deteriorated throughout the follow-up period. Despite some improvements at 6 months after vector administration in the low- and middle-dose groups, only the high-dose group met all of the primary endpoint success criteria. Of note, patients in the high-dose group displayed an increase in both symptomatic (New York Heart Association class, Minnesota Living with Heart Failure Questionnaire score) as well as objectively measurable (6-minute walk test, $VO_2$max, cardiac function biomarkers, and cardiac remodeling) parameters. Compared with placebo, there was also a significant reduction in the rate and duration of cardiovascular hospitalizations in patients that were treated with a high dose of AAV1-SERCA2a. All beneficial effects of gene transfer were further confirmed throughout the 12 months of follow-up. Moreover, the hazard ratios for recurrent clinical events after adjustment for terminal events indicated risk reductions of 60%, 56%, and 88% for patients who had received the low, middle, and high doses of AAV1-SERCA2a, respectively. These promising results have underscored the possibilities for a possible enhanced biological effect of SERCA2a overexpression in the setting of chronic heart failure. However, the low sample size may have led to overestimation of the results, and additional trials are required for confirmation of the observed effects. Also, long-term protection will be assessed in the current cohort to provide additional insights into the potential for gene therapy to treat advanced heart failure.

## 16.3 ADENO-ASSOCIATED VIRAL VECTORS AS A TECHNOLOGICAL PLATFORM FOR CLINICAL TRANSLATION

As discussed in the preceding paragraphs, the development of clinical gene therapy programs has relied heavily on the development of AAV vectors. Indeed, several advantageous features of AAV have identified AAV vectors as the prime candidate for a wide spectrum of different applications. Whereas some aspects of vector development apply to all of these applications, other features may be restricted to unique diseases or organs. For reasons of clarity and conciseness, the following paragraphs will focus predominantly on features that affect gene transfer to the liver. Therefore, we will first discus general aspect of the biology of AAV vectors and, whenever applicable, their development as gene therapy

vehicles for liver-directed gene transfer will be highlighted. For an in-depth discussion of alternative gene transfer vector systems and target organs, readers are referred elsewhere [1,26,31,77–79].

### 16.3.1 Adeno-Associated Viral Vectors

Adeno-associated viral vectors are small, nonenveloped, single-stranded DNA viruses belonging to the *Dependovirus* genus of the *Parvoviridae* family. After their initial identification as contaminants in laboratory preparations of recombinant adenoviral vectors more than three decades ago [80], six distinct AAV serotypes (AAV1–AAV6) were identified. Additional studies have identified an expanded family of AAV serotypes from various source materials based on the recovery of latent forms of the genome using PCR techniques. To date, 11 different serotypes comprising more than 120 capsid variants have been isolated from humans, NHPs, cows, dogs, horses, and birds [81–85]. These serotypes are categorized into six different phylogenetic clades representing a broad distribution of potential AAV biology [81]. Of these serotypes, AAV2 has been studied the most extensively and has served as a paradigm to explore AAV biology.

Compared with the genome size and structure of other viral vectors, the AAV genome is relatively small and simple, at approximately 4.7 kb, and contains only two genes, *rep* and *cap*, that encode four Rep proteins and three capsid proteins, flanked by two 145-base inverted terminal repeats (ITRs). Recombinant and replication-defective AAV vectors used for gene transfer experiments retain only the ITRs, which are necessary for replication, packaging, and integration, but all viral coding sequences are entirely removed [86].

Several favorable features have made AAV a valid candidate as a gene therapy vector. Most important, AAVs have not been associated with human disease and are thus generally considered to be nonpathogenic. This is partly reflected by the fact that most AAV variants are poor activators of both innate and adaptive immune responses. Moreover, AAVs are naturally replication defective and require co-infection with a helper virus, such as adenovirus or herpes virus, in order to replicate. In addition, AAV transduce both dividing and nondividing cells and have been shown to maintain long-term expression in the absence of chromosomal integration. Whereas wild-type AAV possess the ability for site-specific integration into the human genome, this feature is absent from recombinant AAV vectors, and integration into the host genome is random but relatively infrequent [87]. Finally, AAV2 vectors have an average diameter of 22 nm [88]. This is considerably smaller than other common viral vectors used for liver gene transfer [89,90]. Several studies have underscored the critical role of the diameter of fenestrae in restricting the passage of vectors according to their size [90,91]. Considering this potential importance of the diameter of sinusoidal fenestrae on hepatocyte transduction, the relatively small size of AAV vectors should allow efficient transendothelial passage into the space of Disse [79,89,92]. Based on the features discussed, AAV quickly moved into the forefront of the gene therapy field and a myriad of studies were initiated.

### 16.3.2 Serotype Development

The AAV capsid constitutes the primary interface between host and vector and is presumed to be the key determinant of vector tropism. The capsid is also a determinant of innate and adaptive immune responses. In the past decade, the Wilson laboratory has spearheaded the identification and characterization of novel alternative AAV serotypes [81,82]. Recent studies have underscored the supremacy of these novel serotypes *in vivo*, including increased hepatocyte transduction, reduced prevalence of neutralizing antibodies, and diminished capsid immune responses [54,93,94]. Among different serotypes evaluated to date, the most salient increase in liver transduction and transgene expression has been observed for AAV8 and AAV9 [95,96]. In general, transgene expression levels after intravenous injection in mice are increased 10- to 100-fold compared with AAV2 [82]. However, the observed advantages of AAV8 over AAV2 seem to be reduced in larger animal species such as dogs and NHPs [97,99].

Mechanisms underlying the distinct biological performance of different AAV serotypes have been attributed to differences in viral entry, trafficking, uncoating, and genome processing [100,101]. Interestingly, Cunningham et al. showed that following systemic administration of AAV8 to the liver, all dsAAV8 genomes are formed within 3 days [102]. In contrast, AAV2 shows a gradual increase of dsAAV2 genomes over a period of several weeks [103]. Nakai et al. showed that hepatocyte transduction by AAV8 does not differ after either intraportal injection or systemic administration via the tail vein and up to 90% hepatocyte transduction could be achieved [104]. Finally, AAV8 has proven to be more potent in inducing transgene-specific immune tolerance, which has been attributed to the fact that AAV8, and possibly AAV9, does not bind to heparan sulphate proteoglycan, resulting in a lower transduction of antigen-presenting cells (APCs) [105]. AAV8 has also been used with great success in different animal models of metabolic diseases [106–109].

One additional factor that may determine vector tropism *in vivo* is the use of different primary receptors between serotypes. Heparan sulphate proteoglycan has been shown to be the primary receptor for AAV2 *in vivo* [110], acting in combination with several co-receptors, including $\alpha_v\beta_5$ integrin [111], $\alpha_5\beta_1$ integrin [112], fibroblast growth factor receptor 1 [113], and hepatocyte growth factor receptor [114]. Contrastingly, AAV4 and AAV5 have been shown to primarily interact with either O-linked or N-linked $\alpha_{2-3}$ sialic acid, respectively [115]. The specific cell surface receptor for AAV5 has been identified as the platelet-derived growth factor receptor [116]. Finally, the laminin receptor has been shown to act as a receptor for multiple AAV serotypes, including AAV2, AAV3, AAV8, and AAV9 [117]. Although this might provide valuable insights into the biology of distinct AAV serotypes, the exact influence of AAV receptors in their biodistribution remains unknown.

Based on the growing knowledge of structure–function relationships of the AAV capsid, several groups have attempted to modify the AAV2 capsid to increase transduction efficiency or alter biodistribution *in vivo*. In this respect,

Vandenberghe et al. have previously demonstrated that ablation of heparin-binding motifs can modulate the antigenicity of AAV2 particles by reducing transduction of APCs [105]. Alternatively, agents that modulate the ubiquitin–proteasome system have been shown to augment nuclear trafficking of viral particles with a concurrent increase in AAV transgene expression [118]. Recently, Zhong et al. indicated that mutations of surface-exposed tyrosine residues on the AAV2 capsid can result in decreased phosphorylation and ubiquination, thereby inhibiting proteasome-mediated degradation and resulting in a 10-fold increase of transgene expression levels in mice [119]. Whereas these studies underscore the malleability of the AAV capsid, it seems unlikely that this strategy will be sufficient to overcome other disadvantageous features of AAV2, such as the high prevalence of neutralizing antibodies in humans.

In an attempt to overcome these hurdles, recent studies have focused on directed evolution techniques to screen larger AAV mutant libraries for the presence of desirable features. Using this methodology, Mahesri et al. have recently reported the selection of AAV2 variants that are more resistant to neutralizing antibodies [120]. Alternatively, Li et al. reported successful AAV capsid genome shuffling using serotype capsid sequences from AAV1, AAV2, AAV8, and AAV9 *cap* genes to select novel AAV variants for increased transduction of melanoma cells [121]. Whereas directed evolution clearly holds great promise for the generation and selection of novel AAV vectors, complete characterization of these novel variants will require extensive additional *in vivo* experiments.

### 16.3.3 Transcriptional Vector Optimization

Because the observed response in the FIX clinical trial was dose dependent, great effort has been put into increasing the therapeutic index of AAV vectors. The only viral sequences in the vector are the two ITRs, which do not seem to influence hepatocyte transduction [122]. Therefore, one important factor has been the optimization of the eukaryotic DNA sequences comprising the expression cassettes. A remarkable increase in the ratio of transgene expression per vector genome has since been achieved by using stronger promoter–enhancer combinations, by codon optimization, and by using self-complementary double-stranded AAV (scAAV) vectors. The latter is based on the ability to encode both the coding and complementary strands within a single construct, obviating the need for second strand annealing or synthesis after vector uncoating, which is the rate-limiting step in initiating transgene expression [51,123]. However, in a recent series of experiments, Gao et al. have shown that the improved transduction efficiency of hepatocytes by scAAV can be highly context dependent [124,125]. Possible explanations for these observations include differences in transcriptional activity between ssAAV and scAAV genomes, differences in vector design, and the presence of neutralizing antibodies in NHPs. Irrespective of the underlying cause of these discrepancies, the inferred 50% reduction in vector packaging capacity has restricted the use of scAAV to a very limited number of transgenes.

The most pronounced improvements in transgene expression have been obtained by using different promoter–enhancer combinations. Although unique studies performing side-by-side comparisons are rare, several studies have indicated that the most potent promoters for hepatocyte-restricted expression are the *albumin* promoter [126], the α1-antitrypsin promoter [127], and the thyroxine-binding globuline promoter [99,108]. Franco et al. showed that the synthetic 790 bp LSP promoter, consisting of the 475 bp thyroxine-binding globuline promoter and two copies of the 96 bp bikunin/$\alpha_1$-microglobulin enhancer, resulted in 60-fold higher transgene expression levels compared with the constitutive cytomegalovirus enhancer/chicken β-actin promoter [107]. Other strategies to improve the efficiency of transgene expression have been aimed at optimization of transcriptional sequences by inclusion of an optimized Kozak consensus sequence or via codon optimization of the transgene sequence. A consensus Kozak sequence has been shown to be important and required for the optimal translation of mammalian genes by enhancing the recognition by eukaryotic ribosomes [128]. Similarly, rare codon usage may result in translational stall caused by tRNA unavailability, and overrepresented codon pairs may stall because of ribosome pausing [129]. Conversely, underrepresented codon pairs may enhance translational efficiency of a transcript. The rate of translation can also affect protein folding, and it has been suggested that codon usage may be a mechanism to regulate translation speed and thus allow stepwise folding of individual protein domains [130]. Although *lege artis* comparisons in gene therapy experiments are scarce, several studies have shown significant improvements in transgene expression by combining these techniques [52,106,131,132]. Finally, several studies have documented a significant increase in transgene expression after inclusion of the woodchuck hepatitis posttranscriptional element (WPRE) [133,134]. Although the specific mechanism of this enhancement is not well described, this may involve the facilitation of RNA processing and RNA export [135,136]. However, a potential oncogenic activity of the WPRE sequence, because of potential promoter activity and the possibility of expressing a 60-amino-acid woodchuck hepatitis virus X protein-derived peptide, might preclude the incorporation of this DNA sequence in clinical applications [137].

### 16.3.4 Selection of Animal Models for Liver-Directed Gene Transfer

A critical factor in progress toward clinical applications of gene therapy is the availability of animal models that accurately predict outcomes in humans. Unfortunately, extrapolation of data from murine studies to larger animal species has been cumbersome. Complications observed in larger animal species include reduced hepatocyte transduction, a higher prevalence of neutralizing antibodies against the vector, and a higher susceptibility for cellular immune responses toward the transgene or the AAV capsid.

Notably, several studies have indicated that whereas AAV8 is more potent in targeting the liver than AAV2, the observed increment in transgene expression in larger animal models is markedly reduced compared with mice. Based on data from dogs, it was previously estimated that AAV8 is about 1.5- to 4.3-fold more efficient

than AAV2 and two- to sevenfold more efficient than AAV5 for hepatic gene transfer [99]. Interestingly, Davidoff et al. showed that in mice, hepatocyte transduction by AAV was also five- to 13-fold higher in males compared with females, irrespective of the promoter, transgene sequence, or mouse strain [138]. Similar findings have since been observed in several mouse models of metabolic diseases [139–141]. Unfortunately, differences in liver transduction between males and females in larger species have remained largely unaddressed, mostly because of low sample sizes. As such, this might partly contribute to the relatively high degree of variation observed in higher species. An additional critical factor that should be taken into account is that whereas pre-exposure to AAV8 can be low in humans, the prevalence of neutralizing antibodies against this serotype are much higher in NHPs because they are the natural hosts. As such, even very low levels of neutralizing antibodies that are below the detection limit of current modes of analysis have been implicated in reduced hepatocyte transduction in NHPs. However other causes should also be taken into account, and in this respect, Gao et al. recently showed substantial differential processing of AAV genomes in mice and NHPs [142]. Nevertheless, gene transfer in both the FIX clinical trials resulted in clinically relevant transgene expression levels of 10–12%, which is comparable to dogs injected with an identical doses. Thus, it seems that if immunologic complications can be prevented, data obtained in dogs and possibly NHPs might be directly translatable to humans. In this respect, Nathwani et al. recently documented an almost linear dose-response curve after intravenous administration of an scAAV8 vector in NHPs [143]. Remarkably, gene transfer with a dose of $2 \times 10^{12}$ v.g./kg was sufficient to transduce almost 100% of hepatocytes, resulting in supraphysiological transgene expression levels for up to 5 years.

Taken together, these data underscore the great progress that has been made in evaluating the efficiency of AAV vectors in cats, dogs, and NHPs. However, even when stringent precautions are taken, some studies have shown unexpected complications for reasons that remain largely enigmatic. Identification of additional factors that may influence gene transfer efficiency in larger animals is of paramount importance for improving clinical applicability.

## 16.4 CHALLENGES IN HEPATOCYTE-DIRECTED GENE THERAPY

The liver is the largest internal organ of the human body and plays a central role in protein synthesis, detoxification processes, and metabolism. In addition, the liver is highly vascularized, with almost 30% of the total blood volume passing through the liver every minute, thereby facilitating efficient secretion of therapeutic transgenes produced by transduced hepatocytes. As such, parenchymal liver cells (i.e., hepatocytes) constitute a key target for therapeutic strategies aimed at correction of inborn errors of metabolism. A global overview of preclinical studies using AAV gene transfer for the correction of metabolic diseases has been described previously [16,144]. Here, we discuss the most important remaining hurdles that might impede efficient clinical translation (Fig. 16.2).

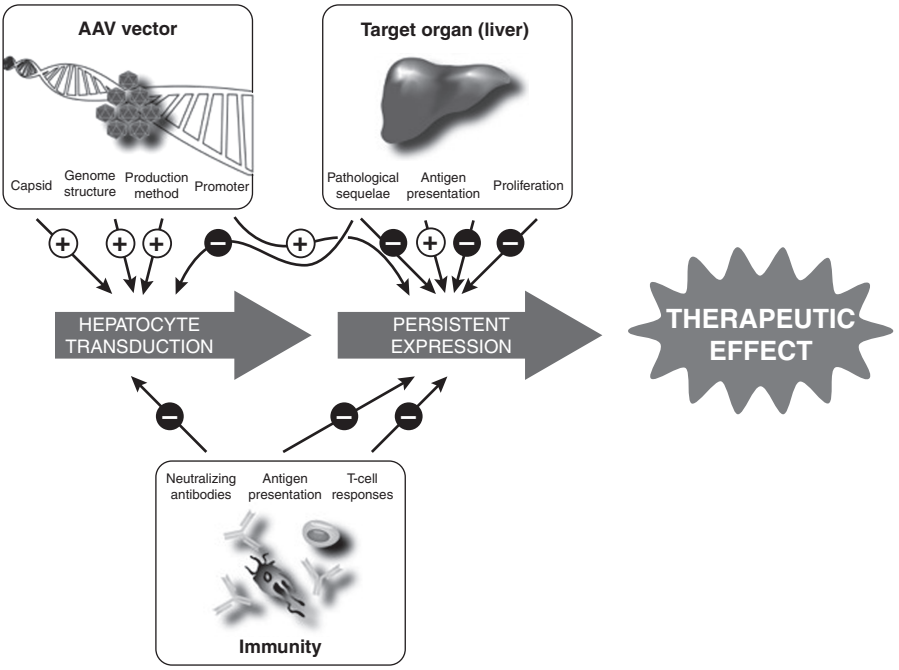


**FIGURE 16.2** Key determinants of efficient and persistent liver gene transfer with adeno-associated viral (AAV) vectors. Sustained therapeutic corrections depend on efficient hepatocyte transduction and the maintenance of persistent transgene expression. Whether or not these goals are met depends on an intricate interplay of different factors. Preexisting neutralizing antibodies greatly impair hepatocyte transduction. In addition, pathological sequelae of the liver, such as inflammation or pseudocapillarization, may also interfere with hepatocyte transduction or contribute to the priming of reactive T cells. Whereas hepatic transgene expression may induce a tolerogenic state, immune activation can also lead to the priming of a cytotoxic T-cell response or the formation of inhibitory antibodies. In pediatric patients, fast liver growth leads to a concomitant decline in transgene expression. Finally, vector design greatly influences the outcome of liver gene transfer strategies. (Adapted from Ref. [16].)

### 16.4.1 Role of Pre-exposure to Natural Adeno-Associated Viral Infections

Prior exposure to AAV has been shown in many species, including mice, dogs, NHPs, and humans. These infections result in the formation of complex profiles of serum antibodies capable of binding and neutralizing various AAV serotypes (see Fig. 16.2) [54,145]. Although human data on the effects of neutralizing antibodies on hepatocyte transduction are scarce, the absence of efficient transduction in at least one patient of the AAV2 FIX clinical trial was attributed to the higher level of neutralizing antibodies [48].

Original reports showed that a large proportion of the human population has been exposed to AAV, with at least one-third having neutralizing antibodies against AAV2

[146]. Importantly, Halbert et al. showed that in pediatric patients with cystic fibrosis, neutralizing antibodies against several AAV serotypes are remarkably lower [147]. In a recent study by the Wilson group, the authors analyzed serum samples from 888 persons, spanning four continents and 10 different countries, for the presence of neutralizing antibodies directed against a wide array of AAV serotypes [54]. In most regions, the prevalence of clinically relevant titers of neutralizing antibodies was lowest for AAV7 (15–30%) and AAV8 (15–30%). In contrast, neutralizing antibody titers against AAV1 (20–45%) and AAV2 (30–60%) were significantly more prevalent. The observed differences in seroprevalence against distinct serotypes most likely reflect their natural host tropism. Humans are not commonly pre-exposed to serotypes derived from NHPs, resulting in lower titers of neutralizing antibodies. Finally, a significant degree of co-prevalence of neutralizing antibodies against different serotypes has been observed, indicating the general breadth of neutralizing antibodies against AAV in humans. A survey of human plasma samples on preexisting neutralizing antibodies suggests that 75% of the population of the United States has sufficiently low levels of neutralizing antibodies against AAV8 to achieve a possible clinical effect after systemic gene transfer. However, even these low levels of neutralizing antibodies can significantly inhibit cell transduction after systemic administration. Alternatively, the use of plasmapheresis in diminishing neutralizing antibodies to viruses has been explored [148]. This could be a strategy for converting some of the AAV8-resistant patients to those eligible for AAV8 gene transfer. Alternatively, animal studies have indicated that interference by low titers of neutralizing antibodies could be overcome by a 10-fold increase in the vector dose [149]. Based on these results, it was estimated that high-dose gene transfer with AAV8 would successfully transduce the liver in 50% of human patients (i.e., in patients who are truly AAV8 antibody negative). In addition, temporary depletion of neutralizing antibodies by plasmapheresis was predicted to increase this to roughly 90%.

However, detection of neutralizing antibodies remains somewhat cumbersome because of technological challenges and the lack of standardization. One common assay to quantify the effects of neutralizing antibodies is *in vitro* inhibition of transduction. However, most novel serotypes poorly transduce cells *in vitro*, requiring high multiplicities of infection to achieve detectable transgene expression levels. In addition, interactions of AAV with antibodies in the presence of serum proteins may not be accurately mimicked *in vitro*, further complicating interpretation of these assays. Detection of AAV-binding neutralizing antibodies can also be performed by ELISA, although this method may not accurately predict whether or not transduction by AAV will be inhibited *in vivo*. Alternatively, the effect of neutralizing antibodies on hepatocyte transduction has been assessed by passive transfusions of sera into naïve mice in order to overcome the issue of low sensitivity of the *in vitro* inhibition assay.

In conclusion, although preexisting vector-specific antibodies represent an important hurdle for clinical translation, overcoming pre-immunization may be insufficient for guaranteeing clinically relevant transgene expression levels. Additional methods that increase transgene expression by AAV vectors (see below) remain of paramount importance in advancing toward clinical applications.

### 16.4.2 Immunologic Complications

***16.4.2.1 Immune Responses Toward Adeno-Associated Viral–Encoded Transgene Products*** Immune responses toward AAV-encoded transgene products are determined by an intricate balance between mechanisms of tolerance and immune activation. The end result will be determined by various factors that influence interaction between AAV and the host (see Fig. 16.2). Several mechanisms have been proposed to address the apparent tolerance toward AAV-encoded transgene products after liver-directed gene transfer. A critical factor is the low activation of innate immunity by AAV, which, together with the low degree of APC transduction, results in impaired activation and maturation of APCs as well as reduced expression of co-stimulatory molecules [150]. As a result of poor innate immunity induced by AAV, APCs, such as macrophages and dendritic cells, do not mature properly, thereby reducing their ability to take up, process, and present antigen. Absence of inflammatory signals after AAV exposure reduces antigen presentation on APCs and hepatocytes, resulting in impaired T-cell priming and reduced targeting of hepatocytes by cytotoxic T cells, respectively [151]. Finally, low transduction of APCs by AAV further decreases the amount of available antigen required to prime naive T cells [152]. Several studies have shown that although AAVs are capable of priming potent $CD8^+$ T-cell responses toward transgene, lack of co-stimulatory signals can result in impaired proliferation and cytokine secretion. Thus, in the absence of additional inflammatory signals, AAV results in the generation of immunologic tolerance via mechanisms of ignorance, anergy, or deletion.

However, depending on the context, an appropriate degree of inflammation can be reached that is sufficient to break AAV-induced tolerance, resulting in robust and functional $CD4^+$ and $CD8^+$ T-cell responses. Factors determining whether this threshold is reached include the vector dose, the route of administration, the serotype, preexisting immunity to AAV, the transgene, the promoter, and the host species [153,154]. This might be of particular importance for secreted proteins because uptake by APCs can result in presentation onto MHCI via a process known as "cross-presentation," possibly leading to $CD8^+$ T-cell priming. Importantly, several studies have underscored the more tolerogenic nature of the liver in maintaining persistent high transgene expression after AAV gene transfer. This ability of antigen expression in the liver to induce tolerance has been shown most convincingly in the setting of allogeneic liver transplantations, showing spontaneous acceptance of liver grafts in the absence of immunosuppression as well as induction of tolerance toward other co-transplanted organs [155]. Recent studies using novel AAV variants show that serotype capsids may also play an important role in the activation of immunity [156]. This has largely been attributed to differences in the transduction of dendritic cells by distinct serotypes. Alternatively, differences in innate immune responses elicited by different serotypes could affect their propensity to activate adaptive immunity. Nevertheless, even if the threshold for initiating a T-cell response is reached, mechanisms of active suppression may still come into play that dictate the priming of either effector T cells or suppressors. In this scenario, the micro-environment of the liver has been shown to favor production of cytokines that result

in the generation of a population of $CD4^+$ $CD25^+$ $Foxp3^+$ Treg suppressors [47,157]. However, tolerance to transgene products after AAV gene transfer appears to be more difficult to establish in large animal models. Although long-term expression of green fluorescent protein (GFP) has been easily achieved in different tissues in mice, GFP expression in the livers of macaques is only short lived or significantly reduced even with a liver-specific promoter, and animals developed $CD8^+$ T cells to GFP [124,158]. Future studies will be aimed at further elucidating the underlying mechanism that determines the threshold between establishing tolerance and generating immunity.

***16.4.2.2 Immune Responses Toward Adeno-Associated Viral Capsids*** Whereas preclinical studies with AAV2 in animal models generally resulted in persistent transgene expression, the FIX clinical trial highlighted the intricate interplay between the vector and the human immune system. Although the underlying mechanisms of these observations remain elusive, several theories have tried to address these observations [154,159]. Mingozzi et al. put forward the hypothesis that preexposure to AAV in humans may have resulted in the generation and maintenance of a small pool of $CD8^+$ memory T cells directed against the AAV2 capsid (Fig. 16.3) [159]. Because AAV2 is endemic in humans and pre-exposure occurs mostly in combination with infection of an adenovirus, this combination may trigger the innate immune system more avidly than AAV2 alone, thus exceeding the threshold for active immunity toward AAV2. Although studies substantiating these claims in humans are scarce, epidemiologic data indicate that up to 60% of the population might harbor a capsid-specific memory T-cell population [49]. After reexposure to AAV2 capsid, these memory T cells are readily reactivated, resulting in elimination of transduced hepatocytes that present peptides derived from capsid protein. One common problem in testing this hypothesis has been the inability to model the elimination of transduced hepatocytes by cytotoxic T cells in animal models. The absence of pre-exposure to AAV2 in most animal models would explain why vigorous T-cell responses toward capsid were not observed in preclinical studies. Preexisting immunity in mice has been simulated using immunization by adenoviral vectors expressing AAV capsid proteins as well as direct delivery of AAV vectors. Although T cells generated using these methodologies have been shown to be fully functional, they are unable to target and eliminate hepatocytes that have been transduced by AAV *in vivo* [94,160–162]. Nonetheless, presentation of the exogenous capsid derived from input AAV particles onto MHCI, followed by priming of a cytotoxic T-cell response, has been shown to occur in both mice and cynomolgus macaques [105,160,163]. Possible explanations for the observed higher threshold for cytotoxic T-cell clearance of hepatocytes in mice include increased responsiveness of human T cells to T-cell receptor stimulation, more efficient homing of human T cells to the liver, lack of efficient cross-presentation of AAV capsid antigen in mouse hepatocytes, and the fact that in humans cytotoxic T cells originate from a pool of memory $CD8^+$ T cells in contrast to AAV-primed $CD8^+$ T cells that are activated in mice [49].

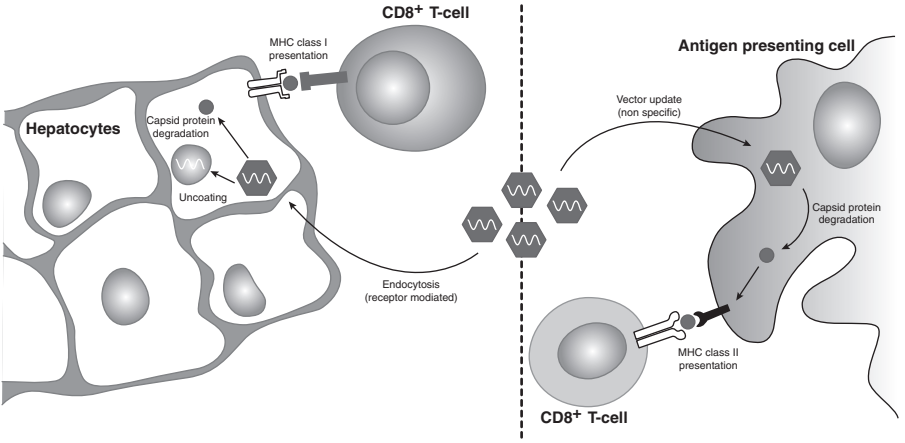


**FIGURE 16.3** Working hypothesis for the observed immunological activation in the AAV2 FIX clinical trial. Mingozzi and King [1] have put forward the "memory T-cell" hypothesis to explain the transient expression of factor IX (FIX) protein after systemic administration of adeno-associated viral 2 (AAV2) vectors. A proportion of humans have been shown to maintain a small population of memory $CD8^+$ T cells after natural infections with AAV2. In the context of the gene therapy trial, systemic reexposure to the AAV2 vector resulted in vector uptake by APCs. Capsid proteins are presented onto major histocompatibility (MHC) class II molecules, resulting in the activation of $CD4^+$ T cells and the production of inflammatory cytokines. This leads to an expansion of the memory $CD8^+$ T-cell pool. Concomitantly, transduction of hepatocytes occurs by receptor-mediated endocytosis in clathrin-coated pits. After endosomal escape, capsid proteins are ubiquitylated and proteosomally degradaded. The resulting peptides can be loaded onto MHC class I molecules via a "cross-priming" mechanism and are subsequently presented on the cell surface. As such, hepatocytes are flagged as targets for recognition by activated cytotoxic $CD8^+$ T cells, which may subsequently lead to the elimination of transduced cells and the loss of transgene expression.

Whatever the cause may be, based on the lack of data substantiating the capsid memory T-cell response, alternative hypotheses have been put forward. Vandenberghe et al. proposed a mechanism in which a specific heparin-binding motif on the AAV2 capsid favors transduction of human dendritic cells, thus facilitating activation of capsid specific $CD8^+$ T-cells [105]. As demonstrated in mice and NHPs, native or engineered AAV serotypes devoid of this motif are able to circumvent the generation of capsid-specific $CD8^+$ T-cells [94,105]. However, the usefulness of non–heparin-binding serotypes in overcoming capsid T-cell activation has recently been questioned [161,164]. It has been hypothesized that because of the high degree of homology between AAV serotypes and the conservation of the potential immunoreactive epitopes in AAV2 capsid, exposure to these serotypes may still result in activation of $CD8^+$ memory T cells [49]. Nevertheless, we believe that serotypes with reduced immunogenicity add a safety measure for clinical applications.

Alternatively, packaging of residual plasmid expressing capsid during vector production may have resulted in continuous expression of capsid proteins. However,

to date, it has not been possible to show *de novo* capsid protein expression after AAV transduction *in vitro*, and animal studies have not shown any evidence of capsid expression [165]. In contrast, Gao et al. recently showed low-level expression of *cap* gene in liver in NHPs caused by the presence of both endogenous AAV genomes and those derived from the preparations of the vectors caused by contamination with the *trans*-packaging plasmid [124]. Although the authors could not discern between both sources of cap expression, this seems irrelevant in terms of activating T cells or rendering hepatocytes targets for cytotoxic T cells because no such T cells were activated in this study. Finally, although no $CD8^+$ T-cell response against FIX was detected in the clinical trial, one recent study in mice reported the transcription of cryptic neo-epitopes encoded from an alternative open reading frame of the *FIX* gene [166].

### 16.4.3 Neonatal and Pediatric Gene Therapy

Because most metabolic diseases manifest at a young age, specific interest should be given to factors that may influence the potency of AAV in this population. Importantly, several lines of evidence indicate that hepatocyte transduction may be more efficient in younger patients. Of note, the presence of neutralizing antibodies has been shown to be reduced in pediatric patients [147]. An additional important safety concern is that in most pediatric patients, there is less tissue damage or accumulation of toxic metabolites, possibly reducing the threshold for therapeutic efficiency (see Fig. 16.2). Moreover, this reduced inflammation may be of paramount importance in evading transgene specific immune responses. In this regard, studies in neonatal dogs and cats have shown great efficiency in preventing the onset of pathologies.

A particular concern of pediatric gene transfer strategies is the potential impact of liver growth on transgene DNA persistence. Most AAV genomes reside within the nucleus in an episomal form. As a result, transgene DNA is lost during mitosis when the nuclear envelope is degraded. In the adult mouse liver, individual hepatocytes are slowly dividing, completing a single population turnover every 180 to 400 days [167]. Consequently, a rapid decline of transgene expression has been observed in the context of rapid hepatocyte proliferation after AAV administration into neonatal mice [102,168]. Interestingly, when vector is delivered at serially higher ages up to adulthood, transgene expression is progressively more durable, correlating with the slowing of liver growth [102]. Taken together, these studies indicate that proliferation of hepatocytes results in a nearly complete clearance of AAV genomes from the liver within two population doublings, corresponding to a time frame of 1 to 2 weeks in mice [102,106,168,169].

The remaining question is to which extent these results can be extrapolated to other species. It seems straightforward that hepatocyte proliferation will result in substantial loss of transgene DNA irrespective of the species studied. Nevertheless, this process is likely to be substantially slower in larger models. In this respect, neonatal gene transfer has recently been shown to result in long-term correction in a feline model of mucoplysaccharidosis type VI or Maroteaux-Lamy syndrome (MPS VI) [170]. As

for humans, it is estimated that a first doubling of the liver mass occurs within 9 months after birth followed by a second doubling in mass taking up to 3 to 4 years [171]. By inference from mouse data, Cunningham et al. [102] estimated that at similar levels of episome clearance with each doubling of liver mass, therapeutic transgene expression in humans would be expected to fall substantially by the end of the first year of life if administered at the neonatal stage. Thus, depending on the levels of transgene expression required for therapeutic benefit, vector readministration is likely to be necessary. Several groups have indeed underscored the possibility of using distinct AAV serotypes to circumvent preexisting immunity [99,172,173].

## 16.5 REGULATORY ASPECTS OF GENE THERAPY

The swift progress of various gene therapy applications into the clinic has also highlighted another critical issue within the community: a general lack in regulatory experience and a relative vacuum of regulatory guidelines for these continuously changing gene therapy technologies [174,175]. From a regulatory aspect, gene therapy products represent one of the most complex classes of therapeutics, combining both elements of nucleic acids as well as a complex mix of proteins. As a result, the development of gene therapy products is highly regulated, and development of a gene therapy drug into the clinic is associated with unique challenges. Although this can sometimes be a quite frustrating and overwhelming process, it is imperative for researchers to collaborate with regulatory agencies early on in the developmental process to avoid unnecessary delays in the development of gene therapy products. Indeed, regulatory guidelines are not legally binding but rather represent the current vision of regulatory agencies at that time. As such, regulators should be regarded as collaborators whose comments can significantly help expedite the product development. As will be discussed in the following paragraphs, the concerted goal of these organizations is to ensure the safety of both the study participants and the researchers themselves by ensuring that proper guidelines for vector production (GMP) and the conduct of the clinical trials (GCP) are stringently followed. As such, the regulation of clinical gene therapy trials combines aspects related to both production and characterization of the clinical product, as well as ethical considerations related to study of human participants [11].

### 16.5.1 Guidelines and Regulatory Organizations

Regulatory bodies that are responsible for the oversight of clinical gene therapy trials are organized in several national and supranational organizations [174,176]. In the United States, all clinical trials are supervised by the Department of Health and Human Services. However, the final responsibility is further shared across different regulatory organs, and a centralized oversight is lacking. As such, research subjects fall under the protection of the Office for Human Research Protection (OHRP), but the Food and Drug Administration (FDA) has specific oversight of the therapeutic gene therapy products. Therefore, before proceeding with the clinical use of any

unlicensed gene therapy product, the investigators must submit an Investigational New Drug (IND) application with the FDA. The FDA reviews all facets of gene therapy trials, including preclinical efficacy, toxicology, product manufacture and certification, and the design and conduct of the clinical trial. The OHRP further mandates that all research involving human subjects undergo full review and approval by an institutional review board (IRB). The IRB is responsible for evaluating the research risk for study participants and must review the clinical study protocol and informed consent documentation before initiation of the study.

Of note, any clinical trial that is supported by funding from the National Institutes of Health (NIH) will also be placed under oversight by the NIH, which is performed by the Office of Biotechnological Activities (OBA). However, final responsibilities are further taken over by the Recombinant DNA Advisory Committee (RAC) and the Institutional Biosafety Committee (IBC). The RAC aims to develop specific policies and regulations for performing research involving recombinant DNA molecules, with a specific emphasis on safety and scientific conduct. On the other hand, the role of the IBC is to assess possible risks to the public and to review individual research protocol as well as to ensure compliance with NIH guidelines. Final oversight is performed by the OBA, which is charged with the federal oversight of all IBCs and performs direct monitoring and review via the RAC. Importantly, the RAC does not approve clinical protocols but provides suggestions for revisions to address scientific, ethical, or legal comments. Before initiation of any trial, the FDA will oversee a complete review of the proposed IND.

In Europe, the responsibility for approval and review of clinical gene therapy trials is predominantly the responsibility of the member states of the European Union. The application for approval starts with the submission of the Investigational Medicinal Product Dossier (IMPD) to the responsible national agency (see complete list: http://www.hma.eu). This IMPD contains a complete description of all facets of the proposed trials, including a project description, available preclinical data, study rationale, detailed experimental procedures, materials, assessment of pharmaceutical quality, and a formal biosafety evaluation. In addition, a complete description of the scheduled clinical trial must be submitted to the appropriate ethical committee. Although the European Union has implemented certain guidelines (e.g., Directive 2001/20/EC) to harmonize procedures for authorizing clinical trials across member states, the detailed procedures remain somewhat different [176]. Furthermore, laws and regulations pertaining to the use and deliberate release of genetically modified organisms (GMOs) and the use of stem cells also differ in among member states, further complicating a synchronized program for developing gene therapy products across the European Union. Whereas some member states regard gene transfer vectors as a deliberate release of GMOs (in line with Directive 2001/18/EC), others regard them as GMOs for contained release (according to Directive 2009/41/EC). Although both directives aim to protect the individuals and their environment, these differences in classifications significantly impact on the design and the review period of the application. Unfortunately, Europe lacks a centralized regulatory organ with an independence and accountability similar to the RAC. However, to further assist in the development of gene therapeutics, the European Union has set up specific provisions

such as that gene therapy products can be developed under the orphan drug designation by the European Commission [11]. In addition, the European Union has funded the CliniGene Network of Excellence (NoE) to further support researchers to perform high-quality gene transfer research and its translation into successful clinical trials. Nevertheless, a more centralized expert review within the European Union would be more efficient and less time consuming and could help circumvent a relative lack of critical gene therapy research expertise in certain member states [177].

Although there are some important distinctions in the organization of these regulatory instances, final review of the submission is subject to similar obstacles in both the Unites States as in Europe. Of note, review bodies may lack experience and expertise in the science of gene therapy, increasing the likelihood that oversight could fail to address the most salient issues while giving excessive attention to less important aspects of gene therapy research [177]. Similarly, many (if not most) gene therapy products are being developed by investigators at nonprofit research institutions, which may lack the necessary experience in dealing with regulatory instances. Finally, the increasing clinical experience with viral gene transfer vectors also directly impacts on the regulatory review of future studies. On the positive end, this increasing experience may help create an extensive regulatory framework that can help facilitate future submissions. On the other hand, new studies may also highlight novel pitfalls or complications, underscoring the need for additional safety guidelines and regulations. In light of the complexity of these regulatory aspects of gene therapy products, a continuous reexamination of the ethical issues surrounding the application of gene therapy into clinical practice is warranted.

### 16.5.2 Vector Production and Product Characterization

The shift toward large-animal studies and first generations of clinical trials has led to a significant increase in vector needs. However, production of purified AAV vectors at the required scale has proven to be difficult to scale up, relying on highly technical and costly procedures. To overcome this impediment, several recent innovative production systems have been developed that are compatible with large-scale production for clinical applications [178–180]. These new systems use adenovirus, herpesvirus, and baculovirus hybrids to deliver the AAV genome and *trans*-acting helper functions to producer cells. Irrespective of the modus of production, several recent studies have identified vector purification methods as a key determinant of vector potency *in vivo* [181,182]. Increasing vector purity and potency is predicted to reduce the risk of deleterious immune responses after administration of recombinant AAV to human subjects. Standard purification of AAV vectors requires multiple rounds of CsCl ultracentrifugation. However, these gradients are limited in terms of their loading capacity, and prolonged exposure to CsCl has been reported to compromise AAV vector potency. Large-scale purification of AAV has been based on ion-exchange chromatography, hydrophobic interaction, and affinity purification methods, resulting in increased purity, yields, and scalability. However, these methods are generally serotype dependent and lack the ability to discern between

fully functional AAV particles and empty capsids. In contrast, density centrifugation using iodixanol gradients effectively separates infectious particles from empty capsids, a desirable property for reducing toxicity and unwanted immune responses during preclinical studies [182]. Finally, an important question remains regarding which criteria accurately predict the overall quality of different vector lots. Several parameters have been developed to further characterize purified AAV stocks, including silver staining of SDS-PAGE (sodium dodecyl sulfate polyacrylamide gel electrophoresis) separation of proteins, ratio of rAAV full capsids to infectious particles, protein contamination, and DNA contamination. However, although these analyses focus predominantly on physical features of the vector preparations that are easily standardized, further studies are needed to assess the validity of these parameters as predictors of vector potency and safety *in vivo*.

Before any gene therapy product can be formally tested in clinical trials, the production methods must be standardized in order to comply with the GMP standards. These standards are a set of scientific methods, principles, or practices that should be implemented during production to ensure the consistent manufacturing of safe, pure, and potent products. The emphasis of GMP guidelines lies in documentation and traceability. As discussed by Laurencot et al. [183], the highly complex biological nature of gene therapy vectors represents numerous manufacturing challenges. Therefore, the FDA may not require full GMP compliance during all stages of development but rather an incremental approach during the different phases of development (i.e., the so-called "sliding scale" of GMP) [183]. Nevertheless, manufacturing of gene therapy products according to GMP standards for clinical trials can be one of the biggest hurdles to overcome. For these reasons, it is quite common that this process will be contracted out to a registered GMP facility that already has relevant expertise.

## 16.6 GENERAL CONCLUSIONS AND PERSPECTIVES

The past decades have seen several successes in the gene therapy field. However, optimism must be tempered with caution. Indeed, despite the significant progress, many important obstacles remain that might impede widespread application of gene therapy applications in clinical practice. The success of gene therapy depends on a multitude of parameters. Although some of these parameters are universally relevant for different diseases, others are highly dependent on the disease, the target tissue, or the transgene. In this regard, most studies have underscored the direct impact of the stage of the disease at the time of intervention on the final therapeutic outcome. Therefore, it is likely that optimal age ranges and disease classifications may need to be established, which could be extremely disease specific. However, this also leads to several ethical dilemmas. Given the risk of insertional mutagenesis and genotoxicity of most vector systems, the relative paucity of studies including long-term follow-up precludes adequate risk assessments. In addition, data on viral shedding have indicated the presence of transgene DNA sequences in semen, raising concerns over inadvertent germline transmission [177]. These concerns are of specific interest

when discussing the eligibility of young children for gene therapy approaches. Given the fact that preliminary data indicate that the possible therapeutic effects of gene transfer may be more pronounced when pathological conditions are reversed early in the disease progression, one important question is when gene therapy approaches should be considered. Is it wise to compromise patients safety in order to achieve maximal therapeutic effects or should gene therapy only be considered as a last resort, even if this can limit the possible therapeutic effect [177]? In other words, should priority be given to safety or efficacy?

Alternatively, the development of immune responses directed against the gene therapy vector or the therapeutic transgene remains an important safety issue. Although several factors that can influence the degree of immune stimulation have been identified (e.g., vector dose, route of administration, vector type, type of transgene, target tissue), their respective roles remain largely enigmatic. The immunostimulatory role of viral gene transfer vectors is also of great importance for certain applications that may require repeated vector administrations because the formation of a neutralizing antibody response will preclude efficient transduction upon readministration of the vector. To overcome these immunologic hurdles, several groups are aiming to minimize the immunogenic risk by using lower vector doses or administration of immunosuppressive drug regimens. These strategies are currently under investigation in clinical trials, which will indicate if these interventions also translate into improved clinical benefits.

Finally, one of the remaining hurdles for the gene therapy remains the development of successful business models to encourage the participation of the biotech industry in the development of commercial gene therapy products [184]. Indeed, whereas commercial biotech companies played a important role in the early development of gene therapy products, this interest has gradually faded because of the significant setbacks highlighted in this chapter as well as the high cost and long timelines associated with the development process. As a result, early companies eventually failed to generate sufficient capital to maintain their investments. Now the emerging positive results from clinical gene therapy trials may spark novel interest by the biotech industry. Almost every venture capital firm and biopharmaceutical company is currently reevaluating the possible commercial value of gene therapy products [9]. This new interest can help provide two scarce resources to the gene therapy field: regulatory experience and financial support. Interestingly, important concerns are not related to technical feasibility but rather to the uncertainty of regulatory processes and concerns about the business model. As illustrated by the lengthy and tortuous approval applications for Glybera, achieving the final goal of reaching marketing approval can be very costly and must be pursued without any guarantees for success. Even though there are currently almost no approved gene therapy products in Europe and the United States, the development of gene therapy products is moving forward with enthusiasm. The question thus remains how to develop a model that is robust enough to provide a balance between profitability (for developers and investors) and affordability (for patients and government agencies) [184].

In conclusion, the field of gene therapy has matured significantly over the past decades. Early hype, failures, and tragic events have largely been replaced by the

necessary stepwise progress needed to realize clinical benefits. The fruits of these efforts are now showing in the first generation of successful clinical trials. Of note, the developmental timeline of gene therapy products is comparable to other novel therapies such as bone marrow transplantations and monoclonal antibody therapies. However, amid these invigorating times, it will be of paramount importance to remain focused on the continuous refinement of gene transfer technologies and the identification of hitherto unknown complications. The importance of these refinements is exemplified by more recent successes, such as the use of novel generation lentiviral vectors for the treatment of adrenoleukodystrophy [185] and of AAV8 vectors for the treatment of hemophilia B [50]. Although it is tempting to quickly move such advancements forward into the clinic, great caution must be taken not to impede on benefits to patients in the long term. With increasing experience, the gene therapy field will continue to mature in order to certify an ordered and timely transition from bench to bedside.

## REFERENCES


1. Mingozzi F, High KA. Therapeutic in vivo gene transfer for genetic disease using AAV: progress and challenges. Nat Rev Genet 2011;12(5):341–355.
2. Williams PD, Kingston PA. Plasmid-mediated gene therapy for cardiovascular disease. Cardiovasc Res 2011;91(4):565–576.
3. Schleef M, Blaesen M, Schmeer M, Baier R, Marie C, Dickson G, et al. Production of non viral DNA vectors. Curr Gene Ther 2010;10(6):487–507.
4. Seow Y, Wood MJ. Biological gene delivery vehicles: beyond viral vectors. Mol Ther 2009;17(5):767–777.
5. Lederberg J. Biological future of man. In: Wolstenholme G, editor. Man and His Future. London: Churchill; 1963. p. 265.
6. Tatum EL. Molecular biology, nucleic acids, and the future of medicine. Perspect Biol Med 1966;10(1):19–32.
7. Tatum EL. Molecular biology, nucleic acids, and the future of medicine. In: Reflections on Research and the Future of Medicine. New York: McGraw-Hill; 1967. p. 31–49.
8. Nirenberg MW. Will society be prepared? Science 1967;157(3789):633.
9. Wilson J. The history and promise of gene therapy. Genet Eng Biotechnol News 2011; 31(17):1–2.
10. Gillet JP, Macadangdang B, Fathke RL, Gottesman MM, Kimchi-Sarfaty C. The development of gene therapy: from monogenic recessive disorders to complex diseases such as cancer. Methods Mol Biol 2009;542:5–54.
11. Bamford KB. Clinical trials of GMP products in the gene therapy field. Methods Mol Biol 2011;737:425–442.
12. Anderson WF, Killos L, Sanders-Haigh L, Kretschmer PJ, Diacumakos EG. Replication and expression of thymidine kinase and human globin genes microinjected into mouse fibroblasts. Proc Natl Acad Sci USA 1980;77(9):5399–5403.

13. Miller AD, Jolly DJ, Friedmann T, Verma IM. A transmissible retrovirus expressing human hypoxanthine phosphoribosyltransferase (HPRT): gene transfer into cells obtained from humans deficient in HPRT. Proc Natl Acad Sci USA 1983;80 (15):4709–4713.
14. Van Doren K, Hanahan D, Gluzman Y. Infection of eucaryotic cells by helper-independent recombinant adenoviruses: early region 1 is not obligatory for integration of viral DNA. J Virol 1984;50(2):606–614.
15. Tratschin JD, West MH, Sandbank T, Carter BJ. A human parvovirus, adeno-associated virus, as a eucaryotic vector: transient expression and encapsidation of the procaryotic gene for chloramphenicol acetyltransferase. Mol Cell Biol 1984; 4(10):2072–2081.
16. Jacobs F, Wang L. Adeno-associated viral vectors for correction of inborn errors of metabolism: progressing towards clinical application. Curr Pharm Des 2011;17 (24):2500–2515.
17. High KA. Gene therapy for haemophilia: a long and winding road. J Thromb Haemost 2011;9(Suppl 1):2–11.
18. DiPrimio N, McPhee SW, Samulski RJ. Adeno-associated virus for the treatment of muscle diseases: toward clinical trials. Curr Opin Mol Ther 2010;12(5):553–560.
19. Raper SE, Chirmule N, Lee FS, Wivel NA, Bagg A, Gao GP, et al. Fatal systemic inflammatory response syndrome in a ornithine transcarbamylase deficient patient following adenoviral gene transfer. Mol Genet Metab 2003;80(1–2):148–158.
20. Gaspar HB, Cooray S, Gilmour KC, Parsley KL, Adams S, Howe SJ, et al. Long-term persistence of a polyclonal T cell repertoire after gene therapy for X-linked severe combined immunodeficiency. Sci Transl Med 2011;3(97):97ra79.
21. Gaspar HB, Cooray S, Gilmour KC, Parsley KL, Zhang F, Adams S, et al. Hematopoietic stem cell gene therapy for adenosine deaminase-deficient severe combined immunodeficiency leads to long-term immunological recovery and metabolic correction. Sci Transl Med 2011;3(97):97ra80.
22. Gaspar HB, Aiuti A, Porta F, Candotti F, Hershfield MS, Notarangelo LD. How I treat ADA deficiency. Blood 2009;114(17):3524–3532.
23. Blaese RM, Culver KW, Miller AD, Carter CS, Fleisher T, Clerici M, et al. T lymphocyte-directed gene therapy for ADA-SCID: initial trial results after 4 years. Science 1995;270(5235):475–480.
24. Onodera M, Ariga T, Kawamura N, Kobayashi I, Ohtsu M, Yamada M, et al. Successful peripheral T-lymphocyte-directed gene transfer for a patient with severe combined immune deficiency caused by adenosine deaminase deficiency. Blood 1998;91(1):30–36.
25. Kohn DB, Weinberg KI, Nolta JA, Heiss LN, Lenarsky C, Crooks GM. Engraftment of gene-modified umbilical cord blood cells in neonates with adenosine deaminase deficiency. Nat Med 1995;1(10):1017–1023.
26. Shaw KL, Kohn DB. A tale of two SCIDs. Sci Transl Med 2011;3(97):97ps36.
27. Aiuti A, Slavin S, Aker M, Ficara F, Deola S, Mortellaro A, et al. Correction of ADA-SCID by stem cell gene therapy combined with nonmyeloablative conditioning. Science 2002;296(5577):2410–2413.
28. Aiuti A, Cattaneo F, Galimberti S, Benninghoff U, Cassani B, Callegaro L, et al. Gene therapy for immunodeficiency due to adenosine deaminase deficiency. N Engl J Med 2009;360(5):447–458.

29. Hacein-Bey-Abina S, Le Deist F, Carlier F, Bouneaud C, Hue C, De Villartay JP, et al. Sustained correction of X-linked severe combined immunodeficiency by ex vivo gene therapy. N Engl J Med 2002;346(16):1185–1193.
30. Hacein-Bey-Abina S, Garrigue A, Wang GP, Soulier J, Lim A, Morillon E, et al. Insertional oncogenesis in 4 patients after retrovirus-mediated gene therapy of SCID-X1. J Clin Invest 2008;118(9):3132–3142.
31. Baum C. Gene therapy for SCID-X1: focus on clinical data. Mol Ther 2011;19 (12):2103–2104.
32. Fischer A, Hacein-Bey-Abina S, Cavazzana-Calvo M. Gene therapy for primary adaptive immune deficiencies. J Allergy Clin Immunol 2010;127(6):1356–1359.
33. Howe SJ, Mansour MR, Schwarzwaelder K, Bartholomae C, Hubank M, Kempski H, et al. Insertional mutagenesis combined with acquired somatic mutations causes leukemogenesis following gene therapy of SCID-X1 patients. J Clin Invest 2008; 118(9):3143–3150.
34. Ramezani A, Hawley TS, Hawley RG. Combinatorial incorporation of enhancer-blocking components of the chicken beta-globin 5′HS4 and human T-cell receptor alpha/delta BEAD-1 insulators in self-inactivating retroviral vectors reduces their genotoxic potential. Stem Cells 2008;26(12):3257–3266.
35. Mortellaro A, Hernandez RJ, Guerrini MM, Carlucci F, Tabucchi A, Ponzoni M, et al. Ex vivo gene therapy with lentiviral vectors rescues adenosine deaminase (ADA)-deficient mice and corrects their immune and metabolic defects. Blood 2006;108(9):2979–2988.
36. Thornhill SI, Schambach A, Howe SJ, Ulaganathan M, Grassman E, Williams D, et al. Self-inactivating gammaretroviral vectors for gene therapy of X-linked severe combined immunodeficiency. Mol Ther 2008;16(3):590–598.
37. Manco-Johnson MJ, Abshire TC, Shapiro AD, Riske B, Hacker MR, Kilcoyne R, et al. Prophylaxis versus episodic treatment to prevent joint disease in boys with severe hemophilia. N Engl J Med 2007;357(6):535–544.
38. High KA. The Jeremiah Metzger Lecture: gene therapy for inherited disorders: from Christmas disease to Leber's amaurosis. Trans Am Clin Climatol Assoc 2009;120: 331–359.
39. Herzog RW, Hagstrom JN, Kung SH, Tai SJ, Wilson JM, Fisher KJ, et al. Stable gene transfer and expression of human blood coagulation factor IX after intramuscular injection of recombinant adeno-associated virus. Proc Natl Acad Sci USA 1997;94 (11):5804–5809.
40. Herzog RW, Mount JD, Arruda VR, High KA, Lothrop CD Jr. Muscle-directed gene transfer and transient immune suppression result in sustained partial correction of canine hemophilia B caused by a null mutation. Mol Ther 2001;4(3):192–200.
41. Manno CS, Chew AJ, Hutchison S, Larson PJ, Herzog RW, Arruda VR, et al. AAV-mediated factor IX gene transfer to skeletal muscle in patients with severe hemophilia B. Blood 2003;101(8):2963–2972.
42. Kay MA, Manno CS, Ragni MV, Larson PJ, Couto LB, McClelland A, et al. Evidence for gene transfer and expression of factor IX in haemophilia B patients treated with an AAV vector. Nat Genet 2000;24(3):257–261.
43. Jiang H, Pierce GF, Ozelo MC, de Paula EV, Vargas JA, Smith P, et al. Evidence of multiyear factor IX expression by AAV-mediated gene transfer to skeletal muscle in an individual with severe hemophilia B. Mol Ther 2006;14(3):452–455.

44. Arruda VR, Schuettrumpf J, Herzog RW, Nichols TC, Robinson N, Lotfi Y, et al. Safety and efficacy of factor IX gene transfer to skeletal muscle in murine and canine hemophilia B models by adeno-associated viral vector serotype 1. Blood 2004; 103(1):85–92.
45. Nakai H, Herzog RW, Hagstrom JN, Walter J, Kung SH, Yang EY, et al. Adeno-associated viral vector-mediated gene transfer of human blood coagulation factor IX into mouse liver. Blood 1998;91(12):4600–4607.
46. Niemeyer GP, Herzog RW, Mount J, Arruda VR, Tillson DM, Hathcock J, et al. Long-term correction of inhibitor-prone hemophilia B dogs treated with liver-directed AAV2-mediated factor IX gene therapy. Blood 2009;113(4):797–806.
47. Mingozzi F, Liu YL, Dobrzynski E, Kaufhold A, Liu JH, Wang Y, et al. Induction of immune tolerance to coagulation factor IX antigen by in vivo hepatic gene transfer. J Clin Invest 2003;111(9):1347–1356.
48. Manno CS, Arruda VR, Pierce GF, Glader B, Ragni M, Rasko J, et al. Successful transduction of liver in hemophilia by AAV-Factor IX and limitations imposed by the host immune response. Nat Med 2006;12(3):342–347.
49. Mingozzi F, Maus MV, Hui DJ, Sabatino DE, Murphy SL, Rasko JE, et al. CD8(+) T-cell responses to adeno-associated virus capsid in humans. Nat Med 2007;13(4):419–422.
50. Nathwani AC, Tuddenham EG, Rangarajan S, Rosales C, McIntosh J, Linch DC, et al. Adenovirus-associated virus vector-mediated gene transfer in hemophilia B. N Engl J Med 2011;365(25):2357–2365.
51. McCarty DM, Monahan PE, Samulski RJ. Self-complementary recombinant adeno-associated virus (scAAV) vectors promote efficient transduction independently of DNA synthesis. Gene Ther 2001;8(16):1248–1254.
52. Nathwani AC, Gray JT, Ng CY, Zhou J, Spence Y, Waddington SN, et al. Self-complementary adeno-associated virus vectors containing a novel liver-specific human factor IX expression cassette enable highly efficient transduction of murine and nonhuman primate liver. Blood 2006;107(7):2653–2661.
53. Mingozzi F, Meulenberg JJ, Hui DJ, Basner-Tschakarjan E, Hasbrouck NC, Edmonson SA, et al. AAV-1-mediated gene transfer to skeletal muscle in humans results in dose-dependent activation of capsid-specific T cells. Blood 2009;114(10):2077–2086.
54. Calcedo R, Vandenberghe LH, Gao G, Lin J, Wilson JM. Worldwide epidemiology of neutralizing antibodies to adeno-associated viruses. J Infect Dis 2009;199(3):381–390.
55. den Hollander AI, Roepman R, Koenekoop RK, Cremers FP. Leber congenital amaurosis: genes, proteins and disease mechanisms. Prog Retin Eye Res 2008;27(4):391–419.
56. Jacobson SG, Aleman TS, Cideciyan AV, Sumaroka A, Schwartz SB, Windsor EA, et al. Identifying photoreceptors in blind eyes caused by RPE65 mutations: prerequisite for human gene therapy success. Proc Natl Acad Sci USA 2005;102(17):6177–6182.
57. Acland GM, Aguirre GD, Ray J, Zhang Q, Aleman TS, Cideciyan AV, et al. Gene therapy restores vision in a canine model of childhood blindness. Nat Genet 2001;28(1):92–95.
58. Bainbridge JW, Smith AJ, Barker SS, Robbie S, Henderson R, Balaggan K, et al. Effect of gene therapy on visual function in Leber's congenital amaurosis. N Engl J Med 2008;358(21):2231–2239.
59. Maguire AM, Simonelli F, Pierce EA, Pugh EN Jr, Mingozzi F, Bennicelli J, et al. Safety and efficacy of gene transfer for Leber's congenital amaurosis. N Engl J Med 2008; 358(21):2240–2248.

60. Maguire AM, High KA, Auricchio A, Wright JF, Pierce EA, Testa F, et al. Age-dependent effects of RPE65 gene therapy for Leber's congenital amaurosis: a phase 1 dose-escalation trial. Lancet 2009;374(9701):1597–1605.
61. Acland GM, Aguirre GD, Bennett J, Aleman TS, Cideciyan AV, Bennicelli J, et al. Long-term restoration of rod and cone vision by single dose rAAV-mediated gene transfer to the retina in a canine model of childhood blindness. Mol Ther 2005; 12(6):1072–1082.
62. Vandenberghe LH, Auricchio A. Novel adeno-associated viral vectors for retinal gene therapy. Gene Ther 2012;19(2):162–168.
63. Gaudet D, de Wal J, Tremblay K, Dery S, van Deventer S, Freidig A, et al. Review of the clinical development of alipogene tiparvovec gene therapy for lipoprotein lipase deficiency. Atheroscler Suppl 2010;11(1):55–60.
64. Goldberg IJ, Merkel M. Lipoprotein lipase: physiology, biochemistry, and molecular biology. Front Biosci 2001;6:D388–D405.
65. Gagne C, Brun LD, Julien P, Moorjani S, Lupien PJ. Primary lipoprotein-lipase-activity deficiency: clinical investigation of a French Canadian population. CMAJ 1989;140 (4):405–411.
66. Frossard JL, Steer ML, Pastor CM. Acute pancreatitis. Lancet 2008;371(9607):143–152.
67. Wittrup HH, Tybjaerg-Hansen A, Nordestgaard BG. Lipoprotein lipase mutations, plasma lipids and lipoproteins, and risk of ischemic heart disease. A meta-analysis. Circulation 1999;99(22):2901–2907.
68. Yla-Herttuala S. Endgame: glybera finally recommended for approval as the first gene therapy drug in the European Union. Mol Ther 2012;20(10):1831–1832.
69. Lloyd-Jones D, Adams R, Carnethon M, De Simone G, Ferguson TB, Flegal K, et al. Heart disease and stroke statistics—2009 update: a report from the American Heart Association Statistics Committee and Stroke Statistics Subcommittee. Circulation 2009;119(3):e21–181.
70. Janssens SP. Cardiac bone marrow cell therapy: the proof of the pudding remains in the eating. Eur Heart J 2011;32(14):1697–1700.
71. Byrne MJ, Power JM, Preovolos A, Mariani JA, Hajjar RJ, Kaye DM. Recirculating cardiac delivery of AAV2/1SERCA2a improves myocardial function in an experimental model of heart failure in large animals. Gene Ther 2008;15(23):1550–1557.
72. Kawase Y, Ly HQ, Prunier F, Lebeche D, Shi Y, Jin H, et al. Reversal of cardiac dysfunction after long-term expression of SERCA2a by gene transfer in a pre-clinical model of heart failure. J Am Coll Cardiol 2008;51(11):1112–1119.
73. Hasenfuss G, Reinecke H, Studer R, Meyer M, Pieske B, Holtz J, et al. Relation between myocardial function and expression of sarcoplasmic reticulum $Ca^{2+}$-ATPase in failing and nonfailing human myocardium. Circ Res 1994;75(3):434–442.
74. Gianni D, Chan J, Gwathmey JK, del Monte F, Hajjar RJ. SERCA2a in heart failure: role and therapeutic prospects. J Bioenerg Biomembr 2005;37(6):375–380.
75. Sakata S, Lebeche D, Sakata N, Sakata Y, Chemaly ER, Liang LF, et al. Restoration of mechanical and energetic function in failing aortic-banded rat hearts by gene transfer of calcium cycling proteins. J Mol Cell Cardiol 2007;42(4):852–861.
76. Jessup M, Greenberg B, Mancini D, Cappola T, Pauly DF, Jaski B, et al. Calcium upregulation by percutaneous administration of gene therapy in cardiac disease

(CUPID): a phase 2 trial of intracoronary gene therapy of sarcoplasmic reticulum $Ca^{2+}$-ATPase in patients with advanced heart failure. Circulation 2011;124(3):304–313.

77. Naldini L. Ex vivo gene transfer and correction for cell-based therapies. Nat Rev Genet 2011;12(5):301–315.
78. Brunetti-Pierri N, Ng P. Helper-dependent adenoviral vectors for liver-directed gene therapy. Hum Mol Genet 2011;20(R1):R7–R13.
79. Jacobs F, Wisse E, De Geest B. The role of liver sinusoidal cells in hepatocyte-directed gene transfer. Am J Pathol 2010;176(1):14–21.
80. Atchison RW, Casto BC, Hammon WM. Adenovirus-associated defective virus particles. Science 1965;149:754–756.
81. Gao G, Vandenberghe LH, Alvira MR, Lu Y, Calcedo R, Zhou X, et al. Clades of Adeno-associated viruses are widely disseminated in human tissues. J Virol 2004;78(12):6381–6388.
82. Gao GP, Alvira MR, Wang L, Calcedo R, Johnston J, Wilson JM. Novel adeno-associated viruses from rhesus monkeys as vectors for human gene therapy. Proc Natl Acad Sci USA 2002;99(18):11854–11859.
83. Chiorini JA, Kim F, Yang L, Kotin RM. Cloning and characterization of adeno-associated virus type 5. J Virol 1999; 73(2):1309–1319.
84. Muramatsu S, Mizukami H, Young NS, Brown KE. Nucleotide sequencing and generation of an infectious clone of adeno-associated virus 3. Virology 1996; 221(1):208–217.
85. Gao G, Vandenberghe LH, Wilson JM. New recombinant serotypes of AAV vectors. Curr Gene Ther 2005;5(3):285–297.
86. Stilwell JL, Samulski RJ. Adeno-associated virus vectors for therapeutic gene transfer. Biotechniques 2003;34(1):148–150, 152, 154 passim.
87. Surosky RT, Urabe M, Godwin SG, McQuiston SA, Kurtzman GJ, Ozawa K, et al. Adeno-associated virus Rep proteins target DNA sequences to a unique locus in the human genome. J Virol 1997;71(10):7951–7959.
88. Chen H. Comparative observation of the recombinant adeno-associated virus 2 using transmission electron microscopy and atomic force microscopy. Microsc Microanal 2007;13(5):384–389.
89. Jacobs F, Feng Y, Van Craeyveld E, Lievens J, Snoeys J, De Geest B. Species differences in hepatocyte-directed gene transfer: implications for clinical translation. Curr Gene Ther 2009;9(2):83–90.
90. Snoeys J, Lievens J, Wisse E, Jacobs F, Duimel H, Collen D, et al. Species differences in transgene DNA uptake in hepatocytes after adenoviral transfer correlate with the size of endothelial fenestrae. Gene Ther 2007;14(7):604–612.
91. Lievens J, Snoeys J, Vekemans K, Van Linthout S, de Zanger R, Collen D, et al. The size of sinusoidal fenestrae is a critical determinant of hepatocyte transduction after adeno-viral gene transfer. Gene Ther 2004;11(20):1523–1531.
92. Wisse E, Jacobs F, Topal B, Frederik P, De Geest B 2008., The size of endothelial fenestrae in human liver sinusoids: implications for hepatocyte-directed gene transfer. Gene Ther 15(17):1193–1199.
93. Davidoff AM, Gray JT, Ng CY, Zhang Y, Zhou J, Spence Y, et al. Comparison of the ability of adeno-associated viral vectors pseudotyped with serotype 2, 5, and 8 capsid

proteins to mediate efficient transduction of the liver in murine and nonhuman primate models. Mol Ther 2005;11(6):875–888.

94. Wang L, Figueredo J, Calcedo R, Lin J, Wilson JM. Cross-presentation of adeno-associated virus serotype 2 capsids activates cytotoxic T cells but does not render hepatocytes effective cytolytic targets. Hum Gene Ther 2007;18(3):185–194.
95. Wang L, Wang H, Bell P, McCarter RJ, He J, Calcedo R, et al. Systematic evaluation of AAV vectors for liver directed gene transfer in murine models. Mol Ther 2010; 18(1):118–125.
96. Zincarelli C, Soltys S, Rengo G, Rabinowitz JE. Analysis of AAV serotypes 1-9 mediated gene expression and tropism in mice after systemic injection. Mol Ther 2008;16(6):1073–1080.
97. Gao G, Lu Y, Calcedo R, Grant RL, Bell P, Wang L, et al. Biology of AAV serotype vectors in liver-directed gene transfer to nonhuman primates. Mol Ther 2006;13(1): 77–87.
98. Jiang H, Couto LB, Patarroyo-White S, Liu T, Nagy D, Vargas JA, et al. Effects of transient immunosuppression on adenoassociated, virus-mediated, liver-directed gene transfer in rhesus macaques and implications for human gene therapy. Blood 2006;108 (10):3321–3328.
99. Wang L, Calcedo R, Nichols TC, Bellinger DA, Dillow A, Verma IM, et al. Sustained correction of disease in naive and AAV2-pretreated hemophilia B dogs: AAV2/8 mediated, liver-directed gene therapy. Blood 105(8):3079–3086.
100. Thomas CE, Storm TA, Huang Z, Kay MA. Rapid uncoating of vector genomes is the key to efficient liver transduction with pseudotyped adeno-associated virus vectors. J Virol 2004;78(6):3110–3122.
101. Wang J, Xie J, Lu H, Chen L, Hauck B, Samulski RJ, Xiao W. Existence of transient functional double-stranded DNA intermediates during recombinant AAV transduction. Proc Natl Acad Sci USA 2007;104(32):13104–13109.
102. Cunningham SC, Dane AP, Spinoulas A, Logan GJ, Alexander IE. Gene delivery to the juvenile mouse liver using AAV2/8 vectors. Mol Ther 2008;16(6):1081–1088.
103. Nakai H, Storm TA, Kay MA. Recruitment of single-stranded recombinant adeno-associated virus vector genomes and intermolecular recombination are responsible for stable transduction of liver in vivo. J Virol 2000;74(20):9451–9463.
104. Nakai H, Thomas CE, Storm TA, Fuess S, Powell S, Wright JF, et al. A limited number of transducible hepatocytes restricts a wide-range linear vector dose response in recombinant adeno-associated virus-mediated liver transduction. J Virol 2002;76(22):11343–11349.
105. Vandenberghe LH, Wang L, Somanathan S, Zhi Y, Figueredo J, Calcedo R, et al. Heparin binding directs activation of T cells against adeno-associated virus serotype 2 capsid. Nat Med 2006;12(8):967–971.
106. Cunningham SC, Spinoulas A, Carpenter KH, Wilcken B, Kuchel PW, Alexander IE. AAV2/8-mediated correction of OTC deficiency is robust in adult but not neonatal Spf (ash) mice. Mol Ther 2009;17(8):1340–1346.
107. Franco LM, Sun B, Yang X, Bird A, Zhang H, Schneider A, et al. Evasion of immune responses to introduced human acid alpha-glucosidase by liver-restricted expression in glycogen storage disease type II. Mol Ther 2005;12(5):876–884.

108. Kassim SH, Li H, Vandenberghe LH, Hinderer C, Bell P, Marchadier D, et al. Gene therapy in a humanized mouse model of familial hypercholesterolemia leads to marked regression of atherosclerosis. PLoS One 2010;5(10):e13424.

109. Ziegler RJ, Cherry M, Barbon CM, Li C, Bercury SD, Armentano D, et al. Correction of the biochemical and functional deficits in Fabry mice following AAV8-mediated hepatic expression of alpha-galactosidase A. Mol Ther 2007;15(3):492–500.

110. Summerford C, Samulski RJ. Membrane-associated heparan sulfate proteoglycan is a receptor for adeno-associated virus type 2 virions. J Virol 1998;72(2):1438–1445.

111. Summerford C, Bartlett JS, Samulski RJ. AlphaVbeta5 integrin: a co-receptor for adeno-associated virus type 2 infection. Nat Med 1999;5(1):78–82.

112. Asokan A, Hamra JB, Govindasamy L, Agbandje-McKenna M, Samulski RJ. Adeno-associated virus type 2 contains an integrin alpha5beta1 binding domain essential for viral cell entry. J Virol 2006;80(18):8961–8969.

113. Qing K, Mah C, Hansen J, Zhou S, Dwarki V, Srivastava A. Human fibroblast growth factor receptor 1 is a co-receptor for infection by adeno-associated virus 2. Nat Med 1999;5(1):71–77.

114. Kashiwakura Y, Tamayose K, Iwabuchi K, Hirai Y, Shimada T, Matsumoto K, et al. Hepatocyte growth factor receptor is a coreceptor for adeno-associated virus type 2 infection. J Virol 2005;79(1):609–614.

115. Kaludov N, Handelman B, Chiorini JA. Scalable purification of adeno-associated virus type 2 4, or 5 using ion-exchange chromatography. Hum Gene Ther 2002;13(10):1235–1243.

116. Di Pasquale G, Davidson BL, Stein CS, Martins I, Scudiero D, Monks A, et al. Identification of PDGFR as a receptor for AAV-5 transduction. Nat Med 2003; 9(10):1306–1312.

117. Akache B, Grimm D, Pandey K, Yant SR, Xu H, Kay MA. The 37/67-kilodalton laminin receptor is a receptor for adeno-associated virus serotypes 8 2, 3, and 9. J Virol 2006; 80(19):9831–9836.

118. Duan D, Yue Y, Yan Z, Yang J, Engelhardt JF. Endosomal processing limits gene transfer to polarized airway epithelia by adeno-associated virus. J Clin Invest 2000;105 (11):1573–1587.

119. Zhong L, Li B, Mah CS, Govindasamy L, Agbandje-McKenna M, Cooper M, et al. Next generation of adeno-associated virus 2 vectors: point mutations in tyrosines lead to high-efficiency transduction at lower doses. Proc Natl Acad Sci USA 2008;105(22): 7827–7832.

120. Maheshri N, Koerber JT, Kaspar BK, Schaffer DV. Directed evolution of adeno-associated virus yields enhanced gene delivery vectors. Nat Biotechnol 2006;24 (2):198–204.

121. Li W, Asokan A, Wu Z, Van Dyke T, DiPrimio N, Johnson JS, et al. Engineering and selection of shuffled AAV genomes: a new strategy for producing targeted biological nanoparticles. Mol Ther 2008;16(7):1252–1260.

122. Grimm D, Pandey K, Nakai H, Storm TA, Kay MA. Liver transduction with recombinant adeno-associated virus is primarily restricted by capsid serotype not vector genotype. J Virol 2006;80(1):426–439.

123. Ferrari FK, Samulski T, Shenk T, Samulski RJ. Second-strand synthesis is a rate-limiting step for efficient transduction by recombinant adeno-associated virus vectors. J Virol 1996;70(5):3227–3234.

124. Gao G, Wang Q, Calcedo R, Mays L, Bell P, Wang L, et al. Adeno-associated virus-mediated gene transfer to nonhuman primate liver can elicit destructive transgene-specific T cell responses. Hum Gene Ther 2009;20(9):930–942.
125. Gao GP, Lu Y, Sun X, Johnston J, Calcedo R, Grant R, et al. High-level transgene expression in nonhuman primate liver with novel adeno-associated virus serotypes containing self-complementary genomes. J Virol 2006;80(12):6192–6194.
126. Ziegler RJ, Lonning SM, Armentano D, Li C, Souza DW, Cherry M, et al. AAV2 vector harboring a liver-restricted promoter facilitates sustained expression of therapeutic levels of alpha-galactosidase A and the induction of immune tolerance in Fabry mice. Mol Ther 2004;9(2):231–240.
127. McEachern KA, Nietupski JB, Chuang WL, Armentano D, Johnson J, Hutto E, et al. AAV8-mediated expression of glucocerebrosidase ameliorates the storage pathology in the visceral organs of a mouse model of Gaucher disease. J Gene Med 2006;8(6):719–729.
128. Kozak M. Regulation of translation via mRNA structure in prokaryotes and eukaryotes. Gene 2005;361:13–37.
129. Ill CR, Chiou HC. Gene therapy progress and prospects: recent progress in transgene and RNAi expression cassettes. Gene Ther 2005;12(10):795–802.
130. Marin M. Folding at the rhythm of the rare codon beat. Biotechnol J 2008;3(8):1047–1057.
131. Ward NJ, Buckley SM, Waddington SN, Vandendriessche T, Chuah MK, Nathwani AC, et al. Codon optimisation of human factor VIII cDNAs leads to high level expression. Blood 2011;117(3):798–807.
132. Wu Z, Sun J, Zhang T, Yin C, Yin F, Van Dyke T, et al. Optimization of self-complementary AAV vectors for liver-directed expression results in sustained correction of hemophilia B at low vector dose. Mol Ther 20008;16(2):280–289.
133. Cotugno G, Annunziata P, Tessitore A, O'Malley T, Capalbo A, Faella A, et al. Long-term amelioration of feline mucopolysaccharidosis VI after AAV-mediated liver gene transfer. Mol Ther 2011;19(3):461-469.
134. Wang L, Nichols TC, Read MS, Bellinger DA, Verma IM. Sustained expression of therapeutic level of factor IX in hemophilia B dogs by AAV-mediated gene therapy in liver. Mol Ther 2000;1(2):154–158.
135. Loeb JE, Cordier WS, Harris ME, Weitzman MD, Hope TJ. Enhanced expression of transgenes from adeno-associated virus vectors with the woodchuck hepatitis virus posttranscriptional regulatory element: implications for gene therapy. Hum Gene Ther 1999;10(14):2295–2305.
136. Huang ZM, Yen TS. Role of the hepatitis B virus posttranscriptional regulatory element in export of intronless transcripts. Mol Cell Biol 1995;15(7):3864–3869.
137. Kingsman SM, Mitrophanous K, Olsen JC. Potential oncogene activity of the woodchuck hepatitis post-transcriptional regulatory element (WPRE). Gene Ther 2005; 12(1):3–4.
138. Davidoff AM, Ng CY, Zhou J, Spence Y, Nathwani AC. Sex significantly influences transduction of murine liver by recombinant adeno-associated viral vectors through an androgen-dependent pathway. Blood 2003;102(2):480–488.
139. Mochizuki S, Mizukami H, Ogura T, Kure S, Ichinohe A, Kojima K, et al. Long-term correction of hyperphenylalaninemia by AAV-mediated gene transfer leads to behavioral recovery in phenylketonuria mice. Gene Ther 2004;11(13):1081–1086.

140. Ogawa K, Hirai Y, Ishizaki M, Takahashi H, Hanawa H, Fukunaga Y, et al. Long-term inhibition of glycosphingolipid accumulation in Fabry model mice by a single systemic injection of AAV1 vector in the neonatal period. Mol Genet Metab 2009;96(3):91–96.
141. Sun B, Zhang H, Franco LM, Young SP, Schneider A, Bird A, et al. Efficacy of an adeno-associated virus 8-pseudotyped vector in glycogen storage disease type II. Mol Ther 2005;11(1):57–65.
142. Sun X, Lu Y, Bish LT, Calcedo R, Wilson JM, Gao G. Molecular analysis of vector genome structures after liver transduction by conventional and self-complementary adeno-associated viral serotype vectors in murine and nonhuman primate models. Hum Gene Ther 2010;21(6):750–761.
143. Nathwani AC, Rosales C, McIntosh J, Rastegarlari G, Nathwani D, Raj D, et al. Long-term safety and efficacy following systemic administration of a self-complementary AAV vector encoding human FIX pseudotyped with serotype 5 and 8 capsid proteins. Mol Ther 2011;19(5):876–885.
144. Alexander IE, Cunningham SC, Logan GJ, Christodoulou J. Potential of AAV vectors in the treatment of metabolic disease. Gene Ther 2008;15(11):831–839.
145. Murphy SL, Li H, Mingozzi F, Sabatino DE, Hui DJ, Edmonson SA, et al. Diverse IgG subclass responses to adeno-associated virus infection and vector administration. J Med Virol 2009;81(1):65–74.
146. Chirmule N, Propert K, Magosin S, Qian Y, Qian R, Wilson J. Immune responses to adenovirus and adeno-associated virus in humans. Gene Ther 1999;6(9):1574–1583.
147. Halbert CL, Miller AD, McNamara S, Emerson J, Gibson RL, Ramsey B, et al. Prevalence of neutralizing antibodies against adeno-associated virus (AAV) types 2 5, and 6 in cystic fibrosis and normal populations: implications for gene therapy using AAV vectors. Hum Gene Ther 2006;17(4):440–447.
148. Monteilhet V, Saheb S, Boutin S, Leborgne C, Veron P, Montus MF, et al. A 10 patient case report on the impact of plasmapheresis upon neutralizing factors against adeno-associated virus (AAV) types 1 2, 6, and 8. Mol Ther 2011;19(11):2084–2091.
149. Hurlbut GD, Ziegler RJ, Nietupski JB, Foley JW, Woodworth LA, Meyers E, et al. Preexisting immunity and low expression in primates highlight translational challenges for liver-directed AAV8-mediated gene therapy. Mol Ther 2010;18(11):1983–1994.
150. Zaiss AK, Liu Q, Bowen GP, Wong NC, Bartlett JS, Muruve DA. Differential activation of innate immune responses by adenovirus and adeno-associated virus vectors. J Virol 2002;76(9):4580–4590.
151. Somanathan S, Breous E, Bell P, Wilson JM. AAV vectors avoid inflammatory signals necessary to render transduced hepatocyte targets for destructive T cells. Mol Ther 2010;18(5):977–982.
152. Zhang Y, Chirmule N, Gao G, Wilson J. CD40 ligand-dependent activation of cytotoxic T lymphocytes by adeno-associated virus vectors in vivo: role of immature dendritic cells. J Virol 2000;74(17):8003–8010.
153. Mays LE, Wilson JM. The complex and evolving story of T cell activation to AAV vector-encoded transgene products. Mol Ther 2011;19(1):16–27.
154. Vandenberghe LH, Wilson JM. AAV as an immunogen. Curr Gene Ther 2007;7(5):325–333.
155. Lerut J, Sanchez-Fueyo A. An appraisal of tolerance in liver transplantation. Am J Transplant 2006;6(8):1774–1780.

156. Mays LE, Vandenberghe LH, Xiao R, Bell P, Nam HJ, Agbandje-McKenna M, et al. Adeno-associated virus capsid structure drives CD4-dependent $CD8^+$ T cell response to vector encoded proteins. J Immunol 2009;182(10):6051–6060.
157. Cao O, Furlan-Freguia C, Arruda VR, Herzog RW. Emerging role of regulatory T cells in gene transfer. Curr Gene Ther 2007;7(5):381–390.
158. Wang L, Calcedo R, Wang H, Bell P, Grant R, Vandenberghe LH, et al. The pleiotropic effects of natural AAV infections on liver-directed gene transfer in macaques. Mol Ther 2010;18(1):126–134.
159. Mingozzi F, High KA. Immune responses to AAV in clinical trials. Curr Gene Ther 2007;7(5):316–324.
160. Li C, Hirsch M, Asokan A, Zeithaml B, Ma H, Kafri T, et al. Adeno-associated virus type 2 (AAV2) capsid-specific cytotoxic T lymphocytes eliminate only vector-transduced cells coexpressing the AAV2 capsid in vivo. J Virol 2007;81(14):7540–7547.
161. Li H, Murphy SL, Giles-Davis W, Edmonson S, Xiang Z, Li Y, et al. Pre-existing AAV capsid-specific $CD8^+$ T cells are unable to eliminate AAV-transduced hepatocytes. Mol Ther 2007;15(4):792–800.
162. Siders WM, Shields J, Kaplan J, Lukason M, Woodworth L, Wadsworth S, et al. Cytotoxic T lymphocyte responses to transgene product, not adeno-associated viral capsid protein, limit transgene expression in mice. Hum Gene Ther 2009;20(1):11–20.
163. Pien GC, Basner-Tschakarjan E, Hui DJ, Mentlik AN, Finn JD, Hasbrouck NC, et al. Capsid antigen presentation flags human hepatocytes for destruction after transduction by adeno-associated viral vectors. J Clin Invest 2009;119(6):1688–1695.
164. Herzog RW. Immune responses to AAV capsid: are mice not humans after all? Mol Ther 2007;15(4):649–650.
165. Hauck B, Murphy SL, Smith PH, Qu G, Liu X, Zelenaia O, et al. Undetectable transcription of cap in a clinical AAV vector: implications for preformed capsid in immune responses. Mol Ther 2009;17(1):144–152.
166. Li C, Goudy K, Hirsch M, Asokan A, Fan Y, Alexander J, et al. Cellular immune response to cryptic epitopes during therapeutic gene transfer. Proc Natl Acad Sci USA 2009;106(26):10770–10774.
167. Magami Y, Azuma T, Inokuchi H, Kokuno S, Moriyasu F, Kawai K, et al. Cell proliferation and renewal of normal hepatocytes and bile duct cells in adult mouse liver. Liver 2002;22(5):419–425.
168. Wang Z, Zhu T, Qiao C, Zhou L, Wang B, Zhang J, et al. Adeno-associated virus serotype 8 efficiently delivers genes to muscle and heart. Nat Biotechnol 2005;23(3):321–328.
169. Nakai H, Yant SR, Storm TA, Fuess S, Meuse L, Kay MA. Extrachromosomal recombinant adeno-associated virus vector genomes are primarily responsible for stable liver transduction in vivo. J Virol 2001;75(15):6969–6976.
170. Cotugno G, Annunziata P, Tessitore A, O'Malley T, Capalbo A, Faella A, et al. Long-term amelioration of feline mucopolysaccharidosis VI after AAV-mediated liver gene transfer. Mol Ther 2011;19(3):461–469.
171. Stocker JT, Dehner LP. Pediatric Pathology. Philadelphia: Lippincott, Williams and Wilkins; 2002.
172. Nathwani AC, Gray JT, McIntosh J, Ng CY, Zhou J, Spence Y, et al. Safe and efficient transduction of the liver after peripheral vein infusion of self-complementary AAV

vector results in stable therapeutic expression of human FIX in nonhuman primates. Blood 2007;109(4):1414–1421.

173. Weinstein DA, Correia CE, Conlon T, Specht A, Verstegen J, Onclin-Verstegen K, et al. Adeno-associated virus-mediated correction of a canine model of glycogen storage disease type Ia. Hum Gene Ther 2010;21(7):903–910.
174. Cornetta K. Regulatory issues in human gene therapy. Blood Cells Mol Dis 2003;31 (1):51–56.
175. Kent A, King NM, Cohen-Haguenauer O. Toward a proportionate regulatory framework for gene transfer: a patient group-led initiative. Hum Gene Ther 2011;22(2):126–134.
176. Gonin P, Buchholz CJ, Pallardy M, Mezzina M. Gene therapy bio-safety: scientific and regulatory issues. Gene Ther 2005;12(Suppl 1):S146–S152.
177. King NM, Cohen-Haguenauer O. En route to ethical recommendations for gene transfer clinical trials. Mol Ther 2008;16(3):432–438.
178. Zhang H, Xie J, Xie Q, Wilson JM, Gao G. Adenovirus-adeno-associated virus hybrid for large-scale recombinant adeno-associated virus production. Hum Gene Ther 2009;20 (9):922–929.
179. Clement N, Knop DR, Byrne BJ. Large-scale adeno-associated viral vector production using a herpesvirus-based system enables manufacturing for clinical studies. Hum Gene Ther 2009;20(8):796–806.
180. Virag T, Cecchini S, Kotin RM. Producing recombinant adeno-associated virus in foster cells: overcoming production limitations using a baculovirus-insect cell expression strategy. Hum Gene Ther 2009;20(8):807–817.
181. Ayuso E, Mingozzi F, Montane J, Leon X, Anguela XM, Haurigot V, et al. High AAV vector purity results in serotype- and tissue-independent enhancement of transduction efficiency. Gene Ther 2010;17(4):503–510.
182. Lock M, Alvira M, Vandenberghe LH, Samanta A, Toelen J, Debyser Z, et al. Rapid, simple, and versatile manufacturing of recombinant adeno-associated viral vectors at scale. Hum Gene Ther 2010;21(10):1259–1271.
183. Laurencot CM, Ruppel S. Regulatory aspects for translating gene therapy research into the clinic. Methods Mol Biol 2009;542:397–421.
184. Philippidis A. Developing a balanced business model for gene therapy. Hum Gene Ther 2011;22(6):645–646.
185. Cartier N, Hacein-Bey-Abina S, Bartholomae CC, Veres G, Schmidt M, Kutschera I, et al. Hematopoietic stem cell gene therapy with a lentiviral vector in X-linked adrenoleukodystrophy. Science 2009;326(5954):818–823.

# Nucleic Acid Vaccines

RACHEL BUGLIONE-CORBETT, JOHN SUSCHAK,
SHIXIA WANG, and SHAN LU

## 17.1 INTRODUCTION

In 1796, Edward Jenner made the first demonstration of human vaccination against a pathogen using cowpox to inoculate subjects against smallpox. In the centuries that followed, smallpox has been eradicated in human populations and the principles and effectiveness of vaccination have been firmly established, yet the emergence of new pathogens has constantly challenged scientists to develop new vaccination technologies for improved safety and efficacy.

In many cases, new vaccine technology surges forward as an epidemic looms and an urgent demand develops. In the first half of the 20th century, poliovirus infection spread throughout the world. By early 1950, poliovirus infection in the United States left thousands dead and thousands more disabled. In response to this public health threat, Jonas Salk's inactivated virus vaccine and Albert Sabin's live-attenuated virus vaccine were developed in 1952 and 1958, respectively, drastically controlling the incidence of polio worldwide, further confirming the power of these two types of the first-generation vaccines [1].

With improved public health surveillance and better diagnostic tests, additional emerging viral infections were discovered after World War II, including a number of hepatitis viruses. Taking advantage of the recombinant protein technology, subunit-based hepatitis B virus (HBV) vaccines were developed in the 1980s, which proved to be a successful example of the second generation of vaccines [2–4].

However, the emergence of new pathogens, such as human immunodeficiency virus (HIV), around the same time when an HBV recombinant protein vaccine was licensed indicated that additional new vaccination technology was needed. Subunit protein-based HIV vaccines alone failed to protect in phase III efficacy trials [5,6], and the concern about the safety and efficacy of the first-generation vaccines

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*,
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

(live-attenuated and inactivated vaccines) prevented much effort in using them for HIV vaccine development.

## 17.2 HISTORY OF NUCLEIC ACID VACCINES

### 17.2.1 Early Discovery of DNA Vaccines

In the early 1990s, the demonstration of effective DNA immunization in animal models changed the paradigm of vaccine technology. For the first time, genes encoding antigens, instead of antigens themselves, were demonstrated to elicit a broad spectrum of immune responses against antigens from a variety of pathogens. The efficacy of DNA vaccination was demonstrated first in small animal models and nonhuman primates (NHPs) against pathogens including HIV-1 [7–10], rabies [11,12], HBV [13–15], bovine herpesvirus 1 [16], tuberculosis [17], and influenza [18–21]. Immunogenicity and efficacy have since been tested in human trials, including vaccines against pathogens that include HIV-1 [22–26], HBV [27], and malaria [28–31], as well as in antitumor DNA vaccinations [32,33].

DNA vaccination provides several advantages over other established vaccination strategies. The antigen of interest is produced endogenously and presented to the immune system without danger of toxicity or infection, a common concern with live-attenuated and sometimes with inactivated viral vaccines. *In vivo* production of antigens allows for easy access of antigens to both class I and class II major histocompatibility complex (MHC) molecules. As a result, both $CD4^+$ and $CD8^+$ T-cell responses are generated. Although previous generations of vaccines have been largely geared toward eliciting a specific humoral response, vaccine-induced cell-mediated immunity (CMI), consisting of $CD4^+$ and $CD8^+$ T cells, is critical for protection against a variety of pathogens, such as HIV-1 and malaria. A strong T-helper (Th) cell immune response is also critical for the induction of high-quality antibody responses via the induction of antigen-specific B-cell responses.

A series of early studies by Heeney et al. demonstrated a correlation between balanced, vaccine-specific $T_h1$ and $T_h2$ $CD4^+$ T-cell responses and protection against viral challenge in NHPs immunized with recombinant HIV-1 Env gp120 protein vaccines [34–36]. A strong T-helper cell immune response is also critical for the induction of high quality antibody responses via the induction of antigen-specific B-cell responses. Cytotoxic T cells have been correlated with reduced viremia and viral replication in infected NHPs, suggesting their crucial contribution to viral control and protective immunity [37–39]. Because of the very nature and life cycle of *Plasmodium falciparum* malaria, by which the parasite accesses the bloodstream, hepatocytes, and erythrocytes, an ideal vaccine would induce a broad immune humoral and CMI response to provide protective immunity at multiple levels and at each life cycle stage [40]. For both HIV-1 and malaria, DNA vaccine technology provides a unique method by which to induce diverse immune responses required for protective immunity.

In 1992, Tang et al. made an early contribution to the field by immunizing mice with plasmid DNA encoding the human growth hormone (hGH) gene under transcriptional control of either the cytomegalovirus (CMV) promoter or the human β-actin promoter. To confirm the application of this immunization technique to other protein sources, mice were alternatively immunized with DNA plasmid encoding human α 1-antitrypsin under control of the CMV promoter. They used a ballistic gene gun immunization strategy, whereby gold beads coated with target DNA were propelled directly into the skin using a handheld ballistic device, which is described in more detail later in this chapter [41]. Immunized animals developed an antigen-specific humoral response after just one DNA immunization. The primary response can be further augmented by subsequent DNA immunizations. This initial study made important strides in demonstrating the immunogenic potential of DNA vaccines, leading the way to immunization against additional pathogens in small animals.

Almost simultaneously, Fynan et al. demonstrated successful inoculation of mice and chickens with DNA vaccines expressing hemagglutinin (HA) antigen of influenza virus, which led to protection against flu lethal challenge. The study also explored a variety of DNA immunization methods, demonstrating the superiority of intravenous (IV), intramuscular (IM), and mucosal routes as compared with intradermal (ID), intraperitoneal (IP), and subcutaneous (SC) injections [18].

In another early study report, mice receiving IM plasmid DNA vaccine encoding influenza nucleoprotein (NP), an internal and more conserved viral component, developed NP-specific cytotoxic lymphocyte (CTL) responses and elevated NP-specific antibody titers. Antigen-specific activation of both arms of the immune response suggested successful cross-presentation of antigen to both MHC class I and II molecules. Immunized mice were protected against viral challenge and demonstrated accelerated viral clearance of both homologous and heterologous viral strains [20]. The role of CMI was confirmed by the protection of mice from heterologous viral challenge by the adoptive transfer of immunized splenocytes [42]. Together, these studies demonstrated high levels of immunogenicity for DNA vaccines in small animal models.

The utility of DNA vaccination was also demonstrated against other pathogens. A one-time, low-dose DNA immunization of mice and guinea pigs with herpes simplex virus-2 (HSV-2) glycoproteins B or D produced specific neutralizing antibodies (NAb), and animals were protected against viral challenge [43]. Rabbits inoculated with DNA encoding the major capsid protein L1 of human papillomavirus (HPV) via IM injection produced NAb responses specific to conformational viral epitopes, conferring protection against viral challenge [44]. The success of HPV vaccination in a small animals particularly stands out not only because of potential protection against infection in humans but also because of the implication of protection against genital carcinomas, which are sequelae of HPV infection.

DNA vaccination has been studied extensively in the context of HIV-1. IM immunization in mice with DNA plasmid constructs expressing the gp160 form of HIV-1 envelope (Env) produced strong antigen-specific antibodies with neutralizing capabilities, as compared with mice immunized with recombinant gp160 protein. In

addition, immunized mice demonstrated lymphocyte proliferation and a broad spectrum of IgG isotypes in immune sera. Vaccine-induced antibodies prevented Env binding to $CD4^+$ T cells, supporting the concept that DNA immunization generated immunity against conformational epitopes on the target pathogen [10]. The breadth and quality of humoral responses, as well as cross-presentation to both $CD4^+$ and $CD8^+$ T cells, in mice immunized with Env DNA vaccines was confirmed by later studies [45]. Furthermore, strong neutralizing activity was demonstrated even with low anti-Env titers, and immunization of mice with a transgene encoding HIV-1 capsid protein yielded persistent antigen-specific antibody responses [46].

The immunogenicity of HIV-1 Env DNA vaccines in rhesus macaques was also demonstrated, especially with the use of gene gun or electroporation immunization methods. Although specific NAb and CTL responses were elicited, partial protection was achieved even in a highly difficult to protect challenge model using uncloned SIV251 viral stock. Infection was attenuated and viral load was reduced in immunized animals compared with naïve animals [47]. Chimpanzees immunized with DNA encoding HIV-1 antigens Env, Rev, and Gag/Pol also produced vaccine antigen-specific antibody and CTL responses and were protected partially against heterologous challenge [7].

Although several versions of prophylactic, recombinant protein-based vaccines for HBV have been approved for routine clinical use, the presence of 350 million people chronically infected with HBV worldwide suggests a need to develop additional therapeutic vaccines to supplement existing drug treatment, which is unable to achieve full and long-term control in such a population [48]. Along with influenza and HIV-1, HBV was among the earliest pathogens studied in the context of DNA vaccination and the first model pathogen studied in chimpanzees. A rapid, robust, and long-lasting humoral response, including potent NAb, as well as indications of T-helper responses, was elicited in mice receiving IM DNA plasmids encoding hepatitis B surface antigen (HBsAg) [14,15]. The immunogenicity of HBV DNA vaccination was confirmed in chimpanzees [13]. HBsAg-specific antibody titers elicited by DNA immunization in clinical trials with healthy human volunteers reached levels sufficient for human protection [14].

Mice immunized with IM plasmid DNA encoding rabies virus glycoprotein demonstrated a broad, specific immune response of CTLs, helper T cells, and NAb, which protected animals against viral challenge [12]. NHPs have also been successfully vaccinated with DNA encoding rabies virus glycoprotein and protected against lethal challenge [11].

In the early 1990s, the World Health Organization (WHO) used the term "nucleic acid vaccines" as the official name for DNA vaccination, with the consideration that RNA may also be used as the delivery vehicle for a "gene-based" vaccination approach. Indeed, in recent years, there has been active work using RNA as vaccines. A recent study by Geall et al. [49] demonstrated the efficacy of immunization with self-amplifying RNA encoding replication machinery of alpha viruses along with genes encoding alpha virus protein antigens, or respiratory syncytial virus (RSV) F antigen, as an alternative approach to DNA vaccination or viral vector methods. Mice were injected IM with self-amplifying RNAs alone or in conjunction with a

lipid nanoparticle (LNP) delivery system, resulting in the generation of RSV-specific antibody responses. Similarly, immunized rats were protected against RSV challenge, with protection correlating to the induction of RSV-specific IgG and high neutralizing titers, as well as reduced viral load [49]. Additionally, Petsch et al. [50] have demonstrated the application of mRNA vaccination against influenza A infection. Pigs and ferrets were immunized with an mRNA encoding full-length influenza A/PuertoRico/8/1934 HA. Immunization resulted in humoral and CMI, which was cross-protective against multiple antigens. These vaccine-induced immune responses were correlated to protective immunity [50].

In this chapter, the main focus is on DNA vaccines because significantly more experience has been accumulated with DNA vaccines compared with RNA vaccines.

### 17.2.2 Challenges and New Progress of DNA Vaccination in Humans

Human studies of DNA vaccines were started shortly after successful small animal and NHP studies. However, promising protective immunity observed in small animals and NHPs was not observed in human studies when DNA vaccines were used alone. Experience accumulated in the past two decades of clinical studies demonstrated that DNA vaccines had a high safety and tolerability profile, but the overall immunogenicity of DNA vaccines was quite low [51,52]. Initial studies of DNA immunization with HIV-1 Env and Rev in asymptomatic HIV-1–infected patients indicated only low levels of antigen-specific CTL and antibody responses, and no changes in viral load or lymphocyte counts were observed. More importantly, immune responses were not persistent [25]. In contrast, DNA immunization with HA antigen of influenza H1 subtype was able to elicit protective levels of antibody responses in humans [53–55]. This result confirmed what was observed with HBV S DNA vaccines [56]. Interestingly, both flu HA and HBV S proteins are highly immunogenic antigens, and in both studies, a gene gun was used to deliver HA and HBV DNA vaccines; other human studies mainly used traditional needle injection. Overall, low immunogenicity of DNA vaccines in humans has led to attempts to optimize DNA vaccine constructs, formulations, and delivery approaches.

The major novel immunization regimen to emerge in human studies consisted of a prime-boost strategy. This strategy focuses on improved elicitation of both T-cell and humoral immunity in human models. Two general categories of prime boost have been widely studied. A DNA prime followed by a viral vector boost consists of DNA "priming" inoculations followed by a "boost" with a recombinant viral vector expressing antigens homologous to the DNA prime [57,58]. The goal of this approach is to induce strong T-cell immunity. Several viral vectors have been used in preclinical and clinical trials, including vaccinia virus [59–63], adenovirus [64–69], and fowlpox [70]. A clinical study by McConkey et al. in 2003 [60] demonstrated the utility of a DNA prime encoding a *P. falciparum* malaria antigen followed by a boost with replication-defective modified vaccinia Ankara (MVA). The elicited T-cell responses were substantially higher with the prime-boost regimen compared with either component alone, and partial protection, characterized by delayed parasitemia, was conferred [60].

Another approach uses DNA priming followed by boosting with the same antigen in the form of recombinant protein, with the goal of enhanced antigen-specific humoral responses as well as CMI [58,71–73]. This regimen has been extensively examined in the context of HIV-1 vaccination in both preclinical [74–78] and human studies [73]. In a phase I clinical trial, a multiclade HIV-1 Env DNA prime protein boost vaccine regimen demonstrated robust antigen-specific T-cell and broad NAb responses in humans [73].

It is important to note that although DNA priming with subsequent protein boosting is efficacious, simultaneous immunization with matched DNA and protein has an unexpected effect. Recent work from Bin Wang's group demonstrated that mice receiving plasmid DNA encoding a viral antigen together with recombinant protein antigen had reduced antigen-specific effector T-cell responses, as well as induction of suppressor T cells [79]. This effect occurs only when the DNA and protein are matched in targeting the same antigen; this effect was not observed with antigens from different viruses. Also, the authors emphasized that such a "suppressing effect" was not observed for antibody responses. In fact, this immune suppression is antigen-specific, inducing a subset of specific regulatory T cells that is $CD4^{+}CD25^{-}$ and produces Foxp3, interleukin-10 (IL-10), and interferon-γ (IFN-γ) [80]. Although this strategy should be avoided in protective immunization, the induction of antigen-specific immunosuppression may actually be a useful strategy in suppressing undesirable immune responses, as in cases of autoimmune disease and transplant.

The field of DNA vaccination has made substantial strides in the past 2 years, offering a strategy to combat pathogens against which immunity remains elusive. In addition, this new vaccine technology offers a simpler, less expensive alternative to established vaccines. This progress in DNA vaccine research has lead to a vast expanse of clinical trials for vaccines against influenza, HIV-1, malaria, and HPV, as well as potential bioterrorism threats such as anthrax, severe acute respiratory syndrome, and Ebola. In particular, DNA prime-boost strategies have improved responses in human vaccines [81].

## 17.3 DNA PLASMIDS AS A NEW GROUP OF BIOLOGICAL DRUG PRODUCTS

Vaccines are traditionally not classified as drugs because vaccines prevent the establishment of a disease but do not treat well-established or chronic disease. However, the concept of vaccines has been expanded greatly in recent years because "therapeutic vaccines" are being tested to control or cure illnesses. Similarly, the DNA vaccination approach has been explored to achieve either preventive or therapeutic purposes. In addition, the discovery of DNA plasmids as an effective delivery vehicle also raised possibility of further applications using DNA plasmid to deliver cytokines or other biologically active molecules and the utilization of CpG sequences encoded within a DNA plasmid or oligo as an immunologically active adjuvant in the following section, selected examples are presented for all of these applications.

### 17.3.1 DNA Plasmids as Preventative Vaccines

***17.3.1.1 Influenza*** Two types of licensed seasonal influenza vaccines are currently used to immunize humans against infection, both of which are composed of three strains of influenza predicted by the WHO to be in circulation for a particular year. The first, approved for use in all populations, is an inactivated virus vaccine. The second, restricted in children, elderly adults, pregnant women, and other groups, is a live-attenuated viral vaccine. Although existing vaccination strategies protect vaccinees against the predicted subtypes, certain characteristics of influenza present challenges. Influenza strains undergo antigenic shifts as they circulate, altering viral antigens and leading to escape from vaccine-induced immunity. The ease and speed with which DNA vaccines can be produced suggests that seasonally circulating strains may be targeted with greater speed and accuracy by DNA vaccination. In addition, there was hope that DNA vaccination could generate CMI responses that were unaffected by antigenic reassortment. A phase I clinical trial in 2006 demonstrated the immunogenicity of an influenza DNA vaccine using the HA antigen [54]. Volunteers with preexisting low, but positive, titers against HA received gene gun DNA immunization encoding HA from the H3N2 strain A/Panama/2007/99. Vaccinees receiving the highest dose of DNA, 4 μg, produced significant levels of antigen-specific protective antibody responses [54].

In a recent study from the National Institutes of Health's Vaccine Research Center (VRC), volunteers received one (at day 0) or two (at day 0 and week 4) IM H5 DNA immunizations. An H5N1 monovalent inactivated vaccine was subsequently administered at 4 or 24 weeks after the DNA prime immunization. DNA priming followed by an inactivated vaccine boost 24 weeks later yielded protective HA-specific antibody titers in 81% of vaccinees, a fourfold increase compared with individuals receiving only inactivated vaccine. Individuals receiving a prime boost also generated NAb responses. This study demonstrated the benefit of DNA priming in enhancing the efficacy of an inactivated influenza vaccine [82].

***17.3.1.2 Human Immunodeficiency Virus 1*** Recombinant HIV-1 Env gp120 protein vaccines have been studied with minimal success in humans. A phase III clinical trial in high-risk individuals demonstrated that gp120 protein immunization induced a humoral response but failed to impact rate of infection, viral load, and other parameters [5,6]. DNA priming combined with a boost of either viral vector or recombinant protein has been much more efficacious in several human trials.

The EuroVacc02 phase I trial, an international effort, consisted of two IM DNA primes at weeks 0 and 4 followed by two poxvirus vector NYVAC boosts at weeks 20 and 24 [83]. Both constructs contained two HIV-1 antigen inserts derived from a clade C viral isolate, one insert encoding fused Gag, Pol, and Nef, and the other encoding Env gp120. Broad antigen-specific $CD4^+$ and $CD8^+$ T-cell responses were observed in 90% of vaccinees at weeks 26 and 28 at the conclusion of the immunization regimen. These responses were persistent at 72 weeks. In addition, strong anti-Env IgG antibody responses were present in the majority of individuals receiving the full regimen. T-cell and antibody responses were superior in individuals

receiving a prime-boost regimen compared with viral vector alone. However, no NAbs were detected. In addition to demonstrating the ability to elicit a promising immune response against HIV-1, the EuroVacc02 trial also demonstrated the crucial role of DNA priming in eliciting humoral and CMI. Although the lack of NAb activity indicates that additional work is needed, EuroVacc02 has achieved the strongest immunogenicity for T-cell responses in a DNA prime-viral vector boost strategy up to that point [83,84]. Previous trials using a DNA prime with MVA boost have demonstrated much lower antibody and T-cell responses [85].

An HIV-1 DNA prime-protein boost vaccine strategy using a formulation called DP6-001 was the first such regimen to be tested in a phase I clinical trial. Volunteers received three IM or ID DNA vaccinations at weeks 0, 4, and 12 followed by two protein boosts at weeks 20 and 28 [73,81]. DNA immunization consisted of five Env gp120 plasmids from clades A, B, C, and E, as well as one plasmid encoding clade C Gag. Protein boosts consisted of the recombinant Env antigens expressed by the DNA transgene, adjuvanted with QS-21. Although antibody responses were unimpressive at the end of three DNA vaccinations with low-dose DNA prime, after only one protein boost, 100% of volunteers produced high gp120-specific antibody titers ($1 \times 10^4$–$10^5$). Not only were these antibody titers maintained for more than 6 months, but immunized sera also demonstrated broad neutralizing activity against homologous and heterologous viral isolates from genetically diverse strains. In addition, vaccine-elicited T-cell responses were potent and polyfunctional. The DP6-001 trial is a landmark study in part because of the ability of responders to produce broad NAb against multiple subtypes of the HIV-1 virus in addition to balanced humoral and CMI responses.

***17.3.1.3 Malaria*** As a parasite, malaria represents a unique but important target for vaccination, with an estimated 216 million cases of infection in 2010 [86]. Many of the current vaccine strategies are aimed at generating a T-cell response against liver-stage parasites of the *P. falciparum* subtype. In various studies, DNA plasmid encoding *P. falciparum* circumsporozoite protein (PfCSP) was administered either three times at multiple doses via IM [30], or three times at a high dose of 2500 μg via IM or ID using a Biojector [31]. In general, vaccine-induced CMI responses were very low. Although the vaccine formulations were safe, antibody responses were not detectable in any of the vaccinated individuals [29].

DNA prime-MVA viral vector boost strategies have yielded more success in humans. A DNA prime-MVA boost using thrombospondin-related adhesion protein as well as multiple T-cell and B-cell epitopes (ME-TRAP) from *P. falciparum* was evaluated in clinical trials [87]. Volunteers received IM or gene gun DNA vaccination. An additional group received ID boosting with MVA expressing ME-TRAP [60]. Individuals who received the MVA boost immunization produced significantly higher specific T-cell responses than those receiving DNA vaccination alone. The combination prime-boost regimen conferred partial protection manifested as delayed parasitemia after heterologous sporozoite challenge. A subsequent study found that the DNA ME-TRAP prime and MVA-boost was threefold more immunogenic in African volunteers than in volunteers in the United Kingdom [62]. MVA ME-TRAP

boosting also enhanced the immune response of a prime with recombinant fowlpox vector encoding ME-TRAP [61].

***17.3.1.4 Hepatitis B*** Simple DNA vaccination has been remarkably successful in the HBV model. In a milestone study, a DNA plasmid encoding HBsAg as well as an additional antigen, pre-S2 from the strain *adw*, was administered via gene gun at weeks 0, 8, and 16 to naïve volunteers. By the third vaccination, all immunized volunteers produced high antibody titers and strong T-cell responses characterized by both $CD8^+$ T cells and $CD4^+$ T cells with a $T_h1$ bias. Additionally, increased DNA dosage enhanced antibody titers [56]. A subsequent study of DNA vaccination with HBsAg via gene gun was performed in individuals who demonstrated poor responses to conventional HBV vaccination. DNA vaccination produced positive antibody titers in the majority of vaccinees after only one immunization, demonstrating an advantage of DNA vaccination over conventional strategies [88].

### 17.3.2 DNA Plasmids as Therapeutic Vaccines

***17.3.2.1 Chronic Hepatitis B*** The utility of DNA vaccination has been explored in therapeutic as well as prophylactic contexts. Despite an existing HBV vaccine, more than 370 million people worldwide are chronically infected, likely because of poor T-cell activation. In a phase I clinical trial, 10 chronically infected patients who responded poorly to conventional treatment received three IM DNA vaccinations with a plasmid-encoding viral surface proteins. After the well-tolerated immunization method, sera viral loads decreased in most recipients, and patients displayed increased peripheral blood mononuclear cell (PBMC) proliferative capabilities upon stimulation with HBV antigen. In addition, the vaccine elicited antigen-specific T-cell responses in several patients [89,90]. This study suggests an alternative to conventional therapy for patients with chronic infection refractory to treatment.

***17.3.2.2 Human Papillomavirus*** The US Food and Drug Administration (FDA) recently licensed a prophylactic vaccine against HPV and associated cervical cancers. Gardasil (Merck & Co., Inc., Whitehouse Station, NJ) is a virus-like particle-based quadrivalent vaccine against four strains of HPV that result in 70% of cervical cancers, including HPV-16 and HPV-18 [91]. DNA vaccination presents a promising approach for a therapeutic strategy for HPV infection and associated genital dysplasia. In a phase I clinical study, patients received four IM immunizations with ZYC101, which consists of DNA encoding HPV-16 viral protein E7. The highest dose of DNA produced transient resolution in several individuals; however, atypical cells and dysplasia were detected as early as 15 weeks after vaccination. Specific T-cell responses were increased in a few patients, two of whom demonstrated no evidence of virus in the weeks after vaccination [92]. A subsequent study with ZYC101 in patients with high-grade cervical dysplasia showed that patients receiving the middle dosage of 100 μg demonstrated decreased histologic evidence of disease and undetectable viral load. These patients also produced HPV-specific IgA. Overall, immunized individuals exhibited improved antigen-specific

T-cell responses against HPV [93]. A study of ZYC101a, modified vaccine formulation to include viral proteins E6 and E7 from both HPV-16 and HPV-18, found that in immunized women younger than 25 years of age, dysplasia resolution was significantly higher than in untreated groups [94].

***17.3.2.3 Cancer Vaccines*** Generally, tumor-associated antigens are weakly immunogenic, although spontaneous immunity is generated at low levels, which may lead to a beneficial immune response or more destructive tolerance. Passive immunotherapy has been shown to be successful in the treatment of cancer in studies of antibody treatment for B-cell tumors or subsets of breast cancer [95]. DNA vaccination offers a promising strategy to deliver tumor-specific antigens or other immunomodulatory molecules to the immune system to enhance and more specifically target cancerous lesions.

Clinical trials have been performed with the administration of intratumoral Leuvectin (Vical, San Diego, CA), a well-tolerated DNA–lipid complex that encodes human IL-2. In a therapeutic setting for prostate cancer previously unresponsive to treatment, increased T-cell infiltration and lymphocyte proliferation were observed. Two-thirds of patients demonstrated a transient decrease in serum prostate-specific antigen, a marker of disease [96]. Patients with metastatic renal cell carcinoma receiving intratumoral Leuvectin demonstrated increased IL-2 expression and improved CTL tumor infiltration. A small percentage of patients experienced partial responses, and 23% of patients experienced disease stabilization over a median of 8 months after treatment [97].

Another phase I trial used a DNA plasmid encoding tyrosinase in patients with stage IV melanoma administered by four infusions into lymph nodes. Although expression of the plasmid was rarely detectable and no clinical responses were observed, the patients receiving the highest dosage level preferentially demonstrated prolonged survival [98].

Clinical trials have also focused on intratumoral injection of an allogenic MHC class I gene, HLA-B7, formulated in a cationic lipid vector (Allovectin-7, Vical, San Diego, CA). Immunized patients were negative for HLA-B7. Patients with metastatic melanoma received between one and three injections at single lesions at multiple doses at 2-week intervals. Expression of the plasmid was confirmed in 93% of patients, and 50% of patients demonstrated tumor responses in terms of decreased nodule size by more than 25%. However, a number of adverse effects were observed, including hemorrhage and hypotension, indicating the necessity of further study [99]. Treatment was well tolerated in a subsequent study in which patients with metastatic renal cancer receiving intratumoral Allovectin-7 demonstrated increased levels of tumor-infiltrating lymphocytes, and plasmid expression was detected in more than half of immunized patients [100].

One promising application of therapeutic cancer DNA vaccines is the production of patient-specific formulations. B-cell malignancies characteristically express immunoglobulin Id, a tumor-specific antigen that can be recognized by host antibodies as well as T cells. Given the rapidity with which DNA vaccine formulations can be prepared compared with protein formulations, it is feasible to identify and

clone patient-specific variable regions of Id and prepare a corresponding DNA vaccine [101].

In a phase I/II trial, patients with follicular B-cell lymphoma who were in remission after chemotherapy received DNA plasmids encoding the idiotypic variable immunoglobulin chains from each patient's tumor. Patients received three monthly immunizations at a range of doses. Initial results were weak, with no humoral responses detected and minimal evidence of T-cell stimulation. All patients received a second series of IM or ID immunizations via the Biojector injection device, and 50% demonstrated anti-Id humoral and T-cell responses cross-reactive against other patients' tumors. A third series of injections included DNA plasmid-encoding human granulocyte-macrophage colony-stimulating factor (GM-CSF), but this did not substantially boost existing responses [102]. Despite modest antitumor immune responses, DNA vaccines remain a promising candidate for patient-specific immunotherapy.

### 17.3.3 DNA Plasmids as Adjuvants

Clinical trials have demonstrated the unexpectedly weak immunogenicity of naked DNA plasmids delivered IM in humans compared with animal models. As a result, as DNA vaccine technology and clinical studies have progressed, there has been an extensive effort focused on the optimization of DNA formulations, regimens, and delivery methods. Optimization and delivery methods are discussed later, but the development of molecular and genetic adjuvants encoded along with target antigen in DNA plasmids have become an active field of study.

***17.3.3.1 Genetic Adjuvants*** Plasmids encoding cytokines, chemokines, or co-stimulatory molecules may be co-administered with DNA vaccines to enhance or modulate the vaccine-induced response. A range of inflammatory and helper T-cell cytokines have been studied in small animal models in conjunction with DNA vaccination [103,104]; however, a few in particular stand out for their potential in human vaccination.

The $T_h1$ cytokine, IL-2, activates $CD4^+$ and $CD8^+$ T cells, as well as natural killer cells. Co-administration with DNA vaccines has produced enhanced humoral and CMI responses in mice, although the adjuvant effect in NHPs has been less promising [105]. Interestingly, fusing IL-2 to the Fc portion of IgG prolongs the half-life and augments the cellular immune response to an SIV/HIV DNA vaccine, improving protection after challenge [106].

IL-12 and IL-15 have also been successful in adjuvanting DNA vaccines. Small animal studies have demonstrated the ability of IL-12 to drive a $T_h1$ response and activate T cells in the context of DNA vaccines for influenza [107] and HIV-1 [108,109]. Conversely, prostate cancer models in rhesus macaques imply a suppression of humoral responses [110]. Another $T_h1$ cytokine, IL-15, plays a role in the activation and proliferation of $CD8^+$ T cells [111]. In small animals, co-immunization of IL-15 plasmid augments vaccine-specific CMI with a $T_h1$ bias in models of HSV [112], influenza [111], and HIV-1 [113]. Rhesus macaques co-immunized with

IL-15 plasmid and a DNA vaccine against simian human immunodeficiency virus (SHIV) demonstrated improved induction of effector $CD4^+$ and $CD8^+$ T cells, as well as protection from SHIV challenge and more rapidly controlled infection [114]. In addition, co-administration of IL-15 and its receptor, IL-15Rα, has been shown to further augment immune responses [115,116].

The use of GM-CSF in conjunction with DNA vaccination has been more widely studied but also controversial. In mice, co-administration of DNA plasmid encoding GM-CSF enhanced antigen-specific antibody and helper T-cell responses in models of rabies [117], HSV [118], and HIV-1 [108]. However, results were less impressive in NHPs and clinical studies. In one study, patients were immunized with a mixed formulation of five malarial DNA vaccines and plasmid-encoded GM-CSF via IM injection. T-cell responses to MHC class I– and class II–restricted epitopes were examined, and it was found that the GM-CSF adjuvant resulted in a negative effect on class I–restricted T-cell responses, and in fact, none of the volunteers were protected from challenge [119].

GM-CSF has also been studied in the context of DNA vaccines against HIV-1. The GEO-D03 vaccine consists of non-infectious HIV-1 virus-like particles (VLPs) that express in a single transcript gp160 Env, Gag, protease (PR), reverse transcriptase (RT), Tat, and Vpu. The VLP also encodes for human GM-CSF. In NHPs immunized with a GEO-D03 prime followed by an MVA boost, coexpression of GM-CSF in the DNA prime conferred significant improvement in protection against infection compared with priming without GM-CSF. Additionally, DNA immunization with GM-CSF demonstrated improved quality of NAb responses and antibody-dependent cellular toxicity [120].

The study of these and other cytokines in a clinical setting is ongoing and should shed further light on the ability of genetic adjuvants to augment the immunogenicity of DNA vaccines.

***17.3.3.2 Cytidine Polyguanosine Nucleotides*** Bacterial DNA contains unmethylated regions of cytidine polyguanine nucleotides, or CpG motifs, which are recognized by Toll-like receptor 9 (TLR-9). The inclusion of CpG motifs in a plasmid DNA backbone that also encodes a vaccine-targeted antigen has been investigated for its potential inherent adjuvant effect. As a TLR-9 agonist, CpG has a role in inducing proinflammatory cytokines, including IL-12 and type I INFs, leading to activation and priming of CTLs [121]. In mice, antigen-specific immunity was correlated with unmethylated CpG motifs, and methylation of these motifs reduced immunogenicity [122]. TLR-9 activation contributed to, but was not required for, immune response to DNA [123]. Furthermore, in mice, CpG motifs contributed to immunogenicity of DNA priming but not to subsequent boosting [121,124].

Although NHPs have demonstrated successful immunity to HBV after vaccination with CpG oligonucleotides in conjunction with recombinant protein [125], an HBsAg DNA vaccine optimized for CpG content demonstrated poor responses in chimpanzees [126]. A CpG oligonucleotide product, CpG7909 (Coley Pharmaceuticals, Wellesley, MA) was licensed to Pfizer as PF3512676 in 2005, primarily

for use as monotherapy or in combination with chemotherapy drugs in cancer treatment. C7909 was also licensed for study to GlaxoSmithKline in 2002 as VaxImmune and to Chiron Corporation (Novartis, Basel, Switzerland) in 2006. Extensive clinical trials have been undertaken to study the utility of CpG alone or in combination with other agents, most notably in cancer treatment. The utility of optimized CpG motifs in DNA vaccine constructs remains to be fully elucidated in humans.

### 17.3.4 DNA Plasmids as Veterinary Vaccines

Since 2005, four DNA products have been licensed for prophylactic and therapeutic purposes in veterinary medicine. Licensure for veterinary use is an enormous step for DNA technology, further confirming the quality of these strategies in large animals and implicating the potential of these vaccines in human disease. In 2005, the first DNA vaccine for West Nile virus (WNV) was licensed for use in horses. Extensive preclinical work confirmed effective protection and strong NAb titers in mice and horses [127], as well as birds [128], using a single IM injection of recombinant plasmid encoding the premembrane and envelope proteins of WNV. A phase I clinical trial of this vaccine in human volunteers was well tolerated and induced specific T-cell and antibody responses [129].

Almost immediately after the equine vaccine licensure, a DNA vaccine was licensed for farmed Atlantic salmon against infectious hematopoietic necrosis virus (IHNV). The plasmid DNA encoded IHNV glycoprotein produced NAb activity and protected against viral challenge [130]. In immunized fish, both antibody and T-cell responses were stimulated along with the induction of type I IFN [131–133]. This vaccine has particular importance in fish farming, both in terms of reducing animal infection and increasing the quantity and quality of farmed fish.

A therapeutic DNA vaccine for malignant melanoma in dogs was given conditional approval in 2007 and was licensed for veterinary medicine in 2010. The DNA plasmid encodes human tyrosinase, a marker of melanoma. The use of a human antigen allows for the production of antibody responses against dog tyrosinase while avoiding the difficulty of tolerance to a self-antigen. Vaccinated dogs produced a strong humoral response against tumor antigen [134]. In addition, dogs with advanced stage metastatic melanoma demonstrated a clinical response in metastases and improved long-term survival with regression of melanoma [135]. This therapeutic use of DNA constructs, effective in even an advanced stage of disease, suggests a clear path toward the treatment of human disease.

The fourth DNA product, licensed in 2007 for use in swine, is directed towards gene therapy rather than vaccination. A plasmid encoding growth hormone–releasing hormone yields an increase in growth hormone. Pregnant swine received IM injections followed by electroporation, with the goal of increasing the quantity and quality of live piglets in a litter and reducing fetal mortality [136,137]. Although not a model of vaccination applicable to human disease, this is an important demonstration of the efficacy of electroporation and DNA-based gene therapy in a large animal.

## 17.4 OPTIMIZATION OF DNA VACCINES

As discussed, the key issue facing the use of DNA vaccines is their immunogenicity in larger animals and humans. Many gene transfer methods work well *in vitro*, but their *in vivo* efficiency is impaired. DNA vaccination is highly immunogenic in mice and smaller mammals but, to date, has only resulted in low-level immune responses in NHPs and humans [138–140]. The lack of strong cellular and humoral responses has hindered the development of DNA vaccines. Fortunately, DNA vaccine plasmids are easily manipulated, providing many possibilities for increasing antigen expression and immunogenicity. Several methods exist for optimizing plasmid transcriptional elements and protein production. The goal of this section is to discuss the factors that affect DNA vaccine efficacy and how they may be altered to improve the immune response.

### 17.4.1 Optimization of Transcription

The use of a plasmid backbone offers many advantages with regard to optimization. Much work has been done to identify methods that increase both gene transcription and protein expression. Of particular interest is the gene promoter. Initial studies established that microbial promoters are poor inducers of mammalian gene expression. Early research focused on oncogenic, virally derived promoters. The first promoter to be investigated was the simian virus 40 (SV40) promoter [141,142]. The SV40 promoter was used until researchers identified the Rous sarcoma virus promoter as having the capacity to drive higher transgene expression rates [141]. These promoters were effective, but some concern existed because of their link to cancer.

Later, noncarcinogenic sources garnered interest. The most common choice is the human CMV immediate early promoter because it yields the highest level of transgene expression. A comparison of the CMV immediate early promoter to a plasmid containing the endogenous AKV murine leukemia long terminal repeat (LTR) demonstrated that the CMV plasmid elicited higher antigen-specific responses [143]. The CMV promoter also offers two additional advantages. It provides for constitutive expression of the transgene within a wide range of mammalian cells, and it does not suffer from downstream read-through, preventing transcription of the plasmid backbone. In general, virus-derived promoters provide for greater transgene expression *in vivo* than do eukaryotic promoters.

In addition to the use of a strong viral promoter, several other transcriptional elements may be inserted into the plasmid open reading frame to enhance promoter activity. Similar to the previously mentioned promoters, many transactivator genes are derived from viral enhancer regions. The inclusion of transcriptional enhancers and transactivators, such as the CMV intron A and LTRs, have been shown to boost the antigen-specific response above levels seen with the CMV promoter alone [144,145]. It is believed that the inclusion of Intron A enhances the rate of polyadenylation and nuclear transport of mRNA [146]. Inclusion of strong promoters or enhancers increases the rate of transcriptional initiation [146,147].

Another factor affecting plasmid expression is the efficiency of primary RNA transcript processing and polyadenylation. Polyadenylation sequences have varying efficiency with regard to primary RNA transcript processing, significantly affecting transgene expression. For example, the SV40 polyadenylation sequence is less efficient than bovine growth hormone polyadenylation sequences when expressed in mouse liver. This may be rectified by the addition of a second SV40 enhancer downstream of the original SV40 polyadenylation signal [148].

### 17.4.2 Enhancing Protein Production

*In vivo* protein expression is required for DNA vaccine efficacy. Consequently, improved expression of the transgene antigen is expected to boost the immune response. Codon optimization is a highly effective strategy for increasing protein production [149–151]. Numerous organisms have biases in their codon usage. This discrepancy exists because transfer RNAs (tRNAs) are not at equal levels within all cells. Consequently, certain codons are negatively selected in particular cell lines. The use of an antigenic codon sequence that uses more abundant tRNAs increases translation efficiency, resulting in increased antigen expression and a higher antigenic dose. Increased antigenic protein expression leads to enhanced CTL and humoral responses.

Another common method of improving transgene expression is the optimization of the start site for protein synthesis. The sequences surrounding the mRNA start codon impact its recognition by eukaryotic ribosomes. Many expression vectors carry modified Kozak consensus sequences ($^{-6}$GCCA/GCCAUGG$^{4+}$) that allow for optimal expression in mammalian cells [152]. Not all eukaryotic genes possess Kozak sequences; therefore, inclusion in the plasmid may augment expression levels. There is a major caveat to increased protein synthesis because it is necessary to ensure translation terminates, preventing read through of the bacterial genome. As such, many plasmid constructs encode extended poly A tails and double stop-codons. These modifications help stabilize mRNA and function to disrupt ribosome–mRNA interactions.

Finally, the inclusion of a leader sequence can add another dimension to plasmid optimization. Leader sequences can enhance the stability of mRNA and increase the efficiency of translation. Many native leader sequences found on antigenic proteins hinder cleavage and lower the rate of folding inside the endoplasmic reticulum [148,153]. To counteract this problem, the human tissue plasminogen activator (tPA) has become a standard inclusion in plasmid design. Human tPA inclusion results in a marked increase in protein expression and immunogenicity [148,154].

### 17.4.3 Delivery of DNA Vaccines

As discussed, the goal of DNA immunization is to transfect host cells, allowing for both MHC class I and II antigen presentation. Therefore, it is not surprising that the method and route of immunization have significant effects on DNA vaccine efficacy.

Although a wide range of delivery approaches are currently being investigated for DNA vaccines, they can be divided into two groups. The "chemical delivery" approaches share the common features of a liquid formulation, but the facilitating reagents vary. In contrast, the "physical delivery" approaches take advantage of a particular physical force (such as air pressure or electrical pulse) to deliver the DNA plasmids into targeted tissues or cells. This section discusses the benefits and drawbacks of each delivery route and addresses possible methods for increasing their effectiveness.

***17.4.3.1 Chemical Delivery Methods*** Several chemical formulations have been used to boost DNA vaccine efficacy. The most basic and commonly used approach simply involves IM injection of vaccine plasmid diluted in phosphate buffered saline (PBS) or saline. The benefit of this approach is that it does not require additional manipulations before vaccine administration. Unfortunately, because plasmid-based vaccines rely on endocytosis and pinocytosis for transfection, much of the injected DNA is wasted. Exogenous DNA either diffuses away or is degraded in the extracellular space. This has resulted in the need to significantly increase the amount of plasmid DNA required for cell transfection. Human trials have required doses of plasmid DNA in the range of 500 to 2500 μg [30,73].

To this end, several laboratories have developed chemical methods that not only enhance transfection efficiency but also have an adjuvant effect. The use of liposomes as carrier molecules has recently become a popular delivery method. One formulation that has shown a clear benefit in human trials is Vaxfectin (Vical, San Diego, CA). Vaxfectin is a proprietary formulation of cationic lipid (GAP-DMORIE) and a neutral phospholipid (DPyPE) that self-assembles into liposomes when mixed. IM injection of a Vaxfectin–influenza NP mix increased antibody titers 20-fold compared with vaccine alone [155,156]. Boosting of antibody titers did not result in a diminished cytotoxic T-cell response. Likewise, inclusion of Vaxfectin in a *P. falciparum* vaccine enhanced the IFN-γ production [157]. An ensuing human trial involving DNA plasmids formulated against the influenza H5 HA, NP, and M2 genes reported response rates and antibody titers in the range of inactivated protein-based H5 vaccines [158].

The other development in DNA vaccine formulation is the use of biodegradable polymeric microparticles as a carrier system, thereby protecting the vaccine plasmid from extracellular deoxyribonucleases. Microspheres, consisting of amphiphilic molecules, have proven especially efficient at delivering DNA to cells. In addition to protecting plasmid DNA from nucleases, microspheres promote the sustained release of vaccine without repeat dosing. High-molecular-weight cationic polymers have proven even more effective than cationic liposomes in aggregating DNA. Plasmid DNA immobilized within the biodegradable polymeric microspheres (chitosan-coated alginate microspheres with diameters ranging from 20 to 500 μm) can induce both systemic and mucosal immune responses [159]. Microspheres may be delivered orally or intraperitoneally, where they can transfect dendritic cells, yielding higher levels of dendritic cell activation. The benefits of such formulations have been shown in mice, NHPs, and humans [160–164]. These results suggest that

microparticle-based delivery systems are capable of enhancing DNA vaccine immunogenicity and boosting cellular and humoral immunity.

***17.4.3.2 Physical Delivery Methods*** Different from the chemical delivery approaches, the physical delivery approaches use different kinds of external physical means to "force" the delivery of DNA vaccine plasmids.

As previously mentioned, the most common route of DNA immunization is IM injection of DNA diluted in saline, which has the benefit of requiring no special delivery system and can be delivered without specific tools or instruments. A similar method is ID injection, in which DNA plasmid is delivered between the layers of skin. Both IM and ID injection have been demonstrated to be $T_h1$ biased, favoring a strong CTL-based response. Notably, it has been demonstrated that the direct transfection of antigen-presenting cells (APCs) is key to successful immunization [165,166]. This has led to the use of transcutaneous microneedle injectors that bypass the upper stratum corneum skin layer and deliver the vaccine to an area rich in APCs, particularly Langerhans cells. Likewise, subsequent boosting immunizations may be delivered to increase cellular and humoral responses. At the same time, more sophisticated instruments have been developed to improve the transfection efficiency of DNA immunization.

Electroporation has long been used *in vitro* to increase transfection efficiency. The application of an electrical pulse to cells temporarily opens pores in the lipid bilayer, allowing DNA to traverse the cellular membrane down the concentration gradient. Subsequent analysis proved that electroporation also functions *in vivo* with minimal cell damage. Naked plasmid is injected into the tissue, and an electric pulse is applied via needle or caliper-type electrodes. This has the added benefit of recruiting inflammatory cells to the site of injection. Electroporation has been shown to increase gene expression up to 1000-fold compared with animals receiving only IM injection. Additionally, electroporation may be applied to many tissue and cell types [167].

Another technique involves propelling DNA plasmid-coated gold beads into the skin using a gene gun. The Helios gene gun (Bio-Rad Laboratories, Irvine, CA) uses high-pressure, helium gas-powered particle bombardment to directly transfect target cells or tissues. This method was initially shown to be capable of generating immune responses in reporter mice [41]. Human clinical trials involving hepatitis B transgenes have established that gene gun immunization is capable of inducing protective immunity in humans [56,88]. Several differences exist between IM and gene gun immunization. Reports have shown that significantly less plasmid DNA is required for generating protective immunity by gene gun immunization in mice [18,73]. The direct transfection of target cells also favors a $T_h2$ immune response, as opposed to the $T_h1$ bias of IM injection. It is a potentially important issue that only a small amount of DNA can be coated onto the gold beads. However, at this point, the main reason to prevent wide use of this technology is the limited access of this technology because of intellectual property rights ownership.

Other needle-free injection systems have been investigated. So-called pneumatic or jet injectors have shown promise in both animal and human clinical trials

[28,168–172]. Jet injectors function by injecting a high-pressure, narrow stream of injection liquid into the epidermis of patients. Similar to the gene gun, it is powered by compressed air or gas. These devices are capable of delivering medication via the IM, ID, or SC routes. Several studies have demonstrated enhanced immune responses compared with needle and syringe delivery. The major concern with previous needle-free injection systems was the possibility of disease transmission between patients. Because the jet injector breaks the skin barrier, there is a small, but measurable, chance that biological material may be transferred between users [173]. Current model designs have rectified this by using single-use syringes, thereby eliminating the potential for cross-contamination. Improvements in functionality and safety of jet injectors will make them a strong candidate for use in mass vaccination.

## 17.5 UNPROVEN SAFETY CONCERNS

Traditional vaccines rely on the injection of live-attenuated virus, inactivated viral particles, or recombinant proteins to induce immunity. Inactivated particles and recombinant proteins tend to yield a decreased CTL-based response. Live-attenuated vaccines carry some risk of infection. For example, live-attenuated vaccines have a small possibility of reverting to a pathogenic form and may cause disease in immunocompromised patients. Conversely, DNA vaccines have little opportunity to cause disease because they are often constructed with several safety features. Because DNA vaccine plasmids are nonlive, nonreplicating constructs, there is little risk of pathogenesis. In addition, fears of adventitial agents being transferred from tissue culture lines are bypassed. Likewise, they do not require the use of toxic treatments to inactivate infectious particles, as do killed vaccines. These factors have established DNA vaccines as intrinsically safer than conventional vaccination methods. This section discusses possible safety concerns and describes the fact that no safety concern has been experimentally observed.

### 17.5.1 Integration into Host Genome

The possibility for integration of plasmid DNA into the recipient's genome has been a major focus in early days of DNA vaccine discovery. Integration of plasmid genes could occur either randomly or through homologous recombination, leading to activation of oncogenes, the inactivation of tumor suppressor genes, or mutational shifts in chromosomal DNA. To combat this possibility, several laboratories use suicide vectors that induce cellular apoptosis in transfected cells. This ensures that any integration events are not carried through vertical transmission.

Initial integration studies were undertaken with plasmids encoding either influenza or HIV genes [174,175]. Only extrachromosomal plasmid DNA was detectable after IM injection, suggesting that random integration occurred at extremely low frequencies, fewer than one to eight copies in 150,000 nuclei. Further studies involving plasmid DNA expressing *P. falciparum* circumsporozoite protein (*Pf*CSP) demonstrated similarly low levels of integration [176]. Several groups have

investigated the role of route of immunization in plasmid integration. Upon IM injection, early time points established a low level of systemic plasmid DNA that dissipated quickly. Plasmid was occasionally detected in the reproductive organs (ovaries and testes); however, all DNA was extrachromosomal, leaving little risk for germline transmission [78,175]. Because such frequencies are three orders of magnitude lower than the spontaneous mutation rate of gene-inactivating mutations, many researchers believe that the risk of mutation associated with plasmid integration is negligible.

In contradiction to these studies, one report by Wang et al. was able to identify four independent integration events after electroporation of mice immunized with a plasmid encoding the mouse erythropoietin gene [177]. Wang et al. used repeat-anchored integration capture (RAIC) polymerase chain reaction to detect covalent junctions of plasmid–genomic DNA sequences. Their study demonstrated that electroporation markedly increased plasmid tissue levels. The integration sites appeared to be random, with only short sequences of homology. These studies identify integration events as a possible area of concern despite their low level of frequency. None of these issues have been reproduced by others in preclinical or clinical trials; however, current techniques are not refined enough to conclusively identify the number of integrations associated with DNA vaccination. Further monitoring in clinical studies may help to clarify any negative effects that plasmid integration may generate.

### 17.5.2 Autoimmune Responses

The second major concern raised regarding DNA vaccination is the induction of autoimmunity. The native immunostimulatory nature of DNA plasmids has led to some concern that anti-DNA immune responses may develop. In particular, the ability of unmethylated CpG to activate innate immunity pathways has led to such questions [178]. Investigation of the ability of CpG motifs to induce autoimmunity in mice demonstrated that the production of anti-DNA antibodies was not significantly increased above spontaneous rates [179,180]. Similarly, many plasmids have other immunomodulatory molecules inserted to boost the immune response. The inclusion of cytokine-encoding transgene sequences may result in systemic activation of APCs and T cells. In addition, there is some concern about the inclusion of tPA leader sequences. Some researchers have expressed trepidation that the attachment of a human leader sequence may direct the formation of antibodies against recipient DNA.

It has been demonstrated in rabbits that the inclusion of certain transgene sequences may increase the risk of anti-DNA humoral responses. Low-level, transient antibody responses were seen in animals immunized with a vaccine encoding the HIV-1 RT and HIV-1 nef genes. The authors suggest that these responses were the result of protein–nucleic acid complexes, but there is no evidence to suggest that such responses are the same as in real clinical autoimmune diseases because no positive laboratory biomarkers have been reported [181]. A clinical study performed in humans demonstrated the possibility of mild reactogenicity in the DNA

prime-protein boost vaccination model when a potent adjuvant is used [182]. Superficial skin tissue–based vasculitis developed in two vaccinees receiving three HIV-1 Env DNA primes followed by one Env protein boost adjuvanted by QS-21. The rapid increase in antibody responses after the first protein boost suggests that DNA priming resulted in a strong humoral response, possibly because of increased cytokine production induced by the adjuvant.

To date, there is no convincing evidence that DNA vaccination induces systemic autoimmune disease. Preclinical studies in NHPs and early human trials have not produced any detectable increases in anti-DNA or antinuclear antibodies. Some evidence of reactogenicity in DNA prime-protein boost models has been seen, but this may be attributable to the inclusion of a strong adjuvant or high immunogenicity when DNA is used to prime the immune responses. In summary, there is no solid evidence that DNA vaccines may direct the formation of anti-DNA antibodies at reproducible or detectable levels.

## 17.6 MANUFACTURING PROCESS

DNA vaccination offers a new and credible way to treat humans and animals. As DNA vaccines become more widespread, new methods for optimizing the manufacturing process are in development. Two major issues facing vaccine manufacturers are the quantity and quality of DNA preparations. Quantity issues focus on the ability to produce the kilogram quantities of plasmid DNA required for mass immunization. The old methods of small-scale production are no longer viable because of the increasing demand for effective vaccines. Equally important is the need to increase product purity to meet the specifications of regulatory agencies. Before developing a manufacturing protocol, it is recommended that a critical evaluation of the vaccine plasmid and production be undertaken.

### 17.6.1 Composition of Vaccine Elements

Although plasmid DNA is generally produced in *Escherichia coli*, several *in vitro* production methods have also been investigated. *In vitro* DNA production can generate high plasmid yield; however, these methods require costly and specialized equipment and reagents. Therefore, the majority of production facilities still use large-scale bioreactors. The generation and purification of plasmid DNA presents unique challenges because of its similarity to host cell nucleic acids. Several factors in the plasmid design affect DNA yield. This section reviews these factors and addresses possible courses of action.

Plasmid DNA conformation greatly affects product yield. Proper plasmid conformation is crucial for efficient production. Optimal production requires high levels of supercoiled plasmid because it is believed that supercoiled DNA is therapeutically more effective than are the open-circle, linear, and nicked forms of DNA. Unfortunately, the undesirable forms are difficult to separate from supercoiled

product. In addition, the amount of supercoiled DNA can be decreased by the inclusion of particular gene sequences. Such sequences result in a large percentage of the total plasmid remaining in the open, circular form. Therefore, it is necessary to optimize the plasmid conformation before beginning production.

Plasmid stability is another important factor in manufacturing. Certain sequences decrease plasmid stability. For instance, palindromic sequences and inverted repeats are extremely unstable, but adenosine-thymidine rich regions render the plasmid susceptible to endonuclease nicking. Alternating regions of pyrimidine and purine sequences can be unstable in large batch fermentations [183]. Oligopyrimidine and oligopurine sequences can increase dimer formation [184]. Of particular importance are long regions of CpG repeat sequences [185]. Because these regions are beneficial in generating an innate immune response, a comprehensive investigation of sequence stability is required. Optimal plasmid DNA production may require a reduction in immunostimulatory ligands to reach full production capacity.

Component orientation also affects plasmid function. Many prokaryotic gene sequences negatively affect eukaryotic gene expression [186]. Other prokaryotic genes can bind eukaryotic transcription factors, further inhibiting gene expression [187]. Inclusion of selective marker genes can also limit production yields. It has been demonstrated that CMV promoter expression can be dramatically lowered if inserted in an improper orientation [188]. These factors must be accounted for during the plasmid design phase to ensure that high production levels are met.

### 17.6.2 Fermentation

Two goals are key in designing a fermentation process: high volumetric yield and high concentrations of supercoiled plasmid. Optimal fermentation plans have a small, economical volume that does not waste media or cells. Three strategies are favored in the mass production of DNA plasmids: batch fermentation, fed-batch fermentation, and continuous fermentation. Each process will be examined in turn.

***17.6.2.1 Batch Fermentation*** The simplest method of culturing requires the addition of all nutrients necessary for cell growth at the time of inoculation. Because all nutrients are present immediately after inoculation, the growth rate will remain essentially static. However, as it is sometimes necessary to reduce the growth rate to obtain optimal plasmid production, the growing temperature may be lowered depending on the type of media used [189].

***17.6.2.2 Fed-Batch Fermentation*** The most efficient—and therefore common—method of fermentation involves adjusting nutrient concentrations during cultivation. Limiting amounts of nutrients are added throughout cultivation so as to control growth rate. This method has the added benefit of yielding higher biomass because nutrients are almost completely consumed, precluding the possibility of reaching inhibitory substrate concentrations. Several established fed-batch fermentation protocols are in use today.

*Merck Process* The Merck Process (WO2005078115) involves selecting for highly productive clonal subtypes of the transformed *E. coli* [190]. Transformed *E. coli* are cultivated and selected for on blood agar and chemically defined agar plates.

*Boehringer Ingelheim Process* The Boehringer Ingelheim Process (WO2005097990) uses exponential feeding for the fed-batch phase to sustain biomass growth [191]. Then feeding is switched to a linear constant program to lower specific growth rates and increase plasmid accumulation. Use of the JM108 strain of *E. coli* (WO2005098002) results in high plasmid yields and increased supercoiled plasmid concentrations [192].

*Nature Technology Corporation Process* The Nature Technology Corporation Process (WO2006023546) involves reduced temperatures during the fed-batch phase [193]. This serves to limit the growth rate and maintain plasmid stability while the biomass grows efficiently. The temperature is then increased, and additional nutrients are added. Plasmid yields increase considerably after the temperature shift. Surprisingly, cells are able to tolerate significantly higher quantities of plasmid after the temperature shift.

***17.6.2.3 Continuous Fermentation*** The third cultivation method involves harvesting cells and adding nutrients concurrently. As sterile nutrients are added to the bioreactor, an equivalent amount of converted nutrient solution is removed. This results in some of the microorganisms also being removed. Under steady state conditions, cell loss caused by overflow must be balanced by bacterial outgrowth.

### 17.6.3 Purification

Large-scale production of plasmid DNA presents many challenges that must be considered before production begins. Several issues also exist in the downstream processing of plasmid DNA. The goal of the purification process is to remove all impurities from the cell lysates, leaving only highly purified plasmid DNA. Purification methods are difficult and expensive but are necessary for producing a safe, clinical-grade product. The majority of manufacturing costs occur during the purification phase, primarily because of the difficulty of separating plasmid DNA from host cell DNA and the relatively small amount of plasmid DNA produced from a fermentation batch. The type of plasmid used in fermentation is a key aspect in designing a purification process.

Before beginning any purification, it is critical that any purity requirements be identified. Commercial plasmid products generally have lower quality needs than do vaccine-related products. The chief impurities in any plasmid DNA preparation are chromosomal DNA, protein, RNA, and endotoxins that may be present in *E. coli*. These impurities may be removed through several methods, including nuclease treatment, chromatography, selective precipitation, and tangential flow filtration. It is often desirable to use more than one purification method at each step. For example, both triple helix chromatography and hydrophobic interaction chromatography may be used to remove genomic DNA. Unfortunately, this process is not ideal because many

chromatography columns have poor binding capacity for plasmid DNA, lowering the overall yield. Conversely, high-capacity ion exchange columns have poor binding capacity for genomic DNA. Therefore, it is best to limit the number of purification steps to those absolutely required to meet preestablished purity requirements.

### 17.6.4 Bulk Plasmid Product Release Testing

Testing of bulk plasmid preparations should include plasmid concentration and conformation species identification. As discussed previously, it is recommended that all preparations contain high levels of supercoiled DNA to increase immunogenicity. Similarly, an examination of bulk plasmid preparations should include testing for bacterial host cell contaminants, including RNA, DNA, and protein (<1%). These scales are not strict rules because it is anticipated that contaminant levels will diminish as the development program moves forward. The FDA also recommends Limulus Amebocyte Lysate testing to quantify the presence of bacterial endotoxins within the preparation. Endotoxin levels should not exceed 40 EU/mg of plasmid. In general, the FDA suggests that producers of DNA vaccines establish strict acceptance criteria and that each vaccine lot be tested for compliance.

## 17.7 REGULATORY CONSIDERATIONS

This section is prepared only as a general introduction to the clinical testing of DNA vaccines. Regulatory agencies or experts with experience in updated regulatory requirements should be consulted for such studies.

The US FDA defines DNA vaccines as purified plasmid preparations containing one or more DNA sequences capable of inducing or promoting an immune response against a pathogen [194]. DNA vaccines are considered biological products as described in Section 351 of the Public Health Service Act (PHS ACT) (42 U.S.C.262) and are regulated by the Center for Biologics Evaluation and Research (CBER). Many of the following issues are included in the points of consideration.

The FDA recommends that a summary of all components used during manufacturing be included in the final product application. This includes plasmid construction, intermediate recombinant cloning procedures, and plasmid sequencing. It is also suggested that the genotype of source bacterial cells be described. In particular, analysis of the master cell bank and working cell banks should be included to ensure that products are free of contaminants. Overall, the manufacturing process should be described sufficiently so as to allow for a detailed and clear understanding of the product's safety.

### 17.7.1 Preclinical Immunogenicity and Safety

The FDA requires that all new vaccines be subjected to preclinical safety testing (21 CFR 312.23(a)(8)). All good laboratory practice (GLP) regulations (21 CFR Part 58)

should be applied to production batches intended for clinical use. GLP regulations also apply to any nonclinical studies that are designed to support investigational new drugs (INDs) and biological license applications. Also, preclinical safety studies should be applied to every DNA vaccine–adjuvant formulation under consideration. It is strongly recommended that CBER be consulted before submission of any INDs to ensure that proper preclinical safety studies were completed.

Evaluations of immunogenicity and reactogenicity should be designed to assess any systemic toxicity. Such studies should be done in a relevant animal model to ensure that antigen-specific humoral and CMI responses are within acceptable levels so as to prevent autoimmunity. Preclinical immunogenicity studies are not subject to GLP guidelines unless they are also being used as safety studies. When a persistent immune response is generated against endogenous or self-antigens, it is recommended that any potential adverse effects be studied in a relevant animal model. Such responses should also be monitored in any future clinical trials.

## 17.8 SUMMARY

The use of safe and effective vaccines has become a fundamental aspect of global health. To this end, immunization with a nucleic acid–based vaccine has led to new considerations in vaccination strategies. At least in preclinical studies, antigen-encoding DNA plasmids are capable of eliciting a broad spectrum of immune responses against antigens from a variety of pathogens.

Although DNA vaccination provides several advantages over other established vaccination strategies, further optimization is necessary before it becomes the predominant strategy in human patients. DNA vaccination has shown great promise in prime-boost strategies and may provide a solution for pathogens that were previously thought unpreventable or untreatable. Their adaptability and ease of manipulation may prove to be the aspect that pushes them to the front line of vaccination. It appears that DNA vaccines are on the precipice of redefining the field of vaccinology.

## REFERENCES

1. Norrby E, Prusiner SB. Polio and Nobel prizes: looking back 50 years. Ann Neurol 2007;61(5):385–395.
2. Brown SE, Stanley C, Howard CR, Zuckerman AJ, Steward MW. Antibody responses to recombinant and plasma derived hepatitis B vaccines. Br Med J (Clin Res Ed) 1986;292 (6514):159–161.
3. Emini EA, Ellis RW, Miller WJ, McAleer WJ, Scolnick EM, Gerety RJ. Production and immunological analysis of recombinant hepatitis B vaccine. J Infect 1986;13(Suppl A):3–9.
4. McAleer WJ, Buynak EB, Maigetter RZ, Wampler DE, Miller WJ, Hilleman MR. Human hepatitis B vaccine from recombinant yeast. Nature 1984;307(5947):178–180.

5. Flynn NM, Forthal DN, Harro CD, Judson FN, Mayer KH, Para MF; rgp120 HIV Vaccine Study Group. Placebo-controlled phase 3 trial of a recombinant glycoprotein 120 vaccine to prevent HIV-1 infection. J Infect Dis 2005;191(5):654–665.
6. Gilbert PB, Peterson ML, Follmann D, Hudgens MG, Francis DP, Gurwith M, et al. Correlation between immunologic responses to a recombinant glycoprotein 120 vaccine and incidence of HIV-1 infection in a phase 3 HIV-1 preventive vaccine trial. J Infect Dis 2005;191(5):666–677.
7. Boyer JD, Ugen KE, Wang B, Agadjanyan M, Gilbert L, Bagarazzi ML, et al. Protection of chimpanzees from high-dose heterologous HIV-1 challenge by DNA vaccination. Nat Med 1997;3(5):526–532.
8. Lu S, Arthos J, Montefiori DC, Yasutomi Y, Manson K, Mustafa F, et al. Simian immunodeficiency virus DNA vaccine trial in macaques. J Virol 1996;70 (6):3978–3991.
9. Lu S, Santoro JC, Fuller DH, Haynes JR, Robinson HL. Use of DNAs expressing HIV-1 Env and noninfectious HIV-1 particles to raise antibody responses in mice. Virology 1995;209(1):147–154.
10. Wang B, Ugen KE, Srikantan V, Agadjanyan MG, Dang K, Refaeli Y, et al. Gene inoculation generates immune responses against human immunodeficiency virus type 1. Proc Natl Acad Sci USA 1993;90(9):4156–4160.
11. Lodmell DL, Ray NB, Parnell MJ, Ewalt LC, Hanlon CA, Shaddock JH, et al. DNA immunization protects nonhuman primates against rabies virus. Nat Med 1998;4 (8):949–952.
12. Xiang ZQ, Spitalnik S, Tran M, Wunner WH, Cheng J, Ertl HC. Vaccination with a plasmid vector carrying the rabies virus glycoprotein gene induces protective immunity against rabies virus. Virology 1994;199(1):132–140.
13. Davis HL, McCluskie MJ, Gerin JL, Purcell RH. DNA vaccine for hepatitis B: evidence for immunogenicity in chimpanzees and comparison with other vaccines. Proc Natl Acad Sci USA 1996;93(14):7213–7218.
14. Davis HL, Michel ML, Whalen RG. DNA-based immunization induces continuous secretion of hepatitis B surface antigen and high levels of circulating antibody. Hum Mol Genet 1993;2(11):1847–1851.
15. Michel ML, Davis HL, Schleef M, Mancini M, Tiollais P, Whalen RG. DNA-mediated immunization to the hepatitis B surface antigen in mice: aspects of the humoral response mimic hepatitis B viral infection in humans. Proc Natl Acad Sci USA 1995;92(12):5307–5311.
16. Cox GJ, Zamb TJ, Babiuk LA. Bovine herpesvirus 1: immune responses in mice and cattle injected with plasmid DNA. J Virol 1993;67(9):5664–5667.
17. Ulmer JB, Liu MA, Montgomery DL, Yawman AM, Deck RR, DeWitt CM, et al. Expression and immunogenicity of *Mycobacterium tuberculosis* antigen 85 by DNA vaccination. Vaccine 1997;15(8):792–794.
18. Fynan EF, Webster RG, Fuller DH, Haynes JR, Santoro JC, Robinson HL. DNA vaccines: protective immunizations by parenteral, mucosal, and gene-gun inoculations. Proc Natl Acad Sci USA 1993;90(24):11478–11482.
19. Robinson HL, Hunt LA, Webster RG. Protection against a lethal influenza virus challenge by immunization with a haemagglutinin-expressing plasmid DNA. Vaccine 1993;11(9):957–960.

20. Ulmer JB, Donnelly JJ, Parker SE, Rhodes GH, Felgner PL, Dwarki VJ, et al. Heterologous protection against influenza by injection of DNA encoding a viral protein. Science 1993;259(5102):1745–1749.
21. Yankauckas MA, Morrow JE, Parker SE, Abai A, Rhodes GH, Dwarki VJ, et al. Long-term anti-nucleoprotein cellular and humoral immunity is induced by intramuscular injection of plasmid DNA containing NP gene. DNA Cell Biol 1993;12(9):771–776.
22. Calarota S, Bratt G, Nordlund S, Hinkula J, Leandersson AC, Sandström E, et al. Cellular cytotoxic response induced by DNA vaccination in HIV-1-infected patients. Lancet 1998;351(9112):1320–1325.
23. Calarota SA, Kjerrström A, Islam KB, Wahren B. Gene combination raises broad human immunodeficiency virus-specific cytotoxicity. Hum Gene Ther 2001;12(13):1623–1637.
24. Calarota SA, Leandersson AC, Bratt G, Hinkula J, Klinman DM, Weinhold KJ, et al. Immune responses in asymptomatic HIV-1-infected patients after HIV-DNA immunization followed by highly active antiretroviral treatment. J Immunol 1999;163(4):2330–2338.
25. MacGregor RR, Boyer JD, Ugen KE, Lacy KE, Gluckman SJ, Bagarazzi ML, et al. First human trial of a DNA-based vaccine for treatment of human immunodeficiency virus type 1 infection: safety and host response. J Infect Dis 1998;178(1):92–100.
26. Weber R, Bossart W, Cone R, Luethy R, Moelling K. Phase I clinical trial with HIV-1 gp160 plasmid vaccine in HIV-1-infected asymptomatic subjects. Eur J Clin Microbiol Infect Dis 2001;20(11):800–803.
27. Conry RM, Curiel DT, Strong TV, Moore SE, Allen KO, Barlow DL, et al. Safety and immunogenicity of a DNA vaccine encoding carcinoembryonic antigen and hepatitis B surface antigen in colorectal carcinoma patients. Clin Cancer Res 2002;8(9):2782–2787.
28. Epstein JE, Gorak EJ, Charoenvit Y, Wang R, Freydberg N, Osinowo O, et al. Safety, tolerability, and lack of antibody responses after administration of a PfCSP DNA malaria vaccine via needle or needle-free jet injection, and comparison of intramuscular and combination intramuscular/intradermal routes. Hum Gene Ther 2002;13(13):1551–1560.
29. Le TP, Coonan KM, Hedstrom RC, Charoenvit Y, Sedegah M, Epstein JE, et al. Safety, tolerability and humoral immune responses after intramuscular administration of a malaria DNA vaccine to healthy adult volunteers. Vaccine 2000;18(18):1893–1901.
30. Wang R, Doolan DL, Charoenvit Y, Hedstrom RC, Gardner MJ, Hobart P, et al. Induction of antigen-specific cytotoxic T lymphocytes in humans by a malaria DNA vaccine. Science 1998;282(5388):476–480.
31. Wang R, Epstein J, Baraceros FM, Gorak EJ, Charoenvit Y, Carucci DJ, He et al. Induction of CD4(+) T cell-dependent CD8(+) type 1 responses in humans by a malaria DNA vaccine. Proc Natl Acad Sci USA 2001;98(19):10817–10822.
32. Tagawa ST, Lee P, Snively J, Boswell W, Ounpraseuth S, Lee S, et al. Phase I study of intranodal delivery of a plasmid DNA vaccine for patients with Stage IV melanoma. Cancer 2003;98(1):144–154.
33. Timmerman JM, Singh G, Hermanson G, Hobart P, Czerwinski DK, Taidi B, et al. Immunogenicity of a plasmid DNA vaccine encoding chimeric idiotype in patients with B-cell lymphoma. Cancer Res 2002;62(20):5845–5852.
34. Heeney JL. The critical role of CD4(+) T-cell help in immunity to HIV. Vaccine 2002;20 (15):1961–1963.

35. Heeney JL, Teeuwsen VJ, van Gils M, Bogers WM, De Giuli Morghen C, Radaelli A, et al. Beta-chemokines and neutralizing antibody titers correlate with sterilizing immunity generated in HIV-1 vaccinated macaques. Proc Natl Acad Sci USA 1998;95(18):10803–10808.
36. Heeney JL, van Gils ME, van der Meide P, de Giuli Morghen C, Ghioni C, Gimelli M, et al. The role of type-1 and type-2 T-helper immune responses in HIV-1 vaccine protection. J Med Primatol 1998;27(2–3):50–58.
37. Gallimore A, Cranage M, Cook N, Almond N, Bootman J, Rud E, et al. Early suppression of SIV replication by CD8+ nef-specific cytotoxic T cells in vaccinated macaques. Nat Med 1995;1(11):1167–1173.
38. Klenerman P, Phillips RE, Rinaldo CR, Wahl LM, Ogg G, May RM, et al. Cytotoxic T lymphocytes and viral turnover in HIV type 1 infection. Proc Natl Acad Sci USA 1996;93(26):15323–15328.
39. Schmitz JE, Kuroda MJ, Santra S, Sasseville VG, Simon MA, Lifton MA, et al. Control of viremia in simian immunodeficiency virus infection by CD8+ lymphocytes. Science 1999;283(5403):857–860.
40. Doolan DL, Hoffman SL. Multi-gene vaccination against malaria: a multistage, multi-immune response approach. Parasitol Today 1997;13(5):171–178.
41. Tang DC, DeVit M, Johnston SA. Genetic immunization is a simple method for eliciting an immune response. Nature 1992;356(6365):152–154.
42. Fu TM, Friedman A, Ulmer JB, Liu MA, Donnelly JJ. Protective cellular immunity: cytotoxic T-lymphocyte responses against dominant and recessive epitopes of influenza virus nucleoprotein induced by DNA immunization. J Virol 1997;71 (4):2715–2721.
43. McClements WL, Armstrong ME, Keys RD, Liu MA. Immunization with DNA vaccines encoding glycoprotein D or glycoprotein B, alone or in combination, induces protective immunity in animal models of herpes simplex virus-2 disease. Proc Natl Acad Sci USA 1996;93(21):11414–11420.
44. Donnelly JJ, Martinez D, Jansen KU, Ellis RW, Montgomery DL, Liu MA. Protection against papillomavirus with a polynucleotide vaccine. J Infect Dis 1996;3(2):314–320.
45. Fuller DH, Haynes JR. A qualitative progression in HIV type 1 glycoprotein 120-specific cytotoxic cellular and humoral immune responses in mice receiving a DNA-based glycoprotein 120 vaccine. AIDS Res Hum Retroviruses 1994;10(11):1433–1441.
46. Lu S, Santoro JC, Fuller DH, Haynes JR, Robinson HL. Use of DNAs expressing HIV-1 Env and noninfectious HIV-1 particles to raise antibody responses in mice. Virology 1995;209(1):147–154.
47. Lu S, Arthos J, Montefiori DC, Yasutomi Y, Manson K, Mustafa F, et al. Simian immunodeficiency virus DNA vaccine trial in macaques. J Virol 1996;70(6):3978–3991.
48. World Health Organization. Hepatitis B Fact Sheet; 2008. Available from: http://www.who.int/mediacentre/factsheets/fs204/en/.
49. Geall AJ, Verma A, Otten GR, Shaw CA, Hekele A, Banerjee K, et al. Nonviral delivery of self-amplifying RNA vaccines. Proc Natl Acad Sci USA 2012;109(36):14604–14609.
50. Petsch B, Schnee M, Vogel AB, Lange E, Hoffmann B, Voss D, et al. Protective efficacy of *in vitro* synthesized, specific mRNA vaccines against influenza A virus infection. Nat Biotechnol 2012;30(12):1210–1216.

51. Geall AJ, Verma A, Otten GR, Shaw CA, Hekele A, Banerjee K, et al. Enhancement of cellular immune response in HIV-1 seropositive individuals: a DNA-based trial. Clin Immunol 1999;90(1):100–107.
52. Boyer JD, Cohen AD, Vogt S, Schumann K, Nath B, Ahn L, et al. Vaccination of seronegative volunteers with a human immunodeficiency virus type 1 env/rev DNA vaccine induces antigen-specific proliferation and lymphocyte production of beta-chemokines J Infect Dis 2000;181(2):476–83.
53. Dean HJ, Chen D. Epidermal powder immunization against influenza. Vaccine 2004;23 (5):681–686.
54. Drape RJ, Macklin MD, Barr LJ, Jones S, Haynes JR, Dean HJ. Epidermal DNA vaccine for influenza is immunogenic in humans. Vaccine 2006;24(21):4475–4481.
55. Jones S, Evans K, McElwaine-Johnn H, Sharpe M, Oxford J, Lambkin-Williams R, et al. DNA vaccination protects against an influenza challenge in a double-blind randomised placebo-controlled phase 1b clinical trial. Vaccine 2009;27(18):2506–2512.
56. Roy MJ, Wu MS, Barr LJ, Fuller JT, Tussey LG, Speller S, et al. Induction of antigen-specific CD8+ T cells, T helper cells, and protective levels of antibody in humans by particle-mediated administration of a hepatitis B virus DNA vaccine. Vaccine 2000;19 (7–8):764–778.
57. Robinson HL. Prime boost vaccines power up in people. Nat Med 2003;9(6):642–643.
58. Lu S. Heterologous prime-boost vaccination. Curr Opin Immunol 2009;21(3):346–351.
59. Hanke T, Goonetilleke N, McMichael AJ, Dorrell L. Clinical experience with plasmid DNA- and modified vaccinia virus Ankara-vectored human immunodeficiency virus type 1 clade A vaccine focusing on T-cell induction. J Gen Virol 2007;88(Pt 1):1–12.
60. McConkey SJ, Reece WH, Moorthy VS, Webster D, Dunachie S, Butcher G, et al. Enhanced T-cell immunogenicity of plasmid DNA vaccines boosted by recombinant modified vaccinia virus Ankara in humans. Nat Med 2003;9(6):729–735.
61. Moorthy VS, Imoukhuede EB, Keating S, Pinder M, Webster D, Skinner MA, et al. Phase 1 evaluation of 3 highly immunogenic prime-boost regimens, including a 12-month reboosting vaccination, for malaria vaccination in Gambian men. J Infect Dis 2004;189 (12):2213–2219.
62. Moorthy VS, Pinder M, Reece WH, Watkins K, Atabani S, Hannan C, et al. Safety and immunogenicity of DNA/modified vaccinia virus Ankara malaria vaccination in African adults. J Infect Dis 2003;188(8):1239–1244.
63. McShane H, Brookes R, Gilbert SC, Hill AV. Enhanced immunogenicity of CD4(+) T-cell responses and protective efficacy of a DNA-modified vaccinia virus Ankara prime-boost vaccination regimen for murine tuberculosis. Infect Immun 2001;69 (2):681–686.
64. Casimiro DR, Chen L, Fu TM, Evans RK, Caulfield MJ, Davies ME, et al. Comparative immunogenicity in rhesus monkeys of DNA plasmid, recombinant vaccinia virus, and replication-defective adenovirus vectors expressing a human immunodeficiency virus type 1 gag gene. J Virol 2003;77(11):6305–6313.
65. Casimiro DR, Tang A, Chen L, Fu TM, Evans RK, Davies ME, et al. Vaccine-induced immunity in baboons by using DNA and replication-incompetent adenovirus type 5 vectors expressing a human immunodeficiency virus type 1 gag gene. J Virol 2003;77 (13):7663–7668.

66. Perkins SD, O'Brien LM, Phillpotts RJ. Boosting with an adenovirus-based vaccine improves protective efficacy against Venezuelan equine encephalitis virus following DNA vaccination. Vaccine 2006;24(17):3440–3445.
67. Seaman MS, Xu L, Beaudry K, Martin KL, Beddall MH, Miura A, et al. Multiclade human immunodeficiency virus type 1 envelope immunogens elicit broad cellular and humoral immunity in rhesus monkeys. J Virol 2005;79(5):2956–2963.
68. Shiver JW, Fu TM, Chen L, Casimiro DR, Davies ME, Evans RK, et al. Replication-incompetent adenoviral vaccine vector elicits effective anti-immunodeficiency-virus immunity. Nature 2002;415(6869):331–335.
69. Wu L, Kong WP, Nabel GJ. Enhanced breadth of CD4 T-cell immunity by DNA prime and adenovirus boost immunization to human immunodeficiency virus Env and Gag immunogens. J Virol 2005;79(13):8024–8031.
70. Pancholi P, Liu Q, Tricoche N, Zhang P, Perkus ME, Prince AM. DNA prime-canarypox boost with polycistronic hepatitis C virus (HCV) genes generates potent immune responses to HCV structural and nonstructural proteins. J Infect Dis 2000;182(1):18–27.
71. Lu S. Combination DNA plus protein HIV vaccines. Springer Semin Immunopathol 2006;28(3):255–265.
72. Wang R, Epstein J, Charoenvit Y, Baraceros FM, Rahardjo N, Gay T, et al. Induction in humans of CD8+ and CD4+ T cell and antibody responses by sequential immunization with malaria DNA and recombinant protein. J Immunol 2004;172(9):5561–5569.
73. Wang S, Kennedy JS, West K, Montefiori DC, Coley S, Lawrence J, et al. Cross-subtype antibody and cellular immune responses induced by a polyvalent DNA prime-protein boost HIV-1 vaccine in healthy human volunteers. Vaccine 2008;26(31):3947–3957.
74. Barnett SW, Rajasekar S, Legg H, Doe B, Fuller DH, Haynes JR, et al. Vaccination with HIV-1 gp120 DNA induces immune responses that are boosted by a recombinant gp120 protein subunit. Vaccine 1997;15(8):869–873.
75. Letvin NL, Montefiori DC, Yasutomi Y, Perry HC, Davies ME, Lekutis C, et al. Potent, protective anti-HIV immune responses generated by bimodal HIV envelope DNA plus protein vaccination. Proc Natl Acad Sci USA 1997;94(17):9378–9383.
76. Pal R, Kalyanaraman VS, Nair BC, Whitney S, Keen T, Hocker L, et al. Immunization of rhesus macaques with a polyvalent DNA prime/protein boost human immunodeficiency virus type 1 vaccine elicits protective antibody response against simian human immunodeficiency virus of R5 phenotype. Virology 2006;348(2):341–353.
77. Pal R, Wang S, Kalyanaraman VS, Nair BC, Whitney S, Keen T, et al. Polyvalent DNA prime and envelope protein boost HIV-1 vaccine elicits humoral and cellular responses and controls plasma viremia in rhesus macaques following rectal challenge with an R5 SHIV isolate. J Med Primatol 2005;34(5–6):226–236.
78. Wang S, Pal R, Mascola JR, Chou TH, Mboudjeka I, Shen S, et al. Polyvalent HIV-1 Env vaccine formulations delivered by the DNA priming plus protein boosting approach are effective in generating neutralizing antibodies against primary human immunodeficiency virus type 1 isolates from subtypes A, B, C, D and E. Virology 2006;350 (1):34–47.
79. Jin H, Kang Y, Zheng G, Xie Q, Xiao C, Zhang X, et al. Induction of active immune suppression by co-immunization with DNA- and protein-based vaccines. Virology 2005;337(1):183–191.

80. Kang Y, Jin H, Zheng G, Du X, Xiao C, Zhang X, et al. Co-inoculation of DNA and protein vaccines induces antigen-specific T cell suppression. Biochem Biophys Res Commun 2007;353(4):1034–1039.
81. Bansal A, Jackson B, West K, Wang S, Lu S, Kennedy JS, Goepfert PA. Multifunctional T-cell characteristics induced by a polyvalent DNA prime/protein boost human immunodeficiency virus type 1 vaccine regimen given to healthy adults are dependent on the route and dose of administration. J Virol 2008;82(13):6458–6469.
82. Ledgerwood JE, Wei CJ, Hu Z, Gordon IJ, Enama ME, Hendel CS, et al.; VRC 306 Study Team. DNA priming and influenza vaccine immunogenicity: two phase 1 open label randomised clinical trials. Lancet Infect Dis 2011;11(12):916–924.
83. Harari A, Bart PA, Stöhr W, Tapia G, Garcia M, Medjitna-Rais E, et al. An HIV-1 clade C DNA prime, NYVAC boost vaccine regimen induces reliable, polyfunctional, and long-lasting T cell responses. J Exp Med 2008;205(1):63–77.
84. Lu S. Immunogenicity of DNA vaccines in humans. Hum Vaccin 2008;4(6):449–452.
85. Hanke T, Goonetilleke N, McMichael AJ, Dorrell L. Clinical experience with plasmid DNA- and modified vaccinia virus Ankara-vectored human immunodeficiency virus type 1 clade A vaccine focusing on T-cell induction, J Gen Virol 2007;88(Pt 1):1–12.
86. World Health Organization. Malaria Fact Sheet; 2011. Available from: http://www.who.int/mediacentre/factsheets/fs094/en/.
87. Moorthy VS, McConkey S, Roberts M, Gothard P, Arulanantham N, Degano P, Schneider J, et al. Safety of DNA and modified vaccinia virus Ankara vaccines against liver-stage *P. falciparum* malaria in non-immune volunteers. Vaccine 2003;21(17-18): 1995–2002.
88. Rottinghaus ST, Poland GA, Jacobson RM, Barr LJ, Roy MJ. Hepatitis B DNA vaccine induces protective antibody responses in human non-responders to conventional vaccination. Vaccine 2003;21(131):4604–4608.
89. Mancini-Bourgine M, Fontaine H, Bréchot C, Pol S, Michel ML. Immunogenicity of a hepatitis B DNA vaccine administered to chronic HBV carriers. Vaccine 2006;24 (21):4482–4489.
90. Mancini-Bourgine M, Fontaine H, Scott-Algara D, Pol S, Bréchot C, Michel ML. Induction or expansion of T-cell responses by a hepatitis B DNA vaccine administered to chronic HBV carriers. Hepatology 2004;40(4):874–882.
91. Wu TC. Therapeutic human papillomavirus DNA vaccination strategies to control cervical cancer. Eur J Immunol 2007;37(2):310–314.
92. Klencke B, Matijevic M, Urban RG, Lathey JL, Hedley ML, Berry M, et al. Encapsulated plasmid DNA treatment for human papillomavirus 16-associated anal dysplasia: a Phase I study of ZYC101. Clin Cancer Res 2002;8(5):1028–1037.
93. Sheets E. Immunotherapy of human cervical high-grade cervical intraepithelial neoplasia with microparticle-delivered human papillomavirus 16 E7 plasmid DNA. Am J Obstet Gynecol 2003;188(4):916–926.
94. Garcia F, Petry KU, Muderspach L, Gold MA, Braly P, Crum CP, et al. ZYC101a for treatment of high-grade cervical intraepithelial neoplasia. Obstet Gynecol 2004;103 (2):317–326.
95. Rice J, Ottensmeier CH, Stevenson FK. DNA vaccines: precision tools for activating effective immunity against cancer. Nat Rev Cancer 2008;8(2):108–120.

96. Belldegrun A, Tso CL, Zisman A, Naitoh J, Said J, Pantuck AJ, et al. Interleukin 2 gene therapy for prostate cancer: phase I clinical trial and basic biology. Hum Gene Ther 2001;12(8):883–892.
97. Galanis E, Burch PA, Richardson RL, Lewis B, Pitot HC, Frytak S, et al. Intratumoral administration of a 1,2-dimyristyloxypropyl-3-dimethylhydroxyethyl ammonium bromide/dioleoylphosphatidylethanolamine formulation of the human interleukin-2 gene in the treatment of metastatic renal cell carcinoma. Cancer 2004;101(11):2557–2566.
98. Tagawa ST, Lee P, Snively J, Boswell W, Ounpraseuth S, Lee S, et al. Phase I study of intranodal delivery of a plasmid DNA vaccine for patients with Stage IV melanoma. Cancer 2003;98(1):144–154.
99. Stopeck AT, Hersh EM, Akporiaye ET, Harris DT, Grogan T, Unger E, et al. Phase I study of direct gene transfer of an allogeneic histocompatibility antigen, HLA-B7, in patients with metastatic melanoma. J Clin Oncol 1997;15(1):341–349.
100. Rini BI, Selk LM, Vogelzang NJ. Phase I study of direct intralesional gene transfer of HLA-B7 into metastatic renal carcinoma lesions. Clin Cancer Res 1999;5(10):2766–2772.
101. Stevenson FK, Zhu D, King CA, Ashworth LJ, Kumar S, Thompsett A, et al. A genetic approach to idiotypic vaccination for B cell lymphoma. Ann NY Acad Sci 1995;772: 212–226.
102. Timmerman JM, Singh G, Hermanson G, Hobart P, Czerwinski DK, Taidi B, et al. Immunogenicity of a plasmid DNA vaccine encoding chimeric idiotype in patients with B-cell lymphoma. Cancer Res 2002;62(20):5845–5852.
103. Abdulhaqq SA, Weiner DB. Weiner, DNA vaccines: developing new strategies to enhance immune responses. Immunol Res 2008;42(1–3):219–232.
104. Laddy DJ, Weiner DB. From plasmids to protection: a review of DNA vaccines against infectious diseases. Int Rev Immunol 2006;25(3–4):99–123; J Virol 2000;74(7):3427–3429.
105. Kim JJ, Yang JS, VanCott TC, Lee DJ, Manson KH, Wyand MS, et al. Modulation of antigen-specific humoral responses in rhesus macaques by using cytokine cDNAs as DNA vaccine adjuvants. J Virol 2000;74(7):3427–3429.
106. Barouch DH, Santra S, Schmitz JE, Kuroda MJ, Fu TM, Wagner W, et al. Control of viremia and prevention of clinical AIDS in rhesus monkeys by cytokine-augmented DNA vaccination. Science 2000;290(5491):486–492.
107. Iwasaki A, Stiernholm BJ, Chan AK, Berinstein NL, Barber BH. Enhanced CTL responses mediated by plasmid DNA immunogens encoding costimulatory molecules and cytokines. J Immunol 1997;158(10):4591–601.
108. Kim JJ, Ayyavoo V, Bagarazzi ML, Chattergoon MA, Dang K, Wang B, et al. *In vivo* engineering of a cellular immune response by coadministration of IL-12 expression vector with a DNA immunogen. J Immunol 1997;158(2):816–826.
109. Tsuji T, Hamajima K, Fukushima J, Xin KQ, Ishii N, Aoki I, et al. Enhancement of cell-mediated immunity against HIV-1 induced by coinnoculation of plasmid-encoded HIV-1 antigen with plasmid expressing IL-12. J Immunol 1997;158(8):4008–4013.
110. Kim JJ, Yang JS, Dang K, Manson KH, Weiner DB. Engineering enhancement of immune responses to DNA-based vaccines in a prostate cancer model in rhesus macaques through the use of cytokine gene adjuvants. Clin Cancer Res 2001;7(3 Suppl):882s–889s.

111. Kutzler MA, Robinson TM, Chattergoon MA, Choo DK, Choo AY, Choe PY, et al. Coimmunization with an optimized IL-15 plasmid results in enhanced function and longevity of CD8 T cells that are partially independent of CD4 T cell help. J Immunol 2005;175(1):112–123.

112. Sin JI, Kim JJ, Boyer JD, Ciccarelli RB, Higgins TJ, Weiner DB. *In vivo* modulation of vaccine-induced immune responses toward a Th1 phenotype increases potency and vaccine effectiveness in a herpes simplex virus type 2 mouse model. J Virol 1999;73 (1):501–509.

113. Xin KQ, Hamajima K, Sasaki S, Tsuji T, Watabe S, Okada E, et al. IL-15 expression plasmid enhances cell-mediated immunity induced by an HIV-1 DNA vaccine. Vaccine 1999;17(7–8):858–866.

114. Boyer JD, Robinson TM, Kutzler MA, Vansant G, Hokey DA, Kumar S, et al. Protection against simian/human immunodeficiency virus (SHIV) 89.6P in macaques after coimmunization with SHIV antigen and IL-15 plasmid. Proc Natl Acad Sci USA 2007;104 (47):18648–18653.

115. Lu S. Gene-based adjuvants: a new meaning. Hum Gene Ther 2009;20(10):1101–1102.

116. Kraynyak KA, Kutzler MA, Cisper NJ, Laddy DJ, Morrow MP, Waldmann TA, et al. Plasmid-encoded interleukin-15 receptor alpha enhances specific immune responses induced by a DNA vaccine *in vivo*. Hum Gene Ther 2009;20(10):1143–1156.

117. Xiang Z, Ertl HC. Manipulation of the immune response to a plasmid-encoded viral antigen by coinoculation with plasmids expressing cytokines. Immunity 1995;2(2):129–135.

118. Sin JI, Kim JJ, Ugen KE, Ciccarelli RB, Higgins TJ, Weiner DB. Enhancement of protective humoral (Th2) and cell-mediated (Th1) immune responses against herpes simplex virus-2 through co-delivery of granulocyte-macrophage colony-stimulating factor expression cassettes. Eur J Immunol 1998;28(11):3530–3540.

119. Wang R, Richie TL, Baraceros MF, Rahardjo N, Gay T, Banania JG, et al. Boosting of DNA vaccine-elicited gamma interferon responses in humans by exposure to malaria parasites. Infect Immun 2005;73(5):2863–2672.

120. Hellerstein M, Xu Y, Marino T, Lu S, Yi H, Wright ER, Robinson HL. Co-expression of HIV-1 virus-like particles and granulocyte-macrophage colony stimulating factor by GEO-D03 DNA vaccine. Hum Vaccin Immunother 2012;8(11):1654–1658.

121. Tudor D, Dubuquoy C, Gaboriau V, Lefèvre F, Charley B, Riffault S. TLR9 pathway is involved in adjuvant effects of plasmid DNA-based vaccines. Vaccine 2005;23 (10):1258–1264.

122. Klinman DM, Yamshchikov G, Ishigatsubo Y. Contribution of CpG motifs to the immunogenicity of DNA vaccines. J Immunol 1997;158(8):3635–3639.

123. Spies B, Hochrein H, Vabulas M, Huster K, Busch DH, Schmitz F, et al. Vaccination with plasmid DNA activates dendritic cells via Toll-like receptor 9 (TLR9) but functions in TLR9-deficient mice. J Immunol 2003;171(11):5908–5912.

124. Rottembourg D, Filippi CM, Bresson D, Ehrhardt K, Estes EA, Oldham JE, et al. Essential role for TLR9 in prime but not prime-boost plasmid DNA vaccination to activate dendritic cells and protect from lethal viral infection. J Immunol 2010;184 (12):7100–7107.

125. Davis HL, Suparto II, Weeratna RR, Jumintarto, Iskandriati DD, Chamzah SS, et al. CpG DNA overcomes hyporesponsiveness to hepatitis B vaccine in orangutans. Vaccine 2000;18(18):1920–1924.

126. Payette PJ, Ma X, Weeratna RD, McCluskie MJ, Shapiro M, Engle RE, et al. Testing of CpG-optimized protein and DNA vaccines against the hepatitis B virus in chimpanzees for immunogenicity and protection from challenge. Intervirology 2006;49(3):144–151.
127. Davis BS, Chang GJ, Cropp B, Roehrig JT, Martin DA, Mitchell CJ, et al. West Nile virus recombinant DNA vaccine protects mouse and horse from virus challenge and expresses *in vitro* a noninfectious recombinant antigen that can be used. J Virol 2001;75(9):4040–4047.
128. Kilpatrick AM, Dupuis AP, Chang GJ, Kramer LD. DNA vaccination of American robins (*Turdus migratorius*) against West Nile virus. Vector Borne Zoonotic Dis 2010;10 (4):377–380.
129. Martin JE, Pierson TC, Hubka S, Rucker S, Gordon IJ, Enama ME, et al. A West Nile virus DNA vaccine induces neutralizing antibody in healthy adults during a phase 1 clinical trial. J Infect Dis 2007;196(12):1732–1740.
130. Anderson ED, Mourich DV, Fahrenkrug SC, LaPatra S, Shepherd J, Leong JA. Genetic immunization of rainbow trout (*Oncorhynchus mykiss*) against infectious hematopoietic necrosis virus. Mol Mar Biol Biotechnol 1996;5(2):114–122.
131. Acosta F, Collet B, Lorenzen N, Ellis AE. Expression of the glycoprotein of viral haemorrhagic septicaemia virus (VHSV) on the surface of the fish cell line RTG-P1 induces type 1 interferon expression in neighbouring cells. Fish Shellfish Immunol 2006;21(3):272–278.
132. Purcell MK, Nichols KM, Winton JR, Kurath G, Thorgaard GH, Wheeler P, et al. Comprehensive gene expression profiling following DNA vaccination of rainbow trout against infectious hematopoietic necrosis virus. Mol Immunol 2006;43(13):2089–2106.
133. Yasuike M, Kondo H, Hirono I, Aoki T. Difference in Japanese flounder, *Paralichthys olivaceus* gene expression profile following hirame rhabdovirus (HIRRV) G and N protein DNA vaccination. Fish Shellfish Immunol 2007;23(3):531–541.
134. Bergman PJ, McKnight J, Novosad A, Charney S, Farrelly J, Craft D, et al. Long-term survival of dogs with advanced malignant melanoma after DNA vaccination with xenogeneic human tyrosinase: a phase I trial. Clin Cancer Res 2003;9(4):1284–1290.
135. Liao JC, Gregor P, Wolchok JD, Orlandi F, Craft D, Leung C, et al. Vaccination with human tyrosinase DNA induces antibody responses in dogs with advanced melanoma. Cancer Immun 2006;6: 8.
136. Khan AS, Bodles-Brakhop AM, Fiorotto ML, Draghia-Akli R. Effects of maternal plasmid GHRH treatment on offspring growth. Vaccine 2010;28(8):1905–1910.
137. Thacker EL, Holtkamp DJ, Khan AS, Brown PA, Draghia-Akli R. Plasmid-mediated growth hormone-releasing hormone efficacy in reducing disease associated with *Mycoplasma hyopneumoniae* and porcine reproductive and respiratory syndrome virus infection. J Anim Sci 2006;84(3):733–742.
138. Zhang L, Yang Y, Yang X, Zhao J, Yang J, Liu F, et al. T cell epitope-based peptide-DNA dual vaccine induces protective immunity against *Schistosoma japonicum* infection in C57BL/6J mice. Microbes Infect 2008;10(3):251–259.
139. Kodihalli S, Goto H, Kobasa DL, Krauss S, Kawaoka Y, Webster RG. DNA vaccine encoding hemagglutinin provides protective immunity against H5N1 influenza virus infection in mice. J Virol 1999;73(3):2094–2098.
140. Laddy DJ, Yan J, Corbitt N, Kobasa D, Kobinger GP, Weiner DB. Immunogenicity of novel consensus-based DNA vaccines against avian influenza. Vaccine 2007;25 (16):2984–2989.

141. Gorman CM, Merlino GT, Willingham MC, Pastan I, Howard BH. The Rous sarcoma virus long terminal repeat is a strong promoter when introduced into a variety of eukaryotic cells by DNA-mediated transfection. Proc Natl Acad Sci USA 1982;79 (22):6777–6781.

142. Moreau P, Hen R, Wasylyk B, Everett R, Gaub MP, Chambon P. The SV40 72 base repair repeat has a striking effect on gene expression both in SV40 and other chimeric recombinants. Nucleic Acids Res 1981;9(22):6047–6068.

143. Galvin TA, Muller J, Khan AS. Effect of different promoters on immune responses elicited by HIV-1 gag/env multigenic DNA vaccine in *Macaca mulatta* and *Macaca nemestrina*. Vaccine 2000;18(23):2566–2583.

144. Barouch DH, Yang ZY, Kong WP, Korioth-Schmitz B, Sumida SM, Truitt DM, et al. A human T-cell leukemia virus type 1 regulatory element enhances the immunogenicity of human immunodeficiency virus type 1 DNA vaccines in mice and nonhuman primates. J Virol 2005;79(14):8828–8834.

145. Chapman BS, Thayer RM, Vincent KA, Haigwood NL. Effect of intron A from human cytomegalovirus (Towne) immediate-early gene on heterologous expression in mammalian cells. Nucleic Acids Res 1991;19(14):3979–3986.

146. Huang MT, Gorman CM. Gorman, Intervening sequences increase efficiency of RNA 3′ processing and accumulation of cytoplasmic RNA. Nucleic Acids Res 1990;18(4):937–947.

147. Proudfoot NJ. How RNA polymerase II terminates transcription in higher eukaryotes. Trends Biochem Sci 1989;14(3):105–110.

148. Xu ZL, Mizuguchi H, Ishii-Watabe A, Uchida E, Mayumi T, Hayakawa T. Optimization of transcriptional regulatory elements for constructing plasmid vectors. Gene 2001;272 (1–2):149–156.

149. Gustafsson C, Govindarajan S, Minshull J. Codon bias and heterologous protein expression. Trends Biotechnol 2004;22(7):346–353.

150. Ko HJ, Ko SY, Kim YJ, Lee EG, Cho SN, Kang CY. Optimization of codon usage enhances the immunogenicity of a DNA vaccine encoding mycobacterial antigen Ag85B. Infect Immun 2005;73(9):5666–5674.

151. Uchijima M, Yoshida A, Nagata T, Koide Y. Optimization of codon usage of plasmid DNA vaccine is required for the effective MHC class I-restricted T cell responses against an intracellular bacterium. J Immunol 1998;161(10):5594–5599.

152. Kozak M. Recognition of AUG and alternative initiator codons is augmented by G in position +4 but is not generally affected by the nucleotides in positions +5 and +6. EMBO J 1997;16(9):2482–2492.

153. Li Y, Luo L, Thomas DY, Kang CY. The HIV-1 Env protein signal sequence retards its cleavage and down-regulates the glycoprotein folding. Virology 2000;272(2):417–428.

154. Malin AS, Huygen K, Content J, Mackett M, Brandt L, Andersen P, et al. Vaccinia expression of *Mycobacterium tuberculosis*-secreted proteins: tissue plasminogen activator signal sequence enhances expression and immunogenicity of *M. tuberculosis* Ag85. Microbes Infect 2000;2(14):1677–1685.

155. Hartikka J, Bozoukova V, Ferrari M, Sukhu L, Enas J, Sawdey M, et al. Vaxfectin enhances the humoral immune response to plasmid DNA-encoded antigens. Vaccine 2001;19(15–16):1911–1923.

156. Reyes L, Hartikka J, Bozoukova V, Sukhu L, Nishioka W, Singh G, et al. Vaxfectin enhances antigen specific antibody titers and maintains Th1 type immune responses to plasmid DNA immunization. Vaccine 2001;19(27):3778–3786.
157. Sedegah M, Rogers WO, Belmonte M, Belmonte A, Banania G, Patterson NB, et al. Vaxfectin enhances both antibody and *in vitro* T cell responses to each component of a 5-gene *Plasmodium falciparum* plasmid DNA vaccine mixture administered at low doses. Vaccine 2010;28(17):3055–3065.
158. Smith LR, Wloch MK, Ye M, Reyes LR, Boutsaboualoy S, Dunne CE, et al. Phase 1 clinical trials of the safety and immunogenicity of adjuvanted plasmid DNA vaccines encoding influenza A virus H5 hemagglutinin. Vaccine 2010;28(13):2565–2572.
159. Alexakis T, Boadi DK, Quong D, Groboillot A, O'Neill I, Poncelet D, et al. Microencapsulation of DNA within alginate microspheres and crosslinked chitosan membranes for *in vivo* application. Appl Biochem Biotechnol 1995;50(1):93–106.
160. Herrmann JE, Chen SC, Jones DH, Tinsley-Bown A, Fynan EF, Greenberg HB, et al. Immune responses and protection obtained by oral immunization with rotavirus VP4 and VP7 DNA vaccines encapsulated in microparticles. Virology 1999;259(1):148–153.
161. Kaur R, Rauthan M, Vrati S. Immunogenicity in mice of a cationic microparticle-adsorbed plasmid DNA encoding Japanese encephalitis virus envelope protein. Vaccine 2004;22(21–22):2776–2782.
162. Otten GR, Schaefer M, Doe B, Liu H, Srivastava I, Megede JZ, et al. Enhanced potency of plasmid DNA microparticle human immunodeficiency virus vaccines in rhesus macaques by using a priming-boosting regimen with recombinant proteins. J Virol 2005;79(13):8189–8200.
163. Wang F, He XW, Jiang L, Ren D, He Y, Li DA, Sun SH. Enhanced immunogenicity of microencapsulated multiepitope DNA vaccine encoding T and B cell epitopes of foot-and-mouth disease virus in mice. Vaccine 2006;24(12):2017–2026.
164. Minigo G, Scholzen A, Tang CK, Hanley JC, Kalkanidis M, Pietersz GA, et al. Poly-L-lysine-coated nanoparticles: a potent delivery system to enhance DNA vaccine efficacy. Vaccine 2007;25(7):1316–1327.
165. Porgador A, Irvine KR, Iwasaki A, Barber BH, Restifo NP, Germain RN. Predominant role for directly transfected dendritic cells in antigen presentation to CD8+ T cells after gene gun immunization. J Exp Med 1998;188(6):1075–1082.
166. Condon C, Watkins SC, Celluzzi CM, Thompson K, Falo LD Jr. DNA-based immunization by *in vivo* transfection of dendritic cells. Nat Med 1996;2(10):1122–1128.
167. Rols MP, Delteil C, Golzio M, Dumond P, Cros S, Teissie J. *In vivo* electrically mediated protein and gene transfer in murine melanoma. Nat Biotechnol 1998;16(2):168–171.
168. Choi AH, Smiley K, Basu M, McNeal MM, Shao M, Bean JA, et al. Protection of mice against rotavirus challenge following intradermal DNA immunization by Biojector needle-free injection. Vaccine 2007;25(16):3215–3218.
169. Ledgerwood JE, Hu Z, Gordon IJ, Yamshchikov G, Enama ME, Plummer S, et al.; VRC 304 and VRC 305 Study Teams. Influenza virus h5 DNA vaccination is immunogenic by intramuscular and intradermal routes in humans. Clin Vaccine Immunol 2012;19 (11):1792–1177.
170. Aguiar JC, et al. Enhancement of the immune response in rabbits to a malaria DNA vaccine by immunization with a needle-free jet device. Vaccine 2001;20(1-2): 275–280.

171. Trimble C, Lin CT, Hung CF, Pai S, Juang J, He L, et al. Comparison of the CD8+ T cell responses and antitumor effects generated by DNA vaccine administered through gene gun, biojector, and syringe. Vaccine 2003;21(25–26):4036–4042.

172. Rogers WO, Baird JK, Kumar A, Tine JA, Weiss W, Aguiar JC, et al. Multistage multiantigen heterologous prime boost vaccine for *Plasmodium knowlesi* malaria provides partial protection in rhesus macaques. Infect Immun 2001;69(9):5565–5572.

173. Hoffman PN, Abuknesha RA, Andrews NJ, Samuel D, Lloyd JS. A model to assess the infection potential of jet injectors used in mass immunisation. Vaccine 2001;19(28–29):4020–4027.

174. Nichols WW, Ledwith BJ, Manam SV, Troilo PJ. Potential DNA vaccine integration into host cell genome. Ann NY Acad Sci 1995;772:30–39.

175. Manam S, Ledwith BJ, Barnum AB, Troilo PJ, Pauley CJ, Harper LB, et al. Plasmid DNA vaccines: tissue distribution and effects of DNA sequence, adjuvants and delivery method on integration into host DNA. Intervirology 2000;43(4–6):273–281.

176. Martin T, Parker SE, Hedstrom R, Le T, Hoffman SL, Norman J, et al. Plasmid DNA malaria vaccine: the potential for genomic integration after intramuscular injection. Hum Gene Ther 1999;10(5):759–768.

177. Wang Z, Troilo PJ, Wang X, Griffiths TG, Pacchione SJ, Barnum AB, et al. Detection of integration of plasmid DNA into host genomic DNA following intramuscular injection and electroporation. Gene Ther 2004;11(8):711–721.

178. Krieg AM. CpG motifs in bacterial DNA and their immune effects. Annu Rev Immunol 2002;20:709–760.

179. Klinman DM, Takeshita F, Kamstrup S, Takeshita S, Ishii K, Ichino M, et al. DNA vaccines: capacity to induce auto-immunity and tolerance. Dev Biol (Basel) 2000;104: 45–51.

180. Mor G, Singla M, Steinberg AD, Hoffman SL, Okuda K, Klinman DM. Do DNA vaccines induce autoimmune disease? Hum Gene Ther 1997;8(3):293–300.

181. Isaguliants MG, Iakimtchouk K, Petrakova NV, Yermalovich MA, Zuber AK, Kashuba VI, et al. Gene immunization may induce secondary antibodies reacting with DNA. Vaccine 2004;22(11–12):1576–1585.

182. Kennedy JS, Co M, Green S, Longtine K, Longtine J, O'Neill MA, et al. The safety and tolerability of an HIV-1 DNA prime-protein boost vaccine (DP6-001) in healthy adult volunteers. Vaccine 2008;26(35):4420–4424.

183. Cooke JR, McKie EA, Ward JM, Keshavarz-Moore E. Impact of intrinsic DNA structure on processing of plasmids for gene therapy and DNA vaccines. J Biotechnol 2004;114 (3):239–254.

184. Kato M. Polypyrimidine/polypurine sequence in plasmid DNA enhances formation of dimer molecules in *Escherichia coli*. Dimerization of plasmid DNA in *Escherichia coli*. Mol Biol Rep 1993;18(3):183–187.

185. Bichara M, Schumacher S, Fuchs RP. Genetic instability within monotonous runs of CpG sequences in *Escherichia coli*. Genetics 1995;140(3):897–907.

186. Peterson DO, Beifuss KK, Morley KL. Context-dependent gene expression: cis-acting negative effects of specific procaryotic plasmid sequences on eucaryotic genes. Mol Cell Biol 1987;7(4):1563–1567.

187. Kushner PJ, Baxter JD, Duncan KG, Lopez GN, Schaufele F, Uht RM, et al. Eukaryotic regulatory elements lurking in plasmid DNA: the activator protein-1 site in pUC. Mol Endocrinol 1994;8(4):405–407.

188. Williams JA, Luke J, Johnson L, Hodgson C. pDNAVACCultra vector family: high throughput intracellular targeting DNA vaccine plasmids. Vaccine 2006;24(21):4671–4676.

189. Carnes AE. Fermentation design for the manufacture of therapeutic plasmid DNA. BioProcess Intl 2005;3:36–44.

190. Chartrain M, Bentley LK, Krulewicz BA, Listner KM, Sun W, Lee CB. Process for large scale production of plasmid DNA by *E. coli* fermentation. Patent W005078115. 2005.

191. Huber H, Weigl G, Buchinger W. Fed-batch fermentation process and culture medium for the production of plasmid DNA in *E. coli* on a manufacturing scale. Patent WO2005097990. April 6, 2005.

192. Huber H, Pacher C, Necina R, Kollman F, Reinisch C. A method for producing plasmid DNA on a manufacturing scale uses *Escherichia coli* K-12 strain JM108. The process results in high yield and homogeneity of plasmid DNA. Patent WO2005098002. April 6, 2005.

193. Carnes AE, Hodgson CP, Williams JA. Inducible *Escherichia coli* fermentation for increased plasmid DNA production. Biotechnol Appl Biochem 2006;45(Pt 3):155–166.

194. Klinman DM, Klaschik S, Tross D, Shirota H, Steinhagen F. FDA guidance on prophylactic DNA vaccines: analysis and recommendations. Vaccine 2010;28 (16):2801–2805.

# Multifunctional Polymeric Nanosystems for RNA Interference Therapy

ARUN IYER, SHANTHI GANESH, QIONG L. ZHOU, and MANSOOR AMIJI

## 18.1 INTRODUCTION

RNA-interference (RNAi) is one of the most powerful tools for selective silencing of specific aberrant genes for therapeutic management of diseases such as cancer, diabetes, and inflammatory disorders. However, the labile nature of nucleic acid constructs coupled with the difficulty in selectively delivering them to the target tissues, cells, and the intracellular cellular machinery within the cells has been a primary hindrance for the clinical translation of RNAi. Multifunctional nanosystems provide distinct advantages over other delivery strategies based on their ability to selectively target diseased tissues, such as solid tumors, based on passive and active targeting principles. Such nanosystems can incorporate multiple therapeutic payloads, such as small molecule drugs, small interfering RNAs (siRNA), and microRNAs (miRNA) as well as imaging and contrast agents for simultaneous disease detection, quantifying siRNA and miRNA gene expression and knockdown, examining *in vivo* fate and sites of localization and excretion, and exploring therapeutic potentials *in vivo*. This chapter describes the principles and mechanisms based on the RNAi phenomenon and the barriers to its successful clinical translation. Furthermore, the chapter discusses the modes of delivery that have been used, including the advantages of multifunctional polymeric nano-systems with illustrative examples covering their use in the management of diseases such as cancer, inflammation, and other chronic diseases.

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*,
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

## 18.2 RNA INTERFERENCE THERAPY

### 18.2.1 Principles and Mechanism of Gene Silencing

Gene silencing is a mechanism by which cells can switch off gene expression and therefore prevent protein production. Gene silencing can occur either at the transcriptional or posttranscriptional level. For posttranscriptional gene silencing, the mRNA of particular gene is blocked or degraded commonly by a mechanism called RNAi (Fig. 18.1). The RNAi phenomenon was discovered in 1998 by Andrew Fire and Craig Mello [1], who were awarded the Nobel Prize in 2006. RNAi depends on two small ribonucleic acid molecules, siRNA and miRNA. siRNAs consist of 21-base pair double-stranded RNAs that are generated from long double-stranded RNAs on their cleavage by an RNase III enzyme called Dicer [2,3]. Each siRNA is then separated into single strands, and the antisense strand of the double-stranded RNA is incorporated into the RNA-induced silencing complex (RISC) [4–6] that binds to a complementary sequence of an mRNA molecule. The RNAi (mRNA cleavage) in cytoplasm is induced by Argonaute proteins [7,8], a catalytic component of the RISC complex. miRNAs are 21- to 23-nucleotide-long RNAs and transcribed from their own genes by RNA polymerase II. A miRNA is expressed from a long RNA-coding gene as a primary transcript known as pri-miRNA, which is cleaved by the RNase III enzyme, Drosha [9] to a 65-nucleotide stem-loop structure called pre-miRNA. The pre-miRNA is then exported from the cell nucleus into the cytoplasm, where its loop is removed by Dicer. The mature miRNAs are produced and integrated into RISC. Thus, miRNAs and siRNAs share the same cellular machinery downstream of their initial processing [10]. Binding of a miRNA–Argonaute protein complex to the seed region of 3′ untranslated region (UTR) of an mRNA silences its expression [11]. Most miRNAs trigger general degradation of the target mRNA [12]. In some cases, the miRNA may block the translation of the mRNA into protein [13]. The human gene contains more than 500 miRNAs and each miRNA can repress hundreds of genes, regulating almost every cellular process [11].

### 18.2.2 Small Interfering RNA and Micro RNA as Therapeutic Agents

RNAi can also be activated exogenously by introducing synthetic siRNAs and miRNAs directly into the cytoplasm of cells of interest. The selective and robust effect of siRNA on gene expression makes it an invaluable research tool for functional genomics and as a potential therapeutic agent. siRNA has the ability to knock down or silence virtually any human gene, which greatly expands the number of therapeutic targets that can be explored compared with small molecules. Thus, siRNA can be applied for diseases caused by abnormal gene overexpression or mutation, such as cancer and diabetes treatment.

Aberrant miRNA expression has been linked to many diseases (e.g., the let-7 miRNA prevents proliferation of cancer stem cells [14], but miRNA 143 promotes differentiation of adipocytes) [15,16]. miR 375 regulates insulin secretion in pancreatic islet cells [17], miRNA 122 participates in liver disease [18,19], and miRNA 103/107 is overexpressed in liver biopsies from patients with alcoholic liver

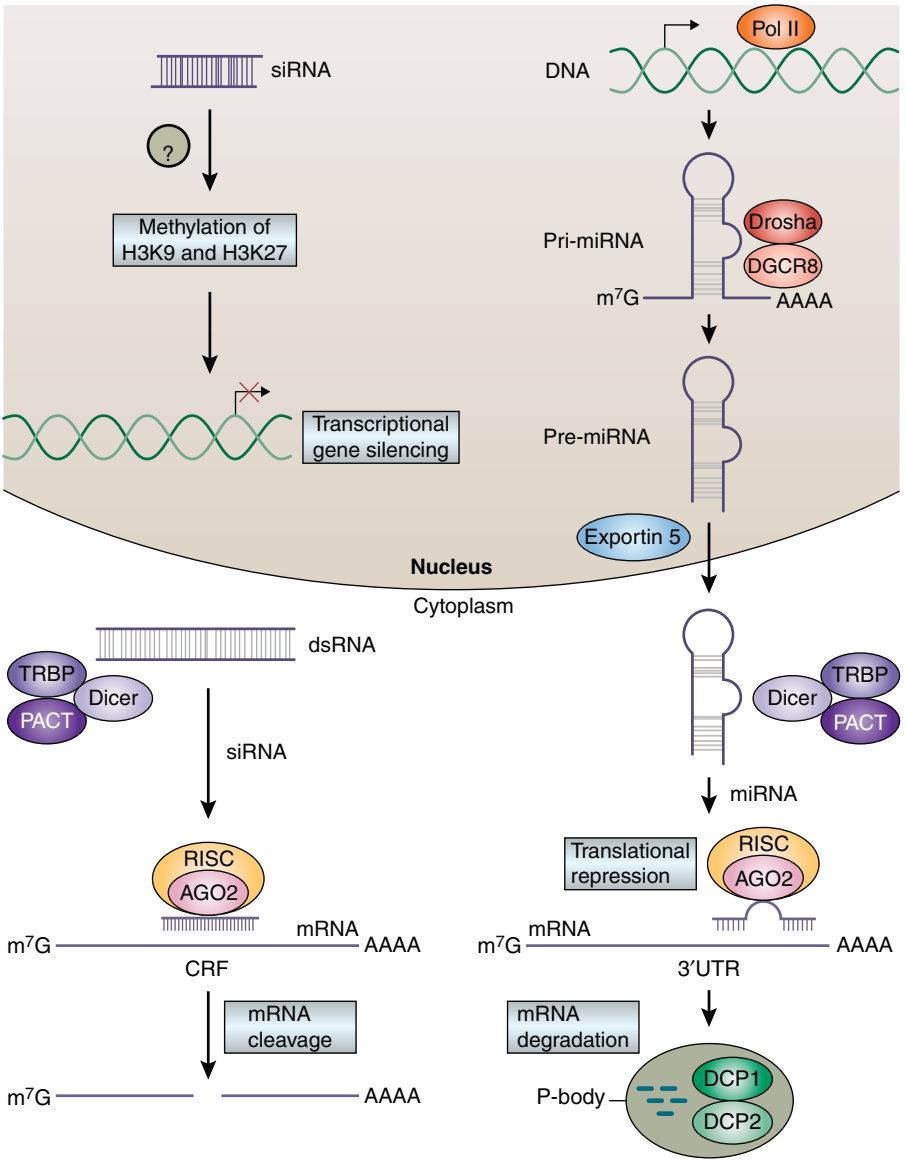


**FIGURE 18.1** Mechanism of RNA interference (RNAi)-mediated gene silencing. AGO2: Argonaute RISC catalytic component 2; DCP1/2: mRNA-degrading enzymes; PACT: protein activator of the interferon-induced protein kinase; RISC: RNA-induced silencing complex; TRPB: typtophan synthase beta subunit; UTR: untranslated region. (Reprinted with permission from Ref. [135].)

disease and nonalcoholic fatty liver disease [20]. Thus, miRNAs and their mimics and inhibitors have great potentials as therapeutic agents.

Together, RNAi represents an important therapeutic strategy. Clinical trials with siRNAs are already in progress [21]. However, the widespread use of siRNAs, miRNAs, and their inhibitors as therapeutic agents for disease prevention and treatment requires the development of safe and effective delivery vehicles.

### 18.2.3 Challenges in Systemic Delivery of RNA-Interference Therapeutics

Systemic delivery of RNAi therapeutics to the targeted tissue and cells has been proven challenging and has many barriers. Intravascular barriers include engulfment of naked siRNAs by macrophages or degradation by serum nucleases [22]. Extracellular nanoparticles, on the other hand, encounter degradation by immune cells and extracellular matrix (ECM). Particles that escape these barriers preferentially traffic to tumors by the enhanced permeability and retention (EPR) effect (Figs. 18.2 and 18.3).

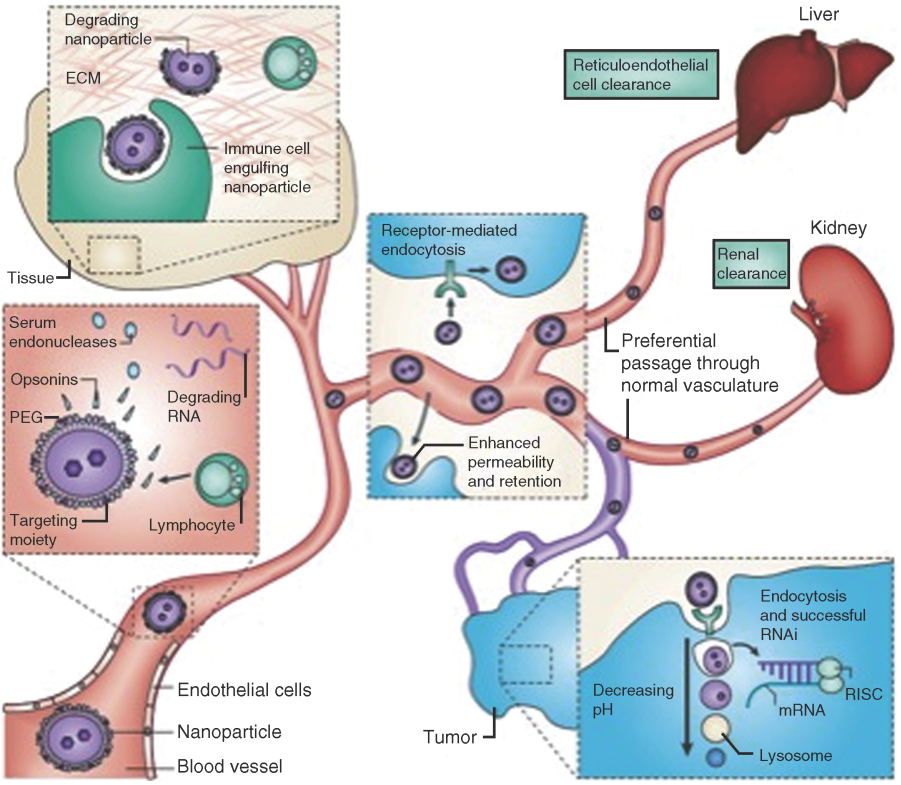


**FIGURE 18.2** Biological barriers of RNA-interference (RNAi) delivery. ECM: extracellular matrix; PEG: polyethylene glycol; RISC: RNA-induced silencing complex. (Reprinted with permission from Ref. [136].)

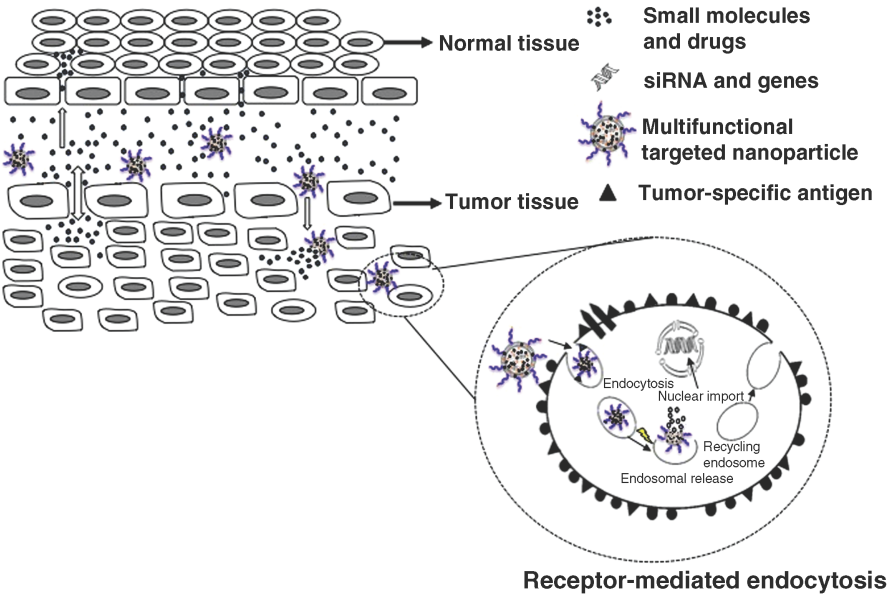


**FIGURE 18.3** Passive and active drug targeting. The schematic shows the passive and active targeting mechanisms of multifunctional nanoparticles and the difference in the vasculature of normal and tumor tissues; drugs and genes diffuse freely in and out of the normal and tumor blood vessels because of their small size, and thus the effective drug concentration in the tumor falls rapidly with time. However, macromolecular drugs and nanoparticles can passively target tumors because of the leaky vasculature or the enhanced permeability and retention (EPR) effect. However, they cannot diffuse back into the bloodstream because of their large size, leading to enhanced tumor accumulation and retention. Targeting ligands such as antibodies or peptides decorated on the surface of nanoparticles can selectively bind to cell surface receptors or antigens overexpressed by tumor or target cells and can be taken up by receptor-mediated endocytosis (active targeting). After being internalized, the drugs and small interfering RNAs (siRNAs) released from the nanoparticles within the cells can exert the necessary action.

Naked siRNAs with large molecular weight (~13 kDa) and net negative charge can neither readily enter the cells nor easily escape from the lumen of intracellular membranes. Even with direct injection into a targeted site, such as the vitreous humor of the eye, or inhalation directly into the lung, siRNAs must evade nuclease activities, penetrate the interstitial space of cell matrices, cross cell plasma membranes, escape from endosomes, and then load into the RISC complex within the cell cytoplasm to be effective [23]. Another problem is that double-stranded RNAs induce undesirable side effects, such as an inflammatory response. Thus, unprotected and unmodified native siRNA have not been used for gene silencing in directed *in vivo* applications.

Current delivery systems can be categorized into physical methods, conjugation methods, and viral and nonviral carrier-mediated methods. Physical techniques can enhance siRNA uptake at a specific tissue site by pressure or mechanical measures. Terminal-modified siRNAs can enhance their resistance to nucleases in serum and

tissue, and modification with a suitable ligand can achieve targeted delivery [21]. Nonviral carrier systems include nanoparticles, liposomes, lipoplex, polyplex, and receptor-targeted nanocomplex. These nonviral systems have shown improved siRNA encapsulation, delivery to targeted tissue, and more effective silencing with less cellular toxicity and immune responses.

This chapter illustrates in detail viral and nonviral delivery systems with an emphasis on multifunctional polymeric nanosystems for siRNA delivery.

## 18.3 DELIVERY SYSTEMS FOR RNA-INTERFERENCE

### 18.3.1 Viral Delivery Systems

In theory, RNAi can be triggered in many different ways. Most of the proposed clinical applications incorporate the double-stranded RNA molecules (siRNAs) that are typically 19 to 23 nucleotides in length and consequently have a molecular weight of 13 to 15 κDa and 38 to 46 negative charges [24]. As a result, they do not easily penetrate hydrophobic cellular membranes without assisting carriers. In addition unprotected siRNAs are generally rapidly degraded by serum nucleases. However, the intracellular entry and translocation into cytoplasm (or nucleus), where the RNAi machinery is located, is a prerequisite for gene silencing activity. This brings the necessity of the delivery vehicle, which ideally should facilitate intracellular entry into target cells and facilitate endosomal or lysosomal escape into the cytoplasm. Although it has its own hurdles, combining RNAi with viral gene therapy vectors appears to be a powerful approach when a persistent suppression of target is mandatory [24]. Viral approaches facilitate effective siRNA delivery into many difficult-to-transfect cell types, including primary cells. Designing viral RNAi vectors became feasible only after the discovery of RNAi machinery through promoter driven expression of short hairpin RNAs (shRNAs) [25]. This strategy involves cloning an oligonucleotide containing the siRNA sequence into plasmid or viral vectors to endogenously express shRNA. This should be further processed in the cytoplasm to siRNA. Because these shRNA expression cassettes are extremely small, they have the advantage of being packaged into even the smallest known viral vectors.

The use of retroviral vectors was among the first vectors used as transfer vehicles for hairpin RNA–expressing plasmids [26]. This approach involves the expression of a shRNA transcript that is processed endogenously to a siRNA. These retroviruses, by integrating their genome stably into the host cell DNA, effectively induce gene silencing and allow long-term, stable reduction of gene expression, which is critical for understanding and studying how reducing expression of disease-relevant genes cause disease phenotypes over extended period of time [24]. Because the retroviral entry and genome integration do not require viral protein synthesis, all of the viral genes in the genome can be replaced with foreign sequences. Because these viruses are likely to produce therapeutic siRNA, the viral envelope seems like a main factor capable of defining cellular tropism for this type of delivery. Tropism of these

particles can be modified through incorporation of alternative viral envelopes. Also, to maximize the safety and efficacy, retroviral RNAi vectors can be targeted to the cell types requiring therapeutic intervention.

Adenoviruses are frequently used for experimental gene therapy and had been popular for gene therapy trials because of their attractive features, including infection of both dividing and nondividing cells, high levels of transgene expression, and the ability to grow in high titers [27]. These viruses replicate in the nucleus of the mammalian cells but do not integrate into the host genome. In addition, these adenoviruses can package about 8 to 10 kb of foreign DNA. These viruses very efficiently enter the cells by a receptor-mediated pathway. The viral fiber protein binds to the cell surface receptors such as CAR (in case of group A adenovirus serotype) or CD46 (or group B serotype) followed by a secondary interaction in which the penton base protein interacts with an integrin and results in entry into the host cells [28]. These adenoviral vectors are generally a common delivery method to introduce shRNA-expression cassettes into target cells. Several groups have reported the use of adenoviral vectors for transduction of RNAi-based therapy using a hairpin-encoding adenoviral system.

Adeno-associated viral (AAV) vectors are another class of promising vectors for siRNA delivery based on several advantages over existing viral vectors. They have also been tested in clinical studies in multiple tissues [29]. They are not known to be pathogenic and have a broad range of possible target cells, including nondividing cells. Their small single-stranded DNA genomes contain only two genes, and these can be easily replaced with foreign ones, which leave the terminal inverted terminal repeats (ITRs) to allow high-level expression of the insert. Given the packaging limit of 5 kb, AAV accommodates at least eight individual shRNA expression cassettes [30,31]. Pseudotyping of AAV allows delivery of shRNA cassette to specific cells or tissues. In addition, AAV vectors show low reactivity with cellular immune responses. These are also nonintegrating and therefore pose only minimum risks of insertional mutagenesis.

Although viral vector systems seem to be a major delivery system for shRNA, a number of limitations are associated with each of them. One of the major disadvantages of adenoviral vectors is a problem of repeated administration, which triggers a strong immune response, potentially limiting the effectiveness [24]. Additionally, the accumulation of adenoviral vectors in the liver leads to severe liver toxicity. Retroviral vectors, on the other hand, also suffer from limitations such as an inability to infect nondividing cells, silencing or position effects on gene expression, and low titers. Adenoviral and AAV vectors are nonintegrating and therefore pose a minimal risk of insertional mutagenesis. But the genetic information seems to be less stably conserved and may be lost during repetitive cell division [24]. Although these vectors efficiently transduce dividing and nondividing cells, they may be inadequate for long-term gene replacement therapy.

### 18.3.2 Nonviral Delivery Systems

Although the viral vectors are very efficient delivery vehicles and have distinct advantages over plasmid DNA delivery, they can be toxic and immunogenic, can

potentially activate oncogenes, or can deactivate tumor suppressor genes. Because of these issues, the nonviral synthetic nanoparticle system has become an increasingly popular alternative for siRNA and miRNA delivery. Therefore, synthetic lipids and polymers have been developed to offer alternatives to viral vectors for nucleic acid delivery applications. These nonviral nanoparticle systems are designed and formulated very carefully to have appropriate dimensions for cell entry and endocytosis. They are also built to avoid stimulation of the immune system. Clearance of larger molecular mass lipids or polymers requires them to be biodegradable. The use of biodegradable materials would have linkages that can be cleaved inside the cell to reduce toxicity. Some of the commonly used such delivery systems include lipid-based carriers, polymer systems, peptides and proteins, and conjugates [32,33]. The most commonly and widely used lipid systems are liposomes, which been used for delivery of nucleic acids for more than 20 years. Liposomes form unilammellar or multilammelar structures. They are globular vesicle in nature with an aqueous core and phospholipid bilayer [32,33]. Because of this amphipathic nature, they efficiently incorporate wide variety of hydrophilic and hydrophobic drugs. Liposomes are created using natural body constituents such as lipids, sterols, and so on and are biocompatible and biodegradable. Cationic lipids are the most commonly used components of lipid-based siRNA delivery systems, and they consists of three major parts—a cationic head group, a lipophilic tail group, and a connecting linker. The structure of the head group, the carbon chain length of the tail group, and the nature of the linker mainly contribute to the transfection efficiency and the toxicity of the siRNA carriers. Transfection efficiency and endosome escape are known to be enhanced with a smaller hydrophilic head group and bulky alkyl side chains [33]. These lipid-based carriers are being used successfully to deliver siRNA to targets such as liver, endothelium, and solid tumors. Although these cationic liposomes are the most popular siRNA delivery agents so far, there are some concerns still remain regarding their safety for therapeutic use because certain cationic lipids have been reported to show severe cytotoxicity *in vitro* and *in vivo* [34]. Neutral leptosomes are another class of liposomes that have already shown good safety and efficacy records in humans. Once such example that received Food and Drug Administration (FDA) approval for human use is Doxil [32]. The poly(ethylene) glycol (PEG)–modified long-circulating liposomes are clinically approved delivery system for doxorubicin and therefore represent a viable option for delivering siRNA in humans [35]. Although this has been currently used for treatment of number of different cancer indications, its further development was limited after cationic liposome development. Although cationic liposomes have been used to efficiently deliver siRNA *in vitro* and *in vivo*, with the realization that significant toxicities associated with cationic charged molecules, neutral liposomes have reemerged as promising carriers for siRNA. However, the initial incorporation of siRNA and stability seem to still be problematic in these neutral systems, and these limitations are currently being addressed by researchers using novel strategies. In addition, it has been reported that with repeat dosing regimens of PEG-modified liposomes, the antibodies specific for PEG-modified particles have been generated, which will end up compromising the efficacy in the long run [36,37].

Conjugating small molecules, peptides, proteins, polymers, and aptamers to siRNA comprise another class of carriers and are known to improve the stability, cellular internalization, or cell-specific active targeting delivery [32,35,38]. Cholesterol-conjugated siRNA, upon systemic administration, silenced the expression of an endogenous gene encoding apolipoprotein (apo) B in the liver and jejunum and reduced the plasma apo B levels as well as the total cholesterol level, but the naked siRNA did not show any activity in these tissues [39]. However, it has been demonstrated that these conjugates need to be administered at very high doses to be active in mice. Surprisingly, no off-target effect or immune stimulation was reported despite of the higher doses used. However, an extensive cell surface binding of lipophilic oligonucleotide conjugates has been reported to interfere nonspecifically with normal cellular processes. This actually necessitates a detailed safety evaluation of this approach. Endosomolytic agent (amphipathic poly[vinyl] ether)–modified siRNA, upon systemic administration, effectively reduced the expression levels of two endogenous genes in mouse liver [40]. The endosomolytic agent is known to enhance the endosomal escape after endocytosis. Conjugating antibody is another way of increasing the specificity of siRNA delivery and minimize the off-target effects. One such example was the delivery of F105-P–siRNA conjugate to HIV-infected cells [41]. This did not trigger the interferon responses but reduced the HIV replication in a difficult-to-transfect HIV infected CD4 T cells. One more way of increasing the cell-specific delivery is by linking siRNA to an RNA aptamer or aptamer–siRNA chimeras. Upon intravenous injections, PEG-modified aptamer chimeras resulted in a significant regression of prostate-specific membrane antigen (PSMA)–expressing tumors [42]. Although the studies suggested the potential use of aptamers for targeted delivery of siRNA drugs, nuclease susceptibility of aptamer ligands may limit their application in systemic treatments, and the possibility of premature degradation of siRNA in the endosomal compartment because of the lack of endosomal escape machinery may require additional functionality.

Cationic polymers also readily bind and condense large nucleic acids and serve as efficient transfection agents. Because of this cationic nature, they were also shown to stimulate endocytosis and endosome escape. These polymers are widely divided into two classes, natural polymers (chitosan, atelocollagen, and cyclodextrin) and synthetic polymers such as polyethylenimine (PEI), poly-L-lysine (PLL), and dendrimers [33].

Cyclodextrin, a natural polymer, has been developed as a delivery vehicle to efficiently deliver siRNA to tumors [43]. β-Cyclodextrin–containing polycations have been self-assembled with adamentane–PEG5000 and transferrin ligand to target transferrin receptors on the cell surface. These nanoparticles seemed to be stable in the bloodstream. Unlike the cationic liposomes, transferrin targeted cyclodextrin polymer particles did not illicit any immunostimulatory effects in mouse models. They also demonstrated good safety profile in nonhuman primates [44]. Not only did they demonstrate good efficacy in number of different preclinical models, but they also showed activity and efficient gene silencing in human cancer patients in clinic [45]. PEI is a synthetic polymer, and it has been broadly investigated for nucleic acid delivery. PEG-modified PEI with arginine–glycine–aspartic acid (RGD) ligand has been extensively studied by Temming et al. [46], for siRNA delivery to target

angiogenesis [46]. Because the commonly used branched PEI is generally known to cause severe toxicity, the use of higher molecular weight unmodified PEI is likely to be limited. Several groups have been successful in minimizing the toxicity by using low-molecular-weight or biodegradable PEI, although this favors the plasmid delivery rather than the siRNA delivery in mouse models. Another approach was also initiated to synthesize a ketalized PEI carrier. Branches of PEI were modified with acid-degradable amine groups through ketal linkages [47]. These bonds rapidly break at a low pH to release the siRNA. Hybrid of these approaches suggests the possibility of a nontoxic PEI carrier to deliver siRNA. As an alternative to PEI carriers, poly(lactic-*co*-glycolic acid) (PLGA) is another extensively studied copolymer for delivery. Compared with cationic polymers and lipids, it has the advantage of lower toxicity. Tertiary amine modified polyvinyl alcohol backbones grafted to PLGA yielded rapid degradation properties for siRNA release. Incorporation of PEI into PLGA nanoparticle further improved the loading and activity. Dendrimers are another class of synthetic, cationic macromolecules that are being used as a carrier for small molecular drugs and large biomolecules [33]. Although the most advanced delivery system at present includes a cationic group for ionic interaction with siRNA (PEG for steric hindrance, an endosomolytic group for endosome escape, and a targeting ligand for specific delivery), several researchers have been working to come up with a noncondensing lipid polymer agent as an siRNA or miRNA carrier to minimize the toxicity effects coming from the cationic charged molecules.

## 18.4 MULTIFUNCTIONAL NANOSYSTEMS FOR DELIVERY OF RNA-INTERFERENCE THERAPEUTICS

### 18.4.1 Nanoparticle-Based Targeted Delivery

Encapsulation of oligonucleotides and siRNAs in nanoparticles can offer several major advantages over conventional modes of delivery in the management of diseases such as cancer [48,49]. For instance, apart from protecting the payload from degradation in the bloodstream, nanoparticles and liposomal delivery systems can ferry the drugs and oligonucleotides selectively to solid tumor tissues based on a phenomenon called the EPR effect discovered by Matsumura and Maeda three decades ago [50].

Tumor cells divide and multiply at phenomenal rates and posses devastating invasive potentials. When the tumor lesion reaches approximately 1 to 2 mm in diameter, the tumor starts to form "neovasculatures" by a process called angiogenesis or development of new blood vessels, which starts to cater to the nutritional and oxygen demands of the growing tumor [51]. The tumor blood vessels are also found to be aberrant and "leaky." Furthermore, tumor tissues often lack a lymphatic drainage system [50,52,53]. The tumor microenvironment is thus highly disorganized in architecture [54]. Also, most solid tumors have excessive production of vascular permeability mediators such as vascular endothelial growth factor (VEGF),

bradykinin, prostaglandins, matrix metalloproteinases (MMPs), nitric oxide, and peroxynitrite [55–58]. The elevated levels of the vascular permeability mediators coupled with the anatomical and pathophysiological abnormalities of the tumor vasculature leads to extensive extravasation of blood plasma components, macromolecules, and nanoparticles into the tumor interstitium [59]. It is thus possible to achieve "passive tumor targeting" and a very high local concentration of the nanoparticles in the tumor tissues. In general, drugs and genes encapsulated in macromolecular delivery systems such as polymer conjugates, liposomes, polymeric nanoparticles, and micelles are thus able to exploit this unique phenomenon to accumulate selectively in solid tumor tissues [60–64].

The EPR effect has been found to be more pronounced if the nanoparticulate delivery systems are designed to evade the mononuclear phagocytic system (MPS), thus prolonging their circulation half-life in the blood. In this regard, grafting of amphiphilic polymers such as PEGs to the nanoparticles has been found to be very effective in providing stealth characteristics that enable MPS escape, thereby rendering a long plasma circulation time [65] (Fig. 18.3).

As described earlier, passive targeting itself can be a highly effective mechanism for delivering nanoparticles loaded with oligonucleotides and siRNAs for moderate and highly vascularized solid tumors; however, their delivery to avascular, hypovascular, or necrotic regions of tumors still remains a great challenge [66]. Also, passive targeting using the EPR effect does not apply for diseases other than solid tumors and some cases of inflammatory disorders. In this regard, "active ligand-based targeting" using antibodies or peptides that can selectively bind to receptors or antigens overexpressed on the surface of diseased cells and tumors can be more promising [67] (see Fig. 18.3). Furthermore, the delivery efficiency and target specificity of passively targeted nanosystems can be improved significantly when tumor targeting agents are also made part of the nanodelivery systems [68]. In our group, we have developed one such nanosystem that utilizes epidermal growth factor receptor (EGFR) peptide conjugated nanoparticles for active targeting to multidrug-resistant ovarian cancer cells that overexpress EGFR [69,70]. In another example of active targeting, RGD tripeptide was coupled to nanoparticles to target $\alpha_5\beta_5$ or $\alpha_5\beta_3$ integrin receptors that are overexpressed on vascular endothelial cells of angiogenic blood vessels of proliferating tumors [71]. Werner et al. [72]. have used folate-based active targeting to deliver nanoparticles to ovarian cancer that overexpress folate receptors. Such particles are internalized by active transport via receptor-mediated endocytosis [73]. Nucleic acid constructs such as aptamers that can selectively bind to PSMA on prostate cancer cells are also being explored as active targeting ligands for diseases such as prostate cancers [74].

Active targeting using antibodies [75] and peptides coupled to nanoparticles thus in effect provides a more selective secondary targeting via binding or intracellular trafficking of the siRNA–drug payload to the target cells [67,76] after the primary (passive) targeting based on the EPR effect (see Fig. 18.3). Such multifunctional smart delivery systems are currently intensely pursued and have the potential to revolutionize nanoparticle-based RNAi therapeutics development [77].

### 18.4.2 Advantages of Multifunctionality

The aim of developing a multifunctional nanoparticulate delivery system is to achieve several goals using a single "nanoconstruct." For instance, a nanoparticle delivery system can be designed with a low, moderate, or high level of complexity depending on the requirements of the target disease or function in the body such as:

1. Protection of the payload (e.g., siRNA) from degradation on systemic delivery and elimination of off-target effects associated with therapeutic agents [78]
2. Enabling long circulation half-life *in vivo* by surface functionalization of nanoparticles with amphiphilic molecules such as PEG that can help evade the immune system and reduce clearance by the MPS [79,80]
3. Taking advantage of passive tumor targeting strategies based on the EPR effect [81]
4. Loading or encapsulating the nanoparticles with single (and often large doses of) therapeutic agent or co-deliver multiple agents such as siRNA and drugs that works synergistically to have a compounded effect [82,83]. In this regard, polymeric nanoparticle–based delivery also offers the flexibility to devise the system to have spatial and temporal control or controlled erosion characteristics based on the nature and choice of the biodegradable polymer used for constructing the nanosystems [84–86]. For example, the delivery vehicle can be engineered to release one drug or oligonucleotide concomitantly with another agent or, if desired, after a time lapse or in a controlled fashion based on the requirement of the disease target [84]. In this regard, in our laboratory, we are developing a combinatorial-designed library of novel biodegradable and biocompatible polymers that can self-assemble in the presence of oligonucleotides and drugs to form stable nanoparticles as shown in Figure 18.4 [87,88]. The advantage of such systems is the diverse choice of polymers with varying functionality that could be screened simultaneously based on a "mix and match" type screening to arrive at the right choice of materials that can be tailored for specific disease targeting [87,88].
5. The nanoparticles could also be surface decorated with targeting ligands such as antibodies or peptides that can help locate, bind, or internalize the nanoparticles into the diseased cells or can be incorporated with agents that can trigger the release of the payload (drugs or genes) in specific location or facilitate the drug escape from organelles (e.g., lysosomes) within the cells [63,89–92].
6. Targeted and nontargeted nanoparticles can also be designed to aid intracellular distribution of the payload that can be critical for many drugs and genes whose sites of action are generally located in the cytoplasm or the nucleus [93]. For example, plasmid DNA, oligonucleotides, miRNA, siRNA, peptides, and proteins have their targets located in the cytoplasm or nucleus. Thus, there is a need for these molecules to be transported inside the cell [93,94]. More importantly, intracellular trafficking of drugs and genes is more challenging for refractory diseases such as cancers that are equipped with

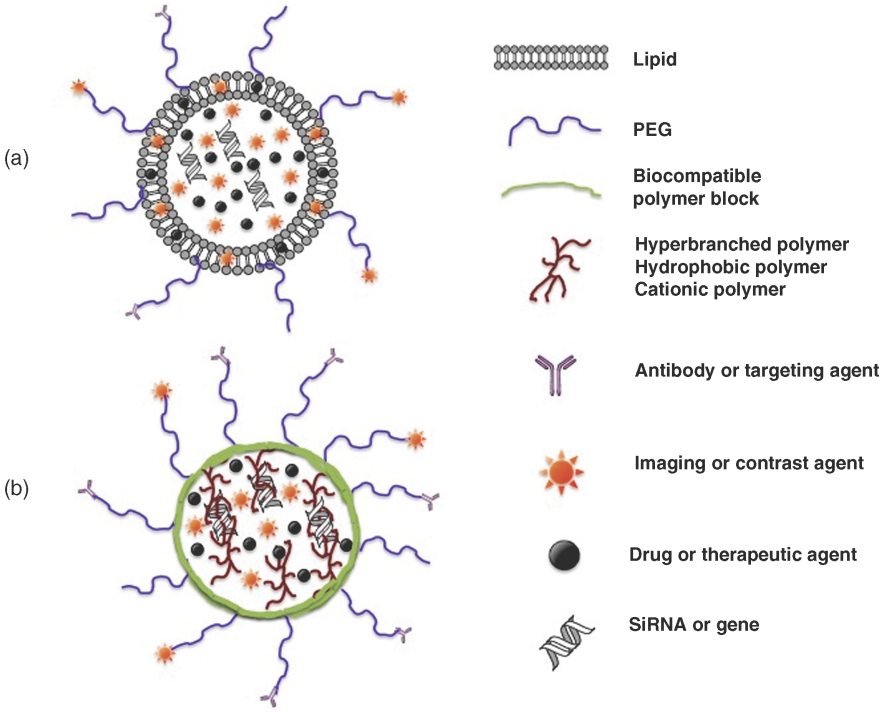


**FIGURE 18.4** Multifunctional nanosytems for targeted therapeutics. Schematic representation of a multifunctional targeted immunoliposome (a) and a multifunctional self-assembled targeted nanoparticle (b) that can be formulated from their corresponding building blocks. The multifunctional nanosystems can be engineered by self-assembly of polymers or lipids containing various functionalities such as amphiphilic block copolymers to facilitate formation of stable nanoparticles loaded with drugs, genes, and imaging agents; antibody and peptides for targeting; and surface functionalized with amphiphilic poly(ethylene) glycol (PEG) chains for enabling long circulation half-life *in vivo*. siRNA: small interfering RNA.

active drug efflux pumps such as P-glycoprotein and multidrug-resistant proteins that are responsible for developing multidrug resistance (MDR) [95]. In such cases, nanoparticle-based delivery systems have shown to evade the drug efflux pathway, thus increasing the intracellular delivery efficiency of drugs and genes [96,97].

7. Nanoparticles can be tagged with imaging agents such as optical dyes, radiotracers, and contrast agents for simultaneous disease detection or visualization and treatment [98,99]. Such multifunctional delivery systems (see Fig. 18.4), have the advantage of performing multiple tasks such as visualizing the sites of localization of nanoparticles, monitoring the *in vivo* fate, efficiently delivering the payload and measuring dose response in patients [100]. Such functions can thus be performed noninvasively and recorded simultaneously in real time using a single nanoconstruct.

## 18.5 ILLUSTRATIVE EXAMPLES OF NANOPARTICLE-MEDIATED SMALL INTERFERING RNA DELIVERY

### 18.5.1 Small Interfering RNA Delivery in Cancer

As discussed in prior sections, gene therapy based on the RNAi mechanism has been well established and has currently become a major area of research that holds great promise for the management of diseases such as cancers. Although siRNAs have high efficiency and specificity, the major hurdles that still remain are its delivery to target tumor cells and intracellular trafficking after localization in the sites of interest [101].

Researchers from Park's group have developed a polyelectrolyte complex micelle-based VEGF siRNA delivery system for antiangiogenic gene therapy [102]. The complexation of the PEG-conjugated VEGF siRNA was achieved by charge interaction with PEI. This complex formed a stable core-shell nanostructure with the PEI/siRNA forming the central core and PEG chains forming the corona. Intravenous (IV) and intratumoral injection of the micelles in mice demonstrated significant inhibition of VEGF expression in the tumor tissue and suppressed tumor growth without detectable toxicities [102]. Furthermore, to confirm the feasibility of siRNA-based gene therapy and imaging *in vivo*, optical imaging was undertaken using a Cy5.5-labeled siRNA. The Cy5.5-labeled siRNA containing polyelectrolyte complex micelles predominantly accumulated in the tumor, affirming the feasibility of this approach [102].

To overcome the dose-limiting side effects of conventional chemotherapeutic agents and the therapeutic failure caused by MDR, we designed and evaluated a novel biocompatible lipid-modified dextran-based self-assembled polymeric nanosystem that could encapsulate MDR-1 siRNA as well as anticancer drugs such as doxorubicin [49,103]. Furthermore, to image the transfection efficacy of siRNA-loaded nanoparticles on cell lines, we used an EGFP (green fluorescence protein) expressing BHK-21 cells and performed confocal fluorescence microscopy to visualize the downregulation of EGFP. It was observed that the EGFP siRNA was efficiently incorporated into cells and effectively inhibited the expression of EGFP in a dose-dependent manner. Similarly, the MDR1 siRNA-loaded dextran nanoparticles efficiently suppressed P-gp expression in the drug-resistant osteosarcoma cell lines. A combination therapy of the MDR1 siRNA-loaded nanocarriers with doxorubicin revealed pronounced increase in doxorubicin uptake in the nucleus (assessed by confocal fluorescence microscopy imaging) of multidrug-resistant cells. These results demonstrate that our approach may be useful for reversing drug resistance by increasing the amount of drug accumulation in MDR cells. Thus, delivery using nontoxic and biocompatible dextran-based nanosystems offers a versatile platform for incorporation of multiple payloads for simultaneous imaging and delivery of therapeutic molecules that can be translatable for human clinical applications.

In another study, Huh et al. [104] designed a siRNA delivery system containing glycol chitosan biodegradable and biocompatible polymer (GC) and PEI. The

polymers were conjugated with 5β-cholanic acid (CA) to stabilize the nanoparticles and endow them with tumor-homing ability. The nanoparticles were formed by mixing GC-CA and PEI-CA to form self-assembled nanostructures (GC–PEI nanoparticles) caused by the strong hydrophobic interactions of 5β-cholanic acids in the polymers. The cationic charge on the GC–PEI nanoparticles was used to complex with the negatively charged red fluorescence protein (RFP) gene silencing siRNA designed to inhibit RFP expression. *In vitro* studies with RFP expressing B16F10 tumor cell incubated with siRNA–GC–PEI nanoparticles revealed time-dependent cellular uptake of the nanoparticles and led to specific mRNA knock down. More importantly, the siRNA-loaded nanoparticles displayed significant inhibition of RFP gene expression in RFP/B16F10-bearing mice models, demonstrating their potential as a promising vector for siRNA delivery [104].

Medarova et al. have developed a dual-purpose probe for *in vivo* delivery and simultaneous imaging of siRNA accumulation in tumors by using high-resolution magnetic resonance imaging (MRI) and near-infrared (NIR) *in vivo* optical imaging techniques [105]. The siRNA was covalently conjugated to nanoparticles consisting of NIR dye conjugated magnetic nanoparticles. Furthermore, the nanoparticles were decorated with specific membrane translocation peptide that enables their intracellular delivery and cytosolic availability [105]. Taken together, the study demonstrates the feasibility of tracking the tumor uptake and silencing ability of the siRNA conjugated nanoparticles *in vivo* [105].

In another study, Kumar et al. [106] used superparamagnetic iron oxide nanoparticles as a construct for delivery of siRNA to solid tumors and for combined optical and MR imaging. The nanoconstruct, apart from performing its function as a vehicle for siRNA delivery, also served as a tool to study basic tumor biology and therapy [106]. The nanoconstruct consisted of magnetic nanoparticles (for MRI), a Cy3 fluoropore (for optical imaging), a targeting peptide (EPPT) that could home to a specific antigen (uMUC-1) overexpressed by majority of human breast cancer cells, and a synthetic siRNA that downregulates the *BIRC5* antiapoptotic gene in breast cancer cells [106]. The nanoconstruct demonstrated specific tumor uptake and significant downregulation of the *BIRC5* gene in animal tumor models. More importantly, the tumor uptake could be visualized both by MRI and NIR optical imaging [106].

### 18.5.2 Small Interfering RNA Delivery in Inflammatory Diseases

Nanoparticle- and microparticle-based oral gene silencing using siRNA-encapsulated polymeric systems has been very promising for inflammatory disorders. In a recent study in our group, we evaluated the downregulation of tumor necrosis factor α (TNF-α) by oral RNAi therapy. Control (scrambled sequence) or TNF-α–specific siRNA were encapsulated in type B gelatin nanoparticles and further entrapped in poly (epsilon-caprolactone) (PCL) microspheres to form a nanoparticles-in-microsphere oral system (NiMOS). It was observed that successful gene silencing led to decreased colonic levels of TNF-α, suppressed expression of other proinflammatory cytokines (e.g., interleukin [IL]-1β, interferon [IFN]-γ) and chemokines (MCP-1), an increase in body weight, and reduced tissue myeloperoxidase activity in murine models. Results

of this study established the clinical potential of a NiMOS-based oral TNF-α gene silencing system for the treatment of inflammatory bowel disease as demonstrated in an acute colitis model [107].

A cell cycle regulatory molecule, cyclin D1 (CyD1), is a well-known therapeutic target for cancer. It is also strongly overexpressed at the sites of inflammation. To study the role of cyclin CyD1 in inflammation, Peer et al. [108] devised a strategy to selectively silence CyD1 in leukocytes *in vivo*. A targeted nanoparticle system (tsNPs) was loaded with CyD1–siRNA. Then a specific antibody to β-7 integrin (β-7 I) was used to target specific subset of leukocyte involved in gut inflammation. It was observed that the systemic administration of β-7 I–tsNPs silenced the CyD1 in leukocytes and reversed experimentally induced colitis in animal models by suppressing the leukocyte production and T-helper cell 1 cytokine expression. The study revealed that CyD1 could be an antiinflammatory target, and the use of similar targeted nanoparticle-loaded siRNAs could be explored for other therapeutic settings [108].

siRNAs have also shown major implications for the management of diseases such intestinal inflammation. Wilson et al. has shown TNF-α silencing using siRNAs loaded in thioketal nanoparticles (TKNs) for targeted inhibition of gene expression in the intestines of murine models with ulcerative colitis [109]. In this case, protection of the siRNA during delivery and its targeting to the intestinal tissues was found to be critical because the systemic depletion of cytokines could cause adverse side effects. Although various delivery vehicles have been developed to orally deliver therapeutics to intestinal tissue, none of these strategies has demonstrated the ability to protect siRNAs from the harsh environment of the gastrointestinal tract and target its delivery to inflamed intestinal tissue. In this regard, the delivery vehicle developed by Wilson et al. [109], termed TKNs, could direct orally delivered siRNA to sites of intestinal inflammation and thus inhibit gene expression in inflamed intestinal tissue. TKNs were made from a polymer, poly-(1,4-phenyleneacetone dimethylene thioketal), that could degrade selectively in response to reactive oxygen species (ROS). Therefore, when delivered orally, TKNs could trigger the release of siRNA in response to the abnormally high levels of ROS specific to sites of intestinal inflammation. The study thus provides concrete evidence that orally administered TKNs loaded with siRNA against the proinflammatory cytokine TNF-α could diminish TNF-α messenger RNA levels in the colon and protect mice from ulcerative colitis [109].

To maintain the integrity of blood–brain barrier (BBB), Bonoiu et al. [110] have evaluated MMP-9 gene silencing using a quantum dot (QD)–siRNA nanoplex delivery. MMPs, particularly MMP-9, are found to be involved in the neuroinflammation processes leading to disruption of the BBB, worsening the condition of neurologic diseases such as HIV-1 AIDS, dementia, and cerebral ischemia. In this regard, nonviral vectors such as nanoparticles have been proposed to have great potentials for their treatment. Bonoiu et al. [110] found that QDs complexed with MMP-9-siRNAs formed a stable nanoplex that could be used for studying the specificity and delivery efficiency for downregulating the expression of MMP-9 gene in brain microvascular endothelial cells (BMVEC), which constituted the BBB. It is hypothesized that silencing MMP-9 gene expression in BMVECs and leukocytes

may help prevent breakdown of the BBB and in turn inhibit the subsequent invasion of central nervous system (CNS) by infected and inflammatory cells. Indeed, it was observed that silencing of MMP-9 gene expression resulted in decrease in endothelial permeability and upregulation of ECM proteins such as collagen I, IV, and V as a result of reduction of transendothelial resistance across the BBB in the *in vit*ro BBB model [110]. MMP-9 gene silencing also resulted in an increased expression of the gene tissue inhibitor of metalloproteinase-1 (TIMP-1). These results corroborate the existence of a delicate balance between the MMP-9 levels and its natural inhibitor TIMP-1 and its role in maintaining the basement membrane integrity. The study by Bonoiu et al. [110] reinforces the role of novel nanoparticle-based siRNA delivery systems in modulating the MMP-9 activity in MMP-9–producing cells that may prevent neuroinflammation and preserve the integrity of the BBB [110].

### 18.5.3 Small Interfering RNA Delivery in Other Diseases

The RNAi mechanism offers the potential for new and effective therapies for a wide variety of diseases other than cancers, such as rheumatoid arthritis. It is well known that TNF-α plays a key role in the progression of rheumatoid arthritis and is an important target for antirheumatic therapies [111]. TNF-α expression can be silenced with siRNA, but its efficacy depends on efficient and safe siRNA delivery vectors. Jensen et al. [112] aimed to identify polymeric nanocarriers for anti–TNF-α siRNA with optimal efficacy and minimal side effects *in vitro*. It was observed that dextran nanogels and polyamidoamine (PAMAM) dendrimers of generations 4 and 7 mediated concentration-dependent gene silencing with minor toxicity [112]. These systems are thus expected to be suitable for systemic delivery of siRNA.

Although the clinical development of new RNAi-based technology for treating diseases such as pulmonary disorders is still limited, there is a wide scope for polymeric and liposomal delivery systems for treatment of such diseases. Sparks et al. [113] have recently reported the development of a structurally versatile class of cationic lipopolyamines designed particularly for siRNA delivery, which showed high levels of target transcript knockdown in a range of cell types *in vitro*. A primary benefit of their lipid system was the ease with which they could be covalently modified with functional molecules. On systemic administration in animal models, the nanoparticles prepared from stearylamine lipid modified with methoxy-polyethyleneglycol (m-PEG) containing a siRNA targeting the caveolin-1 (Cav-1) transcript caused up to 60% reduction of the Cav-1 transcript, specifically in lung tissue after 48 hours compared with treatment with nonsilencing siRNA [113]. Moreover, the modification of the nanoparticles with mPEG showed reduced toxicity associated with IV administration, demonstrating the potential use of using lipid-based RNAi nanotherapeutics for treating pulmonary disease [113].

In 2009, Woodrow et al. [114] were the first to report the use of biodegradable polymeric nanoparticles as safe and effective delivery vehicles for siRNA-induced gene silencing in the genital tract. Although siRNA delivery using conventional liposome-based systems have been useful for protection against infectious disease and gene silencing in the genital tract, commercially available lipids are found to be

toxic to the epithelial mucosa. In their alternative approach, Woodrow et al. [114] used nanoparticles derived entirely from FDA-approved materials. A bolus high dose of siRNA-loaded nanoparticles administered topically to the mouse female reproductive tract resulted in efficient and prolonged gene silencing [114]. It was observed that the knockdown of gene expression was seen not only in the vaginal lumen but also in the uterine horns that are located at distal sites from the topically administered area. In addition, the nanoparticles could penetrate deep into the epithelial tissue, demonstrating their pronounced gene silencing effect [114].

Zhang et al. [115] have reported that siRNA targeting the NS1 gene delivered using nanoparticles may provide an effective inhibition of respiratory syncytial virus (RSV) infection in humans. The role of NS1 gene in modulating RSV infection was studied using a plasmid-borne siRNA targeting the NS1 gene (siNS1) [115]. It was observed that RSV replication was reduced in A549 cells, but not interferon-deficient Vero cells, transfected with siNS1. siNS1-induced upregulated expression of interferon and interferon-inducible genes in A549 cells. It was also reported that there was overproduction of type 1 interferon and induced differentiation of naive $CD4^+$ T cells to T-helper type 1 (TH1) cells in siNS1-transfected human dendritic cells upon RSV infection [115]. Intranasal administration of siNS1 nanoparticles before or after RSV infection showed remarkably decreased virus titers in the lung and decreased inflammation and airway reactivity compared with controls in mice models [115]. This study demonstrated that inhibition of RSV infection using siRNA nanoparticles targeting the viral NS1 gene could be achieved *in vivo* [115].

It is well known that HIV infection of the CNS is the principle cause of HIV-associated dementia in adults and encephalopathy in children. In this regard, therapies using siRNA encapsulated in nanoparticles may facilitate transport across the BBB. Researchers from the Muñoz–Fernández group have used the 2G-NN16 amino-terminated carbosilane dendrimer as a method for delivering siRNA to HIV-infected human astrocytes [116]. It was reported that a gradual time-controlled degradation of the 2G-NN16 dendrimer resulted in the liberation of its siRNA cargo between 12 and 24 h [116]. Furthermore, the siRNA/2G-NN16 dendriplexes were able to transfect human astrocytes successfully even after crossing the BBB in an *in vitro* model. More importantly, the transfected siRNA was observed to exert a biologic effect because the dendriplexes exhibited downregulation of the housekeeping gene GAPDH and reduced the replication of HIV-1 strains X4-HIV NL4-3 and R5-HIV BaL in human astrocytes [116]. These results demonstrate that dendrimers could be successfully used for siRNA gene silencing without causing apparent toxicities.

Huang and King [117] have developed an immobilized siRNA nanoparticle-based platform technology for efficient gene silencing in targeted cells in the circulation especially for diseases derived from disorders of blood cells. To create targeted liposome-loaded siRNA delivery systems, P-selectin was covalently attached to the surface of PEG-modified liposomes. The targeted liposomes were found to effectively absorb onto the inner surface of micro-renathane tubing. The liposome-coated surfaces could not only capture target cells from physiological shear flow and efficiently deliver the encapsulated siRNA into adherent cells but also

exhibit pronounced silencing of the targeted gene neutrophil elastase [117]. The device developed by Huang and King could further create a high and localized concentration for siRNA delivery in the circulatory system, providing circulating target cells (e.g., stem cells) adequate time to interact with therapeutic materials [117].

## 18.6 ILLUSTRATIVE EXAMPLES OF NANOPARTICLE-MEDIATED MICRO RNA DELIVERY

### 18.6.1 miRNA and Antagomir Delivery in Cancer

When oncogenic miRNAs are upregulated, they block the tumor suppressor genes and lead to tumor formation. On the other hand, when there is downregulation or loss of miRNAs with suppressor functions, they may increase the translation of oncogenes and hence formation of tumors [118–120]. One of the strategies to target overexpressed miRNA expression in cancer is the use of oligonucleotides or antagomirs to block the expression. Antagomirs are a novel class of chemically modified stable oligonucleotides consisting of 20 to 24 bases that are used to block the functions of endogenous miRNAs. The other one involves the use of miRNA mimics to substitute for the loss of expression of a tumor suppressor miRNA. The key challenge in achieving effective miRNA therapeutics in those cases is the development of an efficient delivery system that can specifically deliver antagomirs or miRNA mimics to target cells in living animals. To overcome these delivery hurdles, nonviral and viral strategies have been explored. The nonviral strategies include encapsulation of chemically modified oligonucleotides in liposomes or polymers [118]. Liposomes are composed of a phospholipid bilayer with an enclosed aqueous compartment. During complexation, they interact with oligonucleotides to form complexes that are stabilized by electrostatic interactions. These liposomes protect the oligonucleotides from degradation by nucleases and increase the circulation half-life.

Let-7 was one of the best studied miRNAs and was found to be altered in the majority of human lung cancers [121,122]. Bader and his group [121] identified this tumor suppressor miRNA in majority of non–small cell lung cancer (NSCLC) cells and patient samples and used the concept to reintroduce this miRNA mimics to cells and tumors to reactivate cellular pathways to drive a therapeutic response. Studies have shown that the reduced let-7 was significantly associated with shortened postoperative survival times and that overexpression resulted in the inhibition of lung cancer cell growth [123]. Using a neutral lipid delivery system, Bader et al. efficiently delivered the miRNA mimic to cells and demonstrated significant growth inhibition and proliferation [123]. They also demonstrated lung tumor growth inhibition in NSCLC tumors (H460) that lack the expression of let-7 using the same delivery system [123]. In addition, it was demonstrated that let-7 works against the oncogene K-ras that has crucial role in proliferation of lung cancer cells, especially in the cells with K-ras mutations. Given this information, this group

also demonstrated significant tumor growth inhibition and antitumor efficacy in a genetically engineered K-ras mutated oncogenic mouse model, a model that very closely resembles actual human lung cancer [124].

EGFR overexpression is often found in majority of the NSCLC samples. Although the EGFR-tyrosine kinase inhibitor has a striking effect in NSCLC with longer survival rates, the efficacy is attributed to the development of EGFR-TKI–acquired resistance because of additional mutations. Thus, a method to suppress the restored EGFR pathway is needed to overcome the resistance. Rai et al. have developed an approach to deliver miRNA (miRNA7) that targets multiple sites in the 3′ UTR of EGFR mRNA to suppress the EGFR expression [125]. Compared with the conventional targeting therapies such as antibodies and inhibitors, the miRNA strategy seems to be independent of unexpected conformational changes caused by secondary mutations. They used a cationic liposome system to deliver miRNA7 expressing plasmid and demonstrated significant inhibition of cell growth in number of different lung cancer cells. Delivery of miRNA7 also resulted marked tumor regression and EGFR suppression in mouse xenograft lung models, which ultimately overcomes the EGFR-tyrosine kinase inhibitor resistance in lung cancer cells and tumors [125]. Another group has developed a transferrin targeted, protamine containing liposomes, to deliver miRNA. miR29b is a member of miR29 family, which downregulates the cellular expression of antiapoptotic Mcl-1 proteins. The formulation was designated as Tf-LPmR, has protamine to increase the delivery efficiency and transferrin ligand for receptor-mediated internalization in leukemia cells (K562) that overexpress transferrin receptors. It was demonstrated that this Tfr-targeted system containing miR29b liposome formulation resulted enhanced biological effects on the suppression of target gene's expression (efficient inhibition of Mcl-1 expression at the mRNA and protein levels) compared with the nontargeted or nonprotamine-containing liposomes [125]. Huang's group demonstrated efficient miRNA delivery using a tumor-targeted delivery system into experimental lung metastasis of murine B16F10 melanoma. They developed a liposome–polycation–hyaluronic acid nanoparticle system modified with a tumor targeting single chain antibody fragment (scFv) for systemic delivery [126]. Delivery of miR34a using this system demonstrated apoptosis, inhibition of survivin expression, and downregulation of the mitogen-activated protein kinase pathway in B16F10 cells. Systemic delivery of miR34a using this system significantly downregulated survivin expression in the metastatic tumor and reduced tumor load in the lung. Apart from using liposome systems, a number of groups have also worked on a polymer delivery system for the same purpose. One such example was using PLGA-based nanoparticles. PLGA was extensively studied for its ability to deliver different therapeutic agents. This polymer-based particles were shown to escape from the endosomal compartment to the cytoplasmic compartment and release their contents over longer periods of time. These features rendered PLGA particles as potential tool for efficient oligonucleotide or miRNA delivery. A group from China has modified the biodegradable PLGA nanoparticles with a polyplexed PEI coating in which the PLGA acts as the core and cationic PEI as the shell with an aim of encapsulating negatively charged nucleic acids. With the help of cationic shell, they efficiently encapsulated

miR-26a, a cell cycle suppressor, and demonstrated efficient delivery to human hepatocellular carcinoma cells such as HepG2 cells [127]. This significantly increased the expression levels of miR-26a and inhibited the cell cycle progression by induction of G1 phase arrest in transfected HepG2 cells. These PLGA nanoparticles also showed a better safety profile than PEIs and liposomes. In another instance, Sekar et al. [128] encapsulated chemically modified antagomirs in PLGA nanoparticles and showed efficient internalization and sustained release into cells. They also demonstrated tumor regression with the delivery of antagomirs. Their study results indicated that antagomir delivery is not only efficient in crossing cell membranes but also maintained the threshold of intracellular antagomir levels by its slow release profile for prolonged therapeutic effect.

A highly branched dendritic polymer including PAMAM has attracted interest as a nucleic acid delivery vector [129]. The presence of primary amine groups on their branched surface bind nucleic acids, compact it into polyplexes, and promote cellular uptake in wide variety of cultured cells. Primary amines on the surface also make it possible to conjugate suitable ligands such as transferrin for tumor or brain delivery. Overexpression of miR21 plays a key role in majority of cancers. Downregulation of miR21 therefore leads to repression of cell growth, increased cellular apoptosis, and cell cycle arrest, which can theoretically enhance the chemotherapeutic effect in cancer. Ren et al. [130] used this PAMAM dendrimer polymer as a carrier to co-deliver anti-miR21 and 5-fluorouracil to achieve delivery of miRNA to human glioblastoma cells and enhance the cytotoxicity of 5-fluorouracil (5-FU). With the help of positively charged primary amine groups, the miRNA gets encapsulated easily. Because of the presence of well-defined cavities and open architecture in PAMAM, guest molecules such as 5-FU get encapsulated into the macromolecular interior through hydrophobic interactions. Through their charge-based interactions, 5-FU-PAMAM could conjugate with miR21 [130]. With this system, the authors demonstrated significant improvement in cytotoxicity of 5-FU and a dramatic increase in the apoptotic percentage of U251 cells. They also showed that this co-delivery brought down the migration ability of tumor cells, suggesting that the co-delivery system may have an important clinical application in the treatment of miR21-overexpressing glioblastoma.

### 18.6.2 miRNA and Antagomir Delivery in Other Diseases

miRNA and antagomir delivery is also being used in number of diseases other than cancer. The development of antisense oligonucleotide-mediated (antimiR) knock-down and miRNA-mimic mediated overexpression are the techniques that are being followed. Researchers from University of California San Diego (UCSD) identified a new way to regulate the uncontrolled growth of blood vessels [131]. This uncontrolled growth often leads to vascular malformations and hemangiomas, which may become life threatening. The researchers found the switch mechanism that converts the normal blood vessels from the resting state to proliferative state and found a small miRNA (miR132) responsible for controlling the switch. They identified a complementary miRNA that binds to and neutralizes the original miRNA, thereby turning

**TABLE 18.1 MicroRNA Therapeutics in Development**

| miRNA Drug | Indication | Status | Company |
|---|---|---|---|
| let-7 mimic | Lung cancer | Preclinical | Mirna |
| miR-15 | Post-MI remodel | Preclinical | MiRagen, Inc. |
| miR-21 antagomir | Fibrosis | Preclinical | Regulus, Inc. |
| miR-33 antagomir | Cholesterol | Preclinical | Regulus, Inc. |
| miR-34 mimic | Prostate cancer | Preclinical | Mirna/Assuragen |
| miR-92 antagomir | Cardiovascular | Preclinical | MiRagen, Inc. |
| miR-122 antagomir | Hepatitis C | Phase 2 | Santaris A/S |
| miR-122 antagomir | Hepatitis C | Preclinical | Regulus, Inc. |
| miR-208 antagomir | Chronic heart disease | Preclinical | MiRagen, Inc. |
| miR-451 antagomir | Polycythemia | Preclinical | MiRagen, Inc. |

MI: myocardial infarction; miRNA: microRNA.
Adapted from http://microRNA-bioanalytics.de.

off the angiogenic switch controlling disease severity. A lipid-based, cyclic RGD-containing nanoparticle was designed to deliver miRNA or anti-miRNA to the diseased or proliferative blood vessels that express $\alpha v\beta 3$ [131]. The scientists also found out that by delivering more of this miRNA, it may be possible to promote new blood vessels development in patients who have sustained tissue damage from stroke, heart attacks, or diabetes.

Studies have shown that stroke alters expression levels of multiple miRNAs in neural progenitor cells. miRNA 124, a preferentially expressed miRNA in neurons has been known to be one of them. Liu et al. [132] showed that there was substantial reduction of miRNA 124 in neural progenitor cells after stroke and that this was inversely associated with activation of Notch signal [132]. This group efficiently introduced this miRNA124 into progenitor cells using a noncovalent peptide (MPG)–based nanoparticle system and demonstrated that this inhibited neural progenitor cell proliferation and promoted the neuronal differentiation of the progenitor cells [133]. MPG is a 27-residue-long primary amphipathic peptide that forms stable complexes with miRNA [133]. This peptide has three distinct domains, which arean N-terminal hydrophobic motif required for interaction with lipid moiety of cell membrane, a hydrophilic domain involved in the interaction with nucleic acids, and a linker domain that improves the flexibility and integrity of hydrophobic and hydrophilic domains.

Hwang and his colleagues used brain-targeted nanoparticles to deliver miR124a to brain [134]. They used the rabies virus glycoprotein (RVG)–labeled nontoxic disulfide linkage in branched polyethyleneimine (SSPEI) nanomaterial to deliver miR124a to neurons. This group has demonstrated enhanced accumulation of miR124a using a cy5.5-miR124a/RVG-SSPEI. One challenge of using this delivery system for miRNA delivery *in vivo* involves its passage through BBB and delivery of drugs to the CNS. Mannitol played a critical role in supporting BBB opening to facilitate RVG-SSPEI–mediated transport of the miRNA to the brain. So, after the role of specific miRNA in disease pathogenesis is established, selecting anti-miRNA or miRNA mimics and appropriate delivery strategies will be the next critical steps

for successful therapy. It is hoped that miRNA therapeutics with the help of an effective and safe delivery system will soon lead to effective treatments for undruggable targets. The biotech industries try to make the best of it by identifying miRNAs for appropriate therapeutic applications. Table 18.1 shows some of the miRNA mimics or antagomers listed by the companies, in preclinical and clinical development to substitute or downregulate overexpressed miRNA.

## 18.7 CONCLUSIONS

The discovery of RNAi-mediated gene silence heralds a powerful new strategy for treating human diseases. Despite the tremendous efforts from academia and industry in providing a glimpse of this breakthrough discovery in molecular biology, there still remains a great challenge in its clinical translation into viable therapies. Although there are several obstacles for successful RNAi therapeutics, a major hurdle is in overcoming the delivery barriers to the targeted tissue and cells without disrupting their effectiveness and trigging unwanted immune responses. Toward this end, we have discussed some of recent developments in solving some of the delivery issues for successful RNAi therapies *in vitro* and *in vivo*. Although each system has its own merits and shortcomings, multifunctional nanosystems have shown tremendous potentials for RNAi therapy. Thus, the further advancements in gene delivery research will greatly hinge on the continued progresses in conjugation and formulation technologies that can pave the way forward for the development of next generation of mature technologies for their successful clinical translation.

## REFERENCES

1. Fire A, Xu S, Montgomery MK, Kostas SA, Driver SE, Mello CC. Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. Nature 1998;391(6669):806–811.
2. Macrae IJ, Zhou K, Li F, Repic A, Brooks AN, Cande WZ, et al. Structural basis for double-stranded RNA processing by Dicer. Science 2006;311(5758):195–198.
3. Cenik ES, Fukunaga R, Lu G, Dutcher R, Wang Y, Tanaka Hall TM, et al. Phosphate and R2D2 restrict the substrate specificity of Dicer-2, an ATP-driven ribonuclease. Mol Cell 2011;42(2):172–184.
4. Zamore PD, Tuschl T, Sharp PA, Bartel DP. RNAi: double-stranded RNA directs the ATP-dependent cleavage of mRNA at 21 to 23 nucleotide intervals. Cell 2000;101 (1):25–33.
5. Vermeulen A, Behlen L, Reynolds A, Wolfson A, Marshall WS, Karpilow J, et al. The contributions of dsRNA structure to Dicer specificity and efficiency. RNA 2005;11 (5):674–682.
6. Gregory RI, Chendrimada TP, Cooch N, Shiekhattar R. Human RISC couples microRNA biogenesis and posttranscriptional gene silencing. Cell 2005;123(4):631–640.
7. Cenik ES, Zamore PD. Argonaute proteins. Curr Biol 2011;21(12):R446–R449.

8. Bartel DP. MicroRNAs: target recognition and regulatory functions. Cell 2009;136 (2):215–233.
9. Lee Y, Ahn C, Han J, Choi H, Kim J, Yim J, et al. The nuclear RNase III Drosha initiates microRNA processing. Nature 2003;425(6956):415–419.
10. Ambros V. The functions of animal microRNAs. Nature 2004;431(7006):350–355.
11. Broderick JA, Zamore PD. MicroRNA therapeutics. Gene Ther 2011;18(12):1104–1110.
12. Guo H, Ingolia NT, Weissman JS, Bartel DP. Mammalian microRNAs predominantly act to decrease target mRNA levels. Nature 2010;466(7308):835–840.
13. Baek D, Villen J, Shin C, Camargo FD, Gygi SP, Bartel DP. The impact of microRNAs on protein output. Nature 2008;455(7209):64–71.
14. Gunaratne PH. Embryonic stem cell microRNAs: defining factors in induced pluripotent (iPS) and cancer (CSC) stem cells? Curr Stem Cell Res Ther 2009;4(3):168–177.
15. Xie H, Lim B, Lodish HF. MicroRNAs induced during adipogenesis that accelerate fat cell development are downregulated in obesity. Diabetes 2009;58(5):1050–1057.
16. Esau C, Kang X, Peralta E, Hanson E, Marcusson EG, Ravichandran LV, et al. MicroRNA-143 regulates adipocyte differentiation. J Biol Chem 2004;279 (50):52361–52365.
17. Poy MN, Eliasson L, Krutzfeldt J, Kuwajima S, Ma X, Macdonald PE, et al. A pancreatic islet-specific microRNA regulates insulin secretion. Nature 2004;432(7014):226–230.
18. Jopling CL, Norman KL, Sarnow P. Positive and negative modulation of viral and cellular mRNAs by liver-specific microRNA miR-122. Cold Spring Harb Symp Quant Biol 2006;71:369–376.
19. Young DD, Connelly CM, Grohmann C, Deiters A. Small molecule modifiers of microRNA miR-122 function for the treatment of hepatitis C virus infection and hepatocellular carcinoma. J Am Chem Soc 2010;132(23):7976–7981.
20. Trajkovski M, Hausser J, Soutschek J, Bhat B, Akin A, Zavolan M, et al. MicroRNAs 103 and 107 regulate insulin sensitivity. Nature 2011;474(7353):649–653.
21. Yuan X, Naguib S, Wu Z. Recent advances of siRNA delivery by nanoparticles. Expert Opin Drug Deliv 2011;8(4):521–536.
22. Czech MP, Aouadi M, Tesz GJ. RNAi-based therapeutic strategies for metabolic disease. Nat Rev Endocrinol 2011;7(8):473–484.
23. Peer D, Lieberman J. Special delivery: targeted therapy with small RNAs. Gene Ther 2011;18(12):1127–1133.
24. Sliva K, Schnierle BS. Selective gene silencing by viral delivery of short hairpin RNA. Virol J 2010;7:248.
25. Brummelkamp TR, Bernards R, Agami R. A system for stable expression of short interfering RNAs in mammalian cells. Science 2002;296(5567):550–553.
26. Mann R, Mulligan RC, Baltimore D. Construction of a retrovirus packaging mutant and its use to produce helper-free defective retrovirus. Cell 1983;33(1):153–159.
27. Relph KL, Harrington KJ, Pandha H. Adenoviral strategies for the gene therapy of cancer. Semin Oncol 2005;32(6):573–582.
28. Wu E, Nemerow GR. Virus yoga: the role of flexibility in virus host cell recognition. Trends Microbiol 2004;12(4):162–169.
29. Koerber JT, Jang JH, Schaffer DV. DNA shuffling of adeno-associated virus yields functionally diverse viral progeny. Mol Ther 2008;16(10):1703–1709.

30. Grimm D, Kay MA. Therapeutic short hairpin RNA expression in the liver: viral targets and vectors. Gene Ther 2006;13(6):563–575.
31. Grimm D, Pandey K, Kay MA. Adeno-associated virus vectors for short hairpin RNA expression. Methods Enzymol 2005;392:381–405.
32. David S, Pitard B, Benoît JP, Passirani C. Non-viral nano-systems for systemic siRNA delivery. Pharmacol Res 2010;62(2):100–114.
33. Wang J, Lu Z, Wientjes MG, Au JL. Delivery of siRNA therapeutics: barriers and carriers. AAPS J 2010;12(4):492–503.
34. Pecot CV, Calin GA, Coleman RL, Lopez-Berestein G, Sood AK. RNA interference in the clinic: challenges and future directions. Nat Rev Cancer 2011;11(1):59–67.
35. Akhtar S, Benter IF. Nonviral delivery of synthetic siRNAs *in vivo*. J Clin Invest 2007;117(12):3623–3632.
36. Ishida T, Wang X, Shimizu T, Nawata K, Kiwada H. PEG-modified liposomes elicit an anti-PEG IgM response in a T cell-independent manner. J Control Release 2007; 122(3):349–355.
37. Semple SC, Harasym TO, Clow KA, Ansell SM, Klimuk SK, Hope MJ. Immunogenicity and rapid blood clearance of liposomes containing polyethylene glycol-lipid conjugates and nucleic Acid. J Pharmacol Exp Ther 2005;312(3):1020–1026.
38. Jeong JH, Mok H, Oh YK, Park TG. siRNA conjugate delivery systems. Bioconjug Chem 2009;20(1):5–14.
39. Soutschek J, Akinc A, Bramlage B, Charisse K, Constien R, Donoghue M, et al. Therapeutic silencing of an endogenous gene by systemic administration of modified siRNAs. Nature 2004;432(7014):173–178.
40. Wolff JA, Rozema DB. Breaking the bonds: non-viral vectors become chemically dynamic. Mol Ther 2008;16(1):8–15.
41. Song E, Zhu P, Lee SK, Chowdhury D, Kussman S, Dykxhoorn DM, et al. Antibody mediated *in vivo* delivery of small interfering RNAs via cell-surface receptors. Nat Biotechnol 2005;23(6):709–717.
42. Dassie JP, Liu XY, Thomas GS, Whitaker RM, Thiel KW, Stockdale KR, et al. Systemic administration of optimized aptamer-siRNA chimeras promotes regression of PSMA-expressing tumors. Nat Biotechnol 2009;27(9):839–849.
43. Hu-Lieskovan S, Zhang J, Wu L, Shimada H, Schofield DE, Triche TJ. EWS-FLI1 fusion protein up-regulates critical genes in neural crest development and is responsible for the observed phenotype of Ewing's family of tumors. Cancer Res 2005;65(11):4633–4644.
44. Heidel JD, Yu Z, Liu JY, Rele SM, Liang Y, Zeidan RK, et al. Administration in non-human primates of escalating intravenous doses of targeted nanoparticles containing ribonucleotide reductase subunit M2 siRNA. Proc Natl Acad Sci USA 2007;104 (14):5715–5721.
45. Davis ME, Zuckerman JE, Choi CH, Seligson D, Tolcher A, Alabi CA, et al. Evidence of RNAi in humans from systemically administered siRNA via targeted nanoparticles. Nature 2010;464(7291):1067–1070.
46. Temming K, Schiffelers RM, Molema G, Kok RJ. RGD-based strategies for selective delivery of therapeutics and imaging agents to the tumour vasculature. Drug Resist Updat 2005;8(6):381–402.
47. Leng Q, Woodle MC, Lu PY, Mixson AJ. Advances in systemic siRNA delivery. Drugs Future 2009;34(9):721.

48. Potineni A, Lynn DM, Langer R, Amiji MM. Poly(ethylene oxide)-modified poly(beta-amino ester) nanoparticles as a pH-sensitive biodegradable system for paclitaxel delivery. J Control Release 2003;86(2–3):223–234.
49. Susa M, Iyer AK, Ryu K, Choy E, Hornicek FJ, Mankin H, et al. Inhibition of ABCB1 (MDR1) expression by an siRNA nanoparticulate delivery system to overcome drug resistance in osteosarcoma. PLoS One 2010;5(5):e10764.
50. Matsumura Y, Maeda H. A new concept for macromolecular therapeutics in cancer chemotherapy: mechanism of tumoritropic accumulation of proteins and the antitumor agent smancs. Cancer Res 1986;46:6387–6392.
51. Folkman J. Angiogenesis in cancer, vascular, rheumatoid and other disease. Nat Med 1995;1(1):27–31.
52. Greish K, Fang J, Inutsuka T, Nagamitsu A, Maeda H. Macromolecular therapeutics: advantages and prospects with special emphasis on solid tumour targeting. Clin Pharmacokinet 2003;42(13):1089–1105.
53. Padera TP, Kadambi A, di Tomaso E, Carreira CM, Brown EB, Boucher Y, et al. Lymphatic metastasis in the absence of functional intratumor lymphatics. Science 2002;296(5574):1883–1886.
54. Ghosh K, Thodeti CK, Dudley AC, Mammoto A, Klagsbrun M, Ingber DE. Tumor-derived endothelial cells exhibit aberrant Rho-mediated mechanosensing and abnormal angiogenesis *in vitro*. Proc Natl Acad Sci USA 2008;105(32):11305–11310.
55. Maeda H, Wu J, Sawa T, Matsumura Y, Hori K. Tumor vascular permeability and the EPR effect in macromolecular therapeutics: a review. J Control Release 2000;65(1–2):271–284.
56. Senger DR, Galli SJ, Dvorak AM, Perruzzi CA, Harvey VS, Dvorak HF. Tumor cells secrete a vascular permeability factor that promotes accumulation of ascites fluid. Science 1983;219(4587):983–985.
57. Dvorak HF, Nagy JA, Feng D, Brown LF, Dvorak AM. Vascular permeability factor/vascular endothelial growth factor and the significance of microvascular hyperpermeability in angiogenesis. Curr Top Microbiol Immunol 1999;237:97–132.
58. Senger DR, Galli SJ, Dvorak AM, Perruzzi CA, Harvey VS, et al. Tumor cells secrete a vascular permeability factor that promotes accumulation of ascites fluid. Science 1983;219(4587):983–985.
59. Maeda H. The enhanced permeability and retention (EPR) effect in tumor vasculature: the key role of tumor-selective macromolecular drug targeting. Adv Enzyme Regul 2001;41:189–207.
60. Maeda H, Bharate GY, Daruwalla J. Polymeric drugs for efficient tumor-targeted drug delivery based on EPR-effect. Eur J Pharm Biopharm 2009;71(3):409–419.
61. Duncan R, Ringsdorf H, Satchi-Fainaro R. Polymer therapeutics—polymers as drugs, drug and protein conjugates and gene delivery systems: past, present and future opportunities. J Drug Target 2006;14(6):337–341.
62. Danhier F, Feron O, Préat V. To exploit the tumor microenvironment: passive and active tumor targeting of nanocarriers for anti-cancer drug delivery. J Control Release 2010;148(2):135–146.
63. Ganta S, Devalapally H, Shahiwala A, Amiji M. A review of stimuli-responsive nanocarriers for drug and gene delivery. J Control Release 2008;126(3):187–204.

64. Iyer A, Khaled G, Fang J, Maeda H. Exploiting the enhanced permeability and retention effect for tumor targeting. Drug Discov Today 2006;11(17–18):812–818.
65. Davis SS. Biomedical applications of nanotechnology—implications for drug targeting and gene therapy. Trends Biotechnol 1997;15(6):217–224.
66. Kuszyk BS, Corl FM, Franano FN, Bluemke DA, Hofmann LV, Fortman BJ, et al. Tumor transport physiology: implications for imaging and imaging-guided therapy. Am J Roentgenol 2001;177(4):747–753.
67. Sugahara KN, Teesalu T, Karmali PP, Kotamraju VR, Agemy L, Girard OM, et al. Tissue-penetrating delivery of compounds and nanoparticles into tumors. Cancer Cell 2009;16(6):510–520.
68. Marcucci F, Lefoulon F. Active targeting with particulate drug carriers in tumor therapy: fundamentals and recent progress. Drug Discov Today 2004;9(5):219–228.
69. Milane L, Duan ZF, Amiji M. Pharmacokinetics and biodistribution of lonidamine/paclitaxel loaded, EGFR-targeted nanoparticles in an orthotopic animal model of multi-drug resistant breast cancer. Nanomedicine 2011;7(4):435–444.
70. Milane L, Duan Z, Amiji M. Development of EGFR-targeted polymer blend nanocarriers for combination paclitaxel/lonidamine delivery to treat multi-drug resistance in human breast and ovarian tumor cells. Mol Pharm 2011;8(1):185–203.
71. Lue N, Ganta S, Hammer DX, Mujat M, Stevens AE, Harrison L, et al. Preliminary evaluation of a nanotechnology-based approach for the more effective diagnosis of colon cancers. Nanomedicine (Lond) 2010;5(9):1467–1479.
72. Werner ME, Karve S, Sukumar R, Cummings ND, Copp JA, Chen RC, et al. Folate-targeted nanoparticle delivery of chemo- and radiotherapeutics for the treatment of ovarian cancer peritoneal metastasis. Biomaterials 2011;32(33):8548–8554.
73. Rihova B. Receptor-mediated targeted drug or toxin delivery. Adv Drug Deliv Rev 1998;29(3):273–289.
74. Farokhzad OC, Karp JM, Langer R. Nanoparticle-aptamer bioconjugates for cancer targeting. Expert Opin Drug Deliv 2006;3(3):311–324.
75. Iyer AK, Lan X, Zhu X, Su Y, Feng J, Zhang X, et al. Novel human single chain antibody fragments that are rapidly internalizing effectively target epithelioid and sarcomatoid mesotheliomas. Cancer Res 2011;71(7):2428–2432.
76. Iyer AK, Su Y, Feng J, Lan X, Zhu X, Liu Y, et al. The effect of internalizing human single chain antibody fragment on liposome targeting to epithelioid and sarcomatoid mesothelioma. Biomaterials 2011;32(10):2605–2613.
77. Benyettou F, Lalatonne Y, Chebbi I, Di Benedetto M, Serfaty JM, Lecouvey M, et al. A multimodal magnetic resonance imaging nanoplatform for cancer theranostics. Phys Chem Chem Phys 2011;13(21):10020–10027.
78. O'Brien ME, Wigler N, Inbar M, Rosso R, Grischke E, Santoro A, et al. Reduced cardiotoxicity and comparable efficacy in a phase III trial of PEG-modified liposomal doxorubicin HCl (CAELYX/Doxil) versus conventional doxorubicin for first-line treatment of metastatic breast cancer. Ann Oncol 2004;15(3):440–449.
79. Otsuka H, Nagasaki Y, Kataoka K. PEG-modified nanoparticles for biological and pharmaceutical applications. Adv Drug Deliv Rev 2003;55(3):403–419.
80. Dobrovolskaia MA, McNeil SE. Immunological properties of engineered nanomaterials. Nat Nanotechnol 2007;2(8):469–478.

81. Maeda H, Matsumura Y. EPR effect based drug design and clinical outlook for enhanced cancer chemotherapy. Adv Drug Deliv Rev 2011;63(3):129–130.
82. van Vlerken LE, Duan Z, Little SR, Seiden MV, Amiji MM. Biodistribution and pharmacokinetic analysis of paclitaxel and ceramide administered in multifunctional polymer-blend nanoparticles in drug resistant breast cancer model. Mol Pharm 2008; 5(4):516–526.
83. Cao N, Cheng D, Zou S, Ai H, Gao J, Shuai X. The synergistic effect of hierarchical assemblies of siRNA and chemotherapeutic drugs co-delivered into hepatic cancer cells. Biomaterials 2011;32(8):2222–2232.
84. Bhavsar MD, Amiji MM. Gastrointestinal distribution and *in vivo* gene transfection studies with nanoparticles-in-microsphere oral system (NiMOS). J Control Release 2007;119(3):339–348.
85. Uhrich KE, Cannizzaro SM, Langer RS, Shakesheff KM. Polymeric systems for controlled drug release. Chem Rev 1999;99(11):3181–3198.
86. Dillen K, Vandervoort J, Van den Mooter G, Ludwig A. Evaluation of ciprofloxacin-loaded Eudragit RS100 or RL100/PLGA nanoparticles. Int J Pharm 2006;314(1):72–82.
87. Abeylath SC, Ganta S, Iyer AK, Amiji M. Combinatorial-designed multifunctional polymeric nano-systems for tumor-targeted therapeutic delivery. Acc Chem Res 2011; 44(10):1009–1017.
88. Abeylath SC, Amiji MM. 'Click' synthesis of dextran macrostructures for combinatorial-designed self-assembled nanoparticles encapsulating diverse anticancer therapeutics. Bioorg Med Chem 2011;19(21):6167–6173.
89. Ghotbi Z, Haddadi A, Hamdy S, Hung RW, Samuel J, Lavasanifar A. Active targeting of dendritic cells with mannan-decorated PLGA nanoparticles. J Drug Target 2011; 19(4):281–292.
90. Choi CH, Alabi CA, Webster P, Davis ME. Mechanism of active targeting in solid tumors with transferrin-containing gold nanoparticles. Proc Natl Acad Sci USA 2010; 107(3):1235–1240.
91. Choi KY, Chung H, Min KH, Yoon HY, Kim K, Park JH, et al. Self-assembled hyaluronic acid nanoparticles for active tumor targeting. Biomaterials 2010;31(1):106–114.
92. Pecot CV, Calin GA, Coleman RL, Lopez-Berestein G, Sood AK. RNA interference in the clinic: challenges and future directions. Nat Rev Cancer 2011;11(1):59–67.
93. Torchilin VP. Recent approaches to intracellular delivery of drugs and DNA and organelle targeting. Annu Rev Biomed Eng 2006;8:343–375.
94. Panyam J, Labhasetwar V. Targeting intracellular targets. Curr Drug Deliv 2004; 1(3):235–247.
95. Szakacs G, Paterson JK, Ludwig JA, Booth-Genthe C, Gottesman MM. Targeting multidrug resistance in cancer. Nat Rev Drug Discov 2006;5(3):219–234.
96. Devalapally H, Duan Z, Seiden MV, Amiji MM. Modulation of drug resistance in ovarian adenocarcinoma by enhancing intracellular ceramide using tamoxifen-loaded biodegradable polymeric nanoparticles. Clin Cancer Res 2008;14(10):3193–3203.
97. van Vlerken LE, Duan Z, Seiden MV, Amiji MM. Modulation of intracellular ceramide using polymeric nanoparticles to overcome multidrug resistance in cancer. Cancer Res 2007;67(10):4843–4850.
98. McCarthy JR, Weissleder R. Multifunctional magnetic nanoparticles for targeted imaging and therapy. Adv Drug Deliv Rev 2008;60(11):1241–1251.

99. Jones EF, He J, VanBrocklin HF, Franc BL, Seo Y. Nanoprobes for medical diagnosis: current status of nanotechnology in molecular imaging. Curr Nanosci 2008;4(1):17–29.
100. Gindy ME, Prud'homme RK. Multifunctional nanoparticles for imaging, delivery and targeting in cancer therapy. Expert Opin Drug Deliv 2009;6(8):865–878.
101. Shegokar R, Al Shaal L, Mishra PR. siRNA delivery: challenges and role of carrier systems. Pharmazie 2011;66(5):313–318.
102. Kim SH, Jeong JH, Lee SH, Kim SW, Park TG. Local and systemic delivery of VEGF siRNA using polyelectrolyte complex micelles for effective treatment of cancer. J Control Release 2008;129(2):107–116.
103. Susa M, Iyer AK, Ryu K, Hornicek FJ, Mankin H, Amiji MM, et al. Doxorubicin loaded polymeric nanoparticulate delivery system to overcome drug resistance in osteosarcoma. BMC Cancer 2009;9:399.
104. Huh MS, Lee SY, Park S, Lee S, Chung H, Choi Y, et al. Tumor-homing glycol chitosan/polyethylenimine nanoparticles for the systemic delivery of siRNA in tumor-bearing mice. J Control Release 2010;144(2):134–143.
105. Medarova Z, Pham W, Farrar C, Petkova V, Moore A. *In vivo* imaging of siRNA delivery and silencing in tumors. Nat Med 2007;13(3):372–377.
106. Kumar M, Yigit M, Dai G, Moore A, Medarova Z. Image-guided breast tumor therapy using a small interfering RNA nanodrug. Cancer Res 2010;70(19):7553–7561.
107. Kriegel C, Amiji M. Oral TNF-alpha gene silencing using a polymeric microsphere-based delivery system for the treatment of inflammatory bowel disease. J Control Release 2011;150(1):77–86.
108. Peer D, Park EJ, Morishita Y, Carman CV, Shimaoka M. Systemic leukocyte-directed siRNA delivery revealing cyclin D1 as an anti-inflammatory target. Science 2008; 319(5863):627–630.
109. Wilson DS, Dalmasso G, Wang L, Sitaraman SV, Merlin D, Murthy N. Orally delivered thioketal nanoparticles loaded with TNF-alpha-siRNA target inflammation and inhibit gene expression in the intestines. Nat Mater 2010;9(11):923–928.
110. Bonoiu A, Mahajan SD, Ye L, Kumar R, Ding H, Yong KT, et al. MMP-9 gene silencing by a quantum dot-siRNA nanoplex delivery to maintain the integrity of the blood brain barrier. Brain Res 2009;1282:142–155.
111. Feldmann M, Maini RN. Lasker Clinical Medical Research Award. TNF defined as a therapeutic target for rheumatoid arthritis and other autoimmune diseases. Nat Med 2003;9(10):1245–1250.
112. Jensen LB, Griger J, Naeye B, Varkouhi AK, Raemdonck K, Schiffelers R, et al. Comparison of Polymeric siRNA nanocarriers in a murine LPS-activated macrophage cell line: gene silencing, toxicity and off-target gene expression. Pharm Res 2012; 29(3):669–682.
113. Sparks J, Slobodkin G, Matar M, Congo R, Ulkoski D, Rea-Ramsey A, et al. Versatile cationic lipids for siRNA delivery. J Control Release 2011;158(2):269–276.
114. Woodrow KA, Cu Y, Booth CJ, Saucier-Sawyer JK, Wood MJ, Saltzman WM. Intravaginal gene silencing using biodegradable polymer nanoparticles densely loaded with small-interfering RNA. Nat Mater 2009;8(6):526–533.
115. Zhang W, Yang H, Kong X, Mohapatra S, San Juan-Vergara H, Hellermann G, et al. Inhibition of respiratory syncytial virus infection with intranasal siRNA nanoparticles targeting the viral NS1 gene. Nat Med 2005;11(1):56–62.

116. Jimenez JL, Clemente MI, Weber ND, Sanchez J, Ortega P, de la Mata FJ, et al. Carbosilane dendrimers to transfect human astrocytes with small interfering RNA targeting human immunodeficiency virus. BioDrugs 2010;24(5):331–343.
117. Huang Z, King MR. An immobilized nanoparticle-based platform for efficient gene knockdown of targeted cells in the circulation. Gene Ther 2009;16(10):1271–1282.
118. Garzon R, Marcucci G, Croce CM. Targeting microRNAs in cancer: rationale, strategies and challenges. Nat Rev Drug Discov 2010;9(10):775–789.
119. McManus MT. MicroRNAs and cancer. Semin Cancer Biol 2003;13(4):253–258.
120. Visone R, Croce CM. MiRNAs and cancer. Am J Pathol 2009;174(4):1131–1138.
121. Bader AG, Brown D, Winkler M. The promise of microRNA replacement therapy. Cancer Res 2010;70(18):7027–7030.
122. Johnson SM, Grosshans H, Shingara J, Byrom M, Jarvis R, Cheng A, et al. RAS is regulated by the let-7 microRNA family. Cell 2005;120(5):635–647.
123. Trang P, Medina PP, Wiggins JF, Ruffino L, Kelnar K, Omotola M, et al. Regression of murine lung tumors by the let-7 microRNA. Oncogene 2010;29(11):1580–1587.
124. Trang P, Wiggins JF, Daige CL, Cho C, Omotola M, Brown D, et al. Systemic delivery of tumor suppressor microRNA mimics using a neutral lipid emulsion inhibits lung tumors in mice. Mol Ther 2011;19(6):1116–1122.
125. Rai K, Takigawa N, Ito S, Kashihara H, Ichihara E, Yasuda T, et al. Liposomal delivery of MicroRNA-7-expressing plasmid overcomes epidermal growth factor receptor tyrosine kinase inhibitor-resistance in lung cancer cells. Mol Cancer Ther 2011;10(9):1720–1727.
126. Chen Y, Zhu X, Zhang X, Liu B, Huang L. Nanoparticles modified with tumor-targeting scFv deliver siRNA and miRNA for cancer therapy. Mol Ther 2010;18(9):1650–1656.
127. Liang GF, Zhu YL, Sun B, Hu FH, Tian T, Li SC, et al. PLGA-based gene delivering nanoparticle enhance suppression effect of miRNA in HePG2 cells. Nanoscale Res Lett 2011;6:447.
128. Sekar NM, Sekar TV, Paulmurugan R. World Molecular Imaging Congress, San Diego, September 7–10, 2011.
129. Rahbek UL, Nielsen AF, Dong M, You Y, Chauchereau A, Oupicky D, et al. Bioresponsive hyperbranched polymers for siRNA and miRNA delivery. J Drug Target 2010;18(10):812–820.
130. Ren Y, Kang CS, Yuan XB, Zhou X, Xu P, Han L, et al. Co-delivery of as-miR-21 and 5-FU by poly(amidoamine) dendrimer attenuates human glioma cell growth *in vitro*. J Biomater Sci Polym Ed 2010;21(3):303–314.
131. Murphy EA, Majeti BK, Barnes LA, Makale M, Weis SM, Lutu-Fuga K, et al. Nanoparticle-mediated drug delivery to tumor vasculature suppresses metastasis. Proc Natl Acad Sci USA 2008;105(27):9343–9348.
132. Liu XS, Chopp M, Zhang RL, Tao T, Wang XL, Kassis H, et al. MicroRNA profiling in subventricular zone after stroke: MiR-124a regulates proliferation of neural progenitor cells through Notch signaling pathway. PLoS One 2011;6(8):e23461.
133. Crombez L, Charnet A, Morris MC, Aldrian-Herrada G, Heitz F, Divita G. A non-covalent peptide-based strategy for siRNA delivery. Biochem Soc Trans 2007;35(Pt 1):44–46.

134. Hwang do W, Son S, Jang J, Youn H, Lee S, Lee D, L. et al. A brain-targeted rabies virus glycoprotein-disulfide linked PEI nanocarrier for delivery of neurogenic microRNA. Biomaterials 2011;32(21):4968–4975.

135. Kim DH, Rossi JJ. Strategies for silencing human disease using RNA interference. Nat Rev Genet 2007;8(3):173–184.

136. Pecot CV, Calin GA, Coleman RL, Lopez-Berestein G, Sood AK. RNA interference in the clinic: challenges and future directions. Nat Rev Cancer 2011;11(1):59–67.

# Advent and Maturation of Regenerative Medicine

ABNER M. MHASHILKAR and ANTHONY ATALA

## 19.1 INTRODUCTION

The goals for regenerative medicine (RM) are focused toward developing innovative tools in modern medicine. Recent progress has led to the development and clinical implementation of newer technologies, ranging from diagnostics to therapeutics for a wide variety of metabolic, degenerative, and infectious diseases. These include specific applications toward age- and genetic-related human and animal dysfunctions and disorders.

The worldwide RM landscape is changing and rapidly evolving with novel technological advances emerging not only from research centers in North America, Europe, and Australia but also from translational, clinical, and manufacturing groups in Asia and Latin America. The focus of this review is to present, in a nutshell, the current status of adult stem cell therapy and tissue engineering (TE), including the challenges faced, progresses attained, and an optimistic view of the future. We will also provide our views on how the next 20-plus years will potentially affect these RM-driven innovations and how they are applied toward a variety of disease treatments, cures, and diagnostics.

## 19.2 REGENERATIVE MEDICINE: LOOKING BACK

After the breakthrough sequencing of the human genome a decade ago [1], personalized RM, including stem cell therapies and TE technologies, has generated an immense amount of interest, momentum, and hype. As a result, there has been an exponential increase in the expectations for translation of these technologies, both on a national and global scale. This is because RM offers potential solutions to many unmet clinical needs and challenges and provides a variety of innovative stem cell therapy and TE tools to repair, regenerate, and rebuild tissues and organs [2].

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*,
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

## 19.3 TRANSPLANTATION MEDICINE: HISTORICAL PERSPECTIVES

To date, the observational studies encompassing several decades have inferred that, unlike amphibians, which can regenerate limbs, humans react to injury via responses geared toward wound closure and healing and are thus unable to undergo similar regeneration processes. This healing process is further decreased by aging as well as the challenges associated with chronic, acute, and genetic diseases and dysfunctions [3]. In the 18th century, clinicians attempted organ transplantation procedures in animals and humans in order to replace tissues and organs that had been damaged by injury or disease, but the results were poor. In the early 1950s, the evolving fields of surgery and a new understanding of immune-suppression processes united to catalyze rapid advancement of transplantation medicine. After a string of many unsuccessful attempts, organ transplants became a reality (see Table 19.1, which represents a series

**TABLE 19.1 History of Organ Transplantation and New Related Procedures**

| Year | Organ or Procedure | Site or Professionals Involved |
|---|---|---|
| 1914 | Skin grafting | World War I |
| 1954 | Kidney | Brigham and Women's Hospital, Boston, MA |
| 1956 | Bone marrow transplantation | Dr. Donnall Thomas in a recipient twin |
| 1962 | Successful kidney | Drs. Murray and Hume at Brigham and Women's Hospital, Boston |
| 1963 | Successful lung | Dr. Hardy at the University of Mississippi Medical Center |
| 1966 | Pancreas and kidney | Drs. Lillehei and Kelly at the University of Minnesota, MN |
| 1967 | Liver | Dr. Starzl at the University of Colorado, Denver, CO |
| 1967 | Heart | Dr. Barnard, Groote Schuur Hospital, Cape Town, South Africa |
| 1968 | Heart—USA | Dr. Shumway, Stanford University Hospital, Stanford, CA |
| 1981 | Heart and lung | Dr. Reitz, Stanford University Hospital, Stanford, CA |
| 1983 | Single lung | Dr. Cooper, Toronto General Hospital, Toronto, Canada |
| 1983 | Cyclosporin | FDA approves cyclosporin, successful antirejection drug |
| 1984 | NOTA | National Organ Transplant Act authorizes OPSs |
| 1986 | Double lungs | Dr. Cooper, Toronto General Hospital, Toronto, Canada |
| 1988 | Viaspan | FDA approval, extends the preservation of donated livers |
| 1989 | First living-related liver | Dr. Broelsch, University of Chicago Medical Center, Chicago, IL |
| 1989 | Small intestine | Dr. Goulet, Paris, France |
| 1990 | First living-related lung | Dr. Starnes, Stanford University Medical Center, Stanford CA |
| 1993 | Xenografting baboon–human liver | Drs. Todo, Tzakis, Fung, and Starzl, Pittsburgh, PA |
| 1994 | Prograf (FK506) | Immunosuppression drug |
| 1995 | CellCept and Neoral | Formulation of cyclosporine with fewer side effects |
| 1998 | Hand | Drs. Owen and Dubernard, Lyon, France |
| 2005 | Partial face | Drs. Devauchelle and Dubernard, Amiens, France |
| 2010 | Full face | Barcelona's Vall d'Hebron hospital |

FDA: Food and Drug Administration; OPS.

of organ transplantation procedures successfully accomplished in the past 60 years). Breakthroughs included the first kidney transplant in 1954 between identical twins [4,5], the first liver and lung transplants in 1967 [6–9], and a pancreatic transplant in 1966 [10]. This work culminated with the first heart transplant in 1967 [11]. The field of kidney transplantation took a leap in the 1990s, when about 11,000 kidney transplants were carried out in 1996 alone. About 3000 of these involved kidneys recovered from living donors [12]. Transplantation of other organs such as the liver, heart, lungs, pancreas, and small bowel are now considered an accepted part of medical treatment. In addition, other breakthroughs, including bone marrow and corneal transplants, have been made [13]. In all of these cases, mature, fully functional organs were transplanted [14–17].

Another milestone in this process was the discovery of hematopoietic stem cells (HSCs). For the first time, a group of cells, rather than a fully functional organ, could be transplanted in order to "seed" the marrow and reconstitute all hematopoietic cells in the patient [18]. Other important medical breakthroughs in the past 20 years, such as tissue typing and the development of immunosuppressive drugs, have allowed for more successful organ transplants and longer survival times for transplant recipients [19,20]. The most notable development in this area was Jean Borel's discovery of an immunosuppressive drug called cyclosporine in the mid-1970s. This drug was approved for commercial use (in the United States) in 1983 [21,22].

Despite these significant milestones in transplant medicine and the tremendous impact they have had on many lives, the use of transplantation is limited because there simply are not enough organs to transplant [18]. Unfortunately, the need for organ transplants continues to exceed the organ supply, and more and more names are being added to the transplant waiting lists all around the world.

## 19.4 BONE MARROW TRANSPLANTATION AND HEMATOPOIETIC CELL TRANSPLANTATION

The past 60 years has seen the increased use of bone marrow and HSC transplantation to treat a variety of malignant and nonmalignant diseases. Early observations in radiation biology led to the finding that bone marrow is extremely sensitive to the damaging effects of radiation. However, it was subsequently observed that infusion of healthy marrow could rescue lethally irradiated animals. This led to the concept of destroying diseased bone marrow using whole-body radiation and subsequently rescuing the patient with healthy bone marrow to cure the disease. Some 200 allogeneic bone marrow transplants were performed in this manner in the 1950s, although there were no long-term successes until studies of transplantation using identical twin donors helped to illustrate the challenges associated with immune rejection and graft-versus-host disease (GVDH). These transplants spawned several effective research studies in the field [23]. The first successful bone marrow transplant procedures took place shortly thereafter in infants with X-linked lymphopenic immune deficiency and Wiskott-Aldrich syndrome [24,25]. These were followed closely by success stories in patients with aplastic anemia [26] and leukemia [27].

In 1988, the world's first umbilical cord blood transplant (UCBT) was performed successfully in a patient with Fanconi anemia. The cells used in this transplant were collected at the birth of the patient's sibling [28]. The 5-year leukemia-free survival rate in 503 children with acute lymphoblastic leukemia who received a transplant of umbilical cord blood (UCB) that was mismatched for either one or two human leukocyte antigens (HLAs) was similar when compared with the survival rate of 282 children who received bone marrow transplants [29]. In 1992, a patient was successfully transplanted with cord blood instead of bone marrow for the treatment of leukemia.

The limited number of cells in cord blood traditionally restricted its use to children and small adults. However, research into cord blood expansion and the use of multiple units for a single patient have allowed this cell source to become routinely available to many more patients. Over the past decade, the use of cord blood has rapidly expanded, and there has been a scientifically based realization that cord blood is an effective alternative for bone marrow. Advantages of cord blood appear to include a shorter procurement time, fewer issues with GVHD, and a smaller number of cells needed to obtain engraftment [30–32]. Despite considerable earlier skepticism initially by both scientists and clinical specialists in the field, UCB has now become one of the most commonly used sources of HSCs for allogeneic transplantation. Today an estimated 600,000 UCB units have been banked, and 20,000 UCB units have been distributed worldwide for both adults and children with life-threatening malignant and nonmalignant diseases. Table 19.2 gives a historical perspective on cord blood and its diverse applications and shows the exponential rise in cord blood banking worldwide [33–36].

Umbilical cord blood banks (UCBBs) have been established for related or unrelated UCBT. The progress in the field of UCBT has paralleled the huge interest in establishing and developing CBB worldwide. Today more than 400,000 cord blood grafts are available in more than 50 CBBs (http://www.bmdw.org and http://www.wmda.org). This development is due to the organization of international registries for outcome data collection such as the National Marrow Donor Program (NMDP; http://www.nmdp.org) [37]; the Center for International Blood and Marrow Transplant Research (CIBMTR; http://www.cibmtr.org); and other CBB networks, including Netcord (http://www.netcord.org) and Eurocord (http://www.eurocord.org), which is an international registry operating on behalf of the European Blood and Marrow Transplant (EBMT) group and includes European and non-European centers.

The NMDP, the largest donor registry, has facilitated unrelated donor hematopoietic cell transplants for more than 30 years. These unrelated donor bone marrow transplants have been used in both children and adults with acute leukemia, with roughly 10,000 recipients to date. Around 300,000 cord units are currently stored in banks worldwide. The NMDP contains typing information on more than 4 million potential donors and can search international registries as well. Bone Marrow Donors Worldwide (BMDW) is a collective database of 59 registries in 43 countries and 37 cord blood registries from 21 countries; 11,386,999 potential donors and cord units were available as of July 2007 [37–40].

**TABLE 19.2 Historical Perspective on Cord Blood Banking and Applications [30–32]**

| | |
|---|---|
| 1. First HLA-identical sibling cord blood transplant in Fanconi Anemia | Gluckman et al. (1989) |
| 2. Optimization of UCB collection and storage | Rubinstein et al. (1993–1995) |
| 3. First unrelated mismatched cord blood transplant in children | Kurtzbach (1996) |
| 4. First unrelated cord blood transplant in adults | Gluckman et al. (1997) |
| 5. In HLA-identical sibling transplants, cord blood gave delayed engraftment, less graft-versus-host effect, and good survival | Rocha et al. (2000) |
| 6. Description of criteria of donor choice based on number of cells and possibility of using mismatched cord blood | Wagner (1995) and Rubinstein (1998) Laughlin (2001) and Gluckman (2004) |
| 7. Development of cord blood banks (CBB) (related and unrelated use | Paris, Dusseldorf, New York, and Milan |
| 8. Creation of the Eurocord Netcord network | Wernet (2004) |
| 9. Compared to matched, unrelated bone mismatched cord blood gave similar long-term leukemia-free survival in children | Rocha (2001–2004) Takahashi (2007), Eapen et al. (2007), and Laughlin et al. (2004) |
| 10. Improvement of results mostly in adults by double cord blood transplants and nonmyeloablative conditioning regimens | Hwang et al. (2007) |
| 11. Isolation of nonhematopoietic stem cells from cord blood as a first step for RM | Kogler et al. (2004) |
| 12. Development of cord blood banks | Worldwide |

The largest UCBB in the United States is the New York Cord Blood Bank, which has collected more than 35,000 units. The UCBB in Spain now has more than 28,000 specimens collected and processed, and the Australian cord registry has more than 18,000 specimens. Several additional UCBBs are organized throughout the United States, Europe, and Asia, which are collecting units much like a blood bank solicits volunteer blood donations. Also, as a stand-alone example, Taburit (http://www.taburit.org) is a successful UCBB company based in Israel and has banked 3000 units to date.

In addition to bone marrow and cord blood, peripheral blood stem cells (PBSCs) have gained popularity as a source of stem cells since their initial introduction in the 1980s [41]. Over the past 40 years, HSC transplantation has been well established as a therapeutic option against inherited disorders [42–44]. This was soon followed by the identification and definition of autologous [45,46] and allogeneic [47,48] approaches when further developments in HLA typing [49,50] and supportive care were made, and this has led to changes in the available sources of stem cells in the past 30 years. Transplants using PBSCs are not described at length in this chapter, but further information can be obtained through several other resources, including the references provided earlier.

Also, it is worthwhile to mention that many new businesses are on the rise using private, for-profit, UCBBs, which often solicit funds from expectant mothers to pay for the collection and storage of cord blood after delivery. Apart from UCBBs, several additional new banking that have sprung up in this field, including those that store adipose cells, placental cells, dental cells, and several other cell types.

## 19.5 EMERGENCE OF STEM CELL THERAPY

Stem cell therapy falls under the umbrella of RM and is fast emerging as an innovative way to treat a variety of diseases and repair a large number of injuries. It provides enhanced scope to tackle many degenerative diseases caused by the premature death or malfunction of specific cell types and the body's failure to replace or restore them. The cell therapy industry has grown exponentially and is now international in scope, representing $1 billion dollars per year turnover, and includes multifaceted technologies and therapeutics ranging from tissue and organ transplantation to advanced stem cell therapies, including TE and organogenesis. Readers are invited to study Tables 19.3 and 19.4, which represent stem cell research

**TABLE 19.3 Historical Perspective on Stem Cell Research Independent observations by the authors and also through several other search directives including http://stemcellhistory.com/.**

| Year | Findings |
|---|---|
| 1950 | First animals made by *in-vitro* fertilization (IVF) |
| 1960 | Teratocarcinomas determined to originate from embryonic germ cells in mice |
| 1960 | Embryonal carcinoma cells (EC) identified as a kind of stem cell |
| 1960s | Phenomenon of neurogenesis in human adults shown by Altman and Das |
| 1968 | The first human egg is fertilized *in vitro* |
| 1970 | EC cells injected into mouse blastocysts make chimeric mice |
| 1970 | Cultured SC cells are explored as models of embryonic development in mice |
| 1974 | Toronto scientists James Till & Ernest McCulloch establish the foundation for stem cell science |
| 1974 | 93rd Congress bans all federally funded fetal tissue research |
| 1974 | National Commission |
| 1975 | Protection of human subjects of biomedical and behavioral research |
| 1975 | Ethics Advisory Board established for fetal and fetal tissue research that originate from abortions |
| 1978 | First IVF baby is born in England |
| 1978 | Hematopoietic stem cells shown in cord blood |
| 1981 | Mouse ES cells are derived from the inner cell mass of blastocysts |
| 1981 | Mouse ES cells are grown *in vitro*. ES cells injected into mice form teratomas |
| 1984 | Pluripotent, clonal cells called embryonal carcinoma (EC) cells are developed |
| 1988 | When exposed to retinoic acid, these cells differentiate into neuron-like cells and other cell types |

**TABLE 19.3** (***Continued***)

| Year | Findings |
|---|---|
| 1988 | Federal Panel approves funding of embryo research |
| 1989 | A clonal line of human embryonal carcinoma cells derived yielding tissues from all three germ layers |
| 1990 | President George H.W. Bush vetos bill |
| 1993 | President Clinton Executive Order lifts moratorium and reverses order |
| 1994 | Human blastocysts are generated and the inner cell mass is maintained in culture |
| 1994 | ES-like cells form in the center and retain stem cell-like morphology |
| 1995 | Congress bans federal funding for research on embryos through Dickey-Wicker Amendment |
| 1995 | Nonhuman primate ES cells are maintained *in vitro* from the inner cell mass of monkeys |
| 1998 | James Thomson isolates human embryonic stem cells |
| 1999 | HHS legal opinion OKs research on hESC lines |
| 2000 | Scientists derive human ES cells from the inner cell mass of blastocysts |
| 2000 | ES proliferates *in vitro* and form all three germ layers and teratomas in immune-deficient mice |
| 2001 | President Bush prohibits federal funding of human embryonic stem cell research |
| 2004 | President's council on bioethic: "Monitoring Stem Cell Research" |
| 2004 | California Institute of Regenerative Medicine (CIRM): Prop. 71 worth $3 billion for 10 years |
| 2005 | National Academies (NAS) releases "Guidelines for Human Embryonic Stem Cell Research" |
| 2005 | President's Council on Bioethics: "Alternative Sources of Pluripotent Stem Cells" |
| 2006 | ISSCR guidelines for the "Conduct of Human Embryonic Stem Cell Research" |
| 2007 | New National Academies guidelines |
| 2007 | President Bush calls for work on alternate sources |
| 2007 | Yamanaka and Thomson independently derive iPS cells |
| 2008 | Only 16 of 21 lines eligible for federal funding were ethically derived |
| 2008 | NAS release new guidelines |
| 2008 | ISSCR releases guidelines for clinical translation |
| 2009 | President Obama reverses 2001 Executive Order |
| 2010 | Geron initiates clinical trial of human embryonic stem cell-based therapy |
| 2010 | Advanced cell technology wins FDA approval to test stem cell therapy for degenerative eye disease |
| 2011 | Geron discontinues clinical trial |

and development and the commercial products available, respectively. Cell therapy has also been designated as a disruptive technology [51] with its evolving roots in blood transfusion, bone marrow and organ transplantation, tissue banking, and reproductive *in vitro* fertilization. Cell-based therapies have progressed from the first recorded human to human blood transfusion by James Blundell (Guy's Hospital, London, UK) through to the advanced cellular therapies of today. This has been a 200-year journey based initially on clinical trial and error and only recently has

**TABLE 19.4 Current Approved Stem Cell-Based Commercial Products Compiled by the authors.**

| Company | Product | Product Information |
|---|---|---|
| Advanced BioHealing-Shire), CA | Dermagraft | Human fibroblast—dermal substitute |
| Altrika, UK | MySkin, Cryoskin | Allogeneic keratinocytes scaffold |
| Alphatec Spine, CA | PureGen | Osteoprogenitor cell allograft |
| BioDLogics, TN | BioDfactor | Micro scaffold—Bone and void filler |
| Arthro Kinetics, Germany | CaReS | Matrix-based autologous chondrocyte |
| Avita Medical, Australia | ReCell® Spray-On Skin | Combination of kertinocytes and melanocytes |
| Bio-Tissue, FL | Prokera, AmioGraft | Amniotic membrane bioscaffold |
| BioTissue Technologies, Germany | BioSeed-C, Chondrotissue | Autologous three-dimensional chondrocyte graft |
| Cytori, CA | Celution System | Adipose stem cells separation system |
| Dendreon, Seattle, WA | Provenge | Autologous immunotherapy—prostate cancer |
| Euroderm, Germany | Epidex, EpiGraft | Epidermal or epidermodermal skin substitutes |
| Fidia Farmaceuitici, Italy | Hyalograft 3D, Laserskin, Hyalograft C | Hyaluronic acid-based gels |
| Fibrocell, PA | laViv | Collagen fibroblast—anti-wrinkle |
| Genzyme-Sanofi Company), MA | Epicel, Carticel, MACI Implant | Cultured chondrocytes |
| Japan Tissue Engineering Co. | J-TEC Epidermis, J-TEC Cartilage, J-TEC Corneal Epithelium | Cultured epidermis, cultured corneal Epithelium cultured cartilage |
| Living Cell Technologies, Australia/NZ | DIABECELL | Porcine, insulin-producing islet cells |
| NuTech, AL | NuCel | Immunoprivileged amniotic member-transplant |
| NuVasive, CA | Osteocel Plus | Allograft bone graft |
| Organogenesis, MA | Appligraf | Bi layered skin substitue |
| Orthofix, TX | Trinity Evolution | Osteoprogenitors + dimineralized bone allograft |
| TiGenix, Brussels, Belgium | ChrondroCelect | Autologous cartilage cells |
| Therakos, NJ | Therakos Photopheresis | Lymphocyte-based immunotherapy |

extensive progress been made through laboratory sciences. These scientific studies have provided further understanding, newer values, and unique challenges to justify this field as a stand-alone distinct industry. The variety of cell therapeutics and translational technologies is expanding rapidly and currently is highly diversified and following broad-based approaches. The number of different types of cells used for therapies is also rapidly expanding, and the future holds even greater promise. In just 2009, the market for cell-based therapies in the United States was targeted toward about 100 million patients for indications including heart disease, diabetes, neurodegenerative diseases, and musculoskeletal disorders. It has been estimated that this field would generate more than $2 billion by 2012 and will generate more than $5 billion by 2014.

## 19.6 EMERGENCE OF TISSUE ENGINEERING AND SCAFFOLDS

As recently as the early 1980s, the term "tissue engineering" was loosely applied to the use of prosthetic devices and the surgical manipulation of tissues. That decade also saw the inaugural Keystone meeting in Colorado titled "Tissue Engineering," which was sponsored by the National Science Foundation. Boston was the birthplace of dedicated TE focus that combined the principles of biomedical engineering with biological sciences such as cell and molecular biology. The term "tissue engineering" was actually coined in a publication titled, "Functional Organ Replacement: The New Technology of Tissue Engineering" in *Surgical Technology International* in 1991 [52]. TE is the regeneration of biological tissue through the use of cells or biological scaffold materials (or both), and it offers the potential to create a paradigm shift in medicine. The goal is to develop new forms of therapy that allow the repair or regeneration of cells, tissues, and organs that have lost their function because of diseases, injuries, or age- or genetic-related defects. Potential applications of TE include protocols for bone; cartilage; soft tissue; skin; cardiovascular; neuronal; and other tissues such as bladder, kidney, liver, lungs, and others [53–55]. Parallel scientific advances in biomaterials, stem cells, growth and differentiation factors, and biomimetic environments have helped us to generate tissues in the laboratory from combinations of engineered extracellular matrices ("scaffolds"), cells, and biologically active molecules.

Combining stem cells with scaffolds was shown to be a viable approach for engineering tissues and for effective cellular delivery. In 1987, the word "biomaterial" was coined and defined as "a non-viable material used in a medical device, intended to interact with biological systems" (European Society of Biomaterials Conference, 1987). The field has advanced considerably since that time, and a series of innovative implantable scaffolds created with a wide variety of specific biomaterials have been developed. Such scaffolds are currently being explored to promote the viability and regulated differentiation of stem cells, and their role is to provide a structural base for regenerating tissue with the tensile compliance required for the planned implantation environment. The next paragraphs will briefly review the most commonly used scaffolds (also see Table 19.5, which

**TABLE 19.5 The World of Scaffolds and Processing Technologies**

| Natural Scaffolds | Artificial Scaffolds |
|---|---|
| Collagen, fibrin, fibrion, silk, agarose | Polyglycolic Acid (PGA), Poly (lactic acid) (PLA) |
| Glycosaminoglycans (GAGS), HA | PLGA, PLLA, PDLA, PDLLA, Poly(e-caprolactone) |
| Chitosan, Alginates, Agarose, Starch | Polyurethanes, Polymethanes, Polyorthoesters |
| Keratin, Decellularized organs/grafts | Polyanhydrides, ceramic, hydroxyapatites |

| Scaffold Processing: Modulating Scaffold Structure, Mechanical Tensile Strength, Permeability, and Porosity Features | | |
|---|---|---|
| Solvent Casting/Particulate Leaching | Freeze-Drying | Phase Inversion |
| High Pressure-Based Technologies | Fiber Bonding | Melt-based Technologies |
| Fused Deposition Modeling (FDM) | Electrospinning | 3D Printing |
| Selective Laser Sintering (SLS) | Stereolithography | Shape Depositing Manufacturing |
| Robotic Microassembly | Decellularization | |

lists different types of scaffolds and the processing methodologies currently used to create them) [56].

Many interesting studies worldwide make use of the ”scaffold and cells” method of TE. These include examples of bioartificial liver devices (using living hepatocytes), an artificial pancreas, an artificial kidney, and artificial bladders. This approach is also used to engineer cartilage, bone, skin, heart valves, and trachea, to name a few. A wide range of biomaterials, including both natural and synthetic biomaterials and scaffolds, has been developed for different applications [56].

Scaffolds constructed from natural materials include different derivatives of the naturally occurring extracellular matrix (ECM) proteins, such as collagen, fibrin polysaccharide–derived materials, chitosan, glycosaminoglycans, hyaluronic acid, and others. Collagen (which can be isolated from a variety of tissues, such as skin, tendon, or bone, and is usually of animal origin) has been shown to effectively support the culture of some stem cells, such s human mesenchymal stem cells derived from different tissues, and several proof-of-concept studies have been performed for a variety of applications, including engineering bone, ligaments, and cartilage [57,58]. Fibrin, which is derived from the enzymatic cleavage of fibrinogen by thrombin, is associated with a huge body of work describing the effective use of fibrin-based scaffolds in a variety of applications [59–61]. Another protein that has been investigated for use in generating tissue-engineered scaffolds is silk. Alternatively, silk fibroin has been shown to have a slower degradation rate than silk itself as well as desirable mechanical properties for engineering bone, cartilage, and ligament tissues [62,63]. Peptide-based biomaterials consist of short sequences of amino acids, which can produce self-assembling scaffolds. These scaffolds can potentially combine the functionality of protein-based scaffolds by using motifs derived from naturally occurring proteins with the reproducibility of synthetic scaffolds. Many of the peptide-based biomaterials can self-assemble into three-dimensional (3D)

scaffolds through the use of amphiphilic peptides. In addition, peptide-based scaffolds have been used for cell 3D cultures of neural stem cells (NSCs), and these constructs have served as intramyocardial microenvironments for endothelial cells, neurite outgrowth, and active synapse formation [64–66]. Finally, the polysaccharide-based biomaterials include agarose, alginate, hyaluronan, and chitosan, which have also been used for the culture and differentiation of stem cells [67–69].

Several synthetic biomaterials are currently used in a number of TE applications, and these provide an alternative to natural materials for the culture of stem cells and provide effective platforms for TE. The most commonly used synthetic materials, as seen in Table 19.5, include polyglycolic acid (PGA), poly(lactic acid) (PLA), polycaprolactone (PCL), and several other derivatized copolymers that have been shown to have a wide variety of applications in different TE models [70–73]. There are also the ceramic-based biomaterials, including bioactive glass and hydroxyapatite, a naturally occurring ceramic material. A synthetic version of hydroxyapatite hydrogel has also been investigated for use as a potential replacement for bone tissue because of its strength and biocompatibility [74–77].

Another scaffold material under intense investigation is acellular matrix, which is derived from the decellularization of various tissues. To create these, the cellular remnants are removed from an organ, leaving behind the ECM. This approach is attractive because the resulting matrix is similar or identical to that found in natural tissue in terms of topography, composition, microstructure, and biotensile properties. Decellularized matrices maximize the regeneration of both the structural and functional aspects of tissues and organs. Simple tissues as well as complicated organs have been decellularized, and optimized protocols have been developed to completely remove the cellular components while keeping the ECM intact [78–83]. Table 19.5 lists various processing approaches to formulate these scaffolds, but detailed descriptions of each of these processes are beyond the scope of this review. Nonetheless, it is worthwhile to mention the production of nanofiber scaffolds via electrospinning because more and more innovative scaffolds are being created using this method because it provides better biocompatibility, mechanical properties, porosity, permeability, and environment-specific compatibility [84–86].

The examples described here illustrate the importance of the extracellular environment in determining cell behavior in order to preserve function, survival, viability, and potency. However, as a result of both successful and unsuccessful trials in the field, the TE industry has benefited from a heavy dose of reality, and a number of lessons have been learned. Each of the above-mentioned scaffolds presents its own sets of advantages and disadvantages, and none are free of drawbacks at the present time. Several groups are further fine tuning biological scaffolds toward effective therapeutic use [87].

## 19.7 REGENERATIVE MEDICINE: LOOKING FORWARD

Health care management is constantly evolving, and continuous debates are ongoing on a global scale to manage care in both developed and developing countries. These

discussions have resulted in changes that have actually led to better health care, but the increased survival and lifespan resulting from these changes have led to an increase in the prevalence of diseases associated with aging. Many significant human diseases arising from the loss or dysfunction of specific cell types in the body, such as Parkinson's disease, diabetes, and cancer, are becoming increasingly common. In a remarkably short span of time, stem cells and the RM field have attracted an extraordinary level of interest from scientists, clinicians, industry, and the general public. The advances in the field and potential applications of stem cell research findings in RM are truly turning into global enterprises. Clinical translation of RM-based technology is now the key focus of several groups whose goals are the scientific advancement and practical therapeutic applications for the benefit of patients worldwide. This complex translational pipeline requires the interplay of research investigators, operational experts, investors, and clinicians who can work closely, and discussions with regulatory agencies within the organization, state, and federal (Food and Drug Administration [FDA]) levels; patient advocacies; ethics committees; manufacturing operations; and logistics are required for effective planning and higher quality successful outcomes.

Stem cell research and RM offer unique opportunities for developing new therapeutic approaches to prevent and treat these debilitating and life-threatening diseases, and they also provide new ways to explore fundamental questions of biology. The current optimism over potential stem cell therapies is driven by new understandings of genetics and developmental biology. Gradually, the curative and regenerative potential that lies in harnessing stem cells is being realized as the clinical applications of adult stem cells are investigated and used in the treatment of cancer; autoimmune, metabolic, and organ-related diseases; immunodeficiencies; and multiple age- and genetic-related anomalies.

## 19.8 CELL THERAPY, TISSUE ENGINEERING, AND THE GLOBAL LANDSCAPE OF REGENERATIVE MEDICINE

### 19.8.1 Stem Cell Therapies

Stem cell therapies are a new medical frontier. Currently, innovative work is being carried out on a global scale, and in some cases, the resulting therapeutics are already benefiting patients. These therapies involve a variety of adult progenitor cell types from a variety of tissues, including bone marrow, umbilical cord, umbilical cord matrix (Wharton's jelly), peripheral blood, adipose, amniotic fluid, placenta, endometrium, tooth pulp, eye limbus, hair follicles, and many other sources. The field is currently in its infancy, and as a result, it is facing challenges, including fragmentation, as studies with similar goals and approaches have shown that the variability in cell type production and effectiveness is not easily evaluated. However, recent research has validated multiple alternative sources of pluripotent stem cells, and this has led to a rapid development of adult cell reprogramming that has made this approach much safer for patients and brought the hope of pluripotent stem cell

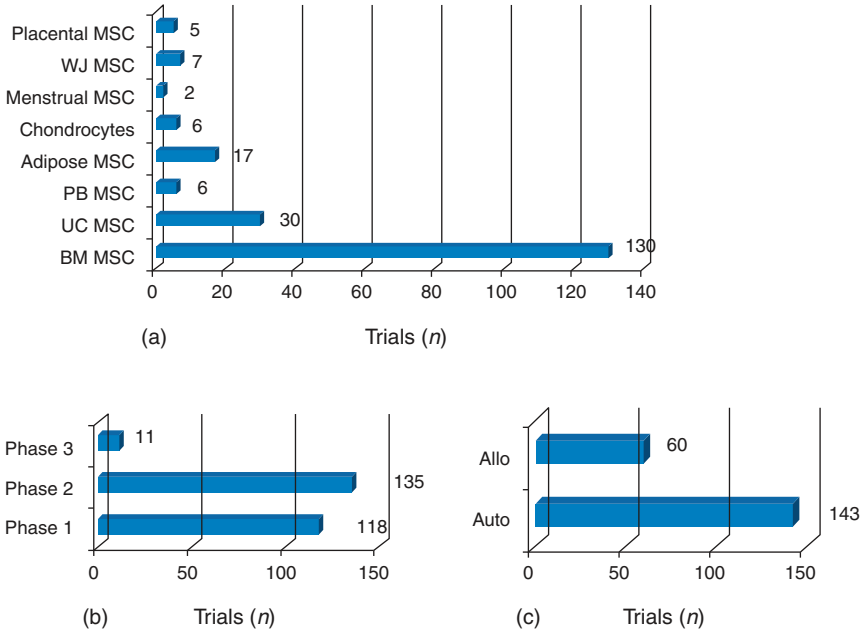


**FIGURE 19.1** Clinical trials with mesenchymal stem (stromal) cells (MSCs). (a), Tissue types used in MSC trials. (b), Different phases of MSC trials. (c), Autogenic (Auto) versus allogeneic (Allo) MSC trials. BM: bone marrow; PB: peripheral blood; UC: umbilical cord; WJ: Wharton's Jelly. (Adapted from http://www.clinicaltrials.gov.)

therapeutics much closer to a practical reality. As an example, Figure 19.1 depicts the number of ongoing trials using mesenchymal stem cells from different tissues and their results. Cell sources, clinical stage, and targeted indications (Table 19.6) are also presented.

As depicted in Figure 19.1 and Table 19.6, the number of different diseases and dysfunctional areas that are being targeted with cell and tissue therapy is also growing quickly, but there are several areas, including cardiovascular, orthopedic, and neuronal disorders as well as GVHD and diabetes, that are being investigated in a more aggressive format. Table 19.7 gives a glimpse of clinical studies being pursued in just the neurologic arena using both neuronal and non–neuronal-derived stem cells [88,89]. NSCs, their progenitors, and differentiated cell derivatives along with other stem cell–based clinical strategies have steadily moved forward to target central nervous system (CNS) diseases and injuries. Current phase I and II trials using stem cells in the CNS are the vanguard for the widely anticipated next generation of regenerative therapies and as such are pioneering the stem cell therapy process [88].

Stem cell science is extraordinarily promising. There have been great advances in treating diseases and conditions of the hematopoietic system using blood-forming stem cells, and these show us just how powerful stem cell therapies can be. Scientists

**TABLE 19.6 MSC Trials in Different Indications as Listed on www.clinicaltrials.gov**

| | | | |
|---|---|---|---|
| Acute burns | 2 | Liver Transplant | 2 |
| Acute Myocardial Infarction (AMI) | 22 | Maxillary Bone cyst | 2 |
| ALS | 4 | Middle cerebral artery infarct | 1 |
| Ankylosing spondylitis | 1 | Multiple Sclerosis (MS) | 4 |
| Aplastic Anemia | 3 | Multiple system atrophy | 1 |
| Alzheimer's Disease | 2 | Neuroblastoma | 2 |
| Articular Cartilage Repair | 7 | Neuromyelitis optica | 1 |
| Atherosclerosis | 3 | Organ Transplantation | 2 |
| Autism | 1 | Osteoarthritis | 10 |
| Bone Fracture | 14 | Osteonecrosis | 5 |
| Buerger's Disease | 2 | Parkinson's Disease (PD) | 3 |
| Cancer | 2 | Perianal Fistula | 1 |
| CABG | 1 | Peritonitis | 1 |
| Chronic allograft nephropathy | 1 | Post-Renal Transplantation | 5 |
| COPD | 2 | Primary Biliary Cirrhosis | 1 |
| Crohn's Disease | 7 | Promote Transplantation | 1 |
| Critical Limb Ischemia | 6 | Pulmonary Fibrosis | 3 |
| Deompensated Cirrhosis | 4 | Romberg's Disease | 1 |
| Diabetes-Type 1 | 11 | Sickle Cell Disease | 2 |
| Diabetes-Type 2 | 4 | Sjögren's syndrome | 1 |
| Diabetic Foot | 2 | SLE | 2 |
| Epidermolysis bullosa | 1 | Spinal Cord Injury | 11 |
| Erectile Dysfunction | 1 | Systemic Cirrhosis | 1 |
| Graft versus Host Disease (GVHD) | 16 | Ulcerative Colitis | 2 |
| Hepatic Artery | 2 | Ventricular Dysplasia | 1 |
| HIV | 1 | Wilson Cirrhosis | 1 |
| Idiopathic dilated Cardiomyopathy | 2 | | |
| Ischemic Stroke | 5 | | |
| Liver Cirrhosis | 7 | | |

all over the world are researching ways to harness stem cells and use them to learn about, diagnose, and treat various diseases and conditions. Every day, scientists are working on new ways to shape and control different types of stem cells in ways that are bringing us closer to new treatments. Many potential treatments are currently being tested in animal models, and some have already been brought to clinical trials. There is tremendous optimism that stem cell therapies will someday be available to treat a wide range of human diseases and conditions. The future is bright in this industry, and global groups are focused on achieving several curative milestones in the field. They are also developing stem cell–based therapies for effective disease management, immunomodulation, cosmetic, bioaesthetic, and effective drug delivery functions. Some specific examples of areas of current research in RM include projects such as replacement of skin for burn patients and patients with nonhealing wounds, pressure sores, or diabetic foot ulcers. There is also tremendous interest in bone and cartilage regeneration; bladder repair; vascular TE; nerve regeneration;

**TABLE 19.7 Application of Human Stem Cells for Different Human Neuropathies**

| Group/Company | Type of Study/Indication | Type of Cells |
|---|---|---|
| | *A. Using Neuronal-Derived Stem Cells* | |
| Advanced Cell Technologies Inc., CA | Phase I/II: Stargardt's muscular dystrophy | Allo-RPEs |
| | PhaseI/II: AMD | Allo-RPEs |
| Califiornia Stem cell Inc., CA | Phase I: Spinal muscular dystrophy | Allo-HuESC der motor neurons |
| Geron Corp. CA | Phase I: Spinal cord injury (trial discontinued) | Allo-hu ESC der oligodendrocytes |
| NeuralStem Inc., MD | Phase I: ALS | Allo-fetal-spinal der HuNSCs |
| NeuroGeneration Inc., CA | Phase I: Advanced Parkinson's disease | Auto-HuNSCs |
| ReNeuron Ltd., UK | Phase I: Stable ischemic stroke | Gen. mod. Allo-fetal-HuNSCs |
| Stem Cells Inc., CA | Phase I: Pelizaeus Merzbacher disease | Allo-HuCNS-SCs |
| | *B. Using Non-Neuronal-Derived Stem Cells* | |
| Aldagen Inc., CA | Phase II: Postacute ischemic stroke | Auto-adult HuBMSCs |
| Athersys Inc., OH | Phase I: Acute ischemic stroke | Allo-adult HuB-mutistem cells |
| BrainStorm Cell Therapeutics Inc. Israel | Phase I/II: ALS | Auto-adult BMSCs |
| China Medical University Hospital | Phase II: Chronic ischemic stroke | Auto-adult-HuPBSCs |
| Duke University, NC | Phase I/II: Spastic celebral palsy | Auto-adult-HuUCBS |
| J&J | Phase I: Acute ischemic stroke | Allo-adult-HuUCBs-derived cells |
| Memorial Hermann Healthcare System, TX | Phase I: Chronic spinal cord injury | Auto-HuBMCs |
| SanBio Inc., CA | Phase I/II: Chronic ischemic stroke | Allo-adult HuMSCs |
| Stemedica Inc., CA | PhaseI/II: Chronic ischemic stroke | Allo-adult BMSCs |
| TCA Cellular Therapy LLC, LA | Phase I: ALS | Auto-adult BMSCs |
| University of Texas, TX | Phase I: Acute TBI | Auto-Human UCBs |

AMD, age-related macular degeneration; RPE, retinal pigmented epithelial cells; Hu, human, ESC, embryonic stem cells; ALS, amyotrophic lateral sclerosis (Lou Gerhigs disease), NSC, Neuronal stem cells; BMSCs, bone marro-derived mesenchymal stem cells; PBSCs, peripheral blood stem cells; UCB, umbilical cord blood stem cells; TBI, traumatic brain injury.

restoration of peripheral nerve or spinal cord after injury; restoration of the cells of the CNS after tumor resection, disease, or injury; and regeneration of pancreatic tissue to produce insulin. By all measures, we are quickly entering a new era in the treatment of disease and injury: the era of stem cell therapies. Progress in this field was initially slowed because the use of embryonic stem cells raised serious ethical issues regarding the destruction of embryos and often gave rise to cancer and other side effects. Since the discovery of adult stem cells, research has moved forward at an accelerated pace and continues rapidly in all directions.

### 19.8.2 Tissue Engineering

Tissue engineering and cell therapy comprise a market for regenerative products that is expected to grow in the United States from $9.9 billion in 2011 to almost $32 billion by 2018 [90]. This market spans many specialties, the biggest of which are orthopedic and spine applications, cardiac and vascular disease, a wide range of neurologic disorders, diabetes, inflammatory diseases, and dental injuries. Competition in the field of TE and cell therapy is intense, with new companies being founded every day. The current industry focus is on disease areas that have the most patients and the largest unmet needs. These include cancer therapies, diabetes, orthopedics, and cardiovascular products, although several groups are also focusing their studies on diseases with smaller patient populations and a less crowded playing field, including Alzheimer's and Parkinson's diseases, multiple sclerosis, other neurodegenerative diseases, and ophthalmology. Key factors expected to influence the market for RM are continued political actions; venture capital; private equity and endowment funding; government funding; and particularly, positive clinical trial results. In addition, industry, pharmaceutical involvement, and investments and an increasing awareness among both physicians and the general public are driving this market. Several technologies in this segment, such as bone and skin grafts, are already being marketed. Growth in these markets will be sustained as the population ages and continues to require these products. Applications of RM in cardiology and vascular medicine will also be developed, and these will be rapidly adopted into clinical practice over the next decade. These include treatments of coronary artery disease, myocardial infarction, congestive heart failure, dysfunctional heart valves, peripheral vascular disorders, and abdominal aortic aneurysms [51,91–93].

However, challenges persist with interfaces between tissues and devices, including approaches intended to improve biocompatibility and integration between cells and novel devices, materials that self-assemble within the body, the fate of biomaterials and implant materials within the environment of the body, and optimal interaction between cells and the morphology and topography of surfaces and coatings. Also, there is a continuous need to look at newer varieties of synthetic scaffolds that are more effective in promoting 3D cell growth and vascularization and that provide a structure for effective assembly in a biofunctional tissue (or ultimately an organ). Possible features of novel materials include multiple characteristics, including the ability to react to external inputs in a controlled manner, the ability to attract endogenous progenitor cells using molecular and chemical cues, and the

ability to react and adapt to changes in physiological parameters. In addition, a number of studies have been designed to determine the optimal porosity and permeability of biomaterials, which facilitate vascularization and innervation of the new tissue with suitable mechanical integrity. Key scientific hurdles still remain in cell therapy and TE, such as graft rejection, and attempts to make the product less or nonimmunogenic are underway. In terms of replacement organs, new organs could be made outside the body or grown within the body. The latter situation may be more achievable. Regardless, creating entirely new organs remains challenging, although replacement hair, skin, teeth, and pancreatic islets may be achieved in the next 10 years, and the kidney is likely to be the first major organ to be created outside the body [51,91–93].

## 19.9 GLOBAL REGENERATIVE MEDICINE AND MEDICAL TOURISM

The current financial crisis has brought economic tightening and challenges raising capital for many industries, including the global RM landscape. This includes nonprofit and for-profit biotech groups and organizations. The health care business around the world is also changing as less expensive options for treatment are provided and cost and time-effective offshore locations for clinical trials and manufacturing are developed. The forces that are shaping the evolution of health care also include integration of worldwide personalized medicine with a lower cost and higher quality of care provided by accredited personnel and institutions that are supported by reliable and transparent information technologies. This rapidly changing environment is driven by unmet clinical needs, with the United States as the largest market trailed by Japan and Europe. In terms of growth, Europe is projected to march ahead of other world markets, propelled by renewed activity in Germany, the United Kingdom, and Belgium, among others [94,95].

Regenerative medicine is touted as the next-generation revolutionizing mode of modern medicine. Active support of governments around the world and technological breakthroughs are expected to result in high growth for the RM industry in the near future. RM also has the potential to emerge as one of the largest drivers of growth for the global economy. The global aging population is plagued with various age-related dysfunctions, and this population provides a readily available market for novel therapies based on stem cell therapy, TE, and various other tools of RM. Unlike other industries, the RM industry has reached its present stage mainly through private participation, but now that trend is changing as government and endowment funding is on the rise. China has set up a US$1.0 billion fund for research into this field. Other major players, such as Australia, India, Brazil, Argentina, South Korea, Singapore, and Malaysia, have been obtaining active support from their respective governments through funding or beneficial policies. The US government support mechanism for RM-driven technologies is the highest in the world, and it includes multiple federal and state-run granting mechanisms.

It can be a difficult process to ensure that policies reflect the desire for progress in disease management while maintaining the collective views of a country. For

example, research on stem cells is still fraught with ethical issues, both within and among nations. The alterations in policy that are occurring around the world clearly reveal a belief in the exciting therapeutic potential of RM, but this is coupled with moral criticisms of the investigation into the stem cell [94].

Big pharma, after maintaining a long-term watching stance, is also gradually investing in RM biotechnologies, indicating a significant, although slow-paced, shift in attitude. Much of the concern surrounding corporate investments was due to an abundance of unproven clinical trial data and the ethical implications associated with this type of research. With more phase I and II clinical trials advancing to later phases, big pharma is now showing more interest in RM. A distinctive prototype is Athersys, which entered into a global agreement with Pfizer ($6 million in initial research funding and $105 million after reaching the first commercial milestone) to develop and market its MultiStem technologies designed to treat inflammatory bowel disease, acute myocardial infarction, and GVHD. Cephalon (now Teva) made a deal to acquire a 20% stake in the Australian RM company Mesoblast, which is currently in phase II trials of its leading cardiovascular treatment product, Revascor. This alliance focuses on commercializing the adult stem cell products used to treat degenerative CNS and cardiovascular conditions. Obtaining worldwide rights for specific Mesoblast products, Cephalon fronted $130 million, with another $220 million to follow. More big pharma companies are trying to develop or obtain approaches of targeting, acquiring, and initiating RM programs.

According to New Global Industry Analysts, Inc., global investments in adult stem cells research are forecast to reach US$2.4 billion by the year 2015. Technological advancements, the rising number of research groups engaged in stem cell research activities, the broadening of these activities, and substantial investments from governments in leading global markets are the key factors driving growth in stem cell research worldwide. Additionally, the market is driven by the rising incidence of life-threatening diseases, large unmet needs in medicine, and the rising costs of drug development.

Despite the common economic downturn, there are signs that RM industry continues to grow in different parts of the world. Manufacturing, clinical studies, and so-called "medical tourism" with the aim of obtaining access to stem cell and tissue-engineered products are on the rise in areas such as China, India, South Korea, and Thailand. Using India as an example, hundreds of patients are now streaming into stem cell therapy clinics all over India even though there is still a big gap between the reality and the future potential in the field worldwide. In India, more than 25 public and several private research institutions are authorized to conduct stem cell research. The only stem cell therapy treatment tried and tested so far is bone marrow transplantation, allowed by the Indian Council of Medical Research. One reason for the fuss and hype surrounding stem cell therapy is its potential to address a plethora of medical conditions. Nevertheless, several hospitals and clinics across major cities in India have treatments for neuronal, cardiac, and reproductive areas of medicine, the largest number being in CNS diseases and soft tissue repair. In 2007, India brought out a set of guidelines for conducting stem cell research, but the

guidelines are not legally binding and have no power to curb clinical implementation of stem cell therapy. The situation is changing now as the government is intervening and underlining policies for RM registration and clinical trial practices. Fifteen clinical trials have been officially registered so far, and efforts have been implemented to curb or stop unregistered work. India has taken up the challenge, and the next 5 to 10 years will be the most defining period for the nascent stem cell industry in India, as the National Apex Committee for Stem Cell Research and Therapy (NAC-SCRT) has been set up by the Union Ministry of Health. This is a stepping stone for the Indian industries, entrepreneurs, and companies to take stem cell research and cell-based therapies to next level. The development of the NAC-SCRT has been very timely in India, and it will play a crucial role in fostering cooperation and productivity from the emerging new companies in the stem cell industry. This should serve as a good model for other international programs in southeast Asia and Latin America (Inter Press Service, India, 2011).

## 19.10 REGULATORY HURDLES

As discussed earlier, medical tourism to obtain unproven, nonlegalized, and unregulated RM technologies is a big problem not just in India but also in other attractive medical touristic sites in Eastern Europe, Latin America, the Middle East, and southeast Asia. These areas must be regulated, and the United States has implemented several initiatives similar to the ones followed by the Indian government.

The FDA recently acknowledged that clinical research has become increasingly global and issued guidance to help sponsors demonstrate that foreign clinical trials conformed to good clinical practice (GCP) standards and the agency's regulations, especially groups that did not conduct their foreign trials under an Investigational New Drug Application (IND). The guidance, *FDA Acceptance of Foreign Clinical Studies Not Conducted Under an IND Frequently Asked Questions*, follows up on a final rule (21 C.F.R. §312.120) issued in April 2008 (7 MRLR 262, 5/7/08) that made good GCP the standard of compliance for foreign clinical studies that are not being conducted under an IND, thereby replacing the Declaration of Helsinki the principles in the regulations. Although sponsors are not required to conduct foreign clinical studies under an IND, they must ensure that the study meets all the rule's requirements to use the study as support for an IND or application for marketing approval. The final rule requires review and approval by an independent ethics committee and informed consent from subjects along with ethical and data integrity standards for clinical studies, which appeared in the *Federal Register* (73 Fed. Reg.22800,4/28/08MedicalResearchLaw&PolicyReport).

The International Society for Stem Cell Research (ISSCR) and the International Society for Cellular Therapy have taken a critical position regarding "stem cell tourism" and unregulated commercial "cell therapies" around the world by launching guidelines dedicated to tracking commercial clinics and providing recommendations for stem cell therapy seekers that emphasize the following key points: (1) there are

several adult tissue types of stem cells with limited potentials to regenerate, (2) one type of cell therapy will not work on a wide variety of conditions, (3) there are very few accepted stem cell therapies, (4) clinics advertising through ad-hoc patient testimonials are suspicious, (5) a regulated and effective cell therapy process takes scientific innovation and a longer time, (6) a treatment is not necessarily safe even if it is autologous, and (7) an experimental treatment offered for sale is not the same as a clinical trial and can just be an expensive gouging of desperate patients. Responsibly conducted clinical trials are critical to the development of new treatments because they allow us to learn whether these treatments are safe and effective. The ISSCR supports participation in responsible clinical trials after careful consideration of the issues highlighted on its site (http://www.iscr.org) and after discussion with a trusted and accredited physician (or physicians).

Tissue-engineered and RM product development processes are often combinations of cells, scaffolds, and other factors and are complex in both structure and function. This introduces challenges for product developers to establish efficient manufacturing and characterization techniques to ensure that these products are safe and effective before clinical trials are conducted in humans. It is critical for early stage RM product developers to establish an FDA dialogue and relationship for their valuable feedback and advice highlighting key scientific considerations for strategic planning applicable specifically to their products [96].

Other critical components for effective development of RM therapies, not included in this review, pertain to establishment of proof of concept, statistically relevant preclinical animal (*in vivo*) studies, intellectual proprietary protection, optimization of process development, current good manufacturing practices, and building a clinical team to conduct the planned study in a cost-effective manner. These are all important pieces of the developmental puzzle needed to overcome translational barriers. Also, a worthwhile effort will be to develop clinical companion tools and diagnostic means that can evaluate the functional efficacy of the therapeutic modality in a clinical setup.

It has been 13 years since the discovery of human embryonic stem cells (hESCs). Companies such as Geron and Advanced Cell Technology (ACT) were leading multiple clinical programs, but unfortunately, Geron's hESC program to treat spinal cord injury was discontinued, and continual challenges in the hESC translation field remain because of ethical and practical issues with these cells. However, a preliminary report from ACT was published describing hESC-derived retinal pigment epithelium cell transplantation into patients with age-related macular degeneration with no signs of hyperproliferation, tumorigenicity, ectopic tissue formation, or apparent rejection after 4 months [97].

It is also worth mentioning the growing field of cancer stem cells (CSCs). Many groups currently aim to identify specific biomarkers of CSCs, especially the ones that play important roles in metastasis and tissue- or organ-specific invasion. Other features of CSCs are that they are highly tumorigenic and pluripotent and have the ability to self-renew. Many CSCs have been shown to possess multidrug resistance capacities, and they likely play a role in cancer relapse [98]. Further study of these cells will also help in devising new anticancer therapeutics in the near future.

## 19.11 THE FUTURE AND CHALLENGES

The RM product pipeline will hopefully open new venues and alternatives to organ transplantation for Americans on chronic mechanical and extracorporeal therapies. The field is continuously growing. Baby boomers will nearly double the number of Americans older than age 65 years by 2025, and the number of Americans without health care insurance will exceed 18% by 2015. US health care expenditures on treatment of acute illness, chronic illness, and long-term support of individuals in chronically debilitated states may double from $3 to $6 trillion, thus approaching 40% of US gross domestic product (GDP). If stem cell–based tissue regeneration enhances quality of life for aging Americans, potentially $3 to $5 trillion dollars of the US GDP can be redirected.

The RM field has evolved significantly in the past 15 years. We are now at a point where it appears that this industry will play a significant role in how we treat a range of diseases in the future. Major corporations such as Pfizer, Teva, Roche, Sanofi, Novartis, and Johnson & Johnson have created internal RM programs and are actively forming collaborations with researchers and companies in the field. Although a key focus for the field remains the need to demonstrate clinical efficacy through execution of large, well-designed clinical trials necessary to gain market approval, an equally important challenge is developing commercialization models for how RM will be delivered to the general population in a cost-effective way. The other evolving component for future growth in the field will be the development of efficient, effective product development and commercialization strategies.

To build tissues and organs, scientists are now beginning to understand the complexity of the many ways tissues can interact with each other. To achieve the envisioned goals of RM, a strong federally directed initiative is needed and is hopefully underway now to ensure that these fundamental research paths are realized and that translation to clinical applications is quickly and efficiently achieved. Current well-funded and focused translational research sponsorship efforts from federal agencies such as the National Institutes of Health, DOD, and AFIRM and state-based funding will potentially help scientist achieve functional repair of damaged tissues within 10 years and possibly full organ regeneration within 20 years. There is a continuous need for a more active role of the FDA in recognizing the challenges of these new medical technologies because well-defined guidelines, transparent support outreach, and education of researchers regarding the regulation process are required. In addition, more effective evaluation, support, and advisory roles are required for complex combinatorial products. This commitment to the future is encouraging, and the FDA remains actively involved in multiple innovative RM developmental paths.

Other countries have already embarked upon national initiatives of their own with hopes of making RM a reality for themselves. Several members of the European Union, including Great Britain, Germany, and Sweden, as well as Japan, China, India, and Australia have all begun making strong national commitments with hopes of achieving their own advances in RM technology.

Also of importance is the consolidation of synergistic efforts by representations from nonprofit, for-profit communities, patient advocacy groups, and government agencies for reimbursement and other outreach efforts to general public. US forums such as the Genetic Policy Institute and the Alliance for Regenerative Medicine (http://www.alliancerm.org) specifically represent the interests of RM companies, institutions, patients, policymakers, and researchers. The Regenerative Medicine Promotion Act of 2011 (HR 1862) was introduced in the House of Representatives on May 12, 2011, by Reps. DeGette (D-CO) and Bilbray (R-CA) with provisions including creation of a multiagency Regenerative Medicine Coordinating Council within the Department of Health & Human Services, establishing grant programs to accelerate the availability of life-saving RM therapies and research tools, funding of critical regulatory research at the FDA, and a detailed assessment of federal activities in RM as well as progress compared with national programs in other countries.

Today the RM industry is exclusively focused on translating science into commercial products, thus integrating the science into the health care system with effective production of automated support and servicing tools. Although there are many challenges, there is a strong base of public support for these modern RM opportunities to create curative therapeutics for a brighter future in disease management [99–102].

## REFERENCES

1. Jasny BR, Zahn LM. Genome-sequencing anniversary. A celebration of the genome, part I. Science 2011;331(6017):546.
2. Sekine H, Shimizu T, Okano T. Myocardial tissue engineering: toward a bioartificial pump. Cell Tissue Res 2011;347(3):775–782.
3. Muneoka M, Han M, Gardiner DM. Regrowing human limbs. Sci Am 2008; 298(4):56–63.
4. Murray JE, Merrill JP, Harrison JH. Renal homotransplantation in identical twins. 1955. J Am Soc Nephrol 2001;12(1):201–204.
5. Toledo-Pereyra LH, Toledo AH. 1954. J Invest Surg 2005;18(6):285–290.
6. Perera MT, Mirza DF, Elias E. Liver transplantation: issues for the next 20 years. J Gastroenterol Hepatol 2009;24(Suppl 3):S124–S131.
7. Starzl TE, Fung JJ. Themes of liver transplantation. Hepatology 2010;51(6):1869–1884.
8. Kirby TJ, Mehta A. Lung transplantation: state of the art. Appl Cardiopulm Pathophysiol 1992;4(4):263–271.
9. Patel N, Criner GJ. Transplantation in chronic obstructive pulmonary disease. COPD 2006;3(3):149–162.
10. Vardanyan M, Parkin E, Gruessner C, Rodriguez Rilo HL. Pancreas vs. islet transplantation: a call on the future. Curr Opin Organ Transplant 2010;15(1):124–130.
11. Tuna IC, Jamieson SW. Human heart and lung transplantation. Adv Surg 1989;22: 251–276.
12. Shapiro R, Sarwal MM. Pediatric kidney transplantation. Pediatr Clin North Am 2010;57 (2):393–400, table of contents.

13. Ezhkova E, Fuchs E. Regenerative medicine: an eye to treating blindness. Nature 2010;466(7306):567–568.
14. Kahan BD. Fifty years in the vineyard of transplantation: looking back. Transplant Proc 2011;43(8):2853–2859.
15. Morris PJ. Transplantation—a medical miracle of the 20th century. N Engl J Med 2004;351(26):2678–2680.
16. Murray JE. The 50th anniversary of the first successful human organ transplant. Rev Invest Clin 2005;57(2):118–119.
17. Watson CJ, Dark JH. Organ transplantation: historical perspective and current practice. Br J Anaesth 2012;108(Suppl 1):i29–i42.
18. Rudge CJ, Buggins E. How to increase organ donation: does opting out have a role? Transplantation 2012;93(2):141–144.
19. Chinen J, Buckley RH. Transplantation immunology: solid organ and bone marrow. J Allergy Clin Immunol 2010;125(2 Suppl 2):S324–S335.
20. Beitinjaneh A, Burns LJ, Majhail NS. Solid organ transplantation in survivors of hematopoietic cell transplantation: a single institution case series and literature review. Clin Transplant 2010;24(4):E94–E102.
21. Sansone F, Rinaldi M. Cyclosporine monotherapy in cardiac transplantation: review of the literature. Transplant Rev (Orlando) 2011;25(4):131–135.
22. Meier-Kriesche HU, Li S, Gruessner RW, Fung JJ, Bustami RT, Barr ML, et al. Immunosuppression: evolution in practice and trends, 1994–2004. Am J Transplant 2006;6(5 Pt 2):1111–1131.
23. de la Morena MT, Gatti RA. A history of bone marrow transplantation. Hematol Oncol Clin North Am 2011;25(1):1–15.
24. Cleland SY, Siegel RM. Wiskott-Aldrich syndrome at the nexus of autoimmune and primary immunodeficiency diseases. FEBS Lett 2011;585(23):3710–3714.
25. Ochs HD, Notarangelo LD. X-linked immunodeficiencies. Curr Allergy Asthma Rep 2004;4(5):339–348.
26. Okamoto Y, Kodama Y, Nishikawa T, Yamaki Y, Mougi H, Masamoto I, et al. Successful bone marrow transplantation for children with aplastic anemia based on a best-available evidence strategy. Pediatr Transplant 2010;14(8):980–985.
27. MacMillan ML, Davies SM, Nelson GO, Chitphakdithai P, Confer DL, King RJ, et al. Twenty years of unrelated donor bone marrow transplantation for pediatric acute leukemia facilitated by the National Marrow Donor Program. Biol Blood Marrow Transplant 2008;14(9 Suppl):16–22.
28. Bielorai B, Hughes MR, Auerbach AD, Nagler A, Loewenthal R, Rechavi G, et al. Successful umbilical cord blood transplantation for Fanconi anemia using pre-implantation genetic diagnosis for HLA-matched donor. Am J Hematol 2004; 77(4):397–399.
29. Barrett AJ. Bone marrow transplantation for acute lymphoblastic leukaemia. Baillieres Clin Haematol 1994;7(2):377–401.
30. Brunstein CG. Umbilical cord blood transplantation for the treatment of hematologic malignancies. Cancer Control 2011;18(4):222–236.
31. Lin SJ, Yan DC, Lee YC, Hsiao HS, Lee PT, Liang YW, et al. Umbilical cord blood immunology: relevance to stem cell transplantation. Clin Rev Allergy Immunol 2012; 42(1):45–57.

32. Rao M, Ahrlund-Richter L, Kaufman DS. Concise review: cord blood banking, transplantation and induced pluripotent stem cell: success and opportunities. Stem Cells 2012;30(1):55–60.
33. Wall DA. Regulatory issues in cord blood banking and transplantation. Best Pract Res Clin Haematol 2010;23(2):171–177.
34. Butler MG, Menitove JE. Umbilical cord blood banking: an update. J Assist Reprod Genet 2011;28(8):669–676.
35. McKenna D, Sheth J. Umbilical cord blood: current status & promise for the future. Indian J Med Res 2011;134(3):261–269.
36. Arrojo IP, Lamas Mdel C, Verdugo LP, Alfaro PR, Pena RR, Gordo FS, et al. Trends in cord blood banking. Blood Transfus 2012;10(1):95–100.
37. Confer D, Robinett P. The US National Marrow Donor Program role in unrelated donor hematopoietic cell transplantation. Bone Marrow Transplant 2008;42(Suppl 1):S3–S5.
38. Welte K, Foeken L, Gluckman E, Navarrete C. International exchange of cord blood units: the registry aspects. Bone Marrow Transplant 2010;45(5):825–831.
39. Majhail NS, Murphy EA, Denzen EM, Ferguson SS, Anasetti C, Bracey A, et al. The National Marrow Donor Program's Symposium on Hematopoietic Cell Transplantation in 2020: a health care resource and infrastructure assessment. Biol Blood Marrow Transplant 2012;18(2):172–182.
40. Bolan CD, Hartzman RJ, Perry EH, Trainor L, Miller J, Miller R, et al. Donation activities and product integrity in unrelated donor allogeneic hematopoietic transplantation: experience of the National Marrow Donor Program. Biol Blood Marrow Transplant 2008;14(9 Suppl):23–28.
41. Korbling M, Freireich EJ. Twenty-five years of peripheral blood stem cell transplantation. Blood 2011;117(24):6411–6416.
42. Bijou F, Ivanovic Z, Boiron JM, Nicolini F. Hematopoietic stem cells mobilization: state of the art in 2011 and perspectives. Transfus Clin Biol 2011;18(5–6):503–515.
43. Walsh CE, Nienhuis AW, Samulski RJ, Brown MG, Miller JL, Young NS, et al. Phenotypic correction of Fanconi anemia in human hematopoietic cells with a recombinant adeno-associated virus vector. J Clin Invest 1994;94(4):1440–1448.
44. Risitano AM, Maciejewski JP, Selleri C, Rotoli B. Function and malfunction of hematopoietic stem cells in primary bone marrow failure syndromes. Curr Stem Cell Res Ther 2007;2(1):39–52.
45. Rettig MP, Ansstas G, DiPersio JF. Mobilization of hematopoietic stem and progenitor cells using inhibitors of CXCR4 and VLA-4. Leukemia 2012;26(1):34–53.
46. Moreb JS, Salmasinia D, Hsu J, Hou W, Cline C, Rosenau E. Long-term outcome after autologous stem cell transplantation with adequate peripheral blood stem cell mobilization using Plerixafor and G-CSF in poor mobilizer lymphoma and myeloma patients. Adv Hematol 2011;2011:517–561.
47. Zhang H, Chen J, Que W. Allogeneic peripheral blood stem cell and bone marrow transplantation for hematologic malignancies: meta-analysis of randomized controlled trials. Leuk Res 2012;36(4):431–437.
48. Bensinger WI. Allogeneic transplantation: peripheral blood vs. bone marrow. Curr Opin Oncol 2012;24(2):191–196.
49. Eng HS, Leffell MS. Histocompatibility testing after fifty years of transplantation. J Immunol Methods 2011;369(1–2):1–21.

50. Cecka JM. Current methodologies for detecting sensitization to HLA antigens. Curr Opin Organ Transplant 2011;16(4):398–403.
51. Mason C, Brindley DA, Culme-Seymour EJ, Davie NL. Cell therapy industry: billion dollar global business with unlimited potential. Regen Med 2011;6(3):265–272.
52. Langer R, Vacanti JP. Tissue engineering. Science 1993;260(5110):920–926.
53. Atala A. Engineering organs. Curr Opin Biotechnol 2009;20(5):575–592.
54. Raya-Rivera A, Esquiliano DR, Yoo JJ, Lopez-Bayghen E, Soker S, Atala A. Tissue-engineered autologous urethras for patients who need reconstruction: an observational study. Lancet 2011;377(9772):1175–1182.
55. Olson JL, Atala A, Yoo JJ. Tissue engineering: current strategies and future directions. Chonnam Med J 2011;47(1):1–13.
56. Carletti E, Motta A, Migliaresi C. Scaffolds for tissue engineering and 3D cell culture. Methods Mol Biol 2011;695:17–39.
57. Chan BP, Hui TY, Yeung CW, Li J, Mo I, Chan GC. Self-assembled collagen-human mesenchymal stem cell microspheres for regenerative medicine. Biomaterials 2007; 28(31):4652–4666.
58. Lee CH, Singla A, Lee Y. Biomedical applications of collagen. Int J Pharm 2001; 221(1–2):1–22.
59. Catelas I, Sese N, Wu BM, Dunn JC, Helgerson S, Tawil B. Human mesenchymal stem cell proliferation and osteogenic differentiation in fibrin gels *in vitro*. Tissue Eng 2006; 12(8):2385–2396.
60. Ahmed TA, Dare EV, Hincke M. Fibrin: a versatile scaffold for tissue engineering applications. Tissue Eng Part B Rev 2008;14(2):199–215.
61. Shaikh FM, Callanan A, Kavanagh EG, Burke PE, Grace PA, McGloughlin TM. Fibrin: a natural biodegradable scaffold in vascular tissue engineering. Cells Tissues Organs 2008;188(4):333–346.
62. Wang Y, Kim HJ, Vunjak-Novakovic G, Kaplan DL. Stem cell-based tissue engineering with silk biomaterials. Biomaterials 2006;27(36):6064–6082.
63. Zhang X, Reagan MR, Kaplan DL. Electrospun silk biomaterial scaffolds for regenerative medicine. Adv Drug Deliv Rev 2009;61(12):988–1006.
64. Holmes TC, de Lacalle S, Su X, Liu G, Rich A, Zhang S. Extensive neurite outgrowth and active synapse formation on self-assembling peptide scaffolds. Proc Natl Acad Sci USA 2000;97(12):6728–6733.
65. Davis ME, Motion JP, Narmoneva DA, Takahashi T, Hakuno D, Kamm RD, et al. Injectable self-assembling peptide nanofibers create intramyocardial microenvironments for endothelial cells. Circulation 2005;111(4):442–450.
66. Gelain F, Bottai D, Vescovi A, Zhang S. Designer self-assembling peptide nanofiber scaffolds for adult mouse neural stem cell 3-dimensional cultures. PLoS One 2006;1: e119.
67. Ma PX. Scaffolds for tissue fabrication. In: Materials Today. Kidlington, UK: Elsevier; 2004, p. 30–40.
68. Zilla PP, Greisler HP. Tissue engineering of vascular prosthetic grafts. R.G. Landes Co; 1999.
69. Macchiarini P, Jungebluth P, Go T, Asnaghi MA, Rees LE, Cogan TA, et al. Clinical transplantation of a tissue-engineered airway. Lancet 2008;372(9655):2023–2030.

70. Choi YS, Park SN, Suh H. Adipose tissue engineering using mesenchymal stem cells attached to injectable PLGA spheres. Biomaterials 2005;26(29):5855–5863.
71. Buxton AN, Zhu J, Marchant R, West JL, Yoo JU, Johnstone B. Design and characterization of poly(ethylene glycol) photopolymerizable semi-interpenetrating networks for chondrogenesis of human mesenchymal stem cells. Tissue Eng 2007;13(10):2549–2560.
72. Bhang SH, Lim JS, Choi CY, Kwon YK, Kim BS. The behavior of neural stem cells on biodegradable synthetic polymers. J Biomater Sci Polym Ed 2007;18(2):223–239.
73. Ashton RS, Banerjee A, Punyani S, Schaffer DV, Kane RS. Scaffolds based on degradable alginate hydrogels and poly(lactide-co-glycolide) microspheres for stem cell culture. Biomaterials 2007;28(36):5518–5525.
74. Kim IL, Mauck RL, Burdick JA. Hydrogel design for cartilage tissue engineering: a case study with hyaluronic acid. Biomaterials 2011;32(34):8771–8782.
75. Macaya D, Spector M. Injectable hydrogel materials for spinal cord regeneration: a review. Biomed Mater 2012;7(1):012001.
76. Ricchetti ET, Aurora A, Iannotti JP, Derwin KA. Scaffold devices for rotator cuff repair. J Shoulder Elbow Surg 2012;21(2):251–265.
77. Martin RA, Yue S, Hanna JV, Lee PD, Newport RJ, Smith ME, et al. Characterizing the hierarchical structures of bioactive sol-gel silicate glass and hybrid scaffolds for bone regeneration. Philos Transact A Math Phys Eng Sci 2012;370(1963):1422–1443.
78. Badylak SF, Taylor D, Uygun K. Whole-organ tissue engineering: decellularization and recellularization of three-dimensional matrix scaffolds. Annu Rev Biomed Eng 2011; 13:27–53.
79. Wolf MT, Daly KA, Reing JE, Badylak SF. Biologic scaffold composed of skeletal muscle extracellular matrix. Biomaterials 2012;33(10):2916–2925.
80. Hoshiba T, Lu H, Kawazoe N, Chen G. Decellularized matrices for tissue engineering. Expert Opin Biol Ther 2010;10(12):1717–1728.
81. Wu S, Liu Y, Bharadwaj S, Atala A, Zhang Y. Human urine-derived stem cells seeded in a modified 3D porous small intestinal submucosa scaffold for urethral tissue engineering. Biomaterials 2011;32(5):1317–1326.
82. Baptista PM, Siddiqui MM, Lozier G, Rodriguez SR, Atala A, Soker S. The use of whole organ decellularization for the generation of a vascularized liver organoid. Hepatology 2011;53(2):604–617.
83. Jordan JE, Williams JK, Lee SJ, Raghavan D, Atala A, Yoo JJ. Bioengineered self-seeding heart valves. J Thorac Cardiovasc Surg 2012;143(1):201–208.
84. Semino CE, Kasahara J, Hayashi Y, Zhang S. Entrapment of migrating hippocampal neural cells in three-dimensional peptide nanofiber scaffold. Tissue Eng 2004; 10(3–4):643–655.
85. Szentivanyi AL, Zernetsch H, Menzel H, Glasmacher B. A review of developments in electrospinning technology: new opportunities for the design of artificial tissue structures. Int J Artif Organs 2011;34(10):986–997.
86. Chen DW, Liao JY, Liu SJ, Chan EC. Novel biodegradable sandwich-structured nanofibrous drug-eluting membranes for repair of infected wounds: an *in vitro* and *in vivo* study. Int J Nanomed 2012;7:763–771.
87. Place ES, Evans ND, Stevens MM. Complexity in biomaterials for tissue engineering. Nat Mater 2009;8(6):457–470.

88. Aboody K, Capela A, Niazi N, Stern JH, Temple S. Translating stem cell studies to the clinic for CNS repair: current state of the art and the need for a Rosetta Stone. Neuron 2011;70(4):597–613.
89. Sobani ZA, Quadri SA, Enam SA. Stem cells for spinal cord regeneration: current status. Surg Neurol Int 2010;1:93.
90. Driscoll P. Tissue engineering, cell therapy, and transplantation: products, technologies, and market opportunities. 2010
91. Rao M, Condic ML. Musings on genome medicine: is there hope for ethical and safe stem cell therapeutics? Genome Med 2009;1(7):70.
92. Brindley DA, Reeve BC, Sahlman WA, Bonfiglio GA, Davie NL, Culme-Seymour EJ, et al. The impact of market volatility on the cell therapy industry. Cell Stem Cell 2011;9 (5):397–401.
93. Kemp P. History of regenerative medicine: looking backwards to move forwards. Regen Med 2006;1(5):653–669.
94. Husing B, Buhrlen B, Gaisser F. Package 1—analysis of the actual market situation—mapping of industry and products. Human tissue engineered products—today's markets and future prospects final report. Fraunhofer Institute for Systems and Innovation Research, 2003, p. 1–133
95. Helble M. The movement of patients across borders: challenges and opportunities for public health. Bull World Health Organ 2011;89(1):68–72.
96. Lee MH, Arcidiacono JA, Bilek AM, Wille JJ, Hamill CA, Wonnacott KM, et al. Considerations for tissue-engineered and regenerative medicine product development prior to clinical trials in the United States. Tissue Eng Part B 2009;16(1):41–54.
97. Schwartz SD, Hubschman JP, Heilwell G, Franco-Cardenas V, Pan CK, Ostrick RM, et al. Embryonic stem cell trials for macular degeneration: a preliminary report. Lancet 2012;379(9817):713–720.
98. Prud'homme GJ. Cancer stem cells and novel targets for antitumor strategies. Curr Pharm Des 2012;18(19):2838–2849.
99. Atala A. Engineering tissues, organs and cells. J Tissue Eng Regen Med 2007;1(2):83–96.
100. Vacanti JP. Tissue engineering and the road to whole organs. Br J Surg 2012;99(4):451–453.
101. Schwartz RS. The politics and promise of stem-cell research. N Engl J Med 2006; 355(12):1189–1191.
102. MacKay G. Marcus Evans Conference, London, UK, 2006.

# Conventional and Novel Container Closure and Delivery Systems

JOSEPH WONG and MAHESH V. CHAUBAL

## 20.1 INTRODUCTION

### 20.1.1 Biologics: Market Growth for Biologics Delivery Technologies

Recent years have seen continued growth in the area of biologics with a majority of pharmaceutical companies developing pipeline in next-generation biologic products. The majority of biologics are administered via injectable routes of administration and often require frequent injections. Hence, delivery devices that can reduce discomfort associated with injectable delivery (both to the caregiver and the patient) are of interest as part of the lifecycle management of biologic products.

Growth in the area of biologics is expected to be fueled further with the anticipated advent of biosimilar products. In February 2012, the draft guidances on scientific and quality considerations in demonstrating biosimilarity to a reference product were provided jointly by Center for Biologics Evaluation and Research (CBER) and Center for Drug Evaluation and Research (CDER) [1,2]. The Food and Drug Administration's (FDA's) totality-of-the-evidence approach proceeds in a stepwise fashion by first demonstrating that the biological product is highly similar to the reference product analytically. Animal studies, including toxicity evaluation, are next followed by clinical evaluations on pharmacokinetics (PK), pharmacodynamics (PD), and immunogenicity. The first licensed interchangeable biological product will be granted a period of exclusivity. Under "Biosimilarity and Interchangeability," biosimilar products may have a different delivery device or container (e.g., prefilled syringe) compared with the reference product (e.g., vial) as long as compatibility is demonstrated [3].

From the perspective of both the innovator companies as well as biosimilars, differentiation of the product using novel delivery devices is critical. Furthermore,

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*,
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

success of such products depends on an in-depth understanding of the interactions between the drug and the device. This chapter attempts to discuss some key features of delivery devices for biologics (both conventional and novel) that are important for consideration in product development.

### 20.1.2 Biological Product Attributes

The acceptability, in terms of safety, effectiveness, and quality, of a biologic product, is inseparably linked to its chemical nature, dosage and frequency, route of administration, formulation and dosage form, manufacturing process, and container and administration set or device. Safety or immunogenicity [4], in particular for biologics, is the primary concern. Protein aggregation caused by various factors (including undesirable interactions with device components) could elicit undesirable immunologic responses. Appropriate formulation design can help to improve on stability and preventing degradation, denaturation, and aggregation.

Because of the importance of the device, especially for biologic products, proteins or vaccine products in devices are treated as combination products.

### 20.1.3 Combination Products

Given that a majority of biologics are administered parenterally, the use of devices to improve on the ease of injection, and therefore lowering the labor burden for clinicians and increasing patient compliance in self-administration, has received significant attention. Products incorporating a drug formulated in an advanced device are now being considered as combination products. Prefilled syringes and insulin injector pens are examples of combination products. Biological products in solution formulation are frequently packaged as prefilled syringes [5].

Per the FDA's Office of Combination Products (OCP), which was established on December 24, 2002 [6], the definition of combination products includes (1) a drug and a device; (2) a biological product and a device; (3) a drug and a biological product; or (4) a drug, a device, and a biological product. The examples of FDA-approved combination products are available on the FDA's website [7]. Some examples are tositumomab and iodine $^{131}$I tositumomab (tositumomab covalently linked $^{131}$I) and the drug-eluting stent for clogged heart arteries (zotarolimus).

Drug-compatible containers are required to provide barrier properties to prevent degradation of the biologic as well as functional properties to maintain its utility over shelf life. The container should provide protection from light (if necessary), gas transmission (e.g., water, oxygen), and microbial contamination. The interaction of the biologics with leachables must not trigger safety concerns, and the marketing pack, a combination of primary (drug contact) and secondary (e.g., overpouch) packaging, must provide the performance as designed [8]. The FDA has recognized the greater potential for adverse effects associated with identity, strength, quality, purity, or potency to occur during storage or shipping for protein products. From a functional perspective, the device should provide reproducible dose delivery over its entire duration of use.

### 20.1.4 Guidance for Industry

The OCP strongly encourages the developer of a combination product to conduct early dialogue with the agency to resolve issues on preclinical and clinical testing [9] such as pharmacokinetic or dose range studies, for example, or to avoid duplicate studies. Guidance on early development considerations for innovative combination products is available [10]. Either with a new chemical entity or existing drug or biologic, the safety and efficacy of a combination product in terms of new use, indication, dosing, target population, route of administration, and so on need to be addressed in clinical investigation in addition to drug or biologic device compatibility and manufacturing.

The OCP does not regulate products. Rather, it designates a FDA Center (CDER, CBER, or Center for Devices and Radiological Health [CDRH]) for regulation for a combination product based on the product's primary mode of action (PMOA) [11]. The PMOA is defined as the most important therapeutic action of a combination product. For example, the drug is determined to be secondary to the stent in a drug-eluting stent, and the application is therefore assigned to CDRH for evaluation. For a prefilled syringe, the drug (or biological product) component has the primary role in therapy, and the syringe serves only to deliver the drug. Therefore, CDER (or CBER) reviews the application [12]. In some cases, CDER may request consultation from CDER. However, manufacturer of a combination product may request for designation because each center requires different current good manufacturing process program or adverse event reporting, for example.

## 20.2 GENERAL DEVELOPMENT APPROACH OF A COMBINATION PRODUCT (DRUG OR BIOLOGIC WITH DEVICE)

As mentioned earlier, biologics are typically administered as injections. Limited exceptions include proteins [13] such as DNase (Pulmozyme) delivered via nebulization for the treatment of cystic fibrosis and Exubera, an inhaled insulin, for type II diabetes, although the latter product was subsequently withdrawn from the market [14].

The nasal route has also attracted attention because of its potential for delivery to the central nervous system as well as for being an effective route for vaccination and for systemic delivery; the nasal route provides a large surface area for drug absorption ($\sim$160 $cm^2$). Examples of biologics approved for nasal delivery include FluMist influenza vaccine and low-molecular-weight peptides (e.g., Syntocinon, oxytocin [MW 1.0 kDa] for induction of labor; Miacalcin, salmon calcitonin [MW 3.4 kDa] for treatment of Paget's disease) [15].

For achieving sustained release after intramuscular injection, a peptide can be formulated in microspheres composed of polylactic glycolic acid (PLGA) (e.g., Lupron Depot [leuprorelin] and Zoladex [goserelin]) [16]. For each of these alternate routes of delivery, the use of an appropriate device becomes critical for effective delivery. For example, a nasal spray device is required for the effective administration of the FluMist vaccine. Similarly, a specialized injection device (SafeSystem

Syringe) was developed for the injection of the PLGA implant under the skin for Zoladex. Exubera involved an advanced delivery disperser that aerosolized dried insulin particles efficiently to deposit insulin to the appropriate region of the lungs. As suggested from these examples, each device has its own attributes that are specific to the route of administration or even to the specific molecule. Although each device has its route-specific functionality, we will attempt to discuss some of the attributes that could be generically applied to devices that are used for delivery of biologics or protein products.

The FDA has provided an overview of medical devices and regulatory pathways:

> Combination products raise regulatory challenges because they involve components that were normally regulated under different types of authorities and often by different FDA Centers. Differences in regulatory pathways for each component can affect the regulatory processes for all aspects of product development and management, including preclinical testing, clinical investigation, marketing applications, manufacturing and quality control, adverse event reporting, promotion and advertising, and post-approval modifications [17].

In the case of a device, a draft guidance was published for pen, jet, and related injectors for use with drug and biological products [18]. Typically, the device itself will go through premarket notification submission (510K) or premarket approval. In addition to the disclosure of the device design to perform accurate and safe delivery of single or repeated doses, human factor design consideration is also required. Simplicity is the key to accommodate the variability in the end-user population (e.g., elderly adults with arthritis) and environment (e.g., storage, use in high attitude). A device designed to incorporate human cognitive and ergonomic considerations should improve on patient compliance. Interestingly, the least burdensome provisions of the FDA Modernization Act apply to all devices and only the device components of combination products [19].

For the drug or biologic component of combination product that includes a device, a New Drug Application (NDA) or a Biological Licensing Application (BLA) is usually the route for marketing application. The demonstration of drug or biologic compatibility with the device during sterilization, storage, and use (e.g., potential biologic aggregation under pressurized delivery) is required. For prefilled syringe, the drug or biologic formulation or concentration can also impact on syringeable (e.g., viscosity) and extract chemicals from the device therefore affecting performance [20].

### 20.2.1 Approved Biologics in Conventional Delivery Packaging

Protein products are usually packaged as lyophilized cake for improvement of stability or dissolution if the drug formulation is stable in glass vials. Table 20.1 shows the currently US-approved protein products that are packaged in ready-to-use prefilled syringes or pens or in flexible containers (e.g., bag or blow-fill-seal [BFS] plastic ampoule) for room temperature, refrigerated, or frozen storage. Some packaging involves the requirement of using a filter needle in the preparation process, such as ranibizumab injection for treatment of patient with neovascular

**TABLE 20.1 Approved Biologics in Ready-to-Use Packaging Systems***

| Drug | Formulation and pH | Packaging | Administration Route |
|---|---|---|---|
| Adalimumab | Contains 40 mg adalimumab, 4.93 mg sodium chloride, 0.69 mg monobasic sodium phosphate dihydrate, 1.22 mg dibasic sodium phosphate dihydrate, 0.24 mg sodium citrate, 1.04 mg citric acid monohydrate, 9.6 mg mannitol, 0.8 mg polysorbate 80; and water for injection, USP at a pH of 5.2 | Prefilled syringe | SC |
| Albumin | For example, the 25% product contains 25 g of albumin (human) in 100 ml with no preservative and stabilized with *N*-acetyltryptophan and sodium caprylate at physiological pH; the sodium content is 145 ± 15 mEq/l | Multilayered plastic container (polyethylene solution contact layer) | IV |
| $\alpha_1$-Proteinase inhibitor ($\alpha_1$-PI) | The solution contains 2% active $\alpha_1$-PI in a phosphate-buffered saline solution | Each carton of GLASSIA contains a single-use vial containing 1 g of functional $\alpha$1-PI in 50 ml of solution and a sterile filter needle | IV |
| Anakinra | Each 1-ml syringe contains 0.67 ml (100 mg) of anakinra in a solution (pH 6.5) containing sodium citrate (1.29 mg), sodium chloride (5.48 mg), disodium EDTA (0.12 mg), and polysorbate 80 (0.70 mg) in water for injection, USP | Prefilled syringe | SC |
| Calcitonin | Each 3.7-ml unit contains calcitonin–salmon 2200 IU/ml, sodium chloride, citric acid, phenylethyl alcohol, benzyl alcohol, polysorbate 80, HCl/NaOH and purified water, USP | Glass bottle | Nasal spray |
| Catumaxomab | One 0.1-ml prefilled syringe contains 10 μg of catumaxomab in addition to sodium citrate, citric acid monohydrate, polysorbate 80, and water for injection | Prefilled syringe | IV |
| Certolizumab pegol [21] | Each 1-ml single-use prefilled syringe contains certolizumab pegol (200 mg), sodium acetate (1.36 mg), sodium chloride (7.31 mg), and water for injection, USP at a pH of ≈4.7 | Plastic prefilled syringe | SC |

(*continued*)

**TABLE 20.1** (*Continued*)

| Drug | Formulation and pH | Packaging | Administration Route |
|---|---|---|---|
| Choriogonadotropin-alfa | Each 0.5 ml prefilled syringe contains 257.5 μg of choriogonadotropin- alfa, 28.1 mg mannitol, 505 μg 85% O-phosphoric acid, 103 μg L-methionine, 51.5 μg Poloxamer 188 and water for injection at pH 6.5 to 7.5 | Prefilled syringe | SC |
| Darbepoetin-alfa | Each syringe contains 10 to 500 μg darbepoetin-alfa.; the excipients per milliliter are polysorbate 80 (0.05 mg), sodium chloride (8.18 mg), sodium phosphate dibasic anhydrous (0.66 mg), and sodium phosphate monobasic monohydrate (2.12 mg) in water for injection, USP (pH 6.2 ± 0.2) | Prefilled syringe and SureClick prefilled pen | IV, SC |
| Denileukin diftitox | Each 2-ml vial contains 300 μg of denileukin diftitox, citric acid (20 mM), EDTA (0.05 mM), and polysorbate 20 (<1%) in water for injection, USP at a pH of 6.9 to 7.2 | Denileukin Diftitox is supplied as a frozen solution in a single-use vial; store frozen at or below −10°C | IV |
| Denosumab | Each 1-ml single-use prefilled syringe contains 60 mg denosumab (60 mg/ml solution), 4.7% sorbitol, 17 mM acetate, 0.01% polysorbate 20, water for injection, USP, and sodium hydroxide to a pH of 5.2 | Prefilled syringe | SC |
| Dornase-alfa | The 2.5-ml solution in a single-use BFS ampoule contains 1.0 mg/ml dornase-alfa, 0.15 mg/ml calcium chloride dihydrate, and 8.77 mg/ml sodium chloride at a pH of 6.3 | BFS plastic ampoule [22] for nebulization | Inhale |
| Epoetin-α | Contains 100 to 40,000 units of epoetin-α in different syringe sizes and glycine (5 mg/ml), polysorbate 80 (0.3 mg/ml), sodium chloride at 1.7 to 5.8 mg, sodium phosphate–monobasic dihydrate at 0.35 to1.16 mg, sodium phosphate–dibasic dihydrate at 0.67 to 2.22 mg, sodium citrate at less than 5 mmol, and water for injection; the prefilled syringes are fitted with the PROTECS needle guard device to help prevent needle stick injuries after use | Prefilled syringe with PROTECS needle guard device | IV |
| Etanercept | Contains a 50-mg/ml solution of etanercept with 1% sucrose, 100 mM sodium chloride, 25 mM L-arginine hydrochloride, and 25 mM sodium phosphate at a pH of 6.3 ± 0.2 | Prefilled syringe and SureClick autoinjector | SC |

**TABLE 20.1** (*Continued*)

| Drug | Formulation and pH | Packaging | Administration Route |
|---|---|---|---|
| Filgrastim | Each syringe container either 300 or 480 μg filgrastim; the 480-μg/0.8 ml vial, for example, contains 480 μg filgrastim, 0.472 mg acetate, 40.0 mg sorbitol, 0.032 mg polysorbate 80, 0.028 mg sodium in water for injection, USP at a pH of 4.0 | Prefilled syringe | IV, SC |
| Follitropin-alfa | Contains 30, 41, or 75 μg follitropin-alfa, 60 mg/ml sucrose, 3.0 mg/ml m-cresol, 1.1 mg/ml disodium hydrogen phosphate dihydrate, 0.45 mg/ml sodium dihydrogen phosphate monohydrate, 0.1 mg/ml methionine, 0.1 mg/ml poloxamer 188 at a pH of 6.7 to 7.3 | Prefilled pen | SC |
| Golimumab [23] | Each 0.5-ml prefilled syringe contains 50 mg of the golimumab antibody, 0.44 mg of L-histidine and L-histidine monohydrochloride monohydrate, 20.5 mg of sorbitol, 0.08 mg of polysorbate 80, and water for injection at a pH of ≈5.5 | Prefilled glass syringe | SC |
| Goserelin | Goserelin acetate implant (1.5-mm diameter cylinder) is supplied as a sterile and totally biodegradable D,L-lactic and glycolic acids copolymer (12.82–14.76 mg/dose) impregnated with goserelin acetate equivalent to 10.8 mg of goserelin preloaded in a SafeSystem syringe; the aluminum foil laminate pouch is light and moisture proof, and the unit is stored at room temperature | Prefilled syringe to inject the implant | SC |
| Hematopoietic progenitor cord blood cells | Each unit contains a minimum of $5.0 \times 10^8$ total nucleated cells with a minimum of $1.25 \times 10^6$ viable $CD34^+$ cells, suspended in 10% DMSO and 1% dextran 40 at the time of cryopreservation | Plastic bag stored at or below −150°C | IV |
| Human insulin and combinations or analogs | Insulin example: Contains insulin lispro 100 units, 16 mg glycerin, 1.88 mg dibasic sodium phosphate, 3.15 mg metacresol, zinc oxide content adjusted to provide 0.0197 mg zinc ion, trace amounts of phenol, and water for injection at a pH of 7.0 to 7.8 | Prefilled pen | IV, SC |

(*continued*)

**TABLE 20.1** (*Continued*)

| Drug | Formulation and pH | Packaging | Administration Route |
|---|---|---|---|
| Icatibant [24] | Each 3-ml prefilled syringe contains equivalent to a 30-mg icatibant, sodium chloride, glacial acetic acid at a pH of ≈5.5 | Prefilled syringe | SC |
| Influenza vaccine (seasonal) | Each 0.5-ml suspension in prefilled syringe contains 45 μg hemagglutinin (three strains), octoxynol-10 (TRITON X-100) ≤0.085 mg, α-tocopherol hydrogen succinate ≤0.1 mg, and polysorbate 80 (Tween 80) ≤0.415 mg; 0.2 ml prefilled syringe for intranasal spray | Prefilled syringe | IM, intranasal spray |
| Interferon α-2a | Each 0.5-ml syringe contains 11.1, 22.2, 33.3 μg of interferon α-2a, recombinant; 3.605 mg sodium chloride; 0.1 mg polysorbate 80; 5 mg benzyl alcohol; and 0.385 mg ammonium acetate | Prefilled syringe | SC |
| Interferon α-2b | The multidose pens provide 3, 5, or 10 million IU per 0.2 ml; each ml contains 7.5 mg sodium chloride, 1.8 mg sodium phosphate dibasic, 1.3 mg sodium phosphate monobasic, 0.1 mg edetate disodium, 0.1 mg polysorbate 80, and 1.5 mg *m*-cresol | Multidose prefilled pen | SC |
| Interferon β-1a [25] | Each 0.5-ml (30-mcg dose) syringe contains 30 μg of interferon β-1a; 0.79 mg sodium acetate trihydrate, USP; 0.25 mg glacial acetic acid, USP; 15.8 mg arginine hydrochloride, USP; and 0.025 mg polysorbate 20 in water for injection, USP at a pH of ≈4.8. Each 0.5-ml (0.5-cc) syringe contains either 22 μg or 44 μg of interferon β-1a, 2 or 4 mg albumin (human) USP, 27.3 mg mannitol USP, 0.4 mg sodium acetate, water for injection USP; also available in other prefilled syringes with slightly different formulations (pH of 3.5–4.5) | Prefilled syringe, prefilled autoinjector (approved February 2012), lyophilized vial for reconstitution | IM, SC |
| Interferon β-1b | Lyophilized vial comes with its diluent in a prefilled syringe | Prefilled syringe | IM, SC |
| Leuprolide acetate for depot suspension | The prefilled dual-chamber syringe contains sterile lyophilized microspheres of leuprolide acetate incorporated in a biodegradable lactic acid polymer; for example, the 3-month 22.5-mg dose is administered as a single IM injection | Prefilled dual-chamber syringe | IM |

**TABLE 20.1** (*Continued*)

| Drug | Formulation and pH | Packaging | Administration Route |
|---|---|---|---|
| Liraglutide | Each 3-ml prefilled pen contains liraglutide 18 mg, disodium phosphate dihydrate 1.42 mg; propylene glycol 14 mg, phenol 5.5 mg; and water for injection | Prefilled pen | SC |
| Methoxy PEG–epoetin-β [26] | Single-use vials 50 to 1000 μg/ml; prefilled syringes containing 50 to 250 μg and 400 to 800 μg methoxy PEG–epoetin-β in 0.3 and 0.6 ml, respectively, are formulated in sodium phosphate, sodium sulfate, mannitol, methionine, and poloxamer 188 at a pH of 6.2 ± 0.2 | Prefilled syringe | IV, SC |
| Peginterferon-α-2a | Each 0.5-ml prefilled syringe contains pegylated interferon-α-2a (180 μg), NaCl (4.0 mg), polysorbate 80 (0.025 mg), benzyl alcohol (5.0 mg), sodium acetate trihydrate (1.03085 mg), acetic acid (0.0231 mg) at a pH of 6.0 ± 0.5 | Prefilled syringe | SC |
| Peginterferon-α-2b | A dual-chamber glass cartridge containing lyophilized peginterferon alfa-2b in one chamber and a second chamber containing sterile water for injection USP to a volume of 0.5 ml<br>The lyophilized powder contains 67.5, 108, 162, or 202.5 μg of peginterferon alfa-2b; 1.013 mg dibasic sodium phosphate anhydrous; 1.013 mg monobasic sodium phosphate dihydrate; 54 mg sucrose; and 0.0675 mg polysorbate 80 | Dual-chamber glass cartridge | SC |
| Sipuleucel-T | 250 ml suspension containing a minimum of 50 million autologous $CD54^{+}$ cells activated with PAP-GM-CSF in lactated Ringer's injection, USP | Infusion bag (suspension) | IV |
| Somatropin | Intra-Mix example: a two-chamber cartridge with the front chamber contains powder mixture of recombinant somatropin 5.8 mg glycine 2.2 mg, mannitol 1.8 mg, sodium dihydrogen phosphate anhydrous 0.32 mg, and disodium phosphate anhydrous 0.31 mg; the rear chamber contains 0.3% | Intra-Mix: the powder and the diluent are mixed when the front and rear chambers are screwed together | SC |

(*continued*)

**TABLE 20.1** (*Continued*)

| Drug | Formulation and pH | Packaging | Administration Route |
|---|---|---|---|
| | m-Cresol and mannitol 45 mg in 1.14 ml water for injection; the reconstituted pH is ≈6.7<br>MiniQuick example: A single-use syringe device containing a two-chamber cartridge; the front chamber contains recombinant somatropin 0.22 to 2.2 mg, glycine 0.23 mg, and mannitol 1.14 mg, sodium dihydrogen phosphate 0.05 mg, and disodium phosphate anhydrous 0.027 mg and the rear chamber contains mannitol 12.6 mg in water for injection 0.275 ml; the reconstituted pH is ≈6.7<br>Nu Spin 10 contains 10 mg somatropin, 17.4 mg sodium chloride, 5 mg phenol, 4 mg polysorbate 20, and 10 mM sodium citrate at a pH of ≈6.0 | | |
| Streptococcal 13-valent vaccine | Each 0.5 ml contains ≈2.2 μg of each of the *Streptococcus pneumoniae* serotypes saccharides, 4.4 μg of 6B saccharides, 34 μg $CRM_{197}$ carrier protein, 100 μg polysorbate 80, 295 μg succinate buffer, and 125 μg aluminum as aluminum phosphate adjuvant | Prefilled syringe (suspension) | IM |
| Teriparatide | Each ml contains 250 μg teriparatide, 0.41 mg glacial acetic acid, 0.1 mg sodium acetate (anhydrous), 45.4 mg mannitol, 3 mg metacresol, and water for injection at a pH of 4 | Prefilled delivery device | SC |
| Ustekinumab | Each 45-mg ustekinumab prefilled syringe, for example, contains L-histidine and L-histidine monohydrochloride monohydrate (0.5 mg), polysorbate 80 (0.02 mg), and sucrose (38 mg) to fill to a final volume of 0.5 ml at a pH of 5.7 to 6 | Prefilled syringe | SC |

*Some of these listed drugs were not approved via OCP, as OCP was established on December 24, 2002. Insulin, for example, delivered via a pen is regulated by CDER.
BFS: blow-fill-seal; DMSO: dimethyl sulfoxide; EDTA: ethylenediaminetetraacetic acid; GM-CSF: granulocyte-macrophage colony-stimulating factor; IM: intramuscular; IV: intravenous; PAP: prostatic acid phosphatase; PEG: polyethylene glycol; SC: subcutaneous; USP: United States Pharmacopeia.

(wet) age-related macular degeneration. Others are designed for inhalation, for example, via nebulization for Pulmozyme. As inhaled insulin, Exubera, which was approved via OCP, is no longer available in the US market; it is not listed in this table. However, the packaging materials suitable for packaging powder for inhalation are discussed later.

### 20.2.2 Current Protein Drug Delivery Systems

The short half-lives typically associated with proteins and peptides often lead to the need for frequent injections. Hence, delivery devices that can ease the injection process, reduce the pain of injection, and improve on patient compliance are desirable. The current drug delivery systems designed for peptides and proteins are directed more to patient and clinician convenience and compliance (self-administration) presented in the form of prefilled syringes or pens. The ready-to-use large-volume parenteral in the form of premix in a flexible container is also available for intravenous (IV) administration in clinical settings.

## 20.3 CONTAINERS AND PRODUCT QUALITY

Product containers act as a barrier for light, oxygen, and moisture for protein products, wherein the protein may be formulated as solution, dispersion, or solid (including lyophilized cake). The moisture barrier is of particular importance for lyophilized products. However, escape of moisture from solution or dispersion through the container can affect the concentration and therefore dosing of these potent pharmaceuticals. A common remedy includes the use of an overpouch for containers that do not provide sufficient water vapor barrier properties. A foil overpouch may act as a barrier for light, oxygen, and moisture, as in the case of Zoladex (goserelin acetate, synthetic analogue of luteinizing hormone–releasing hormone, an implant).

### 20.3.1 Glass Containers

United States Pharmacopeia (USP) describes three types of glass, Types I, II and III, all of which are suitable for parenterals as long as compatibility is demonstrated. Type I glass is the most commonly used material for parenteral products. Although believed to be inert, Type I glass can participate in several interactions with the drug. These interactions could be with leachables from the glass surface (e.g., di- and trivalent cations released at neutral to high pH) or directly with the drug through electrostatic forces that can cause hydrophilic proteins to bind to the hydrophilic glass surfaces [27,28]. Glass can be purposely treated to turn the surface hydrophobic, such as Schott Type I Plus coated with silicon dioxide ($SiO_2$).

In one study, the interaction with containers, including untreated glass, siliconed glass, sulfur-treated glass, and plastic vials (polyester, polypropylene, and nylon) was found not to be correlated with molecular weight or isoelectric point for six proteins

studied [29]. Among the surface materials studied, glass appeared to bind the least amount of protein. Furthermore, adsorption to container surfaces appeared not to be an issue for proteins above 5 to 20 μg/ml. This finding was generally in agreement with a separate study [30], which indicated that whereas a 2.5-ppm immunoglobulin G (IgG) solution stored in a 50-ml ethylene–vinyl acetate (EVA) bag with 200 $cm^2$ surface area lost 32% (or 0.2 $\mu g/cm^2$ adsorption), a 50-ppm solution lost only 4% (or 0.5 $\mu g/cm^2$ adsorption).

***20.3.1.1 Protein Adsorption to Siliconed Glass Syringe or Vial Surface*** As mentioned in the previous section, hydrophobic silicone coating is sometimes used to reduce the adsorption of low-concentration potent biologics to the hydrophilic glass surface [31]. Silicone oil is also ubiquitously applied to various drug solution contact surfaces in devices such as on rubber stoppers (to prevent the sticking of the stoppers in bulk sorting and capping of vials during manufacturing of the parenteral products after solution aseptic filling), syringe barrels to allow smooth glide, and even to needle surfaces.

However, silicone oil on a device surface may be detrimental to the biologic because it has been identified as a possible contributing factor for protein aggregation. This silicone oil–mediated mechanism appears to be supported by the results of a separate study that investigated the effects of agitation, temperature, pH, and ionic strength in the induction of protein aggregation in the presence of silicone oil. A monolayer of protein was reported to absorb to silicone oil. The pH at a value close to the iso-electric point (PI) of the protein, and the increasing in ionic strength caused protein destabilization. It was also observed in this study that isothermal incubation in the presence of silicone oil did not affect protein aggregation except after shaking [32]. Thus, the combination of agitation and silicone oil is concluded to cause a synergistic stimulation of protein aggregation. However, it may be oversimplified to associate agitation and the presence of a hydrophobic surface with protein aggregation. For example, insulin dimerization occurs through interaction of the two hydrophobic regions of an insulin monomer. Destabilization of insulin is initiated with the adsorption of monomer to a hydrophobic surface. It is the dimerization at higher overall insulin concentration that reduces the concentration of hydrophobic insulin monomer and thereby reduces its destabilization [33]. Denaturation at the hydrophobic surface is unlikely for the hydrophilic hexamer, which is conformationally stable. Again, insulin hexamer is formed at a high concentration, such as 100 unit/ml in commercial products, and especially in the presence of zinc and m-cresol [34]. At a low infusion rate of 0.25 ml/min and at a low concentration of 0.1 unit/ml, the loss of regular human insulin to adsorption to some IV infusion bag and tubing was substantial [35]. Thus, the polymeric state of a protein in a formulation, in some cases, may affect adsorption to silicone-coated surfaces. Removal of silicone oil is a good solution for reducing protein aggregation. However, silicone oil is required for the functionality of the container, syringe, or device. The removal of silicone can therefore pose a different set of problems.

Application of a protein formulation strategy may resolve or reduce the container adsorption issue. If protein adsorption to glass occurs via electrostatic interaction, the

addition of a surfactant such as polysorbate 20 could reduce protein loss for susceptible proteins [36]. If the container surface is coated with silicone oil, polysorbate 20 may also completely inhibit silicone oil induced protein loss via agitation [32,36].

***20.3.1.2 Glass Container Delamination*** Glass is a highly viscous, supercooled liquid that for practical purposes behaves like an amorphous solid. When exposed to harsh processing conditions followed by sustained exposure to corrosive formulations, glass can form flakes at the surface, a phenomenon termed *delamination*.

*Glass Delamination Mechanism* In recent years, there has been a rise in product recalls because of glass delamination. Table 20.2 provides a list of biologics that have been subject to recent recalls. Although the issue appears to be multicausal, several proposed mechanisms of glass delamination and the methods for detection have been published [37]. In general, a number of factors are involved, including the pH of the product formulation, glass composition, method of glass container production, surface area treatment, and so on. The chemical compositions of some type I glass compositions are available as technical data from manufacturers and can be downloaded from the Internet [38]. Either leaching and ion exchange or glass dissolution was involved in the

**TABLE 20.2 Formulations of Recently Recalls of Biologics Caused by Glass Delamination**

| Drug | Manufacturer | Sterilization | pH | Composition |
|---|---|---|---|---|
| Epogen | Amgen | Aseptic | 6.9 | Epogen is formulated as a sterile, colorless liquid in an isotonic sodium chloride–sodium citrate buffered solution or a sodium chloride–sodium phosphate buffered solution for IV or SC administration<br>**Single-dose, preservative-free vial:** each 1 ml of solution contains 2000, 3000, 4000 or 10,000 units of epoetin-α; 2.5 mg albumin (human); 5.8 mg sodium citrate; 5.8 mg sodium chloride; and 0.06 mg citric acid in water for injection, USP (pH 6.9 ± 0.3); this formulation contains no preservatives<br>**Single-dose, preservative-free vial:** 1 ml (40,000 units/ml); each 1 ml of solution contains 40,000 units of epoetin-α, 2.5 mg albumin (human), 1.2 mg sodium phosphate monobasic monohydrate, 1.8 mg sodium phosphate dibasic anhydrate, 0.7 mg sodium citrate, 5.8 mg sodium chloride, and 6.8 mcg citric acid in water for injection, USP (pH 6.9 ± 0.3); this formulation contains no preservatives |

(*continued*)

**TABLE 20.2** (***Continued***)

| Drug | Manufacturer | Sterilization | pH | Composition |
|---|---|---|---|---|
| | | | | **Multidose, preserved vial:** 2 ml (20,000 units, 10,000 units/ml); each 1 ml of solution contains 10,000 units of epoetin-α, 2.5 mg albumin (human), 1.3 mg sodium citrate, 8.2 mg sodium chloride, 0.11 mg citric acid, and 1% benzyl alcohol as preservative in water for injection, USP (pH 6.1 ± 0.3) |
| | | | | **Multidose, preserved vial:** 1 ml (20,000 units/ml); each 1 ml of solution contains 20,000 units of epoetin-α, 2.5 mg albumin (human), 1.3 mg sodium citrate, 8.2 mg sodium chloride, 0.11 mg citric acid, and 1% benzyl alcohol as preservative in water for injection, USP (pH 6.1 ± 0.3) |
| Hylenex | Halozyme | Aseptic | 7.4 | Hylenex recombinant (hyaluronidase human injection) is supplied as a sterile, clear, colorless, nonpreserved, ready-for-use solution; each milliliter contains 150 USP units of recombinant human hyaluronidase with 8.5 mg sodium chloride, 1.4 mg dibasic sodium phosphate, 1.0 mg albumin human, 0.9 mg edetate disodium, 0.3 mg calcium chloride, and sodium hydroxide added for pH adjustment; Hylenex recombinant has an approximate pH of 7.4. |

IV: intravenous; SC: subcutaneous; USP: United States Pharmacopeia.

first step in disruption of the glass matrix. In general, a leaching or ion-exchange process was the dominant process in an acidic solution. In an alkaline solution, the dominant process was glass dissolution. At a neutral pH, both leaching and ion exchange and glass dissolution occurred [39]. The recommended pH range for a solution formulation is between 4 and 8.

In the leaching and ion-exchange process, the sodium ions in the glass exchange with hydrogen ions in the solution, leading to an accumulation of hydrogen ions in the glass matrix. In basic solution, glass dissolution occurs; it also occurs with the release of sodium ions from the glass into the solution. The increase of ions on the glass surface decreases the hydrolytic resistance of glass, further driving ion leaching [40]. Another proposed mechanism involves the diffusion of molecular water into silicate matrix of glass that contains alkali oxides and then displaces sodium ions form glass into the solution. In the dissolution process, each $SiO_2$ molecule reacts with two molecules of water that has been diffused into the matrix to form silicic acid ($H_4SiO_4$). At a pH of 8, $H_4SiO_4$ reacts with hydroxyl ion to form soluble silicate ($H_3SiO_4^-$), thereby driving the equilibrium for further dissolution of $H_4SiO_4$ [41].

It is also observed that the first step in the glass delamination often occurs near the drug solution fill line. The precursor to glass delamination is pitting. As pitting increases, the borders of multiple pits overlap and cause pieces of glass to flake or delaminate [42]. Microscopic techniques are generally used in the detection of flakes from the delamination of glass surface [37,43].

*Effects of Product Formulation on Glass Delamination* In addition to pH as discussed earlier, the selection of buffer may also contribute to glass delamination. Anions and cations may exacerbate glass delamination. Citrate, for example, could form soluble complexes with silica in a neutral solution [44]. Parenteral solutions are usually formulated iso-osmotic or isotonic. An increase in ionic concentration could decrease the hydrolytic resistance of glass, which could further drive ion leaching from the glass surface [40]. The formulations of some recalled products are listed in Table 20.2.

*Glass Processing and Product Storage* Protein products in prefilled syringes and glass vials are produced via sterile filtration. However, syringes and glass vials still need to go through depyrogenation and sterilization steps before filling of the protein or peptide solutions can proceed. Depyrogenation does not appear to cause generation of glass flakes because unprocessed vials also showed erosion at in controlled studies [45].

In general, delamination increases with increasing storage temperature and time. Treatment of the glass surface could reduce glass delamination. The coating of $SiO_2$ turned the hydrophilic glass surface to hydrophobic and therefore offered some protection against delamination. Plastic prefilled syringes may also be used to avoid the issue of glass delamination.

### 20.3.2 Containers for Light-Sensitive Products

Type I glass is made of borosilicate, and the solution contact surface is hydrophilic for containers that have not been surface treated/coated. For providing ultraviolet and visible light protection, iron only in oxidized state ($Fe^{3+}$) and manganese are incorporated in the amber Type I glass. Type II is soda-lime glass that is suitably dealkalized. Both Type I and II glasses are suitable for packaging neutral and acidic solutions and need suitability evaluation for alkaline products. Type III glass is also soda-lime glass. It has been published that sulfur in the reduced state ($S^{2-}$) provides ultraviolet and visible light protection in amber Type III glass – amber Type III glass still contains relatively low amount of $Fe^{2+}$ and $Fe^{3+}$. Since $S^{2-}$ is not soluble in Type I glass, amber Type I glass may not be sterilized by gamma irradiation, as the $Fe^{3+}$ acts as an electron trap and causes darkening of the amber Type I glass [46]. Bacitracin for injection, a polypeptide antibiotic, is recommended to be stored at 2 to 8°C and protected from direct sunlight. However, bacitracin for injection is packaged in a clear glass vial and stored in a paper carton. This marketing pack should provide adequate light protection per ICH (Q1B) guidelines [47]. As an alternative, some laboratory chemical suppler, for example, indeed offers bacitracin in an amber vial. A recent commentary suggested changes to the guideline to include photostability of drugs under conditions of patient use [48]. As such, one approach of product life

cycle management can simply be repackaging current biologics in a carefully screened and compatible amber glass container.

In addition to vial products, a number of protein product prefilled syringes are required to be protected from light. Because of the need to inspect a parenteral solution before injection, amber glass is not an appropriate choice for prefilled syringe products. As an example, epoetin-α prefilled syringe has a storage temperature requirement of between 2 to 8°C with no freezing or shaking. The syringe also needs to be protected from exposure to light. Hence, the epoetin-α formulation is packaged in a clear syringe inside a paper carton.

### 20.3.3 Hydrophobic Surface Interaction: Flexible Containers

The USP states the requirements for plastic containers [49]. For flexible containers, coextrusion technologies allow the production of multilayer films with external and solution contact layers. Both of these layers will usually provide the structural strength of the container and the middle layer(s), if needed, can provide the gas and moisture barriers. Coextrusion is the extrusion of multiple layers of material using a single extrusion head or die. Containers made of multiple thin layers of films with different properties for structure and barrier functionalities are likely to provide superior functionality relative to a simple layer of thicker material. However, coextrusion of dissimilar material layers may result in poor adhesion. High-density polyethylene (PE), for example, has poor adhesion to ethylene vinyl alcohol. An extrudable adhesive such as polyurethane may be needed to bond the layers. The common polymers for the external and solution contact layers are polypropylene, PE, polystyrene, or their combinations. The common materials offering oxygen and moisture barriers are polyvinylidene chloride, ethylene vinyl alcohol, nylon, and metalized films [50]. In the case of ethylene vinyl alcohol, hydrolytic degradation by moisture occurs, and this material is therefore not as commonly used in containers for liquid or dispersion formulations.

The oxygen transmission rate (OTR) measured as milliliter per film thickness (mil) per 100 $in^2$ per day is usually measured at 73°F and at a relative humidity for the specified storage condition. For polyvinylidene chloride and ethylene vinyl alcohol, the OTR is below 0.1.

The water vapor transmission rate (WVTR), in some cases, is measured at 100°F and at 90% relative humidity, but test conditions for a final product may vary. For polyvinylidene chloride and ethylene vinyl alcohol, the WVTR values (g/mil/100 $in^2$/24 h) were 0.22 and 1.4, respectively [51]. Polychlorotrifluoroethylene has a WVTR of lower than 0.03 and is used in blister packaging for powder for inhalation.

In terms of adsorption to IV sets and containers, there are publications stating that insulin is adsorbed to glass and plastic materials such as polyvinyl chloride (PVC), PE, and other polyolefins [52]. Estimates of adsorption (loss) of insulin are in the 20 to 30% range. The adsorption extent was also reported to be inversely proportional to insulin concentration. The mechanism for insulin adsorption to surfaces has been published [27]. The surface of an insulin monomer is predominately hydrophobic amino acids. Dimerization involves the weak van der Waal's force and hydrogen bonding in the pleated sheet region [53]. At high concentration and especially in the

presence of zinc ion ($Zn^{2+}$), the monomers associate into hexamers that are hydrophilic.

For hydrophobic surfaces such as PVC [54], insulin monomer absorbs onto the PVC surface via hydrophobic interaction. Albumin, a high-molecular-weight protein that exhibits a high affinity for substrate, may be used to displace adsorbed insulin, which has a smaller molecular weight and is therefore faster in diffusion in the competitive adsorption process [27].

High-density polyethylene (HDPE), a common polymer used in plastic containers, also has a hydrophobic surface. The adsorption of albumin (HSA), for example, onto HDPE was found to be consistent with the formation of a close-packed monomolecular layer and independent of HSA concentration. HSA is bound to HDPE by dispersion forces; the adsorption was irreversible probably caused by HSA conformational changes. Adsorption stops after HSA occupies all the adsorption sites on the HDPE surface [55,56].

Just like HDPE, linear low-density polyethylene (LLDPE) used in certain flexible container is also known to be hydrophobic. The contact angle for LLDPE is about 100 degrees [57]. Contact angle is a measurement of hydrophobicity; contact angles above 90 degrees are defined as hydrophobic. In a separate study, a stable reading of contact angle for water on PE surface of about 100 degrees could be achieved in about 10 min [58].

For PVC, the published contact angle is 97.7 degrees [59]. LLDPE is therefore similar to hydrophobicity of PVC, if not slightly more, and should not be expected to be acceptable in insulin monomeric adsorption [52]. A short-term adsorption study of bovine insulin (2.5 μg/ml in phosphate buffer saline at a pH of 7.4) to PVC and LLDPE was published [30]. As expected, bovine insulin adsorption to LLDPE (0.10 $\mu g/cm^2$) was actually higher than that to PVC (0.07 $\mu g/cm^2$).

### 20.3.4 Protein Aggregation via Interaction with Tungsten in a Container or Device

Devices such as prefilled syringes have additional material considerations that should be evaluated. For example, tungsten leachables in a prefilled syringe is a known culprit for protein aggregation [60]. A tungsten pin is typically used during the syringe manufacturing process to produce the needle hole at the base of the syringe. This process leads to the deposition of vaporized tungsten species, such as tungsten oxide, and glass materials on the solution contact surface of the syringe barrel. The conclusion from the published paper indicated that the extent of the protein aggregation was dependent on

1. Tungsten concentration and specific tungsten species
2. Protein concentration and characteristics of a protein
3. pH of the buffered formulation

In general, the lower the pH and the higher the tungsten concentration, the more protein aggregation; a reaction between an α-helix protein and tungsten followed a

constant stoichiometry. Electrostatic interaction, rather than unfolding or hydrophobicity, between protein in native structure and tungsten molecules appears to be the mechanism for aggregation.

## 20.4 RECENTLY MARKETED AND EMERGING CONTAINER PLATFORMS

### 20.4.1 Nonsiliconed Plastic Syringe

The silicone oil–free, cyclic olefin polymer [61] prefillable syringe, for example, is marketed to be suitable for biologics. This prefillable syringe is claimed to reduce the extent of protein aggregation following agitation, shipment, and storage conditions [62]. The functionality (lubricity and seal integrity) of this syringe piston is maintained via lamination with a polymer film (e.g., fluorocarbon polymer) instead of silicone oil. This polymer film serves to lower protein adsorption and act as a barrier to leachables. The manufacturing of the syringe components does not include tungsten pins. The plastic syringe is also breakage resistant.

### 20.4.2 Flexible Plastic Containers

Given some of the challenges associated with glass containers, there has been an interest in evaluating the use of flexible plastic containers for ready-to-use biologic solution products. FLEXBUMIN (human albumin 25%) is the first FDA-approved product of a biologic solution in a flexible plastic container suitable for room temperature storage. Flexible containers offer various benefits for protein solutions, including eliminating delamination risks, resolving potential glass incompatibility issues, and improving ease of administration by reducing the steps in reconstitution.

## 20.5 BLOW-FILL-SEAL PLASTIC AMPOULE

Blow-fill-seal technology can be used in packaging particle-free, sterile fill injectable solutions. This BFS technology may also be used to package drug solutions for other routes of administration, such as eye drops, and even for packaging of food and cosmetic products. There have also been announcements that prefilled syringes, with or without pistons, could be manufactured using this cost-effective technology. The steps for washing, depyrogenation, and sterilization of the container by gamma irradiation are eliminated.

One example of a biologic in a plastic BFS ampoule is Dornase Alfa for nebulization. BFS, on one hand, has the advantage of molding, sterile filling, and sealing the biologic in a simplified manufacturing process. However, the extrusion of the container requires the heating of the polyolefin resin (e.g., polypropylene and PE) to up to 160°C. In addition to volatile solvent and leachable protein interactions, the relatively higher temperature experienced in manufacturing may also contribute to

protein instability or eliciting immune responses, especially for large, complex, or temperature-sensitive proteins [22].

## 20.6 NONPARENTERAL DELIVERY

Devices have been developed for multiple nonparenteral administration of biologics. Pulmonary delivery devices have been in common use and are considered out of scope for this chapter. However, readers are referred to an excellent review by Byron [63]. Delivery devices have also been used for transdermal and nasal delivery of biologics. Principally, all of the considerations in this chapter are applicable to such devices as well. For transdermal devices, an FDA guidance has been developed to evaluate the completeness of dose delivered (and residual drug in device).

## 20.7 DEVICE FUNCTIONALITY TESTING

The two major functions of any drug delivery device are to protect the drug from degradation and to deliver the right dose every time. The earlier sections discuss various aspects of compatibility between the drug and the device surface. Additional consideration should be provided for the compatibility of the drug during functional operation of the device. For example, gas-powered devices (for jet injectors) may lead to stress during drug administration, which may be detrimental to a protein drug. Functionality of the device to ensure consistent delivery over the entire shelf life is a key consideration.

Yet another consideration during device development is the human factor aspects. For example, prefilled syringes with larger than normal flanges can improve ease of use for patients with arthritis. The piston breakaway force and the glide force for a syringe are often evaluated as critical product attributes during aging studies wherein the forces have to stay within limits previously determined using human factor studies. For complex devices, detailed verification and validation studies are required to demonstrate that the device performs per its specifications over the entire shelf life of the product.

## 20.8 RECOMMENDATIONS

Any new drug delivery device needs to take into account the interactions between the drug and the container or device, as well as the functionality of the device. On the assumption that the route of administration for a drug or biologic in a device such as inhalation or injection is decided, the following steps are recommended for the choice of a container for a combination product:

1. *Characterization of the protein or peptide at the concentration and the associated formulation intended for use:* It is important to understand the protein state, as well as hydrophilicity or hydrophobicity at the intended

concentration and pH. All of these parameters lead to the understanding of formulation flow (viscosity), which may affect functionality of the device such as syringeability, for example.

2. *Evaluation of containers' key quality attributes:* The hydrophilicity or hydrophobicity of the surface of a container needs to be examined. Removal of silicone oil coating or use of a silicone oil–free container may be preferable under certain circumstances. If the delivery systems involve mechanical pumping, agitation in the presence of silicone oil may cause aggregation. The addition of a surfactant such as polysorbate 20 may be an option in formulating the drug if a metal component is present in the device. However, a high level of surfactants that may be used to stabilize a protein may extract components from the device, causing incompatibilities.
3. *Choosing different containers:* A short-term design of experiment is usually conducted to screen out potential containers in terms of performance, such as depth of the insertion or injector or orifice size, and compatibility. If container cost constitutes a relative low percentage of the cost of the combination product, the strategy is to first screen a silicone oil–free, cyclic olefin polymer prefillable syringe for an injectable just to avoid the potential issues with silicone oil, tungsten, and extractives. All in all, the drug–biologic compatibility assessments with the device during sterilization, storage, and use (e.g., potential biologic aggregation under pressurized delivery) are required.
4. *Human factor considerations:* Simplify the device design to accommodate human cognitive and ergonomic considerations for accurate and safe delivery of the dose to improve on patient compliance.

A careful strategy must be adopted to assess protein aggregation via interaction with the selected container, which may result in unwanted immunogenicity. The ultimate goal is to market a safe, effective, easy-to-use, and cost-effective combination product to provide mutual benefits to both patients and the developer.

## REFERENCES

1. Food and Drug Administration. Draft Guidance for Industry, Scientific Considerations in Demonstrating Biosimilarity to a Reference Product, FDA CDER/CBER; February 2012.
2. Food and Drug Administration. Draft Guidance for Industry, Quality Considerations in Demonstrating Biosimilarity to a Reference Product, FDA CDER/CBER; February 2012.
3. Food and Drug Administration. Guidance for Industry on Biosimilars: Q&As Regarding Implementation of the BPCI Act of 2009, FDA CDER/CBER; February 2012.
4. van de Weert M, Moller E. Immunogenicity of biopharmaceuticals: causes, methods to reduce immunogenicity, and biosimilars. In Biotechnology: Pharmaceutical Aspects, Vol. III, Immunogenicity of Biopharmaceuticals. New York: Springer; 2008, p. 97–111.
5. Food and Drug Administration, Office of Combination Products. Available from: http://www.fda.gov/CombinationProducts/AboutCombinationProducts/ucm101496.htm and http://www.fda.gov/CombinationProducts/AboutCombinationProducts/ucm118332.htm.

6. Food and Drug Administration. Available from: http://www.fda.gov/AboutFDA/ReportsManualsForms/Reports/PerformanceReports/CombinationProducts/ucm117857.htm.
7. Food and Drug Administration. Available from: http://www.fda.gov/CombinationProducts/AboutCombinationProducts/ucm101598.htm.
8. Food and Drug Administration. Guidance for Industry, Container Closure Systems for Packaging of Human Drugs and Biologics, FDA CDER/CBER; May 1999.
9. Food and Drug Administration. Draft Guidance for Industry and FDA Staff: FAQs on Pre-Clinical and Clinical Research on Combination Products, Combination Products Coalition: Guidance Document Submission; February 27, 2009.
10. Food and Drug Administration. Guidance for Industry and FDA Staff: Early Development Considerations for Innovative Combination Products, FDA OCP; September 2006.
11. Kramar M. Combination products: challenges and progress. Regul Aff Focus 2005; August: 30–35.
12. Lauritsen K, Nguyen T. Combination products regulation at the FDA. Clin Pharmacol Ther 2009;85(5):468–470.
13. Choi W, Murthy G, Edwards D, Langer R, Klibanov A. Inhalation delivery of proteins from ethanol suspensions. Proc Natl Acad Sci USA 2001;98(20):11103–11107.
14. Kaufman M. Inhaled insulin: troubled drug rises from the ashes. Hospitalist 2009;13(1):10.
15. Ozsoy Y, Gungor S, Cevher E. Nasal delivery of high molecular weight drugs. Molecules 2009;14(9):3754–3779.
16. Toguchi H. Sterility assurance of microspheres, J Control Release 1999;62(1–2):51–55.
17. Food and Drug Administration. Available from: http://www.fda.gov/AboutFDA/CentersOffices/OfficeofMedicalProductsandTobacco/CDRH/CDRHTransparency/ucm203018.htm.
18. Food and Drug Administration. Draft Guidance for Industry and FDA Staff: Technical Considerations for Pen, Jet, and Related Injectors Intended for Use with Drugs and Biological Products, FDA CDRH/CDER/CBER/OCP; April 2009.
19. Food and Drug Administration. The Least Burdensome Provisions of the FDA Modernization Act of 1997: Concept and Principles; Final Guidance for FDA and Industry, FDA CDRH/CBER; October 4, 2002.
20. Adler M. Challenges in the development of pre-filled syringes for biologics from a formulation scientist's point of view. Am Pharm Rev 2012;15(1):96,98–101.
21. Huges B. 2008 FDA drug approvals. Nat Rev Drug Discov 2009;8(2):93–96.
22. Liu W, Lam P, Faulhaber S, Sane S. Biopharmaceutical manufacturing using blow-fill-seal technology. BioPharm Int 2011;24(7):22–29.
23. Kling J. Fresh from the biologics pipeline. Nat Biotechnol 2011;29(3):197–200.
24. Magon K. New drug approvals FDA 2011. Available from: http://knol.google.com/k/krishan-maggon/new-drug-approvals-fda-2011/3fy5eowy8suq3/148#FDA_New_Drug_Approvals_2011.
25. Gokarn Y, Kosky A, Kras E, McAuley A, Remmele A Jr. Excipients for protein drugs. In: Kayare A, Chaubal M, editors. Excipient Development for Pharmaceutical, Biotechnology, and Drug Delivery Systems. New York: Informa; 2006, p. 291–331.
26. Huges B. 2007FDA drug approvals: a year of flux. Nat Rev Drug Discov 2008;7(2):107–109.
27. Henry M, Dupont-Gillain C, Bertrand P. Characterization of insulin adsorption in the presence of albumin by time-of-flight secondary ion mass spectrometry and X-ray photoelectron spectroscopy. Langmuir 2008;24(2):458–464.

28. Nilsson P, Nylander T, Havelund S. Adsorption of insulin on solid surfaces in relation to the surface properties of the monomeric and oligomeric forms. J Coll Interface Sci 1991;144(1):145–152.
29. Burke C, Steadman B, Volkin D, Tsai P, Bruner M, Middaugh C. The adsorption of proteins to pharmaceutical containers surfaces. Int J Pharm 1992;86(1):89–93.
30. Rabinow B, Ding Y, Qin C, McHalsky M, Schneider J, Ashline K, et al. Biomaterials with permanent hydrophilic surfaces and low protein adsorption properties. J Biomater Sci 1994;6(1):91–109.
31. Jones L, Kaufmann A, Middaugh C. Silicone oil induced aggregation of proteins. J Pharm Sci 2005;94(4):918–927.
32. Thirumangalathu R, Krishnan S, Ricci M, Brems D, Randolph T, Carpenter J. Silicone oil- and agitation-induced aggregation of a monoclonal antibody in aqueous solution. J Pharm Sci 2009;98(9):3167–3181.
33. Sluzky V, Tamada J, Klibanov A, Langer R. Kinetics of insulin aggregation in aqueous solutions upon agitation in the presence of hydrophobic surfaces. Pro Natl Acad Sci USA 1991;88(21):9377–9381.
34. Whittingham J, Edwards D, Antson A, Clarkson J, Dodson G. Interactions of phenol and *m*-cresol in the insulin hexamer, and their effect on the association properties of B28 Pro → Asp insulin analogues. Biochem 1998;37(33):11516–11523.
35. Ling J, Hu M, Hagerup T, Campbell R. Lispro insulin: adsorption and stability in selected intravenous devices. Diabetes Educator 1999;25(2):237–245.
36. Duncan M, Lee J, Warchol M. Influence of surfactant upon protein/peptide adsorption to glass and polypropylene. Int J Pharm 1995;120(2):179–188.
37. Bales F, Lathrop L, Kopp A, Mitchell B. Detecting Siliceous Flakes in Glass Containers Using Microscopic Techniques. West Pharmaceutical Services; 2010.
38. Technical Data, Glass, Wheaton. Available from: http://www.wheaton.com/b2cwhn/wheatonsci/pdfs/Tech_glass.pdf.
39. Borchert S, Ryan M, Davison R, Speed W. Accelerated extractable studies of borosilicate glass containers. J Parenter Sci Technol 1989;43(2):67–79.
40. Schwarzenbach M, Reimann P, Thommen V, Hegner M, Mumenthaler M, Schwob J, et al. Topological structure and chemical composition of inner surfaces of borosilicate vials. PDA J Parenter Sci Technol 2004;58(3):169–175.
41. Iacocca R, Allgeier M. Corrosive attack of glass by a pharmaceutical compound. J Mater Sci 2007;42(3):801–811.
42. Case study—Packaging/Pharma, Pre-testing for Glass Delamination in Pharmaceutical Vials. Westmont, IL: McCrone Associates; 2011.
43. Roseman T, Brown J, Scothorn W. Glass for parenteral products: a surface view using the scanning electron microscope. J Pharm Sci 1976;65(1):22–29.
44. Bacon F, Raggon F. Promotion of attack on glass and silica by citrate and other anions in neutral solution. J Am Ceram Soc 1959;42(4):199–205.
45. Iacocca R, Toltl N, Allgeier M, Bustard B, Dong X, Foubert M, et al. Factors affecting the chemical durability of glass used in the pharmaceutical industry. AAPS Pharm Sci Tech 2010;11(3):1340–1349.
46. Wong J, Papadopoulos P, Rebbeck C, Schwabe S, Lischewski G, Melchert J, et al. Effect of terminal sterilization by irradiation on amber Type I and Type III glass containers

containing veterinary oxytetracycline suspension. PDA J Pharm Sci Technol 2004; 58(1):6–14.

47. International Conference on Harmonization: Guideline for the photostability testing of new drug substances and products. Fed Reg 1997;62(95):27115–27122.
48. Baertschi S, Alsante K, Tønnesen H. A critical assessment of the ICH guideline on photostability testing of new drug substances and products (Q1B): recommendation for revision. J Pharm Sci 2010;99(7):2934–2940.
49. Chapter 661. Containers—Plastics, USP34/NF29. The United States Pharmacopeial Convention, Rockville, MD; 2011, p. 259–263.
50. Forman C. Plastics for barrier packaging. In Market Research Report, PLS014F. Wellesley, MA: BCC Research; 2012.
51. Mitchell E, Bossaert B. High performance polypropylene for the packaging market. Tappi J 1992;75(6):79–81.
52. Trissel L. Handbook on Injectable Drugs. 16th ed. Bethesda, MD: American Society of Health-System Pharmacists; 2010, p. 895–897.
53. Lougheed W, Woulfe-Flanagan H, Clement J, Albisser A. Insulin aggregation in artificial delivery systems. Diabetologia 1980;19(1):1–9.
54. Djordjevic D, Wiedmann M, McLandsborough L. Microtiter plate assay for assessment of *Listeria monocytogenes* biofilm formation. App Environ Microbiol 2002;68(6):2950–2958.
55. Brynda E, Houska M, Kalal J. Adsorption of human fibrinogen and human serum albumin onto polyethylene. Ann Biomed Eng 1980;8(3):245–252.
56. Hlady V, Buijs J. Protein adsorption on solid surfaces. Curr Opin Biotechnol 1996; 7(1):72–77.
57. Gufeng C, Yi Z, Yafei Z, Jiarui X. Grafting of PEG400 onto the surface of LLDPE/SMA film. Front Chem Eng China 2007;1(2):128–131.
58. Fu Z, Liu M, Xu J, Wang Q, Fan Z. Surface modification of liner low-density polyethylene film by amphiphilic graft copolymers based on poly(high α-olefin)-graft-poly(ethylene glycol). J Appl Polymer Sci 2011;119(2):1111–1121.
59. Yamamoto K, Kimura T, Nam K, Funamoto S, Ito Y, Shiba K, et al. Synthetic polymer-tissue adhesion using an ultrasonic scalpel. Surg Endosc 2011;25(4):1270–1275.
60. Jiang Y, Nashed-Samuel Y, Li C, Liu W, Pollastrini J, Mallard D, et al. Tungsten-induced protein aggregation: solution behavior. J Pharm Sci 2009;98(12):4695–4710.
61. Shin J, Park J, Liu C, He J, Kim S. Chemical structure and physical properties of cyclic olefin copolymers (IUPAC technical report). Pure Appl Chem 2005;77(5):801–814.
62. Esfandiary R, Joshi S, Vilivalam V, Middaugh C. Characterization of protein aggregation and adsorption on prefillable syringe surfaces. West Pharmaceutical Services.
63. Byron P. Drug delivery devices issues in drug development. Proc Am Thorac Soc 2004; 1(4):321–328.

# Controlled-Release Systems for Biologics

MAYURA OAK, RHISHIKESH MANDKE, BUDDHADEV LAYEK, GITANJALI SHARMA, and JAGDISH SINGH

## 21.1 INTRODUCTION

Recent progress in genetic engineering and pharmaceutical technology has resulted in identification and commercial availability of a number of new biologics with distinct therapeutic activities [1]. The "biologics," in practical term, encompass a variety of products such as recombinant proteins, vaccines, blood and blood components, allergenics, somatic cells, tissues, and gene therapy [2]. The market of recombinant DNA (rDNA)–derived protein and peptide therapeutics has seen a sustained growth and is expected to increase with a compound annual growth rate of more than 10% in 2013 [3]. Monoclonal antibodies (mAbs), another class of biologics, are considered as a second wave of innovation next to recombinant proteins [4]. The demand for mABs is also continuously increasing and should be estimated in billions of dollars per year. The demand for vaccines has also shown a continuous rise with a predicted annual growth of approximately 6.5% [5].

The major concerns about delivering biologics, especially peptides and proteins, are their unique physicochemical properties, delicate structures, and short biological half-lives. Understanding stability issues associated with biologics and their degradation mechanisms is essential for developing successful therapeutics. Most of these protein- and peptide-based pharmaceuticals are administered by frequent injections either in solution or suspension dosage forms. This approach is rather inconvenient for patients, often requires medical personnel or hospitalization, and is therefore expensive. Similar to proteins, mAbs have a delicate structure that makes them more susceptible to conformational changes during formulation of injection and storage. Typical delivery routes for mAbs are intravenous (IV) and subcutaneous (SC) and involve several doses, thus emphasizing the need for controlled-release formulations.

Gene therapy, another area of biologics, holds a great promise when the impaired gene is replaced with its correct copy for the normal functioning of affected cell. Although it has excellent therapeutic potential, the issues involving safety,

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*,
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

therapeutic efficacy, duration of desired effect, and ability to deliver gene to the correct target cell make it difficult to gain clinical acceptance. In the case of vaccines, the need for multiple doses raises a patient compliance issue; therefore, development of a more efficient single administration delivery system is of prime importance. Thus, translating these biologics into convenient medicines requires parallel development of drug delivery technologies.

Controlled drug delivery can be helpful to address the drawbacks associated with traditional methods of administration of biologics. Recently, development of an effective controlled-release system to deliver biologics in a stable and biologically active form has drawn greater attention. Controlled delivery could allow a drug to be delivered at a controlled rate and maintained in a desired therapeutic window for days to months or even for years. The primary objective of controlling the drug delivery is to attain effective drug level for a longer duration, thus reducing the injection frequency and the possibility of accidently administering higher or lower doses.

The most widely used routes for sustained release systems are SC and intramuscular (IM) [6]. For most of the formulations of biologics, delivery via parenteral route shows strong growth prospects and therefore is of prime importance. The significant benefits of parenteral controlled-release systems over conventional delivery systems for peptides and proteins, including mAbs, are minimal side effects and toxicity caused by maintenance of a constant drug level, controlled-release rate and duration, reduced cost because of optimum use of the drug administered, the need for fewer injections, and increased patient compliance by reducing the frequency of dosing, leading to improved patient comfort. Using polymeric controlled-release systems for DNA delivery can potentially help to reduce the limitations of achieving safe and adequate therapeutic effect associated with gene therapy. Several natural and synthetic polymeric vectors have been used as controlled-release systems for DNA delivery.

In the case of conventional vaccines, the major cause of immunization failure is patient negligence toward booster doses of vaccines. Controlled vaccine delivery could be helpful in developing sustained and persistent immune responses because it allows incorporation of larger doses of antigen that can be released continuously for prolonged periods, eliminating the need of booster doses [7]. DNA vaccine is another type of vaccine delivery method in which gene encoding for protein antigen is administered into host cells, enabling antigen production to occur *in vivo* [8]. This continuous and sequential presentation of antigen to the immune system is effective in eliciting a prophylactically important immune response. However, in addition to the aforementioned advantages, development of controlled-release delivery systems for biologics still faces certain obstacles such as low drug loading; physical and chemical instability; and low biological activity caused by processing conditions, instability during storage, sterility issues, and a significant burst effect or dose dumping after being placed in the body [9,10].

### 21.1.1 Polymeric Controlled Delivery Systems

Many research groups have focused on the development of polymeric delivery systems for the delivery of biologics at a controlled rate as per the treatment need.

Various natural and synthetic polymers have been explored as delivery vehicles to achieve controlled release and at the same time preserve the stability of incorporated drugs during storage and release [11]. The polymeric delivery systems could be the matrix or reservoir type. A drug is dispersed or dissolved in the matrix-type delivery system, and its release depends on diffusion, polymer swelling, degradation and erosion; in case of a reservoir-type delivery system, a drug gets surrounded by polymer, forming a diffusion barrier. Therefore, the rate of release depends on the dissolution of drug into polymeric phase and diffusion through polymer walls [12]. The polymer-based delivery systems are broadly classified into particulate delivery systems, gels, suspensions, solutions, emulsions, implants, and drug-eluting stents. Polymers of a biodegradable nature are gaining importance because they undergo enzymatic or nonenzymatic degradation *in vivo* and produce biocompatible, non-toxic byproducts that are further metabolized or eliminated from the body [9]. Additionally, the delivery systems containing biodegradable polymers do not need surgical explantation.

Because of the uncertain purity and high cost of natural polymers, synthetic biodegradable polymers are increasingly used for controlled delivery of proteins and peptides. These include polyesters (lactide and glycolide polymers), poly(caprolactones), poly(phosphoesters), polyanhydrides, block copolymers, poly(amino acids), poly(urethanes), and so on [9,13]. Among these polymers, poly(lactic acid) (PLA), poly(lactic-*co*-glycolic acid) (PLGA), and their copolymers have been extensively used as controlled delivery vehicles because of their biodegradable and biocompatible nature.

## 21.2 CHOICE OF A SUITABLE DELIVERY SYSTEM

The choice of a delivery system depends on the clinical outcome of the therapeutics. The designing of a controlled delivery system involves concurrent consideration of various aspects of vehicles or polymers, drugs, release mechanisms, release duration, and administration routes [14]. These factors are interdependent and have a profound effect on advancement of the biologics from bench to clinic. Figure 21.1 summarizes the various factors which need to be considered while developing a controlled delivery system for any biotherapeutic agent [11].

### 21.2.1 Characteristics of Polymeric Delivery Systems

The following characteristics are desirable for polymeric delivery systems:

1. Polymers should degrade slowly over the period of time that corresponds with their function [13,15].
2. The delivery system should be easy to formulate and should have enough mechanical stability for desired release duration [9,15].
3. Polymer degradation products should be nontoxic, biocompatible, and easily absorbed or eliminated from the body [13,15].

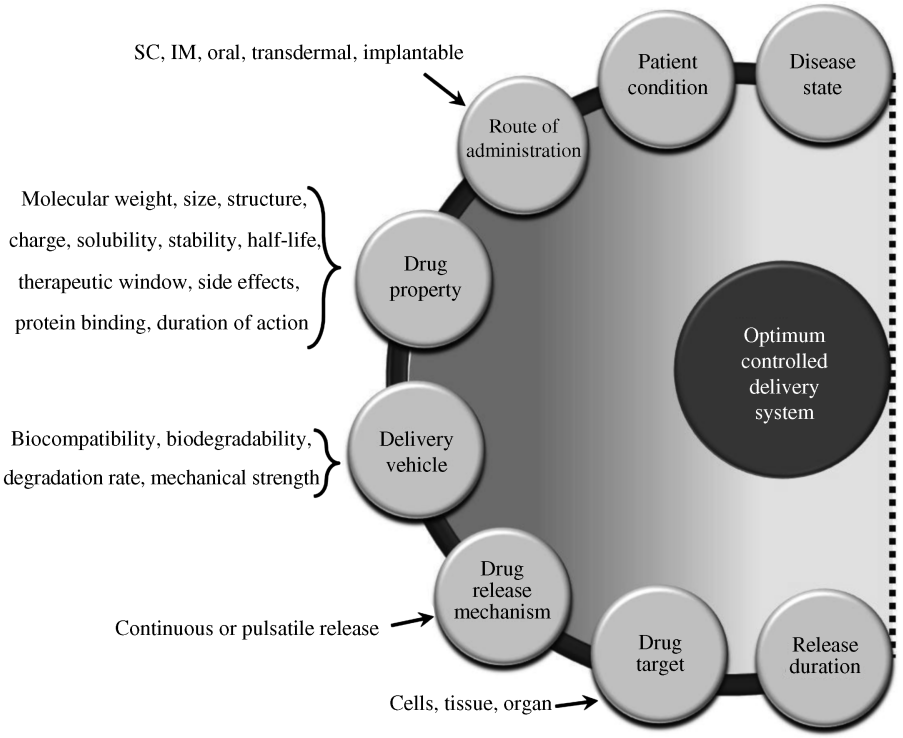


**FIGURE 21.1** Factors to be considered during development of a controlled delivery system. IM: intramuscular; SC: subcutaneous.

4. The delivery system should not produce any immune reaction or inflammation after injection or implantation [15].
5. The delivery system should not alter the stability and therapeutic efficacy of the incorporated drug molecules.
6. Most importantly, the delivery system should release the incorporated molecule for the designated period at a controlled rate in a stable and biologically active form.
7. In the case of DNA, the delivery system should not release the contents until it reaches the target site, and it should be able deliver the gene of interest to the specific target cell.
8. The delivery system should be able to cross the cell membrane barrier and reach the nucleus and deliver the gene to the nucleus upon reaching; thus it should show a delayed-release profile.
9. The delivery system should have good DNA condensation and buffering ability to withstand various degradation enzymes and to protect DNA from non-specific endolysosomal degradation.

### 21.2.2 Stability of Biologics

The stability of biologics is one of the most important issues when developing a controlled-release system. To ascertain the stability of biologics, especially proteins and peptides, during fabrication, storage, and release, various analytical techniques have been used. The secondary and tertiary structural integrity of the protein or peptide can be determined using circular dichroism spectroscopy and differential scanning calorimeter, and the primary structural integrity can be analyzed using high-performance liquid chromatography and mass spectroscopic techniques. Gel electrophoresis and size exclusion chromatography can be used to establish the presence and nature of aggregates of proteins and peptides formed during formulation or encapsulation, release, and storage. Specific bioassays are useful to verify the bioactivity of the protein or peptide released from the delivery system. In the case of mAbs, their surface interactions can be assessed using quartz crystal microbalance [16]. To determine the stability and structural integrity of the gene or DNA incorporated in the delivery system, the commonly used techniques include agarose gel electrophoresis and ultraviolet spectroscopy; their biological functionality can be evaluated by various cell transfection experiments [17].

## 21.3 FEASIBLE SYSTEMS FOR BIOLOGICS

A brief outline of some of the feasible parenteral controlled delivery systems used for delivering biologics has been described here. Table 21.1 presents findings of various controlled delivery systems for biotherapeutics.

### 21.3.1 Microspheres

Biodegradable microsphere is the most common feasible systems for parenteral delivery of biologics. Microspheres are small spherical particles with a diameter of 1 to 1000 μm. The biodegradable microsphere delivery systems for proteins and peptides help to maintain stability against chemical and enzymatic degradation, increase their plasma half-lives, and achieve a sustained release profile. Similarly, biodegradable microspheres for nucleic acid delivery offer enhancement in oligonucleotide pharmacologic effect by increasing the circulation lifetime and stability of the therapeutic gene and avoids potential risks of using viral vectors. One of the key factors in the design of injectable microsphere is the choice of suitable polymers. Biodegradable polymers are chosen over the nonbiodegradable counterpart because of the aforementioned advantages of biodegradable polymers. Different techniques used for microsphere manufacturing include spray drying, double emulsification solvent evaporation, and phase separation–coacervation. Each of them has its specific merits and demerits. The selection of a particular method of encapsulation depends on criteria that include optimal drug loading, minimal exposure of labile proteins to strong denaturing solvents, stability, controlled-release profile, high yield with desired particle size, intra- and interbatch reproducibility, and micromeritic

**TABLE 21.1 Controlled Delivery Systems for Bioactive Agents**

| Delivery Systems | Biologics Delivered |
|---|---|
| Microparticles | HIV-1 D-GPEi plasmid/pCMV-Luc plasmid [20], pCMV-Luc/pCI-Neo plasmid [21], BSA/pEGFP-N1 plasmid [22], HBV S/V1012 plasmid [23], SPf 66 malaria vaccine [24], hepatitis B antigen [25], lysozyme [26], insulin [27,28], calcitonin [29], hGH [30], IGF-1 [31] |
| Microcapsules | Insulin [32,33], FITC-BSA [34], ovalbumin antigen [35], BSA [36,37] |
| Nanoparticles | pDNA for mouse IL-10 [38], pCMV/beta-gal and pEGFP-C1 plasmid [39], luciferase reporter plasmid [40], IL-1Ra plasmid DNA, influenza HA vaccine [41], BMPs [42], VEGF [43] |
| Polymeric hydrogels and *in situ* gel-forming systems | Luciferase pDNA [44], pCMV-SPORTB-gel [45], pLUC/pEGFP [46], IL-2 [47], calcitonin [48], insulin [49,50], GH [51], steroidal hormones [52] |
| Nanogels | Antisense ODN,53 IL-10 [54] |
| Peptide nanofiber hydrogels | EGF [55], ODN [56], lysozyme [16], BSA [16,57], IgG [57,58], IGF-1 [58], VEGF [58], insulin [59] |
| Implants, Matrices | PDGF-BB [60], PDGF-encoding plasmid [61], VEGF [62], bFGF [63], rh-INF alpha-2a (IFN-α) [64], *Clostridium tetani* and *Clostridium novy* toxoids [65] |
| Liposomes | Insulin [66], leuprolide [66], met-enkephalin [66], octreotide [66], cytokines (IL-1, IL-2, IL-6, GM-CSF, and IFN-γ) 67] |
| Suspensions | Zinc-insulin suspension [68], rHir [69], plasma-derived factor IX, and INF-α [70] |
| Vesicular phospholipid gels | Erythropoietin [71], cetrorelix [72] |
| Resealed erythrocytes [73] | Insulin, erythropoietin, IL-2, IL-3, IFN-α, vaccine subunits, toxoids, model antigens |
| Composite delivery systems | TGF-β [74], BSA [75,76], BDNF and VEGF [77], HSP27 [78] |

BDNF: brain-derived neurotrophic factor; bFGF: basic fibroblast growth factor; BMP: bone morphogenetic protein; BSA: bovine serum albumin; CMV: cytomegalovirus; D-GPEi plasmid: codon-optimized DNA plasmid containing the Gag, Pol, and Env gene; EGF: epidermal growth factor; FITC: fluorescein isothiocyanate; FITC-BSA: FITC-conjugated BSA; GH: growth hormone; GM-CSF: granulocyte-macrophage colony-stimulating factor; HA: hemagglutinin; HBV: hepatitis B virus; hGH: human growth hormone; HIV-1: human immunodeficiency virus type 1; HSP: heat shock protein; IFN: interferon; IGF: insulin-like growth factor; IgG: immunoglobulin G; IL: interleukin; ODN: oligonucleotide; pCI-Neo: SIY antigen-encoding plasmid; pCMV Luc: luciferase-encoding plasmid driven by the CMV promoter; PDGF-BB: platelet-derived growth factor-BB; pDNA: plasmid DNA; pEGFP: enhanced green fluorescent protein-encoding plasmid; pLUC: luciferase-encoding plasmid; rHir: recombinant hirudin; spf66: synthetic peptide vaccine against *Plasmodium falciparum* malaria; TGF: transforming growth factor; VEGF: vascular endothelial growth factor.

propeprties [18]. Other than these considerations, an important requirement of microsphere delivery system is syringeability through hypodermic needles. Several microsphere preparations available commercially include Lupron depot, Posilac, Trelstar depot, and Decapeptyl [19]. Although a number of microsphere formulations are on the market, these injectable particulate delivery systems

possess two major limitations, initial burst release (rapid release of encapsulated drug) and very slow release after the initial burst period, which is known as "lag time" [10]. Whereas the initial burst release may cause adverse drug reactions, the lag time results in ineffective treatment because of an insufficient dose. Therefore, great attention is directed toward eliminating the occurrence of initial burst release and lag period or time to make these delivery systems clinically acceptable. Other factors such as incomplete release and instability of delicate biotherapeutics during fabrication and release are also important during the development of effective microparticle-based delivery.

### 21.3.2 Microcapsules

Although more complex than microspheres, microcapsules are receiving increasing interest as potential drug delivery systems. Microcapsules consist of two essential parts, the core and the coat. The core usually includes bioactive agents (protein, peptides, DNA, or entire cells) either in aqueous solution or in solid state with or without additives. The coat forms an inert diffusion barrier surrounding the core and should have desired coating properties such as solubility, permeability, flexibility, strength, and stability. The most commonly used biodegradable coating materials for microencapsulation include alginate [79], PLGA [32], collagen [80], and chitosan [81,82]. The incorporated drug can escape through the capsule coat by rupturing the wall, dissolution, melting, or diffusion [83]. Although several methods can be used to prepare microcapsules such as the double-emulsion method, organic phase separation, spray drying, and supercritical fluid technique [13,84,85], the stability of the active ingredient dictates the choice of a suitable method. A unique and novel method of formulating reservoir-type microcapsules has shown promise for encapsulating proteins and peptides. This method uses the interfacial phase separation technique, which helps to preserve the stability of proteins and peptides by preventing their exposure to an aqueous–organic interface [86].

### 21.3.3 Nanoparticles

Nanoparticles are defined as colloidal dispersion or solid particulate systems having a size in the range of 1 to 1000 nm [87,88]. The therapeutic agent can be distributed within the nanoparticle matrix or adsorbed or attached onto the matrix. Nanoparticles, including nanospheres and nanocapsules, are usually prepared by similar methods as those described earlier for microparticles, but in this case, the manufacturing parameters are tuned so that the particle size will remain in the nanometer range [13]. The commonly used methods for nanoparticle preparation are solvent evaporation, solvent diffusion, nanoprecipitation, ionic gelation or coacervation, double emulsification, and polymer coating using a variety of materials (e.g., polysaccharides, proteins, and synthetic polymers [PLGA, PLA, polycaprolactone]). By changing the preparation method, nanocapsules or nanospheres can be obtained. Nanocapsules exhibit a typical core-shell assembly wherein the drug is confined to a reservoir or within a cavity enclosed by a uniform polymer membrane or coating; in

nanospheres, the drug is homogenously dispersed in the matrix [89]. The aim of designing a nanoparticulate delivery system is to control the particle size, surface characteristics, and release profile of the incorporated biotherapeutic agent to maintain site-specific action and dosing regimen. Because of their small dimensions, nanoparticles can easily enter through the cell membrane compared with microparticles. The cellular uptake of nanoparticles is inversely proportional to their small dimensions; for example, particles with an average diameter of about 100 nm are phagocyted by cells 2.5 times more than particles having an average diameter of 1 μm [90]. Biodegradable polymeric nanoparticles have become an important delivery system because of their ability to maintain the stability of encapsulated drugs and to deliver a vast variety of therapeutic agents, including hydrophilic and hydrophobic drugs, peptides, proteins, vaccines, and other bioactive agents, to various organs for a prolonged period of time [91–94]. But the major problems associated with nanoparticulate delivery systems include particle aggregation, low loading capacity, high burst release, and clearance from the blood circulation by phagocytosis. To overcome these limitations, nanoparticles are often modified with polymeric surfactants or other modifiers to reduce aggregation, prevent burst release, and improve stability in the presence of plasma proteins [95–97].

### 21.3.4 Hydrogels

In recent years, hydrogel-based systems have been receiving increasing interest for the controlled delivery of hydrophilic biological macromolecules such as peptides [98,99], proteins [57,100], and DNA [101,102] because of their capacity to provide a protective environment and ability to control the diffusion by modulating the extent of cross-linking within the polymer matrix. A hydrogel can be defined as a cross-linked polymeric network that has the capacity to swell without dissolving in an aqueous environment. Hydrogels are usually stabilized through physical or chemical cross-linking between polymeric units forming a gel backbone and have porous morphology. Generally, at equilibrium, the hydrogels consist of about 60–90% fluid and only 10–30% polymer. Their high water content makes them compatible with proteins, living tissues, and DNA because of their structural resemblance to the various components of the body [103]. Classification of hydrogels is based on several features summarized by Hamidi et al. [104]. The classification can be based on the nature of polymers, method of preparation, physical structure, and responsiveness to external stimuli. Depending on the polymer source, hydrogels can be classified into three groups: natural, synthetic, and hybrid polymer–based hydrogels. Natural polymers may be neutral (dextran, agarose, pullulan), cationic (chitosan, polylysine,), anionic (alginic acid, hyaluronic acid, pectin, chondroitin sulfate, dextran sulfate), or amphipathic (collagen, fibrin, carboxymethyl chitin) in nature [104]. Although natural polymers possess inherent biocompatibility, their use is also coupled with a number of drawbacks, such as poor mechanical strength, large batch-to-batch variation, and contamination issues. Therefore, synthetic polymers have been increasingly used for preparation of hydrogels. Among the synthetic polymers, “smart polymers” containing various hydrophobic and hydrophilic blocks

have gained tremendous attention. These polymers can be tailormade with different physical and chemical properties, hydrophobicity or hydrophilicity, degradation rate, and duration as per the delivery need. Smart polymers can be classified according to the external stimulus they respond to such as temperature, phase, pH, light, enzyme, ionic strength, electromagnetic radiation, and so on. These polymers show instantaneous changes in their physicochemical properties in response to slight alterations in their microenvironment [105]. Injectable *in situ*–forming hydrogels offer several benefits over their implantable counterparts. For instance, they permit easy homogenous encapsulation of biotherapeutic agent and can form hydrogel *in situ*, avoiding the need for surgical implantation and explantation [51,105–107].

A number of phase-sensitive and temperature-sensitive polymers have received attention as parenteral controlled delivery vehicles for various protein- and peptide-based therapeutics. The advantages of smart polymer–based delivery systems include easy manufacturing, high loading capacity, and injectability [105]. In 1990, Ravivarapu et al. [108] patented the approach of dispersing drug molecules in biodegradable polymers dissolved in pharmaceutically accepted water miscible solvents. This depot-forming delivery system was known as Atrigel, and today seven products have been approved by the Food and Drug Administration based on the Atrigel technology [109].

Another class of smart polymers that has been extensively studied for controlled delivery of biologics is thermosensitive polymers. These polymers change their solubility in response to changes in temperature, and this property has been used to deliver various proteins and peptides. The major advantages of using these polymers as controlled delivery vehicles include no need of organic solvents that could damage the delicate structure of proteins or peptides, easy to prepare, injectable, biodegradable, and biocompatible; also, they do not need any surgical procedures for implantation and explantation [110,111]. The most widely used biodegradable polymers include PLGA and poly(ethylene) glycol (PEG)–based thermosensitive systems (Regel), Pluronics (PF-127), poly(organophosphazenes), and *N*-isopropyl acrylamide (NIPAAm) [111]. Hybrid polymers, the third type, are synthesized by combining both natural and synthetic polymers in an attempt to exploit the benefits of both classes. They can be prepared by blending or covalently crosslinking procedures. Some of these polymers include collagen-acrylate, alginate-acrylate, P(PEG-co-peptides), and alginate-g-(PEO–PPO–PEO) [104].

### 21.3.5 Nanogels

Although the main focus is on macroscopic hydrogels, nanogels and microgels are becoming increasingly popular in the field of controlled drug delivery for biologics. These colloidal systems have properties that are similar to macrogels, and they show additional benefits over nanoparticles. The combined characteristics of nanoparticles (nanometer in size) and hydrogels make them exclusive carriers for proteins, peptides, and DNA. Nanogel synthesis and drug loading are simple procedures and do not involve any organic solvent or harsh fabrication procedures that are detrimental to therapeutic agents. They have high drug loading capacity

and good transport and permeation capabilities because of their extremely small size (10–200 nm), they entrap and protect drug substances from immune cells, and they are easy to modify for targeting specific cells or tissues [53,112,113]. The widely studied hydrophobic core–forming polymers used for nanogel preparation include PLA, PLGA, polycaprolactone (PCL), polypropylene glycol, and polyhydroxybutyrate (PHB) [113]. Various natural polymers (dextrin, chitosan) and smart polymer–based nanogels have been successfully used for the delivery of DNA and proteins [54,114,115]. Some recent studies show the successful delivery of oligonucleotides across the blood–brain barrier (using a nanogel approach [53].

### 21.3.6 Peptide Hydrogels

Peptide hydrogel consists of periodic alternating hydrophobic and hydrophilic amino acids, which upon exposure to electrolyte solutions undergo spontaneous self-organization into nanosize fibers. These nanofibers further form highly hydrated hydrogels with the pore size ranging from 5 to 200 nm [16]. They also possess most appealing properties such as avoidance of toxic cross-linking agents during preparation, ease of injectability, nonimmunogenic nature, biodegradability, biocompatibility, and degradation to nontoxic end products [16]. These nanofibers hydrogels have been widely used to deliver proteins, peptides, DNA, and cells and as three-dimensional scaffolds for tissue engineering [55–58,116,117]. The release of protein from the peptide network is governed by charge and the size of encapsulated protein and their steric interactions with peptide network [57].

### 21.3.7 Liposomes

Liposomes are microscopic spherical vesicles that consist of an aqueous core enclosed by a lipid bilayers (or multilayer) of natural or synthetic phospholipids. The size of liposomes varies from as small as 20 nm (small unilamellar vesicles [SUVs]) to several micrometers (multilamellar vesicles [MLVs]) [118,119]. Because of the biphasic nature of liposomes, they can be used for delivery of both lipophilic and hydrophilic molecules, and depending on their solubility and partition coefficient, drug molecules get entrapped either in the aqueous phase or liposomal hydrophobic environment. The manufacturing protocol and the choice and amount of bilayer components dictate the release kinetics and encapsulation efficiency of a liposome. The drug-entrapping efficiency of liposomes depends on several factors, including the size of liposomes, types of phospholipids used for liposome preparation, and physicochemical properties of the encapsulated agent. For example, small SUVs have poor drug-loading capacity compared with MLVs. However, large unilamellar vesicles provide balance between size and drug-entrapping capacity. The established safety and efficacy profile of liposomes makes them attractive candidates for the delivery of hydrophobic or hydrophilic molecules, proteins, peptides, and vaccines. Furthermore, liposomes can be used for active targeting by attaching specific ligands to their surfaces. The most important feature of liposome-based delivery systems is attributed to the fact that the encapsulated

water-soluble therapeutic agent remains in the aqueous core with liposomal membrane protecting it against various destabilizing factors. Multivesicular liposomes or Depofoam particles have been used to circumvent the issues associated with conventional liposomes, resulting in sustained release delivery of protein or peptide drugs [66]. Liposomes are usually administered via the IV route, which poses certain issues. Liposomal interaction with high-density lipoproteins causes an early release of the encapsulated biotherapeutic agent [121]. In addition, because of the rapid uptake by the reticuloendothelial system, the overall circulation time of liposomes gets reduced [122,123]. PEG grafting on the liposome surface can help to counteract the issue with serum protein opsonization and interaction with mononuclear phagocytic cells. This helps increase the retention time of liposomes in the blood [124]. Liposomes are often combined with other controlled delivery systems to obtain the advantageous properties of both. Polymer–liposome composite systems are developed so as to provide enhanced stability to the liposomes and to control the *in vivo* release profile of encapsulated bioactive agents [125].

### 21.3.8 Vesicular Phospholipid Gels

This delivery system consists of concentrated semisolid, aqueous phospholipid dispersions of vesicular morphology [71,126]. They are usually prepared by a process known as "forced hydration" of phospholipids in the presence of a comparatively small amount of water [71]. The size, morphology, and structure of the vesicles depend on the phospholipid content [127,128]. Vesicular phospholipid gel (VPGs)s can be used as intermediates for liposomes and are converted into liposome dispersions upon dilution. Some of the advantages of VPGs over traditional liposomes include better encapsulation efficiency because more drug gets entrapped or localized within and in between the vesicles of polymer matrix. Also, they are easy to manufacture and scale up; there is no use of organic solvent, eliminating harmful effects on delicate proteins and peptides; and the delivery matrix is biocompatible [71].

### 21.3.9 Suspensions

This is the most common and widely used dosage form in parenteral controlled delivery systems. After SC administration, the drug incorporated in aqueous or oil suspension forms a depot at the site of injection, which acts as a reservoir. The incorporated molecule is released slowly and continuously depending on its aqueous solubility and dissolution of particles into surrounding tissue fluid [129]. One of the classic examples of this delivery system is a insulin or zinc–insulin suspension used for the treatment of diabetes.

### 21.3.10 Implantable Delivery Systems

Implants are small rod-shaped or cylindrical devices that are either used for systemic or local long-term delivery of biologics and small molecule drugs [9]. For systemic

delivery, they are usually placed intramuscularly or in SC tissue from where the drug enters into blood circulation; for local effects, implants are placed in specific sites in the body. The benefits of using implants for controlled delivery include easy manufacturing, tunable drug release, and ability to be removed in case of any adverse effects. A few nonbiodegradable implants are commercially available for long-term controlled delivery of steroids [130–133]. Implantable systems provide many advantages such as prolonged activity and targeting. Among the diverse technology available for the controlled delivery of biologics, delivery systems based on nonbiodegradable, hydrophobic polymers are the most successful. For example, the Norplant system is a reversible contraceptive device consisting of six small matchstick-sized capsules that are placed just beneath the skin of the upper arm for the controlled delivery of progestin over a 5-year period [134]. A new approach of osmotic implantable delivery systems that was described in 1955 has become a widely accepted tool for controlled delivery of various biotherapeutics agents [135]. An osmotically driven subcutaneously implantable delivery system has been designed for the delivery of a peptide, leuprolide, over a year-long duration for palliative treatment of prostate cancer [136]. Another method includes an SC infusion patch containing an exterior protein or peptide pump reservoir that is designed to release protein continuously to achieve a constant level for a prolonged period [137]. Some of the important characteristics of an ideal pump-based delivery system include an accurate and precise delivery rate, easy to program depending on the need of the patient, reliable, safe, free from inflammatory reactions after implantation, no leakage or overdosing, long reservoir life, easy to use by both the patient and health care professional, and sufficient long battery life [138]. Although nonbiodegradable implants are developed for this purpose, their patient acceptance is low because of the need for a large-bore needle for implantation, need for medical professionals, discomfort, and the surgical insertion and retrieval procedures involved. Additionally, the cost of the implant production is high, and the instability of encapsulated drug during manufacturing is still a major issue. In the case of implants, fibrous capsule formation is the most common foreign body reaction observed that gives rise to undesirable effects on drug release from the implant [139]. Biodegradable implants provide a benefit of a single procedure (no explantation), and a large number of PLGA- and PLA-based implants have been evaluated for protein and peptide delivery [140,141]. These implants are degraded into acidic byproducts and are naturally eliminated in the body. Recently, an agarose hydrogel–based implant system has been developed by Toussaint et al. [142] for the delivery of DNA vaccine in cattle.

### 21.3.11 Resealed Erythrocytes

Red blood cells, or erythrocytes, are the most abundant cells in the body and have the potential for controlled delivery of various biotherapeutics. They are biocompatible and biodegradable carriers, are nonimmunogenic, and possess very long circulation half-lives. They can protect the incorporated drug from immunologic detection and enzymatic degradation, do not require any chemical modification for entrapment of

drug molecules, and have controllable lifespan. It is also possible to obtain ideal zero-order release kinetics using resealed erythrocytes [11,73]. The drug release from the erythrocyte can take place by (1) passive diffusion; (2) membrane-associated carrier transport; (3) phagocytosis by macrophages, leading to accumulation of drug into macrophages followed by slow release; or (4) accumulation into lymph nodes after SC administration with subsequent hemolysis and slow release of the incorporated molecule [143].

### 21.3.12 Composite Delivery Systems

A conventional single delivery system may suffer from some restrictions for the delivery of therapeutic agents; therefore, to compensate for the disadvantages associated with a single system, combinations of two or more delivery systems have been explored. With the use of composite delivery systems, it is possible to tune the release period and duration as per the need. The most widely studied combination delivery systems for controlled protein delivery include microparticle–hydrogel, liposome–hydrogel, and microcapsule–nanoparticle [74–78,144,145]. The limitations of liposomal delivery system, such as instability, short half-life, and rapid clearance from the body, can be overcome by using a liposome– hydrogel approach. The new composite delivery system can help to improve liposome stability, and the drug release can be controlled while preserving the bioactivity of the incorporated molecule [146].

## 21.4 FUTURE DIRECTIONS

It is evident by the progress in new drug delivery technologies that much attention has been devoted toward the way in which biologics are delivered. Present efforts in the field of delivery of biologics have accelerated and widened the possibility of their use with therapeutic agents for treatment of various chronic disorders. From the times when achieving a zero-order release profile was just an imaginary concept and controlled release technology was known only to a few scientists, it has now been established widely and has gained significant applications in a variety of fields ranging from gene therapy to tissue engineering. Increasing numbers of publications, patents, and commercial availability of controlled delivery products substantiate their significant advancement. Many pharmaceutical industries are now focusing on the development of new delivery systems for controlling the release of novel biologic agents. The concept of using specific targeting moieties and feedback control mechanisms has shown tremendous growth over the past few decades. The two most appealing controlled delivery systems that will be preferred in the future include targeted delivery systems and feedback-controlled delivery systems. Another significant advancement in the field of controlled delivery is the use of smart polymers. These polymer-based delivery systems are able to control the release of incorporated molecules in response to external stimuli, rendering them ideal for delivery of various biologics, including proteins and peptides, genes, and vaccines as per the needs of a patient [105].

## REFERENCES


1. Degim IT, Çelebi N. Controlled delivery of peptides and proteins. Curr Pharm Des 2007;13(1):99–117.
2. Rader RA. (Re)defining biopharmaceutical. Nat Biotech 2008;26(7):743–751.
3. Hu G. Understanding the basics of peptide and protein production. Bioprocess Int 2010;8(4):22–25.
4. Reichert J, Pavlou A. Monoclonal antibodies market. Nat Rev Drug Discov 2004;3(5): 383–384.
5. Venkataramanan M. Spoonful of medicine: biologics market predicted to see 6.5% yearly growth in demand, spurred by new cancer and infectious disease therapies. Available from: http://blogs.nature.com/spoonful/2011/10/biologics_market_predicted_to_1.html.
6. Rhee YS, Park CW, DeLuca PP, Mansour HM. Sustained-release injectable drug delivery systems. Pharm Tech 2010;11:16–22.
7. Saroja C, Lakshmi P, Bhaskaran S. Recent trends in vaccine delivery systems: a review. Int J Pharma Investig 2011;1(2):64–74.
8. Garmory HS, Perkins SD, Phillpotts RJ, Titball RW. DNA vaccines for biodefense. Adv Drug Deliv Rev 2005;57(9):1343–1361.
9. Shi Y, Li LC. Current advances in sustained-release systems for parenteral drug delivery. Expert Opin Drug Deliv 2005;2(6):1039–1058.
10. Wang J, Wang BM, Schwendeman SP. Characterization of the initial burst release of a model peptide from poly(D,L-lactide-co-glycolide) microspheres. J Control Release 2002;82(2–3):289–307.
11. Leung S-HS, Robinson JR, Lee VHL. Parenteral products. In: Lee VHL, Robinson JR, editors. Controlled Drug Delivery, Fundamentals and Applications. 2nd ed. New York: Marcel Dekker; 1987, p. 433–480.
12. Kost J, Langer R. Responsive polymeric delivery systems. Adv Drug Deliv Rev 2001; 46(1–3):125–148.
13. Jain RA. The manufacturing techniques of various drug loaded biodegradable poly (lactide-co-glycolide) (PLGA) devices. Biomaterials 2000;21(23):2475–2490.
14. Park K, Mrsny RJ. Controlled drug delivery: present and future. In: Park K, Mrsny RJ, editors. Controlled Drug Delivery Designing Technologies for the Future. Washington, DC: American Chemical Society; 2000, p. 2–12.
15. Ulery BD, Nair LS, Laurencin CT. Biomedical applications of biodegradable polymers. J Polym Sci B Polym Phys 2011;49(12):832–864.
16. Koutsopoulos S, Unsworth LD, Nagai Y, Zhang S. Controlled release of functional proteins through designer self-assembling peptide nanofiber hydrogel scaffold. Proc Natl Acad Sci USA 2009;106(12):4623–4628.
17. Walter E, Moelling K, Pavlovic J, Merkle HP. Microencapsulation of DNA using poly (DL-lactide-co-glycolide): stability issues and release characteristics. J Control Release 1999;61(3):361–374.
18. Sinha VR, Trehan A. Biodegradable microspheres for protein delivery. J Control Release 2003;90(3):261–280.

19. Mishra N, Goyal A, Khatri K, Vaidya B, Paliwal R, Rai S, et al. Biodegradable polymer based particulate carrier(s) for the delivery of proteins and peptides. Antiinflamm Antiallergy Agents Med Chem 2008;7:240–251.
20. Zhou X, Liu B, Yu X, Zha X, Zhang X, Wang X, et al. Controlled release of PEI/DNA complexes from PLGA microspheres as a potent delivery system to enhance immune response to HIV vaccine DNA prime/MVA boost regime. Eur J Pharm Biopharm 2008;68(3):589–595.
21. Nguyen DN, Raghavan SS, Tashima LM, Lin EC, Fredette SJ, Langer RS, et al. Enhancement of poly(orthoester) microspheres for DNA vaccine delivery by blending with poly(ethylenimine). Biomaterials 2008;29(18):2783–2793.
22. Ho J, Huang Y, Danquah MK, Wang H, Forde GM. Synthesis of biodegradable polymer-mesoporous silica composite microspheres for DNA prime-protein boost vaccination. Eur J Pharm Sci 2010;39(5):412–420.
23. Zhou X, Liu B, Yu X, Zha X, Zhang X, Chen Y, et al. Controlled release of PEI/DNA complexes from mannose-bearing chitosan microspheres as a potent delivery system to enhance immune response to HBV DNA vaccine. J Control Release 2007;121 (3):200–207.
24. Rosas JE, Pedraz JL, Hernández RM, Gascón AR, Igartua M, Guzmán F, et al. Remarkably high antibody levels and protection against *P. falciparum* malaria in *Aotus* monkeys after a single immunisation of SPf66 encapsulated in PLGA microspheres. Vaccine 2002;20(13–14):1707–1710.
25. Saini V, Jain V, Sudheesh MS, Jaganathan KS, Murthy PK, Kohli DV. Comparison of humoral and cell-mediated immune responses to cationic PLGA microspheres containing recombinant hepatitis B antigen. Int J Pharm 2011;408(1–2):50–57.
26. Tran V-T, Karam J-P, Garric X, Coudane J, Benoît J-P, Montero-Menei CN, et al. Protein-loaded PLGA-PEG-PLGA microspheres: a tool for cell therapy. Eur J Pharm Sci 2012;45(1–2):128–137.
27. Manoharan C, Singh J. Insulin loaded PLGA microspheres: effect of zinc salts on encapsulation, release, and stability. J Pharm Sci 2009;98(2):529–542.
28. Manoharan C, Singh J. Evaluation of polyanhydride microspheres for basal insulin delivery: effect of copolymer composition and zinc salt on encapsulation, *in vitro* release, stability, *in vivo* absorption and bioactivity in diabetic rats. J Pharm Sci 2009;98(11):4237–4250.
29. Prabhu S, Sullivan JL, Betageri GV. Comparative assessment of *in vitro* release kinetics of calcitonin polypeptide from biodegradable microspheres. Drug Deliv 2002;9(3):195–198.
30. Capan Y, Jiang G, Giovagnoli S, Na K-H, DeLuca P.P. Preparation and characterization of poly(D,L-lactide-co-glycolide) microspheres for controlled release of human growth hormone. AAPS PharmSciTech 2003;4(2):147–156.
31. Spiller KL, Liu Y, Holloway JL, Maher SA, Cao Y, Liu W, et al. A novel method for the direct fabrication of growth factor-loaded microspheres within porous nondegradable hydrogels: controlled release for cartilage tissue engineering. J Control Release 2012;157(1):39–45.
32. Yamaguchi Y, Takenaga M, Kitagawa A, Ogawa Y, Mizushima Y, Igarashi R. Insulin-loaded biodegradable PLGA microcapsules: initial burst release controlled by hydrophilic additives. J Control Release 2002;81(3):235–249.

33. Zheng J, Yue X, Dai Z, Wang Y, Liu S, Yan X. Novel iron–polysaccharide multilayered microcapsules for controlled insulin release. Acta Biomater 2009;5(5):1499–1507.
34. Kim BS, Oh JM, Kim KS, Seo KS, Cho JS, Khang G, et al. BSA-FITC-loaded microcapsules for *in vivo* delivery. Biomaterials 2009;30(5):902–909.
35. De Temmerman M-L, Rejman J, Vandenbroucke RE, De Koker S, Libert C, Grooten J, et al. Polyelectrolyte LbL microcapsules versus PLGA microparticles for immunization with a protein antigen. J Control Release 2011;158(2):233–239.
36. Sah H, Toddywala R, Chien YW. The influence of biodegradable microcapsule formulations on the controlled release of a protein. J Control Release 1994;30(3):201–211.
37. Anandhakumar S, Nagaraja V, Raichur AM. Reversible polyelectrolyte capsules as carriers for protein delivery. Colloids Surf B Biointerfaces 2010;78(2):266–274.
38. Basarkar A, Singh J. Poly(lactide-*co*-glycolide)-polymethacrylate nanoparticles for intramuscular delivery of plasmid encoding interleukin-10 to prevent autoimmune diabetes in mice. Pharm Res 2008;26(1):72–81.
39. Park Y-M, Shin B-A, Oh I-J. Poly(L-lactic acid)/polyethylenimine nanoparticles as plasmid DNA carriers. Arch Pharm Res 2008;31(1):96–102.
40. Perez C, Sanchez A, Putnam D, Ting D, Langer R, Alonso M. Poly(lactic acid)-poly (ethylene glycol) nanoparticles as new carriers for the delivery of plasmid DNA. J Control Release 2001;75(1–2):211–224.
41. Okamoto S, Yoshii H, Akagi T, Akashi M, Ishikawa T, Okuno Y, et al. Influenza hemagglutinin vaccine with poly(γ-glutamic acid) nanoparticles enhances the protection against influenza virus infection through both humoral and cell-mediated immunity. Vaccine 2007;25(49):8270–8278.
42. Bessa PC, Machado R, Nürnberger S, Dopler D, Banerjee A, Cunha AM, et al. Thermoresponsive self-assembled elastin-based nanoparticles for delivery of BMPs. J Control Release 2010;142(3):312–318.
43. Golub JS, Kim Y-tae, Duvall CL, Bellamkonda RV, Gupta D, Lin AS, et al. Sustained VEGF delivery via PLGA nanoparticles promotes vascular growth. Am J Physiol Heart Circ Physiol 2010;298(6):H1959–H1965.
44. Li Z, Ning W, Wang J, Choi A, Lee P-Y, Tyagi P, et al. Controlled gene delivery system based on thermosensitive biodegradable hydrogel. Pharm Res 2003;20(6):884–888.
45. Quick DJ, Anseth KS. DNA delivery from photocrosslinked PEG hydrogels: encapsulation efficiency, release profiles, and DNA quality. J Control Release 2004; 96(2):341–351.
46. Wieland JA, Houchin-Ray TL, Shea LD. Non-viral vector delivery from PEG-hyaluronic acid hydrogels. J Control Release 2007;120(3):233–241.
47. Qiao M, Chen D, Hao T, Zhao X, Hu H, Ma X. Injectable thermosensitive PLGA-PEG-PLGA triblock copolymers-based hydrogels as carriers for interleukin-2. Pharmazie 2008;63(1):27–30.
48. Tang Y, Singh J. Thermosensitive drug delivery system of salmon calcitonin: *in vitro* release, *in vivo* absorption, bioactivity and therapeutic efficacies. Pharm Res 2010; 27(2):272–284.
49. Al-Tahami K, Oak M, Mandke R, Singh J. Basal level insulin delivery: *in vitro* release, stability, biocompatibility, and *in vivo* absorption from thermosensitive triblock copolymers. J Pharm Sci 2011;100(11):4790–4803.

50. Oak M, Singh J. Controlled delivery of basal level of insulin from chitosan-zinc-insulin-complex-loaded thermosensitive copolymer. J Pharm Sci 2012;101(3):1079–1096.
51. Chen S, Singh J. Controlled release of growth hormone from thermosensitive triblock copolymer systems: *in vitro* and *in vivo* evaluation. Int J Pharm 2008;352(1-2):58–65.
52. Chen S, Pederson D, Oak M, Singh J. *In vivo* absorption of steroidal hormones from smart polymer based delivery systems. J Pharm Sci 2010;99(8):3381–3388.
53. Vinogradov SV, Batrakova EV, Kabanov AV. Nanogels for oligonucleotide delivery to the brain. Bioconjug Chem 2004;15(1):50–60.
54. Carvalho V, Castanheira P, Madureira P, Ferreira SA, Costa C, Teixeira JP, et al. Self-assembled dextrin nanogel as protein carrier: controlled release and biological activity of IL-10. Biotechnol Bioeng 2011;108(8):1977–1986.
55. Schneider A, Garlick JA, Egles C. Self-assembling peptide nanofiber scaffolds accelerate wound healing. PLoS One 2008;3(1):e1410.
56. Bulut S, Erkal TS, Toksoz S, Tekinay AB, Tekinay T, Guler MO. Slow release and delivery of antisense oligonucleotide drug by self-assembled peptide amphiphile nanofibers. Biomacromolecules 2011;12(8):3007–3014.
57. Branco MC, Pochan DJ, Wagner NJ, Schneider JP. The effect of protein structure on their controlled release from an injectable peptide hydrogel. Biomaterials 2010;31(36):9527–9534.
58. Liu Y, Wu M, Lin B, Zhao X. Controlled release of fuctional proteins IGF-1, aFGF and VEGF through self-assembling peptide nanofiber hydrogel. Sheng Wu Yi Xue Gong Cheng Xue Za Zhi 2011;28(2):310–313.
59. Nishimura A, Hayakawa T, Yamamoto Y, Hamori M, Tabata K, Seto K, et al. Controlled release of insulin from self-assembling nanofiber hydrogel, PuraMatrix™: application for the subcutaneous injection in rats. Eur J Pharm Sci 2012;45(1–2):1–7.
60. Sakiyama-Elbert SE, Das R, Gelberman RH, Harwood F, Amiel D, Thomopoulos S. Controlled-release kinetics and biologic activity of platelet-derived growth factor-BB for use in flexor tendon repair. J Hand Surg Am 2008;33(9):1548–1557.
61. Shea LD, Smiley E, Bonadio J, Mooney DJ. DNA delivery from polymer matrices for tissue engineering. Nat Biotech 1999;17(6):551–554.
62. des Rieux A, Ucakar B, Mupendwa BPK, Colau D, Feron O, Carmeliet P, et al. 3D systems delivering VEGF to promote angiogenesis for tissue engineering. J Control Release 2011;150(3):272–278.
63. Kim MS, Bhang S-H, Yang HS, Rim NG, Jun I, Kim SI, et al. Development of functional fibrous matrices for the controlled release of basic fibroblast growth factor to improve therapeutic angiogenesis. Tissue Eng Part A 2010;16(10):2999–3010.
64. Mohl S, Winter G. Continuous release of rh-interferon alpha-2a from triglyceride matrices. J Control Release 2004;97(1):67–78.
65. Lofthouse S, Kajihara M, Nagahara S, Nash A, Barcham G, Sedgmen B, et al. Injectable silicone implants as vaccine delivery vehicles. Vaccine 2002;20(13–14):1725–1732.
66. Ye Q, Asherman J, Stevenson M, Brownson E, Katre NV. DepoFoam technology: a vehicle for controlled delivery of protein and peptide drugs. J Control Release 2000;64 (1–3):155–166.
67. Anderson PM, Hanson DC, Hasz DE, Halet MR, Blazar BR, Ochoa AC. Cytokines in liposomes: preliminary studies with IL-1, IL-2, IL-6, GM-CSF and interferon-γ. Cytokine 1994;6(1):92–101.

68. Schneider S, Weber R, Luippold G. Blood glucose profiles in diabetic rodents using different insulin preparations. Arzneimittelforschung 2004;54(12):842–846.
69. Gietz U, Arvinte T, Mader E, Oroszlan P, Merkle HP. Sustained release of injectable zinc-recombinant hirudin suspensions: development and validation of *in vitro* release model. Eur J Pharm Biopharm 1998;45(3):259–264.
70. Knepp VM, Muchnik A, Oldmark S, Kalashnikova L. Stability of nonaqueous suspension formulations of plasma derived factor IX and recombinant human alpha interferon at elevated temperatures. Pharm Res 1998;15(7):1090–1095.
71. Tian W, Schulze S, Brandl M, Winter G. Vesicular phospholipid gel-based depot formulations for pharmaceutical proteins: development and *in vitro* evaluation. J Control Release 2010;142(3):319–325.
72. Grohganz H, Tho I, Brandl M. Development and *in vitro* evaluation of a liposome based implant formulation for the decapeptide cetrorelix. Eur J Pharm Biopharm 2005;59 (3):439–448.
73. Hamidi M, Zarrin A, Foroozesh M, Mohammadi-Samani S. Applications of carrier erythrocytes in delivery of biopharmaceuticals. J Control Release 2007;118(2):145–160.
74. Bian L, Zhai DY, Tous E, Rai R, Mauck RL, Burdick JA. Enhanced MSC chondrogenesis following delivery of TGF-β3 from alginate microspheres within hyaluronic acid hydrogels *in vitro* and *in vivo*. Biomaterials 2011;32(27):6425–6434.
75. Yang J, Huo DQ, Hou CJ, Zhang GP, Yang LM, Zhang YC, et al. Fabrication of degradable microsphere/PNIPAAm hydrogel combination systems for protein delivery. Adv Mater Res 2010;160–162:1072–1076.
76. Lee J, Lee KY. Injectable microsphere/hydrogel combination systems for localized protein delivery. Macromol Biosci 2009;9(7):671–676.
77. Wang Y, Wei YT, Zu ZH, Ju RK, Guo MY, Wang XM, et al. Combination of hyaluronic acid hydrogel scaffold and PLGA microspheres for supporting survival of neural stem cells. Pharm Res 2011;28(6):1406–1414.
78. Lee J, Tan CY, Lee S-K, Kim Y-H, Lee KY. Controlled delivery of heat shock protein using an injectable microsphere/hydrogel combination system for the treatment of myocardial infarction. J Control Release 2009;137(3):196–202.
79. de Vos P, Faas MM, Strand B, Calafiore R. Alginate-based microcapsules for immunoisolation of pancreatic islets. Biomaterials 2006;27(32):5603–5617.
80. Hsu FY, Tsai SW, Wang FF, Wang YJ. The collagen-containing alginate/poly (L-lysine)/alginate microcapsules. Artif Cells Blood Substit Immobil Biotechnol 2000;28(2):147–154.
81. Aiedeh K, Gianasi E, Orienti I, Zecchi V. Chitosan microcapsules as controlled release systems for insulin. J Microencapsul 1997;14(5):567–576.
82. Silva CM, Ribeiro AJ, Figueiredo M, Ferreira D, Veiga F. Microencapsulation of hemoglobin in chitosan-coated alginate microspheres prepared by emulsification/internal gelation. AAPS J 2005;7(4):E903–E913.
83. Shekhar K, Madhu MN, Pradeep B, Banji D. A review on microencapsulation. Int J Pharm Sci Rev Res 2010;5(2):58–62.
84. Taluja A, Bae YH. Role of a novel excipient poly(ethylene glycol)-b-poly(L-histidine) in retention of physical stability of insulin at aqueous/organic interface. Mol Pharm 2007;4 (4):561–570.

85. Freitas S, Merkle HP, Gander B. Ultrasonic atomisation into reduced pressure atmosphere—envisaging aseptic spray-drying for microencapsulation. J Control Release 2004;95(2):185–195.
86. Park J, Ye M, Park K. Biodegradable polymers for microencapsulation of drugs. Molecules 2005;10(1):146–161.
87. Basarkar A, Singh J. Nanoparticulate systems for polynucleotide delivery. Int J Nanomedicine 2007;2(3):353–360.
88. Mohanraj V, Chen Y. Nanoparticles—a review. Trop J Pharm Res 2006;5(1):561–573.
89. Letchford K, Burt H. A review of the formation and classification of amphiphilic block copolymer nanoparticulate structures: micelles, nanospheres, nanocapsules and polymersomes. Eur J Pharm Biopharm 2007;65(3):259–269.
90. Singh M, Sharma R, Banerjee UC. Biotechnological applications of cyclodextrins. Biotechnol Adv 2002;20(5–6):341–359.
91. Panyam J, Zhou W-Z, Prabha S, Sahoo SK, Labhasetwar V. Rapid endo-lysosomal escape of poly(DL-lactide-co-glycolide) nanoparticles: implications for drug and gene delivery. FASEB J 2002;16(10):1217–1226.
92. Carino GP, Jacob JS, Mathiowitz E. Nanosphere based oral insulin delivery. J Control Release 2000;65(1–2):261–269.
93. Lemoine D, Préat V. Polymeric nanoparticles as delivery system for influenza virus glycoproteins. J Control Release 1998;54(1):15–27.
94. Tobío M, Sánchez A, Vila A, Soriano I I, Evora C, Vila-Jato JL, Alonso MJ. The role of PEG on the stability in digestive fluids and *in vivo* fate of PEG-PLA nanoparticles following oral administration. Colloids Surf B Biointerfaces 2000;18(3–4):315–323.
95. Bhadra D, Bhadra S, Jain P, Jain NK. Pegnology: a review of PEG-ylated systems. Pharmazie 2002;57(1):5–29.
96. Cun D, Jensen DK, Maltesen MJ, Bunker M, Whiteside P, Scurr D, et al. High loading efficiency and sustained release of siRNA encapsulated in PLGA nanoparticles: quality by design optimization and characterization. Eur J Pharm Biopharm 2011;77(1):26–35.
97. Lee M, Kim SW. Polyethylene glycol-conjugated copolymers for plasmid DNA delivery. Pharm Res 2005;22(1):1–10.
98. Dorkoosh FA, Coos Verhoef J, Ambagts MHC, Rafiee-Tehrani M, Borchard G, Junginger HE. Peroral delivery systems based on superporous hydrogel polymers: release characteristics for the peptide drugs buserelin, octreotide and insulin. Eur J Pharm Sci 2002;15(5):433–439.
99. Andreopoulos FM, Persaud I. Delivery of basic fibroblast growth factor (bFGF) from photoresponsive hydrogel scaffolds. Biomaterials 2006;27(11):2468–2476.
100. Hirakura T, Yasugi K, Nemoto T, Sato M, Shimoboji T, Aso Y, et al. Hybrid hyaluronan hydrogel encapsulating nanogel as a protein nanocarrier: new system for sustained delivery of protein with a chaperone-like function. J Control Release 2010;142(3):483–489.
101. Chun KW, Lee JB, Kim SH, Park TG. Controlled release of plasmid DNA from photocross-linked pluronic hydrogels. Biomaterials 2005;26(16):3319–3326.
102. Kasper FK, Seidlits SK, Tang A, Crowther RS, Carney DH, Barry MA, et al. *In vitro* release of plasmid DNA from oligo(poly(ethylene glycol) fumarate) hydrogels. J Control Release 2005;104(3):521–539.

103. Lee KY, Mooney DJ. Hydrogels for tissue engineering. Chem Rev 2001;101(7): 1869–1879.

104. Hamidi M, Azadi A, Rafiei P. Hydrogel nanoparticles in drug delivery. Adv Drug Deliv Rev 2008;60(15):1638–1649.

105. Al-Tahami K, Singh J. Smart polymer based delivery systems for peptides and proteins. Recent Pat Drug Deliv Formul 2007;1(1):65–71.

106. Roy K, Wang D, Hedley ML, Barman SP. Gene delivery with in-situ crosslinking polymer networks generates long-term systemic protein expression. Mol Ther 2003;7(3):401–408.

107. Singh S, Singh J. Phase-sensitive polymer-based controlled delivery systems of leuprolide acetate: *in vitro* release, biocompatibility, and *in vivo* absorption in rabbits. Int J Pharm 2007;328(1):42–48.

108. Ravivarapu HB, Moyer KL, Dunn RL. Sustained suppression of pituitary-gonadal axis with an injectable, *in situ* forming implant of leuprolide acetate. J Pharm Sci 2000;89 (6):732–741.

109. Dunn R. Application of the ATRIGEL® implant drug delivery technology for patient-friendly, cost-effective product development. Drug Deliv Technol 2003;3(6):38–44.

110. Oak M, Singh S, Tang Y, Al-Tahami K, Singh J. Thermosensitive polymers for controlled delivery of proteins. In: Li S, Tiwari A, Prabhan M, Aryal S, editors. Smart Polymer Materials for Biomedical Applications. New York: Nova Science Publishers; 2010, p. 73–86.

111. Tang Y, Oak M, Mandke R, Layek B, Sharma G, Singh J. Thermosensitive polymers for controlled delivery of hormones. In: Zilberman M, editor. Active Implants and Scaffolds for Tissue Regeneration. Heidelberg: Springer; 2011, p. 457–479.

112. Shidhaye S, Lotlikar V, Malke S, Kadam V. Nanogel engineered polymeric micelles for drug delivery. Curr Drug Ther 2008;3:209–217.

113. Yallapu MM, Jaggi M, Chauhan SC. Design and engineering of nanogels for cancer treatment. Drug Discov Today 2011;16(9–10):457–463.

114. Bhuchar N, Sunasee R, Ishihara K, Thundat T, Narain R. Degradable thermoresponsive nanogels for protein encapsulation and controlled release. Bioconjugate Chem 2011;23 (1):75–83.

115. Lee JI, Kim HS, Yoo HS. DNA nanogels composed of chitosan and pluronic with thermo-sensitive and photo-crosslinking properties. Int J Pharm 2009;373(1–2):93–99.

116. Webber MJ, Kessler JA, Stupp SI. Emerging peptide nanomedicine to regenerate tissuesand organs. J Intern Med 2009;267: 71–88.

117. Miller RE, Kopesky PW, Grodzinsky AJ. Growth factor delivery through self-assembling peptide scaffolds. Clin Orthop Relat Res 2011;469(10):2716–2724.

118. Lasic DD. Stealth liposomes. In: Benita S. editor. Microencapsulation: Methods and Industrial Applications. New York: Marcel Dekker; 1996, p. 297–328.

119. Yang L, Alexandridis P. Physicochemical aspects of drug delivery and release from polymer-based colloids. Curr Opin Colloid Interface Sci 2000;5(1–2):132–143.

120. Malam Y, Loizidou M, Seifalian AM. Liposomes and nanoparticles: nanosized vehicles for drug delivery in cancer. Trends Pharmacol Sci 2009;30(11):592–599.

121. Fonseca MJ, Alsina MA, Reig F. Coating liposomes with collagen (Mr 50,000) increases uptake into liver. Biochim Biophys Acta 1996;1279(2):259–265.

122. Poste G, Bucana C, Raz A, Bugelski P, Kirsh R, Fidler IJ. Analysis of the fate of systemically administered liposomes and implications for their use in drug delivery. Cancer Res 1982;42(4):1412–1422.
123. Fidler IJ, Raz A, Fogler WE, Kirsh R, Bugelski P, Poste G. Design of liposomes to improve delivery of macrophage-augmenting agents to alveolar macrophages. Cancer Res 1980;40(12):4460–4466.
124. Papahadjopoulos D, Allen TM, Gabizon A, Mayhew E, Matthay K, Huang SK, et al. Sterically stabilized liposomes: improvements in pharmacokinetics and antitumor therapeutic efficacy. Proc Natl Acad Sci USA 1991;88(24):11460–11464.
125. Machluf M, Apte RN, Regev O, Cohen S. Enhancing the immunogenicity of liposomal hepatitis B surface antigen (HBsAg) by controlling its delivery from polymeric microspheres. J Pharm Sci 2000;89(12):1550–1557.
126. Brandl M. Vesicular phospholipid gels. In: Weissig V, editor. Liposomes. Totowa, NJ: Humana Press; 2010, p. 205–212.
127. Brandl M. Vesicular phospholipid gels: a technology platform. J Liposome Res 2007;17 (1):15–26.
128. Brandl M, Massing U. Vesicular phospholipid gels. In: Gregoriadis G, editor. Liposome Technology. New York: Informa Healthcare; 2006, p. 241–260.
129. Bari H. A prolonged release parenteral drug delivery system—an overview. Int J Pharm Sci Rev Res 2010;3(1): 1–11.
130. Reiner IJ, Castracane VD. Office device to aid in Implanon insertion. Contraception 2011;84(5):445–446.
131. Cullins VE. Injectable and implantable contraceptives. Curr Opin Obstet Gynecol 1992;4(4):536–543.
132. Ladipo OA, Akinso SA. Contraceptive Implants. Afr J Reprod Health 2005;9(1):16–23.
133. Histrelin Hydrogel Implant—Valera: Histrelin implant, LHRH-Hydrogel implant, RL 0903, SPD 424. Drugs R&D 2005;6(1):53–55.
134. Shoupe D, Mishell DR. Norplant: subdermal implant system for long-term contraception. Am J Obstet Gynecol 1989;160(5):1286–1292.
135. Hill A, Geißler S, Weigandt M, Mäder K. 2011. Controlled delivery of nanosuspensions from osmotic pumps: zero order and non-zero order kinetics. J Control Release 2012;158 (3):403–412.
136. Fowler JE, Flanagan M, Gleason DM, Klimberg IW, Gottesman JE, Sharifi R. Evaluation of an implant that delivers leuprolide for 1 year for the palliative treatment of prostate cancer. Urology 2000;55(5):639–642.
137. Cleland JL, Daugherty A, Mrsny R. Emerging protein delivery methods. Curr Opin Biotechnol 2001;12(2):212–219.
138. Ranade V. Drug delivery systems. 4. Implants in drug delivery. J Clin Pharmacol 1990;30 (10):871–889.
139. Long PH. Medical devices in orthopedic applications. Toxicol Pathol 2008;36(1):85–91.
140. García JT, Jesús Dorta M, Munguía O, Llabrés M, Fariña JB. Biodegradable laminar implants for sustained release of recombinant human growth hormone. Biomaterials 2002;23(24):4759–4764.

141. Santoveña A, García JT, Oliva A, Llabrés M, Fariña JB. A mathematical model for interpreting *in vitro* rhGH release from laminar implants. Int J Pharm 2006;309 (1–2):38–43.

142. Toussaint JF, Dubois A, Dispas M, Paquet D, Letellier C, Kerkhofs P. Delivery of DNA vaccines by agarose hydrogel implants facilitates genetic immunization in cattle. Vaccine 2007;25(7):1167–1174.

143. Gothoskar AV. Resealed erythrocytes: a review. Pharm Tech 2004;28:140–158.

144. Dai C, Wang B, Zhao H, Li B, Wang J. Preparation and characterization of liposomes-in-alginate (LIA) for protein delivery system. Colloids Surf B Biointerfaces 2006;47 (2):205–210.

145. Finotelli PV, Da Silva D, Sola-Penna M, Rossi AM, Farina M, Andrade LR, et al. Microcapsules of alginate/chitosan containing magnetic nanoparticles for controlled release of insulin. Colloids Surfaces B Biointerfaces 2010;81(1):206–211.

146. Mufamadi MS, Pillay V, Choonara YE, Du Toit LC, Modi G, Naidoo D, et al. A review on composite liposomal technologies for specialized drug delivery. J Drug Deliv 2011;2011:939851.

# Routes of Delivery for Biological Drug Products

DARRELL J. IRVINE, XINGFANG SU, and BRANDON KWONG

## 22.1 INTRODUCTION

Unlike conventional small molecule drugs, many of which can be formulated effectively with excipients to avoid degradation in the stomach and that exhibit reasonably efficient uptake through the gastrointestinal (GI) tract, biological drugs may exhibit lower stability and a greater sensitivity to enzymatic degradation, making oral delivery problematic. Thus, the majority of biologics are currently administered through subcutaneous (SC) or intramuscular (IM) injection or via intravenous (IV) infusion. Recent advances in delivery of traditional biologics include methods to increase the acceptable volume of drug solutions that can be administered subcutaneously. In addition, a number of alternative routes of administration for biologic drug products are being intensively investigated at the preclinical and clinical stages, such as intranasal, pulmonary, transcutaneous, and other routes, with some first examples of products recently licensed. This chapter reviews current methods in use for marketed biologics and advanced approaches undergoing clinical testing.

## 22.2 CONVENTIONAL ADMINISTRATION OF BIOLOGICAL DRUG PRODUCTS

### 22.2.1 Systemic Administration of Biologics

Intravenous delivery has traditionally been the standard route of administration for biologic drugs such as antibody or recombinant protein therapies. Although conventional small molecule drugs are frequently delivered orally, providing the highest levels of cost effectiveness, patient comfort and compliance, and ease of administration, biologic drugs have proven to be extremely difficult to deliver in a similar manner. The acidic pH and digestive enzymes present in the GI tract can result in

*Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*,
Edited by Shayne C. Gad
Copyright © 2010 John Wiley & Sons, Inc.

significant degradations of biologic drugs. In addition, biologics can be subject to an extremely poor rate of absorption through the GI tract, a problem that may only be exacerbated by their formulation with carriers or vehicles aimed at resisting degradation [1]. On the other hand, IV delivery of biologics provides the principal advantages of (1) achieving 100% initial bioavailability with no loss of drug at the point of entry into the body and (2) the immediate dissemination of the drug through the systemic circulation, rapidly accessing all vascularized tissues throughout the body. Furthermore, IV administration allows for significantly higher injection volumes (and consequently higher therapeutic dosages) than other parenteral routes such as SC or IM injections. The following section describes examples of biologic drugs currently approved for IV administration and examines the disadvantages associated with IV delivery that have motivated the development of alternative (both local and systemic) routes of administration.

***22.2.1.1 Clinically Approved Biologics for Intravenous Delivery*** Monoclonal antibodies (mAbs) comprise the majority of the biologic drugs currently approved for IV administration. The first mAb therapy to achieve clinical approval from the Food and Drug Administration (FDA) was muromonab-CD3 (Orthoclone OKT3) in 1985, prescribed to prevent acute allograft rejection in organ transplant patients by suppressing the antitransplant T-cell response [2]. After the successful introduction of muronomab-CD3, a wide variety of other mAbs have now reached market approval in the United States and globally. Representative examples of mAbs approved for IV infusion are outlined below in Table 22.1 (part A) along with brief descriptions of the biological activity of each antibody and the indications for which they are prescribed. These include basiliximab, daclizumab, infliximab, and a large number of other mAbs that have been developed predominantly for the purpose of cancer therapy, beginning with the 1997 approval of rituximab, an anti-CD20 antibody [3]. Although certain anticancer mAbs directly target and bind to surface proteins expressed on tumor cells (e.g., rituximab, alemtuzumab, trastuzumab) [4–7], others target immune pathways and angiogenic factors in order to inhibit the progression of tumors (e.g., ipilimumab, bevacizumab) (see Table 22.1). In addition to pure mAbs, a small number of antibody–toxin conjugates using chemical agents or radioisotopes have also been approved for the treatment of various lymphomas (causing targeted cytotoxicity), as well as a limited number of recombinant cytokines such as interleukin-2 (IL-2, Proleukin) and interferon $\alpha$-2b (IFN-$\alpha$-2b), which mediate immunostimulatory antitumor responses.

Whether biologic therapies are used to treat hematologic (e.g., lymphoma) or nonhematologic indications (e.g., solid tumors and autoimmune disorders), IV infusion enables the rapid systemic distribution of the antibody or biologic agent without loss of bioavailability. Moreover, IV injected therapeutics can be rapidly dispersed to solid tumors in internal organs that may otherwise be inaccessible for local administration without surgery, as well as undetectable micrometastases that may have spread throughout the body. From a practical perspective, antibody therapy for cancer frequently requires large therapeutic doses (e.g., 8 mg/kg for trastuzumab; 375 mg/m$^2$ for rituximab) [8,9], necessitating an injection volume for complete

**TABLE 22.1 Representative Biologic Therapies Administered by Conventional Routes**

| Drug Name | Biological Agent | Indications |
|---|---|---|
| | | *A. Intravenous Infusion* |
| Muromonab-CD3 | Anti-CD3 | Organ transplant rejection |
| Basiliximab | Anti-IL-2Rα | Organ transplant rejection |
| Infliximab | Anti-TNFα | Rheumatoid arthritis, Crohn's disease, ulcerative colitis, psoriasis, ankylosing spondylitis |
| Rituximab | Anti-CD20 | Non-Hodgkin's lymphomas, rheumatoid arthritis |
| Alemtuzumab | Anti-CD52 | B-cell chronic lymphocytic leukemia |
| Cetuximab | Anti-EGFR | Colorectal cancer, head and neck cancer |
| Trastuzumab | Anti-HER2 | Metastatic breast cancer |
| Bevacizumab | Anti-VEGF | Colorectal cancer, lung cancer, glioblastoma, metastatic renal cell cancer, non–small cell lung cancer, age-related macular degeneration |
| Ipilimumab | Anti-CTLA-4 | Advanced melanoma |
| Brentuximab vedotin | Anti-CD30/MMAE conjugate | Hodgkin's lymphoma, anaplastic large cell lymphoma |
| Iodine Tositumomab | Anti-CD20/iodine-131 ($^{131}$I) | Follicular lymphoma |
| Gammagard Liquid | Human immune globulin | Primary immunodeficiencies |
| Proleukin | IL-2 | Melanoma, renal cell cancer |
| Abatacept | CTLA-4/Fc fusion | Rheumatoid arthritis |
| Intron A | IFN-α-2b | Malignant melanoma |
| Denileukin diftitox | IL-2/Diphtheria toxin conjugate | Cutaneous T-cell lymphoma |
| Pegloticase | PEGylated uricase | Chronic gout |
| Hemacord | Hematopoietic progenitor cells | Hematopoietic/immunologic reconstitution |
| Provenge | Activated autologous leukocytes | Hormone-refractory prostate cancer |
| | | *B. Subcutaneous or Intramuscular Injection* |
| Adalimumab | Anti–TNF-α | Rheumatoid arthritis, juvenile arthritis, Crohn's disease, ankylosing spondylitis, psoriasis |
| Golimumab | Anti–TNF-α | Rheumatoid arthritis, psoriasis, ankylosing spondylitis |
| Denosumab | Anti-RANKL | Osteoporosis, bone metastases from solid tumors |
| Certolizumab pegol | PEGylated anti–TNF-α | Rheumatoid arthritis, Crohn's disease |

(*continued*)

**TABLE 22.1** (*Continued*)

| Drug Name | Biological Agent | Indications |
|---|---|---|
| Gammagard Liquid | Human immune globulin | Primary immunodeficiencies |
| Etanercept | TNF antagonist | Rheumatoid arthritis, juvenile arthritis, ankylosing spondylitis, psoriasis |
| Anakinra | IL-1R antagonist | Rheumatoid arthritis |
| Abatacept | CTLA-4/Fc fusion | Rheumatoid arthritis |
| ActImmune | IFN-γ-1b | Chronic granulomatous disease, osteopetrosis |
| Roferon-A | IFN-α-2a | Chronic hepatitis C, hairy cell leukemia, chronic myelogenous leukemia |
| Intron A | IFN-α-2b | Chronic or acute hepatitis C, chronic hepatitis B, hairy cell leukemia, follicular lymphoma, Kaposi's sarcoma |
| PEGasys | PEGylated IFN-α-2a | |
| PEG-Intron | PEGylated IFN-α-2b | Chronic hepatitis C, chronic hepatitis B |
| Avonex | IFN-β-1a | Chronic hepatitis C |
| Betaseron | IFN-β-1b | Multiple sclerosis |
| Allergen immunotherapy | Allergen vaccination | Multiple sclerosis<br>Allergic rhinitis, allergic asthma, conjunctivitis, insect allergy |
| | | *C. Local Site Injection* |
| Pegaptinib | PEGylated anti-VEGF aptamer | Age-related macular degeneration (eye) |
| Ranibizumab | Anti-VEGF-A | Age-related macular degeneration (eye) |
| Catumaxomab | Anti-CD3/Anti-EpCAM | Malignant ascites in metastatic cancers (intraperitoneal) |
| Carticel | Autologous chondrocytes | Cartilage replacement (knee) |
| Laviv | Autologous fibroblasts | Cosmetic filling of wrinkles (facial) |
| | | *D. Oral Ingestion* |
| None approved in the United States | | |

EGFR: epidermal growth factor receptor; EpCAM: Epithelial cell adhesion molecule; IFN: interferon; IL: interleukin; PEG: poly(ethylene glycol); RANKL: Receptor activator of nuclear factor kappa B ligand; TNF: tumor necrosis factor; VEGF: vascular endothelial growth factor.

solubilization of the antibody that can only be administered IV. Notably, Proleukin therapy requires a significantly lower dosage than antibody therapies (0.037 mg/kg per dose) [10], but the systemic infusion of Proleukin is associated with severe adverse effects that can be potentially fatal (particularly capillary leak syndrome in internal organs, causing hypotension and reduced organ perfusion), thus requiring inpatient administration. Nevertheless, despite a clear motivation for developing alternative routes of injection, IV delivery remains the only approved route of administration for Proleukin, as SC dosing regimens aimed at optimizing therapeutic efficacy with minimal systemic toxicity have yet to achieve clinical approval [11]. Therefore, IV delivery remains the standard and optimal route of administration for the biologic therapy of cancer.

***22.2.1.2 Intravenous Delivery of Cell-Based Therapies*** There are currently only four approved biologic drug products that are classified as "cellular therapies" by the FDA (distinct from cell or tissue transplantations) [12]. Two of these approved cellular products, Hemacord and Provenge, are administered via IV infusion. Hemacord involves the transplantation of hematopoietic progenitor cells, isolated from donor cord blood, into patients requiring hematopoietic or immunologic reconstitution caused by hematopoietic disorders or a previous myeloablative therapy. Because the recommended minimum dose is 25 million cells/kg [12], IV infusion represents the only feasible option for the injection volume required, as well as for rapid systemic dispersion of the cell therapy. Similarly, Provenge, which consists of autologous leukocytes that are antigen primed and activated *ex vivo* for the treatment of metastatic prostate cancer, has a minimum dose of 50 million cells suspended in a 250-ml volume, so IV delivery remains the only practical option [12]. Furthermore, IV delivery enables the activated antitumor leukocytes to reach the prostate and lymphoid organs as quickly as possible (without necessitating surgery) while potentially allowing for targeting of metastases of the primary prostate tumor.

***22.2.1.3 Limitations of Intravenous Delivery of Biologics*** The case of Proleukin described illustrates one of the primary limitations of IV delivery: the potentially toxic side effects in nontarget organs that can be subjected to high levels of intravenously administered biologic drugs. Another recombinant cytokine therapy that has proven extremely difficult to optimize in terms of balancing efficacy with minimal toxicity has been the use of tumor necrosis factor α (TNF-α) as an immunostimulatory anticancer treatment. TNF-α is believed to act preferentially on tumor endothelium, inducing hyperpermeability that results in hemorrhagic necrosis of the tumor tissue [13]. However, systemic exposure to high levels of TNF-α can cause severe toxicities such as hypotension and septic shock–like syndrome [14]. For this reason, the use of TNF-α in cancer therapy has been limited to isolated limb perfusions (ILPs), approved in Europe for the treatment of soft tissue sarcomas and metastatic melanoma in the extremities. The procedure consists of surgically clamping major blood vessels and ligating collateral vessels to or from the limb followed by perfusion of TNF-α in combination with chemotherapy into the isolated tumor-bearing limb. Although TNF-α ILP therapy has achieved robust

response rates [13], this treatment is limited by the clinical specialization required to provide such therapy and its applicability only to cancers located in limb extremities. More generally, this illustrates a clear motivation to develop alternative forms or routes of administration, to increase the specificity of therapy delivery to localized tumor tissues, and to reduce the severity of systemic toxicity.

Another limitation of the IV infusion is the potential for rapid clearance of the therapeutic agent. Clearance of biologics from the circulation can occur either by renal filtration or nonspecific binding and uptake (e.g., endothelial pinocytosis) [15]. A short circulatory half-life can necessitate frequent injections, from weekly to daily doses, thus increasing the patient's discomfort, overall expense of the treatment regimen, and risks of patient noncompliance. Although antibody therapies (whole IgG) can persist in circulation for days or weeks because of their large size and recycling via neonatal Fc receptors (FcRns), recombinant cytokines and antibody fragments (e.g., FAb′) commonly persist for much shorter circulation times (typically minutes to hours) [16–18]. To increase the plasma half-life, a strategy known as PEGylation was developed, in which the hydrophilic polymer poly(ethylene glycol) (PEG) is covalently conjugated to the therapeutic. This increases the hydrodynamic size of the therapeutic molecule, thereby reducing its rate of renal clearance and significantly enhancing persistence in the circulation, up to several hundred-fold (Fig. 22.1a and b) [19,20]. Moreover, PEGylation has been found to reduce the immunogenicity of biologic drugs and their nonspecific binding to nontarget cells, decreasing the incidence of infusion-site reactions [21]. Successful examples of FDA-approved PEGylated biologics include Pegloticase (uric acid–metabolizing enzyme) for the treatment of chronic gout; PEGasys (INFα-2a), an antiviral therapy used to treat hepatitis B and C; PEGintron (INFα-2b) for hepatitis C; and Certolizumab pegol (anti-TNF-α FAb′ fragment) used to treat autoimmune Crohn's disease. Of these, only Pegloticase has been approved for IV administration to date, but the others have been approved for SC injection and continue to be investigated for possible regimens of IV delivery [22].

A variety of other promising strategies are currently under preclinical investigation for increasing the serum half-life or tuning the pharmacokinetics (and subsequent efficacy) of antibody therapies. These include the engineering of increased affinity of the Fc region of IgG antibodies to neonatal FcRn and the engineering of net charge on antibodies [23]. IgG binding to the FcRn enhances serum half-life by enabling antibodies to be internalized into cells, endosomally recycled to the cell surface, and released back into plasma, in essence protecting the antibody from proteolytic or clearance processes that occur extracellularly during plasma circulation. Preclinical studies have now established that improving FcRn affinity can increase serum half-life in nonhuman primates (Fig. 22.1c and d) [24,25].

### 22.2.2 Subcutaneous Administration of Biologics

The major problems of toxicity and rapid clearance after the IV delivery of biologics are both related to the near-immediate systemic availability of the infused therapy. In the successful examples noted earlier, such rapid availability may be necessary to

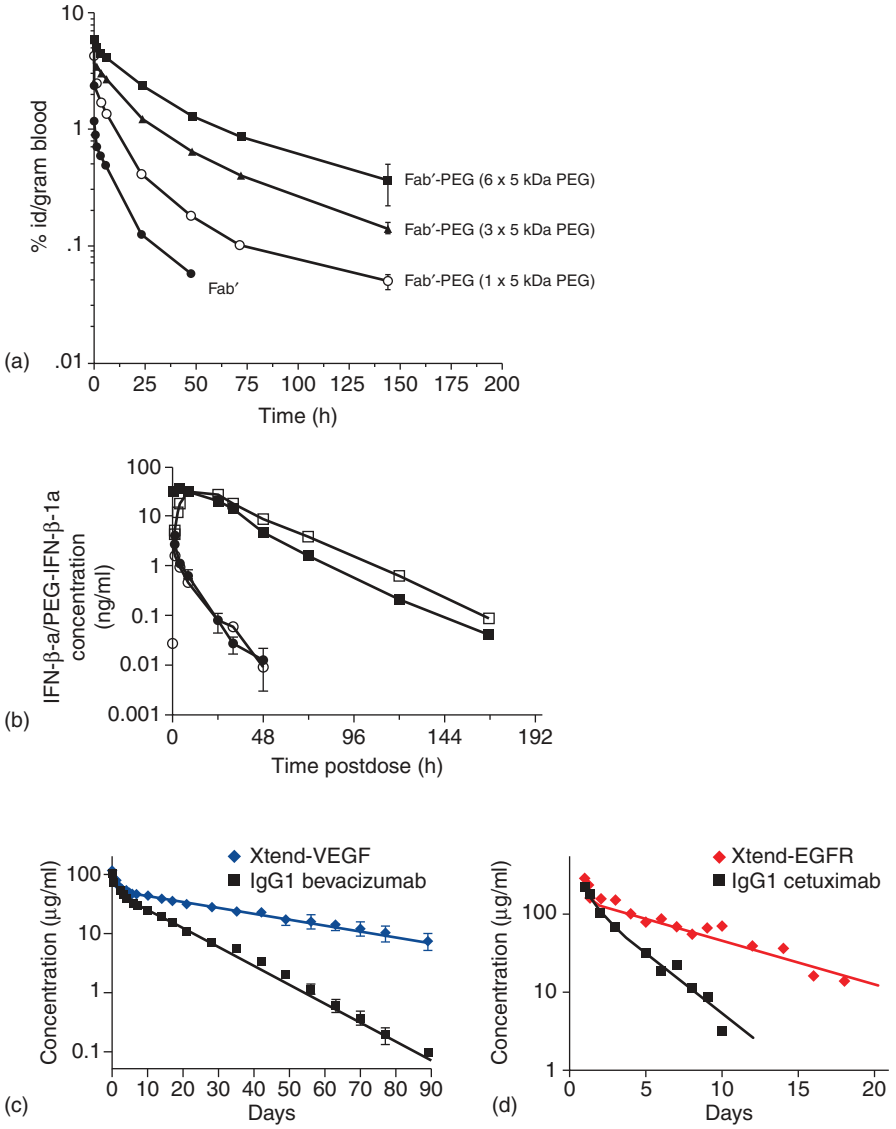


**FIGURE 22.1** Strategies for prolonging the *in vivo* half-life of biologics. (a) Conjugation of different chain lengths of polyethylene glycol (PEG) to a radiolabeled Fab' fragment increases its serum half-life after intravenous (IV) injection in rats [17]. (b) Comparison of the serum half-lives of PEGylated (square symbols) or non-PEGylated (circle symbols) interferon-β-1a (IFN-β-1a) in rhesus monkeys after intramuscular (filled symbols) or subcutaneous (open symbols) administration [18]. (c, d) Increased serum half-lives of anti–vascular endothelial growth factor (VEGF) and anti–epidermal growth factor receptor (EGFR) antibodies that were engineered (Xtend-VEGF, Xtend-EGFR) for enhanced affinity to neonatal Fc receptor (FcRn) [25].

achieve therapeutic efficacy; however, many biologic drugs benefit from a slower pharmacokinetic distribution through the body (while maintaining systemic availability). In addition, IV infusion is accompanied by the risk of catheter-mediated infection, causes significant pain and discomfort, and requires supervised inpatient administration. For more patient-friendly administration and to enable self-administration, SC delivery of biologics is of great interest. A number of licensed biologics are currently administered SC, and the success of this route of administration has been critical for biologic therapies used in managing chronic disease states or symptoms. These include a variety of autoimmune diseases (e.g., rheumatoid arthritis, ulcerative colitis, Crohn's disease, others) and other chronic indications such as chronic granulomatous disease or primary immunodeficiencies in which continuous biologic therapy is essential to compensate for chronically immunocompromised patients (refer to Table 22.1, part B). The ability to self-administer therapy via SC injection, as opposed to requiring the continual inpatient administration of IV infusions, provides an enormous benefit to patient quality of life, particularly for biologic drugs with short circulatory half-lives that may require frequent dosing.

Another context in which the use of SC injection is able to improve on IV delivery is in the case of biologic drugs with severe toxicity after systemic infusion. For example, recombinant cytokine therapies such as α and β INFs can provide therapeutic benefits in a number of diseases, including various cancers, hepatitis B and C, and multiple sclerosis [26–28], but the IV delivery of such therapies can lead to adverse events after rapid systemic dissemination of the therapeutic. SC administration provides an advantage over IV infusion by creating a relative depot effect at the site of injection: the time required for the therapeutic to drain into systemic circulation effectively prolongs the half-life of the agent *in vivo* while simultaneously lowering the peak drug level experienced by various compartments, thus reducing the frequency of dosing required as well as reducing the frequency or severity of adverse events [29]. The following section briefly reviews examples of biologic drugs in which SC injection has been approved as an alternative to or improvement upon IV infusion and discusses some unique limitations associated with the SC delivery route.

***22.2.2.1 Clinically Approved Biologics for Subcutaneous Injection*** For biologic drugs that have short half-lives *in vivo*, the availability of a formulation for self-administered SC injection provides an obvious benefit to patient quality of life. This is particularly true for the management of chronic or age-related diseases requiring a lifelong duration of therapy. For example, Anakinra (Kineret) therapy of rheumatoid arthritis requires a daily injection of 100 mg per dose [30]. Similarly, a number of TNF-α antagonists have been developed and licensed for self-administered SC injection in the treatment of rheumatoid arthritis, Crohn's disease, and other chronic autoimmune conditions (see Table 22.1, part B). In some cases, biologic therapies were first approved for IV administration before subsequent approval for self-administered SC injection has been obtained. For example, Abatacept (Orencia) and Gammagard Liquid (human immune globulin infusion) were initially prescribed

for periodic IV infusion until subsequent studies established that SC administration could provide comparable efficacy with a noninferior safety profile, including measures of the frequency and severity of infusion-site reactions, pain or tolerability, frequency of serious adverse events, and immunogenicity [31]. As a result, both abatacept and Gammagard were approved in 2011 for self-administered SC injection.

Recombinant cytokines such as α and β IFNs have proven successful in the clinical treatment of a wide variety of conditions (see Table 22.1). Whereas IFNs β-1a (Avonex) and β-1b (Betaseron) are administered solely via SC injection (for the treatment of multiple sclerosis) [32,33], the administration route for IFN-α-2a or -2b depends on the disease under treatment. For example, IFN-α-2a and -2b (either in PEGylated or non-PEGylated formulations) are only approved for SC injection in the treatment of hepatitis C, but IFN-α-2b is administered in a mixed regimen (IV induction followed by SC maintenance) for high-risk melanoma patients. While the toxicity of IFN-α-2b therapy is significant, particularly during the IV dosing phase, studies have suggested that the higher peak levels of plasma concentration obtained by IV delivery are necessary for the overall therapeutic efficacy of the treatment [34,35]. This example highlights the difficulty in achieving a tolerable but efficacious therapy with such potent biologic drugs and how tuning the regimen of drug delivery can help to optimize biodistribution and pharmacokinetics in order to reach the required therapeutic window.

***22.2.2.2 Limitations of Subcutaneous Delivery of Biologics*** Although SC injections can be used to achieve a flatter pharmacokinetic profile (a lower peak plasma concentration but a more prolonged duration at an efficacious level), a limitation of the SC route is the injection volume that can be administered without pain in humans, at a maximum of approximately 2 to 2.5 ml of fluid [29]. For regimens of some biologics, such as therapeutic antibodies in oncology applications, the most concentrated form of the drug that can be prepared in a stable, injectable form requires a dosing volume of 5 ml, making traditional SC injection problematic [36]. To overcome this volume limitation, one strategy currently in phase 3 clinical trials is the use of recombinant human hyaluronidase to locally digest hyaluronic acid, a key polysaccharide component of the extracellular matrix (ECM) in connective tissues [37]. When hyaluronidase is co-injected with biologics, it generates nanoscale porosity in the local ECM allowing rapid draining of fluid from the injection site and permitting much larger volumes of solution to be injected without pain; the matrix self-repairs within about 24 h, making the alterations only transient [38]. For example, preclinical studies in animal models have demonstrated that the addition of hyaluronidase to SC injections of a tracer dye, a fluorescently labeled dextran, and fluorescently labeled latex nanoparticles enhanced the immediate dispersion of these model agents [37,38]. Several approved mAbs in oncology (Herceptin, rituximab) that are currently given as IV infusions are being tested with this modified administration strategy to improve patient compliance with prolonged (>1 year) maintenance therapies that are indicated in conditions such as early breast cancer treatment [36]. Thus, even traditional administration routes are benefiting

from innovations in drug delivery or drug administration that may be broadly applicable to many therapeutics.

Although SC (or IM) injections enable patients to self-administer biologic therapies, thus removing the inconvenience and expense of inpatient care, the potential severity of local injection site reactions (and associated pain) remains a significant limitation of this delivery route. Common injection site reactions include erythema, edema, hematoma, swelling, pruritus, local infection, and pain [30,31,39]. Occasionally, more severe symptoms caused by anaphylactic reactions can also occur. Because SC therapies usually require frequent dosing (daily to weekly), the repeated manifestation of these symptoms often leads to poor patient adherence to the prescribed regimen of therapy. Even in the absence of significant pain, psychological reactions associated with needle injections can greatly reduce patient compliance, particularly in the case of pediatric patients [40,41]. This has motivated ongoing studies to develop alternatives to traditional needle injection for the SC delivery of biologics.

The use of automated injection devices (autoinjectors) was first developed for insulin injection for patients with chronic diabetes. Autoinjectors have now been shown to improve patient compliance in a variety of self-administered therapies compared with traditional, manually injected syringe-and-vial formulations [41]. Besides improving the convenience and ease of injection, autoinjectors such as the RebiJect and BetaJect injectors (for the SC IFN-β therapy of multiple sclerosis patients) or the SmartJect injector (for anti-TNF therapy of rheumatoid arthritis patients) have demonstrated lower incidences of injection site reactions and reduced pain [39]. Needle-free injection devices for SC delivery represent a further advance in delivery technology, such as the "cool.click" device developed for the administration of human growth hormone therapy in patients with growth hormone deficiency. The needle-free cool.click device uses a high-pressure nozzle to disperse a liquid formulation of growth hormone through the skin into the SC tissue [40]. Use of this device was particularly preferred among pediatric patients and led to overall increased compliance to prescribed regimens of therapy [42].

Despite continuing efforts to develop new technologies for SC delivery, the majority of SC biologics still require manual injection. Given the limitations associated with this delivery route, this has motivated increasing investigation into alternative delivery routes for currently established SC therapies. For example, sublingual delivery has been extensively studied as an alternative to SC injection for allergen immunotherapy of patients with allergic rhinitis or asthma [43,44]. Although the comparative efficacy of sublingual versus SC administration of allergen immunotherapy has not been fully established, sublingual therapy is highly attractive because of its improved safety profile and pain-free delivery [43]. Additional alternatives to SC delivery, such as transdermal patches, intranasal or pulmonary administration, or other mucosal routes are also discussed in the following sections.

### 22.2.3 Local Injection of Biologics

The majority of parenterally delivered biologics are injected intravenously, subcutaneously, or intramuscularly to allow systemic distribution of the drug to all

vascularized tissues. Although these delivery routes are advantageous for diseases and clinical conditions that affect internal organs or widespread tissues throughout the body, they can also result in suboptimal levels and kinetics of local drug concentration in the case of highly localized diseases. For example, diseases of the brain are especially difficult to treat via the systemic delivery of biologic drugs because biologics (and even the majority of small molecule drugs) are too large and hydrophilic to pass through the low permeability of the blood–brain barrier (BBB). Furthermore, the systemic delivery of biologic drugs can result in adverse side effects in off-target organs, which might be completely avoided by a locally restricted therapy. Therefore, for a small number of indications in which the affected tissues are accessible (with or without surgery), the local injection of biologics has been approved or is currently under development.

***22.2.3.1 Locally Injected Biologic Drug Therapies*** For the treatment of highly localized diseases, the direct injection of biologics into the affected tissue can be highly advantageous (Table 22.1, part C). A primary motivation for the local injection of biologic drugs is to minimize systemic exposure and the subsequent toxicities that may result. An example of this is the intravitreal injection (directly into the eye) of anti–vascular endothelial growth factor (VEGF) antagonists for the treatment of age-related macular degeneration (AMD). Pegaptinib (Macugen), an RNA aptamer that targets VEGF, and the anti-VEGF mAbs bevacizumab and ranibizumab have all demonstrated the ability to slow the progression of AMD [45]. However, the systemic infusion of VEGF antagonists can potentially disrupt the normal functions of VEGF in healthy vasculature throughout the body, leading to increased risks of thromboembolic events, hypertension, and impaired wound healing [45]. Early clinical studies showed that intravitreal injections of VEGF antagonists had an improved safety profile compared with IV infusion, with reduced frequency of systemic adverse events [46]. Therefore, intravitreal injection remains the only approved route of administration for these VEGF antagonists in the treatment of AMD.

Another important motivation for the local injection of biologic therapy is to maximize the local concentration of therapy, improving the efficacy of treatment on a specific tissue or organ. For example, the bispecific antibody catumaxomab (anti-CD3/anti-epithelial cell adhesion molecule EPCAM) is injected intraperitoneally for the treatment of malignant ascites resulting from epithelial, gastric, or ovarian cancers [47]. The EPCAM specificity targets EPCAM-expressing epithelial tumor cells that have invaded the peritoneal cavity. Importantly, EPCAM expression is found in 70–100% of cases of malignant ascites, and the (non-tumor) peritoneal cavity itself is lined with EPCAM-negative mesothelial cells, enabling a high level of specificity of tumor binding [48,49]. Furthermore, preclinical pharmacokinetic studies confirmed that intraperitoneal (IP) administration of catumaxomab produced high local concentrations of the antibody in the ascites fluid while significantly limiting systemic exposure (<5% detectable in plasma), a critical finding given the potential toxicity of systemic anti-CD3 stimulation [50,51]. Subsequent clinical studies demonstrated potent therapeutic efficacy in activating antitumor immune

responses, eliminating tumor cells from the ascites fluid, and prolonging overall survival. This led to the approval of Catumaxomab for IP injection in the United States and Europe, with an escalating four-dose regimen up to a maximum dose of 150 μg [47]. In contrast, a phase I study of intravenously administered catumaxomab for the treatment of non–small cell lung cancer observed systemic dose-limiting toxicities at a dose of 5 μg, thus demonstrating a clear advantage for local injection [52].

Another indication where the local delivery of biologic drugs is currently under extensive study is in the treatment of glioblastoma and other brain malignancies. Effective therapy of brain tumors is especially difficult because of the tight endothelial cell junctions that form the BBB, which blocks the diffusion of molecules between circulatory vessels and cerebrospinal fluid. Although some small hydrophobic drugs may be able to penetrate this barrier, biologic agents such as recombinant proteins and antibodies are too large and hydrophilic, so the systemic infusion of such therapies is unable to reach brain malignancies [53]. Therefore, drug delivery into the brain requires surgical assistance: by the local implantation of controlled-release devices, by catheter-mediated infusion into the brain, or by "convection-enhanced delivery" (a gradient-driven delivery via intracranially implanted microcatheters) [54,55]. A variety of targeted toxin–biologic conjugates have now been tested clinically and preclinically using these methods of local intracerebral infusion, including a bispecific ligand-directed toxin that binds to tumor-overexpressed endothelial growth factor (EGF) receptor and urokinase plasminogen activator receptor, and a recombinant fusion protein of the cytokine TGF-α (also binding EGF receptor) with a *Pseudomonas* exotoxin [56,57].

Local injection is also under investigation as a delivery method for the biologic treatment of osteoarthritis and other joint disorders. After its success in the treatment of rheumatoid arthritis, Anakinra, a recombinant IL-1 receptor antagonist, has been studied in preclinical animal models of osteoarthritis, and intraarticular administration (direct injection into the joint) was found to be the most effective route of delivery [58,59]. A recent clinical trial in patients with knee osteoarthritis determined that intraarticular Anakinra, although well tolerated, had no significant clinical benefit compared with placebo [60]. However, other studies have demonstrated the efficacy of intraarticularly injected IL-1 receptor antagonists in reducing postoperative pain and inflammation while improving knee function in patients who underwent the surgical treatment of knee injuries [61,62]. Thus, the use of intraarticular delivery for local antiinflammatory biologic agents remains a promising strategy warranting continued investigation.

***22.2.3.2 Local Administration of Cell-Based Therapies*** The recent FDA approvals of Provenge, an autologous cell-based cancer vaccine against prostate tumors, and Hemacord, a progenitor cell therapy derived from cord blood for the treatment of hematopoietic disorders, are two respective examples in which the systemic dissemination of cellular therapies via IV delivery can be advantageous or essential. However, the majority of cell transplantation therapies, for a variety of diseases and indications, have historically been administered directly at the site of disease, dating back to the development of tissue or organ transplant procedures. This is largely

because the majority of cells, when infused systemically, lack the necessary homing capacity or ability to transmigrate across endothelium in order to reach their intended tissue. Presently, there are two FDA-approved cellular biologic therapies requiring local injection: Laviv and Carticel (refer to Table 22.1). In the case of stem cell–based therapies, many of which are under ongoing research, an additional concern of systemic administration is the potential risk of nonterminally differentiated, tumorigenic cells taking residence in improper anatomical tissues [63,64]. Therefore, local administration is the primary route of delivery for a growing number of stem cell therapies that are now under preclinical or clinical study.

Numerous preclinical studies in stem or progenitor cell therapies for cardiac repair have demonstrated cellular infiltration into damaged myocardial tissue, differentiation into or regeneration of cardiomyocytes, and overall improvements in cardiac function in animal models of ischemic heart disease [65]. Although the IV infusion of autologous mesenchymal stem cells has shown the ability of achieve myocardial engraftment and subsequent therapeutic efficacy in animal models, clinical benefits have yet to be demonstrated in human trials by this injection route, possibly because of the entrapment of cells in the lungs and an inability to access the damaged myocardial tissue [66,67]. On the other hand, a large number of clinical trials are currently ongoing using direct intramyocardial injections or intracoronary vein injections of various stem and progenitor cell populations, and earlier clinical studies of locally administered cell therapies have already demonstrated promising improvements in clinical outcome [65,68,69].

Other examples of stem cell–based therapies for treating highly localized diseases include the recent, well-publicized trials of the first-ever embryonic stem (ES) cell–derived treatments attempted in human patients. The first of these was GRNOPC1 (from Geron) in which oligodendrocytes derived from human ES cells were injected into the spinal cords of patients with spinal cord injuries [70]. The second trial (from Advanced Cell Technology) involved the injection of retinal pigment epithelium cells (also derived from human ES cells) into the eyes of patients with Stargardt's macular dystrophy [71]. ES cell–derived therapies carry an even greater level of risk than adult stem cell or progenitor cell therapies because ES cells possess the highest degree of pluripotency and therefore the greatest likelihood of forming a tumor or teratoma. These examples highlight the unique challenges faced by cell-based treatments, which must balance the ability of cells to engraft into their proper tissue after systemic delivery against the challenge and cost of highly specialized surgical procedures for localized infusion therapy.

### 22.2.4 Oral Administration of Biologics

The oral delivery of therapeutic drugs is highly attractive because of its convenience and ease of administration. Although a vast number of small molecule drugs have been formulated for oral intake and absorption through the GI tract, the successful delivery of biologics such as proteins, peptides, and small interfering (siRNA) has yet to be achieved. The primary difficulties preventing effective oral delivery of biologics include the presence of proteolytic enzymes and the low pH found in the GI

tract, which can cause the degradation of biologics; the low permeability of the epithelial layer and mucosa lining the GI tract, which hinders the absorption of biologics into the bloodstream; and the hydrophilicity and macromolecular size of most biologic agents, which limits their passive diffusion transcellularly through the epithelial layer. Therefore, a variety of strategies have been proposed and studied, predominantly in the preclinical setting, in an attempt to overcome these limitations.

***22.2.4.1 Particle Carriers for the oral Delivery of Biologics*** Microparticle and nanoparticle carriers have been investigated extensively for the encapsulation and oral delivery of proteins for the dual purposes of protecting their cargo from the degradative environment of the GI tract as well as enhancing uptake across the intestinal epithelium. Historically, much of the research into particle-mediated oral delivery has been performed in the context of the development of subunit antigen vaccines [1,72]. Although live-attenuated or killed whole-cell vaccines have shown some success via oral administration (reaching global market approval in the cases of polio, cholera, rotavirus, and others), oral subunit antigen vaccines currently have yet to achieve market approval [72].

It has long been known that whole cells such as bacteria are able to adhere strongly to intestinal epithelium, thereby inducing uptake into gut-associated lymphoid tissue (GALT) and subsequent immune priming for a successful vaccination. Therefore, it was hypothesized that particulate oral delivery of protein or peptide antigens could also be used to improve the uptake of antigens into GALT (relative to soluble antigens administered orally), especially if combined with the use of surface targeting ligands or the presentation of Toll-like receptor (TLR) ligands to mimic whole pathogens [72]. A wide variety of particle systems have been studied for oral vaccine delivery, including lipid-based particles such as liposomes, polymeric particles based on synthetic materials such as poly(lactic-*co*-glycolic acid) (PLGA), and hydrogel particles based on naturally occurring polysaccharides such as alginate or chitosan [1,73,74]. Strategies for improving particle uptake have included the use of specific antibody targeting [75], the incorporation of adhesive bacterial proteins [76], and the use of immune-stimulatory adjuvants such as TLR ligands [77], and enteric coatings have been used to surround microparticles for increased protection against degradation [78–80]. However, despite convincing demonstrations of enhanced uptake into GALT, oral delivery systems for subunit vaccines have yet to achieve sufficient levels of mucosal immune protection in both preclinical and clinical studies compared with the oral administration of whole-cell vaccines [1]. Insufficient protection of the antigen in the GI tract and an inability to sustain release of the antigen after oral administration are likely critical factors responsible for this lack of efficacy.

Similarly, particulate systems have been investigated for the encapsulation and oral administration of biologic therapeutics. For example, there has been a major focus on developing strategies for the oral delivery of insulin, which is one of the most commonly prescribed protein drugs. An effective oral administration of insulin would provide enormous benefits in convenience and quality of life for diabetic patients that currently require frequent needle injections for self-dosing. However,

when administered orally in free (soluble) form, insulin and other protein drugs cannot be effectively absorbed paracellularly from the GI tract because of the highly tight permeability of the intestinal epithelium [81]. Furthermore, even if protein drugs are endocytosed into epithelial mucosal cells at the apical membrane (from the GI tract), there is minimal release or secretion of the drug at the basolateral membrane for absorption into the bloodstream [82,83]. This has prompted the investigation of particle carriers such as PLGA or other polymeric nanoparticles [84], chitosan nanoparticles [85–87], polymeric micelles [89], and others [89,90]. Encapsulation within such carriers can improve the delivery of insulin and other protein drugs by mediating uptake into, and more importantly, translocation through the epithelial mucosal lining [91]. The addition of targeting moieties to the particle surface can further improve the internalization and transcellular passage of drug-loaded particles across the intestinal epithelium [92]. Nevertheless, despite preclinical successes in animal models, the effective oral delivery of proteins via particle carriers remains elusive partly because of the poor loading efficiency of hydrophilic macromolecules into many such particle formulations, the inability of these nanoparticles to maintain colloidal stability in the harsh environment of the GI tract, and the inability to sustain or control drug release.

***22.2.4.2 Co-delivery of Modulatory Agents to Improve Absorption from the Gastrointestinal Tract*** A second general strategy for improving the oral delivery of biologics involves the co-administration of modulatory compounds aimed at either directly enhancing uptake into epithelial cells, increasing the permeability of the epithelial layer (to allow paracellular absorption), or inhibiting proteolytic activity in the GI tract. Using a diverse array of agents such as cell-penetrating peptides, detergent-like molecules, or mucoadhesive polymers, these effects have been demonstrated in both *in vitro* models of epithelial cell layers and *in vivo* animal models.

Cell-penetrating peptides (CPPs) have been widely investigated for intracellular delivery in a variety of applications, including cancer therapy, gene transfection, and drug delivery [90]. For the purpose of oral drug delivery, cell-penetrating peptides such as the TAT peptide, penetratin, and oligoarginine can theoretically enhance uptake by increasing the membrane permeability of epithelial cells in the intestinal lumen. For example, this has been demonstrated for the delivery of insulin and other protein drugs into *in vitro* cultured monolayers of Caco-2 epithelial cells [93,94] or by direct injection into intestinal loops [95–97]. In these studies, the protein drugs and CPPs have been combined either by covalent conjugation or by noncovalent association before delivery. Investigations of the overall absorption efficacy and subsequent bioavailability of the administered biologic after an actual oral delivery in animal models remains ongoing [94].

Another strategy for permeation enhancement involves the co-delivery of detergent-like or lipid-based molecules, such as the salts of fatty acids, pluronic acid, or small molecule hydrophobic compounds [98–101]. These agents can act either by directly increasing the permeability of cell membranes, thus enhancing epithelial cell uptake of the biologic, or by destabilizing the tight junctions between cells in the epithelial layer, thus increasing paracellular absorption of the drug. Another study

demonstrated the use of a toxin (zonula occludens) to reversibly weaken these epithelial cell tight junctions and enhance the absorption of orally administered insulin [102]. However, the nonspecific action of such agents poses the risk of causing severe disruptions to the anatomical GI barrier, which could then lead to significant toxicity because of the unintended passage of GI-resident pathogens or other compounds into the bloodstream [103]. As a potential strategy to mitigate this risk, a recent study described the development of a controlled-release device aimed at optimizing the delivery kinetics of an absorption enhancer (sodium glycocholate) relative to insulin delivery, although it remains to be tested in an *in vivo* model of oral administration [104].

Finally, protease inhibitors have been used in oral co-delivery with protein drugs to combat the effect of proteolytic enzymes present in the GI tract and the epithelial mucosal layer. For example, inhibitory compounds such as camostat mesilate and bacitracin can be simply co-administered with the protein drug of interest [105–108], protease inhibitors can be co-encapsulated with the biologic drug in a carrier [104], or the particle carrier itself can incorporate protease-inhibiting compounds directly in its composition while being loaded with a biologic drug cargo [87]. However, as with the use of permeation enhancers, the nonspecific action of protease inhibitors in the GI tract has the potential for complications, particularly if used at high doses, and as a result, this strategy has not yet achieved clinical success or approval [103]. Therefore, the passive avoidance of proteolytic degradation by using inert, protective nanoparticle carriers may be more likely to reach clinical translation in the future. Overall, despite extensive studies over the past three decades into the oral delivery strategies described here, the promising results observed in preclinical research have yet to develop into successful clinical strategies for oral biologic drug administration.

## 22.3 ALTERNATIVE ROUTES OF ADMINISTRATION

Although traditional routes of administration for biologics have been very successful in the clinic, these routes have limitations. Although it can be the best choice for drug formulation and pharmacokinetic and pharmacodynamic reasons, IV dosing is impractical for repeated or chronic use for non–life-threatening conditions. SC injection is simple and convenient but can have complications of injection site reactions (e.g., up to 70% of patients in the case of IL-1R antagonists) [109], which are often related to mast cell degranulation and can be caused by the solution vehicle, the high concentrations of protein typically administered, or both [30]. Finally, the biology of many diseases and candidate drugs may logically suggest that other sites of administration may be more effective for treatment, such as airway delivery of biologics for treatment of respiratory disorders or transcutaneous delivery of drugs targeting skin infections. Thus, a variety of preclinical and clinical efforts are underway to develop alternative routes of administration in a broad spectrum of biological drug products, and a few of them have already made their way to the market. Another reason for such efforts is the noninvasiveness of some alternative delivery routes for systemic applications.

### 22.3.1 Airway Delivery of Biologics

Airway delivery has frequently been proposed as the most feasible alternative to parenteral injections because of the high permeability of the nasal and pulmonary epithelium [109]. Other advantages include the large surface area, extensive vascularization, thin air–blood barrier, relatively low enzymatic activity, and no first-pass hepatic metabolism, which helps to increase bioavailability while reducing required drug doses and side effects [110–112]. Pulmonary delivery not only enables local delivery of therapeutics targeting respiratory diseases but can also access the systemic circulation where the rapid drug absorption rate can result in plasma drug profiles that sometimes parallel those from IV injections [110,112]. Finally, the noninvasive nature of the airway delivery route also helps increase patient compliance.

In 1993, the first protein administered via inhalation, the recombinant human DNase (Pulmozyme, Genentech, Inc.), was approved by the FDA for the treatment of cystic fibrosis. Today several inhaled insulin products are in development [113–115]. The furthest in development of these formulations is inhaled human insulin powder (Exubera, Pfizer), studied in adult patients with type 1 or type 2 diabetes [114,115] and approved in the United States in 2006 for the treatment of adult patients with diabetes. Unfortunately, a lack of sales led to the removal of Exubera from the market after only about 1 year [116,117]. Clinical trials by other companies (Alkermes with Eli Lilly and Aradigm with Novo Nordisk) were also discontinued for commercial reasons, although Mannkind's Afrezza Technosphere Insulin is currently pending approval by the FDA [116,117]. Other proteins and peptides under study for pulmonary administration in clinical trials include glucagon-like peptide for diabetes mellitus, IFN-$\alpha$ for antitumor therapy, IL-2 for various cancers, and Sargramostin for metastatic cancer and sarcoma [117]. In 2006, the European Medicines Agency granted orphan designation to an inhaled human IL-2 produced by Immunservice GmbH for the treatment of renal cell carcinoma [117].

Pulmonary gene delivery can mediate direct access to lung epithelial cells and correct specific genetic abnormalities in lung diseases such as cystic fibrosis, chronic obstructive pulmonary disease, asthma, and pulmonary fibrosis and possibly for the treatment of lung cancer, pulmonary hypertension, and acute lung injury [118]. Pulmonary delivery offers several advantages over systemic administration, such as improved nucleic acid stability because of lower nuclease activity in the airways than in the serum, enhancing retention in the lungs and reducing the dose required and hence any systemic toxic effects. With the recent advances in RNA interference (RNAi), two of the four siRNA drugs currently in phase II clinical trials are delivered intranasally or by inhalation [119,120].

For vaccination, FluMist (MedImmune, Gaithersburg, MD, USA), a live-attenuated influenza vaccine administered via a nasal spray and other inhaled spray-dried formulations containing whole inactivated virus or split subunit vaccine, have been developed for influenza prevention [121,122]. In the early 1990s, about 4 million children were immunized against measles with a home-built system that incorporated a jet nebulizer from IPI Medical Products (Chicago, IL, USA). Aerosolized vaccine against measles provided a stronger and more durable boosting response

than injection in school-age children and is subject to testing by the World Health Organization in mass immunization campaigns [123,124]. Similarly, an inhaled measles and rubella vaccine [125], a triple vaccine (measles, mumps, and rubella) [126], a dry powder formulation of live-attenuated measles vaccine [127], and inhaled vaccines for protection against inhaled bioterrorism agents such as anthrax and tularemia are under development [128,129].

***22.3.1.1 Barriers to Airway Delivery of Biologics*** The airways possess several features to protect against foreign molecules, including physical or mechanical (epithelial barrier and mucociliary clearance), chemical (enzymes), and immunologic defense mechanisms (macrophages) [130,131]. These must be overcome to ensure efficient delivery of biologics.

*Mucociliary Clearance* Most foreign particulates are efficiently trapped in the mucus layer by steric obstruction or adhesion. The limited permeability of molecules through the mucus barrier leads to their rapid clearance by coordinated movement of cilia that removes mucus from the delivery site toward the pharynx. The ability of the drug to penetrate the mucus barrier depends on particle charge, solubility, lipophilicity, and size [132,133]. As such, strategies to alter drug characteristics and overcome mucociliary clearance include the use of prodrugs as well as mucoadhesive drug delivery systems based on materials such as chitosan and alginate to promote retention and enhance penetration of therapeutic molecules [138].

*Enzymatic Degradation in the Epithelial and Alveolar Lining Fluid* Biologics delivered via the airways are susceptible to chemical degradation by enzymes present in the epithelial and alveolar lining fluid. The proteolytic activity in the epithelial lining fluid presents a major barrier to pulmonary delivery of therapeutic peptides because these compounds are considered to be absorbed primarily by the paracellular route [134,135]. Enzymatic hydrolysis of small natural peptides can be very high, resulting in low bioavailabilities unless they are chemically engineered (blocked) to inhibit peptidases [136]. Notably, as molecular mass increases (leading to formation of tertiary and quaternary structures), peptidase hydrolysis is inhibited, and bioavailabilities of natural proteins are consequently increased [136,137]. Modifications of drug molecules, incorporation of enzymatic inhibitors or the use of carrier particles can also help prevent enzymatic degradation of biological molecules.

*Macrophages* Macrophages in the airways, particularly alveolar macrophages located on the alveolar surface, are known to rapidly engulf foreign particles by phagocytosis as a defense mechanism. Protein therapeutics that are taken up by macrophages are eventually degraded in the phagolysosomes of the cells [138,139]. Clearance by alveolar macrophages has been demonstrated as an important mechanism of elimination of inhaled macromolecules, particularly for moderate-sized to large proteins, after delivery to the lungs and before absorption into the systemic circulation. Depletion of alveolar macrophages by liposome-encapsulated dichloromethylene diphosphonate causes a several-fold enhancement in systemic absorption of IgG

and human chorionic gonadotropin after intratracheal instillation in rats [139]. Molecules such as insulin are formulated either as liquids or in highly water-soluble aerosol particles that dissolve rapidly in the lungs and thereby largely avoid macrophage degradation [140].

*Epithelium and Tight Junctions* Underlying the mucus layer, airway epithelial cells form a tight ciliated barrier connected by tight junctions that blocks fluid flow and control the transport of ions and solutes through the intercellular space [141]. The permeation of hydrophilic substances of high molecular weight and ionic species is limited [142], with the molecular weight cutoff of the tight junctions of alveolar type I cells at 0.6 nm [143]. Toxin molecules and carrier systems based on materials such as chitosan have been proposed to reversibly open tight junctions to enhance adsorption of therapeutic molecules [144,145].

***22.3.1.2 Aerosolization of Biologics*** For biologics to be delivered to the lungs by inhalation, they must be formulated as an aerosol. Aerosol preparations are dispersions or suspensions of solid materials or liquid droplets in a gaseous medium. Drug delivered by aerosol is deposited in the airways by gravitational sedimentation, inertial impaction, and diffusion, where larger particles are deposited by the first two mechanisms in the airways and smaller particles reach the peripheral region of the lungs by diffusion [146]. For distribution throughout the lung, drug particles should be between 1 and 5 μm in diameter; for delivery into the alveolar region, 1 and 3 μm in diameter is optimum. Whereas particles that are too large cannot easily traverse the bronchial network, particles that are too small descend into the deep lungs inefficiently and are more readily lost during exhalation [146].

The efficiency of pulmonary administration is influenced not only by the aerosol formulation but also by the choice of delivery device. Three types of pulmonary delivery devices are commercially available: jet or ultrasonic nebulizers, metered-dose inhalers (MDIs), and dry powder inhalers (DPIs). There are advantages to each based on the stability of the individual molecule and patient needs. Currently, DPIs and nebulizers are being used to disperse aerosol versions of high-molecular-weight proteins. Efforts are increasingly made to adapt biologics for DPs because they frequently are unstable in liquids and costly.

*Nebulizers* Aqueous inhalation formulations typically are administered with the aid of a nebulizer, which breaks up the formulation into inhalable aerosol droplets. Nebulizers deliver relatively large volumes of drug solutions and suspensions for inhalation and are frequently used for drugs that are stable or cannot be easily formulated into MDIs or DPIs. They are also used for drugs such as proteins in which the therapeutic dose is too large for delivery by either MDI or DPI [147].

Nebulizers have evolved from jet nebulizers to ultrasonic nebulizers and, more recently, vibrating mesh nebulizers, offering smaller sizes, improved aerosol performance, and delivery efficiency [148,149]. Jet nebulizers use the motion of compressed gases at high speed through an orifice to produce aerosol droplets in the respirable size range and are most commonly used clinically. Issues include

changes in drug concentration because of evaporation of liquid, which also decreases the liquid temperature. Ultrasonic nebulizers use high-energy ultrasonic waves generated from a piezoelectric crystal to convert liquid into a spray. Ultrasonic nebulizers, however, may be less suited to the delivery of labile biological molecules that may be damaged by direct exposure to the ultrasonic energy during nebulization. Vibrating mesh nebulizers have been introduced in recent years to address the problems associated with their ultrasound and jet predecessors. These nebulizers work by using a piezoelectric crystal with a lower frequency to vibrate a plate or mesh in contact with the liquid, thus avoiding the direct transfer of ultrasound energy to the solution. This approach typically generates a low-viscosity aerosol with small size particles that deliver the drug to the lung in a highly efficient manner, resulting in a substantial decrease in the dose required. They have been successfully used to aerosolize peptides and proteins such as insulin as well as DNA fragments [150].

Nebulizers have been used for the pulmonary delivery of proteins in the past decade but have not proven 100% successful to date. The shear force generated during nebulization can lead to fragmentation of biomolecules [151]. Furthermore, during nebulization, only a small fraction of the energy available in the nebulizer is used to generate the new surface necessary in the production of droplets. The excess energy is converted to heat, causing the temperature of the liquid within the nebulizers to increase; this heat generation may cause chemical degradation of proteins [152]. The viscosity of the solution affects how readily it can be nebulized, so nebulizers may not be useful for dispersing viscous formulations. For instance, because of its viscosity, the concentration of DNA that can be readily nebulized is less than 5 μg/ml. Nebulization of solute and adhesion of DNA to plastic surfaces results in less than 10% of the DNA in a nebulizer cup's being emitted from the device [153,154].

*Metered-Dose Inhalers* In MDIs, the active substance is suspended or solubilized in a propellant formulation, which contains at least one liquefied gas in a pressurized canister that is sealed with a metering valve. Actuation through the metering valve causes rapid expansion of the propellant, generating aerosol droplets to deliver a metered dose of the medication, which is directed by a suitable adaptor or actuator for administration via oral or nasal inhalation [166].

The main concern in the administration of biologics with MDIs is the maintenance of the conformational stability of proteins in hydrofluoroalkane (HFA) propellants. Using Raman spectroscopy, Quinn et al. demonstrated the structural integrity of lysozyme was maintained in HFA propellants [155]. Kos Pharmaceuticals has successfully completed a 4-week phase IIa human clinical trial evaluating MDI-delivered inhaled insulin therapy for patients with type 2 diabetes. The results of this trial showed that inhaled insulin is comparable to the market-leading injectable insulin in controlling blood glucose levels while also reducing blood lipids such as low-density lipoprotein cholesterol and triglycerides. Kos's inhaled insulin had also demonstrated bioavailability of up to 23% in previous phase I trial [156]. Thus, MDI administration may be a route of interest for further development in the delivery of biologics.

*Dry Powder Inhalers* DPIs are aerosol systems in which drugs are inhaled as clouds of fine particles. The drug is either preloaded in an inhalation device or prefilled in the form of hard capsules or foil blister discs that are loaded into a device before use. DPIs present some advantages over nebulizers and MDIs, especially for the administration of peptide and protein therapeutics to the lungs because of improved stability of formulations as a result of the dry state [157–159].

Size control, which impacts drug distribution, can pose a significant challenge with dry powder formulations, given that the properties of powders can differ widely because of variances in chemical composition and morphology of individual particles. This is often further compounded by the hygroscopic nature of many powders, which tend to absorb moisture leading to agglomeration: Particle size can still be altered during transit in the humid atmosphere of the respiratory tract. As a result of strong interparticle forces, drug delivery by DPI is also highly dependent on respiratory flow rate, which varies greatly [160]. Much research has been done to address this problem. Beyond the use of surfactants to elevate the activation energy of agglomeration by adding electrostatic or steric barriers between particles [161,162], drug nanoparticles are also designed to reduce the interparticle attraction forces and to improve the DPI performance by blending with carrier materials, forming Trojan particles or controlled nanoparticle agglomerates [160,163].

Despite these problems, many proteins have been successfully formulated as DPIs. Chan et al. prepared powders of recombinant human deoxyribonuclease (rhDNase) alone and with sodium chloride (NaCl) by spray drying where they observed better dispersal properties with the later [164]. Lucas et al. evaluated the use of carrier-based dry powder aerosols for pulmonary delivery of bovine serum albumin (BSA) and saw improved inhalation performance with mixtures prepared using BSA–maltodextrin and lactose [165]. It was concluded that fine-particle excipients can be used to improve the performance of carrier-based protein dry powder aerosols.

***22.3.1.3 Intranasal Delivery of Biologics*** Intranasal delivery has grown increasingly attractive as a noninvasive delivery route for a diverse range of therapeutic agents for local, systemic, and central nervous system (CNS) action. Targets for intranasal administration include compounds such as peptides, proteins, and polar drugs with low stability in GI fluids, poor intestinal absorption, or extensive hepatic first-pass elimination [166]. Nasal delivery also provides a way to bypass the BBB for direct drug delivery to the CNS in addition to being considered for the administration of vaccines [166,167]. Intranasal administration also offers several practical advantages both from the perspective of patients (painlessness and ease of delivery) and pharmaceutical companies (sterilization of nasal preparations is not required) [168,169].

However, some limitations should be considered when formulating biologics for intranasal delivery. First, in addition to the physicochemical properties of drugs and characteristics of their formulations, a variety of physiologic and pathological conditions related to the nasal mucosa may compromise the extent of nasal drug absorption and therapy efficacy [166,168,170]. Second the low volume of the nasal cavity restricts the volume administered to about 100 to 150 μl [170]. Nevertheless, strategies to

**TABLE 22.2 Representative Biologic Therapies Administered by Intranasal Route**

| Drug Name | Biological Agent | Indications |
|---|---|---|
| | *A. Marketed* | |
| Miacalcin | Calcitonin | Postmenopausal osteoporosis |
| Desmospray | Desmopressin | Diabetes insipidus |
| Suprecur | Buserelin | Prostate cancer |
| Syntocinon | Oxytocin | Labor induction, lactation stimulation |
| Synarel | Nafarelin | Endometriosis |
| Flu Mist | Live attenuated influenza virus | Flu prevention |
| | *B. In Clinical Trials* | |
| | Human growth hormone | Growth hormone deficiency |
| | Insulin | Diabetes mellitus |
| | Adenovirus-vectored influenza | Flu prevention |
| | Proteosome-influenza | Flu prevention |
| | Group B meningococcal native outer membrane vesicle | Meningococcal meningitis prevention |
| | Live-attenuated respiratory syncytial virus | Respiratory disease prevention |
| | Live attenuated parainfluenza 3 virus | Respiratory disease prevention |

overcome these drawbacks have led to successfully marketed intranasal biologic drugs and vaccines and others in clinical trials as outlined in Table 22.2 [116,171,172].

*Systemic Delivery* Intranasal administration is an effective alternative route to systematically deliver drugs that are administered conventionally via oral or parenteral routes. Furthermore, relative to multiple parenteral applications, repetitive intranasal administration is more practical for chronic applications, especially when oral delivery results in insufficient drug bioavailability, which is often the case for biologics such as peptides and protein with poor oral absorption. However, for currently marketed intranasal peptide and protein products, the formulations generally do not include permeation enhancers, resulting in relatively low bioavailability because of their relatively high molecular weight [181]. The bioavailability of these macromolecules can be improved by using permeation enhancers.

*Nasal Vaccine Delivery* Most pathogens invade the body through mucosal surfaces. Among other mucosal sites, nasal delivery is especially attractive for immunization because the nasal epithelium is characterized by reasonable permeability, low enzymatic activity, and the presence of a significant number of immunocompetent cells. Relative to parenteral routes, nasal vaccination is also beneficial because it is able to induce both mucosal (sIgA) and systemic (IgG) immune responses. In the upper airways, systemic and local immunologic responses are mainly mediated by the nasal-associated lymphoid tissue (NALT, also known as Waldeyer's ring in humans) situated

underneath the nasal epithelium, which is composed of clusters of dendritic cells, T cells, and B cells [173]. Notably, after nasal immunization, secretory immunoglobulin A can also be detected in other mucosal secretions (e.g., vaginal mucosa), which may be important against pathogens transmitted through other mucosal sites, such as human immunodeficiency virus [174] and hepatitis B virus [175]. Despite these encouraging characteristics, intranasal delivery of free antigens alone may not elicit an immune response in the absence of an effective adjuvant. Effective nasal immunization thus requires an effective antigen or a potent mucosal adjuvant or carrier. To enhance potency, immunostimulants such as TLR ligands, toxin-based adjuvants, or cytokines can be co-delivered [173]. In addition, encapsulation of antigens into bioadhesive particles can further improve nasal vaccine administration. These antigen-loaded particles can also incorporate targeting ligands, adjuvants, or endosomal escape mediators to help generate and enhance immunity [176,197].

*Central Nervous System Delivery* Improved delivery to the brain via the nasal route has been reported for therapeutic hormones, peptides, and proteins [177–181]. However, it should be noted that there are also cases in which evidence was found lacking for preferential delivery to the brain via intranasal delivery [182–184]. Although the clinical potential of this delivery route is still the subject of controversial debates, there is considerable interest in exploring the route to deliver peptide and protein for treatment of common intracerebral diseases such as Alzheimer's disease [185] and obesity [186].

*Strategies to Increase Nasal Drug Absorption* The bioavailability of nasally administered drugs is generally limited by low drug solubility, rapid enzymatic degradation in the nasal cavity, poor membrane penetration, and rapid mucociliary clearance. As such, various strategies have been devised to overcome these limitations, including the use of prodrugs, enzymatic inhibitors, absorption enhancers, development of mucoadhesive delivery systems, and new pharmaceutical formulations.

Prodrugs can be used to modify a compound's lipophilicity, improving solubility and ultimately increasing its transport across a biological membrane [209]. In addition, the potential of prodrugs to protect peptides from nasal enzymatic degradation has been suggested for increasing the bioavailability of intranasally administered peptides and proteins [168,170]. Enzymatic inhibition can also be achieved using certain absorption enhancers. For example, disodium ethylenediaminetetraacetate has been shown to reduce enzymatic degradation of a β sheet-breaker peptide used for the treatment of Alzheimer's disease [187]. Strategies to modify the cargo molecule to confer resistance against enzymatic degradation have also been attempted to increase the nasal bioavailability of calcitonins [188,189]. Such strategies, however, would warrant toxicologic studies to evaluate the safety of the modified molecules.

Mucociliary clearance is a major limiting factor for nasal delivery because it reduces the time available for drug absorption. To improve drug absorption by prolonging the contact time between drug and nasal mucosa, a range of mucoadhesive drug delivery systems based on materials such as chitosan and alginate have been developed in preclinical studies, some of which can also open epithelial tight

junctions and inhibit proteolytic enzymes [190]. Nasal formulations containing liposomes, microspheres, and nanoparticles have also been developed extensively for intranasal drug delivery. These systems can contain, besides the drug, enzymatic inhibitors, nasal absorption enhancers, or mucoadhesive polymers, to improve the stability, membrane permeation, and retention time in the nasal cavity. Liposomes were found to enhance nasal absorption of peptides such as insulin and calcitonin by increasing their membrane penetration [191,192]. Wang et al. [193] have investigated aminated gelatin microspheres as a nasal drug delivery system for insulin and observed a significant hypoglycemic effect on intranasal administration as a dry powder to rats. Studies have also suggested that nanoparticle systems may be ideally suited for the delivery of nasal vaccines [194–196].

### 22.3.2 Transcutaneous Delivery of Biologics

A wide range of biologics such as protein-, peptide-, and DNA-based drugs and vaccines are poorly suited to oral delivery because of their high degree of first-pass metabolism in the liver. The transdermal route has the potential to overcome this limitation and provide an efficient way of administering such molecules. The skin represents the largest and most easily accessible organ of the human body and provides a convenient and patient-friendly interface for drug administration. In addition, the transcutaneous route offers promise for achieving sustained and controlled drug delivery. Transcutaneous vaccine delivery could improve immune responses by targeting delivery to immunogenic Langerhans cells in the skin and has attracted increasing interest.

Advances in transdermal delivery systems have been categorized into three generations of development [197]. First-generation transdermal delivery systems have seen a steady increase in clinical use for delivery of drugs that are permissible to the skin via passive diffusion. Second-generation delivery systems using chemical enhancers, noncavitational ultrasonography, and iontophoresis to enhance permeability and drive transcutaneous transport have also led to clinical products. Third-generation delivery systems using microneedles, thermal ablation, microdermabrasion, electroporation, and cavitational ultrasonography to circumvent the barrier presented by the stratum corneum are being actively pursued. Examples of transcutaneous delivery of biologics in clinical development include human growth hormone via thermal ablation (phase 1, TransPharma Medical Teva Pharmaceuticals, Lod, Israel/Petach Tikva, Israel), insulin using thermal ablation (phase 1, Altea Therapeutics, Atlanta, GA, USA) or vesicular carriers (phase 2, Phosphagenics, Melbourne, Australia), parathyroid hormone (PTH) with microneedles (phase 2, Zosano, Fremont, CA, USA), as well as a microneedle-based influenza vaccine (completion of phase 3 clinical trials and submission for registration in Europe by Sanofi Pasteur, Paris, France, and Becton Dickinson, Franklin Lakes, NJ, USA) [197].

***22.3.2.1 Barriers to Transcutaneous Drug Delivery*** Despite representing an attractive target for drug delivery, the functional properties of the skin that make

it an excellent barrier also serve to limit the access of drugs and vaccines into and across the epidermis. The skin is especially adept at keeping out high-molecular-weight compounds, rendering it unrealistic for therapeutic levels of drugs and vaccines to be attained by mere passive absorption. Therefore, an understanding of the barrier properties of the skin will complement the design of strategies to overcome this barrier safely and reversibly in order for transcutaneous delivery to achieve its anticipated potential for the administration of biological products. The skin is composed of several anatomically distinct layers. The stratum corneum, composed of lipids and dead epidermal cells (corneocytes), is the outermost superficial layer (10–20 μm thick) and provides the principal barrier to the percutaneous absorption of exogenous compounds as well as to water loss [198]. The corneocytes are flat anucleated squamous cells composed mainly of cross-linked keratin surrounded by a cell envelope composed of cross-linked proteins and a covalently bound lipid envelope [198]. The stratum corneum lipids are composed primarily of ceramides, cholesterol, and fatty acids that are assembled into multilamellar bilayers, resulting in a very densely packed matrix that serves the primary barrier function [200]. Underlying the stratum corneum is the viable epidermis (50–100 μm thick), which is responsible for regeneration of the stratum corneum. The dermis (1–2 mm thick) lies directly adjacent to the epidermis and provides mechanical support for the skin. A rich capillary bed for systemic drug absorption also sits just below the dermal–epidermal junction.

Molecules may cross the stratum corneum via an intercellular or a transcellular route. Moreover, transport across the stratum corneum and even the epidermis can also occur through shunt pathways such as hair follicles, sebaceous glands, and sweat ducts. Transcutaneous transport is largely controlled by stratum corneum lipid bilayers and is highly anisotropic and size dependent [201]. Because of the impermeable character of the cornified layer, the major route of penetration resides in the tortuous pathway around the corneocytes, where hydrophilic molecules travel through the lipid head group regions and lipophilic molecules travel through the lipid tails regions. Transport between bilayers can also occur at bilayer–bilayer interfaces or other sites of structural disorganization. Cutaneous penetration of hydrophilic substances is limited by the convoluted intercellular space as well as the hydrophobicity of the lipidic components present in the stratum corneum, which imposes structural and solubility constraints for solution and diffusion within stratum corneum lipid bilayers. For systemic delivery, after its transport through the stratum corneum, the drug still need to partition into the aqueous epidermis and finally into the capillary network of the dermis. Hence, a transcutaneous drug candidate must possess both lipophilic and hydrophilic tissue spaces. Molecules too hydrophobic will fail to partition into the SC, and molecules too lipophilic will encounter difficulties in moving through the subsequent aqueous layer in the epidermis [202].

***22.3.2.2 Transdermal Patches*** Despite the significant barrier properties of skin, it was found that small lipophilic drugs exhibit significant permeability, which led to the development of transdermal patches in the 1970s. Today there exist a number of transdermal patches for drugs such as clonidine, fentanyl, lidocaine, nicotine,

nitroglycerin, oestradiol, oxybutynin, scopolamine, and testosterone, as well as combination patches for contraception and hormone replacement [203]. For the delivery of biologics, patches are currently in development for transcutaneous immunization against influenza in elderly and travelers' diarrhea (phase II clinical trial, Iomai Corporation) and various other diseases (EasyVax, Vaxin, Inc.) [204,205].

***22.3.2.3 Penetration Enhancers*** To circumvent the barrier imposed by the skin to transcutaneous delivery of biologics, a wide range of penetration enhancers have been investigated for enhancing transcutaneous penetration. Chemical penetration enhancers can increase skin permeability by various mechanisms, including enhancing solubility, increasing partitioning into the stratum corneum, opening dense protein structure, fluidizing the crystalline structure of stratum corneum, and causing dissolution of stratum corneum lipids [206,207]. Many effective chemical enhancers disrupt the highly ordered bilayer structures of the intracellular lipids found in stratum corneum by inserting amphiphilic molecules into these bilayers to disorganize molecular packing or by extracting lipids using solvents and surfactants to create nanosized lipid-packing defects. Chemical enhancers help increase skin permeability and provide an additional driving force for transport across skin by increasing drug partitioning into the skin, thereby increasing the concentration gradient in the skin driving diffusion of molecules farther into the skin. However, increased permeation enhancement typically correlates with increased skin irritation, and the difficulty of localizing their effects to the stratum corneum to avoid irritation or toxicity to living cells in the deeper skin has largely constrained their application.

Decades of extensive research has yielded several different classes of penetration enhancers, including surfactants, fatty acids and esters, terpenes, and solvents [208]. Ethanol and propylene glycol are often used as solubilizers and can extract lipids in the stratum corneum to facilitate permeation. Oleic acid works by accumulating in the stratum corneum and inducing fluidization and phase separation within the lipid membrane [209]. Surfactants are potent transport enhancers by virtue of their ability to solubilize or extract lipids and denature keratin. Besides off-the-shelf compounds, attempts have also been made to design and synthesize novel chemical penetration enhancers such as lauracapram (Azone) [210,211].

Combinations of chemical enhancers offer further new opportunities in transcutaneous formulations to increase both safety and efficacy. In addition to increasing transcutaneous transport in a synergistic manner, combination of enhancers can also reduce the required dose of each enhancer, making it possible to achieve increased skin permeability in the stratum corneum while avoiding irritation in deeper tissue. However, the rational design of enhancer combinations is limited by the lack of mechanistic information on the interactions between individual chemical enhancers and the stratum corneum. This bottleneck has been addressed with high-throughput methods for screening transdermal formulations to discover previously unknown enhancer mixtures [212]. Dramatically increased enhancement with low skin irritation potential was found, for example, for a combination of sodium laureth sulfate (an anionic surfactant) and phenyl piperazine (a compound with aromatic nitrogen) at concentrations of 0.35% and 0.15% by weight, respectively, in a 1:1 mixture of ethanol

and phosphate-buffered saline. *In vitro* screening results were validated with *in vivo* delivery of a peptide (leuprolide acetate) to hairless rats. These results suggest that combinations of chemical enhancers may succeed for delivery of macromolecules where individual enhancers have generally failed. Such an approach is currently being pursued in industry.

More recently, peptides have been examined as enhancers of skin permeability. Although lipophilic derivatives of peptides have been reported to exhibit increased skin permeability [213], co-delivery of certain peptide sequences has also been shown to enhance protein and siRNA delivery [214,215]. In one such study, phage display was used to screen a library of peptides, which yielded a synthetic peptide sequence that, when coadministered with the cargo, increased transdermal delivery of insulin in diabetic rats [216]. Further analysis suggested that the main pathway of transport was via hair follicles. Another study has shown that a natural pore-forming peptide, magainin, increases skin permeability by targeting bilayer disruption in stratum corneum lipids but not in deeper tissue [217]. The magainin peptide was only effective when applied in combination with a surfactant chemical enhancer, which served a dual purpose of increasing skin permeability to the drug as well as magainin. In addition to peptides, trypsin has also been shown to enhance the transcutaneous delivery of insulin [218].

Colloidal carriers such as emulsions, micelles, dendrimers, liposomes, and deformable vesicles have also been used as chemical penetration enhancers with supramolecular structure that can enhance skin permeability, drug solubilization in the formulation, and drug partitioning into the skin [219–221]. Their supramolecular size generally precludes penetration into deeper tissues and thereby localizes the effects to the stratum corneum. For example, the incorporation of surfactants in these carriers can fluidize the lipid layers in the skin and allow for the formation of depots in the stratum corneum for prolonged effect. In the case of deformable vesicles, their flexible nature has been suggested to enable them to squeeze through pores in the skin. A highly deformable liposome formulation is currently in clinical trials for insulin delivery [197]. Recently, the percutaneous delivery of protein drugs mediated by co-administrating with gold nanoparticles was also reported [222]. This method was then used toward a transcutaneous vaccine study in which immune response in mice was elicited by topically co-administrating antigens with gold nanoparticles.

***22.3.2.4 Ballistic Intradermal Delivery*** After a rise and fall in popularity in the 1950s, high-velocity jet injectors are receiving renewed interest for controlled, needle-free injection of drug solutions or particles across the skin and into deeper tissue. Current commercial jet injectors use compressed gas to generate high-velocity micron-sized ballistic droplets or particles that impinge on and penetrate the skin barrier. Solid particles such as DNA-coated gold particle vaccines [223] or droplets such as insulin solution [224] can be delivered into the epidermis to target particular cells or into deeper dermis for subsequent absorption into blood capillaries. Ballistic delivery using jet injectors combines the advantages of transdermal and parenteral drug delivery methods, allowing direct access into the skin tissue while alleviating the issues of patient phobia, injury, cross-contamination, and waste disposal associated with the use of needles.

Jet injection is the first and only approved needle-free protein delivery system for transcutaneous delivery to date. Liquid jet injectors have been used to deliver several macromolecules, including vaccines [225], insulin [224], growth hormones [226], erythropoietin [227], and INF [228]. Insulin delivery by jet injections has been shown to be as effective as needles in the control of blood glucose levels [229,230]. They represent the first commercial needle-free method for insulin delivery (Antares Vision, Choice, Antares, previously Mediject) and the only such method until the recent FDA approval of inhaled insulin. The first growth hormone injector was also introduced by the same company in Europe in 1994. Increased and earlier maximum blood concentrations were found after jet injection of recombinant human growth hormone compared with needle injections [231]. Other macromolecules, such as low-molecular-weight heparin [232] for treatment of deep vein thrombosis and botulinum toxin [233] for the treatment of plantar hyperhidrosis have also been delivered by jet injections in humans. Liquid jet injections were successfully adopted for immunizing humans using vaccines against a large number of diseases, including measles, smallpox, cholera, hepatitis B, influenza, and polio [234]. However, their usage decreased drastically after evidence of contamination originating from blood and splashed liquid in the multiuse injectors was discovered [235]. This can be avoided by the development of single-use, disposable jet injectors. Currently, jet injections of liquids are being pursued for protein, DNA, and live-vectored vaccines by BioJect [236].

*DNA Vaccine Delivery via Gold Particles* The skin can function as an effective induction site for both mucosal and systemic immune responses. Indeed, the administration of vaccines to the skin by ballistic delivery has been shown to induce immune responses in both blood and mucosal tissues [237,238]. DNA can be delivered by jet injectors in a solid form via precipitation onto gold particles (precipitation occurs before delivery), which confers greater stability of DNA relative to liquid formulations and can, in some cases, even allows for the stable storage of DNA at room temperature and deployment without a cold chain. The use of jet injectors to propel particles coated with antigen-encoding DNA plasmids into skin is of great interest for DNA-based vaccination. This is particularly motivated by the immune competency of skin tissue, attributed to the presence of antigen-presenting cells (APCs) such as Langerhans cells. Ballistic delivery can be used to directly transfect APCs, which can then present the translated antigen directly to the immune system. Antigen can also be expressed and secreted by transfected keratinocytes and taken up by resident APCs through the exogenous antigen presentation pathway. Numerous past studies demonstrate that such particle-mediated DNA vaccination induces protective immune responses both in preclinical studies in animal models and clinical trials [239–244]. More recently, a phase I clinical trial was conducted to evaluate a monovalent influenza DNA vaccine containing the HA gene from A/Panama/2007 /99 delivered by particle-mediated epidermal delivery [245]. Treatment-related reactions in patients were mild to moderate and included skin reactions at the vaccine site. These results provided a preliminary indication of the safety and immunogenicity of a prototype epidermal DNA vaccine for influenza. Development of representative annual and avian influenza vaccines is now underway. Other vaccines in clinical trials include biolistic microparticle injection

(PowderMed) for HIV (phase I) and herpes simplex type 2 (HSV2, phase I) as well as a cancer vaccine (phase I) [236].

***22.3.2.5 Ultrasound-Mediated Transdermal Delivery*** Ultrasonic waves at various frequencies in the range of 20 kHz to 16 MHz have been used to enhance skin permeability and facilitate transcutaneous drug delivery, commonly known as sonophoresis. The use of a therapeutic and high frequencies of ultrasound ($\geq$0.7 MHz) for sonophoresis dates back to as early as the 1950s, but low-frequency sonophoresis (20–100 kHz) has only been explored more recently during the past two decades since greater penetration enhancement effect was observed at low frequencies. Although they both use ultrasound to increase the skin penetration of permeants, the associated mechanisms by which they mediate transcutaneous delivery are distinctly different.

Ultrasound-mediated enhancement of transcutaneous transport is thought to be mediated by a combination of thermal and mechanical alterations within the skin tissue. An increase in temperature due to absorption of ultrasound by the skin can increase skin permeability by (1) increasing the kinetic energy and diffusivity of drug compounds, (2) dilating points of entry of the skin (e.g., hair follicles and sweat glands), (3) facilitating drug absorption, and (4) enhancing circulation of blood in the treated area (in *in vivo* experiments) [246,247]. However, the contribution caused by thermal effects is likely to be insignificant in real clinical settings because any thermal effects will have to be minimized to avoid potential damage to the skin tissue. Beyond the elevation of skin temperature, ultrasound is also reported to induce cavitational effects caused by the formation, oscillation, and collapse of gaseous cavities in or near the skin [248]. Specifically, collapse of cavitation bubbles near the stratum corneum is hypothesized to disrupt its structure because of cavitation-generated shock waves or microjets and was determined to be primarily responsible for the efficacy of sonophoresis. In low-frequency sonophoresis (LFS), the primary mechanism of enhancement has been shown to be transient acoustic cavitation above the skin surface, usually occurring in the coupling medium (the liquid present between the ultrasound transducer and the skin). In high-frequency sonophoresis (HFS), cavitation within the skin, either in skin appendages or at locations near the keratinocytes of the stratum corneum, is the main mechanism of skin permeability enhancement [248], although other mechanisms may play more minor roles. Although lower frequencies induce greater enhancements, the transport at low frequencies was found to be localized to discrete regions termed localized *transport pathways* [249] because of the heterogeneous occurrence of cavitation.

In general, HFS is more effective for enhancing the penetration of low-molecular-weight compounds, such as nonsteroidal antiinflammatory drugs and topical steroids into the skin. HFS is also more readily tested in human trials because of its long track record of safety and use in patients for physical therapy. Conversely, LFS has displayed a greater capacity to deliver high-molecular-weight compounds, including proteins, hormones, vaccines, and nanoparticles. However, clinical uses of LFS are relatively rare, with only a few current applications in patients because of the lack of historical precedence of safety. Examples of biological molecules delivered transcutaneously by the use of LFS are summarized in Table 22.3 [250].

**TABLE 22.3 Representative Biological Molecules Delivered Transcutaneously with Low-Frequency Sonophoresis**

| Molecule | Skin model | |
|---|---|---|
| | *In Vitro* | *In Vivo* |
| Bovine serum albumin | Rat full-thickness skin | Rat |
| Erythropoietin | Human heat-stripped skin | Rat |
| Heparin | Pig full-thickness skin | Mouse, rat, rabbit, pig |
| Insulin | Human full-thickness skin, human heat-stripped skin | Mouse |
| Interferon-γ | | |
| Oligonucleotides | Human full-thickness skin | |
| Tetanus toxoid | Pig full-thickness skin | |

Clinical studies have also been performed on transdermal insulin delivery. Glucose-clamp studies in type I diabetic volunteers showed that therapeutic insulin doses can be delivered across ultrasonically permeabilized skin [251]. Ultrasound skin permeation was accomplished using the SonoPrep skin permeation device (Sontra Medical Corp.) which delivers ultrasound at 55 kHz to the skin through an aqueous medium that consists of a phosphate buffered saline solution and 1% sodium lauryl sulfate.

***22.3.2.6 Microneedles*** In the past decade, microneedles have been explored as a way to facilitate transcutaneous delivery of biologics by piercing through the stratum corneum barrier and directly creating transient delivery channels in the skin without stimulating underlying pain nerves. Microneedles can be hollow, similar to traditional hypodermic needles but fabricated on the micro scale, or solid structures coated with drugs [252,253]. They are generally several hundred microns in diameter and range from 100 to 1000 μm in length. Microneedles have been fabricated with various materials such as metals, silicon, polymers, and glass. Notably, microneedles can be manufactured at lower cost using biodegradable polymers that will dissolve in the skin over time, thus circumventing the issues of breakage in skin and the generation of biohazardous sharp medical waste [254–256]. Solid and hollow microneedles have been used as a pretreatment to create micron-sized holes in the skin and have been shown to increase the transport of a variety of small molecules, proteins, and nanoparticles [257]. Alternatively, drug formulations have been coated on or encapsulated within microneedles for rapid or controlled release of peptides, oligonucleotides, plasmid DNA, virus particles, and nanoparticles upon insertion in the skin. Arrays of hollow needles can also be fabricated to flow drugs into the skin using simple diffusion or a pump system and have been used to deliver insulin and vaccines by infusion [257].

Recently, Harvey et al. [258] reported on the pharmacokinetics and pharmacodynamics for microneedle-based intradermal delivery of several classes of approved protein drugs compared with standard SC administration in the swine model using hollow stainless steel microneedles. Delivery of proteins using microneedles resulted in faster systemic availability and increased maximal drug concentration over SC

delivery for all proteins tested. The same group has previously used microneedles to administer numerous vaccines in preclinical models [259–264], and these devices have also progressed through phase I to III clinical trials for influenza vaccines [265–268] with product licensure obtained in Europe [269].

Phase 1 and 2 clinical trials were also recently conducted by Zosano Pharma (formerly Macroflux, ALZA Corp) to evaluate microneedle-based delivery of PTH for treatment of osteoporosis, their lead product candidate in development [270]. The studies suggested that their microneedle patch system was able to deliver a consistent and therapeutically relevant PTH pharmacokinetic profile. Based on encouraging phase 2 safety and efficacy data, the program is advancing into a pivotal phase 3 clinical study. More recently, the same delivery system was also tested for the delivery of erythropoietin in rats (Fig. 22.2) [271]. Previous clinical and preclinical studies have

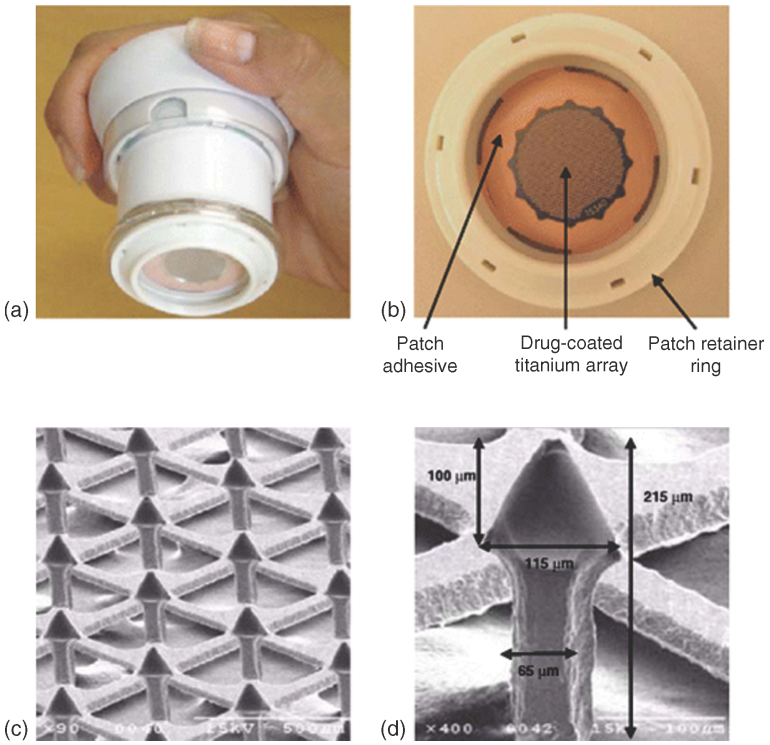


**FIGURE 22.2** Example of a transdermal microneedle patch delivery system: (a) ZP-patch and applicator (Zosano Pharma, Inc.). (b) Five-$cm^2$ adhesive patch with titanium microneedle array (2 $cm^2$) in applicator ring. (c) 90× magnification of erythropoietin (EPO)-coated microneedles, EPO coated at 3000 IU (30 μg)/3 $cm^2$ array. (d) Front view of an individual EPO-coated microneedle (400× magnification). The length of each microneedle is 215 μm; the thickness is 25 μm; and the length and the width of the microneedle arrowhead are 100 and 115 μm, respectively, with a tip angle of 60 degrees. The microneedle shaft is 115 μm long and 65 μm wide [271].

also demonstrated the capability of their drug-coated microneedle patch technology to deliver peptides, recombinant proteins, and vaccine antigen [272–275].

## 22.4 SUMMARY AND OUTLOOK

The steadily increasing fraction of newly approved drugs that are biologics provides a strong motivation for continued development of new and improved methods and routes of administration to obtain greater efficacy and ease of administration with reduced side effects. Traditional methods of administration via IV or SC injection remain the approved ones for the majority of approved biologics, but diverse preclinical and clinical efforts and recent successfully licensed products making use of alternative routes will likely continue to become more prevalent in the coming few years. Promising technologies include the use of recombinant hyaluronidases to facilitate pain-free large-volume SC injections, the development of diverse methods for transcutaneous drug delivery (e.g., ultrasound, jet injectors, and microneedles), and the successful development of aerosol formulations of protein drugs for pulmonary delivery.

Challenges still remain that limit the efficacy of biologics relative to small molecule drugs. Efficient oral delivery of peptide and protein drugs remains a holy grail of drug development, but general and robust strategies remain elusive. As discussed in this chapter, drug delivery materials such as polymeric and lipid nanoparticles or microparticles can be used to encapsulate and protect biologics from the degradative environment of the GI tract, but these particulate drug carriers are inefficiently absorbed across the gut epithelial lining and are rapidly cleared from the digestive tract, limiting uptake of the drug cargo into the tissues. Efforts to target encapsulated drug cargos to the gut epithelium (e.g., via ligands that bind to target receptors on the epithelium) are still in early preclinical stages, and the ability of such approaches to substantially improve oral delivery efficiencies of biologics remains to be proven. Likewise, pulmonary delivery of biologics has reached the stage of licensed drug approvals, but the efficiency of drug delivery by this route may still be improved, and the long-term safety of drug administration to the pulmonary tissues will only become clear with clinical experience over time. The failure of inhaled insulin to attract a viable patient population also serves as cautionary tale for the development of alternative routes of administration. New methods of drug administration will likely need to offer substantial therapeutic gains, ease of use, or lowered side effects to gain widespread acceptance. However, ongoing innovation in the fields of pharmaceutical sciences and drug development suggest that advanced methods for biologics delivery will continue to evolve and enhance the well-being of patients while improving patient compliance.

## ACKNOWLEDGMENTS

This work was supported in part by the Ragon Institute of MIT, MGH, and Harvard University; the Gates Foundation, the NIH (AI095109 and AI091693); and the

Department of Defense (contracts W911NF-07-D-0004 and W911NF-07-D-0004). DJI is an investigator of the Howard Hughes Medical Institute.

## REFERENCES

1. Silin DS, Lyubomska OV, Jirathitikal V, Bourinbaiar AS. Oral vaccination: where we are? Expert Opin Drug Deliv 2007;4(4):323–340.
2. Todd PA, Brogden RN. Muromonab Cd3. A review of its pharmacology and therapeutic potential. Drugs 1989;37(6):871–899.
3. Forero A, Lobuglio AF. History of antibody therapy for non-Hodgkin's lymphoma. Semin Oncol 2003;30(6 Suppl 17):1–5.
4. Ravandi F, O'Brien S. Alemtuzumab. Expert Rev Anticancer Ther 2005;5(1):39–51.
5. Reagan JL, Castillo JJ. Ofatumumab for newly diagnosed and relapsed/refractory chronic lymphocytic leukemia. Expert Rev Anticancer Ther 2011;11(2):151–160.
6. Hudis CA. Trastuzumab—mechanism of action and use in clinical practice. N Engl J Med 2007;357(1):39–51.
7. Harandi A, Zaidi AS, Stocker AM, Laber DA. Clinical efficacy and toxicity of anti-EGFR therapy in common cancers. J Oncol 2009;2009: 567486.
8. National Cancer Institute. Trastuzumab. Available from: http://www.cancer.gov/cancer-topics/druginfo/trastuzumab.
9. National Cancer Institute. Rituximab. Available from: http://www.cancer.gov/cancer-topics/druginfo/rituximab.
10. National Cancer Institute. Aldesleukin. Available from: http://www.cancer.gov/cancer-topics/druginfo/aldesleukin.
11. Ladenstein R, Potschger U, Siabalis D, Garaventa A, Bergeron C, Lewis IJ, et al. Dose finding study for the use of subcutaneous recombinant interleukin-2 to augment natural killer cell numbers in an outpatient setting for stage 4 neuroblastoma after megatherapy and autologous stem-cell reinfusion. J Clin Oncol 2011;29(4):441–448.
12. Food and Drug Administration. Vaccines, Blood, and Biologics—Approved Products. Available from: http://www.fda.gov/BiologicsBloodVaccines/CellularGeneTherapyProducts/ApprovedProducts/default.htm.
13. van Horssen R, Ten Hagen TL, Eggermont AM. TNF-alpha in cancer treatment: molecular insights, antitumor effects, and clinical utility. Oncologist 2006;11(4):397–408.
14. Grunhagen DJ, de Wilt JH, ten Hagen TL, Eggermont AM. Technology insight: utility of TNF-alpha-based isolated limb perfusion to avoid amputation of irresectable tumors of the extremities. Nat Clin Pract Oncol 2006;3(2):94–103.
15. Thurber GM, Schmidt MM, Wittrup KD. Antibody tumor penetration: transport opposed by systemic and antigen-mediated clearance. Adv Drug Deliv Rev 2008;60(12):1421–1434.
16. Konrad MW, Hemstreet G, Hersh EM, Mansell PW, Mertelsmann R, Kolitz JE, Bradley EC. Pharmacokinetics of recombinant interleukin 2 in humans. Cancer Res 1990;50(7):2009–2017.
17. Chapman AP, Antoniw P, Spitali M, West S, Stephens S, King DJ. Therapeutic antibody fragments with prolonged in vivo half-lives. Nat Biotechnol 1999;17(8):780–783.

18. Hu X, Olivier K, Polack E, Crossman M, Zokowski K, Gronke RS, et al. In vivo pharmacology and toxicology evaluation of polyethylene glycol-conjugated interferon beta-1a. J Pharmacol Exp Ther 2011;338(3):984–996.
19. Fishburn CS. The pharmacology of PEGylation: balancing PD with PK to generate novel therapeutics. J Pharm Sci 2008;97(10):4167–4183.
20. Bailon P, Won CY. PEG-modified biopharmaceuticals. Expert Opin Drug Deliv 2009; 6(1):1–16.
21. Rivkin A. Certolizumab pegol for the management of Crohn's disease in adults. Clin Ther 2009;31(6):1158–1176.
22. Winter TA, Wright J, Ghosh S, Jahnsen J, Innes A, Round P. Intravenous CDP870, a PEGylated Fab′ fragment of a humanized antitumour necrosis factor antibody, in patients with moderate-to-severe Crohn's disease: an exploratory study. Aliment Pharmacol Ther 2004;20(11–12):1337–1346.
23. Boswell CA, Tesar DB, Mukhyala K, Theil FP, Fielder PJ, Khawli LA. Effects of charge on antibody tissue distribution and pharmacokinetics. Bioconjug Chem 2010;21 (12):2153–2163.
24. Yeung YA, Leabman MK, Marvin JS, Qiu J, Adams CW, Lien S, et al. Engineering human IgG1 affinity to human neonatal Fc receptor: impact of affinity improvement on pharmacokinetics in primates. J Immunol 2009;182(12):7663–7671.
25. Zalevsky J, Chamberlain AK, Horton HM, Karki S, Leung IW, Sproule TJ, et al. Enhanced antibody half-life improves in vivo activity. Nat Biotechnol 2010;28(2):157–159.
26. Grace MJ, Cutler DL, Bordens RW. Pegylated IFNs for chronic hepatitis C: an update. Expert Opin Drug Deliv 2005;2(2):219–226.
27. Ascierto PA, Kirkwood JM. Adjuvant therapy of melanoma with interferon: lessons of the past decade. J Transl Med 2008;6(1):62.
28. Paolicelli D, Direnzo V, Trojano M. Review of interferon beta-1b in the treatment of early and relapsing multiple sclerosis. Biologics 2009;3:369–376.
29. Lobo ED, Hansen RJ, Balthasar JP. Antibody pharmacokinetics and pharmacodynamics. J Pharm Sci 2004;93(11):2645–2668.
30. Kaiser C, Knight A, Nordstrom D, Pettersson T, Fransson J, Florin-Robertsson E, et al. Injection-site reactions upon Kineret (Anakinra) administration: experiences and explanations. Rheum Int 2012;32(2):295–299.
31. Genovese MC, Covarrubias A, Leon G, Mysler E, Keiserman M, Valente R, et al. Subcutaneous abatacept versus intravenous abatacept: a phase IIIb noninferiority study in patients with an inadequate response to methotrexate. Arthritis Rheum 2011;63 (10):2854–2864.
32. McCormack PL, Scott LJ. Interferon-beta-1b: a review of its use in relapsing-remitting and secondary progressive multiple sclerosis. CNS Drugs 2004;18(8):521–546.
33. Sanford M, Lyseng-Williamson KA. Subcutaneous recombinant interferon-1a (Rebif): a review of its use in the treatment of relapsing multiple sclerosis. Drugs 2011;71 (14):1865–1891.
34. Kirkwood JM, Strawderman MH, Ernstoff MS, Smith TJ, Borden EC, Blum RH. Interferon alfa-2b adjuvant therapy of high-risk resected cutaneous melanoma: the Eastern Cooperative Oncology Group Trial EST 1684. J Clin Oncol 1996;14(1):7–17.
35. Chiarion-Sileni V, Del Bianco P, Romanini A, Guida M, Paccagnella A, Dalla Palma M, et al. Tolerability of intensified intravenous interferon alfa-2b versus the ECOG 1684

schedule as adjuvant therapy for stage III melanoma: a randomized phase III Italian Melanoma Inter-Group Trial (IMI - Mel.A.) [ISRCTN75125874]. BMC Cancer 2006;6:44.

36. Ratner M. Roche plans for more convenient-to-use Herceptin and Rituxan. Nature Biotechnol 2010;28(4):298.
37. Bookbinder LH, Hofer A, Haller MF, Zepeda ML, Keller GA, Lim JE, et al. A recombinant human enzyme for enhanced interstitial transport of therapeutics. J Control Release 2006;114(2):230–241.
38. Frost GI. Recombinant human hyaluronidase (rHuPH20): an enabling platform for subcutaneous drug and fluid administration. Expert Opin Drug Deliv 2007;4(4):427–440.
39. Lugaresi A. Addressing the need for increased adherence to multiple sclerosis therapy: can delivery technology enhance patient motivation? Expert Opin Drug Deliv 2009; 6(9):995–1002.
40. Brearley C, Priestley A, Leighton-Scott J, Christen M. Pharmacokinetics of recombinant human growth hormone administered by cool.click 2, a new needle-free device, compared with subcutaneous administration using a conventional syringe and needle. BMC Clin Pharmacol 2007;7:10.
41. Verdun di Cantogno E, Russell S, Snow T. Understanding and meeting injection device needs in multiple sclerosis: a survey of patient attitudes and practices. Patient Prefer Adherence 2011;5:173–180.
42. Desrosiers P, O'Brien F, Blethen S. Patient outcomes in the GHMonitor: the effect of delivery device on compliance and growth. Pediatr Endocrinol Rev 2005;2(Suppl 3):327–331.
43. Cox LS, Larenas Linnemann D, Nolte H, Weldon D, Finegold I, Nelson HS. Sublingual immunotherapy: a comprehensive review. J Allergy Clin Immunol 2006;117(5):1021–1035.
44. Greenberger PA, Ballow M, Casale TB, Platts-Mills TA, Sampson HA. Sublingual immunotherapy and subcutaneous immunotherapy: issues in the United States. J Allergy Clin Immunol 2007;120(6):1466–1468.
45. van Wijngaarden P, Qureshi SH. Inhibitors of vascular endothelial growth factor (VEGF) in the management of neovascular age-related macular degeneration: a review of current practice. Clin Exp Optom 2008;91(5):427–437.
46. Moshfeghi AA, Rosenfeld PJ, Puliafito CA, Michels S, Marcus EN, Lenchus JD, et al. Systemic bevacizumab (Avastin) therapy for neovascular age-related macular degeneration: twenty-four-week results of an uncontrolled open-label clinical study. Ophthalmology 2006;113(11):2002, e2001–2012.
47. Bokemeyer C. Catumaxomab-trifunctional anti-EPCAM antibody used to treat malignant ascites. Expert Opin Biol Ther 2010;10(8):1259–1269.
48. Okamoto S, Ito K, Sasano H, Moriya T, Niikura H, Terada Y, et al. Ber-Ep4 and anti-calretinin antibodies: a useful combination for differential diagnosis of various histological types of ovarian cancer cells and mesothelial cells. Tohoku J Exp Med 2005;206 (1):31–40.
49. Passebosc-Faure K, Li G, Lambert C, Cottier M, Gentil-Perret A, Fournel P, et al. Evaluation of a panel of molecular markers for the diagnosis of malignant serous effusions. Clin Cancer Res 2005;11(19 Pt 1):6862–6867.
50. Belau A, Pfisterer J, Wimberger P, Kurzeder C, Du Bois A, Sehouli J, et al. Randomized, multicenter, two-dose level, open-label, phase IIA study with the intraperitoneally infused trifunctional bispecific antibody catumaxomab (anti-EpCAM × anti-CD3) to

select the better dose level in platinum refractory epithelial ovarian cancer patients. J Clin Oncol 2007;25(18 Suppl):5556.

51. Ruf P, Kluge M, Jager M, Burges A, Volovat C, Heiss MM, et al. Pharmacokinetics, immunogenicity and bioactivity of the therapeutic antibody catumaxomab intraperitoneally administered to cancer patients. Br J Clin Pharmacol 2010;69(6):617–625.
52. Sebastian M, Passlick B, Friccius-Quecke H, Jäger M, Lindhofer H, Kanniess F, et al. Treatment of non-small cell lung cancer patients with the trifunctional monoclonal antibody catumaxomab (anti-EpCAM × anti-CD3):a phase I study. Cancer Immunol Immunother 2007;56(10):1637–1644.
53. Raza SM, Pradilla G, Legnani FG, Thai QA, Olivi A, Weingart JD, et al. Local delivery of antineoplastic agents by controlled-release polymers for the treatment of malignant brain tumours. Expert Opin Biol Ther 2005;5(4):477–494.
54. Low SC, Nunes SL, Bitonti AJ, Dumont JA. Oral and pulmonary delivery of Fsh-Fc fusion proteins via neonatal Fc receptor-mediated transcytosis. Hum Reprod 2005;20 (7):1805–1813.
55. White E, Bienemann A, Taylor H, Hopkins K, Cameron A, Gill S. A phase I trial of carboplatin administered by convection-enhanced delivery to patients with recurrent/ progressive glioblastoma multiforme. Contemp Clin Trials 2012;33(2):320–331.
56. Sampson JH, Akabani G, Archer GE, Berger MS, Coleman RE, Friedman AH, et al. Intracerebral infusion of an EGFR-targeted toxin in recurrent malignant brain tumors. Neuro-oncology 2008;10(3):320–329.
57. Oh S, Tsai AK, Ohlfest JR, Panoskaltsis-Mortari A, Vallera DA. Evaluation of a bispecific biological drug designed to simultaneously target glioblastoma and its neovasculature in the brain. J Neurosurg 2011;114(6):1662–1671.
58. Caron JP, Fernandes JC, Martel-Pelletier J, Tardif G, Mineau F, Geng C, et al. Chondroprotective effect of intraarticular injections of interleukin-1 receptor antagonist in experimental osteoarthritis. Suppression of collagenase-1 expression. Arthritis Rheum 1996;39(9):1535–1544.
59. Zhang B, Hu J, Man C, Zhu S. Effect of intra-articular administration of interleukin 1 receptor antagonist on cartilage repair in temporomandibular joint. J Craniofac Surg 2011;22(2):711–714.
60. Chevalier X, Goupille P, Beaulieu AD, Burch FX, Bensen WG, Conrozier T, et al. Intraarticular injection of Anakinra in osteoarthritis of the knee: a multicenter, randomized, double-blind, placebo-controlled study. Arthritis Rheum 2009;61(3):344–352.
61. Brown C, Toth A, Magnussen R. Clinical benefits of intra-articular Anakinra for persistent knee effusion. J Knee Surg 2011;24(1):61–65.
62. Kraus VB, Birmingham J, Stabler TV, Feng S, Taylor DC, Moorman CT 3rd, et al. Effects of intraarticular IL1-Ra for acute anterior cruciate ligament knee injury: a randomized controlled pilot trial (NCT00332254). Osteoarthritis Cartilage 2012;20(4):271–278.
63. Anisimov SV, Morizane A, Correia AS. Risks and mechanisms of oncological disease following stem cell transplantation. Stem Cell Rev 2010;6(3):411–424.
64. Ben-David U, Benvenisty N. The tumorigenicity of human embryonic and induced pluripotent stem cells. Nat Rev Cancer 2011;11(4):268–277.
65. Templin C, Luscher TF, Landmesser U. Cell-based cardiovascular repair and regeneration in acute myocardial infarction and chronic ischemic cardiomyopathy: current status and future developments. Int J Dev Biol 2011;55(4–5):407–417.

66. Barbash IM, Chouraqui P, Baron J, Feinberg MS, Etzion S, Tessone A, et al. Systemic delivery of bone marrow-derived mesenchymal stem cells to the infarcted myocardium: feasibility, cell migration, and body distribution. Circulation 2003;108(7):863–868.
67. Price MJ, Chou CC, Frantzen M, Miyamoto T, Kar S, Lee S, et al. Intravenous mesenchymal stem cell therapy early after reperfused acute myocardial infarction improves left ventricular function and alters electrophysiologic properties. Int J Cardiol 2006;111(2):231–239.
68. Schächinger V, Erbs S, Elsässer A, Haberbosch W, Hambrecht R, Hölschermann H, et al. Repair-AMI Investigators. Improved clinical outcome after intracoronary administration of bone-marrow-derived progenitor cells in acute myocardial infarction: final 1-year results of the REPAIR-AMI trial. Eur Heart J 2006;27(23):2775–2783.
69. Assmus B, Fischer-Rasokat U, Honold J, Seeger FH, Fichtlscherer S, Tonn T, et al. Transcoronary transplantation of functionally competent BMCs is associated with a decrease in natriuretic peptide serum levels and improved survival of patients with chronic postinfarction heart failure. Circ Res 2007;100(8):1234–1241.
70. Lebkowski J. GRNOPC1: The world's first embryonic stem cell-derived therapy. Interview with Jane Lebkowski. Regen Med 2011;6(6 Suppl):11–13.
71. Medina RJ, Archer DB, Stitt AW. Eyes open to stem cells: safety trial may pave the way for cell therapy to treat retinal disease in patients. Stem Cell Res Ther 2011;2(6):47.
72. Lavelle EC, O'Hagan D. Delivery systems and adjuvants for oral vaccines. Expert Opin Drug Deliv 2006;3(6):747–762.
73. Colombo P, Sonvico F, Colombo G, Bettini R. Novel platforms for oral drug delivery. Pharm Res 2009;26(3):601–611.
74. Li Z, Zhang L, Sun W, Ding Q, Hou Y, Xu Y. Archaeosomes with encapsulated antigens for oral vaccine delivery. Vaccine 2011;29(32):5260–5266.
75. Pappo J, Ermak T, Steger H. Monoclonal antibody-directed targeting of fluorescent polystyrene microspheres to Peyer's patch M cells. Immunology 1991;73(3):277.
76. Hussain N, Florence AT. Utilizing bacterial mechanisms of epithelial cell entry: invasin-induced oral uptake of latex nanoparticles. Pharm Res 1998;15(1):153–156.
77. McCluskie MJ, Davis HL. Oral, intrarectal and intranasal immunizations using CPG and non-CPG oligodeoxynucleotides as adjuvants. Vaccine 2000;19(4-5):413–422.
78. Rubinstein A, Tirosh B, Baluom M, Nassar T, David A, Radai R, et al. The rationale for peptide drug delivery to the colon and the potential of polymeric carriers as effective tools. J Control Release 1997;46(1–2):59–73.
79. Vogel K, Kantor J, Wood L, Rivera R, Schlom J. Oral immunization with enterocoated microbeads induces antigen-specific cytolytic T-cell responses. Cell Immunol 1998;190 (1):61–67.
80. Hosny EA, Al-Shora HI, Elmazar M. Oral delivery of insulin from enteric-coated capsules containing sodium salicylate: effect on relative hypoglycemia of diabetic beagle dogs. Int J Pharm 2002;237(1–2):71–76.
81. Bendayan M, Ziv E, Gingras D, Ben-Sasson R, Bar-On H, Kidron M. Biochemical and morpho-cytochemical evidence for the intestinal absorption of insulin in control and diabetic rats. Comparison between the effectiveness of duodenal and colon mucosa. Diabetologia 1994;37(2):119–126.
82. Bastian S, Walton P, Ballard F, Belford D. Transport of IGF-I across epithelial cell monolayers. J Endocrinol 1999;162(3):361.

83. Agarwal V, Khan MA. Current status of the oral delivery of insulin. Pharm Technol 2001;10:76–90.
84. Zhang X, Sun M, Zheng A, Cao D, Bi Y, Sun J. Preparation and characterization of insulin-loaded bioadhesive PLGA nanoparticles for oral administration. Eur J Pharm Sci 2012;45(5):632–638.
85. Pan Y, Li Y, Zhao H, Zheng J, Xu H, Wei G, et al. Bioadhesive polysaccharide in protein delivery system: chitosan nanoparticles improve the intestinal absorption of insulin in vivo. Int J Pharm 2002;249(1–2):139–147.
86. Ma Z, Lim TM, Lim LY. Pharmacological activity of peroral chitosan-insulin nanoparticles in diabetic rats. Int J Pharm 2005;293(1–2):271–280.
87. Su FY, Lin KJ, Sonaje K, Wey SP, Yen TC, Ho YC, et al. Protease inhibition and absorption enhancement by functional nanoparticles for effective oral insulin delivery. Biomaterials 2012;33(9):2801–2811.
88. Aliabadi HM, Lavasanifar A. Polymeric micelles for drug delivery. Expert Opin Drug Deliv 2006;3(1):139–162.
89. Morishita M, Goto T, Nakamura K, Lowman AM, Takayama K, Peppas NA. Novel oral insulin delivery systems based on complexation polymer hydrogels: single and multiple administration studies in type 1 and 2 diabetic rats. J Control Release 2006;110(3):587–594.
90. Morishita M, Peppas NA. Is the oral route possible for peptide and protein drug delivery? Drug Discov Today 2006;11(19–20):905–910.
91. Desai MP, Labhasetwar V, Walter E, Levy RJ, Amidon GL. The mechanism of uptake of biodegradable microparticles in Caco-2 cells is size dependent. Pharm Res 1997;14 (11):1568–1573.
92. Russell-Jones G. Use of targeting agents to increase uptake and localization of drugs to the intestinal epithelium. J Drug Target 2004;12(2):113–123.
93. Liang JF, Yang VC. Insulin-cell penetrating peptide hybrids with improved intestinal absorption efficiency. Biochem Biophys Res Commun 2005;335(3):734–738.
94. Khafagy ES, Morishita M. Oral biodrug delivery using cell-penetrating peptide. Adv Drug Deliv Rev 2012;64(6):531–539.
95. Morishita M, Kamei N, Ehara J, Isowa K, Takayama K. A novel approach using functional peptides for efficient intestinal absorption of insulin. J Control Release 2007;118(2):177–184.
96. Kamei N, Morishita M, Eda Y, Ida N, Nishio R, Takayama K. Usefulness of cell-penetrating peptides to improve intestinal insulin absorption. J Control Release 2008;132 (1):21–25.
97. Kamei N, Morishita M, Ehara J, Takayama K. Permeation characteristics of oligoarginine through intestinal epithelium and its usefulness for intestinal peptide drug delivery. J Control Release 2008;131(2):94–99.
98. Palin K, Phillips A, Ning A. The oral absorption of cefoxitin from oil and emulsion vehicles in rats. Int J Pharm 1986;33(1–3):99–104.
99. Mesiha M, Plakogiannis F, Vejosoth S. Enhanced oral absorption of insulin from desolvated fatty acid-sodium glycocholate emulsions. Int J Pharm 1994;111(3):213–216.
100. Goldberg M, Gomez-Orellana I. Challenges for the oral delivery of macromolecules. Nat Rev Drug Discov 2003;2(4):289–295.

101. Wang Y, Yu L, Han L, Sha X, Fang X. Difunctional pluronic copolymer micelles for paclitaxel delivery: synergistic effect of folate-mediated targeting and pluronic-mediated overcoming multidrug resistance in tumor cell lines. Int J Pharm 2007;337(1–2):63–73.
102. Fasano A, Uzzau S. Modulation of intestinal tight junctions by zonula occludens toxin permits enteral administration of insulin and other macromolecules in an animal model. J Clin Invest 1997;99(6):1158.
103. Müller G. Oral protein therapy for the future—transport of glycolipid-modified proteins: vision or fiction. Pharmacology 2010;86(2):92–116.
104. Del Curto MD, Maroni A, Palugan L, Zema L, Gazzaniga A, Sangalli ME. Oral delivery system for two-pulse colonic release of protein drugs and protease inhibitor/absorption enhancer compounds. J Pharm Sci 2011;100(8):3251–3259.
105. Yamamoto A, Taniguchi T, Rikyuu K, Tsuji T, Fujita T, Murakami M, et al. Effects of various protease inhibitors on the intestinal absorption and degradation of insulin in rats. Pharm Res 1994;11(10):1496–1500.
106. Marschütz MK, Caliceti P, Bernkop-Schnürch A. Design and in vivo evaluation of an oral delivery system for insulin. Pharm Res 2000;17(12):1468–1474.
107. Agarwal V, Nazzal S, Reddy IK, Khan MA. Transport studies of insulin across rat jejunum in the presence of chicken and duck ovomucoids. J Pharm Pharmacol 2001;53 (8):1131–1138.
108. Agarwal V, Reddy IK, Khan MA. Polymethyacrylate based microparticulates of insulin for oral delivery: preparation and in vitro dissolution stability in the presence of enzyme inhibitors. Int J Pharm 2001;225(1–2):31–39.
109. Mertens M, Singh JA. Anakinra for rheumatoid arthritis: a systematic review. The J Rheumatol 2009;36(6):1118–1125.
110. Cryan S-A, Sivadas N, Garcia-Contreras L. In vivo animal models for drug delivery across the lung mucosal. Adv Drug Deliv Rev 2007;59(11):1133–1151.
111. Scheuch G, Siekmeier R. Novel approaches to enhance pulmonary delivery of proteins and peptides. J Physiol Pharmacol 2007;58:615–625.
112. Sung JC, Pulliam BL, Edwards DA. Nanoparticles for drug delivery to the lungs. Trends Biotechnol 2007;25(12):563–570.
113. Hollander PA, Blonde L, Rowe R, Mehta AE, Milburn JL, Hershon KS, et al. Efficacy and safety of inhaled insulin (Exubera) compared with subcutaneous insulin therapy in patients with type 2 diabetes—results of a 6-month, randomized, comparative trial. Diabetes Care 2004;27(10):2356–2362.
114. Patton JS, Bukar JG, Eldon MA. Clinical pharmacokinetics and pharmacodynamics of inhaled insulin. Clin Pharmacokinet 2004;43(12):781–801.
115. Quattrin T, Belanger A, Bohannon NJV, Schwartz SL. Efficacy and safety of inhaled insulin (Exubera) compared with subcutaneous insulin therapy in patients with type 1 diabetes—results of a 6-month, randomized, comparative trial. Diabetes Care 2004;27 (11):2622–2627.
116. Antosova Z, Mackova M, Kral V, Macek T. Therapeutic application of peptides and proteins: parenteral forever? Trends Biotechnol 2009;27(11):628–635.
117. Andrade F, Videira M, Ferreira D, Sarmento B. Nanocarriers for pulmonary administration of peptides and therapeutic proteins. Nanomedicine 2011;6(1):123–141.
118. Kolb M, Martin G, Medina M, Ask K, Gauldie J. Gene therapy for pulmonary diseases. Chest 2006;130(3):879–884.

119. Lam JK-W, Liang W, Chan H-K. Pulmonary delivery of therapeutic siRNA. Adv Drug Deliv Rev 2012;64(1):1–15.
120. Merkel OM, Kissel T. Nonviral pulmonary delivery of sIRNA. Acc Chem Res 2012;45 (7):961–970.
121. Smith DJ, Bot S, Dellamary L, Bot A. Evaluation of novel aerosol formulations designed for mucosal vaccination against influenza virus. Vaccine 2003;21(21–22):2805–2812.
122. McCarthy MW, Kockler DR. Trivalent intranasal influenza vaccine, live. Ann Pharmacother 2004;38(12):2086–2093.
123. Dilraj A, Sukhoo R, Cutts FT, Bennett JV. Aerosol and subcutaneous measles vaccine: measles antibody responses 6 years after re-vaccination. Vaccine 2007;25 (21):4170–4174.
124. Omer SB, Hiremath GS, Halsey NA. Respiratory administration of measles vaccine. Lancet 2010;375(9716):706–708.
125. Bennett JV, de Castro JF, Valdespino-Gomez JL, Garcia-Garcia MD, Islas-Romero R, Echaniz-Aviles G, et al. Aerosolized measles and measles-rubella vaccines induce better measles antibody booster responses than injected vaccines: randomized trials in Mexican schoolchildren. Bull WHO 2002;80(10):806–812.
126. de Castro JF, Bennett JV, Rincon HG, Munoz M, Sanchez L, Santos JI. Evaluation of immunogenicity and side effects of triple viral vaccine (MMR) in adults, given by two routes: subcutaneous and respiratory (aerosol). Vaccine 2005;23(8):1079–1084.
127. Burger JL, Cape SP, Braun CS, McAdams DH, Best JA, Bhagwat P, et al. Stabilizing formulations for inhalable powders of live-attenuated measles virus vaccine. J Med Pulm Drug Deliv 2008;21(1):25–34.
128. Conlan JW, Shen H, KuoLee R, Zhao XG, Chen WC, et al. Aerosol-, but not intradermal-immunization with the live vaccine strain of *Francisella tularensis* protects mice against subsequent aerosol challenge with a highly virulent type a strain of the pathogen by an alpha beta T cell- and interferon gamma-dependent mechanism. Vaccine 2005; 23(19):2477–2485.
129. Huang J, Mikszta JA, Ferriteri MS, Jiang G, Harvey NG, Dyas B, et al. Intranasal administration of dry powder anthrax vaccine provides protection against lethal aerosol spore challenge. Hum Vaccine 2007;3(3):90–93.
130. Twigg HL. Pulmonary host defenses. J Thorac Imaging 1998;13(4):221–233.
131. Nicod LP. Pulmonary defence mechanisms. Respiration 1999;66(1):2–11.
132. Bhat PG, Flanagan DR, Donovan MD. The limiting role of mucus in drug absorption: drug permeation through mucus solution. Int J Pharm 1995;126(1–2):179–187.
133. Rubin BK. Therapeutic aerosols and airway secretions. J Aerosol Med Pulm Drug Deliv 1996;9(1):123–130.
134. Wall DA, Lanutti AT. High levels of exopeptidase activity are present in rat and canine bronchoalveolar lavage fluid. Int J Pharm 1993;97(1–3):171–181.
135. Evans JP, Tudball N, Dickinson PA, Farr SJ, Kellaway IW. Transport of a series of D-phenylalanine-glycine hexapeptides across rat alveolar epithelia in vitro. J Drug Target 1998;6(4):251–259.
136. Patton JS, Nagarajan S, Clark A. Pulmonary absorption and metabolism of peptides and proteins. Respir Drug Deliv 1998;VI(1):17–24.
137. Adjei LA, Carrigan PJ. Pulmonary bioavailability of Lh-Rh analogs: some pharmaceutical guidelines. J Biopharm Sci 1992;3247–254.

138. Patton JS, McCabe JG, Hansen SE, Daugherty AL. Absorption of human growth hormone from the rat lung. Biotechnol Ther 1989;1(3):213–228.
139. Lombry C, Edwards DA, Preat V, Vanbever R. Alveolar macrophages are a primary barrier to pulmonary absorption of macromolecules. Am J Physiol Lung Cell Mol Physiol 2004;286(5):L1002–L1008.
140. Patton JS, Byron PR. Inhaling medicines: delivering drugs to the body through the lungs. Nat Rev Drug Discov 2007;6(1):67–74.
141. Summers QA. Inhaled Drugs and the lung. Clin Exp Allergy 1991;21(3):259–268.
142. Cryan SA. Carrier-based strategies for targeting protein and peptide drugs to the lungs. AAPS J 2005;7(1):E20–E41.
143. Agu RU, Ugwoke MI, Armand M, Kinget R, Verbeke N. The lung as a route for systemic delivery of therapeutic proteins and peptides. Respir Res 2001;2(4):198–209.
144. Smith J, Wood E, Dornish M. Effect of chitosan on epithelial cell tight junctions. Pharm Res 2004;21(1):43–49.
145. Salama NN, Eddington ND, Fasano A. Tight junction modulation and its relationship to drug delivery. Adv Drug Deliv Rev 2006;58(1):15–28.
146. Heyder J, Gebhart J, Rudolf G, Schiller CF, Stahlhofen W. Deposition of particles in the human respiratory tract in the size range 0.005–15 μm. J Aerosol Sci 1986;17(5):811–825.
147. Taylor KMG, McCallion ONM. Ultrasonic nebulisers for pulmonary drug delivery. Int J Pharm 1997;153(1):93–104.
148. Labiris NR, Dolovich MB. Pulmonary drug delivery. Part II. The role of inhalant delivery devices and drug formulations in therapeutic effectiveness of aerosolized medications. Br J Clin Pharmacol 2003;56(6):600–612.
149. Yeo LY, Friend JR, McIntosh MP, Meeusen ENT, Morton DAV. Ultrasonic nebulization platforms for pulmonary drug delivery. Expert Opin Drug Deliv 2010;7(6):663–679.
150. Frijlink HW, de Boer AH. Trends in the technology driven development of new inhalation devices. Drug Discov Today Technol 2005;2(1):47–57.
151. Ip AY, Arakawa T, Silvers H, Ransone CM, Niven RW. Stability of recombinant consensus interferon to air-jet and ultrasonic nebulization. J Pharm Sci 1995;84 (10):1210–1214.
152. Cipolla DC, Clark AR, Chan HK, Gonda I, Shire SJ. Assessment of aerosol delivery systems for recombinant human deoxyribonuclease. STP Pharma Sci 1994;4(1):50–62.
153. Lentz YK, Anchordoquy TJ, Lengsfeld CS. Rationale for the selection of an aerosol delivery system for gene delivery. J Aerosol Med 2006;19(3):372–384.
154. Birchall J. Pulmonary delivery of nucleic acids. Expert Opin Drug Deliv 2007; 4(6):575–578.
155. Quinn EA, Forbes RT, Williams AC, Oliver MJ, McKenzie L, Purewal TS. Protein conformational stability in the hydrofluoroalkane propellants tetrafluoroethane and heptafluoropropane analysed by Fourier transform Raman spectroscopy. Int J Pharm 1999;186(1):31–41.
156. Benfait C. Kos Reports Achievement of New Research and Development Milestones. Kos Press Release; 2004.
157. Timsina MP, Martin GP, Marriott C, Ganderton D, Yianneskis M. Drug delivery to the respiratory tract using dry powder inhalers. Int J Pharm 1994;101(1–2):1–13.

158. Niven R. Dry powder formulations for inhalation: preface. Adv Drug Deliv Rev 1997;26 (1):1–2.
159. Codrons V, Vanderbist F, Verbeeck RK, Arras M, Lison D, Preat V, et al. Systemic delivery of parathyroid hormone (1-34) using inhalation dry powders in rats. J Pharm Sci 2003;92(5):938–950.
160. Daniher DI, Zhu J. Dry powder platform for pulmonary drug delivery. Particuology 2008;6(4):225–238.
161. Van Eerdenbrugh B, Van den Mooter G, Augustijns P. Top-down production of drug nanocrystals: nanosuspension stabilization, miniaturization and transformation into solid products. Int J Pharm 2008;364(1):64–75.
162. Tam JM, Engstrom JD, Ferrer D, Williams RO III, Johnston KP. Templated open flocs of anisotropic particles for pulmonary delivery with pressurized metered dose inhalers. J Pharm Sci 2010;99(7):3150–3165.
163. Rogueda PGA, Traini D. The nanoscale in pulmonary delivery. Part 2. Formulation platforms. Expert Opin Drug Deliv 2007;4(6):607–620.
164. Chan HK, Clark A, Gonda I, Mumenthaler M, Hsu C. Spray dried powders and powder blends of recombinant human deoxyribonuclease (RHDnase) for aerosol delivery. Pharm Res 1997;14(4):431–437.
165. Lucas P, Anderson K, Staniforth JN. Protein deposition from dry powder inhalers: fine particle multiplets as performance modifiers. Pharm Res 1998;15(4):562–569.
166. Illum L. Nasal drug delivery: possibilities, problems and solutions. J Control Release 2003;87(1–3):187–198.
167. Graff CL, Pollack GM. Nasal drug administration: potential for targeted central nervous system delivery. J Pharm Sci 2005;94(6):1187–1195.
168. Arora P, Sharma S, Garg S. Permeability issues in nasal drug delivery. Drug Discovery Today 2002;7(18):967–975.
169. Rapoport A, Winner P. Nasal delivery of antimigraine drugs: clinical rationale and evidence base. Headache 2006;46:S192–S201.
170. Costantino HR, Illum L, Brandt G, Johnson PH, Quay SC. Intranasal delivery: physicochemical and therapeutic aspects. Int J Pharm 2007;337(1–2):1–24.
171. Pires A, Fortuna A, Alves G, Falcao A. Intranasal drug delivery: how, why and what for? J Pharm Pharm Sci 2009;12(3):288–311.
172. Bitter C, Suter-Zimmermann K, Surber C. Nasal drug delivery in humans. Curr Probl Dermatol 2011;40:20–35.
173. Slutter B, Hagenaars N, Jiskoot W. 2008. Rational design of nasal vaccines. J Drug Target 16(1):1–17.
174. Brave A, Hallengard D, Schroder U, Blomberg P, Wahren B, Hinkula J. Intranasal immunization of young mice with a multigene HIV-1 vaccine in combination with the N3 adjuvant induces mucosal and systemic immune responses. Vaccine 2008;26(40):5075–5078.
175. Borges O, Cordeiro-da-Silva A, Tavares J, Santarem N, de Sousa A, Borchard G, et al. Immune response by nasal delivery of hepatitis B surface antigen and codelivery of a CPG ODN in alginate coated chitosan nanoparticles. Eur J Pharm Biopharm 2008;69 (2):405–416.
176. Csaba N, Garcia-Fuentes M, Alonso MJ. Nanoparticles for nasal vaccination. Adv Drug Deliv Rev 2009;61(2):140–157.

177. Dufes C, Olivier JC, Gaillard F, Gaillard A, Couet W, Muller JM. Brain delivery of vasoactive intestinal peptide (VIP) following nasal administration to rats. Int J Pharm 2003;255(1–2):87–97.
178. Banks WA, During MJ, Niehoff ML. Brain uptake of the glucagon-like peptide-1 antagonist exendin(9-39) after intranasal administration. J Pharmacol Exp Ther 2004;309(2):469–475.
179. Ross TM, Martinez PM, Renner JC, Thorne RG, Hanson LR, Frey WH. Intranasal administration of interferon beta bypasses the blood-brain barrier to target the central nervous system and cervical lymph nodes: a non-invasive treatment strategy for multiple sclerosis. J Neuroimmunol 2004;151(1–2):66–77.
180. Thorne RG, Pronk GJ, Padmanabhan V, Frey WH. Delivery of insulin-like growth factor-I to the rat brain and spinal cord along olfactory and trigeminal pathways following intranasal administration. Neuroscience 2004;127(2):481–496.
181. Banks WA, Morley JE, Niehoff ML, Mattern C. Delivery of testosterone to the brain by intranasal administration: comparison to intravenous testosterone. J Drug Target 2009;17 (2):91–97.
182. van den Berg MP, Merkus P, Romeijn SG, Verhoef JC, Merkus F. Uptake of melatonin into the cerebrospinal fluid after nasal and intravenous delivery: studies in rats and comparison with a human study. Pharm Res 2004;21(5):799–802.
183. van den Berg MP, Verhoef JC, Romeijn SG, Merkus F. Uptake of estradiol or progesterone into the CSF following intranasal and intravenous delivery in rats. Eur J Pharm Biopharm 2004;58(1):131–135.
184. Yang Z, Huang Y, Gan GF, Sawchuk RJ. Microdialysis evaluation of the brain distribution of stavudine following intranasal and intravenous administration to rats. J Pharm Sci 2005;94(7):1577–1588.
185. Reger MA, Watson GS, Green PS, Baker LD, Cholerton B, Fishel MA, et al. Intranasal insulin administration dose-dependently modulates verbal memory and plasma amyloid-beta in memory-impaired older adults. J Alzheimers Dis 2008;13(3):323–331.
186. Hallschmid M, Benedict C, Schultes B, Perras B, Fehm H-L, Kern W, et al. Towards the therapeutic use of intranasal neuropeptide administration in metabolic and cognitive disorders. Regul Peptides 2008;149(1-3):79–83.
187. Greimel A, Bernkop-Schnuerch A, Del Curto MD, D'Antonio M. Transport characteristics of a beta sheet breaker peptide across excised bovine nasal mucosa. Drug Dev Ind Pharm 2007;33(1):71–77.
188. Watanabe Y, Mizufune Y, Utoguchi N, Endo K, Matsumoto M. Studies of drug delivery systems for a therapeutic agent used in osteoporosis. II. Enhanced absorption of elcatonin from nasal mucosa in rabbits. Biol Pharm Bull 1998;21(11):1191–1194.
189. Na DH, Youn YS, Park EJ, Lee JM, Cho OR, Lee KR, et al. Stability of PEGylated salmon calcitonin in nasal mucosa. J Pharm Sci 2004;93(2):256–261.
190. Ugwoke MI, Agu RU, Verbeke N, Kinget R. Nasal mucoadhesive drug delivery: background, applications, trends and future perspectives. Adv Drug Deliv Rev 2005;57(11):1640–1665.
191. Law SL, Huang KJ, Chou VHY, Cherng JY. Enhancement of nasal absorption of calcitonin loaded in liposomes. J Liposome Res 2001;11(2-3):165–174.
192. Jain AK, Chalasani KB, Khar RK, Ahmed FJ, Diwan PV. Muco-adhesive multivesicular liposomes as an effective carrier for transmucosal insulin delivery. J Drug Target 2007;15 (6):417–427.

193. Wang J, Tabata Y, Morimoto K. Aminated gelatin microspheres as a nasal delivery system for peptide drugs: evaluation of in vitro release and in vivo insulin absorption in rats. J Control Release 2006;113(1):31–37.
194. Jung T, Kamm W, Breitenbach A, Hungerer KD, Hundt E, Kissel T. Tetanus toroid loaded nanoparticles from sulfobutylated poly(vinyl alcohol)-graft-poly(lactide-co-glycolide): evaluation of antibody response after oral and nasal application in mice. Pharm Res 2001;18(3):352–360.
195. Debin A, Kravtzoff R, Santiago JV, Cazales L, Sperandio S, Melber K, et al. Intranasal immunization with recombinant antigens associated with new cationic particles induces strong mucosal as well as systemic antibody and CTL responses. Vaccine 2002;20(21–22):2752–2763.
196. Nagamoto T, Hattori Y, Takayama K, Maitani Y. Novel chitosan particles and chitosan-coated emulsions inducing immune response via intranasal vaccine delivery. Pharm Res 2004;21(4):671–674.
197. Prausnitz MR, Langer R. Transdermal drug delivery. Nat Biotechnol 2008;26(11):1261–1268.
198. Elias PM, Friend DS. Permeability barrier in mammalian epidermis. J Cell Biol 1975;65 (1):180–191.
199. Elias PM, Menon GK. Structural and lipid biochemical correlates of the epidermal permeability barrier. Adv Lipid Res 1991;24:1–26.
200. Bouwstra JA, Honeywell-Nguyen PL, Gooris GS, Ponec M. Structure of the skin barrier and its modulation by vesicular formulations. Prog Lipid Res 2003;42(1):1–36.
201. Marrink SJ, Berendsen HJC. Permeation process of small molecules across lipid membranes studied by molecular dynamics simulations. J Phys Chem 1996;100 (41):16729–16738.
202. Potts RO, Guy RH. Predicting skin permeability. Pharm Res 1992;9(5):663–669.
203. Prausnitz MR, Mitragotri S, Langer R. Current status and future potential of transdermal drug delivery. Nat Rev Drug Discov 2004;3(2):115–124.
204. Frech SA, DuPont HL, Bourgeois AL, McKenzie R, Belkind-Gerson J, Figueroa JF, et al. Use of a patch containing heat-labile toxin from *Escherichia coli* against travellers' diarrhoea: a phase II, randomised, double-blind, placebo-controlled field trial. Lancet 2008;371(9629):2019–2025.
205. El-Kamary SS, Billington M, Deitz S, Colby E, Rhinehart H, Wu Y, et al. Safety and tolerability of the EasyVax™ clinical epidermal electroporation system in healthy adults. Mol Ther 2012;20(1):214–220.
206. Williams AC, Barry BW. Skin absorption enhancers. Crit Ref Ther Drug Carrier Syst 1992;9(3-4):305–353.
207. Finnin BC, Morgan TM. Transdermal penetration enhancers: applications, limitations, and potential. J Pharm Sci 1999;88(10):955–958.
208. Williams AC, Barry BW. Penetration enhancers. Adv Drug Deliv Rev 2004;56(5):603–618.
209. Ongpipattanakul B, Burnette RR, Potts RO, Francoeur ML. Evidence that oleic acid exists in a separate phase within stratum corneum lipids. Pharm Res 1991;8(3):350–354.
210. Takanashi Y, Higashiyama K, Komiya H, Takayama K, Nagai T. Thiomenthol derivatives as novel percutaneous absorption enhancers. Drug Dev Ind Pharm 1999;25(1):89–94.
211. Akimoto T, Nagase Y. Novel transdermal drug penetration enhancer: synthesis and enhancing effect of alkyldisiloxane compounds containing glucopyranosyl group. J Control Release 2003;88(2):243–252.

212. Karande P, Jain A, Mitragotri S. Discovery of transdermal penetration enhancers by high-throughput screening. Nat Biotechnol 2004;22(2):192–197.
213. Setoh K, Murakami M, Araki N, Fujita T, Yamamoto A, Muranishi S. Improvement of transdermal delivery of tetragastrin by lipophilic modification with fatty acids. J Pharm Pharmacol 1995;47(10):808–811.
214. Kim YC, Ludovice PJ, Prausnitz MR. Transdermal delivery enhanced by antimicrobial peptides. J Biomed Nanotechnol 2010;6(5):612–620.
215. Hsu T, Mitragotri S. Delivery of sIRNA and other macromolecules into skin and cells using a peptide enhancer. Proc Natl Acad Sci USA 2011;108(38):15816–15821.
216. Chen YP, Shen YY, Guo X, Zhang CS, Yang WJ, Ma ML, et al. Transdermal protein delivery by a coadministered peptide identified via phage display. Nat Biotechnol 2006;24(4):455–460.
217. Kim Y-C, Ludovice PJ, Prausnitz MR. Transdermal delivery enhanced by magainin pore-forming peptide. J Control Release 2007;122(3):375–383.
218. Li YZ, Quan YS, Zang L, Jin MN, Kamiyama F, Katsumi H, et al. Transdermal delivery of insulin using trypsin as a biochemical enhancer. Biol Pharm Bull 2008;31 (8):1574–1579.
219. Kogan A, Garti N. Microemulsions as transdermal drug delivery vehicles. Adv Colloid Interface Sci 2006;123369–385.
220. Cosco D, Celia C, Cilurzo F, Trapasso E, Paolino D. Colloidal carriers for the enhanced delivery through the skin. Expert Opin Drug Deliv 2008;5(7):737–755.
221. Sinico C, Fadda AM. Vesicular carriers for dermal drug delivery. Expert Opin Drug Deliv 2009;6(8):813–825.
222. Huang Y, Yu F, Park Y-S, Wang J, Shin M-C, Chung HS, et al. Co-administration of protein drugs with gold nanoparticles to enable percutaneous delivery. Biomaterials 2010;31(34):9086–9091.
223. Lin MTS, Pulkkinen L, Uitto J, Yoon K. The gene gun: current applications in cutaneous gene therapy. Int J Dermatol 2000;39(3):161–170.
224. Bremseth DL, Pass F. Delivery of insulin by jet injection: recent observations. Diabetes Technol Ther 2001;3(2):225–232.
225. Tezel A, Paliwal S, Shen ZC, Mitragotri S. Low-frequency ultrasound as a transcutaneous immunization adjuvant. Vaccine 2005;23(29):3800–3807.
226. Bareille P, MacSwiney M, Albanese A, DeVile C, Stanhope R. Growth hormone treatment without a needle using the Preci-Jet 50 transjector. Arch Dis Child 1997;76(1):65–67.
227. Suzuki T, Takahashi I, Takada G. Daily subcutaneous erythropoietin by jet injection in pediatric dialysis patients. Nephron 1995;69(3):347–347.
228. Brodell RT, Bredle DL. The treatment of palmar and plantar warts using natural alpha-interferon and a needleless injector. Dermatol Surg 1995;21(3):213–218.
229. Weller C, Linder M. Jet injection of insulin vs syringe-and-needle method. JAMA 1966;195(10):844.
230. Lindmayer I, Menassa K, Lambert J, Moghrabi A, Legendre L, Legault C, et al. Development of new jet injector for insulin therapy. Diabetes Care 1986;9(3):294–297.
231. Verhagen A, Ebels JT, Dogterom AA, Jonkman JHG. Pharmacokinetics and pharmacodynamics of a single-dose of recombinant human growth-hormone after subcutaneous

administration by jet-injection: comparison with conventional needle-injection. Eur J Clin Pharmacol 1995;49(1–2):69–72.

232. Hollingsworth SJ, Hoque K, Linnard D, Corry DG, Barker SGE. Delivery of low molecular weight heparin for prophylaxis against deep vein thrombosis using a novel, needle-less injection device (J-Tip®). Ann R Coll Surg Engl 2000;82(6):428–431.
233. Vadoud-Seyedi J. Treatment of plantar hyperhidrosis with botulinum toxin type A. Int J Dermatol 2004;43(12):969–971.
234. Weniger BG. Jet injection of vaccines: overview and challenges for mass vaccination with jet injections (JIS). Available from: http://www.hhs.gov/nvpo/meetings/dec2003/Contents/ThursdayPM/Weniger.pdf.
235. Mitragotri S. Innovation: current status and future prospects of needle-free liquid jet injectors. Nat Rev Drug Discov 2006;5(7):543–548.
236. Kendall M. Engineering of needle-free physical methods to target epidermal cells for DNA vaccination. Vaccine 2006;24(21):4651–4656.
237. Chen DX, Periwal SB, Larrivee K, Zuleger C, Erickson CA, Endres RL, et al. Serum and mucosal immune responses to an inactivated influenza virus vaccine induced by epidermal powder immunization. J Virol 2001;75(17):7956–7965.
238. Fuller DH, Rajakumar PA, Wilson LA, Trichel AM, Fuller JT, Shipley T, et al. Induction of mucosal protection against primary, heterologous simian immunodeficiency virus by a DNA vaccine. J Virol 2002;76(7):3309–3317.
239. Tang DC, Devit M, Johnston SA. Genetic immunization is a simple method for eliciting an immune response. Nature 1992;356(6365):152–154.
240. Fynan EF, Webster RG, Fuller DH, Haynes JR, Santoro JC, Robinson HL. DNA vaccines: protective immunizations by parenteral, mucosal, and gene-gun inoculations. Proc Natl Acad Sci USA 1993;90(24):11478–11482.
241. Degano P, Sarphie DF, Bangham CRM. Intradermal DNA immunization of mice against influenza a virus using the novel Powder Ject® system. Vaccine 1998; 16(4):394–398.
242. Roy MJ, Wu MS, Barr LJ, Fuller JT, Tussey LG, Speller S, et al. Induction of antigen-specific $CD8^+$ T cells, T helper cells, and protective levels of antibody in humans by particle-mediated administration of a hepatitis B virus DNA vaccine. Vaccine 2000; 19(7–8):764–778.
243. Chen DX, Weis KF, Chu QL, Erickson C, Endres R, Lively CR, et al. Epidermal powder immunization induces both cytotoxic T-lymphocyte and antibody responses to protein antigens of influenza and hepatitis B viruses. J Virol 2001;75(23):11630–11640.
244. Fuller DH, Loudon P, Schmaljohn C. Preclinical and clinical progress of particle-mediated DNA vaccines for infectious diseases. Methods 2006;40(1):86–97.
245. Drape RJ, Macklin MD, Barr LJ, Jones S, Haynes JR, Dean HJ. Epidermal DNA vaccine for influenza is immunogenic in humans. Vaccine 2006;24(21):4475–4481.
246. Byl NN. The use of ultrasound as an enhancer for transcutaneous drug delivery: phonophoresis. Phys Ther 1995;75(6):539–553.
247. Simonin JP. On the mechanisms of in-vitro and in-vivo phonophoresis. J Control Release 1995;33(1):125–141.
248. Mitragotri S, Edwards DA, Blankschtein D, Langer R. Mechanistic study of ultrasonically-enhanced transdermal drug delivery. J Pharm Sci 1995;84(6):697–706.

249. Kushner J, Blankschtein D, Langer R. Experimental demonstration of the existence of highly permeable localize transport regions in low-frequency sonophoresis. J Pharm Sci 2004;93(11):2733–2745.

250. Polat BE, Blankschtein D, Langer R. Low-frequency sonophoresis: application to the transdermal delivery of macromolecules and hydrophilic drugs. Expert Opin Drug Deliv 2010;7(12):1415–1432.

251. Mitragotri S, Kost J. Low-frequency sonophoresis: a review. Adv Drug Deliv Rev 2004;56(5):589–601.

252. Prausnitz MR. Microneedles for transdermal drug delivery. Adv Drug Deliv Rev 2004;56 (5):581–587.

253. Gill HS, Prausnitz MR. Coated microneedles for transdermal delivery. J Control Release 2007;117(2):227–237.

254. Park JH, Allen MG, Prausnitz MR. Biodegradable polymer microneedles: fabrication, mechanics and transdermal drug delivery. J Control Release 2005;104(1):51–66.

255. Park JH, Allen MG, Prausnitz MR. Polymer microneedles for controlled-release drug delivery. Pharm Res 2006;23(5):1008–1019.

256. Lee JW, Park J-H, Prausnitz MR. Dissolving microneedles for transdermal drug delivery. Biomaterials 2008;29(13):2113–2124.

257. McAllister DV, Wang PM, Davis SP, Park JH, Canatella PJ, Allen MG, et al. Microfabricated needles for transdermal delivery of macromolecules and nanoparticles: fabrication methods and transport studies. Proc Natl Acad Sci USA 2003;100 (24):13755–13760.

258. Harvey AJ, Kaestner SA, Sutter DE, Harvey NG, Mikszta JA, Pettis RJ. Microneedle-based intradermal delivery enables rapid lymphatic uptake and distribution of protein drugs. Pharm Res 2011;28(1):107–116.

259. Dean CH, Alarcon JB, Waterston AM, Draper K, Early R, Guirakhoo F, et al. Cutaneous delivery of a live, attenuated chimeric flavivirus vaccine against Japanese encephalitis (ChimeriVax™-JE) in non-human primates. Hum Vaccine 2005;1(3):106–111.

260. Mikszta JA, Sullivan VJ, Dean C, Waterston AM, Alarcon JB, Dekker JP, et al. Protective immunization against inhalational anthrax: a comparison of minimally invasive delivery platforms. J Infect Dis 2005;191(2):278–288.

261. Mikszta JA, Dekker J.P.III, Harvey NG, Dean CH, Brittingham JM, Huang J, et al. Microneedle-based intradermal delivery of the anthrax recombinant protective antigen vaccine. Infect Immun 2006;74(12):6806–6810.

262. Alarcon JB, Hartley AW, Harvey NG, Mikszta JA. Preclinical evaluation of microneedle technology for intradermal delivery of influenza vaccines. Clin Vaccine Immunol 2007;14(4):375–381.

263. Morefield GL, Tammariello RF, Purcell BK, Worsham PL, Chapman J, et al. An alternative approach to combination vaccines: intradermal administration of isolated components for control of anthrax, botulism, plague and staphylococcal toxic shock. J Immune Based Ther Vaccines 2008;6:5.

264. Huang J, D'Souza AJ, Alarcon JB, Mikszta JA, Ford BM, Ferriter MS, et al. Protective immunity in mice achieved with dry powder formulation and alternative delivery of plague F1-V vaccine. Clin Vaccine Immunol 2009;16(5):719–725.

265. Holland D, Booy R, De Looze F, Eizenberg P, McDonald J, Karrasch J, et al. Intradermal influenza vaccine administered using a new microinjection system produces superior

immunogenicity in elderly adults: a randomized controlled trial. J Infect Dis 2008;198 (5):650–658.

266. Leroux-Roels I, Vets E, Freese R, Seiberling M, Weber F, Salamand C, et al. Seasonal influenza vaccine delivered by intradermal microinjection: a randomised controlled safety and immunogenicity trial in adults. Vaccine 2008;26(51):6614–6619.
267. Saville M, Marsh G, Hoffenbach A. Improving seasonal and pandemic influenza vaccines. Influenza Other Respi Viruses 2008;2(6):229–235.
268. Beran J, Ambrozaitis A, Laiskonis A, Mickuviene N, Bacart P, Calozet Yet al. Intradermal influenza vaccination of healthy adults using a new microinjection system: a 3-year randomised controlled safety and immunogenicity trial. BMC Med 2009;7:13.
269. Intanza®, the first intradermal influenza vaccine, receives the European marketing authorisation. Available from: http://www.medicalnewstoday.com/articles/140580.php.
270. Daddona PE, Matriano JA, Mandema J, Maa Y-F. Parathyroid hormone (1-34)-coated microneedle patch system: clinical pharmacokinetics and pharmacodynamics for treatment of osteoporosis. Pharm Res 2011;28(1):159–165.
271. Peters EE, Ameri M, Wang X, Maa Y-F, Daddona PE. Erythropoietin-coated ZP-microneedle transdermal system: preclinical formulation, stability, and delivery. Pharm Res 2012;29(6):1618–1626.
272. Matriano JA, Cormier M, Johnson J, Young WA, Buttery M, Nyam K, et al. Macroflux® microprojection array patch technology: a new and efficient approach for intracutaneous immunization. Pharm Res 2002;19(1):63–70.
273. Cormier M, Johnson B, Ameri M, Nyam K, Libiran L, Zhang DD, et al. Transdermal delivery of desmopressin using a coated microneedle array patch system. J Control Release 2004;97(3):503–511.
274. Sathyan G, Sun Y-N, Weyers R, Daddona P, Staehr P, Gupta S. Macroflux® desmopressin transdermal delivery system: pharmacokinetics and pharmacodynamic evaluation in healthy volunteers. AAPS J 2004;6(S1):665.
275. Widera G, Johnson J, Kim L, Libiran L, Nyam K, Daddona PE, et al. Effect of delivery parameters on immunization to ovalbumin following intracutaneous administration by a coated microneedle array patch system. Vaccine 2006;24(10):1653–1664.